

Nadica Miljković

Analiza biosignala sa praktičnim primerima u programskom jeziku R

Univerzitet u Beogradu – Elektrotehnički fakultet
Beograd 2024.

Nadica Miljković

Analiza biosignala sa praktičnim primerima u programskom jeziku R
elektronski udžbenik

Recenzenti

Dr Predrag Pejović, redovni profesor, Univerzitet u Beogradu
Dr Nenad Filipović, redovni profesor, Univerzitet u Kragujevcu
Dr Miloš Ajčević, docent, Univerzitet u Trstu

Nastavno-naučno veće Elektrotehničkog fakulteta odobrilo je objavljivanje
ove nastavne literature odlukom br. 1275/7 od 2. jula 2024. godine.

Izdavač

Univerzitet u Beogradu – Elektrotehnički fakultet
Bulevar kralja Aleksandra 73, 11000 Beograd

Štampa

Nadica Miljković, Beograd, 2024

Tiraž

50 primeraka (CD)

ISBN 978-86-7225-098-5

NAPOMENA: Elektronski udžbenik je podeljen pod otvorenom licencom
Kreativne zajednice (eng. *Creative Commons*) [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) (Autorstvo,
međunarodna licenca), osim u delovima gde je drugačije naznačeno.
Umnožavanje i fotokopiranje udžbenika je dozvoljeno bez prethodne
saglasnosti i bez pismenog odobrenja Autorke i izdavača.



Sadržaj

Predgovor.....	6
1 Osnovni pojmovi i korisne informacije za rad u programskom jeziku R.....	9
1.1 Tipovi podataka u R-u.....	26
1.1.1 O nedostajućim vrednostima.....	37
1.1.2 Uvoz podataka u R.....	51
1.2 Operatori i kontrolne strukture u R-u.....	72
1.2.1 Kontrolne funkcije.....	82
1.3 Funkcije u R-u.....	87
1.3.1 Leksička pravila pretrage argumenata.....	93
2 Automatske metode za analizu biosignala.....	96
2.1 Uticaj subjektivnih faktora na automatsku analizu biosignala.....	102
2.1.1 Preregistracija kao primer dobre prakse.....	103
2.1.2 HARKovanje kao primer loše prakse.....	105
2.1.3 Kvalitativni pristup je važniji od kvantitativnog?.....	106
2.1.4 Pristrasnost u razumevanju rezultata.....	109
2.2 Pretprocesiranje biosignala za eliminaciju šuma.....	110
2.2.1 Merenje šuma.....	111
2.2.1 Priprema podataka za dalju obradu.....	114
2.2.1.1 Primer pripreme podataka za merenje opsega pokreta.....	115
2.2.1.2 Primena dplyr paketa za pripremu podataka za dalju obradu	122
2.2.1.3 Drugi primeri pripreme podataka za dalju obradu.....	138
2.2.1.4 Kratka istorija merenja pokreta.....	146
2.2.2 Eliminacija šuma na primeru EMG biosignala.....	149
2.2.2.1 Distorzija faze signala usled filtriranja.....	163
2.2.2.2 Smernice za odabir filtra.....	166
2.2.2.3 Primena Batervortovog filtra za eliminaciju šuma.....	174
2.2.2.4 Primena MA filtra za računanje obvojnice.....	176
2.2.3 Normalizacija EMG biosignala.....	190
2.3 Izdvajanje karakterističnih obeležja.....	194
2.3.2 Izdvajanje karakterističnih obeležja iz EMG signala.....	202
2.3.2.1 Određivanje početka i kraja mišićne kontrakcije.....	203
2.3.2.2 Određivanje medijana frekvencije EMG signala.....	209
2.3.2 Delineacija u obradi EKG biosignala za izdvajanje karakterističnih obeležja.....	214
2.3.3 Završna razmatranja o izdvajanju obeležja.....	222
3 Simulirani biosignali.....	225
3.1 Generisanje pseudoslučajnih brojeva u R-u.....	227
3.1.1 Generisanje belog i obojenog šuma.....	233

3.2 Generisanje simuliranih EMG signala.....	237
3.2.1 Osobine EMG signala.....	239
3.2.2 Generisanje EMG biosignala.....	250
3.2.3 (Ne)stacionarnost simuliranih biosignala.....	255
4 Eksplorativna analiza biosignala.....	261
4.1 Vizuelizacija signala.....	264
4.1.1 Istorijski primeri vizuelizacije podataka.....	266
4.1.2 Osnovne funkcije za prikaz grafika u R-u.....	271
4.1.3 Preporuke za vizuelizaciju biosignala.....	277
4.1.3.1 Primeri dobre prakse u vizuelizaciji biosignala.....	283
4.1.3.2 Primeri loše prakse u vizuelizaciji biosignala.....	287
4.1.4 Primeri grafika u ggplot paketu.....	290
4.1.4.1 Prikaz vremenskih serija u ggplot paketu.....	308
4.1.4.2 Prikaz zavisnosti biomarkera u ggplot paketu.....	313
4.1.4.3 Prikaz kros-korelacionih matrica za ocenu preslušavanja biosignala i interaktivni grafici.....	318
4.1.5 Odabrani primeri grafika i drugi R paketi.....	323
4.1.5.1 Toplotne mape (eng. heat map).....	325
4.1.5.2 Spektrogram biosignala.....	330
4.1.5.3 Grafik vremenske linije.....	334
5 Statistička analiza biosignala.....	336
5.1 Analiza podataka na primeru studije spavanja.....	337
5.2 Osnovni proračun t-testa na primeru podataka o težini biljaka.....	351
5.2.1 Nulta i alternativna hipoteza – praktično objašnjenje.....	360
5.3 Analiza varijanse na primeru podataka o težini biljaka.....	362
5.4 Praktično objašnjenje p vrednosti i problem sa p vrednošću.....	364
6 Dobre prakse u analizi biosignala i u radu u programskom jeziku R.....	372
6.1 Otvoreni podaci i slobodan softver.....	372
6.1.1 Osnovni elementi deljenja softverskog koda.....	373
6.1.2 Otvoreni biosignali: pronalaženje, ponovno korišćenje i deljenje.....	380
6.2 Kako izbeći greške?.....	383
6.3 Generisanje automatskih izveštaja.....	390
6.4 Estetika softverskog koda.....	395
6.5 Indeksiranje u R-u.....	399
7 Umesto zaključka.....	402
7.1 Veštačka inteligencija kao pomoć u programiranju.....	403
7.2 Budući pravci u analizi biosignala.....	407
Literatura.....	410
Dodatak A – Spisak lekcija iz TOBS predmeta.....	447
Dodatak B – Spisak vežbi iz TOBS predmeta.....	449
Dodatak C – R Studio razvojno okruženje.....	451

Instalacija (Ubuntu operativni sistem).....	451
Instalacija (Vindovs operativni sistem).....	453
Kratko upustvo za rad u R Studio razvojnom okruženju.....	458

Predgovor

Udžbenik je namenjen studentkinjama i studentima koji slušaju predmet Tehnike obrade biomedicinskih signala na master – akademskim studijama Elektrotehničkog fakulteta, Univerziteta u Beogradu (ETF). Takođe, delovi udžbenika sadrže i elemente koji mogu biti od koristi za predmete Metode analize elektrofizioloških signala i Kliničko inženjerstvo koji se nude studentima na 4. godini osnovnih akademskih studija na ETF-u, ali i svima koje zanima obrada signala, nauka o podacima i biomedicinsko inženjerstvo.

Programski jezici, kao i prirodni, prolaze niz promena tokom svog trajanja u određenom vremenskom periodu, jer ne predstavljaju ništa drugo do alatku koja je razvijena za potrebe izvršenja zadataka na mašini. Bez ikakvih pretenzija da se programski jezik R večno zadrži i održi u upotrebi, kao i bez želje da to postane jedini izbor za analizu biosignala, primeri prikazani u ovom udžbeniku su predstavljeni u programskom jeziku R, ali se primena prikazanih metoda obrade signala može uz manje izmene realizovati u drugim programskim jezicima. Cilj ovog udžbenika nije se razvije zavisnost ka bilo kom programskom jeziku, a ponajmanje tromost, na koju se često nailazi u tehnološkom svetu i akademskoj zajednici, da se ono što je prvo naučeno nikada ne menja i da se ne radi na sopstvenom razvoju i rastu kroz kontinuirano učenje i upotrebu novih alata, za koje je Autorka sigurna da će biti bolji i jednostavniji za upotrebu u budućnosti.

Ima više razloga zašto je R odabran kao programski jezik za analizu biosignala: (1) to je jedan od najpopularnijih (trenutno) programskih jezika u oblasti nauke o podacima i bioinformatiki (koja nije isto što i biomedicinsko inženjerstvo, ali povezanost između ove dve oblasti se ne može isključiti), (2) programski jezik R je slobodan softver, te ispunjava četiri slobode koje su definisane licencom koja je primerena za upotrebu u akademskoj zajednici i (3) pokazano je da ovaj programski jezik relativno brzo usvajaju i oni bez iskustva u programiranju, te se očekuje da će studenti završne godine osnovnih akademskih studija i studenti master studija moći kroz samo jedan kurs uz minimalni napor da savladaju rad u R-u, te da ostatak vremena mogu da posvete analizi biosignala.

U duhu otvorene nauke, slobodnog softvera i otvorenih podataka za koje se Autorka na Elektrotehničkom fakultetu, Univerziteta u Beogradu i globalno već duži niz godina zalaže, ovaj udžbenik je podeljen pod otvorenom licencom

Kreativne zajednice (eng. *Creative Commons*) CC BY 4.0¹ (Autorstvo, međunarodna licenca) u izvornom otvorenom formatu (sa ekstenzijom .odt) i u pdf-u, te kao takav nikada neće predstavljati opterećenje za ionako opterećeni studentski budžet. Studenti mogu slobodno štampati sadržaj udžbenika za svoje potrebe. Dodatno, elektronsko izdanje udžbenika je pogodnije za čitanje, jer udžbenik sadrži niz hiperlinkova i fusnota, odnosno napomena, koje pružaju dodatna pojašnjenja za praćenje izloženog sadržaja. Takođe, prpratni kodovi i materijali su dostupni na stranici <https://github.com/NadicaSm/TOBS> i podeljeni su pod slobodnom softverskom licencom GNU GPL verzija 3² (eng. *General Public License*). Svi koji žele da primere prikazane u ovom udžbeniku koriste za svoje potrebe bilo u nauci ili nastavi to mogu da urade i jedini uslov je da citiraju udžbenik na odgovarajući način. Udžbenik je pisan neutralnim tehničkim jezikom sa akcentom na upotrebu pasiva gde god je to bilo moguće i svi pojmovi su pisani u muškom gramatičkom rodu, a odnose se podjednako na ženski i muški rod u tekstu.

Pobrojati sve kojima je Autorka zahvalna što je ovaj udžbenik ugledao svetlost dana nije nimalo jednostavan zadatak, te se Autorka unapred izvinjava svima koje je eventualno propustila da spomene. Iako nisu direktno doprineli pisanju udžbenika, redovni profesori Univerziteta u Beogradu i Univerziteta u Ljubljani Predrag Pejović, Jaka Sodnik i Tomislav Šekara su bili stalni podstrek i podrška u profesionalnom životu Autorke, tako da bez rezerve oni zauzimaju prvo mesto u ovoj skromnoj zahvalnici velikim naučnicima i ljudima, iako se sigurno nikome od njih ovako pompezne reči neće dopasti. Svesno prihvatajući rizik da im ton može biti intiman i neformalan, Autorka im se od sveg srca zahvaljuje!

Drugo, a ne manje važno mesto u zahvalnici pripada studentima, posebno onima koji su svojim entuzijazmom i radoznalošću u oblasti analize biosignala inspirisali Autorku da preispita svoja gledišta i način na koji prenosi znanja budućim biomedicinskim inženjerima i akademskim građanima. Posebno, Autorka se zahvaljuje dr Nenadu B. Popoviću koji je doktorirao pod njenim mentorsvom, kao i doktorandu Iliji Tanaskoviću na neiscrpnim diskusijama iz oblasti obrade biosignala. Naročitu zahvalnost za čitanje udžbenika i više nego korisne sugestije u raznim fazama pisanja Autorka duguje doktorandima Iliji Tanaskoviću i Smilji Stokanović, kao i studentkinji master akademskih studija Minji Vuković.

1 <https://creativecommons.org/licenses/by/4.0/>, pristupljeno 29.07.2023.

2 <https://www.gnu.org/licenses/licenses.html#GPL>, pristupljeno 29.07.2023.

Autorka je mišljenja da obrazovni i naučnoistraživački rad na fakultetima ima svoje prednosti u odnosu na naučnoistraživački rad koji je u velikoj meri zastupljen na institutima. Razlog tome je što se saradnici i profesori stalno vraćaju na osnove gradiva koje predaju i koje većina naučnika uzima zdravo za gotovo. Saradnici i profesori imaju tu privilegiju da kroz neprestano prenošenje osnovnih znanja i iznalaženje kreativnih načina za prenos znanja studentima, preispituju sebe i druge, te da pristupaju postojećim znanjima na različite načine što oplemenjuje njihov naučnoistraživački rad, a može čak i da uzdrma postojeće prakse iz temelja ili pak da ih učvrsti. Sa druge strane, raznovrsno naučnoistraživačko i industrijsko iskustvo koje Autorka poseduje u velikoj meri utiče na znanja i iskustva koja su u ovom udžbeniku preneti studentima. Iz tog razloga, Autorka se zahvaljuje ljudima koji su imali viziju multidisciplinarnog pristupa nauci, te su vrlo rado sa Autorkom uplovili u analizu biosignala sa svojih stanovišta medicinskih lekara i psihologa. Mogućnost da se problem sagleda očima onih koji ga rešavaju i vide na drugačiji način je oblikovao ovaj udžbenik na mnoge načine. Za intenzivnu, dinamičnu i prijateljsku saradnju koja je razvijena sa Laboratorijom za ispitivanje individualnih razlika (LIRA) na Filozofskom fakultetu, Univerziteta u Beogradu, Autorka se zahvaljuje svim članovima Laboratorije, a posebno redovnom profesoru Goranu Kneževiću i naučnoj savetnici Ljiljani B. Lazarević Valerjev. Za upliv u oblast otvorene nauke koja je u velikoj meri i oblikovala ovaj udžbenik, Autorka je osobito zahvalna pionirki otvorene nauke u Srbiji Milici Ševkušić, višem diplomiranom bibliotekaru. Na kraju, Autorka je beskrajno zahvalna recenzentima udžbenika (profesoru Pejoviću, profesoru Filipoviću i docentu Ajčeviću) što su izdvojili svoje vreme i učinili Autorki tu čast i zadovoljstvo da daju svoju stručnu ocenu udžbenika.

U Beogradu, 12. maja 2024. godine

Autorka

Dr Nadica Miljković, vanredna profesorka na Elektrotehničkom fakultetu,
Univerziteta u Beogradu i gostujuća istraživačica
na Elektrotehničkom fakultetu, Univerziteta u Ljubljani

1 Osnovni pojmovi i korisne informacije za rad u programskom jeziku R

Programski jezik R³ su kreirali Ros Ihaka i Robert Džentlmen sa Univerziteta u Oklandu na Novom Zelandu⁴ davne⁵ 1991. godine. Naziv jezika označava početna slova imena njegovih tvoraca. Zvanično, programski jezik R je objavljen 1996 godine [2], a godinu dana pre toga je postao slobodan softver. Kao osnova za programski jezik R poslužio je programski jezik S (od eng. *Statistics*) koji je razvijen u Belovim laboratorijama 1976. godine. Iako R i S imaju sličnu sintaksu, različita im je semantika. Sintaksa određuje pravila po kojima se reči kombinuju u rečenici za prirodne jezike tj. kako se reči kombinuju ili kako se pišu komande programskog jezika. Sa druge strane, semantika je oblast koja se bavi značenjem reči, odnosno koju funkcionalnost u programskom jeziku ima određena komanda ili reč. Prema tome, poznavanje jezika S može biti od pomoći da se savlada programski jezik R, ali se programski kod ne može direktno kopirati iz jednog u drugi programski jezik, odnosno nisu kompatibilni. Za sve one koji se prvi put sreću sa programskim jezikom R, istorijat samog jezika nije od prevelike važnosti, ali jeste za razumevanje primene R-a u praksi. Naime, primarna primena jezika S je u statistici, pa je R nasledio ovu namenu. Međutim, kako se programski jezik R razvijao i usavršavao, sa posebnim osvrtom na porast broja paketa⁶, R se sada osim u statistici primenjuje i u analizi signala i mašinskom učenju, a vrlo je popularan za vizuelizaciju. [3-9]

3 Postoji više načina kako se na srpskom jeziku izgovara programski jezik R. Neki inženjeri i analitičari potenciraju upotrebu srpskog “R”, neki ga izgovaraju kao “ER”, a u upotrebi su i američko “/ɑ:r/” i britansko “/ɑ:/” (gde se r i ne čuje). Nema pravila ni ograničenja kako se može izgovarati naziv programskog jezika. Na TOBS predmetu na Elektrotehničkom fakultetu, Univerziteta u Beogradu se koristi američko “/ɑ:r/”.

4 <https://www.auckland.ac.nz/en.html>, pristupljeno 30.07.2023.

5 U pojedinim akademskim krugovima se programski jezik R smatra novinom u odnosu na, na primer, široko zastupljen Matlab (The Mathworks, Natick, USA) koji se kao komercijalni softver pojavio 1984. godine [1] što nije preterano velika razlika, ali dovoljna da bar 7 generacija saradnika i nastavnika stekne teško promenljivu naviku da za analizu signala koristi isključivo Matlabu pri tom odbijajući da nauče i primene druge alate.

6 Paketi u programskom jeziku predstavljaju funkcionalne celine. Negde se nazivaju paketima (eng. *package*), negde bibliotekama (eng. *library*), alatkama (eng. *tool*, *toolbox*) ili modulima (eng. *module*). U R-u se koristi engleska reč *library*, dok je dominantno i bez posebnog razloga u ovom udžbeniku u većini slučajeva korišćen naziv paket.

Iako je tvorcima programskog jezika R bila logična i već viđena praksa da programski jezik nazovu samo jednim slovom (slučaj programskog jezika S), to može predstavljati problem kada se naziv programskog jezika koristi za pretragu na internetu. Iskustvo pokazuje da je u većini slučajeva potrebno koristiti “*R programming*” prilikom pretrage da bi se nedvosmisleno naglasilo da se pretražuju izvori koji se odnose na programski jezik R. Bez želje da se ovaj udžbenik koristi za reklamiranje pretraživača kompanije Gugl⁷, ali uzimajući u obzir činjenicu da korišćenje Gugla jeste uobičajena praksa, posebno među studentima, postoje dodatne smernice koje mogu da pomognu u pretraživanju materijala na Gugl platformi⁸ (sa naglaskom na korišćenje naprednih opcija i operatora) za ograničen broj pretraga, jer je za specijalizovane izvore značajno bolje koristiti odgovarajuće baze podataka [10-11]. Ove smernice sa Berkli Univerziteza i sa Džon Hopkins Univerziteta [10-11] mogu biti od velike koristi kako za pretragu materijala za programiranje u R-u, tako i za ostale pretrage.

R je slobodan softver i podeljen je pod GNU⁹ GPL 2.0 (eng. *General Public License*) licencom, što znači da je izvorni kod programskog jezika, kao i funkcije koje dolaze sa osnovnim R paketima i sa paketima koji su podeljeni sa istom ili sličnom licencom dostupni svima [3]. Svaki slobodan softver i sve licence slobodnog softvera ispunjavaju četiri uslova tj. slobode koje je definisala Fondacija za slobodni softver (eng. *Free Software Foundation*, skraćeno FSF) f13¹⁰:

- sloboda 0. da se program pokreće u bilo koju svrhu
- sloboda 1. da se program proučava i prilagođava (pristup izvornom kodu je obavezan)
- sloboda 2. da se dele kopije programa i
- sloboda 3. da se program unapredi i da se deli unapređena verzija (pristup izvornom kodu je obavezan).

Odličan pregledan rad na srpskom jeziku, posebno namenjen početnicima je predstavljen na Elektrotehničkom fakultetu, Univerziteta u Beogradu 2020.

7 <https://about.google/>, pristupljeno 29.07.2023.

8 <https://www.google.com/>, pristupljeno 29.07.2023.

9 GNU je rekurzivna skraćenica od *GNU is Not Unix* koja je nastala sa vizijom da se stvori slobodan operativni sistem koji će biti nalik Unix-u, ali ne baš isti kao Unix, jer Unix nije bio slobodan softver. Više informacija se može naći na internet stranici Fondacije za slobodan softver (eng. *Free Software Foundation*, skraćeno FSF) [12].

10 Preporuka je da se za razumevanje koncepta slobodnog softvera pogleda predavanje Ričarda Stalmana Free software, free society: Richard Stallman at TEDxGeneva 2014, https://youtu.be/Ag1AKIl_2GM, pristupljeno 29.07.2023.

godine u okviru PSSOH konferencije (Primena Slobodnog Softvera i Otvorenog Hardvera) i dostupan je pod otvorenom licencom na internetu [14]. Poznavanje licenci je od posebnog značaja za rad sa softverom u savremenom digitalnom dobu. Ako je programski kod podeljen preko interneta, a nije mu dodeljena nikakva licenca, onda je upotreba tog koda ilegalna. Ovo je važno i za one koji dele softver i za one koji ga preuzimaju. Dodatno, neke licence onemogućavaju komercijalnu upotrebu, pa je legalno njihovo korišćenje u akademskom okruženju, ali nije ako neko radi u privatnoj kompaniji te nije moguće ugraditi preuzeti kod u komercijalni proizvod.

Slobode koje je definisala FSF, a koje su licencom dodeljene programskom jeziku R iskoristila je korporacija Majkrosoft¹¹, te je iz izvornog koda R jezika razvijen novi programski jezik Microsoft R-Open koji je takođe otvoren softver u skladu sa originalnom R licencom [15]. Ova praksa je potpuno legitimna, a dodatno ukazuje na potencijal programskog jezika koji je prepoznala ova privatna kompanija i to za primenu u oblasti statističke analize kompleksnih¹² podataka, prediktivnog modeliranja i mašinskog učenja [15]. U 2021. godini je najavljeno da će Majkrosoft povući Microsoft R-Open, ali ne zbog gubitka interesovanja za oblast i programiranje u R-u, već zbog migracije razvijenih alata ka originalnom R projektu [17-18]. Iz kog razloga je došlo do ove migracije, nije poznato, ali sigurno je da postoji niz prednosti prilikom razvoja slobodnog softvera i softvera otvorenog koda¹³ koje je i Majkrosoft prepoznao, ali i niz specifičnosti koji se ogleda kroz organizaciju zajednice koja doprinosi takvom projektu. Odličan primer održivosti slobodnog softvera je dao Italo Vinjoli tokom plenarnog predavanja održanog tokom pete PSSOH¹⁴

11 <https://www.microsoft.com/sr-latn-rs/>, pristupljeno 30.07.2023.

12 Statistika kompleksnih podataka predstavljena kao “*big data statistics*” u [15], ali kako postoje akademska okruženja koja nisu naklonjena terminu “*big data*”, ovaj udžbenik prati takve preporuke, te se u njemu neće pojavljivati termin “veliki podaci” odnosno engleski termin “*big data*” [16].

13 Fondacija za slobodni softver koristi isključivo termin slobodan softver, dok je inicijativa za otvoreni kod (eng. *Open Source Initiative*) insistirala na terminu softver otvorenog koda (eng. *open-source software*). Postojao je veliki jaz između ove dve organizacije koji se oslikavao i u terminima koje su koristile. Međutim, kako se danas koriste iste licence i za slobodan softver i za softver otvorenog koda (koje su u skladu sa četiri slobode koje je definisala Fondacija za slobodni softver), ovaj sukob se smatra završenim. Da bi se naglasilo da sukob više ne postoji, u upotrebi je FOSS skraćenica od eng. *Free and Open-Source Software* ili FLOSS od eng. *Free/Libre and Open-Source Software*. *Libre* je španska reč za slobodu na kojoj je Fondacija za slobodni softver insistirala kako bi se naglasilo da slobodan softver nije isto što i besplatan softver (engleska reč *free* označava oba pojma) [13-14]. U ovom udžbeniku se dominantno koristi termin slobodan softver.

14 <https://pssoh.etf.bg.ac.rs/>, pristupljeno 30.07.2023.

konferencije u Beogradu u oktobru 2022. godine na primeru LibreOffice¹⁵ projekta [619].

U obrazovanju, primena slobodnog softvera ima veliki značaj. Pored toga što ustanova i država (ako se radi o držanom univerzitetu) nemaju dodatnih finansijskih izdataka, dostupnost izvornog koda (koju garantuju sloboda 1 i sloboda 3. slobodnog softvera [13]) je presudna za obrazovne programe. Na primer, neka je zadatak da studenti razumeju kako radi *findpeaks()* funkcija koja ima za cilj detekciju lokalnih ekstremuma na signalu. Ova funkcija se može koristiti za detekciju R pikova na elektrokardiografskom signalu¹⁶ što predstavlja osnovni pristup u detekciji srčanog ritma i metodama koje imaju za cilj ispitivanje varijabilnosti srčanog ritma. Ako bi studenti koristili slobodan softver, onda im je pristup izvornom kodu *findpeaks()* funkcije omogućen, za razliku od toga ako koriste vlasnički softver. Primer funkcije za detekciju lokalnih ekstremuma je jednostavan, ali iza zaključanih vrata (nedostupnost izvornog koda) mogu se naći i pogrešne i tačne nepoznate procedure. Studenti, u tom slučaju, nisu u mogućnosti da nauče kako program funkcioniše niti saradnici i nastavnici mogu da im to pokažu, što je loša akademska praksa.

Sloban softver ne dovodi do tzv. neželjenog zaključavanja (od eng. *user-lock-in* ili *vendor-lock-in*) koji predstavlja situaciju u kojoj korisnici postaju zavisni od tehnologije i proizvoda jednog dobavljača, usled čega ne mogu da pređu na drugi proizvod [21]. Na primer, moguće je da kompanija softverski proizvod (u ovom slučaju programski jezik i razvojno okruženje) nudi besplatno ili po relativno maloj ceni, te na taj način proizvod postane logičan izbor korisnika ili akademske zajednice. Nakon što je kompanija sigurna da su korisnici stekli navike ili da bi prelazak na drugi sistem doveo do dodatnih troškova (koji ne uključuju samo dodatna finansijska sredstva za kupovinu drugog programskog jezika, već, na primer, i vreme i stručnost da se prethodno razvijeni programski kodovi prevedu i prilagode novom programskom jeziku), softverski proizvod se menja i postaje dostupan isključivo uz dodatnu doplatu. Na taj način kompanija može da određuje i menja cenu proizvoda zarad ostvarivanja profita. Neželjeno zaključavanje može biti štetno po institucionalnu autonomiju i akademske slobode, pa je čak označeno kao praksa koja može da šteti ugledu univerziteta u društvu prema deklaraciji Lige univerziteta Evrope (eng. *League of European Research Universities*, skraćeno LERU¹⁷) [22]. Stav LERU je da se omogući

15 <https://www.libreoffice.org/>, pristupljeno 30.07.2023.

16 Ovde je detekcija R pikova data kao ilustrativni primer, ali je potrebno naglasiti da se u praksi detekcija R pikova vrši primenom specijalizovanih algoritama, među kojima je najpoznatiji Pan-Tompkins algoritam [20].

17 <https://www.leru.org/>, pristupljeno 27.05.2024.

slobodan pristup rezultatima istraživanja na univerzitetima, kao i podacima, algoritmima, metapodacima i softverskim alatima koji su rezultat istraživanja, što je moguće isključivo uz slobodan softver [22].

Inicijativa za prelazak sa vlasničkog na slobodni softver u akademskom okruženju se nametnula u velikoj meri biomedicinskim inženjerima, slično kako su se prakse otvorene nauke nametnule psiholozima nakon fantastičnog i istovremeno poražavajućeg otkrića u studiji iz 2015. godine koje je ukazalo na potrebu za primenom principa otvorene nauke kako bi se omogućila reproducibilnost¹⁸ objavljenih studija i naučnih rezultata [23]. Primera radi, jedan od tvoraca SciPy biblioteke [24] za programski jezik Pajton [25] je Trevis Olifant. On je takođe učestvovao u kreiranju i NumPy biblioteke [26], ali i Anakonda distribucije za Pajton i R programske jezike (Anaconda, Inc., Austin, SAD). Trevis je 90ih godina prošlog veka za potrebe rada na svojoj doktorskoj disertaciji iz oblasti biomedicinskog inženjerstva prvi put koristio slobodan programski jezik Pajton [27]. Nakon toga, njegova karijera je dobila potpuni zaokret – od biomedicinskog inženjerstva ka slobodnom softveru. On nije jedini primer. Profesor Tobajas Knop je još jedan biomedicinski inženjer koji je počeo ne samo da intenzivno koristi, već i da razvija slobodan softver i to za programski jezika Džulija (Julia¹⁹) [28-29]. Obojica su Matlab (The Mathworks, Natick, SAD) zamenili slobodnim softverom.

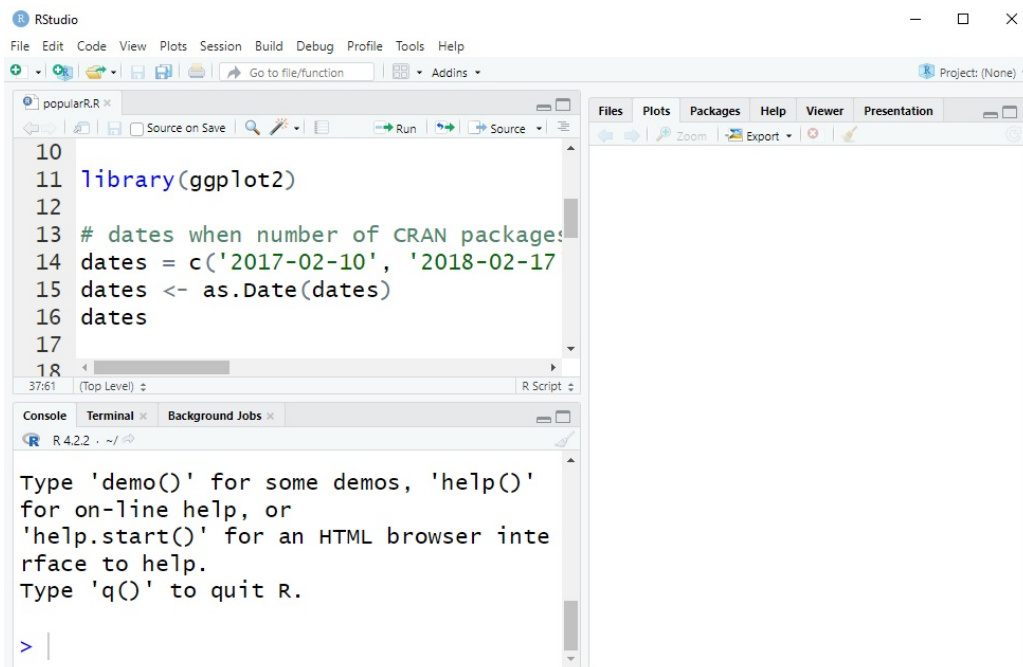
Iako se tradicionalno upotreba R programskog jezika svodi na primenu u statistici, već duže od deceniju R se dominantno koristi u oblasti nauke o podacima²⁰ (eng. *data science*), te je predložena i promena definicije programskog jezika R koja sada uključuje nauku o podacima, a ne isključivo statistiku [3]. Danas se za R kaže da je *lingua franca* nauke o podacima,

18 Reproducibilnost od eng. *reproducibility* bi bilo preciznije prevesti na srpski jezik kao reproduktivnost, ali kako još uvek nije ustaljen odgovarajući termin na srpskom jeziku, to se u ovom udžbeniku koristi isključivo termin reproducibilnost.

19 <https://julialang.org/>, pristupljeno 28.07.2023.

20 U nekim oblastima, istraživači, analitičari i inženjeri ne koriste termin nauka o podacima (eng. *data science*). Kako je ovo široko rasprostranjen termin i kako Elektrotehnički fakultet, Univerziteta u Beogradu i Fakultet tehničkih nauka, Univerziteta u Beogradu imaju zajednički master program Napredne informacione tehnologije u digitalnoj transformaciji koji eksplicitno eksploatiše termin nauka o podacima (<https://www.dsi.rs/master-40-it-biznis/programi/?study-programme=%7B%22module%22%3A448%7D&submit=OK>, pristupljeno 28.07.2023), to je rešeno da se termin nauka o podacima zadrži u udžbeniku. Bez obzira da li će se čitaoci opredeliti za korišćenje termina ili ne, činjenica je da je termin prilično maglovit, jer pored toga što neko ko se bavi naukom o podacima ne mora da bude naučnik, a postoji relativno veliki broj raznovrsnih poslova i metoda koje ovaj termin obuhvata da je već bilo predloga za redefinisavanje i menjanje termina nauka o podacima [30].

odnosno opšteprihvaćeni jezik za oblast nauke o podacima [3]. Uprkos predrasudama oko isključive primene programskog jezika R u statistici, u njemu postoji niz manje poznatih, ali veoma efikasnih i korisnih paketa. Primer je serial paket [31] koji se može koristiti za merenje signala, jer sadrži procedure koje vrše učitavanje i upis podataka u binarnom i ASCII (eng. *American Standard Code for Information Interchange*) formatu preko RS232/RS422/RS485 ili drugih virtuelnih serijskih portova na računaru. Na TOBS predmetu se ovaj paket koristi za merenje srčanog pulsa primenom infracrvenog senzora i Arudino mikrokontrolerske pločice.



Slika 1, Izgled R Studio razvojnog okruženja. Radi bolje preglednosti veličina slova tj. font uređivača (eng. *Editor*) je uvećan. Za one koji su upoznati sa drugim razvojnim okruženjima (kao na primer: Matlab, GNU Octave, Pajčarm, Spajder), ovakav raspored prozora ne bi trebalo da predstavlja iznenađenje.

Za razliku od programskog jezika Pajton (Python, Python Software Foundation, Wilmington, DE, USA) za koji postoji veliki broj integrisanih razvojnih okruženja (eng. *Integrated Development Environment*, skraćeno IDE) kao što su Spajder (Spyder²¹) i Pajčarm (PyCharm²²), za R je široko rasprostranjen samo jedno razvojno okruženje i to R Studio²³ (Posit Co, Boston, USA). Početni

21 <https://www.spyder-ide.org/>, pristupljeno 28.07.2023.

22 <https://www.jetbrains.com/pycharm/>, pristupljeno 28.07.2023.

23 <https://posit.co/download/rstudio-desktop/>, pristupljeno 28.07.2023.

ekran R Studio okruženja je prikazan na Sl. 1. Jedna od prednosti R-a i R Studija je što se vrlo jednostavno i brzo instaliraju na računaru. Da bi se koristile odgovarajuće metode i algoritmi koje su razvili članovi zajednice, jedna od prvih komandi koju treba savladati je komanda za instalaciju paketa.

Najveći broj R paketa je smešten na CRAN-u (eng. *Comprehensive R Archive Network*) koji predstavlja centralni zvanični softverski repozitorijum R Fondacije [32]. Kao pandan CRAN-u, postojao je i MRAN (eng. *Microsoft R Application Network*) koji je razvijala i do 2023. održavala Majkrosoft kompanija sve dok nije usledila migracija ka CRAN-u [17-18], a R paketi se mogu naći i na sajtu Bioconductor²⁴ projekta koji je posebno važan za one koji se bave ili planiraju da se bave bioinformatikom. Takođe, R paket se mogu naći i na internet platformama zasnovanim na primeni Git sistema za distribuirano verzionisanje i kolaboracioni razvoj softvera kao što su GitHub²⁵ i BitBucket²⁶. Zanimljivo je da je razvoj Git sistema usko povezan sa razvojem slobodnog softvera, jer se upravo u takvim okruženjima gde programski kod razvija čitava zajednica javila potreba za upravljanjem razvojem softvera [33]. Kako doći do stabilne verzije nekog softvera ili kako izvesti testiranje softvera na čijem razvoju rade desetine, stotine ili pak hiljade programera? Odgovor je Git, koji omogućava da se prate izmene izvornog koda unutar kolekcija datoteka (eng. *files*) i/ili fascikli (eng. *folder*) sa odgovarajućim metapodacima²⁷ [34].

Za rad u R-u potrebni su paketi, tako da će svako ko koristi ovaj programski jezik morati da prođe kroz proceduru instalacije paketa. CRAN repozitorijum poseduje interne politike tj. regulative kojima su uređeni minimalni standardi koji se odnose na metapodatke i kvalitet paketa koji se postavljaju u repozitorijum [35]. Sa druge strane, internet platforme kao što je GitHub pružaju veliku slobodu korisnicima i Autorima paketa i ne unose nužno rigorozne provere i recenzije koda, te se korišćenje paketa koji su slobodno dostupni na ovakvim platformama savetuje uz odgovarajući oprez i proveru ispravnosti preuzetog softvera.

24 <https://www.bioconductor.org/>, pristupljeno 28.07.2023.

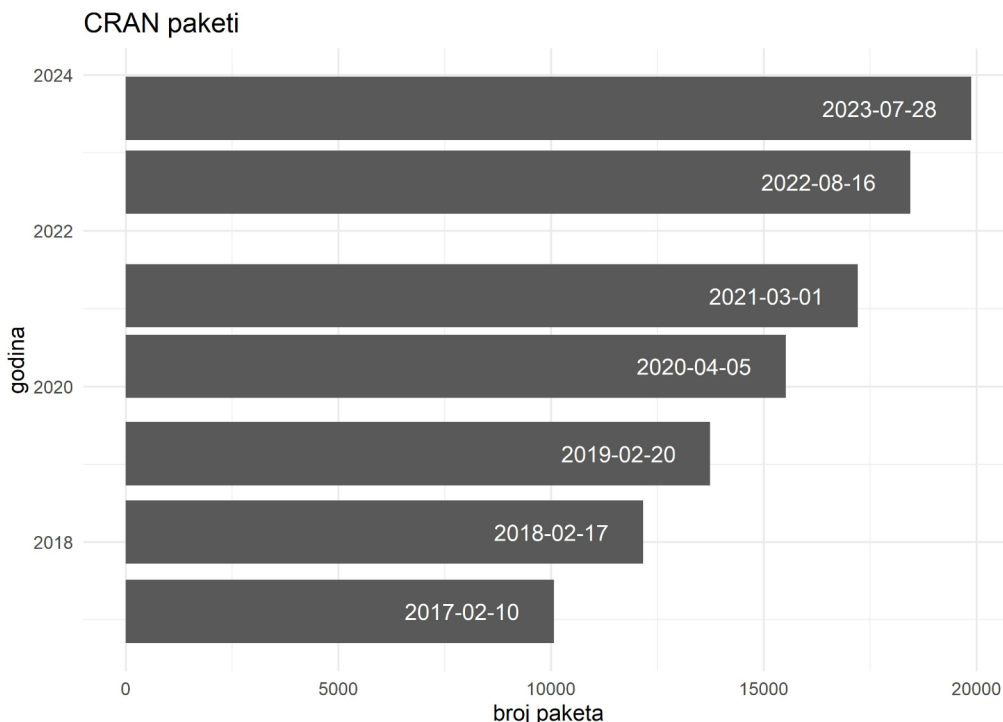
25 <https://github.com/>, pristupljeno 28.07.2023.

26 <https://bitbucket.org/product>, pristupljeno 28.07.2023.

27 Jednostavno rečeno, metapodaci su podaci o podacima. Ako se ovaj udžbenik smatra podatkom, onda su primeri metapodataka: ime i prezime Autorke, naslov udžbenika, broj strana, ISBN broj. Za softverski kod, to mogu biti vreme poslednje izmene, sadržaj izmene, slobodni komentari programera, ime i prezime ili samo korisničko ime programera [34].

Broj paketa na CRAN-u raste iz godine u godinu. Na GitHub profilu²⁸ je za potrebe kursa Tehnike obrade biomedicinskih signala (TOBS) na master-akademske studijama na Elektrotehničkom fakultetu, Univerziteta u Beogradu, podeljen broj CRAN paketa od početka držanja nastave iz TOBS predmeta koji je periodično uređivan i dopunjavan. Rezultat tog R koda je prikazan na Sl. 2.

CRAN paketi, a generalno svi R paketi, mogu pored specijalizovanih funkcija posedovati i podatke ili isključivo podatke. Primer paketa koji se sastoji isključivo iz podataka je ISwR paket (eng. *Introductory Statistics with R*) [36] koji sadrži primere iz udžbenika za statistiku [37]. Jedan primer podataka iz ISwR paketa predstavlja skup *lung* (srp. pluća) koji sadrži 18 merenja kapaciteta pluća na 6 ispitanika primenom tri različite metode [36].



Slika 2, Broj R paketa na CRAN repozitorijumu od početka držanja TOBS kursa na Elektrotehničkom fakultetu, Univerziteta u Beogradu. Jedina izmena u odnosu na <https://github.com/NadicaSm/simpleRgraph> (pristupljeno 28.07.2023.) jeste što su oznake na osama i naslov grafika ispisani na srpskom jeziku.

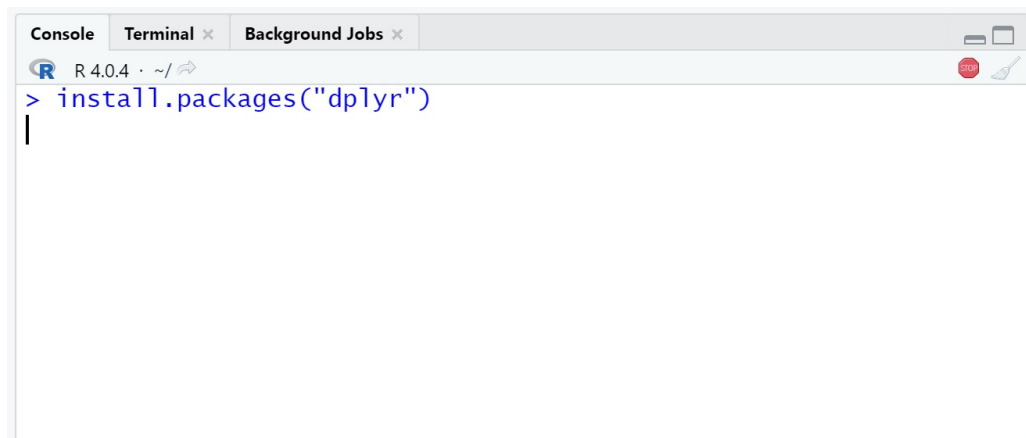
28 <https://github.com/NadicaSm/simpleRgraph>, pristupljeno 28.07.2023.

Da bi se podaci i metode koje su ugrađene u neki od paketa mogli koristiti potrebno je da se najpre instalira željeni paket. Na Sl. 3 je prikazano kako se instalira dplyr paket (eng. *A Grammar of Data Manipulation*) [38] sa CRAN repozitorijuma primenom:

```
install.packages()
```

funkcije. Po ugledu na pip biblioteku u Pajtonu za instalaciju paketa [39], koja nakon odgovarajuće komande kojom se želi instalacija odabranog paketa sa zvaničnog PyPI Pajton repozitorijuma (eng. *Python Package Index*) omogućava automatski pristup internetu, preuzimanje paketa i instalaciju (ali i testiranje, instaliranje zavisnih i potrebnih dodatnih paketa i druge operacije), slično radi i *install.packages()* funkcija u R-u. Prilikom koršćenja funkcionalnosti nekog od R paketa (koji nije osnovni paket) potrebno je pozvati funkciju *library()*. Funkcija *install.packages()* se koristi isključivo za instaliranje paketa sa CRAN-a, a kako je dplyr dostupan i na GitHub internet platformi, to se isti paket može instalirati pozivanjem komande:

```
install_github("hadley/dplyr").
```



Slika 3, Prikaz konzole u okviru R Studio razvojnog okruženja sa pokrenutom komandom za instaliranje dplyr paketa.

Iako su impresivni i broj i porast broja softverskih paketa na CRAN repozitorijumu za slobodni programski jezik R (Sl. 2), nisu svi uslovi rada sa slobodnim softverom idealni. Mana svih programskih jezika koji su podeljeni pod licencama slobodnog softvera je što u slučaju da ne postoji metoda koju je neko prethodno implementirao u okviru paketa ili dostupnih funkcija, onda se mora samostalno razviti ili je potrebno naći nekog ko će da razvije programski

kod koji obavlja željenu funkciju [3]. Dodatno, u literaturi se često pojavljuje nedostatak programskog jezika R koji se odnosi na upravljanje memorijskim resursima prilikom analize signala [3]. Naime, promenljive u R-u su smeštene u fizičkoj memoriji te se može desiti da se ne može lokalno na računaru realizovati analiza signala, jer ne postoji dovoljno računarskih resursa. Međutim, postoje određeni paketi u R-u koji mogu da optimizuju korišćenje memorije, te je u većini slučajeva RA memorija (eng. *Random Access Memory*, skraćeno RAM²⁹) na računaru je dovoljno velika.

U članku “*RAM is eating big data*” (srp. “Memorija RAM-a nadmašuje memoriju podataka”, slobodan prevod sa engleskog na srpski jezik) je pokazano da je brzina porasta RA memorije veća od brzine kojom se uvećava memorija koju zauzimaju skupovi podataka koji se koriste u nauci o podacima, te da mnoga analitička predviđanja o prevelikoj količini podataka nisu nužno tačna [40]. Naravno, treba uzeti u obzir da nisu svi podaci isti ni po tipu ni po veličini memorije koju zauzimaju na računaru, ali je činjenica je da se i računarski resursi menjaju. Kada ne postoji dovoljno memorije na računaru ili kada se ta memorija ne može proširiti, rešenje je da se koriste klasteri računara kojima se pristupa preko interneta. Danas je takav način analize poznat pod nazivom računarstvo u oblaku (eng. *cloud computing*), ali i raspodeljeno računarstvo (eng. *distributed computing*) [41-42]. Ova tehnologija omogućava, osim izvršenja računarskih zadataka sa visokom performansom (eng. *High-Performance Computing*, skraćeno HPC), korišćenje i memorijskih resursa i infrastrukture. Korišćenje ovakvih resursa, iako na prvi pogled deluje maglovito (sve se dešava u oblaku, van računara korisnika) zahteva veliku količinu resursa, te može biti i relativno skupo. Na tržištu su zastupljeni Majkrosoft Ejžr (Microsoft Azure)³⁰ i Amazon³¹, ali inženjeri koji se bave analizom podataka i analitičari se opredeljuju i za klastere i servise dostupne pod licencama slobodnog softvera kao što je Apači Hadup (Apache Hadoop)³². Projekat Apači Hadup koristi Apači licencu koju je Fondacija za slobodni softver kategorizovala kao licencu slobodnog softvera [43]. Metode zasnovane na distribuiranom računarstvu i računarstvu u oblaku se neće koristiti u udžbeniku, jer metode analize i podaci predstavljeni u ovom udžbeniku ne prevazilaze resurse personalnog računara. Međutim, posebno sa osvrtom na trenutno popularne metode analize video signala i primene računarske vizije (eng. *computer vision*) u biomedicinskom inženjerstvu, nije isključeno da će čitaoci u

29 Pravilno je RAM ili RA memorija, ali nije pravilno RAM memorija, jer je M skraćeno od memorija (eng. *Memory*), te su prema tome obe skraćenice u optičaju.

30 <https://azure.microsoft.com/en-us>, pristupljeno 29.07.2023.

31 <https://aws.amazon.com/what-is-cloud-computing/>, pristupljeno 29.07.2023.

32 <https://hadoop.apache.org/>, pristupljeno 29.07.2023.

svojoj praksi naići na problem koji će zahtevati pristup ovakvim resursima. Metode zasnovane na merenju i analizi video signala su važne za razvoj i širu primenu merenja na daljinu (eng. *remote measurements*) osnovnih fizioloških parametara kao što su, na primer, srčani ritam, temperatura tela i frekvencija disanja. Beskontaktno merenje na daljinu je oblast koja će još dugo biti u fokusu u biomedicinskom inženjerstvu. Primera radi, da bi podstakli razvoj u ovoj oblasti urednici časopisa *Biomedical Signal Processing and Control* su u 2023. godini otvorili posebno izdanje časopisa i poziv za radove u oblasti fizioloških merenja na daljinu pod nazivom *Remote Physiological Measurement: Novel algorithms and application areas* (srp. Fiziološka merenja na daljinu: Novi algoritmi i primene, slobodan prevod sa engleskog jezika na srpski) [44].

Za čitaoce koji budu u situaciji da koriste ovakve resurse, važan je podatak da u Republici Srbiji postoje servisi koji uključuju korišćenje superračunara i to na dva mesta – u državnom data centru u Kragujevcu koji je otvoren 2020. godine i u Institutu za fiziku, Univerziteta u Beogradu. Ovakvi centri su od strateškog značaja, jer je važno da se prilikom korišćenja resursa dostupnih na internetu obrati pažnja da se u oblak³³ ili na udaljena mesta ne šalju podaci koji su poverljivi i koji mogu sadržati podatke o ličnosti (osim ako čuvanje poverljivosti nije uređeno posebnom uredbom), što je čest slučaj sa podacima u biomedicinskom inženjerstvu. Prema saznanjima Autorke, superračunar u Kragujevcu je dostupan istraživačima za korišćenje, dok je Institut za fiziku trenutno zatvoren za saradnju sa istraživačima u Republici Srbiji, što ne znači da se ta situacija neće promeniti u skorijoj budućnosti. UNESCO (eng. *The United Nations Educational, Scientific and Cultural Organization*) preporuke o otvorenoj nauci uključuju i preporuke o deljenju istraživačkih resursa kao što su oprema i instrumenti u sekciji *Open science infrastructure* (srp. Infrastruktura za otvorenu nauku, slobodan prevod sa engleskog na srpski jezik) [45].

Zemlje u regionu i države Evropske unije već imaju nacionalne registre istraživačke opreme koja je dostupna za korišćenje istraživačima po principu otvorenog pristupa istraživačkoj infrastrukturi (primer je registar opreme Hrvatske zaklade za znanost [46]), a ovo se posebno odnosi na kapitalnu

33 Fondacija za slobodan softver nije blagonaklona prema nazivu oblak (eng. *Cloud*), te je deo njihove kampanje slogan “*There is no cloud – only other people’s computers.*” (srp. Ne postoji oblak, samo računari koji pripadaju drugim ljudima, slobodan prevod sa engleskog na srpski jezik), <https://www.fsf.org/resources/stickers>, pristupljeno 30.07.2023.

opremu³⁴. Te bi i Institut za fiziku, trebalo uskoro da usvoji slične dobre prakse otvorene nauke.

Pored toga što u nekim situacijama može biti potrebno da se istraže dodatni resursi za čuvanje podataka i realizaciju analize prilikom korišćenja R programskog jezika, nedavno je ukazano na još jednu manu R-a. Naime, R (kao uostalom i drugi programski jezici slobodnog softvera) nije pogodan za kreiranje programskih kodova koji ispunjavaju uslove računarske ponovljivosti (eng. *computational reproducibility*). Nacionalne akademije za nauku, inženjerstvo i medicinu u SAD definišu računarsku reproducibilnost kao mogućnost da se ponovi analiza i da se dobiju isti rezultati [47]. Računarska reproducibilnost podrazumeva da programski kod daje isti rezultat za pokretanje na istom ili različitom računaru, pod različitim operativnom sistemima, u različitim vremenskim intervalima³⁵ i sl. U prirodnim naukama, posebno tamo gde se primenjuje mašinsko učenje, pa samim tim i u biomedicinskom inženjerstvu, računarska reproducibilnost je vrlo važan aspekt, jer omogućava verifikaciju i proveru postignutih rezultata [49-50]. Logično je pretpostaviti da ako postoji softverski kod koji kao rezultat daje odluku da li je, na primer, neko bolestan ili nije ili kakvu terapiju je najbolje primeniti, da nije poželjno da taj kod daje različite rezultate za različita pokretanja koda, da se ponaša nepredvidivo ili da ga u pojedinim situacijama nije ni moguće pokrenuti. Za sada su predloženi standardi u naučnim radovima koje bi trebalo pratiti za kreiranje softverskog koda koji daje reproducibilni rezultat [49]. Važno je spomenuti da su R i SAS (SAS Institute Inc., Cary, NC, SAD) programski jezici najčešće korišćeni od strane farmaceutskih kompanija za analizu signala i prikaz kliničkih studija na osnovu kojih regulatorna tela kao što je FDA (eng. *Food and Drug Administration*) u SAD odlučuju, na primer, o puštanju u promet novih lekova ili terapijskih uređaja [51]. SAS Institut i R Fondacija su izdale posebna namenska dokumenta za ispunjavanje kriterijuma za pripremu softvera za regulatorna tela za odobrenje novih medicinskih uređaja ili lekova u farmaceutskoj industriji [52-53]. Ova uputstva na prvom mestu imaju u fokusu statistička proračunavanja i elemente koji se odnose na

34 Kapitalna oprema je osnovno sredstvo za rad za koje postoji oslobađanje od poreza. To bi u ovom slučaju značilo da ako se država odrekla poreza i na taj način uložila u opremu za istraživače, logično je da se uložena sredstva iskoriste na ekonomičan način posredstvom deljenja sa drugim ustanovama i pojedincima (koji kao poreski obveznici imaju pravo na takvu opremu), prema uslovima koje definiše vlasnik opreme. NAPOMENA: Ovo je slobodna Autorkina interpretacija, a ne zvanično pravno i ekonomsko stručno tumačenje.

35 Vremenski interval je relativno pojam u računarstvu. Jedan od aspekata koji je spomenut u intervjuu sa prof. Predragom Pejovićem sa Univerziteta u Beogradu – Elektrotehničkog fakulteta povodom računarske reproducibilnosti jeste šta bi moglo da se desi sa programskim kodovima koji se trenutno koriste u narednih 200 godina [48].

kvalitet softvera, a samo u manjoj meri direktno ili indirektno spominje se računarsku reproducibilnost. To je razumljivo, jer su tek u razvoju uputstva i alati za računarsku reproducibilnost. Posebno je kritičan aspekt računarske reproducibilnosti algoritama mašinskog učenja³⁶ i trenutno postoji više inicijativa za rešavanje ovog problema. Na primer, FARR RCN³⁷ organizacija (eng. *FAIR in ML, AI Readiness, & Reproducibility Research Coordination Network*) i više radnih i interesnih grupa unutar RDA³⁸ organizacije (eng. *Research Data Alliance*).

Na internet stranici bloga Data Colada³⁹ u 2021. i 2022. godini pojavila su se dva odlična članka u kojima su prikazani problemi koji mogu da nastanu kada se koristi programski jezik R, a koji su uzrokovani time što R kod nije računarski reproducibilan [54-55]. Ovi članci su na ilustrativan način ukazali na problem i što je još pohvalnije na moguća rešenja. Posebno je istaknut primer kada dolazi do narušavanja računarske reproducibilnosti usled izmena i unapređenja R paketa (eng. *Update*) koje čine paket nekompatibilnim unazad. Iako je stalno unapređenje postojećih paketa i programskih jezika pohvalno i poželjno, jer sa novijim verzijama dostupni alati i softverska rešenja postaju bolja i efikasnija, to dolazi sa ogromnom manom – narušava se računarska reproducibilnost. Praktično, ako je R kod napisan danas možda za godinu dana neće raditi sa istim paketom ili hoće ali neće proizvesti isti rezultat, jer se paket promenio u međuvremenu. To se, na primer, desilo 2016. godine sa *distinct()* funkcijom unutar dplyr paketa. U analizi preko 2000 dostupnih R skripti u repozitorijumima iz 2022. godine, pokazano je da za više od 50% nije bilo moguće pokrenuti, čak ni posle instaliranja paketa i popravki koje su uključile podešavanje aktivnog direktorijuma [56]. Deo grešaka je nastao upravo zato što nije bilo više moguće učitati paketa [55-56], te računarsku reproducibilnost treba vrlo ozbiljno shvatiti, posebno tamo gde su odluke kritične po ljudsko zdravlje (u farmaciji i medicini) ili gde je potrebna rigorozna metodologija da bi se došlo do novih otkrića (akademsko okruženje).

Postoji više načina kako može da se spreči ili u najbolju ruku smanji problem računarske reproducibilnosti, a jedan je da se koristi paket groundhog sa CRAN-a [57]. Korišćenjem funkcija ovog paketa se svaki put kada se kod koji je kreiran u nekom vremenskom trenutku pokreće sa verzijama paketa koje su bile dostupne u vreme kreiranja koda primenom funkcije:

36 Hipotetički, ako algoritam nije reproducibilan, moglo bi se desiti da jednog dana klasifikuje osobu kao pacijenta, a drugog dana istu osobu kao zdravog ispitanika.

37 <https://www.farr-rcn.org/about>, pristupljeno 31.07.2023.

38 <https://www.rd-alliance.org/>, pristupljeno 31.07.2023.

39 <http://datacolada.org/>, pristupljeno 30.07.2023.

groundhog.library()

sa datumom kao ulaznim parametrom. Ova funkcija je itekako korisna, jer za razliku od klasičnog programiranja u R-u kada se paket prvo mora instalirati pozivom *install.packages()*, a potom i aktivirati primenom *library()* funkcije, *groundhog.library()* funkcija automatski instalira potrebne pakete, te se kod na taj način minimizuje i povećava nivo apstrakcije. Primera radi, sledeći kod za paket *dplyr* omogućava reproducibilnost:

```
if (!require("groundhog")) install.packages("groundhog")
library(groundhog)
groundhog.day = "2023-07-30"
pkgs = c('dplyr')
groundhog.library(pkgs, groundhog.day, tolerate.R.version = '4.1.2')
```

Prva linija koda u prethodnom primeru omogućava da se proverí da li je *groundhog* paket već instaliran na računaru, jer ako nije pokreće se instalacija, a ako jeste nema pokretanje instalacije paketa iznova (predstavlja dobru praksu u pisanju programa u R programskom jeziku). Da bi se ovo automatski realizovalo za veći broj paketa, potrebno je liniju koda:

```
pkgs = c('dplyr')
```

zameniti tako da se umesto paketa *dplyr* dodaju i drugi paketi koji su ovde ilustrativno predstavljeni sa 'paket1' i 'paket2':

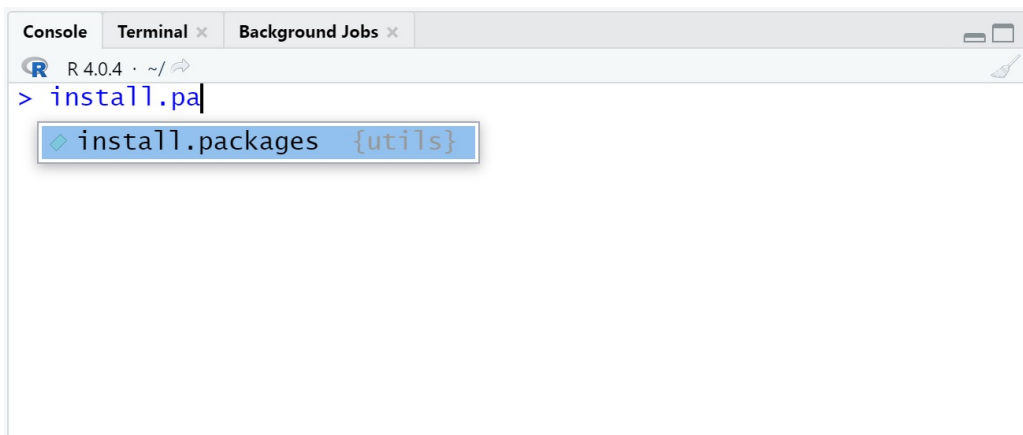
```
pkgs = c('dplyr', 'paket1', 'paket2')
```

Ima i drugih rešenja ovog problema, a svakako se nameće da je dobra dokumentacija koda (koja uključuje deljenje metapodataka kao što su datum kada je kod kreiran i verzije programskog jezika R, R Studio okruženja i paketa koje su korišćene) može da doprinese dugotrajnoj reproducibilnosti koda. Pravolinijsko rešenje ovog problema u računarstvu je upotrebom Docker kontejnera⁴⁰ (Docker, Inc, Palo Alto, SAD) koji uz pomoć virtuelizacije operativnog sistema omogućavaju da se kreira okruženje koje je jednako razvojnom za dalje testiranje koda. Za kraj, važno je spomenuti da je *groundhog* paket koristan u slučaju da neki paket nije više dostupan na CRAN-u ili da se ne održava [55]. Iako je vrlo koristan, *groundhog* paket nije svemoguć, jer postoje i drugi uzroci koji mogu dovesti do narušavanja računarske

40 <https://www.docker.com/resources/what-container/>, pristupljeno 30.07.2023.

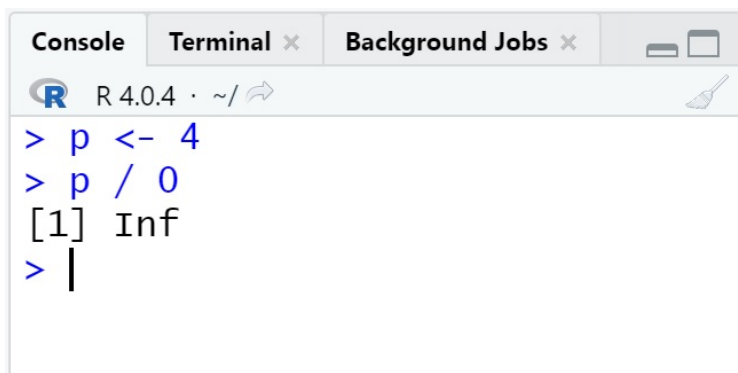
reproducibilnosti, kao što su: (1) neodgovarajuća dokumentacija programskih kodova, podataka i procedura (eng. *Workflow*), (2) nedostajuće funkcije i delovi koda i (3) nedostajući podatak o licenci pod kojom je kod podeljen [58]. U studiji iz 2018. godine u kojoj su Viktorija Stoden i saradnici predstavili ove nedostatke za računarsku reproducibilnost, na uzorku koji je ispitan pokazano je da preko 70% postojećih kodova ima jedan od navedenih uzroka i nije računarski reproducibilan [58].

Korišćenje slobodnog softvera kao što je R nije samo po sebi dovoljno za računarsku reproducibilnost, ali dobar odabir programskog jezika predstavlja stabilnu polaznu osnovu za pisanje reproducibilnog softvera. Od prednosti R-a izdvaja se *fast/rapid prototyping*, odnosno brže komponavanje programskih kodova. Posebno je privlačna ova prednost za početnike, bilo da se radi o početnicima u programiranju ili o početnicima u R-u. Ovo se ostvaruje kroz jednostavnije komande i ugrađene funkcije koje izvršavaju kompleksniji i veći broj operacija. U većini slučajeva, korišćenje takvih funkcija je efikasnije sa stanovišta računarskih resursa u odnosu na situaciju kada se programira takva funkcija samostalno. To je postignuto tako što su Autori paketa i funkcija višeg nivoa koristili metode za optimizaciju koda ili su kod pisali u drugim bržim programskim jezicima, kao što je jezik C [3]. Najveći broj korisnika programskog jezika R nisu softverski inženjeri i ne bave se razvojem programskih jezika, te je za njih ovo vrlo poželjna osobina gotovih funkcija zbog relativno brzog dolaska do rešenja tj. funkcionalnog koda. Dodatna pogodnost je automatsko popunjavanje koje je prikazano na Sl. 4 za pokretanje komande za instaliranje paketa u R Studio okruženju.



Slika 4, Prikaz konzole u okviru R Studio razvojnog okruženja sa komandom za instaliranje dplyr paketa sa automatskim popunjavanjem.

Dodatno, moglo bi se reći da je R programski jezik otporan na greške u poređenju sa drugim programskim jezicima. Primera radi, u R-u je moguće deljenje sa nulom i vrlo često na mestima gde bi došlo do greške i ujedno do zaustavljanja izvršavanja skripte u drugom programskom jeziku, R može javiti samo upozorenje, dok se izvršavanje koda nastavlja. Rezultat deljenja sa nulom u programskom jeziku R je prikazano na Sl. 5.

The image shows a screenshot of the R Studio interface. At the top, there are three tabs: 'Console', 'Terminal', and 'Background Jobs'. The 'Console' tab is active. Below the tabs, the R logo and version 'R 4.0.4' are visible, along with a home icon and a refresh icon. The console contains the following text:

```
> p <- 4
> p / 0
[1] Inf
> |
```

Slika 5, Prikaz konzole u okviru R Studio okruženja sa komandom koja omogućava deljenje sa nulom promenljive p koja se razlikuje od nule. Primiti da u ovom slučaju R ne javlja ni grešku ni upozorenje, već za rezultat daje beskonačnu vrednost tj. *Inf* od eng. *Infinitive*.

Osnove programskog jezika R koje su prikazane u ovom udžbeniku nemaju za cilj da uče studente programiranju. Pretpostavlja se da su studenti tokom osnovnih akademskih studija savladali sve osnove programskih jezika generalno. Udžbenik ima za cilj da pomogne studentima da razumeju osnovne osobine programskog jezika R sa ciljem da mogu samostalno da analiziraju biosignale u programskom okruženju i da kreiraju profesionalne kodove i izveštaje prateći dobre prakse u industrijskom i akademskom okruženju. Najveći deo ovog poglavlja koje se bavi osnovama rada u R-u zasnovano je na udžbeniku profesora Rodžera D. Penga (Rodger D. Peng) sa Džon Hopkins Univerziteta⁴¹ iz SAD [3] koji takođe drži i online kurs pod nazivom “*R Programming*” koji je deo specijalizacije iz oblasti nauke o podacima. Kurs je dostupan na Coursera⁴² platformi za učenje na daljinu [59]. Svima koji polože kurs iz TOBS predmeta i onima kojima se dopada ova knjiga, a žele još da prodube svoja znanja, preporučuje se slušanje ovog kursa koje je besplatno.

Iskustvo je pokazalo da učenje novog programskog jezika ima sličnosti i sa učenjem prirodnih jezika. Na početku procesa treba očekivati da će se dešavati

41 John Hopkins University, <https://www.jhu.edu/>, pristupljeno 27.05.2024.

42 <https://www.coursera.org/>, pristupljeno 30.07.2023.

da se pomeša kod već poznatog programskog jezika i novog jezika koji se uči, kao i da neće svaki segment koji je potrebno savladati biti podjednake težine⁴³. Međutim, to je prolazna faza učenja te će vrlo brzo uporni korisnici moći nezavisno da koriste sve programske jezike koje znaju.

Studenti koji su pohađali TOBS predmet na Elektrotehničkom fakultetu, Univerziteta u Beogradu su najčešće imali problema sa sličnostima i razlikama između R-a i GNU Octave (GNU Octave, Copyright (C), The Octave Project Developers) [60] i Matlab programskih jezika. Uobičajeno, greške su nastajale prilikom indeksiranja nizova i prilikom kreiranja nizova. Međutim, kako se u R-u pretežno koriste tabelarni podaci tipa *data frame*, a ne nizovi i matrice kao u Matlab-u i GNU Octave-u to se vrlo brzo ovakve greške prevaziđu i ne ponavljaju. Jedna od najvećih razlika je u znaku dodele vrednosti, što je ilustrovano na Sl. 5 prilikom dodele vrednosti 4 promenljivoj p . Iako je u R-u omogućeno da se kao ekvivalent znaku dodele vrednosti može koristiti znak jednakosti (=), znak dodele vrednosti (<-) je logičniji izbor, jer se promenljivoj p dodeljuje vrednost 4, odnosno 4 se upisuje u memorijsku lokaciju promenljive p , te je zapis prirodniji. Nema pravila šta je bolje koristiti, ali je važno da studenti koji se opredele da koriste znak jednakosti mogu da razumeju kod koji je pisao neko ko koristi znak dodele vrednosti (<-) u R-u. Za one koji znaju Matlab / GNU Octave i žele da nauče programski jezik R (ili obrnuto), postoji knjiga iz 2015. godine koja je specifično namenjena takvim korisnicima i koja pomoću karakterističnih sličnosti i razlika ovih programskih jezika omogućava korisnicima da jednostavnije savladaju novi programski jezik [61]. Takođe, knjiga [61] je zvanično preporučena na stranici Matlaba⁴⁴.

Sledeća potpoglavlja sadrže konkretna uputstva za rad u R-u i pregled osnovnih funkcionalnosti koje su u velikoj meri realizovane po ugledu na udžbenik [3], ali prilagođene studentima Elektrotehničkog fakulteta, Univerziteta u Beogradu i primeni R-a u biomedicinskom inženjerstvu. Svim studentima koji su došli do ove stranice i još uvek nisu sigurni da li im se predmet dopada ili su sigurni da ne žele da slušaju TOBS predmet, pravo je vreme da od toga odustanu. TOBS je izborni predmet – studenti biraju nastavnika i gradivo, a ne obrnuto. Za sve čitaoce koji se zgražavaju nad idejom slobodnog softvera, dobrih praksi u pisanju programskih kodova i nad primenom R-a u nauci o podacima i biomedicinskom inženjerstvu, sledi upozorenje da će ovi koncepti biti prikazani

43 Ovde treba imati na umu individualne razlike. Kod svake osobe postoje elementi koji su jednostavniji ili kompleksniji za učenje. Iskustva drugih su korisna, ali ne treba se uvek na njih oslanjati.

44 R and Matlab, <https://www.mathworks.com/academia/books/r-and-matlab-hiebeler.html>, pristupljeno 30.07.2023.

vrlo detaljno i temeljno, te je ovo pravo vreme da i oni odustanu od čitanja udžbenika i da sebe ne izlažu dodatnoj nelagodnosti. Studenti i čitaoci kojima se nakon uvodnog dela dopada ideja da se dalje usavršavaju i vide priliku da nauče i iskoriste u praksi nova saznanja, savetuje se da što pre uzmu računar u ruke i da instaliraju R programski jezik i R Studio razvojno okruženje na svom računaru. Dodatak B ovog udžbenika sadrži kompletan spisak vežbi sa TOBS predmeta, a prezentacija za prvu vežbu se sastoji iz upustva za instalaciju. Dodatak C sadrži detaljno upustvo za rad u R Studio razvojnom okruženju. Kodovi koji su prikazani u ovom udžbeniku su javno dostupni pod licencom slobodnog softvera na <https://github.com/NadicaSm/TOBS>. Srećno!

1.1 Tipovi podataka u R-u

U R-u se podaci dele na atomske tj. osnovne objekte (numeričke, logičke, celobrojne i kompleksne), nizove, liste, faktore (kategoričke podatke), nedostajuće vrednosti i tabelarne podatke (eng. *data frame*). Za primenu R-a u nauci o podacima i biomedicinskom inženjerstvu, tabelarni podaci su od najvećeg značaja⁴⁵. U R-u objekti mogu imati svoje atribute tj. metapodatke (podatke o podacima) koji služe da se opiše sadržaj promenljivih. Najveća praktična korist atributa je što mogu na jednostavan način da povežu fizički smisao sa promenljivom. Primeri takvih atributa su ime promenljive, ime(na) dimenzija, dimenzije i klase. Ako bi se pregledali metapodaci promenljive koje nemaju unete atribute, u R-u bi bilo ispisano NULL što bi značilo da atribut ne postoji.

Brojevi u R-u su predstavljeni podrazumevano kao realni brojevi, a za eksplicitan prikaz celobrojnih vrednosti potrebno je dodati slovo L brojnoj vrednosti. Za razliku od većine programskih jezika, u R-u je deljenje nulom validna operacija koja za rezultat daje *Inf* tj. beskonačnost (Sl. 5). Važi i obrnuto – deljenje sa *Inf* daje za rezultat nulu. Iako deljenje nulom ima svojih prednosti, biomedicinski inženjeri moraju voditi računa da greška ne propagira kroz program usled toga što je R neće javiti. Slično kao i u drugim jezicima, u R-u postoji NaN (eng. *Not A Number*) i može se dobiti kao rezultat određenih operacija kao što je 0/0.

Logičke promenljive se mogu označiti na više načina, a u R-u je uobičajen zapis velikim slovima TRUE i FALSE. Jedna od osnovnih odlika R-a je *fast prototyping* ili brzo generisanje/komponovanje koda, te su u R-u validni i

⁴⁵ Može se reći da je ovaj tip podataka jedna od osnovnih odlika R-a, kao što su matrice za Matlab i GNU Octave ili liste za Pajton programski jezik.

skraćeni zapisi logičkih promenljivih: T i F. Iako je ugodno koristiti skraćene zapise, preporuka je da se koristi kompletan zapis logičkih promenljivih zbog povećane čitljivosti koda.

Niz podataka može sadržati isključivo podatke koji su istog tipa, a može se definisati na više načina:

- korišćenjem operatora dvotačka “:” za celobrojne vrednosti⁴⁶ (eng. *integer*),
- korišćenjem *vector()*⁴⁷ funkcije kojom se čak može kreirati i prazan niz,
- primenom *c()* funkcije od eng. *combine* koja je najzastupljenija u R-u za kreiranje nizova⁴⁸,
- *seq()* funkcijom kojom se mogu kreirati sekvence realnih brojeva (pogodna za kreiranje vremenske ose biosignala) i
- na druge načine (pogledati funkcije *rep()* ili *seq_along()*).

U slučaju da se dogodi da se prilikom kreiranja niza zadaju vrednosti koje pripadaju različitim klasama doći će do implicitne tj. automatske konverzije koja se obavlja po receptu:

Logic → *Numeric* → *Character*

Primeru radi, na Sl. 6 je prikazano šta će se desiti ako se definišu tri promenljive tipa niz (*niz1*, *niz2* i *niz3*) sa elementima koji pripadaju različitim klasama. Primitivni da je za prikaz sadržaja promenljive dovoljno samo da se ispiše naziv te promenljive u konzoli⁴⁹.

46 U R-u nije moguće kao u Matlabu i GNU Octave-u definisati nizove realnih brojeva pomoću “:” operatora, već isključivo nizove koji sadrže celobrojne vrednosti.

47 Uobičajeno se u literaturi za R programski jezik koriste otvorena i zatvorena zagrada kako bi se naznačilo da se u tekstu govori o funkcijama, te će na taj način biti predstavljene funkcije u ovom udžbeniku. Takođe, pozivanje funkcija bez ulaznih parametara ima isti zapis.

48 Ova funkcija je već bila iskorišćena u uvodnom delu poglavlja za formiranje niza naziva paketa nad kojima se želi primeniti groundhog paket koji omogućava računarsku reproducibilnost.

49 Za razliku od Matlabu i GNU Octave-a (isto važi i za programske jezike C i C++) u kojima se mora dodati operator “;” tačka-zarez kako bi se sprečio ispis sadržaja komande (ili čak omogućio pravilan poziv komande), u R-u to nije slučaj. Nije potrebno dodavati ovaj operator u R programima. Time se ujedno skraćuje vreme potrebno za pisanje koda.

```
Console Terminal × Background Jobs ×
R 4.0.4 · ~/ ↵
> niz1 <- c(TRUE, FALSE, 15, 13, "c", "tobs")
> niz1
[1] "TRUE" "FALSE" "15"    "13"    "c"     "tobs"
> class(niz1)
[1] "character"
> niz2 <- c(TRUE, FALSE, 15, 13)
> niz2
[1] 1 0 15 13
> class(niz2)
[1] "numeric"
> niz3 <- c(TRUE, FALSE, "c", "tobs")
> niz3
[1] "TRUE" "FALSE" "c"     "tobs"
> class(niz3)
[1] "character"
> |
```

Slika 6, Implicitna konverzija za elemente tri niza. Obratiti pažnju na korišćenje funkcije `class()` za proveru klase elemenata u nizu. Primetiti da je prilikom ispisa sadržaja nizova u konzoli korišćen implicitni prikaz pozivom imena promenljive, a da nije korišćen eksplicitni prikaz pomoću `print()` funkcije.

Zadavanje vrednosti koje pripadaju različitim klasama jednom nizu je teorijska pretpostavka koja se u praksi nikada ne dešava iako se spominje u udžbenicima za učenje programskih jezika [3]. Međutim, može se desiti da rezultat neke operacije koji se smešta u nekom momentu dobije vrednost koja nije očekivana, te da onda nastupi konverzija. Preporuka je da se ovakve greške izbegavaju i da se prilikom debugovanja (procedure verifikacije koda koja služi za otklanjanje grešaka) obavezno proverava sadržaj okruženja u kome su smeštene promenljive tj. klase promenljivih i vrednosti koje se u njima nalaze. Pored implicitne, postoji i eksplicitna konverzija. U radu sa biomedicinskim signalima i podacima, konverzija se koristi najčešće prilikom učitavanja podataka kada je potrebno prilagoditi tip podataka planiranoj analizi. O tome će biti kasnije reči. Funkcije koje se koriste za konverziju imaju prefiks “*as*”, te, na primer, za konverziju u faktorsku promenljivu koristi se funkcija:

`as.factor()`

U slučaju da nije moguće razrešiti kako da se izvrši konverzija, kao, na primer, iz karaktera u kompleksne brojeve, R neće javiti grešku, nastaviće sa

izvršavanjem koda, javiće samo upozorenje, a rezultat će biti popunjen sa NA (eng. *Not Available*), odnosno nedostajućim vrednostima. Kako bi se testirala ova funkcionalnost R-a potrebno je pokrenuti sledeći kod za prethodno kreiran niz koji se nalazi u promenljivoj *niz3*:

```
as.logical(niz3).
```

Jedna od pogodnosti koja je uključena u novije verzije R Studio razvojnog okruženja je automatsko prepoznavanje stringova koji predstavljaju nazive boja. Izgled R Studio razvojnog okruženja za definisanje niza koji se sastoji od naziva boja je prikazan na Sl. 7. O vizuelizaciji signala i odgovarajućim grafičkim prikazima biće više detalja prikazano u kasnijim poglavljima, ali je ovde važno napomenuti da je za izbor boja i kombinacije boja korisno pogledati katalog boja “*Colors in R*”⁵⁰ sa Kolumbija univerziteta u SAD koje je pripremila profesorka Jing Vei.

```
c("aquamarine2", "red", "blue", "chocolate",  
  "goldenrod", "deeppink")
```

Slika 7, Primer defnisanja niza koji sadrži nazive boja u R Studio razvojnem okruženju. Primititi da je moguće prikazati poziv neke funkcije ili komandu u R-u jednostavnim prelaskom u novi red. Razvojno okruženje automatski dodaje tabulaciju tj. početak novog reda koji je uvučen kako bi kod bio čitljiviji.

Poznato je da za razliku od nizova, lista može sadržati podatke koji pripadaju različitim klasama (klasu tj. tip podatka moguće je proveriti primenom:

```
class()
```

funkcije – pogledati upotrebu ove funkcije na Sl. 6) i kreira se istoimenom:

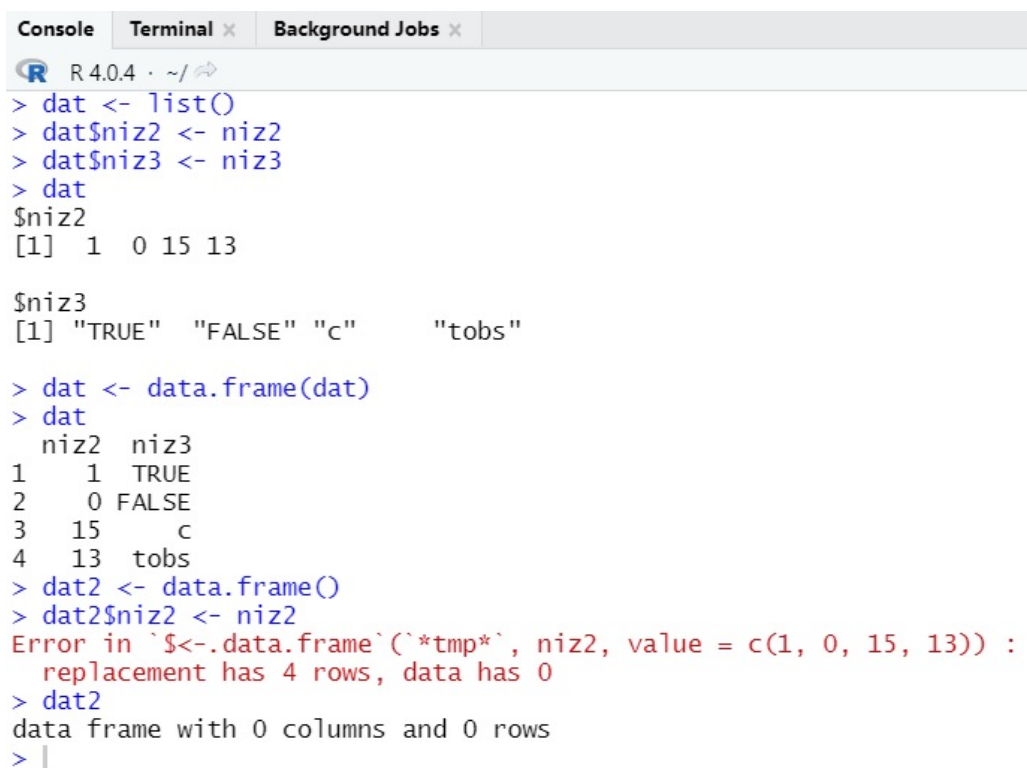
```
list()
```

funkcijom. Kako se u R-u se najčešće koriste tabelarni podaci tipa *data frame*, za kreiranje novih *data frame*-ova, mogu se koristiti liste, kao što je prikazano na Sl. 8. Naravno, ovo nije jedini način kako se može kreirati novi *data frame*, ali jeste relativno jednostavan. Nema posebnog razloga za ovakav pristup, osim ako se ne želi postići dodatna manipulacija podacima. Jedna od prednosti listi je što ne moraju svi elementi liste da budu iste dužine te se tokom analize pruža

50 <http://www.stat.columbia.edu/~tzheng/files/Rcolor.pdf>, pristupljeno 30.07.2023.

prilika korisniku da zadrži podatke tipa liste umesto tipa *data frame*-a, a elementi *data frame*-a se mogu jednostavnije definisati kao što je prikazano na Sl. 8. Ovde je važno dati dve napomene za početnike:

- Pored *class()* funkcija, može se koristiti i *str()* funkcija za proveru strukture objekta. Pored primene na objektima, *str()* se može primeniti i na funkcijama.
- Na Sl. 8 je prikazano korišćenje operatora “\$” za pristup kolonama tabelarnih podataka o kome će kasnije biti više reči.



```
Console Terminal x Background Jobs x
R 4.0.4 · ~/
> dat <- list()
> dat$niz2 <- niz2
> dat$niz3 <- niz3
> dat
$niz2
[1] 1 0 15 13

$niz3
[1] "TRUE" "FALSE" "c" "tobs"

> dat <- data.frame(dat)
> dat
  niz2 niz3
1    1 TRUE
2    0 FALSE
3   15    c
4   13 tobs
> dat2 <- data.frame()
> dat2$niz2 <- niz2
Error in `<-data.frame`(`*tmp*`, niz2, value = c(1, 0, 15, 13)) :
  replacement has 4 rows, data has 0
> dat2
data frame with 0 columns and 0 rows
> |
```

Slika 8, Primer kreiranja *data frame*-a pomoću liste i *data.frame()* funkcije koje imaju za elemente nizove prikazane na Sl. 6. Obratiti pažnju da je značajno intuitivnije kreiranje tabelarnog tipa podataka (eng. *data frame*) i dodavanje odgovarajućih kolona tabelarnih podataka primenom operatora “\$” pomoću liste.

Logično proširenje nizova je matrica, a matrice u R-u su predstavljene kao nizovi sa atributima tj. metapodatkom o dimenziji. Atribut predstavlja celobrojnu vrednost dužine dva (sa ove dve dimenzije su predstavljene vrste i kolone), a mogu se pogledati primenom funkcije:

attributes()

Za pristup pojedinim elementima treba imati na umu da su matrice u R-u *column-wise* tj. prvo se popunjavaju kolone, pa tek onda redovi. Malo je verovatno da će neko u praksi kreirati matricu tako što će ručno unositi element po element, a posebno ne u R-u u kome je primarno zastupljen rad sa tabelarnim podacima. Međutim, postoje funkcije koje mogu koristiti matrice. Na primer, ako su odbirci višekanalnog EMG signala smešteni u matricu tako što je u svakom redu po jedan kanal, ako se primeni filter funkcija filtriraće se kolone koje apsolutno ništa ne znače i od kojih svaka sadrži po jedan odbirak. Ovo je greška koja se može vrlo lako javiti i u Matlabu ili GNU Octave-u, te se svakako savetuje da se proverí sadržaj međurezultata prilikom evaluacije koda.

```
> m1 <- matrix(37:44, ncol = 4, nrow = 2)
> m1
      [,1] [,2] [,3] [,4]
[1,]  37  39  41  43
[2,]  38  40  42  44
> dimnames(m1)
NULL
> dimnames(m1) <- list(c("vrsta1", "vrsta"),
+                      c("kol1", "kol2", "kol3", "kol4"))
> m1
      kol1 kol2 kol3 kol4
vrsta1  37  39  41  43
vrsta   38  40  42  44
> attributes(m1)
$dim
[1] 2 4

$dimnames
$dimnames[[1]]
[1] "vrsta1" "vrsta"

$dimnames[[2]]
[1] "kol1" "kol2" "kol3" "kol4"
```

Slika 9, Primer kreiranja matrice i realizacije osnovnih operacija nad matricama za pregled i dodelu atributa u konzoli R Studio okruženja.

Kao i nizovi, matrice se mogu kreirati na više načina. Prazne matrice je jednostavno realizovati primenom:

matrix()

funkcije, ali i spajanjem u redove i kolone postojećih nizova uz funkcije *cbind()* od eng. *column bind* ili *rbind()* od eng. *raw bind*. Matrice se mogu kreirati i dodelom dimenzije nizovima primenom funkcije *dim()*. Primer kreiranja matrice sa pregledom i dodelom atributa matricama je prikazan na Sl. 9 (primer je realizovan po ugledu na vežbu 2 iz TOBS predmeta – materijali za vežbe i spisak vežbi su dati u Dodatku B na kraju udžbenika).

Kategoričke promenljive koje se najčešće nazivaju faktorima ili faktorskim promenljivima su veoma važne za analizu biomedicinskih podataka, ali i za primenu mašinskog učenja. Poznato je da algoritmi mašinskog učenja, a posebno oni koji su zasnovani na *bootstrap* metodi koja predstavlja proces reodabiranja tj. rekombinovanja podataka kao što je to metod slučajnih šuma (eng. *Random Forest*, skraćeno RF) mogu davati neočekivano lošije rezultate, ako prethodno nije izvršena odgovarajuća konverzija promenljivih u kategoričke promenljive [62]. Tipičan primer podataka koje je potrebno predstaviti pomoću kategoričke promenljive u biomedicinskom inženjerstvu je podatak o polu⁵¹ ispitanika. Kategoričke promenljive tj. faktori se mogu kreirati pomoću funkcije:

factor()

pri čemu treba obratiti pažnju na *levels* ulazni parametar koji služi da se na željeni način rasporede vrednosti kategoričkih promenljivih. Na Sl. 10 je prikazano kako se kreira faktorska promenljiva koja prikazuje radne dane u sedmici. U praksi, najčešće se *factor()* promenljiva i ne koristi, već se ulazni podaci, ako za to postoji potreba, transformišu eksplicitno u kategoričku promenljivu primenom *as.factor()* funkcije.

51 Ne treba pol (eng. *sex*) primenjivati kao sinonim za rod (eng. *Gender*), jer je u udžbeniku korišćena opšteprihvaćena i tradicionalna podela na pol i rod gde pol podrazumeva biološke karakteristike, a rod unutrašnju svest osobe o sopstvenom identitetu [63]. Za detalje o terminološkim i suštinskim razlikama preporučuje se [64]. Ako se radi o medicinskoj studiji, pol može biti vrlo često važan parametar, upravo zbog anatomskih i funkcionalnih razlika muškog i ženskog tela. Za veliki broj medicinskih parametara postoje posebni opsezi za muškarce i žene, koji se nameću zbog prirodnih razlika. Novijeg datuma su uputstva za objavljivanje biomedicinskih studija u naučnim časopisima. Primera radi, časopis *Computers in Biology and Medicine* se oslanja na SAGER (od eng. *The Sex and Gender Equity in Research*) [65] preporuke za objavljivanje studija koje se bave razlikama u polu i rodu, gde se posebno preporučuje da naučnici naglase koju definiciju pola i roda su usvojili u naučnom radu koji objavljuju (<https://www.elsevier.com/journals/computers-in-biology-and-medicine/0010-4825/guide-for-authors>, pristupljeno 30.07.2023.).


```

> f1 <- as.factor(c("pon", "ut", "sr", "cet", "pet"))
> f1
[1] pon ut sr cet pet
Levels: cet pet pon sr ut
> f1 <- factor(c("pon", "ut", "sr", "cet", "pet"),
+             levels = c("pon", "ut", "sr", "cet", "pet"),
+             ordered = TRUE)
> f1
[1] pon ut sr cet pet
Levels: pon < ut < sr < cet < pet

```

Slika 10, Kreiranje faktorske/kategoričke promenljive i ilustracija korišćenja *levels* argumenta za promenu redosleda nivoa tj. hronološki redosled dana u nedelji. Skraćenice pon, ut, sr, cet i pet odgovaraju radnim danima u sedmici: ponedeljak, utorak, sreda, četvrtak i petak.

Za skladištenje tabelarnih podataka (koji mogu sadržati i numeričke i kategoričke podatke po kolonama) u R-u postoji *data frame* koji predstavlja osnovni i najvažniji tip podataka u R-u. Za pojednostavljen rad sa *data frame* tipom podataka razvijen je dplyr (od eng. *deep layer*) paket [38]. *Data frame* se definiše kao lista u kojoj svaki element ima jednaku dužinu, a podaci mogu pripadati različitim tipovima osnovnih/atomskih promenljivih. Atributi predstavljaju najznačajniji deo *data frame*-a, te se većina operacija u R-u (što je velika prednost ovog programskog jezika) obavlja nad podacima iz *data frame*-a pozivom koji uključuje naziv kolona i pomoću operatora “\$”. Već je pokazano (Sl. 8) kako se pomoću *list()* funkcije može kreirati *data frame*, ali važno je spomenuti da postoji i specijalizovana *data.frame()* funkcija.

Za pristup pojedinim elementima *data frame*-a koriste se odgovarajući operatori, a najčešće su u upotrebi “[]”, “[[]]”, kao i najzastupljeniji “\$” [20]. Upotreba operatora “\$” je prikazana na Sl. 11. Jednostruke pravougaone zagrade “[]” služe da se odredi podskup elementa koji je isti tip podataka kao originalni (na primer, da se odredi lista koja je podskup liste) ili da se pristupi pojedinim elementima u nizu. Kada se radi sa matricama, operator “[]” može imati ulazni parametar *drop* koji je podrazumevano TRUE (rezultat operacije indeksiranja je niz), ali ako se postavi na FALSE, tada rezultat indeksiranja postaje matrica. Dvostruke uglaste zagrade “[[]]” se koriste za određivanje podskupa dimenzije jednog elementa u listi ili *data frame*-a, a podaci mogu i ne moraju biti istog tipa kao originalni podaci (to zavisi od tipa podataka konkretnog elementa). Ovaj operator je isti kao “\$” operator s tom razlikom što “\$” operator može da izdvoji više od jednog elementa i moguće im je pristupiti po imenu. Operator “\$” se ne može koristiti sa podacima tipa string, ali se može koristiti sa imenom kolona što je prikazano na Sl. 11.

```

Console Terminal x Background Jobs
R 4.0.4 · ~/ ↗
> dat$niz2
[1] 1 0 15 13
> naziv_kolone = "niz2"
> dat$naziv_kolone
NULL
> |

```

Slika 11, Rad sa operatorom “\$” za izdvajanje kolona iz *data frame*-a po imenu. Primititi da se podatak tipa string koji ima naziv kolone ne može koristiti sa ovim operatorom, kao i da R ne javlja grešku, već samo kao rezultat operacije daje NULL. Iako je ignorisanje ovakvih operacija (ne izaziva prekid rada programa) poželjna osobina, jer čini R kod robusnim na neočekivane rezultate, ujedno je i mana, jer je na korisniku da spreči da do ovakvih rezultata ne dođe.

```

Console Terminal x Background Jobs x
R 4.0.4 · ~/ ↗
> if (!require("ISwR")) install.packages("ISwR")
> library(ISwR)
> pod <- stroke
> head(pod, 2)
  sex      died      dstr age dgn coma diab minf han  dead  obsmonths
1 Male 1991-01-07 1991-01-02  76 INF  No  No  Yes  No  TRUE  0.1633987
2 Male      <NA> 1991-01-03  58 INF  No  No  No  No FALSE 59.6078431
> class(pod$sex)
[1] "factor"
> sum(pod$sex == "Female")
[1] 510
> length( pod$sex[pod$sex == "Female"] )
[1] 510
> |

```

Slika 12, Logičko indeksiranje. Primer je preuzet sa testa iz TOBS predmeta koji je održan 10.04.2017. godine za zadatak “Iz biblioteke "ISwR" izdvojiti podatke iz studije "stroke". Koliko ima žena snimljenih u ovoj studiji?”. Prikazana su dva načina koja koriste logička indeksiranja. Naravno, ima ih i više. Primititi da prva linija koda označava *if* strukturu o kojoj će kasnije biti više reči, a da ovde služi za instaliranje paketa u slučaju da ne postoje na računaru na kome se pokreće kod. Funkcija *head()* služi za prikaz početnih redova tabelarnih podataka. Podrazumevano prikazuje 6 redova, ali ovde su radi jednostavnosti prikazana samo dva reda (zato je i ulazni parametar *head()* funkcije postavljen na 2).

Izdvajanje kolone *data frame*-a po imenu (pogledati Sl. 11) doprinosi i većoj čitljivosti koda i već spomenutom *fast prototyping*-u koji odlikuje R programski jezik. Indeksi koji se koriste za izdvajanje pojedinih elemenata mogu, kao i u većini programskih jezika, biti numerički i logički. Ovi drugi su posebno važni, za izdvajanje pojedinačnih elemenata ili grupa koje ispunjavaju određeni uslov.

Na Sl. 12 dat je primer logičkog indeksiranja sa podacima iz ISwR paketa [36] koje ima za cilj da proveri koliko je osoba ženskog pola prisutno u skupu podataka koji se analizira. *Fast prototyping* se ogleda i u delimičnom poklapanju naziva kolona u R-u, tj. naziv kolone se može samo delimično uneti u R, a programski jezik R će prepoznati o kojoj koloni je reč. S obzirom na postojanje opcije automatskog popunjavanja, teško da će ikada biti potrebno da se određene kolone izdvajaju primenom delimičnog poklapanja po imenu.

Posebna klasa podataka u R-u je namenjena za rad sa datumom i vremenom što u nekim situacijama može da pojednostavi analizu podataka, ako su, na primer, umesto godina ispitnika dostupni datumi ili godine rođenja i datumi ili godine merenja; ili za rad sa vremenskim oznakama (eng. *timestamps*) koje su značajne prilikom sinhronizacije merenja sa više uređaja. Datumi su u R-u predstavljeni pomoću *Date* klase podataka i svi datumi su smešteni kao redni brojevi dana od 01.01.1970. kome odgovara redni broj 1. Ako string ima odgovarajuću strukturu moguće je primeniti *as.Date()* funkcije da bi se dobio podatak klase *Date*. Na Sl. 13 ilustrovan je rad sa funkcijama za datum i jednostavna operacija računanja broja proteklih dana između dva datuma. Takođe, na Sl. 13 je prikazana upotreba funkcije *Sys.time()* za trenutno vreme (postoji i funkcija *Sys.Date()* za trenutni sistemski datum na računaru).

Za predstavljanje vremena postoje dve dostupne klase podataka *POSIXct* sa celobrojnim vrednostima i *POSIXlt* kao lista koja sadrži metapodatke (npr. dani u nedelji). Slično kao datumi, vremena su smeštena kao redni brojevi sekundi od 01.01.1970. Postoje specifične funkcije koje se mogu koristiti sa ovim tipovima podataka kao što su *weekdays()* i *months()*. Ovde je važno spomenuti funkciju *Sys.time()* koja ima mogućnost da daje trenutno vreme u formi klase *POSIXct*. U praksi, jednostavno se ovladava svim operacijama i funkcijama za rad sa vremenom, a najveći problem (ili izazov⁵²) predstavlja konverzija postojećih datuma i vremena u podacima u format kakav postoji u R-u. Kako bi se efikasnije realizovala transformacija u željenu klasu može se koristiti funkcija *strptime()*. Iako ova funkcija može da automatizuje promenu formata, ona ipak ne može da reši sve probleme, a posebno ne one koji mogu nastati kao rezultat pogrešnog ručnog unosa. Na žalost u takvoj situaciji najverovatnije se

52 U trenutku pisanja ovog udžbenika, u velikom broju kompanija u Srbiji i šire praksa je da se reč “problem” ne koristi, već da se umesto ove reči usvoji “izazov” u svakodnevnoj komunikaciji. Autorka udžbenika smatra da je to prolazna i beskorisna korporacijska kultura, te da bez obzira što je usvojena u industriji ne treba da nalazi uporište u akademiji. Sa druge strane studenti treba da budu pripremljeni za rad u industriji, posebno prilikom pisanja izveštaja i predstavljanja rezultata analize biosignala, te je ovo možda korisna informacija za njihov rad u novom kolektivu.

samo ručno greška može ispraviti ili se vrednost može automatski proglasiti za nedostajuću. U nekim slučajevima nije ni moguće utvrditi pogrešan unos. Iako je u R-u odlučeno da se godine broje od 01.01.1970. treba znati da se u različitim kalendarima pojavljuje i nulta godina. U lat. *Anno Domini* eri koju označavaju julijanski i gregorijanski kalendari nulta godina ne postoji. Međutim, u astronomskoj predstavi godina i u ISO 8601:2004⁵³ standardu za reprezentaciju datuma i vremena postoji 0-ta godina⁵⁴.

```
> Sys.time()
[1] "2023-07-30 16:01:47 CEST"
> datum1 <- as.Date("2023-06-16")
> datum2 <- as.Date("2023-07-30")
> datum2 - datum1
Time difference of 44 days
> |
```

Slika 13, Korišćenje funkcija u R-u za rad sa datumom, Računato je proteklo vreme u danima od trenutka poslednje stabilne verzije programskog jezika R do trenutka pisanja ovog dela udžbenika. Verzije bilo kog slobodnog softvera, a posebno programskog jezika R su vrlo česte, što se može videti iz priloženog primera. Čitaocima se ostavlja da provere u trenutku čitanja udžbenika i pokretanja ovog koda broj dana od poslednje stabilne verzije.

Nazivi promenljivih bi trebalo, kao i u bilo kom drugom programskom jeziku, da budu opisni, te da na taj način doprinesu ukupnoj čitljivosti koda. Očekivano, nazivi promenljivih mogu biti jedno slovo, gde se pri tom razlikuju velika i mala slova, odnosno *y* i *Y* predstavljaju različite promenljive. U R-u se objekat i funkcija mogu zvati isto, jer je različiti prostor rezervisan za funkcije i promenljive. [3]

Kako je u ranijim primerima već pokazano u R-u postoje posebne oznake i tipovi podataka za nedostajuće vrednosti. Nedostajuće vrednosti mogu biti NA ili NaN (od eng. *Not A Number*). Dok NA vrednosti imaju sopstvenu klasu (celobrojne vrednosti ili *integer*, karakteri i sl), NaN jeste NA, ali konverzija iz NaN u NA nije moguća. Vrlo često korišćene funkcije tokom procesa čišćenja podataka o kom će kasnije biti više reči su:

`is.na()` i

53 ISO 8601:2004(en) Data elements and interchange formats – Information interchange – Representation of dates and times, <https://www.iso.org/obp/ui/#iso:std:iso:8601:ed-3:v1:en>, pristupljeno 30.07.2023.

54 https://en.wikipedia.org/wiki/Year_zero, pristupljeno 30.07.2023.

`is.nan()`

koje omogućavaju da se proveri na datom skupu da li NA i NaN postoje u datom skupu. Rezultat ovih funkcija je logička promenljiva tj. niz logičkih promenljivih. Rad sa nedostajućim vrednostima ne može zaobići nijednog biomedicinskog inženjera koji se bavi analizom podataka. Postojanje ovog tipa promenljivih značajno olakšava rad u R-u. S obzirom na značaj koji ima pravilna manipulacija nedostajućim podacima naredno potpoglavlje je posvećeno nedostajućim vrednostima i analizi podataka uprkos nedostajućim vrednostima.

1.1.1 O nedostajućim vrednostima

Nedostajuće vrednosti je nemoguće izbeći, a sa njima nije jednostavno analizirati podatke. One predstavljaju prvi u nizu problema sa kojima se sreću biomedicinski inženjeri, jer vrlo pažljivo treba pristupiti radu sa nedostajućim vrednostima. Iskustvo je pokazalo da se na kraju analize signala treba vratiti na ovaj početni korak te proveriti kako rukovanje nedostajućim vrednostima (bilo da se radi o smanjenju skupa podataka ili zameni nedostajućih vrednosti) utiče ili ne utiče na krajnji rezultat analize. U najboljem slučaju, rezultat ne bi trebalo da se promeni, ali to na žalost nije uvek slučaj. Dobra praksa nalaže da se provera uticaja nedostajućih vrednosti na rezultat analize uvek detaljno dokumentuje.

Pokazano je u literaturi da za podatke gde je procenat nedostajućih vrednosti manji od 10% ili 15% nema potrebe primenjivati nikakvu metodu zamene nedostajućih vrednosti, već da je dovoljno raditi analizu na preostalim 90% ili 85% podataka [66]. Drugi Autori preporučuju da se podaci mogu brisati ako nedostaje manje od 5% podataka i da ih tada nije potrebno menjati aproksimativnim vrednostima [67]. Ovu praktičnu preporuku treba uzeti sa rezervom, jer u zavisnosti od veličine skupa podataka (u biomedicinskom inženjerstvu se obično radi sa merenjima na ispitanicima, a takva merenja su zahtevna vremenski, finansijski, nekada i zdravstveno, jer u određenim situacijama pacijenti nemaju strpljenja ni snage da učestvuju u istraživanjima, a dodatno merenja mogu predstavljati rizik od širenja infektivnih bolesti i po istraživače i po pacijente) te 15% ne mora da bude uvek zanemariv deo podataka. Relativno velikim skupom podataka se najčešće smatra skup sa više od 100 parametara (kolona u *data frame*-u) i sa više od 100 000 observacija (vrsta u *data frame*-u) [66]. Proces popunjavanja, odnosno zamene nedostajućih vrednosti drugim vrednostima, može se realizovati primenom neke statističke

metode (zamena srednjom vrednošću, medijanom, *mode* zamena⁵⁵ ili primenom regresije) ili primenom metoda mašinskog učenja (na primer, primenom neuralne mreže) [66]. Zamena nedostajućih vrednosti procenjenim vrednostima se naziva na engleskom *Missing Value Imputation* (MVI). Očekivano, metoda zamene srednjom vrednošću je jednostavna i među najčešće korišćenim statističkim metodama, dok je k-NN (eng. *k-Nearest Neighbors*) najpopularnija metoda mašinskog učenja za zamenu nedostajućih vrednosti posebno u radu sa medicinskim podacima i sa podacima koji su dostupni na UCI repozitorijumu (eng. *UC Irvine Machine Learning Repository*)⁵⁶ [66]. Ovo donekle i ne iznenađuje, jer je intuitivno k-NN metoda odgovara proceni nedostajućih vrednosti na osnovu najbližih susednih vrednosti⁵⁷ [66]. U poslednje vreme sve je popularniji algoritam mašinskog učenja koji uključuje zamenu nedostajućih vrednosti, a zasniva se na gradijentnom pojačavanju i primeni stabala odlučivanja. Naziv ovog algoritma je XGBoost (od eng. *Extreme Gradient Boosting*). XGBoost se pokazao superiornim u poređenju sa drugim metoda, a posebno u okviru takmičenja koja je organizovala Kegl⁵⁸ internet platforma [69]. XGBoost realizuje stabla odlučivanja, tako što za nedostajuće podatke primenjuje tzv. *Sparsity-aware Split Finding* algoritam za pretragu podrazumevanog pravca u stablu za nedostajuću vrednost [70-71]. Očekivano, na CRAN-u postoji paket *xgboost*⁵⁹ koji je namenjen primeni ovog algoritma u programskom jeziku R [72].

Jedan od projekata iz TOBS predmeta u školskoj 2022/2023. godini je uključio ispitivanje uspešnosti klasifikacije ispitanika sa bolom u leđima primenom XGBoost algoritma [70]. Kao evaluacioni parametar odabran je F1 parametar koji služi za merenje tačnosti modela [70]. Rezultati primene XGBoost algoritma za binarnu klasifikaciju skupa podataka sa pseudoslučajno unetim nedostajućim vrednostima na 5%, 10%, 30%, 50%, 70% i 90% podataka su prikazani u Tabeli 1. Procedura je ponovljena 10 puta [70], a rezultati svih

55 *Mode* zamena ili zamena modalnom vrednošću se koristi kod kategoričkih podataka, gde se nedostajuće vrednosti zamenjuju vrednošću koja se najčešće pojavljuje u skupu podataka. Na primer, ako se radi o nedostajućim podacima o polu ispitanika koji može biti ili muški ili ženski, onda se proverava da li ima više muškaraca ili žena i nedostajuća vrednost se zamenjuje vrednošću koja je prisutnija u skupu podataka. Predstavlja najjednostavniju i jednu od najpopularnijih metoda zamene nedostajućih vrednosti. [66, 68]

56 Na dan 24.08.2022. UCI sadrži 622 skupa podataka koji se mogu koristiti sa algoritmima mašinskog učenja. Izvor: <https://archive.ics.uci.edu/ml/index.php>, pristupljeno 24.08.2022.

57 Za procenu rastojanja između susednih vrednosti često je u upotrebi Euklidsko rastojanje za numeričke promenljive, ali se mogu definisati ili koristiti druge metode procene rastojanja.

58 Kegl je platforma za takmičenja iz oblasti nauke o podacima i mašinskog učenja, kao i platforma za deljenje podataka i koda, <https://www.kaggle.com/>, pristupljeno 30.07.2023.

59 <https://cran.r-project.org/package=xgboost>, pristupljeno 31.07.2023.

dobijenih F1 parametara su usrednjeni, tako da su u Tabeli 1 prikazane srednja vrednost i standardna devijacija za svaki od procenata nedostajućih podataka. Analizirani su podaci koji su podeljeni pod slobodnom Apači 2.0 licencom⁶⁰ i koji su javno dostupni na Kegl platformi koji služe za predikciju bola u leđima i sadrže ukupno 310 opservacija tj. 310 ispitanika za koje je potrebno izvršiti predikciju [73].

Tabela 1, Uspešnost klasifikacije ispitanika sa bolom u leđima primenom XGBoost [70].

Procentat nedostajućih vrednosti	F1 parametar	
	Srednja vrednost	Standardna devijacija
5%	0.859	0.014
10%	0.862	0.017
30%	0.850	0.024
50%	0.825	0.017
70%	0.799	0.028
90%	0.757	0.041

Rezultati prikazani u Tabeli 1 [70] ubedljivo ilustruju mogućnost XGBoost algoritma da radi sa relativno velikim brojem nedostajućijih podataka na manjem inicijalnom skupu (ukupno 310 opservacija). Treba imati na umu da je ovo binarni klasifikator te da treba očekivati veću uspešnost, ali svakako razlike koje se dobijaju u F1 parametru za 5% i 90% nedostajućih podataka čine XGBoost vrlo obećavajućim algoritmom. Da ne bi bilo zabune oko toga šta predstavlja relativno veliki skup, osim slobodne procene koja je već navedena (100 000 opservacija), trenutno ne postoji jasna granica. Ovo je ujedno jedno od najčešćih pitanja koje muči sve koji primenjuju algoritme mašinskog učenja.

Za primenu statističkih testova, moguće je izračunati snagu testa koja može pomoći u planiranju studije. Sa druge strane, za primenu mašinskog učenja nad podacima, još uvek ne postoji odgovor na pitanje koliko je podataka dovoljno, a možda i neće postojati. Razlozi su: (1) nije svaki algoritam mašinskog učenja isti (neki rade sa relativno manjim kao što je XGBoost [69], a neki sa relativno većim brojem podataka) i (2) nisu svi podaci isti (tip podataka ne uključuje samo da li su parametri numerički, kategoričke promenljive, slike i sl, već kako su odabrana obeležja, da li je bilo uključeno pretprocesiranje podataka i priroda podataka). Neki od skorijih primera koji su ovde dati isključivo ilustrativno, a ne instrukcijski su:

60 <https://www.apache.org/licenses/LICENSE-2.0>, pristupljeno 30.07.2023.

- primena algoritama dubokog učenja (eng. *deep learning*) za procenu kardio-vaskularne starosti je realizovana na 34 432 opservacija [74]
- primena algoritama dubokog učenja na 1427 ispitanika i 7590 opservacija koje su predstavljene kao parovi podataka koji su korišćeni za primenu algoritama – algoritam je primenjen sa ciljem predikcije odbacivanja srčanog grafta [75] i
- primena konvolucionih neuralnih mreža na EKG signalima za određivanje pola i godina starosti ispitanika na 499 727 pacijenata [76].

Ako se za trenutak mašinsko učenje i računarski zahtevne tehnike ostave po strani, jedna od najjednostavnijih metoda zamene numeričkih nedostajućih vrednosti jeste zamena srednjom vrednošću podataka koji su dostupni. Da bi se primenila ovako jednostavna metoda, potrebno je uvesti, a najbolje proveriti odgovarajuće pretpostavke. Prva je da podaci imaju normalnu raspodelu, a druga da nedostaju po slučajnom redosledu. Primer takvih podataka mogu biti godine ispitanika u randomiziranoj studiji sa relativno velikim brojem ispitanika⁶¹. Ako podaci ne nedostaju po slučajnom redosledu (na primer, moguće je da iz nekog razloga nisu unete godine za ispitanike iz neke od starosnih grupa), onda se metoda zamene srednjom vrednošću ne može primeniti, jer zamena srednjom vrednošću može biti pristrasna. Dodatna mana zamene srednjom vrednošću u slučaju da jesu ispunjene pretodne dve pretpostavke jeste što takva zamena ne dodaje novu informaciju postojećem skupu i može u daljoj analizi doprineti potcenjivanju greške. Nedostajući podaci mogu smanjiti verodostojnost dobijenih rezultata analize signala, zbog smanjene statističke snage testova (proračuni se vrše na značajno manjem skupu podataka) ili čak do pristrasnih (eng. *biased*) procena i netačnih rezultata. [77]

Za procenu najbolje metode za zamenu nedostajućih vrednosti mogu se umetati nedostajućih vrednosti, odnosno mogu se brisati odabrane vrednosti, te na taj način se porediti primenjene metode za popunjavanje nedostajućih vrednosti. Ovakav pristup podrazumeva da su tačne vrednosti (eng. *ground truth*) koje nedostaju dostupne, te da je moguće proceniti grešku koju unose tehnike za zamenu nedostajućih vrednosti. U R-u se ovakva simulacija ili eskperiment može jednostavno sprovesti primenom:

`sample()`

61 Ovde se pod relativno velikim brojem ispitanika podrazumeva broj ispitanika koji je dovoljan da se proverii pretpostavka da podaci imaju Gausovu raspodelu.

funkcije za pseudoslučajan odabir simuliranih indeksa za nedostajuće vrednosti. Primer primene ove funkcije prikazan je na Sl. 14. Dobra praksa nalaže da se u više pokušaja simuliraju nedostajuće vrednosti kako bi se omogućila nepristrasna procena, slično kao u slučaju kros-validacije za primenu metoda mašinskog učenja. Ovakav pristup je primenjen i u projektu čiji su rezultati prikazani u Tabeli 1, samo što je 10 ponavljanja relativno mali broj ponavljanja, obzirom da se ova procedura može automatizovati. [66, 70]

```
> d <- 1:20
> sample(d, 5)
[1] 6 15 8 17 14
> sample(d[d < 5], 5, replace = TRUE)
[1] 2 1 3 4 3
> |
```

Slika 14, Primer primene *sample()* funkcije u R-u za pseudoslučajan odabir 5 celih brojeva iz skupa/niza *d*. Primititi da se sa *sample()* funkcijom može kombinovati logičko indeksiranje, kao i da postoji dodatni argument funkcije – *replace*. Takođe, kako se radi o pseudoslučajnom odabiranju, za različita pokretanja dobiće se različiti rezultati.

Brisanje nedostajućih vrednosti jeste najsigurniji način da se ne utiče na analizu i da ona ne bude pristrasna. Međutim, potrebno je prvo proceniti koliko vrednosti nedostaje i koje vrednosti nedostaju. Ako je procenjeno da nema previše nedostajućih vrednosti, tada se one mogu isključiti u celini (eng. *listwise deletion*) ili po parovima (eng. *pairwise deletion*). Ako ipak nedostaje veći broj podataka, tada se pored zamene srednjom vrednošću (eng. *mean substitution*) ili medijanom⁶² (eng. *median substitution*), može koristiti zamena pomoću regresije (eng. *regression imputation*). U zameni pomoću regresije koriste se postojeće vrednosti tako da se napravi predikcija na osnovu koje se popunjavaju nedostajuće vrednosti. Postoje neslaganja u literaturi za primenu regresije, te i njene kritike. Naime, regresija predstavlja zamenu nedostajućih vrednosti procenjenim vrednostima na osnovu ostalih promenljivih u skupu. Ovaj način ima prednost kada se poredi sa metodama za delimično ili potpuno brisanje opservacija, jer zadržava oblik raspodele podataka i ne menja značajno varijansu. Mana ove metode je, kao i kod zamene srednjom vrednošću, što se nikakva nova informacija ne dodaje skupu i što je ugrožena procena greške analize. Kada se poredi zamena regresijom i srednjom vrednošću, potrebno je

62 Zamena medijanom se radi kada nije ispunjen uslov o Gausovoj raspodeli promenljive čije opservacije nedostaju. Takođe, medijana je robusna u odnosu na ekstremne vrednosti u odnosu na srednju vrednost, pa može da bude dobra zamena srednjoj vrednosti ako u skupu postoje relativno velike razlike između opservacija.

spomenuti i to da je regresija računarski zahtevnija, te u pojedinim situacijama treba uzeti i takva ograničenja u obzir. [68, 77]

Ovaj udžbenik nema za cilj da predstavi pojedinačno sve metode za zamenu nedostajućih podataka, već da studentima približi problem rada sa merenim podacima, kao i da pruži smernice za dalji rad sa podacima koji poseduju nedostajuće vrednosti. U ovom polju postoji pored preporuka i veliki broj kritika, te podacima koji su na raspolaganju treba pristupiti obazrivo. Na primer, u anesteziologiji je široko rasprostranjena metoda za zamenu nedostajućih podataka u vremenskim serijama primenom LOCF (eng. *Last Observation Carried Forward*) tehnike koja uključuje zamenu tekuće promenljive poslednjom merenom vrednošću. Iako LOCF ima smisla, posebno ako se pretpostavi stacionarna priroda signala na relativno kratkim segmentima, zvanična preporuka je da se ova metoda ne koristi, osim ako ne postoje odgovarajući naučni dokazi koji bi potkrepili njenu primenu [77]. Jedan od preporučenih načina za zamenu nedostajućih podataka je višestruka zamena nedostajućih vrednosti (eng. *multiple imputation*) gde se istovremeno menjaju sve nedostajuće vrednosti koje su dobijene na osnovu procene varijabilnosti i/ili nesigurnosti ostalih podataka [77]. Pokazano je da se u naučnim časopisima u psihologiji pored brisanja nedostajućih vrednosti često koristi i višestruka zamena kao primer dobre prakse u radu sa podacima [68]. LOCF se koristi i za zamenu nedostajućih vrednosti u vremenskim serijama. Samo tu se mora voditi računa o broju nedostajućih vrednosti i mogućim izmenama. Ako je signal meren sa frekvencijom odabiranja 1000 Hz⁶³ tokom 60 s i ako nedostaju poslednjih 50 odbiraka, onda tih 50 odbiraka čine 0.05 s (0.083% od ukupnog trajanja signala). U ovom slučaju verovatno LOCF zamena neće uneti promenu na signalu, a postavlja se i pitanje zašto uopšte zadržati <0.1% podataka. Dodatno, u radu sa realnim signalima, posebno sa biosignalima, šum nije retka pojava, te može doći do situacije gde je poslednja izmerena vrednost značajno veća ili manja od ostalih vrednosti na signalu. Ovaj problem se može prevazići zamenom nedostajućih vrednosti medijanom u nekom od intervala. Čak i da ne postoji šum ili da zamena medijanom reši problem šuma kod nestacionarnih biomedicinskih signala (kao, na primer, elektroencefalogram, EEG) može se narušiti nestacionarnost, jer je signal takav da nikada nema konstantnu vrednost. Iako je ovo narušavanje stacionarnosti relativno malo, zavisi od primenjene metode analize EEG signala da li će se LOCF zamena odraziti na informaciju koja je sadržana u signalu. Iz svih navedenih razlika, uobičajeno se

63 Iako je Hz ustaljena jedinica za frekvenciju odabiranja, nije u potpunosti tačna. Tačnije bi bilo navesti da je biosignal meren sa 1000 odbiraka po sekundi. Skraćenice koje se u praksi sreću kao zamena za Hz za frekvenciju odabiranja su S/s, SPS, SS od eng. *samples per second* [78].

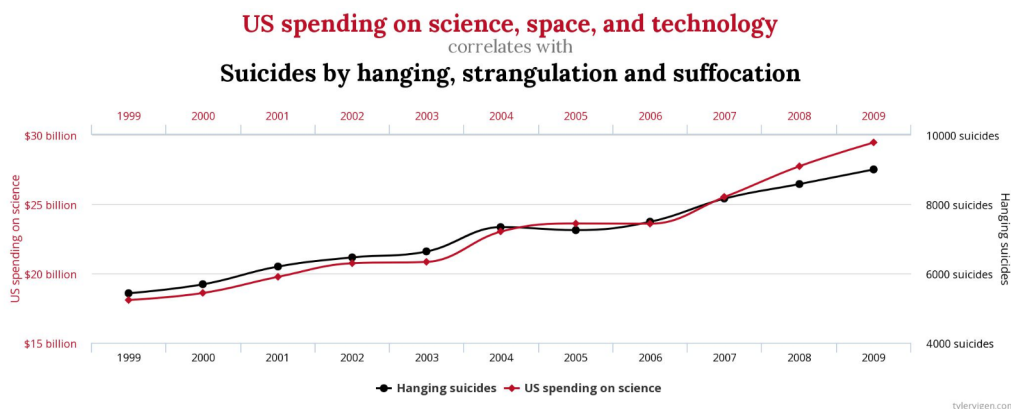
koristi brisanje odbiraka na delu signala koji nedostaje. Široko je rasprostranjena metoda brisanja odbiraka u analizi elektrogastrograma (EGG) koji predstavlja električnu aktivnost glatkih mišića želuca. Kako je EGG signal relativno malih amplituda (izvor signala je prilično udaljen od površinskih srebro/srebro-hlorid Ag/AgCl elektroca) i kako je relativno malih frekvencija (uobičajen rad želuca je na 2-4 ciklusa u minutu, odnosno cpm od eng. *cycles per minute*), to je signal vrlo podložan šumu. Kako ne postoje odgovarajuće metode za eliminaciju takvog šuma, u upotrebi je brisanje odbiraka pod uslovom da je trajanje šuma relativno kratko u odnosu na trajanje signala [79]. Manipulacija kratkim vremenskim intervalima na EGG signalu nije isto što i rad sa nedostajućim vrednostima u pravom smislu, jer instrument jeste izmerio neke vrednosti, ali se zbog prisustva šuma i ne poznavanja vrednosti odbiraka u trenutku šuma, ti intervali tretiraju kao nedostajuće vrednosti.

Pored šuma, u vremenskim serijama može doći do drugih razloga za nedostajanje odbiraka. Pokazano je da korišćenje odevnih tehnologija (eng. *wearables*) usled bežične komunikacije mogu da dovedu do gubitka odbiraka – u nekim slučajevima do 70% [80]. Ovakve podatke nije jednostavno analizirati, a pre analize treba doneti procenu o njihovoj upotrebljivosti tj. procenu da li je analiza uopšte potrebna ili merenja treba ponoviti. Pored osnovnih metoda koje su prikazane u ovom radu za rad sa nedostajućim vrednostima u tabelarnim podacima, treba spomenuti i metode za oporavak vremenskih serija signala (eng. *signal recovery*) ili rekonstrukciju signala (eng. *signal reconstruction*) koje se zasnivaju na gradijentnim iterativnim procedurama [81], a koje mogu biti veoma korisne kada dođe do gubitaka odbiraka signala.

Zamena slučajnim vrednostima (eng. *hot deck imputation*) je takođe prisutna u praksi. Ovde reč slučajno treba uslovno shvatiti, jer se radi o nasumičnom odabiru promenljive iz određenog skupa vrednosti. Na primer, ako nedostaje visina ispitanika, potpuno slučajnim odabirom vrednosti može se desiti da visina osobe bude jednaka 10 mm ili 10 m, što bi moglo imati katastrofalni uticaj na rezultat analize ako se, na primer, ispituje zavisnost visine ispitanika od kapaciteta pluća. Ovaj drastični primer ukazuje da za odabir odgovarajuće metode zamene nedostajućih vrednosti, u opštem slučaju, ne samo što je potrebno poznavanje promenljivih i njenih fizičkih ograničenja, već i cilj analize signala. Za kategoričke promenljive, zamena slučajnim vrednostima je daleko jednostavnija i široko je zastupljena u anketnim istraživanjima [68], u odnosu na numeričke podatke.

Posledica pogrešne slike o podacima se može stvoriti ako se, na primer, nakon zamene srednjom vrednošću ili zamene regresijom, računa korelaciona analiza.

Rezultat takve analize može dati lažno visoke vrednosti korelacija [68]. Čak i uz sve mere predostrožnosti, statistički testovi ne mogu da pomognu, ako dođe do lažnih rezultata.



Slika 15, Primer lažne korelacije. Slika je podeljena pod CC licencom, a preuzeta je sa <http://www.tylervigen.com/spurious-correlations> (pristupljeno 31.07.2023). Savetuje se čitaocima da posete blog na kome se nalazi grafik (zbog rezolucije koja je ovde smanjena usled formata udžbenika i drugih vizuelizacija). Lažna korelacija između prikazane dve promenljive je 99.79%. Preporučuje se i knjiga u kojoj su prikazani i drugi primeri lažnih kros-korelacija [82].

Lažne kros-korelacije koje nastaju primenom metoda za zamenu nedostajućih vrednosti se mogu sprečiti. Međutim, postoje lažne korelacije koje se ne mogu sprečiti, jer su prisutne u podacima. Na primer, jedna od lažnih kros-korelacija jeste primer linearne zavisnosti sa kros-korelacionim koeficijentom od 99.79% između finansijskih ulaganja u Sjedinjenim Američkim Državama (SAD) za nauku, kosmos i tehnologiju u milijardama i broja samoubistava pomoću vešanja, davljenja i gušenja (Sl. 15) koja je prikazana u blogu Tajlera Vigena⁶⁴ i u knjizi [82]. Iznenađujuće visok kros-korelacioni koeficijent (99.79%) je dobijen za podatke koji nisu manipulisani ni na koji način. Ovakve situacije su retke, ali se dešavaju. Tada je važno poznavati podatke uzročno-posledične veze da bi se zaključilo da li je dobijen kros-korelacioni koeficijent tačan ili lažan. Na primeru ulaganja u nauku i vrste samoubistava logično je da nema povezanosti. Međutim, problem se javlja ako se tu nađu veličine koje mogu da budu povezane, a ne postoji dodatni način da se proverí njihova povezanost. Zato je važno primeniti rigorozne metode analize, uvesti kad god je moguće dodatne provere i biti obazriv prilikom donošenja zaključaka na osnovu podataka. Nezavisna verifikacija naučnih rezultata od strane drugih grupa koja

64 <https://www.tylervigen.com/about>, pristupljeno 31.07.2023.

bi imala za cilj proveru ponovljivosti rezultata, ali i računarske reproducibilnosti je od suštinskog značaja za savremenu nauku [23].

Preporuka je da se prilikom analize podataka utvrdi razlog nedostajućih podataka. Ako to nije moguće, onda dobra praksa nalaže da se na odgovarajući način opišu i procene nedostajući podaci [68]. U slučaju da je moguće proceniti uzrok nedostajućih podataka, sledeći korak je da se donese odluka da li će se podaci brisati (isključiti iz analize), popunjavati srednjom vrednošću sličnih promenljivih, da li će preuzeti vrednost medijane umesto srednje vrednosti i sl. Postoji više razloga zašto podaci mogu da nedostaju, u literaturi su ti razlozi kategorizovani kao [66, 77, 83]:

- MCAR (eng. *Missing Completely at Random*) je situacija u kojoj su potpuno slučajno raspodeljeni nedostajući podaci. Ovo je, donekle idealna situacija i za simuliranje i za primenu neke od metoda za procenu nedostajućih vrednosti. Međutim, u praksi se retko dešava. Primer ove situacije jeste kada dolazi do prekida u komunikaciji ili tehničkih problema sa prenosom i čuvanjem podataka.
- MAR (eng. *Missing at Random*) je situacija u kojoj iako su slučajno raspodeljeni nedostajući podaci, razlog zašto podaci nedostaju zavisi od neke promenljive u skupu podataka. Zbog zavisnosti sa drugim promenljivima, MAR nedostajanje može imati za posledicu pogrešan zaključak iz analize podataka.
- MNAR (eng. *Missing Not at Random*) je situacija u kojoj podaci nedostaju po slučajnom rasporedu, a u njihovo nedostajanje je u direktnoj vezi sa analizom i rezultatom istraživanja. Razlozi mogu biti potpuno opravdani. Na primer, ako su pacijenti u međuvremenu preminuli oni nisu mogli biti dostupni za ponovljeno merenje (eng. *follow up*), a to može biti važan parametar za kliničku studiju. Takođe, MNAR predstavlja sve slučajeve koji se ne mogu predstaviti sa MCAR i MAR kategorijama, te opisuje najkompleksniju situaciju sa kojom se inženjeri mogu sresti u analizi.

Najčešće se pomoću dostupnih informacija o merenju/prikupljanju podataka zaključuje zbog čega podaci nedostaju (usled MCAR, MAR ili MNAR). U slučaju da nema dovoljno informacija na osnovu kojih se može zaključiti zašto podaci nedostaju, a pri tom je važno da se zna da li nedostajući podaci imaju slučajnu raspodelu, moguće je koristiti Litlov MCAR test [68]. Jedan od rezultata testa je p vrednost, odnosno verovatnoća (od eng. *Probability*), kojom se utvrđuje statistička značajnost slučajne raspodele. Za slučaj da test nije značajan tj. da verovatnoća p nije manja od α , gde se za α često uzimaju

vrednosti od 0.05 (5%), 0.01 (1%) ili 0.001 (0.1%), onda se može zaključiti da nedostajući podaci ne odstupaju od slučajne raspodele.

```
> x <- airquality[1:6,]
> x
  Ozone solar.R wind Temp Month Day
1    41    190  7.4   67     5   1
2    36    118  8.0   72     5   2
3    12    149 12.6   74     5   3
4    18    313 11.5   62     5   4
5    NA     NA 14.3   56     5   5
6    28     NA 14.9   66     5   6
> is.na(x$solar.R)
[1] FALSE FALSE FALSE FALSE  TRUE  TRUE
> sum(is.na(x$solar.R))
[1] 2
> x[complete.cases(x),]
  Ozone solar.R wind Temp Month Day
1    41    190  7.4   67     5   1
2    36    118  8.0   72     5   2
3    12    149 12.6   74     5   3
4    18    313 11.5   62     5   4
> |
```

Slika 16, Primer korišćenja funkcija za rad sa NA vrednostima. Korišćeni su `airquality` ugrađeni podaci u R-u. Primititi kako se `sum()` funkcija može koristiti u kombinaciji sa `is.na()` funkcijom za procenu da u koloni `Solar.R` ima dve nedostajućih vrednosti.

Pristup u kome se proverava raspodela nedostajućih vrednosti i uzrok nedostajućih vrednosti se ne primenjuje uvek. Za studente elektrotehnike, ovde je verovatno najbolje uporediti ovu situaciju sa računanjem merne nesigurnosti u električnim merenjima. Koliko detaljno će se pristupiti računu merne nesigurnosti zavisi od važnosti merenja – jedno je ako se proverava otpornost otpornika za realizaciju električnog kola, a potpuno drugo kada se meri etalonska vrednost⁶⁵ [78]. Slično važi i za nedostajuće vrednosti – u zavisnosti od podataka i cilja analize biraju se pogodne metode, a nekada više njih. Bez obzira na odabrane metode analize i zamene, uvek se vrši procena nedostajućih

65 Etaloni su izbačeni iz Međunarodne definicije osnovnih jedinica, ali ovde je samo ilustrativno korišćen primer merenja vrednosti etalona sa ciljem ilustracije merenja od relativno velike važnosti.

vrednosti, pa se tek onda pristupa ili izbacivanju NA vrednosti iz skupa podataka ili njihovoj zameni.

U R-u je moguće koristiti *is.na()* funkciju koja dalje logički rezultat koje su vrednosti NA, ali i:

complete.cases()

koja omogućava da se nađu svi objekti tj. subjekti za koje ne postoji nijedan nedostajući podatak. Funkcija *complete.cases()* se ne koristi za rad isključivo sa podacima tipa *data frame*, već i sa matricama i listama. Primer primene ove dve funkcije dat je na Sl. 16.

Brisanje podataka se radi tako što se ne briše samo jedan podatak, već, na primer, ako jedan parametar nedostaje za jednog pacijenta (jedna vrsta u *data frame*-u), tada se izostavlja tj. briše cela vrsta/opservacija (nakon brisanja podataka, rezultujući podaci imaju dve vrste manje na Sl. 16). Analiza se dalje sprovodi na skupu podataka koji ima kompletne opservacije (eng. *complete cases*). Ovo je posebno zgodno ako ima dovoljno opservacija i ako je MCAR razlog nedostajućih podataka. Postoji i drugačiji tip delimičnog isključenja nedostajućih podataka iz analize (eng. *pairwise deletion*) koji se najčešće primenjuje kada se ispituje jedan ili više parametara iz podskupa podataka i kada se želi ispitati međusobna zavisnost parametara ili trend, a svojstvena je MCAR ili MAR podacima. Iako predstavlja blažu varijantu od brisanja potpunih opservacija, ovaj pristup analizi je i dalje osetljiv na relativno veliki broj nedostajućih podataka. Na kraju, u slučaju da prevelik broj podataka nedostaje u nekoj koloni, jedno od rešenja bi bilo da se kompletna kolona izostavi iz dalje analize. [68, 77]

U kliničkim studijama najčešći razlozi zašto podaci nedostaju su:

- usled nedostupnosti pacijenata ili zdravih ispitanika da učestvuju u istraživanju prilikom ponovljenog merenja ili prilikom naknadog prikupljanja podataka (razlog može da bude i odustajanje zdravih ispitanika ili pacijenata na koje imaju pravo po Helsinškoj deklaraciji [84]),
- usled nedostatka opreme ili kvara na opremi tokom merenja ili prilikom ponovljenih merenja⁶⁶,

⁶⁶ U biomedicinskom inženjerstvu se ređe koristi termin eksperiment u radu sa ljudima. Uobičajeno se na ljudima meri, ali se eksperimentiše na životinjama. Terminologija nije ista u svim oblastima, te je u psihologiji termin "eksperiment" ustaljen za ispitivanja na ljudima.

- usled nedostupnosti istraživača (inženjera ili lekara),
- usled gubitka podataka (bilo da se radi o računarskoj korupciji datoteka, krađi podataka ili nemarnosti),
- nedostatak finansija može predstavljati veliki problem prilikom realizacije naučnih studija, te se može desiti da prekid finansiranja dovede do prestanka prikupljanja podataka i
- drugi faktori.

U svakom istraživanju je potrebno preduzeti odgovarajuće mere tako da se ne dođe do situacija koje su prethodno opisane, što je i jedan od uslova za odborenje projektnih prijava (upravljanje rizicima), ali situacija nekada može biti u potpunosti nepredvidiva. Na primer, nedavna pandemija koja je u mnogim delovima sveta izazvala zaključavanje (eng. *lock down*) i prelazak na rad od kuće je u velikoj meri kao vanredna okolnost (fr. *force majeure*) uticala na prikupljanje podataka. Iako o tome još uvek nema puno izvora, u 2022. godini održan je *Euroasian Open Science Salon*⁶⁷ u Kini na kome su izneti problemi sa prikupljanjem podataka sa kojima su se susreli naučnici tokom pandemije, a bilo je i drugih primera. Najbolji način da se upravlja nedostajućim podacima jeste da se spreči njihovo pojavljivanje dobrim dizajnom i rukovođenjem tokom kliničke studije, kao i blagovremenim izmenama u planu sprovođenja istraživanja. Na primer, to uključuje kontinualnu procenu nedostajućih vrednosti tokom istraživanja, te po potrebi i povećanje broja uključenih zdravih subjekata i pacijenata u studiji. [77]

U zaključku ovog potpoglavlja: postoji relativno veliki broj opcija kako se mogu nedostajuće vrednosti zameniti. Najjednostavnije metode uključuju zamenu vrednosti koje su označene sa NA ili imaju prazna polja u tabelarnim podacima srednjom vrednošću za numeričke podatke, zamenu nedostajućih vrednosti onom koja se najčešće pojavljuje za kategoričke podatke ili korišćenjem prediktivnih modela da bi se odredile sa određenim stepenom tačnosti najbolje aproksimativne vrednosti koje se mogu koristiti za zamenu nedostajućih vrednosti. Vremenske serije podležu nešto drugačijim metodama zamene nedostajućih vrednosti, ali su principi tretiranja podataka isti.

Ove razlike su terminološke, jer svako merenje/eksperimenti na ljudima i životinjama podležu odgovarajućoj međunarodnoj regulativi od kojih je najvažnija Helsinška deklaracija [84], ali i nacionalnim i institucionalnim regulativama. Na Univerzitetu u Beogradu u Kodeksu profesionalne etike Univerziteta u Beogradu godina LIV, broj 193, 10. jul 2016. godine u članu 19. su definisane smernice i obaveze istraživača na Univerzitetu u Beogradu prilikom korišćenja ljudi i životinja u naučnom istraživanju (<https://bg.ac.rs/files/sr/univerzitet/univ-propisi/Kodeks-profesionalne-etike.pdf>, pristupljeno 31.07.2023).

67 <https://www.bjexchange.cn/index/Ennews/index?id=23>, pristupljeno 31.07.2023.

Generalno, prednosti zamene nedostajućih vrednostima su što nijedna opservacija u tom slučaju nije izostavljena iz inicijalnog skupa podataka, te se skup podataka ne mora smanjivati, a mana je što kada nedostaje relativno veliki broj vrednosti (bilo da se radi o opservacijama ili o vremenskim odbircima biosignala) može biti zametno odrediti kojim vrednostima bi se one menjale, a da ta zamena ne utiče na rezultat analize. [67]

Bilo da se opservacije brišu zato što vrednosti nedostaju ili zato što su zašumljene i izuzete iz analize (eng. *outlier*), dobra praksa nalaže da se metoda analize ponovi sa i bez obrisanih vrednosti, kad god je to izvodljivo. Najbolji način za isključenje podataka je da se iskoriste praktična saznanja (primer visine ispitanika od 10 mm i 10 m je već predstavljen ranije u ovom potpoglavlju). Kada ne postoje prethodna saznanja o mogućim opsezima parametra koji se analizira uobičajena praksa nalaže da se vrednosti veće od tri interkvartilna opsega od trećeg kvartila izuzimaju iz analize [85-87]. Nekada se ovakva numerička ograničenja u opsezima parametara kombinuju sa postojećim praktičnim saznanjima.

```
# excluding outliers in time to fixate (ttf)
# note: please, skip this part if you want get results with outliers
s <- summary(dat$ttf)
s
iq <- IQR(dat$ttf, na.rm = T)
dat$ttf[dat$ttf > (s[5]+3*iq)] <- NA
dat$ttf[dat$ttf < (s[2]-3*iq)] <- NA

dat$ttf[dat$ttf > 500] <- NA # additional rule
```

Slika 17, R kod za isključenje *outlier*-a iz analize TTF parametra. Kod je prikazan u izvornom obliku i sa komentarima na engleskom jeziku [85-86, 88]. Primititi da su u ovom kodu korišćene *IQR()* i *summary()* funkcije. Obe funkcije se koriste za računanje interkvartilnih intervala.

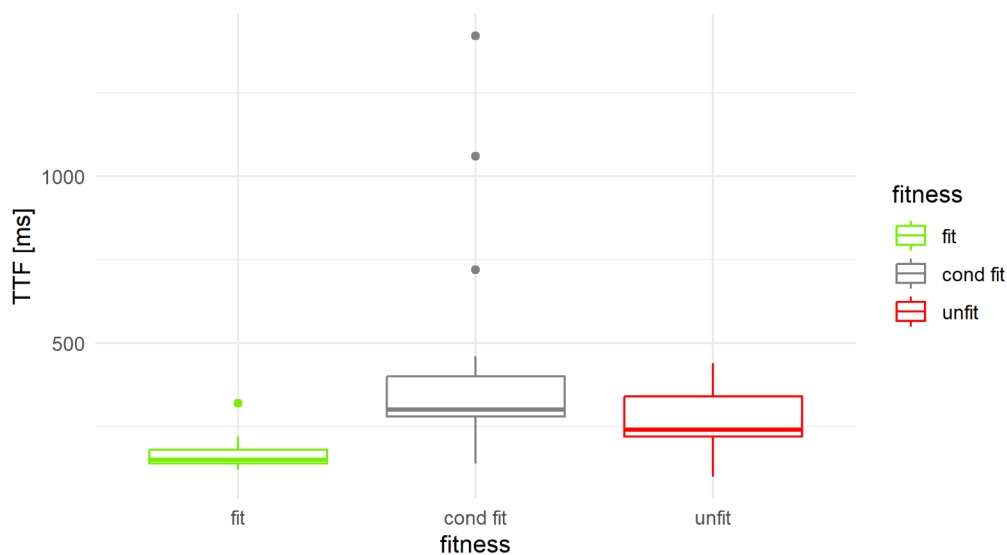
Na primer, u studiji iz 2023. godine [85], analizirani su parametri dobijeni iz video signala merenih pomoću uređaja za praćenje pokreta oka (eng. *eye tracker*). Studija je sprovedena na neurološkim pacijentima u Univerzitetском rehabilitacionom institutu Soča⁶⁸, a cilj je bio da se analiziraju pokreti oka neuroloških pacijenata tokom upravljanja simuliranim vozilom⁶⁹. U studiji je predstavljen TTF (eng. *Time To Fixate*) parametar, odnosno vreme koje je proteklo od pojave događaja u simulaciji do trenutka kada je pacijent usmerio pogled na mesto događaja. R programski kod za statističke proračune u ovoj

68 <https://www.uri-soca.si/>, pristupljeno 31.07.2023.

69 Korišćen je Nervtech simulator vožnje (Nervtech d.o.o, Ljubljana, Slovenija), <https://www.nervtech.com/>, pristupljeno 31.07.2023.

studiji je javno dostupan na Zenodo⁷⁰ repozitorijumu [85-86, 88], a isečak iz koda je prikazan na Sl. 17.

Grafici na kojima se vizuelno porede parametri pre i posle proglašenja određenih TTF vrednosti nedostajućim su dati na Sl. 18-19. Pored kriterijuma koji se zasniva na kvartilima za određivanje *outlier*-a (TTF parametar je prvi put predstavljen za neurološke pacijente u ovom radu, te nisu bili dostupni opsezi TTF parametra koje je bilo moguće koristiti), korišćen je i dodatni kriterijum [85-86] koji ograničava TTF tako da ne može biti veće od 500 ms - kriterijum koji je baziran na saznanjima o tome koliko treba da bude vreme reakcije (eng. *reaction time*) za koje je poznato da je veće od TTF parametra (Sl. 17).

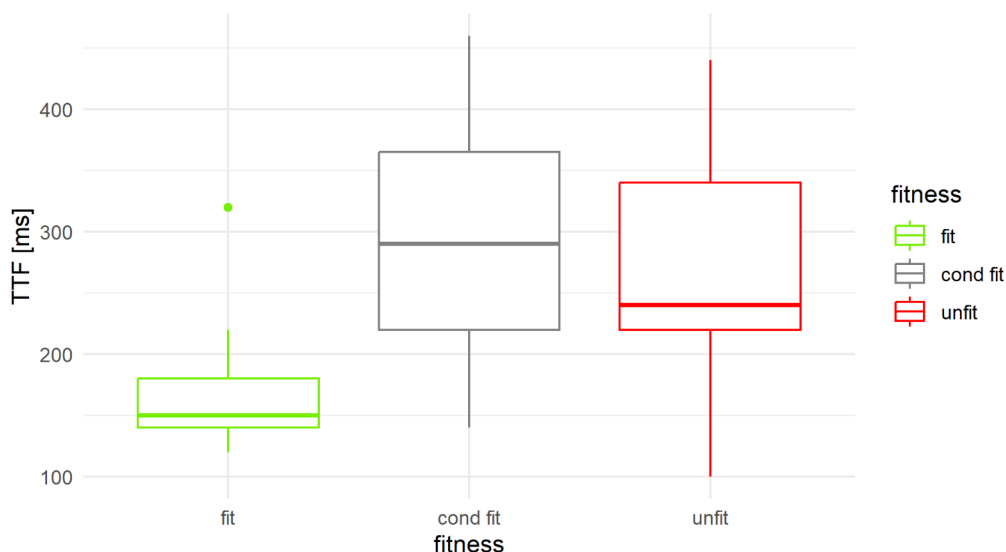


Slika 18, TTF (eng. *Time To Fixate*) parametri za tri grupe pacijenata (fit, cond fit i unfit – sposobni, uslovno sposobni i nisu sposobni za vožnju) [85-86, 88]. Grafici su prikazani za sve opservacije. Kako je R programski kod za iscrtavanje ovih grafika podeljen na engleskom jeziku uz publikaciju [85], to oznake na engleskom jeziku nisu menjane za potrebe grafičkog prikaza u udžbeniku. Slika je podeljena pod CC BY 4.0 DEED licencom (<https://creativecommons.org/licenses/by/4.0/>), pristupljeno 28.05.2024).

Iako se grafici na Sl. 18 i Sl. 19 razlikuju, međusobni odnosi tri grupe TTF parametara su isti (najveće vrednosti za TTF parametar su u grupi cond fit pacijenata – uslovno sposobni za vožnju, a najmanje u fit grupi – pacijenti za koje je multidisciplinarna komisija zaključila da su sposobni za vožnju). Ovi odnosi između tri grupe TTF parametara su potvrđeni i statističkom analizom

70 <https://zenodo.org/>, pristupljeno 31.07.2023.

koja je prikazana u [85]. Za prikaz grafika na Sl. 18 i Sl. 19, iskorišćen je javno podeljen R kod [85-86, 88], a svi zainteresovani mogu preuzeti kod i reprodukovati iste grafike. U narednim poglavljima biće detaljno objašnjeno kako se koriste osnovne funkcije u R-u za prikaz podataka, koji su paketi preporučeni za vizuelizaciju signala, kao i kako odabrati odgovarajući tip grafika, boje i kako pravilno označiti osa grafika.



Slika 19, TTF (eng. *Time To Fixate*) parametri za tri grupe pacijenata (fit, cond fit i unfit – sposobni, uslovno sposobni i nisu sposobni za vožnju) [85-86, 88]. Grafici su prikazani za opservacije na kojima su uklonjeni *outlier*-i primenom dve metode (poređenje sa interkvartilima i pod pretpostavkom da TTF ne može biti veće od 500 ms). Kako je R programski kod za iscrtavanje ovih grafika podeljen na engleskom jeziku uz publikaciju [85], to oznake na engleskom jeziku nisu menjane za potrebe grafičkog prikaza u udžbeniku. Slika je podeljena pod CC BY 4.0 DEED licencom (<https://creativecommons.org/licenses/by/4.0/>, pristupljeno 28.05.2024).

1.1.2 Uvoz podataka u R

Sada kada su savladane osnovne operacije za manipulaciju podacima, a posebno nedostajućim podacima, potrebno je savladati tehnike za uvoz podataka u R. Najčešće korišćene funkcije za uvoz podataka su:

`read.table()` i `read.csv()`

jer se podaci najčešće nalaze u datotekama sa `.csv` (eng. *comma separated value*) ili `.txt` (eng. *plain text*) ekstenzijama. Formatu podataka koji se koriste su od neprocenjive važnosti za trajno skladištenje podatka, kao i za računarsku

reproducibilnost. Vlasnički (eng. *proprietary*) zatvoreni formati podataka, koliko god da su široko rasprostranjeni, ne treba da budu izbor formata za trajno čuvanje podataka. Najvažniji razlog je što vlasnički formati predstavljaju proizvod komercijalnih kompanija, te u najvećem broju slučajeva nisu otvoreni⁷¹, a kompanija zadržava pravo da otvori ili zatvori format koji je razvila. Ako kompanija koja format održava donese odluku da se format više neće koristiti, u tom slučaju neće postojati način da se do podataka dođe, odnosno da se pristupi datotekama u kojima se nalaze podaci. Takvi podaci će biti trajno zaključani, odnosno izgubljeni. Primeri široko rasprostranjenih, vlasničkih, zatvorenih formata su .docx i .xlsx [90]. Sva dokumenta kao što su, na primer, projektni izveštaj, podaci dobijeni iz kliničke studije, diplomski rad i dokumenta državnih institucija (obrasci za diplome, odluke, izvorni tekstovi regulativa/Zakona/akata i dr.) nisu i ne mogu biti trajno arhivirani u zatvorenim vlasničkim formatima, jer postoji realna opasnost od njihovog trajnog zaključavanja. Veći broj međunarodnih i nacionalnih organizacija je izdao smernice za izbor formata za trajno arhiviranje podataka, a među njima su DANS⁷² i OpenAIRE⁷³ [89-90]. Ove smernice su posebno važne za istraživačke podatke u nauci, ali bi ih trebalo primeniti i u privatnom i državnom sektoru. Uputstvo na srpskom jeziku za formate podataka je dostupno na Serbia.RMD⁷⁴ internet stranici projekta Univerziteta u Beogradu.

U upotrebi su dva termina koji su usko povezani sa trajnim zaključavanjem podataka, a predstavljaju osnovu pismenosti analitičara i biomedicinskih inženjera koji se bave obradom podataka. Prvi termin je digitalna zastarelost (eng. *digital obsolescence*) koja se ne odnosi samo na otvorenost formata, već i na nedostupnost hardvera i softvera za čitanje digitalnog medijuma [91]. To bi mogli da budu podaci sačuvani na flopi disketi⁷⁵ koja već dve decenije nije više u upotrebi. Ovo je jedan od najvećih problema savremenog društva, a vrlo često se označava kao mračno digitalno doba u literaturi (eng. *digital dark age*), jer trenutno dostupni medijumi za čuvanje digitalnih podataka nisu pouzdani [91], te je pitanje da li će podaci kojima danas raspolažemo biti dostupni budućim generacijama kroz, na primer, 2000 godina na planeti Zemlji. Drugi termin je digitalni suverenitet: "Digitalni suverenitet je sposobnost države ili federacije država da obezbede digitalne tehnologije koje smatraju kritičnim za svoje

71 Postoji određen broj vlasničkih formata koji su otvoreni, a detalji se mogu pogledati u [89].

72 <https://dans.knaw.nl/en/>, pristupljeno 01.08.2023.

73 <https://www.openaire.eu/>, pristupljeno 01.08.2023.

74 <https://rdm.open.ac.rs/index.php/organize-data/formats>, pristupljeno 01.08.2023. (planirana je migracija materijala sa Serbia.RDM sajta na sajt Zajednice otvorene nauke Srbije, <https://oscs.open.ac.rs/>, pristupljeno 01.08.2023).

75 https://en.wikipedia.org/wiki/Floppy_disk, pristupljeno 01.08.2023.

blagostanje, konkurentnost i sposobnost delovanja, i da budu u stanju da ih razvijaju ili da ih nabave iz drugih ekonomskih oblasti bez jednostrane strukturne zavisnosti.” [92-93]. Jednostavno rečeno, da bi neka ustanova ili država posedovale digitalni suverenitet potrebno je da koriste otvorene formate, jer će u protivnom pristup njihovoj arhivi (podacima) da zavisi od onih koji su kreirali formate. Jedan od najdrastičnijih primera posledica upotrebe zatvorenih formata i uticaja vlasničkih formata na nauku i društvo je činjenica da je usvojen međunarodni dogovor da se promene imena 27 ljudskih gena, jer je Majkrosoft Eksel⁷⁶ (Microsoft, Redmond, Washington, USA) editor za tabelarne podatke uporno konvertovao nazive ovih gena u datume [94].

Bez obzira što u svom imenu funkcije *read.table()* i *read.csv()* sadrže nazive ekstenzija datoteka koje se učitavaju, obe mogu služiti za čitanje datoteka sa .txt i sa .csv ekstenzijama. Ono što dominantno razlikuje ove dve funkcije jesu podrazumevane vrednosti ulaznih parametara. Praktično, nije jednostavno, a ni potrebno da se sve te vrednosti pamte, već uvek treba proveriti rezultat koji se dobija, pa se po potrebi vratiti na komandu za učitavanje podataka i promeniti ili funkciju ili ulazne parametre funkcije. Kao rezultat ove dve funkcije dobijaju se podaci tipa *data frame*.

Ulazni parametri *read.table()* funkcije za čitanje podataka (koje ne treba pamtiti napamet, već samo imati na umu šta se sve može podesiti) su: *file* (ime datoteke umesto koje se može uneti i veb adresa na kojoj se nalazi datoteka), *header* (logička promenljiva koja označava postojanje zaglavlja sa osnovnim metapodacima), *sep* od eng. *separator* (služi da se naznači koji skup karaktera se u datoteci koristi za razdvajanje kolona), *colClasses* (služi da se definiše tip podataka koji se učitavaju), *nrows* (maksimalan broj vrsta koji se učitava za slučaj da se datoteka delimično učitava), *comment.char* (unos karaktera koji u datoteci označava početak komentara), *skip* (broj linija na početku datoteke koje treba preskočiti prilikom učitavanja podataka), *stringAsFactors* (da li bi promenljive koje se učitavaju trebalo sačuvati u *data frame*-u kao kategoričke promenljive, što se može uraditi i naknadno kada su podaci već učitani) i drugi ulazni parametri. Savet je da se provere podrazumevane vrednosti ulaznih parametara i tip ulaznih parametara. Za obrnut proces tj. čuvanje podataka u datoteku koristi se funkcija *write.table()*. Kako bi se pregledali svi ulazni parametri i pomoć u radu sa bilo kojim funkcijama u R-u se može jednostavno dobiti korišćenjem operatora “?” kako je prikazano na Sl. 20. Stranica sa uputstvom za korišćenje funkcije se otvara u posebnom Help prozoru u okviru R Studio razvojnog okruženja.

76 <https://www.microsoft.com/en-us/microsoft-365/excel>, pristupljeno 01.08.2023.

Ne samo za funkcije koje se koriste prilikom učitavanja podataka, već i generalno za sve funkcije u R-u, moguće je promeniti podrazumevanu vrednost nekog od ulaznih parametara komandom *options()*, a primer korišćenja ove funkcije je:

```
options(stringAsFactors = FALSE)
```

```
> ?read.table  
> ?read.csv  
> ?write.table  
> |
```

Slika 20, Primer korišćenja operatora “?” u konzoli R Studio okruženja za pregled najčešće korišćenih funkcija za učitavanje podataka iz datoteka i čuvanje podataka u datotekama u programskom jeziku R.

U većini slučajeva nije preporučljivo korišćenje komande *options()*, jer u slučaju da se programski kod deli ili ponovo pokreće na drugom računaru, može se desiti da se ovo podešavanje ne podeli, jer je, na primer, realizovano u konzoli, a nije sačuvano u programskom kodu, te se može ugroziti računarska ponovljivost programskog koda.

Ako su podaci dostupni u vlasničkim formatima kojima je potrebno pristupiti u R programskom jeziku, moguće im je pristupiti primenom specijalizovanih paketa. Na primer, studija koju je objavila Una Pale sa saradnicima prikazuje novu metodu analize impedansnog kardiograma (eng. *Impedance Cardiography*, skraćeno ICG ili IKG), a uz naučni rad [95], Autori su objavili i podatke [96]. Pale i saradnici su se opredelili da IKG podatke podele sa .mat ekstenzijom. Datoteke sa .mat ekstenzijom predstavljaju binarni Matlab format [97]. Najčešće je razlog za odabir formata koji nisu široko zastupljeni, a koji su usko vezani za softver, to što je za analizu ili prikupljanje signala korišćen isti softver. DANS organizacija preporučuje korišćenje široko rasprostranjenih otvorenih formata prilikom arhiviranja i deljenja istraživačkih podataka [90]. Za pristup podacima sa .mat ekstenzijom, potrebno je instalirati CRAN paket R.matlab [98] i iskoristiti:

```
readMat()
```

funkciju za učitavanje podataka sa .mat ekstenzijom, kao što je to urađeno u nedavnoj studiji [99-101] u kojoj je predstavljena nova metoda za delineaciju IKG signala koja je testirana, između ostalog, i na podacima sa .mat ekstenzijom [96].

Ovaj udžbenik nema za cilj da predstavi sve funkcije za čitanje podataka u R-u, već da pruži osnovne smernice za rad kako bi studenti bili u mogućnosti da samostalno koriste ugrađene i najčešće korišćene funkcije, ali i funkcije ili pakete koji će tek biti razvijeni. Neki od parova funkcija za čitanje i upis podataka u datoteku su:

readLines()/writeLines()

za rad sa datotekama bez ograničenja u ekstenziji i za učitavanje/ispis proizvoljnog dela datoteke (broj linija za učitavanje/ispis određuje korisnik):

source()/dump()

za učitavanje ili upisivanje R koda tj. funkcija⁷⁷:

dget()/dput()

se koristi za rad sa R objektima (češće se R objekti čuvaju u datoteke ili sa .csv ili sa .txt ekstenzijom):

load()/save()

za rad sa datotekama koje sadrže binarne podatke (ove ekstenzije su uobičajene kada se vrši ušteda memorije), a potom:

unserialize()/serialize()

su funkcije nižeg nivoa sa rad sa binarnim formatima (eng. *low-level*) i ovde o njima neće biti reči. [3]

Problem smeštanja podataka u RA memoriju računara u programskom jeziku R prilikom pristupa podacima iz datoteke može se delimično rešiti procenom potrebne memorije prilikom učitavanja podataka, kako je predloženo u [3].

⁷⁷ Retko se koriste, jer se učitavanje funkcija u radni direktorijum obavlja pomoću drugih alata u R Studio razvojnom okruženju, a koje omogućavaju opcije korisničkog interfejsa.

Najpre je potrebno proceniti memoriju podataka koji su smešteni u datoteku, a moguće je podesiti i tip podataka koji se učitavaju kako bi se ubrzao proces čitanja primenom argumenta *colClasses* unutar funkcije za učitavanje podataka iz datoteke. Obe procedure (procena memorije i promena tipa argumenata) mogu da pomognu da se podaci ili učitaju deo po deo ili da se smeste u postojeću memoriju na računaru.

Pre upliva u analizu podataka koja zauzimaju relativno veliki⁷⁸ memorijski prostor, potrebno je proveriti koliko je memorije dostupno na računaru, da li se koriste i druge aplikacije pored programskog jezika R, da li postoji više aktivnih korisnika na računaru i da li se koristi 32-o bitni ili 64-oro bitni operativni sistem. Dostupnost memorijskog prostora nije najjednostavnije odrediti, posebno u uslovima u kojima se ona menja u vremenu. Jednostavan način (eng. *dirty way*) predložen u [3] je ovde opisan na primeru datoteke koja sadrži elektromiogram (EMG).

Neka tekstualna datoteka sadrži 30 000 vrsta i 20 kolona numeričkih podataka koji potiču od višekanalnog merenja EMG signala (sa 20 kanala) u trajanju od 30 s sa frekvencijom odabiranja od 1000 odbiraka u sekundi⁷⁹. Za merenje EMG signala primenom površinskih Ag/AgCl (srebro/srebro-hlorid) elektroda 1000 Hz je sasvim dovoljna frekvencija odabiranja po Nikvist-Šenon-Koteljnikovoj⁸⁰ teoremi odabiranja, jer maksimalni frekvencijski opseg EMG signala može biti do 500 Hz [104-105]. U automatici, preporuka je da se za frekvenciju odabiranja bira vrednost koja je bar 5 puta veća od najveće frekvencije na signalu (preporučeno je da odnos bude od 5 do 30 [106] ili od 5 do 10 [107]) kako bi se omogućila verna digitalna predstava analognog signala. Međutim, u biomedicinskom inženjerstvu se za vrednost maksimalne frekvencije na signalu koristi teorijski maksimum, gde je pritom frekvencija na kojoj signal ima maksimalnu magnitudu značajno manja (u slučaju EMG

78 Misli se na podatke dovoljno velike da mogu da zauzimaju postojeće resurse na računaru, što nije moguće kvantitativno navesti u udžbeniku, jer zavisi od računara na kome se realizuje analiza.

79 Iako je za frekvenciju odabiranja uobičajeno da se koristi Hz, to nije u potpunosti tačno. Pogledati još jednom fusnotu 63.

80 Iako je u literaturi i na kursovima na Elektrotehničkom fakultetu, Univerziteta u Beogradu uobičajeno da se koristi naziv Nikvist-Šenonova teorema, pravi naziv ove teoreme je Vitaker-Koteljnikov-Rabe-Šenon-Someja-Nikvist (eng. Whittaker-Kotelnikov-Raabe-Shannon-Someya-Nyquist), jer je više naučnika istovremeno došlo do istog saznanja. U nekim izvorima je predloženo da se koristi u literaturi naziv teorema odabiranja, kako se ne bi neko od naučnika nepravedno izostavio [102-103]. Enciklopedija Vikipedija sadrži odličan tehnički napisan tekst na ovu temu, https://en.wikipedia.org/wiki/Nyquist%E2%80%93Shannon_sampling_theorem, pristupljeno 03.08.2023.

signala to predstavlja okolinu 150 Hz [104-105]). Kada bi se poredila vrednost od 1000 Hz sa 150 Hz, videlo bi se da je ispunjen uslov iz automatike, jer je 6.67^{81} veće od minimalno preporučenih 5 puta [106-107].

Memorija koja je potrebna za smeštanje *data frame*-a koji sadrži 20-o kanalni EMG signal sa 30 000 odbiraka je otprilike jednaka $30\,000 \times 20 \times 8\text{ B} = 4\,800\,000\text{ B}$. Dobijena vrednost od 4.8 MB bi trebalo da bude manja od polovine RA memorije na računaru [3]. Ako je taj uslov ispunjen onda učitavanje podataka ne predstavlja nikakvo opterećenje za računar. U slučaju da uslov nije ispunjen ne ostaje puno opcija osim da se učita deo po deo podataka iz datoteke ili da se promeni računar. Jedna od opcija, ako se pokrene kod koji učitava podatke koji zauzimaju više do polovine RA memorije je da se uradi reset (eng. *reboot*) računara ili da se zaustavi R proces na računaru. Memoriju je moguće proceniti i korišćenjem gotovih funkcija u R-u kao što je:

object.size()

funkcija. Ova funkcija se nalazi u R.utils paketu [109] (koji je automatski učitana prilikom pokretanja R Studio okruženja). Postoje i druge realizacije ove funkcije, ali ono što je važno znati je da će se dobiti razlika u primeni jednostavnog načina za računanje potrebnog memorijskog prostora [3] u odnosu na primenu *object.size()* funkcije [109-110]. Naime, i prazni nizovi u R-u zauzimaju određen memorijski prostor (48 B) što se može, na primer, proveriti pozivom:

object.size(numeric())

jer memoriju zauzimaju i metapodaci, pokazivači na objekat i alocirani prostor za objekat koji se kreira [110]. Postoje i određene uštede u alociranju memorijskog prostora u R-u [110] koje se odnose na rad sa stringovima koji se ponavljaju, ali oni u ovom udžbeniku nisu od posebnog značaja.

Preporučen paket za učitavanje signala iz datoteka je *readr* paket [111], posebno kada ne postoji procena o memoriji koju će zauzeti podaci ili kada se očekuje rad sa relativno velikom memorijom podataka. Ovaj paket je razvijen sa ciljem da se unaprede često korišćene *read.csv()* i *read.table()* funkcije koje u *readr*

81 Autorka je svesna činjenice da se u srpskom jeziku decimalne vrednosti brojeva predstavljaju sa decimalnim zarezom “,”. Međutim, kako se u R-u podrazumevano koristi decimalna tačka “.”, radi jednostavnosti i saobraznosti odabrano je da se realni brojevi u ovom udžbeniku predstavljaju sa decimalnom tačkom, kako je to prikazano i za odnos frekvencije odabiranja sa maksimalnom frekvencijom na signalu.

paketu [111] nose naziv *read_csv()* i *read_table()*. Pored brzine učitavanja podataka koja je poboljšana, ističe se posebna pogodnost u korišćenju ovih funkcija u indikaciji procesa učitavanja (eng. *progress meter*). U ovom udžbeniku će se koristiti funkcije iz osnovnog R paketa za aučitavanje podataka relativno malih dimenzija, te readr paket neće dalje biti korišćen. Zbog svoje efikasnosti i dodatne funkcionalnosti, readr paket se svakako preporučuje za dalju upotrebu.

```
> con <- url("https://automatika.etf.bg.ac.rs/sr/13m051tobs")
> x <- readLines(con)
Warning message:
In readLines(con) :
  incomplete final line found on 'https://automatika.etf.bg.ac.rs/sr/13m051tobs'
> head(x, 10)
[1] "<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
 \"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd\">"
[2] ""
[3] "<html xmlns=\"http://www.w3.org/1999/xhtml\" xml:lang=\"sr-yu\"
 lang=\"sr-yu\">"
[4] ""
[5] "<head>"
[6] "  <script type=\"text/javascript\">"
[7] "var siteurl='/';"
[8] "var tplurl='/templates/ja_university/';"
[9] "var isRTL = false;"
[10] "</script>"
> close(con)
> |
```

Slika 21, Primer korišćenja *url()* funkcije za preuzimanje HTML sadržaja internet stranice TOBS predmeta (<https://automatika.etf.bg.ac.rs/sr/13m051tobs>, pristupljeno 03.08.2023.) koji je korišćen u vežbi 3 TOBS predmeta. Spisak svih vežbi sa prapatnim materijalima je dat u Dodatku B na kraju udžbenika.

Rad sa podacima koji su sačuvani u tekstualnim datotekama je najčešći slučaj u praksi [3]. Neke od prednosti rada su mogućnost ručnog menjanja podataka (koja ujedno može da bude i mana), metapodaci su sačuvani i čitljivi, ali i u slučaju potrebe mogu se pregledati i popraviti delovi u kojima je došlo do korupcije podataka. Mane rada sa tekstualnim datotekama su da ovakvi podaci zauzimaju više memorije od ostalih formata (što predstavlja jedan od glavnih razloga što su biosignali dostupni u Physionet bazi [112] podataka sačuvani u binarnom formatu) i činjenica da metapodaci često nemaju odgovarajuću strukturu te mogu biti samo delimično čitljivi. Prema tome, binarni formati su često u upotrebi zbog svoje efikasnosti tj. zato što zauzimaju relativno mali memorijski prostor u odnosu na ostale formate. Takođe, bolji su od tekstualnog formata koji može da uzrokuje smanjenje numeričke preciznosti u podacima.

Profesor Peng ističe da su podaci u tekstualnim datotekama deo *Unix* filozofije tj. tehničke kulture [3]. Ova kultura podrazumeva da je univerzalni interfejs u tekstualnoj formi (kao u *Unix-u*), da tip podataka određuje konačnu aplikaciju, a ne algoritam. Za sve koje više zanima o *Unix* filozofiji mogu pogledati knjigu *Umetnost Unixa programiranja*⁸² [113].

```
> dat <- read.table(url("http://www.statsci.org/data/oz/ms212.txt"),
+                   header = TRUE)
> head(dat)
  Height weight Age Gender Smokes Alcohol Exercise Ran Pulse1
1    173    57  18     2      2      1         2  2     86
2    179    58  19     2      2      1         2  1     82
3    167    62  18     2      2      1         1  1     96
4    195    84  18     1      2      1         1  2     71
5    173    64  18     2      2      1         3  2     90
6    184    74  22     1      2      1         3  1     78
  Pulse2 Year
1     88   93
2    150   93
3    176   93
4     73   93
5     88   93
6    141   93
> |
```

Slika 22, Primer korišćenja *url()* funkcije u kombinaciji sa *read.table()* funkcijom za čitanje sadržaja tekstualne datoteke direktno sa interneta. Podaci koji su učitani su prikazani pomoću *head()* funkcije za prikaz prvih 6 vrsta iz datoteke “ms212.txt” u konzoli R Studio razvojnog okruženja. Podaci i primer čitanja podataka su preuzeti iz vežbe 5 TOBS predmeta, a koji su dostupni na <http://www.statsci.org/data/oz/ms212.txt>, pristupljeno 03.08.2023. [114]. Spisak svih vežbi sa propratnim materijalima je dat u Dodatku B na kraju udžbenika.

Podaci se mogu učítavati i direktno sa interneta primenom:

url()

funkcije. Za čitanje bilo koje internet stranice moguće je dobiti HTML (eng. *Hyper Text Markup Language*) na izlazu. Osim pristupa i mogućnosti čuvanja statičkog sadržaja internet stranice (primer iz vežbe 3 TOBS predmeta je prikazan na Sl. 21) i direktnog pristupa datotekama koje se nalaze na internet stranicama (primer iz vežbe 5 iz TOBS predmeta je prikazan na Sl. 22) ovaj način čitanja sadržaja internet stranica može poslužiti i za preuzimanje informacija koje sadrže mrežni paketi. Primiti na Sl. 21 prilikom preuzimanja statičkog HTML sadržaja sa sajta TOBS predmeta, da je korisno koristiti *url()* u

82 Ovo je slobodan prevod naslova sa engleskog na srpski jezik. Naslov knjige u originalu je “*The art of Unix programming*” [113].

kombinaciji sa `readLines()` funkcijom za kreiranu vezu, kao i da je potrebno zatvoriti je nakon završnog čitanja sa sajta primenom funkcije `close()`.

Dobra praksa nalaže da internet stranice moraju zaštititi mrežne pakete, te primena `url()` funkcije za preuzimanje mrežnih paketa, posebno automatski kreiranim R kodom nije dobra, a u većini situacija nije ni dozvoljena. Preuzimanje mrežnih paketa bi moglo da se realizuje kod internet stranica kod kojih dolazi do stalne promene sadržaja. Ako se čitanje sadržaja internet stranice realizuje u kraćim vremenskim intervalima od intervala u kojima se sadržaj stranice menja (u skladu sa teoremom odabiranja), onda je moguće preuzeti sadržaj stranice koji inače ne bi bio dostupan u jednom vremenskom trenutku. Primer stranice na kojoj se podaci periodično menjaju, a njihove vrednosti ispusuju numerički na stranici je stranica Moj Doktor⁸³ Ministarstva zdravlja, Republike Srbije. Presretanje mrežnih paketa je poznato pod nazivom *sniffing attack*⁸⁴ (njuškanje⁸⁵) i podrazumeva niz metoda koje prevazilaze mogućnosti korišćenja `url()` funkcije u programskom jeziku R. Jedan primer loše prakse je testiran i na Elektrotehničkom fakultetu, Univerziteta u Beogradu⁸⁶ kada je za potrebe preuzimanja sadržaja sa internet stranice na kojoj u intervalima od 1 s prikazivani podaci sa senzora temperature i vlažnosti vazduha primenom `url()` funkcije u R-u. Ovaj primer preuzimanja podataka je prikazan u prezentaciji “IoT⁸⁷ & električna merenja” za predavanje po pozivu [115]. Problem ovakvog pristupa podacima je razrešen tako što podaci više nisu numerički prikazivani na internet stranici, već samo u formi grafika. Ima i drugih rešenja.

Ako su podaci, koje je potrebno učitati, dostupni na internetu, onda im je moguće direktno pristupiti sledećom komandom `read.csv(url("http://...csv"))`, kao što je prikazano na Sl. 22. Ovde bi trebalo voditi računa o ponovljivosti i računarskoj reproducibilnosti koda. Naime, internet stranice mogu biti podložne promenama, tako da je sigurnije preuzeti “ručno” podatke i smestiti ih u lokalnu memoriju računara. Ovo ne važi ako su podaci dostupni u nekom od repozitorijuma i podeljeni pod FAIR (od eng. *Findability, Accessibility, Interoperability, Reusability*) principima [116]. FAIR principi predstavljaju najvažnije uslove da podaci koji se javno dele budu pronađeni, citirani i

83 <https://live.mojdoktor.gov.rs/>, pristupljeno 03.08.2023.

84 https://en.wikipedia.org/wiki/Sniffing_attack, pristupljeno 03.08.2023.

85 Direktnan prevod termina *sniffing* sa engleskog jezika je njuškanje.

86 Ovo testiranje je izvedeno uz saglasnost svih aktera (administratora, vlasnika sajta i podataka) i u obrazovne svrhe.

87 Skraćenica od eng. *Internet of Things*, odnosno internet stvari.

upotrebljeni⁸⁸. Jedan od osnovnih uslova za ispunjenje FAIR principa je postojanje trajnih identifikatora (eng. *persistent identifier*). Najpoznatiji trajni identifikator je DOI (eng. *Digital Object Identifier*), ali ima i drugih. Na primer, baza doktorskih disertacija u Republici Srbiji NaRDuS⁸⁹ (od eng. *National Repository of Dissertations in Serbia*) koje su dostupne javno u skladu sa amandmanima Zakona o visokom obrazovanju iz 2014. godine, ne koristi DOI trajne identifikatore, već handle.net⁹⁰ identifikatore.

Procenjeno je da na internetu postoji veliki broj tzv. trulih linkova ili internet grobova (eng. *link rot/death*⁹¹). Dobra praksa nalaže da se prilikom navođenja hiperlinkova koji nisu trajni identifikatori, uvek doda datum pristupa određenom linku kako je to urađeno u ovom udžbeniku. Za veliki broj citatnih stilova to je i uslov za navođenje izvora sa interneta. Prema tome, ako je nešto dostupno na sajtu, pa makar to bila i zvanična internet stranica predmeta ili Fakulteta, ne mora da znači da će zauvek ostati dostupno na internetu. Pored trajnih identifikatora, postoje i druge opcije da sadržaj sa interneta ostane trajno⁹² dostupan. Primer je internet arhiva pod nazivom *Wayback Machine*⁹³, a uputstvo za korišćenje ove arhive je dostupno na stranici “*Wikipedia:Link rot*”⁹⁴.

Kako je najsigurniji način za pristup podacima čije je učitavanje pomoću *url()* funkcije prikazano na Sl. 22 [114] da se podaci preuzmu u .txt formatu i da se učitaju sa računara, lokalno, to treba dati par napomena kako se tako sačuvanim podacima pristupa. Da bi R kod mogao da pročita podatke pozivom funkcije *read.table()*, potredno je postaviti radni direktorijum tamo gde je sačuvan kod i gde su sačuvani podaci ili je potrebno u nazivu datoteke dodati direktorijum. Na Sl. 23 je prikazano kako se učitavaju podaci preuzeti sa internet stranice <http://www.statsci.org/data/oz/ms212.txt>⁹⁵ (pristupljeno 03.08.2023.) [114] kada su postavljeni u radnom direktorijumu *Downloads*.

88 Preporučeni izvor sa uputstvom za deljenje istraživačkih podataka je sajt Serbia.RDM projekta na Univerzitetu u Beogradu (<https://rdm.open.ac.rs/>, pristupljeno 03.08.2023.) - u budućnosti se očekuje migracija sadržaja Serbia.RDM projekta na internet stranicu Zajednice otvorene nauke Srbije (<https://oscs.open.ac.rs/>, pristupljeno 03.08.2023).

89 <https://nardus.mpn.gov.rs/>, pristupljeno 03.08.2023.

90 <https://hdl.handle.net/>, pristupljeno 03.08.2023.

91 https://en.wikipedia.org/wiki/Link_rot, pristupljeno 03.08.2023.

92 Nije predmet ovog udžbenika, ali bi trebalo spomenuti da se pojam trajno kod arhiviranja i prilikom korišćenja trajnih identifikatora odnosi najčešće na decenije, a ne na milenijume.

93 <https://archive.org/web/>, pristupljeno 03.08.2023.

94 https://en.wikipedia.org/wiki/Wikipedia:Link_rot, pristupljeno 03.08.2023.

95 Za preuzimanje datoteke u .txt formatu direktno sa internet stranice, potrebno je odabranom internet pregledaču odabrati desnim klikom miša opciju “Save Page As” ili “Sačuvaj kao” ili koristiti prečicu sa tastature Ctrl + S.

Datoteka koja se nalazi u direktorijumu koji nije radni direktorijum R Studio razvojnog okruženja se može učitati primenom *read.table()* funkcije tako što se pored naziva datoteke unese i njena lokacija na računaru. U primeru sa Sl. 23 to je “C:\Users\BioEng\Downloads”. Primititi da je ovo lokacija fascikle koju daje računar i koja koristi kosu crtu “\” za koju će R javiti grešku. Potrebno je uneti lokaciju datoteke sa kosom crtom koja je usmerena u suprotnom smeru: “/” kako bi podaci mogli da se učitaju (Sl. 23). Potom je moguće pregledati sadržaj rezultujućeg *data frame*-a pozivom *head()* funkcije (parametar 2 na Sl. 23 označava broj vrsta za prikaz podataka u konzoli). Drugi način je da se koristi funkcija:

setwd()

koja postavlja radni direktorijum na željenu lokaciju na računaru (od eng. *set working directory*). Za pregled trenutnog radnog direktorijuma, može se koristiti funkcija:

getwd()

kao što je prikazano na Sl. 23.

```
> dat <- read.table("C:\Users\BioEng\Downloads\ms212.txt",
Error: '\U' used without hex digits in character string starting ""C:\U"
> dat <- read.table("C:/Users/BioEng/Downloads/ms212.txt",
+                   header = TRUE) # ispravno
> head(dat, 2)
  Height Weight Age Gender Smokes Alcohol Exercise Ran Pulse1 Pulse2 Year
1    173     57  18     2      2        1         2    2     86     88   93
2    179     58  19     2      2        1         2    1     82    150   93
> setwd("C:/Users/BioEng/Downloads")
> getwd()
[1] "C:/Users/BioEng/Downloads"
> if (!require("rstudioapi")) install.packages("rstudioapi")
Loading required package: rstudioapi
> setwd(dirname(rstudioapi::getSourceEditorContext())$path)
> |
```

Slika 23, Učitavanje podataka iz tekstualne datoteke “ms212.txt” koja je sačuvana lokalno na računaru. Za automatsko podešavanje direktorijuma korišćene su funkcije *setwd()* i *getSourceEditorContext()* iz *rstudioapi* paketa [117].

Na kraju pomoću `rstudioapi` paketa⁹⁶, moguće je uvek postaviti automatski radni direktorijum na lokaciju gde je snimljen R kod [117]. Poziv funkcije iz `rstudioapi` paketa:

```
rstudioapi::getSourceEditorContext()
```

je isti kao:

```
getSourceEditorContext()
```

Jedina razlika između ova dva poziva je što je u prvom slučaju naglašeno da funkcija pripada `rstudioapi` paketu pomoću operatora sa dve dvotačke `::`. Postojanje ovog operatora je vrlo važno, jer sa oko 20 000 paketa trenutno dostupnih u CRAN repozitorijumu (Sl. 2), nemoguće je da svaki paket poseduje funkcije sa različitim imenima. Naglašavanje poziva funkcije se može obaviti i primenom:

```
library()
```

funkcije koja omogućava da se paket aktivira. Naime, ako postoji veći broj funkcija sa istim imenom, a funkcija se na nekom mestu u kodu pozove bez operatora `::`, pozvaće se funkcija iz poslednji paket koji je aktiviran primenom `library()` komande. Prirodno se nameće da je dobra praksa da se u kodovima u kojima se koristi relativno veliki broj paketa uvek sa operatorom `::` naglasi iz kog se paketa poziva neka od funkcija.

U slučaju da su podaci smešteni u komprimovanim datotekama⁹⁷ (na primer, otvoreni `gzip`⁹⁸ i `bzip`⁹⁹ formati), nije potrebno da ih korisnik raspakuje ručno, već se i taj proces može automatizovati u programskom jeziku R. Na sličan način kao što se to radi sa internet adresom pomoću `url()` funkcije, moguće je “otvoriti konekciju” za rad sa komprimovanim datotekama. Nakon čitanja podataka, konekciju bi trebalo zatvoriti. U najvećem broju slučajeva otvaranje i zatvaranje konekcije se obavlja u pozadini. Otvaranje se obavlja sa `readLines()`

96 Kada se koristi dvotačka operator `::` nije potrebno uključiti funkcionalnost nekog paketa primenom `library()` funkcije. Prema tome, prilikom korišćenja automatskog podešavanja radnog direktorijuma pozivom funkcije iz `rstudioapi` paketa nije potrebno da se koristi `library()` funkcija, ali je potrebno da se ovaj paket instalira na računaru [117].

97 Komprimovanje podataka se radi sa ciljem da se smanji memorija koju podaci zauzimaju na računaru.

98 <http://gzip.org/>, pristupljeno 03.08.2023.

99 <https://sourceware.org/bzip2/>, pristupljeno 03.08.2023.

ili sa `writeLines()`, a zatvaranje sa funkcijom `close()`. Iako je rad sa bazama podataka, namenjen primarno softverskim inženjerima i inženjerima računarske tehnike i informatike, nije na odmet da i biomedicinski inženjeri budu upoznati sa osnovnim elementima rada sa bazama podataka, jer nije isključeno da će u praksi biti u prilici da pristupaju bazama podataka. U R-u je moguće koristiti funkcionalnost SQL¹⁰⁰ (eng. *Structures Query Language*) programskog jezika i obrnuto. U `dplyr` paketu, postoji funkcija `translate_sql()` koja se koristi za automatsko prevođenje R koda u SQL¹⁰¹ kod [38]. Preporučeni primer pristupa bazi podataka i rada sa relacionim bazama u programskom jeziku R na TOBS predmetu je sadržan u video lekciji¹⁰² DataCamp¹⁰³ kursa [118]. Video lekcija je posebno preporučena studentima i čitaocima koji se nikada ranije nisu sreli sa bazama podataka jer u lekciji [118] objašnjeno šta su baze podataka na primeru relacionih tabela zaposlenih i njihovog učinka.

Signali se mogu, direktno sa senzora, uvesti/izmeriti u programski jezik R primenom serial paketa sa CRAN-a [31]. Ova biblioteka, kako je već rečeno, služi za merenje signala sa hardvera koji koristi RS232/RS422/RS485 ili bilo koji drugi virtuelni serijski interfejs. Na TOBS predmetu je ova biblioteka korišćena za merenje signala u programskom jeziku R, primenom UNO R3 mikrokontrolerske pločice i Arudino¹⁰⁴ razvojnog okruženja (Arduino, Monza, Lombardia, Italy).

UNO R3 nije ništa drugo do Ardiuno UNO pločica, ali koja nije proizvedena i puštena u promet od strane ovlašćenih Arduino proizvođača i distributera. Potpuno je legalna praksa proizvodnje ovakvih pločica, jer je Arduino hardver u stvari otvoren hardver¹⁰⁵ i podeljen je pod otvorenom CC licencom [119]. Kod ovakvih mikrokontrolerskih pločica nije dopušteno da se koristi zaštitni znak ni naziv Arduino (eng. *Trademark*) [119-120].

100 <https://en.wikipedia.org/wiki/SQL>, pristupljeno 04.08.2023.

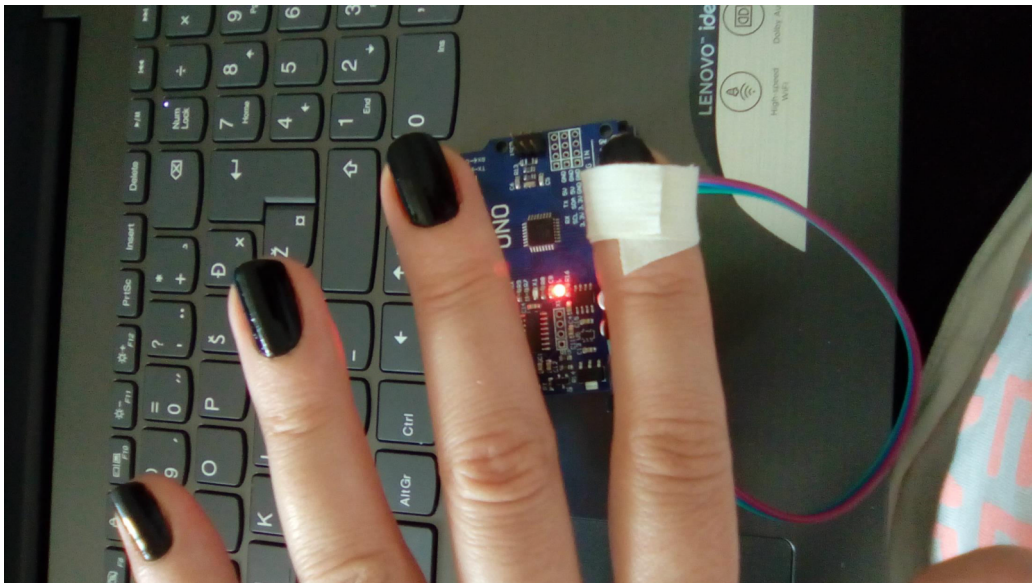
101 Izgovara se sikvel ili eskjuel. Oba su opšteprihvćena.

102 <https://youtu.be/9OSB9pmlJpI>, pristupljeno 04.08.2023. [132]

103 <https://www.datacamp.com/>, pristupljeno 04.08.2023.

104 <https://www.arduino.cc/>, pristupljeno 04.08.2023.

105 Otvoren hardver prati principe koji su pandan slobodnom softveru. Na engleskom jeziku više je termina u upotrebi: *Open Hardware*, *Open Science Hardware*, *Open Design*, *Free Hardware Design*, *Free Hardware* i drugi. Za razliku od softvera koji se može jednostavno kopirati i štiti slobodnim licencama, kod hardvera je nešto komplikovaniji slučaj, jer je za kopiranje hardvera ipak potrebno izdvojiti odgovarajuća sredstva i nabaviti odgovarajuće komponente [14]. Bez obzira na ova ograničenja za odgovarajuće definisanje licence otvorenog hardvera, postoji niz licenci koje se mogu koristiti od kojih su najpoznatije CERN licence koje je razvila CERN (eng. *The European Organization for Nuclear Research*, <https://www.home.cern/>, pristupljeno 04.08.2023.) organizacija za nuklearna istraživanja.



Slika 24, Senzor pulsa postavljen na kažiprst leve ruke i povezan sa UNO R3 mikrokontrolerskom pločicom i laptop računarem za demonstraciju merenja u programskom jeziku R primenom serial paketa [31].

U ovom udžbeniku, predstavljen je princip merenja biosignala u programskom jeziku R primenom senzora za merenje pulsa¹⁰⁶. Kako bi se izmerio signal sa Arduino ili UNO R3 mikrokontrolerske pločice, najpre je potrebno proveriti na kom portu je povezan uređaj sa računarem (za UNO R3 na Sl. 24 korišćen je USB¹⁰⁷ port za povezivanje mikrokontrolerske pločice sa računarem). Primer koji je prikazan u ovom udžbeniku je primer iz lekcije 9 TOBS predmeta (spisak svih lekcija i trajni identifikatori za pristup odgovarajućim prezentacijama su dati na kraju ovog udžbenika) i realizovan je na operativnom sistemu Vindovs¹⁰⁸, ali ga je moguće realizovati i pod drugim operativnim sistemima, isto kao što je moguće koristiti programski jezik R i razvojno okruženje R Studio sa drugim operativnim sistemima. U R kodu potrebno je podesiti na kom COM (eng. *Communication*) portu je povezana mikrokontrolerska pločica, te je zato važno proveriti na računaru koji COM port se koristi. COM port na Vindovs operativnom sistemu može se videti u *Control Panel* → *Panel* → *Device Manager* → *Ports*, ali i u sekciji *Devices and Printers* na računaru. Ima više načina kako da se proveri COM port na koji je povezana mikrokontrolerska pločica, a ovde su navedena samo dva.

106 <https://pulsesensor.com/>, pristupljeno 03.08.2023.

107 Od eng. *Universal Serial Bus*, <https://en.wikipedia.org/wiki/USB>, pristupljeno 03.08.2023.

108 Windows OS (od eng. *Operating System*), <https://www.microsoft.com/en-us/windows>, pristupljeno 04.08.2023.

Merenje sa senzora pulsa je realizovano preko analognog kanala A0 mikrokontrolerske pločice UNO R3 sa frekvencijom odabiranja od 100 Hz. Ovde neće biti reči o programiranju u Arduino okruženju niti o tome kako se postiže konstanta frekvencija odabiranja, osim napomene da se nikako ne sme koristiti funkcija *delay()*, već tajmeri da bi se postiglo merenje signala sa analognog kanala sa konstantnom frekvencijom odabiranja. Studenti se upućuju na materijale za predmet Merni sistemi u računarstvu (MSR¹⁰⁹) na Elektrotehničkom fakultetu, Univerziteta u Beogradu i na udžbenik za MSR predmet [78], kao i knjigu sa osnovama programiranja u Arduino okruženju [121]. Kod za merenje sa konstantnom frekvencijom odabiranja je takođe dat na sajtu TOBS predmeta u datoteci pod nazivom “AnalogReadBezKasnjenja.ino”¹¹⁰.

Procedura za merenje signala sa senzora pulsa u programskom jeziku R je sledeća:

- senzor za merenje pulsa se poveže sa UNO R3 pločicom (izlaz iz senzora se poveže na A0 port, a za napajanje senzora se koriste namenski pinovi sa DC (eng. *Direct Current*) napajanjana računaru)
- UNO R3 pločica se preko USB porta poveže sa laptop računarom¹¹¹ koji nije priključen na eksterno napajanje,
- pokrene se Arduino kod za merenje signala sa frekvencijom odabiranja od 100 Hz (“AnalogReadBezKasnjenja.ino”),
- Arduino kod¹¹² se “spusti” u mikrokontroler na UNO R3 pločici pozivom komande *Upload*, nakon što se podesi odabir hardvera i COM porta iz padajućeg menija *Tools*,
- potom se pokrene R kod koji omogućava pristup serijskom portu realizacijom odgovarajuće konekcije (Sl. 25) – broj COM porta potrebno je podesiti i u programskom jeziku R,

109 <https://automatika.etf.bg.ac.rs/sr/13e053msr>, pristupljeno 04.08.2023.

110 https://automatika.etf.bg.ac.rs/images/FAJLOVI_srpski/predmeti/izborni_kursevi_os/biomedicinsko_inzenjerstvo/TOBS/predavanja/analogReadBezKasnjenja.zip, pristupljeno 04.08.2023.

111 Računar ne bi trebalo priključiti na napajanje, jer senzor i pločica ne poseduju galvansku izolaciju. Potrebno je koristiti baterijsko napajanje laptop računara. Ovaj udžbenik se ne bavi merenjem biomedicinskih signala, a za više informacija o galvanskom razdvajanju u medicinskim instrumentima pogledati [122]. Dodatno, senzor koji se ovde koristi radi po principu generisanja i detekcije svetlosti, tako da nije u električnom kontaktu sa ispitanikom.

112 Rezultat koda bi trebalo da bude ispis vrednosti odbiraka signala sa analogno digitalnog A/D (eng. *Analog to Digital*) konvertora na serijski port.

- podaci koji se čitaju sa serijskog porta su tekstualnog tipa, pa ih je potrebno konvertovati u numeričke vrednosti i
- obavezno je zatvaranje konekcije na kraju koda, kako bi serijski port bio slobodan da mu se pristupi ponovo iz R-a ili nekog drugog programskog jezika (Sl. 25).

```

library(serial)
con <- serialConnection(name = "test_con",
                        port = "COM4",
                        mode = "9600,n,8,1",
                        newline = 1,
                        translation = "cr")

open(con)

stopT <- Sys.time() + 7
foo <- ""
textSize <- 0
while(Sys.time() < stopT){
  newText <- read.serialConnection(con)
  if(0 < nchar(newText))
  {
    foo <- paste(foo, newText)
  }
}
cat("\r\n", foo, "\r\n", file = "podaci.txt")
close(con)

dat <- read.table("podaci.txt")
plot(dat$V1, type = "l")

```

Slika 25, Realizacija koda u programskom jeziku R za merenje signala tj. čitanje vrednosti odbiraka signala sa serijskog porta.

Kod koji omogućava merenje signala u R-u je prikazan na Sl. 25 i prilagođen je za potrebe TOBS predmeta iz [123]. Predloženi kod se može koristiti sa bilo kojim drugim hardverom koji vrši ispis merenog signala na serijskom portu. Najpre je potrebno da se formira serijska veza primenom funkcije:

serialConnection()

u kojoj je jedan od najvažnijih parametara naziv porta (ovde je postavljen na COM4). Nakon toga se otvara veza za komunikaciju sa serijskim portom primenom funkcije:

open()

koju obavezno na kraju prati funkcija za zatvaranje serijskog porta:

close()

Nakon inicijalizacije tri promenljive *n*, *foo* i *textSize*, pokreće se *while* petlja (kontrolne strukture će biti predstavljene u narednim poglavljima). Untar petlje najvažnija je funkcija:

read.serialConnection()

koja omogućava da se pročita sadržaj sa serijskog porta. Ostale funkcije koje su korišćenje u kodu sa Sl. 25 služe za konverziju podataka i čuvanje merenih odbiraka u datoteku “podaci.txt”¹¹³ i za vremensko ograničenje *while* petlje na 7 s. Poslednje dve linije koda služe za čitanje signala iz tekstualne datoteke i za jednostavan prikaz signala na grafiku primenom ugrađene:

plot()

funkcije iz osnovnog R paketa (eng. *base* biblioteka). Na grafiku je prikazana prva kolona iz *data frame*-a *dat*, a parametar *type* kome je dodeljena vrednost “l” (od eng. *line*) označava tip prikaza, odnosno linijski prikaz. Za vrednost “p” (od eng. *points*) odbirci na grafiku neće biti spojeni linijama već će biti predstavljeni tačkama, a za vrednost “b” (od eng. *both*) odbirci na grafiku će biti predstavljeni i sa tačkama i spojeni linijama. Za ostale moguće vrednosti parametra *type*, trebalo bi pogledati uputstvo za korišćenje *plot()* funkcije, pozivom iz komandne linije: *?plot* nakon čega će se otvoriti uputstvo u Help kartici R Studio razvojnog okruženja ili će korisnik u Help kartici imati mogućnost da bira između više *plot()* funkcija, ako postoji više paketa sa istoimenom funkcijom. Realizacija *plot()* funkcije za različite vrednosti parametra *type* odgovara *plot()* i *stem()* funkcijama u Matlabu i GNU Octave-u za koje se nekada kaže da omogućavaju kontinualni i diskretni prikaz signala, odnosno odbiraka signala. Da ne bi došlo do zabune, ne postoji kontinualni

113 Zajedno sa R kodom koji je podeljen uz ovaj udžbenik (<https://github.com/NadicaSm/TOBS>) i u kome su realizovani svi primeri dati u udžbeniku, podeljeni su i podaci mereni sa senzori pulsa koji se nalaze u datoteci “podaci.txt”.

prikaz na računaru, jer su signali sa kojima se radi uvek digitalni (u ovom slučaju direktno mereni sa A/D konvertora). Najtačniji prikaz takvih signala je uz *type* = "p", jer opcija *type* = "l" iako ostavlja vizuelni utisak kontinualnog prikaza nije ništa drugo do spajanje postojećih odbiraka pravim linijama (isto važi za *plot()* i *stem()* u Matlabu/GNU Octave-u, ali i za slične funkcije u drugim programskim jezicima). Međutim, za signale koji su odabirani sa relativno velikim brojem odbiraka (50 Hz, 100 Hz ili više tokom, na primer, 10 s merenja) takav prikaz bi bio prenatrpan informacijama i ne bi bilo moguće videti pojedinačne odbirke na grafiku, te je tzv. kontinualni prikaz primereniji. Za manji broj odbiraka koji se prikazuje na jednom grafiku (na primer, 10 ili 20 odbiraka), najinformativnija je varijanta prikaza za parametar *type* = "b".

Radi jednostavnog prikaza, u ovom slučaju ose nisu obeležene ni u kodu sa Sl. 25 ni na Sl. 26. Iako će kasnije biti reči o pravilima prikaza signala, važno je naglasiti da se ose uvek obeležavaju. Ovde to pravilo nije ispoštovano iz dva razloga – prvi je jer je cilj bio da se prikaže kako se najjednostavnije koristi *plot()* funkcija, a drugi jer bi takvi elementi učinili kod nečitkim na trenutnom nivou znanja čitaoca. Vrednost ordinate¹¹⁴ (tzv. *y* ose) je napon (originalno u Arduinu su bili kvanti), a vrednost apscise (tzv. *x* ose) je redni broj odbiraka. Kvanti potiču od rezolucije A/D konvertora UNO R3 mikrokontrolerske pločice koja iznosi 10 bita. Praktično, naponski opseg A/D konvertora se deli sa 2^{10} mogućih i konačnih naponskih nivoa koji se mere [78]. Vrednosti koje daje A/D konvertor na UNO R3 pločici su u opsegu od 0 do 1023, što odgovara naponu od 0 V do 5 V, te je potrebno jednostavnom proporcijom najpre ove vrednosti u kvantima pretvoriti u napon¹¹⁵. Potpuno legitimno bi bilo i da se merena vrednost normalizuje od 0 do 1, jer je za merenje koje je ovde prikazano važan oblik signala (iz koga bi trebalo izdvojiti informaciju o srčanim otkucajima), a pri tom nije važno kolika je amplituda signala. Za razliku od amplitude, vrednost *x* ose je značajna u određivanju srčanog ritma, jer se rastojanje između dva srčana otkucaja meri na ovoj osi.

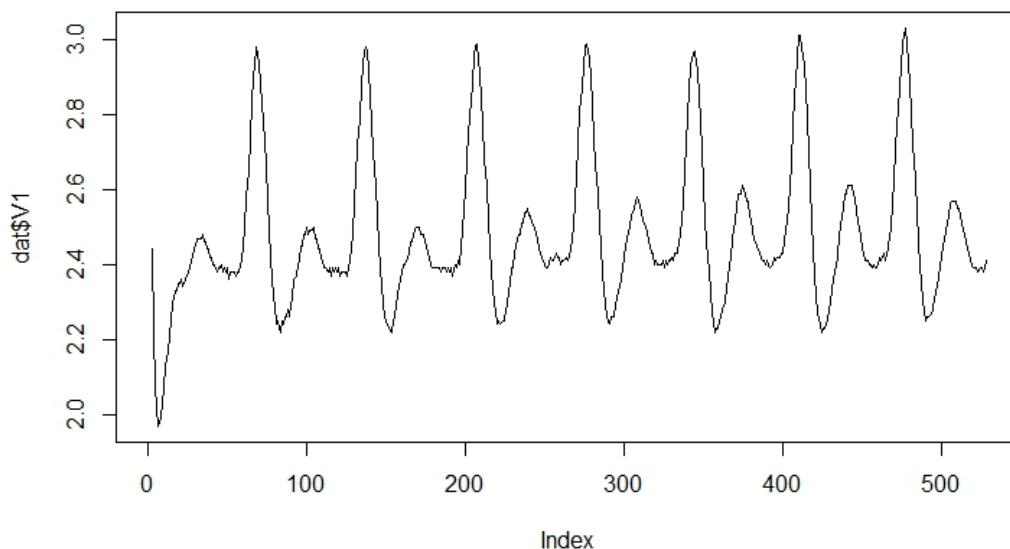
Na CRAN repozitorijumu dostupan je i Rduino paket [124] koji se može koristiti kao direktan interfejs sa mikrokontrolerima, bez korišćenja serijskog porta. Za kraj, ovo nije jedini način da se mere signali u programskom jeziku R, kao ni da se koristi programski jezik R u kodu za merenje signala. Moguće je i u Arduino razvojnom okruženju pristupati programskom jeziku R, odnosno moguće je koristiti R kod. Za te potrebe koristi se R server [125]. Pored

114 U inženjerstvu je neprimereno reći *x* osa i/ili *y* osa, a da prethodno nije objašnjeno da li se ovi nazivi odnose na ordinatu ili apscisu.

115 U Arduino razvojnom okruženju postoji i ugrađen primer koda pod nazivom "ReadAnalogVoltage.ino" u kome se može pogledati odgovarajuća relacija.

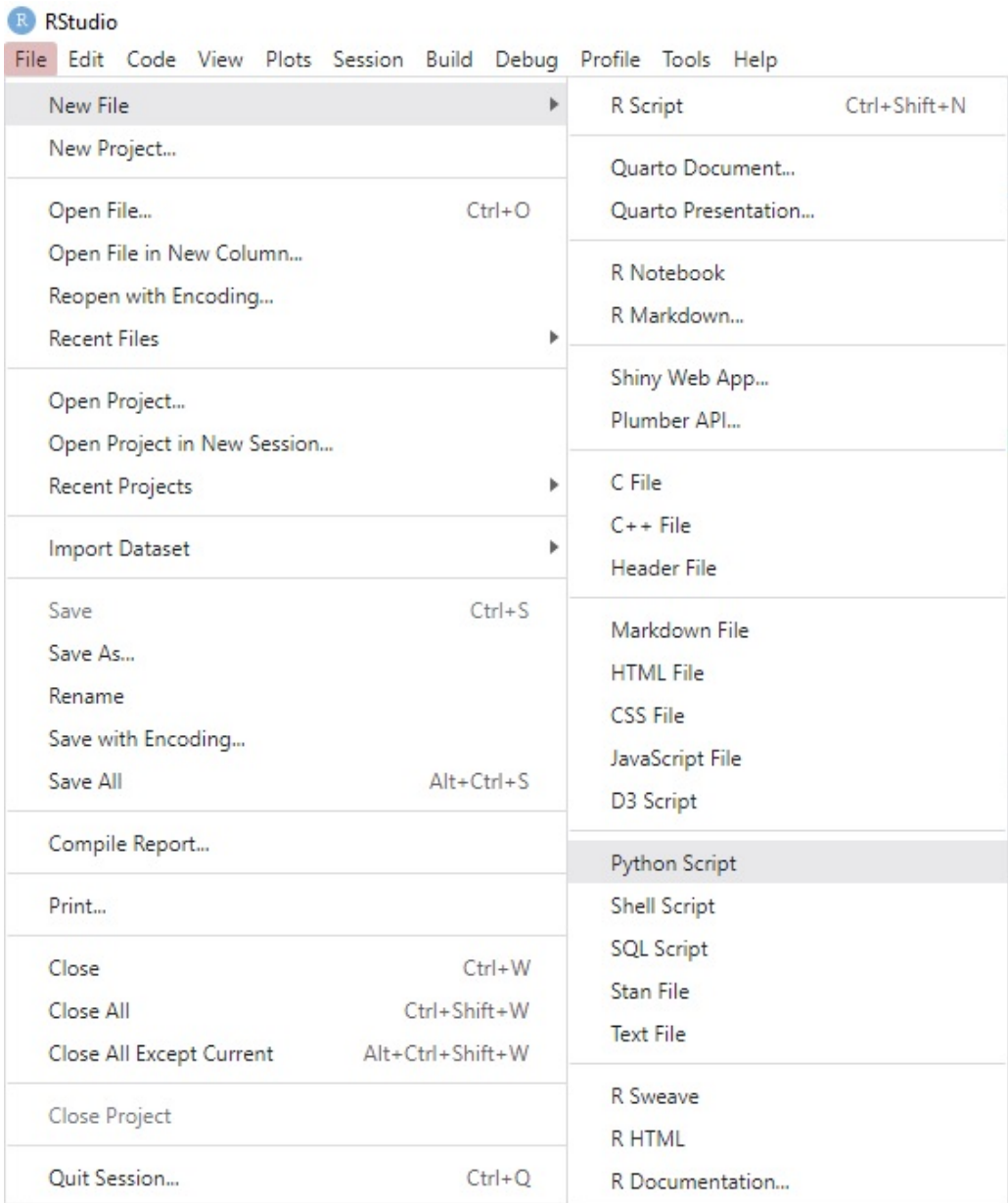
povezivanja Arduino i R programskih okruženja, moguće je povezati, na primer, R i Pajton programske jezike na sledeće načine:

- U CRAN repozitorijumu postoji *reticulate* paket koji omogućava da se Pajton koristi u okviru programskog jezika R uz automatsku konverziju tipova podatka iz jednog programskog jezika u drugi [126]. Ovaj paket omogućava da se u okviru programskog koda realizuju sekcije sa Pajton programskim jezikom (eng. *chunk*).
- U Pajtonu postoji *subprocess* modul koji se može koristiti kao interfejs ka drugim programskim jezicima uključujući R programski jezik [127].
- Takođe, *rpy2* – Pajton paket se može koristiti kao direktan interfejs ka programskom jeziku R [128].
- Razvojno okruženje R Studio nije namenjeno isključivo programskom jeziku R, iako se za njega dominantno koristi. Namenjeno je i drugim programskim jezicima uključujući Pajton. Primer kreiranja Pajton koda tj. skripte u R Studio okruženju je dat na Sl. 27. Detaljno uputstvo za rad u R Studio okruženju se može naći u Dodatku C na kraju udžbenika.



Slika 26, Prikaz signala iz datoteke “podaci.txt” za signal meren sa senzora pulsa. Kod koji generiše grafik prikazan na ovoj slici je dat na Sl. 25.

Iako ovde nije prikazana iscrpna lista koje sve mogućnosti postoje za realizaciju interfejsa između programskih jezika, jasno je da sa ovakvim alatima mnogobrojni tekstovi koji se mogu naći na internetu, a koji su usmereni ka tome koji je programski jezik bolji postaju besmisleni, jer se uvek može napraviti sprega, odnosno interfejs između programskih jezika.



Slika 27, Izgled R Studio razvojnog okruženja. Odabirom padajućeg menija *File* → *New File* moguće je odabrati *Python Script* opciju i koristiti R Studio razvojno okruženje za rad u Pajton programskom jeziku. Primititi koje sve opcije postoje, a posebno da se R Studio može koristiti za razvoj kodova u programskom jeziku SQL za rad sa bazama podataka.

Umesto da studenti gube vreme na poređenje programskih jezika i donošenje odluke koji programski jezik treba da nauče, uvek je dobro rešenje da se nauči što više programskih jezika. U praktičnim projektima u industrijskom i

akademsom okruženju, uvek se za realizaciju analize signala koristi više od jednog programskog jezika. Nekada nije ni moguće realizovati projekat na drugi način (na primer, ako se pristupa podacima u relacionim bazama podataka), a nekada je to stvar izbora pojedinačnih učesnika projekta. Iako se R i Pajton mogu koristiti u skoro istim situacijama, postoje elementi koji su bolje razvijeni u R-u, kao i oni koji su bolje razvijeni u Pajtonu. Dupliranje truda (eng. *reinventing the wheel*) da se nešto razvije u programskom jeziku R, ako već postoji u programskom jeziku Pajton, apsolutno je besmisleno.

Jedan od najkompleksnijih primera uvoza podataka je uvoz iz .pdf¹¹⁶ (od eng. *Portable Document Format*) datoteka [131]. U R-u je za potrebe čuvanja sadržaja iz .pdf dokumenta u mašinski čitljiv format razvijen alat pdftools [132]. Usled nekonzistentnog formatiranja takvih dokumenata (na primer, ako je .pdf dokument nastao spajanjem više drugih dokumenata ili više dokumenata koji su izvorno realizovani u različitim okruženjima), nekada se ne može koristiti paket pdftools, te se mora pristupiti ručnom kopiranju podataka radi postizanja mašinske čitljivosti¹¹⁷ [133].

Na kraju ovog potpoglavlja treba ukazati na to da je učitavanje podataka tek prvi korak u automatskoj analizi signala. “*It’s one thing for data to be captured, and quite another for them to surrender their secrets.*”¹¹⁸ [134].

1.2 Operatori i kontrolne strukture u R-u

Najčešće korišćeni operator u R-u je znak dodele vrednosti koji predstavlja kombinaciju srednje crte i znaka manje od “<-”. Ovaj operator se može zameniti i znakom jednakosti koji je uobičajen u drugim programskim jezicima

116 <https://www.adobe.com/acrobat/about-adobe-pdf.html>, pristupljeno 08.08.2023.

117 Mašinski čitljivi su oni podaci koje je moguće automatski obraditi na računaru. Oni moraju da imaju odgovarajuću strukturu da bi bili na odgovarajući način učitani u programskom kodu. Vrlo je česta zabuna da su svi digitalni dokumenti i podaci mašinski čitljivi [129]. U biomedicinskom inženjerstvu je sa primenom metoda dubokog učenja (eng. *deep learning*) došlo do toga da se analiziraju čak skenirani EKG nalazi koji su originalno bili odštampani na milimetarskom papiru. Iako takva analiza može da pruži korisne rezultate, predstavlja potpuno pogrešan pristup u analizi podataka. Sličan je primer upotrebe metoda dubokog učenja za digitalizaciju takvih snimaka [130], ali se postavlja pitanje zašto bi neko digitalizovao papirne zapise, ako današnji EKG uređaji omogućavaju čuvanje podataka u mašinski čitljivim podacima? Uvek je bolje analizirati mašinski čitljive podatke. Transformaciju iz digitalnog u mašinski čitljiv format treba primenjivati samo u slučaju da ne postoji druga dostupna metoda za pristup podacima.

118 Srp. “*Jedno je imati podatke, a potpuno drugo razumeti njihove tajne.*”, slobodan prevod sa engleskog na srpski jezik [134].

“=” [3]. Međutim, već je naglašeno da operator dodele vrednosti verno predstavlja operaciju koju izvodi, a to je dodeljivanje vrednosti operandu/promenljivoj koja se nalazi sa leve strane znaka koja je jednaka rezultatu operacije sa desne strane znaka. Kada se jednom usvoji ovakav zapis teže je vratiti se na standardan pristup dodeli vrednosti u drugim programskim jezicima. Za one koji žele da usvoje rad sa ovim intuitivnim operatorom dodele vrednosti, potrebno je da znaju da se on ne može koristiti unutar poziva funkcije kada se dodeljuju vrednosti argumentima funkcije, jer je za te potrebe moguće koristiti isključivo znak jednakosti “=”. Za one koji žele da koriste znak jednakosti za dodelu vrednosti, važno je da znaju čemu služi operator “<-” kako bi mogli da razumeju postojeći R kod.

Pored operatora za određivanje podskupa koji su predstavljeni u prethodnom podpoglavlju, kompletna lista aritmetičkih i logičkih operatora u R-u se može naći u otvorenom udžbeniku “*R Programming: Zero to Pro*”¹¹⁹ [135] i u drugoj literaturi. Ovde neće biti prikazan rad sa svim operatorima, već sa odabranim operatorima koji mogu koristiti studentima za generalno upoznavanje sa radom u R-u, kao i za primenu u analizi podataka. Operatori se mogu koristiti za vektorske operacije nad matricama i nizovima. Primena vektorskih operatora smanjuje potrebu za korišćenjem kontrolnih struktura, a slična funkcionalnost postoji i u Matlabu/GNU Octave-u. Na Sl. 28 je prikazan primer korišćenja aritmetičkih i logičkih operatora za rad sa dva niza x i y koji su jednakih dužina. Takođe, prikazano je kako se može implicitno prikazati sadržaj promenljivih x i y u konzoli R Studio razvojnog okruženja samo pozivom naziva promenljive. Eksplicitan prikaz se postiže primenom funkcije:

print()

čiji je osnovni argument naziv promenljive. Rezultat aritmetičke operacije je niz iste klase kao i ulazni nizovi, a rezultat logičke operacije je niz logičkih promenljivih (Sl. 28). Poseban slučaj predstavljaju nizovi različitih dužina. Rad sa takvim nizovima je prikazan na Sl. 29. Primetiti da za razliku od drugih programskih jezika R ne javlja grešku i generiše rezultat koji je jednak dužini većeg niza, tako što replicira niz manje dužine, što je objašnjeno i u poruci upozorenja (eng. *Warning message*) u konzoli R Studio okruženja (Sl. 29). U primerima prikazanim na Sl. 28 i na Sl. 29 korišćeni su nizovi sa istim nazivima: x i y . Iako je u prvom primeru alocirano četiri mesta za niz y , u drugom primeru je ista promenljiva imala dužinu tri. Ovo je još jedna od važnih

119 Direktan link to otvorenog udžbenika je <https://r02pro.github.io/> (pristupljeno 05.08.2023), a spisak operatora se nalazi u potpoglavlju 2.9 “*Summary of Operators*”, <https://r02pro.github.io/summary-of-operators.html>, pristupljeno 05.08.2023.

osobina za rad u programskom jeziku R, a primer promene sadržaja promenljivih je prikazan na Sl. 30.

```
> x <- 1:4; y <- 5:8
> x
[1] 1 2 3 4
> y
[1] 5 6 7 8
> x + y
[1] 6 8 10 12
> x > 2
[1] FALSE FALSE TRUE TRUE
> x > 100
[1] FALSE FALSE FALSE FALSE
> x >= 3
[1] FALSE FALSE TRUE TRUE
> y == 6
[1] FALSE TRUE FALSE FALSE
> x * y
[1] 5 12 21 32
> y / x
[1] 5.000000 3.000000 2.333333 2.000000
> |
```

Slika 28, Primena operatora za vektorske operacije nad nizovima u konzoli R Studio okruženja.

Iako niz karaktera “gold” označava zlatnu boju (eng. *gold*), ovde promenljiva nije prikazana u zlatnoj boji (kao što je to ilustrovano na Sl. 7), jer je na Sl. 30 prikazan rezultat iz konzole R Studio okruženja, a na Sl. 7 je prikazan isečak iz R koda tj. skripte sa ekstenzijom .R. Vektorske operacije se mogu dalje primeniti i na druge tipove podataka. Množenje matrica je ilustrovano na Sl. 31. Primititi da se za množenje matrica element po element koristi operator “*”, dok se za množenje matrica koristi operator “%*%”. Na Sl. 31 se koristi i:

rep()

funkcija (od eng. *replicate*) iz osnovnog R paketa (eng. *base* biblioteka) koja omogućava da se jedna vrednost ponavlja željeni broj puta (ovde se vrednost 10 ponavlja 4 puta).

```

> x <- 1:4; y <- 1:3
> x - y
[1] 0 0 0 3
Warning message:
In x - y : longer object length is not a multiple of shorter object length
> x * y
[1] 1 4 9 4
Warning message:
In x * y : longer object length is not a multiple of shorter object length
> |

```

Slika 29, Primena operatora za vektorske operacije nad nizovima različitih dužina u konzoli R Studio okruženja. Primer je preuzet iz 3. lekcije za TOBS predmet. Spisak svih lekcija i reference ka odgovarajućim prezentacijama su date u Dodatku A na kraju udžbenika.

```

> x <- 1:50
> x
 [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15
[16] 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
[31] 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45
[46] 46 47 48 49 50
> x <- 1
> x
[1] 1
> x <- "gold"
> print(x)
[1] "gold"
> x <- 'gold'
> x
[1] "gold"
> |

```

Slika 30, Primer promene sadržaja promenljive *x*. Primititi upotrebu funkcije *print()* za eksplicitno prikazivanje sadržaja promenljive u konzoli R Studio okruženja. Takođe, primititi da je nizove karaktera moguće definisati podjednako primenom jednostrukih i dvostrukih znaka navoda.

Logički i aritmetički operatori se mogu primeniti i nad podacima koji pripadaju klasama za datum i vreme s tim što se mora voditi računa da operandi pripadaju istim klasama. O vremenskim zonama i prestupnim godinama (eng. *leap year*) ne treba voditi računa, jer je to realizovano automatski. [3]

Kontrolne strukture u R-u su:

if/else, for, while, repeat, break, next i *return*

za koje se kao početak i kraj u R-u koriste vitičaste zagrade “{}”, što je slično kao u programskom jeziku C, ali značajno opterećuje kod za razliku od programskog jezika Pajton u kome se ovakve oznake početka i kraja strukture ne koriste. U slučaju da je zapis ovih struktura u jednom redu, onda nije potrebno u R-u koristiti zagrade za početak i kraj strukture. Na primer, u R-u je moguće koristiti skraćeni zapis uslovne *if/else* strukture primenom operatora za dodelu vrednosti promenljivoj kao što je to prikazano na Sl. 32. U R-u *else* deo uslovne strukture nije obavezan iako sva tri primera na Sl. 32 sadrže ovaj deo. Dobra praksa nalaže da se uvek koristi oznake za početak i kraj strukture.

```
> x <- matrix(1:4, 2, 2)
> y <- matrix(rep(10, 4), 2, 2)
> x
      [,1] [,2]
[1,]    1    3
[2,]    2    4
> y
      [,1] [,2]
[1,]   10   10
[2,]   10   10
> x * y
      [,1] [,2]
[1,]   10   30
[2,]   20   40
> x %*% y
      [,1] [,2]
[1,]   40   40
[2,]   60   60
> |
```

Slika 31, Množenje matrica u programskom jeziku R.

For petlja može imati sličan skraćeni zapis kao i *if/else* struktura. Dodatno, uz ovu petlju može se koristiti funkcija:

seq_along()

koja omogućava da se *for* petlja izvršava element po element (Sl. 33). Očekivano, nije preporučljivo koristiti veliki broj ugnježenih *for* struktura [3], ali one mogu biti korisne kod automatizacije analize signala. Na Sl. 33 prikazani su primeri realizacije *for* petlje sa punim zapisom (sa oznakama za početak i kraj strukture), kao i sa skraćenim zapisom i korišćenjem ključne reči *letter*. Ni u jednom primeru sa Sl. 33 nije korišćena oznaka *i* za iterator u *for*

petlji. Iako je uobičajeno u programskim jezicima da se kao iterator koristi *i*, to nije poželjno ni u jednom programskom jeziku [136].

```
> x <- 3
> if (x > 3) {
+   y <- "veceOd3"
+ } else {
+   y <- "manjeIliJednako3"
+ }
> y
[1] "manjeIliJednako3"
> x <- 2
> y <- if (x > 3) {
+   "veceOd3"
+ } else {
+   "manjeIliJednako3"
+ }
> y
[1] "manjeIliJednako3"
> x <- 7
> y <- if (x > 3) "veceOd3" else "manjeIliJednako3"
> y
[1] "veceOd3"
> |
```

Slika 32, Korišćenje uslovne *if/else* strukture u programskom jeziku R. Prikazana su tri načina za realizaciju uslovne strukture: (1) klasičan način, (2) dodela vrednosti strukture promenljivoj *y* i (3) skraćeni zapis strukture u jednom redu [3].

U R-u se mogu definisati kompleksni brojevi te se *i* koristi za oznaku imaginarne jedinice tj. kvadratnog korena od -1, te nije poželjno da se koristi za ime promenljive niti brojača u petlji [136]. U programskom jeziku R se za kompleksne brojeve koristi imaginarno *i*, dok se u programskim jezicima Matlab i GNU Octave za imaginarnu jedinicu podjednako koriste *j* i *i*, te se ova preporuka odnosi i na *j*. Najverovatniji razlog zašto je u programskom jeziku R samo imaginarno *i* u upotrebi to što je R nastao za statističke, odnosno matematičke proračune [3], dok su drugi programski jezici kao Matlab primarno bili namenjeni inženjerima i naučnicima¹²⁰.

120 <https://www.mathworks.com/discovery/what-is-matlab.html>, pristupljeno 05.08.2023.

```

> for (ind in 1:5) {
+   print(ind)
+ }
[1] 1
[1] 2
[1] 3
[1] 4
[1] 5
> for (ix in 1:3) print(ix)
[1] 1
[1] 2
[1] 3
> x <- c("t", "o", "b", "s")
> for (ix in seq_along(x)) print(x[ix])
[1] "t"
[1] "o"
[1] "b"
[1] "s"
> for (letter in x) print(letter)
[1] "t"
[1] "o"
[1] "b"
[1] "s"
> |

```

Slika 33, Korišćenje uslovne *for* strukture u programskom jeziku R.

Na Sl. 34 prikazano je kako se koriste kompleksni brojevi u programskom jeziku R. Posebno je naglašeno par osobina programskog jezika R koje su ilustrovane na Sl. 34:

- Kvadratni koren od -1 primenom funkcije *sqrt()* će kao rezultat dati NaN vrednosti i neće za rezultat dati imaginarnu jedinicu, osim ako -1 nije kompleksni broj.
- Da bi -1 bilo kompleksni broj, moguće je ili izvršiti konverziju primenom funkcije *as.complex()* ili odgovarajućim zapisom kompleksnog broja -1 u formi: $-1 + 0i$.
- Zapis kompleksnih brojeva se piše tako što se pored numeričke vrednosti doda imaginarna jedinica *i* (na primer, $0i$ i $1i$).
- U programskom jeziku R isključivo se koristi *i* kao imaginarna jedinica. Korišćenje simbola *j* će dati grešku (eng. *Error*) i prekinuće se izvršavanje R koda.

- Kvadriranje se u R-u može vršiti na dva načina – primenom operatora “^” i primenom operatora “**”.
- Kvadriranje imaginarne jedinice se vrši tako što se ispred *i* doda numerička vrednost 1. Činjenica da R pretražuje objekat pod nazivom *i*, ako se kvadrira njegova vrednost pozivom komande *i**2* sprečava da dođe do greške da se imaginarna jedinica pomeša sa kompleksnom vrednošću, ali svakako treba izbegavati definisanje iteratora korišćenjem oznaka za imaginarne jedinice [136]. Na Sl. 33 predložene su intuitivne oznake za iteratore *for* petlje: *ix* i *ind*.

```

> sqrt(-1)
[1] NaN
Warning message:
In sqrt(-1) : NaNs produced
> sqrt(as.complex(-1))
[1] 0+1i
> sqrt(-1 + 0i)
[1] 0+1i
> 2 + 3i
[1] 2+3i
> 2 + 3j
Error: unexpected symbol in "2 + 3j"
> i**2
Error: object 'i' not found
> 1i^2
[1] -1+0i
> |

```

Slika 34, Kompleksni brojevi u R-u. Ilustracija zašto ne treba koristiti *i* kao iterator *for* petlje i dat je pregled osnovnih operacija nad kompleksnim brojevima.

Na Sl. 35 dat je isečak iz R koda koji je podeljen zajedno sa naučnom publikacijom za delinaciju IKG signala [99-100]. Kod sa Sl. 35 predstavlja klasičan pristup automatizaciji procesa učitavanja celog skupa podataka (tzv. serijska obrada, odnosno eng. *batch processing*). U datotekama “IDN1.txt”, “IDN2.txt”, ..., “IDN20.txt” su smešteni podaci mereni na 20 ispitanika. Dobra praksa analize podataka nalaže da se metoda koja se razvija na početku testira na jednom reprezentativnom primeru signala, a da se kasnije izvrši automatizacija celog procesa za sve ispitanike. Najjednostavniji način da se to postigne jeste da se u okviru niza definišu nazivi svih datoteka, a da se kasnije kroz *for* petlju koja ima isti broj iteracija kao broj ispitanika ili u opštem slučaju merenih signala analizira svaki ponaosob. Pored niza datoteka u kojima se nalaze signali u primeru sa Sl. 35, kreiran je i niz datoteka sa .csv ekstenzijom u kojima se nalaze anotacije (lokacije lokalnih maksimuma za signale u

odgovarajućim datotekama). Takođe, primetiti da je u primeru koda sa slike komentarom¹²¹ označen segment čitanja podataka, ali i da su nazivi promenljivih opisni: *files2read* od eng. *files to*¹²² *read*, odnosno datoteke za učitavanje¹²³.

```
# reading data
files2read <- c("IDN1.txt", "IDN2.txt", "IDN3.txt", "IDN4.txt", "IDN5.txt",
               "IDN6.txt", "IDN7.txt", "IDN8.txt", "IDN9.txt", "IDN10.txt",
               "IDN11.txt", "IDN12.txt", "IDN13.txt", "IDN14.txt", "IDN15.txt",
               "IDN16.txt", "IDN17.txt", "IDN18.txt", "IDN19.txt", "IDN20.txt")
annot2read <- c("IDN1.csv", "IDN2.csv", "IDN3.csv", "IDN4.csv", "IDN5.csv",
               "IDN6.csv", "IDN7.csv", "IDN8.csv", "IDN9.csv", "IDN10.csv",
               "IDN11.csv", "IDN12.csv", "IDN13.csv", "IDN14.csv", "IDN15.csv",
               "IDN16.csv", "IDN17.csv", "IDN18.csv", "IDN19.csv", "IDN20.csv")

A30 <- array()
DE30 <- array()
A150 <- array()
DE150 <- array()
N <- array()
missed <- array()
```

```
for (ix in 1:length(annot2read)) {
```

Slika 35, Primer korišćenja *for* petlje za automatizaciju obrade signala u skripti "ICGanalysis.R". Primer je preuzet iz [99-100]. Sa desne strane skripte nalazi se svetlo siva horizontalna linija koja služi kao smernica za pisanje koda tj. za odgovarajuću i preporučunu širinu koda. O tome će biti više reči u kasnijim poglavljima, a detaljno uputstvo za rad u R Studio okruženju nalazi se u Dodatku C na kraju udžbenika.

While petlja se najčešće izbegava u analizi signala, jer može rezultovati beskonačnom petljom. Međutim, u namenskim računarskim sistemima (eng. *embedded systems*) i u električnim merenjima, beskonačne petlje mogu biti korisne. Iako se R ređe koristi za merenje signala, moguće ga je jednostavno povezati sa serijskim portom, kao što je prikazano u prethodnom potpoglavlju. Pored toga što je merenje u R-u realizovano sa *for* petljom i unapred predviđenim brojem odbiraka (pogledati Sl. 25), moglo je bez ikakve zadržke biti realizovano primenom beskonačne *while()* petlje, mada je najčešća pojava da se u R-u odredi vremenski interval u kome se želi izmeriti signal (na primer, 5 s i 10 s). U Arudino programskom okruženju, ovo merenje jeste realizovano primenom beskonačne petlje.

Na Sl. 36 su prikazana dva primera *while* strukture. Primetiti da uzastopno pokretanje ove dve strukture, bez ponovne inicijalizacije promenljive *iterator*,

121 Komentar u R-u se označava sa tarabom: "#".

122 Izgovara se na engleskom jeziku isto kao broj dva (eng. *two*), te se često kao neformalan zapis (posebno u programiranju) menja sa brojem dva u skraćenom zapisu promenljivih.

123 Slobodan prevod sa engleskog jezika na srpski jezik.

druga *while* petlja ne daje nikakav rezultat i čak ne inkrementira *iterator*. U R-u ne postoji operator za inkrementiranje, ali se može definisati posebna funkcija [3]. Operatori su, po definiciji, funkcije koje na ulazu imaju jedan ili dva argumenta i koje se mogu koristiti bez zagrada [137]. Dodatno, u R-u je moguće definisati i binarne operatore po želji (eng. *custom operator*), ali takvi elementi prevazilaze gradivo predstavljeno u ovom udžbeniku i mogu se naći u literaturi koja detaljno obrađuje programski jezik R [137].

```
> iterator <- 0
> while (iterator < 5) {
+   iterator <- iterator + 1
+   print(iterator)
+ }
[1] 1
[1] 2
[1] 3
[1] 4
[1] 5
> iterator
[1] 5
> while (iterator < 5 && iterator >=0) {
+   iterator <- iterator + 1
+   print(iterator)
+ }
> iterator
[1] 5
> |
```

Slika 36, Primeri korišćenja *while* strukture u programskom jeziku R za konsekutivno tj. uzastopno izvršavanje dve *while* petlje.

Na Sl. 36 može se primetiti da je unutar logičkog uslova *while* petlje korišćen logički operator “&&”¹²⁴. Na žalost, za razliku od Pajton programskog jezika u kome su slično prirodnom jeziku definisani ovi operatori (*and*, *or* i *not*), u R-u to nije slučaj. Logički operatori su: “&” (logičko i – poredi element¹²⁵ po element), “|” (logičko ili – poredi element po elemente), “&&” (logičko i – poredi samo prve elemente), “||” (logičko ili – poredi samo prve elemente), “!” (logička negacija – negira element po element)¹²⁶. Do sada nije spominjan, ali u

124 Logički operatori u programskom jeziku R su veoma slični logičkim operatorima u programskom jeziku Julia (<https://docs.julialang.org/en/v1/manual/mathematical-operations/>, pristupljeno 05.08.2023).

125 Ovde se pod elementima misli na elemente niza.

126 https://www.tutorialspoint.com/r/r_operators.htm, pristupljeno 05.08.2023.

slučaju da ga je potrebno koristiti za moduo se u R-u koristi operator “%%”. Prvi primer sa Sl. 36 nije praktično upotrebljiv, jer se može jednostavnije realizovati primenom *for* strukture – primena *while* strukture ima smisla kada nije unapred poznat broj iteracija. Za veći broj uslova unutar kontrolne strukture, vrši se proveravanje sa leva na desno [3]. Struktura:

repeat

se najčešće koristi kada je neophodno izvršavati operaciju dok uslov nije ispunjen i tada se poziva komanda:

break

- ova komanda je pogodna za estimaciju/procenu parametara kod metoda analize koje koriste optimizaciju. Treba napomenuti da je *repeat* petlju moguće zameniti i *while* i *for* petljama, kao i što se *break* (srp. prekid) može koristiti i sa drugim petljama. Funkcije za rad sa strukturama uključuju i:

next()

funkciju koja služi da se preskoči jedna ili više iteracija u kombinaciji sa *if/else* strukturom, ali i:

return()

funkciju koja na svom izlazu daje konačnu vrednosti i omogućava izlazak iz petlje.

1.2.1 Kontrolne funkcije

Pored kontrolnih struktura postoje i kontrolne funkcije u R-u koje se ujedno koriste za vektorizaciju drugih funkcija i kao zamena za kontrolne strukture. Najčešće se koriste funkcije:

sapply() i *lapply()*,

ali postoje i:

apply(), *tapply()*, *mapply()* i *split()*

od kojih se *split()* koristi u kombinaciji sa *lapply()* funkcijom kada za tim ima potrebe. Kontrolna funkcija *lapply()* se koristi za rad sa listama (od eng. *list apply*), a za slučaj da ulazni podataka nije lista interno se primenjuje *as.list()* funkcija [3]. Korisna je ako je potrebno da se neka operacija primeni na sve elemente pojedinačno (na primer, da se izračuna srednja vrednost svih elemenata u listi). Izlaz iz *lapply()* funkcije je, kao i ulaz, lista.

```
> x <- list(prvi = 1:10, drugi = c(NA, 2, 5, 9, 10))
> lapply(x, mean)
$prvi
[1] 5.5

$drugi
[1] NA

> lapply(x, mean, na.rm <- TRUE)
Error in mean.default(x[[i]], ...) : 'trim' must be numeric of length one
> lapply(x, mean, na.rm = TRUE)
$prvi
[1] 5.5

$drugi
[1] 6.5
```

Slika 37, Primer primene *lapply()* funkcije nad elementima liste za vektorizaciju funkcije za određivanje srednje vrednosti signala *mean()*. Parametar *na.rm* funkcije *mean()* služi da označi da li prilikom računanja srednje vrednosti treba ignorisati nedostajuće podatke tipa NA.

Argumenti funkcije, tj. operacije koja se primenjuje na svaki element liste, se prosleđuju *lapply()* funkciji kao njeni ulazni argumenti (na primer, za računanje srednje vrednosti, bez nedostajućih vrednosti, kada se prosleđuje *na.rm* argument). Prvi argument je lista na kojoj se primenjuje operacija, potom sledi naziv funkcije, pa argumenti funkcije koja se treba primeniti na sve elemente pojedinačno. Primer primene *lapply()* funkcije je prikazan na Sl. 37. Pored toga što je pokazano kako se argumenti funkcije *mean()* koja se vektorizuje, pokazano je kako se za dodelu vrednosti argumentima funkcije ne može koristiti operator dodele vrednosti “<-”, već se mora koristiti operator jednakosti “=”, kako je već ranije objašnjeno u udžbeniku.

Može se izbeći korišćenje kontrolnih funkcija i primenom kontrolnih struktura doći do istog rezultata. Međutim, ove funkcije su realizovane uz odgovarajuću optimizaciju, tako da ih je poželjno koristiti za rad u programskom jeziku R [3]. Kako je jedan od argumenata *lapply()* funkcije druga funkcija koju treba primeniti na sve elemente liste, to poziv *lapply()* funkcije može kao argument imati definisanu funkciju. Kako definisana funkcija nema ime, to se naziva anonimna funkcija. Često se anonimna funkcija koristi za operacije

indeksiranja, kao što je izdvajanje prve kolone za svaki od elemenata liste kao što je prikazano na Sl. 38.

```
> x <- list(prvi = matrix(1:10, 2, 5), drugi = matrix(1:10, 5, 2))
> x
$prvi
  [,1] [,2] [,3] [,4] [,5]
[1,]  1   3   5   7   9
[2,]  2   4   6   8  10

$druzi
  [,1] [,2]
[1,]  1   6
[2,]  2   7
[3,]  3   8
[4,]  4   9
[5,]  5  10

> lapply(x, function(pom) pom[,1])
$prvi
[1] 1 2

$druzi
[1] 1 2 3 4 5
```

Slika 38, Primer primene *lapply()* funkcije sa anonimnom funkcijom za indeksiranje (izdvajanje prve kolone) svakog od elemenata liste.

Funkcija *sapply()* radi isto što i *lapply()* sa jednom razlikom što je rezultat *sapply()* funkcije jednostavniji. Ako je rezultat lista gde je svaki element jedan broj, *sapply()* pojednostavljuje izlaz tako što daje niz brojeva kao rezultat. Za rezultat u obliku liste čiji su svi elementi iste dužine, onda *sapply()* na izlazu daje matricu. Na kraju, ako ne postoji način kako da se uprosti rezultat, onda je izlaz iz *sapply()* isti kao izlaz iz *lapply()* funkcije. Primititi na Sl. 38 da je izlaz *lapply()* funkcije uvek lista. Iako je ovo poželjna osobina *sapply()* funkcije, nekada bi trebalo izbegavati njeno korišćenje jer nije uvek pogodno da u kodu postoji funkcija za koju se ne zna koji tip podataka će dati na izlazu. U *purrr* paketu koji je dostupan na CRAN-u postoje i druge slične funkcije koje zamenjuju kontrolne strukture, gde je najčešće korišćena familija *map()* funkcija u analizi podataka [138].

Pre primene *lapply()*, u upotrebi je često *split()* funkcija, koja razdvaja objekte liste ili elemente *data frame*-a na podskupove prema određenom kriterijumu. Ova funkcija ima logički argument *drop* koji definiše da li prilikom podele na podskupove mogu da postoje prazni nivoi. U kombinaciji sa:

gl()

od eng. *generate levels* funkcijom¹²⁷ moguće je generisati nivoe kategoričkih promenljivih tj. faktora. Ovo je vrlo jednostavna funkcija i njena primena prikazana je na Sl. 39 za generisanje tri nivoa od po 5 elemenata. Ova funkcija radi vrlo slično kao *rep()* funkcija (čiji je primer korišćenja prikazan ranije na Sl. 31 – množenje matrica), ali može biti korisna jer omogućava veću čitljivost koda i brže pisanje programa, što jeste jedna od osnovnih osobina programskog jezika R.

U slučaju kompleksnije podele na podskupove (kada postoje podskupovi podskupova) može se koristiti i funkcija *interaction()* ili *list()* u kombinaciji sa *gl()* funkcijom. Na primeru podataka iz *datasets* paketa [139-140] je prikazan primer korišćenja *split()* i *sapply()* funkcija na Sl. 39. Spisak svih dostupnih skupova podataka u ovom paketu se može dobiti primenom:

```
library(help = "datasets")128
```

komande kako je prikazano na Sl. 39. Za korišćenje ovih podataka nije potrebno posebno učitavati *datasets* paket primenom *library()* funkcije. Podaci koji su korišćeni su *PlantGrowth* skupa koji sadrži dve kolone *weight* (težina suve, odnosno dehidrirane materije biljke, u g i *group* (jedna kontrolna *ctrl* od eng. *control* i dve u kojima su biljke eksperimentalno tretirane postupcima dodavanja veštačkog đubriva *trt1* i *trt2* od eng. *Treatment*)¹²⁹ [139-140].

Kontrolna funkcija koja se ređe koristi je *tapply()*, a služi da se operacije primene na pojedinačne elemente nizova ili tabela, a podešavanjem argumenta *simplify*, funkcija *tapply()* može imati isti rezultat kao *sapply()*. U radu sa nizom podataka (to može biti vrsta ili kolona matrice) primenjuje se *apply()* kontrolna funkcije najčešće u kombinaciji sa anonimnom funkcijom. Kombinacija *apply()*

127 Intuitivno se nameće zaključak da je ovakvu funkciju korisno imati pri ruci prilikom podele podataka na obučavajući (eng. *training*) i test skup za primenu mašinskog učenja. Međutim, u R-u postoje paketi sa funkcijama koje omogućavaju na jednostavnije načine da se dođe do odgovarajuće podele podataka za primenu metoda mašinskog učenja. Ova funkcija se primarno primenjuje u eksplorativnoj analizi podataka kada je potrebno vizuelno ili računanjem statističkih parametara uočiti razlike između grupa podataka. Na primer, kada se porede srednje vrednosti srčanog ritma za različite starosne grupe ispitanika korisno je primeniti *gl()* funkciju.

128 Poziv ove funkcije otvara detaljnije informacije o paketu u posebnom prozoru i može se pozivati za bilo koji paket.

129 Ovaj eksperiment prati dobru praksu randomiziranog eksperimentalnog dizajna [139-140], koja je posebno važna u kliničkom inženjerstvu i u medicini zasnovanoj na dokazima [141].

sa `mean()` ili `sum()` ima isti rezultat kao `rowSums()`, `rowMeans()`, `colSums()` i `colMeans()` [3]. Za rad sa matricama postoji i:

`mapply()`

kontrolna funkcija.

```
> gl(3,5)
 [1] 1 1 1 1 1 2 2 2 2 2 3 3 3 3 3
Levels: 1 2 3
> library(help = "datasets")
> head(PlantGrowth, 3)
  weight group
1   4.17  ctrl
2   5.58  ctrl
3   5.18  ctrl
> unique(PlantGrowth$group)
 [1] ctrl trt1 trt2
Levels: ctrl trt1 trt2
> dat <- split(PlantGrowth$weight, PlantGrowth$group)
> class(dat)
 [1] "list"
> rez <- sapply(dat, mean, na.rm = TRUE)
> rez
  ctrl  trt1  trt2
5.032 4.661 5.526
> |
```

Slika 39, Primer primene `gl()` funkcije, kao i `split()` funkcije u kombinaciji sa `sapply()` funkcijom za računanje srednje vrednosti tri grupe biljaka iz `PlantGrowth` skupa podataka dostupnih u `datasets` paketu [139-140].

Primer vektorizacije `rep()` funkcije pomoću `mapply()` je prikazan na Sl. 40. Inače, funkcija `rep()` omogućava da se prosleđen argument replicira željeni broj puta. Matrice se ne koriste dominantno u analizi podataka u programskom jeziku R, ali je primer sa Sl. 40 značajan za potrebe razumevanja vektorizacije funkcija u R-u, jer se bi inače bilo nemoguće pokrenuti funkcije za argumente koji su predstavljeni nizovima. Ne treba zaboraviti da argumenti `rep()` funkcije postaju argumenti `mapply()` funkcije, inače će R javiti grešku i prekinuti izvršavanje programa (Sl. 40).

```

> rep(2)
[1] 2
> rep(2, 4)
[1] 2 2 2 2
> rep(c(3, 5, 4), 12)
[1] 3 5 4 3 5 4 3 5 4 3 5 4 3 5 4 3 5 4 3 5 4 3 5 4 3 5 4 3 5 4 3 5 4
> rep(1:3, 3)
[1] 1 2 3 1 2 3 1 2 3
> rep(3, 1:3)
Error in rep(3, 1:3) : invalid 'times' argument
> mapply(rep(3, 1:3))
Error in rep(3, 1:3) : invalid 'times' argument
> mapply(rep, 3, 1:3)
[[1]]
[1] 3

[[2]]
[1] 3 3

[[3]]
[1] 3 3 3

```

Slika 40, Primena *mapply()* funkcije za vektorizaciju *rep()* funkcije.

1.3 Funkcije u R-u

Pretpostavlja se da je pojam funkcije poznat studentima master akademskih studija, a primer realizacije anonimne funkcije u programskom jeziku R je već prikazan na Sl. 38 u prethodnom potpoglavlju. Ovo potpoglavlje ima za cilj da ukaže na dodatne smernice za pisanje funkcija u R-u, kao i kada ih treba, a kada ne treba pisati.

Funkcije se pišu kada postoji kod koji se ponavlja, kada ima smisla podeliti programski kod na manje celine radi efikasnije organizacije (na primer, moguće je podeliti programski kod na sledeće celine: učitavanje podataka, eliminacija šuma, vizuelizacija, izdvajanje obeležja i statistička analiza) i zbog interfejsa tj. apstrakcije postojećeg koda. Funkcije u R-u predstavljaju objekte prve klase tj. klase funkcije i mogu biti prosleđene drugim funkcijama (kao argumenti, ali češće kao operacija koja se izvodi nad argumentima). Funkcije mogu biti ugnježdene. Da bi se proverila klasa funkcija treba iskoristiti:

```
class()
```

kao *class(mean)*, *class(sum)* ili sl. Rezultat će biti da je klasa *mean()* i *sum()* funkcija "function". [3]

Kao i za strukture, za označavanje početka i kraja funkcije, koriste se vitičaste zagrade “{}”. U R-u nije uobičajeno čuvanje funkcija u posebnoj datoteci sa .R ekstenzijom¹³⁰, ali je obavezno učitavanje funkcija (pokretanjem tela funkcije) kako bi se ona izlistala u radnom direktorijumu i bila dostupna za upotrebu.

Pozivom prazne funkcije u konzoli dobiće se NULL vrednost. Vrednost koju vraća funkcija je jednaka promenljivoj koja poslednja preuzima vrednost u telu funkcije. Da bi se izbegla nedvosmislenost, može se koristiti komanda:

```
return()
```

unutar funkcije u kojoj se definiše izlaz funkcije. Funkcije ne moraju nužno imati ni ulazne ni izlazne argumente¹³¹. Ako su uključeni argumenti u definiciji funkcije, onda se oni nazivaju formalnim argumentima, te se ne mogu izostaviti prilikom poziva funkcije. Ako se izostave formalni argumenti, javiće se greška. Svi formalni argumenti jedne funkcije se mogu izlistati pomoću:

```
formals()
```

komande. [3]

Primeri funkcija su dati na Sl. 41. Pokazano je kako je moguće definisati praznu funkciju $f1()$ koja na svom izlazu daje NULL, potom kako je moguće definisati funkciju $f2()$ bez argumenata i kako se definiše i koristi funkcija $f3()$ sa formalnim argumentom. Primiti za funkciju $f3()$ na Sl. 41:

- Nije moguće realizovati poziv funkcije koja nije definisana (učitana u radno okruženje (eng. *Environment*)).
- Nije moguće pozvati (iskoristiti funkciju $f3()$) pre nego što je učitana u radno okruženje.
- R će javiti grešku prilikom poziva funkcije $f3()$ bez definisanja ulaznog argumenta ar .
- Iako R javlja grešku za poziv funkcije bez formalnog argumenta (“*Error in print(ar)*”), izvršava se prethodna komanda u telu funkcije.

130 Nije ni zabranjeno. To je stvar izbora.

131 Radi jezičkog bogatstva, u udžbeniku je korišćen i termin parametar da se označe ulazni argumenti funkcija u R-u.


```

> f1 <- function() {}
> f2 <- function() {
+   print("Primer funkcije.")
+ }
> f1()
NULL
> f2()
[1] "Primer funkcije."
> f3()
Error in f3() : could not find function "f3"
> f3 <- function(ar) {
+   print("Primer funkcije sa jednim argumentom.")
+   print(ar)
+ }
> f3()
[1] "Primer funkcije sa jednim argumentom."
Error in print(ar) : argument "ar" is missing, with no default
> f3(ar = 37)
[1] "Primer funkcije sa jednim argumentom."
[1] 37
> |

```

Slika 41, Primeri funkcija u R-u.

Definisanje funkcije $f4()$ sa podrazumevanim argumentom je dato na Sl. 42. Funkcija:

$formals(f4)$

kao rezultat prikazuje da postoji formalni argument ar čija je podrazumevana vrednost 5 (Sl. 42). Za funkcije koje nemaju formalne argumente, rezultat poziva $formals()$ komande bi bio NULL, a za funkcije koje nemaju podrazumevane vrednosti argumenata $formals()$ komanda bi dala samo naziv argumenta.

Definisanje funkcija u programskom jeziku R je vrlo intuitivno i jednostavno, kako je pokazano na Sl. 41 i Sl. 42. Međutim, jedna razlika programskog jezika R u odnosu na druge programske jezike koji su često zastupljeni u obradi signala (kao što su Matlab i GNU Octave) jeste što su sve funkcije, kao i promenljive izlistane u okruženju u kom su definisane. Na Sl. 43 prikazano je kako su funkcije $f1()$, $f2()$, $f3()$ i $f4()$ koje su definisane na Sl. 41 i Sl. 42, smeštene u globalnom R okruženju (eng. *Global environment*). Primititi da se pored naziva svake od funkcija u kartici Environment R Studio okruženja,

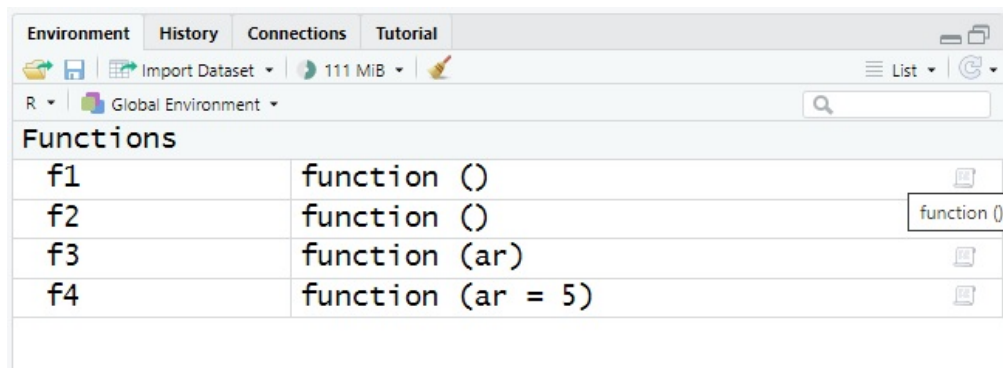
nalazi ikonica. Prelaskom strelice računarskog miša preko te ikonice se ispisuje “function()” (Sl. 43), a klikom na ikonicu se pokreće komanda:

`View(f2)`

koja otvara novu karticu pod nazivom Function u R Studio okruženju i prikazuje telo funkcije.

```
> f4 <- function(ar = 5) {  
+   print("Primer funkcije sa podrazumevanim argumentom.")  
+   print(ar)  
+ }  
> f4()  
[1] "Primer funkcije sa podrazumevanim argumentom."  
[1] 5  
> formals(f4)  
$ar  
[1] 5  
  
> |
```

Slika 42, Primer funkcije sa podrazumevanim argumentom i korišćenje komande `formals()` u R-u.



Slika 43, Prikaz globalnog okruženja u R-u nakon definisanja funkcija `f1()`, `f2()`, `f3()` i `f4()` na Sl. 41 i Sl. 42.

Slaganje/podudaranje argumenata (eng. *argument matching*) za funkcije predstavlja jednu od osnovnih odlika programskog jezika R i vrlo je intuitivno rešeno. U R-u postoje dva načina za slaganje argumenata: (1) prema poziciji i (2) prema imenu argumenta. Preporuka je da se koristi podudaranje argumenata po imenu, a ne po poziciji, jer pojedini argumenti mogu imati podrazumevane vrednosti (eng. *default values*), pa nema puno smisla da se podrazumevani

argumenti unose u R kod. Primeri podudaranja argumenata su prikazani u poglavlju koje se odnosi na generisanje veštačkih biosignala, a ovde o njima neće biti više reči. [3]

Važna osobina programskog jezika R je tzv. lenjost (eng. *lazy evaluation*) prilikom izvršavanja operacija za poziv neke funkcije [3]. Ova osobina se može primetiti kada u telu funkcije postoji neka greška (na primer, nije prosleđen odgovarajući argument), jer tada će se R funkcija izvršiti do trenutka kada se dođe do greške (što je ilustrovano na Sl. 41 za poziv $f3()$ funkcije bez formalnog argumenta *ar*). Ova osobina može biti veoma korisna za debugovanje¹³² koda koji se nalazi u telu funkcije. Nije preporučljivo da se kod koji nije prethodno proveren i testiran stavlja u telo funkcije, jer se lokalnim argumentima (argumenti koji su definisani u telu funkcije) ne može pristupiti iz globalnog okruženja.

```
> ?rnorm
> noviRnorm <- function(mean = 2.5, ...) {
+   rnorm(mean = 2.5, ...)
+ }
> x <- noviRnorm(300, sd = 1, mean = 2.5)
> mean(x)
[1] 2.40854
> length(x)
[1] 300
> |
```

Slika 44, Primer definisanja nove *rnorm()* funkcije sa argumentom “...” u R-u.

Specifičnost programskog jezika R se ogleda u argumentu “...” koji se koristi kada se svi argumenti postojeće funkcije prosleđuju novoj funkciji – u tom slučaju “...” označava sve postojeće argumente funkcije koji se prosleđuju [3]. Ovakva prosleđivanja su uobičajena ako se želi definisati nova funkcija sa promenjenim parametrima (na primer, *rnorm()* funkcija u kojoj će srednja vrednost biti podrazumevano jednaka vrednosti 2.5). Primer definisanja nove funkcije *noviRnorm()* je dat na Sl. 44, a stranica sa opisom *rnorm()* funkcije se dobija korišćenjem operatora “?”. *rnorm()* funkcija generiše pseudoslučajne brojeve (na Sl. 44 to je 300 pseudoslučajnih brojeva) koji prate Gausovu raspodelu i pripada stats paketu koji se ne mora učitavati u R-u korišćenjem funkcije *library()* [142].

132 Otklanjanje grešaka odnosno bagova u kodu.

Sa druge strane, ovakav argument je koristan za funkcije kod kojih se unapred (pre samog poziva) ne zna tačan broj argumenata. Primeri su *paste()* i *cut()* funkcije koje služe za manipulaciju stringovima. U slučaju da se koristi argument "...", treba obratiti pažnju na podudaranje, jer se posle ovog argumenta funkcije ne mogu pozivati sa delimičnim, već isključivo sa potpunim podudaranjem po imenu argumenta (što se vidi u pozivu *noviRnorm()* funkcije na Sl. 44). *mean()* funkcija na Sl. 44 je iskorišćena da se proveri da li srednja vrednosti generisanih pseudoslučajnih brojeva odgovara unetom argumentu, a *length()* funkcija služi za proveru dužine niza, odnosno da li je generisano zadatih 300 pseudoslučajnih brojeva. Kako se radi o pseudoslučajnim brojevima, srednja vrednost ne može biti jednaka tačno 2.5, već postoji neko odstupanje. Po Centralnoj graničnoj teoremi za dovoljno veliki broj merenja, svaka raspodela teži Gausovoj raspodeli [78]. Prema tome, za što veći broj generisanih psudoslučajnih brojeva srednja vrednost će biti bliža zadatoj i obrnuto.

```
> mean
function (x, ...)
UseMethod("mean")
<bytecode: 0x000001c6fd605398>
<environment: namespace:base>
> mean <- function(x) x**2
> mean
function(x) x**2
> mean(3)
[1] 9
> search()
 [1] ".GlobalEnv"          "tools:rstudio"
 [3] "package:stats"       "package:graphics"
 [5] "package:grDevices"  "package:utils"
 [7] "package:datasets"   "package:methods"
 [9] "AutoLoads"          "package:base"
> mean <- "tobs"
> mean
[1] "tobs"
> |
```

Slika 45, Primer definisanja funkcije i promenljive sa istim imenom kao postojeća funkcija u R-u. Prikazan je i redosled pretrage primenom *search()* komande.

Kada se definiše nova funkcija koja ima isto ime kao postojeća funkcija, na primer, nova *filter()* funkcija, postavlja se pitanje kako R “zna” koja je funkcija

pozvana. U tom slučaju, primenjuje se procedura koja uključuje sledeće korake [3]:

- proverava se globalno okruženje u kome se traži objekat pod istim imenom kao funkcija (može se pogledati redosled pretrage promenljive nakon poziva *search()* komande kao na Sl. 45),
- proveravaju se funkcije u svim paketima koji se pretražuju po abecednom redosledu i na kraju
- se pretražuje unutar osnovnog (eng. *base*) paketa.

U globalnom okruženju može postojati samo jedna instanca sa jednim imenom (što je pokazano na primeru *mean()* funkcije i *mean* promenljive, u trenutku kada se definiše promenljiva funkcija se ne vidi kako je prikazano na Sl. 45). Redosled po kom se paketi pretražuju moguće je menjati na jednostavan način, a to je primenom odgovarajućeg redosleda za aktivaciju paketa u R kodu primenom *library()* funkcije ili primenom dvotačka operatora “::” kako je već navedeno u prethodnim potpoglavljima. Praksu definisanja funkcija i promenljivih sa istim imenima, osim ako se ne pravi novi paket, treba izbegavati.

1.3.1 Leksička pravila pretrage argumenata

Svaka funkcija ima tri dela [3]:

- formalne parametre (argumente koje korisnik definiše prilikom poziva funkcije),
- telo funkcije koje predstavlja sekvencu komandi i izraza koji se izvršavaju prilikom poziva funkcije i
- okruženje u kome “živi” funkcija tj. okruženje u kome je funkcija definisana.

Okruženje predstavlja kolekciju simbola i njihovih vrednosti. Svako okruženje ima svoje matično okruženje i može imati više pomoćnih okruženja (eng. *child environment*). Jedino okruženje koje nema svoje matično okruženje je prazno okruženje. Uobičajeno, funkcije i promenljive se definišu u globalnom okruženju (čiji je sadržaj moguće pregledati u R Studio razvojnom okruženju kao na Sl. 43). U slučaju da se definiše funkcija unutar funkcije, onda je

okruženje ugnježdene funkcije¹³³ druga funkcija, a ne globalno okruženje. Pomoću komande:

`ls()`

može se proveriti koji objekti su definisani u okruženjima (za pokretanje koda sa Sl. 45, to bi trebalo da bude samo promenljiva *mean*). Najjednostavniji način prikaza u konzoli kom okruženju pripada funkcija je da se otkuca ime funkcije u konzoli (bez zagrada), jer sve funkcije imaju i pokazivače ka okruženju u kome su definisane. Ovo je pokazano za funkciju *mean()* koja računa srednju vrednosti niza, a koja je definisana u base okruženju (Sl. 45).

Svaka funkcija može imati tri vrste promenljivih [3]:

- formalne promenljive (prosleđuju se kao argumenti funkcijama, mogu imati podrazumevane vrednosti, primeri su dati na Sl. 41 i Sl. 42),
- lokalne promenljive (definisane su unutar funkcije) i
- slobodne promenljive na koje se odnose pravila leksičke pretrage da bi se utvrdilo u kom okruženju su promenljive definisane.

R koristi leksička, odnosno statička pravila pretrage (eng. *Lexical/Static Scoping*), za pretragu slobodnih promenljivih [3]. Funkcija, pored formalnih argumenata, može imati i slobodne promenljive (one koje su definisane unutar funkcije ili se koriste unutar funkcije bez prethodne definicije). U slučaju da se promenljiva koristi unutar funkcije, a nije jedan od argumenata i nije definisana unutar funkcije, mora se primeniti leksičko pravilo pretrage argumenata da bi se došlo do odluke koju vrednost dodeliti takvoj slobodnoj promenljivoj. Prema leksičkim pravilima pretrage, najpre se pretražuje okruženje u kome je funkcija definisana, a potom se prelazi u matično (eng. *parent*) okruženje. Ako se ni tamo ne nađe promenljiva, pretraga se nastavlja do globalnom okruženja tj. okruženja najvišeg nivoa (eng. *top-level environment*). Nakon toga, pretraga se nastavlja do okruženja najnižeg nivoa (eng. *down-level*) tj. do praznog (eng. *empty*) okruženja. U slučaju da se ne nađe promenljiva, R će javiti da je došlo do greške. Generalno, treba izbegavati slobodne promenljive, osim u posebnim slučajevima kreiranja optimizacionih procedura jer optimizacione funkcije za svoje argumente mogu imati druge funkcije koje se koriste za pretragu maksimuma ili minimuma, što je detaljno prikazano u [3]. Primer primene leksičke pretrage date je na Sl. 46. Za leksičku pretragu unutar funkcije

¹³³ Podsećanja radi, u programskom jeziku C nije moguće definisati funkciju unutar funkcije, te ove odlike programskog jezika jesu od značaja, posebno za one koji već imaju iskustvo korišćenja drugih programskih jezika.

funkcija2() argument *arg1* preuzima vrednost 2 (iz okruženja gde je funkcija definisana, a ne iz okruženja u kome je funkcija pozvana), te je prema tome za poziv funkcije *funkcija1(10)* rezultat jednak $5 \times 10 \times 2 = 100$, a za dinamičku pretragu slobodna promenljiva bi bila preuzeta iz okruženja iz koga je funkcija pozvana te bi rezultat bio: $5 \times 10 \times 5 = 250$.

Leksičko pretraživanje nije karakteristično samo za R, već i za druge programske jezike (na primer, Perl, Pajton i Lisp) [3]. Upravo je posledica leksičkog okruženja da su svi objekti smešteni u memoriji računara. Više informacija o leksičkom pretraživanju sa primerima u R-u i razlozima zašto je u R-u odabrano leksičko, a ne dinamičko pretraživanje može se naći u [4, 143]¹³⁴.

```
> x <- 2
> funkcija1 <- function(arg1) {
+   x <- 5
+   x * funkcija2(arg1)
+ }
> funkcija2 <- function(arg1) {
+   arg1 * x
+ }
> funkcija1(10)
[1] 100
> |
```

Slika 46, Primer primene leksičke pretrage za argumentom *x* unutar funkcije *funkcija2()* koji nije definisan u telu funkcije (lokalno).

¹³⁴ Odlično poređenje leksičke sa dinamičkom pretragom dato je i u enciklopediji Vikipedija, [https://en.wikipedia.org/wiki/Scope_\(computer_science\)#Lexical_scope_vs._dynamic_scope](https://en.wikipedia.org/wiki/Scope_(computer_science)#Lexical_scope_vs._dynamic_scope), pristupljeno 07.08.2023.

2 Automatske metode za analizu biosignala

Signal je funkcija koja prenosi informaciju o ponašanju nekog sistema ili o osobinama nekog fenomena, a pored prirodnih signala, postoje i veštački, odnosno generisani signali [144]¹³⁵. Dodatno, signal može biti funkcija jedne ili više promenljivih koje sadrže korisnu informaciju [145]. Za signal se kaže da je biosignal ili biološki signal ili biomedicinski signal¹³⁶, ako je snimljen na živom sistemu i sadrži informaciju o stanju i ponašanju tog sistema [145]. Osim jednodimenzionih signala, koji su najčešće snimljeni kao vremenske serije, postoje i višedimenzioni signali¹³⁷ [145]. EMG signal meren matričnim elektrodama sadrži, pored amplitude EMG signala, koja je promenljiva po jednoj dimenziji (vremenu) i prostornu dimenziju određenu lokacijom EMG elektroda iznad regije od interesa, koja može pokrivati jedan ili više mišića. Primer višedimenzionog EMG signala merenog na mišiću *vastus lateralis* (lat.) za dva subjekta pod različitim uslovima merenja (različiti intenzitet aktivacije mišića u odnosu na maksimalnu kontrakciju) je prikazan na Sl. 47 [146].

U biomedicinskom inženjerstvu se najčešće analiziraju elektrofiziološki, odnosno električni biosignali, od kojih su neki već ranije spomenuti u udžbeniku, a primeri elektrofizioloških signala su [145]:

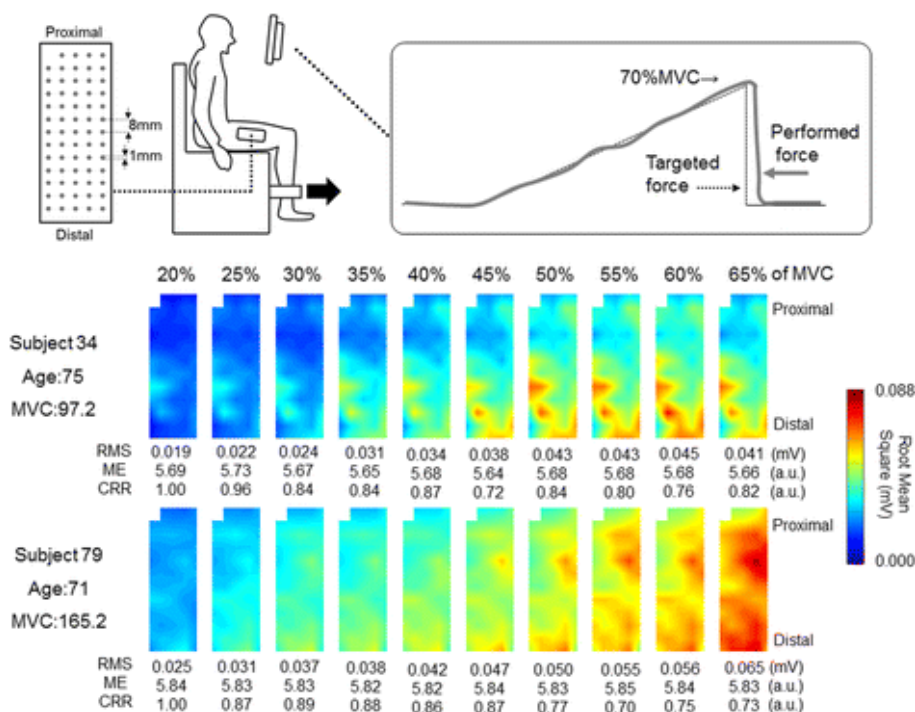
- elektroencefalogram (EEG, eng. *Electroencephalogram*) – električni signal mereni sa skalpa glave koji potiču iz mozga,
- elektrokardiogram (EKG, eng. *Electrocardiogram* - ECG) – električni signal srca,
- elektromiogram (EMG, eng. *Electromyogram*) – električni signal mišića,

135 Definicija na engleskom jeziku: “A signal is a function that conveys information about the behavior of a system or attributes of some phenomenon. Signals occur naturally and they are also synthesized.” [144]

136 U ovom udžbeniku se koristi dominantno termin biosignal (eng. *biosignal*), ali se biološki signal (eng. *biological signal*) i biomedicinski signal (eng. *biomedical signal*) mogu koristiti i koriste se u literaturi kao sinonimi.

137 Definicija na engleskom jeziku: “A signal is a function of one or several variables that carries useful information. A signal is said to be biological if it is recorded from a living system, and conveys information about the state or behavior of that system ... Signals can be either one-dimensional, if they depend on a single variable such as time, or multidimensional if they depend on several variables such as spatial coordinates.”. [145]

- elektrookulogram (EOG, eng. *Electrooculogram*) – predstavlja razliku između električnih potencijala na prednjem i zadnjem delu oka i u korelaciji je sa pokretima oka,
- elektrodermalna aktivnost (EDA, eng. *Electrodermal Activity*) ili galvanski refleks kože (GSR, eng. *Galvanic Skin Response*) – električni signal provodnosti kože (bioimpedansa),
- elektrogastrogram (EGG, eng. *Electrogastrogram*) – električni signal glatkih mišića želuca i
- elektrohistogram (EHG, eng. *Electrohysterogram*) – električni signal materice.

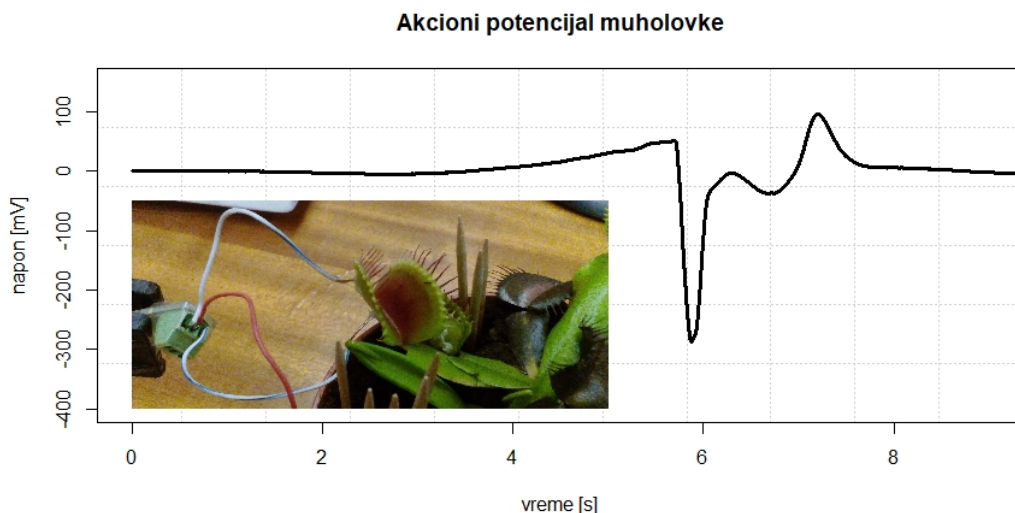


Slika 47, Primer višedimenzionog EMG biosignala za dva ispitanika od 71 i 75 godina za signale merene sa mišića *vastus lateralis* (lat.) primenom matrice EMG elektrode koja je prikazana u gornjem levom uglu. EMG signali su prikazani pomoću topografskih mapa za različite intenzitete mišićnih kontrakcija u odnosu na maksimalnu voljnu kontrakciju (20%, 25%, 30%, ..., 65%). Slika je preuzeta iz [146] i podeljena pod CC BY 4.0 licencom (<https://creativecommons.org/licenses/by/4.0/>, pristupljeno 07.08.2023).

Nazivi elektrofizioloških signala se završavaju na “gram” kao, na primer, elektromiogram. Naziv uređaja se završava sa “graf” (elektromiograf), a naziv metode na “grafija” (elektromiografija). Skraćenica EMG označava istovremeno elektromiogram, elektromiograf i elektromiografiju, te je pravilno pisati za električne signale mišića EMG, ali i EMG signal. Isto se odnosi i na

ostale elektrofiziološke signale sa završetkom na “gram”, a isti princip se primenjuje i na srpskom i na engleskom jeziku. [147]

Medicinski signal je uži pojam od biosignala, jer se medicinski signali mere u medicini isključivo na ljudima, dok biosignali mogu biti mereni na bilo kom živom organizmu. Biosignali se mogu meriti na biljkama. Na primer, biljka mesožderka (eng. *venus fly trap*, lat. *dionaea muscipula*) ima mogućnost generisanja akcionog potencijala koji je moguće izmeriti [148-149]. Akcioni potencijal biljke mesožderke je meren i na Elektrotehničkom fakultetu, Univerziteta u Beogradu, a grafik koji je zajedno sa fotografijom prikazan u programskom jeziku R, dat je na Sl. 48. Jedan od prvih EMG signala je izmeren na slonu [150 u radu koji je objavljen 1921. godine, što ima smisla obzirom na opremu koja je bila na raspolaganju s početka 20. veka i na dimenzije mišića kod slona.



Slika 48, Akcioni potencijal¹³⁸ meren na biljci mesožderki na Elektrotehničkom fakultetu, Univerziteta u Beogradu. R kod, fotografija biljke i podaci su dostupni u okviru GitHub repozitorijuma <https://github.com/NadicaSm/TOBS>. Ovaj primer je prikazan u prezentaciji iz 2017. godine o TOBS predmetu¹³⁹. Kod za generisanje slike je prikazan u narednim poglavljima, a ovde je važno skrenuti pažnju da je slika sa fotografijom pripremljena u R-u i da je R odlično okruženje za kombinovane, odnosno kompozitne ilustracije, koje mogu da uključuju grafike i slike, po ugledu na Sl. 47 [146].

138 Na slici se vidi jedan akcioni potencijal.

139 Prezentacija je dostupna na internet stranici TOBS predmeta, https://automatika.etf.bg.ac.rs/images/FAJLOVI_srpski/predmeti/izborni_kursevi_os/biomedicinsko_inzenjerstvo/TOBS/TOBS_00_Umesto_uvoda.pdf, pristupljeno 14.08.2023.

Pojedine klasifikacije biosignala u biomedicinskom inženjerstvu, među biosignale ne ubrajaju rezultat merenja opsega pokreta [151], ali ako se uzme u obzir da svaki signal koji se odnosi na aktivnost živog organizma predstavlja biosignal, onda su opseg pokreta meren elektrogoniometrom (senzor za merenje ugla koji se zasniva na primeni mernih traka [78]), pokreti oka mereni uređajima za praćenje pokreta oka (eng. *eye tracker*), broj koraka meren pedometrom koji se zasniva na primeni akcelerometara, primeri biosignala. Signali mereni istim sensorima (na primer, akcelerometrima) mogu biti u direktnoj vezi sa elektrofiziološkim signalima, te se signal mehaničke kontrakcije srca – balistokardiogram (eng. *Ballistocardiography*, BCG ili BKG) ubraja u biosignale [152]. Iako se izučava kao posebna oblast, te neće biti predmet u ovom udžbeniku, medicinska slika jeste vrsta biosignala [145]. Takođe, govorni signal jeste biosignal.

Bez obzira da li su predstavljeni kao vremenske serije ili ne, biosignali nisu uvek dovoljni da se dobije kompletna informacija o zdravlju osobe, bilo u dijagnostici ili prilikom ocene terapijskih modaliteta. Kompletan skup podataka može da uključi demografske podatke, kliničke parametre i rezultate laboratorijske analize [134]. U literaturi je uobičajen naziv biomedicinski podaci¹⁴⁰ za sveobuhvatne kompleksne skupove. Pored ovog termina, u literaturi se može sresti i termin kompleksni biomedicinski podaci koji pokriva dva tipa biomedicinskih podataka. Prvo značenje je što se kompleksni biomedicinski podaci koriste za opisivanje nelinearnih i nestacionarnih pojava koje se javljaju u vremenskim serijama biosignala, a koje često mogu da sadrže tzv. skrivenu informaciju¹⁴¹ (eng. *hidden information*) [112, 153]. Drugo značenje se odnosi na podatke relativno velikih dimenzija (eng. *Big Data, Large Datasets*) [131, 154]. Primer analize skrivene informacije je prikazan u istraživanju naučnika sa MIT-ja¹⁴² (eng. *Massachusetts Institute of Technology*) koji su analizom podataka sa odeljenja intenzivne nege u periodu od 10 godina došli do zaključka da nijedan pacijent sa uznapredovalom bolešću jetre kome je bila potrebna dijaliza nije preživeo otpust [134]. Raniji pogled na ovu bolest je bio da se komplikacija otkazivanja bubrega kod takvih pacijenata može lečiti dijalizom, ali se ispostavilo da je to za pacijente sa bolešću jetre kritična i nepremostiva tačka, što je važan podatak, posebno u određivanju dostupnih

140 Striktno govoreći i signali su podaci, ali ustaljeno je da se vremenske serije nazivaju signalima, a da se skupovi podataka koji sadrže druge informacije (na primer: demografski podaci, laboratorijske analize [134]) nazivaju podacima.

141 Ovde se ne radi o kodovanim informacijama, već o informacijama koje se vide standardnim i jednostavnim metodama vizuelizacije i analize na signalu, te zahtevaju realizaciju specifičnih algoritama [112].

142 <https://www.mit.edu/>, pristupljeno 08.08.2023.

terapija, ali i dogovoru potencijalne palijativne nege sa pacijentima i njihovim porodicama [134].

Biosignali su u odnosu na ostale signale specifični po tome što oslikavaju rad kompleksnog sistema tj. živog organizma/organa/tkiva, a obzirom da se najčešće mere primenom neinvazivnih mernih metoda biosignali imaju relativno male amplitude (reda mV ili μV), te su bez obzira na primenu biomedicinskih pojačavača podložni šumu [151]. Pre primene odgovarajuće metode za eliminaciju šuma, važno je znati koji element kompleksnog sistema se ispituje i šta je cilj analize. Na primer, ako je cilj analize da se detektuju RR intervali (rastojanja između susednih R pikova) na EKG signalu, onda se filter propusnik opsega (eng. *band pass filter*) može primeniti tako da se sačuva manji frekvencijski opseg, nego u analizi koja ima za cilj da se detektuju svi relevantni intervali na EKG signalu, odnosno PQRST segment. Prema tome, iz ugla analize signala koja ima za cilj ocenu RR intervala, P, Q, S i T talasi predstavljaju šum. Ovo je jedan od osnovnih principa koji se primenjuje u Pan-Tompkins metodi za određivanje srčanog ritma i detekciju R pikova, odnosno QRS segmenata [20]. Sledi da ono što je šum u jednoj analizi, može biti korisna informacija u drugoj analizi, te svaki slučaj analize biosignala treba posmatrati pojedinačno.

Klasična teorijska podela iz digitalne obrade signala (eng. *digital signal processing*) deli signale na determinističke i stohastičke [155-156]. Dok se deterministički signali mogu predstaviti egzaktnim matematičkim modelom, kod stohastičkih signala postoji stepen slučajnosti (eng. *random signal*) [155]. Međutim, u prirodi ne postoji jasna granica između stohastičkih i determinističkih signala. Biosignali se generalno mogu klasifikovati kao pretežno stohastički, jer opisuju kompleksne biološke sisteme na koje utiče niz unutrašnjih i spoljašnjih faktora koje nije jednostavno opisati matematičkim modelom. Detaljan upliv u stohastičke sisteme prevazilazi gradivo pokriveno ovim udžbenikom i ovde se podrazumeva da su studenti već stekli osnovna znanja iz stohastičkih sistema i estimacije na osnovnim akademskim studijama. U radu sa biosignalima, nije jednostavno povući granicu između determinističkih i stohastičkih biosignala. Primera radi, EKG signal se vrlo često uzima kao primer determinističkog signala, jer je složenoperiodičan. Na prvi pogled jeste. Ali, varijabilnost srčanog ritma koja se određuje iz rastojanja QRS kompleksa ili R pikova na susednim PQRST kompleksima ima izrazito stohastičku prirodu, posebno kod zdravih ispitanika [157]. Pored stohastičke prirode, dinamika varijabilnosti srčanog ritma ima i osobine haosa [158]. Takođe, PQRST kompleksi i klinički parametri koji odgovaraju EKG signalu (kao što su na primer: PR interval, PR segment, ST interval, and ST segment

[159-160]) menjaju se u vremenu kao rezultat fizioloških procesa u organizmu, te se o eksplicitnom determinizmu ne može govoriti. Drugi primer stohastičkog signala je EEG. Međutim, kada se pogledaju metode koje se koriste za analizu EEG signala, primećuje se da se one zasnivaju na ispitivanju sinusoidalnih ritmova (tzv. alfa, beta, gama i delta ritmovi) u kratkim vremenskim intervalima, odnosno na intervalima na kojima je signal stacionaran¹⁴³ [161]. Kada bi se posmatrao vremenski interval u kome je prisutan, na primer, samo alfa ritam, teško da bi se takav signal mogao klasifikovati kao stohastički. Takvi intervali su obično trajanja oko 0.25 s i nazivaju se kvazi-stacionarni intervali ili tobože stacionarni [161]. Slično se može zaključiti i za druge signale. Vrlo oprezno treba pristupiti teorijskoj podeli na stohastičke i determinističke signale, jer realni signali mogu imati osobine koje se ne mogu ignorisati niti uklopiti u jasno razgraničene kategorije. Generalno, za sve biosignale važi da imaju tzv. 3N prirodu, odnosno da su po prirodi nestacionarni, nelinearni i da imaju šuma (od eng. *Nonstationary, Nonlinear, Noisy*) [161]. Kako je ovde pokazano, u zavisnosti od situacije i cilja analize signala, biosignal može imati prirodu šuma (3N), ali i ne mora.

Analiza determinističkih i stohastičkih biosignala, u opštem slučaju, usvojen je termin koji se koristi kao sinonim za obradu signala, a uključuje i analizu i sintezu. To se jednostavno može objasniti na primeru Furijeove transformacije (odgovara analizi, jer razdvaja signal na elementarne komponente – sinusoide tj. prostoperiodične komponente) i inverzne Furijeove transformacije (odgovara sintezi, jer korišćenjem elementarnih elemenata kao što su sinusoide iznova generiše, odnosno sintetiše biosignal) [162]. Slično važi i za transformaciju talasića (eng. *wavelet transformation*). Prema tome, obrada signala obuhvata i analizu i sintezu signala, ali je ustaljen termin i sinonim za obradu signala izraz analiza signala.

Često se sreće i termin analitika (eng. *data analytics*) koji se smatra sinonimom za analizu podataka (eng. *data analysis*), ali ovi termini označavaju donekle različite metode. Naime, analitika je širi pojam od analize, te uključuje analizu podataka, ali i druge metode da bi se donele odgovarajuće odluke, te se često termin analitika koristi van naučno-istraživačke prakse [163], iako su metode koje se primenjuju iste onima koje se sreću u nauci kao što su pronalaženje zavisnosti, predikcija/predviđanje stanja i donošenje odluka. Jedan od najvažnijih elemenata koji se ističe u analitici je izveštaj koji je razumljiv donosiocima odluka koji ne moraju nužno posedovati odgovarajuća znanja o podacima i primenjenim metodama analize, ali čije odluke bi trebalo da se

143 Stacionaran signal je signal čije se statističke osobine ne menjaju u vremenu [161].

zasnivaju na rezultatima analitičara. Iako je termin analitičar uobičajen van akademije, sa akademskog stanovišta jedina razlika između onih koji se bave analizom podataka i analitikom, jeste što analitika može da uključi prikupljanje podataka. Termini se verovatno ne smatraju sinonimima zbog poslova koji se razlikuju, jer se u industriji od analitičara očekuje finalni izveštaj sa preporukama kako da se poslovi unaprede, optimizuju ili promene na osnovu podataka i analize podataka, dok se od onih koji analiziraju podatke očekuje da primene specifične metode obrade podataka, što može, ali ne mora nužno voditi direktnoj praktičnoj primeni. U ovom udžbeniku se neće koristiti termin analitika, ali bi ovo objašnjenje razlika i sličnosti trebalo studentima da približi buduće poslove za koje im znanje o analizi podataka može biti korisno i ključno. I analiza i analitika pripadaju oblasti nauke o podacima¹⁴⁴. Primer primene analitike u medicini i zdravstvu je obrada podataka koja uključuju biosignale, protok informacija na internet stranicama bolnica i drugih medicinskih ustanova, analizu administrativnih podataka (kao što su zakazivanje termina i plaćanje računa) sa ciljem donošenja konkretnih preporuka rukovodiocima¹⁴⁵ ustanova za unapređenje poslovanja i efikasniju negu i zdravstvenu zaštitu [164].

2.1 Uticaj subjektivnih faktora na automatsku analizu biosignala

Iako ovaj udžbenik sadrži delove koji su posvećeni programiranju u programskom jeziku R i radu u R Studio okruženju, uputstva nisu ista kao u standardnim knjigama o programiranju, jer je udžbenik obogaćen odabranim primerima obrade biosignala. Između ostalog, kod je predstavljen na slikama radi jednostavnijeg referenciranja, što nije uobičajeno za udžbenike iz programiranja. Dodatno, kako su različiti aspekti razmatrani o automatskoj

144 Iako termin nauke o podacima sadrži reč “nauka”, ova oblast nije bukvalno i isključivo naučna, već predstavlja oblast izučavanja podataka primenom, između ostalih, i naučnih metoda.

145 U srpskom jeziku su se previše odomaćili termini preuzeti iz engleskog jezika i na žalost osim javnih kritika rodno-ravnopravnog jezika, kritika na račun termina iz engleskog jezika gotovo da nema ili su potpuno izostale u Republici Srbiji. Vrlo često se može čuti da je neko menadžer (od eng. *manager*) umesto rukovodilac, ejčar (od eng. *Human Resource*, skraćeno HR) umesto službenik za ljudske resurse, tim lider (od eng. *team leader*) umesto rukovodilac sektora, da se radi o kompjuterskoj (od eng. *computer*) umesto o računarskoj analizi signala, da je neko senior (od eng. *senior*) ili junior (od istoimene engleske reči) umesto stariji i mlađi saradnik i sl. Ovaj udžbenik je pisan u duhu srpskog jezika sa namerom da se očuva srpski jezik u biomedicinskom inženjerstvu, ali i u inženjerstvu i računarstvu generalno, a posebno imajući u vidu da je tehnički jezik danas engleski jezik te da se uticaj engleskog jezika na srpski jezik ne može i neće u potpunosti izbeći.

obradi biosignala, to se ovde treba osvrnuti na par činjenica koje bi o analizi biosignala jedan biomedicinski inženjer trebalo da zna.

Analizu signala sprovode ljudi, te nije od presudnog značaja samo kako se sprovodi analiza na računaru, već i kako određene elemente analize (na primer: aritmetičke operacije) i rezultate analize procesira ljudski mozak. Poznato je da istraživači koji mere signale često imaju pristrasnost u analizi signala i tendenciju da uoče elemente i osobine koje signal ne poseduje. Sa druge strane, oni koji isključivo vrše analizu signala su u prilici da uoče sve manjkavosti mernog postupka jer poseduju neophodnu objektivnost. Ali, s obzirom da nisu prisustvovali postupku merenja njima mogu nedostajati odgovarajuća znanja o povezanosti informacija na signalu sa uslovima merenja.

U slučaju da ista osoba meri i analizira podatke, dva su načina da se izbegne pristrasnost: (1) ostavljanjem vremenske distance, odnosno puštanjem da protekne određeno vreme od merenja od obrade signala i (2) uključenjem stručnjaka koji ne učestvuju u merenju, već u analizi signala (nekada se primenjuje pristup da više osoba analizira iste signale ne bi li se utvrdilo da li su zaključci do kojih se dolazi pouzdani i ne zavise od osobe koja analizira signale). Situacija u kojoj inženjeri koji analiziraju podatke ne učestvuju u merenju, dovodi do toga da se njihovi zaključci o podacima tretiraju kao kritika timovima koji su merili podatke. Profesionalnost u komunikaciji je jedini način da se zaobiđe ovaj problem – ne postoji savršena merna metoda i umesto da se vreme troši na istragu ko je odgovoran za propuste (koja jeste važna, ali ne sme da bude primarna u radu sa podacima), potrebno je efikasno iskoristiti dobijene rezultate. Nakon donetih zaključaka, vreme bi trebalo posvetiti preporukama za druge istraživače ili pak za ponavljanje mernog procesa.

2.1.1 Preregistracija kao primer dobre prakse

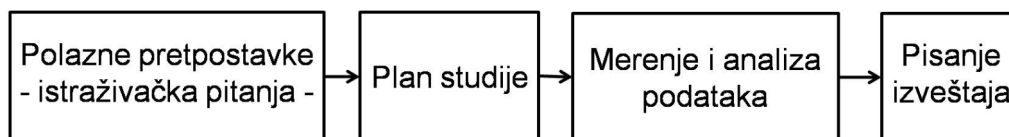
Jedna od najboljih praksi u metodološkom pristupu jeste preregistracija koja je na žalost prema Autorkinim saznanjima u primeni u samo oblasti psihologije. Preregistracija je primer dobre prakse u nauci, a principi kojima se služi mogu se primeniti i u industriji. Osnovni koraci u ovoj metodi su prikazani na Sl. 49:

- najpre se postave polazne pretpostavke i istraživačka pitanja (na primer: da li se broj treptaja menja usled pospanosti),
- potom se pripremi kompletan plan studije koji uključuje i protokol merenja signala i metodu analize signala sa razradom svih mogućih detalja (na primer: podaci će biti mereni primenom EOG metode na 20

zdravih ispitanika tokom dana i rano ujutru kada se probude, signali će biti prikupljeni tokom 20 minuta sa odgovarajućim parametrima (uređaj za merenje signala, frekvencija odabiranja, rezolucija A/D konverzije), a podaci će biti analizirani primenom metode za detekciju lokalnih maksimuma kako bi se izdvojili svi treptaji na EOG signalu na kome su prethodno otklonjeni svi šumovi, potom će se primeniti odgovarajući statistički testovi kako bi se utvrdilo da li postoji statistički značajna razlika između broja treptaja kada su ispitanici pospani odmah nakon buđenja i kada nisu tokom dana),

- merenje i analiza signala se sprovode nakon što postoji detaljan plan i za merenje i za analizu signala i
- na kraju se piše kompletan izveštaj.

Faza pisanja izveštaja (poslednji korak u blok dijagramu prikazanom na Sl. 49) ne pretpostavlja pisanje izveštaja od početka, već samo pisanje rezultata, diskusije i zaključka studije. Tokom planiranja studije pretpostavlja se da su već napisani uvod (koji sadrži polaznu pretpostavku ili pretpostavke) i ukupna metoda rada.



Slika 49, Blok dijagram koraka za objavljivanje preregistrovane studije. Slika je inspirisana blok dijagramom sa internet stranice Centra za otvorenu nauku (eng. *Centre for Open Science*) <https://www.cos.io/initiatives/registered-reports>, CC BY-NC-SA 4.0, pristupljeno 06.08.2023.

U nauci je praksa preregistracije prepoznata i od strane naučnih časopisa, pa se plan studije šalje na recenziju (ocenu od strane stručnjaka, odnosno drugih naučnika) pre nego što se realizuju merenja i primeni metoda analize signala. Ovo daje mogućnost da nezavisni stručnjaci nepristrasno, profesionalno i objektivno ocene ne samo protokol merenja, već i metodu analize pre nego što se počne sa izradom studije. Po pravilu, odluka da li je plan studije sadržan u delimično¹⁴⁶ napisanom izveštaju se ne donosi odmah po slanju izveštaja u časopis, već se vraća na korekciju (recenziju) Autorima. Recenzent može da predloži da se izmeni metodologija i da se ukupnoj analizi priključi parametar koliko su ispitanici sna imali tokom noći i tokom dana, jer to može da utiče na njihovu pospanost. Nakon toga, studija se sprovodi i kompletan izveštaj, odnosno rad ponovo prolazi proces recenzije (koji sada ima za cilj samo da

¹⁴⁶ Delimično, zato što se ne može napisati kompletan izveštaj pre merenja i analize signala (u izveštaju odnosno naučnom radu ili preregistrovanom izveštaju od eng. *registered report* nedostaju rezultati, diskusija i zaključak).

oceni predstavljanje rezultata i strukturu diskusije i zaključka) i na kraju se objavljuje u časopisu (poslednji korak objavljivanja nije prikazan na Sl. 49). Procenjeno je da trenutno više stotina naučnih časopisa primenjuje ovakvu praksu. Preregistracija pomaže da se izbegnu zamke koje su već opisane prilikom analize signala, ali i stavlja akcenat na metodologiju i plan istraživanja. Dodatno, povećava se transparentnost i omogućava publikovanje negativnih rezultata. Na primer, ako rezultati analize pokažu da nije moguće odrediti stepen pospanosti zdravih ispitanika EOG merenjima, takav rad bi u tradicionalnom pristupu naučnim publikacijama bilo skoro nemoguće objaviti u naučnom časopisu, a ipak je važno da drugi istraživači čitaju o negativnim rezultatima kako bi se znala ograničenja EOG metode. [165]

2.1.2 HARKovanje kao primer loše prakse

Jedna od loših praksi koja je prepoznata u naučnim publikacijama, a koja bi se sprečila primenom preregistracije je HARKovanje (eng. *HARKing* od eng. *Hypothesizing After the Results are Known*) [165-166]. Na primeru korišćenja EOG signala za procenu pospanosti zdravih ispitanika, to bi značilo da su merenja prvo izvršena, da je na primer¹⁴⁷ zaključeno da frekvencija treptaja nije u korelaciji sa pospanošću, ali da amplituda treptaja koja se meri primenom EOG signala jeste. HARKovanje bi značilo da se onda promene hipoteze i da se odmah na početku navede kako se za frekvenciju treptanja pretpostavlja da nije od značaja, ali da amplituda jeste, umesto da se diskutuje zašto polazna pretpostavka o broju treptaja nije dokazana.

Postoje izuzeci kada je HARKovanje opravdano (na primer, da se izmeni metoda i da se proširi analiza signala na procenu amplitude treptaja, jer taj parametar nije bilo moguće proceniti pre realizacije merenja i pre eksplorativne analize podataka koja uključuje vizuelizaciju merenih signala), ali svakako HARKovanje treba izbegavati u većini praktičnih situacija [166]. Iako je ovakva praksa svojstvena naučnim radovima, postoji puno razloga da se bilo koja studija (u akademiji ili industriji) sprovodi na sličan način uz korišćenje spoljnih stručnjaka (u slučaju naučnih časopisa to su recenzenti) pre nego što se realizuju merenje i analiza signala.

¹⁴⁷ Broj treptaja se menja sa pospanošću – veći zamor ili pospanost dovode do povećanog broja treptaja koji traju duže, ali postoji uticaj individualnih razlika i spoljašnjih faktora kao što su fizički zamor i nivo pospanosti koji mogu uticati na broj treptaja [167-168]. To je na primeru EOG signala i pokazano u okviru diplomskog rada na Elektrotehničkom fakultetu, Univerziteta u Beogradu studenta Miljana Dilparića u 2016. godini. Ovde je takva studija predstavljena hipotetički za potrebe definisanja preregistrovane studije i prakse HARKovanja.

2.1.3 Kvalitativni pristup je važniji od kvantitativnog?

Bez obzira na sve dobre prakse koje postoje i bez obzira na upotrebu automatskih procedura realizovanih u programskim jezicima za obradu signala, treba imati na umu upotrebnu vrednost i razumljivost rezultata. Upotrebna vrednost je usko povezana sa primenom rezultata i najviše se odnosi na kliničke studije i primenu terapijskih metoda ili lekova na pacijentima. Ovo će detaljnije biti objašnjeno na pojednostavljenom primeru ocene terapijske metode. Prilikom ispitivanja uticaja električne stimulacije na urinarnu inkontinenciju nije dovoljno porediti samo kvantitativne parametre (broj inkontinentnih epizoda), već je potrebno porediti i parametre koji određuju kvalitet života. Urinarna inkontinencija se najjednostavnije definiše kao nevoljan gubitak urina [169]. Koliko se kvalitet života i mogućnost ispitanika da obavljaju svakodnevne aktivnosti bez korišćenja dodatne zaštite (na primer, korišćenje pelena za odrasle) razlikuje ako je posle tretmana broj inkontinentnih epizoda tokom dana smanjen sa 10 na dve? Kvantitativno, ovo je veliko poboljšanje od čak 5 puta (koje bi se sigurno iskazalo i kao statistički značajno ako bi se statistički testovi primenili na većem broju ispitanika sa sličnim ishodom), ali kvalitativno ne mora da bude. Ako pacijent želi da izlazi van kuće bez dodatne zaštite, onda sigurno nije. Sa druge strane, ako pacijent očekuje od terapije da se broj neprijatnih inkontinentnih događaja smanji na minimum, onda jeste.

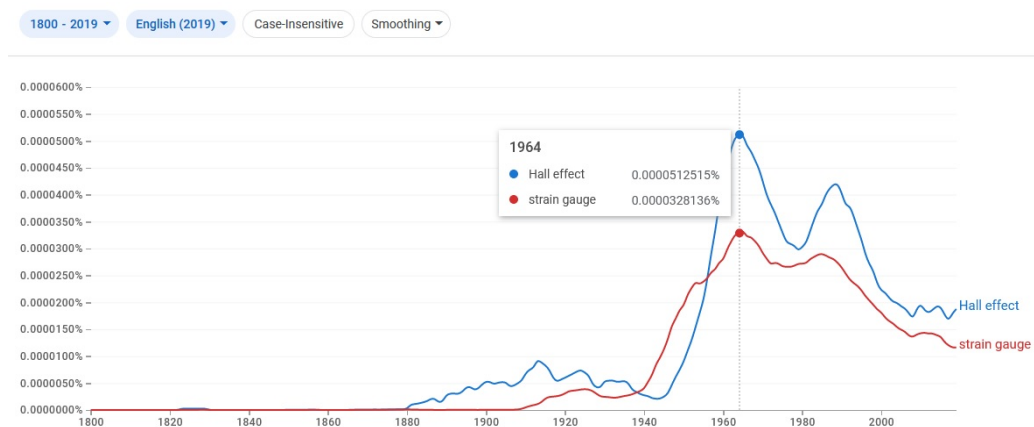
Kvalitet života je vrlo subjektivna ocena i može zavisi od niza faktora, a primarno od individualnog doživljaja kvaliteta života [170]. Svetska zdravstvena organizacija (SZO, eng. *World Health Organization*, WHO¹⁴⁸) je razvila instrument / upitnik, kao i odgovarajuće preporuke za njegovu primenu u različitim kulturama [171]. Upitnik sadrži na pitanja kao što su: “*How much do you worry about your health?*” (srp. Koliko brinete za svoje zdravlje?, slobodan prevod sa engleskog na srpski jezik), a odgovori su ponuđeni u formi Likertove skale¹⁴⁹ (za navedeno pitanje to su: “1. *Not at all*, 2. *A little*, 3. *A moderate amount*, 4. *Very much*, 5. *An extreme amount*”, srp. 1. Uopšte ne brinem, 2. Malo brinem, 3. Umereno brinem, 4. Veoma brinem, 5. Brinem u velikoj meri, slobodan prevod sa engleskog na srpski jezik) [171].

Nakon ovakve kritike kvantitativnih mera i automatske analize signala na računaru, skeptični studenti mogu logično postaviti pitanje da li je uopšte automatska analiza signala potrebna u kliničkom okruženju? Odgovor je da, naravno da jeste. Prvi razlog je to što je takva metoda nezamenjiva da se utvrdi

148 <https://www.who.int/>, pristupljeno 30.05.2024.

149 https://en.wikipedia.org/wiki/Likert_scale, pristupljeno 30.05.2024.

da li uopšte električna stimulacija utiče na inkontinenciju. Ako bi se merio kvalitet života pacijenta, možda se nikada ne bi došlo do odgovora na ovo pitanje. Sa druge strane, kvalitet života definiše primenljivost u praksi, a ne naučni dokaz. Prema tome, možda električna stimulacija nije dostigla nivo da bude primenjiva u praksi, ali zato jeste dokazana njena uspešnost u lečenju inkontinencije. Logično je da su se naučnici u svojim studijama opredelili da koriste i kvantitativne i kvalitativne mere kako bi utvrdili i efikasnost i praktičnu upotrebljivost električne stimulacije [169-173]. To što je zaključak preglednog rada iz 2017. godine [173] da električna stimulacija ima dokazanu efikasnost, ali da nije praktično korisna, ne treba da spreči naučnike da i dalje istražuju ovu metodu i da je unapređuju.



Slika 50, Poređenje broja reči u knjigama iz Gugle korpusa za Holov efekat i za mernu traku od 1800. do 2019. godine. Iako će o graficima biti više reči u narednim poglavljima, ovde bi trebalo spomenuti da je ovo tip interaktivnog grafika koji omogućava da se prelaskom strelice računarskog miša prikažu vrednosti za godine koje su odabrane (na slici je prikazano kako je moguće pogledati vrednosti za 1964. godinu). Grafik je moguće jednostavno generisati preko veb Ngram alata: <https://books.google.com/ngrams/> (pristupljeno 30.05.2024).

Da li je primereno ocenjivati svaku metodu samo sa stanovišta njene praktične upotrebljivosti ili ne? Može li se za svaku metodu u bilo kom trenutku oceniti njena praktična upotrebljivost? Odgovori na oba pitanja je isti: Ne.

Danas se u svetu (kroz fondove finansiranja i u naučnim časopisima), uključujući Republiku Srbiju akcenat stavlja na primenjena istraživanja. Može se reći da takav stav odgovara industrijalizaciji, ali nauka ne mora uvek biti primenjiva u praksi ili bar ne mora biti odmah primenjiva u praksi. Zato je važno da se kroz javna finansiranja podrže i naučna istraživanja koja imaju za cilj da prošire saznanja. Niz istorijskih primera dokazuje ovu tvrdnju. Holov efekat je otkrio Edvin Hol u 19. veku i predstavio u naučnom radu 1879. godine

[161]. Ovo naučno otkriće je našlo primenu tek sa pojavom integrisanih kola u 20. veku, a jedna od prvih primena je iz 1969. za konstrukciju računarske tastature [174]. Dakle, trebalo je da prođe 90 godina da bi Holovo naučno otkriće dobilo svoju prvu praktičnu primenu. Holovi pretvarači su danas toliko zastupljeni da se čak može govoriti o većoj popularnosti Holovih pretvarača od mernih traka ne samo u biomedicinskom inženjerstvu, već u svim granama u kojima se koriste senzori.

Za ilustraciju koliko je Holovo otkriće značajno i kakva je njegova današnja primena u odnosu na primenu mernih traka, iskorišćena je Gugl alatka Books Ngram Viewer¹⁵⁰ koja služi za poređenje popularnosti tekstualnih stringova u knjigama koje su dostupne u Guglovom korpusu. Na Sl. 50 prikazan je grafik iz Ngram alata za poređenje ključnih reči eng. *Hall effect* (srp. Holov efekat) i eng. *strain gauge* (srp. merna traka¹⁵¹). Sl. 50 jasno prikazuje popularnost Holovog efekta u odnosu na merne trake u periodu od 1800. do 2019. godine. Generalno, saradnja nauke i privrede je vrlo upitna po mnogim elementima [175], a ovaj ilustrativni primer samo pokazuje da ako industrija diktira nauci pravce istraživanja i razvoja (kao što je to nažalost danas slučaj), ni industrija ne može od toga imati koristi u budućnosti. Naime, Hol nikada ne bi bio finansiran za primenu koju možda nije ni predvideo, a industrija danas ne bi raspolagala Holovim sensorom da je primenjena nauka bila dominantno zastupljena u akademskim krugovima.

Gugl alatka prikazana na Sl. 50 omogućava da svi koji su zainteresovani mogu da preuzmu javno dostupne podatke pod slobodnom CC-BY 3.0 licencom¹⁵², te da sami kreiraju grafik. Međutim, kako se radi o veoma malom skupu podataka koji je ovde od interesa (pretraga samo dve ključne reči), u CRAN repozitorijumu je dostupan paket ngramr [176] koji omogućava da se preuzmu podaci za odabrane ključne reči i da se dalje manipuliše podacima, odnosno da se prikažu podaci korišćenjem funkcija u programskom jeziku R.

150 https://books.google.com/ngrams/graph?content=Hall+effect%2Cstrain+gauge&year_start=1800&year_end=2019&corpus=en-2019&smoothing=3, pristupljeno 06.08.2023.

151 Odabrana je merna traka za poređenje sa Holovim efektom, jer je merna traka jedna od najpoznatijih komponenti bez koje se ne može zamisliti nijedan kurs iz električnih merenja, a koja služi za konstrukciju velikog broja senzora [98].

152 <https://storage.googleapis.com/books/ngrams/books/datasetsv3.html>, pristupljeno 06.08.2023.

2.1.4 Pristrasnost u razumevanju rezultata

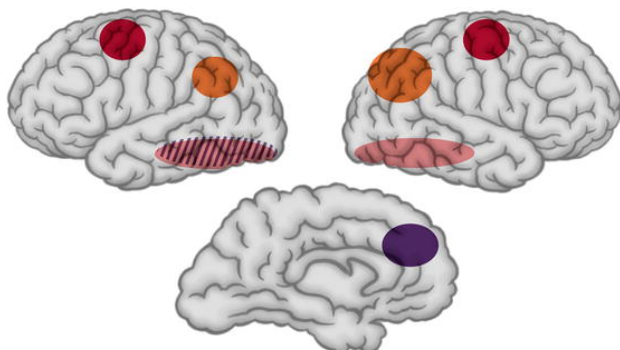
Razumljivost rezultata ili pristrasnost u razumevanju numeričkih vrednosti je problem prepoznat u psihologiji. U 2023. godini je jedan članak u Politici podsetio na istraživanja izraelskih naučnika nobelovca Danijela Kanemana i Amosa Tverskog da ljudi imaju kognitivnu pristrasnost. Praktično, mi jedino razumemo šta znači 0% (događaj se sigurno neće desiti), 100% (događaj će se sigurno desiti) i 50% (podjednake su šanse da se događaj dogodi ili ne dogodi). Posebno je ispitivano kako naučnici tumače rezultate svojih istraživanja i statističkih testova, ali i kako pacijenti donose odluku kada treba da odaberu odgovarajuću terapiju na osnovu dostupne statistike o mortalitetu i uspešnosti terapije. [177-180]

Pored kognitivne pristrasnosti, trenutno su samo ljudi u stanju da pišu programske kodove i da analiziraju signale, te drugi izbor ne postoji. Naravno potrebno je očuvati svest o zamkama pristrasnosti koje mogu biti kamen spoticanja u dolasku do rezultata analize ili nedolasku do njih. Čuveni citat nobelovca i ekonomiste Ronalda Kosa glasi “Ako dovoljno mučite podatke, priznaće bilo šta”¹⁵³ upravo ukazuje na značaj uloge onoga ko sprovodi analizu. Primena neodgovarajućih metoda za obradu signala može imati katastrofalne posledice po konačan rezultat, odnosno može se dobiti informacija koja nikada nije ni bila sadržana u signalu. Da li je pitanje predikcije cena na berzi tačno ili ne, možda nije od direktnog uticaja na život ljudi, ali pitanje da li je predikcija za pacijente da moraju ili ne moraju na operaciju ili da li će preživeti ili ne tretman koji im se nudi ima vrlo direktan uticaj na nečije zdravlje i očuvanje života. Analiza biosignala sa sobom nosi odgovornost, te bi profesionalizam i objektivnost uvek trebalo da su na prvom mestu ili kako je to rekao pokojni profesor Elektrotehničkog fakulteta, Univerziteta u Beogradu Vladislav Đ. Jovanović “Najvažnija stvar u svakom merenju jeste čestitost.” [181]. Iako bi generalno proširenje bilo “u svakom poslu”, ovde je ovaj citat primeren, jer se itekako primenjuje na obradu signala.

Na kraju ovog potpoglavlja, valja spomenuti zanimljiva istraživanja na spoju neuronauka i psihologije koja su usmerena ka identifikaciji regija u mozgu koje su aktivne tokom zadatka logičke operacije poređenja brojeva (rezultati studija su dati na Sl. 51). Kako bi se utvrdilo koje su regije aktivne tokom kognitivnog zadatka poređenja brojeva realizovana su istraživanja koja su uključila sinhrono

153 Slobodan prevod sa engleskog na srpski jezik, a u originalu na engleskom jeziku citat glasi: “If you torture the data long enough, it will confess to anything.”, https://en.wiktionary.org/wiki/if_you_torture_the_data_long_enough_it_will_confess_to_anything, pristupljeno 08.08.2023.

merenje EEG signala i medicinske slike primenom fMRI (eng. *Functional Magnetic Resonance Imaging*) metode. Proces koji je ilustrovan na Sl. 51 se naziva mentalna hronometrija i predstavlja prostornu i vremensku organizaciju kognitivnih operacija. [182]



Vision (150 ms)

-  Arabic Digits
-  Arabic Digits, Spelled Numbers
-  Comparison (190 ms)
-  Movement (330 ms)
-  Error correction (470 ms)

Slika 51, Regije mozga koje su aktivne tokom kognitivnog zadatka poređenja brojeva. Iz <http://biology.plosjournals.org/perlserv/?request=get-document&doi=10.1371/journal.pbio.0030051>, dostupno pod CC BY-SA 2.5 licencom, <https://en.wikipedia.org/w/index.php?curid=4776115>.

2.2 Pretprocesiranje biosignala za eliminaciju šuma

Analiza biosignala predstavlja dinamičnu i privlačnu oblast u konstantnom razvoju. Izazovi koji se predstavljaju pred biomedicinske inženjere su uslovljeni kako kompleksnom prirodom biosignala, tako i njihovom osetljivošću na šum. Za razliku od signala koji se mere u drugim oblastima (automatiki, elektronici ili telekomunikacijama), signali u biomedicinskom inženjerstvu su manjih amplituda (vrlo često i nekoliko puta manjih od šuma) i nižeg frekvencijskog sadržaja (većina biosignala ima frekvencijski sadržaj koji se poklapa sa mrežnim napajanjem ili sadržaj koji je vrlo blizak DC (eng. *Direct Current*) komponenti). Ovde nije predmet razmatranja, ali je važno spomenuti da se slični izazovi sa šumom koji ima značajno veće amplitude od signala sreću i u

energetici. Na primer, prilikom merenja parcijalnog pražnjenja (eng. *partial discharge*) uobičajeno se mora rešiti postojanje šuma i nepovoljan odnos SNR kao neizostavan deo analize signala u energetici [183-184].

Šum u biomedicinskom signalu se nekada naziva artefakt – ovaj termin nema isto značenje kao artefakt¹⁵⁴ koji je u upotrebi u arheologiji i označava predmet koji je napravio ili prepravio čovek, kao i proizvode u tehnološkom razvoju [185]. Smetnja, interferencija (eng. *interference*), artefakt i šum su sinonimi. Po definiciji, šum je bilo koji signal koji sadrži informaciju koja nije od interesa [186]. U literaturi na engleskom jeziku, ova tri sinonima imaju ustaljenu upotrebu, te se za šum pokreta koristi termin *movement artifact*, za šum napajanja *power interference*, a za slučajni šum *random noise* [79]. Postoje i drugi šumovi na biosignalima kao što su šum koji je posledica alajzinga, termalni šum i šum koji nastaje kao posledica pomeranja elektroda (nešto je drugačiji od šuma pokreta, jer često pomeraj elektroda rezultuje gubitkom kontakta elektrode sa kožom što se manifestuje na drugačiji način od npr. pomeranja kablova kada kontakt može biti stabilan) [79, 151]. Alajzing nastaje usled pogrešne rekonstrukcije signala kod čijeg merenja nije ispunjen uslov teoreme odabiranja i nije specifičan isključivo za biosignale, već kod svih merenja kod kojih postoji proces digitalizacije [78]. Termalni šum se ređe analizira u obradi signala, jer je usko povezan sa elektronikom biomedicinskih uređaja, a najčešće nastaje na otpornicima i pojačavačima [187].

2.2.1 Merenje šuma

Jedan od najvećih problema u pretprocesiranju biosignala (koje ima za cilj eliminaciju šuma) predstavlja šum napajanja na 50 Hz ili 60 Hz ili takozvanu elektromagnetnu smetnju (eng. *ElectroMagnetic Interference*, skraćeno EMI), jer se ne može u potpunosti eliminisati prilikom merenja, za razliku od šuma pokreta koji se može minimizovati uz odgovarajuće protokole sa smanjenim opsegom i brojem pokreta ispitanika tokom merenja [79]. Eliminacija šuma pokreta je očigledna u ambulatornom merenju EKG signala gde se ispitanik postavlja u ležeći položaj, a klinički EKG signal se meri tokom mirovanja. To nije slučaj tokom testa opterećenja (ergometrija srca, eng. *stress test*) kada se od pacijenta očekuje fizička aktivnost (vožnja sobnim biciklom ili korišćenje pokretne trake pod različitim nagibom kako bi se upravljalo opterećenjem i procenio uticaj fizičkog napora na srce), te se metode za otklanjanje šuma EKG signala posebno sa ciljem odgovarajuće delineacije (detekcije karakterističnih

154 <https://sr.wikipedia.org/wiki/Artefakt>, pristupljeno 09.08.2023.

tačaka) još uvek razvijaju i unapređuju, ali i razlikuju od uslova merenja EKG signala [188].

Iako je šum bolje sprečiti prilikom merenja nego ga kasnije eliminisati, to nije uvek moguće kao što je pokazano na primeru testa opterećenja. Jedini način da se spreči EMI jeste da se primeni Faradejev kavez. Međutim, smeštati ispitanike u Faradejev kavez bi značajno umanjilo prenosivnost uređaja, onemogućilo primenu odevnih tehnologija i bežične komunikacije, ali i smanjilo udobnost ispitanika prilikom merenja¹⁵⁵. Primena personalizovanog Faradejevog kaveza je gotovo iščezla iz upotrebe u medicini 30ih godina prošlog veka, a kavez je korišćen jer nije bilo moguće izmeriti EEG signale koji su relativno malih amplituda (reda μV) [189]. Danas to nije slučaj. Faradejevi kavezi su danas u upotrebi za EMI izolaciju prostorija u kliničkom okruženju. Na primer, moguće je izolovati od EMI Faradejevim kavezom cele prostorije u kojima se sprovode merenja elektrofizioloških signala u klinikama – primer takvog uređaja je MU-COPPER^{TM156} (Holland Shielding Systems B.V, Dordrecht, The Netherlands).

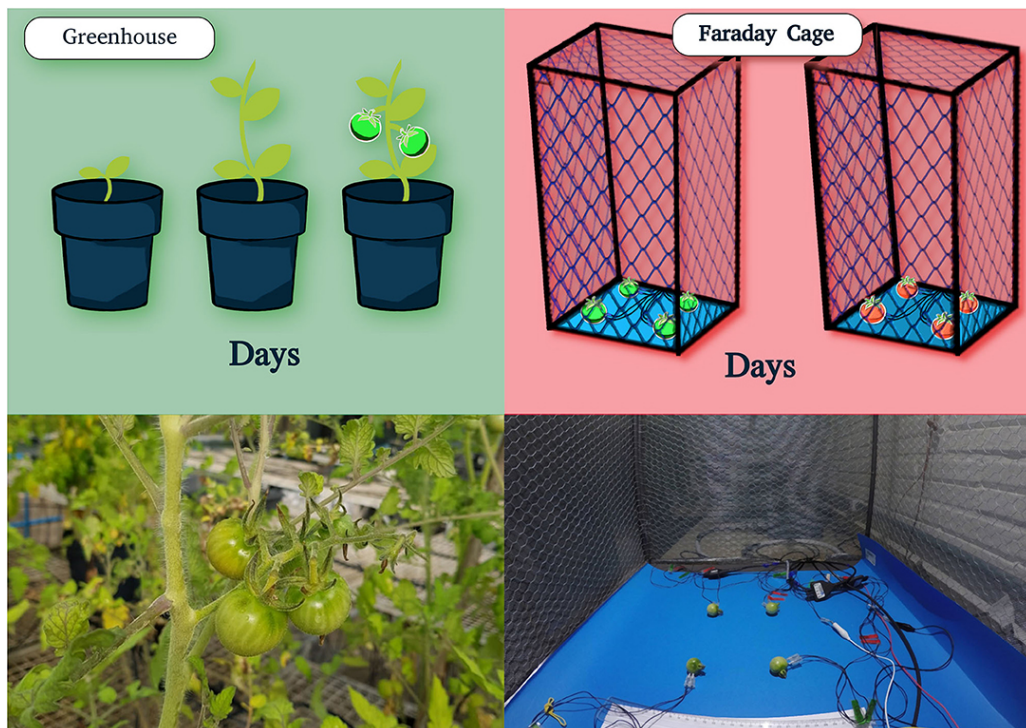
U praksi je danas neizbežno korišćenje Faradejevog kaveza za merenja biosignala koji su manji od μV , kao što su biosignali biljaka. Na primer, za merenje promene elektrofizioloških potencijala u različitim fazama sazrevanja voća kao što je paradajz. Da bi se izmerili električni potencijali paradajza od faze kada nastaje plod zelene boje pa do sazrevanja (faza svetlo crvene boje kada 60% do 90% ploda nije prekriveno zelenom bojom), Resig i saradnici su koristili Faradejev kavez (Sl. 52). Bez obzira što je merenje izvršeno u Faradejevom kavezu, pored para elektroda koji je bio postavljen u svaki paradajz na rastojanju od po 1 cm, jedan par nepovezanih elektroda je tokom trajanja eksperimenta ostao u kavezu. Razlog je merenje šuma iz okruženja, jer elektrofiziološki potencijali paradajza imaju veoma male amplitude ($< \mu\text{V}$) koje mogu biti i za više redova veličine manje od šuma. Signal sa otvorenih, odnosno nepovezanih elektroda je poslužio da bi se proverilo da li je šum unutar Faradejevog kaveza bio stabilan tokom merenja, kao i da bi se sa sigurnošću promene merene na paradajzu mogle pripisati elektrofiziološkim procesima unutar voća. [190]

Merenja elektrofizioloških potencijala biljaka su u velikom pivoju, pa su u razvoju i uređaji za merenje elektrofizioloških potencijala biljaka bez

155 Verovatno dodatna obrazloženja pažljivim čitaocima nisu potrebna, ali nijednom ispitaniku se ne bi dopalo da ga stave u metalni kavez tokom dijagnostičke procedure.

156 <https://faradaycages.com/medical/eeg-eog-ecg-emg-cmap-and-erg-rooms->, pristupljeno 09.08.2023.

Faradejevog kaveza. Primer takvog uređaja je PhytlSigns koji je takođe testiran na paradajzu [191]. Za razliku od merenja elektrofizioloških potencijala plodova paradajza [190], uređaj PhytlSigns je korišćen za merenje električnih signala sa stabljike paradajza koji su osetljivi na dnevne promene intenziteta svetlosti iz okruženja [191].



Slika 52, Merenje elektrofizioloških potencijala na paradajzu primenom Faradejevog kaveza. Na gornjim panelima prikazane su skice dve faze: (1) faza rasta voća do trenutka pojave zelenih plodova (gore levo) i (2) faza merenja i sazrevanja u Faradejevom kavezu (gore desno). Na donjim panelima prikazane su fotografije koje odgovaraju fazama prikazanim na skicama na gornjim panelima. U svaki paradajz unutar Faradejevog kaveza postavljene su po dve elektrode na međusobnom rastojanju od po 1 cm. Slika je dostupna pod CC BY 4.0 licencom, <https://creativecommons.org/licenses/by/4.0/>, pristupljeno 07.08.2023. [190]

Primer merenja biosignala sa paradajza [190] u kome postoji dodatni kanal za merenje ambijentalnog šuma ili šuma okruženja Faradejevog kaveza je odličan pristup u planiranju realizacije obrade signala. Kad god je moguće, potrebno je prikupiti što više informacija o šumu ili izmeriti šum. To je uobičajeno kod merenja EEG signala. Na elektrodama postavljenim u frontalnom režnju skalpa (blizu čela na glavi) pored EEG signala može se izmeriti i šum koji nastaje dominantno od treptaja. Kako bi se taj šum na odgovarajući način filtrirao, zajedno, odnosno sinhrono sa EEG signalima, meri se i EOG signal [192]. Iako

su inicijalno u upotrebi bile linearne metode za otklanjanje EOG šuma iz EEG signala [192], vremenom su se razvijale adaptive metode eliminacije šuma koje više odgovaraju nestacionarnoj prirodi EEG i EOG signala [193], ali i metode koje isključuju dodatno EOG merenje [194]. Bez obzira što dodatni kanal za merenje EOG šuma, ne mora uvek da bude nužno potreban [194] i što se električni potencijal paradajza može izmeriti bez primene Faradejevog kaveza [191], kvantitativna procena šuma ili dodatno merenja šuma spada u dobru praksu.

2.2.1 Priprema podataka za dalju obradu

Vrlo često je kao pretkorak analize signala potrebno primeniti tehnike čišćenja podataka (eng. *data cleaning*) ili metode za pripremu podataka za dalju obradu na računaru. Ovaj pretkorak se ređe sreće u analizi vremenskih serija biosignala, ali je obavezan u slučaju da podaci sadrže nedostajuće podatke. Priprema podataka za dalju obradu koristi se za analizu vremenskih serija u situaciji kada su signali mereni primenom bežičnog prenosa koji je uobičajen za odevne tehnologije. Iako primena bežične komunikacije omogućava manje šuma pokreta merenih biosignala koji uglavnom nastaje zbog pomeranja kablova koji su posledica pokreta ispitanika i istovremeno može doprineti da se uređaji realizuju bez potrebe za galvanskom izolacijom, primena ovakve komunikacije ima svoju cenu. To je loš kvalitet podataka usled paketa koji se gube¹⁵⁷ u komunikaciji koji može da bude 30% ili više, iako je tačnost dobijenih obeležja bila relativno pouzdana [80].

U sistemima u kojima se vrši merenje većeg broja biosignala potrebno je sinhronizovati sve merne uređaje kako bi se izvršila odgovarajuća analiza signala. Najčešće se sinhronizacija u biomedicinskim uređajima jednostavno postiže primenom zajedničkog A/D konvertora za merenje biosignala. Međutim, kada se koriste bežični sistemi koji su osnova odevnih tehnologija ili kada je u upotrebi distribuiran merni sistem, onda se mora koristiti sinhronizacija [195]. Takva merenja obično dolaze sa datotekama koje sadrže podatke merene sa različitih uređaja i sa različitim vremenskim potpisima događaja (eng. *timestamp*). Postoji više načina sinhronizacije pomoću vremenskih potpisa događaja, a jedan je na osnovu potrošnje energije bežičnog sistema zasnovanog na Blututu (eng. *Bluetooth*) [196]. Uobičajeno je da se

¹⁵⁷ Ovde treba razlikovati situacije u kojima nedostaju podaci zbog interne predobrade unutar samog senzora, koja se, na primer, kod Empatica E4 (Empatica, Cambridge, Massachusetts, United States) uređaja u obliku narukvice zasniva na proveru kvaliteta signala ili parametra koji se dobija iz senzora (<https://support.empatica.com/hc/en-us/articles/360030058011-E4-data-IBI-expected-signal>, pristupljeno 10.08.2023).

postupak sinhronizacije nekada sprovodi nakon merenja, kako bi podatke bilo moguće analizirati (npr. definisanje jedinstvene vremenske ose). Ova procedura se ubraja u pripremu podataka za dalju obradu. U nekim situacijama to nije potrebno uraditi, jer se obrada signala ne sprovodi na sinhrono snimljenim odbircima, već je samo važan trenutak početka ili kraja događaja od interesa. Međutim, u određenim situacijama je važno da se dobiju u potpunosti sinhrono vremenske ose, a podaci mogu biti mereni i sa različitim frekvencijama odabiranja. Volumen srca¹⁵⁸ (eng. *cardiac output*) je moguće proceniti primenom istovremenog merenja IKG i EKG signala, te je sinhronizacija presudna [197]. U slučaju da frekvencija odabiranja nije ista za signale, a dalja obrada to zahteva, moguće je primeniti tehnike za smanjenje (eng. *down sampling*) ili povećanje (eng. *up sampling*) frekvencije odabiranja ili obe istovremeno. Jeftiniji senzori (eng. *low cost*) imaju često varijabilnu periodu odabiranja [199] koja je usko povezana sa energetsom potrošnjom koju treba minimizovati posebno u merenjima pomoću mobilnih telefona [200-201]. Jedna od posledica merenja bežičnim prenosivim uređajima je i promenljivo kašnjenje [196] koje može dovesti do toga da je frekvenciju odabiranja i performanse odabiranja mernog sistema potrebno proceniti, tek nakon merenja, neposredno pre obrade signala [202].

2.2.1.1 Primer pripreme podataka za merenje opsega pokreta

U ovom poglavlju će biti prikazana primena metode za pripremu podataka za dalju obradu opsega pokreta koja pripada oblasti kinetike i osteokinetike. Kinematika je oblast koja se bavi kretanjem, a osteokinematika se bavi kretanjem koštanog sistema u prostoru. U opštem slučaju, opseg pokreta (eng. *Range of Motion*, ROM) predstavlja linearnu ili ugaonu razdaljinu koji objekat ili subjekat može da pređe [203-205].

U biomedicinskom inženjerstvu, ROM se odnosi na razdaljinu i pravac zgloba između dve pozicije koje se definišu kao fleksija, ekstenzija, adukcija, abdukcija i rotacija zgloba, a predstavlja maksimalan opseg koji se može izvesti. Termin je vrlo često zastupljen u rehabilitaciji i fizikalnoj terapiji, te se razlikuju sledeće vrste ROM-a [203-205]:

- AROM (eng. *Active Range of Motion*) – aktivni opseg pokreta,
- PROM (eng. *Passive Range of Motion*) – pasivni opseg pokreta i

158 Uobičajena metoda merenja volumena srca je primenom arterijskog katetera, ali je invazivna i nije pogodna, na primer, za primenu u istraživanjima iz psihologije koja su usmerena na kvantitativnu procenu emocija, jer sa sobom nosi rizik od mogućih komplikacija i nelagodnost tokom merenja [197-198].

- AAROM (eng. *Active Assistive Range of Motion*) – aktivni opseg pokreta uz asistenciju.

AROM nastaje kada ispitanici voljno pomeraju deo tela bez asistencije za razliku od PROM-a koji nastaje kada terapeut ili spoljni rehabilitacijski uređaj primeni spoljnu silu koja dovodi do pomeranja dela tela bez učešća voljnog pomeranja. AAROM se nalazi na granici između AROM i PROM i predstavlja voljno pomeranje zgloba uz asistenciju (manuelnu ili mehaničku), odnosno uz primenu spoljne sile. ROM se koristi u oceni terapijskih modaliteta (najčešće se sprovodi isključivo AROM u tim slučajevima), ali predstavlja i način na koji se sprovodi fizikalna terapija (AAROM i PROM). U kliničkom okruženju treba voditi računa o kontraindikacijama (kada se ROM ne primenjuje), ograničenjima ROM-a (kada ROM nije koristan) i odgovarajućim procedurama za njihovo pravilno sprovođenje u procesu rehabilitacije pacijenata. [203-205]

Ova merenja su posebno zastupljena u rehabilitaciji pacijenata nakon povrede nastale usled moždanog udara (eng. *stroke*) ili usled preloma u koštanom sistemu [203-205]. U laboratorijskim uslovima za merenje ROM se uobičajeno koriste elektrogoniometri ili IR (infracrvene, eng. *infrared*) kamere, ali u kliničkom okruženju zastupljene su pasivne metode koje se zasnivaju na korišćenju pokretnih uglomera (goniometara) koji predstavljaju isključivo mehaničke instrumente¹⁵⁹ [204]. Elektrogoniometar ili električni uglomer se koristi za merenje ugla između dva rigidna tela. Osnovu ovih senzora čine merne trake koje se vezuju u merni most, iako postoje realizacije sa potenciometrima [207]. Da bi se na odgovarajući način izmerio ugao nekog zgloba primenom elektrogoniometra potrebno je pre početka merenja, kalibrisati instrument. Za odgovarajuće merenje ROM, potrebno je na odgovarajući način postaviti goniometar na zglob čiji se ROM meri.

159 Ima više razloga zašto tehnološki naprednije rešenje nije u primeni. Iako su relativno malih dimenzija i predstavljaju prenosiv tip instrumenta, elektrogoniometri su relativno osetljivi na habanje i koriste baterijsko napajanje. Iako su digitalni toplomeri zastupljeni na klinikama, nije jednostavno održavati uređaje sa baterijskim napajanjem u kliničkom okruženju, jer je potrebno stalno nadgledanje stanja baterije i njeno punjenje. Rezolucija elektrogoniometara je manja od mehaničkih goniometara (0.1° za Biometrics goniometre, <https://www.biometricsltd.com/goniometer.htm>, pristupljeno 09.08.2023), ali tačnost je samo 2° što nije značajno manje u odnosu na mehaničke instrumente. IR kamere [206] su precizne, tačne i sa sobom nose značajnu količinu informacija (više od ROM-a, jer mogu da mere pokrete tela u 3D prostoru), ali zahtevaju odgovarajuću osvetljenost, nisu portabilne, a kalibracija je vremenski zahtevna za pregled u kliničkom okruženju. IR kamere snimaju položaj markera koji su postavljeni na telo ispitanika i uobičajeno se koriste u kombinaciji sa platformama za merenje sile koje vrši stopalo tokom hoda ili trčanja na podlogu (OptiTrack, NaturalPoint Inc, Corvallis, OR, USA, <https://optitrack.com/applications/movement-sciences/>, pristupljeno 10.08.2023).



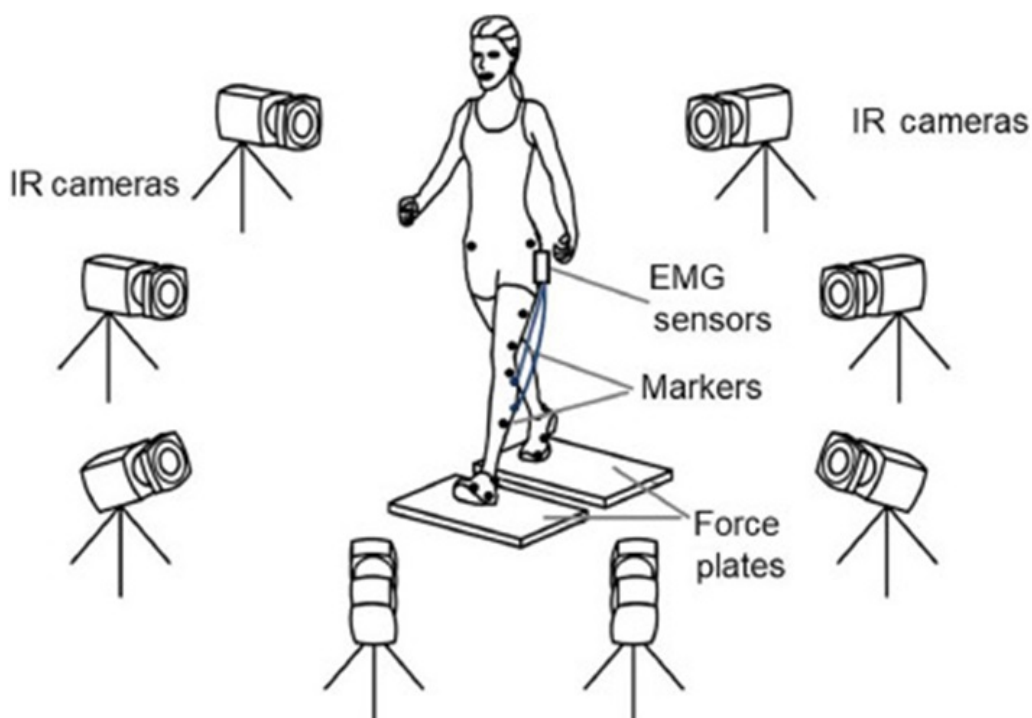
Slika 53, Postavka elektrogoniometara za merenje opsega pokreta ROM-a. Na levom panelu je prikazana neispravna postavka elektrogoniometara za merenje ROM-a u skočnom zglobu, a na desnom panelu je prikazana ispravna postavka za merenje ROM-a. Fotografije uslikane na Elektrotehničkom fakultetu, Univerziteta u Beogradu.

Na Sl. 53 prikazana je pogrešna i ispravna postavka dvoosnih elektrogoniometara SG110A¹⁶⁰ (Biometrics Ltd., Newport, UK) za merenje ROM u skočnom zglobu. Neispravna postavka je prikazana na levom panelu na Sl. 53 i ogleda se u neodgovarajućem načinu postavljanja senzora (pravac postavljanja ne prati ugao pod kojim se nalazi skočni zglob) i u tome što trake nisu postavljene na odgovarajući način (senzor je nedovoljno pričvršćen za nogu, tako da može da dođe do pomeranja senzora i pogrešnog merenja). Upravo suprotan primer je prikazan na desnom panelu na Sl. 53. Na obe slike je odeća ispitanika primerena uslovima merenja, odnosno nije preširoka jer u slučaju da je senzor postavljen preko širokih pantalona, došlo bi do pomeranja senzora i očitavanja koje ne odražava ugao u skočnom zglobu. Ista postavka senzora se može koristiti za merenje ROM tokom dorzifleksije¹⁶¹ ili plantarne fleksije¹⁶², ali i tokom kretanja kao što su hod i trčanje.

160 <https://www.biometricsltd.com/goniometer.htm>, pristupljeno 09.08.2023.

161 Predstavlja pokret odizanja prstiju stopala dok peta dodiruje podlogu. ROM za dorzifleksiju zavisi od toga pod kojim je uglom koleno u sedećem položaju, pa je potrebno standardizovati položaj ispitanika tokom merenja. [208]

162 Pokret podizanja pete od poda dok prsti stopala dodiruju podlogu. Kao i za dorzifleksiju, i za plantarnu fleksiju položaj tela (posebno ugao u kolenu tokom sedenja) može da utiče na merenje ROM-a. [208]



Slika 54, Skica biomehaničke laboratorije. Slika je preuzeta sa sajta <https://engcourses-uofa.ca/books/ortho/gait-analysis/> i licencirana pod CC BY NC SA 4.0 licencom. *IR cameras* (eng.) su IR kamere, *EMG sensors* (eng.) su EMG senzori koji služe za merenje električne aktivnosti mišića, *Markers* (eng.) su markeri koji se postavljaju na telo ispitanika, a čija se pozicija prati primenom IR kamera i *Force plates* (eng.) su podloge za merenje sile na kontaktu stopala sa podlogom.

Pored profesionalnih istraživačkih instrumenata za merenje ROM, u razvoju su i mobilne aplikacije za korišćenje akcelerometara koje su dostupne u mobilnim telefonima (primer aplikacije je Goniometer Pro App (G-pro¹⁶³), DigiflexLabs – više informacija o tome koji su senzori na mobilnim telefonima korišćeni i kako je signal transformisan u ugao nije dostupno). Pored merenja ROM, elektrogoniometri i drugi merni sistemi omogućavaju da se prate uglovi u zglobovima tokom pokreta. Elektrogoniometri se mogu koristiti i za kontrolisano izvršavanje pokreta sa ciljem ocene oporavka pacijenata nakon moždanog udara [209]. Nije samo ROM od značaja za rehabilitaciju, već i trajektorija i električna aktivnost mišića tokom kontrolisanih pokreta [209].

ROM se može meriti kao jedna diskretna vrednost koja najčešće označava opseg pokreta, mada bi po teoriji električnih merenja, a posebno sa stanovišta

163 https://play.google.com/store/apps/details?id=com.digiflexlabs.GoniometerPro&hl=en_US, pristupljeno 10.08.2023.

računanja merne nesigurnosti tipa A [78, 210] takvo merenje trebalo ponoviti više puta. Dodatni razlog je što maksimalni ugao zavisi od motivacije ili trenutnih okolnosti (uticaj bola kod pacijenata), slično kao kod korišćenja maksimalne voljne kontrakcije za normalizaciju amplitude EMG signala – što je detaljno prikazano u narednim poglavljima. Teorijski, ROM se može meriti u dužim vremenskim intervalima, te se može analizirati vremenska serija signala. Praktično, to nije korisno, jer ROM ne bi bio ponovljiv zbog nastanka zamora na mišićima. Osim u retkim slučajevima kada se ROM procenjuje iz merenog ugla u dužim vremenskih intervalima [211], snimanje ROM-a u dužim vremenskim intervalima nije uobičajeno.

Posebno opremljene laboratorije sa IR kamerama mogu postići značajno veću preciznost za merenje uglova u odnosu na goniometre, ali njihova mana je što je potrebno postaviti markere (pasivne ili aktivne) na ispitanika i što su pokreti ograničeni prostorno [206]. Na Sl. 54 prikazana je skica sistema sa IR kamerama, podlogama za merenje sile i EMG sensorima, dok je na Sl. 55 prikazana fotografija sistema za merenje primenom optičkog sistema koji se nalazi u Laboratoriji za informacione tehnologije (LaIT¹⁶⁴, slo. *Laboratorij za informacijske tehnologije*) na Elektrotehničkom fakultetu, Univerziteta u Ljubljani¹⁶⁵.

U slučaju merenja sa elektrogoniometrima koji pružaju značajno veću nezavisnost u pogledu kretanja u prostoru, za merenje kontakta pete sa podlogom sa ciljem određivanja faze hoda tj. faze zamaha (eng. *swing*) i faze oslonca (eng. *stance*) koriste se kontaktni senzori ili senzori pritiska (eng. *Force Sensing Resistor*)¹⁶⁶, dok se za merenje sa IR kamerama koriste platforme za merenje sile. FSR senzori nisu najpogodniji za merenja sile tj. pritiska kao jedinice po površini, ali mogu poslužiti kao pouzdani detektori kontakta (beleženje početka i kraja segmenta)¹⁶⁷ tj. određivanje da li postoji kontakt pete sa podlogom ili ne). Osim što FSR senzori nemaju dobru ponovljivost i poseduju histerezis [212], kada se postave u obuću ili na đon sa spoljne strane obuće kako bi se merio kontakt stopala sa podlogom podložni su različitim oštećenjima, naročito u slučaju hoda u kome se osoba oslanja na ivicu stopala ili kada hoda na prstima ili sl.

164 <http://www.lait.fe.uni-lj.si/>, pristupljeno 10.08.2023.

165 <https://fe.uni-lj.si/>, pristupljeno 10.08.2023.

166 Iako FSR skraćeniica označava da senzor meri silu, takav senzor u stvari meri pritisak kao silu po jedinici površine, te to treba imati na umu prilikom njegovog korišćenja u istraživanju.

167 Koje se na engleskom označava kao *ON/OFF* stanje.



Slika 55, Laboratorija za informacione tehnologije (LaIT) na Elektrotehničkom fakultetu, Univerziteta u Ljubljani – Qualisys¹⁶⁸ optički sistem kamera (eng. *motion capture* ili *mocap*) postavljen na zidovima i plafonu laboratorije. Fotografisano aprila 2024. godine. Uređaj je korišćen u nizu naučnoistraživačkih studija sa ciljem evaluacije drugih senzora [213-215], za biomehaničku procenu tenisera [216] i za druge primene.

U biomedicinskom inženjerstvu je proučavanje ljudskog pokreta od posebnog značaja, a ovde je prikazan samo jedan segment biomehanike. Kao i biomedicinsko inženjerstvo, biomehanika je interdisciplinarna oblast koja primenjuje mehaniku za rešavanje bioloških problema i koja se bavi proučavanjem, analizom i ocenom ljudskog pokreta [207]. Merenje i analiza uglova je samo jedan segment biomehanike. Za procenu uglova mogu se koristiti merenja sa akcelerometara ili merenja primenom IMU (eng. *Inertial Measurement Unit*) koje sadrže tri vrste senzora: (1) akcelerometre (za merenje ubrzanja), (2) žiroskope (za merenje ugaone brzine) i (3) magnetometre (za merenje pravca magnetnog polja). Inicijalno su se IMU sastojale iz akcelerometara i žiroskopa koji su tipično troosni, ali kasnije su dodati i magnetometri i sada su IMU sa sva tri senzora uobičajeni [217]. Indirektna merenja uglova primenom IMU bi trebalo uzeti sa rezervom, jer većina senzora zahteva kalibraciju i nije pogodna za primenu u svim uslovima merenja [207, 218-219]. Sistemi u kojima se ne radi ovakva transformacija, već se pokret posmatra u novom prostoru stanja i poredje različiti obrasci za zdrave ispitanike ili za pacijente, daleko je prirodniji pristup u analizi pokreta, a posebno u prepoznavanju aktivnosti [220-221]. Pod prostorom stanja misli se na opis

168 <https://www.qualisys.com/>, pristupljeno 28.04.2024.

pokreta pomoću veličina koja senzor meri: na primer, definisanje pokreta u prostoru stanja ubrzanja i ugaonih brzina.

```
> dat <- read.csv("PrimerPodatakaProsireno.csv")
> head(dat, 3)
  skocniZglob.koleno.pol.zdrav
1          98;108;M;da
2          87;115;F;da
3          73;110;F;ne
> dat <- read.csv("PrimerPodatakaProsireno.csv", sep = ";")
> head(dat, 3)
  skocniZglob koleno pol zdrav
1          98    108  M   da
2          87    115  F   da
3          73    110  F   ne
> class(dat$pol)
[1] "character"
> class(dat$zdrav)
[1] "character"
> dat$pol <- as.factor(dat$pol)
> dat$zdrav <- as.factor(dat$zdrav)
> class(dat$pol)
[1] "factor"
> class(dat$zdrav)
[1] "factor"
> |
```

Slika 56, Prikaz R koda u konzoli R Studio okruženja za učitavanje podataka iz datoteke “PrimerPodatakaProsireno.csv” i konverziju kolona *pol* i *zdrav* u kategoričke promenljive. Prikazano je kako se određuje ulazni parametar i koristi *head()* funkcija za proveru da li su podaci dobro učitani.

Za ilustraciju rada dplyr paketa za pripremu podataka za dalju obradu [38] sa ROM podacima, korišćeni su podaci u datoteci “PrimerPodatakaProsireno.csv”¹⁶⁹ koji predstavljaju simulirane podatke za ROM zdravih ispitanika i pacijenata (datoteka se nalazi u GitHub repozitorijumu <https://github.com/NadicaSm/TOBS>) koji sadrži sledeće kolone: (1) ROM u skočnom zglobu u stepenima (naziv kolone je *skocniZlob*), (2) ROM u kolenu u stepenima (naziv kolone je *koleno*), (3) *pol* ispitanika (naziv kolone je *pol*, a postoje dve vrednosti M (od eng. *Male* za muški *pol*) i F (od eng. *Female* za ženski *pol*) za muški i ženski *pol*) i (4) kolona u kojoj su dati

169 Radi se o proširenoj datoteci simuliranih ROM-a koji su pripremljeni za potrebe realizacije vežbi iz TOBS predmeta. Spisak svih vežbi i referenci ka njima dat je u Dodatku B na kraju udžbenika.

podaci da li se radi o zdravim ispitanicima ili pacijentima (naziv kolone je *zdrav*, a moguće vrednosti su *da* i *ne* za zdrave subjekte i pacijente). Na Sl. 56 prikazan je R kod za učitavanje ROM-a iz datoteke, kao i za konverziju podataka u kolonama *pol* i *zdrav* u kategoričke promenljive.

2.2.1.2 Primena dplyr paketa za pripremu podataka za dalju obradu

Za čišćenje ili pripremu podataka za dalju obradu pogodno je instalirati dplyr paket, ako nije instaliran na računaru (na Sl. 3 je prikazana komanda za instalaciju paketa, postoji i komanda koja pre primene funkcije za instalaciju proverava da li paket već postoji, a njena upotreba je prikazana na Sl. 23 u funkciji za odabir radnog direktorijuma) koji se koristi za rad sa *data frame* podacima [38]. Da bi se dplyr funkcije primenile na podacima moraju biti ispunjeni uslovi od kojih su sledeća dva najvažniji:

- postoji jedna opservacija po vrsti (ili je jedan ispitanik predstavljen u jednoj vrsti, a različiti podaci u kolonama ili je jedno merenje u jednom vremenskom intervalu¹⁷⁰ prikazano u vrsti, a različiti kanali ili biosignali po kolonama) i
- svaka kolona predstavlja neku mernu promenljivu.

Sa stanovišta biomedicinskih inženjera i načina kako inženjeri generalno posmatraju signale na računaru, ovo su potpuno nebitne pretpostavke. Međutim, u saradnji sa drugim strukama može doći do zabune prilikom zajedničkog rada na projektu analize signala. Naime, u statistici, psihologiji i medicini, ovi uslovi predstavljaju jedini način na koji se podaci predstavljaju tabelarno i prikaz na ovaj način je posebno važan ako je potrebno pregledati tabele vizuelno. Za biomedicinske inženjere potpuno je nevažno šta je u vrstama, a šta u kolonama¹⁷¹. Međutim, da bi uskladio prikaz podataka koji je uobičajen u drugim strukama i pravolinjski primenjivali korisni alati iz dplyr paketa i drugih paketa u R-u, trebalo bi usvojiti ove pretpostavke. Operator “\$” za indeksiranje tabelarnih podataka (pogledati Sl. 8 i Sl. 11 za korišćenje operatora) izdvaja elemente po kolonama, pa preporuke za smeštanje podataka u promenljivu *data frame* po ova dva uslova imaju praktičnog smisla.

170 Ovom vremenskom intervalu u opštem slučaju odgovara vremenski potpis događaja.

171 Operacija transponovanja matrica u R-u se realizuje pomoću funkcije *t()* iz osnovnog paketa, a postoje i specijalizovane funkcije za transponovanje elemenata liste, *data frame*-a i tabela podataka.

Svako ko poznaje osnove programiranja u bilo kom jeziku, vrlo brzo će sa upoznavanjem rada u dplyr paketu uočiti da dplyr ne unosi novu funkcionalnost u odnosu na osnovne funkcije koje postoje u R-u ili bilo kom drugom programskom jeziku. Međutim, postojeće funkcije su dodatno poboljšane i većina je razvijana u C++ programskom jeziku, te se brže izvršavaju [38]. Osnovne funkcije za rad sa dplyr paketom su:

- *select()* za odabir podskupa kolona,
- *filter()* za odabir podskupa vrsta na osnovu logičkog uslova,
- *arrange()* za promenu redosleda vrsta,
- *rename()* služi da se preimenuju promenljive,
- *mutate()* koristi se kada je potrebno da se dodaju nove kolone ili da se promene postojeće,
- *summarise()* za automatsko generisanje opisne statistike (eng. *descriptive statistics*) na skupu podataka (na primer, srednja vrednost i standardna devijacija).

Najčešće se koristi *filter()* funkcija u radu sa podacima. Treba voditi računa, jer funkcija pod istim imenom postoji i u osnovnom R paketu (eng. *base package*). Pojava istoimenih funkcija se ne može izbeći sa velikim brojem dostupnih paketa, te su postojeći načini da se naglasi da se funkcija poziva iz određenog paketa već pokazani u udžbeniku: (1) poziv *library()* funkcije neposredno pre primene odgovarajućeg paketa i preporučeni način (2) korišćenjem operatora "...".

U dplyr paketu postoji drugačija verzija osnovne:

print()

funkcije koja može sprečiti štampanje prevelike količine podataka u konzoli. Ređe se koristi, jer su u upotrebi *head()* i *tail()* funkcije koje se koriste za prikaz par vrsta u konzoli bilo da se broji od prve ili poslednje vrste u *data frame*-u (pogledati Sl. 12 i Sl. 22). U velikom broju slučajeva nema puno smisla da se štampaju nepregledne vrste podataka u konzoli. Za slučaj da je potrebno pregledati podatke u R-u, preporučuje se:

View()

funkcija koja ne pripada dplyr paketu, ali je korisna jer omogućava da se podaci prikažu tabelarno.

```

> library(dplyr)
> dat1 <- filter(dat, dat$skocniZglob < 65)
> class(dat1)
[1] "data.frame"
> dat1
  skocniZglob koleno pol zdrav
1          53     71  M    ne
2          60     91  M    ne
> dat2 <- dplyr::filter(dat, skocniZglob < 65)
> dat2 == dat1
      skocniZglob koleno  pol zdrav
[1,]          TRUE  TRUE TRUE  TRUE
[2,]          TRUE  TRUE TRUE  TRUE
> identical(dat1, dat2)
[1] TRUE
> |

```

Slika 57, Primena *filter()* funkcije za izdvajanje svih elemenata iz *data frame*-a koji je kreiran učitavanjem simuliranih podataka prikazan na Sl. 56 za izdvajanje svih opservacija koje imaju opseg pokreta u skočnom zglobu manji od 65° [222].

Funkcija *filter()* je slična *subset()* funkciji iz osnovnog R paketa, a moguće je napraviti podskup *data frame*-a čiji elementi ispunjavaju neki uslov (na primer, ostaviti samo one elemente čiji je ROM u određenim granicama). Unutar *filter()* funkcije se najčešće koriste jedan ili više logičkih uslova. Na Sl. 57 prikazano je kako se koristi *filter()* funkcija sa simuliranim¹⁷² ROM podacima u datoteci “PrimerPodatakaProsireno.csv”. Pokazano je kako je moguće identifikovati broj merenja opsega pokreta (ROM) u članku koji se nalaze ispod 65° što je usvojeno kao minimum u literaturi iako postoje individualne razlike [222].

Na Sl. 57 predstavljena su dva poziva *filter()* funkcije koja rezultuju *data frame*-ovima *dat1* i *dat2*. Prvi poziv je bez operatora dvotačka “::” i sa operatorom “\$” za kreiranje podskupa unutar *filter()* funkcije. Operator za podskup se ne mora koristiti u *dplyr* funkcijama, jer je prvi argument funkcije *data frame*, pa se u funkciji automatski naziv kolone pripisuju prvom argumentu. Da bi se utvrdila jednakost rezultujućih tabelarnih podataka *dat1* i *dat2* korišćen je operator jednakosti “==” koji je za rezultat vratio poređenje pojedinačnih elemenata, ali i funkcija:

172 Posebno poglavlje udžbenika je posvećeno simuliranim podacima u biomedicinskom inženjerstvu.

identical()

koja je kao rezultat vratila jednu logičku vrednost kao rezultat poređenja oba *data frame*-a.

```
> head(dat, 3)
  skocniZglob koleno pol zdrav
1          98    108  M    da
2          87    115  F    da
3          73    110  F    ne
> tail(select(dat, koleno:zdrav), 3)
  koleno pol zdrav
27    128  M    da
28    119  M    ne
29    125  M    da
> tail(select(dat, -(koleno:zdrav)), 3)
  skocniZglob
27          75
28          75
29          91
> tail(select(dat, skocniZglob, koleno), 3)
  skocniZglob koleno
27          75    128
28          75    119
29          91    125
> |
```

Slika 58, Primena *select()* funkcije za kreiranje novih *data frame*-ova za koje su prikazane poslednje tri opservacije primenom *tail()* funkcije. Primetiti primenu logičkih uslova za odabir kolona.

Sa funkcijom *select()* moguće je kreirati podskup *data frame*-a koji za rezultat daje novi *data frame* sa manjim brojem kolona. Kao argumenti koriste se nazivi kolona, a moguće je i isključiti podskup podataka. Primer kolona koje se isključuju su visina i težina ispitanika, ako se u daljoj analizi koristi samo indeks telesne mase (eng. *Body Mass Index*, skraćeno BMI)¹⁷³. Za detalje potrebno je pogledati uputstvo za funkciju, a posebno proučiti argumente

173 Ovo je uobičajen primer pripreme podataka za primenu metoda analize. BMI se dobija iz visine i težine ispitanika, te ako se sve tri kolone zadrže u podacima može doći do veoma visokih kros-korelacija među visinom i BMI-jem, što dalje može da utiče na ukupnu obradu signala – predikcije koje se dobijaju primenom algoritama mašinskog učenja mogu biti osetljive na kros-korelacije u skupu obeležja, jer su visoke kros-korelacije posledica redundantnosti obeležja [223].

ends_with i *starts_with*. Ilustracija primene ove funkcije na simuliranim podacima za izdvajanje pojedinih kolona u novi *data frame* je data na Sl. 58.

```
> head(dat, 3)
  skocniZglob koleno pol zdrav
1          98   108  M   da
2          87   115  F   da
3          73   110  F   ne
> head(arrange(dat, koleno), 3)
  skocniZglob koleno pol zdrav
1          53    71  M   ne
2          76    85  M   ne
3          78    88  M   ne
> tail(arrange(dat, koleno), 3)
  skocniZglob koleno pol zdrav
27          70   131  M   da
28          75   133  M   da
29          77   137  M   da
> head(arrange(dat, desc(koleno)), 3)
  skocniZglob koleno pol zdrav
1          77   137  M   da
2          75   133  M   da
3          70   131  M   da
> |
```

Slika 59, Uređivanje podataka po opsegu ugla u kolenu od najmanjeg do najvećeg ugla za podatke sa Sl. 56. Za uređivanje podataka po suprotnom redosledu (od najvećeg do najmanjeg) koristi se *desc()* funkcija iz *dplyr* paketa u kombinaciji sa *arrange()* funkcijom.

Funkcija *select()* se može iskoristiti za kreiranje novog *data frame*-a koji sadrži manji broj kolona. Međutim, u praksi se mogu i brisati kolone iz *data frame*-a. Za brisanje kolona u *data frame*-u nije neophodno koristiti *dplyr* biblioteku, jer je dovoljno odgovarajućoj koloni dodeli vrednost *NULL*.

Za uređivanje podataka po rastućem ili opadajućem redosledu, koristi se *arrange()* funkcija iz *dplyr* paketa. Treba voditi računa za primenu ove funkcije na podacima tipa vremenskih serija (npr. EMG, EEG signali) gde nije od posebnog značaja uređivanje podatka, a može da izazove greške u obradi signala. Za tabelarne biomedicinske podatke koji se mogu preurediti po polu, težini, visini i sličnim parametrima ima smisla primeniti ovu funkciju. Posebno je pogodna funkcija za tzv. ukrštanje podataka. Na primer, u psihologiji se uz

odgovarajuće šifre (ID od eng. *identity / identification*) ispitanika određuju obeležja iz biosignala, a u drugim datotekama se određuju rezultati psiholoških testova uz šifre ispitanika, te je najbolji način spajanja ove dve baze podataka u jednu primenom *arrange()* funkcije. Uređivanja podataka po opsegu ugla u kolenu je prikazano na Sl. 59. U simuliranom skupu podataka ugao u kolenu je u opsegu od 71° do 137°, što je u skladu sa podacima iz literature u kojima je pokazano da ugao u kolenu zavisi od niza činilaca (na primer, od godina ispitanika), ali da se ROM za koleno kreće od 130° do 150° [224].

```
> head(dat, 3)
  skocniZglob koleno pol zdrav
1          98    108  M   da
2          87    115  F   da
3          73    110  F   ne
> head(rename(dat, ugaoKoleno = koleno), 3)
  skocniZglob ugaoKoleno pol zdrav
1          98         108  M   da
2          87         115  F   da
3          73         110  F   ne
> head(mutate(dat, koleno = koleno - mean(koleno)), 3)
  skocniZglob      koleno pol zdrav
1          98 -6.0689655  M   da
2          87  0.9310345  F   da
3          73 -4.0689655  F   ne
> head(mutate(dat, koleno = koleno*pi/180), 3)
  skocniZglob      koleno pol zdrav
1          98  1.884956  M   da
2          87  2.007129  F   da
3          73  1.919862  F   ne
> |
```

Slika 60, Primena *rename()* i *mutate()* funkcija nad podacima čije je učitavanje prikazano na Sl. 56. U okviru *rename()* funkcije pokazano je kako je moguće koloni *koleno* promeniti ime u *ugaoKoleno*, a na primeru primene *mutate()* funkcije pokazano je kako je moguće centrirati promenljivu *koleno* i izvršiti transformaciju uglova u stepenima u uglove u radianima. Broj π je predstavljen kao *pi* u R-u.

Kako se operacije na *data frame* podacima primenjuju uz nazive kolona, to je poželjno da nazivi kolona imaju logična i opisna imena za šta se može koristiti *rename()* funkcija. Za slučaj da je potrebno transformisati neku promenljivu, na primer, centriranjem promenljive (oduzimanjem srednje vrednosti od pojedinačnih elemenata tj. transformacijom signala koja omogućava da srednja vrednost skupa bude jednaka nuli) koristi se promenljiva *mutate()*. Primeri korišćenja ove dve funkcije su prikazani na Sl. 60. Ovde neće biti prikazana, ali

u dplyr paketu pored *mutate()*, postoji i *transmutate()* funkcija pomoću koje se mogu iz niza promenljive koja je transformisana, izostaviti netransformisane promenljive.

```

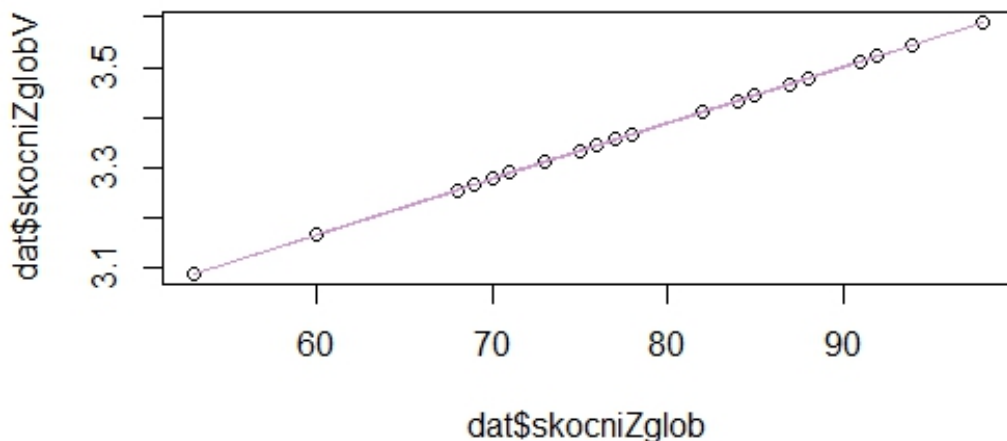
> dat$skocniZglobV <- dat$skocniZglob/90 + 2.5
> head(dat$skocniZglob, 3)
[1] 98 87 73
> head(dat$skocniZglobV, 3)
[1] 3.588889 3.466667 3.311111
> plot(dat$skocniZglob, dat$skocniZglobV)
> lines(dat$skocniZglob, dat$skocniZglobV, col = "plum3")
> A <- matrix(c(dat$skocniZglob[1], dat$skocniZglob[2],
+              1, 1), nrow = 2, ncol = 2)
> p <- matrix(c(dat$skocniZglobV[1], dat$skocniZglobV[2]),
+            nrow = 2, ncol = 1)
> A
      [,1] [,2]
[1,]   98    1
[2,]   87    1
> solve(A)
      [,1] [,2]
[1,] 0.09090909 -0.09090909
[2,] -7.90909091  8.90909091
> solve(A) %*% p # ili solve(A, p)
      [,1]
[1,] 0.01111111
[2,] 2.50000000
> |

```

Slika 61, Primer korišćenja linearnog modela za transformaciju jedinica sa elektrogoniometra za podatke čije je učitavanje prikazano na Sl. 56.

ROM podaci se predstavljaju u stepenima, ali se mere u V sa A/D konvertora, te ih je potrebno transformisati pre dalje analize. U ovoj fazi analize se podaci sa A/D konvertora transformišu tako da imaju odgovarajuće jedinice (u slučaju ROM to su stepeni). Ovde će ilustracije radi biti prikazano kako se od uglova u skočnom zglobu i u kolenu dobijaju naponi mereni na A/D konvertoru, ali u praksi je situacija obrnuta. Napon koji se meri je u opsegu od 0.5 V do 4.5 V, a ugao je u opsegu od -180° do 180° [225], te je potrebno realizovati jednostavnu linearnu transformaciju da se uglovi predstave kao napon meren na A/D konvertoru. Kako između napona i ugla u stepenima postoji linearna zavisnost (na osnovu uputstva proizvođača i poznavanja karakteristika senzora), to se ova transformacija može jednostavno izvršiti kao što je pokazano na Sl. 61.

Linearna zavisnost je u ovom slučaju ilustrovana na grafiku koji je prikazan na Sl. 62. Međutim, to nije uvek slučaj. Ako je odnos nelinearan, kao za merenje temperature termistorom [78], onda je moguće za radni opseg pretpostaviti linearni model.



Slika 62, Prikaz linearne zavisnosti ugla u skocnom zglobu (*dat\$skocniZglob*) u odnosu na napon koji se meri na A/D konvertoru (*dat\$skocniZglobV*).

Za predstavljanje linearnog modela dovoljno je izabrati dve tačke. Na Sl. 61 prikazan je kod u programskom jeziku R koji koristi funkciju:

```
solve()
```

za rešavanje linearne algebarske jednačine. Iako je na prvi pogled ovo trivijalno, linearni model je široko zastupljen u praksi, jer je jednostavan i nema smisla prelaziti na komplikovanije modele, ako se pokaže da je linearni model dovoljno dobar. U nauci trenutno preovlađuju određeni uticaji da je kompleksnije rešenje bolje rešenje. Dva su problema sa kompleksnim rešenjima: (1) teško ih je razumeti, objasniti i predvideti i (2) retko su isplativa (ili je potrebno uložiti računarsko ili ljudsko vreme, a sasvim je moguće da je jednostavan pristup najefikasniji).

Na primeru skupa podataka koji je predstavljen u ovom poglavlju ne postoje nedostajući podaci, ali se mogu uključiti, kako bi se ilustrovao rad sa nedostajućim podacima. Funkcije za računanje srednje vrednosti, standardne devijacije, opsega numeričkih vrednosti i druge se ne koriste sa podacima koji poseduju nedostajuće vrednosti (NA), jer je tada i rezultat NA. Da bi se *mean()*, *sd()*, *range()* ili slične funkcije primenile na podacima koji imaju NA, potrebno

je postaviti argument *na.rm* (eng. *remove NA*) TRUE. Primer računanja srednje vrednosti ROM-a za skočni zglob za pacijente i zdrave ispitanike je prikazan na Sl. 63. Takođe, na Sl. 63 prikazano je kako se računa broj NA što je važan korak pre primene obrade signala, jer procena broja nedostajućih vrednosti može da utiče na dalju analizu.

Na primer, može se zaključiti da poređenje razlika između pacijenata i zdravih nije moguće, jer je nakon izbacivanja NA vrednosti ostao samo podatak za jednog pacijenta, što nikako nije reprezentativni uzorak. Primititi da je korišćeno indeksiranje pomoću funkcije:

c()

čija je upotreba prikazana na Sl. 6, kako bi se za vrednosti opsega pokreta u skočnom zglobu na indeksima 1, 3, 4, 5, i 7 postavile nedostajuće vrednosti. Ako se nad takvim podacima računa srednja vrednost ROM-a u skočnom zglobu za zdrave ispitanike dobija se NA vrednost. Nakon postavljanja *na.rm* argumenta na vrednost TRUE, dobija se srednja vrednost za zdrave ispitanike od 82.57°, a za pacijente od 70.40°, što je očekivano jer smanjenje opsega može da ukaže na povredu nervnog sistema kod pacijenata. Za računanje broja nedostajućih vrednosti korišćena je kombinacija funkcija *sum()* i *is.na()* koje za rezultat daju tri nedostajuća podatka za zdrave ispitanike od ukupno 17 zdravih ispitanika i dva nedostajuća podatka za pacijente od ukupno 12 pacijenata (Sl. 63).

Manipulacija podacima kao deo pripreme za dalju obradu koja je prikazana na Sl. 61-63 nije uključila upotrebu funkcija iz *dplyr* paketa. Međutim, ovi koraci se mogu primeniti u kombinaciji sa funkcijama *dplyr* paketa da bi se podaci pripremili za obradu na odgovarajući način i da bi se procenio njihov sadržaj.

Ređe korišćenja funkcija iz *dplyr* paketa je *group_by()* koja služi da se na podskupu podataka primeni odgovarajuća funkcija (u primeru na Sl. 64 je dat primer korišćenja *group_by()* i *summarise()* funkcija). U praksi je najčešće jednostavnije formirati nove *data frame*-ove koji odgovaraju grupama podataka (na primer, jedan koji sadrži podatke o pacijentima i drugi koji sadrži podatke o zdravim ispitanicima). Bez obzira na alternativni pristup, ne treba izgubiti iz vida *group_by()* funkciju, jer takav pristup štedi memoriju na računaru (kreira se manji broj promenljivih).

```

> dat$skocniZglob[c(1, 3, 4, 5, 7)] <- NA
> mean(dat$skocniZglob[dat$zdrav == "da"])
[1] NA
> mean(dat$skocniZglob[dat$zdrav == "da"], na.rm = T)
[1] 82.57143
> mean(dat$skocniZglob[dat$zdrav == "ne"], na.rm = T)
[1] 70.4
> sum(is.na(dat$skocniZglob[dat$zdrav == "da"]))
[1] 3
> sum(dat$zdrav == "da")
[1] 17
> sum(is.na(dat$skocniZglob[dat$zdrav == "ne"]))
[1] 2
> sum(dat$zdrav == "ne")
[1] 12
> |

```

Slika 63, R kod za unos vrednosti ROM-a u skočnom zglobu na NA korišćenjem indeksiranja i primenom *c()* funkcije. Dodatno, pokazano je kako se može izračunati broj zdravih i pacijenata, kao i broj nedostajućih vrednosti ROM-a za skočni zglob za zdrave i pacijente.

```

> datNovo <- group_by(dat, zdrav)
> class(datNovo)
[1] "grouped_df" "tbl_df"      "tbl"        "data.frame"
> head(datNovo, 3)
# A tibble: 3 × 4
# Groups:   zdrav [2]
  skocniZglob koleno pol   zdrav
  <int> <int> <chr> <chr>
1      NA    108 M     da
2      87    115 F     da
3      NA    110 F     ne
> summarize(datNovo, skocniZglob = median(skocniZglob))
# A tibble: 2 × 2
  zdrav skocniZglob
  <chr> <int>
1 da      NA
2 ne      NA
> summarize(datNovo, skocniZglob = median(skocniZglob, na.rm = T))
# A tibble: 2 × 2
  zdrav skocniZglob
  <chr> <dbl>
1 da      83
2 ne      70
> |

```

Slika 64, Korišćenje *group_by()* i *summarize()* funkcije iz *dplyr* paketa za podatke čije je učitavanje prikazano na Sl. 56. Primititi da je jedan od ulaza *summarize()* funkcije funkcija *median()* čiji se argument *na.rm* može proslediti sa ciljem da se izračuna medijana podataka nad vrednostima koji nisu NA.

U poglavlju u kome je predstavljen rad sa nedostajućim vrednostima navedeno je zašto je medijanu korisno računati na podacima na kojima postoje vrednosti koje se značajno razlikuju od ostalih, a koje najverovatnije potiču od šuma. ROM koji se analizira u ovom poglavlju nema te osobine, ali je ilustracije radi na Sl. 64 prikazano kako se za rezultat funkcije *group_by()* računa medijana ROM-a u skočnom zglobu za pacijente i za zdrave ispitanike.

Rezultat *group_by()* funkcije nije *data frame*, već *tibble* koji je vrlo sličan *data frame*-u. Uobičajeno je korišćenje *tibble* tipa podataka u radu sa R paketima koji omogućavaju manipulaciju podacima i postoji poseban paket *tibble* na CRAN-u [226] koji sadrži funkcije za konverziju iz *data frame*-a u *tibble* podatke. Neke od odlika *tibble* podataka su [227]:

- mogu imati nestandardna imena (na primer: njihovo ime može sadržati razmak),
- automatski sprečavaju da se u konzoli odštampa velika količina podataka,
- nema delimičnog podudaranja imena kolone, te je određivanje podskupa striktnije nego kod *data frame* podataka i
- kolona može biti tip liste (što je odlično, jer u tom slučaju podaci koji su smešteni u kolonama mogu biti vremenske serije signala, te *tibble* u potpunosti odgovara potrebama u biomedicinskom inženjerstvu).

Osim što je korisno znati da podaci tipa *tibble* postoje i da se nad njim mogu izvoditi slične operacije kao nad *data frame* podacima, u ovom udžbeniku neće biti dalje reči o njima. Jedan od najvažnijih operatora koji dolazi sa *dplyr* paketom je *pipeline*¹⁷⁴ operator u oznaci “%>%” koji omogućava da se transformacija nad podacima realizuje unutar jedne sekvence. Upotreba *pipe* (eng. *pipeline*) operatora može doprineti uštedi memorijskog prostora, jer je moguće izbeći kreiranje novih promenljivih tj. međurezultata analize. Ne treba izgubiti iz vida da su za proveru rada realizovanih metoda od ključne važnosti međurezultati, posebno u fazi testiranja. Drugi način da se postigne funkcionalnost koju pruža *pipeline* operator jeste primenom ugnježenih funkcija. U slučaju *pipeline*-a, programski kod je elegantniji, razumljiviji i koristi manje argumenata. Ako je zapis ugnježenih funkcija $f_1()$, $f_2()$ i $f_3()$ nad argumentom x sledeći:

¹⁷⁴ Koliko je Autorki poznato nema odgovarajućeg prevoda za *pipeline* operator u računarskoj tehnici i informatici. U slobodnom prevodu sa engleskog to bi bio cevovod, dok termin formalno označava skup procesirajućih elemenata koji su povezani redno (izlaz prvog elementa je ulaz drugog itd.) [https://en.wikipedia.org/wiki/Pipeline_\(software\)](https://en.wikipedia.org/wiki/Pipeline_(software)), pristupljeno 16.08.2023].

$f_3(f_2(f_1(x)))$

onda se zapis primenom pipeline operatora svodi na:

$f_1(x) \%>\% f_2() \%>\% f_3()$.

Pipeline može započeti nazivom *data frame*-a na kome se vrši analiza, a rezultat *pipeline*-a se može dodeliti promenljivoj po želji. U slučaju da je potrebno nad *data frame*-om koji sadrži jednu kolonu primeniti jednostavne operacije, kao što je određivanje srednje vrednosti, to se na kraju *pipeline*-a koristi *unclass()* funkcija. Rezultat *pipeline*-a može biti *tibble*. Na Sl. 65 prikazana je upotreba *pipeline* operatora

```
> dat %>% group_by(zdrav) %>% summarize(median(skocniZglob, na.rm = T))
# A tibble: 2 × 2
  zdrav `median(skocniZglob, na.rm = T)`
  <chr>                                <dbl>
1 da                                     83
2 ne                                     70
> |
```

Slika 65, Prikaz korišćenja *pipeline* operatora za kod koji je prikazan na Sl. 64.

Važno je napomenuti da ovde nije detaljno prikazana *dplyr* gramatika, odnosno ovaj paket sadrži još funkcija koje se mogu koristiti za manipulaciju podacima [38]. Dodatno, za one koji žele samostalno da se dalje usavršavaju u ovoj oblasti preporučuje se *tidyverse*¹⁷⁵ kolekcija paketa koji uključuje za verziju *tidyverse* 1.3.0 sledeće pakete:

- *ggplot2* paket koji se primarno koristi za vizuelizaciju signala [228], a toliko je popularan da su u Pajtonu¹⁷⁶ i Matlabu¹⁷⁷ [229] razvijani paketi koji imaju sličnu funkcionalnost kao *ggplot2*,
- *dplyr* paket koji se koristi za manipulaciju podacima i pripremu podataka za dalju obradu na računaru [38],
- *tidyr* paket sadrži funkcije za uređivanje podataka u standardnu formu koju ima *data frame* ili *tibble* (obeležja su u kolonama, a pojedinačna merenja u vrstama) [230],
- *readr* paket koji sadrži dodatnu funkcionalnost za učitavanje (paket je prikazan u potpoglavlju Uvoz podataka u R) [111],

175 Skup paketa pod nazivom *tidyverse* se u R-u instalira pozivom *install.packages()* funkcije koja onda instalira automatski sve pakete sadržane u *tidyverse* kolekciji, <https://www.tidyverse.org/>, pristupljeno 16.08.2023.

176 *ggpy* paket za Pajton, <https://github.com/yhat/ggpy>, pristupljeno 16.08.2023.

177 *gramm* paket za Matlab, <https://github.com/piermorel/gramm>, pristupljeno 16.08.2023.

- purrr paket koji se zasniva na principima funkcionalnog programiranja¹⁷⁸, omogućava jednostavan rad sa funkcijama koji podrazumeva pisanje koda bez korišćenja kontrolnih struktura [232],
- tibble paket koji sadrži procedure za rad sa podacima tipa *tibble* koji je ranije prikazan u ovom poglavlju, a sadržaj *tibble* podatka je prikazan na Sl. 64 i Sl. 65 [226-227],
- stringr paket sadrži osnovne funkcije sa rad sa stringovima koje se uobičajeno sreću u pripremi podataka za dalju analizu [233] i
- forcats paket je sličan stringr paketu, ali je specijalizovan za rad sa kategoričkim promenljivima (eng. *factor variable*) [234].

Postoje i drugi paketi koji se moraju posebno aktivirati u R-u: pozivom *library(tidyverse)* aktiviraće se paketi koji su ovde nabrojani. Ovaj skup paketa su razvili iskusni naučnici i analitičari, među kojima je i profesor Hadli Vikam¹⁷⁹ sa Novog Zelanda koji je rukovodeći analitičar u R Studio kompaniji (sada Posit¹⁸⁰), a istovremeno i profesor na Univerzitetu u Oklandu¹⁸¹, Stanford Univerzitetu¹⁸² i Rajs Univerzitetu¹⁸³. Ono što se posebno vezuje za pakete koje je prof. Vikam razvio ili učestvovao u njihovom kreiranju jeste to što R paketi na intuitivan način pomažu da se primene dobre prakse u obradi signala i što je pisanje koda brže¹⁸⁴. Primera radi, ako neko dobije podatke koje treba da analizira i krene sa primenom dplyr paketa, funkcije koje postoje u dplyr paketu će ponuditi odlično rešenje i smernice za dobre prakse u obradi signala, ali i u računarstvu. Svi paketi imaju odlike funkcionalnog programiranja, te su kompaktne, smanjuju linije koda i u većini slučajeva dovode brže do rezultata.

Za praktičnu primenu, najbolje je proći kroz primere korišćenja ovih funkcija u R-u. U nastavku sledi par primera kako je moguće primenom funkcija iz dplyr paketa izvršiti manipulaciju ROM podataka i pripremiti ih za dalju analizu.

178 Funkcionalno programiranje je tip programiranja karakterističan za programski jezik R, ali i druge programske jezike. Nekada se naziva i primenjeno programiranje (eng. *functional/applicative programming*), a programski jezici koji imaju ove osobine funkcionalni jezici (eng. *functional languages*). Programski kod u tom slučaju može biti napisan brže, sažetije (sa manje linija koda) i da odgovaraju jezicima višeg nivoa (izrazi su sličniji prirodnim jezicima ili tekstovima sa matematičkim izrazima). Naravno, neke od osobina treba uzeti sa rezervom, jer je vrlo subjektivno šta znači brže pisanje programskog koda. [231]

179 <https://hadley.nz/>, pristupljeno 16.08.2023.

180 <https://posit.co/>, pristupljeno 16.08.2023.

181 University of Auckland, <https://www.stat.auckland.ac.nz>, pristupljeno 16.08.2023.

182 Stanford University, <https://datalab.stanford.edu/>, pristupljeno 16.08.2023.

183 Rice University, <https://statistics.rice.edu/>, pristupljeno 16.08.2023.

184 Na primer, daleko je jednostavnije pozvati funkciju *library(tidyverse)*, nego pojedinačno pozivati *library()* funkciju za sve pakete koji su sadržani u tidyverse kolekciji.

ROM za zdrave ispitanike u skočnom zglobu je od 65° do 75° [222], te je na Sl. 66 provereno koliko ispitanika ima ugao manji od 65° , a ROM u kolenu za zdrave ispitanike u u opsegu od 130° do 150° [224], te je provereno koliko ima ispitanika sa uglom manjim od 130° . Razlog zašto je ovo urađeno jeste da se proveriti autentičnost podataka. Ako se ROM merio na zdravima i na pacijentima, onda je očekivano da će ROM kod pacijenata biti manji od ROM-a kod zdravih ispitanika, kao i da neće postojati ROM manji od 0° . Rezultati prikazani na Sl. 66 pokazuju da ne postoje opsezi pokreta ni za skočni zglob ni za koleno koji su negativni, ali zabrinjavajuće je to što veliki broj ispitanika (25 njih od ukupno 29 ispitanika, od kojih su 12 pacijenti) ima ROM u kolenu manji od 130° što je manje od očekivanog opsega pokreta kod zdravih ispitanika [224]. Međutim, kada se pogledaju sve ROM vrednosti vidi se da postoji ROM u kolenu u okolini 130° . Kako ova granica nije donja granica zdravih osoba, već predstavlja tipičan ROM, to se može videti u poslednjoj liniji koda na Sl. 66 da, ako se granica promeni na 90% od 130° , broj ispitanika se skoro prepolovio (umesto 25 sada ih je 13).

Dalja priprema podataka uključuje procenu nedostajućih vrednosti koja je ilustrovana na Sl. 63. Ovde je prikazano samo kako jednostavna zamena nedostajućih vrednosti kao što su zamena medijanom¹⁸⁵ utiče na ukupan rezultat. Kako u datom skupu nema nedostajućih vrednosti one su na Sl. 67 dodate na odabranim mestima, odnosno na indeksima [1, 2, 3, 4, 19, 20, 21, 22].

Pre dodavanja nedostajućih vrednosti sa ciljem simulacije situacije u kojoj se one javljaju, predstavljene su medijane svih ROM vrednosti u kolenu (119°), kao i zasebno za zdrave ispitanike (125°) i za pacijente (103°). Međutim, kada se na 8 mesta (od ukupno 29 ispitanika, što čini 27.6% nedostajućih vrednosti) doda medijana koja je izračunata na celom skupu, podaci za zdrave ispitanike postaju isti kao za sve ispitanike (123°), ali se menjaju i za pacijente (sa 103° na 121°) što nije zanemarivo. U čemu je problem? Prvo, nedostaje relativno veliki broj podataka, iako je pokazano da u korišćenju odevnih tehnologija taj procenat nedostajućih podataka jeste uobičajena praksa [80]. Drugo, kada je određena medijana (promenljiva *kol* na Sl. 67) to je urađeno za ceo skup i nisu pojedinačno posmatrani zdravi ispitanici i pacijenti.

¹⁸⁵ Odabrana je zamena medijanom, jer bi za zamenu srednjom vrednošću trebalo proveriti da li podaci imaju Gausovu funkciju gustine verovatnoće ili ne [78]. U kasnijim poglavljima udžbenika je objašnjeno kako je moguće prikazati i analizirati podatke koji imaju i koji nemaju Gausovu funkciju gustine verovatnoće.

```

> dat <- read.csv("PrimerPodatakaProsireno.csv", sep = ";")
> sum(dat$zdrav == "da")
[1] 17
> sum(dat$zdrav == "ne")
[1] 12
> sum(dat$skocniZglob < 0)
[1] 0
> sum(dat$skocniZglob < 65)
[1] 2
> sum(dat$koleno < 0)
[1] 0
> sum(dat$koleno < 130)
[1] 25
> dat$koleno
[1] 108 115 110 120 124 103 126 130 115 101 126 123 96 124 133 85 126 71 93
[20] 91 131 137 119 127 88 114 128 119 125
> sum(dat$koleno < 130*0.9)
[1] 13
> |

```

Slika 66, Procena sadržaja ROM podataka korišćenjem *sum()* funkcije i logičkog indeksiranja.

```

> dat <- read.csv("PrimerPodatakaProsireno.csv", sep = ";")
> median(dat$koleno)
[1] 119
> median(dat$koleno[dat$zdrav == "da"])
[1] 125
> median(dat$koleno[dat$zdrav == "ne"])
[1] 103
> dat$koleno[c(1, 2, 3, 4, 19, 20, 21, 22)] <- NA
> dat$koleno[c(1:4, 19:22)] <- NA
> kol <- median(dat$koleno, na.rm = TRUE)
> kol
[1] 123
> dat$koleno[c(1:4, 19:22)] <- kol
> median(dat$koleno)
[1] 123
> median(dat$koleno[dat$zdrav == "da"])
[1] 123
> median(dat$koleno[dat$zdrav == "ne"])
[1] 121
> |

```

Slika 67, Manipulacija nedostajućim vrednostima za ROM u kolenu kod zdravih ispitanika i pacijenata. Primetiti da su u konzoli ilustrovana dva načina indeksiranja *data frame*-a od kojih jedan uključuje operator ":" koji može da radi sa celim brojevima u R-u.

Ako se procedura zamene nedostajućih vrednosti ponovi tako što će medijana ROM-a u kolenu za zdrave ispitanike da bude iskorišćena da se popune nedostajuće vrednosti za zdrave ispitanike, a isto to se uradi za pacijente,

rezultati su značajno bolji i ne odstupaju previše od inicijalnih srednjih vrednosti (Sl. 68). U ovom primeru zamene nedostajućih vrednosti bilo je jednostavno da se donese ovakva odluka, jer jedini parametar koji bi mogao da utiče na vrednost ROM-a se nalazi u koloni *zdrav*. Međutim, ako postoji više parametara koji mogu uticati na vrednost promenljive koja nedostaje, situacija nije jednostavna, a najkompleksniji slučaj je kada zavisnost između parametara nije poznata. Ovo je posebno važno za naučna istraživanja koja se ne bi ni izvodila da je rezultat unapred poznat ili da su parametri i njihova zavisnost unapred poznati.

```
> dat <- read.csv("PrimerPodatakaProsireno.csv", sep = ";")
> dat$koleno[c(1:4, 19:22)] <- NA
> kolZ <- dat %>% filter(zdrav == "da") %>% select(koleno) %>%
+   unlist() %>% median(na.rm = TRUE)
> kolZ
[1] 125
> kolP <- dat %>% filter(zdrav == "ne") %>% select(koleno) %>%
+   unlist() %>% median(na.rm = TRUE)
> kolP
[1] 105
> dat$koleno[is.na(dat$koleno) == TRUE & dat$zdrav == "da"] <- kolZ
> dat$koleno[is.na(dat$koleno) == TRUE & dat$zdrav == "ne"] <- kolP
> median(dat$koleno)
[1] 123
> median(dat$koleno[dat$zdrav == "da"])
[1] 125
> median(dat$koleno[dat$zdrav == "ne"])
[1] 105
> |
```

Slika 68, Ispravan način da se realizuje zamena nedostajućih vrednosti za slučaj ilustrovan na Sl. 67. U ovom primeru su korišćene funkcije dplyr paketa što je rezultovalo svedenijim i jednostavnijim kodom.

U literaturi na engleskom jeziku mogu se sresti termini kao što su *data wrangling* ili *data munging* za manipulaciju podacima. Oba termina obuhvataju procedure za procenu kvaliteta podataka i odgovarajuću modifikaciju kako bi se podaci mogli koristiti za dalju analizu. Iako na prvi pogled ovaj korak može delovati beznačajno, trebalo bi napomenuti da analitičari i istraživači koji analiziraju podatke, najveći deo svoga vremena provedu u manipulaciji, odnosno pripremi podataka za dalju analizu. [235]

2.2.1.3 Drugi primeri pripreme podataka za dalju obradu

Do sada su na više mesta spominjane odevne tehnologije (eng. *wearables*). Njihova je primena naročito privlačna, posebno sa stanovišta realizacije merenja, jer su neinvazivna i relativno nenametljiva. Obrada signala merenih odevnim tehnologijama ume da bude problematična, posebno sa stanovišta kvaliteta signala. U fazi pripreme podataka za obradu signala, kvalitet je jedan od važnih parametara koji se ispituje, ne samo za odevne tehnologije, već i kod signala koji su relativno manjih amplituda. Na primer, nije uobičajeno ispitivanje kvaliteta EKG ili EMG signala, ali EHG, EGG i fetalni EKG su veoma podložni šumovima, te im je neretko potrebno proceniti kvalitet [79, 236-237]. Kvalitet signala zavisi od merne metode i mernog uređaja. Iako se kvalitetom može manipulirati na računaru (na primer, filtriranjem šuma moguće je poboljšati kvalitet signala), nije moguće dodati nepostojeću informaciju signalu na računaru. Ako informacija nije izmerena, ona se ne može i ne sme izmisliti, odnosno naknadno dodati signalu.

Ovaj udžbenik se ne bavi merenjem biosignala, ali pojmovi invazivnost i nenametljivost (eng. *unobtrusive*) mogu biti pogrešno protumačeni, pa će ovde biti objašnjeni. Procedura je neinvazivna ako ne postoji prolazak kroz kožu (na primer, kod merenja EMG signala primenom iglenih elektroda postoji prolazak kroz kožu) i ako ne postoji fizički ulazak u telo (ugradnja pejsmejкера zahteva fizički ulazak uređaja kroz telo), te prema tome rentgensko zračenje i EKG nisu invazivne procedure prema MedlinePlus medicinskoj enciklopediji Nacionalne biblioteke medicine u SAD [238]. Sa druge strane, u ovoj enciklopediji prihvaćena je definicija invazivne procedure u kojoj postoji prolazak kroz telo (eng. *the body is invaded*) ili je došlo do ulaska igle, cevke, uređaja ili sonde u telo [239]. Striktno govoreći, nije potrebno da dođe do probijanja kože koje je takvo da će izazvati mehaničku povredu tkiva da bi merenje bilo invazivno, jer ultrazvuk može da prolazi kroz kožu i unutrašnje organe. To je ujedno osnovni princip rada ultrazvučne dijagnostike [240]. Prema ovom rigoroznom shvatanju invazivne procedure, merenje električne provodnosti kože (EDA) je procedura koja je invazivna jer struja relativno malih amplituda (reda mA) prolazi kroz kožu kako bi se izmerila promena provodnosti, dok merenje EKG signala nije jer se nikakva električna (u opštem slučaju fizička) energija se ne isporučuje telu. Nenametljivost mernog sistema podrazumeva da merni sistem ne unosi neprijatnosti tokom obavljanja svakodnevnih aktivnosti i da je metoda prihvatljiva od strane ispitanika [241]. Beskontaktna merenje su nenametljiva jer se nikakav senzor ne postavlja na ispitanika. Video snimak lica ispitanika sa kamere može da se obradi tako da se ili posmatra promena crvene boje u video signalu [242-243] ili da se primene takve metode da se detektuju pokreti koji

nisu vidljivi ljudskom oku [244] da bi se merio puls ispitanika. Ovo je u potpunosti beskontaktno i neinvazivno merenje, ali ako se zahteva da ispitanik sve vreme bude u kadru kamere, onda može biti nametljivo. Za ispitanike koji tokom merenja rade na računaru ili koji voze automobile (ili simulatore vožnje i letenja), ovo nije nametljivo, ali za one koji obavljaju svoje aktivnosti napolju (odlazak u šetnju, kupovinu ili sl), onda jeste nametljivo merenje. Odevne tehnologije su kontaktne, uglavnom neinvazivne (već je objašnjeno zašto je EDA invazivna) i pretežno nenametljive¹⁸⁶. Ovo su sve dobre osobine za uređaj koji meri biosignale, ali da bi se dobilo na udobnosti ispitanika, gubi se na kvalitetu. Pod kvalitetnim merenjem se podrazumeva merenje u kome nema izgubljenih podataka i na koje šum nije uticao u značajnoj meri te informacija od interesa jeste sačuvana. Posebno poglavlje se bavi eliminacijom šuma u ovom udžbeniku, zato što to jeste jedan od najvažnijih koraka u obradi biosignala. Međutim, u pripremi podataka za dalju obradu, nekada je potrebno proceniti šum. Kvalitativne procedure podrazumevaju prikaz grafika i vizuelnu inspekciju prisustva šuma [79], dok kvantitativne uključuju računanje parametara koji se mogu koristiti za njegovu ocenu [237, 245].

Vizuelizaciji signala je posvećeno posebno poglavlje u ovom udžbeniku, te ovde neće biti dalje reči o ovoj vrsti procene kvaliteta signala. Najjednostavniji način da se kvalitet signala proceni kvantitativno je da se izračuna odnos signal-šum (eng. *signal to noise ratio*, skraćeno SNR). Ali, kako izmeriti šum? Primeri merenja šuma su već dati na početku potpoglavlja o pretprocesiranju biosignala za eliminaciju šuma gde je navedeno da se šum može izmeriti tako što se u Faradejevom kavezu za merenje potencijala paradajza koristi dodatni kanal [190] ili kada se tokom merenja EEG signala meri i EOG [194]. Međutim, bez obzira što je dodatnim merenjem priroda šuma poznata i dalje nije poznat njen udeo u samom signalu koji se može menjati. Zato se koriste različite metode procene SNR od kojih neke uključuju izdvajanje šuma i korisnog signala u frekvencijskom opsegu [237]. Da bi se procenio SNR primenjuje se formula data u Relaciji (1) gde je sa \widehat{SNR} označena procena SNR¹⁸⁷, a sa $P(s)$ snaga signala i sa $P(n)$ snaga šuma.

186 Ideja odevnih tehnologija jeste da senzori budu smešteni u odeći tako da nema potrebe za dodatnim mernim sistemima. U tom smislu, postojanje senzora ne bi trebalo da ometa svakodnevne aktivnosti ili izaziva neprijatnost ispitanicima.

187 Za merenje odnosa amplituda izraz za procenu SNR se množi sa 20, umesto sa 10 [Lecture 8, Digital Communication Systems, MIT, https://ocw.mit.edu/courses/6-02-introduction-to-eecs-ii-digital-communication-systems-fall-2012/20e59def7c26c7e1bf48369cec8d6810_MIT6_02F12 lec08.pdf, pristupljeno 12.04.2024].

$$\widehat{SNR} = 10 \log_{10} \frac{P(s)}{P(n)} \quad \text{Relacija (1)}$$

SNR se izražava u decibelima (dB): veće vrednosti označavaju da je šum je manje izražen (što je poželjna situacija) i obrnuto. Na žalost, ove se vrednosti retko mogu direktno proceniti i najčešće se procenjuje udeo šuma na nekom frekvencijskom opsegu u odnosu na opseg gde se očekuje signal od interesa. Logično, ovakav pristup nije idealan, jer se u frekvencijskom domenu frekvencijski opsezi na kome se nalaze šum i signal mogu preklapati.

Postoji niz paketa u R-u koji se bave analizom šuma i njegovom eliminacijom. Ovde će biti predstavljen u kratkim crtama specijalizovan R paket za rad sa Empatica E4 narukvicom (Empatica Inc., Cambridge MA, USA) koja je primer odevnog uređaja [245]. Biosignali mereni odevnim tehnologijama omogućavaju kontinualni monitoring i realizaciju biološke povratne sprege (biofeedback od eng. *Biofeedback*) [245]. Odevne tehnologije se, između ostalog, mogu koristiti za digitalnu fenotipizaciju¹⁸⁸ (eng. *digital phenotyping*), koja podrazumeva korišćenje digitalnih uređaja za merenje individualnih karakteristika u različitim vremenskim intervalima [245-246]. Termin digitalna fenotipizacija je prvi put uveden 2016. godine, u radu u kome je predstavljeno kako se mobilni telefoni mogu koristiti za dobijanje korisnih informacija u klinici i istraživanju, a koji su napisali Toros i saradnici [247]. Drugi termin za digitalnu fenotipizaciju je individualno merenje biosignala (eng. *personal sensing*), a pored upotrebe odevnih tehnologija podrazumeva i merenje mobilnim telefonima ili drugim prenosivim (eng. *portable*) uređajima za primenu u psihijatriji za prevenciju, ranu dijagnostiku, nadgledanje i terapiju [246]. Prednost koju su omogućile odevne tehnologije se najviše ogleda u tome što se merenja iz laboratorija sele u okruženja u realnom životu [245]. Široko zastupljen je primer pedometra (uređaja za merenje broja koraka) primenom mobilnih telefona (tzv. pametnih¹⁸⁹

188 Da li je ovaj termin samo prolazna moda ili oblast koja će ostati duže zastupljena u biomedicinskom inženjerstvu, ostaje da se vidi. Autorka nema jasan stav povodom ovog termina, osim što može naglasiti činjenicu da je prirodno da kako se oblasti razvijaju (a oblast merenja biosignala primenom odevnih tehnologija i pametnim telefonima je tek u povoju, posebno u oblasti kliničke primene) onda je logično da se uvodi i nova terminologija. U Onela laboratoriji na Harvard Univerzitetu u SAD glavni pravac istraživanja je posvećen digitalnoj fenotipizaciji, te ako je ova fenotipizacija dovoljna da se zbog nje formira cela laboratorija, pretpostavka je da će termin duže ostati prisutan u biomedicinskom inženjerstvu (<https://www.hsph.harvard.edu/onnella-lab/research/>, pristupljeno 19.08.2023).

189 U inženjerstvu preovlađuje naklonjenost ka pripisivanju ljudskih osobina predmetima, uređajima i softverskim metodama. Pametni telefoni, pametni senzori, pametne kuće i veštačka inteligencija su samo neki od primera. Iako naširoko zastupljeni, ovi termini ni

mobilnih telefona) koji je primarno namenjen upotrebi van stambenog prostora i laboratorijskih uslova merenja. Međutim, čak i pedometar može biti validan klinički parametar (na primer, koliko je osoba aktivna tokom dana, sedmice, meseca). Biosignali mereni primenom odevnih tehnologija mogu se koristiti u medicini za dijagnostiku, nadgledanje (eng. *monitoring*) bioloških markera i za ocenu terapije [245]. Na žalost, merenja sa ovih uređaja nisu standardizovana (ne postoje formalna uputstva za njihovu obradu), te predstavljaju veliki izazov u praksi. Dodatno, za razliku od merenja u kontrolisanim uslovima, bilo u laboratorijama ili u kliničkom okruženju, ova merenja su podložna različitim uticajima, posebno velikom prisustvu šuma [245]. Pored ovih inženjerskih problema, postoje i druga pitanja/problemi u radu sa odevnim tehnologijama [245-246]:

- Problem privatnosti – podaci su najčešće smešteni u oblaku, odnosno na nečijem računaru. Ko ima pristup podacima? Da li su podaci dovoljno zaštićeni (da li postoje ovlašćenja za pristup podacima i da li su preduzete bezbednosne mere da se spreči neovlašćeno korišćenje)? Ko je vlasnik tih podataka i kakva su prava korisnika? Da li su podaci anonimizovani?
- Problem netransparentnosti – veliki broj odevnih tehnologija ima vlasnički (eng. *proprietary*), odnosno tajni način analize signala, kako bi se dobio konačan rezultat koji je najčešće u obliku kvantitativnog parametra koji nije pogodan za medicinsku upotrebu. Ovo nije kritika kompanija koje proizvode takve uređaje jer oni mogu imati svojih razloga (kao što su poslovna tajna ili patenti), ali time se dodatno ograničava upotrebljivost biosignala u kliničkim istraživanjima i medicinskoj praksi. Osim toga, komercijalni uspeh je moguć i uz otvorene tehnologije [249], ali kada su ugovori jednom potpisani, ovakvu praksu nije jednostavno promeniti unutar kompanije.

Bez obzira na to što nacionalni i međunarodni zakoni regulišu većim delom problem privatnosti, ne treba ispustiti iz vida činjenicu da su medicinski podaci najosetljiviji lični podaci. U Republici Srbiji Zakon o zaštiti podataka o ličnosti (ZZPL “Sl. glasnik RS”, br. 87/2018¹⁹⁰) određuje pravni okvir za rad sa podacima o ličnosti kojih se moraju pridržavati svi koji rade sa takvim

približno ne odgovaraju sistemima koje opisuju, jer niti telefon može biti pametan (a ni glup), niti postoji veštačka inteligencija. Studenti Elektrotehničkog fakulteta, Univerziteta u Beogradu bi trebalo da su se već susreli sa ovim terminima [78, 248] i njima bi trebalo da je u potpunosti jasno koji sistemi se opisuju na ovaj način, ali čitaocima koji nemaju formalno inženjersko obrazovanje ovo može zvučati senzacionalno ili pak banalno, pa je dodatno pojašnjenje u obliku fusnote 189 bilo neophodno.

podacima, dok je u Evropskoj uniji to Opšta regulativa za zaštitu podataka o ličnosti Evropske unije (GDPR od eng. *General Data Protection Regulation*¹⁹¹). Sa druge strane, transparentnost je osnovni princip na kome se zasnivaju principi otvorene nauke prema UNESCO definiciji [45]. Transparentnost ne uključuje samo dostupnost metoda koje se koriste, već i podataka. Na izgled ovo su dva oprečna zahteva. Kako otvoriti podatke i učiniti ih dostupnim na internetu, a istovremeno iz zaštititi od zloupotrebe? Rešenje je u anonimizaciji podataka. Anonimizacija se radi sa ciljem da se onemogući direktan i/ili indirektan način dolaska do identiteta osobe čiji su podaci anonimizovani. Logičan način anonimizacije uključuje brisanje ličnih podataka (JMBG¹⁹² broja, imena i prezimena, adrese stanovanja, broja medicinskog dosijea/kartona i sl). Međutim, neki lični podaci nemaju tako očigledno lični karakter. Primera radi, ako se realizuje anonimna anketa koja se prosledi svim zaposlenima na Elektrotehničkom fakultetu, Univerziteta u Beogradu i ako se postavi pitanje datuma rođenja, velika je šansa da će identiteti svih zaposlenih koji su popunili anketu biti poznata. Ako se reši da se to pitanje svede samo na godinu rođenja, ali se postavi pitanje o pripadnosti katedri, polu i zvanju, opet je godina rođenja identifikacioni parametar. Dakle, podaci se ne posmatraju pojedinačno, već se gleda njihova međusobna povezanost sa ostalim podacima u nekoj studiji. Jedan od načina da se ovo reši jeste da ispitanici zaokruže ponuđene odgovore za opseg godina (20-25, 25-30, 30-35, itd). Druga varijanta je da se istraživači obavežu da čuvaju podatke koje su prikupili, ali da ih anonimizuju, odnosno smeste u kategorije, ako ih objavljuju. Ovo je samo jedna od tehnika anonimizacije. Ima i drugih. Autorki nije poznato da li u programskom jeziku R postoji poseban paket koji vrši anonimizaciju podataka, ali se takva procedura može realizovati korišćenjem paketa koji su već prikazani u ovom udžbeniku. Za slučaj da se studenti i čitaoci opredele da koriste dostupan slobodan softver, preporuka je da koriste Amnesia¹⁹³ alat (Institut za upravljanje informacionim sistemima, istraživački centar "Atina"¹⁹⁴, Marusi, Grčka) koji je razvijen u okviru OpenAIRE¹⁹⁵ projekta koji je finansirala Evropska unija¹⁹⁶. Ova pitanja

190 https://www.paragraf.rs/propisi/zakon_o_zastiti_podataka_o_licnosti.html, pristupljeno 19.08.2023.

191 <https://gdpr-info.eu/>, pristupljeno 19.08.2023.

192 JMBG se primenjuje u Republici Srbiji i označava jedinstven matični broj građana. U drugim zemljama postoje drugi i/ili slični identifikatori.

193 <https://amnesia.openaire.eu/>, pristupljeno 19.08.2023.

194 Institute for the Management of Information Systems, Research Center "Athena", <https://www.imsi.athenarc.gr/en>, pristupljeno 23.02.2024.

195 <https://www.openaire.eu/>, pristupljeno 19.08.2023.

196 Za više detalja preporučuje se članak Poverljivost sa Serbia.RMD sajta (<https://rdm.open.ac.rs/index.php/13-deljenje/25-personal-data>, pristupljeno 19.08.2023).

Očekuje se da sadržaj sa ovog sajta bude preseljen na internet stranicu Zajednice otvorene

neće biti dalje razmatrana u ovom udžbeniku, ali važno je da studenti razumeju da postoje podaci koje ne treba deliti (ili se radi o poverljivim podacima kao što su vojni ili medicinski podaci ili su to podaci koje nije moguće jednostavno anonimizovati¹⁹⁷). Upravo iz ovog razloga, Evropska komisija je u uputstvu za FAIR podatke spremljenom za H2020 projektni program (Horizont 2020), donela preporuku da je potrebno da “podaci budu otvoreni koliko god je moguće i zatvoreni koliko god je potrebno”¹⁹⁸ [250]. Postoji niz prednosti deljenja otvorenih i uređenih (FAIR) podataka, ali kreatori H2020 uputstva su svesni da nije sve podatke moguće, a ni bezbedno deliti. Biosignali mereni primenom odevnih tehnologija za sada uključuju podatke koji se mogu relativno jednostavno anonimizovati, a wearables paket ni ne uključuje metode anonimizacije [245].

Prednost wearables¹⁹⁹ paketa [245] je što ga mogu koristiti i stručnjaci sa odgovarajućim znanjem iz oblasti obrade signala, ali i lekari i medicinski tehničari za pripremu za obradu i dalju obradu biosignala sa Empatica E4 senzorske narukvice²⁰⁰. Empatica je čest izbor u naučnoistraživačkim studijama²⁰¹, jer omogućava preuzimanje sirovih (eng. *raw*) podataka, koji su od posebnog značaja za istraživače, jer njima nisu potrebni rezultati algoritama koji nisu poznati, odnosno koji predstavljaju tzv. crne kutije ili na eng. *black box* modele. Isprva, R nije bio realizovan za potrebe obrade vremenskih serija biosignala, ali sa pojavom specijalizovanih paketa, sada je sve veća primena R-a u obradi signala, uključujući obradu signala merenih primenom odevnih tehnologija [99-100, 245, 251].

nauke Srbije (<https://oscs.open.ac.rs/>, pristupljeno 19.08.2023).

197 Jedan od primera je deidentifikacija slika mozga sa magnetne rezonance (eng. *Medical Resonance Imaging*, skraćeno MRI) ili CT skenera (eng. *Computer Tomography*), jer je ovakvu medicinsku sliku moguće koristiti za identifikaciju ispitanika. Trenutno su različiti softverski alati u razvoju za deidentifikaciju slika glave/mozga [265], ali tu postoje dva problema: (1) potrebno je omogućiti povratak na originalnu sliku ili omogućiti zaštićeno čuvanje originalnih medicinskih slika i (2) potrebno je sprečiti da metode za deidentifikaciju, odnosno za anonimizaciju dovedu do smanjenja informacije koja ima dijagnostički značaj.

198 Ovo je slobodan prevod sa engleskog na srpski jezik. U originalu, preporuka je “*As open as possible, as closed as necessary.*” [250].

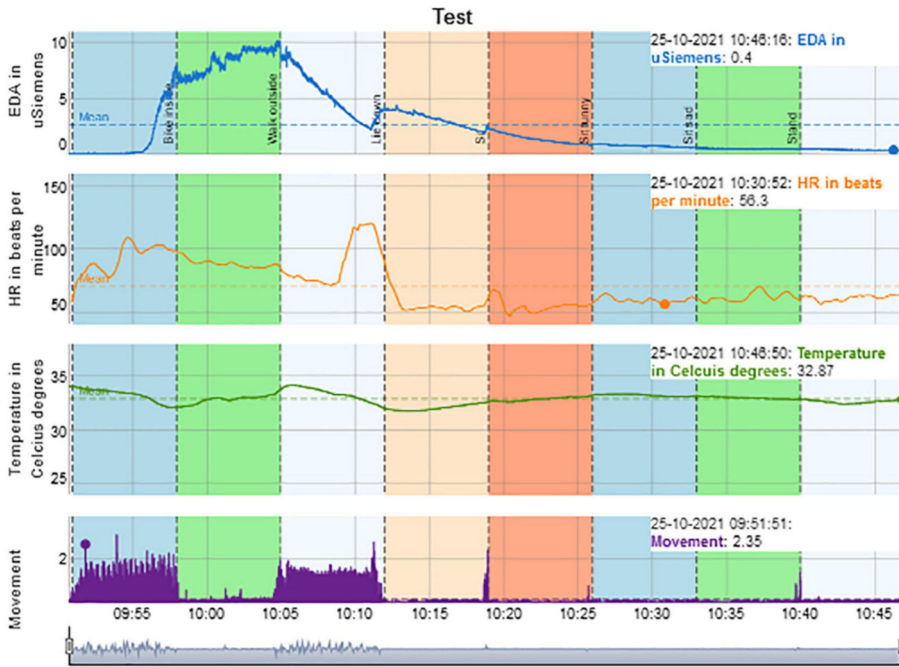
199 <https://cran.r-project.org/package=wearables>, dostupno na CRAN-u i pristupljeno 19.08.2023.

200 Autori wearables paketa tvrde da je moguće proširiti i relativno jednostavno prilagoditi ovaj paket kako bi mogao da se koristi sa drugim odevnim uređajima.

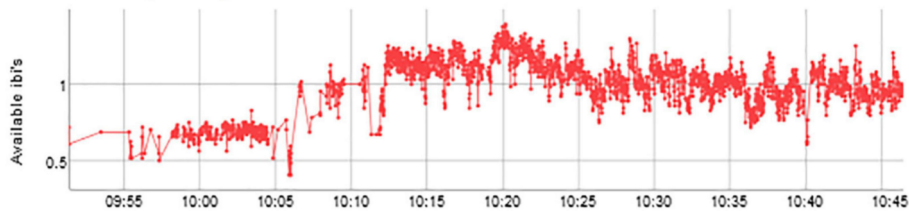
201 Na žalost, Empatica kompanija je najavila da Empatica E4 narukvica, koja je dominantno korišćena u naučnoistraživačkim studijama, neće više biti podržana, <https://support.empatica.com/hc/en-us/articles/6093302244637-E4-Sunsetting-FAQs> (pristupljeno 30.05.2024).

Dynamic report

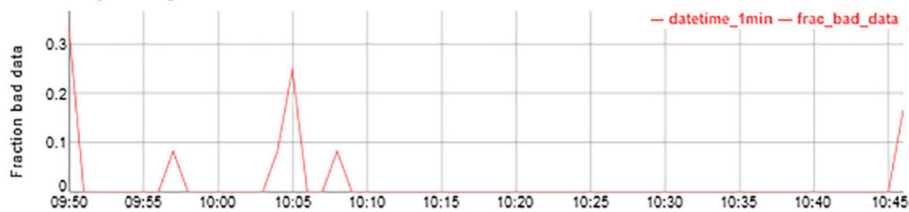
Timeseries plot



IBI Data quality



Data quality



Slika 69, Primer rezultata primene metoda za manipulaciju i vizuelizaciju biosignalima sa Empatica E4 senzora u wearables R paketu [245]. Slika je preuzeta iz [245] i licencirana pod međunarodnom CC BY 4.0 licencom, <https://creativecommons.org/licenses/by/4.0/>, pristupljeno 19.08.2023.

Pored povećanja i smanjenja frekvencije odabiranja (eng. *up sampling* i *down sampling*), biosignali mereni odevnim tehnologijama mogu uključiti i različite metode za detekciju prisustva šuma i procene kvaliteta merenja, sinhronizaciju događaja na signalu sa vremenskim potpisom događaja ili uslovima merenja, kao i obaveznu anonimizaciju. Nekada se za naizgled vrlo jednostavan korak, kao što je detekcija šuma (koji jeste moguće vizuelno i ručno odrediti) koriste vrlo sofisticirane metode kao što su algoritmi mašinskog učenja. [79, 237, 245]

U ovom udžbeniku neće biti prikazan proces promene frekvencije odabiranja iako je uključen u paket wearables [245]. Generalno, za biosignale dobijene primenom odevnih tehnologija, najčešće se koristi promena frekvencije odabiranja u procesu pripreme podataka za dalju obradu signala. Proces smanjenja frekvencije odabiranja se zove decimacija, a proces povećanja frekvencije odabiranja interpolacija [252]. Praćenje promene frekvencijskog sadržaja signala nakon decimacije i interpolacije je uobičajena procedura, a osnovni cilj je da se proveriti kakve promene ove procedure unose u signal. Paket wearables koristi proverene metode (iz naučnih publikacija) za pripremu podataka za dalju obradu, te primenjuje mašinski algoritam metode potpornih vektora (eng. *Support Vector Machine*, skraćeno SVM) za identifikaciju segmenata sa i bez šuma [245].

SVM u wearables paketu ima opciju za dodavanje treće klase gde se ne može odrediti prisustvo šuma (eng. *unclear* klasa) za primenu na EDA biosignalima merenim sa Empatica E4 senzorom [245]. Sa druge strane, jednostavna metoda je primenjena za proveru detektovanog srčanog ritma, odnosno samo je izvršena provera da li se puls nalazi u opsegu od 40 do 180 bita u minuti [245]. Kako se uobičajeno parametri računaju na relativno kratkim intervalima (na primer, 10 s, 1 minut ili 5 minuta) [237, 240-245], to wearables paket uključuje i deo za izdvajanje segmenata na kojima će dalja obrada biti primenjena (eng. *data cutter* alatka). Na Sl. 69 je prikazan primer rezultata manipulacije i analize biosignala sa Empatica E4 senzora primenom wearables paketa [245]. Na gornjem panelu na Sl. 69 se može primetiti rezultat podele na segmente koji su označeni različitim bojama intervala (prikazani su redom odozgo na dole EDA, srčani ritam u oznaci HR (eng. *Heart Rhythm*), temperatura kože i biosignali koji opisuju pokrete ispitanika merene primenom akcelerometara). Na srednjem i donjem panelu na Sl. 69 su prikazani kvalitet IBI podataka koji opisuju rastojanje između dva srčana otkucaja (eng. *Inter-Beat Interval*) i udeo podataka čiji je kvalitet označen kao loš u odnosu na ukupnu količinu podataka.

Detaljno korišćenje wearables paketa neće biti prikazano u ovom udžbeniku, ali se preporučuje čitaocima da detaljnije istraže ovaj paket koji je realizovan kao

Shiny aplikacija pod nazivom “E4 dashboard”²⁰² (interaktivna veb aplikacija koja se može pokretati i primenom internet pretraživača bez prethodne instalacija R-a i R Studio okruženja) i koji koristi automatsku obradu signala ili tzv. serijsku obradu (eng. *batch processing*). Serijska obrada u okviru wearables paketa omogućava da se automatski učitaju i obrade podaci mereni na relativno velikom broju ispitanika (nije potrebno ručno učitavati pojedinačne datoteke, analiza se sprovodi na svim ispitanicima sa jednim pokretanjem programa). [245]

2.2.1.4 Kratka istorija merenja pokreta

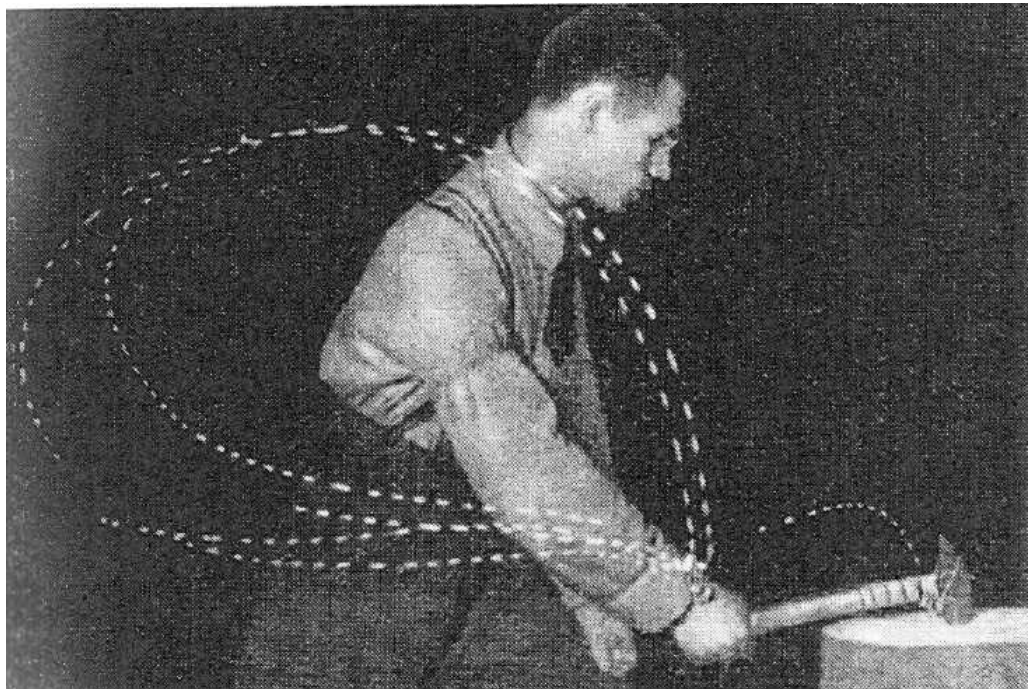
Merenje biosignala primenom odevnih tehnologija, minijaturnih prenosivih uređaja i mobilnih telefona je karakteristika istraživanja u biomedicinskom inženjerstvu 21. veka (delimično i 20. veka), ali proučavanje ljudskog ponašanja, posebno izučavanje upravljanja kretanjem (ROM podacima je započeto ovo poglavlje udžbenika), ima svoje korene u daljoj prošlosti. Primenu mehaničkih principa za rešavanje problema u biologiji su, prema Dejvidu A. Vinteru²⁰³, prvi primenili Leonardo da Vinči, Galilej, Lagranž, Bernuli, Ojler i Jung. Biomehanika je multidisciplinarna i interdisciplinarna²⁰⁴ oblast koja koristi saznanja iz fizike, hemije, matematike, fiziologije i anatomije. Pored primene u medicini, na primer, u oceni pokreta kod osobe kod kojih su amputirane ruka ili noga, biomehanika ima primenu i u sportu. Primena računara za merenje podataka tokom pokreta i za proračun kvantitativnih ocena

202 <https://github.com/PcdLf/wearables>, pristupljeno 19.08.2023.

203 Dejvid A. Vinter je bio profesor emeritus na Univerzitetu u Vaterlu u Kanadi (University of Waterloo, <https://uwaterloo.ca/>, pristupljeno 21.08.2023). Bio je jedan od najznačajnijih naučnika u svetu iz oblasti biomehanike i proučavanja kretanja, a posebno ljudskog hoda (https://en.wikipedia.org/wiki/David_A._Winter, pristupljeno 21.08.2023). Pisao je sjajne udžbenike iz biomehanike, motorne kontrole, ravnoteže, posture i hoda. Njegov udžbenik o biomehanici i motornoj kontroli ljudskog pokreta je naširoko poznat [207], a na sajtu Univerziteta u Vaterlu se može naći lista njegovih najznačajnijih publikacija [253]. Danas je nezamislivo baviti se ljudskim hodom, a prethodno ne pročitati knjige ili radove profesora Vintera.

204 Iako se termini multidisciplinarno i interdisciplinarno koriste intuitivno, važno je naznačiti u čemu je razlika između ta dva termina. Takvo razgraničenje može biti od značaja čitaocima koji se bave ili planiraju da se bave naučnoistraživačkim radom, jer veliki broj konkursa za finansiranje se poziva na jedan ili oba termina. U multidisciplinarnom pristupu, više istraživačkih grupa radi na istom problemu, ali iz svog ugla. Na primer, EMG signale mere i lekari i inženjeri sa ciljem da iz njih dobiju korisne informacije za ocenu stanja mišića, mišićno-skeletenog sistema ili o senzorno-motornoj kontroli. Interdisciplinarno dovodi do stvaranja zajedničkog znanja koje je rezultat zajedničkog rada, ali i do razvoja novih disciplina. Primera radi, lekari i inženjeri rade zajedno sa ciljem korišćenja EMG signala u oceni oporavka pacijenata nakon povrede nervnog sistema. Postoje i konvergentna i transdisciplinarna istraživanja. [254-255]

(kvalitativne ocene mogu biti rezultat posmatranja kretanja ispitanika u prostoru, ali i posmatranje, odnosno vizuelna inspekcija grafika, na kome su predstavljeni mereni biosignali) je značajno pojednostavila ocenu pokreta u biomehanici. [207]



Slika 70, Ciklogram tokom obrade metala dletom i čekićem – Aleksej Gastev u laboratoriji Centralnog instituta za rad. Izvor: nepoznat, sopstveno delo iz knjige “Radne instalacije”, CIT, 1924, javno vlasništvo, <https://commons.wikimedia.org/w/index.php?curid=15483992>.

Nezamislivo je uploviti u uzbudljiva istraživanja senzorno-motorne kontrole, a ne spomenuti jednog od pionira ove oblasti – sovjetskog naučnika Nikolaja Bernštajna²⁰⁵ koji je prvi uveo pojam biomehanike. U nekim izvorima se italijanski naučnik Đovani Alfonso Boreli²⁰⁶ pominje kao otac biomehanike²⁰⁷, jer je došao do značajnih otkrića tako što je poredio kretanje životinja sa

205 https://en.wikipedia.org/wiki/Nikolai_Bernstein, pristupljeno 21.08.2023.

206 https://en.wikipedia.org/wiki/Giovanni_Alfonso_Borelli, pristupljeno 21.08.2023.

207 Boreli je bio darovit renesansni naučnik, te njegova istraživanja nisu isključivo vezana za biomehaniku – smatra se i prvim idejnim tvorcem ronilačke opreme. Na žalost nije bio poznat za svoga života, ali je uz pomoć švedske kraljice Kristine njegov rad postao objavljen i poznat u nauci. Kristina je bila visoko obrazovana žena koja je pomagala naučnike i umetnike, a ostala je upamćena njena izjava da je potrebna veća hrabrost za udaju, nego za odlazak u rat (https://en.wikipedia.org/wiki/Christina,_Queen_of_Sweden, pristupljeno 21.08.2023). [256]

mašinama. Naravno, uz dozu preterivanja, počeci nastanka biomehanike se mogu povezati i sa prvim pisanim tragovima u Egiptu koji potkrepljuju tvrdnju o interesovanju za ljudsku anatomiju od pre oko 3700 godina. [256-257]

Bernštajnov rad je utro put savremenom pristupu biomehanici, te je zato i odabran za predstavljanje u ovom poglavlju. Maks Plank institut za istoriju nauke²⁰⁸ je 2009. godine objavio sjajan članak pod nazivom “Fiziologija klavira”²⁰⁹ u kome je predstavljen rad neurofiziologa Nikolaja Bernštajna na biomehaničkom objašnjenju virtuoznog izvođenja klavirskih kompozicija.

Izvođenje brzih i paralelnih oktava na klaviru nije nimalo jednostavan fizički zadatak za one koji žele da nauče da sviraju klavir profesionalno. Bernštajn je u saradnji sa Nacionalnim institutom za muzikologiju u Rusiji razvio uređaj koji je mogao da beleži pokrete pijanista tokom izvođenja paralelnih oktava. To je uspeo da uradi postavljanjem sijalica na ruku pijanista i snimanjem pokreta kamerom. Pažljivim čitaocima verovatno nije promaklo kako se danas u svetu mere uglovi u zglobovima i druge mehaničke karakteristike pokreta prikazane na Sl. 54 i Sl. 55 u udžbeniku. Proučavanjem trajektorija sijalica u pojedinačnim prozorima (eng. *frame*), Bernštajn je došao do zaključka da samo brzi pokreti i voljne vibracije ruke mogu da izvedu ovakve kompozicije, a da je prethodno primenjivana strategija prilikom vežbanja da se pokret najpre izvrši polako, a potom da se ubrzava bio pogrešan. Bernštajn je kasnije realizovao uređaj i princip analize signala primenio uspešno u drugim oblastima – kod pokreta sa postavljenim ortozama²¹⁰, u sportu i ergonomiji. Princip po kome je zasnovana metoda profesora Bernštajna za merenje pokreta je bio da su ljudski pokreti brzi te se ne mogu videti golim okom, a za kvantitativni opis pokreta potreban je tačan i precizan uređaj. Bernštajn je sa saradnicima razvio uređaj koji su nazvali ciklograf. Primer jednog ciklograma (signala merenog primenom ciklografa) je prikazan na Sl. 70. [258-259, 261]

Brzina kojom su kamere merile je bila 200 slika u minuti – dobijeni ciklogrami su se koristili za primenu u sportu kako bi se na odgovarajući način planiralo obučavanje atletičara [261], ali i za poboljšanje produktivnosti radnika (Sl. 70) [258-259]. U biomehanici je poznat “problem stepeni slobode” (od eng. *degrees*

208 Max Planck Institute for the History of Science, <https://www.mpiwg-berlin.mpg.de/>, pristupljeno 19.08.2023.

209 Slobodan prevod sa engleskog jezika na srpski jezik. U originalu, naziv članka je “*Physiology of the Piano*” [258-259].

210 Neuralne proteze su asistivni uređaji koji omogućavaju zamenu ili ponovno uspostavljanje funkcije koja je izgubljena nakon povrede nervnog sistema (eng. “*Neural prostheses (NPs) or assistive devices that restore functions lost as a result of neural damage*”). [260].

*of freedom problem*²¹¹) čijoj je formulaciji, između ostalih naučnika, doprineo i Bernštajn. On je formulisao problem koji se odnosi na način na koji nervni sistem određuje kako će neki pokret biti izvršen, zato što postoji relativno veliki broj stepeni slobode koji omogućavaju da se jedan pokret izvede na više načina. Bernštajn je izneo pretpostavku da se organizacija upravljanja pokretima obavlja unutar podsistema ili sinergija, kao i da postoje nivoi kontrole ispod i iznad sinergija. Ovo je značajno pojednostavilo problem upravljanja pokretima u ljudskom telu, jer umesto da se posmatraju pojedinačni stepeni slobode, u biomehanici se posmatraju globalnije funkcionalne jedinice – sinergije. [262]

Pored opsega pokreta, za studije koje je Bernštajn izvodio važno je da postoji informacija u vremenu. Ne mora nužno svaki pokret da sadrži pun opseg u zglobovima, posebno usled postojanja redundantnih stepeni slobode koji je objašnjen postojanjem sinergija [262]. Merenje uglova u zglobovima i drugih biosignala od značaja za proučavanje ljudskog kretanja ili posture, od kojih je najznačajniji EMG signal, važno je i u ergonomiji. Kako bi se poboljšali uslovi rada i smanjio negativan uticaj okruženja na zaposlene koji može da potiče od alata koji se koriste na radnom mestu i položaja tela (rad u sedećem ili stojećem položaju, neodgovarajuća stolica i dr), važno je da se na kvantitativan način analiziraju pokreti i položaj tela [105]. Na primer, u studiji koja je realizovana na Stomatološkom fakultetu, Univerziteta u Beogradu²¹² u saradnji sa drugim institucijama uključujući Elektrotehnički fakultet, Univerziteta u Beogradu, i koja je bila pretežno fokusirana na EMG signale uz merenja sa akcelerometara koji su bili postavljeni na telo stomatologa je zaključeno da je najbolje da stomatolozi tokom rada kombinuju sedeći i stojeći položaj kako bi sprečili pojavu mišićno-skeletnih oštećenja [263]. Povrede usled nepravilnog položaja tela ili zamora koji može da nastane tokom rada mogu se javiti ne samo kod stomatologa, već i kod onih koji provode dugo vremena za računarnom²¹³, upravljaju prevoznim sredstvima, rade u postrojenjima za preradu mesa, rade na industrijskim zaptivačima automobila, rudari, magacioneri, slikari, građevinski radnici, vatrogasci, medicinsko osoblje i drugi, te je potreba za računarskom analizom i odgovarajućim mernim sistemima u ergonomiji velika [264].

2.2.2 Eliminacija šuma na primeru EMG biosignala

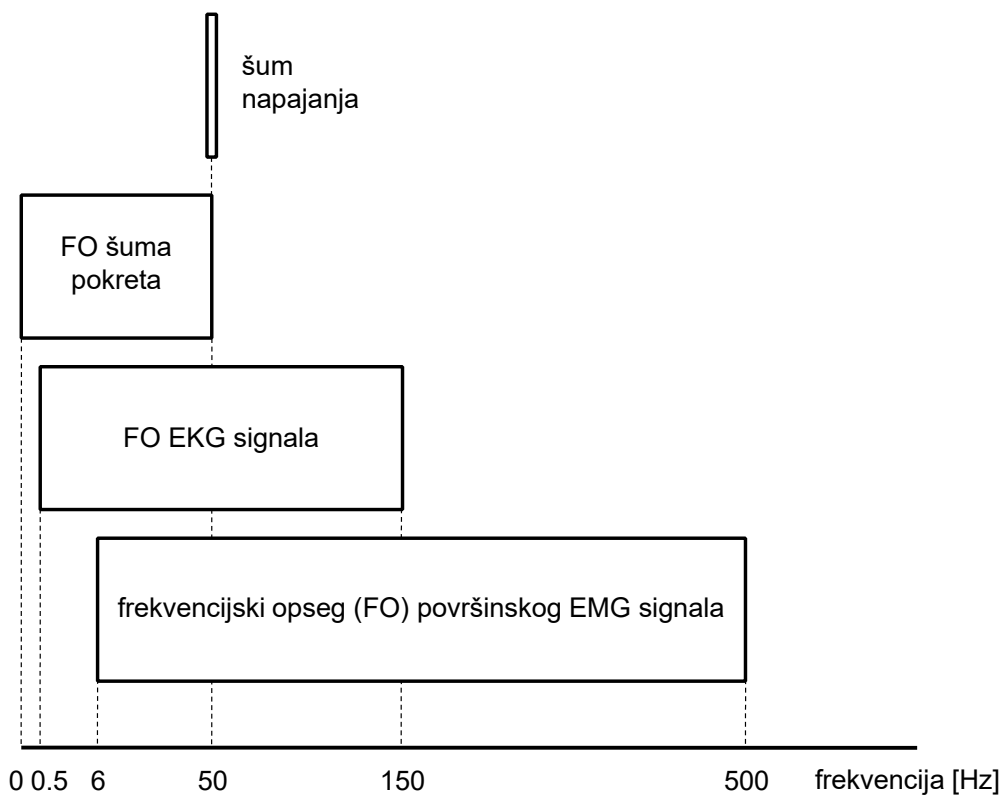
Kao što je bilo pokazano na primeru ergonomijske studije za ocenu pokreta stomatologa tokom pregleda u sedećem i stojećem položaju [263], merenje

211 https://en.wikipedia.org/wiki/Degrees_of_freedom_problem#cite_note-bernstein-1, pristupljeno 22.08.2023.

212 <https://www.stomf.bg.ac.rs/>, pristupljeno 22.08.2023.

213 Ako ovaj udžbenik čitate na računaru, možda je vreme da napravite pauzu i prošetate?

elektromiografskih signala je korisno u studijama merenja pokreta u biomehanici. Značaj EMG signala se ogleda u činjenici što ovaj signal predstavlja sumu akcionih potencijala mišića koja je rezultat kontrakcije skeletnih mišića čija je funkcija da omogući kretanje i održavanje ravnoteže u ljudskom telu [265]. Pored merenja električnih potencijala skeletnih mišića, moguće je meriti i električne potencijale glatkih mišića. Merenjem sume akcionih potencijala glatkih mišića želuca dobija se EGG signal koji je od značaja u gastroenterologiji i u drugim oblastima u kojima se proučava rad želuca [266]. Analogno sa ostalim elektrofiziološkim metodama, kao što su EKG, EEG i EGG, neinvazivni EMG je moguće meriti primenom površinskih srebro/srebro-hlorid (Ag/AgCl) elektroda [104-105], pa se u literaturi na engleskom jeziku EMG signal označava i sa skraćenicom sEMG od eng. *surface electromyogram*²¹⁴.



Slika 71, Grafička ilustracija preklapanja frekvencijskih opsega (FO) EMG signala sa EKG signalom, šumom pokreta i šumom napajanja (50 Hz ili 60 Hz).

214 Površinski elektromiogram, slobodan prevod sa engleskog na srpski jezik.

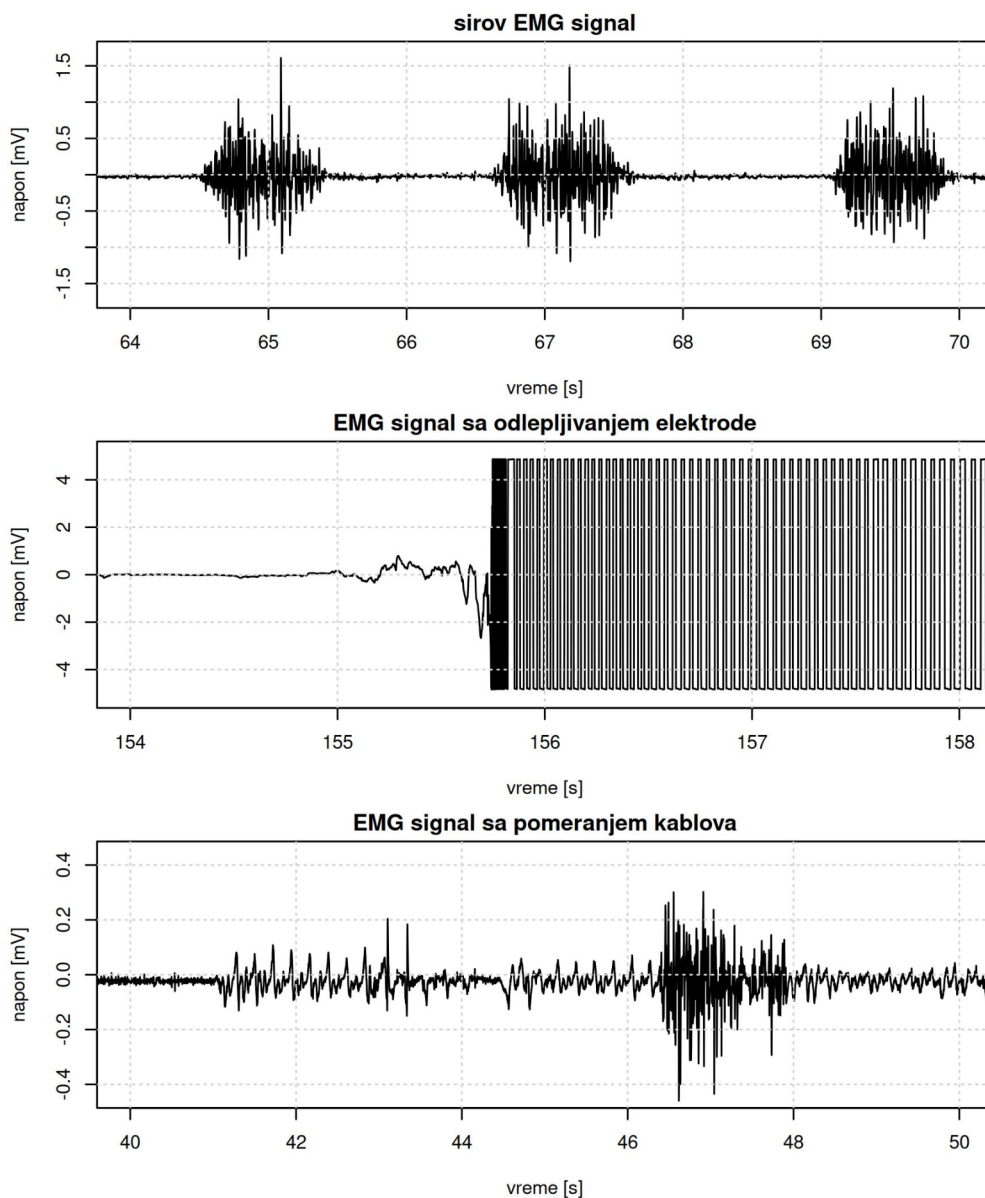
Kao i ostali biosignali, EMG signal može sadržati značajnu količinu šuma i to je pretežno rezultat relativno malih amplituda, te nepovoljnog SNR. Na EMG signal može uticati šum napajanja od 50 Hz ili 60 Hz, mada u značajno manjoj meri nego, na primer, na EEG signale, te se ređe u uređajima za merenje EMG signala nalazi noć²¹⁵ (eng. *notch*) filtar. Ova vrsta šuma se nekada naziva i AC šum od eng. *Alternating Current*, dok se DC šumom od eng. *Direct Current* naziva šum koji potiče dominantno od potencijala polaziracije elektroda. Frekvencijski opseg EMG signala je od 20 Hz do 500 Hz [265] ili od 6 Hz do 500 Hz [105], te se koristan frekvencijski opseg signala poklapa sa šumom napajanja, što je prikazano na skici na Sl. 71 i na grafiku na Sl. 72 na kome je prikazan EMG signal sa uočljivim šumom napajanja.

EMG signali su reda par mV, mada kod merenja na sportistima signal može biti značajno veći [104-105]. Za digitalizaciju EMG signala primenom A/D konvertora, potrebno je primeniti pojačanje (najčešće se koristi instrumentacioni pojačavač sa pojačanjem u opsegu od 100 do 5000 puta za površinsko merenje EMG signala). U zavisnosti od A/D konvertora koji se primenjuje moguće je naći kompromis između pojačanja i rezolucije A/D konvertora. Naime, što je rezolucija A/D konvertora veća, to je moguće izmeriti manje promene na signalu. Primera radi, rezolucija Arudino UNO mikrokontrolerske pločice je 10 bita²¹⁶ na opsegu analognog signala od 5 V, te je najmanja moguća vrednost koju je moguće izmeriti, odnosno najmanja promena na signalu, koja će biti digitalizovana jednaka $5 \text{ V} / 2^{10} = 4.88 \text{ mV}$ [78], dok bi bilo moguće izmeriti najmanju moguću vrednost od $76.29 \mu\text{V}$ u slučaju da je rezolucija 16 bita. Sledi da nije potrebno relativno veliko povećanje EMG signala u slučaju da postoji relativno velika rezolucija A/D konvertora²¹⁷.

215 Noć filter je filter nepropusnik opsega (eng. *stopband filter*) u relativno uskom propusnom opsegu – u slučaju šuma napajanja, ovaj šum uklanja jednu frekvenciju (50 Hz ili 60 Hz) u frekvencijskom opsegu. Praktično, filtar uklanja i relativno malu okolinu ciljane frekvencije, a širinu nepropusnog opsega prilikom projektovanja filtra je moguće podesiti.

216 Rezolucija Arudino mikrokontrolerskih pločica je detaljno prikazana u uputstvu za *analogRead()* funkciju na Arduino sajtu, <https://www.arduino.cc/reference/en/language/functions/analog-io/analogread/>, pristupljeno 05.09.2023.

217 Iako je većina biomedicinskih inženjera saglasna da se EMG signal može izmeriti primenom A/D konvertora relativno visoke rezolucije (na primer, primenom zvučne kartice na računaru rezolucije 32 bita), bez pojačavačkih kola, Autorka do sada nije u praksi imala prilike da vidi nijednu takvu realizaciju. Autorka iz svog iskustva može samo da potvrdi da je uz A/D konvertor rezolucije 16 bita moguće koristiti EMG pojačavač sa pojačanjem od 100 puta.



Slika 72. EMG signal na gornjem panelu, EMG signal sa šumom napajanja na srednjem panelu (od 155.8 s) i EMG signal sa šumom pokreta na donjem panelu (kablovi su pomerani ritmički tokom merenja, kada je izvršena jedna kontrakcija EMG mišića od 46 s do 48 s). Prikazani EMG signali su mereni za potrebe pisanja udžbenika na Elektrotehničkom fakultetu, Univerziteta u Beogradu, a detaljnije informacije su dostupne u tekstu.

Iako amplituda EMG signala zavisi od veličine mišića (što jeste naročito uočljivo kada se meri EMG na mišićima sportista), ona zavisi i od drugih parametara kao što su vrsta pokreta, intenzitet kontrakcije, impedansa na

kontaktu elektroda-koža, debljina adipoznog (masnog) tkiva i impedansa kože [105]. Obrada biosignala koji su mereni iglenim i implantibilnim elektrodama neće biti opisana u ovom udžbeniku. Razlog je što su osobine invazivno merenih biosignala drugačije od osobina površinski merenih biosignala: (1) mogu biti značajno većih amplituda (elektrode su postavljene direktno na izvor signala, a ne kao kod površinskog merenja preko kože) i (2) mogu sadržati značajno veći frekvencijski opseg (do 2 kHz ili čak do 10 kHz), jer ne postoji prostorno filtriranje signala. Pojam prostornog filtriranja se jednostavno može objasniti kapacitivnim osobinama kože i potkožnog tkiva. Najjednostavniji kondenzator je sastavljen iz dve provodne ploče između kojih se nalazi izolator [78], što je model tkiva kroz koje prolazi signal od izvora (mišićno tkivo) do mernog mesta (kontakt elektroda-koža na površini kože) u kome se smenjuju provodni i neprovodni elementi²¹⁸). Tkivo koje se nalazi između izvora signala i mernog mesta ima osobine jednostavnog filtra – RC kola [78] koje ostvaruje prostorno filtriranje EMG signala u slučaju da se signal meri površinskim elektrodama. Ovo potpoglavlje ima za cilj da pokaže kako se mogu realizovati metode za eliminaciju šuma na površinski merenom EMG signalu. Sa druge strane, implantibilni signal meren iglenim ili implantibilnim elektrodama je u manjoj meri podložan šumu, te sa stanovišta eliminacije šuma i manje interesantan. EMG signali mereni površinskim elektrodama, pored šuma napajanja (pogledati EMG signal sa šumom napajanja na Sl. 72), mogu biti kontaminirani i drugim šumovima, od kojih je najčešće zastupljen u dinamičkim studijama – šum pokreta.

Dominantno, šum pokreta nastaje usled pomeranja kablova tokom merenja. Da bi se smanjio šum pokreta, moguće je koristiti bežične sisteme sa predpojačavačima. Naime, pojačavački stepen bi trebalo da bude što bliže mernom mestu. Ako se pomeraju kablovi (u opštem slučaju konektori) koji se nalaze nakon pojačavačkog stepena (koji u opštem slučaju može da sadrži i analogni filter), onda je SNR povoljniji i većih vrednosti u odnosu na pomeranje kablova pre pojačavačkog stepena. Iako se šum smanjuje primenom bežičnih sistema, on se ne može u potpunosti eliminisati (posebno kada su studije izrazito dinamičkog karaktera – na primer, kada postoji fizički kontakt dela tela, odeće ili elektronike sa elektrodama tokom merenja [267]).

Pored toga što upotreba bežičnih sistema može ublažiti problem nastajanja šuma pokreta, ona ga ne može rešiti. Sa druge strane, upotreba bežičnih sistema unosi druge probleme u mernom uređaju (kao što su pouzdanost protokola za

218 Mišići i nervi spadaju u ekscitabilna tkiva, odnosno mogu se smatrati provodnicima [105].

Sa druge strane, adipozno/masno tkivo ne provodi električnu struju, te se može smatrati u ovom slučaju izolatorom.

prenos podataka i kompleksnost sistema koja se ogleda u merenju višekanalnog EMG signala i sinhronizaciji merenja sa drugim uređajima) [80, 268]. Generalno, puno je studija u poslednjih par godina objavljeno o kvalitetu signala koji su mereni bežično primenom odevnih tehnologija. Iako su među najpopularnijim i najzastupljenijim u oblasti, odevne tehnologije predstavljaju najveći izazov za biomedicinske inženjere, jer zadatak dobijanja korisne informacije iz tako merenih signala nije nimalo jednostavan. [269-271]

Na Sl. 72 dat je primer EMG signala merenog bez šuma (na gornjem panelu na Sl. 72) za tri kontrakcije mišića *biceps brachii* (lat.) leve ruke kada nije bilo izrazitih šumova tokom merenja. Na srednjem panelu na Sl. 72 prikazano je kako se menja signal sa odlepljivanjem elektrode u otprilike 155.8 s. Na EMG signalu prikazanom na srednjem panelu nije snimljena nijedna mišićna kontrakcija, a promena talasnog oblika signala neposredno pre odlaska pojačavača u zasićenje usled odlepljivanja elektrode oko 155 s potiče od šuma pokreta (pomeranje kablova i elektrode) [267]. Treći panel prikazuje jednu mišićnu kontrakciju koja se nalazi otprilike u opsegu od 46 s do 48 s, dok su preostale varijacije bazne linije signala prikazane u relativno periodičnim trenucima posledica namernog pomeranja kablova (trzanje kablova). Na Sl. 72 prikazani su signali mereni sa EMG predpojačavačima koji imaju kraće kablove od elektrode do predpojačavača i duže kablove od predpojačavača do A/D konvertora (Biovision Ltd., Wehrheim, Germany)²¹⁹. Prema tome, šum prikazan na srednjem i donjem panelu na Sl. 72 potiče od jednih i drugih kablova koji su, kako je već objašnjeno, različito osetljivi na artefakte pokreta.

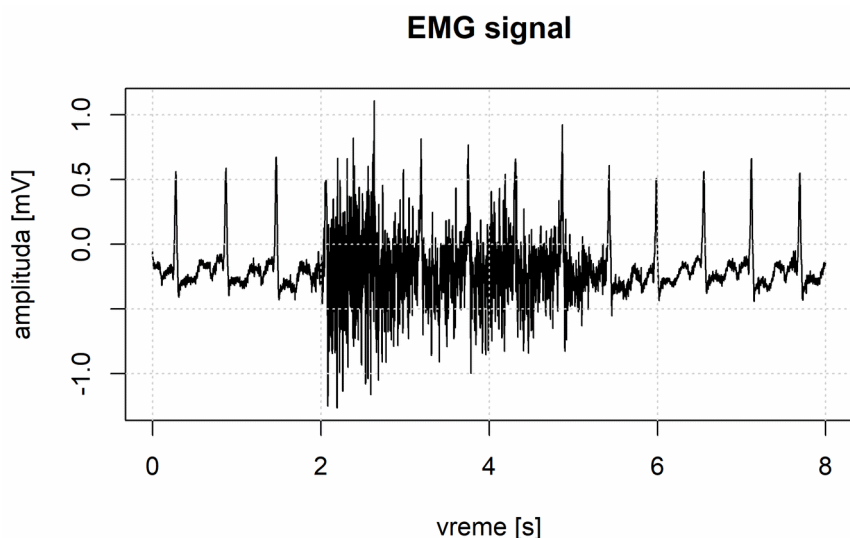
Većina komercijalno dostupnih uređaja za merenje EMG signala sa ciljem filtriranja šuma pokreta već ima ugrađen filter visokopropusnik (eng. *high pass*) [104]. Šum pokreta je relativno niskog frekvencijskog sadržaja (do 20 Hz), ali pomeranje kablova može biti u opsegu do čak 50 Hz kako je ilustrovano na dijagramu na Sl. 71. [272]. Na Sl. 72 na donjem panelu vidljive su kratke visokofrekventne promene signala koje se mogu pogrešno protumačiti kao EMG signali, jer se šum pokreta može pogrešno tumačiti kao refleksna aktivacija mišića [267]. Ovo je samo jedan od primera zašto je dobro poznavati prirodu signala i šuma, kao i uslove merenja za odgovarajuću analizu biosignala.

Kada se EMG signal meri sa mišića koji se nalazi u blizini srca (praktično na svim mestima na trupu), često se javlja šum čije je poreklo u električnoj aktivnosti srca. U literaturi se može sresti i naziv EKG šum, ali važno je

219 http://www.biovision.eu/biovision1_en.htm, pristupljeno 05.09.2023.

spomenuti da se EKG, u skladu sa standardizacijom i kliničkom upotrebom, dobija primenom Ajthovenovog trougla, a da se ovde radi o fiziološkom preslušavanju signala, te nije pravilno reći EKG šum. Ali nije ni pogrešno, jer taj signal ipak predstavlja transformaciju EKG signala. Frekvencijski opseg EKG signala koji je definisan industrijskim standardom se kreće od 0.05 Hz (ili 0.5 Hz) do 100 Hz (ili 150 Hz) [273], te se EKG u potpunosti poklapa sa EMG signalom kako je prikazano na Sl. 71.

Za razliku od šumova napajanja i pokreta koji su ili u potpunosti razdvojeni u frekvencijskom domenu od opsega u kome se nalazi EMG signal ili se poklapaju na zanemarivo malom opsegu (jedna frekvencija sa relativno malom okolinom u slučaju šuma napajanja), EKG šum nije takav. Činjenica da je kompletno preklapljen sa EMG signalom, značajno otežava eliminaciju EKG šuma iz EMG signala. Zbog preklapanja u frekvencijskim opsezima nije moguće primeniti linearno filtriranje, odnosno metode koje dobro razdvajaju frekvencijske opsege signala. Međutim, kombinacija određenih tehnika, kao što su filter medijana i frakcioni izvod mogu značajno poboljšati efekte filtriranja, uz ograničenje da krajnji cilj nije dobijanje čistog EMG signala, već obvojnice na koju ne utiče u značajnoj meri EKG šum [274].



Slika 73. EMG signal sa desnog mišića *Pectoralis major* koji se nalazi na prednjoj strani grudnog koša. Signal i R kod su dostupni na GitHub repozitorijumu (<https://github.com/NadicaSm/satRday-Belgrade-2018>, pristupljeno 14.09.2023.) [251].

Na Sl. 73 prikazan je primer signala merenog na mišiću *Pectoralis major* sa leve strane tela. Na EMG signalu tokom relaksacije i tokom jedne kontrakcije

koja je izvršena tako što je ispitanik podigao ruke ispred sebe u ravni sa podlogom vidljivi su otkucaju srca (Sl. 73). Iako ovakvo preklapanje dva biosignala predstavlja veliki izazov u analizi, a posebno u eliminaciji šuma (u slučaju da se meri EMG signal, EKG signal se smatra šumom), ono ujedno predstavlja i veliku mogućnost za multimodalna merenja. Primera radi, ovakvo preklapanje se može iskoristiti da se primenom jednog merenog kanala izdvoje informacije o dva biosignala EMG-u i EKG-u, a sličan princip je predložen za EGG i EKG signale [79, 268].

EMG signali koji su prikazani na Sl. 72 su mereni na Elektrotehničkom fakultetu, Univerziteta u Beogradu za potrebe pisanja ovog udžbenika. Merenja su izvršena u skladu sa Helsinškom deklaracijom i u skladu sa članom 19. “Korišćenje ljudi i životinja u naučnom istraživanju” Kodeksa profesionalne etike Univerziteta u Beogradu²²⁰. Signali su sačuvani u dve datoteke “emg-sve.txt” i “emg-kabl.txt” koje su podeljene na GitHub repozitorijumu²²¹ kao prateći materijal uz ovaj udžbenik. U datotekama se nalazi 16 kolona, jer je primenom softvera za akviziciju signala [104] moguće meriti do 16 signala sinhrono. U obe datoteke merenje je izvršeno samo na prvom kanalu, te je signal dostupan isključivo u generičkoj koloni pod nazivom *V1*. Da bi se signal iscrtao u vremenskom domenu, potrebno je: (1) formirati vremensku osu (signal je meren sa frekvencijom odabiranja od 1000 Hz) i (2) po potrebi izvršiti skaliranje amplitude kako bi amplituda bila prikazana u realnom opsegu (pojačanje je bilo postavljeno na 1000 puta). Na Sl. 74 prikazana je skripta koja omogućava da se definišu vremenske ose za oba signala koji su smešteni u različitim datotekama. Za realizaciju vremenske ose, moguće je koristiti funkciju:

`seq()`

koja generiše nizova brojeva (u ovom slučaju realnih brojeva, odnosno vreme u s, koje odgovara svakom pojedinačnom odbirku). Tri ulazna parametra funkcije su korišćena: (1) početni vremenski trenutak u sekundama koji je u ovom slučaju postavljen na 0²²², (2) trajanje merenja EMG signala u s – primetiti da je

220 Kodeks profesionalne etike Univerziteta u Beogradu iz 2016. godine, <https://bg.ac.rs/files/sr/univerzitet/univ-propisi/Kodeks-profesionalne-etike.pdf>, pristupljeno 14.09.2023.

221 Iako je ranije već dat naziv repozitorijuma, korisno je spomenuti ga na više mesta u udžbeniku: <https://github.com/NadicaSm/>.

222 Istovremeno pravilno je i nepravilano pisati jedinicu uz vrednost 0, jer 0 znači da neka vrednost nije postojeća (0 m znači da je to nepostojeća dužina, pa se jednostavno može pisati samo 0). Međutim, u nekim situacijama ovo je opravdano, kao što je, na primer, dovođenje mernog mosta u ravnotežu ili posmatranje da li postoje odstupanja od ravnoteže

trajanje umanjeno za jednu periodu ($1/fs$, gde je fs frekvencija odabiranja) i (3) korak (argument *by*), odnosno perioda odabiranja (Sl. 74). Primiti da na Sl. 75 ne postoji dodatno skaliranje amplitude signala, a razlog tome je što je pojačanje bilo 1000, te se jedinice odmah mogu pisati u mV, iako je signal izmeren sa A/D konvertora u V. Za pojačanje od 2000, sve amplitude bi se morale podeliti sa 2 kako bi jedinice bile prikazane u mV. Ovo je posebno važno i ne sme se prikazati pogrešna amplituda signala. Postoje i standardni dodatni elementi u prikazu biosignala kao što je milimetarska pozadina u kliničkoj kardiologiji, ali o njima ovde neće biti reči.

```
> datEMG <- read.table("emg-sve.txt")
> datEMGkab1 <- read.table("emg-kab1.txt")
> fs <- 1000
> timeAxis <- seq(0, length(datEMG$V1)/fs - 1/fs, by = 1/fs)
> timeAxisKab1 <- seq(0, length(datEMGkab1$V1)/fs - 1/fs, by = 1/fs)
> |
```

Slika 74. Definisane vremenske ose za dva EMG signala koji su smešteni u različitim datotekama. Za svaki od signala je potrebno definisati posebno vremensku osu. Promenljive koje sadrže vremenske ose su *timeAxis* i *timeAxisKab1*.

Signali su prikazani i grafik je sačuvan u JPEG (eng. *Joint Photographic Experts Group*) formatu (Sl. 75). Za čuvanje grafika u JPEG formatu, koristi se istoimena funkcija:

jpeg()

koja kao ulazne parametre ima naziv datoteke, dimenzije slike u inčima i rezoluciju (Sl. 75). Nakon poziva ove funkcije se grafik koji je potrebno sačuvati u JPEG formatu iscrta. Da bi se na jednoj slici prikazalo više grafika (pogledati argument *mfrow*) koristi se funkcija:

par()

koja na Sl. 75 omogućava da se prikaže tri grafika u tri vrste i jednoj koloni, a ukupna komanda se završava sa:

dev.off()

funkcijom koja ima za cilj da kontroliše prikaz više grafika, odnosno da ponovo postavi početna podešavanja (resetuje) koja su promenjena pomoću *par()*

(u pozitivnom ili negativnom smeru) – u tom slučaju korisno je staviti jedinicu.

funkcije (takođe, ova komanda je važna kako bi se pravilno izvršila funkcija *jpeg()* za čuvanje slike u JPEG formatu).

```
jpeg(file = "emg.jpg", width = 6, height = 7,
      units = 'in', res = 300)
par(mar = c(4.2, 4.2, 1.7, 1.7))
par(mfrow = c(3, 1))
plot(timeAxis, datEMG$V1, type = "l", xlim = c(64, 70),
      ylim = c(-1.7, 1.7),
      xlab = "vreme [s]", ylab = "napon [mV]",
      main = "sirov EMG signal")
grid()
plot(timeAxis, datEMG$V1, type = "l", xlim = c(154, 158),
      ylim = c(-5.2, 5.2),
      xlab = "vreme [s]", ylab = "napon [mV]",
      main = "EMG signal sa odlepljivanjem elektrode")
grid()
plot(timeAxisKabl, datEMGkabl$V1, type = "l", xlim = c(40, 50),
      ylim = c(-0.45, 0.45),
      xlab = "vreme [s]", ylab = "napon [mV]",
      main = "EMG signal sa pomeranjem kablova")
grid()
dev.off()
```

Slika 75. Skripta koja omogućava prikaz EMG signala na grafiku. Reprodukcijski prikaz koda rezultuje signalom prikazanim na Sl. 72.

Dodatno, na Sl. 75 je primenjen poziv funkcije *par()* sa argumentom *mar*:

```
par(mar = c(4.2, 4.2, 1.7, 1.7))
```

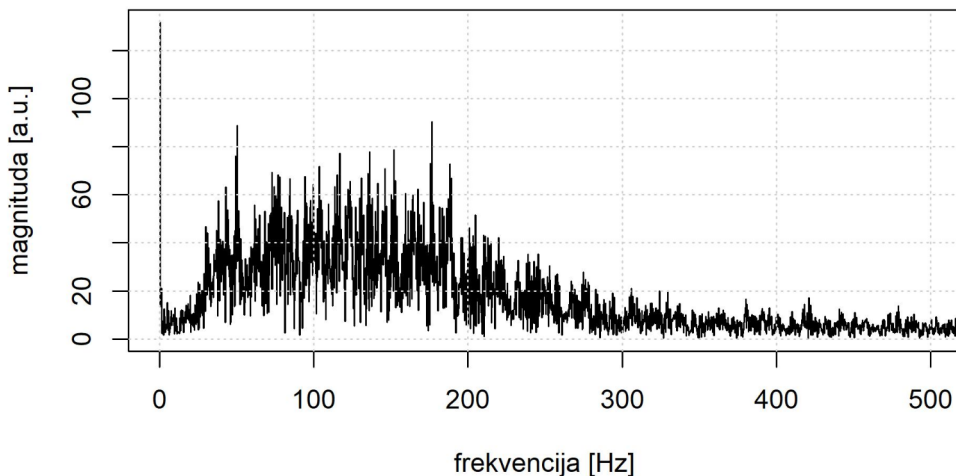
Podrazumevana vrednost *mar* argumenta je [5.1, 4.1, 4.1, 2.1]. Ovaj niz brojeva definiše broj linija od margine koji se neće koristiti kao aktivna površina grafika u formi [donja margina, leva margina, gornja margina, desna margina]. U prikazanom primeru na Sl. 75 broj linija je smanjen svuda osim na levoj margini kako bi grafik bio prijatniji za posmatranje i kako bi što veća površina bila iskorišćena.

Za one koji imaju iskustva u GNU Octave ili Matlab programskim jezicima, funkcija *par()* je slična *subplot()* funkciji i *hold on* komandi. Dodatna podešavanja grafika su moguća, jer na Sl. 72 sve ose (apscisa i ordinata na sva tri grafika) imaju iste oznake, te ima smisla ostaviti samo na grafiku na donjem panelu oznake za ose, kako bi se oslobodio prostor na slici. Za te potrebe, mogu se koristiti sledeće komande:

```
xaxt = "n", yaxt = "n",  
axes = FALSE,  
xlab = "", ylab = "",
```

unutar `plot()` funkcije koje omogućavaju redom da se na osama ne prikažu brojne vrednosti (za slučaj da je na svim graficima prikazan isti opseg u vremenskom domenu), da se ne prikažu same granice, kao i da se ne prikažu nazivi osa.

Prikaz signala u vremenskom domenu je jedan od prvih koraka koji je potrebno primeniti da bi se proverilo da li je signal dobro učitani u R okruženje, da li postoji prisustvo šuma koje se može utvrditi vizuelnom inspekcijom, da li su ose dobro unete za prikaz signala i sl. Sledeći korak, pre primene tehnika za eliminaciju šuma je najčešće prikaz signala u frekvencijskom domenu. Za prikaz brze Furijeove transformacije (eng. *Fast Fourier Transform*, skraćeno FFT), kao i za primenu linearnih filtara, u R programskom jeziku se koristi signal paket [275]. FFT sirovog (neobrađenog) EMG signala koji je prikazan na Sl. 72 je data na Sl. 76, a skripta u R-u je prikazana na Sl. 77.



Slika 76. FFT EMG signala koji je u vremenskom domenu prikazan na gornjem panelu na Sl. 72.

Uočljivo je da je dominantni frekvencijski sadržaj EMG signala između 100 Hz i 150 Hz (Sl. 76) – ovo je važan podatak jer postoji literatura iz oblasti biomedicinskog inženjerstva u kojoj je pogrešno naveden značajno viši opseg na kome se nalazi najveća spektralna gustina snage EMG signala. Kako to obično biva, greške u udžbenicima propagiraju, pa se skreće pažnja studentima.

```

> library(signal)
> podFFT <- fft(datEMG$V1[(64*fs):(70*fs)])
> podMag <- Mod(podFFT)
> podMag <- podMag[1:length(podMag)/2]
> fosa <- 1:length(podMag)/
+   (length(datEMG$V1[(64*fs):(70*fs)])/fs)
> jpeg(file = "emg-fft.jpg", width = 6.5, height = 4,
+   units = 'in', res = 300)
> plot(fosa, podMag, type = "l", xlim = c(0, 500),
+   xlab = "frekvencija [Hz]",
+   ylab = "magnituda [a.u.]")
> grid()
> dev.off()
null device
      1
> |

```

Slika 77. R skripta za prikaz FFT EMG signala koji je u vremenskom domenu prikazan na gornjem panelu na Sl. 72.

Jedinice za magnitudu²²³ se obično ne stavljaju, te je na Sl. 76 uneto a.u. (eng. *Arbitrary Unit*) za oznaku jedinice magnitude što u prevodu na srpski jezik znači proizvoljna jedinica. Jedinice za magnitudu su mV/Hz ili mVs (ili u mV što zavisi od dodatnih skaliranja i realizacije *fft()* funkcije u programskom jeziku). Potrebno je primetiti da se za prikaz magnitude EMG signala na Sl. 77 koristi funkcija:

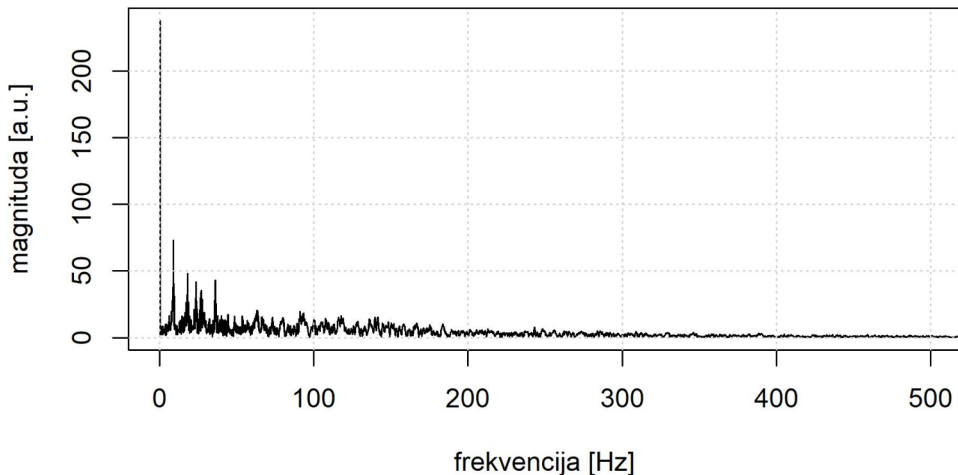
Mod()

koja na svom izlazu daje moduo kompleksnog broja (rezultat FFT funkcije je kompleksan broj). Takođe, za prikaz magnitude se iscrtava samo polovina spektra, jer je koristan deo spektra do Nikvistove frekvencije, odnosno do $fs/2$.

Kako bi studenti stekli iskustvo u posmatranju uticaja šuma na magnitudu merenog signala, na Sl. 78 je prikazana FFT od EMG signala kod koga je došlo do pomeranja kablova. Na Sl. 78 je jasno uočljiv šum bazne linije (najpre kroz jednosmernu ili DC komponentu) koja postoji i na Sl. 76 samo je manje magnitude, a potom i kroz relativno visoku spektralnu gustinu snage signala na nižim učestanostima koje ne predstavljaju uobičajen FFT EMG signala. Umesto da se najveća spektralna gustina snage signala nalazi oko 130 Hz kao na Sl. 76,

²²³ U pojedinim izvorima magnituda se naziva amplitudom bez obzira na skaliranja unutar *fft()* funkcije. Oba su ispravna, ali će u ovom udžbeniku, radi jednostavnijeg praćenja, amplituda biti korišćena isključivo za signal u vremenskom domenu, a magnituda za signal u frekvencijskom domenu.

na Sl. 78 najveća spektralna gustina snage signala je na značajno manjim frekvencijama i to ispod 50 Hz. Iako svi istraživači koji analiziraju EMG signale pre primene metoda za eliminaciju šuma pogledaju kako signal izgleda u vremenskom i frekvencijskom domenu, postoji saglasnost da filtriranje signala, posebno filtrom visokopropusnikom, treba uvek izvršiti. Filtar visokopropusnik nije neophodan za signal čiji je FFT prikazan na Sl. 76, ali ako jedna realizacija istog merenja ima oblik signala kao na Sl. 78 (sa izraženim šumom pokreta na nižim frekvencijama), onda ima smisla uniformisati metodu eliminacije šuma i odgovarajući filter primeniti na celom skupu merenih signala (tzv. serijska obrada ili eng. *batch processing*). Još izraženiji šum pokreta je zajedno sa šumom napajanja na 50 Hz i pripadajućim harmoničima (koji su posebno izraženi na 100 Hz i na 200 Hz) prikazan na Sl. 79.



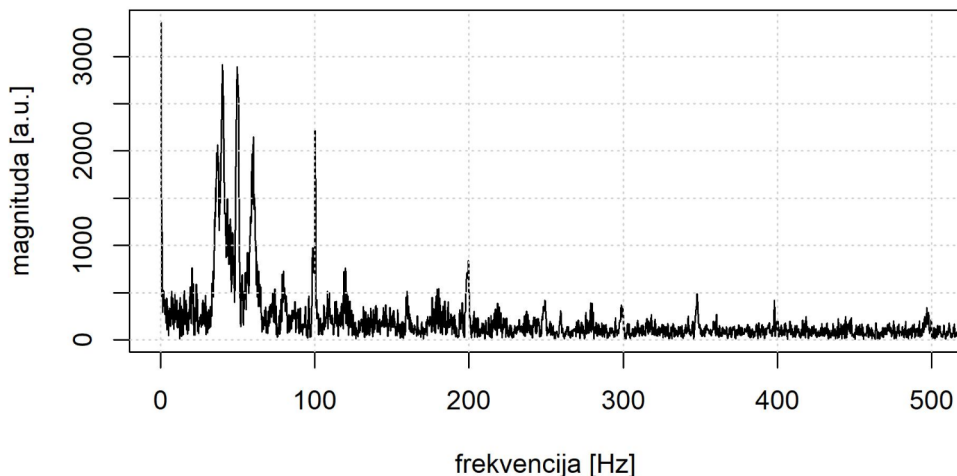
Slika 78. FFT EMG signala koji je u vremenskom domenu prikazan na donjem panelu na Sl. 72 (efekat pomeranja kablova je ilustrovan ovim primerom).

U praksi, pored filtra visokopropusnika nad svim merenim EMG signalima se primenjuje još filter ili čak češalj filter kako bi se uklonile ne samo frekvencija na 50 Hz, već i svi harmonici koji su vidljivi u spektru signala. Ovi šumovi se relativno jednostavno filtriraju primenom linearnih filtera iz signal paketa [275] u programskom jeziku R. Funkcije za analizu signala po ugledu na rad u Matlabu i GNU Octave-u su dostupne u signal paketu [275]. Za odgovarajuću primenu *filter()* funkcije potrebno je uneti argumente koji predstavljaju koeficijente funkcije prenosa u Z domenu. U opštem slučaju, za filtre sa konačnim (eng. *Finite Impulse Response*, skraćeno FIR²²⁴) i beskonačnim (eng.

²²⁴ Prednost FIR sistema odnosno filtera u odnosu na IIR je što se mogu ostvariti karakteristike filtera koje su uz IIR ostvarive samo aproksimativno (na primer: linearna faza) [276].

Infinite Impulse Response, skraćeno IIR) impulsnim odzivom funkcija prenosa $H(z)$ u Z domenu ima oblik kao u Relaciji (2).

$$H(z) = \frac{Y(z)}{X(z)} = \frac{b(1) + b(2)z^{-1} + \dots + b(n_b+1)z^{-n_b}}{1 + a(2)z^{-1} + \dots + a(n_a+1)z^{-n_a}} \quad \text{Relacija (2)}$$



Slika 79. FFT EMG signala koji je u vremenskom domenu prikazan na srednjem panelu na Sl. 72 (uticaj šuma na 50 Hz je ilustrovan ovim primerom).

Funkcija *filter()* iz signal paketa [275] prati konvenciju programskih jezika GNU Octave i Matlab. U Relaciji (2) sa $Y(z)$ je označena Z transformacija signala $y(t)$ koji predstavlja izlaz iz filtra, a sa $X(z)$ je predstavljena Z transformacija signala $x(t)$ koji se dovodi na ulaz filtra. Sa $b(1), b(2), \dots, b(n_b+1)$ i sa $1, a(2), \dots, a(n_a+1)$ su predstavljeni koeficijenti filtra. Da bi se signal filtrirao, najpre je potrebno uneti koeficijente filtra, a potom proslediti signal u vremenskom domenu. Potrebno je obratiti pažnju da se rezultat u programskom jeziku R, ali i u GNU Octave i Matlab programskim jezicima dobija u vremenskom domenu, a ne u frekvencijskom domenu. Dve su napomene važne za filtriranje signala: (1) *filter()* funkciju ne treba primeniti, već *filtfilt()* funkciju i (2) koeficijente filtra ne treba ručno unositi (osim u retkim situacijama kada se raspolože funkcijom prenosa u Z domenu kao što je to prikazano u Relaciji (2)).

2.2.2.1 Distorzija faze signala usled filtriranja

Funkcija:

filtfilt()

se koristi u praksi zato što filtrira signal u dva prolaza i dva različita pravca. Ovakva realizacija filtra omogućava da se eliminiše distorzija faze signala koja nastaje kao rezultat filtriranja [275]. Iako bi teorijski ovo ispravljanje faze signala trebalo da bude idealno, u praksi to nije slučaj i ne dolazi do idealne korekcije faze signala. Sveukupno, primena *filtfilt()* funkcije će poboljšati fazu, ali ne savršeno. Kako bi se ilustrovao rad ove funkcije, prikazan je izvorni kod *filtfilt()* funkcije u R-u na Sl 80.

```
filtfilt <- function(filt, ...) UseMethod("filtfilt")

filtfilt.default <- function(filt, a, x, ...) {
  y = filter(filt, a, c(x, numeric(2 * max(length(a),
length(filt)))))
  y = rev(filter(filt, a, rev(y)))[seq_along(x)]
  y
}
```

Slika 80, Isečak izvornog koda (eng. *source code*) funkcije *filtfilt()* u R programskom jeziku, <https://rdr.io/cran/signal/src/R/filtfilt.R> (pristupljeno 23.02.2024).

Na Sl. 80 je za filtriranje u različitom smeru iskorišćena funkcija:

rev()

koja omogućava da se niz podataka prikaže u obrnutom redosledu (niz [1, 2, 3, 4, 5, 6, 7] bi nakon primene *rev()* funkcije bio transformisan u niz [7 6 5 4 3 2 1]). Filtriranje signala unosi kašnjenje, odnosno izlazni signal iz filtra je pomeren u vremenskom domenu u odnosu na ulazni signal. U slučaju da je faza signala linearna²²⁵, pomeraj je konstantan i korekcija kašnjenja se u naknadnoj (eng. *offline*) obradi signala može uvesti jednostavnim pomeranjem signala u vremenskom domenu. Faza signala je linearna u funkciji od frekvencije, a samo kod FIR filtera je moguće postići linearnu fazu, dok kod IIR filtera je moguće postići samo aproksimaciju linearnosti. U slučaju da se kašnjenja razlikuju u

²²⁵ https://en.wikipedia.org/wiki/Linear_phase, pristupljeno 23.02.2024.

odnosu na frekvencijski sadržaj signala (ili frekvencijske opsege) dolazi do distorzije faze. Kod signala koji su filtrirani se najčešće distorzija faze uočava tako što signal ima promenjen oblik u vremenskom domenu. Na primer, karakteristična talasnost PQRSST segmenata na EKG signalu menja svoj oblik, što nije beznačajno, jer se na taj način EKG zdrave osobe može pomešati sa EKG-om pacijenta ili obrnuto. Srećom, u većini slučajeva je fazna frekvencijska karakteristika, ili faza, pretežno nelinearna u nepropusnom opsegu, te se nelinearnost faze u odnosu na frekvenciju toleriše (kao što je to slučaj kod Batervortovog filtra). Kod filtara sa linearnom fazom mogu se javiti diskontinuiteti (na π ili 2π radijana), ali takve pojave ne narušavaju linearnost faze, odnosno faza je u tom slučaju deo po deo linearna (eng. *piecewise linear*), a diskontinuitet može biti posledica računanja faze signala. Funkcija *filtfilt()* omogućava kompenzaciju kašnjenja ili tzv. filtriranje sa nultom fazom (eng. *zero-phase filtering*). Filtar koji ima nultu fazu je poseban slučaj filtra sa linearnom fazom. Praktično, faza 0 znači da ne postoji ni distorzija faze, ali ni kašnjenje. Funkcija *filtfilt()* implementirana kao nekauzalna metoda (zavisi od budućih stanja) i ne može se primeniti u realnom vremenu, što je intuitivno jasno sa Sl. 80.²²⁶ [276-277]

Veoma je važno da se napomene da funkcija *filtfilt()* unosi, pored promene faze, kvadriranje magnitude i dupliranje reda filtra. Prema tome, ako funkcija prenosa u Z domenu $H(z)$ (Rel. 1) za projektovan filter sa odgovarajućim koeficijentima je trećeg reda, nakon primene *filtfilt()* funkcije, onda će biti 6. reda [237, 275, 278]. Kako bi se postigla ponovljivost rezultata, potrebno je da se u svim izveštajima i radovima u kojima se opisuje metodologija²²⁷

226 Za dodatna razmatranja o *filtfilt()* funkciji i fazi FIR i IIR filtara, pogledati predavanje "Metode za detekciju QRS kompleksa" za predmet Metode analize elektrofizioloških signala koji se drži na Elektrotehničkom fakultetu, Univerziteta u Beogradu (https://automatika.etf.bg.ac.rs/images/FAJLOVI_srpski/predmeti/izborni_kursevi_os/biomedicinsko_inzenjerstvo/OS4_OF4_MS1_MAS/2023%20nastava/09MAS23%20Metode%20za%20detekciju%20QRS%20kompleksa.pdf, pristupljeno 23.02.2024). Posebno obratiti pažnju na slajdove 30-59 koji su dodati i ažurirani u 2023/2024. školskoj godini.

227 Prilikom pisanja naučnih radova ili završnih studentskih radova, pa i pojedinih izveštaja i seminarskih radova, u upotrebi je IMRaD (eng. *Introduction, Methods, Results, and Discussion*) struktura rada [279]. Postoje i kritike ovakve organizacije radova koje se posebno zasnivaju na tome da IMRaD struktura ne prati način dolaska do rezultata [280]. Savet je studentima da se pridržavaju važećih pravilnika i uputstava svojih mentora pri izradi diplomskih i master radova, kao i doktorskih disertacija. Međutim, savet kolegama profesorima i istraživačima je da primene kritički osvrt, posebno imajući u vidu da je nemoguće ispoštovati IMRaD strukturu kada se u naučnim publikacijama predstavlja metodologija, pregled radova ili kada se piše kritički rad (odgovor uredniku), te da sami utiču na dalju evoluciju naučnih publikacija u svom okruženju.

eliminacije šuma navede da li je korišćena *filtfilt()* funkcija prilikom filtriranja signala.

Za razliku od FFT funkcije koja u svim programskim jezicima ima istu realizaciju, jer se zasniva na FFTW algoritmu (eng. *The Fastest Fourier Transform in the West*) koji je podeljen kao slobodan softver i koji su kreirali Mateo Frigo i Stiven Džonson sa Tehnološkog instituta u Masačusetsu²²⁸ [281], *filtfilt()* funkcija ima različite realizacije u programskim jezicima. Dodatno, FFTW algoritam je poznat i po tome što predstavlja jedan od primera dobre prakse dvojnog licenciranja softvera²²⁹ [14]. Naime, kada se softver deli pod otvorenom kopileft licencom kao što je GNU GPL (eng. *General Public License*) u tom slučaju se ne može koristiti u komercijalnim programskim jezicima ni u komercijalnom softveru. Jedini način da se softver zaštićen pod kopileft licencom iskoristi u komercijalnom softveru koji nije slobodan softver jeste primena metode dvojnog licenciranja [14]. Tako su u Matlabu bili primorani da traže dozvolu od FFTW Autora kako bi FFTW bio dodat Matlabu. Uz komercijalnu/dodatnu licencu, odnosno dozvolu, za koju su Autori FFTW algoritma dobili i novčanu nadoknadu, FFTW/FFT je postala dostupna i u Matlabu. Kada se primenjuje *filtfilt()* funkcija, Matlab i R programski jezici neće dati iste rezultate što može da ugrozi računarsku reproducibilnost, te je važno znati ovu činjenicu. Sve skripte programskih jezika koji nose oznaku slobodnog softvera su dostupne (funkcija *filtfilt()* u R-u se može detaljno proučiti kao što je pokazano na Sl. 80), dok je dostupnost skripti za funkcije vlasničkih programskih jezika ograničena i prepuštena volji privatnih kompanija²³⁰. Na žalost realizacija *filtfilt()* funkcije u programskom jeziku R nije idealna što je detaljno objašnjeno u [275] i pokazano kasnije kroz odziv filtra na jediničnu funkciju tzv. odskočni odziv, ali daje sasvim zadovoljavajuće rezultate za primenu u analizi biosignala.

228 Massachusetts Institute of Technology, <https://www.mit.edu/>, pristupljeno 16.09.2023.

229 FFTW, <https://tlo.mit.edu/technologies/fftw-fastest-fourier-transform-west>, pristupljeno 16.09.2023.

230 U Matlabu je moguće pristupiti relativno velikom broju izvornih kodova (ne smeju se svi koristiti bez posebne dozvole) primenom komande “edit”, te se, na primer, kucanjem komande “edit filtfilt” u komandnom prozoru Matlabu može videti da je osnovna razlika u primeni ove funkcije što Matlab realizacija smanjuje tranzijente dodatnom manipulacijom signala i postavljanjem početnih uslova (The Mathworks, Natick, SAD). Treba napomenuti da postojanje mogućnosti da se vidi izvorni kod ne povlači sa sobom pravo na promenu koda i njegovo dalje korišćenje, jer ne postoje licence slobodnog softvera pod kojima je omogućen pristup izvornom kodu.

2.2.2.2 Smernice za odabir filtra

Pre filtriranja biosignala, potrebno je izabrati tip filtra. Postoje tri ključna faktora za izbor filtra: (1) red filtra – za veći red filtra razdvajanje opsega je bolje, odnosno karakteristika je bliža idealnoj, koja vrši savršeno odsecanje u odabranoj frekvenciji, (2) frekvencijski opseg kojim se definiše jedna frekvencija odsecanja ili dve (u zavisnosti da li se realizuje niskopropusnik, visokopropusnik, nepropusnik opsega ili propusnik opsega) – za odabir frekvencije odsecanja važno je znati odlike signala i šuma i (3) tip filtra – potrebno je odabrati odgovarajuću frekvencijsku karakteristiku. Ovde nisu pobrojane sve karakteristike filtra koje se koriste u projektovanju u digitalnoj obradi signala, ali jesu najvažnije. Tri ključna parametra koji su nabrojani predstavljaju ulaz specijalizovanih funkcija koje na svom izlazu daju koeficijente filtra u Z domenu, te ih je moguće koristiti kasnije sa *filtfilt()* funkcijom. Dodatni detalji digitalne obrade signala sa prikazom najvažnijih prozorskih funkcija prevazilaze gradivo koje je pokriveno ovim udžbenikom, te se čitaoci upućuju na predmete osnovnih studija i specijalizovanu literaturu [276-277, 282-284].

Ovde će biti prikazano kako je za filtriranje šuma pokreta na EMG signalu potrebno realizovati filter u programskom jeziku R. Ali, pre prelaska na primer, potrebno je podsetiti studente na par elemenata iz digitalne obrade signala koji su važni za razumevanje praktičnih koncepata u ovom udžbeniku. U obradi signala uobičajeno se koriste frekvencijske metode koje rezultuju frekvencijskom karakteristikom. Pod krovnim nazivom frekvencijska karakteristika definišu se amplitudska (ili magnitudska) frekvencijska karakteristika i fazna frekvencijska karakteristika koje prikazuju zavisnost amplitude/magnituda i faze u odnosu na frekvenciju. U sistemima upravljanja, frekvencijska metoda je veoma pogodna, jer se može jednostavno eksperimentalno dobiti kada se sistem pobudi prostoperiodičnim signalom, jer se na izlazu tada mere amplituda i faza u stacionarnom režimu. Dodatno, za filtre su, pored frekvencijske karakteristike, važni impulsni odziv (kako se sistem ponaša kada se na ulazu nalazi Dirakov impuls/funkcija²³¹ kao pobuda) i odziv na jediničnu funkciju (odskočni odziv – ponašanje sistema kada se na ulazu nalazi Hevisajdova funkcija²³²), kao i provera lokacije nula i polova funkcije prenosa. Frekvencijske karakteristike filtera su najčešće korisne u obradi signala, jer omogućavaju da se na jasan način predstavi kako filter ili sistem u opštem slučaju utiču na frekvencijski sadržaj korisnog signala i šuma. [186, 276, 285-286]

231 https://en.wikipedia.org/wiki/Dirac_delta_function, pristupljeno 08.03.2024.

232 https://en.wikipedia.org/wiki/Heaviside_step_function, pristupljeno 08.03.2024.

Ne postoji recept po kome se može odabrati red filtra, ali upravo zato je oblast obrade signala zanimljiva i zahtevna. Potrebno je iskustvo, a nekada ni iskustvo nije dovoljno, već je potrebno primeniti različite filtre dok se ne donese konačna odluka koji će se filter koristiti. Pored toga, treba uzeti u obzir da korišćenje *filtfilt()* funkcije duplo uvećava red, kao i da postoji veza između frekvencije odsecanja i reda filtra. Na primer, pod pretpostavkom da su šum i koristan deo signala razdvojeni (ne postoji preklapanje u frekvencijskom domenu), te ako je šum u opsegu do 10 Hz, a koristan deo signala u opsegu od 20 Hz i ako se frekvencija odsecanja filtra visokopropusnika postavi na 15 Hz onda bi manji red filtra bio dovoljan. Ali, ako se frekvencija odsecanja postavi na 20 Hz onda je potreban veći red filtra kako ne bi došlo do filtriranja/eliminacije korisnog dela spektra signala. Kako bi se uklonio šum pokreta može se koristiti filter 2. reda sa frekvencijom odsecanja u 10 Hz kako su preporučili Melo i saradnici za filtriranje EMG signala [287]. Obzirom da će se koristiti filtriranje u oba smera sa ciljem eliminacije distorzije faze signala, to će red rezultujućeg filtra biti jednak 4. Frekvencijsku karakteristiku digitalnog filtra moguće je prikazati primenom funkcije:

freqz()

iz signal paketa. Funkcija *freqz()* se primenjuje nad koeficijentima filtra u Z domenu koji se mogu dobiti primenom specijalizovanih funkcija. Na Sl. 81 dat je R kod za prikaz frekvencijske karakteristike tri Batervortova filtra za 2, 4. i 8. red. Da bi se dobili koeficijenti koji odgovaraju funkciji prenosa u Z domenu za Batervortov filter, potrebno je koristiti funkciju:

butter()

koja ima tri osnovna ulazna parametra:

- red filtra,
- frekvenciju odsecanja normalizovanu na Nikvistovu frekvenciju (za 10 Hz, potrebno je pisati $10/(fs/2)$, ako je sa fs označena frekvencija odabiranja kao na Sl. 81) i
- tip filtra (na Sl. 81 je prikazan kod za generisanje visokopropusnika tj. *high pass* filtra).

Rezultujući grafik koji prikazuje uticaj reda filtra na frekvencijsku karakteristiku Batervortovog filtra visokopropusnika sa frekvencijom odsecanja od 10 Hz je prikazan na Sl. 82.

```

fs <- 1000 # frekvencija odabiranja
filt2 <- butter(2, 10/(fs/2), "high")
filt4 <- butter(4, 10/(fs/2), "high")
filt8 <- butter(8, 10/(fs/2), "high")
f2 <- freqz(filt2, Fs = fs)
f4 <- freqz(filt4, Fs = fs)
f8 <- freqz(filt8, Fs = fs)

par(mar = c(4, 4, 0.1, 0.1))
plot(f2$f, abs(f2$h), type="l", lwd=3, lty=1, # kontinualna linija (eng. solid)
     xlab="frekvencija [Hz]", ylab="magnituda [a.u.]",
     ylim=c(0, 1), xlim=c(0, 40))
par(new = T)
plot(f4$f, abs(f4$h), type="l", lwd=3, lty=2, # isprekidana linija (eng. dashed)
     xlab="frekvencija [Hz]", ylab="magnituda [a.u.]",
     ylim=c(0, 1), xlim=c(0, 40))
par(new = T)
plot(f8$f, abs(f8$h), type="l", lwd=3, lty=3, # tačkasta linija (eng. dotted)
     xlab="frekvencija [Hz]", ylab="magnituda [a.u.]",
     ylim=c(0, 1), xlim=c(0, 40))
grid()
legend(0, 1, legend=c('2. red','4. red','8. red'),
      lty=1:3, lwd=3)

```

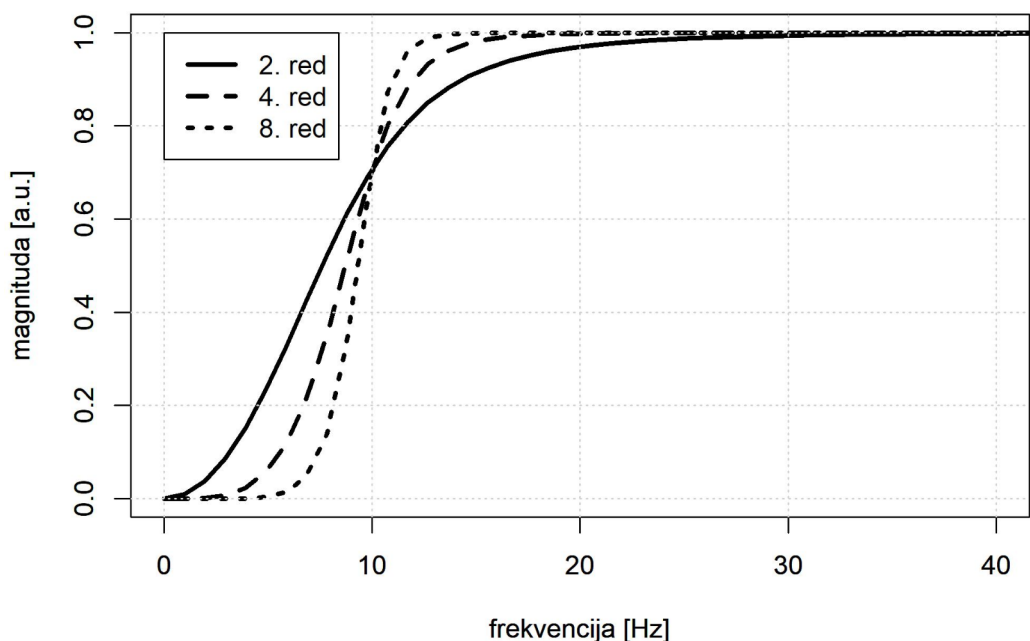
Slika 81, R kod za prikaz frekvencijske karakteristike Batervortovog filtra 2, 4. i 8. reda. Pored argumenta *lty* (od eng. *line type*) kojim se određuje tip linije unutar *plot()* funkcije dodati su komentari radi pojašnjenja tipa linije za prikaz grafika. Za više informacija pogledati <http://www.sthda.com/english/wiki/line-types-in-r-lty>, pristupljeno 23.02.2024.

Za filtriranje biosignala u najvećoj meri se koristi Batervortov²³³ filter koji ima ravnu karakteristiku u propusnom opsegu (eng. *flat magnitude in the passband*²³⁴) [288]. Naravno, nije neophodo koristiti isključivo Batervortov filter, jer postoje situacije u kojima su drugi filtri primereniji. Na primer, Čebiševljev filter²³⁵ postiže bolje razdvajanje propusnog i nepropusnog dela, te se obično koristi za izdvajanje karakterističnih opsega EEG signala, jer se alfa i beta talasi graniče u frekvencijskom domenu: alfa je u opsegu od 8 Hz do 12 Hz, dok je beta u opsegu od 12 Hz do 35 Hz [289], pa je potrebno dobro razdvajanje opsega na zadatoj frekvenciji. Primititi kako filtri nižeg reda (Sl. 82) lošije razdvajaju opsege na frekvenciji odsecanja od 10 Hz za razliku od filtera višeg reda. Dakle, za filtriranje signala, ali i eliminaciju šuma generalno, nema jasnog uputstva ni recepta, postoje samo smernice za primenu odgovarajućih metoda, gde treba imati na umu pozitivne i negativne strane svake od metoda, ali i prirodu signala.

233 Stiven Batervort je prvi predložio filter koji je dobio naziv po njemu (https://en.wikipedia.org/wiki/Stephen_Butterworth, pristupljeno 16.09.2023).

234 https://en.wikipedia.org/wiki/Butterworth_filter, pristupljeno 16.09.2023.

235 Čebiševljev filter je dobijen iz polinoma Pafnutija Čebiševa, te je po Pafnuti Čebiševu filter i dobio ime (https://en.wikipedia.org/wiki/Pafnuty_Chebyshev, pristupljeno 16.09.2023).



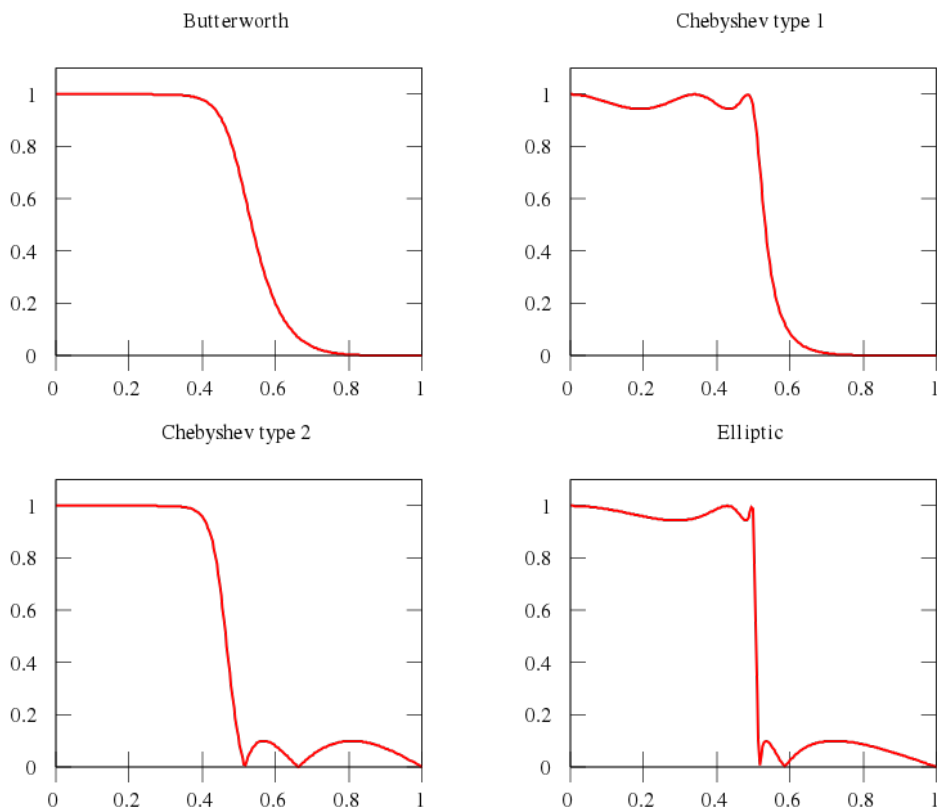
Slika 82, Uporedni prikaz frekvencijskih karakteristika Batervortovog filtra visokopropusnika 2, 4. i 8. reda sa frekvencijom odsecanja na 10 Hz. R kod koji generiše prikazan grafik je dat na Sl. 81. Iako je odabrano da je frekvencija odabiranja 1000 Hz (frekvencijska karakteristika filtra se posmatra do 500 Hz), ovde je zumiran samo deo karakteristike do 40 Hz.

Kada se uporedi sa drugim tipovima filtara, Batervortov filter istog reda sa Čebiševim filtrom tipa 1 i tipa 2, kao i sa eliptičkim filtrom, Batervort filter ima najlošije razdvajanje propusnog i nepropusnog opsega, ali ima ravnu magnitudsku frekvencijsku karakteristiku u propusnom i nepropusnom opsegu (Sl. 83). Za razliku od Batervortovog filtra, Čebiševljevi filtri tipa 1 i tipa 2 imaju ravnu karakteristiku ili u propusnom ili u nepropusnom delu opsega (Sl. 83), dok eliptički filter ima najbolje razdvajanje opsega, ali nema ravnu karakteristiku ni u propusnom ni u nepropusnom opsegu. Ovo dalje znači da se bolje razdvajanje propusnih opsega, bilo da se radi o eliminaciji šuma (eliminacija interferencije na 50 Hz) ili izdvajanju segmenta od značaja (alfa opseg EEG signala) može postići primenom Čebiševljevih filtara.

Na Sl. 83 je ilustrovano kako nije jednostavno postići kompromis prilikom izbora tipa digitalnih filtara – dok eliptički filter od svih prikazanih na slici ima najbolje razdvajanje opsega pri tom nema ravnu karakteristiku u oba opsega (nepropusnom i propusnom). Odabir filtra se svodi na staru izreku da kompromis vodi ka sigurnom nezadovoljstvu obe strane²³⁶. U tom smislu,

²³⁶ Iako se sličan izreka prepisuje Lariju Dejvidu “A good compromise is when both parties are dissatisfied” (https://en.wikiquote.org/wiki/Talk:Henry_Clay, pristupljeno 16.09.2023),

Batervortov filter predstavlja dobar kompromis između razdvajanja frekvencijskih opsega i ravne karakteristike, te će se ovde prikazati realizacija Batervortovog filtra 2. reda sa frekvencijom odsecanja 10 Hz za filtriranje šuma pokreta [287].



Slika 83, Frekvencijske karakteristike za Batervortov, Čebišev tipa 1 i tipa 2, kao i za eliptički filter. Autor slike je Alesio Damato – sopstveno delo. Slika je podeljena pod CC BY-SA 3.0 licencom, <https://commons.wikimedia.org/w/index.php?curid=427738>, pristupljeno 16.09.2023. Obe ose (apscisa i ordinata) su normalizovane na svim graficima.

Jedna od mana Batervortovog filtra je što može da izazove artefakt talasanja/zvonjenja (eng. *ringing*), jer ima relativno loš odziv na jediničnu funkciju tj. ima loš odskočni odziv (eng. *step response*) što se može videti na Sl. 84. Efekat talasanja je veći kada na signalu postoje nagli prelazi (odgovor na jediničnu funkciju zato daje nagli prelaz na početku rezultujućeg signala kao što se vidi na levom panelu na Sl. 84 ili i na početku i na kraju u slučaju dva prolaza filtra primenom *filtfilt()* funkcije što se vidi na desnom panelu na Sl.

originalna referenca ka ovom citatu još uvek nije pronađena.

84). Ovaj fenomen se takođe naziva i Gibsov efekat ili Gibsove oscilacije²³⁷, a vidljive su naročito na amplitudskoj frekvencijskoj karakteristici MA filtra, jer na njihovu pojavu utiču nagle promene na signalu (kod MA filtra to je pravougaoni oblik prozora). Nepoželjne su i posebno izražene u neposrednoj blizini granične frekvencije. Jedan od načina da se prevaziđe ovaj problem jeste da se koriste druge vrste filtra kao što su Beselov ili Gausov filter (misli se na oblik magnitudske karakteristike, odnosno frekvencijske karakteristike), ali i da se koristi filtriranje u oba smera primenom *filtfilt()* funkcije čiji je efekat prikazan na Sl. 84. [276, 290]

Na Sl. 84 je prikazano kako *filtfilt()* funkcija drugačije utiče na efekte talasanja odskočnog odziva. Obratiti pažnju da može doći do razlika prilikom korišćenja *filtfilt()* funkcije u drugim programskim jezicima, kako je već navedeno u prethodnim potpoglavljima, te će odziv na jediničnu funkciju biti drugačiji ako se ova analiza uradi u, na primer, u GNU Octave-u. Kada se ispituje efekat *filtfilt()* funkcije u odnosu na *filter()* funkciju u GNU Octave-u, razlike u odskočnim odzivima su drastične te je odmah uočljiv učinak *filtfilt()* funkcije, dok realizacija u R-u nije idealna. Čitaoci mogu samostalno uneti dodatne izmene u *filtfilt()* funkciju otvorenog koda (Sl. 80) kako bi dodatno popravili ovu karakteristiku.

Jednostavan način da se dobije odziv filtra na jediničnu funkciju jeste da se filtrira signal sastavljen od jedinica (kompletan R kod dat je u pratećoj R skripti, a ovde su prikazane samo najvažnije funkcije). Za primenu *filter()* funkcije potrebno je pokrenuti sledeći kod:

```
filt2 <- butter(2, 10/(fs/2), "high"); filter(filt2, c(0.5, rep(1, n-1)))
```

dok je za predstavljanje odziva na jediničnu funkciju potrebno dodati *filtfilt()* funkciju za filtriranje signala:

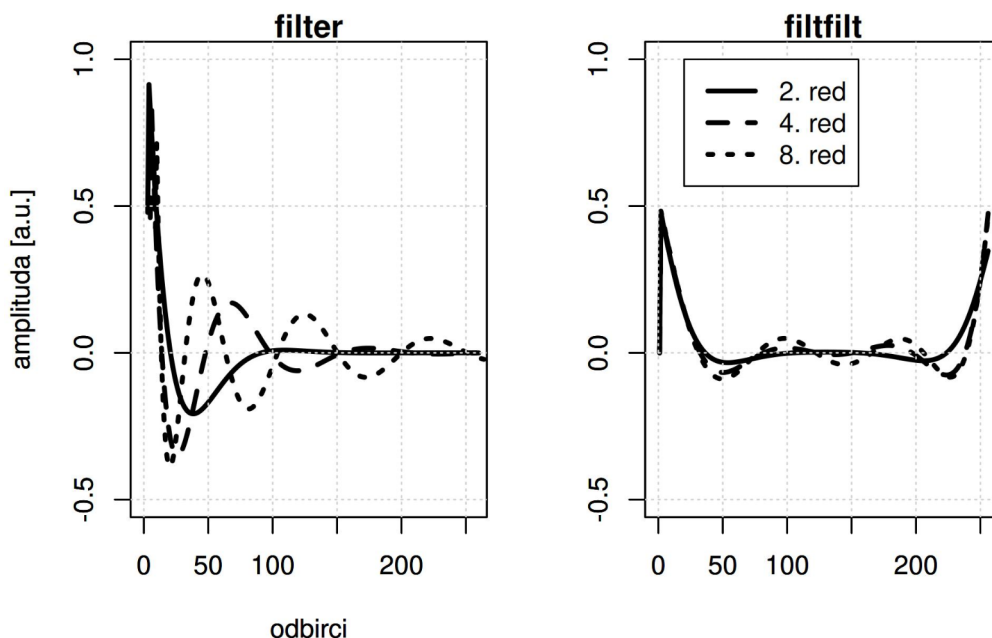
```
filt2 <- butter(2, 10/(fs/2), "high"); filtfilt(filt2, c(0.5, rep(1, n-1)))
```

Primetiti da se koristi *rep()* funkcija koja služi da se kreira niz sastavljen iz jedinica (u ovom slučaju je odabran broj tačaka za odziv na Sl. 84 od $n = 2^8$), kao i da je prikazan filter 2. reda visokopropusnik sa frekvencijom odsecanja od 10 Hz (*fs* predstavlja frekvenciju odabiranja). Korišćenje *rep()* funkcije je objašnjeno u ranijim poglavljima. Ovde je za potrebe ilustracije prikazan filter visokopropusnik, ali se najčešće efekat talasanja u odzivu na jediničnu funkciju

237 https://en.wikipedia.org/wiki/Gibbs_phenomenon, pristupljeno 30.05.2024.

ispituje za filter niskopropusnik. Takođe, na Sl. 84 su odzivi prikazani linijama, ali u praksi se najčešće prikazuju sa tačkama (Autorka je odabrala linijski pristup da bi Sl. 84 bila u skladu sa Sl. 82). Takođe, obratiti pažnju da signal koji se filtrira ima 0.5 u koordinatnom početku, odnosno u nuli, što je i podrazumevana računarska realizacija Hevisajdove funkcije u drugim programskim jezicima (Matlab/GNU Octave). Međutim, u signal paketu u R-u ne postoji posebna Hevisajdova funkcija (za te potrebe mogu se koristiti drugi R paketi), ali se zato ona jednostavno realizuje kao:

```
c(0.5, rep(1, n-1))
```

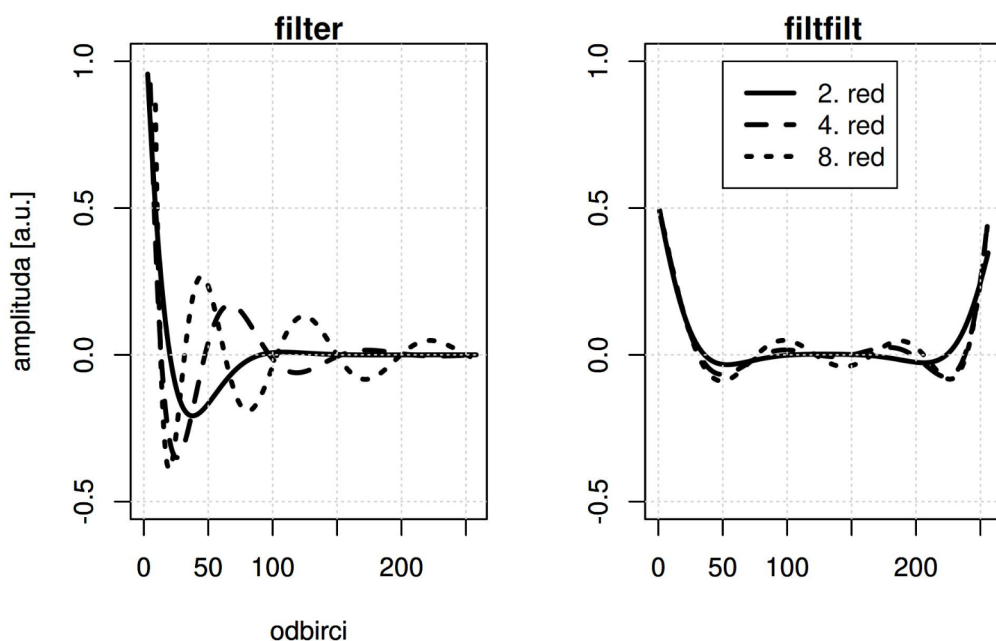


Slika 84, Odskočni odziv Batervortovog filtra na jediničnu funkciju korišćenjem *filter()* i *filtfilt()* funkcija. Kao na Sl. 82, korišćeni su Batervort filteri visokopropusnici 2, 4. i 8. reda sa frekvencijom odsecanja na 10 Hz i frekvencijom odabiranja od 1000 Hz. U ovom slučaju Hevisajdova funkcija je jednaka 0.5 u 0.

Na Sl. 84 je prikazan odskočni odziv na ovu realizaciju Hevisajdove funkcije (sa vrednošću 0.5 u 0), dok je na Sl. 85 prikazan odskočni odziv u slučaju da je Hevisajdova funkcija u 0 jednaka 1 (gde je uočljiviji efekat *filtfilt()* funkcije na odskočni odziv u odnosu na odzive na Sl. 84).

Vrlo jednostavna ilustracija generisanog ulaza (jedinične funkcije) Batervortovog filtra daje uvid kako se sistem ponaša, kao i kakve promene filter može da unese u signal. Prilikom odabira odgovarajuće metoda za analizu

signala ili filtra za eliminaciju šuma, važno je simulirati različite situacije. Pored odskočnog, može se posmatrati i impulsni odziv (manje je interesantan u analizi biomedicinskih signala), ali i ponašanje sistema u slučaju simuliranih biosignala [291], što je tema posebnog poglavlja u ovom udžbeniku. Promene koje se unose u signal primenom metoda analize mogu biti opasnije po rezultat analize od šuma, te bi trebalo posvetiti pažnju razumevanju i preciznom dokumentovanju efekata koje na signal ostavljaju filtri.

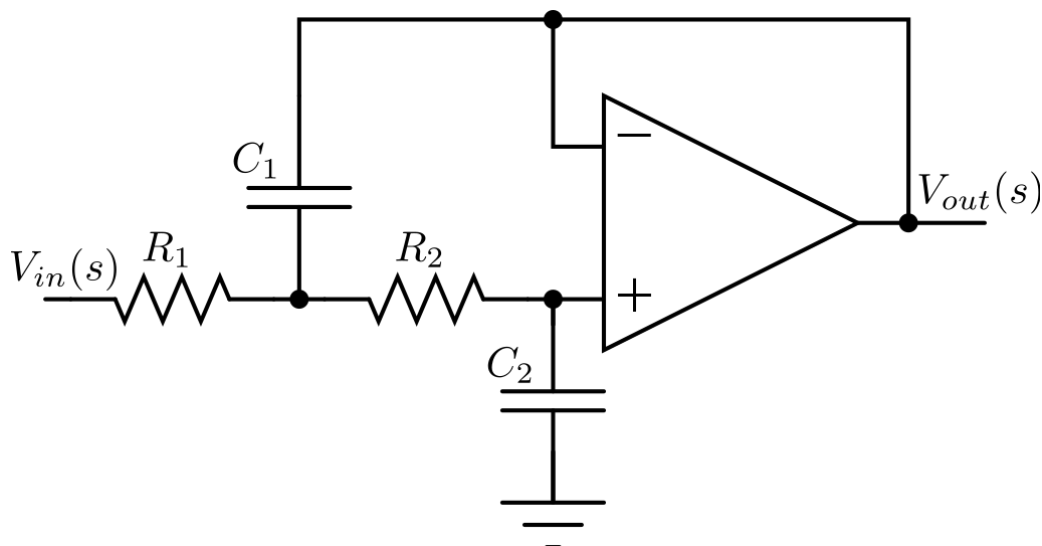


Slika 85, Odskočni odziv Batervortovog filtra na jediničnu funkciju korišćenjem *filter()* i *filtfilt()* funkcija. Kao na Sl. 82, korišćeni su Batervort filtri visokopropusnici 2, 4. i 8. reda sa frekvencijom odsecanja na 10 Hz i frekvencijom odabiranja od 1000 Hz. U ovom slučaju Hevisajdova funkcija je jednaka 1 u 0.

Pored softverske realizacije, Batervortov filter se može realizovati i hardverski, a na Sl. 86 je prikazana realizacija filtra drugog reda u Salen-Kej topologiji²³⁸ koja koristi aktivne i pasivne električne komponente. Jednostavnim odabirom otpornika i kondenzatora (R_1 , R_2 , C_1 i C_2) moguće je podesiti frekvenciju odsecanja filtra. Funkcija prenosa filtra $H(s)$ kao odnos transformacije izlaznog $V_{out}(s)$ i ulaznog napona $V_{input}(s)$ je prikazana u Relaciji (3). Za više informacija o realizaciji analognih filtara studenti mogu pogledati [292].

²³⁸ https://en.wikipedia.org/wiki/Sallen%E2%80%93Key_topology, pristupljeno 23.02.2024.

$$H(s) = \frac{V_{out}(s)}{V_{input}(s)} = \frac{1}{1 + C_2(R_1 + R_2)s + C_1 C_2 R_1 R_2 s^2} \quad \text{Relacija (3)}$$



Slika 86, Hardverska realizacija Batervortovog filtra 2. reda. Autora Đulijama Simarda – sopstveno delo, javno vlasništvo, <https://commons.wikimedia.org/w/index.php?curid=4275147>

2.2.2.3 Primena Batervortovog filtra za eliminaciju šuma

Striktno gledano, u digitalnoj obradi signala ne postoji filter visokopropusnik, jer su digitalni signali ograničeni Nikvistovom frekvencijom, odnosno polovinom frekvencije odabiranja, te je filter visokopropusnik u stvari filter propusnik opsega. Takođe, instrumentacioni pojačavači koji se koriste u biomedicinskom inženjerstvu za merenje elektrofizioloških signala imaju propusni frekvencijski opseg na kome mogu da pojačavaju signale, tako da i oni uvode ograničenja u maksimalnoj frekvenciji/opsegu frekvencija na signalu. Na primer, INA121 instrumentacioni pojačavač čija je osnovna primena za pojačanje fizioloških signala (EKG, EMG, EEG i signal disanja) u zavisnosti od pojačanja propušta frekvencijski opseg signala od 10 kHz do 1 MHz²³⁹.

EMG signal je za potrebe ilustracije u ovom udžbeniku filtriran sa dva filtra: Batervortovim visokopropusnikom sa ciljem eliminacije artifakta pokreta i Batevortovim nepropusnikom opsega od 49 Hz do 51 Hz sa ciljem eliminacije elektromagnetnih smetnji. Nakon što se dobiju koeficijenti za funkciju prenosa filtra u Z domenu, primenom *butter()* funkcije, filtriranje se vrši ili primenom

²³⁹ INA121, <https://pdf1.alldatasheet.com/datasheet-pdf/view/847621/TI1/INA121.html>, uputstvo, pristupljeno 24.02.2024.

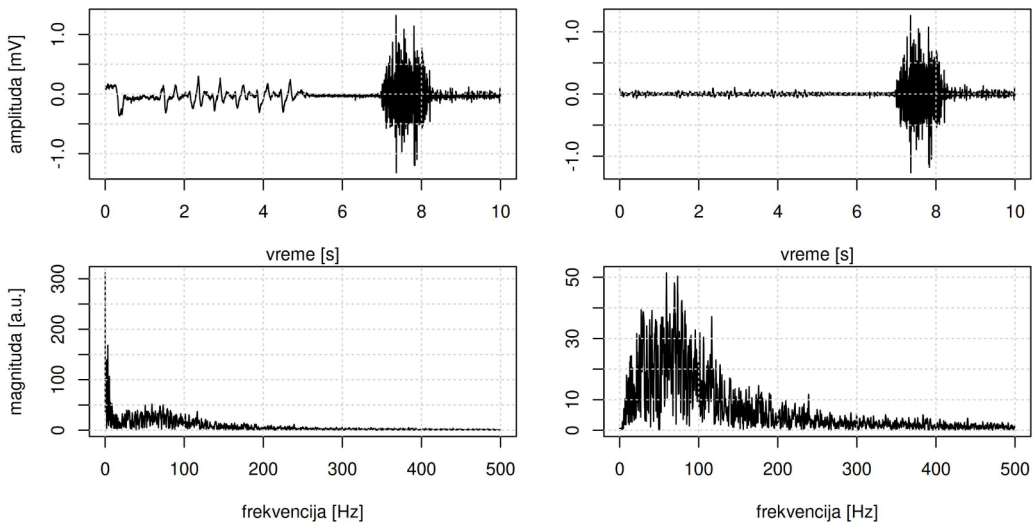
`filter()` ili `filtfilt()` funkcije. Realizacija filtera u programskom jeziku R je prikazana na Sl. 87.

```
library(signal)
datEMGkablAll <- read.table("emg-sve.txt")
fs <- 1000
datEMGkabl <- datEMGkablAll$V1[(4*fs):(14*fs)]
timeAxisKabl <- seq(0, length(datEMGkabl)/fs - 1/fs, by = 1/fs)

fb <- butter(2, 10/(fs/2), "high")
f50 <- butter(3, c(49/(fs/2), 51/(fs/2)), "stop")
datEMGkablFilt <- filtfilt(f50, filtfilt(fb, datEMGkabl))
```

Slika 87, Filtriranje EMG signala sa dva filtera sa ciljem eliminacije šuma koji potiču od artefakta pokreta i elektromagnetne interferencije na 50 Hz.

Rezultat primene `filtfilt()` funkcije na EMG signal (Sl. 87) je sačuvan u promenljivoj `datEMGkabl` koja sadrži filtrirani signal u vremenskom domenu koji je smešten u promenljivoj `datEMGkablFilt`. Na Sl. 88 prikazan je signal u vremenskom i frekvencijskom domenu pre i posle filtriranja. Može se primetiti da filter nepropusnik opsega na 50 Hz nije ni bio potreban, jer signal nije inicijalno sadržao šum na 50 Hz. Sa druge strane filter visokopropusnik je uklonio veći deo šuma, ali šum nije nestao u potpunosti.



Slika 88, Prikaz signala pre filtriranja u vremenskom domenu (gornji panel levo) i u frekvencijskom domenu (donji panel levo), kao i signala nakon filtriranja u vremenskom domenu (gornji panel desno) i u frekvencijskom domenu (donji panel desno). Parametri filtera su dati u tekstu i na Sl. 87.

Kada bi se uvećao deo spektra filtriranog signala u okolini 50 Hz primetili bi se efekti filtra nepropusnika na spektru signala. Primetiti da EMG signal nakon ~250 Hz nema značajnijeg sadržaja (na tim frekvencijama ne postoji ni koristan deo signala, a ni šum), te nije bilo potrebno primeniti filter niskopropusnik. U slučajevima kada je potrebno da se primeni filter propusnik opsega, postoje dve realizacije koje bi trebalo uzeti u obzir: (1) kroz jedan poziv filtra propusnika opsega i (2) kaskadnom/rednom realizacijom filtra visokopropusnika i filtra niskopropusnika, odnosno eng. *low pass* filtra. Kaskadna realizacija, odnosno odvojena realizacija niskopropusnika i visokopropusnika je fleksibilnija, jer pored toga što se mogu nezavisno birati redovi filtera, mogu se birati i tipovi filtera u skladu sa potrebama i prikazom tipova filtera, odnosno njihovih amplitudskih ili magnitudskih (u opštem slučaju frekvencijskih) karakteristika prikazanih na Sl. 83. Pored ovog načina konstrukcije filtra propusnika opsega, postoje i drugi u literaturi a svode se na to da je sve tipove filtera moguće kreirati primenom filtra niskopropusnika. [286, 293]

2.2.2.4 Primena MA filtra za računanje obvojnice

Za one koji se prvi put susreću sa EMG signalom, trebalo bi spomenuti da EMG signal u vremenskom domenu “podseća” na audio signal. Zbog sličnosti sa audio signalom, za analizu EMG signala se mogu koristiti i metode koje se tradicionalno koriste za obradu audio signala kao što su Markovljevi modeli [294]. EMG signal, u opštem slučaju i tehnika merenja EMG signala, se može koristiti za medicinska istraživanja (u ortopediji, hirurgiji, funkcionalnoj neurologiji, kao i u analizi hoda i držanja tela od eng. *posture*), za rehabilitaciju (u post-hirurškom oporavku, neurološkoj rehabilitaciji, fizikalnoj terapiji i u drugim oblicima terapije), za ergonomiju (u prevenciji rizika, ergonomskom dizajnu, sertifikaciji proizvoda i u ergonomskoj analizi) i za sport (u biomehanici, analizi pokreta, sportskoj rehabilitaciji i u kondicionom treniranju/obučavanju) [105]. Ukratko, EMG se može koristiti za procenu stanja nervno-mišićnog sistema, ali i za terapiju.

Posebno je interesantna primena EMG signala u biološkoj povratnoj sprezi ili biofidbeku (eng. *biofeedback*), gde ispitanik ima mogućnost da vidi ili sirov signal u realnom vremenu ili neki od parametara (na primer, srednju vrednost obvojnice ili srednju vrednost ispravljenog signala), te da u realnom vremenu koriguje mišićnu kontrakciju i da na taj način direktno utiče na neuralne mehanizme. Biofidbek omogućava da povratna informacija u ljudskom organizmu bude ili pojačana ili dodata uticajem spoljašnjih faktora. Osim EMG signala, za biofidbek se može koristiti i EKG signal, a najčešće u formi srčanog ritma. Iz iskustva Autorke daleko je jednostavnije upravljati EMG signalima

nego srčanim ritmom na voljnom nivou (korišćenjem biofidbek tehnike), a zanimljiva je situacija u kojoj se od ispitanika traži da relaksiraju mišić umesto da ga kontrakuju, jer je praktično pokazano da je relaksacija kompleksniji zadatak od aktivacije. [295-297]

Jedan od najčešćih koraka u analizi EMG signala, bilo da se informacija od interesa koristi za terapiju ili dijagnostiku/monitoring, je računanje obvojnice EMG signala koja se, zajedno sa ispravljenim EMG signalom, može koristiti i za kvalitativnu/vizuelnu procenu karakteristika mišićne kontrakcije [298]. Uređaji za merenje EMG signala iz 80ih godina 20. veka su sadržali dodatna analogna kola za određivanje obvojnice²⁴⁰ (eng. *envelope*) EMG signala, najčešće primenom integratora [299]. Memorija uređaja u računarima 80ih godina prošlog veka je iz današnjeg ugla bila skromna, te je ovo bio jedan od načina uštede memorijskog prostora. Naime, obvojnica je značajno nižeg frekvencijskog opsega, te ju je moguće digitalizovati sa manjom frekvencijom odabiranja, što rezultuje manjim brojem odbiraka i najvažnije manjom memorijom koju je potrebno skladištiti. U poglavlju u kome su dati primeri uvoza podataka u R pokazano je da je za skladištenje 30 s EMG signala sa 20 kanala koji su digitalizovani sa frekvencijom odabiranja od 1 kHz, potrebno 4.58 MB, dok je na flopi disketi koja se u to vreme koristila za skladištenje podataka moglo da stane 1.44 MB²⁴¹ [300]. Kada se signal digitalizuje primenom A/D konvertora kako bi se omogućila dalja obrada na računaru (digitalna obrada signala), bez obzira da li je prethodno filtriran ili pojačan, takav signal se naziva neprocesiran/neobrađen EMG signal ili sirov (od eng. *raw*) EMG signal [105].

Iz obvojnice EMG signala se može relativno jednostavno odrediti početak i kraj mišićne kontrakcije, kao najčešće korišćena obeležja u vremenskom domenu [105]. Da bi se dobila obvojnica EMG signala potrebno je dvostrano ispraviti EMG signal. Prilikom ispravljanja EMG signala menja se srednja vrednost signala. Pre ispravljanja, srednja vrednost EMG signala je jednaka 0 (ovo je teorijski tačno, ali šum, kao što je artefakt pokreta, može uneti promenu srednje vrednosti²⁴²), a posle ispravljanja srednja vrednost je veća ili manja od 0. Nakon

240 U srpskom jeziku se na žalost odomaćio i termin *anvelopa*. Savetuje se studentima da koriste termin *obvojnica*, ali radi jednostavnijeg praćenja literature na srpskom jeziku potrebno je znati da je *anvelopa* takođe zastupljena, iako se radi o pogrešno prevedenom terminu.

241 Ovo je ujedno i razlog najveće kritike termina "*Big Data*", jer ono što je danas trivijalno odnosno sadrži relativno malo memorije, 80ih i 90ih godina 20. veka se moglo smatrati velikim podacima.

242 Šum napajanja (50 Hz ili 60 Hz) ima srednju vrednost 0, te je važno napomenuti da ne unosi svaki šum promenu srednje vrednosti EMG signala.

ispravljanja signala, da bi se dobila obvojnica EMG signala, može se primeniti neka od tehnika koja omogućava glačanje EMG signala (eng. *smoothing*) kao [105]:

- filtriranje niskopropusnikom,
- usrednjavanje signala primenom pokretnih prozora pomoću linearnih filtara (npr. primenom pravougaonih prozora kao što je eng. *Moving Average*, skraćeno MA) ili primenom nelinearnih filtara (npr. računanjem medijane na prozorima koji se primenjuju u vremenskom domenu) i
- integracija (uobičajena je za hardversku realizaciju računanja obvojnice preko sistema prekidača kojima je omogućeno punjenje i pražnjenje kondenzatora [299]).

U ovom udžbeniku je odabran primer linearnog filtra sa pokretnim usrednjavanjem (MA) sa pravougaonim prozorom, koji se primenjuje u vremenskom domenu za izdvajanje obvojnice. U praksi, Autorka je uvek koristila filtar niskopropusnik za izdvajanje obvojnice. Međutim, u zavisnosti od SNR-a i uslova merenja pogodnije mogu biti i druge metode (na primer, medijana filtar [274]). MA filtar je odabran iz dva razloga: (1) linearan je, te mu se može odrediti funkcije prenosa i jednostavno ga je objasniti analitički i (2) predstavlja jedan od najčešće korišćenih filtara u elektrotehnici i računarskoj tehnici²⁴³ zbog svoje jednostavnosti (ranije je već prikazana alatka Ngram Viewer koja sadrži ugrađen MA filtar za prikaz rezultata pretrage po ključnim rečima). Filtar sa pokretnim usrednjavanjem (MA filtar) se koristi kako bi se smanjile fluktuacije na signalu (koje u biomedicinskim signalima najčešće potiču od visokofrekventnog šuma) tokom nekog vremenskog perioda (rezultat MA filtra zavisi od prošlih podataka) [67]. Učinak MA filtra je analogan filtru niskopropusniku, a mana MA filtra je što ne vrši dobro razdvajanje signala u frekvencijskom domenu, ali je zato u stanju da filtrira šum koji se u frekvencijskom spektru poklapa sa signalom u slučaju da postoji šum čija je srednja vrednost jednaka nuli. Iako je srednja vrednost EMG signala jednaka nuli, MA filtar se primenjuje na dvostrano ispravljenom signalu (apsolutna vrednost signala), te u tom slučaju EMG signal nema srednju vrednost jednaku nuli za razliku od šuma.

Najjednostavniji način ilustracije rada MA filtra je dat u Relaciji (4) koja prikazuje zavisnost izlaznog/filtriranog signala y od ulaznog signala x za

²⁴³ MA se primenjuje i u drugim oblastima kao što su ekonomija i okeanografija. Klasičan primer iz ekonomije uključuje kreditni račun gde kamata zavisi od tromesečnog ili šestomesečnog euribora (evropska međubankarska stopa).

filtriranje MA filtrom na prozoru dužine M odbiraka. Primetiti da je u Relaciji (4) prikazan način kako se dolazi do jednog odbirka izlaznog signala $y[i]$, te se proces dalje ponavlja pomeranjem prozora dužine M u vremenskom domenu [301].

$$y[i] = \frac{1}{M} \sum_{j=0}^{M-1} x[i+j] \quad \text{Relacija (4)}$$

Dakle, da bi se dobio 13. odbirak izlaznog filtriranog signala y na prozoru dužine $M = 5$, potrebno je pronaći srednju vrednost 5 odbiraka ulaznog signala x , kao što je to prikazano u Relaciji (5).

$$y[13] = \frac{x[13] + x[12] + x[11] + x[10] + x[9]}{5} \quad \text{Relacija (5)}$$

Alternativno, Relacija (5) se može predstaviti tako da se tekuća tačka (13. odbirak) računa simetrično kao u Relaciji (6). Ovakva simetrična realizacija bi odgovarala promeni granica sume u Relaciji (5) od $-(M - 1) / 2$ do $(M - 1) / 2$. Jednostavno se dolazi do zaključka da MA filter predstavlja konvoluciju ulaznog signala x sa pravougaonom funkcijom koja u 5 tačaka ima vrednost $1/5$, dok na ostalim mestima ima vrednost 0 (pravougaoni prozor od [... 0 0 0 0 1/5 1/5 1/5 1/5 1/5 0 0 0 ...] čija je površina jednaka 1). [301]

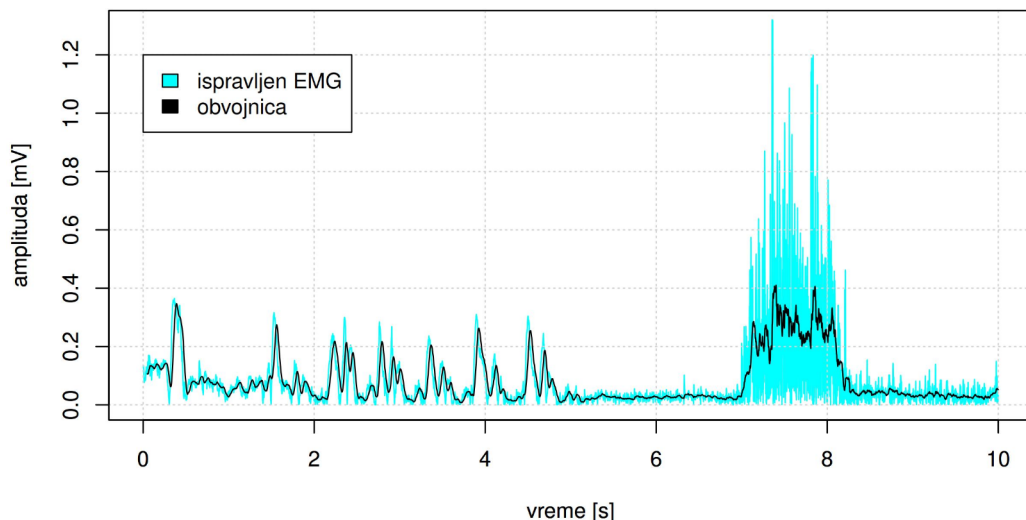
$$y[13] = \frac{x[11] + x[12] + x[13] + x[14] + x[15]}{5} \quad \text{Relacija (6)}$$

Funkcija prenosa MA filtra u Z domenu u oznaci $H[z]$ je data u Relaciji (7) gde su sa $Y[z]$ i sa $X[z]$ prikazane Z transformacije izlaznog $y[i]$ i ulaznog $x[i]$ signala iz filtra.

$$H[z] = \frac{Y[z]}{X[z]} = \frac{1}{M} \sum_{k=0}^{M-1} z^{-k} \quad \text{Relacija (7)}$$

Postoji više načina kako se može realizovati MA filter u programskom jeziku R. Na Sl. 89 je prikazana obvojnica EMG signala dobijena primenom MA filtra širine 50 odbiraka (50 ms za frekvenciju odabiranja od 1000 Hz) zajedno sa ispravljenim EMG signalom na kome postoji šum od pomeranja kablova. Kod za realizaciju MA filtra dat je na Sl. 90. Kako prikazani segment EMG signala na Sl. 89 nije filtriran visokopropusnikom to se vidi da je šum uticao na obvojnici EMG signala, te su se na obvojnici javile lažne aktivacije do oko 5 s

koje potiču od pomeranja kablova. Ovakvi šumovi bi se mogli izbeći korišćenjem medijana filtra (umesto srednje vrednosti, računa se medijana na prozorima – nelinearan filter).



Slika 89, Primena MA filtra za računanje obvojnice na nefiltriranom EMG signalu. Širina MA filtra je 50 odbiraka, odnosno 50 ms, a kod za računanje obvojnice je dat na Sl. 90.

```
datEMGkabl <- abs(datEMGkabl) # ispravljanje signala  
ma50 <- rep(1/50, 50)  
datEMGkablMA <- stats::filter(datEMGkabl, ma50, sides = 1)
```

Slika 90, R kod za računanje EMG obvojnice primenom MA filtra širine 50 ms.

Da bi se primenio MA filter najpre je kreiran niz brojeva od 50 elemenata gde su svi elementi iste vrednosti 1/50 (Sl. 89). Ovaj niz je iskorišćen zajedno sa *filter()* funkcijom iz stats paketa – jednog od osnovnih R paketa koja realizuje konvoluciju pravougaone funkcije sa EMG signalom. Pomoću *sides* argumenta može se definisati simetrična i asimetrična realizacija MA filtra. Na Sl. 90 prikazana je simetrična realizacija (*sides* = 1), a za asimetričnu realizaciju potrebno je postaviti argument *sides* na 2. Detalji ove realizacije su prikazani u uputstvu za studente TOBS predmeta²⁴⁴, ali i u knjigama o programiranju u R-u [302]. Kada bi se prikazali MA filteri za asimetričnu i simetričnu realizaciju primetilo bi se kašnjenje između ovih obvojnica na grafiku, što očekivano kada

244 https://automatika.etf.bg.ac.rs/images/FAJLOVI_srpski/predmeti/izborni_kursevi_os/biomedicinsko_inzenjerstvo/TOBS/Implementacija%20MA%20u%20R-u.pdf, pristupljeno 02.02.2024.

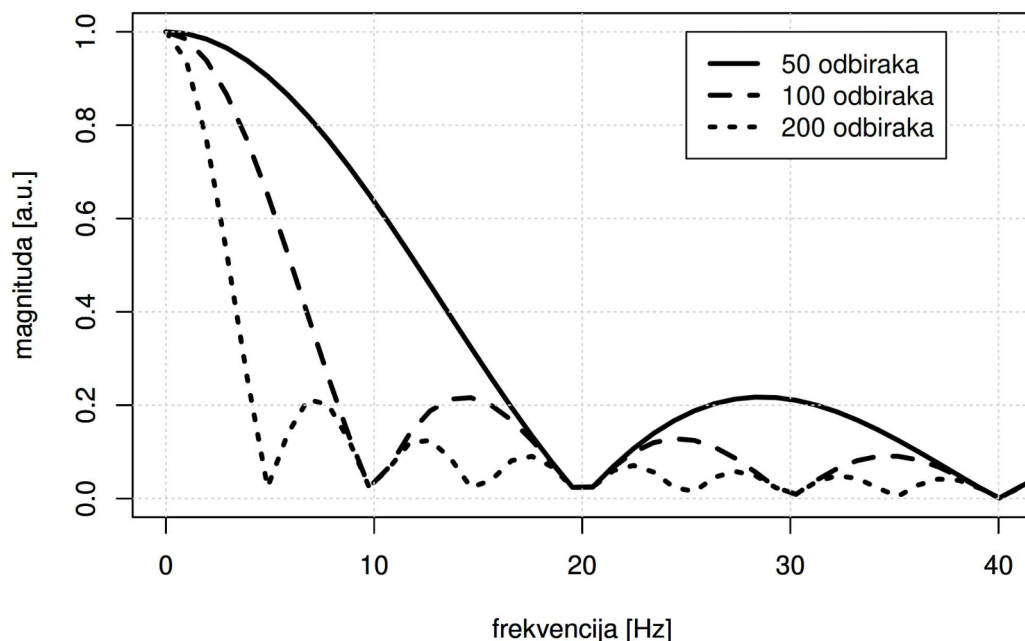
se pogleda razlika u granicama u kojima se kreće suma za $H[z]$ datim u Relacijama (5) i (6).

MA filtar se u R-u može realizovati i primenom Z transformacije MA filtra (Relacija (7)), te je potrebno odrediti a i b koeficijente (po ugledu na Relaciju (2)), a potom primeniti `filter()` funkciju iz signal paketa (naravno poželjnije je primeniti `filtfilt()` funkciju, ali je na Sl. 91 prikazana primena `filter()` funkcije radi poređenja). Rezultujuće EMG obvojnice, bez obzira na metodu kojom su dobijene su identične kao na Sl. 89.

```
a50 <- c(1, rep(0, 50 - 1))  
b50 <- rep(1/50, 50)
```

```
datEMGkablMA2 <- signal::filter(b50, a50, datEMGkabl)
```

Slika 91, R kod za računanje obvojnice primenom transformacije u Z domenu $H[z]$.

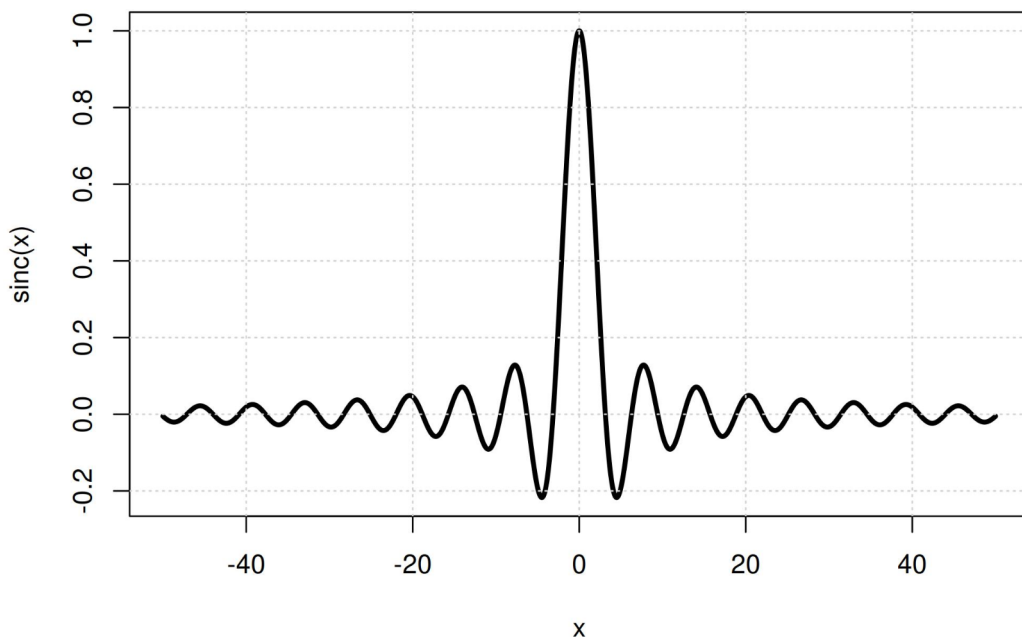


Slika 92, Frekventijska karakteristika tri MA filtra (širina prozora je 50, 100 i 200 odbiraka, odnosno 50 ms, 100 ms i 200 ms za uobičajenu frekvenciju odabiranja EMG signala od 1 kHz).

Pored dva prikazana načina realizacije MA filtra (Sl. 90-91), u signal paketu postoje specijalizovane funkcije za realizaciju MA filtra. Generalno, sa osnovnim MA filtrom dva su problema za primenu na EMG signalima: (1)

frekvencijska karakteristika koja loše razdvaja opsege u frekvencijskom domenu i (2) odabir širine prozora MA filtra tako da se dobije odgovarajuća obvojnica, ali da se ne filtrira koristan deo signala (za prozor širine 10 s, odnosno 10 000 odbiraka, za frekvenciju odabiranja od 1 kHz dobila bi se konstantna vrednost). Osnovno ograničenje MA filtra je što se primenjuje pod pretpostavkom da je signal stacionaran u okviru prozora koji je odabran, te je za odgovarajuću primenu MA filtra potrebno dobro poznavanje karakteristika signala. [186, 286]

Na Sl. 92 su prikazane frekvencijske karakteristike tri MA filtra gde je uočljivo da MA filter ne može da razdvoji frekvencijske opsege na odgovarajući način, kao i da se uočava značajno postojanje Gibsovih oscilacija za sve prikazane širine prozora. Jedan od razloga je što konvolucija pravougaonim signalom u vremenskom domenu (kako je ranije pokazano) rezultuje množenjem sa sinc funkcijom u frekvencijskom domenu i karakteristikom koja nije ravna u propusnom i nepropusnom frekvencijskom opsegu. [301]



Slika 93, Prikaz sinc funkcije u R-u za opseg ulaznog signala od -50 do 50.

Pre nego što bude pokazano kako se još mogu realizovati filtri za glačanje, treba objasniti pojam sinc funkcije koja je predstavljena u Relaciji (8).

$$\text{sinc}(x) = \begin{cases} \frac{\sin(x)}{x}, & x \neq 0 \\ 1, & x = 0 \end{cases} \quad \text{Relacija (8)}$$

Na Sl. 93 je prikazana sinc funkcija, a skripta u R-u za njenu realizaciju je jednostavna²⁴⁵:

```
sinc <- sin(x)/x; sinc[sinc == Inf] <- 1;
```

Frekvencijska karakteristika pravougaone funkcije u oznaci $H[f]$ kakva se koristi za primenu MA filtra sadrži sinc funkciju kao što je prikazano u Relaciji (9). Sa M je označen broj tačaka u kome se računa MA filter (širina prozora u vremenskom domenu).

$$H[f] = \frac{\sin(\pi f M)}{M \sin(\pi f)} \quad \text{Relacija (9)}$$

Obratiti pažnju da sinc funkcija ne može imati vrednost Inf (iako je u R-u ova operacija dozvoljena), kao i da većina udžbeničke literature prilikom navođenja sinc funkcije ne definiše šta se dešava prilikom deljenja sa nulom. Isto važi i za funkciju prenosa MA filtra koja za $f = 0$ ima vrednost 1 što se i vidi na Sl. 92 [301]. Signal filtriran MA filtrom ima Furijeovu transformaciju koja predstavlja konvoluciju Furijeove transformacije originalnog/nefiltriranog signala i funkcije koja sadrži sinc funkciju (množenje u vremenskom domenu odgovara konvoluciji u frekvencijskom domenu). Posledica toga je efekat bočnog curenja (eng. *side lobe leakage effect*) visokofrekventnih komponenti u rezultujućem signalu. Ovaj efekat se uočava na Sl. 92. Očekuje se da će ove visokofrekvencijske komponente (kao, na primer, komponenta na 15 Hz na Sl. 92) biti filtrirane MA filtrom koji je po prirodi niskopropusnik, ali u praksi to se ne dešava što ukazuje na ograničenje MA filtra. [303]

Realizacija pravougaonog MA filtra odgovara definiciji konvolucije. Međutim, nije pravougaoni MA filter jedini filter koji se koristi u praksi. Postoji više implementacija MA filtra – neke od njih mogu biti i rekurzivne. Pod rekurzivnim realizacijama misli se na filtre čiji tekući izlaz zavisi od prethodnih izlaza. U ovom udžbeniku prikazane su isključivo nerekurzivne realizacije MA filtra u Relacijama (4)-(9), ali postoje i rekurzivne realizacije. Sa ciljem razumevanja kako je rekurzivna forma povezana sa nerekurzivnom ovde je data

²⁴⁵ Iako u programskom jeziku R nije potrebno koristiti operator “;” kao u Matlabu, GNU Octave-u ili C-u, ovde je korišćen da bi se skratio kod odnosno da bi se dve komande izvršile u jednoj liniji koda.

ilustracija. Ako bi se uzela Relacija (5) kojom je određen 13. odbirak rezultujućeg filtriranog signala i odredio 14. odbirak, onda bi mogla da se piše rekurzivna forma 14. odbirka kao u Relaciji (10). Jednostavno se dolazi do opšteg slučaja. Važna odlika rekurzivne implementacije je u brzini primene filtra, što je posebno važno ako se radi trenutna obrada signala (eng. *online*).

$$y[13] = \frac{x[13] + x[12] + x[11] + x[10] + x[9]}{5} \quad \text{Relacija (10)}$$

$$y[14] = \frac{x[14] + x[13] + x[12] + x[11] + x[10]}{5}$$

$$y[14] = \frac{y[13] - x[9] + x[14]}{5}$$

Jedan od načina da se “popravi” frekvencijska karakteristika MA filtra (Sl. 91) jeste da se koriste srodni filtri, odnosno umesto pravougaonog koriste se prozorske funkcije, kao što su Hanova, Kajzerova i Hamingova (ili Hemingova). Druge opcije podrazumevaju korišćenje više prolazaka MA filtra kroz signal (tako se postiže efekat trougaonog ili Gausovog prozora u zavisnosti od broja prolazaka MA filtra kroz signal) ili da se koriste adaptivni prozori MA filtra. Značaj prozorske funkcije i MA filtra je i u tome što digitalni signali imaju konačnu dužinu, ali se teorijski mogu predstaviti kao rezultat množenja prozorske funkcije i signala beskonačne dužine u vremenskom domenu. [276, 301, 303-305]

Dobra osobina MA filtra je što poseduje linearnu fazu je sadržana i kod Hanovog filtra u vremenskom domenu. Dodatno, spektar signala koji se množi sa Hanovim prozorom²⁴⁶ u vremenskom domenu $Y_{HANN}[k]$ se može predstaviti preko spektra izlaza iz MA filtra $X_{MA}[k]$ kao što je prikazano u Relaciji (11). Sumiranje spektara kako je prikazano u Relaciji (11) ima za posledicu smanjenje magnitude prvog bočnog luka na magnitudskoj karakteristici Hanovog filtra u odnosu na MA filter (Sl. 94). Posledica ove kombinacije spektara je i to što se može efikasnije realizovati hardverski. [276, 306]

²⁴⁶ Naziv prozora je od austrijskog meteorologa Džulijasa fon Hana (https://en.wikipedia.org/wiki/Hann_function, pristupljeno 08.03.2024). Nekad se funkcija/prozor/filtar naziva fon Hanov, Hanov, a nekad Haningov (što je pogrešno, ali ustaljeno). Treba ga razlikovati Hamingovog prozora koji je naziv dobio po američkom matematičaru Ričardu Hamingu (https://en.wikipedia.org/wiki/Richard_Hamming, pristupljeno 08.03.2024). Mada, Autorka priznaje da nazivi Hanova i Haningova prozorska funkcija (koje su iste) česte unose zabunu da se radi o dve različite funkcije.

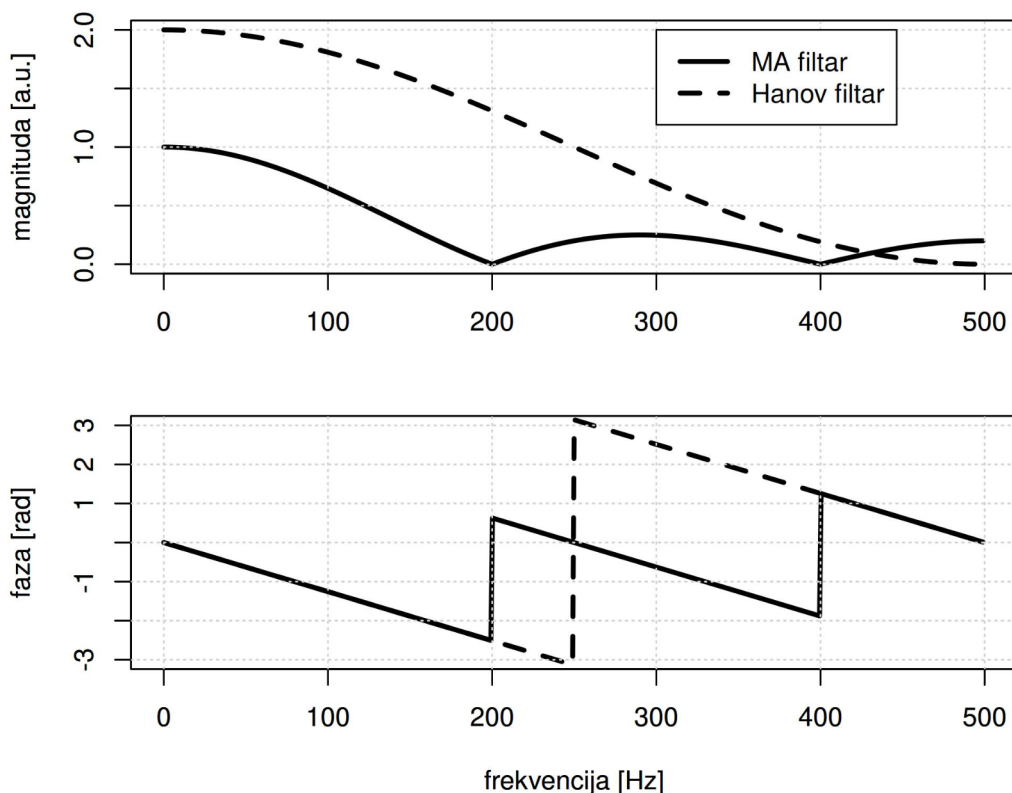
$$Y_{HANN}[k] = \frac{1}{2}(X_{MA}[k] - \frac{1}{2}X_{MA}[k-1] + X_{MA}[k+1]) \quad \text{Relacija (11)}$$

Za širinu prozora od 5 odbiraka, izlazni signal nakon filtriranja pokretnim Hanovim prozorom za odbirak 13 ima oblik kao u Relaciji (12) [186, 286].

$$y[13] = \frac{1}{4}(x[13] + 2x[12] + x[11]) \quad \text{Relacija (12)}$$

Z transformacija Hanovog prozora koji je dat u Relaciji (12) rezultuje funkcijom prenosa koja je data u Relaciji (13).

$$H[z] = \frac{1}{4}(1 + 2z^{-1} + z^{-2}) \quad \text{Relacija (13)}$$



Slika 94, Poređenje magnitudske i fazne karakteristike MA filtra i Hanovog filtra.

Ako se koristi *filter()* funkcija iz stats paketa za konvoluciju EMG signala i prozora u vremenskom domenu, onda se može jednostavno dobiti Hanov prozor primenom:

hanning()

funkcije prema Relacijama (12)-(13). Jednostavnom primenom *freqz()* funkcije mogu se uporediti frekvencijski odziv i faza dva filtra (MA sa pravougaonim prozorom i linearnog filtra sa Hanovim prozorom) sa širinom prozora od 5 odbiraka koje su prikazane na Sl. 94. Oblik frekvencijske karakteristike se razlikuje kod Hanovog prozora u odnosu na klasičan pravougaoni prozor kod MA filtra, odnosno karakteristika Hanovog filtra je povoljnija, jer bolje razdvaja frekvencijske opsege i smanjuje uticaj sinc funkcije.

Iako ovde nije prikazano korišćenje Kajzerove prozorske funkcije, smatra se jednom od najpoželjnijih za primenu u analizi signala, jer je analitički izraz za ovu funkciju fleksibilan, te se može podesiti dodatni parametar, pored dužine prozorske funkcije. Dodatnim parametrom se omogućava da se istovremeno utiče na širinu glavnog loba i na slabljenje prvog bočnog loba [307-309]. Kako se u većini primena u biomedicinskom inženjerstvu koristi Hanov prozor i kako je njegova analitička predstava jednostavnija, u ovom udžbeniku je prikazana primena Hanovog prozora. Za prikaz frekvencijske karakteristike na Sl. 94 korišćena je funkcija:

*abs()*²⁴⁷

za računanje modula²⁴⁸, odnosno za prikaz amplitudske frekvencijske karakteristike, dok je za prikaz fazne frekvencijske karakteristike (odnosno za argument kompleksnog broja) korišćena je:

atan2()

funkcija iz osnovnog R paketa. Funkcije *abs()* i *atan2()* su primenjene na frekvencijski odziv koji je dobijen primenom:

freqz()

²⁴⁷ U R-u se pored *abs()* funkcije može koristiti i *Mod()* funkcija, a moguće je izračunati i kvadratni koren od zbira kvadriranog imaginarnog i kvadriranog kompleksnog dela da bi se dobio moduo odnosno magnitudska frekvencijska karakteristika.

²⁴⁸ U upotrebi su termini moduo, modul i apsolutna vrednost u literaturi [310].

funkcije i koji rezultuje kompleksnim brojevima (slično kao i Furijeova transformacija signala). Za kompleksan broj $z=x+iy$ atan2 funkcija ima oblik kao u Relaciji (14). U R-u se za poziv funkcije za argumente koji su jednaki 0 dobija nula, jer je tako realizovana atan2 funkcija. U literaturi se pogrešno koristi arctan funkcija koja može da dovede do problema da se izgubi informacija o znaku uređenog para (x, y) kojim je određen kompleksan broj. [310]

$$\operatorname{atan}2(y, x) = \begin{cases} \arctan\left(\frac{y}{x}\right) & \text{za } x > 0 \\ \arctan\left(\frac{y}{x}\right) + \pi & \text{za } x < 0 \text{ i } y \geq 0 \\ \arctan\left(\frac{y}{x}\right) - \pi & \text{za } x < 0 \text{ i } y < 0 \\ +\frac{\pi}{2} & \text{za } x = 0 \text{ i } y > 0 \\ -\frac{\pi}{2} & \text{za } x = 0 \text{ i } y < 0 \\ \text{nedefinisan} & \text{za } x = 0 \text{ i } y = 0 \end{cases} \quad \text{Relacija (14)}$$

Kako bi se pravilno predstavile fazne karakteristike (bez diskontinuiteta) sa Sl. 94 potrebno je dodatno iskoristiti funkciju:

unwrap()

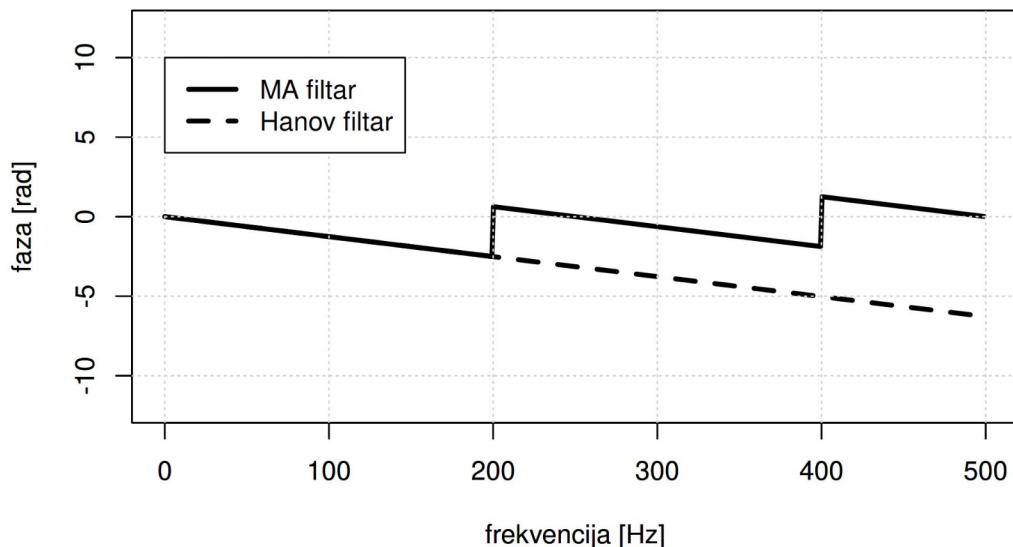
iz signal paketa. Za primenu ove funkcije dobijaju se faze MA i Hanovog filtra kao na Sl. 95. Uloga *unwrap()* funkcije je da smanji diskontinuitet koji nastaje kao posledica računanja faze: ako je diskontinuitet veći ili jednak π radijana, onda se vrši pomeranje faze za $\pm 2\pi$ sve dok diskontinuitet ne bude manji od π radijana²⁴⁹. Primećuje se da diskontinuiteti na Hanovoj faznoj frekvencijskoj karakteristici ne postoje nakon primene ove funkcije, dok MA filter ima deo po deo linearnu fazu [186, 286]. Postoje i drugi načini prikaza faze signala od kombinovanja atan2 i *unwrap()* funkcija. Na primer, iz diferencijalne geometrije kao odnos polarnih i Dekartovih koordinata, ali to prevazilazi gradivo pokriveno ovim udžbenikom²⁵⁰. Međutim, zanimljivo je spomenuti da je analitički oblik Hanovog prozora oblika *hann[n]* u diskretnoj tački n za

249 <https://octave.sourceforge.io/octave/function/unwrap.html>, pristupljeno 08.03.2024. i <https://www.mathworks.com/help/matlab/ref/unwrap.html>, pristupljeno 08.03.2024

250 Pogledati detalje u https://en.wikipedia.org/wiki/Winding_number, pristupljeno 08.03.2024.

dužinu prozora $N + 1$ je dat u Relaciji (15) (ovaj oblik se naziva Hanova funkcija²⁵¹).

$$hann[n] = 0.5 \left[1 - \cos\left(\frac{2\pi n}{N}\right) \right] = \sin^2\left(\frac{\pi n}{N}\right) \quad \text{Relacija (15)}$$



Slika 95, Poređenje fazne karakteristike MA filtra i Hanovog filtra sa primenom *unwrap()* funkcije.

Umesto primene *hanning()* funkcije u R-u je moguće koristiti Relaciju (16) da bi se realizovao Hanov prozor²⁵². Dalje, Relacija (15) se može generalizovati u formi $w[n]$ kao što je prikazano u Relaciji (16) [306], gde je za Hanov prozor parametar a_0 jednak 0.5, dok je za Hamingov ili Hemingov prozor taj parametar jednak 25/46. Forma data u Relaciji (16) može se dalje generalizovati da pokriva veći broj prozorskih funkcija.

$$w[n] = a_0 - (1 - a_0) \cos\left(\frac{2\pi n}{N}\right) \quad \text{Relacija (16)}$$

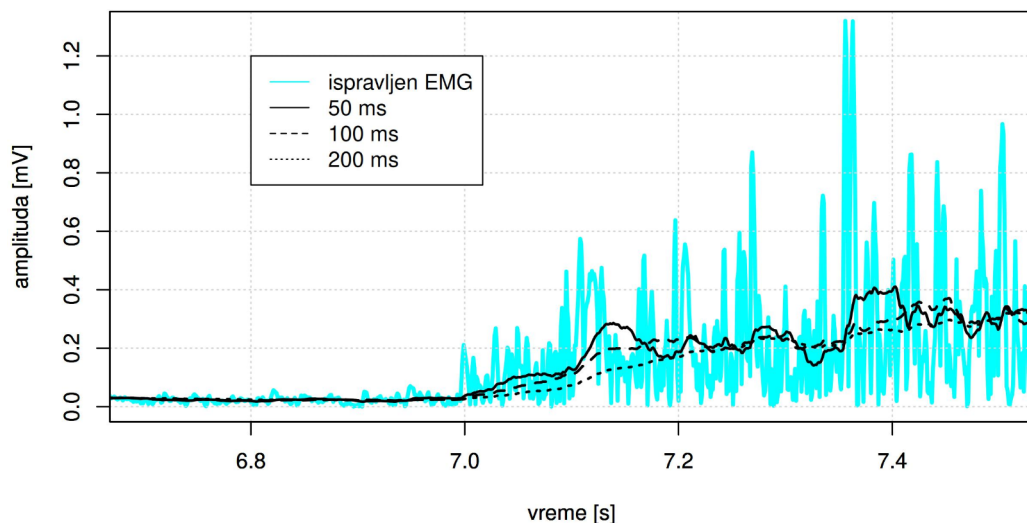
Za praktičnu primenu MA filtra ili srednje kvadratne greške tj. eng. *Root Mean Square*, skraćeno RMS ili medijane ili bilo koje druge metode glačanja koja koristi pokretne prozore, treba naći kompromisno rešenje za širinu prozora. Ako je prozor preuzak (najmanja širina prozora je $M = 1$) efekat filtra tj.

251 https://en.wikipedia.org/wiki/Window_function, pristupljeno 08.03.2024.

252 Pogledati kod koji je podeljen uz udžbenik za dva načina realizacije Hanovog prozora u R-u.

peglanja/glačanja se neće ni primetiti. Ako je prozor preširok, na primer, od par hiljada odbiraka što odgovara u vremenskom domenu intervalu od par sekundi (pod pretpostavkom da je frekvencija odabiranja $f_s = 1000$ Hz) onda će se izgubiti značajna informacija. U par sekundi može se desiti i par mišićnih kontrakcija, ako je ispitanik izvršavao relativno brzo pokrete. U praksi se koriste prozori od 100 ms do 200 ms za kreiranje obvojnice signala na kojima se, kao što je to već ranije ukazano, pretpostavlja da je signal stacionaran [105].

Za razliku od usrednjavanja signala u vremenu, mogu se usrednjavati i ansamblu merenog EMG signala. Kod usrednjavanja ansambla, može se usrednjavati samo vremenski ponovljiv oblik (ili EMG signal koji sadrži repetitivnu aktivaciju, koja se pri tom mora vremenski normalizovati tako da svaki segment bude iste dužine) kao što je to slučaj sa evociranim EMG potencijalima. Obe metode (usrednjavanje prozorama i usrednjavanje kroz ansamble) filtriraju visokofrekventni šum. Pored EMG signala, sinhrono usrednjavanje tj. usrednjavanje po ansamblima je naročito pogodno za filtriranje EKG signala kada je potrebno odrediti karakteristične tačke na EKG signalu koje su podložne šumu čiji se frekvencijski sadržaj poklapa sa frekvencijskim sadržajem EKG signala. Osim EKG signala, ovaj način analize se koristi za IKG signale (impedansni kardiogram, eng. *Impedance Cardiography*) [311].



Slika 96, Primer primene MA filtra sa tri različita širine prozora (50 ms, 100 ms i 200 ms) zajedno sa ispravljenim EMG signalom prikazanim u cijan boji.

Primeri obvojnice i ispravljenog EMG signala za tri MA filtra su prikazani na Sl. 96. Primetiti da je za veći prozor (200 ms) glačanje veće (obvojnica sadrži manje visokofrekventnih komponenti u odnosu na filtre sa manjom širinom

prozora), ali i maksimalna vrednost obvojnice je tada manja (posledica glačanja ili peglanja signala od eng. *smoothing*). Takođe, kada bi se određivao početak mišićne kontrakcije, mogli bi se dobiti različiti rezultati za račun primenjen na tri različite obvojnice EMG signala prikazane na Sl. 96. Kao što je već spomenuto, računanje početka i kraja mišićne kontrakcije su najčešće računata obeležja EMG signala koja su povezana sa obvojnicom EMG signala [105]. Primena metoda za glačanje ispravljenog EMG signala nije jedina transformacija koju je potrebno primeniti za računanje obvojnice. Prema SENIAM²⁵³ (eng. *Surface ElectroMyoGraphy for the Non-Invasive Assessment of Muscles*) protokolu poželjno je postaviti elektrode za merenje mišićne aktivnosti na odgovarajuće mesto [312], ali to nije dovoljno da bi se amplituda mogla koristiti za dalju analizu i odgovarajuću interpretaciju, te se amplituda EMG signala mora na odgovarajući način normalizovati [313].

2.2.3 Normalizacija EMG biosignala

Analiza EMG signala ima najveću primenu u medicini (za dijagnostiku, terapiju, rehabilitaciju i istraživanja u oblasti ocene rada senzorno-motornog sistema), ergonomiji i sportu [104-105, 263, 298, 314]. Međutim, čak ni jednostavno poređenje amplituda EMG signala koji su mereni na dva različita ispitanika se ne može realizovati bez prethodne predobrade (filtriranje i normalizacija), jer se impedansa kože, dimenzija mišića, fizičke i kondicione karakteristike ispitanika razlikuju i utiču na amplitudu EMG signala. Da stvar bude još komplikovanija, ako se EMG signal meri sa jednog mišića na istom ispitaniku, ponovljivost merenja može da zavisi od mesta postavke elektroda (ako se ne vodi računa o međunarodnim SENIAM i drugim preporukama za postavljanje elektroda [312-314]), te se željeni uticaj leka ili terapije ne može sagledati na odgovarajući način.

Dodatno, kritika upotrebljivosti EMG signala je usmerena i ka drugim aspektima kao što su nelinearna zavisnost sile koja se proizvodi prilikom pokreta i mišićne aktivnosti tj. amplitude EMG signala. Prema tome, EMG signal se ne može koristiti za procenu sile, osim ako se ne radi o relativno uskom i kontrolisanom opsegu sile koja se proizvodi na taj način. Dodatno, još uvek ne postoje jasne smernice za odgovarajući odabir mišića za proučavanje nekog pokreta. Uglavnom se koriste znanja iz anatomije da bi se odabrali mišići sa kojih će se meriti EMG signali, ali pokreti i radnje mogu biti veoma kompleksni²⁵⁴ da bi se opisali samo merenjem EMG signala sa par mišića.

253 <http://www.seniam.org/>, pristupljeno 25.02.2024.

254 Ovo se posebno odnosi na aktivacije sinergistične aktivacije mišića. Na primer, ako bi se merile aktivacije mišića tokom stajanja na jednoj nozi, logično je da bi mišići noge koja je

Pored toga što je EMG tehnika za merenje mišićne aktivnosti jedna od najzastupljenijih u biomedicinskim laboratorijama širom sveta, jer su merni uređaji relativno jednostavni i jer je ovu metodu jednostavno primeniti, treba napomenuti da površinski EMG nekada ne opisuje verno mišićnu aktivnost. Razlog tome može da bude relativno pomeranje mišića²⁵⁵ u odnosu na merno mesto na koži. Zato je opravdano preispitivati proceduru merenja da bi se utvrdilo da li su mereni signali pouzdani. Da situacija bude gora, ne postoji jedinstvena procedura normalizacije EMG signala, jer kako je pokazano, amplituda EMG signala zavisi od niza faktora. U okviru ovog poglavlja biće objašnjeni osnovni elementi pristupa normalizaciji EMG signala, koja je kompleksnija za razumevanje, ali vrlo jednostavna za primenu. EMG normalizacija se svodi na određivanje koeficijenta kojim će se normirati/pomnožiti obvojnica EMG signala (istim koeficijentom se deli svaki odbirak EMG obvojnice) i koja će omogućiti da se obvojnica predstavi normirano u opsegu od 0 do 1 ili do 100%²⁵⁶. Koncept normalizacije je od ključne važnosti za pravilnu analizu i interpretaciju EMG signala. [298]

Ukratko, normalizacija EMG signala predstavlja metodu kalibracije EMG signala. Pored toga, što je EMG signal potrebno normalizovati, ako je zadatak da se pored amplitude, u velikom broju situacija je potrebno izmeriti i referentnu EMG aktivnost tokom mirovanja (eng. *baseline*). S obzirom da normalizacija EMG signala uvodi skaliranje signala, osim uticaja na amplitudu obvojnice EMG signala, nema uticaja na oblik obvojnice EMG signala.

Najzastupljeniji način normalizacije je normalizacija EMG signala u odnosu na maksimalnu voljnu aktivaciju mišića (MVK/MVC, eng. *Maximal Voluntary Contraction*). MVK se meri u odvojenoj sesiji²⁵⁷ i poželjno je da se mišićna kontrakcija koja ima za cilj da postane MVK ponovi bar tri ili četiri puta. Razlog tome je da istraživač može da uporedi da li dolazi do smanjenja

oslonjena na pod bili pod većim opterećenjem. Međutim, kako bi vreme prolazilo postepeno bi se ti mišići zamarali i drugi mišići bi preuzeli funkciju održavanja zauzetog položaja tela (pa čak i mišići noge koja je u vazduhu odnosno koja nema oslonac). Ovo je primer sinergistične reakcije mišića u telu.

255 Iako je relativno jednostavno dostupan, mišić nadlaktice *biceps brachii* je jedan od mišića koji se pomera tokom pokreta u odnosu na merno mesto. Biceps na nadlaktici nije jedini mišić koji se pomera relativno u odnosu na mesto merenja, jer je pokazano da svi skeletni mišići imaju tu osobinu. [298]

256 U zavisnosti od načina normiranja, može se desiti da opseg normalizovane obvojnice bude veći od 100%.

257 Poželjno je da se merenje realizuje u istom danu sa istom postavkom elektroda, ali sa vremenskim razmakom između merenja MVK i kontrakcije od interesa. Ovo je potrebno kako potencijalni nastanak zamora ne bi uticao na kontrakciju od interesa.

amplitude EMG signala dok sila ostaje ista ili se povećava. To znači da je došlo do zamora i da jeste MVK izvršena. Ova činjenica se nikako ne može proveriti u naknadnoj (eng. *offline*) analizi, osim u slučaju da je snimljena i sila koja je rezultat mišićne aktivacije. Nakon toga se koristi MVK kontrakcija da se izračuna koeficijent – najčešće maksimum obvojnice EMG signala u mV kojim će se podeliti svaki odbirak obvojnice EMG signala koji se normalizuje. U tom slučaju, normalizovana mišićna kontrakcija u formi obvojnice se predstavlja u procentima u odnosu na MVK.

Međutim, tokom MVK može postojati problem motivacije ispitanika da izvrši maksimalnu voljnu kontrakciju [315], kao i problem da ispitanik uopšte izvrši MVK je to može biti opasno po zdravlje ispitanika/pacijenta ili se mogu javiti neugodnost i bol [316]. U studiji u kojoj je meren EMG ispitanika sa bolom u leđima primenjena je normalizacija na maksimalnu referentnu vrednost, a ne na MVK [317]. Postoji niz načina kako je moguće izmeriti referentnu vrednost, a primeri metoda su [105]²⁵⁸:

- ponovljiv pokret ili postura koji za rezultat daju ponovljiv EMG signal (na primer, da svaki ispitanik u stojećem položaju sa ispravljenim leđima pruži ruke ispred sebe paralelno sa podlogom [274]),
- iako nije pogodna zbog varijabilnosti EMG amplitude iz dana u dan i iz sesije u sesiju, nekada se koristi merenje prihvatljivog maksimalnog truda (eng. *Acceptable Maximal Effort*, skraćeno AME) [317] i
- moguće je uniformisati opterećenje ako, na primer, svi ispitanici imaju zadatak da u zadatom položaju u ruci drže teg iste težine.

Istraživač koji meri EMG signal i istovremeno pruža otpor ka izvršenju pokreta koji za rezultat daje maksimalnu mišićnu kontrakciju, može da utvrdi da li je motivacija ispitanika za izvršenje pokreta dovoljno velika, tako što će primetiti da se sila koju proizvodi osoba povećala ili ostala na istom nivou, dok je amplituda EMG signala počela da se smanjuje. Ovakva situacija znači da je osoba dostigla MVK, ali da je zbog trajanja kontrakcije došlo do zamora, odnosno do situacije kada se motorne jedinice postepeno isključuju i kada se amplituda EMG signala smanjuje. Zamor je, pored smanjenja amplitude EMG signala u vremenskom domenu, praćen pomerajima u frekvencijskom domenu, odnosno smanjenjem medijana frekvencije i pomeranjem frekvencijskog sadržaja u niži opseg frekvencija. Razlog tome je jednostavan – prilikom

²⁵⁸ U usmenoj komunikaciji sa naučnicima u oblasti, zaključeno je da ne postoji jasan način normalizacije signala, kao i da oni sa najviše iskustva i znanja vrlo često koriste više metoda ili testiraju željenu metodu normalizacije na referentnom merenju na zdravim ispitanicima. Slična tvrdnja je potvrđena i u radu Besoma i saradnika [316].

povećanja sile koju mišić proizvodi uključuje se veći broj motornih jedinica čiji se akcioni potencijali sumiraju i tako nastaje EMG signal. Nemaju sve motorne jedinice isti udeo u ukupan frekvencijski sadržaj EMG signala, te se poslednje uključuju one čija je funkcija povezana sa višim frekvencijama na EMG signalu. Kada dođe do zamora, prvo se isključuju motorne jedinice koje daju doprinos višim frekvencijama, pa nižim. Zbog toga se usled povećanja intenziteta kontrakcije i usled zamora javljaju promene u frekvencijskom domenu na EMG signalu. Problem kod detekcije zamora je što pomeranje EMG signala u frekvencijskom domenu može biti izazvano i zamorom i voljnim smanjenjem intenziteta mišićne kontrakcije, te se za detekciju zamora koriste dodatni senzori. To su najčešće senzori sile koji omogućavaju da se proceni da nije došlo do smanjenja sile koju mišić proizvodi na direktan ili indirektan način. [104-105, 318]

Mnogi naučnici i istraživači se zalažu da se u oblasti normalizacije EMG signala izvrši odgovarajuće uparivanje MVK procedure i procedure za merenje EMG signala koji je potrebno analizirati [319-320]. Na primer, ako je cilj izmeriti EMG signal tokom izometrične kontrakcije, onda je potrebno i normalizaciju realizovati tokom izometrične kontrakcije i obrnuto. Izometrična kontrakcija²⁵⁹ nastaje ako mišić proizvodi akcione potencijale u statičkom položaju, odnosno ne dolazi do skraćanja mišića ni do pokreta [321]. U studiji ocene mišićnih aktivacija iz EMG signala tokom hoda u slučaju perturbacija, korišćena je normalizacija u odnosu na EMG signale tokom hoda (anizometrična aktivacija mišića) bez perturbacija koji su mereni u istoj sesiji (postavka elektroda je bila ista tokom oba merenja) [322], što je u skladu sa preporukama za normalizaciju EMG signala merenih u dinamičkim uslovima [298, 316]. Međutim, nije uvek moguće izvršiti ovakvo uparivanje, posebno ako se meri eksplozivna snaga kod sportista tokom skoka [323]. Takođe, nisu svi istraživači u saglasnosti sa ovim preporukama, te neki isključivo preporučuju izometričnu kontrakciju za normalizaciju EMG signala [318]. Detaljan pregled metoda na normalizaciju EMG signala dali su Suoza i Travers [324], kao i Besomi i saradnici [316].

Pored normalizacije amplitude, postoji i vremenska normalizacija EMG signala. Vremenska normalizacija je korisna za analizu ljudskog hoda, odnosno kada je potrebno segmentirati EMG signal na faze hoda: na fazu zamaha ili eng. *swing* i na fazu oslonca ili na eng. *stance*. Da bi se izvršila vremenska normalizacija, na primer, sa ciljem sinhronog usrednjavanja EMG obvojnica za pojedinačne

259 Primer izometrične kontrakcije je statična vežba zvana daska (eng. *plank* ili *frontal plank*) u kojoj dolazi do aktivacije mišića, ali nema pokreta. Više na [https://en.wikipedia.org/wiki/Plank_\(exercise\)](https://en.wikipedia.org/wiki/Plank_(exercise)), pristupljeno 10.03.2024.

korake tokom hoda, najpre je neophodno da se izvrši detekcija početka i kraja mišićne kontrakcije na EMG signalu. Segmentacija na korake, pored detekcije početka i kraja mišićne kontrakcije može se izvršiti i primenom dodatnog senzora pritiska postavljenog na peti, te se u tom slučaju vrši segmentacija na fazu zamaha kada se noga nalazi u vazduhu i fazu oslonca kada je peta u kontaktu sa podlogom. Najčešće se koristi inicijalni kontakt pete sa podlogom za računanje početka i kraja koraka. Obzirom da se izdvojeni segmenti EMG signala razlikuju po dužini (sadrže nejednak broj odbiraka), najpre se vrši njihova podela na jednak broj intervala. Svaki interval predstavlja određen procenat od ukupne dužine izdvojenog segmenta, te se usrednjavanje vrši po ovim segmentima, odnosno računa se srednja vrednost iz svih segmenata. [104-105, 325-326].

Cilj ovog potpoglavlja nije da studenti nauče da mere normalizovane EMG signale za šta je potrebna odgovarajuća obuka, već je cilj da razumeju koja ograničenja postoje u izmerenim EMG signalima kako bi ih na odgovarajući način analizirali. Normalizacija je neophodan korak da se smanji uticaj fizioloških i anatomskih karakteristika mišića i okolnog tkiva [316]. Na kraju, ne treba zaboraviti da, iako se EMG signal jednostavno meri (elektrofiziološki pojačavački sistem se može relativno jednostavno realizovati), važi da je “elektromiografija jednostavna za upotrebu, ali i za zloupotrebu” (slobodan prevod na srpski jezik iz [104]²⁶⁰).

2.3 Izdvajanje karakterističnih obeležja

Kada je šum uklonjen i kada su podaci transformisani na odgovarajući način, sledeći korak u obradi signala uključuje izdvajanje obeležja (eng. *feature extraction*). Proces izdvajanja obeležja se retko kada obavlja na sirovom, tj. nefiltriranom signalu, a može da uključi prethodnu transformaciju podataka. Na primer, izdvajanje obeležja kao što su početak i kraj mišićne kontrakcije neće biti izvršeno na filtriranom EMG signalu, već na transformisanom EMG signalu tj. obvojnici EMG signala.

Kao pretkorak u izdvajanju obeležja može se primeniti tehnika dekompozicije signala. Na Sl. 97 dat je primer procedure za transformaciju signala primenom

260 Citat u originalu glasi “Electromyography is too easy to use and consequently too easy to abuse.” [318]. Takođe, slični citati se mogu naći i u radu pod naslovom “Electromyography: its use and misuse in physical education” (slobodan prevod na srpski: Elektromiografija: upotreba i zloupotreba u fizičkom obrazovanju) [327].

empirijske dekompozicije (eng. *Empirical Mode Decomposition*, skraćeno EMD) [328]. Da bi se izvršila EMD, potrebni su sledeći koraci u obradi signala:

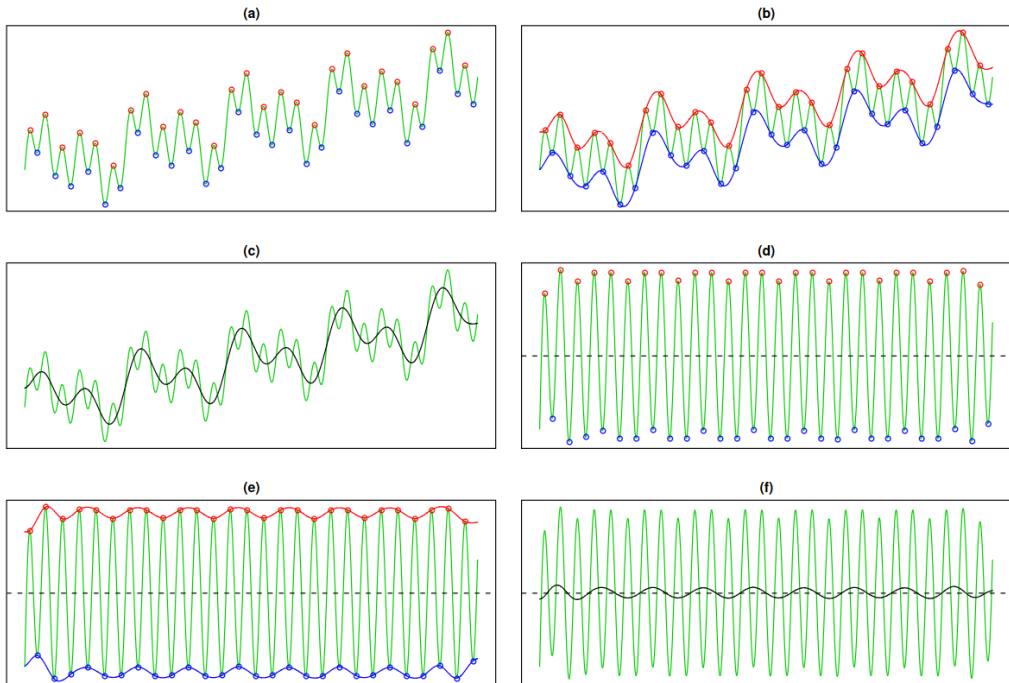
- Vršiti se detekcija lokalnih ekstremuma, odnosno minimuma i maksimuma (panel a) na Sl. 97).
- Na osnovu lokalnih minimuma i maksimuma kreiraju se dve obvojnice (panel b) na Sl. 97). Ovaj proces se može obaviti na više načina, ali je najčešće u upotrebi kubična²⁶¹ interpolacija.
- Srednja vrednost dobijenih obvojnica se oduzima od originalnog signala (panel c) na Sl. 97). Rezultujuća srednja vrednost signala ima niži frekvencijski sadržaj od preostalog dela signala, što je jedan od razloga zašto je EMD metoda slična filtriranju primenom banke filtara, čemu u prilog ide i činjenica da preostali deo signala prolazi kroz više iteracija. Najjednostavnija definicija banke filtara jeste da je to skup filtara koji su sastavljeni sa ciljem da se signal rastavi na frekvencijske podopsege²⁶² [329].
- EMD procedura se ponavlja (paneli d)-f) odgovaraju koracima prikazanim na panelima a)-c) na Sl. 97).

Drugi naziv za EMD tehniku je prosejavanje (eng. *sifting*), jer ponavljanje procesa prikazanog na Sl. 97 podseća na razdvajanje signala na komponente sa različitim frekvencijskim udelima, što je slično usitnjavanju, odnosno prosejavanju. Privlačnost ove metode se ogleda u tome što je pogodna za analizu nestacionarnih i nelinearnih vremenskih serija što odgovara prirodi većine biosignala. Međutim, EMD je kritikovana u literaturi, jer je uočen [328, 330-331]:

- nedostatak teorijskih osnova da se objasne osnovne funkcije (eng. *Intrinsic Mode Function*, skraćeno IMF),
- efekat EMD analize na signal (u slučaju prostoperiodičnih signala efekat se jednostavno objašnjava, ali u slučaju realnih stohastičkih signala to je znatno kompleksniji zadatak),
- koji kriterijum odabrati za zaustavljanje procedure prosejavanja i
- na koji način realizovati interpolaciju lokalnih ekstremuma.

261 Nemoguće je ne primetiti da u modeliranju i interpolaciji, broj tri predstavlja “magičan broj”. Verovatni razlog za njegovu čestu upotrebu u analizi i modeliranju signala je što omogućava veću fleksibilnost od korišćenja drugog reda, a manju računarsku kompleksnost od četvrtog. Autorki nije poznato da li je neko istraživao uzrok korišćenja trećeg reda.

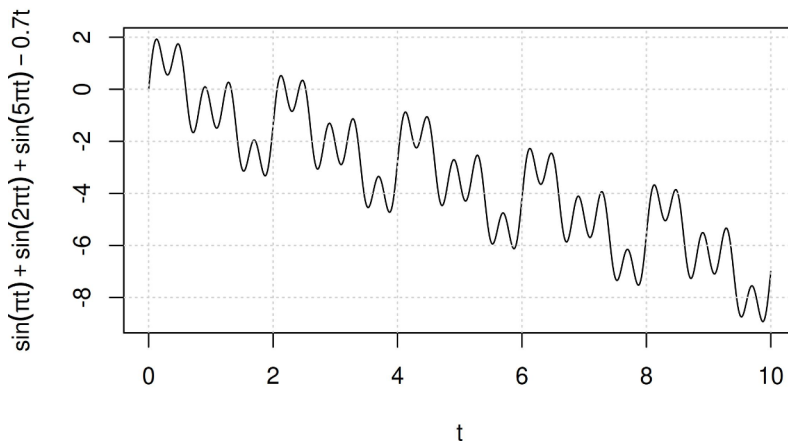
262 Zanimljivo je istaći da, pored toga što deluje apsurdno, banka filtara se može realizovati bez filtara. Na primer, korišćenjem operatora za smanjenje i povećanje frekvencije odabiranja (eng. *downsampling* i eng. *upsampling*). [329]



Slika 97, Procedura EDA tehnike: a) određivanje lokalnih ekstremuma (na slici su lokalni maksimumi označeni crvenom bojom, a lokalni minimumi plavom), b) određivanje obvojnica kroz lokalne maksimume (crvena linija) i kroz lokalne minimume (plava linija), c) rezultat usrednjavanja dve obvojnice sa panela b) rezultuje jednom obvojnicom prikazanom crnom bojom, d) proces određivanja lokalnih ekstremuma iz koji je prikazan na panelu a) se nastavlja na signalu kome je oduzeta usrednjena obvojnica, e) i f) paneli odgovaraju b) i c) panelima. Izvor: članak objavljen u R časopisu sa otvorenim pristupom [328].

Na Sl. 98 prikazan je primer kombinacije tri prostoperiodična signala sa linearnim trendom nad kojima je primenjena EMD tehnika. Rezultat EMD metode u kome su prikazane osnovne funkcije (IMF) i ostatak, odnosno rezidual je dat na Sl. 99. Jedna od mana EMD procedure se može jasno uvideti na krajevima osnovnih funkcija prve i druge IMF (Sl. 99) koji je posledica primene EMD procedure. Na ovaj način rezultujuće IMF imaju različitu energiju izdvojenih komponenti u odnosu na energiju originalnog signala. Za rešavanje ovog problema i drugih prepoznatih mana EMD procedure, predložene su modifikacije originalne EMD procedure koje su detaljno prikazane u preglednom radu Leja i saradnika [330].

Ako bi se ista EMD procedura, koja je primenjena na signal sa Sl. 98, primenila na EMG signal meren sa mišića *Pectoralis* (lat.) u kome postoji šum od EMG signala (pogledati Sl. 73), dobio bi se rezultat kao na Sl. 100.



Slika 98, Primer zbira tri prostoperiodična signala i linearnog trenda.

Odabran primer primene EMD tehnike na EMD signal je ranije već prikazan u [251]²⁶³ sa podeljenim R programom koji omogućava primenu EMD procedure korišćenjem EMD paketa koji je dostupan na CRAN-u²⁶⁴ [328]. Rezultat na Sl. 99 je očekivan i ukazuje na to da metoda prosejanja dobro razdvaja osnovne komponente (prostoperiodični signal i linearni trend), rezultat koji se dobija kada se EMD primeni na EMG signal sa EKG šumom (Sl. 100) nije tako jednostavan za tumačenje. Delimično, nije ni očekivan, jer se ne vide jasno IMF koje potiču od EMG signala, a koje od EKG signala. Međutim, ako se pogledaju prve dve IMF (gornji paneli na Sl. 100) može se primetiti da oni ne poseduju udeo od EKG signala, te da se njihovom kombinacijom može dobiti filtriran deo EMG signala na kome nema EKG šuma i na kome se mogu jednostavno odrediti početak i kraj mišićne kontrakcije, što nije zanemariva korist od EMD metode.

U Relaciji (17) je prikazana EMD procedura rasejanja, gde je sa $x(t)$ označen ulazni signal, sa $imf(t)$ osnovna funkcija, dok je sa $r(t)$ označen je rezidual, tj. ostatak EMD procedure.

$$x(t) = \sum_{i=1}^n imf_i(t) + r(t) \quad \text{Relacija (17)}$$

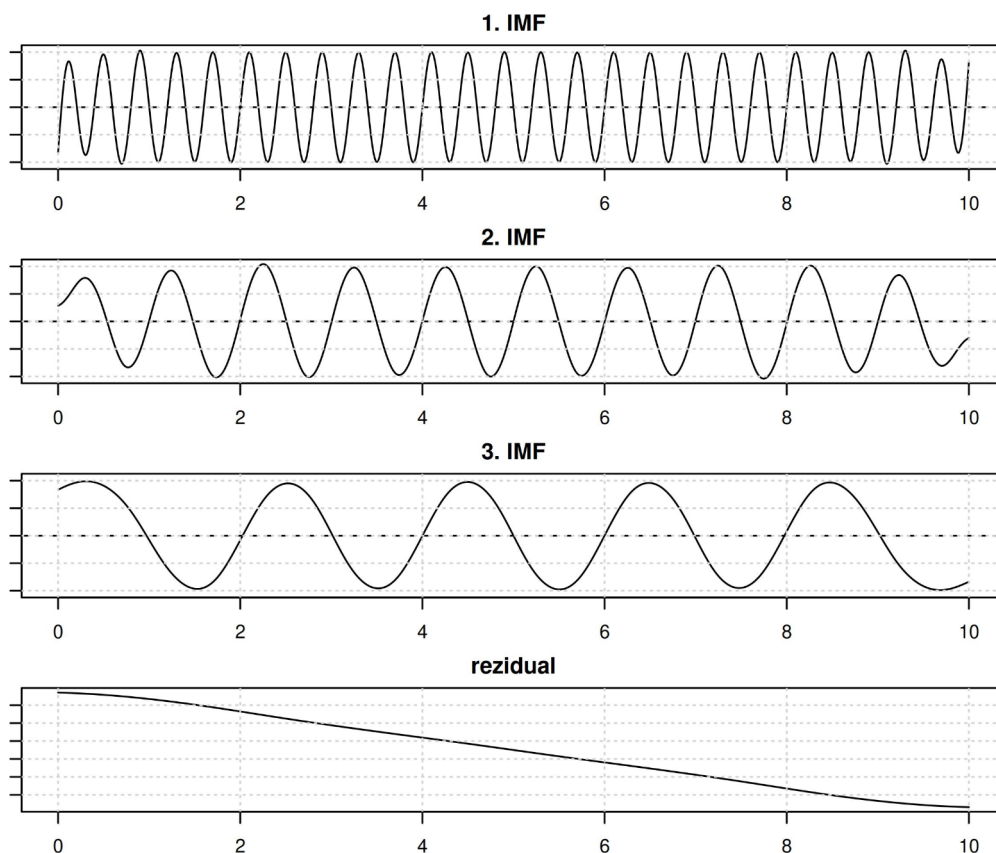
263 R kod za apstrakt i prezentaciju su dostupni na <https://github.com/NadicaSm/satRday-Belgrade-2018/>, pristupljeno 10.03.2024.

264 <https://cran.r-project.org/package=EMD>, pristupljeno 10.03.2024. NAPOMENA: U zavisnosti da li su Autori CRAN paketa dali koje reference treba citirati, način navođenja se može razlikovati. Kako u slučaju EMD paketa u okviru preporučenih referenci nema linka ka CRAN repozitorijumu, naknadno je dodan link ka CRAN repozitorijumu u fusnoti.

Na Sl. 99 i Sl. 100 rezultat EMD metode je dobijen korišćenjem funkcije:

emd()

u R-u. Izlaz *emd()* funkcije je predstavljen u formi liste koja sadrži tri elementa: (1) matricu koja sadrži pojedine odbirke osnovnih funkcija (IMF), (2) niz koji sadrži odbirke reziduala i (3) numerički podatak koji predstavlja broj IMF-a koji je izdvojen.

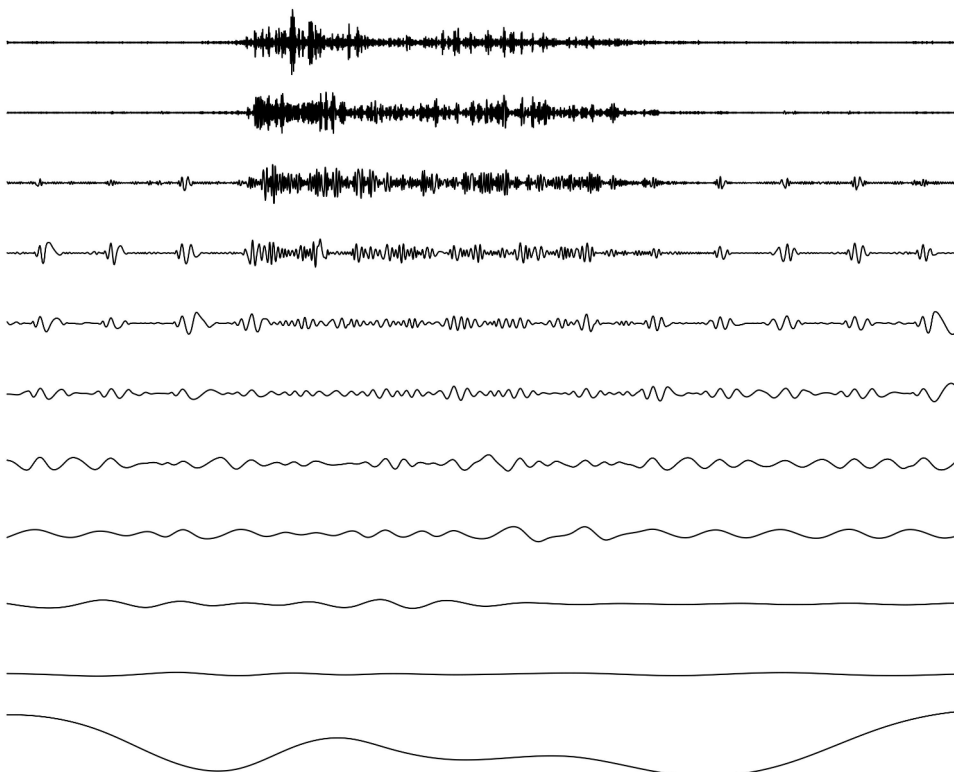


Slika 99, Rezultat primene EMD metode na signal $x(t)$ koji je prikazan na Sl. 98. Skraćenica IMF označava osnovnu funkciju od eng. *Intrinsic Mode Function*.

Argumenti *emd()* funkcije koji su korišćeni za primenu EMD nad sumom prostoperiodičnih funkcija i nad EMG signalom koji sadrži EKG šum su: (1) odbirci signala, (2) vremenska osa i (3) metodu za smanjenje ivičnih efekata koji se naročito vide na Sl. 99 za prvu i drugu IMF, dok su ostali argumenti postavljeni na podrazumevane vrednosti. Postoji više metoda koje se mogu

koristiti za smanjenje ivičnih efekata EMD metode (eng. *boundary*), te *emd()* funkcija nudi sledeće opcije [328]:

- “none” što je samo po sebi jasno da ne postoji nikakva dodatna procedura za smanjenje ivičnih efekata,
- “wave” metoda konstruiše talas definisan sa dve uzastopne ekstremne vrednosti i dodaje talase na oba kraja signala,
- “symmetric” produžava signal na oba kraja simetrično,
- “periodic” produžava originalan signal na oba kraja periodično i
- “evenodd” je procedura koju su predložili Zeng i He [332], a koja uključuje dodavanje odbiraka sa obe strane signala tako da dodati deo ima obvojnici čija je srednja vrednost jednaka nuli čime se ne menja ukupna energija signala.



Slika 100, Rezultat primene EMD tehnike na EMG signal sa *Pectoralis* (lat.) mišića koji je prikazan na Sl. 73. Na donjem panelu je prikazan rezidual, odnosno ostatak.

Za ilustraciju EMD metode na Sl. 99 i Sl. 100 korišćena je metoda zaustavljanja procedure rasejanja koja je označena sa “wave”. Ova metoda ne uklanja ivične efekte u potpunosti, ali kada se uporedi EMD procedura sa opcijom “none” i sa

opcijom “wave”, jasno se uočavaju veći ivični efekti na rezidualu, kao i na osnovnim funkcijama nižeg frekvencijskog sadržaja u slučaju da je odabrana opcija “none”.

Pored EMD tehnike, postoje i druge transformacije koje je moguće primeniti na biosignalima u zavisnosti od toga šta se ispituje. Jedna od zanimljivih metoda analize EMG biosignala podrazumeva izdvajanje sinergije iz višekanalnog merenja EMG-a. Nikolaj Bernštajn je 1967. godine predložio da se pokreti u kojima učestvuje veliki broj mišića mogu objasniti modularnom organizacijom nervno-mišićnog sistema. Bernštajnova hipoteza se zasniva na tome da postoji relativno mali broj mišićnih sinergija koje karakterišu veliki broj kompleksnih pokreta u prostoru. Mišićne sinergije predstavljaju kontrolne mehanizme koji omogućavaju aktivaciju većeg broja mišića. Postoji više načina da se identifikuju mišićne sinergije, ali sve metode uključuju sledeće korake [333]:

- filtriranje EMG signala sa ciljem eliminacije šuma,
- ispravljanje i filtriranje signala sa ciljem dobijanja obvojnice EMG signala,
- normalizaciju EMG amplitude (ovaj korak se u nekim situacijama može preskočiti),
- odabir broja sinergija koje će biti izdvojene primenom odgovarajuće analize i
- primenu faktorske analize za izdvajanje pojedinih sinergija.

Kako proračun mišićnih sinergija uključuje niz koraka koji mogu imati različite parametre, od kojih su primeri filtriranja signala sa ciljem eliminacije šuma i izdvajanje obvojnice prikazani u ovom udžbeniku, to se mora voditi računa o tome da pretprocesiranje može imati značajan uticaj na konačan rezultat, odnosno na broj i prirodu sinergija [333].

Faktorska analiza koja se koristi za izdvajanje sinergija uključuje primenu statističkih metoda sa ciljem da se na odgovarajući način kvantifikuje varijabilnost korelisanih promenljivih. Rezultat faktorske analize su transformisane promenljive, tzv. faktori, kojih ima manje od broja polaznih promenljivih (broj obvojnica koji odgovara različitom broju mišića na kojima su mereni EMG signali), a koji zadržavaju varijabilnost originalnog skupa. [334]

Primeri faktorskih metoda koje se koriste za izdvajanje mišićnih sinergija su [333]:

- nenegativna matrična faktorizacija (eng. *Nonnegative Matrix Factorization*),
- metoda glavnih komponenti (eng. *Principal Component Analysis*, skraćeno PCA) i
- metoda nezavisnih komponenti (eng. *Independent Component Analysis*, skraćeno ICA).

Izdvajanje broja sinergija se može obaviti na više načina i u velikoj meri zavisi od primenjene faktorske analize. U upotrebi su zaključivanje na osnovu varijanse, minimizacija varijanse na skupu podataka, kriterijumi koji se zasnivaju na računanju sopstvenih vrednosti i klasterizacija [333]. Postoje i drugi primeri primene metoda faktorske analize. ICA metoda se može koristiti za filtriranje EKG artifakta²⁶⁵ iz EEG signala kod novorođenih beba, jer postoji pretpostavka statističke nezavisnosti EKG i EEG signala [335, 337]. Pored ICA metode, PCA tehnika je verovatno najčešće korišćena multivarijantna metoda u biomedicinskom inženjerstvu. Pokazala se korisnom u smanjenju dimenzionalnosti obeležja dobijenih iz EMG signala (koji su mereni sa ukupno 10 kanala) sa ciljem klasifikacije različitih pokreta ruke sa samo dve izdvojene glavne komponente (eng. *Principal Component*, PC) [338]. Odabir dve glavne komponente je doprinio smanjenju skupa od 5 puta (inicijalno su bili dostupni EMG signali sa 10 kanala) što je dalo relativno zadovoljavajuće rezultate. Međutim, pokazano je da se smanjenju dimenzija mora postupiti sa više pažnje. Naime, ako se mesto dve iskoriste tri PC za klasifikaciju pokreta, dobijaju se statistički značajno bolji rezultati klasifikacije pokreta ruku [339]. Takođe, pokazano je da optimalan broj glavnih komponenti zavisi od broja pokreta koje je potrebno klasifikovati primenom PCA na merene EMG signale [340]. Među najpoznatijim istraživanjima primene PCA tehnike su radovi u kojima su predstavljeni dokazi postojanja pet mišićnih sinergija na osnovu višekanalne analize EMG signala tokom hoda koja je uključila čak preko 30 mišića koji su aktivni tokom hoda [341-342]. Najčešće je prvi korak u primeni faktorske analize, računanje kros-korelacionih koeficijenata i prikaz kros-korelacione matrice kako bi se videlo da li postoji potreba za faktorskom analizom, odnosno da li skup podataka sadrži redundantnu informaciju [341]. Metode prikaza kros-korelacionih matrica, kao i primera kada je poželjno koristiti ovakvu

²⁶⁵ Ovo je vrlo specifična situacija, jer se EKG artefakti koji potiču od pulsacija u kranijalnim krvnim sudovima ne javljaju kod odraslih osoba, zato što postoji prostorno filtriranje takvih signala. Kod novorođenih beba, lobanja još uvek nije formirana, te se krvni sudovi nalaze u blizini memnog mesta [335]. Električni potencijal srca stvara električno polje duž celog tela, te uticaj EKG artifakta na EEG signal postoji i kod odraslih ispitanika, ali se takav uticaj razmatra isključivo u specijalizovanim ispitivanjima [336], dok se zanemaruje usled prostornog filtriranja kod merenja EEG signala sa površine skalpa kod odraslih osoba.

vizuelizaciju su detaljno predstavljene u poglavlju koje se bavi eksplorativnom analizom podataka.

2.3.2 Izdvajanje karakterističnih obeležja iz EMG signala

Ako je normalizacija EMG signala sprovedena na odgovarajući način, onda se može proceniti amplituda EMG signala u odabranim vremenskim intervalima. Jedan način procene amplitude je računanje srednje vrednosti na segmentu obvojnice EMG signala. Segmentacija se može vršiti inicijalno prema protokolu koji se primenjuje, a onda se za određene segmente mogu računati srednje vrednosti EMG obvojnice kao obeležja za dalju analizu (na primer, za klasifikaciju pokreta ruku primenom višekanalnog EMG signala) [339-340].

Umesto srednje vrednosti obvojnice EMG signala ili usrednjene obvojnice EMG signala, može se računati integral EMG signala, tj. obvojnice na određenom vremenskom intervalu. Integracija signala proizvodi rezultat čije su jedinice Vs (ili mVs), jer se računa površina ispod signala na kome se vrši integracija. Najčešće se ovaj korak izvodi na ispravljenom signalu (ili na obvojnici EMG signala) i koristi se da se kvantifikuje intenzitet mišićne kontrakcije. Očekivano, intenzitet mišićne kontrakcije zavisi od primenjene tehnike normalizacije. [298]

Za računanje površine ispod obvojnice EMG signala moguće je koristiti funkciju:

cumtrapz() ili *trapz()*

funkciju iz *pracma* (eng. *Practical Numerical Math Functions*) paketa [343]. Kada se funkcija *cumtrapz()* primeni na 8 s, odnosno 8000 odbiraka EMG signala iz datoteke "EMGpectoralis.txt" [251] onda se dobija rezultat kao na Sl. 101. Obe funkcije koriste trapezoidnu integraciju, s tim što *cumtrapz()* računa kumulativni integral, te izlaz iz ove funkcije ima isti broj odbiraka kao EMG signal nad kojim se vrši integracija (Sl. 101). Logično, maksimalna vrednost koju daje *cumtrapz()* funkcija je jednaka vrednosti koju daje *trapz()* funkcija.

Integracija se može postići i hardverski, što je ranije bio običaj u uređajima za merenje EMG signala, a danas se ona najčešće vrši softverski nad digitalnom reprezentacijom merenog signala. Integraljena usrednjena vrednost predstavlja 63.7% od polovine PTP (eng. *Peak To Peak*) vrednosti [78]. Nije poželjno da se analiza oslanja na ekstremne vrednosti: izraz koji povezuje integraljenu vrednost dvostrano ispravljenog signala je u upotrebi za prostoperiodične

signale, ali ne i za stohastičke biosignale, jer PTP vrednost signala može da bude pod uticajem ekstremnih vrednosti. Za karakterizaciju amplitude EMG signala ili obvojnice u slučaju da je signal nastao kao rezultat voljne mišićne kontrakcije, de Luka je preporučio da se koristi RMS (eng. *Root Mean Square*) vrednost koja predstavlja 70.7% od polovine PTP vrednosti signala (odnosno amplitude prostoperiodičnog signala) [78, 318]. RMS metoda je pogodnija za EMG signal kao rezultat voljne kontrakcije, jer ima direktan fizički značaj – predstavlja energiju signala i može se koristiti u izrazu za SNR (pogledati Relaciju (18)), dok za integral i srednju vrednost signala ne postoji direktan fizički značaj.

```
> dat <- read.table("EMGpectoralis.txt")
> emg <- dat$V1[1:8000] # 8 s sa prvog kanala
> max(cumtrapz(abs(emg)))
[1] 2054.361
> trapz(abs(emg))
[1] 2054.361
> length(emg)
[1] 8000
> length(cumtrapz(abs(emg)))
[1] 8000
> |
```

Slika 101, R kod koji prikazuje rezultat *cumtrapz()* i *trapz()* funkcija za integraciju ispravljenog EMG signala, kao i relaciju između ove dve funkcije²⁶⁶.

U Relaciji (18) dat je izraz za računanje RMS vrednosti EMG signala emg_{RMS} za digitalni signal dužine n čiji su pojedinačni odbirci označeni sa emg_i .

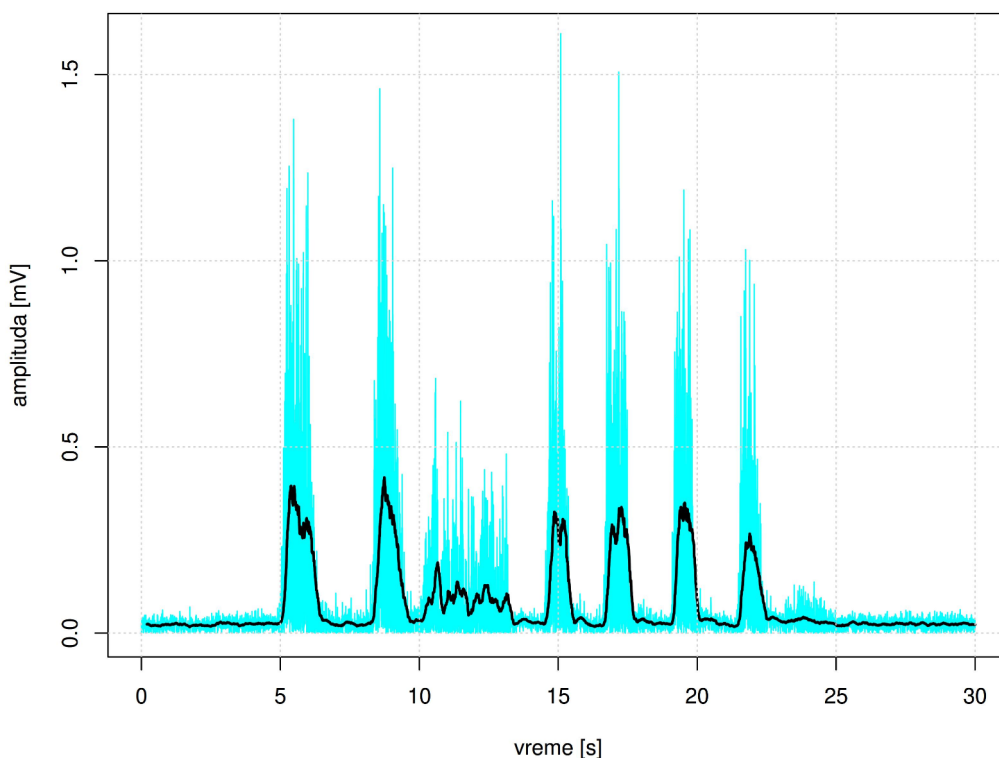
$$emg_{RMS} = \sqrt{\left(\frac{1}{n} \sum_{i=1}^n emg_i^2[n]\right)} \quad \text{Relacija (18)}$$

2.3.2.1 Određivanje početka i kraja mišićne kontrakcije

Jedan od primera analize signala, gde je važno proceniti amplitudu EMG signala, odnosno promenu amplitude EMG signala, jeste određivanje trenutka u

²⁶⁶ Ovde je radi ilustracije EMG signal dvostrano ispravljen, odnosno izračunata je njegova apsolutna vrednost (primenom *abs()* funkcije). Međutim, ovo je apsolutno pogrešan pristup, jer je EMG signal pre računanja apsolutne vrednosti potrebno filtrirati. Da bi se studenti uverili u ovu činjenicu dovoljno je da prikažu ispravljen EMG signal primenom sledeće linije koda: *plot(abs(emg), type = "l")*.

kome je nastupila mišićna kontrakcija i trenutka kada se završila. Usled različitih faktora koji utiču na amplitudu signala, to nije jednostavno uraditi, a nekada se istraživači opredeljuju da odaberu prag (eng. *threshold*) kao udeo u maksimalnoj voljnoj kontrakciji. U praksi, većina istraživača izbegava da prag određuje na osnovnu ekstremnih vrednosti ili vrednosti koje mogu biti pod velikim uticajem šuma, te se najčešće prag određuje kao umnožak standardne devijacije EMG signala. [105]



Slika 102, Prikaz ispravljenog EMG signala i obvojnice EMG signala iz datoteke “emg-sve.txt” za interval od 50 s do 80 s. Obvojnica je računata za MA filter širine 200 ms. Kako je signal snimljen sa minimalnim uticajem šuma, to nije bilo potrebno dodatno filtrirati signal pre određivanja obvojnice. U praksi, za serijsko procesiranje, uvek je preporučljivo filtrirati signale na prethodno ustaljen način, jer se u suprotnom proces ne može automatizovati, odnosno mora se koristiti vizuelna inspekcija signala (šum i priroda signala mogu se razlikovati od ispitanika do ispitanika i od sesije do sesije).

Detekcija početka mišićne kontrakcije primenom praga nije tako jednostavan zadatak kako se na prvi pogled može učiniti. Na primer, pitanje je kako najefikasnije definisati prag. Najčešće se signal pored, sa dva do tri puta većom vrednošću od standardne devijacije bazne linije, što bi trebalo da je vrednost tek nešto veća od bazne linije, odnosno od elektrofiziološke aktivnosti kada mišić

nije aktivan [316]. Dodatno, poželjno je definisati najmanji mogući interval, tj. trajanje mišićne kontrakcije, što takođe ne mora da bude jednostavan zadatak. Na Sl. 102 je dat primer EMG signala sa kontrakcijama različitog intenziteta i različitog trajanja koje mogu predstavljati izazov u definisanju vremenskog intervala i praga za određivanje početka i kraja mišićne kontrakcije.

Iz pracma paketa može se koristiti funkcija:

findpeaks()

za pretragu lokalnih ekstremuma za detekciju početka i kraja mišićne kontrakcije. Iako ova funkcija po svojoj definiciji ne određuje početak i kraj mišićne kontrakcije, uz odgovarajuće modifikacije, može se iskoristiti za detekciju početka mišićne kontrakcije kao što je prikazano na Sl. 103. Ovde neće biti prikazano, ali jednostavnom izmenom koda gde bi se obvojnica predstavila u obrnutom redosledu, ponovo bi se mogli odrediti trenuci početka mišićne kontrakcije (koji bi sada predstavljali kraj mišićne kontrakcije). Niz podataka se može prikazati u obrnutom redosledu jednostavnom primenom *rev()* funkcije čija je upotreba prikazana na Sl. 80 na kojoj je predstavljen izvorni kod za *filtfilt()* funkciju.

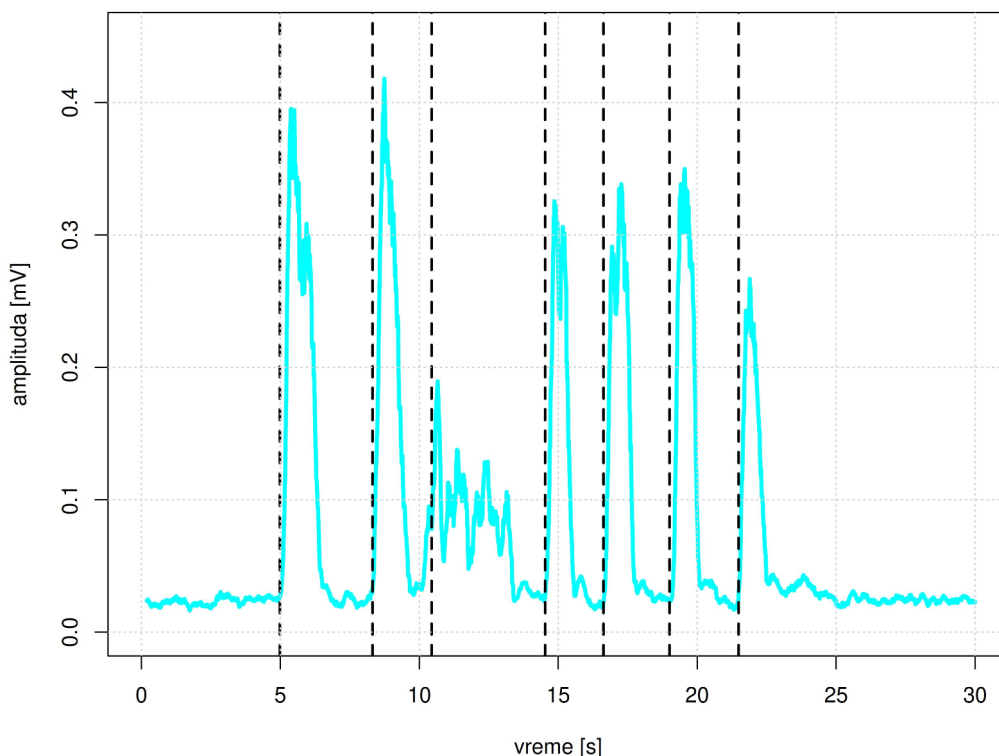
Modifikacija *findpeaks()* funkcije je posebno pripremljena za ovaj udžbenik, te će ovde biti objašnjeni elementi na koje bi trebalo obratiti pažnju na Sl. 103 i Sl. 104:

- funkciju *findpeaks()* bi trebalo primeniti na nizu, a ne na *data frame*-u ili nekom drugom tipu podataka, te je zato iskorišćena *as.vector()* funkcija,
- prvih 200, tačnije 199, odbiraka je prazno, zbog primene MA filtra, te je potrebno koristiti signal bez tih odbiraka,
- prag je definisan u funkciji standardne devijacije EMG obvojnice (promenljiva *prag*),
- najvažniji korak pre primene funkcije za pretragu lokalnih ekstremuma jeste da se sve vrednosti koje su veće od praga postave na vrednost praga – tako se omogućava da lokalni ekstremum bude određen u vremenskom trenutku u kom se nalazi početak EMG obvojnice²⁶⁷, što za posledicu ima duplo više pikova, odnosno duplo više mišićnih kontrakcija u rezultatu *findpeaks()* funkcije (14 umesto 7),

²⁶⁷ Uobičajeno se kaže početak mišićne kontrakcije za početak obvojnice EMG signala, odnosno početak električne aktivnosti mišića. Precizno, prvo se pojavi električna aktivnost u mišiću, a nakon određenog vremenskog intervala (elektromehaničko kašnjenje) nastaje pokret (mehanički odgovor).

- u predloženoj realizaciji visina pika odgovara standardnoj devijaciji, odnosno polovini definisanog praga, minimalno rastojanje između pikova (koje odgovara pretpostavljenom minimalnom rastojanju između dve obvojnice je jednako $2s$, odnosno dva puta više odbiraka od fs – frekvencije odabiranja) i argument *nups* kojim se definiše minimalan broj koraka pre nego što se postigne pik je postavljen na 0, jer važno odrediti početak obvojnice EMG signala.

Treća kontrakcija po redu je drugačijeg intenziteta i trajanja i nije je tako jednostavno detektovati. U većini praktičnih situacija, kontrakcije čiji se početak i kraj određuju su sličnih intenziteta i trajanja, tako da se problem koji je ilustrovan na Sl. 102 i Sl. 103 retko sreće u praksi. Na Sl. 104 je prikazan R kod koji je korišćen za detekciju početka mišićne kontrakcije EMG signala (Sl. 102 i Sl. 103).



Slika 103, Prikaz obvojnice EMG signala (kao na Sl. 102) sa detektovanim početkom mišićne kontrakcije primenom *findpeaks()* funkcije iz *pracma* paketa.

Rezultat detekcije početka mišićnih kontrakcija je određen za 7 kontrakcija na Sl. 103, ali prikazan rezultat na Sl. 104 uključuje 14 ekstremuma. Poreklo 14

lokalnih ekstremuma je u tome što je oblik signala promenjen i u načinu na koji su birani parametri *findpeaks()* funkcije. Rezultat *findpeaks()* funkcije ima 4 kolone (broj vrsta zavisi od broja detektovanih tačaka) u kojima su smešteni:

- u prvoj koloni je amplituda detektovanog pika,
- u drugoj koloni je indeks na kome se nalazi detektovan pik,
- u trećoj koloni je dat početak pika i
- u četvrtoj koloni je dat kraj detektovanog ekstremuma.

Metode za detekciju početka obvojnice EMG signala koje se zasnivaju na pragu jesu relativno brze i mogu se primeniti u realnom vremenu ili kvazi realnom vremenu. Njihova velika mana je što ne postoji standardizovan način kako odabrati parametre za primenu metode (visina praga i rastojanje između dve kontrakcije). Takođe, odabir parametara igra značajnu ulogu, jer je metoda veoma osetljiva na promenu ulaznih argumenata. Znatiželjni studenti mogu iskoristiti podeljen programski kod da promene bilo koji od ulaznih parametara te da se sami uvere u osetljivost metode za detekciju početka mišićne kontrakcije primenom praga.

Da bi se dodale vertikalne linije na Sl. 103 primenjena je funkcija *abline()* na Sl. 104 koja omogućava dodavanje vertikalnih i/ili horizontalnih linija (granica) na grafiku (može da se koristi i za linearnu regresiju). Ova funkcija pripada osnovnom grafičkom paketu i koristi se nakon poziva *plot()* funkcije. Mišićne kontrakcije nižeg intenziteta predstavljaju vrlo izazovne zadatke za detekciju početka i kraja mišićne kontrakcije. Pored toga, tip mišićne kontrakcije može predstavljati izazov, jer ako povećanje intenziteta amplitude traje veoma dugo, onda se može javiti problem izdvajanja takve kontrakcije. Trenutno, ne postoji standard u odabiru metode za detekciju početka i kraja mišićne kontrakcije. Kada se ispituju nove tehnike za detekciju početka i kraja mišićne kontrakcije postoje dva uobičajena pristupa: (1) da se koristi manuelna detekcija od strane stručnjaka kao nominalna vrednost (eng. *ground truth*) i (2) da se koriste simulirani biosignali. Pored velikog poverenja u manuelnu detekciju koja se najčešće obavlja uz vizuelnu inspekciju (označavanjem početka i kraja mišićne kontrakcije na grafiku), ova metoda nije poželjna jer nema dovoljno dobru reproducibilnost, a i vremenski je veoma zahtevna. Dodatno, algoritmi za detekciju mogu da uključuju različite korake za pretprocesiranje, ali i za postprocesiranje. Primeri pretprocesiranja su korišćenje filtera u frekvencijskom i vremenskom domenu ili čak kvadriranje²⁶⁸, a primer postprocesiranja je

268 Kvadriranje je uobičajena tehnika koja se primenjuje u algoritmima za detekciju pika. Koristi se da se signal transformiše tako da pikovi budu što izraženiji (predstavlja pretkorak u primeni detektora koji se zasnivaju na poređenju sa amplitudskim pragom). [20]

donošenje naknadne odluke da li neki trenutak predstavlja ili ne predstavlja verodostojan početak/kraj mišićne kontrakcije, tj. da li se nalazi u fiziološkim granicama koje su očekivane za primenjen tip mernog protokola. [344-345]

```
> emgObvojnica <- as.vector(datEMGkablMA)
> emgObvojnica <- emgObvojnica[200:length(emgObvojnica)]
> prag <- 2*sd(emgObvojnica)
> emgObvojnica[emgObvojnica > prag] <- prag
> onoff <- findpeaks(emgObvojnica, minpeakheight = prag/2,
+                   minpeakdistance = 2*fs, nups = 0)
> onoff
```

	[,1]	[,2]	[,3]	[,4]
[1,]	0.1839326	5018	4983	5018
[2,]	0.1839326	5018	5018	5018
[3,]	0.1839326	8321	8318	8321
[4,]	0.1839326	8321	8321	8321
[5,]	0.1839326	10446	10443	10446
[6,]	0.1839326	10446	10446	10446
[7,]	0.1839326	14532	14529	14532
[8,]	0.1839326	14532	14532	14532
[9,]	0.1839326	16633	16627	16633
[10,]	0.1839326	16633	16633	16633
[11,]	0.1839326	19061	19004	19061
[12,]	0.1839326	19061	19061	19061
[13,]	0.1839326	21500	21492	21500
[14,]	0.1839326	21500	21500	21500

Slika 104, R kod za računanje početka mišićnih kontrakcija. Rezultat detekcije je prikazan na Sl. 103.

Kako EMG signal vizuelno i po svojim karakteristikama podseća na govor (audio signal), to se često koristi i Tajger-Kajzerov energetska operator koji je proporcionalan trenutnoj amplitudi signala i frekvenciji za detekciju početka kontrakcije. Trenutno je veoma aktuelan razvoj novih metoda za detekciju početka i kraja mišićne kontrakcije na osnovu algoritama mašinskog učenja, ali je još uvek dominantno zastupljena metoda poređenja sa pragom. Razlog tome je što su metode koje se zasnivaju na pragu jednostavne, brze i pouzdane. U zavisnosti od vrste praga koji se koristi, postoje: (1) metode sa jednim pragom (kao što je prikazano na Sl. 103 primenom *findpeaks()* funkcija sa Sl. 104), (2) metode sa dva praga i (3) metode zasnovane na primeni adaptivnog praga. Iznenadjuć je rezultat poređenja ove tri metode koji je ukazao da nema

značajnih razlika u tačnosti detekcije praga, osim u brzini primene koja je poželjna kod trenutne detekcije (metoda sa jednim pragom i metoda sa adaptivnim pragom su se pokazale kao značajno brže u poređenju sa metodom sa dva praga). Dodatno, metode koje primenjuju jedan prag kao što je pokazano na primeru na Sl. 103 su računarski najmanje zahtevne. Veličina praga se može odrediti na osnovu standardne devijacije (Sl. 104), u zavisnosti od amplitude maksimalne voljne ili samo od maksimalne amplitude merenog EMG signala. Za prag određen primenom standardne devijacije, u zavisnosti od tipa signala i mišićnih kontrakcija, može se uzeti da je prag jednak jednoj do tri standardne devijacije amplitude EMG signala. [345]

Po iskustvu Autorke, detekcija početka sa jednim pragom jeste često u upotrebi, ali u slučaju različitog trajanja mišićnih kontrakcija, jedna od najboljih metoda za detekciju početka mišićne kontrakcije se zasniva na određivanju koeficijenta pravca kroz dve tačke. Naime, na obvojnici EMG signala se dve odabrane tačke (na odgovarajućem rastojanju) pomeraju sekvencijalno po obvojnici u vremenu i računa se koeficijent pravca prave kroz dve tačke. U momentu kada koeficijent pređe neki prag (koji se najčešće jednostavno određuje, jer je porast koeficijenta pravca značajno veći u trenutku početka kontrakcije u odnosu na relaksaciju), moguće je odrediti početak kontrakcije. Slična metoda se primenjuje u analizi ventilacionih pragova (eng. *ventilatory thresholds*) tokom fizičkog napora [346].

2.3.2.2 Određivanje medijana frekvencije EMG signala

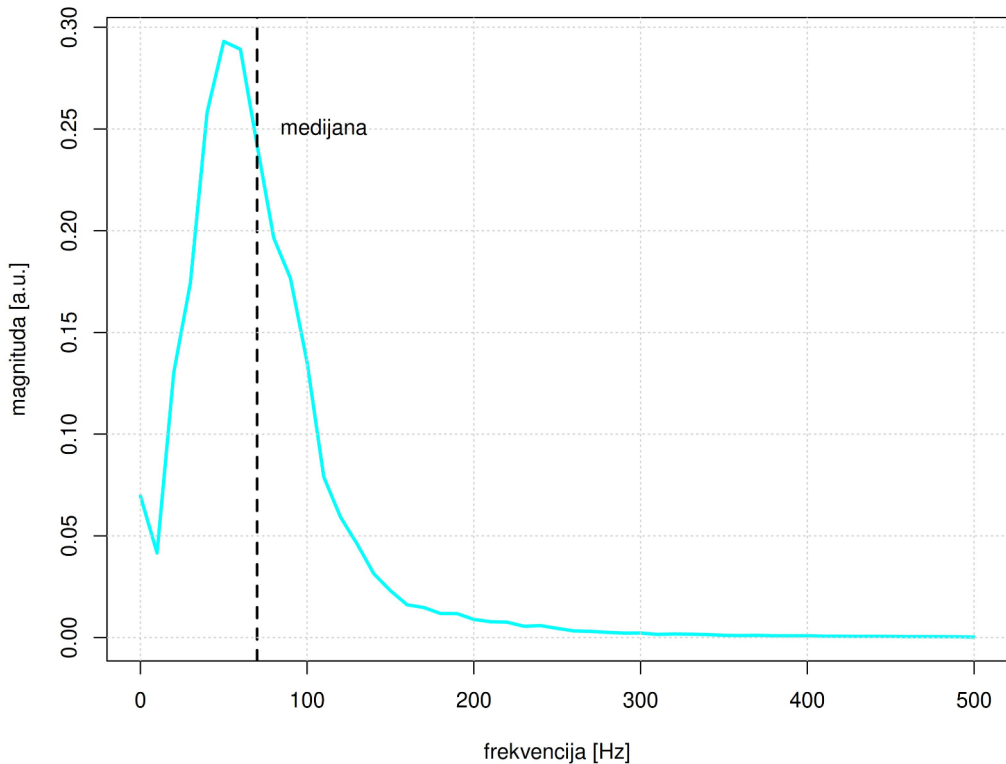
Pored izdvajanja obeležja u vremenskom domenu, EMG signal se može parametrizovati i u frekvencijskom domenu. Uobičajen način predstavljanja EMG signala, kao i drugih biosignala, jeste primenom Velčove metode²⁶⁹ za prikaz spektralne gustine snage signala (eng. *Power Spectral Density*, skraćeno PSD). Spektralna gustina snage EMG signala iz datoteke "emg-sve.txt" za signal u opsegu od 50 s do 80 s od početka merenja je prikazana na Sl. 105. Da bi se dobio Velčov spektar signala (odnosno Velčova procena spektralne gustine snage EMG signala) korišćena je funkcija:

welchPSD()

iz *bspec* paketa (od eng. *Bayesian Spectral Interface*) [347-348]. Velčova metoda za procenu spektralne gustine snage signala je pogodna za

269 https://en.wikipedia.org/wiki/Welch%27s_method, pristupljeno 22.03.2024.

nestacionarne signale [349], što odgovara prirodi većine biosignala, odnosno elektrofizioloških signala.



Slika 105, Medijana EMG signala obeležena na spektralnoj gustini snage²⁷⁰ EMG signala.

Velčova metoda koristi prozorske funkcije i zasniva se na usrednjavanju periodograma. Najpre se izdvajaju segmenti signala primenom prozorskih funkcija sa preklapanjem od 50% i Hamingovim prozorom. Funkcija u R-u podrazumevano primenjuje Tukijev prozor. Za svaki od izdvojenih segmenata se računa brza Furijeova transformacija. Na kraju, sve Furijeove transformacije se sinhrono usrednjavanju ili se računaju njihove medijane. Korišćenje *welchPSD()* funkcije za računanje medijana frekvencije (odnosno medijane) EMG signala (koja je prikazana na Sl. 105) je dato na Sl. 106. Primititi da je potrebno odabrati *seglength* parametar (dužina prozora) za primenu *welchPSD()* na Sl. 106.

Podrazumeva se da korisnik u R-u ili bilo kom drugom programskom jeziku može samostalno da definiše prozorsku funkciju, preklapanje prozora u vremenskom domenu, kao i metodu sinhronog usrednjavanja ili sinhronog

²⁷⁰ U literaturi se koristi i termin spektar snage signala.

računanja medijane, te korišćenje gotovih paketa nije potrebno, ali jeste korisno. Studenti se upućuju na R dokumentaciju, specifično na uputstvo za korišćenje `bspec` paketa [347-348]. Na kraju, Velčova metoda je posebno zanimljiva, jer u odnosu na klasičnu Furijeovu transformaciju ili procenu spektralne gustine snage signala primenom Furijeove transformacije, predstavlja uglačanu verziju spektralne gustine snage signala (Sl. 105). S toga se može smatrati, iako to nije baš najpravilnije, filtriranom verzijom spektralne gustine snage signala.

```
fs <- 1000
datEMGkablAll <- read.table("emg-sve.txt")

datEMGkabl <- datEMGkablAll$V1[(50*fs):(80*fs)]
timeAxisKabl <- seq(0, length(datEMGkabl)/fs - 1/fs, by = 1/fs)

# računanje Velčovog spektra snage
datEMGkabl <- as.ts(datEMGkabl, frequency = fs)
psdw <- welchPSD(datEMGkabl, seglength = 100)

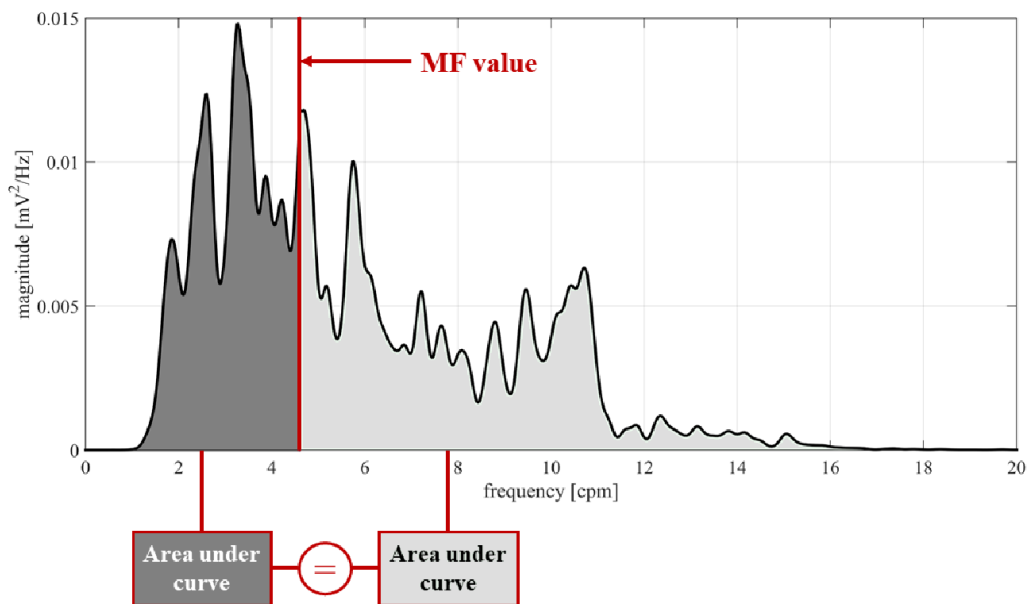
# određivanje medijane
psdw05 = sum(psdw$power)/2;
psdwmed = 0;
ind <- 1
while (psdwmed <= psdw05) {
  psdwmed = psdwmed + psdw$power[ind];
  ind <- ind + 1
}

fmed <- psdw$frequency[ind]*fs
```

Slika 106, R kod za računanje medijane EMG signala koja je prikazana sa Velčovom procenom spektralne gustine snage signala na Sl. 105.

Slika 107 prikazuje spektralnu gustinu snage EGG signala na kojoj je ilustrovano kako je moguće odrediti medijanu (eng. *MF value*). Medijana predstavlja frekvenciju na kojoj se spektralna gustina snage deli na dva dela jednake površine (eng. *Area under curve*). Pored toga što EGG signal predstavlja električnu aktivnost glatkih mišića želuca, ne može se analizirati na isti način kao EMG signal. Kod EMG signala pomeranje spektralne gustine snage signala u desno (ka višim frekvencijama) se dešava tokom većeg intenziteta mišićne kontrakcije, dok se pomeranje ka višim frekvencijama kod EGG signala dešava kada želudačna aktivnost postaje ubrzana (tzv. tahigastrija)

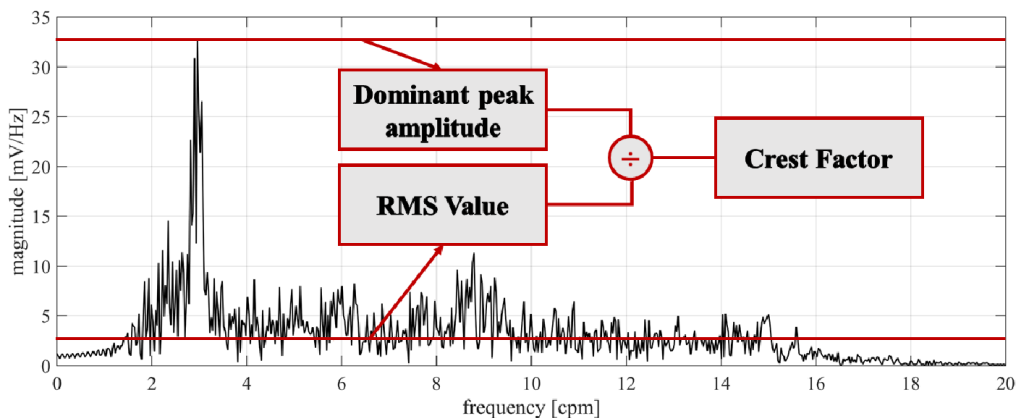
[266]. Dakle, porast medijane ima različito značenje kod EMG i EGG signala, jer se promena medijane EGG signala može koristiti da se opiše promena oblika spektra i disritmična aktivnost želuca.



Slika 107, Ilustrativni primer računanja medijane na spektru EGG signala. Slika je podeljena pod otvorenom licencom i dostupna u [266]. Primiteti da je umesto u Hz, frekvencija prikazana u jedinicama cpm od eng. *cycles per minute* – razlog je što su frekvencije veoma male, pa se pribegava korišćenju alternativnih jedinica.

Procena medijana frekvencije u spektru snage signala je vrlo korisna, ako se EGG signal koristi u oceni mučnine nastale tokom korišćenja virtuelne realnosti ili simulatora vožnje [79, 266, 268-269, 350-351]. Medijana frekvencija se takođe pokazala korisnom za karakterizaciju gustine spektralne snage elektrohisterograma sa ciljem predikcije prevremenog porođaja sa statistički značajnom razlikom [352]. Za karakterizaciju spektralne gustine snage signala ili Furijeovu transformaciju biosignala, mogu se koristiti i druge mere/parametri, pored medijane. Na Sl. 108 je prikazano kako se za procenu istaknutosti dominantnog pika EGG signala koji je određen normalnom ritmičnom stomačnom aktivnošću (normogastrija) i koji se nalazi u opsegu od 2 cpm (eng. *cycles per minute*) do 4 cpm [79, 266, 268-269] može koristiti vršni faktor šuma (eng. *Crest Factor*, skraćeno CS). Da bi se izračunao CS potrebno je podeliti magnitudu, tj. amplitudu pika sa srednje kvadratnom vrednošću okoline pika (RMS). Za istaknutije pikove CS je veće i obrnuto. Računanje parametara za karakterizaciju oblika Furijeove transformacije i spektralne gustine snage biosignala primenom frekvencije medijane i vršnog faktora šuma

su od značaja u objektivnoj kvantifikaciji događaja koji predstavljaju uzrok posmatranih promena.



Slika 108, Ilustrativni primer računanja vršnog faktora šuma (eng. *Crest Factor*) za Furijeovu transformaciju EGG signala. Slika je podeljena pod otvorenom licencom i dostupna u [266]. Skraćenica RMS označava srednje kvadratnu vrednost (eng. *Root Mean Square*).

Pored Velčove metode, za procenu spektra snage signala, može se kvadrirati magnitudska frekvencijska karakteristika signala dobijena iz Furijeove transformacije signala, a u upotrebi su i druge metode. Zanimljivo je spomenuti metodu koju je moguće primeniti u vremenskom domenu – brojanje prolazaka kroz nulu. Ima dva razloga zašto je istorijski brojanje prolazaka kroz nulu bila vrlo uobičajena metoda za analizu EMG signala. Jedan je što brojanje prolazaka kroz nulu ne zavisi od amplitude signala (ne mora se normalizovati amplituda EMG signala), a drugi je zato što je vremenski manje zahtevna od metoda u frekvencijskom domenu. Drugi razlog je omogućio i implementaciju u realnom vremenu sa ciljem zamene frekvencije medijane sa brzinom prolaska signala kroz nulu (eng. *Zero-Crossing Rate*, ZCR) [353]. Interpretacija ZCR je jednostavna: ako postoji veći broj prolazaka kroz nulu tada je frekvencijski sadržaj signala pomeren ka višim frekvencijama i obrnuto.

Medijana frekvencija je jednostavan i koristan parametar u analizi EMG signala, ali i drugih biosignala. Njena najveća praktična korist u analizi EMG signala je za detekciju zamora. Zbog karakterističnih uslova u kojima se zamor javlja, ovakvu vrstu analize treba pažljivo sprovesti [354]. Pored toga što na prvi pogled, EMG signal ostavlja utisak biosignala koga je jednostavno analizirati, to u stvarnosti nije tako. EMG signal nastaje kao rezultat prenosa akcionih potencijala iz nervnog sistema, te istovremeno može služiti za ocenu mišićno-skeletnog i nervnog sistema. Dodatno, višekanalno merenje EMG signala ima viši nivo kompleksnosti, posebno ako se uzmu u obzir sinergističke

mišićne kontrakcije tokom različitih složenih fizičkih aktivnosti (npr. plivanje, trčanje, hod i skakanje).

2.3.2 Delineacija u obradi EKG biosignala za izdvajanje karakterističnih obeležja

Detekcija početka i kraja mišićne kontrakcije, kao i metode za određivanje medijane i vršnog faktora šuma koriste se za računanje karakterističnih tačaka/obeležja na signalu u vremenskom i frekvencijskom domenu. Najpoznatiji primer iz analize biosignala za detekciju karakterističnih tačaka je primer delieancije EKG biosignala. Jednostavna funkcija za pretragu lokalnih ekstremuma – *findpeaks()* funkcija radi sasvim zadovoljavajuće, ali samo u kontrolisanim uslovima, odnosno u uslovima sa smanjenim uticajem šuma tj. većim SNR. U studiji u kojoj su mereni EKG signali u sedećem položaju na zdravim ispitanicima (studentima psihologije), jedini zadatak ispitanika je bio da slušaju audio signal i da sede mirno, te nije bilo potrebe za korišćenjem neke druge metode [355]. Ali, bilo je potrebno uvesti vizuelnu inspekciju detektovanih pikova – iako predstavlja manuelnu (u inženjerskom slengu ruralnu) metodu, vizuelna inspekcija je i dalje zlatni standard u analizi biosignala [79, 266]. Najčešće²⁷¹ korišćena metoda za detekciju dominantnog lokalnog ekstremuma na EKG signalu (R pika) je Pan-Tompinksova metoda iz 1985. godine [20] koja uključuje sledeće korake u detekciji R pikova, odnosno QRS kompleksa, složenoperiodičnog EKG signala:

- filtriranje filtrom propusnikom opsega,
- diferenciranje signala,
- kvadriranje signala,
- primenu integratora (filtriranje MA filtrom) za glačanje signala,
- primenu dva adaptivna praga i
- eliminaciju kašnjenja koje je nastalo filtriranjem signala.

Značajna je primena Pan-Tompkins algoritma za detekciju R pikova u realnom vremenu, ali i robusnost u različitim situacijama kada se menja odnos

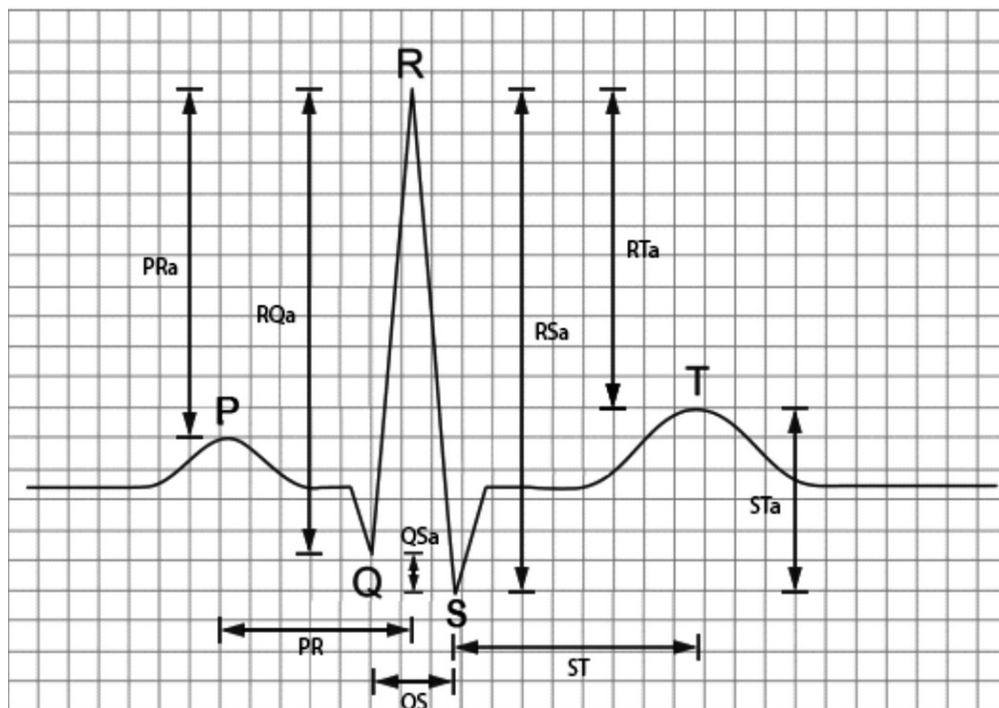
²⁷¹Na dan 13. aprila 2024. godine rad u kome su Pan i Tompkins predstavili metodu za detekciju R pika, odnosno QRS kompleksa je citiran 9160 puta (izvor – Google Scholar, https://scholar.google.com/scholar?cites=2948543122161795956&as_sdt=2005&scioldt=0,5&hl=en, pristupljeno 13.04.2024).

signal/šum (SNR) ili kada dođe do promene srčanog ritma. Pan-Tompkins metoda predstavlja polaznu osnovu za detekciju lokalnih ekstremuma svih biosignala koji imaju izražene pikove (na primer, detekcija lokalnog ekstremuma IKG signala). Takođe, vremenom su se razvile razne modifikacije ovog algoritma, ali osnova je ostala ista [20]:

- filtriranje propusnikom opsega se obavlja u što užem propusnom opsegu – cilj ovog filtriranja nije da se sačuva kompletna morfologija EKG signala (koja je ilustrativno prikazana na Sl. 109), već da se sačuva informacija od značaja (QRS kompleks),
- diferenciranje signala se obavlja sa ciljem izdvajanja pika, odnosno sa ciljem povećanja istaknutosti R pika – ako se pogleda karakterističan oblik EKG signala na Sl. 109 primećuje se da je brzina promene signala najveća u okolini R pika (postoji najveći nagib na signalu) u odnosu na okolinu P i T talasa,
- kvadriranje signala je logičan korak koji se obavlja sa ciljem da se poveća istaknutost pika, jer kvadriranje omogućava da se postignu dva efekta na modifikovanom EKG signalu: (1) da reprezentacija signala bude pozitivna, što olakšava automatsko definisanje praga i (2) da razlike između amplituda EKG signala postanu još veće²⁷²,
- glačanje signala predstavlja pripremu za primenu praga na signalu – ako signal nije dovoljno uglačan, primena praga/pragova može rezultovati lažnim detekcijama pikova,
- primena praga je uobičajena i jednostavna procedura (postiže se primenom *findpeaks()* funkcije u R-u i drugim programskim jezicima), ali kako bi se izbegli uticaji šumova na signalu i signal prilagodio promenama srčanog ritma (koji je određen pojavom R pikova, odnosno dužinom rastojanja između susednih R pikova – RR intervala) Pan-Tompkins metoda koristi dva adaptivna praga i
- kako se detekcija pikova obavlja na transformisanom signalu, koji više ne podseća na EKG signal, to je važno da se mesta detekcije postave tačno na QRS intervale, odnosno R pikove tako što će se proceniti

²⁷² Za slučaj da nije intuitivno jasno, povećanje razlike između dva pika primenom kvadriranja je jednostavno objasniti na primerima dva broja: ako je jedan pik amplitude 3 (a.u.), a drugi amplitude 2, njihov odnos je $3/2$ (1.50), međutim posle kvadriranja odnos postaje $9/4$ (2.25).

kašnjenje koje uvode filtri (za izdvajanje propusnog opsega od interesa i za glačanje signala) – treba imati na umu da je ovo jednostavno ako se radi o filtrima konstantne ili linearne faze, a da je praktično nemoguće izvesti ovaj korak, ako filtri imaju nelinearnu fazu, te treba voditi računa o odabiru filtara u Pan-Tompkins metodi.



Slika 109, Delineacija EKG signala i odabrana obeležja [355]. Podeljeno pod CC BY 4.0 DEED licencom, <https://creativecommons.org/licenses/by/4.0/>, pristupljeno 12.04.2024.

Zanimljivo je istaći da ovo nije prva metoda koja je uključila računanje prvog izvoda kao pretprocesiranje za detekciju pika. Brzina promene električne aktivnosti EKG signala nema praktičnog značaja, koliko je Autorki poznato, ali brzina promene pokreta oka ima fiziološki značaj. Još 1981. Bahili i saradnici su predložili metodu za detekciju pikova, odnosno sakada, iz signala pokreta oka primenom prvog izvoda – umesto pozicije oka koristili su brzinu pokreta oka, jer je takva metoda omogućila tačniju i precizniju detekciju pika sakade, kao i početka i kraja pika [356-357]. Sakada predstavlja brz pokret oba oka u istom pravcu između fiksacija oka na neku tačku u prostoru [358]. Intenzivno

se koriste u psihologiji, jer omogućavaju da se na jednostavan način utvrdi kako subjekti gledaju scenu od interesa [359-360].



Slika 110, Trajektorija pokreta oka (desni panel) ispitanika koji je gledao u lice koje je prikazano na levom panelu. Izvor: Autor: SpooSpa. Derivativ: Simon Viktória – od Face of SpooSpa.jpg, Podeljeno pod CC BY-SA 2.0 licencom, <https://commons.wikimedia.org/w/index.php?curid=8711778>, pristupljeno 14.04.2024.

Tipično, sakade se javljaju prilikom čitanja teksta i to kada se pogledom prelazi u novi red. Za merenje ovih signala mogu se koristiti dva tipa uređaja: (1) EOG koji omogućava da se meri kretanje oka u horizontalnoj i vertikalnoj ravni, a princip merenja ovog signala se svodi na model oka kao dipola i (2) primenom uređaja za praćenje pokreta oka (eng. *eye tracker device*) koji rade na principu praćenja pokreta zenica i snimanja scene koju gleda ispitanik. Primer trajektorije pokreta oka prilikom gledanja lica je prikazan na Sl. 110.

Bilo da se detektuje R pik EKG signala, određuje pik na trajektoriji pokreta oka (sakada) ili se vrši detekcija pika nekog drugog signala, može se koristiti prvi izvod signala. Za detekciju ekstremuma nižeg intenziteta, kao što je tzv. *dicrotic notch* na fotopletizmogramu ili tzv. B tačka na impedansnom kardiogramu, u upotrebi je izvod drugog reda [99-100, 186, 286]. Međutim, primena izvoda može biti štetna u obradi signala, jer može značajno pojačati šum. Očekivano, sa povećanjem reda izvoda, povećava se i šum. Jedan od načina da se reši

problem sa prvim izvodom, jeste da se izvod uopšte ne koristi i da se iskoriste druge metode kao što je adaptivno skaliranje amplitude signala [99-100] ili da se iskoristi frakcioni račun (eng. *fractional calculus*) [237, 274]. Razlomljeni izvod (ili izvod sa frakcionim redom ili frakcioni izvod od eng. *fractional order derivative*) predstavlja generalizovan izvod signala [361]. Za razliku od standardnog izvoda signala koji koriste kao red celobrojne vrednosti (na primer, izvod prvog reda i izvod drugog reda), frakcioni izvod se može definisati za realne brojeve (na primer, frakcioni izvod reda 0.3). Ova generalizacija je korisna u primeni na biosignalima jer omogućava veći manevarski prostor da se pomire dva oprečna zahteva: (1) da se maksimalno istakne pik i (2) da se maksimalno smanji pojačanje šuma primenom izvoda. Frakcioni izvod se pokazao korisnom u analizi biosignala. Na primer, ako je potrebno razdvojiti EKG signal trudnice od EKG signala fetusa koji su snimljeni postavljanjem površinskih elektroda na stomak trudnice, onda je moguće izdvojiti EKG signal fetusa (koji je značajno nižih amplituda od EKG signala trudnice) primenom frakcionog izvoda. Pokazano je da je sa frakcionim izvodom reda ~ 0.5 moguće transformisati signal tako da R pikovi EKG signala fetusa postanu veći od R pikova EKG signala trudnice, što je posledica toga da je EKG signal fetusa značajno brži od EKG signala trudnice [237, 362].

Jedna od najpoznatijih formi frakcionog izvoda je Grunvald-Letnikova forma koja je korišćena za detekciju R pikova EKG signala fetusa u okviru modifikovane Pan-Tompkins metode [237, 362]. Prvi izvod signala / funkcije $f(x)$ u oznaci $f'(x)$ se može definisati kao u Relaciji (19).

$$f'(x) = \frac{f(x+h) - f(x)}{h} \quad \text{Relacija (19)}$$

U Relaciji (19) je sa h označena $h = x_2 - x_1$. Generalizovana forma izvoda celobrojnog reda n u oznaci $f^{(n)}$ je prikazana u Relaciji (20).

$$f^{(n)}(x) = \frac{1}{h^n} \sum_{m=0}^n (-1)^m \binom{n}{m} f(x - mh) \quad \text{Relacija (20)}$$

Potpuna generalizacija izvoda n -tog reda Grunvald-Letnikovom formom za frakcioni red α u oznaci $f^{(\alpha)}(x)$ je data u Relaciji (21), gde je sa $\Gamma()$ označena Gama funkcija, dok t i a predstavljaju gornju i donju granicu.

Relacija (21)

$$f^{(\alpha)}(x) = \frac{1}{h^\alpha} \sum_{m=0}^{\frac{t-a}{h}} (-1)^m \frac{\Gamma(\alpha+1)}{m! \Gamma(\alpha-m+1)} f(x-mh)$$

Za detaljniji prikaz frakcionog računa studenti su upućuju na dodatne reference i materijale u [95, 237, 362-366], kao i na predmet Distribuirani i frakcioni sistemi upravljanja (13M051DIF²⁷³) na master akademskim studijama na Elektrotehničkom fakultetu, Univerziteta u Beogradu. Koliko je Autorki poznato, u R-u ne postoji za sada poseban paket koji se bavi frakcionim računom. Međutim, veliki broj skripti je već dostupno na internetu u drugim programskim jezicima, te ih je jednostavno implementirati u R-u.

Matlab funkcija *glfdiff()*²⁷⁴ predstavlja primer jednostavne implementacije frakcionog izvoda koja se pravolinjski prevodi u R kod [367-368]. Za test signal (originalni signal) dat u Relaciji (22) koja je ilustrativno korišćena u [369], rezultat primene izvoda sa frakcionim redovima 0.2, 0.4, 0.6, 0.8, 1.0 i 1.2 (primetiti da je uključen i celobrojni prvi izvod) zajedno sa originalnim signalom je prikazan na Sl. 111. Kako bi se direktno poredili efekti koje frakcioni izvodi imaju na signalu – prikaz svih signala/funkcija je normalizovan na maksimalnu vrednost na Sl. 111. Uočava se da su izvodi signala nižeg reda bliski po obliku originalnom signalu, što ukazuje na selektivnost koja se postiže primenom frakcionog izvoda u odnosu na celobrojni izvod. Takođe, moguće je uočiti da primena frakcionih izvoda unosi manji šum na signalu.

$$f(t) = 1 + \sin(0.5 * 2 \pi t) + 0.4 \sin(2 \pi t) - 0.2 \sin(4 * 2 \pi t) \quad \text{Relacija (22)}$$

Kako funkcija *glfdiff()* nije podeljena pod otvorenom licencom, njena realizacija u R-u nije dostupna u skripti kao propratni materijal udžbenika, a studenti mogu funkciju iz Matlaba da prevedu u R. Jedini element *glfdiff()* funkcije koji može biti zbunjujuć je:

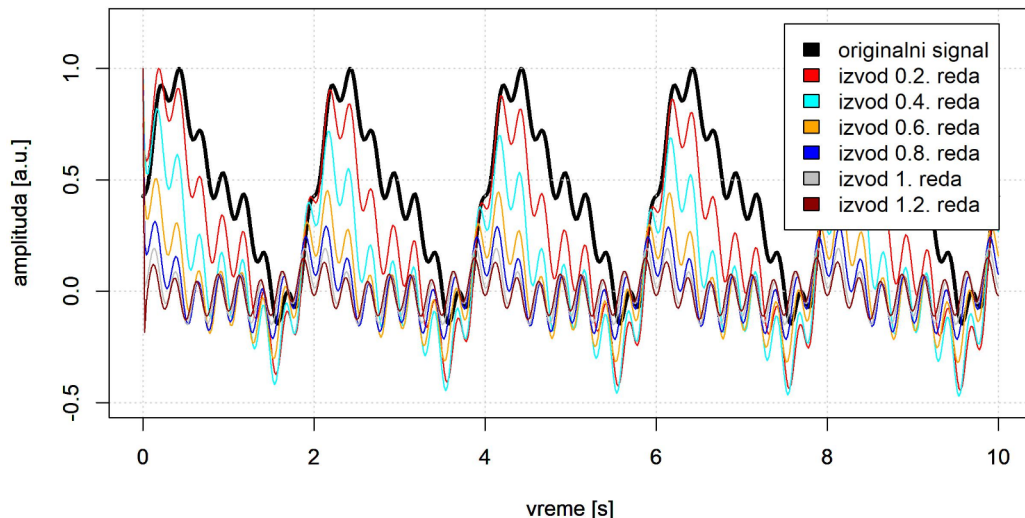
```
*[f(ind:-1:1)]
```

i koji predstavlja matrično množenje (zbog [] u Matlabu) sa nizom *f* u kome je dat obrnut redosled poziva indeksa (od indeksa *ind* do 1), što se u R-u može realizovati pomoću funkcije *rev()* za prikaz elemenata u obrnutom redosledu i primenom operatora za matrično množenje (%*%):

273 https://www.etf.bg.ac.rs/sr-lat/fis/karton_predmeta/13M051DIF-2019, pristupljeno 14.04.2024.

274 <https://www.mathworks.com/matlabcentral/mlc-downloads/downloads/submissions/60874/versions/1/previews/fractional/glfdiff.m/index.html>, pristupljeno 01.06.2024.

%*% rev(f[1:ind])



Slika 111, Primer frakcionih izvoda originalnog signala datog u Relaciji (22). Signali (uključujući i originalni signal) su za potrebe poređenja na grafiku normalizovani na maksimalnu vrednost.

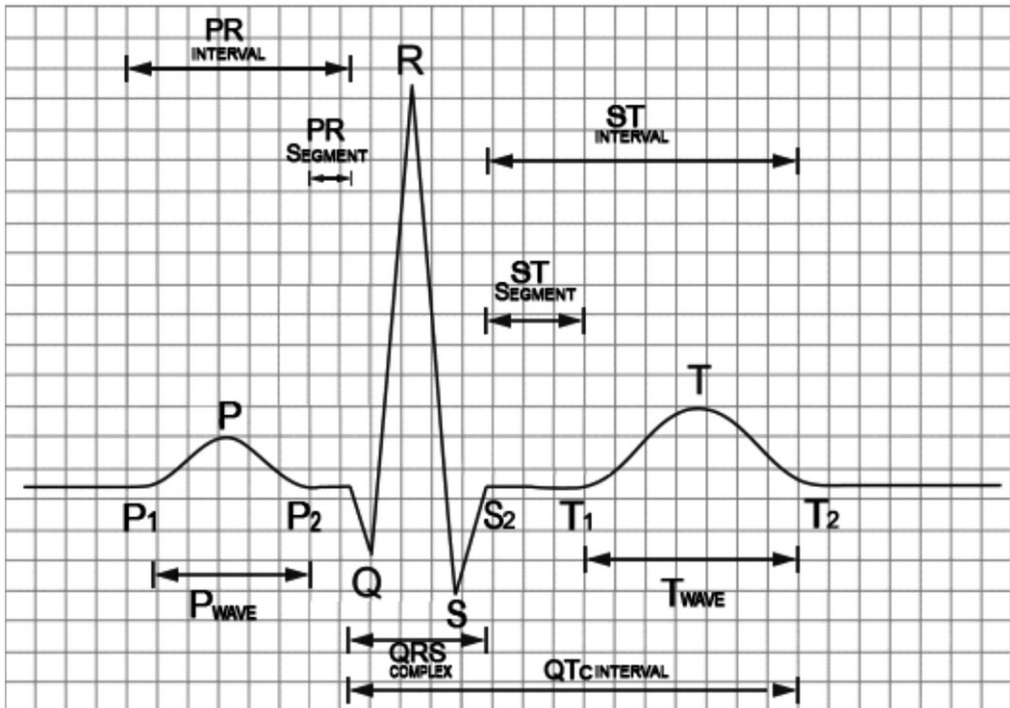
Nezavisno od metode pretprocesiranja EKG signala, proces delineacije EKG signala je složen, jer obuhvata detekciju svih karakterističnih tačaka na EKG signalu (P, Q, R, S i T pikovi, Sl. 109), te je detekcija R pika samo početni korak u delineaciji signala [355, 370-371]. Pored detekcije karakterističnih pikova, oni koji analiziraju EKG signale se mogu susresti sa još dva izazova koja nisu na odgovarajući način predstavljena u literaturi:

- Postoje klinički relevantni parametri i oni koji to nisu. Ovo je logična podela, ali nije uvek u naučnim i stručnim radovima naglašena i očigledna. Prema principima medicine zasnovane na dokazima, svaki parametar za koji je u kliničkoj studiji statistički pokazano da predstavlja indikator nekog oboljenja, taj se parametar smatra klinički značajnim i koristi se u kardiološkoj praksi. Prikaz klinički značajnih intervala/parametara je dat na Sl. 112. Parametri koji su prikazani na Sl. 109 su korišćeni u naučnoj literaturi (na primer, za biometrijsku identifikaciju ispitanika²⁷⁵), ali ne postoji fiziološki osnov koji bi ukazao

275 Relativno veliki broj biomarkera u ljudskom telu, uključujući i elektrofiziološke signale, a posebno EKG, ima biometrijske karakteristike [370-371]. Drugim rečima, svaka osoba ima jedinstven EKG signal, kao što ima jedinstven otisak prsta.

kako su ti parametri povezani sa zdravljem ispitanika i sa električnom aktivnošću srca.

- Izoelektrična linija na EKG signalu predstavlja trenutak u kome nema električne aktivnosti srca. Koristi se kao polazna osnova za merenje relativne amplitude R pika i nalazi se pre početka P talasa na EKG signalu – tačnije u opsegu od 180 ms do 130 ms pre pika P talasa [355, 372]. Logično je pretpostaviti da se filtriranjem signala, odnosno uklanjanjem bazne linije izoelektrični deo postavlja u bliskoj okolini nule, ali to nije tačno. Na Sl. 113 je ilustrovano kako za dva različita ispitanika (ID7 i ID9 iz otvorene baze podataka [355, 373-374]) izoelektrična linija ima različite amplitude iako je primenjen isti filter za filtriranje bazne linije (Batervortov niskopropusnik 3. reda sa filtriranjem u oba smera i frekvencijom odsecanja na 1 Hz). Dakle, relativna amplituda signala (razlika amplituda izoelektrične linije i R pika) nije ista kao apsolutna amplituda signala (amplituda R pika) kod svih ispitanika, te se izoelektrična linija mora detektovati. Ako se ne vodi računa o tome da izoelektrična linija nakon filtriranja nije u bliskoj okolini nule, relativna greška može biti 10% i veća.

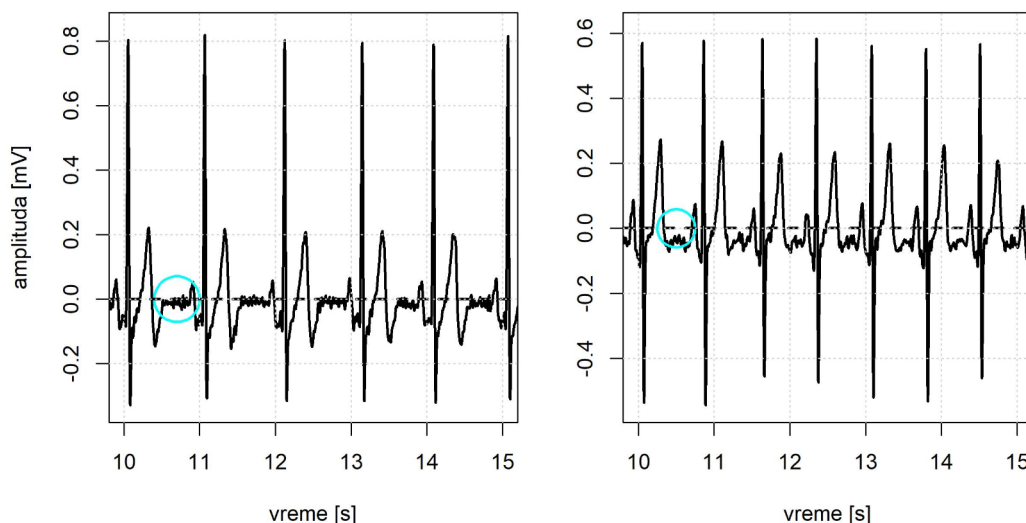


Slika 112, Delineacija EKG signala i odabrana obeležja [355]. Podeljeno pod CC BY 4.0 DEED licencom, <https://creativecommons.org/licenses/by/4.0/>, pristupljeno 12.04.2024.

Za prikaz kružnica na Sl. 113 koje ukazuju na odstupanja u detekciji izoelektrične linije EKG signala, korišćena je funkcija:

`draw.circle()`

iz paketa `plotrix`²⁷⁶ [375] koji sadrži napredne funkcije za prikaz grafika, oznake na osama i skaliranje boja. Kompletna delineacija EKG signala prevazilazi gradivo pokriveno ovim udžbenikom, a čitaoci se upućuju na bazu signala koja sadrži 70 ispitanika i R skripte za delineaciju signala i izdvajanje obeležja koja imaju i koja nemaju klinički značaj [355, 373-374].



Slika 113, Prikaz EKG signala za dva ispitanika (na levom panelu izoelektrična linija je u okolini nule, dok je izoelektrična linija negativna na EKG signalu prikazanom na desnom panelu – korišćene su kružnice da se označe primeri segmenata sa izoelektričnom linijom). Korišćeni su EKG signali mereni na ispitanicima ID7 i ID9 [355, 373-374].

2.3.3 Završna razmatranja o izdvajanju obeležja

Pored detekcije pikova, frakcioni račun može da služi za razdvajanje fizioloških signala – na primer, razdvajanje EKG od EMG signala prilikom merenja EMG signala merenih sa mišića trupa [274, 369, 376]. U slučaju da se EMG signal meri u blizini grudnog koša, dolazi do kontaminacije signala sa EKG signalom i artefaktom disanja. Primer ove kontaminacije je EMG signal meren na mišiću *Pectoralis major*, kao što je to prikazano u ranijim poglavljima. Umesto da se EKG šum filtrira iz EMG signala u frekvencijskom domenu (što i nije moguće u potpunosti realizovati, jer se spektri signala preklapaju), moguće je filtrirati

²⁷⁶ <https://cran.r-project.org/package=plotrix>, pristupljeno 01.06.2024.

EMG signal u vremenskom domenu. Dva najčešće korišćena filtra su MA (eng. *Moving Average*) ili filter medijane. MA filter je linearan i njegovu funkciju prenosa moguće je predstaviti u Z domenu (kao što je pokazano u odeljku koje je bilo namenjeno predobradi signala), dok je medijana filter nelinearan i njegova funkcija prenosa se ne može kvantitativno odrediti. Medijana filter nije do sada spominjan, ali se pokazao veoma korisnim za filtriranje pikova koji potiču od EKG signala. U kombinaciji sa frakcionim izvodima, ovaj filter daje zadovoljavajuće rezultate [274]. Za razdvajanje EKG od EMG signala, što se smatra jednim od najvećih problema u obradi oba signala, moguće je koristiti i analizu nezavisnih komponenti (eng. *Independent Component Analysis*, skraćeno ICA), adaptivne filtre, transformaciju talasića (eng. *Wavelet Transformation*) i druge. U praksi je pokazano da je najuspešnija metoda razdvajanja ova dva signala kada se primene odgovarajuće mere prevencije tokom merenja, te se ili premeste merne elektrode ili izmeni protokol merenja. Međutim, izmenjen protokol ne garantuje da će cilj analize biti postignut, te je razvoj metoda za razdvajanje biosignala od velikog značaja. U idealnom slučaju, bilo bi moguće meriti biosignale jednokanalnim uređajem, a koristiti ih sve u analizi [79, 268]. Takav koncept bi omogućio da se EKG ne smatra šumom na EMG signalu ili obrnuto, već da se oba signala mogu koristiti u daljoj analizi. Ovo važi i za druge biosignale (na primer: merenje EGG i EKG signala, merenje EKG signala trudnice i fetusa).

Obeležja se najčešće izdvajaju u vremenskom (karakterizacija vremenskih i amplitudskih karakteristika) i u frekvencijskom domenu, ali mogu se izdvojiti i u drugim domenima. Na primer, za karakterizaciju EKG signala sa ciljem klasifikacije individualnih karakteristika, korišćen je i geometrijski domen, odnosno predstava signala u faznom prostoru stanja primenom Poinkareovih dijagrama [355]. Za diferencijaciju između EKG signala zdravih ispitanika i pacijenata koji imaju srčanu slabost, mera simetrije, odnosno asimetrije Poinkareovog dijagrama, može se takođe koristiti kao obeležje EKG signala, odnosno obeležje varijabilnosti srčanog ritma [377]. Pored linearnih obeležja, kao što su RMS, medijana frekvencija i broj prolazaka signala ili njegove autokorelacije kroz nulu, postoje i nelinearna obeležja signala. Pretpostavka na kojoj se zasniva korišćenje jeste da su biomedicinski signali, kao što je elektrohisterogram, rezultat nelinearnih dinamičkih bioloških sistema [352]. Primeri nelinearnih obeležja su procena/estimacija maksimalnog Ljapunovog eksponenta, procena fraktalne i korelacione dimenzije, kao i entropija²⁷⁷ uzorka (eng. *sample entropy*). Potrebno je razumeti praktični značaj ovih obeležja da bi

²⁷⁷ Termin entropija se sreće u termodinamici i u teoriji informacija. Za karakterizaciju vremenskih serija koristi se entropija iz teorije informacija ili Šenonova entropija po Klodu Šenonu koji je prvi predstavio ovaj koncept 1948. godine [378].

ona bila primenjena na odgovarajući način. Linearna i nelinearna obeležja se ne koriste samo u obradi biosignala. Entropija se pokazala korisnom u ekonomiji – primenom aproksimativne entropije (eng. *approximate entropy*) nad finansijskim vremenskim serijama je moguće predvideti trendove na berzi tokom kriznih situacija, jer se u tim slučajevima obrasci ponavljaju [379]. Entropija se u obradi biosignala koristi da se proceni predvidljivost vremenskih serija, odnosno u kojoj meri je signal rezultat slučajnog procesa [352]. U analizi biosignala koriste se entropija uzorka i aproksimativna entropija, ali i unakrsna entropija uzorka (eng. *cross-sample entropy*) [380]. Za vremenske serije koje više odgovaraju stohastičkom procesu, entropija je veća u odnosu na vremenske serije koje imaju pretežno determinističke karakteristike. Primeri korišćenja entropije (mere stohastičnosti vremenskih serija) u analizi biosignala su:

- za karakterizaciju signala merenih sa akcelerometara postavljenih na ručni zglob ispitanika tokom korišćenja simulatora vožnje, gde je pokazano da Šenonova entropija predstavlja jedno od najinformativnijih obeležja [381];
- kako bi se ispitalo da li je moguće na osnovu električne aktivnosti materice predvideti da li će doći do prevremenog porođaja, korišćena je entropija uzorka, a rezultati su pokazali da je entropija uzorka zajedno sa medijana frekvencijom dala najrobusnije rezultate i statistički značajne razlike u najvećem broju ispitanih slučajeva [352];
- za potrebe prepoznavanja EGG signala koji su mereni tokom mučnine usled korišćenja simulatora vožnje i kojima je naknadno dodat simuliran šum sa različitim SNR, entropija uzorka i spektralna entropija su se izdvojile u grupi najrobusnijih i najvažnijih parametara za dalju primenu mašinskog učenja i statističke analize [382];
- unakrsna entropija uzorka između RR intervala na EKG signalu i respiratornog signala je korišćena za karakterizaciju stanja pacijenata sa srčanom slabošću, tj. insuficijencijom, a pokazano je da je unakrsna entropija uzorka ukazala na moguće interakcije između RR intervala i respiratornog signala između parametara kod zdravih ispitanika i različitih grupa pacijenata [383].

Na kraju, unakrsna entropija predstavlja samo jedno u nizu spregnutih obeležja koje se može koristiti za ocenu interakcije dva različita biosignala. Nelinearna interakcija između regulatornih mehanizama kardiovaskularnog i kardiorespiratornog sistema su korisna u izdvajanju informacija koje su značajne za razumevanje funkcionisanja zdravog organizma i u razumevanju patoloških pojava. [384]

3 Simulirani biosignali

Generisanje ili simuliranje veštačkih, odnosno sintetičkih biosignala ili test signala može biti korisno za proveru realizovanih metoda. Sati i saradnici su koristili simulirane kontrolne signale da bi se proverilo kakvi rezultati primene predložene metode se dobijaju za vremenske serije sa izraženijom i manje izraženom autokorelacijom [385]. Za simuliranje vremenske serije sa manje izraženom autokorelacijom (na osnovu segmenta signala ne može se ništa pretpostaviti o kompletnom signalu) korišćen je beli šum (signal koji ima konstantan spektar snage i podjednako je zastupljen na svim frekvencijama [386]), dok je za simulaciju vremenske serije sa većom autokorelacijom korišćen Braunov ili crveni šum koji je nastao integraljenjem belog šuma [386]. Na ovaj način je bilo moguće stvoriti pozitivne i negativne kontrolne serije kako bi se metoda proverila. Sličan pristup se primenjuje i u drugim oblastima. Ako se na testira metoda za procenu simetrije slike ili topografske mape, onda je korisno da se kreiraju simetrične i asimetrične simulirane slike, da bi se proverilo da li mera simetrije ili asimetrije daje očekivane rezultate [377].

Korist od simuliranih biosignala je višestruka. Ne samo da omogućava kreiranje pozitivnih i negativnih kontrolnih primera, ona može poslužiti i za uvećavanje broja postojećih vremenskih serija. Da bi se ispitala uspešnost automatskih algoritama za detekciju promena na signalu, potrebno je imati pristup raznovrsnim signalima koji se mogu naći samo u sveobuhvatnim i relativno velikim bazama podataka, što nije uvek izvodljivo. Posebno je problematično naći takve baze, ako biosignali sadrže informaciju koja je posledica retkih bolesti. Ako je potrebno evaluirati algoritam za rad sa različitim srčanim ritmovima, jednostavnije je proizvesti simulirani EKG signal ili simulirane RR intervale na kojima bi se proverila metoda analize signala. Praktično je skoro nemoguće snimiti srčani ritam sa određenom preciznošću u relativno velikom rasponu na ispitanicima, te su simulirani ritmovi i/ili EKG signali od velike koristi. Dodatno, uvećavanje skupova ima značaj u realizaciji algoritama mašinskog učenja, jer je potreban relativno veliki broj elemenata da bi se na odgovarajući način proverila efikasnost ovih algoritama [147, 387]. Ne samo da se takvi signali mogu koristiti u algoritmima mašinskog učenja, već mogu poslužiti da se reši problem neuravnoteženosti klasa u mašinskom učenju (eng. *class disbalance/imbalance*) koji je posebno izražen kod binarnih klasifikatora [388]. Algoritmi mašinskog učenja podešavaju svoje parametre koristeći podatke (proces obučavanja), te ako je potrebno klasifikovati podatke u dve

kategorije: oni koji odgovaraju zdravima i oni koji odgovaraju pacijentima, a pri tom zdravih ispitanika ima u relativno velikom broju, a pacijenata značajno manje, očekivano je da će algoritam bolje predviđati podatke zdravih osoba. Kako bi se sprečila ova situacija, moguće je tehnikama simulacije signala uvećati broj pacijenata u skupu za obučavanje.

U analizi vremenskih serija biosignala, najčešće se koriste simulirani biosignali i simulirani šumovi na signalu, kako bi se ispitao uticaj šuma na rezultat analize, što u realnim situacijama nije uvek moguće izmeriti. Korišćenjem metoda za generisanje sintetičkih biosignala i sintetičkog šuma izvodljivo je kontrolisati odnos signal-šum (SNR) vrlo precizno, te ispitati uticaj šuma na rezultat analize signala. Na primer, ako je potrebno odrediti uspešnost metode filtriranja EMG signala koji je zašumljen EKG signalom (ili obrnuto zavisi šta je cilj merenja ili cilj analize signala²⁷⁸) u zavisnosti od SNR, ne postoji drugi način do da se simulira jedan ili oba signala. Ovakav pristup je veoma koristan, ako je potrebno testirati metodu za izdvajanje korisne informacije tako što će se uporediti filtriran signal sa originalnim signalom koji nije zašumljen, što se ne može izvesti bez prethodnog pristupa originalnom signalu koji je “meren” bez šuma, što se postiže isključivo simulacijom. [274, 392-393]

Dobra praksa nalaže da generisanje sintetičkog signala podrazumeva modeliranje fiziološkog procesa koji za rezultat daje generisan signal. Tako razvijen model doprinosi boljem razumevanju prirode biosignala i fizioloških procesa, pored toga što je koristan u oceni metoda analize signala. Naravno, to nije uvek moguće postići, te metode za generisanje veštačkih biosignala, mogu biti predmet kritike relativno velikog broja istraživača. Razlog tome je što postoje metode koje, iako omogućavaju da se dobiju signali odgovarajućih osobina, ne prate njihove prirodne odlike, te postoji ograničenje u korišćenju takvih signala u analizi. Na primer, najjednostavniji model za simulaciju EEG signala jeste da se EEG signal predstavi kao suma periodičnih talasa (sinusoida) različitih frekvencija kako bi se opisali karakteristični talasni oblici na signalu

278 Prilikom merenja EMG signala sa mišića trupa, može se javiti u većoj ili manjoj meri izražen šum koji potiče od EKG signala (razlog ovome je elektrofiziološko preslušavanje, eng. *crosstalk*). Koliko će signal biti kontaminiran ovim šumom zavisi od udaljenosti od srca, impedanse kože i tkiva, ali i od drugih anatomskih odlika ispitanika. Slično, prilikom merenja EKG signala (za merenje u kliničkim uslovima podrazumeva se ležeći položaj na leđima [389]), aktivacija mišića trupa će izazvati šum na EKG signalu. Oba signala mogu biti korisni, ali oba mogu predstavljati i šum. Primera radi, ako se meri fetalni EKG primenom samolepljivih površinskih Ag/AgCl elektroda, onda se EKG trudnice i EMG koji potiče od mišića trbušnog zida i materice mogu tumačiti kao šum [237]. Postoje situacije u kojima se više signala meri sa jednog kanala, a kasnije se tehnikama obrade signala ti signali razdvajaju, te oba predstavljaju korisnu informaciju [79, 390-391].

[394-395]. Međutim, ovakvi modeli često ne uzimaju u obzir da ne postoji izvor prostoperiodičnog signala u mozgu, te su predloženi realističniji modeli za generisanje izvornih talasa u mozgu [396]. Prostoperiodični napon se može koristiti i za simulaciju šuma napajanja od 50 Hz ili 60 Hz. Jedna od dobrih odlika u simulaciji biosignala i šuma jeste što su šumovi i komponente signala aditivne, te se jednostavno generisanje simuliranih zašumljenih biosignala završava sabiranjem vremenskih serija. Ovo je donekle i očekivano, jer elektrofiziološki signali, kakav je EMG, predstavljaju sumu akcionih potencijala na mestu postavljanja elektrode [318].

Značaj sintetičkih biosignala je u tome što se analiza mnogih fenomena i procena uspešnosti tehnika obrade signala ne bi mogla sprovesti na drugačiji način. Dodatno, značaj generisanih biosignala se ogleda i u povećanoj reproducibilnosti i robusnosti rezultata istraživanja. Rezultat analize je reproducibilan ako se korišćenjem istih podataka i iste tehnike analize može doći do istih rezultata. Robustan rezultat se definiše kao rezultat u kome se korišćenjem istih podataka, a primenom različitog procesa rada (na primer, metoda implementirana u Matlabu i druga metoda implementirana u programskom jeziku R) dolazi do sličnih ili identičnih rezultata. Termin robusnost označava da rezultat ne zavisi, na primer, od programskih jezika ili operativnog sistema na kome se analiza izvršava. Preduslov za reproducibilna i robusna istraživanja jeste da postoji dostupan skup podataka. U slučaju da takav skup nije podeljen na internetu i nije dostupan pod otvorenom licencom, jedini način da se evaluacija izvrši na istom skupu, jeste da se podaci simuliraju. [397]

Ovo poglavlje sadrži tehnike za generisanje pseudoslučajnih brojeva u R-u. Pseudoslučajni brojevi su korisni za razumevanje odabranog modela za generisanje EMG signala. Metode koje su u ovom poglavlju opisane, a neke i praktično realizovane, mogu se primeniti za generisanje drugih biosignala, a primer EMG signala je odabran, jer već postoji CRAN paket specifično namenjen generisanju EMG signala u R-u.

3.1 Generisanje pseudoslučajnih brojeva u R-u

Ugrađena funkcija u R-u *rnorm()* se može koristiti za generisanje pseudoslučajnih²⁷⁹ (eng. *pseudorandom*) brojeva, prema Gausovoj/normalnoj raspodeli. Podrazumevane vrednosti argumenata *rnorm()* funkcije su: (1) *mean*

²⁷⁹ Tačan termin je pseudoslučajan generator brojeva ili pseudoslučajan broj, a ne slučajan broj, jer se na računaru u okviru programskih jezika koriste procedure koje ne mogu da generišu u potpunosti slučajne brojeve.

(srednja vrednost generisanih slučajnih brojeva) koja je jednaka 0 i (2) *sd* (standardna devijacija generisanih slučajnih brojeva) čija je podrazumevana vrednost 1. Za uspešno pokretanje *rnorm()* funkcije, potrebno je uneti obavezan parametar *n* koji predstavlja broj generisanih pseudoslučajnih vrednosti. Za generisanje 10 brojeva sa podrazumevanom srednjom vrednošću i standardnom devijacijom primenom ove funkcije može se pozvati komanda:

```
rnorm(n = 10, mean = 0, sd = 1)
```

Pored prikazanog načina, ova funkcija se može pozvati i na više načina, jer R prepoznaje podudaranje argumenata u funkciji po poziciji, po imenu ili po oba parametra, kako je već pokazano u prethodnim poglavljima. Da bi se u R-u utvrdilo da li je došlo do podudaranja po poziciji ili po imenu ili u kombinaciji (što se nikako ne preporučuje), primenjuje se sledeća procedura:

- proverava se da li postoji potpuno podudaranje za neki od argumenata po imenu (proverava se da li na primeru *rnorm()* funkcije postoji u pozivu funkcije naziv argumenata *n*, *mean* i *sd*),
- proverava se da li postoji delimično podudaranje za neki od argumenata po imenu (na primer, *me*, *mea*, *m* i *s*) i na kraju
- proverava se da li postoji poziciono podudaranje (pozicija argumenata je data po redosledu *n*, *mean*, *sd*).

Za funkcije se može koristiti:

```
args()
```

komanda da se dobije lista svih argumenata ili *str()* da se dobije struktura objekta (uloga *str()* funkcije je već objašnjena u prethodnim poglavljima). Na primeru *rnorm()* funkcije, relativno je jednostavno pozvati funkciju na odgovarajući način, jer postoji samo tri ulazna parametra. To ne mora biti situacija u opštem slučaju, jer funkcije za generisanje pseudoslučajnih brojeva mogu imati relativno veliki broj parametara, pa je preporuka da se koristi podudaranje po imenima, a ne po poziciji [3]. Ako se neki od argumenata postavi na NULL vrednost, u tom slučaju je moguće promeniti ulogu funkcije, ako za tim postoji potreba naravno.

Iako *rnorm()*, i slične funkcije za generisanje pseudoslučajnih brojeva za različite raspodele, imaju za cilj da, u najvećoj mogućoj meri, obezbede slučajnost prilikom ponovljenih poziva funkcije za isti broj ulaznih parametara, u nekim situacijama je poželjno da kod koji se oslanja na generisanje

pseudoslučajnih brojeva daje isti ili sličan rezultat sa ciljem postizanja reproducibilnosti i robusnosti rezultata analize [397]. U biomedicinskom inženjerstvu se *rnorm()* funkcija može direktno koristiti, ako postoji potreba da se generiše sintetički šum sa normalnom raspodelom ili da se ponovi rezultat mašinskog učenja prilikom, na primer, podele podataka na obučavajući i test skup prema psudoslučajnom odabiru elemenata. Da bi se postigla reproducibilnost takvog koda potrebno je koristiti funkciju:

set.seed()

čiji je obavezni ulazni argument celobrojna vrednost. Kada se poziv ove funkcije postavi neposredno pre poziva funkcije za generisanje pseudoslučajnih brojeva, onda će svaki ponovljen poziv nezavisno od R verzije, operativnog sistema ili računara na kome se kod pokreće, poziv *rnorm()* funkcije dati za rezultat istu sekvencu brojeva. Ovo je vrlo važno da bi se obezbedili kontrolisani uslovi za simulaciju pseudoslučajnih brojeva na računaru. Primer korišćenja *set.seed()* funkcije u kombinaciji sa *rnorm()* funkcijom je prikazan na Sl. 114.

Pored toga što se *set.seed()* funkcija može koristiti za testiranja i što omogućava da kod bude reproducibilan i robusan, može biti korisna i za uklanjanje grešaka, tzv. debugovanje od eng. *debugging*. Pored ulaznog parametra koji je ceo broj (Sl. 114), *set.seed()* funkcija ima i druge ulazne parametre²⁸⁰ (*kind*, *normal.kind* i *sample.kind*). Ovi parametri definišu tip pseudoslučajnog generatora, koji je podrazumevano Mersene Twister (eng. *Mersenne Twister*)²⁸¹ [398-399].

Na Sl. 114 se može primetiti da R daje brojeve sa relativno velikim brojem decimala. Jedan od načina da se rezultat ograniči na određen broj decimala jeste da se koristi funkcija:

options()

koja omogućava podešavanje niza parametara u programskom jeziku R, a jedan od parametara je i broj decimala za prikaz u konzoli. Na Sl. 115 je prikazano kako se *options()* funkcija može koristiti za ograničen broj decimala *summary()*

280 Za prikaz ulaznih argumenata i primere korišćenja svih funkcija u R-u, može se koristiti *Help* stranica sa uputstvima, primerima i referencama.

281 Izvorni kod je dostupan na <http://www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/VERSIONS/C-LANG/c-lang.html>, pristupljeno 01.06.2024.

funkcije. Za podatke koji imaju normalnu funkciju gustine verovatnoće, medijana i srednja vrednost su jednake (na Sl. 115 bi trebalo da obe budu 0), ali to nije slučaj, jer je relativno mali broj pseudoslučajnih brojeva generisan (10).

```
> rnorm(5)
[1] -1.483552013  0.430408603  0.007923051  0.885742737 -0.409657598
> rnorm(5)
[1]  0.06094591 -1.16648312 -0.01304479  1.19594189 -0.02123129
> set.seed(38); rnorm(5)
[1] -0.2535911 -1.0556027  0.6864966  0.0251569 -1.6718410
> set.seed(38); rnorm(5)
[1] -0.2535911 -1.0556027  0.6864966  0.0251569 -1.6718410
> rnorm(5)
[1] -1.483552013  0.430408603  0.007923051  0.885742737 -0.409657598
> set.seed(45); rnorm(5)
[1]  0.3407997 -0.7033403 -0.3795377 -0.7460474 -0.8981073
> set.seed(45); rnorm(5)
[1]  0.3407997 -0.7033403 -0.3795377 -0.7460474 -0.8981073
```

Slika 114, Primer generisanja pseudoslučajnih brojeva sa normalnom raspodelom korišćenjem *set.seed()* funkcije. Preporuka je studentima da na svojim računarima pokrenu isti kod i da uporede rezultat. Primititi da je radi jednostavnijeg zapisa većeg broja komandi u jednom redu korišćen operator dvotačka “;”, kao i da se svakoj *rnorm()* funkciji mora pridružiti jedna *set.seed()* funkcija.

```
> y <- rnorm(10); y
[1]  0.8141  0.6344 -0.0063 -1.3791 -0.5215  1.5286  0.9550 -0.1791
[9] -0.3441  0.2597
> summary(y)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-1.38  -0.30   0.13   0.18   0.77   1.53
> options(digits = 2)
> summary(y)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-1.38  -0.30   0.13   0.18   0.77   1.53
```

Slika 115, Korišćenje *options()* funkcije za ograničenje u prikazu broja decimala osnovnih parametara deskriptivne statistike (primenom *summary()* funkcije koja može biti korisna za proveru sadržaja numeričkih promenljivih) za 10 generisanih pseudoslučajnih brojeva koji imaju Gausovu funkciju gustine verovatnoće.

Osim *rnorm()* funkcije za generisanje pseudoslučajnih brojeva sa normalnom raspodelom, u upotrebi su *dnorm()* za generisanje funkcije gustine verovatnoće²⁸² (*d* od eng. *density*), *pnorm()* za generisanje funkcije raspodele verovatnoće (eng. *cumulative distribution*, *p* od eng. *probability*) i *qnorm()* za

282 Za nazive funkcija u R-u korišćene su skraćenice, a prevod na engleski funkcije gustine verovatnoće (fgv) je *probability density function* (skraćeno pdf).

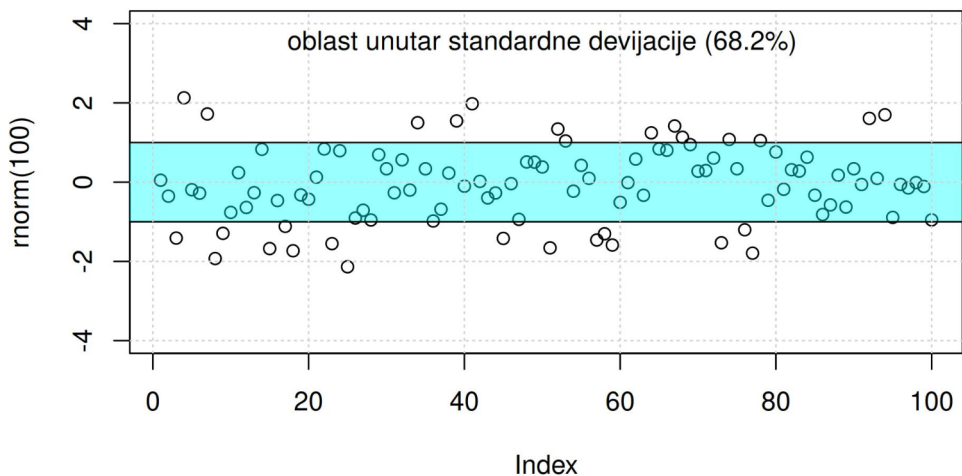
računanje kvantila (q od eng. *quantile*). Na Sl. 116 je prikazan kod koji generiše dva grafika – jedan predstavlja prikaz generisanih 100 brojeva koji se uglavnom gomilaju oko oblasti standardne devijacije (Sl. 117), dok drugi predstavlja teorijsku funkciju gustine verovatnoće za srednju vrednost 0 i standardnu devijaciju 1 nastalu primenom *dnorm()* funkcije na odabranom intervalu [-4, 4] (Sl. 118).

```
> plot(rnorm(100), ylim = c(-4, 4))
> polygon(c(-4, 104, 104, -4), c(1, 1, -1, -1),
+        col = rgb(0, 1, 1, alpha = 0.4))
> grid()
> text(50, 3.5, "oblast unutar standardne devijacije (68.2%)")
> plot(seq(-4, 4, by = 0.1),
+      dnorm(seq(-4, 4, by = 0.1), mean = 0, sd = 1), type = "l",
+      lwd = 2)
> grid()
> polygon(c(-1, 1, 1, -1), c(0.42, 0.42, -0.1, -0.1),
+        col = rgb(0, 1, 1, alpha = 0.4))
> text(0, 0.2, "68.2%")
```

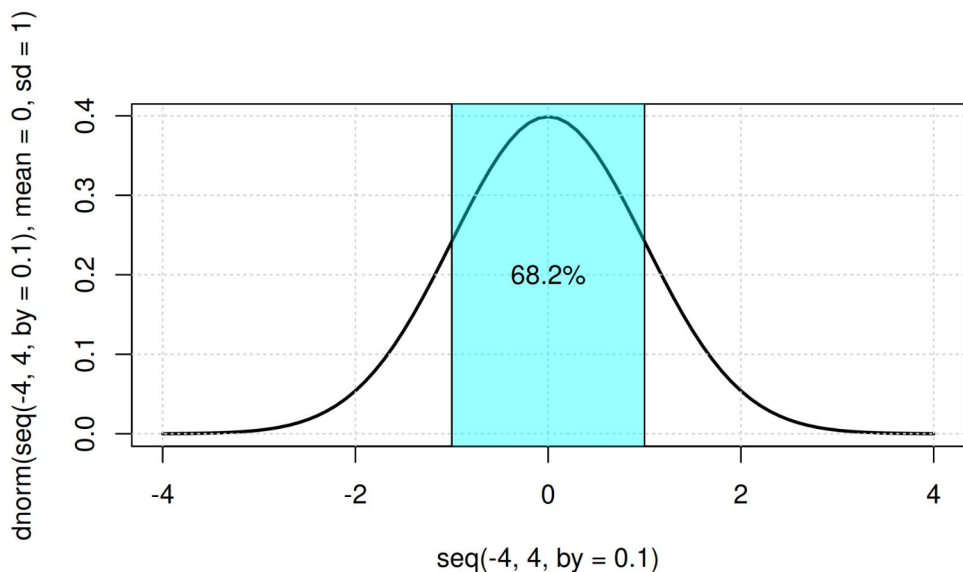
Slika 116, R kod za prikaz grafika datim na Sl. 117 i na Sl. 118 korišćena je funkcija *polygon()* za isticanje oblasti na grafiku (granice funkcije su odabrane tako da se prostiru preko celog grafika), za ispis teksta korišćena je funkcija *text()*, a za prikaz transparentne boje korišćena je *rgb()* funkcija sa parametrom *alpha* koji je normalizovan na vrednosti u opsegu od 0 do 1.

Pored generisanja pseudoslučajnih brojeva i šumova, nekada je od značaja pseudoslučajno odabiranje pojedinačnih elemenata iz postojećeg skupa. Za te potrebe može se koristiti *sample()* funkcija koja se primenjuje na bilo koji objekat čiji broj elemenata n ispunjava uslov $n > 1$. Za upravljanje odabiranjem, tj. da li jedna vrednost može više puta da bude odabrana, postoji argument *replace* čija je podrazumevana vrednost FALSE. Funkcija *set.seed()* se može koristiti u kombinaciji sa *sample()* funkcijom za reproducibilno odabiranje. a u okviru *sample()* funkcije se može odabrati način na koji će elementi biti pododabirani (ova funkcionalnost je dostupna za R verzije od 3.6). Primer primene je slučajna podela ispitanika na više grupa, kao što je randomizacija u kliničkim studijama. Randomizovana kontrolisana klinička ispitivanja predstavljaju zlatni standard u evaluaciji novih metoda i najrigorozniji su naučni pristup za ocenu efikasnosti terapije ili intervencije u kliničkom okruženju [400]. Randomizacija se odnosi na slučajnu podelu ispitanika, dok su ispitivanja kontrolisana jer je potrebno obezbediti kontrolu svih faktora (osim onog koji se ispituje) u svim grupama ispitanika [401]²⁸³.

283 O istoriji i značaju kliničkih studija, studenti se upućuju na knjigu [114].



Slika 117, Prikaz 100 generisanih pseudoslučajnih brojeva koji imaju Gausovu funkciju gustine verovatnoće sa označenim intervalom koji je obuhvaćen standardnom devijacijom za kod na Sl. 116.



Slika 118, Prikaz korišćenja *dnorm()* funkcije za predstavljanje normalne funkcije gustine verovatnoće u opsegu [-4, 4] za kod na Sl. 116.

Za razliku od *dnorm()* funkcije čiji je način upotrebe prikazan na Sl. 116 koja daje funkciju gustine verovatnoće, *pnorm()* funkcija se može koristiti da se generiše funkcija raspodele verovatnoće ili konkretna vrednost unutar funkcije. Primer korišćenja *pnorm()* i *qnorm()* funkcije je prikazan na Sl. 119.


```

> pnorm(mean = 0, sd = 1, 0.5)
[1] 0.69
> qnorm(0.69, mean = 0, sd = 1)
[1] 0.5
> pnorm(mean = 0, sd = 1, 1)
[1] 0.84
> pnorm(mean = 0, sd = 1, 4)
[1] 1
> pnorm(mean = 0, sd = 1, 400)
[1] 1

```

Slika 119, Primer korišćenja *pnorm()* i *qnorm()* funkcija.

Za vrednost 0.5 unutar populacije koja ima Gausovu²⁸⁴ raspodelu sa srednjom vrednošću 0 i standardnom devijacijom 1, postoji verovatnoća od 0.69 (69%) da će se taj broj naći u opsegu $[-\infty, 0.5]$ (pogledati rezultat *pnorm()* funkcije na Sl. 119). Ako se pitanje postavi na drugačiji način, odnosno u kom opsegu postoji verovatnoća od 0.69, odgovor se dobija primenom *qnorm()* funkcije i to je 0.5 kao što je prikazano na Sl. 119, odnosno opseg $[-\infty, 0.5]$. Na Sl. 119 se vidi da se određen broj vrednosti nalazi izvan opsega koji je obuhvaćen standardnom devijacijom, te je verovatnoća da će se pseudoslučajni broj naći u opsegu $[-\infty, 1]$ jednaka 0.84, dok je verovatnoća 100% (ili 1) da će se broj naći u opsezima $[-\infty, 4]$ i $[-\infty, 400]$ (Sl. 119). Za druge raspodele moguće je koristiti slične funkcije. Za binomijalnu raspodelu: *dbinom()*, *pbinom()*, *qbinom()* i *rbinom()*.

3.1.1 Generisanje belog i obojenog šuma

Gausovom raspodelom u biomedicinskom inženjerstvu moguće je simulirati beli ili obojeni šum. Beli šum sadrži podjednaku zastupljenost svih frekvencija u frekvencijskom opsegu [386]. Naravno, nije moguće očekivati pojavu svih frekvencija na digitalnom belom šumu, jer digitalni signal ima konačnu periodu odabiranja, te je propusni opseg određen vremenskom rezolucijom. Na primer, ako se generiše beli šum za vremensku osu na kojoj je rezolucija, tj. perioda odabiranja 1 ms, to je frekvencijski sadržaj takvog signala do maksimalno $f = (1 / 1 \text{ ms}) / 2 = 500 \text{ Hz}$ (odb/s). Beli šum koji ima Gausovu raspodelu se može generisati primenom *rnorm()* funkcije.

²⁸⁴ Formalno, postoje funkcija gustine verovatnoće (čiji oblik može odgovarati Gausovoj ili normalnoj, binomijalnoj, uniformnoj ili drugim raspodelama), a postoji i funkcija raspodele verovatnoće koja je monotono neopadajuća odnosno kumulativna (eng. *cumulative probability distribution*, skraćeno cpd) [78]. Uobičajeno se kaže da podaci imaju Gausovu raspodelu, ali se misli na to da funkcija gustine verovatnoće ima oblik Gausovog zvona.

```

dat <- read.table("puls.txt")
fs <- 1000

ecg <- dat$V1[1:(10*fs)]
set.seed(14)
bs <- 0.11 * rnorm(10*fs)

ecg_sum <- ecg + bs
vreme <- seq(0, length(ecg)/fs - 1/fs, by = 1/fs)

plot(vreme, ecg_sum, type = "l", xlab = "vreme [s]",
      ylab = "amplituda [mV]", xlim = c(0, 10), ylim = c(-0.7, 0.7),
      col = "cyan")
par(new = TRUE)
plot(vreme, ecg, type = "l", xlab = "vreme [s]",
      ylab = "amplituda [mV]", xlim = c(0, 10), ylim = c(-0.7, 0.7),
      col = "black")
grid()
legend(7.6, 0.7, legend = c("EKG sa belim šumom", "EKG"),
       fill = c("cyan", "black"))

```

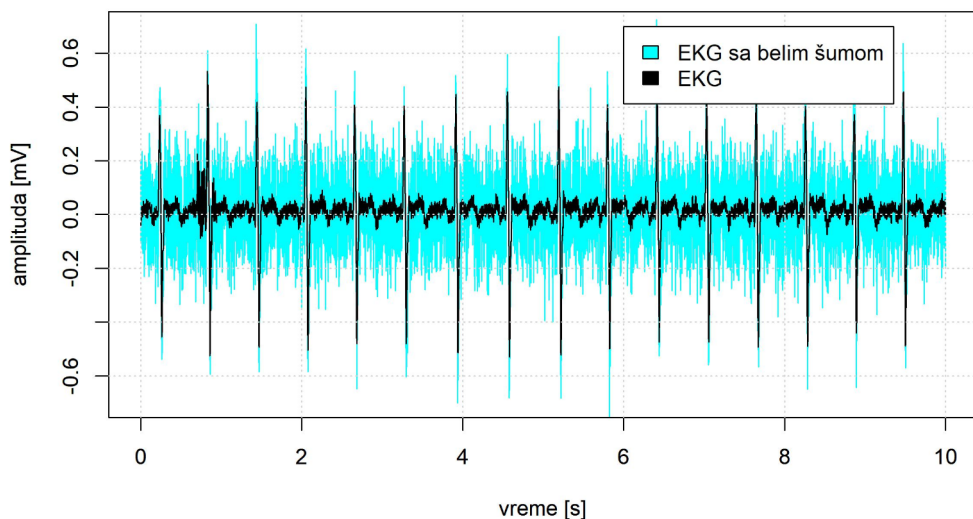
Slika 120, Programski kod koji omogućava generisanje belog šuma *bs* u trajanju od 10 s (sa *fs* je označena frekvencija odabiranja) koji je dodat EKG signalu (promenljiva *ecg*) kako bi se dobio semi-sintetički signal *ecg_sum*. Podaci su javno dostupni (pogledati fusnote i tekst za detalje).

Na Sl. 120 je prikazano kako je moguće generisati beli šum primenom *rnorm()* funkcije. Takođe, pokazano je kako izgleda kada se šum doda snimljenom EKG signalu na Sl. 121. Za generisanje sintetičkog aditivnog šuma sa slike i semi-sintetičkog zašumljenog EKG signala korišćena je funkcija *set.seed()*, te svako može pokretanjem koda sa Sl. 120 da generiše željene grafike. Šum koji je prikazan na Sl. 121 je aditivni, a EKG signal je preuzet sa sajta TOBS predmeta²⁸⁵ i dostupan je u GitHub repozitorijumu²⁸⁶ u kome je podeljen R kod sa realizovanim procedurama i metodama prikazanim u ovom udžbeniku. Signali u datoteci “puls.txt” (Sl. 120) su mereni sa frekvencijom odabiranja od 1000 Ss i pojačanjem od 1000 puta (za EKG signal²⁸⁷) uz poštovanje svih međunarodnih i etičkih principa. Važno je napomenuti da, iako je korišćen elektrofiziološki pojačavač, meren signal u datoteci “puls.txt” nije EKG u pravom smislu te reči, jer su elektrode postavljene u blizini srca, a ne prema standardnim odvodima i Ajthovenovom trouglu.

285 https://automatika.etf.bg.ac.rs/images/FAJLOVI_srpski/predmeti/izborni_kursevi_os/biomedicinsko_inzenjerstvo/TOBS/vezbe/puls.txt, pristupljeno 12.02.2024.

286 <https://github.com/NadicaSm/TOBS>

287 Meren je takođe i PPG signal koji se nalazi na drugom kanalu, ali taj signal je bez jedinica, odnosno u datoteci “puls.txt” se nalazi signal direktno meren sa AD konvertora koji prethodno nije pojačan.

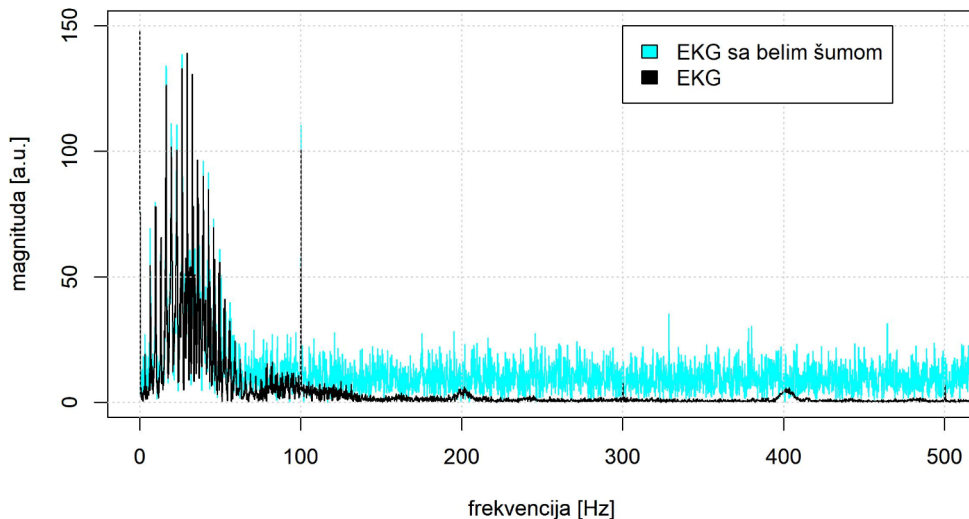


Slika 121, EKG signal sa dodatim belim šumom (u cijan boji) i EKG signal koji je meren na ispitaniku. Kod koji je korišćen za prikaz ove slike prikazan je na Sl. 120.

Sa Sl. 121 se vidi da postoji uticaj belog šuma na izgled EKG signala u vremenskom domenu, ali da su svi R pikovi na signalu vidljivi. Ova razlika se može pogledati i u frekvencijskom domenu kako je prikazano na Sl. 122. Uočljiva je nepromenljiva energija na delu signala koji potiče od belog šuma (za frekvencije veće od 100 Hz) koja odgovara belom šumu po definiciji²⁸⁸. Dodatno, EKG signal koji je ovde prikazan nije dodatno filtriran, te se na njemu uočava šum napajanja, čiji je harmonik na 100 Hz posebno izražen. Takođe, uočljiva je i relativno velika DC komponenta (Sl. 122).

Obojeni šum se može dobiti, ako se generisani šum filtrira u željenom opsegu. Iako se ne mogu direktno povezati sa specifičnim šumom tokom merenja, beli i obojeni šumovi se mogu koristiti i generalno da predstavljaju bilo koji šum, a koji ima dovoljno veliki propusni opseg, kao što je artefakt pokreta i EMG šum na EKG signalu [402]. Da bi se što više dočarala realna situacija, beli šum se, osim bojenja, može podvrgnuti i amplitudskoj modulaciji, kao što je pokazano u [393]. Za simulaciju artefakta pokreta, napredne metode uključuju autoregresioni model (eng. *Autoregressive Model*), Markovljeve modele (eng. *Markov chain*) i rekurentne neuralne mreže (eng. *Recurrent Neural Network*) [403].

²⁸⁸ Spektar belog šuma se može koristiti i za eksperimentalno merenje propusnog opsega elektrofizioloških pojačavača [104].



Slika 122, Prikaz Furijeove transformacije merenog EKG signala (crna linija) i semi-sintetičkog EKG signala kome je dodat beli šum (cijan boja). Prikaz ovih signala u vremenskom domenu dat je na Sl. 121.

Poseban tip obojenog šuma predstavlja roze ili pink šum (fraktalni ili frakcioni šum) kod koga je spektar snage signala obrnuto proporcionalan frekvenciji signala [404]. Pink šum se koristi za opis bioloških sistema [405], a posebno u obradi EEG signala [406]. U R-u postoji CRAN paket `tuneR` u kome se funkcija `noise()` može koristiti za generisanje, između ostalih, i pink šuma [407]. Na Sl. 123 je pokazano kako se može primenom funkcije:

`noise()`

iz `tuneR` paketa generisati pink šum. Rezultat koji se dobija primenom ove funkcije je klase `Wave`, a na Sl. 123 je prikazano koji su elementi te klase kojima se može pristupiti primenom operatora `@`. Frekvencija odabiranja generisanog šuma je podrazumevano 44100 Hz, što je i očekivano jer ovaj paket je primarno namenjen obradi audio signala. Kada se prikaže Furijeova transformacija signala dobija se prikaz kao na donjem panelu na Sl. 124 (primetiti da magnituda pink šuma opada sa porastom frekvencije).

U R-u se za prikaz vremenskih serija (generisani i mereni biosignali i šumovi) mogu koristiti i specijalizovane funkcije, objekti i klase, ali njima nije posvećena posebna pažnja u ovom udžbeniku. Ova funkcionalnost je nasleđena iz programskog jezika S [408]. Ukratko, moguće je u okviru jednog specijalizovanog objekta skladištiti podatke, vreme početne opservacije, vreme

poslednje opservacije i frekvenciju odabiranja. Za više informacija, može se pogledati R dokumentacija za najčešće korišćene funkcije za rad sa ovim objektima kao što su *ts()* i *print.ts()* za kreiranje i štampanje/formatiranje objekata iz klase vremenskih serija.

```
> pinksum <- noise(kind = c("pink"))
> pinksum

Wave object
  Number of Samples:      44100
  Duration (seconds):     1
  Samplingrate (Hertz):  44100
  Channels (Mono/Stereo): Mono
  PCM (integer format):  FALSE
  Bit (8/16/24/32/64):   32

> fs <- pinksum@samp.rate
> vreme <- seq(0, length(pinksum@left)/fs - 1/fs, by = 1/fs)
> pinksumFFT <- Mod(fft(pinksum@left))
> pinksumFFT <- pinksumFFT[1:length(pinksumFFT)/2]
> fosa <- 1:length(pinksumFFT) / (length(pinksum@left)/fs)
> # prikaz pink šuma sa Furijeovom transformacijom na grafiku
> par(mfrow = c(2,1))
> plot(vreme, pinksum@left, type = "l", col = "pink",
+      ylab = "amplituda [a.u.]", xlab = "vreme [s]")
> grid()
> plot(fosa, pinksumFFT, type = "l", xlab = "frekvencija [Hz]",
+      ylab = "magnituda [a.u.]", col = "pink")
> grid()
```

Slika 123, Primena *noise()* funkcije iz *tuneR* paketa za generisanje pink šuma. Pokazani su elementi *Wave* klase, kao i način računanja i prikaza Furijeove transformacije pink šuma.

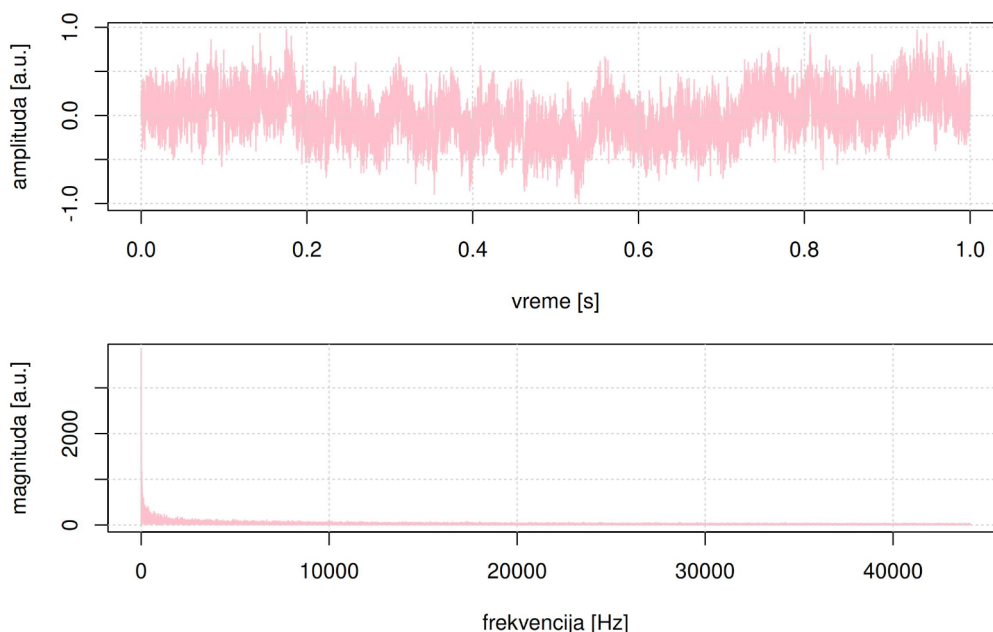
3.2 Generisanje simuliranih EMG signala

Na CRAN-u je od 2015. godine dostupan *biosignalEMG* paket za simulaciju EMG signala [409-410]. Pored funkcija za generisanje EMG signala, čiji rezultat pripada posebnoj *emg* klasi u R-u, ovaj paket uključuje funkcije za obradu EMG signala, ali i EMG/ENG signala merenih na životinjama. Funkcije za obradu signala se mogu koristiti za ispravljanje, integraljenje i usrednjavanje EMG signala, za automatsku detekciju aktivacije mišića iz EMG signala i prikaz rezultata. Primeri funkcija za obradu EMG signala uključuju:

- *rectification()* – omogućava operaciju dvostranog ili jednostranog ispravljanja EMG signala²⁸⁹,

289 Ovde se radi o analogiji sa jednostranim i dvostranim ispravljanjem signala prilikom konstrukcije analognih mernih instrumenata [78]. Hardverska realizacija dvostranog ispravljanja (koja odgovara apsolutnoj vrednosti EMG signala) je bila često korišćena u

- *integration()* – vrši za integraljenje EMG signala koje se može koristiti za računanje obeležja ili za obvojnice,
- *envelope()* – generiše obvojniciu EMG signala primenom MA (eng. *Moving Average*) filtra i
- *eaemg()* – omogućava sinhrono usrednjavanje (eng. *ensemble average*) EMG signala, a ima namenu u sinhronom usrednjavanju voljnih kontrakcija bez prethodne predobrade, jer sintetički EMG signal sadrži tačnu informaciju o aktivaciji u obliku niza logičkih vrednost.



Slika 124, Gornji panel – prikaz generisanog pink šuma u vremenskom domenu. Donji panel – prikaz generisanog pink šuma u frekvencijskom domenu. Frekvencija odabiranja generisanog pink šuma je 44100 Hz (pogledati kod na Sl. 123).

Paket `biosignalEMG` sadrži i druge funkcije koje se koriste, na primer, za filtriranje i određivanje početka i kraja mišićne kontrakcije. Sve ove funkcije se mogu realizovati korišćenjem signal paketa i drugih funkcija/paketa u R-u, ali specifičnost ovog `biosignalEMG` metoda se ogleda u tome što omogućava generisanje EMG signala, kao i što sadrži merene podatke. Podeljeni podaci su mereni na mački (sa povredom nervnog sistema) sa tri mišića (EMG) i na kičmenoj moždini (ENG). Deljenje podataka koji su mereni na životinjama je

uređajima za merenje EMG signala kao pretkorak za računanje obvojnice EMG signala 80ih godina prošlog veka. Razlog tome je što je bilo jednostavnije sačuvati u memoriji uređaja ili računara obvojniciu za koju je potrebno manje memorijskog prostora, jer ima manju frekvenciju odabiranja od tzv. sirovog (eng. *raw*) EMG signala.

etički, jer se sve životinje moraju žrtvovati nakon eksperimenata²⁹⁰, te postoji veliki značaj i upotrebna vrednost ovog paketa. U narednom potpoglavlju će biti pokazan model koji koristi biosignalEMG paket za generisanje EMG signala. Da bi se razumeo način na koji je realizovan model, potrebno je razumeti osobine EMG signala.

3.2.1 Osobine EMG signala

Kada je rezultat voljne kontrakcije, EMG signal je nestacionaran i stohastički signal. U relativno kratkim vremenskim intervalima se može smatrati stacionarnim signalom (u intervalima do oko 1500²⁹¹ ms). Srednja vrednost EMG signala je jednaka nuli, a signal ima Gausovu funkciju gustine verovatnoće, ali to važi samo za stacionarne EMG signale, tj. za EMG signale u relativno kratkim vremenskim intervalima za vreme kojih su osobine signala konstantne. Opšte je prihvaćeno da je EMG signal stohastički signal. Međutim, to zavisi od intervala i porekla signala koji se posmatra. Na primer, ako se posmatra aktivnost mišića *Tibialis Anterior* (lat. TA) prilikom hoda, onda se može primetiti periodična ponovljivost koja odgovara brzini hoda²⁹². Ta periodičnost postaje još uočljivija ako se prikazuje obvojnica EMG signala. Sa druge strane, ako se pogleda aktivnost mišića tokom drugih pokreta koji nisu periodični ili ponovljivi ili se prikaže signal u kraćim vremenskim intervalima (tokom jedne aktivacije TA mišića kao na Sl. 125) tada se jasno uočava stohastička priroda signala. [104, 411-412]

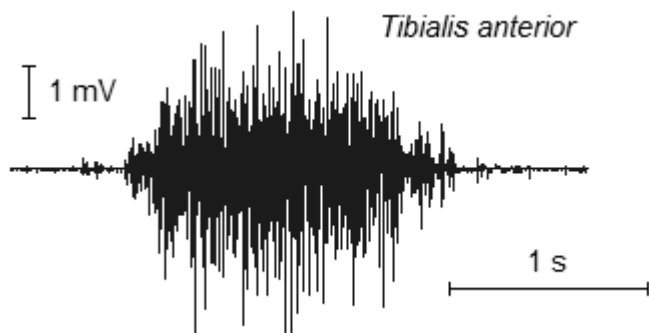
U prethodnim primerima, pokazan je samo EMG signal tokom voljne mišićne kontrakcije, a postoji i EMG signal snimljen kao evocirani potencijal (Hofmanov ili H-refleks i M talas) u odgovoru na stimulaciju nerva koji predstavlja deterministički signal. Usled razlike u prirodi signala, za analizu evociranih EMG potencijala, koriste se drugačije metode analize. Na primer, često se za filtriranje evociranih potencijala (bilo EMG ili EEG signala) koristi sinhrono usrednjavanje, odnosno usrednjavanje po ansamblu (eng. *ensemble average*), koje se veoma retko primenjuje u analizi EMG signala koji su rezultat voljne kontrakcije. Jedini slučaj kada se primenjuje sinhrono usrednjavanje u analizi takvih signala jeste kada je potrebno predstaviti usrednjene obvojnice

290 Postoje biomedicinski inženjeri koji pojam eksperiment smatraju neprimerenim kada se vrše testiranja na ljudima [lična komunikacija], ali odobravaju primenu ovog termina u merenjima ili testiranjima na životinjama. U drugim oblastima, kao što je psihologija, je eksperiment vrlo uobičajen termin u radu sa ljudima.

291 Interval na kome je EMG signal stacionaran varira u literaturi, jer u velikoj meri zavisi od toga kako je EMG signal meren i na kom mišiću je meren.

292 Pretpostavlja se da je hod relativno konstantne brzine.

EMG signala merene tokom ponovljenih pokreta kao što su ciklične aktivacije mišića tokom hoda. [105]



Slika 125, Primer voljne mišićne kontrakcije za jednu kontrakciju na mišiću *Tibialis anterior*. Slika je preuzeta iz prezentacije za [104]. Sadržaj u [104] je podeljen pod otvorenom licencom.

Ilustracija metode sinhronog usrednjavanja je prikazana na Sl. 126. Za signale čiji su odbirci $a_n, b_n, c_n \dots z_n$ mogu se odrediti srednje vrednosti po odbircima redom $a_1, a_2 \dots a_n$, potom po $b_1, b_2 \dots b_n$, sve do $z_1, z_2 \dots z_n$ i dobiti rezultujući filtriran signal ($\Sigma a_i, \Sigma b_i, \Sigma c_i \dots \Sigma z_i$) koji je prikazan na donjem desnom panelu na Sl. 126. Preduslovi za primenu ove metode su slični, kao i preduslovi za primenu MA filtra, a filtriranje se primenjuje u vremenskom domenu, odgovara filtru niskopropusniku i najvažnije: efikasno je za primenu u slučaju kada se frekvencijski opseg šuma i korisnog signala poklapaju. [413]

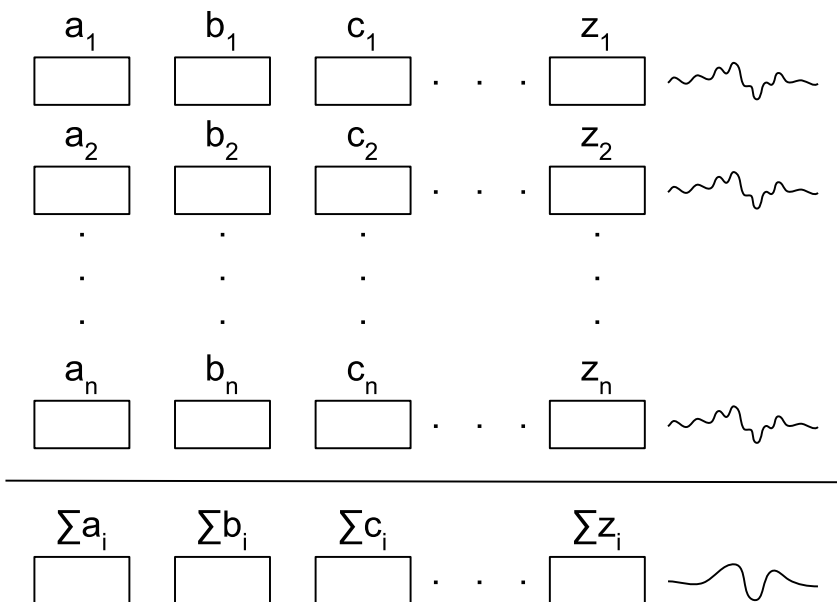
Sinhrono usrednjavanje je jednostavno izvesti, ako postoji jasno označen početak signala (u slučaju evociranih potencijala to je trenutak kada se isporučuje stimulus. Za merenje H-refleksa to je početak električne stimulacije nerva, za EMG signal meren tokom hoda, to može biti kontakt pete sa podlogom²⁹³) i ako se zna trajanje signala²⁹⁴, odnosno ako za svaki signal postoji isti broj odbiraka ($a, b, c \dots z$ na Sl. 126). Međutim, to ne važi za hod. Iako je hod automatizovana radnja, ne može svaki korak da traje podjednako, te je potrebno normalizovati signale u vremenu [105]. Da bi se to uradilo, signal²⁹⁵ se

293 Primititi da je za sinhrono usrednjavanje potrebno koristiti dodatne senzore ili sinhronizaciju. Ovo je neophodno da bi se znalo u kom trenutku je započeta električna stimulacija. Da bi se dobila informacija u kom trenutku je peta stopala ostvarila kontakt sa podlogom, potrebno je koristiti senzore sile, najčešće kapacitivne ili rezistivne. [104]

294 Na primer, H-refleks i M talas uvek isto traju i javljaju se u istim trenucima na jednom ispitaniku. Njihovo kašnjenje u odnosu na stimulacioni impuls zavisi od anatomskih karakteristika (visine ispitanika koja je povezana sa dužinom nerva) i od provođenja nerva (ovde se najveće razlike uočavaju za zdrave ispitanike i pacijente).

295 Nije uobičajeno i nema smisla da se sinhrono usrednjavanje radi na sirovim EMG signalima, već se radi na obvojnica EMG signala.

u vremenu deli na segmente. Najčešće se uzima od 2% do 5% od ukupnog trajanja za segmente na kojima se radi usrednjavanje u vremenskom domenu bez preklapanja prozora [414]. Na taj način se postiže da, signali koji se usrednjavaju, imaju isto trajanje (isti broj odbiraka) što za posledicu ima smanjenu rezoluciju signala u vremenskom domenu.



Slika 126, Realizacija sinhronog usrednjavanja. Na desnom panelu su prikazane skice signala koje su za potrebe prikaza u ovom udžbeniku nacrtane rukom, odnosno predstavljaju isključivo ilustraciju realnih signala. Sa pravougaonicima su označeni pojedinačni odbirci signala, dok su signali mereni tokom različitih pokreta ili realizacija evociranih potencijala dati u vrstama.

Pre nego što se primeni sinhrono usrednjavanje na obvojnici EMG signala, koji je rezultat voljne kontrakcije mišića, karakteristike EMG signala se mogu posmatrati i izdvojiti na filtriranom EMG signalu. Svaki EMG signal koji je rezultat voljne kontrakcije ima dva karakteristična segmenta – jedan koji potiče od relaksacije mišića (obuhvata spontanu elektrofiziološku aktivnost) i drugi koji potiče od kontrakcije (predstavlja zbir akcionih potencijala koje generišu motorne jedinice). Kada se prikažu funkcije gustine verovatnoće na ovim segmentima dobija se oblik Gausovog zvona. [410]

Praktična ispitivanja raspodele EMG odbiraka, pokazala su da EMG signali, pored Gausove raspodele, mogu imati Laplasovu raspodelu. Takođe, pokazano je da raspodela odbiraka EMG signala zavisi od SNR-a (za veći SNR raspodela

više “liči” na Gausovu) i tipa šuma koji postoji na signalu (u slučaju EKG šuma na EMG signalu, dobila se Laplasova raspodela), te da oblik raspodele može da ukaže na postojanje šuma, kao i na tip šuma na EMG signalu. U slučaju da je EMG signal meren sa relativno velikim SNR i da je šum prethodno filtriran, odbirci EMG signala imaju Gausovu raspodelu, a osobine Gausove raspodele (oblik zvona) su najizraženije za EMG signal meren sa ekstenzora²⁹⁶ i za EMG signal tokom kontrakcije relativno velikog intenziteta. [411]

Mnogi realni signali nemaju Gausovu raspodelu. Razlog tome je, pored prirode signala, ograničen broj odbiraka kojima se raspolaže prilikom analize digitalizovanih signala, jer takvi signali ne ispunjavaju Centralnu graničnu teorem (CGT) prema kojoj slučajne promenljive konvergiraju ka Gausovoj raspodeli [415]. Gausova raspodela odbiraka signala je poželjna zbog toga što konvolucija, proizvod i suma dve Gausove raspodele i dalje imaju Gausovu raspodelu. Dodatno, poželjna osobina je i što se najvažniji parametri raspodele jednostavno određuju (srednja vrednost i standardna devijacija²⁹⁷). Da bi se procenilo da li EMG signal ima (i pod kojim uslovima) ili nema Gausovu raspodelu, mogu se koristiti grafičke metode (na primer: histogram, QQ (eng. *Quantile Quaantile*) grafik, prikaz funkcije gustine verovatnoće) i statističke, tj. kvantitativne metode.

Ako se koristi histogram za procenu da li raspodela signala odgovara Gausovoj raspodeli, onda je obavezan parametar (ulazni argument odgovarajuće funkcije) koji određuje broj opsega/binova za prikaz histograma. Ovaj parametar se bira prema broju odbiraka koji se prikazuju (pogledati detalje u [78]). Ako je broj intervala manji, histogram će grubo prikazati funkciju gustine raspodele verovatnoće (u najgorem slučaju za jedan interval, fgv će odgovarati uniformnoj raspodeli). U slučaju da je broj intervala veći, može se desiti da neki od intervala budu prazni. Ako je broj odbiraka koji se posmatra označen sa n onda je broj intervala b dat u Relaciji (23) [78].

$$b = \sqrt{n} + 1$$

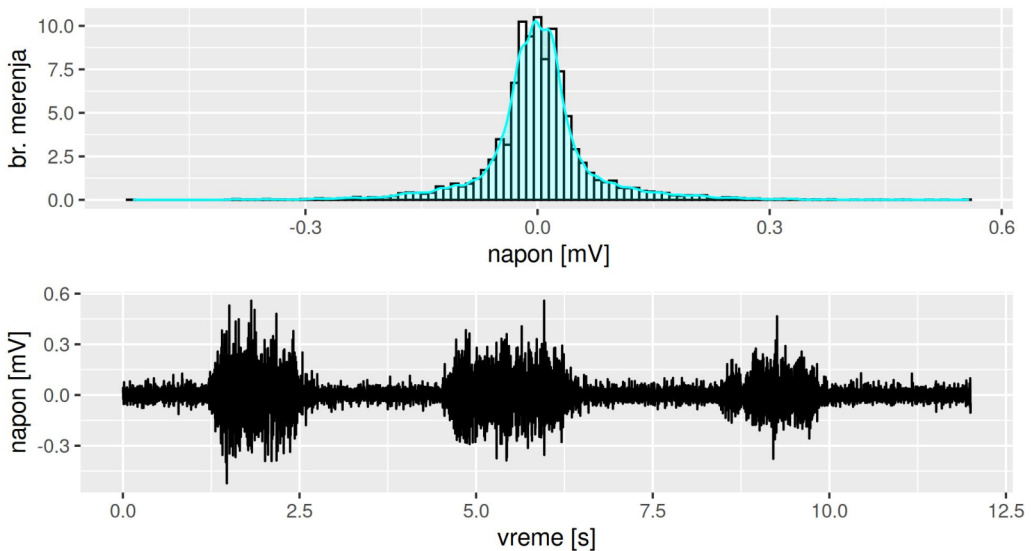
Relacija (23)

Histogram je koristan, jer se sa njega može dobiti informacija o obliku funkcije gustine verovatnoće testiranih odbiraka, tj. da se proveriti da li histogram ima oblik Gausovog zvona. EMG signal meren u dužim vremenskim intervalima

²⁹⁶ Ekstenzor je mišić koji ima ulogu u povećanju ugla u zglobu, <https://www.britannica.com/science/extensor-muscle> (pristupljeno 01.06.2024).

²⁹⁷ Iako se u praksi često računa standardna devijacija za bilo koji skup odbiraka ili numeričkih vrednosti, to bi trebalo raditi isključivo ako vrednosti za koje se računa standardna devijacija ispunjavaju Gausovu raspodelu.

ima Gausovu raspodelu verovatnoće prema CGT. Međutim, EMG signal koji je nastao kao rezultat voljne kontrakcije ima bimodalnu raspodelu, tj. predstavlja kombinaciju dve Gausove funkcije gustine verovatnoće (jedna odgovara kontrakciji mišića, a druga relaksaciji). Na Sl. 127 je prikazana bimodalna raspodela EMG signala koji je meren na mišiću podlaktice zdravog ispitanika za tri kontrakcije na Univerzitetu u Beogradu – Elektrotehničkom fakultetu i koji je sačuvan u dokumentu “EMG.csv”.



Slika 127, Na gornjem panelu prikazan je histogram EMG signala čiji je grafik u vemenskom domenu prikazan na donjem panelu. Takođe, pored histograma EMG signala cijan bojom je prikazana i funkcija gustine verovatnoće.

Oblik histograma na Sl. 127 jeste sličan Gausovom zvonu, ali ima nešto širi središnji deo – ovakva deformacija Gausove krive je uobičajena kada je raspodela nastala kombinacijom odbiraka dve raspodele. Dodatno, primetiti na donjem panelu na Sl. 127, da signali za koje je prikazan histogram nemaju DC komponentu niti izražene šumove koji potiču od artifakta pokreta. Iako je tema vizuelizacije detaljno predstavljena u sledećem poglavlju i to na primeru ggplot2 paketa [228], ovde je ggplot2 preliminarno iskorišćen za prikaz Sl. 127 - odgovarajući kod je prikazan na Sl. 128. Za prikaz histograma, određen je najpre broj intervala b (Sl. 128), a na osnovu njega i širina intervala u oznaci bw (eng. *binwidth*), kao količnik razlike minimalne i maksimalne vrednosti napona EMG signala i broja intervala.

Primetiti na Sl. 128 da je kao ulaz u:

`ggplot()`

funkciju korišćen naziv *data frame*-a, a da je sa funkcijom:

`aes()`

označeno koja kolona u *data frame*-u odgovara kojoj osi na grafiku. Za prikaz histograma, funkcije gustine verovatnoće i EMG signala u vremenskom domenu korišćene su specijalizovane funkcije `geom_histogram()`, `geom_density()` i `geom_line()`. Boje na grafiku su određene korišćenjem naziva na engleskom jeziku (*cyan* za cijan boju i *black* za crnu), a moguće je koristiti i heksadecimalni zapis za boje (#000000 umesto *black* na Sl. 128). Pomoćne funkcije `xlab()` i `ylab()` su korišćene za nazive osa, a parametar *alpha* je korišćen da bi površina funkcije gustine verovatnoće na Sl. 127 bila providna.

```
library(ggplot2)
library(gridExtra)
sig <- read.csv("EMG.csv")
fs <- 1000
sig$ vreme <- seq(0, (length(sig$EMG)/fs - 1/fs), 1/fs)

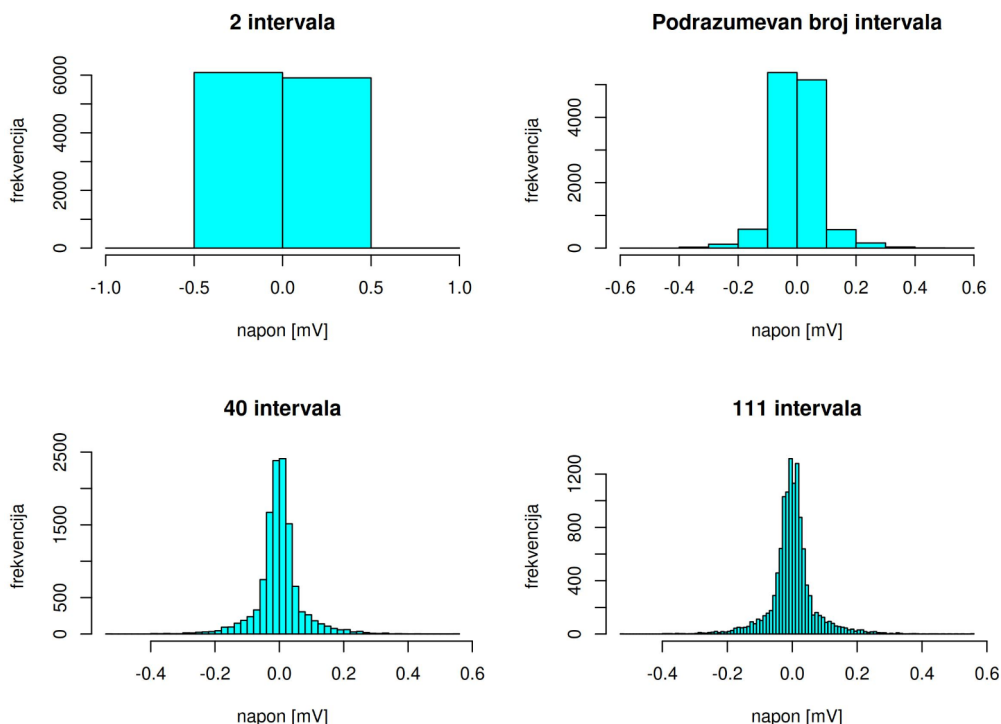
b <- round(sqrt(length(sig$EMG)) + 1)
bw <- (max(sig$EMG) - min(sig$EMG)) / b
emgHist <- ggplot(sig, aes(x=EMG)) +
  geom_histogram(aes(y = ..density..), binwidth = bw,
                 color = "#000000", fill = "white") +
  geom_density(alpha = 0.2, color = "cyan", fill = "cyan") +
  xlab("napon [mV]") + ylab("br. merenja")
emgTime <- ggplot(sig, aes(x=vreme, y=EMG)) +
  geom_line(color = "black") + xlab("vreme [s]") +
  ylab("napon [mV]")
g <- arrangeGrob(emgHist, emgTime)
ggsave("EMGHist.jpg", g, dpi = 300, units = "in",
        width = 6.5, height = 3.58)
```

Slika 128, R kod za prikaz histograma i EMG signala u vremenskom domenu na Sl. 127 primenom `ggplot2` paketa.

Na kraju, korišćenjem `gridExtra` paketa [416] omogućeno je da se dva grafika `emgHist`²⁹⁸ i `emgTime` prikažu na jednom panelu, a onda i funkcija:

`ggsave()`

da se grafik sačuva u željenom formatu sa odgovarajućom rezolucijom i unetim dimenzijama. Primititi da `ggplot2` grafici imaju slojevit način prikaza (Sl. 128), jer se za dodavanje pojedinačnih elemenata na grafiku koristi operator “+”.



Slika 129, Prikaz histograma EMG signala za različite intervale, uključujući preporučeni broj intervala od 111 koji je prikazan na donjem desnom panelu.

Moguće je prikazati histograme korišćenjem:

`hist()`

funkcije iz osnovnog R paketa, čiji ulazni parametar može biti broj intervala (*b*). Histogrami za različit broj intervala su prikazani na Sl. 129, a kod za njihov

²⁹⁸ U `ggplot2` paketu grafici se mogu sačuvati u odgovarajućim promenljivim koje su instance `ggplot` klase.

prikaz je dat na Sl. 130. Ako je broj intervala relativno mali, onda histogram ima oblik uniformne raspodele, dok za veći broj intervala histogram ima oblik Gausovog zvona, tačnije bimodalne raspodele, koja nastaje kombinacijom podataka koji imaju dve raspodele (Sl. 129).

Grafička provera funkcije gustine verovatnoće se može, pored histograma, izvršiti pomoću QQ grafika (eng. *Quantile Quantile plot*, odnosno *QQ plot*) [417] primenom *qqnorm()* funkcije. Ovaj grafik služi da se proveri poklapanje postojeće funkcije gustine verovatnoće (fgv) sa teorijski pretpostavljenom fgv. U slučaju da fgv merenih signala odgovara pretpostavljenoj (Gausovoj) raspodeli, onda postoji potpuno (ili relativno veliko) poklapanje sa crvenom linijom (označava $x = y$, gde su sa x i y predstavljene koordinate na aspcisi i ordinati) kao što je prikazano na Sl 131.

QQ grafik na osama ima kvantile funkcija gustine verovatnoće koje se porede, te ako podaci imaju pretežno normalnu funkciju gustine verovatnoće, QQ grafik će prikazati linearnu zavisnost. Da bi se prikazala linija $x = y$ koja prolazi kroz prvi i treći kvantil, koristi se funkcija:

qqline()

dok se za prikaz QQ grafika koristi ugrađena R funkcija:

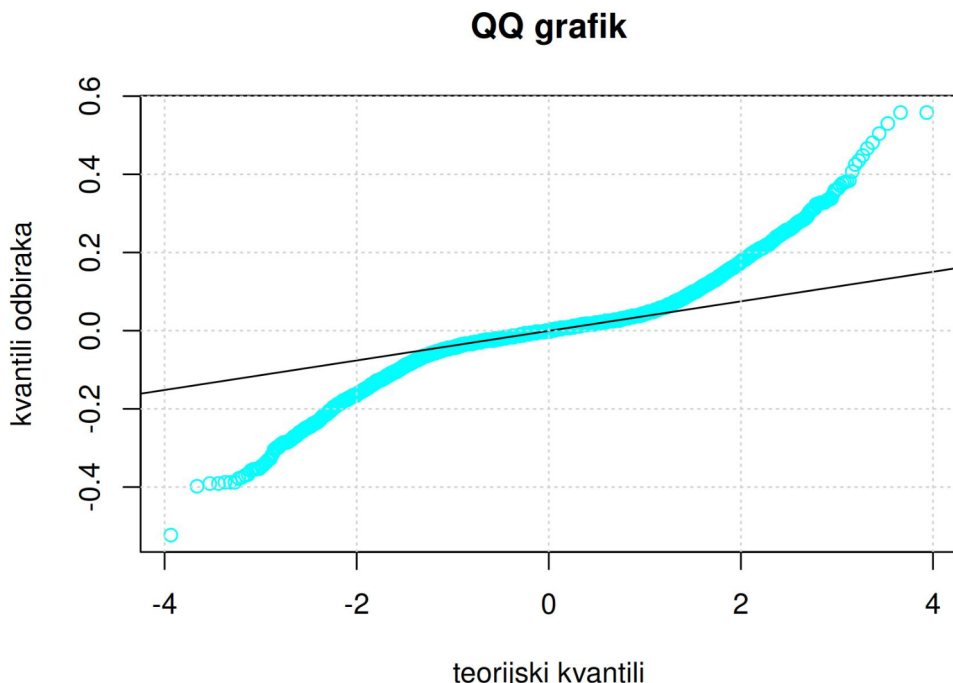
qqnorm()

```
par(mfrow = c(2, 2))
hist(sig$EMG, 2, xlab = "napon [mV]", ylab = "frekvencija",
     main = paste(as.character(2), "intervala", sep = " "), col = "cyan")
hist(sig$EMG, xlab = "napon [mV]", ylab = "frekvencija",
     main = "Podrazumevan broj intervala", col = "cyan")
hist(sig$EMG, 40, xlab = "napon [mV]", ylab = "frekvencija",
     main = paste(as.character(40), "intervala", sep = " "), col = "cyan")
hist(sig$EMG, b, xlab = "napon [mV]", ylab = "frekvencija",
     main = paste(as.character(b), "intervala", sep = " "), col = "cyan")
```

Slika 130, Kod koji omogućava prikaz četiri histograma za EMG signale, ako se koristi različit broj intervala (eng. *bins*). Za oblikovanje naslova spajanjem stringova korišćena je funkcija *paste()*.

Svako odstupanje od centralne linije ukazuje na “neslaganje” između realne i teorijski pretpostavljene fgv. Naravno, uz odgovarajuća znanja kako se dobija QQ grafik (Sl. 131), moguće ga je jednostavno realizovati primenom osnovnih funkcija u R programskom jeziku. Međutim, nema potrebe za tim, jer su razvijene gotove funkcije. U posebnim slučajevima može se raditi

parametrizacija ovog grafika [418], ali u opštem slučaju on se koristi za vizuelno poređenje fgv. Pored grafičkog prikaza moguće je i kvantitativno primenom statističkih testova odrediti da li neki podaci imaju Gausovu funkcije raspodele verovatnoće, što je detaljno objašnjeno u poglavlju Statistička analiza biosignala.

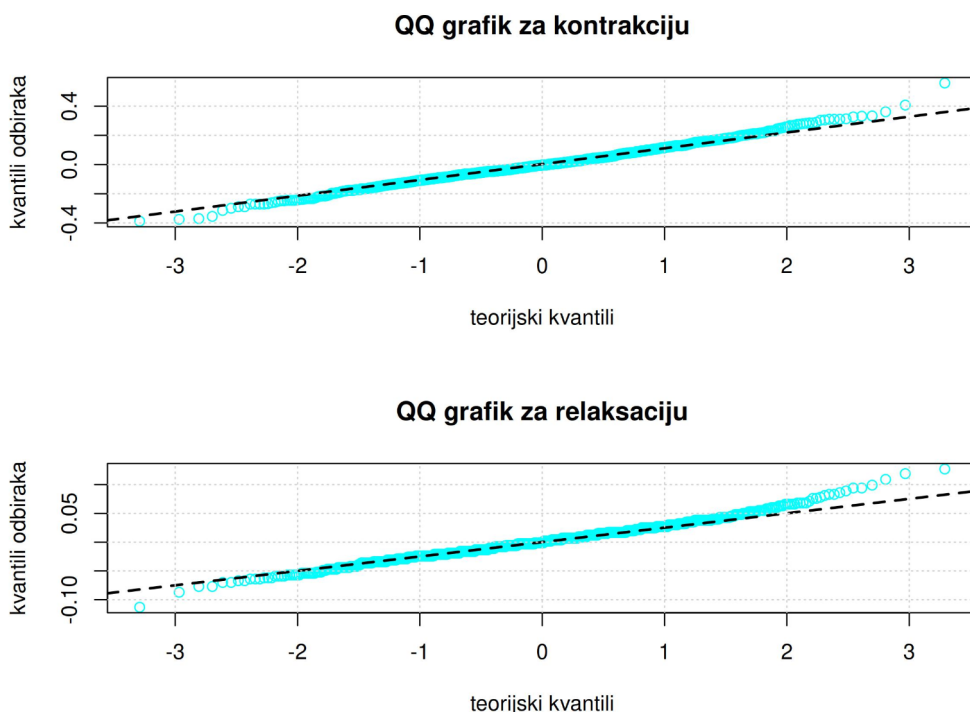


Slika 131, QQ grafik za ceo EMG signal koji uključuje tri kontrakcije.

Za procenu raspodele odbiraka EMG signala primenom QQ grafika ima smisla pokazati QQ grafik za ceo segment merenog EMG signala (kao na Sl. 131) koji uključuje tri mišićne kontrakcije, kao i za odabrane segmente tokom relaksacije i tokom kontrakcije. Na Sl. 132 je prikazan kod, a na Sl. 131 je prikazan QQ grafik za ceo EMG signal.

```
qqnorm(sig$EMG, col = "cyan", main = "QQ grafik",
        xlab = "teorijski kvantili", ylab = "kvantili odbiraka")
qqline(sig$EMG)
grid()
```

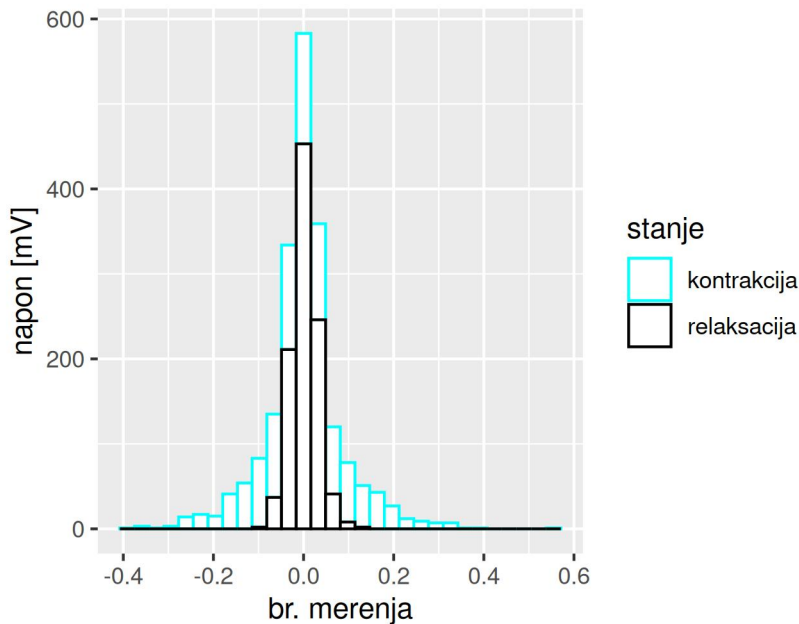
Slika 132, Primer korišćenja `qqnorm()` i `qqline()` funkcija za prikaz QQ grafika na celom EMG signalu.



Slika 133. QQ grafici za odbirke EMG signala koji odgovaraju relaksaciji i mišićnoj kontrakciji.

Na Sl. 131 je uočljivo da QQ grafik, odnosno međusobna relacija teorijskih i realnih kvantila na odbircima EMG signala ne prati centralnu liniju, te da EMG signal nema Gausovu raspodelu. Međutim, ako se predstave dva QQ grafika gde jedan odgovara kontrakciji EMG signala (od 5 s do 6 s) i tokom relaksacije (do prve sekunde trajanja signala) kao na Sl. 133 može se uočiti da oba QQ grafika imaju solidno poklapanje sa pretpostavljenom Gausovom raspodelom. Citaoci se upućuju da samostalno istraže R kod koji je korišćen da bi se prikazali QQ grafici za relaksaciju i kontrakciju kao na Sl. 133.

Da bi se poredili histogrami EMG signala koji su mereni tokom relaksacije i tokom kontrakcije, mogu se takođe koristiti funkcije ggplot2 paketa. Na Sl. 134 je prikazano takvo poređenje histograma. Primetiti da je standardna devijacija za relaksaciju značajno manja od standardne devijacije za kontrakciju (histogram za odbirke merene tokom kontrakcije je širi). Ovo je očekivan rezultat. R kod koji omogućava da se prikaže grafik na Sl. 134 je dat na Sl. 135.



Slika 134, Poređenje histograma EMG odbiraka tokom kontrakcije i tokom relaksacije.

```

emg <- list()
emg$napon <- c(sig$EMG[1:fs], sig$EMG[(5*fs):(6*fs-1)])
emg$vreme <- c(sig$vreme[1:fs], sig$vreme[(5*fs):(6*fs-1)])
emg$stanje <- c(rep("relaksacija", fs), rep("kontrakcija", fs))
emg <- data.frame(emg)

ggplot(emg, aes(x = napon, color = stanje)) +
  geom_histogram(fill = "white") +
  scale_color_manual(values = c("cyan", "black")) +
  xlab("br. merenja") + ylab("napon [mV]")

```

Slika 135, R kod koji omogućava prikaz dva histograma pomoću funkcija u ggplot2 paketu.

Da bi se koristio ggplot2 za prikaz histograma, potrebno je reorganizovati podatke u okviru novog *data frame*-a pod nazivom *emg* koji sadrži kolonu *stanje* u kojoj se nalazi informacija da li je EMG odbirak snimljen tokom relaksacije ili tokom kontrakcije. Dodatno, kako bi se upravljalo prikazom boja na grafiku korišćena je funkcija za ručno dodavanje skaliranje boja:

scale_color_manual()

Histogram EMG signala ne predstavlja uobičajen način analize EMG signala, ali može biti korišćen za ocenu morfologije EMG signala i izdvajanje obeležja

iz EMG signala koji su mereni sa mišića osoba koje su obolele od Parkinsonove bolesti [419]. Ovaj prilaz se pokazao uspešnim u poređenju impulsnog oblika EMG signala koji je karakterističan kod pacijenata kod kojih je dijagnostifikovana Parkinsonova bolest sa oblikom EMG signala koji su mereni na zdravim osobama. Dakle, primenom histograma moguća je uspešna diferencijacija između ispitanika koji su oboleli od Parkinsonove bolesti i zdravih ispitanika. Postoje i drugi radovi u kojima su istraživači koristili histogram za prepoznavanje različitih funkcionalnih pokreta iz EMG signala za kontrolu neuralne proteze [420]. Komercijalno dostupni softver za analizu EMG signala, takođe, omogućava prikaz histograma EMG signala²⁹⁹.

3.2.2 Generisanje EMG biosignala

Primer simuliranog EMG signala dat je na Sl. 136, a predstavlja rezultat primene modela iz [421]. Model je realizovan na primeru heteroskedastične metode (eng. *heteroscedasticity*)³⁰⁰, koja pretpostavlja da su podaci od interesa sastavljeni iz dva ili više skupa podataka sa istom raspodelom, ali sa slučajnom varijansom (proces odgovara slučajnom procesu sa promenljivom varijansom). Ovakav pristup u potpunosti odgovara modelu EMG signala koji je sastavljen iz više celina (relaksacije i kontrakcije) koje imaju različite Gausove fgv. Obe raspodele imaju srednju vrednost 0, a razlikuju im se varijanse (što je očekivano) Kada se pogleda vremenski prikaz EMG signala i histogrami odbiraka EMG signala (Sl. 127, Sl. 129 i Sl. 134) primećuje se da je amplituda EMG signala tokom aktivacije primetno veća od amplitude EMG signala tokom relaksacije). Na Sl. 136. je prikazan primer generisanog EMG signala primenom funkcije iz biosignalEMG paketa:

syntheticemg()

dok je na Sl. 137 prikazan kod za prikaz signala i pokretanje funkcije sa podrazumevanim parametrima. Parametri koji se mogu zadati ovoj funkciji su [421]:

- broj odbiraka signala (*n.lenght.out*) – podrazumevana vrednost je 10000,
- standardna devijacija odbiraka tokom kontrakcije mišića (*on.sd*) – podrazumevana vrednost je 1,

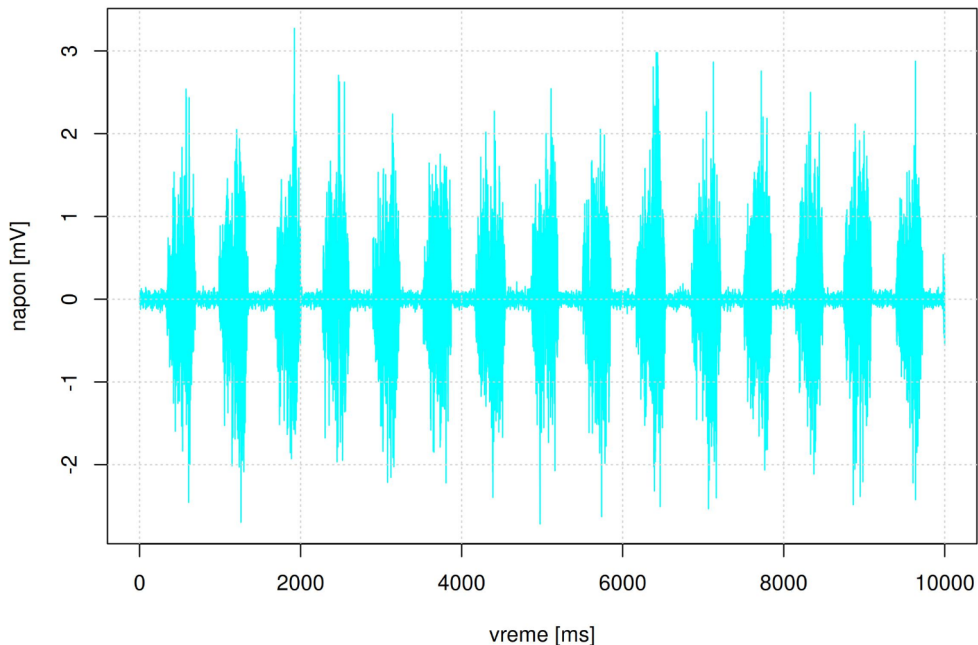
299 Biopac, <https://www.biopac.com/application/emgelectromyography/advanced-feature/histogram-analysis/>, pristupljeno 16.02.2024.

300 Heteroscedasticity, https://en.wikipedia.org/wiki/Homoscedasticity_and_heteroscedasticity, pristupljeno 01.06.2024.

- srednja vrednost trajanja kontrakcije u odbircima (*on.duration.mean*) – podrazumevana vrednost je 350,
- standardna devijacija trajanja kontrakcije u odbircima (*on.duration.sd*) – podrazumevana vrednost je 10,
- standardna devijacija odbiraka tokom relaksacije mišića (*off.sd*) – podrazumevana vrednost je 0.05,
- srednja vrednost trajanja relaksacije u odbircima (*off.duration.mean*) – podrazumevana vrednost je 300,
- standardna devijacija trajanja relaksacije u odbircima (*off.duration.sd*) – podrazumevana vrednost je 20,
- indikacija kada počinje i kada se završava aktivna faza (*on.mode.pos*) – podrazumevana vrednost je 0.75,
- parametar koji utiče na oblik signala u vremenskom domenu (*shape.factor*) – podrazumevana vrednost je 0.5,
- frekvencija odabiranja za generisanje EMG signala (*samplingrate*) – nema podrazumevanu vrednost, ali je usvojeno na Sl. 136 da je 1000 Hz,
- jedinice u kojima će biti amplituda generisanog EMG signala (*units*) – nema podrazumevanu vrednost, ali je usvojeno na Sl. 136 da su mV i
- ime promenljive za prikaz generisanog EMG signala na graficima (*data.name*) – podrazumevana vrednost argumenta, odnosno naziv promenljive je “Synthetic EMG”.

Na osnovu predstavljenih ulaznih parametara *syntheticemg()* funkcije, kao i na osnovu osobina EMG signala koje su pokazane u prethodnom potpoglavlju, može se doći do jednostavnog zaključka da je model EMG signala realizovan kroz dve Gausove raspodele – jedna koja odgovara odbircima tokom kontrakcije, a druga koja odgovara odbircima tokom relaksacije. Nakon toga, ti odbirci su spojeni u vremenskom domenu i rezultat je generisani EMG kao na Sl. 136. Bez obzira što ulazni parametri uključuju frekvenciju odabiranja i jedinice za amplitudu EMG signala, generisani signal je veštački signal i može mu se dodeliti bilo koja frekvencija odabiranja i bilo koja amplituda. Kako je cilj da se generiše signal koji oponaša realan EMG, onda ima smisla da se koriste vrednosti koje odgovaraju realnim signalima koji su mereni površinskim elektrodama u skladu sa postojećim preporukama [422]. Ukratko, ne bi bilo pogrešno da je na obe ose na Sl. 136 ispisana bilo koja jedinica ili proizvoljna jedinica, odnosno a.u. skraćunica. Model sintetičkog EMG signala u oznaci X_i se može predstaviti Relacijom (24).

$$X_i \sim N(0 | b_i \sigma_a^2 + (1 - b_i) \sigma_s^2) \quad \text{Relacija (24)}$$



Slika 136, Prikaz generisanog EMG signala sa podrazumevanim parametrima.

```
library(biosignalEMG)
```

```
semg <- syntheticemg()
plot(semg$values, xlab = "vreme [ms]", ylab = "napon [mV]",
      type = "l", col = "cyan")
grid()
```

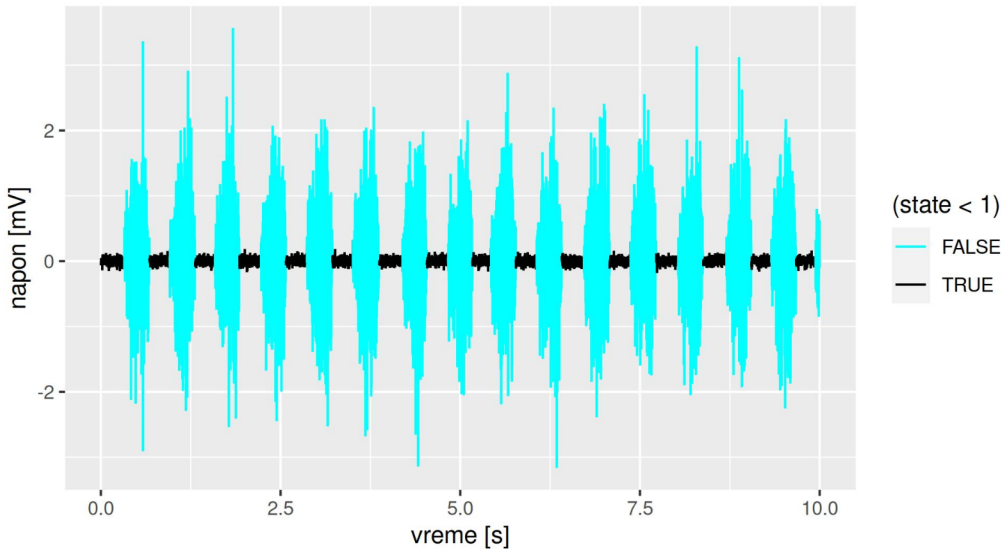
Slika 137, R kod za generisanje sintetičkog EMG signala primenom funkcije *syntheticemg()* iz biosignalEMG paketa.

Različiti trenuci aktivacije u modelu su predstavljeni sekvencom logičkih vrednosti $b = \{b_1, b_2, \dots, b_n\}$ i to na sledeći način: ako je $b_i = 1$ onda je mišić aktivan, a ako je $b_i = 0$ onda je mišić relaksiran. U Relaciji (24) su sa σ_a^2 i σ_s^2 označene standardne devijacije za Gausovu raspodelu EMG odbiraka tokom aktivacije i tokom relaksacije mišića, a sa N je označena Gausova, tj. normalna raspodela. Iz Relacije (24) je uočljivo da mišić ne može istovremeno biti aktivan i relaksiran, jer je trenutak u kome je mišić relaksiran ($1 - b_i$) suprotan trenutku u kome je mišić aktivan (b_i). Ovako generisan EMG signal X_i ima srednju vrednost 0, a varijansa signala je određena varijansom koja je data u Relaciji (25).

$$\sigma_i^2 = b_i \sigma_a^2 + (1 - b_i) \sigma_s^2$$

Relacija (25)

Predstavljeni model EMG signala se može koristiti za generisanje sintetičkih ENG signala koji nisu nastali kao rezultat električne stimulacije, odnosno koji nisu deterministički [421].



Slika 138, Sintetički generisan EMG signal prikazan pomoću ggplot2 funkcija sa označenim intervalima aktivacije i relaksacije.

```
emgDf <- list()
emgDf$values <- semg$values
emgDf$state <- semg$on.off
emgDf$vreme <- (1:length(emgDf$values))/1000 # u sekundama
emgDf <- data.frame(emgDf)
```

```
ggplot(emgDf, aes(x = vreme, y = values)) +
  geom_line(aes( group = 1, color = (state < 1) )) +
  xlab("vreme [s]") + ylab("napon [mV]") +
  scale_color_manual(values = c("cyan", "black"))
```

Slika 139, Prikaz sintetičkog EMG signala tokom aktivacije i relaksacije kao na Sl. 138, primenom ggplot2 paketa.

Nije pogodno da sekvenca b bude isključivo logička, već bi bilo zgodno da odgovara normalizovanim vrednostima od 0 do 1, jer realna kontrakcija ne nastaje u jednom trenutku, već postoji prelazni period, odnosno tranzijent.

Otuda i veliki problem u detekciji početka i kraja mišićne kontrakcije koji su neobično dobro uočljivi na simuliranom signalu (Sl. 138). Iako realni signal može, u pojedinim situacijama, da ima tako dobro razdvojene aktivacije i relaksacije, to nije često slučaj [105, 423]. Za prikaz grafika kao na Sl. 138 korišćen je ggplot2 paket, a boje za prikaz relaksacije i kontrakcije su odabrane u odnosu na logički uslov (Sl. 139). Takođe, generisan EMG signal je zapakovan u odgovarajući *data frame* kako bi se omogućio prikaz signala pomoću ggplot2 funkcija koje su specijalizovane da rade sa *data frame* podacima.

Moguće unapređenje biosignalEMG metode za simulaciju EMG signala bi se moglo postići korišćenjem rampe za aktivaciju mišića, tj. realizacijom funkcije koja bi simulirala postepeno uključanje motornih jedinica u mišiću i postepeno povećanje amplitude EMG signala. Jedan takav sistem realizovan je u master radu koji je predstavljao nastavak rada na projektu na predmetu Tehnike obrade biomedicinskih signala na Elektrotehničkom fakultetu, Univerziteta u Beogradu [424]. Detekcija početka i kraja mišićne kontrakcije je jedna od najizazovnijih metoda, posebno u prisustvu šuma i u slučaju mišićnih kontrakcija relativno nižeg intenziteta [345, 425], te predstavljen model ima veliku upotrebnu vrednost za evaluaciju algoritama. Iskustvo je pokazalo da pored adaptivnog praga, najbolji rezultat daju metode koje su zasnovane na detekciji nagiba koji se računa na odabranim parametrima u okviru prozorskih funkcija, kao što je već objašnjeno u ranijim poglavljima [344, 426].

```

i <- 0
b <- numeric()
semg <- numeric()
while (i < n.length.out) {
  n.off <- round(rnorm(1, off.duration.mean, off.duration.sd))
  if (n.off < 1)
    n.off <- 1
  n.on <- round(rnorm(1, on.duration.mean, on.duration.sd))
  if (n.on < 1)
    n.on <- 1
  b <- c(b, rep(0, n.off), rep(1, n.on))
  semg <- c(semg, rnorm(n.off, 0, off.sd), c(seq(off.sd/on.sd, 1, length.out = round(on.mode.pos *
    n.on)), seq(1, off.sd/on.sd, length.out = n.on - round(on.mode.pos *
    n.on)))^(shape.factor) * rnorm(n.on, 0, on.sd))
  i <- i + n.off + n.on
}

```

Slika 140, Deo koda iz funkcije *syntheticemg()* iz biosignalEMG paketa [409].

Ako se pogleda izvorni kod funkcije *syntheticemg()*³⁰¹ iz biosignalEMG³⁰² paketa za simulaciju EMG signala (Sl. 140), može se primetiti da je proces generisanja veoma jednostavan i da uključuje korišćenje *rnorm()* funkcije za generisanje pseudoslučajnih odbiraka koji prate Gausovu raspodelu. Predstavljen model EMG signala ne uključuje dodavanje šuma na signalu, ali to je moguće postići generisanjem šuma odvojeno od EMG signala, pa potom dodavanjem šuma signalu, odnosno njihovim sabiranjem.

Prikazani model EMG signala je moguće proširiti tako da omogućava generisanje simuliranih višekanalnih signala. Ovaj posao nije baš jednostavan, jer mora uključiti i prostornu modulaciju amplitude, kao i kašnjenja pojedinačnih signala [104]. Iako bi ova modifikacija doprinela dodatnoj kompleksnosti metode, ona nije nemoguća. Na kraju, važno je napomenuti da je EMG signal koji predstavlja rezultat kompleksnog procesa, te na njega utiču anatomske i fiziološke karakteristike mišića, kontrolni mehanizmi nervnog sistema, ali i karakteristike mernog uređaja [265]. Usled toga, model koji podrazumeva samo karakteristike signala iz oblasti obrade signala se mora uzeti sa određenom rezervom, jer predstavljeni model za generisanje EMG signala ne omogućava da se:

- odaberu anatomske i fiziološke karakteristike mišića čiji se signal simulira,
- izvrši izbor mišića na telu čoveka za koji se simulira EMG signal,
- osim vremena početka i kraja mišićne kontrakcije, drugi uticaji kontrolnog signala iz nervnog sistema nisu uzeti u obzir prilikom realizacije predstavljenog modela,
- merni uređaji nisu uzeti u obzir (na primer: pojačanje, frekvencijski opseg i druge karakteristike hardverskih filtara, oblik elektroda – koncentrične, u nizu, matrične i sl), ali se uticaj ovih parametara može jednostavno implementirati u model.

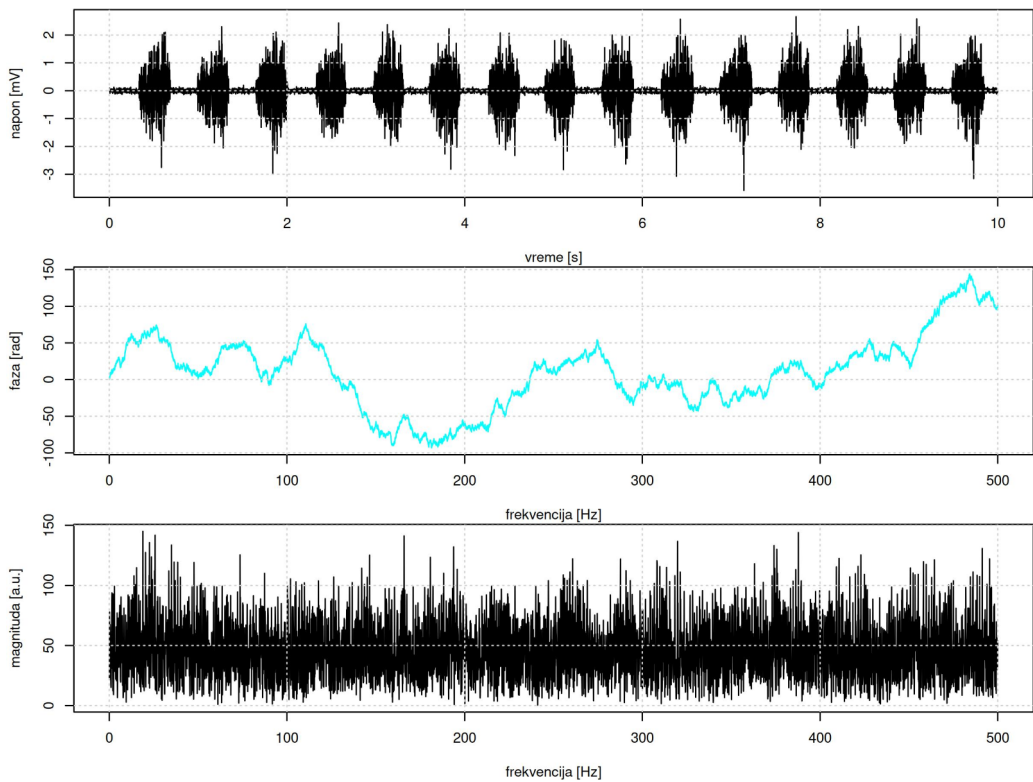
3.2.3 (Ne)stacionarnost simuliranih biosignala

Na primeru modela EMG signala pokazano je kako se nestacionarnost signala može predstaviti heteroskedastičnim modelom koji podrazumeva da smene intervala (relaksacija i aktivacija mišića) na EMG signalu odgovaraju različitim

301 Pristup izvornom kodu je omogućen preko stranice paketa u CRAN repozitorijumu. Najčešće je izvorni kod moguće preuzeti u komprimovanim datotekama.

302 Pored biosignalEMG, postoje i drugi paketi u R-u i funkcije koje se mogu koristiti za simulaciju EMG signala [427].

raspodelama. Kada bi se, pored magnitude prikazala i faza³⁰³ ovako generisanog signala, uočilo bi se da faza signala ima širok raspon u radijanima kada se primeni *unwrap()* funkcija, te sledi da je u fazi EMG signala sadržana značajna informacija, što je već ranije pokazano na primeru govornog signala [428-429]. Potpuno suprotan slučaj je iskorišćen za generisanje sintetičkog EKG signala, odnosno za generisanje varijabilnosti srčanog ritma, u kome je pretpostavljeno da je varijabilnost srčanog ritma stacionarna, da je važna samo amplitudska frekventijska karakteristika, te da se faza može simulirati uniformnom raspodelom [430].

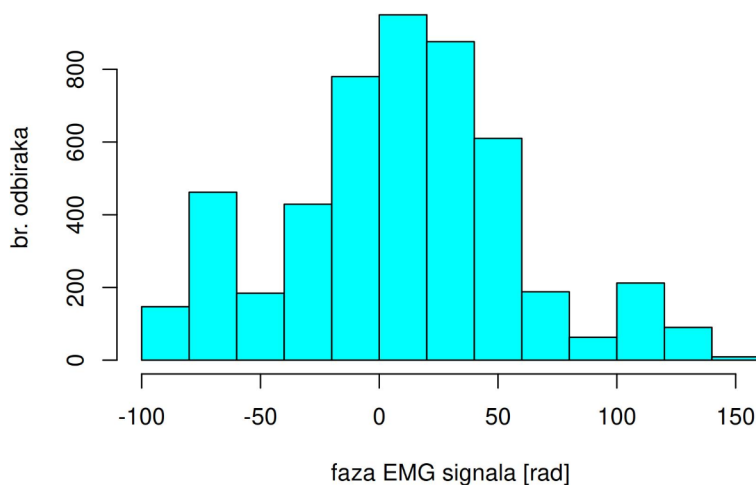


Slika 141, Prikaz simuliranog EMG signala (gornji panel) sa faznom frekventijskom karakteristikom (srednji panel) i amplitudskom frekventijskom karakteristikom signala (donji panel).

Pomoću funkcija iz signal paketa [275] iscrtane su magnituda i faza simuliranog EMG signala (Sl. 141), a pomoću *hist()* funkcije pokazano je da raspodela faze “ličić” na Gausovu raspodelu (Sl. 142). U zavisnosti od toga kako je realan signal snimljen, mogu se dobiti različiti oblici faze na realnim

303 Pod magnitudom i fazom se podrazumevaju magnitudska/amplitudska frekventijska karakteristika i fazna frekventijska karakteristika.

signalima. Predstavljena metoda za simuliranje EMG signala koristi beo šum za simuliranje signala u vremenskom domenu. Iz tog razloga, nisu sve frekvencije prisutne na realnom EMG signalu podjednako kao što je prikazano na magnitufskoj karakteristici na Sl. 141. Takođe, iz prikaza magnitude realnog signala može se uočiti istaknut šum na 50 Hz, kao i nešto manje izražena DC komponenta (Sl. 78). Najveću snagu EMG signal ima između 100 Hz i 200 Hz, a frekvencijski sadržaj nije posebno uočljiv na frekvencijama koje su približno jednake 500 Hz kao što je to slučaj sa simuliranim signalom kada se uporede prikazi na Sl. 141 i Sl. 78.



Slika 142, Histogram faze simuliranog EMG signala na kome se vidi da raspodela faze ima oblik Gausovog zvona.

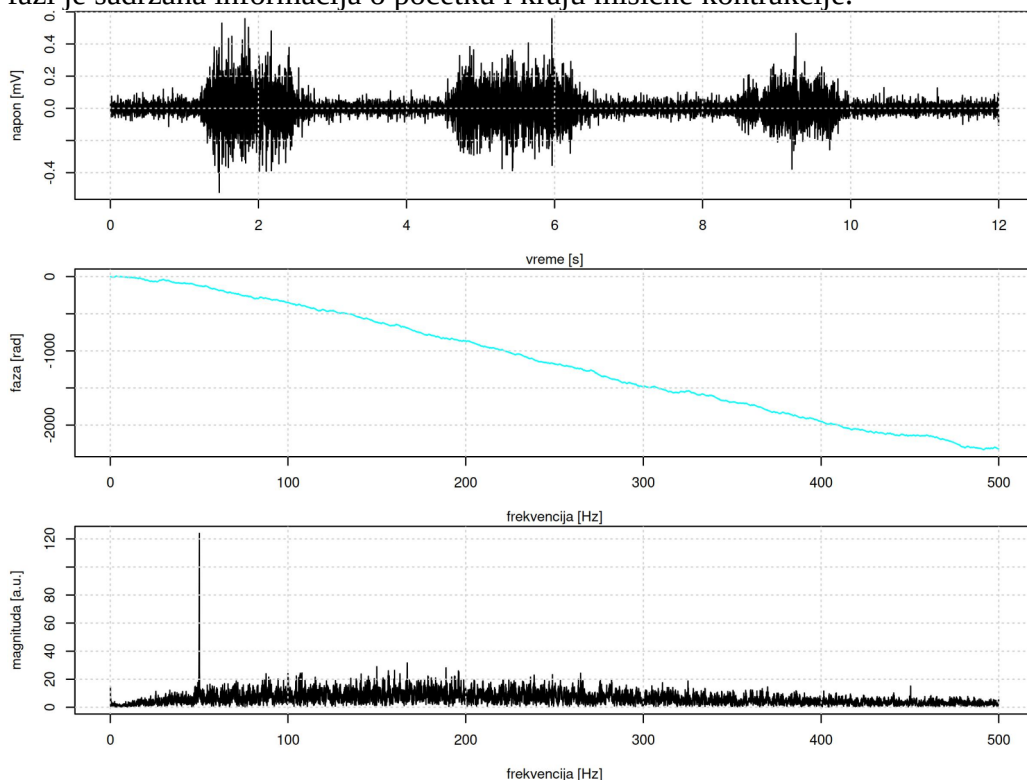
Kako bi se ukazalo na značaj faze EMG signala, moguće je primeniti funkciju za inverznu Furijeovu transformaciju:

ifft()

kao na Sl. 144. U prvom slučaju, rekonstruisan je signal sa realnom fazom i realnom magnitudom (*emgNov*), u drugom slučaju rekonstruisan je signal sa realnom magnitudom i sa fazom koja je dobijena iz uniformne raspodele (*emgNov2*), a u trećem slučaju je faza realna, a magnituda rekonstruisana iz uniformne raspodele (*emgNov3*).

Slično kao u GNU Octave-u, potrebno je da magnituda i faza za koje se primenjuje *ifft()* funkcija budu dvostrani (Sl. 144). Dodatno, *ifft()* funkcija kao rezultat daje kompleksne brojeve, te je potrebno primeniti funkciju *Re()* za

izdvajanje realnog dela da bi se dobio signal u vremenskom domenu. Grafici rekonstruisanih signala u vremenskom domenu su prikazani na Sl. 145 i sa njih je uočljivo da je faza od presudnog značaja za rekonstrukciju tri aktivacije. U fazi je sadržana informacija o početku i kraju mišićne kontrakcije.



Slika 143, Realni EMG signal sa magnitudom i fazom.

```

emgMag <- c(emgMag, emgMag)
emgFaza <- c(emgFaza, emgFaza)

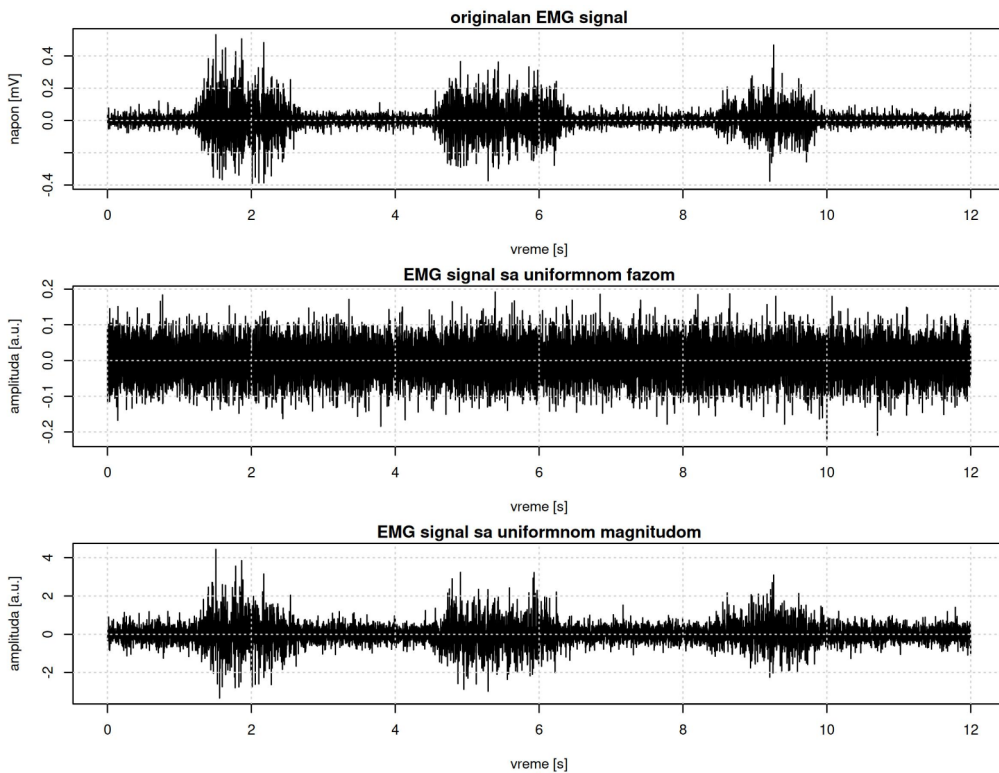
emgNovFFT = emgMag * exp(1i*emgFaza);
emgNov = Re(fft(emgNovFFT));

emgNovFFT2 = emgMag * exp(1i*runif(length(emgFaza),
                                     min = min(emgFaza), max = max(emgFaza)));
emgNov2 = Re(fft(emgNovFFT2));

emgNovFFT3 = runif(length(emgMag), min = min(emgMag), max = max(emgMag)) *
               exp(1i*emgFaza);
emgNov3 = Re(fft(emgNovFFT3));

```

Slika 144, R kod za rekonstrukciju signala u vremenskom domenu primenom *fft()* funkcije za inverznu Furijeovu transformaciju.



Slika 145, Rekonstrukcija EMG signala. Gornji panel: signal nastao primenom inverzne Furijeove transformacije iz originalne faze i originalne magnitude. Srednji panel: signal nastao rekonstrukcijom iz originalne magnitude i uniformno raspodeljene faze. Donji panel: signal nastao iz originalne faze i uniformno raspodeljene magnitude.

U fazi signala je sadržana informacija o njegovoj nestacionarnosti. Da se rekonstrukcija kojim slučajem vršila na segmentima na kojima postoji relaksacija ili kontrakcija relativno konstantnog intenziteta, u tom slučaju bi rekonstrukcija sa uniformno raspodeljenim odbircima faze bila moguća. Takvi segmenti EMG signala su stacionarni. Slična rekonstrukcija je predstavljena u [429] gde je pokazano da faza koja je rezultat dvodimenzione Furijeove transformacije može biti ključna za rekonstruisanje slika. Iako se faza signala ne koristi često u analizi signala, ona ima ključnu ulogu u analizi video signala i korišćena je za izdvajanje karakterističnih obeležja i lokalizaciju događaja [428, 431]. Na primer, faza EMG signala se pokazala korisnom u diferenciranju mišićno-skeletnih oboljenja [432].

Pored EMG signala, heteroskedastični princip i model signala u frekvencijskom domenu, se mogu primeniti za generisanje drugih sintetičkih biosignala. Za generisanje sintetičkog EGG signala [147], faza stacionarnih EGG signala je

modelirana kao skup pseudoslučajnih brojeva koji imaju uniformnu raspodelu, dok je spektar signala modeliran Gausovim funkcijama. Za modeliranje spektra snage varijabilnosti srčanog ritma, Gausove funkcije su korišćene za modeliranje magnitude, dok je faza modelirana uniformnom raspodelom stacionarnosti varijabilnosti srčanog ritma [430].

4 Eksplorativna analiza biosignala

Eksplorativna analiza signala (eng. *Exploratory Data Analysis*, skraćeno EDA) je oblast nauke o podacima koja uključuje primenu metoda za vizuelizaciju i deskriptivnu statistiku signala. Veoma je korisna za uočavanje ponovljivih obrazaca u podacima. Za razliku od tradicionalnih statističkih metoda koje se zasnivaju na definisanju hipoteza, pa se mogu smatrati sudskim ili kvazi-sudskim procesom (donosi se odluka da li je nešto istina ili ne), EDA se smatra detektivskim poslom u radu sa podacima, jer ne postoji rigidna pretpostavka za njenu primenu. [433]

U ovom udžbeniku, posebna pažnja je posvećena EDA oblasti koja uključuje vizuelizaciju signala. Iako u nauci o podacima, vizuelizacija signala predstavlja važan deo, zapostavljena je u obradi biosignala – nema puno literature iz oblasti biomedicinskog inženjerstva koja se bavi problemom prikaza informacije. Tipovi grafika, koji se često koriste u vizuelnom prikazu biosignala, mogu se podeliti u sledeće grupe:

- Dijagram rasejanja ili međusobne zavisnosti dve promenljive za koji je opšteprihvaćen engleski termin *scatter plot*. Primeri takvih grafika su:
 - Prikaz linearne/nelinearne zavisnosti dve promenljive (na primer, zavisnost sistolnog krvnog pritiska od indeksa telesne mase i procena njihove povezanosti).
 - Prikaz signala u vremenskom i frekvencijskom domenu (zavisnost amplitude ili magnitude od vremena ili frekvencije).
 - Funkcija gustine verovatnoće – primenjuje se u oceni raspodele merenih parametara ili signala, ali i u električnim merenjima [78].
 - QQ grafik, koji se koristi za ocenu funkcije gustine verovatnoće signala, prikazuje međusobnu zavisnost realne od pretpostavljene raspodele.
- Prikaz barova – pogodan za grafičko predstavljanje diskretnih grupa³⁰⁴ parametara signala. Tipovi barova se mogu razlikovati u zavisnosti od vrste informacije koju prikazuju, a mogu se kombinovati i sa drugim prikazima:

³⁰⁴ Na računaru se uvek prikazuju diskretne grupe. Međutim, ovde se misli na manji broj grupa parametara – na primer, razlike u srčanom ritmu pušača i nepušača su pogodnije za prikaz preko barova, dok amplituda signala u diskretnim vremenskim trenucima (prikaz biosignala u vremenskom domenu) nije pogodna za prikaz pomoću barova, jer sadrži relativno veliki broj grupa. Primera radi, 10 s signala merenog sa frekvencijom odabiranja od 1000 Hz sadrži 10000 odbiraka i ima 10000 diskretnih vrednosti.

- *Box* grafik ili kutijasti/pravougaoni grafik – koristi se kada je potrebno prikazati sumarno podatke za koje se ne zna funkcija gustine raspodele verovatnoće (ili za koje je poznato da nemaju Gausovu funkciju gustine raspodele verovatnoće), sa vidljivim osnovnim parametrima kao što su medijana, prvi i treći kvartil. Poznat je i pod nazivom na engleskom jeziku *whisker plot*³⁰⁵ zbog dodatne informacije o opsegu podataka (prikazuje minimalnu i maksimalnu vrednost).
- Bar sa standardnim devijacijama (eng. *error bar* ili *dynamite plot*) – koristi se za prikaz srednje vrednosti grupa podataka i prikaz standardne devijacije/merne nesigurnosti tipa A (najčešće sa faktorom proširenja 1, ali je u upotrebi i faktor proširenja 2 [78]). Treba voditi računa prilikom korišćenja ovog grafika, jer je prilično kritikovan (neki naučnici i profesori smatraju da ga ne bi trebalo nikada koristiti i da je bolji pristup prikaz pojedinačnih parametara/biosignala/biomarkera za svaku grupu koja se posmatra³⁰⁶). Međutim, ako se ovaj grafik ipak koristi, onda to može biti u sledeće dve situacije:
 - kada je poznato da je funkcija gustine verovatnoće Gausova i
 - kada je cilj da se računanjem i prikazom standardne devijacije i srednje vrednosti utvrdi disperzija podataka, ali u tom slučaju se ne može pridružiti odgovarajuća verovatnoća vrednosti standardne devijacije (68.2% je verovatnoća koja odgovara Gausovoj raspodeli [78]).
- Histogram predstavlja posebnu vrstu bar grafika i za razliku od ostalih bar grafova, nije poželjno histogram prikazati horizontalno, jer on služi za ocenu funkcije gustine verovatnoće (Gausovo zvono se ne može dobro vizuelno proceniti kada je rotirano za 90 stepeni). Za sve ostale barove važi da je njihov horizontalni prikaz prijatiji za oko od vertikalnog položaja barova [434].
- Violinski grafik (eng. *violin plot*) kombinuje prikaz funkcije gustine verovatnoće i prikaz bar grafika u diskretnim grupama. Koristan je prilikom poređenja funkcija gustine verovatnoće između više kategorija podataka. Simetričan izgled grafika (rotirane funkcije

305 Ovaj naziv je manje uobičajen za *box plot*, a u bukvalnom prevodu sa engleskog *whisker* znači dlaka ili brkovi.

306 Pogledati blog tj. otvoreno pismo prof. Irizarija sa Harvard Univerziteta (Fakultet za javno zdravlje, <https://www.hsph.harvard.edu/profile/rafael-a-irizarry/>, pristupljeno 01.05.2024.) iz 2019 godine urednicima naučnih časopisa kako treba izbaciti ove grafike iz upotrebe, <https://simplystatistics.org/posts/2019-02-21-dynamite-plots-must-die/>, pristupljeno 01.05.2024.

prikazane kao lik u ogledalu oko jedne ose) je pogodan za posmatranje i uočavanje razlika u raspodelama, a naziv je dobio jer podseća na muzički instrument.

- Prikaz grafika vremenske linije (eng. *timeline plot*) je posebna vrsta diskretnog prikaza kada je na apscisi postavljena informacija o vremenu, dok se na ordinati nalazi tip događaja.
- Višedimenzioni grafici, od kojih su najčešće u upotrebi oni sa dve dimenzije (2D) sa dve ose (apscisa i ordinata, odnosno tzv. *x*-osa i *y*-osa), i kod kojih su dodatne dimenzije (informacije) prikazane kroz boju, šrafuru, oblik tačaka ili na neki drugi način. Korišćenje 3D grafika (sa tri ose) se ne preporučuje, jer ih nije jednostavno tumačiti i ponekad informacija može biti pogrešno protumačena³⁰⁷. Primeri 2D grafika sa pridruženom dodatnom dimenzijom/informacijom su:
 - Spektrogram biosignala koji omogućava vremensko-frekvencijski prikaz signala. Uobičajeno, na apscisi sadrži vreme, na ordinati frekvenciju i bojom je označena magnituda. Prikaz je koristan u analizi nestacionarnih signala, posebno EEG signala [395].
 - *Heat map* ili toplotna mapa (na engleskom se sreće i kao jedna reč *heatmap*) predstavlja tip grafika koji prikazuje informaciju kodovanu bojama. Može da sadrži i elemente dendrograma (dijagrama koji povezuje slične elemente, a koristi se u klasterizaciji, odnosno kategorizaciji prikazanih parametara). Ovaj grafik je dobar pretkorak u klasterizaciji podataka, a može se koristiti i za uočavanje povezanosti i istaknutosti pojedinih vrednosti (na primer, u slučaju da se prati izraženost / ekspresija gena [435]).

U ovom poglavlju će biti predstavljeni osnovni grafici sa osvrtom na realizaciju u R-u. Jedna od najvažnijih prednosti R-a, u odnosu na druge programske jezike, je što omogućava jednostavno kreiranje dobrih grafičkih prikaza koji su pri tom reproducibilni čemu doprinosi činjenica da je R slobodan softver [245]. Ugrađene R funkcije, koje su do sada u velikoj meri korišćene, nisu uvek idealne za grafički prikaz informacije iz sledećih razloga:

- Izgled osnovnih R grafika nije idealan. Na primer, grafik na kome je prikazan roze šum na Sl. 124 ima puno praznog prostora dva panela i stepen iskorišćenosti ukupnog prostora koji zauzima korisna informacija u odnosu na margine nije idealan (gruba procena je da predstavlja nešto

³⁰⁷ Za konkretne primere zašto treba izbegavati višedimenzioni prikaz, pogledati listu 10 najgorih grafika koju je pripremio Prof. Karl Broman sa Univerziteta Viskonsin-Medison u SAD, https://www.biostat.wisc.edu/~kbroman/topten_worstgraphs/, pristupljeno 01.05.2024.

više do 50%, dok bi idealno bilo da je preko 80% ili čak 90%). Margine se mogu dodatno podesiti dodatnim podešavanjima, ali je zametno.

- Zahtevaju relativno duge skripte za jednostavne primere grafika. Prilikom prikaza više grafika, na jednom ili na više panela, mora se koristiti *par()* funkcija, a nazivi osa svakog od grafika i granice osa se moraju ručno podesiti za sve signale koji se prikazuju.
- Za izvoz grafika veće rezolucije potrebno je koristiti više od jedne dodatne komande kako bi se podesila rezolucija, dimenzija slike i sl.

Iz svih navedenih razloga, u ovom poglavlju će dominantno biti prikazana vizuelizacija signala u ggplot paketu (gg od eng. *Grammar of Graphics*).

4.1 Vizuelizacija signala

Vizuelizacija signala je jedan od najvažnijih koraka u analizi signala. U pojedinim situacijama vizuelizacija signala je dovoljna da se prikaže informacija od interesa te dalja analiza nije potrebna [436]. Iako na prvi pogled, vizuelizacija ostavlja utisak jednostavnog zadatka u analizi signala, za odgovarajuću primenu vizuelizacije signala, potrebno je dobro predznanje o osobinama signala i grafika da bi se izdvojila i vizuelno uočila informacija od interesa. Pored izdvajanja korisne informacije, koja predstavlja cilj analize, vizuelizacija signala se koristi i za proveru realizovanih metoda analize signala.

Danas je nezamislivo sprovesti analizu biosignala bez računara, odnosno bez automatskih računarskih metoda. Vizuelna evaluacija je preporučena i u važećim kliničkim standardima za EKG signale [437]. U slučaju EGG signala, situacija je još složenija, jer standardi ne postoje, a vrlo često se sprovedi ne samo vizuelna evaluacija, već i vizuelna identifikacija šuma pokreta [79]. Vizuelizacija predstavlja veliki izazov za inženjere, jer pored tehničkih znanja, zahteva i poznavanje dizajna i osećaj za lepo. To je odlična prilika za biomedicinske inženjere da prikažu svoju multidisciplinarnost i da usavrše postojeća znanja usvajanjem novih. Pored toga, svaka analiza signala bi trebalo da započne posmatranjem signala. Logično se nameće zaključak da fenomen koji nije uočljiv okom na signalu se ne može kvantifikovati na odgovarajući način. Primera radi, ako se posmatra EKG signal, može se primetiti da su na njemu uočljive karakteristične tačke, odnosno ponovljiv oblik signala opisan PQRST tačkama, te se rastojanja između karakterističnih tačaka mogu ucrtati rukom i metoda da se pronađu se može jednostavno pretočiti u kod:

- lokalni ekstremumi koji odgovaraju pojedinačnim otkucajima srca (R pikovi) se mogu izdvojiti odgovarajućim tehnikama i

- ostale karakteristične tačke se nalaze u neposrednoj okolini R pika (pre ili posle pojave R pika), pa se jednostavno realizuju automatske metode za detekciju ostalih pikova tako što se u okolini R pika definiše vremenski prozor u kome se amplituda EKG signala poredi sa pragom.

Naravno, postoje i druge metode za određivanje karakterističnih obeležja EKG signala (delineacija signala), ali opisana metoda je intuitivno jasna i direktno se dobija posmatranjem EKG signala u vremenskom domenu. Generalno, realizacija metoda za analizu signala, koje su inspirisane vizuelnim prikazom signala, se komplikuje u slučaju da informacija nije jednostavno uočljiva u vremenskom domenu. U tom slučaju se primenjuju odgovarajuće transformacije signala. Na primeru EKG signala, te transformacije bi bile korišćenje transformisanog skupa podataka (prikaz varijabilnosti srčanog ritma računanjem isključivo RR intervala) ili prelazak u frekvencijski domen (ili vremensko-frekvencijski domen) da bi se proučilo poreklo šuma. Klasičan primer transformacije signala sa ciljem uočavanja promena na signalu, prelaskom u frekvencijski domen, je uočavanje zamora na EMG signalu, a primer prelaska u vremensko-frekvencijski domen je uočavanje smena ritmova na EEG signalu. Nakon odabrane metode transformacije biosignala, dalja analiza se nastavlja ili u transformisanom skupu informacija ili povratkom u vremenski domen. U slučaju analize varijabilnosti srčanog ritma ili u slučaju ocene zamora, analizu je pogodno nastaviti u novom prostoru stanja. Međutim, treba biti oprezan sa vremensko-frekvencijskom reprezentacijom signala. Naime, u literaturi se mogu naći primeri analize spektrograma EEG signala primenom metoda za obradu slike³⁰⁸, ali takav pristup, iako daje validne naučne rezultate, nije opravdan. Tri su razloga da studenti ne primenjuju ovu praksu: (1) informacija od interesa može se izdvojiti u vremenskom domenu (uključuje filtriranje pojedinih ritmova EEG signala i nakon toga primenu praga za određivanje početka i kraja pojedinog ritma) [438], (2) da bi se signal predstavio u vremensko-frekvencijskom domenu informacija od interesa je smanjena tj. vremenska rezolucija je smanjena na račun frekvencijske i obrnuto [395] i (3) obrada slike, koja bi morala da se primeni nad spektrogramom EEG signala, je računarski zahtevna i dodatno onemogućava primenu u realnom vremenu, ako za tim postoji potreba.

Najteži zadatak za biomedicinske inženjere se javlja u slučaju da je informacija sakrivena. Termin skriveno treba uzeti sa rezervom, jer je svaka informacija sakrivena dok se ne pronađe uoči (vizuelno ili primenom računarske metode za

308 Kako cilj udžbenika nije omaložavanje bilo čijeg rada, ovde neće biti citirani primeri takvih naučnih publikacija.

automatsku detekciju). Za izdvajanje tzv. “skrivenih” informacija najčešće su u upotrebi nestandardne metode kao što su metode za ocenu nelinearnih dinamičkih sistema [377, 439-440] ili primenom metoda veštačke inteligencije [441]. Takođe, dolazak do skrivene informacije je povezan sa relativno velikim skupovima podataka koji su arhivirani u dužim vremenskim periodima na relativno velikom broju pacijenata. Primer takve baze podataka je MIMIC-III baza (od eng. *Medical Information Mart for Intensive Care*) koja sadrži podatke pacijenata sa intenzivne nege i koja sadrži relativno veliki broj podataka (elektrofiziološki signali, rezultati laboratorijskih testova i dr.) dobijenih na preko 60000 pacijenata [134, 442]. Korisnu informaciju, na tako velikom skupu podataka, nije jednostavno vizuelno predstaviti, pa se zato ona smatra “skrivenom”. U analizi skrivene informacije, ne sme se izgubiti iz vida činjenica da se informacija već izmerenom signalu ne može dodati, već isključivo oduzeti u procesu analize. Sa ciljem ilustracije značaja vizuelizacionih metoda, u sledećem potpoglavlju dati su primeri znamenitih istorijskih grafika, a studenti mogu sami da odrede na pokazanim primerima koja je skrivena informacija prikazana.

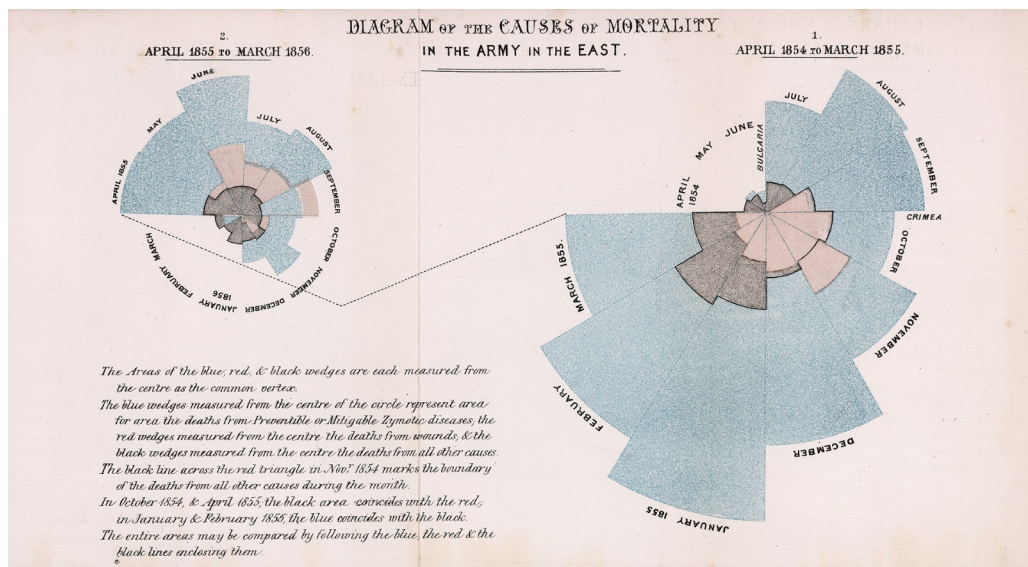
4.1.1 Istorijski primeri vizuelizacije podataka

U ovom potpoglavlju odabrana su tri istorijska primera primene vizuelizacije za rešavanje praktičnog problema. Prvi primer je vezan za Florens Najtingejl³⁰⁹ koja je bila poznata i kao “Dama sa lampom”³¹⁰. Florens je odigrala značajnu ulogu u modernizaciji medicine i primeni metoda zasnovanih na dokazima u medicinskoj praksi. Jedna od njenih zasluga jeste uvođenje reformi u sanitetskoj službi i nezi pacijenata u bolnicama. Mere koje je uvela za unapređenje higijene u vojnim bolnicama u Britaniji tokom Krimskog rata doprinele su smanjenju smrtnosti. Da bi ilustrovala svoj doprinos smanjenoj smrtnosti pacijenata, usled bolje higijene (u februaru 1855. godine je u jednoj britanskoj bolnici u Turskoj u kojoj je uvela radikalne mere (smrtnost je bila 43% da bi se smanjila na 2% u junu iste godine), ona je podnela stotine stranica izveštaja Kraljevskoj komisiji za zdravlje vojske (eng. *Royal Commission on the Health of the Army*) koji su pored teksta i tabela sadržali dijagrame u boji. Grafik koji je pripremila/osmislila za tu priliku, da na odgovarajući način predstavi informaciju, naziva se danas Najtingejlin diagram ruže (eng.

309 https://en.wikipedia.org/wiki/Florence_Nightingale, pristupljeno 01.05.2024.

310 Na engleskom su Florens Najtingejl zvali *The Lady with the Lamp*, jer je noću obilazila ranjenike u bolnicama noseći lampu [443]. Iako su slični, nadimak *Lady in the Dark* nije posvećen Florens Najtingejl, već je Dama u mraku naziv pesme hevi metal benda Sabaton koja je posvećena Milunki Savić, srpskoj junakinji i ženi sa najviše vojnih odlikovanja (https://sr.wikipedia.org/wiki/Milunka_Savić, pristupljeno 11.05.2024).

Nightingale rose diagram), cirkularni histogram (eng. *circular histogram*) ili polarni dijagram (eng. *polar area diagram*)³¹¹. Ovakva prezentacija rezultata (Sl. 146) doprinela je da Kraljevska komisija uvede izmene koje su omogućile da se osnuje Kraljevska medicinska škola i uvede redovno praćenje parametara koji utiču na zdravlje pacijenata. Florensin doprinos medicini zasnovanoj na dokazima je veliki u istoriji medicine, a ne samo britanske vojske. Kružni dijagrami koje je Florens osmislila imali su 12 delova u kojima je prikazana smrtnost pacijenata, a boja je iskorišćena da se kodira uzrok smrti. [141]



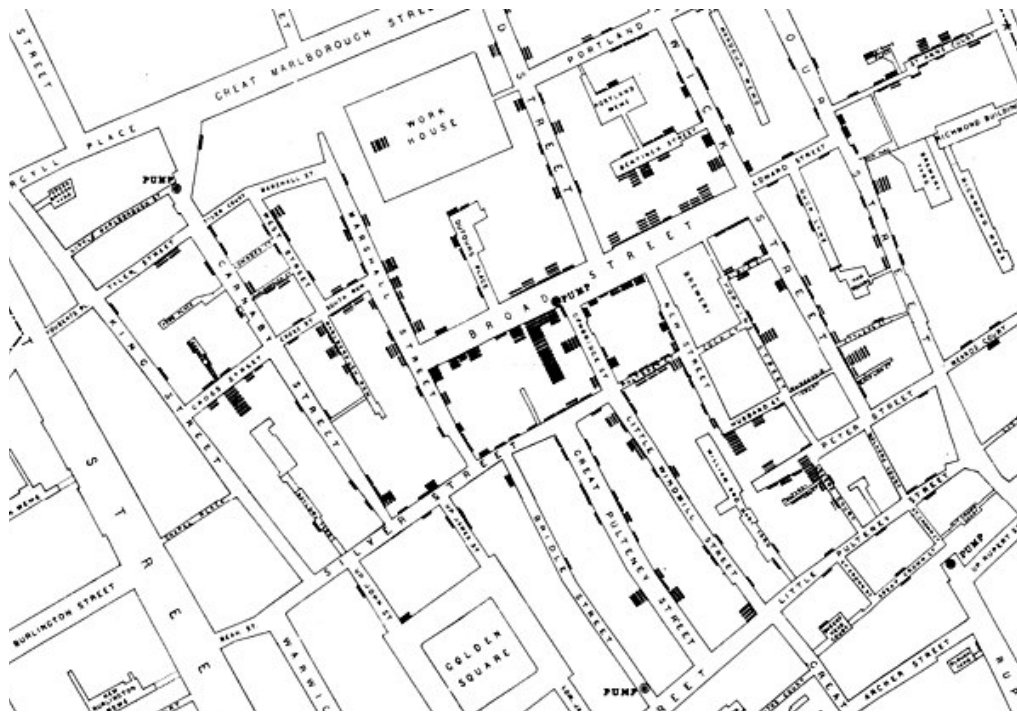
Slika 146, Dijagram uzroka mortaliteta u vojnim bolnicama koji je kreirala Florens Najtingejl, <https://www.davidrumsey.com/luna/servlet/s/h6xid2>, Javno vlasništvo, <https://commons.wikimedia.org/w/index.php?curid=1474443>.

Drugi popularan primer, uključuje prikaz obolelih od kolere u Londonu koji je napravio Džona Snou³¹², uz pretpostavku da je epidemiju kolere izazvala kontaminirana voda. Da bi validnim činjenicama potkrepio svoju sumnju, ucrtao je u mapu Londona (Sl. 147) blokove koji su označavali broj smrti i označio mesta svih česama. Vrlo brzo je uočeno da su sve smrti bile povezane sa česmom u Brod ulici (eng. *Broad Street*). Detaljnom istragom, Džon Snou je čak utvrdio da je ista česma bila razlog i udaljenijih smrti, a da stanari pojedinih stambenih objekata u blizini česme nisu oboleli od kolere, jer su koristili druge izvore vode. Ovaj grafik je omogućio Snou da ubedi vlasti da zatvore česmu, kao i da se u naučnim i medicinskim krugovima napusti uverenje da kolere izaziva “loš vazduh”. Iako je bio poznat akušer i jedan od prvih lekara koji su

311 https://en.wikipedia.org/wiki/Pie_chart#Polar_area_diagram, pristupljeno 01.05.2024.

312 https://en.wikipedia.org/wiki/John_Snow, pristupljeno 01.05.2024.

primenjivali anesteziju, Snou je danas primarno upamćen po svom doprinosu epidemiologiji. [141]

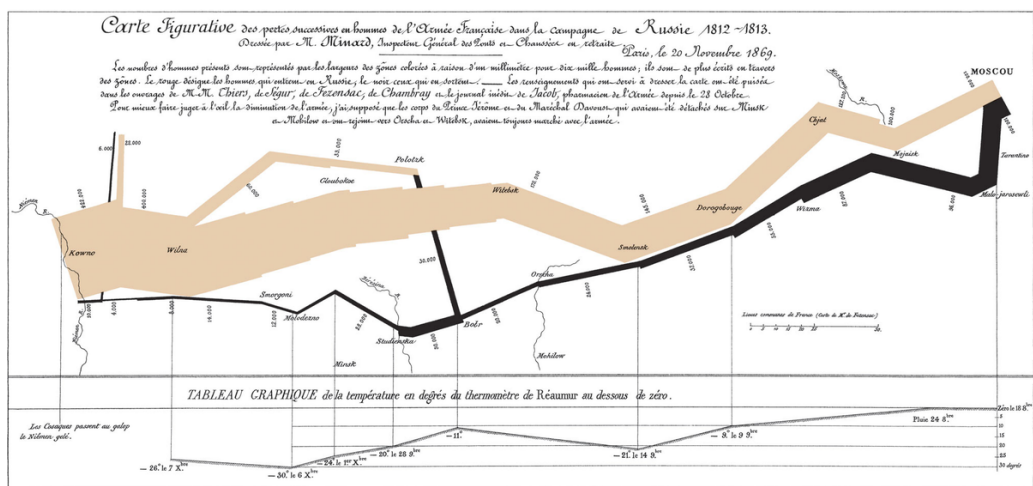


Slika 147, Deo mape koju je kreirao Džon Snou 1854. godine. Česme su označene nazivom PUMP dok je broj smrtnih slučajeva ucrtan blokovima crne boje. Primetno je da je najveći broj smrti u okolini česme u ulici označenoj sa *Broad Street* u kojoj se nalazila česma sa zaraženom vodom. Modifikovana originalna slika koju je ilustrovao Čefins i koja je nakadno digitalizovana, Javno vlasništvo, <https://commons.wikimedia.org/w/index.php?curid=2278605>.

Treći primer je proglašen, prema nekim izvorima, za najbolji grafik svih vremena, a predstavlja mapu Napoleonovog vojnog pohoda koju je kreirao Šarl Minar³¹³ [434]. Ovaj kompliment je dodatno upotpunjen činjenicom da je grafik nastao je u vreme kada računari nisu ni postojali, te verovatno iz tog razloga izaziva divljenje svih koji se bave naukom o podacima danas. Radi se o prikazu Napolenovog vojnog pohoda na Rusiju iz 1812. i 1813. godine (Sl. 148) koji se završio katastrofalnim porazom. Mapa je najviše cenjena zbog svedenog prikaza i zbog velikog broja informacija koje su prikazane u 2D prostoru. Svedenost se najviše ogleda u tome što je za detalje korišćen samo potreban minimum (iako grafik za pozadinu ima geografsku kartu, ona je vrlo štura, jer su istaknuti samo najznačajniji detalji kao što je Moskva – najudaljenija geografsko mesto do kog je Napoleon sa svojom vojskom stigao). Informacije

313 https://en.wikipedia.org/wiki/Charles_Joseph_Minard, pristupljeno 01.05.2024.

koje su prikazane na Sl. 148 uključuju: (1) broj vojnika (tokom ofanzive/napada je prikazan bež bojom, dok je tokom defanzive/povlačenja prikazan crnom bojom, a na pojedinim mestima na mapi postavljene su i kvantitativne procene na osnovu kojih se vidi da je Napoleon u pohod poveo 422000 vojnika, a vratio se sa 10000 vojnika), (2) geografsku udaljenost i mesta kroz koja je prošla vojska na šturoj geografskoj mapi, (3) meteorološke uslove tokom povlačenja vojske u zimskim mesecima, odnosno dijagram promene temperature tokom povlačenja u donjem delu grafika na Sl. 148 (povlačenje vojske je prikazano crnom bojom), (4) geografsku širinu i geografsku dužinu, (5) pravac kretanja vojske (bež bojom je određen pravac napredovanja vojske ka Moskvi, a crnom bojom pravac povlačenja) i (6) pojedine ucrtane lokacije su povezane sa datumima izdvojenih događaja tokom pohoda na Rusiju (datumi su vidljivi pored temperatura na dijagramu temperature u donjem delu grafika). [44]



Slika 148, Mapa koju je kreirao Šarl Minar i koja opisuje Napoleonov neuspešni vojni pohod na Rusiju 1812. i 1813. godine. Od Šarla Minara (1781–1870) – Javno vlasništvo, <https://commons.wikimedia.org/w/index.php?curid=297925>³¹⁴

Minarova mapa Napolenovog poraza u Rusiji izaziva još veće divljenje, ako se uzme u obzir koliki je napor Minar uložio da decenijama nakon vojnog pohoda prikupi podatke iz dostupnih izvora, jer je mapa nastala 1869. godine. 57 godina nakon početka pohoda (Sl. 148). Zanimljivo je istaći da se na mapi

314 Udžbenik pisan tako da se jednostavno čita u elektronskoj formi, ali se savetuje studentima da pogledaju mapu na većem ekranu u punoj rezoluciji kako bi uočili sve potrebne detalje. Moderna varijanta mape je prevedena na engleski jezik i dostupna na https://upload.wikimedia.org/wikipedia/commons/thumb/d/db/Redrawing_of_Minard%27s_Napoleon_map.svg/1920px-Redrawing_of_Minard%27s_Napoleon_map.svg.png, pristupljeno 01.05.2024.

moгу jednostavno uočiti detalji vojne taktike “spaljena zemlja”³¹⁵, ali i činjenica da, iako je došlo do jedne od najvećih bitaka u zapadnom svetu (bitka kod Borodina³¹⁶), to nije izazvalo značajne izmene u broju vojnika, kao što je to slučaj sa taktičkim nastupom ruske vojske u prvom delu pohoda [445]. Iako zanimljive, ove činjenice se ne mogu iskoristiti u vizuelizaciji biosignala, ali način prikaza može. Na primer, mapa ima za cilj da predstavi stradanje Napoleonove vojske i njegov neuspeh koji se ogleda u velikim gubicima tokom povlačenja vojske, te je povlačenje prikazano najdramatičnije na mapi – crnom bojom (Sl. 148). Svetle nijanse, bež boje na Sl. 148, su iskorišćene za važnu, ali ne tako dramatičnu informaciju o broju vojnika tokom ofanzive. To bi se moglo iskoristiti za poređenje signala merenih sa dva uređaja – na primer, ako se ocenjuje uređaj B, a kao referentan (eng. *ground truth*) se koristi uređaj A, onda ima smisla napraviti takav grafik u kome bi signali sa uređaja B bili prikazani jarkim bojama, a signali sa uređaja A (koji samo služi kao referentna vrednost) svetlim bojama koje se diskretno razlikuju od pozadine grafika.

Odabrani istorijski primeri nisu u direktnoj vezi sa biosignalima, ali predstavljene strategije vizuelizacije se mogu primeniti na biosignale, na izdvojena obeležja iz biosignala i na biomedicinske podatke. U primeru vizuelizacije Florens Najtingejl, to je prikaz pre i posle događaja od interesa (izmene u broju smrtnih slučajeva nakon uvođenja higijene u bolnicama), te se sličan princip može primeniti u slučaju analize EGG signala za ocenu mučnine (prikaz parametara/biosignala tokom relaksacije i tokom korišćenja proizvoda virtuelne realnosti (eng. *virtual reality*, skraćeno VR) kada se očekuje promene na signalu uzrokovane pojavom mučnine kod ispitanika). U slučaju grafičke reprezentacije Džona Snoa, to je uključivanje demografskih parametara ili drugih značajnih faktora, koji nisu inicijalno zastupljeni u originalnom skupu podataka niti su bili uključeni u protokol. U prikazu epidemiološko-geografske mape Londona 19. veka, ključni element je bio lokacija česme, a u biomedicinskom inženjerstvu se može desiti da je razlika koja se uočava posledica pola ili nekih drugih anatomskih ili fizioloških karakteristika ispitanika (puls je kod žena veći nego kod muškaraca, razlike između polova postoje i u kapacitetu pluća). Treći primer, iako nema veze sa zdravstvom, ukazuje na mogućnosti koje pruža 2D prikaz za predstavljanje relativno velikog broja informacija. Posebno treba istaći značaj trećeg primera u oblasti koja je nedavno dobila na popularnosti i odnosi se na pripovedanje uz pomoć podataka (eng. *data storytelling*³¹⁷). Autorka se nada da će studenti, uz odgovarajuće funkcije i pakete u R-u, koji su

315 https://en.wikipedia.org/wiki/Scorched_earth, pristupljeno 02.05.2024.

316 https://en.wikipedia.org/wiki/Battle_of_Borodino, pristupljeno 02.05.2024.

317 https://en.wikipedia.org/wiki/Data_and_information_visualization, pristupljeno 02.05.2024.

prikazani u narednim poglavljima, biti u mogućnosti da kreiraju odgovarajuće vizuelizacije biosignala.

Istorijskih primera vizuelizacije je bilo i na Elektrotehničkom fakultetu, Univerziteta u Beogradu. U 2018. godini, sasvim slučajno je pronađen (prof. Milan Bjelica je pronašao infografike), iza jednog ormara u kabinetu Fakulteta, grafik dijagrama struja u automatskoj telefonskoj centrali u Beogradu koji su zabeleženi neposredno pre i nakon ubistva kralja Aleksandra Karađorđevića u Marselju u Francuskoj 1934. godine. Izložbu pronađenih infografika su u 2018. godini organizovali PTT muzej i Fakultet tokom PSSOH konferencije. Poster koji je za tu priliku pripremila studentkinja Mina Petrović sa fotografijama grafika koje je napravio prof. Miodrag Tasić su dostupni na Zenodo repozitorijumu [446].

4.1.2 Osnovne funkcije za prikaz grafika u R-u

Osnovna funkcija u programskom jeziku R koja se koristi za prikaz podataka je:

plot().

Za razliku od Matlab/GNU Octave programskih jezika u kojima *plot()* funkcija odgovara tzv. kontinualanom prikazu podataka (misli se na uslovno kontinualan prikaz, jer se na računaru analiziraju i prikazuju isključivo digitalni signali), ova funkcija prikazuje podatke diskretno (po ugledu na *stem()* funkciju u Matlab/GNU Octave programskim jezicima). Za kontinualni prikaz digitalnih podataka, potrebno je parametru funkcije *type* dodeliti vrednost "1" od eng. *line* što označava linijski prikaz. Da ne bude zabune oko uslovno kontinualnog prikaza, jer funkcije za takav prikaz signala ne koriste npr. metode polinomijalnog fitovanja, već se pojedini odbirci signala spajaju pravim linijama (što se može videti ako se dovoljno uveća grafik). Pored tipa linije, ulazni argumenti *plot()* funkcije mogu služiti da se označe ose, doda naslov grafika, definišu granice osa, odabere tip markera, boje, debljina linije, tip linije i dr.

U "primerPodataka.csv"³¹⁸ datoteci se nalaze simulirani signali koji odgovaraju ROM podacima na skočnom zglobo i u kolenu. Primenom jednostavnih i ugrađenih R funkcija (Sl. 149), moguće je prikazati na apscisi (tzv. "x-osi") redni broj merenja, a na ordinati (tzv. "y-osi") simulirani ugao učitani iz prve kolone datoteke "primerPodataka.csv". Za dodavanje mreže tj. *grid*-a, potrebno

318 Podaci su dostupni u GitHub repozitorijumu koji je podeljen uz udžbenik.

je koristiti posebnu istoimenu funkciju *grid()*. Dobra praksa nalaže da *grid* mora biti svetlije boje i/ili imate tanje linije od linije koja se koristi za prikaz signala, jer informacija koju nosi *grid* nije od značaja za informaciju koju sadrži signal. Uloga takve mreže je da opsezi i lokalne vrednosti budu uočljiviji. Na Sl. 149 je prikazano korišćenje podrazumevane mreže.

```
dat <- read.table("primerPodataka.csv", sep = ";")
head(dat)
redniBr <- 1:length(dat$V1)

plot(redniBr, dat$V1,
      xlab = "ID", ylab = "ugao u stepenima",
      main = "Uglovi (simulirani podaci)")
grid()
```

Slika 149, Kratka skripta koja ilustruje primenu osnovnih R funkcija za prikaz simuliranih uglova iz prve kolone učitanih podataka. Rezultat prikaza grafika dat je na Sl. 150.

Kako bi se prikazao veći broj signala na jednom grafiku koristi se funkcija:

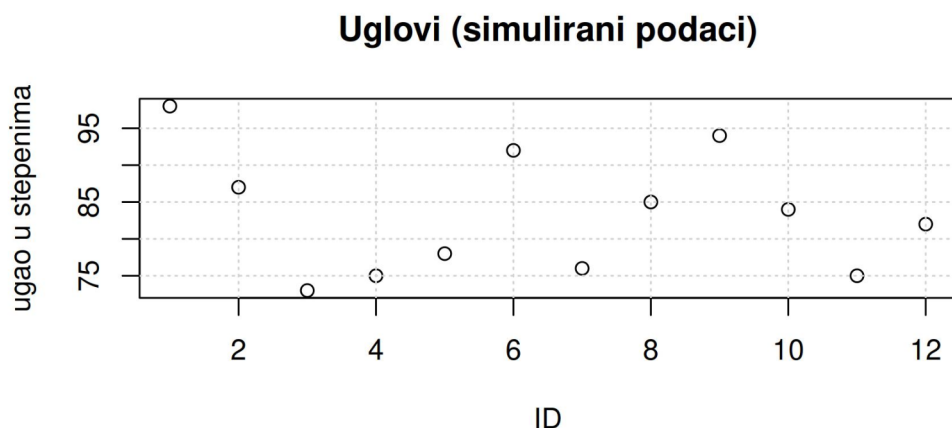
par()

a za dodavanje linija koristi se funkcija *abline()* čija je upotreba prikazana u ranijim poglavljima.

Na Sl. 151 je prikazana skripta koja omogućava da se prikažu podaci iz celog *data frame*-a koji je učitani (dve kolone V1 i V2 koje prikazuju simulirane uglove u skočnom zglobu i kolenu) na odgovarajući način (diskretno i kontinulano) i da se dodaju granice za poređenje signala, kao i legenda koja označava na kom zglobu su mereni ROM signali. Kako podrazumevani grafici u R-u imaju bespotreban prazan prostor, jedan od načina da se prostor popuni korisnom informacijom jeste da se koristi *par()* funkcija za podešavanje margina (prvi red u skripti na Sl. 151 sadrži funkciju kojom se podešavaju donja, leva, gornja i desna marginu). Kako bi se četiri grafika prikazalo na jednom dijagramu, potrebno je:

- koristiti *par(new = TRUE)* što odgovara *hold on* komandi u Matlab/GNU Octave programskim jezicima – primetiti da kod osnovnih R funkcija postoji potreba da za pozivom ove funkcije pre svakog (osim prvog) grafika koji se dodaje i

- potrebno je da se nazivi osa (*xlab* i *ylab* promenljive unutar *plot()* funkcije), naslov grafika (*main* argument *plot()* funkcije) i granice za apsiscu i ordinatu (*xlim* i *ylim* argumenti *plot()* funkcije) pozivaju kod svakog grafika koji se prikazuje na osi.



Slika 150, Prikaz simuliranih uglova u kolenu za različite ispitanika. Grafik je dobijen primenom jednostavnih R funkcija, a kod koji omogućava reprodukciju grafika je prikazan na Sl. 149. Na GitHub je dostupan i deo koda za čuvanje slike u željenoj rezoluciji. Ovde, radi jednostavnosti, nije prikazan.

Ovo je bespotrebno ponavljanje linija koda, ali da bi se izbeglo ponavljanje, potrebno je koristiti specijalizovane pakete za prikaz signala. Za prikaz legende, koristi se istoimena funkcija:

legend()

koja nema mogućnost automatskog prepoznavanja odabranih boja u okviru *plot()* funkcija, već se moraju ručno uneti sledeći argumenti:

- lokacija agende (na Sl. 151 je odabrano da bude u gornjem desnom uglu, ali pored automatskog podešavanja lokacije kroz stringove, moguće je uneti i koordinate),
- nizovi (dobijeni primenom *c()* funkcije) za natpis na legendi i odabrane boje,
- tip linije koji će se prikazati (*lty* argument od eng. *line type*) – za broj 1 to je standardna linija, dok je broj 2 isprekidana (pogledati uputstvo u R-u za druge tipove linija),

- debljina linije (*lwd* argument od eng. *line width*) – 1 je podrazumevano, a za vrednosti veće od 1 linije su punije (koliko punije zavisi od računara) i
- argument *cex* koji definiše uvećanje ili smanjenje zapisa (ovde je izabran tako da su slova u legendi 20% manja od slova na osama).

```

par(mar = c(4.2, 4.2, 1.7, 1.7))
plot(redniBr, dat$V1,
      xlab = "ID", ylab = "ugao u stepenima", main = "ROM",
      xlim = c(1, 12), ylim = c(60, 140))
par(new=TRUE)
plot(redniBr, dat$V1, type = "l", lwd = 2,
      xlab = "ID", ylab = "ugao u stepenima", main = "ROM",
      xlim = c(1, 12), ylim = c(60, 140))
par(new = TRUE)
plot(redniBr, dat$V2, col = "cyan",
      xlab = "ID", ylab = "ugao u stepenima", main = "ROM",
      xlim = c(1, 12), ylim = c(60, 140))
par(new=TRUE)
plot(redniBr, dat$V2, type = "l", col = "cyan", lwd = 2,
      xlab = "ID", ylab = "ugao u stepenima", main = "ROM",
      xlim = c(1, 12), ylim = c(60, 140))
grid()
legend("topright", c('skočni zglob', 'koleno'),
      col=c("black", "cyan"), lty = 1, cex = 0.8, lwd = 2)
abline(h = 120, col = "black", lwd = 2, lty = 2)

```

Slika 151, Skripta za odgovarajući prikaz simuliranih ROM podataka. Rezultat je prikazan na Sl. 152.

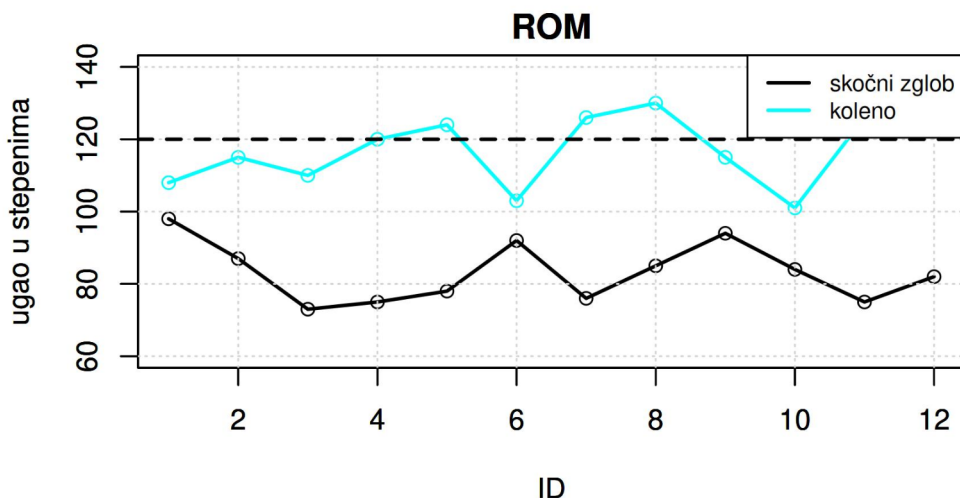
Da bi se dodala linija na grafik, moguće je koristiti *abline()* funkciju koja pored klasičnih elemenata za boju, tip i debljinu linije, sadrži i argument *h* (od eng. *horizontal*) ili argument *v* (od eng. *vertical*) za dodavanje horizontalnih i vertikalnih linija. Primetiti na Sl. 151 da za *abline()* funkciju nije potrebno dodati *par(new=TRUE)*, jer se linija automatski dodaje na grafiku koji je prethodno iscrtan. Ako postoji potreba da se prikaže više grafičkih elemenata na jednoj slici, moguće je koristiti *mfrow* argument *par()* funkcije kako je to pokazano u ranijim poglavljima.

Argument *col* (od eng. *color* ili *colour*, u zavisnosti da li je zapis na britanskom ili američkom engleskom jeziku) unutar *plot()*, *legend()* i *abline()* funkcija na

Sl. 151 služi da se podesi odgovarajuća boja na grafiku. Bojama se može pristupiti preko RGB šeme³¹⁹ (eng. *Red Green Blue*, odnosno crvena, zelena i plava) ili preko imena boje unosom odgovarajućeg stringa (kao što je prikazano za stringove “cyan” i “black” na Sl. 151). Da bi se dobio spisak naziva boja u R-u potrebno je u konzoli pozvati funkciju:

```
colors()
```

a da bi se izlistali simboli kojima je moguće označiti tačke na grafiku (na Sl. 151 su prikazani podrazumevani kružići) potrebno je u Help kartici potražiti “pch”. Dobra praksa nalaže da se boje štede, tj. da se informacija koja može biti kodovana na neki drugi način ne koduje bojama bez preke potrebe. Na Sl. 151 je bilo moguće koristiti promenu tipa linije i tipa tačke, odnosno kružić zameniti nekim drugim simbolom kako bi se izbeglo korišćenje boje.



Slika 152, Prikaz simuliranih ROM podataka – skripta koja omogućava prikaz ovog grafika data je na Sl. 151.

Iz primera vizuelizacije Napoleonovog pohoda na Rusiju, može se videti kako je svaki element iskorišćen za informaciju. Prema tome, ako se prikazuju dva grafika, nijanse sive će biti sasvim dovoljne da se među njima uočila razlika, a boja se može sačuvati za neku drugu informaciju (na primer, za prikaz podatka da li su signali mereni na zdravim ispitanicima ili na pacijentima). Takođe, informaciju ne treba ponavljati. Odnosno, bilo bi bespotrebno na Sl. 152 prikazati ugao na kolenu istovremeno isprekidanom linijom i drugom bojom (cijan). Informacije je potrebno istaći, ali ih ne treba ponavljati, jer se na taj

319 https://en.wikipedia.org/wiki/RGB_color_model, pristupljeno 02.05.2024.

način bespotrebno troše resursi koji se mogu iskoristiti za drugi tip informacije. Klasičan primer je bar grafik u kome je na apscisi napisano šta koji bar prikazuje, ali koji istovremeno ima različite obojene barove. Dodatno, oblike tačaka na grafiku treba birati primereno. Različiti oblici dolaze sa različitim dimenzijama, te je tačkama koje treba prikazati tako da su uočljivije potrebno pridružiti relativno veće simbole.

Ako je ipak potrebno koristiti više boja na jednom grafiku, onda je važno odabrati boje koje daju dobre kontraste kada se štampaju u nijansama sive i boje koje, pored različitih nijansi, imaju i različitu osvetljenost. Školski primer loše odabranih boja su crvena i plava, koje se koriste kao podrazumevane boje i u R-u, Pajtonu, Matlabu, GNU Octave-u i drugim programskim jezicima. Takve kombinacije boja je potrebno izbegavati, ako postoji mogućnost da se grafik kasnije predstavi u nijansama sive. Dobra strana crveno-plave kombinacije jeste što većina ljudi koji su slepi za boje vide razliku, pa je dobro rešenje da se iskoristi svetlost, te da se na grafiku predstavi tamno crvena sa svetlo plavom bojom, ili obrnuto, ako se već koriste te dve boje. Dakle, dobar odabir boje ima i humanu dimenziju, jer je procenjeno da veliki broj ljudi ima neki oblik daltonizma, te postoji velika verovatnoća da onaj kome je grafik namenjen neće biti u mogućnosti da vidi boje koje su odabrane za prikaz signala [447].

Kako bi se pojednostavio odabir boja i kombinacija boja prof. Vei (Ying Wei³²⁰) sa Kolumbija Univerziteta je pripremila katalog R boja sa nazivima za sve nijanse, koji su za svaku preporuku, jer omogućavaju jednostavan odabir boja za grafičku prezentaciju³²¹. Da bi se ilustrovalo koliko su elementi slika važni, prema novoj direktivi Evropske unije (eng. *European Accessibility Act* broj 2019/882³²²) za knjige izdate od stranih izdavača, one se ne mogu prodavati ako nisu prilagođene osobama sa hendikepom tj. ako ne uključuju narative za slike (eng. *alt text* ili *alternative text*) koje se nalaze u knjigama (planirano je da direktiva stupi na snagu 28. juna 2025. godine), što je u skladu i sa međunarodnim standardom za pristup informacijama (eng. *Web Content Accessibility Guidelines*, skraćeno WCAG³²³). Uzimajući u obzir ove direktive. i te kako ima smisla koristiti odgovarajuće boje i imati na umu dostupnost/pristup grafika svima (što se odnosi i na narativ o korišćenim bojama). Do sada je pažljivim čitaocima postalo jasno da vizuelizacija signala nije samo par linija koda koje su nevažne u odnosu na vreme provedeno u

320 <https://yingweistat.com/>, pristupljeno 02.05.2024.

321 <http://www.stat.columbia.edu/~tzheng/files/Rcolor.pdf>, pristupljeno 02.05.2024.

322 <https://ec.europa.eu/social/main.jsp?catId=1202&intPageId=5581&langId=en>, pristupljeno 02.05.2024.

323 <https://www.w3.org/WAI/standards-guidelines/wcag/>, pristupljeno 02.05.2024.

razumevanju fizioloških pojava, protokola za merenje signala, čišćenju/filtriranju i analizi signala, već vizuelizaciji treba posvetiti vreme i obratiti pažnju na detalje. Da bi se dobio dobar grafik, nekad je potrebno više radnih dana posvetiti pažljivom odabiru svih elemenata i kreirati, po potrebi, više desetina verzija dok se ne dobije konačna verzija grafika. Pored najboljeg grafika na svetu (Napoleonov neuspešan pohod na Rusiju), Šarl Minar je uspeo da napravi još samo 50 grafika za života [448] i to većinu nakon odlaska u penziju, što je ubedljiva ilustracija koliko je vremena potrebno uložiti u vizuelizaciju.

4.1.3 Preporuke za vizuelizaciju biosignala

Vizuelizacija je proces transformacije podataka u vizuelnu informaciju koja omogućava da se u podacima ili obeležjima, koja su dobijena iz podataka, uoče korelacije, veze, anomalije ili obrasci. Usko je povezana sa statistikom, psihologijom i računarskom tehnikom. Jednostavno rečeno, vizuelizacija je proces grafičkog predstavljanja informacija kako bi se ostvatio uvid u povezanost, obrasce, sličnosti i razlike u podacima pomoću oblika, boja, pozicije i dimenzija. Da bi se došlo do odgovarajuće vizuelizacije, potrebno je očistiti podatke, formatirati podatke, ispitati ih, analizirati i evaluirati rezultat. [67]

Danas, upotreba programskih jezika za vizuelizaciju i računara za komunikaciju su doprineli da grafici ne moraju nužno da predstavljaju statički prikaz (ili da budu uramljeni na zidu – što je bila sudbina Minarovih grafika koji su korišćeni kao pozadina na portretima ministara u Francuskoj [449]), mogu biti dinamički, interaktivni i animirani. Računarstvo u oblaku (eng. *cloud computing*³²⁴) je takođe doprinelo sve većoj primeni vizuelizacije, jer više nema prepreka za pristup udaljenim računarskim resursima. Primeri takvih grafika mogu se naći na internetu [67]:

- Gugl alatka Ngram Viewer³²⁵ omogućava uvid u interaktivne dijagram i poređenje popularnosti ključnih reči u knjigama u vremenskim intervalima, a Gugl Trends³²⁶ omogućava da se pretraži i poredi zastupljenost pojedinih termina u različitim oblastima, određenim vremenskim intervalima i regijama,

324 https://en.wikipedia.org/wiki/Cloud_computing, pristupljeno 02.05.2024.

325 <https://books.google.com/ngrams/>, pristupljeno 02.05.2024.

326 <https://trends.google.com/trends/>, pristupljeno 02.05.2024.

- GapMinder³²⁷ je interaktivno grafičko okruženje za istraživanje statističkih geografskih historijskih podataka³²⁸ koje je dostupno pod otvorenom CC licencom,
- Jedan od primera je grafik Nju Jork Tajmsa koji prikazuje kako različite grupe ljudi provode vreme tokom dana u Americi³²⁹ [450].
- Primer iz Srbije je “Moj doktor”³³⁰ internet stranica Ministarstva zdravlja, Republike Srbije koja u realnom vremenu prikazuje (grafički i numerički), na svaku sekundu, broj zahteva, pregleda, uputa, kao i broj recepata po lekovima. Drugi primer iz Srbije predstavlja portal eNauka³³¹ koji pruža uvid u interaktivnu statistiku kvantitativnih pokazatelja naučnoistraživačkih rezultata istraživača u Republici Srbiji.

Za one koji žele da se bave interaktivnim veb prikazima grafika, preporuka je da pogledaju materijale koji su dostupni na Gugl stranici za programere³³², a koji uključuju programski interfejs za programiranje (eng. *Application Programming Interface*, skraćeno API³³³). Da bi se jednostavno pokazalo čemu API može da služi, odabran je primer grafika (Sl. 153) koji prikazuje popularnost ključne reči “virtual reality” u periodu od 1920. do 2019. godine u alatki Gugl Akademik (eng. *Google Scholar*³³⁴) [451].

Grafik prikazan na Sl. 153 je nastao tako što je ručno na Gugl Akademiku izvršena pretraga za odabranu ključnu reč i za opseg godina (po dve godine od 1920. do 2019. godine), a podaci o broju rezultata, koji se dobijaju ručnim podešavanjem parametara pretrage, su uneti ručno u .csv tabelu. Potom, ti podaci su prikazani u programskom jeziku R sa ciljem da se vidi kako je od početka 90ih godina 20. veka naglo poraslo interesovanje za vituelnu realnost (maksimalan broj rezultata je dat za period od 2010. do 2012. godine), a da je kasnije (posle 2012. godine) pojavljivanje ključne reči naglo opalo. U

327 GapMinder je neprofitna Fondacija (https://en.wikipedia.org/wiki/Gapminder_Foundation, pristupljeno 03.05.2024.) čiji je jedan od osnivača prof. Hans Rosling koji je održao čuveno predavanje u okviru TED (eng. *Technology, Entertainment, Design*) događaja u 2006. godini, https://www.ted.com/talks/hans_rosling_the_best_stats_you_ve_ever_seen?language=en, pristupljeno 03.05.2024.

328 <https://www.gapminder.org/tools>, pristupljeno 02.05.2024.

329 <https://archive.nytimes.com/www.nytimes.com/interactive/2009/07/31/business/20080801-metrics-graphic.html>, pristupljeno 02.05.2024.

330 <https://live.mojdoktor.gov.rs/>, pristupljeno 02.05.2024.

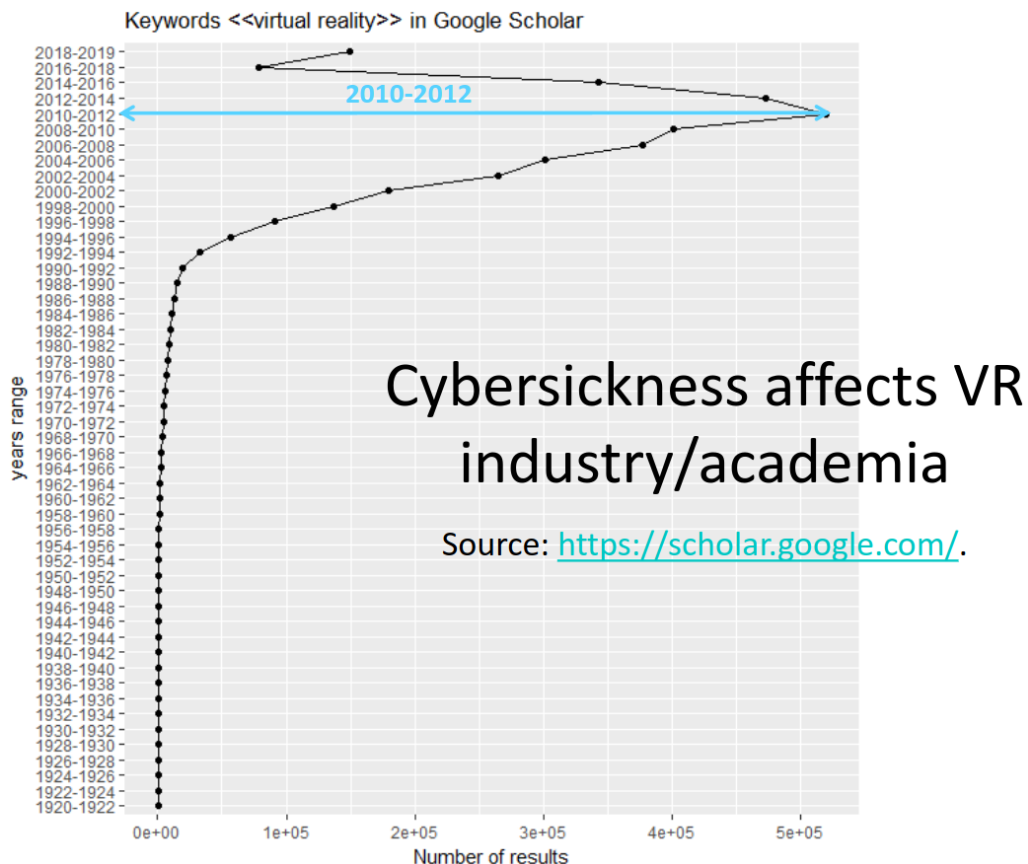
331 <https://enauka.gov.rs/>, pristupljeno 03.05.2024.

332 <https://developers.google.com/chart>, pristupljeno 02.05.2024.

333 <https://en.wikipedia.org/wiki/API>, pristupljeno 02.05.2024.

334 <https://scholar.google.com/>, pristupljeno 02.05.2024.

prezentaciji je pretpostavljeno da bi to mogla biti posledica mučnine koju izaziva virtualna realnost [451].



Slika 153, Prikaz popularnosti ključne reči “virtual reality” za predavanje po pozivu održano na Elektrotehničkom fakultetu, Univerziteta u Ljubljani. Prezentacija je podeljena pod otvorenom CC BY licencom, a grafik je pripremljen u programskom jeziku R. [451]

U vizuelizaciji nema eksplicitnih pravila, jer je ona na prvom mestu stvar stila³³⁵. Međutim, postoje preporuke iz prakse koje predstavljaju osnov za odgovarajući prikaz signala. Pored toga, što grafik treba da sadrži tačnu i preciznu informaciju, postoje dodatni elementi na koje treba obratiti pažnju. Netačna informacija bi bila kada bi se na grafiku prikazao EMG signal u opsegu od -5 V do 5 V što predstavlja izlaz iz A/D konvertora, ali ne odgovara signalu koji je pojačan 1000 puta, te je tačan prikaz EMG signala u opsegu od -5 mV do 5 mV. Neprecizan prikaz EMG signala bi bio da se ose ne obeleže ili

335 “O ukusima ne treba raspravljati” u originalu na lat. “*De gustibus non est disputandum*” [452].

da se vremenska osa predstavi u odbircima. Prikaz u odbircima je tačan, ali informacija o frekvenciji odabiranja nedostaje: za signal odabiran sa frekvencijom odabiranja od 1000 Hz, 2000 odbiraka predstavlja interval od 2 s, a za signal odabiran sa 2000 Hz, 2000 odbiraka predstavlja 1 s, što nije zanemarivo.

Vizuelizacija signala je važna u procesu filtriranja šuma (da bi se osiguralo da je filter dobro realizovan, da je šum isfiltriran, da bi se u frekventijskom i vremenskom domenu uočilo prisustvo šuma i sl). Ali, važna je i u procesu izdvajanja obeležja. Ako se R pikovi ne bi videli na EKG signalu, ni pre ni posle filtriranja, onda ne bi bilo moguće ni detektovati ih. Svaki stručnjak u obradi signala zna da se ne može odgovarajućim obeležjem kvantifikovati informacija, ako informacija prethodno nije bila vidljiva golim okom na grafiku. Tu se naravno javlja drugi problem – kako na najbolji način predstaviti informaciju? Na primer, ako se EKG signal predstavi u intervalu od 24 sata, onda je sigurno da se pojedinačni PQRST segmenti neće uočiti na grafiku koji je prikazan na standardnom računarskom monitoru. Dodatno, ako je informacija od interesa uočljiva u frekventijskom domenu, onda vizuelizacija u vremenskom domenu neće dati dobar rezultat i obrnuto. U nekim situacijama, analitičari, studenti i naučnici se susreću sa tzv. “skrivenim” informacijama koje nije jednostavno uočiti karakterističnim tehnikama za prikaz signala, kao što je ranije navedeno.

Preporuke za prikaz grafika su [453]:

- da se nikada ne koristi za prikaz signala kružni grafik (eng. *pie chart*³³⁶), jer može pogrešno da prikaže udeo u procentima³³⁷ koji se može prikazati preko bar grafika (takav grafik mora uvek imati granice od 0³³⁸ do 100%, tj. ne treba ga uvećavati, odnosno zumirati),
- da se ne koriste boje koje nemaju dobar kontrast, posebno ako postoji mogućnost da će se grafik štampati u nijansama sive,
- da se ne koriste boje koje ljudi koji imaju smetnje da uoče razlike u bojama (ovoga se, naravno, ne može niko uvek pridržavati, jer je standardna forma boja na semaforima zelena za prolaz i crvena za

336 https://en.wikipedia.org/wiki/Pie_chart, pristupljeno 02.05.2024.

337 Iako je grafik koji je Florense Najtingejl prikazala bio tip kružnog grafika, treba uzeti u obzir da ona nije koristila kružni prikaz za predstavljanje udela u procentima, već je koristila ravnomernu podelu kružnog grafika na 12 delova (12 meseci), te se njen grafik mora posmatrati na drugačiji način.

338 Dobra praksa je da se uz vrednost 0 ne piše jedinica, jer 0 predstavlja nepostojeću vrednost.

nedozvoljen prolaz³³⁹, a postoje grafici koji crpe inspiraciju iz sličnih konvencija),

- ne treba koristiti 3D prikaz, ako postoji 2D prikaz koji omogućava da se prikaže ista informacija, jer se ne može na papiru (forma koja se često koristi za grafik) dobro predstaviti treća dimenzija,
- ne treba koristiti boje za informaciju koja je već prikazana na neki drugi način i
- ne treba koristiti grafik koji sakriva ili delimično sakriva informaciju od interesa.

Dodatne preporuke za pisanje izveštaja, realizaciju sajta, pripremu grafika, pripremu prezentacija i predstavljanje grafika su [454]:

- najvažniji elementi bi trebalo da su na vrhu strane ili sa leve strane, jer većina ljudi tako čita tekst,
- detalje koje su važni treba istaći na odgovarajući način (u slučaju Minarove mape to su bile uočljive boje),
- pokazano je da ljudi čitaju u F formi, te je potrebno obratiti pažnju na početna slova redova,
- ako neki deo izveštaja, slajda ili grafika liči na logo ili reklamu (eng. *banner look*), postoji mogućnost da taj deo niko neće ni pročitati,
- obavezno je uključiti kratak pregled ključnih elemenata ili rezime (eng. *abstract* ili eng. *summary*) – na primer, nije dovoljno ispod slike na kojoj je prikazan EMG signal napisati samo “EMG signal”, mnogo bolje je “EMG signal koji je meren tokom ponovljenog pokreta ruke na mišiću *biceps brachii* za podizanje tega od 1 kg ...” ili slično,
- uvek je poželjno istaći konačan rezultat – najbolja situacija je kada je rezultat pozitivan te se može pisati “Metoda A ima uspešnost 68%, a nova metoda B uspešnost od 94%” ili ako to nije situacija “Nova metoda B ima manju uspešnost u poređenju sa metodom A (za 5%), ali se metoda B izvršava duplo brže na računaru” i
- iako je poželjno da prikaz i objašnjenje rezultata budu dopadljivi, oni na prvom mestu treba da budu tačni i profesionalni.

U [67] je nabrojano 10 osnovnih elemenata u vizuelizaciji podataka, od kojih su mnogi zasnovani na radu poznatih Autora u polju informacione vizuelizacije. Ovde su ti standardi preneti u skraćenom prikazu i sa kratkim osvrtom na metode u biomedicinskom inženjerstvu:

339 Daltonisti nisu u mogućnosti da razlikuju crvenu i zelenu boju [453].

- Format grafika je od značaja, jer može pomoći da se optimizuje kvalitet grafika na različitim platformama. Na primer, za grafike koji se štampaju bi trebalo proveriti kako bi izgledali u nijansama sive, a njihova poželjna rezolucija je oko 300 dpi (eng. *dots per inch*) dok su poželjne ekstenzije tiff, eps, psd ili pfd. Ovde se misli i na odnos stranica grafika, jer je odnos od 16:9 poželjni za šire ekrane, dok je za uže ekrane poželjan odnos 4:3.
- Odabir boja igra značajnu ulogu. Boje treba dodavati samo u slučaju da se ukazuje na razlike među podacima i na taj način donosi nova informacija prikazu. Dobra praksa nalaže da je potrebno izbegavati tematske i dekorativne elemente, kako na dijagramima tako i u prezentacijama. Preporuka je da se vodi računa i o kulturološkom značenju boja, kao i o tome da postoje osobe koje su slepe za boje, pa bi grafike generalno trebalo prilagoditi svima. Preporučuje se relativno visok kontrast koji bi i osobama sa oštećenim vidom omogućio da uoče informaciju na grafiku.
- Tekst i oznake na osama bi trebalo koristiti kao dodatnu informaciju za željeno razumevanje grafika. Nekada su ove oznake i deo tehničkog standarda. Recimo, nije primereno da se vremenske serije biosignala predstave sa apcismom u odbircima, kada je uobičajen i poželjan prikaz koji može pružiti dodatne informacije (na primer, da se omogući vizuelna procena srčanog ritma sa vremenskog dijagrama EKG signala). Potrebno je objasniti sve skraćenice u nazivu ispod slike (ako se slika nalazi u izveštaju) ili u uglu slike (ako je grafik napravljen kao zasebna celina).
- Čitljivost bi trebalo da je dobra, a postiže izborom odgovarajuće veličine, pravca i boje fonta. Posebno bi trebalo smanjiti upotrebu ili u potpunosti izbeći pisanje teksta pod uglom koji ne prati naslov grafika i/ili ostatak teksta. Tip fonta može takođe biti od značaja, te je za tekst na graficima i slikama poželjno koristiti sans-serifne fontove, kao što su Verdana, Ariel i Helvetika, jer je pokazano da se ovi tipovi fonta prijatniji za čitanje. U izveštajima bi veličina fonta trebalo da bude oko 12, a na slajdovima prezentacija bar 16.
- Inkrementi na osama bi trebalo da budu logički i poželjno ekvidistantni.
- Integritet podataka je vrlo važan element svakog grafika. Primer zloupotrebe podataka je kada se dve veličine (na primer, dva bara na grafiku) predstavljaju nesrazmerno (umesto da se fizički oslikava razlika od 20%, barovi se predstave tako da se oslikava razlika od 200%). Za takve grafike je definisan i faktor laži (eng. *lie factor*) koji bi u slučaju prikaza razlike od 20% kao 200% bio veći od 1 tačnije $200\% / 20\% = 10$ [434]. Takođe, treba izbegavati trodimenzionalni (3D) prikaz u

slučaju da se informacija može prikazati u dve dimenzije (2D). U zavisnosti od vrednosti, koje promenljive mogu fizički da uzimaju, potrebno je voditi računa o opsegu koji se prikazuje na ordinati.

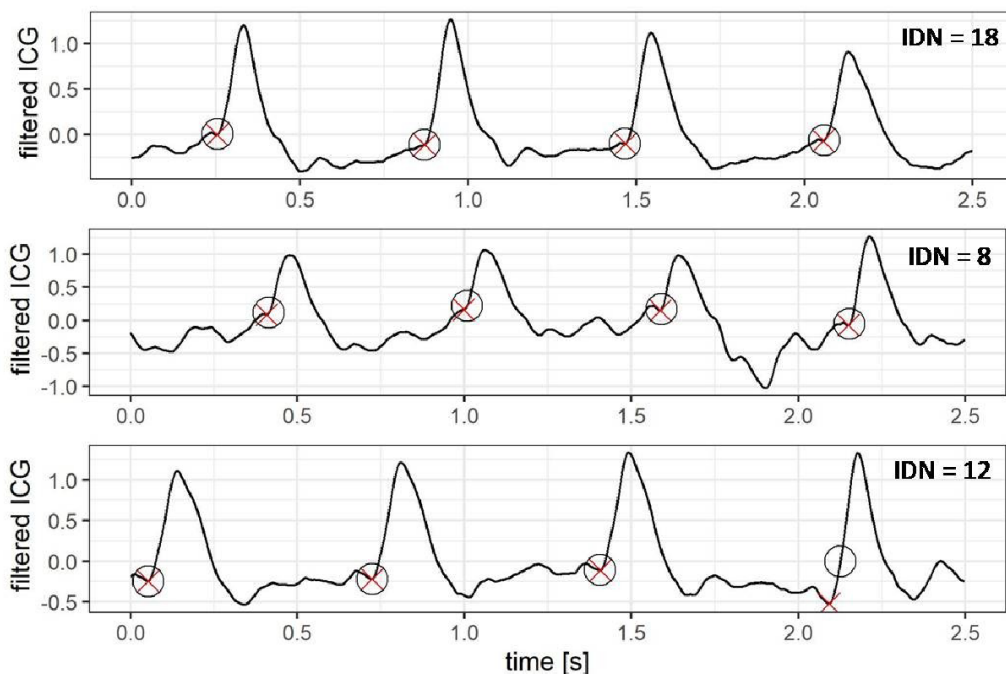
- Višak elemenata (eng. *chartjunk*) je potrebno obrisati sa grafika. Ovo se posebno odnosi na dekoracije (na primer, senka, boja ili dizajn pozadine grafika) i na *grid* koji se prikazuje u pozadini, a koji je previše istaknut. Višak elemenata se može odnositi i na tekst, te treba imati na umu pravila za kreiranje prezentacija, kao što je pravilo 1-6-6 koje glasi da treba, po jednom slajdu, imati jednu ideju, maksimalno 6 linija i optimalno 6 reči po liniji. Sa *grid*-om treba biti oprezan iz dva razloga. Prvi je što *grid* u nekim situacijama mora da postoji, a drugi ako već postoji trebalo bi da je svetlije boje od samog grafika te da odmah bude jasno posmatraču da se radi o pomoćnom elementu koji sa sobom ne nosi nikakvu informaciju. Primer obaveznog *grid*-a u medicini je prikaz EKG signala na milimetarskom papiru (pogledati Sl. 109 i Sl. 112).
- Količina podataka koja se prikazuje na grafiku bi trebalo da bude ograničena. Jedna od loših praksi je da se prikaže previše elemenata na grafiku, tj. previše informacija/podataka koje mogu da pokvare preglednost i čitljivost.
- Sa druge strane, za razliku od prethodnog standarda, treba voditi računa da se ne smanjuje količina prikazanih podataka, posebno u slučaju kada se informacija menja. Na taj način, treba obezbediti kvalitet informacije kroz najmanju, ali neophodnu količinu podataka koja se prikazuje.
- Izvor podataka bi trebalo uvek navesti/citirati. Ovde se ne misli na preuzimanje slika uz dozvolu ili slika koje poseduju odgovarajuće slobodne licence kao što su CC licence, već na izvor podataka koji je korišćen da bi se informacija predstavila na grafiku. Ta informacija može da stoji i u nazivu slike na kojoj je grafik predstavljen. Ovo se posebno odnosi na grafike koji su predstavljeni kao zasebne celine.

Preporuke iz odabranih izvora [67, 453-454], koje su predstavljene u ovom poglavlju, sadrže niz preklapljenih informacija, koje sigurno nisupromakle pažljivim čitaocima. Ova preklapanja bi trebalo da dodatno istaknu značaj vizuelizacije u obradi signala. Radi ilustracije, u narednom potpoglavlju prikazani su odabrani primeri dobre prakse u vizuelizaciji signala koji su u skladu sa navedenim preporukama.

4.1.3.1 Primeri dobre prakse u vizuelizaciji biosignala

Za najbolje grafike važi pravilo da ih treba planirati prema tome kakvu poruku treba da prikažu. Na primer, za potrebe predstavljanja nove metode za detekciju

karakteristične B tačke na EKG signalu, iskorišćena je grafička ilustracija uspešnosti predložene metode u odnosu na usvojenu tačnu metodu (eng. *ground truth*) – manuelnu detekciju B tačaka kao na Sl. 154 [99-100].

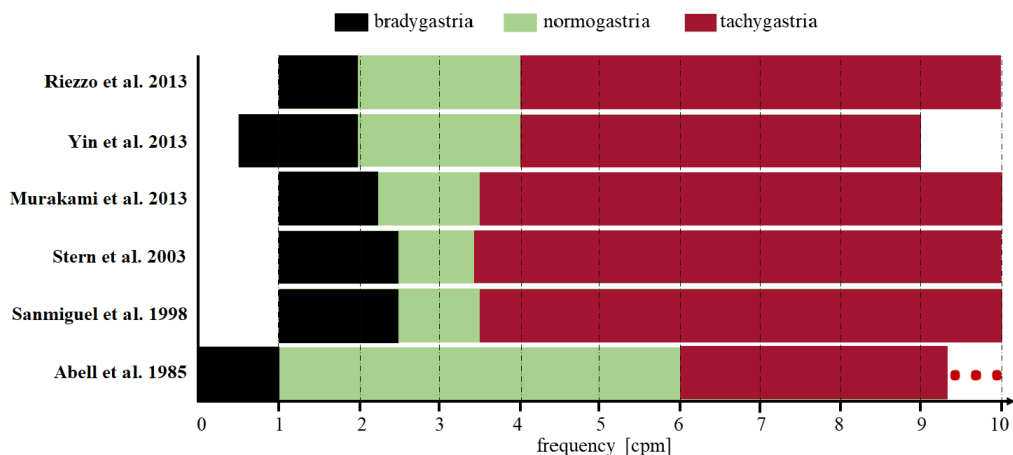


Slika 154, Prikaz IKG signala sa detektovanim karakterističnim B tačkama. Grafik je pripremljen u programskom jeziku R. Sa IDN označeni su različiti subjekti, kružićima su označene lokacije B tačaka na IKG signalu (eng. *Impedance Cardiogram*, skraćeno ICG), dok su crvenim krstićima označene lokacije koje je izdvojio automatski algoritam [99-100].

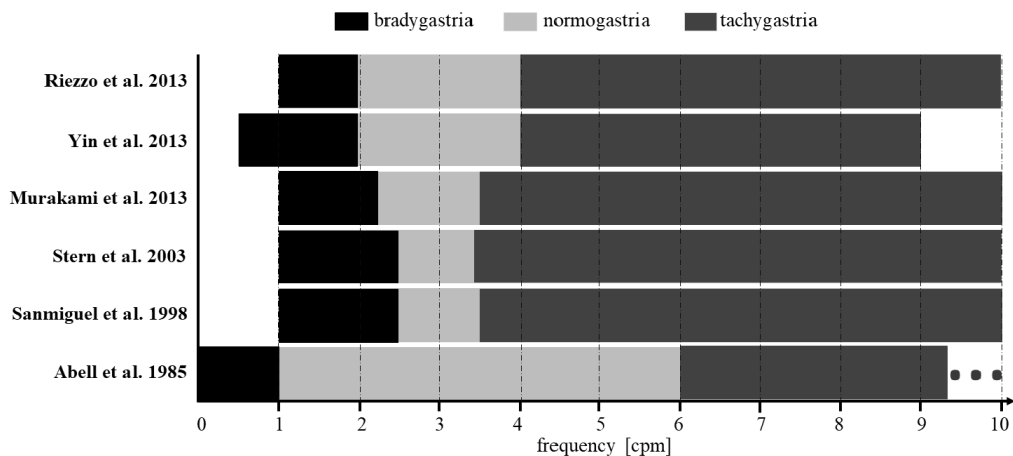
Za ilustraciju metode za detekciju B tačke na Sl. 154, izdvojeni su primeri IKG signala na tri ispitanika (IDN = 18, IDN = 8 i IDN = 12). Na svakom grafiku je prikazano samo četiri otkucaja srca (radi bolje preglednosti), te su tačne vrednosti lokacije B tačke označene kružićem, a lokacije koje je predloženi algoritam automatski detektovao predstavljene su crvenim krstićima. U slučaju da dolazi do poklapanja lokacije B tačke, crveni krstić bi trebalo da bude upisan unutar kružnice. Ovo je intuitivan primer vizuelizacije poređenja dve metode za detekciju karakterističnih tačaka na biosignalu.

Sledeći primer predstavlja pregled literature za određivanje frekvencijskih opsega različitih ritmova na EGG signalu (Sl. 155). Važan parametar na slici predstavlja odstupanje u granicama ritmova u različitim publikacijama koje je ilustrovano bar dijagramom i predstavljeno bojama. Spori ritmovi EGG signala su na Sl. 155 prikazani crnom bojom, normalni ritam svetlo zelenom bojom i

brzi ritmovi tamno crvenom bojom. Iako kombinacija zelene i crvene boje nije idealna, jer ljudi koji su slepi za boje ne mogu da razlikuju crvenu i zelenu, ali ako se ovakva slika predstavi u nijansama sive (Sl. 156) jasno je da je osvetljenost drugačija i da je moguće razlikovati tri ritma glatkih mišića želuca.

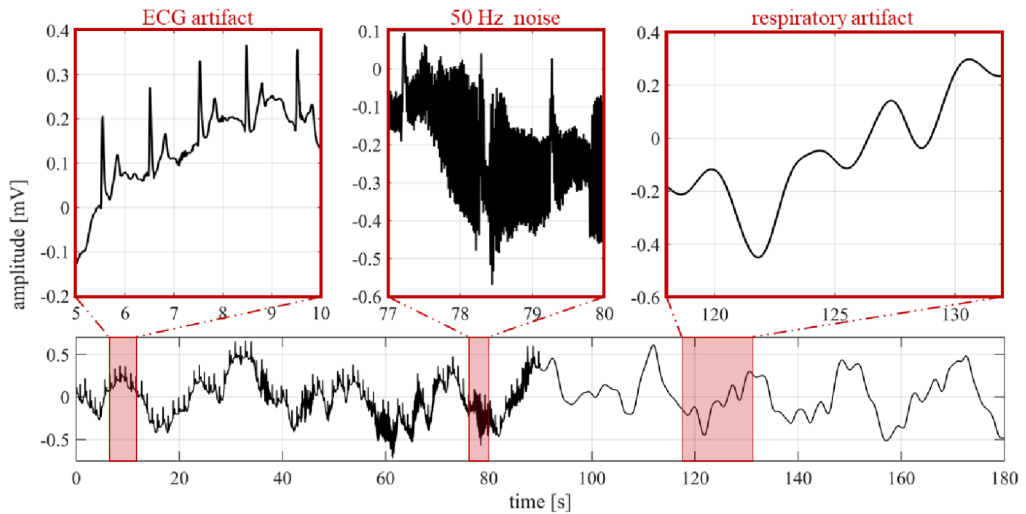


Slika 155, Poređenje različitih frekvencijskih opsega za tri ritma – spor, normalan i brz ritam kod EGG signala (nazivi ritmova na engleskom jeziku: *bradygastria*, *normogastria* i *tachygastria*) u literaturi. Slika je podeljena pod otvorenom licencom i dostupna u [266].



Slika 156, Grafik sa Sl. 155 prikazan u nijansama sive.

Iako grafik na Sl. 155 ne prikazuje izračunate parametre iz biosignala. činjenica da je na apscisi prikazana frekvencija i da su različiti ritmovi prikazani bojama, dok su na ordinati date odgovarajuće reference, čini da ovaj 2D grafik prikazuje 3D informaciju koja na vrlo efekatan način ukazuje na neslaganja u literaturi.

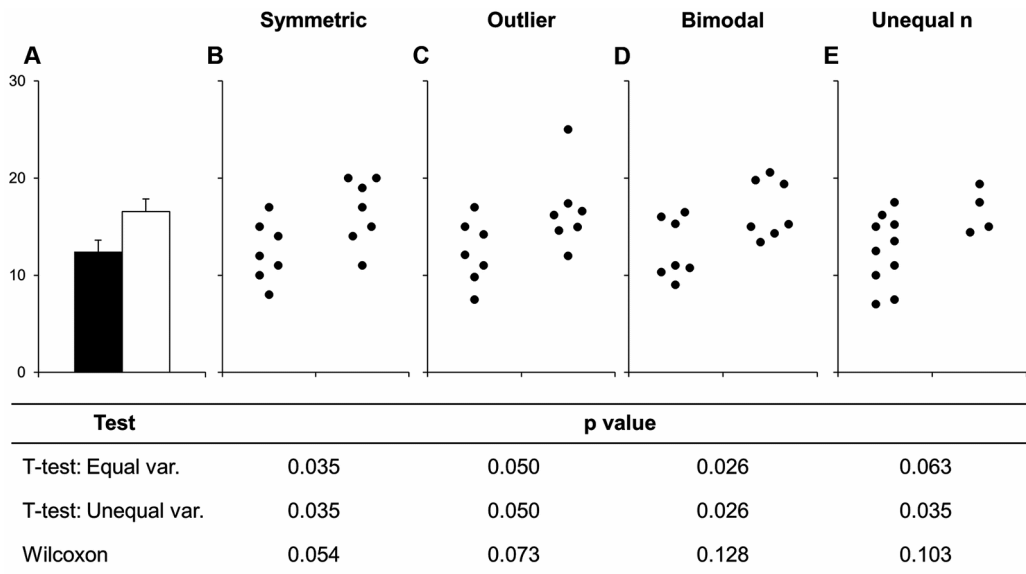


Slika 157, Prikaz tri različita šuma na EGG signalu koji su izdvojeni na gornjem panelu, a EGG signal u vremenskom domenu sa osenčenim segmentima je prikazan na donjem panelu. Slika je podeljena pod otvorenom licencom i dostupna u [266].

Na Sl. 157 je prikazan EGG signal u vremenskom domenu na donjem panelu, koji u pojedinim vremenskim trenucima (koji su osenčeni crvenom bojom) sadrži različite šumove. Kako bi se jasnije uočili šumovi, oni su izdvojeni u odabranim osenčenim opsezima i uvećani na tri gornja panela koja prikazuju s leva na desno: EKG šum, šum napajanja na 50 Hz i šum koji nastaje kao rezultat disanja. Najznačajnija informacija (šum) na Sl. 157 je prikazana na gornjem, a ne na donjem panelu. Ovakva predstava informacije odgovara prikazu na Minarovom dijagramu. Naime, na Minarovom dijagramu na Sl. 148 ne dominira grafik temperature, već je taj grafik predstavljen diskretno u donjem delu slike. Iako nije bilo potrebe za korišćenjem crvene boje na Sl. 157, ova boja ipak omogućava da postane jasno da se u prvih 20 s na signalu javlja EKG šum, oko 80 s da se javlja šum na 50 Hz i da se oko 120 s na EGG signalu javlja artifakt disanja. Crvena boja je poslužila da se povežu gornji i donji paneli, tako da intuitivno postane jasno da se radi o uvećanim prikazima segmenata signala. Dodatno, svi grafici u [266], odakle je preuzet grafik za prikaz na Sl. 157, su predstavljeni sa odabranom šemom boja, koja uključuje tamno crvenu nijansu sa Sl. 157.

4.1.3.2 Primeri loše prakse u vizuelizaciji biosignala

Grafici bi trebalo da su čitki i jasni [67], a da bi se ilustrovalo značaj ovih osobina, u ovom potpoglavlju, dati su primeri loše prakse³⁴⁰ u vizuelizaciji biosignala. Primer loše prikazanog grafika, predstavljen je na levom krajnjem panelu na Sl. 158 u formi bar grafika koji prikazuje srednju vrednost i standardnu devijaciju.

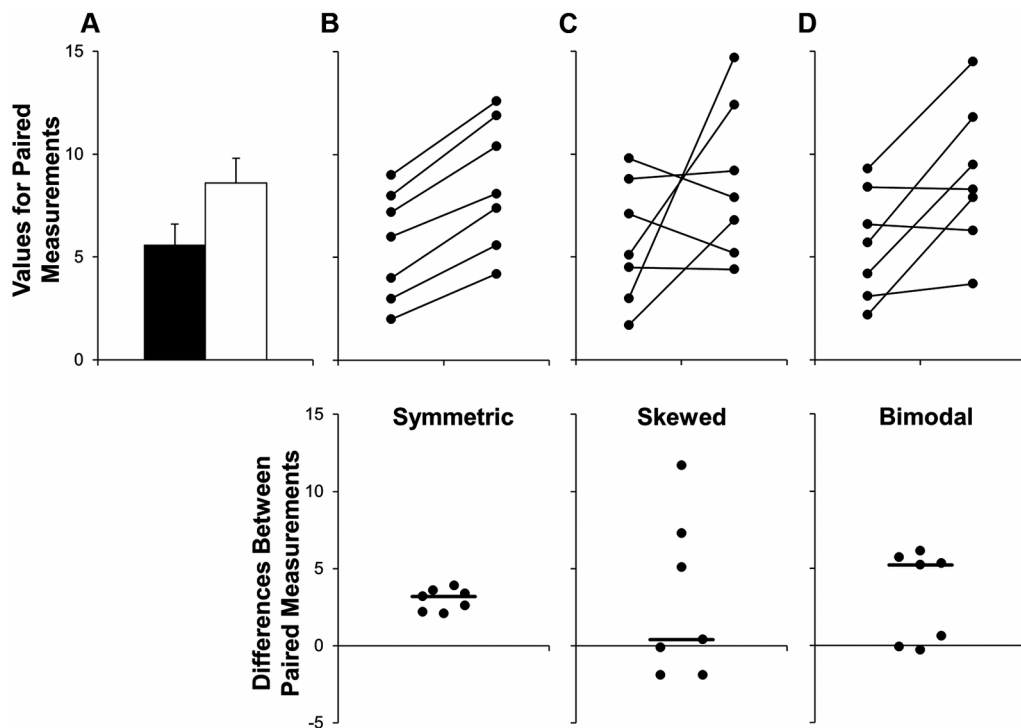


Slika 158, Ilustracija raspodele podataka (prikazanih tačkama za simetrične raspodele, raspodele sa autlajerima, bimodalna raspodela i poređenje dve raspodele koje nemaju isti broj odbiraka). Slika je preuzeta iz [455] i podeljena pod slobodnom CC BY licencom.

Sl. 158 ilustruje kako *error bar* ne prikazuje dobro informaciju, jer može predstaviti podatke sa različitim raspodelama (koje se ne vide u prikazu pomoću *error bar*-a), te ga ne treba koristiti da se prikažu podaci bez prethodne provere raspodele [455]. Takođe, na Sl. 158 u donjem panelu je pokazano da se za različite raspodele koriste odgovarajući statistički testovi, što se ne vidi prilikom prikaza predstavljenog u gornjem desnom uglu, jer skupovi podataka B, C, D i E imaju isti *error bar* grafik. Slično važi i za podatke kod kojih se posmatraju promene pre i posle intervencije, jer *error bar* će biti isti, odnosno neće preneti informaciju o pojedinačnim promenama uparenih merenja kao što je ilustrovano na Sl. 159 [455]. Na primer, kada bi se posmatrala promena dominantne frekvencije na EGG signalu kod subjekata pre i posle jela, bilo bi

³⁴⁰ Vizuelizacija podataka predstavlja jednim svojim delom i umetničku disciplinu, te je nekada dobro ili loše u očima posmatrača. U ovom poglavlju fokus je na loše prenetoj informaciji, te primer loše prakse treba shvatiti na taj način.

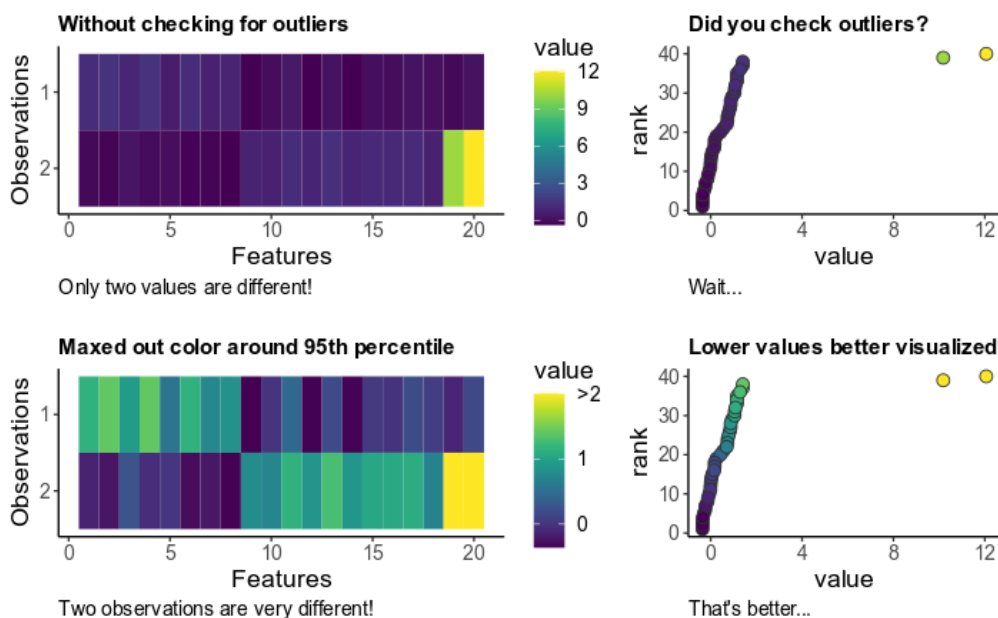
korisno pokazati da li u svakom pojedinačnom slučaju dolazi do porasta dominantne frekvencije. Na panelu B na Sl. 159 je pokazano kako uvek dolazi do povećanja, te bi tada bilo opravdano koristiti bar grafik na panelu A kada razlike imaju simetričnu raspodelu. U svim ostalim situacijama, iako bar grafik opisuje promene na podacima, informacija nije ista, te nije opravdano korišćenje bar grafika.



Slika 159, Ilustracija razlike parametara iz dva skupa kada razlike imaju simetričnu, asimetričnu i bimodalnu raspodelu, te prikaz u desnom uglu (panel A) nije odgovarajući, ako nije poznata raspodela razlika uparenih podataka (posmatranje parametara pre i posle intervencije). Slika je preuzeta iz [455] i podeljena pod slobodnom CC BY licencom.

Kada se koriste 2D grafici da se prikaže 3D informacija, osim izbora boje, važno je u kom opsegu se boja prikazuje. Na Sl. 160 [456] je pokazano kako je moguće primenom iverograma (eng. *eventogram*) prikazati kako se razlikuju 20 obeležja u dva skupa. Međutim, ako na skupu postoje autlajeri, onda će opseg boja biti takav da neće biti moguće da se uoče razlike između dva skupa koji se posmatraju, jer će kontrast boja dominantno biti iskorišćen za autlajere (gornji panel na Sl. 160). Ako se opseg boja rasporedi na vrednostima koje nisu ekstremne, onda se jednostavno dolazi do zaključka (donji panel na Sl. 160) da iverogrami za dva skupa, odnosno dve opservacije (eng. *observation*) jesu različiti. Sličan pristup se koristi za prikaz kros-korelacionih koeficijenata, ali

treba napomenuti da se za kros-korelacione koeficijente ovakve izmene skala ne primenjuju, jer je skala za kros-korelacije normalizovana od -1 do 1, odnosno u slučaju apsolutnih vrednosti kros-korelacionih koeficijenata od 0 do 1.



Slika 160, Prikaz inventograma u slučaju postojanja autlajera (gornji panel) i bez autlajera (donji panel). Slika je preuzeta iz [456] i podeljena je pod slobodnom MIT licencom³⁴¹.

Neki od problema sa lošim prikazom podataka mogu se izbeći pažljivim odabirom alata za vizuelizaciju, ali nijedan alat ne može zameniti kritičko mišljenje i savesno predstavljanje rezultata analize biosignala. U sledećem potpoglavlju, dat je prikaz korišćenja jednog od najpoznatijih R paketa čijom se primenom mogu dobiti kvalitetni i sadržajni grafici. Kako nijedan udžbenik ne može biti dovoljno iscrpan da se predstavi kompletan skup tehnika za vizuelizaciju, studenti su upućuju na pretragu kataloga na internetu koji mogu biti odličan izvor inspiracije za vizuelizaciju rezultata analize biosignala. Preporučeni izvor je STHDA³⁴² internet stranica (od eng. *Statistical Tools for High-throughput Data Analysis*) čiji je sadržaj podeljen pod CC BY NC SA licencom, a STHDA katalog za bar grafike je dostupan na: <http://www.sthda.com/english/wiki/ggplot2-barplots-quick-start-guide-r-software-and-data-visualization> (pristupljeno 03.05.2024).

341 Informacija je dostupna na GitHub stranici na kojoj su prikazani primeri loše vizuelizacije podataka, <https://github.com/cxli233/FriendsDontLetFriends>, pristupljeno 03.05.2024.

342 <http://www.sthda.com/english/>, pristupljeno 03.05.2024.

4.1.4 Primeri grafika u ggplot paketu

Pored ggplot paketa, postoji niz drugih paketa u R-u i drugim programskim jezicima koji se mogu koristiti za vizuelizaciju signala. Na primer, za prikaz podataka u veb aplikacijama najčešće su u upotrebi JavaScript biblioteke, dok se u R-u se, pored ggplot, odnosno ggplot2³⁴³ paketa [228], koriste i RGraphics [458], car [459], lattice [460], ndtv [461], ggvis [462], plotly [463], ggthemes [464] i Shiny [465] biblioteke, ali i druge specijalizovane kao što su ggcorplot [466] i timevis [467]. Za vizuelizaciju u Pajtonu korisne su biblioteke geoplotlib [468], pandas [469], altair [470] i druge [471], dok su najpoznatijih biblioteke u JavaScript-u d3.js [472], charts.js [473], dimple.js [474] i druge [475]. Ovo nije beznačajna informacija, jer se u okviru R-a mogu koristiti JavaScript i Pajton kodovi – R Studio razvojno okruženje pruža mogućnost integracije drugih programskih jezika u R okruženje (Pogledati Dodatak C ovog udžbenika), a postoje i dodatni paketi koji omogućavaju takvu funkcionalnost. Neki od osnovnih kriterijuma za izbor softvera za vizuelizaciju podataka uključuju mogućnost deljenja grafika, interoperabilnost sa drugim paketima i programskim jezicima, izbor više opcija za prikaz grafika, mogućnost analize podataka i jednostavnost (kompaktne funkcije za podešavanje stilskih elemenata). Naravno, nemoguće je ne spomenuti i licence, jer je korisno da se radi sa paketima koji su podeljeni uz slobodne licence, posebno u akademskom okruženju. [3, 67, 476]

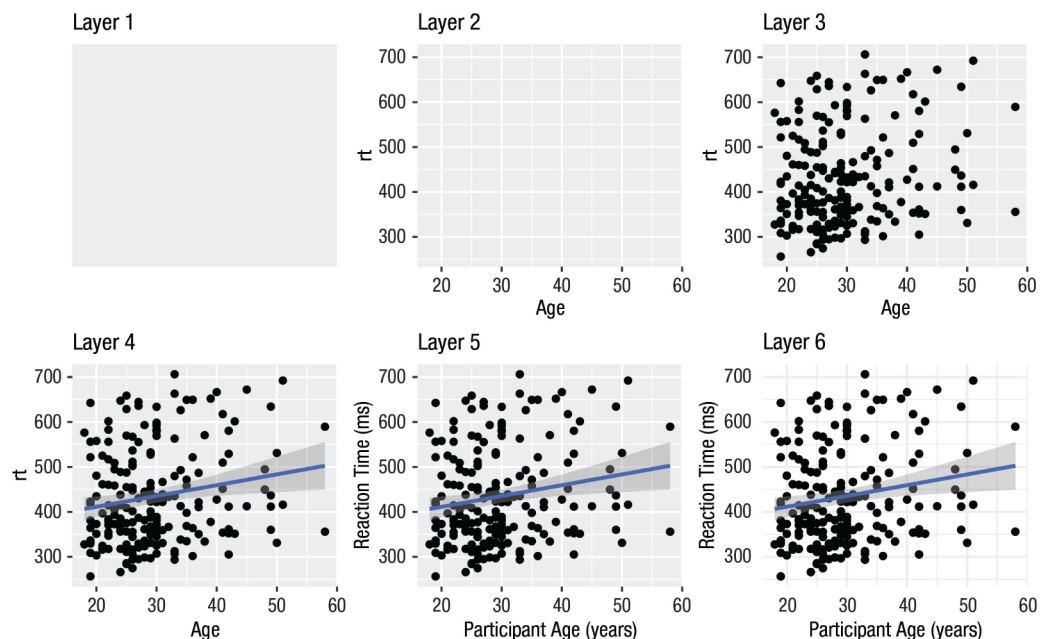
Osnovna odlika ggplot grafika jeste što postoji mogućnost da se grade kroz slojeve (eng. *layers*) što je ilustrovano na Sl. 161. Za dodatno pojašnjenje slojevitog prikaza grafika, koristi se produkcija crtanih filmova, jer da bi se dobio crtani film potrebno je pripremiti pozadinu, pa tek onda karaktere/avatare koji se dodaju na tu pozadinu. [477]. Slojevi koji su prikazani na Sl. 161 odgovaraju sledećim podešavanjima u ggplot2 paketu:

- Prvi sloj (eng. *Layer 1*) je određen prostorom za prikaz grafika,
- Drugi sloj (eng. *Layer 2*) ima dodate oznake na osama sa brojnim vrednostima i sa nazivom promenljivih,
- Treći sloj (eng. *Layer 3*) sadrži prikaz pojedinačnih vrednosti na grafiku (tip grafika je *scatter plot*),
- Četvrti sloj (eng. *Layer 4*) se sastoji iz dodate linearne zavisnosti dve promenljive (u skupu u kome se ispituje zavisnost vremena reakcije i starosti ispitanika) sa standardnom devijacijom,

343 Naziv paketa koji se koristi u R skriptama ima broj 2 u sebi (ggplot2), jer istorijski predstavlja drugu verziju koja se zadržala do danas, jer se sva ažuriranja vrše pod nazivom ggplot2 [457].

- Peti sloj (eng. *Layer 5*) ima dodate nazive osa sa jedinicama – vreme reakcije (eng. *Reaction Time*) je prikazano u ms, dok je starost/uzrast ispitanika (eng. *Participant Age*) predstavljena u godinama i
- Šesti sloj (eng. *Layer 6*) sadrži podešavanje stila, te je odabran stil sa svetlijom pozadinom od onog koji je dodat na drugom sloju.

Svaki od slojeva može biti sačuvan u zasebnu promenljivu, te menjan dalje, po potrebi, u čemu se ogleda fleksibilnost ggplot paketa u odnosu na osnovne funkcije u R-u.



Slika 161, Prikaz kreiranja grafika u R-u po slojevima. Na slici je ilustrovano kako se dobija konačna verzija grafika u ukupno 6 slojeva (eng. *layer*). Grafik je preuzet iz [477] i podeljen pod CC BY 4.0 licencom.

U biomedicinskom inženjerstvu se pretežno koriste numerički podaci, bilo da se radi o vremenskim serijama ili transformacijama vremenskih serija (npr. prikaz signala u spektralnom domenu), a koriste se i kategoričke promenljive za grupisanje ispitanika (podele po starosti, polu i drugim kriterijumima). Pored ova dva tipa podataka, u praksi se mogu sresti i geoprostorni podaci (predstavljaju parove geografske širine i dužine, a koriste se za prikaz mapa), tekstulani podaci, slike, video i drugi tipovi podataka. Za prikaz grafika, koristi se od jedne do par linija koda u okviru R-a. Međutim, iza konačnog izbora dobrog grafika i par linija koda, stoji mukotrpan posao i puno sati rada. Na primer, da bi se tokom analize podataka (prilikom izbora odgovarajuće metode

za analizu) uočile međusobne zavisnosti koje možda i neće biti prikazane u finalnom izveštaju ili da bi se odredio najpogodniji način prikaza, potrebno je u fazi ispitivanja (istraživanja podataka) pripremiti relativno veliki broj grafika (nekad i par stotina) [67]. Automatizacija prikaza informacije na grafiku, shodno tome, ne treba da bude prioritarna, jer automatizacija može onemogućiti vizuelnu proveru procedura i rezultata analize.

Nakon instalacije `ggplot2` paketa i njegove aktivacije primenom `library()` funkcije, moguće je istražiti neke od osnovnih opcija za rad pokretanjem ugrađenih primera. Primeri grafika sa ugrađenim podacima dobijaju se pozivom funkcije:

```
example("geom_point")
```

koja omogućava da se navigacijom sa tastature pregleda veći broj *scatter* grafika koji prikazuju međusobnu zavisnost odabranih parametara. Jedan od ugrađenih skupova, koji se koristi u ilustrativnim primerima, se odnosi na kvalitet i cenu dijamanta na tržištu ("diamonds" skup). Da bi se, na primer, prikazao bar grafik zavisnosti načina na koji su sečeni dijamanti (rez ili eng. *cut*) od čistoće (eng. *clarity*), može se koristiti sledeća linija koda:

```
ggplot(diamonds) + geom_bar(aes(x = cut, fill = clarity))
```

koja rezultuje grafikom prikazanim na Sl. 162. Primititi da je iskorišćenost prostora značajno bolja nego kod klasičnih R funkcija. Dodatno, primititi da je grafik prikazan sa:

```
ggplot()
```

funkcijom koja postavlja osnovni, odnosno prvi sloj (kada bi se prikazao samo rezultat ove funkcije sa ili bez *data frame*-a kao argumentom dobio bi se samo prvi sloj kao na Sl. 161). Funkcija:

```
geom_bar()
```

se koristi za iscrtavanje bar grafika i kao argument ima:

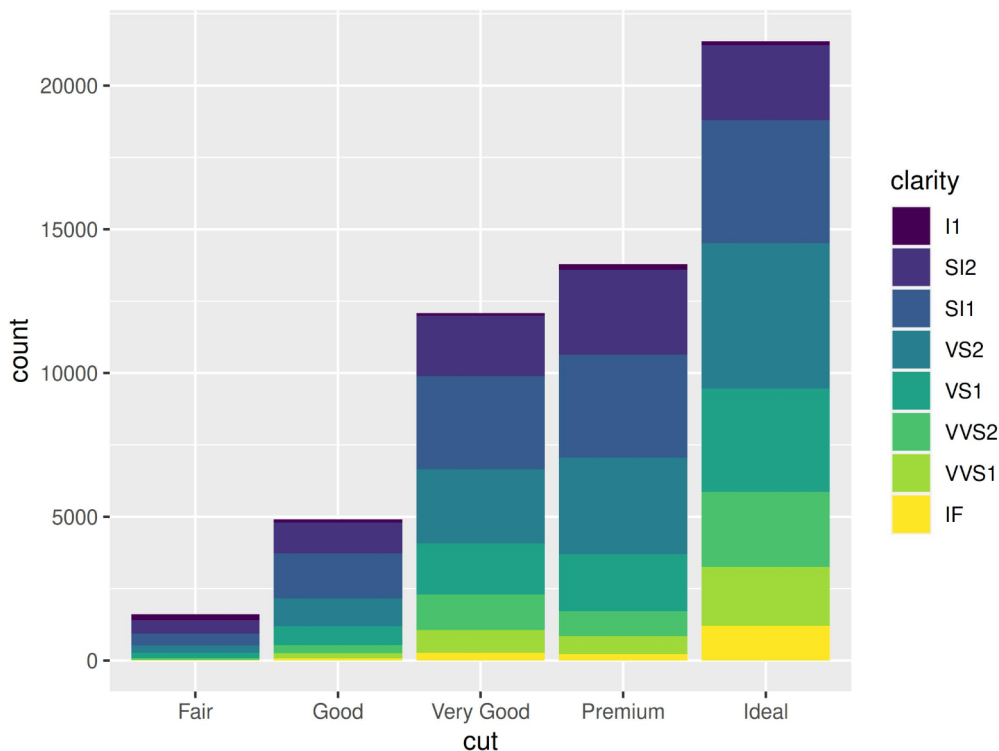
```
aes()
```

funkciju koja služi za estetsko mapiranje grafika, odnosno služi da se odredi šta se nalazi na kojoj osi i koja promenljiva će biti kodirana bojom. Estetika grafika

se odnosi na sve vidljive elemente na grafiku, kao što su pozicija na osama (apscisi i ordinati), boje (unutar oblika ili kao ivice različitih oblika), oblik, tip linije i veličina objekata/linija na grafiku. Ako bi se kojim slučajem promenio redosled promenljivih unutar *aes()* funkcije kao:

```
ggplot(diamonds) + geom_bar(aes(x = clarity, fill = cut ))
```

dobio bi se grafik kao na Sl. 163.



Slika 162, Prikaz bar grafika za zavisnost reza od čistoće dijamenta iz ugrađenog skupa "diamonds".

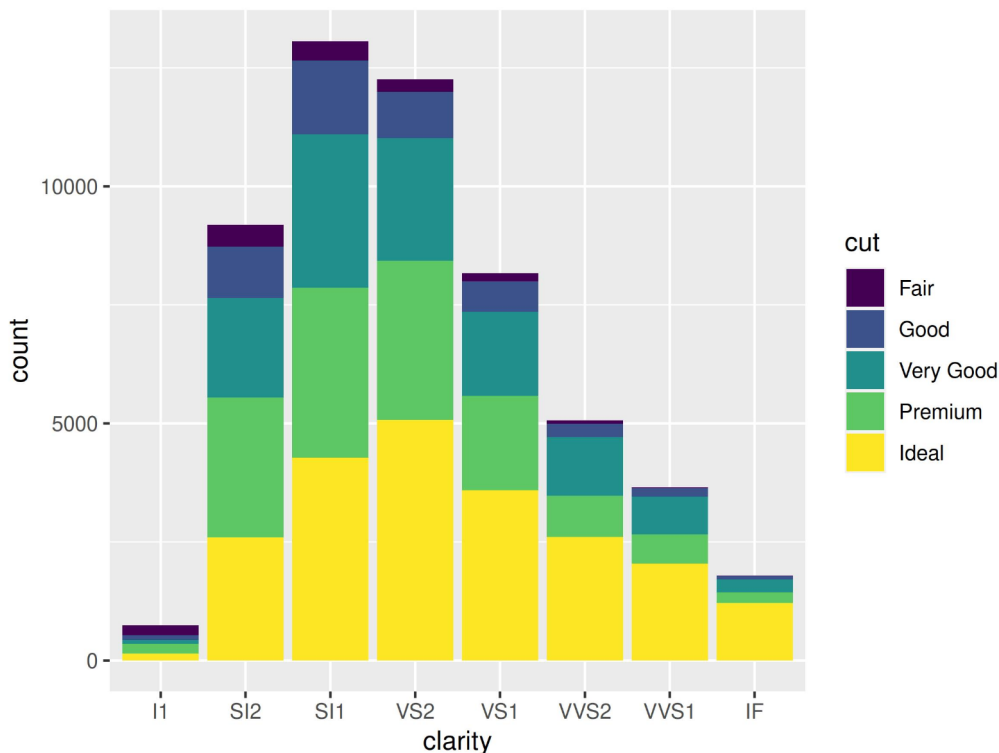
U okviru ggplot2 paketa, postoji familija *geom_* funkcija koje služe za iscrtavanje grafika, odnosno prikaz nekog podskupa estetskih elemenata (ovde je iskorišćena *geom_bar()* funkcija). Za prikaz zavisnosti promenljivih (tzv. na eng. *scatter plot*³⁴⁴), koristi se funkcija *geom_point()*, a za prikaz vremenskih serija ili linijskih trendova³⁴⁵ koristi se *geom_line()*, dok se za prikaz *box plot*³⁴⁶

344 https://en.wikipedia.org/wiki/Scatter_plot, pristupljeno 02.02.2024.

345 https://en.wikipedia.org/wiki/Line_chart, pristupljeno 02.02.2024.

346 https://en.wikipedia.org/wiki/Box_plot, pristupljeno 02.02.2024.

grafika koristi `geom_boxplot()`. Ako je potrebno prikazati više grafika na jednoj slici, potrebno je jednostavno primeniti operator za sabiranje “+”.

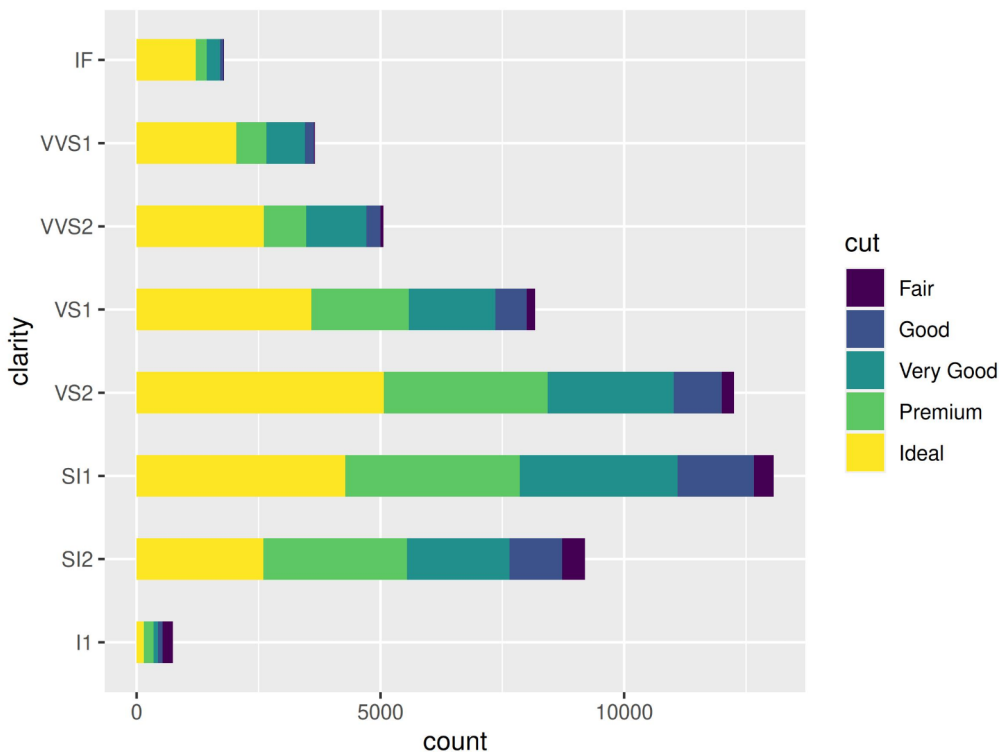


Slika 163, Prikaz bar grafika za zavisnost čistoće od reza dijamenta iz ugrađenog skupa “diamonds”.

Iz primera dva bar grafika (Sl. 162 i Sl. 163), jednostavno se nameće da ggplot omogućava dolazak do brzih rezultata, te je to najveći razlog zašto je jedan od omiljenih paketa. Zato i ne čudi što postoji posebna realizacija ggplot paketa za Pajton [478] i Matlab [229] i to u više varijanti³⁴⁷. Visok nivo apstrakcije, fleksibilnost i jednostavnost su osnovne odlike ggplot paketa. Bazični elementi svih ggplot2 grafika su podaci, *aesthetic mapping*, geometrijski objekti, statističke transformacije, skaliranje, koordinatni sistem, pozicioniranje i *faceting* (slobodan prevod sa engleskog je oblikovanje). Ovaj paket nudi veliki broj gotovih funkcija koje istovremeno vrše odgovarajuće proračune i prikazuju rezultate na grafiku. Jedna od najpogodnijih osobina ggplot paketa je što poseduje ugrađene metode za linearno i polinomijalno fitovanje, koje će biti objašnjene kasnije na primeru. Jedan od najvažnijih operatora za kombinovanje

³⁴⁷ Making plots with plotnine (aka ggplot), https://monashdatafluency.github.io/python-workshop-base/modules/plotting_with_ggplot/, pristupljeno 04.05.2024.

ggplot2 funkcija jeste operator “+” i koji omogućava slojevito kreiranje grafika (eng. *layer by layer*) i koji se ujedno koristi za prikaz više grafika na jednoj slici.



Slika 164, Prikaz rotiranog bar grafika sa povećanim prostorom između barova za podatke iz ugrađenog skupa “diamonds”.

Kod prikaza bar grafika bi trebalo voditi računa da razmak između grafika bude oko dva puta manji od širine bara. Za horizontalne barove je pokazano da se najbolje koriste za rangiranja i da ima smisla prirediti aranžiranje podataka, ali i predstavljanje horizontalnog umesto vertikalnog prikaza, jer ljudsko oko bolje poredi barove po visini nego po dužini. [69, 434]

Da bi se prikaz grafika sa Sl. 163 rotirao, potrebno je koristiti funkciju/sloj:

```
coord_flip()
```

dok je za povećanje razlika između barova, potrebno koristiti argument *width* unutar *geom_bar()* funkcije koji se postavlja na vrednosti od 0 do 1

(normalizovano). Za *width* parametar od 0.5 (50%) dobija se slika na kojoj je širina bara ista kao razmak između barova (Sl. 164).

Da bi se grafici koji se dobiju primenom ggplot paketa sačuvali sa relativno velikom rezolucijom (kako su prikazani u ovom udžbeniku) potrebno je iskoristiti funkciju:

ggsave()

koja kao ulazne parametre ima naziv datoteke sa ekstenzijom u kojoj će grafik biti sačuvan, rezoluciju slike, dimenziju slike i boju pozadine³⁴⁸. Bar grafik³⁴⁹ ili bar dijagram (koji se dobija primenom *geom_bar()* funkcije u ggplot paketu) omogućava prikaz podataka pomoću pravougaonika koji imaju dužinu ili širinu proporcionalnu veličini koja se prikazuje. Ovaj dijagram služi za poređenje tipova podataka po različitim kategorijama. Grafiku se mogu dodati površine u raznim bojama da se predstave i drugi podaci (pogledati Sl. 162 do Sl. 164), ali grafik se može i zumirati da bi se bolje predstavile razlike u dužini ili širini pravougaonika. Međutim, poželjno da kod predstavljanja procenata, opseg pravougaonika na slici bude predstavljen opseg od 0% do 100% kako bi se jednostavno videle i relativne i apsolutne razlike.

Grafici koji prikazuju detaljnije informacije o podacima (minimalnu, maksimalnu vrednost, medijanu i autlajere) su *box plot* i violinski grafik (eng. *violin plot*). Pored toga što detaljnije prikazuje informaciju, *box plot* omogućava prikaz osnovnih parametara za signal za koji nije poznata funkcija gustine verovatnoće ili je poznato da nije Gausova raspodela. U ovom udžbeniku se neće detaljno prolaziti kroz sve moguće elemente grafika, već kroz neke osnovne primere koji se mogu relativno jednostavno koristiti u praksi. Da bi se ilustrovao prikaz *box plot*-a, koji je poželjniji za predstavljanje karakteristika podataka od bar grafika, posebno kada nije poznata raspodela odbiraka i ako ih ima relativno malo, biće iskorišćeni podaci iz ISwR paketa iz skupa “lung” (od eng. *lung volume*) u kojima su sadržani podaci o kapacitetu pluća za 6 ispitanika na kojima je kapacitet pluća meren sa tri metode koje su označene sa A, B i C [36, 479]. Da bi se na odgovarajući način napravilo poređenje osnovnih R funkcija i ggplot funkcija, na Sl. 165 je prikazano korišćenje dva načina za prikaz *box plot*-a, dok su na Sl. 166 prikazani odgovarajući grafici.

348 Za detalje primene *ggsave()* funkcije pogledati Help stranicu u R Studio razvojnom okruženju.

349 https://en.wikipedia.org/wiki/Bar_chart, pristupljeno 02.02.2024.


```

# osnovne R funkcije
boxplot(volume ~ method, data = lung,
         main = "Kapacitet pluća", xlab = "Metoda merenja",
         ylab = "Zapremina [L]")
grid()

# ggplot funkcije
g <- ggplot(lung, aes(x=method, y=volume)) +
  ggtitle("Kapacitet pluća") + xlab("Metoda merenja") +
  ylab("Zapremina [L]")
g + geom_boxplot()

```

Slika 165. R skripta za prikaz *box plot*-a primenom osnovnih R funkcija i primenom ggplot funkcija. Primiti slojevitost ggplot grafika i način korišćenja operatora “+” za dodavanje slojeva (tip grafika, natpisi na osama i naslov grafika).

Sa Sl. 166 se vidi da je raspodela ispitanika merenih A metodom simetrična dok za druge dve metode, raspodela nije simetrična. Jedna od opcija je da se zajedno sa *box plot* grafikom, ili posebno ako je dat relativno mali broj merenja, prikažu i pojedinačna merenja. U ggplot paketu to se jednostavno postiže dodavanjem novog sloja na sledeći način:

```
g + geom_boxplot() + geom_point() + theme_minimal()
```

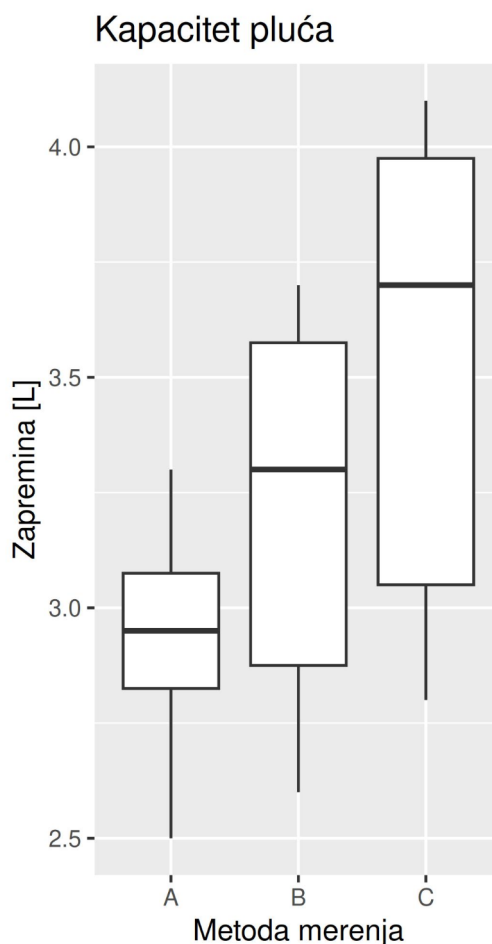
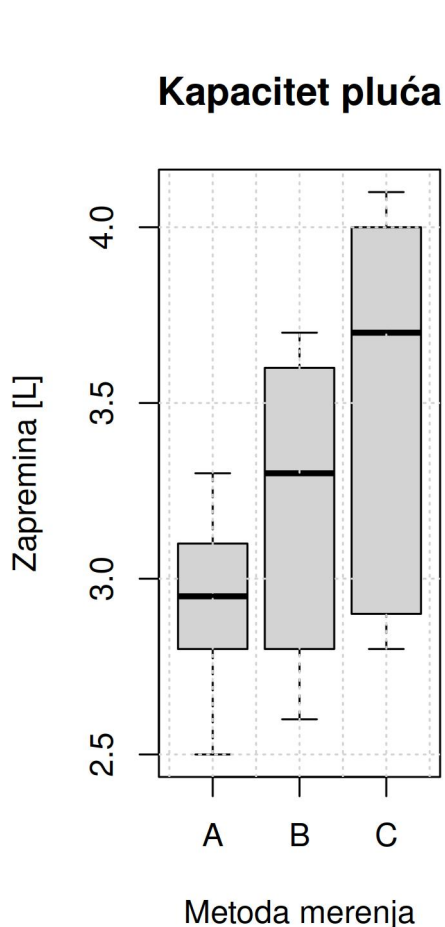
Funkcija:

```
theme_minimal()
```

pripada familiji funkcija koje počinju sa *theme_* i koje služe da se preuzmu gotova stilska podešavanja, odnosno teme, a *minimal*³⁵⁰ se odnosi na temu koja odgovara minimalističkom stilu³⁵¹. Rezultat je prikazan na Sl. 167.

350 Pored teme koja odgovara minimalističkom prikazu (eng. *minimal*), postoje druge teme kao što su siva (eng. *grey*) ili klasična (eng. *classic*). Kompletan spisak tema sa primerima grafika je dostupan na <https://ggplot2.tidyverse.org/reference/ggtheme.html> (pristupljeno 04.05.2024).

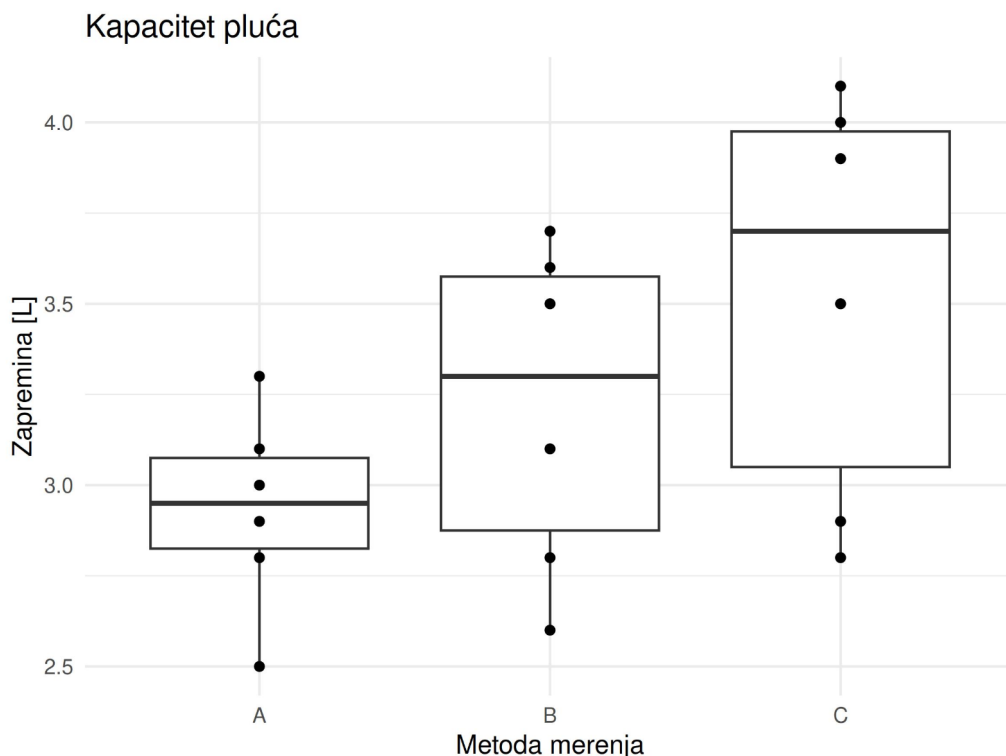
351 Minimalizam kao umetnički pravac odbacuje sve tradicionalne elemente i svodi se na redukciju prikaza (<https://sr.wikipedia.org/wiki/Minimalizam>, pristupljeno 04.05.2024). U tom smislu, ne mogu se uzimati u obzir lični izraz, politika, religija i kultura (koje odbacuje minimalizam kao umetnički pravac), ali se tehnički minimalizam svodi na sveden broj elemenata u grafičkom prikazu. U uređenju prostora, podrazumeva se samo neophodan broj elemenata (komada nameštaja) koji je neophodan za funkcionisanje, što se može direktno porediti sa vizuelizacijom.



Slika 166, Box plot grafici primenom funkcija iz osnovnog R paketa (levi panel) i ggplot funkcija (desni panel).

Na Sl. 165 je dat kod koji omogućava prikaz grafika na Sl. 166. Primetiti u kodu, da promenljiva *g* pripada ggplot klasi podataka, te da je moguće grafik sačuvati kao promenljivu radi kasnijih izmena ili prikaza. *Box plot* služi za prikaz numeričkih podataka, te je na grafiku, konvencijom, dogovoren prikaz kvantila, srednje vrednosti i medijane. Da bi se detaljnije proverio oblik raspodele, umesto *box plot*-a može se koristiti *violin plot*, jer oblik grafika odgovara obliku raspodela na podacima. Za poređenje metoda A, B i C za merenje kapaciteta pluća primenom violinskog grafika, potrebno je promenljivoj *g* dodati sloj koji odgovara *violin* grafiku i temu po izboru:

```
g + geom_violin() + theme_classic()
```



Slika 167, *Box plot* pripremljen u *ggplot* paketu sa pojedinačnim realizacijama merenja kapaciteta pluća. Iskorišćen je minimalistički stil za prikaz elemenata na grafiku.

Violinski grafik na Sl. 168 je prikazan pogrešno, jer je funkcija gustine verovatnoće procenjena samo na osnovu 6 odbiraka što je malo [456]. Međutim, ono što je dobro prikazano na Sl. 168 i ostalim graficima jeste što svi grafici imaju jasno označene ose sa odgovarajućim jedinicama. Takođe, sa prikazanih grafika se može zaključiti da se tri metode (A, B i C) razlikuju što u principu nije očekivano – pretpostavka je da sve metode dobro mere i da nema mnogo odstupanja od referentne vrednosti, ali podaci u “lung” skupu pokazuju upravo suprotno.

Poređenje različitih metoda za merenje biosignala je uobičajen zadatak u EDA oblasti. Za poređenje dve metode koristi se specijalizovan Bland-Altmanov (BA) grafik, koji služi za prikaz slaganja između dva različita merenja iste veličine [480]. Svrha BA grafika je da pruži uvid u pojedinačne razlike merenja u odnosu na srednju vrednost, kao i da prikaže postojanje pristrasnosti (eng. *bias*) u pojedinim metodama koje nije moguće uočiti korišćenjem drugih tehnika vizuelizacije. Za prikaz BA grafika postoji specijalizovan paket sa

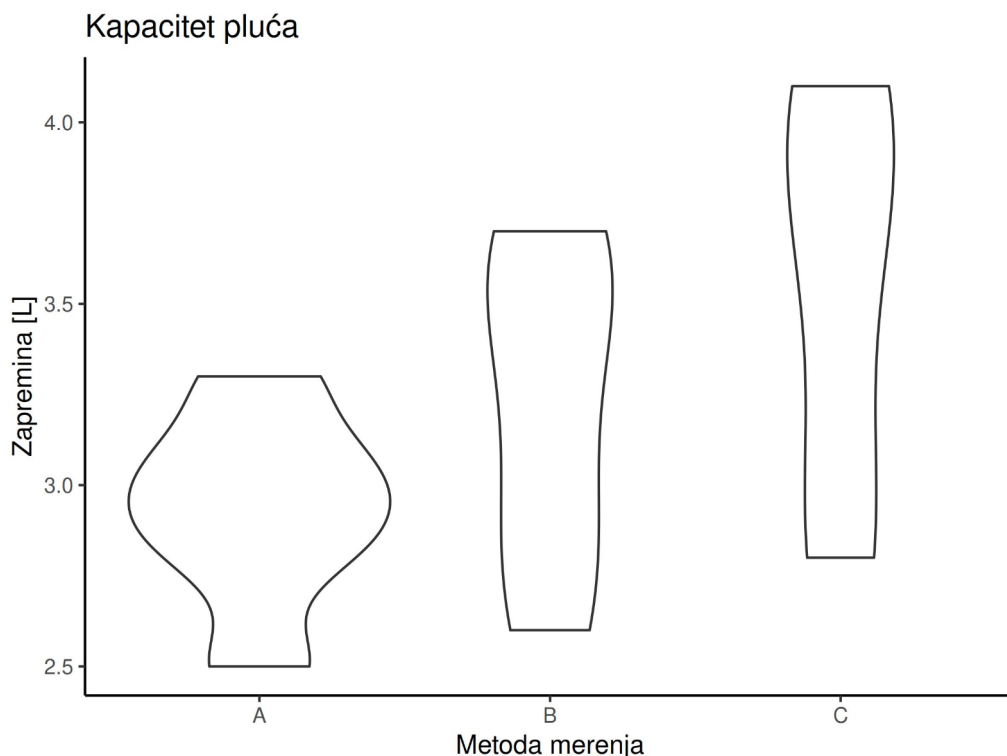
CRAN-a koji je moguće koristiti u kombinaciji sa ggplot paketom – blandr [481]. U okviru ovog paketa, da bi se prikazao BA grafik koristi se funkcija:

blandr.statistics()

koja omogućava da se izračunaju parametri deskriptivne statistike (na primer, srednje vrednosti u grupama i pojedinačne razlike parametara). Nakon toga, primenjuje se funkcija, koja omogućava da se predstavi BA grafik:

blandr.plot.ggplot()

Kratka skripta za prikaz BA grafika, za poređenje metoda A i B, je prikazana na Sl. 169, a BA grafik je prikazan na Sl. 170.



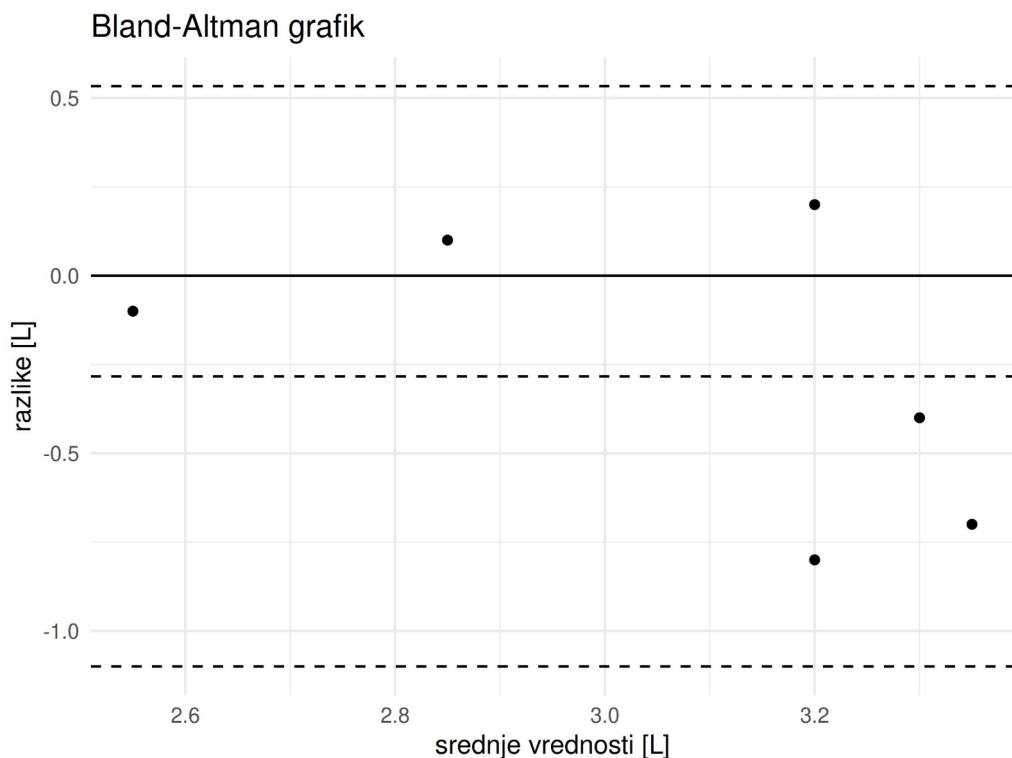
Slika 168, Violin plot za merenje kapaciteta pluća sa tri metode (A, B i C) sa klasičnom temom – korišćenjem sloja *theme_classic()*.

Na Sl. 170 je na apscisi prikazana srednja vrednost dva merenja dok je na ordinati prikazana razlika pojedinačnih merenja primenom metode A i metode B. Što su pojedinačne vrednosti (predstavljene sa tačkama) dalje od horizontalne

linije, koja je prikazana punom linijom u 0, to znači da se merenja primenom dva instrumenta više razlikuju. Za pozitivne razlike, tačke su iscrtane iznad linije (dva merenja na Sl. 170) dok su negativne razlike prikazane ispod linije (4 merenja na Sl. 170). Ovakav rezultat se tumači kao da postoji pristrasnost u metodi B u odnosu na referentnu metodu A za merenje kapaciteta pluća, jer su pojedinačna odstupanja većinski negativna.

```
ba <- blandr.statistics(lung$volume[lung$method == "A"],
                      lung$volume[lung$method == "B"])
blandr.plot.ggplot(ba, plotTitle = "Bland-Altman grafik",
                  ciDisplay = FALSE , ciShading = FALSE) +
  xlab("srednje vrednosti [L]") + ylab("razlike [L]") +
  theme_minimal()
```

Slika 169, R skripta koja omogućava da se generiše BA grafik prikazan na Sl. 170.



Slika 170, Bland-Altman grafik za poređenje metoda A i B za merenje kapaciteta pluća. Skripta za prikaz grafika je data na Sl. 169.

Kao i kod violinskog grafika, potrebno je pažljivo pristupiti tumačenju BA grafika, jer je grafik prikazan samo za merenja na 6 ispitanika. Isprekidanim

linijama na BA grafiku predstavljen je interval poverenja, a prikaz se može menjati podešavanjem argumenata u `blendr.plot.ggplot()` funkciji. Na Sl. 169 su argumenti `ciDisplay` i `ciShading` postavljeni na negativnu logičku vrednosti (FALSE) te je na Sl. 170 predstavljen sveden BA grafik bez detaljnih oznaka intervala poverenja, senki na grafiku i parametara koji opisuju pristranost dve metode.

Najveći problem u prethodnim prikazima je relativno mali broj dostupnih merenja, te će za dalju ilustraciju ggplot funkcija biti korišćen skup podataka Blood1 pod naslovom “Blood pressure, weight and smoking status”³⁵² iz baze podataka koja je podeljena pod GPL-3 licencom³⁵³ [482-483]. Podaci su mereni/dobijeni na 500 ispitanika i sadrže tri kolone: (1) vrednosti sistolnog krvnog pritiska u mmHg (*SystolicBP*), (2) kategoriju *Smoke* (da li je ispitanik pušač ili nije pušač) i (3) kategoriju *Overwt* (da li ispitanik pripada jednoj od tri kategorije *obese*, *normal* ili *overweight*, odnosno da li ispitanik pripada grupi gojaznih ili grupama sa normalnom i prekomernom težinom).

Kada se prikaže violinski grafik za Blood1 podatke na Sl. 171, dobija se simetrična raspodela samo za *normal* grupu ispitanika. Predstavljanjem pojedinačnih odbiraka za pušače i nepušače, moguće je dodati novu dimenziju grafiku, te se, na primer, sada jasno uočava da ispitanici čiji je indeks telesne mase u granicama prosečnog indeksa zdrave osobe (BMI kategorija *normal*) i koji su nepušači imaju pretežno nizak sistolni krvni pritisak, što je u skladu sa podacima iz literature da pušači imaju povišen krvni pritisak [484].

R skripta, koja omogućava prikaz *violin* grafika za relativno veliki broj uzoraka, je data na Sl. 172. Dodatni sloj, koji do sada nije objašnjen, odnosi se na odabir boja pojedinačna merenja sistolnog krvnog pritiska koja odgovaraju ispitanicima u zavisnosti od toga da li su pušači ili nepušači. Promena boja se postiže primenom funkcije iz familije `scale_color_`:

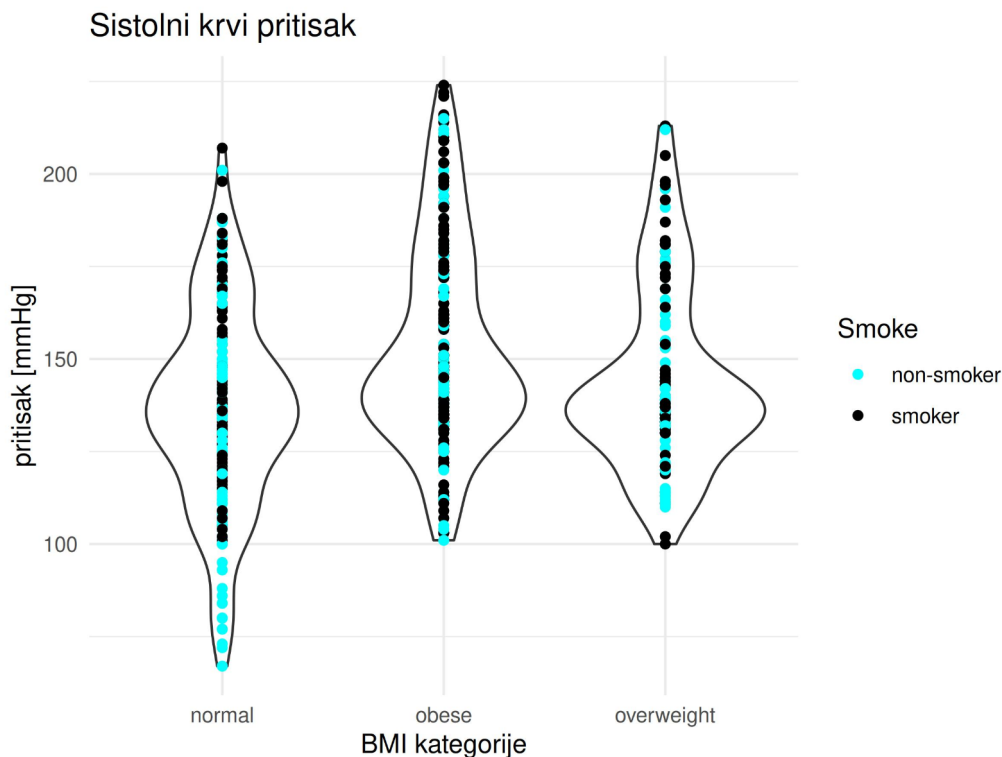
`scale_color_manual()`

za izbor boja ili šeme boja. U ovom slučaju su boje odabrane ručno sa ciljem da se postigne dobar kontrast u osvetljenju i u slučaju da se grafik prikazuje u

352 Podacima se može direktno pristupiti preko <https://raw.githubusercontent.com/vincentarelbundock/Rdatasets/master/csv/Stat2Data/Blood1.csv> (pristupljeno 05.05.2024), dok je opis podataka dostupan na <https://vincentarelbundock.github.io/Rdatasets/doc/Stat2Data/Blood1> (pristupljeno 05.05.2024).

353 <https://vincentarelbundock.github.io/Rdatasets/>, pristupljeno 05.05.2024.

nijansama sive, kao i sa željom da se održi perceptualna uniformnost (cijan boja je korišćena dominantno na graficima u udžbeniku, a ovde označava pritisak meren kod nepušača u odnosu na crnu boju koja ima negativniju percepciju od cijan boje i označava pritisak meren kod pušača). [485]



Slika 171, Prikaz violinskog grafika za odnos sistolnog krvnog pritiska za tri BMI (eng. *Body Mass Index*) kategorije ispitanika. Tačkama su prikazana pojedinačna merenja sistolnog pritiska na ispitanicima, bojom je predstavljeno da li su ispitanici pušači ili ne.

```
p1 <- ggplot(dat, aes(x = Overwt, y = SystolicBP)) +
  ggtitle("Sistolni krvi pritisak") +
  ylab("pritisak [mmHg]") +
  xlab("BMI kategorije") + theme_minimal()
p1 + geom_violin() + geom_point(aes(col = Smoke)) +
  scale_color_manual(values = c("cyan", "black"))
```

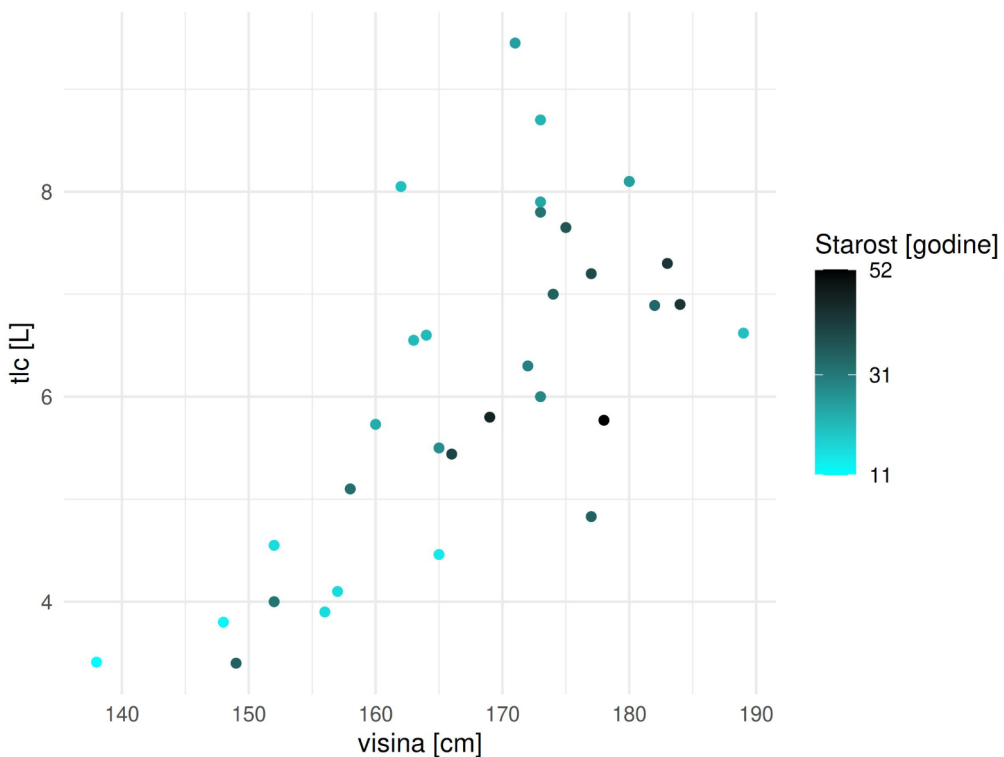
Slika 172. R skripta sa ggplot funkcijama za prikaz violinskog grafika na Sl. 171.

Pored skaliranja boja, *scale_* familija funkcija se može koristiti za skaliranje pozicije, veličine i oblika, kao i za odabir tipa linije. Uobičajeni argumenti

`scale_()` funkcija su: ime (nazivi osa i/ili legende), granice za vrednosti koje se prikazuju na skali, tzv. *breaks*, odnosno lokacije na osi, na kojima bi trebalo postaviti oznake, labele (od eng. *label*) koje se odnose na nazive koji se pojavljuju kod svake anotacije, kao i drugi argumenti. Na primer, pomoću:

`scale_color_continuous()`

se podešava da se boje postave na vrednosti *low* ili *high* (ako se ovi argumenti ne unesu dobija se podrazumevana plava šema boja). Sa ciljem ilustracije skaliranja iskorišćeni su podaci iz `tlc` skupa podataka (eng. *Total Lung Capacity*) iz ISwR paketa za prikaz grafika zavisnosti dve promenljive na Sl. 173 za koju je data kratka skripta na Sl. 174.



Slika 173, Zavisnost visine ispitanika od ukupnog kapaciteta pluća (eng. *total lung capacity*, skraćeno `tlc`), kao i u zavisnosti od starosti ispitanika. Prikazani su podaci iz “`tlc`” skupa podataka koji je dostupan u ISwR paketu. Očekivano, pretežno stariji ispitanici imaju veći `tlc` i viši su u poređenju sa mlađim ispitanicima.

Kontinulano skaliranje boja se ne preporučuje, ako su razlike na koje bi skaliranje trebalo da ukaže relativno diskretne, odnosno ne bi trebalo vršiti

zloupotrebu manipulacijom bojama na ovaj način kako bi se pokazale veće razlike od onih koje zaista postoje u podacima. Sa druge strane, ako postoje značajne razlike u podacima koji se posmatraju, onda je važno da se te razlike istaknu skaliranjem. Intuitivno se nameće zaključak da skaliranje nije poželjna opcija u slučaju da postoji relativno veliki broj grupa podataka koji se posmatra. Iako se na grafiku na Sl. 173 primećuje da postoji očekivani trend tako da stariji ispitanici imaju veću visinu i veći kapacitet pluća, iako se radi o pacijentima koji čekaju transplantaciju srca i pluća [36], kontinualno nijansiranje po godinama nije dalo dovoljno pregledne rezultate. Poželjno bi bilo da su ispitanici podeljeni u kategorije, koje su definisane opsezima godina, te bi onda i razlike u boji bile uočljivije.

```
ggplot(tlc, aes(x = height, y = tlc)) +  
  geom_point(aes(col = age)) +  
  scale_color_continuous(name = "Starost [godine]",  
                          breaks = c(11, 31, 52),  
                          labels = c(11, 31, 52),  
                          low = "cyan", high = "black") +  
  xlab("visina [cm]") + ylab("tlc [L]") +  
  theme_minimal()
```

Slika 174, R skripta za prikaz grafika zavisnosti tri promenljive na Sl. 173.

Izbor boja nije bio sporan kod Blood1 skupa podataka, jer su podaci već bili podeljeni u kategorije (po indeksu telesne mase i statusu, odnosno da li su pušači ili nisu). Sa violin grafika na Sl. 171 se može videti da je samo jedna od funkcija gustina verovatnoće simetrična, pa se može pretpostaviti da simetričnost ne postoji ni na celom skupu podataka (za sve sistolne pritiske). Da bi se to proverilo, može se prikazati histogram sa pridruženom funkcijom gustine verovatnoće kao na Sl. 175.

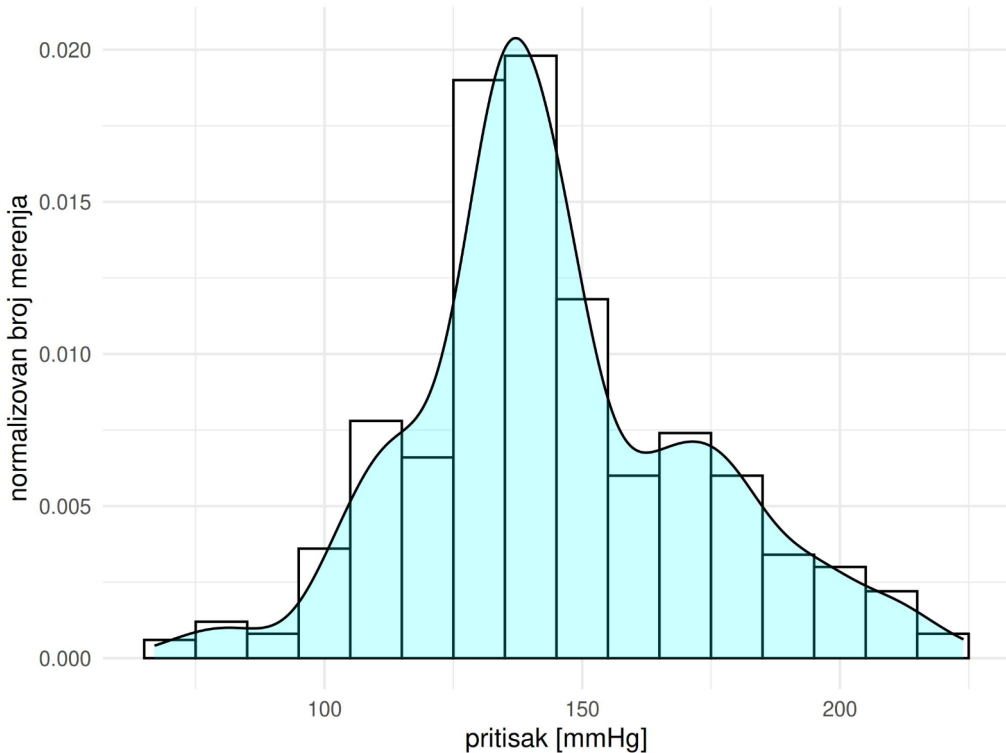
Preporuka za prikaz histograma je da se koristi broj binova koji je jednak kvadratnom korenu broja merenja uvećanim za 1 [78]. U okviru:

geom_hist()

funkcije postoji parametar *binwidth* koji se može proračunati jednostavnom podelom opsega pritiska sa preporučenim brojem binova [78]. Radi jednostavnosti, histogram za Sl. 175 je prikazan za 10 binova (Sl. 176). Da bi se prikazala funkcija gustine verovatnoće na Sl. 175, je prikazano kako je moguće za te potrebe koristiti funkciju:

`geom_density()`

na takav način da se ne ugrozi vidljivost histograma, odnosno korišćenjem alfa (α) parametra kojim se definiše nivo transparentnosti (na Sl. 175 je postavljen na 20%, odnosno na 0.2) u prikazu funkcije gustine verovatnoće.



Slika 175, Prikaz histograma sa funkcijom gustine verovatnoće za sve podatke iz studije “Blood 1”.

```
p2 <- ggplot(dat, aes(x = SystolicBP)) +  
  xlab("pritisak [mmHg]") + ylab("normalizovan broj merenja")  
p3 <- p2 + geom_histogram(aes(y = ..density..), binwidth = 10,  
  color = "black", fill = "white") +  
  geom_density(alpha = .2, fill = "cyan") + theme_minimal()  
print(p3)
```

Slika 176, R skripta za prikaz histograma i funkcije gustine verovatnoće za podatke iz skupa Blood1. Rezultat skripte je prikazan na Sl. 175.

Histogram je jedan od najkorisnijih grafika u oceni funkcije gustine verovatnoće na eksperimentalnim podacima, a prvi histogram je nacrtao Karl

Pirson³⁵⁴ koji se bavio biostatistikom i koji je imao veliki uticaj na razvoj statističkih testova i deskriptivne statistike. Između ostalog, Pirson je razvio i Pirsonov korelacioni koeficijent koji meri linearnu povezanost između dve promenljive [486]. Na Sl. 175 se može videti da, očekivano, funkcija gustine verovatnoće ne odgovara Gausovom zvonu, da nije simetrična i da je najverovatnije multimodalna (nastala kombinacijom podataka koji imaju različite raspodele).

Sistolni krvni pritisak za pojedinačne kategorije, ali i ukupno za sve tri grupe ispitanika, ne bi trebalo predstavljati srednjom vrednošću i standardnom devijacijom, osim u slučaju da se želi pokazati kolika je varijabilnost pritiska za različite kategorije. Za te potrebe, koristi se grafik greške ili *error bar* ili *dynamite plot* koji je, kako je do sada navedeno, prilično i sa razlogom kritikovan. U otvorenom pismu urednicima časopisa pod nazivom “Open letter to journal editors: Dynamite plots must die”, prof. Rafael Irizari sa Harvard Univerziteta je obrazložio svoje viđenje zašto nikada ne treba koristiti *error bar* u naučnim publikacijama [487]. Bez obzira na kritike, ovo je često korišćen grafik i ovde je prikazano kako se može koristiti ggplot paket za njegov prikaz. Pre primene funkcije:

```
geom_errorbar()
```

potrebno je, na odgovarajući način, transformisati podatke u formi *data frame*-a tako da sadrži kolone sa srednjim vrednostima i standardnim devijacijama po različitim grupama. Primer korišćenja dplyr paketa za izdvajanje srednje vrednosti pritiska je:

```
meanN <- filter(dat, Overwt == "normal") %>% select(SystolicBP) %>%  
unlist %>% mean
```

Kada se izračunaju srednje vrednosti (*meanN*, *meanOw* i *meanOb*) i standardne devijacije po kategorijama (*sdN*, *sdOw* i *sdOb*), moguće je kreirati novi *data frame* pod nazivom *pod* (Sl. 177) koji je potrebno prikazati na grafiku. Funkcija *geom_errorbar()* ne prepoznaje automatski u kojoj koloni se nalaze standardne devijacije, a u kojoj koloni se nalaze srednje vrednosti pritiska, te je potrebno ručno uneti granice za prikaz standardne devijacije, kao na Sl. 177. Primenom parametra *lwd* menja se debljina linije za prikaz standardne devijacije, dok se

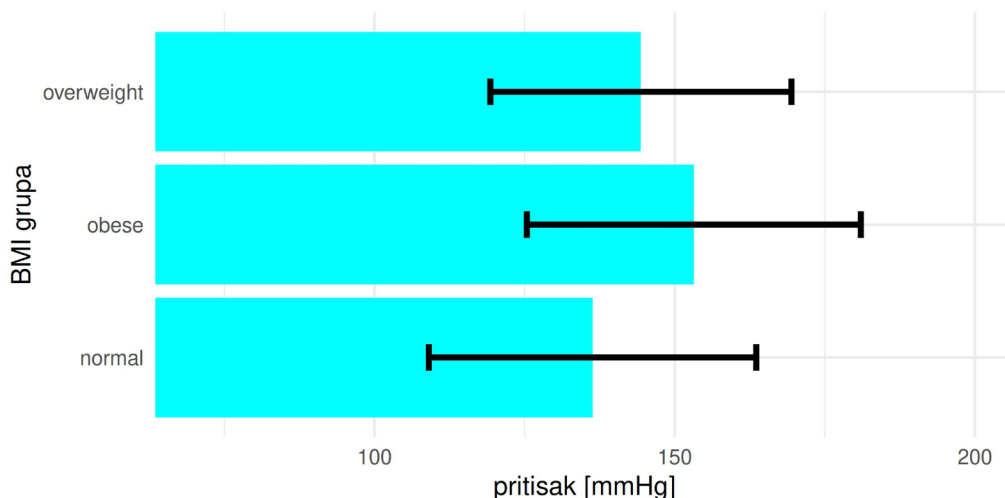
354 Karl Pearson (1857-1936), https://en.wikipedia.org/wiki/Karl_Pearson, pristupljeno 02.02.2024.

primenom parametra *width* menja dužina na krajevima grafika koja se koristi za prikaz standardne devijacije na Sl. 178.

```
pod <- list("numeric")
pod$BMI <- as.factor(c("normal", "obese", "overweight"))
pod$mn <- c(meanN, meanOw, meanOb)
pod$sd <- c(sdN, sdOw, sdOb)
pod <- as.data.frame(pod)
pod

barE <- ggplot(pod, aes(x = BMI, y = mn)) +
  xlab("BMI grupa") + ylab("pritisak [mmHg]") +
  geom_bar(stat = "identity", fill = "cyan")
barE + geom_errorbar(aes(ymin = mn - sd, ymax = mn + sd), width = 0.2, lwd = 1.2) +
  coord_flip(ylim = c(70, 200)) + theme_minimal()
```

Slika 177, R skripta koja se koristi za prikaz grafika greške primenom ggplot funkcija.



Slika 178, Grafik greške ili *error bar* za poređenje srednje vrednosti i standardnih devijacija sistolnog krvnog pritiska za tri različite grupe ispitanika na osnovu indeksa telesne mase (eng. *Body Mass Index*, skraćeno BMI).

4.1.4.1 Prikaz vremenskih serija u ggplot paketu

U većini slučajeva, biosignali su predstavljeni vremenskim serijama (na primer: EKG, EGG, PPG i EEG signali). Pored osnovnih R funkcija, ggplot paket se takođe može koristiti za prikaz vremenskih serija. Prikaz biosignala u vremenskom i frekvencijskom domenu nije ništa do običan grafik zavisnosti, odnosno na eng. *scatter plot*³⁵⁵, jer se prikazuje zavisnost amplitude od

355 https://en.wikipedia.org/wiki/Scatter_plot, pristupljeno 02.06.2024.

vremena, odnosno magnitude od frekvencije. Međutim, prva asocijacija na *scatter plot* bi bila da se posmatraju dva obeležja (ovo je jednostavniji slučaj u odnosu na kros-korelacione matrice koje sadrže veliki broj obeležja³⁵⁶). Kod prikaza linijskog grafika (biosignal u vremenskom ili frekvencijskom domenu), u literaturi se mogu sresti termini diskretno i kontinualno za prikaz grafika [67]. Termin kontinualno treba uzeti sa rezervom, jer se na računaru analiziraju isključivo diskretizovani podaci. Međutim, kako kontinualna predstava jednostavnog spajanja svih tačaka koji su uključeni u vremensku seriju nekog signala ili pak neki drugi prikaz odaje utisak kontinualnosti, s toga je ovaj termin usvojen u praksi. Treba voditi računa da je nekada zgodno da se podaci prikažu i kontinualno i diskretno, posebno ako ne postoji veliki broj tačaka, a potrebno je uočiti odgovarajući trend u podacima. Nekada se termin kontinualno koristi i za podatke kako bi se naglasilo da se radi o podacima koji nisu kategorički [67].

U ovom poglavlju, korišćeni su signali iz “puls.txt” datoteke u kojoj se nalazi signal nalik EKG³⁵⁷-u (meren EKG elektrodama i primenom elektrofizioloških pojačavača, ali elektrode nisu postavljene na takav način da se ne mere klasični odvođi kao što se to radi primenom Ajthovenovog trougla [488-489]) i signal pulsa ispitanika meren sa infracrvenim (IR) senzorom (fotopletizmograf od eng. *photoplethysmography*, skraćeno PPG). Oba signala su merena primenom istog analogno digitalnog (A/D) konvertora sa frekvencijom odabiranja od 1000 Hz, a mereni su na zdravom ispitaniku na Elektrotehničkom fakultetu, Univerziteta u Beogradu. Ispitanik je dao dobroboljnu pisanu saglasnost u skladu sa Helsinškom deklaracijom [490], a merenja su izvršena u skladu sa članom 19. Kodeksa profesionalne etike Univerziteta u Beogradu (Glasnik Univerziteta u Beogradu, br. 193/16 i 229/21³⁵⁸). Cilj merenja je bio da se pokaže da se konstruisanim PPG senzorom može meriti puls ispitanika, odnosno cilj je bio validacija PPG senzora u odnosu na komercijano dostupan elektrofiziološki pojačavač. Kako bi se vremenske serije prikazale na ggplot grafiku, poželjno je da se *data frame* reorganizuje na odgovarajući način, tako da u kolonama budu vremenski odbirci (jednog pa drugog signala), odbirci signala (odabran je

356 U praksi, ako postoji veliki broj obeležja mora se prikazati popriličan broj grafika zavisnosti (eng. *scatter plot*) za različite kombinacije obeležja. Ovakva vizuelizacija sigurno nije nešto što treba prikazati u tekstu naučnog rada, diplomskog ili master rada, ali jeste zgodan međurezultat u kome onaj ko analizira signale može steći dodatni uvid u njihove međusobne zavisnosti. Ponekad se ti grafici prikazuju u dodatku na kraju rada (eng. *Appendix*) ili u dodatnim materijalima (eng. *Supplementary materials*).

357 U daljem tekstu će ovaj signal biti označen kao EKG signal.

358 Kodeks profesionalne etike Univerziteta u Beogradu je dostupan na <https://bg.ac.rs/files/sr/univerzitet/univ-propisi/Kodeks-profesionalne-etike2021.pdf>, pristupljeno 06.05.2024.

redosled EKG pa PPG) i u trećoj kategorička promenljiva koja označava tip signala (“ekg” i “ppg”). Način kreiranja nove promenljive je prikazan na Sl. 179.

```
> pod <- list()
> pod$ vreme <- rep(seq(0, length(dat$V1)/fs - 1/fs, by = 1/fs), 2)
> pod$odbirci <- c(dat$V1, dat$V2)
> pod$signal <- c(rep("ekg", length(dat$V1)),
+                rep("ppg", length(dat$V2)))
> pod <- as.data.frame(pod)
> head(pod, 3)
  vreme odbirci signal
1 0.000 -0.033   ekg
2 0.001 -0.026   ekg
3 0.002 -0.030   ekg
```

Slika 179, Reorganizacija *data frame*-a za prikaz vremenskih serija primenom *ggplot* paketa u R-u.

Nakon što je kreiran novi *data frame* pod nazivom *pod*, moguće je kroz par slojeva *ggplot* grafika:

```
ggplot(pod, aes(x = vreme, y = odbirci)) + geom_line(aes(col = signal)) +
  ylab("amplituda [a.u.]") + xlab('vreme [s]') +
  xlim(0, 10) + theme_minimal()
```

predstaviti signale kao na Sl. 180. Jedinice za amplitudu nisu dodate, jer EKG ima pojačanje od 1000 puta (što znači da bi se na osi moglo napisati da su jedinice u mV), ali PPG nema jedinicu (signal je dobijen direktno sa A/D konvertora, te bi se moglo samo pisati da je jedinica u V). Prikaz grafika na Sl. 180 ne odgovara preporukama za dobru vizuelizaciju, jer boje nisu odgovarajuće (dodate su automatski podrazumevane boje: crvena i plava) i PPG signal je drugačije razmere te se jedva primećuje na slici.

Kako bi se unapredio prikaz grafika na Sl. 180, postoji mogućnost dodavanja *faceting* opcije u *ggplot2* paketu koja se odnosi na tzv. *layout panels* i služi da se prikaže više panela na jednom grafiku primenom sledećeg sloja:

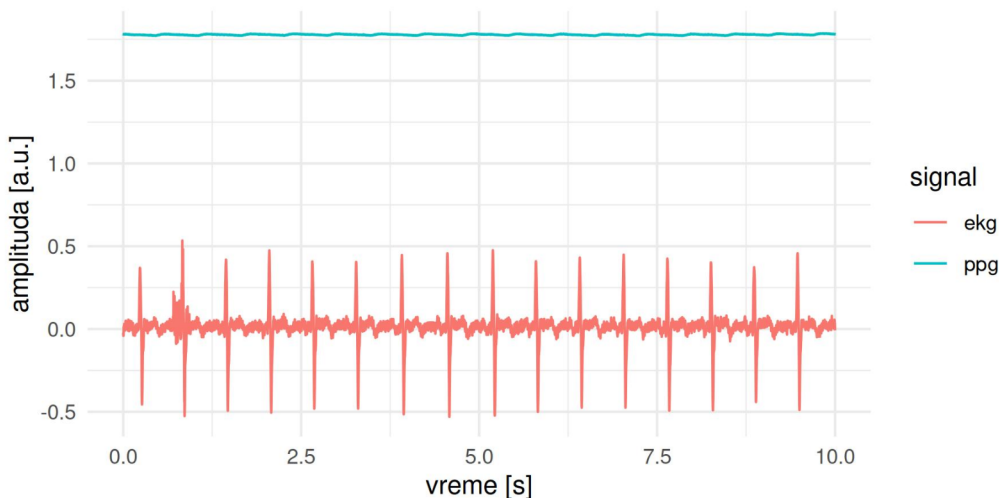
```
facet_wrap(~signal)
```

koji u kombinaciji sa skraćenim prikazom vremenske ose (do 7 s umesto do 10 s) rezultuje grafikom na Sl. 181. *Faceting* se odnosi na podelu grafika na

prozore³⁵⁹, što je korisno kod prikaza većeg broja signala ili podskupa nekog signala u slučajevima kada je potrebno porediti te signale. Za podelu grafika na prozore koristi se i funkcija:

`facet_grid()`

koja za dva signala daje isti rezultat, ali je treba zameniti sa `facet_wrap()` kod prikaza tri ili više signala, jer je omogućena bolja iskorišćenost prostora na grafiku.



Slika 180, Izgled PPG i EKG signala za prvih 10 s. Prikaz odgovara `hold on`; komandi u Matlab / GNU Octave programskim jezicima ili `par(new = TRUE)` opciji za osnovni prikaz u R-u.

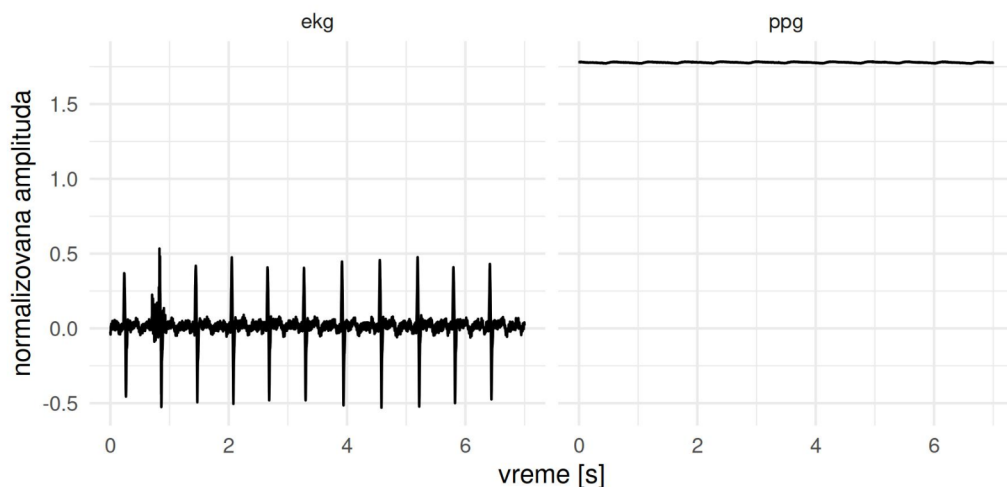
Iako je prikaz biosignala na Sl. 181 nešto pregledniji i dalje nije idealan, jer je cilj vizuelizacije utvrditi da li PPG signal može da meri puls korišćenjem elektrofiziološkog pojačavača za merenje EKG signala, što nije uočljivo na grafiku. Jedan od načina da se signali prikažu tako da se mogu porediti događaji koji se mere jeste da se primeni normalizacija amplitude. Za prikaz signala na Sl. 182 i Sl. 183 iskorišćena je sledeća transformacija:

```
pod$odbirci <- c(dat$V1/mean(dat$V1), ((dat$V2 - mean(dat$V2))*4000)+40)
```

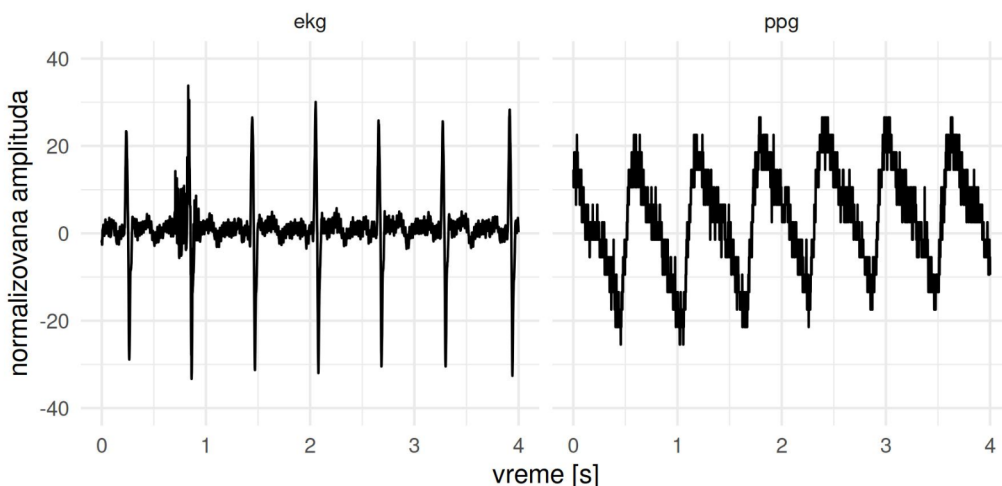
Odabrana transformacija nema mnogo praktičnog smisla, osim što omogućava da prikaz bude odgovarajući. Primetiti da je na Sl. 183 promenjena boja prikaza grafika, tako da je EKG signal u sličnoj boji kao i pozadina panela, dok je PPG

³⁵⁹ U Matlab / GNU Octave programskim jezicima ovakav prikaz odgovara `subplot()` funkciji. Željena funkcionalnost se može postići korišćenjem `par()` funkcije u R-u.

signal prikazan sa dobrim kontrastom. Ovakav prikaz nosi poruku da je cilj posmatranje PPG signala koji se poredi sa referentnim merenjem (EKG signalom – kao u slučaju Minarove mape).



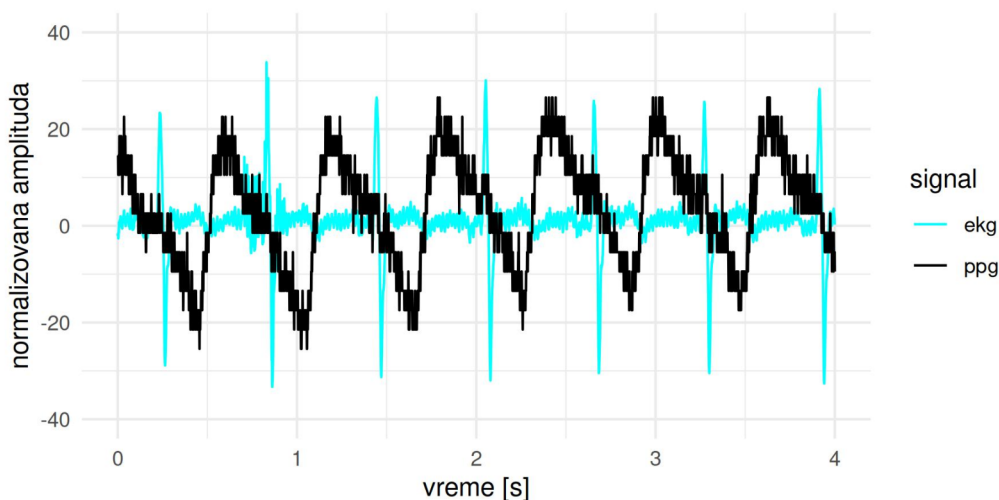
Slika 181, Prikaz EKG i PPG signala na odvojenim panelima primenom *facet_wrap()* funkcije.



Slika 182, Prikaz EKG i PPG signala pomoću *facet* opcije na dva panela za jednostavno poređenje kod sinhronog merenja.

EKG i PPG signali prikazani na Sl. 180, Sl. 181, Sl. 182 i Sl. 183 nisu filtrirani, a na njima se uočava visokofrekventan šum koji potiče od mišićne aktivnosti i šuma napajanja, te bi imalo smisla prilikom predstavljanja vremenskih serija prikazati filtrirane signale. Takođe, na PPG signalu koji je relativno malog

opsega uočava se stepenasta promena amplitude koja odlikuje signal sa šumom kvantizacije. Dodatno, kašnjenje pika PPG signala u odnosu na R pik EKG signala, što se jasno vidi na Sl. 183, je očekivano i naziva se elektromehaničko kašnjenje [491].



Slika 183, Istovremeni prikaz EKG i PPG signala nakon transformacije tj. normalizacije amplitude ovih signala.

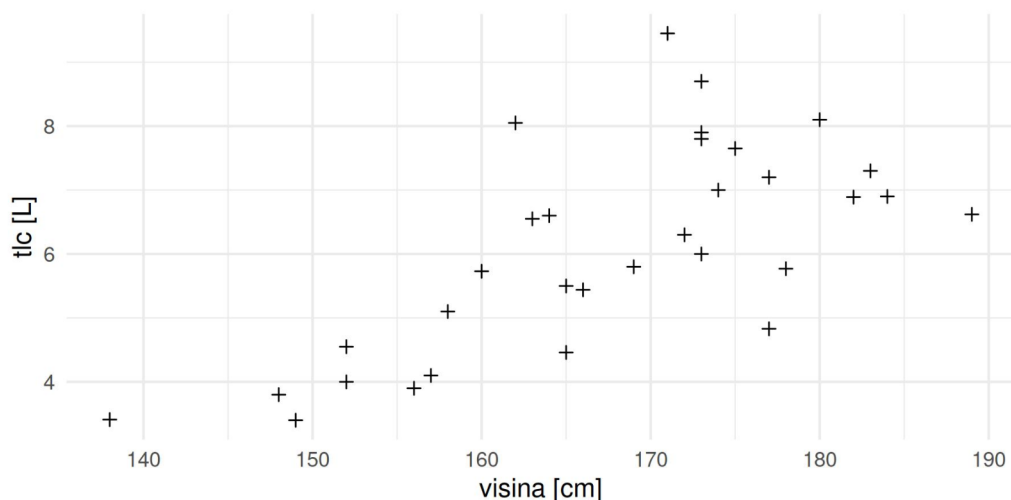
4.1.4.2 Prikaz zavisnosti biomarkera u ggplot paketu

Na Sl. 184 je prikazana zavisnost visine ispitanika od ukupnog kapaciteta pluća (podaci `tlc` u ISwR paketu [36]), primenom sledećeg ggplot koda:

```
gr <- ggplot(tlc, aes(x = height, y = tlc)) + geom_point(shape = 3) +
  xlab("visina [cm]") + ylab("tlc [L]") +
  theme_minimal()
print(gr)
```

Primititi, da nije moguće prikazati grafik snimljen u promenljivu `gr`, u slučaju da se vrednost ggplot grafika čuva unutar promenljive. Za prikaz grafika, potrebno je iskoristiti direktno (primenom `print()` funkcije) ili indirektno štampanje (pokretanjem linije koda sa nazivom promenljive). Za odabir tipa odbiraka može se koristiti argument `shape`³⁶⁰ unutar `geom_point()` funkcije.

³⁶⁰ Podrazumevan oblik odbirka na grafiku je kružić (`shape` ima vrednost 0). Za kompletan spisak oblika tačaka pogledati <http://www.sthda.com/english/wiki/ggplot2-point-shapes>, pristupljeno 06.05.2024.



Slika 184, Prikaz zavisnosti visine ispitanika od ukupnog kapaciteta pluća (eng. *total lung capacity*, skraćeno tlc).

Uočava se da zavisnost nije u potpunosti linearna (Sl. 184), ali da bi se potencijalno mogla aproksimirati linearnim modelom. U literaturi je pokazano da je visina ispitanika robustan parametar za predikciju kapaciteta pluća [492], dok je još 1961. godine utvrđeno da kapacitet pluća i visina na treći stepen imaju relativno veliku i značajnu kros-korelaciju [493], te je rezultat prikazan na Sl. 184 očekivan³⁶¹. Kod ispitivanja relacija između dve promenljive, kao što je to slučaj sa visinom ispitanika i kapacitetom pluća, potrebno je uzeti u obzir i praktični značaj parametara kada se posmatraju ovakvi grafici. U ovom slučaju, pitanje je da li ima smisla da visina ispitanika i kapacitet pluća imaju linearnu zavisnost? Neka od konkretnih pitanja/pretpostavki ovakve i sličnih analiza su:

- Da li je linearna zavisnost očekivana i zašto?
- Da li je nelinearna zavisnost očekivana i zašto?
- Da li je dobijen očekivan rezultat?
- Ako nije dobije očekivan rezultat, zašto?

Grafik zavisnosti dve promenljive je korisna alatka da se odgovori na postavljena pitanja, ali dodatno ima smisla iskoristiti ggplot funkciju:

`geom_smooth()`

³⁶¹ Potrebno je rezultate i poređenje sa Sl. 184 uzeti sa dozom rezerve. Prvi razlog je što su podaci iz tlc studije mereni na pacijentima koji su na listi čekanja za transplantaciju srca i pluća [36]. Drugi razlog je što su podaci iz literature mereni na zdravim ispitanicima, a ne na starijim osobama i pacijentima [492-493].

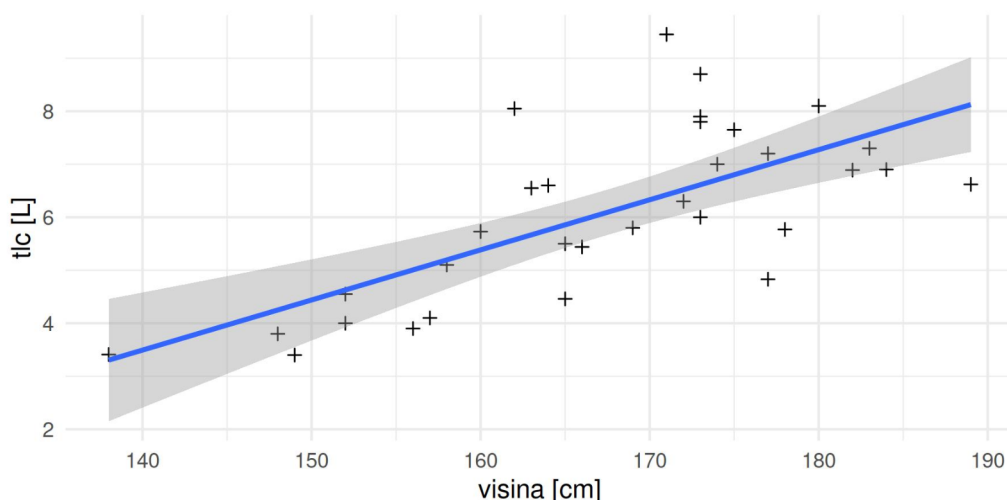
koja služi da se jednostavnim dodavanjem sloja:

```
gr + geom_smooth(method = lm)
```

prikaže procenjen linearni model (lm od eng. *linear model*) na grafiku ili da se dodavanjem sloja:

```
gr + geom_smooth(method = loess)
```

prikaže potencijalna nelinearna zavisnost primenom lokalne polinomijalne regresije (loess od eng. *Local Polynomial Regression Fitting*).

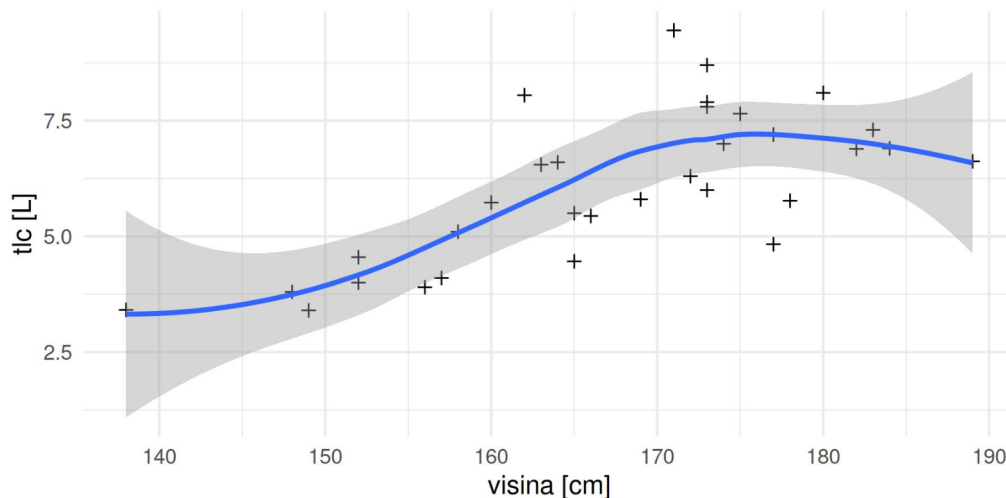


Slika 185, Procena linearnog modela za zavisnost visine ispitanika od kapaciteta pluća.

Alternativno, da bi se proverilo postojanje linearne zavisnosti, odnosno nezavisnosti, može se koristiti ugrađena funkcija *lm()* u programskom jeziku R u kombinaciji sa *predict()* funkcijom. Kako su ovo uobičajeni koraci u analizi parametara, to je provera linearnosti modela ugrađena u ggplot paket. Primećuje se da je na Sl. 185 prikazan i interval poverenja na slici koji odgovara intervalu od 95% verovatnoće u linearnom modelu.

Linearni model je dobra polazna tačka, ali za slučaj da je utvrđeno da linearna zavisnost nije izražena, moguće je pristupiti drugim metodama modeliranja relacija između dve promenljive. Na Sl. 186 je prikazano šta se dobija ako se primeni loess metoda, odnosno fitovanje lokalnom polinomijalnom regresijom. Ako se uporede grafici na Sl. 185 i Sl. 186 može se doći do zaključka da loess model više odgovara zavisnosti visine ispitanika od kapaciteta pluća u

poređenju sa linearnim modelom. Takođe, može se primetiti da ukupan skup podataka ima promenljivu zavisnost – u nekim segmentima ona deluje skoro savršeno linearno (kao od oko 160 cm do oko 170 cm), dok u nekim ni loess model ne može idealno da predstavi ovu zavisnost (kao u granicama blizu 140 cm i 190 cm).

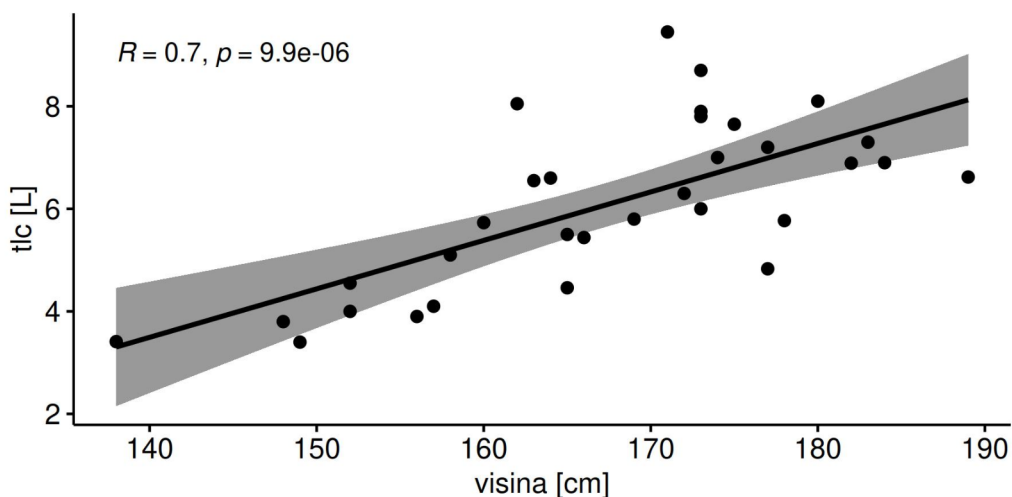


Slika 186, Procena polinomijalnog modela za zavisnost visine ispitanika od kapaciteta pluća.

Iako nije tema ovog poglavlja, linearna zavisnost se kvantitativno ispituje korišćenjem kros-korelacionih koeficijenata. U upotrebi su Spirmanov neparametarski test i Pirsonov parametarski test³⁶² za ispitivanje koeficijenta kros-korelacije. Iako se za ispitivanje linearne zavisnosti koristi neparametarski Spirmanov test³⁶³ za ispitivanje koeficijenta korelacije, to nije dobar pristup, jer ovaj parametar ocenjuje monotonu zavisnost između dve promenljive koja ne mora nužno biti linearna. Pirsonov koeficijent je osetljiv na autlajere, dok Spirmanov koeficijent nije. Dodatno, Pirsonov koeficijent kros-korelacije se zasniva na poređenju srednjih vrednosti i standardnih devijacija koje se mogu koristiti za računanje Gausove raspodele (pretpostavka je pre primene Pirsonovog kros-korelacionog koeficijenta da podaci imaju Gausovu raspodelu), a Spirmanov se bazira na poređenju medijana i kvartila koji se mogu računati za bilo koju raspodelu. S obzirom da je Spirmanov koeficijent manje osetljiv na ekstremume, on ne predstavlja nužno linearan model, već monoton model koji bi potencijalno mogao da se linearizuje (ako za tim postoji potreba). [436, 494]

362 https://en.wikipedia.org/wiki/Pearson_correlation_coefficient, pristupljeno 07.05.2024.

363 https://en.wikipedia.org/wiki/Spearman%27s_rank_correlation_coefficient, pristupljeno 07.05.2024.



Slika 187, Prikaz Pirsonovog kros-korelacionog koeficijenta R sa verovatnoćom p i linearnog modela između visine ispitanika i kapaciteta pluća primenom ggpubr paketa.

Za ispitivanje zavisnosti primenom koeficijenata kros-korelacije, postoji poseban paket koji je povezan za ggplot paketom: ggpubr paket [495]. Na Sl. 187 je prikazana linearna zavisnost sa intervalom poverenja od 95% (što odgovara prikazu na Sl. 1185) koji se dobija primenom ggplot paketa. Dodatna informacija koja je prikazana na grafiku na Sl. 187 je to što ggpubr paket uključuje prikaz kros-korelacionog koeficijenta R i verovatnoće p (koja ukazuje na statistički značajnu linearnu zavisnost visine ispitanika od kapaciteta pluća), primenom sledećeg koda:

```
ggscatter(tlc, x = "height", y = "tlc",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "pearson",
          xlab = "visina [cm]", ylab = "tlc [L]")
```

Iako `ggscatter()` funkcija iz ggpubr paketa je vrlo slična `ggplot()` funkciji, uočava se izostanak slojevitog funkcionalnog programiranja, jer nema operatora "+". Dodatno, argumenti `ggscatter()` funkcije intuitivno ukazuju da se mogu odabrati pojedinačno elementi koji će se prikazati na slici, kao što je, na primer, metoda kros-korelacije (argument `cor.method`). Za primenu osnovnih R funkcija za računanje kros-korelacionog koeficijenta, koriste se:

`cor()` i `cor.test()`

funkcije, gde je jedina razlika što *cor.test()* funkcija kao rezultat daje i varovatnoću p , koja omogućava da se proceni da li je zavisnost dve promenljive statistički značajna. Argumentom *method*, moguće je odabrati tip korelacije (*spearman* ili *pearson*). Primenom ove dve funkcije, dobija se da je za visinu ispitanika i kapacitet pluća, Pirsonov koeficijent 0.695, dok je Spirmanov 0.672, kao i da su oba statistički značajna ($p < 0.001$). Provera usaglašenosti modela sa unetom pretpostavkom se može obaviti i proverom reziduala, odnosno razlike između merenih/dostupnih vrednosti i predikcije modela. Ako je model dobro odabran, reziduali bi trebalo, pored ostalog, da imaju srednju vrednost nula i da budu normalno raspodeljeni [496], ali ako se njihovom vizuelizacijom utvrdi da postoji određen šablon ili obrazac umesto da se reziduali pojavljuju po redosledu koji nije pseudoslučajan, to znači da model nije dobro odabran [497].

U slučaju da je potrebno ispitati kros-korelacione koeficijente za veći broj promenljivih, 2D grafici zavisnosti gube na smislu, jer je moguće iskoristiti samo boju za treću dimenziju. Ali, ako postoji četiri ili više parametara, onda je poželjnije koristiti kros-korelacione matrice i grafički prikaz kros-korelacionih koeficijenata. Ovo je važan pretkorak kod primene metoda za smanjenje dimenzionalnosti skupa parametara, ispitivanje uticaja većeg broja promenljivih na konačan ishod (u terapiji ili dijagnostici) i za primenu metoda mašinskog učenja. Na primer, pokazano je da modeli mašinskog učenja kao što je model slučajnih šuma (eng. *Random Forest*, skraćeno RF) su veoma osetljivi na kros-korelacije inicijalnog skupa. Srećom, kod RF modela relativno velika kros-korelacija između obeležja ne utiče na uspešnost klasifikacije (koja je najčešće merena tačnošću klasifikacije i F1 parametrom), ali utiče na važnost obeležja [498]. U biomedicinskom inženjerstvu, kros-korelacija se koristi i za ocenu preslušavanja kod biosignala što je detaljnije prikazano u sledećem potpoglavlju.

4.1.4.3 Prikaz kros-korelacionih matrica za ocenu preslušavanja biosignala i interaktivni grafici

Preslušavanje elektrofizioloških signala (eng. *crosstalk*³⁶⁴) nastaje kada se jedan signal od interesa meri u neposrednoj blizini drugog signala od interesa. U ovom udžbeniku je već prikazan primer takvog merenja, gde je EKG signal smatran šumom tj. nepoželjnom komponentom na EMG signalu. Međutim, u opštem slučaju oba signala mogu biti od interesa za konačnu analizu. Takođe, preslušavanje može da se javi prilikom merenja EMG signala sa mišića koji se

364 <https://en.wikipedia.org/wiki/Crosstalk>, pristupljeno 06.05.2024.

nalazi u neposrednoj blizini drugih mišića; kada se površinskim elektrodama na stomaku trudnice mere istovremeno EKG trudnice, EKG fetusa i električna aktivnost materice; kada se meri EEG signal u frontalnoj regiji može doći do preslušavanja sa EOG signalom; a preslušavanje se javlja i kada se meri EEG kod novorođenih beba, zbog izloženosti elektroda krvnim sudovima, jer je proces okoštavanja u toku, što za posledicu ima činjenicu da je EEG signal kontaminiran EKG signalom tj. srčanim otkucajima; ili kada se meri višekanalni EEG signal može doći do preslušavanja usled toga što se elektrode nalaze u neposrednoj blizini. [237, 274, 337, 340, 499]

Treba napomenuti da je preslušavanje u obradi i merenju signala širi pojam nego preslušavanje elektrofizioloških biosignala, te može označavati i druge pojave. Takođe, u biomedicinskom inženjerstvu je prisutan poseban tip preslušavanja, a to je preslušavanje u pejsmejkeru (eng. *pacemaker crosstalk*³⁶⁵) i nastaje kada se stimuliše srčani mišić u jednoj komori, a to se detektuje kao signal u drugoj komori, odnosno na mestu na kome ne bi trebalo da postoje spolja generisani električni signali. [500]

U literaturi je preslušavanje u velikoj meri pokriveno za EMG signale [501-502], a jedan od načina da se proceni preslušavanje EMG signala jeste kros-korelacija. Naime, ako je došlo do preslušavanja očekuje se da kros-korelacioni koeficijent bude veći, a ako nije došlo do preslušavanja očekuje se da je kros-korelacioni koeficijent manji [503]. Za ilustraciju kros-korelacionih matrica, korišćeni su podaci iz studije u kojoj je meren 8-kanalni EMG signal sa mišića podlaktice sa ciljem prepoznavanja pokreta ruke u uslovima rotacije i translacije niza EMG elektroda [340, 504]. U ovom slučaju je posebno značajno proveriti kros-korelaciju parametara, jer elektrode nisu bile ciljano postavljene na mišiće prateći SENIAM protokol [312], već je cilj merenja bio da se pokaže da je moguće kružnom postavkom elektroda pokriti odgovarajući broj mišića podlaktice. Očekivano, više susednih elektrode je pokrivalo površinu istog mišića ili istih mišića, te za klasifikaciju pokreta nije bilo potrebno koristiti parametre EMG signala sa svih 8 kanala, već je za redukciju dimenzionalnosti korišćenja analiza glavnih komponenti (eng. *Principal Component Analysis*, skraćeno PCA). Da bi se utvrdilo da li je PCA metoda opravdano primenjena, moguće je proveriti postojanje kros-korelacija između susednih elektroda. Za ilustraciju postojanja kros-korelacija korišćena je datoteka "ID1_PS_P1.txt" iz dostupne baze podataka [340, 504].

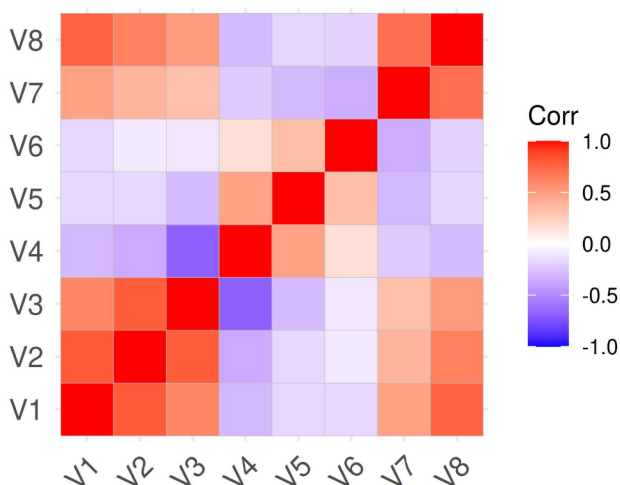
365 https://en.wikipedia.org/wiki/Pacemaker_crosstalk, pristupljeno 06.05.2024.

Za potrebe iscrtavanja kros-korelacionih matrica, postoji paket srodan ggplot paketu – ggcorrplot [466]. Da bi se primenile funkcije ovog paketa za prikaz kros-korelacionih matrica, potrebno je da se prvo izračuna matrica. Ako su podaci učitani u *data frame* pod nazivom *dat*, onda se primenom:

```
corM <- cor(dat)
```

dobija promenljiva *corM* dimenzija 8×8 koju je potrebno predstaviti na grafiku tako da se boje iskoriste za prikaz vrednosti kros-korelacionog koeficijenta. Da bi se dobila matrica prikazana na Sl. 188 potrebno je izvršiti samo funkciju:

```
ggcorrplot(corM)
```



Slika 188, Kros-korelaciona matrica za merenje EMG signala sa 8 kanala koji su označeni sa V1, V2, ..., V8. Podaci su preuzeti iz [340, 504].

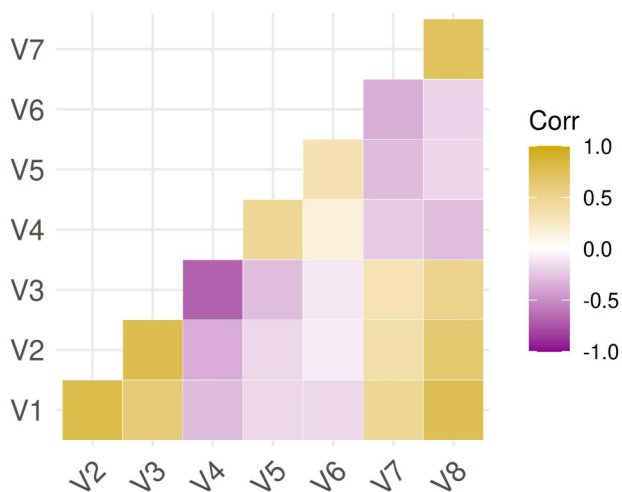
Matrica prikazana na Sl. 188 je simetrična, te je dovoljno prikazati samo deo slike iznad ili ispod dijagonale. Dodatno, za odabir boja koje mogu da razlikuju i ljudi koji imaju problema sa prepoznavanjem boja, postoje specijalizovani R paketi sa preporučenim paletama boja (eng. *friendly color palette*) [505]. Da bi se dobio prikaz samo polovine matrice sa odgovarajućim nijansama boje, kao na Sl. 189, potrebno je koristiti:

```
ggcorrplot(corM, type = "lower", outline.col = "white",
           colors = c("darkmagenta", "white", "gold3"))
```


u R-u podešavanjem argumenata *type* (odabran je “lower” za prikaz vrednosti matrice ispod dijagonale), *outline.col* (boja za oznake iznad dijagonale), kao i *colors* za odabir tri boje (maksimum i minimum kros-korelacionih koeficijenata, kao i za vrednosti u okolini 0). Kada se izračuna matrica, sa svim mogućim kombinacijama kros-korelacionih koeficijenata, ona se može predstaviti jednostavnim funkcijama u R-u za prikaz slike:

image()

ali je značajno bolje da se koristi specijalizovan paket za kroz korelacione matrice, odnosno *ggcorrplot*, jer pruža mnogo više opcija, uključujući prikaz kolor bara.

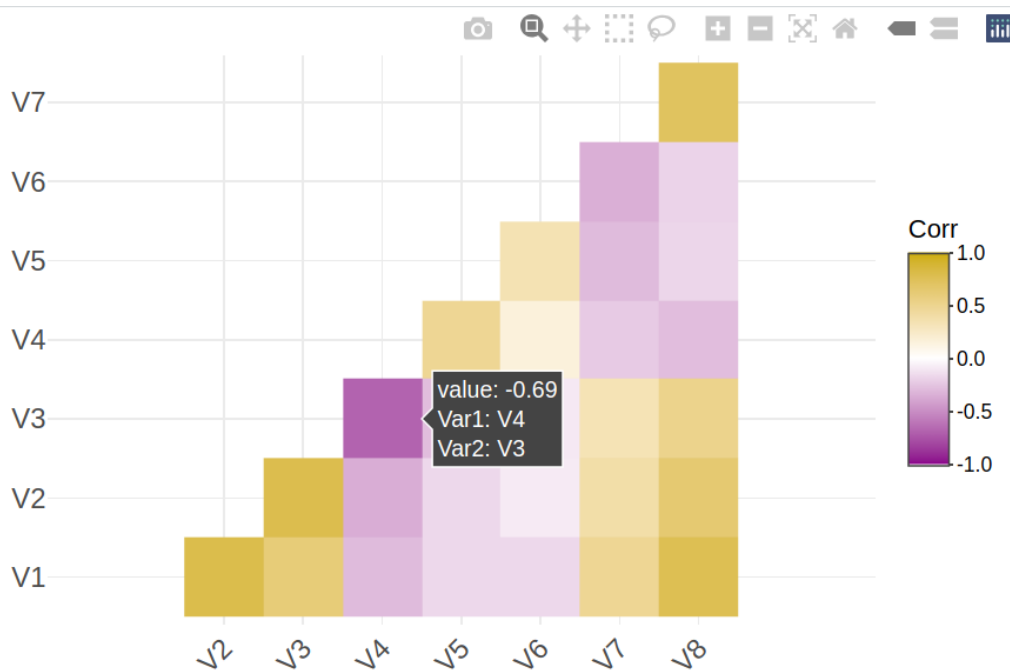


Slika 189, Kros-korelaciona matrica za merenje EMG signala sa 8 kanala sa Sl. 188 koji su označeni sa V1, V2, ... , V8. Prikazana je samo polovina matrice (ispod glavne dijagonale), a boje su odabrane tako da je omogućeno da ih jednostavno razlikuju i ljudi koji imaju problem da razlikuju boje.

Postoji mogućnost podešavanja unutar *ggcorrplot()* funkcije, tako da se u svakom kvadratu upišu vrednosti kros-korelacija (argument *lab* treba postaviti na TRUE vrednost unutar funkcije)³⁶⁶. Dodatno, moguće je da se interaktivno pregledaju vrednosti kros-korelacionih koeficijenata (postavljanjem strelice miša iznad odgovarajućeg polja, moguć je prikaz vrednosti kros-korelacionog

³⁶⁶ Ostala podešavanja izgleda kros-korelacione matrice se mogu pogledati na <http://www.sthda.com/english/wiki/ggcorrplot-visualization-of-a-correlation-matrix-using-ggplot2>, pristupljeno 06.05.2024.

koeficijenta u tom polju, kao na Sl. 190). Interaktivnost je popularna, jer omogućava da se prikaže veće količine informacija na grafiku.



Slika 190, Prikaz interaktivne kros-korelacione matrice za prikaz sa Sl. 189. Meni za zumiranje, čuvanje slike i druga podešavanja interaktivnog prikaza je predstavljen u gornjem desnom uglu.

Da bi grafici, koji se proizvedu primenom funkcija u ggplot2 paketu, postali interaktivni potrebno je pozvati sledeću komandu u R-u:

```
ggplotly(ggplot2::last_plot())
```

iz plotly paketa [463], te će grafik postati interaktivan, kao što je ilustrovano na Sl. 190. Interaktivnost se ogleda u zumiranju, prikazu brojnih vrednosti promenljivih koje su predstavljene na apscisi i ordinati samo prelaskom računarskog miša preko određenog dela grafika. Na Sl. 190 je ilustrovano kako se prelaskom miša preko polja koje ima vrednost kros-korelacije -0.69 dobija prikaz vrednosti u polju, kao i koordinate odabranog polja (V3 i V4). Primećuje se da su elektrode V1 i V2, kao i V7 i V8, ali i V3 i V4 verovatno bile postavljene na istim mišićima, jer relativno veći kros-korelacioni koeficijenti ukazuju na potencijalno veće preslušavanje (prikazani tamnijim nijansama ljubičaste i zlatne). Ostali parovi elektroda nisu pokazali tako velike kros-korelacione koeficijente verovatno zato što nisu bile susedne, odnosno nisu bile

postavljene u neposrednoj blizini (kao V2 i V5), i verovatno iz razloga što su bile postavljene preko kosti (mesto na nadlaktici gde nema izražene mišićne mase).

4.1.5 Odabrani primeri grafika i drugi R paketi

Važan element u predstavljaju grafika su kompozitne slike koje mogu da sadrže prikaz signala, proračunate parametre, dodatne oznake posebnih elemenata, brojne vrednosti, ali i skice i/ili fotografije. Jedan takav primer (merenje akcionog potencijala kod biljaka) je prikazan na Sl. 48. Za prikaz ovakvog grafika potrebno je koristiti jpeg paket [506], koji se može koristiti za učitavanje slika (u ovom slučaju fotografije biljke mesožderke) primenom funkcije:

```
slika <- readJPEG("muholovka.jpg")
```

dok se za prikaz slike na grafiku koristi funkcija:

```
rasterImage()
```

iz osnovnog R paketa (graphics) kojom se učitana slika smešta na odgovarajuće mesto (unose se koordinate *xleft*, *ybottom*, *xright* i *ytot* kao argument funkcije). Kod koji omogućava da se dobija Sl. 48 je prikazan na Sl. 191.

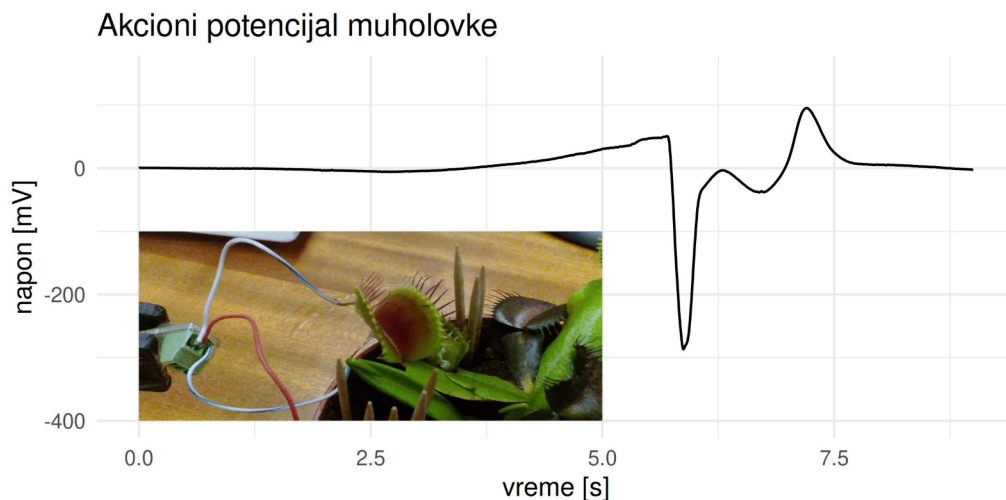
```
dat <- read.table("plant.csv")
fs <- 1000 # frekvencija odabiranja
dat$time <- seq( 0, (length(dat$V1) - 1)/fs, by = 1/fs )

# prikaz kompozitne slike (pojačanje je bilo 100 puta)
plot(dat$time, 10*dat$V1, type = "l",
      xlab = "vreme [s]", ylab = "napon [mV]",
      main = "Akcioni potencijal muholovke",
      xlim = c(0, 9), ylim = c(-400, 150), lwd = 2)
grid()
rasterImage(slika, 0, -400, 5, -100)
```

Slika 191, R skripta za prikaz kompozitne slike (pogledati Sl. 48).

Funkcije koje se koriste za prikaz vremenske ose, kao i za učitavanje podataka (frekvencija odabiranja je bila 1000 Hz, a pojačanje 100 puta i podaci su

smešteni u datoteci “plant.csv”) su pokazane u ranijim poglavljima. Ista funkcionalnost dodavanja slika (eng. *raster*) postoji u ggplot paketu koji omogućava još pregledniju prezentaciju kompozitne slike, koja prikazuje merenje električnih potencijala biljke muholovke (Sl. 192).



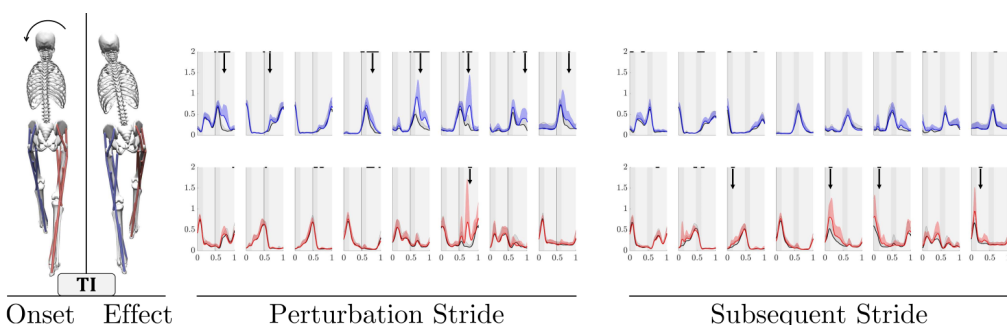
Slika 192, Kompozitni prikaz električnih potencijala merenih na biljci muholovki zajedno sa fotografijom merne aparature korišćenjem funkcija iz ggplot paketa. Pogledati alternativni prikaz na Sl. 48. korišćenjem funkcija iz osnovnog R paketa (Sl. 191).

Sloj na Sl. 192, koji do sada nije bio objašnjen, u okviru ggplot funkcionalnosti se dobija primenom:

```
annotation_raster(slika, 0, 5, -400, -100)
```

funkcije. Primititi da je raspored koordinata unutar funkcije *annotation_raster()* različit od rasporeda koordinata funkcije *rasterImage()*. Kombinacijom fotografija i grafika može se efikasnije iskoristiti prostor i slika učiniti jednostavnijom za tumačenje. Najčešće se u biomedicinskom inženjerstvu, kako je ovde i pokazano, kombinuju fotografija procedure merenja (bilo da se radi o mernoj opremi ili o tipu pokreta/fenomena koji se meri) i grafik biosignala ili biomarkera. Dobar primer kompozitne slike koja pokazuje razliku između profila EMG signala za levu i desnu nogu tokom različitih perturbacija tokom hoda je dat na Sl. 193 iz preprinta³⁶⁷ [322].

³⁶⁷ Preprint je verzija naučnog članka koja je dostupna preko interneta, ali se smatra neobjavljenom. Predstavlja preliminarnu verziju članka koju još nisu pregledali naučni recenzenti i koju nije odobrio nijedan naučni časopis. Međutim, sa ciljem što ranije diseminacije, odnosno širenja znanja među naučnicima, preprint je vrlo popularan. Takođe,



Slika 193, Prikaz dela kompozitne slike EMG profila. Za kompletan uvid u sliku, pogledati preprint [322] kao primer dobre prakse. Preprint i slika su podeljeni pod otvorenom licencom CC BY 4.0.

Intuitivno je jasno na Sl. 193 koji EMG profili (usrednjene obvojnice) prikazuju mišiće sa leve, a koji sa desne nove. Takođe, vidi se razlika između srednjih panela koji prikazuju korak (eng. *stride*) u kome je došlo do perturbacije (eng. *perturbation stride*), kao i sledeći korak (eng. *subsequent stride*).

4.1.5.1 Toplotne mape (eng. *heat map*)

Najsličniji grafik kros-korelacionim matricama je toplotna mapa³⁶⁸ ili eng. *heat map*³⁶⁹. Za prikaz kompleksnih podataka, koji imaju relativno veliki broj dimenzija, koristi se toplotna mapa. Postoji više paketa u R-u koji sadrže specijalizovane funkcije za prikaz toplotnih mapa, a na Sl. 194 je prikazana toplotna mapa kros-korelacionih koeficijenata za 8-kanalno EMG merenje signala sa mišića podlaktice (koji su dostupni u datoteci "ID1_PS_P1.txt" iz [340, 504]) primenom funkcije koja dolazi sa osnovnim R paketima:

```
heatmap()
```

koja je pozvana samo sa jednim argumentom - promenljivom *corM*. Mapi prikazanoj na Sl. 194, moguće je promeniti skalu boja i dodati oznake na osama primenom sledećih komandi koje rezultuju prikazom na Sl. 195:

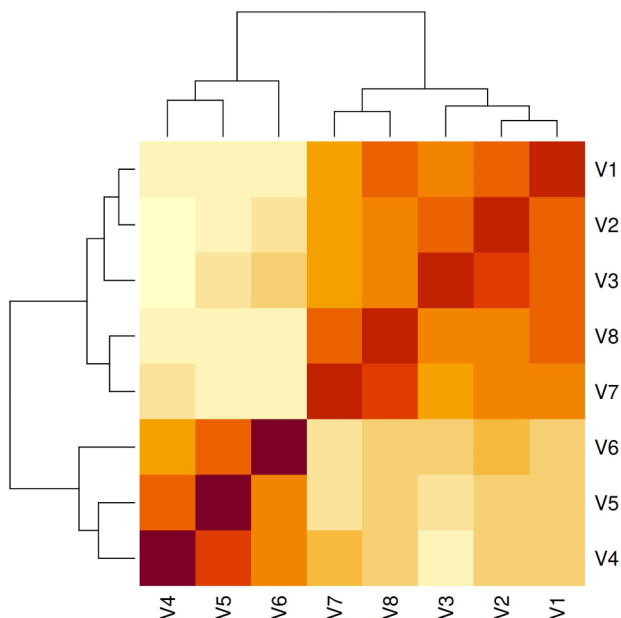
```
mojeBoje <- colorRampPalette(c("cyan", "black"))(20)
```

može da bude zaštita od zlonamernih recenzenata koji odugovlače recenziju da bi u svojim laboratorijama ponovili isto istraživanje. Prvi put se sa primenom takve prakse počelo u oblasti fizike i matematike, da bi se danas psihologija i posebno biologija izdvojile kao oblasti sa velikim brojem preprinta [507]. Najčešće korišćen preprint server za inženjere je arXiv (<https://arxiv.org/>, pristupljeno 07.05.2024).

368 Prevod na srpski nije uobičajen i ređe se koristi.

369 https://en.wikipedia.org/wiki/Heat_map, pristupljeno 02.06.2024.

`heatmap(corM, xlab = "EMG kanali", ylab = "EMG kanali", col = mojeBoje)`



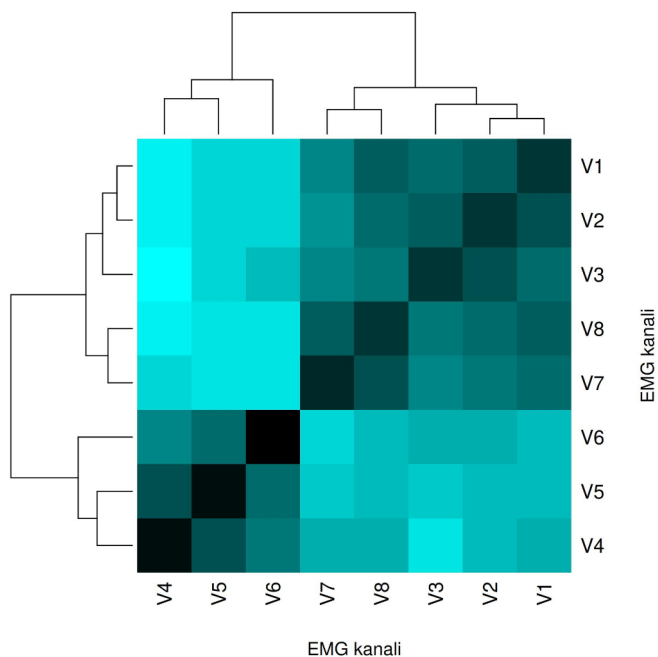
Slika 194, Toplotna mapa za kros-korelacione koeficijente za 8-kanalno merenje EMG signala [340, 504].

Primititi da je za potrebe kreiranja palete boja, iskorišćena ugrađena funkcija `colorRampPallete()` čiji je argument niz dve boje koje omogućavaju da se kreira paleta sa 20 skaliranih boja. Jedna od odlika toplotnih mapa je postojanje dendrograma koji omogućavaju inicijalnu klasifikaciju, odnosno grupisanje kros-korelacionih koeficijenata sa sličnim vrednostima. Takođe, kao rezultat grupisanja tj. klasterizacije izdvajaju se dve grupe kros-korelacija za EMG signale koje međusobno više korelišu:

- u jednoj grupi su V1, V2, V3, V7 i V8, a
- u drugoj grupi su V4, V5 i V6.

Ovakav rezultat grupisanja ima smisla, ako se uzme u obzir da je niz EMG elektroda pokrio dva veća mišića (fleksor i ekstenzor šake), kao i da je njihova aktivacija tokom izvođenja hvata odgovarala recipročnoj inhibiciji (kada je jedan mišić aktivan drugi nije) [340, 504]. Takođe, važno je spomenuti da su V1 i V8 susedni kanali, jer su EMG elektrode bile postavljene kružno, što potkrepljuje činjenicu da se radi o fleksoru i ekstenzoru. Druga, takođe validna pretpostavka bi bila, da je zbog velike sile koju je ispitanik razvio došlo do kokontrakcije (istovremene aktivacije fleksora i ekstenzora nadlaktice), te se u

prvoj grupi nalaze EMG signali sa kanala koji pokrivaju mišićnu regiju, dok su u drugoj grupi EMG signali sa kanala koji nisu pokrili eskcitabilno tkivo. Da bi se utvrdilo koja od navedenih pretpostavki je tačna, potrebno je pogledati pojedinačne kanale, tip pokreta i anatomske markere za postavku niza elektroda na nadlakticu ispitanika.



Slika 195, Toplotna mapa sa označenim osama i sa promenjenom paletom boja u odnosu na Sl. 194.

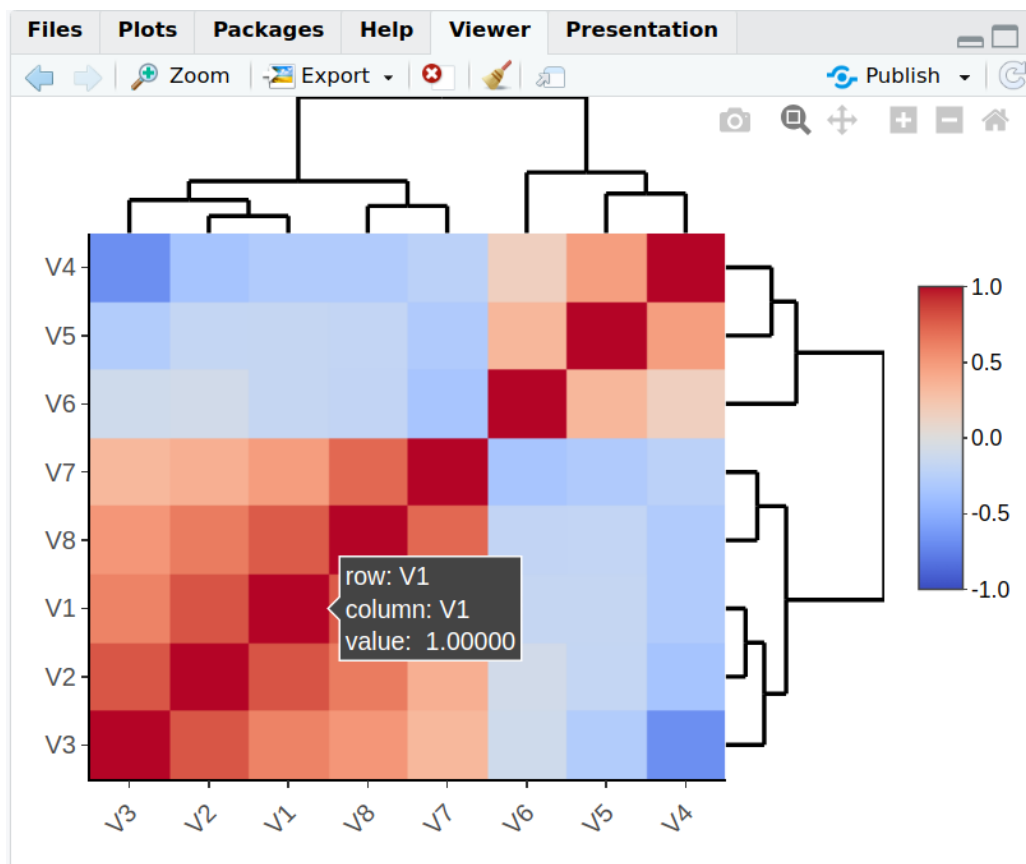
Toplotna mapa primenjuje se za reorganizaciju podataka i dodatni uvid u relacije između obeležja. Kako toplotna mapa omogućava prikaz hijerarhijske klasterizacije³⁷⁰ (može se koristiti kao pretkorak prilikom primene klasterizacije) pomoću dendrograma, za njenu primenu je važan ulazni parametar *Reordering*. U slučaju rada sa višedimenzionim signalima i sa relativno velikim brojem obeležja, u kojima je potrebno pronaći odgovarajuće relacije i odnose (van domena kros-korelacije), toplotne mape predstavljaju dobar izbor. Prikaz toplotnih mapa je posebno koristan u oblasti bioinformatike za prikaz ekspresije gena (eng. *gene expression*) [435].

³⁷⁰ Primeri toplotnih mapa se mogu naći na <https://www.atlassian.com/data/charts/heatmap-complete-guide>, pristupljeno 07.05.2024.

Dva elementa nedostaju toplotnim mapama na Sl. 194 i Sl. 195: (1) prikaz skale boja, kako bi bilo jasno šta znače boje na grafiku i (2) interaktivnost. Oba elementa se mogu dodati primenom funkcije:

```
heatmaply_cor(corM)
```

iz heatmaply paketa [508] kao što je ilustrovano na Sl. 196.



Slika 196, Interaktivna mapa za kros-korelacione koeficijente 8-kanalnog EMG signala dobijena primenom funkcije iz heatmaply paketa.

Jedna od najpoznatijih toplotnih mapa predstavlja prikaz uticaja vakcinacije na broj obolelih od morbila, odnosno malih boginja u SAD, za podatke koji su dostupni na internet adresi Tycho³⁷¹ projekta [509]. U Vol Strit časopisu se u 2015. godini pojavio članak koji je predstavio niz sjajnih interaktivnih vizuelizacija pomoću toplotnih mapa koje su ilustrovale značaj uvođenja

371 <https://www.tycho.pitt.edu/data/>, pristupljeno 07.05.2024.

vakcinacije za niz infektivnih bolesti, uključujući male boginje³⁷² [510]. Od tada se pojavio niz pokušaja³⁷³ da se rekreiraju ove mape, sa posebnim osvrtom na to poboljšanje originalnog prikaza iz [510]. Dva načina popravke grafika su dominantno bila zastupljena: (1) da se prikažu jasno oznake na osama koje odgovaraju nazivima država (da se ne preskaču oznake i da se predstave puni nazivi država) i (2) da se uvede posebna oznaka, boja ili šrafura, koja bi ukazala da neke države nisu imale dostupnu statistiku (originalni grafik je lažno ukazivao da u tim državama nikada nije ni bilo obolelih od malih boginja). Jedan od rimejka, koji je prikazan u ovom udžbeniku na Sl. 197, je pripremio prof. Irizari kao deo materijala za udžbenik [131]. Grafik na Sl. 197 nedvosmisleno ukazuje na dramatično smanjenje broja obolelih nakon primene vakcine protiv malih boginja³⁷⁴. Iako vertikalna linija na toplotnoj mapi na Sl. 197, prikazuje godinu kada je uvedena vakcina (1963. godina), tek 5 godina kasnije, tj. 1968. je predstavljena nova unapređena verzija vakcine [511], koja je verovatno i izazvala najdramatičniji pad u broju obolelih na Sl. 197. Kreatori heatmaply paketa su u inicijalnom radu u kome je predstavljen paket prikazali sopstvenu verziju rekreirane toplotne mape na kojoj se vidi efekat vakcinacije protiv malih boginja [508]. Njihova mapa je dostupna u dodatnim materijalima za grafik, interaktivna je i prikazana je sa dendrogramima. Dvodiomenziona ili 2D vizuelizacija je korisna u analizi biosignala, a za dalje usavršavanje čitaoci se upućuju na viridis paket³⁷⁵ [512] koji sadrži različite panele boja koje mogu biti korisne za prikaz toplotnih mapa.

372 <http://graphics.wsj.com/infectious-diseases-and-vaccines/>, pristupljeno 07.05.2024.

373 Ovde su izlistani samo primeri rekreiranja grafika za epidemiološke podatke u SAD pre i posle uvođenja vakcine protiv malih boginja:

<https://benjaminlmoore.wordpress.com/2015/04/09/recreating-the-vaccination-heatmaps-in-r/> (pristupljeno 07.05.2024), <https://community.jmp.com/t5/The-Plot-Thickens/Measles-Heat-Map-Remakes/ba-p/20989> (pristupljeno 07.05.2024.) i

<https://blogs.sas.com/content/sastraining/2015/02/17/how-to-make-infectious-diseases-look-better/> (pristupljeno 07.05.2024).

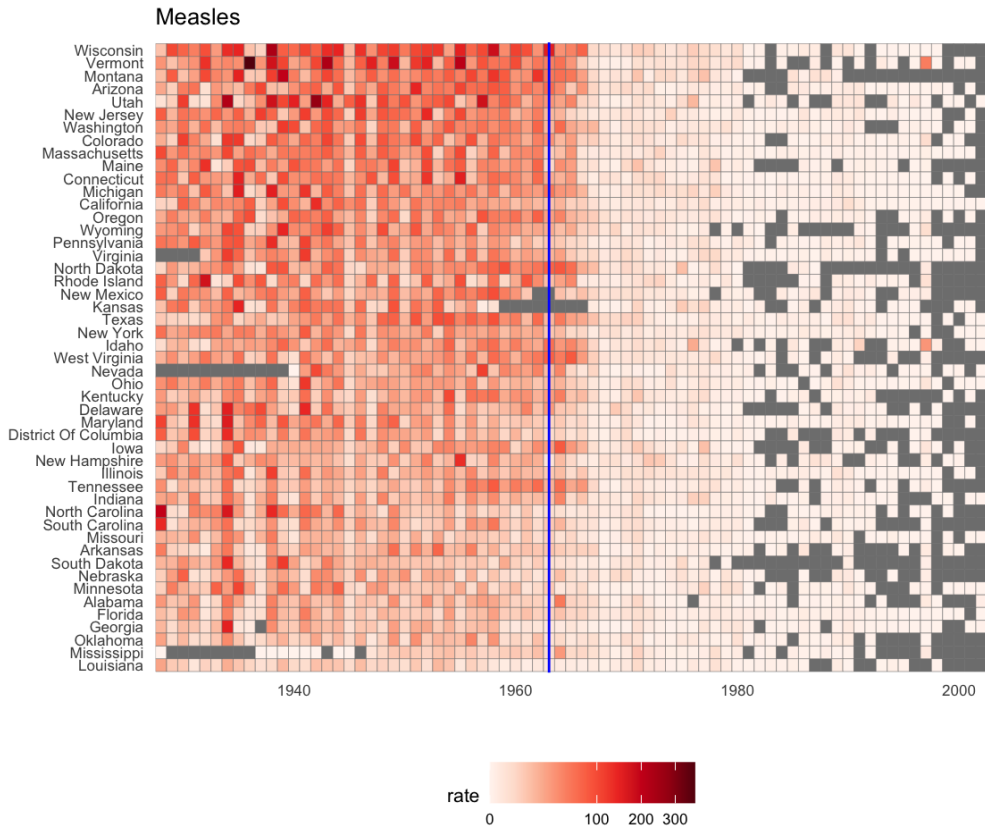
374 https://en.wikipedia.org/wiki/Measles_vaccine, pristupljeno 07.05.2024.

375 Uputstvo za korišćenje viridis paketa sa primerima skala je dat u

<https://cran.r-project.org/web/packages/viridis/vignettes/intro-to-viridis.html>, pristupljeno 05.05.2024.

4.1.5.2 Spektrogram biosignala

Pored toplotnih mapa, postoje drugi specijalizovani grafici koji se koriste u analizi podataka. Na primer, spektrogram se koristi za vizuelizaciju nestacionarnih signala u vremensko-frekvencijskom domenu. Promene koje se mogu uočiti na spektrogramima su, na primer, postojanje mišićne kontrakcije (promena frekvencijskog sadržaja u različitim vremenskim intervalima usled kontrakcije mišića).

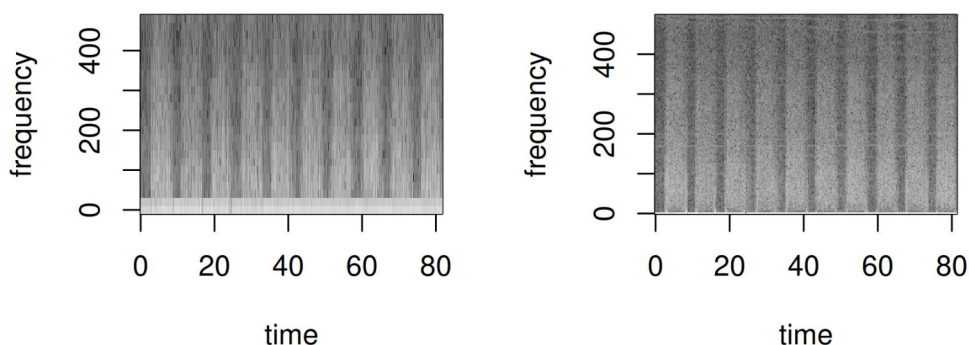


Slika 197, Prikaz broja obolelih od boginja (morbila) nakon uvođenja vakcine 1963. godine. Slika je preuzeta iz udžbenika [131] koji je podeljen pod otvorenom CC-BY-NC-SA 4.0 licencom³⁷⁶. Direktno pristup slici je moguć na https://rafalab.dfci.harvard.edu/dsbook-part-1/dataviz/dataviz-in-practice_files/figure-html/vaccines-plot-1.png, pristupljeno 07.05.2024.

Najčešće se za prikaz spektrograma prikazuju dva grafika: (1) jedan spektrogram se prikazuje sa boljom vremenskom rezolucijom (spektrogram se

³⁷⁶ <https://github.com/rafalab/dsbook-part-1/tree/main>, pristupljeno 07.05.2024.

računa na kraćim vremenskim intervalima) i (2) drugi spektrogram se predstavlja sa boljom frekvencijskom rezolucijom (računa se na prozorima koji su odabrani u relativno dugim vremenskim intervalima). Iako je zastupljeno u nekim naučnim radovima, veoma je pogrešno izdvajati obeležja iz spektrograma, zato što spektrogrami služe da se proverí kakve su frekvencijske promene u vremenu na relativno grub način. Iz tog razloga, nije opravdano koristiti metode za obradu slike na spektrogramima, jer da bi se prikazale istovremeno informacija o vremenu i frekvenciji, jedna od njih se mora smanjiti. Određivanje vremenskog intervala na kome će se računati spektrogram je stvar kompromisa – ako je cilj da se dobro vide početak i kraj mišićne kontrakcije, onda se frekvencijski sadržaj neće videti u dobroj rezoluciji i obrnuto. Ovaj tip grafika služi isključivo da se vizuelno predstavi frekvencijski i vremenski sadržaj signala, istovremeno na jednom grafiku, ali nikako sa ciljem da se spektrogram analizira primenom metoda koje se koriste u obradi slike.



Slika 198, Spektrogram EMG signala sa prvog kanala iz datoteke "ID1_PS_P1.txt" [340, 504] za širinu prozora od 50 ms, odnosno 50 odbiraka, na levom panelu i za širinu prozora od 500 ms, odnosno 500 odbiraka, na desnom panelu.

Za odabran vremenski prozor (na kome se pretpostavlja da je signal stacionaran), primenjuje se kratkotrajna Furijeova transformacija (eng. *Short-Time Fourier Transform*, skraćeno STFT) iz koje se dobijaju podaci o magnitudi ili amplitudi (koja je kodirana bojom) i frekvenciji koja se najčešće prikazuje na ordinati, dok je vreme prikazano na apscisi. Na Sl. 198 su prikazana dva spektrograma za EMG signal tako što je na levom panelu spektrogram čija je vremenska rezolucija bolja (prozor je širine 50 ms, odnosno odbiraka³⁷⁷), a za spektrogram prikazan na desnom panelu bolja je frekvencijska

³⁷⁷ U literaturi se može često sresti podatak o broju tačaka na kome je računata STFT, ali je to pogrešan pristup, jer broj tačaka može da odredi kratke ili duge intervale u zavisnosti od

rezolucija (odabran je prozor širine 500 ms). Rezolucija je, u slučaju merenja, najmanja promena koju može da prikazuje ili registruje instrument [78]. U slučaju spektrograma, to je najmanji vremenski korak/prozor Δt , odnosno najmanja promena frekvencija Δf , koji se može prikazati. Frekvencijska rezolucija je jednaka opsegu frekvencija koji može prikazati dati prozor i jednaka je fs/N , gde je fs frekvencija odabiranja, a N broj odbiraka na prozoru. Na prikazanim spektrogramima na Sl. 198, koji su dobijeni primenom:

specgram(dat\$V1, n = 50³⁷⁸, Fs = fs)

funkcije iz signal paketa [275], se uočavaju frekvencije u Hz na ordinati i vreme u s na apscisi. Oznake i proračuni vremena su automatske, zato što je jedan od ulaznih argumenata (pored niza podataka i broja odbiraka prozora n) i frekvencija odabiranja (u oznaci Fs). Takođe, pored podešavanja osa, ovom spektrogramu nedostaje i informacija o skaliranju boje (legenda pored grafika). Na žalost, spektrogramima iz signal paketa, nije moguće dodati bar sa bojama. Ali, moguće je koristiti druge funkcije sa interneta koje imaju tu funkcionalnost. Na primer. moguće je korišćenjem funkcije:

color.bar()

sa GitHub-a³⁷⁹ dodati naknadno bar sa bojama spektrogramima. Pored signal paketa, postoji i niz drugih specijalizovanih paketa za vremensko-frekvencijsku analizu signala. Za prikaz spektrograma. može se koristiti funkcija:

ggspectro()

iz seewave paketa koji je kompatibilan sa ggplot paketom i koji se dominantno koristi u analizi zvuka [513].

frekvencije odabiranja. U konkretnom primeru, sasvim slučajno je frekvencija odabiranja 1 kHz, te je broj odbiraka isti kao i broj ms. Studenti bi trebalo uvek da, u izveštajima, daju informaciju o dužini intervala u s (ili kao u ovom udžbeniku u ms). Zabuna verovatno nastaje usled toga što funkcije za iscrtavanje spektrograma primaju, kao informaciju, broj odbiraka, a ne dužinu vremenskog intervala.

378 Za iscrtavanje spektrograma na desnom panelu na Sl. 198 korišćena je širina prozora od 500 odbiraka.

379 <https://gist.github.com/johncolby/993830#file-colorbar-r>, pristupljeno 07.05.2024.

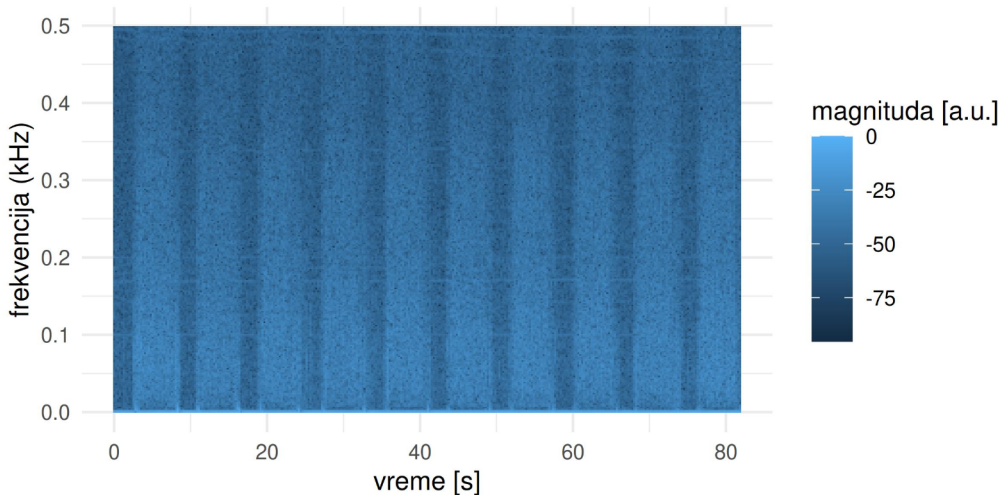
```

ggspectro(dat$V1, f = fs, tlab = "vreme [s]", wl = 500,
          wn = "hamming", ovlp = 50,
          flab = "frekvencija (kHz)",
          alab = "magnituda [a.u.]") +
  geom_tile(aes(fill = amplitude)) + theme_minimal()

```

Slika 199, Primer korišćenja funkcije iz seewave paketa [513] za prikaz spektrograma EMG signala koji se nalazi u promenljivoj *dat\$V1*.

Primenom jednostavnog koda koji je prikazana na Sl. 199 (gde je pored širine intervala određene argumentom *wl*, moguće odrediti tip prozorske funkcije postavljanjem argumenta *ovlp* od eng. *overlap*, ali i dodati odgovarajuće oznake osa), dobija se spektrogram signala na Sl. 200.



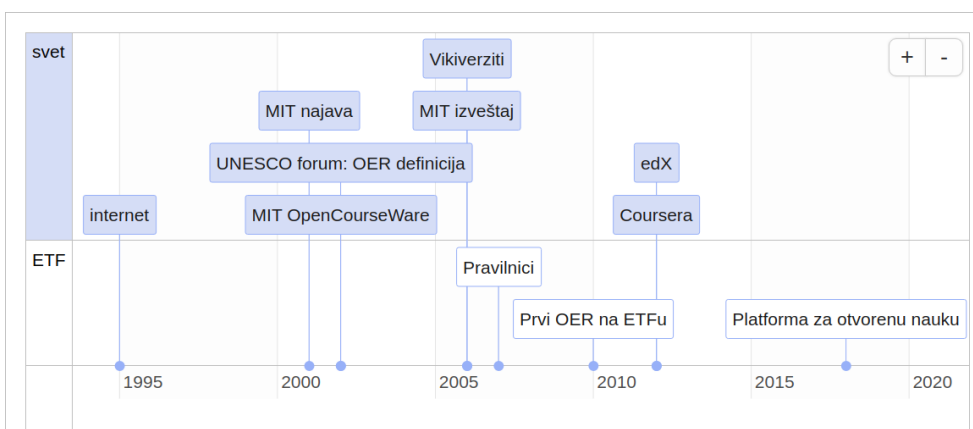
Slika 200, Spektrogram EMG signala dobijen primenom funkcije iz seewave paketa i primenom koda na Sl. 199.

Na svim slikama, na kojima je prikazan spektrogram EMG signala, moguće je uočiti 10 mišićnih kontrakcija koje su vidljive kao povećanje frekvencijskog sadržaja (tamnija boja na spektrogramima na Sl. 198 i na Sl. 200). Takođe, primetno je da su u kvazi-ekvidistantnim intervalima realizovane mišićne kontrakcije, što je u saglasnosti sa protokolom studije za čije su potrebe mereni EMG signali [340, 504]. Kako jedini izvor nestacionarnosti, koji se uočava, potiče od postojanja mišićne kontrakcije, jednostavno se dolazi do zaključka da spektrogram nije idealan za analizu, tj. vizuelizaciju EMG signala. Najčešće se spektrogrami primenjuju kod analize EEG signala, jer oni imaju izraženu stacionarnost [514], kao i u analizi govornih signala [515]. Slično kao i u

prikazu medicinske slike, skaliranje boja. odnosno zumiranje. se može primeniti i na spektrograme, tako da manje promene budu uočljive golim okom [240].

4.1.5.3 Grafik vremenske linije

Iako nije često zastupljen u analizi biosignala, jedan zanimljiv prikaz grafika, odražava događaje na vremenskoj liniji (eng. *timeline*). Na Sl. 201 je prikazan grafik vremenske linije koji na apscisi sadrži godine i koji uporedo prikazuje najvažnije događaje u svetu i na Elektrotehničkom fakultetu, Univerziteta u beogradu (ETF) u oblasti otvorenih udžbenika odnosno otvorenih nastavnih materijala (eng. *Open Educational Resources*, skraćeno OER³⁸⁰). Grafik je realizovan u R-u primenom *timevis* paketa [467], a u sličnoj formi je predstavljen na PSSOH konferenciji u 2018. godini [516], dok je kod dostupan na GitHub profilu PSSOH konferencije³⁸¹.



Slika 201, Izmenjena slika poređenja najvažnijih događaja u oblasti otvorenih nastavnih materijala (posebno se misli na udžbenike) u svetu i na Elektrotehničkom fakultetu, Univerziteta u Beogradu. Skraćenice: ETF (Elektrotehnički fakultet, Univerziteta u Beogradu), MIT (*Massachusetts Institute of Technology*, odnosno Tehnološki institut u Masačusetsu), OER (eng. *Open Educational Resources*, odnosno otvoreni nastavni materijali) i UNESCO (eng. *The United Nations Educational, Scientific and Cultural Organization*, odnosno Organizacija Ujedinjenih nacija za obrazovanje, nauku i kulturu). Coursera i edX su platforme za učenje na daljinu. Više u [516]. Slobodan softver za reprodukciju slike je dostupan u <https://github.com/pssoh/Electronic-textbooks-at-ETF-2018/>, pristupljeno 03.05.2024.

Na Sl. 202 je prikazan R kod koji omogućava predstavljanje vremenske linije događaja kao na Sl. 201. Paket *timevis* omogućava da se prikaže interaktivni grafik i namenjen je kreiranju interaktivnih veb aplikacija kroz *Shiny* paket u R-

380 https://en.wikipedia.org/wiki/Open_educational_resources, pristupljeno 07.05.2024.

381 <https://github.com/pssoh/Electronic-textbooks-at-ETF-2018/>, pristupljeno 03.05.2024.

u. Iz tog razloga, nije moguće koristiti `ggsave()` ni `jpeg()` funkcije koje se koriste za čuvanje slika sa željenom rezolucijom. Primetiti oznake “+” i “-” u gornjem desnom uglu na Sl. 201 koje omogućavaju da se smanji ili uveća prikaz vremenske linije. Takođe, `timevis` paket dodaje crvenu vertikalnu liniju za tekući vremenski trenutak što bi se moglo videti smanjenjem prikaza slike i povećanjem vremenskog intervala prikazanog na Sl. 201. Kreiranje grafika vremenske linije je relativno jednostavno (Sl. 202), jer se za prikaz grafika koristi istoimena funkcija:

`timevis()`

čiji su ulazni parametri *data frame* koji sadrži neophodne podatke (na Sl. 202 je prikazano kako je kreiran *data frame* pod nazivom *dat*), oznaka za grupe (ako ih ima) i stil (ako se želi menjati podrazumevani stil). Jedna od mana ovog paketa je što, za iskusne korisnike (slično kao i *Shiny* paket za kreiranje interaktivnih veb aplikacija), nije dovoljno fleksibilan (komande su relativno visokog nivoa, te nije jednostavno izmeniti neka osnovna podešavanja), dok je za manje iskusne korisnike odličan, jer omogućava da se samo uz par linija koda postigne odgovarajuća funkcionalnost.

```
dat <- data.frame(id = 1:11, content = c("internet", "MIT najava",
    "UNESCO forum: OER definicija", "MIT OpenCourseWare",
    "MIT izveštaj", "Vikiverziti", "Coursera", "edX",
    "Pravilnici", "Prvi OER na ETFu", "Platforma za otvorenu nauku"),
  start = c("1995-01-01", "2001-01-01", "2002-01-01", "2002-01-01", "2006-01-01",
    "2006-01-01", "2012-01-01", "2012-01-01", "2007-01-01", "2010-01-01",
    "2018-01-01"),
  end = c(rep(NA, 11)),
  group = c(rep(1, 8), rep(2, 3)),
  style = c(rep(NA, 8), rep("background-color: white;", 3)))

timevis(dat, groups = data.frame(id = 1:2, content = c("svet", "ETF"),
  style = c("color: black; background-color: #d5ddf6;",
    "color: black; background-color: white;")))
```

Slika 202, R skripta za prikaz grafika vremenske linije kao na Sl. 201. Originalni kod je dostupan na GitHub profilu PSSOH konferencije: <https://github.com/pssoh/Electronic-textbooks-at-ETF-2018/>, pristupljeno 03.05.2024. Skripta koja je podeljena uz ovaj udžbenik, kao i prikaz na Sl. 201 su neznatno promenjeni.

5 Statistička analiza biosignala

Poznavanje i razumevanje statističkih metoda je deo osnovne pismenosti lekara i biomedicinskih inženjera. Bez razumevanja statističkih testova i njihovih rezultata, nije moguće zaključiti da li je, na primer, neki lek efikasan za lečenje bolesti, da li je uređaj uspešan u dijagnostici ili terapiji, kao ni da li je nova terapija bolja od prethodne. Da bi se donela odgovarajuća odluka da li je potrebno, na primer, uvrstiti novu dijagnostičku ili terapijsku metodu u kliničku praksu, potrebna su odgovarajuća klinička ispitivanja i statistička analiza. Naravno, statistička analiza nije svemoguća i postoji niz problema koji se, praktično, mogu javiti prilikom primene postojećih testova na podacima u praksi, ali za sada predstavlja najbolje sredstvo kojim se raspolaže i na kome se temelji medicina zasnovana na dokazima (MZD ili na eng. *Evidence-based Medicine*, skraćeno i EBM), a posebno numerički deo MZD, kao i postojeći zdravstveni sistem u svetu [517].

Kako bi se dobili validni zaključci, potrebno je pažljivo odabrati statističku metodu/test, koji odgovara praktičnom problemu i uključiti sve elemente koji mogu biti od značaja. Na primer, ako metoda (novi lek) koja se ispituje pokazuje statistički značaj tako što utiče na smanjenje arterijskog krvnog pritiska od 2 mmHg u proseku, to je odličan numerički pokazatelj. Međutim, takva metoda nema nikakvu primenu u praksi, jer sa praktičnog stanovišta, smanjenje krvnog pritiska od samo 2 mmHg ne utiče značajno na zdravlje ljudi. Još drastičniji primer je pokazan u randomizovanom kliničkom ispitivanju, koje je sprovedeno sa ciljem ispitivanja uspešnosti primene funkcionalne električne stimulacije za tretman inkontinencije kod žena [518]. Inkontinencija je stanje u kome osoba ne može da kontroliše eliminaciju urina [519]. U ovakvoj studiji, postoji potreba da se paralelno ocenjuju subjektivni i objektivni parametri oporavka istovremeno (kako je već navedeno u prethodnim poglavljima). Primera radi, ako električna stimulacija doprinosi oporavku tako da se dnevni broj inkontinentalnih epizoda smanji sa 10 na jednu epizodu kod većine pacijenata (ili u ovom slučaju pacijentkinja) to može dovesti do statistički značajne razlike u broju epizoda nakon terapije, ali to i dalje ne utiče značajno na kvalitet života pacijenata jer epizode, odnosno epizoda i dalje postoji na dnevnom nivou. Iz tog razloga je važno koristiti i subjektivne i objektivne kriterijume [518]. Dok su objektivni kriterijumi, koji su najčešće zasnovani na statističkim testovima, važni (posebno sa inženjerskog gledišta), subjektivni ukazuju na praktični značaj i potencijalnu primenu terapijske metode u praksi.

Analiza sama po sebi nije cilj, već zaključak, koji se može doneti koristeći sve informacije koje su na raspolaganju, što uključuje i rezultate sprovedene analize. Pored toga, konačna odluka se najčešće donosi na osnovu i objektivnih i subjektivnih kriterijuma, te je važno sprovesti detaljnu analizu podataka. Na društvenoj mreži Tviter (sada X, X Co, San Francisko, Kalifornija, SAD) postoji odličan ilustrativni primer kako i zašto treba sprovesti detaljnu analizu podataka, čak i ako na prvi pogled podaci deluju trivijalno [520]. Cilj analize, koja je grafički prikazana u [520] je da se odgovori na relativno jednostavno pitanje: Da li ima više pasa ili mačaka kao kućnih ljubimac? Jednostavnim prebrojavanjem se dolazi do zaključka da ima više pasa (50000) od mačaka (20000). Međutim, dodatnom proverom se utvrđuje da podatke treba očistiti (ili pretprocesirati) da se uklone duplikati. Nakon toga, odnos se menja i ima više mačaka (15000) od pasa (10000). Dalje, obzirom da podaci dolaze iz različitih zemalja, ukupan broj (nakon otklonjenih duplikata) pasa i mačaka, kao i njihov odnos se menja, ako se svaka zemlja posmatra pojedinačno. Ovakva analiza može dalje da dovede do izmena u odnosu broja pasa i mačaka te primer grafik u [520] odlično ilustruje kako čak i jednostavnim podacima treba pristupiti temeljno i uzeti u obzir različite elemente koji mogu uticati na konačan rezultat.

5.1 Analiza podataka na primeru studije spavanja

U ovom udžbeniku je prikazana primena statističkih testova za ispitivanje uticaja dužine sna na zdravlje ispitanika, ispitivanjem promene vremena reakcije usled redukovanog sna. Pored statističke analize podataka iz studije spavanja, u ovom poglavlju je pokazano kako pripremiti podatke za analizu uz pomoć *dplyr* paketa [38]. Deo podataka iz studije pod nazivom “*Reaction times in a sleep deprivation study*” (srp. Uticaj redukovanog sna na vreme reakcije³⁸²) [521] je dostupan u GitHub repozitorijumu³⁸³, a opis podataka je dostupan na internet stranici Autora koji su podelili podatke³⁸⁴. Podaci su podeljeni u datoteci sa .csv ekstenzijom, a njihova struktura kada se otvore u Libre Calc-u (Libre Office, The Document Foundation, Berlin, Nemačka) je prikazana na Sl. 203.

Vreme reakcije je jedna od najčešćih mera koja se primenjuje u eksperimentalnoj psihologiji, ali i u drugim oblastima. Postoji više načina da se izmeri vreme reakcije. U jednostavnim merenjima, postoji samo jedan stimulus

382 Slobodan prevod sa engleskog na srpski jezik.

383 <https://vincentarelbundock.github.io/Rdatasets/datasets.html>, pristupljeno 02.06.2024.

384 <https://vincentarelbundock.github.io/Rdatasets/doc/lme4/sleepstudy.html>, pristupljeno 02.04.2024.

i jedan odgovor. Na primer, ako ispitanik pritisne taster u trenutku kada čuje zvuk to je primer jednostavnog vremena reakcije (na ovaj način je mereno vreme reakcije u studiji [521]). U merenjima reakcije, u kojima je cilj prepoznavanje (eng. *Recognition*) stimulusa, potrebno je reagovati na odabrane stimuluse, a druge ignorisati. Ako bi zadatak bio da ispitanik treba da pritisne taster kada vidi crveno svetlo, a da ignoriše zeleno svetlo (odnosno da u tom trenutku ne pritiska taster), to bi značilo da se meri prepoznavanje stimulusa, pored vremena reakcije. Prepoznavanje simbola i prepoznavanje tonova takođe spada u merenja prepoznavanja stimulusa. U studijama izbora, zadatak je da se reaguje na različite načine kada naiđu različiti stimulusi (na primer: taster se pritiska jednom kada se uključi crveno svetlo, a dva puta kada se uključi zeleno svetlo). Vreme reakcije se meri na stimuluse koji se prikazuju po pseudoslučajnom redosledu, kako ne bi došlo do privikavanja osobe na stimulus, tj. adaptacije. Međutim, postoje i studije u kojima redosled stimulusa nije pseudoslučajan. U takvim situacijama, može se izmeriti kraće vreme reakcije kako se broj ponavljanja povećava (ispitanici duže vežbaju i dolazi do adaptacije na stimulus). Uobičajeno vreme reakcije kod zdravih i mladih ljudi (studentska populacija), na jednostavan svetlosni stimulus je 190 ms, a na zvuk 160 ms. Vreme reakcije zavisi od niza parametara, kao što su intenzitet stimulusa, godine, pol, umor, konzumacija alkohola, tip ličnosti, stres, inteligencija i povrede centralnog nervnog sistema. [522]

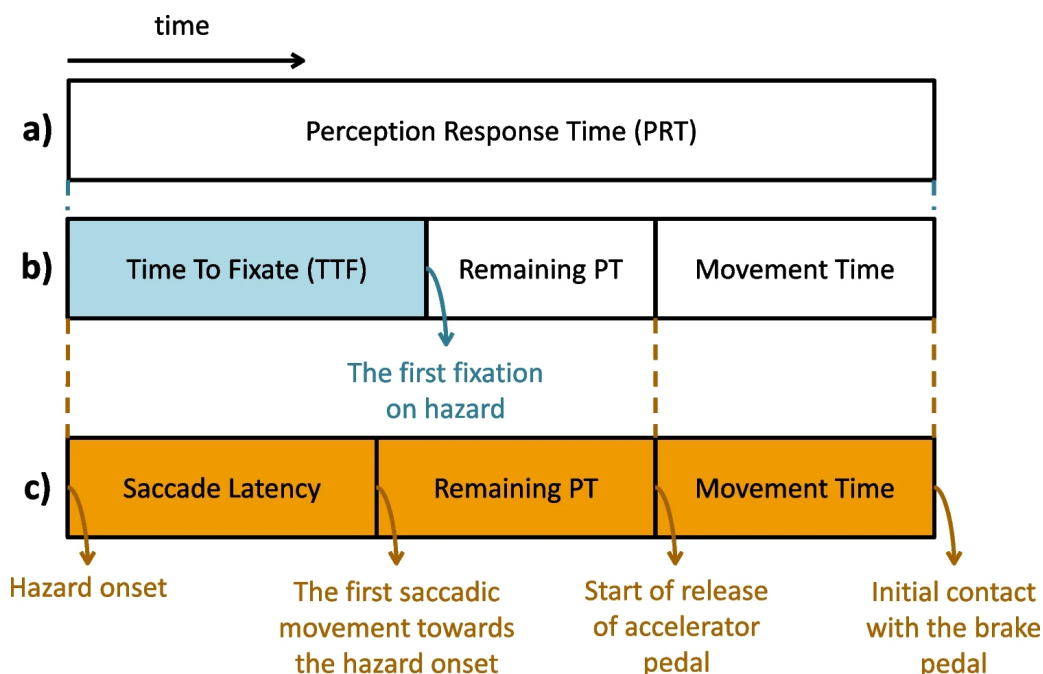
rownames	Reaction	Days	Subject
1	249.56	0	308
2	258.7047	1	308
3	250.8006	2	308
4	321.4398	3	308
5	356.8519	4	308
6	414.6901	5	308
7	382.2038	6	308
8	290.1486	7	308
9	430.5853	8	308

Slika 203, Podaci u studiji proučavanja uticaja redukovanog sna na vreme reakcije.

Ukratko, vreme reakcije (eng. *Reaction Time*) ili vreme percepcije i reakcije (eng. *Perception Response Time*, skraćeno PRT) je vreme koje protekne od trenutka kada se primeti stimulus (na primer, prepreka na putu tokom vožnje

automobila ili pojava crvenog svetla/stop znaka u simulaciji vožnje sa instrukcijom da ispitanik zaustavi vozilo) do trenutka kada započne usporavanje kretanja vozila pritiskom na kočnicu. PRT parametar je kritičan i za planiranje saobraćajnica, a usvojeno vreme u Evropi za PRT je 2 s, dok je u Americi 2.5 s. [85, 523-527]

Generalno, PRT se može podeliti na par komponenti ,od kojih najvažnije dve predstavljaju fazu percepcije stimulusa (senzorna komponenta) i fazu reagovanja na stimulus (motorna komponenta), kao što je prikazano na Sl. 204. Komponente, na koje se PRT može podeliti, trenutno se neznatno razlikuju u naučnoj literaturi kako je i prikazano na Sl. 204 [85, 359, 528-529]. Najveća razlika u pristupima je u izdvajanju senzorne komponente (paneli b) i c) na Sl. 204) koja može da uključi vreme potrebno da se pogled usmeri ka stimulusu (eng. *Saccade Latency*) ili vreme za koje se pogled fiksira na stimulus (eng. *Time to First Fixation*). Ono što je zajedničko za oba pristupa, u podeli PRT na komponente, jeste da postoji vremenski interval u kome se vrši obrada informacija u nervnom sistemu (eng. *Remaining Processing Time*, na Sl. 204 označeno kao *Remaining PT*), kao i vremenski interval potreba da se izvrši pokret (eng. *Movement Time*) na Sl. 204.



Slika 204, Vreme percepcije i reakcije (eng. *Perception Response Time*, skraćeno PRT) sa komponentama. Slika je podeljena pod otvorenom licencom i preuzeta iz [85].

Studija, čiji su podaci na raspolaganju za analizu u ovom poglavlju [521] je realizovana na 66 zdravih ispitanika (studentata) koji su podeljeni u grupe. U zavisnosti u kojoj su grupi, studenti su 7 dana bili izloženi sa po 3, 5, 7 ili 9 sati sna. Nakon perioda od 7 dana, svi ispitanici su tri dana spavali po 8 sati (što se smatra normalnom dužinom sna tokom dana, a u ovoj studiji je predstavljeno kao faza oporavka). Ukupno trajanje ispitivanja je bilo 13 dana, jer su prva dva dana korišćena za obučavanje i privikavanje, dok je trećeg dana izvršeno referentno merenje (eng. *baseline*). Za procenu uticaja sna na senzorno-motornu kontrolu ispitanika, odnosno za merenje vremena reakcije, korišćen je PVT test (eng. *Psychomotor Vigilance Task*) u kome se meri vreme reakcije na vizuelni stimulus. Podaci koji su dostupni predstavljaju usrednjene vrednosti PVT merenja po ispitanicima. U bazi, koja je podeljena, na raspolaganju je 180 usrednjenih PVT merenja sa 18 ispitanika i uključeni su samo ispitanici koji su spavali 3 sata dnevno. PVT je meren tako što je vizuelni stimulus predstavljen otprilike 10 puta u minuti tokom 10 minuta (interval između dva stimulusa je bio u opsegu od 2 s do 10 s). Ispitanici su držali uređaj u ruci, tokom testiranja (eng. *hand-held device*), a bilo je potrebno da palcem pritisnu taster kada dioda zasvetli (vizuelni stimulus). Ovi testovi su ponovljeni više puta tokom dana, a na raspolaganju (u otvorenom domenu) su usrednjeni rezultati merenja [521].

Pored podeljenih podataka, studija je uključila merenje i analizu polisomnografskih podataka (EEG, EMG, EKG i EOG signali su mereni). Polisomnografija je medicinska procedura u kojoj se realizuje sinhrono merenje većeg broja parametara (uključujući i elektrofiziološke signale) sa ciljem dijagnostike poremećaja sna [530]. PVT, kao standardizovan test, koristi se za merenje vremena reakcije ispitanika na vizuelni stimulus, pritiskom tastera, a ima praktični značaj za ocenu kvaliteta sna i dijagnostiku patologija koje su u vezi sa poremećajem spavanja [531]. Aplikacija ovog tipa se može relativno jednostavno realizovati, a njena primena je široko rasprostranjena i nije namenjan isključivo primeni u studijama spavanja. PVT test se, na primer, koristi u međunarodnoj svemirskoj stanici, a NASA (eng. *National Aeronautics and Space Administration*) njime ocenjuje neurobihevijoralne sposobnosti kosmonauta [532]. Merenje PVT testom je važno sprovesti u strogo kontrolisanim uslovima, jer na rezultat, pored smanjenog sna, mogu uticati i drugi faktori (na primer: unos kofeina, fizička iscrpljenost i lekovi) [532].

Podaci iz studije spavanja koji su sačuvani u datoteci "sleepstudy.csv" sadrže 4 kolone i ukupno 180 merenja. Prva kolona odgovara rednom broju merenja, u drugoj se nalazi vreme reakcije u ms, u trećoj koloni je smešten redni broj dana koji su ispitanici proveli tokom ispitivanja, dok se u četvrtoj koloni nalazi identifikacioni broj (ID) subjekta (Sl. 203). *Reaction* kolona sadrži usrednjenu

vrednost svih dnevnih testova izraženu u ms. Za početak je potrebno učitati podatke u R okruženje, a potom proveriti strukturu i tip podataka po kolonama. Takođe, potrebno je proveriti da li ima nedostajućih vrednosti. Ovi koraci uključuju upotrebu relativno jednostavnih funkcija, a prikazani su na Sl. 205. Slična analiza bi se mogla postići primenom *str()* funkcije (od eng. *structure*). Kako nema nedostajućih podataka (Sl. 205), može se pristupiti daljoj analizi.

```
> dat <- read.csv("sleepstudy.csv")
> head(dat)
  rownames Reaction Days Subject
1         1 249.5600   0     308
2         2 258.7047   1     308
3         3 250.8006   2     308
4         4 321.4398   3     308
5         5 356.8519   4     308
6         6 414.6901   5     308
> dim(dat)
[1] 180  4
> class(dat$rownames)
[1] "integer"
> class(dat$Reaction)
[1] "numeric"
> class(dat$Days)
[1] "integer"
> class(dat$Subject)
[1] "integer"
> sum(is.na(dat$Reaction))
[1] 0
> |
```

Slika 205, R skripta za učitavanje podataka i prikaz dimenzije i osnovnih parametara *data frame*-a u kome se nalaze osnovni podaci iz studije spavanja.

Sledeći korak uključuje računanje parametara deskriptivne (ili opisne) statistike po danima, jer su dani od interesa u ovoj studiji (očekuje se da između prva tri dana razlika ne postoji, ali da se vreme reakcije povećava kako broj dana sa redukovanim snom raste). Najjednostavniji način da se dobije opisna statistika jeste da se koristi:

summary()

funkcija, kao što je prikazano na Sl. 206 i preporučeno u [67]. Prikaz parametara deskriptivne statistike ima najviše smisla za parametar *Reaction*, ali

za ostala obeležja, osim toga što se može videti da je maksimalan broj redova 180 (*rownames* kolona), kao i da broj dana je u opsegu od 0 do 9 (*Days* kolona), nema mnogo smisla.

```
> summary(dat)
  rownames      Reaction      Days      Subject
Min.   : 1.00   Min.   :194.3   Min.   :0.0   Min.   :308.0
1st Qu.: 45.75   1st Qu.:255.4   1st Qu.:2.0   1st Qu.:331.0
Median : 90.50   Median :288.7   Median :4.5   Median :336.0
Mean    : 90.50   Mean    :298.5   Mean    :4.5   Mean    :341.3
3rd Qu.:135.25   3rd Qu.:336.8   3rd Qu.:7.0   3rd Qu.:352.0
Max.    :180.00   Max.    :466.4   Max.    :9.0   Max.    :372.0
> |
```

Slika 206, Primer primene *summary()* funkcije za podatke iz studije spavanja.

Za prikaz opisnih statističkih podataka po danima, mogu se koristiti funkcije iz *dplyr* paketa, ali postoje i drugi načini. Čitaoci se upućuju na funkciju:

by()

koja se može koristiti u kombinaciji sa *summary()* funkcijom za računanje i prikaz parametara deskriptivne statistike.

```
> filter(dat, Days == 0) %>% select(Reaction) %>% unlist %>% mean()
[1] 256.6518
> filter(dat, Days == 1) %>% select(Reaction) %>% unlist %>% mean()
[1] 264.4958
> filter(dat, Days == 9) %>% select(Reaction) %>% unlist %>% mean()
[1] 350.8512
> pod <- dat %>% group_by(Days) %>%
+   summarize(Average = mean(Reaction)) %>% as.data.frame()
> pod
  Days Average
1     0 256.6518
2     1 264.4958
3     2 265.3619
4     3 282.9920
5     4 288.6494
6     5 308.5185
7     6 312.1783
8     7 318.7506
9     8 336.6295
10    9 350.8512
```

Slika 207, Prikaz srednjih vrednosti vremena reakcije po danima.

Neki od osnovnih parametara opisne statistike su minimalna i maksimalna vrednost, srednja vrednost, standardna devijacija i prvi i treći kvartil³⁸⁵ i računaju se po kolonama (Sl. 206). Postoji više načina da se dplyr paket iskoristi za prikaz parametara opisne statistike (na primer, za prikaz srednje vrednosti vremena reakcije po danima), a dva pristupa su prikazana na Sl. 207. U prvom pristupu je potrebno pozvati bar 9 linija koda (za svaki dan pojedinačno se računaju srednje vrednosti) ili realizovati poziv funkcije kroz petlju, dok je u drugom pristupu potrebno pozvati samo jednu liniju koda (rezultujući *data frame* sadrži dve kolone – jednu u kojoj su prikazani dani i drugu u kojoj su prikazane usrednjene vrednosti vremena reakcije u ms).

Primititi da, iako je srednja vrednost dovoljno ubedljiva da smanjeno trajanje sna utiče na povećanje vremena reakcije, podaci prikazani na Sl. 207 ne mogu se koristiti kao dokaz za ovu tvrdnju, iz jednostavnog razloga, jer ne sadrže standardnu devijaciju. Da bi se dodala kolona koja prikazuje standardnu devijaciju u novom *data frame*-u pod nazivom *pod*, potrebno je realizovati kod kao na Sl. 208.

```
> pod$SD <- dat %>% group_by(Days) %>%
+   summarize(SD = sd(Reaction)) %>% select(SD) %>% unlist()
> pod
```

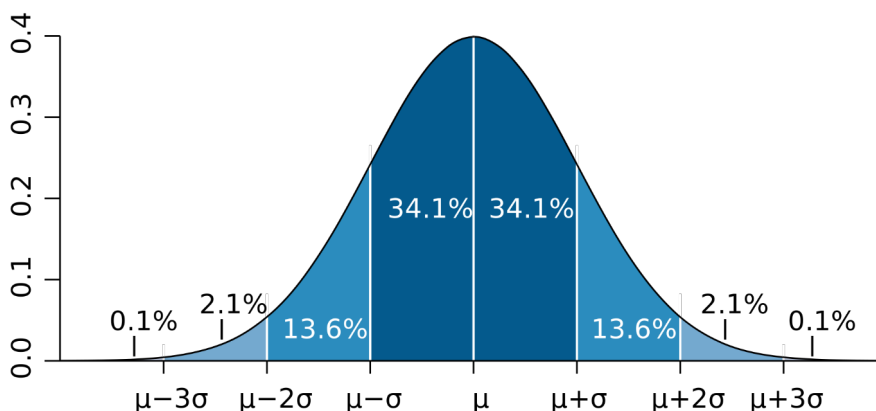
	Days	Average	SD
1	0	256.6518	32.12945
2	1	264.4958	33.43033
3	2	265.3619	29.47342
4	3	282.9920	38.85774
5	4	288.6494	42.53789
6	5	308.5185	51.76962
7	6	312.1783	63.17372
8	7	318.7506	50.10396
9	8	336.6295	60.19972
10	9	350.8512	66.98616

Slika 208, prikaz novog *data frame*-a *pod* sa dodatom standardnom devijacijom u koloni *SD* za vreme reakcije po danima.

Sa Sl. 208 je jasno da iako za 9. dan postoji najveća srednja vrednost za vreme reakcije (*Reaction*) i standardna devijacija (*SD*) je najveća, te bi trebalo pažljivo pristupiti poređenju vremena reakcije po danima. Dodatno, računanje srednje vrednosti i standardne devijacije ima smisla isključivo, ako su podaci normalno raspodeljeni, odnosno ako im je funkcija gustine verovatnoće normalna, tj.

385 Za detaljno objašnjenje šta su kvantili, kvantili, decili i percentili, čitaoci se upućuju na članak “Mere varijacije” na portalu e-Statistika [533].

Gausova [78]. U slučaju da je raspodela normalna, može se dodeliti odgovarajuća verovatnoća prikazu rezultata za, na primer, 9. dan, odnosno može se pisati da je vreme reakcije za sve ispitanike 9. dana bilo 350.85 ± 66.99 ms. Ovakav zapis znači da postoji verovatnoća od 68.2% da se novo vreme reakcije (ako bi se, na primer, merilo za još jednog 19. ispitanika) za 9. dan nalazi u opsegu $[350.85 - 66.99, 350.85 + 66.99]$ ms, odnosno u opsegu $[283.86, 417.84]$ ms. Verovatnoća od 68.2% je osobina Gausove funkcije gustine verovatnoće kao što je prikazano na Sl. 209 gde je sa μ označena srednja vrednost, dok je sa σ označena standardna devijacija. Za više informacija pogledati udžbenik [78], a posebno metodu za računanje merne nesigurnosti tipa A kada je funkcija gustine verovatnoće Gausova.



Slika 209, Od Ainali – Sopstveno delo, podeljeno pod licencom CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=3141713>

Bez prethodnog poznavanja funkcije gustine verovatnoće, rezultat koji je predstavljen sa srednjom vrednošću i standardnom devijacijom, ne može imati dodeljenu verovatnoću da će se novoizmerena vrednost naći u odgovarajućem opsegu (u slučaju vremena reakcije za 9. dan to je $[283.86, 417.84]$ ms). Međutim, čak i kada podaci nemaju normalnu raspodelu, standardna devijacija može biti koristan parametar prilikom prikaza rezultata, jer sadrži informaciju o tome koliko se pojedinačne vrednosti međusobno razlikuju.

U statistici se najčešće za proveru raspodele podataka, odnosno za proveru da li su podaci normalno raspodeljeni ili ne, koristi Šapiro-Vilk test (eng. *Shapiro-Wilk's test*) koji su Šapiro i Vilk predstavili 1965. godine [534]. U R-u se za primenu ovog testa može koristiti funkcija:

`shapiro.test()`

Detalji realizacije Šapiro-Vilk testa prevazilaze obim gradiva koji je pokriven udžbenikom, a studenti su upućuju na originalni rad u kome je ovaj test prvi put predstavljen [534]. Osnovna pretpostavka, odnosno hipoteza u ovom testu, je da su podaci normalno raspodeljeni. Ako to nije ispunjeno, rezultujuća p vrednost (p od eng. *probability* ili verovatnoća na srpskom jeziku) će biti manja ili jednaka od neke granične vrednosti (najčešće je to 0.05³⁸⁶, ali u naučnoj literaturi se koriste 0.01 i 0.001 za granicu odlučivanja). Ako se za vremena reakcije po danima pretpostavi da je granična vrednost za p jednaka 0.01, onda se jednostavno primenom *shapiro.test()* funkcije može proveriti da li je $p > 0.01$ za sve dane (pogledati Sl. 210).

```
> shapiro.test(dat$Reaction[dat$Days == 9])$p.value
[1] 0.0134231
> shapiro.test(dat$Reaction[dat$Days == 8])$p.value
[1] 0.8185571
> shapiro.test(dat$Reaction[dat$Days == 7])$p.value
[1] 0.3724405
> shapiro.test(dat$Reaction[dat$Days == 6])$p.value
[1] 0.584681
> shapiro.test(dat$Reaction[dat$Days == 5])$p.value
[1] 0.9270848
> shapiro.test(dat$Reaction[dat$Days == 4])$p.value
[1] 0.8427033
> shapiro.test(dat$Reaction[dat$Days == 3])$p.value
[1] 0.918987
> shapiro.test(dat$Reaction[dat$Days == 2])$p.value
[1] 0.9935835
> shapiro.test(dat$Reaction[dat$Days == 1])$p.value
[1] 0.3880334
```

Slika 210, Primena Šapiro-Vilk testa za proveru normalnosti podataka, odnosno da li vreme reakcije po danima ima Gausovu funkciju gustine verovatnoće.

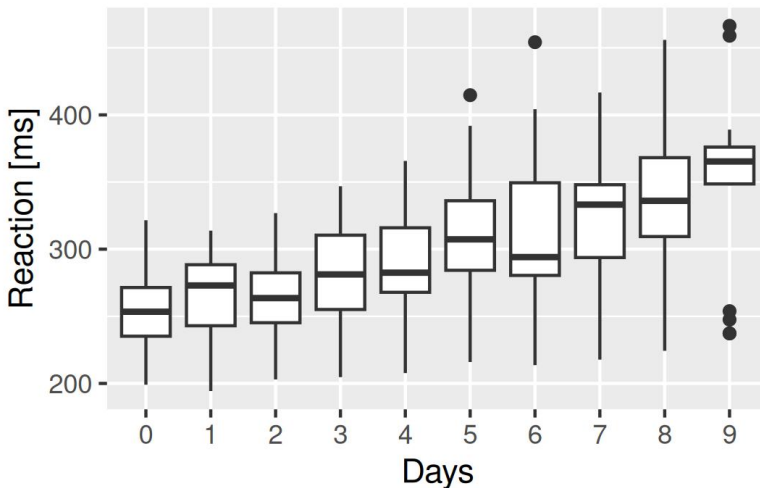
Kako je rezultat Šapiro-Vilk testa predstavljen sa p vrednostima koje su sve veće od 0.01, zaključuje se da su vremena reakcije po danima normalno raspodeljena. Na Sl. 210 je korišćen operator "\$" za izdvajanje p vrednosti, radi kompaktnijeg zapisa. Takođe, p vrednosti nisu iste za podatke po danima i p vrednost za podatke koji su mereni 9. dana je najmanja, te je bilo moguće dobiti rezultat (za granicu od 0.05) da podaci mereni 9. dana nisu normalno raspodeljeni (jer je $p < 0.05$). Dalje, da bi se grafički predstavile razlike po danima, moguće je iskoristiti ggplot2 paket. U poglavlju o vizuelizaciji detaljno

386 Broj 0.05 nema jedinicu, ali kada bi se predstavio u procentima označava 5%. Generalno, u statistici je uobičajeno da se umesto procenata brojne vrednosti prikazuju normalizovane na 1. Analogno, normalizovanim vrednostima od 0.01 i 0.001 odgovaraju 1% i 0.1%.

je objašnjeno kako se *error bar* koristi za podatke sa normalnom raspodelom, ali ovde će biti prikazan *box plot* iako je Šapiro-Vilk test pokazao da podaci jesu normalno raspodeljeni³⁸⁷. Za poziv sledeće komande:

```
ggplot(dat, aes(x = as.factor(Days), y = Reaction)) +  
  geom_boxplot() + xlab("Days") + ylab("Reaction [ms]")
```

i primenom *ggsave()* funkcije za čuvanje grafika sa odgovarajućom rezolucijom u odabranom formatu, dobija se grafik kao na Sl. 211. Grafički prikaz podataka ukazuje da postoje nedvosmisleno razlike između vremena reakcije 0. i 9. dana, ali da li su ove razlike statistički značajne ili nisu? Za odgovor na ovo pitanje, potrebno je primeniti odgovarajući statistički test.



Slika 211, *Box plot* za vreme reakcije (*Reaction*) po danima (*Days*) za 18 ispitanika na kojima je mereno vreme reakcije u studiji spavanja [521].

Statistički testovi se dele na parametarske i neparametarske testove [535]. Ovde neće detaljno biti reči o svim statističkim testovima, već samo o osnovnim metodama koje se, po iskustvu Autorke, najčešće sreću u praksi. Parametarski testovi, između ostalog, podrazumevaju da je raspodela podataka normalna, dok neparametarski testovi nemaju ovu pretpostavku. Prema tome, provera hipoteze da li su podaci normalno raspodeljeni predstavlja važan pretkorak u statističkom proračunu, jer utiče na izbor tipa testa (parametarski ili

³⁸⁷ Kao i za prikaz srednje vrednosti i standardne devijacije, *error bar* se koristi za podatke koji imaju normalnu raspodelu. Može se koristiti da prikaže meru disperzije/rasipanja podataka oko srednje vrednosti (bez odgovarajuće verovatnoće), ali nije preporučljivo.

neparametarski). U slučaju da su ispunjeni uslovi za primenu parametarskih testova, njima treba dati prednost usled veće statističke snage [535].

Poređenjem samo srednjih vrednosti, dolazi se do zaključka da je vreme reakcije veće 9. dana (350.85 ms) u odnosu na 0. dan (256.65 ms) i to za apsolutnu razliku od 94.20 ms, odnosno za relativnu razliku (u odnosu na 0. dan) od 36.7% ili (u odnosu na 9. dan) od 26.8%. Ove relativne razlike nisu zanemarive, a u literaturi je usvojeno da relativna razlika koja je veća od 10% može da ukazuje na značajno smanjenje/povećanje, dok razlika koja je manja od 5% najverovatnije ne ukazuje na značajne promene između dve grupe parametara [536]. Naravno, dodatna informacija o praktičnom značaju parametra koji se posmatra uvek mora biti uzeta u obzir (kao u primeru terapijske metode koja smanjuje srčani krvni pritisak za samo 2 mmHg). Da bi se numerički pokazalo da li dobijene relativne razlike za vreme reakcije jesu ili nisu statistički značajne, može se koristiti t-test primenom funkcije:

t.test()

koja daje rezultat kao na Sl. 212. Studentima se savetuje da probaju da pozovu ovu funkciju tako što će zameniti mesta prvog i drugog parametra – zamena mesta ne utiče na konačan rezultat (da li postoji ili ne postoji statistički značajna razlika), osim na vrednost *t* statistike. Takođe primetiti na Sl. 212 da se dobijaju različite *p* vrednosti kada se parametar *paired* postavi na vrednost *TRUE*, u odnosu na podrazumevanu *FALSE* vrednost. Bez obzira na ovu izmenu, dobijene *p* vrednosti su manje od graničnih 0.05, 0.01 i 0.001, te ukazuju na postojanje statistički značajnih razlika između vremena reakcije za 0. i 9. dan.

Pravilno korišćenje t-testa, u ovom slučaju, je prikazano kada je parametar *paired* postavljen na *TRUE* vrednost (logička jedinica u Bulovoj algebri), jer su podaci dobijeni iz istog merenja (t-test za zavisne uzorke), odnosno na istim ispitanicima. Kada bi se, kojim slučajem, poredili rezultati na različitim ispitanicima (jedna grupa koja je imala san uobičajenog trajanja od 8 sati sa drugom grupom ispitanika koja je imala san od 3 sata), onda bi bilo potrebno da se parametar *paired* postavi na *FALSE* vrednost (t-test za nezavisne uzorke).

Povratkom na početak analize, moguće je zaključiti da vremena reakcije koja su merena 9. dana nemaju normalnu raspodelu (ako se *p* vrednost poredi sa pragom od 0.05), te u tom slučaju nije moguće primeniti t-test. Za primenu t-testa, potrebno je da obe promenljive koje se porede imaju normalnu raspodelu. U slučaju da jedna od promenljivih nema normalnu raspodelu, koristio bi se

neparametarski Men-Vitni-Vilkoksonov³⁸⁸ (eng. *Mann-Whitney-Wilcoxon*) test [382, 537]. Pored neparametarskog Men-Vitni-Vilkoksonovog testa, u upotrebi je i neparametarski Kruskal-Valisov (eng. *Kruskal-Wallis*) test [538]. Osnovna razlika između ova dva neparametarska testa je u tome što se Kruskal-Valisov test može jednostavno prilagoditi poređenju između više od dve grupe podataka, dok im je zajedničko da se najčešće koriste za nezavisne skupove podataka i da koriste sumarne ocene ranga u proračunima [539]. Detalji o osnovnim testovima prevazilaze gradivo predstavljeno u ovom udžbeniku, pa se studenti upućuju na odlične izvore na srpskom i engleskom jeziku [535, 540-542].

```
> t.test(dat$Reaction[dat$Days == 9], dat$Reaction[dat$Days == 0])
```

Welch Two Sample t-test

```
data: dat$Reaction[dat$Days == 9] and dat$Reaction[dat$Days == 0]
t = 5.3794, df = 24.429, p-value = 1.507e-05
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 58.09198 130.30685
sample estimates:
mean of x mean of y
350.8512 256.6518
```

```
> t.test(dat$Reaction[dat$Days == 9], dat$Reaction[dat$Days == 0],
+        paired = TRUE)
```

Paired t-test

```
data: dat$Reaction[dat$Days == 9] and dat$Reaction[dat$Days == 0]
t = 6.9576, df = 17, p-value = 2.311e-06
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 65.63465 122.76419
sample estimates:
mean difference
 94.19942
```

Slika 212, Prikaz rezultata t-testa u dva slučaju (pravilno, u slučaju primene testa za zavisne uzorke, kada je *paired* parametar postavljen na vrednost *TRUE* i nepravilno, u slučaju primene testa za nezavisne uzorke, kada je *paired* parametar postavljen na podrazumevanu *FALSE* vrednost).

388 Može se naći u naučnoj i stručnoj literaturi i pod nazivima *Mann-Whitney U* test ili *Wilcoxon rank-sum* test [539].

Pored podele testova na parametarske i neparametarske, postoji podela testova i na testove za jedan uzorak, dva uzorka i testove za više uzoraka (Kruskal-Valisov test se koristi za tri i više uzoraka). A, kako je već pokazano, nezavisni (eng. *unpaired*) i zavisni (eng. *paired*) testovi se primenjuju na podatke koji imaju dva ili više uzoraka, dok se t-test za jedan uzorak (eng. *one sample t-test*) primenjuje kada postoji poređenje uzorka iz populacije sa očekivanim vrednostima koje su unapred poznate. [535]. Na primer, t-test za jedan uzorak se može koristiti kada se želi proveriti da li su sistolni i dijastolni krvni pritisci na uzorku muškaraca (uzorak iz zdrave populacije) jednaki 122 mmHg i 74.5 mmHg, što je u opsegu krvnog pritiska koja je određena na populaciji od 43800 muškaraca u studiji koju je publikovala Bojton sa saradnicima 1947. godine [543]. Cilj primene t-testa za jedan uzorak bi bio da se proverí da li su sistolni i dijastolni pritisci na novom uzorku u skladu sa podacima iz literature.

```
> wilcox.test(dat$Reaction[dat$Days == 9], dat$Reaction[dat$Days == 0],  
+           paired = TRUE)
```

```
Wilcoxon signed rank exact test
```

```
data: dat$Reaction[dat$Days == 9] and dat$Reaction[dat$Days == 0]
```

```
V = 170, p-value = 1.526e-05
```

```
alternative hypothesis: true location shift is not equal to 0
```

Slika 213, Men-Vitni-Vilkoksonov neparametarski test za poređenje vremena reakcije 0. i 9. dana pod pretpostavkom da podaci nemaju normalnu raspodelu.

Funkcija:

wilcox.test()

u programskom jeziku R omogućava da se za izbor *paired* parametra pokreću dva različita testa: ako je *paired* jednak FALSE (logička nula u Bulovoj algebri), u tom slučaju test je jednak Man-Vitnijevom testu. Za više informacija, potrebno je pogledati R dokumentaciju za osnovni paket stats³⁸⁹.

Pre detaljnog objašnjenja t-testa, biće dat kratak osvrt na značaj statističkih testova u praksi, na primeru studije spavanja i poređenja vremena reakcije 0. i 9. dana. Jedno je pokazati da su parametri iz jedne grupe (vreme reakcije mereno 0. dana) razlikuju u odnosu na drugu grupu parametara (vreme reakcije mereno 9. dana), ali je potpuno druga dimenzija u analizi ako se pokaže da su

389 <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/wilcox.test>, pristupljeno 02.06.2024.

razlike na dva skupa podataka takve da su statistički značajne. To znači da se korišćenjem matematičkog, preciznije statističkog aparata, pokazuje da razlika nastala između dve grupe parametara nije slučajna, već da postoji i da je značajna. Ovakvi testovi omogućavaju da se dođe do korisnih rezultata i validnih zaključaka. Belenki je sa saradnicima [521] pokazao da se vreme reakcije ispitanika menjaju u odnosu na to koliko dugo traje restrikcija sna. Takođe, pokazano je da na oporavak od restrikcije sna utiče broj sati provedenih u snu tokom noći, jer je u grupi koja je spavala najkraće (3 sata) primećeno najveće povećanje vremena reakcije, kao i najsporiji oporavak. Na osnovu ovih podataka, došlo se do zaključka da se ljudski mozak prilagođava uslovima smanjenog sna, tako što se smanjuje operativna kapacitivnost sve dok ne dođe do potpunog oporavka [521]. Rezultati merenja i statističkih testova su omogućili da se detaljnije prouči restrikcija sna, jer nedostatak sna može negativno da utiče na zdravlje, bezbednost, produktivnost i kvalitet života [544-546]. Statistička analiza, u naučnom radu je uključila drugačiji pristup od onog koji je ovde pokazan [521]. Ovo je očekivano, jer je veći broj rezultata (na više ispitanika, tokom različitih PVT testova u danu) bio na raspolaganju. Zanimljivo je spomenuti da je bilo potrebno proveriti za različite grupe ispitanika (oni koji su spavali 3 sata dnevno, 5 sati itd.) da li su postojale statistički značajne razlike u godinama ispitanika i u polu ispitanika. Ovo je važno, jer godine i pol ispitanika mogu da utiču na uticaj restrikcije koja se ogleda u povećanju vremena reakcije. Statistički testovi koji su pokazani u ovom poglavlju mogu se koristiti za ovakva ispitivanja. Dodatno, uobičajeno je da se umesto analize vremena reakcije koristi brzina reakcije, te se jednostavnom transformacijom podataka: $1000/Reaction$ može dobiti brzina u jedinici koja odgovara broju reakcija po s, gde je sa *Reaction* označeno vreme reakcije u ms. Ukupno, rezultati studije koju su izveli Belenki o saradnici [521] su bili u skladu sa Hornovom hipotezom da je 4.5 sati sna potrebno dnevno da bi se zadovoljila osnovna fiziološka potreba [547]. Dodatno, rezultati PVT testa u studiji [521] su ukazali na to da smanjen broj sati sna, kroz par dana, može da izazove kontinualno smanjenje budnosti i ukupnog učinka, dok dužina sna koja zadovoljava minimalno trajanje neće dovesti do trajnijih promena, već će se postići asimptotski stabilno stanje. Rezultati studije su ukazali na uticaj redukovane sna kroz duže vremenske periode, koji mogu da ostave trajnije posledice po vreme reakcije. U analizi dostupnih podataka iz ove studije, potvrđeni su ovi navodi, jer je primenom odgovarajućeg testa (t-test za zavisne uzorke) dobijena statistički značajna razlika između 0. i 9. dana koja je pri tom potvrđena i grafički korišćenjem *box plot* grafika (ili grafikona kutije kako je nazvan u [535]).

5.2 Osnovni proračun t-testa na primeru podataka o težini biljaka

T-test će biti detaljno objašnjen na primeru PlantGrowth podataka o težini biljaka u tri grupe iz ISwR paketa [36]³⁹⁰. Ako se učitaju podaci o težini biljaka u R i prikažu *error bar* dijagrami za sve tri grupe, postavlja se pitanje kako utvrditi da li je razlika koja se uočava posledica slučajnosti, odnosno kako uporediti srednje vrednosti prikazane na Sl. 214, kao i *error bar* koji je prikazan na Sl. 215 koja se dobija primenom koda na Sl. 216 Kako nije poželjno podatke za koje ne postoji potvrda o normalnoj raspodeli prikazivati primenom *error bar* grafika, alternativni prikaz sa *box plot*-om je predstavljen na Sl. 217.

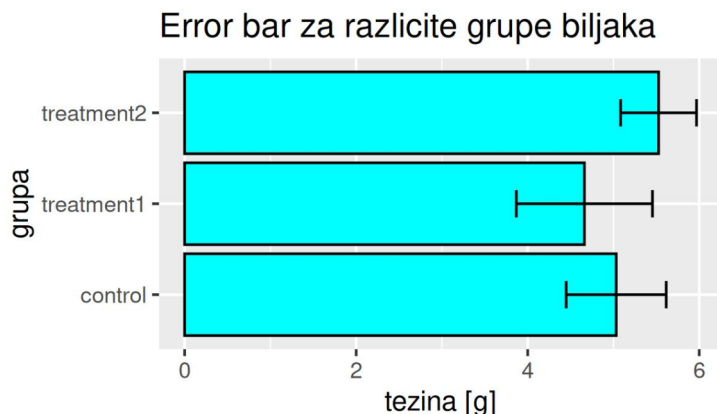
```
> dat <- PlantGrowth
> ctDat <- filter(dat, group == "ctrl")
> t1Dat <- filter(dat, group == "trt1")
> t2Dat <- filter(dat, group == "trt2")
> pod <- list("numeric")
> pod$mean <- c(mean(ctDat$weight), mean(t1Dat$weight),
+              mean(t2Dat$weight))
> pod$group <- c("control", "treatment1", "treatment2")
> pod$sd <- c(sd(ctDat$weight), sd(t1Dat$weight),
+            sd(t2Dat$weight))
> pod <- as.data.frame(pod)
> pod
  X.numeric. mean      group      sd
1  numeric 5.032   control 0.5830914
2  numeric 4.661 treatment1 0.7936757
3  numeric 5.526 treatment2 0.4425733
```

Slika 214, Srednje vrednosti i standardne devijacije za tri grupe biljaka (*control*, *treatment1* i *treatment2*) koje su sačuvane u promenljivoj *pod*. Biblioteka koju je potrebno učitati da bi se iskoristili ugrađeni podaci je *IswR* [36], a biblioteka koju je potrebno učitati za manipulaciju podacima je *dplyr* [38].

Box plot koji je prikazan na Sl. 217 se dobija jednostavnom izmenom koda prikazanom na Sl. 216 za prikaz *error bar* grafika. Radi jednostavnijeg računa, dalja statistička analiza će biti prikazana samo za dve grupe biljaka – *control* (*ctrl*) i *treatment2* (*trt2*). Pod pretpostavkom da, težine biljaka u ova dva uzorka imaju normalnu raspodelu, može se koristiti t-test za dva nezavisna uzorka. Takođe, u tom slučaju je opravdano koristiti *error bar* za prikaz podataka. U literaturi se vrlo često na *error bar* graficima prikazuju statistički značajne razlike, te je na Sl. 218 prikazano kako se te razlike mogu predstaviti pomoću

390 Ovaj primer je prikazan na predavanju u okviru odgovarajućih lekcija za TOBS predmet na Univerzitetu u Beogradu – Elektrotehničkom fakultetu koje su navedene na kraju ovog udžbenika u dodatku.

error bar grafika za dve grupe biljaka. Kod koji omogućava da se generiše Sl. 218 je prikazan na Sl. 219.

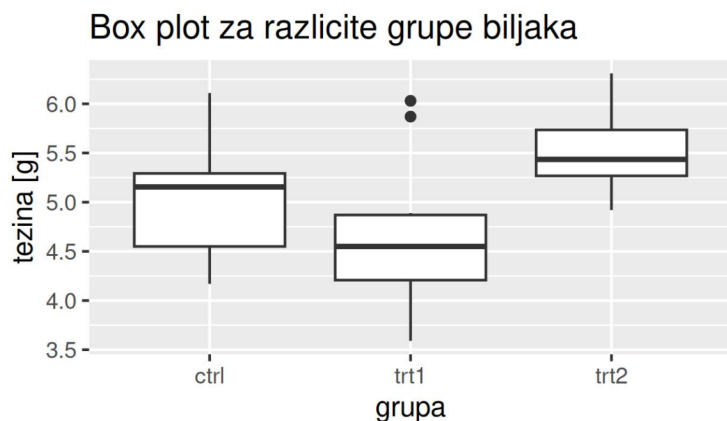


Slika 215, Error bar za različite grupe biljaka.

```
> barE <- ggplot(pod, aes(x = group, y = mean)) +
+   ggtitle("Error bar za različite grupe biljaka") +
+   xlab("grupa") + ylab("tezina [g]") +
+   geom_bar(stat = "identity",
+           fill = "cyan", col = "black")
> barE + geom_errorbar(aes(ymin = mean - sd, ymax = mean + sd),
+                   width = 0.3) + coord_flip()
> ggsave('error_bar_biljke.jpg', dpi = 300)
```

Saving 4.08 x 2.38 in image

Slika 216, R kod potreban za iscrtavanje *error bar*-a primenom ggplot2 paketa. Primititi da je bilo potrebno pripremiti podatke na odgovarajući način kao što je to prikazano na Sl. 177.

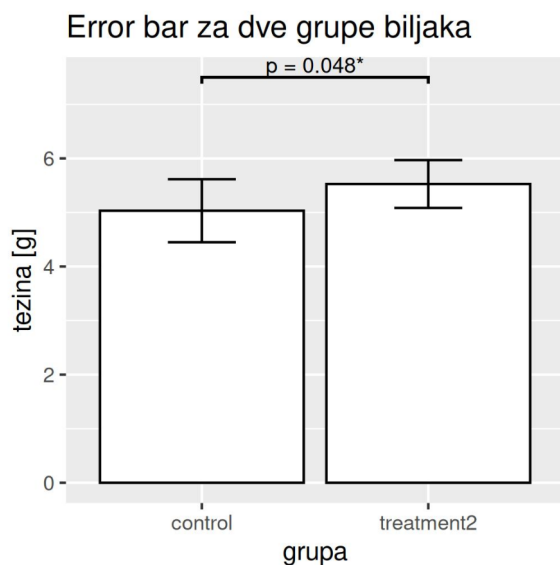


Slika 217, Box plot grafik koji prikazuje težinu biljaka u tri različite grupe.

Primititi da je za dodavanje p vrednosti na grafiku na Sl. 218 korišćena funkcija (Sl. 219):

`add_pvalue()`

iz `ggprism` paketa [548]. Pored dodavanja p vrednosti na *error bar* grafiku. Uobičajeno se koriste i oznake “*”, “***” i “****” da se ukaže na to da li je p vrednost manja od 0.05, 0.01 ili od 0.001. Za detalje korišćenja `ggprism` funkcija, čitaoci se upućuju na odgovarajuća uputstva sa primerima na CRAN sajtu³⁹¹. Sa Sl. 218 se može primititi da, iako je bliska graničnoj vrednosti od 0.05, p vrednost je manja od 0.05, odnosno $p = 0.048$. Ovakav rezultat ukazuje na to da postoji statistički značajna razlika između kontrolne grupe biljaka (*control*) i grupe biljaka pod nazivom *treatment2*. Ako se analiza vrati par koraka unazad, pre primene t-testa, mogla bi se porediti razlika srednjih vrednosti težine biljaka u ove dve grupe (dobija se da je to 0.494 g što odgovara prikazu na Sl. 215).



Slika 218, *Error bar* za dve grupe biljaka sa prikazom p vrednosti t-testa za dva nezavisna uzorka na grafiku.

Razlika srednjih vrednosti uzoraka od 0.494 g i uzorak, jesu slučajne promenljive, jer kada bi se uzeo drugi uzorak, razlika verovatno ne bi bila ista. Uzorak predstavlja deo populacije i služi da se opiše na odgovarajući način

391 Adding p -values, <https://cran.r-project.org/web/packages/ggprism/vignettes/pvalues.html>, pristupljeno 27.01.2024.

populacija. Retko kada se u statističkoj analizi raspolaže populacijom. Primera radi, ako je potrebno odrediti prosečnu visinu svih ljudi u Srbiji, bilo bi potrebno izmeriti visinu svakog stanovnika, što je zametan i nepotreban posao, posebno ako se uzme u obzir da broj stanovnika i njihova visina nisu konstantne. Prema tome, ima smisla da se visina ljudi u Srbiji ispituje na uzorku. Reprezentativni uzorak za celu zemlju sigurno ne uključuje 5 ispitanika koji su iz istog grada i pri tom u srodstvu. Takođe, nije isto ako se uzme 5 ispitanika ženskog ili 5 ispitanika muškog pola. Prema tome, nije jednostavno, ali jeste od velikog značaja, odabrati odgovarajući reprezentativni uzorak za populaciju koja se ispituje. Da bi se ilustrativno pokazalo kako uzorak može i ne mora da opisuje populaciju, može se pretpostaviti da grupe biljaka iz ispitivane dve grupe predstavljaju populacije. Kako te populacije imaju po 10 biljaka u svakoj grupi, neka uzorak čini 50% grupe, odnosno po 5 biljaka. Kada se iskoristi funkcija *sample()* jasno se vidi na Sl. 220 da razlika ne samo da može da se menja, već se menja i odnos između dve grupe.

```
> pod2 <- list("numeric")
> pod2$mean <- c(mean(ctDat$weight), mean(t2Dat$weight))
> pod2$group <- c("control", "treatment2")
> pod2$sd <- c(sd(ctDat$weight), sd(t2Dat$weight))
> pod2 <- as.data.frame(pod2)
> result <- t.test(ctDat$weight, t2Dat$weight, paired = F)$p.value
> df_p_val <- data.frame(group1 = "control", group2 = "treatment2",
+   label = round(result, 3), y.position = 7.5, p.adj.signif = "*")
> barE <- ggplot(pod2, aes(x = group, y = mean)) +
+   ggtitle("Error bar za dve grupe biljaka") +
+   xlab("grupa") + ylab("tezina [g]") +
+   geom_bar(stat = "identity",
+     fill = "white", col = "black")
> barE + geom_errorbar(aes(ymin = mean - sd, ymax = mean + sd),
+   width = 0.3) + # coord_flip() +
+   add_pvalue(df_p_val, label = "p = {label}{p.adj.signif}") #,
> #coord.flip = TRUE) # za rotaciju koordinata
> ggsave('error_bar_p_biljke.jpg', dpi = 300)
Saving 2.9 x 2.38 in image
```

Slika 219, Kod za prikaz *error bar*-a sa dodatom *p* vrednošću koji je prikazan na Sl. 218.

Za veliki broj ponovnih pokretanja koda na Sl. 220, primećuje se da, u većini slučajeva, razlika jeste negativna, odnosno težina biljaka u grupi *treatment2* je veća, ali isto tako, primećuje se da i te razlike variraju u relativno velikom opsegu. Da bi se primenio odgovarajući statistički test i da bi se proverilo da li se iz podataka o uzorku mogu doneti zaključci o populaciji, potrebno je postaviti početnu pretpostavku odnosno 0. hipotezu na sledeći način:

Ne postoji razlika između težine biljaka u kontrolnoj grupi (*control*) i u grupi pod nazivom *treatment2*. U tom slučaju bi se za različite uzorke u ove dve grupe dobile vrednosti koje imaju Gausovu raspodelu sa srednjom vrednoću 0.

```
> mean(sample(ctDat$weight, 5)) - mean(sample(t2Dat$weight, 5))
[1] -0.964
> mean(sample(ctDat$weight, 5)) - mean(sample(t2Dat$weight, 5))
[1] -0.302
> mean(sample(ctDat$weight, 5)) - mean(sample(t2Dat$weight, 5))
[1] -0.642
> mean(sample(ctDat$weight, 5)) - mean(sample(t2Dat$weight, 5))
[1] -0.37
> mean(sample(ctDat$weight, 5)) - mean(sample(t2Dat$weight, 5))
[1] 0.044
```

Slika 220, Uticaj pseudoslučajnog odabiranja različitih uzoraka iz dve grupe biljaka na razlike u srednjim vrednostima.

Praktično, ovo znači da ako ne postoji statistički značajna razlika između dve grupe, onda bi različiti uzorci imali razliku za ove dve grupe koja je jednaka 0. Prema Centralnoj graničnoj teoremi, razlika srednjih vrednosti uzoraka bi težila nuli, a rezultujuće razlike bi imale Gausovu raspodelu sa srednjom vrednošću 0. Statistički testovi služe da se proveri ovako postavljena, tzv. nulta hipoteza. Za primenu t-testa, potrebno je izračunati test statistiku ili t statistiku. Kako postoji pretpostavka da je raspodela Gausova, onda se može izračunati standardna greška za razliku dva uzorka. U Relaciji (26) je prikazana standardna greška razlike dve aritmetičke sredine za uzorke x i y dimenzija M i N i za pripadajuće varijanse (kvadrate standardnih devijacija) u oznakama σ_x^2 i σ_y^2 .

$$ts = \frac{\bar{Y} - \bar{X}}{\sqrt{\frac{\sigma_x^2}{M} + \frac{\sigma_y^2}{N}}} \sim N(0, 1) \quad \text{Relacija (26)}$$

Drugim rečima, t statistika u oznaci ts (Relacija (26)) je jednaka količniku razlike srednjih vrednosti \bar{X} i \bar{Y} za dva uzorka koji je podeljen standardnom devijacijom razlike dva uzorka. Očekuje se da t statistika takođe teži Gausovoj raspodeli sa srednjom vrednošću 0 i jediničnom varijansom u oznaci $N(0, 1)$ kako je prikazano u Relaciji (26), u slučaju da je nulta hipoteza tačna, naravno pod pretpostavkom da je na raspolaganju uzorak odgovarajuće dimenzije (da u

uzorku ima dovoljno elemenata i da su M i N dovoljno veliki). Primenom R koda na Sl. 221 se može jednostavno izračunati t statistika za odbrane dve grupe biljaka.

Da bi se iz t statistike dobila verovatnoća, odnosno p vrednost primenjenog t -testa, potrebno je primeniti $pnorm()$ funkciju koja omogućava da se proveri kolika je verovatnoća da promenljiva ima vrednost veću od t statistike. Inače, $pnorm()$ funkcija generiše funkciju raspodele verovatnoće (eng. *Cumulative Probability Distribution*) koja se može predstaviti Relacijom (27) gde je sa $f(x)$ predstavljena funkcija gustine verovatnoće [78].

$$F(ts) = \int_{-\infty}^{ts} f(x) dx \quad \text{Relacija (27)}$$

Može se reći da funkcija raspodele verovatnoće za t statistiku ts u oznaci $F(ts)$ odgovara verovatnoći nalaženja ts za različite uzorke u intervalu $[-\infty, ts]$ [78]. Verovatnoća p je na Sl. 221 proračunata na obrnut način, odnosno odgovara verovatnoći da se ts za različite uzorke ne nađe u intervalu $[-\infty, ts]$ odnosno da se dobije verovatnoća koja ukazuje da li je 0. hipoteza ostvariva. U ovom slučaju, verovatnoća da je nulta hipoteza ostvariva je 0.033, što je manje od uobičajene granice od 0.05, te se 0. hipoteza, koja tvrdi da razlike ne postoje, odbacuje i dolazi se do zaključka da ova dva skupa biljaka imaju statistički značajne razlike u težinama. Granica za odlučivanje od 0.05 ovde ima praktični značaj, jer ukazuje na to da bi se za ponovljena merenja uzoraka u 5% slučajeva javili uzorci na kojima nema statistički značajne razlike.

```
> N = length(ctDat$weight)
> tstat <- ( mean(t2Dat$weight) - mean(ctDat$weight) )
+ sqrt( var(ctDat$weight)/N + var(t2Dat$weight)/N )
> tstat
[1] 2.13402
> p <- 2 * (1 - pnorm(tstat))
> p
[1] 0.03284111
> p < 0.05
[1] TRUE
```

Slika 221, Računanje t statistike (promenljiva $tstat$) za dva uzorka, odnosno za dve grupe biljaka sa različitim težinama. P vrednost se dobija iz $pnorm()$ funkcije i množi sa dva, jer je t -test realizovan dvosmerno (nije unapred poznato da li je razlika srednjih vrednosti pozitivna ili negativna).

U zavisnosti kako su postavljene 0. i alternativna pretpostavka, odnosno hipoteza, može se koristiti jednosmerni (eng. *one-tailed test*) ili dvosmerni test (eng. *two-tailed test*). Dvosmerni test je češće u upotrebi i koristi se kada nije izvesno pre testiranja da li će ishod, odnosno razlika srednjih vrednosti uzoraka koja se testira biti pozitivna ili negativna [549]. Iako jednosmerni t-test daje duplo manju p vrednost (pogledati Sl. 221), koristan je jer se ne može uvek predvideti u kom se smeru je posmatrana promena. Ovo se najbolje vidi na Sl. 217 na kojoj su prikazane sve grupe biljaka. Logično je pretpostaviti da *treatment1* i *treatment2* predstavljaju težinu biljaka koje su dodatno tretirane sa ciljem poboljšanja prinosa, te je očekivano da će se u obe grupe naći biljke veće težine od težine biljaka u kontrolnoj grupi, što nije slučaj.

Jedan važan korak je preskočen u prikazanoj analizi razlike srednjih vrednosti težine biljaka, a to je da nije provereno da li podaci iz ove dve grupe imaju normalnu raspodelu. Jednostavnom primenom Šapiro-Vilk testa ili drugim metodama za ocenu raspodele, može se zaključiti da podaci imaju Gausovu raspodelu.

```
> t.test(ctDat$weight, t2Dat$weight, paired = F)

Welch Two Sample t-test

data: ctDat$weight and t2Dat$weight
t = -2.134, df = 16.786, p-value = 0.0479
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.98287213 -0.00512787
sample estimates:
mean of x mean of y
 5.032     5.526
```

Slika 222, Prikaz rezultata ugrađenog t-testa primenom *t.test()* funkcije u R-u koji podrazumeva Studentovu t-raspodelu.

U praksi se retko sreću uzorci sa Gausovom raspodelom, posebno ako sadrže, kao u prikazanom primeru, samo 10 elemenata. Zato se uobičajeno raspodela, koja ima delimična poklapanja sa Gausovom raspodelom, aproksimira Studentovom t-raspodelom. Računanje verovatnoće p za Studentovu t-raspodelu je nešto kompleksnije, te neće biti detaljno prikazano u ovom udžbeniku. Na Sl. 222 je prikazan rezultat ugrađenog t-testa koji se zasniva na pretpostavci Studentove t-raspodele.

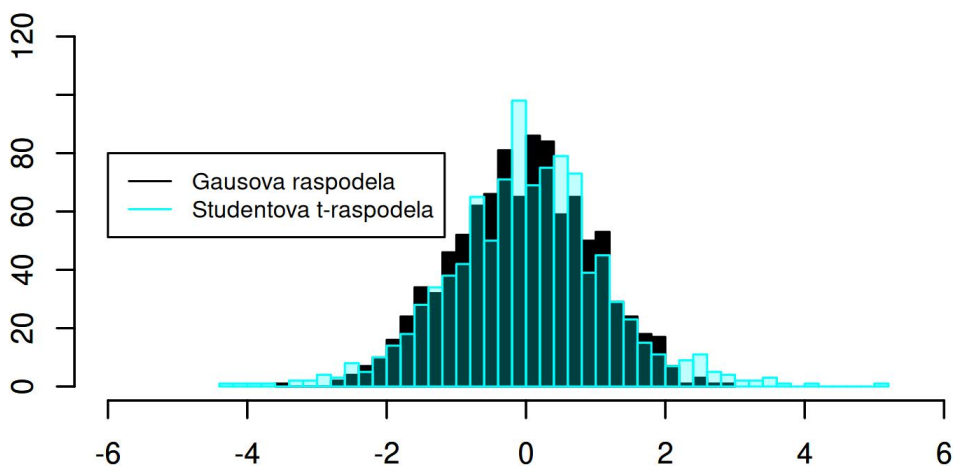
Primititi da je rezultat poređenja dve grupe isti ako se primeni ugrađena funkcija za t-test u R-u (Sl. 222), odnosno da je p vrednost i dalje manja od granice 0.05, te je i u komandnom prozoru ispisano da je tačna alternativna pretpostavka kojom je izraženo tvrđenje da postoji statistički značajna razlika između dva posmatrana uzorka. Kada bi se ugrađena funkciji za t-test pozvala tako da grupe zamene mesta, dobila bi se pozitivna t statistika, što je logično ako se pogleda Relacija (26) kojom je određena t statistika. Dodatno, potrebno je primititi da izraz za t statistiku u Relaciji (26) postaje veći za veći broj elemenata u uzorku, te treba imati na umu da t statistika zavisi od broja uzoraka, kao i p vrednost. Što je veći uzorak, to su veće i t statistika i p vrednost. Takođe, t statistika se može interpretirati kao odnos signala i šuma. Kada bi se prikazale raspodele za dva uzorka za koje se računa t statistika, jednostavno se dolazi do zaključka da razlika srednjih vrednosti uzoraka definiše očekivan rezultat (signal ili korisnu informaciju), a da varijanse predstavljaju rasipanje oko te srednje vrednosti (šum) [550]. Što je imenilac veći, to znači da je raspodela šira, te može doći do preklapanja raspodela i manjih razlika između uzoraka koji se ispituju. Razlika između t-testa za normalnu i t-testa za Studentovu raspodelu je u p vrednosti koja je nešto veća za ugrađeni t-test i iznosi 0.048 (Sl. 222).

Parametar df na Sl. 222 označava broj stepeni slobode u Studentovoj t-raspodeli. Dodatno, rezultat t-testa sa Sl. 222 sadrži i interval poverenja od 95% (eng. *confidence interval*) i srednje vrednosti oba uzorka (eng. *mean of*). Interval poverenja ima praktičnog značaja, jer se tumači kao verovatnoća (od 95% na Sl. 222) da će se razlika srednjih vrednosti dva uzorka, odnosno razlika težine iz dve grupe biljaka, naći u intervalu od -0.983 g do -0.005 g, ako bi se uzorkovanje ponovilo. Dakle, postoji 5% verovatnoća da razlika uzoraka ne bude u procenjenom intervalu. Promenom ulaznih parametara t-testa u R-u moguće je promeniti interval poverenja (koji ne mora nužno da bude 95%, iako je to uobičajen interval u naučnim i stručnim publikacijama). Očekivano, sa promenom verovatnoće intervala menja se i interval, odnosno što je verovatnoća veća, to je interval u kome se može naći razlika srednjih vrednosti dva uzorka veći.

Studentova t-raspodela je veoma korisna, kada se ne može potvrditi postojanje Gausove raspodele, a pri tom raspodela “podseća” na Gausovu. Naravno, ovo je posledica toga što uzorak ima konačan broj elemenata, ali prema Centralnoj graničnoj teoremi, Studentova t-raspodela će za povećanje broja elemenata u skupu postati Gausova raspodela [78]. Praktično je korisnije (i ekonomičnije) raditi sa Studentovom t-raspodelom, jer veličina uzoraka ne može uvek (najčešće iz ekonomskih razloga/ograničenja) da bude velika. Studentovu t-

raspodelu je uveo Vilijam Sili Goset³⁹² 1907. godine u svom radu u časopisu Biometrika [551]. Goset je radio za Ginis pivaru (eng. *Guinness*) i raspodela, na žalost, danas ne nosi njegovo ime, jer je firma zabranjivala korišćenje prezimena svojim zaposlenima u publikacijama da bi sprečila odavanje poverljivih informacija, te je Goset objavljivao svoje radove pod pseudonimom Student [552]. Ako je sa Γ označema Gama funkcija, sa v broj stepeni slobode, onda se gustina funkcije verovatnoće za Studentovu t-raspodelu za promenljivu x u oznaci $p(x, v)$ definiše kao u Relaciji (28) [78]:

$$p(x, v) = \frac{1}{\sqrt{\pi v}} \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)} \left(1 + \frac{x^2}{v}\right)^{-\frac{(v+1)}{2}} \quad \text{Relacija (28)}$$



Slika 223, Histogrami za Gausovu raspodelu i Studentovu t-raspodelu verovatnoće.

Što je broj stepeni slobode v veći to je Studentova t-raspodela sličnija Gausovoj raspodeli. Da bi se ilustrovala razlika i sličnost između Gausove i Studentove t-raspodele, u R-u je moguće generisati pseudoslučajne brojeve prema jednoj i drugoj raspodeli i uporediti, na primer, histograme. Za generisanje pseudoslučajnih brojeva može se koristiti funkcija:

`rt()`

392 William Sealy Gosset, https://en.wikipedia.org/wiki/William_Sealy_Gosset, pristupljeno 02.06.2024.

Na Sl. 223 su prikazani histogrami za 1000 generisanih pseudoslučajnih brojeva za dve raspodele: Gausovu i Studentovu t-raspodelu (sa brojem stepeni 10). Uočava se da Studentova t-raspodela, kao i Gausova raspodela, ima centralnu tendenciju (gomilanje ponovljenih merenja, odnosno elemenata u uzorku, oko centralne vrednosti), kao i da je sličnog zvonastog simetričnog oblika kao i Gausova raspodela. Studenti se upućuju na R kod koji je podeljen uz udžbenik kako bi mogli da generišu grafik sa Sl. 223. Takođe, moguće je videti efekte Centralne granične teoreme, ako se menja broj generisanih pseudoslučajnih brojeva, kao i broj stepeni slobode za Studentovu t-raspodelu.

5.2.1 Nulta i alternativna hipoteza – praktično objašnjenje

U prethodnim primerima, pokazano je kako je moguće u programskom jeziku R na odabranim podacima, realizovati statističke testove. Proces realizacije statističkih testova obuhvata sledeće korake [535]:

- Postavljanje nulte i alternativne hipoteze. Na primeru posmatrane razlike u težini sušene materije za dve grupe biljaka, nulta hipoteza se sastoji iz tvrđenja da ne postoji razlika između srednjih vrednosti težina za dve grupe biljaka, a alternativna hipoteza da postoji razlika. Kod primene Šapiro-Vilk testa, nulta hipoteza podrazumeva da posmatrani skup ima normalnu raspodelu.
- Određuje se nivo značajnosti testa, odnosno granica α , sa kojom se poredi p vrednost (verovatnoća koja se dobija kao jedan od rezultata testiranja). Najčešće se u biomedicinskom inženjerstvu, koristi granica od 0.05 ili 5%, ali se mogu primeniti i granice od 0.01 (1%) i od 0.001 (0.1%). Na *error bar* graphicima se mogu ucrtati p vrednosti, te se najčešće³⁹³ za $p < 0.05$ koristi oznaka “*”, za $p < 0.01$ oznaka “**” i za $p < 0.001$ ³⁹⁴ koristi se oznaka “***”.
- Sprovodi se račun odabranog statističkog testa. Na izbor statističkog testa utiče cilj analize (na primer: da li se želi izvršiti poređenje uzoraka ili testiranje raspodele uzoraka), priroda uzorka (na primer: ako uzorci koji se poredi imaju Gausovu raspodelu, onda se biraju parametarski, a u suprotnom neparametarski testovi³⁹⁵) i broj uzoraka koji se poredi

393 Dobra praksa podrazumeva da se u nazivu slike objasni koja oznaka (“*”, “**” i “***”) odgovara kojoj graničnoj vrednosti α sa kojom se poredi p vrednost statističkog testa.

394 Rezultat statističkog testa tj. p vrednost se najčešće zaokružuje na tri decimale [535].

395 Takođe, nije isto ako se poredi podaci koji su kategorički (na primer, sadrže informaciju o polu ispitanika) i podaci koji su numerički (na primer, puls ispitanika). Kategorički podaci nisu predmet razmatranja u ovom udžbeniku, jer se u obradi biosignala najčešće radi sa numeričkim podacima, a kategorije se koriste za odvajanje uzoraka (poređenje pulsa

(detalji poređenja više uzoraka primenom testova za analizu varijanse su prikazani u sledećem potpoglavlju).

- Na osnovu rezultata testa, donosi se odluka o prihvatanju ili odbacivanju 0. hipoteze. Na primer, ako je $p < 0.05$ kod primenjenog Šapiro-Vilk testa, to znači da podaci nemaju normalnu raspodelu. Ako je $p < 0.05$ za poređenje godina ispitanika ženskog i muškog pola, to znači da dve grupe ispitanika imaju statistički različite starosti (godine života)³⁹⁶. Generalno, ako je $p < \alpha$, onda se odbacuje 0. hipoteza i usvaja se alternativna hipoteza. I obrnuto.

Rezultat statističkih testova može biti da se dolazi do sledećeg zaključka [535]:

- odbacuje se 0. hipoteza koja se često u literaturi označava sa H_0 i prihvata se alternativna hipoteza u oznaci H_1 ili
- prihvata se H_0 i odbacuje se H_1 .

U slučaju da se prihvata alternativna hipoteza, koja uvek tvrdi da postoje razlike (za Šapiro-Vilk test to su razlike između raspodele uzorka i normalne raspodele), onda rezultat može biti dvosmerna alternativna hipoteza (kao rezultat dvosmernog ili *two-tailed* testa), kada nije poznato da li je razlika pozitivna ili negativna, i jednosmerna alternativna hipoteza (kao rezultat jednosmernog ili *one-tailed* testa), kada je poznata kog je znaka razlika koja se ispituje.

Striktno gledano, odbacivanje jedne hipoteze ne mora nužno da vodi prihvatanju druge hipoteze. Statistički testovi se sprovode za testiranje 0. hipoteze i sa početnom pretpostavkom da je 0. hipoteza tačna i proračun statističkih testova se ne zasniva na pretpostavci da je alternativna hipoteza tačna [517]. Ovo jeste važna razlika, ali za praktičnu primenu, ako postoje samo dve opcije – jedna da nema statistički značajnih razlika, a druga da ih ima, onda je primereno očekivati da su 0. i alternativna hipoteza isključive (kada je jedna tačna druga nije i obrnuto).

ispitanika kod žena i muškaraca).

396 Ovakav tip analize se najčešće sprovodi kao predobrada podataka. Ako bi se poredio puls ispitanika za različite polove, važno je da uzorci muškaraca i žena budu što sličnijih karakteristika, kako bi poređenje pulsa po polovima bilo validno (na primer: da su istog zdravstvenog stanja, da su približno istih godina i sl). U suprotnom, drugi faktori (koji nisu povezani sa polom) bi mogli da utiču na konačan rezultat i da dovedu do pogrešnih zaključaka.

5.3 Analiza varijanse na primeru podataka o težini biljaka

I t-test za normalnu raspodelu i t-test za Studentovu t-raspodelu daju isti konačan rezultat (bez obzira na razliku u p vrednostima, koja je očekivana, za koju je očekivano da je veća za Studentovu t-raspodelu, jer je ta raspodela samo aproksimacija Gausove raspodele) – težine biljaka iz *treatment2* grupe su statistički značajno veće u odnosu na težine biljaka iz kontrolne grupe. Međutim, skup podataka ne sadrži samo ova dva, već tri uzorka. Iako bi logično bilo da se, u takvim slučajevima, primene tri t-testa: za poređenje tri grupe međusobno i dobiju tri p vrednosti, to bi bio pogrešan način. Kada se poredi srednja razlika između više grupa, koristi se ANOVA test (od eng. *Analysis of Variance*). Trebalo bi napomenuti da, iako naziv testa ukazuje na analizu varijansi, test zapravo poredi srednje vrednosti na više grupa/uzoraka [553]. Konkretno, u ovom primeru se koristi *one-way* ANOVA ili na srpskom jednofaktorska analiza varijanse.

Jednofaktorska analiza se sprovodi, ako se podaci mogu podeliti po jednom principu kategorizacije, što se ovde odnosi na postojanje, odnosno vrstu dodatnog tretmana biljaka. Dakle, jedina podela je na kontrolnu grupu i na dve grupe biljaka koje su prošle dva različita tretmana. Kada bi postojalo, na primer, 5 grupa gde su biljke u 4. i 5. grupi bile izložene istim uslovima kao biljke iz grupe *treatment1* i *treatment2*, ali koje su za razliku od prve tri grupe imale duplo više svetla, onda bi bilo potrebno analizu sprovesti po grupama u odnosu na tretman i u odnosu na količinu svetla (dva faktora), te bi trebalo primeniti dvofaktorsku analizu (eng. *two-way* ANOVA).

Nulta hipoteza koja se postavlja za primenu jednofaktorske analize varijanse, jeste da između srednjih vrednosti težina tri grupe biljaka ne postoji statistički značajna razlika, odnosno sve tri grupe imaju istu srednju vrednost. Umesto t statistike, kod ANOVA testa se računa F statistika koja predstavlja odnos između razlika između grupa u odnosu na varijanse u grupama. Na sličan način kao t statistika, manja F statistika ukazuje da ne postoje značajne razlike između posmatranih uzoraka [553].

Ako se primeni ANOVA test na podacima koji su smešteni u promenljivoj *dat* (pogledati Sl. 224), onda se može primeniti funkcija:

```
aov()
```

na način kako je to prikazano na Sl. 224. Rezultat ANOVA testa (Sl. 224) ima vrednost F statistike od 4.846 i ukazuje na statistički značajne razlike, jer je procenjena p vrednost 0.0159 (< 0.05).

```
> model <- aov(weight ~ group, data = dat)
> summary(model)
          Df Sum Sq Mean Sq F value Pr(>F)
group      2  3.766   1.8832   4.846 0.0159 *
Residuals 27 10.492   0.3886
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Slika 224, Primena ANOVA testa na podacima za tri različite grupe biljaka.

Kako bi se na odgovarajući način tumačili rezultati ANOVA testa na Sl. 224, potrebno je koristiti testove višestrukih poređenja (eng. *pairwise comparison*). Bez dodatnih testova, rezultat ANOVA ukazuje samo na postojanje statistički značajnih razlika između nekih grupa, ali nije poznato između kojih grupa [554]. Ovdje će biti prikazano korišćenje Takijevog testa (eng. *Tukey post-hoc test*) [553]. Rezultat primene Takijevog testa za višestruko poređenje je dat na Sl. 225. Nakon primene Takijevog testa, postaje jasno da postoji statistički značajna razlika samo između dve grupe biljaka (p vrednost je jednaka 0.012 što je manje od 0.05) koje su prošle dva tretmana. Rezultat primene ANOVA testa se razlikuje od rezultata primene t-testa za dva nezavisna uzorka, jer Takijev test za višestruko poređenje nije pokazao statistički značajnu razliku između kontrolne grupe (ctrl) i *treatment2* grupe (trt2) – p vrednost je 0.198 (Sl. 225). Ovakav rezultat je očekivan i ovdje je prikazan, jer ukazuje na značaj korišćenja odgovarajućih testova kada se porede dve ili kada se poredi tri ili više grupa, jer rezultati mogu biti različiti.

```
> TukeyHSD(model, conf.level=.95)
Tukey multiple comparisons of means
 95% family-wise confidence level

Fit: aov(formula = weight ~ group, data = dat)

$group
      diff      lwr      upr      p adj
trt1-ctrl -0.371 -1.0622161 0.3202161 0.3908711
trt2-ctrl  0.494 -0.1972161 1.1852161 0.1979960
trt2-trt1  0.865  0.1737839 1.5562161 0.0120064
```

Slika 225, Primena Takijevog testa za višestruko poređenje.

U slučaju da je bilo potrebno primeniti neparametarski test (kada podaci nemaju Gausovu raspodelu u svim uzorcima), test koji odgovara jednofaktorskoj analizi varijanse je Kruskal-Valisov test. Šta bi se desilo kada bi na istim podacima o težini biljaka bio primenjen neparametarski Kruskal-Valisov test (eng. *Kruskal-Wallis rank sum test*) umesto jednofaktorskog testa varijanse? Na Sl. 226 je prikazana primena testa sa rezultatima. Statistički značajna razlika postoji između istih grupa podataka, ali p vrednost je nešto veća. Po ugledu na p vrednost kada je pretpostavljena Gausova i kada je pretpostavljen Studentova t -raspodela, ovo je očekivan rezultat. Generalno, neparametarski testovi rezultuju većom p vrednošću od parametarskih testova na istim podacima. Naravno, dalja analiza uključuje upotrebu odgovarajućeg testa za višestruko poređenje. Iako u ovom udžbeniku, nije predstavljena iscrpna lista statističkih testova koji su na raspolaganju, trebalo bi da je osnovni princip na kome se zasniva postavljanje 0. hipoteze i kasnije odbacivanje ili potvrda iste, kao i princip primene i odabira osnovnih testova, dovoljno jasan da studenti mogu sami da pretražuju literaturu i usavršavaju svoja znanja u oblasti statističke analize. U nastavku poglavlja su data završna praktična razmatranja, koja treba uzeti u obzir prilikom korišćenja statističkih testova.

```
> kruskal.test(weight ~ group, data = dat)
```

```
Kruskal-Wallis rank sum test
```

```
data: weight by group
```

```
Kruskal-Wallis chi-squared = 7.9882, df = 2, p-value = 0.01842
```

Slika 226, Primer primene Kruskal-Valis testa.

5.4 Praktično objašnjenje p vrednosti i problem sa p vrednošću

Da bi se ukazalo na praktični značaj p vrednosti, može se izvesti jednostavan eksperiment sa pseudogenerisanim uzorcima. Ovakva vrsta eksperimenta je pogodna, jer se može uticati na broj elemenata u uzorku, kao i na njihovu raspodelu. Na primer, neka je generisano dva uzorka od po 1000 pseudoslučajnih elemenata koji prate normalnu raspodelu sa srednjom vrednošću 0 i standardnom devijacijom 1 u oznaci $p1$ i $p2$. Primenom dvostranog t -testa za nezavisne uzorke, očekivano bi bilo da se dobije p vrednost koja je veća od 0.05 (Sl. 227), odnosno ne postoji statistički značajna razlika između ova dva uzorka. U skladu sa očekivanjima, dobija se p vrednost na od 0.108 (Sl. 227) koja je veća od 0.05, ali nije jednaka 1, odnosno 100%.

Šta bi se desilo da se iznova i iznova generišu pseudoslučajni brojevi na ovaj način? Dobile bi se različite p vrednosti koje bi, u većini slučajeva, bile veće od 0.05. Međutim, očekivano bi bilo da se ponekad dobije statistički značajna razlika, iako ona ne postoji. Da bi se proverila ova tvrdnja, može se koristiti Monte Karlo simulacija.

```
> set.seed(1)
> p1 <- rnorm(1000)
> set.seed(2)
> p2 <- rnorm(1000)
> t <- t.test(p1, p2, paired = FALSE)
> t$p.value
[1] 0.108287
```

Slika 227, Primena t-testa na dva nezavisna uzorka koji se sastoje od 1000 generisanih pseudoslučajnih brojeva koji prate normalnu raspodelu sa srednjom vrednošću 0 i standardnom devijacijom 1. Primititi da je, u predstavljenom primeru, korišćena *set.seed()* funkcija koja omogućava da se za ponovljena pokretanja na različitim računarima i u različitim vremenskim trenucima generišu iste pseudoslučajne vrednosti.

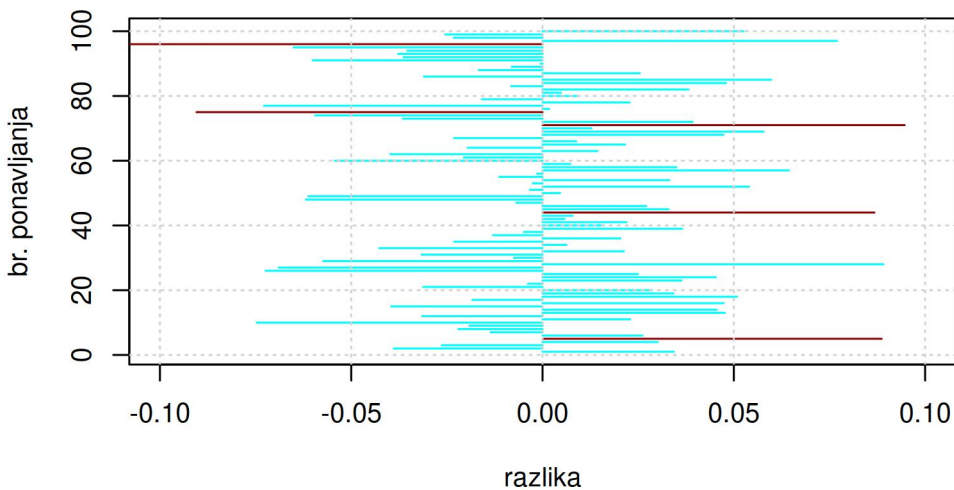
Monte Karlo metoda ili Monte Karlo simulacija je računarska metoda, koja se sastoji iz ponovljenog generisanja pseudoslučajnih brojeva, a pored ostalih primena, omogućava simulaciju merenja manipulaciju podacima koji se dobijaju iz različitih raspodela kako bi se omogućilo testiranje pretpostavki [555]. Ovde će Monte Karlo simulacija biti korišćena da se generišu različiti uzorci pseudoslučajnih brojeva, odnosno procedura prikazana na Sl. 227 će biti ponovljena 100 puta da bi se videlo da li su sve p vrednosti veće od 0.05. Za Monte Karlo simulaciju, 100 ponavljanja je relativno mali broj ponavljanja, ali ovde je odabran sa ciljem da se prikaže broj odstupanja direktno u procentima.

Na Sl. 228 je prikazan R kod za realizaciju Monte Karlo simulacije, dok je na Sl. 229 prikazan rezultujući grafik. Unutar *for* petlje na Sl. 228, u promenljivoj *ip*, smešten je brojač u kome se smešta informacija o tome koliko puta je vrednost t-testa bila manja od 0. Za svaku od 100 realizacija t-testa, prikazana je razlika srednjih vrednosti dva uzorka na grafiku na Sl. 229. Opseg razlika je bio od oko -0.1 do 0.1 za pseudoslučajne brojeve koji su generisani tako da prate Gausovu raspodelu, čija je srednja vrednost 0 i standardna devijacija 1. Takođe, pozitivne i negativne razlike su prikazane na odgovarajući način i to cijan (zelenkasto-plavom) bojom u slučaju da rezultat t-testa nije pokazao statistički značajne razlike. Međutim, kada je t-test pokazao statistički značajne razlike, one su ucrtane na grafiku na Sl. 229 tamno crvenom bojom. Može se izbrojati sa grafika da statistički značajnih razlika ima 5, što je i rezultat koji je ispisano u konzoli na Sl. 228 za promenljivu *ip*. Takođe, primećuje se da su sve

razlike prikazane tamno crvenom bojom veće od razlika koje su prikazane cijan bojom. Ako bi se dalje Monte Karlo simulacija ponovila još par puta, moglo bi se uočiti da se broj statistički značajnih razlika menja, ali da je uvek oko 5, što odgovara verovatnoći od 0.05.

```
> int <- 100
> plot(0, 0, type = "l", ylim = c(1, 100),
+      xlim = c(-0.1, 0.1) , xlab = "razlika",
+      ylab = "br. ponavljanja", main = "Simulacija")
> ip <- 0
> for (interval in 1:int) {
+   p1 <- rnorm(1000)
+   p2 <- rnorm(1000)
+   df <- mean(p1) - mean(p2)
+   t <- t.test(p1, p2, paired = FALSE)
+   color <- ifelse(t$p.value < 0.05, "darkred", "cyan")
+   ip <- ifelse(color == "darkred", ip <- ip + 1, ip <- ip)
+   lines(c(0, df), c(interval, interval), col = color)
+ }
> grid()
> ip
[1] 5
```

Slika 228, Realizacija Monte Karlo simulacije za 100 ponovljenih generisanja pseudoslučajnih uzoraka i njihovo poređenje primenom t-testa za nezavisne uzorke.



Slika 229, Grafički prikaz rezultata Monte Karlo simulacije za R kod na Sl. 228.

Sa Sl. 228 i Sl. 229 dolazi se do zaključka da, u statističkim proračunima, može doći do greške, iako je verovatnoća pojave greške relativno mala (u 5 od 100 slučajeva). U statistici postoje dva tipa greške [536, 556]:

- Tip I greške (poznata i kao α greška ili lažno pozitivna greška od eng. *False Positive*) – kada ne postoji statistički značajna razlika, ali analiza ukazuje suprotno (kao što je prikazano na Sl. 229 u Monte Karlo simulaciji).
- Tip II greške (poznata i kao β greška ili lažno negativna od eng. *False Negative*) – kada postoji statistički značajna razlika, ali sprovedena analiza daje suprotan rezultat.

Nemoguće je izbeći greške i one se najčešće dešavaju jer su podaci zašumljeni i jer uzorak nije reprezentativni za populaciju koju predstavlja. Takođe, iako šum na podacima ili pristrasnost u prikupljanju mogu da utiču na tip I i tip II greške, takvi izvori se ne klasifikuju kao tip I i tip II. [556]

Kako se pojava grešaka ne može izbeći, treba uložiti napor da se one smanje odgovarajućim odabirom uzorka. Pored osnovne pretpostavke za primenu t-testa, da su i uzorak i populacija normalno raspodeljeni, postoji još pretpostavki za primenu t-testa koje se odnose na uzorak: da su varijanse na podacima slične i da postoji isti ili sličan broj elemenata u uzorcima koji se porede. Takođe, važno je napomenuti da je praktično pokazano da t-test odlično radi za oko 20 do 30 elemenata po uzorku [550]. Primer sa težinom biljaka, koji ima 10 elemenata po uzorku, nije najpoželjnija situacija u praksi, posebno ne u biomedicinskom inženjerstvu. Da bi se sprečio pogrešan odabir elemenata u uzorku, mnogi naučni časopisi postavljaju uslove prilikom objavljivanja istraživačkih rezultata. Na primer, u časopisu *Physiological Measurement* nemoguće³⁹⁷ je objaviti rezultate, koji prikazuju poređenje merenja na grupama ispitanika, ako u svakoj grupi ne postoji 30 i više ispitanika³⁹⁸. Kao razlozi se navode mala snaga statističkih testova (smanjena moć odlučivanja i poverenja u statistički test) i smanjena klinička i medicinska verodostojnost prikazanih rezultata. Uputstva, kako predstaviti rezultate statističkih testova su uobičajena za naučne časopise, ali i druge tehničke i analitičke izveštaje. Na primer, u časopisu *Behavior Research Methods* se očekuje od Autora da provere statističku snagu testova koje koriste³⁹⁹. Pravilno je planirati broj ispitanika u kliničkoj studiji ili u studijama biomedicinskog inženjerstva unapred. Na žalost,

397 Nemoguće je možda preteška reč, zato što je Urednički odbor časopisa *Physiological Measurement* ostavio mogućnost, da im se Autori obrate, u slučaju da je u grupama ispitivano manje od 31 ispitanika uz odgovarajuće obrazloženje zašto je tako urađeno. Ali, to su izuzeci.

398 <https://publishingsupport.iopscience.iop.org/journals/physiological-measurement/about-physiological-measurement/>, pristupljeno 28.01.2024.

399 <https://link.springer.com/journal/13428/submission-guidelines#Instructions%20for%20Authors%20Statistical%20Guidelines>, pristupljeno 28.01.2024.

to nije uvek pravolinijski proces i ne postoji brzi način dolaska do potrebnog broja elemenata u uzorku, posebno što se merenja međusobno razlikuju i što je ljudski organizam kompleksan sistem, te treba uzeti puno faktora u obzir. Važno je primeniti razuman pristup u planiranju prikupljanja podataka i objasniti zašto je odabran odgovarajući broj ispitanika. Sa jedne strane, postoji problem malog uzorka koji je nedovoljan da se prikaže da li promena/e između uzoraka postoji/e ili ne, a sa druge postoji problem ako se iskoristi relativno veliki uzorak.

Kada se pogleda Relacija (26) za t statistiku, jednostavno se dolazi do zaključka da bi, za relativno veliki broj elemenata, t statistika bila veća, te broj uzoraka može uticati na pojavu relativno male p vrednosti i statistički značajne razlike između posmatranih uzoraka, iako razlika u stvarnosti ne mora biti značajna. Sledi, da ako postoji relativno mali uzorak, na kome je moguće odrediti statistički značaj, da je to dobar rezultat, jer postoji relativno velika verovatnoća da će sa povećanjem uzorka smanjiti p vrednost. Prema tome, efekat veličine uzorka nije zanemariv. Dodatno, statistički značajna razlika ne mora nužno da označava i praktično značajan napredak, kao što je objašnjeno na primeru terapije inkontinencije, na početku poglavlja. Iako su statistički testovi i p vrednosti koje se dobijaju kao rezultat testa, značajan alat u medicini i biomedicinskom inženjerstvu, ipak ih treba uzeti sa rezervom.

Da bi se prevazišao problem veličine uzorka, postoji niz tehnika koje su razvijene. Na CRAN-u se mogu naći i specijalizovani paketi kao što je `pwr` paket [557] koji nudi funkcije za računanje snage testa na osnovu Kohenovih uputstava koja se smatraju zlatnim standardom [558]. Snaga testa se može posmatrati kao verovatnoća o odbijanju H_0 hipoteze, u slučaju da je alternativna hipoteza tačna. Računanje snage testa je dobar pristup u prikupljanju podataka i analizi signala. Pored toga, predstavlja dobru etičku praksu, posebno ako se studije izvode na eksperimentalnim životinjama⁴⁰⁰, a može da uštedi vreme i resurse. Na primer, za pokazivanje učinka terapije, možda je potrebno ispitati uticaj terapije na zdravlje ispitanika za po 40 ispitanika u grupi, što je mnogo manje nego inicijalnih 100 koji su možda bili u planu za ulazak u kliničku studiju. Na ovaj način se smanjuje vreme potrebno za dijagnostičku ocenu, resursi potrebni da se terapija ponudi većem broju pacijenata i sl. U svakom slučaju, bilo kakav plan je bolji od odsustva planiranja veličine uzorka. Prilikom planiranja, na raspolaganju su i praktične preporuke kao, na primer, preporuke Kvintana, ali i Labordea i saradnika za veličinu uzorka za ispitivanje

400 Kada se izvode eksperimenti na eksperimentalnim životinjama, nakon završenog eksperimenta životinje se žrtvuju. Generalno, važi princip da bi trebalo izbeći upotrebu eksperimentalnih životinja kad god je to moguće [559].

varijabilnosti srčanog ritma [560-561]. Najveći izazov, u sprovođenju ovakvih analiza je kako odrediti veličinu efekta (eng. *effect size*), jer kada ne postoje preporuke iz literature, ni primeri podataka dobijeni iz preliminarnih istraživanja (eng. *pilot study*), jedino što ostaje jeste da se odredi proizvoljna veličina koja je najčešće jednaka 10% [556].

Veličina efekta se takođe koristi kada se želi dodatna potvrda rezultata t-testa, jer očigledno određeno nepoverenje u p vrednost postoji. Najčešće su u upotrebi Kohenova d vrednost (eng. *Cohen's d*) kao parametarska mera (kada podaci imaju normalnu ili približno normalnu raspodelu) i neparametarska Klifova δ (eng. *Cliff's delta*). U R-u postoji paket *effsize* [562], čijom se primenom jednostavno mogu dobiti parametarska i neparametarska veličina efekta. Za potrebe kvantitativne ocene, na primer, na obeležja dobijena iz EGG signala koja se koriste za odluku da li postoji mučnina ili ne, korišćena su originalna obeležja i obeležja dobijena iz veštački zašumljenih signala sa različitim SNR. Uticaj šuma na obeležja je proveravan parametarskim i neparametarskim statističkim testovima, ali i testovima za ocenu veličine efekta u studiji [382]. R kod za ovu analizu sa obeležjima je dostupan na GitHub platformi [563]. Za uzorke x i y dužina M i N i za njihove varijanse (kvadrate standardnih devijacija) u oznaci σ_x^2 i sa σ_y^2 Kohenov d parametar se definiše prema Relaciji (29).

$$d = \frac{\bar{Y} - \bar{X}}{\sqrt{\frac{(M-1)\sigma_x^2 + (N-1)\sigma_y^2}{(N+M-2)}}} \quad \text{Relacija (29)}$$

U Relaciji (29) je sa \bar{X} i \bar{Y} predstavljena srednja vrednost za dva uzorka. Moguće je uočiti koliko je izraz za Kohenovo d sličan izrazu za t statistiku datim u Relaciji (26), a može se jednostavno izvesti odnos između t statistike i Kohenovog d parametra [558, 564]. Veće Kohenovo d , označava da je promena uzorka veća i obrnuto. Kohen je predložio da se malim efektom smatra kada je $d = 0.2$, srednjim kada je $d = 0.5$, a velikim efektom kada je $d = 0.8$. Međutim, predložene vrednosti su proizvoljne i ne treba ih shvatiti bukvalno, jer od problema koji se analizira će zavisiti koliko veličinu efekta treba uzeti u obzir. U slučaju analize varijabilnosti srčanog ritma, predložene su izmenjene veličine efekta, pa se za granice umesto 0.2, 0.5 i 0.8 uzimaju 0.25, 0.50 i 0.90 [560]. Važna osobina Kohenovog d parametra jeste da ne zavisi od broja elemenata u uzorku, kao statistički testovi i to da ga je najbolje primeniti je za uzorke sličnih ili istih dimenzija (Relacija (29)).

Primenom `effsize` paketa na podatke dobijene u studiji spavanja, kada se pored vremena reakcije za 0. i 9. dan, dobija se da je Kohenovo d jednako -1.793 ili 1.793, što predstavlja veliku vrednost (> 0.8) [558, 564]. Kako razlike srednjih vrednosti u brojiocu, za računanje Kohenovog d , nisu apsolutne, logično je da se u zavisnosti koji su podaci stavljeni na prvo mesto, a koji na drugo kao argumenti funkcije u:

```
cohen.d()
```

dobiće se Kohenovo d sa promenjinim znakom (Sl. 230). U R-u se, pored vrednosti Kohenovog d , ispisuje u zagradi i veličina efekta (eng. *large* na Sl. 230), ali i interval poverenja sa verovatnoćom od 95% za Kohenovo d . U primeru studije spavanja, nedvosmisleno je pokazano i primenom statističkih testova i primenom Kohenovog d , da postoji značajna razlika između vremena reakcije koja su merena 0. i 9. dana studije.

```
> library(effsize)
> dat <- read.csv("sleepstudy.csv")
> cohen.d(dat$Reaction[dat$Days == 0], dat$Reaction[dat$Days == 9])
```

Cohen's d

```
d estimate: -1.793145 (large)
95 percent confidence interval:
      lower      upper
-2.5952229 -0.9910673
```

```
> cohen.d(dat$Reaction[dat$Days == 9], dat$Reaction[dat$Days == 0])
```

Cohen's d

```
d estimate: 1.793145 (large)
95 percent confidence interval:
      lower      upper
 0.9910673  2.5952229
```

Slika 230, Računanje Kohenovog d efekta u studiji spavanja za poređenje 0. i 9. dana.

Dobra je praksa prikazati efekat veličine, a pored toga preporuka je da treba uvek prikazati p vrednost koja se dobija (nije isto, ako je p vrednost nešto manja 0.05 ili ako je za red veličine manja od granične vrednosti). Dodatno, poželjno je pored p vrednosti prikazati i t statistiku (ili F statistiku ili odgovarajuće poređenje razlike u srednjim vrednostima, kao i varijanse

uzoraka). Takođe, mnogi Autori savetuju prikaz intervala poverenja koji je važan rezultat statističkog testa [436] i koji pomaže interpretaciji rezultata. Na kraju, jedna od osnovnih kritika p vrednosti jeste zabluda da se rezultat neke analize može predstaviti isključivo jednim brojem, kao i da su 0. i alternativna hipoteza nužno isključive (kada je jedna tačna druga nije i obrnuto). Ovo su neki od glavnih argumenata za korišćenje intervala poverenja, pored p vrednosti, prilikom predstavljanja rezultata statističkih testova. Treba pažljivo pristupiti oceni statističkih testova, ali i kritici p vrednosti, jer dok se ne pronade bolje rešenje, testovi za ispitivanje statističke značajnosti će ostati u upotrebi. [536, 565]

6 Dobre prakse u analizi biosignala i u radu u programskom jeziku R

Odabrane smernice za analizu biosignala, u programskom jeziku R, su predstavljene u ovom poglavlju: rad sa otvorenim podacima i slobodnim softverom, osnovne tehnike u detekciji i otklanjanju grešaka (debugovanje) u programskom kodu sa osvrtom na dobre i loše prakse, metode za generisanje automatskih izveštaja, estetika programskog koda i osvrt na indeksiranje u R-u (i drugim programskim jezicima). Spisak dobrih praksi nije konačan ni iscripan, a ima za cilj da ponudi alate i metode za praktični rad u oblasti analize biosignala.

6.1 Otvoreni podaci i slobodan softver

Deljenje podataka i softverskog koda, pored društvene koristi koja je usmerena ka ubrzanom razvoju biomedicinskog inženjerstva i drugih grana, donosi i ličnu prepoznatljivost i povećanu vidljivost inženjera koji dele podatke i softver. Medicinski softver, koji se koristi u kliničkoj praksi, kao i hardver, mora proći odgovarajuću sertifikaciju. Na primer, Epl aplikacija ECG App za merenje, analizu i prikaz EKG signala je dobila FDA⁴⁰¹ (eng. *Food and Drug Administration*) dozvolu za korišćenje u medicinske svrhe, uz određena ograničenja⁴⁰². Za ovakva odobrenja medicinskog softvera, potrebno je izvršiti niz testiranja uređaja za koji bi trebalo da pokažu da li je koristan u dijagnostici. Za softver u razvoju i za softver koji se koristi u istraživačke svrhe, nije potrebno takvo odobrenje, ali bi trebalo i dobra je praksa da, kada se softver za analizu biosignala deli javno, dodati napomenu da se takav softver ne može koristiti u medicinske svrhe.

Otvoreni podaci se mogu naći u repozitorijumima opšte namene (na primer, u Zenodu⁴⁰³ ili Figshare⁴⁰⁴) ili u specijalizovanim repozitorijumima. Za biosignale

401 FDA u Americi obavlja sličnu funkciju kao Agencija za lekove i medicinska sredstva Srbije (ALIMS, <https://www.alims.gov.rs/o-agenciji/>, pristupljeno 01.02.2024.) s tim što FDA ima daleko širi spektar delovanja, a posebno je važna njena uloga u odobravanju i testiranju novih medicinskih uređaja za izlazak na tržište.

402 https://www.accessdata.fda.gov/cdrh_docs/reviews/DEN180044.pdf, pristupljeno 01.02.2024.

403 <https://zenodo.org/>, pristupljeno 11.05.2024.

404 <https://figshare.com/>, pristupljeno 11.05.2024.

se dominantno koristi PhysioNet⁴⁰⁵ repozitorijum. Sa druge strane, softver se dominantno nalazi na internet platformama koje koriste Git tehnologije – kao što su GitHub⁴⁰⁶, BitBucket⁴⁰⁷ i GitLab⁴⁰⁸. Postoje i specijalizovane platforme za deljenje softverskog koda kao što je *Software Heritage*⁴⁰⁹ ili repozitorijumi koji su zasnovani na Dataverse⁴¹⁰ veb aplikaciji koja je razvijena na Harvard Univerzitetu, a koja omogućava da se istovremeno dele podaci i prateći kod koji služi za analizu podataka [566].

6.1.1 Osnovni elementi deljenja softverskog koda

Višegodišnja kampanja Fondacije za slobodan softver u Evropi glasi “*Public Money, Public Code*”⁴¹¹ i odnosi se na odgovorno upravljanje javnim sredstvima, tj. novcem poreskih obveznika. Ova kampanja polazi od prirodne pretpostavke da ako su, na primer, naučna istraživanja finansirana sredstvima poreskih obveznika, da onda poreski obveznici imaju pravo pristupa rezultatu koji je nastao i proizvodu koju su oni platili⁴¹². Ovo se ne odnosi samo na softverski kod koji se koristi u analizi biosignala, već generalno na softver finansiran iz budžetskih sredstava.

U naučnoj zajednici postoje i dalje podeljena mišljenja u vezi deljenja koda, od toga da deljenje softverskog koda treba da bude obavezno za sve naučne publikacije za koje se zalažu pobornici pokreta za otvorenu nauku⁴¹³, do mišljenja da postoje specijalizovane softverske skripte koje ne bi trebalo da sadrže obavezu deljenja koda. Ovo se posebno odnosi na softver koji se koristi u medicinske svrhe. Na žalost, argumenti za to da se medicinski softver ne učini javno dostupnim su isključivo materijalne prirode i odnose se na procenu da industrija ne može da ostvari profit, ako je njihov softver dostupan konkurenciji. [567]

405 <https://physionet.org/>, pristupljeno 11.05.2024.

406 <https://github.com/>, pristupljeno 11.05.2024.

407 <https://bitbucket.org/>, pristupljeno 11.05.2024.

408 <https://about.gitlab.com/>, pristupljeno 11.05.2024.

409 <https://www.softwareheritage.org/>, pristupljeno 11.05.2024.

410 <https://en.wikipedia.org/wiki/Dataverse>, pristupljeno 11.05.2024.

411 Direktan prevod sa engleskog nema puno smisla, ali u slobodnom prevodu sa engleskog jezika, ovaj slogan znači, da ako se koriste budžetska sredstva za kreiranje softvera, onda bi softver trebalo da bude otvoren svima koji su doprineli budžetu (poreskim obveznicima tj. javnosti).

412 <https://publiccode.eu/>, pristupljeno 11.05.2024.

413 Otvorenu nauku nikako ne treba posmatrati isključivo kroz otvoreni pristup naučnim publikacijama (eng. *Open Access*), a koju pojedini časopisi koriste za materijalnu korist.

Po mišljenju Autorke, segmenti u kojima treba pažljivo deliti softver u medicini i biomedicinskom inženjerstvu se odnose na regulativu i upotrebljivost, kao i na moguću osetljivost podataka (ako na primer, podaci sadrže lične informacije koje mogu odati identitet ispitanika). Ako je softver sertifikovan u posebnim uslovima, na primer, uz korišćenje određene vrste senzora i protokola, a pri tome testiran samo za jednu grupu ispitanika ili pacijenata, onda se deljenjem tog koda ne sme uvesti zablude da je softver efikasan za dijagnostiku ili lečenje svih oboljenja, već da je potrebna dalja sertifikacija. Ovaj princip podjednako se primenjuje i na otvoreni hardver [119]. Upotrebljivost se isto, odnosi na softver koji je neodvojiv od podataka koji mogu biti osetljivi ili zaštićeni pod licencom koja ne omogućava deljenje. To je situacija sa unapred obučanim (eng. *pretrained*) modelima mašinskog i dubokog učenja. Prema tome, prilikom diseminacije algoritama mašinskog i dubokog učenja treba obratiti pažnju na kompleksnost takvih algoritama, jer se oni ne moraju nužno sastojati isključivo iz softverskog koda, već i iz dijagrama računarskog toka (eng. *computational workflow*) i podataka [50].

Da bi se softver podelio, potrebno je ispuniti nekoliko preduslova, koji uključuju poznavanje pravnih osnova, kao što su intelektualna svojina i odabir licence, ali i poštovanje dobrih praksi kao što su FAIR principi. Prvobitno, FAIR principi su definisani za digitalne objekte sa posebnim akcentom na podatke [116]. Međutim, softver dolazi sa većom kompleksnošću. Softver se može unaprediti kroz različite verzije (podaci se samo jednom mere i iako se skup može uvećavati, oni se ne mogu promeniti), može imati kompleksne zavisnosti sa drugim softverskim kodovima ili čak sa hardverom, a softver može i generisati podatke (kao što je bilo pokazano u poglavlju o generisanju biosignala). Iz tih razloga su uloženi napor da se FAIR principi prilagode za softver [568].

Slično, kao i za deljenje podataka pod FAIR principima, potrebna su dva osnovna elementa: (1) repozitorijum koji omogućava da se ispune FAIR principi i (2) unos dovoljnog broja metapodataka (opisnih podataka o podacima). Objašnjenje elemenata koje repozitorijum treba da ispuni da bi deljen softver bio FAIR prevazilazi gradivo pokriveno ovim udžbenikom i nije ni potrebno budućim inženjerima, ali razjašnjenje nekih osnovnih termina je prikazano u udžbeniku, sa ciljem da se studenti osposobe da sami pronalaze i dele softverski kod. Repozitorijum⁴¹⁴ predstavlja uređenu zbirku digitalnih dokumenata. U kontekstu otvorene nauke, ta dokumenta, u najvećem broju slučajeva, predstavljaju naučne radove, ali mogu biti i drugi dodatni materijali

414 <https://sr.wikipedia.org/sr/Repozitorijum>, pristupljeno 11.05.2024.

(eng. *supplementary material*), softverski kod ili projektna dokumentacija hardvera. Najčešće su u upotrebi institucionalni ili nacionalni repozitorijumi, koji ispunjavaju FAIR principe o čemu se staraju nadležne osobe na fakultetima, institutima ili ministarstvima. Pored toga, postoje i opšti repozitorijumi kao što je Zenodo u čije je održavanje i uvođenje uključen veći broj međunarodnih tela. Da bi repozitorijumi ispunili FAIR principe, potrebno je da omoguće dodelu trajnih identifikatora⁴¹⁵ (eng. *persistent identifier*⁴¹⁶) i mogućnost da se metapodaci slažu u prepoznatljive obrasce/šeme⁴¹⁷ koje omogućavaju ljudsku i mašinsku čitljivost. Podaci i softver koji su deponovani na ovaj način u repozitorijume, mogu biti pretraživi na internetu i trajno dostupni (trajnost je relativni pojam, ali je omogućena kroz upotrebu trajnih identifikatora⁴¹⁸). Potrebno je naglasiti, da za postupak otvaranja biosignala u repozitorijumima, standardne šeme nisu dovoljne da se opišu svi potrebni metapodaci. Ovo je posebno važno, zbog ponovne upotrebljivosti podataka (R od eng. *Reusable* u okviru FAIR principa). Za biosignale postoje specijalizovani formati, koji sadrže metapodatke kao što su pojačanje elektrofizioloških pojačavača, frekvencija odabiranja i broj kanala sa kojih su mereni biosignali. Najpopularniji format je EDF⁴¹⁹ (eng. *European Data Format*), odnosno novija verzija EDF+, koji je dopunjen dodatnim elementima (odnos tačke i zarez, mogućnost dodavanja anotacija i dr). Inicijalno, EDF je nastao sa ciljem arhiviranja polisomnografskih studija, odnosno biosignala merenih tokom spavanja. Kasnije su nastajali i drugačiji formati metapodataka za potrebe arhiviranja i deljenja metapodataka [569]. Nije jednostavno odabrati odgovarajući format, ali metapodatke treba dodati, poznajući najbolja iskustva iz prakse. Potrebno je prvo koristiti deljene podatke i softverski kod, da bi se razumelo koje su sve informacije potrebne za njihovo ponovno korišćenje.

Sledeća stavka koju je potrebno rešiti prilikom deljenja softvera i za koju bi repozitorijumi trebalo da imaju podršku, ali ne i rešenje, jeste odabir odgovarajuće licence (izbor licence treba da izvrše Autori). Slobodne licence se mogu podeliti u par kategorija: po restriktivnosti, po tome da li su namenjene tome da se pridruže programskim skriptama, bibliotekama/paketima ili mrežnim aplikacijama, dokumentaciji, slikama i projektnoj dokumentaciji

415 Primeri trajnih identifikatora su DOI (eng. *Digital Object Identifier*) ili *handle* identifikator.

416 https://en.wikipedia.org/wiki/Persistent_identifier, pristupljeno 11.05.2024.

417 Najpoznatiji standard/šema za uređivanje metapodataka koja omogućava mašinsku čitljivost je *Dublin Core* (https://en.wikipedia.org/wiki/Dublin_Core, pristupljeno 11.05.2024).

418 Za navođenje internet stranica ili referenci sa trajnim identifikatorima nije neophodno navoditi datum kada je pristupljeno internet stranici, jer se u slučaju postojanja trajnog identifikatora smatra da adresa stranice nije promenljiva.

419 <https://www.edfplus.info/>, pristupljeno 11.05.2024.

istraživačkog hardvera. Restriktivne licence imaju pridruženu kopyleft⁴²⁰ osobinu (eng. *Copyleft*) koja obavezuje onoga ko dalje koristi modificovanu verziju softvera mora da je zaštititi pod istom licencom. Najpoznatija i verovatno najčešće korišćena licenca slobodnog softvera GNU GPL ima kopyleft osobinu. Autorka dominantno prilikom deljenja softverskog koda koristi ovu licencu, ali za deljenje softverskih biblioteka/paketa poželjnije je koristiti GNU *Lesser GPL* dok je licenca GNU *Affero GPL* namenjena deljenju mrežnih aplikacija pod kopyleft uslovima. [14]

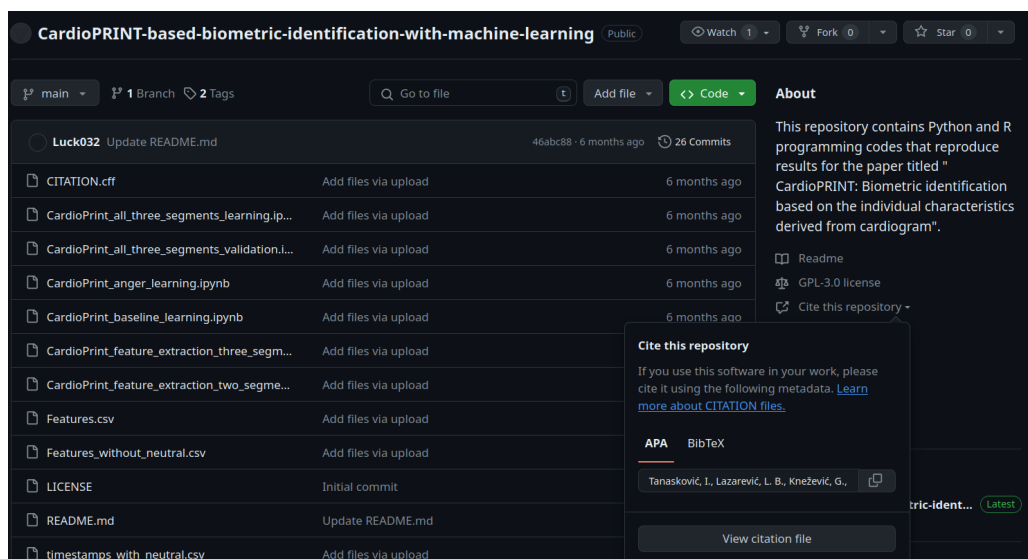
Kada se podaci dele preko GitHub-a, ne postoji mogućnost da se iskoristi *Dublin Core* šema za metapodatke, niti da se dobije DOI, ali je uspostavljena integracija između GitHub-a i Zenoda tako da se može za GitHub repozitorijum dobiti odgovarajući DOI na Zenodu⁴²¹. Metapodaci na Zenodu se moraju popuniti ručno, ali ova mogućnost je otvorila prostor da se bez dodatnih troškova deli softverski kod u skladu sa FIAR principima. Postoje i druge opcije kako se repozitorijumi na GitHub-u mogu promeniti tako da se omogući da zadovolje FAIR principe [119]. Zašto je Git⁴²² softver toliko popularan? Zato što se koristi za kontrolu verzija distribuiranog softvera [33]. Kontrola verzija podrazumeva da je omogućeno čuvanje svih izmena jedne ili više datoteka/skripti u kojima se nalazi programski kod. Distribuirana kontrola znači da više ljudi (nekad i par stotina) može istovremeno učestvovati u kreiranju softverskog koda. Git se ne mora koristiti distribuirano, može se koristiti i lokalno na računaru. GitHub je ništa drugo do internet platforma koja omogućava da se Git koristi distribuirano preko interneta, kao i da se u okviru platforme mogu kreirati korisnički profili i različiti unosi kroz projekte tj. repozitorijume. Ne moraju nužno svi GitHub projekti biti otvoreni niti je u svim slučajevima kod podeljen pod slobodnim licencama. Naprotiv, ako ne postoji podatak o licenci pod kojom je nešto vidljivo na GitHub-u, onda se automatski primenjuje kopirajt i kod se ne sme koristiti ni preuzimati (iako je to omogućeno na internetu). Ovo važi za sve materijale koji su dostupni na internetu (slike, tekstovi, video klipovi i dr). Dodatno, popularnost GitHub-a je tolika da su ostale platforme trenutno u njegovoj senci – u 2023. GitHub je imao 100 miliona korisnika [570]. Moć koju je GitHub, na taj način, postigao je dovela u opasnost da će veći deo softverskog koda, kao deo samo jedne platforme, prerasti opšte dobro i skliznuti u drugi poslovni model, isključivo

420 Krativna zajednica (eng. *Creative Commons*, skraćeno CC) je ovo osobinu, koja ponekad zbunjuje zbog svoje jezičke sličnosti sa kopirajtom (eng. *Copyright*), nazvala deljenje pod istim uslovima (eng. *Share Alike*, skraćeno SA).

421 Više u <https://docs.github.com/en/repositories/archiving-a-github-repository/referencing-and-citing-content> (pristupljeno 12.05.2024).

422 <https://en.wikipedia.org/wiki/Git>, pristupljeno 12.05.2024.

orijentisan na materijalnu korist [571]. To je verovatno jedan od razloga zašto se profesori, koji predaju GitHub, češće ograničavaju na Git softver u lekcijama, navodeći GitHub kao samo jedan od primera internet platforme i ostavljajući studentima izbor platforme. U trenutku pisanja ovog udžbenika, R Studio razvojno okruženje ima mogućnost implementacije Git softvera za verzionisanje i mogućnost direktnog povezivanja sa GitHub-om, ali je ova funkcionalnost još uvek u fazi testiranja (informacija od 12.05.2024). Za trenutno najbolje prakse integracije Git-a i R-a, pogledati otvoreni udžbenik [572]. Ovde je GitHub odabran zbog iskustva koje Autorka ima u njemu, ali naravno da studenti treba samostalno da donesu odluku koju će platformu koristiti za deljenje softvera. Za dalje usavršavanje u oblasti Git softvera, Autorka preporučuje knjige [33, 131, 572] i materijale sa otvorenog MIT kursa koji je posvećen nedostajućem gradivu sa studija računarstva [573].



Slika 231, U meniju sa desne strane, nakon sekcije *About*, postoji deo koji omogućava da se automatski izveze citat GitHub projekta i koji je označen sa *Cite this repository*. Direktn link do ove stranice je <https://github.com/Luck032/CardioPRINT-based-biometric-identification-with-machine-learning> (pristupljeno 12.05.2024.) gde studenti mogu pogledati stranicu u većoj rezoluciji.

Kada je softverski kod podeljen sa slobodnom licencom, na nekoj od internet platformi za kontrolu verzija i distribuirano programiranje, u skladu sa FAIR principima, to ne znači da je posao završen. Važan element je citiranje koda. Na GitHub-u postoji mogućnost da se automatski doda mogućnost izvoženja citata dodavanjem datoteke za citiranje sa ekstenzijom *.cff*. Na Sl. 231 dat je primer jednog repozitorijuma na GitHub-u koji je podeljen uz preprint [371] i koji

sadrži .cff datoteku. Treba napomenuti, da posao dodavanja .cff datoteke, nije u potpunosti automatizovana te da je potrebno ručno napisati .cff skriptu⁴²³. Problem sa ovakvim načinom citiranja je što nije moguće uneti više od jedne reference u .cff skriptu, te se zato savetuje svima koji dele kod i/ili podatke i/ili hardver na bilo kojoj platformi da dodaju uputstvo za citiranje. Studija za koju je podeljen kod na Sl. 231 zahteva citiranje četiri referenci, ali je to naglašeno u tekstu (pogledati Sl. 232).

How to Cite this Repository?

If you find provided code and signals useful for your own research and teaching class, please cite the following references:

1. Tanasković, I., Lazarević, L. B., Knežević, G., Milosavljević, N., Dubljević, O., Bjegojević, B., & Miljković, N. (2023). CardioPRINT-based Biometric Identification with Machine Learning (Version 1.0) [Computer software]. <https://github.com/Luck032/CardioPRINT-based-biometric-identification-with-machine-learning>, <https://doi.org/10.5281/zenodo.10204894>
2. Tanasković, I., Lazarević, L. B., Knezević, G., Milosavljević, N., Dubljević, O., Bjegojević, B., & Miljković, N. (2023, November 24). CardioPRINT: Biometric identification based on the individual characteristics derived from cardiogram. PsyArXiv preprint. <https://doi.org/10.31234/osf.io/bau7j>
3. Bjegojević B, Milosavljević N, Dubljević O, Purić D, Knežević G. In pursuit of objectivity: Physiological Measures as a Means of Emotion Induction Procedure Validation. Empirical Studies in Psychology 2020:17.
4. Tanasković, I., Lazarević, L. B., Knežević, G., Milosavljević, N., Dubljević, O., Bjegojević, B., & Miljković, N. (2023). Dataset for CardioPRINT-based Biometric Identification [Dataset]. Version 1.0. <https://doi.org/10.5281/zenodo.1020495>

Slika 232, Uputstvo za citiranje GitHub repozitorijuma dostupno na <https://github.com/Luck032/CardioPRINT-based-biometric-identification-with-machine-learning> (pristupljeno 12.05.2024).

Citiranje podeljenog softverskog koda je obaveza onoga ko koristi taj kod i u skladu je sa relativno velikim brojem licenci koje se koriste. Isto se odnosi i na programske jezike, posebno one koji su podeljeni pod slobodnom licencom. Da bi se citirao R, u naučnom ili stručnom radu (da, naravno i u diplomskim i master radovima treba primeniti isti princip), može se koristiti funkcija:

citation()

koja daje rezultat kao na Sl. 233. Da bi se proverilo pod kojom licencom je zaštićen R kod može se koristiti funkcija:

423 Savetuje se studentima da koriste interaktivne alate za uređivanje .cff datoteka, kako bi jednostavnije generisali skriptu. Može se koristiti, na primer, uređivač koji nudi GitHub <https://citation-file-format.github.io/> (pristupljeno 12.05.2024).

*licence()*⁴²⁴

koja će kao rezultat dati licencu i dodatno objašnjenje, ako postoje elementi programskog jezika koji su pokriveni različitim verzijama licence. R je zaštićen sa GNU GPL licencom.

```
> citation()
```

To cite R in publications use:

```
R Core Team (2023). _R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
```

A BibTeX entry for LaTeX users is

```
@Manual{,  
  title = {R: A Language and Environment for Statistical Computing},  
  author = {{R Core Team}},  
  organization = {R Foundation for Statistical Computing},  
  address = {Vienna, Austria},  
  year = {2023},  
  url = {https://www.R-project.org/},  
}
```

We have invested a lot of time and effort in creating R, please cite it when using it for data analysis. See also 'citation("pkgname")' for citing R packages.

Slika 233, Uputstvo za citiranje programskog jezika R u naučnim i stručnim radovima. Rezultat je dobijen u konzoli R Studio razvojnog okruženja kao rezultat funkcije *citation()*.

Specifični R paketi imaju različite reference, koje se mogu videti, kada se na CRAN-u pogleda njihov repozitorijum. Na primer, na CRAN stranici za *ggplot2*⁴²⁵ paket, postoji deo koji se odnosi na licencu (*License*) i deo koji se odnosi na citiranje (*Citation*), pa je relativno jednostavno i citirati ovaj paket i proveriti licencu za slučaj da se želi menjati i ponovo upotrebiti podeljen kod. Međutim, to na žalost nije slučaj sa svim paketima podeljenim na CRAN-u, te se onda korisnici moraju snaći kako da sami citiraju odgovarajući paket, odnosno softversku biblioteku. Sve što je navedeno, treba koristiti za sve programske jezike koji su podeljeni pod licencama slobodnog softvera (na primer: GNU Octave i Pajton). A, da li treba citirati R Studio? Ne u formi reference, ali je potrebno navesti u stručnim i naučnim radovima isto kao druge

424 U R-u je validna funkcija koja ima naziv kao i američka verzija reči *license()* za licencu.

425 <https://cran.r-project.org/web/packages/ggplot2/>, pristupljeno 12.05.2024.

softverske ili hardverske proizvode. Evo primera kako se navodi R Studio na engleskom jeziku:

R Studio Integrated Development Environment (Posit, Boston, USA)

i na srpskom jeziku:

R Studio integrisano razvojno okruženje (Posit, Boston, SAD)

Pored toga, što treba referencirati programski jezik, biblioteke i razvojno okruženje, potrebno je navesti i verziju u kojoj je razvijana skripta koja se deli ili čiji se rezultati predstavljaju u naučnim ili stručnim radovima. Uobičajeno se stavlja samo verzija programskog jezika, jer se pretpostavlja da su sve ostale verzije (biblioteka i okruženja) odabrane tako da su ažurne i kompatibilne sa verzijom programskog jezika. Da bi se dobila verzija programskog jezika R, potrebno je iskoristiti sledeću funkciju:

R.Version()

čiji se rezultat može sačuvati u promenljivoj, po izboru.

6.1.2 Otvoreni biosignali: pronalaženje, ponovno korišćenje i deljenje

Iako u deljenju softvera i podataka postoje sličnosti, procesure nisu u potpunosti iste. Na primer, kao licence za otvoreno deljenje podataka, koriste se dominantno licence Kreativne zajednice (CC licence). Jedan od načina da se pronađu podaci je da se pretražuju specijalizovane baze podataka/repozitorijumi (na primer, PhysioNet, Kegl⁴²⁶ i repozitorijum dostupnih podataka za mašinsko učenje UCI⁴²⁷) ili da se vrši opšti pregled u internet pretraživačima. Specijalizovan alat je Gugl alatka za pretragu podataka *Google Dataset Search*⁴²⁸, u kojoj se može pomoću ključnih reči doći do otvorenih biosignala. Najveći problem, prilikom otvaranja biosignala jeste što u velikom broju slučajeva ti podaci sadrže lične podatke, te ih nije moguće otvoreno deliti pre primene anonimizacije podataka. Podaci koji se dela prema Zakonu o zaštiti podataka o ličnosti ("Sl. glasnik RS", br. 87/2018) u Republici Srbiji⁴²⁹ i prema Opštoj

426 <https://www.kaggle.com/>, pristupljeno 12.05.2024.

427 <https://archive.ics.uci.edu/>, pristupljeno 12.05.2024.

428 <https://datasetsearch.research.google.com/>, pristupljeno 12.05.2024.

429 https://www.paragraf.rs/propisi/zakon_o_zastiti_podataka_o_licnosti.html, pristupljeno 12.05.2024.

regulativi za zaštitu podataka o ličnosti u Evropskoj uniji (eng. *General Data Protection Regulation*, skraćeno GDPR) ne smeju sadržati informacije koje mogu direktno ili indirektno otkriti nečiji identitet. Iako u Republici Srbiji nije na snazi GDPR, ova regulativa se odnosi na sve građane Evropske unije, a kako se studije često izvode u saradnji sa drugim institucijama, onda ima smisla poznavati regulative drugih zemalja.

Nakon procesa anonimizacije i nakon razrešenja pripadnosti podataka⁴³⁰, podaci se mogu deliti javno, pod otvorenom licencom. Za neke podatke, to je jednostavno, te je potrebno obrisati lične podatke, kao što su ime i prezime i sl. Međutim, postoje situacije u kojima to nije jednostavno. Ako bi se prosledio upitnik nastavnom osoblju Fakulteta, sa pitanjem ko je koje godine rođen (što je javni podatak za sve nastavnike Fakulteta) i na kojoj Katedri radi, kao i koliko godina iskustva ima i kog je pola, na skoro svim fakultetima u Srbiji, bi bilo u potpunosti jasno o kojoj osobi se radi. Iako nijedan podatak ne odaje pojedinačno jednu osobu, njihova kombinacija to čini, te je potrebno preduzeti odgovarajuće mere. Najjednostavnija metoda je da se obrišu pojedini podaci, ali se može desiti da se tada gubi korisna informacija. U tom slučaju, postoji opcija da se umesto, na primer, godine rođenja napiše opseg godina rođenja, te da se doda odgovarajuća neodređenost u skupu koja bi sprečila da se otkrije identitet osobe na koju se podaci odnose. Druga varijanta je da redosled podataka bude različit, te da, na primer, ukupan broj osoba muškog i ženskog pola bude isti, ali da konkretan podatak ne bude povezan sa osobom kojoj pripada. Da bi se pojednostavio princip anonimizacije podataka, razvijeni su i namenske softverske aplikacije koje se mogu koristiti za anonimizaciju kao što je aplikacija Amnezija⁴³¹, koja je nastala kao rezultat OpenAIRE projekta (jedan od rezultata ovog projekta, u saradnji sa Evropskom agencijom za nuklearna istraživanja, tj. CERN-om⁴³² je i Zenodo repozitorijum). Detalji o korišćenju Amnezija softverske aplikacije na srpskom jeziku se mogu naći na sajtu projekta koji je realizovao Tim za upravljanje istraživačkim podacima (Serbia.RDM tim od eng. *Research Data Management*) sa Univerziteta u Beogradu⁴³³.

Upravljanje istraživačkim podacima (RDM) je važan korak u pravilnom deljenju podataka [574]. Slično, postoji i upravljanje istraživačkim softverom,

430 U slučaju da studenti dele podatke, moraju prethodno to raditi u dogovoru sa predmetnim nastavnicima ili mentorima.

431 <https://github.com/dTsitsigkos/Amnesia>, pristupljeno 14.05.2024.

432 Od eng. *The European Organization for Nuclear Research*, <https://home.cern/>, pristupljeno 13.05.2024.

433 <https://rdm.open.ac.rs/index.php/sharing-publishing/security>, pristupljeno 14.05.2024.

za koje postoje uputstva u Holandiji na inicijativu više institucija [575]. Iako se ovaj udžbenik bavi analizom signala, a ne prikupljanjem i deljenjem podataka, važno je znati dobre prakse u deljenju istraživačkih podataka, kako bi se otvoreni istraživački podaci odabrali na odgovarajući način. Kvalitetno podeljeni podaci bi trebalo da poštuju FAIR principe. Veb portal Serbia.RDM⁴³⁴ Tima sa Univerziteta u Beogradu nudi informacije o različitim elementima upravljanja istraživačkim podacima na srpskom jeziku. Ukratko, upravljanje istraživačkim podacima (eng. *research data stewardship*) predstavlja dugotrajnu i održivu brigu o podacima, te je neizostavni deo kliničkih istraživanja [576].

Jedan od osnovnih elemenata upravljanja istraživačkim podacima je objašnjen u ranijim poglavljima, a odnosi se na odabir formata za čuvanje podataka, međurezultata i rezultata analize. Osnovni princip jeste da ne treba birati zatvorene, odnosno komercijalne formate. Primer loše odabranog formata je .xlsx. Način kako upravljati ovim formatom, tj. recept za njegovo otvaranje i izmene nije javno dostupan i u rukama je privatne kompanije, koja može i ne mora da postoji u narednim godinama i koja može i ne mora da odobri otvaranje takvih formata u budućnosti. Otvoreni format⁴³⁵ služi za smeštanje digitalnih podataka i definisan je pomoću objavljene specifikacije, koju u najvećem broju slučajeva održava regulatorno telo i koju može implementirati svako ko to želi, bez posebne dozvole ili uslova pod kojima je moguće korišćenje formata [577]. U praksi se često sreće zabluda, da je dobro dokumentovan format isto što i otvoren format. To nije tačno, jer otvoren format mora da ispuni uslov da može da ga otvori bilo ko i bilo kada. U ovom udžbeniku, posebna pažnja je data otvorenim formatima, odnosno podacima koji su dostupni u datotekama sa ekstenzijama koje pripadaju otvorenim formatima, kao što su .csv i .txt.

Koraci koje je potrebno preduzeti, pre merenja podataka, uključuju različite aspekte upravljanja istraživačkim podacima, ali i pribavljanje Etičke dozvole, kao i registraciju randomiziranih kontrolisanih studija (na primer, na portalu namenjenom kliničkom istraživanjima⁴³⁶). Jedan od koraka, pre planiranja upravljanja podacima i softverom (ali i pre plana statističke analize – snagu testa je potrebno proceniti pre nego što se odredi broj ispitanika na kojima će merenje da se izvrši) jeste da se preispita korišćenje postojećih podataka ili otvorenih podataka, kako bi se izbegli dodatni i nepotrebni troškovi

434 <https://rdm.open.ac.rs/>, pristupljeno 14.05.2024.

435 https://en.wikipedia.org/wiki/Open_file_format, pristupljeno 14.05.2024. Lista otvorenih formata je dostupna na https://en.wikipedia.org/wiki/List_of_open_file_formats, pristupljeno 14.05.2024.

436 <https://clinicaltrials.gov/>, pristupljeno 14.05.2024.

istraživanja. Iako ovde nisu od posebnog interesa, jer je glavni cilj udžbenika da predstavi metode za analizu i korišćenje već dostupnih podataka, ova znanja predstavljaju osnovu za rad u praksi biomedicinskih inženjera. [576]

6.2 Kako izbeći greške?

U analizi podataka i programiranju, greške se često dešavaju i neizbežne su. Važno je uočiti i korigovati greške na vreme i na odgovarajući način. Takođe, postoji uspostavljen proces analize podataka, koji može biti od pomoći, kako bi se tokom realizacije izbegle greške ili njihova pojava smanjila na minimum. Odgovorno upravljanje greškama je poželjno i predstavlja potrebnu vrlinu biomedicinskih inženjera. Podeljen softverski kod donosi dodatnu pogodnost, jer omogućava da stručna i naučna zajednica učestvuje u testiranju i javnoj oceni, odnosno recenziji rešenja. Za preuzimanje koda, sa izvora koji nisu pouzdani (kod koji nije prošao recenziju ili proveru – kao što je većina koda dostupna na GitHub-u), potrebna je doza opreza, jer se mogu preuzeti i koristan deo koda i greške.

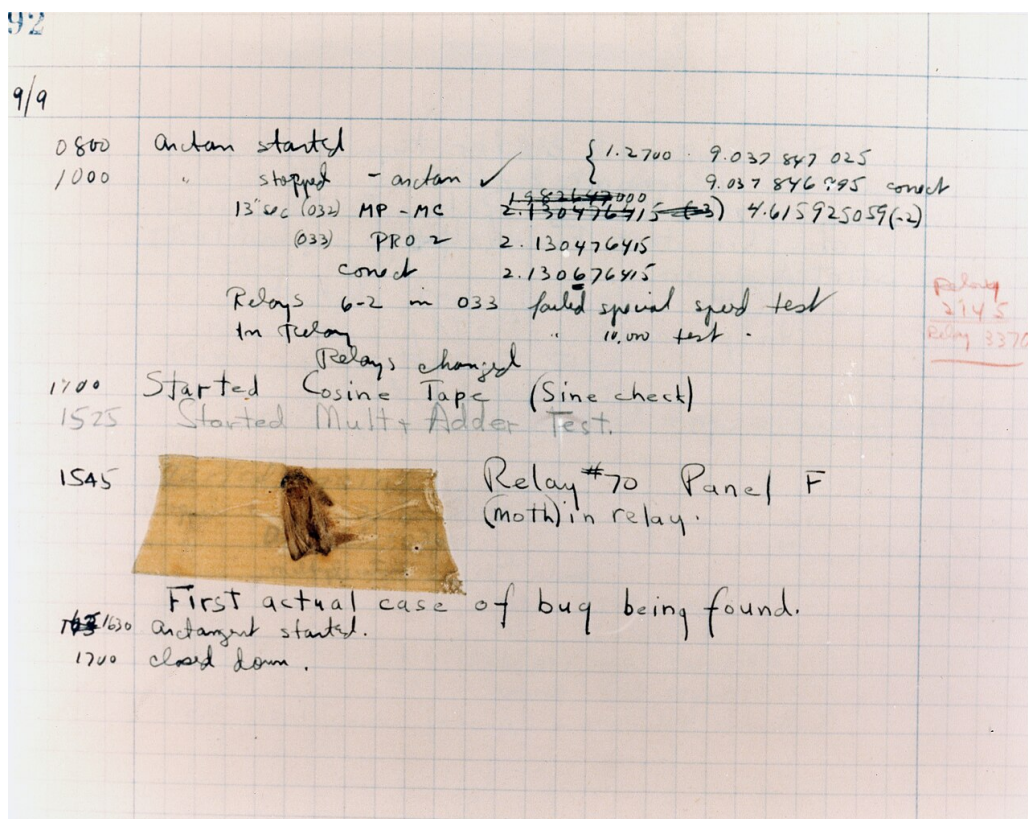
Greškama se može upravljati primenom metode debugovanja⁴³⁷ programskog koda. Prema definiciji sa Vikipedije, debugovanje je metoda pronalaska uzroka problema i rešavanje situacije koja nastaje kao rezultat projektovanog koda i sprečava željeni rad programske skripte. Prvo korišćenje ovog termina, u računarskoj tehnici i informatici je povezano sa Mark II računarom na Harvard Univerzitetu (Sl. 234) [578].

Kada se program ne izvršava na željeni način, R automatski može ispisati u konzoli: (1) poruku (eng. *message*) koja ne prekida izvršavanje koda, ali ispisuje obaveštenje kao rezultat *message()* funkcije, (2) upozorenje (eng. *Warning*), koje je vrlo slično poruci, sa razlikom što je obaveštenje rezultat *warning()* funkcije, (3) grešku (eng. *Error*), kada se ispisuje obaveštenje da je došlo do greške i zaustavlja program (obaveštenje i zaustavljanje su rezultat *stop()* funkcije) i (4) uslov (eng. *condition*) koji korisnik definiše, tj. određuje pod kojim uslovom/ima dolazi do obaveštenja i/ili zaustavljanja programa.

Greške se ne mogu u potpunosti izbeći, ali se mogu smanjiti tako što će se preduzeti preventivni koraci: (1) potrebno je proveriti vrednosti ulaznih argumenata funkcije (da li su očekivane?), (2) proveriti način na koji je funkcija pozvana (postoji mogućnost da je došlo do kombinacije podudaranja po mestu i po imenu argumenata), (3) uporediti očekivan izlaz sa dobijenim izlazom

⁴³⁷ <https://en.wikipedia.org/wiki/Debugging>, pristupljeno 14.05.2024.

funkcije (na primer: da li je signalu filtrirana niskofrekventna komponenta ili nije za slučaj da se primenjuje filter visokopropusnik), (4) da li je dobijeni rezultat realan (da li je vrednost koja se dobija moguća ili ne, na primer, da li je rastojanje RR intervala EKG signala 1 ms ili 1 min ili 1 s), (5) da li je moguće ponoviti rezultat na drugom računaru (ovaj aspekt je važan, posebno sa stanovišta računarske reproducibilnosti, a za ponovljivost rezultata treba voditi računa da se kod generisanja pseudoslučajnih brojeva koristi obavezno *set.seed()* funkcija) i (6) pre rešavanja problema razgraničiti jasno šta radi, šta daje očekivane rezultate, a šta ne radi i druge tehnike, koje mogu zavisiti od konkretnih primera podataka i metoda analize signala.



Slika 234, Fotografija je dobijena ljubaznošću Mornaričkog vojnog centra u SAD (Naval Surface Warfare Center, Dahlgren, VA., 1988. - U.S. Naval Historical Center Online Library Photograph NH 96566-KN). Fotografija je dostupna u javnom vlasništvu, <https://commons.wikimedia.org/w/index.php?curid=165211>

U ovom udžbeniku je prikazan kratak pregled postojećih alata za debugovanje u R-u, a studenti se upućuju na detalje koji su prikazani u udžbeniku prof. Penga [3]. U R-u su na raspolaganju, sledeći alati za debugovanje:

- *traceback()* funkcija koja daje informaciju o nastanku greške do tzv. korene funkcije (eng. *call stack*) i poziva se odmah nakon linije koda u kojoj je došlo do greške,
- *debug()* funkcija koja omogućava da se izvršavanje koda liniju po liniju obavlja u *debug mod*-u (u režimu za otklanjanje i ispitivanje greške programskog koda), a koristi se nekada u kombinaciji sa *undebug()* funkcijom za izlazak iz *debug* moda,
- *browser()* funkcija koja omogućava upravljanje linijama koda na način sličan kao *debug()*,
- *trace()* funkcija koja služi da se kod za debugovanje doda u funkciju, a posebno je važna za proveru funkcija koje su sastavni deo paketa i
- *recover()* funkcija koja može da modifikuje poruku greške (eng. *error behavior*).

Postoje oprečna mišljenja, po pitanju alata za debugovanje, jedni ih smatraju korisnim, a drugi ne, jer automatizacija procesa debugovanja nije idealna, te mora da postoji ljudski faktor u odlučivanju [3]. Bilo da se studenti opredele da ih koriste ili ne, potrebno je da iz tih alata usvoje osnovni pristup kako se prilazi rešavanju problema u kodu, a to su: (1) pokretanje koda liniju po liniju, (2) provera međurezultata štampanjem u konzoli, prikazom na grafiku, proverom da li su ispunjeni logički uslovi i sl, (3) pretraga korene greške kroz izlaze i argumente funkcija (slično kao međurezultati, samo se prate i elementi koji nisu od značaja za analizu), (4) kod realizovan unutar funkcije proveravati van funkcije i (5) ne dodavati korisnički interfejs pre nego što se proveru u potpunosti funkcionalnost koda.

Pored specijalizovanih alata, odnosno funkcija za proces debugovanje koda, mogu se koristiti i alati za profilisanje koda (u kombinaciji ili nezavisno od funkcija za debugovanje). Prema definiciji sa Vikipedije, alati za profilisanje koda u programskim jezicima⁴³⁸, omogućavaju da se na sistematičan način prouči koliko je vreme izvršavanja različitih delova koda. Dve su situacije kada profilisanje koda može biti potrebno: (1) kada je potrebno da se kod izvršava kraće (u slučaju da se primenjuju metode koje su vremenski zahtevne) i (2) kada se realizuju metode za obradu u realnom vremenu, pa je cilj ušteda vremena (brzo reagovanje) ili pak ušteda baterije⁴³⁹. Predmet ovog udžbenika, nije ni jedna ni druga analiza, ali studentima će verovatno biti korisno da znaju šta profilisanje znači. Profilisanje se najčešće koristi za optimizaciju koda. Najvažnije je primeniti profilisanje, tek kad je kod napisan čitko, testiran i

438 [https://en.wikipedia.org/wiki/Profiling_\(computer_programming\)](https://en.wikipedia.org/wiki/Profiling_(computer_programming)), pristupljeno 16.05.2024.

439 Na primer, takvu vrstu profilisanja bi trebalo da koriste biomedicinski uređaji čiju bateriju nije moguće često menjati – kao što su pejsmejkeri i defibrilatori.

debugovan, a nikako u fazi razvoja. Optimizaciju tj. povećanje efikasnosti koda ne treba zasnivati na pretpostavkama, već na objektivnim činjenicama, tj. kvantitativnim parametrima kakve može da pruži procedura profilisanja. U R-u postoje osnovne funkcije za profilisanje kao što je:

Rprof()

funkcija, ali i specijalizovani paketi za profilisanje. *Rprof()* pokreće proceduru za profilisanje koda i koristi se u kombinaciji sa:

summaryRprof()

funkcijom, čije samo ime kaže da ima ulogu da sumarno predstavi rezultat *Rprof()*, koji je nakon završene procedure profilisanja sačuvan u datoteci *Rprof.out*, u radnom direktorijumu. Za zaustavljanje procedure profilisanja, koristi se:

Rprof(NULL)

komanda. Važno je napomenuti, da je mana ove procedure, to što je vreme odabiranja 0.02 s, te ako je program brži od ovog vremena, prirodno se nameće zaključak da profilisanje tada nije ni potrebno. Za više informacija o ovoj proceduri, studenti se upućuju na udžbenik [3] i prateću dokumentaciju za R programski jezik. Ilustracije radi, za pokretanje jednostavnog koda:

Rprof()

set.seed(38)

p <- rnorm(45)

Rprof(NULL)

s <- summaryRprof()

View(s)

dobija se rezultat kao na Sl. 235, a u radnom direktorijumu je snimljena datoteka pod nazivom “Rprof.out” u kojoj su sadržani podaci prikazani na Sl. 235. Profilisanje se ne savetuje inženjerima koji se bave analizom biosignala (najbolje je profilisanje ostaviti stručnjaku softverskog inženjerstva, u trenutku kada je metoda za analizu u potpunosti realizovana i kada sigurno neće biti daljih izmena). Profesor emeritus Donald Ervin Knut⁴⁴⁰, sa Stanford

440 https://en.wikipedia.org/wiki/Donald_Knuth, pristupljeno 16.05.2024.

Univerziteta i dobitnik Tjuringove nagrade⁴⁴¹ je rekao da je prerana optimizacija izvor svih (ili većine) problema u programiranju⁴⁴² (“*Premature optimization is the root of all evil (or at least most of it) in programming!*”) [579]. Međutim, nekada može biti produktivno korišćenje funkcija koje ostvaruju ulogu štoperice, sa ciljem realizacije grube procene efikasnosti koda. Za to se mogu koristiti funkcije kao što su:

`system.time()`

tj. `system.time({})`. Argument ove funkcije je kod, čije se vreme izvršavanja meri u sekundama kao objekat klase `proc_time`.

Name	Type	Value
• s	list [4]	List of length 4
• by.self	list [1 x 4] (S3: data.frame)	A data.frame with 1 row and 4 columns
self.time	double [1]	0.02
self.pct	double [1]	100
total.time	double [1]	0.02
total.pct	double [1]	100
• by.total	list [16 x 4] (S3: data.frame)	A data.frame with 16 rows and 4 columns
total.time	double [16]	0.02 0.02 0.02 0.02 0.02 0.02 ...
total.pct	double [16]	100 100 100 100 100 100 ...
self.time	double [16]	0.02 0.00 0.00 0.00 0.00 0.00 ...
self.pct	double [16]	100 0 0 0 0 0 ...
sample.interval	double [1]	0.02
sampling.time	double [1]	0.02

Slika 235, Prikaz rezultata `summaryRprof()` funkcije za kod koji sadrži dve funkcije: `set.seed()` i `rnorm()`.

Za slučaj da postoji greška u kodu, ova funkcija meri vreme koje prođe do pojave greške tj. zaustavljanja koda. Postoje tri vrste proteklog vremena, koje se meri na računaru [3]:

441 Tjuringova nagrada se smatra ekvivalentnom Nobelovoj nagradi za računarsku tehniku i informatiku (https://en.wikipedia.org/wiki/Turing_Award, pristupljeno 16.05.2024).

442 U direktnom prevodu sa engleskog, citat bi glasio da je optimizacija izvor zla u programiranju. Međutim, to nije u potpunosti tačno, jer zli i dobri programi ne postoje, takve mogu biti samo namere inženjera. Često, inženjeri elektrotehnike i računarstva imaju sklonost (prof. Knut, pa ni Autorka udžbenika nisu izuzeci) da dodeljuju ljudske osobine objektima koji ih ne poseduju. Zato danas, postoje veštačka inteligencija, pametne kuće, pametni senzori, zli ili dobri programi i sl. Radi se isključivo o jezičkom stilu u struci, koji je, na žalost, često pogrešno shvaćen u javnosti.

- sistemsko vreme (eng. *system time*) je vreme koje je potrebno procesoru za izvršavanje koda i zavisi od operativnog sistema,
- korisničko vreme (eng. *user time*, skraćeno UT) je vreme koje je potrebno procesoru (u smislu izvršavanja instrukcija) i takođe, zavisi od operativnog sistema i
- proteklo tj. realno vreme (eng. *elapsed time*, skraćeno ET ili eng. *wall clock time*).

Za jednostavne komande, UT je približno isto kao ET. Kada CPU (eng. *Central Processing Unit*) ima čekanje, tada je ET veće od UT, a kada postoji procesor sa više jezgara (eng. *multi-core processor*), tada se može desiti da ET bude manje od UT. Mana *system.time()* funkcije je što ona meri sva tri protekla vremena zbirno, te se ne može koristiti da se odredi koji deo koda se duže ili kraće izvršava. Međutim, korisnik može definisati deo koda čije vreme izvršavanja želi da izmeri primenom ove metode. Ova funkcija se ne sme koristiti u kombinaciji sa funkcijom *Rprof()* za profilisanje koda. Za jednostavnu realizaciju štoperice, može se testirati jednostavan kod (Sl. 236), čiji rezultat ukazuje da je sistemsko vreme najmanje (0.002 s), dok su korisničko i realno vreme slični ($UT < ET$), što znači da su ovo jednostavne komande, kada CPU ima relativno malo čekanje/kašnjenje.

```
> system.time({
+   set.seed(38)
+   p <- rnorm(4500000)
+ })
   user  system elapsed
 0.131   0.002   0.133
```

Slika 236, Realizacija štoperice *system.time()* za procenu vremena potrebnog za izvršavanje *set.seed()* i *rnorm()* funkcija.

Pored toga, što se prilikom otklanjanja greške, mogu koristiti specijalizovane funkcije ili procedure (koje u najvećoj meri zavise od provere međurezultata metoda analize signala), pomoć se može potražiti i na internetu. Programski jezik R, uključujući i R Studio razvojno okruženje, se ističu vrlo aktivnom otvorenom zajednicom, te da je dolazak do iskusnih korisnika programskog jezika, kao i do razvojnog tima iznenađujuće jednostavan. Prvi korak je *Help* stranica, kojoj se relativno jednostavno može pristupiti, u R Studio okruženju. Drugi korak je pretraga na internetu. Međutim, pretraživači vrše automatsku pretragu, te informacije treba tražiti na odgovarajući način, koji odstupa od pravila u prirodnom jeziku, posebno ako postoji konkretno pitanje na koje se

traži odgovor⁴⁴³. Ako se studenti opredele za korišćenje Gugle alata, preporuka je da se koriste specijalizovani pretraživači, od kojih su neki već spomenuti u ranijim poglavljima:

- naučne literature (Gugl Akademik, <https://scholar.google.com/>, pristupljeno 16.05.2024),
- patenata (<https://patents.google.com/>, pristupljeno 16.05.2024),
- knjiga (<https://books.google.com/>, pristupljeno 16.05.2024),
- podataka (<https://datasetsearch.research.google.com/>, pristupljeno 16.05.2024),
- trendova tj. pojmova koje drugi pretražuju (<https://trends.google.com/trends/?geo=RS>⁴⁴⁴, pristupljeno 16.05.2024.)
i
- pretraga pojmova u knjigama (<https://books.google.com/ngrams>, pristupljeno 16.05.2024.), čiji je rezultat već prikazan u ovom udžbeniku na Sl. 50.

Pored nabrojanih, postoji i niz drugih alata, koji mogu biti od koristi za pretragu kao, na primer, pretraga po slici (Gugl Lens, <https://lens.google/>, pristupljeno 16.05.2024).

Treći korak/metoda pretrage rešenja je direktan kontakt sa kreatorima biblioteka ili članovima R Zajednice. Najpoznatija konferencija sa temom R programskog jezika je useR!, ali postoje i druge⁴⁴⁵. U Srbiji je 2018. godine održana satRday⁴⁴⁶ konferencija, a od 2018. se godišnje održava PSSOH⁴⁴⁷ konferencija na Elektrotehničkom fakultetu, Univerziteta u Beogradu. Od 2009. godine postoji i časopis⁴⁴⁸ sa radovima, u otvorenom pristupu, koji je posvećen programiranju u R-u, a koji je osnovala i održava ga R Fondacija⁴⁴⁹.

Interaktivna diskusija na blogovima oko problema u programiranju, može da pomogne. Za uključenje u diskusije na blogovima, prvo i najvažnije pravilo je

443 Dobar primer korišćenja dostupnih alata za pretragu na internetu (obratiti pažnju na napredne opcije pretrage u Guglu) je data na sajtu biblioteke na Berkli Univerzitetu u Kaliforniji: <https://guides.lib.berkeley.edu/GoogleTips> (pristupljeno 17.05.2024).

444 Link odgovara lokaciji u Republici Srbiji, ali uz izmenu ključnih reči na kraju linka može se zameniti geografsko područje za koje se primenjuje pretraga trendova.

445 Lista konferencija koja se odnosi na programski jezik R je data na sajtu R projekta: <https://www.r-project.org/conferences/> (pristupljeno 17.05.2024).

446 <https://belgrade2018.satrdays.org/>, pristupljeno 17.05.2024.

447 <https://pssoh.etf.bg.ac.rs/>, pristupljeno 17.05.2024.

448 R Journal, <https://journal.r-project.org/>, pristupljeno 17.05.2024.

449 <https://www.r-project.org/foundation/>, pristupljeno 17.05.2024.

da se proveri da li je neko već postavio isto ili slično pitanje, što je obično istaknuto u obaveštenjima na blogu. Naravno, treba voditi računa o poštovanju sagovornika i držati se standardnih bonton pravila u komunikaciji. Takođe, treba sa rezervom uzeti odgovore koje nije moguće proveriti.

Pored slučajnih, tj. nenamernih grešaka koje su bile predmet ovog potpoglavlja, postoje i greške koje su učinjene svesno, najčešće sa ciljem da se postigne profesionalni ili lični uspeh u nauci, kao što su dobijanje finansiranja ili unapređenje u karijeri [580]. U obradi signala se odnosi na promenu podataka, sa ciljem da se dobiju željeni rezultati. Jedino rešenje za ovaj problem je promena načina razmišljanja, u kome bi naučni integritet zauzimao najvišu statusnu poziciju, dok bi materijalni motivi ili motivi ličnog napredovanja bili u drugom planu [580]. Pokrenuto je više inicijativa koje su dominantno usmerene ka naučnicima i akademskim krugovima da se transparentno i javno objave informacije o greškama⁴⁵⁰.

6.3 Generisanje automatskih izveštaja

U okviru R Studio razvojnog okruženja, moguće je otvoriti i sačuvati skripte sa različitim ekstenzijama. Do sada je bilo reči o .R ekstenzijama, ali je veoma korisna i .Rmd (od eng. R *Markdown*) ekstenzija, koja se koristi za istovremeno pisanje koda i generisanje automatskih izveštaja sa rezultatima. *Markdown*⁴⁵¹ je programski interfejs za editovanje teksta u jednostavnim tekstulanim okruženjima [582] i vrlo je sličan HTML-u. Pored uređivanja R dokumenata, koji kombinuju tekst i skripte, koristi ga GitHub, *Swift*⁴⁵² i drugi.

Ako se odabere kreiranje nove datoteke, koja je tipa R *Markdown* u R Studio okruženju, pojaviće se prozor za unos osnovnih metapodataka (naslov dokumenta, Autor, datum, kao i željeni format za izvoz) kao na Sl. 237. Kada se klikne na OK taster, otvoriće se generički dokument sa ekstenzijom .Rmd kao na Sl. 238 sa padajućim menijem Knit⁴⁵³. Ako se odabere opcija za generisanje

450 Jedna nekompletna lista pretencioznih grešaka u nauci je dostupna na Vikipediji, https://en.wikipedia.org/wiki/List_of_scientific_misconduct_incidents (pristupljeno 17.05.2024).

451 <https://en.wikipedia.org/wiki/Markdown>, pristupljeno 17.05.2024.

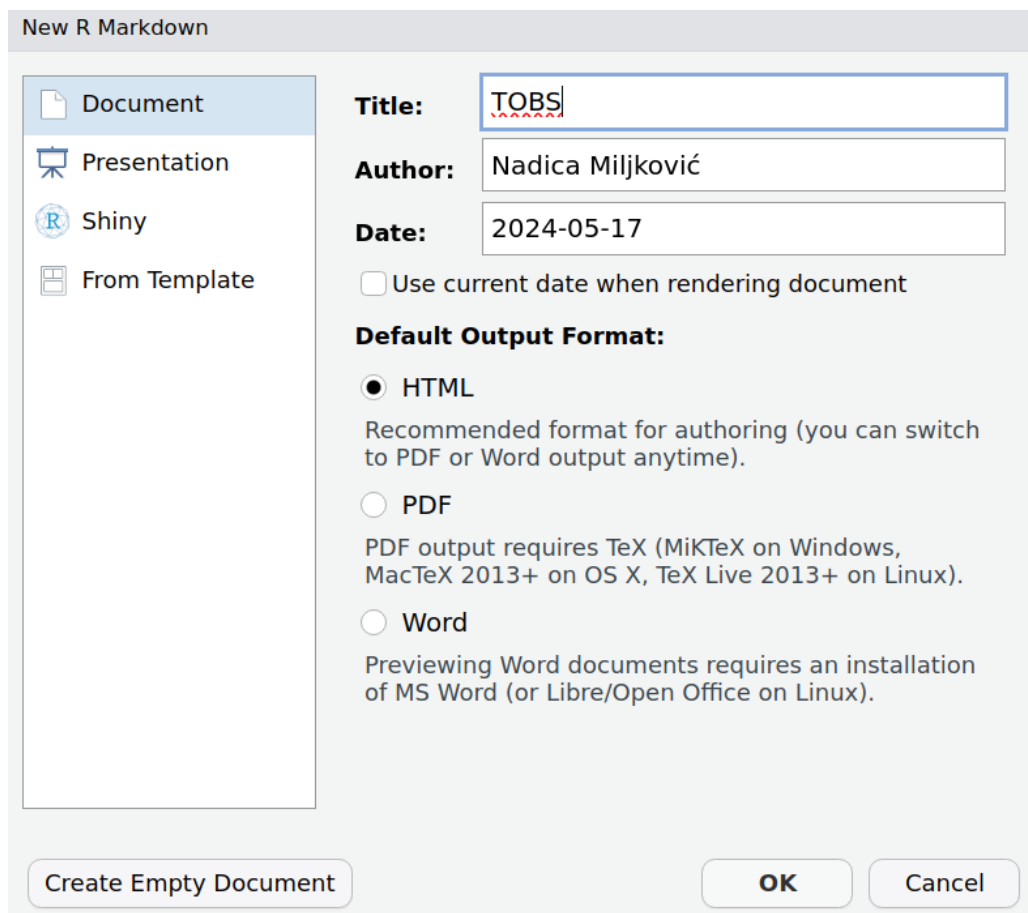
452 *Swift* je popularno programsko okruženje za kreiranje Epl (eng. *Apple*, <https://www.apple.com/>, pristupljeno 17.05.2024.) aplikacija, <https://www.swift.org/>, pristupljeno 17.05.2024.

453 U bukvalnom prevodu sa engleskog jezika, *knit* znači štrikanje/heklanje, odnosno u konektstu programskih okruženja: generisanje automatskog izveštaja.

izveštaja sa .pdf ekstenzijom, u R-u biti ispisana poruka, da je potrebno dodatno instalirati pokrenuti iz paketa tinytex funkciju [582-583]:

```
tinytex::install_tinytex()
```

koji omogućava da se funkcionalnost Lateha⁴⁵⁴ iskoristi u programskom jeziku R.

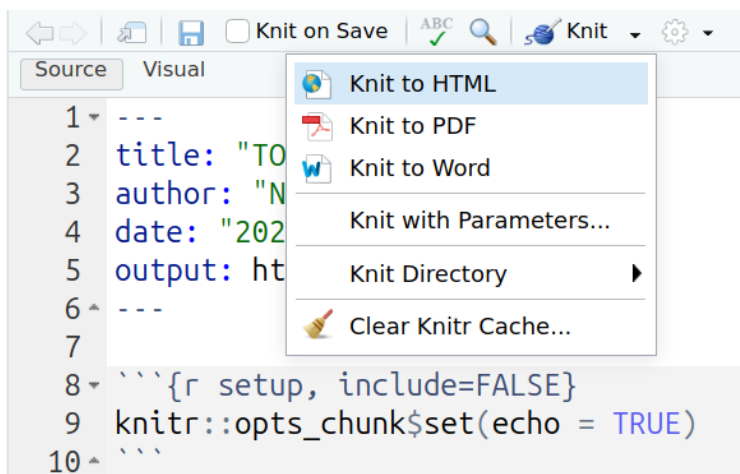


Slika 237, Kreiranje R *markdown* dokumenta za generisanje automatskog izveštaja.

Na Sl. 238 se vidi, da postoji zaglavlje (eng. *header*) dokumenta sa odgovarajućim metapodacima koje sadrži naslov, ime Autora, datum i podrazumevanu formu generisanog izveštaja. Ovaj deo *Markdown* skripte u R-u se naziva *YAML*⁴⁵⁵ i specifičan je za R programski jezik. Nakon zaglavlja, korisnik može pisati slobodan tekst, a kodovi se dodaju kao posebni

⁴⁵⁴ <https://www.latex-project.org/>, pristupljeno 17.05.2024.

elementi/delovi skripte (eng. *chunk*). Delovi skripte, koji sadrže izvršne elemente (eng. *executable chunks*), mogu sadržati različite opcije, koje se uglavnom odnose na to da li će u generisanom izveštaju biti prikazan R kod ili ne, kao i da li će biti prikazane poruke greške, upozorenja i obaveštenja i sl. Za ilustraciju generisanja izveštaja u R-u, iskorišćen je u nešto izmenjenom obliku primer sa 10. vežbe TOBS predmeta pod nazivom “Automatsko generisanje izveštaja”⁴⁵⁶.



Slika 238, Korišćenje padajućeg menija *Knit* u okviru *.Rmd* skripte u R-u.

Iskustvo je pokazalo da se R *Markdown* uči relativno brzo, kao i HTML iz dva razloga: (1) interfejs i komande su jednostavnije od onih u programskim jezicima i (2) postoji dobra podrška na internetu sa relativno velikim brojem primera. Pored R *Markdown* opcije, postoji mogućnost da se radi u R *Notebook* dokumentima, odnosno beležnicama. Osnovna razlika između R *Markdown* skripte i beležnice je što je unutar beležnice omogućeno interaktivno izvršavanje pojedinih elemenata, koji sadrže R kod, te je prema tome beležnica još poželjnije okruženje za rad u R-u od *Markdown* klasičnog dokumenta⁴⁵⁷. Za razliku od korisničkog interfejsa aplikacija za obradu biosignala, *Markdown* skripte se mogu pisati od samog početka kreiranja metoda za analizu signala. Na Sl. 239 i S. 240 je prikazana osnovna struktura *Markdown* skripti, sa označenim osnovnim elementima, dok je na Sl. 240 jasno uočljiva razlika

455 Slično kao GNU, YAML je rekurzivna skraćenica od eng. *YAML Ain't Markup Language* što znači da ovo nije klasična karakteristika *Markdown* skripti, kakve su, na primer, dostupne u GitHub-u (<https://github.com/hao203/rmarkdown-YAML>, pristupljeno 17.05.2024).

456 Spisak svih vežbi dat je na kraju udžbenika u Dodatku.

457 http://uc-r.github.io/r_notebook, pristupljeno 17.05.2024.

između teksta i elemenata koda, tj. *chunk*-a, koji su uobičajeno prikazani na osenčenoj pozadini. Takođe, svaki element koda se može testirati ili se njegov prikaz može smanjiti korišćenjem alata u gornjem desnom uglu *chunk*-a. Tekst može sadržati reči ili rečenice ispisane kurzivom ili podebljane (kao na Sl. 239), a može sadržati i formule za koje se koristi sličan unos kao u Latehu.

Još jedan *Chunk*, ima ih dva na slici.

Italic tj. kurziv

formula

Zaokružena je strelica za testiranje *Chunk*-a.

Hiperlink

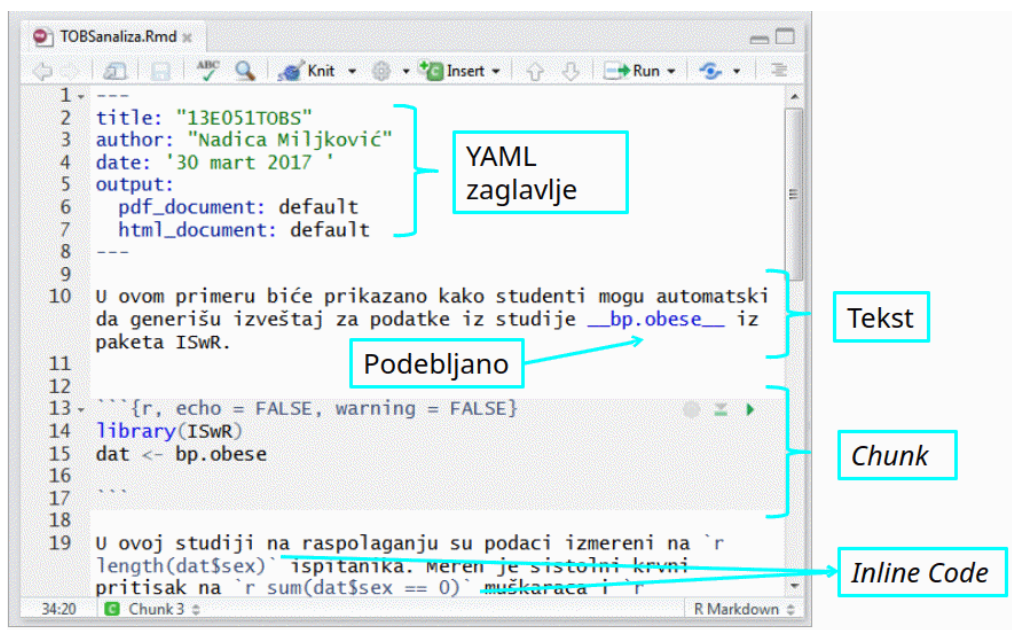
Slika 239, Označeni delovi R *Markdown* dokumenta.

Posebna pogodnost *Markdown* dokumenata je što postoji mogućnost da se kratke komande integrišu u R skript (eng. *Inline Code*), te se u finalnom izveštaju, mogu uneti brojne vrednosti u tekstu, koje su rezultat računa. Na primer, može se pisati koliko ima ispitanika i koliko je dug učitani vemenski interval biosignala. Ovo ne bi imalo puno smisla, da u *Markdown* skriptama nije omogućena interaktivnost, odnosno mogućnost odabira parametara na početku generisanja izveštaja. U najvećem broju slučajeva, to su nazivi datoteka koje je potrebno učitati u programski jezik. Postojanje promenljivih parametara se unosi u okviru *YAML* zaglavlja. Na primer, sledeće linije koda:

params:

filename: emg-kabl.txt

u okviru *YAML* dela, omogućavaju da se prilikom generisanja izveštaja, otvara poseban prozor, u kome se može odabrati datoteka po izboru (podrazumevana datoteka je "emg-kabl.txt").



Slika 240, YAML zaglavlje i osnovni elementi R *Markdown* dokumenta.

Markdown se, uz malo modifikacija i dodatne pakete, može koristiti za pisanje udžbenika, a za potrebe pisanja knjiga u R-u [135] razvijeni su specijalizovani paketi kao što je *bookdown* paket [584-585]. Postoje i drugi primeri korisnih alata za pisanje, odnosno komponovanje udžbenika, kao što je *Sfinks*⁴⁵⁸ generator dokumentacije, u kome je napisan odličan fakultetski udžbenik za programiranje u programskom jeziku Pajton [586-588].

Pored interaktivnih izveštaja, korišćenjem *Markdown* dokumenata i beležnica, u R-u se mogu kreirati specijalizovane veb aplikacije, korišćenjem *Shiny* paketa [465]. Prikaz jednostavne *Shiny* aplikacije, koja se koristi na vežbama za TOBS predmet je dat na Sl. 241. Struktura *Shiny* koda se sastoji iz dve osnovne funkcije:

ui()

koja je zadužena za izgled korisničkog interfejsa, kao i:

server()

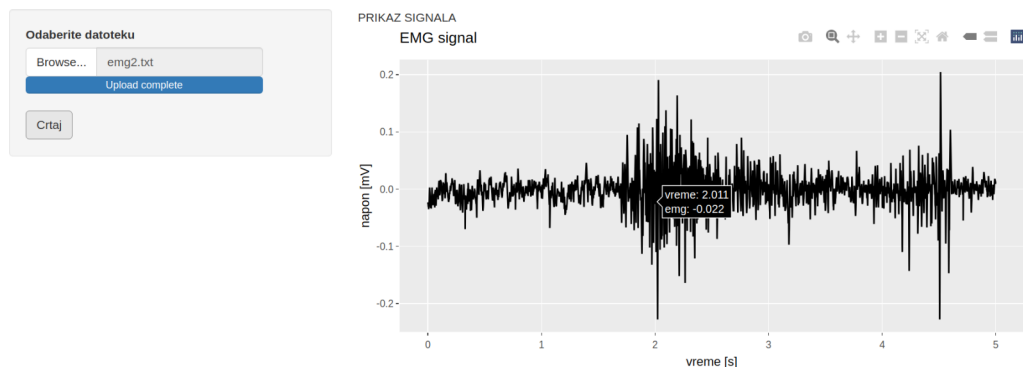
458 Shinx, Documentation generator, <https://www.sphinx-doc.org/en/master/>, pristupljeno 26.04.2024.

funkcije, koja omogućava predloženu funkcionalnost. Pokrenutu aplikaciju lokalno na računaru (kako je prikazano na Sl. 241) je moguće pokrenuti na nekom serveru, te bi se onda izveštaju, tj. aplikaciji, pristupalo preko interneta. Ovo je veoma zgodno za kompaktno deljenje rezultata obrade signala, kao i za funkcionalnost koda. Iako je *Shiny* veoma popularan u R-u, nije tako omiljen u programerskoj zajednici. Naime, jednostavne funkcije, iako omogućavaju brzo dolaženje do funkcionalne aplikacije su veoma visokog nivoa, te se mnoge operacije obavljaju u pozadini i iskusni programeri ne mogu tako jednostavno da manipulišu *Shiny* okruženjem. Međutim, *Shiny* ima minimalističku filozofiju tidyverse grupe paketa⁴⁵⁹, te se savetuje njegovo korišćenje za vizuelizaciju biomedicinskih signala.

Vezba 12

Tehnike obrade biomedicinskih signala

13E051TOBS - Tehnike obrade biomedicinskih signala
Univerzitet u Beogradu - Elektrotehnički fakultet



Slika 241, Primer jednostavne *Shiny* aplikacije za učitavanje datoteke po izboru i interaktivni prikaz biosignala u vremenskom primenom ggplot i plotly paketa.

6.4 Estetika softverskog koda

Nema striktnih pravila ni standarda koji se tiču estetike koda, jer je estetika lični izbor. Ipak, jednostavno je prepoznati kod koji prati dobre smernice, u odnosu na onaj koji je nečitak, nerazumljiv i nepregledan. Neke od smernica za pisanje pristojnog koda⁴⁶⁰ su [3]:

459 Na internetu postoji niz primera *Shiny* veb aplikacija sa veoma korisnim i dobrim vizuelizacijama, a neke od njih su dostupne na R Studio sajtu:

<https://www.rstudio.com/products/shiny/>, pristupljeno 17.05.2024.

460 Pristojni i nepristojni kodovi ne postoje, a ovde se misli na čitkost koda.

- potrebno je pisati kod u tekstualnom editoru, a nikako u konzoli⁴⁶¹ (u R-u je uobičajeno pokretanje koda liniju po liniju, istovremenim pritiskom Ctrl i Enter tastera sa tastature),
- korišćenje razmaka poboljšava reorganizaciju blokova koda (moguće je automatski podesiti razmake odabirom iz padajućeg menija *Tools* sledeće opcije *Global Options / Code / Editing*) – na Sl. 242 je prikazan deo koda za prikaz EMG signala sa razmacima i bez njih (čitljivost je приметно bolja u kodu u kome postoje razmaci),
- “širinu” koda je potrebno ograničiti, a preporuka je da se koristi oko 80 kolona (prikaz vertikalne linije u uređivaču, tj. editoru R Studio okruženja, koja ograničava širinu koda je prikazana na Sl. 243),
- promenljivama treba dati opisne nazive (na primer: *frekvencija_odabiranja*, *frekvencijaOdabiranja*, *fOd*, *fs* ili *fod*),
- potrebno je dobro dokumentovati kod, na primer, pisanjem komentara nakon znaka “#”,
- ne treba preterivati u dužini i broju komentara, ali ne treba izostaviti važne informacije za razumevanje softverskog koda (posebno bi trebalo voditi računa da “Manje je više”),
- ograničiti “dužinu” funkcija i koda na veličinu prozora (zavisi od monitora i veličine slova, tj. fonta koji se koristi),
- funkcije treba da imaju logičke nazive i uloge, itd.

Usvajanje standarda je proces, te je za primenu nekih pobrojanih praksi potrebna vežba. R Studio okruženje omogućava dodatne olakšice, kao što je to vertikalna linija u editoru (Sl. 243), koja ima funkciju “sveske u malim linijama” i koja služi da se nauči pisanje u osnovnoj školi.

Postoje četiri ustaljene konvencije za davanje imena promenljivima (funkcijama, programskim skriptama, klasama, interfejsima i metodama) u programskim jezicima [589]. U nastavku su izlistane konvencije sa primerom naziva za promenljivu, koja nosi informaciju o frekvenciji odabiranja:

- *Camel Case* – naziv se sastoji iz više reči koje su spojene, svaka nova reč ima veliko slovo, dok je prvo slovo malo, primer naziva je *frekvencijaOdabiranja* (ovo je ujedno najpopularniji način davanja imena na koji je Autorka nailazila u praksi),
- *Pascal Case* – je sličan kao Camel Case, sa razlikom što je prvo slovo veliko, kao na primer: *FrekvencijaOdabiranja*,

⁴⁶¹ Radi jednostavnosti i ilustracije izlaza u konzoli, pojedini primeri, u ovom udžbeniku, su ilustrovani u konzoli, ali gde god je to bilo moguće, sačuvan je redosled komandi u pratećoj skripti.

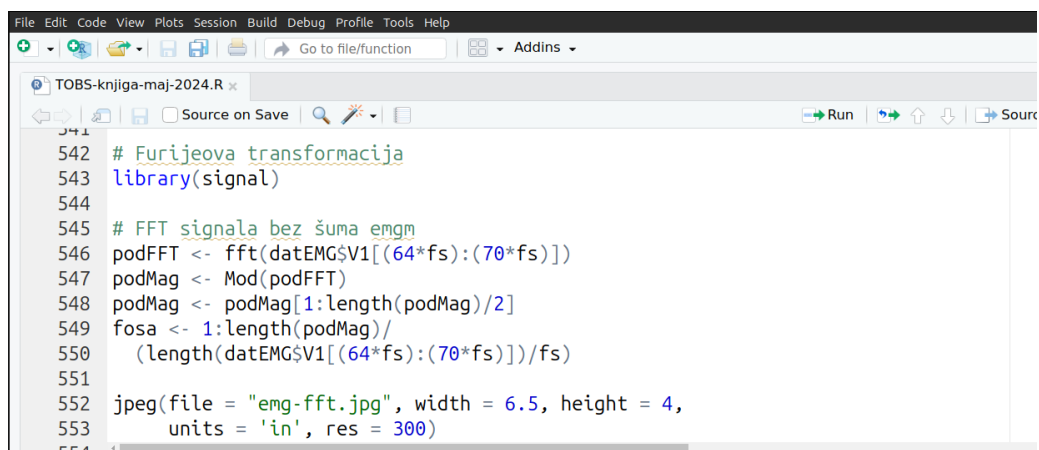
- *Snake Case* – koristi donje crte za razdvajanje reči, koje su sve napisane malim slovima, te je rezultujuće ime *frekvencija_odabiranja* (jedna je od zastupljenijih konvencija) i
- *Kebab Case* – je veoma sličan *Snake Case* konvenciji (uobičajen je kod programiranja u HTML-u, ali i kod definisanja URL), te bi za frekvenciju odabiranja promenljiva bila nazvana: *frekvencija-odabiranja*.

```
jpeg(file = "emg-fft.jpg", width = 6.5, height = 4,
      units = 'in', res = 300)
plot(fosa, podMag, type = "l", xlim = c(0, 500),
      xlab = "frekvencija [Hz]",
      ylab = "magnituda [a.u.]")
grid()
dev.off()
```

```
jpeg(file = "emg-fft.jpg", width = 6.5, height = 4,
      units = 'in', res = 300)
  plot(fosa, podMag, type = "l", xlim = c(0, 500),
        xlab = "frekvencija [Hz]",
        ylab = "magnituda [a.u.]")
  grid()
dev.off()
```

Slika 242, Deo koda u kome su iskorišćene osnovne R funkcije za prikaz grafika, kao i korišćenje *jpeg()* i *dev.off()* funkcija koje omogućavaju čuvanje grafika u .jpg formatu. Na gornjem panelu je prikazan grafik sa kodom u kome nisu korišćeni razmaci, dok donji panel prikazuje čitkiju varijantu, u kojoj su korišćeni razmaci, te ovaj deo koda zadovoljava estetske standarde.

Verovatno su se studenti, do sada, sreli sa nekom od konvencija, a sasvim je moguće da ih već koriste u praksi. U ovom udžbeniku, nije korišćen nijedan od prethodnih primera za naziv promenljive koja sadrži informaciju o frekvenciji odabiranja, zato što je uobičajeno da se u praksi ona označava sa skraćenicama *fs* ili *fod*, što je i dalje intuitivno jasno, a značajno kraće. Dobra praksa u davanju naziva promenljivim, jeste da se za promenljive logičkog tipa koristi naziv koji sadrži pitanje ili iskaz [590]. Na primer, ako je vrednost promenljive *ImaGresaka* tačna, onda greške postoje i obrnuto.



```
542 # Furijeova transformacija
543 library(signal)
544
545 # FFT signala bez šuma emgm
546 podFFT <- fft(datEMG$V1[(64*fs):(70*fs)])
547 podMag <- Mod(podFFT)
548 podMag <- podMag[1:length(podMag)/2]
549 fosa <- 1:length(podMag)/
550   (length(datEMG$V1[(64*fs):(70*fs)]/fs)
551
552 jpeg(file = "emg-fft.jpg", width = 6.5, height = 4,
553       units = 'in', res = 300)
```

Slika 243, Na desnom delu prozora se nalazi vertikalna linija svetlo sive boje, koja označava poželjnu “širinu” koda.

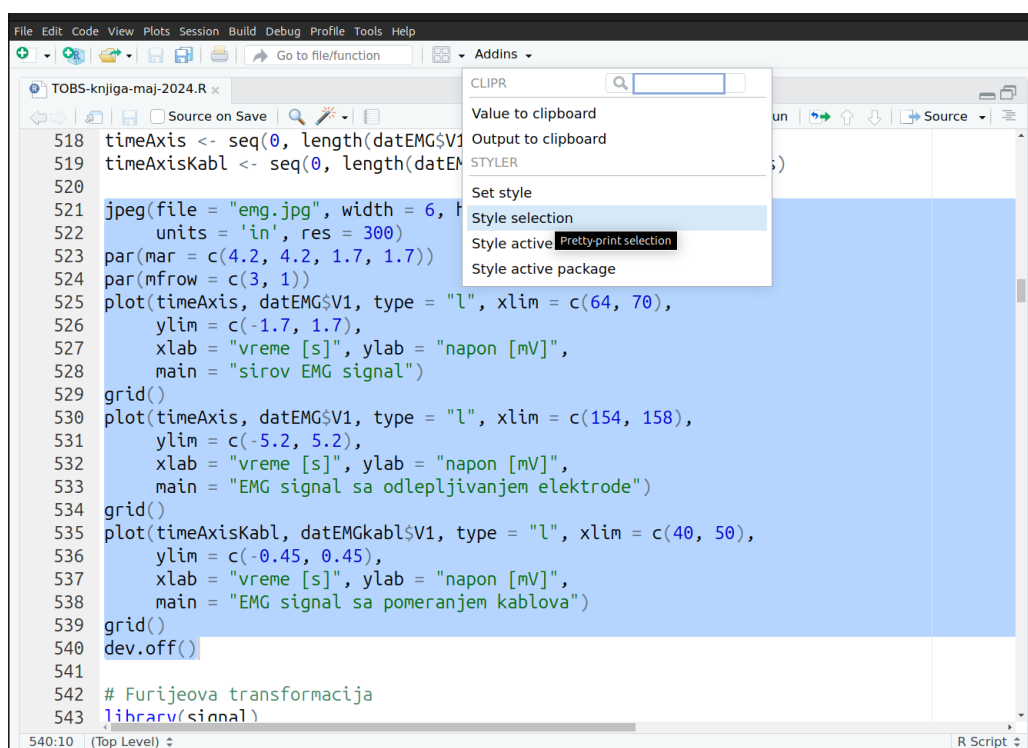
Pored konvencije za nazive promenljivih (eng. *naming conventions*⁴⁶²), postoje konvencije za pisanje skripti (eng. *coding conventions*⁴⁶³) i različiti stilovi programiranja (eng. *programming style*⁴⁶⁴). Odluka o tome koju konvenciju ili stil primeniti, trebalo bi da bude individualna, odnosno svako treba za sebe da odluči koju će konvenciju i stil preuzeti. Jedno je sigurno – treba izbegavati mešanje stilova i konvencija u individualnom pristupu. Za one koji su napisali svojih prvih par hiljada linija koda, verovatno to više nije upitno, a veoma jednostavno se mogu prepoznati različite osobe samo gledanjem koda (verovatno postoje karakteristike u ponašanju i sklonosti koje određuju individualni stil). Na primer, Autorka uvek koristi za dodelu vrednosti ili logičke uslove, razmak pre i posle znaka jednakaosti, a jedina situacija u kojoj to ne primenjuje je dodela vrednosti argumenata prilikom poziva funkcije za funkcije, koje imaju relativno veliki broj argumenata. Kod individualnog stila, veliku ulogu imaju navike. Kako je navike teško promeniti, u većim timovima, gde više osoba pojedinačno razvija kod, koji se kasnije integriše u jednu celinu, razvijeni su posebni alati za integraciju koda. U R-u postoje dva paketa koja se mogu koristiti za proveru i korekciju u stilu programiranja: (1) `lintr` [591] - ovaj paket proverava stil, ali može da pronalazi i sintaksne greške i (2) `styler` [592] - paket koji se koristi za formatiranje R koda (deo je `tidyverse`⁴⁶⁵ skupa paketa).

462 [https://en.wikipedia.org/wiki/Naming_convention_\(programming\)](https://en.wikipedia.org/wiki/Naming_convention_(programming)), pristupljeno 17.05.2024.

463 https://en.wikipedia.org/wiki/Coding_conventions, pristupljeno 17.05.2024.

464 https://en.wikipedia.org/wiki/Programming_style, pristupljeno 17.05.2024.

465 <https://www.tidyverse.org/blog/2017/12/styler-1.0.0/>, pristupljeno 17.05.2024.



Slika 244. Pokretanje opcije za stilizovanje označenog koda u R Studio okruženju primenom styler paketa [592].

Za ilustraciju primera, kako radi styler paket, iskorišćen je deo koda, podeljen uz ovaj udžbenik, u kome je izvršena izmena stila. Kada se instalira styler paket, otvara se u okviru R Studio razvojnog okruženja padajući meni *Addins*. Na Sl. 244 je prikazano kako se koristi *Addins* opcija (koja postaje aktivna nakon instalacije paketa) za stilizovanje odabranog dela koda iz padajuće menija, dok je na Sl. 245 prikazan rezultat stilizovanja koda.

6.5 Indeksiranje u R-u

Konvencije o davanju nazivima promenljivim, najčešće uključuju preporuku da jedno slovo može biti naziv promenljive, u slučaju da se radi o iteratoru, koji se koristi za indeksiranje. Međutim, i tu treba voditi računa. Na primer, u članku o čestim programskim greškama u programskom jeziku Matlab [136], skrenuta je pažnja na jednu lošu praksu, koja po Autorkinim saznanjima propagira kroz skoro sve udžbenike programiranja, a to je indeksiranje korišćenjem *i* i *j* naziva za iteratore. Naime, ne treba koristiti *i* i *j* kao nazive za indekse *for* petlje (ili druge nazive u opštem slučaju), jer se ove oznake koriste u većini programskih

jezika za definisanje kompleksnih brojeva. Kreatori R-a su statističari, a kreatori Matlab-a su automatičari, te možda iz tog razloga se u Matlabu koriste *i* i *j* za označavanje kompleksnih brojeva, a u R-u samo postoji kompleksno *i* (matematičari, za razliku od inženjera, ne koriste kompleksno *j*). Bez obzira što se ne koristi kompleksno *j* u R-u, ova dva slova bi trebalo izbegavati prilikom indeksiranja petlje. Ovo je velika promena za nekoga ko od srednje, a sada i osnovne škole u Srbiji, uči programiranje i ko je napisao stotine *for* petlji sa pogrešnim indeksima (Autorka pretpostavlja da je to slučaj sa svim studentima), ali vredi pomučiti se da se ta navika promeni. Greške koje mogu da nastanu, u takvoj situaciji, nije jednostavno otkriti, jer nijedan programski jezik, za sada, ne prepoznaje automatski takav način indeksiranja kao pogrešan. Preporučuje se da se koriste druge oznake za indeksiranje unutar petlje, kao na primer, *ind*, *ix*, *id* ili *k*.

```

518 timeAxis <- seq(0, length(datEMG$V1)/fs - 1/fs, by = 1/fs)
519 timeAxisKabl <- seq(0, length(datEMGkabl$V1)/fs - 1/fs, by = 1/fs)
520
521 jpeg(
522   file = "emg.jpg", width = 6, height = 7,
523   units = "in", res = 300
524 )
525 par(mar = c(4.2, 4.2, 1.7, 1.7))
526 par(mfrow = c(3, 1))
527 plot(timeAxis, datEMG$V1,
528   type = "l", xlim = c(64, 70),
529   ylim = c(-1.7, 1.7),
530   xlab = "vreme [s]", ylab = "napon [mV]",
531   main = "sirov EMG signal"
532 )
533 grid()
534 plot(timeAxis, datEMG$V1,
535   type = "l", xlim = c(154, 158),
536   ylim = c(-5.2, 5.2),
537   xlab = "vreme [s]", ylab = "napon [mV]",
538   main = "EMG signal sa odlepljivanjem elektrode"
539 )
540 grid()
541 plot(timeAxisKabl, datEMGkabl$V1,
542   type = "l", xlim = c(40, 50),
543   ylim = c(-0.45, 0.45)

```

Slika 245, Stilizovan deo koda sa Sl. 244, za podrazumevani stil u okviru styler paketa [592].

Podrazumevani osnovni indeks (eng. *default base index*) u R-u (kao i u Matlabu i GNU Octave-u) je 1. Sasvim slučajno, u industriji su češće u upotrebi programski jezici, kod kojih je osnovni indeks jednak 0 i njih ima više (na primer, C i Pajton), pa treba biti spreman na ovu promenu prilikom korišćenja

novih programskih jezika⁴⁶⁶. Posebno, treba obratiti pažnju prilikom prelaska u programski jezik R, iz drugih programskih jezika, jer R neće javiti grešku, ako se niz indeksira sa 0. Neće javiti grešku, ni ako je indeks negativan, te uvek treba proveriti da li je indeksiranje realizovano na odgovarajući način i bez greške, koja se ne može dijagnostifikovati automatski.

⁴⁶⁶ Iscrpna lista programskih jezika i podrazumevanih osnovnih indeksa koji se koriste je dostupna na Vikipedija stranici “Comparison of programming languages: Cross-reference list”, https://en.wikipedia.org/wiki/Comparison_of_programming_languages_%28array%29#Array_system_cross-reference_list (pristupljeno 17.05.2024).

7 Umesto zaključka

Živopisna diskusija na internetu⁴⁶⁷, koja samo povremeno jenjava, na temu da li koristiti i učiti programski jezik R ili Pajton, za bavljenje naukom o podacima, nije od posebnog značaja ni za analizu biosignala, ni za korišćenje ova dva programska jezika. Od nedavno, postoje i pitanja da li je bolje koristiti Julia⁴⁶⁸ [29, 594] programski jezik ili R. U srpskoj akademskoj zajednici primerenije je upitati da li koristiti R ili Matlab koji je trenutno verovatno najzastupljeniji na našim univerzitetima.

Nema pravog odgovora na ova pitanja i neće ih ni biti. Ako se o programskim jezicima toliko govori, znači da ih ljudi intenzivno koriste, što je jedno od glavnih merila prilikom izbora programskog jezika [3]. Primereno pitanje jeste da li se opredeliti za softver koji ima slobodnu licencu ili za softver koji je vlasnički. Prirodno se nameće odgovor, da slobodan softver, koji nema nikakvih ograničenja, čiji je izvorni kod (eng. *source code*) dostupan i koji nema za cilj da stvori zavisnost kod korisnika (eng. *vendor lock-in*), treba koristiti u akademskom okruženju. Sva ostala pitanja su stvar navike, ukusa ili politike u radnoj ili obrazovnoj sredini. Postoji čak i knjiga namenjena iskusnim korisnicima Matlaba da jednostavnije savladaju R [61] i obrnuto. Sigurno da, zavisno učenje Matlaba i R-a nije idealan pristup, jer podrazumeva podilaženje korisničkim navikama, što posebno može biti problematično u akademskom okruženju, gde je cilj da studenti nauče osnove analize biosignala i programiranja, kao i da steknu sposobnost da usvajaju nova znanja, a ne da se čvrsto drže usvojenih rutina. S toga, ovaj udžbenik iako obiluje primerima u programskom jeziku R, treba najviše da posluži, da se usvoje dobre prakse u programiranju, sa ciljem realizacije metoda analize biosignala, kako bi se kasnije te prakse iskoristile u bilo kom programskom jeziku. Autorkin stav po pitanju programskih jezika je donekle neutralan, jer je njena preporuka da koristi ko šta želi, s tim da prednost, posebno u akademskom okruženju, treba dati slobodnom softveru [3, 595]. Na kraju, poređenje programskih jezika je bespredmetno. Poznavanje većeg broja programskih jezika, čini osobu

467 Primeri članaka na temu da li treba učiti R ili Pajton koji se mogu naći na internetu su: <https://www.datacamp.com/blog/python-vs-r-for-data-science-whats-the-difference>, <https://www.ibm.com/blog/python-vs-r/>, <https://bioinformatics.ccr.cancer.gov/btep/r-or-python-which-should-i-learn/> i <https://www.coursera.org/articles/python-or-r-for-data-analysis>, pristupljeno svim linkovima 29.01.2024.

468 Julia je programski jezik koji je nastao 2012. godine. Po sintaksi je vrlo sličan Pajtonu, ali značajno brži (tu je više nalik na C i C++ programske jezike). Odgovori na pitanja kako i zašto je nastao Julia programski jezik se nalaze u članku [593].

bogatijom za dodatna znanja i iskustva. Veći problem od izbora programskog jezika, u analizi biosignala, predstavlja odabir pogodnih metoda za obradu signala. Ne postoje dobre, loše, prave ni pogrešne metode za analizu biosignala, statističku ocenu parametara ili primenu metoda veštačke inteligencije. Kada bi se mogao sastaviti jednostavan algoritam, po kome bi bilo moguće odabrati metodu za neki konkretan slučaj, onda biomedicinski inženjeri ne bi ni bili neophodni. Tehničari ili čak računari, bi mogli samostalno da obavljaju zadatke analize, kako bi došli do informacije od interesa u biosignalu. Svaka metoda poseduje mane i prednosti, a od pristupa problemu, kreativnosti, posvećenosti i od iskustva biomedicinskog inženjera zavisi koliko će odgovarajuća metoda biti uspešna. Ovaj udžbenik nije imao za cilj da prikaže iscrpan skup metoda koje se primenjuju u praksi, već da ukaže na način razmišljanja i praktični značaj analize biosignala kroz odabrane metode i njihovu implementaciju u programskom jeziku R. Dodatno, R je prikazan u ovom udžbeniku, jer poseduje osobine koje njegovu primenu u analizi biosignala čine pogodnijom, a konkretni razlozi su: (1) R poseduje odličan kvalitet grafika i dobar izbor metoda za vizuelizaciju, (2) postoji aktivna zajednica i veliki izbor materijala na internetu, (3) česte su novije verzije softvera, što je inače odlika slobodnog softvera [3], (4) R omogućava tzv. *fast prototyping* tj. maksimalan rezultat uz minimalan broj linija koda i (5) R jeste, bar za sada, *lingua franca*⁴⁶⁹ među programskim jezicima [3].

7.1 Veštačka inteligencija kao pomoć u programiranju

Nemoguće je ostati imun na nove alate, koji su na dohvat ruke i čija upotreba je postala sve učestalija, ma koliko oni kontraverzni bili. ChatGPT (od eng. *Chat Generative Pre-trained Transformer*) je softverska aplikacija koja je razvijena sa ciljem da oponaša ljudsku konverzaciju (eng. *chat bot*), pomoću metoda za mašinsko generisanje teksta, a razvila ju je kompanija OpenAI⁴⁷⁰ (San Francisko, Kalifornija, SAD). Alati zasnovani na računarskim algoritmima, koji

469 Sa latinskog u slobodnom prevodu, *lingua franca* znači da je to zajednički jezik ljudima koji govore različite jezike. Detalje pogledati, u Oksfordskom rečniku: <https://www.oxfordlearnersdictionaries.com/definition/english/lingua-franca?q=lingua+franca>, pristupljeno 29.01.2024.

470 <https://openai.com/>, pristupljeno 29.01.2024.

koriste prilagodljivu logiku⁴⁷¹ i koji učestvuju u konverzaciji su tu da ostanu, kakva god sudbina da čeka ChatGPT aplikaciju⁴⁷².

Jednom kada je sadržaj na internetu postao naširoko dostupan i kada je Vikipedija projekat poprimio današnje razmere, niko se više i ne seća da je ranije jedini način da se dođe do brzih odgovora⁴⁷³ na pitanja, bilo korišćenje rečnika i enciklopedija, u najboljem slučaju u kućnoj, a češće u komšijskoj, gradskoj ili školskoj (fakultetskoj) biblioteci. Brz pristup informacijama i kombinovanje dostupnih informacija, na ljudima razumljiv način su dobre strane konverzijskih računarskih alata. Međutim, postoje i loše strane. Mašinski algoritmi nemaju mogućnost kritičkog mišljenja, te ako postoje oprečni zaključci u naučnim radovima, takvi algoritmi će zauzeti jednu stranu, navesti nepostojeće izvore ili ih ne navesti u potpunosti, što dalje može da vodi najvećem akademskom i etičkom prekršaju – plagijatu. Odlični primeri zašto treba pravilno koristiti ChatGPT alat u pisanju naučnog rada su prikazani u video uputstvu Karen Meki [597].

Primer pogrešne upotrebe ChatGPT-ja je zabeležen 2023. godine u Njujorku, kada je dva advokata kaznio sudija jer su koristili ChatGPT kao pomoć u pripremi materijala za suđenje, bez dodatne provere informacija koje su dobili. Tako su advokati, na sudu, koristili nepostojeće sudske slučajeve, koje je ChatGPT izmislio. U presudi advokatima je pisalo da nema ničeg lošeg u korišćenju alata kao što je ChatGPT, ali da oni koji saopštavaju informacije imaju odgovornost za tačnost tih informacija⁴⁷⁴. Zaključak sudije iz ovog slučaja, može se primeniti i u akademiji, jer se nikome ne može zabraniti korišćenje ChatGPT-ja i sličnih aplikacija, ali svako ima odgovornost za informacije koje prenosi, posebno u diplomskim radovima, master radovima, naučnim radovima i drugim pisanim materijalima. Sa razvojem softvera za

471 Ovakvi računarski algoritmi se u literaturi nazivaju veštačka inteligencija ili AI od eng. *Artificial Intelligence*.

472 U decembru 2023. godine, Njujork Tajms je tužio OpenAI i Majkrosoft uz tvrdnju da su na neodgovarajući koristili materijale/članke Njujork Tajmsa, <https://www.theguardian.com/media/2023/dec/27/new-york-times-openai-microsoft-lawsuit> (pristupljeno 31.01.2024). U januaru 2024. godine, OpenAI je odgovorio na ovu tužbu negirajući navode, a u vreme pisanja ovog udžbenika dalji tok ovoj slučaju još uvek nije poznat, <https://www.cnbc.com/2024/01/08/openai-responds-to-new-york-times-lawsuit.html> (pristupljeno 31.01.2024).

473 Pored generisanja brzih odgovora, postoje aplikacije koje omogućavaju brzo sumiranje rezultata pretrage ili analize teksta, kao što je Elicit (Ought, SAD), <https://elicit.com/>, pristupljeno 31.01.2024. [596]

474 <https://www.theguardian.com/technology/2023/jun/23/two-us-lawyers-fined-submitting-fake-court-citations-chatgpt>, pristupljeno 31.01.2024.

generisanje teksta koji oponaša ljudski govor, razvili su se i alati za prepoznavanje mašinski generisanog teksta, kao što je GPTZero⁴⁷⁵ koji je nastao na Princeton Univerzitetu⁴⁷⁶ sa osnovnim ciljem da detektuje plagijarizam⁴⁷⁷.

Iako naziv kompanije OpenAI, koja je proizvela ChatGPT, ukazuje na otvorenost (eng. *Open*), ova aplikacija nije otvorena, jer osim dela softverskog koda, kompanija nije podelila skup podataka na kome je algoritam obučavan, kod za obučavanje i hiperparametre⁴⁷⁸ modela [50, 599]. Zato je jedini dostupan način, da se prouči kako ChatGPT, radi kroz testiranje. To se može postići postavljanjem pitanja, a na koja su odgovori poznati, te proučavanjem kakve (tačne, netačne, potpune, nepotpune) odgovore na svom izlazu daje ChatGPT. Testiranje algoritama mašinskog učenja, primenom ulaznih veličina kojima se namenski manipuliše (na primer: generisane ili izmenjene slike) se koristi u oblasti razumljive veštačke inteligencije, koja se bavi objašnjenjem rada algoritama (eng. *Explainable Artificial Intelligence*, skraćeno XAI) [600]. Kako XAI nije uvek moguće postići⁴⁷⁹, nekada je jedini dostupan način, za postizanje razumljivosti algoritama veštačke inteligencije, da se primeni namensko testiranje. Tehnička kompleksnost, čini XAI metode teško razumljivim, ali postoje specijalizovani alati, koji mogu povećati razumljivost algoritama veštačke inteligencije.

Alati za generisanje teksta su se pokazali korisnim, u procesu parafraziranja teksta i razvoju jezičkih veština na engleskom jeziku, te su posebno važni Autorima, čiji prvi jezik nije engleski jezik. Veliki broj izdavača, a među njima i Elsevier⁴⁸⁰, imaju objavljena vrlo detaljna uputstva, za koje svrhe se mogu (na primer: da se poveća čitkost teksta), a za koje se ne mogu (na primer: formiranje zaključaka na osnovu pregleda literature ili dobijenih rezultata) koristiti alati zasnovani na veštačkoj inteligenciji⁴⁸¹. Dodatno, većina izdavača

475 <https://gptzero.me/>, pristupljeno 31.01.2024.

476 Princeton University, <https://www.princeton.edu/>, pristupljeno 31.01.2024.

477 <https://www.businessinsider.com/ai-gptzero-startup-founder-building-leading-text-detector-2023-7>, pristupljeno 31.01.2024.

478 Postoje dve vrste parametara u modelima mašinskog učenja: (1) parametri koji se postavljaju na osnovu podataka u procesu obučavanja (slično kao što se određuju parametri modela kod linearne regresije) i (2) parametri koje podešava programer pre početka procedure obučavanja – ovi parametri se nazivaju hiperparametri [371, 598].

479 Za one kojima oblast mašinskog učenja nije dovoljno bliska – poznat je način realizacije algoritama veštačke inteligencije, ali nije tako jednostavno objasniti u kakvoj je vezi izlaz algoritma sa ulaznim parametrom ili parametrima, kao što je to jednostavno prikazati kod linearnih modela.

480 Elsevier, <https://www.elsevier.com/>, pristupljeno 31.01.2024.

481 <https://www.elsevier.com/about/policies-and-standards/the-use-of-generative-ai-and-ai-assisted-technologies-in-writing-for-elsevier>, pristupljeno 31.01.2024.

insistira na transparentnosti u procesu pisanja rada, pa su Autori u obavezi, da napišu izjavu, u okviru teksta naučnog ili stručnog rada, sa detaljnim objašnjenjem na koji način su ChatGPT ili slični alati, korišćeni u procesu pisanja rada⁴⁸². Iako su i izdavači naučnih časopisa, uočili prednosti ovih alata, to ne znači da alatima treba slepo verovati, jer postoji opasnost od korišćenja sumnjivih i nerazumljivih fraza [601]⁴⁸³.

Pored generisanja teksta, alati veštačke inteligencije se mogu koristiti i za pisanje/generisanje programskog koda. Dobre i loše strane ChatGPT-ja, za detekciju grešaka u programskom kodu, pomoć u otklanjanju grešaka (debugovanje), predikciju grešaka i objašnjenje grešaka su prikazali Suramiri i Šakor u preglednom članku 2023. godine [602]. ChatGPT se može koristiti i za generisanje delova koda, ali i za pripremu odgovora na tehnička pitanja i automatsko generisanje dokumenata, kako je Bisvas pokazao na primeru programiranja pomoću alata veštačke inteligencije u programskom jeziku Pajton [603]. Za programiranje, u programskom jeziku R, postoji više paketa koji su zasnovani na ChatGPT-ju [604]. Na primer, `air` paket [605] se može koristiti prilikom programiranja u R-u. Funkcija:

`howto()`

omogućava da se generiše sekvenca koda, a pomoću:

`whatis()`

funkcije je moguće dobiti opis, u prirodnom jeziku, sa detaljnim objašnjenjem funkcije koda. Za korišćenje `air` paketa, potrebno je imati registrovan API⁴⁸⁴ (eng. *Application Programming Interface*) na OpenAI sajtu⁴⁸⁵.

482 Primer takve izjave, dat je u radu [371].

483 Na preporuci za ovu referencu, Autorka je zahvalna Milici Ševkušić, višoj diplomiranoj bibliotekarki iz Instituta tehničkih nauka Srpske akademije nauka i umetnosti (SANU). Takođe, Autorka je zahvalna i na ukazanoj praktičnoj upotrebljivosti i opasnostima koje su povezane sa korišćenjem alata veštačke inteligencije, u okviru održanog neformalnog kursa iz oblasti Tehničkih i etičkih aspekta prilikom pisanja naučnog rada, koji je održan 2023. godine i čijom je organizacijom rukovodila Milica Ševkušić.

484 Iako je na srpskom jeziku, skraćenica API uobičajena, ovde je detaljnije objašnjena. API je aplikacija koja kreira interfejs između dva ili više programska koda/jezika/uređaja. Na primer, za komunikaciju između sata koji meri puls (tzv. pametan sat, na eng. *smart watch*) i aplikacije na mobilnom telefonu, potreban je API. Dobra praksa (u duhu slobodnog softvera) je da se API deli otvoreno i javno, ali na žalost, u velikom broju slučajeva, to je usluga, odnosno proizvod koji se naplaćuje.

485 <https://github.com/soumyaray/air>, pristupljeno 31.01.2024.

Osim za pomoć u programiranju, Chat GPT se može koristiti i za asistenciju, za prevođenje softverskog koda iz jednog u drugi programski jezik. Postoji specijalizovana *Shiny* aplikacija CodeLingo⁴⁸⁶, koji je razvila kompanija Analytica⁴⁸⁷ (Vašington, SAD), u kojoj je moguće odabrati verziju Chat GPT-ja koja se koristi za prevođenje koda, moguće je automatski detektovati programski jezik u kome je prekopiran ili postavljen (eng. *upload*) kod napisan, a trenutno je omogućeno prevođenje za veći broj programskih jezika (R, Pajton, Skala, C++, PHP, Julia, Matlab i dr). Iako je ovo vrlo pogodna funkcionalnost, treba imati na umu da neke opcije, koje je moguće koristiti u jednom programskom jeziku, ne postoje u drugom, te automatsko prevođenje može da rezultuje isključivo dobrom polaznom tačkom i preporukom, a kompletno prevođenje mora sprovesti čovek, nikako mašina.

7.2 Budući pravci u analizi biosignala

Oblasti ne postoje i ne razvijaju se nezavisno, odnosno međusobno su povezane. Da bi se predvideo pravac u jednoj oblasti, potrebno je sagledati sve oblasti zajedno, a to nije moguće. Pitanje da li je i potrebno.

Obrada biosignala i biomedicinsko inženjerstvo, u celini, su potrebni društvu, ali ne treba očekivati da će se na svakom koraku pojaviti prilika za rad u oblasti, niti su poslodavci uvek svesni da bi posao u njihovoj firmi mogli odlično obavljati upravo biomedicinski inženjeri. Obrada signala je široka oblast, a ko ume da analizira biosignale, verovatno će se snaći i sa drugim manje ili više srodnim signalima. U Tabeli 2 je prikazan rezultat jednog eksperimenta pretrage po ključnim rečima, koji je predstavljen studentima na predmetu Metode analize elektrofizioloških signala⁴⁸⁸ na Elektrotehničkom fakultetu, Univerziteta u Beogradu. Rezultat pretrage je pokazao da se broj radova, koji sadrže ključne reči obrada biosignala, skoro izjednačio sa brojem onih koji sadrže termin biomedicinsko inženjerstvo, dok je odnos između oblasti obrade signala skoro duplo veći od oblasti obrade biosignala.

Rezultati prikazani u Tabeli 2 ukazuju na to da oblast obrade biosignala čini skoro polovinu od ukupne oblasti obrade signala. Iako su rezultati predstavljeni u Tabeli 2, samo procena i moraju se uzeti sa rezervom, jer ne predstavljaju

486 <https://analytica.shinyapps.io/codelingo/>, pristupljeno 02.02.2024.

487 <https://analyticadss.com/>, pristupljeno 02.02.2024.

488 Šifra predmeta je 13E054MAS, a internet stranica predmeta -

<https://automatika.etf.bg.ac.rs/sr/13e054mas> i karton predmeta

https://www.etf.bg.ac.rs/fis/karton_predmeta/13E054MAS-2013, linkovima je pristupljeno 31.01.2024.

iscrpnu analizu svih mogućih ključnih reči, ipak se ne mogu u potpunosti ignorisati. Obrada biosignala je kompleksnija od obrade signala, zato što su biosignali: (1) malih amplituda (nekada i niskih frekvencija, kao EGG signal), te su podložni relativno velikom uticaju šuma, koji nije uvek tako jednostavno ukloniti i (2) rezultat su rada kompleksnog živog sistema.

Tabela 2, Broj rezultata pretrage po ključnim rečima u Google Scholar alatu⁴⁸⁹. Rezultati prikazuju okviran broj pronađenih dokumenata za ključne reči “*biomedical processing*”, “*biomedical engineering*” i “*signal processing*”

[https://automatika.etf.bg.ac.rs/images/FAJLOVI_srpski/predmeti/izborni_kursevi_os/biomedicinsko_inzenjerstvo/OS4_OF4_MS1_MAS/2023%20nastava/01MAS23UvodniCas.pdf, pristupljeno 31.01.2024.]

od školske 2016/2017. do 2023/2024. godine.

školska godina	ključne reči		
	<i>biomedical signal processing</i>	<i>biomedical engineering</i>	<i>signal processing</i>
2016/2017	2 250 000	2 800 000	4 340 000
2017/2018	2 430 000	3 040 000	4 880 000
2018/2019	2 350 000	2 770 000	5 220 000
2019/2020	2 510 000	2 940 000	5 410 000
2020/2021	2 430 000	3 090 000	5 460 000
2021/2022	2 650 000	3 270 000	5 550 000
2022/2023	2 820 000	3 460 000	6 100 000
2023/2024	3 450 000	3 460 000	6 890 000

U udžbeniku je kroz više primera pokazano koliko su znanje o biosignalu koji se analizira (za šta je, na primer, potrebno predznanje o mernoj metodi, o izvoru signala i šuma) i poznavanje dobrih praksi u programiranju (kao na primer, računarska reproducibilnost) usko povezani sa rezultatom analize. Da li je ovaj udžbenik pokrio kompletnu oblast obrade biosignala? Ne, ni približno. A, da li je udžbenik pokrio sve izazove sa kojima se mogu sresti biomedicinski inženjeri u svojoj karijeri prilikom obrade biosignala? Nije ni to. Generacije, koje ovih dana odlaze u penziju, upisivale su Fakultet u vreme kada je oblast digitalne obrade signala bila tek u povoju, kada ni internet nije postojao, a kamoli pristup raznovrsnim biomedicinskim signalima. Literatura se nabavljala poštom, bila je skupa i do nje se teško i sporo dolazilo. Ti inženjeri su, tokom svoje karijere (van studentskih klupa), savladali rad na računarima i usvojili niz novih veština. Nijedan udžbenik, kurs, a ni Fakultet, ne mogu da pokriju sve izazove i probleme u jednoj oblasti, ali Autorka se nada da su studenti, uz pomoć ovog udžbenika i gradiva na TOBS predmetu, u stanju da:

⁴⁸⁹ <https://scholar.google.com/>, pristupljeno 31.01.2024.

- primene naučeno u oblasti analize signala (ili analitike) generalno,
- usvoje znanja o programskom jeziku R i njegovoj primeni u obradi biosignala, ali i način na koji mogu da pristupe usvajanju drugih alata,
- izvuku pouke iz prikazanih primera analize i vizuelizacije biosignala,
- uspeju da samostalno primene tehnike koje su prikazane u udžbeniku, ali i da na osnovu prikazanih tehnika, steknu veštine da usvoje i razumeju nove metode, sa kojima će se susresti u budućnosti,
- da su razvili mogućnost kritičkog razmišljanja i naučnog preispitivanja, prilikom odabira metoda za obradu biosignala ili kreiranja novih metoda i
- da su usvojili dobre prakse i spremnost da se uhvate u koštac sa novim praksama i problemima, pred kojima će se naći u budućnosti, prilikom obrade biosignala.

Razvoj novih ili unapređenje postojećih metoda za merenje biosignala, za numerička proračunavanje, kao i pojava novih primena u biomedicinskom inženjerstvu značajno utiču na razvoj oblasti obrade biosignala. A, za odgovor na pitanje kakva će budućnost oblasti obrade biosignala biti, znaće se tek sutra.

Literatura

- [1] Chonacky, N., & Winch, D. (2005). Reviews of maple, mathematica, and matlab: Coming soon to a publication near you, *Computing in Science & Engineering*, 7(2), 9-10. <https://doi.org/10.1109/MCSE.2005.39>
- [2] Ihaka, R., & Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5(3), 299-314. <https://doi.org/10.1080/10618600.1996.10474713>
- [3] Peng, R. D. (2016). *R programming for data science* (pp. 86-181). Victoria, BC, Canada: Leanpub.
- [4] R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing. <https://www.R-project.org/>, pristupljeno 27.05.2024.
- [5] Giorgi, F. M., Ceraolo, C., & Mercatelli, D. (2022). The R language: An engine for bioinformatics and data science. *Life*, 12(5), 648. <https://doi.org/10.3390/life12050648>
- [6] Ihaka, R. (2017). The r project: A brief history and thoughts about the future. *Univ. Auckl*, 4, 22.
- [7] Ihaka, R. (1998). R: Past and future history. *Computing Science and Statistics*, 392396.
- [8] Ihaka, R. (1997). New R Version for Unix, *stat.ethz.ch*, <https://stat.ethz.ch/pipermail/r-announce/1997/000014.html>, pristupljeno 27.05.2024.
- [9] Marić, M., Stojanović, S., et al. (2005). Sintaksa i semantika programskih jezika, Obrazovni softver, Univerzitet u Beogradu – Matematički fakultet, <https://edusoft.matf.bg.ac.rs/csharp/Verzija2005/racunar4.html>, pristupljeno 27.05.2024.
- [10] Google Research Made Easy: Advanced Searching in Google, Berkeley Library, University of California, <https://guides.lib.berkeley.edu/GoogleTips>, pristupljeno 28.07.2023.
- [11] Find Books, Articles and More: Use Google for Research, Sheridan Libraries, Johns Hopkins University, <https://guides.library.jhu.edu/find/google>, pristupljeno 28.07.2023.
- [12] Stallman, R. (2021). Linux and the GNU system, Free Software Foundation, <https://www.gnu.org/gnu/linux-and-gnu.html>, pristupljeno 28.07.2023.
- [13] Stallman, R. (2002). Free software, free society: Selected essays of Richard M. Stallman. GNU Press, <https://www.gnu.org/philosophy/fsfs/rms-essays.pdf>, pristupljeno 29.07.2023.
- [14] Pejović, P., Miljković, N., Cvetanović, M., & Ševkušić, M. (2021). Licence slobodnog softvera i otvorenog hardvera-kratko uputstvo za nestrpljive. *Zbornik treće nacionalne konferencije sa međunarodnim učešćem pod nazivom Primena slobodnog softvera i otvorenog hardvera PSSOH 2020*, 7-14. <https://doi.org/10.5281/zenodo.4210351>
- [15] Microsoft R Open 3.2.2 for Revolution R Enterprise 8.0.1, <https://www.microsoft.com/en-us/download/details.aspx?id=51205>, pristupljeno 28.07.2023.
- [16] Davenport, T. H. (2014). Stop using the term “big data”, Deloitte Insights, <https://www2.deloitte.com/us/en/insights/topics/analytics/big-data-buzzword.html>, pristupljeno 28.07.2023.
- [17] Jayachandran, U. (2023). Microsoft R Application Network retirement, Tech Community Microsoft,

- <https://techcommunity.microsoft.com/t5/azure-sql-blog/microsoft-r-application-network-retirement/ba-p/3707161>, pristupljeno 28.07.2023.
- [18] Rowland-Jones, J. (2021). Looking to the future for R in Azure SQL and SQL Server, Cloud Blogs Microsoft, <https://cloudblogs.microsoft.com/sqlserver/2021/06/30/looking-to-the-future-for-r-in-azure-sql-and-sql-server/>, pristupljeno 28.07.2023.
- [19] Vignoli, I. (2022). Sustainability of OSS Projects (the Experience of LibreOffice). *Primena slobodnog softvera i otvorenog hardvera* (in English "Applicaton of Free Software and Open Hardware") (PSSOH), Belgrade, Serbia. Zenodo. <https://doi.org/10.5281/zenodo.7244312>
- [20] Pan, J., & Tompkins, W. J. (1985). A real-time QRS detection algorithm. *IEEE Transactions on Biomedical Engineering*, (3), 230-236. <https://doi.org/10.1109/TBME.1985.325532>
- [21] Sjoerdsma, B. (2016). *Dealing with vendor lock-in* (Bachelor's thesis, University of Twente).
- [22] LERU Data Statement, League of European Research Universities LERU Office, Leuven, Belgium, 2021, https://www.leru.org/files/LERU-Data-Statement_12.2021.pdf, pristupljeno 28.07.2023.
- [23] Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>
- [24] Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., ... & Van Mulbregt, P. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3), 261-272. <https://doi.org/10.1038/s41592-019-0686-2>
- [25] Van Rossum, G., & Drake, F. L. (2009). *Python/C Api manual-python 3*. CreateSpace.
- [26] Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... & Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825), 357-362. <https://doi.org/10.1038/s41586-020-2649-2>
- [27] Oliphant, T. (2021). NumPy, SciPy, Anaconda, Python & Scientific Programming, Lex Fridman Podcast, <https://youtu.be/gFEE3w7F0ww>, pristupljeno 27.05.2024.
- [28] The Julia Community. (2022). Why we use Julia, 10 years later, Julia Blog, <https://julialang.org/blog/2022/02/10years/>, pristupljeno 27.05.2024.
- [29] Bezanson, J., Karpinski, S., Shah, V. B., & Edelman, A. (2012). Julia: A fast dynamic language for technical computing. *ArXiv Preprint*, <https://doi.org/10.48550/arXiv.1209.5145>
- [30] Nield, T. (2019). Data Science Has Become Too Vague: Let's Specialize and Break It Up! Towards Data Science, <https://towardsdatascience.com/data-science-has-become-too-vague-538899bab57>, pristupljeno 28.07.2023.
- [31] Seilmayer, M. (2020). serial: The Serial Interface Package, CRAN, <https://cran.r-project.org/web/packages/serial/index.html>, pristupljeno 28.07.2023.
- [32] Wickham, H., & Bryan, J. (2023). *R packages*. O'Reilly Media, Inc.
- [33] Chacon, S., & Straub, B. (2014). *Pro git* (p. 456). Springer Nature. Poglavlje: 1.2 Getting Started – A Short History of Git, <https://git-scm.com/book/en/v2/Getting-Started-A-Short-History-of-Git>, pristupljeno 28.07.2023.
- [34] Loeliger, J., & McCullough, M. (2012). *Version Control with Git: Powerful tools and techniques for collaborative software development*. O'Reilly Media, Inc.
- [35] CRAN Repository Maintainers, CRAN Repository Policy, <https://cran.r-project.org/web/packages/policies.html>, pristupljeno 29.07.2023.

- [36] Dalgaard, P., Dalgaard, M. P., & Dalgaard, P. (2015). Package 'ISwR'. *Energy*, 8, 1. <https://cran.r-project.org/package=ISwR>, pristupljeno 27.05.2024.
- [37] Dalgaard, P. (2008). *Introductory statistics with R*. Springer-Verlag.
- [38] Wickham, H., Francois, R., Henry, L., & Müller, K. (2015). *Dplyr: A Grammar of Data Manipulation*. R Found. Stat. Comput. R package version 0.4. 3, CRAN, <https://CRAN.R-project.org/package=dplyr>. pristupljeno 28.07.2023.
- [39] Project pip, Python Package Index - PyPI. (n.d.). Python Software Foundation. <https://pypi.org/project/pip/>, pristupljeno 28.07.2023.
- [40] Lambba, C. (2015). RAM is eating big data – Size of data-sets used for analytic's, <https://www.linkedin.com/pulse/ram-eating-big-data-size-data-sets-used-analytics-chaaranpall-lambba>, pristupljeno 29.07.2023.
- [41] Ray, P. P. (2017). An introduction to dew computing: definition, concept and implications. *IEEE Access*, 6, 723-737. <https://doi.org/10.1109/ACCESS.2017.2775042>
- [42] Van Steen, M., & Tanenbaum, A. (2002). *Distributed systems principles and paradigms*. *Network*, 2(28), 1.
- [43] Apache Software Foundation. (2008). Apache License v2.0 and GPL Compatibility. <http://www.apache.org/licenses/GPL-compatibility.html>, pristupljeno 27.05.2024.
- [44] Remote Physiological Measurement: Novel algorithms and application areas. (2023). Special Issue in Biomedical Signal Processing and Control Journal, <https://www.sciencedirect.com/journal/biomedical-signal-processing-and-control/about/call-for-papers>, pristupljeno 29.07.2023.
- [45] UNESCO, UNESCO Recommendation on Open Science, United Nations Educational, Scientific and Cultural Organization, Paris, France, SC-PCB-SPP/2021/OS/UROS, <https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en>, pristupljeno 29.07.2023.
- [46] Registar i obilježavanje opreme, Hrvatska zaklada za znanost, <https://www3.hrzz.hr/default.aspx?id=2277>, pristupljeno 29.07.2023.
- [47] National Academies of Sciences, Engineering, and Medicine. (2019). *Reproducibility and Replicability in Science*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/25303>
- [48] Miljković, N. (2021). Interview with Prof. Predrag Pejović on Computational Reproducibility for CURE-FAIR RDA WG Subgroup 3 (Version 1), Zenodo. <https://doi.org/10.5281/zenodo.4469568>
- [49] Heil, B. J., Hoffman, M. M., Markowitz, F., Lee, S. I., Greene, C. S., & Hicks, S. C. (2021). Reproducibility standards for machine learning in the life sciences. *Nature Methods*, 18(10), 1132-1135. <https://doi.org/10.1038/s41592-021-01256-7>
- [50] Miljković, N., & Milovanović, V. (2023). An Emerging Trend: Open and Not So Open Licenses for Machine and Deep Learning, *The Second Serbian International Conference on Applied Artificial Intelligence (SICAAI)*, University of Kragujevac, Serbia, Kragujevac.
- [51] Dalton, M. (2019). Submission of Software Programs to Regulatory Agencies, Paper SA04, PhUSE US Connect, pp. 1-10.
- [52] Matthews, C. I., & Shilling, B. C. (2008). *Validating Clinical Trial Data Reporting with SAS*. SAS Institute.
- [53] R: Regulatory compliance and validation issues a guidance document for the use of R in regulated clinical trial environments. *The R Foundation for Statistical Computing*, 2021, pp. 1-25, <https://www.r-project.org/doc/R-FDA.pdf>, pristupljeno 30.07.2023.

- [54] Simonsohn, U. (2021). [95] Groundhog: Addressing the treat that R poses to reproducible research, Datacolada Blog, <http://datacolada.org/95>, pristupljeno 30.07.2023.
- [55] Simonsohn, U. (2022). [100] Groundhog 2.0: Further addressing the treat R poses to reproducible research, Datacolada Blog, <http://datacolada.org/100>, pristupljeno 30.07.2023.
- [56] Trisovic, A., Lau, M. K., Pasquier, T., & Crosas, M. (2022). A large-scale study on research code quality and execution. *Scientific Data*, 9(1), 60. <https://doi.org/10.1038/s41597-022-01143-6>
- [57] Simonsohn, U., & Gruson. (2023). groundhog: Version-Control for CRAN, GitHub, and GitLab Packages, CRAN, <https://cran.r-project.org/package=groundhog>, pristupljeno 30.07.2023.
- [58] Stodden, V., Krafczyk, M. S., & Bhaskar, A. (2018). Enabling the verification of computational results: An empirical evaluation of computational reproducibility. In *Proceedings of the First International Workshop on Practical Reproducible Evaluation of Computer Systems* (pp. 1-5). <https://doi.org/10.1145/3214239.3214242>
- [59] Peng, R. D. (2023). R Programming, Johns Hopkins University, Coursera, <https://www.coursera.org/learn/r-programming>, pristupljeno 30.07.2023.
- [60] Eaton, J. W., Bateman, D., Hauberg, S., & Wehbring, R. (2014). GNU Octave (Version 3.8. 1): A high-level interactive language for numerical computations. Computer software. CreateSpace, <http://www.gnu.org/software/octave/doc/interpreter>, pristupljeno 30.07.2023.
- [61] Hiebeler, D. E. (2018). *R and Matlab*. CRC Press.
- [62] Strobl, C., Boulesteix, A. L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, 8(1), 1-21. <https://doi.org/10.1186/1471-2105-8-25>
- [63] Prince, V. (2005). Sex vs. gender. *International Journal of Transgenderism*, 8(4), 29-32.
- [64] Oakley, A. (2015). *Sex, gender and society*. Ashgate Publishing, Ltd.
- [65] Heidari, S., Babor, T. F., De Castro, P., Tort, S., & Curno, M. (2016). Sex and gender equity in research: rationale for the SAGER guidelines and recommended use. *Research Integrity and Peer Review*, 1(1), 1-9. <https://doi.org/10.1186/s41073-016-0007-6>
- [66] Lin, W. C., & Tsai, C. F. (2020). Missing value imputation: a review and analysis of the literature (2006–2017). *Artificial Intelligence Review*, 53, 1487-1509. <https://doi.org/10.1007/s10462-019-09709-4>
- [67] Sosulski, K. (2018). *Data visualization made simple: insights into becoming visual*. Routledge.
- [68] Oblaković, M., Sokolovska, V., & Dinić, B. (2015). Tretmani nedostajućih podataka. *Primenjena psihologija*, 8(3), 289-309. <https://doi.org/10.19090/pp.2015.3.289-309>
- [69] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- [70] Đukić, D. (2023). Primena metoda mašinskog učenja za predikciju bola u leđima, projekat iz TOBS (Tehnike obrade biomedicinskih signala) predmeta na Univerzitetu u Beogradu – Elektrotehničkom fakultetu, <https://automatika.etf.bg.ac.rs/sr/13m051tobs>, pristupljeno 28.05.2024.
- [71] Stokanović, S., Đukić, D., & Miljković, N. (2024). Robustness of XGBoost Algorithm to Missing Features for Binary Classification of Medical Data. In *2024 23rd*

- International Symposium INFOTEH-JAHORINA (INFOTEH)* (pp. 1-6). IEEE. <https://doi.org/10.1109/INFOTEH60418.2024.10495929>
- [72] Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., & Li, Y. (2023). xgboost: Extreme Gradient Boosting, CRAN, <https://cran.r-project.org/package=xgboost>, pristupljeno 26.08.2023.
- [73] Canniford, W. (2018). Predicting lower back pain using caret in R – Lower back pain symptoms dataset, Kaggle, Version 20, <https://www.kaggle.com/code/willcanniford/predicting-lower-back-pain-using-caret-in-r/notebook#Evaluating-our-models-and-predictions>, pristupljeno 30.07.2023.
- [74] Libiseller-Egger, J., Phelan, J. E., Attia, Z. I., Benavente, E. D., Campino, S., Friedman, P. A., ... & Clark, T. G. (2022). Deep learning-derived cardiovascular age shares a genetic basis with other cardiac phenotypes. *Scientific Reports*, 12(1), 22625. <https://doi.org/10.1038/s41598-022-27254-z>
- [75] Adedinsewo, D., Hardway, H. D., Morales-Lara, A. C., Wieczorek, M. A., Johnson, P. W., Douglass, E. J., ... & Yamani, M. (2023). Non-invasive detection of cardiac allograft rejection among heart transplant recipients using an electrocardiogram based deep learning model. *European Heart Journal-Digital Health*, 4(2), 71-80. <https://doi.org/10.1093/ehjdh/ztad001>
- [76] Attia, Z. I., Friedman, P. A., Noseworthy, P. A., Lopez-Jimenez, F., Ladewig, D. J., Satam, G., ... & Kapa, S. (2019). Age and sex estimation using artificial intelligence from standard 12-lead ECGs. *Circulation: Arrhythmia and Electrophysiology*, 12(9), e007284. <https://doi.org/10.1161/CIRCEP.119.007284>
- [77] Kang, H. (2013). The prevention and handling of the missing data. *Korean Journal of Anesthesiology*, 64(5), 402-406. <https://doi.org/10.4097/kjae.2013.64.5.402>
- [78] Miljković, N. (2016). *Metode i instrumentacija za električna merenja*, Elektrotehnički fakultet, Univerziteta u Beogradu, <https://doi.org/10.5281/zenodo.1335250>
- [79] Miljković, N., Popović, N. B., & Sodnik, J. (2024). Electrogastrogram Signal Processing: Techniques and Challenges with Application for Simulator Sickness Assessment. In *Biomedical Signal Processing* (pp. 62-89). CRC Press. <http://dx.doi.org/10.1201/9781003201137-4>
- [80] Martinez, G. J., Mattingly, S. M., Mirjafari, S., Nepal, S. K., Campbell, A. T., Dey, A. K., & Striegel, A. D. (2020). On the quality of real-world wearable data in a longitudinal study of information workers. In *2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)* (pp. 1-6). IEEE. <https://doi.org/10.1109/PerComWorkshops48775.2020.9156113>
- [81] Stanković, L., Daković, M., & Vujović, S. (2017). Reconstruction of sparse signals in impulsive disturbance environments. *Circuits, Systems, and Signal Processing*, 36, 767-794. <https://doi.org/10.1007/s00034-016-0334-3>
- [82] Vigen, T. (2015). *Spurious correlations*. Hachette UK.
- [83] Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581-592. <https://doi.org/10.1093/biomet/63.3.581>
- [84] General Assembly of the World Medical Association. (2014). World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *The Journal of the American College of Dentists*, 81(3), 14-18.
- [85] Miljković, N., & Sodnik, J. (2023). Effectiveness of a time to fixate for fitness to drive evaluation in neurological patients. *Behavior Research Methods*, 1-16. <https://doi.org/10.3758/s13428-023-02177-3>
- [86] Motnikar, L., Stojmenova, K., Štaba, U. Č., Klun, T., Robida, K. R., & Sodnik, J. (2020). Exploring driving characteristics of fit-and unfit-to-drive neurological patients:

- A driving simulator study. *Traffic Injury Prevention*, 21(6), 359-364. <https://doi.org/10.1080/15389588.2020.1764547>
- [87] Walfish, S. (2006). A review of statistical outlier methods. *Pharmaceutical Technology*, 30(11), 82.
- [88] Miljković, N., & Sodnik, J. (2023b). Parameters for Statistical Evaluation of Time to Fixate Effectiveness for Assessment of Fitness to Drive (Version 2) [Data set], *Zenodo*. <https://doi.org/10.5281/zenodo.7963337>
- [89] Venkataraman, S. & Moura, P. (2020). Data formats for preservation: What you need to know when creating a DMP, *Zenodo*, <https://doi.org/10.5281/zenodo.4041512>
- [90] File formats R.5.5. Version 1.1, 30.03.2023, DANS, the Dutch national centre of expertise and repository for research data, <https://dans.knaw.nl/en/file-formats/>, pristupljeno 01.08.2023.
- [91] Deljanin, S. (2012). Digital obsolescence, *INFOtheca-Journal of Informatics & Librarianship*, 13(1).
- [92] Vignoli, I. (2021). Digital Sovereignty, *PSSOH Conference*, Belgrade, Serbia, *Zenodo*, <https://doi.org/10.5281/zenodo.5713734>
- [93] Vignoli, I. (2023). Softver otvorenog koda i otvoreni standardi za digitalni suverenitet, *Balkan Magazin*, <http://www.balkanmagazin.net/tehnologija/cid132-330792/softver-otvorenog-koda-i-otvoreni-standardi-za-digitalni-suverenitet>, pristupljeno 01.08.2023.
- [94] Vincent, J. (2020). Scientists rename human genes to stop Microsoft Excel from misreading them as dates, *The Verge*, <https://www.theverge.com/2020/8/6/21355674/human-genes-rename-microsoft-excel-misreading-dates>, pristupljeno 28.05.2024.
- [95] Pale, U., Müller, N., Arza, A., & Atienza, D. (2021). ReBeatICG: Real-time low-complexity beat-to-beat impedance cardiogram delineation algorithm. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (pp. 5618-5624). IEEE. <https://doi.org/10.1109/EMBC46164.2021.9630170>
- [96] Pale, U., Müller, N., Arza, A., & Atienza, D. (2021). ReBeatICG database (v20210503-01) [Data set], *Zenodo*. <https://doi.org/10.5281/zenodo.4725433>
- [97] MAT-File Versions: Overview of MAT-File Versions, R2023a, Mathworks, https://www.mathworks.com/help/matlab/import_export/mat-file-versions.html, pristupljeno 01.08.2023.
- [98] Bengtsson, H., Jacobson, A., Riedy, J., Bengtsson, M. H., LazyLoad, T. R. U. E., & ByteCompile, T. R. U. E. (2018). Package ‘R. matlab’. *R package version*, 3(2). <https://cran.r-project.org/package=R.matlab>, pristupljeno 01.08.2023.
- [99] Miljković, N., & Šekara T. B. (2023). Software for Detection of B-point in Impedance Cardiogram with Dataset Consisting of 20 Healthy Participants (Version 3), *Zenodo*. <https://doi.org/10.5281/zenodo.7968424>
- [100] Miljković, N., & Šekara, T. B. (2022). A New Weighted Time Window-based Method to Detect B-point in ICG. *arXiv preprint arXiv:2207.04490*. <https://arxiv.org/ftp/arxiv/papers/2207/2207.04490.pdf>
- [101] Miljković, N., Dubljević, O., Bjeogojević, B., & Milosavljević, N. (2020). Biosignals: Measurement and Analysis with Applications in Psychology. *Empirical Studies in Psychology*, 26.
- [102] Higgins, J. R. (1996). *Sampling theory in Fourier and signal analysis: Foundations*. Oxford Science Publications.

- [103] Meijering, E. (2002). A chronology of interpolation: from ancient astronomy to modern signal and image processing. *Proceedings of the IEEE*, 90(3), 319-342. <https://doi.org/10.1109/5.993400>
- [104] Miljković, N. (2013). Metode i instrumentacija za procenu aktivnosti motornog sistema na osnovu elektromiografskih signala, doktorska disertacija, *Elektrotehnički fakultet, Univerziteta u Beogradu*, https://hdl.handle.net/21.15107/rcub_nardus_5831
- [105] Konrad, P. (2005). The ABC of EMG. *A practical introduction to kinesiological electromyography*, 1(2005), 30-5. <http://www.noraxon.com/wp-content/uploads/2014/12/ABC-EMG-ISBN.pdf>, pristupljeno 03.08.2023.
- [106] Levine, W. S. (2018). *The Control Handbook (three volume set)*. CRC press.
- [107] Fadali, M. S., & Visioli, A. (2012). *Digital control engineering: Analysis and design*. Academic Press.
- [108] R Core Team, utils – The R Utils Package, <https://rdocumentation.org/packages/utils/versions/3.6.2>, pristupljeno 06.08.2023.
- [109] Bengtsson, H. (2023). Rutils: Various Programming Utilities [R package R. utils version 2.12. 3], CRAN, <https://cran.r-project.org/package=R.utils>, pristupljeno 28.05.2024.
- [110] Wickham, H. (2019). *Advanced R*. CRC press. Memory, <http://adv-r.had.co.nz/memory.html>, pristupljeno 06.08.2023.
- [111] Wickham, H., Hester, J., Francois, R., Bryan, J., Bearrows, S., & Posit, P. B. C. (2023). Package ‘readr’. *Read Rectangular Text Data*. <https://cran.r-project.org/package=readr>, pristupljeno 03.08.2023.
- [112] Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., ... & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation*, 101(23), e215-e220. <https://doi.org/10.1161/01.CIR.101.23.e215>
- [113] Raymond, E. S. (2003). *The art of Unix programming*. Addison-Wesley Professional.
- [114] Wilson, R. J. (2023). Pulse Dataset Information, Department of Mathematics, University of Queensland, <http://www.statsci.org/data/oz/ms212.html>, pristupljeno 03.08.2023.
- [115] Miljković, N. (2019). IoT & električna merenja, predavanje održano tokom EESTech Challenge manifestacije na Univerzitetu u Beogradu – Elektrotehničkom fakultetu, https://automatika.etf.bg.ac.rs/images/FAJLOVI_srpski/predmeti/izborni_kursevi_os/biomedicinsko_inzenjerstvo/TOBS/NadicaMiljkovicEESTEC27032019.pdf, pristupljeno 03.08.2023.
- [116] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 1-9. <https://doi.org/10.1038/sdata.2016.18>
- [117] Ushey, K., Allaire, J. J., Wickham, H., & Ritchie, G. (2023). RStudio, rstudioapi: Safely Access the RStudio API, version 0.15.0, CRAN, <https://cran.r-project.org/package=rstudioapi>, pristupljeno 03.08.2023.
- [118] DataCamp, R Tutorial: Connect To A Database, 10.11.2016, <https://youtu.be.OSB9pmlJpI>, pristupljeno 04.08.2023.
- [119] Miljković, N., Trisovic, A., & Peer, L. (2021). Towards FAIR Principles for Open Hardware. PSSOH Conference, Belgrade, Serbia, pp. 90–101. <https://doi.org/10.5281/zenodo.6506428>

- [120] Licensing for products based on Arduino, Arduino CC, <https://support.arduino.cc/hc/en-us/articles/4415094490770-Licensing-for-products-based-on-Arduino>, pristupljeno 04.08.2023.
- [121] Smith, A. G. (2011). Introduction to arduino. <https://www.introtoarduino.com/downloads/IntroArduinoBook.pdf>, pristupljeno 04.08.2023.
- [122] Becchetti, C., & Neri, A. (2013). *Medical instrument design and development: From requirements to market placements*. John Wiley & Sons.
- [123] How to Read Data from Serial Port in R, Stack Overflow, <https://stackoverflow.com/questions/34522535/how-to-read-data-from-serial-port-in-r>, pristupljeno 04.08.2023.
- [124] Suh, C., & Hoff, P. (2017). Rduino: A Microcontroller Interface, CRAN, <https://cran.r-project.org/package=Rduino>, pristupljeno 03.08.2023.
- [125] Rserve – Binary R server, Rforge.net, <https://www.rforge.net/Rserve/index.html>, pristupljeno 04.08.2023.
- [126] Kalinowski, T., Ushey, K., Allaire, J. J., Rstudio, Tang, Y., Eddelbuettel, D., Lewis, B., Keydana, S., Hafen, R., & Geelnard, M. (2023). reticulate: Interface to “Python”, CRAN, <https://cran.r-project.org/package=reticulate>, pristupljeno 05.08.2023.
- [127] subprocess – Subprocess management, <https://docs.python.org/3/library/subprocess.html#module-subprocess>, pristupljeno 05.08.2023.
- [128] Python → R brdge, rpy2, <https://pypi.org/project/rpy2/>, pristupljeno 05.08.2023.
- [129] Open Data Handbook: Machine readable, <http://opendatahandbook.org/glossary/en/terms/machine-readable/>, pristupljeno 08.08.2023.
- [130] Li, Y., Qu, Q., Wang, M., Yu, L., Wang, J., Shen, L., & He, K. (2020). Deep learning for digitizing highly noisy paper-based ECG records. *Computers in Biology and Medicine*, 127, 104077. <https://doi.org/10.1016/j.compbiomed.2020.104077>
- [131] Irizarry, R. A. (2019). *Introduction to data science: Data analysis and prediction algorithms with R*. CRC Press. <http://rafalab.dfci.harvard.edu/dsbook/>, pristupljeno 08.08.2023.
- [132] Ooms, J. (2023). pdftools: Text Extraction, Rendering and Coverting PDF Documents, CRAN, <https://cran.r-project.org/package=pdfutils>, pristupljeno 08.08.2023.
- [133] Miljković, N., & Stanković, B. (2019). Gender gap in electrical engineering at the University of Belgrade (1923-2010): Analysis of graduates' structure using R. PSSOH Conference, Belgrad, Serbia, pp. 22–25. <https://doi.org/10.5281/zenodo.3520315>
- [134] Taylor, N. E. (2016). Intensive Analysis, MIT Spectrum, <https://spectrum.mit.edu/fall-2016/intensive-analysis>, pristupljeno 08.08.2023.
- [135] Feng, Y., & Zhu, J. (2022). R Programming: Zero to Pro, published with bookdown R package, <https://r02pro.github.io/>, pristupljeno 05.08.2023.
- [136] Shure, L. (2008). Programming Patterns: Some Common MATLAB Programming Pitfalls and How to Avoid Them, Technical Article, MathWorks, <https://www.mathworks.com/company/newsletters/articles/programming-patterns-some-common-matlab-programming-pitfalls-and-how-to-avoid-them.html>, pristupljeno 05.08.2023.
- [137] Adler, J. (2010). *R in a nutshell: A desktop quick reference*. O'Reilly Media, Inc.
- [138] Wickham, H., & Henry, L. (2023). *purrr: Functional Programming Tools*. <https://purrr.tidyverse.org/>, <https://github.com/tidyverse/purrr>, pristupljeno 05.08.2023.

- [139] R Core Team, & Contributors, The R Datasets Package, <https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/00Index.htm>, pristupljeno 05.08.2023.
- [140] Dobson, A. J. (2013). *Introduction to statistical modelling*. Springer.
- [141] Singh, S., & Ernst, E. (2008). *Trick or treatment: The undeniable facts about alternative medicine*. WW Norton & Company. (knjiga je prevedena na srpski jezik "Napitak za boljitak").
- [142] R Core Team, stats-package: The R Stats Package, <https://rdrr.io/r/stats/stats-package.html>, pristupljeno 07.08.2023.
- [143] Gentleman, R., & Ihaka, R. (2000). Lexical scope and statistical computing. *Journal of Computational and Graphical Statistics*, 9(3), 491-508.
- [144] Priemer, R. (1991). *Introductory signal processing* (Vol. 6). World scientific.
- [145] Delgutte, B. (2007). Course materials for HST.582J / 6.555J / 16.456J, Biomedical Signal and Image Processing, MIT OpenCourseWare (<http://ocw.mit.edu>), Massachusetts Institute of Technology, https://ocw.mit.edu/courses/hst-582j-biomedical-signal-and-image-processing-spring-2007/resources/11_intro/, pristupljeno 07.08.2023.
- [146] Watanabe, K., Kouzaki, M., Ogawa, M., Akima, H., & Moritani, T. (2018). Relationships between muscle strength and multi-channel surface EMG parameters in eighty-eight elderly. *European Review of Aging and Physical Activity*, 15(1), 1-10. <https://doi.org/10.1186/s11556-018-0192-z>
- [147] Miljković, N., Milenić, N., Popović, N. B., & Sodnik, J. (2024). Data augmentation for generating synthetic electrogastrogram time series. *Medical & Biological Engineering & Computing*, 1-13. <https://doi.org/10.1007/s11517-024-03112-0>
- [148] Volkov, A. G., Vilfranc, C. L., Murphy, V. A., Mitchell, C. M., Volkova, M. I., O'Neal, L., & Markin, V. S. (2013). Electrotonic and action potentials in the Venus flytrap. *Journal of Plant Physiology*, 170(9), 838-846. <https://doi.org/10.1016/j.jplph.2013.01.009>
- [149] Di Palma, J. R., Mohl, R., & Best Jr, W. (1961). Action potential and contraction of *Dionaea muscipula* (Venus flytrap). *Science*, 133(3456), 878-879. <https://doi.org/10.1126/science.133.3456.878>
- [150] Forbes, A., Cobb, S., & Cattell, M. (1921). An electrocardiogram and an electromyogram in an elephant. *American Journal of Physiology-Legacy Content*, 55(3), 385-389. <https://doi.org/10.1152/ajplegacy.1921.55.3.385>
- [151] Dey, N., Ashour, A. S., Mohamed, W. S., Nguyen, N. G., Dey, N., Ashour, A. S., ... & Nguyen, N. G. (2019). Biomedical signals. *Acoustic Sensors for Biomedical Applications*, 7-20. https://doi.org/10.1007/978-3-319-92225-6_2
- [152] Kim, J. M., Hong, J. H., Cho, M. C., Cha, E. J., & Lee, T. S. (2007). Wireless biomedical signal monitoring device on wheelchair using noncontact electro-mechanical film sensor. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 574-577). IEEE. <https://doi.org/10.1109/iembs.2007.4352355>
- [153] Faes, L., Chon, K. H., & Nollo, G. (2008). A method for the time-varying nonlinear prediction of complex nonstationary biomedical signals. *IEEE Transactions on Biomedical Engineering*, 56(2), 205-209. <https://doi.org/10.1109/TBME.2008.2008726>
- [154] Banerjee, A., Chakraborty, C., Kumar, A., & Biswas, D. (2020). Emerging trends in IoT and big data analytics for biomedical and health care technologies. *Handbook of data science approaches for biomedical engineering*, 121-152. <https://doi.org/10.1016/B978-0-12-818318-2.00005-2>

- [155] Garg, V. K., & Wang, Y. C. (2005). Signal types, properties, and processes. In *The Electrical Engineering Handbook* (pp. 951-956). Academic Press. <https://doi.org/10.1016/B978-012170960-0/50069-4>
- [156] Proakis, J. G., & Manolakis, D. G. (2007). *Digital Signal Processing*, Upper Saddle River. New Jersey, 7458.
- [157] Soliński, M., Gierakowski, J., & Żebrowski, J. (2016). Modeling heart rate variability including the effect of sleep stages. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 26(2). <http://dx.doi.org/10.1063/1.4940762>
- [158] Garfinkel, A., Spano, M. L., Ditto, W. L., & Weiss, J. N. (1992). Controlling cardiac chaos. *Science*, 257(5074), 1230-1235. <https://doi.org/10.1126/science.1519060>
- [159] Electrophysiology, T. F. O. T. E. S. O. C. T. N. A. S. O. P. (1996). Heart rate variability: standards of measurement, physiological interpretation, and clinical use. *Circulation*, 93(5), 1043-1065. <https://doi.org/10.1161/01.CIR.93.5.1043>
- [160] Wagner, G. S., Lim, T. H., Strauss, D. G., & Simlund, J. (2008). Interpretation of the normal electrocardiogram. *Marriott's Practical Electrocardiography*, 9, 50.
- [161] Klonowski, W. (2009). Everything you wanted to ask about EEG but were afraid to get the right answer. *Nonlinear Biomedical Physics*, 3(1), 1-5. <https://doi.org/10.1186/1753-4631-3-2>
- [162] Osgood, B. (2002). Lecture notes for EE 261 the Fourier transform and its applications. <https://see.stanford.edu/materials/lsoftae261/book-fall-07.pdf>, pristupljeno 08.08.2023.
- [163] Kidd, C., & Hornay, R. (2021). Data Analytics vs Data Analysis: What's The Difference?, Machine Learning & Big Data Blog, <https://www.bmc.com/blogs/data-analytics-vs-data-analysis/>, pristupljeno 08.08.2023.
- [164] Saranya, P., & Asha, P. (2019). Survey on big data analytics in health care. In *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)* (pp. 46-51). IEEE. <https://doi.org/10.1109/ICSSIT46314.2019.8987882>
- [165] The Turing Way Community, Registered Reports, <https://the-turing-way.netlify.app/communication/dif-articles/reg.html>, pristupljeno 06.08.2023.
- [166] Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, 2(3), 196-217. https://doi.org/10.1207/s15327957pspr0203_4
- [167] Tag, B., Vargo, A. W., Gupta, A., Chernyshov, G., Kunze, K., & Dingler, T. (2019). Continuous alertness assessments: Using EOG glasses to unobtrusively monitor fatigue levels In-The-Wild. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-12). <https://doi.org/10.1145/3290605.3300694>
- [168] Stern, J. A., Boyer, D., & Schroeder, D. (1994). Blink rate: a possible measure of fatigue. *Human Factors*, 36(2), 285-297. <https://doi.org/10.1177/001872089403600209>
- [169] Sokol, E. R. (2007). Chapter 24 - Urinary Incontinence. In A. I. Sokol & E. R. Sokol (Eds.), *General Gynecology* (pp. 583-607). Mosby. <https://doi.org/https://doi.org/10.1016/B978-032303247-6.10024-3>
- [170] Orley, J., & Kuyken, W. (1994). *Quality of life assessment: International perspectives* (pp. 41-60). Berlin: Springer-Verlag. <https://doi.org/10.1007/978-3-642-79123-9>
- [171] World Health Organization. (2004). *The world health organization quality of life (WHOQOL)-BREF* (No. WHO/HIS/HSI Rev. 2012.02). World Health Organization. <https://www.who.int/tools/whoqol>, pristupljeno 06.08.2023.
- [172] Correia, G. N., Pereira, V. S., Hirakawa, H. S., & Driusso, P. (2014). Effects of surface and intravaginal electrical stimulation in the treatment of women with stress urinary

- incontinence: randomized controlled trial. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 173, 113-118. <https://doi.org/10.1016/j.ejogrb.2013.11.023>
- [173] Stewart, F., Berghmans, B., Bø, K., & Glazener, C. M. (2017). Electrical stimulation with non-implanted devices for stress urinary incontinence in women. *Cochrane Database of Systematic Reviews*, (12). <https://doi.org/10.1002/14651858.CD012390.pub2>
- [174] Hall, E. H. (1879). On a new action of the magnet on electric currents. *American Journal of Mathematics*, 2(3), 287-292.
- [175] Dajkstra, E. (1997). Snage akademske delatnosti, Trag Nula, prevod originalnog članka pod nazivom "The strengths of the academic enterprise", <http://trag.nula.rs/post/edsgerdajkstra-snage-akademske-delatnosti/>, pristupljeno 14.08.2023. https://doi.org/10.1007/978-3-642-60858-2_14
- [176] Carmody, S. (2023). ngramr: Retrieve and Plot Google n-Gram Data, CRAN, <https://cran.r-project.org/package=ngramr>, pristupljeno 06.08.2023.
- [177] Krstajić, D. (2023). Pristrasnost u razumevanju verovatnoće, Kulturni dodatak str. 9, Politika, https://www.rcc.org.rs/Pristrasnost_verovatnoce.pdf, pristupljeno 06.08.2023.
- [178] Tversky, A., & Kahneman, D. (1989). Rational choice and the framing of decisions. In *Multiple criteria decision making and risk analysis using microcomputers* (pp. 81-126). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-74919-3_4
- [179] Kahneman, D., & Tversky, A. (2013). Prospect theory: An analysis of decision under risk. In *Handbook of the fundamentals of financial decision making: Part I* (pp. 99-127). https://doi.org/10.1142/9789814417358_0006
- [180] Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, 76(2), 105. <https://psycnet.apa.org/doi/10.1037/h0031322>
- [181] Jovanović, V. Đ. (1967). Vežbe u laboratoriji za električna merenja, deseto neizmenjeno izdanje, Naučna knjiga, Beograd.
- [182] Posner, M. I. (2005). Timing the brain: Mental chronometry as a tool in neuroscience. *PLoS Biology*, 3(2), e51. <https://doi.org/10.1371/journal.pbio.0030051>
- [183] Gaouda, A. M., El-Hag, A., Abdel-Galil, T. K., Salama, M. M. A., & Bartnikas, R. (2008). On-line detection and measurement of partial discharge signals in a noisy environment. *IEEE Transactions on Dielectrics and Electrical Insulation*, 15(4), 1162-1173. <https://doi.org/10.1109/TDEI.2008.4591239>
- [184] Najafipour, A., Babae, A., & Shahrtash, S. M. (2013). Comparing the trustworthiness of signal-to-noise ratio and peak signal-to-noise ratio in processing noisy partial discharge signals. *IET Science, Measurement & Technology*, 7(2), 112-118. <https://doi.org/10.1049/iet-smt.2012.0113>
- [185] Stambolić, M., urednik. (1986). Rečnik književnih termina, Beograd: Nolit
- [186] Rangayyan, R. M. (2015). Filtering for removal of artifacts. in *Biomedical Signal Analysis*, John Wiley & Sons, Ltd, pp. 91-231. <https://doi.org/10.1002/9781119068129.ch3>
- [187] Li, J. (2011). Thermal noise analysis in ECG applications. *Texas Instruments*, 1-6. <https://www.ti.com/lit/an/sbaa185/sbaa185.pdf>, pristupljeno 09.08.2023.
- [188] Bortolan, G., Christov, I., Simova, I., & Dotsinsky, I. (2015). Noise processing in exercise ECG stress test for the analysis and the clinical characterization of QRS and T wave alternans. *Biomedical Signal Processing and Control*, 18, 378-385. <https://doi.org/10.1016/j.bspc.2015.02.003>

- [189] Arns, M., & Sterman, M. B. (2019). *Neurofeedback: How it all started*. Brainclinics Insights.
- [190] Reissig, G. N., Oliveira, T. C., Costa, Á. V. L., Parise, A. G., Pereira, D. R., & Souza, G. M. (2021). Machine learning for automatic classification of tomato ripening stages using electrophysiological recordings. *Frontiers in Sustainable Food Systems*, 5, 696829. <https://doi.org/10.3389/fsufs.2021.696829>
- [191] Tran, D., Dutoit, F., Najdenovska, E., Wallbridge, N., Plummer, C., Mazza, M., ... & Camps, C. (2019). Electrophysiological assessment of plant status outside a Faraday cage using supervised machine learning. *Scientific Reports*, 9(1), 17073. <https://doi.org/10.1038/s41598-019-53675-4>
- [192] Verleger, R., Gasser, T., & Möcks, J. (1982). Correction of EOG artifacts in event-related potentials of the EEG: Aspects of reliability and validity. *Psychophysiology*, 19(4), 472-480. <https://doi.org/10.1111/j.1469-8986.1982.tb02509.x>
- [193] Maddirala, A. K., & Shaik, R. A. (2016). Removal of EOG artifacts from single channel EEG signals using combined singular spectrum analysis and adaptive noise canceler. *IEEE Sensors Journal*, 16(23), 8279-8287. <https://doi.org/10.1109/JSEN.2016.2560219>
- [194] Kumar, P. S., Arumuganathan, R., Sivakumar, K., & Vimal, C. (2008). Removal of ocular artifacts in the EEG through wavelet transform without using an EOG reference channel. *Int. J. Open Problems Compt. Math*, 1(3), 188-200. [http://ftp.gwdg.de/pub/EMIS/journals/IJOPCM/files/IJOPCM\(vol.1.3.2.D.8\).pdf](http://ftp.gwdg.de/pub/EMIS/journals/IJOPCM/files/IJOPCM(vol.1.3.2.D.8).pdf), pristupljeno 09.08.2023.
- [195] Li, J., Quintin, E., Wang, H., McDonald, B. E., Farrell, T. R., Huang, X., & Clancy, E. A. (2023). Application-Layer Time Synchronization and Data Alignment Method for Multichannel Biosignal Sensors Using BLE Protocol. *Sensors*, 23(8), 3954. <https://doi.org/10.3390/s23083954>
- [196] Rheinländer, C. C., & Wehn, N. (2016). Precise synchronization time stamp generation for Bluetooth low energy. In *2016 IEEE SENSORS* (pp. 1-3). IEEE. <https://doi.org/10.1109/ICSENS.2016.7808812>
- [197] De Maria, A. N., & Raisinghani, A. (2000). Comparative overview of cardiac output measurement methods: has impedance cardiography come of age?. *Congestive Heart Failure*, 6(2), 60-73.
- [198] Cacioppo, J. T., Tassinary, L. G., & Berntson, G. (Eds.). (2007). *Handbook of psychophysiology*. Cambridge university press.
- [199] Yang, Z., Wang, Z., Zhang, J., Huang, C., & Zhang, Q. (2015). Wearables can afford: Light-weight indoor positioning with visible light. In *Proceedings of the 13th annual international conference on mobile systems, applications, and services* (pp. 317-330). <https://doi.org/10.1145/2742647.2742648>
- [200] Williamson, J., Liu, Q., Lu, F., Mohrman, W., Li, K., Dick, R., & Shang, L. (2015). Data sensing and analysis: Challenges for wearables. In *The 20th Asia and South Pacific Design Automation Conference* (pp. 136-141). IEEE. <https://doi.org/10.1109/ASPDAC.2015.7058994>
- [201] Rachuri, K. K., Mascolo, C., Musolesi, M., & Rentfrow, P. J. (2011). Sociablesense: exploring the trade-offs of adaptive sampling and computation offloading for social sensing. In *Proceedings of the 17th annual international conference on Mobile computing and networking* (pp. 73-84). <https://doi.org/10.1145/2030613.2030623>
- [202] Guna, J., Jakus, G., Pogačnik, M., Tomažič, S., & Sodnik, J. (2014). An analysis of the precision and reliability of the leap motion sensor and its suitability for static and dynamic tracking. *Sensors*, 14(2), 3702-3720. <https://doi.org/10.3390/s140203702>

- [203] Clarkson, H. M. (2000). *Musculoskeletal assessment: joint range of motion and manual muscle strength*. Lippincott Williams & Wilkins.
- [204] Indrawati, I., & Sajidin, M. (2018). Active, passive, and active-assistive range of motion (ROM) exercise to improve muscle strength in post stroke clients: A systematic review. In *9th International Nursing Conference (INC 2018)*.
- [205] Jones, M. (1991). *Therapeutic Exercise: Foundations and techniques* | *Therapeutic Exercise: Foundations and techniques*, Carolyn Kisner MS PT, Lynn Allen Colby MS PT. FA Davies (USA)(Eds.).
- [206] Van der Kruk, E., & Reijne, M. M. (2018). Accuracy of human motion capture systems for sport applications; state-of-the-art review. *European Journal of Sport Science*, 18(6), 806-819. <https://doi.org/10.1080/17461391.2018.1463397>
- [207] Winter, D. A. (2009). *Biomechanics and motor control of human movement*. John Wiley & Sons.
- [208] Hicks, J. H. (1953). The mechanics of the foot: I. The joints. *Journal of Anatomy*, 87(Pt 4), 345.
- [209] Kojović, J., Miljković, N., Janković, M. M., & Popović, D. B. (2011). Recovery of motor function after stroke: a polymyography-based analysis. *Journal of Neuroscience Methods*, 194(2), 321-328. <https://doi.org/10.1016/j.jneumeth.2010.10.006>
- [210] BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML. (1993). *Guide to the Expression of Uncertainty in Measurement*. British Standards Institution.
- [211] Fuller, J. R., Lomond, K. V., Fung, J., & Côté, J. N. (2009). Posture-movement changes following repetitive motion-induced shoulder muscle fatigue. *Journal of Electromyography and Kinesiology*, 19(6), 1043-1052. <https://doi.org/10.1016/j.jelekin.2008.10.009>
- [212] FSR 402 Data Sheet, FSR 400 Series Round Force Sensing Resistor, Interlink Electronics, Sensor Technologies, <https://www.trossenrobotics.com/productdocs/2010-10-26-DataSheet-FSR402-Layout2.pdf>, pristupljeno 10.08.2023.
- [213] Obradović, S., & Stančin, S. (2023). Inertial Measurement Units' Reliability for Measuring Knee Joint Angle during Road Cycling. *Electronics*, 12(3), 751. <https://doi.org/10.3390/electronics12030751>
- [214] Jakus, G., Guna, J., Tomažič, S., & Sodnik, J. (2014). Evaluation of leap motion controller with a high precision optical tracking system. In *Human-Computer Interaction. Advanced Interaction Modalities and Techniques: 16th International Conference, HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014, Proceedings, Part II 16* (pp. 254-263). Springer International Publishing. https://doi.org/10.1007/978-3-319-07230-2_25
- [215] Umek, A., & Kos, A. (2021). Validation of MEMS Accelerometer for Rapid Hand Movement Measurement. *Procedia Computer Science*, 187, 530-537. <https://doi.org/10.1016/j.procs.2021.04.095>
- [216] Umek, A., Tomažič, S., & Kos, A. (2019). Application for impact position evaluation in tennis using UWB localization. *Procedia Computer Science*, 147, 307-313. <https://doi.org/10.1016/j.procs.2019.01.269>
- [217] Ahmad, N., Ghazilla, R. A. R., Khairi, N. M., & Kasi, V. (2013). Reviews on various inertial measurement unit (IMU) sensor applications. *International Journal of Signal Processing Systems*, 1(2), 256-262. <http://www.ijspss.com/uploadfile/2013/1128/20131128022014877.pdf>, pristupljeno 10.08.2023.

- [218] Dejnabadi, H., Jolles, B. M., & Aminian, K. (2005). A new approach to accurate measurement of uniaxial joint angles based on a combination of accelerometers and gyroscopes. *IEEE Transactions on Biomedical Engineering*, 52(8), 1478-1484. <https://doi.org/10.1109/TBME.2005.851475>
- [219] Williamson, R., & Andrews, B. J. (2001). Detecting absolute human knee angle and angular velocity using accelerometers and rate gyroscopes. *Medical & Biological Engineering & Computing*, 39, 294-302. <https://doi.org/10.1007/BF02345283>
- [220] Ignatov, A. (2018). Real-time human activity recognition from accelerometer data using Convolutional Neural Networks. *Applied Soft Computing*, 62, 915-922. <https://doi.org/10.1016/j.asoc.2017.09.027>
- [221] Bayat, A., Pomplun, M., & Tran, D. A. (2014). A study on human activity recognition using accelerometer data from smartphones. *Procedia Computer Science*, 34, 450-457. <https://doi.org/10.1016/j.procs.2014.07.009>
- [222] Brockett, C. L., & Chapman, G. J. (2016). Biomechanics of the ankle. *Orthopaedics and Trauma*, 30(3), 232-238. <https://doi.org/10.1016/j.mporth.2016.04.015>
- [223] Khalid, S., Khalil, T., & Nasreen, S. (2014). A survey of feature selection and feature extraction techniques in machine learning. In *2014 science and information conference* (pp. 372-378). IEEE. <https://doi.org/10.1109/SAI.2014.6918213>
- [224] Roach, K. E., & Miles, T. P. (1991). Normal hip and knee active range of motion: the relationship to age. *Physical Therapy*, 71(9), 656-665. <https://doi.org/10.1093/ptj/71.9.656>
- [225] Biometrics DataLINK, NEXGEN Ergonomics: The Global Source for Software & Instrumentation for Ergonomics, Biomechanics & Medicine, <http://www.nexgenergo.com/ergonomics/biodatalink.html>, pristupljeno 13.08.2023.
- [226] Müller, K., Wickham, H., Francois, R., Bryan, J., & Rstudio. (2023). tibble: Simple Data Frames, CRAN, <https://cran.r-project.org/package=tibble>, pristupljeno 16.08.2023.
- [227] Wickham, H., & Grolemund G. R. (2017). R for data science: import, tidy, transform, visualize, and model data. <https://r4ds.had.co.nz/index.html>, pristupljeno 16.08.2023.
- [228] Wickham, H. (2011). ggplot2. *Wiley interdisciplinary reviews: Computational statistics*, 3(2), 180-185. <https://cran.r-project.org/package=ggplot2>, pristupljeno 16.08.2023.
- [229] Morel, P. (2018). Gramm: grammar of graphics plotting in Matlab, *Journal of Open Source Software*, 3(23), 568, <https://doi.org/10.21105/joss.00568>
- [230] Wickham, H., & Wickham, M. H. (2017). Package 'tidyr'. *Easily Tidy Data with 'spread()' and 'gather()' Functions*, CRAN, <https://cran.r-project.org/package=tidyr>, pristupljeno 16.08.2023.
- [231] Hudak, P. (1989). Conception, evolution, and application of functional programming languages. *ACM Computing Surveys (CSUR)*, 21(3), 359-411. <https://doi.org/10.1145/72551.72554>
- [232] Mailund, T. (2022). Functional Programming: purrr. In *R 4 Data Science Quick Reference: A Pocket Guide to APIs, Libraries, and Packages* (pp. 89-110). Berkeley, CA: Apress, CRAN, <https://cran.r-project.org/package=purrr>, pristupljeno 16.08.2023.
- [233] Wickham, H. (2010). stringr: modern, consistent string processing. *R Journal*, 2(2), 38, CRAN, <https://cran.r-project.org/package=stringr>, pristupljeno 16.08.2023.
- [234] Wickham, H., & Rstudio. (2023). forcats: Tools for Working with Categorical Variables (Factors), CRAN, <https://cran.r-project.org/package=forcats>, pristupljeno 16.08.2023.
- [235] Kandel, S., Heer, J., Plaisant, C., Kennedy, J., Van Ham, F., Riche, N. H., ... & Buono, P. (2011). Research directions in data wrangling: Visualizations and transformations for

- usable and credible data. *Information Visualization*, 10(4), 271-288. <https://doi.org/10.1177/1473871611415994>
- [236] Vlemminx, M. W., de Lau, H., Vullings, R., Peters, C. H., & Oei, S. G. (2015). Electrohysterography. A promising alternative for monitoring contractions. *Nederlands Tijdschrift Voor Geneeskunde*, 159, A8535-A8535.
- [237] Tanasković, I., & Miljković, N. (2023). A new algorithm for fetal heart rate detection: Fractional order calculus approach. *Medical Engineering & Physics*, 104007. <https://doi.org/10.1016/j.medengphy.2023.104007>
- [238] Noninvasive term, MedlinePlus, National Library of Medicine, <https://medlineplus.gov/ency/article/002269.htm>, pristupljeno 17.08.2023.
- [239] Invasive term, MedlinePlus, National Library of Medicine, <https://medlineplus.gov/ency/article/002384.htm>, pristupljeno 17.08.2023.
- [240] Suetens, P. (2017). *Fundamentals of medical imaging*. Cambridge University Press.
- [241] Wang, J., Bauer, J., Becker, M., Bente, P., Dasenbrock, L., Elbers, K., ... & Wolf, K. H. (2014). A novel approach for discovering human behavior patterns using unsupervised methods. *Zeitschrift für Gerontologie und Geriatrie*, 47(8). <https://doi.org/10.1007/s00391-014-0806-7>
- [242] Miljković, N., & Trifunović, D. (2014). Pulse rate assessment: Eulerian video magnification vs. electrocardiography recordings. In *12th Symposium on Neural Network Applications in Electrical Engineering (NEUREL)* (pp. 17-20). IEEE. <https://doi.org/10.1109/NEUREL.2014.7011447>
- [243] Wu, H. Y., Rubinstein, M., Shih, E., Guttag, J., Durand, F., & Freeman, W. (2012). Eulerian video magnification for revealing subtle changes in the world. *ACM transactions on graphics (TOG)*, 31(4), 1-8. <https://doi.org/10.1145/2185520.2185561>
- [244] Balakrishnan, G., Durand, F., & Guttag, J. (2013). Detecting pulse from head motions in video. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3430-3437). https://openaccess.thecvf.com/content_cvpr_2013/papers/Balakrishnan_Detecting_Pulse_from_2013_CVPR_paper.pdf, pristupljeno 17.08.2023.
- [245] de Looft, P., Duursma, R., Noordzij, M., Taylor, S., Jaques, N., Scheepers, F., ... & Koldijk, S. (2022). Wearables: an R package with accompanying shiny application for signal analysis of a wearable device targeted at clinicians and researchers. *Frontiers in Behavioral Neuroscience*, 16, 856544. <https://doi.org/10.3389/fnbeh.2022.856544>
- [246] Huckvale, K., Venkatesh, S., & Christensen, H. (2019). Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety. *NPJ Digital Medicine*, 2(1), 1-11. <https://doi.org/10.1038/s41746-019-0166-1>
- [247] Torous, J., Kiang, M. V., Lorme, J., & Onnela, J. P. (2016). New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. *JMIR Mental Health*, 3(2), e5165. <https://doi.org/10.2196%2Fmental.5165>
- [248] Miljković, N. (2018). IZAZOV 2017: Zbornik zadataka i odabranih studentskih rešenja. Elektrotehnički fakultet, Univerzitet u Beogradu, <https://doi.org/10.5281/zenodo.1334788>
- [249] Hecker, F. (1999). Setting up shop: The business of open-source software. *IEEE Software*, 16(1), 45-51. <https://doi.org/10.1109/52.744568>
- [250] European Commission. (2016). H2020 Programme: Guidelines on FAIR Data Management in Horizon 2020, Version 3.0, https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf, pristupljeno 19.08.2023.

- [251] Miljković, N. (2018). Digital biosignal processing with R. satRday Conference, R Consortium, Belgrade, Serbia, October 27, 2018, <https://belgrade2018.satrdays.org/> i <https://github.com/NadicaSm/satRday-Belgrade-2018>, pristupljeno 19.08.2023.
- [252] Milić, Lj. (Ed.). (2009). *Multirate Filtering for Digital Signal Processing: MATLAB Applications: MATLAB Applications*. IGI Global.
- [253] Kinesiology remembers professor David Winter, University of Waterloo, 2012, <https://uwaterloo.ca/health/news/kinesiology-remembers-professor-david-winter>, pristupljeno 21.08.2023.
- [254] sepower2. (2020). The Difference Between Multidisciplinary, Interdisciplinary, and Convergence Research, Research Development Office, <https://research.ncsu.edu/rdo/the-difference-between-multidisciplinary-interdisciplinary-and-convergence-research/>, pristupljeno 21.08.2023.
- [255] Morton, L. W., Eigenbrode, S. D., & Martin, T. A. (2015). Architectures of adaptive integration in large collaborative projects. *Ecology and Society*, 20(4). <https://www.jstor.org/stable/26270306>, pristupljeno 21.08.2023.
- [256] Pope, M. H. (2005). Giovanni Alfonso Borelli—the father of biomechanics. *Spine*, 30(20), 2350-2355.
- [257] Manolova, A. (2018). Chapter 1: History and rise of biomechanics, Sciences du sport, <https://www.sci-sport.com/en/theory/chapter-1-history-and-rise-of-biomechanics.php>, pristupljeno 21.08.2023.
- [258] Kursell J. (2009). N° 6 Physiology of the Piano, Max Planck Institute for the History of Science, <https://www.mpiwg-berlin.mpg.de/news/features/features-feature6>, pristupljeno 19.08.2023.
- [259] Kursell, J. (2006). Piano Mécanique and Piano Biologique: Nikolai Bernstein's Neurophysiological Study of Piano Touch. *Configurations*, 14(3), 245-273. <https://doi.org/10.1353/con.0.0015>
- [260] Prochazka, A., Mushahwar, V. K., & McCreery, D. B. (2001). Neural prostheses. *The Journal of Physiology*, 533(1), 99-109. <https://doi.org/10.1111/j.1469-7793.2001.0099b.x>
- [261] Oger T. (2012). Nikolai Bernstein. Courtesy of Andrei Smirnov – YouTube <https://youtu.be/yDxPjBqWuM>, pristupljeno 22.08.2023.
- [262] Rosenbaum, D. A., & Collyer, C. E. (Eds.). (1998). *Timing of behavior: Neural, psychological, and computational perspectives*. MIT Press.
- [263] Pejčić, N., Đurić-Jovičić, M., Miljković, N., Popović, D. B., & Petrović, V. (2016). Posture in dentists: Sitting vs. standing positions during dentistry work: An EMG study. *Srpski Arhiv za Celokupno Lekarstvo*, 144(3-4), 181-187. <http://dx.doi.org/10.2298/sarh1604181P>
- [264] Vieira, E. R., & Kumar, S. (2004). Working postures: a literature review. *Journal of Occupational Rehabilitation*, 14, 143-159. <https://doi.org/10.1023/B:JOOR.0000018330.46029.05>
- [265] De Luca, C. (2006). Electromyography. *Encyclopedia of medical devices and instrumentation*. <https://doi.org/10.1002/0471732877.emd097>
- [266] Popović, N. B. (2021). Methods for assessment of electrical activity of smooth muscles, doktorska disertacija, Univerzitet u Beogradu – Elektrotehnički fakultet, https://hdl.handle.net/21.15107/rcub_nardus_18615
- [267] Merlo, A., & Campanini, I. (2010). Technical aspects of surface electromyography for clinicians. *The Open Rehabilitation Journal*, 3(1). <http://dx.doi.org/10.2174/1874943701003010098>

- [268] Miljković, N., & Sodnik, J. (2023). Towards objective assesment of driving simulation sickness: Pros and cons of stomach electrical activity. *Proceedings of the 22nd Driving Simulation and Virtual Reality Conference and Exhibition*, pp. 81-88, Antibes, France.
- [269] Gruden, T., Stojmenova, K., Sodnik, J., & Jakus, G. (2019). Assessing drivers' physiological responses using consumer grade devices. *Applied Sciences*, 9(24), 5353. <https://doi.org/10.3390/app9245353>
- [270] Böttcher, S., Vieluf, S., Bruno, E., Joseph, B., Epitashvili, N., Biondi, A., ... & Loddenkemper, T. (2022). Data quality evaluation in wearable monitoring. *Scientific Reports*, 12(1), 21412. <https://doi.org/10.1038/s41598-022-25949-x>
- [271] Menghini, L., Gianfranchi, E., Cellini, N., Patron, E., Tagliabue, M., & Sarlo, M. (2019). Stressing the accuracy: Wrist-worn wearable sensor validation over different conditions. *Psychophysiology*, 56(11), e13441. <https://doi.org/10.1111/psyp.13441>
- [272] Boyer, M., Bouyer, L., Roy, J. S., & Campeau-Lecours, A. (2023). Reducing Noise, Artifacts and Interference in Single-Channel EMG Signals: A Review. *Sensors*, 23(6), 2927. <https://doi.org/10.3390/s23062927>
- [273] Tereshchenko, L. G., & Josephson, M. E. (2015). Frequency content and characteristics of ventricular conduction. *Journal of Electrocardiology*, 48(6), 933-937. <https://doi.org/10.1016%2Fj.jelectrocard.2015.08.034>
- [274] Miljković, N., Popović, N., Djordjević, O., Konstantinović, L., & Šekara, T. B. (2017). ECG artifact cancellation in surface EMG signals by fractional order calculus application. *Computer Methods and Programs in Biomedicine*, 140, 259-264. <https://doi.org/10.1016/j.cmpb.2016.12.017>
- [275] signal developers (2014). *signal: Signal processing*, CRAN, <http://r-forge.r-project.org/projects/signal/>, pristupljeno 14.09.2023.
- [276] Popović, M. V. (1999). Digitalna obrada signala, Nauka, Beograd.
- [277] Semmlow, J. L., & Griffel, B. (2014). Biosignal and Medical Image Processing, 3rd Edn. Hoboken.; Oppenheim, A.V., and R.W. Schafer, Discrete-Time Signal Processing, Prentice-Hall, 1989.
- [278] Gustafsson, F. (1996). Determining the initial states in forward-backward filtering. *IEEE Transactions on Signal Processing*, 44(4), 988-992. <https://doi.org/10.1109/78.492552>
- [279] Sollaci, L. B., & Pereira, M. G. (2004). The introduction, methods, results, and discussion (IMRAD) structure: A fifty-year survey. *Journal of the Medical Library Association*, 92(3), 364.
- [280] Wu, J. (2011). Improving the writing of research papers: IMRAD and beyond. *Landscape Ecology*, 26, 1345-1349. <https://doi.org/10.1007/s10980-011-9674-3>
- [281] Frigo, M., & Johnson, S. G. (1997). The Fastest Fourier Transform in the West. <https://dspace.mit.edu/bitstream/handle/1721.1/149858/MIT-LCS-TR-728.pdf?sequence=1&isAllowed=y>, pristupljeno 16.09.2023.
- [282] Mitra, S. K. (2001). *Digital signal processing: a computer-based approach*. McGraw-Hill Higher Education.
- [283] Milić, Lj., Dobrosavljević, Z., & Paunović, Đ. (1999). *Uvod u digitalnu obradu signala*. Elektrotehnički fakultet, Beograd.
- [284] Oppenheim, A.V., & R.W. Schafer. (1989). Discrete-Time Signal Processing, Prentice-Hall
- [285] Stojić, M. (1978). *Kontinualni sistemi automatskog upravljanja*, Nauka, Beograd
- [286] Rangayyan, R. M., & Krishnan, S. (2024). *Biomedical signal analysis*. John Wiley & Sons.

- [287] Mello, R. G., Oliveira, L. F., & Nadal, J. (2007). Digital Butterworth filter for subtracting noise from low magnitude surface electromyogram. *Computer Methods and Programs in Biomedicine*, 87(1), 28-35. <https://doi.org/10.1016/j.cmpb.2007.04.004>
- [288] Selesnick, I. W., & Burrus, C. S. (1998). Generalized digital Butterworth filter design. *IEEE Transactions on Signal Processing*, 46(6), 1688-1694. <https://doi.org/10.1109/78.678493>
- [289] Abhang, P. A., Gawali, B. W., & Mehrotra, S. C. (2016). *Introduction to EEG-and speech-based emotion recognition*. Academic Press.
- [290] Widmann, A., Schröger, E., & Maess, B. (2015). Digital filter design for electrophysiological data—a practical approach, *Journal of Neuroscience Methods*, 250, 34-46. <https://doi.org/10.1016/j.jneumeth.2014.08.002>
- [291] de Cheveigné, A., & Nelken, I. (2019). Filters: when, why, and how (not) to use them. *Neuron*, 102(2), 280-293. <https://doi.org/10.1016/j.neuron.2019.02.039>
- [292] Prutchi, D., & Norris, M. (2005). *Design and development of medical electronic instrumentation: A practical perspective of the design, construction, and test of medical devices*. John Wiley & Sons.
- [293] Beyerer, J., León, F. P., & Frese, C. (2015). *Machine vision: Automated visual inspection: Theory, practice and applications*. Springer.
- [294] Meng, M., She, Q., Gao, Y., & Luo, Z. (2010). EMG signals based gait phases recognition using hidden Markov models. In *The 2010 IEEE International Conference on Information and Automation* (pp. 852-856). IEEE. <https://doi.org/10.1109/ICINFA.2010.5512456>
- [295] Giggins, O. M., Persson, U. M., & Caulfield, B. (2013). Biofeedback in rehabilitation. *Journal of Neuroengineering and Rehabilitation*, 10, 1-11. <https://doi.org/10.1186/1743-0003-10-60>
- [296] Stuckey, S. J., Jacobs, A., & Goldfarb, J. (1986). EMG biofeedback training, relaxation training, and placebo for the relief of chronic back pain. *Perceptual and Motor Skills*, 63(3), 1023-1036. <https://doi.org/10.2466/pms.1986.63.3.1023>
- [297] Prinsloo, G. E., Rauch, H. L., & Derman, W. E. (2014). A brief review and clinical application of heart rate variability biofeedback in sports, exercise, and rehabilitation medicine. *The Physician and Sports Medicine*, 42(2), 88-99. <https://doi.org/10.3810/psm.2014.05.2061>
- [298] Clarys, J. P. (2000). Electromyography in sports and occupational settings: An update of its limits and possibilities. *Ergonomics*, 43(10), 1750-1762. <https://doi.org/10.1080/001401300750004159>
- [299] Reaz, M. B. I., Hussain, M. S., & Mohd-Yasin, F. (2006). Techniques of EMG signal analysis: detection, processing, classification and applications. *Biological Procedures Online*, 8, 11-35. <https://doi.org/10.1251%2Fbpo115>
- [300] Buchanan, W. J. (1997). 32 - File Management. In W. J. Buchanan (Ed.), *Software Development for Engineers* (pp. 513-522). Butterworth-Heinemann. <https://doi.org/https://doi.org/10.1016/B978-034070014-3/50077-3>
- [301] Smith, S. W. (1999). *The Scientist and Engineer's Guide to Digital Signal Processing*, California Technical Publishing, https://www.analog.com/en/resources/technical-books/scientist_engineers_guide.html, pristupljeno 24.02.2024
- [302] Teetor, P. (2011). *R cookbook: Proven recipes for data analysis, statistics, and graphics*, O'Reilly Media, Inc.
- [303] Li, C. (2022). *Time Series Data Analysis in Oceanography: Applications Using MATLAB*. Cambridge University Press.

- [304] Mendonça, T. R., Martins, C. H., Pinto, M. F., & Duque, C. A. (2015). Variable window length applied to a modified hanning filter for optimal amplitude estimation of power systems signals. In *2015 IEEE Power & Energy Society General Meeting* (pp. 1-5). IEEE. <https://doi.org/10.1109/PESGM.2015.7285859>
- [305] Mak, D. K. (2006). *Mathematical techniques in financial market trading*. World Scientific.
- [306] Harris, F. J. (1978). On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE*, 66(1), 51-83.
- [307] Kaiser, J., & Schafer, R. (1980). On the use of the I 0-sinh window for spectrum analysis. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28(1), 105-107.
- [308] Oppenheim, A. V. (1999). *Discrete-time signal processing*. Pearson Education India.
- [309] Lee, B. H., & Kuo, S. M. (2001). *Real-Time Digital Signal Processing: Implementations, Applications and Experiments with the TMS320C55X*. Hoboken, NJ, USA: Wiley.
- [310] Pejović, P. (2021). Nikvistov kriterijum stabilnosti, *Zenodo*. <https://doi.org/10.5281/zenodo.5142695>
- [311] Kelsey, R. M., Reiff, S., Wiens, S., Schneider, T. R., Mezzacappa, E. S., & Guethlein, W. (1998). The ensemble-averaged impedance cardiogram: An evaluation of scoring methods and interrater reliability. *Psychophysiology*, 35(3), 337-340. <https://doi.org/10.1017/S0048577298001310>
- [312] Hermens, H. J., Freriks, B., Disselhorst-Klug, C., & Rau, G. (2000). Development of recommendations for SEMG sensors and sensor placement procedures. *Journal of Electromyography and Kinesiology*, 10(5), 361-374. [https://doi.org/10.1016/S1050-6411\(00\)00027-4](https://doi.org/10.1016/S1050-6411(00)00027-4)
- [313] Hermens, H. J., Freriks, B., Merletti, R., Stegeman, D., Blok, J., Rau, G., ... & Hägg, G. (1999). European recommendations for surface electromyography. *Roessingh Research and Development*, 8(2), 13-54.
- [314] Merletti, R., Hermens, H., & Kadefors, R. (2001). European community projects on surface electromyography. In *2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Vol. 2, pp. 1119-1122). IEEE. <https://doi.org/10.1109/IEMBS.2001.1020387>
- [315] Peacock, B., Westers, T., Walsh, S., & Nicholson, K. (1981). Feedback and maximum voluntary contraction. *Ergonomics*, 24(3), 223-228. <https://doi.org/10.1080/00140138108559236>
- [316] Besomi, M., Hodges, P. W., Clancy, E. A., Van Dieën, J., Hug, F., Lowery, M., ... & Tucker, K. (2020). Consensus for experimental design in electromyography (CEDE) project: Amplitude normalization matrix. *Journal of Electromyography and Kinesiology*, 53, 102438. <https://doi.org/10.1016/j.jelekin.2020.102438>
- [317] Djordjevic, O., Konstantinović, L., Miljković, N., & Bijelić, G. (2015). Relationship between electromyographic signal amplitude and thickness change of the trunk muscles in patients with and without low back pain. *The Clinical Journal of Pain*, 31(10), 893-902. <https://www.doi.org/10.1097/AJP.0000000000000179>
- [318] De Luca, C. J. (1997). The use of surface electromyography in biomechanics. *Journal of Applied Biomechanics*, 13(2), 135-163. <https://doi.org/10.1123/jab.13.2.135>
- [319] Halaki, M., & Ginn, K. (2012). Normalization of EMG signals: to normalize or not to normalize and what to normalize to. *Computational intelligence in electromyography analysis-a perspective on current applications and future challenges*, 10, 49957.

- [320] Knutson, L. M., Soderberg, G. L., Ballantyne, B. T., & Clarke, W. R. (1994). A study of various normalization procedures for within day electromyographic data. *Journal of Electromyography and Kinesiology*, 4(1), 47-59. [https://doi.org/10.1016/1050-6411\(94\)90026-4](https://doi.org/10.1016/1050-6411(94)90026-4)
- [321] Cauchi, R., Camilleri, K. P., Saliba, M. A., & Attard, J. (2021). Isometric and anisometric contraction relationships with surface electromyography. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (pp. 6368-6371). IEEE. <https://doi.org/10.1109/embc46164.2021.9630490>
- [322] Mohseni, O., Berry, A., Schumacher, C., Seyfarth, A., Vallery, H., & Sharbafi, M. A. (2024). Mediolateral Upper-body Gyroscopic Moment Perturbations During Walking: Exploring Muscle Activation Patterns and Control Strategies. <https://doi.org/10.21203/rs.3.rs-3881620/v1>
- [323] van Soest, A. J., & Bobbert, M. F. (1993). The contribution of muscle properties in the control of explosive movements. *Biological Cybernetics*, 69(3), 195-204. <https://doi.org/10.1007/BF00198959>
- [324] Sousa, A. S., & Tavares, J. M. R. (2012). Surface electromyographic amplitude normalization methods: A review. *Electromyography: New developments, procedures and applications*.
- [325] Winter, D. A., & Yack, H. J. (1987). EMG profiles during normal human walking: stride-to-stride and inter-subject variability. *Electroencephalography and Clinical Neurophysiology*, 67(5), 402-411. [https://doi.org/10.1016/0013-4694\(87\)90003-4](https://doi.org/10.1016/0013-4694(87)90003-4)
- [326] Shiavi, R., Frigo, C., & Pedotti, A. (1998). Electromyographic signals during gait: criteria for envelope filtering and number of strides. *Medical & Biological Engineering & Computing*, 36, 171-178. <https://doi.org/10.1007/BF02510739>
- [327] Cavanagh, P. R. (1974). Electromyography: its use and misuse in physical education. *Journal of Health, Physical Education, Recreation*, 45(5), 61-64. <https://doi.org/10.1080/00221473.1974.10614161>
- [328] Kim, D., & Oh, H. S. (2009). EMD: A Package for Empirical Mode Decomposition and Hilbert Spectrum. *R Journal*, 1(1), 40. <https://doi.org/10.32614/RJ-2009-002>
- [329] Strang, G., & Nguyen, T. (1996). *Wavelets and filter banks*. SIAM.
- [330] Lei, Y., Lin, J., He, Z., & Zuo, M. J. (2013). A review on empirical mode decomposition in fault diagnosis of rotating machinery. *Mechanical Systems and Signal Processing*, 35(1-2), 108-126. <https://doi.org/10.1016/j.ymssp.2012.09.015>
- [331] Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., ... & Liu, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1971), 903-995. <https://doi.org/10.1098/rspa.1998.0193>
- [332] Zeng, K., & He, M. X. (2004). A simple boundary process technique for empirical mode decomposition. In *IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium* (Vol. 6, pp. 4258-4261). IEEE. <https://doi.org/10.1109/IGARSS.2004.1370076>
- [333] Kieliba, P., Tropea, P., Pirondini, E., Coscia, M., Micera, S., & Artoni, F. (2018). How are muscle synergies affected by electromyography pre-processing?. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(4), 882-893. <https://doi.org/10.1109/TNSRE.2018.2810859>
- [334] Jöreskog, K. G. (2012). Factor analysis as an errors-in-variables model. In *Principals of modern psychological measurement* (pp. 185-196). Routledge.

- [335] De Vos, M., Deburchgraeve, W., Cherian, P. J., Matic, V., Swarte, R. M., Govaert, P., ... & Van Huffel, S. (2011). Automated artifact removal as preprocessing refines neonatal seizure detection. *Clinical Neurophysiology*, 122(12), 2345-2354. <https://doi.org/10.1016/j.clinph.2011.04.026>
- [336] Dirlich, G., Vogl, L., Plaschke, M., & Strian, F. (1997). Cardiac field effects on the EEG. *Electroencephalography and Clinical Neurophysiology*, 102(4), 307-315. [https://doi.org/10.1016/S0013-4694\(96\)96506-2](https://doi.org/10.1016/S0013-4694(96)96506-2)
- [337] Miljković, N., Matic, V., Van Huffel, S., & Popović, M. B. (2010). Independent Component Analysis (ICA) methods for neonatal EEG artifact extraction: Sensitivity to variation of artifact properties. In *10th Symposium on Neural Network Applications in Electrical Engineering* (pp. 19-21). IEEE. <https://doi.org/10.1109/NEUREL.2010.5644041>
- [338] Atzori, M., Gijsberts, A., Heynen, S., Hager, A. G. M., Deriaz, O., Van Der Smagt, P., ... & Müller, H. (2012). Building the Ninapro database: A resource for the biorobotics community. In *2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)* (pp. 1258-1265). IEEE. <https://doi.org/10.1109/BioRob.2012.6290287>
- [339] Isaković, M. S., Miljković, N., & Popović, M. B. (2014). Classifying sEMG-based hand movements by means of principal component analysis. In *2014 22nd Telecommunications Forum Telfor (TELFOR)* (pp. 545-548). IEEE. <https://doi.org/10.1109/TELFOR.2014.7034467>
- [340] Miljković, N., & Isaković, M. S. (2021). Effect of the sEMG electrode (re)placement and feature set size on the hand movement recognition. *Biomedical Signal Processing and Control*, 64, 102292. <https://doi.org/10.1016/j.bspc.2020.102292>
- [341] Ivanenko, Y. P., Poppele, R. E., & Lacquaniti, F. (2006). Spinal cord maps of spatiotemporal alpha-motoneuron activation in humans walking at different speeds. *Journal of Neurophysiology*, 95(2), 602-618. <https://doi.org/10.1152/jn.00767.2005>
- [342] Ivanenko, Y. P., Poppele, R. E., & Lacquaniti, F. (2004). Five basic muscle activation patterns account for muscle activity during human locomotion. *The Journal of Physiology*, 556(1), 267-282. <https://doi.org/10.1113/jphysiol.2003.057174>
- [343] Borchers, H. W., & Borchers, M. H. W. (2019). Package ‘pracma’. *Practical numerical math functions, version, 2*(5), CRAN, <https://cran.r-project.org/package=pracma>, pristupljeno 12.03.2024.
- [344] Staude, G., Flachenecker, C., Daumer, M., & Wolf, W. (2001). Onset detection in surface electromyographic signals: a systematic comparison of methods. *EURASIP Journal on Advances in Signal Processing*, 2001, 1-15. <https://doi.org/10.1155/S1110865701000191>
- [345] Carvalho, C. R., Fernández, J. M., Del-Ama, A. J., Oliveira Barroso, F., & Moreno, J. C. (2023). Review of electromyography onset detection methods for real-time control of robotic exoskeletons. *Journal of NeuroEngineering and Rehabilitation*, 20(1), 141. <https://doi.org/10.1186/s12984-023-01268-8>
- [346] Voet, N. B., Saris, C. G., Thijssen, D. H., Bastiaans, V., Sluijs, D. E., & Janssen, M. M. (2022). Surface electromyography thresholds as a measure for performance fatigability during incremental cycling in patients with neuromuscular disorders. *Frontiers in Physiology*, 13, 821584. <https://doi.org/10.3389/fphys.2022.821584>
- [347] Roever, C. (2015). *bspec: Bayesian spectral inference*. R package version 1.4., CRAN, <https://cran.r-project.org/package=bspec>, pristupljeno 01.06.2024.

- [348] Roever, C., Meyer, R., & Christensen, N. (2011). Modelling coloured residual noise in gravitational-wave signal processing, *Classical and Quantum Gravity*, 28(1), 015010, <http://arxiv.org/abs/0804.3853>
- [349] Welch, P. (1967). The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15(2), 70-73. <https://doi.org/10.1109/TAU.1967.1161901>
- [350] Miljković, N., Popović, N. B., Prodanov, M., & Sodnik, J. (2019). Assessment of sickness in virtual environments. In *Proceedings of the 9th International Conference on Information Society and Technology (ICIST 2019)*. <https://www.eventiotic.com/eventiotic/files/Papers/URL/f412a358-82d8-46aa-b0c9-6ad30de5890d.pdf>, pristupljeno 01.06.2024.
- [351] Popović, N. B., Miljković, N., Stojmenova, K., Jakus, G., Prodanov, M., & Sodnik, J. (2019). Lessons learned: Gastric motility assessment during driving simulation. *Sensors*, 19(14), 3175. <https://doi.org/10.3390/s19143175>
- [352] Fele-Žorž, G., Kavšek, G., Novak-Antolič, Ž., & Jager, F. (2008). A comparison of various linear and non-linear signal processing techniques to separate uterine EMG records of term and pre-term delivery groups. *Medical & Biological Engineering & Computing*, 46, 911-922. <https://doi.org/10.1007/s11517-008-0350-y>
- [353] Inbar, G. F., Paiss, O., Allin, J., & Kranz, H. (1986). Monitoring surface EMG spectral changes by the zero crossing rate. *Medical & Biological Engineering & Computing*, 24, 10-18. <https://doi.org/10.1007/BF02441600>
- [354] Phinyomark, A., Thongpanja, S., Hu, H., Phukpattaranont, P., & Limsakul, C. (2012). The usefulness of mean and median frequencies in electromyography analysis. *Computational intelligence in electromyography analysis - A perspective on current applications and future challenges*, 23, 195-220. <https://www.doi.org/10.5772/50639>
- [355] Boljanić, T., Miljković, N., Lazarevic, L. B., Knezevic, G., & Milašinović, G. (2022). Relationship between electrocardiogram-based features and personality traits: Machine learning approach. *Annals of Noninvasive Electrocardiology*, 27(1), e12919. <https://doi.org/10.1111/anec.12919>
- [356] Nyström, M., & Holmqvist, K. (2010). An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behavior Research Methods*, 42(1), 188-204. <https://doi.org/10.3758/BRM.42.1.188>
- [357] Bahill, A. T., Brockenbrough, A., & Troost, B. T. (1981). Variability and development of a normative data base for saccadic eye movements. *Investigative Ophthalmology & Visual Science*, 21(1), 116-125.
- [358] Cassin, B., Solomon, S., & Rubin, M. L. (1990). Dictionary of Eye Terminology. Gainesville. Florida: Triad Publishing.
- [359] Godwin, H. J., Hout, M. C., Alexdóttir, K. J., Walenchok, S. C., & Barnhart, A. S. (2021). Avoiding potential pitfalls in visual search and eye-movement experiments: A tutorial review. *Attention, Perception, & Psychophysics*, 83(7), 2753-2783. <https://doi.org/10.3758/s13414-021-02326-w>
- [360] Holmqvist, K., Nyström, M., & Mulvey, F. (2012). Eye tracker data quality: What it is and how to measure it. In *Proceedings of the symposium on eye tracking research and applications* (pp. 45-52). <https://doi.org/10.1145/2168556.2168563>
- [361] Lazarević, M. P., Rapaić, M. R., Šekara, T. B., Mladenov, V., & Mastorakis, N. (2014). Introduction to fractional calculus with brief historical background. *Advanced topics on applications of fractional calculus on control problems, system stability and modeling*, 3-16.

- [362] Tanasković, I., & Miljković, N. (2022). Fractional order calculus for fetal pulse detection. In *2022 21st International Symposium INFOTEH-JAHORINA (INFOTEH)* (pp. 1-5). IEEE. <https://doi.org/10.1109/INFOTEH53737.2022.9751327>
- [363] Loverro, A. (2004). Fractional calculus: history, definitions and applications for the engineer. *Rapport technique, Univeristy of Notre Dame: Department of Aerospace and Mechanical Engineering*, 1-28.
- [364] Oldham, K., & Spanier, J. (1974). *The fractional calculus theory and applications of differentiation and integration to arbitrary order*. Elsevier.
- [365] Фихтенгольц, Г. М. (1966). *Курс дифференциального и интегрального исчисления*. Рипол Классик.
- [366] Kilbas, A. A., Srivastava, H. M., & Trujillo, J. J. (2006). *Theory and applications of fractional differential equations* (Vol. 204). Elsevier.
- [367] Хуе, Д. (2024). FOTF Toolbox, MATLAB Central File Exchange, <https://www.mathworks.com/matlabcentral/fileexchange/60874-fotf-toolbox>, pristupljeno 01.06.2024.
- [368] Хуе, Д. Y., & Bai, L. (2022). *Fractional Calculus: Numerical Algorithms and Implementations*. Beijing: Tsinghua University Press.
- [369] Popović, N. B., Miljković, N., Šekara, T. (2018). Two applications of fractional calculus on biosignal processing. *Konferencija Merno-informacione tehnologije 2018 (MIT 2018)*, pp. 1 - 3, Faculty of Technical Sciences, Novi Sad, Republic of Serbia.
- [370] Antić, M., Popović, N. B., Milosavljević, N., Dubljević, O., Bjeogojević, B., & Miljković, N. (2020). CardioPRINT: Individual features hidden in electrocardiogram and impedance-cardiogram. *Empirical Studies in Psychology*, 13. https://reff.f.bg.ac.rs/bitstream/handle/123456789/4340/EIP2020_conf_proceedings.pdf?sequence=1&isAllowed=y#page=13, pristupljeno 01.06.2024.
- [371] Tanasković, I., Lazarević, L., Knezevic, G., Milosavljevic, N., Dubljević, O., Bjeogojevic, B., & Miljković, N. (2023). CardioPRINT: Biometric identification based on the individual characteristics derived from cardiogram, <https://doi.org/10.31234/osf.io/bau7j>
- [372] Koelsch, S., Remppis, A., Sammler, D., Jentschke, S., Mietchen, D., Fritz, T., ... & Siebel, W. A. (2007). A cardiac signature of emotionality. *European Journal of Neuroscience*, 26(11), 3328-3338. <https://doi.org/10.1111/j.1460-9568.2007.05889.x>
- [373] Bjeogojević, B., Milosavljević, N., Dubljević, O., Purić, D., & Knežević, G. (2020). In pursuit of objectivity: Physiological measures as a means of emotion induction procedure validation. *Empirical Studies in Psychology*, pp. 17-19, http://empirijskaistrazivanja.org/wp-content/uploads/2021/04/EIP2020_conf_proceedings.pdf#page=17, pristupljeno 01.06.2024.
- [374] Boljanić, T., Miljković, N., Lazarević B. Lj., Knežević, G., & Milašinović, G. (2021). Surface electrocardiogram (ECG) dataset recorded during relaxation in 70 healthy subjects (Version 1) [Data set], *Zenodo*. <https://doi.org/10.5281/zenodo.5599239>
- [375] Lemon, J. (2006). Plotrix: a package in the red light district of R. *R-News*, 6(4), 8-12.
- [376] Popović, N., Miljković, N., Djordjevic, O., & Šekara, T. B. (2016). Artifact cancellation using median filter, moving average filter, and fractional derivatives in biomedical signals. *Proceedings of the ICFDA*, 68(2), 150-161.
- [377] Paunović, M., Čosić, M., Radovanović, N., & Miljković, N. (2024). New Asymmetry Measure of a Poincaré Plot for Heart Rate Assessment, *International Conference Mathematical Analysis and Applications in Science and Engineering ICMA2SC'24*, in print, Porto, Portugal.

- [378] Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379-423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- [379] Delgado-Bonal, A. (2019). Quantifying the randomness of the stock markets. *Scientific Reports*, 9(1), 12761. <https://doi.org/10.1038/s41598-019-49320-9>
- [380] Bajić, D., Škorić, T., & Japundžić-Žigon, N. (2018). Unakrsna entropija uzoraka transformisanih kardiovaskularnih podataka: pouzdanost procene, 17th *International Symposium INFOTEH-JAHORINA 2018*.
- [381] Šolaja, N., Miljković, N., & Sodnik, J. (2022). Driving activity assessment using accelerometer data. *The 28th International Electrotechnical and Computer Science Conference (ERK)*, Portorož, Slovenia. <https://doi.org/10.5281/zenodo.6036290>
- [382] Jakus, G., Sodnik, J., & Miljković, N. (2022). Electrogastragram-derived features for automated sickness detection in driving simulator. *Sensors*, 22(22), 8616. <https://doi.org/10.3390/s22228616>
- [383] Radovanović, N. N., Pavlović, S. U., Milašinović, G., Kirčanski, B., & Platiša, M. M. (2018). Bidirectional cardio-respiratory interactions in heart failure. *Frontiers in Physiology*, 9, 331043. <https://doi.org/10.3389/fphys.2018.00165>
- [384] Schulz, S., Adochiei, F. C., Edu, I. R., Schroeder, R., Costin, H., Bär, K. J., & Voss, A. (2013). Cardiovascular and cardiorespiratory coupling analyses: A review. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1997), 20120191. <https://doi.org/10.1098/rsta.2012.0191>
- [385] Satti, R., Abid, N. U. H., Bottaro, M., De Rui, M., Garrido, M., Raoufy, M. R., ... & Mani, A. R. (2019). The application of the extended Poincaré plot in the analysis of physiological variabilities, *Frontiers in Physiology*, 10, 116. <https://doi.org/10.3389/fphys.2019.00116>
- [386] Mancini, R., & Carter, B. (2009). Op amps for everyone. Texas Instruments.
- [387] Chen, R. J., Lu, M. Y., Chen, T. Y., Williamson, D. F., & Mahmood, F. (2021). Synthetic data in machine learning for medicine and healthcare. *Nature Biomedical Engineering*, 5(6), 493-497. <https://doi.org/10.1038/s41551-021-00751-8>
- [388] Díez-Pastor, J. F., Rodríguez, J. J., García-Osorio, C. I., & Kuncheva, L. I. (2015). Diversity techniques improve the performance of the best imbalance learning ensembles, *Information Sciences*, 325, 98-117. <https://doi.org/10.1016/j.ins.2015.07.025>
- [389] Kligfield, P., Gettes, L. S., Bailey, J. J., Childers, R., Deal, B. J., Hancock, E. W., ... & Wagner, G. S. (2007). Recommendations for the standardization and interpretation of the electrocardiogram: part I: the electrocardiogram and its technology: a scientific statement from the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society endorsed by the International Society for Computerized Electrocardiology. *Circulation*, 115(10), 1306-1324.
- [390] Pietraszek, S., & Komorowski, D. (2009). The simultaneous recording and analysis both EGG and HRV signals. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 396-399). IEEE. <https://doi.org/10.1109/IEMBS.2009.5333455>
- [391] Popović, N. B., Miljković, N., & Šekara, T. B. (2020). Electrogastragram and electrocardiogram interference: Application of fractional order calculus and Savitzky-Golay filter for biosignals segregation. In *2020 19th International Symposium INFOTEH-JAHORINA (INFOTEH)* (pp. 1-5). IEEE. <https://doi.org/10.1109/INFOTEH48170.2020.9066278>

- [392] Cimeša, Lj., Popović, N., Miljković, N., & Šekara, T. B. (2017). Heart rate detection: Fractional approach and empirical mode decomposition. In *2017 25th Telecommunication Forum (TELFOR)* (pp. 1-4). IEEE. <https://doi.org/10.1109/TELFOR.2017.8249358>
- [393] Jovanović, N., Popović, N. B., & Miljković, N. (2021). Empirical mode decomposition for automatic artifact elimination in electrogastrogram. In *2021 20th International Symposium INFOTEH-JAHORINA (INFOTEH)* (pp. 1-6). IEEE. <https://doi.org/10.1109/INFOTEH51037.2021.9400683>
- [394] Abutaleb, A. S., & Soliman, H. (2023). The Estimation of the Instantaneous Amplitudes and the Instantaneous Frequencies of the EEG when both are Stochastic Processes. *Biomedical Journal of Scientific & Technical Research*, 48(2), 39494-39502. <http://dx.doi.org/10.26717/BJSTR.2023.48.007621>
- [395] Thakor, N. V., & Tong, S. (2004). Advances in quantitative electroencephalogram analysis methods. *Annual Review of Biomedical Engineering*, 6, 453-495. <https://doi.org/10.1146/annurev.bioeng.5.040202.121601>
- [396] Barzegaran, E., Bosse, S., Kohler, P. J., & Norcia, A. M. (2019). EEGSourceSim: A framework for realistic simulation of EEG scalp data using MRI-based forward models and biologically plausible signals and noise. *Journal of Neuroscience Methods*, 328, 108377. <https://doi.org/10.1016/j.jneumeth.2019.108377>
- [397] Reproducibility definitions, <https://the-turing-way.netlify.app/reproducible-research/overview/overview-definitions.html#table-of-definitions-for-reproducibility>, pristupljeno 07.02.2024.
- [398] L'Ecuyer, P. (1999). Good parameters and implementations for combined multiple recursive random number generators, *Operations Research*, 47, 159–164. <https://doi.org/10.1287/opre.47.1.159>.
- [399] Matsumoto, M. & Nishimura, T. (1998). Mersenne Twister: A 623-dimensionally equidistributed uniform pseudo-random number generator, *ACM Transactions on Modeling and Computer Simulation*, 8, 3–30. <https://doi.org/10.1145/272991.272995>
- [400] Hariton, E., & Locascio, J. J. (2018). Randomised controlled trials—the gold standard for effectiveness research. *BJOG: An International Journal of Obstetrics and Gynaecology*, 125(13), 1716. <https://doi.org/10.1111%2F1471-0528.15199>
- [401] Sibbald, B., & Roland, M. (1998). Understanding controlled trials. Why are randomised controlled trials important?. *BMJ: British Medical Journal*, 316(7126), 201. <https://doi.org/10.1136/bmj.316.7126.201>
- [402] Ruha, A., Sallinen, S., & Nissila, S. (1997). A real-time microprocessor QRS detector system with a 1-ms timing accuracy for the measurement of ambulatory HRV. *IEEE Transactions on Biomedical Engineering*, 44(3), 159-167. <https://doi.org/10.1109/10.554762>
- [403] Farago, E., & Chan, A. D. (2021). Motion artifact synthesis for research in biomedical signal quality analysis. *Biomedical Signal Processing and Control*, 68, 102611. <https://doi.org/10.1016/j.bspc.2021.102611>
- [404] Hartmann, W. M. (2004). *Signals, sound, and sensation*. Springer Science & Business Media.
- [405] Szendro, P., Vincze, G., & Szasz, A. (2001). Pink-noise behaviour of biosystems. *European Biophysics Journal*, 30, 227-231. <https://doi.org/10.1007/s002490100143>
- [406] Barry, R. J., & De Blasio, F. M. (2021). Characterizing pink and white noise in the human electroencephalogram. *Journal of Neural Engineering*, 18(3), 034001. <https://doi.org/10.1088/1741-2552/abe399>

- [407] Ligges, U., Krey, S., Mersmann, O., & Schnackenberg, S. (2023). *tuneR: Analysis of Music and Speech*. <https://CRAN.R-project.org/package=tuneR>, pristupljeno 07.02.2024.
- [408] Becker, R. A., Chambers, J. M., & Wilks, A. R. (1988). *The New S Language*. Wadsworth & Brooks/Cole.
- [409] Guerrero, J. A., & Macías-Díaz, J. E. (2018). biosignalEMG: tools for electromyogram signals (EMG) analysis. *R package version*, 2(0), CRAN, <https://cran.r-project.org/package=biosignalEMG>, pristupljeno 07.02.2024.
- [410] Guerrero, J. A., & Macías-Díaz, J. E. (2019). A package for the computational analysis of complex biophysical signals. *International Journal of Modern Physics C*, 30(01), 1950005. <https://doi.org/10.1142/S0129183119500050>
- [411] Thongpanja, S., Phinyomark, A., Quaine, F., Laurillau, Y., Limsakul, C., & Phukpattaranont, P. (2016). Probability density functions of stationary surface EMG signals in noisy environments. *IEEE Transactions on Instrumentation and Measurement*, 65(7), 1547-1557. <https://doi.org/10.1109/TIM.2016.2534378>
- [412] Muthuraman, M., Galka, A., Deuschl, G., Heute, U., & Raethjen, J. (2010). Dynamical correlation of non-stationary signals in time domain—A comparative study. *Biomedical Signal Processing and Control*, 5(3), 205-213. <https://doi.org/10.1016/j.bspc.2010.02.006>
- [413] Semmlow, J. (2017). *Circuits, signals, and systems for bioengineers: A MatLab-based introduction*. Academic Press.
- [414] Kumar, S., & Mitia, A. (1996). *Electromyography in ergonomics*, Taylor & Francis.
- [415] Fischer, H. (2011). *A history of the central limit theorem: From classical to modern probability theory* (Vol. 4). New York: Springer.
- [416] Auguie, B., & Antonov, A. (2017). gridExtra: Miscellaneous Functions for “Grid” Graphics, CRAN, <https://cran.r-project.org/package=gridExtra>, pristupljeno 16.02.2024.
- [417] Wilk, M. B., & Gnanadesikan, R. (1968). Probability plotting methods for the analysis for the analysis of data. *Biometrika*, 55(1), 1-17. <https://doi.org/10.1093/biomet/55.1.1>
- [418] Pleil, J. D. (2016). QQ-plots for assessing distributions of biomarker measurements and generating defensible summary statistics. *Journal of Breath Research*, 10(3), 035001. <https://doi.org/10.1088/1752-7155/10/3/035001>
- [419] Rissanen, S., Kankaanpää, M., Tarvainen, M. P., Nuutinen, J., Tarkka, I. M., Airaksinen, O., & Karjalainen, P. A. (2007). Analysis of surface EMG signal morphology in Parkinson's disease. *Physiological Measurement*, 28(12), 1507. <https://doi.org/10.1088/0967-3334/28/12/005>
- [420] Zardoshti-Kermani, M., Wheeler, B. C., Badie, K., & Hashemi, R. M. (1995). EMG feature evaluation for movement control of upper extremity prostheses. *IEEE Transactions on Rehabilitation Engineering*, 3(4), 324-333. <https://doi.org/10.1109/86.481972>
- [421] Esquivel-Frausto, M. E., Guerrero, J. A., & Macías-Díaz, J. E. (2010). Activity pattern detection in electroneurographic and electromyogram signals through a heteroscedastic change-point method. *Mathematical Biosciences*, 224(2), 109-117. <https://doi.org/10.1016/j.mbs.2010.01.001>
- [422] Merletti, R., & Cerone, G. L. (2020). Tutorial. Surface EMG detection, conditioning and pre-processing: Best practices. *Journal of Electromyography and Kinesiology*, 54, 102440. <https://doi.org/10.1016/j.jelekin.2020.102440>
- [423] Guerrero, J. A., & Macías-Díaz, J. E. (2014). A computational method for the detection of activation/deactivation patterns in biological signals with three levels of electric

- intensity. *Mathematical Biosciences*, 248, 117-127. <https://doi.org/10.1016/j.mbs.2013.12.010>
- [424] Jovanović, J. (2019). Ispitivanje zavisnosti elektromiografskog signala i generisane sile, master rad, Univerzitet u Beogradu – Elektrotehnički fakultet.
- [425] Yang, D., Zhang, H., Gu, Y., & Liu, H. (2017). Accurate EMG onset detection in pathological, weak and noisy myoelectric signals. *Biomedical Signal Processing and Control*, 33, 306-315. <https://doi.org/10.1016/j.bspc.2016.12.014>
- [426] Bairagi, V. K., & Kanwade, A. B. (2021). Novel slope-based onset detection algorithm for electromyographical signals. *International Journal of Biomedical Engineering and Technology*, 36(4), 311-317. <https://doi.org/10.1504/IJBET.2021.116992>
- [427] Borg, F. (2011). Analyzing biosignals using the R freeware (open source) tool. *arXiv preprint arXiv:1103.3624*. <https://doi.org/10.48550/arXiv.1103.3624>
- [428] Tešić, L., Bondžulić, B., Andrić, M., & Pavlović, B. (2017). An experimental study on the phase importance in digital processing of speech signal. *Acta Polytechnica Hungarica*, 14(8), 197-213.
- [429] Oppenheim, A. V., & Lim, J. S. (1981). The importance of phase in signals. *Proceedings of the IEEE*, 69(5), 529-541. <https://doi.org/10.1109/PROC.1981.12022>
- [430] McSharry, P. E., Clifford, G. D., Tarassenko, L., & Smith, L. A. (2003). A dynamical model for generating synthetic electrocardiogram signals. *IEEE Transactions on Biomedical Engineering*, 50(3), 289-294. <https://doi.org/10.1109/TBME.2003.808805>
- [431] Paraskevas, I. (2005). *Phase as a feature extraction tool for audio classification and signal localisation*. University of Surrey (United Kingdom).
- [432] Zhemchuzhkina, T. V., Zlepko, S. M., Nosova, T. V., Semenets, V. V., Kirichek, O. V., Maciejewski, M., & Ormanbekova, A. (2019). Application of EMG-signal phase portraits for differentiation of musculoskeletal system diseases. In *Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments 2019* (Vol. 11176, pp. 906-911). SPIE.
- [433] Tukey, J. W. (1977). *Exploratory data analysis* (Vol. 2, pp. 131-160).
- [434] Tufte, E. R. (2001). *The visual display of quantitative information* (Vol. 2, p. 9). Cheshire, CT: Graphics press.
- [435] Zhao, S., Guo, Y., Sheng, Q., & Shyr, Y. (2014). Advanced heat map and clustering analysis using heatmap3. *BioMed Research International*. <https://doi.org/10.1155/2014/986048>
- [436] Irizarry, R. A., & Love, M. I. (2016). *Data analysis for the life sciences with R*. Chapman and Hall/CRC.
- [437] Eldridge, J., Richley, D., Baxter, S., Blackman, S., Cathal, J. B., Brown, C., ... & Ross, C. (2014). Recording a standard 12-lead electrocardiogram. An approved methodology by the Society of Cardiological Science and Technology (SCST). Clinical guidelines by consensus, <https://pure.ulster.ac.uk/en/publications/recording-a-standard-12-lead-electrocardiogram-an-approved-method-3>, pristupljeno 08.08.2023.
- [438] Pfurtscheller, G., & Aranibar, A. (1979). Evaluation of event-related desynchronization (ERD) preceding and following voluntary self-paced movement. *Electroencephalography and Clinical Neurophysiology*, 46(2), 138-146. [https://doi.org/10.1016/0013-4694\(79\)90063-4](https://doi.org/10.1016/0013-4694(79)90063-4)
- [439] Naït-Ali, A. (Ed.). (2009). *Advanced biosignal processing*. Springer Science & Business Media.
- [440] Despotović, S. Z., & Ćosić, M. (2022). The Morphological Analysis of the Collagen Fiber Straightness in the Healthy Uninvolved Human Colon Mucosa Away from the Cancer. *Frontiers in Physics*, 10, 915644. <https://doi.org/10.3389/fphy.2022.915644>

- [441] Yoon, D., Jang, J. H., Choi, B. J., Kim, T. Y., & Han, C. H. (2020). Discovering hidden information in biosignals from patients using artificial intelligence. *Korean Journal of Anesthesiology*, 73(4), 275. <https://doi.org/10.4097%2Fkja.19475>
- [442] Johnson, A. E., Pollard, T. J., Shen, L., Lehman, L. W. H., Feng, M., Ghassemi, M., ... & Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. *Scientific Data*, 3(1), 1-9. <https://doi.org/10.1038/sdata.2016.35>
- [443] Swenson, K. (2005). *Medical women and Victorian fiction*. University of Missouri Press.
- [444] Corbett, J. (2001). Charles Joseph Minard, Mapping Napoleon's March, 1861. CSISS Classics.
- [445] Miljković, N. (2019). Programski jezik R i njegova primena u nauci o podacima, LiBRE! Magazin, članak u 4 dela.
- [446] Petrović, M. (2018). Posters of historical infographs and telegraph law. Primena slobodnog softvera i otvorenog hardvera (in English "Applicaton of Free Software and Open Hardware") (PSSOH), Belgrade, Serbia, Zenodo. <https://doi.org/10.5281/zenodo.1473009>
- [447] Katsnelson, A. (2021). Fixing figures for colour blindness. *Nature*, 598: 224-225, <https://doi.org/10.1038/d41586-021-02696-z>
- [448] Robinson, A. H. (1967). The thematic maps of Charles Joseph Minard. <https://doi.org/10.1080/03085696708592302>
- [449] Friendly, M. (2002). Visions and re-visions of Charles Joseph Minard. *Journal of Educational and Behavioral Statistics*, 27(1), 31-51. <https://www.jstor.org/stable/3648145>
- [450] Cox, A., Carter, S., Quealy, K., & Schoenfeld, A. (2009). For the unemployed, the day stacks up differently. *New York Times*, <https://www.nytimes.com/2009/08/02/business/02metrics.html>, pristupljeno 02.05.2024.
- [451] Miljković, N. (2019). Biosignals for assessment of cybersickness, Zenodo. <https://doi.org/10.5281/zenodo.2631273>
- [452] Hirsch, E. D., Kett, J. F., Trefil, J., & Trefil, J. S. (2002). *The new dictionary of cultural literacy*. Houghton Mifflin Harcourt.
- [453] Irizzary, R. (2012). Statistics for Genomics: Useful plots and bad plots, <https://www.youtube.com/watch?v=46-t2jOYsyY>, pristupljeno 02.05.2024.
- [454] Birkett, A. (2023). The complete guide to heat maps: How to create and use them, CXL Blog, <https://cxl.com/blog/heat-maps/>, pristupljeno 02.05.2024.
- [455] Weissgerber, T. L., Milic, N. M., Winham, S. J., & Garovic, V. D. (2015). Beyond bar and line graphs: Time for a new data presentation paradigm. *PLoS Biology*, 13(4), e1002128. <https://doi.org/10.1371/journal.pbio.1002128>
- [456] C. Li. (2024). cxli233/FriendsDontLetFriends: FriendsDontLetFriends (v6.2) (v6.2), Zenodo. <https://doi.org/10.5281/zenodo.10802096>
- [457] Freeman, M., & Ross, J. (2018). *Data Science Foundations Tools and Techniques*. Addison-Wesley Professional.
- [458] Murrel, P. (2020). Rgraphics: Data and functions fro the book R Graphics, Third Edition, CRAN, <https://cran.r-project.org/package=RGraphics>, pristupljeno 04.05.2024.
- [459] Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression*, Third edition. Sage, Thousand Oaks CA. <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>, pristupljeno 04.05.2024.
- [460] Sarkar, D. (2008). *Lattice: Multivariate Data Visualization with R*. Springer, New York. ISBN 978-0-387-75968-5, <http://lmdvr.r-forge.r-project.org>, pristupljeno 04.05.2024.

- [461] Bender-deMoll, S., & Morris, M. (2022). ndtv: Network dynamic temporal visualizations, CRAN, <https://cran.r-project.org/package=ndtv>, pristupljeno 04.05.2024.
- [462] Chang, W., & Wickham, H. (2015). ggvis: Interactive Grammar of Graphics, CRAN, <https://CRAN.R-project.org/package=ggvis>, pristupljeno 04.05.2024.
- [463] Sievert, C. (2020). *Interactive Web-Based Data Visualization with R, plotly, and shiny*. Chapman and Hall/CRC. <https://plotly-r.com>, pristupljeno 04.05.2024.
- [464] Arnold, J.B. (2019). ggthemes: Extra Themes, Scales and Geoms for 'ggplot2', CRAN, <https://cran.r-project.org/package=ggthemes>, pristupljeno 04.05.2024.
- [465] Chang, W., Cheng, J., Allaire, J., Xie, Y., & McPherson, J. (2015). Shiny: web application framework for R, CRAN, <https://cran.r-project.org/package=shiny>, pristupljeno 04.05.2024.
- [466] Kassambara, A., & Kassambara, M.A. (2019). ggcorrplot: Visualization of a correlation matrix using 'ggplot2', CRAN, <https://cran.r-project.org/package=ggcorrplot>, pristupljeno 04.05.2024.
- [467] Attali, D., & Almende, B. V. (2022). timevis: Create Interactive Timeline Visualizations in R, CRAN, <https://cran.r-project.org/package=timevis>, pristupljeno 03.05.2024.
- [468] Cuttone, A., Lehmann, S., & Larsen, J. E. (2016). Geoplotlib: A python toolbox for visualizing geographical data. *arXiv preprint arXiv:1608.01933*.
- [469] Reback, J., McKinney, W., Van Den Bossche, J., Augspurger, T., Cloud, P., Klein, A., ... & Seabold, S. (2020). pandas-dev/pandas: Pandas 1.0. 5. *Zenodo*.
- [470] VanderPlas, J., Granger, B., Heer, J., Moritz, D., Wongsuphasawat, K., Satyanarayan, A., ... & Sievert, S. (2018). Altair: interactive statistical visualizations for Python. *Journal of Open Source Software*, 3(32), 1057.
- [471] Adams, C. (2014). *Learning Python data visualization*. Packt Publishing.
- [472] Bostock, M. (2012). D3.js - Data-Driven Documents.
- [473] Contributors. (2024). Chart.js, <https://www.chartjs.org/>, pristupljeno 04.05.2024.
- [474] Contributors. (2024). dimple, <http://dimplejs.org/>, pristupljeno 04.05.2024.
- [475] Dale, K. (2022). *Data Visualization with Python and JavaScript*, O'Reilly Media, Inc.
- [476] Otašević, V., & Kosanović, B. (2020). Primena softvera otvorenog koda kod vizualizacije podataka. PSSOH konferencija, Univerzitet u Beogradu – Elektrotehnički fakultet i Akademska Misao, pp. 34–37. <https://doi.org/10.5281/zenodo.4748508>
- [477] Nordmann, E., McAleer, P., Toivo, W., Paterson, H., & DeBruine, L. M. (2022). Data visualization using R for researchers who do not use R. *Advances in Methods and Practices in Psychological Science*, 5(2), 25152459221074654. <https://doi.org/10.1177/25152459221074654>
- [478] ggpy: ggplot port for Python, <https://github.com/yhat/ggpy>, pristupljeno 04.05.2024.
- [479] Anon. (1977), Exercises in Applied Statistics, Exercise 4.15, Dept. of Theoretical Statistics, Aarhus University.
- [480] Altman, D. G., & Bland, J. M. (1983). Measurement in medicine: the analysis of method comparison studies. *Journal of the Royal Statistical Society Series D: The Statistician*, 32(3), 307-317. <https://doi.org/10.2307/2987937>
- [481] Datta, D. (2017). blandr: A Bland-Altman Method Comparison package for R, CRAN, <https://zenodo.org/doi/10.5281/zenodo.824514>, GitHub, <https://github.com/deepankardatta/blandr>, pristupljeno 05.05.2024.
- [482] Arel-Bundock, V. (2024). Rdatasets: A collection of datasets originally distributed in various R packages. R package version 1.0.0, <https://vincentarelbundock.github.io/Rdatasets>, pristupljeno 05.05.2024.

- [483] Statistical Society of Canada, Case studies for the 2003 annual meeting, <https://ssc.ca/en/case-studies-2003-annual-meeting>, pristupljeno 05.05.2024.
- [484] Omvik, P. (1996). How smoking affects blood pressure. *Blood Pressure*, 5(2), 71-77. <https://doi.org/10.3109/08037059609062111>
- [485] Hadley, W. (2016). *Ggplot2: Elegrant graphics for data analysis*. Springer.
- [486] Pearson, K. (1895). VII. Note on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, 58(347-352), 240-242.
- [487] Irizarry, R. (2019). Open letter to journal editors: dynamite plots must die, SimplyStatistics Blog, <https://simplystatistics.org/posts/2019-02-21-dynamite-plots-must-die/>, pristupljeno 02.02.2024.
- [488] Wilson, F. N., Johnston, F. D., Rosenbaum, F. F., & Barker, P. S. (1946). On Einthoven's triangle, the theory of unipolar electrocardiographic leads, and the interpretation of the precordial electrocardiogram.
- [489] Barnes, A. R., Katz, L. N., Levine, S. A., Pardee, H. E., White, P. D., & Wilson, F. N. (1943). The standardization of electrocardiographic nomenclature: Report of Committee of the American Heart Association. *Journal of the American Medical Association*, 121(17), 1347-1349. <https://doi.org/10.1001/jama.1943.62840170006009>
- [490] Goodyear, M. D., Krleza-Jeric, K., & Lemmens, T. (2007). The declaration of Helsinki. *Bmj*, 335(7621), 624-625., <https://doi.org/10.1136/bmj.39339.610000.BE>
- [491] Zhu, Q., Tian, X., Wong, C. W., & Wu, M. (2021). Learning your heart actions from pulse: ECG waveform reconstruction from PPG. *IEEE Internet of Things Journal*, 8(23), 16734-16748. <https://doi.org/10.1109/JIOT.2021.3097946>
- [492] Bhatti, U., Rani, K., & Memon, M. Q. (2014). Variation in lung volumes and capacities among young males in relation to height. *Journal of Ayub Medical College Abbottabad*, 26(2), 200-202. <http://www.ayubmed.edu.pk/JAMC/26-2/Bhatti.pdf>, pristupljeno 06.05.2024.
- [493] Cook, C. D., & Hamann, J. F. (1961). Relation of lung volumes to height in healthy persons between the ages of 5 and 38 years. *The Journal of Pediatrics*, 59(5), 710-714. [https://doi.org/10.1016/S0022-3476\(61\)80007-3](https://doi.org/10.1016/S0022-3476(61)80007-3)
- [494] Hauke, J., & Kossowski, T. (2011). Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. *Quaestiones Geographicae*, 30(2), 87-93. <https://intapi.sciendo.com/pdf/10.2478/v10117-011-0021-1>, pristupljeno 02.06.2024.
- [495] Kassambara, A., & Kassambara, M. A. (2020). ggpubr: 'ggplot2' based publication ready plots, CRAN, <https://cran.r-project.org/package=ggpubr>, pristupljeno 02.06.2024.
- [496] Verran, J. A., & Ferketich, S. L. (1987). Testing linear model assumptions: Residual analysis. *Nursing Research*, 36(2), 127-129.
- [497] James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). Linear regression. In *An introduction to statistical learning: With applications in python* (pp. 69-134). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-38747-0_3
- [498] Tološi, L., & Lengauer, T. (2011). Classification with correlated features: unreliability of feature ranking and solutions. *Bioinformatics*, 27(14), 1986-1994. <https://doi.org/10.1093/bioinformatics/btr300>
- [499] Popović, N. B., Miljković, N., & Popović, M. B. (2019). Simple gastric motility assessment method with a single-channel electrogastrogram. *Biomedical Engineering/Biomedizinische Technik*, 64(2), 177-185. <https://doi.org/10.1515/bmt-2017-0218>

- [500] Safavi-Naeini, P., & Saeed, M. (2016). Pacemaker troubleshooting: common clinical scenarios. *Texas Heart Institute Journal*, 43(5), 415-418. <https://doi.org/10.14503/THIJ-16-5918>
- [501] Mezzarane, R. A., & Kohn, A. F. (2009). A method to estimate EMG crosstalk between two muscles based on the silent period following an H-reflex. *Medical Engineering & Physics*, 31(10), 1331-1336. <https://doi.org/10.1016/j.medengphy.2009.09.005>
- [502] Van Vugt, J. P. P., & Van Dijk, J. G. (2001). A convenient method to reduce crosstalk in surface EMG. *Clinical Neurophysiology*, 112(4), 583-592. [https://doi.org/10.1016/S1388-2457\(01\)00482-5](https://doi.org/10.1016/S1388-2457(01)00482-5)
- [503] Winter, D. A., Fuglevand, A. J., & Archer, S. E. (1994). Crosstalk in surface electromyography: theoretical and practical estimates. *Journal of Electromyography and Kinesiology*, 4(1), 15-26. [https://doi.org/10.1016/1050-6411\(94\)90023-X](https://doi.org/10.1016/1050-6411(94)90023-X)
- [504] Miljković, N. & Isaković, M. S. (2020). Surface electromyogram (sEMG) dataset recorded from forearm for 9 hand movements and three electrode array positions. [Data set]. *Zenodo*. <https://zenodo.org/doi/10.5281/zenodo.4039549>
- [505] Bryan, J. (2016). *Stat545: Data wrangling, exploration, and analysis with R*. <https://stat545.com/>, pristupljeno 06.05.2024.
- [506] Urbanek, S. (2022). *jpeg: Read and write JPEG images*, CRAN, <https://cran.r-project.org/package=jpeg>, pristupljeno 06.05.2024.
- [507] Cobb, M. (2017). The prehistory of biology preprints: A forgotten experiment from the 1960s. *PLoS Biology*, 15(11), e2003995. <https://doi.org/10.1371/journal.pbio.2003995>
- [508] Galili, T., O'Callaghan, A., Sidi, J., & Sievert, C. (2018). heatmaply: an R package for creating interactive cluster heatmaps for online publishing. *Bioinformatics*, 34(9), 1600-1602. <https://doi.org/10.1093/bioinformatics/btx657>
- [509] Van Panhuis, W. G., Grefenstette, J., Jung, S. Y., Chok, N. S., Cross, A., Eng, H., ... & Burke, D. S. (2013). Contagious diseases in the United States from 1888 to the present. *The New England Journal of Medicine*, 369(22), 2152. <https://doi.org/10.1056%2FNEJMms1215400>
- [510] DeBold, T., & Friedman, D. (2015). Battling infectious diseases in the 20th century: the impact of vaccines. *Wall Street Journal*, <http://graphics.wsj.com/infectious-diseases-and-vaccines/>, pristupljeno 07.05.2024.
- [511] Tulchinsky, T. H. (2018). Maurice Hilleman: Creator of vaccines that changed the world. *Case Studies in Public Health*, 443. <https://doi.org/10.1016%2FB978-0-12-804571-8.00003-2>
- [512] Garnier, S., Ross, N., Rudis, R., Camargo, P. A., Sciaini, M., & Scherer, C. (2024). *viridis(Lite) - Colorblind-Friendly Color Maps for R*. [doi:10.5281/zenodo.4679423](https://doi.org/10.5281/zenodo.4679423), viridis package version 0.6.5, <https://sjmgarnier.github.io/viridis/m> pristupljeno 03.06.2024.
- [513] Sueur, J., Aubin, T., & Simonis, C. (2008). Seewave, a free modular tool for sound analysis and synthesis. *Bioacoustics*, 18(2), 213-226. <https://doi.org/10.1080/09524622.2008.9753600>
- [514] Moura, L. M., Shafi, M. M., Ng, M., Pati, S., Cash, S. S., Cole, A. J., ... & Westover, M. B. (2014). Spectrogram screening of adult EEGs is sensitive and efficient. *Neurology*, 83(1), 56-64. <https://doi.org/10.1212%2FWNL.0000000000000537>
- [515] Borst, J. M. (1956). The use of spectrograms for speech analysis and synthesis. *Journal of the Audio Engineering Society*, 4(1), 14-23. <https://www.aes.org/e-lib/browse.cfm?elib=314>

- [516] Miljković, N. (2018). Otvoreni nastavni materijali: Interna iskustva. PSSOH konferencija, Univerzitet u Beogradu – Elektrotehnički fakultetu i Akademska misao, 16–19. <https://doi.org/10.5281/zenodo.1468045>
- [517] Leung, W. C. (2001). *Statistics and evidence-based medicine for examinations*. Radcliffe Publishing.
- [518] Luber, K. M., & Wolde-Tsadik, G. (1997). Efficacy of functional electrical stimulation in treating genuine stress incontinence: a randomized clinical trial. *Neurourology and Urodynamics: Official Journal of the International Continence Society*, 16(6), 543-551. [https://doi.org/10.1002/\(SICI\)1520-6777\(1997\)16:6%3C543::AID-NAU4%3E3.0.CO;2-E](https://doi.org/10.1002/(SICI)1520-6777(1997)16:6%3C543::AID-NAU4%3E3.0.CO;2-E)
- [519] Urinary Incontinence – Overview, NHS, UK, <https://www.nhs.uk/conditions/urinary-incontinence/>, pristupljeno 02.06.2024.
- [520] Evans, J. (2015), Bad Data Analysis vs. Better Data Analysis, <https://pbs.twimg.com/media/CJAJkJyW8AA9gvm.jpg>, pristupljeno 02.06.2024.
- [521] Belenky, G., Wesensten, N. J., Thorne, D. R., Thomas, M. L., Sing, H. C., Redmond, D. P., ... & Balkin, T. J. (2003). Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: A sleep dose-response study. *Journal of Sleep Research*, 12(1), 1-12. <https://doi.org/10.1046/j.1365-2869.2003.00337.x>
- [522] Kosinski, R. J. (2008). A literature review on reaction time. *Clemson University*, 10(1), 337-344. <https://www.sciencebuddies.org/Files/16478/4/clemson.rt.pdf>, pristupljeno 02.06.2024.
- [523] Chrysler, S. T., Ahmad, O., & Schwarz, C. W. (2015). Creating pedestrian crash scenarios in a driving simulator environment. *Traffic Injury Prevention*, 16(sup1), S12–S17. <https://doi.org/10.1080/15389588.2015.1015001>
- [524] Edwards, C. J., Creaser, J. I., Caird, J. K., Lamsdale, A. M., & Chisholm, S. L. (2003). Older and younger driver performance at complex intersections: Implications for using perception-response time and driving simulation. *Proceedings of the Second International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design* (pp. 33–38). Public Policy Center.
- [525] Green, M. (2000). “How long does it take to stop?” Methodological analysis of driver perception-brake times. *Transportation Human Factors*, 2(3), 195–216. https://doi.org/10.1207/STHF0203_1
- [526] Jurecki, R. S., & Stańczyk, T. L. (2018). Analyzing driver response times for pedestrian intrusions in crash-imminent situations. *2018 XI International Science-Technical Conference Automotive Safety* (pp. 1–7). IEEE. <https://doi.org/10.1109/AUTOSAFE.2018.8373339>
- [527] Olson, P. L., & Sivak, M. (1986). Perception-response time to unexpected roadway hazards. *Human Factors*, 28(1), 91–96. <https://doi.org/10.1177/001872088602800110>
- [528] Ciceri, M., Ruscio, D., Confalonieri, F., Vangi, D., & Virga, A. (2013). Hazard detection in driving context and braking reaction. In *Proceedings EVU 2013* (pp. 101-105). EVU Europe. <https://hdl.handle.net/2158/823315>
- [529] D’Addario, P., & Donmez, B. (2019). The effect of cognitive distraction on perception-response time to unexpected abrupt and gradually onset roadway hazards. *Accident Analysis & Prevention*, 127, 177–185. <https://doi.org/10.1016/j.aap.2019.03.003>
- [530] Ibáñez, V., Silva, J., & Cauli, O. (2018). A survey on sleep assessment methods. *PeerJ*, 6, e4849. <https://doi.org/10.7717/peerj.4849>
- [531] Loh, S., Lamond, N., Dorrian, J., Roach, G., & Dawson, D. (2004). The validity of psychomotor vigilance tasks of less than 10-minute duration. *Behavior Research Methods, Instruments, & Computers*, 36, 339-346. <https://doi.org/10.3758/BF03195580>

- [532] Psychomotor Vigilance Test (PVT) on ISS, NASA Life Sciences Portal, https://nls.nasa.gov/view/lspub/lspda_experiment/f39b0243-b8f5-5a2a-8905-e5c694b44c22, pristupljeno 02.06.2024.
- [533] Kvantili, kvartili, decili i percentili, 2011, <https://e-statistika.rs/kvantili-kvartili-decili-i-percentili>, pristupljeno 17.01.2024.
- [534] Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4), 591-611. <https://doi.org/10.1093/biomet/52.3-4.591>
- [535] Trajković, G., & Bukumirić, Z. (2020). *Medicinska statistika u R programskom okruženju*, Akademski Misao, Beograd.
- [536] Goodman, S. N. (1999). Toward evidence-based medical statistics. 1: The P value fallacy. *Annals of Internal Medicine*, 130(12), 995-1004. <https://doi.org/10.7326/0003-4819-130-12-199906150-00008>
- [537] Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 50-60. <https://doi.org/10.1214%2Faoms%2F1177730491>
- [538] Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260), 583-621. <https://doi.org/10.1080%2F01621459.1952.10483441>
- [539] The Theory Behind Mann-Whitney U tests (A.k.A. Wilcoxon Rank Sum test) & Kruskal-Wallis H Tests, Stata Help, Reed College, <https://www.reed.edu/psychology/stata/analyses/nonparametric/kruskal-wallacetheory.html>, pristupljeno 02.06.2024.
- [540] McDonald, J. H. (2009). *Handbook of biological statistics* (Vol. 2, pp. 6-59). Baltimore, MD: sparky house publishing.
- [541] Mangiafico, S. S. (2015). An R companion for the handbook of biological statistics.
- [542] Goss-Sampson, M. (2019). Statistical analysis in JASP: A guide for students.
- [543] Boynton, R. E., & Todd, R. L. (1947). Blood pressure readings of 75,258 university students. *Archives of Internal Medicine*, 80(4), 454-462. <https://doi.org/10.1001/archinte.1947.00220160033003>
- [544] Bonnet, M. H., & Arand, D. L. (1995). We are chronically sleep deprived. *Sleep*, 18(10), 908-911. <https://doi.org/10.1093/sleep/18.10.908>
- [545] Leger, D. (1994). The cost of sleep-related accidents: a report for the National Commission on Sleep Disorders Research. *Sleep*, 17(1), 84-93. <https://doi.org/10.1093/sleep/17.1.84>
- [546] Spiegel, K., Leproult, R., & Van Cauter, E. (1999). Impact of sleep debt on metabolic and endocrine function. *The Lancet*, 354(9188), 1435-1439. [https://doi.org/10.1016/S0140-6736\(99\)01376-8](https://doi.org/10.1016/S0140-6736(99)01376-8)
- [547] Horne, J. A. (1988). Sleep loss and “divergent” thinking ability. *Sleep*, 11(6), 528-536.
- [548] Dawson, C. (2024). ggprism: A ‘ggplot2’ Extension Inspired by ‘GraphPad Prism’, CRAN, <https://cran.r-project.org/package=ggprism>, pristupljeno 27.01.2024.
- [549] Milojković-Opsenica, D. Statistički testovi, https://www.chem.bg.ac.rs/~dusankam/ORM_10_11/Termin%205_%206%20Statisti%83ki%20testovi.pdf, pristupljeno 27.01.2024.
- [550] Student’s t-test, Bozeman Science, <https://youtu.be/pTmLQvMM-1M>, pristupljeno 28.01.2024.
- [551] Haemycytometer, A. (1907). On the error of counting with, *Biometrika*, 5(3), 351-360. <https://doi.org/10.2307/2331633>

- [552] Ziliak, S. T. (2019). How large are your G-values? Try Gosset's Guinnessometrics when a little "p" is not enough. *The American Statistician*, 73(sup1), 281-290. <https://doi.org/10.1080/00031305.2018.1514325>
- [553] ANOVA in R, Data Novia, <https://www.datanovia.com/en/lessons/anova-in-r/>, pristupljeno 27.01.2024.
- [554] One-Way ANOVA Test in R, STHDA, http://www.sthda.com/english/wiki/one-way-anova-test-in-r-#google_vignette, pristupljeno 28.01.2024.
- [555] Kroese, D. P., Brereton, T., Taimre, T., & Botev, Z. I. (2014). Why the Monte Carlo method is so important today. *Wiley Interdisciplinary Reviews: Computational Statistics*, 6(6), 386-392. <https://doi.org/10.1002/wics.1314>
- [556] Banerjee, A., Chitnis, U. B., Jadhav, S. L., Bhawalkar, J. S., & Chaudhury, S. (2009). Hypothesis testing, type I and type II errors. *Industrial Psychiatry Journal*, 18(2), 127. <https://doi.org/10.4103/0972-6748.62274>
- [557] Champely, S., Ekstrom, C., Dalgaard, P., Gill, J., Weibelzahl, S., Anandkumar, A., Ford, C., Volcic, R., & De Rosario, H. (2020). pwr: Basic Functions for Power Analysis, CRAN, <https://cran.r-project.org/package=pwr>, pristupljeno 28.01.2024.
- [558] Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd Edn. Hillsdale, NJ: Erlbaum.
- [559] Batričević, A. (2012). Etički i pravni okviri za zaštitu dobrobiti eksperimentalnih životinja u Republici Srbiji. *Glasnik Advokatske komore Vojvodine*, 84(11), 726-751.
- [560] Quintana, D. S. (2017). Statistical considerations for reporting and planning heart rate variability case-control studies. *Psychophysiology*, 54(3), 344-349. <https://doi.org/10.1111/psyp.12798>
- [561] Laborde, S., Mosley, E., & Thayer, J. F. (2017). Heart rate variability and cardiac vagal tone in psychophysiological research—recommendations for experiment planning, data analysis, and data reporting. *Frontiers in Psychology*, 8, 238557. <https://doi.org/10.3389/fpsyg.2017.00213>
- [562] Torchiano M (2020). *effsize: Efficient Effect Size Computation*. [doi:10.5281/zenodo.1480624](https://doi.org/10.5281/zenodo.1480624), R package version 0.8.1, CRAN, <https://CRAN.R-project.org/package=effsize>, pristupljeno 28.01.2024.
- [563] Jakus, G., Sodnik, J., & Miljković, N. (2022). NadicaSm/Statistical-Analysis-and-Machine-Learning-for-EKG-based-Nausea-Detection: v1 (Version v1). Version v1, *Zenodo*. <https://doi.org/10.5281/zenodo.7242797>
- [564] Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, 62627. <https://doi.org/10.3389/fpsyg.2013.00863>
- [565] Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance. *Nature*, 567(7748), 305-307. <https://doi.org/10.1038/d41586-019-00857-9>
- [566] King, G. (2007). An introduction to the dataverse network as an infrastructure for data sharing. *Sociological Methods & Research*, 36(2), 173-199. <https://doi.org/10.1177/0049124107306660>
- [567] Melvin, R. L., Barker, S. J., Kiani, J., & Berkowitz, D. E. (2022). Pro-Con Debate: Should Code Sharing Be Mandatory for Publication?. *Anesthesia & Analgesia*, 135(2), 241-245. <https://doi.org/10.1213/ane.0000000000005848>
- [568] Hong, N. P. C., Katz, D. S., Barker, M., Lamprecht, A. L., Martinez, C., Psomopoulos, F. E., ..., Miljković, N., ... & WG, F. (2022). FAIR principles for research software (FAIR4RS principles), *Zenodo*. <https://doi.org/10.15497/RDA00068>
- [569] Brooks, D. (2009). Extensible biosignal metadata a model for physiological time-series data. In *2009 Annual International Conference of the IEEE Engineering in Medicine*

- and *Biology Society* (pp. 3881-3884). IEEE. <https://doi.org/10.1109/IEMBS.2009.5332642>
- [570] Dohmke, T. (2023). 100 million developers and counting. The GitHub Blog, <https://github.blog/2023-01-25-100-million-developers-and-counting/>, pristupljeno 12.05.2024.
- [571] The Problem With Putting All the World's Code in GitHub,. Wired, 2015, <https://www.wired.com/2015/06/problem-putting-worlds-code-github/>, pristupljeno 12.05.2024.
- [572] Bryan, J. (2018). Happy Git and GitHub for the useR. GitHub. <https://happygitwithr.com/>, pristupljeno 03.06.2024.
- [573] The Missing Semester of Your CS Education, Lesson Version Control (Git), Massachusetts Institute of Technology, 2020, <https://missing.csail.mit.edu/2020/version-control/>, pristupljeno 28.07.2023.
- [574] Clare, C., Cruz, M., Papadopoulou, E., Savage, J., Teperek, M., Wang, Y., ... & Yeomans, J. (2019). *Engaging researchers with data management: The cookbook* (p. 170). Open Book Publishers.
- [575] Martinez-Ortiz, C., Martinez Lavanchy, P., Sesink, L., Olivier, B. G., Meakin, J., de Jong, M., & Cruz, M. (2023). Practical guide to Software Management Plans (1.1). Zenodo. <https://doi.org/10.5281/zenodo.7589725>
- [576] Kubben, P., Dumontier, M., & Dekker, A. (2019). Fundamentals of clinical data science. <https://doi.org/10.1007/978-3-319-99713-1>
- [577] Edstrom, D. (2005). Open Document Format for Office Applications (OpenDocument) v1.0, OASIS Standard, <https://web.archive.org/web/20060924102321/http://www.xml.gov/presentations/sun/odf.pdf>, pristupljeno 14.05.2024.
- [578] Porter Adams, V. (1981). Captain Grace M. Hopper: The mother of COBOL, InfoWorld, <https://web.archive.org/web/20190918012636/https://books.google.com/books?id=JT0EAAAAMBAJ&pg=RA1-PA33&lpg=RA1-PA33&focus=viewport#v=onepage&q&f=false>, pristupljeno 14.05.2024.
- [579] Knuth, D. E. (2007). Computer programming as an art. In *ACM Turing award lectures* (p. 1974).
- [580] Chop, R. M., & Silva, M. C. (1991). Scientific fraud: Definitions, policies, and implications for nursing research. *Journal of Professional Nursing*, 7(3), 166-171. [https://doi.org/10.1016/8755-7223\(91\)90051-L](https://doi.org/10.1016/8755-7223(91)90051-L)
- [581] Gruber, J. (2012). Markdown: Syntax. <https://daringfireball.net/projects/markdown/syntax#philosophy>, pristupljeno 17.05.2024.
- [582] Xie, Y. (2024). tinytex: Helper Functions to Install and Maintain TeX Live, and Compile LaTeX Documents. R package version 0.51, <https://github.com/rstudio/tinytex>, pristupljeno 17.05.2024.
- [583] Xie, Y. (2019). TinyTeX: A lightweight, cross-platform, and easy-to-maintain LaTeX distribution based on TeX Live, TUGboat, 40(1), 30–32. <https://tug.org/TUGboat/Contents/contents40-1.html>, pristupljeno 17.05.2024.
- [584] Xie, Y. (2024). *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.39, <https://github.com/rstudio/bookdown>, pristupljeno 17.05.2024.
- [585] Xie, Y. (2016). *bookdown: Authoring Books and Technical Documents with R Markdown*. Chapman and Hall/CRC, Boca Raton, Florida, <https://bookdown.org/yihui/bookdown>, pristupljeno 17.05.2024.

- [586] Brandl, G. (2010). Sphinx documentation. URL <http://sphinx-doc.org/sphinx.pdf>, pristupljeno 26.05.2024.
- [587] Milovanović, V. (2021). Komponovanje računarskih programa, Fakultet inženjerskih nauka, <https://github.com/milovanovic/krp>, pristupljeno 26.04.2024.
- [588] Milovanović, V. (2021). An Engineering Textbook Typesetting Using Sphinx Documentation Generator. PSSOH Conference, pp. 33–40. <https://doi.org/10.5281/zenodo.6506339>
- [589] byteshisper. (2023). Naming conventions: Four popular cases, dev Blog, <https://dev.to/bytewhisper/naming-conventions-4-popular-cases-571k>, pristupljeno 17.05.2024.
- [590] Rules of naming variables, Developer network, https://1c-dn.com/library/rules_of_naming_variables/, pristupljeno 17.05.2024.
- [591] Hester, J., Angly, F., Hyde, R., Chirico, M., Ren, K., Rosenstock, A., & Patil, I. (2024). lintr: A ‘Linter’ for R code, CRAN, <https://cran.r-project.org/package=lintr>, pristupljeno 17.05.2024.
- [592] Müller, K., Walthert, L., & Patil, I. (2024). styler: Non-invasive pretty printing for R code, CRAN, <https://cran.r-project.org/package=styler>, pristupljeno 17.05.2024.
- [593] Bezanson, J., Karpinski, S., Shah, V. B., & Edelman, A. (2012). Why We Created Julia, <https://julialang.org/blog/2012/02/why-we-created-julia/>, pristupljeno 29.01.2024.
- [594] Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. (2017). Julia: A fresh approach to numerical computing. *SIAM Review*, 59(1), 65–98. <https://doi.org/10.1137/141000671>
- [595] Williams, S. (2010). Free as in freedom (2.0): Richard Stallman and the free software revolution, Boston: Free Software Foundation, <https://sagitter.fedorapeople.org/faif-2.0.pdf>, pristupljeno 03.06.2024.
- [596] Kung, J. Y. (2023). Elicit. *The Journal of the Canadian Health Libraries Association*, 44(1), 15. <https://doi.org/10.29173%2Fjchla29657>
- [597] McKee, L. (2023). Can ChatGPT Write Your Next Scientific Paper, *Scientific Writing with Karen*, <https://youtu.be/wnGPt030IG4>, pristupljeno 31.01.2024.
- [598] Géron, A. (2022). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. O’Reilly Media, Inc.
- [599] Contractor, D., McDuff, D., Haines, J. K., Lee, J., Hines, C., Hecht, B., ... & Li, H. (2022, June). Behavioral use licensing for responsible ai. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 778-788). <https://doi.org/10.1145/3531146.3533143>
- [600] Xu, F., Uszkoreit, H., Du, Y., Fan, W., Zhao, D., Zhu, J. (2019). Explainable AI: A Brief Survey on History, Research Areas, Approaches and Challenges. In: *Tang, J., Kan, MY., Zhao, D., Li, S., Zan, H. (eds) Natural Language Processing and Chinese Computing. NLPCC 2019. Lecture Notes in Computer Science*, vol. 11839. Springer, Cham. https://doi.org/10.1007/978-3-030-32236-6_51
- [601] Cabanac, G., Labbé, C., & Magazinov, A. (2021). Tortured phrases: A dubious writing style emerging in science. Evidence of critical issues affecting established journals. arXiv preprint arXiv:2107.06751. arXiv. <https://doi.org/10.48550/arXiv.2107.06751>
- [602] Surameery, N. M. S., & Shakor, M. Y. (2023). Use chat gpt to solve programming bugs. *International Journal of Information Technology & Computer Engineering (IJITC)*, 2455-5290, 3(01), 17-22. <https://doi.org/10.55529/ijitc.31.17.22>
- [603] Biswas, S. (2023). Role of ChatGPT in Computer Programming.: ChatGPT in Computer Programming. *Mesopotamian Journal of Computer Science*, 2023, 8-16. <https://doi.org/10.58496/MJCSC/2023/002>

- [604] Machlis, S. (2023). 8 ChatGPT tools for R programming, Info World, <https://www.infoworld.com/article/3694611/8-chatgpt-tools-for-r-programming.html>, pristupljeno 31.01.2024.
- [605] Ray, S. (2023). air: AI assistant to Write and Understand R Code, CRAN, <https://cran.r-project.org/package=air>, pristupljeno 31.01.2024.
- [606] Banta, J. (2022). How to install R and Rstudio under Ubuntu Linux, YouTube, <https://youtu.be/eE1zh0YVMeM>, pristupljeno 02.02.2024.
- [607] Vesasquez, I. (2021). Three ways to program in Python with R Studio, Posit Blog, <https://posit.co/blog/three-ways-to-program-in-python-with-rstudio/>, pristupljeno 12.05.2024.
- [608] Galili, T., Rowlingson, B., Hejblum, B., Schonbrodt, F., Grothendieck, G., Daroczi, G., & Andrew, H. (2022). installr: Using R to install stuff on Windows OS, CRAN, <https://cran.r-project.org/package=installr>, pristupljeno 12.05.2024.
- [609] Fryda, T. et al. (2024). h2o: Interface for the 'H2O' scalable machine learning platform, CRAN, <https://cran.r-project.org/package=h2o>, pristupljeno 12.05.2024.
- [610] Aiello, S., Eckstrand, E., Fu, A., Landry, M., & Aboyou, P. (2015). Machine Learning with R and H2O, <http://h2o.ai/resources/>, pristupljeno 12.05.2024.
- [611] Aden-Buie, B., et al. (2023). learnr: Interactive tutorials for R, CRAN, <https://cran.r-project.org/package=learnr>, pristupljeno 12.05.2024.

Dodatak A – Spisak lekcija iz TOBS predmeta

U ovom dodatku se nalazi spisak lekcija sa referencama ka prezentacijama za TOBS predmet. Iako je udžbenik namenjen TOBS predmetu, redosled poglavlja u udžbeniku ne odgovara redosledu lekcija na TOBS predmetu. Razlog je jednostavan – tokom semestra, lekcije su organizovane tako da delimično pokrivaju aspekte koji se odnose rad na računaru (programiranje u R-u i upoznavanje sa specifičnim paketima za obradu signala), dok se drugi deo odnosi na odabrane primere obrade signala. Kako bi gradivo koje se prenosi studentima bilo zanimljivije i kako bi ranije mogli da uoče praktični značaj delimično suvoparnih elemenata rada na računaru, ova dva elementa su kombinovana u predavanjima. To nije bio slučaj za udžbenikom.

Veliki broj primera koji su prikazani u prezentacijama i podeljeni pod otvorenom CC BY 4.0 licencom⁴⁹⁰ je preuzet za potrebe ilustracija primera u ovom udžbeniku. Međutim, postoje elementi koji nisu prikazani u prezentacijama, a koji su sadržani u udžbeniku, te se za praćenje predmeta, savetuje studentima da koriste i udžbenik i prezentacije za predavanja.

Spisak lekcija sa prezentacijama za predavanja je:

1. Uvodno predavanje, <https://doi.org/10.5281/zenodo.4573991>
2. Osnovni podaci u programskom jeziku R, <https://doi.org/10.5281/zenodo.4606661>
3. Interfejsi u R-u, primeri biomedicinskih signala i nauka o podacima, <https://doi.org/10.5281/zenodo.4625915>
4. Operatori, kontrolne strukture, dplyr paket i studija spavanja, <https://doi.org/10.5281/zenodo.4642413>
5. Srčani puls, funkcije, estetika, dplyr paket i pretraga, <https://doi.org/10.5281/zenodo.4661754>
6. Kontrolne funkcije, debugovanje, profilisanje koda i filtriranje biosignala, <https://doi.org/10.5281/zenodo.4679295>
7. Generisanje pseudo(slučajnih) brojeva i sintetičkih biosignala, <https://doi.org/10.5281/zenodo.4699316>
8. Analiza EMG-a i vizuelizacija (I deo), <https://doi.org/10.5281/zenodo.4721818>

490 <https://creativecommons.org/licenses/by/4.0/legalcode>, pristupljeno 18.09.2023.

9. Vizuelizacija (II deo) i merenje signala u R-u, <https://doi.org/10.5281/zenodo.4739338>
10. Automatski izveštaji i t-test, <https://doi.org/10.5281/zenodo.4764537>
11. t-test i case study analize podataka, <https://doi.org/10.5281/zenodo.4784060>
12. Shiny veb aplikacija, <https://doi.org/10.5281/zenodo.4894532>

Svi Zenodo repozitorijumi sadrže prezentacije za 2020/2021, 2021/2022, 2022/2023. i 2023/2024. školske godine, dok se na sajtu predmeta⁴⁹¹ nalazi i spisak tri dodatne lekcije koje nisu držane studentima svake godine. Usled pandemije ili drugih nepredviđenih okolnosti, dešavalo se da semestar traje kraće, pa je Autorka ostavila studentima da sami odaberu koju bi lekciju dodatno slušali, u tom slučaju. Ove dodatne lekcije su nedovršene i prezentacije koje su date na sajtu, nisu podeljene na Zenodo repozitorijumu:

1. Topografija za vizuelizaciju podataka i Etika u radu sa podacima
2. PCA (eng. *Principal Component Analysis*) i EMD (eng. *Empirical Mode Decomposition*) realizacija u R-u za obradu signala
3. Topografska analiza, klaterizacija biosignala, standardi za snimanje i etika

U školskoj 2021/2022. godini, realizovana je nastava na daljinu, a sva predavanja su snimljena i dostupna na sajtu TOBS predmeta, kao i u okviru TOBS Jutjub (eng. *YouTube*) kanala⁴⁹².

491 <https://automatika.etf.bg.ac.rs/sr/13m051tobs>, pristupljeno 03.06.2024.

492 https://youtube.com/playlist?list=PLI3SYeiSufnAOXhd_OQ56RBldHMwqt3QP&si=V1ZFCIkYavgAP6VZ, pristupljeno 18.09.2023.

Dodatak B – Spisak vežbi iz TOBS predmeta

U ovom dodatku se nalazi spisak vežbi sa pratećim materijalima, koji su podeljeni na Zenodo repozitorijumu pod otvorenom CC BY licencom, kao i materijali za predavanja prikazani u Dodatku A. Neki od primera su prikazani u udžbeniku, a vežbe u velikoj meri hronološki prate lekcije sa predavanja. Prvih 6 vežbi je organizovano u formi prezentacija sa detaljnim pojašnjenjima za rad, dok je preostalih 6 vežbi predstavljeno u formi dokumenata sa jasno definisanim zadacima. Laboratorijske vežbe se, u okviru TOBS predmeta, izvode samostalno na Fakultetskim ili na sopstvenim laptop računarima.

Zadaci prikazani u vežbama i podeljeni pod otvorenom CC BY licencom, su prilagođeni za potrebe ilustracije primera u udžbeniku. Kao i za predavanja, postoje snimci i uputstva za TOBS vežbe na TOBS Jutjub kanalu, a svaki Zenodo repozitorijum trenutno sadrži prezentacije iz školskih 2020/2021, 2021/2022, 2022/2023. i 2023/2024. godina. Dodatno, na sajtu predmeta, pored naziv vežbe se nalaze predlozi rešenja (i/ili dodatnih signala) kojima se može pristupiti preko sajta TOBS predmeta.

Spisak vežbi za TOBS predmet je:

1. Uvodna vežba, <https://doi.org/10.5281/zenodo.4625577>
2. Osnovni podaci u R-u i učitavanje podataka iz datoteka, <https://doi.org/10.5281/zenodo.4625585>
3. Prikaz podataka i manipulacija nedostajućim vrednostima, <https://doi.org/10.5281/zenodo.4625784>
4. Manipulacija podacima, kontrolne strukture i dplyr paket na primeru studije spavanja, <https://doi.org/10.5281/zenodo.4699127>
5. Funkcija i analiza srčanog pulsa merenog na zdravim ispitanicima, <https://doi.org/10.5281/zenodo.4699135>
6. Kontrolne funkcije, pretprocesiranje biosignala i debugovanje, <https://doi.org/10.5281/zenodo.4699141>
7. Sintetički EMG signal, <https://doi.org/10.5281/zenodo.4764545>
8. Vizuelizacija podataka, <https://doi.org/10.5281/zenodo.4764586>

9. Vizuelizacija i merenje biosignala⁴⁹³,
<https://doi.org/10.5281/zenodo.4764624>
10. Automatsko generisanje izveštaja,
<https://doi.org/10.5281/zenodo.4896534>
11. t-test u stroke studiji i studija spavanja,
<https://doi.org/10.5281/zenodo.4896746>
12. Kreiranje *Shiny* veb aplikacije, <https://doi.org/10.5281/zenodo.4896766>

493 Ovaj deo laboratorijske vežbe se najčešće izvodi kao demonstracija, što zavisi od broja studenata u grupi i prostora koji je na raspolaganju za rad.

Dodatak C – R Studio razvojno okruženje

U ovom dodatku je prikazana osnovna funkcionalnost R Studio razvojnog okruženja, ali i način na koji je moguće instalirati programski jezik R i R Studio okruženje na računarima sa Vindows i Ubuntu operativnim sistemima.

Instalacija (Ubuntu operativni sistem)

Da bi se proverila verzija računara, na kome se instaliraju R programski jezik i R Studio okruženje, potrebno je u komandnom prozoru⁴⁹⁴ Ubuntu ukucati:

```
lsb_release -a
```

komandu. U udžbeniku će biti pokazano kako se instalira R programski jezik sa razvojnim okruženjem na primeru računara koji ima Ubuntu 22.04.1 LTS (eng. *Long-Term Support*) operativni sistem, a uputstvo je pripremljeno u skladu sa uputstvom prikazanim u [606]. Pre pokretanja instalacije, potrebno je pokrenuti ažuriranje (eng. *update*) repozitorijuma, u kome Linuks Ubuntu operativni sistem pretražuje softver, koji je potrebno instalirati unosom komande:

```
sudo apt update-a
```

za šta će biti potrebno da se unese lozinka na računaru. Proces provere verzije operativnog sistema i ažuriranja softvera na računaru sa Ubuntu operativnim sistemom je prikazan na Sl. C1. Nakon toga se pokreće instalacija programskog jezika R:

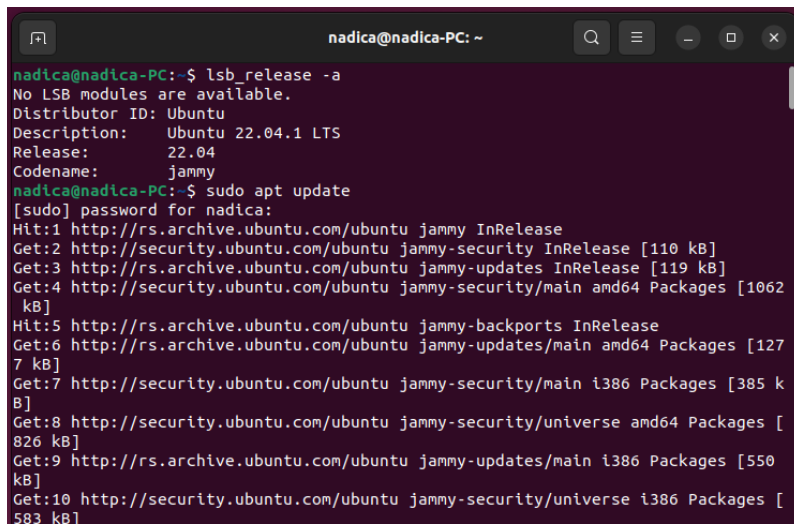
```
sudo apt -y install r-base gdebi-core
```

gde je *gdebi-core* alatka koja se koristi kao pomoć prilikom instaliranja. Sa ove dve komande završava se instalacija programskog jezika R. Nakon toga, potrebno je sa sajta R Studio proizvođača Posit⁴⁹⁵ preuzeti datoteku za

494 Jedan od načina da se otvori komandni prozor jeste da se pretragom iz aplikacija (na Ubuntu 22 operativnom sistemu aplikacije se nalaze u donjem levom uglu ekrana) odabere komandni prozor / terminal. Moguće je koristiti i skraćenicu sa tastature istovremenim pritiskom tastera Ctrl + Alt + T.

495 Posit Co. <https://posit.co/download/rstudio-desktop/>, pristupljeno 02.02.2024.

instaliranje R Studio Desktop okruženja (besplatna verzija) koje je namenjeno odgovarajućem operativnom sistemu (u ovom slučaju to je verzija instalacione datoteke za Ubuntu 22/Debian 12 pod nazivom “rstudio-2023.12.0-369-amd64.deb”).



```
nadica@nadica-PC: ~  
nadica@nadica-PC:~$ lsb_release -a  
No LSB modules are available.  
Distributor ID: Ubuntu  
Description:    Ubuntu 22.04.1 LTS  
Release:        22.04  
Codename:       jammy  
nadica@nadica-PC:~$ sudo apt update  
[sudo] password for nadica:  
Hit:1 http://rs.archive.ubuntu.com/ubuntu jammy InRelease  
Get:2 http://security.ubuntu.com/ubuntu jammy-security InRelease [110 kB]  
Get:3 http://rs.archive.ubuntu.com/ubuntu jammy-updates InRelease [119 kB]  
Get:4 http://security.ubuntu.com/ubuntu jammy-security/main amd64 Packages [1062  
kB]  
Hit:5 http://rs.archive.ubuntu.com/ubuntu jammy-backports InRelease  
Get:6 http://rs.archive.ubuntu.com/ubuntu jammy-updates/main amd64 Packages [127  
7 kB]  
Get:7 http://security.ubuntu.com/ubuntu jammy-security/main i386 Packages [385 k  
B]  
Get:8 http://security.ubuntu.com/ubuntu jammy-security/universe amd64 Packages [826  
kB]  
Get:9 http://rs.archive.ubuntu.com/ubuntu jammy-updates/main i386 Packages [550  
kB]  
Get:10 http://security.ubuntu.com/ubuntu jammy-security/universe i386 Packages [583  
kB]
```

Slika C1, Izgled komandnog prozora (terminala) u Ubuntu operativnom sistemu za izvršenje dve komande: (1) za proveru verzije operativnog sistema i (2) za ažuriranje softvera.

Ako je instalaciona datoteka sačuvana u fascikli, odnosno folderu Preuzimanja (eng. *Downloads*), onda je potrebno u komandnom prozoru postaviti tekuću fasciklu pozivom:

```
cd Downloads496
```

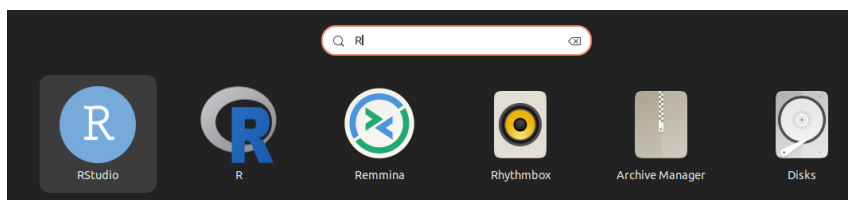
u slučaju da se u fascikli Preuzimanja nalazi instalaciona datoteka. Na kraju, moguće je pokrenuti instalaciju R Studio razvojnog okruženja komandom:

```
sudo gdebi rstudio-2023.12.0-369-amd64.deb497
```

496 Alternativno, terminal u kome je fascikla Preuzimanje/*Downloads* radni direktorijum se može dobiti i ako se najpre otvori fascikla na računaru, a potom se desnim klikom miša na sadržaj fascikle odabere opcija iz padajućeg menija *Open in Terminal*.

497 U terminalu neće raditi skraćenica sa tastature istovremenim pritiskom tastera Ctrl + V za kopiranje sadržaja, odnosno naziva instalacione datoteke, te se preporučuje da se koristi skraćenica Ctrl + Shift + V ili izbor *Paste* opcije iz menija koji se dobija pritiskom na desni klik računarskog miša.

nakon čega se može u komandnom prozoru pojaviti pitanje za saglasnost korisnika za početak instalacije. Jednostavnim unosom “y” ili “N” sa tastature, moguće je započeti ili zaustaviti instalaciju R Studio razvojnog okruženja. Po završetku procesa instaliranja, terminal više nije potreban i može se zatvoriti. Da bi se pristupilo razvojnom okruženju, može se u meniju aplikacija pretražiti R Studio kao na Sl. C2 (meni aplikacija se podrazumevano u Ubuntu 22 operativnom sistemu nalazi u donjem levom uglu ekrana). Desnim klikom na ikonicu R Studio okruženja, prečica do R Studio alata se može dodati u glavni skup alatki (eng. *toolbar*) do koga se jednostavno dolazi iz radne površine (eng. *desktop*).



Slika C2, Prikaz rezultata za pretraživanje po ključnim rečima (u ovom slučaju po slovu “R”) na računaru sa Ubuntu operativnim sistemom. Prvi rezultat (sa leva na desno) je R Studio okruženje, a sledeći je R programski jezik.

Za odgovarajuće korišćenje *install.packages()* funkcije na Ubuntu operativnom sistemu, potrebno je pokrenuti i komandu:

```
sudo apt-get install r-base-dev
```

u terminalu, prema instrukcijama sa CRAN-a⁴⁹⁸.

Instalacija (Vindovs operativni sistem)

U Vindovs operativnom sistemu se može proveriti verzija operativnog sistema u prozoru *System Information*, koji se dobija jednostavnom pretragom na računaru. U odeljku *System Summary* na kartici *System Information*, moguće je pogledati naziv operativnog sistema (*OS Name*). Računar na kome je pripremljeno uputstvo ima operativni sistem kao na Sl. C3 (*Microsoft Windows 10 Education*).

Ima i drugih načina kako se može proveriti verzija operativnog sistema. Na primer, pretragom u Guglu ili nekom drugom internet pretraživaču sa ključnim rečima "install R" otvara se prozor sa predlogom verzije programskog jezika R

498 <https://cran.r-project.org/bin/linux/ubuntu/fullREADME.html>, pristupljeno 16.02.2024.

koju je potrebno instalirati, jer u većini slučajeva pretraživači prepoznaju verziju Vindovs operativnog sistema na računaru.

Item	Value
OS Name	Microsoft Windows 10 Education
Version	10.0.19045 Build 19045
Other OS Description	Not Available
OS Manufacturer	Microsoft Corporation
System Name	PC1
System Manufacturer	FUJITSU
System Model	ESPRIMO P520

Slika C3, Izgled dela kartice *System Information* za uključenu opciju *System Summary*.

Za operativni sistem, u ovom primeru, preuzet je dokument koji je označen na CRAN stranici⁴⁹⁹ kao:

Download R-4.3.2 for Windows (79 megabytes, 64 bit)

Kada se preuzme instalaciona datoteka treba je pokrenuti, odabrati jezik za instalaciju, nakon čega se pojavljuje obavještenje o licenci pod kojom je dostupan programski jezik R (GNU GPL licenca) i bira se fascikla, tj. folder na računaru, u kome će se instalirati R. Preporuka je da to bude generička fascikla:

C:\ProgramFiles\R\R-4.3.2

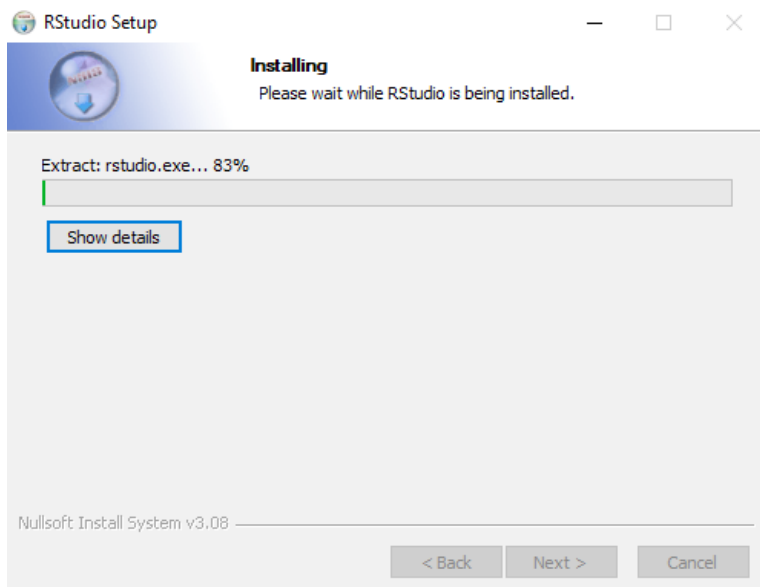
i onda se kroz veoma intuitivni interfejs započinje proces instalacije. Nakon što je R instaliran, potrebno je preuzeti instalacionu datoteku za Vindovs operativni sistem, sa sajta Posit⁵⁰⁰ kompanije. Kada se pokrene instalaciona datoteka, procedura je slična kao u slučaju instalacije R programskog jezika (proces instalacije je predstavljen na Sl. C4). Za odabir fascikle na računaru, preporučuje se slično generičko mesto, kao prilikom instalacije programskom jezika R:

499 <https://cran.r-project.org/bin/windows/base/>, pristupljeno 05.02.2024. (Kako se stalno objavljuju nove verzije programskog jezika R, u trenutku provere ovog linka na sajtu će se naći verzija koja je veća od 4.3.2 verzije koja je bila dostupna na dan 05.02.2024, a instalaciona datoteka koja je preuzeta u trenutku pisanja ovog uputstva je “R-4.3.2-win-exe”.)

500 <https://posit.co/downloads/>, pristupljeno 05.02.2024.

C:\Program Files\RStudio

Osim ako nije drugačije odabrano, pojaviće se ikonica R Studio programskog jezika na Desktopu računara.



Slika C4, Proces instalacije⁵⁰¹ programskog R Studio okruženja na Vindovs operativnom sistemu.

Na sajtu R Studio pretraživača se dobija informacija da je redosled instaliranja takav da se prvo instalira R, pa onda tek razvojno okruženje R Studio (Sl. C5), te se može prvo pristupiti Posit sajtu i mogu se iskoristiti linkovi koji su tamo dati za preuzimanje instalacionih datoteka. Za razliku od programskog jezika Pajton, u kome je korisnicima na raspolaganju veći izbor razvojnih okruženja, u R-u je, prema Autorkinim saznanjima, dominantno zastupljeno R Studio okruženje⁵⁰². Kao i u većini programskih jezika, nije potrebno ni instalirati R i R Studio na računaru, jer je moguće programirati u R-u korišćenjem pretraživača. Na primer, webR⁵⁰³ (Sl. C6) predstavlja razvojno okruženje koje koristi mogućnost programiranja preko pretraživača⁵⁰⁴.

501 U trenutku pisanja uputstva, instalaciona datoteka, koja je preuzeta, nosila je naziv “RStudio-2023.12.1-402.exe” na dan 05.02.2024. Treba imati na umu da će se verzija u budućnosti menjati.

502 Anaconda i Džupajter (eng. *Jupyter*) okruženja se mogu koristiti, ali su manje zastupljena (<https://docs.anaconda.com/free/working-with-conda/packages/using-r-language/>, pristupljeno 12.05.2024).

503 <https://docs.r-wasm.org/web/latest/>, pristupljeno 12.05.2024.

1: Install R

RStudio requires R 3.3.0+. Choose a version of R that matches your computer's operating system.

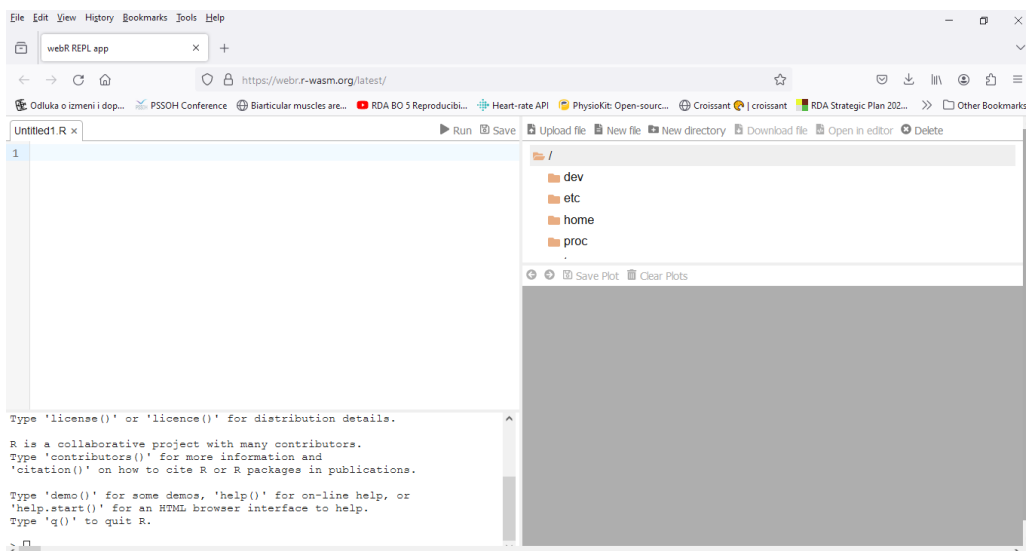
DOWNLOAD AND INSTALL R

2: Install RStudio

DOWNLOAD RSTUDIO DESKTOP FOR WINDOWS

Size: 215.66 MB | SHA-256: D3C03C42 | Version: 2023.12.1+402 | Released: 2024-01-29

Slika C5, Uputstvo za redosled instaliranja R programskog jezika i R Studio okruženja sa sajta Posit kompanije, <https://posit.co/downloads/>, pristupljeno 02.02.2024.

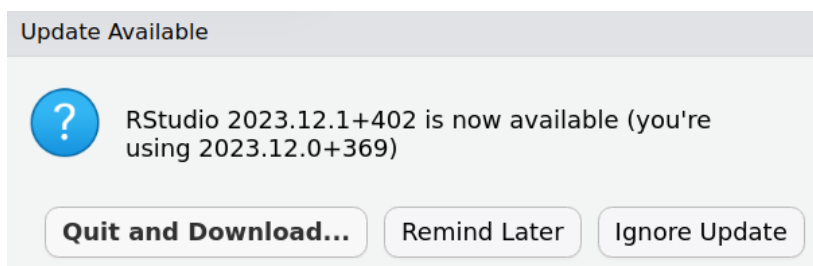


Slika C6, webR razvojno okruženje za programiranje u pretraživaču, bez prethodne instalacije programskog jezika i R Studio okruženja na računaru, <https://webr-r-wasm.org/latest/>, pristupljeno 02.02.2024.

Iako je R Studio razvojno okruženje, prvenstveno namenjeno programiranju u programskom jeziku R, moguće je koristiti i druge programske jezike (ili ih kombinovati sa R-om). Ima tri načina, kako se R Studio može koristiti za programiranje u Pajtonu, koji se dalje mogu istražiti u [607].

U R Studio okruženju, implementirani su automatski podsetnici za ažuriranje (eng. *update*) kao na Sl. C7. Korisnici mogu da odaberu opciju da odlože ažuriranje, da ga pokrenu ili da prosto ignorišu nove verzije.

504 Iako su ovo vrlo jednostavni i korisni alati, mora im se pristupiti sa dozom rezerve, jer je sasvim moguće da nisu sve opcije u okviru pretraživača dostupne kao što bi bile na računaru. Autorka, u svojoj praksi, koristi razvojna okruženja u pretraživačima samo za rad u C++ programskom jeziku, te studenti moraju sami istražiti mogućnosti koje pruža webR.



Slika C7, Ažuriranje R Studio razvojnog okruženja – automatska poruka.

Ažuriranje programskog jezika R u Vindovs operativnom sistemu, može se realizovati automatski korišćenjem installr paketa [608] sa CRAN-a uz funkciju:

`updateR()`

U ranijim verzijama, prema Autorinom iskustvu, ova vrsta ažuriranja nije bila idealna. U slučaju da automatske procedure nisu od koristi, R i R Studio se relativno brzo i jednostavno instaliraju na računaru. Dešava se da se javi problemi tokom ažuriranja paketa i drugih zavisnih softverskih alata (eng. *dependencies*). Na primer, korišćenje H2O paketa [609], koji se koristi za automatsko mašinsko učenje⁵⁰⁵ je usko povezano sa Java verzijom na računaru, te najveći problem koji korisnici imaju jeste što moraju da instaliraju ranije verzije Java softvera, jer pojava novih verzija različitih softvera nije sinhrona. [610]

Problem, koji može da nastane kod prelaska na noviju verziju R programskog jezika jeste kada nisu instalirani drugi paketi, na koje se željeni paket oslanja (eng. *dependencies*). Jedno od rešenja je da se ponovo instalira paket, koji nije mogao da se učita, ali sada sa argumentom `dependencies = TRUE` u okviru `install.packages()` funkcije. Nekada problem može da nastane i kada se menja broj ili vrsta korisničkih naloga na računaru. Za slučaj da je odabrana opcija ažuriranja ponovnom instalacijom R-a i R Studio okruženja, preporučljivo je da se obrišu sve povezane fascikle (eng. *associated folders*).

505 Pored toga što mašinsko učenje, omogućava automatizaciju mnogih poslova u radu sa podacima, ne postoji još uvek (a možda nikada ni neće postojati) algoritam, koji će moći kompletna proračunavanja da radi bez ljudske intervencije. Ovde se pod automatskim mašinskim učenjem, misli na automatizovane rutine za podešavanje hiperparametara, te za jednostavnu realizaciju relativno velikog broja algoritama u relativno kratkom vremenskom intervalu. [609-610]

Kratko upustvo za rad u R Studio razvojnom okruženju

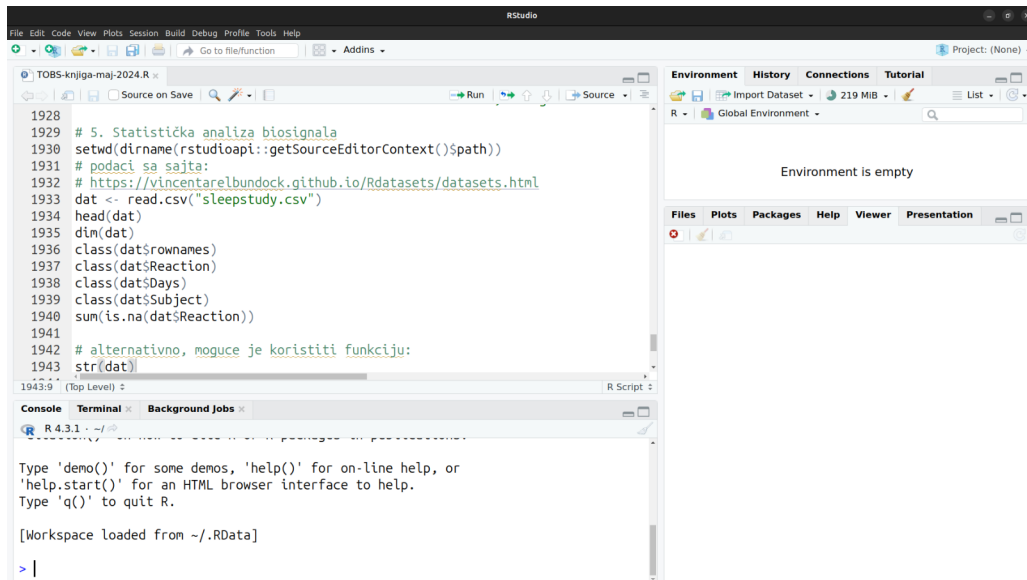
Datoteke sa ekstenzijom .R, u kojima se čuva programski kod, mogu da sadrže komentare pisane ćirilicom, ako se prethodno podesi odgovarajuće enkodiranje koristeći *File* padajući meni i odabirom *Save with encoding* opcije. U slučaju da su komentari pisani na srpskom jeziku (latiničnim ili ćirilničnim pismom), trebalo bi, pre čuvanja skripte sa ekstenzijom .R, da se javi poruka/pitanje sa kakvom vrstom enkodiranja će se čuvati datoteka. Kako je udžbenik pisan latinicom, komentari u R kodu koji je podeljen preko GitHub-a su pisani latiničnim pismom na srpskom jeziku. Takođe, preporuka je da se koristi readr paket prilikom učitavanja podataka iz datoteka [111]. Ako se na odgovarajući način, iskoriste argumenti funkcije iz readr paketa, za učitavanje datoteka, onda je moguće raditi sa tekstualnim podacima, koji su predstavljeni ćirilničnim pismom.

Kada se otvori R Studio razvojno okruženje interfejs je podešen podrazumevano i podseća na interfejs za Matlab, GNU Octave, Pajčarm i Spajder za rad u Pajtonu (Sl. C8). U donjem levom uglu se nalazi prozor sa konzolom⁵⁰⁶, gde postoje tri različite kartice (konzola, terminal i za obavljanje poslova “u pozadini”). U gornjem levom uglu na Sl. C8 je smeštena osnovna skripta (na slici je prikazana skripta sa ekstenzijom .R, ali je moguće uređivati i skripte sa drugim ekstenzijama). U donjem desnom uglu (Sl. C8) se nalazi multifunkcijski prozor čija se kartica *Plots* najčešće koristi za prikaz grafika. Pored *Plots* kartice, moguće je odabrati prikaz datoteka (*Files*), stranice za pomoć (*Help*), listu paketa (*Packages*) ili specijalizovane kartice za pregled prezentacija i drugih rezultata. U gornjem desnom uglu se nalazi kartica okruženje (*Environment*) u kojoj se mogu pregledati promenljive u globalnom okruženju, a alternativni prikaz u ovom prozoru uključuje kartice za pregled prethodnih komandi (*History*), za pregled veza sa bazama podataka ili za rad distribuirane analize podataka (*Connections*) i uputstvo za rad u R-u korišćenjem learnr paketa [611] (kartica *Tutorial*). Raspored prozora prikazan na Sl. C8 nije fiksna i moguće ga je promeniti na više načina.

Jedan način za promenu veličine prozora je prevlačenjem ivica prozora. pomoću levog klika računarskog miša. Takođe, u gornjem desnom uglu svakog od prozora, postoje ikonice da se prozor smanji ili poveća. Dodatno, ispod gornjeg menija, postoji ikonica koja omogućava uređenje prozora (*Pane*

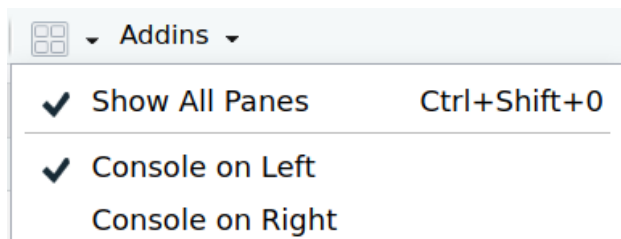
⁵⁰⁶ U nekim programskim jezicima, to je komandni prozor ili šel (eng. *Shell*), a predstavlja prevodilac (eng. *interpreter*) za komande, njihovo izvršavanje i prikaz rezultata.

Layout). Klikom na ikonicu sa četiri kvadrata dobija se padajući meni (Sl. C9) koji se može iskoristiti za uređenje prozora u okviru R Studio okruženja.



Slika C8, Izgled R Studio okruženja sa osnovnim prozorima.

Iako su podrazumevana podešavanja, u potpunosti zadovoljavajuća za prosečan ekran i prosečnog korisnika, vrlo jednostavno se u R Studio interfejsu može pristupiti podešavanjima (padajući meni *Tools* → *Global Options*⁵⁰⁷) ili prečicama koje se mogu koristiti za tastature (*Tools* → *Keyboard Shortcuts*), kako bi rad u ovom okruženju bio udobniji. Prečice se mogu menjati, a u svakom od padajućeg menija (koji neće biti detaljno objašnjen ovde), moguće je u nastavku opcije, videti kako se koriste prečice. Na primer, iz padajućeg menija *Code*, moguće je videti da je prečica sa tastature za komentarisanje ili odkomentarisanje neke linije koda **Ctrl + Shift + C**.



Slika C9, Deo padajućeg menija za uređenje panela u okviru R Studio razvojnog okruženja.

⁵⁰⁷ To je još jedan od načina da se pristupi opciji za uređenje prozora (*Pane Layout* kartica u okviru *Global Options* prozora).

Za one koji su navikli na Matlab/GNU Octave komande:

```
clear all; close all; clc;
```

koje se koriste za brisanje promenljivih, zatvaranje otvorenih prozora za prikaz grafika i za brisanje sadržaja u komandnom prozoru (konzoli), u R-u postoje funkcije koje se mogu koristiti sa istim ciljem, mada to za programiranje u R-u nije uobičajeno. Poziv funkcije:

```
rm(list = ls())
```

omogućava da se brišu sve promenljive, dok se zatvaranje grafika može postići na dva načina:

```
graphics.off()
```

ili :

```
dev.off()
```

a brisanje istorije komandnog prozora je moguće postići korišćenjem funkcije:

```
cat("\014")
```

U R-u se, uobičajeno koristi za brisanje sadržaja konzole, prečica sa tastature Ctrl + L, a za brisanje/zatvaranje prikaza grafika i brisanje promenljivih uobičajeno je koristiti ikonicu sa slikom metlice koja je dostupna u prozorima za pregled promenljivih u okruženju (*Environment*) i za prikaz grafika (*Plots*).

Za pisanje skripti u R Studio okruženju, postoji niz pogodnosti. Jedna se ogleda u tome, što je korisno koristiti skraćenicu sa tastature Ctrl + Enter za pokretanje jedne linije koda, što je ekvivalentno klikom na komandu *Run* u desnom gornjem delu prozora, u kome se skripta nalazi. Dodatno, pre izvršenja pojedinih linija koda, u R Studio prozoru u kome je skripta izvršena dokazi do prikaza greške kao što je prikazano za 262. liniju koda na Sl. C10. Greška je nastala zato što je pogrešno definisan kompleksni broj (korišćeno je kompleksno *j* umesto kompleksnog *i* za definisanje kompleksnih brojeva). Slično, ako se u skripti koja je otvorena u R Studio okruženju koriste funkcije iz biblioteke koja nije ranije instalirana, R Studio automatski prepoznaje da je potrebno instalirati odgovarajuće biblioteke i nudi automatsku opciju za njihovu instalaciju.

```

255 for (letter in x) print(letter)
256
257 # kompleksni brojevi i indeksi for petlje
258 sqrt(-1)
259 sqrt(as.complex(-1))
260 sqrt(-1 + 0i)
261 2 + 3i
262 2 + 3j
263 i**2
264 1i^2

```

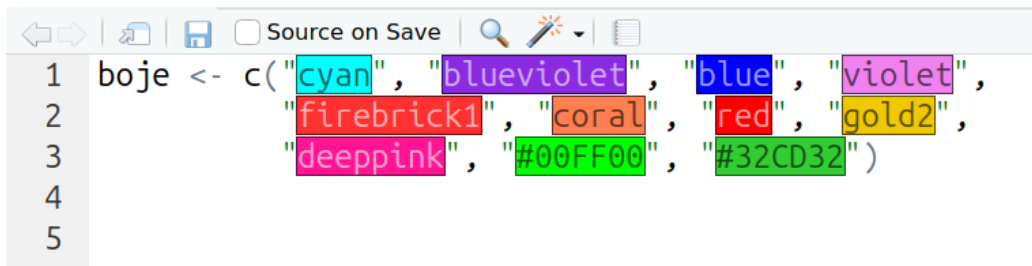
Slika C10, Prikaz automatskog prepoznavanja greške u 262. liniji skripte.

Za strukture, u okviru R programskog jezika, moguće je odabrati takav prikaz da se ne prikaže sadržaj neke od grana, odnosno da se spakuje (eng. *fold*). Na Sl. C11 na levom panelu, prikazana je opcija za pakovanje (*Fold code*), a na desnom panelu prikazan je deo spakovanog koda, koga je jednostavno moguće otpakovati levim klikom računarskog miša na ikonicu koja označava spakovan kod.

<pre> 238 x <- 2 239 y <- if (x > 3) { 240 "veceOd3" 241 } else { 242 "jeIliJednako3" 243 } 244 y </pre>	<pre> 238 x <- 2 239 y <- if (x > 3) { 240 "veceOd3" 241 } else { 242 "jeIliJednako3" 243 } 244 y 245 246 x <- 7 </pre>
---	---

Slika C11, Prikaz opcije za pakovanje koda (*Fold code*) na levom panelu i prikaz ikonice koja predstavlja spakovan kod na desnom panelu.

Dodatna pogodnost u R Studio okruženju je i u tome što dolazi automatski do prikaza boje čiji je string unet u kodu. Ova opcija radi podjednako za prikaz boja po imenu ili za unos heksadecimalnih brojeva koji odgovaraju različitim bojama. Ova osobina R Studio okruženja je prikazana na Sl. C12.



```
1 boje <- c("cyan", "blueviolet", "blue", "violet",  
2          "firebrick1", "coral", "red", "gold2",  
3          "deeppink", "#00FF00", "#32CD32")  
4  
5
```

Slika C12, Prikaz boja u R skripti u okviru R Studio okruženja. Slična funkcionalnost R Studio okruženja je pokazana na Sl. 7.

CIP - Каталогизација у публикацији
Народна библиотека Србије, Београд

54:004(075.8)(0.034.2)
004.43R(075.8)(0.034.2)

МИЉКОВИЋ, Надица, 1986-

Analiza biosignala sa praktičnim primerima u programskom jeziku R
[Електронски извор] : [електронски удџбеник] / Nadica Miljković. - Beograd :
Univerzitet, Elektrotehnički fakultet, 2024 (Beograd : N. Miljković). - 1
elektronski optički disk (CD-ROM) ; 12 cm

Sistemski zahtevi: Nisu navedeni. - Nasl. sa naslovne strane dokumenta. - Tiraž
50. - Napomene i bibliografske reference uz tekst. - Sadrži bibliografiju.

ISBN 978-86-7225-098-5

а) Биомедицина -- Информациона технологија б) Програмски језик "R"

COBISS.SR-ID 148889609
