



Manipulacija podacima, kontrolne strukture i dplyr paket na primeru studije spavanja

Dr Nadica Miljković, vanredna profesorka

kabinet 68, nadica.miljkovic@etf.bg.ac.rs

Datum i vreme: zadaci

```
> # pocetak letnjeg raspusta na ETF
> datum1 <- as.Date("2017-07-06")
> # kraj letnjeg raspusta na ETF
> datum2 <- as.Date("2017-08-20")
> datum2 - datum1
Time difference of 45 days
> |
```

- Koristeći funkciju *Sys.time()* odrediti trenutno vreme i sačuvati rezultat u promenljivoj *x*.
- Potom, proveriti strukturu objekta i klasu objekta *x* i za kraj eksplicitno prikazati rezultat na ekranu (*print()* funkcija).
- Korišćenjem *as.Date()* funkcije za rad sa stringovima kao na slici izračunati npr. koliko je dana prošlo od dana kada ste upisali master akademske studije ili neki drugi. Na slici je dat primer razlike u danima između dva datuma.

Kompleksni brojevi

```
> x <- 2 + 3i
> y <- 1 - 8i
> class(x)
[1] "complex"
> class(y)
[1] "complex"
> x - y
[1] 1+11i
> x + y
[1] 3-5i
> x*y
[1] 26-13i
> is.complex(x*y)
[1] TRUE
> Re(y)
[1] 1
> Im(x)
[1] 3
```

```
> sqrt(-1)
[1] NaN
Warning message:
In sqrt(-1) : NaNs produced
> sqrt(as.complex(-1))
[1] 0+1i
> |
```

- Testirati rad prikazanih funkcija.
- Posebno obratiti pažnju na:
 - *Re()* i *Im()* funkcije za izdvajanje realnog i imaginarnog dela kompleksnog broja.
 - NaN (eng. *Not A Number*) rezultat na slici.

Zadatak #1

```
# zavisnost sd od broja merenja
sdr <- vector(mode = "numeric", length = 10L)
sdr[1] <- abs(sd(rnorm(2)) - 1)
sdr[2] <- abs(sd(rnorm(4)) - 1)
sdr[3] <- abs(sd(rnorm(8)) - 1)
sdr[4] <- abs(sd(rnorm(16)) - 1)
sdr[5] <- abs(sd(rnorm(32)) - 1)
sdr[6] <- abs(sd(rnorm(64)) - 1)
sdr[7] <- abs(sd(rnorm(128)) - 1)
sdr[8] <- abs(sd(rnorm(256)) - 1)
sdr[9] <- abs(sd(rnorm(512)) - 1)
sdr[10] <- abs(sd(rnorm(1024)) - 1)

brMerenja <- c(2, 4, 8, 16, 32, 64, 128, 256, 512, 1024)

plot(brMerenja, sdr)
```

- Zavisnost *sd* (standardne devijacije) od broja merenja realizovati sa *for* petljom.
- Na slici je dat kod zadatka u kome je potrebno uneti izmene tj. dodati *for* petlju.

Zadatak #2

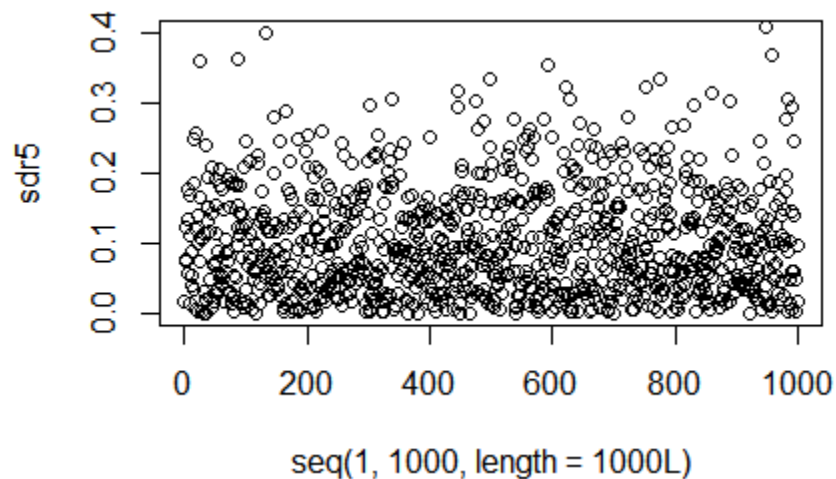
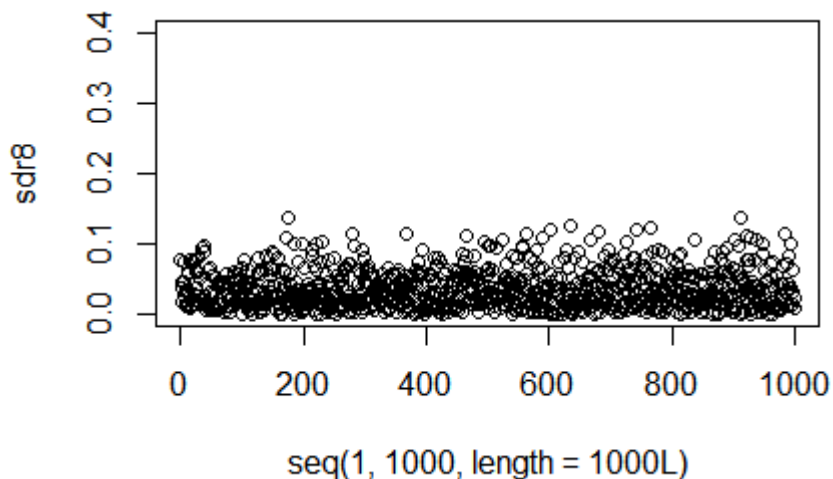
- Pokretanjem koda iz zadatka #1 (*for* petlja) uvek se dobija različito rešenje zbog *rnorm()* funkcije koja prilikom svakog pokretanja daje različit izlaz.
- Kompletan kod je potrebno ponoviti 1000 puta i sačuvati različite vrednosti koje se dobijaju pokretanjem programa za *brMerenja* = 32 i *brMerenja* = 256. Uporediti na graficima ova dva rezultata. Ograničiti prikaz na ordinati grafika (*y* osa) od 0 do 0.4.
- NAPOMENA: Iako je u rešenju poslednjeg zadatka korišćeno slovo “*i*” za iterator u petlji, u R-u (ali i Matlab-u i mnogim drugim programskim jezicima) “*i*” je rezervisano za kompleksne brojeve (pogledati sliku). **Potrebno je koristiti neki drugi naziv za iterator!**

Rešenje #2

```
> # zadatak #2 za for petlju
> sdr <- vector(mode = "numeric", length = 10L)
> brMerenja <- vector(mode = "numeric", length = 10L)
> sdr5 <- vector(mode = "numeric", length = 1000L)
> sdr8 <- vector(mode = "numeric", length = 1000L)
> for (ind in 1:1000) {
+   for (i in 1:length(sdr)) {
+     sdr[i] <- abs(sd(rnorm(2^i))) - 1)
+     brMerenja[i] <- 2^i
+   }
+   sdr5[ind] <- sdr[5]
+   sdr8[ind] <- sdr[8]
+ }
>
> plot(seq(1,1000, length = 1000L), sdr5, ylim = c(0, 0.4))
> plot(seq(1,1000, length = 1000L), sdr8, ylim = c(0, 0.4))
> |
```

Da li je moglo da se izbegne ugnježdavanje petlji?

Zašto je $sdr5 > sdr8$? Proveriti njihove srednje vrednosti i standardne devijacije.



Zadatak #3: *sleep study* (SS)

	X	Reaction	Days	Subject
1	1	249.5600	day 6	308
2	2	258.7047	day 0	308
3	3	250.8006	day 0	308
4	4	321.4398	day 0	308
5	5	356.8519	day 0	308
6	6	414.6901	day 0	308
7	7	382.2038	day 0	308
8	8	290.1486	day 0	308
9	9	430.5853	day 0	308
10	10	466.3535	day 0	308

- Podaci koji se koriste u ovom zadatku su dostupni na sajtu: <https://vincentarelbundock.github.io/Rdatasets/datasets.html> (pristupljeno 17.03.2024), a nalaze se u datoteci “sleepstudy.csv” (**Reaction times in a sleep deprivation study**).
- Učitati podatke u R, potom ih prikazati pomoću funkcije `View()`. Proveriti tip podataka primenom funkcije `class()`.
- Najpre proveriti koliko ima subjekata koji su snimani 0-tog, 1-og, ..., i 9-og dana. *HINT: koristiti `unique()` funkciju.*

Zadatak #4: SS

	Num	Reaction	Days	Subject
1	10	466.3535	9	308
2	20	237.3142	9	309
3	30	247.5153	9	310
4	40	354.0487	9	330
5	50	371.5811	9	331
6	60	253.8644	9	332
7	70	362.0428	9	333
8	80	377.2990	9	334
9	90	237.2466	9	335
10	100	458.0167	0	337

- Preimenovati kolonu u *data frame*-u iz *X* u *Num* i snimiti u istu promenljivu *dat*.
- Preurediti podatke tako da se prvo pojavljuju podaci koji su snimani 9-og dana, a potom podaci koji su snimani 0-og dana. Prikazati podatke pomoću funkcije *View()*. Trebalo bi da se dobije prikaz kao na Sl. Koristiti pipeline.

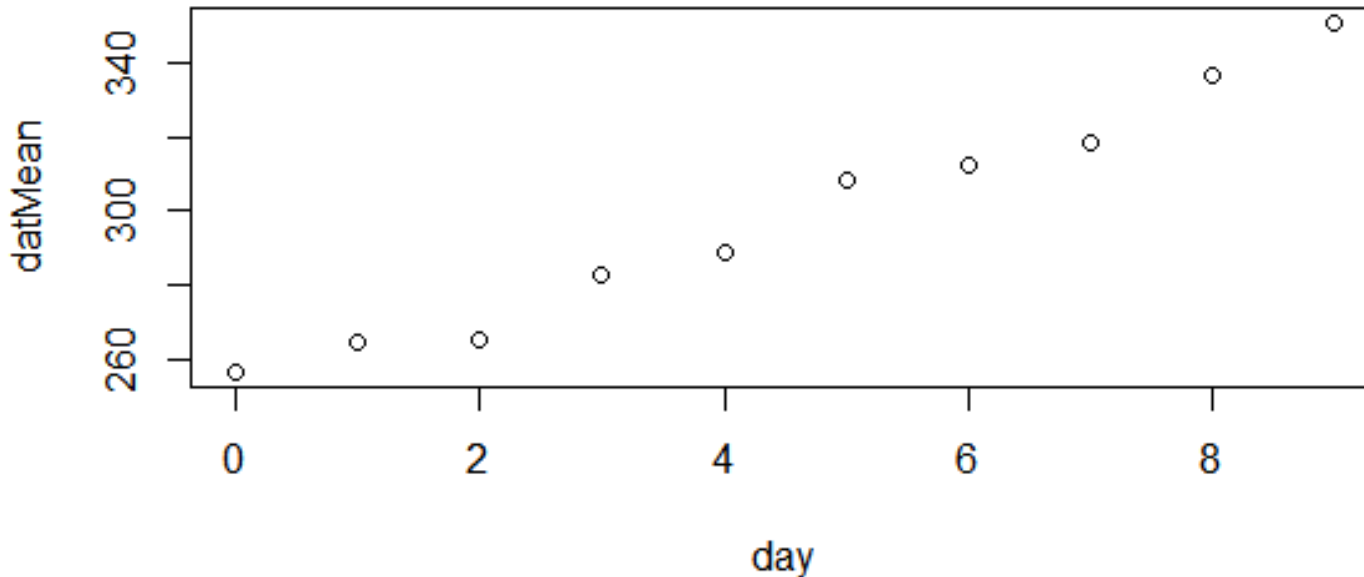
Rešenje: zadatak #4 SS

```
> # zadatak 2
> dat <- rename(dat, Num = X)
> dat <- arrange(dat, desc(Days))
> # ili PIPELINE
> dat <- dat %>% rename(Num = X) %>% arrange(desc(Days))
Error: Unknown variables: X.
> view(dat)
> |
```



Zašto je došlo do greške?

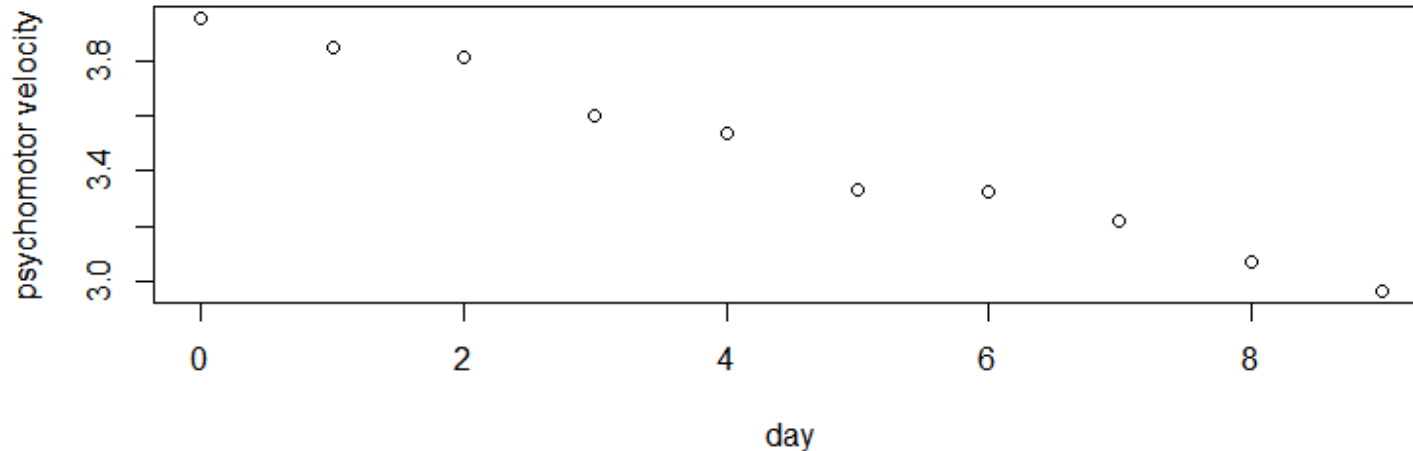
Zadatak #5: SS



- Projektovati matricu *datN* koja sadrži podatke tipa *numeric*, a u kojoj su u vrstama smešteni podaci o različitim danima, a u kolonama podaci koji su mereni na različitim ispitanicima (*Reaction*).
 - *HINT*: koristiti funkcije *filter()*, *select()* i *unlist()*, ali i *for* petlju.
- Potom odrediti srednje vrednosti *Reaction* parametra za svaki dan pojedinačno i snimiti u promenljivu *datMean* (postoji *rowMeans()* funkcija) i prikazati rezultat kao na slici.

Dodatno (zadatak #5)

Averaged data for 18 subject for 3 h of sleep restriction



- Da bi prikazan rezultat bio validan, potrebno je uvesti sledeće korekcije:
 - Prikazati *error bar* grafik sa dodatim standardnim devijacijama ili box plot. **To kada budemo radili ggplot paket ...**
 - *Reaction* parametar je potrebno predstaviti u drugom obliku.
 - Označiti ose na grafiku.
- U radu: Belenky, Gregory, et al. "Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: A sleep dose-response study." *Journal of sleep research* 12.1 (2003): 1-12, navedeno je da se parametar *Reaction* koristi kako bi se izračunala brzina reagovanja prema formuli: $(1/Reaction)*1000$.
- Uneti ovu izmenu i nacrtati prethodni grafik i označiti kao na slici.

Rešenje

```
> datNnew <- (1 / datN)*1000
> datNewMean <- rowMeans(datNnew)
>
> plot(day, datNewMean, xlab = "day", ylab = "psychomotor velocity",
+       main = "Averaged data for 18 subject for 3 h of sleep restriction")
> |
```

- Kako je moguće iskoristiti funkcije dplyr paketa da se realizuje ovo rešenje?
- Koju funkciju je potrebno koristiti?
- U kojim jedinicama su izražene brzine reagovanja?

Zaključak studije

- U radu je navedeno da dolazi do konstantnog smanjenja brzine reagovanja kada postoji san od 3 sata i da je potrebno više od 2 dana za period oporavka.
- To se ne dešava kada je san ograničen na 5 i 7 sati, jer u tom slučaju PVT (eng. *Psychomotor Vigilance Test*) dostiže konstantnu brzinu. Za 9 sati sna je utvrđeno da nema izmena u odnosu na 8 sati sna.
- Zaključci koji su doneti za usrednjene PVT brzine su bili isti kao i za 10% najbržih odgovora po ispitaniku.
- Takođe, ovo označava da način života (broj sati provedenih u snu) može imati trajne posledice po brzinu reagovanja i procese u mozgu.

Funkcije i komande za danas

- `as.Date()`
- `Sys.time()`
- `print()`
- `class()`
- `str()`
- `abs()`
- `sd()`
- `rnorm()`
- `plot()`
- `c()`
- `for`
- `vector()`
- `length()`
- `sqrt()`
- `is.complex()`
- `Re()`
- `Im()`
- `as.complex()`
- `mean()`
- `View()`
- `unique()`
- `read.csv()`
- `%>%()`
- `rename()`
- `arrange()`
- `filter()`
- `select()`
- `unlist()`
- `matrix()`
- `rowMeans()`

Interaktivno vežbanje



- Pokrenuti SWIRL i uraditi lekciju pod nazivom *Sequences of numbers*.
 - NAPOMENA: Pre pokretanja *swirl()* komande potrebno je učitati i biblioteku/paket komandom *library(swirl)*.