

VEŽBA 8

VIZUELIZACIJA PODATAKA

NAPOMENA: Kod sačuvati u datoteci pod nazivom "vezba8.R". Voditi računa o prikazu signala i rezultata: svi grafici bi trebalo da imaju označene ose, naslov, unete jedinice (gde god je to moguće), kao i legende.

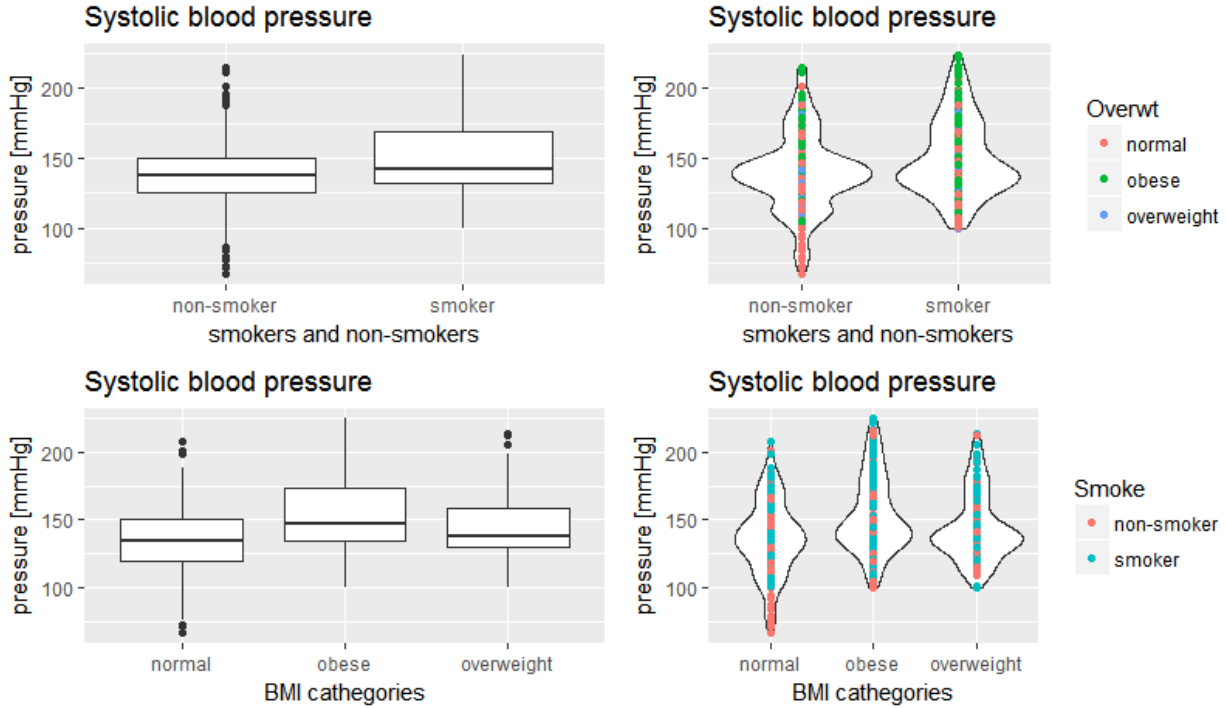
Zadaci za rad:

1. Učitati podatke koji sadrže sistolni krvni pritisak za 500 ljudi sa sledećeg linka: <https://raw.githubusercontent.com/vincentarelbundock/Rdatasets/master/csv/Stat2Data/Blood1.csv>, pristupljeno 10.04.2024. godine (opciono: podaci se mogu sačuvati i u .csv datoteci i učitati iz radnog direktorijuma). Za metapodatke pogledati: <https://vincentarelbundock.github.io/Rdatasets/doc/Stat2Data/Blood1.html> (pristupljeno 10.04.2024), a potom proučiti organizaciju podataka (koristiti *summary()* funkciju).
2. Preimenovati numeričke promenljive u kolonama "Smoke" i "Overwt" u kategoričke promenljive na sledeći način:
 - a. u "Smoke" koloni bi trebalo da postoje vrednosti: "smoker" i "non-smoker",
 - b. u "Overwt" koloni bi trebalo da postoje vrednosti: "normal", "overweight" i "obese". Po želji, sortirati podatke tako da im je redosled "normal", "overweight" i "obese" u ovu kategoričku promenljivu.

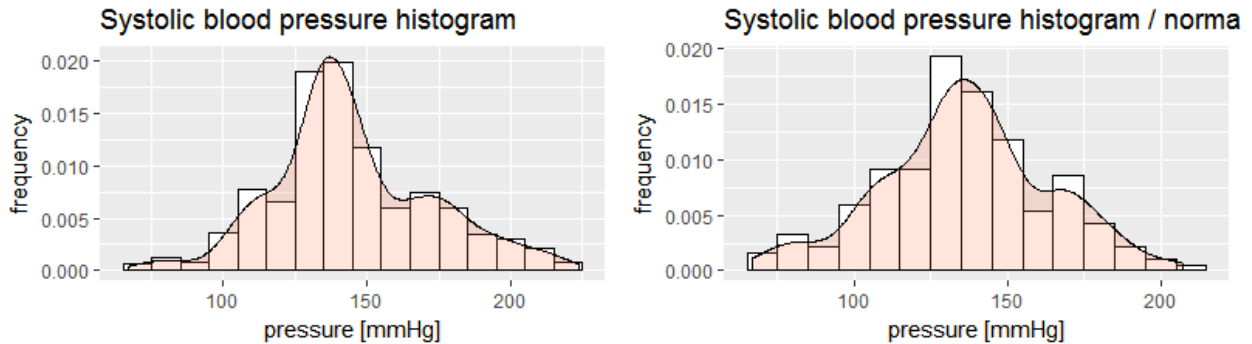
Trebalo bi da se dobije struktura kao na slici:

```
> head(dat)
  X SystolicBP      Smoke      Overwt
1 1           133 non-smoker      obese
2 2           115 non-smoker      normal
3 3           140 non-smoker  overweight
4 4           132 non-smoker      obese
```

3. Prikazati *boxplot* grafike (primenom funkcija u ggplot2 paketu) za sistolni pritisak pušača i nepušača i za sistolni pritisak za tri grupe u "Overwt" koloni podataka. Za sve navedene slučajeve dodati i *violin* grafik. Trebalo bi da se dobiju grafici kao na sledećim slikama:



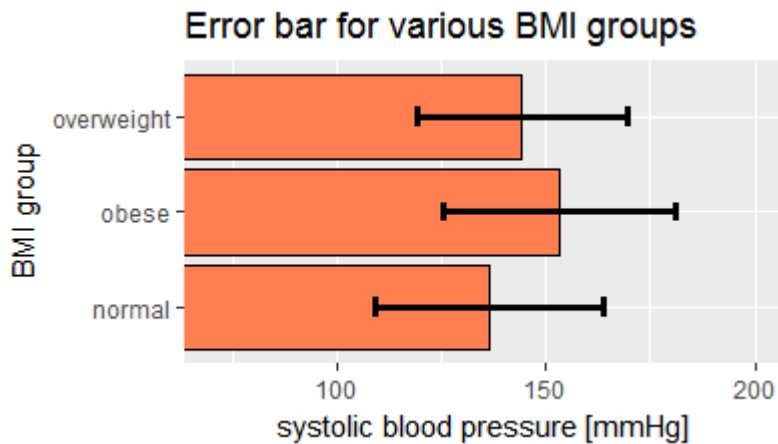
4. Prikazati histogram sistolnog krvnog pritiska zajedno sa procenjenom funkcijom gustine verovatnoće. Dodatno, prikazati histogram i za "Overwt" == "normal". Grafici bi trebalo da izgledaju kao na slici:



5. Prikazati na *bar* dijagramu broj ispitanika za svaku od "Overwt" grupa (opciono: prikazati barove horizontalno). Grafici koji bi trebalo da se dobiju su prikazani na sledećim slikama:



DODATNI ZADATAK (nije obavezan): Prikazati pomoću *error bar* grafika, usrednjene vrednosti sistolnog pritiska sa standardnim devijacijama za sve tri grupe ispitanika (HINT: koristiti funkcije iz *dplyr* paketa za formiranje novog *data frame*-a koji sadrži neophodne podatke).



PAZNIČNI ZADATAK: Instalirati paket *onion* i nacrtati podatke *bunny*.

INFORMACIJA: Za one koji su navikli na Matlab / GNU Octave funkcije "clear all", "close all" i "clc", mogu koristiti sledeće funkcije u R-u: "clear all" == "rm(list = ls())", "close all" == "graphics.off()" ili "close all" == "dev.off()" i "clc" == "cat("\014)". U R-u se uobičajeno koriste za "clc" == prečica sa tastature Ctrl + L, a za "close all" i "clear all" uobičajeno je koristiti ikonisu sa slikom metlice:



OPCIONO: Uraditi SWIRL lekcije pod nazivom: "lapply & sapply" i "vapply & tapply".