

VEŽBA 8

LINEARNA REGRESIJA I VIZUELIZACIJA

NAPOMENA: Sve zadatke realizovati u fajlu "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.

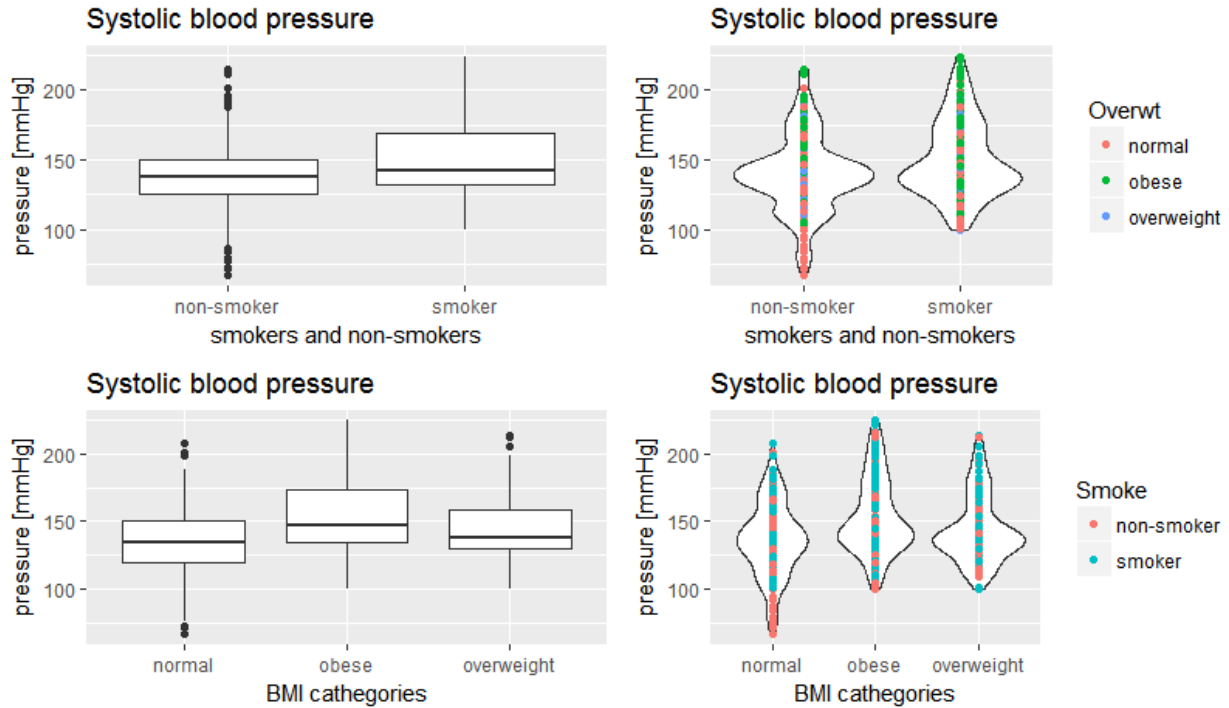
Zadaci za rad:

1. Učitati podatke sistolnog krvnog pritiska za 500 ljudi sa sajta: <https://raw.githubusercontent.com/vincentarelbundock/Rdatasets/master/csv/Stat2Data/Blood1.csv>. Detalji i protokol merenja ovih podataka mogu se naći na linku: <https://vincentarelbundock.github.io/Rdatasets/doc/Stat2Data/Blood1.html>. Proučiti organizaciju ovih podataka (između ostalih, 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". Sortirati podatke tako da im je redosled "normal", "overweight" i "obese" u ovu kategoričku promenljivu.

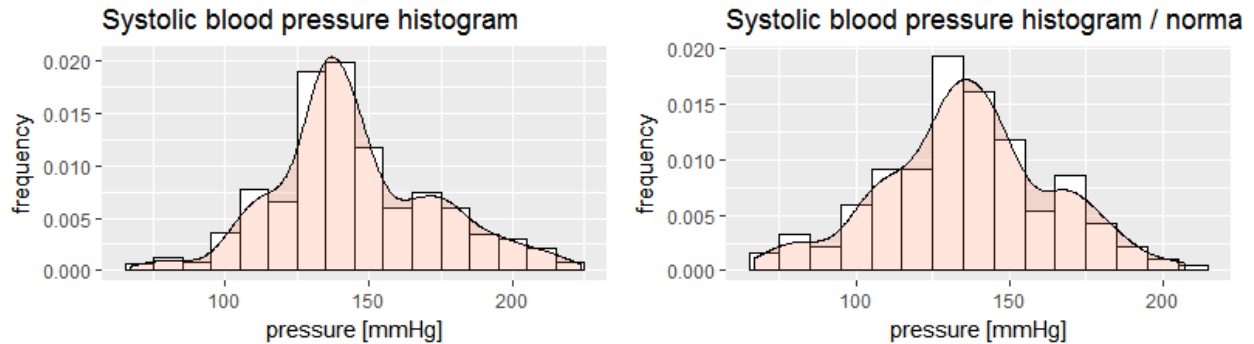
Trebalo bi da se dobije struktura kao na slici:

```
> 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. Obavezno označiti ose. Za sve navedene slučajeve dodati i *violin* grafik. Grafici bi trebalo da izgledaju kao na sledećim slikama:



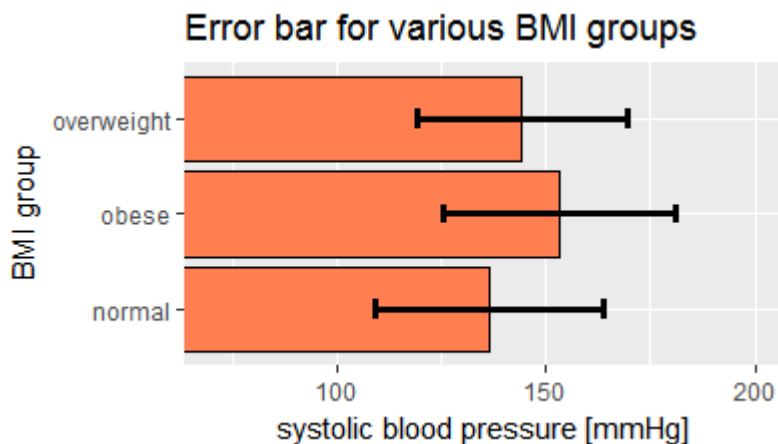
4. Prikazati histogram sistolnog krvnog pritiska zajedno sa procenjenom funkcijom gustine verovatnoće. Dodatno, prikazati isti grafik za "Overwt" == "normal". Grafici bi trebalo da izgledaju kao na slici:



5. Prikazati na *bar* dijagramu broj ispitanika za svaku od "Overwt" grupa (opciono: prikazati barove horizontalno). Grafik koji bi trebalo da se dobije dat je na slici:



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



INFORMACIJA: Za Matlab korisnice i korisnike o funkcijama "clear all", "close all" i "clc" koje se mogu koristiti i u R-u na sledeći način: "clear all" == "rm(list = ls())", "close all" == "graphics.off()" ili "close all" == "dev.off()" i "clc" == "cat("\014)". Za R korisnike, uobičajeno je "clc" == sa tastature Ctrl + L, a za "close all" i "clear all" uobičajeno je koristiti metlicu sa slike:



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