

R



Vežbe 6



Pretvaranje numeričkih podataka u kategorijske

- U programskom jeziku R je moguće određene intervale vrednosti numeričke promenljive pretvoriti u odgovarajuće kategorije

```
bonusi = c(12, .4, 5, 2, 50, 8, 4, 1, 4, .25)

#funkcija cut ce za svaku promenljivu da vrati kom od
#definisanih opsega pripada
opsezi = cut(bonusi, breaks = c(0, 1, 5, max(bonusi)))
opsezi

tabela = table(opsezi)
tabela
```



Pretvaranje numeričkih podataka u kategorijske

- Ovakve podatke je dalje moguće posmatrati grafički preko bar i pie grafika

```
barplot(tabela, col = c("red", "green", "blue"))  
title("Raspodela bonusa")  
  
pie(tabela, col = c("red", "green", "blue"))  
title("Raspodela bonusa")
```



Zadatak 1

- Dat je sledeći skup podataka:

{0.23, 1.8, 5.65, 8.32, 2.4, 10.12, 8.9, 12.3, 16.3, 7.1, 19.8}

- Podeliti ove podatke u kategorije, i to tako da u prvoj kategoriji budu podaci iz prvog kvartila, u drugoj podaci između prvog kvartila i medijane, u trećoj podaci između medijane i trećeg kvartila, a u četvrtoj podaci veći od trećeg kvartila. Rezultate prikazati na bar chart-u

```
podaci = c(0.23, 1.8, 5.65, 8.32, 2.4, 10.12, 8.9, 12.3,
16.3, 7.1, 19.8)

opsezi = cut(podaci, breaks = c(min(podaci), quantile(podaci,
.25), median(podaci), quantile(podaci, .75), max(podaci)))
tabela = table(opsezi)
tabela

barplot(tabela, col = c("red", "green", "blue", "cyan"))
```



Histogram

- Sličan je bar chart-u, koristi se za pregledno prikazivanje velike količine podataka. Podaci koji se prikazuju su obično numerički.

```
podaci = c(29.6, 28.2, 19.6, 13.7, 13.0, 7.8, 3.4, 2.0, 1.9,  
1.0,.7, .4, .4, .3, .3, .3, .3, .2, .2, .2, .1, .1, .1, .1)  
  
#sluzi da prikaze veliku kolicinu podataka na pregledan nacin  
hist(podaci)  
  
#prikazuje linije ispod histograma gde se nalaze podaci  
#posto podaci mogu da se preklapaju, jitter linije pomera po  
#malo, da bi se videlo gde je najvise podataka  
rug(jitter(podaci))
```



Histogram

- Moguće je uticati na broj stubova koji se iscrtavaju na histogramu

```
#moguće je zadati iz koliko stubova se sastoji histogram  
#ovo se uzima kao "preporuka"  
hist(podaci, breaks = 10)  
  
#moguće je zadati koji će opsezi da budu predstavljeni  
hist(podaci, breaks = c(0, 1, 2, 3, 4, 5, 10, 20,  
max(podaci)))
```

Zadatak 2



- Data je sledeća lista podataka:

{12, 3.2, 5.9, 89.1, 92, 12, 3.5, 8.2, 1.3, 0.2, 1.3, 150.3, 220.1}

- Na histogramu predstaviti podatke u sledećim opsezima:

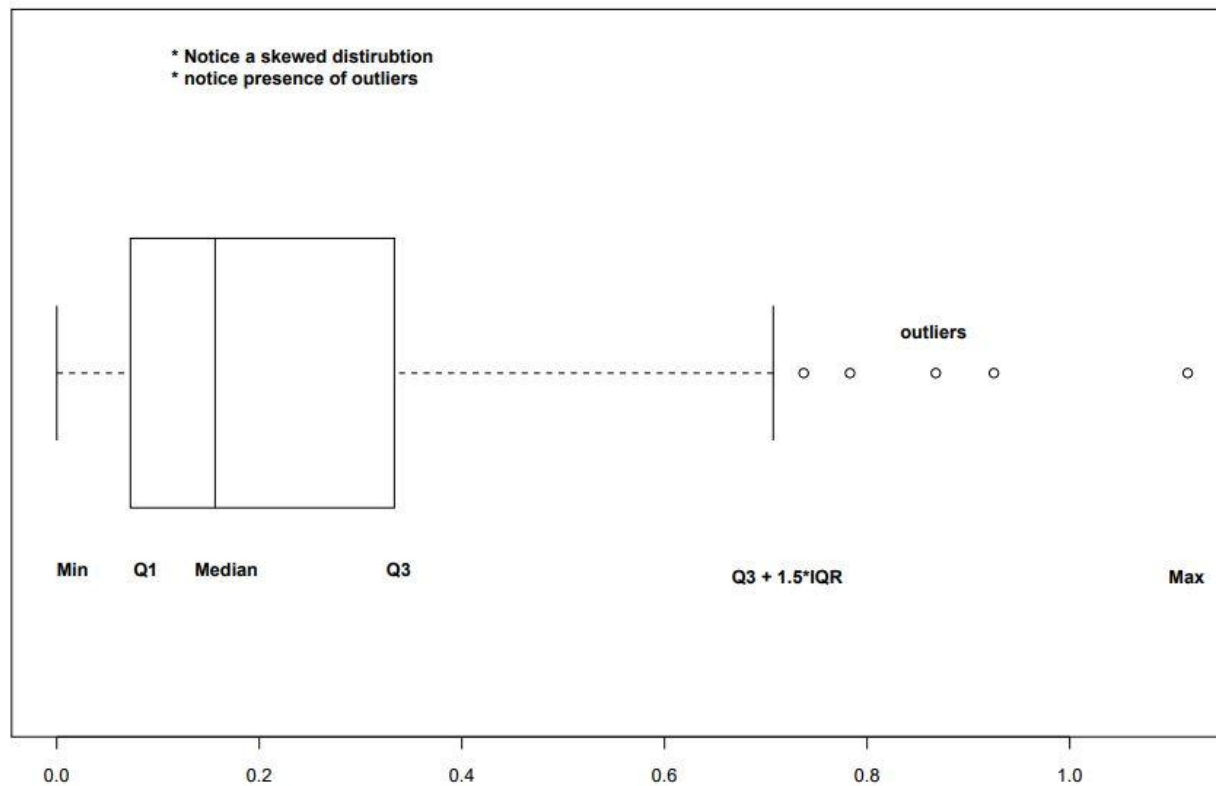
(min, min + 5], [min + 5, mean(podaci)], [mean(podaci), max(podaci))

```
podaci = c(12, 3.2, 5.9, 89.1, 92, 12, 3.5, 8.2, 1.3, 0.2, 1.3, 150.3, 220.1)
hist(podaci, breaks = c(min(podaci), min(podaci) + 5, mean(podaci), max(podaci)))
```

Boxplot



- Boxplot služi da grafički prikaže informacije o skupu koji se analizira: kvartile, interkvartilnu razliku, medijanu, netipične vrednosti, ekstremne vrednosti...





Boxplot

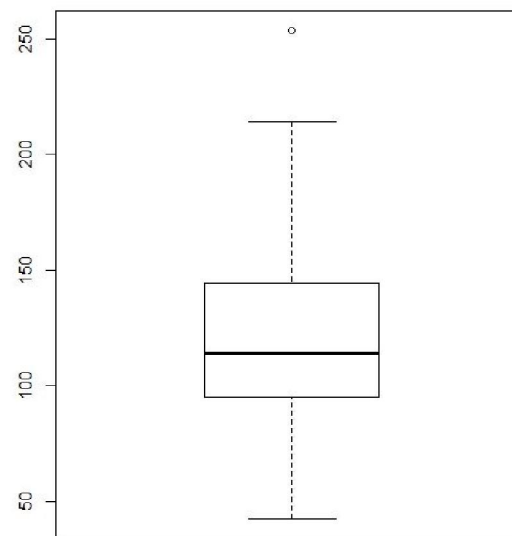
```
#ovaj paket sadrzi neke funkcije koje ce se koristiti,  
#kao i dodatne data frame-ove  
install.packages("UsingR")  
library(UsingR)  
  
movies  
names(movies)  
#pozivom funkcije attach omogucava se da se kolonama  
#data frame-a pristupi po imenima, kao da su promenljive  
attach(movies)  
  
#podrazumevano se boxplot crta vertikalno  
boxplot(current, main = "Trenutno")  
boxplot(gross, main = "Ukupno", horizontal = T)  
  
#funkcija suprotna funkciji attach  
detach(movies)
```



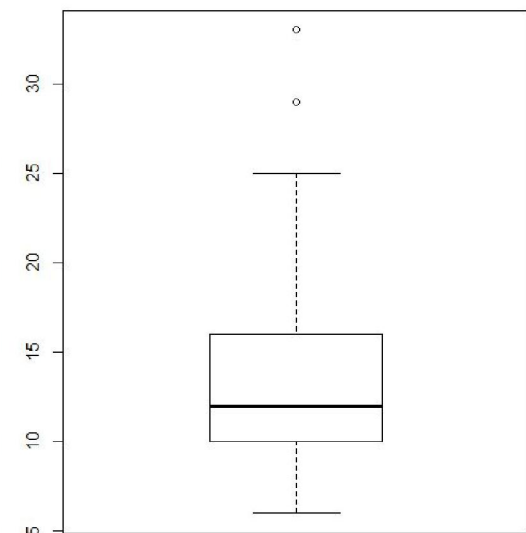
Zadatak 3

- Data su dva ugrađena skupa podataka: „south“ i „aid“. Predstaviti oba ova skupa na boxplot-ovima. Koji od ova dva skupa podataka je simetričniji? Koji ima više netipičnih vrednosti?

```
boxplot(aid)  
boxplot(south)
```



boxplot „aid“



boxplot „south“



Boxplot i histogram

- Funkcija **simple.hist.and.boxplot** na istom grafiku prikazuje histogram i boxplot kreirane na osnovu prosleđenih podataka

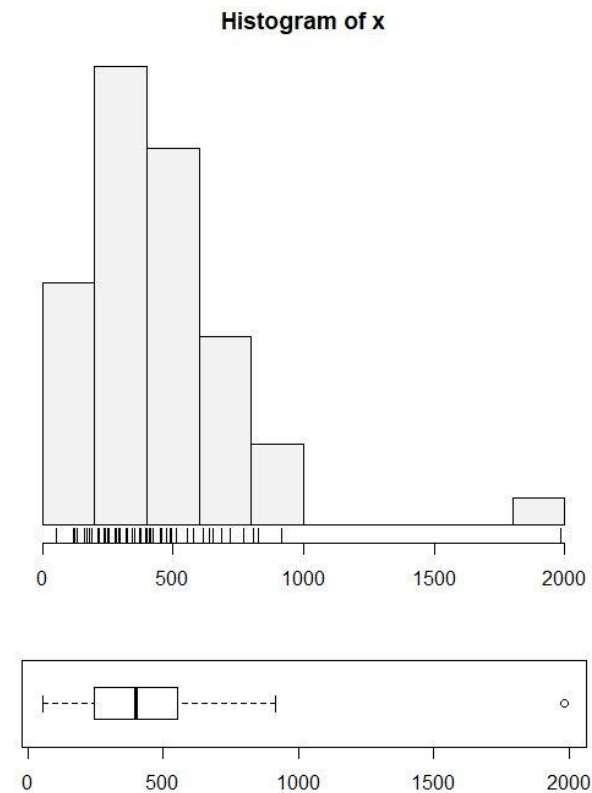
```
podaci = c(0, -.25, .2, .32, -.02, .35, 2, .3, .5, .32, .45,  
.78, 1.0)  
summary(podaci)  
simple.hist.and.boxplot(podaci)
```



Zadatak 4

- Iz data frame-a **crime** na spojenom histogramu i boxplot-u prikazati kolonu „y1983“. Analizirati prikazane rezultate.

```
simple.hist.and.boxplot(crime$y1983)
```



Zadatak 5



- Na histogramu predstaviti podatke iz skupa podataka „math“. Na osnovu grafika oceniti kolika je srednja vrednost, medijana i standardna devijacija. Proveriti pretpostavke odgovarajućim funkcijama programskog jezika R.

```
hist(math)  
  
mean(math)  
  
median(math)  
  
sd(math)
```

