

R



Vežbe 7



Upareni podaci

- Često je interesantno posmatrati vezu između dve promenljive
- Primer: veza između visine i težine, poređenje efekata starog i novog leka...

```
#komanda table ce tabelarno pokazati uparene podatke, kao sto je to
#cinila sa pojedinacnim podacima
smokes = c("Y", "N", "N", "N", "Y", "N", "Y", "Y", "Y", "N", "Y")
amount = c(1, 2, 2, 3, 3, 3, 1, 2, 1, 3, 2)

#ovako dati podaci su vrlo nepregledni
table(smokes, amount)
```

Upareni podaci



- Funkcija **prop.table** daje pregled podataka po frekvencijama, što je dosta preglednije i razumljivije

```
#funkcija prop.table ce nam dati pregled po frekvencijama
podaci = table(smokes, amount)

prop.table(podaci, 1) #po vrstama
prop.table(podaci, 2) #po kolonama
prop.table(podaci) #svi
```



Grafički prikaz

- Ovi podaci se često prikazuju grafički radi lakše analize. Jedan od grafika koji se koristi u ove svrhe jeste barplot

```
#iscrtano po amount
barplot(table(smokes, amount))

#iscrtano po smokes
barplot(table(amount, smokes))

#iscrtani stubci jedan do drugog, sa legendom
smokes = factor(smokes)
barplot(table(smokes, amount), beside = T, legend.text = T)

barplot(table(amount, smokes), main = "table(amount, smokes)", beside =
T, legend.text = c("< 5", "5-10", "> 10"))
```



Zadatak 1

- Dati su sledeći podaci o ispitanicima:

godine_studiranja = {2, 1, 4, 1, 5, 2, 5, 1, 6, 4, 5}

zaposlen = {"N", "N", "Y", "N", "Y", "N", "Y", "N", "Y", "Y", "Y"}

- Prikazati ove podatke na bar plot-u i analizirati dobijeni grafik.

```
godine_studiranja = c(2, 1, 4, 1, 5, 2, 5, 1, 6, 4, 5)
zaposlen = c("N", "N", "Y", "N", "Y", "N", "Y", "N", "Y", "Y", "Y")

barplot(table(b, a), beside = T, legend.text = c("NE", "DA"),
main = "")
barplot(table(a, b), beside = T)
```



- Numerički podaci koji se posmatraju mogu biti razdvojeni na kategorije koje se posmatraju razdvojeno

```
#x je jedna kategorija, y je druga kategorija  
x = c(1, 0, 4, 5, 8, 11, 3, 2, 1, 4)  
y = c(4, 5, 3, 8, 7, 8, 10, 7, 9)  
  
boxplot(x, y)
```



- Ovi podaci takođe mogu biti skupljeni u jednu listu
- U tom slučaju, u posebnoj listi bi morale da postoje naznačene kategorije za svaki od podataka iz prve liste

```
kolicina = c(5, 5, 5, 13, 7, 11, 11, 9, 8, 9, 11, 8, 4, 5, 9, 5, 10, 5, 4, 10)
kategorija = c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2)

#~ - "prikaz kolicine po kategoriji"
boxplot(kolicina ~ kategorija)
```

Zadatak 2



- U jednoj listi date su cene proizvoda, dok je u drugoj data njihova podela na kategorije. Na boxplot-ovima predstaviti cene proizvoda po kategorijama i analizirati dobijene rezultate.

cene = {10, 12, 3, 5, 8, 17, 23, 19, 1, 5, 14, 65, 87, 53, 23, 11, 6, 23, 18, 91, 78, 15}

kategorije = {1, 1, 1, 2, 2, 1, 2, 1, 3, 2, 1, 3, 3, 3, 2, 2, 2, 2, 1, 1, 1, 1}

```
cena = c(10, 12, 3, 5, 8, 17, 23, 19, 1, 5, 14, 65, 87, 53, 23, 11, 6, 23, 18, 91, 78, 15)
kategorija = c(1, 1, 1, 2, 2, 1, 2, 1, 3, 2, 1, 3, 3, 3, 2, 2, 2, 2, 1, 1, 1, 1)

boxplot(cena ~ kategorija)
```

Uparene numeričke varijable



- Poređenje uparenih numeričkih varijabli može se vršiti na različite načine.
- Ukoliko se smatra da su one nezavisne, onda treba porediti njihove raspodele
- Ako se očekuje veza između promenljivih, grafici sa parovima tačaka su bolji izbor

```
library(UsingR)

attach(home)
names(home)

#vrlo tesko je analizirati ovakve podatke
boxplot(old, new)

#funkcija scale vraca skalirane podatke sa dodatnim atributima
#pomocu [,1] se izdvajaju samo podaci
boxplot(scale(old)[,1], scale(new)[,1])
```



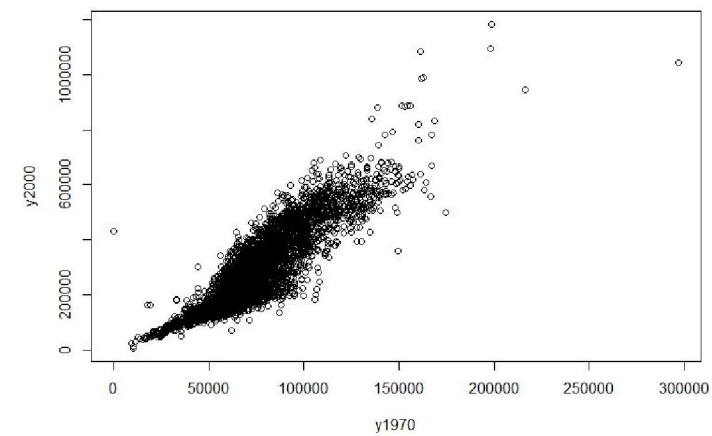
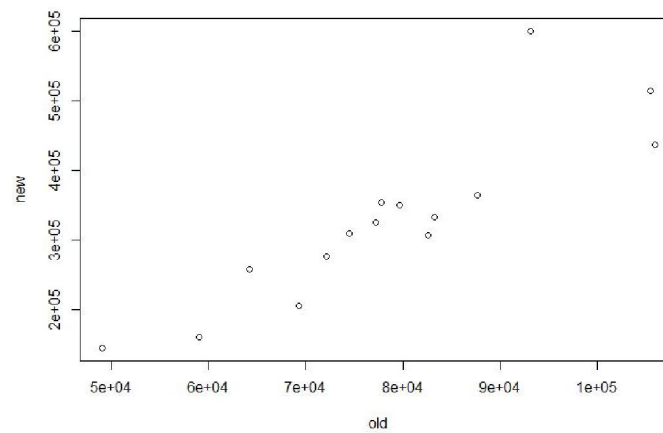
- Gustina raspodele predstavlja verovatnoću da će promenljiva uzeti neku vrednost, i preko nje je moguće vršiti analizu nekog skupa podataka

```
#poredjenje gustine raspodele na grafiku  
attach(homedata)  
  
plot(density(y1970), col = "blue")  
lines(density(y2000), col = "red")
```



- Scatter plot-ovi se koriste da prikažu vezu između parova podataka

```
plot(old, new)  
plot(y1970, y2000)
```

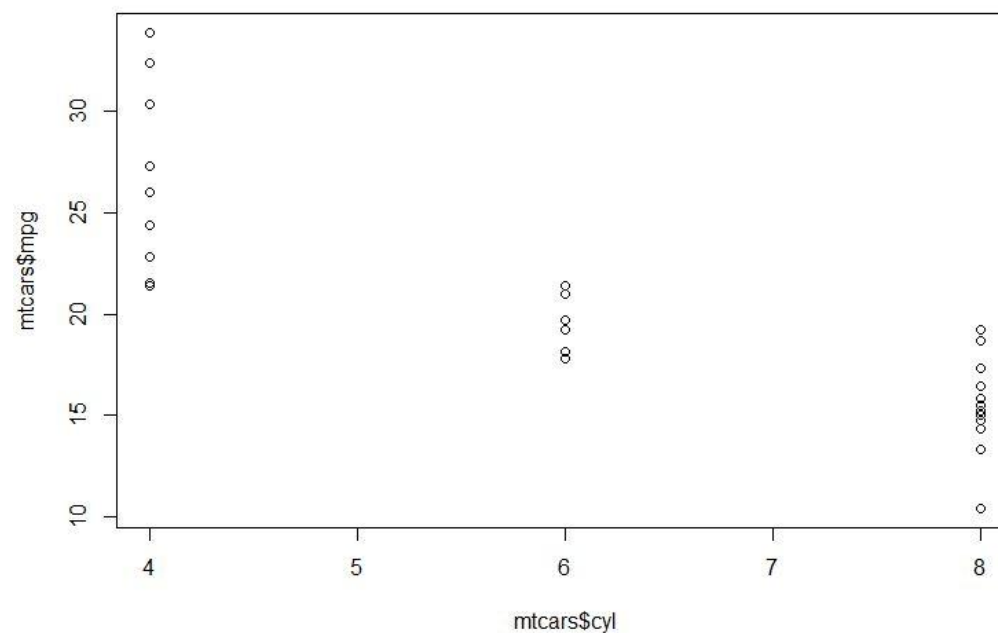




Zadatak 3

- Iz data frame-a mtcars izvući podatke o broju cilindara automobila i pređenog puta po galonu. Da li postoji veza između navedenih veličina?

```
plot(mtcars$cyl, mtcars$mpg)
```





Zadatak 4

- Na slici su date ocene kojima su učenici ocenili profesore na skali od 1 do 5. Na spojenom (stacked) barplot-u predstaviti odgovore na prvo i drugo pitanje, a na grafiku sa odvojenim stupcima predstaviti odgovore na sva tri pitanja.

```
q1 = c(3, 3, 3, 4, 3, 4, 3, 4, 3, 4)
q2 = c(5, 2, 5, 5, 2, 2, 5, 5, 4, 2)
q3 = c(1, 3, 1, 1, 1, 3, 1, 1, 1, 1)

barplot(table(q1, q2), legend.text = c("q1", "q2"))
barplot(matrix(data = c(q1, q2, q3), nrow = 3), beside = T,
legend.text = c("q1", "q2", "q3"))
```

Student	Ques. 1	Ques. 2	Ques. 3
1	3	5	1
2	3	2	3
3	3	5	1
4	4	5	1
5	3	2	1
6	4	2	3
7	3	5	1
8	4	5	1
9	3	4	1
10	4	2	1