

3. вежбе - R

Напредни софтверски алати



Data frame

- Data frame (оквир или база података) представља посебну структуру података у програмском језику R која користи доста карактеристика листа и матрица
- Садржи именоване колоне које се састоје од података истог типа и именоване редове, од којих сваки има јединствено име
- Формирање

```
visina = c(180, 175, 210, 169)
tezina = c(65, 61, 110, 65)
pol = c('Z', 'Z', 'M', 'Z')
baza = data.frame(visina,tezina,pol) #formira bazu
baza
View(baza) # otvara bolji pregled
baza = data.frame(vis=visina,tez=tezina,p=pol) #promena imena kolona
row.names(baza)<-c("Maja","Ana","Bojan","Julija") #promena imena vrsta
```



Додавање података

```
novipodatak <- data.frame(visina=189,tezina=99,pol='M')
baza <- rbind(baza,novipodatak)
rownames(baza)[rownames(baza)==1]<-'Ilija'

godine <- c(42,40,17,24,19)
baza <- cbind(baza,godine)

baza$godine.mes <- baza$godine*12
```



Пристап појединим елементима

- Особине које су представљене вредностима у колонама називаћемо варијаблама
- Приступ колони и врсти

```
baza[, 'visina']
```

```
baza[, 2]
```

```
baza$tezina
```

```
baza[1, ]
```

```
baza[ 'Maja' , ]
```

```
baza[, c(1, 3)]
```

```
baza[1:3, ]
```

```
baza[ 'Maja' , 'visina' ]
```

- Варијаблама се може приступити појединачно након `attach` команде

```
rm(visina) # brise vektor ukoliko postoji u okruzenju
```

```
visina
```

```
attach(baza)
```

```
visina
```

```
detach(baza)
```



Уграђени data frame

- R у својим пакетима садржи одређени број уграђених база, који се углавном користе као демо подаци за упознавање са функцијама
- неке од база: mtcars, iris, ToothGrowth, PlantGrowth, USArrests...
- пакет datasets садржи велики број таквих база, за више информација:

```
library(help = "datasets")
```

или

```
data()
```

```
mtcars #stampa ceo sadrzaj ugradjenog data frame-a
```

```
?mtcars
```

```
head(mtcars)
```

```
tail(mtcars) #prvih 6/ poslednjih 6 elemenata data frame-a
```



Задатак 1

- Из уграђене базе mtcars издвојити у посебне базе аутомобиле са мануелним и аутоматским мењачем (варијабла “am”).
- Пронаћи и штампати податке о аутомобилу који троши најмање горива за аутомобиле са мануелним мењачем и податке о аутомобилу који троши највише горива (за аутомобиле са аутоматским мењачем). Податке одштампати на основу пређеног пута по галону (варијабла “mpg”)
- Одштампати податке о просечном пређеном путу по галону горива свих аутомобила, а затим аутомобила са мануелним мењачем и аутомобила са аутоматским мењачем.



Задатак 1

```
#najmanja potrosnja manuelnih
manuelni = mtcars[mtcars$am == 1,]
najmanji_manuelni = which(manuelni$mpg == max(manuelni$mpg))
print(manuelni[najmanji_manuelni,])

#najveca potrosnja automatika
automatici = mtcars[mtcars$am == 0,]
najveci_automatik = which(automatici$mpg == min(automatici$mpg))
print(automatici[najveci_automatik,])

#prosecne potrosnje
paste("Prosek predjenog puta svih je:", mean(mtcars$mpg))
paste("Prosek predjenog puta manuelnih je:", mean(manuelni$mpg))
paste("Prosek predjenog puta automatika je:", mean(automatici$mpg))
```



Учитавање из фајла

- Команда `read.table` учитава и складишти у базу података

```
baza1=read.table(file="zdravlje.csv",header=TRUE,sep = ",")
```

- Други начин

```
baza2=read.csv(file="zdravlje.csv",header=TRUE,sep = ",") #или
```

```
baza2=read.csv(file.choose(),header=T)
```

- Excel baza

```
library("readxl")
```

```
my_data <- read_excel("my_file.xls") # xls files
```

```
my_data <- read_excel("my_file.xlsx") #xlsx files
```

- Извоз података

```
write.table(baza,file="Nova baza.csv",sep=",")
```

```
write.table(baza,file="Nova baza t.txt",sep=" ")
```



Извоз података

```
write.table(baza, file="Nova baza.csv", sep=",")  
write.table(baza, file="Nova baza t.txt", sep=" ")
```

- Извоз excel фајла

```
install.packages("writexl")  
library("writexl")  
write_xlsx(baza, "nova baza excel.xlsx")
```



Задатак 2

- Из фајла „countries.csv“ учитати податке о БДП-у земаља и процентуалном учешћу индустрије и услуга у БДП-у сваке земље
- На основу процентуалног учешћа у БДП-у израчунати вредност индустрије, услуга и осталих делатности за сваку од земаља.
- Одштампати податке о:
 - земљи која има највећи удео индустрије (у процентима)
 - земљи која има најмањи удео услуга (у процентима)



Задатак 2

```
data = read.csv(file="countries.csv", header=T)
vrednost_industrije = data$gdp * (data$percent_industry/100)
vrednost_usluga = data$gdp * (data$percent_services/100)
ostalo = data$gdp * ((100 - data$percent_industry -
data$percent_services)/100)

#zemlja sa maksimalnim udelom industrije u procentima
industrijalizovane_zemlje = which(data$percent_industry ==
max(data$percent_industry))
print(data[industrijalizovane_zemlje,])

#zemlje sa minimalnim nivoom usluga u procentima
zemlje_bez_usluga = which(data$percent_services ==
min(data$percent_services))
print(data[zemlje_bez_usluga,])
```

