Projekt SAP

Tema 2 - Uloga izvoza i uvoza u gospodarstvu

Pavo Matanović, Karla Baričević, Slavko Boldin

Učitavanje podataka i deskriptivna analiza

Na početku učitavamo podatke i analiziramo kako izgledaju podaci.

```
export.data = read.csv("Export_data.csv")
# head(export.data)

import.data = read.csv("Import_data.csv")
# head(import.data)

gdp.data = read.csv("GDP_data.csv")
# head(gdp.data)

gdp.pc.data = read.csv("GDPpercapita_data.csv")
# head(gdp.pc.data)
```

Sljedeći blok koda generira dataframe sa brojem upisanih podataka te brojem procjena među upisanim podacima.

```
export.loc.cnt = export.data %>% group_by(LOCATION) %>%
    summarise(exp_n = n(), exp_est = sum(Flag.Codes == 'E')) %>%
    arrange(desc(exp_n), exp_est)
import.loc.cnt = import.data %>% group_by(LOCATION) %>%
    summarise(imp_n = n(), imp_est = sum(Flag.Codes == 'E')) %>%
    arrange(desc(imp_n), imp_est)
gdp.loc.cnt = gdp.data %>% group_by(LOCATION) %>%
    summarise(gdp_n = n(), gdp_est = sum(Flag.Codes == 'E')) %>%
    arrange(desc(gdp_n), gdp_est)
gdp.pc.loc.cnt = gdp.pc.data %>% group_by(LOCATION) %>%
    summarise(gdp_pc_n = n(), gdp_pc_est = sum(Flag.Codes == 'E')) %>%
    arrange(desc(gdp_pc_n), gdp_pc_est)
loc.cnt = merge(merge(export.loc.cnt, import.loc.cnt), merge(gdp.loc.cnt, gdp.pc.loc.cnt))
knitr::kable(
  head(arrange(loc.cnt,
               desc(loc.cnt[,2]), desc(loc.cnt[,4]), desc(loc.cnt[,6]), desc(loc.cnt[,8]),
               loc.cnt[,3], loc.cnt[,5], loc.cnt[,7], loc.cnt[,9]), 20),
  caption = "Broj podataka za pojedinu državu"
)
```

Table 1: Broj podataka za pojedinu državu

LOCATION	exp_n	exp_est	imp_n	imp_est	gdp_n	gdp_est	gdp_pc_n	gdp_pc_est
CAN	41	0	41	0	41	0	41	0

LOCATION	exp_n	exp_est	imp_n	imp_est	gdp_n	gdp_est	gdp_pc_n	gdp_pc_est
DNK	41	0	41	0	41	0	41	0
FRA	41	0	41	0	41	0	41	0
CHE	41	1	41	1	41	1	41	1
FIN	41	1	41	1	41	1	41	1
DEU	41	12	41	12	41	12	41	12
SWE	41	14	41	14	41	14	41	14
GBR	41	16	41	16	41	0	41	0
AUT	41	16	41	16	41	16	41	16
BEL	41	16	41	16	41	16	41	16
ESP	41	16	41	16	41	16	41	16
GRC	41	16	41	16	41	16	41	16
IRL	41	16	41	16	41	16	41	16
ITA	41	16	41	16	41	16	41	16
NLD	41	16	41	16	41	16	41	16
ISL	41	16	41	16	41	16	41	17
PRT	41	17	41	17	41	17	41	17
AUS	40	0	40	0	40	0	40	0
NOR	40	0	40	0	40	0	40	0
USA	40	0	40	0	40	0	40	0

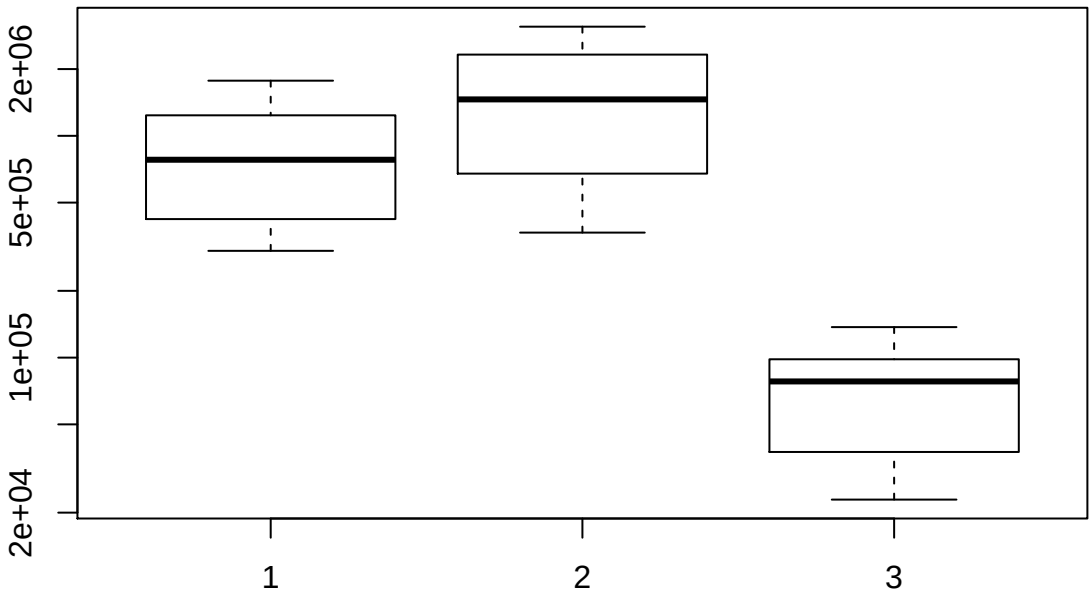
Odabrane drzave

Odabrali smo USA, Njemačku(DEU) i Grčku(GRC) za analizu.

```
usa = data.frame(export.mln_usd = export.data$Value[export.data$LOCATION == "USA"],
                 import.mln_usd = import.data$Value[import.data$LOCATION == "USA"],
                 gdp.mln_usd = gdp.data$Value[gdp.data$LOCATION == "USA"],
                 gdp.pc.usd_cap = gdp.pc.data$Value[gdp.pc.data$LOCATION == "USA"])
usa$net.trade = usa$export.mln_usd - usa$import.mln_usd
deu = data.frame(export.mln_usd = export.data$Value[export.data$LOCATION == "DEU"],
                 import.mln_usd = import.data$Value[import.data$LOCATION == "DEU"],
                 gdp.mln_usd = gdp.data$Value[gdp.data$LOCATION == "DEU"],
                 gdp.pc.usd_cap = gdp.pc.data$Value[gdp.pc.data$LOCATION == "DEU"])
deu$net.trade = deu$export.mln_usd - deu$import.mln_usd
grc = data.frame(export.mln_usd = export.data$Value[export.data$LOCATION == "GRC"],
                 import.mln_usd = import.data$Value[import.data$LOCATION == "GRC"],
                 gdp.mln_usd = gdp.data$Value[gdp.data$LOCATION == "GRC"],
                 gdp.pc.usd_cap = gdp.pc.data$Value[gdp.pc.data$LOCATION == "GRC"])
grc$net.trade = grc$export.mln_usd - grc$import.mln_usd
```

Deskriptivna statistika

```
boxplot(deu$import.mln_usd,
        usa$import.mln_usd,
        grc$import.mln_usd,
        # names = expression(deparse(substitute())),
        log = "y")
```



Vidimo da im se čisti izvoz u mil. USD razlikuje jako čak i na logaritamskoj skali.

To možemo potvrditi i statističkim t testom.

```
t.test(log(usa$import.mln_usd), log(deu$import.mln_usd),
       alternative = "g", var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  log(usa$import.mln_usd) and log(deu$import.mln_usd)
## t = 3.607, df = 76.264, p-value = 0.0002755
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.2795736       Inf
## sample estimates:
## mean of x mean of y
##  14.01251  13.49321
```

```
t.test(log(deu$import.mln_usd), log(grc$import.mln_usd),
       alternative = "g", var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  log(deu$import.mln_usd) and log(grc$import.mln_usd)
## t = 19.365, df = 79.86, p-value < 2.2e-16
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  2.265246      Inf
## sample estimates:
## mean of x mean of y
##  13.49321  11.01499
```