# 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.

```
time = 1979:2019
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
```

```
usa = usa %>% mutate(import.mln_usd,
                     import.growth = import.mln_usd - lag(import.mln_usd),
                     import.growth.percentage = import.growth / lag(import.mln_usd) * 100)
deu = deu %>% mutate(import.mln_usd,
                     import.growth = import.mln_usd - lag(import.mln_usd),
                     import.growth.percentage = import.growth / lag(import.mln_usd) * 100)
grc = grc %>% mutate(import.mln_usd,
                     import.growth = import.mln_usd - lag(import.mln_usd),
                     import.growth.percentage = import.growth / lag(import.mln_usd) * 100)
```

```
interpolate.usa = data.frame(t(rep(NA, length(names(usa)))))
names(interpolate.usa) = names(usa)
interpolate.usa$import.growth.percentage = mean(usa$import.growth.percentage
                                                [time >= 2013 & time < 2019])
interpolate.usa$import.mln_usd = usa$import.mln_usd[nrow(usa)] *
                                 (1 + interpolate.usa$import.growth.percentage / 100)
interpolate.usa$import.growth = interpolate.usa$import.mln_usd - usa$import.mln_usd[nrow(usa)]
usa = rbind(usa, interpolate.usa)
```

```
summary(usa)
```

```
##  export.mln_usd    import.mln_usd     gdp.mln_usd       gdp.pc.usd_cap
##  Min.   : 347872   Min.   : 366207   Min.   : 2627334   Min.   :11672
##  1st Qu.: 591516   1st Qu.: 690188   1st Qu.: 5540294   1st Qu.:22445
##  Median :1185694   Median :1538060   Median : 9346740   Median :33648
##  Mean   :1206334   Mean   :1545274   Mean   :10103023   Mean   :34815
##  3rd Qu.:1762818   3rd Qu.:2326964   3rd Qu.:14517106   3rd Qu.:48004
##  Max.   :2416053   Max.   :3224342   Max.   :20580223   Max.   :62853
##  NA's   :1                           NA's   :1          NA's   :1
##    net.trade       import.growth     import.growth.percentage
##  Min.   :-722881   Min.   :-304448   Min.   :-13.084
##  1st Qu.:-497748   1st Qu.:  32563   1st Qu.:  2.683
##  Median :-274298   Median :  65733   Median :  5.143
##  Mean   :-296964   Mean   :  70801   Mean   :  5.592
##  3rd Qu.: -77556   3rd Qu.: 130815   3rd Qu.:  8.659
##  Max.   :  19122   Max.   : 265511   Max.   : 24.343
##  NA's   :1         NA's   :1         NA's   :1
```

```
summary(deu)
```

```
##  export.mln_usd    import.mln_usd     gdp.mln_usd      gdp.pc.usd_cap
##  Min.   : 277599   Min.   : 303002   Min.   : 736116   Min.   : 9425
##  1st Qu.: 439432   1st Qu.: 421397   1st Qu.:1413237   1st Qu.:17963
##  Median : 789414   Median : 780127   Median :2158516   Median :26510
##  Mean   : 961808   Mean   : 851788   Mean   :2344231   Mean   :28963
##  3rd Qu.:1466601   3rd Qu.:1237522   3rd Qu.:3103958   3rd Qu.:38432
##  Max.   :2019336   Max.   :1773266   Max.   :4632060   Max.   :55737
##
##    net.trade       import.growth     import.growth.percentage
##  Min.   :-27681   Min.   :-119972   Min.   :-9.695
##  1st Qu.: 12702   1st Qu.:  12657   1st Qu.: 2.694
##  Median : 38067   Median :  36879   Median : 4.881
##  Mean   :110020   Mean   :  36700   Mean   : 4.605
##  3rd Qu.:235195   3rd Qu.:  62129   3rd Qu.: 7.926
##  Max.   :294551   Max.   : 143844   Max.   :12.871
##                   NA's   :1         NA's   :1
```

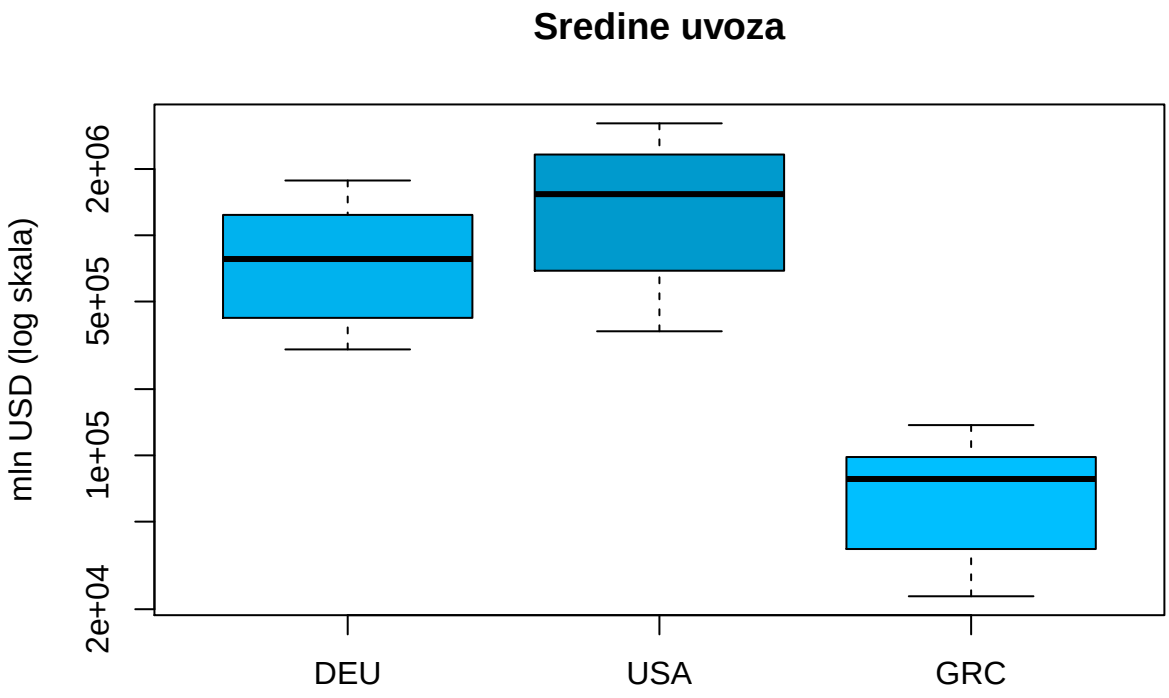
```
summary(grc)
```

```
##  export.mln_usd   import.mln_usd    gdp.mln_usd     gdp.pc.usd_cap
##  Min.   : 20482   Min.   : 22900   Min.   : 76529   Min.   : 7933
##  1st Qu.: 28580   1st Qu.: 37540   1st Qu.:130598   1st Qu.:12811
##  Median : 56743   Median : 78096   Median :198712   Median :18465
##  Mean   : 56376   Mean   : 70093   Mean   :210316   Mean   :19609
##  3rd Qu.: 81461   3rd Qu.: 98300   3rd Qu.:291109   3rd Qu.:26839
```

```
##  Max.   :109674   Max.   :137267   Max.   :341818   Max.   :31172
##
##    net.trade      import.growth      import.growth.percentage
##  Min.   :-44429   Min.   :-27939.2   Min.   :-20.3539
##  1st Qu.:-24525   1st Qu.:   345.2   1st Qu.:  0.5721
##  Median :-11914   Median :  1963.6   Median :  4.2862
##  Mean   :-13717   Mean   :  2054.5   Mean   :  4.1784
##  3rd Qu.: -3016   3rd Qu.:  4379.9   3rd Qu.:  8.5646
##  Max.   :  4592   Max.   : 18180.4   Max.   : 20.1801
##                   NA's   :1          NA's   :1
```

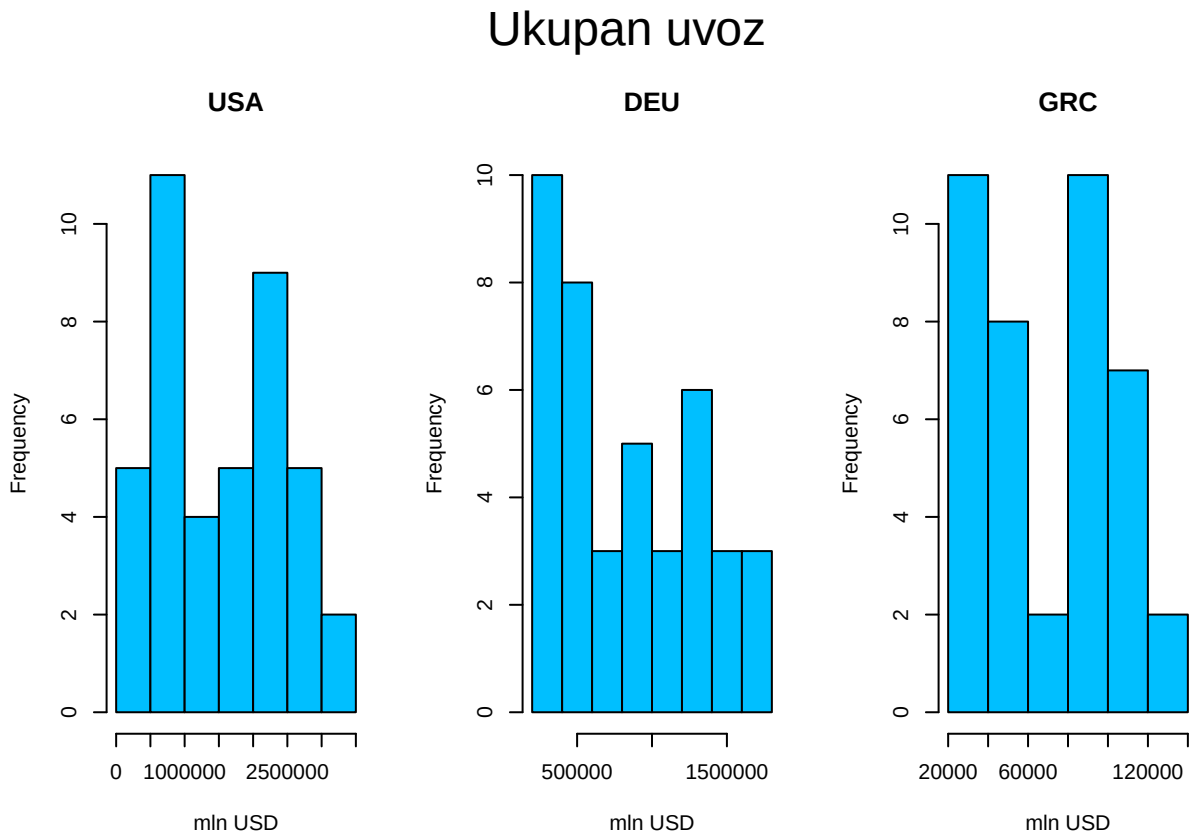
### Deskriptivna statistika

```
boxplot(deu$import.mln_usd,
        usa$import.mln_usd,
        grc$import.mln_usd,
        names = c("DEU", "USA", "GRC"), main = "Sredine uvoza",
        col = c("deepskyblue2", "deepskyblue3", "deepskyblue"),
        ylab = "mln USD (log skala)",
        log = "y")
```




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

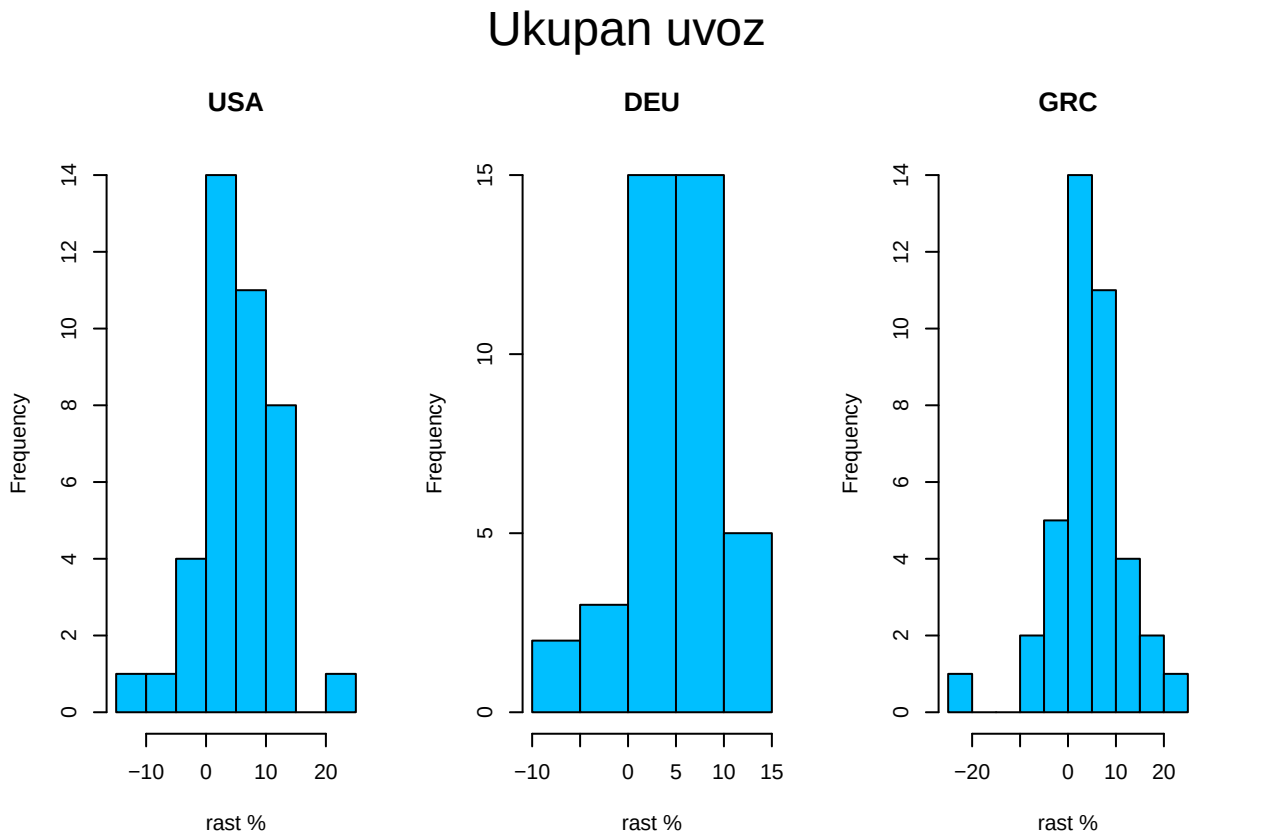
```
par(mfrow = c(1, 3), oma = c(0, 0, 2, 0))
hist(usa$import.mln_usd, main="USA", xlab="mln USD", col="deepskyblue")
hist(deu$import.mln_usd, main="DEU", xlab="mln USD", col="deepskyblue")
hist(grc$import.mln_usd, main="GRC", xlab="mln USD", col="deepskyblue")
mtext("Ukupan uvoz", outer = T, cex = 1.5)
```




Za distribucije ukupnog izvoza ne možemo pretpostaviti normalnost, pa nema smisla raditi parametarske testove.

Usporedimo sada postotnu promjenu izvoza.

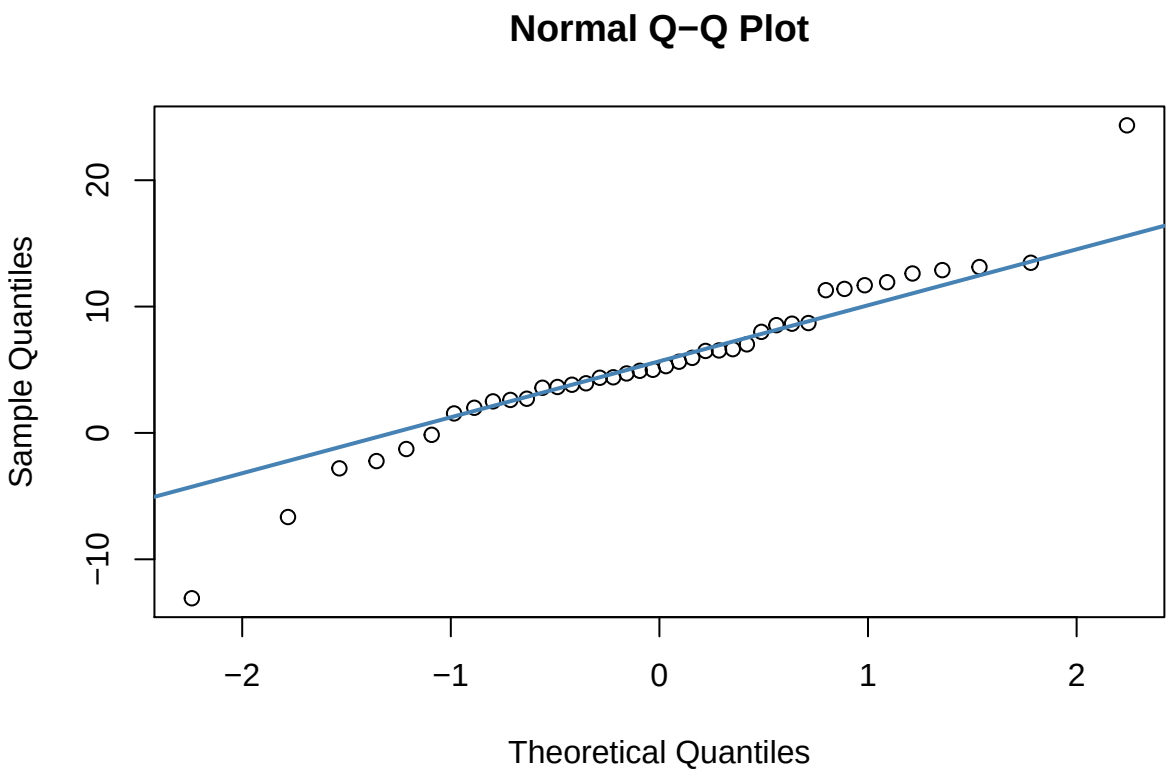
```
par(mfrow = c(1, 3), oma = c(0, 0, 2, 0))
hist(usa$import.growth.percentage, main="USA", xlab="rast %", col="deepskyblue")
hist(deu$import.growth.percentage, main="DEU", xlab="rast %", col="deepskyblue")
hist(grc$import.growth.percentage, main="GRC", xlab="rast %", col="deepskyblue")
mtext("Ukupan uvoz", outer = T, cex = 1.5)
```



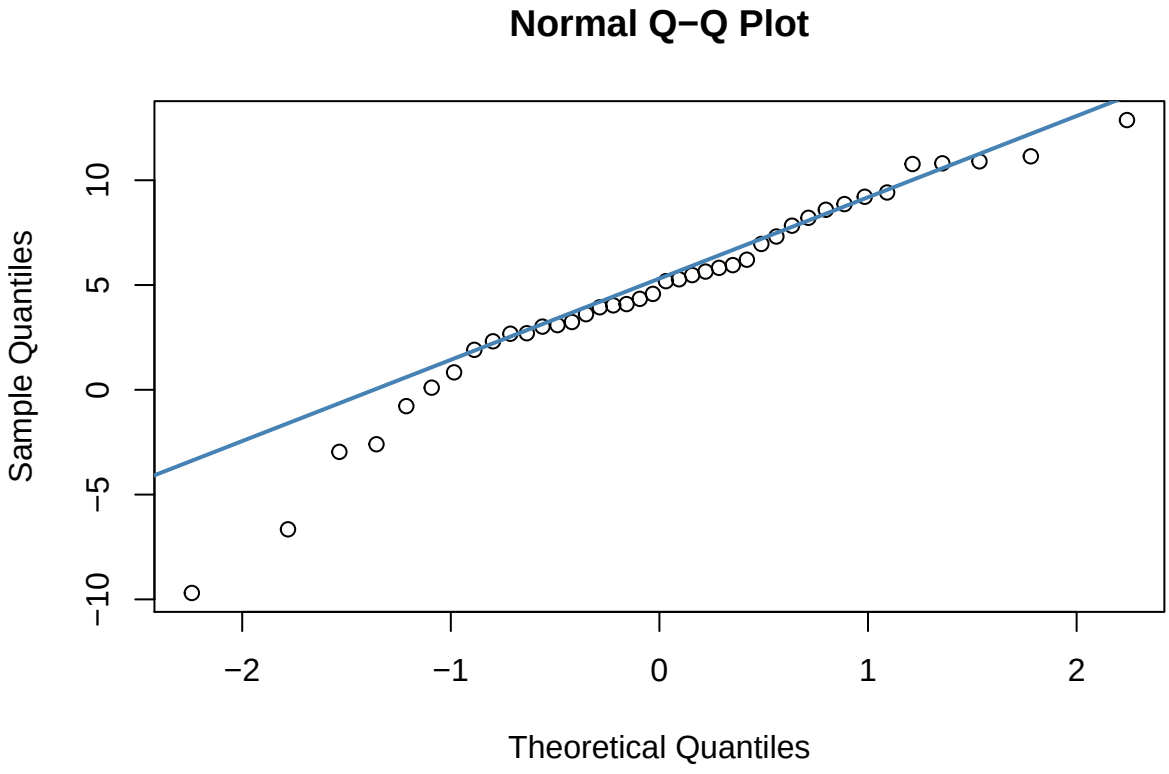

Distribucije nisu previše zakrivljene i imamo dovoljno podataka da možemo pretpostaviti normalnost distribucije.

Taj zaključak potvrđuju i qq plotovi.

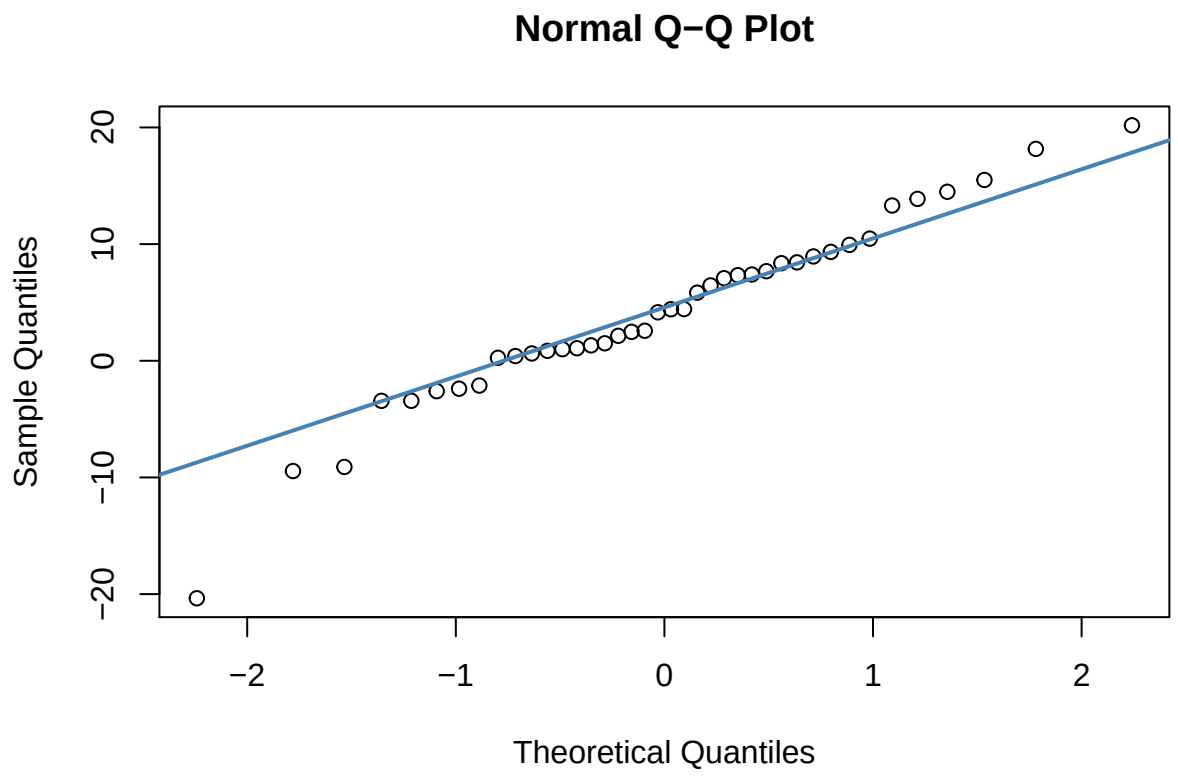
```
qqnorm(usa$import.growth.percentage)
qqline(usa$import.growth.percentage, col = "steelblue", lwd = 2)
```



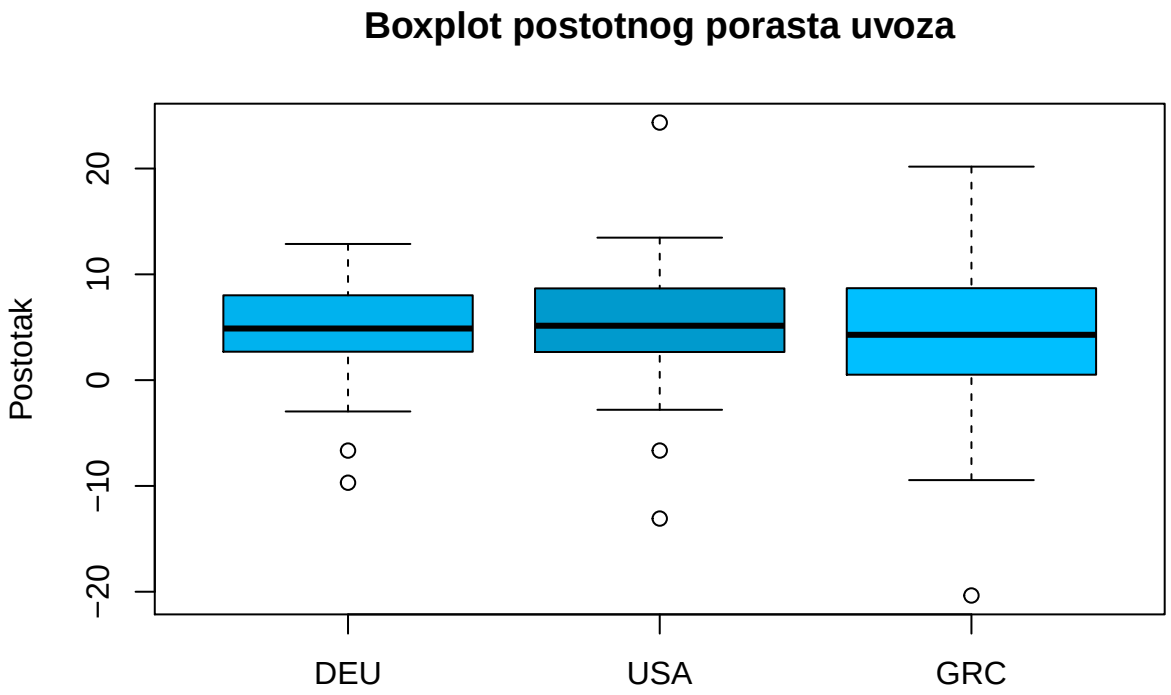

```r
qqnorm(deu$import.growth.percentage)
qqline(deu$import.growth.percentage, col = "steelblue", lwd = 2)
```




```r
qqnorm(grc$import.growth.percentage)
qqline(grc$import.growth.percentage, col = "steelblue", lwd = 2)
```




```
boxplot(deu$import.growth.percentage,
        usa$import.growth.percentage,
        grc$import.growth.percentage,
        names = c("DEU", "USA", "GRC"), main = "Boxplot postotnog porasta uvoza",
        col = c("deepskyblue2", "deepskyblue3", "deepskyblue"),
        ylab = "Postotak")
```




Ovaj plot pokazuje da bi varijable USA i DEU mogle imati istu sredinu. Provjerimo to sa t testom.

```r
t.test(usa$import.growth.percentage[-c(1)], deu$import.growth.percentage[-c(1)],
       alternative = "g")
```

```
##
##  Welch Two Sample t-test
##
## data:  usa$import.growth.percentage[-c(1)] and deu$import.growth.percentage[-c(1)]
## t = 0.79157, df = 72.659, p-value = 0.2156
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -1.089884       Inf
## sample estimates:
## mean of x mean of y
##  5.591600  4.605104
```

```r
deu[-c(1), ] %>% with(hist(import.growth.percentage,
                           breaks = seq(min(import.growth.percentage) - 0.5,
                                        max(import.growth.percentage) + 0.5, length.out = 9)))
```