

Mjerenje uspješnosti investitora na financijskim tržištima

Zadatak je provjeriti ispravnost hipoteze ekiasnih tržišta koja tvrdi da je nemoguće “pobijediti tržište”. Varijabla promatranja je prinos u odnosu na CROBEX.

Većina analize će se provoditi nad spojenim podacima od 2009. godine do 2017. godine. Njihovim spajanjem te promatranjem varijable prinos iznad CROBEXA pretpostavlja se nezavisnost između godina. Ta tvrdnja nije istinita jer će se kasnije pokazati da ima puno ekipa koje su sudjelovale kroz nekoliko godina natjecanja te samim time je njihov noviji rezultat zavisao o prethodnim pokušajima. Također, pretpostavka da su uvjeti bili jednaki svake godine je isto vrlo vjerojatno prekršena. Usprkos svemu tome, rezultat analize ne bi bio puno drugačiji da gornje tvrdnje nisu prekršene.

Donji kod učitava podatke iz svih datoteka te dodaje stupac koji govori o kojoj se godini radi. Miču se prazni reci te se svi stupci koji sadrže postotak pretvaraju u double. Znak posto je maknut te broj koji se sada nalazi u tom stupcu označava postotak. Tipa ako piše 10 to onda znači 10 posto. Kolumna koja sadrži godine je potom pretvorena u factor, tj kategorijski podatak. Takav dataset je spremljen u file dionice.Rdata te ga se lako može učitati iz memorije kasnije.

```
# dionice.2009 = read.csv("podaci/mojedionice2009.csv", header = T, sep = ",")
# dionice.2010 = read.csv("podaci/mojedionice2010.csv", header = T, sep = ",")
# dionice.2011 = read.csv("podaci/mojedionice2011.csv", header = T, sep = ",")
# dionice.2012 = read.csv("podaci/mojedionice2012.csv", header = T, sep = ",")
# dionice.2013 = read.csv("podaci/mojedionice2013.csv", header = T, sep = ",")
# dionice.2014 = read.csv("podaci/mojedionice2014.csv", header = T, sep = ",")
# dionice.2015 = read.csv("podaci/mojedionice2015.csv", header = T, sep = ",")
# dionice.2016 = read.csv("podaci/mojedionice2016.csv", header = T, sep = ",")
# dionice.2017 = read.csv("podaci/mojedionice2017.csv", header = T, sep = ",")
#
# dionice.2009$Godina = 2009
# dionice.2010$Godina = 2010
# dionice.2011$Godina = 2011
# dionice.2012$Godina = 2012
# dionice.2013$Godina = 2013
# dionice.2014$Godina = 2014
# dionice.2015$Godina = 2015
# dionice.2016$Godina = 2016
# dionice.2017$Godina = 2017

# problem koji ovdje nastaje je taj da dionice iz 2009 imaju drugačiji naziv za otp indeks

# help("colnames")
# colnames(dionice.2010)[7]
# colnames(dionice.2009)[7] = colnames(dionice.2010)[7]
# dionice.2009$Prinos.iznad..OTP.indeksnog.fonda

# dionice = rbind(
#   dionice.2009,
#   dionice.2010,
#   dionice.2011,
#   dionice.2012,
#   dionice.2013,
#   dionice.2014,
```

```

#   dionice.2015,
#   dionice.2016,
#   dionice.2017
# )

#
# dionice = na.omit(dionice)
# dionice
#
# my_function("5.34%")
# my_function("5,34%")
# my_function = function(string) {
#   s = gsub("%", "", string)
#   s = gsub(",", ".", s)
#   return(as.numeric(s))
# }
#
# dionice$Prinos.iznad..OTP.indeksnog.fonda = unlist(lapply(dionice$Prinos.iznad..OTP.indeksnog.fonda, my_function))
# dionice$Prinos = unlist(lapply(dionice$Prinos, my_function))
# dionice$Prinos.bez.div. = unlist(lapply(dionice$Prinos.bez.div., my_function))
#
# save(dionice, file = "dionice.Rdata")
# load("dionice.Rdata")
#
# head(dionice)
#
# class(dionice$Prinos.iznad..OTP.indeksnog.fonda)
# class(dionice$Godina)
#
# levels(dionice$Godina)
# head(dionice)
# dim(dionice)
# names(dionice)
#
# dionice$Godina = as.factor(dionice$Godina)
# levels(dionice$Godina)

# save(dionice, file = "dionice.Rdata")
load("dionice.Rdata")

```

Deskriptivna statistika.

Uspoređujući mediane i srednje vrijednosti tokom godina uočava se veliko odstupanje između samih godina te se polako naslućuje da prinos iznad CROBEXA gleano sveukupno naginje prema negativnoj vrijednosti.

```
tapply(dionice$Prinos.iznad..OTP.indeksnog.fonda, dionice$Godina, summary)
```

```

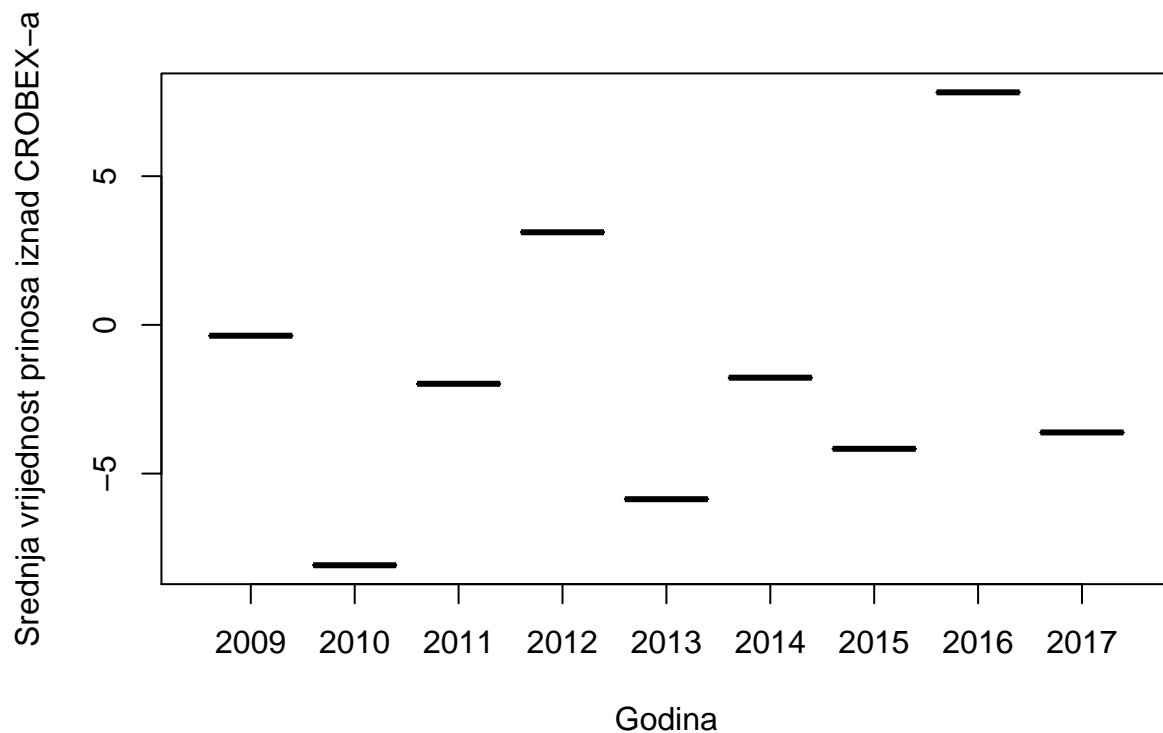
## $`2009`
##      Min.   1st Qu.   Median     Mean  3rd Qu.    Max.
## -43.2600  -7.5650   -0.4800  -0.3647   8.4950   26.8100
##
## $`2010`
##      Min.   1st Qu.   Median     Mean  3rd Qu.    Max.

```

```
## -58.980 -18.660 -8.910 -8.082 0.950 62.950
##
## $`2011`
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## -35.26   -9.77   -3.46   -1.98    5.55   42.71
##
## $`2012`
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## -63.700  -9.020   5.630   3.113  17.230  80.330
##
## $`2013`
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## -59.690 -16.005  -3.920  -5.862   4.475 131.780
##
## $`2014`
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## -50.890 -16.670  -3.870  -1.776   9.900  88.830
##
## $`2015`
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## -49.590 -13.025  -2.870  -4.167   6.425  30.230
##
## $`2016`
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## -27.890  -1.575   4.940   7.818  16.328  64.080
##
## $`2017`
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## -84.540 -15.640  -0.590  -3.617  12.325  61.260
```

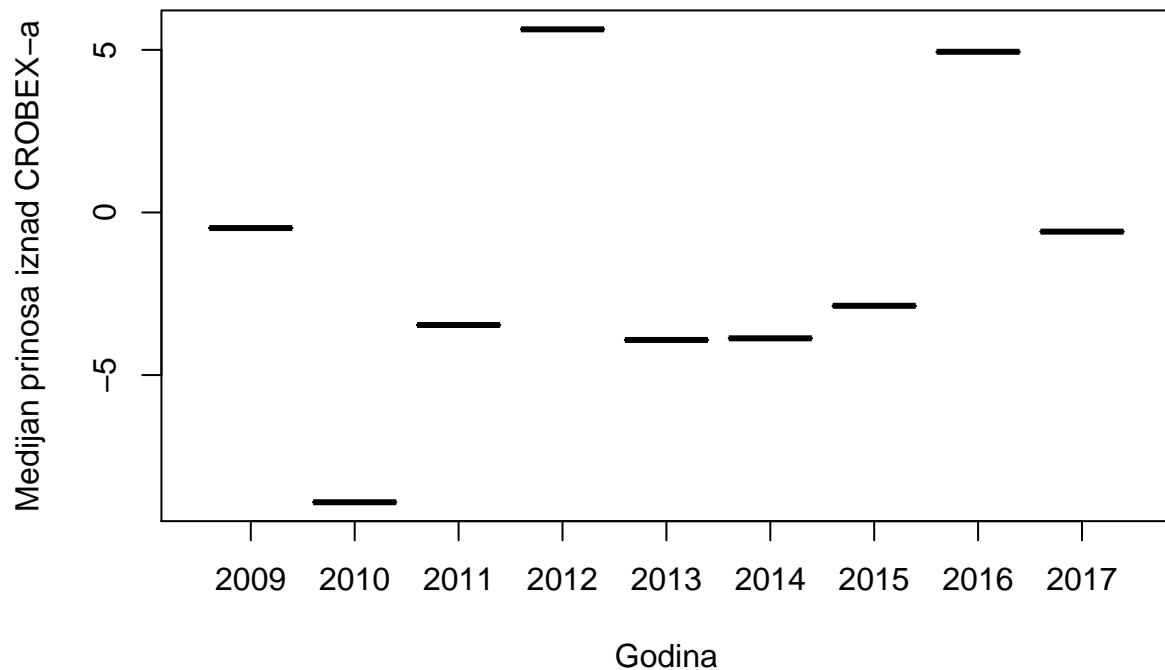
Dole je prikazan graf srednje vrijednosti prinosa iznad CROBEXA kroz godine. Većina srednjih vrijednosti se nalazi u blagome minusu uz izuzetak 2016. godine kada je srednja vrijednost na skoro 8 posto u plusu, 2010. godina je pak u minusu od 8%.

```
mean.by.year = aggregate(dionice[, c(7)], list(dionice$Godina), mean)
plot(mean.by.year, ylab="Srednja vrijednost prinosa iznad CROBEX-a", xlab="Godina")
```



Dole je prikazan graf medijana kroz godine. Većina medijana se također nalazi u blagome minusu no zanimljivo je da se najveći medijan ovog puta nalazi u 2012. godini (najveća srednja vrijednost je bila 2016. godine). Medijan za 2012. godinu iznosi 5.63% dok srednja vrijednost za 2012. godinu iznosi 3.1%, 2010. godina je i dalje u velikom minusu i to skoro za 9% ovaj puta.

```
median.by.year = aggregate(dionice[, c(7)], list(dionice$Godina), median)
plot(median.by.year, ylab="Medijan prinosa iznad CROBEX-a", xlab="Godina")
```

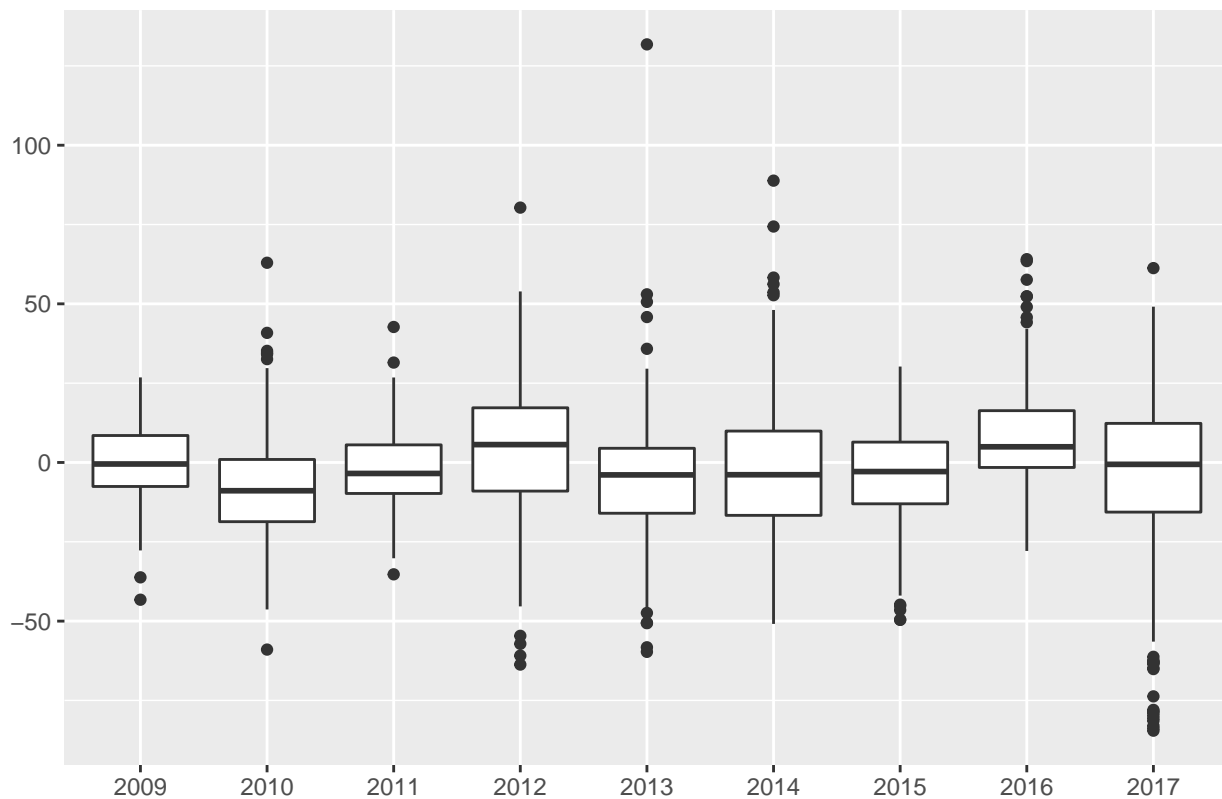


Dvije generalne tvrdnje koje se mogu konstatirati u ovom trenutku su te da je 2010. godina bila najbolja godina za tržište dok su 2012. i 2016. godina bile najgore godine za tržište.

Dole je pak prikaz Boxplotova za svaku od godina. Uočava se stršuća pozitivnu vrijednost u 2013 godini. Neko je tad naime, bio 131% u plusu u prihodu iznad CROBEX indexa.

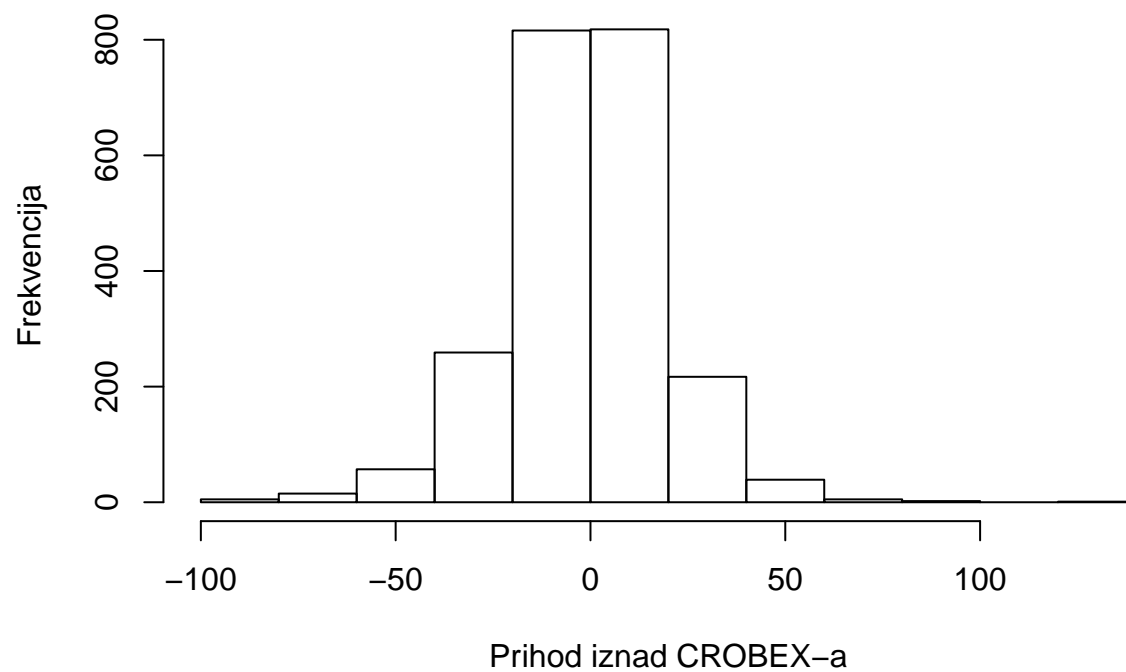
```
require(ggplot2, quietly = T)
ggplot(dionice, aes(x=Godina, y=Prinos.iznad..OTP.indeksnog.fonda)) + geom_boxplot(aes(fill=Prinos.iznad..OTP.indeksnog.fonda))
theme(axis.title.y = element_blank(), axis.title.x = element_blank()) +
ggtitle("Boxplot prikaz prinosa iznad CROBEX-a po godinama")
```

Boxplot prikaz prinosa iznad CROBEX-a po godinama



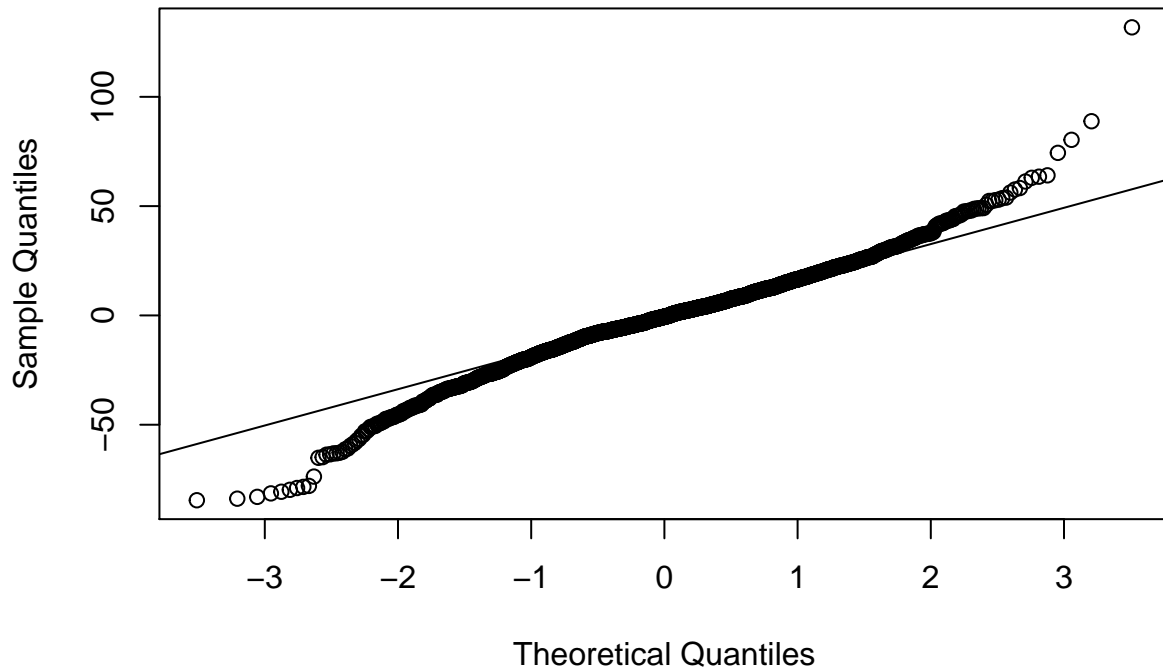
Proučavanje ima li varijabla prinos iznad CROBEX-a normalnu distribucije. Na temelju histograma i kvantil-kvantil prikaza možemo pretpostaviti da se zaista radi o normalnoj distribuciji makar na kvantil-kvantil prikazu vidimo odstupanja u lijevom i desnom repu zbog kojih možemo posumnjati.

```
hist(dionice$Prinos.iznad..OTP.indeksnog.fonda, main=NULL, xlab = "Prihod iznad CROBEX-a", ylab = "Frekvencija")
```



```
qqnorm(dionice$Prinos.iznad..OTP.indeksnog.fonda, main="Kvantil-Kvantil prikaz prinosa iznad CROBEX-a")  
qqline(dionice$Prinos.iznad..OTP.indeksnog.fonda)
```

Kvantil–Kvantil prikaz prinosa iznad CROBEX–a



Medijan promatran kroz sve godine prihoda iznad CROBEX-a je -0.605 dok je srednja vrijednost -1.171. Oboje su negativni što daje naslutiti da tržište pobjeđuje. Najlošiji postotak ostvaren tokom svih godina je onaj od -84% dok je najbolji u plusu za 131%.

```
summary(dionice$Prinos.iznad..OTP.indeksnog.fonda)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -84.540 -11.750  -0.605  -1.171  10.637  131.780
```

Srednja vrijednost podrezanih najgorih 20% i najboljih 20% prihoda iznad CROBEXA također iznosi negativnu vrijednost te joj je iznos jako sličan medijanu.

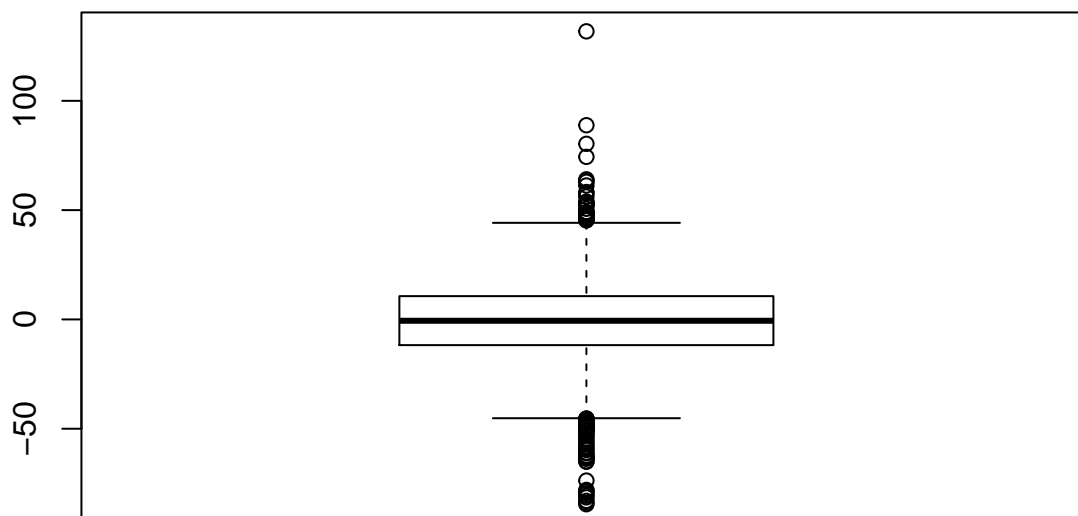
```
otp = dionice$Prinos.iznad..OTP.indeksnog.fonda
mean(otp, trim=0.2)
```

```
## [1] -0.5620864
```

Dole se nalazi boxplot prikaz svih zajedno godina ostvarenog prihoda iznad CROBEX-a.

```
boxplot(otp, main="Pravokutni dijagram prihoda iznad CROBEX-a svih godina zajedno.")
```


Pravokutni dijagram prihoda iznad CROBEX-a svih godina zajedno



Testiranje srednje vrijednosti prihoda iznad CROBEX-a

Provodi se t test s null hipotezom da je srednja vrijednost jednaka nuli, s alternativom da nije jednaka nuli. Koristeći pogrešku prve vrste (α) od 5% te na temelju dobivene p vrijednosti od 0.58% možemo odbaciti null hipotezu. Gledajući 95%-tni interval pouzdanosti koji ide od $[-2.0, -0.33]$ možemo vidjeti da nula definitivno nije unutra te da je srednja vrijednost prihoda populacije negativna. Naslućuje se kako pojedinci generalno gube a tržište pobjeđuje.

```
t.test(otp, mu=0, alternative = "two.sided")
```

```
##
## One Sample t-test
##
## data: otp
## t = -2.7571, df = 2233, p-value = 0.005879
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -2.0040814 -0.3381478
## sample estimates:
## mean of x
## -1.171115
```

Drugi test koji se provodi nad podacima je na medijanima prihoda iznad CROBEX-a kroz godine.

Testira se null hipoteza koja govori da je srednja vrijednost medijana jednaka nuli uz alternativnu hipotezu da je srednja vrijednost medijana različita od nule. Dobivena p-vrijednost je 35% te se ne može odbaciti nulta

hipoteza čak ni uz osjetno veće alphe od 5%, 95-postotni interval pouzdanosti je između -5 i 2 te naginje negativnom iznosu, također srednja vrijednost medijana kroz godine je negativan broj iznosa -1.5.

```
medians = tapply(dionice$Prinos.iznad..OTP.indeksnog.fonda, dionice$Godina, median)
medians
```

```
## 2009 2010 2011 2012 2013 2014 2015 2016 2017
## -0.48 -8.91 -3.46 5.63 -3.92 -3.87 -2.87 4.94 -0.59
```

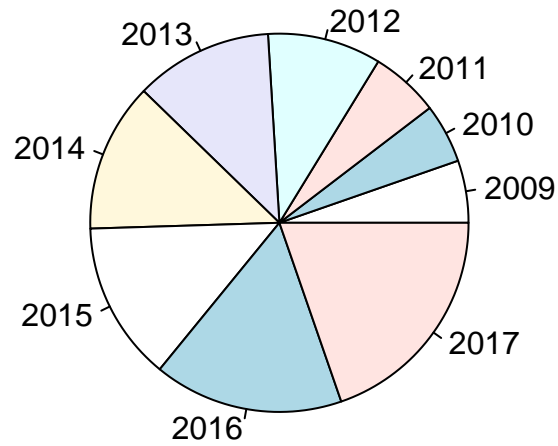
```
t.test(medians, mu=0, alt="two.sided")
```

```
##
## One Sample t-test
##
## data: medians
## t = -0.98924, df = 8, p-value = 0.3515
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -5.007722 2.001055
## sample estimates:
## mean of x
## -1.503333
```

FUN FACT: Broj prijavljenih ekipa raste tokom godina

```
#pie chart
pie(table(dionice$Godina), main = "Strukturni krug broja sudjelovanja po godinama")
```

Strukturni krug broja sudjelovanja po godinama

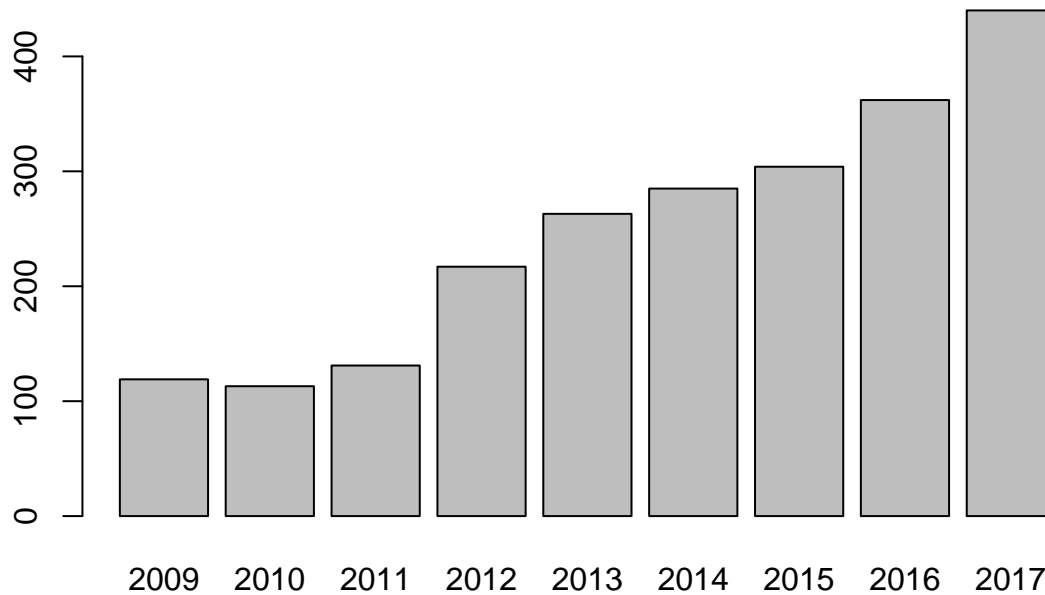


```
#barplot
barplot(table(dionice$Godina), main = "Stupičasti dijagram broja sudjelovanja po godinama")

## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
## conversion failure on 'Stupičasti dijagram broja sudjelovanja po godinama'
## in 'mbcsToSbcs': dot substituted for <c4>

## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
## conversion failure on 'Stupičasti dijagram broja sudjelovanja po godinama'
## in 'mbcsToSbcs': dot substituted for <8d>
```

Stupčasti dijagram broja sudjelovanja po godinama



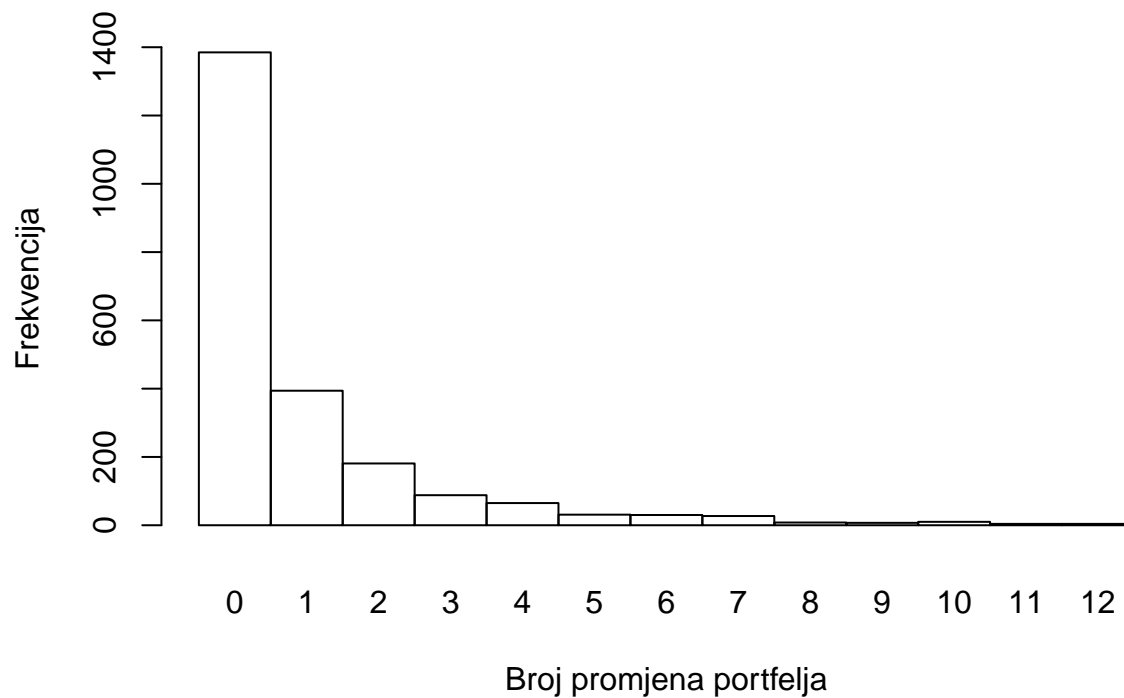
Broj promjena portfelja

Donji kod parsira stupac s dionicama te iz njega uzima broj promjena portfelja svakog natjecatelja kojeg potom dodaje u zasebni stupac zvan Broj.promjena.

```
parser = function(string) {  
  s = strsplit(as.character(string), "\\s+")[[1]][1]  
  s = gsub("[()]", "", s)  
  s = gsub("[ ]", "", s)  
  return(as.numeric(s))  
}  
dionice$Broj.promjena = dionice$Dionice  
dionice$Broj.promjena = unlist(lapply(dionice$Broj.promjena, parser))  
  
# save(dionice, file = "dionice.Rdata")
```

Većina natjecatelja nije sklonio mijenjanju portfelja te je to prikazano histogramom dolje.

```
hist(dionice$Broj.promjena, main=NULL, xlab = "Broj promjena portfelja", ylab = "Frekvencija", breaks =  
axis(side=1, at=seq(min(dionice$Broj.promjena), max(dionice$Broj.promjena), 1), tcl=0, col="white")
```



Korelacija između stupca broj promjena i prihoda iznad CROBEX-a je jako malena (pozitivna) te ne možemo tvrditi da broj promjena povećava šanse za uspjeh.

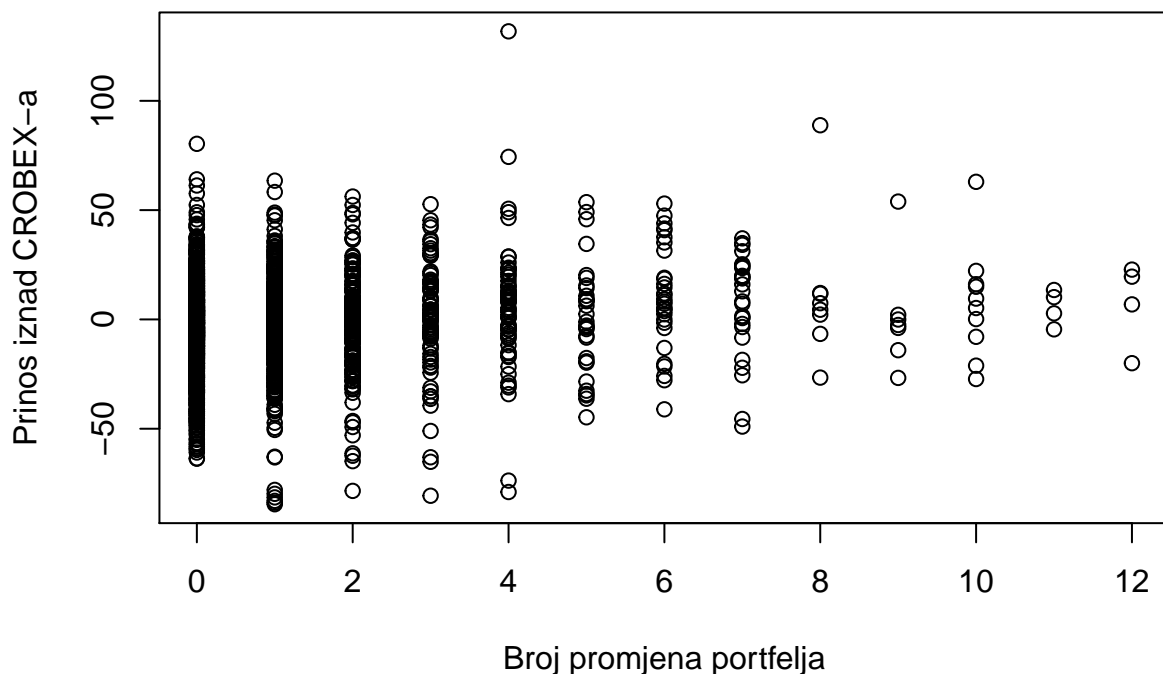
```
cor(dionice$Prinos.iznad..OTP.indeksnog.fonda, dionice$Broj.promjena);
```

```
## [1] 0.1022367
```

```
# cor(dionice$Prinos.iznad..OTP.indeksnog.fonda, dionice$Broj.promjena, method = "kendall")  
# cor(dionice$Prinos.iznad..OTP.indeksnog.fonda, dionice$Broj.promjena, method = "spearman")
```

Dijagram raspršenja koji na x-osi ima broj promjena dok na y-osi ima prihod iznad CROBEX-a je prikazan dole.

```
plot(dionice$Broj.promjena, dionice$Prinos.iznad..OTP.indeksnog.fonda, xlab = "Broj promjena portfelja")
```



Primjećuje se da su medijani i srednje vrijednosti prihoda iznad CROBEX-a gledano posebno po broju mijenjanja portfelja pretežito pozitivni brojevi.

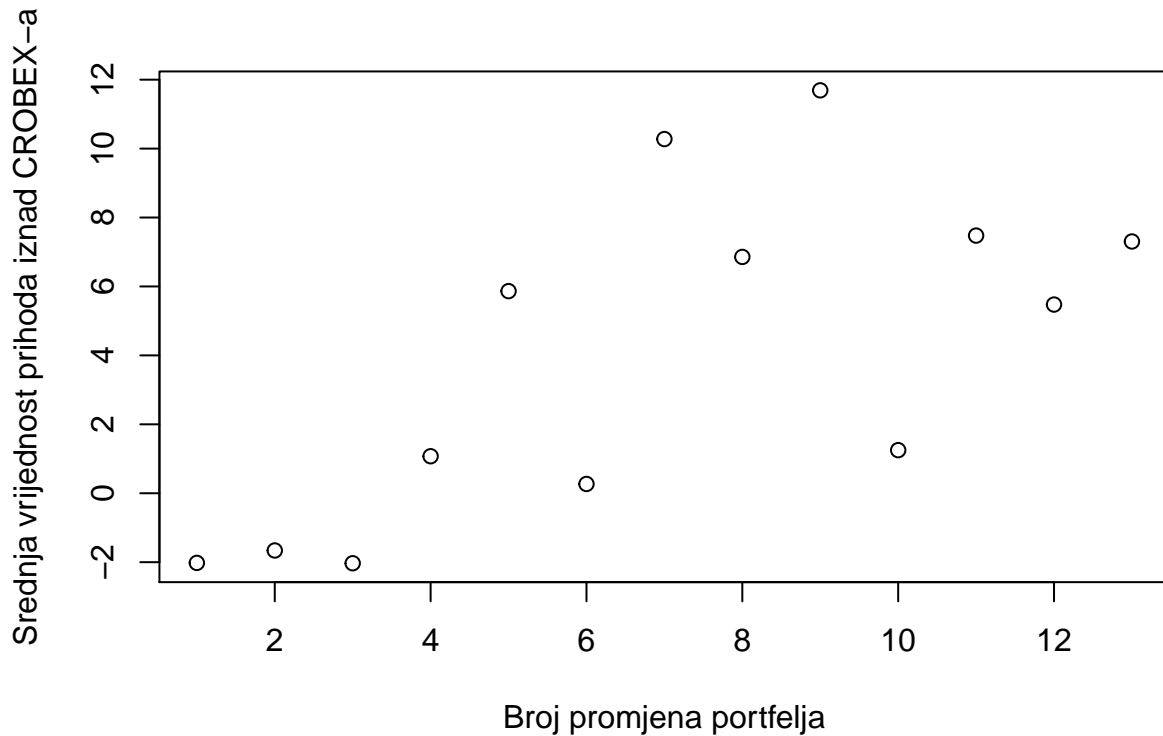
```
tapply(dionice$Prinos.iznad..OTP.indeksnog.fonda, dionice$Broj.promjena, summary)
```

```
## $`0`
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -63.700 -11.640   -1.600   -2.022   8.760   80.330
##
## $`1`
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -84.540 -12.535    0.030   -1.661   11.123   63.470
##
## $`2`
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -78.430 -15.900   -0.840   -2.031    9.880   56.210
##
## $`3`
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -80.650  -7.798    2.025    1.074   16.340   52.710
##
## $`4`
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -78.950  -6.310    8.160    5.864   15.680  131.780
##
## $`5`
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
```

```
## -44.7600 -18.4050 -1.0700 0.2693 12.5850 53.6400
##
## $`6`
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##    -41.12  -1.15    8.70   10.28   28.38   53.00
##
## $`7`
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   -49.010  -2.765  12.860    6.857   23.790   37.090
##
## $`8`
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  -26.6100  -0.0025   5.9400   11.6900  11.8325  88.8300
##
## $`9`
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   -26.75   -8.97   -2.55    1.25    1.04   53.91
##
## $`10`
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  -27.260  -5.923   7.370    7.476   15.820   62.950
##
## $`11`
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   -4.540   0.950   6.485    5.475   11.010   13.470
##
## $`12`
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  -20.0100   0.1575  13.2200    7.3050  20.3675  22.7900
```

Slika srednje vrijednosti po broju promjena prihoda iznad CROBEX-a otkriva zanimljivu činjenicu da je srednja vrijednost negativna samo kada je broj promjena portfelja manji od 3.

```
plot(tapply(dionice$Prinos.iznad..OTP.indeksnog.fonda, dionice$Broj.promjena, mean), xlab = "Broj promjena")
```



Ukazuje se na činjenicu da makar srednja vrijednost rasla s brojem promjena to ne znači puno jer je situacija takva da je većina podataka grupirana na malom broju promjena portfelja. Čak 87.7% sudionika je mijenjalo portfelj manje od 3 puta.

```
tapply(dionice$Prinos.iznad..OTP.indeksnog.fonda, dionice$Broj.promjena, length)
```

```
##      0      1      2      3      4      5      6      7      8      9     10     11     12
## 1385   394   181    88    65    31    30    27     8     7    10     4     4
```

Kako su velike razlike između broja promjena portfelja s puno sudionika i malo sudionika napravljena je linearna regresija koja kao izlaz gleda srednje vrijednosti prinosa iznad CROBEX-a te kao ulaz u linearnu funkciju promatra broj promjena portfelja.

```
mean_broj_promjena = tapply(dionice$Prinos.iznad..OTP.indeksnog.fonda, dionice$Broj.promjena, mean)
mean_broj_promjena_df = data.frame(mean_broj_promjena)
sort(unique(dionice$Broj.promjena))
```

```
## [1] 0 1 2 3 4 5 6 7 8 9 10 11 12
```

```
mean_broj_promjena_df$Broj.promjena = sort(unique(dionice$Broj.promjena))
names(mean_broj_promjena_df)[1] = "Mean"
```

```
fit = lm(Mean~Broj.promjena, data = mean_broj_promjena_df)
fit
```

```
##
```

```
## Call:
```

```
## lm(formula = Mean ~ Broj.promjena, data = mean_broj_promjena_df)
```

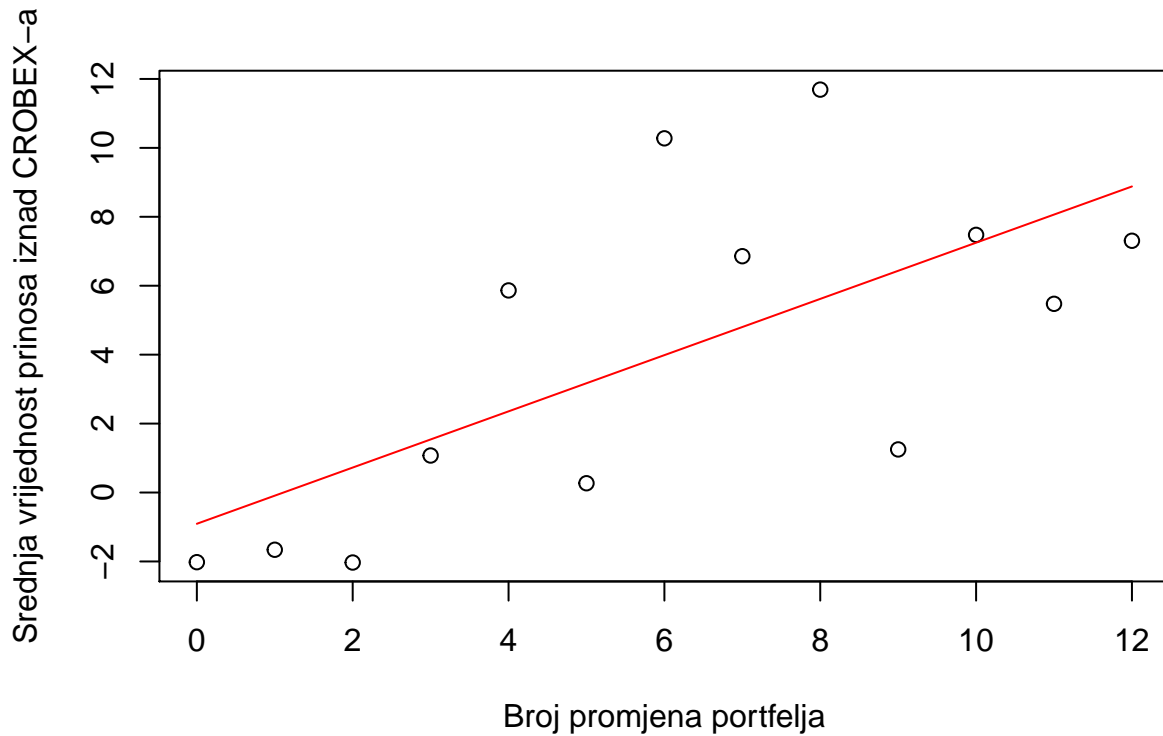
```
##
```



```
## Coefficients:
## (Intercept) Broj.promjena
## -0.9071      0.8156
```

```
#cor(mean_broj_promjena_df$Mean, mean_broj_promjena_df$Broj.promjena)
```

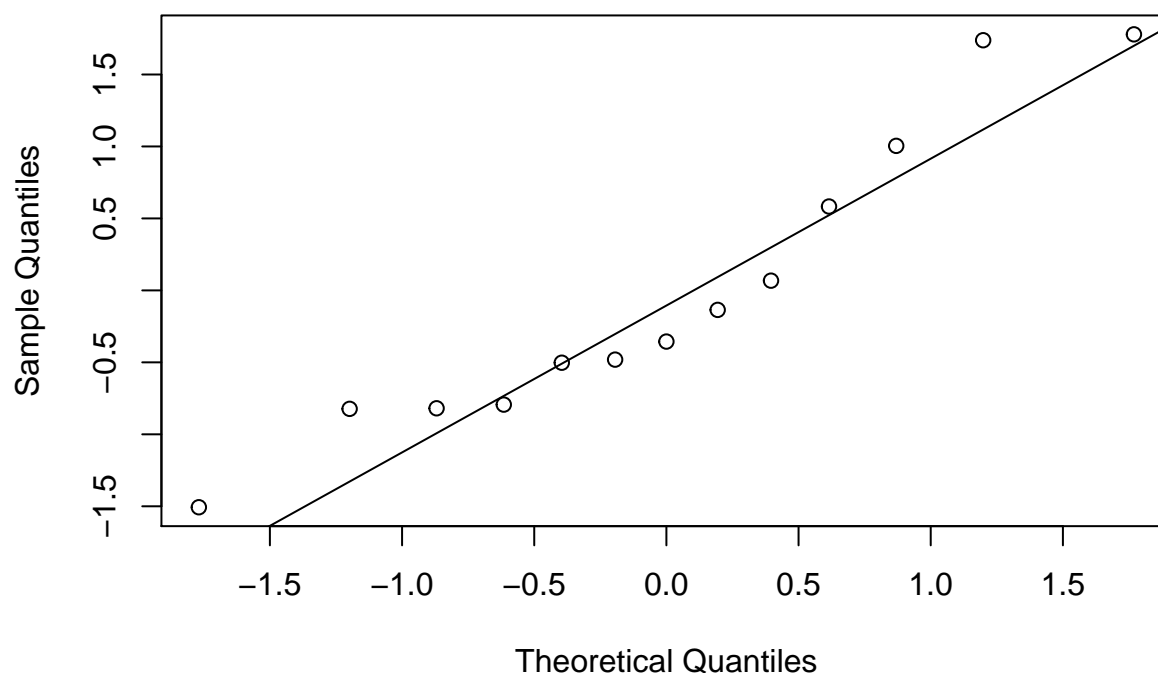
```
plot(mean_broj_promjena_df$Broj.promjena, mean_broj_promjena_df$Mean, xlab = "Broj promjena portfelja",
lines(mean_broj_promjena_df$Broj.promjena, fit$fitted.values, col="red"))
```



Pretpostavka linearne regresije je da reziduali dolaze iz normalne distribucije. Ta pretpostavka se može provjeriti pomoću kvantil kvantil prikaza te je takav prikaz dan ispod. Na njemu se naslučuje da bi reziduali mogli dolaziti iz normalne distribucije.

```
qqnorm(rstandard(fit), main = "Kvantil-Kvantil plot")
qqline(rstandard(fit))
```

Kvantil–Kvantil plot



Statistički test za provjeru normalnosti reziduala je test Kolmogorov-Smirnov. Rezultat testa je p vrijednost od čak 74.6 posto te ona govori da se nulta hipoteza (hipoteza da reziduali dolaze iz normalne distribucije) ne može odbaciti uz pogrešku prve vrste od 5 posto.

```
ks.test(rstandard(fit), 'pnorm')
```

```
##
## One-sample Kolmogorov-Smirnov test
##
## data:  rstandard(fit)
## D = 0.17737, p-value = 0.7458
## alternative hypothesis: two-sided
```

Promatrajući p vrijednost dobivenu F-testom čija je vrijednost iznosa 1.2 posto odlučeno je da se može odbaciti nulta hipoteza koja govori da broj promjena portfelja ne ovisi o srednjoj vrijednosti prinosa iznad CROBEX-a.

```
summary(fit)
```

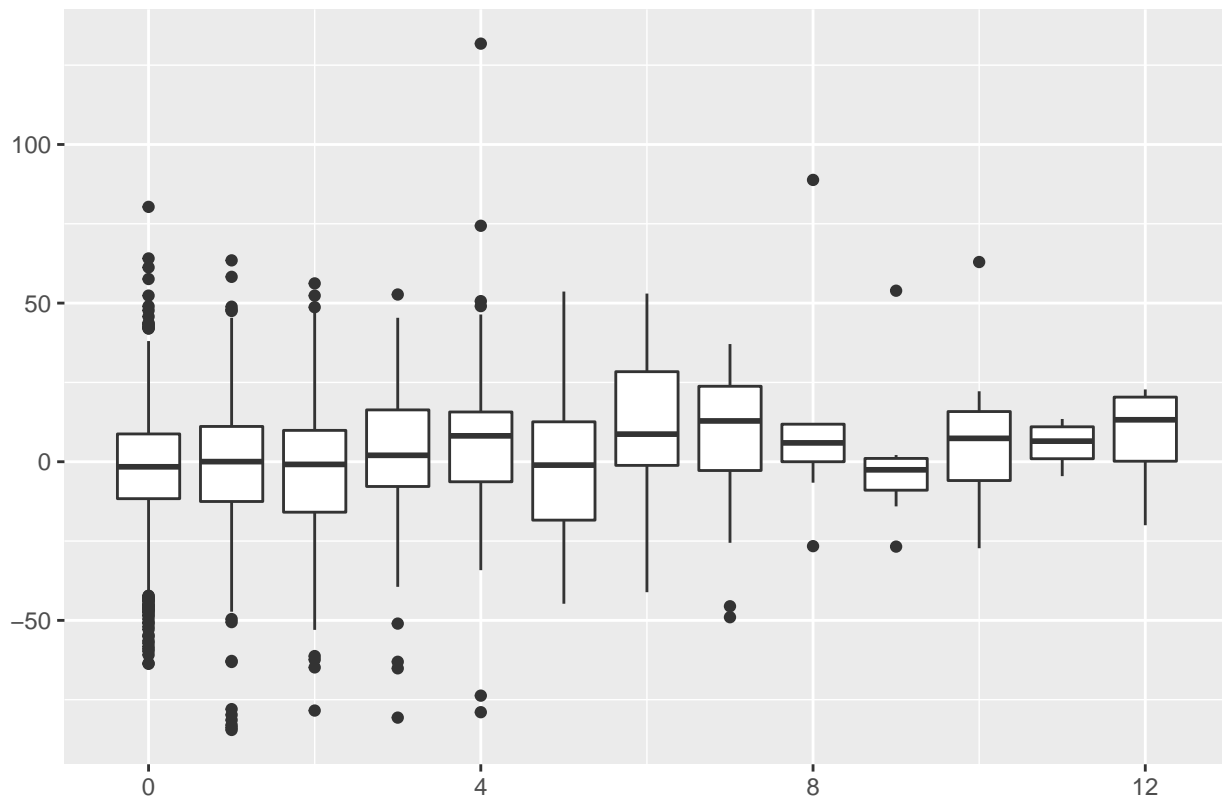
```
##
## Call:
## lm(formula = Mean ~ Broj.promjena, data = mean_broj_promjena_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.183 -2.589 -1.115  2.055  6.291
##
## Coefficients:
```

```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -0.9071     1.9291  -0.470   0.6474
## Broj.promjena  0.8156     0.2728   2.989   0.0123 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.681 on 11 degrees of freedom
## Multiple R-squared:  0.4483, Adjusted R-squared:  0.3981
## F-statistic: 8.937 on 1 and 11 DF,  p-value: 0.01231
```

Boxplotovi prinosa iznad CROBEX-a po broju mijenjanja portfelja pokazani zajedno su dani dole.

```
ggplot(dionice, aes(x=Broj.promjena, y=Prinos.iznad..OTP.indeksnog.fonda, group=Broj.promjena)) + geom_boxplot()
```

Pravokutni dijagram prihoda iznad CROBEX-a po broju promjena portfelja



T test čija nulta hipoteza govori da je srednja vrijednost prihoda iznad CROBEX-a sudionika s nula promjena portfelja jednaka srednjoj vrijednosti onih koji imaju 1 ili više promjenu portfelja. Alternativna hipoteza je ta da onima s više promjena ide bolje. Rezultat testa je p vrijednost od 99.2% čime ne samo da se ne može odbaciti nulta hipoteza u korist alternativne nego je situacija takva da onima s 0 promjena ide bolje. Da je test napravljen s alternativnom hipotezom da onima s nula promjena ide bolje mogla bi se odbaciti nulta hipoteza.

```
nula_promjena = dionice[dionice$Broj.promjena > 0, ]$Prinos.iznad..OTP.indeksnog.fonda
više_od_nula_promjena = dionice[dionice$Broj.promjena == 0, ]$Prinos.iznad..OTP.indeksnog.fonda
length(nula_promjena) + length(više_od_nula_promjena) == length(otp)
```

```
## [1] TRUE
```

```
t.test(nula_promjena, više_od_nula_promjena, alternative = "less")
```

```
##
## Welch Two Sample t-test
##
## data: nula_promjena and više_od_nula_promjena
## t = 2.4091, df = 1460.9, p-value = 0.9919
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf 3.766639
## sample estimates:
## mean of x mean of y
## 0.2162309 -2.0215523
```

Sudionici natjecanja koji se ponavljaju tokom godina

Pretpostavlja se da natjecatelj ulazi u natjecanje svake godine s istim imenom te da ljudi nisu radili više različitih računa kako bi sudjelovali u natjecanju. Te tvrdnje su vjerovatno prekršene.

Nalazimo one čije se ime barem 2 puta ponovilo.

```
name.occurence = data.frame(table(dionice$Naziv.portfelja))
more.than.2.times = name.occurence[name.occurence$Freq > 1, ]
```

U 2234 podataka ima samo 1347 jedinstvenih igrača te od tih 1347 ima 430 igrača koji su bili barem dva puta te njih 917 koji su bili samo jednom.

```
length(name.occurence$Var1); length(more.than.2.times$Var1); length(name.occurence$Var1) - length(more.than.2.times$Var1)
```

```
## [1] 1347
```

```
## [1] 430
```

```
## [1] 917
```

Na temelju histograma broja sudjelovanja uočava se činjenica da velika većina natjecatelja sudjeluje relativno malen broj puta no čak njih 7 je sudjelovalo kroz sve godine trajanja natjecanja.

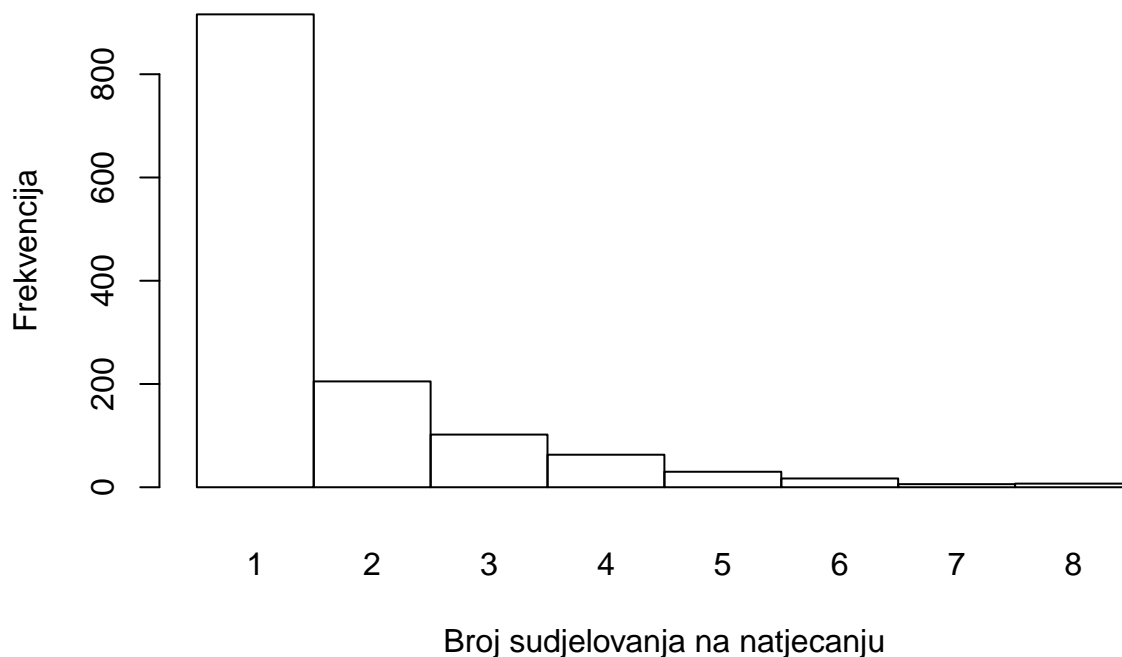
```
colnames(name.occurence) = c('Naziv.portfelja', 'Broj.sudjelovanja')
name.occurence = name.occurence[!name.occurence$Naziv.portfelja == "", ]
table(name.occurence$Broj.sudjelovanja)
```

```
##
```

```
##  1  2  3  4  5  6  7  8
```

```
## 916 205 102 63 30 17 6 7
```

```
hist(name.occurence$Broj.sudjelovanja, breaks = seq(min(name.occurence$Broj.sudjelovanja)- 0.5, max(name.occurence$Broj.sudjelovanja), 1), axis(side=1, at=seq(min(name.occurence$Broj.sudjelovanja), max(name.occurence$Broj.sudjelovanja), 1), t
```



Donosi li iskustvo prednost?

Za one koji imaju više sudjelovanja je prinos iznad CROBEX-a sveden na srednju vrijednost svih njihovih sudjelovanja. Dole u varijabli gusteri je spremljen prinos iznad CROBEX-a od sudionika koji su bili samo jednom na natjecanju dok varijbla starosjedioci opisuje sudionike koji su bili više od jedanput.

```
spojeni_otp = aggregate(dionice$Prinos.iznad..OTP.indeksnog.fonda, by=list(dionice$Naziv.portfelja), FUN=
colnames(spojeni_otp) = c("ID", "otp")
```

```
gusteri = spojeni_otp[!(spojeni_otp$ID %in% more.than.2.times$Var1), ]
starosjedioci = spojeni_otp[spojeni_otp$ID %in% more.than.2.times$Var1, ]
length(gusteri$ID); length(starosjedioci$ID)
```

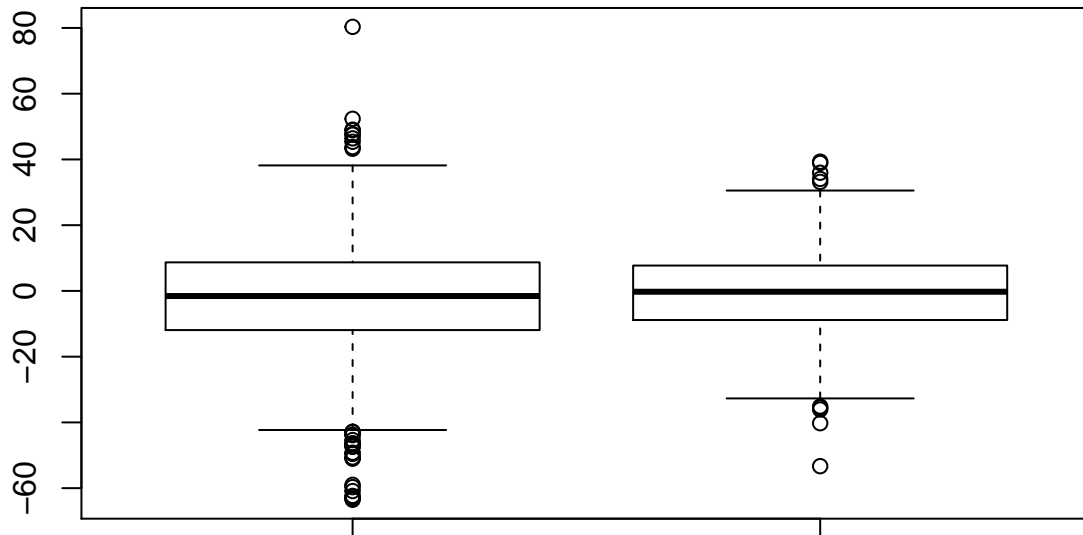
```
## [1] 916
```

```
## [1] 430
```

Uspoređivanje boxplotova guštera i starosjedioca. Na temelju boxplotova ne vidi se neka prevelika razlika osim što se uočava da je medijan starosjedioca malo iznad medijana guštera.

```
boxplot(gusteri$otp, starosjedioci$otp, main="Pravokutni dijagram usporedbe neiskusnih i iskusnih sudionika")
```

Pravokutni dijagram usporedbe neiskusnih i iskusnih sudionika



Pokazuje se da su i medijan i srednja vrijednost prinosa iznad CROBEX-a viši kod onih koji su bili barem 2 puta na natjecanju.

```
summary(gusteri$otp); summary(starosjedioci$otp)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -63.530 -11.910  -1.555   -2.015   8.660   80.330

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -53.310  -8.841  -0.245   -1.038   7.697   39.325
```

Radimo t test s nultom hipotezom da su srednje vrijednosti prinosa iznad CROBEX-a iste kod starosjedioca i kod gusteri te s alternativnom hipotezom da starosjedioci ide bolje. Dobivena je p vrijednost od 13.3% na temelju koje se ne može odbaciti nulta hipoteza uz razinu značajnosti od 5%. Ima razlike među njima no ona nije dovoljno značajna da odbacimo hipotezu da su jednaki.

```
t.test(starosjedioci$otp, gusteri$otp, alternative = "greater")
```

```
##
## Welch Two Sample t-test
##
## data: starosjedioci$otp and gusteri$otp
## t = 1.1127, df = 1112.2, p-value = 0.133
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## -0.4686764      Inf
```

```
## sample estimates:
## mean of x mean of y
## -1.037501 -2.015076
```

Možda ako sudionicima koji su se pojavili više od jedanput na natjecanju spojimo rezultate uzimajući samo njihovu zadnju vrijednost, umjesto srednje vrijednosti pokušaja, dobijemo bolje rezultate. Pretpostavka koja nas navodi na ovako nešto je ta da sudionicima ide bolje i bolje s vremenom kako skupljaju iskustvo.

```
spojeni_otp_zadnja_vrijednost = aggregate(dionice$Prinos.iznad..OTP.indeksnog.fonda, by=list(dionice$Na
colnames(spojeni_otp_zadnja_vrijednost) = c("ID", "otp")

starosjedioci_zadnja_vrijednost = spojeni_otp_zadnja_vrijednost[spojeni_otp_zadnja_vrijednost$ID %in% m
length(gusteri$ID); length(starosjedioci_zadnja_vrijednost$ID)
```

```
## [1] 916
```

```
## [1] 430
```

Napravljen je test srednje vrijednosti prinosa iznad CROBEX-a starosjedioca i gusteri gdje su starosjediocima uzimali zadnje vrijednosti kao njihov rezultat. Alternativna hipoteza je ta da starosjediocima ide bolje. Dobivena je p vrijednost u iznosu od 97.3% te hipotezu da su starosjedioci bolji definitivno ne možemo shvatiti ozbiljno. Da je tu napravljen test s alternativnom hipotezom koja govori da starosjediocima ide lošije onda bi se nulta hipoteza mogla odbaciti uz razinu značajnosti od 3%. Uspoređujući njihove srednje vrijednosti zaista je istina da su starosjedioci ispali lošiji u ovom novom scenariju.

```
t.test(starosjedioci_zadnja_vrijednost$otp, gusteri$otp, alternative = "greater")
```

```
##
## Welch Two Sample t-test
##
## data: starosjedioci_zadnja_vrijednost$otp and gusteri$otp
## t = -1.9295, df = 660.64, p-value = 0.973
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## -4.72827 Inf
## sample estimates:
## mean of x mean of y
## -4.565837 -2.015076
```

Pretpostavljeno je iznad da će sudionicima zadnja vrijednost na natjecanju biti bolja od srednje vrijednosti. T test te tvrdnje s alternativnom hipotezom da je zadnja vrijednost bolja od srednje vrijednosti je napravljen dole. Rezultat tog testa je p vrijednost od 99.57% te ne samo da se ne može odbaciti nulta hipoteza nego ispada da se sudionici nisu poboljšali kroz godine te bolje rezultate dobiju računajući srednju vrijednost od napravljenih rezultata.

```
t.test(starosjedioci_zadnja_vrijednost$otp, starosjedioci$otp, alternative = "greater")
```

```
##
## Welch Two Sample t-test
##
## data: starosjedioci_zadnja_vrijednost$otp and starosjedioci$otp
## t = -2.6321, df = 662.61, p-value = 0.9957
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## -5.736385 Inf
## sample estimates:
## mean of x mean of y
## -4.565837 -1.037501
```

Postoji li korelacija između broja mijenjanja portfelja i broja sudjelovanja na natjecanju?

To jest mijenjaju li starosjedioci češće portfelj. Nekako je intuitivno za pretpostaviti da sudionici koji često dolaze na natjecanje često i mijenjaju svoj portfelj iskušavajući neke nove taktike. Dobivena korelacija između mijenjanja portfelja i frekvencije sudjelovanja je iznosa 0.14. Nije jaka korelacija, no starosjedioci ipak malo češće mijenjaju portfelj.

```
colnames(name.occurence) = c("Naziv.portfelja", "Broj.sudjelovanja")
# dionice_broj_sudjelovanja = merge(dionice, name.occurence, by="Naziv.portfelja")
# dionice_broj_sudjelovanja = dionice_broj_sudjelovanja[with(dionice_broj_sudjelovanja, order(Godina, -

broj_sudjelovanja_naziva = rep(0.0, length(name.occurence$Naziv.portfelja))
names(broj_sudjelovanja_naziva) = name.occurence$Naziv.portfelja
pom = c()
for (naziv in dionice$Naziv.portfelja){
  broj_sudjelovanja_naziva[naziv] = broj_sudjelovanja_naziva[naziv] + 1
  pom = c(pom, broj_sudjelovanja_naziva[naziv])
}
dionice$Broj.Sudjelovanja = pom

cor(dionice$Broj.Sudjelovanja, dionice$Broj.promjena)
```

```
## [1] 0.1440001
```

Korelacija između prinosa iznad CROBEX-a i broja sudjelovanja na natjecanju je iznosa -0.0038 te je zanemariva.

```
cor(dionice$Prinos.iznad..OTP.indeksnog.fonda, dionice$Broj.Sudjelovanja)
```

```
## [1] -0.00386311
```

Najčešće odabrana dionica te dionica koja ima najveću korelaciju s ostvarenim prihodom iznad crobexa je dionica naziva RIVP-R-A. To je dionica koja pripada kategoriji pod turizmi i hoteli. Duštvo koje stoji iza nje je VALAMAR RIVIERA D.D., te ako sam shvatio bave se upravljanjem turističkih mjesta poput hotela, kampova i ljetovališta. Korelacija te dionice iznosi 0.185 s prihodom iznad CROBEX-a te od 2234 sudionika kroz ovih 8 godina njih čak 623 je nju imalo u svojem portfelju.

Prvo moramo parsirati stupac Dionice i dobiti listu stringova(factora) koji predstavljaju dionice

```
parser = function(string) {
  s = strsplit(as.character(string), "\\s+")[[1]][2]
  return (unlist(strsplit(s, "[,]")))
}
parser(dionice$Dionice[1])
```

```
## [1] "PTKM-R-A" "ADRS-P-A" "INA-R-A"
```

```
parser(factor(dionice$Dionice[1]))
```

```
## [1] "PTKM-R-A" "ADRS-P-A" "INA-R-A"
```

```
dionice$Lista.Dionica = lapply(dionice$Dionice, parser)
# dionice$Lista.Dionica[1]
# head(dionice)
# summary(dionice$Lista.Dionica)
```

```
# install.packages("XML")
# install.packages("RCurl")
```



```

# install.packages("qdapTools")

library(qdapTools)
# dionice[1]
binary_dionice = cbind(dionice$RB, dionice$Naziv.portfelja, mtabulate(dionice$Lista.Dionica))

# colSums(binary_dionice)
# binary_dionice[, -c(1,2)] # omitting first and second column
binary_samo_dionice = binary_dionice[, -c(1, 2)]

sort(colSums(binary_samo_dionice), decreasing = T)[1]

## RIVP-R-A
##      623
correlation_dionice_crobex = cor(binary_samo_dionice, dionice$Prinos.iznad..OTP.indeksnog.fonda)
which(correlation_dionice_crobex == max(correlation_dionice_crobex), arr.ind = T)

##      row col
## RIVP-R-A 101  1
correlation_dionice_crobex[101, 1]

## RIVP-R-A
## 0.1845373

```