

Cijene nekretnina

Kristijan Cetina

15.02.2020.

```
#cat("\f"); graphics.off(); rm(list = ls())
setwd("~/Documents/remax-house-pricing")
options(digits=3)
options(scipen=5)
library("gvlma")
library("corrgram")
```

```
## Registered S3 method overwritten by 'seriation':
##   method      from
##   reorder.hclust gclus
df = read.csv("houses_full.csv", sep = ",", header = TRUE, dec=",")
summary(df)
```

```
##                                                                 X
## http://www.remax-svijetnekretnina.com/apartman/akcijska-cijena-novogradnje-u-liznjanu
## http://www.remax-svijetnekretnina.com/apartman/apartman-s-garazom-u-zgradi-s-liftom-blizu-plaze
## http://www.remax-svijetnekretnina.com/apartman/apartman-u-fazani-100m-od-mora-i-s-privatnim-bazenom
## http://www.remax-svijetnekretnina.com/apartman/dvoetazni-apartmansa-tri-terase-i-pogledom-na-more
## http://www.remax-svijetnekretnina.com/apartman/istra-apartman-u-atraktivnome-medulinu-sa-terasom
## http://www.remax-svijetnekretnina.com/apartman/istra-banjole-apartman-s-pogledom-na-more1
## (Other)
##   Lokacija.      Transakcija.      Vrsta.nekretnine.
## Poreč      : 230      Za prodaju:3197      Kuća/Vikendica      :1292
## Medulin    : 147      Građevinsko zemljište : 882
## Rovinj     : 128      Stan/Apartman      : 600
## Vodnjan    : 121      Negrađevinsko zemljište: 291
## Umag       : 115      Poslovni prostor   : 99
## Novigrad   : 82      Ugostiteljski objekti : 19
## (Other)    :2374      (Other)            : 14
## Ukupan.broj.soba.      Toaleti.      Ukupno.katova.
## Min.      : 1      Min.      : 1      Min.      : 1
## 1st Qu.: 3      1st Qu.: 1      1st Qu.: 2
## Median    : 4      Median    : 1      Median    : 2
## Mean      : 6      Mean      : 2      Mean      : 2
## 3rd Qu.: 7      3rd Qu.: 2      3rd Qu.: 3
## Max.      :86      Max.      :86      Max.      :22
## NA's      :1222      NA's      :2651      NA's      :1308
##   Cijena.      Površina.      Okućnica.
## na upit      : 73      150      : 41      :2636
## 120.000 € (891.099 kn) : 49      200      : 39      ima      : 44
## 250.000 € (1.856.456 kn): 37      180      : 33      nema okućnice : 295
## 130.000 € (965.357 kn) : 35      160      : 31      u vlasništvu : 49
```

```

## 160.000 € (1.188.132 kn): 32 120 : 29 zajednička : 72
## 150.000 € (1.113.873 kn): 31 100 : 28 zajednička na korištenje: 101
## (Other) :2940 (Other):2996
## Godina.izgradnje. Nekretnina..stanje.. Stil.gradnje.
## :1566 :1556 :1487
## 1900 god.: 120 dovršena :1330 autohtoni : 356
## 2019 god.: 90 nedovršena: 311 klasičan : 809
## 2018 god.: 73 moderan : 456
## 2017 god.: 58 montažno : 3
## 2007 god.: 53 stara vila: 86
## (Other) :1237
## Razred.energetske.učinkovitosti. Lift. Pristup.invalidima.
## U izradi :1414 :2081 :1732
## :1227 da: 62 da : 394
## B : 164 ne:1054 ne : 692
## Izuzeto od obveze: 124 ne znam : 20
## C : 104 uz manju adaptaciju: 359
## A : 86
## (Other) : 78
## Nekretnina.se.prodaje.iznajmljuje. Visina.plafona.
## :1544 :3177
## djelomično namještena: 152 4 m : 5
## namještaj po dogovoru: 314 3 m : 4
## namještena : 726 6-8,5 m: 2
## nenamještena : 461 2.82 m : 1
## 2.9 m : 1
## (Other): 7
## Namjena.poslovnog.prostora. P.P..sa.izlogom.
## :3103 :3106
## tiha djelatnost : 4 da: 48
## Restoran : 3 ne: 43
## Tiha djelatnost : 3
## Ugostiteljski objekt: 3
## Ugostiteljstvo : 3
## (Other) : 78
## Pristup.kamionima.šleperima. Struja. Parking. Parking..broj..
## :3106 :3106 :1427 :2344
## da: 58 monofazna: 52 da:1537 2 : 250
## ne: 33 trofazna : 39 ne: 233 1 : 220
## 3 : 108
## 4 : 98
## 5 : 50
## (Other): 127
## Natkriveni.parking. Povlašteni.parking. Garaža. Pogled.na.more.
## :1564 :2308 :1509 : 527
## da: 230 da: 167 da: 380 da: 994
## ne:1403 ne: 722 ne:1308 ne:1676
##
##
##
##
## Pogled.na. Udaljenost.od.mora. Ključ.u.agenciji.
## :2402 : 617 :1750
## otvoren : 71 1. red : 57 Da: 147

```

```

## prirodu      : 41  100-300m  : 194      Ne:1300
## more         : 38  2. red   : 75
## Prirodu      : 20  300-500m  : 237
## zelene površine: 18  500-1000m : 320
## (Other)      : 607  više od 1km:1697
##
##                                                     naziv
## Istarska kuća u srednjovjekovnoj jezgri Motovuna (ID: 300041087-35)      : 1
## Istra, Novigrad, Dajla kuća na samoj obali mora (ID: 300441007-49)      : 1
## Istra, zemljište sa pogledom 360°, jedinstveno (ID: 300041016-1025)     : 1
## Istra, Medulin, moderna pametna kuća sa bazenom (ID: 300041049-535)     : 1
## RUSTIKALNA VILA S BAZENOM NA MIRNOJ LOKACIJI BLIZU MORA (ID: 300041026-775): 1
## 1. red do mora/plaže-SAMOSTOJEĆA KUĆA (ID: 300391006-267)               : 1
## (Other)                                             :3191
##
## \n          Građevinska parcela u predivnom zavičaju sjeverne Istre, koju zbog svojeg
## \n          Građevno zemljište u uređenom stambenom naselju udaljenom 5 kilometara od
## \n          Prodaje se penthouse na prvome katu luksuzne novogradnje, planirani dovrš
## \n          Kuća s krovnom terasom - Novigrad: Kuća se nalazi na dobroj i atraktivnoj
## \n          Ova prekrasna Villa moderne arhitekture smjestila se u Vrsaru, izgrađena u
## \n          Poreč, Kukci prodaje se trosoban stan u novogradnji. Nalazi se na drugom l
## (Other)
##
## []
## ['Balkon/lođa/terasa']
## ['Pogled - prekrasan!']
## ['U mirnoj ulici']
## ['Balkon/lođa/terasa', 'Plin']
## ['bez infrastrukture', 'Voda', 'Struja', 'otvoren pogled', 'Zapad', 'Istok', 'U mirnoj ulici', 'U b
## (Other)
## Spavaće.sobe.      Kupaonice.      Površina.okućnice. Površina.balkona.lođe.terase.
## Min.      : 1      Min.      : 1      :1969      :2629
## 1st Qu.: 2      1st Qu.: 1      700      : 25      20      : 42
## Median : 3      Median : 2      200      : 24      10      : 32
## Mean    : 4      Mean    : 3      500      : 24      15      : 32
## 3rd Qu.: 4      3rd Qu.: 3      1.000    : 23      30      : 24
## Max.    :54      Max.    :86      300      : 23      12      : 21
## NA's    :1339    NA's    :1947    (Other):1109    (Other): 417
##          Roštilj.      Dimnjak.      Bazen.      Vrt.
##          :2893      :1656      :2246      :2583
## na korištenje: 12 ima :1286 na korištenje: 19 na korištenje: 47
## nema      : 261 nema: 255 nema      : 780 nema      : 375
## vlastiti   : 31      vlastiti   : 152 vlastiti   : 192
##
##
##
## Vrsta.kuće.vikendice. Tlocrtna.površina.nekretnine. Mogućnost.zamjene.
##          :2169      :2186      :2739
## dvojna    : 158      100      : 56      da: 24
## samostojeća: 626      120      : 40      ne: 434
## u nizu    : 244      110      : 37
##          80      : 35
##          90      : 30
##          (Other): 813
##          Namjena.terena. Površina.građevinskog.dijela.

```

```

##                :2246                :2315
## građevinsko    : 140    00    : 59
## Građevinsko    : 108    0,0    : 27
## stambena       : 74    0,00    : 26
## poljoprivredno : 65    1    : 22
## Građevinsko zemljište: 43    -    : 16
## (Other)        : 521    (Other): 732
## Površina.negrađevinskog.dijela. Instalacije.struje. Instalacije.vode.
##                :2381                :3101                :3098
## 00    : 150                2017 god.: 11    2017 god.: 11
## 1    : 90                2016 god.: 9    2016 god.: 9
## 0,00 : 61                2000 god.: 6    2000 god.: 7
## 0,0   : 56                2007 god.: 6    2007 god.: 7
## 0000  : 33                2008 god.: 6    2011 god.: 7
## (Other): 426                (Other) : 58    (Other) : 58
## Pod..parket.laminat.pločice. Pripojene.prostorije. Fasada.
##                :3100                :3111                :3111
## 2017 god.: 11                Drvarnica: 4    2017 god.: 15
## 2016 god.: 8                1    : 3    2015 god.: 9
## 2000 god.: 6                2    : 3    2016 god.: 8
## 2007 god.: 6                garaža : 3    2000 god.: 5
## 2011 god.: 6                DA    : 2    2012 god.: 5
## (Other) : 60                (Other) : 71    (Other) : 44
## Krov. Struja.zakupljena.snaga. Vanjska.stolarija.
##                :3102                :3180                :3089
## 2017 god.: 11    23 kWh : 2    2016 god.: 12
## 2015 god.: 8    10 kWh : 1    2017 god.: 10
## 2000 god.: 6    13 kWh : 1    2013 god.: 8
## 2006 god.: 6    17 kWh : 1    2012 god.: 7
## 2010 god.: 5    210 kWh: 1    2015 god.: 7
## (Other) : 59    (Other): 11    (Other) : 64
## Unutarnja.stolarija. Kupaonica..kuhinja.
##                :3111                :3100
## 2017 god.: 9    2017 god.: 10
## 2000 god.: 8    2016 god.: 9
## 2016 god.: 8    2000 god.: 8
## 2007 god.: 7    2011 god.: 7
## 2015 god.: 6    2006 god.: 6
## (Other) : 48    (Other) : 57

```

```
str(df)
```

```

## 'data.frame':    3197 obs. of  58 variables:
## $ X                : Factor w/ 3197 levels "http://www.remax-svijetnekretnina.com/..."
## $ Lokacija.        : Factor w/ 348 levels "", "Anžići, Višnjan",...: 323 271 284 258 ...
## $ Transakcija.     : Factor w/ 1 level "Za prodaju": 1 1 1 1 1 1 1 1 1 ...
## $ Vrsta.nekretnine. : Factor w/ 10 levels "Garaža","Građevinsko zemljište",...: 7 8 ...
## $ Ukupan.broj.soba. : int  1 3 12 3 NA NA 7 2 NA 10 ...
## $ Toaleti.         : int  1 1 NA NA NA NA NA NA NA 1 ...
## $ Ukupno.katova.    : int  1 1 3 3 NA NA 2 4 NA 3 ...
## $ Cijena.           : Factor w/ 1250 levels "1.000.000 € (7.425.823 kn)",...: 891 10 ...
## $ Površina.         : Factor w/ 1710 levels "", "1.000 ", "1.002 ",...: 612 1243 1063 ...
## $ Okućnica.         : Factor w/ 6 levels "", "ima", "nema okućnice",...: 3 6 1 3 1 1 ...
## $ Godina.izgradnje. : Factor w/ 180 levels "", "- god.", "00000 god.",...: 124 35 114 ...
## $ Nekretnina..stanje.. : Factor w/ 3 levels "", "dovršena",...: 2 2 2 2 1 1 2 2 1 3 ...

```

```

## $ Stil.gradnje. : Factor w/ 6 levels "", "autohtoni", ...: 4 3 3 3 1 1 3 3 1 2 ...
## $ Razred.energetske.učinkovitosti. : Factor w/ 11 levels "", "A", "A+", "B", ...: 4 11 11 11 1 1 11 3 1 ...
## $ Lift. : Factor w/ 3 levels "", "da", "ne": 3 3 3 3 1 1 2 3 1 3 ...
## $ Pristup.invalidima. : Factor w/ 5 levels "", "da", "ne", "ne znam", ...: 2 5 3 3 1 1 3 3 ...
## $ Nekretnina.se.prodaje.iznajmljuje. : Factor w/ 5 levels "", "djelomično namještena", ...: 4 3 4 2 1 1 ...
## $ Visina.plafona. : Factor w/ 13 levels "", "2.82 m", "2.9 m", ...: 5 1 1 1 1 1 1 1 1 1 ...
## $ Namjena.poslovnog.prostora. : Factor w/ 81 levels "", "APARTMANI", ...: 41 1 1 1 1 1 1 1 1 1 ...
## $ P.P..sa.izlogom. : Factor w/ 3 levels "", "da", "ne": 2 1 1 1 1 1 1 1 1 1 ...
## $ Pristup.kamionima.šleperima. : Factor w/ 3 levels "", "da", "ne": 2 1 1 1 1 1 1 1 1 1 ...
## $ Struja. : Factor w/ 3 levels "", "monofazna", ...: 2 1 1 1 1 1 1 1 1 1 ...
## $ Parking. : Factor w/ 3 levels "", "da", "ne": 2 3 2 2 1 1 2 3 1 2 ...
## $ Parking..broj.. : Factor w/ 41 levels "", " 4", " 2", "1", ...: 38 1 25 4 1 1 1 1 1 1 ...
## $ Natkriveni.parking. : Factor w/ 3 levels "", "da", "ne": 3 3 2 3 1 1 3 3 1 3 ...
## $ Povlašteni.parking. : Factor w/ 3 levels "", "da", "ne": 3 3 2 3 1 1 2 3 1 2 ...
## $ Garaža. : Factor w/ 3 levels "", "da", "ne": 3 2 2 3 1 1 3 3 1 3 ...
## $ Pogled.na.more. : Factor w/ 3 levels "", "da", "ne": 3 3 2 3 3 3 3 2 2 2 ...
## $ Pogled.na. : Factor w/ 425 levels "", "ako se izgradi katnica ima pogled mo... : 1 1 1 1 1 1 1 1 1 1 ...
## $ Udaljenost.od.mora. : Factor w/ 7 levels "", "1. red", "100-300m", ...: 6 7 4 7 7 7 7 6 ...
## $ Ključ.u.agenciji. : Factor w/ 3 levels "", "Da", "Ne": 3 3 3 3 1 1 3 3 1 3 ...
## $ naziv : Factor w/ 3197 levels " Istarska kuća u srednjovjekovnoj jezgri... : 1 1 1 1 1 1 1 1 1 1 ...
## $ opis : Factor w/ 3182 levels "\n\n\nIstra, centralna Istra, imanje sa... : 1 1 1 1 1 1 1 1 1 1 ...
## $ dodatne_oznake : Factor w/ 2748 levels "['Balkon/lođa/terasa', 'Centar', 'U blizini... : 1 1 1 1 1 1 1 1 1 1 ...
## $ Spavaće.sobe. : int NA 1 9 1 NA NA 4 1 NA 8 ...
## $ Kupaonice. : int NA 1 NA NA NA NA NA NA NA 1 ...
## $ Površina.okućnice. : Factor w/ 620 levels "", "1.000 ", "1.003 ", ...: 1 198 1 1 1 1 1 1 ...
## $ Površina.balkona.lođe.terase. : Factor w/ 205 levels "", " 20 balkon, 20 terasa ", ...: 1 134 1 1 1 1 1 1 ...
## $ Roštilj. : Factor w/ 4 levels "", "na korištenje", ...: 1 3 1 3 1 1 1 3 1 1 ...
## $ Dimnjak. : Factor w/ 3 levels "", "ima", "nema": 1 2 2 2 1 1 2 2 1 2 ...
## $ Bazen. : Factor w/ 4 levels "", "na korištenje", ...: 1 3 1 3 1 1 1 3 1 1 ...
## $ Vrt. : Factor w/ 4 levels "", "na korištenje", ...: 1 2 1 3 1 1 1 3 1 1 ...
## $ Vrsta.kuće.vikendice. : Factor w/ 4 levels "", "dvojna", "samostojeća", ...: 1 1 3 1 1 1 1 1 ...
## $ Tlocrtna.površina.nekretnine. : Factor w/ 283 levels "", "0.001 ", "100 ", ...: 1 1 47 1 1 1 276 ...
## $ Mogućnost.zamjene. : Factor w/ 3 levels "", "da", "ne": 1 1 1 3 1 1 3 3 1 3 ...
## $ Namjena.terena. : Factor w/ 227 levels "", "-", "Djelomično građevinsko", ...: 1 1 1 1 1 1 1 1 ...
## $ Površina.građevinskog.dijela. : Factor w/ 572 levels "", "- ", "/ ", "0,0 ", ...: 1 1 1 1 196 17 1 ...
## $ Površina.negrađevinskog.dijela. : Factor w/ 326 levels "", "- ", "/ ", "0,0 ", ...: 1 1 1 1 153 1 1 ...
## $ Instalacije.struje. : Factor w/ 26 levels "", "1988 god.", ...: 1 1 1 1 1 1 1 1 1 1 ...
## $ Instalacije.vode. : Factor w/ 26 levels "", "1988 god.", ...: 1 1 1 1 1 1 1 1 1 1 ...
## $ Pod..parket.laminat.pločice. : Factor w/ 25 levels "", "1998 god.", ...: 1 1 1 1 1 1 1 1 1 1 ...
## $ Pripojene.prostorije. : Factor w/ 70 levels "", "1", "2", "3", ...: 1 1 1 1 1 1 1 1 1 1 ...
## $ Fasada. : Factor w/ 28 levels "", "1989 god.", ...: 1 1 1 1 1 1 1 1 1 1 ...
## $ Krov. : Factor w/ 31 levels "", "1988 god.", ...: 1 1 1 1 1 1 1 1 1 1 ...
## $ Struja.zakupljena.snaga. : Factor w/ 17 levels "", "10 kWh", "13 kWh", ...: 1 1 1 1 1 1 1 1 1 1 ...
## $ Vanjska.stolarija. : Factor w/ 31 levels "", "1988 god.", ...: 1 1 1 1 1 1 1 1 1 1 ...
## $ Unutarnja.stolarija. : Factor w/ 26 levels "", "1988 god.", ...: 1 1 1 1 1 1 1 1 1 1 ...
## $ Kupaonica..kuhinja. : Factor w/ 26 levels "", "1998 god.", ...: 1 1 1 1 1 1 1 1 1 1 ...

```

```
unique( df$Vrsta.nekretnine.)
```

```

## [1] Poslovni prostor      Stan/Apartman          Kuća/Vikendica
## [4] Građevinsko zemljište  Negrađevinsko zemljište Industrijski objekti
## [7] Ugostiteljski objekti  Trgovački objekti      Hoteli i pansioni
## [10] Garaža
## 10 Levels: Garaža Građevinsko zemljište ... Ugostiteljski objekti

```

```

df <- df[df$Vrsta.nekretnine. == "Stan/Apartman",]
df$Vrt. <- NULL
df$Struja.zakupljena.snaga. <- NULL
df$Pripojene.prostorije. <- NULL
df$Vanjska.stolarija. <- NULL
df$Unutarnja.stolarija. <- NULL
df$Kupaonica.kuhinja. <- NULL
df$Instalacije.vode. <- NULL
df$Instalacije.struje. <- NULL
df$Pod.parket.laminat.pločice. <- NULL
df$Fasada. <- NULL
df$Krov. <- NULL
df$Površina.negrađevinskog.dijela. <- NULL
df$Površina.okućnice. <- NULL
df$Površina.građevinskog.dijela. <- NULL
df$Tlocrtna.površina.nekretnine. <- NULL
df$Vrsta.kuće.vikendice. <- NULL
df$Transakcija. <- NULL
df$Namjena.terena. <- NULL
df$Stil.gradnje. <- NULL
df$Mogućnost.zamjene. <- NULL
df$Dimnjak. <- NULL
df$Okućnica. <- NULL
df$Toaleti. <- NULL
df$Roštilj. <- NULL
df$Bazen. <- NULL
df$Nekretnina.se.prodaje.iznajmljuje. <- NULL
df$Nekretnina..stanje.. <- NULL
df$Natkriveni.parking. <- NULL
df$Razred.energetske.učinkovitosti. <- NULL
df$dodatne_oznake <- NULL
df$opis <- NULL
df$Pristup.invalidima. <- NULL
df$naziv <- NULL
df$Pogled.na. <- NULL
df$Pogled.na.more. <- NULL
df$Povlašteni.parking. <- NULL
df$Visina.plafona. <- NULL
df$Namjena.poslovnog.prostora. <- NULL
df$Lift. <- NULL
df$P.P.sa.izlogom. <- NULL
df$Pristup.kamionima.šleperima. <- NULL
df$Ključ.u.agenciji. <- NULL
df$Struja. <- NULL
df$Ukupno.katova. <- NULL
df$Parking. <- NULL
df$Parking..broj.. <- NULL
df$Garaža. <- NULL
df$Udaljenost.od.mora. <- NULL
df$Površina.balkona.lođe.terase. <- NULL
df$Vrsta.nekretnine. <- NULL

df$Površina. <- as.double(gsub(",", ".", df$Površina.))

```

```

df$Godina.izgradnje.[df$Godina.izgradnje. == ""] <- NA
df$Cijena. <- readr::parse_number(as.character(df$Cijena.) ,locale = readr::locale(decimal_mark = ","))

## Warning: 9 parsing failures.
## row col expected actual
## 63 -- a number na upit
## 75 -- a number na upit
## 76 -- a number na upit
## 193 -- a number na upit
## 202 -- a number na upit
## ... ..
## See problems(...) for more details.

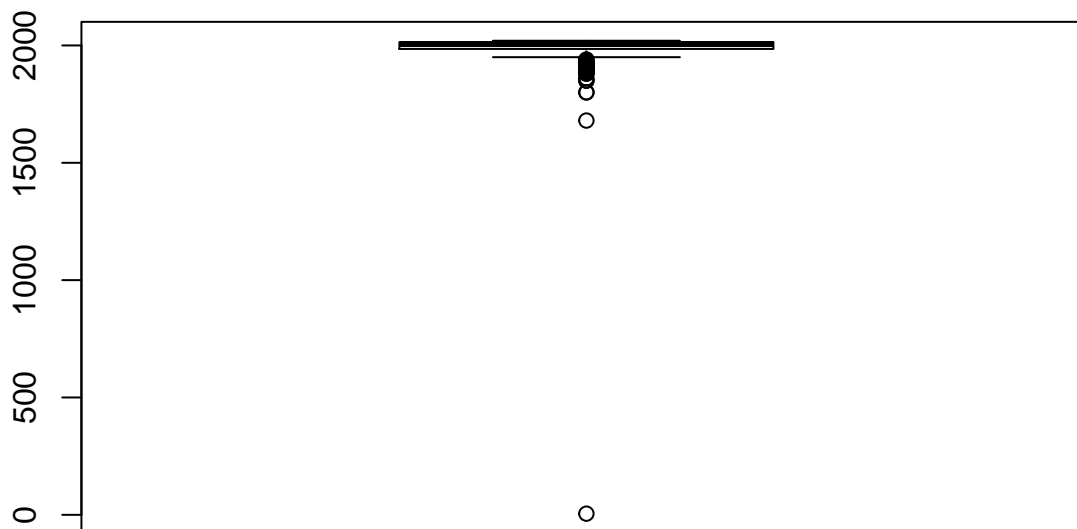
df$Godina.izgradnje. <- readr::parse_number(as.character(df$Godina.izgradnje.) ,locale = readr::locale(

## Warning: 1 parsing failure.
## row col expected actual
## 396 -- a number stara zgrada god.

df <- df[!is.na(df$Cijena.),]

boxplot(df$Godina.izgradnje.)

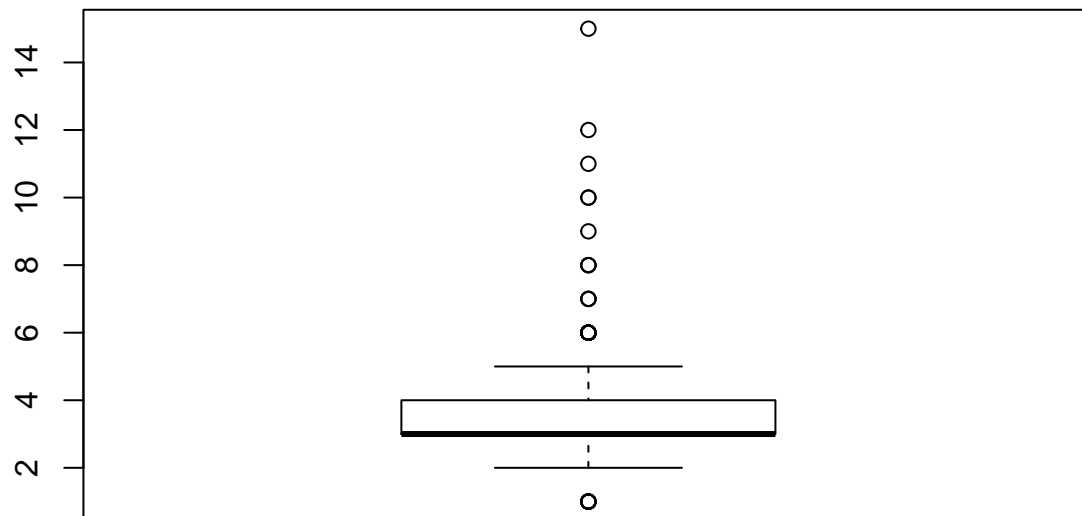
```



```

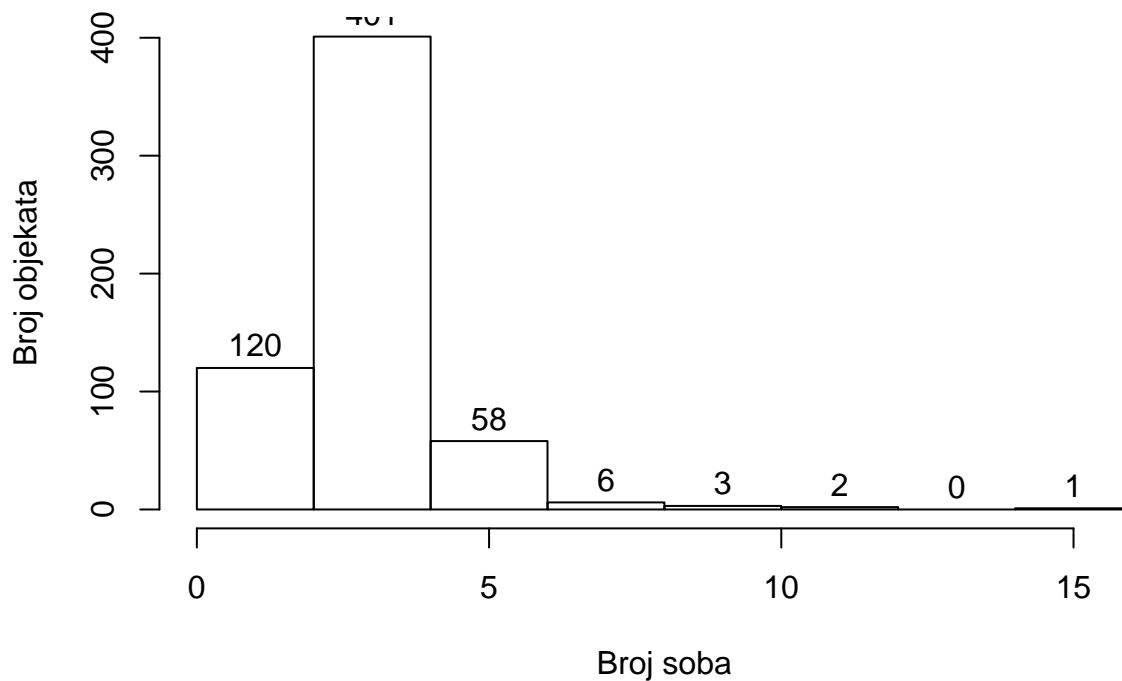
boxplot(df$Ukupan.broj.soba.)

```

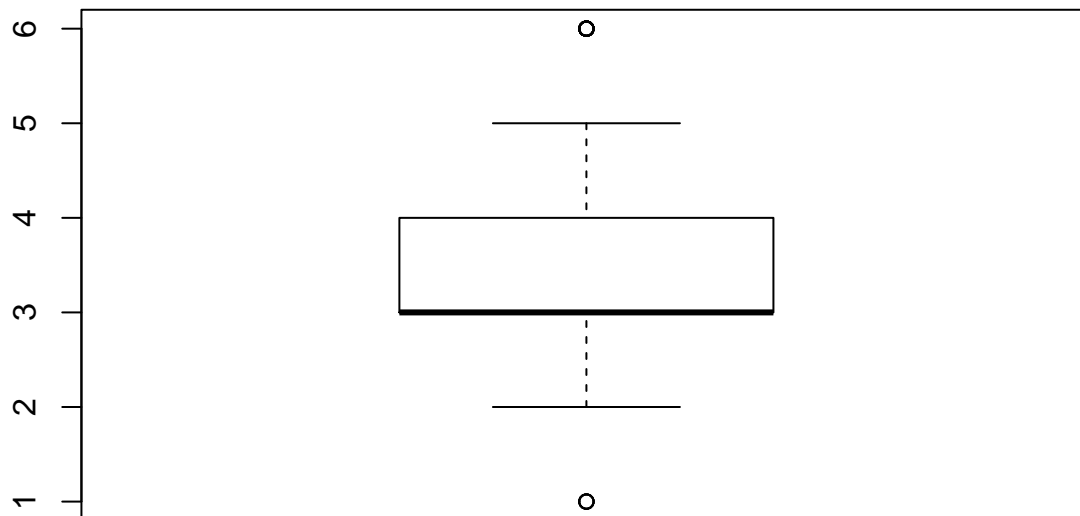


```
h0 <- hist(df$Ukupan.broj.soba., breaks = 6,
  labels = T,
  main = "Distribucija broja soba",
  xlab = "Broj soba",
  ylab = "Broj objekata"
)
```

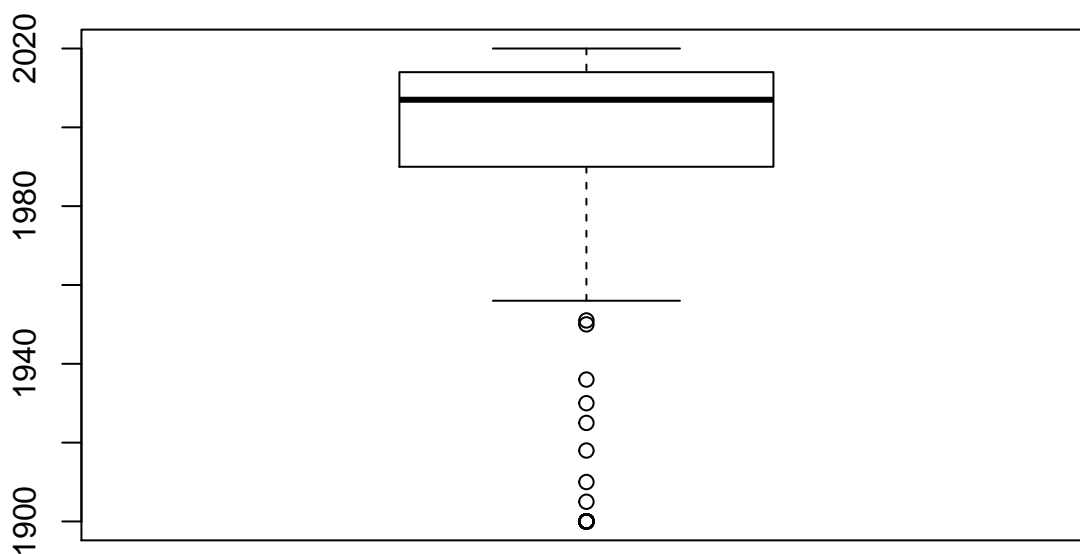
Distribucija broja soba



```
df <- df[!df$Ukupan.broj.soba. > 6,]
df <- df[!df$Cijena. < 20000,]
df <- na.omit(df)
df <- df[!df$Godina.izgradnje. < 1900,]
boxplot(df$Ukupan.broj.soba.)
```

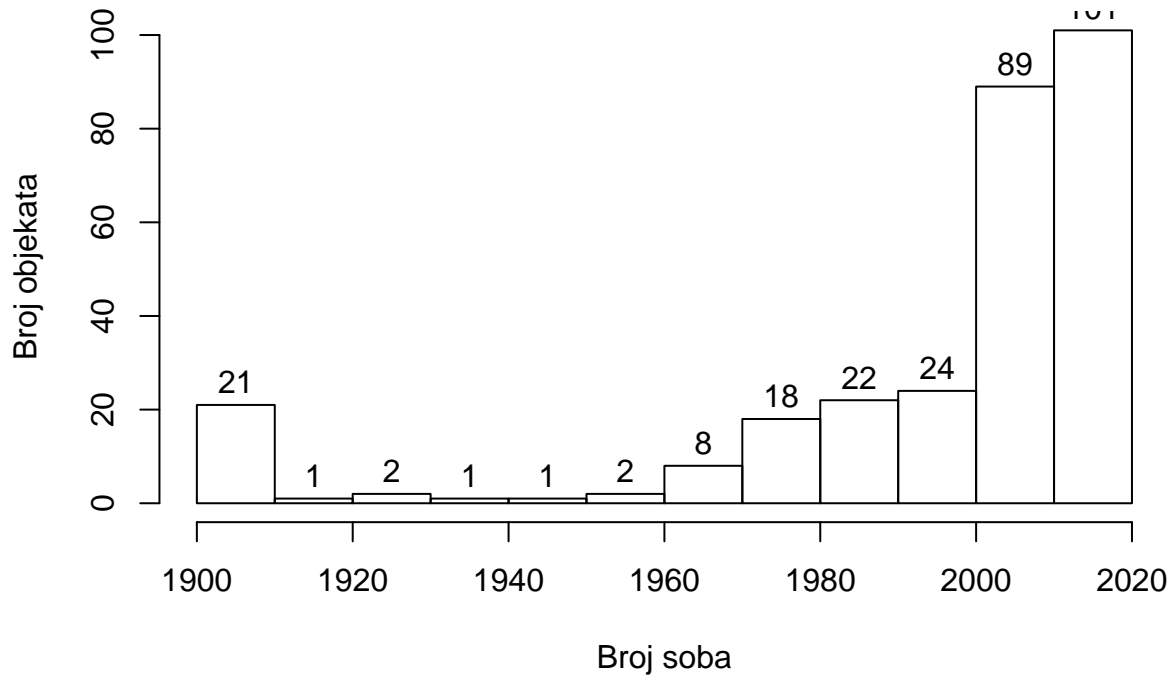



```
boxplot(df$Godina.izgradnje.)
```



```
h1 <- hist(df$Godina.izgradnje.,
  labels = T,
  main = "Distribucija broja soba",
  xlab = "Broj soba",
  ylab = "Broj objekata"
)
```

Distribucija broja soba



```
summary(df)
```

```
##
## http://www.remax-svijetnekretnina.com/apartman/akcijska-cijena-novogradnje-u-liznjanu
## http://www.remax-svijetnekretnina.com/apartman/dvoetazni-apartmansa-tri-terase-i-pogledom-na-more
## http://www.remax-svijetnekretnina.com/apartman/istra-banjole-apartman-s-pogledom-na-more1
## http://www.remax-svijetnekretnina.com/apartman/istra-barbariga-apartman-u-prizemlju-sa-vrtom
## http://www.remax-svijetnekretnina.com/apartman/istra-barbariga-dva-apartmana-sa-velikom-terasom
## http://www.remax-svijetnekretnina.com/apartman/istra-fazana-500m-od-mora-novi-stan-na-odlicnoj-loka
## (Other)
##   Lokacija.   Ukupan.broj.soba.   Cijena.   Površina.
## Poreč      : 40   Min.      :1.0      Min.    : 55000   Min.    : 25.0
## Medulin    : 27   1st Qu.:3.0      1st Qu.:108212   1st Qu.: 57.9
## Novigrad   : 25   Median   :3.0      Median   :138456   Median   : 75.0
## Umag       : 20   Mean     :3.3      Mean     :166520   Mean     : 81.7
## Rovinj     : 19   3rd Qu.:4.0      3rd Qu.:195000   3rd Qu.: 98.0
## Vrsar      : 14   Max.     :6.0      Max.     :766500   Max.     :224.5
## (Other)    :145
## Godina.izgradnje. Spavaće.sobe.   Kupaonice.
## Min.      :1900   Min.      :1.00   Min.     : 1.00
## 1st Qu.:1990   1st Qu.:2.00   1st Qu.: 1.00
## Median    :2007   Median    :2.00   Median    : 1.00
## Mean      :1995   Mean      :2.15   Mean      : 1.37
## 3rd Qu.:2014   3rd Qu.:3.00   3rd Qu.: 2.00
## Max.      :2020   Max.      :5.00   Max.      :11.00
##
```

```
write.csv(df, file = "cleanData.csv")
pairs(df)
```

```
## Warning in strwidth(labels, "user"): conversion failure on 'Površina.' in
## 'mbcsToSbcs': dot substituted for <c5>

## Warning in strwidth(labels, "user"): conversion failure on 'Površina.' in
## 'mbcsToSbcs': dot substituted for <a1>

## Warning in strwidth(labels, "user"): conversion failure on 'Spavaće.sobe.' in
## 'mbcsToSbcs': dot substituted for <c4>

## Warning in strwidth(labels, "user"): conversion failure on 'Spavaće.sobe.' in
## 'mbcsToSbcs': dot substituted for <87>

## Warning in text.default(x, y, txt, cex = cex, font = font): conversion failure
## on 'Površina.' in 'mbcsToSbcs': dot substituted for <c5>

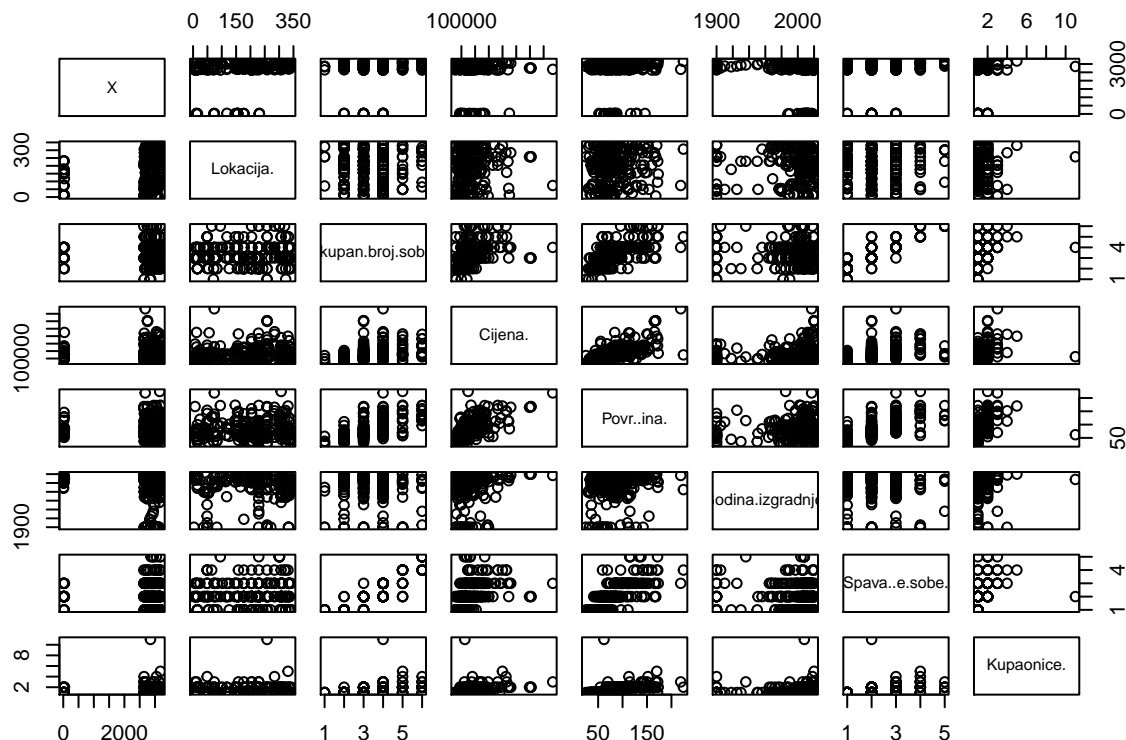
## Warning in text.default(x, y, txt, cex = cex, font = font): conversion failure
## on 'Površina.' in 'mbcsToSbcs': dot substituted for <a1>

## Warning in text.default(x, y, txt, cex = cex, font = font): font metrics unknown
## for Unicode character U+0161

## Warning in text.default(x, y, txt, cex = cex, font = font): conversion failure
## on 'Spavaće.sobe.' in 'mbcsToSbcs': dot substituted for <c4>

## Warning in text.default(x, y, txt, cex = cex, font = font): conversion failure
## on 'Spavaće.sobe.' in 'mbcsToSbcs': dot substituted for <87>

## Warning in text.default(x, y, txt, cex = cex, font = font): font metrics unknown
## for Unicode character U+0107
```



```
set.seed(666999)
idx <- sample(nrow(df), 0.7*nrow(df)) #70-30
trening <- df[idx,]
test <- df [-idx,]
```

```
corrgram(trening, order=TRUE, lower.panel=panel.shade,
         upper.panel=panel.pie, text.panel=panel.txt)
```

```
## Warning in strwidth(labels, "user"): conversion failure on 'Spavaće.sobe.' in
## 'mbcsToSbcs': dot substituted for <c4>

## Warning in strwidth(labels, "user"): conversion failure on 'Spavaće.sobe.' in
## 'mbcsToSbcs': dot substituted for <87>

## Warning in strwidth(labels, "user"): conversion failure on 'Površina.' in
## 'mbcsToSbcs': dot substituted for <c5>

## Warning in strwidth(labels, "user"): conversion failure on 'Površina.' in
## 'mbcsToSbcs': dot substituted for <a1>

## Warning in text.default(x, y, txt, cex = cex, font = font, srt = srt):
## conversion failure on 'Spavaće.sobe.' in 'mbcsToSbcs': dot substituted for <c4>

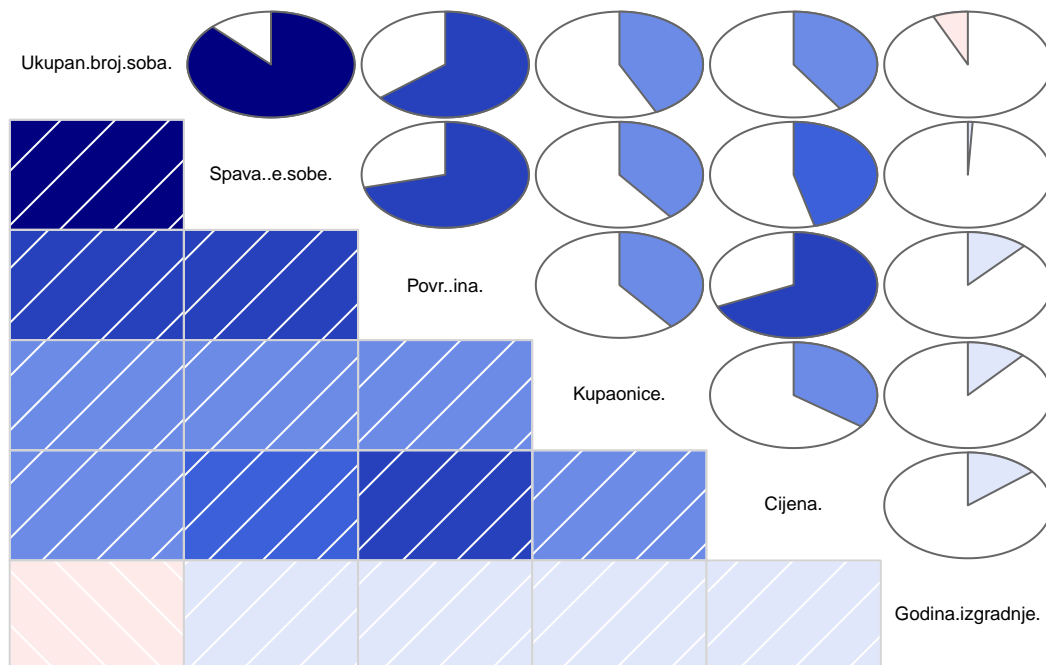
## Warning in text.default(x, y, txt, cex = cex, font = font, srt = srt):
## conversion failure on 'Spavaće.sobe.' in 'mbcsToSbcs': dot substituted for <87>

## Warning in text.default(x, y, txt, cex = cex, font = font, srt = srt): font
## metrics unknown for Unicode character U+0107

## Warning in text.default(x, y, txt, cex = cex, font = font, srt = srt):
## conversion failure on 'Površina.' in 'mbcsToSbcs': dot substituted for <c5>

## Warning in text.default(x, y, txt, cex = cex, font = font, srt = srt):
## conversion failure on 'Površina.' in 'mbcsToSbcs': dot substituted for <a1>

## Warning in text.default(x, y, txt, cex = cex, font = font, srt = srt): font
## metrics unknown for Unicode character U+0161
```



```
cor.test(trening$Cijena., trening$Površina.)
```

```
##
## Pearson's product-moment correlation
```

```
##
## data: trening$Cijena. and trening$Površina.
## t = 13, df = 201, p-value <2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.601 0.750
## sample estimates:
##    cor
## 0.682

model1 <- lm(Cijena.~Spavaće.sobe. + + Površina. , data = trening)
summary(model1)

##
## Call:
## lm(formula = Cijena. ~ Spavaće.sobe. + +Površina., data = trening)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -299536  -33360   -9036   16052  323192
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    10997     14457    0.76   0.45
## Spavaće.sobe.   -6088      8680   -0.70   0.48
## Površina.       2057       210    9.79 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 71700 on 200 degrees of freedom
## Multiple R-squared:  0.467, Adjusted R-squared:  0.462
## F-statistic: 87.6 on 2 and 200 DF, p-value: <2e-16

gvlma(model1)

##
## Call:
## lm(formula = Cijena. ~ Spavaće.sobe. + +Površina., data = trening)
##
## Coefficients:
## (Intercept)  Spavaće.sobe.    Površina.
##      10997      -6088       2057
##
##
## ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS
## USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:
## Level of Significance = 0.05
##
## Call:
## gvlma(x = model1)
##
##              Value p-value              Decision
## Global Stat    329.231 0.00e+00 Assumptions NOT satisfied!
## Skewness      49.846 1.66e-12 Assumptions NOT satisfied!
## Kurtosis      276.753 0.00e+00 Assumptions NOT satisfied!
```

```
## Link Function      2.123 1.45e-01    Assumptions acceptable.
## Heteroscedasticity 0.509 4.76e-01    Assumptions acceptable.
```

```
model2 <- lm(Cijena.~ Lokacija. + Površina., data = trening)
summary(model2)
```

```
##
## Call:
## lm(formula = Cijena. ~ Lokacija. + Površina., data = trening)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -275454  -21012        0   15950  280259
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   -32794      64621  -0.51    0.613
## Lokacija.Babići, Umag           -3528      89829  -0.04    0.969
## Lokacija.Banjole, Medulin       168765     73332   2.30    0.023 *
## Lokacija.Barbariga, Vodnjan      16778     68601   0.24    0.807
## Lokacija.Bibali, Buje           9540     90079   0.11    0.916
## Lokacija.Buje                  -16803     77855  -0.22    0.829
## Lokacija.Busoler, Pula          -31136     89836  -0.35    0.729
## Lokacija.Centar, Pula          -17466     69772  -0.25    0.803
## Lokacija.Červar-Porat, Poreč     37953     73436   0.52    0.606
## Lokacija.Fažana                 78746     70066   1.12    0.263
## Lokacija.Funtana, Vrsar         37322     73492   0.51    0.612
## Lokacija.Fuškuljin, Poreč       -47272     89906  -0.53    0.600
## Lokacija.Gregovica, Pula        -21188     89858  -0.24    0.814
## Lokacija.Ilirija, Pula         -15173     89936  -0.17    0.866
## Lokacija.Kaštanjer, Pula        -7115     89830  -0.08    0.937
## Lokacija.Kaštel, Buje          11578     77802   0.15    0.882
## Lokacija.Kmeti, Umag            782      73335   0.01    0.992
## Lokacija.Kukci, Poreč          -63424     78066  -0.81    0.418
## Lokacija.Ližnjan              20787     67434   0.31    0.758
## Lokacija.Lovrečica, Umag       58945     71008   0.83    0.408
## Lokacija.Medulin              27380     65163   0.42    0.675
## Lokacija.Mušalež, Poreč        -61209     89868  -0.68    0.497
## Lokacija.Novigrad              38451     65601   0.59    0.559
## Lokacija.Peroj, Vodnjan         16612     71046   0.23    0.815
## Lokacija.Pomer, Medulin       176828     77824   2.27    0.024 *
## Lokacija.Poreč                33644     64463   0.52    0.602
## Lokacija.Premantura, Medulin    34424     71094   0.48    0.629
## Lokacija.Radoši kod Žbandaja, Poreč 45931     89919   0.51    0.610
## Lokacija.Rovinj              123519     65358   1.89    0.061 .
## Lokacija.Savudrija, Umag      123432     89813   1.37    0.171
## Lokacija.Šijana, Pula           2010     89813   0.02    0.982
## Lokacija.Širi centar, Pula     -15219     89854  -0.17    0.866
## Lokacija.Šišan, Ližnjan        -601     89821  -0.01    0.995
## Lokacija.Štinjan, Pula         81170     90024   0.90    0.369
## Lokacija.Stoja, Pula           76638     77888   0.98    0.327
## Lokacija.Tar, Tar-Vabriga      -6173     73341  -0.08    0.933
## Lokacija.Umag                  5902     65545   0.09    0.928
## Lokacija.Vabriga, Tar-Vabriga   35292     73338   0.48    0.631
## Lokacija.Valdebek, Pula       -39013     73546  -0.53    0.597
```

```
## Lokacija.Veli vrh, Pula -4060 77781 -0.05 0.958
## Lokacija.Veruda, Pula -23689 71041 -0.33 0.739
## Lokacija.Verudela, Pula -40588 90442 -0.45 0.654
## Lokacija.Višnjan -12010 89813 -0.13 0.894
## Lokacija.Vižinada 90018 91020 0.99 0.324
## Lokacija.Vodnjan -2985 77830 -0.04 0.969
## Lokacija.Vrsar 62022 67373 0.92 0.359
## Lokacija.Vrvari, Poreč 20384 90061 0.23 0.821
## Lokacija.Zambratija, Umag 46540 90079 0.52 0.606
## Površina. 2010 157 12.79 <2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 63500 on 154 degrees of freedom
## Multiple R-squared: 0.678, Adjusted R-squared: 0.577
## F-statistic: 6.75 on 48 and 154 DF, p-value: <2e-16
```

```
gvlma(model2)
```

```
##
## Call:
## lm(formula = Cijena. ~ Lokacija. + Površina., data = trening)
##
## Coefficients:
## (Intercept) Lokacija.Babići, Umag
## -32794 -3528
## Lokacija.Banjole, Medulin Lokacija.Barbariga, Vodnjan
## 168765 16778
## Lokacija.Bibali, Buje Lokacija.Buje
## 9540 -16803
## Lokacija.Busoler, Pula Lokacija.Centar, Pula
## -31136 -17466
## Lokacija.Červar-Porat, Poreč Lokacija.Fažana
## 37953 78746
## Lokacija.Funtana, Vrsar Lokacija.Fuškulín, Poreč
## 37322 -47272
## Lokacija.Gregovica, Pula Lokacija.Ilirija, Pula
## -21188 -15173
## Lokacija.Kaštanjer, Pula Lokacija.Kaštel, Buje
## -7115 11578
## Lokacija.Kmeti, Umag Lokacija.Kukci, Poreč
## 782 -63424
## Lokacija.Ližnjan Lokacija.Lovrečica, Umag
## 20787 58945
## Lokacija.Medulin Lokacija.Mušalež, Poreč
## 27379 -61209
## Lokacija.Novigrad Lokacija.Peroj, Vodnjan
## 38451 16612
## Lokacija.Pomer, Medulin Lokacija.Poreč
## 176828 33644
## Lokacija.Premantura, Medulin Lokacija.Radoši kod Žbandaja, Poreč
## 34424 45931
## Lokacija.Rovinj Lokacija.Savudrija, Umag
## 123519 123432
## Lokacija.Šijana, Pula Lokacija.Širi centar, Pula
```

```
##              2010              -15219
##      Lokacija.Šišanj, Ližnjan      Lokacija.Štinjan, Pula
##              -601              81170
##      Lokacija.Stoja, Pula      Lokacija.Tar, Tar-Vabriga
##              76638              -6173
##      Lokacija.Umag      Lokacija.Vabriga, Tar-Vabriga
##              5902              35292
##      Lokacija.Valdebek, Pula      Lokacija.Veli vrh, Pula
##              -39013              -4060
##      Lokacija.Veruda, Pula      Lokacija.Verudela, Pula
##              -23689              -40588
##      Lokacija.Višnjan      Lokacija.Vižinada
##              -12010              90018
##      Lokacija.Vodnjan      Lokacija.Vrsar
##              -2985              62022
##      Lokacija.Vrvari, Poreč      Lokacija.Zambratija, Umag
##              20384              46540
##      Površina.
##      2010
```

```
## ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS
## USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:
## Level of Significance = 0.05
```

```
## Call:
```

```
## gvlma(x = model2)
```

```
##              Value      p-value      Decision
## Global Stat      456.48 0.0000000000 Assumptions NOT satisfied!
## Skewness          8.61 0.0033429880 Assumptions NOT satisfied!
## Kurtosis          414.66 0.0000000000 Assumptions NOT satisfied!
## Link Function      30.86 0.0000000277 Assumptions NOT satisfied!
## Heteroscedasticity 2.35 0.1253360687 Assumptions acceptable.
```

```
model3 <- lm(Cijena.~Površina. , data = trening)
summary(model3)
```

```
##
```

```
## Call:
```

```
## lm(formula = Cijena. ~ Površina., data = trening)
```

```
##
```

```
## Residuals:
```

```
##      Min      1Q  Median      3Q      Max
## -295959 -34527 -10126  15715  332281
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      6582      12999   0.51    0.61
## Površina.        1953        148  13.24 <2e-16 ***
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## Residual standard error: 71600 on 201 degrees of freedom
```

```
## Multiple R-squared:  0.466, Adjusted R-squared:  0.463
```

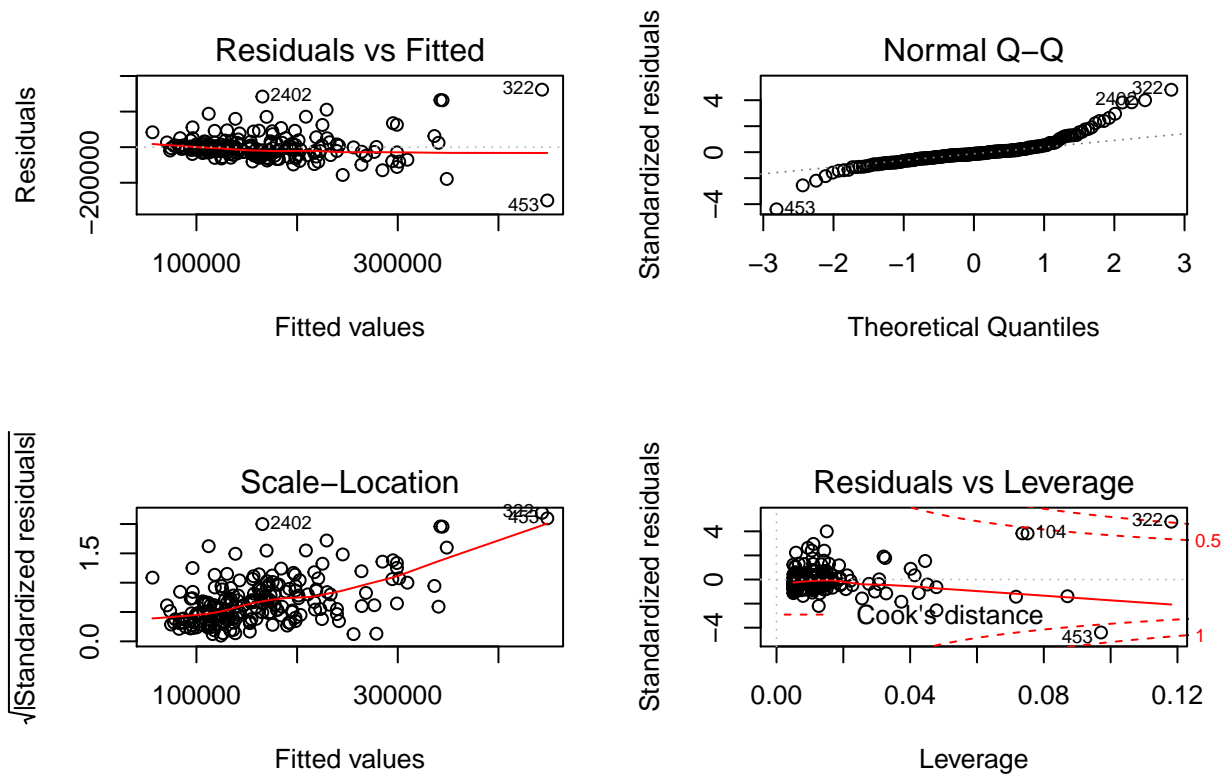


```
## F-statistic: 175 on 1 and 201 DF, p-value: <2e-16
gvlma(model3)

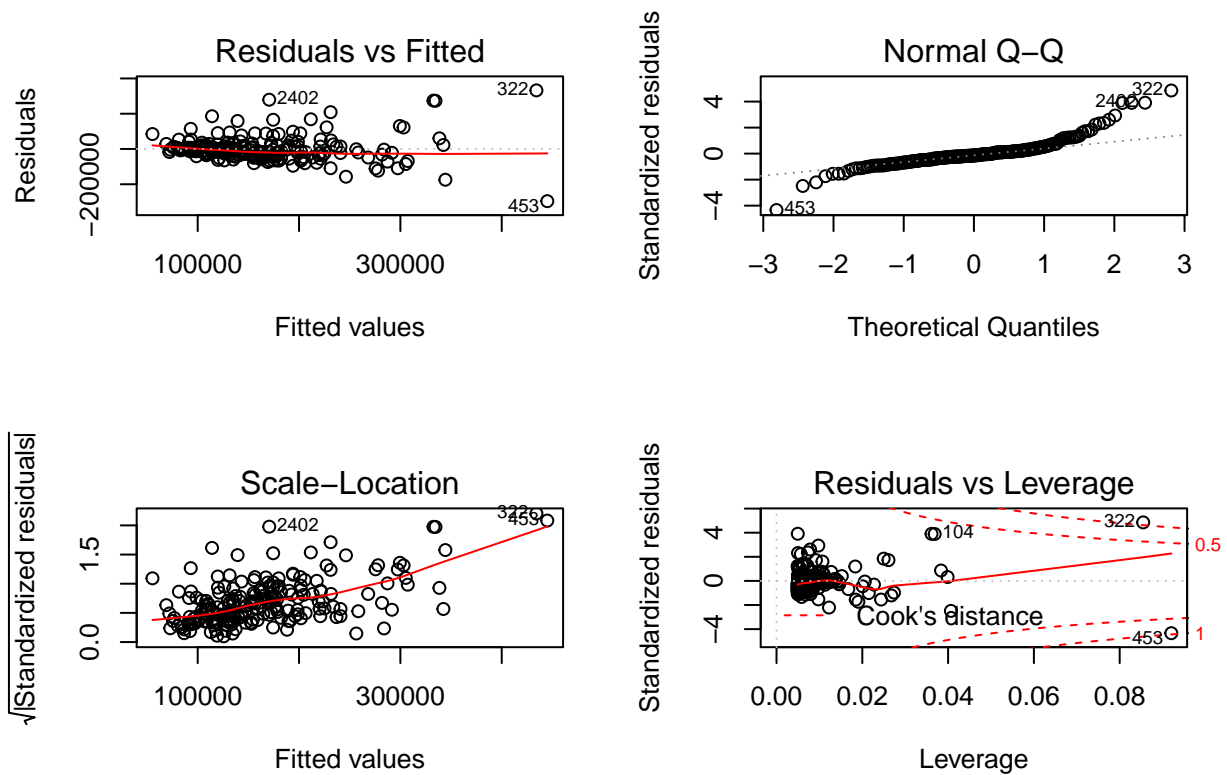
##
## Call:
## lm(formula = Cijena. ~ Površina., data = trening)
##
## Coefficients:
## (Intercept)    Površina.
##          6582          1953
##
##
## ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS
## USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:
## Level of Significance = 0.05
##
## Call:
## gvlma(x = model3)
##
##              Value p-value              Decision
## Global Stat      358.66 0.00e+00 Assumptions NOT satisfied!
## Skewness         54.82 1.32e-13 Assumptions NOT satisfied!
## Kurtosis         300.87 0.00e+00 Assumptions NOT satisfied!
## Link Function      1.89 1.70e-01 Assumptions acceptable.
## Heteroscedasticity 1.08 2.98e-01 Assumptions acceptable.

pred1 <- predict(model1, newdata = test)
#pred2 <- predict(model2, newdata = test)
pred3 <- predict(model3, newdata = test)

par(mfrow = c(2,2)) # 2x2 plot
plot(model1)
```



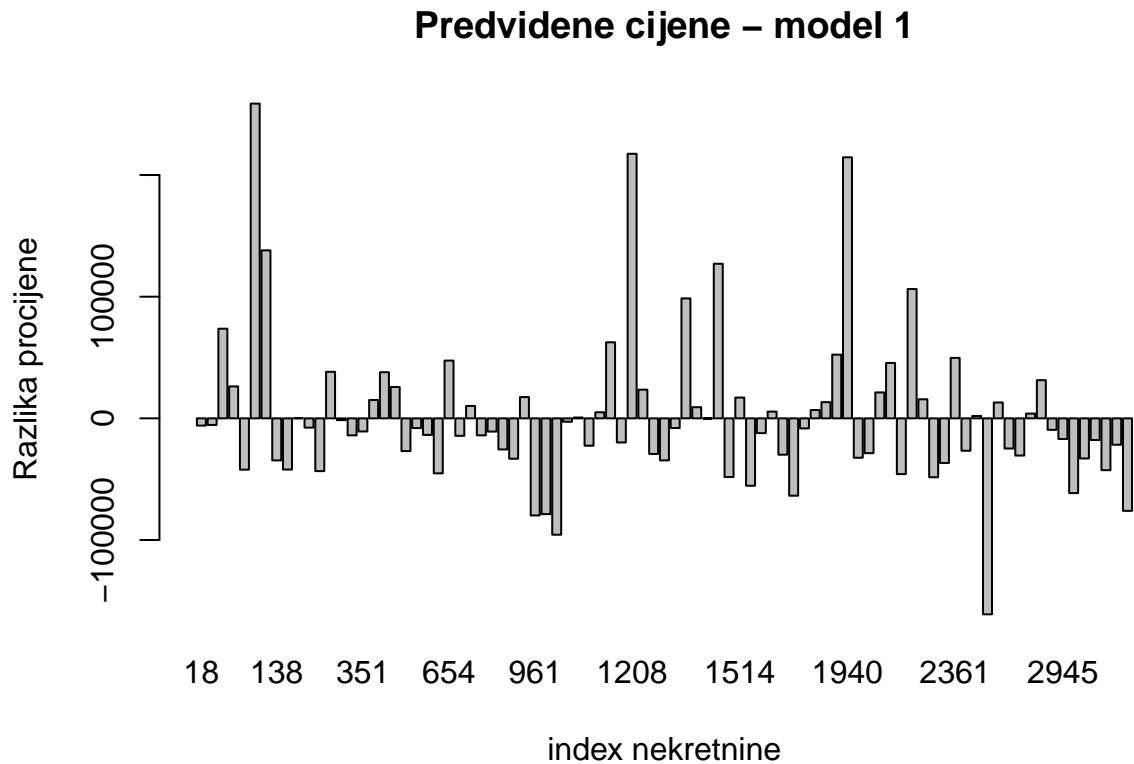
```
plot(model13)
```



```
par(mfrow=c(1,1))
```

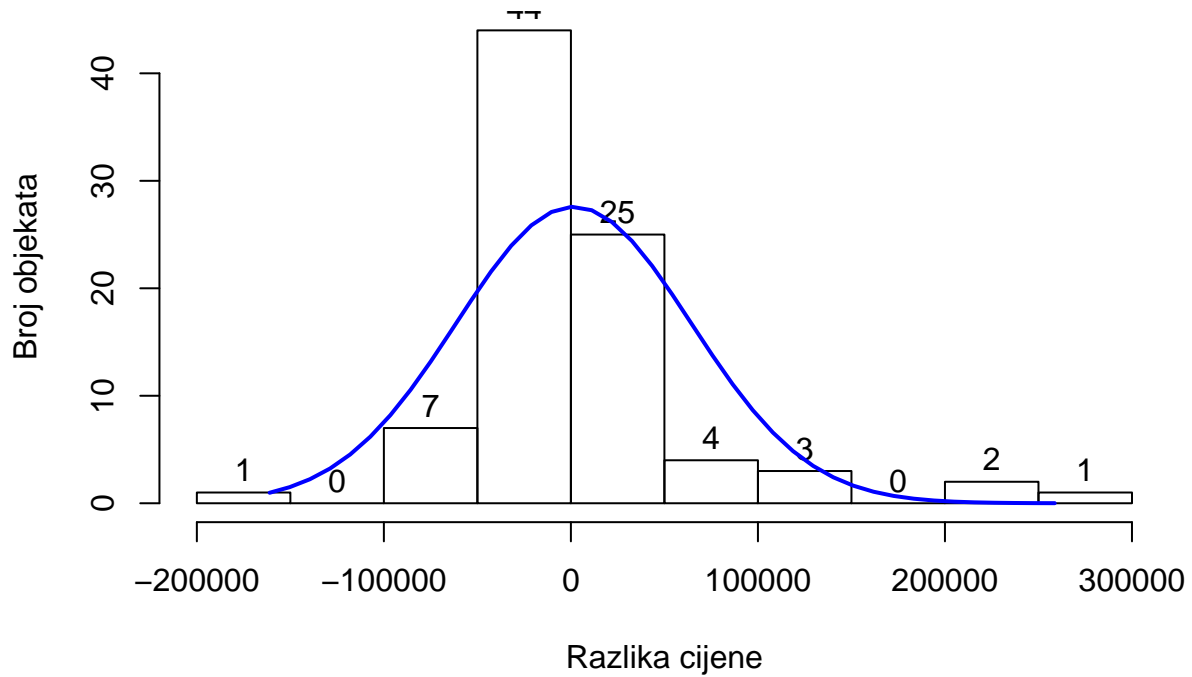
```
t1 <- test$Cijena. - pred1
```

```
barplot(t1,
  main = "Predvidene cijene - model 1",
  xlab = "index nekretnine",
  ylab = "Razlika procijene")
```



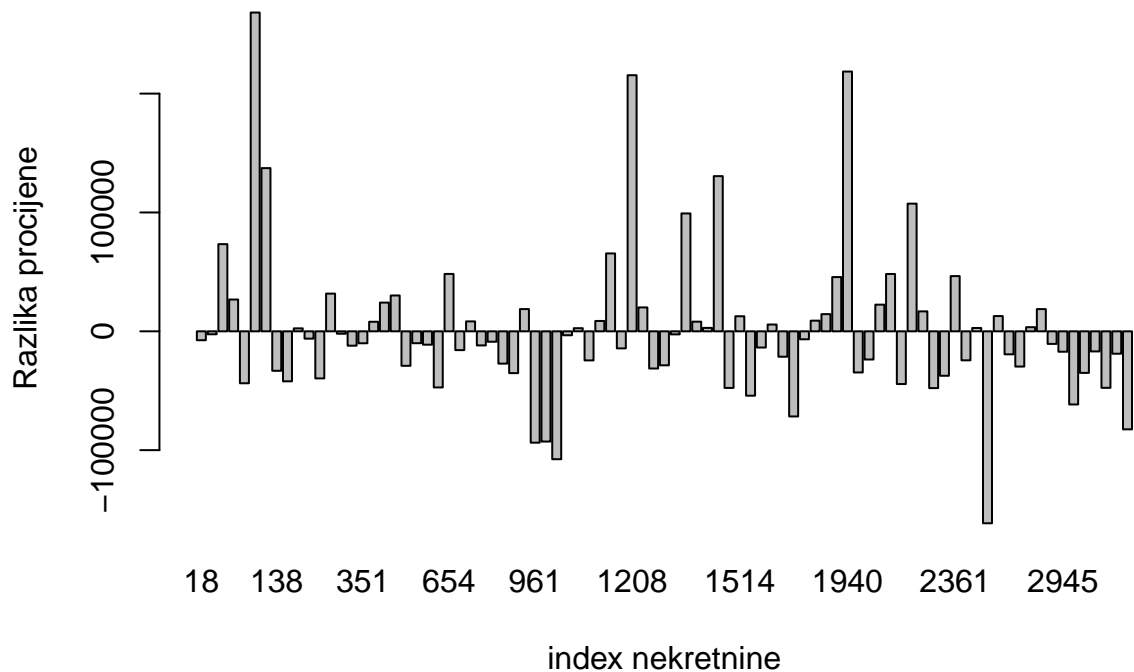
```
h_t1 <- hist(t1,breaks = 10,
  labels = T,
  main = "Distribucije procijene cijena model_1",
  xlab = "Razlika cijene",
  ylab = "Broj objekata"
)
xfit<-seq(min(t1),max(t1),length=40)
yfit<-dnorm(xfit,mean=mean(t1),sd=sd(t1))
yfit <- yfit*diff(h_t1$mids[1:2])*length(t1)
lines(xfit, yfit, col="blue", lwd=2)
```

Distribucije procijene cijena model_1



```
t3 <- test$Cijena. - pred3
barplot(t3,
  main = "Predvidene cijene - model 3",
  xlab = "index nekretnine",
  ylab = "Razlika procijene")
```

Predvidene cijene – model 3

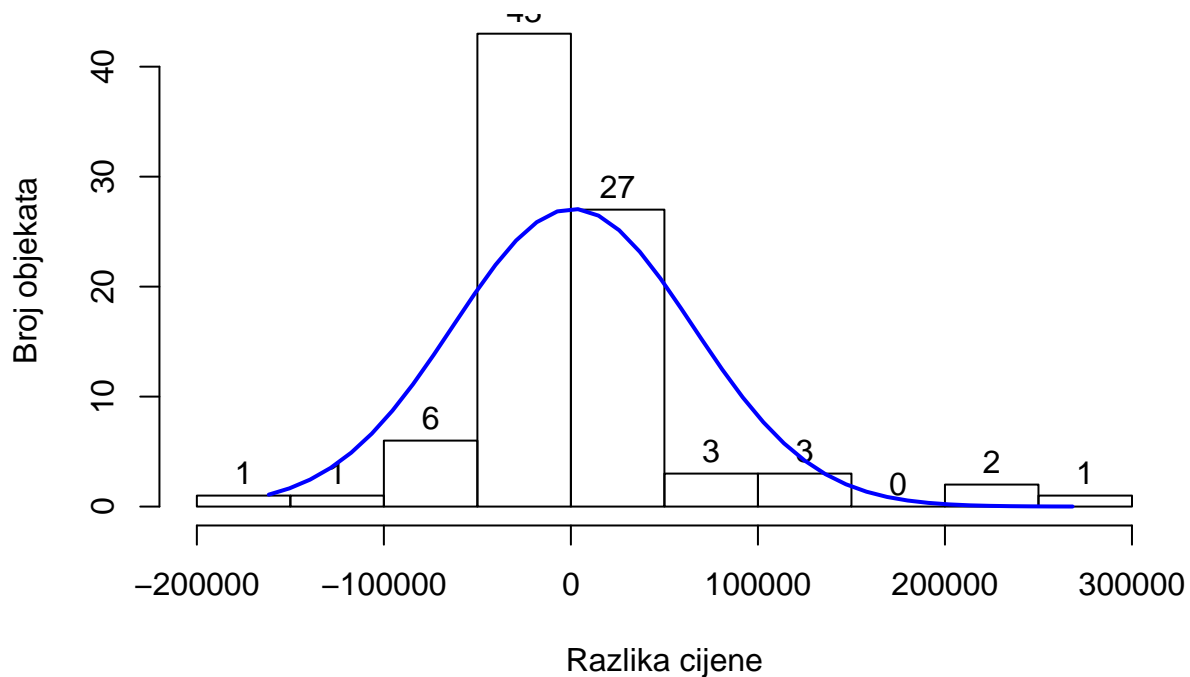


```

h_t3 <- hist(t3,breaks = 10,
             labels = T,
             main = "Distribucije procijene cijena- model 3",
             xlab = "Razlika cijene",
             ylab = "Broj objekata"
)
xfit<-seq(min(t3),max(t3),length=40)
yfit<-dnorm(xfit,mean=mean(t3),sd=sd(t3))
yfit <- yfit*diff(h_t3$mids[1:2])*length(t3)
lines(xfit, yfit, col="blue", lwd=2)

```

Distribucije procijene cijena- model 3



Zaključak i diskusija

Korišteni model ne može dobro predvidjeti cijene nekretnina iz zadanog dataseta. Trebalo bi riješiti poznati problem utjecaja lokacije na cijenu nekretnine. Nadalje, vidi se da model ipak podjednako precijeni kao i podcijeni nekretnine iz testnog skupa podataka s blagim repom prema podcijenjenim