

Homework DA

2024-06-01

Compito di Data Analytics

Gruppo C: ZANNI, JIMÉNEZ LIRIANO, RICCIO, SMOKOVIĆ

Importazione dei dati

```
BankChurners = read.csv("credit-card-customers/BankChurners.csv", stringsAsFactors = T)
HousePrices = read.csv("house-prices/train.csv", stringsAsFactors = T)
```

ANALISI DEL DATASET “BankChurners”

```
str(BankChurners)
```

```
## 'data.frame': 10127 obs. of 23 variables:
## $ CLIENTNUM
## $ Attrition_Flag
## $ Customer_Age
## $ Gender
## $ Dependent_count
## $ Education_Level
## $ Marital_Status
## $ Income_Category
## $ Card_Category
## $ Months_on_book
## $ Total_Relationship_Count
## $ Months_Inactive_12_mon
## $ Contacts_Count_12_mon
## $ Credit_Limit
## $ Total_Revolving_Bal
## $ Avg_Open_To_Buy
## $ Total_Amt_Chng_Q4_Q1
## $ Total_Trans_Amt
## $ Total_Trans_Ct
## $ Total_Ct_Chng_Q4_Q1
## $ Avg_Utilization_Ratio
## $ Naive_Bayes_Classifier_Attrition_Flag_Card_Category_Contacts_Count_12_mon_Dependent_count_Education_Level_Months_Inactive_12_mon_1
## $ Naive_Bayes_Classifier_Attrition_Flag_Card_Category_Contacts_Count_12_mon_Dependent_count_Education_Level_Months_Inactive_12_mon_2
```

```
head(BankChurners)
```

```
## CLIENTNUM Attrition_Flag Customer_Age Gender Dependent_count
## 1 768805383 Existing Customer 45 M 3
## 2 818770008 Existing Customer 49 F 5
## 3 713982108 Existing Customer 51 M 3
## 4 769911858 Existing Customer 40 F 4
```

```

## 5 709106358 Existing Customer          40      M          3
## 6 713061558 Existing Customer          44      M          2
##   Education_Level Marital_Status Income_Category Card_Category Months_on_book
## 1      High School      Married    $60K - $80K          Blue          39
## 2      Graduate      Single    Less than $40K          Blue          44
## 3      Graduate      Married    $80K - $120K          Blue          36
## 4      High School      Unknown    Less than $40K          Blue          34
## 5      Uneducated      Married    $60K - $80K          Blue          21
## 6      Graduate      Married    $40K - $60K          Blue          36
##   Total_Relationship_Count Months_Inactive_12_mon Contacts_Count_12_mon
## 1              5              1              3
## 2              6              1              2
## 3              4              1              0
## 4              3              4              1
## 5              5              1              0
## 6              3              1              2
##   Credit_Limit Total_Revolving_Bal Avg_Open_To_Buy Total_Amt_Chng_Q4_Q1
## 1      12691          777          11914          1.335
## 2      8256          864          7392          1.541
## 3      3418           0          3418          2.594
## 4      3313         2517           796          1.405
## 5      4716           0          4716          2.175
## 6      4010         1247          2763          1.376
##   Total_Trans_Amt Total_Trans_Ct Total_Ct_Chng_Q4_Q1 Avg_Utilization_Ratio
## 1          1144          42          1.625          0.061
## 2          1291          33          3.714          0.105
## 3          1887          20          2.333          0.000
## 4          1171          20          2.333          0.760
## 5           816          28          2.500          0.000
## 6          1088          24          0.846          0.311
##   Naive_Bayes_Classifier_Attrition_Flag_Card_Category_Contacts_Count_12_mon_Dependent_count_Education_Level_Inferred_Socioeconomic_Status
## 1
## 2
## 3
## 4
## 5
## 6
##   Naive_Bayes_Classifier_Attrition_Flag_Card_Category_Contacts_Count_12_mon_Dependent_count_Education_Level_Inferred_Socioeconomic_Status
## 1
## 2
## 3
## 4
## 5
## 6

```

```
summary(BankChurners)
```

```

##      CLIENTNUM          Attrition_Flag Customer_Age      Gender
## Min.   :708082083   Attrited Customer:1627   Min.   :26.00   F:5358
## 1st Qu.:713036770   Existing Customer:8500   1st Qu.:41.00   M:4769
## Median :717926358
## Mean   :739177606
## 3rd Qu.:773143533
## Max.   :828343083
##

```

```

## Dependent_count      Education_Level  Marital_Status      Income_Category
## Min.      :0.000    College      :1013    Divorced: 748    $120K +      : 727
## 1st Qu.:1.000    Doctorate   : 451    Married :4687    $40K - $60K   :1790
## Median :2.000    Graduate    :3128    Single  :3943    $60K - $80K   :1402
## Mean    :2.346    High School :2013    Unknown : 749    $80K - $120K  :1535
## 3rd Qu.:3.000    Post-Graduate: 516                      Less than $40K:3561
## Max.    :5.000    Uneducated  :1487                      Unknown        :1112
##                               Unknown      :1519
## Card_Category  Months_on_book  Total_Relationship_Count
## Blue      :9436    Min.      :13.00    Min.      :1.000
## Gold       : 116    1st Qu.:31.00    1st Qu.:3.000
## Platinum:  20    Median :36.00    Median :4.000
## Silver    : 555    Mean     :35.93    Mean     :3.813
##                               3rd Qu.:40.00    3rd Qu.:5.000
##                               Max.      :56.00    Max.      :6.000
##
## Months_Inactive_12_mon  Contacts_Count_12_mon  Credit_Limit
## Min.      :0.000          Min.      :0.000          Min.      : 1438
## 1st Qu.:2.000          1st Qu.:2.000          1st Qu.: 2555
## Median :2.000          Median :2.000          Median : 4549
## Mean     :2.341          Mean     :2.455          Mean     : 8632
## 3rd Qu.:3.000          3rd Qu.:3.000          3rd Qu.:11068
## Max.     :6.000          Max.     :6.000          Max.     :34516
##
## Total_Revolving_Bal  Avg_Open_To_Buy  Total_Amt_Chng_Q4_Q1  Total_Trans_Amt
## Min.      :  0          Min.      :  3          Min.      :0.0000          Min.      :  510
## 1st Qu.: 359          1st Qu.: 1324          1st Qu.:0.6310          1st Qu.: 2156
## Median :1276          Median : 3474          Median :0.7360          Median : 3899
## Mean     :1163          Mean     : 7469          Mean     :0.7599          Mean     : 4404
## 3rd Qu.:1784          3rd Qu.: 9859          3rd Qu.:0.8590          3rd Qu.: 4741
## Max.     :2517          Max.     :34516          Max.     :3.3970          Max.     :18484
##
## Total_Trans_Ct      Total_Ct_Chng_Q4_Q1  Avg_Utilization_Ratio
## Min.      : 10.00    Min.      :0.0000          Min.      :0.0000
## 1st Qu.: 45.00    1st Qu.:0.5820          1st Qu.:0.0230
## Median : 67.00    Median :0.7020          Median :0.1760
## Mean     : 64.86    Mean     :0.7122          Mean     :0.2749
## 3rd Qu.: 81.00    3rd Qu.:0.8180          3rd Qu.:0.5030
## Max.     :139.00    Max.     :3.7140          Max.     :0.9990
##
## Naive_Bayes_Classifier_Attrition_Flag_Card_Category_Contacts_Count_12_mon_Dependent_count_Education
## Min.      :0.0000077
## 1st Qu.:0.0000990
## Median :0.0001815
## Mean     :0.1599975
## 3rd Qu.:0.0003373
## Max.     :0.9995800
##
## Naive_Bayes_Classifier_Attrition_Flag_Card_Category_Contacts_Count_12_mon_Dependent_count_Education
## Min.      :0.00042
## 1st Qu.:0.99966
## Median :0.99982
## Mean     :0.84000
## 3rd Qu.:0.99990

```

```
## Max. :0.99999
##
```

```
sum(is.na(BankChurners))
```

```
## [1] 0
```

PER LE VARIABILI DI TIPO QUANTITATIVE(integer/numeric)

```
# Customer_age
```

```
eta_media <- mean(BankChurners$Customer_Age) # eta media dei clienti
```

```
round(eta_media, digits = 1) # arrotondando a una cifra decimale
```

```
## [1] 46.3
```

```
sd(BankChurners$Customer_Age) # deviazione standard
```

```
## [1] 8.016814
```

```
quantile(BankChurners$Customer_Age) # quantili
```

```
## 0% 25% 50% 75% 100%
```

```
## 26 41 46 52 73
```

```
# notiamo che l'eta dei clienti sta nell'intervallo dai 26 ai 73 anni
```

```
# con 26 anni il cliente piu giovane e 73 il piu anziano
```

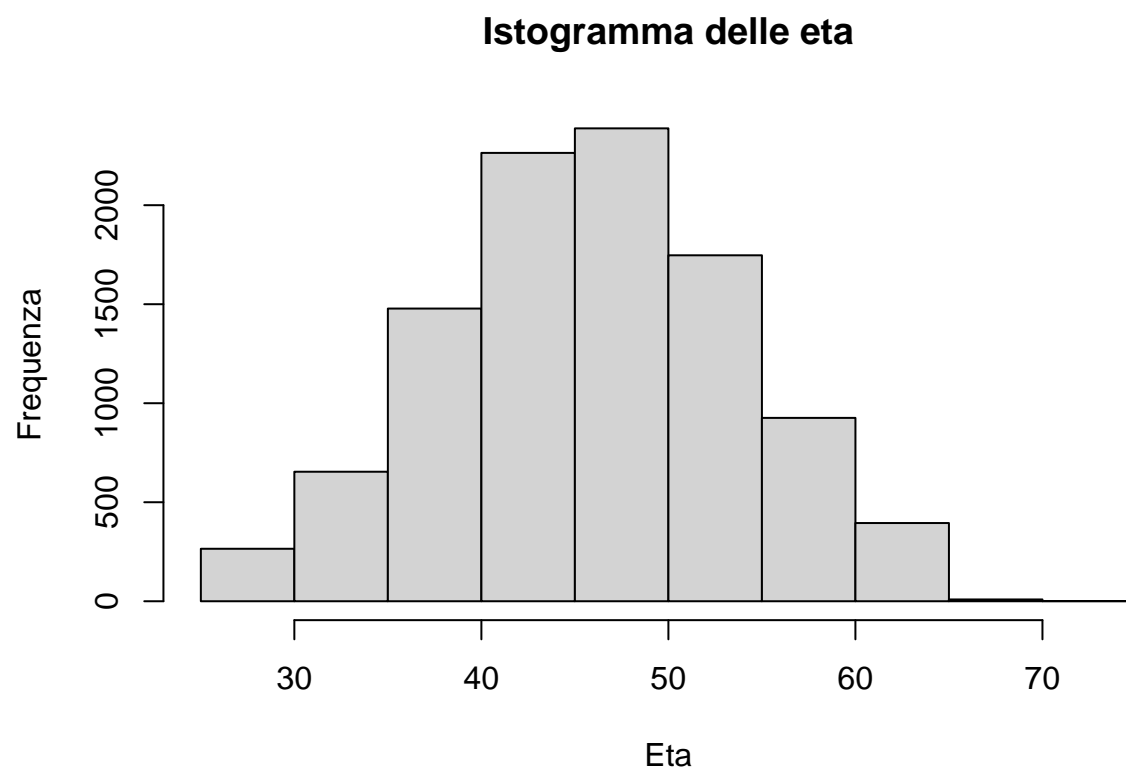
```
summary(BankChurners$Customer_Age) # e lo possiamo notare anche in questo modo
```

```
## Min. 1st Qu. Median Mean 3rd Qu. Max.
```

```
## 26.00 41.00 46.00 46.33 52.00 73.00
```

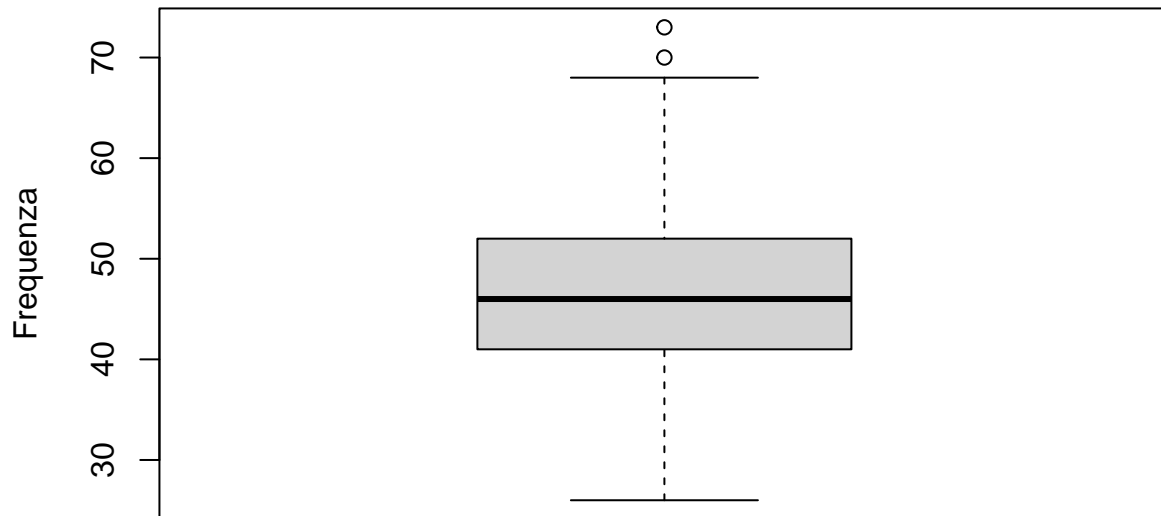
```
# visualizzazione grafica
```

```
hist(BankChurners$Customer_Age, main = "Istogramma delle eta", xlab = "Eta", ylab = "Frequenza")
```



```
boxplot(BankChurners$Customer_Age, main = "Boxplot delle eta", ylab = "Frequenza")
```

Boxplot delle eta



```
# Dependent_count
summary(BankChurners$Dependent_count) #abbiamo da 0 a 5 dipendenti assegnati

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.000   1.000   2.000   2.346   3.000   5.000

round(mean(BankChurners$Dependent_count), digits = 1) #media

## [1] 2.3

median(BankChurners$Dependent_count) #mediana

## [1] 2

sd(BankChurners$Dependent_count) #deviazione standard

## [1] 1.298908

#visualizzazione grafica
hist(BankChurners$Dependent_count, main="Istogramma di Dependent Count", xlab="Numero di Dipendenti",
ylab="Frequenza", breaks=8, col="lightblue")

# notiamo che la maggioranza dei clienti possiede dai 2 ai 3 dipendenti associati

# controllando i numeri:
table(BankChurners$Dependent_count)

##
##      0      1      2      3      4      5
```

```
## 904 1838 2655 2732 1574 424
```

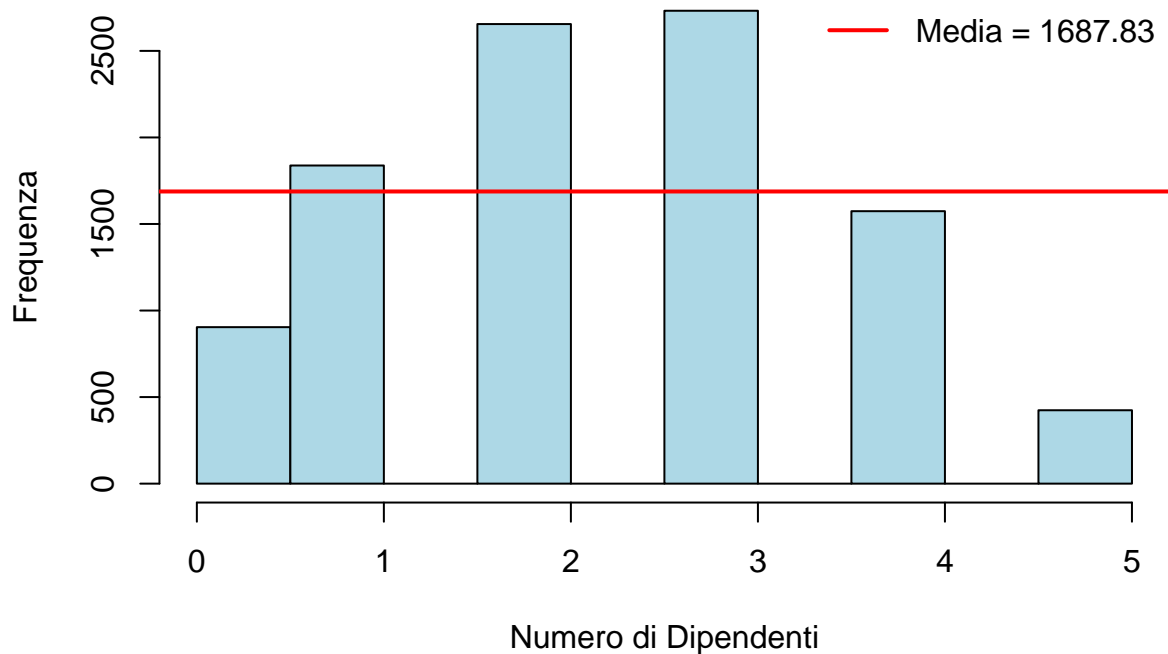
```
# aggiungendo la retta della media delle frequenze:
```

```
media_dip <- mean(table(BankChurners$Dependent_count))
```

```
abline(h = media_dip, col = "red", lwd = 2)
```

```
legend("topright", legend = paste("Media =", round(media_dip, 2)),  
      col = "red", lwd = 2, bty = "n")
```

Istogramma di Dependent Count



```
# Credit_Limit
```

```
summary(BankChurners$Credit_Limit) # limite del credito che va da 1.4k a 34k
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
```

```
##      1438   2555   4549   8632  11068  34516
```

```
media_credito <- round(mean(BankChurners$Credit_Limit), digits = 0) # credito limite medio
```

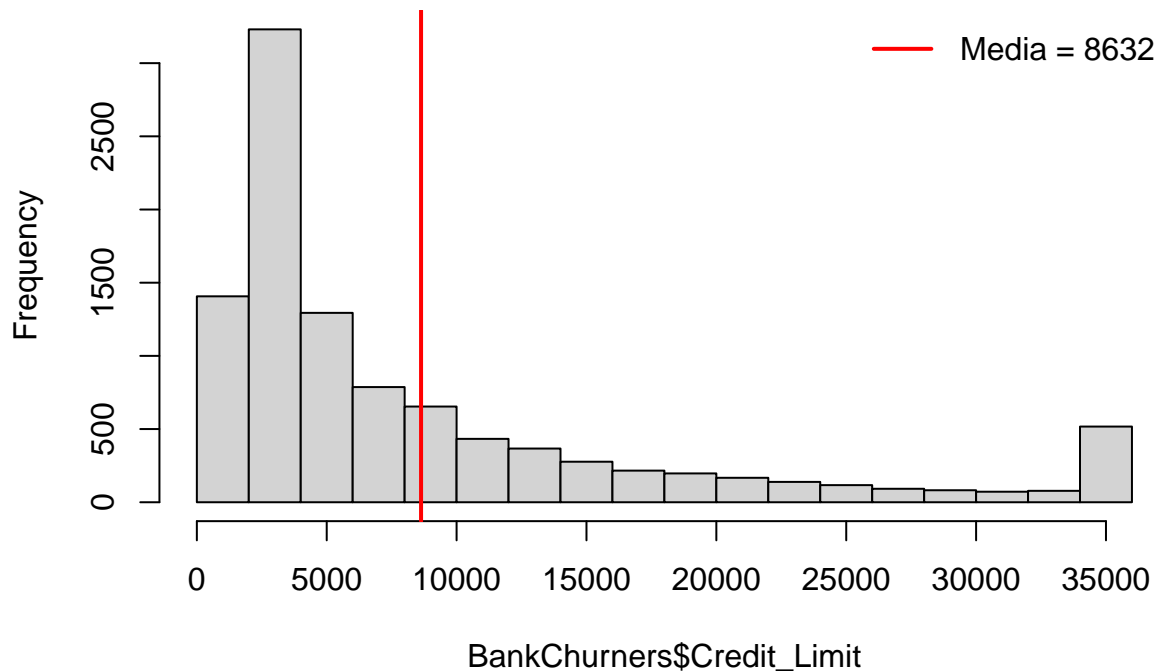
```
# visualizzazione grafica:
```

```
hist(BankChurners$Credit_Limit)
```

```
abline(v = media_credito, col = "red", lwd = 2)
```

```
legend("topright", legend = paste("Media =", media_credito),  
      col = "red", lwd = 2, bty = "n")
```

Histogram of BankChurners\$Credit_Limit



*# notiamo che il grafico e' piu' popolato ovvero ci sono piu' clienti al di sotto della media
sono piu' numerosi i clienti con credito limite massimo (34k) che i clienti nei vari
intervalli tra 10k e 33k*

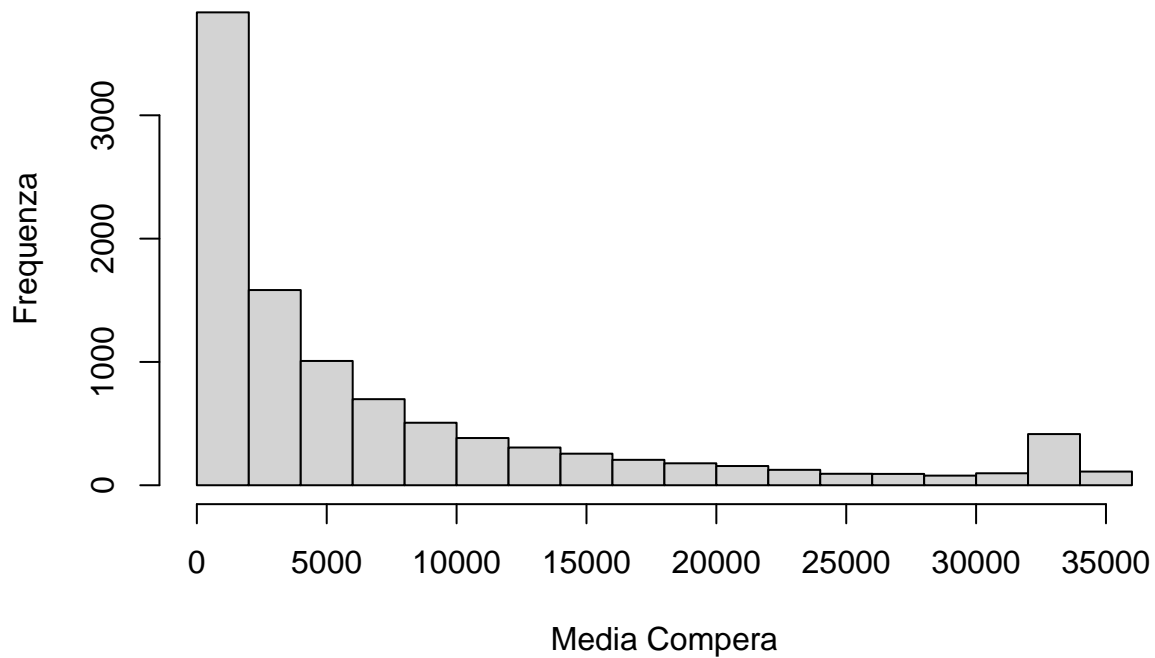
Avg_Open_To_Buy

```
summary(BankChurners$Avg_Open_To_Buy)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##         3   1324    3474    7469   9859   34516
```

```
hist(BankChurners$Avg_Open_To_Buy, main = "Istogramma", xlab = "Media Compera", ylab = "Frequenza")
```


Istogramma



*# notiamo che i valori massimi sono esattamente il credito limite dei clienti
 # il grafico segue un andamento esponenziale verso il basso
 # e' molto simile al grafico visto in precedenza riguardante il credito limite
 # quindi da questi grafici si nota come praticamente tutti i clienti raggiungano
 # o quasi il credito limite*

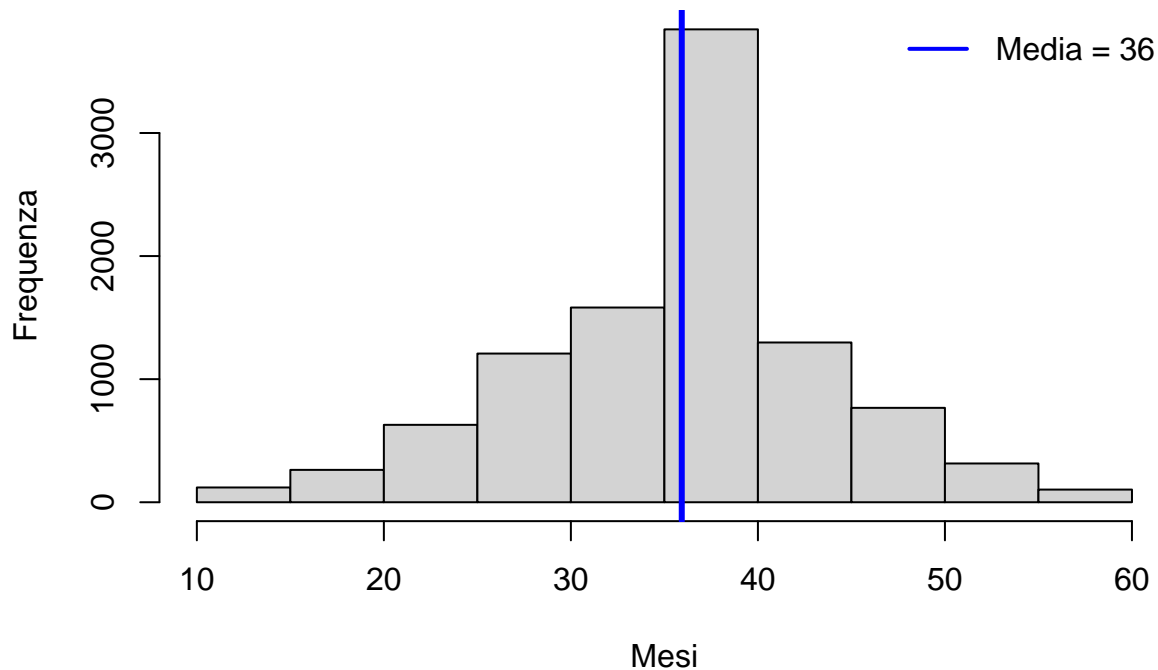
```
#Month_On_Book
summary(BankChurners$Months_on_book)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    13.00  31.00   36.00   35.93  40.00   56.00
```

*#il tempo minimo trascorso con la banca e' di 13 mesi invece il
 # maggior tempo trascorso e' di 56 mesi
 # vedendolo graficamente:*

```
hist(BankChurners$Months_on_book, main = "Istogramma mesi trascorsi",
     xlab = "Mesi", ylab = "Frequenza")
media_mesi <- mean(BankChurners$Months_on_book)
abline(v= media_mesi, col = "blue", lwd = 3)
legend("topright", legend = paste("Media =", round(media_mesi, digits = 0)),
     col = "blue", lwd = 2, bty = "n")
```

Istogramma mesi trascorsi



```
# osserviamo che poche persone restano poco o tanto
# la maggior parte dei clienti resta con la banca tra i 35 e i 40 mesi

#Total_Trans_Amt
summary(BankChurners$Total_Trans_Amt) # transazioni minime e massime negli ultimi 12 mesi

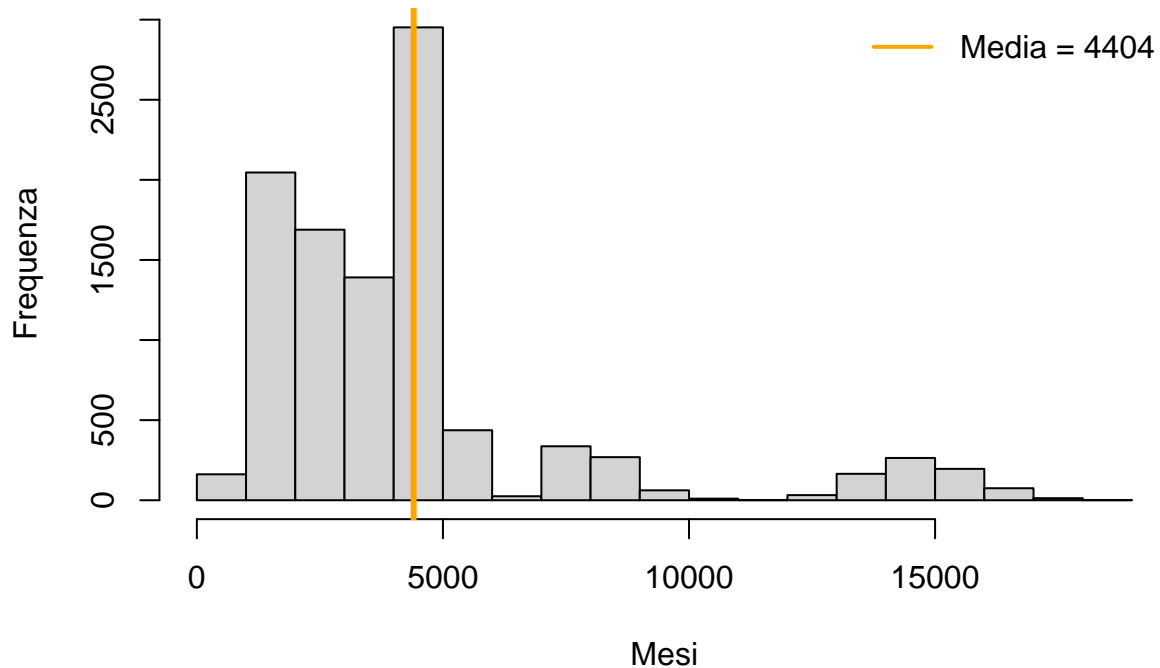
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      510   2156   3899   4404   4741  18484

# minima di 510 dollari
# la massima e' di 18k dollari

# grafico:
hist(BankChurners$Total_Trans_Amt, main = "Istogramma delle transazioni per 12 mesi",
     xlab = "Mesi", ylab = "Frequenza")

# si nota che la maggioranza si tiene sotto i 5k dollari di spesa
# aggiungendo la media:
media_transazioni <- mean(BankChurners$Total_Trans_Amt)
abline(v= media_transazioni, col = "orange", lwd = 3)
legend("topright", legend = paste("Media =", round(media_transazioni, digits = 0)),
      col = "orange", lwd = 2, bty = "n")
```

Istogramma delle transazioni per 12 mesi



*# come detto poco fa una volta superata la media delle transazioni
la frequenza delle persone si abbassa*

PER LE VARIABILI DI TIPO CATEGORIALE(factor)

```
# sesso
BankChurners$Gender <- factor(BankChurners$Gender) # converto in un fattore per fare l'analisi
levels(BankChurners$Gender)
```

```
## [1] "F" "M"
```

```
str(BankChurners$Gender)
```

```
## Factor w/ 2 levels "F","M": 2 1 2 1 2 2 2 2 2 2 ...
```

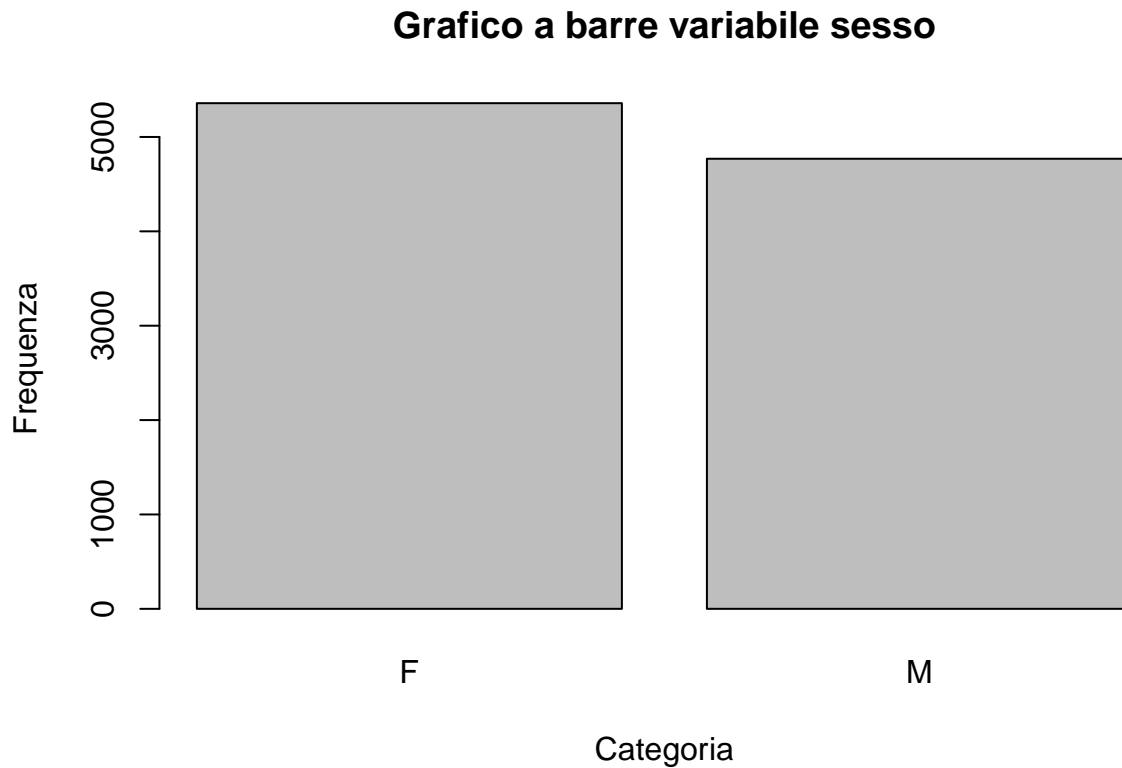
```
# per calcolare la frequenza assoluta
table(BankChurners$Gender)
```

```
##
##      F      M
## 5358 4769
```

```
# marginale
prop.table(table(BankChurners$Gender))
```

```
##
##      F      M
## 0.5290807 0.4709193
```

```
# Grafico a barre
barplot(table(BankChurners$Gender), main="Grafico a barre variabile sesso",
        xlab="Categoria", ylab="Frequenza")
```



```
# Education_Level
BankChurners$Education_Level <- factor(BankChurners$Education_Level) # conversione in fattore
summary(BankChurners$Education_Level)
```

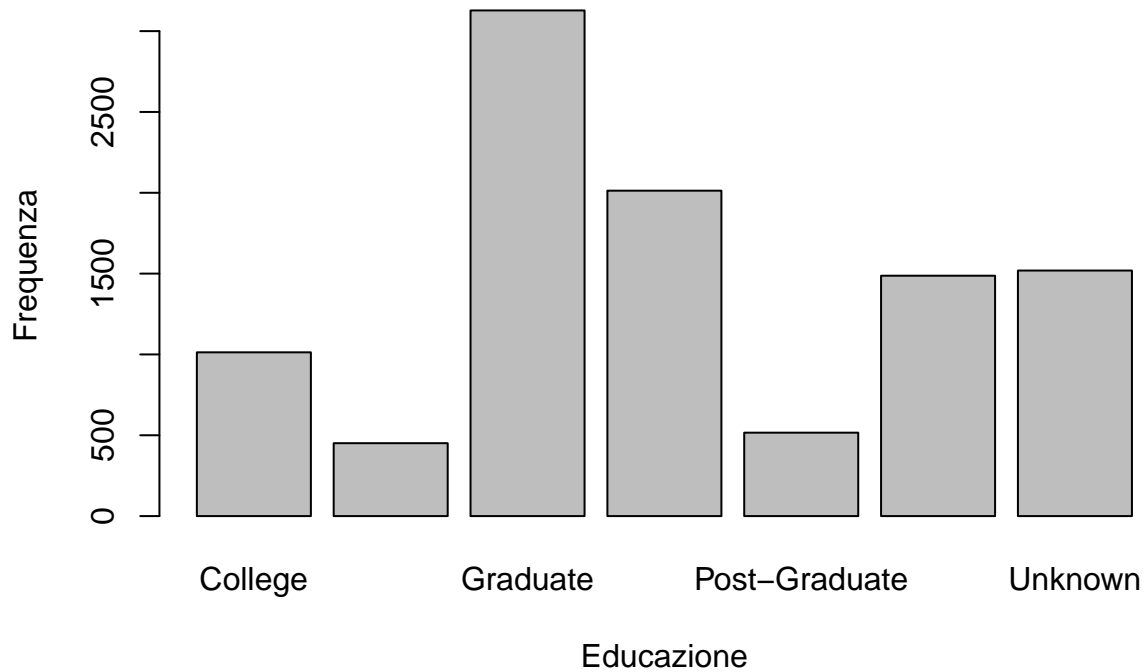
```
##      College      Doctorate      Graduate      High School Post-Graduate
##      1013          451          3128          2013          516
##      Uneducated      Unknown
##      1487          1519
```

```
str(BankChurners$Education_Level)
```

```
## Factor w/ 7 levels "College","Doctorate",...: 4 3 3 4 6 3 7 4 6 3 ...
```

```
barplot(table(BankChurners$Education_Level), main = "Grafico a barre livello educazione",
        xlab = "Educazione", ylab = "Frequenza")
```

Grafico a barre livello educazione



```
#Marital_Status
BankChurners$Marital_Status <- factor(BankChurners$Marital_Status)
str(BankChurners$Marital_Status) # ho 4 livelli(divorziato, sposato, single e stato sconosciuto)

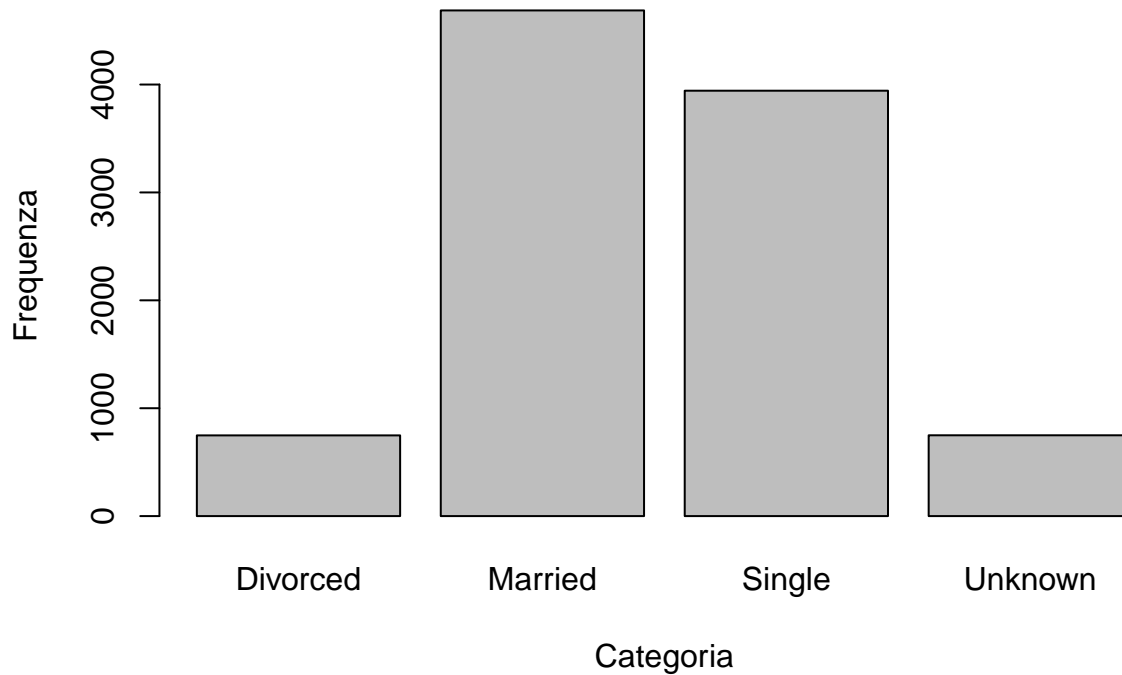
## Factor w/ 4 levels "Divorced","Married",...: 2 3 2 4 2 2 2 2 4 3 3 ...

summary(BankChurners$Marital_Status)

## Divorced Married Single Unknown
##      748    4687    3943     749

barplot(table(BankChurners$Marital_Status), main= "Grafico a barre", xlab="Categoria",
          ylab="Frequenza")
```

Grafico a barre



```
# la maggioranza delle persone e' sposata oppure single
# distribuendo gli stati sconosciuti il risultato sarebbe sempre lo stesso
```

```
round(prop.table(table(BankChurners$Marital_Status)), digits = 3)
```

```
##
## Divorced Married Single Unknown
## 0.074 0.463 0.389 0.074
```

```
# il 46% delle persone e' sposato
# il 38% e' single
# il 16% altro
```

```
# Income_Category
```

```
BankChurners$Income_Category <- factor(BankChurners$Income_Category)
summary(BankChurners$Income_Category)
```

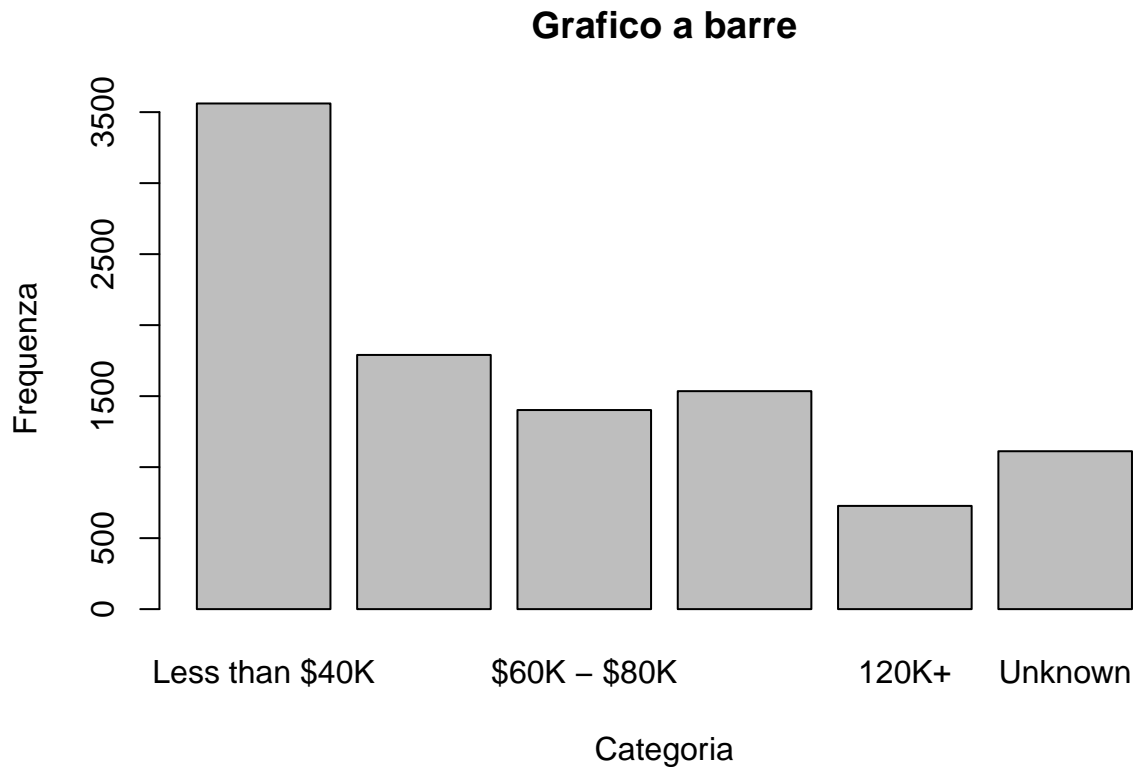
```
## $120K + $40K - $60K $60K - $80K $80K - $120K Less than $40K
## 727 1790 1402 1535 3561
## Unknown
## 1112
```

```
str(BankChurners$Income_Category)
```

```
## Factor w/ 6 levels "$120K +","$40K - $60K",...: 3 5 4 5 3 2 1 3 3 4 ...
```

```
levels(BankChurners$Income_Category) <-
c("120K+", "$40K - $60K", "$60K - $80K", "$80K - $120K", "Less than $40K", "Unknown")
```

```
BankChurners$Income_Category <- factor(
  BankChurners$Income_Category,
  levels = c("Less than $40K", "$40K - $60K", "$60K - $80K", "$80K - $120K", "120K+", "Unknown"))
barplot(table(BankChurners$Income_Category), main= "Grafico a barre",
  xlab="Categoria", ylab="Frequenza")
```



```
# frequenze:
table(BankChurners$Income_Category) # stesso output di summary
```

```
##
## Less than $40K    $40K - $60K    $60K - $80K    $80K - $120K    120K+
##           3561           1790           1402           1535           727
##           Unknown
##           1112
```

```
prop.table(table(BankChurners$Income_Category))
```

```
##
## Less than $40K    $40K - $60K    $60K - $80K    $80K - $120K    120K+
##    0.35163425    0.17675521    0.13844179    0.15157500    0.07178829
##           Unknown
##    0.10980547
```

```
# Card_Category
```

```
BankChurners$Card_Category <- factor(BankChurners$Card_Category)
str(BankChurners$Card_Category)
```

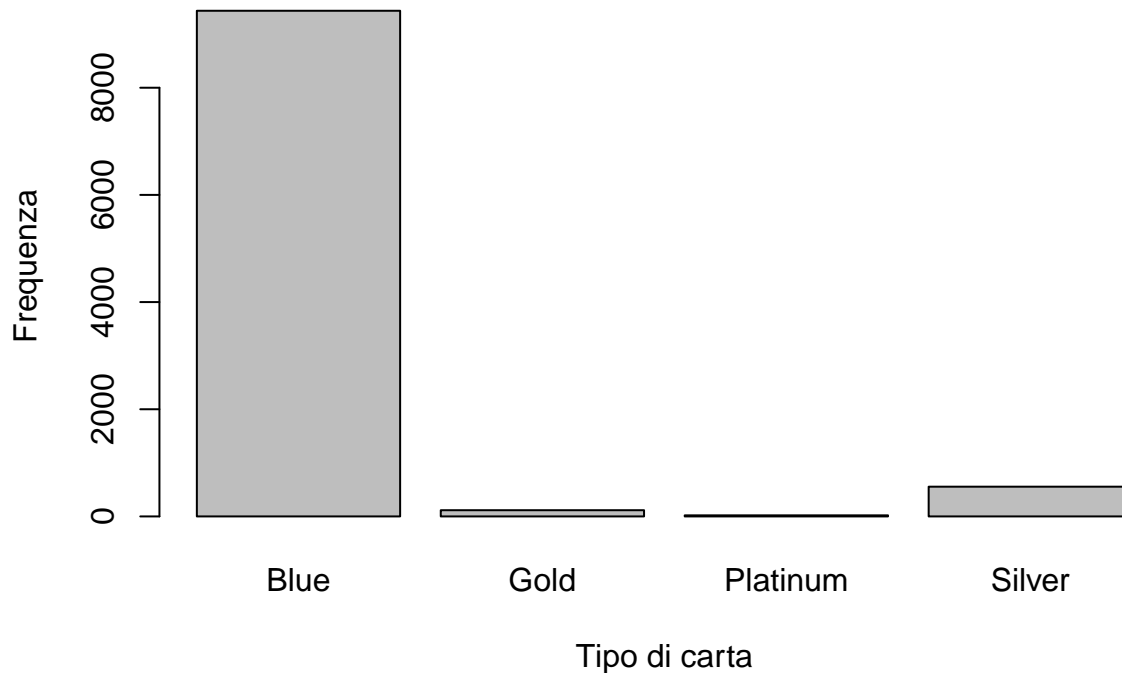
```
## Factor w/ 4 levels "Blue","Gold",...: 1 1 1 1 1 1 2 4 1 1 ...
```

```
summary(BankChurners$Card_Category)
```

```
##      Blue      Gold Platinum   Silver  
##    9436      116        20     555
```

```
barplot(table(BankChurners$Card_Category), main = "Grafico per la categoria delle carte",  
         xlab = "Tipo di carta", ylab = "Frequenza")
```

Grafico per la categoria delle carte



notiamo che la carta di credito di tipo Blue e' nettamente la piu' presente

ANALISI DEL DATASET “HousePrices”

```
HousePricesUnivConf <- read.csv("houseprices-univ.config.csv")
```

```
univarAnalysis <- function(conf, df) {  
  cat("\n\n\\clearpage\n\\newpage")  
  cat("\n\n##", conf["colname"], "\n")  
  if (as.logical(conf["uselog"])) {  
    df[, conf["colname"]] <- log(df[, conf["colname"]])  
  }  
  cat("\n\n Numero di NA: ", sum(is.na(df[, conf["colname"]]), na.rm = T), "\n\n")  
  if (as.logical(conf["filterzeros"])) {  
    cat("\n\n Numero di zeri rimossi: ", sum(df[, conf["colname"]] == 0, na.rm = T), "\n\n")  
    df <- df[df[, conf["colname"]] != 0,]  
  }  
  if (sum(  

```



```

    as.logical(conf["dobarplot"]),
    as.logical(conf["doboxplot"]),
    as.logical(conf["dohist"]),
    as.logical(conf["dodensity"]),
    as.logical(conf["doqqplot"])
  ) > 2) {
    par(mfrow = c(2, 2))
  } else {
    par(mfrow = c(1, 1))
  }
  if (as.logical(conf["dotable"])) {
    print(kable(t(rbind(
      idx = levels(factor(df[, conf["colname"]])),
      abs = table(df[, conf["colname"]]),
      rel = prop.table(table(df[, conf["colname"]]))
    )), col.names = c("Level", "Abs. Freq", "Rel. Freq"), row.names = FALSE))
  } else {
    print(kable(as.array(summary(df[, conf["colname"]])), col.names = c("Stat", "Value")))
  }
  if (as.logical(conf["dobarplot"])) {
    barplot(table(df[, conf["colname"]]), main=NULL)
  }
  if (as.logical(conf["doboxplot"])) {
    boxplot(df[, conf["colname"]], main=NULL)
  }
  if (as.logical(conf["dohist"])) {
    hist(df[, conf["colname"]], freq = F, main=NULL, xlab = NULL)
  }
  if (as.logical(conf["dodensity"])) {
    plot(density(df[, conf["colname"]]), na.rm = T, main="")
  }
  if (as.logical(conf["doqqplot"])) {
    qqnorm(df[, conf["colname"]])
    qqline(df[, conf["colname"]])
  }
  if (!is.na(conf["desc"])) {
    cat("\n\n", conf["desc"], "\n")
  }
}

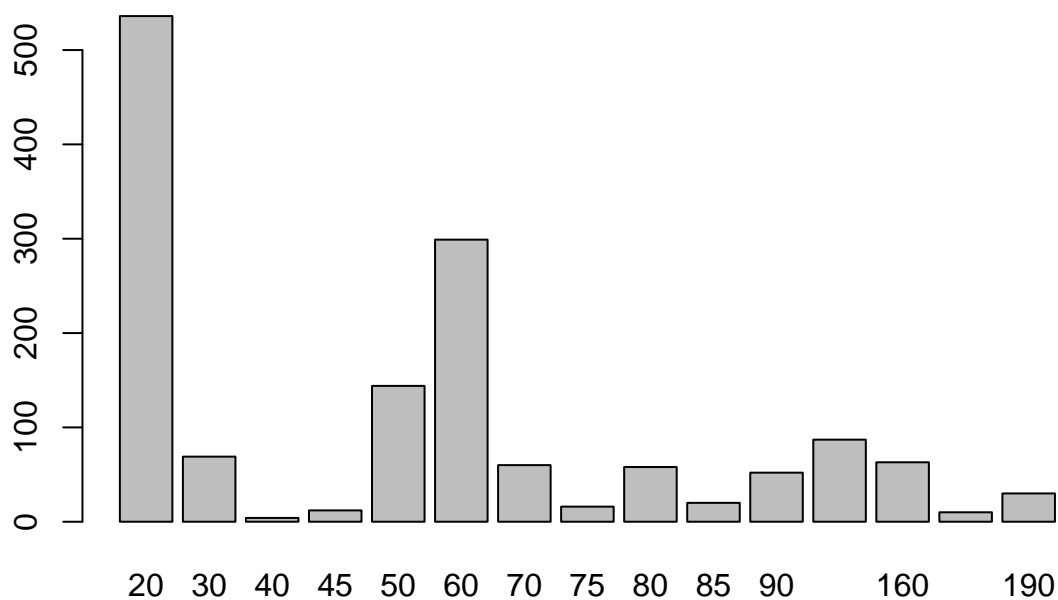
out <- apply(HousePricesUnivConf, 1, univarAnalysis, df = HousePrices)

```

MSSubClass

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
20	536	0.367123287671233
30	69	0.0472602739726027
40	4	0.00273972602739726
45	12	0.00821917808219178
50	144	0.0986301369863014
60	299	0.204794520547945
70	60	0.0410958904109589
75	16	0.010958904109589
80	58	0.0397260273972603
85	20	0.0136986301369863
90	52	0.0356164383561644
120	87	0.0595890410958904
160	63	0.0431506849315069
180	10	0.00684931506849315
190	30	0.0205479452054795

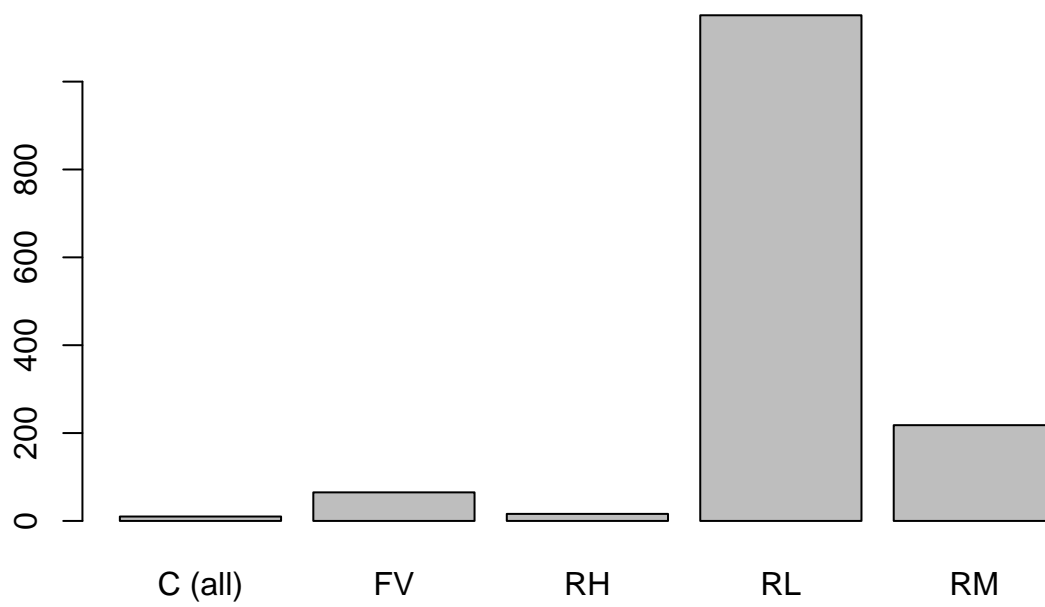


Le case più frequenti sono a 1 o 2 piani successive al 1946

MSZoning

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
C (all)	10	0.00684931506849315
FV	65	0.0445205479452055
RH	16	0.010958904109589
RL	1151	0.788356164383562
RM	218	0.149315068493151

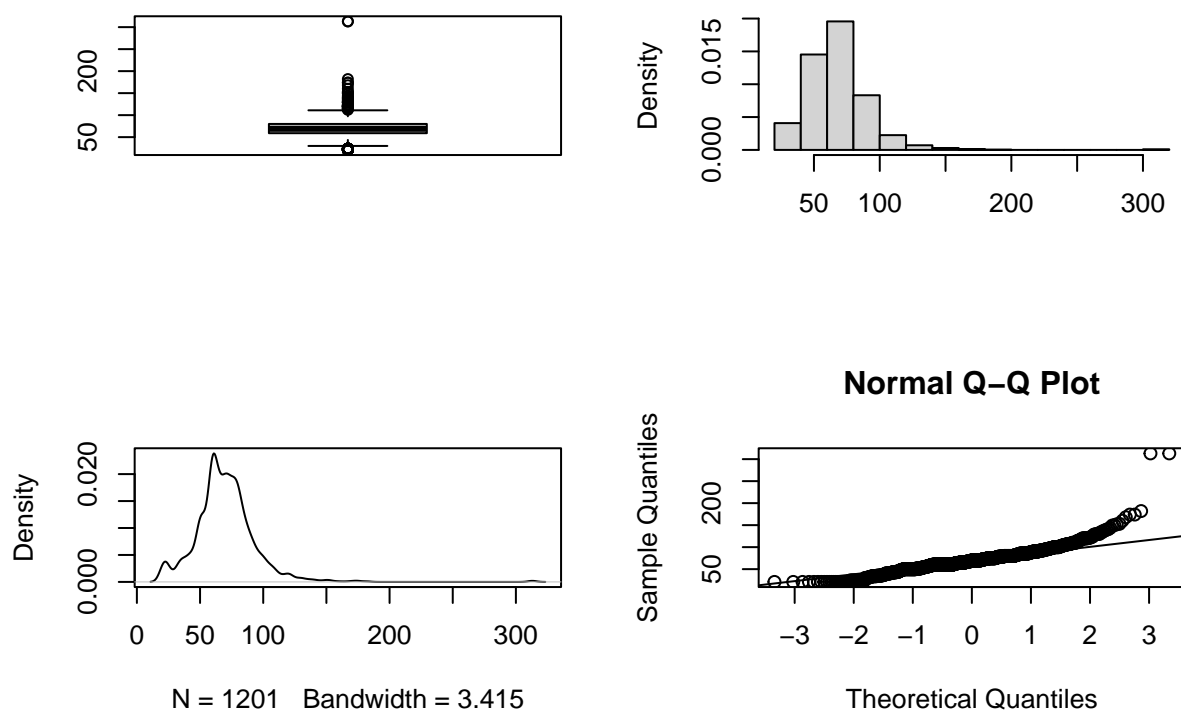


Le case nel dataset sono prevalentemente in zone residenziali a bassa e media densità

LotFrontage

Numero di NA: 259

Stat	Value
Min.	21.00000
1st Qu.	59.00000
Median	69.00000
Mean	70.04996
3rd Qu.	80.00000
Max.	313.00000
NA's	259.00000

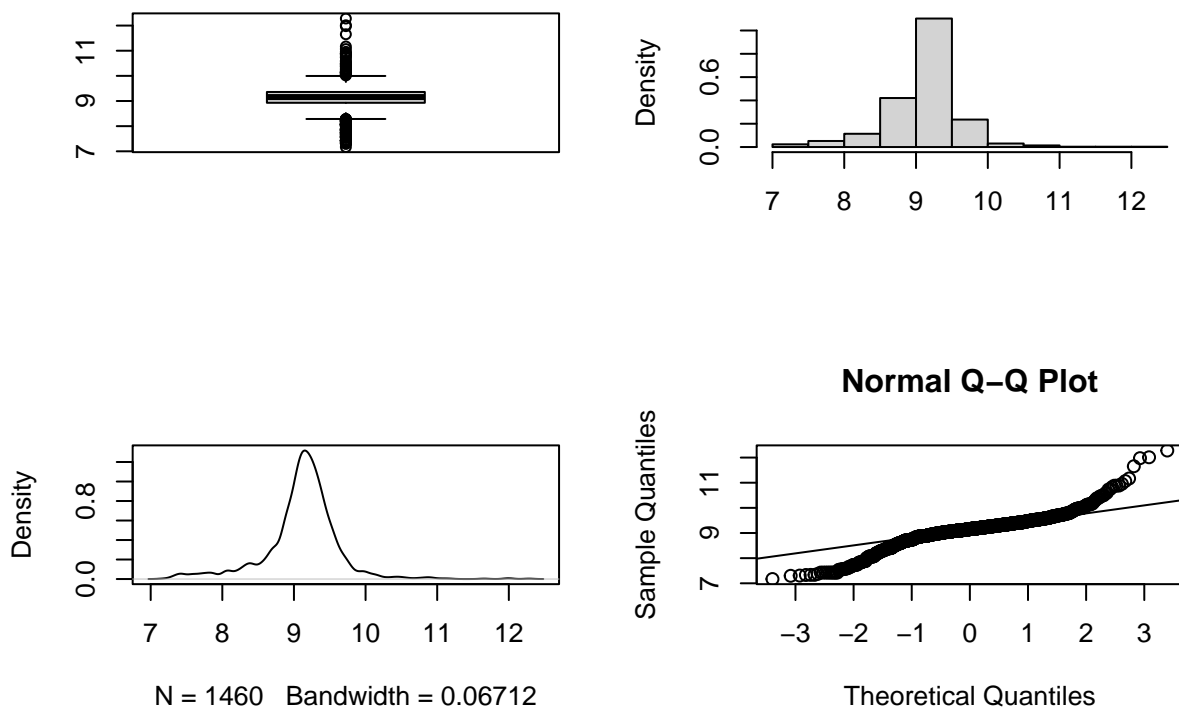


La distribuzione non è normale nelle code. Il boxplot evidenzia che i valori sono solitamente nell'intervallo tra 50 e 100 piedi lineari, spesso possono raggiungere fino ai 200 piedi lineari. Inoltre è presente un outlier superiore a 300 piedi lineari. Confrontando il primo e il terzo quartile si nota che la distribuzione è solo lievemente asimmetrica.

LotArea

Numero di NA: 0

Stat	Value
Min.	7.170120
1st Qu.	8.929766
Median	9.156781
Mean	9.110838
3rd Qu.	9.358890
Max.	12.279532



La distribuzione non è normale nelle code, non lo è neanche il suo logaritmo. La media si attesta intorno ai 10500 piedi quadrati e dai quartili si nota che è asimmetrica.

Street

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
Grvl	6	0.00410958904109589
Pave	1454	0.995890410958904



Solo 6 delle case hanno un accesso non asfaltato sulla strada

Alley

Numero di NA: 1369

Level	Abs. Freq	Rel. Freq
Grvl	50	0.549450549450549
Pave	41	0.450549450549451

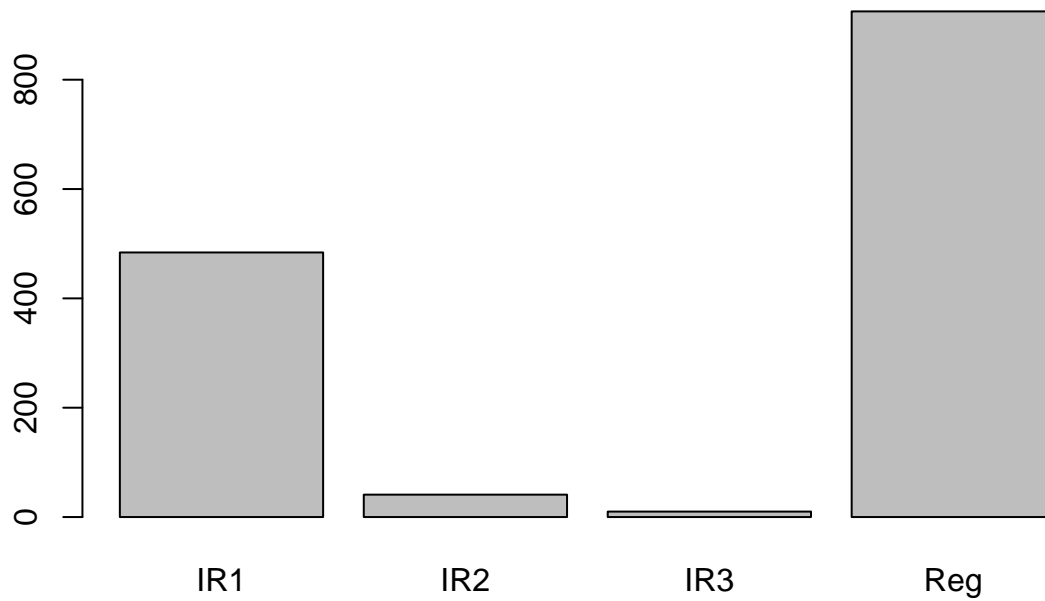


Le case generalmente non hanno un vialetto, tra quelle in cui c'è per metà sono asfaltati

LotShape

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
IR1	484	0.331506849315069
IR2	41	0.0280821917808219
IR3	10	0.00684931506849315
Reg	925	0.633561643835616

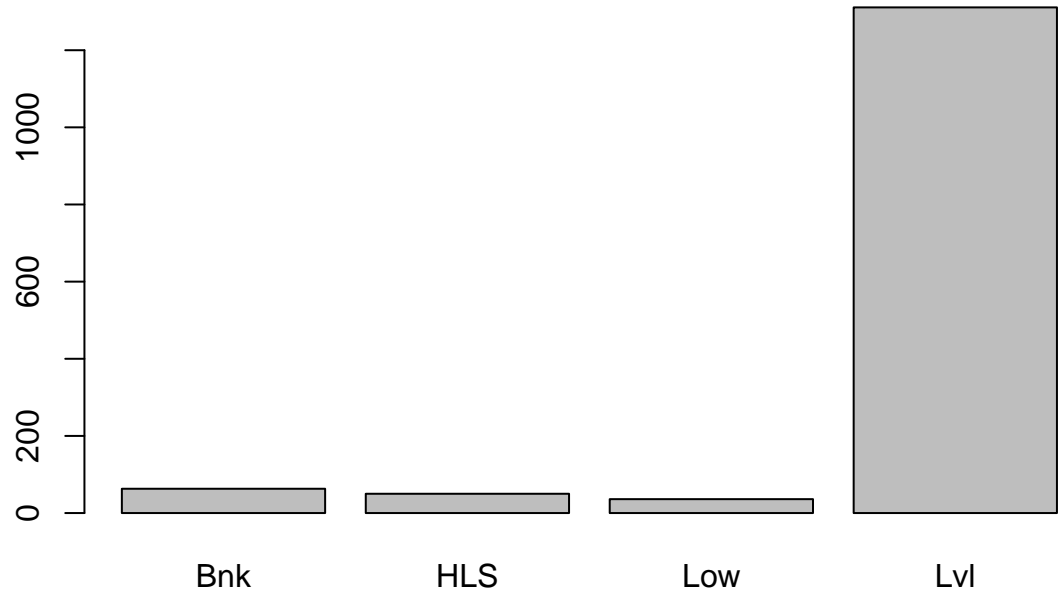


La maggior parte delle proprietà nel dataset sono regolari o presentano lievi irregolarità nella forma

LandContour

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
Bnk	63	0.0431506849315069
HLS	50	0.0342465753424658
Low	36	0.0246575342465753
Lvl	1311	0.897945205479452

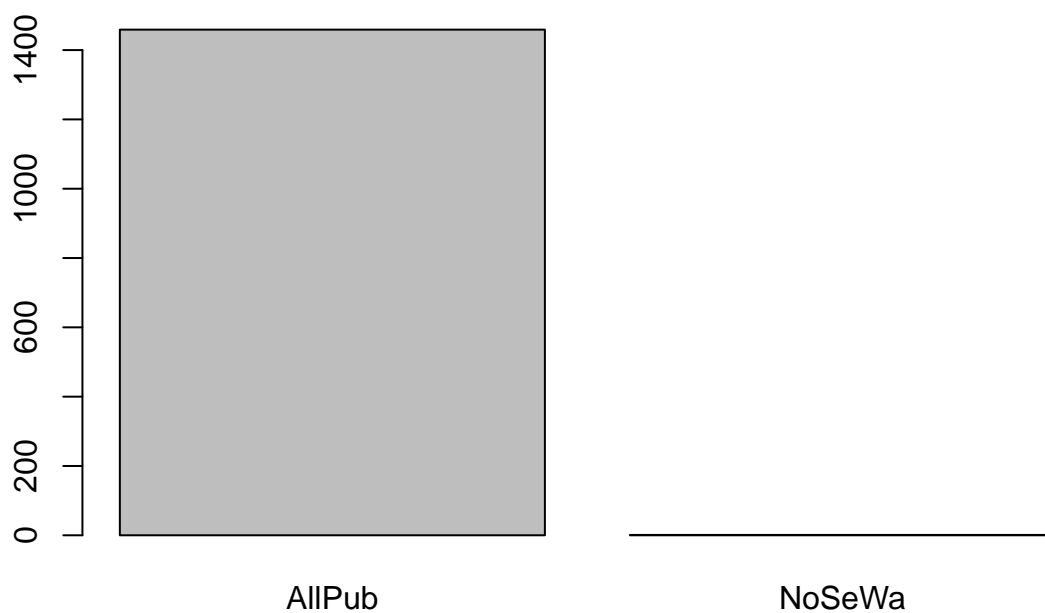


Prevalgono le case costruite in piano

Utilities

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
AllPub	1459	0.999315068493151
NoSeWa	1	0.000684931506849315

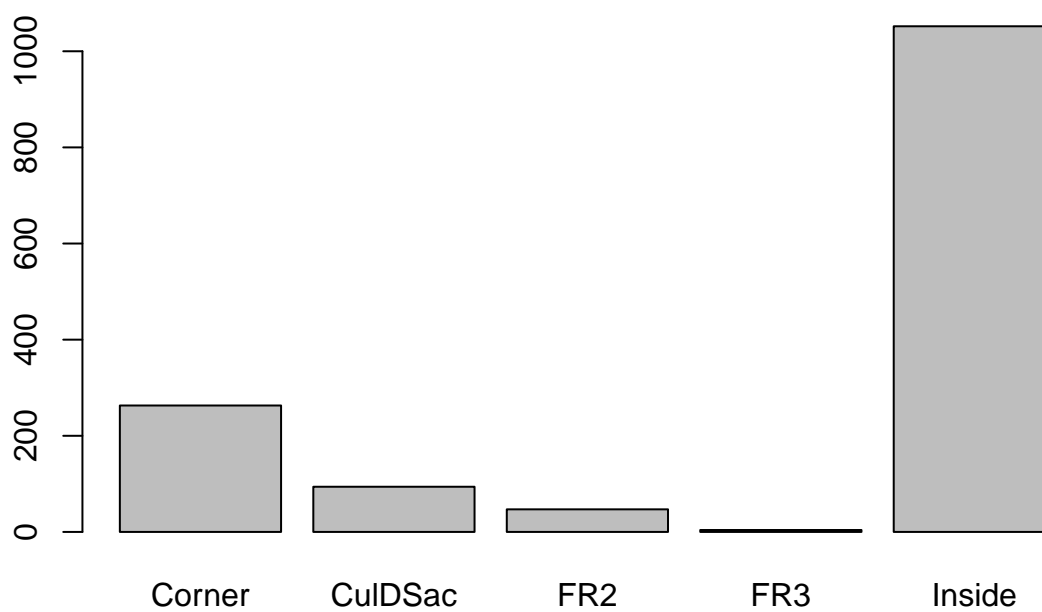


Hanno tutte i servizi essenziali, tranne una

LotConfig

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
Corner	263	0.18013698630137
CulDSac	94	0.0643835616438356
FR2	47	0.0321917808219178
FR3	4	0.00273972602739726
Inside	1052	0.720547945205479

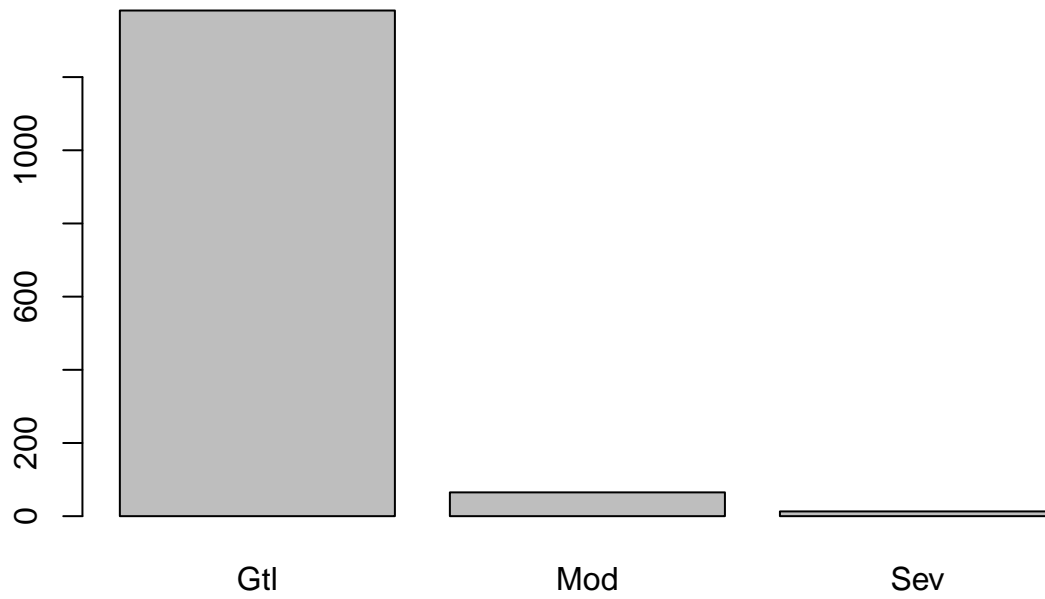


I terreni interni e in angolo sono più frequenti

LandSlope

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
Gtl	1382	0.946575342465753
Mod	65	0.0445205479452055
Sev	13	0.0089041095890411

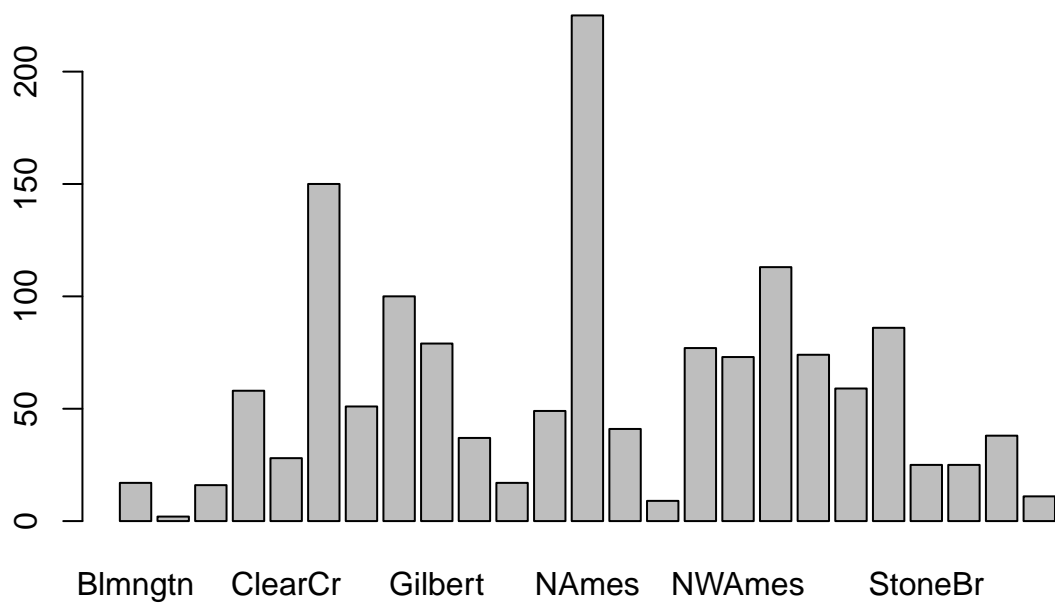


La pendenza del suolo nelle proprietà è lieve nella maggior parte dei casi

Neighborhood

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
Blmngtn	17	0.0116438356164384
Blueste	2	0.00136986301369863
BrDale	16	0.010958904109589
BrkSide	58	0.0397260273972603
ClearCr	28	0.0191780821917808
CollgCr	150	0.102739726027397
Crawfor	51	0.0349315068493151
Edwards	100	0.0684931506849315
Gilbert	79	0.0541095890410959
IDOTRR	37	0.0253424657534247
MeadowV	17	0.0116438356164384
Mitchel	49	0.0335616438356164
NAmes	225	0.154109589041096
NoRidge	41	0.0280821917808219
NPkVill	9	0.00616438356164384
NridgHt	77	0.0527397260273973
NWAmes	73	0.05
OldTown	113	0.0773972602739726
Sawyer	74	0.0506849315068493
SawyerW	59	0.0404109589041096
Somerst	86	0.0589041095890411
StoneBr	25	0.0171232876712329
SWISU	25	0.0171232876712329
Timber	38	0.026027397260274
Veenker	11	0.00753424657534247

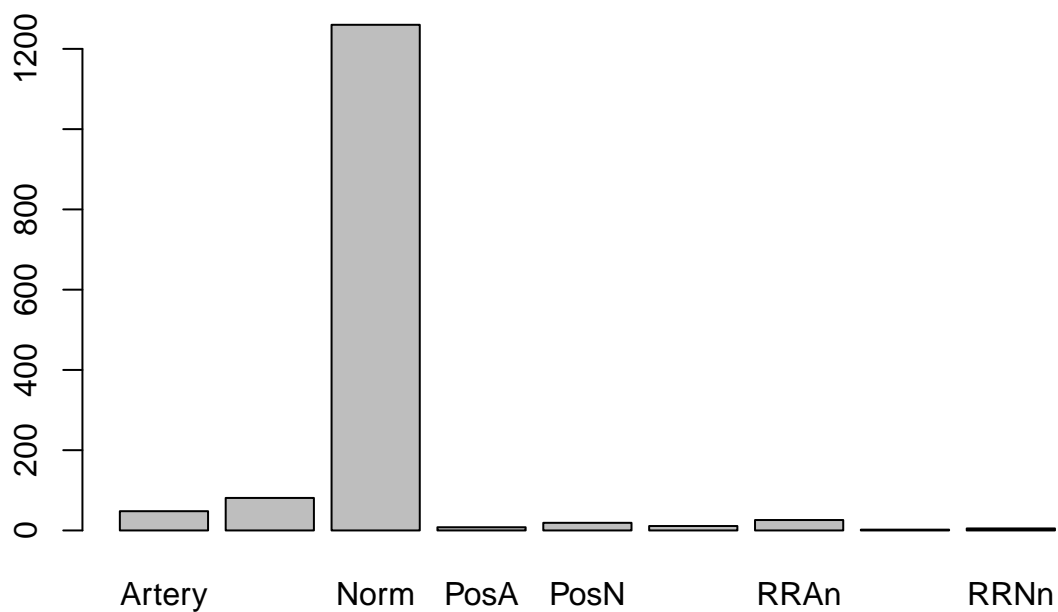


I quartieri più scelti sono North Ames, College Creek e Old Town

Condition1

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
Artery	48	0.0328767123287671
Feedr	81	0.0554794520547945
Norm	1260	0.863013698630137
PosA	8	0.00547945205479452
PosN	19	0.013013698630137
RR Ae	11	0.00753424657534247
RR An	26	0.0178082191780822
RR Ne	2	0.00136986301369863
RR Nn	5	0.00342465753424658

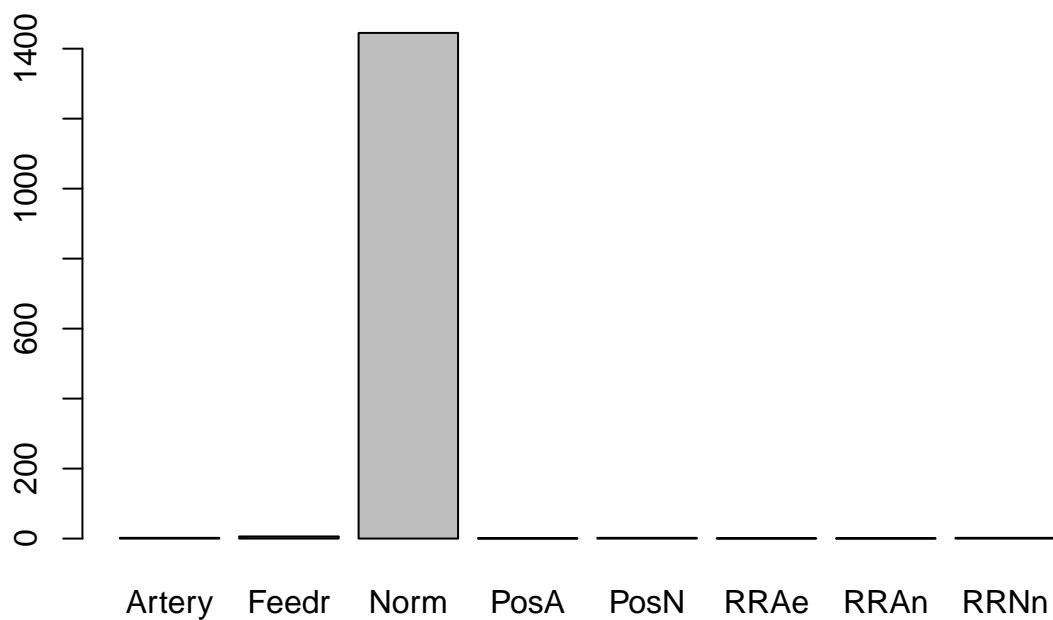


La maggior parte delle case non ha stazioni, arterie o strade di collegamento adiacenti

Condition2

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
Artery	2	0.00136986301369863
Feedr	6	0.00410958904109589
Norm	1445	0.98972602739726
PosA	1	0.000684931506849315
PosN	2	0.00136986301369863
RR Ae	1	0.000684931506849315
RR An	1	0.000684931506849315
RR Nn	2	0.00136986301369863

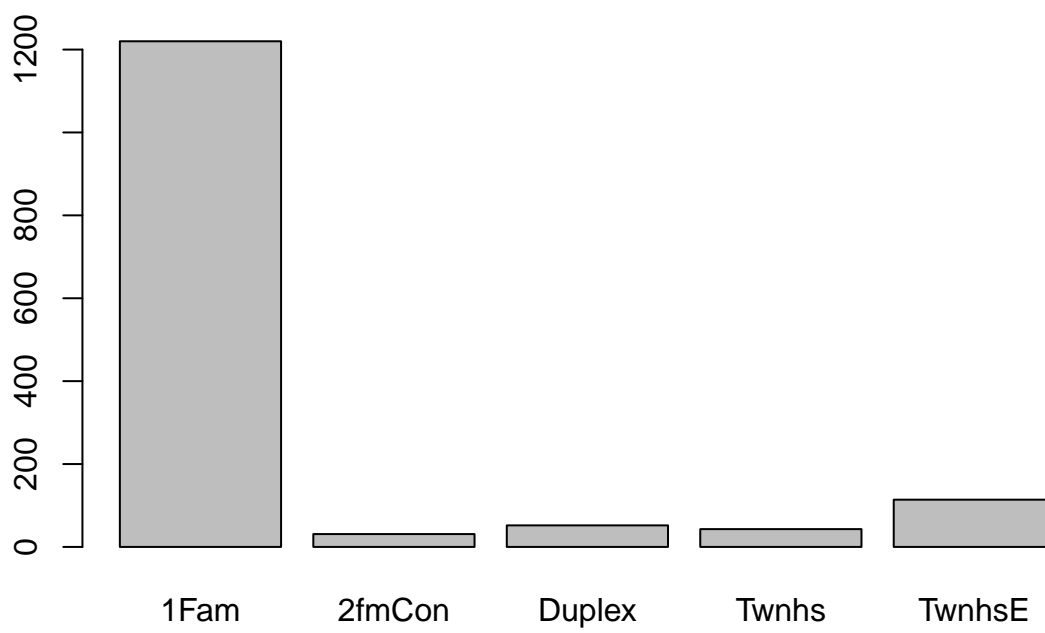


Molto simile al grafico precedente

BldgType

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
1Fam	1220	0.835616438356164
2fmCon	31	0.0212328767123288
Duplex	52	0.0356164383561644
Twnhs	43	0.0294520547945205
TwnhsE	114	0.0780821917808219

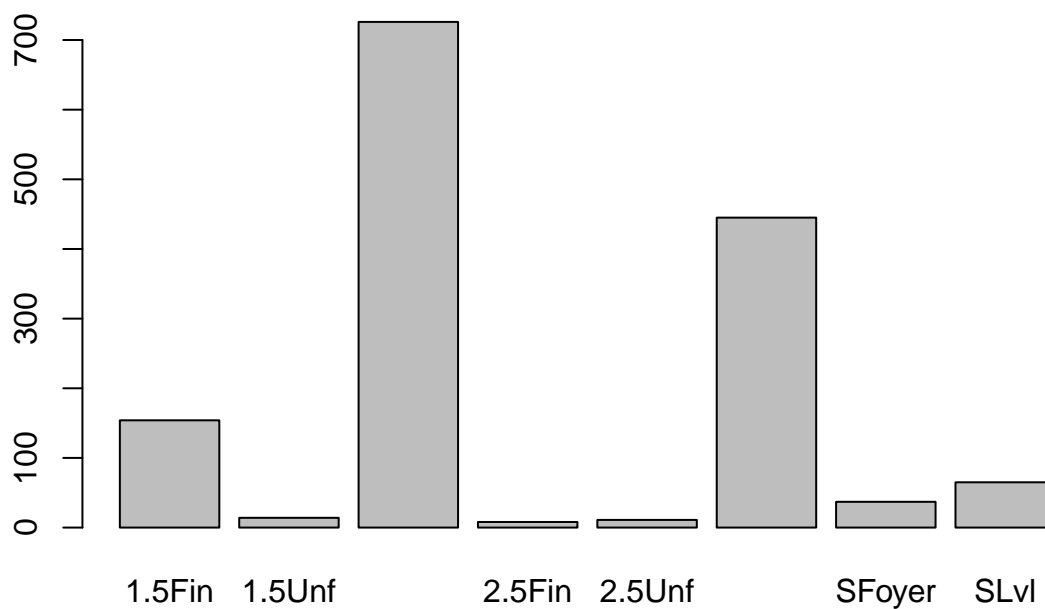


Sono quasi tutte unifamiliari, seguono le villette a schiera

HouseStyle

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
1.5Fin	154	0.105479452054795
1.5Unf	14	0.00958904109589041
1Story	726	0.497260273972603
2.5Fin	8	0.00547945205479452
2.5Unf	11	0.00753424657534247
2Story	445	0.304794520547945
SFoyer	37	0.0253424657534247
SLvl	65	0.0445205479452055

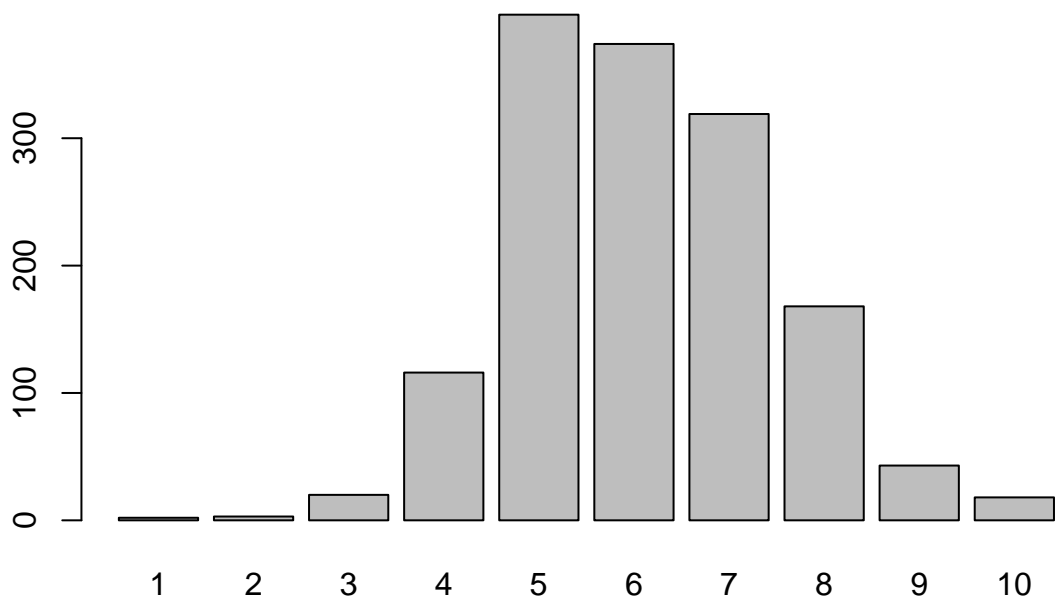


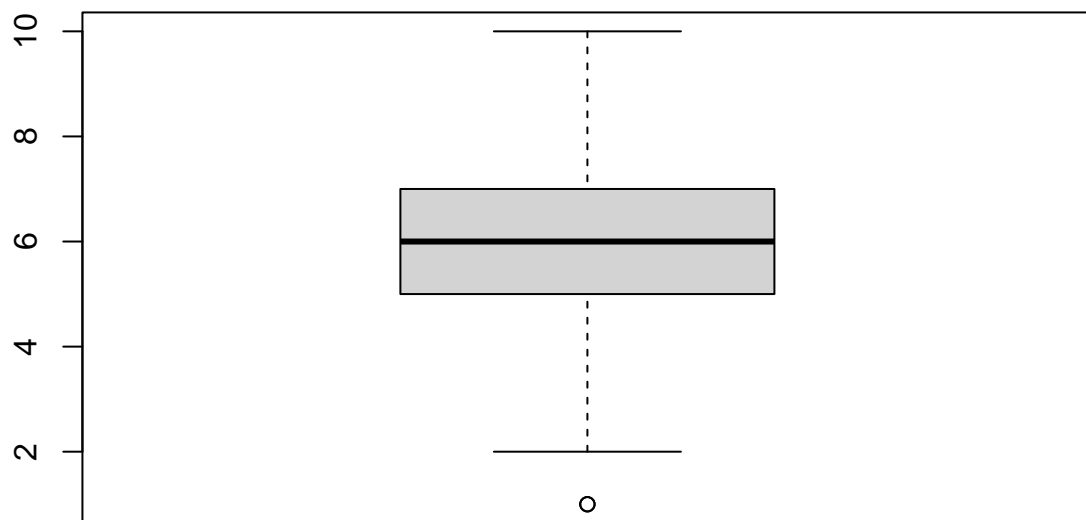
Le case sono prevalentemente a un piano o a due piani

OverallQual

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
1	2	0.00136986301369863
2	3	0.00205479452054795
3	20	0.0136986301369863
4	116	0.0794520547945206
5	397	0.271917808219178
6	374	0.256164383561644
7	319	0.218493150684931
8	168	0.115068493150685
9	43	0.0294520547945205
10	18	0.0123287671232877



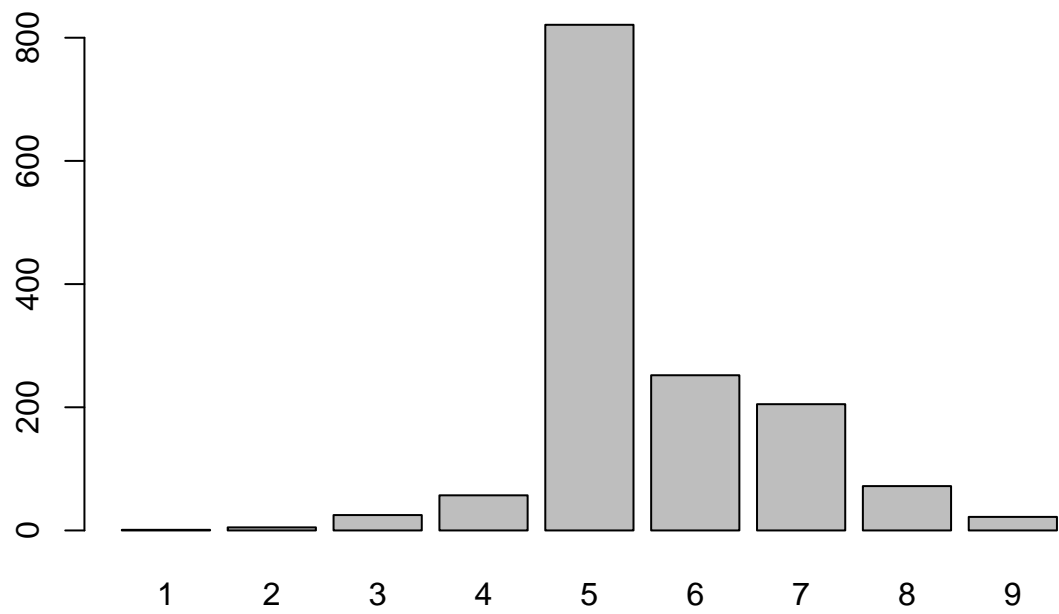


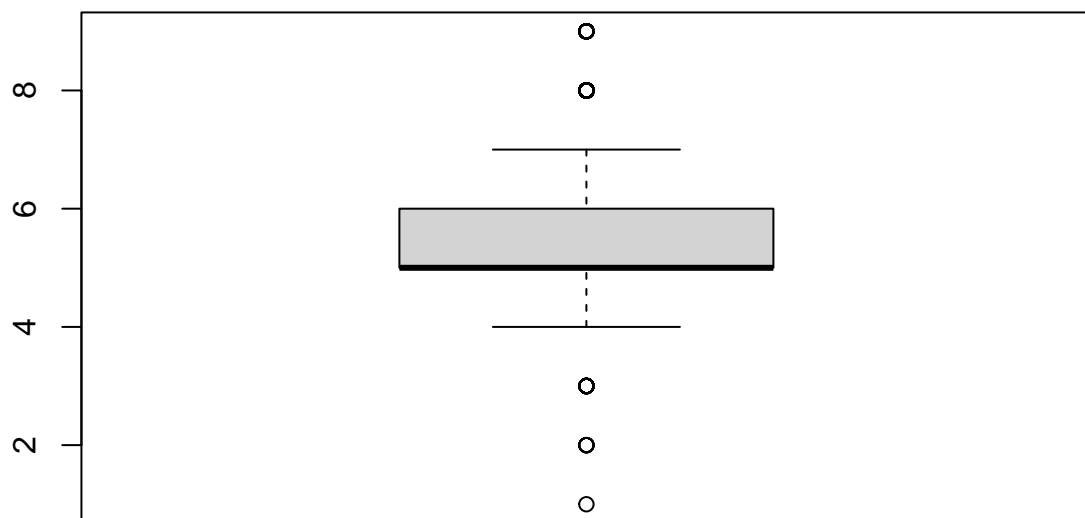
La maggior parte delle case è valutata con una qualità tra 5 e 7 decimi

OverallCond

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
1	1	0.000684931506849315
2	5	0.00342465753424658
3	25	0.0171232876712329
4	57	0.039041095890411
5	821	0.562328767123288
6	252	0.172602739726027
7	205	0.14041095890411
8	72	0.0493150684931507
9	22	0.0150684931506849



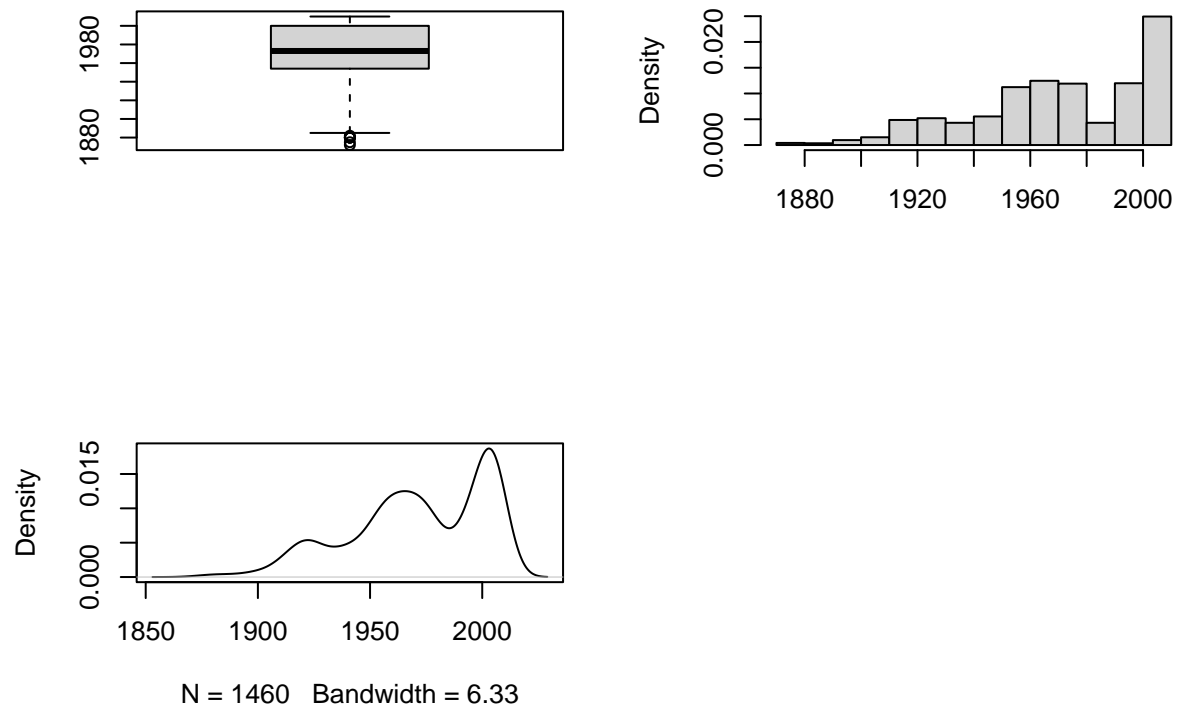


La valutazione delle condizioni solitamente è 5/10 e varia generalmente tra 4 e 7 decimi con alcuni outlier

YearBuilt

Numero di NA: 0

Stat	Value
Min.	1872.000
1st Qu.	1954.000
Median	1973.000
Mean	1971.268
3rd Qu.	2000.000
Max.	2010.000

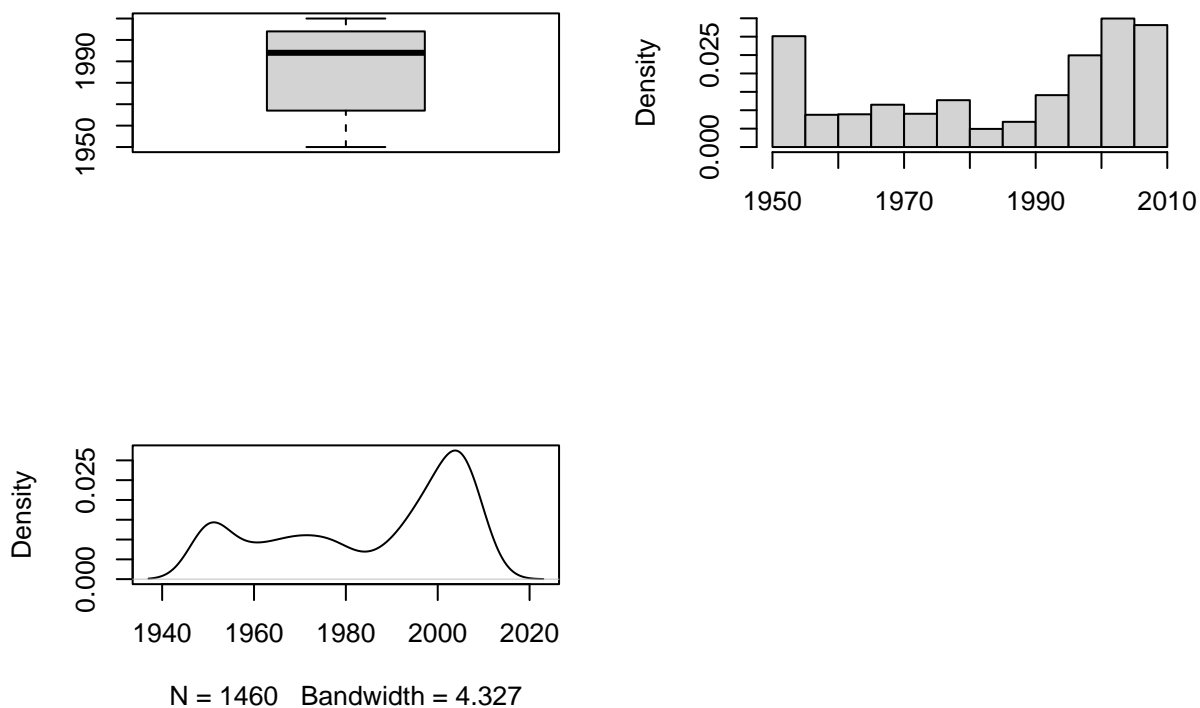


Dal grafico della densità si nota che la distribuzione dei dati ha 3 mode, quindi si evidenziano periodi in cui venivano costruite molte case seguiti da periodi in cui la costruzione di case diminuisce significativamente. Dal boxplot notiamo che la maggior parte delle case è costruita tra il 1954 e il 2000 con alcuni outlier di costruzione antecedente.

YearRemodAdd

Numero di NA: 0

Stat	Value
Min.	1950.000
1st Qu.	1967.000
Median	1994.000
Mean	1984.866
3rd Qu.	2004.000
Max.	2010.000

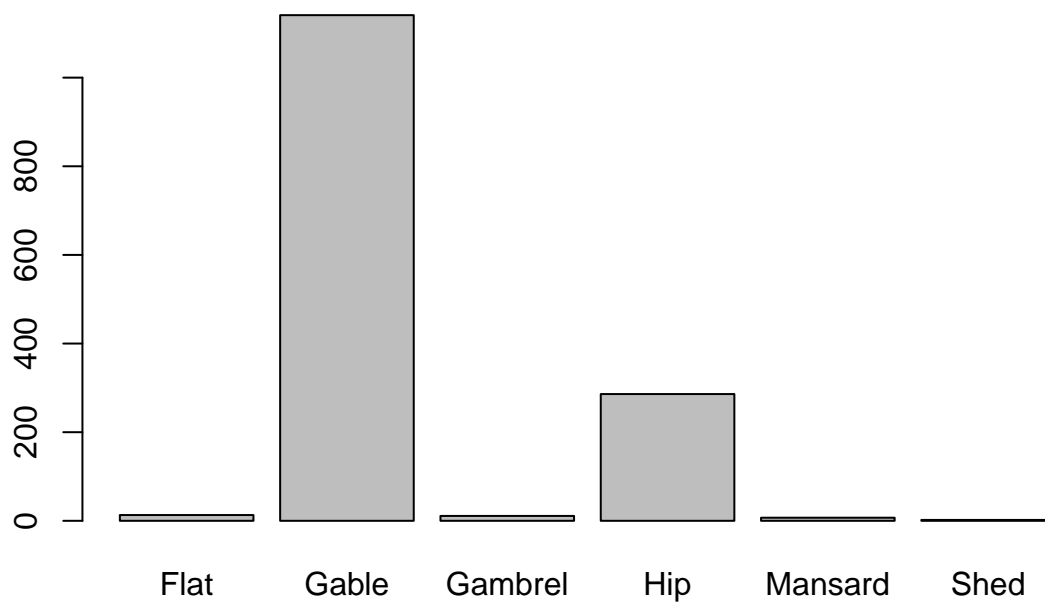


La data dell'ultima ristrutturazione è solitamente recente: tra il 1967 e il 2004, ma c'è anche una grande quantità di edifici che non sono stati ristrutturati dopo gli anni 50'.

RoofStyle

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
Flat	13	0.0089041095890411
Gable	1141	0.781506849315069
Gambrel	11	0.00753424657534247
Hip	286	0.195890410958904
Mansard	7	0.00479452054794521
Shed	2	0.00136986301369863

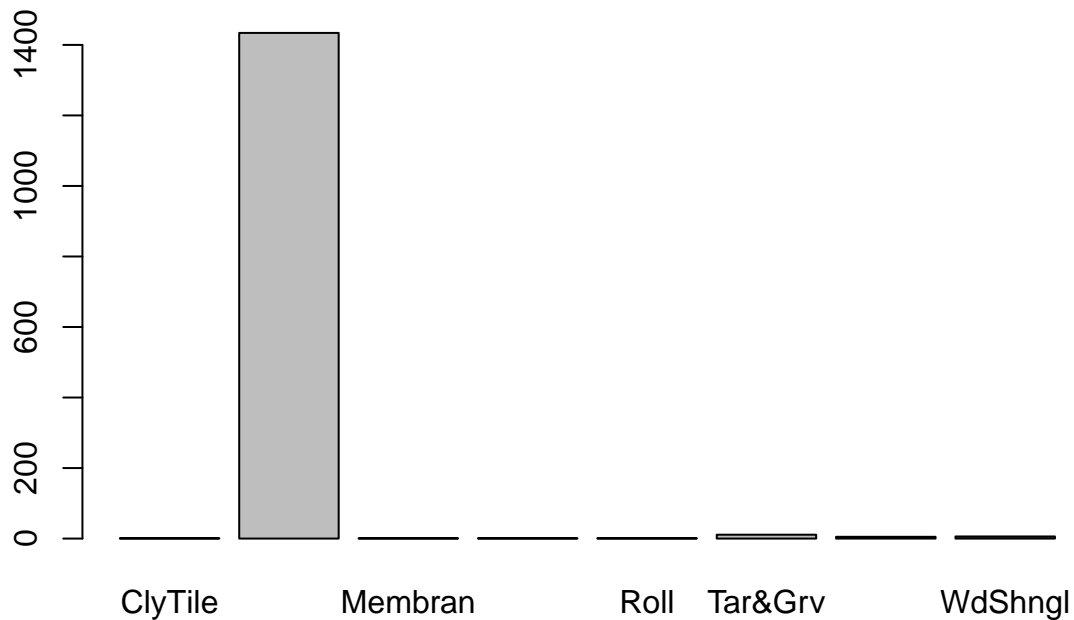


I tetti sono a capanna o a falde

RoofMatl

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
ClyTile	1	0.000684931506849315
CompShg	1434	0.982191780821918
Membran	1	0.000684931506849315
Metal	1	0.000684931506849315
Roll	1	0.000684931506849315
Tar&Grv	11	0.00753424657534247
WdShake	5	0.00342465753424658
WdShngl	6	0.00410958904109589

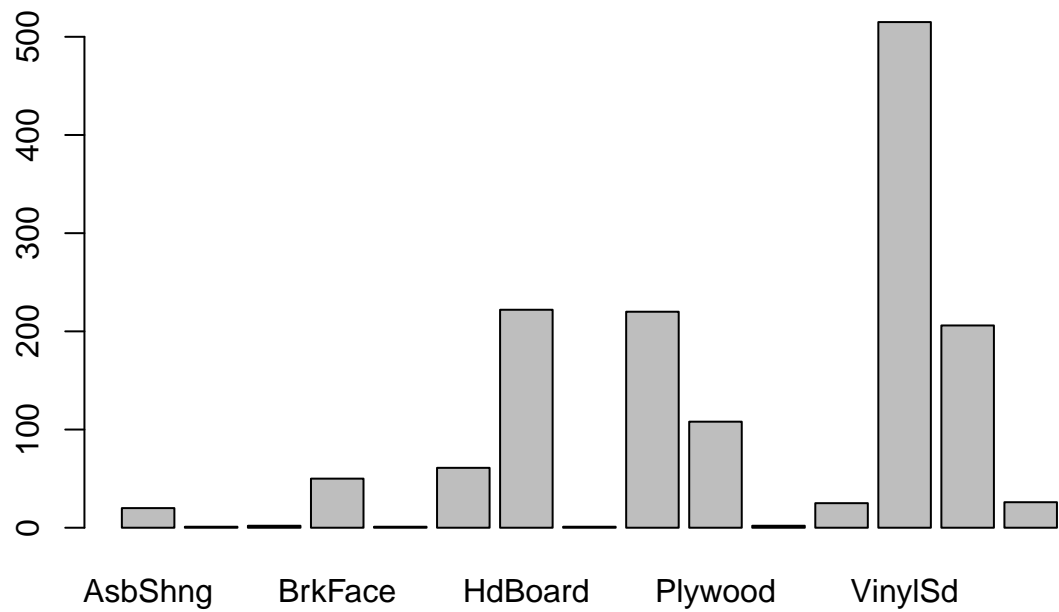


Le coperture sono in tegole standard

Exterior1st

Numero di NA: 0

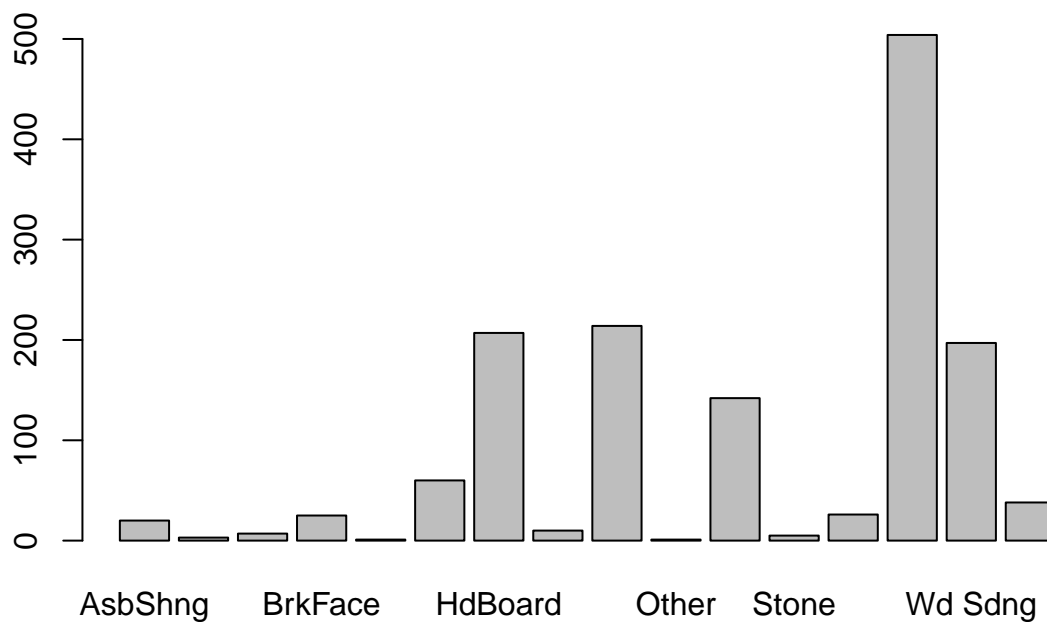
Level	Abs. Freq	Rel. Freq
AsbShng	20	0.0136986301369863
AsphShn	1	0.000684931506849315
BrkComm	2	0.00136986301369863
BrkFace	50	0.0342465753424658
CBlock	1	0.000684931506849315
CemntBd	61	0.0417808219178082
HdBoard	222	0.152054794520548
ImStucc	1	0.000684931506849315
MetalSd	220	0.150684931506849
Plywood	108	0.073972602739726
Stone	2	0.00136986301369863
Stucco	25	0.0171232876712329
VinylSd	515	0.352739726027397
Wd Sdng	206	0.141095890410959
WdShing	26	0.0178082191780822



Exterior2nd

Numero di NA: 0

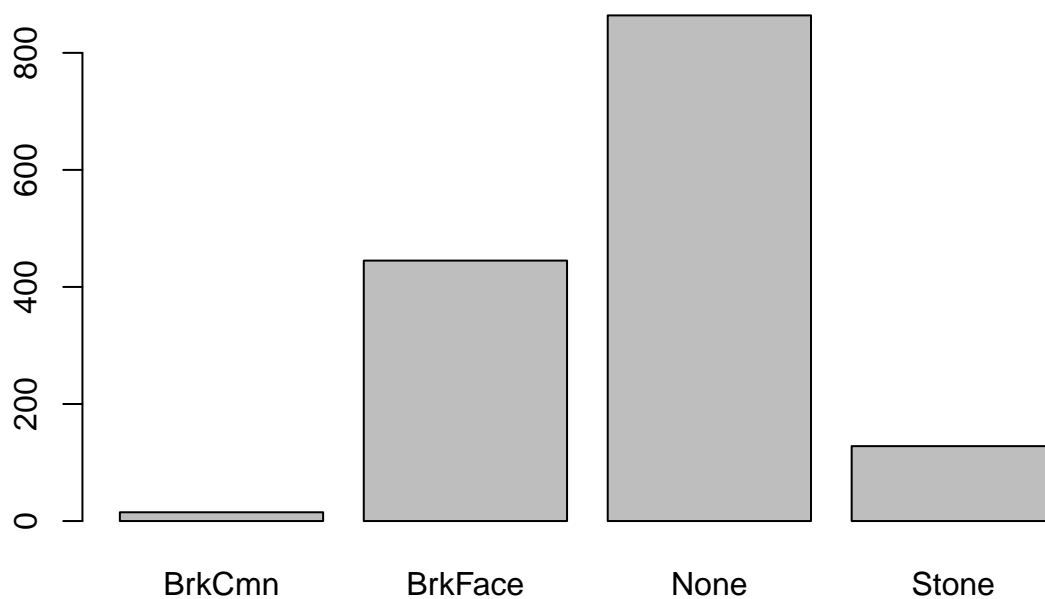
Level	Abs. Freq	Rel. Freq
AsbShng	20	0.0136986301369863
AsphShn	3	0.00205479452054795
Brk Cmn	7	0.00479452054794521
BrkFace	25	0.0171232876712329
CBlock	1	0.000684931506849315
CmentBd	60	0.0410958904109589
HdBoard	207	0.141780821917808
ImStucc	10	0.00684931506849315
MetalSd	214	0.146575342465753
Other	1	0.000684931506849315
Plywood	142	0.0972602739726027
Stone	5	0.00342465753424658
Stucco	26	0.0178082191780822
VinylSd	504	0.345205479452055
Wd Sdng	197	0.134931506849315
Wd Shng	38	0.026027397260274



MasVnrType

Numero di NA: 8

Level	Abs. Freq	Rel. Freq
BrkCmn	15	0.0103305785123967
BrkFace	445	0.306473829201102
None	864	0.59504132231405
Stone	128	0.0881542699724518



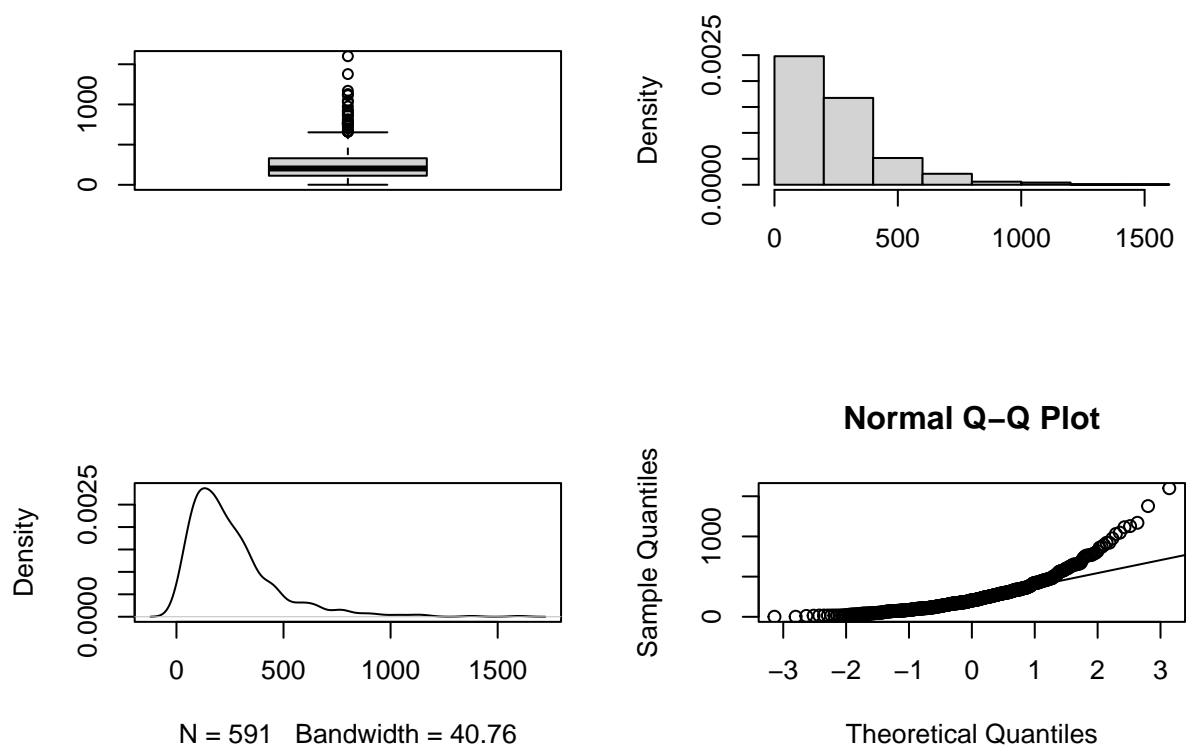
I rivestimenti spesso non sono presenti, qualora fossero presenti sono spesso in mattone e più raramente in pietra

MasVnrArea

Numero di NA: 8

Numero di zeri rimossi: 861

Stat	Value
Min.	1.0000
1st Qu.	113.0000
Median	203.0000
Mean	254.7394
3rd Qu.	330.5000
Max.	1600.0000
NA's	8.0000

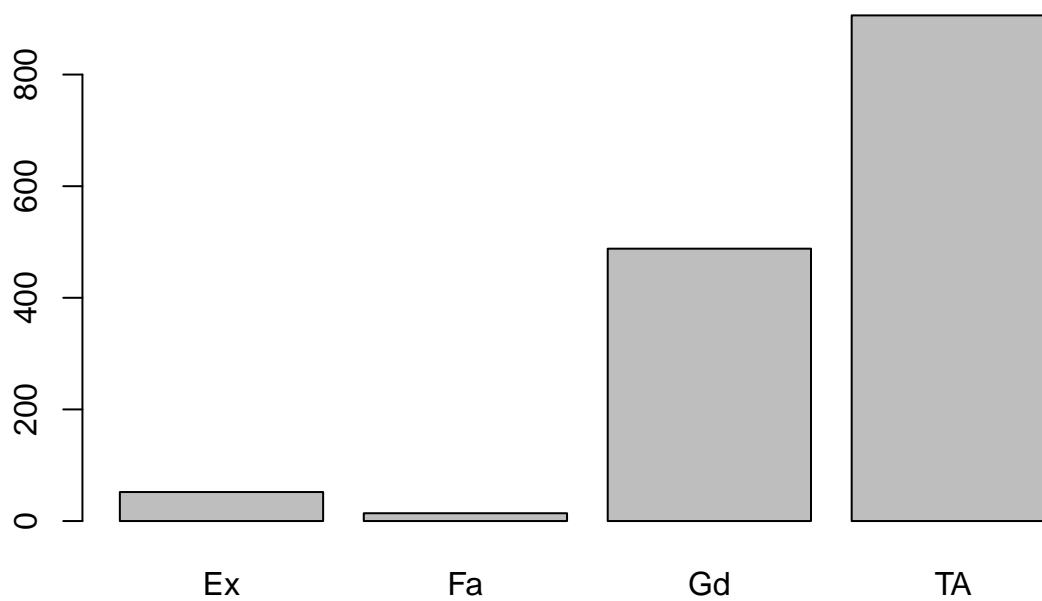


Notiamo che 861 case non hanno rivestimenti dei muri e quindi possiamo non analizzarne la superficie. Escludendo queste notiamo che comunque prevalgono superfici con rivestimenti basse tra i 113 e i 330 piedi quadrati.

ExterQual

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
Ex	52	0.0356164383561644
Fa	14	0.00958904109589041
Gd	488	0.334246575342466
TA	906	0.620547945205479

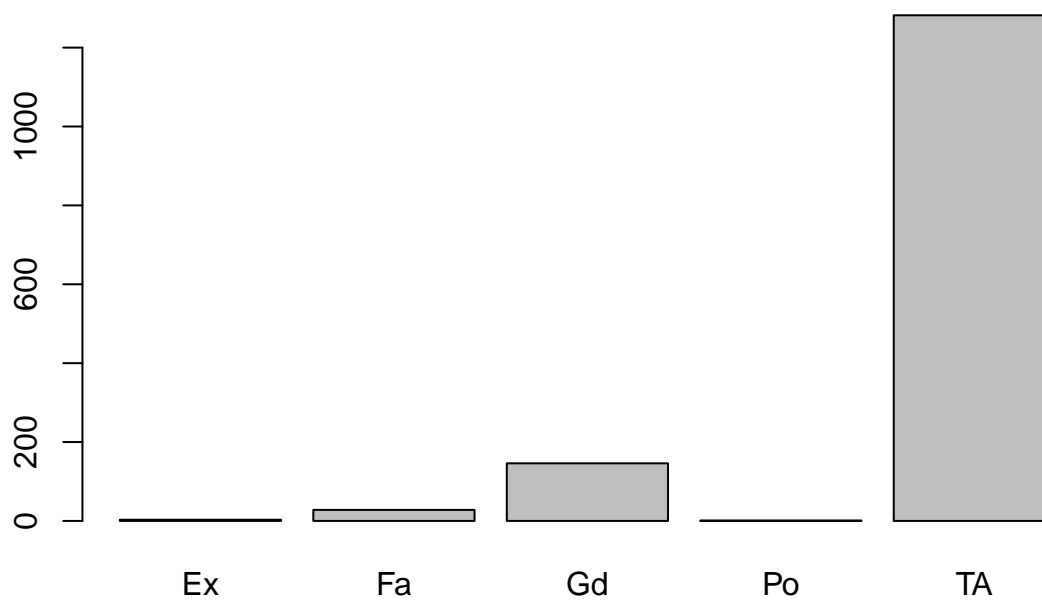


La qualità dei materiali esterni viene generalmente considerata nella media o buona

ExterCond

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
Ex	3	0.00205479452054795
Fa	28	0.0191780821917808
Gd	146	0.1
Po	1	0.000684931506849315
TA	1282	0.878082191780822

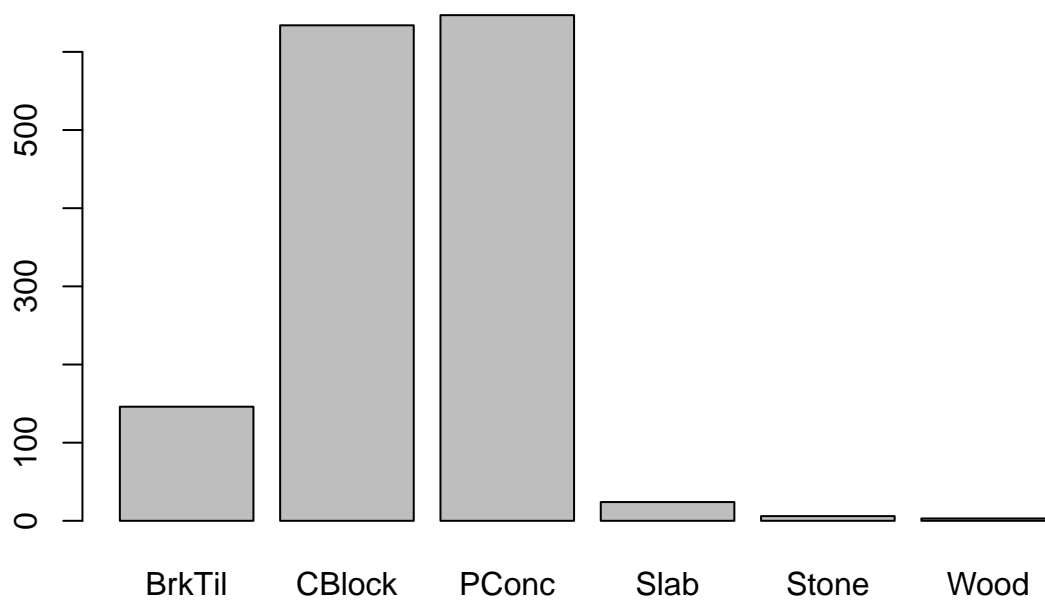


Le condizioni dei materiali esterni sono nella media o buone, ma raramente anche decenti

Foundation

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
BrkTil	146	0.1
CBlock	634	0.434246575342466
PConc	647	0.443150684931507
Slab	24	0.0164383561643836
Stone	6	0.00410958904109589
Wood	3	0.00205479452054795

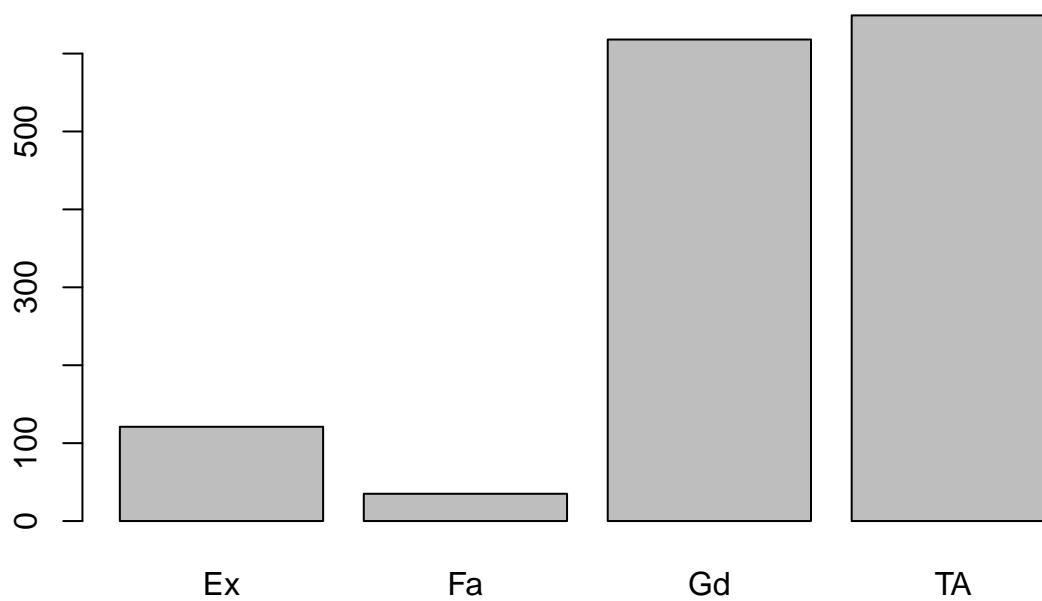


Le fondamenta sono in calcestruzzo o cemento

BsmtQual

Numero di NA: 37

Level	Abs. Freq	Rel. Freq
Ex	121	0.0850316233309909
Fa	35	0.0245959241040056
Gd	618	0.434293745607871
TA	649	0.456078706957133

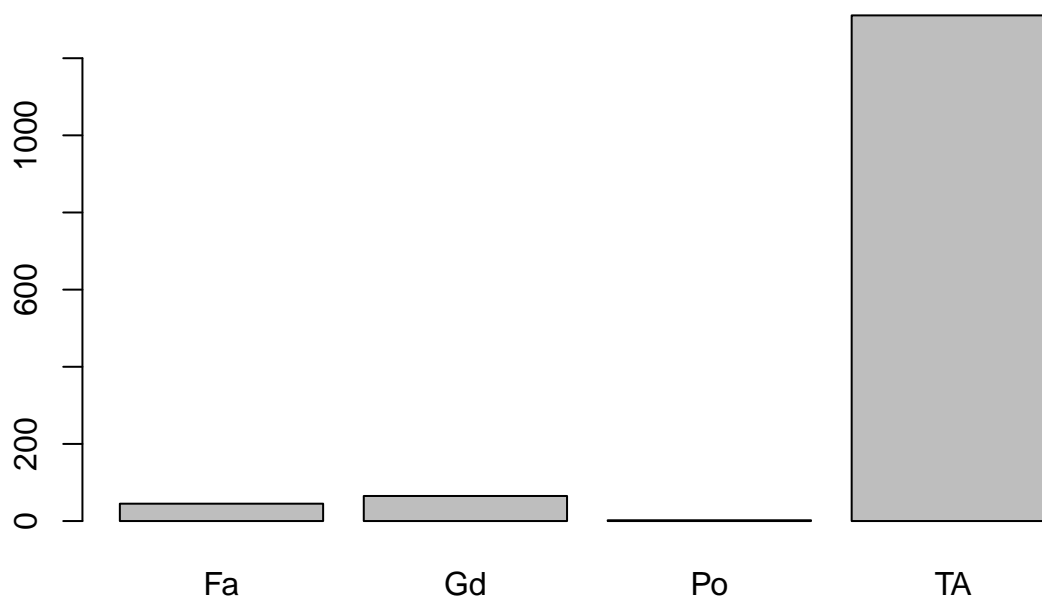


Qualità del seminterrato buona o nella media

BsmtCond

Numero di NA: 37

Level	Abs. Freq	Rel. Freq
Fa	45	0.0316233309908644
Gd	65	0.0456781447645819
Po	2	0.00140548137737175
TA	1311	0.921293042867182

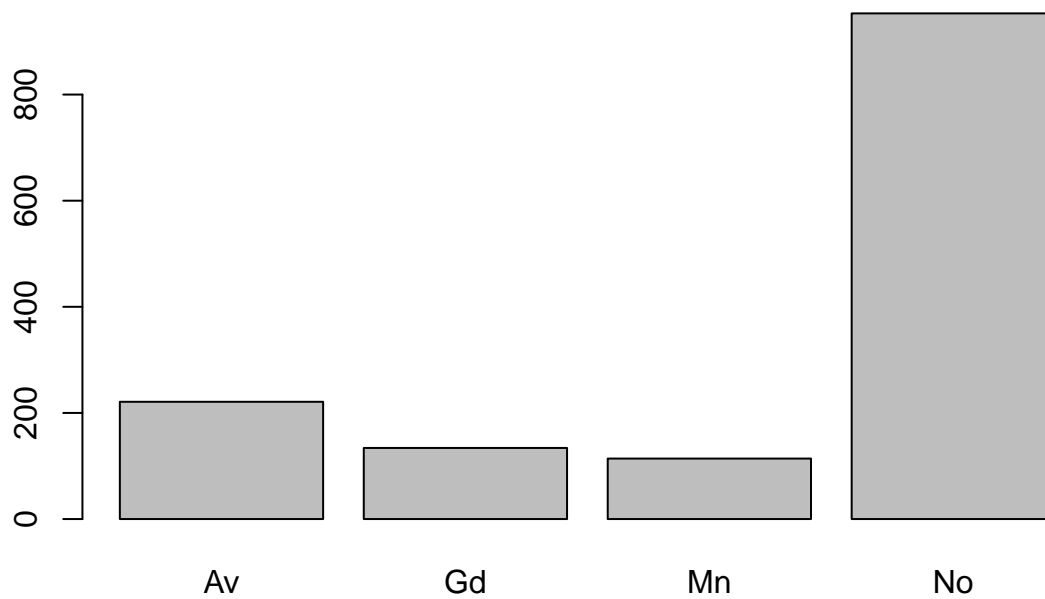


Condizioni del seminterrato nella media

BsmtExposure

Numero di NA: 38

Level	Abs. Freq	Rel. Freq
Av	221	0.155414908579466
Gd	134	0.0942334739803094
Mn	114	0.080168776371308
No	953	0.670182841068917

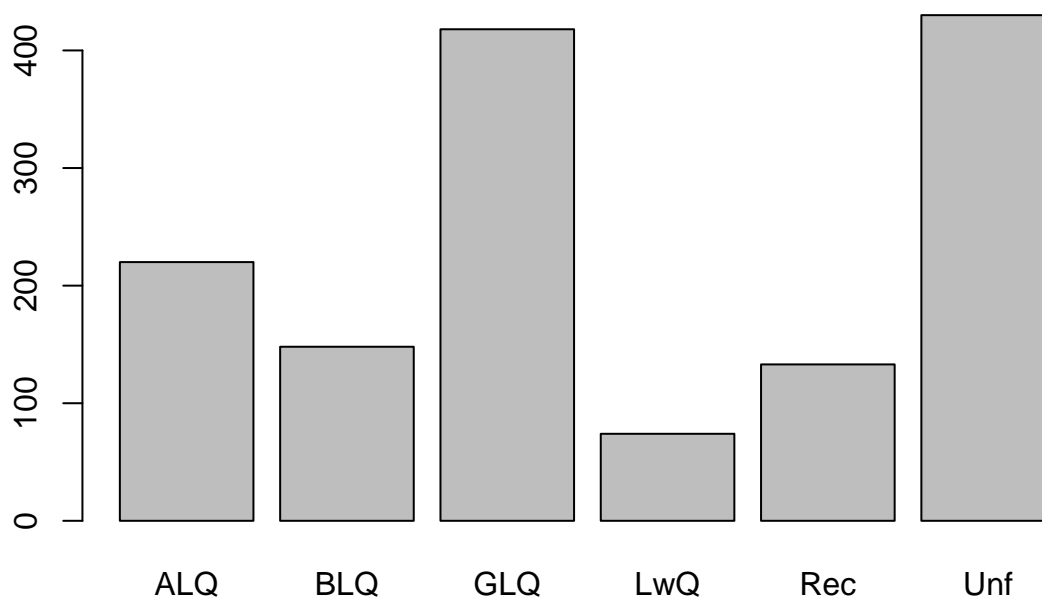


Il seminterrato raramente può essere esposto ma generalmente non lo è

BsmtFinType1

Numero di NA: 37

Level	Abs. Freq	Rel. Freq
ALQ	220	0.154602951510892
BLQ	148	0.104005621925509
GLQ	418	0.293745607870696
LwQ	74	0.0520028109627547
Rec	133	0.0934645115952214
Unf	430	0.302178496134926



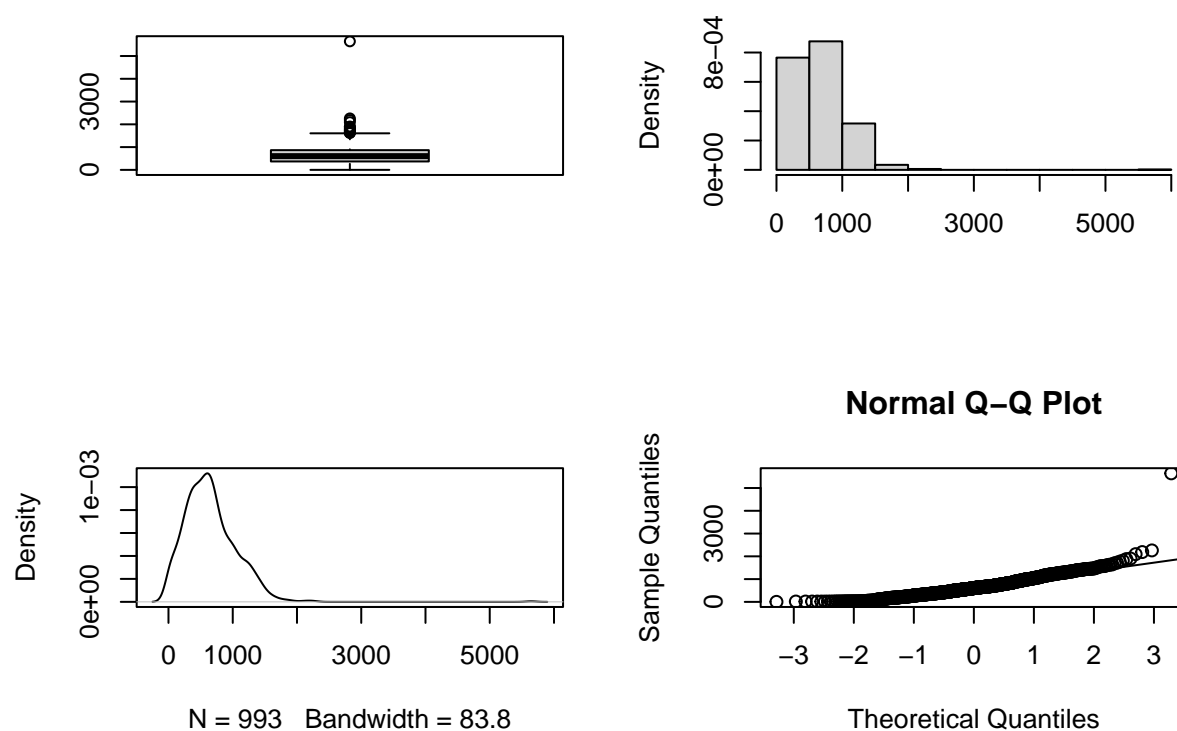
I rating del seminterrato sono disomogenei

BsmtFinSF1

Numero di NA: 0

Numero di zeri rimossi: 467

Stat	Value
Min.	2.00
1st Qu.	371.00
Median	604.00
Mean	652.28
3rd Qu.	867.00
Max.	5644.00

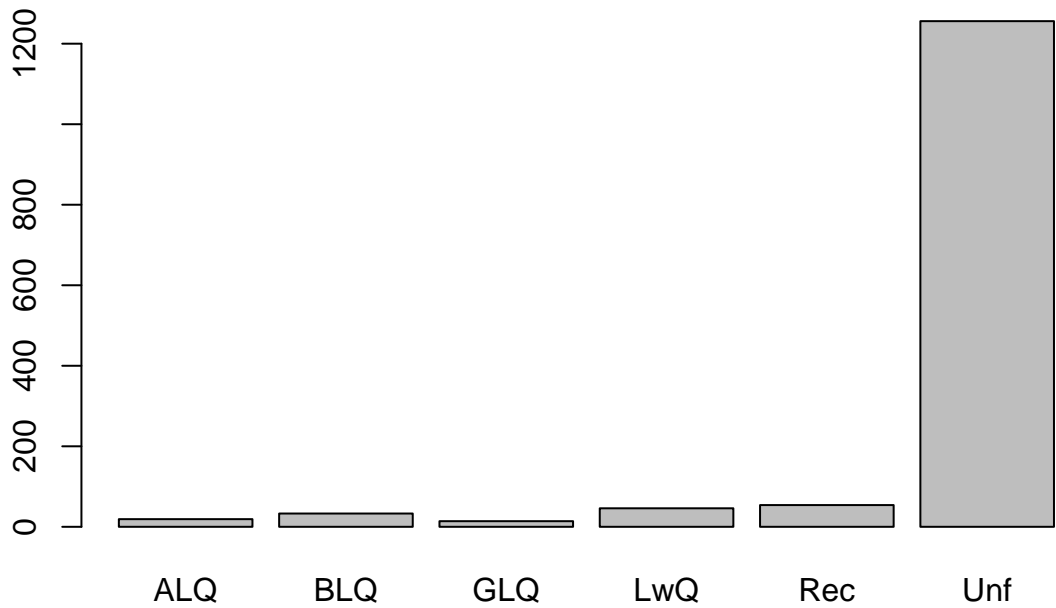


Escludiamo come nel caso precedente dall'analisi della superficie dei seminterrati i valori di superficie pari a 0. Notiamo che le superfici dei seminterrati con la categoria di utilizzo definita in `BsmtFinType1` sono distribuiti tra 371 e 867 piedi quadrati con un outlier massimo di 5644 piedi quadrati.

BsmtFinType2

Numero di NA: 38

Level	Abs. Freq	Rel. Freq
ALQ	19	0.0133614627285513
BLQ	33	0.0232067510548523
GLQ	14	0.00984528832630099
LwQ	46	0.0323488045007032
Rec	54	0.0379746835443038
Unf	1256	0.883263009845288



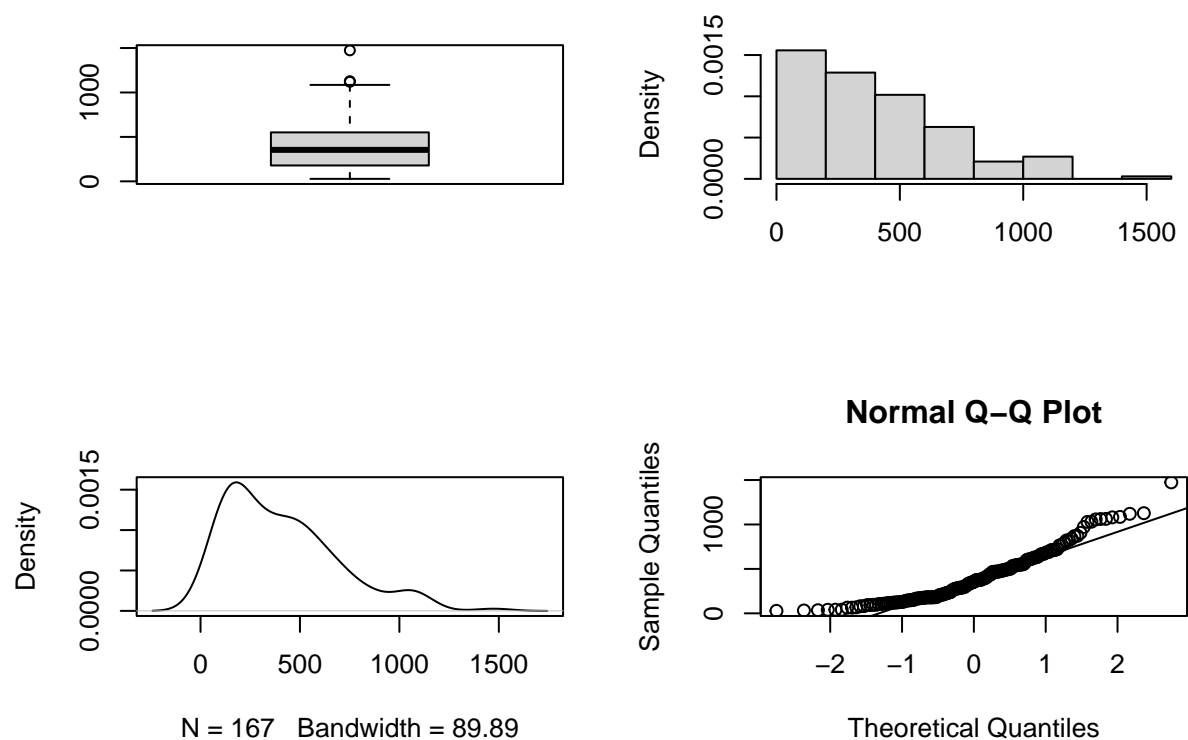
Il secondo seminterrato non è completato

BsmtFinSF2

Numero di NA: 0

Numero di zeri rimossi: 1293

Stat	Value
Min.	28.0000
1st Qu.	178.5000
Median	354.0000
Mean	406.9581
3rd Qu.	551.0000
Max.	1474.0000



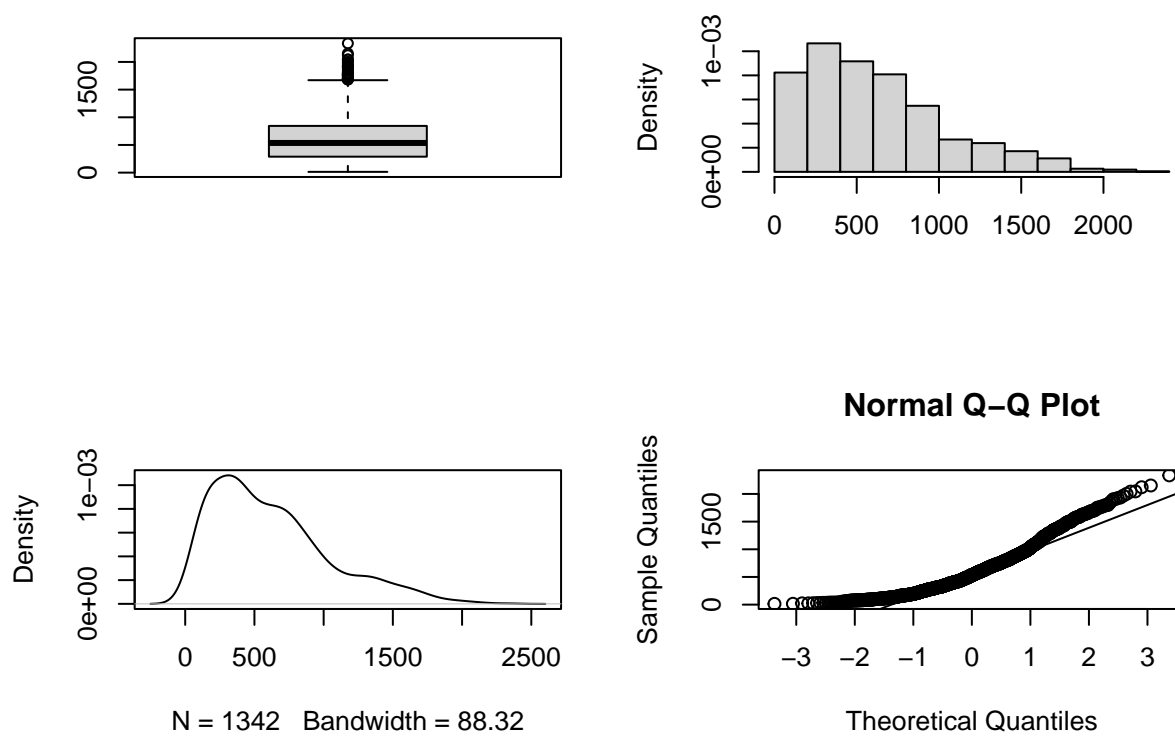
Per il secondo seminterrato notiamo che ci sono ancora più valori nulli e i valori si abbassano notevolmente rispetto alla precedente analisi con primo e terzo quartile 178 e 551.

BsmtUnfSF

Numero di NA: 0

Numero di zeri rimossi: 118

Stat	Value
Min.	14.000
1st Qu.	288.000
Median	536.000
Mean	617.117
3rd Qu.	843.250
Max.	2336.000



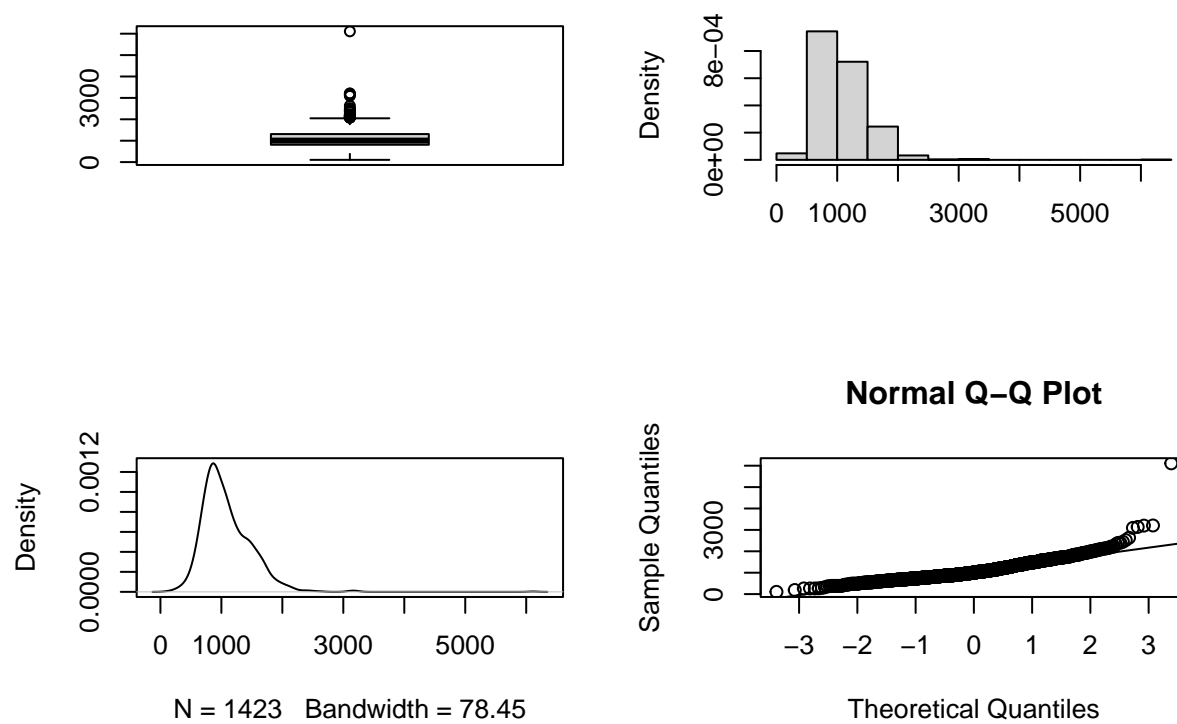
Sono molto più frequenti i seminterrati non finiti, per le superfici di questi continua a valere l'analisi di BsmtFinSF1.

TotalBsmtSF

Numero di NA: 0

Numero di zeri rimossi: 37

Stat	Value
Min.	105.000
1st Qu.	810.500
Median	1004.000
Mean	1084.924
3rd Qu.	1309.500
Max.	6110.000

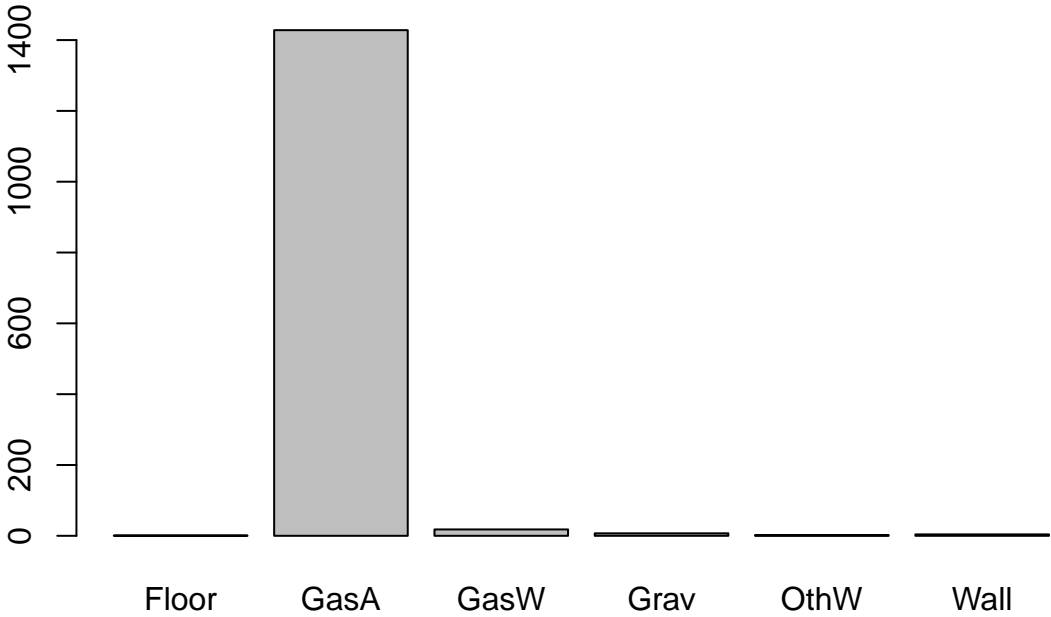


Le superfici totali dei seminterrati sono tra 810 e 1310 piedi quadrati.

Heating

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
Floor	1	0.000684931506849315
GasA	1428	0.978082191780822
GasW	18	0.0123287671232877
Grav	7	0.00479452054794521
OthW	2	0.00136986301369863
Wall	4	0.00273972602739726

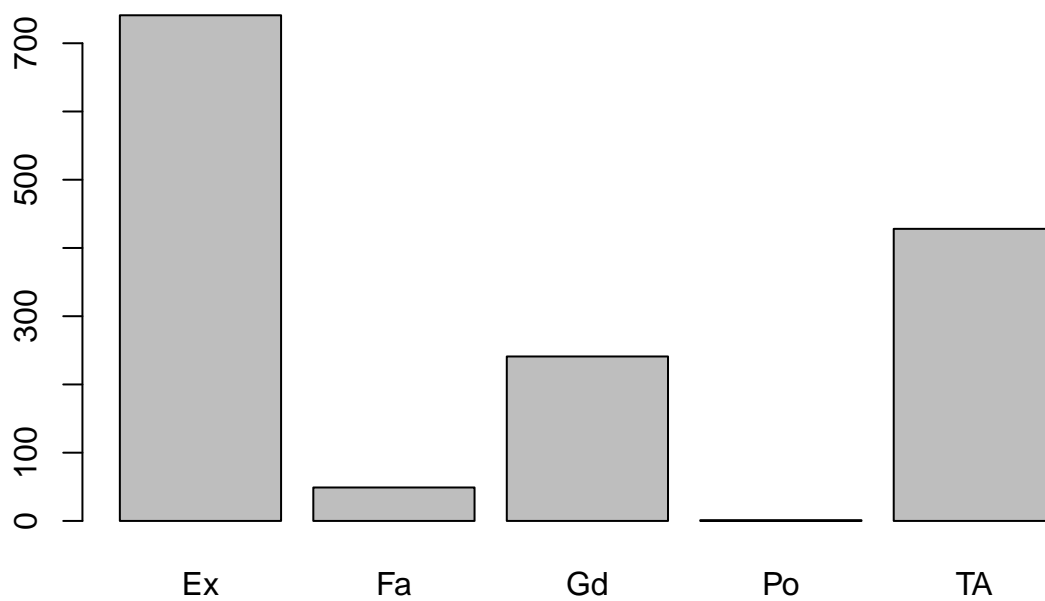


Riscaldamento con gas ad aria calda

HeatingQC

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
Ex	741	0.507534246575343
Fa	49	0.0335616438356164
Gd	241	0.165068493150685
Po	1	0.000684931506849315
TA	428	0.293150684931507

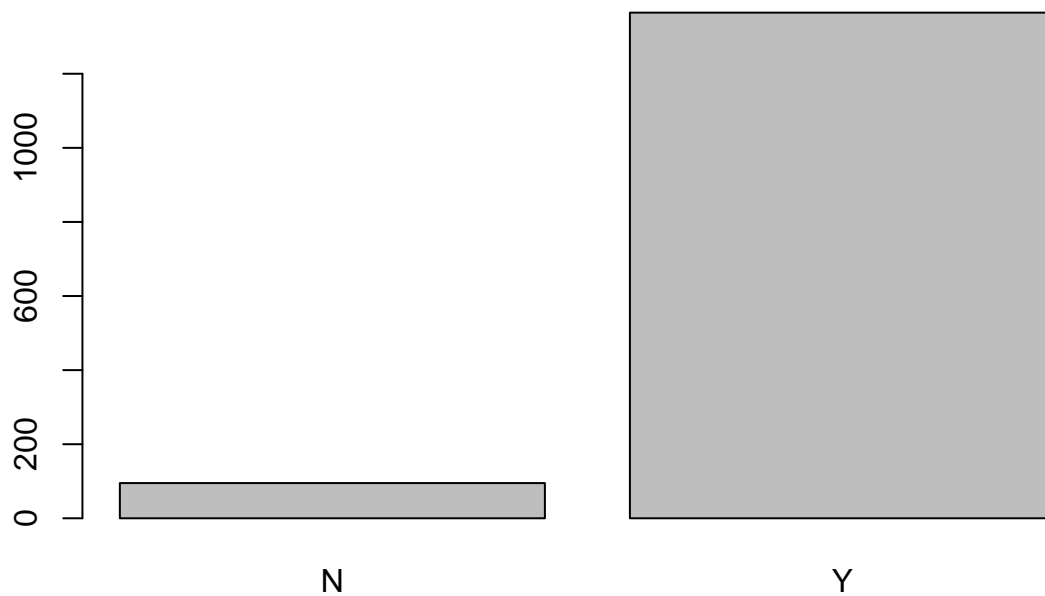


Qualità riscaldamento eccellente, seguita da nella media e da buona

CentralAir

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
N	95	0.0650684931506849
Y	1365	0.934931506849315

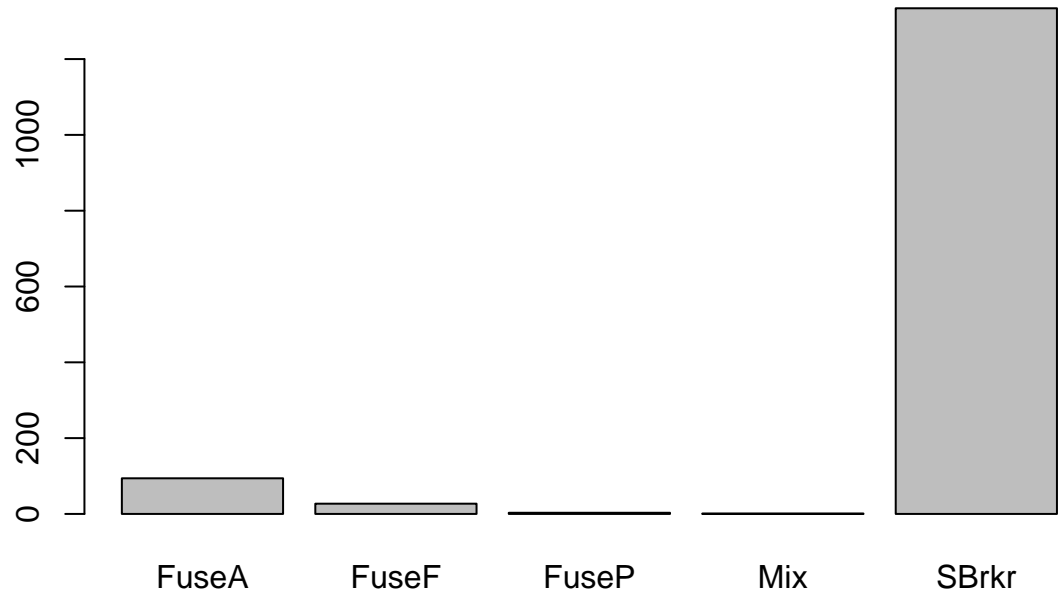


Riscaldamento centralizzato presente

Electrical

Numero di NA: 1

Level	Abs. Freq	Rel. Freq
FuseA	94	0.0644276901987663
FuseF	27	0.0185058259081563
FuseP	3	0.00205620287868403
Mix	1	0.000685400959561343
SBrkr	1334	0.914324880054832

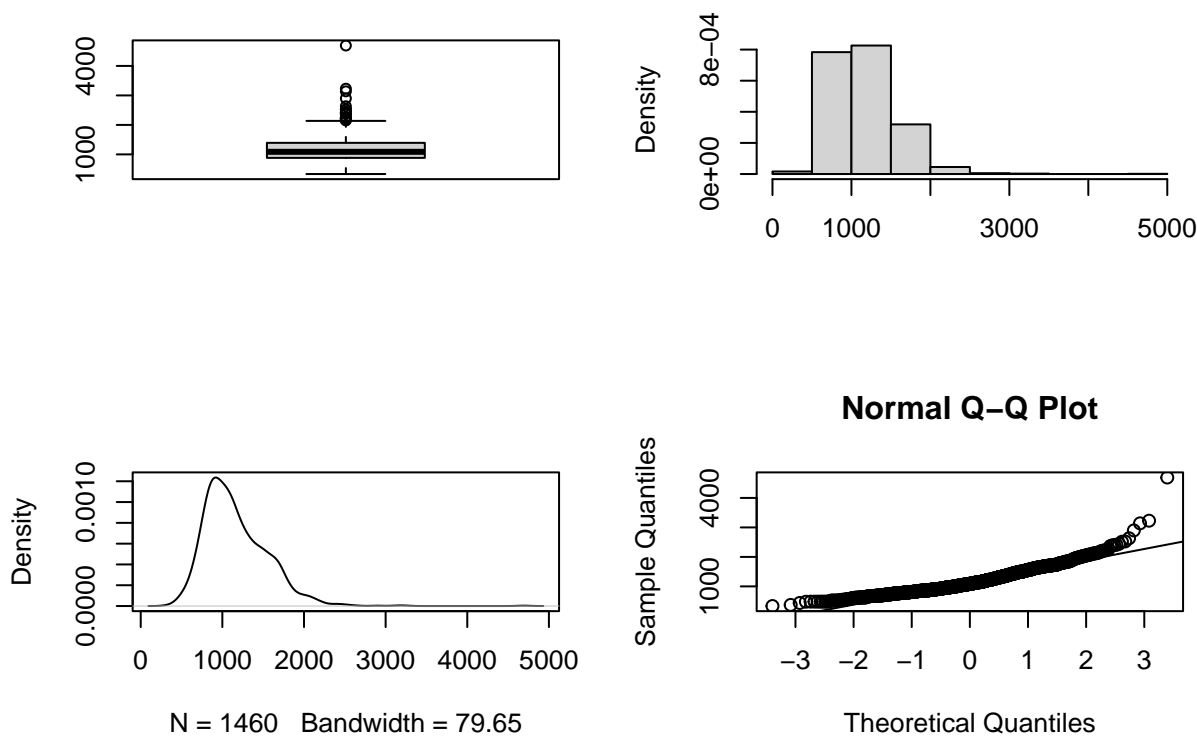


Interruttori elettrici standard

X1stFlrSF

Numero di NA: 0

Stat	Value
Min.	334.000
1st Qu.	882.000
Median	1087.000
Mean	1162.627
3rd Qu.	1391.250
Max.	4692.000

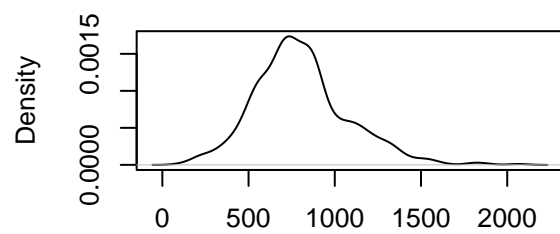
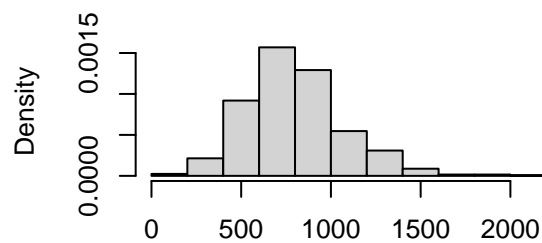
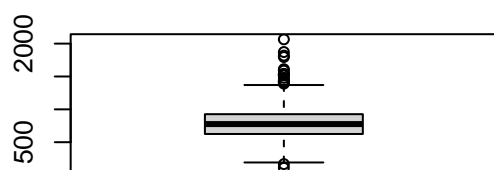


X2ndFlrSF

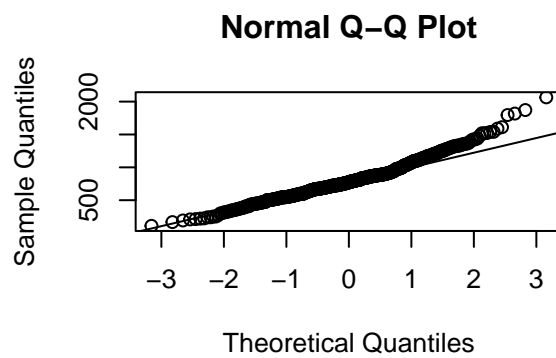
Numero di NA: 0

Numero di zeri rimossi: 829

Stat	Value
Min.	110.0000
1st Qu.	625.0000
Median	776.0000
Mean	802.8669
3rd Qu.	926.5000
Max.	2065.0000



N = 631 Bandwidth = 55.77

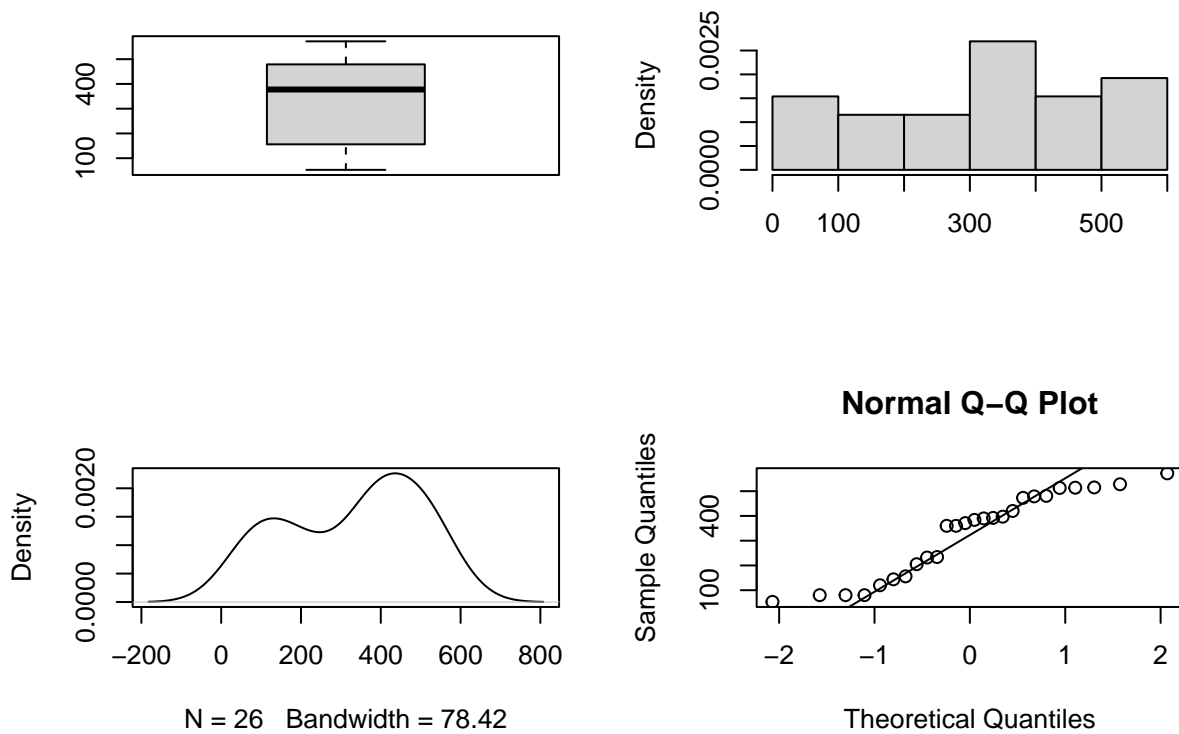


LowQualFinSF

Numero di NA: 0

Numero di zeri rimossi: 1434

Stat	Value
Min.	53.0000
1st Qu.	168.2500
Median	377.5000
Mean	328.1923
3rd Qu.	477.5000
Max.	572.0000

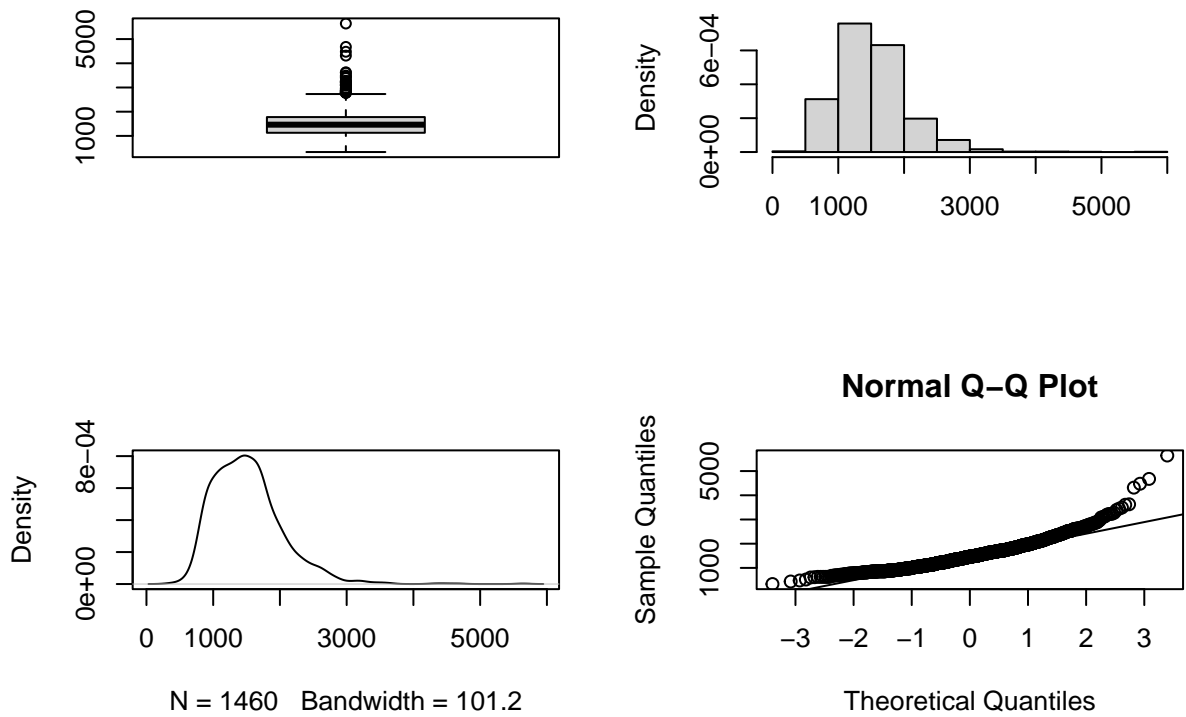


La maggior parte delle case non hanno finiture di bassa qualità. In quelle che c'è la superficie varia principalmente tra 168 e 477 piedi quadrati. Notiamo inoltre che la distribuzione è bimodale.

GrLivArea

Numero di NA: 0

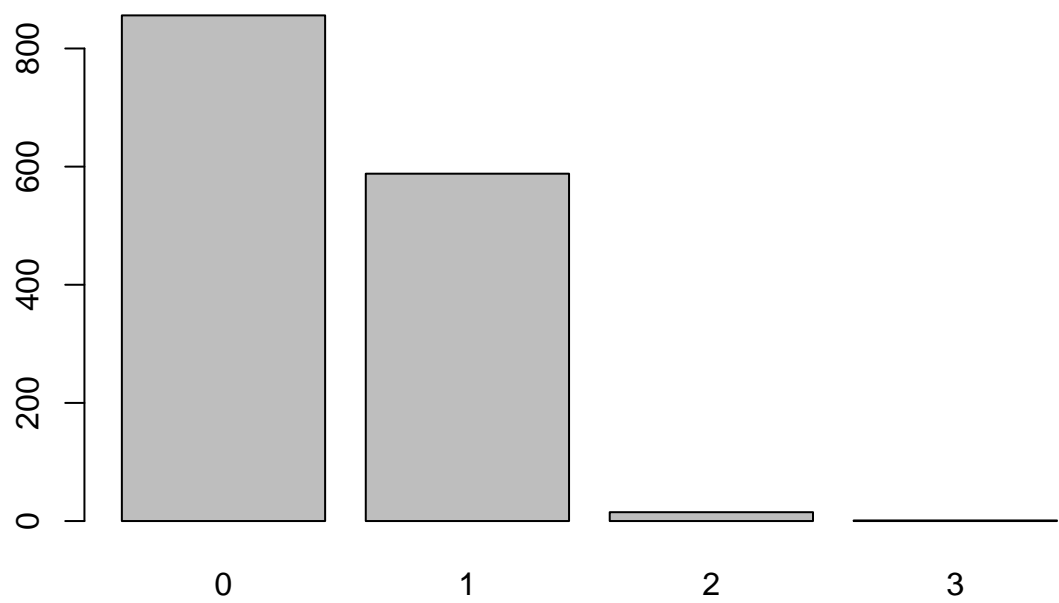
Stat	Value
Min.	334.000
1st Qu.	1129.500
Median	1464.000
Mean	1515.464
3rd Qu.	1776.750
Max.	5642.000



BsmtFullBath

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
0	856	0.586301369863014
1	588	0.402739726027397
2	15	0.0102739726027397
3	1	0.000684931506849315

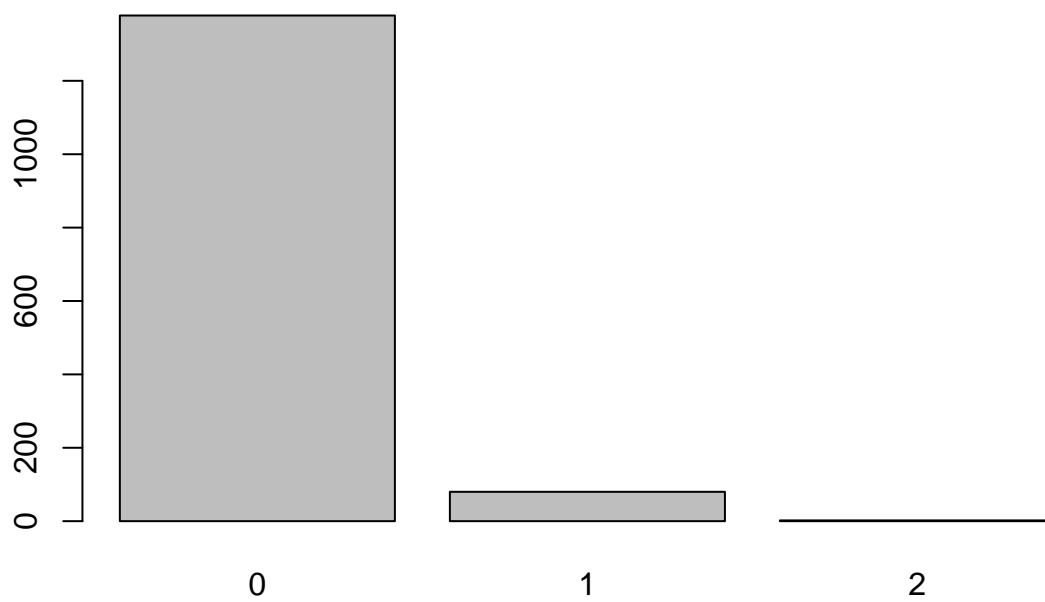


Le case non hanno bagni completi nel seminterrato o ne hanno uno solo

BsmtHalfBath

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
0	1378	0.943835616438356
1	80	0.0547945205479452
2	2	0.00136986301369863

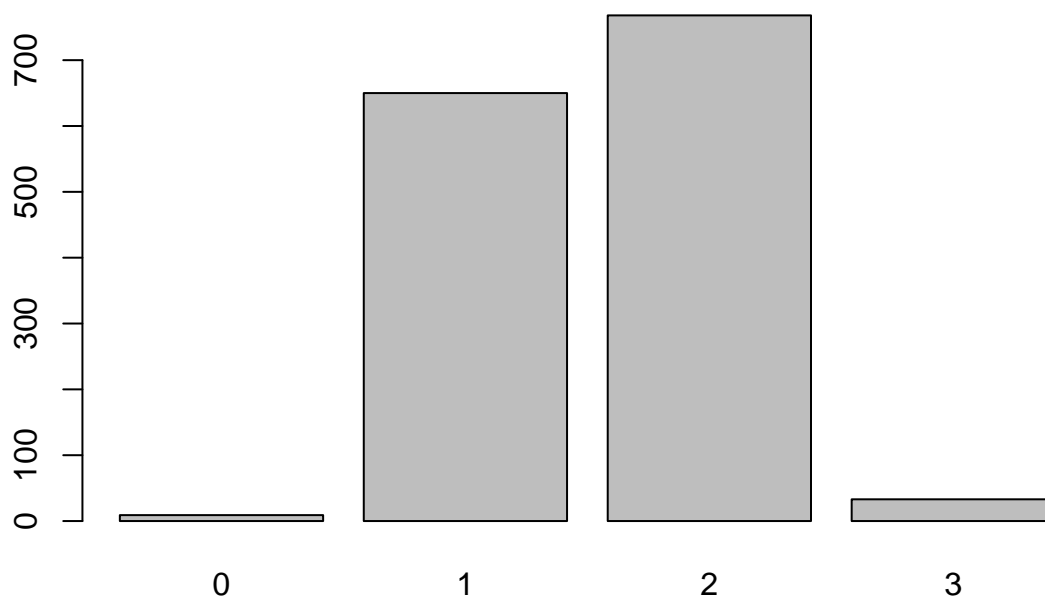


Non hanno mezzi bagni nel seminterrato

FullBath

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
0	9	0.00616438356164384
1	650	0.445205479452055
2	768	0.526027397260274
3	33	0.0226027397260274

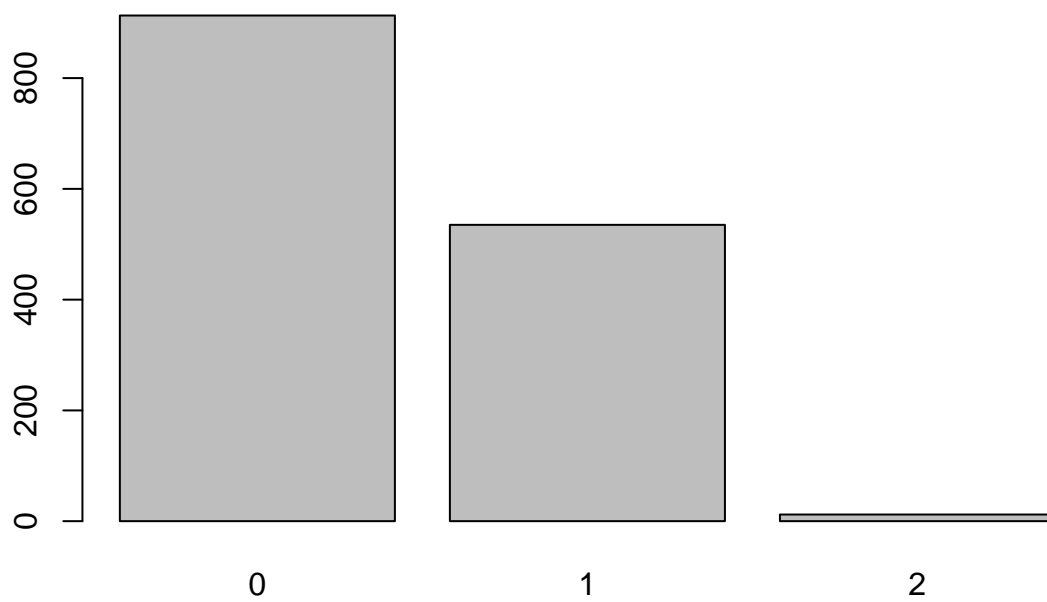


Hanno uno o due bagni completi in casa

HalfBath

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
0	913	0.625342465753425
1	535	0.366438356164384
2	12	0.00821917808219178

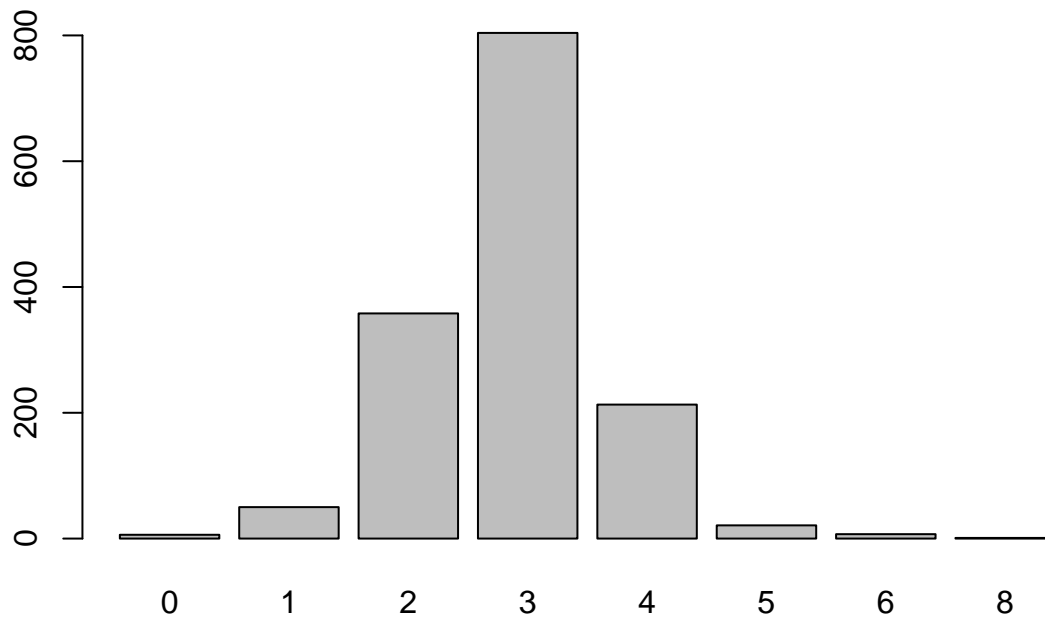


Hanno uno o nessun mezzo bagno in casa

BedroomAbvGr

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
0	6	0.00410958904109589
1	50	0.0342465753424658
2	358	0.245205479452055
3	804	0.550684931506849
4	213	0.145890410958904
5	21	0.0143835616438356
6	7	0.00479452054794521
8	1	0.000684931506849315

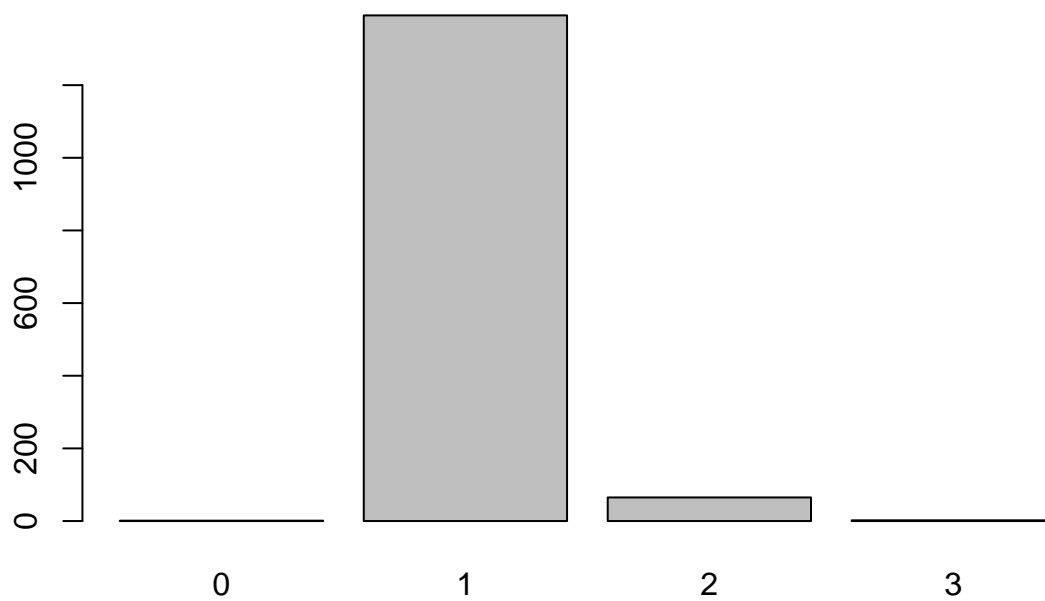


Hanno generalmente tra 2 e 4 stanze da letto

KitchenAbvGr

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
0	1	0.000684931506849315
1	1392	0.953424657534247
2	65	0.0445205479452055
3	2	0.00136986301369863

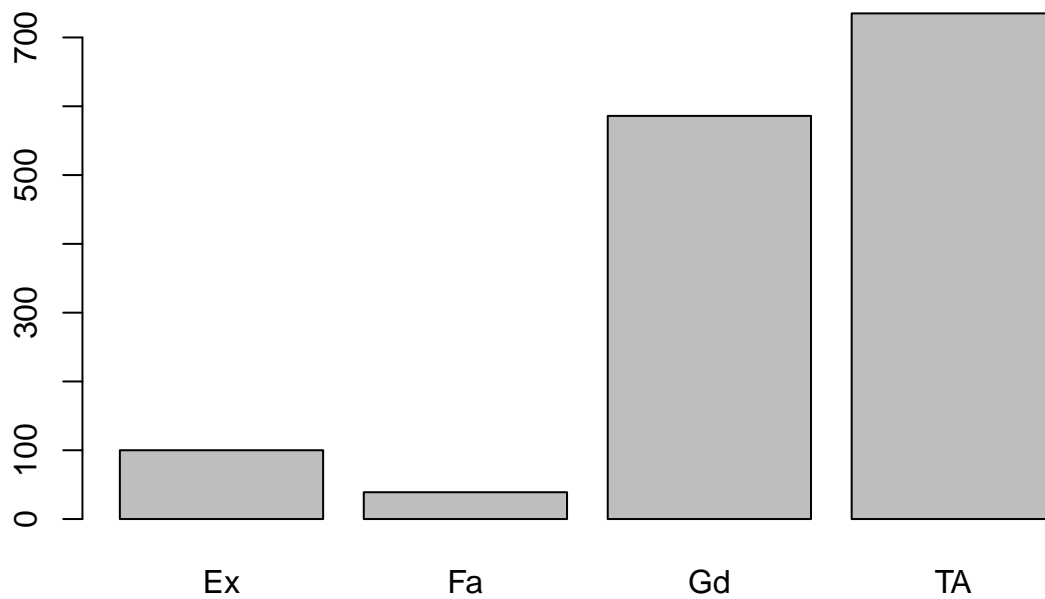


Le case hanno solitamente una sola cucina

KitchenQual

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
Ex	100	0.0684931506849315
Fa	39	0.0267123287671233
Gd	586	0.401369863013699
TA	735	0.503424657534247

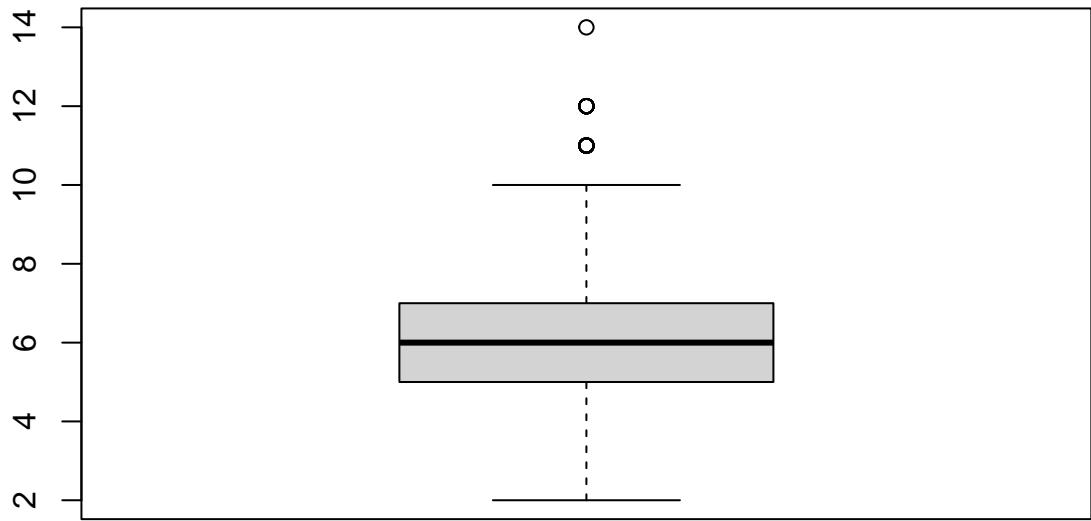


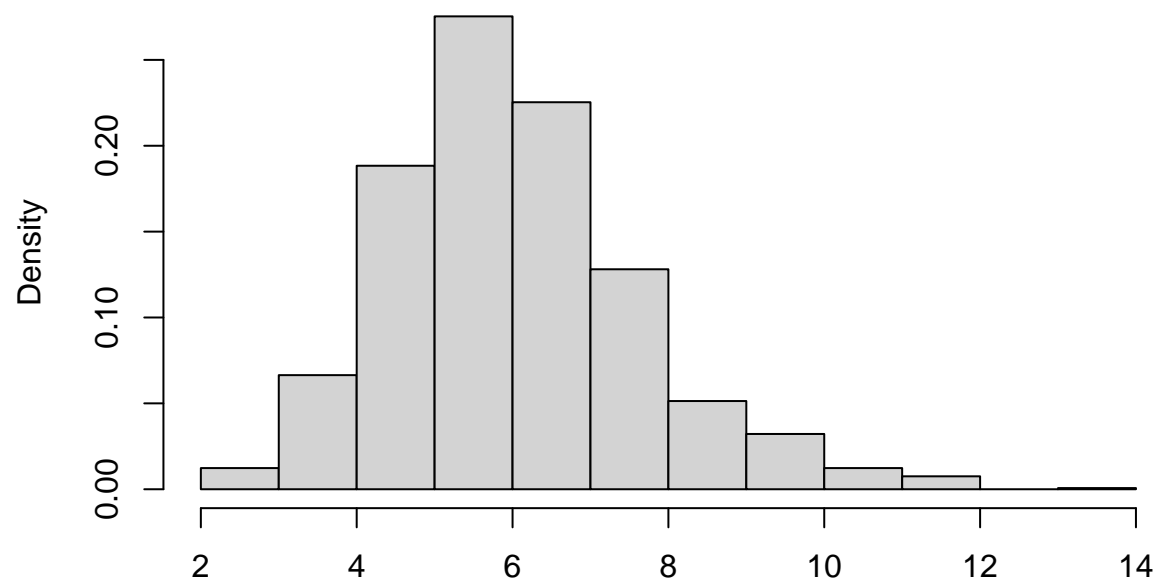
Qualità della cucina nella media o buona

TotRmsAbvGrd

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
2	1	0.000684931506849315
3	17	0.0116438356164384
4	97	0.0664383561643836
5	275	0.188356164383562
6	402	0.275342465753425
7	329	0.225342465753425
8	187	0.128082191780822
9	75	0.0513698630136986
10	47	0.0321917808219178
11	18	0.0123287671232877
12	11	0.00753424657534247
14	1	0.000684931506849315

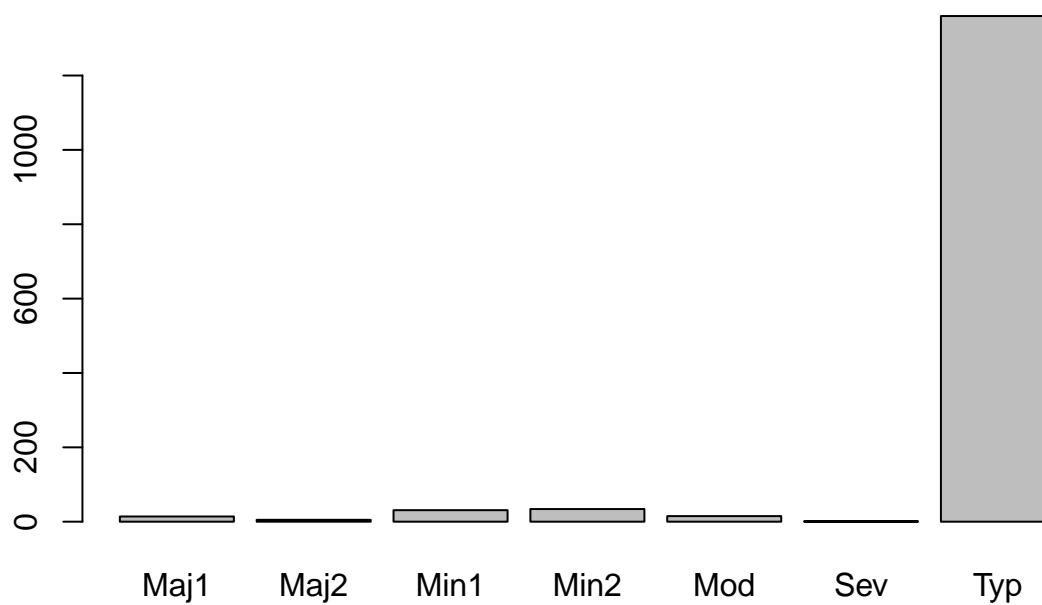




Functional

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
Maj1	14	0.00958904109589041
Maj2	5	0.00342465753424658
Min1	31	0.0212328767123288
Min2	34	0.0232876712328767
Mod	15	0.0102739726027397
Sev	1	0.000684931506849315
Typ	1360	0.931506849315068

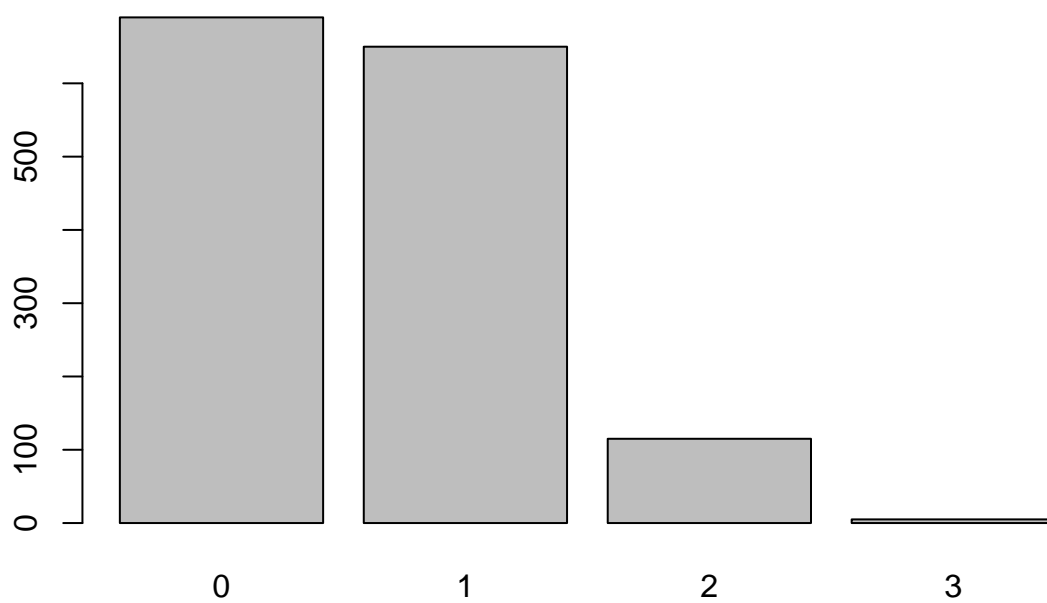


La casa è funzionale non danneggiata

Fireplaces

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
0	690	0.472602739726027
1	650	0.445205479452055
2	115	0.0787671232876712
3	5	0.00342465753424658

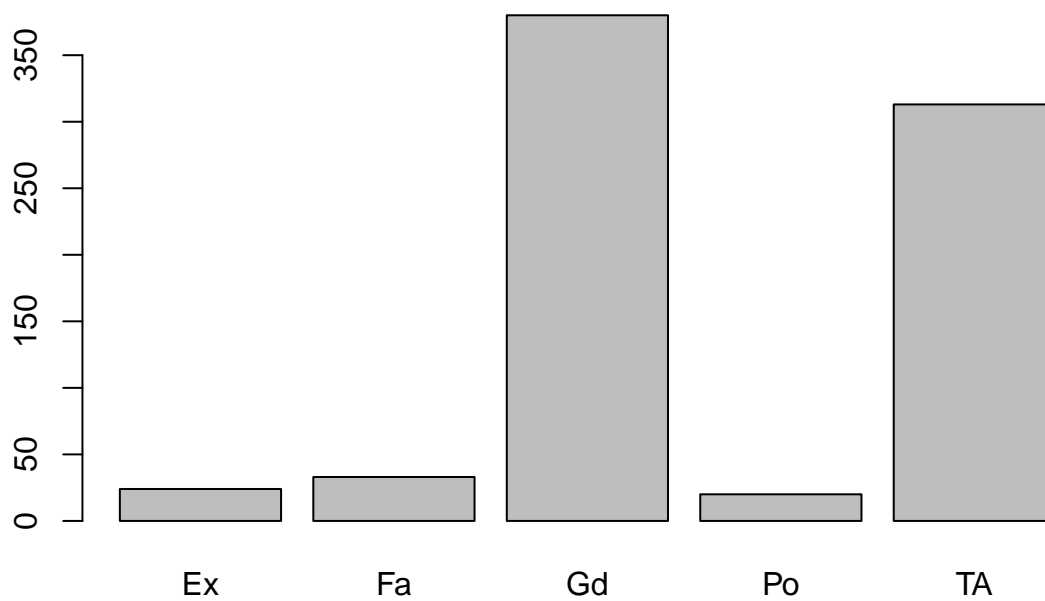


Raramente c'è più di un camino in casa

FireplaceQu

Numero di NA: 690

Level	Abs. Freq	Rel. Freq
Ex	24	0.0311688311688312
Fa	33	0.0428571428571429
Gd	380	0.493506493506494
Po	20	0.025974025974026
TA	313	0.406493506493507

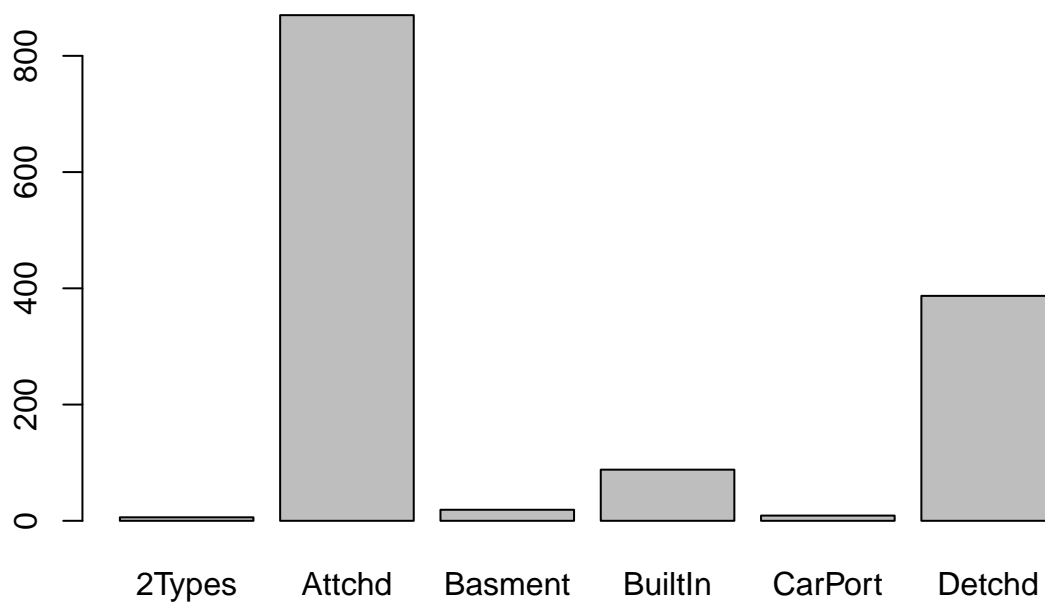


Qualità dei camini, buona o nella media

GarageType

Numero di NA: 81

Level	Abs. Freq	Rel. Freq
2Types	6	0.00435097897026831
Attchd	870	0.630891950688905
Basment	19	0.0137781000725163
BuiltIn	88	0.0638143582306019
CarPort	9	0.00652646845540247
Detchd	387	0.280638143582306

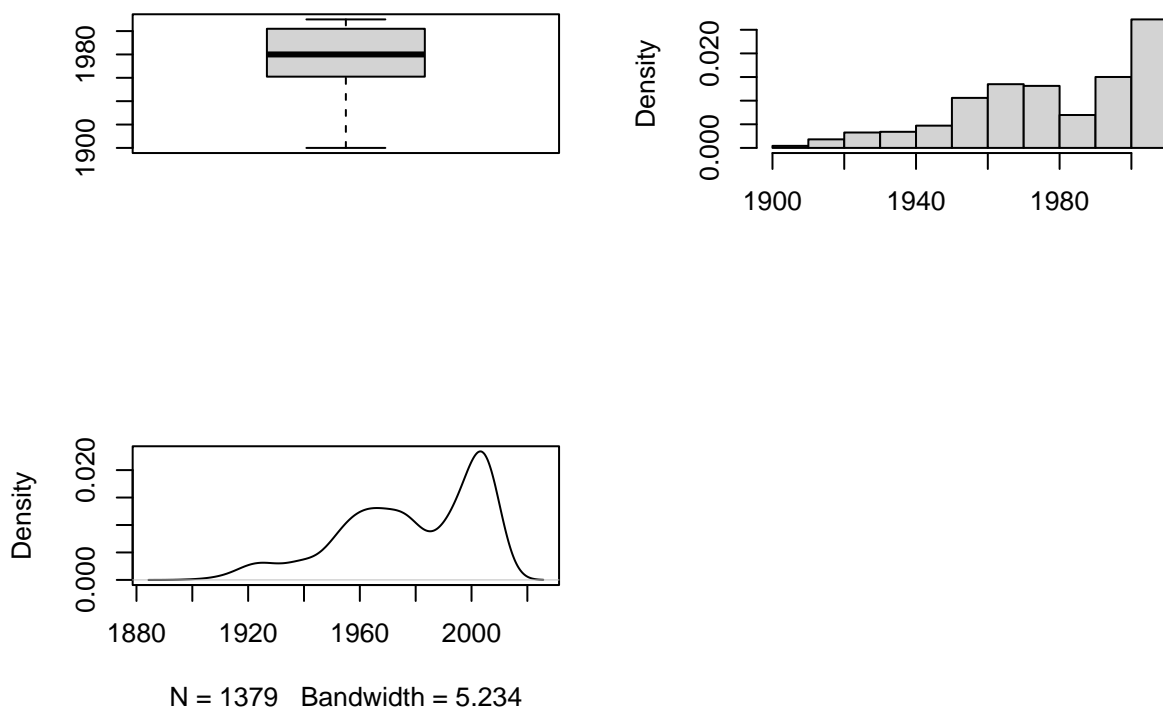


Garage annessi o indipendenti

GarageYrBlt

Numero di NA: 81

Stat	Value
Min.	1900.000
1st Qu.	1961.000
Median	1980.000
Mean	1978.506
3rd Qu.	2002.000
Max.	2010.000
NA's	81.000

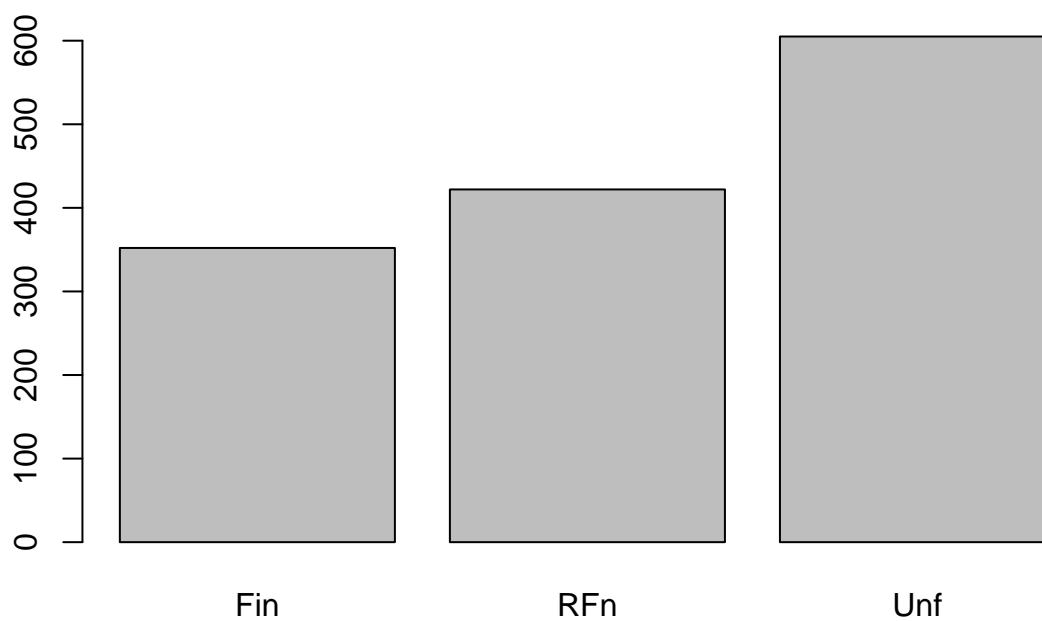


Si evidenzia anche negli anni di costruzione dei garage una distribuzione a più mode. I garage sono soprattutto di costruzione recente con un picco di costruzioni negli anni 2000.

GarageFinish

Numero di NA: 81

Level	Abs. Freq	Rel. Freq
Fin	352	0.255257432922408
RFn	422	0.306018854242204
Unf	605	0.438723712835388

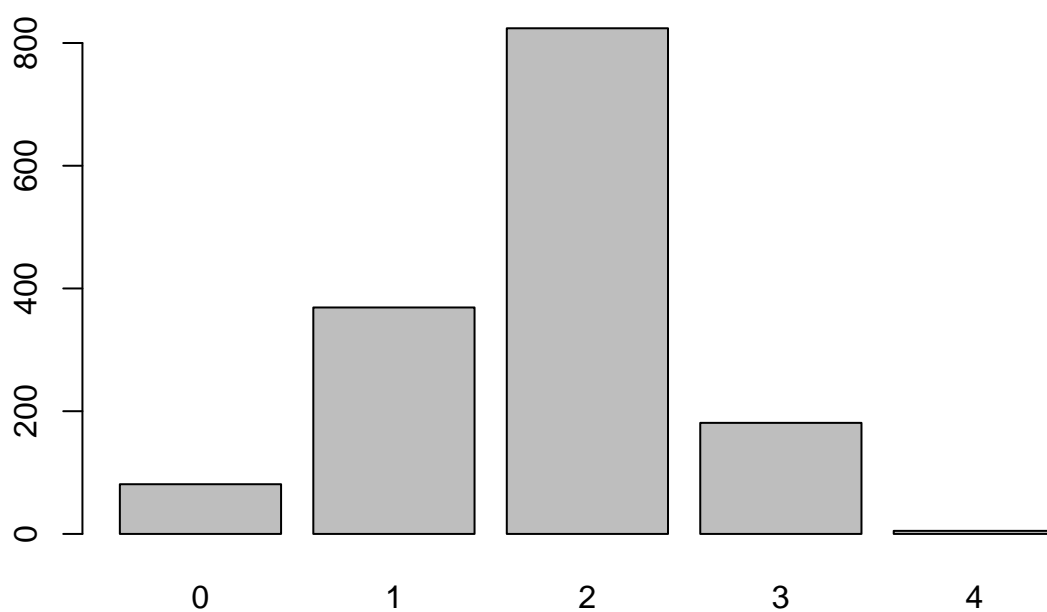


Molti dei garage non sono finiti, altri sono finiti in uno stato grezzo, altri ancora sono finiti

GarageCars

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
0	81	0.0554794520547945
1	369	0.252739726027397
2	824	0.564383561643836
3	181	0.123972602739726
4	5	0.00342465753424658



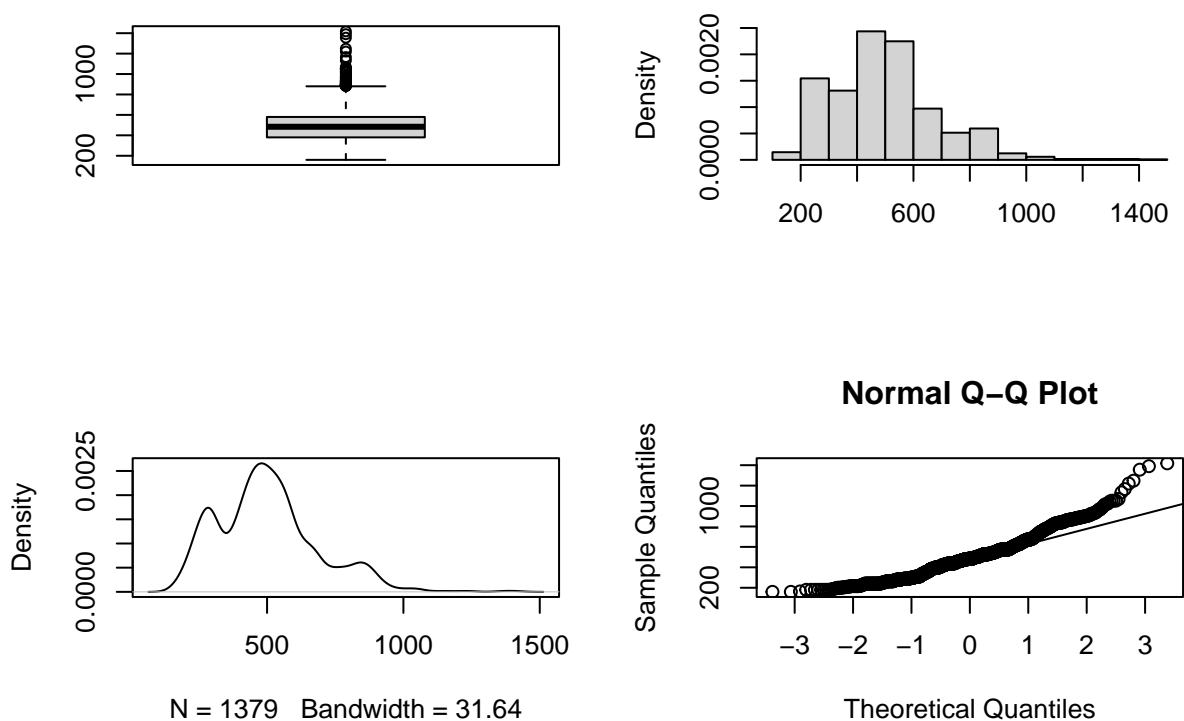
La capienza dei garage contiene 2 auto o meno frequentemente 1 sola

GarageArea

Numero di NA: 0

Numero di zeri rimossi: 81

Stat	Value
Min.	160.0000
1st Qu.	380.0000
Median	484.0000
Mean	500.7621
3rd Qu.	580.0000
Max.	1418.0000

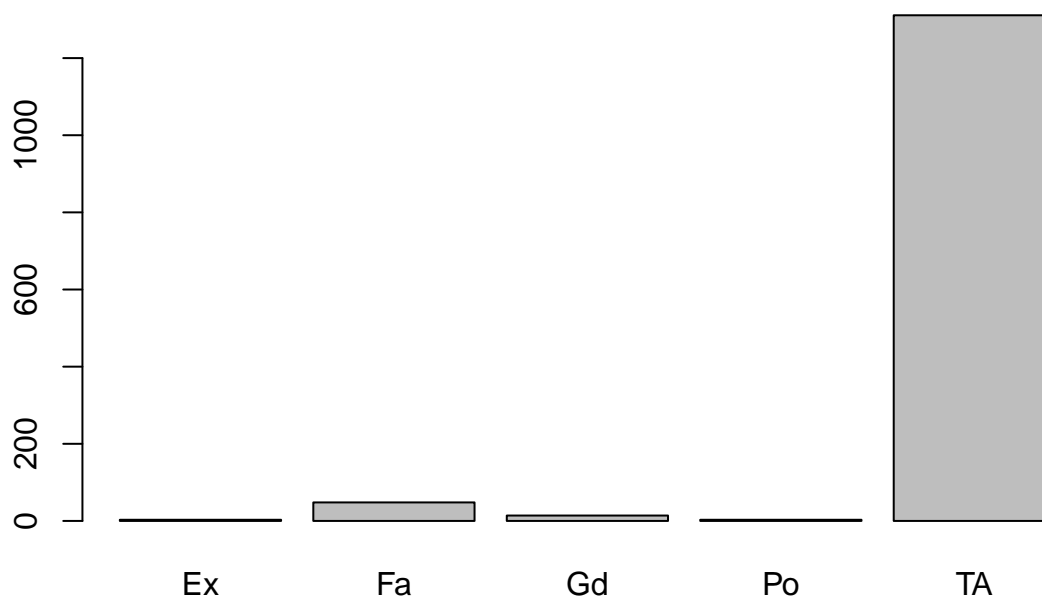


La distribuzione dell'area del garage ha più mode, le mode probabilmente corrispondono con la capienza in numero di auto per cui il garage è stato costruito.

GarageQual

Numero di NA: 81

Level	Abs. Freq	Rel. Freq
Ex	3	0.00217548948513416
Fa	48	0.0348078317621465
Gd	14	0.0101522842639594
Po	3	0.00217548948513416
TA	1311	0.950688905003626

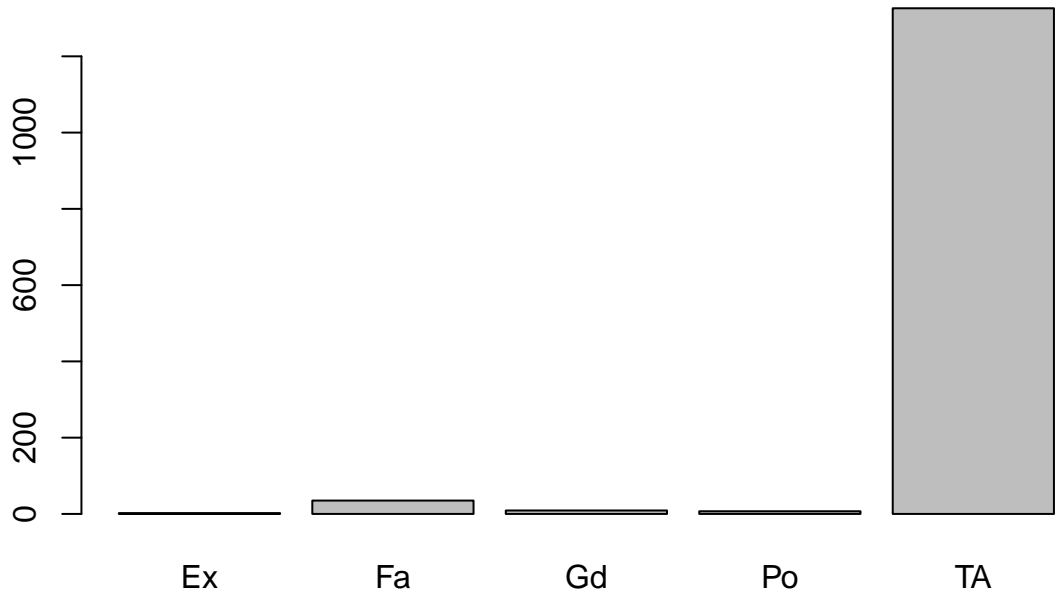


Qualità del garage nella media

GarageCond

Numero di NA: 81

Level	Abs. Freq	Rel. Freq
Ex	2	0.00145032632342277
Fa	35	0.0253807106598985
Gd	9	0.00652646845540247
Po	7	0.0050761421319797
TA	1326	0.961566352429297

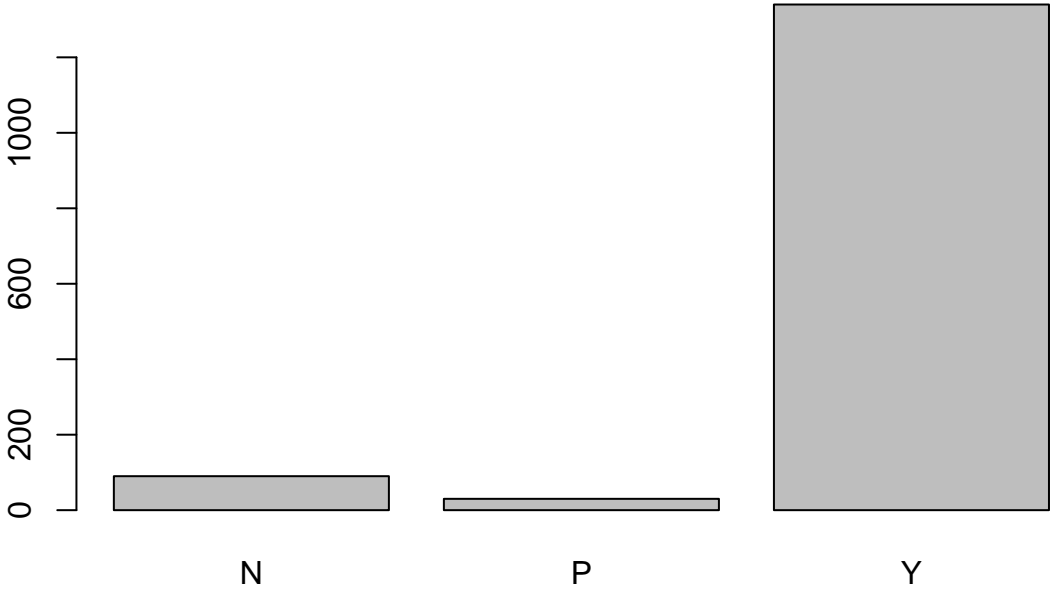


Condizioni del garage nella media

PavedDrive

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
N	90	0.0616438356164384
P	30	0.0205479452054795
Y	1340	0.917808219178082



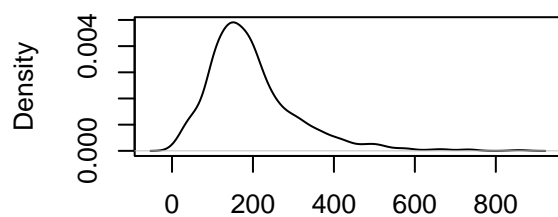
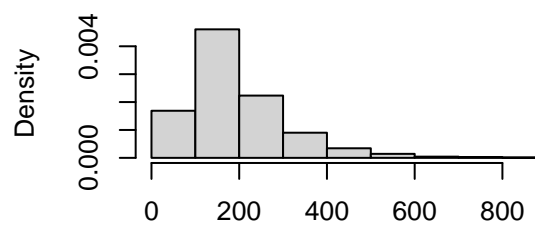
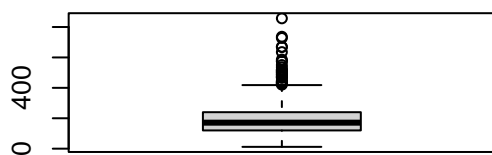
Strada del garage asfaltata

WoodDeckSF

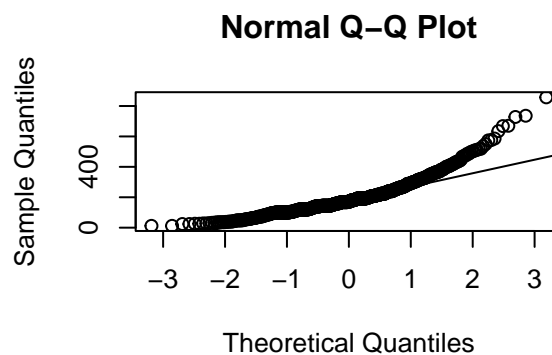
Numero di NA: 0

Numero di zeri rimossi: 761

Stat	Value
Min.	12.0000
1st Qu.	120.0000
Median	171.0000
Mean	196.8484
3rd Qu.	240.0000
Max.	857.0000



N = 699 Bandwidth = 21.75

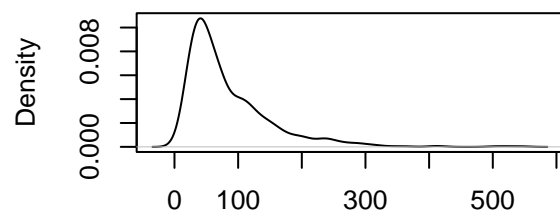
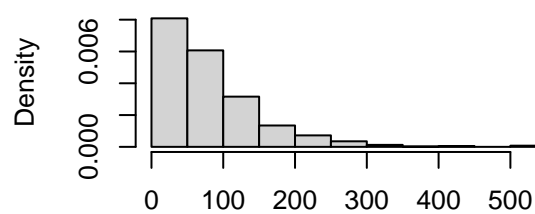
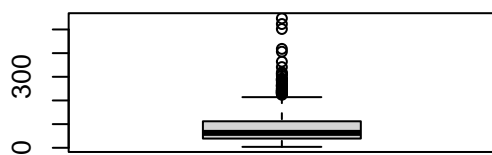


OpenPorchSF

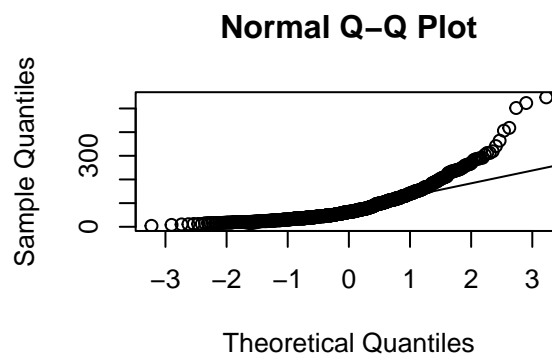
Numero di NA: 0

Numero di zeri rimossi: 656

Stat	Value
Min.	4.00000
1st Qu.	39.00000
Median	63.00000
Mean	84.73134
3rd Qu.	112.00000
Max.	547.00000



N = 804 Bandwidth = 12.86

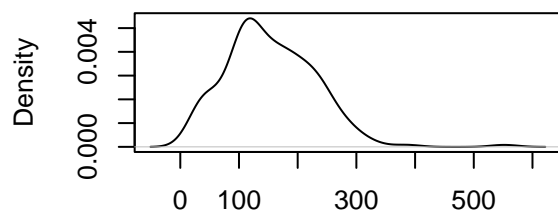
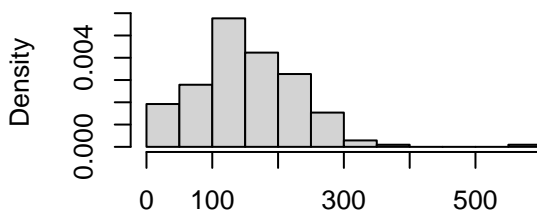
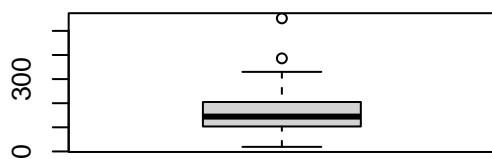


EnclosedPorch

Numero di NA: 0

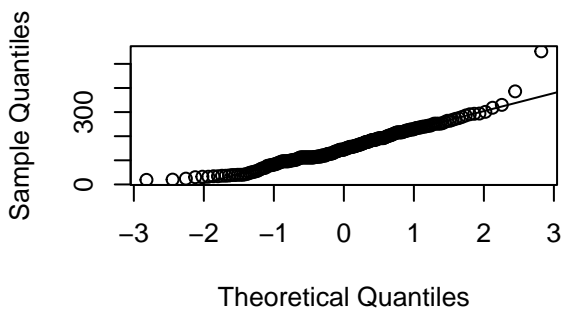
Numero di zeri rimossi: 1252

Stat	Value
Min.	19.000
1st Qu.	104.250
Median	144.500
Mean	154.101
3rd Qu.	205.000
Max.	552.000



N = 208 Bandwidth = 23.27

Normal Q-Q Plot

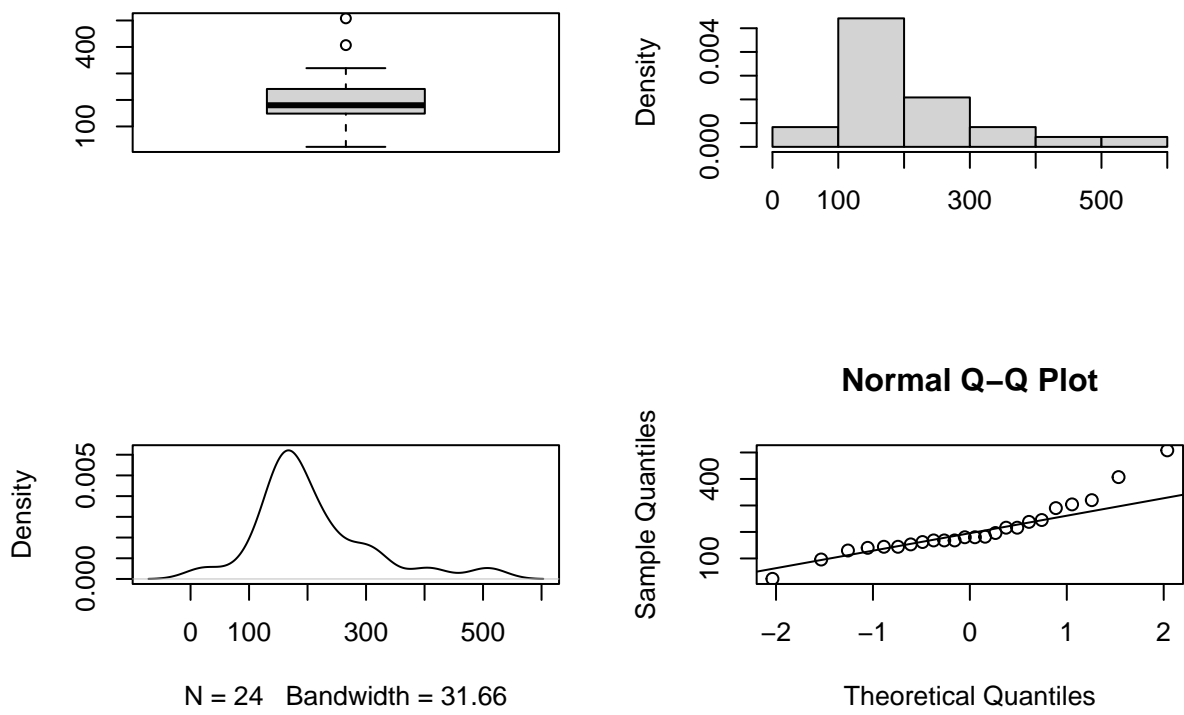


X3SsnPorch

Numero di NA: 0

Numero di zeri rimossi: 1436

Stat	Value
Min.	23.0000
1st Qu.	150.7500
Median	180.0000
Mean	207.4167
3rd Qu.	239.7500
Max.	508.0000

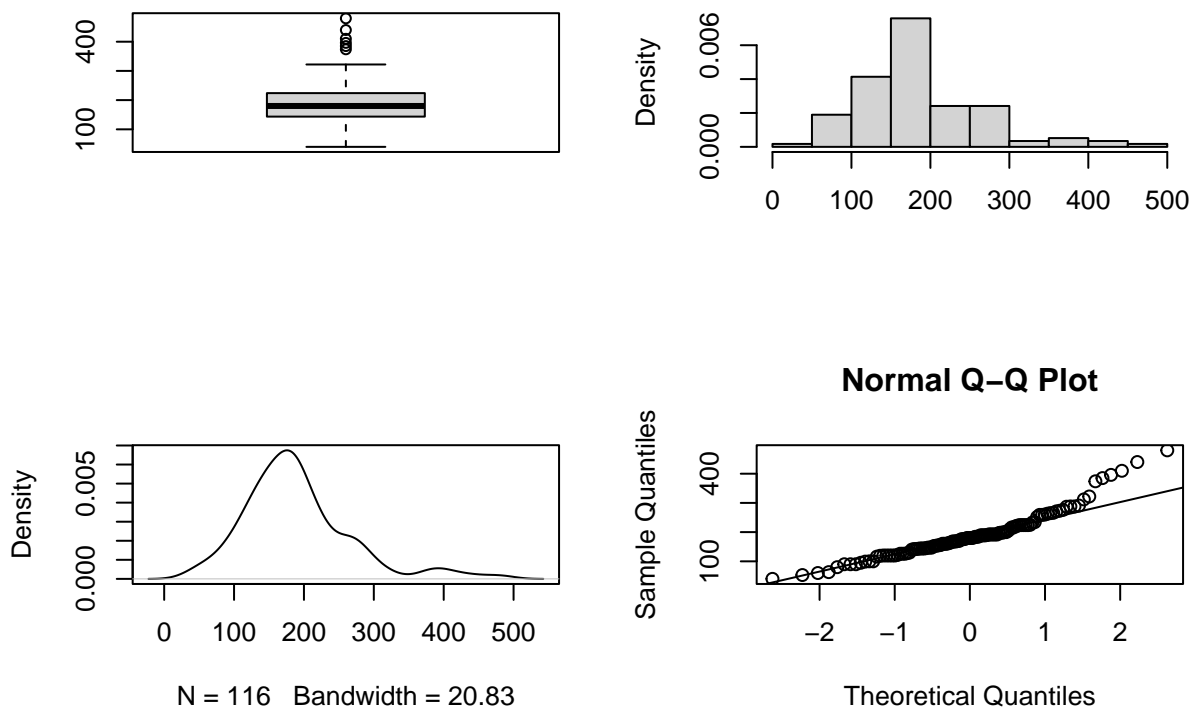


ScreenPorch

Numero di NA: 0

Numero di zeri rimossi: 1344

Stat	Value
Min.	40.0000
1st Qu.	143.7500
Median	180.0000
Mean	189.5603
3rd Qu.	224.0000
Max.	480.0000

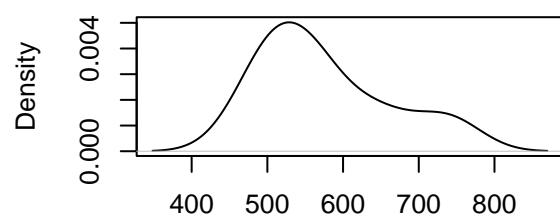
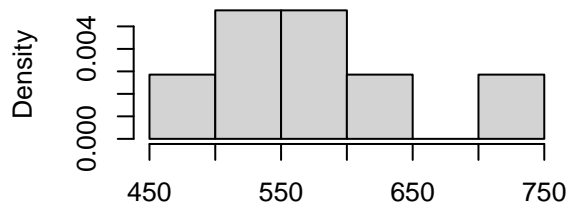
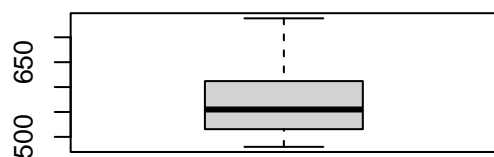


PoolArea

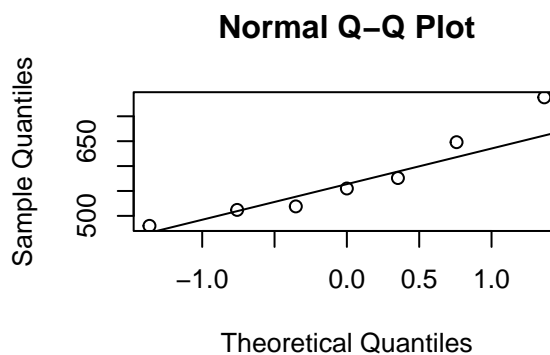
Numero di NA: 0

Numero di zeri rimossi: 1453

Stat	Value
Min.	480.0000
1st Qu.	515.5000
Median	555.0000
Mean	575.4286
3rd Qu.	612.0000
Max.	738.0000



N = 7 Bandwidth = 43.92

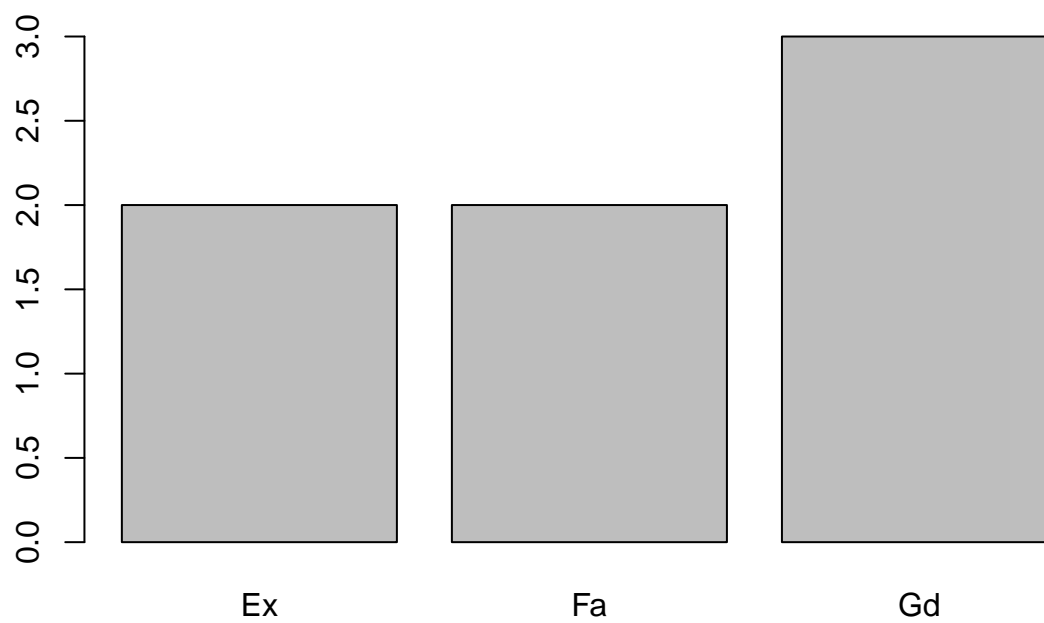


Ci sono solo 7 piscine, troppo pochi valori.

PoolQC

Numero di NA: 1453

Level	Abs. Freq	Rel. Freq
Ex	2	0.285714285714286
Fa	2	0.285714285714286
Gd	3	0.428571428571429

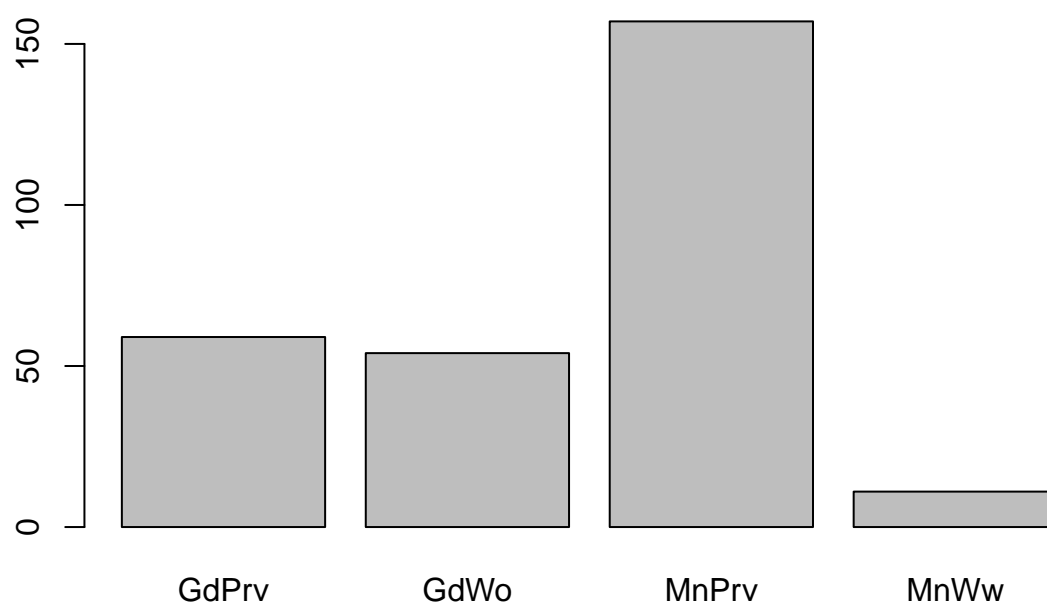


Qualità piscina equamente distribuita tra eccellente, buona e decente

Fence

Numero di NA: 1179

Level	Abs. Freq	Rel. Freq
GdPrv	59	0.209964412811388
GdWo	54	0.192170818505338
MnPrv	157	0.558718861209964
MnWw	11	0.0391459074733096

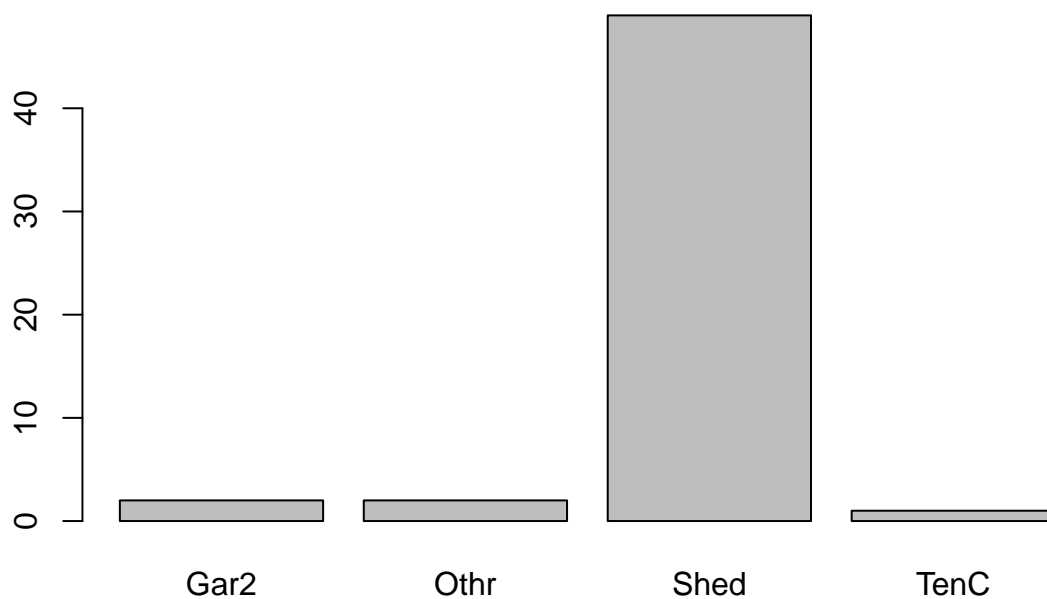


Il confine offre un minimo di privacy

MiscFeature

Numero di NA: 1406

Level	Abs. Freq	Rel. Freq
Gar2	2	0.037037037037037
Othr	2	0.037037037037037
Shed	49	0.907407407407407
TenC	1	0.0185185185185185



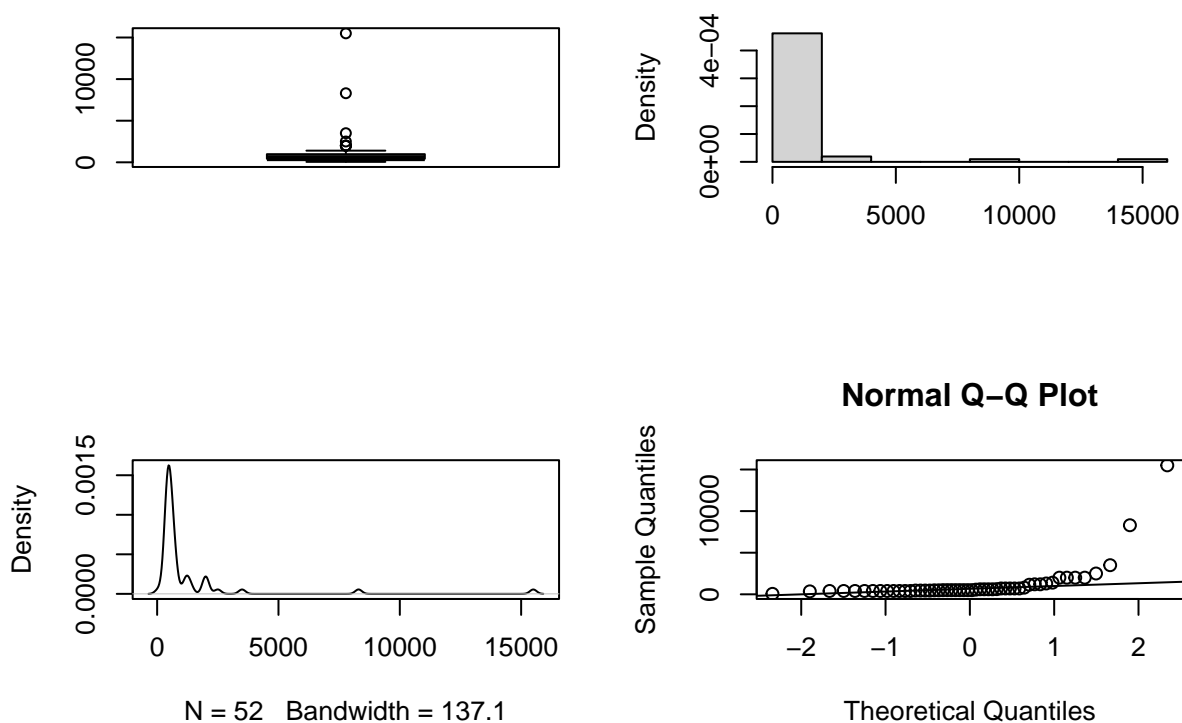
Generalmente non sono presenti altre strutture, avvolte sono presenti capannoni

MiscVal

Numero di NA: 0

Numero di zeri rimossi: 1408

Stat	Value
Min.	54.000
1st Qu.	437.500
Median	500.000
Mean	1221.038
3rd Qu.	887.500
Max.	15500.000

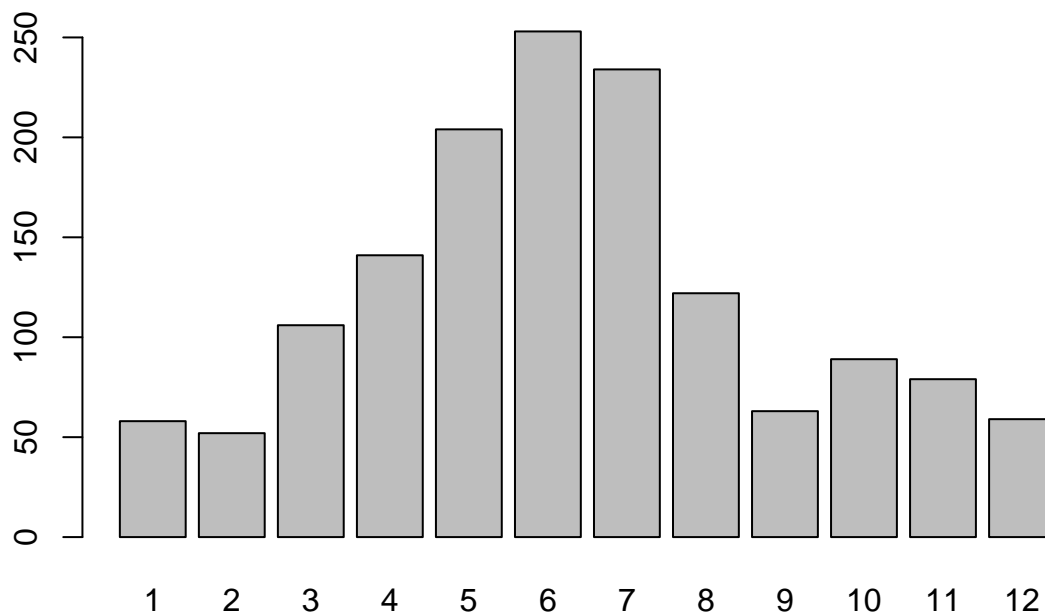


Non analizzo la variabile MiscVal in quanto le dimensioni di diverse caratteristiche della casa non sono comparabili (non posso comparare la grandezza di un campo da tennis con la grandezza di un ascensore)

MoSold

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
1	58	0.0397260273972603
2	52	0.0356164383561644
3	106	0.0726027397260274
4	141	0.0965753424657534
5	204	0.13972602739726
6	253	0.173287671232877
7	234	0.16027397260274
8	122	0.0835616438356164
9	63	0.0431506849315069
10	89	0.060958904109589
11	79	0.0541095890410959
12	59	0.0404109589041096

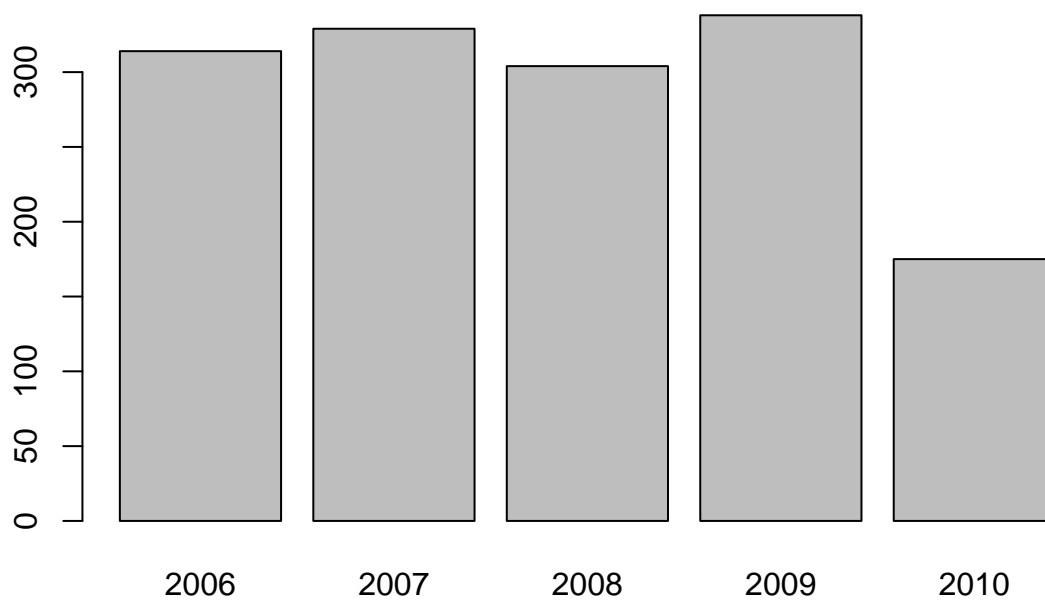


Vengono vendute più case nei mesi primaverili ed estivi rispetto alle stagioni autunnali e invernali

YrSold

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
2006	314	0.215068493150685
2007	329	0.225342465753425
2008	304	0.208219178082192
2009	338	0.231506849315068
2010	175	0.11986301369863

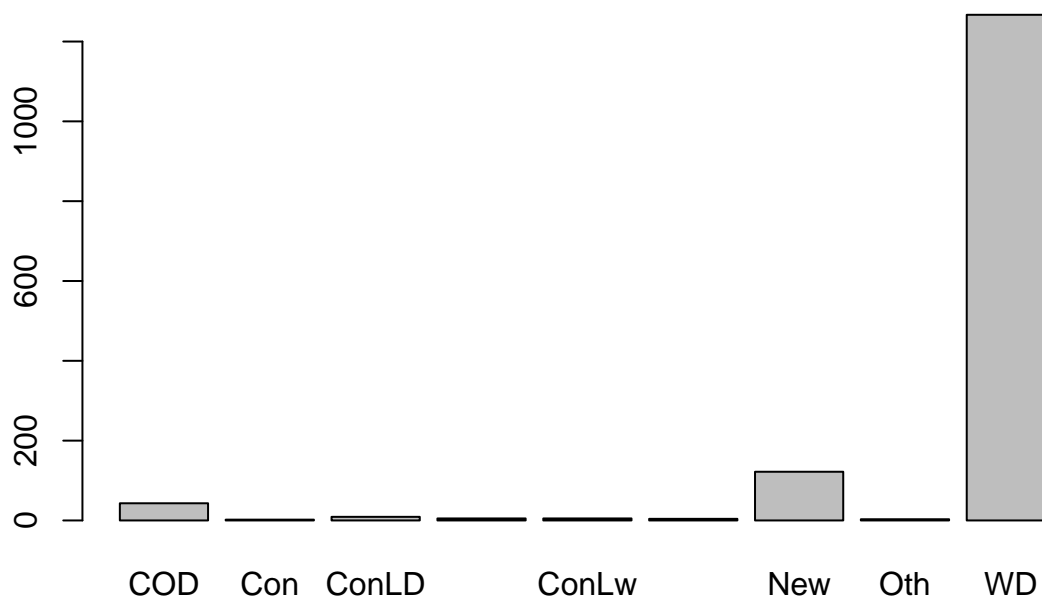


Si evidenzia una decrescita nelle vendite nel 2010, forse dovuta al termine delle osservazioni nel dataset

SaleType

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
COD	43	0.0294520547945205
Con	2	0.00136986301369863
ConLD	9	0.00616438356164384
ConLI	5	0.00342465753424658
ConLw	5	0.00342465753424658
CWD	4	0.00273972602739726
New	122	0.0835616438356164
Oth	3	0.00205479452054795
WD	1267	0.867808219178082

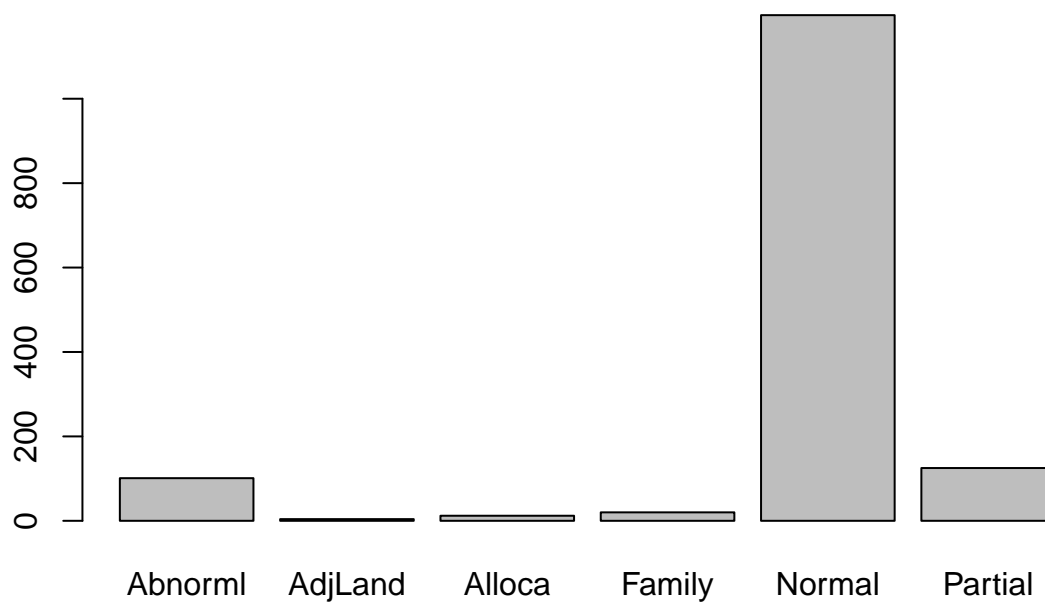


Vendita con atto di garanzia convenzionale

SaleCondition

Numero di NA: 0

Level	Abs. Freq	Rel. Freq
Abnorml	101	0.0691780821917808
AdjLand	4	0.00273972602739726
Alloca	12	0.00821917808219178
Family	20	0.0136986301369863
Normal	1198	0.820547945205479
Partial	125	0.0856164383561644

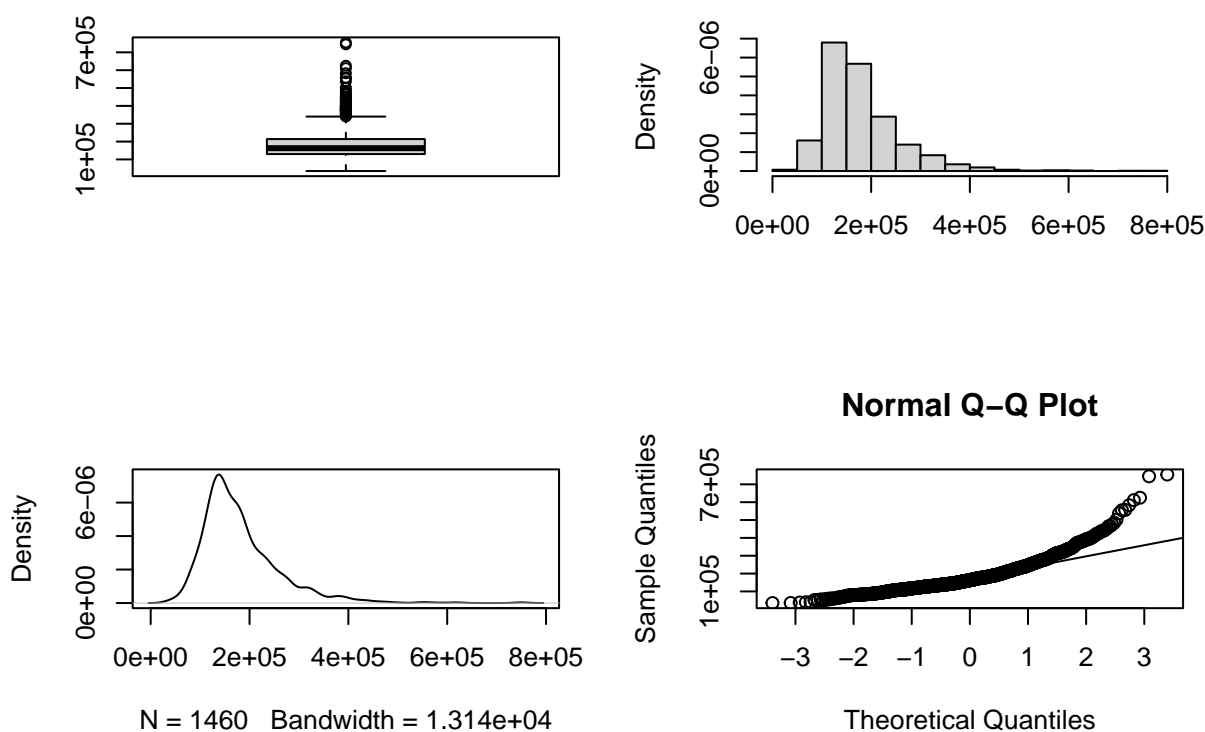


Condizioni di vendita generalmente normali

SalePrice

Numero di NA: 0

Stat	Value
Min.	34900.0
1st Qu.	129975.0
Median	163000.0
Mean	180921.2
3rd Qu.	214000.0
Max.	755000.0



I prezzi di vendita delle case sono nel range tra 34900 e 755000 con una media di circa 181000. La distribuzione è asimmetrica: non ci sono molte case a prezzi bassi, ce ne sono molte invece che superano il terzo quartile.

ANALISI BIVARIATA DEL DATASET “House prices”

```
# CONSIDERANDO LA VARIABILE TARGET(SalePrice):
```

```
HousePrices$SalePrice # prendiamo come variabile target il prezzo della proprieta' in vendita
```

```
##      [1] 208500 181500 223500 140000 250000 143000 307000 200000 129900 118000
##     [11] 129500 345000 144000 279500 157000 132000 149000  90000 159000 139000
##     [21] 325300 139400 230000 129900 154000 256300 134800 306000 207500  68500
##     [31]  40000 149350 179900 165500 277500 309000 145000 153000 109000  82000
##     [41] 160000 170000 144000 130250 141000 319900 239686 249700 113000 127000
```

##	[51]	177000	114500	110000	385000	130000	180500	172500	196500	438780	124900
##	[61]	158000	101000	202500	140000	219500	317000	180000	226000	80000	225000
##	[71]	244000	129500	185000	144900	107400	91000	135750	127000	136500	110000
##	[81]	193500	153500	245000	126500	168500	260000	174000	164500	85000	123600
##	[91]	109900	98600	163500	133900	204750	185000	214000	94750	83000	128950
##	[101]	205000	178000	118964	198900	169500	250000	100000	115000	115000	190000
##	[111]	136900	180000	383970	217000	259500	176000	139000	155000	320000	163990
##	[121]	180000	100000	136000	153900	181000	84500	128000	87000	155000	150000
##	[131]	226000	244000	150750	220000	180000	174000	143000	171000	230000	231500
##	[141]	115000	260000	166000	204000	125000	130000	105000	222500	141000	115000
##	[151]	122000	372402	190000	235000	125000	79000	109500	269500	254900	320000
##	[161]	162500	412500	220000	103200	152000	127500	190000	325624	183500	228000
##	[171]	128500	215000	239000	163000	184000	243000	211000	172500	501837	100000
##	[181]	177000	200100	120000	200000	127000	475000	173000	135000	153337	286000
##	[191]	315000	184000	192000	130000	127000	148500	311872	235000	104000	274900
##	[201]	140000	171500	112000	149000	110000	180500	143900	141000	277000	145000
##	[211]	98000	186000	252678	156000	161750	134450	210000	107000	311500	167240
##	[221]	204900	200000	179900	97000	386250	112000	290000	106000	125000	192500
##	[231]	148000	403000	94500	128200	216500	89500	185500	194500	318000	113000
##	[241]	262500	110500	79000	120000	205000	241500	137000	140000	180000	277000
##	[251]	76500	235000	173000	158000	145000	230000	207500	220000	231500	97000
##	[261]	176000	276000	151000	130000	73000	175500	185000	179500	120500	148000
##	[271]	266000	241500	290000	139000	124500	205000	201000	141000	415298	192000
##	[281]	228500	185000	207500	244600	179200	164700	159000	88000	122000	153575
##	[291]	233230	135900	131000	235000	167000	142500	152000	239000	175000	158500
##	[301]	157000	267000	205000	149900	295000	305900	225000	89500	82500	360000
##	[311]	165600	132000	119900	375000	178000	188500	260000	270000	260000	187500
##	[321]	342643	354000	301000	126175	242000	87000	324000	145250	214500	78000
##	[331]	119000	139000	284000	207000	192000	228950	377426	214000	202500	155000
##	[341]	202900	82000	87500	266000	85000	140200	151500	157500	154000	437154
##	[351]	318061	190000	95000	105900	140000	177500	173000	134000	130000	280000
##	[361]	156000	145000	198500	118000	190000	147000	159000	165000	132000	162000
##	[371]	172400	134432	125000	123000	219500	61000	148000	340000	394432	179000
##	[381]	127000	187750	213500	76000	240000	192000	81000	125000	191000	426000
##	[391]	119000	215000	106500	100000	109000	129000	123000	169500	67000	241000
##	[401]	245500	164990	108000	258000	168000	150000	115000	177000	280000	339750
##	[411]	60000	145000	222000	115000	228000	181134	149500	239000	126000	142000
##	[421]	206300	215000	113000	315000	139000	135000	275000	109008	195400	175000
##	[431]	85400	79900	122500	181000	81000	212000	116000	119000	90350	110000
##	[441]	555000	118000	162900	172500	210000	127500	190000	199900	119500	120000
##	[451]	110000	280000	204000	210000	188000	175500	98000	256000	161000	110000
##	[461]	263435	155000	62383	188700	124000	178740	167000	146500	250000	187000
##	[471]	212000	190000	148000	440000	251000	132500	208900	380000	297000	89471
##	[481]	326000	374000	155000	164000	132500	147000	156000	175000	160000	86000
##	[491]	115000	133000	172785	155000	91300	34900	430000	184000	130000	120000
##	[501]	113000	226700	140000	289000	147000	124500	215000	208300	161000	124500
##	[511]	164900	202665	129900	134000	96500	402861	158000	265000	211000	234000
##	[521]	106250	150000	159000	184750	315750	176000	132000	446261	86000	200624
##	[531]	175000	128000	107500	39300	178000	107500	188000	111250	158000	272000
##	[541]	315000	248000	213250	133000	179665	229000	210000	129500	125000	263000
##	[551]	140000	112500	255500	108000	284000	113000	141000	108000	175000	234000
##	[561]	121500	170000	108000	185000	268000	128000	325000	214000	316600	135960
##	[571]	142600	120000	224500	170000	139000	118500	145000	164500	146000	131500
##	[581]	181900	253293	118500	325000	133000	369900	130000	137000	143000	79500

##	[591]	185900	451950	138000	140000	110000	319000	114504	194201	217500	151000
##	[601]	275000	141000	220000	151000	221000	205000	152000	225000	359100	118500
##	[611]	313000	148000	261500	147000	75500	137500	183200	105500	314813	305000
##	[621]	67000	240000	135000	168500	165150	160000	139900	153000	135000	168500
##	[631]	124000	209500	82500	139400	144000	200000	60000	93000	85000	264561
##	[641]	274000	226000	345000	152000	370878	143250	98300	155000	155000	84500
##	[651]	205950	108000	191000	135000	350000	88000	145500	149000	97500	167000
##	[661]	197900	402000	110000	137500	423000	230500	129000	193500	168000	137500
##	[671]	173500	103600	165000	257500	140000	148500	87000	109500	372500	128500
##	[681]	143000	159434	173000	285000	221000	207500	227875	148800	392000	194700
##	[691]	141000	755000	335000	108480	141500	176000	89000	123500	138500	196000
##	[701]	312500	140000	361919	140000	213000	55000	302000	254000	179540	109900
##	[711]	52000	102776	189000	129000	130500	165000	159500	157000	341000	128500
##	[721]	275000	143000	124500	135000	320000	120500	222000	194500	110000	103000
##	[731]	236500	187500	222500	131400	108000	163000	93500	239900	179000	190000
##	[741]	132000	142000	179000	175000	180000	299800	236000	265979	260400	98000
##	[751]	96500	162000	217000	275500	156000	172500	212000	158900	179400	290000
##	[761]	127500	100000	215200	337000	270000	264132	196500	160000	216837	538000
##	[771]	134900	102000	107000	114500	395000	162000	221500	142500	144000	135000
##	[781]	176000	175900	187100	165500	128000	161500	139000	233000	107900	187500
##	[791]	160200	146800	269790	225000	194500	171000	143500	110000	485000	175000
##	[801]	200000	109900	189000	582933	118000	227680	135500	223500	159950	106000
##	[811]	181000	144500	55993	157900	116000	224900	137000	271000	155000	224000
##	[821]	183000	93000	225000	139500	232600	385000	109500	189000	185000	147400
##	[831]	166000	151000	237000	167000	139950	128000	153500	100000	144000	130500
##	[841]	140000	157500	174900	141000	153900	171000	213000	133500	240000	187000
##	[851]	131500	215000	164000	158000	170000	127000	147000	174000	152000	250000
##	[861]	189950	131500	152000	132500	250580	148500	248900	129000	169000	236000
##	[871]	109500	200500	116000	133000	66500	303477	132250	350000	148000	136500
##	[881]	157000	187500	178000	118500	100000	328900	145000	135500	268000	149500
##	[891]	122900	172500	154500	165000	118858	140000	106500	142953	611657	135000
##	[901]	110000	153000	180000	240000	125500	128000	255000	250000	131000	174000
##	[911]	154300	143500	88000	145000	173733	75000	35311	135000	238000	176500
##	[921]	201000	145900	169990	193000	207500	175000	285000	176000	236500	222000
##	[931]	201000	117500	320000	190000	242000	79900	184900	253000	239799	244400
##	[941]	150900	214000	150000	143000	137500	124900	143000	270000	192500	197500
##	[951]	129000	119900	133900	172000	127500	145000	124000	132000	185000	155000
##	[961]	116500	272000	155000	239000	214900	178900	160000	135000	37900	140000
##	[971]	135000	173000	99500	182000	167500	165000	85500	199900	110000	139000
##	[981]	178400	336000	159895	255900	126000	125000	117000	395192	195000	197000
##	[991]	348000	168000	187000	173900	337500	121600	136500	185000	91000	206000
##	[1001]	82000	86000	232000	136905	181000	149900	163500	88000	240000	102000
##	[1011]	135000	100000	165000	85000	119200	227000	203000	187500	160000	213490
##	[1021]	176000	194000	87000	191000	287000	112500	167500	293077	105000	118000
##	[1031]	160000	197000	310000	230000	119750	84000	315500	287000	97000	80000
##	[1041]	155000	173000	196000	262280	278000	139600	556581	145000	115000	84900
##	[1051]	176485	200141	165000	144500	255000	180000	185850	248000	335000	220000
##	[1061]	213500	81000	90000	110500	154000	328000	178000	167900	151400	135000
##	[1071]	135000	154000	91500	159500	194000	219500	170000	138800	155900	126000
##	[1081]	145000	133000	192000	160000	187500	147000	83500	252000	137500	197000
##	[1091]	92900	160000	136500	146000	129000	176432	127000	170000	128000	157000
##	[1101]	60000	119500	135000	159500	106000	325000	179900	274725	181000	280000
##	[1111]	188000	205000	129900	134500	117000	318000	184100	130000	140000	133700
##	[1121]	118400	212900	112000	118000	163900	115000	174000	259000	215000	140000

```
## [1131] 135000 93500 117500 239500 169000 102000 119000 94000 196000 144000
## [1141] 139000 197500 424870 80000 80000 149000 180000 174500 116900 143000
## [1151] 124000 149900 230000 120500 201800 218000 179900 230000 235128 185000
## [1161] 146000 224000 129000 108959 194000 233170 245350 173000 235000 625000
## [1171] 171000 163000 171900 200500 239000 285000 119500 115000 154900 93000
## [1181] 250000 392500 745000 120000 186700 104900 95000 262000 195000 189000
## [1191] 168000 174000 125000 165000 158000 176000 219210 144000 178000 148000
## [1201] 116050 197900 117000 213000 153500 271900 107000 200000 140000 290000
## [1211] 189000 164000 113000 145000 134500 125000 112000 229456 80500 91500
## [1221] 115000 134000 143000 137900 184000 145000 214000 147000 367294 127000
## [1231] 190000 132500 101800 142000 130000 138887 175500 195000 142500 265900
## [1241] 224900 248328 170000 465000 230000 178000 186500 169900 129500 119000
## [1251] 244000 171750 130000 294000 165400 127500 301500 99900 190000 151000
## [1261] 181000 128900 161500 180500 181000 183900 122000 378500 381000 144000
## [1271] 260000 185750 137000 177000 139000 137000 162000 197900 237000 68400
## [1281] 227000 180000 150500 139000 169000 132500 143000 190000 278000 281000
## [1291] 180500 119500 107500 162900 115000 138500 155000 140000 160000 154000
## [1301] 225000 177500 290000 232000 130000 325000 202500 138000 147000 179200
## [1311] 335000 203000 302000 333168 119000 206900 295493 208900 275000 111000
## [1321] 156500 72500 190000 82500 147000 55000 79000 130500 256000 176500
## [1331] 227000 132500 100000 125500 125000 167900 135000 52500 200000 128500
## [1341] 123000 155000 228500 177000 155835 108500 262500 283463 215000 122000
## [1351] 200000 171000 134900 410000 235000 170000 110000 149900 177500 315000
## [1361] 189000 260000 104900 156932 144152 216000 193000 127000 144000 232000
## [1371] 105000 165500 274300 466500 250000 239000 91000 117000 83000 167500
## [1381] 58500 237500 157000 112000 105000 125500 250000 136000 377500 131000
## [1391] 235000 124000 123000 163000 246578 281213 160000 137500 138000 137450
## [1401] 120000 193000 193879 282922 105000 275000 133000 112000 125500 215000
## [1411] 230000 140000 90000 257000 207000 175900 122500 340000 124000 223000
## [1421] 179900 127500 136500 274970 144000 142000 271000 140000 119000 182900
## [1431] 192140 143750 64500 186500 160000 174000 120500 394617 149700 197000
## [1441] 191000 149300 310000 121000 179600 129000 157900 240000 112000 92000
## [1451] 136000 287090 145000 84500 185000 175000 210000 266500 142125 147500
```

```
str(HousePrices$SalePrice)
```

```
## int [1:1460] 208500 181500 223500 140000 250000 143000 307000 200000 129900 118000 ...
```

```
# analizziamo la variabile:
```

```
summary(HousePrices$SalePrice)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
```

```
## 34900 129975 163000 180921 214000 755000
```

```
# il prezzo delle proprietà varia dai 34k ai 755k dollari
```

```
# il prezzo medio è di 180k dollari
```

```
# MSSubClass
```

```
# questa variabile identifica la tipologia di abitazione
```

```
str(HousePrices$MSSubClass) # ogni numero è specifico per una sola tipologia
```

```
## int [1:1460] 60 20 60 70 60 50 20 60 50 190 ...
```

```
# se vogliamo vedere come essa influenza il prezzo vendita
```

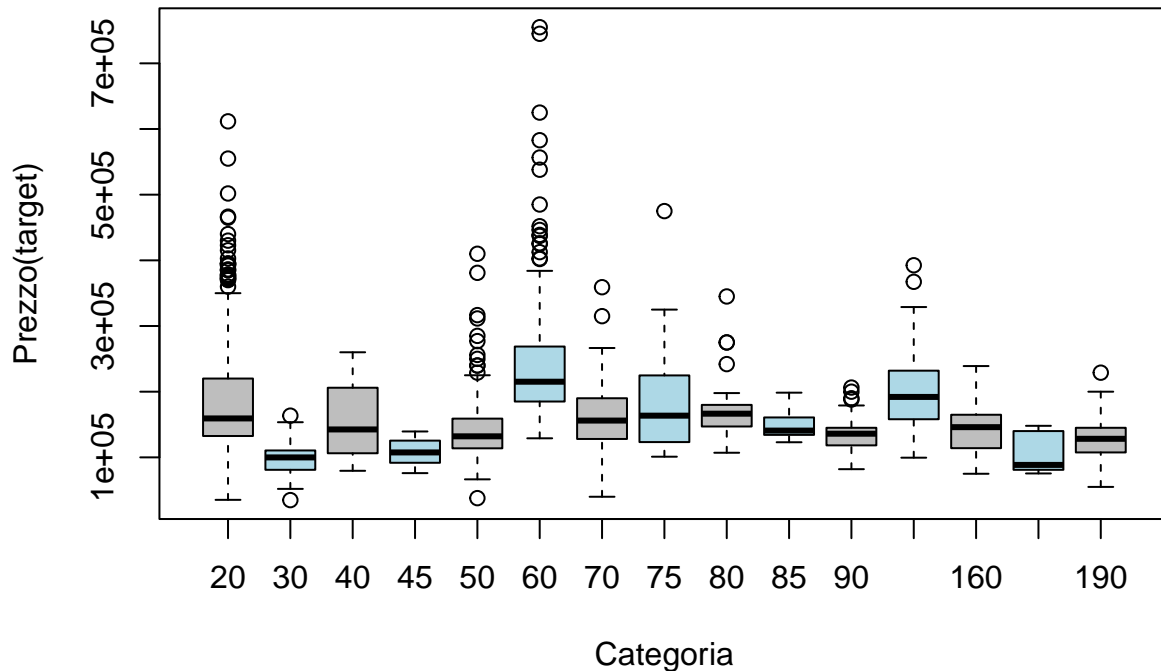
```
# trattando MSSubClass come una variabile categorica:
```

```
boxplot(SalePrice ~ MSSubClass, data = HousePrices,
```



```
main = "Boxplot di SalePrice per ciascun tipo di abitazione",
xlab = "Categoria", ylab = "Prezzo(target)", col = c("grey", "lightblue"))
```

Boxplot di SalePrice per ciascun tipo di abitazione



```
# le proprieta' che raggiungono i prezzi piu' alti sono quelle della categoria 60
# la mediana tende a non essere al centro
# ci sono outliers
# la varianza per certe categorie e' piu' grande per altre meno
```

```
# GarageType
# variabile categorica che indica dove si trova il garage
str(HousePrices$GarageType)
```

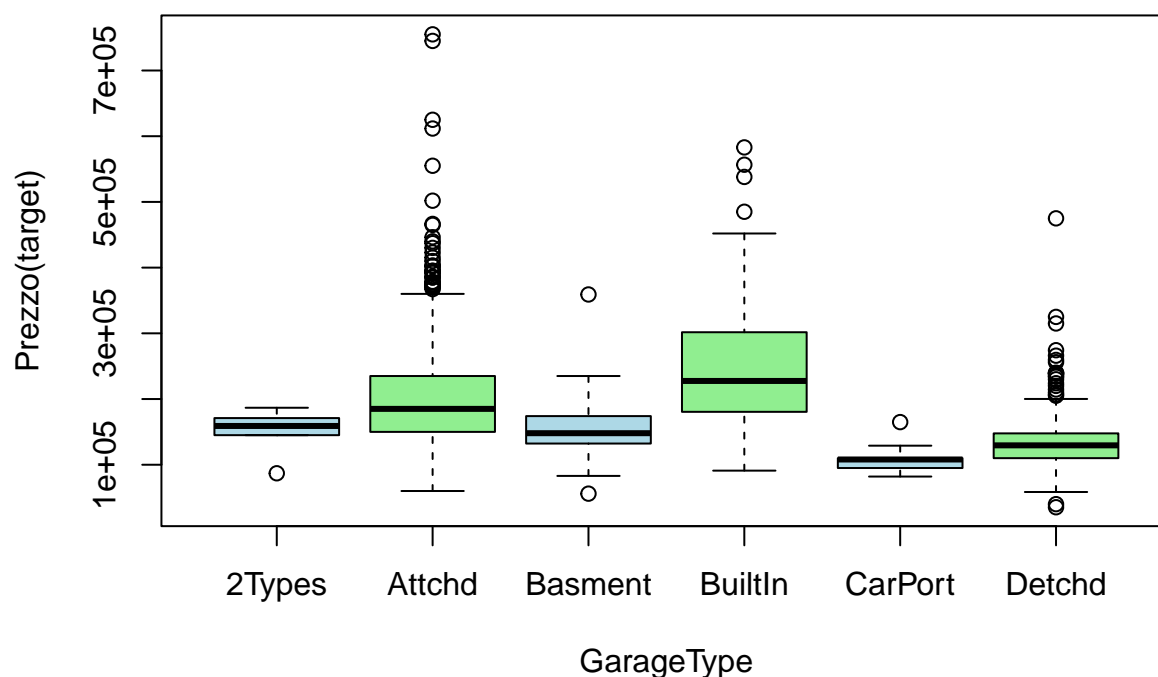
```
## Factor w/ 6 levels "2Types","Attchd",...: 2 2 2 6 2 2 2 2 6 2 ...
```

```
HousePrices$GarageType <- factor(HousePrices$GarageType)
summary(HousePrices$GarageType) # noto che ci sono degli NA's
```

```
## 2Types Attchd Basment BuiltIn CarPort Detchd NA's
##      6      870      19      88      9      387      81
```

```
# graficamente:
boxplot(SalePrice ~ GarageType, data = HousePrices,
        main = "Boxplot di SalePrice per ciascun livello di GarageType",
        xlab = "GarageType", ylab = "Prezzo(target)", col = c("lightblue", "lightgreen"))
```

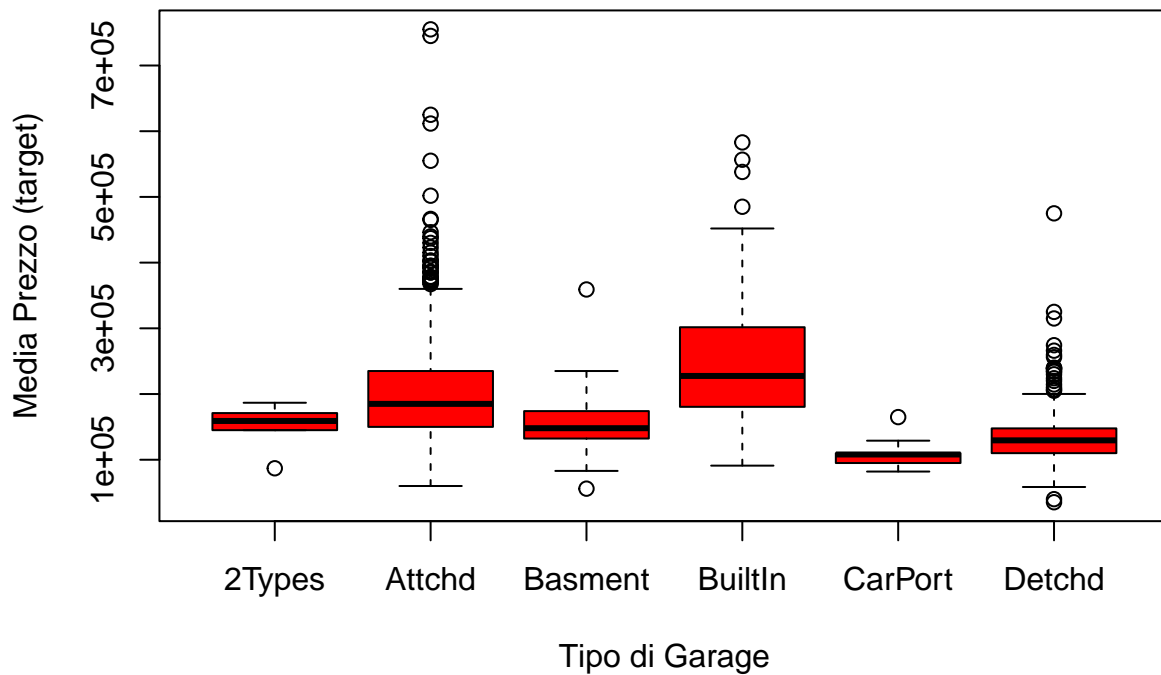
Boxplot di SalePrice per ciascun livello di GarageType



*# la distribuzione e' asimmetrica per quanto riguarda Attchd, Basement, BuiltIn e CarPort
 # invece risulta simmetrica per 2Types e Detchd
 # la varianza piu' alta si trova nella categoria BuiltIn
 # ci sono molti outliers e questo va ad indicare una distribuzione con code lunghe
 # o la presenza di valori anomali.*

```
# la media dei prezzi di vendita per ciascun tipo di garage
media_prezzi <- tapply(HousePrices$SalePrice, HousePrices$GarageType, mean)
# il grafico a barre:
boxplot(HousePrices$SalePrice~HousePrices$GarageType,
  main = "Media dei Prezzi per Tipo di Garage", xlab = "Tipo di Garage",
  ylab = "Media Prezzo (target)", col = "red")
```

Media dei Prezzi per Tipo di Garage



```
# il prezzo medio piu' alto e' quello della categoria BuiltIn
```

```
# PoolArea
```

```
# variabile che indica quanti piedi quadri ha la piscina
```

```
str(HousePrices$PoolArea)
```

```
## int [1:1460] 0 0 0 0 0 0 0 0 0 ...
```

```
summary(HousePrices$PoolArea) # poche proprieta' hanno la piscina
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
```

```
## 0.000  0.000   0.000   2.759  0.000 738.000
```

```
plot(HousePrices$PoolArea, HousePrices$SalePrice, main = "Grafico a dispersione",  
      xlab = "Piedi quadri", ylab = "Prezzo(target)")
```

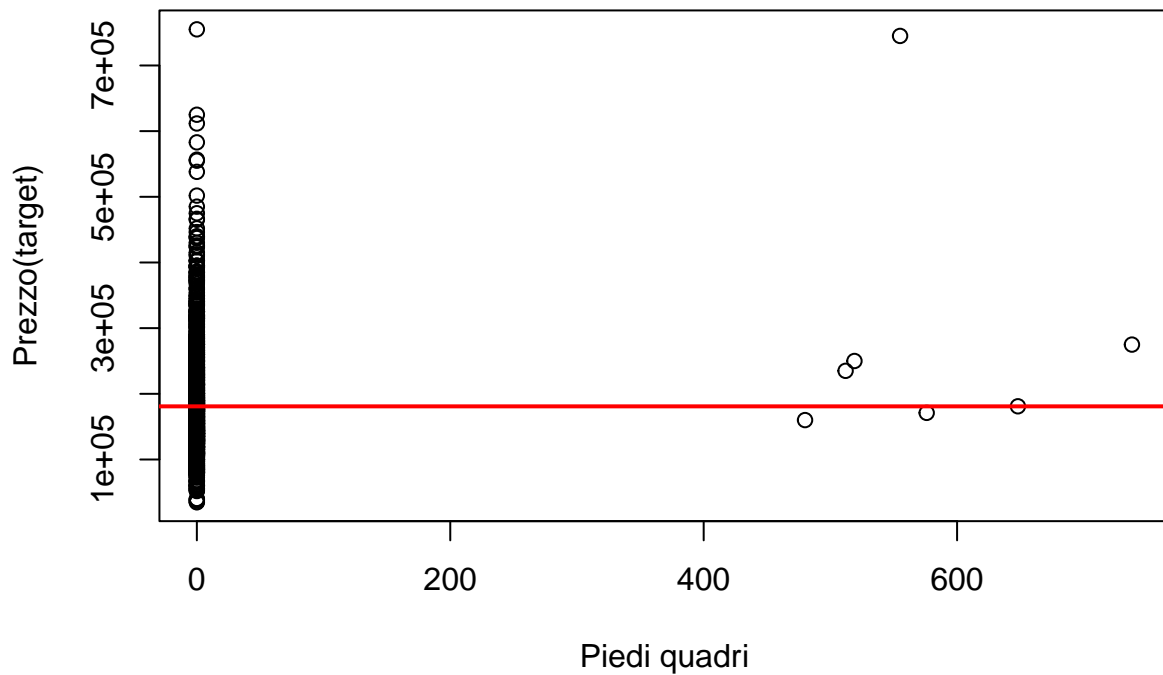
```
# solo una proprieta' con la piscina ha un prezzo molto elevato
```

```
# inoltre se aggiungiamo la retta della media dei prezzi:
```

```
media_prezzi_case <- mean(HousePrices$SalePrice) # che e' 180k
```

```
abline(h = media_prezzi_case, col = "red", lwd = 2)
```

Grafico a dispersione



notiamo che le altre proprieta' con la piscina hanno tutte prezzi vicini alla media

Calcolo la correlazione

```
correlation <- cor(HousePrices$PoolArea, HousePrices$SalePrice, method="pearson")
correlation
```

```
## [1] 0.09240355
```

debole relazione lineare tra le due variabili.

GarageCars

variabile numerica che indica la capienza del garage per le automobili

```
str(HousePrices$GarageCars)
```

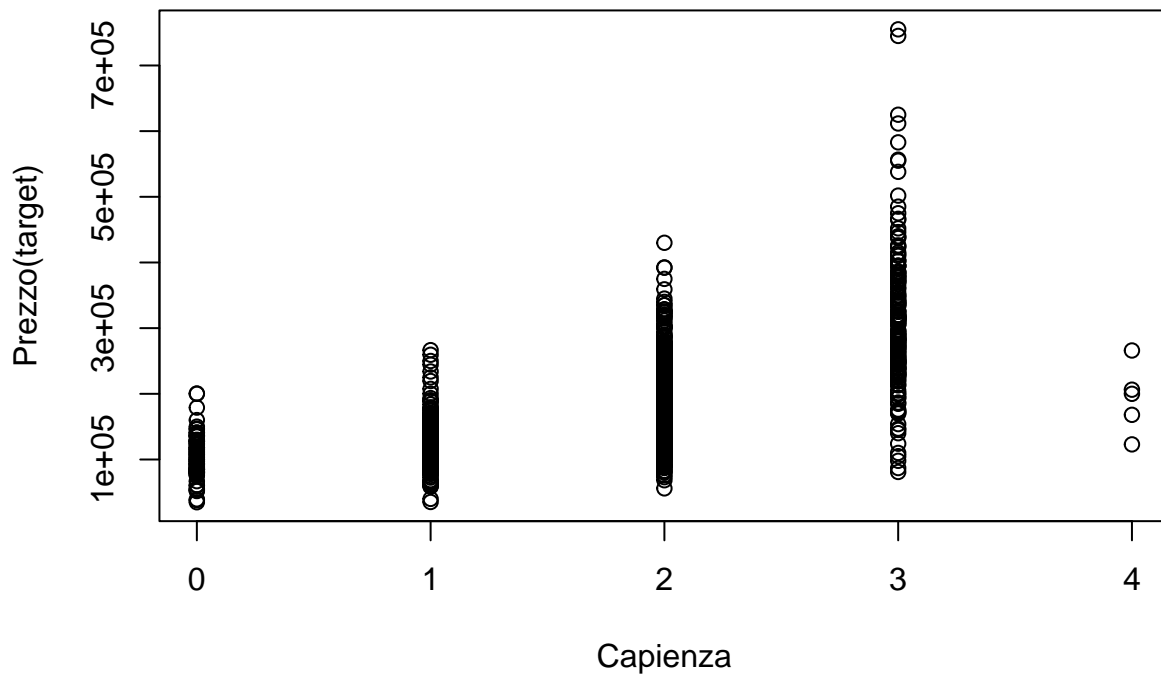
```
## int [1:1460] 2 2 2 3 3 2 2 2 2 1 ...
```

```
summary(HousePrices$GarageCars) # varia da 0(niente garage) a 4
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.000   1.000   2.000   1.767   2.000   4.000
```

```
plot(HousePrices$GarageCars, HousePrices$SalePrice, main = "Grafico a dispersione",
      xlab = "Capienza", ylab = "Prezzo(target)")
```

Grafico a dispersione



*# le proprieta' con la capienza di 3 posti auto sono quelle piu' costose(non tutte ovviamente)
ci sono nettamente meno dati per le case con 4 posti*

```
correlation <- cor(HousePrices$GarageCars, HousePrices$SalePrice, method="pearson")
correlation
```

```
## [1] 0.6404092
```

*# abbastanza forte relazione tra le due variabili
all'aumentare della variabile x aumenta anche la variabile y(target)*

VARIABILI QUANTITATIVE

```
QuantVars = HousePrices[, c("SalePrice", "OverallQual", "GrLivArea",
                             "GarageCars", "GarageArea", "TotalBsmtSF")]
# Correlazione tra le variabili
cor(QuantVars)
```

```
##          SalePrice OverallQual GrLivArea GarageCars GarageArea TotalBsmtSF
## SalePrice    1.0000000  0.7909816  0.7086245  0.6404092  0.6234314  0.6135806
## OverallQual  0.7909816  1.0000000  0.5930074  0.6006707  0.5620218  0.5378085
## GrLivArea    0.7086245  0.5930074  1.0000000  0.4672474  0.4689975  0.4548682
## GarageCars   0.6404092  0.6006707  0.4672474  1.0000000  0.8824754  0.4345848
## GarageArea   0.6234314  0.5620218  0.4689975  0.8824754  1.0000000  0.4866655
## TotalBsmtSF  0.6135806  0.5378085  0.4548682  0.4345848  0.4866655  1.0000000
```

```

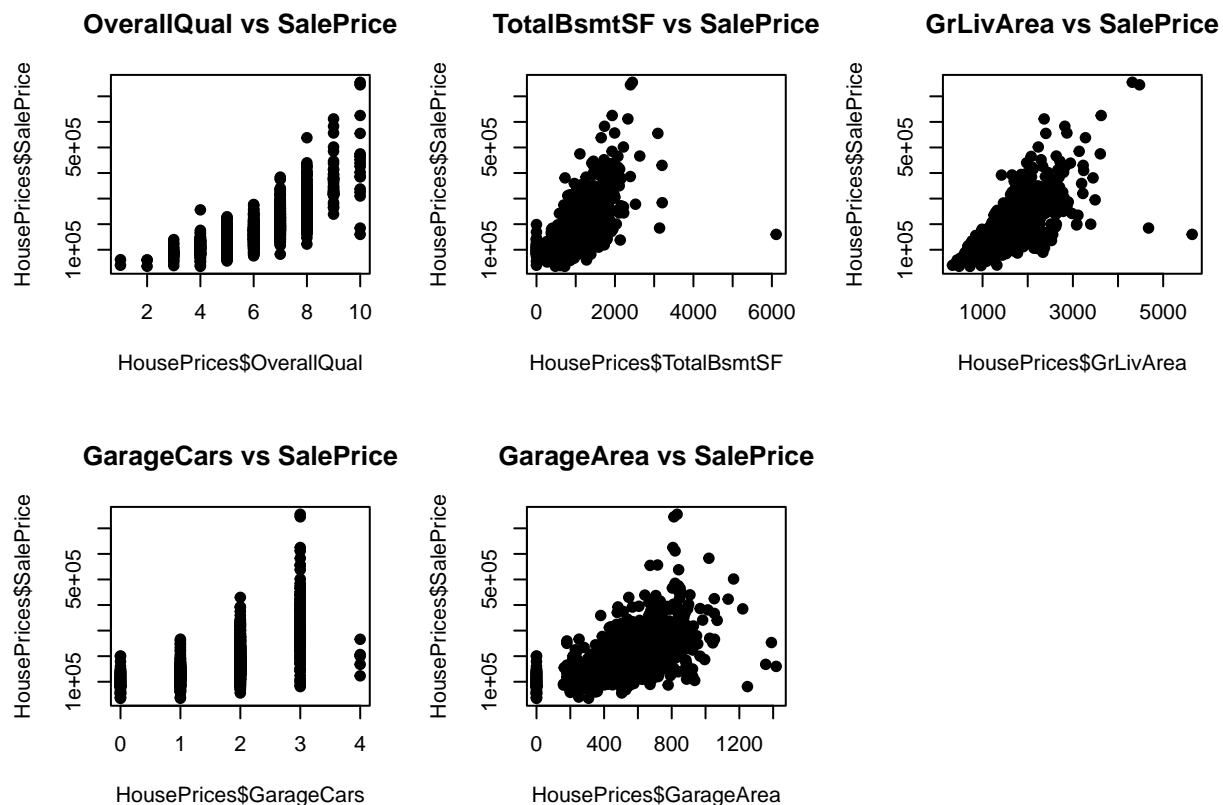
#Scatterplot
par(mfrow=c(2,3))
plot(HousePrices$OverallQual, HousePrices$SalePrice, main="OverallQual vs SalePrice", pch=19)
plot(HousePrices$TotalBsmtSF, HousePrices$SalePrice, main="TotalBsmtSF vs SalePrice", pch=19)
plot(HousePrices$GrLivArea, HousePrices$SalePrice, main="GrLivArea vs SalePrice", pch=19)
plot(HousePrices$GarageCars, HousePrices$SalePrice, main="GarageCars vs SalePrice", pch=19)
plot(HousePrices$GarageArea, HousePrices$SalePrice, main="GarageArea vs SalePrice", pch=19)

#Calcolo covarianza e correlazione
CovCorr = function(x, y) {
  n = length(y)
  Covarianza = sum((x-mean(x))*(y-mean(y))) / (n-1)
  Correlazione = Covarianza / (sqrt(var(x)*var(y)))
  return(list(Covarianza=Covarianza, Correlazione=Correlazione))
}

#Calcolo delle covarianze e correlazioni
Quantitative = c("OverallQual", "GrLivArea", "GarageCars", "GarageArea", "TotalBsmtSF")
for (quant in Quantitative) {
  cat("Analisi per", quant, "\n")
  result = CovCorr(HousePrices[[quant]], HousePrices$SalePrice)
  cat("Covarianza tra", quant, "e SalePrice:", result$Covarianza, "\n")
  cat("Coefficiente di correlazione tra", quant, "e SalePrice:", result$Correlazione, "\n\n")
}

## Analisi per OverallQual
## Covarianza tra OverallQual e SalePrice: 86904.13
## Coefficiente di correlazione tra OverallQual e SalePrice: 0.7909816
##
## Analisi per GrLivArea
## Covarianza tra GrLivArea e SalePrice: 29581867
## Coefficiente di correlazione tra GrLivArea e SalePrice: 0.7086245
##
## Analisi per GarageCars
## Covarianza tra GarageCars e SalePrice: 38020.18
## Coefficiente di correlazione tra GarageCars e SalePrice: 0.6404092
##
## Analisi per GarageArea
## Covarianza tra GarageArea e SalePrice: 10589103
## Coefficiente di correlazione tra GarageArea e SalePrice: 0.6234314
##
## Analisi per TotalBsmtSF
## Covarianza tra TotalBsmtSF e SalePrice: 21384417
## Coefficiente di correlazione tra TotalBsmtSF e SalePrice: 0.6135806

```



Bivariata (tutte le variabili)

```
HousePricesBivConf <- read.csv("houseprices-biv.config.csv")
```

```
bivarAnalysis <- function(conf, df) {
  cat("\n\nnewpage")
  cat("\n\n##", conf["colname"], "\n")
  if (as.logical(conf["uselog"])) {
    df[, conf["colname"]] <- log(df[, conf["colname"]])
  }
  cat("\n\n Numero di NA: ", sum(is.na(df[, conf["colname"]]), na.rm = T), "\n\n")
  if (as.logical(conf["filterzeros"])) {
    cat("\n\n Numero di zeri rimossi: ", sum(df[, conf["colname"]] == 0, na.rm = T), "\n\n")
    df <- df[df[, conf["colname"]] != 0,]
  }
  if (as.logical(conf["docmpboxplot"])) {
    boxplot(log(df[, conf["tgcol"]])~df[, conf["colname"]],
            xlab = conf["colname"], ylab = conf["tgcol"])
  }
  if (as.logical(conf["doanova"])) {
    anova_obj <- aov(log(df[, conf["tgcol"]])~df[, conf["colname"]])
    print(kable(summary(anova_obj)[1][[1]], row.names = FALSE))
    if (as.logical(conf["doetasq"])) {
      etasq_obj <- summary(anova_obj)[1][[1]][[2]][1]/(summary(anova_obj)[1][[1]][[2]][2]+
                                                         summary(anova_obj)[1][[1]][[2]][1])
    }
  }
}
```

```

    cat("\n\n Eta squared: ", etasq_obj, "\n\n")
  }
}
if (as.logical(conf["dosscatterplot"])) {
  plot(df[, conf["colname"]], log(df[, conf["tgcol"]]), xlab = conf["colname"],
        ylab = conf["tgcol"])
  lm_obj <- lm(log(df[, conf["tgcol"]])~df[, conf["colname"]])
  abline(a= lm_obj$coefficients[1], b = lm_obj$coefficients[2], col=2)
}

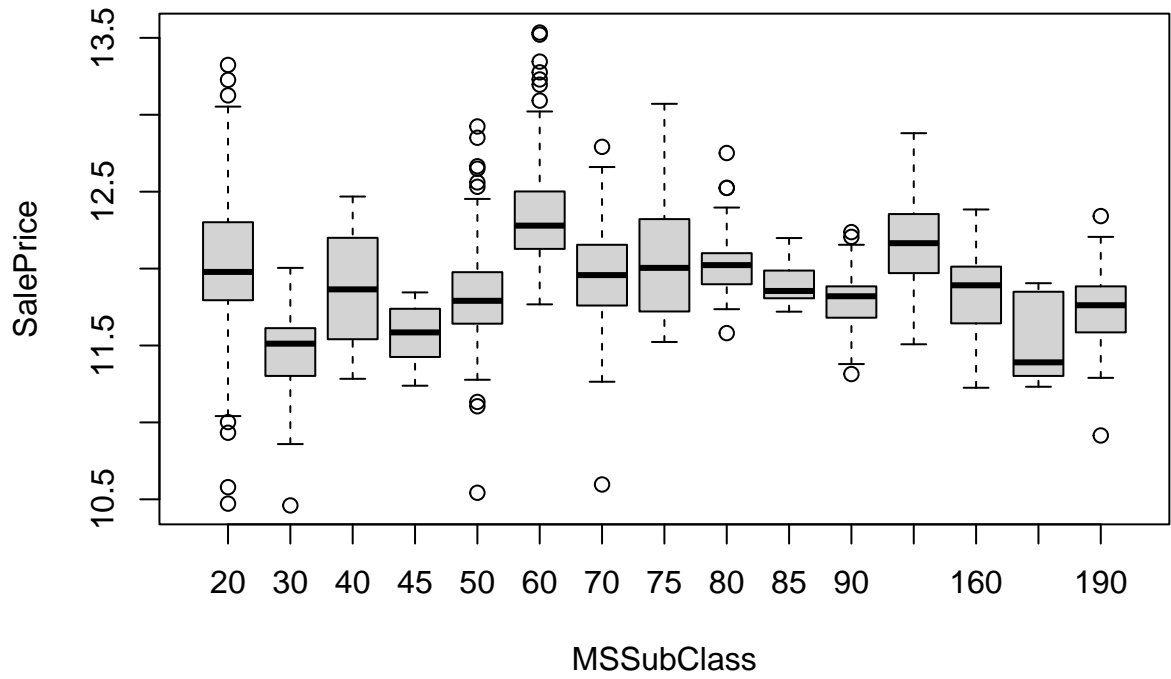
if (as.logical(conf["dosscatterplot"])) {
  plot(df[, conf["colname"]], log(df[, conf["tgcol"]]), xlab = conf["colname"],
        ylab = conf["tgcol"])
  lm_obj <- lm(log(df[, conf["tgcol"]])~df[, conf["colname"]])
  abline(a= lm_obj$coefficients[1], b = lm_obj$coefficients[2], col=2)
  print(kable(summary(lm_obj)$coefficients, row.names = FALSE))
  if (as.logical(conf["dorsq"])) {
    print(kable(cbind(
      cov = cov(df[, conf["colname"]], log(df[, conf["tgcol"]]), use = "complete.obs"),
      cor = cor(df[, conf["colname"]], log(df[, conf["tgcol"]]), use = "complete.obs"),
      rsq = summary(lm_obj)$r.squared
    )))
  }
}
if (!is.na(conf["desc"])) {
  cat("\n\n", conf["desc"], "\n\n")
}
}

out <- apply(HousePricesBivConf, 1, bivarAnalysis, df = HousePrices)

```


MSSubClass

Numero di NA: 0

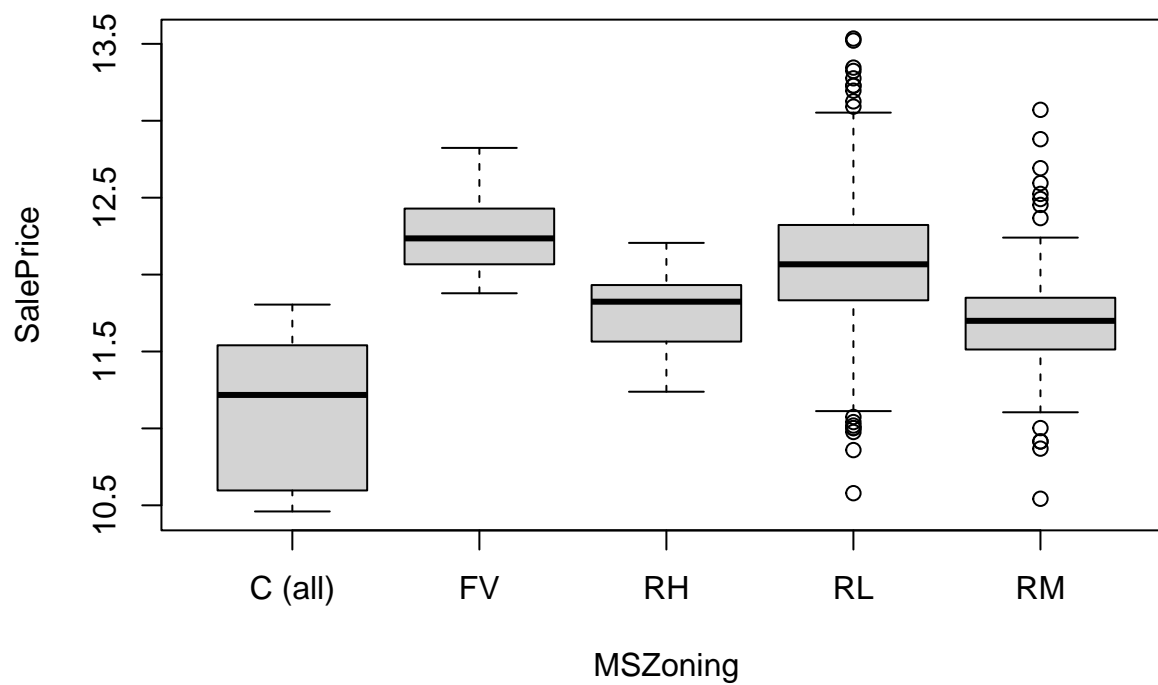


Df	Sum Sq	Mean Sq	F value	Pr(>F)
1	1.273405	1.273408	8.019031	0.0046924
1458	231.527254	0.1587978	NA	NA

Eta squared: 0.005469936

MSZoning

Numero di NA: 0

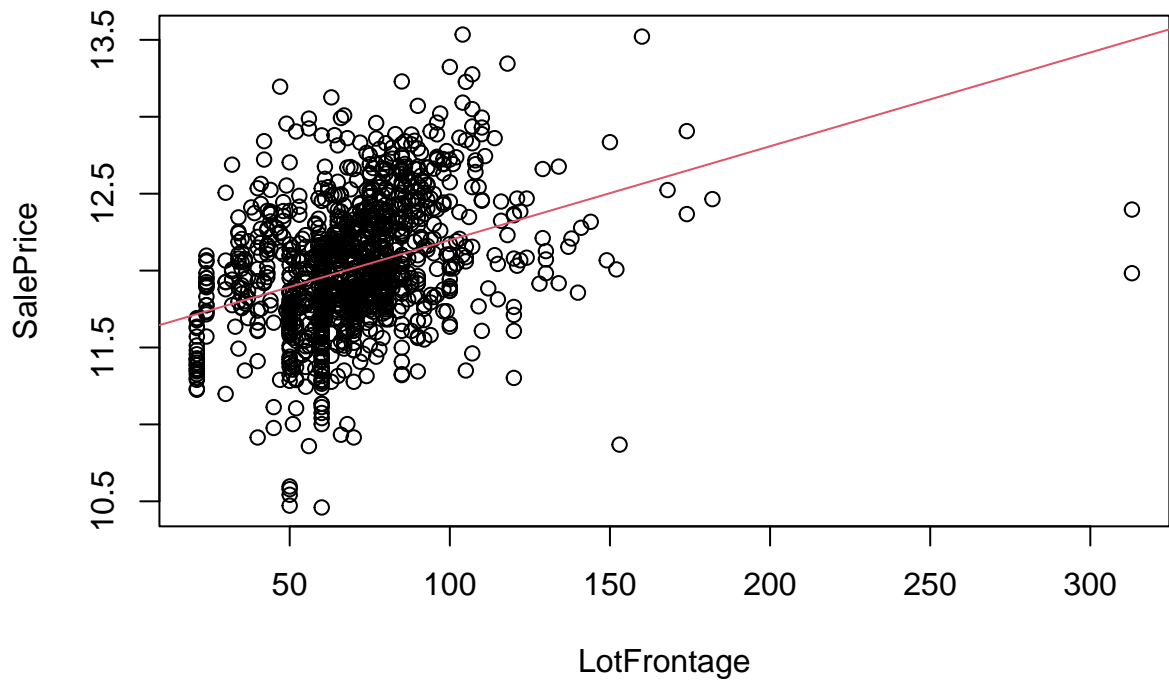


Df	Sum Sq	Mean Sq	F value	Pr(>F)
4	40.93539	10.2338483	77.60784	0
1455	191.86527	0.1318662	NA	NA

Eta squared: 0.1758388

LotFrontage

Numero di NA: 259

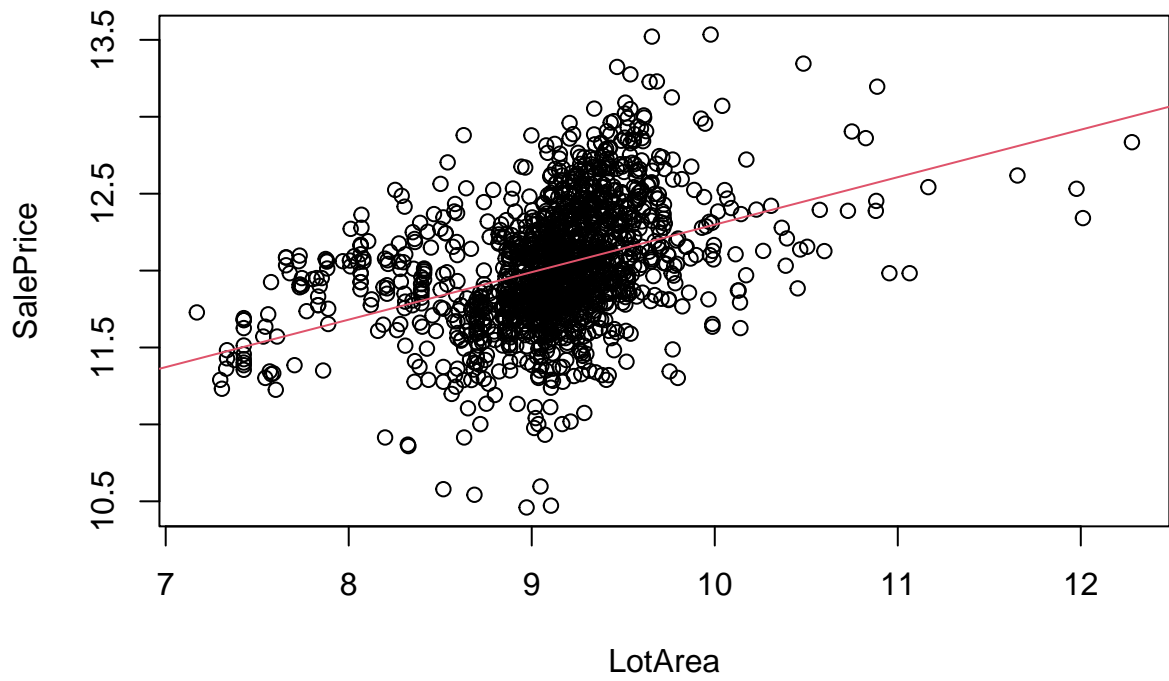


Estimate	Std. Error	t value	Pr(> t)
11.5887337	0.0342787	338.07361	0
0.0060969	0.0004624	13.18612	0

cov	cor	rsq
3.595635	0.3558785	0.1266495

LotArea

Numero di NA: 0

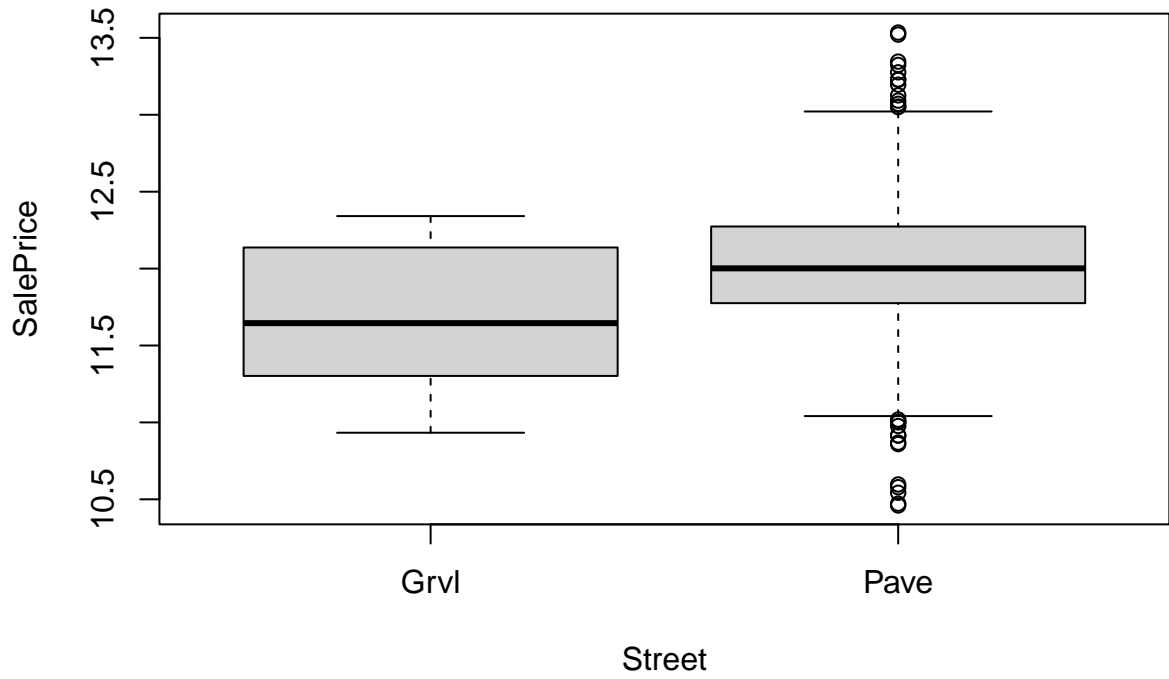


Estimate	Std. Error	t value	Pr(> t)
9.2113297	0.1690958	54.47404	0
0.3087226	0.0185300	16.66068	0

cov	cor	rsq
0.0826612	0.3999177	0.1599342

Street

Numero di NA: 0

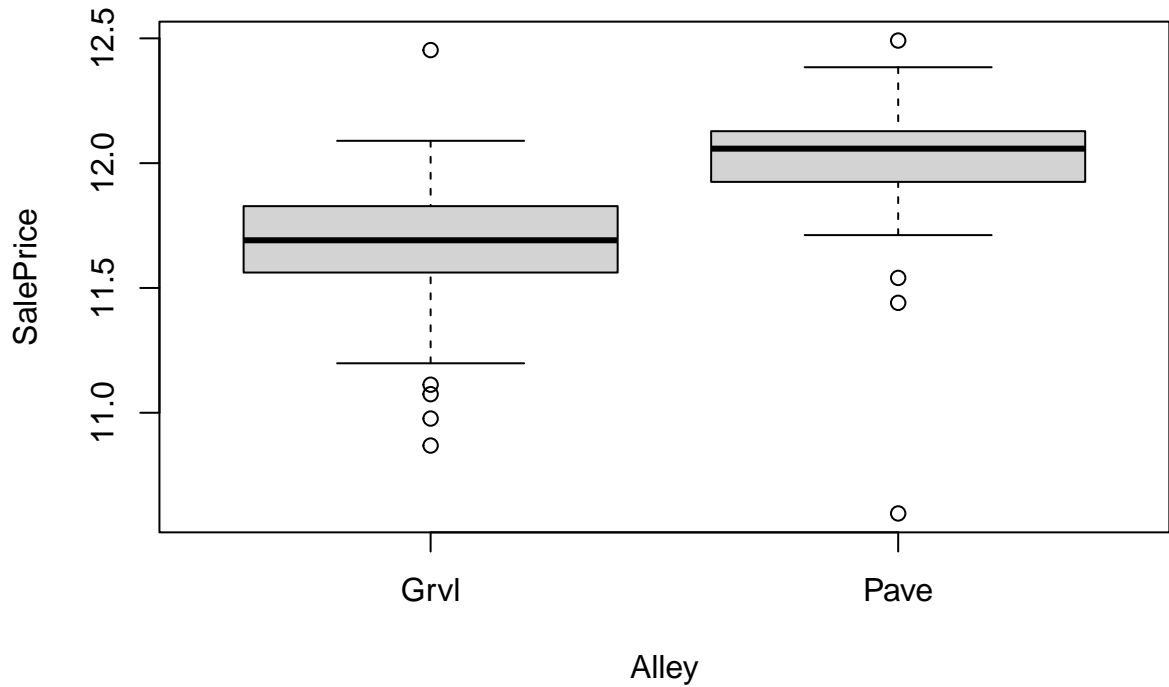


Df	Sum Sq	Mean Sq	F value	Pr(>F)
1	0.7661999	0.7661999	4.814455	0.0283793
1458	232.0344591	0.1591457	NA	NA

Eta squared: 0.003291228

Alley

Numero di NA: 1369

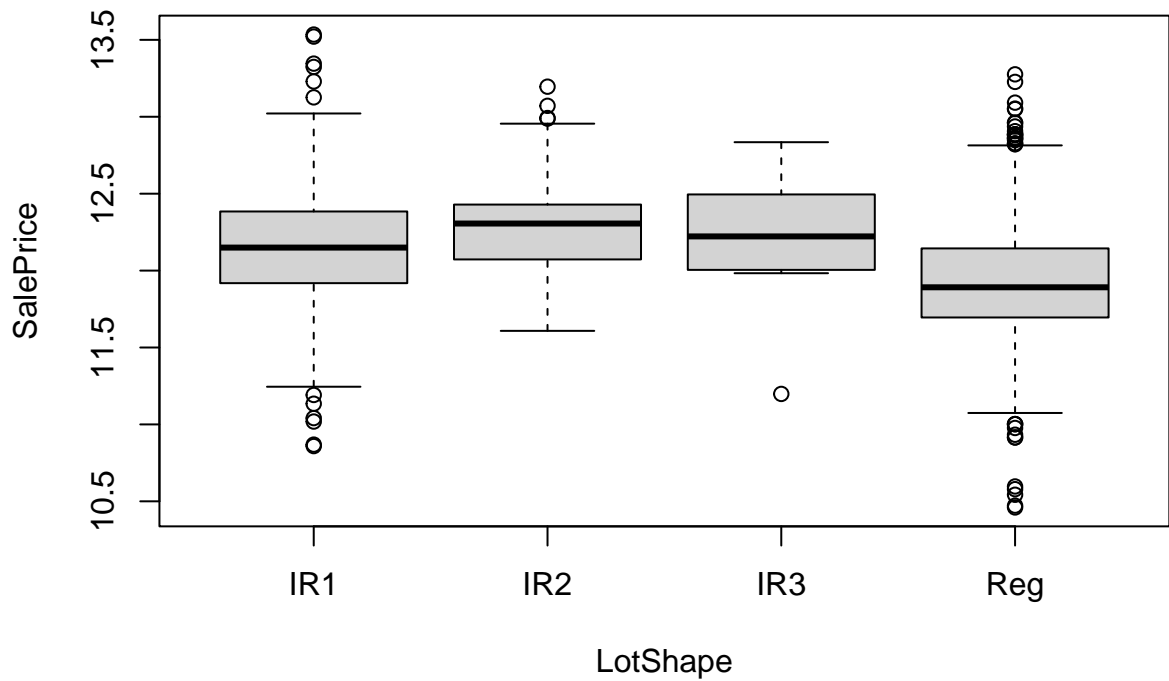


Df	Sum Sq	Mean Sq	F value	Pr(>F)
1	2.356945	2.3569453	27.00739	1.3e-06
89	7.767065	0.0872704	NA	NA

Eta squared: 0.2328075

LotShape

Numero di NA: 0

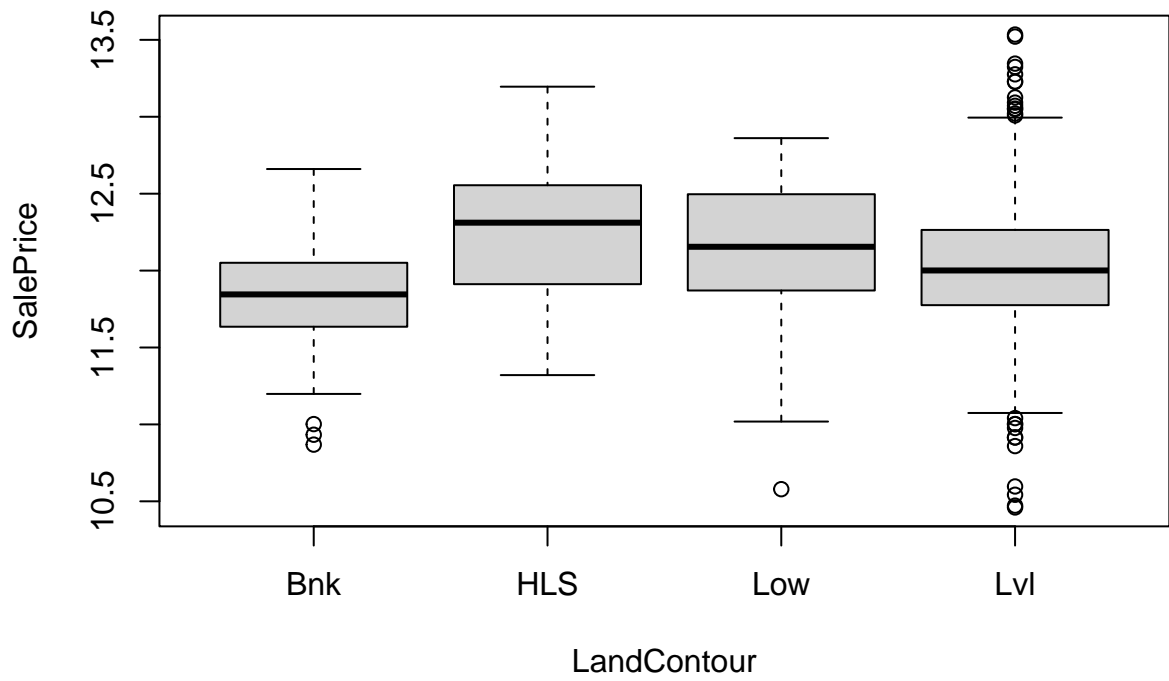


Df	Sum Sq	Mean Sq	F value	Pr(>F)
3	20.4459	6.8152989	46.72876	0
1456	212.3548	0.1458481	NA	NA

Eta squared: 0.08782577

LandContour

Numero di NA: 0

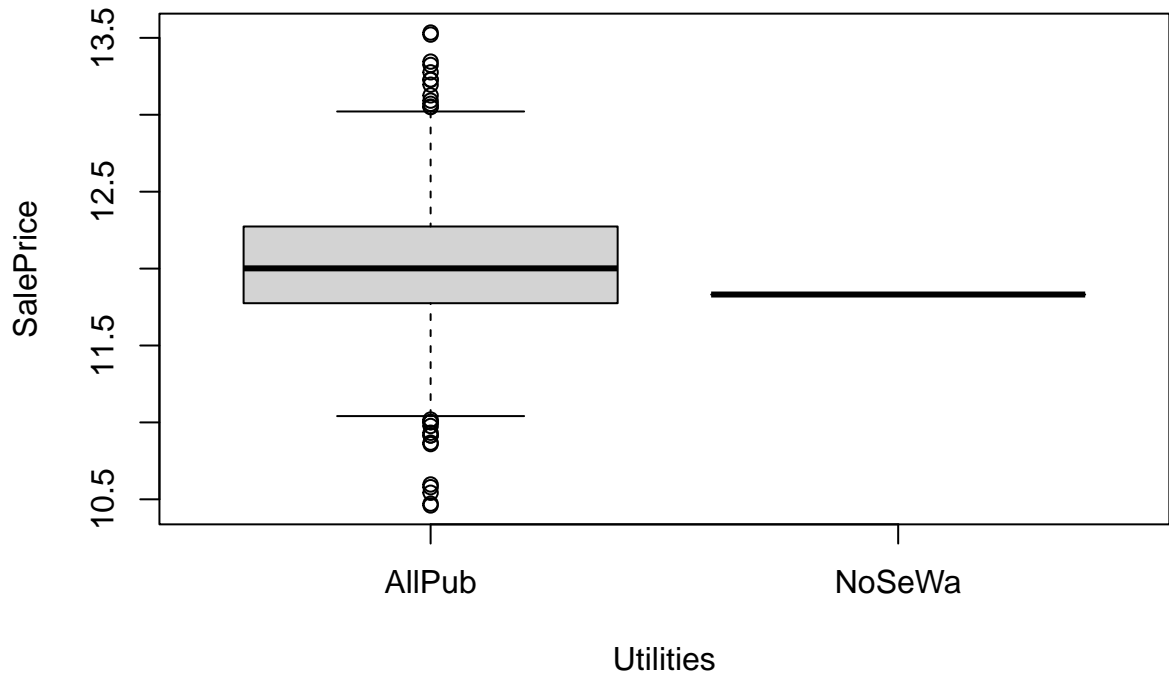


Df	Sum Sq	Mean Sq	F value	Pr(>F)
3	5.967126	1.9890419	12.76727	0
1456	226.833533	0.1557923	NA	NA

Eta squared: 0.02563191

Utilities

Numero di NA: 0

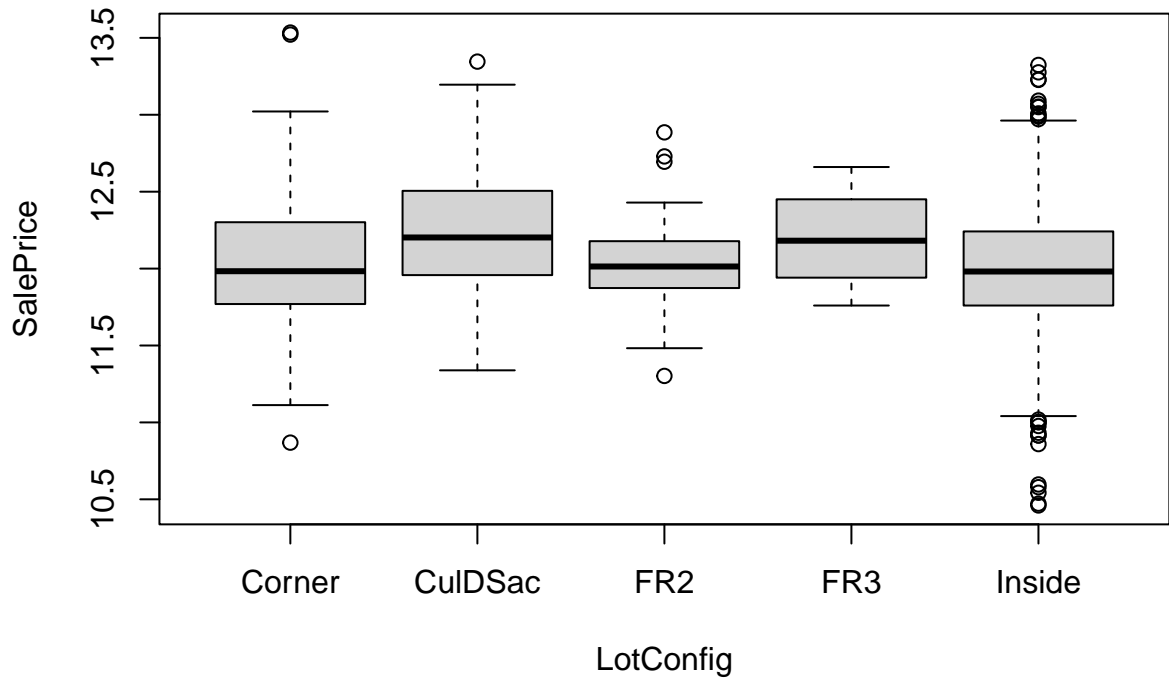


Df	Sum Sq	Mean Sq	F value	Pr(>F)
1	0.0371478	0.0371478	0.2326891	0.6296094
1458	232.7635112	0.1596458	NA	NA

Eta squared: 0.0001595693

LotConfig

Numero di NA: 0

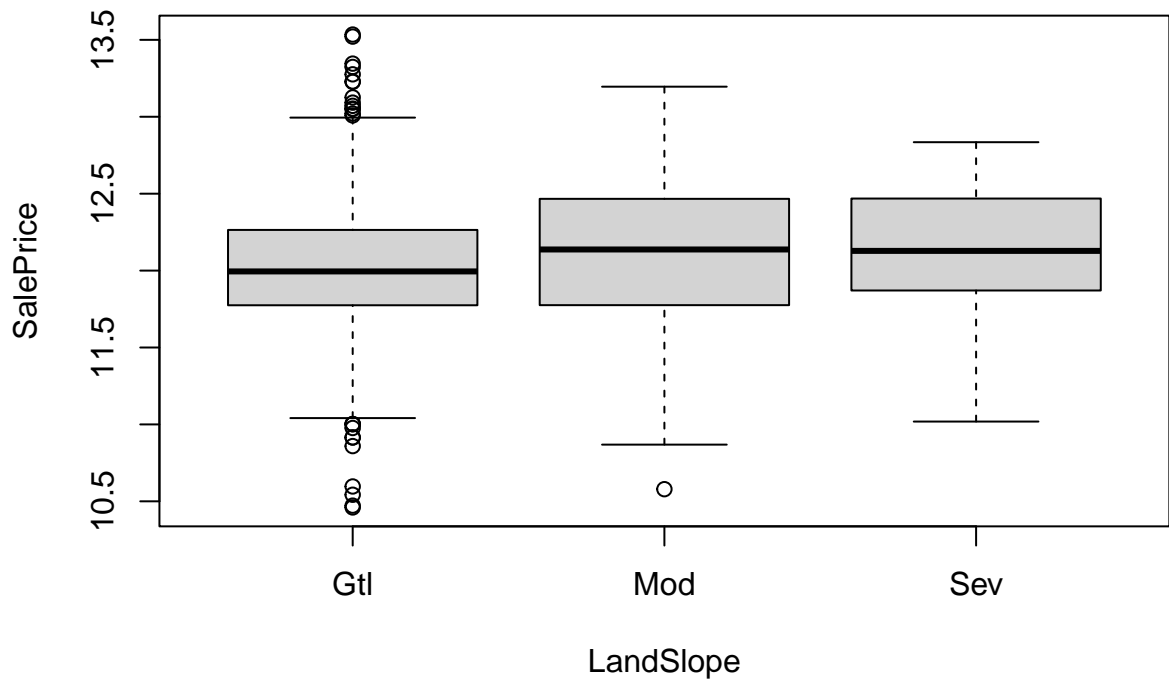


Df	Sum Sq	Mean Sq	F value	Pr(>F)
4	5.433338	1.3583346	8.69244	6e-07
1455	227.367321	0.1562662	NA	NA

Eta squared: 0.02333902

LandSlope

Numero di NA: 0

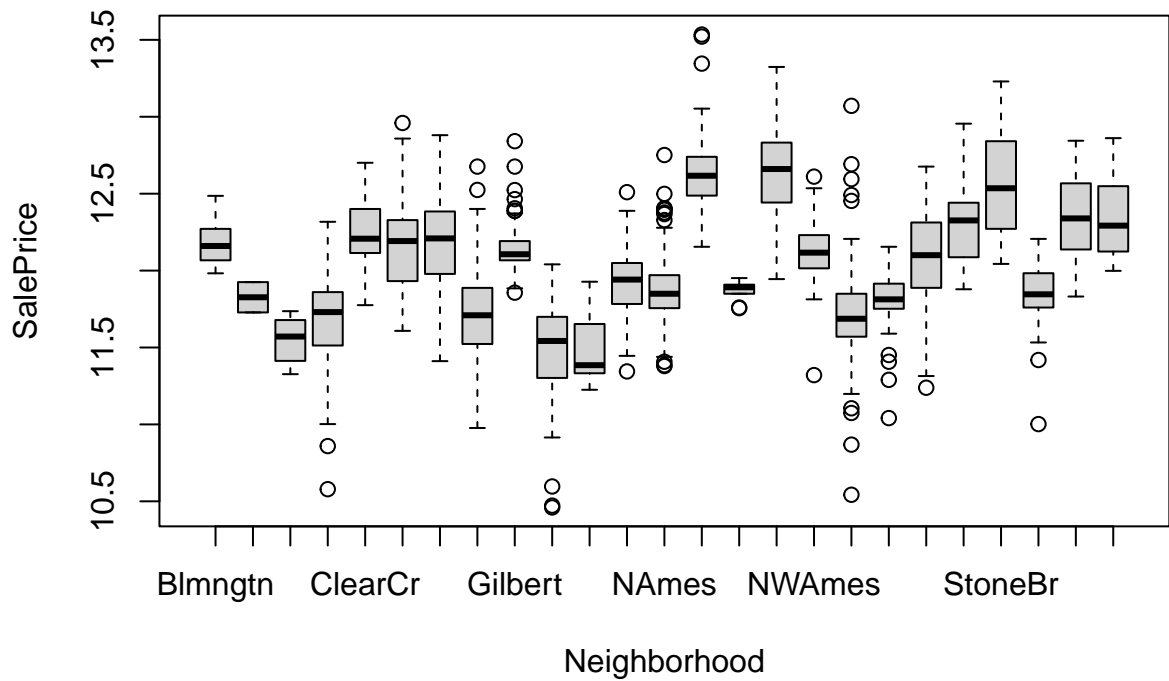


Df	Sum Sq	Mean Sq	F value	Pr(>F)
2	0.3455839	0.1727920	1.083039	0.3388375
1457	232.4550751	0.1595436	NA	NA

Eta squared: 0.001484463

Neighborhood

Numero di NA: 0

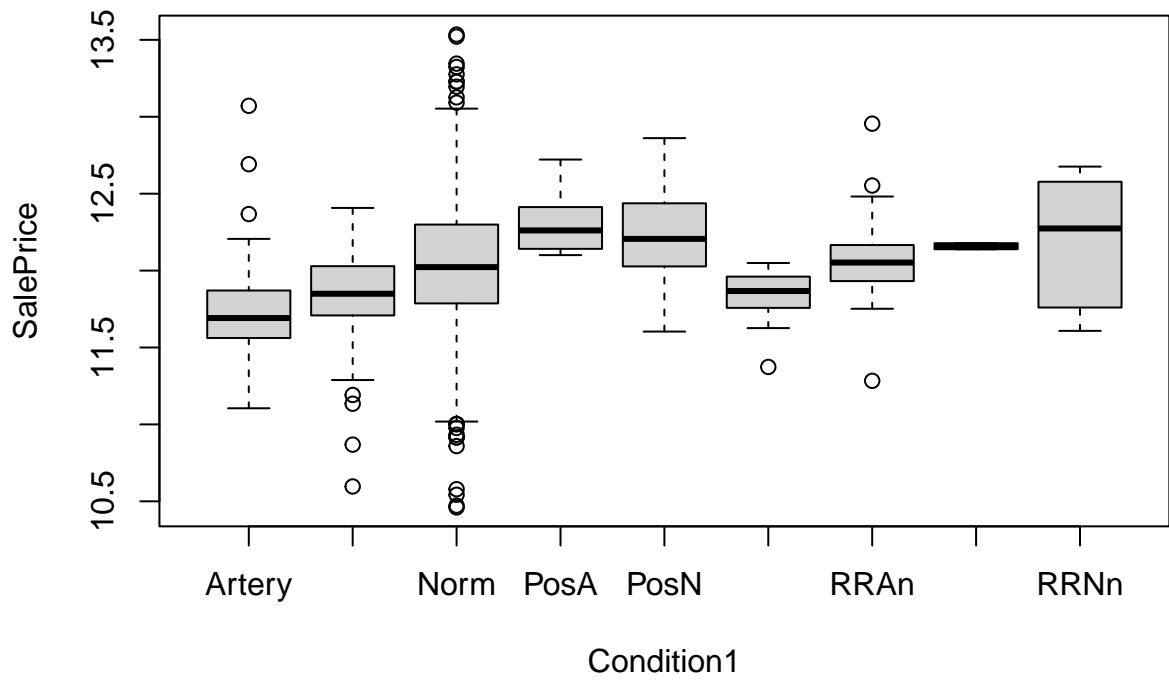


Df	Sum Sq	Mean Sq	F value	Pr(>F)
24	132.88443	5.536851	79.52042	0
1435	99.91623	0.069628	NA	NA

Eta squared: 0.5708078

Condition1

Numero di NA: 0

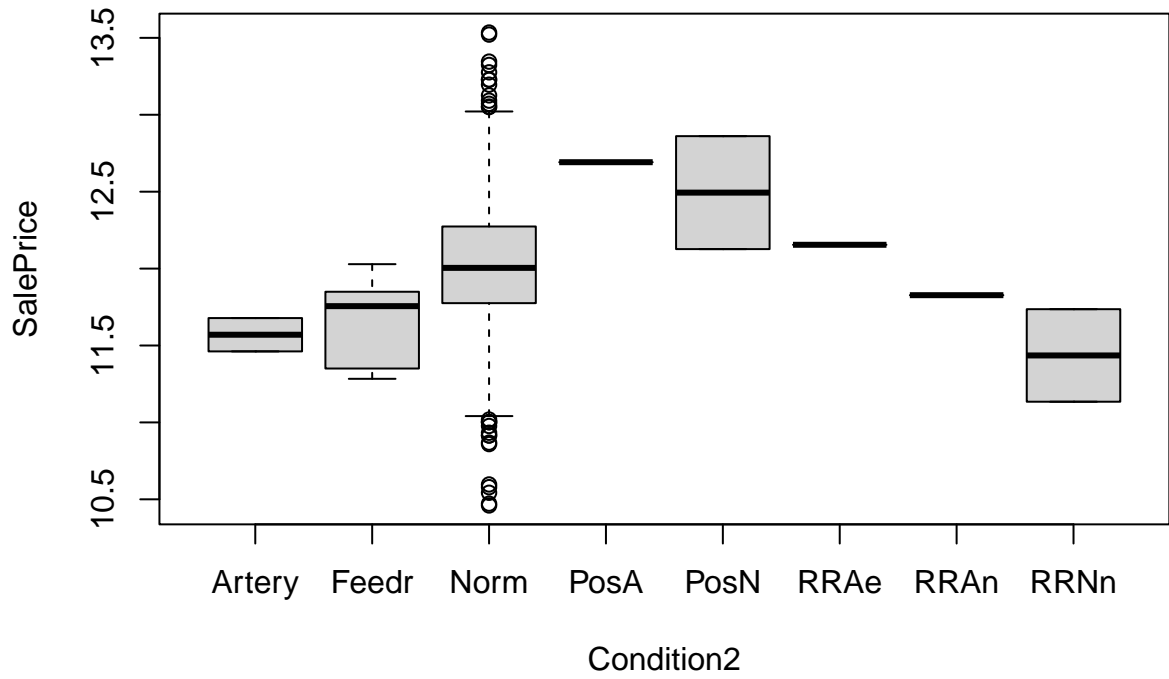


Df	Sum Sq	Mean Sq	F value	Pr(>F)
8	9.878652	1.2348315	8.037522	0
1451	222.922007	0.1536334	NA	NA

Eta squared: 0.04243395

Condition2

Numero di NA: 0

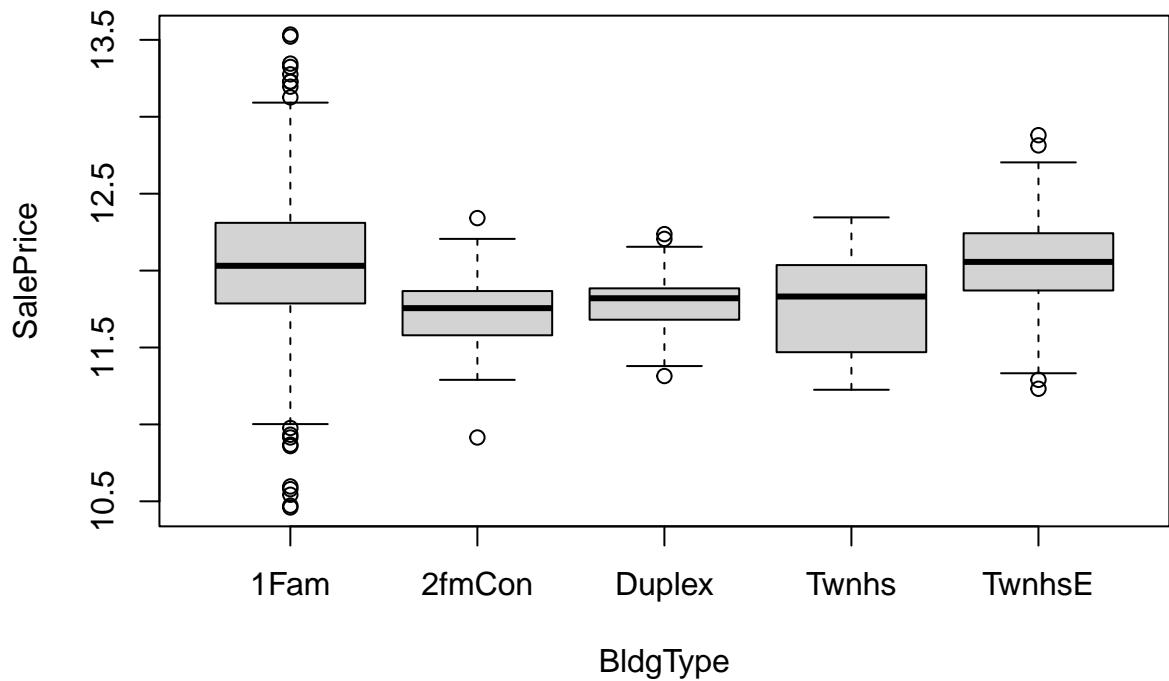


Df	Sum Sq	Mean Sq	F value	Pr(>F)
7	2.802771	0.4003959	2.52774	0.0138204
1452	229.997888	0.1584007	NA	NA

Eta squared: 0.01203936

BldgType

Numero di NA: 0

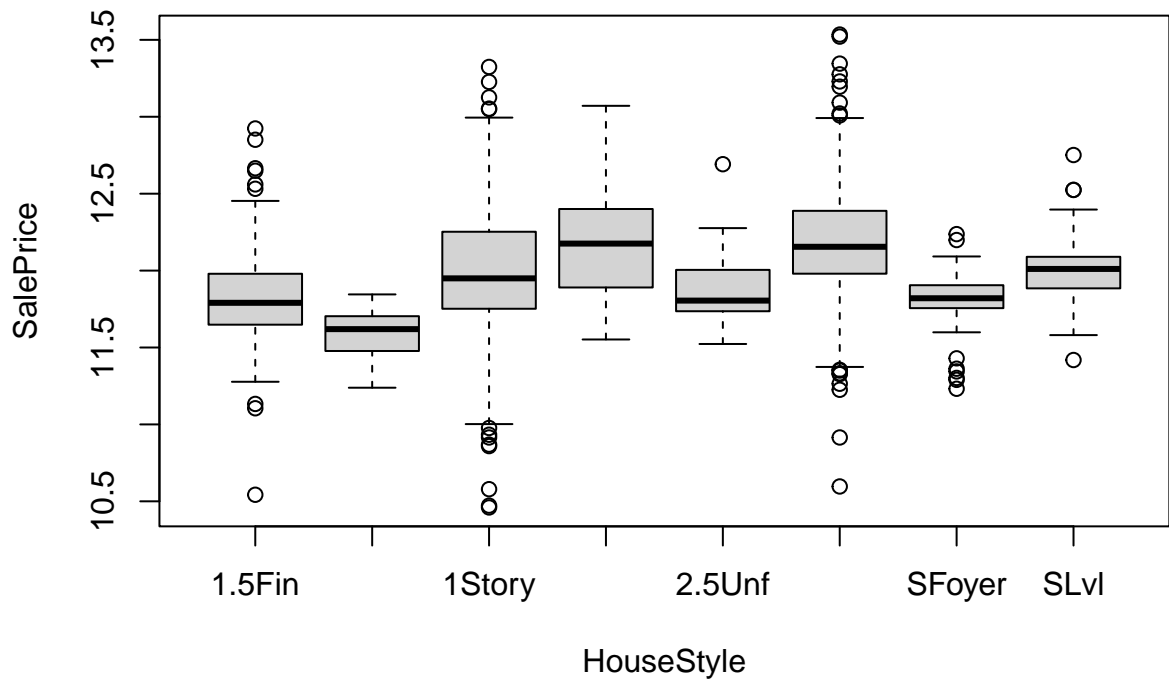


Df	Sum Sq	Mean Sq	F value	Pr(>F)
4	9.344708	2.336177	15.21167	0
1455	223.455951	0.153578	NA	NA

Eta squared: 0.04014039

HouseStyle

Numero di NA: 0

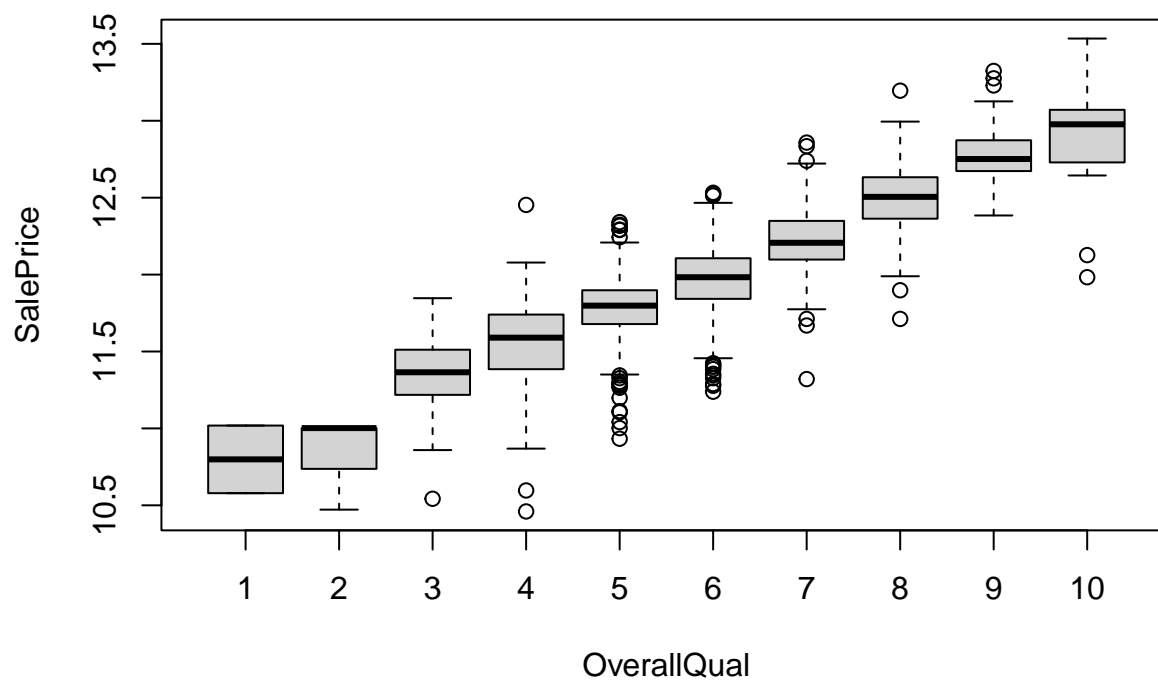


Df	Sum Sq	Mean Sq	F value	Pr(>F)
7	23.76814	3.3954485	23.58576	0
1452	209.03252	0.1439618	NA	NA

Eta squared: 0.1020965

OverallQual

Numero di NA: 0



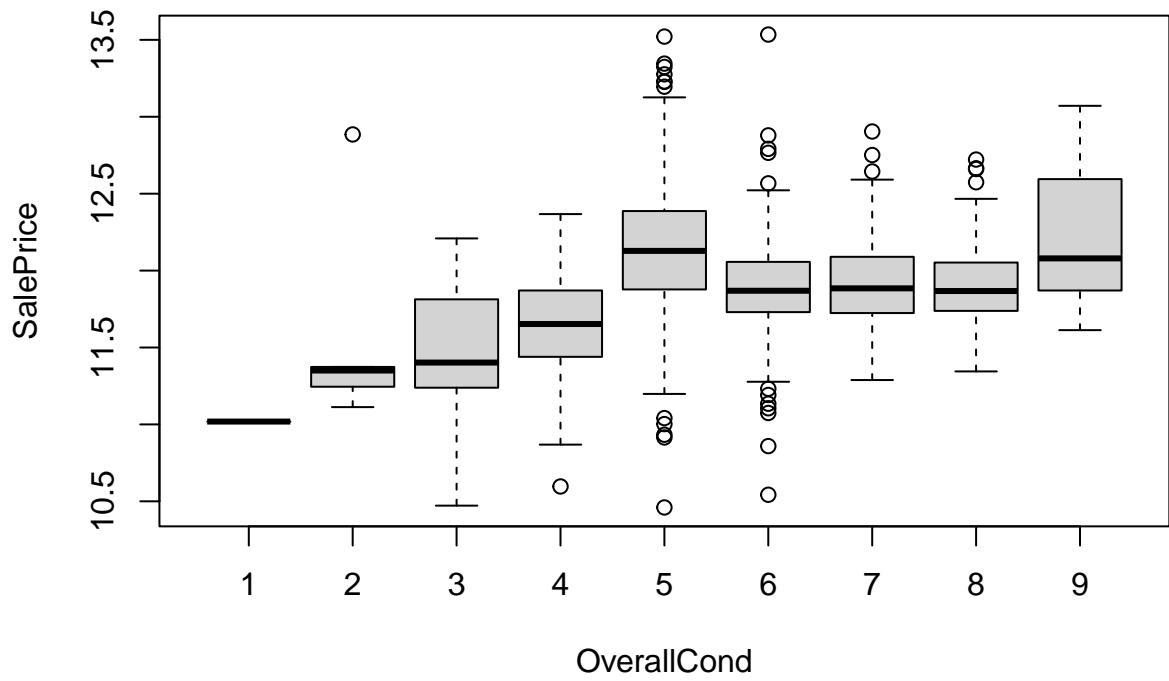
Df	Sum Sq	Mean Sq	F value	Pr(>F)
1	155.46204	155.4620389	2930.795	0
1458	77.33862	0.0530443	NA	NA

Eta squared: 0.6677904

Nel grafico si può notare una relazione positiva tra la qualità complessiva e il prezzo di vendita, questo è confermato dalla covarianza, che è positiva e dall'alto valore del coefficiente di correlazione (0.79). All'aumentare della qualità complessiva, il prezzo di vendita tende ad aumentare

OverallCond

Numero di NA: 0

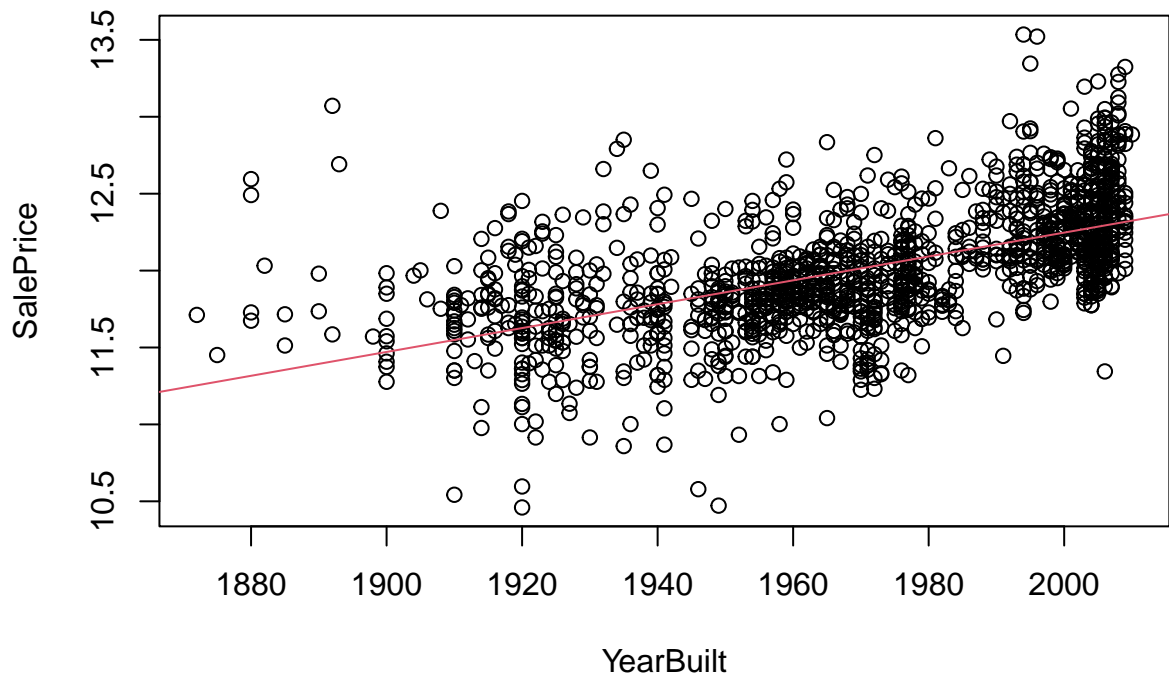


Df	Sum Sq	Mean Sq	F value	Pr(>F)
1	0.316434	0.3164340	1.984482	0.1591323
1458	232.484225	0.1594542	NA	NA

Eta squared: 0.001359249

YearBuilt

Numero di NA: 0

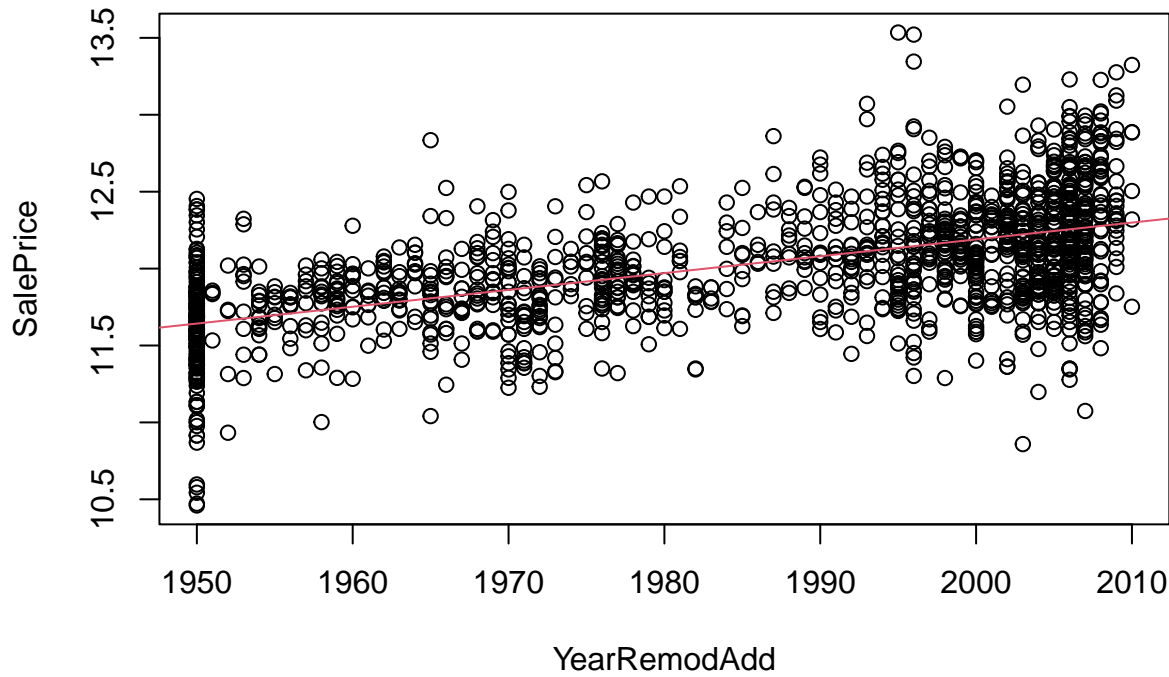


Estimate	Std. Error	t value	Pr(> t)
-3.2685519	0.5530496	-5.910052	0
0.0077577	0.0002805	27.654657	0

cov	cor	rsq
7.076739	0.5865702	0.3440646

YearRemodAdd

Numero di NA: 0

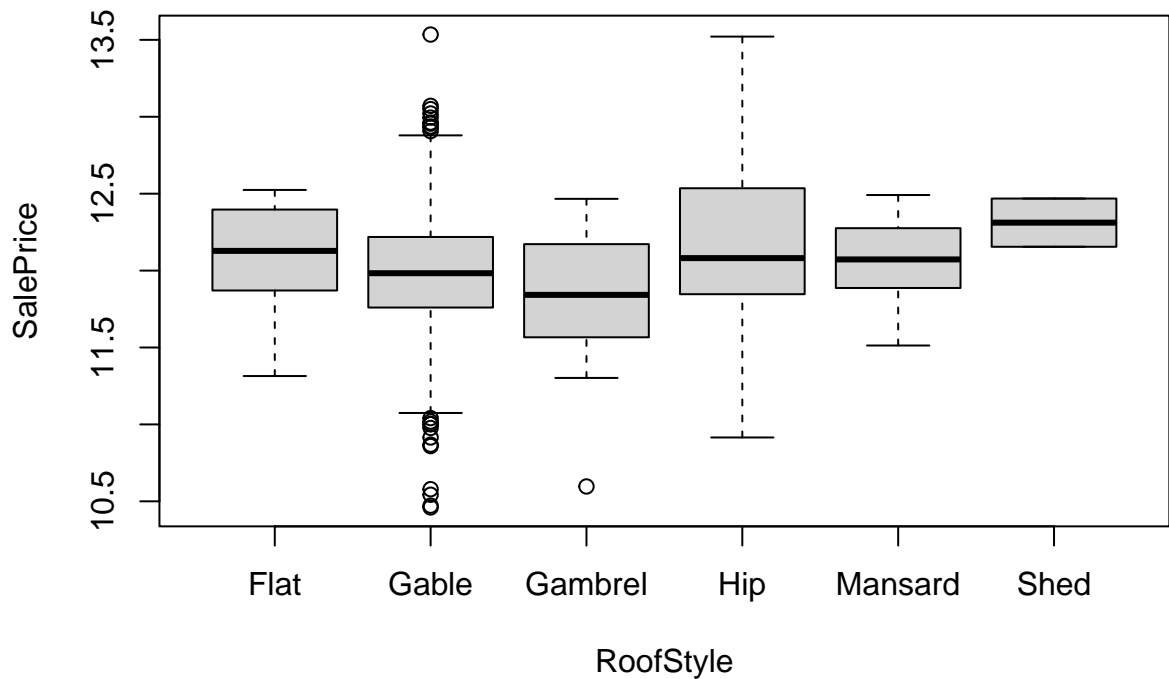


Estimate	Std. Error	t value	Pr(> t)
-9.6973367	0.8294675	-11.69104	0
0.0109435	0.0004179	26.18856	0

cov	cor	rsq
4.664481	0.5656078	0.3199122

RoofStyle

Numero di NA: 0

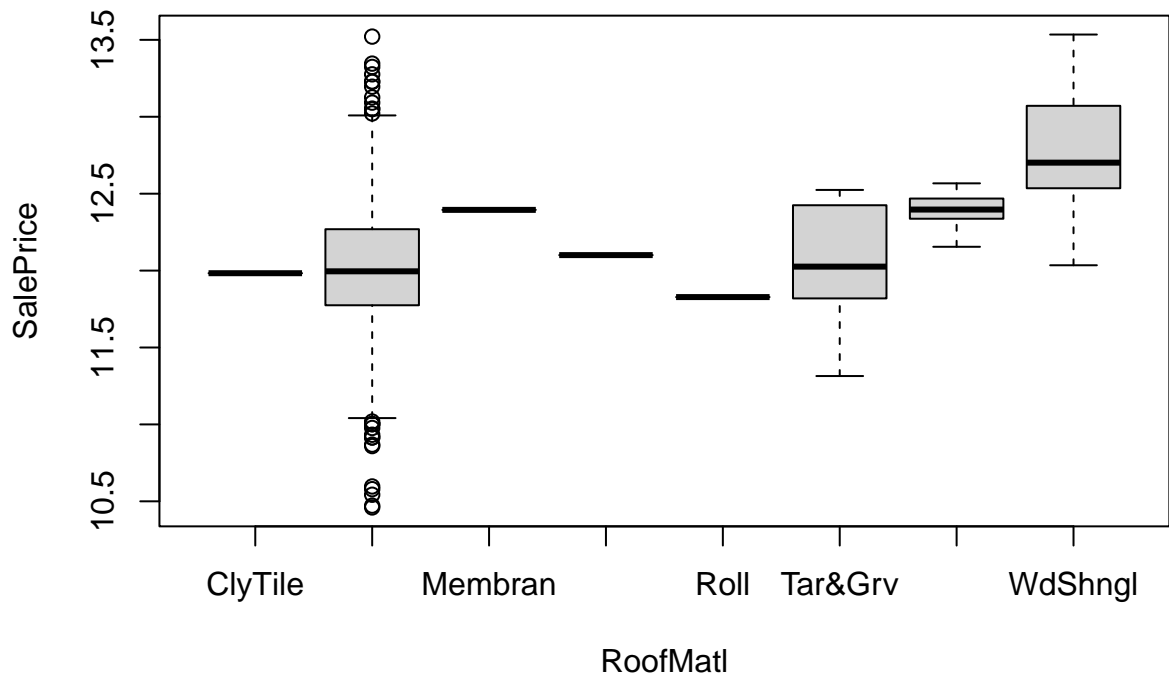


Df	Sum Sq	Mean Sq	F value	Pr(>F)
5	10.03318	2.0066359	13.09728	0
1454	222.76748	0.1532101	NA	NA

Eta squared: 0.04309773

RoofMatl

Numero di NA: 0

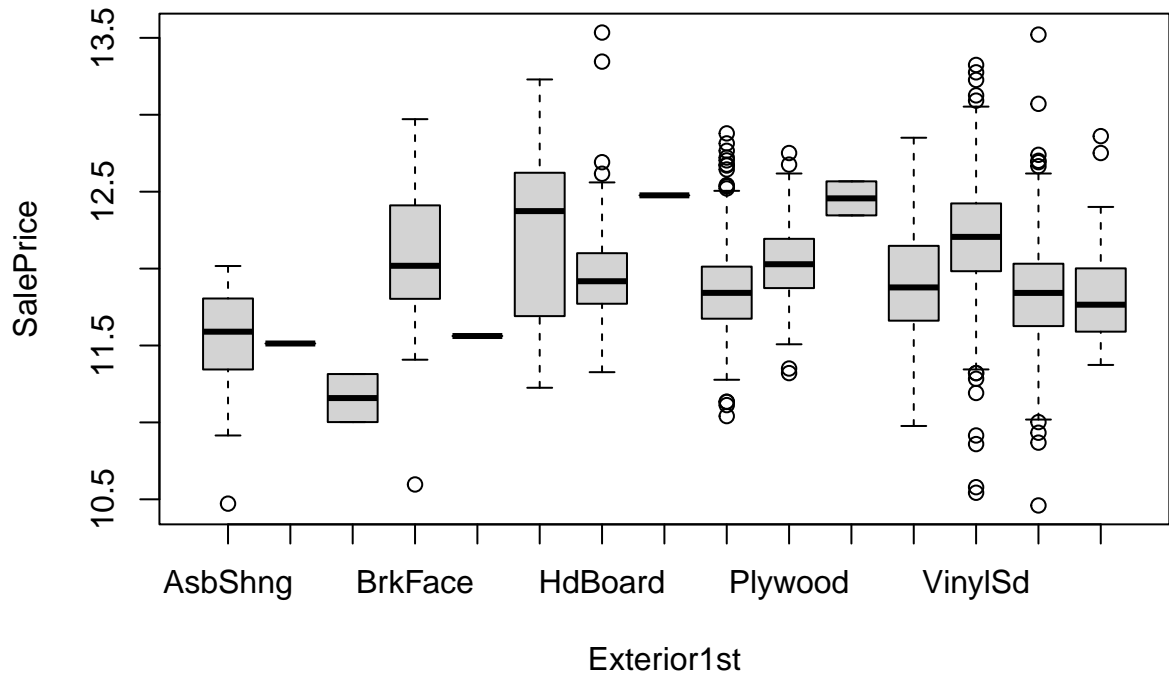


Df	Sum Sq	Mean Sq	F value	Pr(>F)
7	4.167132	0.5953046	3.780645	0.0004504
1452	228.633527	0.1574611	NA	NA

Eta squared: 0.0179

Exterior1st

Numero di NA: 0

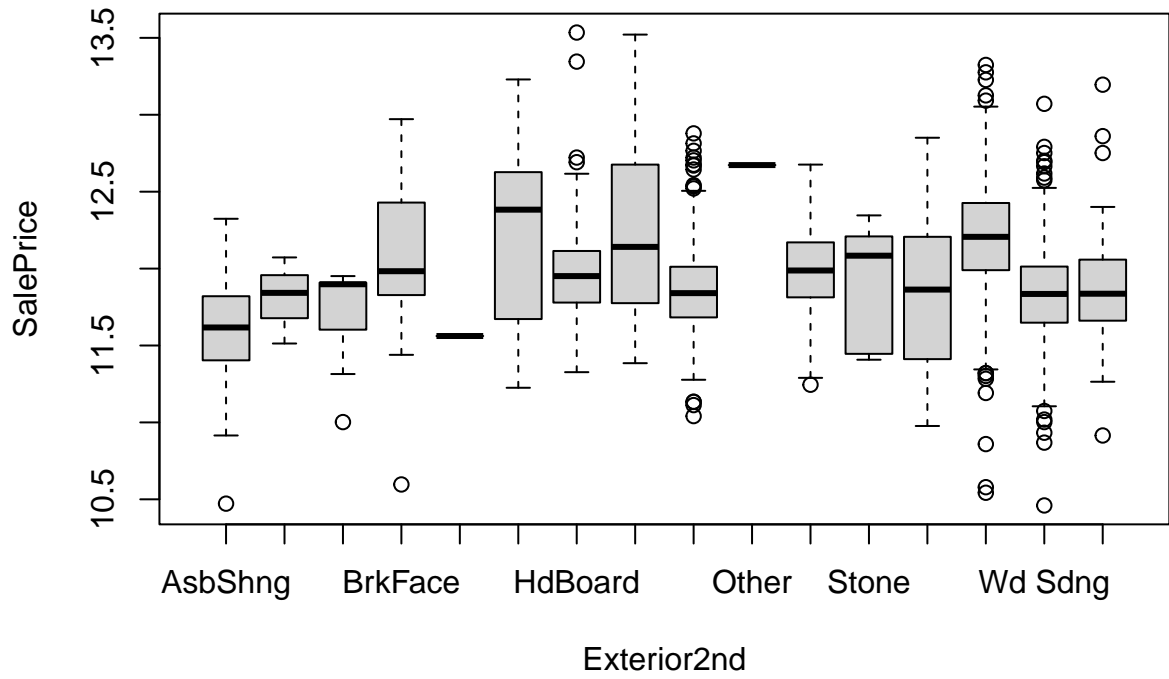


Df	Sum Sq	Mean Sq	F value	Pr(>F)
14	42.26057	3.0186124	22.89227	0
1445	190.54008	0.1318617	NA	NA

Eta squared: 0.1815312

Exterior2nd

Numero di NA: 0

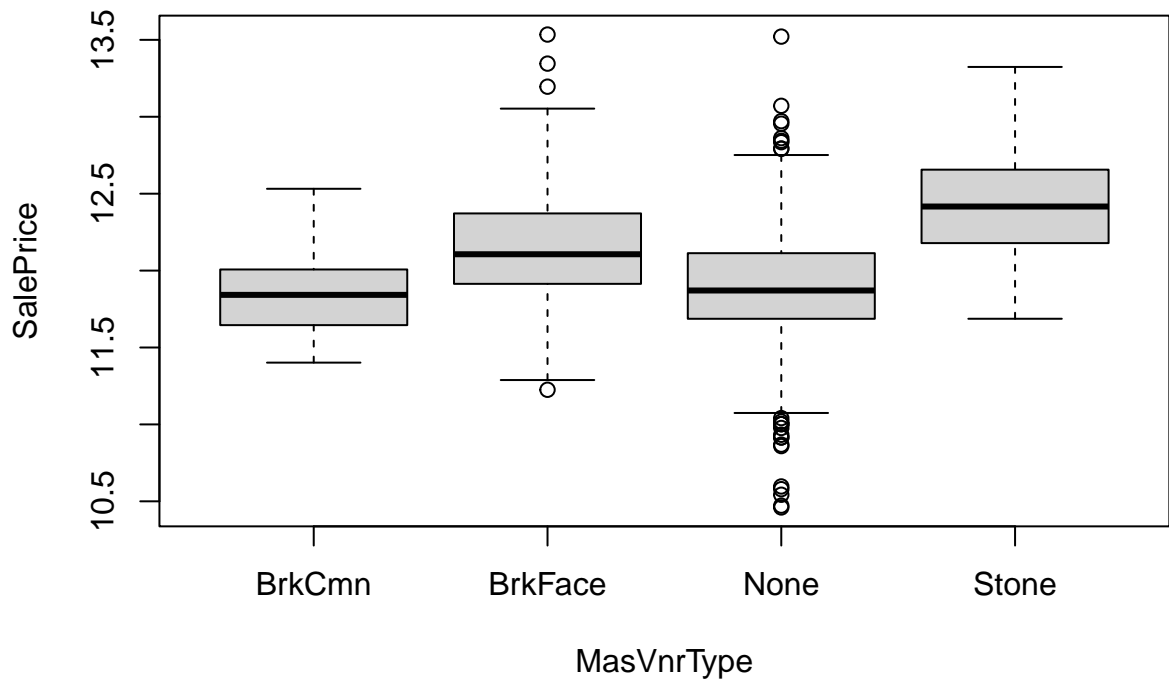


Df	Sum Sq	Mean Sq	F value	Pr(>F)
15	39.91659	2.6611063	19.92201	0
1444	192.88407	0.1335762	NA	NA

Eta squared: 0.1714625

MasVnrType

Numero di NA: 8



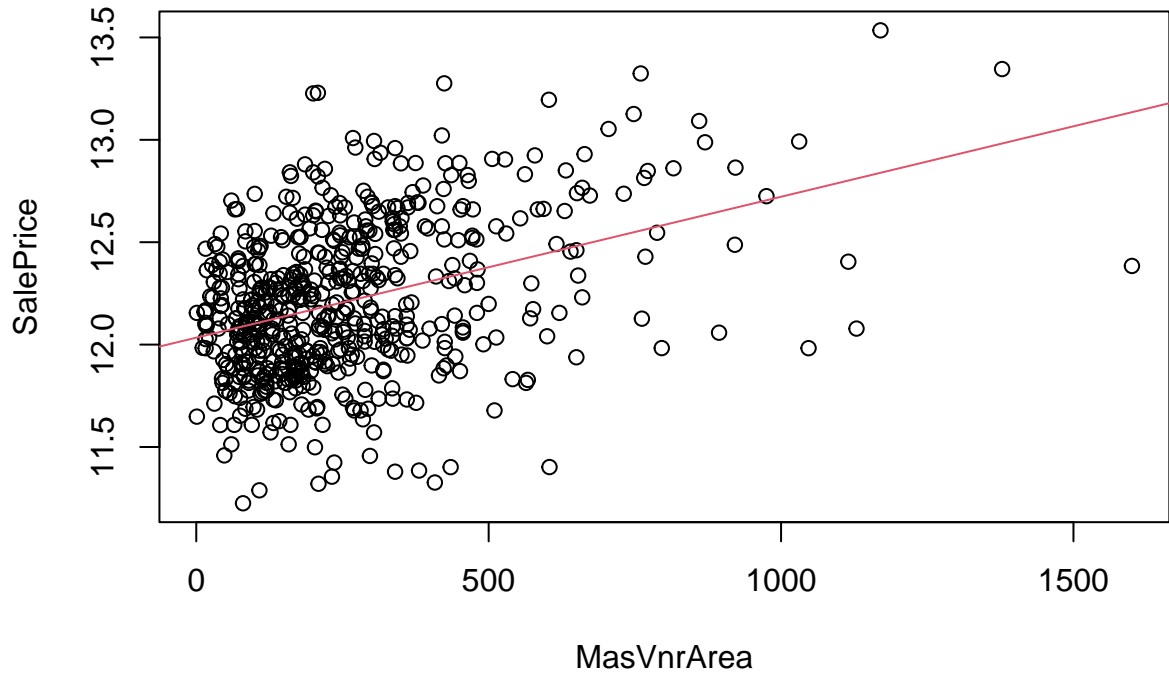
Df	Sum Sq	Mean Sq	F value	Pr(>F)
3	44.5516	14.8505331	115.072	0
1448	186.8706	0.1290543	NA	NA

Eta squared: 0.1925122

MasVnrArea

Numero di NA: 8

Numero di zeri rimossi: 861

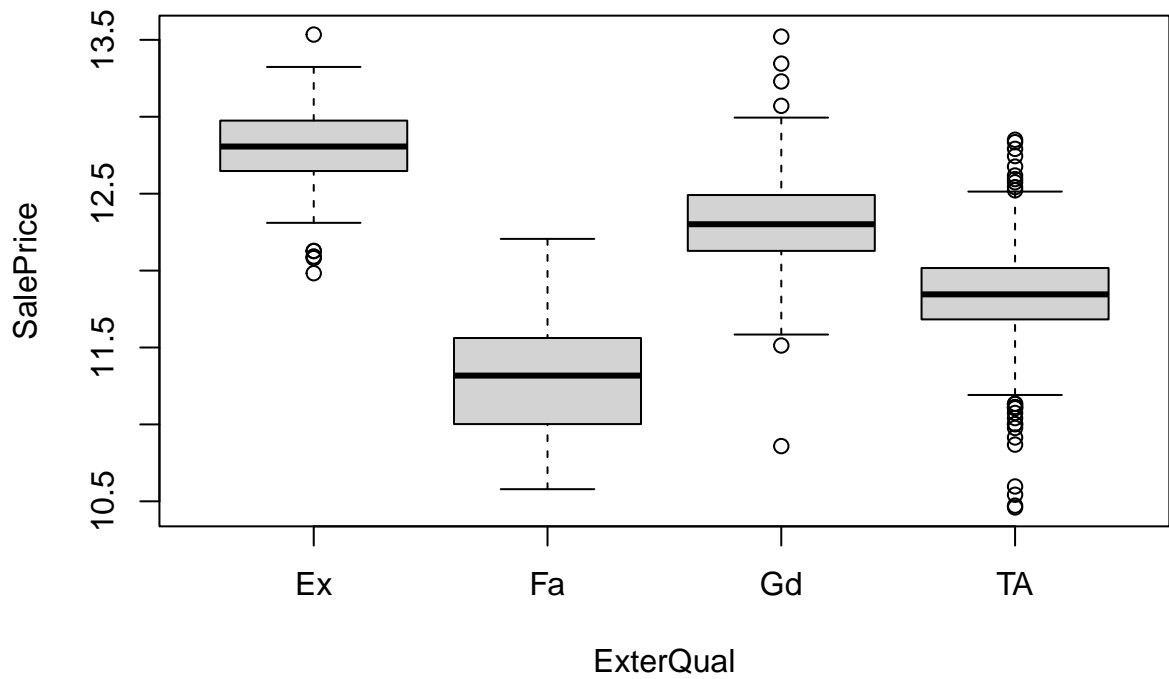


Estimate	Std. Error	t value	Pr(> t)
12.0338613	0.0226276	531.822929	0
0.0006881	0.0000692	9.943031	0

cov	cor	rsq
28.95849	0.379112	0.1437259

ExterQual

Numero di NA: 0

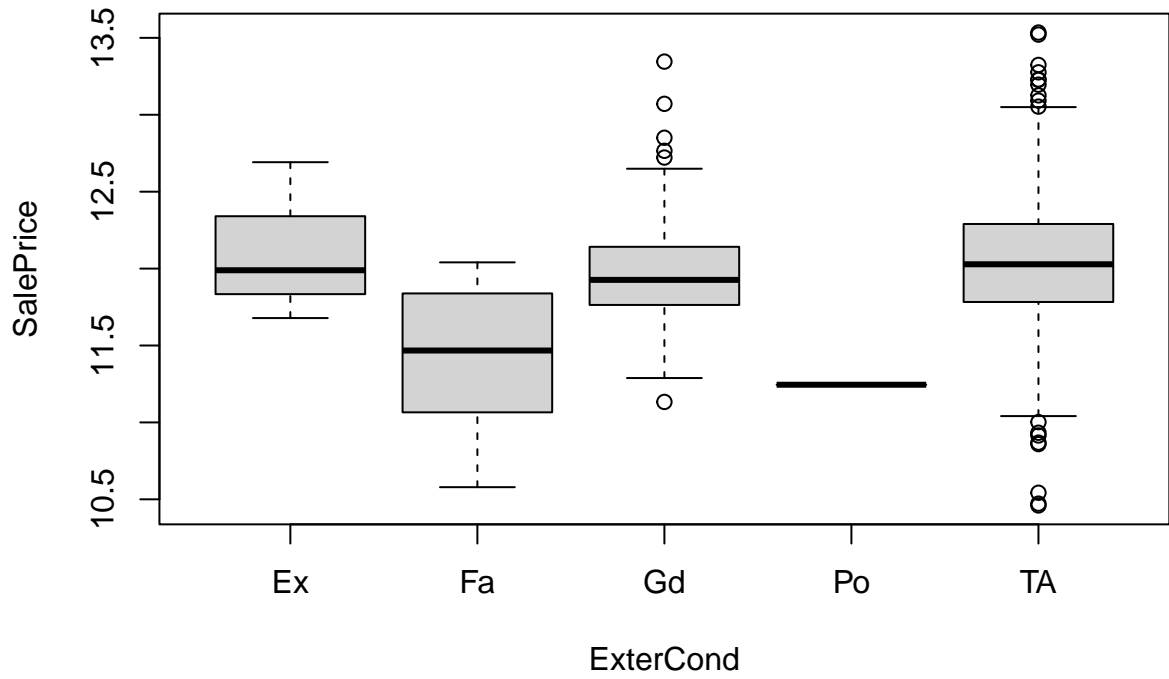


Df	Sum Sq	Mean Sq	F value	Pr(>F)
3	107.3495	35.7831655	415.3034	0
1456	125.4512	0.0861615	NA	NA

Eta squared: 0.461122

ExterCond

Numero di NA: 0

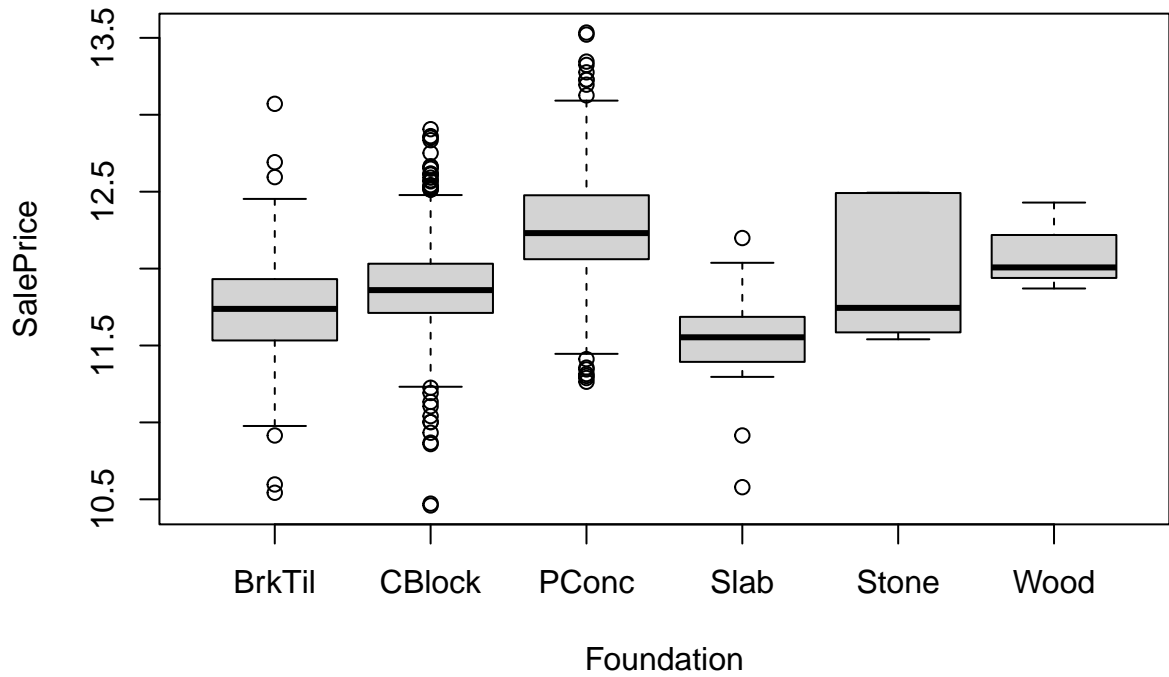


Df	Sum Sq	Mean Sq	F value	Pr(>F)
4	10.59544	2.6488606	17.34474	0
1455	222.20522	0.1527184	NA	NA

Eta squared: 0.04551294

Foundation

Numero di NA: 0

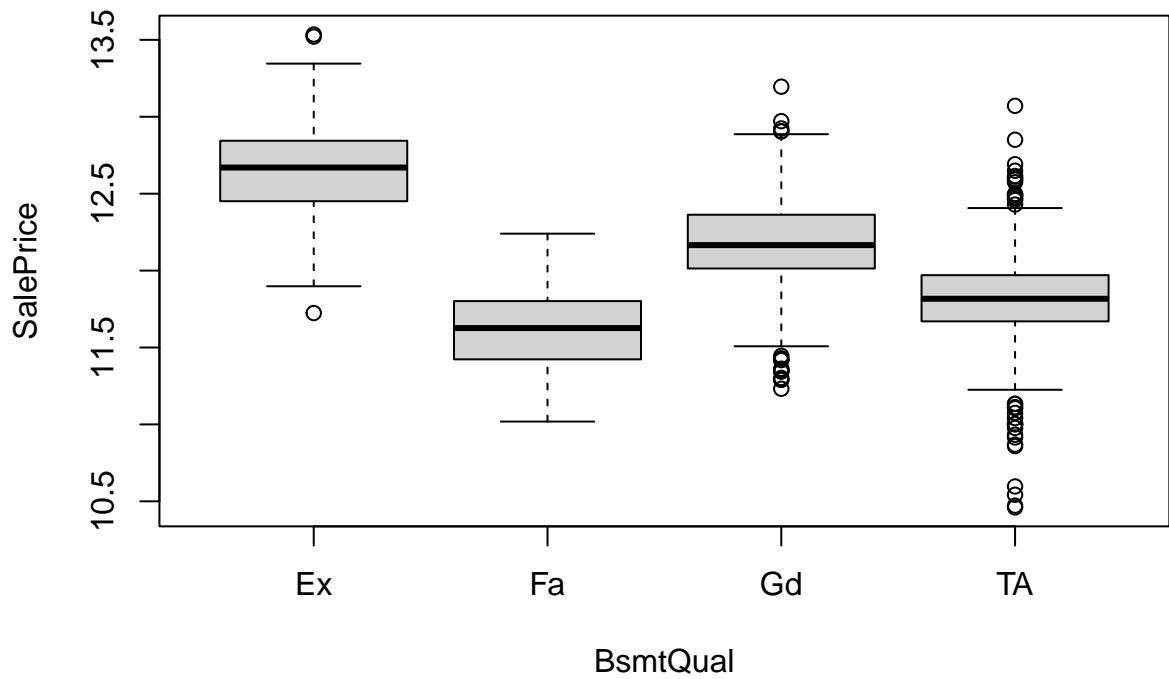


Df	Sum Sq	Mean Sq	F value	Pr(>F)
5	70.69018	14.1380352	126.8067	0
1454	162.11048	0.1114928	NA	NA

Eta squared: 0.3036511

BsmtQual

Numero di NA: 37

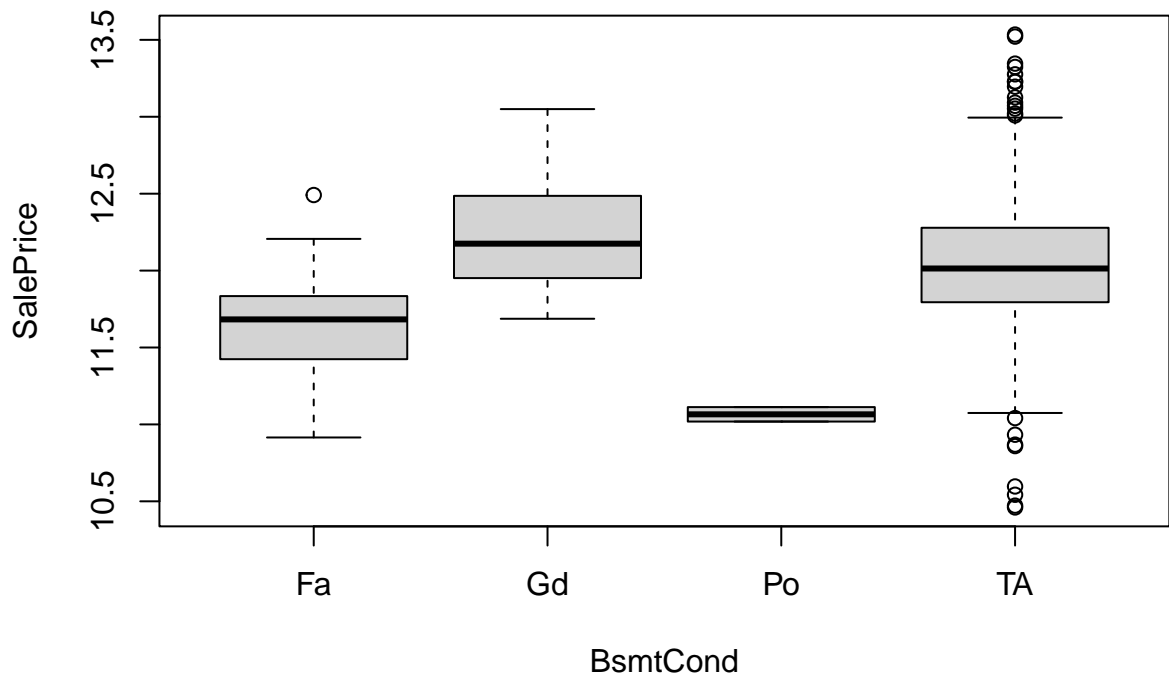


Df	Sum Sq	Mean Sq	F value	Pr(>F)
3	96.01799	32.0059961	364.7272	0
1419	124.52187	0.0877533	NA	NA

Eta squared: 0.435377

BsmtCond

Numero di NA: 37

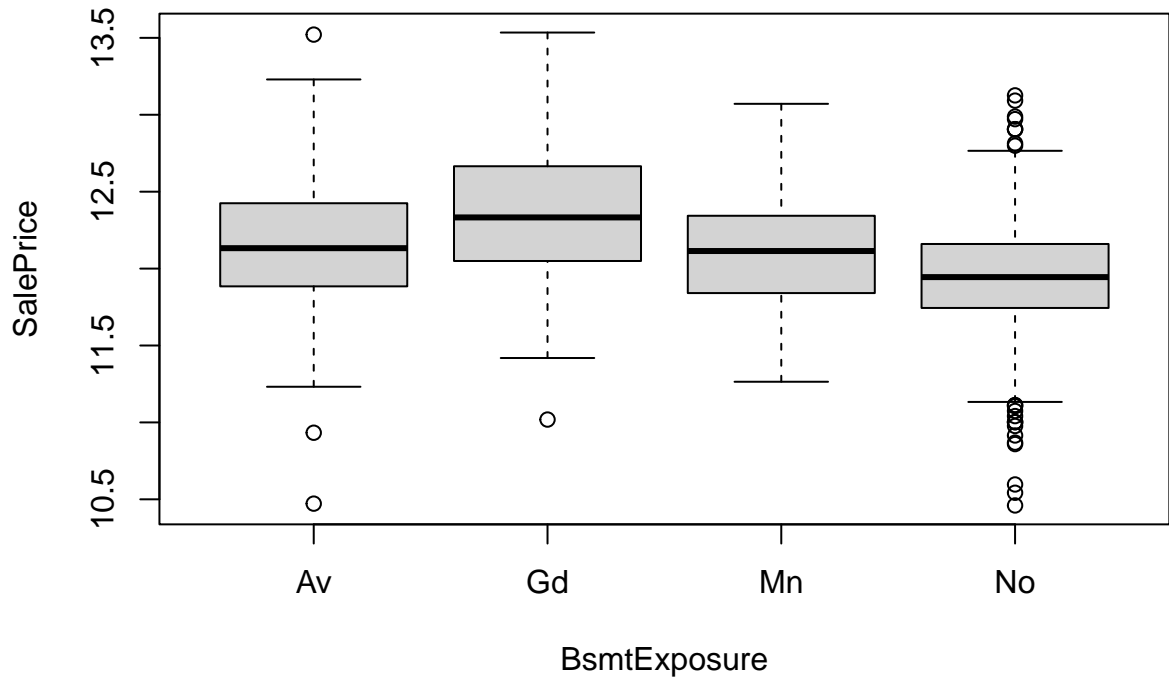


Df	Sum Sq	Mean Sq	F value	Pr(>F)
3	10.89146	3.6304859	24.57285	0
1419	209.64840	0.1477438	NA	NA

Eta squared: 0.04938544

BsmtExposure

Numero di NA: 38

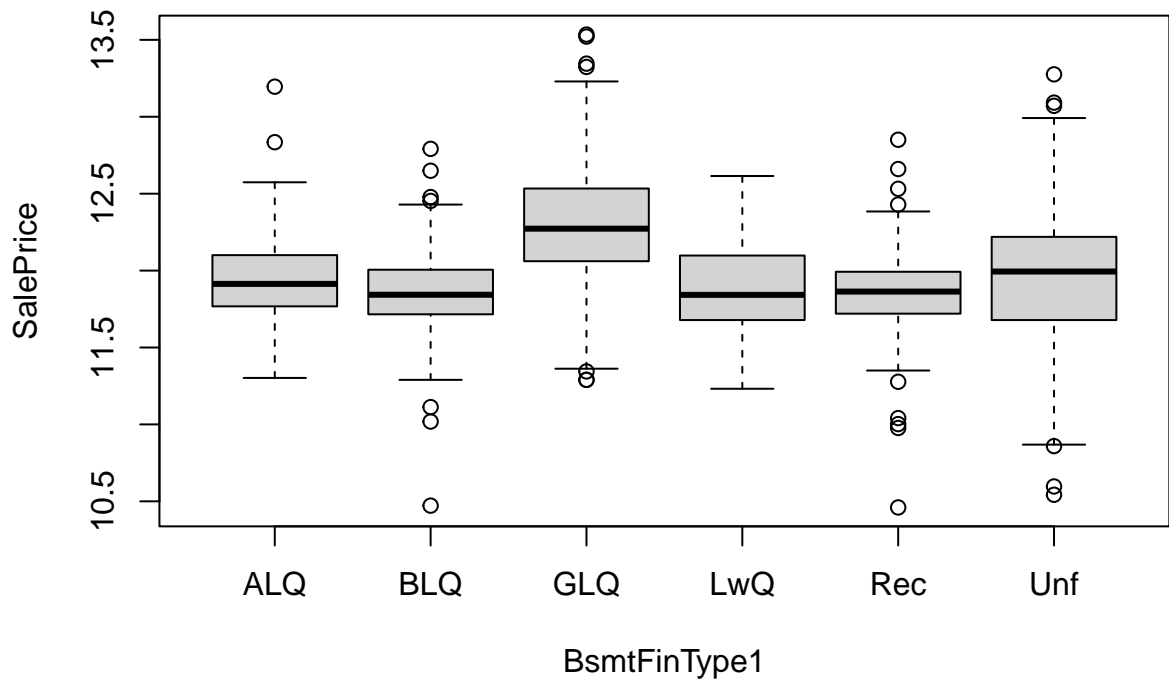


Df	Sum Sq	Mean Sq	F value	Pr(>F)
3	24.29727	8.0990901	58.52712	0
1418	196.22543	0.1383818	NA	NA

Eta squared: 0.1101804

BsmtFinType1

Numero di NA: 37



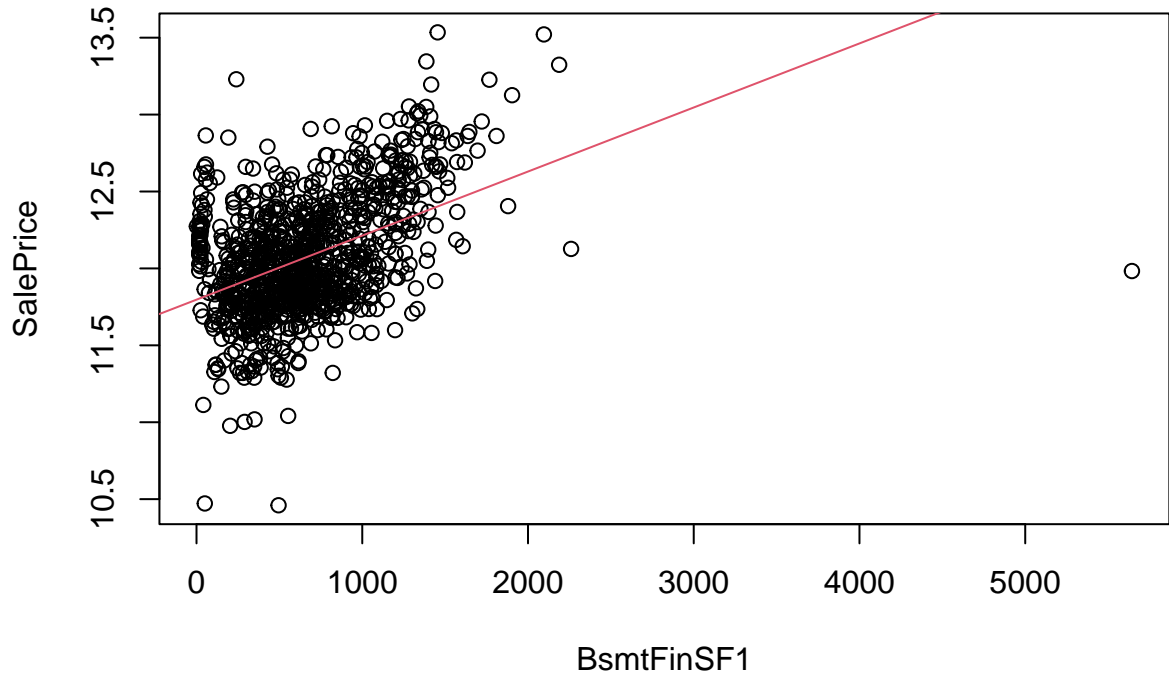
Df	Sum Sq	Mean Sq	F value	Pr(>F)
5	43.72088	8.744177	70.07448	0
1417	176.81898	0.124784	NA	NA

Eta squared: 0.1982448

BsmtFinSF1

Numero di NA: 0

Numero di zeri rimossi: 467

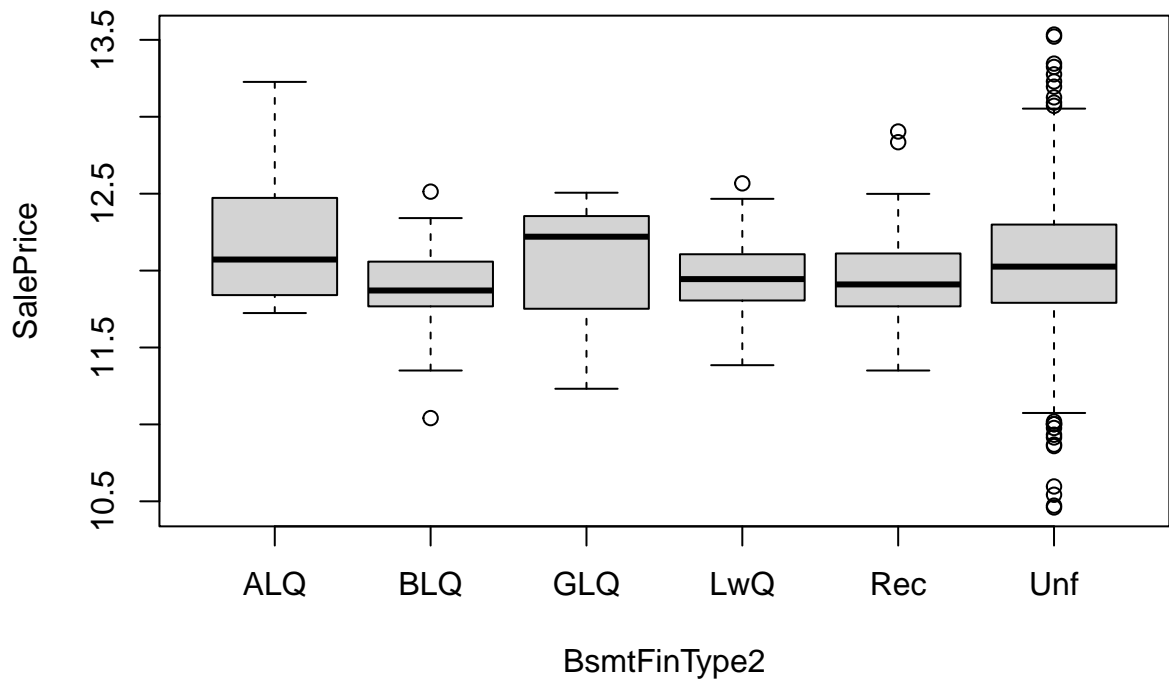


Estimate	Std. Error	t value	Pr(> t)
11.7968493	0.0202810	581.67101	0
0.0004164	0.0000263	15.83867	0

cov	cor	rsq
70.67929	0.4494503	0.2020056

BsmtFinType2

Numero di NA: 38



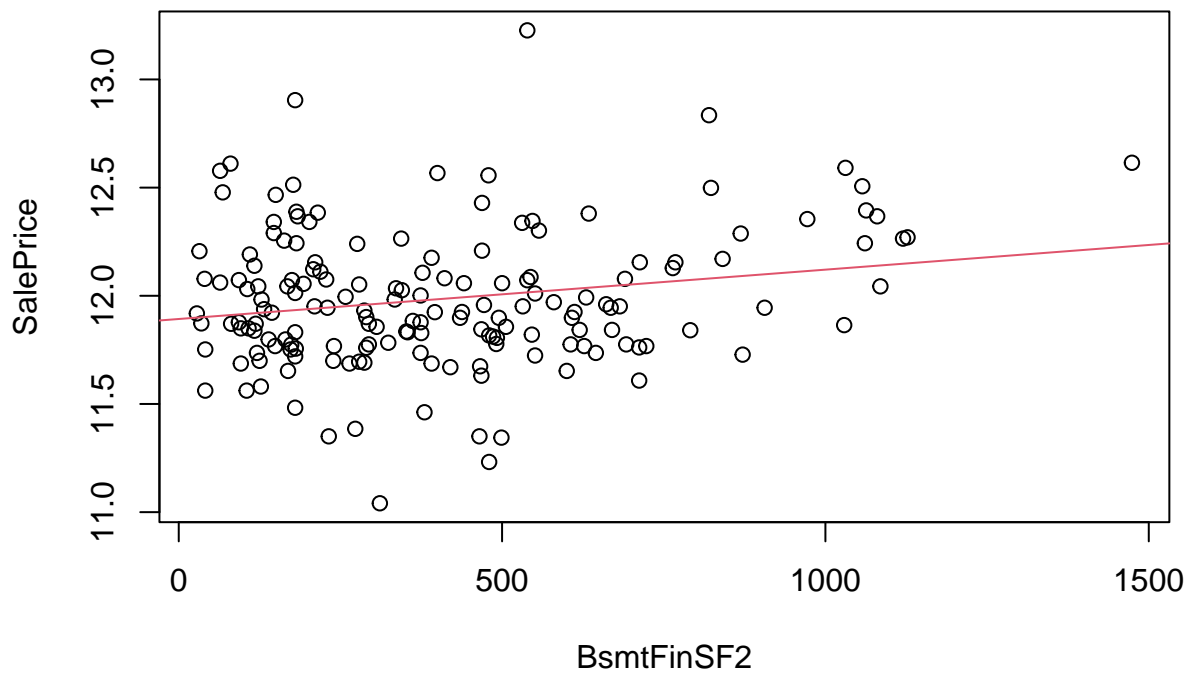
Df	Sum Sq	Mean Sq	F value	Pr(>F)
5	1.55316	0.3106321	2.011076	0.0743771
1416	218.71629	0.1544607	NA	NA

Eta squared: 0.007051184

BsmtFinSF2

Numero di NA: 0

Numero di zeri rimossi: 1293



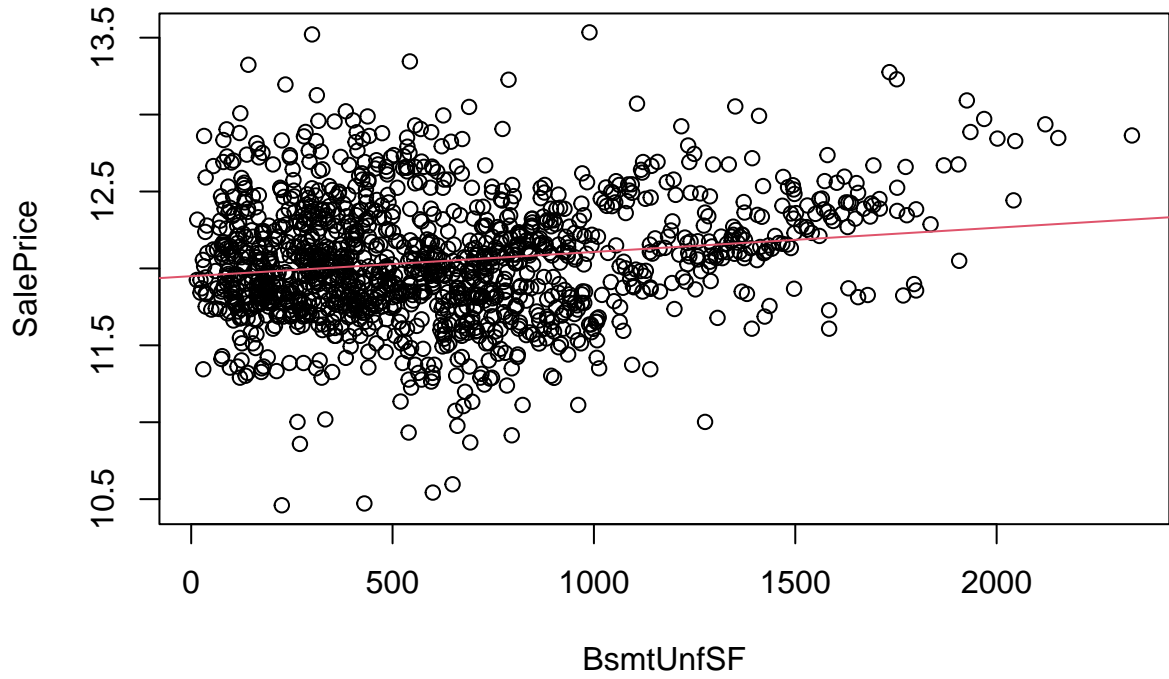
Estimate	Std. Error	t value	Pr(> t)
11.8926979	0.0421459	282.179486	0.0000000
0.0002281	0.0000849	2.685712	0.0079773

cov	cor	rsq
18.51373	0.2046569	0.0418844

BsmtUnfSF

Numero di NA: 0

Numero di zeri rimossi: 118



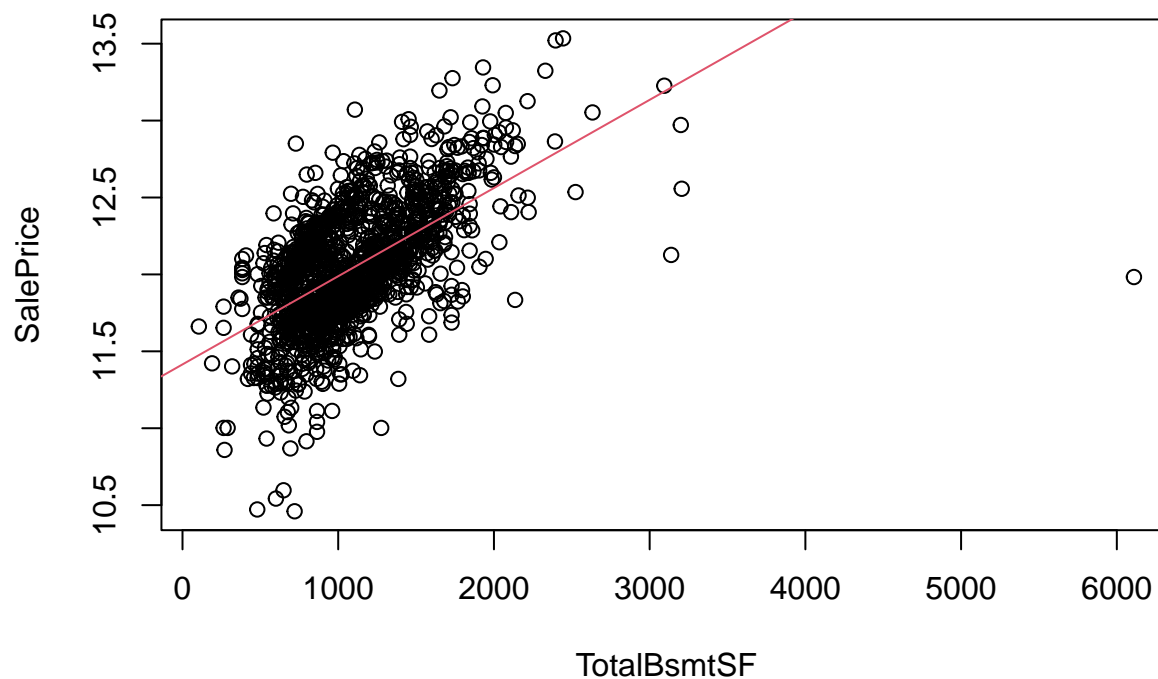
Estimate	Std. Error	t value	Pr(> t)
11.9494533	0.0187359	637.78531	0
0.0001576	0.0000250	6.30853	0

cov	cor	rsq
28.62747	0.1698324	0.028843

TotalBsmtSF

Numero di NA: 0

Numero di zeri rimossi: 37



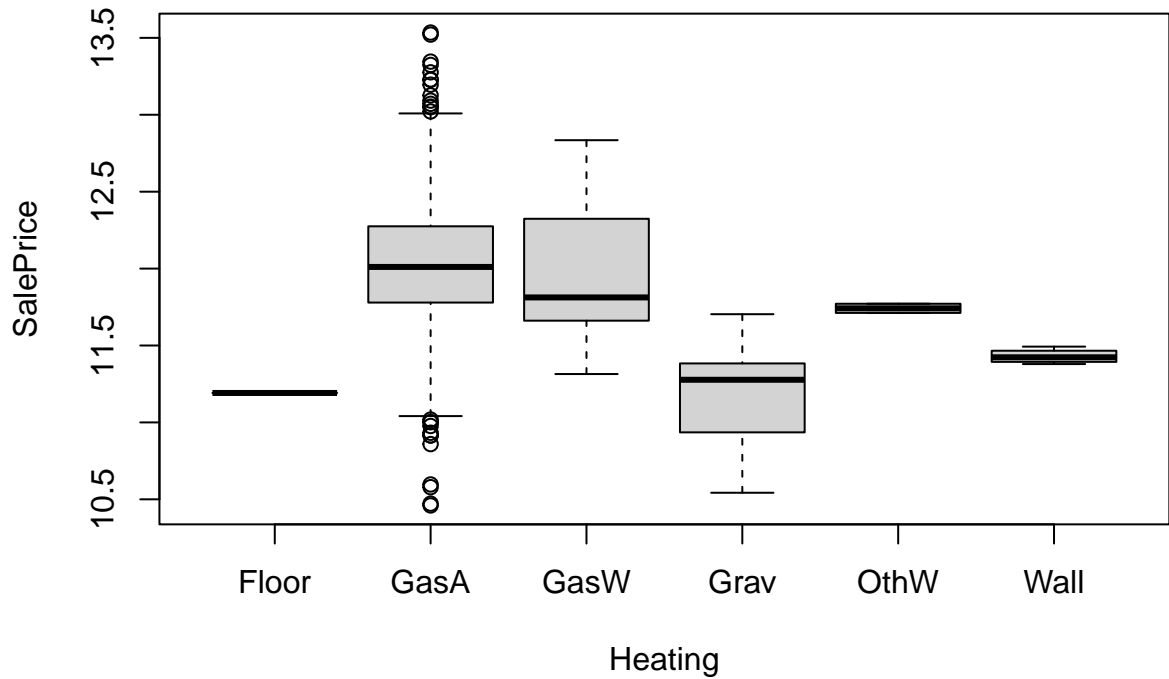
Estimate	Std. Error	t value	Pr(> t)
11.4148506	0.0237577	480.47053	0
0.0005734	0.0000205	27.98441	0

cov	cor	rsq
96.10676	0.5960708	0.3553003

Nel grafico si può notare una relazione positiva tra la superficie totale del seminterrato e il prezzo di vendita. La covarianza è positiva e il coefficiente di correlazione abbastanza alto (0.61). Quindi, le case con i seminterrati più grandi tendono ad avere un prezzo di vendita maggiore

Heating

Numero di NA: 0

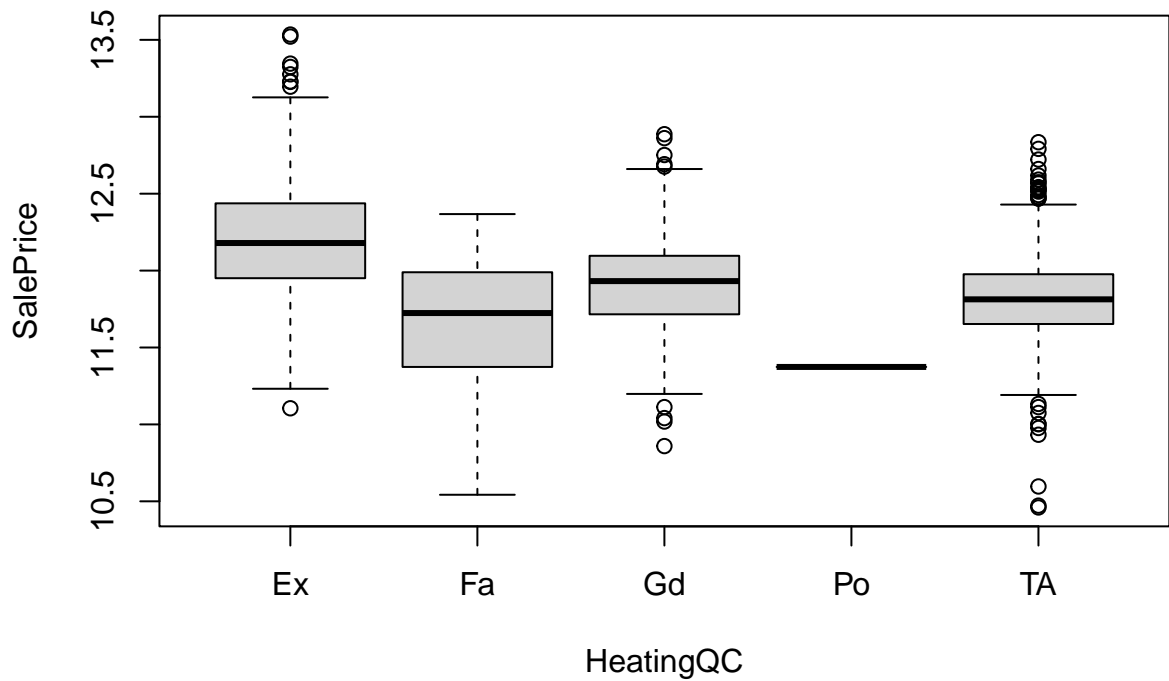


Df	Sum Sq	Mean Sq	F value	Pr(>F)
5	7.669813	1.5339627	9.907046	0
1454	225.130846	0.1548355	NA	NA

Eta squared: 0.03294584

HeatingQC

Numero di NA: 0

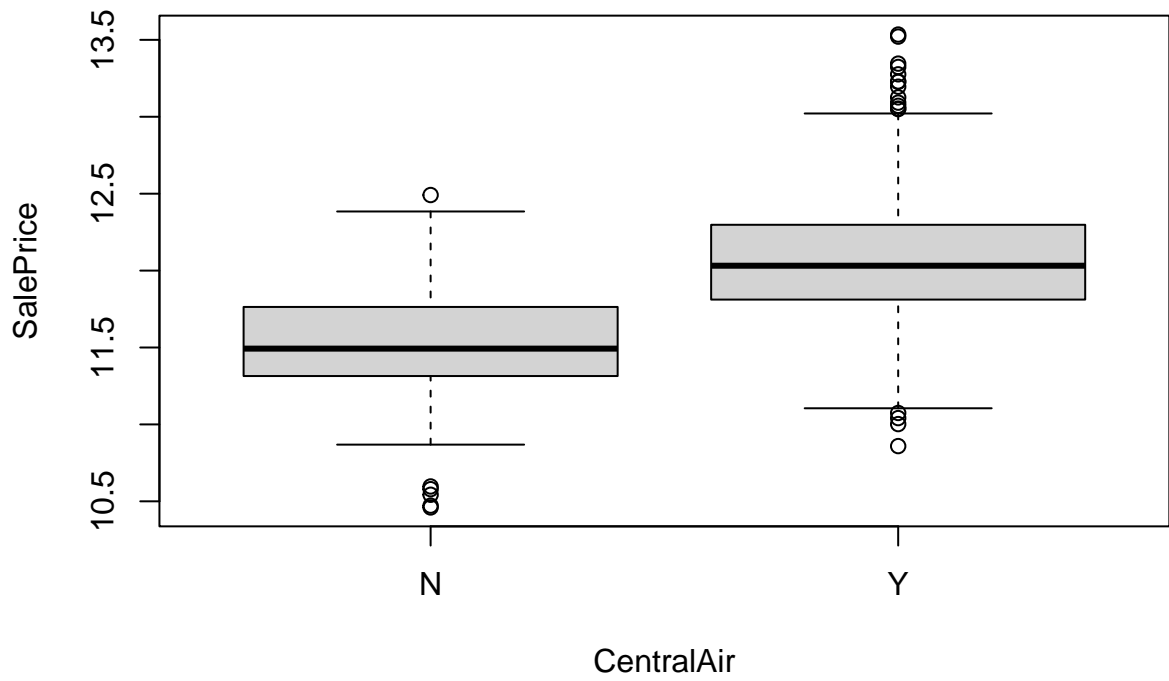


Df	Sum Sq	Mean Sq	F value	Pr(>F)
4	54.3630	13.5907489	110.8204	0
1455	178.4377	0.1226376	NA	NA

Eta squared: 0.2335174

CentralAir

Numero di NA: 0

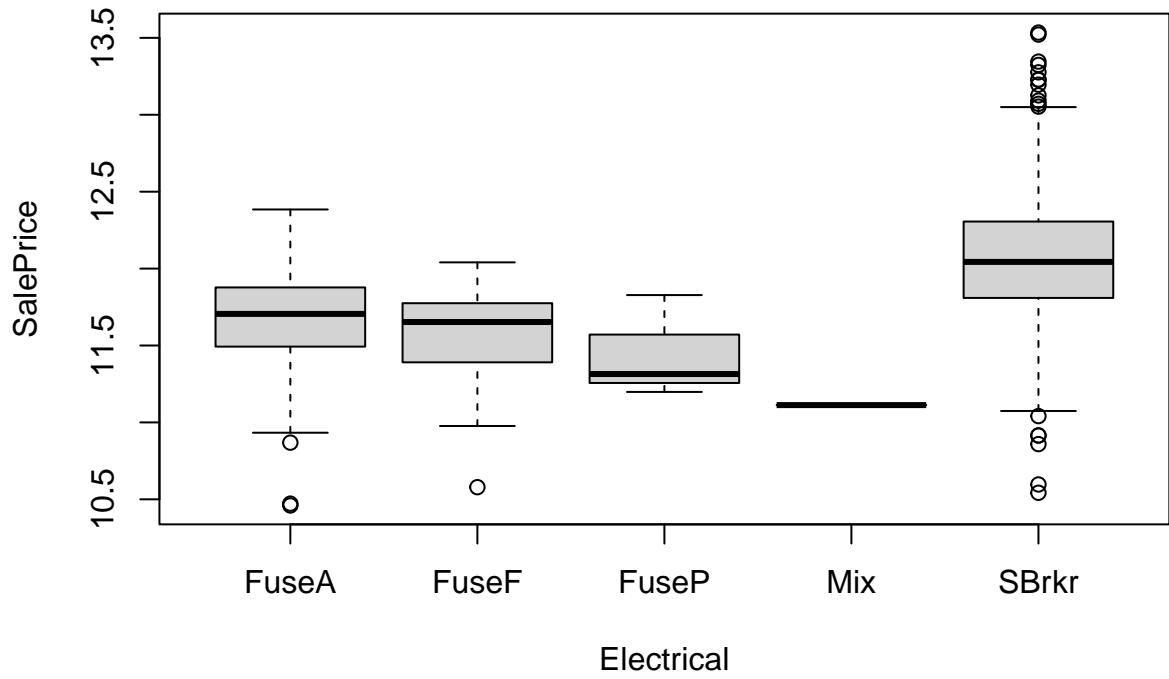


Df	Sum Sq	Mean Sq	F value	Pr(>F)
1	28.77944	28.7794435	205.667	0
1458	204.02122	0.1399322	NA	NA

Eta squared: 0.1236227

Electrical

Numero di NA: 1

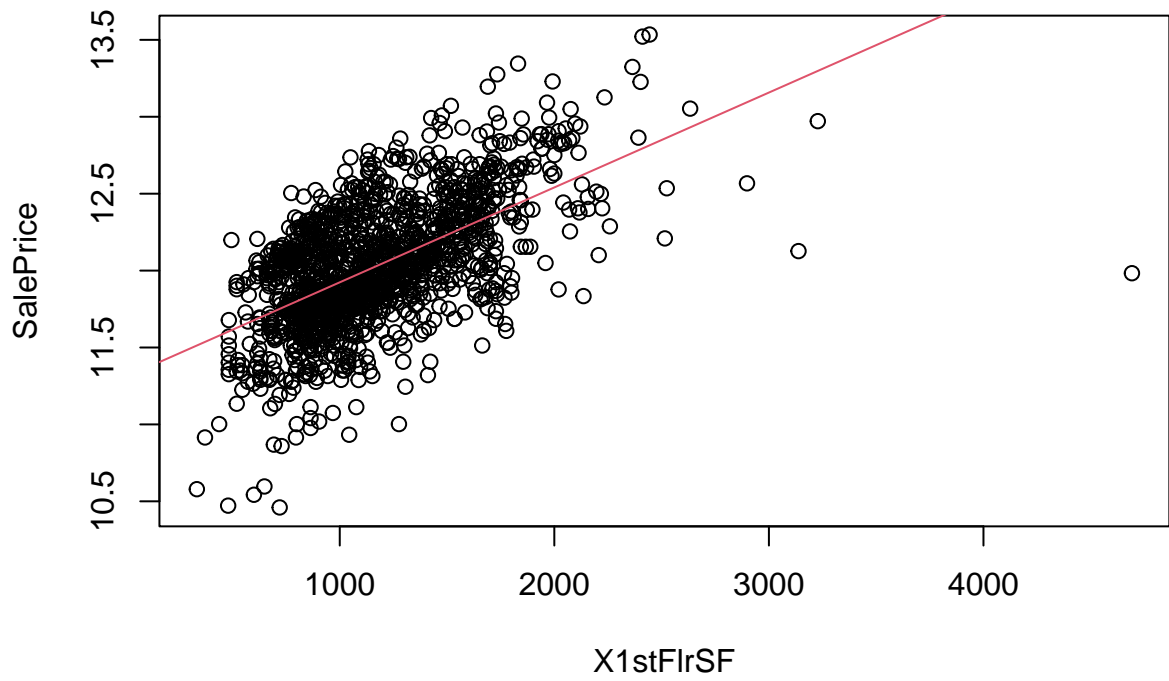


Df	Sum Sq	Mean Sq	F value	Pr(>F)
4	22.47053	5.6176318	38.83437	0
1454	210.33011	0.1446562	NA	NA

Eta squared: 0.09652262

X1stFlrSF

Numero di NA: 0



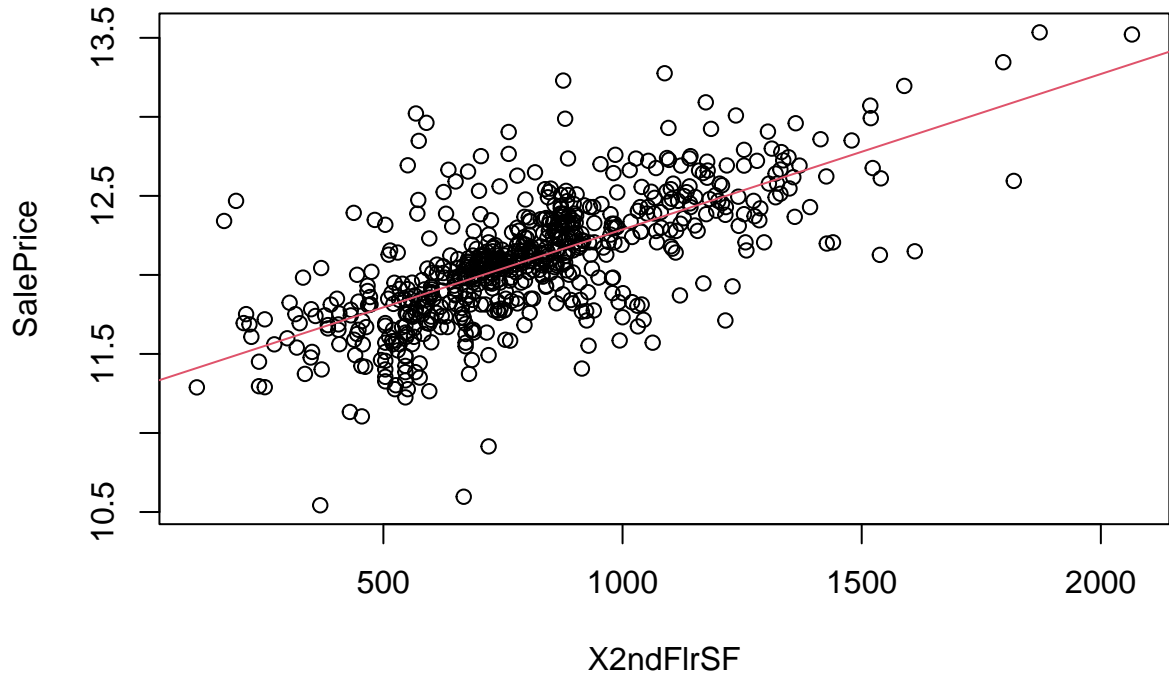
Estimate	Std. Error	t value	Pr(> t)
11.3068890	0.0265979	425.10432	0
0.0006168	0.0000217	28.41363	0

cov	cor	rsq
92.18772	0.596981	0.3563864

X2ndFlrSF

Numero di NA: 0

Numero di zeri rimossi: 829



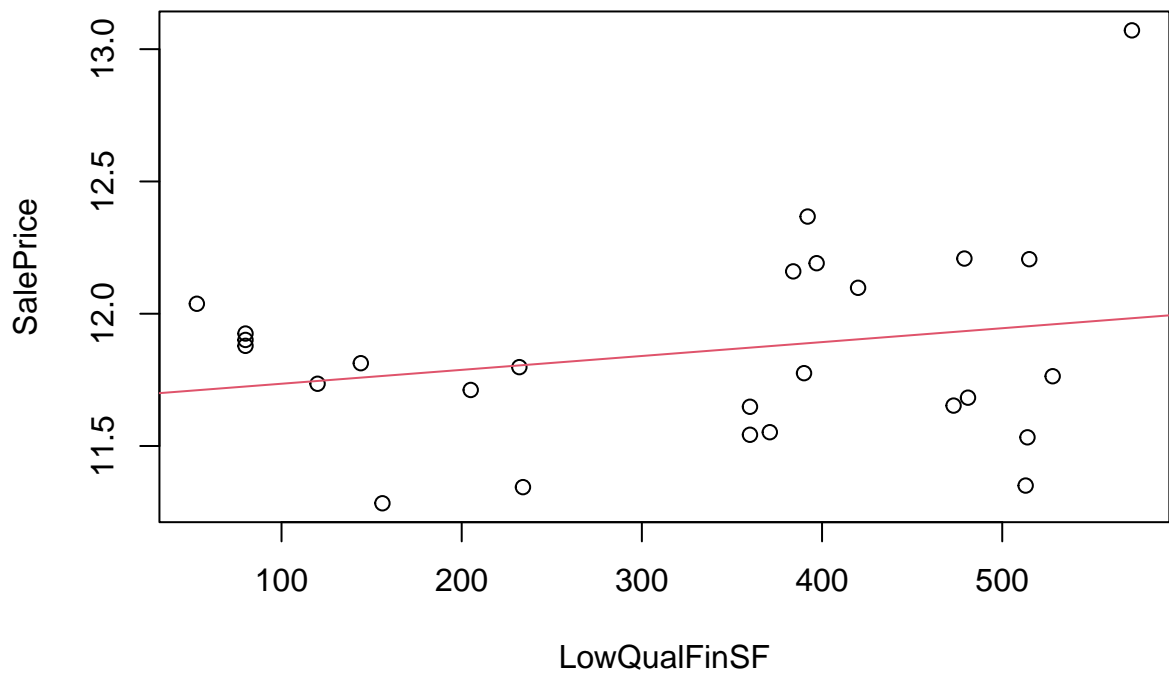
Estimate	Std. Error	t value	Pr(> t)
11.3028540	0.0356718	316.85720	0
0.0009841	0.0000421	23.39586	0

cov	cor	rsq
73.53069	0.682131	0.4653026

LowQualFinSF

Numero di NA: 0

Numero di zeri rimossi: 1434

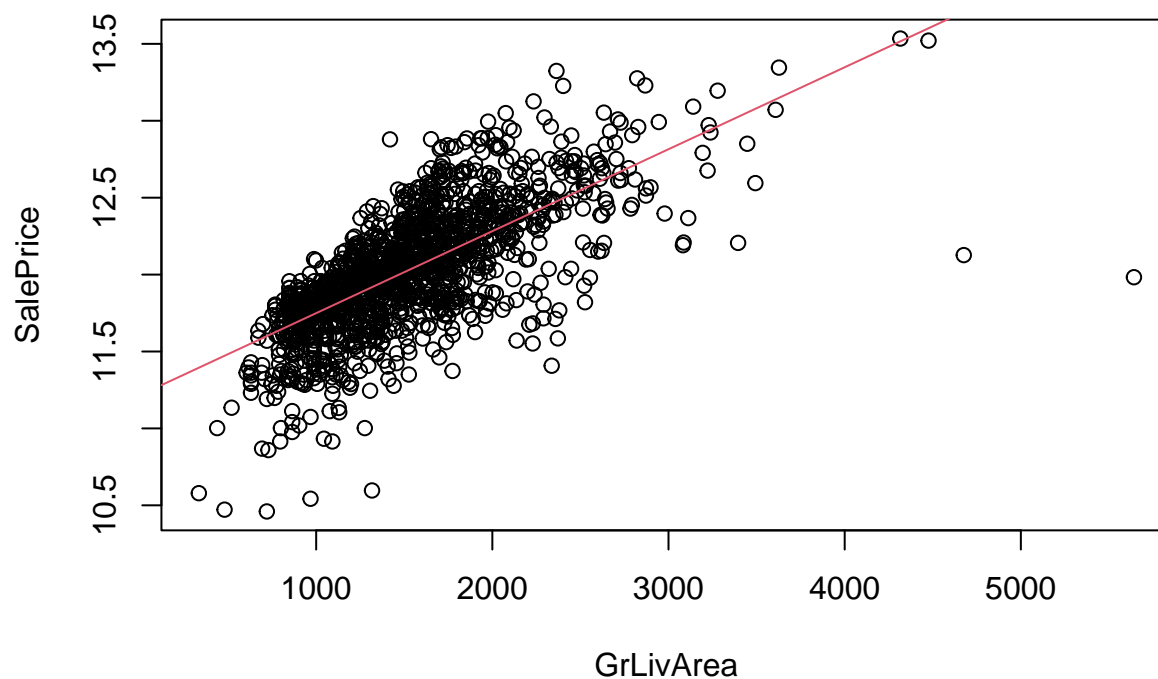


Estimate	Std. Error	t value	Pr(> t)
11.6829123	0.1653328	70.663000	0.0000000
0.0005244	0.0004507	1.163702	0.2559839

cov	cor	rsq
14.65888	0.231109	0.0534114

GrLivArea

Numero di NA: 0



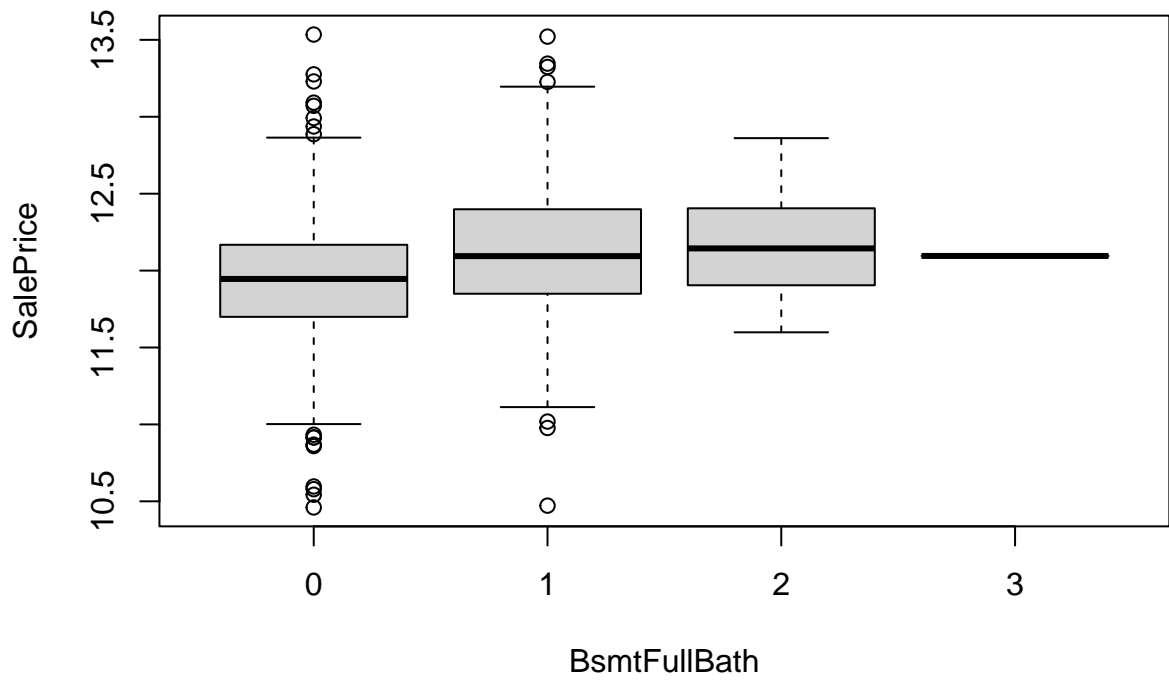
Estimate	Std. Error	t value	Pr(> t)
11.2165815	0.0227743	492.51110	0
0.0005328	0.0000142	37.52491	0

cov	cor	rsq
147.1274	0.7009267	0.4912982

Dal grafico si può vedere che c'è una forte relazione positiva tra l'area abitabile sopra il suolo e il prezzo di vendita, la covarianza è positiva e il coefficiente di correlazione è alto (0.71), indicando una forte correlazione positiva. Case con maggiore area abitabile tendono ad avere prezzi di vendita più elevati

BsmtFullBath

Numero di NA: 0

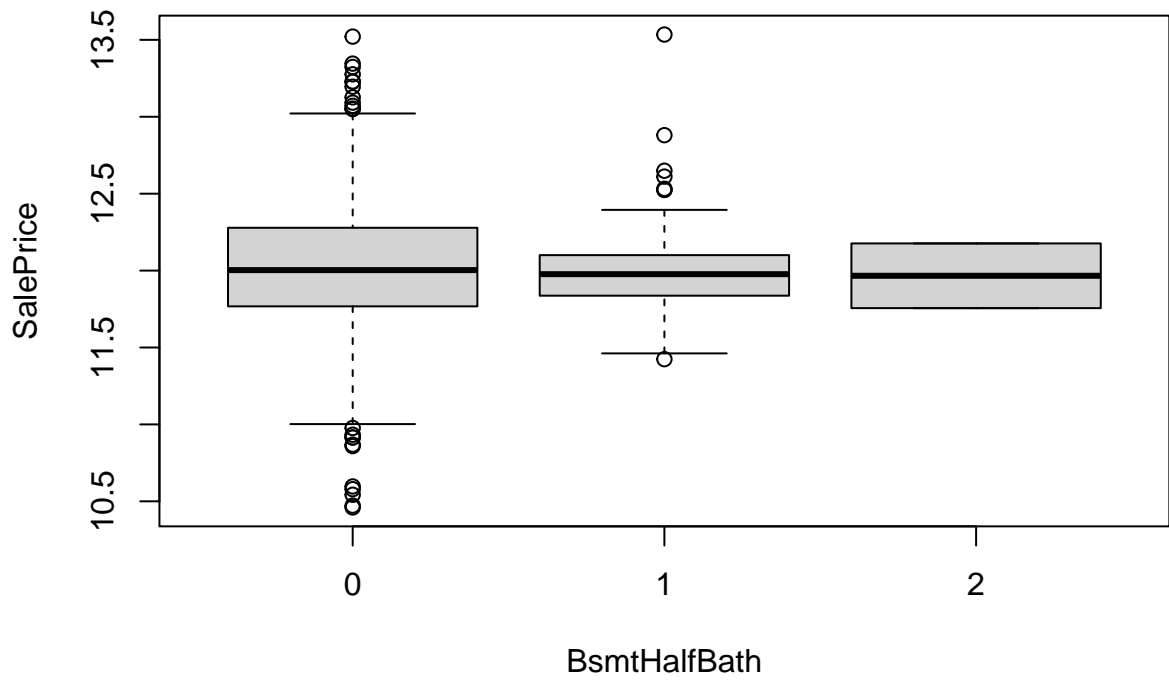


Df	Sum Sq	Mean Sq	F value	Pr(>F)
1	12.9907	12.9907027	86.16736	0
1458	219.8100	0.1507613	NA	NA

Eta squared: 0.05580183

BsmtHalfBath

Numero di NA: 0

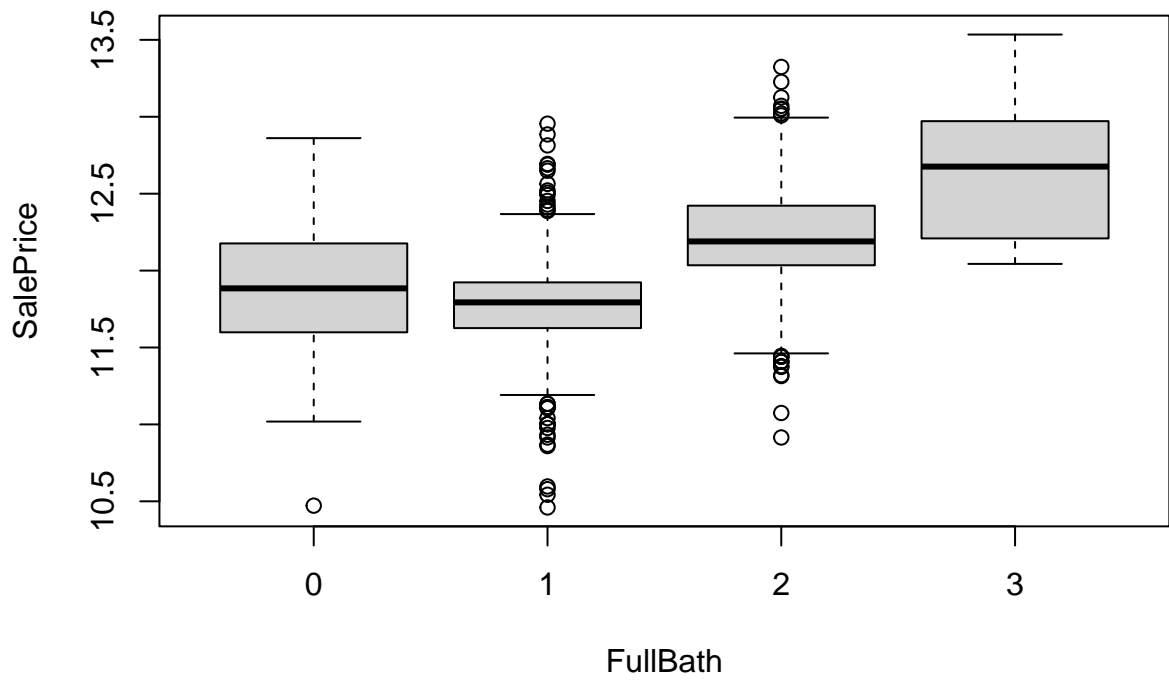


Df	Sum Sq	Mean Sq	F value	Pr(>F)
1	0.0061723	0.0061723	0.0386572	0.8441567
1458	232.7944867	0.1596670	NA	NA

Eta squared: 2.651317e-05

FullBath

Numero di NA: 0

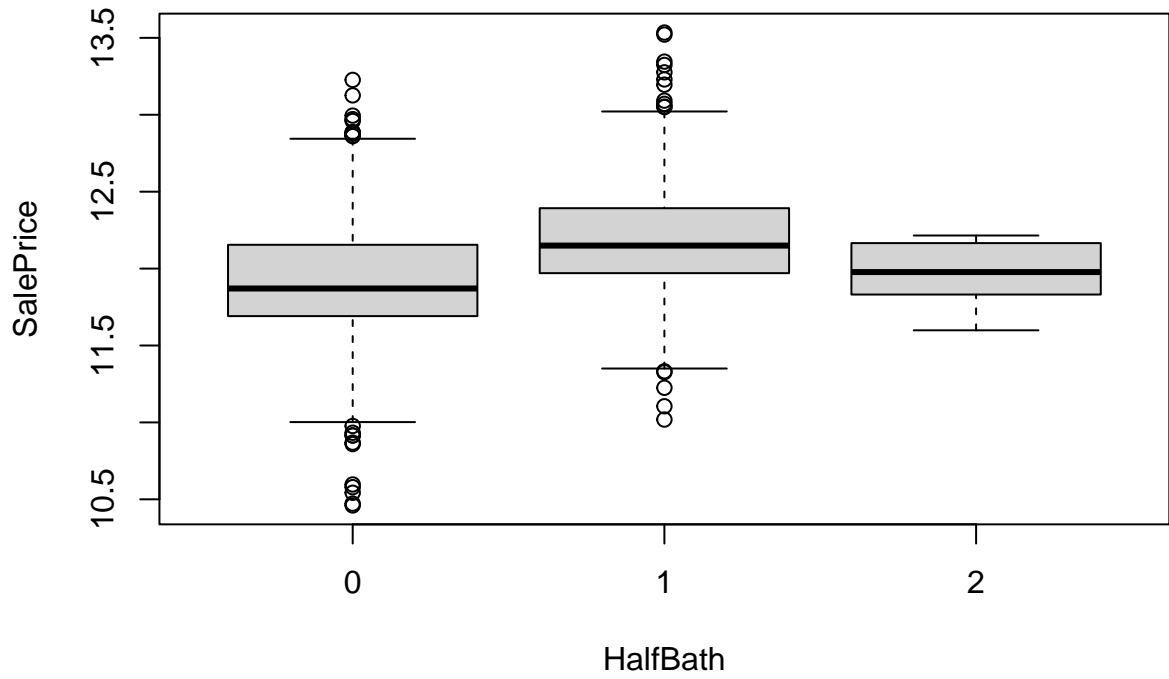


Df	Sum Sq	Mean Sq	F value	Pr(>F)
1	82.3537	82.3536982	798.0998	0
1458	150.4470	0.1031872	NA	NA

Eta squared: 0.353752

HalfBath

Numero di NA: 0

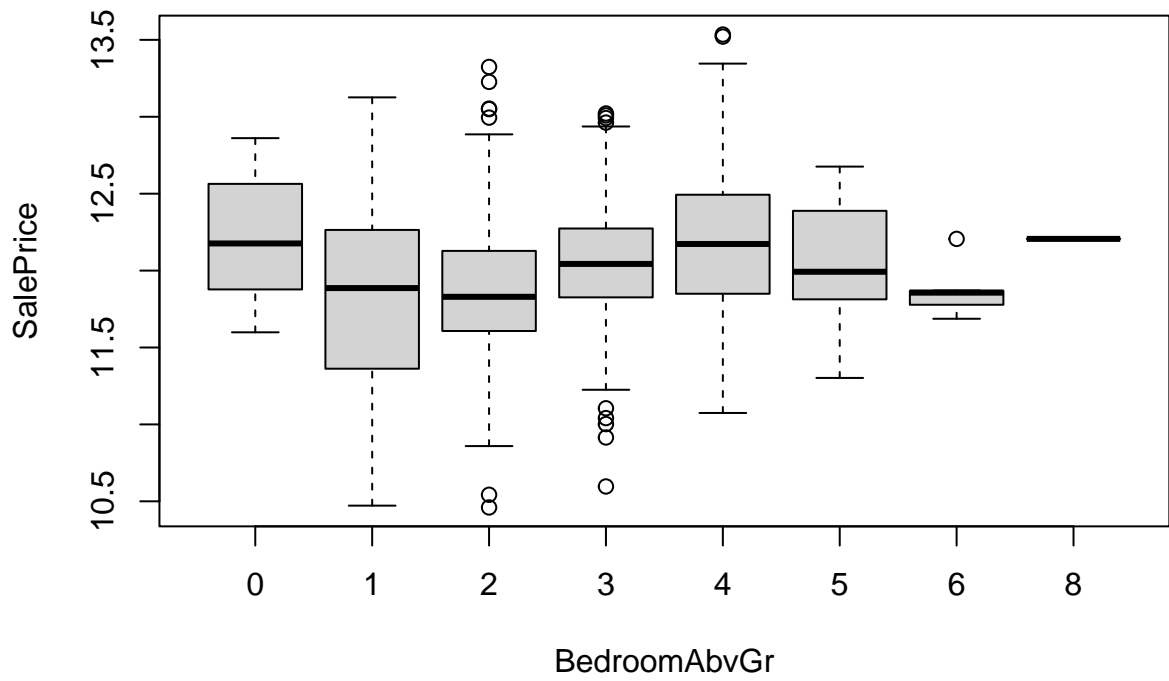


Df	Sum Sq	Mean Sq	F value	Pr(>F)
1	22.95062	22.9506188	159.4567	0
1458	209.85004	0.1439301	NA	NA

Eta squared: 0.09858485

BedroomAbvGr

Numero di NA: 0

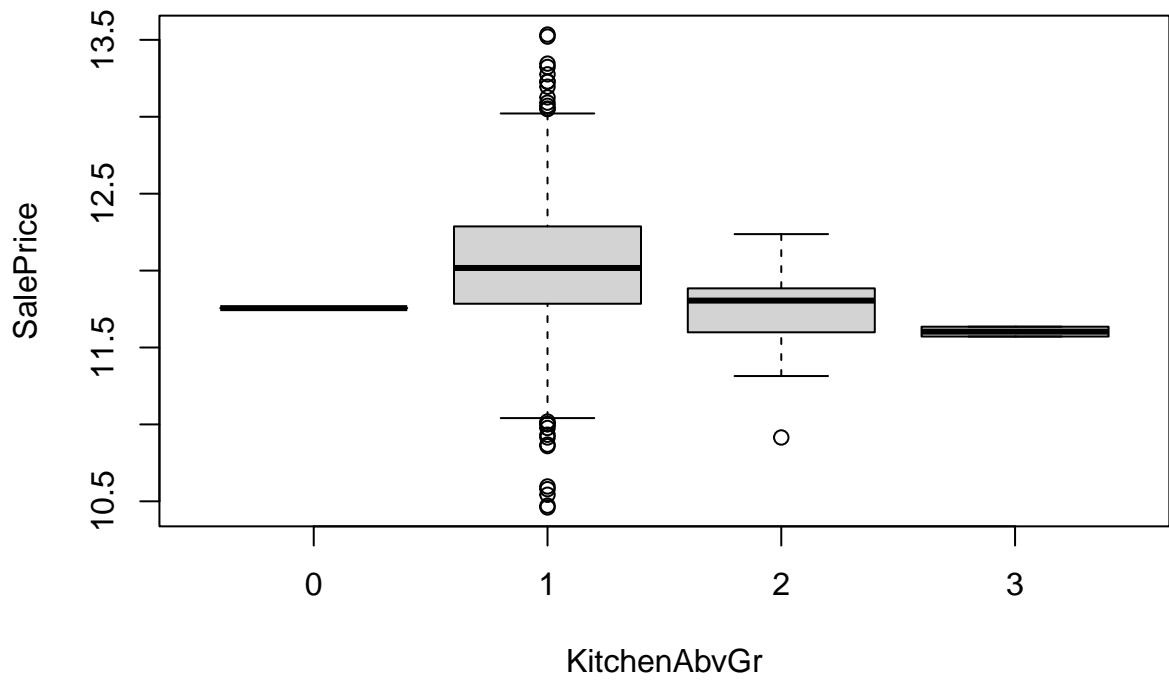


Df	Sum Sq	Mean Sq	F value	Pr(>F)
1	10.17322	10.1732168	66.62499	0
1458	222.62744	0.1526937	NA	NA

Eta squared: 0.04369926

KitchenAbvGr

Numero di NA: 0

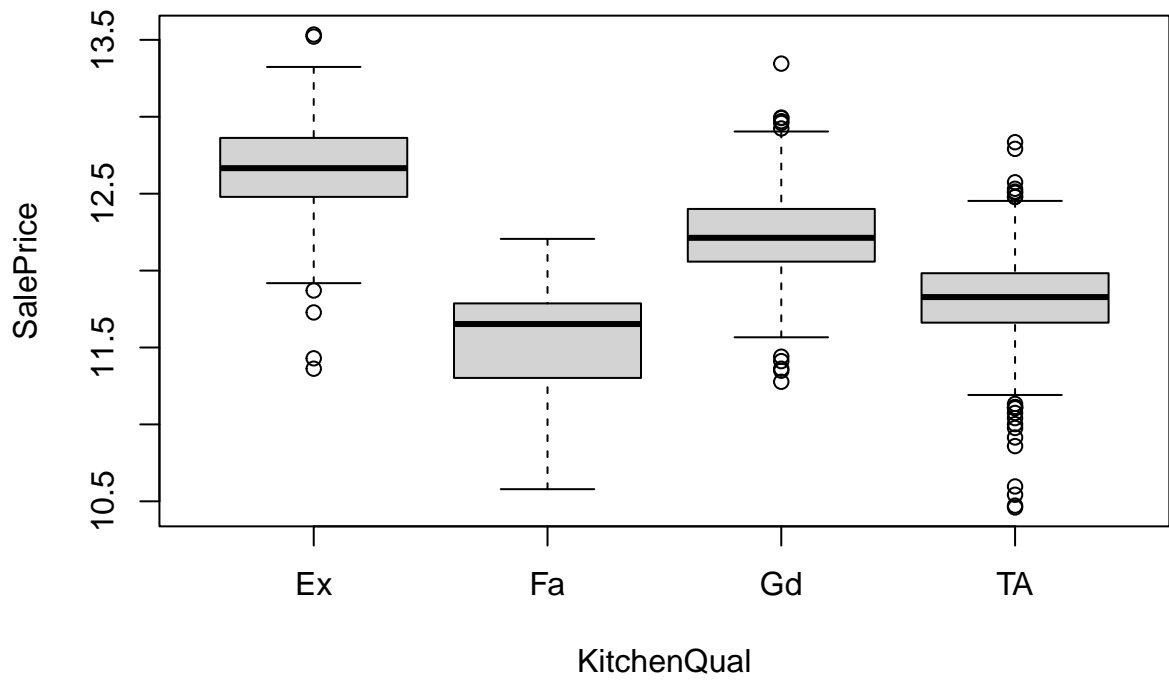


Df	Sum Sq	Mean Sq	F value	Pr(>F)
1	5.068172	5.0681724	32.4477	0
1458	227.732487	0.1561951	NA	NA

Eta squared: 0.02177044

KitchenQual

Numero di NA: 0

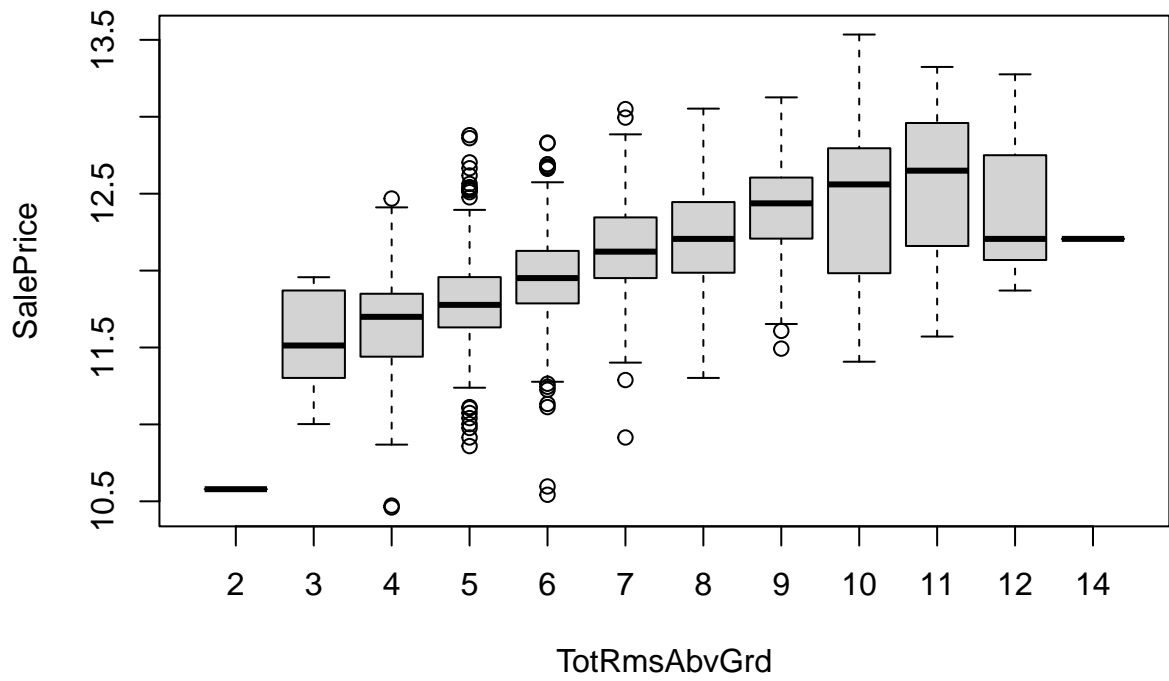


Df	Sum Sq	Mean Sq	F value	Pr(>F)
3	104.2108	34.7369385	393.3202	0
1456	128.5898	0.0883172	NA	NA

Eta squared: 0.4476397

TotRmsAbvGrd

Numero di NA: 0

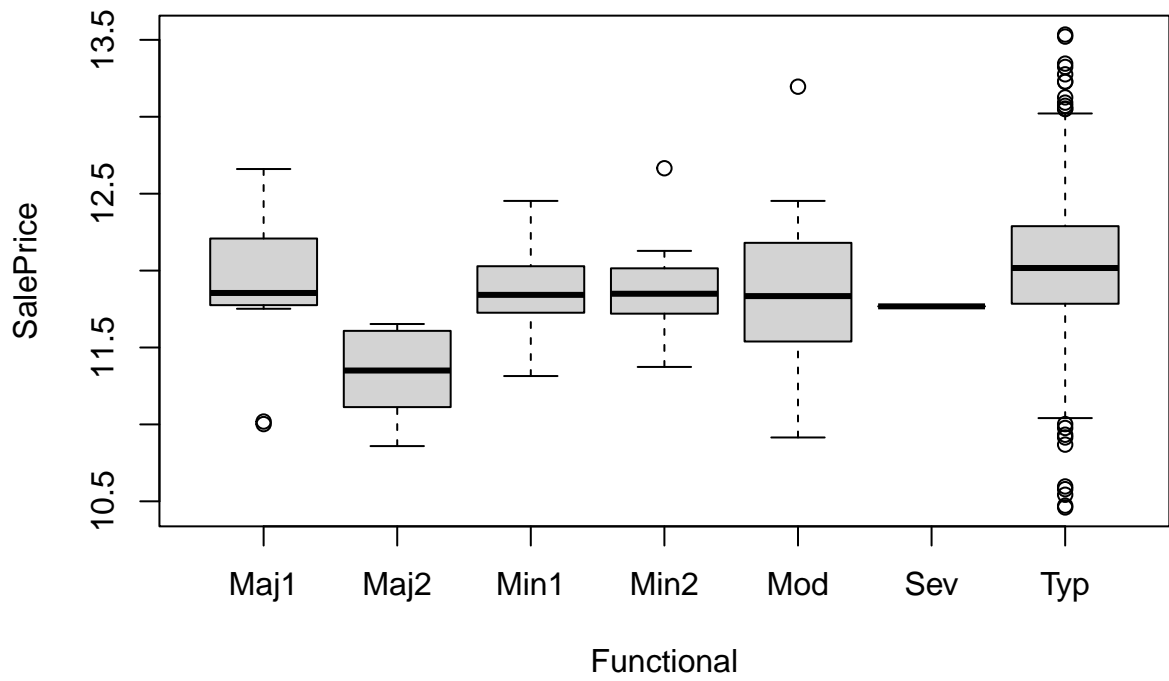


Df	Sum Sq	Mean Sq	F value	Pr(>F)
1	66.48952	66.489519	582.8937	0
1458	166.31114	0.114068	NA	NA

Eta squared: 0.2856071

Functional

Numero di NA: 0

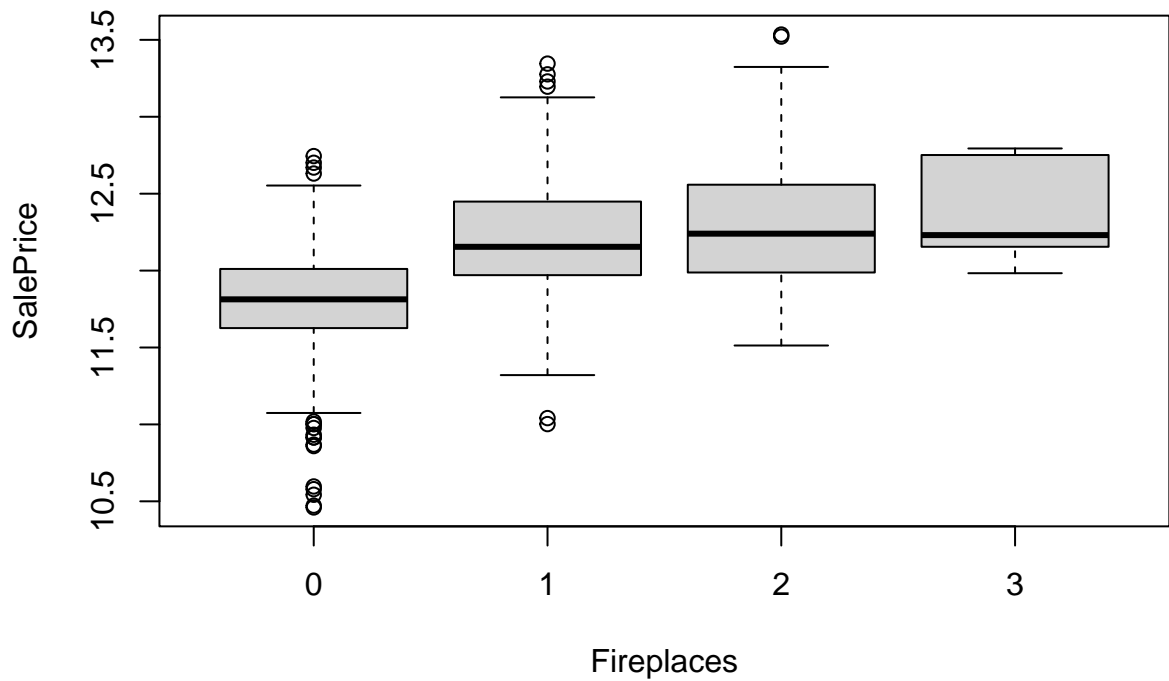


Df	Sum Sq	Mean Sq	F value	Pr(>F)
6	5.538054	0.9230091	5.901244	4.3e-06
1453	227.262605	0.1564092	NA	NA

Eta squared: 0.02378883

Fireplaces

Numero di NA: 0

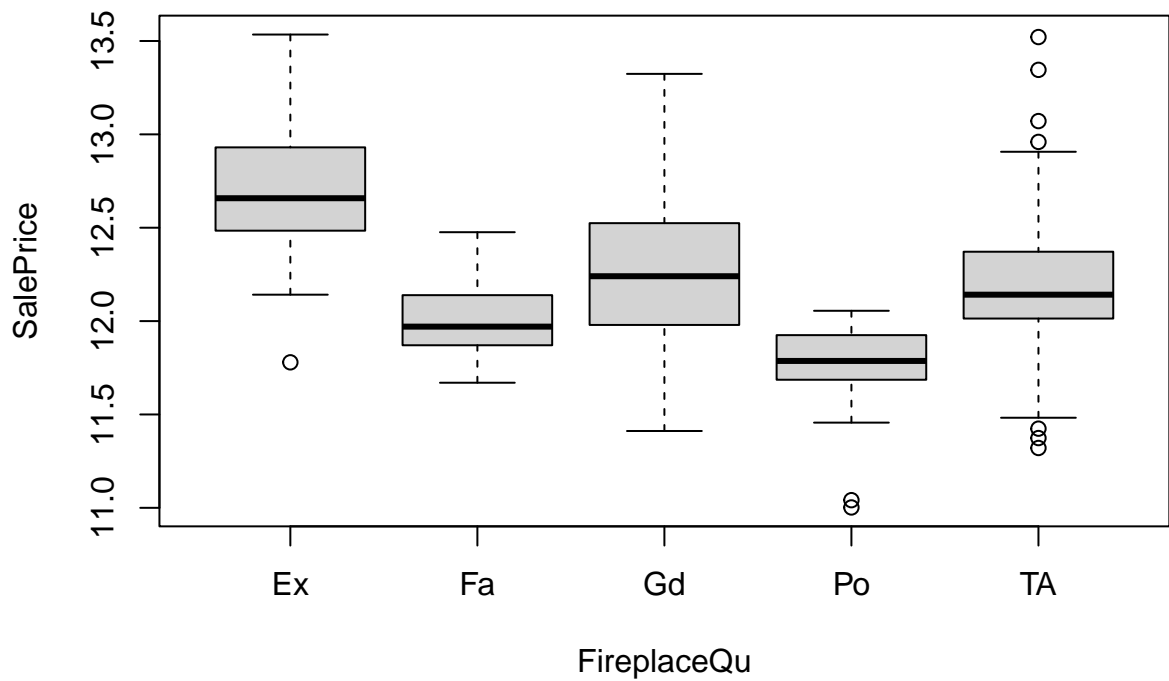


Df	Sum Sq	Mean Sq	F value	Pr(>F)
1	55.7699	55.7698986	459.3129	0
1458	177.0308	0.1214203	NA	NA

Eta squared: 0.2395607

FireplaceQu

Numero di NA: 690

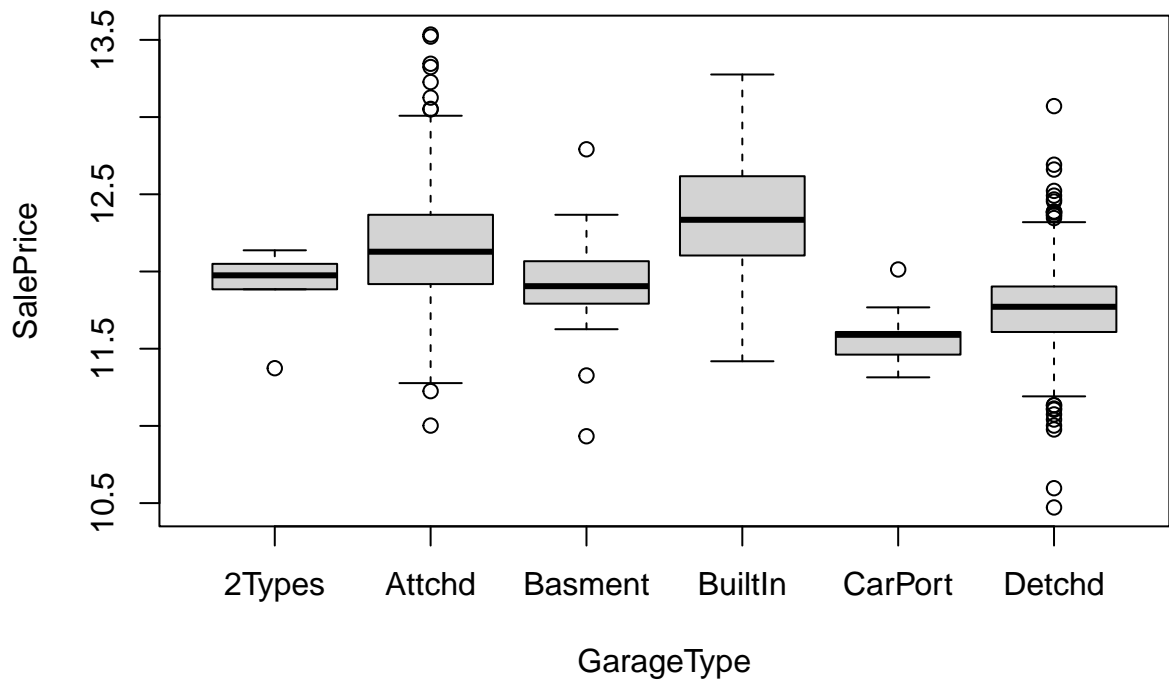


Df	Sum Sq	Mean Sq	F value	Pr(>F)
4	11.81972	2.9549312	25.5853	0
765	88.35240	0.1154933	NA	NA

Eta squared: 0.1179942

GarageType

Numero di NA: 81

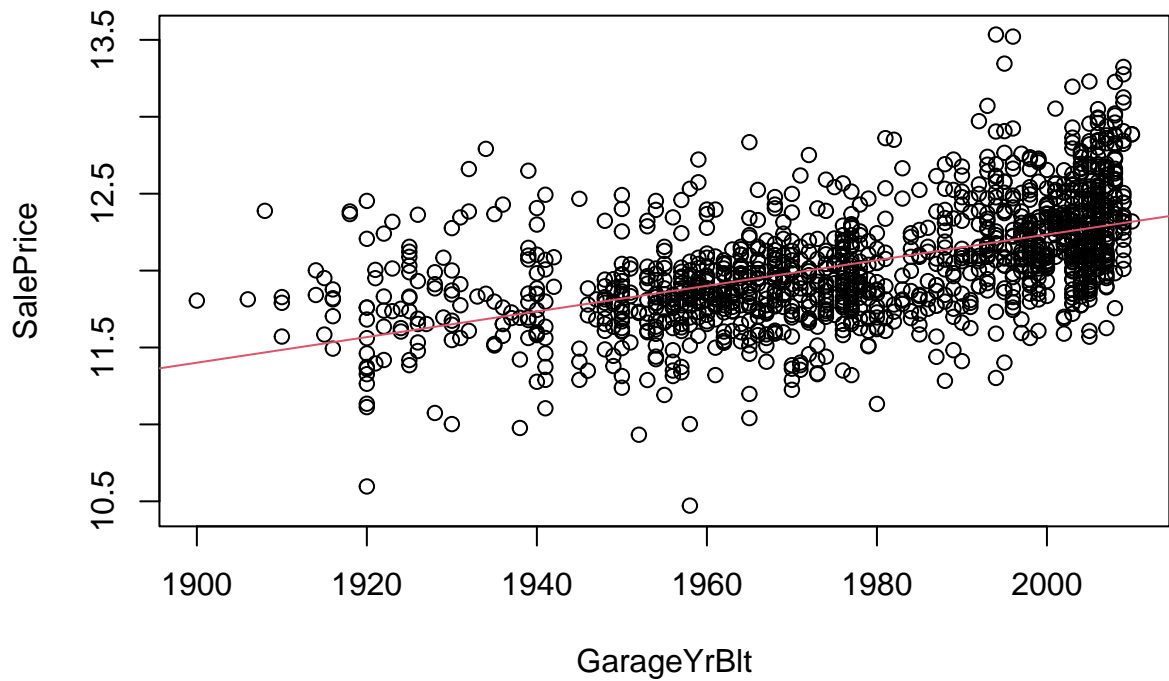


Df	Sum Sq	Mean Sq	F value	Pr(>F)
5	53.61664	10.7233282	101.1629	0
1373	145.53881	0.1060006	NA	NA

Eta squared: 0.2692201

GarageYrBl

Numero di NA: 81

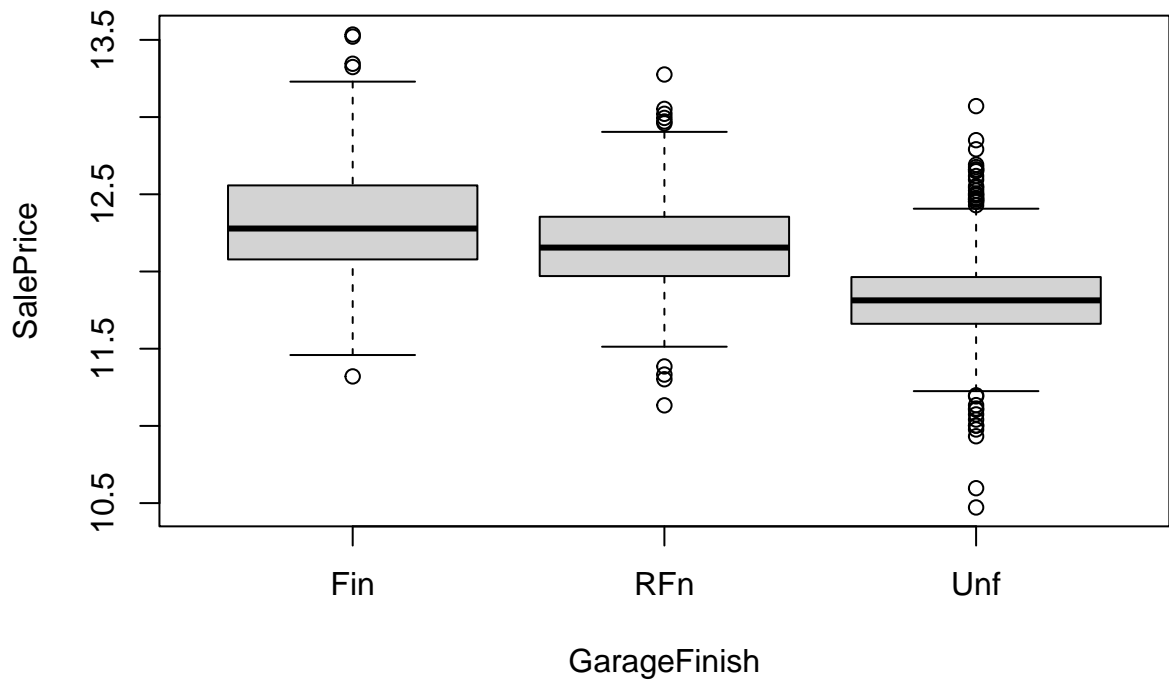


Estimate	Std. Error	t value	Pr(> t)
-4.4281509	0.690467	-6.413269	0
0.0083313	0.000349	23.874774	0

cov	cor	rsq
5.078594	0.5410729	0.2927598

GarageFinish

Numero di NA: 81

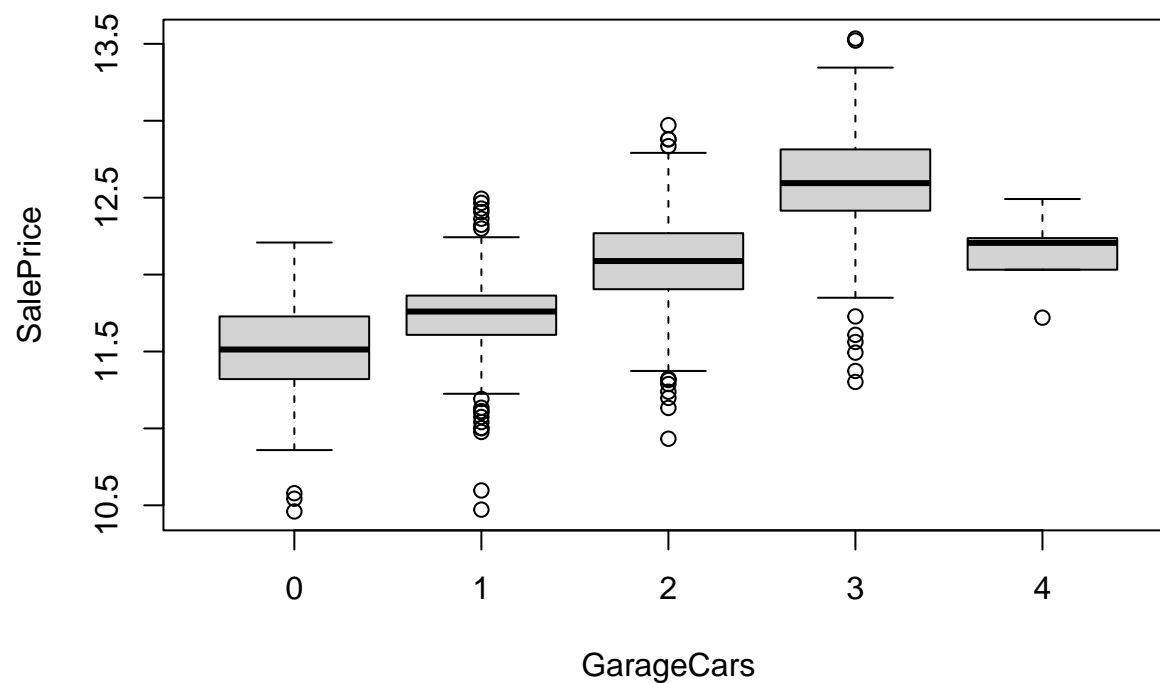


Df	Sum Sq	Mean Sq	F value	Pr(>F)
2	64.41727	32.2086371	328.9274	0
1376	134.73818	0.0979202	NA	NA

Eta squared: 0.3234522

GarageCars

Numero di NA: 0



Df	Sum Sq	Mean Sq	F value	Pr(>F)
1	107.8449	107.8449352	1258.349	0
1458	124.9557	0.0857035	NA	NA

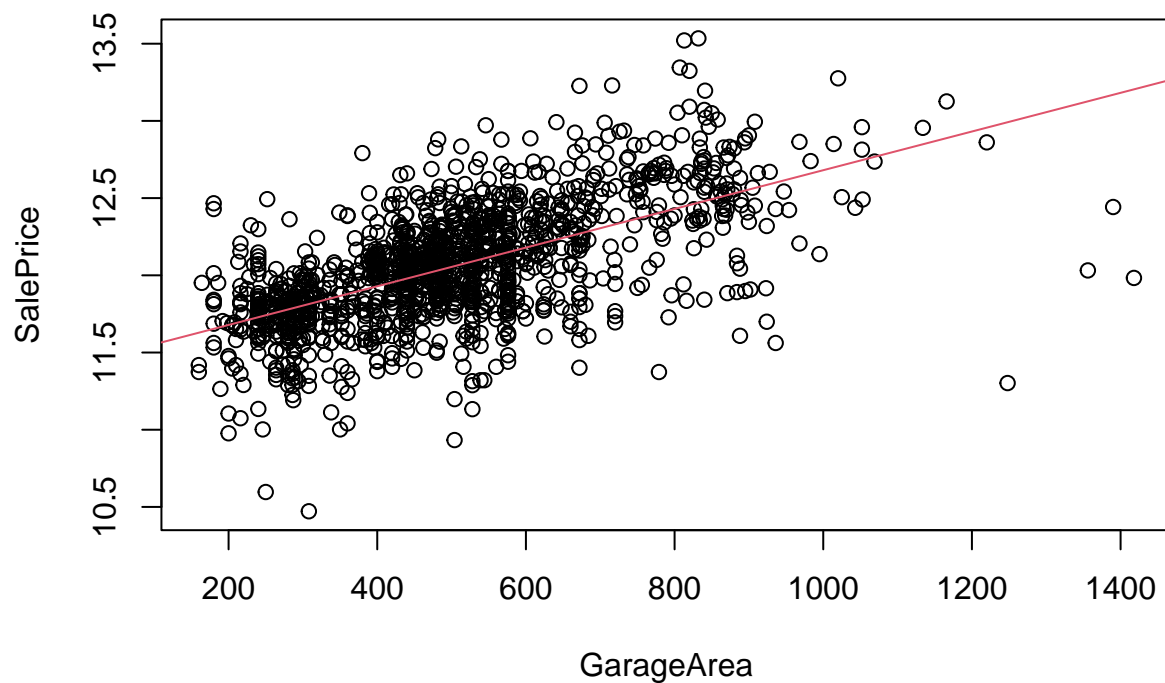
Eta squared: 0.4632501

Nel grafico si nota una relazione positiva tra il numero di auto che il garage può contenere e il prezzo di vendita. La covarianza è positiva e il coefficiente di correlazione alto (0.64). Quindi, il numero di posti auto nel garage è un fattore significativo nel determinare il prezzo di vendita.

GarageArea

Numero di NA: 0

Numero di zeri rimossi: 81



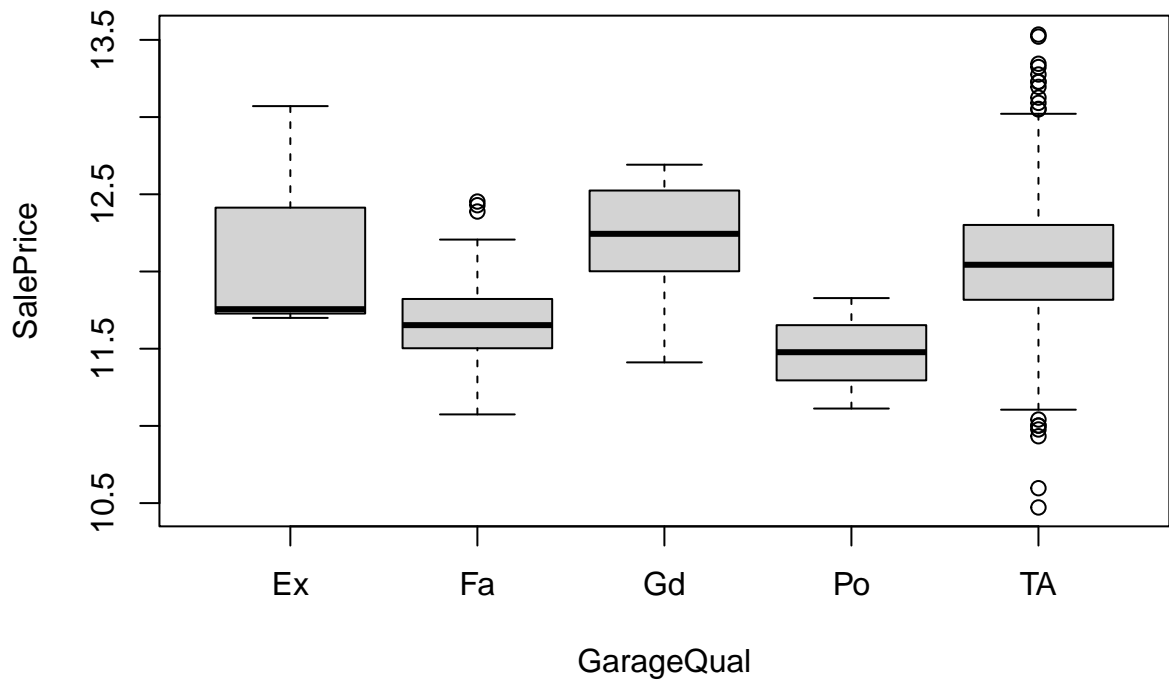
Estimate	Std. Error	t value	Pr(> t)
11.4279755	0.0233064	490.33554	0
0.0012528	0.0000436	28.70635	0

cov	cor	rsq
43.19171	0.6118748	0.3743907

Il grafico mostra una relazione positiva tra l'area del garage e il prezzo di vendita. La covarianza è positiva e il coefficiente di correlazione è moderatamente alto (0.62). Le case con garage più grandi tendono ad avere prezzi di vendita più alti

GarageQual

Numero di NA: 81

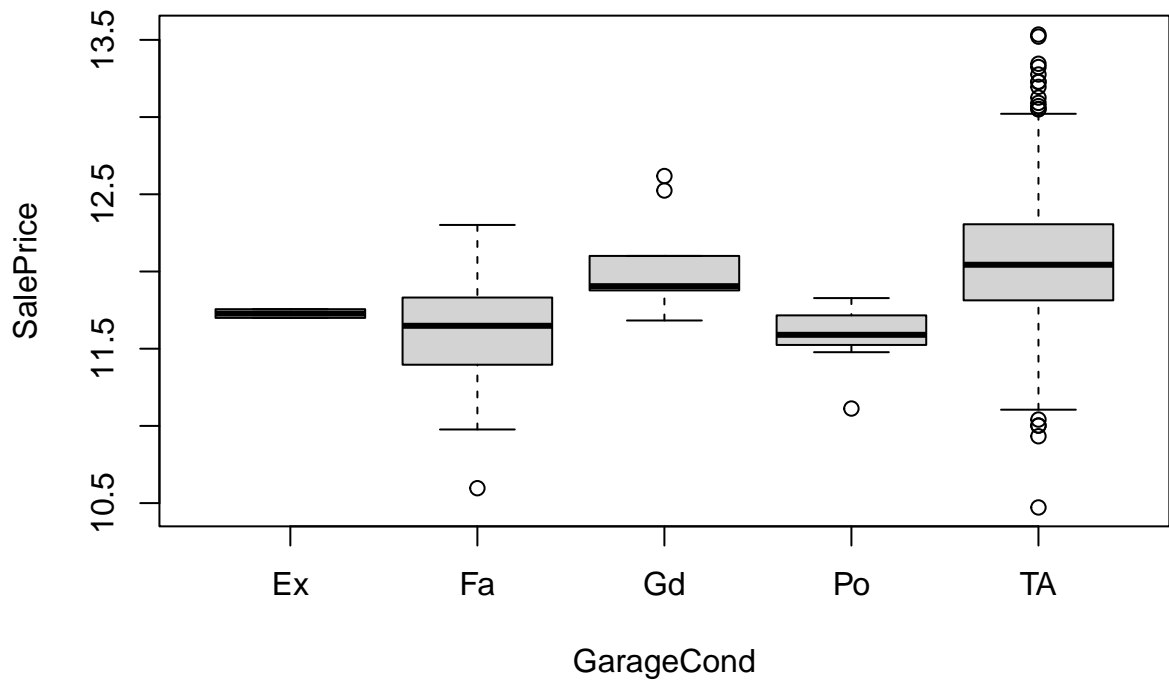


Df	Sum Sq	Mean Sq	F value	Pr(>F)
4	8.619977	2.1549942	15.54021	0
1374	190.535474	0.1386721	NA	NA

Eta squared: 0.04328266

GarageCond

Numero di NA: 81

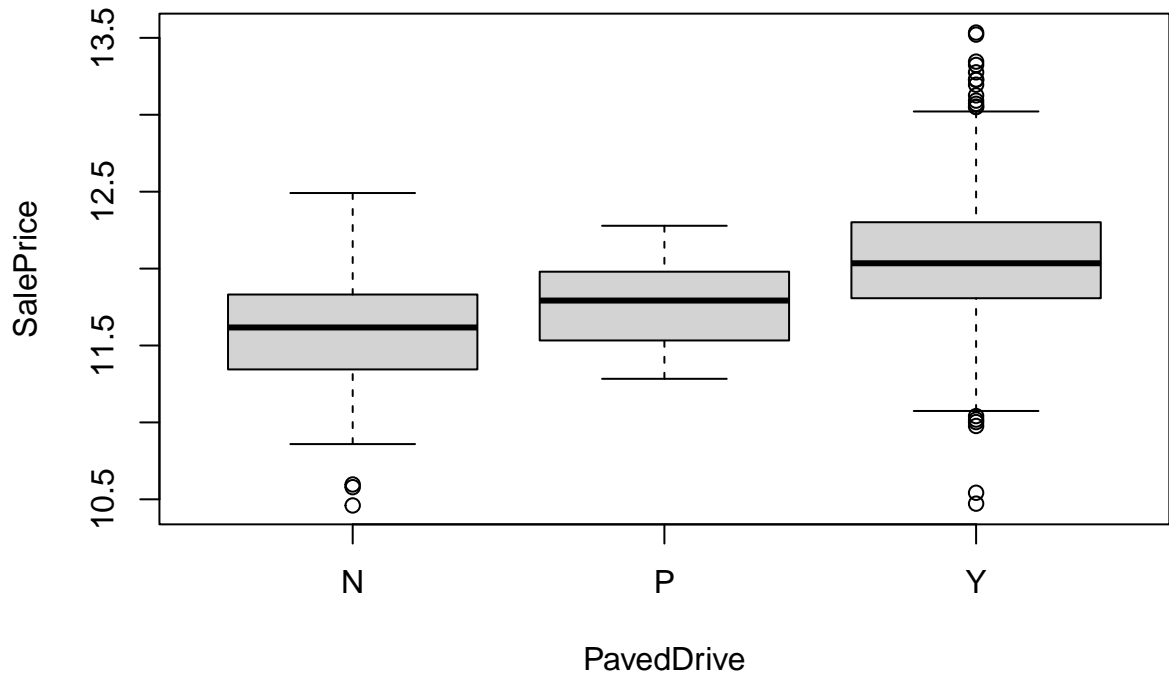


Df	Sum Sq	Mean Sq	F value	Pr(>F)
4	9.504909	2.3762274	17.21554	0
1374	189.650541	0.1380281	NA	NA

Eta squared: 0.04772608

PavedDrive

Numero di NA: 0



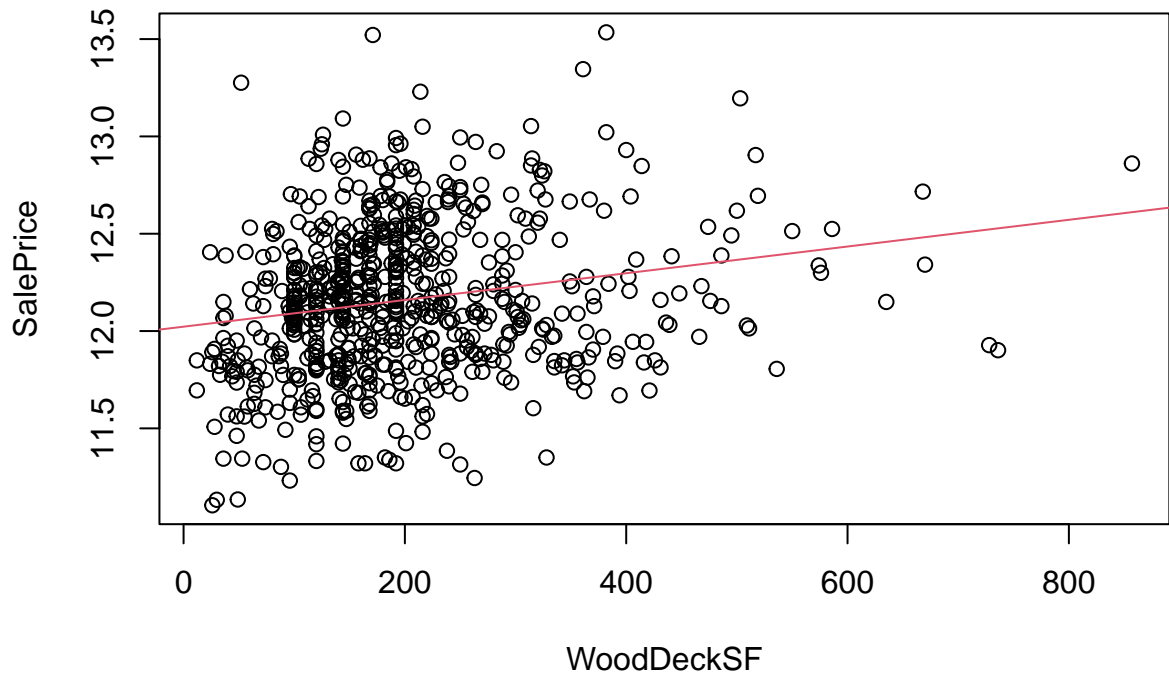
Df	Sum Sq	Mean Sq	F value	Pr(>F)
2	21.70324	10.851618	74.89815	0
1457	211.09742	0.144885	NA	NA

Eta squared: 0.09322669

WoodDeckSF

Numero di NA: 0

Numero di zeri rimossi: 761



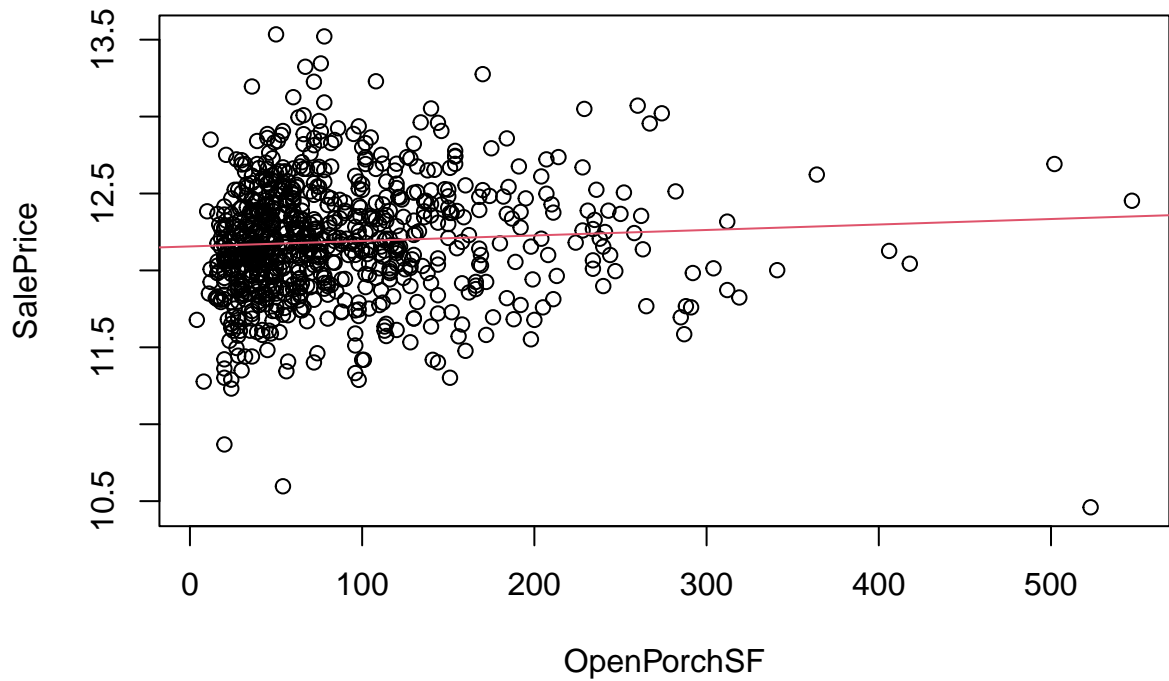
Estimate	Std. Error	t value	Pr(> t)
12.0224039	0.0287189	418.624010	0e+00
0.0006864	0.0001267	5.415747	1e-07

cov	cor	rsq
8.656546	0.2009515	0.0403815

OpenPorchSF

Numero di NA: 0

Numero di zeri rimossi: 656



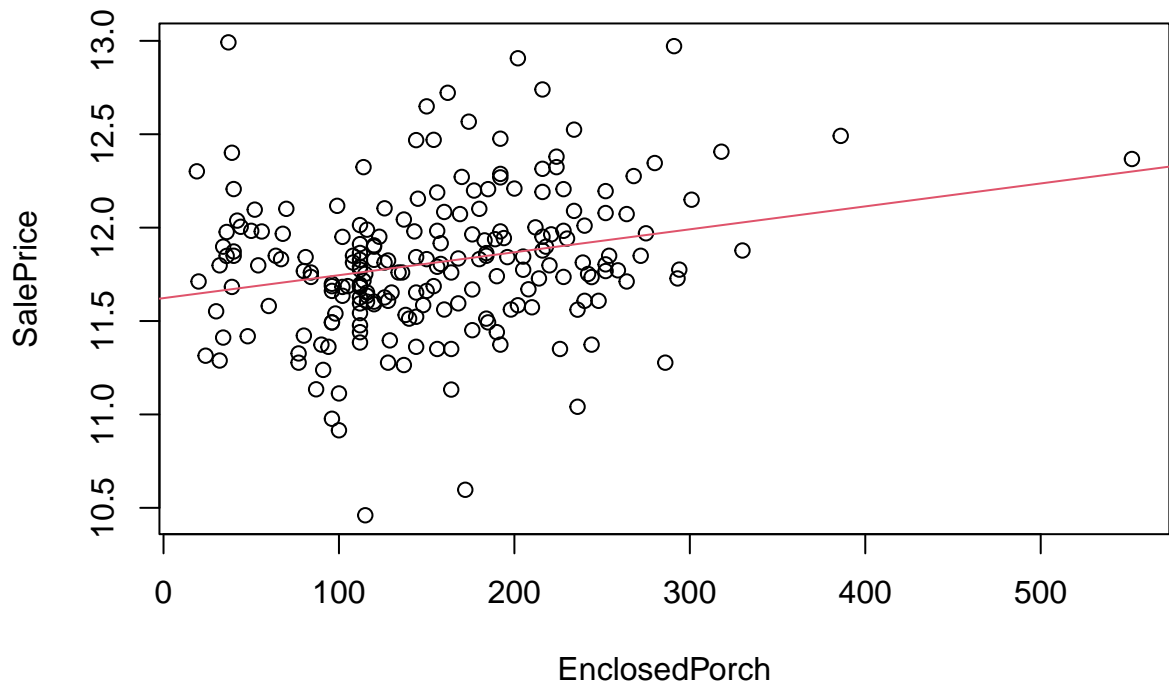
Estimate	Std. Error	t value	Pr(> t)
12.1558703	0.0209807	579.382108	0.0000000
0.0003566	0.0001922	1.855431	0.0639014

cov	cor	rsq
1.692327	0.0653774	0.0042742

EnclosedPorch

Numero di NA: 0

Numero di zeri rimossi: 1252



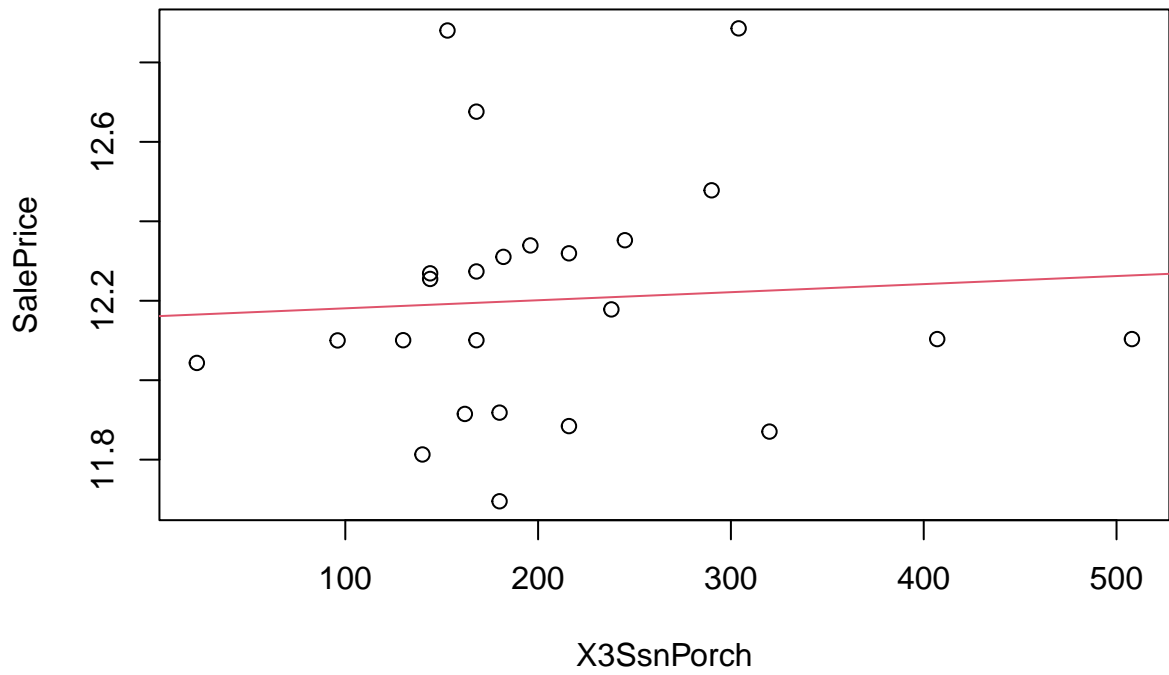
Estimate	Std. Error	t value	Pr(> t)
11.6226857	0.0575475	201.966812	0.0000000
0.0012278	0.0003346	3.669825	0.0003091

cov	cor	rsq
7.203512	0.2477195	0.061365

X3SsnPorch

Numero di NA: 0

Numero di zeri rimossi: 1436



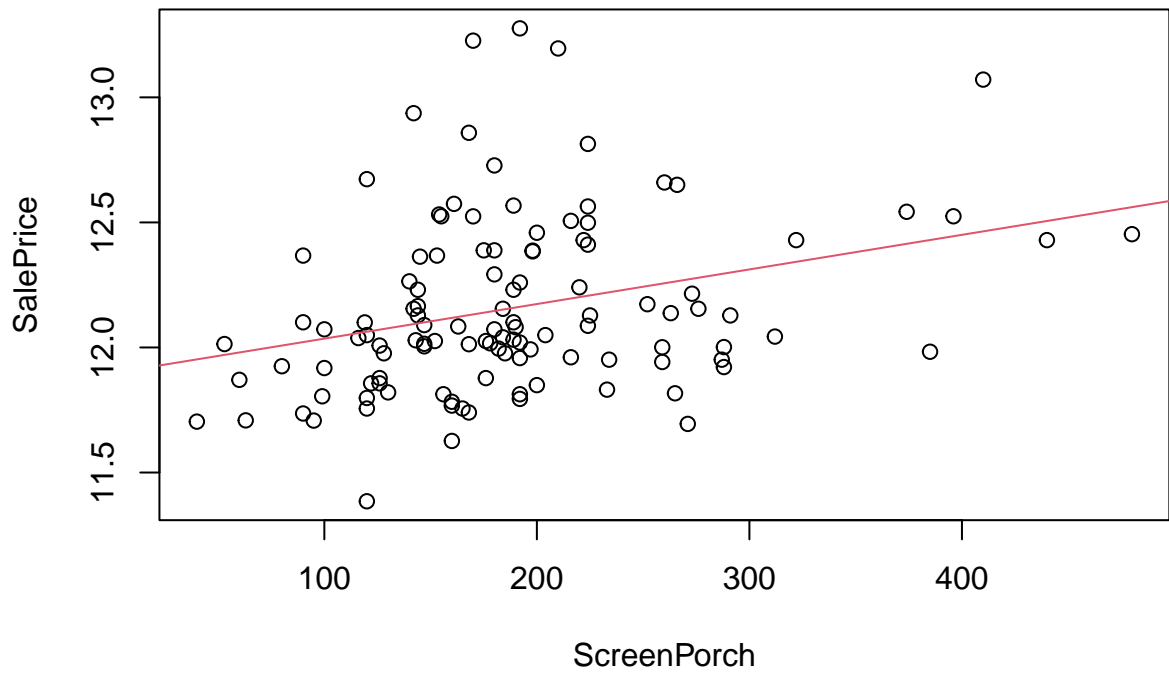
Estimate	Std. Error	t value	Pr(> t)
12.1606351	0.1475058	82.4417654	0.0000000
0.0002031	0.0006410	0.3168337	0.7543567

cov	cor	rsq
2.105639	0.0673956	0.0045422

ScreenPorch

Numero di NA: 0

Numero di zeri rimossi: 1344



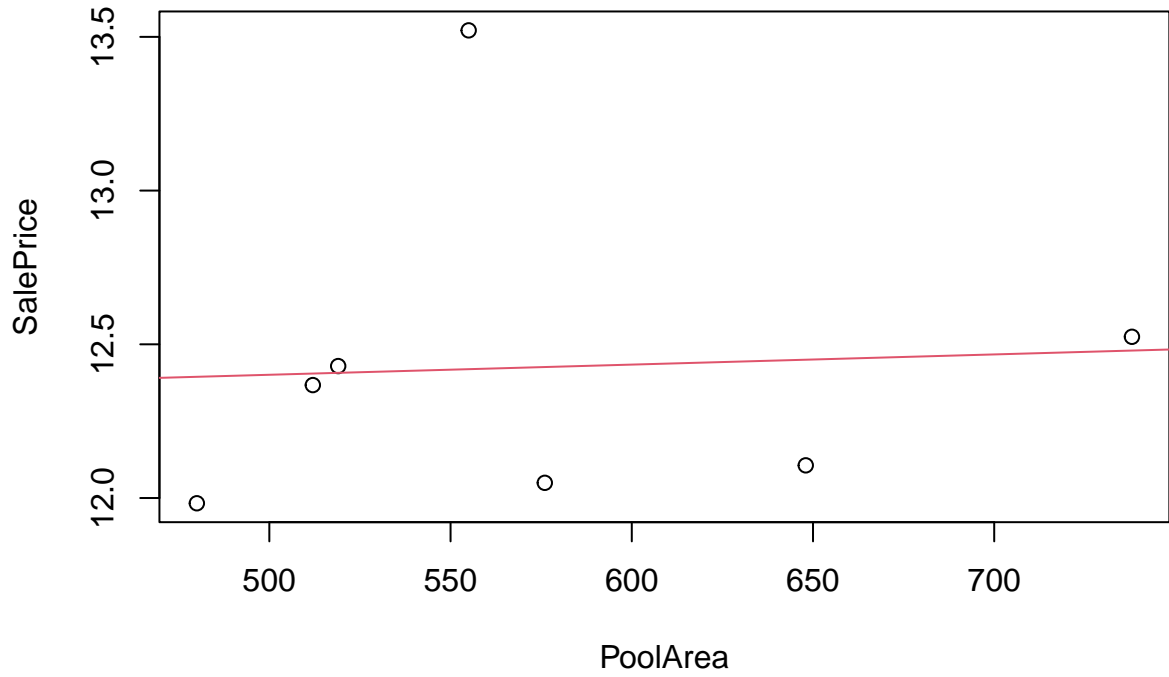
Estimate	Std. Error	t value	Pr(> t)
11.8969482	0.0825312	144.150872	0.0000000
0.0013835	0.0004029	3.433716	0.0008308

cov	cor	rsq
8.406896	0.3061545	0.0937306

PoolArea

Numero di NA: 0

Numero di zeri rimossi: 1453

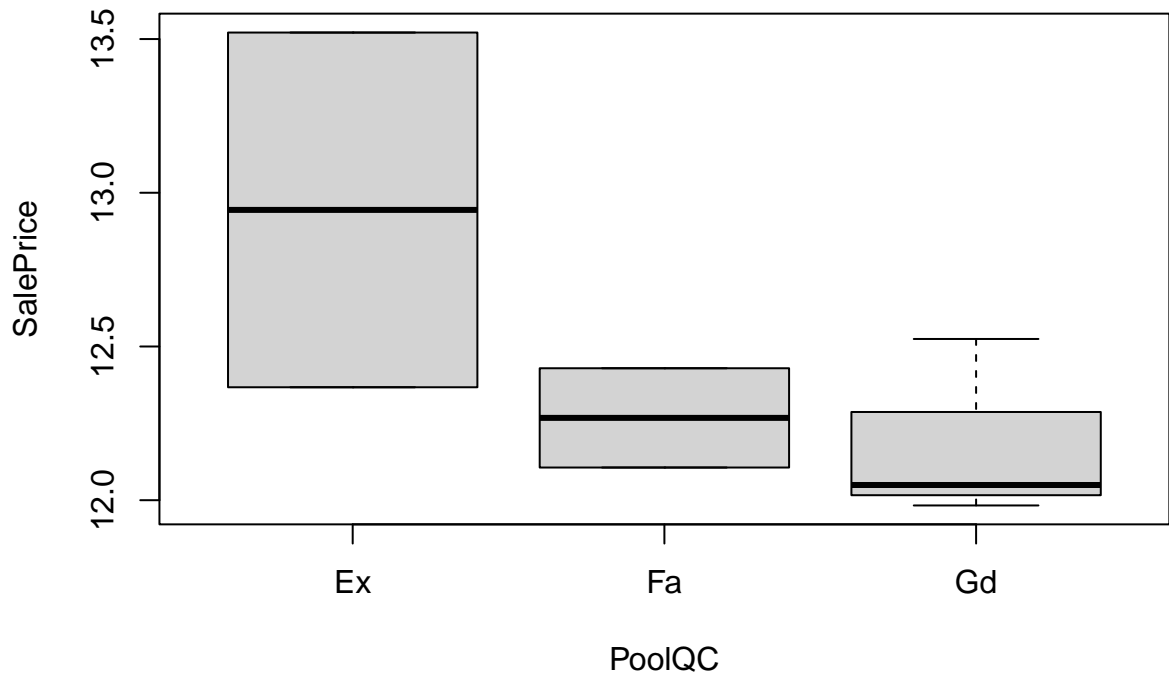


Estimate	Std. Error	t value	Pr(> t)
12.235359	1.5165817	8.0677216	0.0004738
0.000331	0.0026085	0.1268885	0.9039736

cov	cor	rsq
2.671462	0.0566551	0.0032098

PoolQC

Numero di NA: 1453

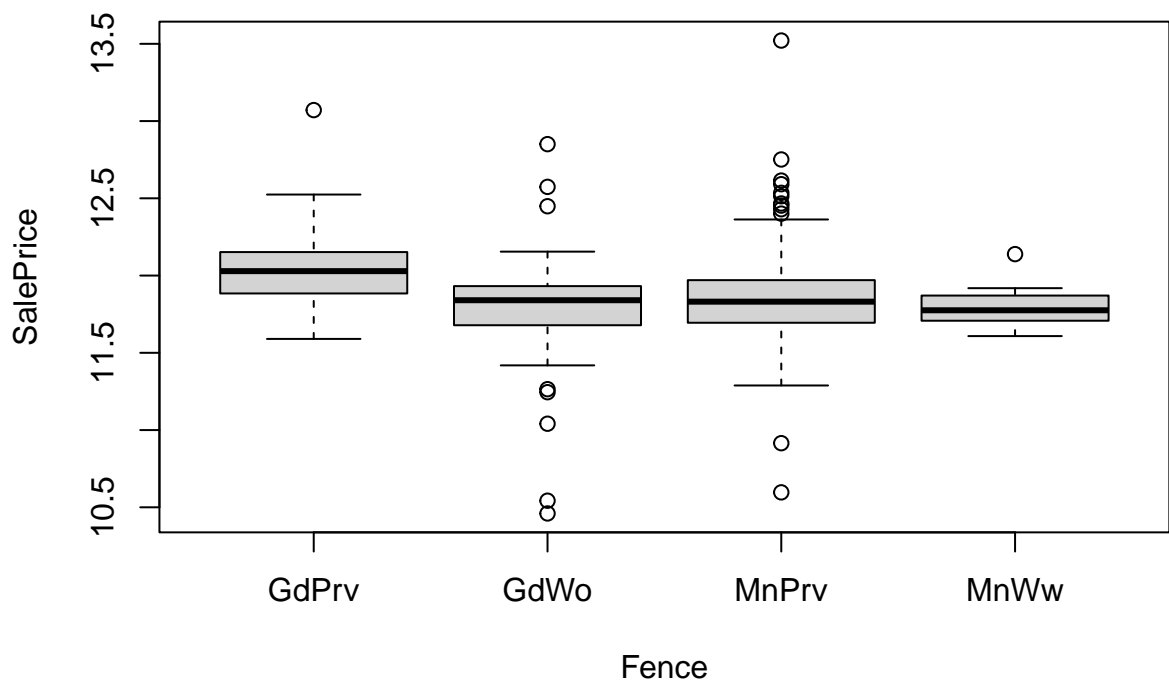


Df	Sum Sq	Mean Sq	F value	Pr(>F)
2	0.7606348	0.3803174	1.705083	0.2913829
4	0.8921965	0.2230491	NA	NA

Eta squared: 0.4602011

Fence

Numero di NA: 1179

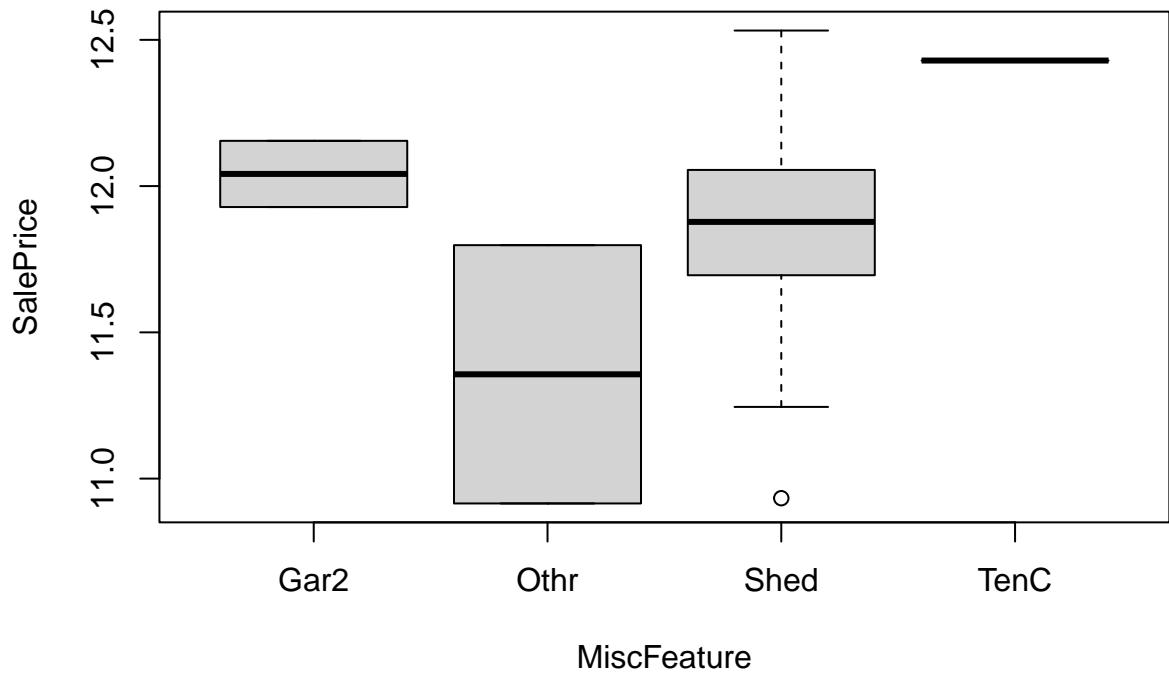


Df	Sum Sq	Mean Sq	F value	Pr(>F)
3	2.557665	0.8525549	8.260153	2.78e-05
277	28.589992	0.1032130	NA	NA

Eta squared: 0.08211419

MiscFeature

Numero di NA: 1406



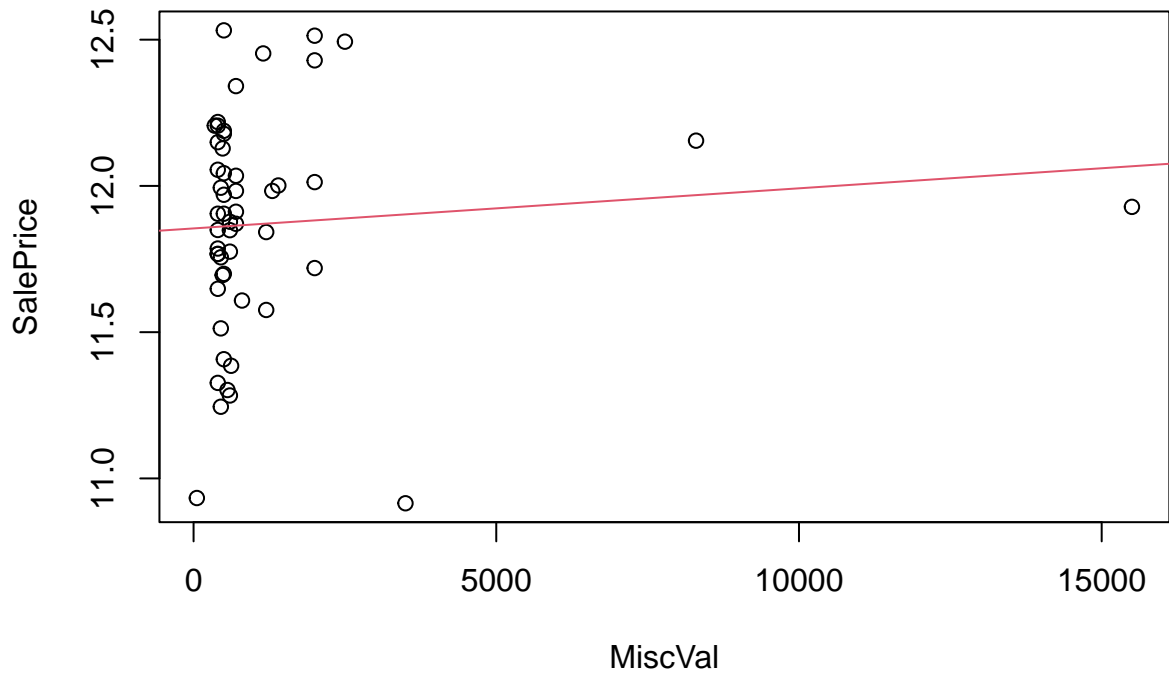
Df	Sum Sq	Mean Sq	F value	Pr(>F)
3	0.8981155	0.2993718	2.383971	0.0803062
50	6.2788467	0.1255769	NA	NA

Eta squared: 0.1251387

MiscVal

Numero di NA: 0

Numero di zeri rimossi: 1408

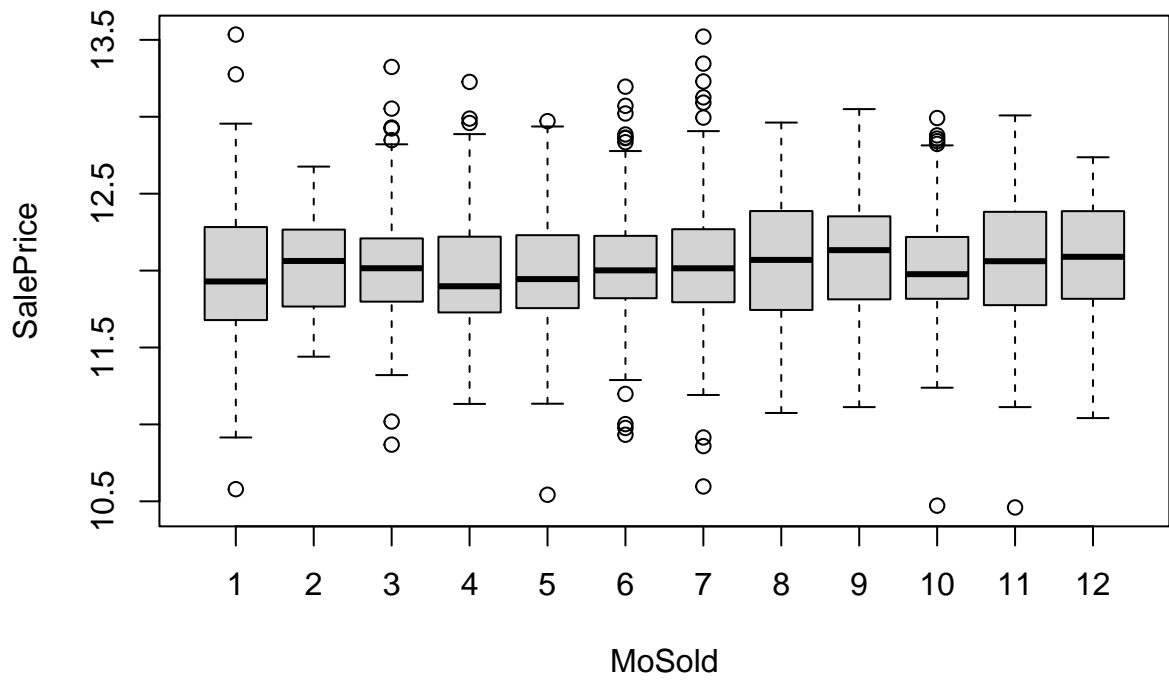


Estimate	Std. Error	t value	Pr(> t)
11.8547305	0.0588506	201.4376691	0.0000000
0.0000137	0.0000223	0.6146438	0.5415771

cov	cor	rsq
76.44975	0.0865972	0.0074991

MoSold

Numero di NA: 0

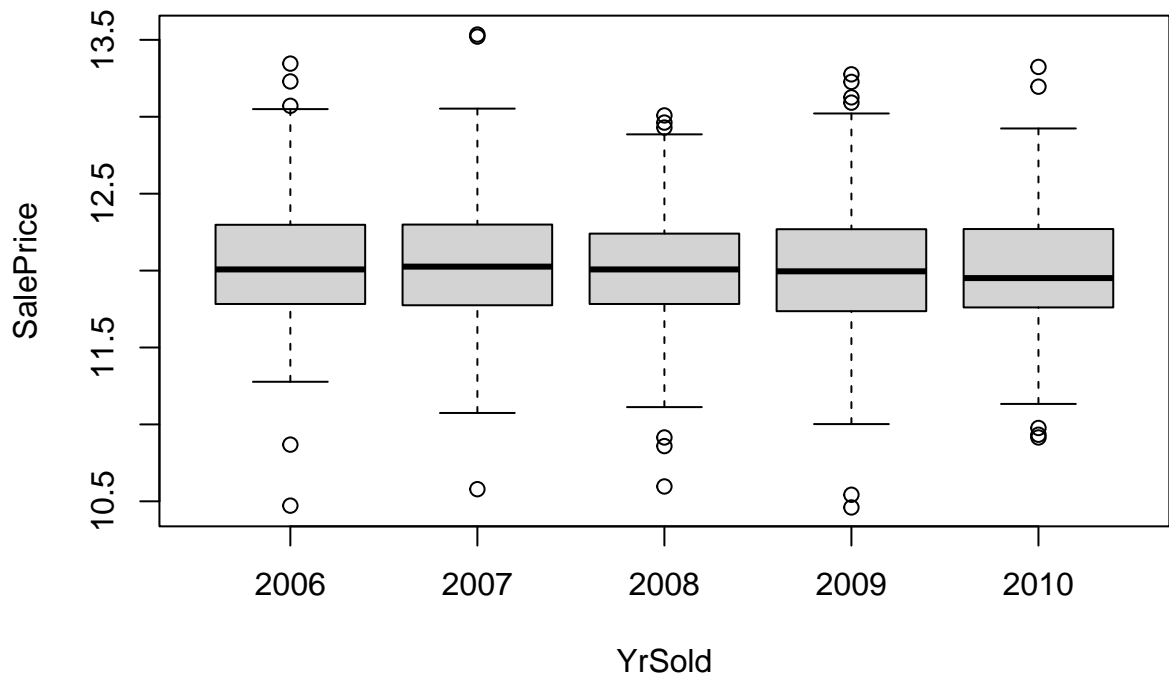


Df	Sum Sq	Mean Sq	F value	Pr(>F)
1	0.7651389	0.7651389	4.807766	0.0284893
1458	232.0355201	0.1591464	NA	NA

Eta squared: 0.00328667

YrSold

Numero di NA: 0

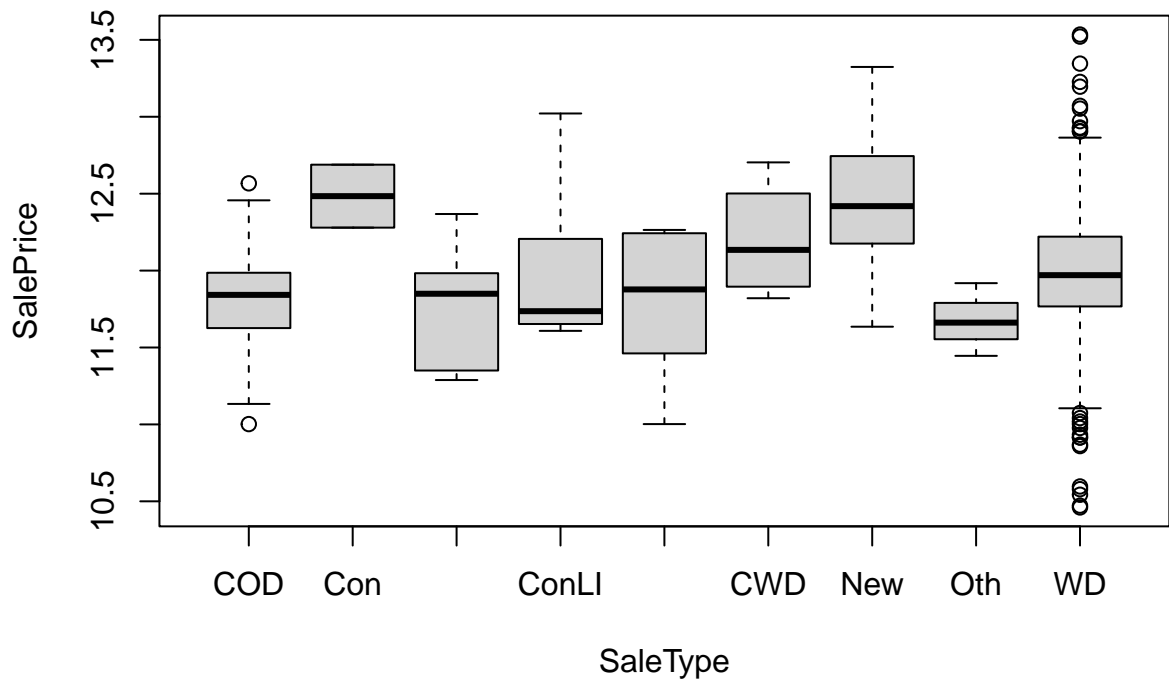


Df	Sum Sq	Mean Sq	F value	Pr(>F)
1	0.3232504	0.3232504	2.02729	0.1547098
1458	232.4774085	0.1594495	NA	NA

Eta squared: 0.001388529

SaleType

Numero di NA: 0

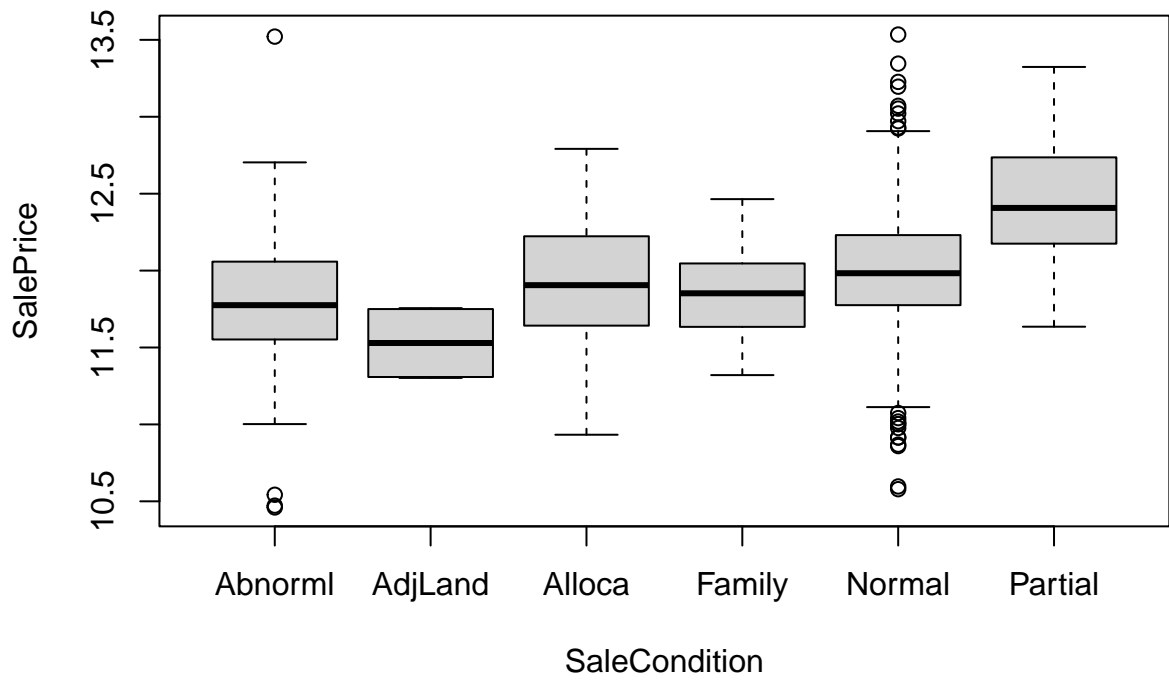


Df	Sum Sq	Mean Sq	F value	Pr(>F)
8	27.96295	3.495369	24.75999	0
1451	204.83771	0.141170	NA	NA

Eta squared: 0.1201154

SaleCondition

Numero di NA: 0



Df	Sum Sq	Mean Sq	F value	Pr(>F)
5	30.0359	6.0071796	43.07671	0
1454	202.7648	0.1394531	NA	NA

Eta squared: 0.1290198