# Ensemble Methods on the Ames Housing dataset

install_tensorflow() install.packages("tensorflow") library(tensorflow) install_tensorflow()

library(tensorflow) tf$constant("Hellow Tensorflow")

## Predicting house prices using Ensemble methods

#### Reading in the dataset

```
dataHouse <- read.csv("housing.csv", colClasses = c(MSSubClass = 'factor',  MoSold ='factor'))
dataHouse$Id <- NULL
```

### Section 1 - Data Cleaning

#### Overall structure and summary statistics of variables.

```
summary(dataHouse)
```

```
##    MSSubClass     MSZoning     LotFrontage       LotArea        Street
##  20     :536   C (all):  10   Min.   : 21.00   Min.   :  1300   Grvl:   6
##  60     :299   FV     :  65   1st Qu.: 59.00   1st Qu.:  7554   Pave:1454
##  50     :144   RH     :  16   Median : 69.00   Median :  9478
##  120    : 87   RL     :1151   Mean   : 70.05   Mean   : 10517
##  30     : 69   RM     : 218   3rd Qu.: 80.00   3rd Qu.: 11602
##  160    : 63                  Max.   :313.00   Max.   :215245
##  (Other):262                  NA's   :259
##   Alley      LotShape  LandContour  Utilities      LotConfig    LandSlope
##  Grvl:  50   IR1:484   Bnk:  63    AllPub:1459   Corner : 263   Gtl:1382
##  Pave:  41   IR2: 41   HLS:  50    NoSeWa:   1   CulDSac:  94   Mod:  65
##  NA's:1369   IR3: 10   Low:  36                  FR2    :  47   Sev:  13
##              Reg:925   Lvl:1311                  FR3    :   4
##                                                  Inside :1052
##
##
##   Neighborhood   Condition1     Condition2     BldgType      HouseStyle
##  NAmes  :225   Norm   :1260   Norm   :1445   1Fam  :1220   1Story :726
##  CollgCr:150   Feedr  :  81   Feedr  :   6   2fmCon:  31   2Story :445
##  OldTown:113   Artery :  48   Artery :   2   Duplex:  52   1.5Fin :154
##  Edwards:100   RRAn   :  26   PosN   :   2   Twnhs :  43   SLvl   : 65
##  Somerst: 86   PosN   :  19   RRNn   :   2   TwnhsE: 114   SFoyer : 37
##  Gilbert: 79   RRAe   :  11   PosA   :   1                 1.5Unf : 14
##  (Other):707   (Other):  15   (Other):   2                 (Other): 19
##   OverallQual      OverallCond      YearBuilt      YearRemodAdd    RoofStyle
##  Min.   : 1.000   Min.   :1.000   Min.   :1872   Min.   :1950   Flat   :  13
```

```
##  1st Qu.: 5.000   1st Qu.:5.000   1st Qu.:1954   1st Qu.:1967   Gable  :1141
##  Median : 6.000   Median :5.000   Median :1973   Median :1994   Gambrel:  11
##  Mean   : 6.099   Mean   :5.575   Mean   :1971   Mean   :1985   Hip    : 286
##  3rd Qu.: 7.000   3rd Qu.:6.000   3rd Qu.:2000   3rd Qu.:2004   Mansard:   7
##  Max.   :10.000   Max.   :9.000   Max.   :2010   Max.   :2010   Shed   :   2
##
##     RoofMatl     Exterior1st   Exterior2nd    MasVnrType    MasVnrArea
##  CompShg:1434   VinylSd:515   VinylSd:504   BrkCmn : 15   Min.   :   0.0
##  Tar&Grv:  11   HdBoard:222   MetalSd:214   BrkFace:445   1st Qu.:   0.0
##  WdShngl:   6   MetalSd:220   HdBoard:207   None   :864   Median :   0.0
##  WdShake:   5   Wd Sdng:206   Wd Sdng:197   Stone  :128   Mean   : 103.7
##  ClyTile:   1   Plywood:108   Plywood:142   NA's   :  8   3rd Qu.: 166.0
##  Membran:   1   CemntBd: 61   CmentBd: 60                 Max.   :1600.0
##  (Other):   2   (Other):128   (Other):136                 NA's   :8
##  ExterQual ExterCond  Foundation  BsmtQual   BsmtCond    BsmtExposure
##  Ex: 52    Ex:   3   BrkTil:146   Ex  :121   Fa  :  45   Av  :221
##  Fa: 14    Fa:  28   CBlock:634   Fa  : 35   Gd  :  65   Gd  :134
##  Gd:488    Gd: 146   PConc :647   Gd  :618   Po  :   2   Mn  :114
##  TA:906    Po:   1   Slab  : 24   TA  :649   TA  :1311   No  :953
##            TA:1282   Stone :  6   NA's: 37   NA's:  37   NA's: 38
##                      Wood  :  3
##
##  BsmtFinType1   BsmtFinSF1     BsmtFinType2   BsmtFinSF2        BsmtUnfSF
##  ALQ :220     Min.   :   0.0   ALQ :  19    Min.   :   0.00   Min.   :   0.0
##  BLQ :148     1st Qu.:   0.0   BLQ :  33    1st Qu.:   0.00   1st Qu.: 223.0
##  GLQ :418     Median : 383.5   GLQ :  14    Median :   0.00   Median : 477.5
##  LwQ : 74     Mean   : 443.6   LwQ :  46    Mean   :  46.55   Mean   : 567.2
##  Rec :133     3rd Qu.: 712.2   Rec :  54    3rd Qu.:   0.00   3rd Qu.: 808.0
##  Unf :430     Max.   :5644.0   Unf :1256    Max.   :1474.00   Max.   :2336.0
##  NA's: 37                      NA's:  38
##   TotalBsmtSF      Heating     HeatingQC CentralAir Electrical     X1stFlrSF
##  Min.   :   0.0   Floor:   1   Ex:741    N:  95     FuseA:  94   Min.   : 334
##  1st Qu.: 795.8   GasA :1428   Fa: 49    Y:1365     FuseF:  27   1st Qu.: 882
##  Median : 991.5   GasW :  18   Gd:241               FuseP:   3   Median :1087
##  Mean   :1057.4   Grav :   7   Po:  1               Mix  :   1   Mean   :1163
##  3rd Qu.:1298.2   OthW :   2   TA:428               SBrkr:1334   3rd Qu.:1391
##  Max.   :6110.0   Wall :   4                        NA's :   1   Max.   :4692
##
##    X2ndFlrSF     LowQualFinSF       GrLivArea     BsmtFullBath
##  Min.   :   0   Min.   :  0.000   Min.   : 334   Min.   :0.0000
##  1st Qu.:   0   1st Qu.:  0.000   1st Qu.:1130   1st Qu.:0.0000
##  Median :   0   Median :  0.000   Median :1464   Median :0.0000
##  Mean   : 347   Mean   :  5.845   Mean   :1515   Mean   :0.4253
##  3rd Qu.: 728   3rd Qu.:  0.000   3rd Qu.:1777   3rd Qu.:1.0000
##  Max.   :2065   Max.   :572.000   Max.   :5642   Max.   :3.0000
##
##   BsmtHalfBath        FullBath        HalfBath       BedroomAbvGr
##  Min.   :0.00000   Min.   :0.000   Min.   :0.0000   Min.   :0.000
##  1st Qu.:0.00000   1st Qu.:1.000   1st Qu.:0.0000   1st Qu.:2.000
##  Median :0.00000   Median :2.000   Median :0.0000   Median :3.000
##  Mean   :0.05753   Mean   :1.565   Mean   :0.3829   Mean   :2.866
##  3rd Qu.:0.00000   3rd Qu.:2.000   3rd Qu.:1.0000   3rd Qu.:3.000
##  Max.   :2.00000   Max.   :3.000   Max.   :2.0000   Max.   :8.000
##
```

```
##   KitchenAbvGr   KitchenQual  TotRmsAbvGrd    Functional    Fireplaces
##  Min.   :0.000   Ex:100      Min.   : 2.000   Maj1:  14   Min.   :0.000
##  1st Qu.:1.000   Fa: 39      1st Qu.: 5.000   Maj2:   5   1st Qu.:0.000
##  Median :1.000   Gd:586      Median : 6.000   Min1:  31   Median :1.000
##  Mean   :1.047   TA:735      Mean   : 6.518   Min2:  34   Mean   :0.613
##  3rd Qu.:1.000               3rd Qu.: 7.000   Mod :  15   3rd Qu.:1.000
##  Max.   :3.000               Max.   :14.000   Sev :   1   Max.   :3.000
##                                               Typ :1360
##  FireplaceQu   GarageType   GarageYrBlt   GarageFinish   GarageCars
##  Ex  : 24    2Types :  6   Min.   :1900   Fin :352     Min.   :0.000
##  Fa  : 33    Attchd :870   1st Qu.:1961   RFn :422     1st Qu.:1.000
##  Gd  :380    Basment: 19   Median :1980   Unf :605     Median :2.000
##  Po  : 20    BuiltIn: 88   Mean   :1979   NA's: 81     Mean   :1.767
##  TA  :313    CarPort:  9   3rd Qu.:2002                3rd Qu.:2.000
##  NA's:690    Detchd :387   Max.   :2010                Max.   :4.000
##              NA's   : 81   NA's   :81
##    GarageArea     GarageQual  GarageCond  PavedDrive   WoodDeckSF
##  Min.   :   0.0   Ex  :   3   Ex  :   2   N:  90     Min.   :  0.00
##  1st Qu.: 334.5   Fa  :  48   Fa  :  35   P:  30     1st Qu.:  0.00
##  Median : 480.0   Gd  :  14   Gd  :   9   Y:1340     Median :  0.00
##  Mean   : 473.0   Po  :   3   Po  :   7              Mean   : 94.24
##  3rd Qu.: 576.0   TA  :1311   TA  :1326              3rd Qu.:168.00
##  Max.   :1418.0   NA's:  81   NA's:  81              Max.   :857.00
##
##   OpenPorchSF     EnclosedPorch      X3SsnPorch      ScreenPorch
##  Min.   :  0.00   Min.   :  0.00   Min.   :  0.00   Min.   :  0.00
##  1st Qu.:  0.00   1st Qu.:  0.00   1st Qu.:  0.00   1st Qu.:  0.00
##  Median : 25.00   Median :  0.00   Median :  0.00   Median :  0.00
##  Mean   : 46.66   Mean   : 21.95   Mean   :  3.41   Mean   : 15.06
##  3rd Qu.: 68.00   3rd Qu.:  0.00   3rd Qu.:  0.00   3rd Qu.:  0.00
##  Max.   :547.00   Max.   :552.00   Max.   :508.00   Max.   :480.00
##
##     PoolArea        PoolQC       Fence      MiscFeature    MiscVal
##  Min.   :  0.000   Ex  :   2   GdPrv:  59   Gar2:   2   Min.   :    0.00
##  1st Qu.:  0.000   Fa  :   2   GdWo :  54   Othr:   2   1st Qu.:    0.00
##  Median :  0.000   Gd  :   3   MnPrv: 157   Shed:  49   Median :    0.00
##  Mean   :  2.759   NA's:1453   MnWw :  11   TenC:   1   Mean   :   43.49
##  3rd Qu.:  0.000               NA's :1179   NA's:1406   3rd Qu.:    0.00
##  Max.   :738.000                                        Max.   :15500.00
##
##      MoSold        YrSold        SaleType     SaleCondition    SalePrice
##  6      :253   Min.   :2006   WD     :1267   Abnorml: 101   Min.   : 34900
##  7      :234   1st Qu.:2007   New    : 122   AdjLand:   4   1st Qu.:129975
##  5      :204   Median :2008   COD    :  43   Alloca :  12   Median :163000
##  4      :141   Mean   :2008   ConLD  :   9   Family :  20   Mean   :180921
##  8      :122   3rd Qu.:2009   ConLI  :   5   Normal :1198   3rd Qu.:214000
##  3      :106   Max.   :2010   ConLw  :   5   Partial: 125   Max.   :755000
##  (Other):400                  (Other):   9
```

There are 45 categorical features, and 35 numerical features. Of which, 19 features have missing values.

```r
colMeans(is.na(dataHouse))
```

```
##    MSSubClass      MSZoning   LotFrontage       LotArea        Street
```

```
## 0.0000000000 0.0000000000 0.1773972603 0.0000000000 0.0000000000
##         Alley      LotShape   LandContour     Utilities     LotConfig
## 0.9376712329 0.0000000000 0.0000000000 0.0000000000 0.0000000000
##     LandSlope  Neighborhood    Condition1    Condition2      BldgType
## 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
##    HouseStyle   OverallQual   OverallCond     YearBuilt  YearRemodAdd
## 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
##     RoofStyle      RoofMatl   Exterior1st   Exterior2nd    MasVnrType
## 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0054794521
##    MasVnrArea     ExterQual     ExterCond    Foundation      BsmtQual
## 0.0054794521 0.0000000000 0.0000000000 0.0000000000 0.0253424658
##      BsmtCond  BsmtExposure  BsmtFinType1    BsmtFinSF1  BsmtFinType2
## 0.0253424658 0.0260273973 0.0253424658 0.0000000000 0.0260273973
##    BsmtFinSF2     BsmtUnfSF   TotalBsmtSF       Heating     HeatingQC
## 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
##    CentralAir    Electrical     X1stFlrSF     X2ndFlrSF  LowQualFinSF
## 0.0000000000 0.0006849315 0.0000000000 0.0000000000 0.0000000000
##     GrLivArea  BsmtFullBath  BsmtHalfBath      FullBath      HalfBath
## 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
##  BedroomAbvGr  KitchenAbvGr   KitchenQual  TotRmsAbvGrd    Functional
## 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
##    Fireplaces   FireplaceQu    GarageType   GarageYrBlt  GarageFinish
## 0.0000000000 0.4726027397 0.0554794521 0.0554794521 0.0554794521
##    GarageCars    GarageArea    GarageQual    GarageCond    PavedDrive
## 0.0000000000 0.0000000000 0.0554794521 0.0554794521 0.0000000000
##    WoodDeckSF   OpenPorchSF EnclosedPorch    X3SsnPorch   ScreenPorch
## 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
##      PoolArea        PoolQC         Fence   MiscFeature       MiscVal
## 0.0000000000 0.9952054795 0.8075342466 0.9630136986 0.0000000000
##        MoSold        YrSold      SaleType SaleCondition     SalePrice
## 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
```

Variables with Missing values %: LotFrontage(17.7%), Alley(93.8%), MasVnrType (0.6%), MasVnrArea (0.6%), BsmtQual (2.5%), BsmtCond (2.5%), BsmtExposure (2.6%), BsmtFinType1 (2.5%), BsmtFinType2 (2.6), Electrical (0.06%), FireplaceQu (47.3%), GarageType (5.5%), GarageYrBlt (5.5%), GarageFinish (5.5%), GarageQual (5.5%), GarageCond (5.6%), PoolQC (99.5%), Fence (80.8%), MiscFeature (96.3%).

```
which(colMeans(is.na(dataHouse)) >= 0.30)
```

```
##       Alley FireplaceQu      PoolQC       Fence MiscFeature
##           6          57          72          73          74
```

```
dataHouse$Alley <- NULL
dataHouse$FireplaceQu <- NULL
dataHouse$PoolQC <- NULL
dataHouse$Fence <- NULL
dataHouse$MiscFeature <- NULL
```

```
NROW(dataHouse[!complete.cases(dataHouse),])/NROW(dataHouse) * 100
```

```
## [1] 25.06849
```

Around 25% rows have one or more missing values now.

*Imputing missing values using Random Forests*

```r
#install.packages("missForest")
library(missForest)
```

```
## Warning: package 'missForest' was built under R version 3.6.3
```

```
## Loading required package: randomForest
```

```
## Warning: package 'randomForest' was built under R version 3.6.3
```

```
## randomForest 4.6-14
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
## Loading required package: foreach
```

```
## Warning: package 'foreach' was built under R version 3.6.3
```

```
## Loading required package: itertools
```

```
## Warning: package 'itertools' was built under R version 3.6.3
```

```
## Loading required package: iterators
```

```
## Warning: package 'iterators' was built under R version 3.6.3
```

```r
set.seed(1)

tempVar <- dataHouse$SalePrice
dataHouse$SalePrice <- NULL

data_Imp <- missForest(dataHouse)
```

```
##   missForest iteration 1 in progress...done!
##   missForest iteration 2 in progress...done!
##   missForest iteration 3 in progress...done!
```

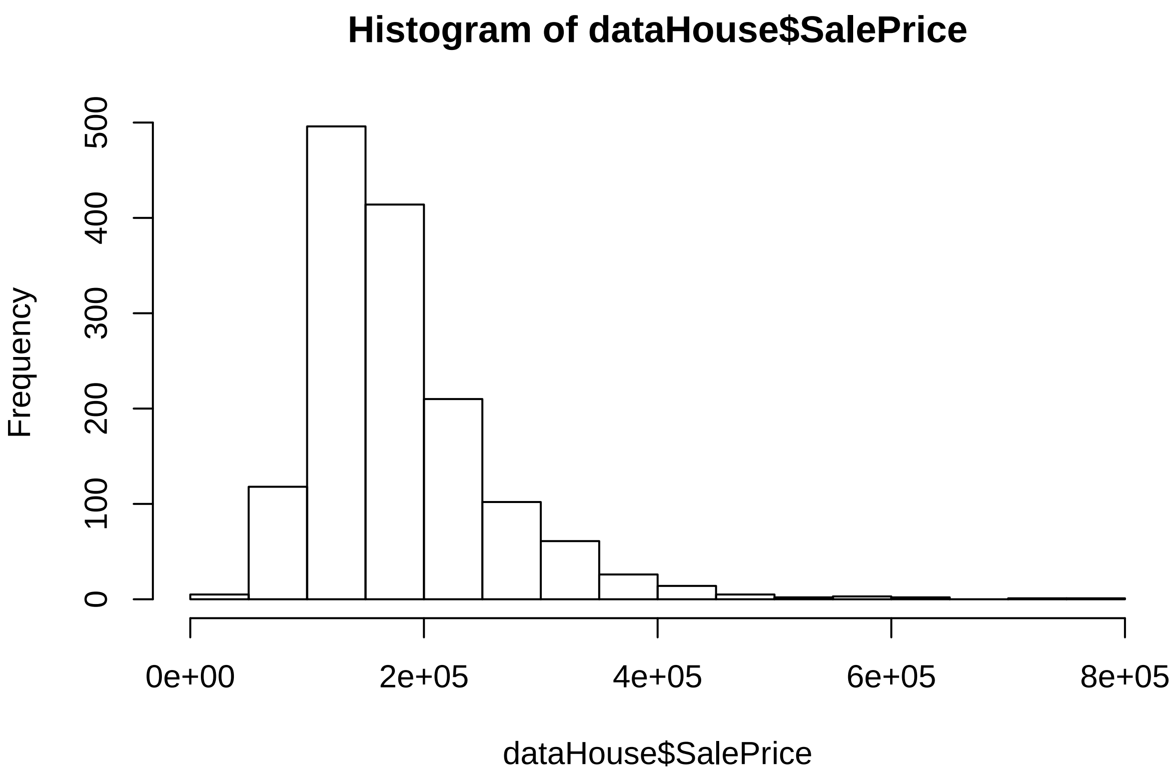
```r
data_Imp$OOBerror
```

```
##       NRMSE         PFC
## 0.008078283 0.039729589
```

```r
dataHouse <- data_Imp$ximp
dataHouse$SalePrice <- tempVar
```

## Section 2 - Data Exploration

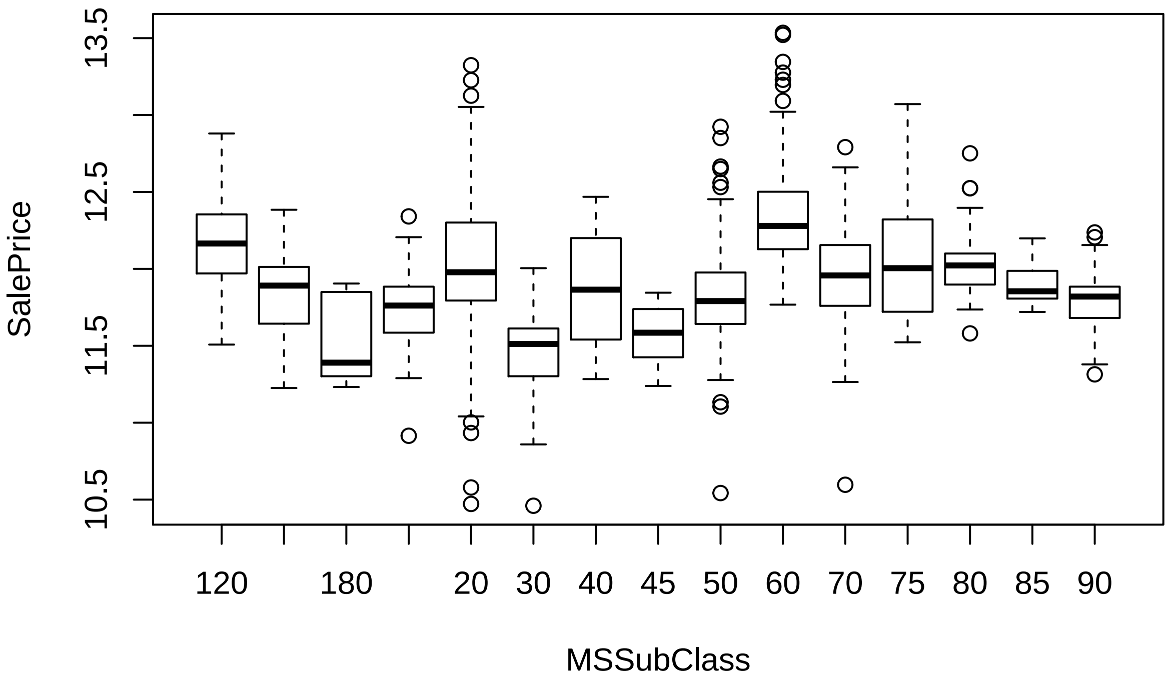
```r
hist(dataHouse$SalePrice)
```

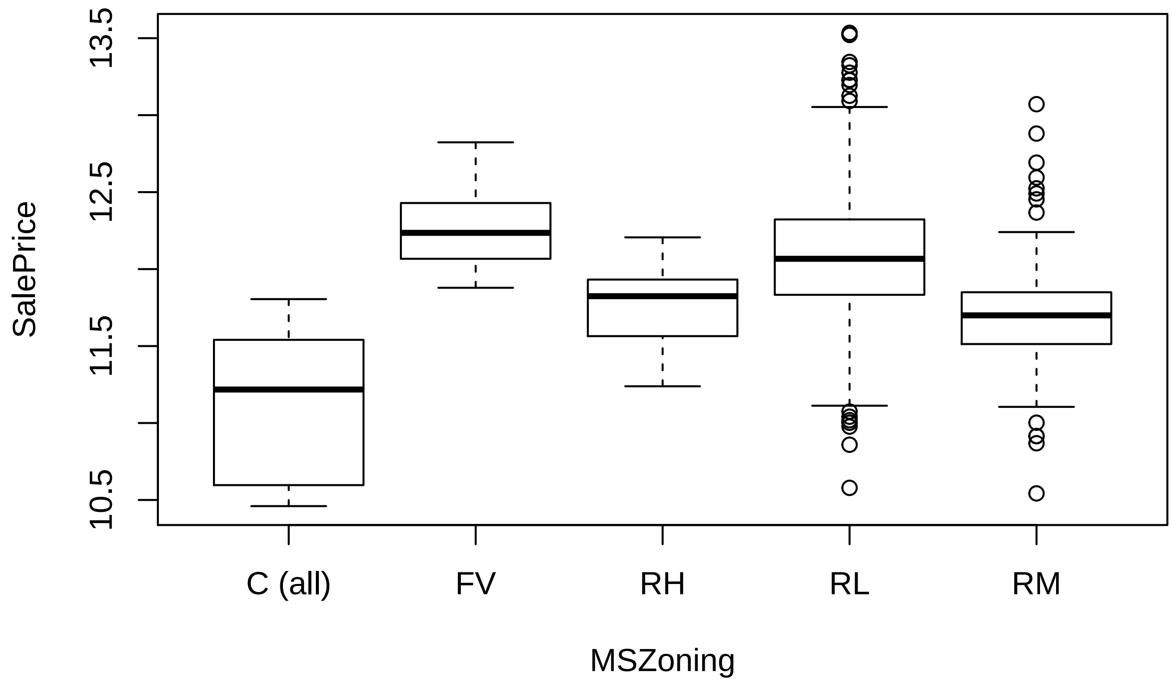



Eyeballing the distribution, we can see that the distribution is postively skewed, with most of the houses being in the lower price range.

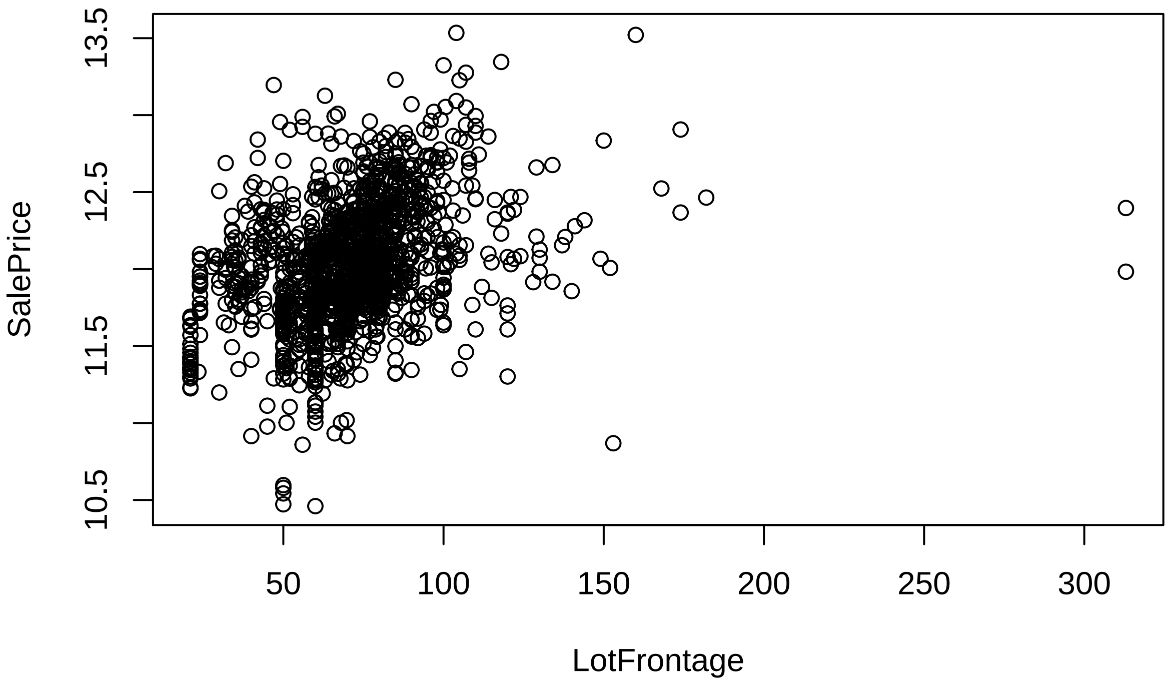
```r
# Log transformation of target variable

dataHouse$SalePrice <- log(dataHouse$SalePrice)
```

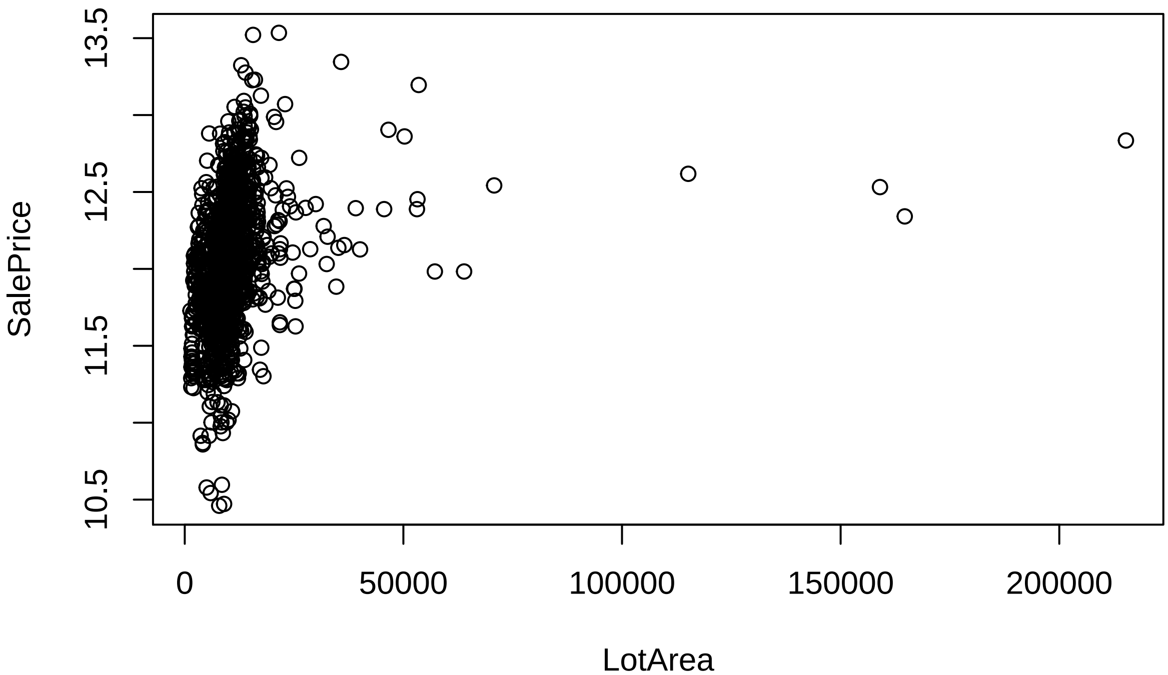
```r
plot(SalePrice ~ ., data= dataHouse)
```

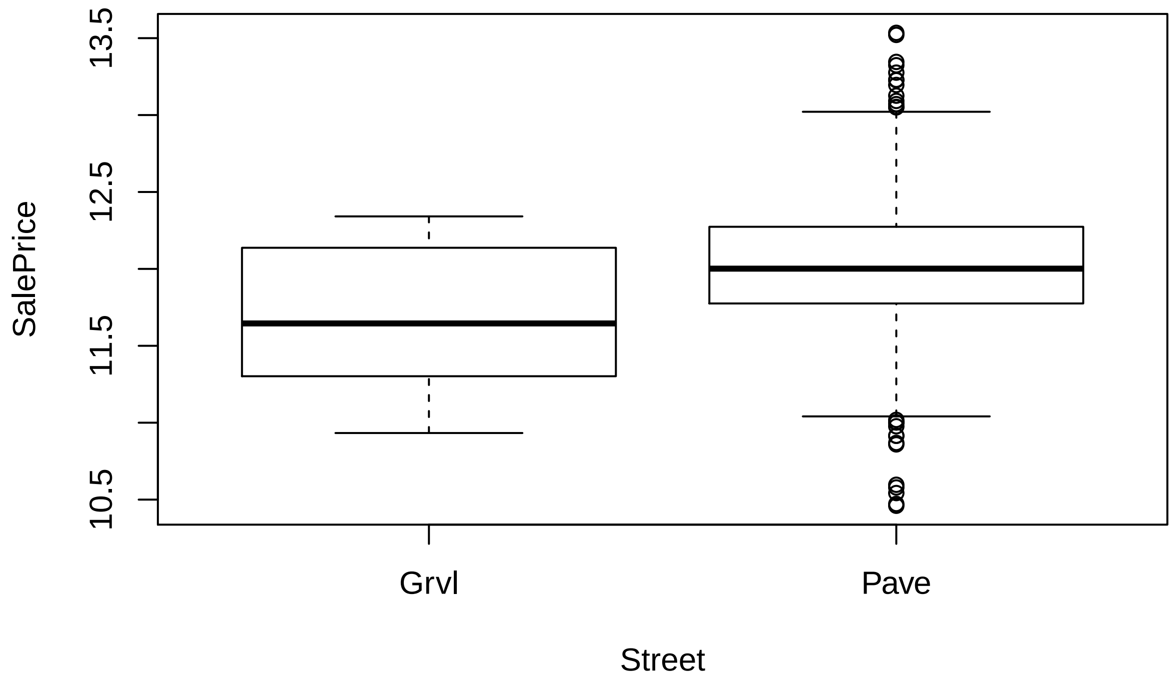

SalePrice
10.5
11.5
12.5
13.5
120
180
20
30
40
45
50
60
70
75
80
85
90
MSSubClass

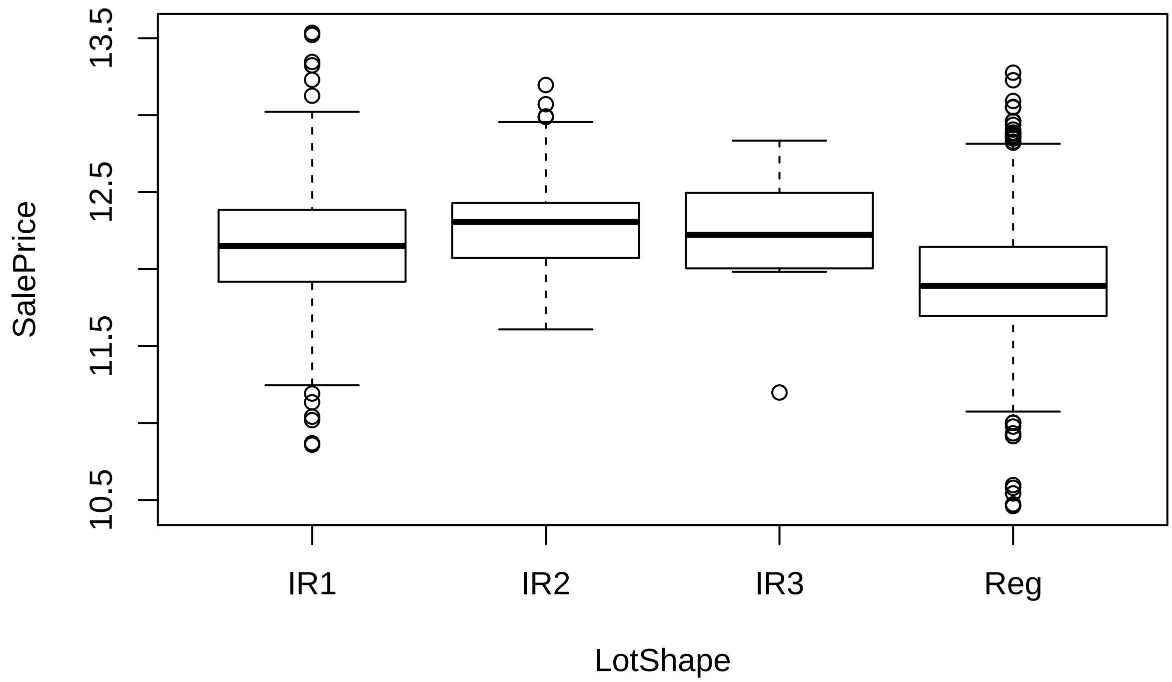


SalePrice
13.5
12.5
11.5
10.5
C (all)
FV
RH
RL
RM
MSZoning

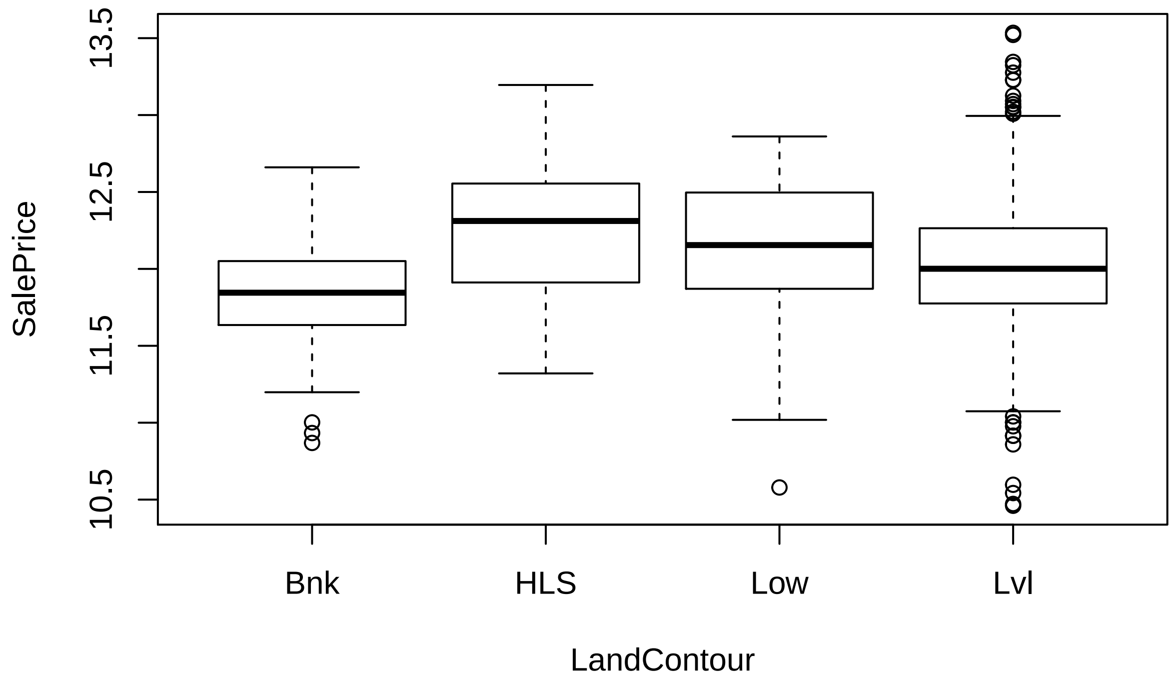

13.5
12.5
11.5
10.5
SalePrice
50
100
150
200
250
300
LotFrontage

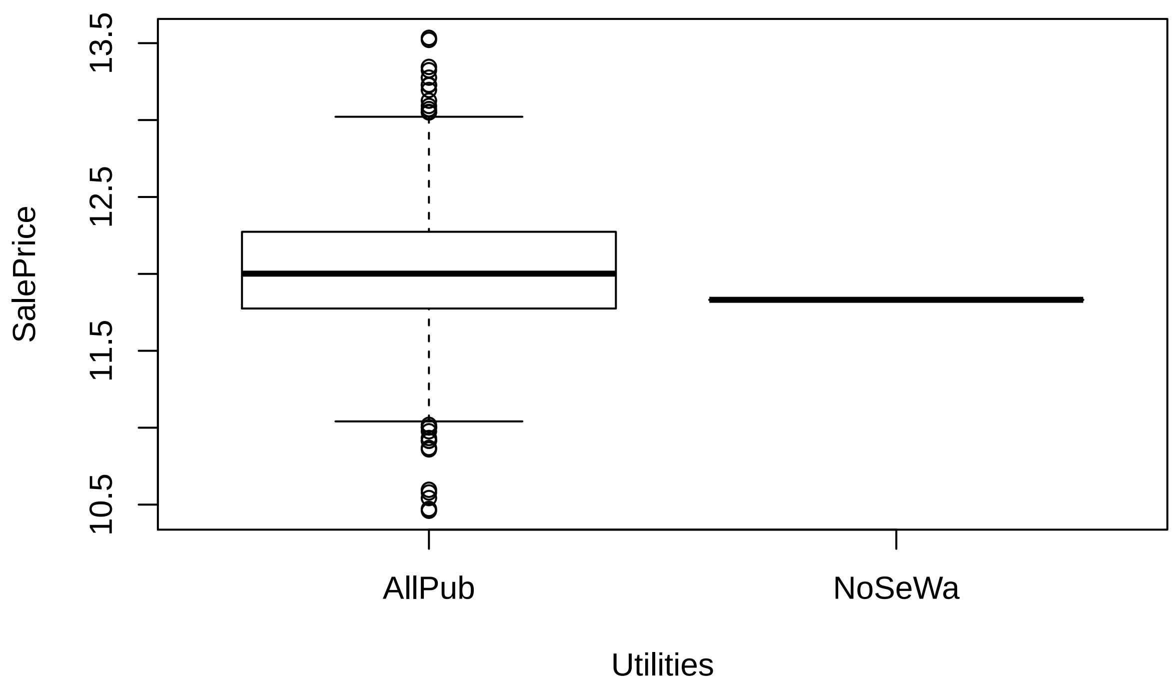

13.5
12.5
11.5
10.5
SalePrice
0
50000
100000
150000
200000
LotArea


13.5
12.5
11.5
10.5
SalePrice
Grvl
Pave
Street

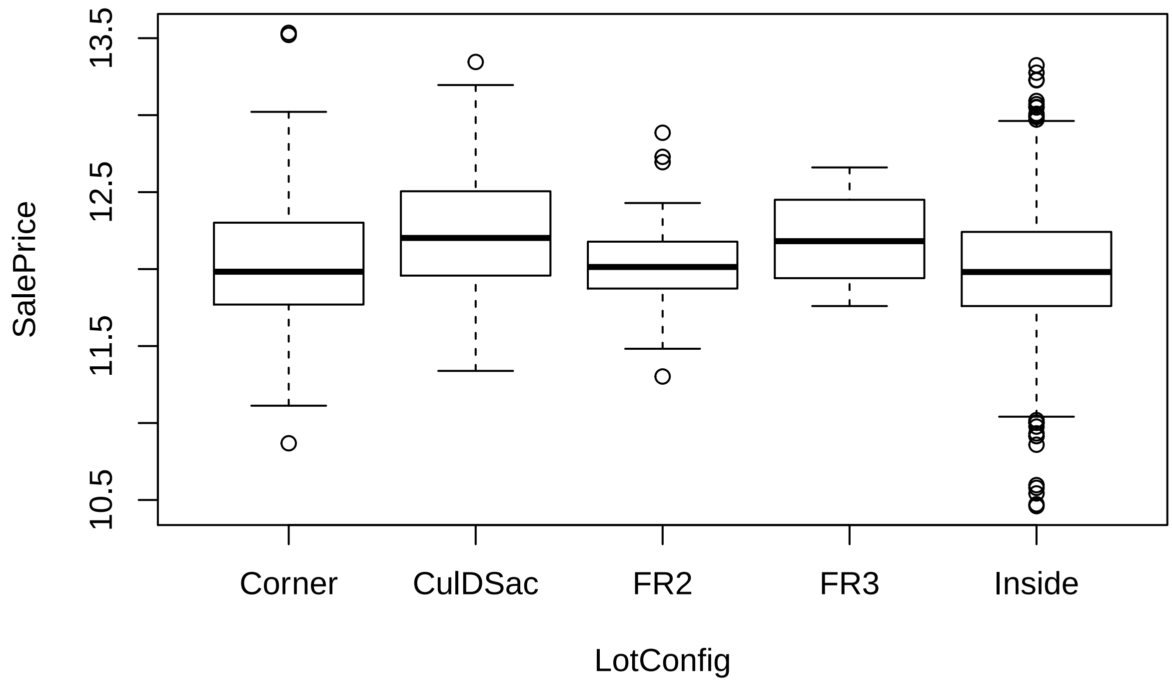


SalePrice
10.5
11.5
12.5
13.5
IR1
IR2
IR3
Reg
LotShape

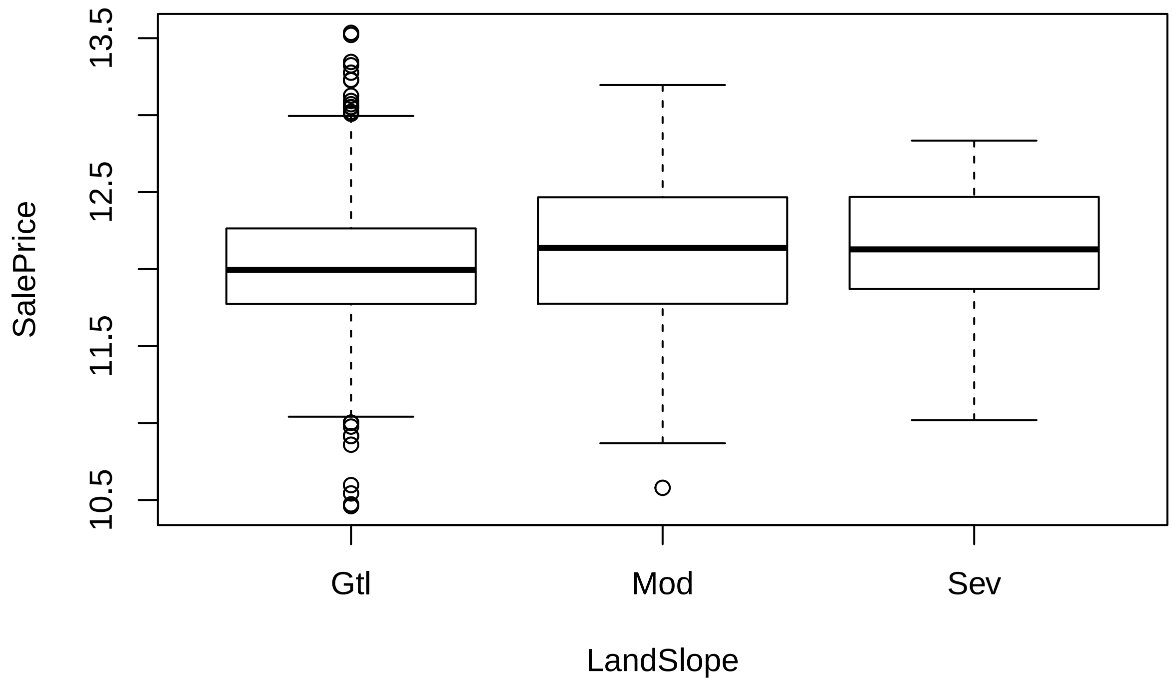

SalePrice
10.5
11.5
12.5
13.5
Bnk
HLS
Low
Lvl
LandContour

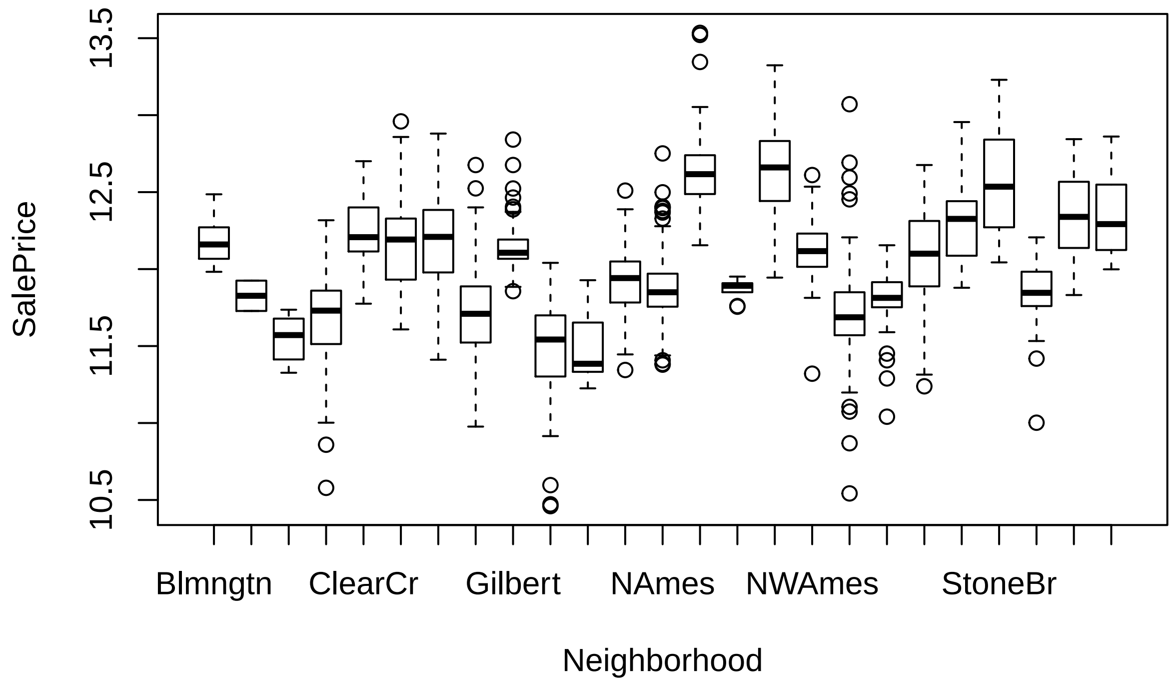


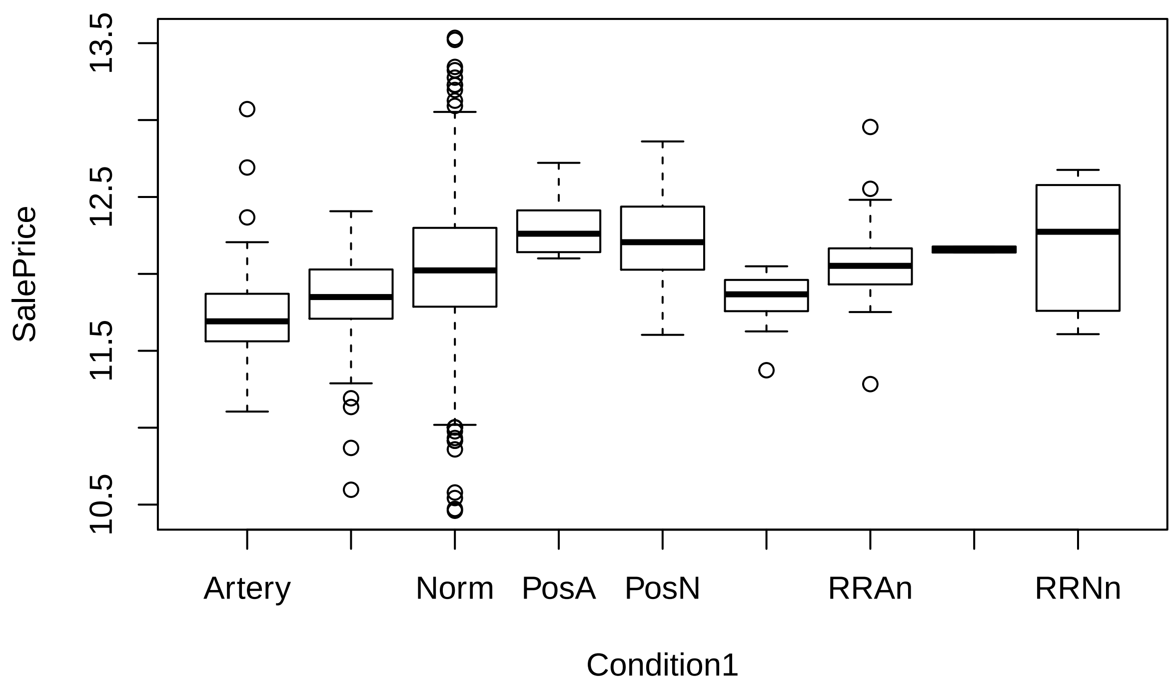
13.5
12.5
11.5
10.5
SalePrice
AllPub
NoSeWa
Utilities

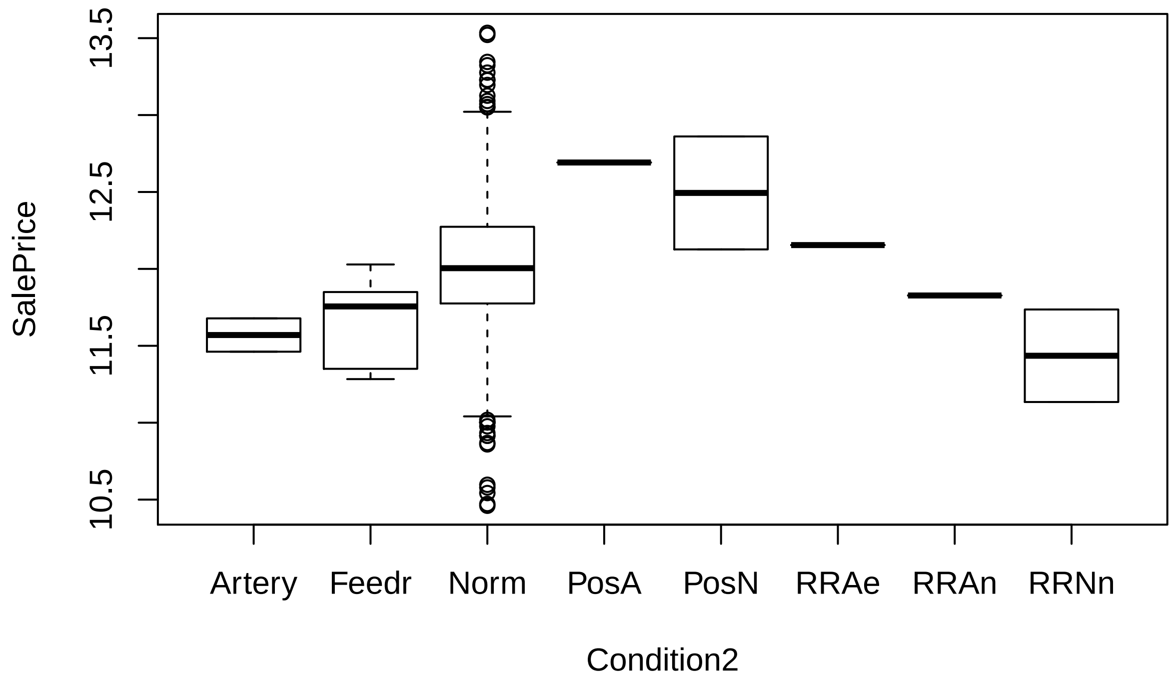


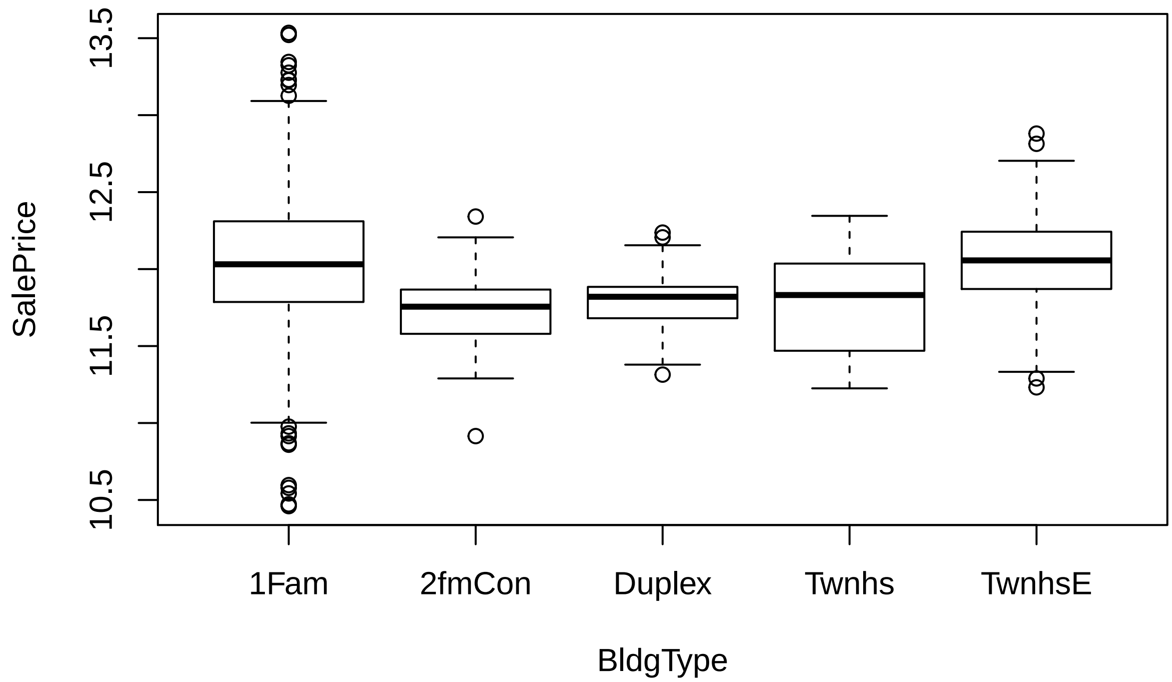
SalePrice
13.5
12.5
11.5
10.5
Corner
CulDSac
FR2
FR3
Inside
LotConfig

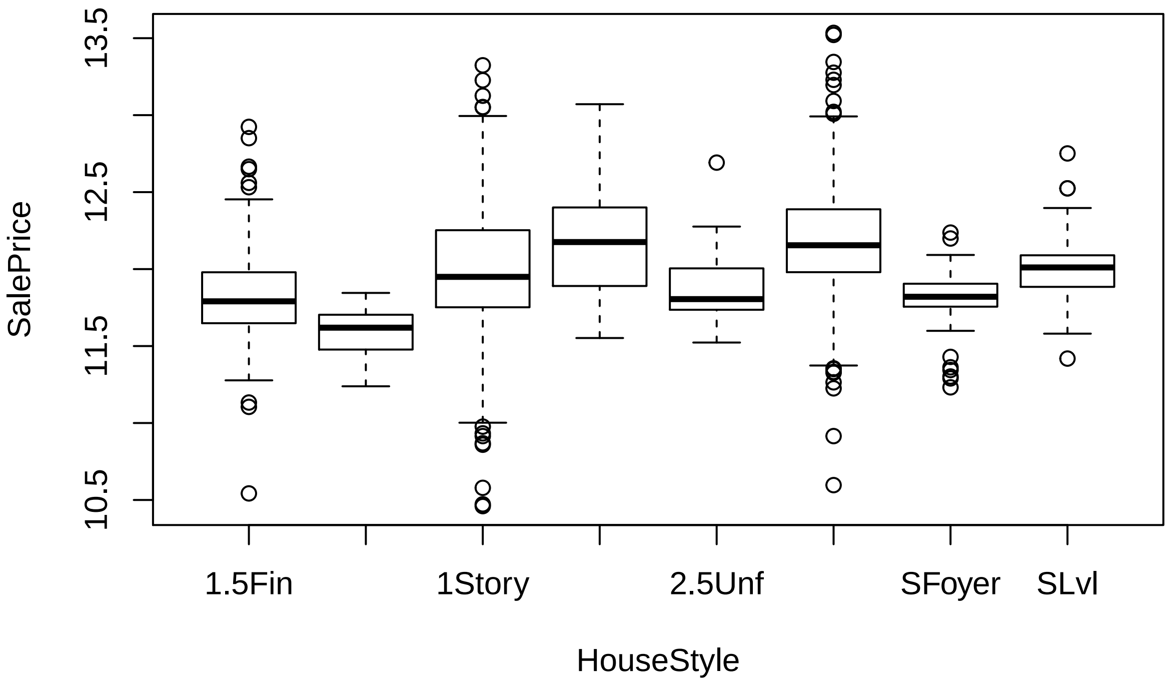

SalePrice
13.5
12.5
11.5
10.5
Gtl
Mod
Sev
LandSlope


13.5
12.5
11.5
10.5
SalePrice
Blmngtn
ClearCr
Gilbert
NAmes
NWAmes
StoneBr
Neighborhood

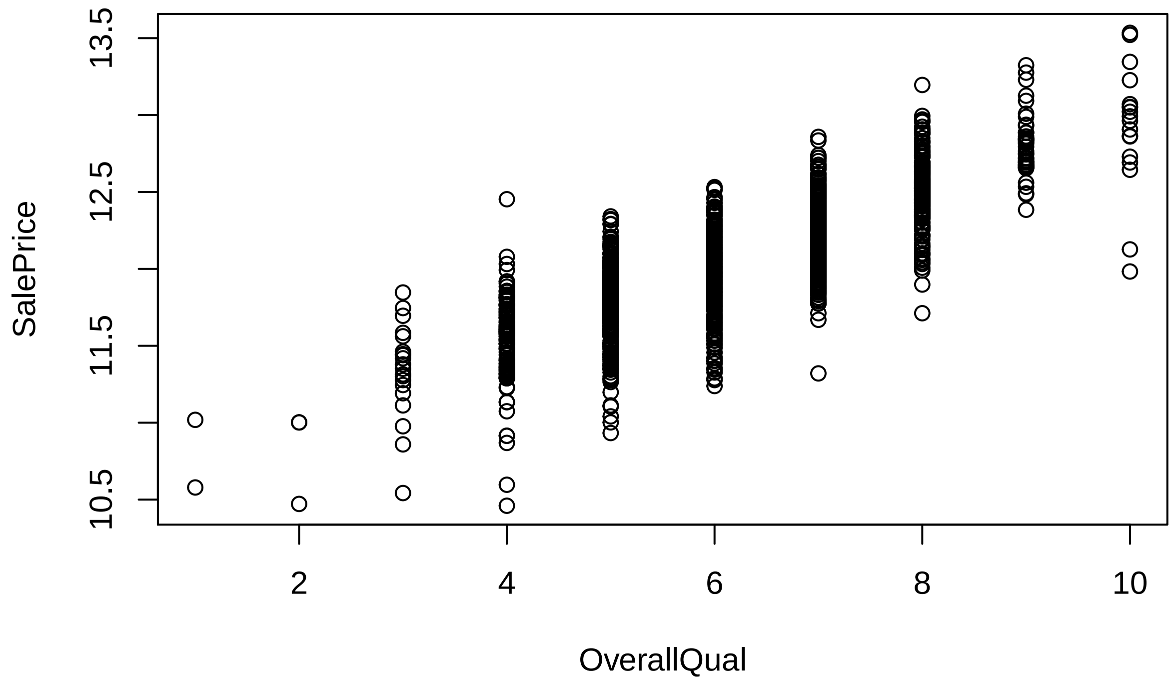

SalePrice
13.5
12.5
11.5
10.5
Artery
Norm
PosA
PosN
RRAn
RRNn
Condition1

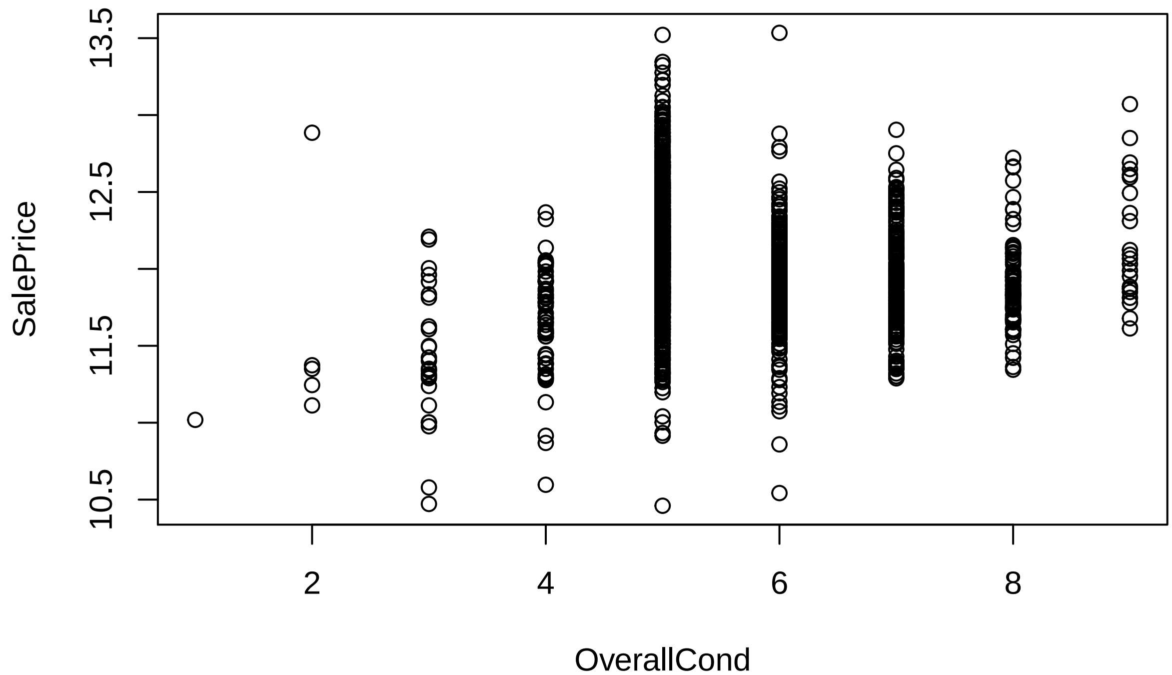

SalePrice
13.5
12.5
11.5
10.5
Artery
Feedr
Norm
PosA
PosN
RRAe
RRAn
RRNn
Condition2

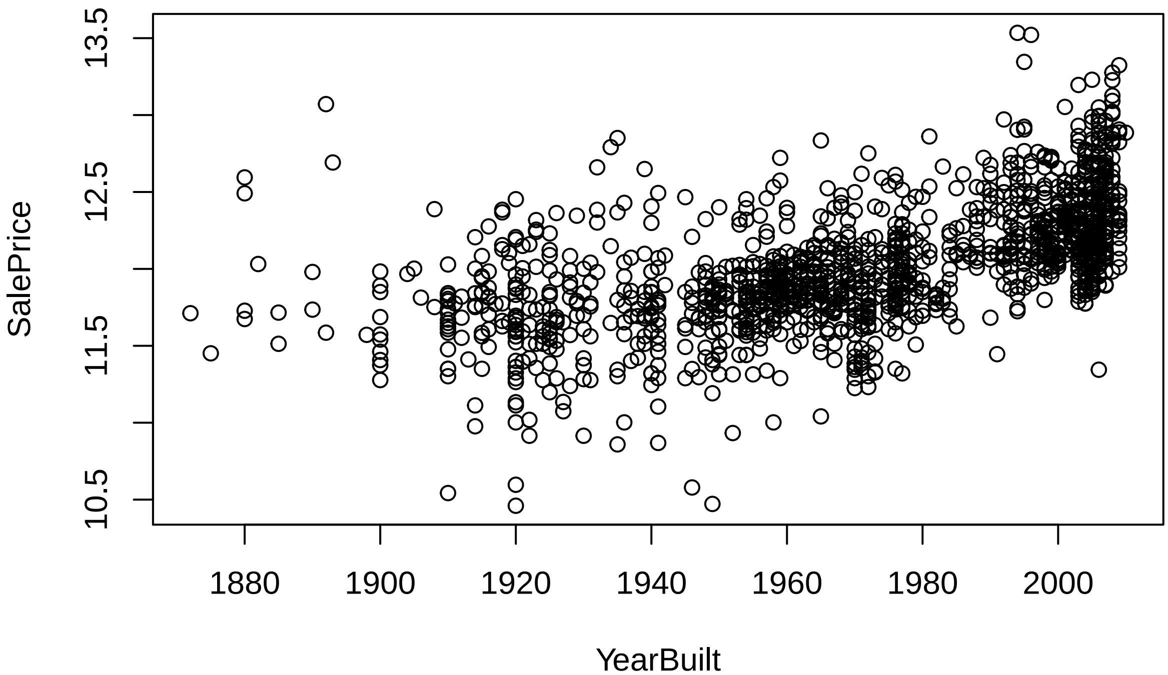

13.5
12.5
11.5
10.5
SalePrice
1Fam
2fmCon
Duplex
Twnhs
TwnhsE
BldgType



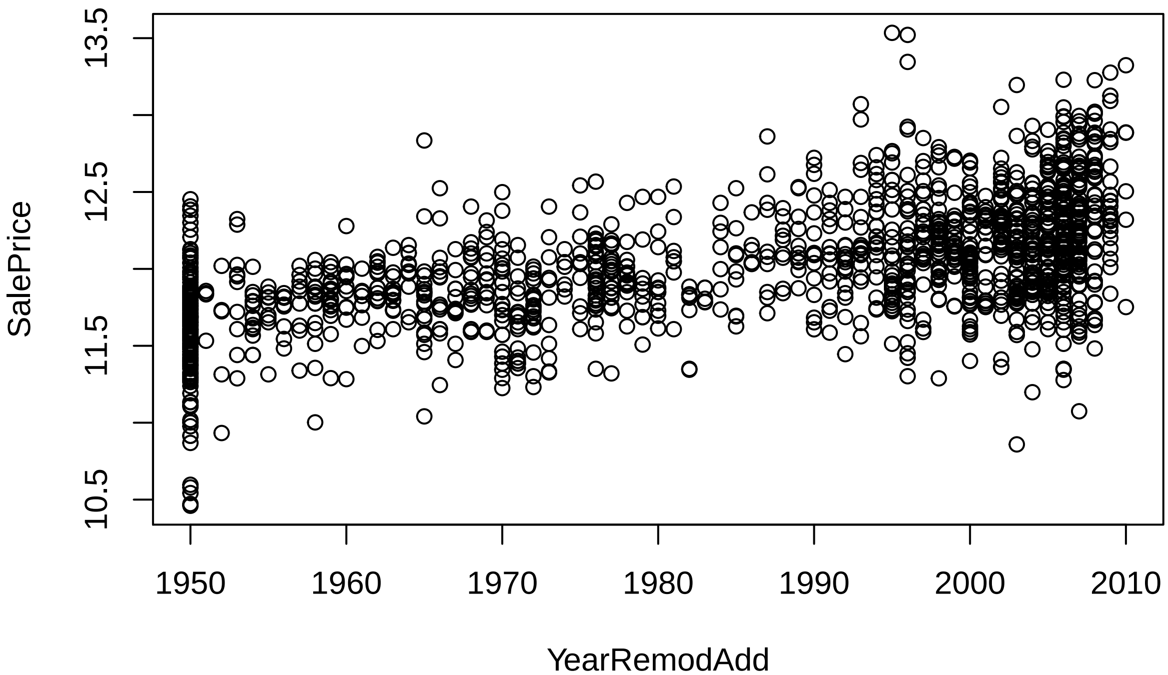
SalePrice
13.5
12.5
11.5
10.5
1.5Fin
1Story
2.5Unf
SFoyer
SLvl
HouseStyle

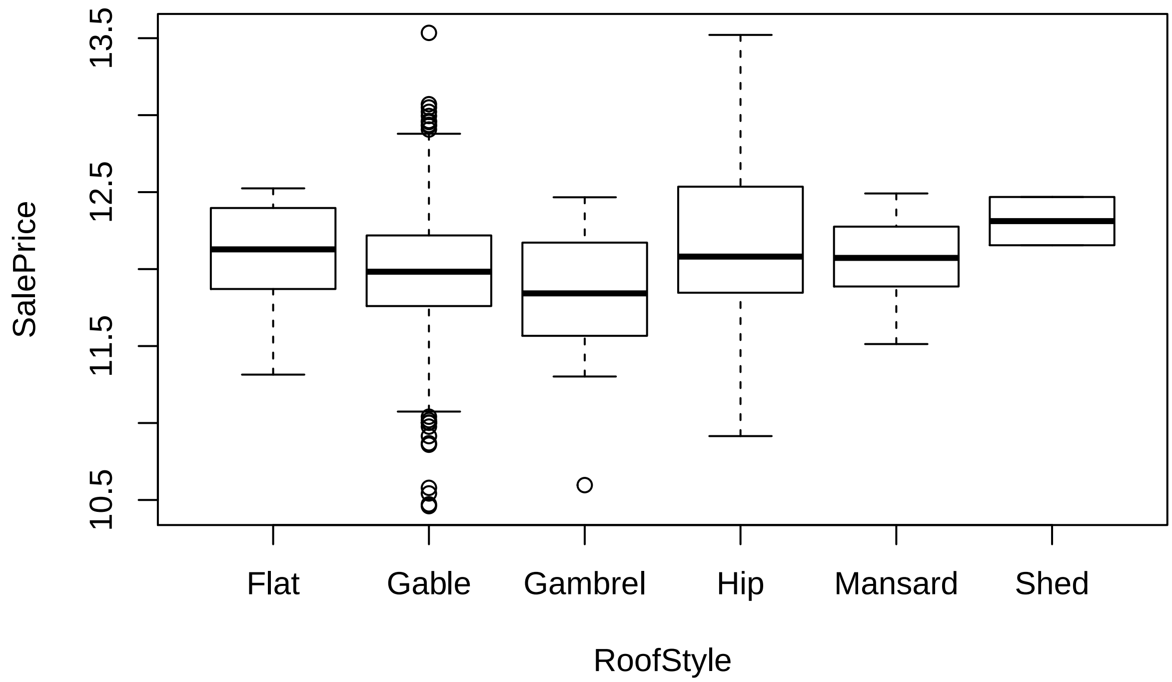

13.5
12.5
11.5
10.5
SalePrice
2
4
6
8
10
OverallQual

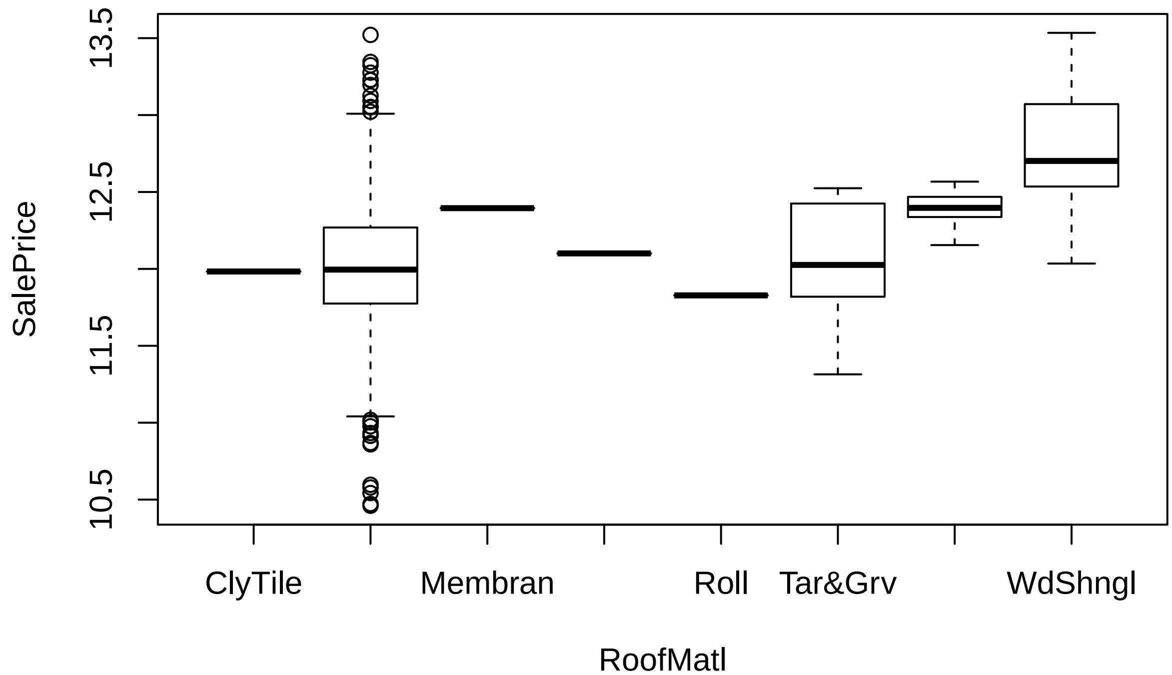

SalePrice
13.5
12.5
11.5
10.5
2
4
6
8
OverallCond

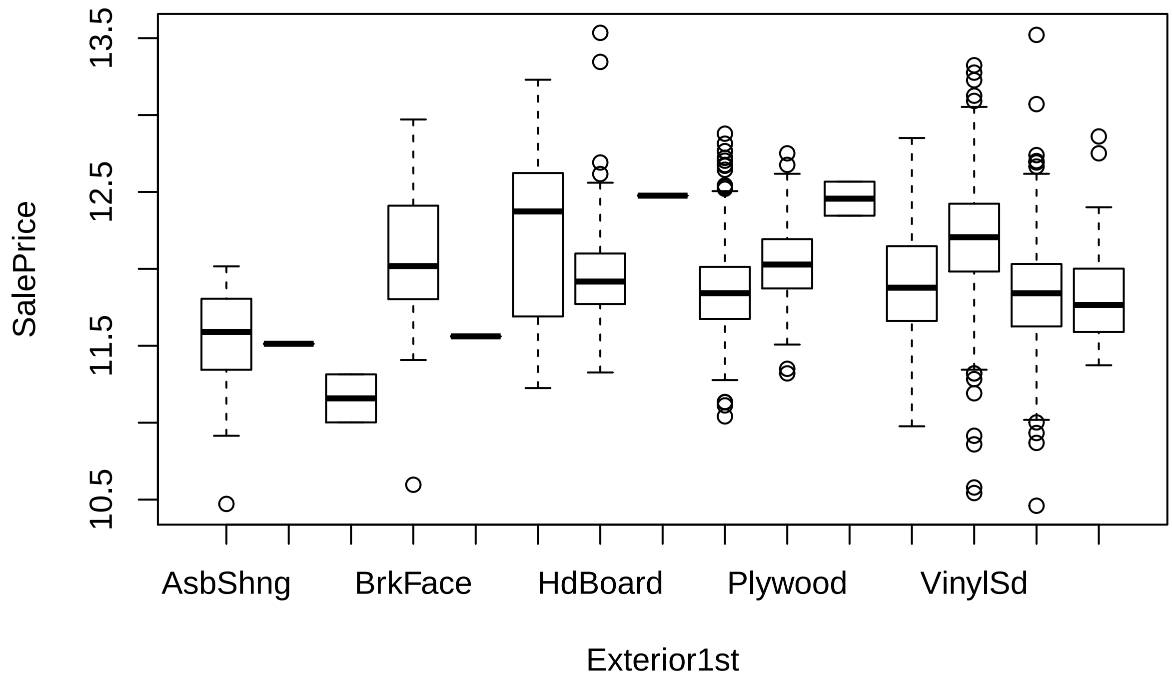


SalePrice
10.5
11.5
12.5
13.5
1880
1900
1920
1940
1960
1980
2000
YearBuilt

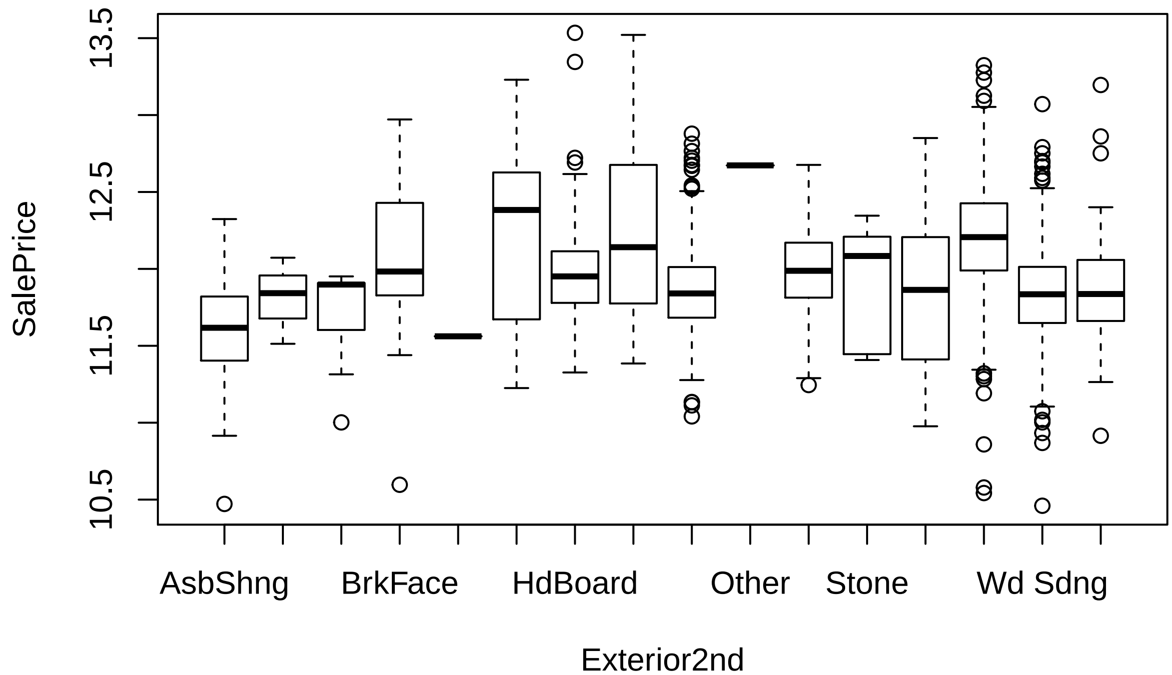

SalePrice
13.5
12.5
11.5
10.5
1950
1960
1970
1980
1990
2000
2010
YearRemodAdd


SalePrice
10.5
11.5
12.5
13.5
Flat
Gable
Gambrel
Hip
Mansard
Shed
RoofStyle

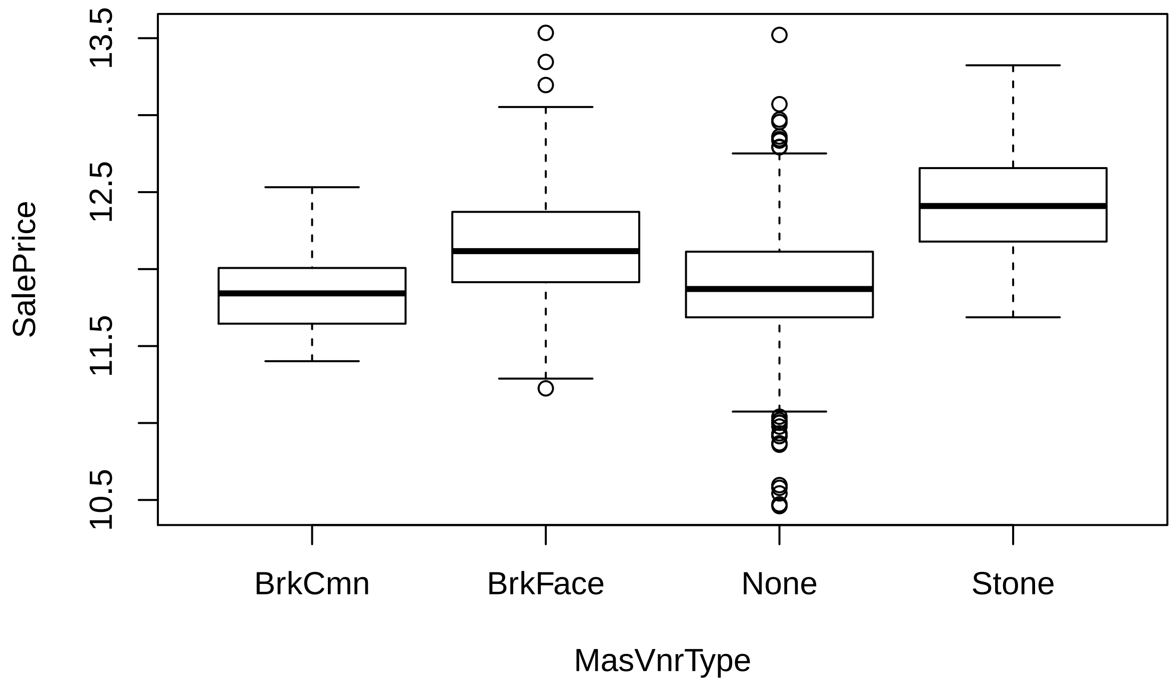

SalePrice
13.5
12.5
11.5
10.5
ClyTile
Membran
Roll
Tar&Grv
WdShngl
RoofMatl

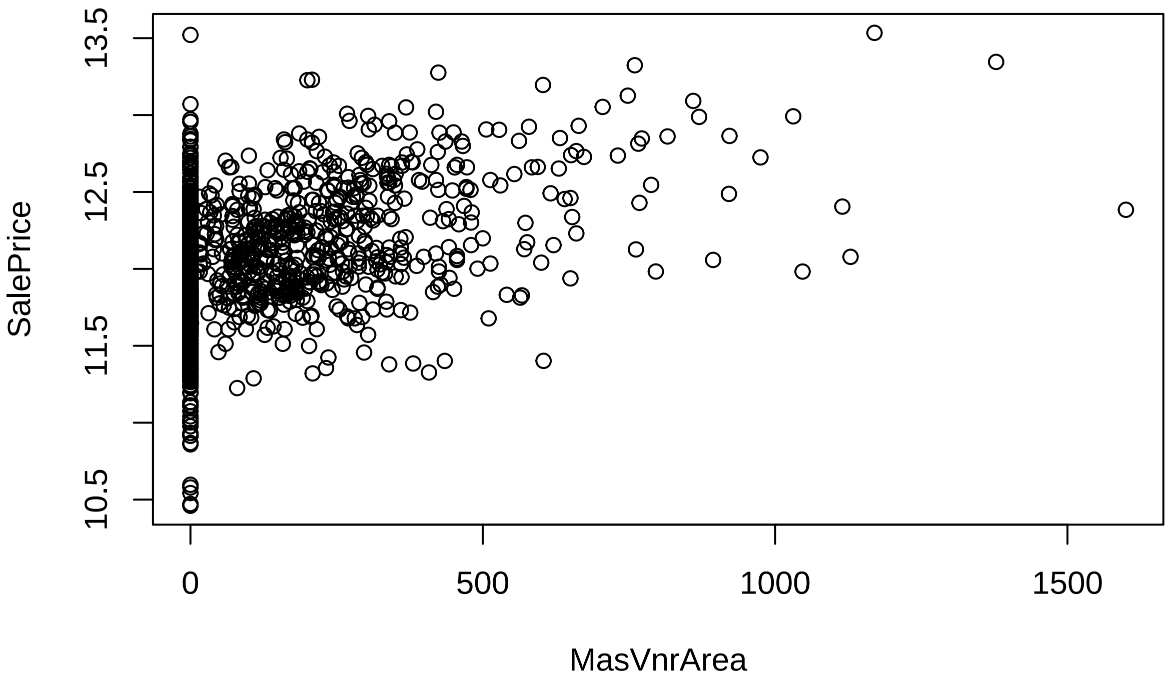


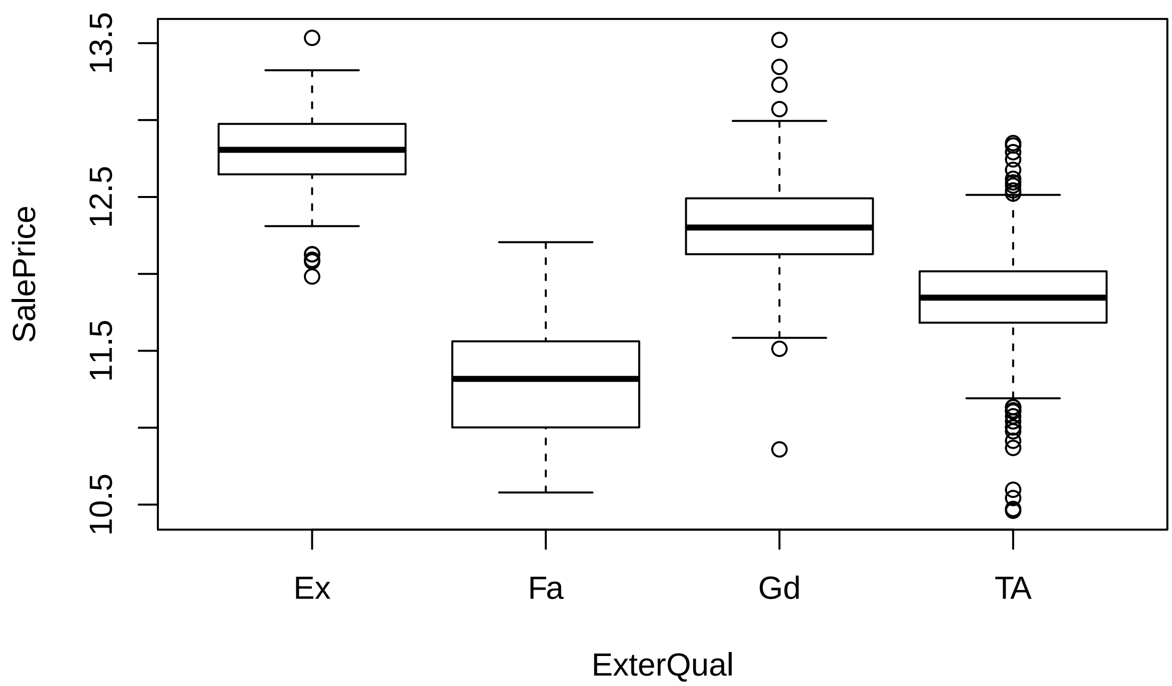
SalePrice
10.5
11.5
12.5
13.5
AsbShng
BrkFace
HdBoard
Plywood
VinylSd
Exterior1st

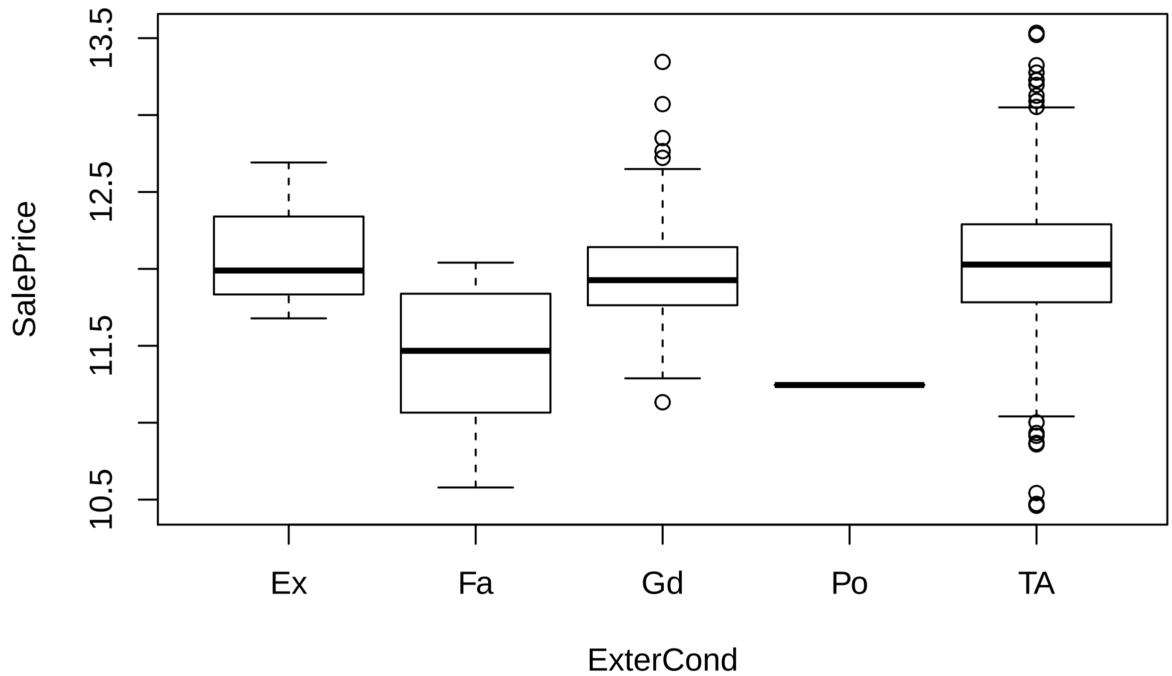


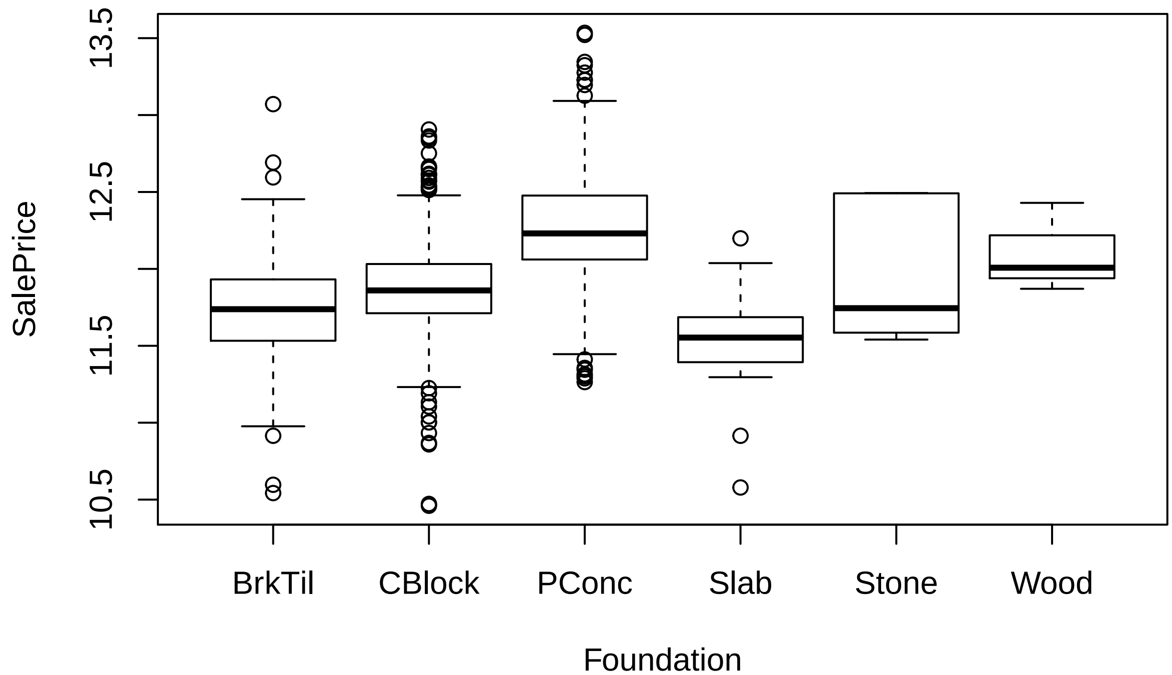
SalePrice
13.5
12.5
11.5
10.5
AsbShng
BrkFace
HdBoard
Other
Stone
Wd Sdng
Exterior2nd


SalePrice
10.5
11.5
12.5
13.5
BrkCmn
BrkFace
None
Stone
MasVnrType

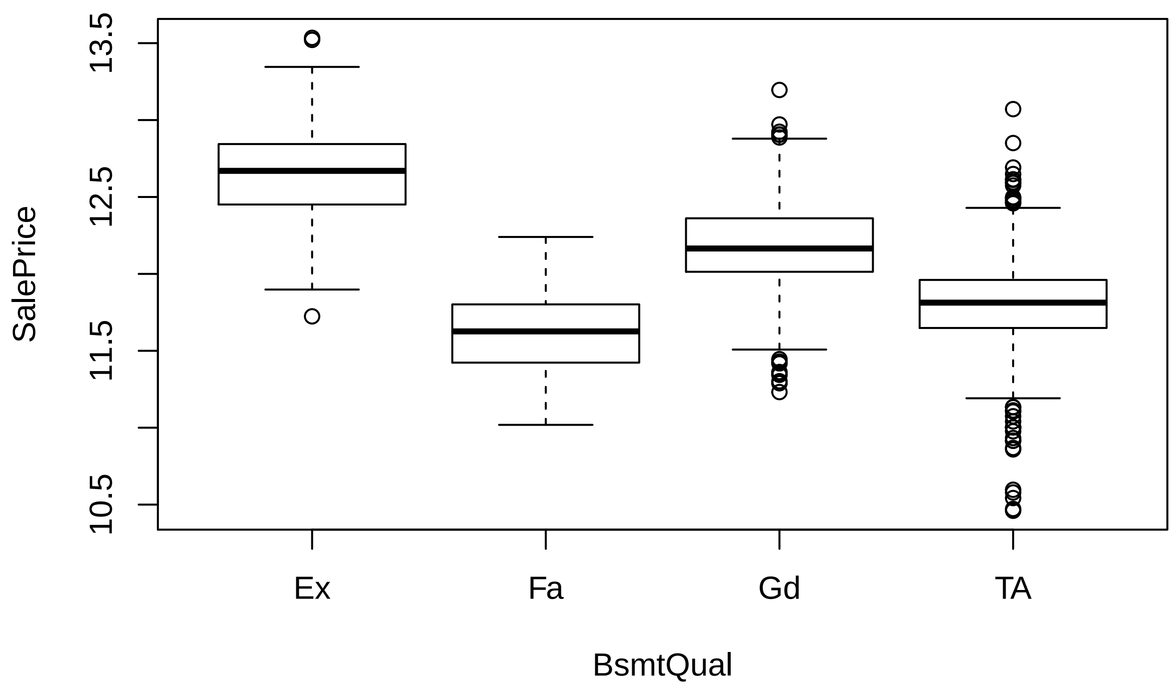

SalePrice
MasVnrArea
10.5
11.5
12.5
13.5
0
500
1000
1500

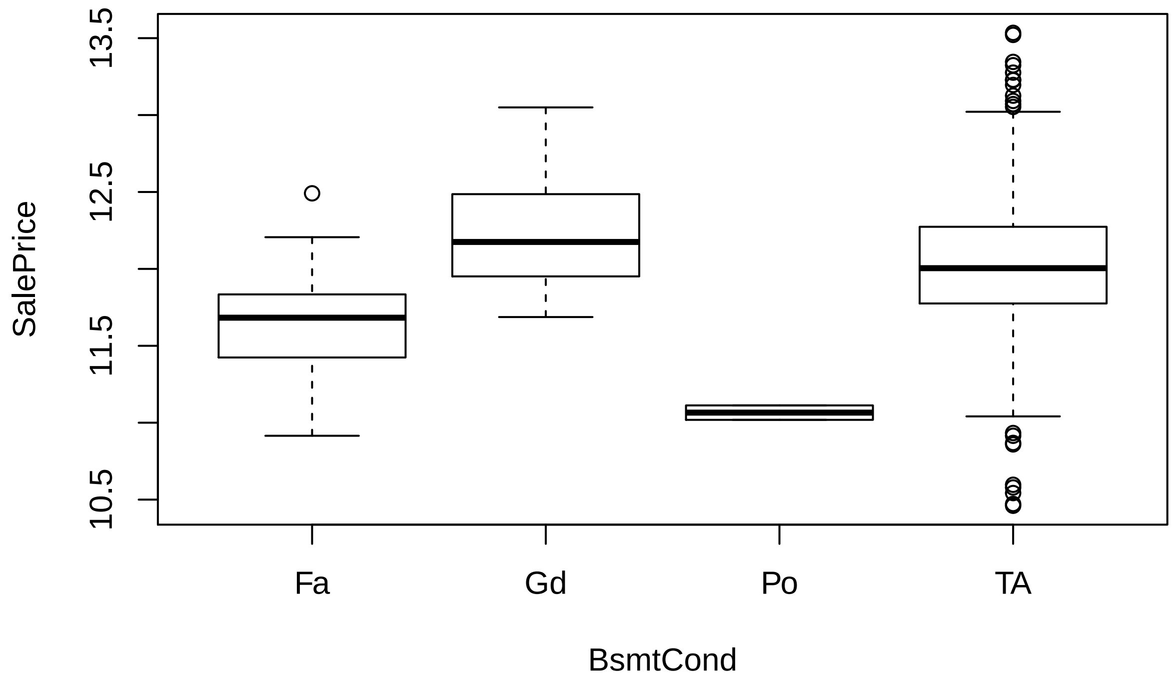

13.5
12.5
11.5
10.5
SalePrice
Ex
Fa
Gd
TA
ExterQual


13.5
12.5
11.5
10.5
SalePrice
Ex
Fa
Gd
Po
TA
ExterCond

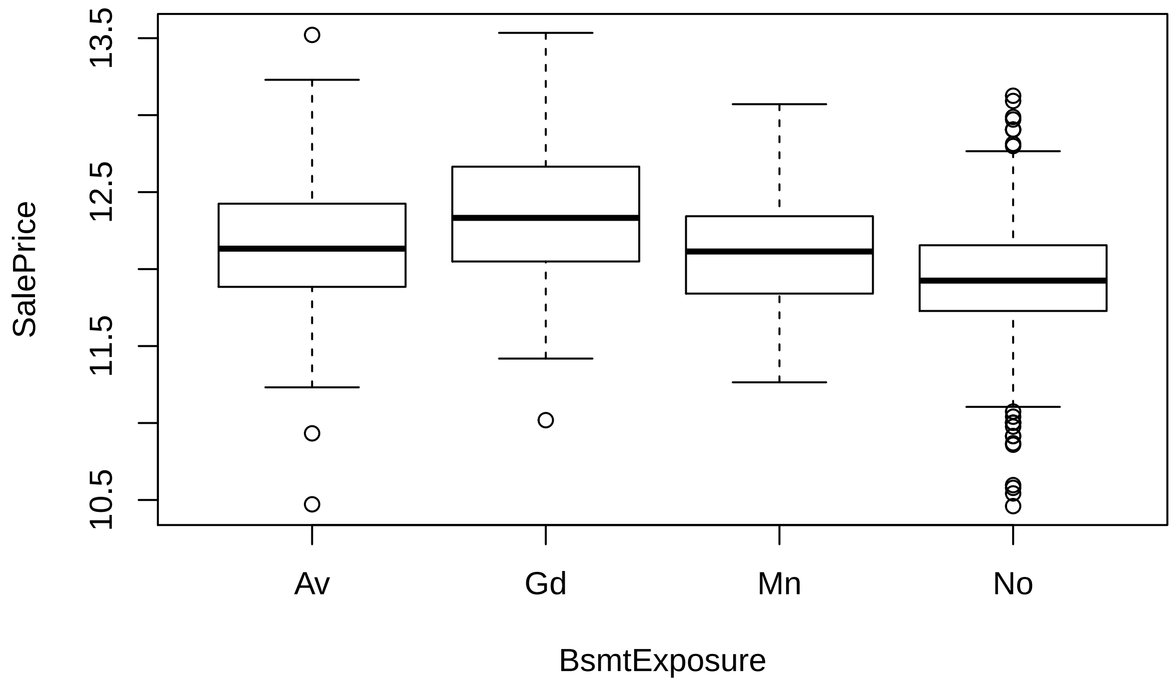


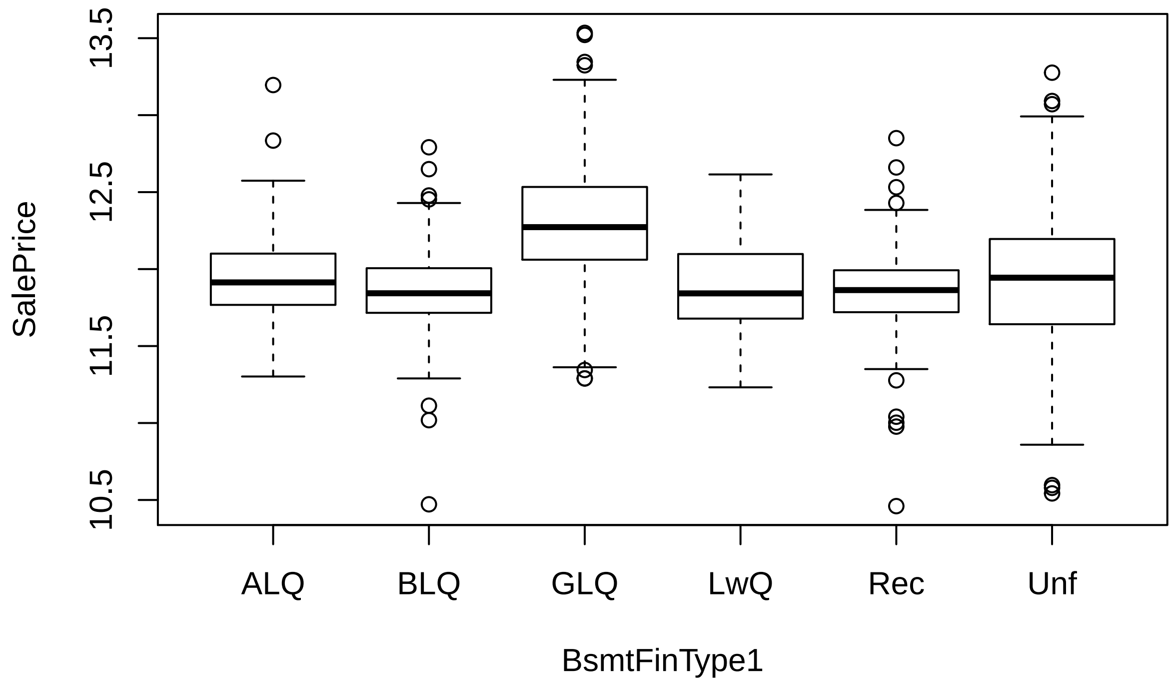
13.5
12.5
11.5
10.5
SalePrice
BrkTil
CBlock
PConc
Slab
Stone
Wood
Foundation

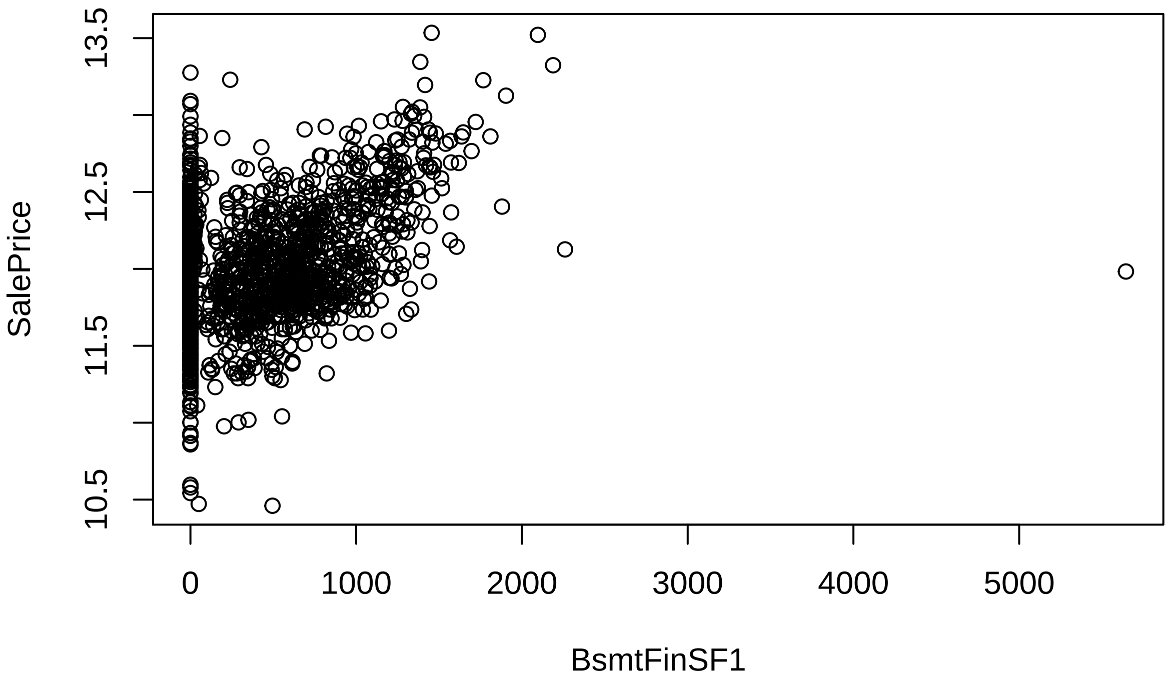

13.5
12.5
11.5
10.5
SalePrice
Ex
Fa
Gd
TA
BsmtQual



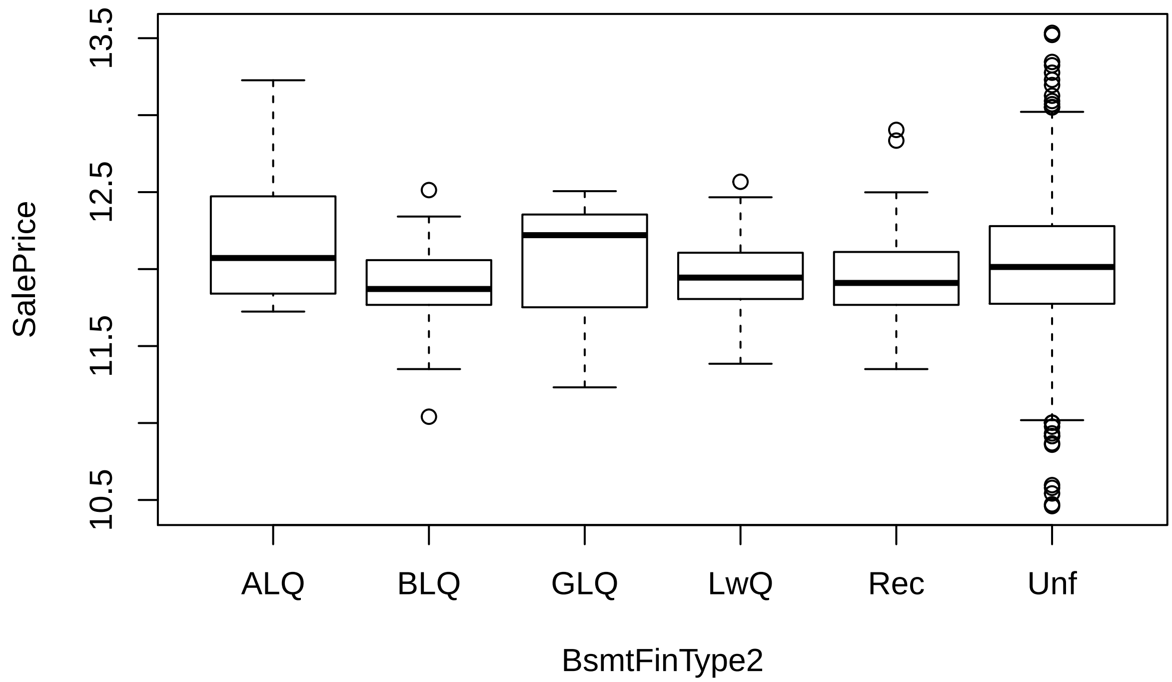
SalePrice
10.5
11.5
12.5
13.5
Fa
Gd
Po
TA
BsmtCond

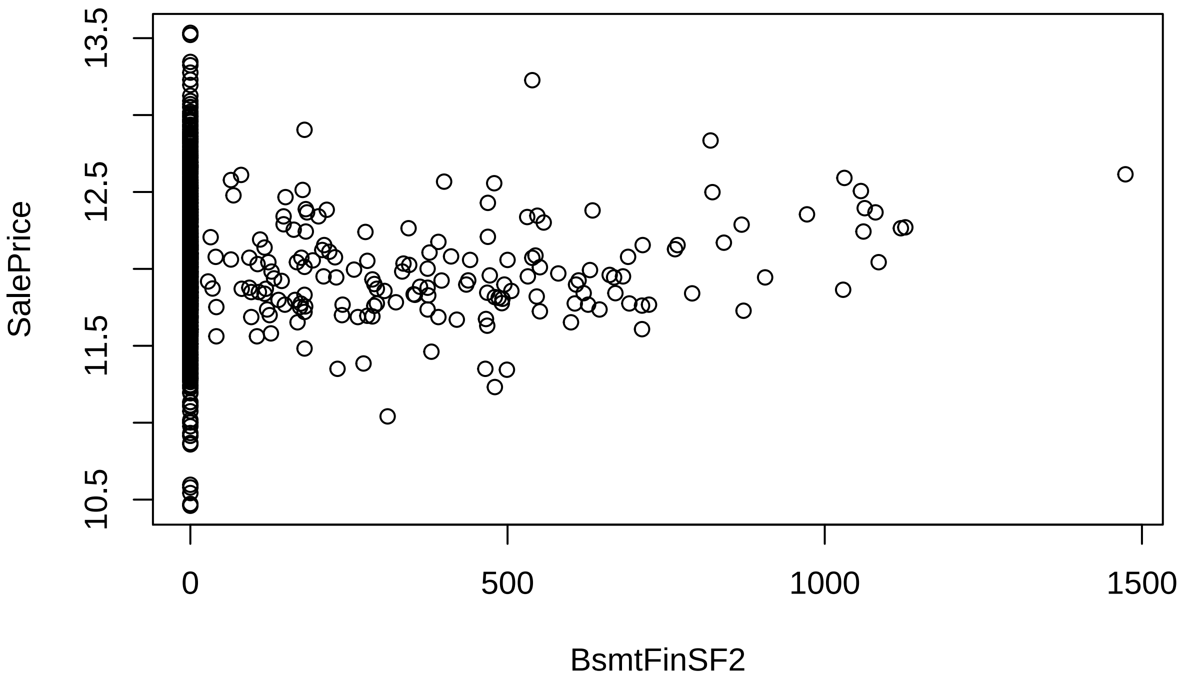

13.5
12.5
11.5
10.5
SalePrice
Av
Gd
Mn
No
BsmtExposure


13.5
12.5
11.5
10.5
SalePrice
ALQ
BLQ
GLQ
LwQ
Rec
Unf
BsmtFinType1

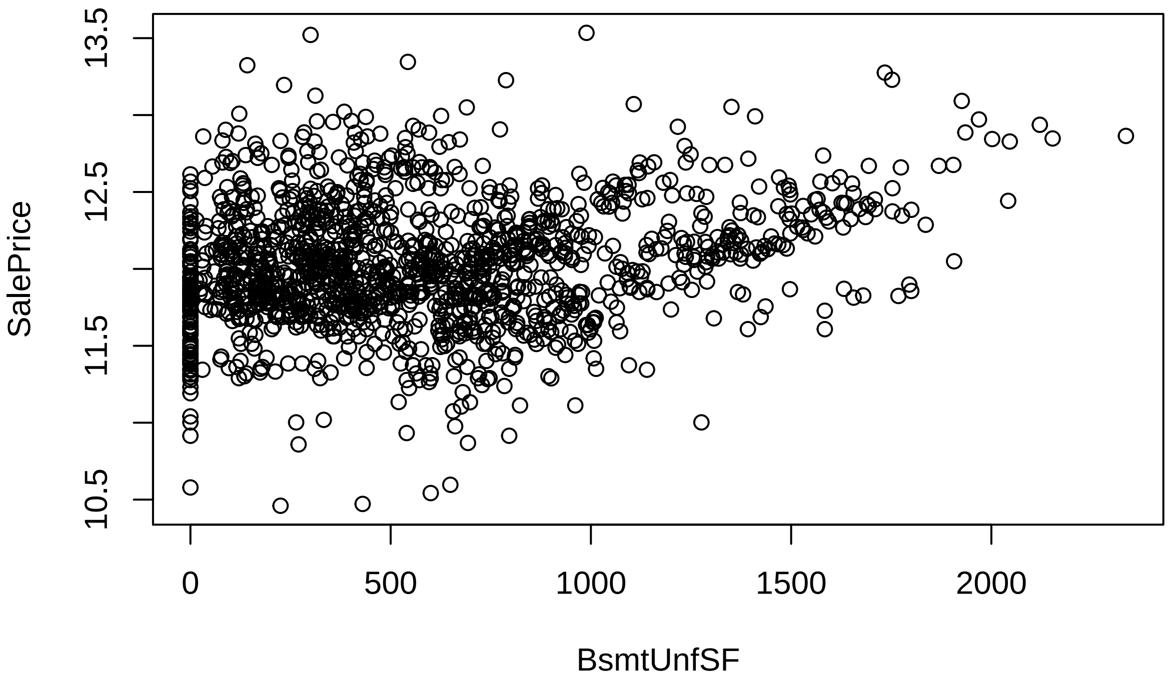


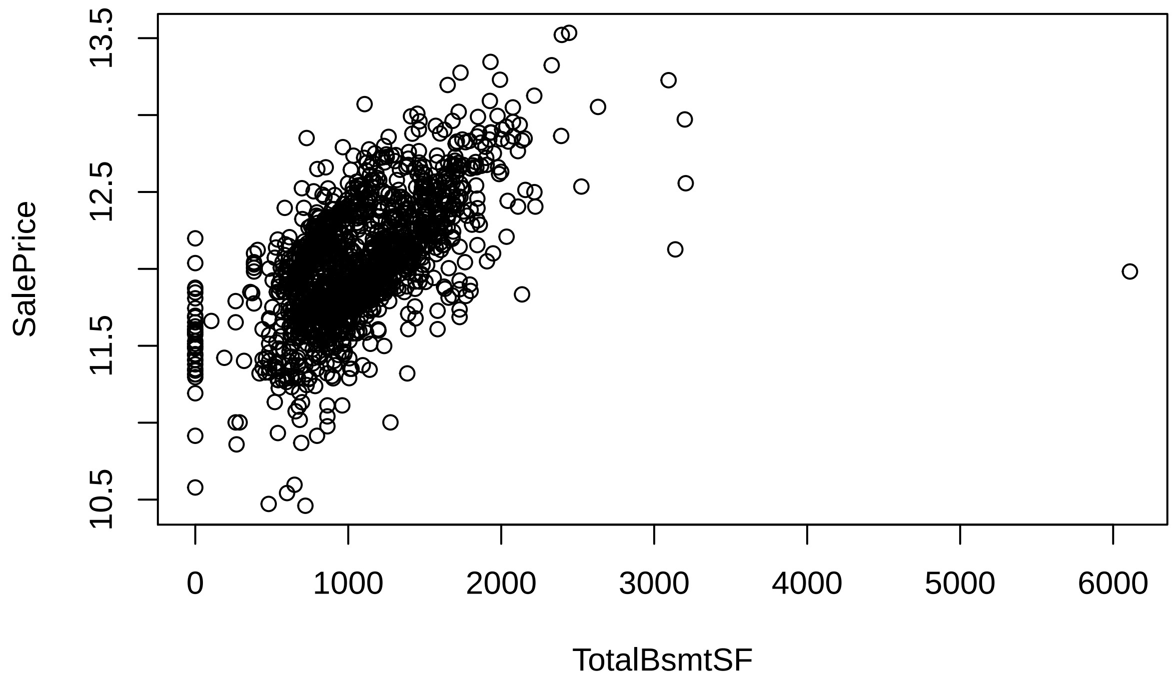
SalePrice
13.5
12.5
11.5
10.5
0
1000
2000
3000
4000
5000
BsmtFinSF1

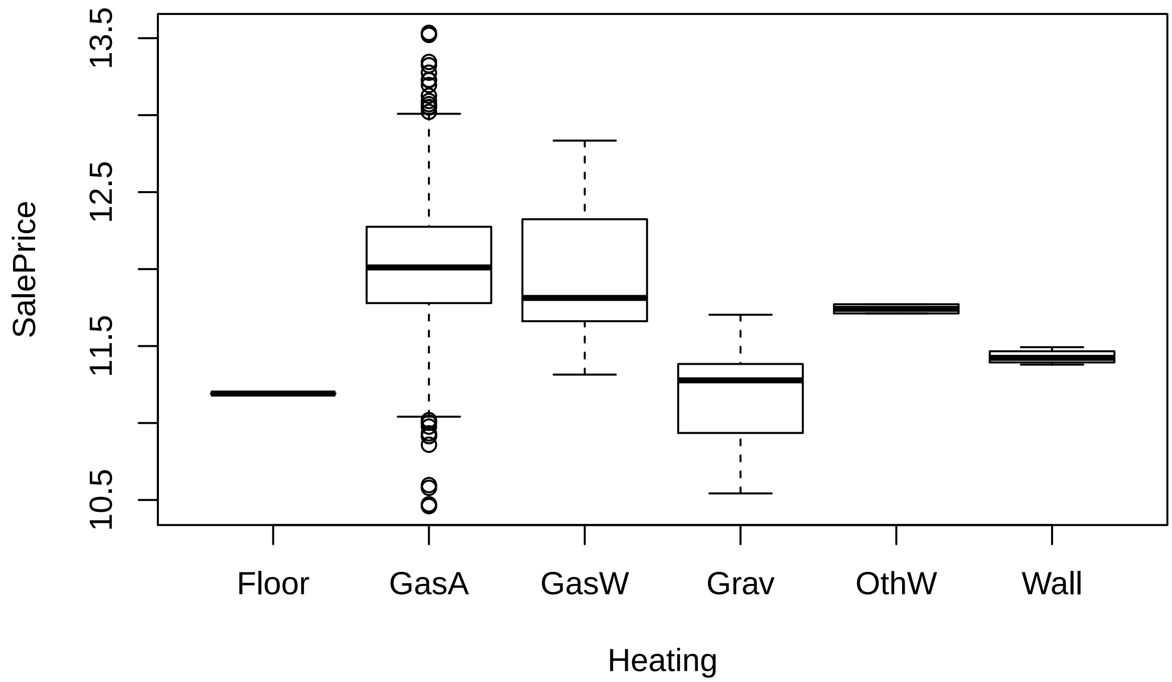

SalePrice
10.5
11.5
12.5
13.5
ALQ
BLQ
GLQ
LwQ
Rec
Unf
BsmtFinType2



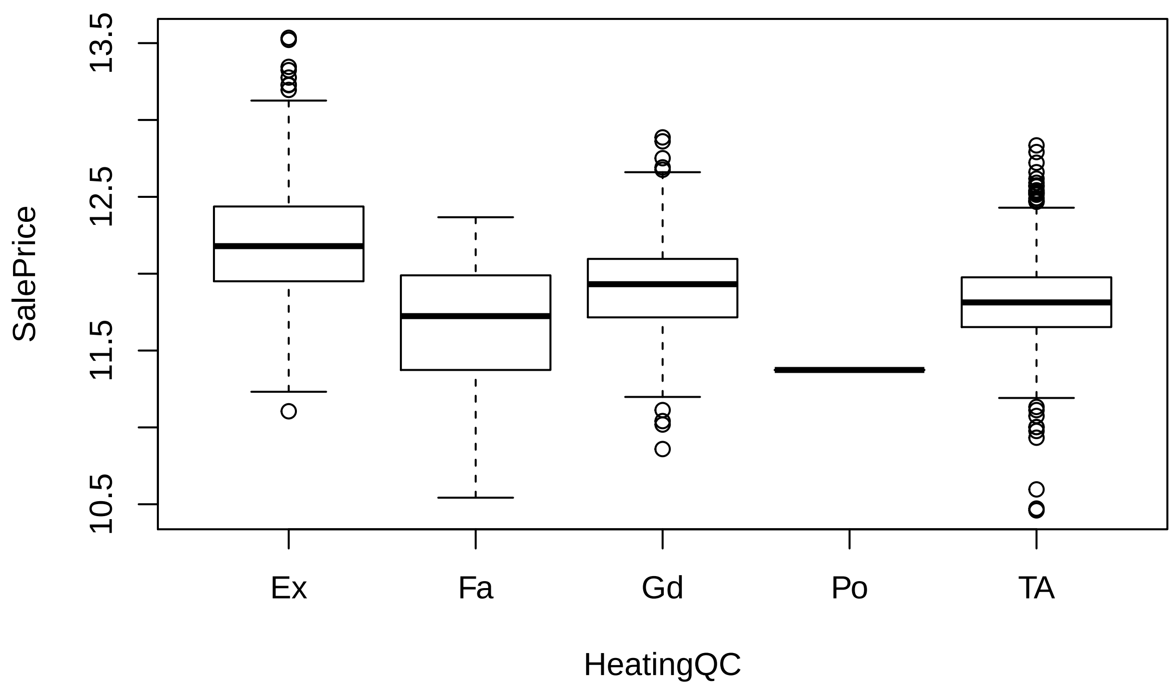
SalePrice
10.5
11.5
12.5
13.5
0
500
1000
1500
BsmtFinSF2

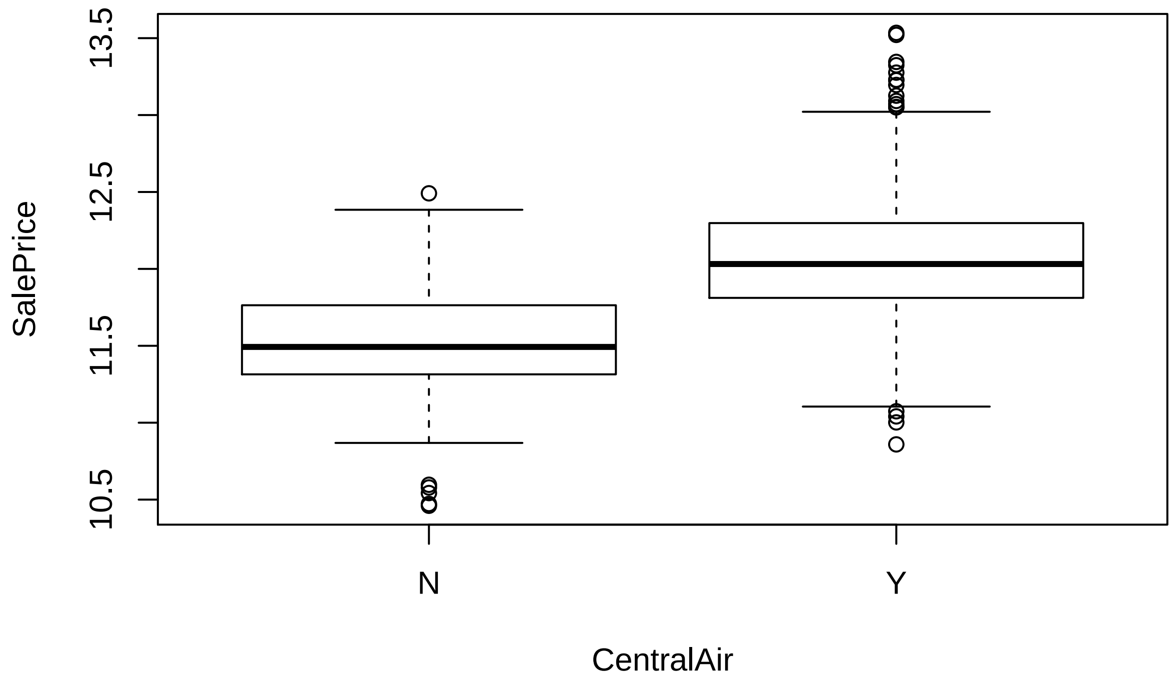

13.5
12.5
11.5
10.5
SalePrice
0
500
1000
1500
2000
BsmtUnfSF

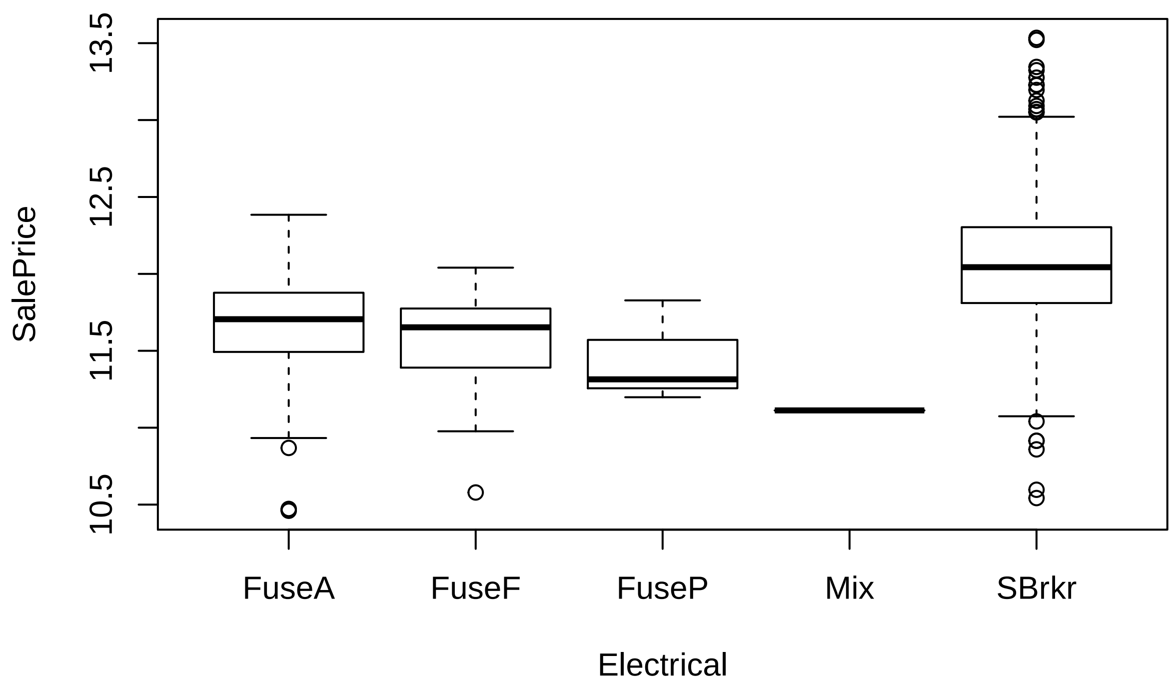

SalePrice
10.5
11.5
12.5
13.5
0
1000
2000
3000
4000
5000
6000
TotalBsmtSF

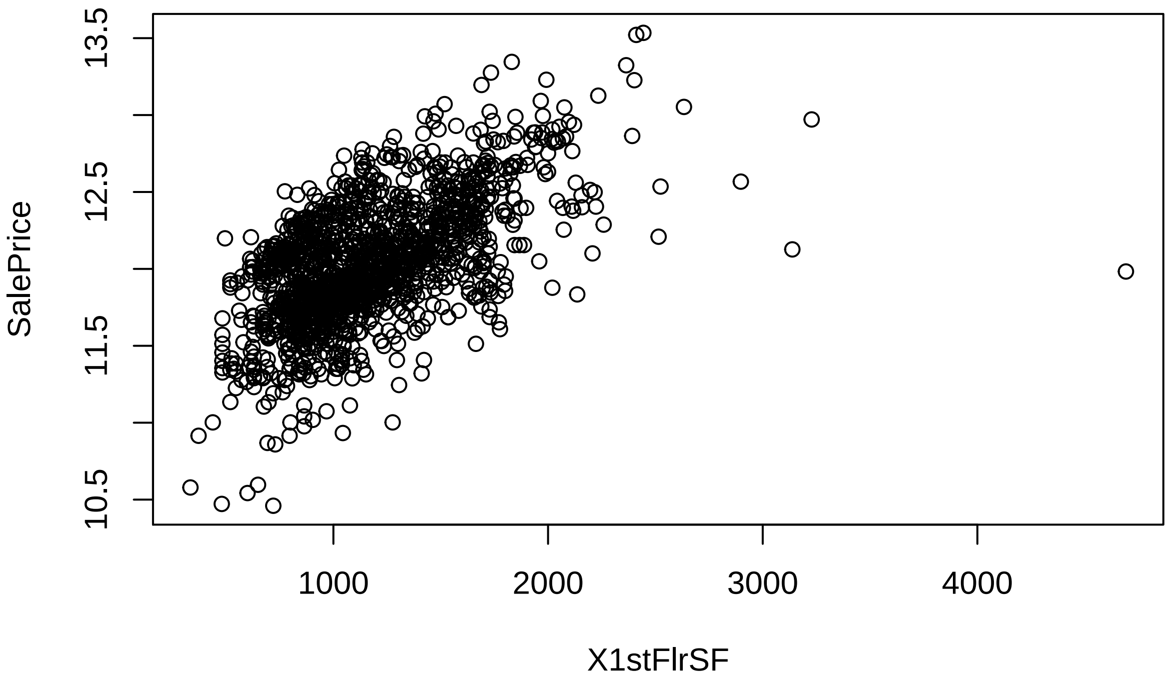


SalePrice
13.5
12.5
11.5
10.5
Floor
GasA
GasW
Grav
OthW
Wall
Heating

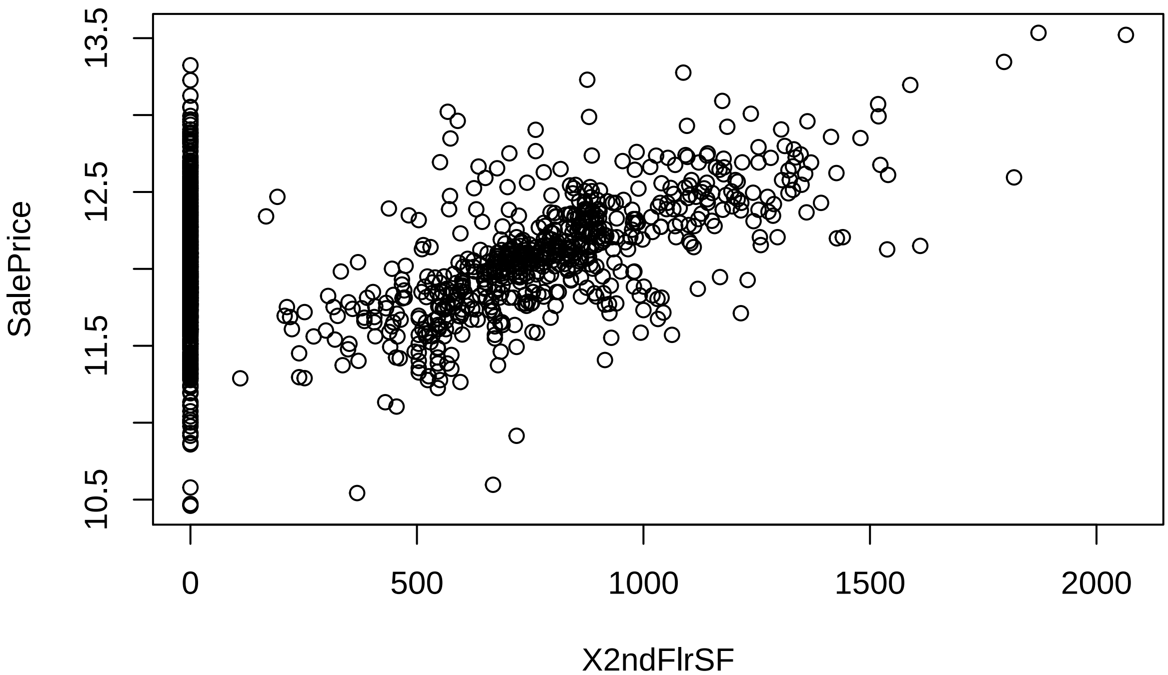

13.5
12.5
11.5
10.5
SalePrice
Ex
Fa
Gd
Po
TA
HeatingQC


SalePrice
13.5
12.5
11.5
10.5
N
Y
CentralAir

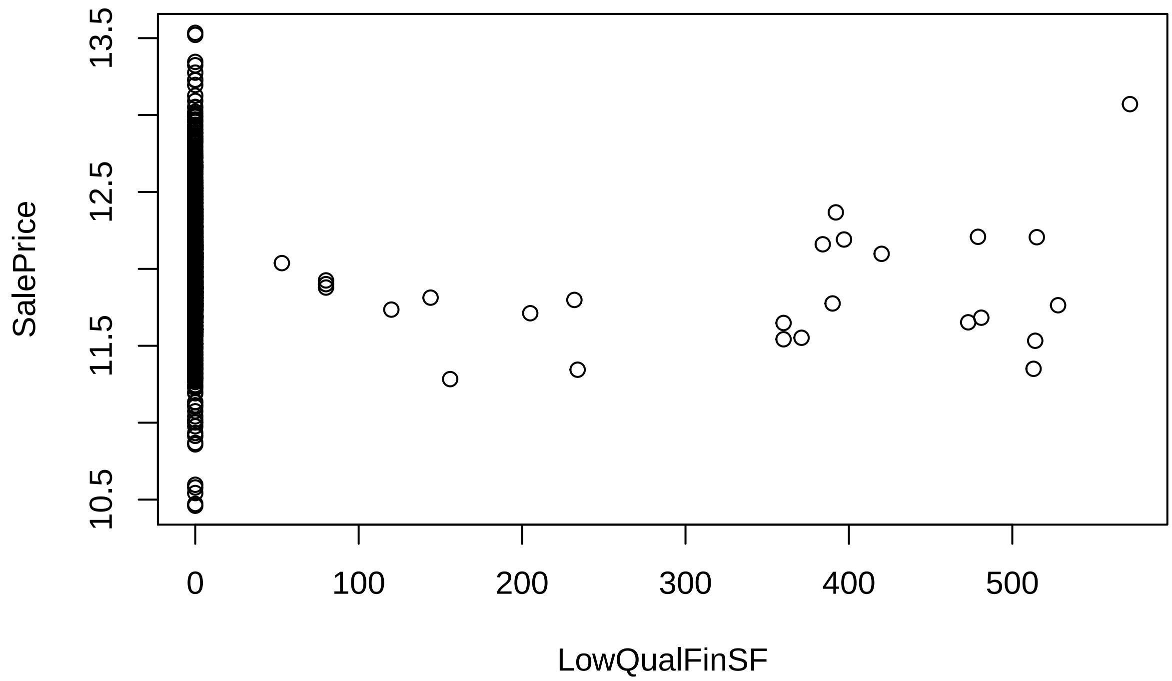

SalePrice
10.5
11.5
12.5
13.5
FuseA
FuseF
FuseP
Mix
SBrkr
Electrical



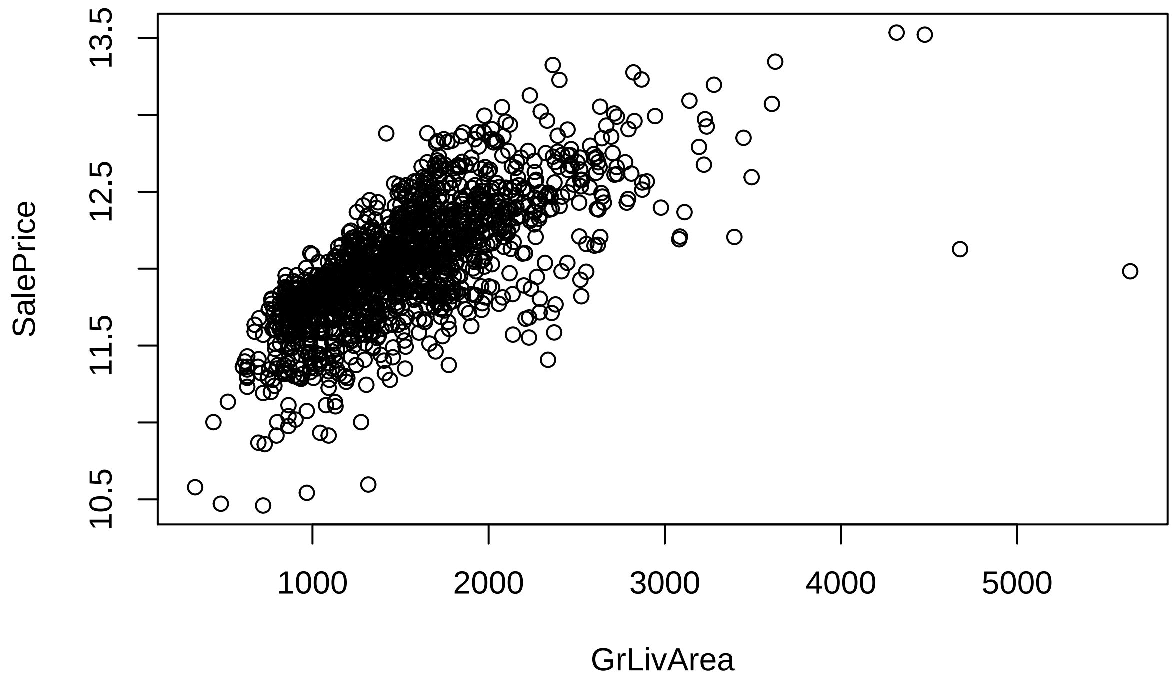
SalePrice
10.5
11.5
12.5
13.5
1000
2000
3000
4000
X1stFlrSF



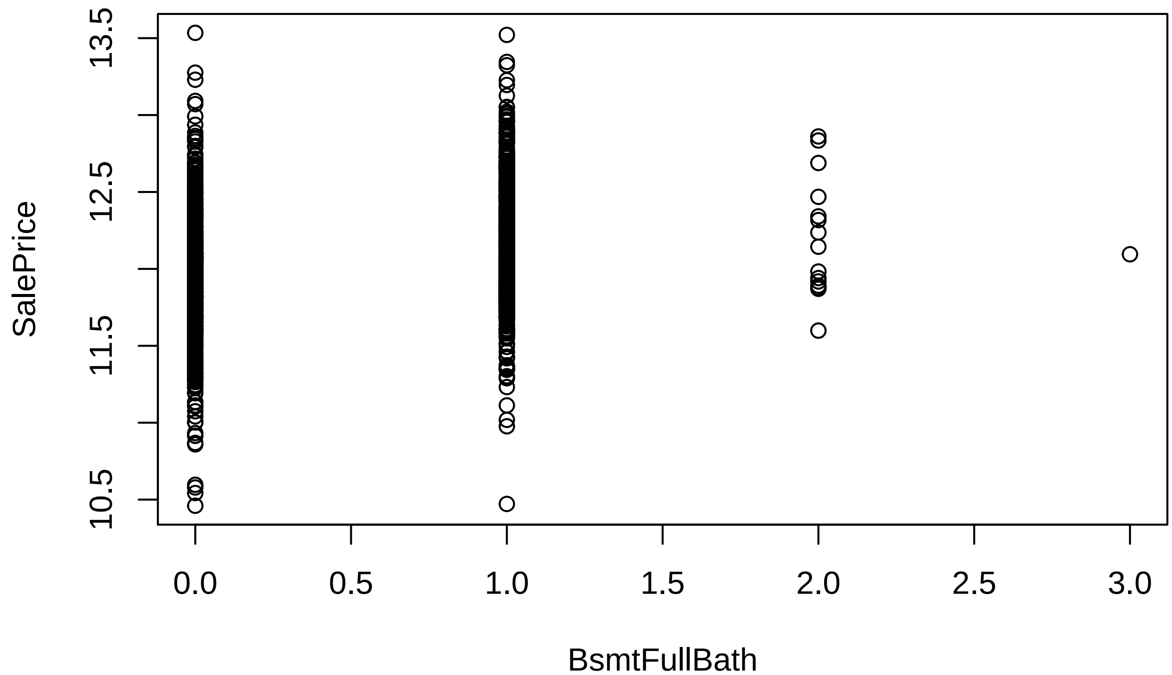
SalePrice
10.5
11.5
12.5
13.5
0
500
1000
1500
2000
X2ndFlrSF

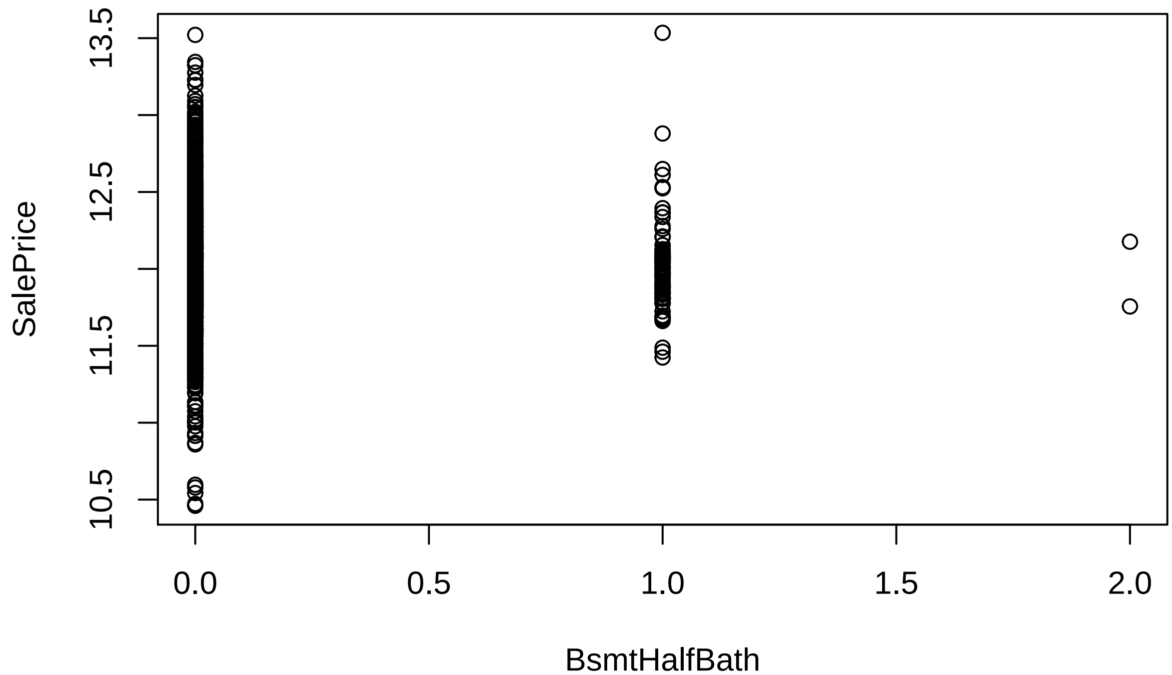


13.5
12.5
11.5
10.5
SalePrice
0
100
200
300
400
500
LowQualFinSF

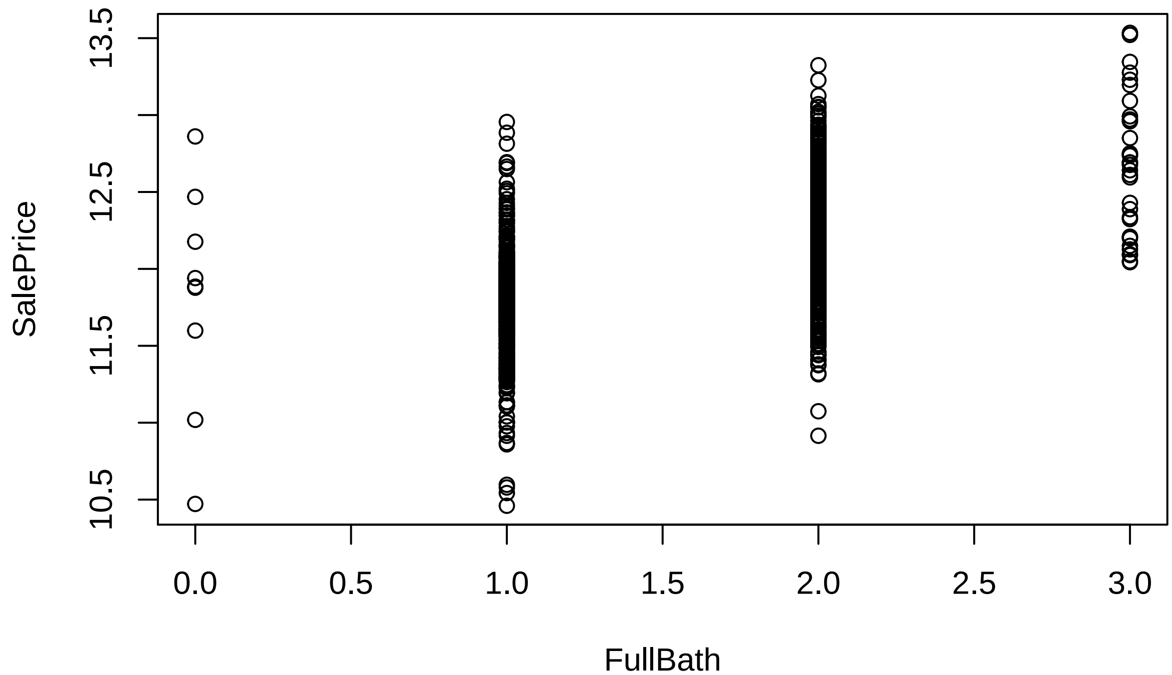


SalePrice
13.5
12.5
11.5
10.5
1000
2000
3000
4000
5000
GrLivArea

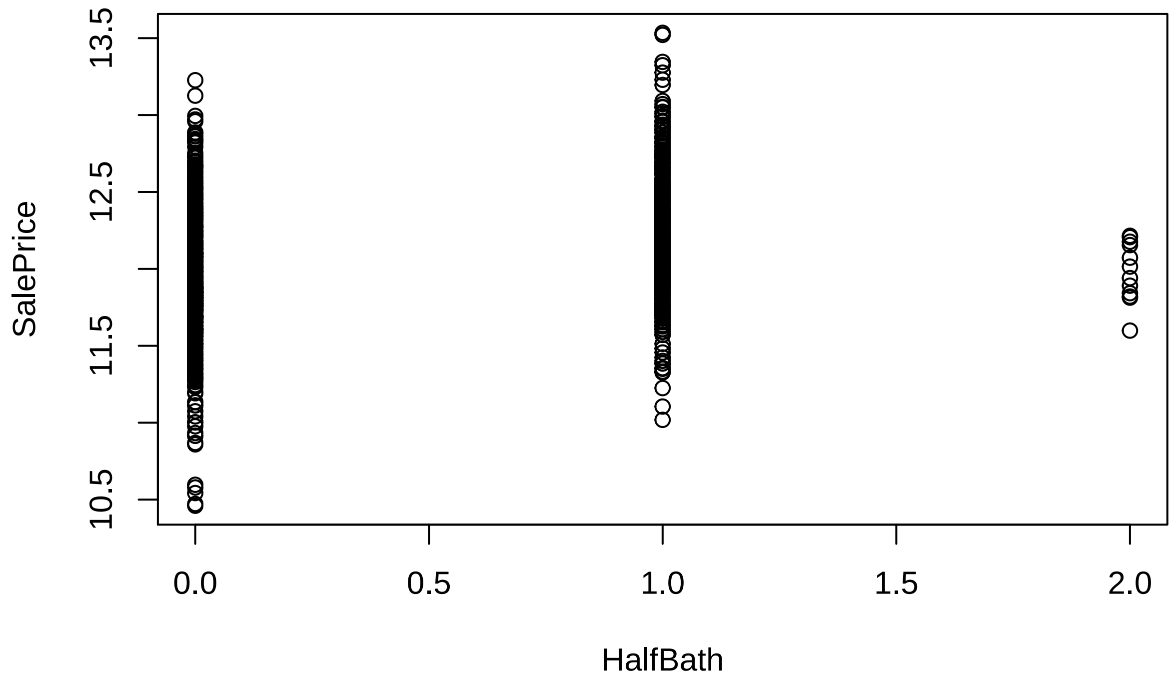

SalePrice
10.5
11.5
12.5
13.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
BsmtFullBath



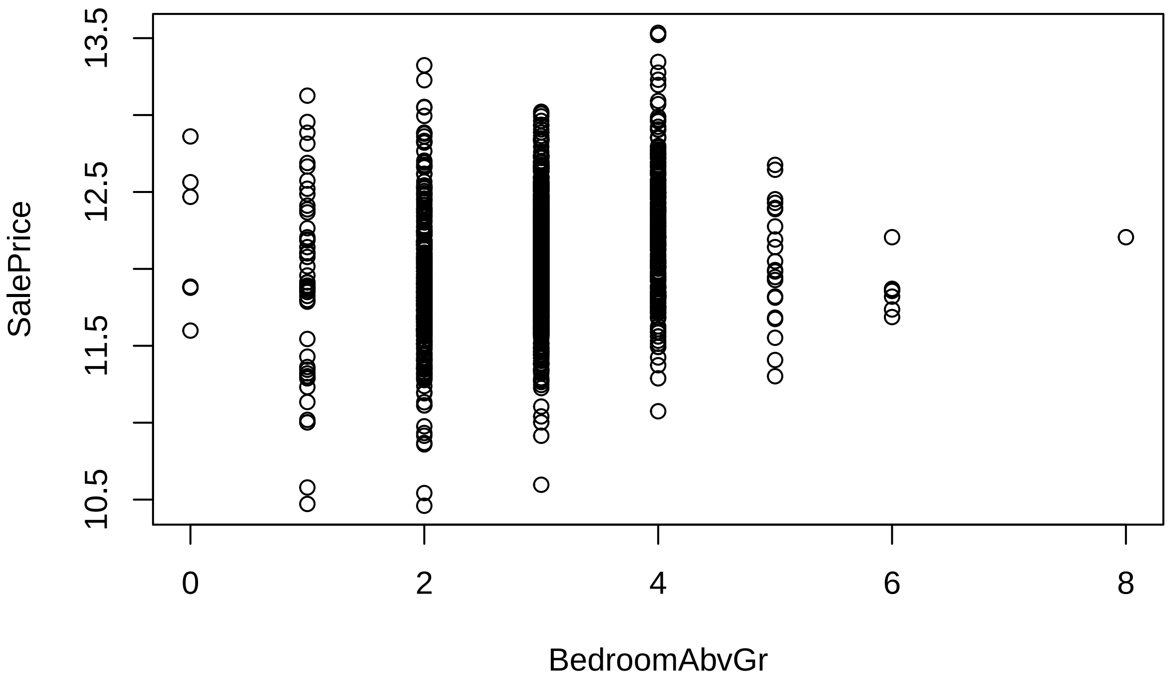
SalePrice
13.5
12.5
11.5
10.5
0.0
0.5
1.0
1.5
2.0
BsmtHalfBath



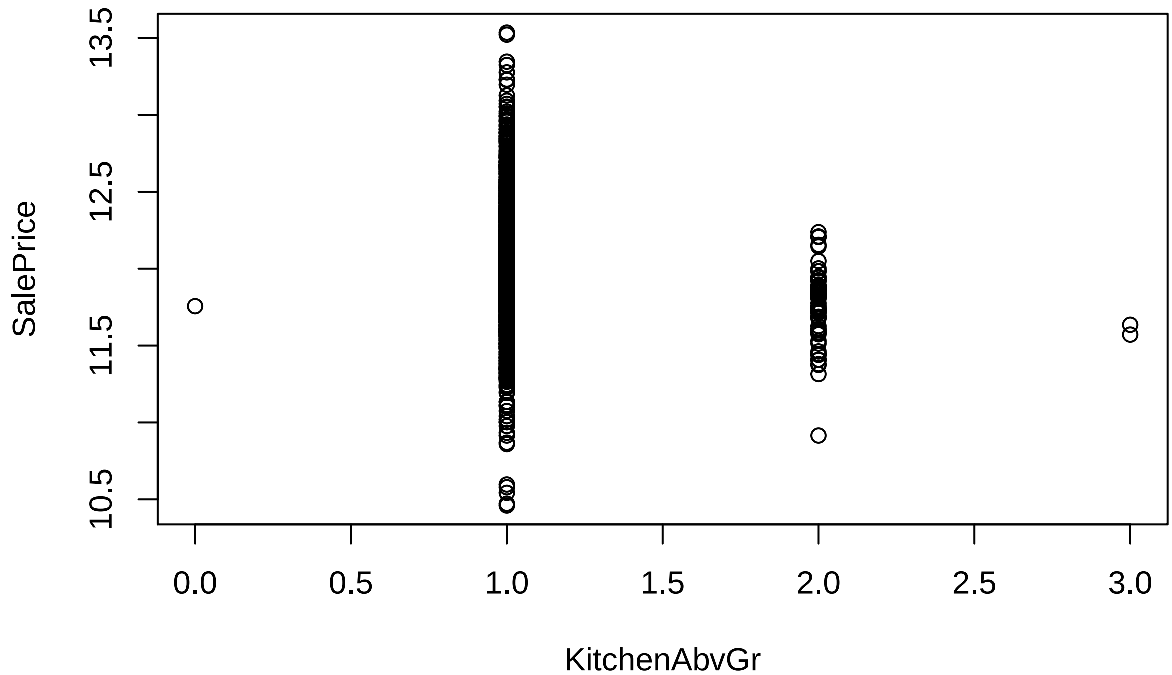
SalePrice
10.5
11.5
12.5
13.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
FullBath

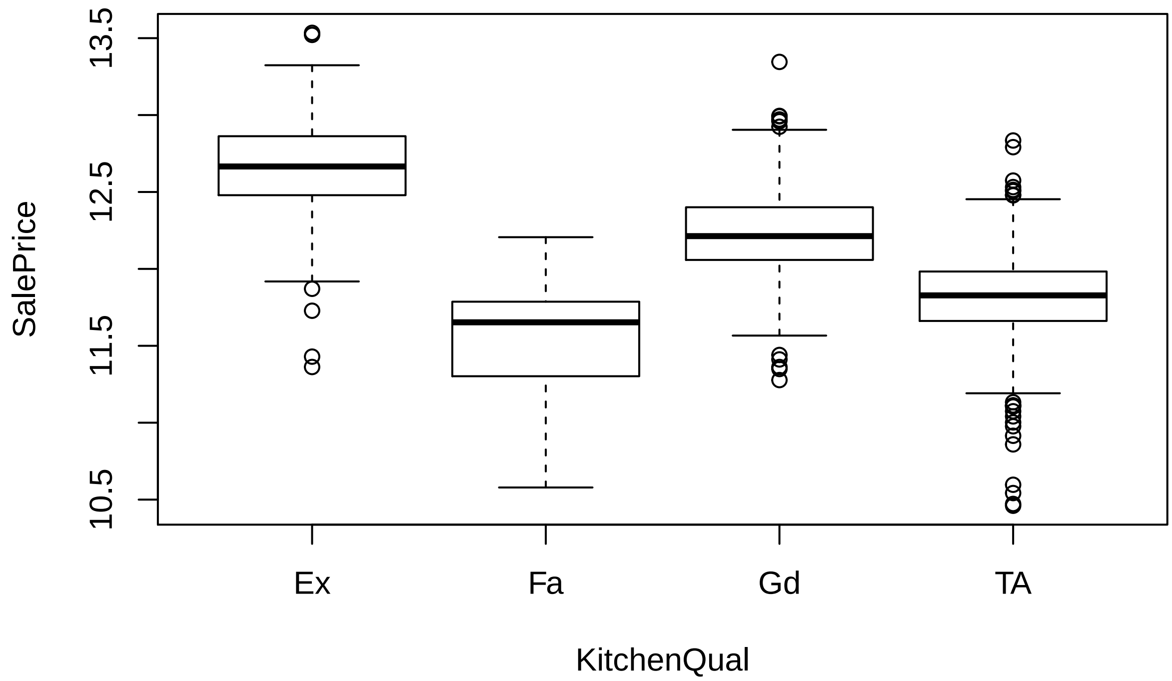


SalePrice
13.5
12.5
11.5
10.5
0.0
0.5
1.0
1.5
2.0
HalfBath

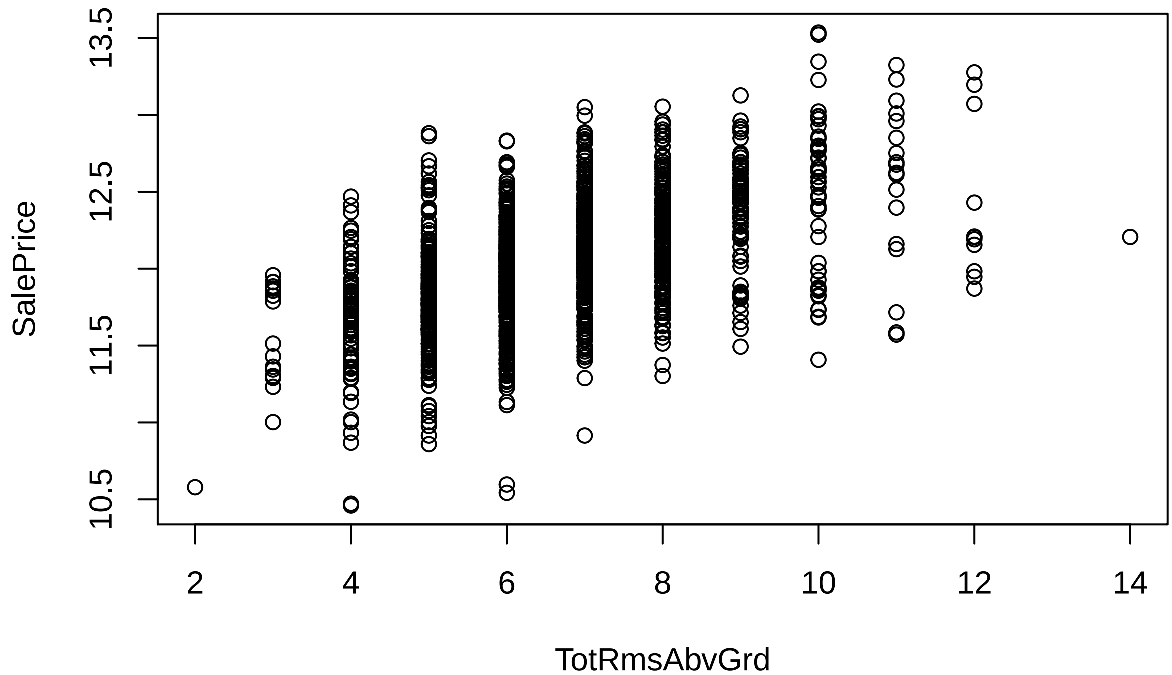


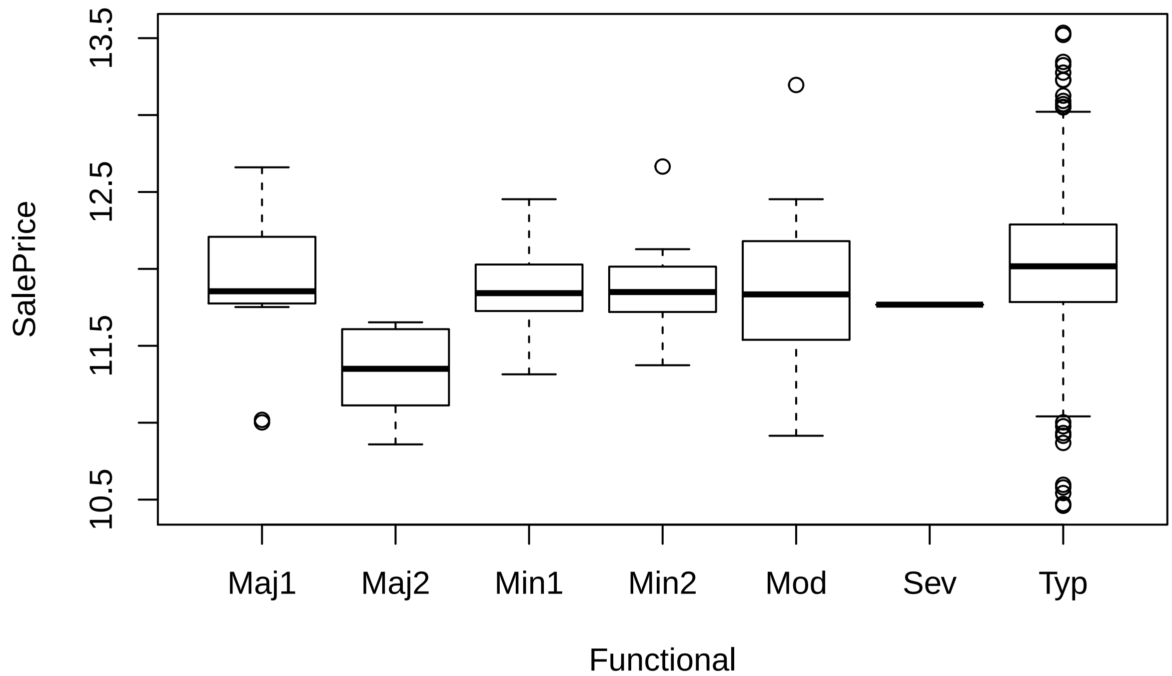
SalePrice
13.5
12.5
11.5
10.5
0
2
4
6
8
BedroomAbvGr

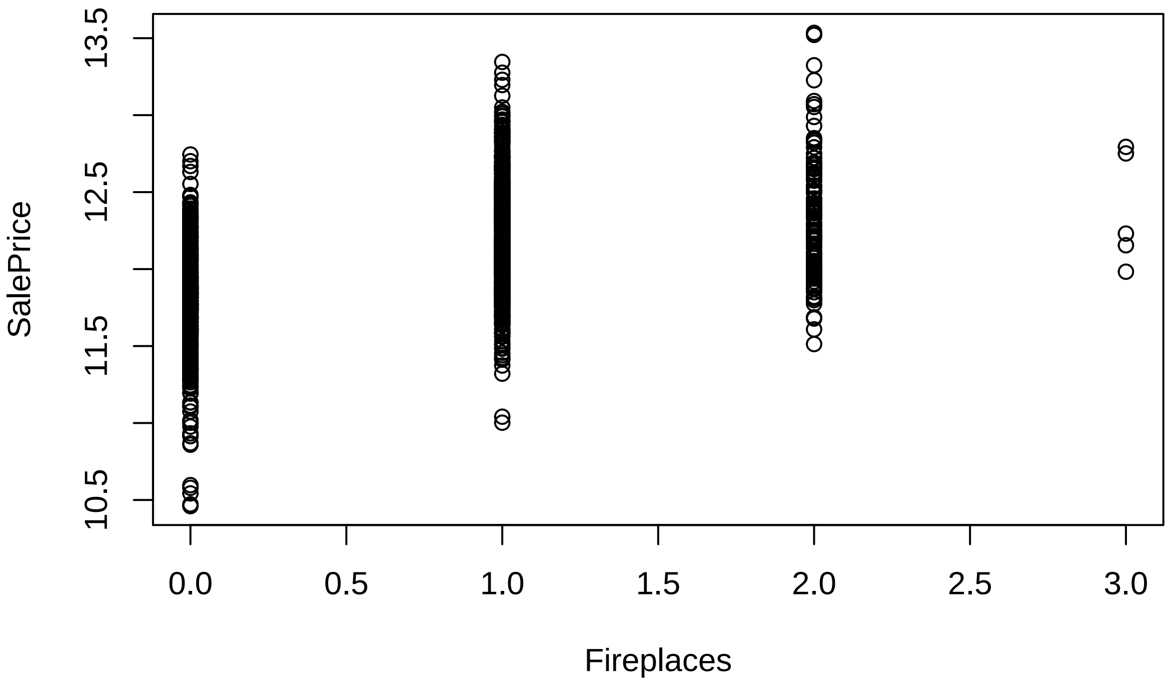

SalePrice
13.5
12.5
11.5
10.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
KitchenAbvGr

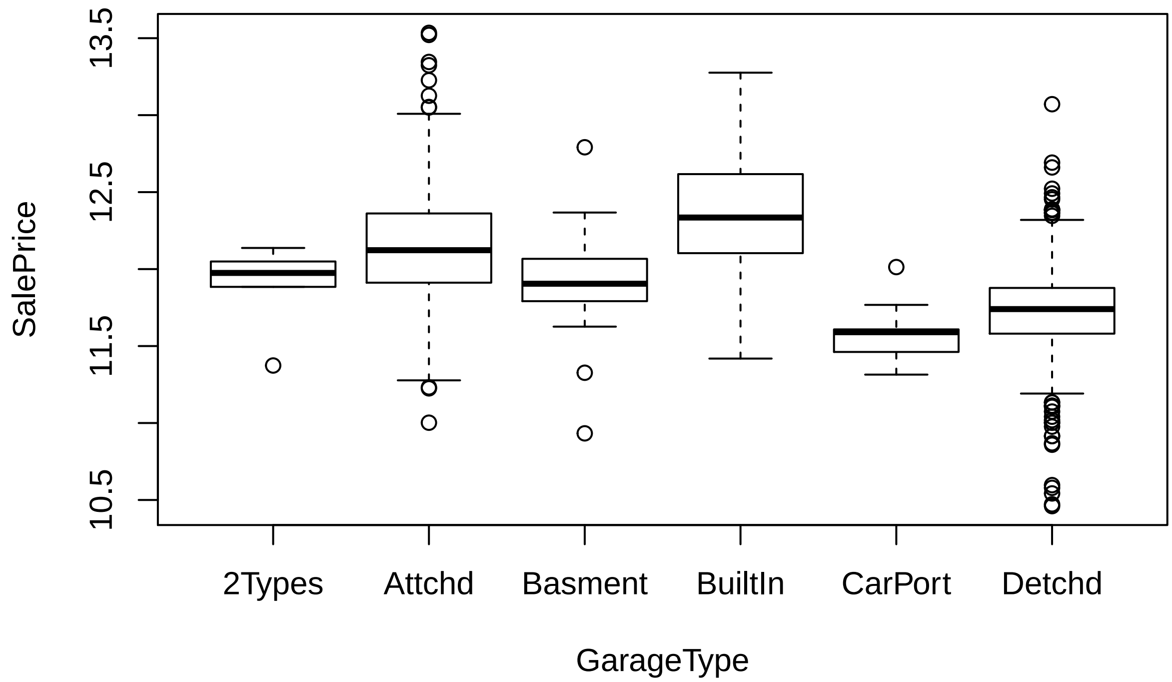


SalePrice
10.5
11.5
12.5
13.5
Ex
Fa
Gd
TA
KitchenQual

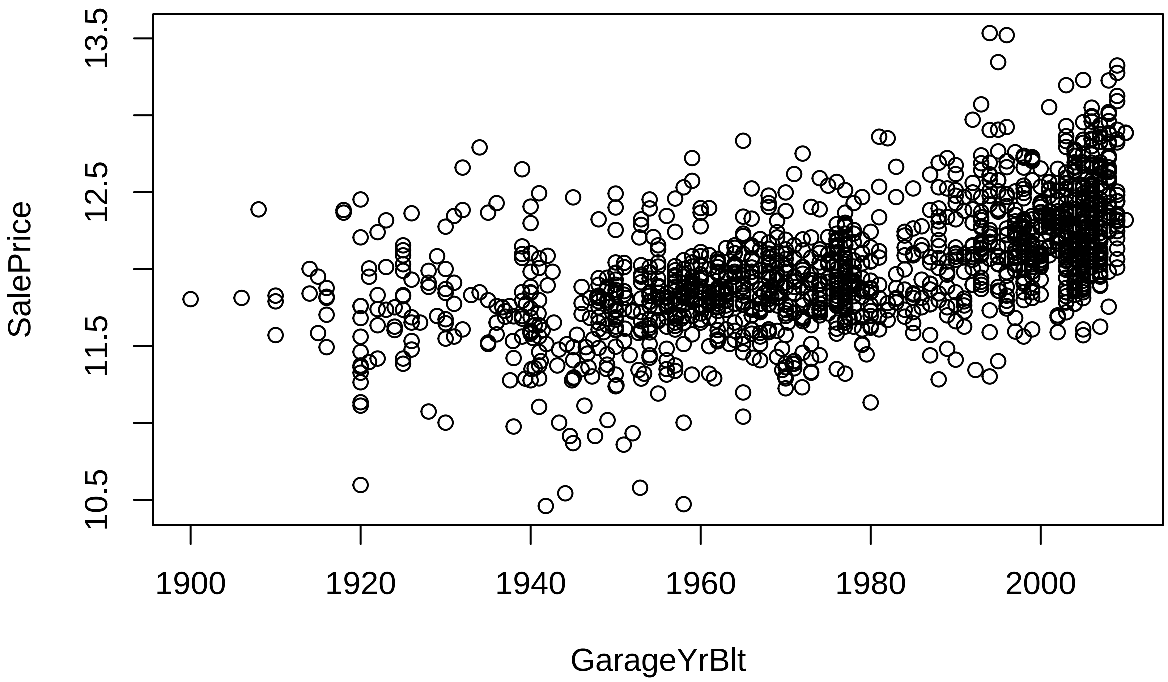

SalePrice
10.5
11.5
12.5
13.5
2
4
6
8
10
12
14
TotRmsAbvGrd

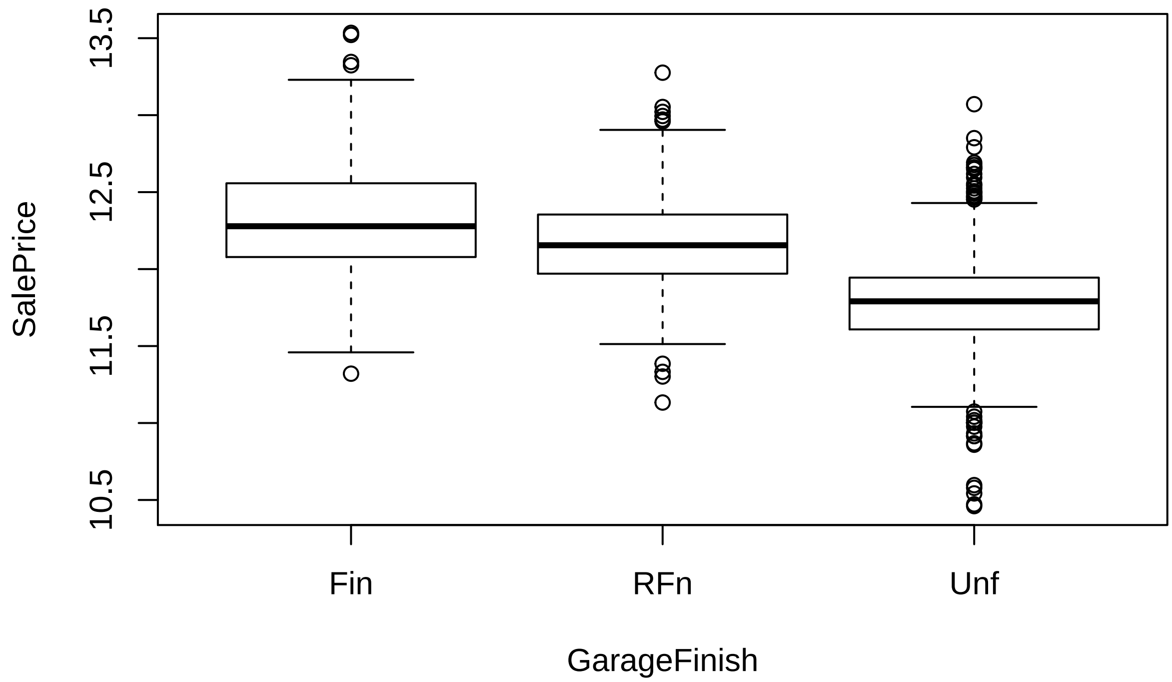

13.5
12.5
11.5
10.5
SalePrice
Maj1
Maj2
Min1
Min2
Mod
Sev
Typ
Functional


13.5
12.5
11.5
10.5
SalePrice
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Fireplaces

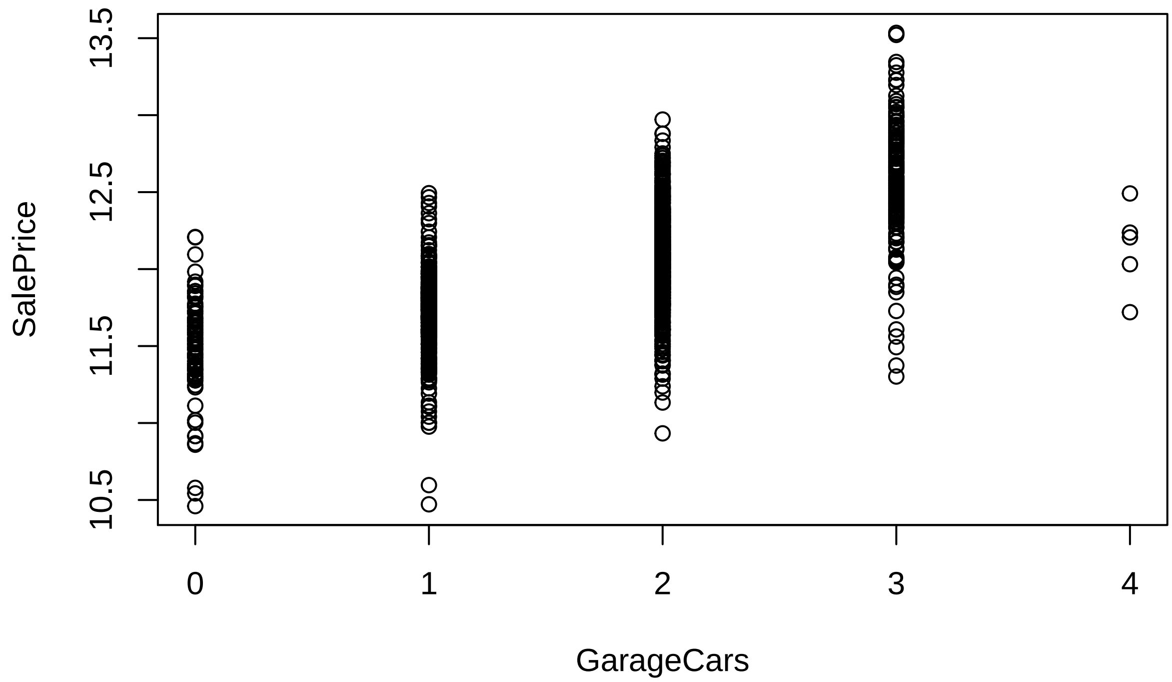


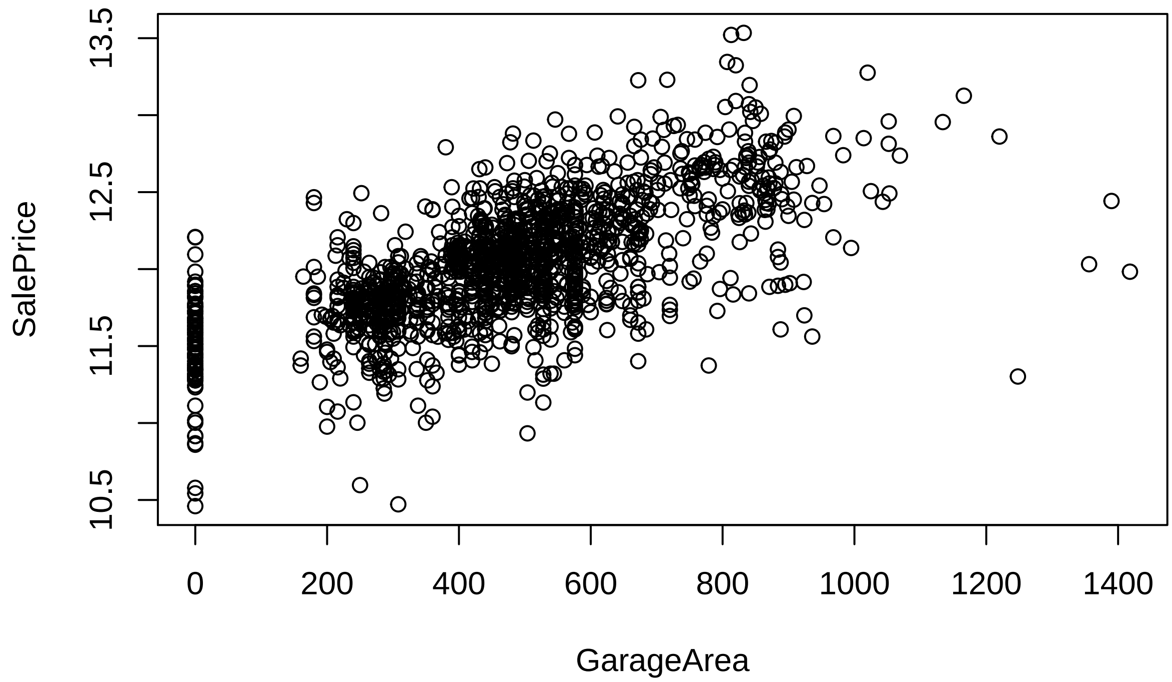
SalePrice
10.5
11.5
12.5
13.5
2Types
Attchd
Basment
BuiltIn
CarPort
Detchd
GarageType

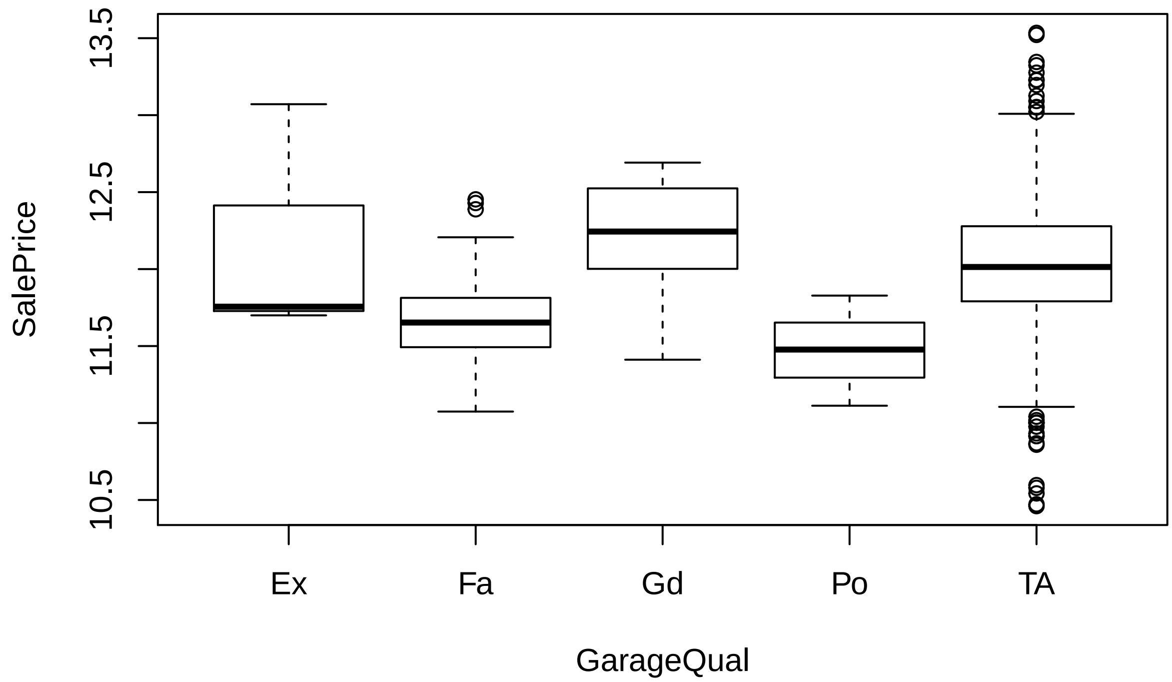

SalePrice
10.5
11.5
12.5
13.5
1900
1920
1940
1960
1980
2000
GarageYrBlt



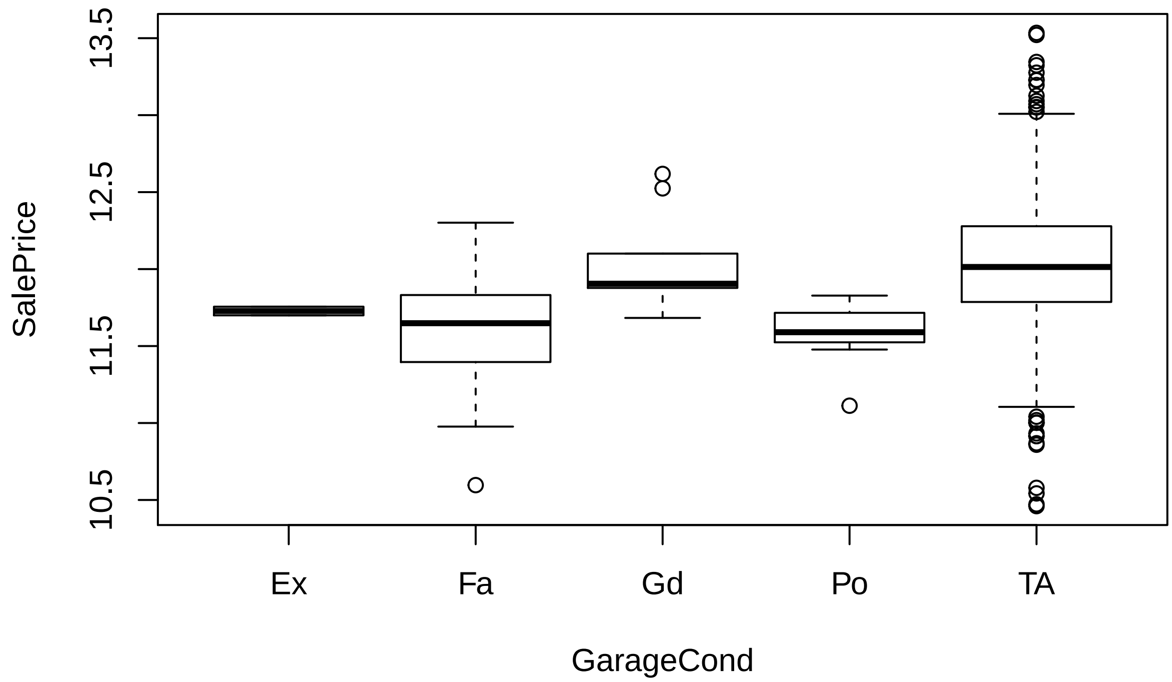
SalePrice
10.5
11.5
12.5
13.5
Fin
RFn
Unf
GarageFinish

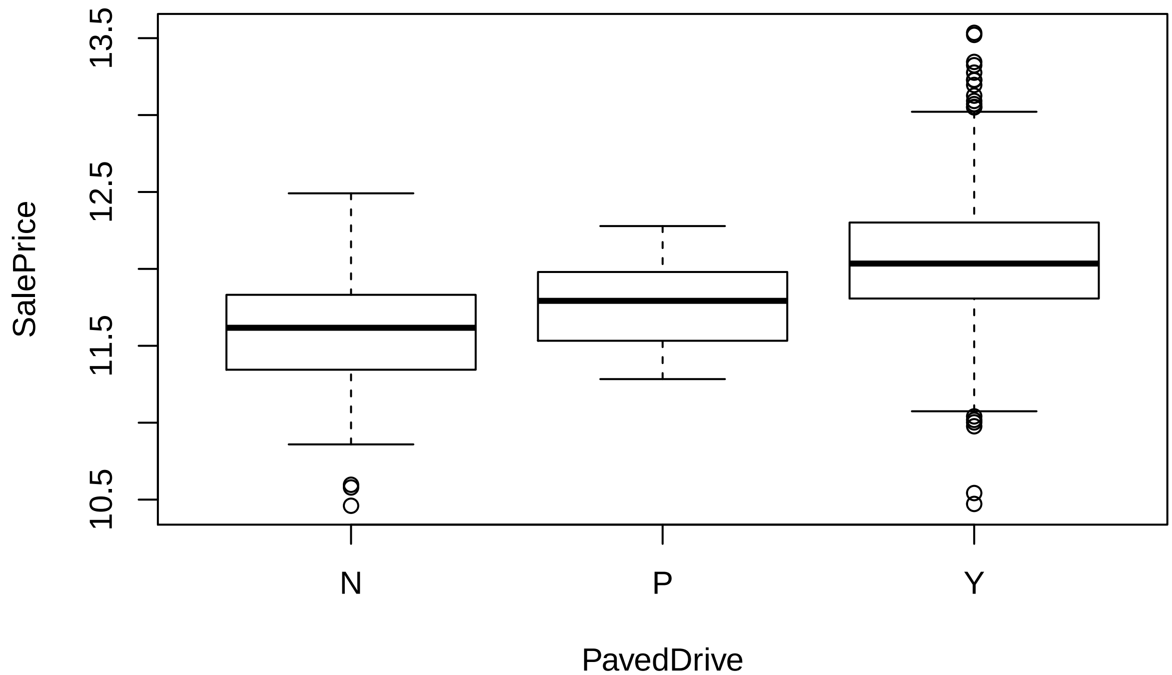

13.5
12.5
11.5
10.5
SalePrice
0
1
2
3
4
GarageCars

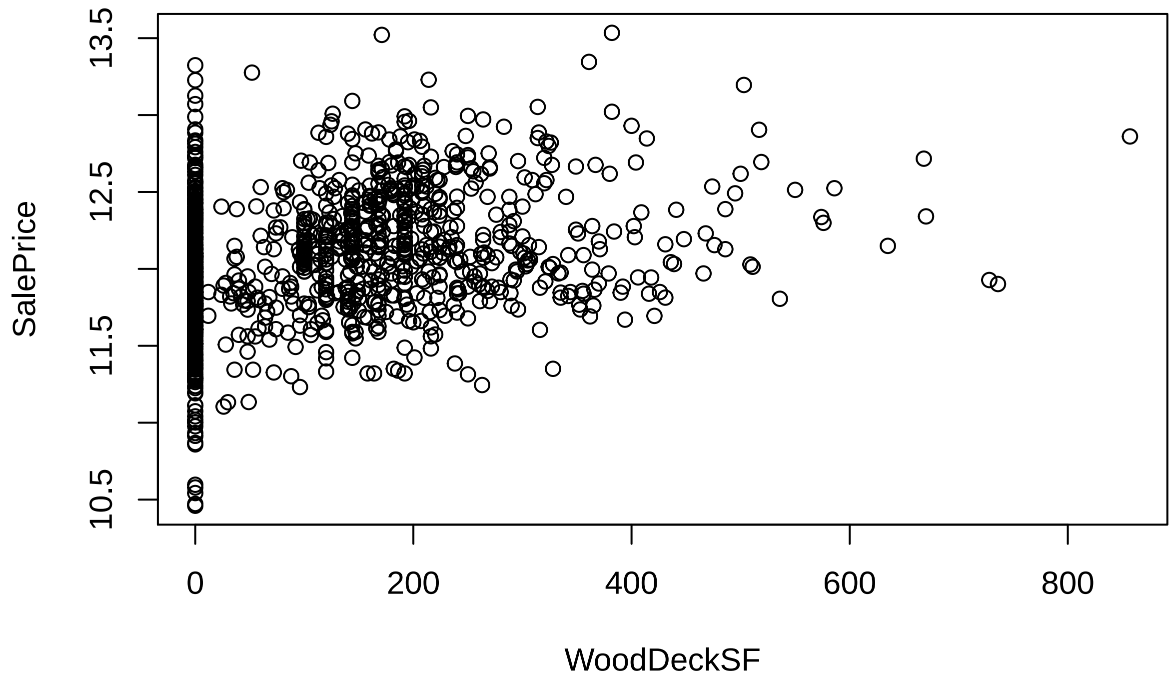

SalePrice
10.5
11.5
12.5
13.5
0
200
400
600
800
1000
1200
1400
GarageArea

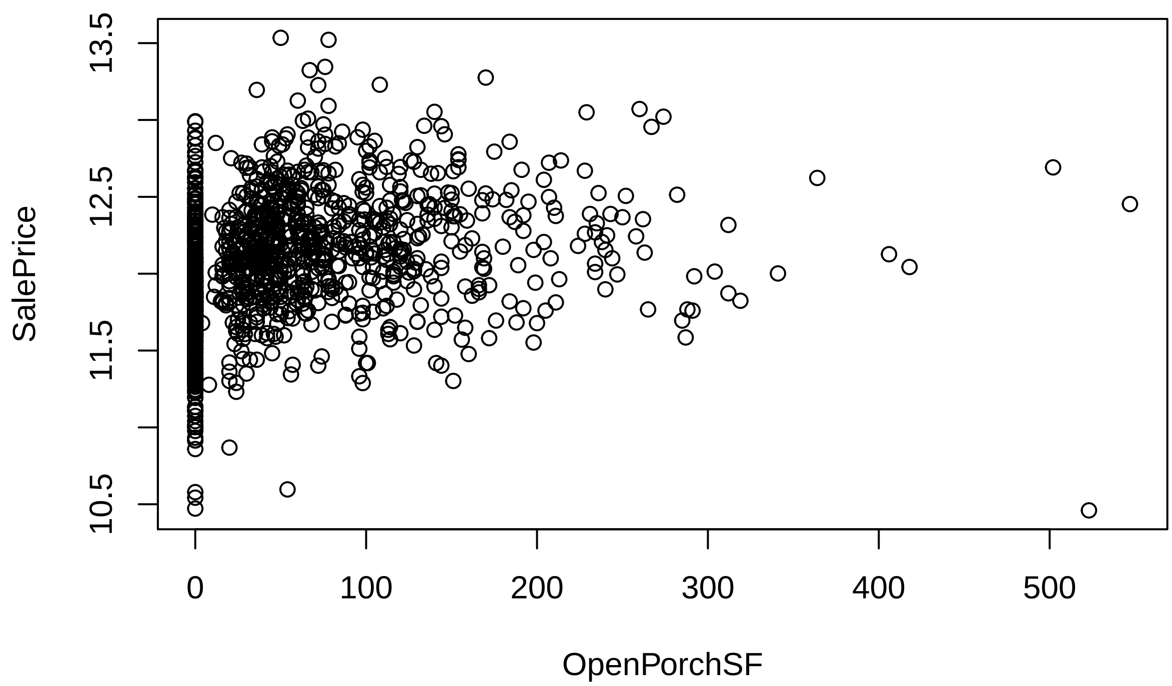


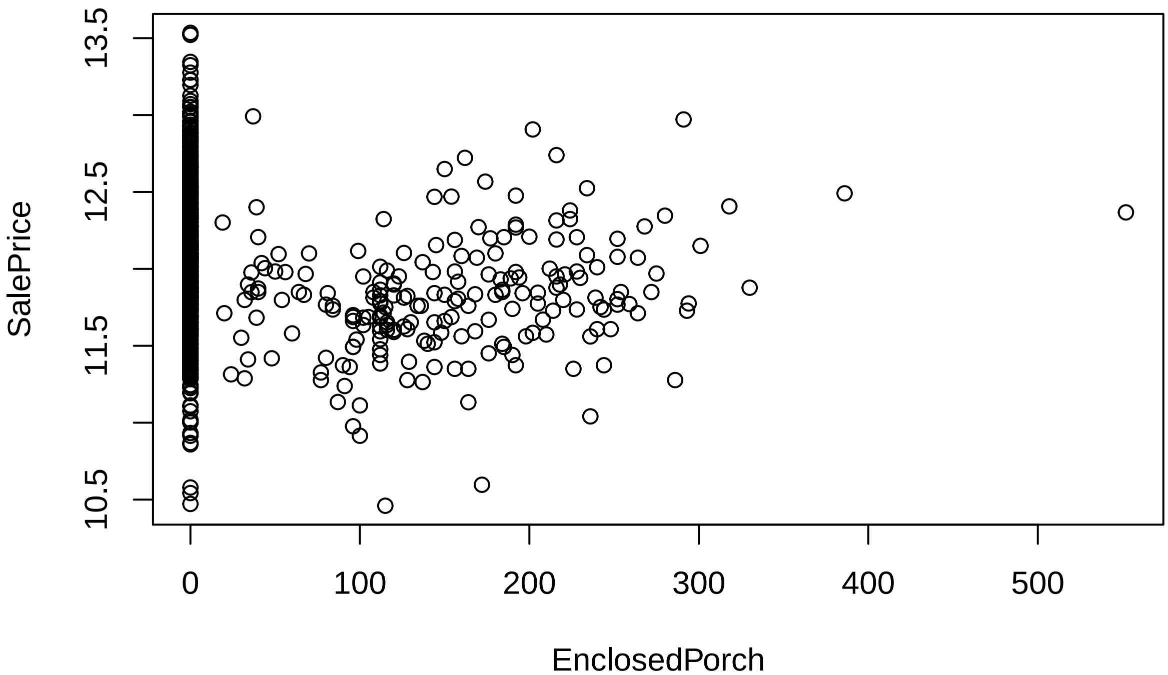
13.5
12.5
11.5
10.5
SalePrice
Ex
Fa
Gd
Po
TA
GarageQual

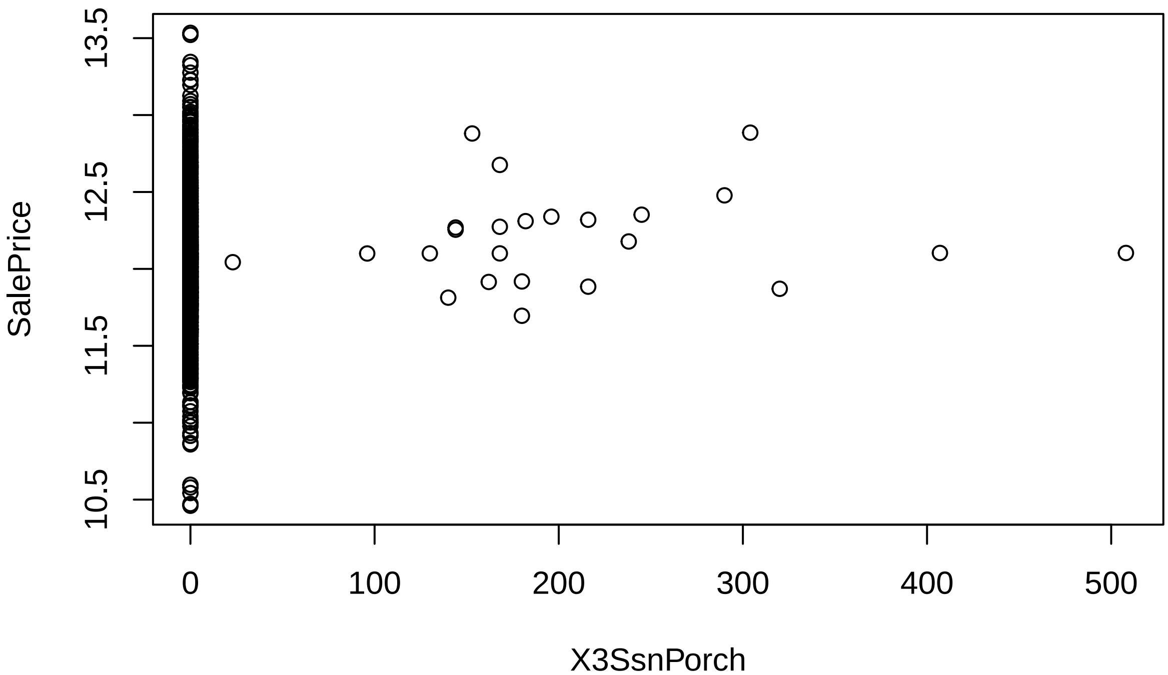

13.5
12.5
11.5
10.5
SalePrice
Ex
Fa
Gd
Po
TA
GarageCond

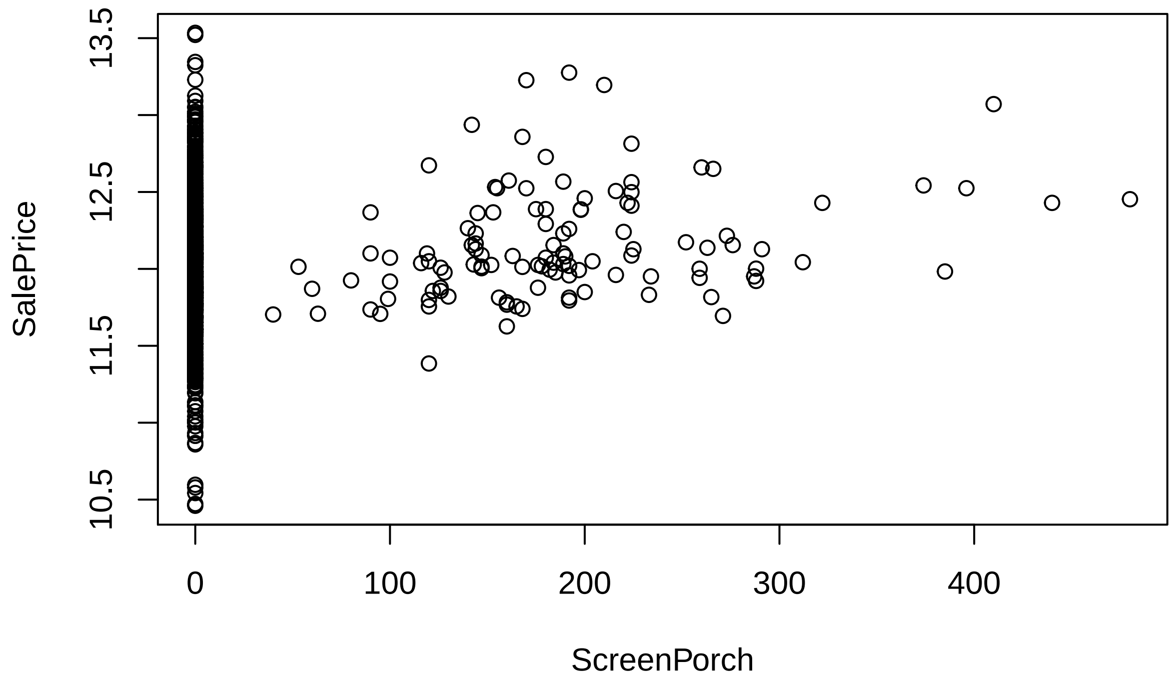

SalePrice
13.5
12.5
11.5
10.5
N
P
Y
PavedDrive


SalePrice
13.5
12.5
11.5
10.5
0
200
400
600
800
WoodDeckSF

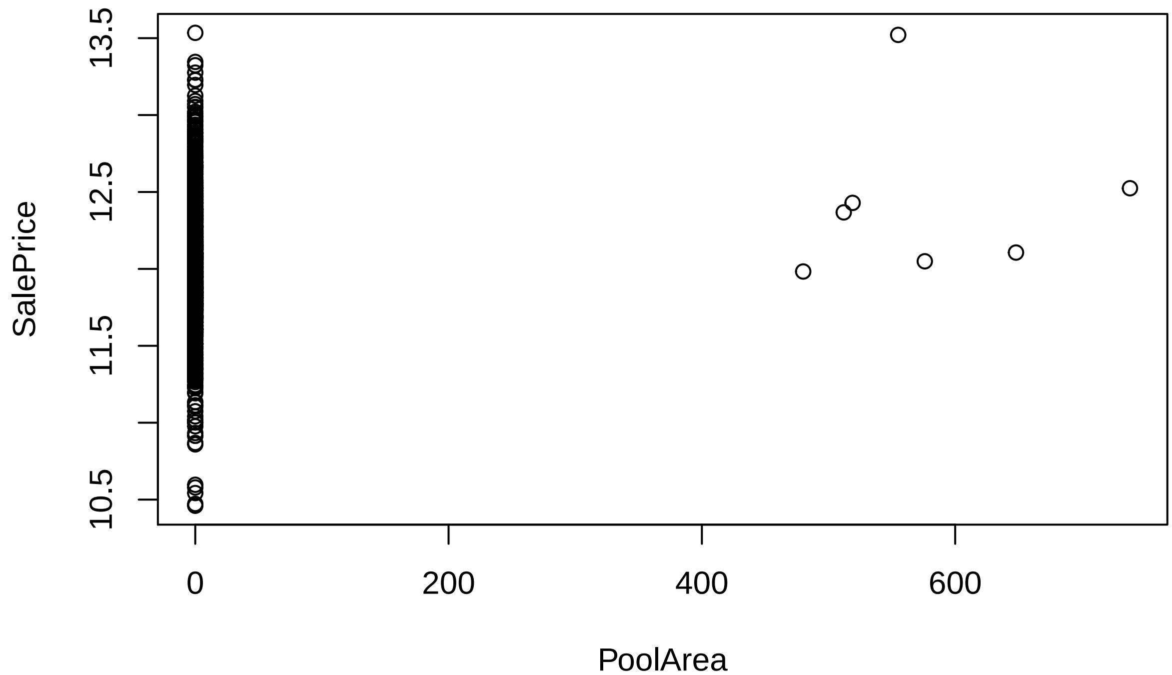

SalePrice
13.5
12.5
11.5
10.5
0
100
200
300
400
500
OpenPorchSF


SalePrice
10.5
11.5
12.5
13.5
0
100
200
300
400
500
EnclosedPorch

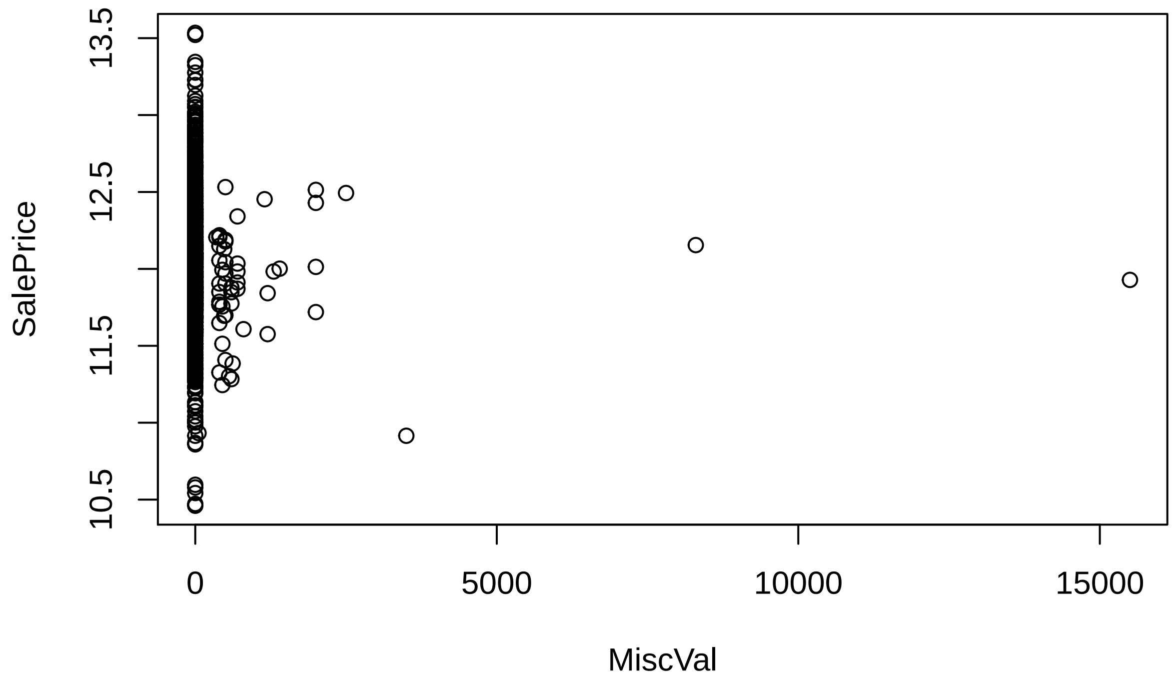

SalePrice
10.5
11.5
12.5
13.5
0
100
200
300
400
500
X3SsnPorch

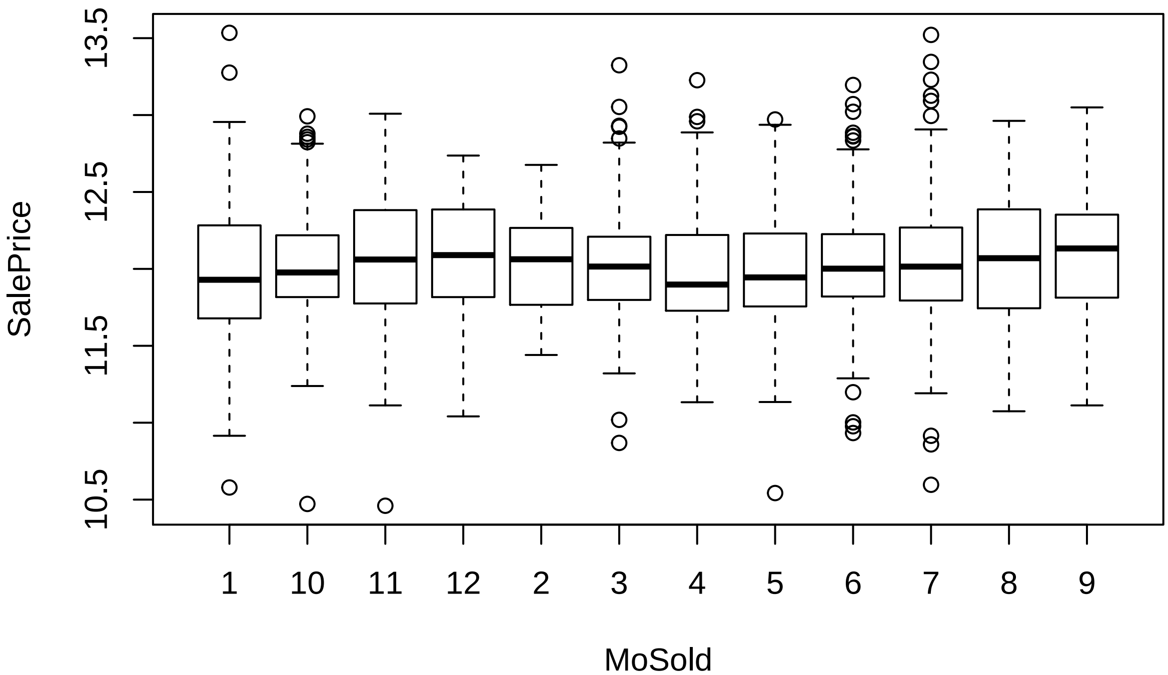


SalePrice
ScreenPorch
13.5
12.5
11.5
10.5
0
100
200
300
400

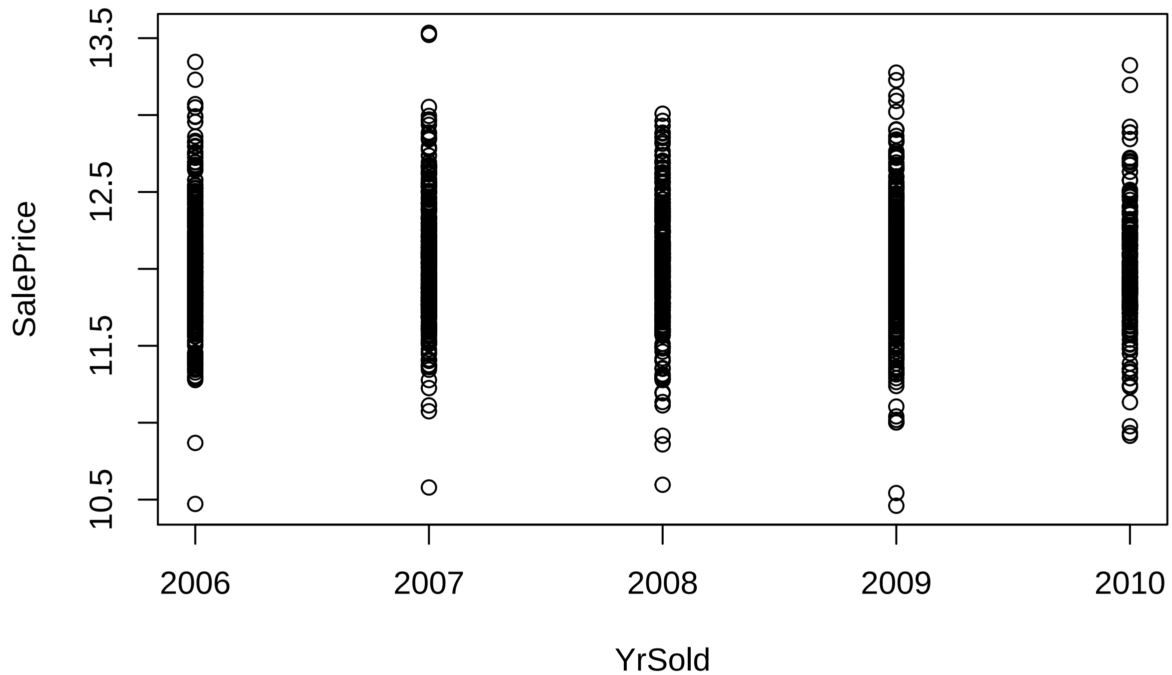


13.5
12.5
11.5
10.5
SalePrice
0
200
400
600
PoolArea

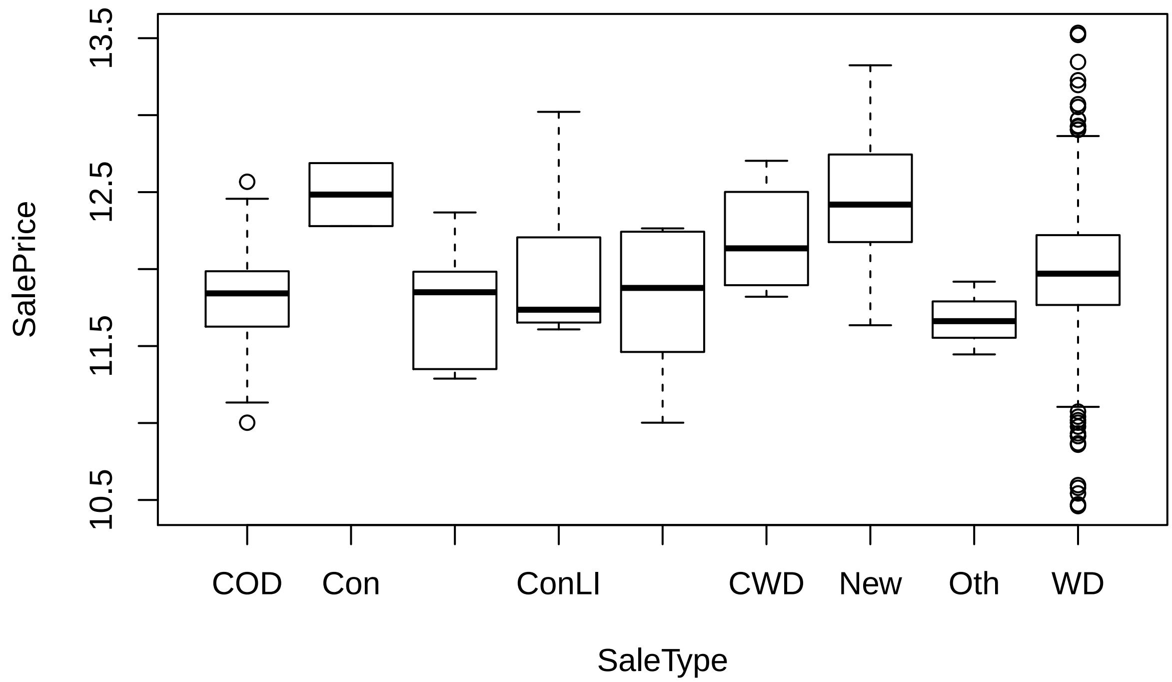

13.5
12.5
11.5
10.5
SalePrice
0
5000
10000
15000
MiscVal

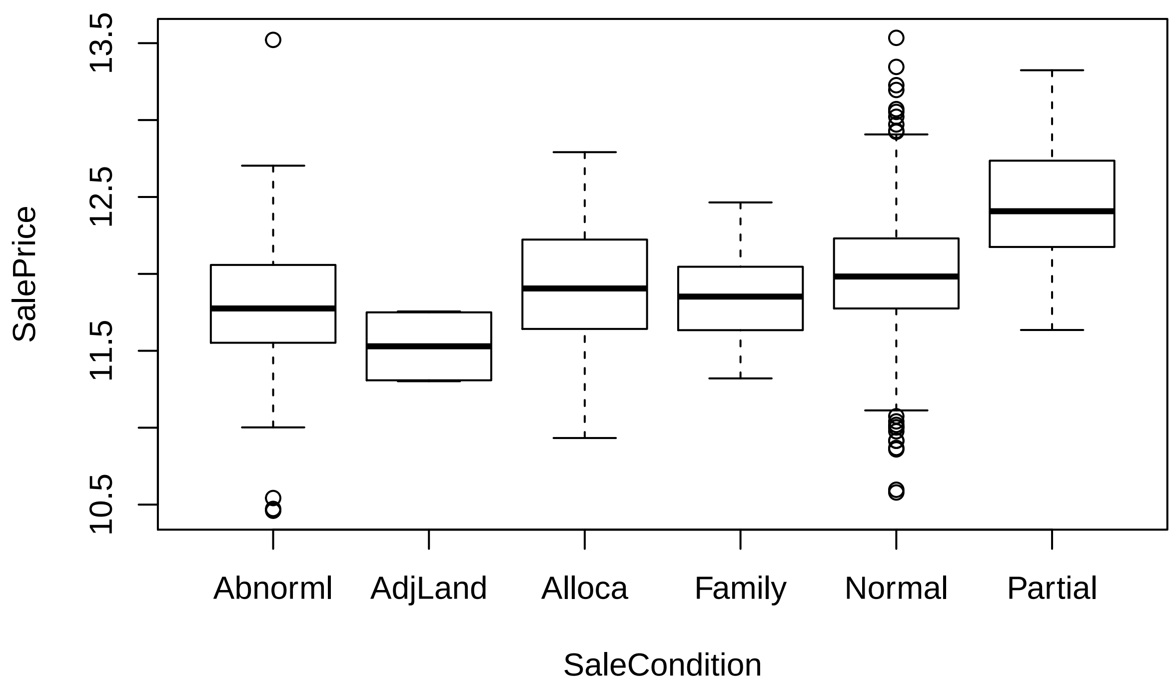


SalePrice
10.5
11.5
12.5
13.5
1
10
11
12
2
3
4
5
6
7
8
9
MoSold



13.5
12.5
11.5
10.5
SalePrice
2006
2007
2008
2009
2010
YrSold


13.5
12.5
11.5
10.5
SalePrice
COD
Con
ConLI
CWD
New
Oth
WD
SaleType




Visually observing the plots, the following variables seems like they might be correlated: OverallQual, TotalBsmtSF, X1stFlrSF, GrLivArea, TotRmsAbvGrd.

### Section 3 - Creating Predictive Models

```
library(caret)
```

```
## Warning: package 'caret' was built under R version 3.6.3
```

```
## Loading required package: lattice
```

```
## Loading required package: ggplot2
```

```
## Warning: package 'ggplot2' was built under R version 3.6.3
```

```
##
## Attaching package: 'ggplot2'
```

```
## The following object is masked from 'package:randomForest':
##
##     margin
```

```r
# Train test split
train.index <- createDataPartition(dataHouse$SalePrice, p = 0.8, list = FALSE)
data_train <- dataHouse[train.index, ]
data_test <- dataHouse[-train.index, ]
```

*Lasso Linear Regression Model*

```r
set.seed(1)

lasso <- train(
  SalePrice ~., data = data_train, method = "glmnet",
  trControl = trainControl("cv", number = 10),
  tuneGrid = expand.grid(alpha = 1, lambda = 10^seq(-3, 3, length = 100)))
```

```
## Warning in nominalTrainWorkflow(x = x, y = y, wts = weights, info = trainInfo, :
## There were missing values in resampled performance measures.
```

Best tuned coefficients:

```r
coeffs <- as.matrix(coef(lasso$finalModel, lasso$bestTune$lambda))
coeffs
```

```
##                                1
## (Intercept)         7.276627e+00
## MSSubClass160      -9.153581e-02
## MSSubClass180       0.000000e+00
## MSSubClass190       0.000000e+00
## MSSubClass20        2.841034e-02
## MSSubClass30       -4.184269e-02
## MSSubClass40        0.000000e+00
## MSSubClass45        0.000000e+00
## MSSubClass50        0.000000e+00
## MSSubClass60        0.000000e+00
## MSSubClass70        0.000000e+00
## MSSubClass75        0.000000e+00
## MSSubClass80        0.000000e+00
## MSSubClass85        0.000000e+00
## MSSubClass90        0.000000e+00
## MSZoningFV          3.904708e-02
## MSZoningRH          0.000000e+00
## MSZoningRL          4.004596e-02
## MSZoningRM          0.000000e+00
## LotFrontage         0.000000e+00
## LotArea             1.352954e-06
## StreetPave          5.237496e-02
## LotShapeIR2         0.000000e+00
## LotShapeIR3        -9.447559e-02
## LotShapeReg         0.000000e+00
## LandContourHLS      0.000000e+00
## LandContourLow      0.000000e+00
## LandContourLvl      0.000000e+00
## UtilitiesNoSeWa    -3.322919e-02
```

```
## LotConfigCulDSac       3.811951e-02
## LotConfigFR2           0.000000e+00
## LotConfigFR3           0.000000e+00
## LotConfigInside        0.000000e+00
## LandSlopeMod           0.000000e+00
## LandSlopeSev           0.000000e+00
## NeighborhoodBlueste    0.000000e+00
## NeighborhoodBrDale     0.000000e+00
## NeighborhoodBrkSide    4.369358e-04
## NeighborhoodClearCr    2.440725e-02
## NeighborhoodCollgCr    0.000000e+00
## NeighborhoodCrawfor    1.176998e-01
## NeighborhoodEdwards   -3.605538e-02
## NeighborhoodGilbert    0.000000e+00
## NeighborhoodIDOTRR    -8.028385e-02
## NeighborhoodMeadowV   -6.511095e-02
## NeighborhoodMitchel    0.000000e+00
## NeighborhoodNAmes      0.000000e+00
## NeighborhoodNoRidge    5.332634e-02
## NeighborhoodNPkVill    0.000000e+00
## NeighborhoodNridgHt    1.140510e-01
## NeighborhoodNWAmes     0.000000e+00
## NeighborhoodOldTown   -3.192463e-02
## NeighborhoodSawyer     0.000000e+00
## NeighborhoodSawyerW    0.000000e+00
## NeighborhoodSomerst    4.725641e-02
## NeighborhoodStoneBr    1.330543e-01
## NeighborhoodSWISU      0.000000e+00
## NeighborhoodTimber     0.000000e+00
## NeighborhoodVeenker    3.015219e-02
## Condition1Feedr       -1.228277e-02
## Condition1Norm         2.783237e-02
## Condition1PosA         0.000000e+00
## Condition1PosN         0.000000e+00
## Condition1RRAe         0.000000e+00
## Condition1RRAn         0.000000e+00
## Condition1RRNe         0.000000e+00
## Condition1RRNn         0.000000e+00
## Condition2Feedr        0.000000e+00
## Condition2Norm         0.000000e+00
## Condition2PosA         0.000000e+00
## Condition2PosN        -4.758618e-01
## Condition2RRAe         0.000000e+00
## Condition2RRAn         0.000000e+00
## Condition2RRNn         0.000000e+00
## BldgType2fmCon         0.000000e+00
## BldgTypeDuplex         0.000000e+00
## BldgTypeTwnhs         -2.758086e-02
## BldgTypeTwnhsE        -1.867469e-02
## HouseStyle1.5Unf       0.000000e+00
## HouseStyle1Story       0.000000e+00
## HouseStyle2.5Fin       0.000000e+00
## HouseStyle2.5Unf       0.000000e+00
## HouseStyle2Story       0.000000e+00
```

```
## HouseStyleSFoyer       0.000000e+00
## HouseStyleSLvl         0.000000e+00
## OverallQual            7.049432e-02
## OverallCond            2.831610e-02
## YearBuilt              7.839586e-04
## YearRemodAdd           9.272254e-04
## RoofStyleGable        -2.644506e-03
## RoofStyleGambrel       0.000000e+00
## RoofStyleHip           0.000000e+00
## RoofStyleMansard       0.000000e+00
## RoofStyleShed          0.000000e+00
## RoofMatlCompShg        0.000000e+00
## RoofMatlMembran        0.000000e+00
## RoofMatlMetal          0.000000e+00
## RoofMatlRoll           0.000000e+00
## RoofMatlTar&Grv        0.000000e+00
## RoofMatlWdShake        0.000000e+00
## RoofMatlWdShngl        5.974352e-02
## Exterior1stAsphShn     0.000000e+00
## Exterior1stBrkComm     0.000000e+00
## Exterior1stBrkFace     3.886782e-02
## Exterior1stCBlock      0.000000e+00
## Exterior1stCemntBd     0.000000e+00
## Exterior1stHdBoard    -5.692639e-03
## Exterior1stImStucc     0.000000e+00
## Exterior1stMetalSd     0.000000e+00
## Exterior1stPlywood     0.000000e+00
## Exterior1stStone       0.000000e+00
## Exterior1stStucco      0.000000e+00
## Exterior1stVinylSd     0.000000e+00
## Exterior1stWd Sdng    -1.033213e-02
## Exterior1stWdShing     0.000000e+00
## Exterior2ndAsphShn     0.000000e+00
## Exterior2ndBrk Cmn     0.000000e+00
## Exterior2ndBrkFace     0.000000e+00
## Exterior2ndCBlock      0.000000e+00
## Exterior2ndCmentBd     0.000000e+00
## Exterior2ndHdBoard     0.000000e+00
## Exterior2ndImStucc     0.000000e+00
## Exterior2ndMetalSd     0.000000e+00
## Exterior2ndOther       0.000000e+00
## Exterior2ndPlywood     0.000000e+00
## Exterior2ndStone       0.000000e+00
## Exterior2ndStucco     -4.102236e-02
## Exterior2ndVinylSd     0.000000e+00
## Exterior2ndWd Sdng     0.000000e+00
## Exterior2ndWd Shng     0.000000e+00
## MasVnrTypeBrkFace      0.000000e+00
## MasVnrTypeNone         0.000000e+00
## MasVnrTypeStone        3.420891e-03
## MasVnrArea             1.505733e-05
## ExterQualFa           -6.924633e-03
## ExterQualGd            0.000000e+00
## ExterQualTA           -6.947233e-03
```

```
## ExterCondFa          -2.405911e-02
## ExterCondGd           0.000000e+00
## ExterCondPo           0.000000e+00
## ExterCondTA           0.000000e+00
## FoundationCBlock      0.000000e+00
## FoundationPConc       1.422058e-02
## FoundationSlab       -3.276404e-02
## FoundationStone       0.000000e+00
## FoundationWood        0.000000e+00
## BsmtQualFa            0.000000e+00
## BsmtQualGd            0.000000e+00
## BsmtQualTA           -1.399450e-02
## BsmtCondGd            0.000000e+00
## BsmtCondPo            0.000000e+00
## BsmtCondTA            0.000000e+00
## BsmtExposureGd        4.772708e-02
## BsmtExposureMn        0.000000e+00
## BsmtExposureNo       -4.829930e-03
## BsmtFinType1BLQ       0.000000e+00
## BsmtFinType1GLQ       1.047856e-02
## BsmtFinType1LwQ       0.000000e+00
## BsmtFinType1Rec       0.000000e+00
## BsmtFinType1Unf      -2.868208e-02
## BsmtFinSF1            0.000000e+00
## BsmtFinType2BLQ      -2.790412e-02
## BsmtFinType2GLQ       0.000000e+00
## BsmtFinType2LwQ       0.000000e+00
## BsmtFinType2Rec       0.000000e+00
## BsmtFinType2Unf       0.000000e+00
## BsmtFinSF2            0.000000e+00
## BsmtUnfSF             0.000000e+00
## TotalBsmtSF           2.773587e-05
## HeatingGasA           0.000000e+00
## HeatingGasW           3.210388e-02
## HeatingGrav          -9.300081e-02
## HeatingOthW           0.000000e+00
## HeatingWall           0.000000e+00
## HeatingQCFa           0.000000e+00
## HeatingQCGd           0.000000e+00
## HeatingQCPo           0.000000e+00
## HeatingQCTA          -1.484642e-02
## CentralAirY           7.121859e-02
## ElectricalFuseF       0.000000e+00
## ElectricalFuseP       0.000000e+00
## ElectricalMix         0.000000e+00
## ElectricalSBrkr       0.000000e+00
## X1stFlrSF             3.309808e-07
## X2ndFlrSF             0.000000e+00
## LowQualFinSF          0.000000e+00
## GrLivArea             1.969026e-04
## BsmtFullBath          4.148992e-02
## BsmtHalfBath          0.000000e+00
## FullBath              2.378506e-02
## HalfBath              1.509926e-02
```

```
## BedroomAbvGr          0.000000e+00
## KitchenAbvGr         -3.777305e-02
## KitchenQualFa         0.000000e+00
## KitchenQualGd         0.000000e+00
## KitchenQualTA        -1.885371e-02
## TotRmsAbvGrd          5.294280e-03
## FunctionalMaj2       -1.052601e-01
## FunctionalMin1        0.000000e+00
## FunctionalMin2        0.000000e+00
## FunctionalMod        -6.941289e-03
## FunctionalSev        -1.711373e-01
## FunctionalTyp         2.555149e-02
## Fireplaces            2.758108e-02
## GarageTypeAttchd      0.000000e+00
## GarageTypeBasment     0.000000e+00
## GarageTypeBuiltIn     0.000000e+00
## GarageTypeCarPort     0.000000e+00
## GarageTypeDetchd     -4.766737e-03
## GarageYrBlt           0.000000e+00
## GarageFinishRFn       0.000000e+00
## GarageFinishUnf      -1.346648e-02
## GarageCars            6.953160e-02
## GarageArea            0.000000e+00
## GarageQualFa          0.000000e+00
## GarageQualGd          1.742896e-02
## GarageQualPo          0.000000e+00
## GarageQualTA          0.000000e+00
## GarageCondFa         -3.960323e-02
## GarageCondGd          0.000000e+00
## GarageCondPo          0.000000e+00
## GarageCondTA          0.000000e+00
## PavedDriveP           0.000000e+00
## PavedDriveY           1.910285e-02
## WoodDeckSF            7.947326e-05
## OpenPorchSF           1.083448e-06
## EnclosedPorch         0.000000e+00
## X3SsnPorch            0.000000e+00
## ScreenPorch           1.837631e-04
## PoolArea             -2.301343e-04
## MiscVal               0.000000e+00
## MoSold10              0.000000e+00
## MoSold11              0.000000e+00
## MoSold12              0.000000e+00
## MoSold2               0.000000e+00
## MoSold3               0.000000e+00
## MoSold4               0.000000e+00
## MoSold5               0.000000e+00
## MoSold6               0.000000e+00
## MoSold7               0.000000e+00
## MoSold8               0.000000e+00
## MoSold9               0.000000e+00
## YrSold                0.000000e+00
## SaleTypeCon           0.000000e+00
## SaleTypeConLD         0.000000e+00
```

```
## SaleTypeConLI         0.000000e+00
## SaleTypeConLw         0.000000e+00
## SaleTypeCWD           0.000000e+00
## SaleTypeNew           7.025603e-02
## SaleTypeOth           0.000000e+00
## SaleTypeWD            0.000000e+00
## SaleConditionAdjLand  0.000000e+00
## SaleConditionAlloca   0.000000e+00
## SaleConditionFamily   0.000000e+00
## SaleConditionNormal   3.169134e-02
## SaleConditionPartial  0.000000e+00
```

```
coeffs[coeffs[,1] ==0, ]
```

```
##        MSSubClass180        MSSubClass190         MSSubClass40
##                    0                    0                    0
##         MSSubClass45         MSSubClass50         MSSubClass60
##                    0                    0                    0
##         MSSubClass70         MSSubClass75         MSSubClass80
##                    0                    0                    0
##         MSSubClass85         MSSubClass90           MSZoningRH
##                    0                    0                    0
##           MSZoningRM          LotFrontage          LotShapeIR2
##                    0                    0                    0
##          LotShapeReg       LandContourHLS       LandContourLow
##                    0                    0                    0
##       LandContourLvl          LotConfigFR2          LotConfigFR3
##                    0                    0                    0
##      LotConfigInside         LandSlopeMod         LandSlopeSev
##                    0                    0                    0
##  NeighborhoodBlueste  NeighborhoodBrDale  NeighborhoodCollgCr
##                    0                    0                    0
##  NeighborhoodGilbert NeighborhoodMitchel    NeighborhoodNAmes
##                    0                    0                    0
##  NeighborhoodNPkVill  NeighborhoodNWAmes   NeighborhoodSawyer
##                    0                    0                    0
##  NeighborhoodSawyerW    NeighborhoodSWISU  NeighborhoodTimber
##                    0                    0                    0
##       Condition1PosA       Condition1PosN       Condition1RRAe
##                    0                    0                    0
##       Condition1RRAn       Condition1RRNe       Condition1RRNn
##                    0                    0                    0
##      Condition2Feedr       Condition2Norm       Condition2PosA
##                    0                    0                    0
##       Condition2RRAe       Condition2RRAn       Condition2RRNn
##                    0                    0                    0
##       BldgType2fmCon       BldgTypeDuplex     HouseStyle1.5Unf
##                    0                    0                    0
##     HouseStyle1Story     HouseStyle2.5Fin     HouseStyle2.5Unf
##                    0                    0                    0
##     HouseStyle2Story     HouseStyleSFoyer       HouseStyleSLvl
##                    0                    0                    0
##     RoofStyleGambrel         RoofStyleHip     RoofStyleMansard
##                    0                    0                    0
```

```
##       RoofStyleShed      RoofMatlCompShg      RoofMatlMembran
##                   0                    0                    0
##       RoofMatlMetal         RoofMatlRoll      RoofMatlTar&Grv
##                   0                    0                    0
##      RoofMatlWdShake  Exterior1stAsphShn  Exterior1stBrkComm
##                   0                    0                    0
##   Exterior1stCBlock   Exterior1stCemntBd  Exterior1stImStucc
##                   0                    0                    0
##  Exterior1stMetalSd   Exterior1stPlywood    Exterior1stStone
##                   0                    0                    0
##   Exterior1stStucco   Exterior1stVinylSd  Exterior1stWdShing
##                   0                    0                    0
##  Exterior2ndAsphShn   Exterior2ndBrk Cmn  Exterior2ndBrkFace
##                   0                    0                    0
##   Exterior2ndCBlock   Exterior2ndCmentBd  Exterior2ndHdBoard
##                   0                    0                    0
##  Exterior2ndImStucc   Exterior2ndMetalSd    Exterior2ndOther
##                   0                    0                    0
##  Exterior2ndPlywood     Exterior2ndStone  Exterior2ndVinylSd
##                   0                    0                    0
##  Exterior2ndWd Sdng   Exterior2ndWd Shng   MasVnrTypeBrkFace
##                   0                    0                    0
##      MasVnrTypeNone          ExterQualGd         ExterCondGd
##                   0                    0                    0
##         ExterCondPo          ExterCondTA    FoundationCBlock
##                   0                    0                    0
##     FoundationStone       FoundationWood          BsmtQualFa
##                   0                    0                    0
##          BsmtQualGd           BsmtCondGd          BsmtCondPo
##                   0                    0                    0
##          BsmtCondTA       BsmtExposureMn     BsmtFinType1BLQ
##                   0                    0                    0
##     BsmtFinType1LwQ      BsmtFinType1Rec          BsmtFinSF1
##                   0                    0                    0
##     BsmtFinType2GLQ      BsmtFinType2LwQ     BsmtFinType2Rec
##                   0                    0                    0
##     BsmtFinType2Unf           BsmtFinSF2           BsmtUnfSF
##                   0                    0                    0
##         HeatingGasA         HeatingOthW         HeatingWall
##                   0                    0                    0
##         HeatingQCFa         HeatingQCGd         HeatingQCPo
##                   0                    0                    0
##     ElectricalFuseF     ElectricalFuseP       ElectricalMix
##                   0                    0                    0
##     ElectricalSBrkr           X2ndFlrSF        LowQualFinSF
##                   0                    0                    0
##        BsmtHalfBath        BedroomAbvGr       KitchenQualFa
##                   0                    0                    0
##       KitchenQualGd      FunctionalMin1      FunctionalMin2
##                   0                    0                    0
##    GarageTypeAttchd   GarageTypeBasment   GarageTypeBuiltIn
##                   0                    0                    0
##   GarageTypeCarPort         GarageYrBlt      GarageFinishRFn
##                   0                    0                    0
```

```
##           GarageArea         GarageQualFa         GarageQualPo
##                    0                    0                    0
##         GarageQualTA         GarageCondGd         GarageCondPo
##                    0                    0                    0
##         GarageCondTA          PavedDriveP        EnclosedPorch
##                    0                    0                    0
##           X3SsnPorch              MiscVal             MoSold10
##                    0                    0                    0
##             MoSold11             MoSold12              MoSold2
##                    0                    0                    0
##              MoSold3              MoSold4              MoSold5
##                    0                    0                    0
##              MoSold6              MoSold7              MoSold8
##                    0                    0                    0
##              MoSold9               YrSold          SaleTypeCon
##                    0                    0                    0
##        SaleTypeConLD        SaleTypeConLI        SaleTypeConLw
##                    0                    0                    0
##          SaleTypeCWD          SaleTypeOth           SaleTypeWD
##                    0                    0                    0
## SaleConditionAdjLand  SaleConditionAlloca  SaleConditionFamily
##                    0                    0                    0
## SaleConditionPartial
##                    0
```

Some variables lambda values were reduced to zero, which means they weren't used in the model for prediction. The lasso models automatically carry out variable selection. Variables shown above were the ones which were reduced to 0.

```
predictions <- predict(lasso,data_test)
RMSE(predictions, data_test$SalePrice)
```

```
## [1] 0.1167734
```

*Ridge Linear Regression Model*

```
set.seed(1)

ridge <- train(
  SalePrice ~., data = data_train, method = "glmnet",
  trControl = trainControl("cv", number = 10),
  tuneGrid = expand.grid(alpha = 0, lambda = 10^seq(-3, 3, length = 100)))
```

```
## Warning in nominalTrainWorkflow(x = x, y = y, wts = weights, info = trainInfo, :
## There were missing values in resampled performance measures.
```

```
predictions <- predict(ridge,data_test)
RMSE(predictions, data_test$SalePrice)
```

```
## [1] 0.1237117
```

*Elastic Net Linear Regression Model*

```
set.seed(1)

enet <- train(
  SalePrice ~., data = data_train, method = "glmnet",
  trControl = trainControl("cv", number = 10),
  tuneGrid = expand.grid(alpha = seq(0,1, length=10), lambda = 10^seq(-3, 3, length = 100)))
```

```
## Warning in nominalTrainWorkflow(x = x, y = y, wts = weights, info = trainInfo, :
## There were missing values in resampled performance measures.
```

```
predictions <- predict(enet,data_test)
RMSE(predictions, data_test$SalePrice)
```

```
## [1] 0.1179759
```

### Random Forest Models

```
set.seed(1)

m_rf <- train(SalePrice ~ ., data = data_train, method = "rf", importance = TRUE,
              trControl = trainControl(method = "cv", number = 10), tuneGrid = expand.grid(mtry = c(2, 4
```

```
predictions <- predict(m_rf,data_test)
RMSE(predictions, data_test$SalePrice)
```

```
## [1] 0.1331681
```

```
varImp(m_rf)
```

```
## rf variable importance
##
##   only 20 most important variables shown (out of 255)
##
##                Overall
## GrLivArea       100.00
## X1stFlrSF        90.38
## OverallQual      86.70
## TotalBsmtSF      85.92
## LotArea          83.21
## BsmtFinSF1       79.45
## GarageArea       78.14
## X2ndFlrSF        76.74
## Fireplaces       70.22
## GarageCars       69.94
## LotFrontage      66.37
## OverallCond      65.92
## ExterQualTA      63.80
## TotRmsAbvGrd     63.62
```

```
## YearBuilt         62.88
## BsmtQualGd        60.58
## HalfBath          58.87
## GarageYrBlt       58.66
## GarageFinishUnf   58.38
## YearRemodAdd      58.05
```

```r
listRFImp <- varImp(m_rf)$importance
listRFImp$Var <- row.names(listRFImp)
row.names(listRFImp) <- NULL
listRFImp <- listRFImp[order(listRFImp$Overall, decreasing = TRUE),][1:20,2]


#as.data.frame(coeffs[coeffs[,1] !=0,])
```

```r
coeffsDF <- as.data.frame(coeffs)
coeffsDF$Var <- row.names(coeffsDF)
row.names(coeffsDF) <- NULL

coeffsDF <- coeffsDF[coeffsDF$Var %in% listRFImp,]
coeffsDF
```

```
##                 1             Var
## 20   0.000000e+00     LotFrontage
## 21   1.352954e-06         LotArea
## 86   7.049432e-02     OverallQual
## 87   2.831610e-02     OverallCond
## 88   7.839586e-04       YearBuilt
## 89   9.272254e-04    YearRemodAdd
## 137 -6.947233e-03     ExterQualTA
## 148  0.000000e+00      BsmtQualGd
## 161  0.000000e+00      BsmtFinSF1
## 169  2.773587e-05     TotalBsmtSF
## 184  3.309808e-07       X1stFlrSF
## 185  0.000000e+00       X2ndFlrSF
## 187  1.969026e-04       GrLivArea
## 191  1.509926e-02        HalfBath
## 197  5.294280e-03    TotRmsAbvGrd
## 204  2.758108e-02      Fireplaces
## 210  0.000000e+00     GarageYrBlt
## 212 -1.346648e-02 GarageFinishUnf
## 213  6.953160e-02      GarageCars
## 214  0.000000e+00      GarageArea
```

By examining the filtered DF, we can see what the Lasso Regression model determined for the top 20 variables that the Random Forest model deemed important. Of the 20 that were identified important by RF, 5 were totally ignored by Lasso, and the one that most important in RF 'GrLivArea', was one of the important in Lasso but not the most important.

**Gradient Boosted Trees Model**

```r
set.seed(1)
```

```r
gbm <- train(
  SalePrice ~., data = data_train, method = "gbm",
  trControl = trainControl("cv", number = 10), preProc = "nzv")
```

```
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##        1        0.1483             nan     0.1000    0.0145
##        2        0.1372             nan     0.1000    0.0119
##        3        0.1276             nan     0.1000    0.0094
##        4        0.1190             nan     0.1000    0.0083
##        5        0.1115             nan     0.1000    0.0079
##        6        0.1056             nan     0.1000    0.0060
##        7        0.0995             nan     0.1000    0.0057
##        8        0.0943             nan     0.1000    0.0051
##        9        0.0895             nan     0.1000    0.0042
##       10        0.0852             nan     0.1000    0.0039
##       20        0.0560             nan     0.1000    0.0018
##       40        0.0325             nan     0.1000    0.0002
##       60        0.0240             nan     0.1000    0.0002
##       80        0.0200             nan     0.1000    0.0000
##      100        0.0181             nan     0.1000   -0.0000
##      120        0.0170             nan     0.1000   -0.0000
##      140        0.0161             nan     0.1000   -0.0000
##      150        0.0157             nan     0.1000   -0.0000
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##        1        0.1466             nan     0.1000    0.0162
##        2        0.1321             nan     0.1000    0.0131
##        3        0.1203             nan     0.1000    0.0108
##        4        0.1085             nan     0.1000    0.0106
##        5        0.0991             nan     0.1000    0.0099
##        6        0.0926             nan     0.1000    0.0057
##        7        0.0858             nan     0.1000    0.0058
##        8        0.0789             nan     0.1000    0.0064
##        9        0.0738             nan     0.1000    0.0052
##       10        0.0687             nan     0.1000    0.0049
##       20        0.0388             nan     0.1000    0.0019
##       40        0.0212             nan     0.1000    0.0002
##       60        0.0169             nan     0.1000   -0.0002
##       80        0.0148             nan     0.1000   -0.0000
##      100        0.0135             nan     0.1000   -0.0001
##      120        0.0127             nan     0.1000   -0.0000
##      140        0.0120             nan     0.1000   -0.0000
##      150        0.0117             nan     0.1000   -0.0001
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##        1        0.1449             nan     0.1000    0.0186
##        2        0.1291             nan     0.1000    0.0155
##        3        0.1154             nan     0.1000    0.0136
##        4        0.1040             nan     0.1000    0.0114
##        5        0.0938             nan     0.1000    0.0093
##        6        0.0856             nan     0.1000    0.0080
##        7        0.0780             nan     0.1000    0.0076
##        8        0.0719             nan     0.1000    0.0053
```

```
##      9        0.0655             nan     0.1000    0.0063
##     10        0.0602             nan     0.1000    0.0044
##     20        0.0320             nan     0.1000    0.0011
##     40        0.0175             nan     0.1000    0.0001
##     60        0.0140             nan     0.1000    0.0000
##     80        0.0124             nan     0.1000    0.0000
##    100        0.0112             nan     0.1000    0.0000
##    120        0.0101             nan     0.1000    0.0000
##    140        0.0094             nan     0.1000   -0.0000
##    150        0.0091             nan     0.1000   -0.0001
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        0.1482             nan     0.1000    0.0139
##      2        0.1373             nan     0.1000    0.0114
##      3        0.1272             nan     0.1000    0.0093
##      4        0.1195             nan     0.1000    0.0068
##      5        0.1125             nan     0.1000    0.0073
##      6        0.1058             nan     0.1000    0.0065
##      7        0.1001             nan     0.1000    0.0060
##      8        0.0947             nan     0.1000    0.0048
##      9        0.0901             nan     0.1000    0.0044
##     10        0.0854             nan     0.1000    0.0040
##     20        0.0567             nan     0.1000    0.0015
##     40        0.0335             nan     0.1000    0.0005
##     60        0.0245             nan     0.1000    0.0002
##     80        0.0203             nan     0.1000    0.0001
##    100        0.0180             nan     0.1000    0.0000
##    120        0.0167             nan     0.1000   -0.0000
##    140        0.0158             nan     0.1000    0.0000
##    150        0.0155             nan     0.1000   -0.0000
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        0.1472             nan     0.1000    0.0166
##      2        0.1326             nan     0.1000    0.0136
##      3        0.1205             nan     0.1000    0.0114
##      4        0.1096             nan     0.1000    0.0105
##      5        0.1006             nan     0.1000    0.0079
##      6        0.0926             nan     0.1000    0.0079
##      7        0.0866             nan     0.1000    0.0047
##      8        0.0800             nan     0.1000    0.0057
##      9        0.0740             nan     0.1000    0.0058
##     10        0.0691             nan     0.1000    0.0043
##     20        0.0406             nan     0.1000    0.0015
##     40        0.0230             nan     0.1000    0.0002
##     60        0.0182             nan     0.1000    0.0001
##     80        0.0161             nan     0.1000   -0.0001
##    100        0.0148             nan     0.1000   -0.0001
##    120        0.0140             nan     0.1000   -0.0001
##    140        0.0133             nan     0.1000   -0.0000
##    150        0.0129             nan     0.1000   -0.0000
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        0.1438             nan     0.1000    0.0176
##      2        0.1284             nan     0.1000    0.0163
```

```
##      3        0.1149             nan     0.1000    0.0135
##      4        0.1034             nan     0.1000    0.0106
##      5        0.0937             nan     0.1000    0.0076
##      6        0.0844             nan     0.1000    0.0092
##      7        0.0770             nan     0.1000    0.0064
##      8        0.0705             nan     0.1000    0.0060
##      9        0.0647             nan     0.1000    0.0053
##     10        0.0596             nan     0.1000    0.0044
##     20        0.0322             nan     0.1000    0.0011
##     40        0.0178             nan     0.1000    0.0001
##     60        0.0143             nan     0.1000   -0.0001
##     80        0.0128             nan     0.1000   -0.0000
##    100        0.0118             nan     0.1000   -0.0000
##    120        0.0110             nan     0.1000   -0.0000
##    140        0.0103             nan     0.1000   -0.0001
##    150        0.0101             nan     0.1000   -0.0000
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        0.1466             nan     0.1000    0.0140
##      2        0.1356             nan     0.1000    0.0115
##      3        0.1262             nan     0.1000    0.0095
##      4        0.1183             nan     0.1000    0.0082
##      5        0.1117             nan     0.1000    0.0067
##      6        0.1051             nan     0.1000    0.0068
##      7        0.0992             nan     0.1000    0.0056
##      8        0.0936             nan     0.1000    0.0054
##      9        0.0887             nan     0.1000    0.0042
##     10        0.0842             nan     0.1000    0.0043
##     20        0.0565             nan     0.1000    0.0018
##     40        0.0338             nan     0.1000    0.0005
##     60        0.0254             nan     0.1000    0.0002
##     80        0.0217             nan     0.1000    0.0000
##    100        0.0197             nan     0.1000   -0.0000
##    120        0.0183             nan     0.1000    0.0000
##    140        0.0175             nan     0.1000   -0.0000
##    150        0.0171             nan     0.1000    0.0000
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        0.1440             nan     0.1000    0.0162
##      2        0.1287             nan     0.1000    0.0140
##      3        0.1164             nan     0.1000    0.0118
##      4        0.1069             nan     0.1000    0.0098
##      5        0.0985             nan     0.1000    0.0068
##      6        0.0909             nan     0.1000    0.0075
##      7        0.0841             nan     0.1000    0.0063
##      8        0.0783             nan     0.1000    0.0054
##      9        0.0727             nan     0.1000    0.0056
##     10        0.0679             nan     0.1000    0.0045
##     20        0.0401             nan     0.1000    0.0017
##     40        0.0230             nan     0.1000    0.0001
##     60        0.0180             nan     0.1000    0.0000
##     80        0.0157             nan     0.1000   -0.0000
##    100        0.0147             nan     0.1000   -0.0001
##    120        0.0137             nan     0.1000   -0.0001
```

```
##    140        0.0130             nan     0.1000   -0.0000
##    150        0.0127             nan     0.1000   -0.0000
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        0.1417             nan     0.1000    0.0175
##      2        0.1259             nan     0.1000    0.0163
##      3        0.1123             nan     0.1000    0.0124
##      4        0.1016             nan     0.1000    0.0112
##      5        0.0921             nan     0.1000    0.0091
##      6        0.0844             nan     0.1000    0.0070
##      7        0.0771             nan     0.1000    0.0066
##      8        0.0704             nan     0.1000    0.0062
##      9        0.0646             nan     0.1000    0.0050
##     10        0.0597             nan     0.1000    0.0043
##     20        0.0327             nan     0.1000    0.0014
##     40        0.0189             nan     0.1000    0.0002
##     60        0.0150             nan     0.1000   -0.0000
##     80        0.0132             nan     0.1000    0.0000
##    100        0.0121             nan     0.1000   -0.0000
##    120        0.0113             nan     0.1000   -0.0000
##    140        0.0105             nan     0.1000   -0.0000
##    150        0.0102             nan     0.1000   -0.0000
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        0.1449             nan     0.1000    0.0140
##      2        0.1341             nan     0.1000    0.0113
##      3        0.1254             nan     0.1000    0.0087
##      4        0.1170             nan     0.1000    0.0081
##      5        0.1102             nan     0.1000    0.0074
##      6        0.1038             nan     0.1000    0.0063
##      7        0.0977             nan     0.1000    0.0054
##      8        0.0923             nan     0.1000    0.0046
##      9        0.0874             nan     0.1000    0.0046
##     10        0.0831             nan     0.1000    0.0035
##     20        0.0555             nan     0.1000    0.0017
##     40        0.0330             nan     0.1000    0.0004
##     60        0.0248             nan     0.1000    0.0002
##     80        0.0214             nan     0.1000    0.0000
##    100        0.0195             nan     0.1000    0.0000
##    120        0.0183             nan     0.1000   -0.0000
##    140        0.0175             nan     0.1000   -0.0001
##    150        0.0171             nan     0.1000   -0.0000
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        0.1430             nan     0.1000    0.0155
##      2        0.1298             nan     0.1000    0.0135
##      3        0.1185             nan     0.1000    0.0113
##      4        0.1091             nan     0.1000    0.0083
##      5        0.1002             nan     0.1000    0.0096
##      6        0.0919             nan     0.1000    0.0081
##      7        0.0853             nan     0.1000    0.0056
##      8        0.0794             nan     0.1000    0.0053
##      9        0.0737             nan     0.1000    0.0058
##     10        0.0690             nan     0.1000    0.0043
```

```
##     20        0.0399             nan     0.1000    0.0019
##     40        0.0226             nan     0.1000    0.0002
##     60        0.0183             nan     0.1000    0.0001
##     80        0.0162             nan     0.1000   -0.0001
##    100        0.0149             nan     0.1000   -0.0000
##    120        0.0140             nan     0.1000   -0.0001
##    140        0.0134             nan     0.1000   -0.0000
##    150        0.0131             nan     0.1000   -0.0001
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        0.1399             nan     0.1000    0.0170
##      2        0.1248             nan     0.1000    0.0145
##      3        0.1124             nan     0.1000    0.0127
##      4        0.1011             nan     0.1000    0.0114
##      5        0.0919             nan     0.1000    0.0079
##      6        0.0838             nan     0.1000    0.0076
##      7        0.0762             nan     0.1000    0.0074
##      8        0.0698             nan     0.1000    0.0064
##      9        0.0648             nan     0.1000    0.0048
##     10        0.0598             nan     0.1000    0.0045
##     20        0.0324             nan     0.1000    0.0012
##     40        0.0186             nan     0.1000    0.0001
##     60        0.0148             nan     0.1000    0.0001
##     80        0.0132             nan     0.1000   -0.0001
##    100        0.0120             nan     0.1000   -0.0000
##    120        0.0111             nan     0.1000   -0.0000
##    140        0.0103             nan     0.1000   -0.0000
##    150        0.0100             nan     0.1000   -0.0000
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        0.1470             nan     0.1000    0.0143
##      2        0.1351             nan     0.1000    0.0108
##      3        0.1252             nan     0.1000    0.0089
##      4        0.1172             nan     0.1000    0.0079
##      5        0.1099             nan     0.1000    0.0073
##      6        0.1034             nan     0.1000    0.0060
##      7        0.0979             nan     0.1000    0.0049
##      8        0.0927             nan     0.1000    0.0049
##      9        0.0886             nan     0.1000    0.0037
##     10        0.0838             nan     0.1000    0.0046
##     20        0.0562             nan     0.1000    0.0020
##     40        0.0335             nan     0.1000    0.0007
##     60        0.0253             nan     0.1000    0.0002
##     80        0.0218             nan     0.1000    0.0001
##    100        0.0200             nan     0.1000    0.0000
##    120        0.0187             nan     0.1000    0.0000
##    140        0.0178             nan     0.1000   -0.0000
##    150        0.0174             nan     0.1000    0.0000
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        0.1445             nan     0.1000    0.0162
##      2        0.1315             nan     0.1000    0.0132
##      3        0.1206             nan     0.1000    0.0109
##      4        0.1097             nan     0.1000    0.0105
```

```
##      5        0.1012             nan     0.1000    0.0080
##      6        0.0938             nan     0.1000    0.0066
##      7        0.0865             nan     0.1000    0.0068
##      8        0.0799             nan     0.1000    0.0066
##      9        0.0741             nan     0.1000    0.0053
##     10        0.0690             nan     0.1000    0.0049
##     20        0.0409             nan     0.1000    0.0013
##     40        0.0240             nan     0.1000    0.0004
##     60        0.0190             nan     0.1000    0.0000
##     80        0.0168             nan     0.1000   -0.0001
##    100        0.0153             nan     0.1000    0.0000
##    120        0.0141             nan     0.1000   -0.0001
##    140        0.0133             nan     0.1000   -0.0000
##    150        0.0130             nan     0.1000   -0.0000
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        0.1426             nan     0.1000    0.0173
##      2        0.1279             nan     0.1000    0.0148
##      3        0.1148             nan     0.1000    0.0130
##      4        0.1041             nan     0.1000    0.0096
##      5        0.0948             nan     0.1000    0.0079
##      6        0.0854             nan     0.1000    0.0086
##      7        0.0783             nan     0.1000    0.0071
##      8        0.0720             nan     0.1000    0.0059
##      9        0.0659             nan     0.1000    0.0055
##     10        0.0607             nan     0.1000    0.0047
##     20        0.0337             nan     0.1000    0.0011
##     40        0.0195             nan     0.1000    0.0002
##     60        0.0156             nan     0.1000    0.0000
##     80        0.0137             nan     0.1000   -0.0000
##    100        0.0124             nan     0.1000   -0.0001
##    120        0.0114             nan     0.1000   -0.0001
##    140        0.0106             nan     0.1000   -0.0000
##    150        0.0103             nan     0.1000   -0.0000
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        0.1516             nan     0.1000    0.0149
##      2        0.1396             nan     0.1000    0.0123
##      3        0.1303             nan     0.1000    0.0093
##      4        0.1218             nan     0.1000    0.0083
##      5        0.1140             nan     0.1000    0.0080
##      6        0.1066             nan     0.1000    0.0069
##      7        0.1008             nan     0.1000    0.0055
##      8        0.0951             nan     0.1000    0.0051
##      9        0.0900             nan     0.1000    0.0050
##     10        0.0852             nan     0.1000    0.0044
##     20        0.0571             nan     0.1000    0.0019
##     40        0.0337             nan     0.1000    0.0006
##     60        0.0247             nan     0.1000    0.0001
##     80        0.0208             nan     0.1000    0.0001
##    100        0.0187             nan     0.1000    0.0000
##    120        0.0174             nan     0.1000    0.0000
##    140        0.0167             nan     0.1000   -0.0000
##    150        0.0164             nan     0.1000    0.0000
```

```
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        0.1497             nan     0.1000    0.0165
##      2        0.1351             nan     0.1000    0.0148
##      3        0.1224             nan     0.1000    0.0118
##      4        0.1112             nan     0.1000    0.0112
##      5        0.1021             nan     0.1000    0.0080
##      6        0.0940             nan     0.1000    0.0078
##      7        0.0864             nan     0.1000    0.0068
##      8        0.0805             nan     0.1000    0.0058
##      9        0.0748             nan     0.1000    0.0057
##     10        0.0703             nan     0.1000    0.0047
##     20        0.0410             nan     0.1000    0.0016
##     40        0.0226             nan     0.1000    0.0002
##     60        0.0178             nan     0.1000    0.0000
##     80        0.0161             nan     0.1000    0.0000
##    100        0.0148             nan     0.1000   -0.0000
##    120        0.0139             nan     0.1000   -0.0001
##    140        0.0133             nan     0.1000   -0.0000
##    150        0.0130             nan     0.1000    0.0000
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        0.1468             nan     0.1000    0.0206
##      2        0.1300             nan     0.1000    0.0145
##      3        0.1163             nan     0.1000    0.0129
##      4        0.1043             nan     0.1000    0.0115
##      5        0.0942             nan     0.1000    0.0097
##      6        0.0854             nan     0.1000    0.0080
##      7        0.0774             nan     0.1000    0.0065
##      8        0.0715             nan     0.1000    0.0056
##      9        0.0660             nan     0.1000    0.0044
##     10        0.0608             nan     0.1000    0.0051
##     20        0.0324             nan     0.1000    0.0014
##     40        0.0178             nan     0.1000    0.0002
##     60        0.0144             nan     0.1000   -0.0000
##     80        0.0131             nan     0.1000   -0.0000
##    100        0.0119             nan     0.1000   -0.0000
##    120        0.0110             nan     0.1000   -0.0001
##    140        0.0105             nan     0.1000   -0.0001
##    150        0.0101             nan     0.1000   -0.0001
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        0.1459             nan     0.1000    0.0140
##      2        0.1347             nan     0.1000    0.0116
##      3        0.1249             nan     0.1000    0.0089
##      4        0.1162             nan     0.1000    0.0084
##      5        0.1095             nan     0.1000    0.0063
##      6        0.1028             nan     0.1000    0.0062
##      7        0.0964             nan     0.1000    0.0061
##      8        0.0915             nan     0.1000    0.0050
##      9        0.0872             nan     0.1000    0.0046
##     10        0.0828             nan     0.1000    0.0042
##     20        0.0558             nan     0.1000    0.0019
##     40        0.0331             nan     0.1000    0.0006
```

```
##     60        0.0247             nan     0.1000    0.0001
##     80        0.0207             nan     0.1000    0.0000
##    100        0.0186             nan     0.1000   -0.0000
##    120        0.0176             nan     0.1000   -0.0001
##    140        0.0166             nan     0.1000    0.0000
##    150        0.0162             nan     0.1000   -0.0000
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        0.1437             nan     0.1000    0.0157
##      2        0.1297             nan     0.1000    0.0138
##      3        0.1177             nan     0.1000    0.0121
##      4        0.1083             nan     0.1000    0.0088
##      5        0.0985             nan     0.1000    0.0092
##      6        0.0901             nan     0.1000    0.0076
##      7        0.0837             nan     0.1000    0.0061
##      8        0.0774             nan     0.1000    0.0058
##      9        0.0720             nan     0.1000    0.0048
##     10        0.0677             nan     0.1000    0.0041
##     20        0.0402             nan     0.1000    0.0014
##     40        0.0223             nan     0.1000    0.0002
##     60        0.0175             nan     0.1000    0.0001
##     80        0.0154             nan     0.1000   -0.0001
##    100        0.0140             nan     0.1000    0.0000
##    120        0.0131             nan     0.1000   -0.0000
##    140        0.0124             nan     0.1000   -0.0000
##    150        0.0121             nan     0.1000   -0.0000
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        0.1405             nan     0.1000    0.0181
##      2        0.1245             nan     0.1000    0.0149
##      3        0.1113             nan     0.1000    0.0125
##      4        0.1007             nan     0.1000    0.0092
##      5        0.0906             nan     0.1000    0.0097
##      6        0.0822             nan     0.1000    0.0076
##      7        0.0758             nan     0.1000    0.0055
##      8        0.0698             nan     0.1000    0.0052
##      9        0.0643             nan     0.1000    0.0052
##     10        0.0595             nan     0.1000    0.0041
##     20        0.0325             nan     0.1000    0.0010
##     40        0.0183             nan     0.1000    0.0002
##     60        0.0146             nan     0.1000   -0.0001
##     80        0.0130             nan     0.1000    0.0000
##    100        0.0118             nan     0.1000   -0.0000
##    120        0.0109             nan     0.1000   -0.0001
##    140        0.0101             nan     0.1000   -0.0001
##    150        0.0098             nan     0.1000   -0.0000
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        0.1433             nan     0.1000    0.0140
##      2        0.1325             nan     0.1000    0.0113
##      3        0.1229             nan     0.1000    0.0090
##      4        0.1148             nan     0.1000    0.0073
##      5        0.1071             nan     0.1000    0.0073
##      6        0.1012             nan     0.1000    0.0053
```

```
##      7        0.0957             nan     0.1000    0.0053
##      8        0.0900             nan     0.1000    0.0052
##      9        0.0851             nan     0.1000    0.0045
##     10        0.0810             nan     0.1000    0.0040
##     20        0.0540             nan     0.1000    0.0012
##     40        0.0320             nan     0.1000    0.0005
##     60        0.0239             nan     0.1000    0.0003
##     80        0.0203             nan     0.1000   -0.0001
##    100        0.0187             nan     0.1000   -0.0000
##    120        0.0174             nan     0.1000    0.0000
##    140        0.0166             nan     0.1000   -0.0001
##    150        0.0163             nan     0.1000   -0.0001
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        0.1414             nan     0.1000    0.0163
##      2        0.1284             nan     0.1000    0.0134
##      3        0.1167             nan     0.1000    0.0112
##      4        0.1065             nan     0.1000    0.0107
##      5        0.0986             nan     0.1000    0.0071
##      6        0.0901             nan     0.1000    0.0084
##      7        0.0835             nan     0.1000    0.0067
##      8        0.0779             nan     0.1000    0.0053
##      9        0.0729             nan     0.1000    0.0048
##     10        0.0674             nan     0.1000    0.0054
##     20        0.0385             nan     0.1000    0.0016
##     40        0.0219             nan     0.1000    0.0003
##     60        0.0176             nan     0.1000    0.0000
##     80        0.0154             nan     0.1000   -0.0000
##    100        0.0142             nan     0.1000   -0.0000
##    120        0.0133             nan     0.1000   -0.0000
##    140        0.0125             nan     0.1000   -0.0000
##    150        0.0122             nan     0.1000   -0.0001
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        0.1394             nan     0.1000    0.0193
##      2        0.1239             nan     0.1000    0.0151
##      3        0.1110             nan     0.1000    0.0127
##      4        0.0997             nan     0.1000    0.0117
##      5        0.0905             nan     0.1000    0.0082
##      6        0.0820             nan     0.1000    0.0083
##      7        0.0742             nan     0.1000    0.0074
##      8        0.0681             nan     0.1000    0.0058
##      9        0.0624             nan     0.1000    0.0060
##     10        0.0582             nan     0.1000    0.0038
##     20        0.0314             nan     0.1000    0.0013
##     40        0.0185             nan     0.1000    0.0001
##     60        0.0146             nan     0.1000    0.0000
##     80        0.0130             nan     0.1000   -0.0000
##    100        0.0118             nan     0.1000   -0.0000
##    120        0.0110             nan     0.1000   -0.0000
##    140        0.0103             nan     0.1000   -0.0000
##    150        0.0100             nan     0.1000   -0.0000
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
```

```
##      1        0.1495             nan     0.1000    0.0143
##      2        0.1376             nan     0.1000    0.0118
##      3        0.1283             nan     0.1000    0.0090
##      4        0.1202             nan     0.1000    0.0084
##      5        0.1127             nan     0.1000    0.0072
##      6        0.1057             nan     0.1000    0.0066
##      7        0.0994             nan     0.1000    0.0057
##      8        0.0941             nan     0.1000    0.0046
##      9        0.0894             nan     0.1000    0.0046
##     10        0.0851             nan     0.1000    0.0041
##     20        0.0581             nan     0.1000    0.0008
##     40        0.0341             nan     0.1000    0.0006
##     60        0.0255             nan     0.1000    0.0002
##     80        0.0215             nan     0.1000    0.0001
##    100        0.0195             nan     0.1000   -0.0000
##    120        0.0182             nan     0.1000    0.0001
##    140        0.0173             nan     0.1000   -0.0001
##    150        0.0170             nan     0.1000   -0.0000
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        0.1464             nan     0.1000    0.0168
##      2        0.1325             nan     0.1000    0.0129
##      3        0.1202             nan     0.1000    0.0103
##      4        0.1106             nan     0.1000    0.0094
##      5        0.1009             nan     0.1000    0.0097
##      6        0.0926             nan     0.1000    0.0077
##      7        0.0855             nan     0.1000    0.0074
##      8        0.0798             nan     0.1000    0.0053
##      9        0.0746             nan     0.1000    0.0047
##     10        0.0696             nan     0.1000    0.0049
##     20        0.0403             nan     0.1000    0.0014
##     40        0.0226             nan     0.1000    0.0003
##     60        0.0179             nan     0.1000    0.0001
##     80        0.0158             nan     0.1000   -0.0001
##    100        0.0148             nan     0.1000   -0.0000
##    120        0.0138             nan     0.1000   -0.0000
##    140        0.0130             nan     0.1000   -0.0000
##    150        0.0128             nan     0.1000   -0.0000
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        0.1453             nan     0.1000    0.0184
##      2        0.1297             nan     0.1000    0.0155
##      3        0.1155             nan     0.1000    0.0130
##      4        0.1041             nan     0.1000    0.0108
##      5        0.0940             nan     0.1000    0.0092
##      6        0.0851             nan     0.1000    0.0079
##      7        0.0778             nan     0.1000    0.0064
##      8        0.0708             nan     0.1000    0.0071
##      9        0.0652             nan     0.1000    0.0054
##     10        0.0603             nan     0.1000    0.0049
##     20        0.0325             nan     0.1000    0.0014
##     40        0.0184             nan     0.1000    0.0001
##     60        0.0151             nan     0.1000    0.0000
##     80        0.0131             nan     0.1000   -0.0000
```

```
##    100        0.0121             nan     0.1000   -0.0001
##    120        0.0113             nan     0.1000   -0.0001
##    140        0.0106             nan     0.1000   -0.0000
##    150        0.0103             nan     0.1000   -0.0000
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        0.1444             nan     0.1000    0.0134
##      2        0.1332             nan     0.1000    0.0115
##      3        0.1237             nan     0.1000    0.0089
##      4        0.1158             nan     0.1000    0.0076
##      5        0.1086             nan     0.1000    0.0073
##      6        0.1024             nan     0.1000    0.0056
##      7        0.0965             nan     0.1000    0.0052
##      8        0.0914             nan     0.1000    0.0051
##      9        0.0865             nan     0.1000    0.0044
##     10        0.0822             nan     0.1000    0.0036
##     20        0.0552             nan     0.1000    0.0019
##     40        0.0329             nan     0.1000    0.0003
##     60        0.0246             nan     0.1000    0.0001
##     80        0.0209             nan     0.1000    0.0001
##    100        0.0189             nan     0.1000    0.0000
##    120        0.0176             nan     0.1000    0.0001
##    140        0.0168             nan     0.1000   -0.0000
##    150        0.0165             nan     0.1000   -0.0000
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        0.1419             nan     0.1000    0.0163
##      2        0.1286             nan     0.1000    0.0136
##      3        0.1166             nan     0.1000    0.0105
##      4        0.1062             nan     0.1000    0.0113
##      5        0.0982             nan     0.1000    0.0090
##      6        0.0901             nan     0.1000    0.0079
##      7        0.0837             nan     0.1000    0.0058
##      8        0.0780             nan     0.1000    0.0051
##      9        0.0721             nan     0.1000    0.0056
##     10        0.0675             nan     0.1000    0.0039
##     20        0.0397             nan     0.1000    0.0013
##     40        0.0228             nan     0.1000    0.0002
##     60        0.0182             nan     0.1000    0.0001
##     80        0.0161             nan     0.1000    0.0000
##    100        0.0149             nan     0.1000   -0.0000
##    120        0.0138             nan     0.1000   -0.0000
##    140        0.0131             nan     0.1000   -0.0000
##    150        0.0127             nan     0.1000    0.0000
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        0.1395             nan     0.1000    0.0185
##      2        0.1243             nan     0.1000    0.0160
##      3        0.1113             nan     0.1000    0.0129
##      4        0.0995             nan     0.1000    0.0107
##      5        0.0901             nan     0.1000    0.0083
##      6        0.0824             nan     0.1000    0.0073
##      7        0.0755             nan     0.1000    0.0066
##      8        0.0692             nan     0.1000    0.0060
```

```
##      9        0.0639             nan     0.1000    0.0053
##     10        0.0593             nan     0.1000    0.0043
##     20        0.0322             nan     0.1000    0.0014
##     40        0.0186             nan     0.1000    0.0000
##     60        0.0149             nan     0.1000   -0.0001
##     80        0.0133             nan     0.1000   -0.0001
##    100        0.0122             nan     0.1000   -0.0000
##    120        0.0114             nan     0.1000   -0.0001
##    140        0.0107             nan     0.1000   -0.0000
##    150        0.0103             nan     0.1000   -0.0000
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        0.1430             nan     0.1000    0.0185
##      2        0.1269             nan     0.1000    0.0145
##      3        0.1141             nan     0.1000    0.0124
##      4        0.1016             nan     0.1000    0.0119
##      5        0.0917             nan     0.1000    0.0102
##      6        0.0835             nan     0.1000    0.0073
##      7        0.0765             nan     0.1000    0.0071
##      8        0.0704             nan     0.1000    0.0050
##      9        0.0647             nan     0.1000    0.0052
##     10        0.0600             nan     0.1000    0.0044
##     20        0.0326             nan     0.1000    0.0014
##     40        0.0189             nan     0.1000   -0.0001
##     60        0.0148             nan     0.1000    0.0001
##     80        0.0132             nan     0.1000   -0.0000
##    100        0.0121             nan     0.1000   -0.0001
##    120        0.0112             nan     0.1000   -0.0000
##    140        0.0106             nan     0.1000   -0.0000
##    150        0.0103             nan     0.1000   -0.0000
```

```r
predictions <- predict(gbm,data_test)
RMSE(predictions, data_test$SalePrice)
```

```
## [1] 0.1209606
```

### Comparison of Models

```r
compare=resamples(list(L=lasso, R=ridge, E=enet, RF = m_rf, G = gbm))
```

```r
summary(compare)
```

```
##
## Call:
## summary.resamples(object = compare)
##
## Models: L, R, E, RF, G
## Number of resamples: 10
##
## MAE
##          Min.    1st Qu.     Median       Mean    3rd Qu.      Max. NA's
## L  0.07816799 0.08674583 0.09519405 0.09323783 0.09970306 0.1044491    0
```

```
## R  0.07670146 0.08862596 0.09988095 0.09506917 0.10156409 0.1050569    0
## E  0.07775151 0.08623393 0.09583900 0.09346221 0.10033290 0.1035560    0
## RF 0.07920558 0.09219023 0.09596471 0.09733000 0.10128513 0.1146040    0
## G  0.08507741 0.08999805 0.09521986 0.09332114 0.09617522 0.1026588    0
##
## RMSE
##          Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
## L  0.09778152 0.1217666 0.1452775 0.1518799 0.1575893 0.2695698    0
## R  0.09841857 0.1292656 0.1474946 0.1518827 0.1600506 0.2393770    0
## E  0.09687368 0.1229803 0.1459079 0.1492083 0.1521539 0.2525771    0
## RF 0.10259145 0.1415651 0.1510390 0.1487340 0.1570593 0.1860757    0
## G  0.10774435 0.1236806 0.1367302 0.1358157 0.1487888 0.1598050    0
##
## Rsquared
##         Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
## L  0.6420894 0.8496630 0.8959360 0.8551290 0.9138795 0.9399385    0
## R  0.6854652 0.8461922 0.8879079 0.8569824 0.8995269 0.9405751    0
## E  0.6622618 0.8557042 0.8957389 0.8610559 0.9161265 0.9415999    0
## RF 0.7961990 0.8514791 0.8764924 0.8746620 0.9014009 0.9420880    0
## G  0.8360557 0.8629792 0.8821800 0.8849289 0.9137745 0.9309447    0
```

Of the 5 models ran over the same dataset, the Gradient Boosted Trees model performed well producing the minimum RMSE compared to other models. The Ridge regression model performed the worst of them all.

**Neural Networks with Dropout**

```r
# Convert Year columns to numeric

cols <- c("YearBuilt", "YearRemodAdd", "GarageYrBlt", "YrSold")
data_train[cols] <- sapply(data_train[cols],as.numeric)
data_test[cols] <- sapply(data_test[cols],as.numeric)
```

```r
# Train validation split

train.index <- createDataPartition(data_train$SalePrice, p = 0.9, list = FALSE)
data_nn_train <- data_train[train.index, ]
data_nn_validation <- data_train[-train.index, ]

# Test set
data_nn_test <- data_test

# Separating and log transforming the target variable
data_nn_train_y <- data_nn_train$SalePrice
data_nn_train$SalePrice <- NULL

data_nn_test_y <- data_nn_test$SalePrice
data_nn_test$SalePrice <- NULL

data_nn_validation_y <- data_nn_validation$SalePrice
data_nn_validation$SalePrice <- NULL
```

```r
# Scaling the data

ind <- sapply(data_nn_train, is.numeric) # Only for numeric
```

```r
col_means_train <- lapply(data_nn_train[ind], mean)
col_stddevs_train <- lapply(data_nn_train[ind], sd)

data_nn_train[ind] <- lapply(data_nn_train[ind], scale)
```

```r
# Scaling validation and testing data using mean and sd

data_nn_validation[ind] <- scale(data_nn_validation[ind], center = col_means_train, scale = col_stddevs_
data_nn_test[ind] <- scale(data_nn_test[ind], center = col_means_train, scale = col_stddevs_train)
```

```r
# One Hot Encoding

library("mltools")
```

```
## Warning: package 'mltools' was built under R version 3.6.3
```

```r
library("data.table")
```

```
## Warning: package 'data.table' was built under R version 3.6.3
```

```r
data_nn_train = as.data.frame(one_hot(as.data.table(data_nn_train)))
data_nn_validation = as.data.frame(one_hot(as.data.table(data_nn_validation)))
data_nn_test = as.data.frame(one_hot(as.data.table(data_nn_test)))
```

```r
library("tfruns")

runs <- tuning_run("house_train.R",
                   flags = list(
                   nodes_hlayer1 = c(600, 500, 400),
                   nodes_hlayer2 = c(500, 250, 100),
                   learning_rate = c(0.01, 0.05, 0.001, 0.0001),
                   batch_size=c(10,20,50,75),
                   epochs=c(30,50,100),
                   activation=c("relu","sigmoid","tanh"),
                   dropout1=c(0.3, 0.4, 0.5),
                   dropout2=c(0.2, 0.3, 0.4)),
                   sample = 0.01
)
```

```r
#runs
#view_run(runs$run_dir[9])

runsHouse <- runs[order(runs$metric_val_loss, decreasing = FALSE),][1,]
```

Best performing with params: nodes_hlayer1 = 400, nodes_hlayer2 = 500, batch_size = 10, activation = sigmoid, learning_rate = $10^{-4}$, epochs = 50 dropout1 = 0.4, dropout2 = 0.4.

The model doesn not overfit since the difference between error is little with the training error = 0.4511 and validation error = 0.0249.

**Model Testing**

```r
# Combine validation and training data
data_nn_train_all <- rbind(data_nn_train, data_nn_validation)
data_nn_train_all_y <- c(data_nn_train_y, data_nn_validation_y)
```

```r
set.seed(1)

model <- keras_model_sequential()
model %>%
  layer_dense(units = 900, activation = runsHouse$flag_activation, input_shape = dim(data_nn_train)[2])
  layer_dense(units = runsHouse$flag_nodes_hlayer1, activation = runsHouse$flag_activation) %>%
  layer_dropout(runsHouse$flag_dropout1) %>%
  layer_dense(units = runsHouse$flag_nodes_hlayer2, activation = runsHouse$flag_activation) %>%
  layer_dropout(runsHouse$flag_dropout2) %>%
  layer_dense(units = 1)

model %>%
  compile(optimizer = optimizer_adam(lr = runsHouse$flag_learning_rate), loss = 'mse')

set.seed(123)

model %>%
  fit(as.matrix(data_nn_train_all), as.matrix(data_nn_train_all_y), batch_size = runsHouse$flag_batch_si
      validation_data = list(as.matrix(data_nn_test), as.matrix(data_nn_test_y)))
```

```r
predictions <- model %>% predict(as.matrix(data_nn_test))

cat('RMSE:', RMSE(predictions, as.matrix(data_nn_test_y)))
```

```
## RMSE: 0.130623
```

Compared with the Gradient Boosted Trees model which had the RMSE = 0.136, this neural network was not able outperform the GBM by having a slightly worse RMSE.