House price

ForozanHeidaryan

2022-12-03

Understanding the Business Question

Price Recommendation for house Data inspection

Dataset Description

File descriptions

data_description.txt - full description of each column, originally prepared by Dean De Cock

but lightly edited to match the column names used here

sample_submission.csv - a benchmark submission from a linear regression on year and month of sale, lot square footage, and number of bedrooms

Data fields

Here's a brief version of what you'll find in the data description file.

SalePrice : the property's sale price in dollars. This is the target variable that you're

trying to predict.

MSSubClass: The building class

MSZoning: The general zoning classification

LotFrontage: Linear feet of street connected to property

LotArea: Lot size in square feet

Street: Type of road access

Alley: Type of alley access

LotShape: General shape of property

LandContour: Flatness of the property

Utilities: Type of utilities available

LotConfig: Lot configuration

LandSlope: Slope of property

Neighborhood: Physical locations within Ames city limits

Condition1: Proximity to main road or railroad

Condition2: Proximity to main road or railroad (if a second is present)

BldgType: Type of dwelling

HouseStyle: Style of dwelling

OverallQual: Overall material and finish quality

OverallCond: Overall condition rating

YearBuilt: Original construction date

YearRemodAdd: Remodel date

RoofStyle: Type of roof

RoofMatl: Roof material

Exterior1st: Exterior covering on house

Exterior2nd: Exterior covering on house (if more than one material)

MasVnrType: Masonry veneer type

MasVnrArea: Masonry veneer area in square feet

ExterQual: Exterior material quality

ExterCond: Present condition of the material on the exterior

Foundation: Type of foundation

BsmtQual: Height of the basement

BsmtCond: General condition of the basement

BsmtExposure: Walkout or garden level basement walls

BsmtFinType1: Quality of basement finished area

BsmtFinSF1: Type 1 finished square feet

BsmtFinType2: Quality of second finished area (if present)

BsmtFinSF2: Type 2 finished square feet

BsmtUnfSF: Unfinished square feet of basement area

TotalBsmtSF: Total square feet of basement area

Heating: Type of heating

HeatingQC: Heating quality and condition

CentralAir: Central air conditioning

Electrical: Electrical system

1stFlrSF: First Floor square feet

2ndFlrSF: Second floor square feet

LowQualFinSF: Low quality finished square feet (all floors)

GrLivArea: Above grade (ground) living area square feet

BsmtFullBath: Basement full bathrooms

BsmtHalfBath: Basement half bathrooms

FullBath: Full bathrooms above grade

HalfBath: Half baths above grade

Bedroom: Number of bedrooms above basement level

Kitchen: Number of kitchens

KitchenQual: Kitchen quality

TotRmsAbvGrd: Total rooms above grade (does not include bathrooms)

Functional: Home functionality rating

Fireplaces: Number of fireplaces

FireplaceQu: Fireplace quality

GarageType: Garage location

GarageYrBlt: Year garage was built

GarageFinish: Interior finish of the garage

GarageCars: Size of garage in car capacity

GarageArea: Size of garage in square feet

GarageQual: Garage quality

GarageCond: Garage condition

PavedDrive: Paved driveway

WoodDeckSF: Wood deck area in square feet

OpenPorchSF: Open porch area in square feet

EnclosedPorch:Enclosed porch area in square feet

3SsnPorch: Three season porch area in square feet

ScreenPorch: Screen porch area in square feet

PoolArea: Pool area in square feet

PoolQC: Pool quality

Fence: Fence quality

MiscFeature: Miscellaneous feature not covered in other categories

MiscVal: $Value of miscellaneous feature

MoSold: Month Sold

YrSold: Year Sold

SaleType: Type of sale

SaleCondition:Condition of sale

read data from file

```
train <- read.csv("train.csv" , header = TRUE )
```

A summary of the data

```
dim(train)
```

```
## [1] 1460   81
```

```
length(unique(train$Id))
```

```
## [1] 1460
```

```
summary(train)
```

```
train <- read.csv("train.csv" , header = TRUE )
```

```
##        Id          MSSubClass       MSZoning         LotFrontage
## Min.   :   1.0   Min.   : 20.0   Length:1460      Min.   : 21.00
## 1st Qu.: 365.8   1st Qu.: 20.0   Class :character   1st Qu.: 59.00
## Median : 730.5   Median : 50.0   Mode  :character   Median : 69.00
## Mean   : 730.5   Mean   : 56.9                      Mean   : 70.05
## 3rd Qu.:1095.2   3rd Qu.: 70.0                      3rd Qu.: 80.00
## Max.   :1460.0   Max.   :190.0                      Max.   :313.00
##                                                     NA's   :259
##     LotArea         Street           Alley           LotShape
## Min.   :  1300   Length:1460      Length:1460      Length:1460
## 1st Qu.:  7554   Class :character   Class :character   Class :character
## Median :  9478   Mode  :character   Mode  :character   Mode  :character
## Mean   : 10517
## 3rd Qu.: 11602
## Max.   :215245
##
##   LandContour        Utilities        LotConfig        LandSlope
## Length:1460        Length:1460      Length:1460      Length:1460
## Class :character   Class :character   Class :character   Class :character
## Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##   Neighborhood      Condition1       Condition2        BldgType
## Length:1460        Length:1460      Length:1460      Length:1460
## Class :character   Class :character   Class :character   Class :character
## Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##   HouseStyle        OverallQual      OverallCond      YearBuilt
## Length:1460        Min.   : 1.000   Min.   :1.000    Min.   :1872
## Class :character   1st Qu.: 5.000   1st Qu.:5.000    1st Qu.:1954
## Mode  :character   Median : 6.000   Median :5.000    Median :1973
##                    Mean   : 6.099   Mean   :5.575    Mean   :1971
##                    3rd Qu.: 7.000   3rd Qu.:6.000    3rd Qu.:2000
##                    Max.   :10.000   Max.   :9.000    Max.   :2010
##
##   YearRemodAdd     RoofStyle         RoofMatl        Exterior1st
## Min.   :1950     Length:1460      Length:1460      Length:1460
## 1st Qu.:1967     Class :character   Class :character   Class :character
## Median :1994     Mode  :character   Mode  :character   Mode  :character
## Mean   :1985
## 3rd Qu.:2004
## Max.   :2010
##
##   Exterior2nd       MasVnrType         MasVnrArea       ExterQual
## Length:1460        Length:1460      Min.   :   0.0   Length:1460
## Class :character   Class :character   1st Qu.:   0.0   Class :character
## Mode  :character   Mode  :character   Median :   0.0   Mode  :character
##                                      Mean   : 103.7
##                                      3rd Qu.: 166.0
##                                      Max.   :1600.0
##                                      NA's   :8
##   ExterCond         Foundation        BsmtQual         BsmtCond
## Length:1460        Length:1460      Length:1460      Length:1460
## Class :character   Class :character   Class :character   Class :character
## Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##   BsmtExposure      BsmtFinType1      BsmtFinSF1       BsmtFinType2
## Length:1460        Length:1460      Min.   :   0.0   Length:1460
## Class :character   Class :character   1st Qu.:   0.0   Class :character
## Mode  :character   Mode  :character   Median : 383.5   Mode  :character
##                                      Mean   : 443.6
##                                      3rd Qu.: 712.2
##                                      Max.   :5644.0
##
##   BsmtFinSF2        BsmtUnfSF        TotalBsmtSF       Heating
## Min.   :   0.00   Min.   :   0.0   Min.   :   0.0   Length:1460
## 1st Qu.:   0.00   1st Qu.: 223.0   1st Qu.: 795.8   Class :character
## Median :   0.00   Median : 477.5   Median : 991.5   Mode  :character
## Mean   :  46.55   Mean   : 567.2   Mean   :1057.4
## 3rd Qu.:   0.00   3rd Qu.: 808.0   3rd Qu.:1298.2
```

```
##   Max.   :1474.00   Max.   :2336.0   Max.   :6110.0
## 
##   HeatingQC         CentralAir         Electrical         X1stFlrSF
## Length:1460        Length:1460        Length:1460        Min.   : 334
## Class :character   Class :character   Class :character   1st Qu.: 882
## Mode  :character   Mode  :character   Mode  :character   Median :1087
##                                                          Mean   :1163
##                                                          3rd Qu.:1391
##                                                          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
## Min.   :0.000   Length:1460        Min.   : 2.000   Length:1460
## 1st Qu.:1.000   Class :character   1st Qu.: 5.000   Class :character
## Median :1.000   Mode  :character   Median : 6.000   Mode  :character
## Mean   :1.047                      Mean   : 6.518
## 3rd Qu.:1.000                      3rd Qu.: 7.000
## Max.   :3.000                      Max.   :14.000
## 
##   Fireplaces     FireplaceQu         GarageType        GarageYrBlt
## Min.   :0.000   Length:1460        Length:1460        Min.   :1900
## 1st Qu.:0.000   Class :character   Class :character   1st Qu.:1961
## Median :1.000   Mode  :character   Mode  :character   Median :1980
## Mean   :0.613                                         Mean   :1979
## 3rd Qu.:1.000                                         3rd Qu.:2002
## Max.   :3.000                                         Max.   :2010
##                                                       NA's   :81
## GarageFinish        GarageCars      GarageArea      GarageQual
## Length:1460        Min.   :0.000   Min.   :   0.0   Length:1460
## Class :character   1st Qu.:1.000   1st Qu.: 334.5   Class :character
## Mode  :character   Median :2.000   Median : 480.0   Mode  :character
##                    Mean   :1.767   Mean   : 473.0
##                    3rd Qu.:2.000   3rd Qu.: 576.0
##                    Max.   :4.000   Max.   :1418.0
## 
##   GarageCond         PavedDrive        WoodDeckSF       OpenPorchSF
## Length:1460        Length:1460        Min.   :  0.00   Min.   :  0.00
## Class :character   Class :character   1st Qu.:  0.00   1st Qu.:  0.00
## Mode  :character   Mode  :character   Median :  0.00   Median : 25.00
##                                       Mean   : 94.24   Mean   : 46.66
##                                       3rd Qu.:168.00   3rd Qu.: 68.00
##                                       Max.   :857.00   Max.   :547.00
## 
##  EnclosedPorch     X3SsnPorch       ScreenPorch        PoolArea
## Min.   :  0.00   Min.   :  0.00   Min.   :  0.00   Min.   :  0.000
## 1st Qu.:  0.00   1st Qu.:  0.00   1st Qu.:  0.00   1st Qu.:  0.000
## Median :  0.00   Median :  0.00   Median :  0.00   Median :  0.000
## Mean   : 21.95   Mean   :  3.41   Mean   : 15.06   Mean   :  2.759
## 3rd Qu.:  0.00   3rd Qu.:  0.00   3rd Qu.:  0.00   3rd Qu.:  0.000
## Max.   :552.00   Max.   :508.00   Max.   :480.00   Max.   :738.000
## 
##    PoolQC             Fence            MiscFeature         MiscVal
## Length:1460        Length:1460        Length:1460        Min.   :    0.00
## Class :character   Class :character   Class :character   1st Qu.:    0.00
## Mode  :character   Mode  :character   Mode  :character   Median :    0.00
##                                                          Mean   :   43.49
##                                                          3rd Qu.:    0.00
##                                                          Max.   :15500.00
## 
##     MoSold           YrSold         SaleType         SaleCondition
## Min.   : 1.000   Min.   :2006   Length:1460        Length:1460
## 1st Qu.: 5.000   1st Qu.:2007   Class :character   Class :character
## Median : 6.000   Median :2008   Mode  :character   Mode  :character
## Mean   : 6.322   Mean   :2008
```

```
##   3rd Qu.: 8.000   3rd Qu.:2009
##   Max.   :12.000   Max.   :2010
##
##     SalePrice
##   Min.   : 34900
##   1st Qu.:129975
##   Median :163000
##   Mean   :180921
##   3rd Qu.:214000
##   Max.   :755000
##
```

Convert categorical variables to factor

```
cat_var <- c("MSZoning" , "Street" , "Alley" ,"LotShape" , "LandContour", "Utilities" , "LotConfig" , "LandSlope" ,
             "Neighborhood" , "Condition1" , "Condition2" , "BldgType" ,
             "HouseStyle", "OverallQual", "OverallCond" , "RoofStyle" , "Exterior1st" , "Exterior2nd" , "MasVnrType",
             "ExterQual" ,"ExterCond" , "Foundation" , "BsmtQual" ,"BsmtCond" ,
             "BsmtExposure" ,"BsmtFinType1" ,"BsmtFinType2" , "Heating" , "HeatingQC" , "CentralAir" , "Electrical" ,
             "KitchenQual" , "Functional" , "Fireplaces" , "FireplaceQu" ,
             "GarageType" , "GarageFinish" , "GarageCars" , "GarageQual" ,"GarageCond" , "PavedDrive" , "PoolQC",
             "Fence","RoofMatl" ,
             "MiscFeature" , "SaleType" , "SaleCondition")
```

```
train[,cat_var]  <-  lapply(train[,cat_var] , factor)
knitr::kable(summary(train))
```

Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	LotConfig	LandSlope	Nei
Min. : 1.0	Min. : 20.0	C (all): 10	Min. : 21.00	Min. : 1300	Grvl: 6	Grvl: 50	IR1:484	Bnk: 63	AllPub:1459	Corner : 263	Gtl:1382	NA
1st Qu.: 365.8	1st Qu.: 20.0	FV : 65	1st Qu.: 59.00	1st Qu.: 7554	Pave:1454	Pave: 41	IR2: 41	HLS: 50	NoSeWa: 1	CulDSac: 94	Mod: 65	Col
Median : 730.5	Median : 50.0	RH : 16	Median : 69.00	Median : 9478	NA	NA's:1369	IR3: 10	Low: 36	NA	FR2 : 47	Sev: 13	Old
Mean : 730.5	Mean : 56.9	RL :1151	Mean : 70.05	Mean : 10517	NA	NA	Reg:925	Lvl:1311	NA	FR3 : 4	NA	Edv
3rd Qu.:1095.2	3rd Qu.: 70.0	RM : 218	3rd Qu.: 80.00	3rd Qu.: 11602	NA	NA	NA	NA	NA	Inside :1052	NA	Sor
Max. :1460.0	Max. :190.0	NA	Max. :313.00	Max. :215245	NA	NA	NA	NA	NA	NA	NA	Gill
NA	NA	NA	NA's :259	NA	NA	NA	NA	NA	NA	NA	NA	(Ot

```
knitr::kable(str(train))
```

```
## 'data.frame':    1460 obs. of  81 variables:
##  $ Id           : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ MSSubClass   : int  60 20 60 70 60 50 20 60 50 190 ...
##  $ MSZoning     : Factor w/ 5 levels "C (all)","FV",..: 4 4 4 4 4 4 4 4 5 4 ...
##  $ LotFrontage  : int  65 80 68 60 84 85 75 NA 51 50 ...
##  $ LotArea      : int  8450 9600 11250 9550 14260 14115 10084 10382 6120 7420 ...
##  $ Street       : Factor w/ 2 levels "Grvl","Pave": 2 2 2 2 2 2 2 2 2 2 ...
##  $ Alley        : Factor w/ 2 levels "Grvl","Pave": NA NA NA NA NA NA NA NA NA NA ...
##  $ LotShape     : Factor w/ 4 levels "IR1","IR2","IR3",..: 4 4 1 1 1 1 4 1 4 4 ...
##  $ LandContour  : Factor w/ 4 levels "Bnk","HLS","Low",..: 4 4 4 4 4 4 4 4 4 4 ...
##  $ Utilities    : Factor w/ 2 levels "AllPub","NoSeWa": 1 1 1 1 1 1 1 1 1 1 ...
##  $ LotConfig    : Factor w/ 5 levels "Corner","CulDSac",..: 5 3 5 1 3 5 5 1 5 1 ...
##  $ LandSlope    : Factor w/ 3 levels "Gtl","Mod","Sev": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Neighborhood : Factor w/ 25 levels "Blmngtn","Blueste",..: 6 25 6 7 14 12 21 17 18 4 ...
##  $ Condition1   : Factor w/ 9 levels "Artery","Feedr",..: 3 2 3 3 3 3 3 5 1 1 ...
##  $ Condition2   : Factor w/ 8 levels "Artery","Feedr",..: 3 3 3 3 3 3 3 3 3 1 ...
##  $ BldgType     : Factor w/ 5 levels "1Fam","2fmCon",..: 1 1 1 1 1 1 1 1 1 2 ...
##  $ HouseStyle   : Factor w/ 8 levels "1.5Fin","1.5Unf",..: 6 3 6 6 6 1 3 6 1 2 ...
##  $ OverallQual  : Factor w/ 10 levels "1","2","3","4",..: 7 6 7 7 8 5 8 7 7 5 ...
##  $ OverallCond  : Factor w/ 9 levels "1","2","3","4",..: 5 8 5 5 5 5 5 6 5 6 ...
##  $ YearBuilt    : int  2003 1976 2001 1915 2000 1993 2004 1973 1931 1939 ...
##  $ YearRemodAdd : int  2003 1976 2002 1970 2000 1995 2005 1973 1950 1950 ...
##  $ RoofStyle    : Factor w/ 6 levels "Flat","Gable",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ RoofMatl     : Factor w/ 8 levels "ClyTile","CompShg",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ Exterior1st  : Factor w/ 15 levels "AsbShng","AsphShn",..: 13 9 13 14 13 13 13 7 4 9 ...
##  $ Exterior2nd  : Factor w/ 16 levels "AsbShng","AsphShn",..: 14 9 14 16 14 14 14 7 16 9 ...
##  $ MasVnrType   : Factor w/ 4 levels "BrkCmn","BrkFace",..: 2 3 2 3 2 3 4 4 3 3 ...
##  $ MasVnrArea   : int  196 0 162 0 350 0 186 240 0 0 ...
##  $ ExterQual    : Factor w/ 4 levels "Ex","Fa","Gd",..: 3 4 3 4 3 4 3 4 4 4 ...
##  $ ExterCond    : Factor w/ 5 levels "Ex","Fa","Gd",..: 5 5 5 5 5 5 5 5 5 5 ...
##  $ Foundation   : Factor w/ 6 levels "BrkTil","CBlock",..: 3 2 3 1 3 6 3 2 1 1 ...
##  $ BsmtQual     : Factor w/ 4 levels "Ex","Fa","Gd",..: 3 3 3 4 3 3 1 3 4 4 ...
##  $ BsmtCond     : Factor w/ 4 levels "Fa","Gd","Po",..: 4 4 4 2 4 4 4 4 4 4 ...
##  $ BsmtExposure : Factor w/ 4 levels "Av","Gd","Mn",..: 4 2 3 4 1 4 1 3 4 4 ...
##  $ BsmtFinType1 : Factor w/ 6 levels "ALQ","BLQ","GLQ",..: 3 1 3 1 3 3 3 1 6 3 ...
##  $ BsmtFinSF1   : int  706 978 486 216 655 732 1369 859 0 851 ...
##  $ BsmtFinType2 : Factor w/ 6 levels "ALQ","BLQ","GLQ",..: 6 6 6 6 6 6 6 2 6 6 ...
##  $ BsmtFinSF2   : int  0 0 0 0 0 0 0 32 0 0 ...
##  $ BsmtUnfSF    : int  150 284 434 540 490 64 317 216 952 140 ...
##  $ TotalBsmtSF  : int  856 1262 920 756 1145 796 1686 1107 952 991 ...
##  $ Heating      : Factor w/ 6 levels "Floor","GasA",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ HeatingQC    : Factor w/ 5 levels "Ex","Fa","Gd",..: 1 1 1 3 1 1 1 1 3 1 ...
##  $ CentralAir   : Factor w/ 2 levels "N","Y": 2 2 2 2 2 2 2 2 2 2 ...
##  $ Electrical   : Factor w/ 5 levels "FuseA","FuseF",..: 5 5 5 5 5 5 5 5 2 5 ...
##  $ X1stFlrSF    : int  856 1262 920 961 1145 796 1694 1107 1022 1077 ...
##  $ X2ndFlrSF    : int  854 0 866 756 1053 566 0 983 752 0 ...
##  $ LowQualFinSF : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ GrLivArea    : int  1710 1262 1786 1717 2198 1362 1694 2090 1774 1077 ...
##  $ BsmtFullBath : int  1 0 1 1 1 1 1 1 0 1 ...
##  $ BsmtHalfBath : int  0 1 0 0 0 0 0 0 0 0 ...
##  $ FullBath     : int  2 2 2 1 2 1 2 2 2 1 ...
##  $ HalfBath     : int  1 0 1 0 1 1 0 1 0 0 ...
##  $ BedroomAbvGr : int  3 3 3 3 4 1 3 3 2 2 ...
##  $ KitchenAbvGr : int  1 1 1 1 1 1 1 1 2 2 ...
##  $ KitchenQual  : Factor w/ 4 levels "Ex","Fa","Gd",..: 3 4 3 3 3 4 3 4 4 4 ...
##  $ TotRmsAbvGrd : int  8 6 6 7 9 5 7 7 8 5 ...
##  $ Functional   : Factor w/ 7 levels "Maj1","Maj2",..: 7 7 7 7 7 7 7 7 3 7 ...
##  $ Fireplaces   : Factor w/ 4 levels "0","1","2","3": 1 2 2 2 2 1 2 3 3 3 ...
##  $ FireplaceQu  : Factor w/ 5 levels "Ex","Fa","Gd",..: NA 5 5 3 5 NA 3 5 5 5 ...
##  $ GarageType   : Factor w/ 6 levels "2Types","Attchd",..: 2 2 2 6 2 2 2 2 6 2 ...
##  $ GarageYrBlt  : int  2003 1976 2001 1998 2000 1993 2004 1973 1931 1939 ...
##  $ GarageFinish : Factor w/ 3 levels "Fin","RFn","Unf": 2 2 2 3 2 3 2 2 3 2 ...
##  $ GarageCars   : Factor w/ 5 levels "0","1","2","3",..: 3 3 3 4 4 3 3 3 3 2 ...
##  $ GarageArea   : int  548 460 608 642 836 480 636 484 468 205 ...
##  $ GarageQual   : Factor w/ 5 levels "Ex","Fa","Gd",..: 5 5 5 5 5 5 5 5 2 3 ...
##  $ GarageCond   : Factor w/ 5 levels "Ex","Fa","Gd",..: 5 5 5 5 5 5 5 5 5 5 ...
##  $ PavedDrive   : Factor w/ 3 levels "N","P","Y": 3 3 3 3 3 3 3 3 3 3 ...
##  $ WoodDeckSF   : int  0 298 0 0 192 40 255 235 90 0 ...
##  $ OpenPorchSF  : int  61 0 42 35 84 30 57 204 0 4 ...
##  $ EnclosedPorch: int  0 0 0 272 0 0 0 228 205 0 ...
##  $ X3SsnPorch   : int  0 0 0 0 0 320 0 0 0 0 ...
##  $ ScreenPorch  : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ PoolArea     : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ PoolQC       : Factor w/ 3 levels "Ex","Fa","Gd": NA NA NA NA NA NA NA NA NA NA ...
##  $ Fence        : Factor w/ 4 levels "GdPrv","GdWo",..: NA NA NA NA NA 3 NA NA NA NA ...
##  $ MiscFeature  : Factor w/ 4 levels "Gar2","Othr",..: NA NA NA NA NA 3 NA 3 NA NA ...
##  $ MiscVal      : int  0 0 0 0 0 700 0 350 0 0 ...
##  $ MoSold       : int  2 5 9 2 12 10 8 11 4 1 ...
```

```
## $ YrSold      : int  2008 2007 2008 2006 2008 2009 2007 2009 2008 2008 ...
## $ SaleType    : Factor w/ 9 levels "COD","Con","ConLD",..: 9 9 9 9 9 9 9 9 9 9 ...
## $ SaleCondition: Factor w/ 6 levels "Abnorml","AdjLand",..: 5 5 5 1 5 5 5 5 1 5 ...
## $ SalePrice   : int  208500 181500 223500 140000 250000 143000 307000 200000 129900 118000 ...
```

Identify missing data

```
mv_summary2      <-  data.frame('variables name' = colnames(train))
mv_summary2$freq  <- apply(train , 2 , function(x)  sum(is.na(x)))
mv_summary2$pers  <- round(mv_summary2$freq / nrow(train) , 3) * 100
mv_summary2_1     <- as.data.frame(mv_summary2[mv_summary2$pers > 0 & mv_summary2$pers <10, ])
mv_summary2_2     <- as.data.frame(mv_summary2[mv_summary2$pers > 10 , ])
knitr::kable(mv_summary2_1)
```

	variables.name	freq	pers
26	MasVnrType	8	0.5
27	MasVnrArea	8	0.5
31	BsmtQual	37	2.5
32	BsmtCond	37	2.5
33	BsmtExposure	38	2.6
34	BsmtFinType1	37	2.5
36	BsmtFinType2	38	2.6
43	Electrical	1	0.1
59	GarageType	81	5.5
60	GarageYrBlt	81	5.5
61	GarageFinish	81	5.5
64	GarageQual	81	5.5
65	GarageCond	81	5.5

```
knitr::kable(mv_summary2_2)
```

	variables.name	freq	pers
4	LotFrontage	259	17.7
7	Alley	1369	93.8
58	FireplaceQu	690	47.3
73	PoolQC	1453	99.5
74	Fence	1179	80.8
75	MiscFeature	1406	96.3

Removing columns that have more than 10% missing value

```
t1 <-  train[,c("LotFrontage","Alley","FireplaceQu" ,"Fence" ,
                "MiscFeature" , "PoolQC")]
```

In my opinion, these columns have high missing values, so it causes problems in our modeling, so I preferred to remove them.

```
train1 <- train[,-which(train %in% t1)]
dim(train1)
```

```
## [1] 1460   75
```

I also removed all the missing values in the rows. This is a good solution to get rid of the missing values

```
train2 <- train1[apply(train1,1,function(x) any(is.na(x))) == F,]
```

For convenience, I divided the data into continuous and discrete parts

```
cat <- train2[,c("MSZoning" , "Street" ,"LotShape" , "LandContour", "Utilities" , "LotConfig" , "LandSlope" ,
                 "Neighborhood" , "Condition1" , "Condition2" , "BldgType" ,
                 "HouseStyle", "OverallQual", "OverallCond" , "RoofStyle" , "Exterior1st" , "Exterior2nd" ,
                 "MasVnrType"      ,
                 "ExterQual" ,"ExterCond" , "Foundation" , "BsmtQual" ,"BsmtCond" ,
                 "BsmtExposure" ,"BsmtFinType1" ,"BsmtFinType2" , "Heating" , "HeatingQC" , "CentralAir" ,
                 "Electrical"  ,
                 "KitchenQual" , "Functional" , "GarageType" , "GarageFinish" , "GarageCars" , "GarageQual" ,
                 "GarageCond" , "PavedDrive" ,
                 "RoofMatl" ,
                 "SaleType" , "SaleCondition" ,"Fireplaces" )]
```

I prefer to convert the columns that contain the date field to age, this helps me more easily determine the relationship between the price and the life of the house.

```
train3 <- train2[,-c(18,19,57,72)]
today <- as.Date("2022"  , format = "%Y")
```

```
train3$ageBuilt <- as.Date(as.character(train2$YearBuilt) , format = "%Y")

train3$ageBuilt    <- as.numeric(today - train3$ageBuilt)

train3$ageBuilt<- round(train3$ageBuilt/365 )
summary(train3$ageBuilt)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   12.00   21.00   46.00   48.97   66.00  142.00
```

```
train3$ageRemodAdd  <- as.Date(as.character(train2$YearRemodAdd), format = "%Y")

train3$ageRemodAdd <- as.numeric(today - train3$ageRemodAdd)

train3$ageRemodAdd <- round(train3$ageRemodAdd / 365 )

summary(train3$ageRemodAdd)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   12.00   18.00   27.50   36.33   54.00   72.00
```

```
train3$GarageageBlt   <- as.Date(as.character(train2$GarageYrBlt),format = "%Y")
train3$GarageageBlt   <- as.numeric(today - train3$GarageageBlt)
train3$GarageageBlt   <-round(train3$GarageageBlt / 365 )

summary(train3$GarageageBlt)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   12.0    20.0    42.0    43.4    60.0   122.0
```

```
#Regarding the sale date, it also helps me to better recognize the rise and fall of prices
train3$ageSold  <- as.Date(as.character(train2$YrSold) , format = "%Y")
train3$ageSold  <- as.numeric(today - train3$ageSold)
train3$ageSold  <- round(train3$ageSold / 365 )
summary(train3$ageSold)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   12.00   13.00   14.00   14.19   15.00   16.00
```

Regarding the sale date, it also helps me to better recognize the rise and fall of prices

```
par(mfrow = c(1,1))
hist(train$SalePrice , breaks = 50,probability = TRUE , col = "#4B0082")
lines(density(train$SalePrice) , col = "red")
```

Histogram of train$SalePrice
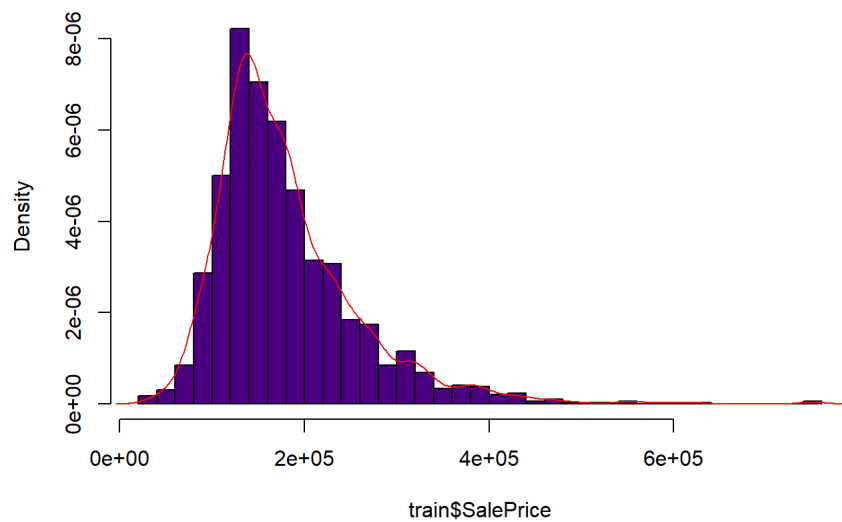

```
qqnorm(train$SalePrice , col = "#4B0082")
qqline(train$SalePrice , col = "red")
```

Normal Q-Q Plot
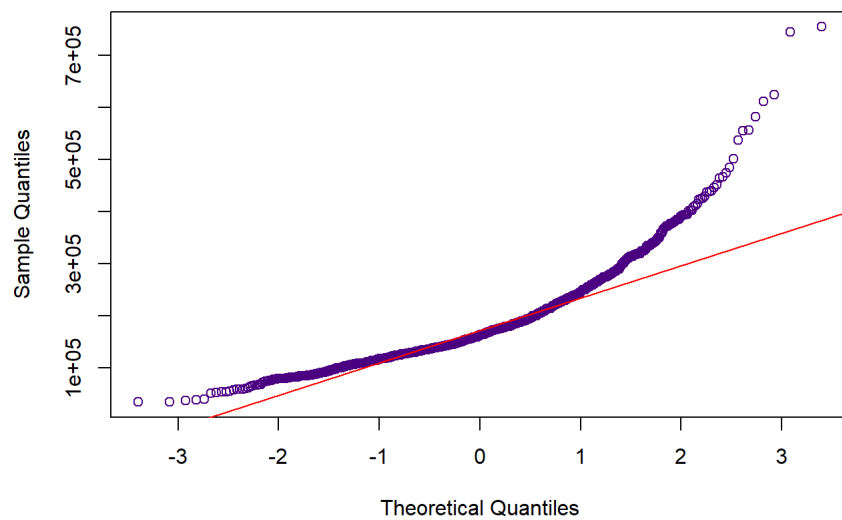

```
par(mfrow = c(1,1))
boxplot(train$SalePrice)
```
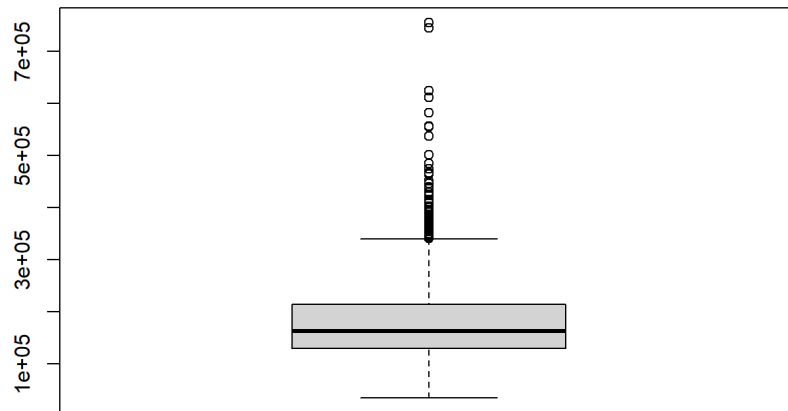
```
library("moments")

jarque.test(train$SalePrice)
```

```
##
##   Jarque-Bera Normality Test
##
## data:  train$SalePrice
## JB = 3438.9, p-value < 2.2e-16
## alternative hypothesis: greater
```

```
#pvalue < 0    -> h0 regect
```

The graphs and statistical tests indicate that the response variable does not follow a normal distribution, which makes sense to me because a smaller percentage of people are very wealthy.I would like to approximate them to a normal distribution, this may help me For this purpose, I use Box Cox conversion

```
library("MASS")
box_result <- boxcox(train$SalePrice ~ 1 , lambda = seq(-5 , 5 , 0.1))
```
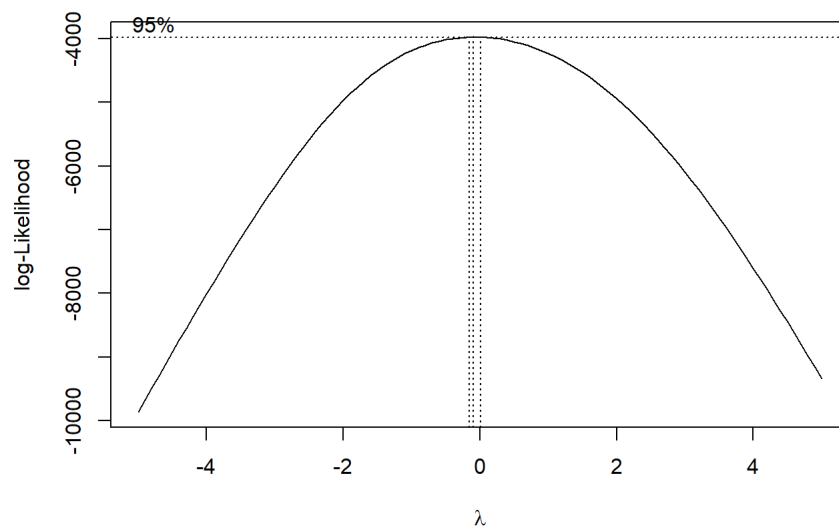
```
box_result <- data.frame(box_result)
lambda <- box_result[which(box_result$y == max(box_result$y)),]
salPrice <- log(train$SalePrice)
```

It is clear that zero is inside this confidence interval Therefore, I use logarithmic transformation

```
hist(salPrice , breaks = 50 , probability = TRUE , col = "#4B0082")
lines(density(salPrice) , col = "red")
```

Histogram of salPrice
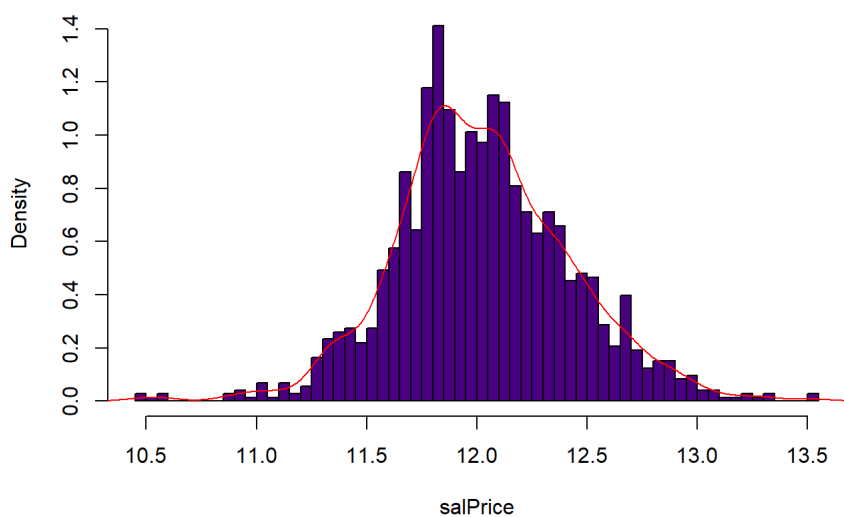

```
qqnorm(salPrice , col = "#4B0082")
qqline(salPrice , col = "red")
```

Normal Q-Q Plot
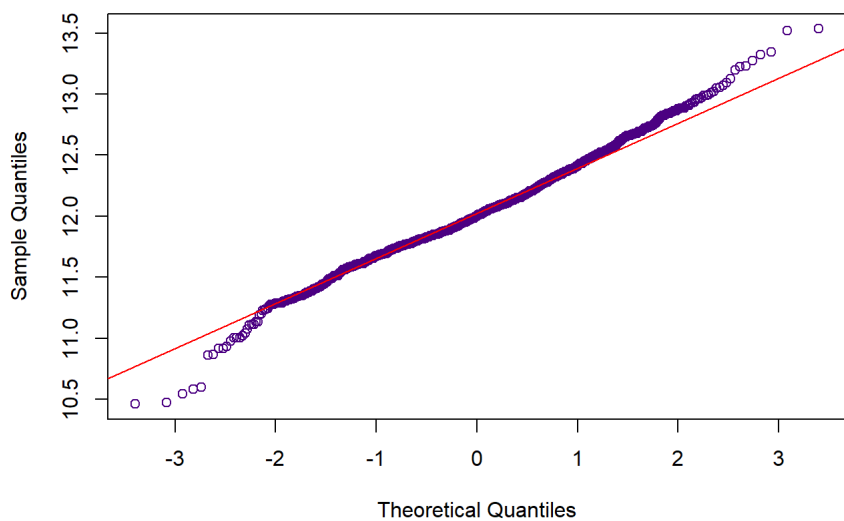

```
jarque.test(salPrice)
```

```
##
##  Jarque-Bera Normality Test
##
## data:  salPrice
## JB = 42.767, p-value = 5.167e-10
## alternative hypothesis: greater
```

```
anscombe.test(salPrice)
```

```
##
##  Anscombe-Glynn kurtosis test
##
## data:  salPrice
## kurt = 3.8027, z = 4.6820, p-value = 2.84e-06
## alternative hypothesis: kurtosis is not equal to 3
```

```
par(mfrow = c(1,1))
boxplot(train3$SalePrice ,main = "plot respont", col = "#4B0082")
```

plot respont
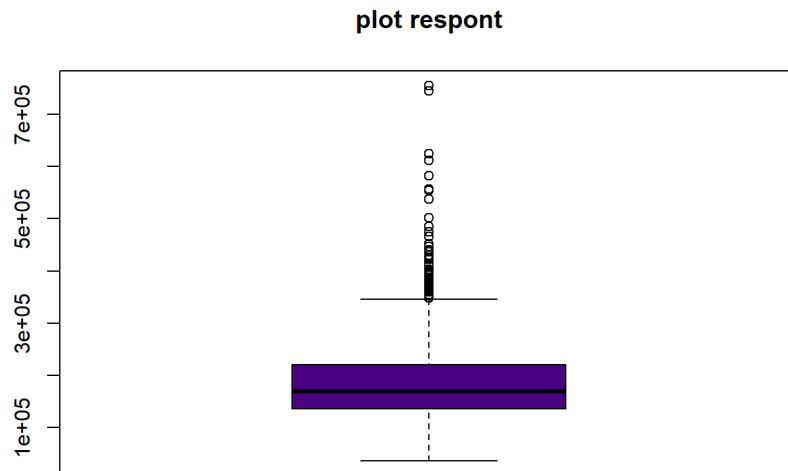
It seems that the data is far from the

normal distribution

```
tukey_u <- quantile(train3$SalePrice , probs = 0.75) + 1.5 * IQR(train3$SalePrice)

sum(train3$SalePrice > tukey_u)
```

```
## [1] 56
```

The number of continuous data is very high, so I will convert them into two parts

```
count <- train3[,-which(train3 %in% cat)]
cunt1 <- count[,c(1:17)]
cunt2 <- count[,c(18:33)]

cunt1 <- cunt1[,-1]
cunt1$SalePrice  <- cunt2$SalePrice
```

Histogram of continuous data

```
par(mar = c(2,2,2,2))
par(mfrow = c(4,4))
for (i in 1:16 ) {
  hist(cunt1[,i] , xlab =    "" ,col = "#4B0082", main = paste("HIST COUNT 1" ,
                                                 names(cunt1[i])))
}
```

HIST COUNT 1 MSSubClass   HIST COUNT 1 LotArea   HIST COUNT 1 MasVnrArea   HIST COUNT 1 BsmtFinSF1

HIST COUNT 1 BsmtFinSF2   HIST COUNT 1 BsmtUnfSF   HIST COUNT 1 TotalBsmtSF   HIST COUNT 1 X1stFlrSF

HIST COUNT 1 X2ndFlrSF   HIST COUNT 1 LowQualFinSF   HIST COUNT 1 GrLivArea   HIST COUNT 1 BsmtFullBath

HIST COUNT 1 BsmtHalfBath   HIST COUNT 1 FullBath   HIST COUNT 1 HalfBath   HIST COUNT 1 BedroomAbvG
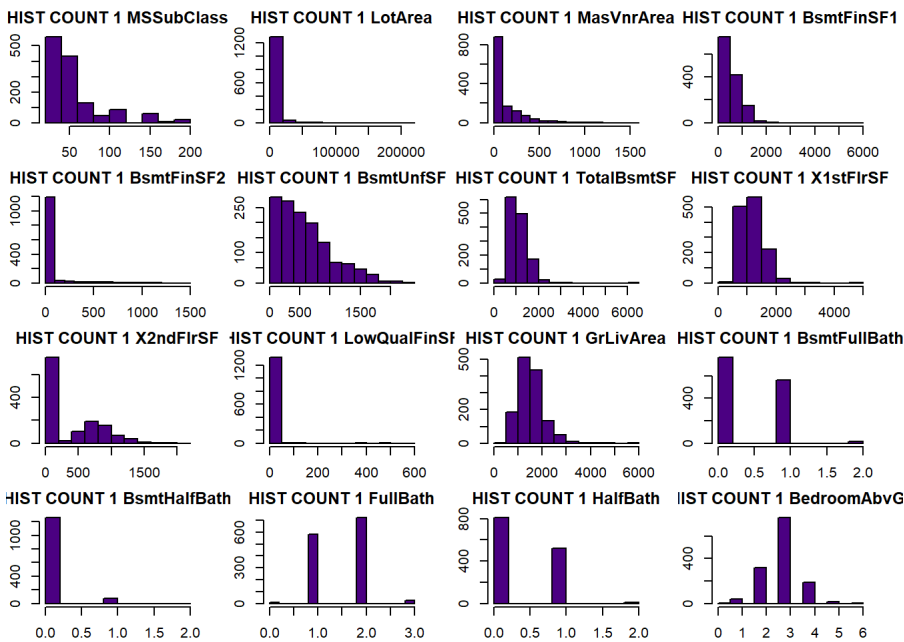
```
par(mar = c(2,2,2,2))
par(mfrow = c(4,4))
for (i in 1:16) {
  hist(cunt2[,i] , xlab = "" , col = "#4B0082" , main = paste("HIST COUNT 2" ,
                                              names(cunt1[i])))
}
```

HIST COUNT 2 MSSubClass   HIST COUNT 2 LotArea   HIST COUNT 2 MasVnrArea   HIST COUNT 2 BsmtFinSF1

HIST COUNT 2 BsmtFinSF2   HIST COUNT 2 BsmtUnfSF   HIST COUNT 2 TotalBsmtSF   HIST COUNT 2 X1stFlrSF

HIST COUNT 2 X2ndFlrSF   HIST COUNT 2 LowQualFinSF   HIST COUNT 2 GrLivArea   HIST COUNT 2 BsmtFullBath

HIST COUNT 2 BsmtHalfBath   HIST COUNT 2 FullBath   HIST COUNT 2 HalfBath   HIST COUNT 2 BedroomAbvG

Checking the correlation of continuous data with the response variable

```
cros_tab <- round(cor(count),2)

library("corrplot")
```

```
## corrplot 0.92 loaded
```

```
par(mfrow = c(1,1))
corrplot(cros_tab)
```
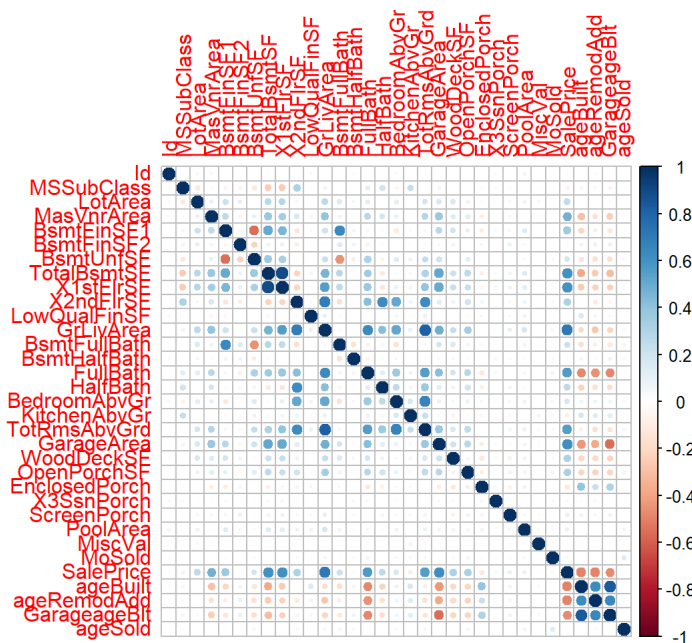
```
par(mar=c(2,2,2,2))
par(mfrow = c(4,4))
for (i in 1:16) {
  plot(cunt1[,i] , cunt2$SalePrice,xlab = "" , main = paste("plot" ,names(cunt1)[i] ))
}
```
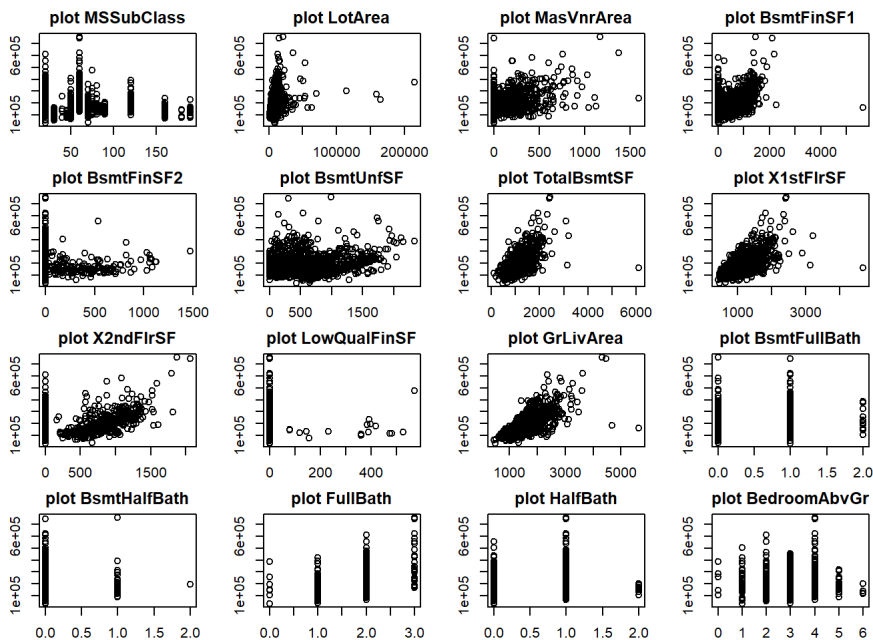


```
par(mar=c(2,2,2,2))
par(mfrow = c(4,4))
for (i in 1:16) {
  plot(cunt2[,i] , cunt2$SalePrice,xlab = "" , main = paste("plot" ,names(cunt2)[i] ))
}
```
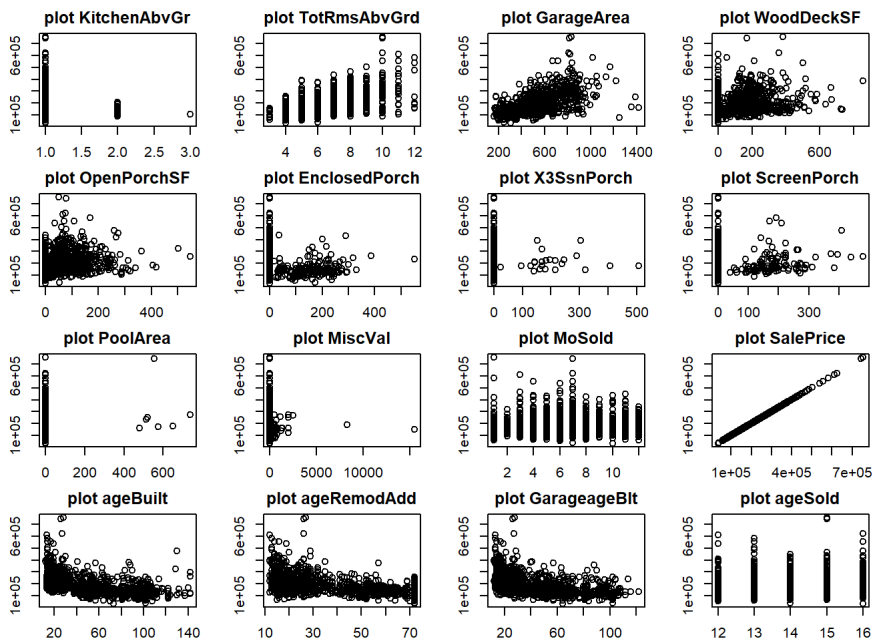
I put the columns that contain discrete variables in one file

```
categori <- train2[,which( train2 %in% cat)]
colnames(categori)
```

```
##  [1] "MSZoning"      "Street"        "LotShape"      "LandContour"
##  [5] "Utilities"     "LotConfig"     "LandSlope"     "Neighborhood"
##  [9] "Condition1"    "Condition2"    "BldgType"      "HouseStyle"
## [13] "OverallQual"   "OverallCond"   "RoofStyle"     "RoofMatl"
## [17] "Exterior1st"   "Exterior2nd"   "MasVnrType"    "ExterQual"
## [21] "ExterCond"     "Foundation"    "BsmtQual"      "BsmtCond"
## [25] "BsmtExposure"  "BsmtFinType1"  "BsmtFinType2"  "Heating"
## [29] "HeatingQC"     "CentralAir"    "Electrical"    "KitchenQual"
## [33] "Functional"    "Fireplaces"    "GarageType"    "GarageFinish"
## [37] "GarageCars"    "GarageQual"    "GarageCond"    "PavedDrive"
## [41] "SaleType"      "SaleCondition"
```

```
dim(categori)
```

```
## [1] 1338   42
```

Examining each variable with the response variable

```
table(train3$MSZoning)
```

```
##
## C (all)      FV      RH      RL      RM
##       8      62      11    1066     191
```

```
boxplot(SalePrice ~ MSZoning , data = train3 , col = "#4B0082" , border = "#7E3817")
```
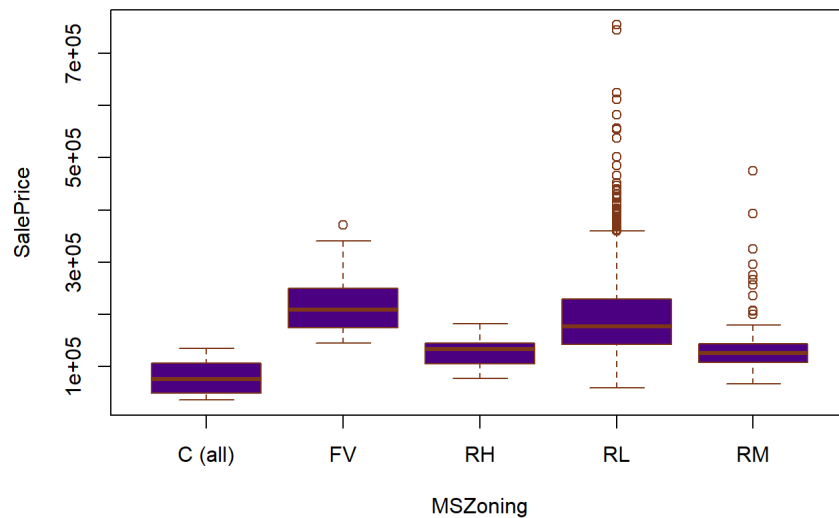
Residential Low Density seems to have the highest price

```
table(train3$Street)
```

```
##
## Grvl Pave
##    5 1333
```

```
boxplot(SalePrice ~  Street , data = train3 , col = "#4B0082" , border = "#7E3817")
```
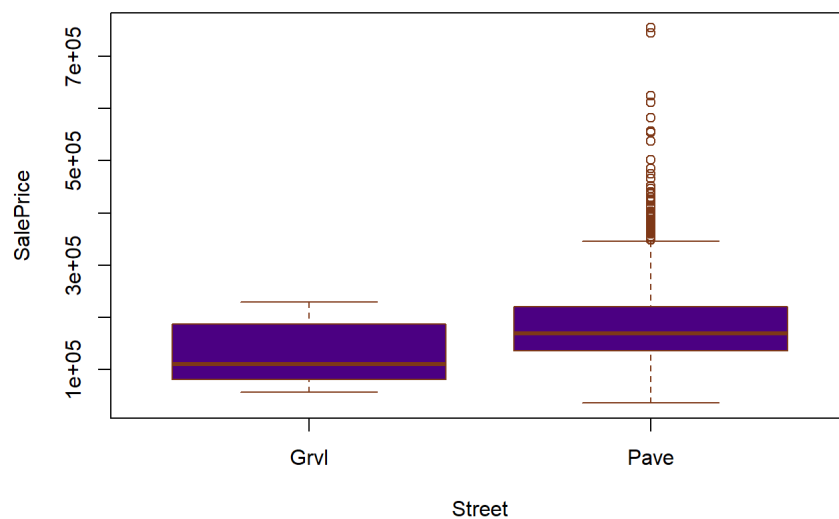

The type of Paved street has a great

effect on the price. It seems that these houses are located in a more expensive area

```
table(train3$LotShape)
```

```
##
## IR1 IR2 IR3 Reg
## 459  40  10 829
```

```
boxplot(SalePrice ~ LotShape , data = train3 , col = "#4B0082" , border = "#7E3817")
```

LotShape
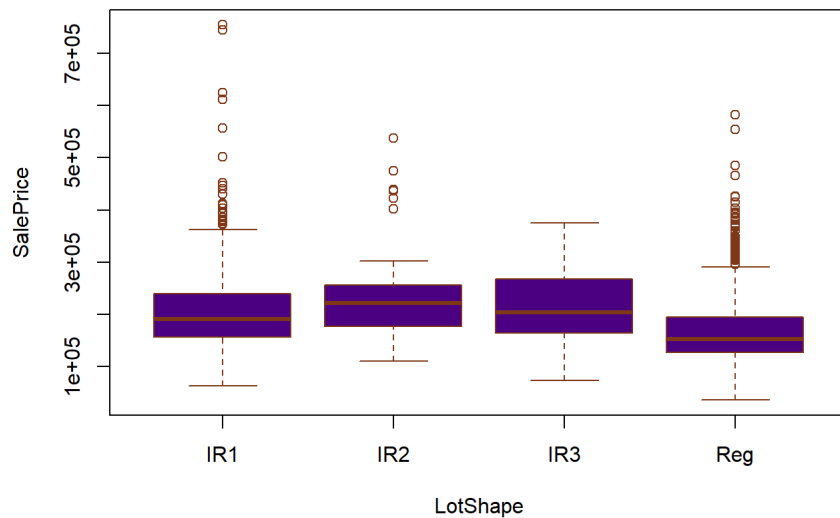
```
table(train3$LandContour)
```

```
##
##  Bnk  HLS  Low  Lvl
##   52   48   32 1206
```

```
boxplot(SalePrice ~ LandContour , data = train , col = "#4B0082" , border = "#7E3817")
```


LandContour
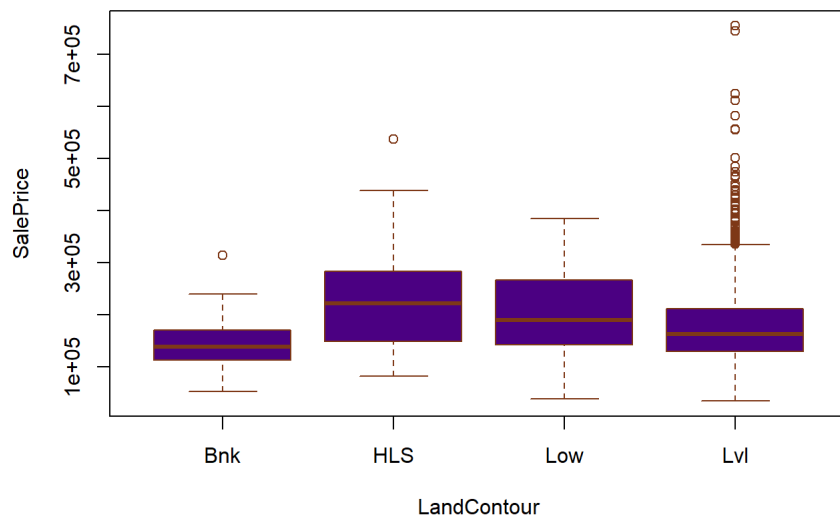
```
table(train3$Utilities)
```

```
##
## AllPub NoSeWa
##   1337      1
```

```
boxplot(SalePrice ~  Utilities , data = train3 , col = "#4B0082" , border = "#7E3817")
```
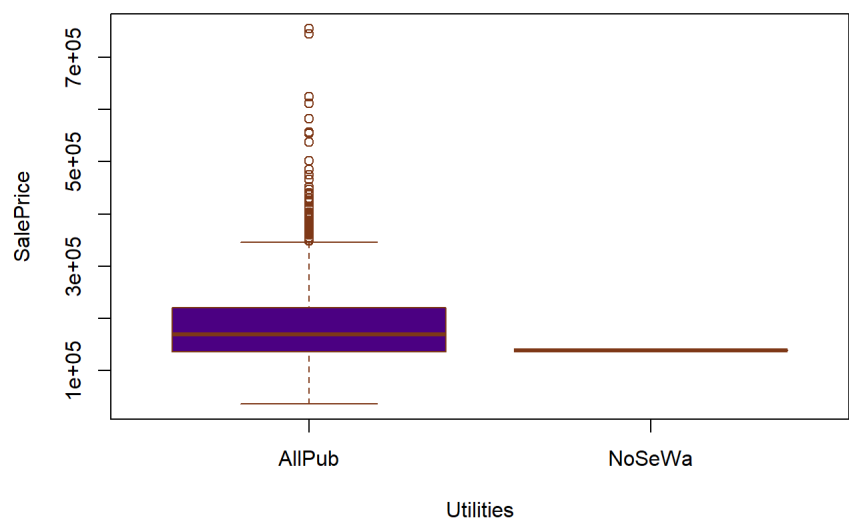
```
table(train3$LotConfig)
```

```
##
##   Corner CulDSac    FR2     FR3  Inside
##      244     90      43       4     957
```
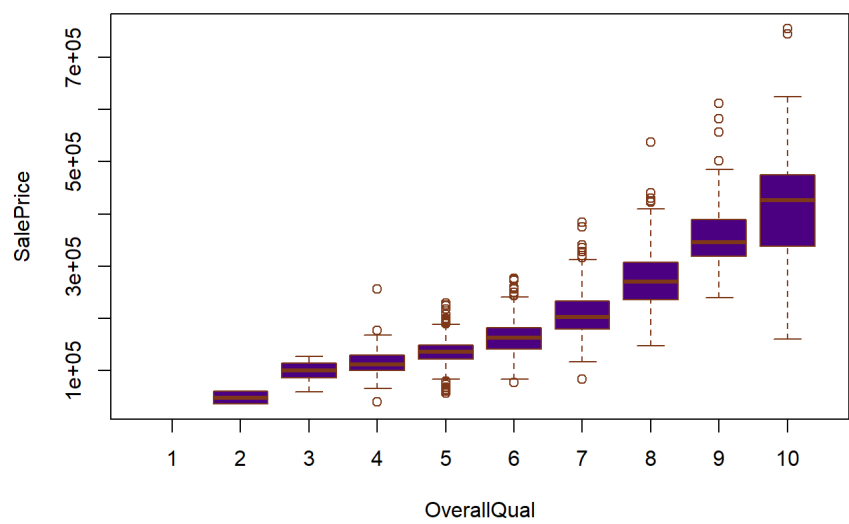
```
boxplot(SalePrice ~  LotConfig , data = train3 , col = "#4B0082" , border = "#7E3817")
```
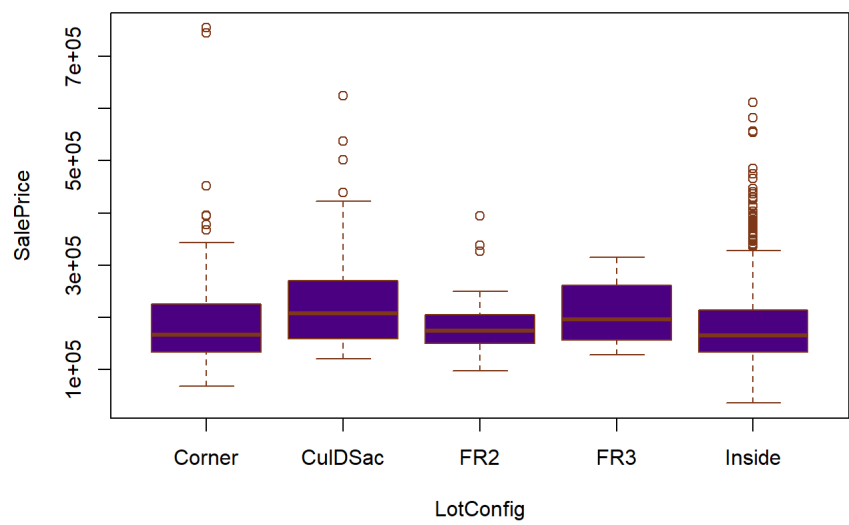


```
table(train3$LandSlope)
```

```
##
##  Gtl  Mod  Sev
## 1265   61   12
```

```
boxplot(SalePrice ~  LandSlope , data = train3 , col = "#4B0082" , border = "#7E3817")
```
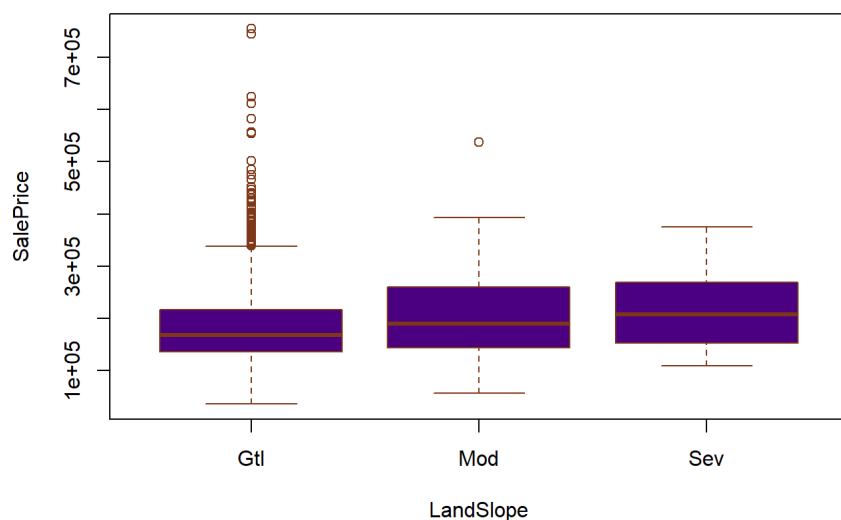
```
table(train3$Neighborhood)
```

```
##
## Blmngtn Blueste  BrDale BrkSide ClearCr CollgCr Crawfor Edwards Gilbert  IDOTRR
##      17       2      15      47      26     146      50      70      77      29
## MeadowV Mitchel   NAmes NoRidge NPkVill NridgHt  NWAmes OldTown  Sawyer SawyerW
##      12      42     209      41       9      75      73     100      69      53
## Somerst StoneBr   SWISU  Timber Veenker
##      83      25      20      37      11
```

```
boxplot(SalePrice ~ Neighborhood , data = train3 , col = "#4B0082" , border = "#7E3817")
```
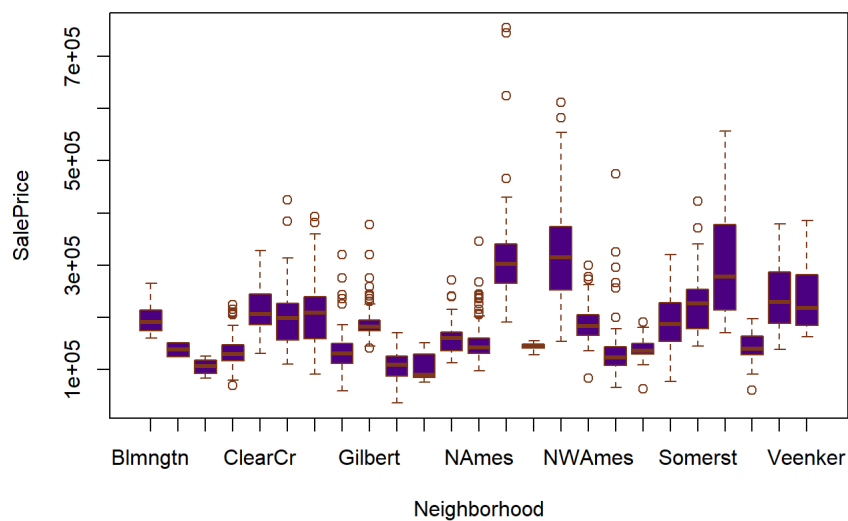


```
table(train3$Condition1)
```

```
##
## Artery  Feedr   Norm   PosA   PosN   RRAe   RRAn   RRNe   RRNn
##     43     63   1162      8     19     10     26      2      5
```

```
boxplot(SalePrice ~ Condition1 , data = train3 , col = "#4B0082" , border = "#7E3817")
```
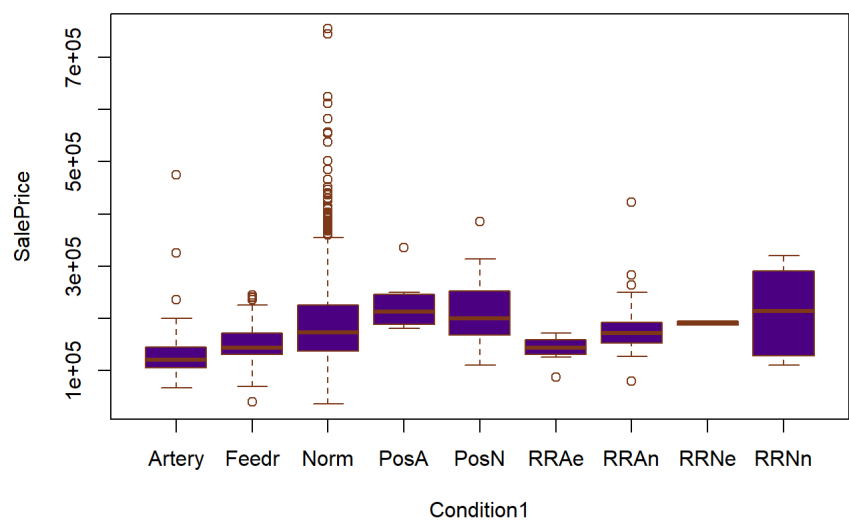
Condition1

```
table(train3$Condition2)
```

```
##
## Artery  Feedr   Norm   PosA   PosN   RRAe   RRAn   RRNn
##     2      5    1324      1      2      1      1      2
```

```
boxplot(SalePrice ~ Condition2 , data = train3 , col = "#4B0082" , border = "#7E3817")
```


Condition2
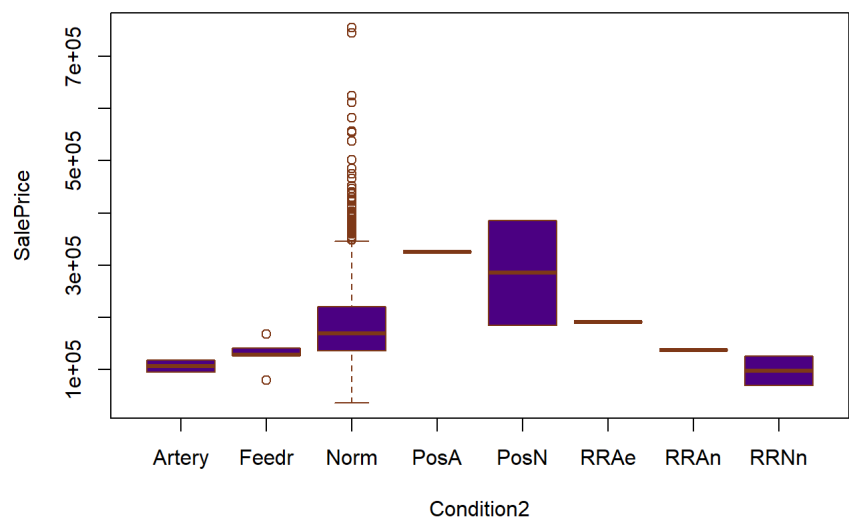
```
table(train3$BldgType)
```

```
##
##   1Fam 2fmCon Duplex  Twnhs TwnhsE
##   1138     22     28     38    112
```

```
boxplot(SalePrice ~ BldgType , data = train3 , col = "#4B0082" , border = "#7E3817")
```
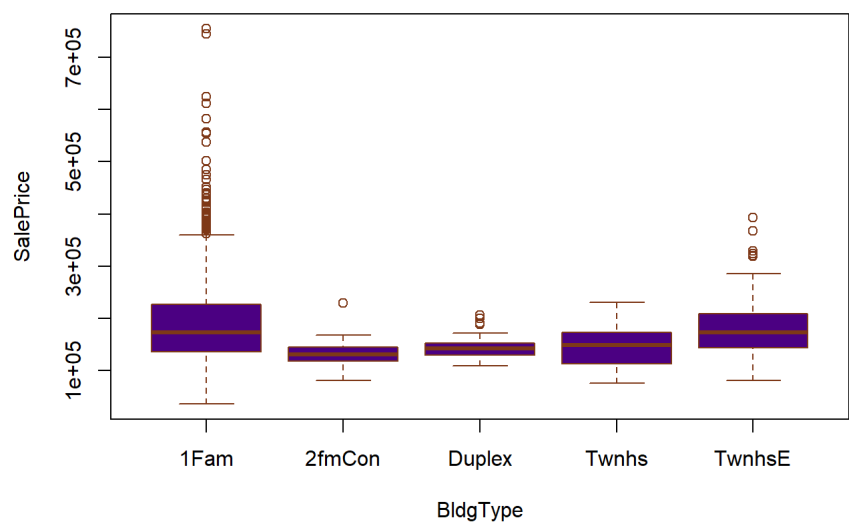
```
table(train3$HouseStyle)
```

```
##
## 1.5Fin 1.5Unf 1Story 2.5Fin 2.5Unf 2Story SFoyer   SLvl
##    134     11    657      6     10    426     30     64
```

```
boxplot(SalePrice ~ HouseStyle , data = train3 , col = "#4B0082" , border = "#7E3817")
```
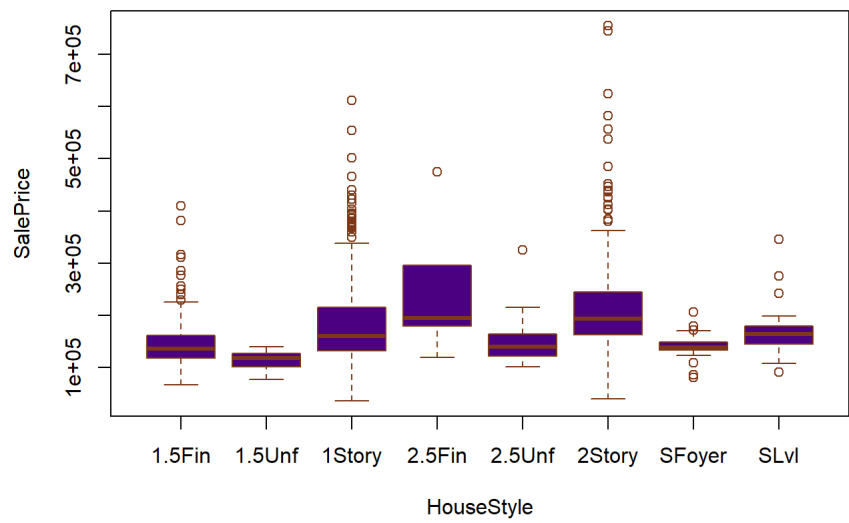


```
table(train3$OverallQual)
```

```
##
##   1   2   3   4   5   6   7   8   9  10
##   0   2   8  81 351 359 312 165  43  17
```

```
boxplot(SalePrice ~ OverallQual , data = train3 , col = "#4B0082" , border = "#7E3817")
```

```
table(train3$OverallCond)
```

```
##
##   1   2   3   4   5   6   7   8   9
##   0   3  15  46 770 233 183  68  20
```

```
boxplot(SalePrice ~ OverallCond , data = train3 , col = "#4B0082" , border = "#7E3817")
```
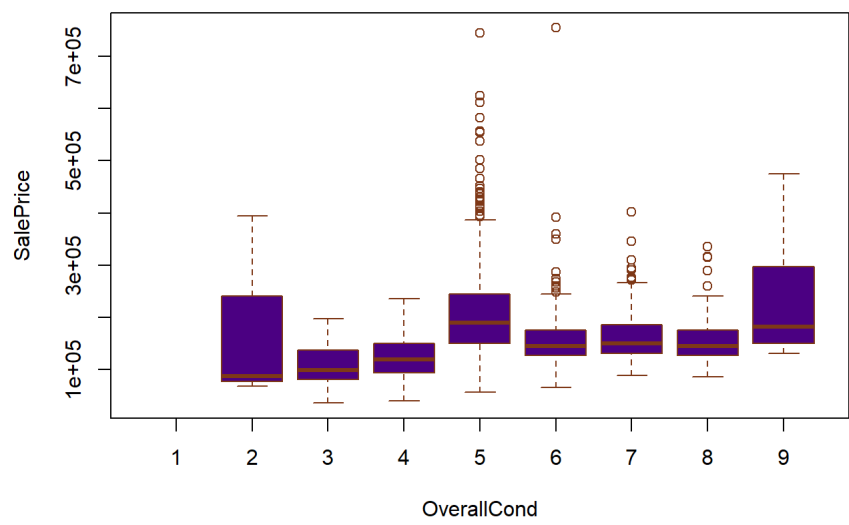


```
table(train3$RoofStyle)
```

```
##
##    Flat   Gable Gambrel     Hip Mansard    Shed
##      11    1037      10     272       6       2
```

```
boxplot(SalePrice ~ RoofStyle , data = train3 , col = "#4B0082"  , border = "#7E3817")
```
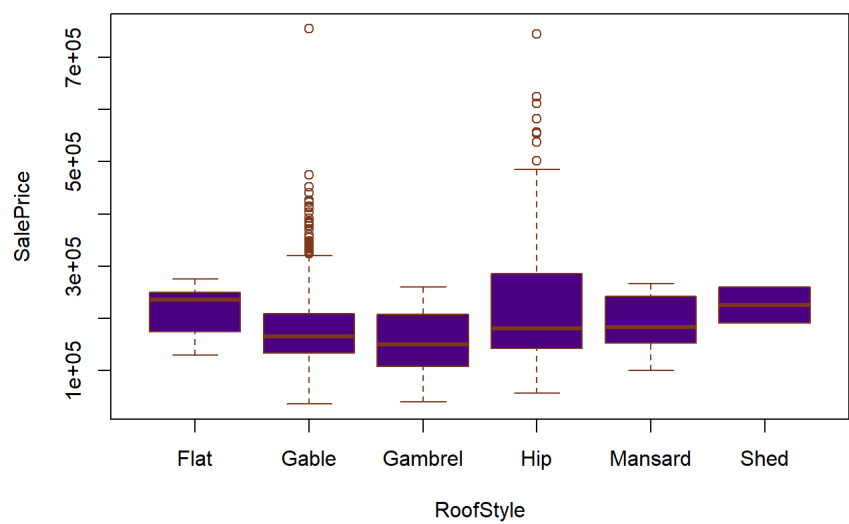
```
table(train3$RoofMatl)
```

```
##
## ClyTile CompShg Membran   Metal    Roll Tar&Grv WdShake WdShngl
##       1    1314       1       1       1       9       5       6
```

```
boxplot(SalePrice ~  RoofMatl , data = train3 , col = "#4B0082" , border = "#7E3817")
```
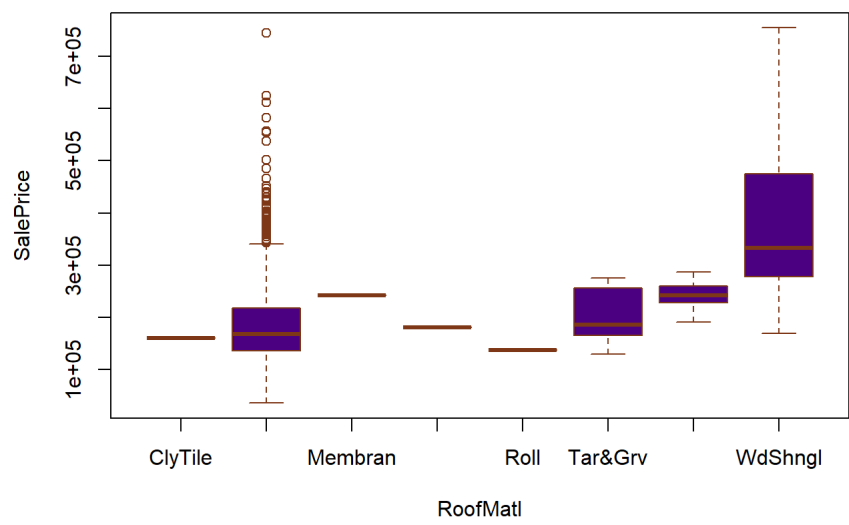


```
table(train3$Exterior1st)
```

```
##
## AsbShng AsphShn BrkComm BrkFace  CBlock CemntBd HdBoard ImStucc MetalSd Plywood
##      15       0       1      44       1      52     211       1     201     100
##   Stone  Stucco VinylSd Wd Sdng WdShing
##       2      21     486     183      20
```

```
boxplot(SalePrice ~ Exterior1st , data = train3 , col = "#4B0082" , border = "#7E3817")
```
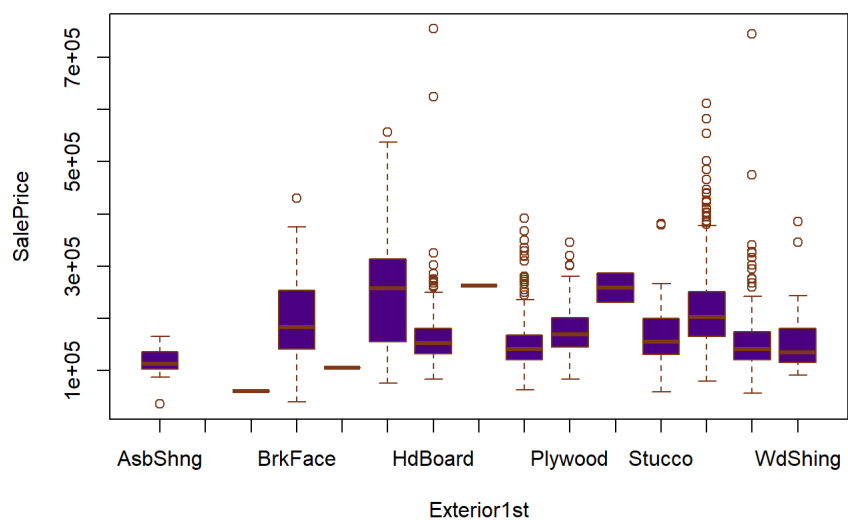
```
table(train3$Exterior2nd)
```

```
##
## AsbShng AsphShn Brk Cmn BrkFace  CBlock CmentBd HdBoard ImStucc MetalSd   Other
##      16       2       6      22       1      51     197      10     197       1
## Plywood   Stone  Stucco VinylSd Wd Sdng Wd Shng
##     127       2      23     475     176      32
```
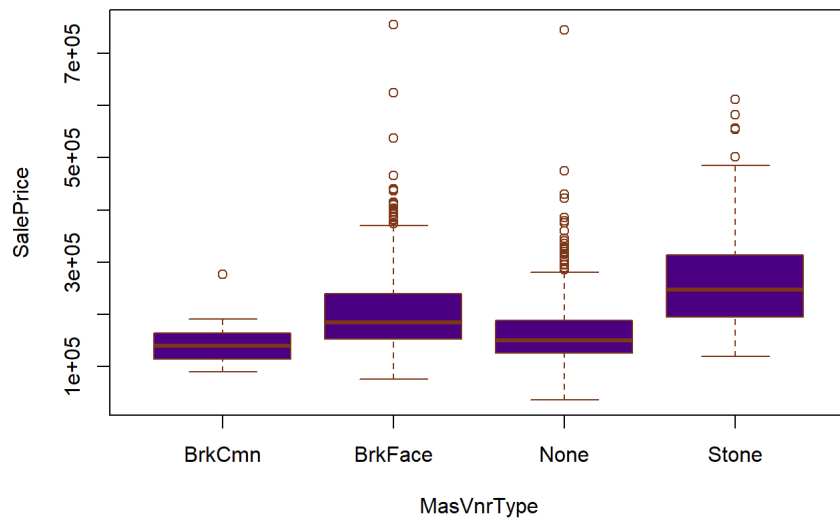
```
boxplot(SalePrice ~ Exterior2nd , data = train3 , col = "#4B0082" , border = "#7E3817")
```
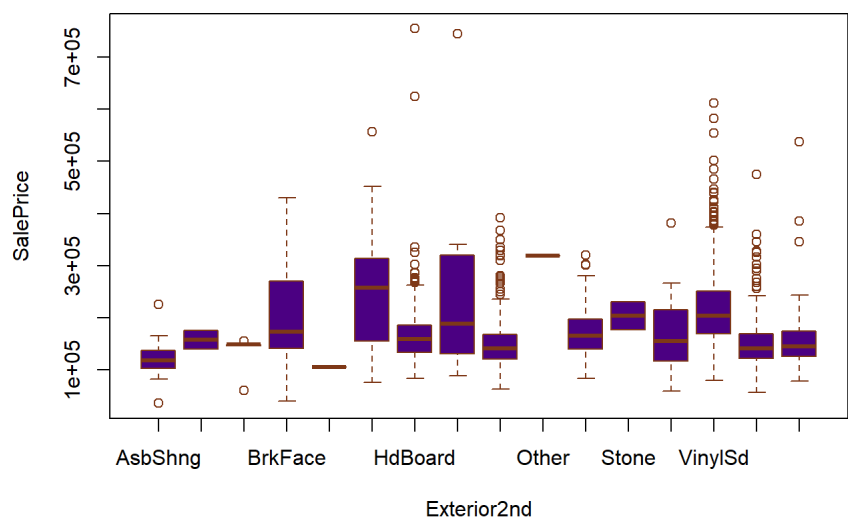


```
table(train3$MasVnrType)
```
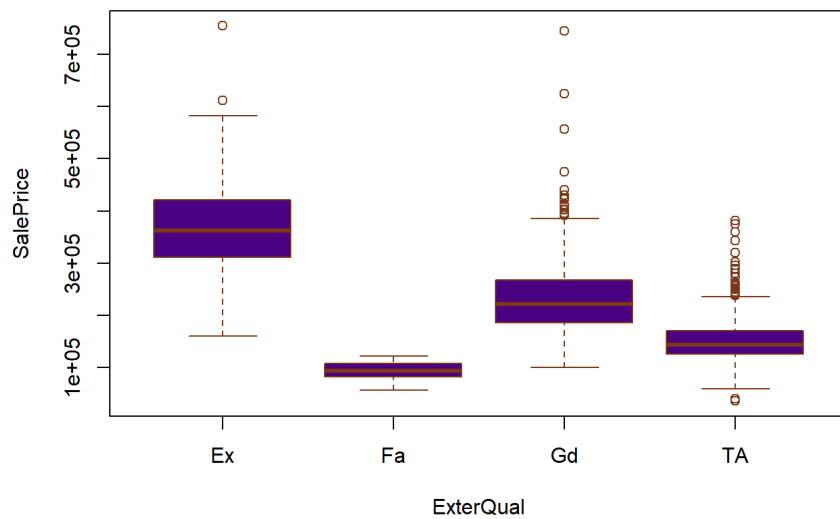
```
##
## BrkCmn BrkFace    None   Stone
##     15     432     763     128
```

```
boxplot(SalePrice ~ MasVnrType , data = train3 , col = "#4B0082" , border = "#7E3817")
```

```
table(train3$ExterQual)
```

```
##
## Ex  Fa  Gd  TA
## 51   7 477 803
```
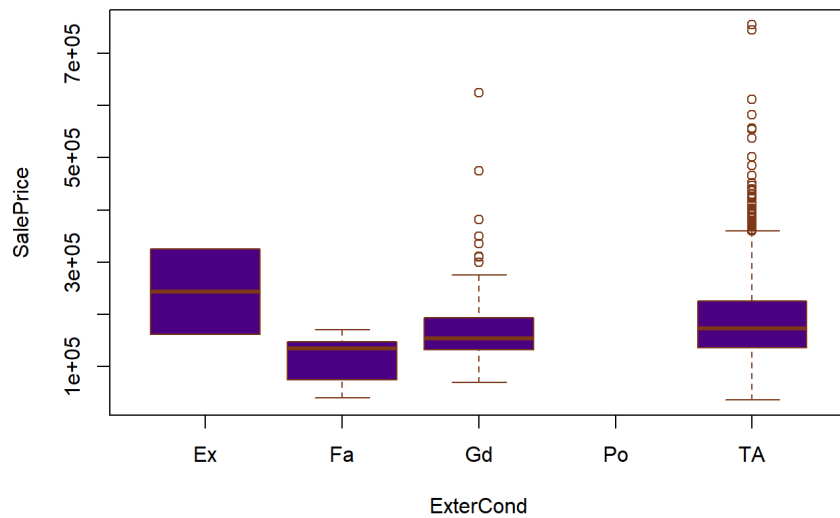
```
boxplot(SalePrice ~ ExterQual ,data = train3 , col = "#4B0082" , border = "#7E3817")
```



```
table(train3$ExterCond)
```

```
##
## Ex   Fa   Gd   Po   TA
##  2   16  137    0 1183
```
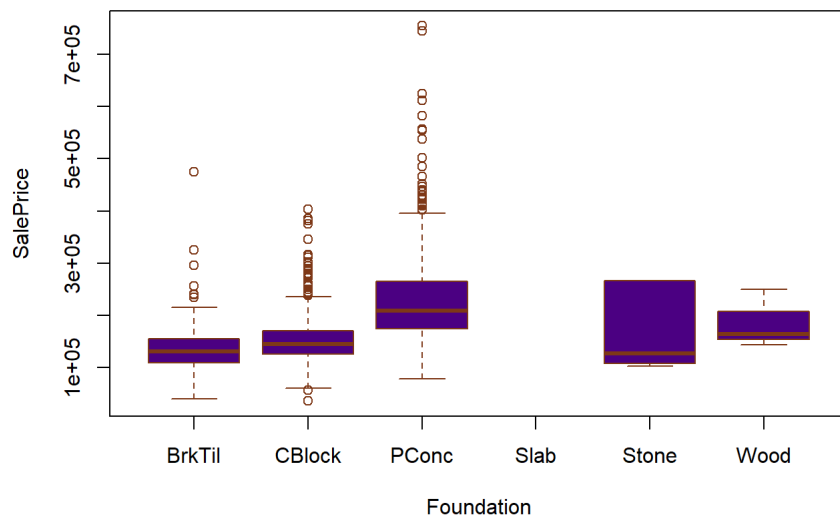
```
boxplot(SalePrice ~ ExterCond , data = train3 , col = "#4B0082" , border = "#7E3817")
```

```
table(train3$Foundation)
```

```
##
## BrkTil CBlock  PConc   Slab  Stone   Wood
##    129    580    620      0      6      3
```
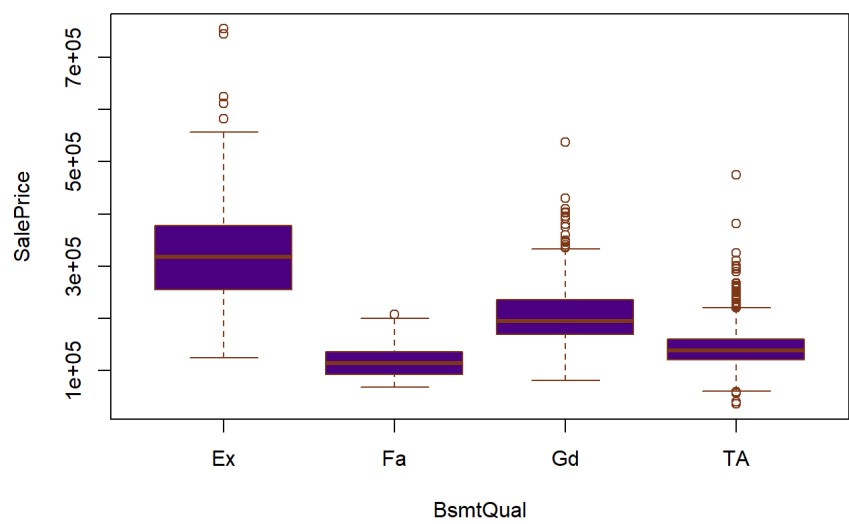
```
boxplot(SalePrice ~ Foundation , data = train3 , col = "#4B0082" , border = "#7E3817")
```



```
table(train3$BsmtQual)
```
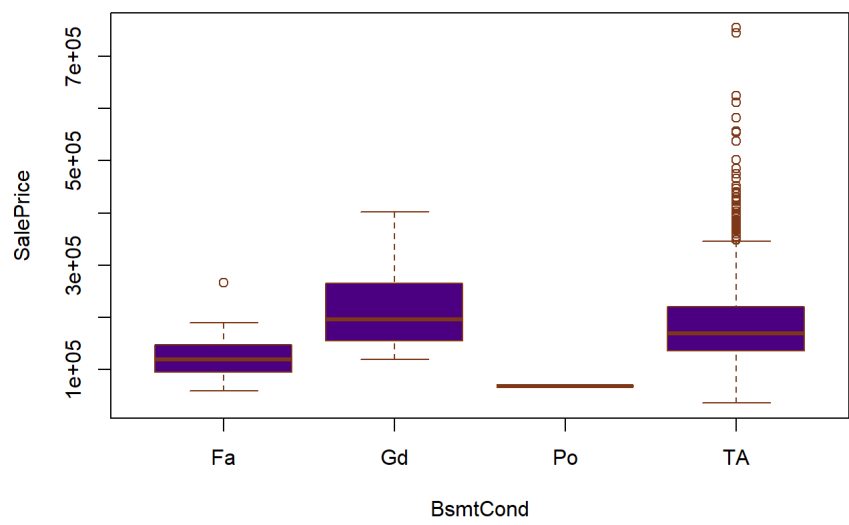
```
##
##  Ex  Fa  Gd  TA
## 120  32 592 594
```

```
boxplot(SalePrice ~ BsmtQual , data = train3 , col = "#4B0082" , border = "#7E3817")
```

```
table(train3$BsmtCond)
```

```
##
##   Fa   Gd   Po   TA
##   38   62    1 1237
```
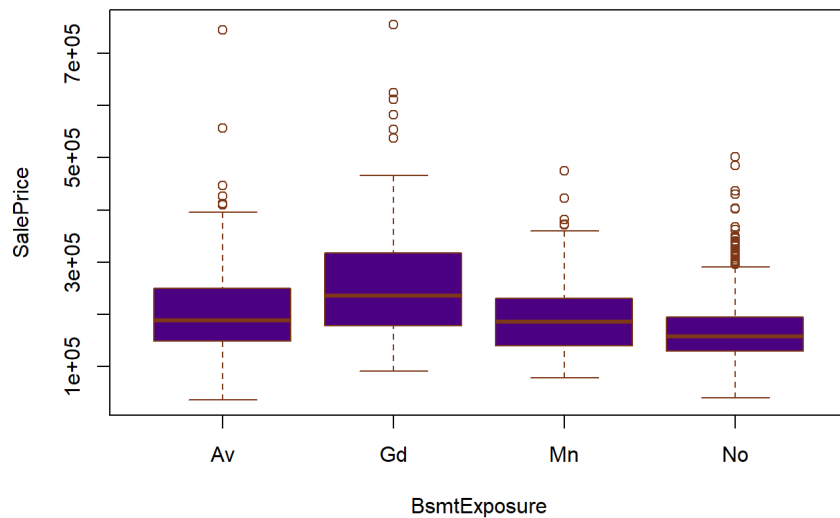
```
boxplot(SalePrice ~ BsmtCond , data = train3 , col = "#4B0082" , border = "#7E3817")
```



```
table(train3$BsmtExposure)
```

```
##
##  Av  Gd  Mn  No
## 213 127 111 887
```
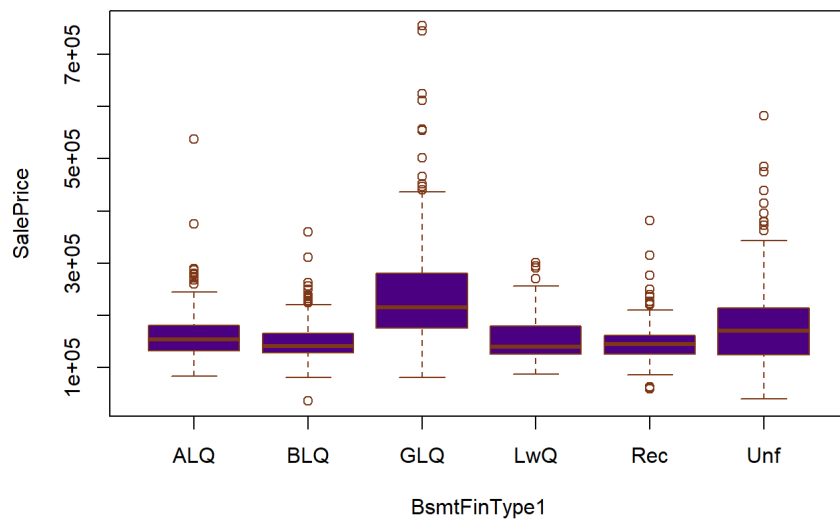
```
boxplot(SalePrice ~ BsmtExposure , data = train3 , col = "#4B0082" , border = "#7E3817")
```

```
table(train3$BsmtFinType1)
```

```
##
## ALQ BLQ GLQ LwQ Rec Unf
## 209 141 402  69 125 392
```
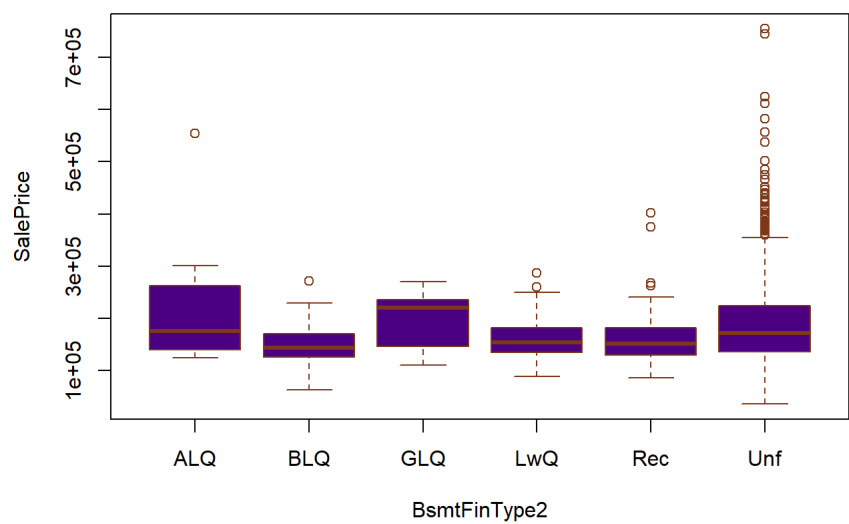
```
boxplot(SalePrice ~ BsmtFinType1 , data = train3 , col = "#4B0082" , border = "#7E3817")
```



```
table(train3$BsmtFinType2)
```

```
##
##  ALQ  BLQ  GLQ  LwQ  Rec  Unf
##   19   32   12   46   53 1176
```
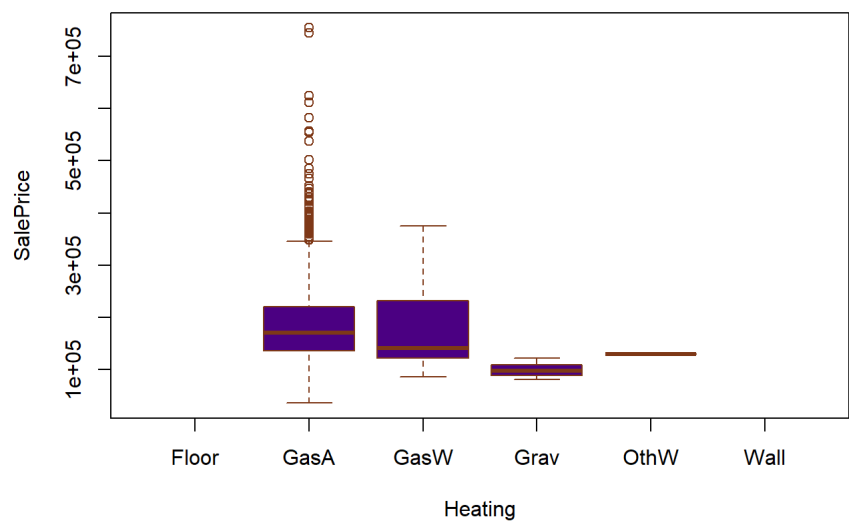
```
boxplot(SalePrice ~ BsmtFinType2 , data = train3 , col = "#4B0082" , border = "#7E3817")
```

```
table(train3$Heating)
```

```
##
## Floor  GasA  GasW  Grav  OthW  Wall
##     0  1318    16     3     1     0
```
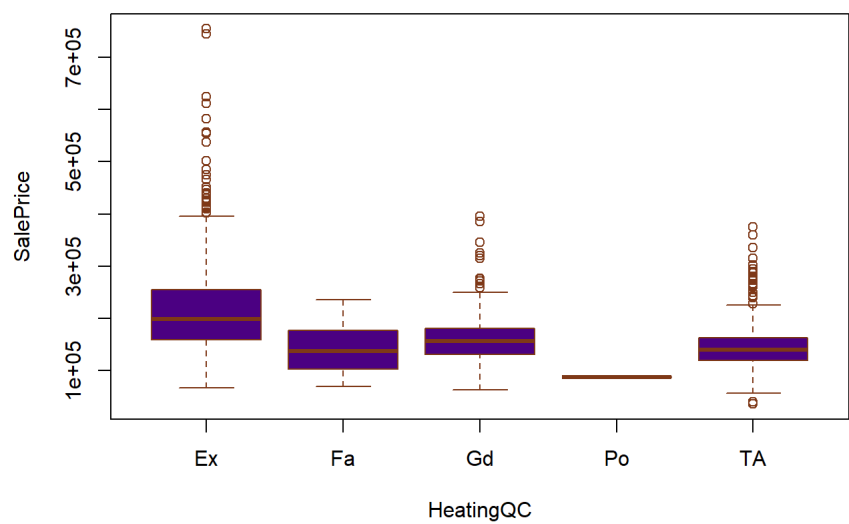
```
boxplot(SalePrice ~ Heating ,data = train3 , col = "#4B0082" , border = "#7E3817")
```



```
table(train3$HeatingQC)
```

```
##
##  Ex  Fa  Gd  Po  TA
## 704  36 217   1 380
```

```
boxplot(SalePrice ~ HeatingQC , data = train3 , col = "#4B0082" , border = "#7E3817")
```

```
table(train3$CentralAir)
```

```
##
##    N    Y
##   61 1277
```

```
boxplot(SalePrice ~ CentralAir ,data = train3 , col = "#4B0082" , border = "#7E3817")
```
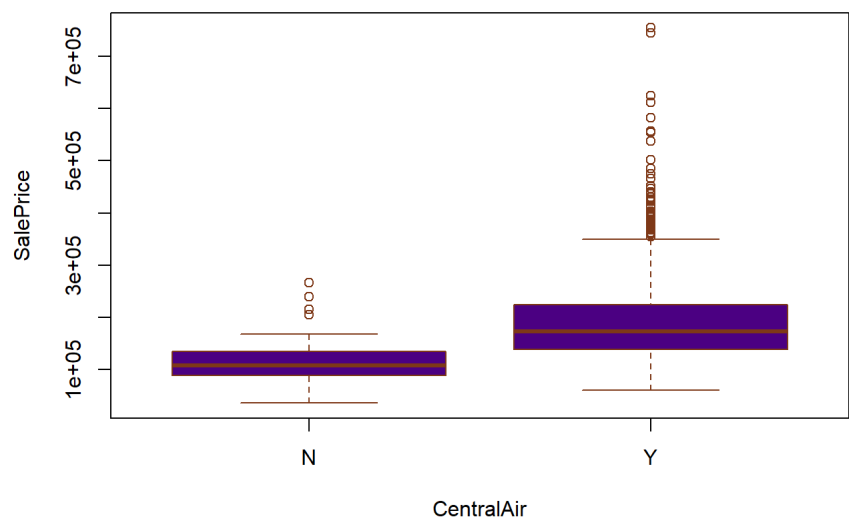


```
table(train3$Electrical)
```

```
##
## FuseA FuseF FuseP   Mix SBrkr
##    76    17     2     1  1242
```

```
boxplot(SalePrice ~ Electrical , data = train3 , col = "#4B0082" , border = "#7E3817")
```

```
table(train3$KitchenQual)
```
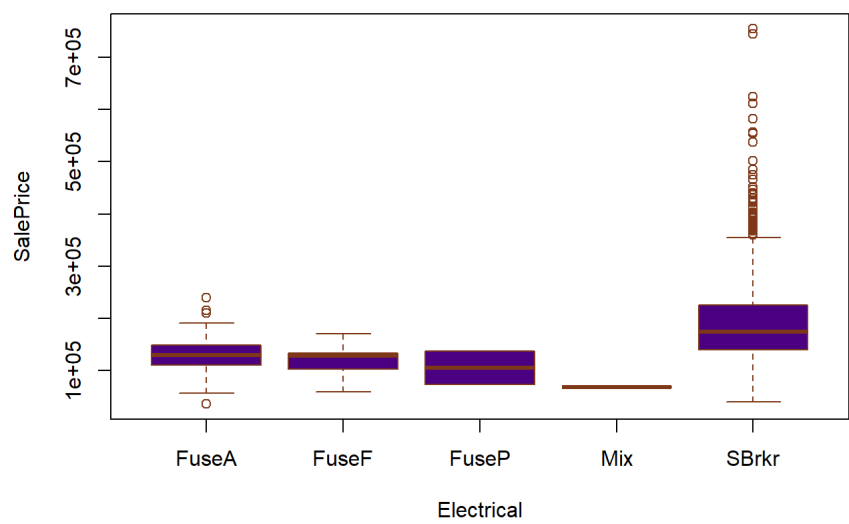
```
##
## Ex  Fa  Gd  TA
## 97  23 568 650
```

```
boxplot(SalePrice ~ KitchenQual , data = train3, col = "#4B0082" , border = "#7E3817")
```
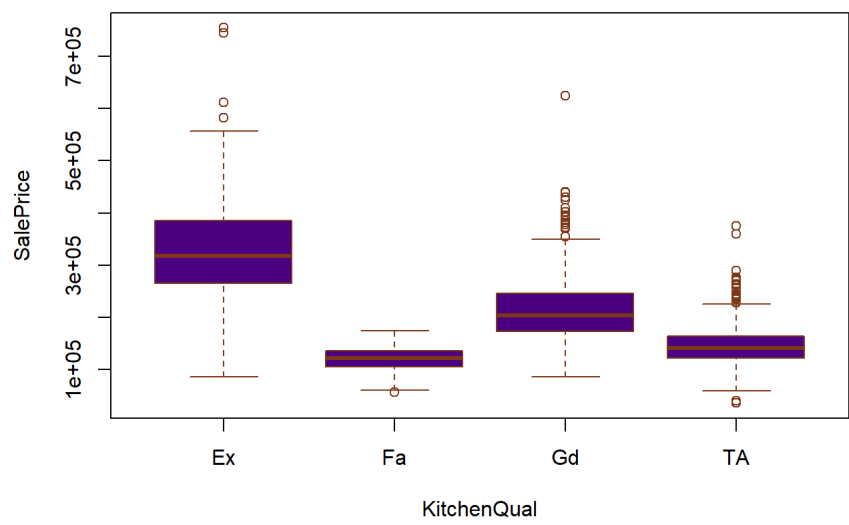


```
table(train3$Functional)
```

```
##
## Maj1 Maj2 Min1 Min2  Mod  Sev  Typ
##   10    4   28   30   11    1 1254
```

```
boxplot(SalePrice ~ Functional, data = train3 , col = "#4B0082" , border = "#7E3817")
```
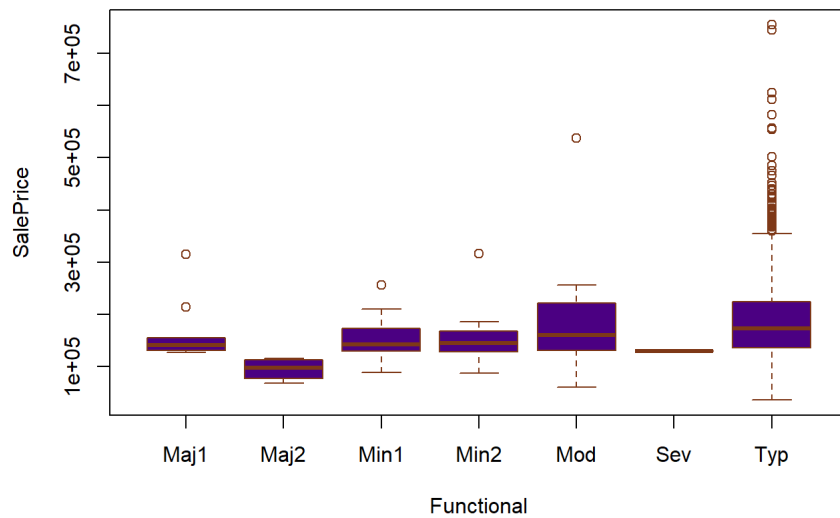
```
table(train3$Fireplaces)
```

```
##
##   0   1   2   3
## 591 631 111   5
```

```
boxplot(SalePrice ~ Fireplaces , data = train3 , col = "#4B0082" , border = "#7E3817")
```



```
table(train3$GarageType)
```

```
##
##  2Types  Attchd Basment BuiltIn CarPort  Detchd
##       6     852      19      85       7     369
```

```
boxplot(SalePrice ~ GarageType , data = train3 , col = "#4B0082" , border = "#7E3817")
```

```
table(train3$GarageFinish)
```

```
##
## Fin RFn Unf
## 345 413 580
```

```
boxplot(SalePrice ~ GarageFinish , data = train3 , col = "#4B0082" , border = "#7E3817")
```



```
table(train3$GarageCars)
```

```
##
##   0   1   2   3   4
##   0 361 793 179   5
```

```
boxplot(SalePrice ~ GarageCars , data = train3 , col = "#4B0082" , border = "#7E3817")
```
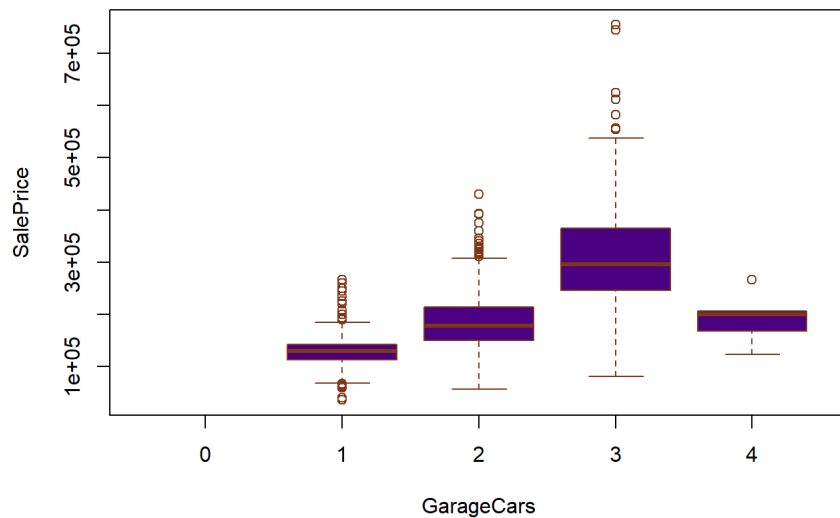
```
table(train3$GarageCond)
```

```
##
##   Ex   Fa   Gd   Po   TA
##    2   33    9    7 1287
```

```
boxplot(SalePrice ~ GarageCond , data = train3 , col = "#4B0082" , border = "#7E3817")
```
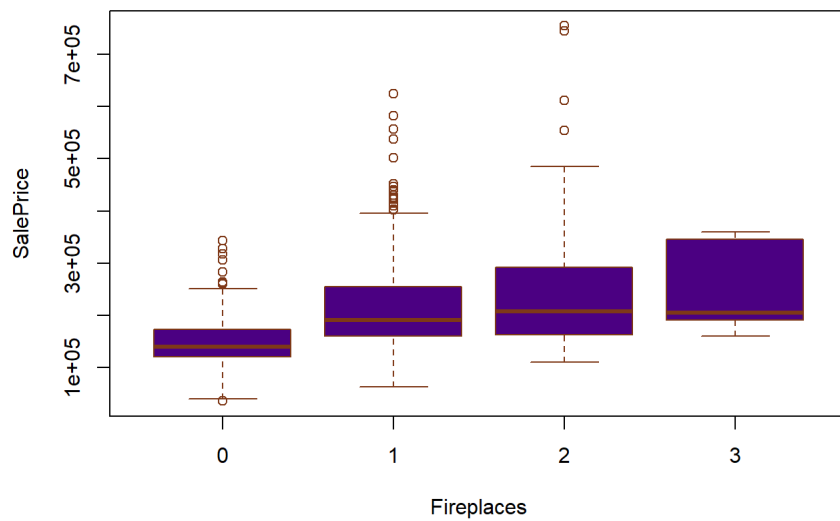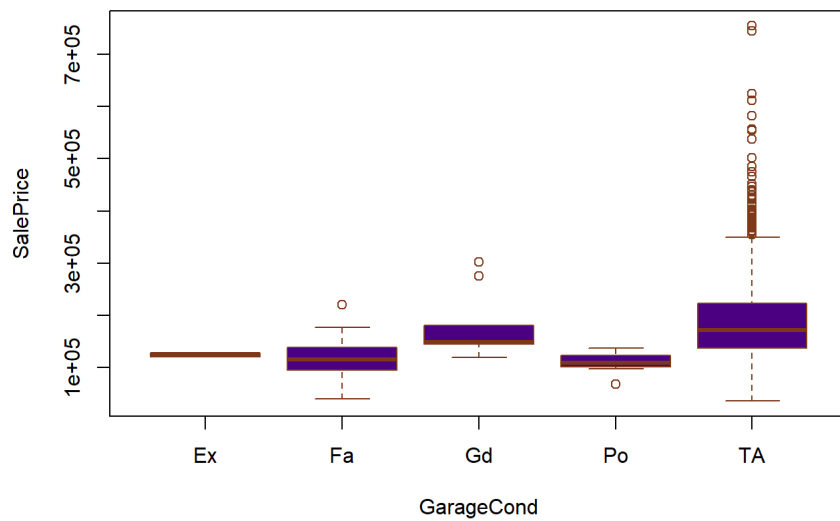


```
table(train3$PavedDrive)
```
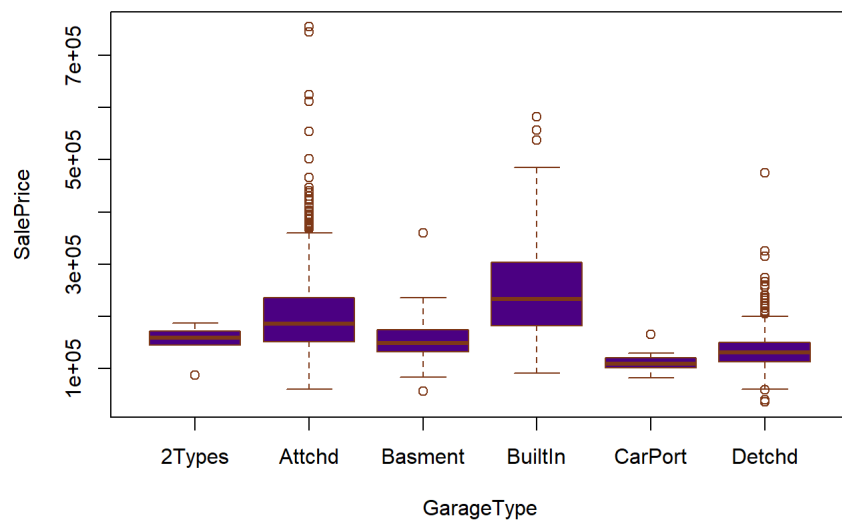
```
##
##    N    P    Y
##   54   27 1257
```

```
boxplot(SalePrice ~ PavedDrive , data = train3 , col = "#4B0082" , border = "#7E3817")
```
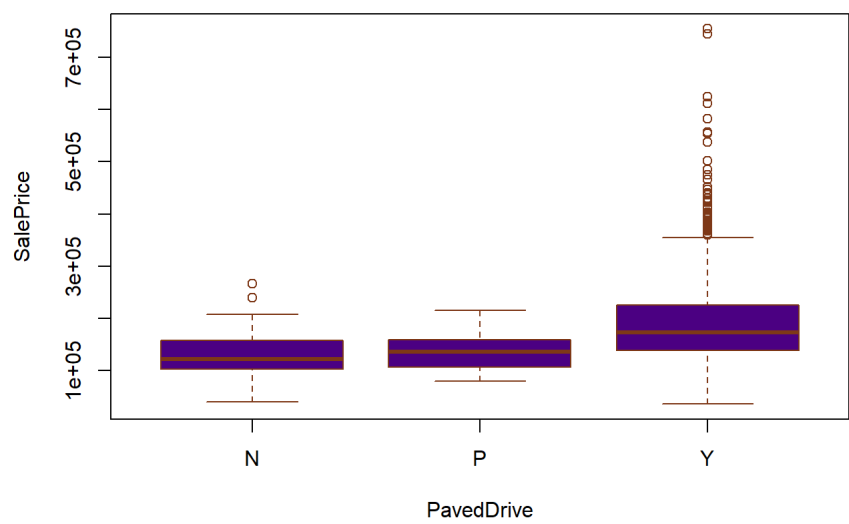
```
table(train3$SaleType)
```

```
##
##   COD   Con ConLD ConLI ConLw   CWD   New   Oth    WD
##    42     2     6     4     4     4   117     1  1158
```
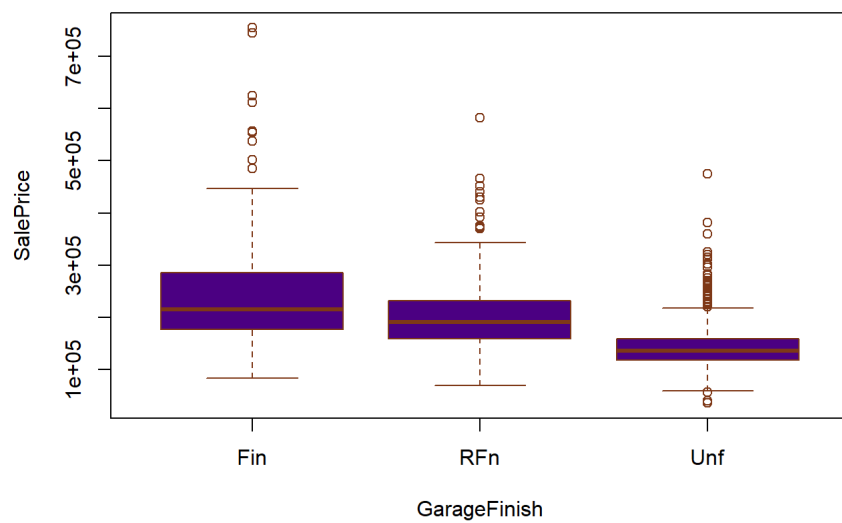
```
boxplot(SalePrice ~ SaleType , data = train3 , col = "#4B0082" , border = "#7E3817")
```
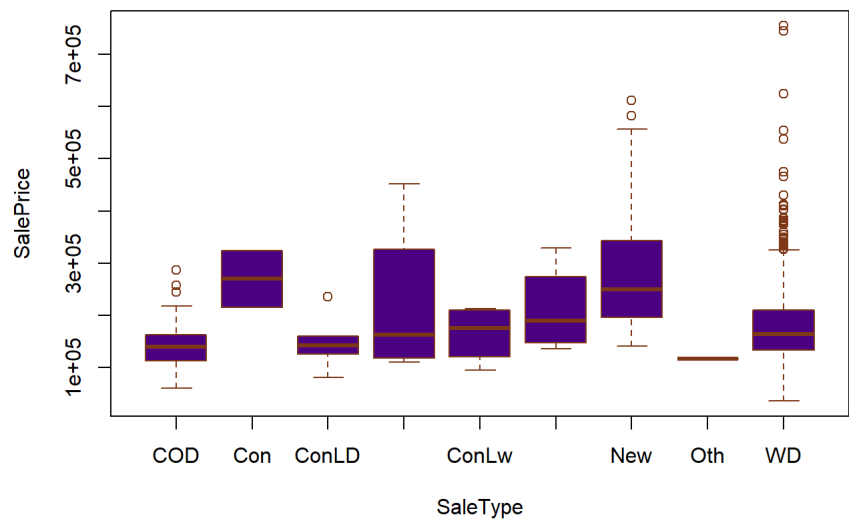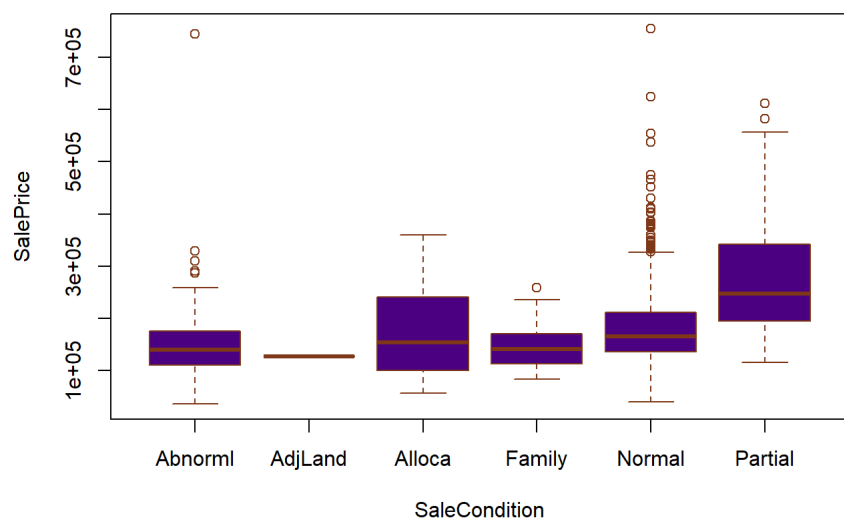


```
table(train3$SaleCondition)
```

```
##
## Abnorml AdjLand  Alloca  Family  Normal Partial
##      86       1       7      20    1104     120
```

```
boxplot(SalePrice ~ SaleCondition , data = train3 , col = "#4B0082" , border = "#7E3817")
```

```
m1 <- lm(SalePrice ~ . ,data = train3 )

summary(m1)
```

```
##
## Call:
## lm(formula = SalePrice ~ ., data = train3)
##
## Residuals:
##     Min     1Q  Median     3Q     Max
## -182117   -8946    118   8735  182117
##
## Coefficients: (4 not defined because of singularities)
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -7.146e+05  6.781e+04 -10.539  < 2e-16 ***
## Id                 4.377e-01  1.600e+00   0.274 0.784447
## MSSubClass        -3.399e+01  8.907e+01  -0.382 0.702841
## MSZoningFV         4.453e+04  1.327e+04   3.356 0.000817 ***
## MSZoningRH         3.768e+04  1.398e+04   2.694 0.007157 **
## MSZoningRL         3.652e+04  1.170e+04   3.121 0.001849 **
## MSZoningRM         3.257e+04  1.103e+04   2.954 0.003201 **
## LotArea            7.732e-01  1.119e-01   6.912 8.12e-12 ***
## StreetPave         4.897e+04  1.583e+04   3.094 0.002024 **
## LotShapeIR2        4.958e+03  4.273e+03   1.160 0.246167
## LotShapeIR3        4.846e+03  8.954e+03   0.541 0.588502
## LotShapeReg        9.651e+02  1.657e+03   0.582 0.560435
## LandContourHLS     6.227e+03  5.546e+03   1.123 0.261765
## LandContourLow    -1.113e+04  6.989e+03  -1.593 0.111415
## LandContourLvl     4.992e+03  4.122e+03   1.211 0.226142
## UtilitiesNoSeWa   -3.559e+04  2.637e+04  -1.350 0.177359
## LotConfigCulDSac   6.750e+03  3.271e+03   2.063 0.039302 *
## LotConfigFR2      -9.238e+03  4.130e+03  -2.237 0.025489 *
## LotConfigFR3      -1.607e+04  1.255e+04  -1.281 0.200615
## LotConfigInside   -1.598e+03  1.832e+03  -0.872 0.383205
## LandSlopeMod       8.688e+03  4.210e+03   2.063 0.039309 *
## LandSlopeSev      -4.411e+04  1.149e+04  -3.838 0.000131 ***
## NeighborhoodBlueste 2.020e+03  1.913e+04   0.106 0.915926
## NeighborhoodBrDale  5.752e+02  1.124e+04   0.051 0.959195
## NeighborhoodBrkSide 7.465e+02  1.006e+04   0.074 0.940842
## NeighborhoodClearCr -8.848e+03  9.591e+03  -0.922 0.356475
## NeighborhoodCollgCr -8.182e+03  7.386e+03  -1.108 0.268201
## NeighborhoodCrawfor  1.434e+04  8.786e+03   1.632 0.102914
## NeighborhoodEdwards -2.073e+04  8.274e+03  -2.506 0.012370 *
## NeighborhoodGilbert -7.041e+03  7.824e+03  -0.900 0.368375
## NeighborhoodIDOTRR  -4.021e+03  1.145e+04  -0.351 0.725485
## NeighborhoodMeadowV -1.496e+04  1.187e+04  -1.261 0.207744
## NeighborhoodMitchel -1.540e+04  8.484e+03  -1.815 0.069726 .
## NeighborhoodNAmes   -1.568e+04  8.016e+03  -1.956 0.050694 .
## NeighborhoodNoRidge  1.975e+04  8.534e+03   2.315 0.020804 *
## NeighborhoodNPkVill  1.066e+04  1.401e+04   0.761 0.446934
## NeighborhoodNridgHt  1.434e+04  7.580e+03   1.891 0.058869 .
## NeighborhoodNWAmes  -1.401e+04  8.232e+03  -1.701 0.089139 .
## NeighborhoodOldTown -9.599e+03  1.011e+04  -0.950 0.342556
## NeighborhoodSawyer  -9.397e+03  8.281e+03  -1.135 0.256698
## NeighborhoodSawyerW -2.951e+03  7.998e+03  -0.369 0.712243
## NeighborhoodSomerst -2.675e+03  9.085e+03  -0.294 0.768506
## NeighborhoodStoneBr  3.286e+04  8.465e+03   3.881 0.000110 ***
## NeighborhoodSWISU   -9.256e+03  1.036e+04  -0.894 0.371645
## NeighborhoodTimber  -8.042e+03  8.244e+03  -0.975 0.329532
## NeighborhoodVeenker  2.837e+03  1.064e+04   0.267 0.789865
## Condition1Feedr      3.471e+03  5.545e+03   0.626 0.531493
## Condition1Norm       1.350e+04  4.497e+03   3.002 0.002745 **
## Condition1PosA       8.959e+03  1.008e+04   0.889 0.374090
## Condition1PosN       9.994e+03  7.649e+03   1.307 0.191634
## Condition1RRAe      -1.325e+04  9.656e+03  -1.372 0.170287
## Condition1RRAn       7.722e+03  7.087e+03   1.089 0.276184
## Condition1RRNe      -5.488e+03  1.748e+04  -0.314 0.753591
## Condition1RRNn       1.249e+04  1.307e+04   0.956 0.339379
## Condition2Feedr      1.475e+02  2.538e+04   0.006 0.995363
## Condition2Norm       4.263e+03  2.211e+04   0.193 0.847109
## Condition2PosA      -1.476e+04  4.351e+04  -0.339 0.734568
## Condition2PosN      -2.580e+05  2.934e+04  -8.794  < 2e-16 ***
## Condition2RRAe      -1.131e+05  4.715e+04  -2.398 0.016661 *
## Condition2RRAn      -2.184e+03  3.275e+04  -0.067 0.946837
## Condition2RRNn       1.268e+04  2.862e+04   0.443 0.657704
## BldgType2fmCon      -6.813e+03  1.418e+04  -0.480 0.630974
## BldgTypeDuplex      -1.145e+04  8.827e+03  -1.297 0.195007
## BldgTypeTwnhs       -1.765e+04  1.053e+04  -1.676 0.094055 .
## BldgTypeTwnhsE      -1.275e+04  9.571e+03  -1.332 0.183149
## HouseStyle1.5Unf     1.221e+04  9.420e+03   1.296 0.195210
## HouseStyle1Story     7.816e+03  4.772e+03   1.638 0.101713
## HouseStyle2.5Fin    -2.223e+04  1.457e+04  -1.526 0.127417
```

```
## HouseStyle2.5Unf    -3.381e+02  1.057e+04  -0.032 0.974490
## HouseStyle2Story    -3.285e+03  3.783e+03  -0.868 0.385426
## HouseStyleSFoyer     4.858e+03  7.026e+03   0.691 0.489474
## HouseStyleSLvl       5.512e+03  5.922e+03   0.931 0.352186
## OverallQual3        -1.571e+04  2.116e+04  -0.743 0.457865
## OverallQual4        -1.809e+04  1.931e+04  -0.937 0.348990
## OverallQual5        -1.560e+04  1.944e+04  -0.802 0.422453
## OverallQual6        -1.346e+04  1.949e+04  -0.691 0.490023
## OverallQual7        -8.562e+03  1.966e+04  -0.436 0.663259
## OverallQual8         4.710e+03  1.987e+04   0.237 0.812616
## OverallQual9         3.648e+04  2.051e+04   1.778 0.075611 .
## OverallQual10        7.783e+04  2.165e+04   3.595 0.000339 ***
## OverallCond3        -6.999e+03  2.101e+04  -0.333 0.739104
## OverallCond4         2.500e+02  2.006e+04   0.012 0.990057
## OverallCond5         9.413e+03  1.974e+04   0.477 0.633515
## OverallCond6         1.612e+04  1.985e+04   0.812 0.416901
## OverallCond7         2.265e+04  1.988e+04   1.140 0.254710
## OverallCond8         2.638e+04  2.014e+04   1.310 0.190502
## OverallCond9         3.772e+04  2.105e+04   1.792 0.073444 .
## RoofStyleGable      -3.535e+01  1.852e+04  -0.002 0.998477
## RoofStyleGambrel     1.990e+03  2.051e+04   0.097 0.922727
## RoofStyleHip         1.368e+02  1.856e+04   0.007 0.994123
## RoofStyleMansard     9.716e+03  2.241e+04   0.434 0.664712
## RoofStyleShed        9.652e+04  3.446e+04   2.801 0.005184 **
## RoofMatlCompShg      6.745e+05  3.537e+04  19.070  < 2e-16 ***
## RoofMatlMembran      7.630e+05  4.969e+04  15.353  < 2e-16 ***
## RoofMatlMetal        7.345e+05  4.847e+04  15.154  < 2e-16 ***
## RoofMatlRoll         6.840e+05  4.353e+04  15.713  < 2e-16 ***
## RoofMatlTar&Grv      6.747e+05  4.023e+04  16.771  < 2e-16 ***
## RoofMatlWdShake      6.649e+05  3.891e+04  17.088  < 2e-16 ***
## RoofMatlWdShngl      7.188e+05  3.648e+04  19.706  < 2e-16 ***
## Exterior1stBrkComm  -3.581e+04  3.403e+04  -1.052 0.292873
## Exterior1stBrkFace   6.028e+03  1.441e+04   0.418 0.675687
## Exterior1stCBlock   -2.787e+04  2.955e+04  -0.943 0.345827
## Exterior1stCemntBd  -5.551e+03  2.038e+04  -0.272 0.785415
## Exterior1stHdBoard  -1.454e+04  1.456e+04  -0.998 0.318264
## Exterior1stImStucc  -4.500e+04  2.804e+04  -1.605 0.108775
## Exterior1stMetalSd  -6.398e+03  1.671e+04  -0.383 0.701945
## Exterior1stPlywood  -1.521e+04  1.443e+04  -1.054 0.291899
## Exterior1stStone    -1.001e+04  2.700e+04  -0.371 0.710803
## Exterior1stStucco   -9.197e+03  1.604e+04  -0.573 0.566566
## Exterior1stVinylSd  -1.356e+04  1.479e+04  -0.917 0.359416
## Exterior1stWd Sdng  -1.218e+04  1.414e+04  -0.861 0.389360
## Exterior1stWdShing  -1.157e+04  1.516e+04  -0.763 0.445747
## Exterior2ndAsphShn   1.822e+04  2.354e+04   0.774 0.438981
## Exterior2ndBrk Cmn   1.829e+04  2.081e+04   0.879 0.379683
## Exterior2ndBrkFace   1.209e+04  1.465e+04   0.826 0.409170
## Exterior2ndCBlock          NA         NA      NA       NA
## Exterior2ndCmentBd   1.359e+04  1.974e+04   0.688 0.491309
## Exterior2ndHdBoard   1.656e+04  1.378e+04   1.202 0.229761
## Exterior2ndImStucc   2.890e+04  1.546e+04   1.870 0.061811 .
## Exterior2ndMetalSd   1.368e+04  1.602e+04   0.854 0.393231
## Exterior2ndOther    -1.186e+04  2.750e+04  -0.431 0.666412
## Exterior2ndPlywood   1.531e+04  1.344e+04   1.139 0.254915
## Exterior2ndStone     5.036e+03  2.435e+04   0.207 0.836219
## Exterior2ndStucco    1.496e+04  1.522e+04   0.983 0.325786
## Exterior2ndVinylSd   1.903e+04  1.402e+04   1.357 0.174932
## Exterior2ndWd Sdng   1.809e+04  1.341e+04   1.349 0.177538
## Exterior2ndWd Shng   1.440e+04  1.396e+04   1.031 0.302735
## MasVnrTypeBrkFace    8.737e+03  6.824e+03   1.280 0.200729
## MasVnrTypeNone       1.023e+04  6.884e+03   1.487 0.137380
## MasVnrTypeStone      1.165e+04  7.189e+03   1.621 0.105341
## MasVnrArea           1.505e+01  5.790e+00   2.600 0.009454 **
## ExterQualFa          1.734e+04  1.571e+04   1.104 0.269956
## ExterQualGd         -3.422e+03  5.224e+03  -0.655 0.512606
## ExterQualTA         -4.531e+03  5.675e+03  -0.798 0.424832
## ExterCondFa         -5.850e+02  2.656e+04  -0.022 0.982430
## ExterCondGd         -8.986e+03  2.551e+04  -0.352 0.724678
## ExterCondTA         -4.038e+03  2.553e+04  -0.158 0.874341
## FoundationCBlock     2.935e+03  3.537e+03   0.830 0.406862
## FoundationPConc      5.431e+03  3.757e+03   1.446 0.148508
## FoundationStone      1.008e+04  1.142e+04   0.883 0.377345
## FoundationWood      -3.276e+04  1.480e+04  -2.214 0.027019 *
## BsmtQualFa          -6.454e+03  6.638e+03  -0.972 0.331101
## BsmtQualGd          -1.033e+04  3.400e+03  -3.039 0.002429 **
## BsmtQualTA          -8.927e+03  4.212e+03  -2.119 0.034289 *
## BsmtCondGd          -5.016e+02  5.592e+03  -0.090 0.928541
## BsmtCondPo           1.195e+04  4.272e+04   0.280 0.779791
## BsmtCondTA           2.892e+03  4.661e+03   0.621 0.535026
```

```
## BsmtExposureGd       1.168e+04  3.070e+03   3.806 0.000149 ***
## BsmtExposureMn      -3.189e+03  3.061e+03  -1.042 0.297700
## BsmtExposureNo      -4.374e+03  2.215e+03  -1.975 0.048518 *
## BsmtFinType1BLQ      1.608e+03  2.862e+03   0.562 0.574275
## BsmtFinType1GLQ      5.912e+03  2.569e+03   2.301 0.021555 *
## BsmtFinType1LwQ     -2.654e+03  3.874e+03  -0.685 0.493428
## BsmtFinType1Rec      9.310e+02  3.095e+03   0.301 0.763581
## BsmtFinType1Unf      3.569e+03  3.031e+03   1.177 0.239341
## BsmtFinSF1           3.456e+01  5.713e+00   6.049 1.99e-09 ***
## BsmtFinType2BLQ     -8.625e+03  7.622e+03  -1.132 0.258074
## BsmtFinType2GLQ     -2.648e+03  9.853e+03  -0.269 0.788210
## BsmtFinType2LwQ     -1.280e+04  7.415e+03  -1.726 0.084611 .
## BsmtFinType2Rec     -9.464e+03  7.104e+03  -1.332 0.183073
## BsmtFinType2Unf     -7.594e+03  7.580e+03  -1.002 0.316655
## BsmtFinSF2           2.644e+01  9.290e+00   2.846 0.004503 **
## BsmtUnfSF            1.579e+01  5.337e+00   2.960 0.003147 **
## TotalBsmtSF                 NA         NA      NA       NA
## HeatingGasW          6.569e+02  7.653e+03   0.086 0.931611
## HeatingGrav          1.023e+04  1.922e+04   0.533 0.594464
## HeatingOthW         -1.006e+04  2.748e+04  -0.366 0.714315
## HeatingQCFa          4.284e+02  5.072e+03   0.084 0.932704
## HeatingQCGd         -2.771e+03  2.139e+03  -1.296 0.195387
## HeatingQCPo          1.174e+04  2.746e+04   0.427 0.669163
## HeatingQCTA         -2.935e+03  2.178e+03  -1.348 0.178027
## CentralAirY          5.440e+03  4.685e+03   1.161 0.245828
## ElectricalFuseF     -6.540e+03  7.315e+03  -0.894 0.371496
## ElectricalFuseP      1.033e+04  2.302e+04   0.449 0.653669
## ElectricalMix               NA         NA      NA       NA
## ElectricalSBrkr     -1.362e+03  3.222e+03  -0.423 0.672612
## X1stFlrSF            5.195e+01  6.133e+00   8.471  < 2e-16 ***
## X2ndFlrSF            6.549e+01  5.842e+00  11.212  < 2e-16 ***
## LowQualFinSF         1.764e+01  2.253e+01   0.783 0.433797
## GrLivArea                   NA         NA      NA       NA
## BsmtFullBath         9.382e+02  2.044e+03   0.459 0.646374
## BsmtHalfBath         1.560e+03  3.093e+03   0.504 0.614234
## FullBath             3.839e+03  2.388e+03   1.607 0.108282
## HalfBath             7.892e+02  2.206e+03   0.358 0.720640
## BedroomAbvGr        -2.660e+03  1.473e+03  -1.805 0.071305 .
## KitchenAbvGr        -1.656e+04  7.406e+03  -2.237 0.025519 *
## KitchenQualFa       -1.535e+04  7.017e+03  -2.188 0.028883 *
## KitchenQualGd       -1.726e+04  3.646e+03  -4.732 2.51e-06 ***
## KitchenQualTA       -1.619e+04  4.096e+03  -3.952 8.26e-05 ***
## TotRmsAbvGrd         9.444e+02  9.845e+02   0.959 0.337620
## FunctionalMaj2      -4.360e+03  1.728e+04  -0.252 0.800823
## FunctionalMin1       6.682e+02  9.331e+03   0.072 0.942921
## FunctionalMin2      -1.372e+02  9.490e+03  -0.014 0.988469
## FunctionalMod       -6.256e+03  1.188e+04  -0.526 0.598654
## FunctionalSev       -4.449e+04  2.994e+04  -1.486 0.137616
## FunctionalTyp        1.209e+04  8.231e+03   1.469 0.142240
## Fireplaces1          9.276e+02  1.781e+03   0.521 0.602602
## Fireplaces2          8.036e+03  3.107e+03   2.586 0.009827 **
## Fireplaces3          2.341e+03  1.229e+04   0.191 0.848940
## GarageTypeAttchd     2.150e+04  1.156e+04   1.859 0.063307 .
## GarageTypeBasment    2.408e+04  1.339e+04   1.798 0.072392 .
## GarageTypeBuiltIn    2.049e+04  1.200e+04   1.707 0.088108 .
## GarageTypeCarPort    2.575e+04  1.652e+04   1.559 0.119317
## GarageTypeDetchd     2.332e+04  1.157e+04   2.015 0.044192 *
## GarageFinishRFn     -9.552e+02  1.984e+03  -0.481 0.630345
## GarageFinishUnf     -1.784e+02  2.440e+03  -0.073 0.941744
## GarageCars2         -8.922e+02  2.698e+03  -0.331 0.740942
## GarageCars3          8.816e+03  4.947e+03   1.782 0.074993 .
## GarageCars4          1.849e+04  1.260e+04   1.467 0.142575
## GarageArea           1.824e+01  8.091e+00   2.254 0.024364 *
## GarageQualFa        -4.916e+04  3.189e+04  -1.541 0.123544
## GarageQualGd        -3.905e+04  3.270e+04  -1.194 0.232591
## GarageQualPo        -7.294e+04  4.066e+04  -1.794 0.073098 .
## GarageQualTA        -4.490e+04  3.154e+04  -1.423 0.154919
## GarageCondFa         3.953e+04  3.622e+04   1.091 0.275320
## GarageCondGd         4.025e+04  3.738e+04   1.077 0.281822
## GarageCondPo         4.563e+04  3.886e+04   1.174 0.240512
## GarageCondTA         4.177e+04  3.583e+04   1.166 0.243897
## PavedDriveP         -3.260e+03  6.270e+03  -0.520 0.603167
## PavedDriveY          1.486e+03  4.237e+03   0.351 0.725811
## WoodDeckSF           9.165e+00  5.956e+00   1.539 0.124127
## OpenPorchSF          8.870e+00  1.230e+01   0.721 0.470957
## EnclosedPorch        6.918e+00  1.309e+01   0.528 0.597335
## X3SsnPorch           4.719e+01  2.254e+01   2.094 0.036496 *
## ScreenPorch          3.755e+01  1.234e+01   3.042 0.002402 **
## PoolArea             6.694e+01  1.846e+01   3.625 0.000302 ***
```

```
## MiscVal                  4.415e-01  1.428e+00   0.309 0.757275
## MoSold                  -4.235e+02  2.549e+02  -1.662 0.096877 .
## SaleTypeCon              2.692e+04  1.746e+04   1.542 0.123370
## SaleTypeConLD            2.016e+04  1.195e+04   1.687 0.091839 .
## SaleTypeConLI           -4.808e+03  1.292e+04  -0.372 0.709919
## SaleTypeConLw            1.025e+03  1.305e+04   0.079 0.937417
## SaleTypeCWD              8.743e+03  1.293e+04   0.676 0.499006
## SaleTypeNew              2.961e+04  1.569e+04   1.887 0.059424 .
## SaleTypeOth              2.012e+04  2.331e+04   0.863 0.388310
## SaleTypeWD              -4.980e+02  4.321e+03  -0.115 0.908262
## SaleConditionAdjLand     3.759e+04  2.545e+04   1.477 0.140077
## SaleConditionAlloca      1.100e+04  1.116e+04   0.985 0.324823
## SaleConditionFamily     -1.491e+03  6.164e+03  -0.242 0.808968
## SaleConditionNormal      5.279e+03  3.098e+03   1.704 0.088696 .
## SaleConditionPartial    -7.685e+03  1.507e+04  -0.510 0.610221
## ageBuilt                -3.402e+02  8.867e+01  -3.837 0.000132 ***
## ageRemodAdd             -6.951e+01  6.188e+01  -1.123 0.261535
## GarageageBlt            -1.718e+00  6.478e+01  -0.027 0.978851
## ageSold                  2.071e+01  5.323e+02   0.039 0.968975
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22360 on 1097 degrees of freedom
## Multiple R-squared:  0.9341, Adjusted R-squared:  0.9197
## F-statistic: 64.81 on 240 and 1097 DF,  p-value: < 2.2e-16
```

```
hist(m1$residuals , probability = TRUE , breaks = 35 ,col = "#4B0082" , border = "#7E3817" )
lines(density(m1$residuals) , col = "red")
```


Histogram of m1$residuals
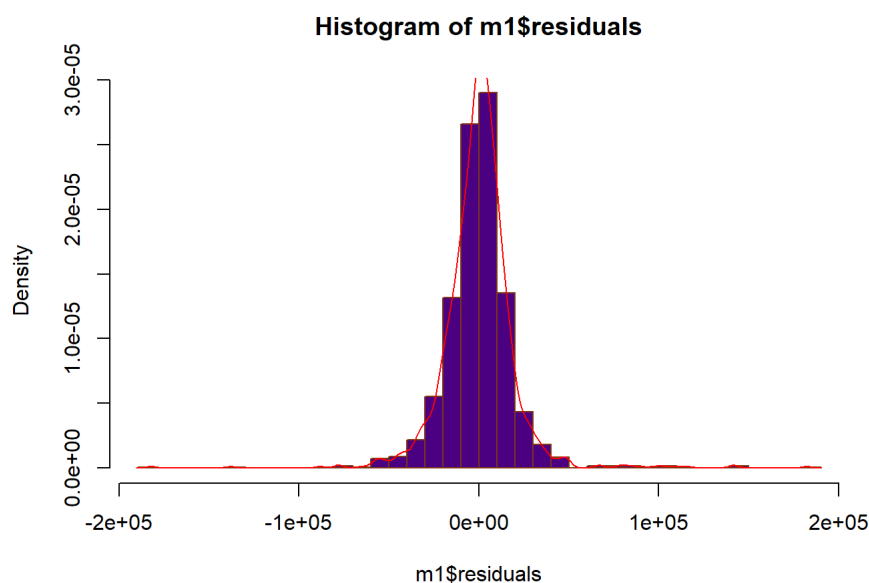
```
qqnorm(m1$residuals , main = "QQplot residuls" ,col = "#4B0082" )
qqline(m1$residuals , col = "red")
```

QQplot residuls
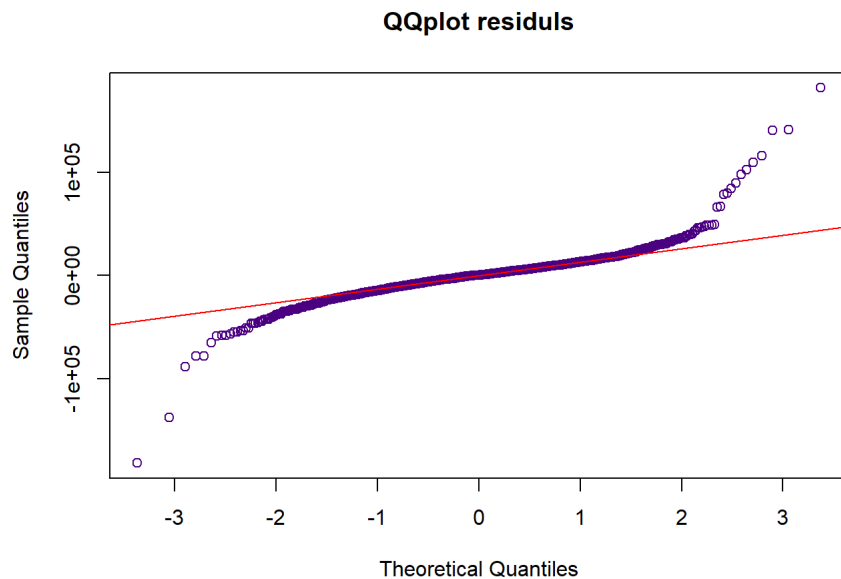

```
jarque.test(m1$residuals)
```

```
##
##  Jarque-Bera Normality Test
##
## data:  m1$residuals
## JB = 17222, p-value < 2.2e-16
## alternative hypothesis: greater
```

```
anscombe.test(m1$residuals)
```

```
##
##  Anscombe-Glynn kurtosis test
##
## data:  m1$residuals
## kurt = 20.532, z = 18.004, p-value < 2.2e-16
## alternative hypothesis: kurtosis is not equal to 3
```