Week2.R

Sinchana Mysore Eshwar

Fri Feb 15 16:31:56 2019

#Import Libraries  
  
library(ggthemes)

## Warning: package 'ggthemes' was built under R version 3.5.2

library(dplyr)

## Warning: package 'dplyr' was built under R version 3.5.2

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(gridExtra)

## Warning: package 'gridExtra' was built under R version 3.5.2

##   
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':  
##   
## combine

library(corrplot)

## Warning: package 'corrplot' was built under R version 3.5.2

## corrplot 0.84 loaded

library(ggplot2)  
library(GGally)

## Warning: package 'GGally' was built under R version 3.5.2

##   
## Attaching package: 'GGally'

## The following object is masked from 'package:dplyr':  
##   
## nasa

library(data.table)

## Warning: package 'data.table' was built under R version 3.5.2

##   
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':  
##   
## between, first, last

library(scales)  
library(MVA)

## Warning: package 'MVA' was built under R version 3.5.2

## Loading required package: HSAUR2

## Warning: package 'HSAUR2' was built under R version 3.5.2

## Loading required package: tools

#Read Training and Testing dataset  
training <- read.csv("D:/MultiAnalysis/Project/house-prices-advanced-regression-techniques/Data.csv.csv", stringsAsFactors=FALSE)  
View(training)  
testing <- read.csv("D:/MultiAnalysis/Project/house-prices-advanced-regression-techniques/test.csv")  
View(testing)  
  
?legend()

## starting httpd help server ...

## done

#Get the dimension and summary of training and testing data set  
dim(training)

## [1] 1460 81

str(training)

## '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 : chr "RL" "RL" "RL" "RL" ...  
## $ 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 : chr "Pave" "Pave" "Pave" "Pave" ...  
## $ Alley : chr NA NA NA NA ...  
## $ LotShape : chr "Reg" "Reg" "IR1" "IR1" ...  
## $ LandContour : chr "Lvl" "Lvl" "Lvl" "Lvl" ...  
## $ Utilities : chr "AllPub" "AllPub" "AllPub" "AllPub" ...  
## $ LotConfig : chr "Inside" "FR2" "Inside" "Corner" ...  
## $ LandSlope : chr "Gtl" "Gtl" "Gtl" "Gtl" ...  
## $ Neighborhood : chr "CollgCr" "Veenker" "CollgCr" "Crawfor" ...  
## $ Condition1 : chr "Norm" "Feedr" "Norm" "Norm" ...  
## $ Condition2 : chr "Norm" "Norm" "Norm" "Norm" ...  
## $ BldgType : chr "1Fam" "1Fam" "1Fam" "1Fam" ...  
## $ HouseStyle : chr "2Story" "1Story" "2Story" "2Story" ...  
## $ OverallQual : int 7 6 7 7 8 5 8 7 7 5 ...  
## $ OverallCond : int 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 : chr "Gable" "Gable" "Gable" "Gable" ...  
## $ RoofMatl : chr "CompShg" "CompShg" "CompShg" "CompShg" ...  
## $ Exterior1st : chr "VinylSd" "MetalSd" "VinylSd" "Wd Sdng" ...  
## $ Exterior2nd : chr "VinylSd" "MetalSd" "VinylSd" "Wd Shng" ...  
## $ MasVnrType : chr "BrkFace" "None" "BrkFace" "None" ...  
## $ MasVnrArea : int 196 0 162 0 350 0 186 240 0 0 ...  
## $ ExterQual : chr "Gd" "TA" "Gd" "TA" ...  
## $ ExterCond : chr "TA" "TA" "TA" "TA" ...  
## $ Foundation : chr "PConc" "CBlock" "PConc" "BrkTil" ...  
## $ BsmtQual : chr "Gd" "Gd" "Gd" "TA" ...  
## $ BsmtCond : chr "TA" "TA" "TA" "Gd" ...  
## $ BsmtExposure : chr "No" "Gd" "Mn" "No" ...  
## $ BsmtFinType1 : chr "GLQ" "ALQ" "GLQ" "ALQ" ...  
## $ BsmtFinSF1 : int 706 978 486 216 655 732 1369 859 0 851 ...  
## $ BsmtFinType2 : chr "Unf" "Unf" "Unf" "Unf" ...  
## $ 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 : chr "GasA" "GasA" "GasA" "GasA" ...  
## $ HeatingQC : chr "Ex" "Ex" "Ex" "Gd" ...  
## $ CentralAir : chr "Y" "Y" "Y" "Y" ...  
## $ Electrical : chr "SBrkr" "SBrkr" "SBrkr" "SBrkr" ...  
## $ 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 : chr "Gd" "TA" "Gd" "Gd" ...  
## $ TotRmsAbvGrd : int 8 6 6 7 9 5 7 7 8 5 ...  
## $ Functional : chr "Typ" "Typ" "Typ" "Typ" ...  
## $ Fireplaces : int 0 1 1 1 1 0 1 2 2 2 ...  
## $ FireplaceQu : chr NA "TA" "TA" "Gd" ...  
## $ GarageType : chr "Attchd" "Attchd" "Attchd" "Detchd" ...  
## $ GarageYrBlt : int 2003 1976 2001 1998 2000 1993 2004 1973 1931 1939 ...  
## $ GarageFinish : chr "RFn" "RFn" "RFn" "Unf" ...  
## $ GarageCars : int 2 2 2 3 3 2 2 2 2 1 ...  
## $ GarageArea : int 548 460 608 642 836 480 636 484 468 205 ...  
## $ GarageQual : chr "TA" "TA" "TA" "TA" ...  
## $ GarageCond : chr "TA" "TA" "TA" "TA" ...  
## $ PavedDrive : chr "Y" "Y" "Y" "Y" ...  
## $ 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 : chr NA NA NA NA ...  
## $ Fence : chr NA NA NA NA ...  
## $ MiscFeature : chr NA NA 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 : chr "WD" "WD" "WD" "WD" ...  
## $ SaleCondition: chr "Normal" "Normal" "Normal" "Abnorml" ...  
## $ SalePrice : int 208500 181500 223500 140000 250000 143000 307000 200000 129900 118000 ...

summary(training)

## 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   
## Length:1460 Length:1460 Length:1460   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## LandSlope Neighborhood Condition1   
## Length:1460 Length:1460 Length:1460   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## Condition2 BldgType HouseStyle OverallQual   
## Length:1460 Length:1460 Length:1460 Min. : 1.000   
## Class :character Class :character Class :character 1st Qu.: 5.000   
## Mode :character Mode :character Mode :character Median : 6.000   
## Mean : 6.099   
## 3rd Qu.: 7.000   
## Max. :10.000   
##   
## OverallCond YearBuilt YearRemodAdd RoofStyle   
## Min. :1.000 Min. :1872 Min. :1950 Length:1460   
## 1st Qu.:5.000 1st Qu.:1954 1st Qu.:1967 Class :character   
## Median :5.000 Median :1973 Median :1994 Mode :character   
## Mean :5.575 Mean :1971 Mean :1985   
## 3rd Qu.:6.000 3rd Qu.:2000 3rd Qu.:2004   
## Max. :9.000 Max. :2010 Max. :2010   
##   
## RoofMatl Exterior1st Exterior2nd   
## Length:1460 Length:1460 Length:1460   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## MasVnrType MasVnrArea ExterQual ExterCond   
## Length:1460 Min. : 0.0 Length:1460 Length:1460   
## Class :character 1st Qu.: 0.0 Class :character Class :character   
## Mode :character Median : 0.0 Mode :character Mode :character   
## Mean : 103.7   
## 3rd Qu.: 166.0   
## Max. :1600.0   
## NA's :8   
## Foundation BsmtQual BsmtCond   
## Length:1460 Length:1460 Length:1460   
## Class :character Class :character Class :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   
## Length:1460 Length:1460 Length:1460   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## MiscVal MoSold YrSold SaleType   
## Min. : 0.00 Min. : 1.000 Min. :2006 Length:1460   
## 1st Qu.: 0.00 1st Qu.: 5.000 1st Qu.:2007 Class :character   
## Median : 0.00 Median : 6.000 Median :2008 Mode :character   
## Mean : 43.49 Mean : 6.322 Mean :2008   
## 3rd Qu.: 0.00 3rd Qu.: 8.000 3rd Qu.:2009   
## Max. :15500.00 Max. :12.000 Max. :2010   
##   
## SaleCondition SalePrice   
## Length:1460 Min. : 34900   
## Class :character 1st Qu.:129975   
## Mode :character Median :163000   
## Mean :180921   
## 3rd Qu.:214000   
## Max. :755000   
##

dim(testing)

## [1] 1459 80

str(testing)

## 'data.frame': 1459 obs. of 80 variables:  
## $ Id : int 1461 1462 1463 1464 1465 1466 1467 1468 1469 1470 ...  
## $ MSSubClass : int 20 20 60 60 120 60 20 60 20 20 ...  
## $ MSZoning : Factor w/ 5 levels "C (all)","FV",..: 3 4 4 4 4 4 4 4 4 4 ...  
## $ LotFrontage : int 80 81 74 78 43 75 NA 63 85 70 ...  
## $ LotArea : int 11622 14267 13830 9978 5005 10000 7980 8402 10176 8400 ...  
## $ 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 1 1 1 1 1 1 1 4 4 ...  
## $ LandContour : Factor w/ 4 levels "Bnk","HLS","Low",..: 4 4 4 4 2 4 4 4 4 4 ...  
## $ Utilities : Factor w/ 1 level "AllPub": 1 1 1 1 1 1 1 1 1 1 ...  
## $ LotConfig : Factor w/ 5 levels "Corner","CulDSac",..: 5 1 5 5 5 1 5 5 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",..: 13 13 9 9 22 9 9 9 9 13 ...  
## $ Condition1 : Factor w/ 9 levels "Artery","Feedr",..: 2 3 3 3 3 3 3 3 3 3 ...  
## $ Condition2 : Factor w/ 5 levels "Artery","Feedr",..: 3 3 3 3 3 3 3 3 3 3 ...  
## $ BldgType : Factor w/ 5 levels "1Fam","2fmCon",..: 1 1 1 1 5 1 1 1 1 1 ...  
## $ HouseStyle : Factor w/ 7 levels "1.5Fin","1.5Unf",..: 3 3 5 5 3 5 3 5 3 3 ...  
## $ OverallQual : int 5 6 5 6 8 6 6 6 7 4 ...  
## $ OverallCond : int 6 6 5 6 5 5 7 5 5 5 ...  
## $ YearBuilt : int 1961 1958 1997 1998 1992 1993 1992 1998 1990 1970 ...  
## $ YearRemodAdd : int 1961 1958 1998 1998 1992 1994 2007 1998 1990 1970 ...  
## $ RoofStyle : Factor w/ 6 levels "Flat","Gable",..: 2 4 2 2 2 2 2 2 2 2 ...  
## $ RoofMatl : Factor w/ 4 levels "CompShg","Tar&Grv",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Exterior1st : Factor w/ 13 levels "AsbShng","AsphShn",..: 11 12 11 11 7 7 7 11 7 9 ...  
## $ Exterior2nd : Factor w/ 15 levels "AsbShng","AsphShn",..: 13 14 13 13 7 7 7 13 7 10 ...  
## $ MasVnrType : Factor w/ 4 levels "BrkCmn","BrkFace",..: 3 2 3 2 3 3 3 3 3 3 ...  
## $ MasVnrArea : int 0 108 0 20 0 0 0 0 0 0 ...  
## $ ExterQual : Factor w/ 4 levels "Ex","Fa","Gd",..: 4 4 4 4 3 4 4 4 4 4 ...  
## $ ExterCond : Factor w/ 5 levels "Ex","Fa","Gd",..: 5 5 5 5 5 5 3 5 5 5 ...  
## $ Foundation : Factor w/ 6 levels "BrkTil","CBlock",..: 2 2 3 3 3 3 3 3 3 2 ...  
## $ BsmtQual : Factor w/ 4 levels "Ex","Fa","Gd",..: 4 4 3 4 3 3 3 3 3 4 ...  
## $ BsmtCond : Factor w/ 4 levels "Fa","Gd","Po",..: 4 4 4 4 4 4 4 4 4 4 ...  
## $ BsmtExposure : Factor w/ 4 levels "Av","Gd","Mn",..: 4 4 4 4 4 4 4 4 2 4 ...  
## $ BsmtFinType1 : Factor w/ 6 levels "ALQ","BLQ","GLQ",..: 5 1 3 3 1 6 1 6 3 1 ...  
## $ BsmtFinSF1 : int 468 923 791 602 263 0 935 0 637 804 ...  
## $ BsmtFinType2 : Factor w/ 6 levels "ALQ","BLQ","GLQ",..: 4 6 6 6 6 6 6 6 6 5 ...  
## $ BsmtFinSF2 : int 144 0 0 0 0 0 0 0 0 78 ...  
## $ BsmtUnfSF : int 270 406 137 324 1017 763 233 789 663 0 ...  
## $ TotalBsmtSF : int 882 1329 928 926 1280 763 1168 789 1300 882 ...  
## $ Heating : Factor w/ 4 levels "GasA","GasW",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ HeatingQC : Factor w/ 5 levels "Ex","Fa","Gd",..: 5 5 3 1 1 3 1 3 3 5 ...  
## $ CentralAir : Factor w/ 2 levels "N","Y": 2 2 2 2 2 2 2 2 2 2 ...  
## $ Electrical : Factor w/ 4 levels "FuseA","FuseF",..: 4 4 4 4 4 4 4 4 4 4 ...  
## $ X1stFlrSF : int 896 1329 928 926 1280 763 1187 789 1341 882 ...  
## $ X2ndFlrSF : int 0 0 701 678 0 892 0 676 0 0 ...  
## $ LowQualFinSF : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ GrLivArea : int 896 1329 1629 1604 1280 1655 1187 1465 1341 882 ...  
## $ BsmtFullBath : int 0 0 0 0 0 0 1 0 1 1 ...  
## $ BsmtHalfBath : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ FullBath : int 1 1 2 2 2 2 2 2 1 1 ...  
## $ HalfBath : int 0 1 1 1 0 1 0 1 1 0 ...  
## $ BedroomAbvGr : int 2 3 3 3 2 3 3 3 2 2 ...  
## $ KitchenAbvGr : int 1 1 1 1 1 1 1 1 1 1 ...  
## $ KitchenQual : Factor w/ 4 levels "Ex","Fa","Gd",..: 4 3 4 3 3 4 4 4 3 4 ...  
## $ TotRmsAbvGrd : int 5 6 6 7 5 7 6 7 5 4 ...  
## $ Functional : Factor w/ 7 levels "Maj1","Maj2",..: 7 7 7 7 7 7 7 7 7 7 ...  
## $ Fireplaces : int 0 0 1 1 0 1 0 1 1 0 ...  
## $ FireplaceQu : Factor w/ 5 levels "Ex","Fa","Gd",..: NA NA 5 3 NA 5 NA 3 4 NA ...  
## $ GarageType : Factor w/ 6 levels "2Types","Attchd",..: 2 2 2 2 2 2 2 2 2 2 ...  
## $ GarageYrBlt : int 1961 1958 1997 1998 1992 1993 1992 1998 1990 1970 ...  
## $ GarageFinish : Factor w/ 3 levels "Fin","RFn","Unf": 3 3 1 1 2 1 1 1 3 1 ...  
## $ GarageCars : int 1 1 2 2 2 2 2 2 2 2 ...  
## $ GarageArea : int 730 312 482 470 506 440 420 393 506 525 ...  
## $ GarageQual : Factor w/ 4 levels "Fa","Gd","Po",..: 4 4 4 4 4 4 4 4 4 4 ...  
## $ 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 140 393 212 360 0 157 483 0 192 240 ...  
## $ OpenPorchSF : int 0 36 34 36 82 84 21 75 0 0 ...  
## $ EnclosedPorch: int 0 0 0 0 0 0 0 0 0 0 ...  
## $ X3SsnPorch : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ ScreenPorch : int 120 0 0 0 144 0 0 0 0 0 ...  
## $ PoolArea : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ PoolQC : Factor w/ 2 levels "Ex","Gd": NA NA NA NA NA NA NA NA NA NA ...  
## $ Fence : Factor w/ 4 levels "GdPrv","GdWo",..: 3 NA 3 NA NA NA 1 NA NA 3 ...  
## $ MiscFeature : Factor w/ 3 levels "Gar2","Othr",..: NA 1 NA NA NA NA 3 NA NA NA ...  
## $ MiscVal : int 0 12500 0 0 0 0 500 0 0 0 ...  
## $ MoSold : int 6 6 3 6 1 4 3 5 2 4 ...  
## $ YrSold : int 2010 2010 2010 2010 2010 2010 2010 2010 2010 2010 ...  
## $ 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 5 5 5 5 5 5 5 ...

summary(testing)

## Id MSSubClass MSZoning LotFrontage   
## Min. :1461 Min. : 20.00 C (all): 15 Min. : 21.00   
## 1st Qu.:1826 1st Qu.: 20.00 FV : 74 1st Qu.: 58.00   
## Median :2190 Median : 50.00 RH : 10 Median : 67.00   
## Mean :2190 Mean : 57.38 RL :1114 Mean : 68.58   
## 3rd Qu.:2554 3rd Qu.: 70.00 RM : 242 3rd Qu.: 80.00   
## Max. :2919 Max. :190.00 NA's : 4 Max. :200.00   
## NA's :227   
## LotArea Street Alley LotShape LandContour  
## Min. : 1470 Grvl: 6 Grvl: 70 IR1:484 Bnk: 54   
## 1st Qu.: 7391 Pave:1453 Pave: 37 IR2: 35 HLS: 70   
## Median : 9399 NA's:1352 IR3: 6 Low: 24   
## Mean : 9819 Reg:934 Lvl:1311   
## 3rd Qu.:11518   
## Max. :56600   
##   
## Utilities LotConfig LandSlope Neighborhood Condition1   
## AllPub:1457 Corner : 248 Gtl:1396 NAmes :218 Norm :1251   
## NA's : 2 CulDSac: 82 Mod: 60 OldTown:126 Feedr : 83   
## FR2 : 38 Sev: 3 CollgCr:117 Artery : 44   
## FR3 : 10 Somerst: 96 RRAn : 24   
## Inside :1081 Edwards: 94 PosN : 20   
## NridgHt: 89 RRAe : 17   
## (Other):719 (Other): 20   
## Condition2 BldgType HouseStyle OverallQual OverallCond   
## Artery: 3 1Fam :1205 1.5Fin:160 Min. : 1.000 Min. :1.000   
## Feedr : 7 2fmCon: 31 1.5Unf: 5 1st Qu.: 5.000 1st Qu.:5.000   
## Norm :1444 Duplex: 57 1Story:745 Median : 6.000 Median :5.000   
## PosA : 3 Twnhs : 53 2.5Unf: 13 Mean : 6.079 Mean :5.554   
## PosN : 2 TwnhsE: 113 2Story:427 3rd Qu.: 7.000 3rd Qu.:6.000   
## SFoyer: 46 Max. :10.000 Max. :9.000   
## SLvl : 63   
## YearBuilt YearRemodAdd RoofStyle RoofMatl Exterior1st   
## Min. :1879 Min. :1950 Flat : 7 CompShg:1442 VinylSd:510   
## 1st Qu.:1953 1st Qu.:1963 Gable :1169 Tar&Grv: 12 MetalSd:230   
## Median :1973 Median :1992 Gambrel: 11 WdShake: 4 HdBoard:220   
## Mean :1971 Mean :1984 Hip : 265 WdShngl: 1 Wd Sdng:205   
## 3rd Qu.:2001 3rd Qu.:2004 Mansard: 4 Plywood:113   
## Max. :2010 Max. :2010 Shed : 3 (Other):180   
## NA's : 1   
## Exterior2nd MasVnrType MasVnrArea ExterQual ExterCond  
## VinylSd:510 BrkCmn : 10 Min. : 0.0 Ex: 55 Ex: 9   
## MetalSd:233 BrkFace:434 1st Qu.: 0.0 Fa: 21 Fa: 39   
## HdBoard:199 None :878 Median : 0.0 Gd:491 Gd: 153   
## Wd Sdng:194 Stone :121 Mean : 100.7 TA:892 Po: 2   
## Plywood:128 NA's : 16 3rd Qu.: 164.0 TA:1256   
## (Other):194 Max. :1290.0   
## NA's : 1 NA's :15   
## Foundation BsmtQual BsmtCond BsmtExposure BsmtFinType1  
## BrkTil:165 Ex :137 Fa : 59 Av :197 ALQ :209   
## CBlock:601 Fa : 53 Gd : 57 Gd :142 BLQ :121   
## PConc :661 Gd :591 Po : 3 Mn :125 GLQ :431   
## Slab : 25 TA :634 TA :1295 No :951 LwQ : 80   
## Stone : 5 NA's: 44 NA's: 45 NA's: 44 Rec :155   
## Wood : 2 Unf :421   
## NA's: 42   
## BsmtFinSF1 BsmtFinType2 BsmtFinSF2 BsmtUnfSF   
## Min. : 0.0 ALQ : 33 Min. : 0.00 Min. : 0.0   
## 1st Qu.: 0.0 BLQ : 35 1st Qu.: 0.00 1st Qu.: 219.2   
## Median : 350.5 GLQ : 20 Median : 0.00 Median : 460.0   
## Mean : 439.2 LwQ : 41 Mean : 52.62 Mean : 554.3   
## 3rd Qu.: 753.5 Rec : 51 3rd Qu.: 0.00 3rd Qu.: 797.8   
## Max. :4010.0 Unf :1237 Max. :1526.00 Max. :2140.0   
## NA's :1 NA's: 42 NA's :1 NA's :1   
## TotalBsmtSF Heating HeatingQC CentralAir Electrical   
## Min. : 0 GasA:1446 Ex:752 N: 101 FuseA: 94   
## 1st Qu.: 784 GasW: 9 Fa: 43 Y:1358 FuseF: 23   
## Median : 988 Grav: 2 Gd:233 FuseP: 5   
## Mean :1046 Wall: 2 Po: 2 SBrkr:1337   
## 3rd Qu.:1305 TA:429   
## Max. :5095   
## NA's :1   
## X1stFlrSF X2ndFlrSF LowQualFinSF GrLivArea   
## Min. : 407.0 Min. : 0 Min. : 0.000 Min. : 407   
## 1st Qu.: 873.5 1st Qu.: 0 1st Qu.: 0.000 1st Qu.:1118   
## Median :1079.0 Median : 0 Median : 0.000 Median :1432   
## Mean :1156.5 Mean : 326 Mean : 3.543 Mean :1486   
## 3rd Qu.:1382.5 3rd Qu.: 676 3rd Qu.: 0.000 3rd Qu.:1721   
## Max. :5095.0 Max. :1862 Max. :1064.000 Max. :5095   
##   
## BsmtFullBath BsmtHalfBath FullBath HalfBath   
## Min. :0.0000 Min. :0.0000 Min. :0.000 Min. :0.0000   
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:1.000 1st Qu.:0.0000   
## Median :0.0000 Median :0.0000 Median :2.000 Median :0.0000   
## Mean :0.4345 Mean :0.0652 Mean :1.571 Mean :0.3777   
## 3rd Qu.:1.0000 3rd Qu.:0.0000 3rd Qu.:2.000 3rd Qu.:1.0000   
## Max. :3.0000 Max. :2.0000 Max. :4.000 Max. :2.0000   
## NA's :2 NA's :2   
## BedroomAbvGr KitchenAbvGr KitchenQual TotRmsAbvGrd   
## Min. :0.000 Min. :0.000 Ex :105 Min. : 3.000   
## 1st Qu.:2.000 1st Qu.:1.000 Fa : 31 1st Qu.: 5.000   
## Median :3.000 Median :1.000 Gd :565 Median : 6.000   
## Mean :2.854 Mean :1.042 TA :757 Mean : 6.385   
## 3rd Qu.:3.000 3rd Qu.:1.000 NA's: 1 3rd Qu.: 7.000   
## Max. :6.000 Max. :2.000 Max. :15.000   
##   
## Functional Fireplaces FireplaceQu GarageType GarageYrBlt   
## Typ :1357 Min. :0.0000 Ex : 19 2Types : 17 Min. :1895   
## Min2 : 36 1st Qu.:0.0000 Fa : 41 Attchd :853 1st Qu.:1959   
## Min1 : 34 Median :0.0000 Gd :364 Basment: 17 Median :1979   
## Mod : 20 Mean :0.5812 Po : 26 BuiltIn: 98 Mean :1978   
## Maj1 : 5 3rd Qu.:1.0000 TA :279 CarPort: 6 3rd Qu.:2002   
## (Other): 5 Max. :4.0000 NA's:730 Detchd :392 Max. :2207   
## NA's : 2 NA's : 76 NA's :78   
## GarageFinish GarageCars GarageArea GarageQual GarageCond   
## Fin :367 Min. :0.000 Min. : 0.0 Fa : 76 Ex : 1   
## RFn :389 1st Qu.:1.000 1st Qu.: 318.0 Gd : 10 Fa : 39   
## Unf :625 Median :2.000 Median : 480.0 Po : 2 Gd : 6   
## NA's: 78 Mean :1.766 Mean : 472.8 TA :1293 Po : 7   
## 3rd Qu.:2.000 3rd Qu.: 576.0 NA's: 78 TA :1328   
## Max. :5.000 Max. :1488.0 NA's: 78   
## NA's :1 NA's :1   
## PavedDrive WoodDeckSF OpenPorchSF EnclosedPorch   
## N: 126 Min. : 0.00 Min. : 0.00 Min. : 0.00   
## P: 32 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.00   
## Y:1301 Median : 0.00 Median : 28.00 Median : 0.00   
## Mean : 93.17 Mean : 48.31 Mean : 24.24   
## 3rd Qu.: 168.00 3rd Qu.: 72.00 3rd Qu.: 0.00   
## Max. :1424.00 Max. :742.00 Max. :1012.00   
##   
## X3SsnPorch ScreenPorch PoolArea PoolQC   
## Min. : 0.000 Min. : 0.00 Min. : 0.000 Ex : 2   
## 1st Qu.: 0.000 1st Qu.: 0.00 1st Qu.: 0.000 Gd : 1   
## Median : 0.000 Median : 0.00 Median : 0.000 NA's:1456   
## Mean : 1.794 Mean : 17.06 Mean : 1.744   
## 3rd Qu.: 0.000 3rd Qu.: 0.00 3rd Qu.: 0.000   
## Max. :360.000 Max. :576.00 Max. :800.000   
##   
## Fence MiscFeature MiscVal MoSold   
## GdPrv: 59 Gar2: 3 Min. : 0.00 Min. : 1.000   
## GdWo : 58 Othr: 2 1st Qu.: 0.00 1st Qu.: 4.000   
## MnPrv: 172 Shed: 46 Median : 0.00 Median : 6.000   
## MnWw : 1 NA's:1408 Mean : 58.17 Mean : 6.104   
## NA's :1169 3rd Qu.: 0.00 3rd Qu.: 8.000   
## Max. :17000.00 Max. :12.000   
##   
## YrSold SaleType SaleCondition   
## Min. :2006 WD :1258 Abnorml: 89   
## 1st Qu.:2007 New : 117 AdjLand: 8   
## Median :2008 COD : 44 Alloca : 12   
## Mean :2008 ConLD : 17 Family : 26   
## 3rd Qu.:2009 CWD : 8 Normal :1204   
## Max. :2010 (Other): 14 Partial: 120   
## NA's : 1

#Since the train and test sets are divided into half we are combing them into one single dataset for further modeling purposes  
housedata<-bind\_rows(training,testing)

## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector

## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector

View(housedata)  
head(housedata)

## Id MSSubClass MSZoning LotFrontage LotArea Street Alley LotShape  
## 1 1 60 RL 65 8450 Pave <NA> Reg  
## 2 2 20 RL 80 9600 Pave <NA> Reg  
## 3 3 60 RL 68 11250 Pave <NA> IR1  
## 4 4 70 RL 60 9550 Pave <NA> IR1  
## 5 5 60 RL 84 14260 Pave <NA> IR1  
## 6 6 50 RL 85 14115 Pave <NA> IR1  
## LandContour Utilities LotConfig LandSlope Neighborhood Condition1  
## 1 Lvl AllPub Inside Gtl CollgCr Norm  
## 2 Lvl AllPub FR2 Gtl Veenker Feedr  
## 3 Lvl AllPub Inside Gtl CollgCr Norm  
## 4 Lvl AllPub Corner Gtl Crawfor Norm  
## 5 Lvl AllPub FR2 Gtl NoRidge Norm  
## 6 Lvl AllPub Inside Gtl Mitchel Norm  
## Condition2 BldgType HouseStyle OverallQual OverallCond YearBuilt  
## 1 Norm 1Fam 2Story 7 5 2003  
## 2 Norm 1Fam 1Story 6 8 1976  
## 3 Norm 1Fam 2Story 7 5 2001  
## 4 Norm 1Fam 2Story 7 5 1915  
## 5 Norm 1Fam 2Story 8 5 2000  
## 6 Norm 1Fam 1.5Fin 5 5 1993  
## YearRemodAdd RoofStyle RoofMatl Exterior1st Exterior2nd MasVnrType  
## 1 2003 Gable CompShg VinylSd VinylSd BrkFace  
## 2 1976 Gable CompShg MetalSd MetalSd None  
## 3 2002 Gable CompShg VinylSd VinylSd BrkFace  
## 4 1970 Gable CompShg Wd Sdng Wd Shng None  
## 5 2000 Gable CompShg VinylSd VinylSd BrkFace  
## 6 1995 Gable CompShg VinylSd VinylSd None  
## MasVnrArea ExterQual ExterCond Foundation BsmtQual BsmtCond BsmtExposure  
## 1 196 Gd TA PConc Gd TA No  
## 2 0 TA TA CBlock Gd TA Gd  
## 3 162 Gd TA PConc Gd TA Mn  
## 4 0 TA TA BrkTil TA Gd No  
## 5 350 Gd TA PConc Gd TA Av  
## 6 0 TA TA Wood Gd TA No  
## BsmtFinType1 BsmtFinSF1 BsmtFinType2 BsmtFinSF2 BsmtUnfSF TotalBsmtSF  
## 1 GLQ 706 Unf 0 150 856  
## 2 ALQ 978 Unf 0 284 1262  
## 3 GLQ 486 Unf 0 434 920  
## 4 ALQ 216 Unf 0 540 756  
## 5 GLQ 655 Unf 0 490 1145  
## 6 GLQ 732 Unf 0 64 796  
## Heating HeatingQC CentralAir Electrical X1stFlrSF X2ndFlrSF LowQualFinSF  
## 1 GasA Ex Y SBrkr 856 854 0  
## 2 GasA Ex Y SBrkr 1262 0 0  
## 3 GasA Ex Y SBrkr 920 866 0  
## 4 GasA Gd Y SBrkr 961 756 0  
## 5 GasA Ex Y SBrkr 1145 1053 0  
## 6 GasA Ex Y SBrkr 796 566 0  
## GrLivArea BsmtFullBath BsmtHalfBath FullBath HalfBath BedroomAbvGr  
## 1 1710 1 0 2 1 3  
## 2 1262 0 1 2 0 3  
## 3 1786 1 0 2 1 3  
## 4 1717 1 0 1 0 3  
## 5 2198 1 0 2 1 4  
## 6 1362 1 0 1 1 1  
## KitchenAbvGr KitchenQual TotRmsAbvGrd Functional Fireplaces FireplaceQu  
## 1 1 Gd 8 Typ 0 <NA>  
## 2 1 TA 6 Typ 1 TA  
## 3 1 Gd 6 Typ 1 TA  
## 4 1 Gd 7 Typ 1 Gd  
## 5 1 Gd 9 Typ 1 TA  
## 6 1 TA 5 Typ 0 <NA>  
## GarageType GarageYrBlt GarageFinish GarageCars GarageArea GarageQual  
## 1 Attchd 2003 RFn 2 548 TA  
## 2 Attchd 1976 RFn 2 460 TA  
## 3 Attchd 2001 RFn 2 608 TA  
## 4 Detchd 1998 Unf 3 642 TA  
## 5 Attchd 2000 RFn 3 836 TA  
## 6 Attchd 1993 Unf 2 480 TA  
## GarageCond PavedDrive WoodDeckSF OpenPorchSF EnclosedPorch X3SsnPorch  
## 1 TA Y 0 61 0 0  
## 2 TA Y 298 0 0 0  
## 3 TA Y 0 42 0 0  
## 4 TA Y 0 35 272 0  
## 5 TA Y 192 84 0 0  
## 6 TA Y 40 30 0 320  
## ScreenPorch PoolArea PoolQC Fence MiscFeature MiscVal MoSold YrSold  
## 1 0 0 <NA> <NA> <NA> 0 2 2008  
## 2 0 0 <NA> <NA> <NA> 0 5 2007  
## 3 0 0 <NA> <NA> <NA> 0 9 2008  
## 4 0 0 <NA> <NA> <NA> 0 2 2006  
## 5 0 0 <NA> <NA> <NA> 0 12 2008  
## 6 0 0 <NA> MnPrv Shed 700 10 2009  
## SaleType SaleCondition SalePrice  
## 1 WD Normal 208500  
## 2 WD Normal 181500  
## 3 WD Normal 223500  
## 4 WD Abnorml 140000  
## 5 WD Normal 250000  
## 6 WD Normal 143000

dim(housedata)

## [1] 2919 81

str(housedata)

## 'data.frame': 2919 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 : chr "RL" "RL" "RL" "RL" ...  
## $ 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 : chr "Pave" "Pave" "Pave" "Pave" ...  
## $ Alley : chr NA NA NA NA ...  
## $ LotShape : chr "Reg" "Reg" "IR1" "IR1" ...  
## $ LandContour : chr "Lvl" "Lvl" "Lvl" "Lvl" ...  
## $ Utilities : chr "AllPub" "AllPub" "AllPub" "AllPub" ...  
## $ LotConfig : chr "Inside" "FR2" "Inside" "Corner" ...  
## $ LandSlope : chr "Gtl" "Gtl" "Gtl" "Gtl" ...  
## $ Neighborhood : chr "CollgCr" "Veenker" "CollgCr" "Crawfor" ...  
## $ Condition1 : chr "Norm" "Feedr" "Norm" "Norm" ...  
## $ Condition2 : chr "Norm" "Norm" "Norm" "Norm" ...  
## $ BldgType : chr "1Fam" "1Fam" "1Fam" "1Fam" ...  
## $ HouseStyle : chr "2Story" "1Story" "2Story" "2Story" ...  
## $ OverallQual : int 7 6 7 7 8 5 8 7 7 5 ...  
## $ OverallCond : int 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 : chr "Gable" "Gable" "Gable" "Gable" ...  
## $ RoofMatl : chr "CompShg" "CompShg" "CompShg" "CompShg" ...  
## $ Exterior1st : chr "VinylSd" "MetalSd" "VinylSd" "Wd Sdng" ...  
## $ Exterior2nd : chr "VinylSd" "MetalSd" "VinylSd" "Wd Shng" ...  
## $ MasVnrType : chr "BrkFace" "None" "BrkFace" "None" ...  
## $ MasVnrArea : int 196 0 162 0 350 0 186 240 0 0 ...  
## $ ExterQual : chr "Gd" "TA" "Gd" "TA" ...  
## $ ExterCond : chr "TA" "TA" "TA" "TA" ...  
## $ Foundation : chr "PConc" "CBlock" "PConc" "BrkTil" ...  
## $ BsmtQual : chr "Gd" "Gd" "Gd" "TA" ...  
## $ BsmtCond : chr "TA" "TA" "TA" "Gd" ...  
## $ BsmtExposure : chr "No" "Gd" "Mn" "No" ...  
## $ BsmtFinType1 : chr "GLQ" "ALQ" "GLQ" "ALQ" ...  
## $ BsmtFinSF1 : int 706 978 486 216 655 732 1369 859 0 851 ...  
## $ BsmtFinType2 : chr "Unf" "Unf" "Unf" "Unf" ...  
## $ 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 : chr "GasA" "GasA" "GasA" "GasA" ...  
## $ HeatingQC : chr "Ex" "Ex" "Ex" "Gd" ...  
## $ CentralAir : chr "Y" "Y" "Y" "Y" ...  
## $ Electrical : chr "SBrkr" "SBrkr" "SBrkr" "SBrkr" ...  
## $ 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 : chr "Gd" "TA" "Gd" "Gd" ...  
## $ TotRmsAbvGrd : int 8 6 6 7 9 5 7 7 8 5 ...  
## $ Functional : chr "Typ" "Typ" "Typ" "Typ" ...  
## $ Fireplaces : int 0 1 1 1 1 0 1 2 2 2 ...  
## $ FireplaceQu : chr NA "TA" "TA" "Gd" ...  
## $ GarageType : chr "Attchd" "Attchd" "Attchd" "Detchd" ...  
## $ GarageYrBlt : int 2003 1976 2001 1998 2000 1993 2004 1973 1931 1939 ...  
## $ GarageFinish : chr "RFn" "RFn" "RFn" "Unf" ...  
## $ GarageCars : int 2 2 2 3 3 2 2 2 2 1 ...  
## $ GarageArea : int 548 460 608 642 836 480 636 484 468 205 ...  
## $ GarageQual : chr "TA" "TA" "TA" "TA" ...  
## $ GarageCond : chr "TA" "TA" "TA" "TA" ...  
## $ PavedDrive : chr "Y" "Y" "Y" "Y" ...  
## $ 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 : chr NA NA NA NA ...  
## $ Fence : chr NA NA NA NA ...  
## $ MiscFeature : chr NA NA 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 : chr "WD" "WD" "WD" "WD" ...  
## $ SaleCondition: chr "Normal" "Normal" "Normal" "Abnorml" ...  
## $ SalePrice : int 208500 181500 223500 140000 250000 143000 307000 200000 129900 118000 ...

summary(housedata)

## Id MSSubClass MSZoning LotFrontage   
## Min. : 1.0 Min. : 20.00 Length:2919 Min. : 21.00   
## 1st Qu.: 730.5 1st Qu.: 20.00 Class :character 1st Qu.: 59.00   
## Median :1460.0 Median : 50.00 Mode :character Median : 68.00   
## Mean :1460.0 Mean : 57.14 Mean : 69.31   
## 3rd Qu.:2189.5 3rd Qu.: 70.00 3rd Qu.: 80.00   
## Max. :2919.0 Max. :190.00 Max. :313.00   
## NA's :486   
## LotArea Street Alley LotShape   
## Min. : 1300 Length:2919 Length:2919 Length:2919   
## 1st Qu.: 7478 Class :character Class :character Class :character   
## Median : 9453 Mode :character Mode :character Mode :character   
## Mean : 10168   
## 3rd Qu.: 11570   
## Max. :215245   
##   
## LandContour Utilities LotConfig   
## Length:2919 Length:2919 Length:2919   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## LandSlope Neighborhood Condition1   
## Length:2919 Length:2919 Length:2919   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## Condition2 BldgType HouseStyle OverallQual   
## Length:2919 Length:2919 Length:2919 Min. : 1.000   
## Class :character Class :character Class :character 1st Qu.: 5.000   
## Mode :character Mode :character Mode :character Median : 6.000   
## Mean : 6.089   
## 3rd Qu.: 7.000   
## Max. :10.000   
##   
## OverallCond YearBuilt YearRemodAdd RoofStyle   
## Min. :1.000 Min. :1872 Min. :1950 Length:2919   
## 1st Qu.:5.000 1st Qu.:1954 1st Qu.:1965 Class :character   
## Median :5.000 Median :1973 Median :1993 Mode :character   
## Mean :5.565 Mean :1971 Mean :1984   
## 3rd Qu.:6.000 3rd Qu.:2001 3rd Qu.:2004   
## Max. :9.000 Max. :2010 Max. :2010   
##   
## RoofMatl Exterior1st Exterior2nd   
## Length:2919 Length:2919 Length:2919   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## MasVnrType MasVnrArea ExterQual ExterCond   
## Length:2919 Min. : 0.0 Length:2919 Length:2919   
## Class :character 1st Qu.: 0.0 Class :character Class :character   
## Mode :character Median : 0.0 Mode :character Mode :character   
## Mean : 102.2   
## 3rd Qu.: 164.0   
## Max. :1600.0   
## NA's :23   
## Foundation BsmtQual BsmtCond   
## Length:2919 Length:2919 Length:2919   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## BsmtExposure BsmtFinType1 BsmtFinSF1 BsmtFinType2   
## Length:2919 Length:2919 Min. : 0.0 Length:2919   
## Class :character Class :character 1st Qu.: 0.0 Class :character   
## Mode :character Mode :character Median : 368.5 Mode :character   
## Mean : 441.4   
## 3rd Qu.: 733.0   
## Max. :5644.0   
## NA's :1   
## BsmtFinSF2 BsmtUnfSF TotalBsmtSF Heating   
## Min. : 0.00 Min. : 0.0 Min. : 0.0 Length:2919   
## 1st Qu.: 0.00 1st Qu.: 220.0 1st Qu.: 793.0 Class :character   
## Median : 0.00 Median : 467.0 Median : 989.5 Mode :character   
## Mean : 49.58 Mean : 560.8 Mean :1051.8   
## 3rd Qu.: 0.00 3rd Qu.: 805.5 3rd Qu.:1302.0   
## Max. :1526.00 Max. :2336.0 Max. :6110.0   
## NA's :1 NA's :1 NA's :1   
## HeatingQC CentralAir Electrical X1stFlrSF   
## Length:2919 Length:2919 Length:2919 Min. : 334   
## Class :character Class :character Class :character 1st Qu.: 876   
## Mode :character Mode :character Mode :character Median :1082   
## Mean :1160   
## 3rd Qu.:1388   
## Max. :5095   
##   
## X2ndFlrSF LowQualFinSF GrLivArea BsmtFullBath   
## Min. : 0.0 Min. : 0.000 Min. : 334 Min. :0.0000   
## 1st Qu.: 0.0 1st Qu.: 0.000 1st Qu.:1126 1st Qu.:0.0000   
## Median : 0.0 Median : 0.000 Median :1444 Median :0.0000   
## Mean : 336.5 Mean : 4.694 Mean :1501 Mean :0.4299   
## 3rd Qu.: 704.0 3rd Qu.: 0.000 3rd Qu.:1744 3rd Qu.:1.0000   
## Max. :2065.0 Max. :1064.000 Max. :5642 Max. :3.0000   
## NA's :2   
## BsmtHalfBath FullBath HalfBath BedroomAbvGr   
## Min. :0.00000 Min. :0.000 Min. :0.0000 Min. :0.00   
## 1st Qu.:0.00000 1st Qu.:1.000 1st Qu.:0.0000 1st Qu.:2.00   
## Median :0.00000 Median :2.000 Median :0.0000 Median :3.00   
## Mean :0.06136 Mean :1.568 Mean :0.3803 Mean :2.86   
## 3rd Qu.:0.00000 3rd Qu.:2.000 3rd Qu.:1.0000 3rd Qu.:3.00   
## Max. :2.00000 Max. :4.000 Max. :2.0000 Max. :8.00   
## NA's :2   
## KitchenAbvGr KitchenQual TotRmsAbvGrd Functional   
## Min. :0.000 Length:2919 Min. : 2.000 Length:2919   
## 1st Qu.:1.000 Class :character 1st Qu.: 5.000 Class :character   
## Median :1.000 Mode :character Median : 6.000 Mode :character   
## Mean :1.045 Mean : 6.452   
## 3rd Qu.:1.000 3rd Qu.: 7.000   
## Max. :3.000 Max. :15.000   
##   
## Fireplaces FireplaceQu GarageType GarageYrBlt   
## Min. :0.0000 Length:2919 Length:2919 Min. :1895   
## 1st Qu.:0.0000 Class :character Class :character 1st Qu.:1960   
## Median :1.0000 Mode :character Mode :character Median :1979   
## Mean :0.5971 Mean :1978   
## 3rd Qu.:1.0000 3rd Qu.:2002   
## Max. :4.0000 Max. :2207   
## NA's :159   
## GarageFinish GarageCars GarageArea GarageQual   
## Length:2919 Min. :0.000 Min. : 0.0 Length:2919   
## Class :character 1st Qu.:1.000 1st Qu.: 320.0 Class :character   
## Mode :character Median :2.000 Median : 480.0 Mode :character   
## Mean :1.767 Mean : 472.9   
## 3rd Qu.:2.000 3rd Qu.: 576.0   
## Max. :5.000 Max. :1488.0   
## NA's :1 NA's :1   
## GarageCond PavedDrive WoodDeckSF OpenPorchSF   
## Length:2919 Length:2919 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 : 26.00   
## Mean : 93.71 Mean : 47.49   
## 3rd Qu.: 168.00 3rd Qu.: 70.00   
## Max. :1424.00 Max. :742.00   
##   
## EnclosedPorch X3SsnPorch ScreenPorch PoolArea   
## Min. : 0.0 Min. : 0.000 Min. : 0.00 Min. : 0.000   
## 1st Qu.: 0.0 1st Qu.: 0.000 1st Qu.: 0.00 1st Qu.: 0.000   
## Median : 0.0 Median : 0.000 Median : 0.00 Median : 0.000   
## Mean : 23.1 Mean : 2.602 Mean : 16.06 Mean : 2.252   
## 3rd Qu.: 0.0 3rd Qu.: 0.000 3rd Qu.: 0.00 3rd Qu.: 0.000   
## Max. :1012.0 Max. :508.000 Max. :576.00 Max. :800.000   
##   
## PoolQC Fence MiscFeature   
## Length:2919 Length:2919 Length:2919   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## MiscVal MoSold YrSold SaleType   
## Min. : 0.00 Min. : 1.000 Min. :2006 Length:2919   
## 1st Qu.: 0.00 1st Qu.: 4.000 1st Qu.:2007 Class :character   
## Median : 0.00 Median : 6.000 Median :2008 Mode :character   
## Mean : 50.83 Mean : 6.213 Mean :2008   
## 3rd Qu.: 0.00 3rd Qu.: 8.000 3rd Qu.:2009   
## Max. :17000.00 Max. :12.000 Max. :2010   
##   
## SaleCondition SalePrice   
## Length:2919 Min. : 34900   
## Class :character 1st Qu.:129975   
## Mode :character Median :163000   
## Mean :180921   
## 3rd Qu.:214000   
## Max. :755000   
## NA's :1459

#Find the total percentage of missing data  
sum(is.na(housedata)/(nrow(housedata)\*nrow(housedata)))

## [1] 0.001810209

#Check for duplicate rows  
unique(nrow(housedata))

## [1] 2919

#Check for N/A values all the columns  
colSums(sapply(housedata,is.na))

## Id MSSubClass MSZoning LotFrontage LotArea   
## 0 0 4 486 0   
## Street Alley LotShape LandContour Utilities   
## 0 2721 0 0 2   
## LotConfig LandSlope Neighborhood Condition1 Condition2   
## 0 0 0 0 0   
## BldgType HouseStyle OverallQual OverallCond YearBuilt   
## 0 0 0 0 0   
## YearRemodAdd RoofStyle RoofMatl Exterior1st Exterior2nd   
## 0 0 0 1 1   
## MasVnrType MasVnrArea ExterQual ExterCond Foundation   
## 24 23 0 0 0   
## BsmtQual BsmtCond BsmtExposure BsmtFinType1 BsmtFinSF1   
## 81 82 82 79 1   
## BsmtFinType2 BsmtFinSF2 BsmtUnfSF TotalBsmtSF Heating   
## 80 1 1 1 0   
## HeatingQC CentralAir Electrical X1stFlrSF X2ndFlrSF   
## 0 0 1 0 0   
## LowQualFinSF GrLivArea BsmtFullBath BsmtHalfBath FullBath   
## 0 0 2 2 0   
## HalfBath BedroomAbvGr KitchenAbvGr KitchenQual TotRmsAbvGrd   
## 0 0 0 1 0   
## Functional Fireplaces FireplaceQu GarageType GarageYrBlt   
## 2 0 1420 157 159   
## GarageFinish GarageCars GarageArea GarageQual GarageCond   
## 159 1 1 159 159   
## PavedDrive WoodDeckSF OpenPorchSF EnclosedPorch X3SsnPorch   
## 0 0 0 0 0   
## ScreenPorch PoolArea PoolQC Fence MiscFeature   
## 0 0 2909 2348 2814   
## MiscVal MoSold YrSold SaleType SaleCondition   
## 0 0 0 1 0   
## SalePrice   
## 1459

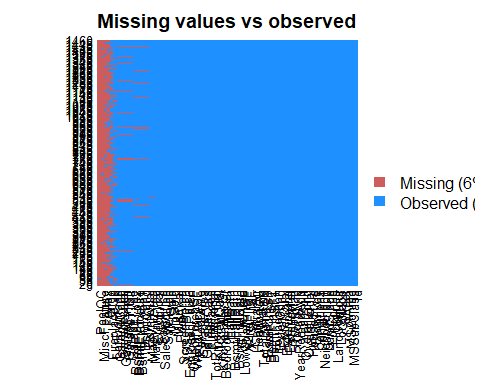
library(Amelia)

## Warning: package 'Amelia' was built under R version 3.5.2

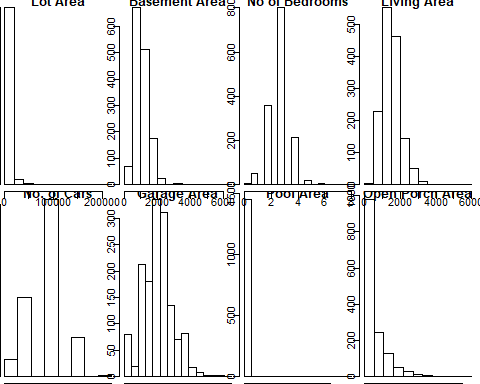
## Loading required package: Rcpp

## ##   
## ## Amelia II: Multiple Imputation  
## ## (Version 1.7.5, built: 2018-05-07)  
## ## Copyright (C) 2005-2019 James Honaker, Gary King and Matthew Blackwell  
## ## Refer to http://gking.harvard.edu/amelia/ for more information  
## ##

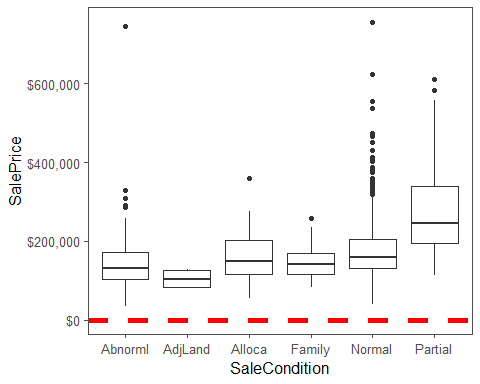
missmap(training, main ="Missing values vs observed")



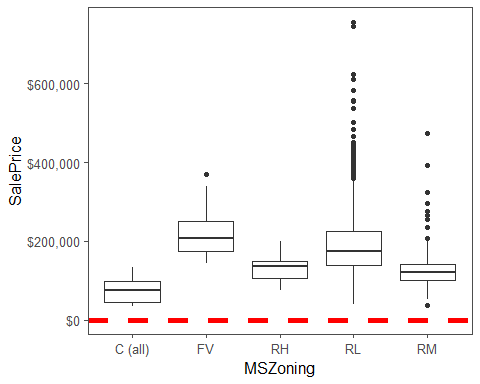
#Choosing the categorical and numerical variables for our analysis  
catvar <- c('MSZoning','Street', 'Neighborhood', 'LandContour','BldgType', 'LandSlope', 'RoofStyle',  
 'HouseStyle','CentralAir','PavedDrive','SaleCondition','OverallCond' )  
numvar<-c('SalePrice','LotArea','TotalBsmtSF','GrLivArea','BedroomAbvGr','TotRmsAbvGrd','GarageCars','GarageArea'  
 ,'OpenPorchSF','EnclosedPorch','WoodDeckSF','PoolArea')  
  
attach(housedata)  
catdf<-housedata[,catvar]  
numdf<-housedata[,numvar]  
  
#Check for distribution of variables that could attribute for higher Sale Price  
par(mfrow=c(2,4))  
hist(training$LotArea,xlab="Lot Area", main="Lot Area")  
hist(training$TotalBsmtSF, xlab="Basement Area", main="Basement Area")  
hist(training$BedroomAbvGr, xlab="No of Bedrooms", main="No of Bedrooms")  
hist(training$GrLivArea, xlab="Living Area",main="Living Area")  
hist(training$GarageCars, xlab="No. of Cars",main="No. of Cars")  
hist(training$GarageArea, xlab="Garage Area",main="Garage Area")  
hist(training$PoolArea, xlab="Pool Area",main="Pool Area")  
hist(training$OpenPorchSF, xlab="Open Porch Area",main="Open Porch Area")



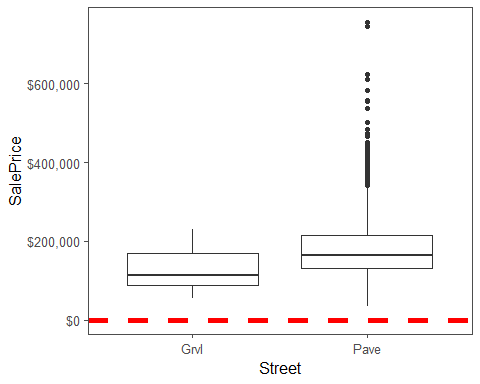
#Check how Sale price varies as per different Sale Conditions  
ggplot(training, aes(x = SaleCondition, y = SalePrice)) +geom\_boxplot() +  
 geom\_hline(aes(yintercept=80),   
 colour='red', linetype='dashed', lwd=2) +  
 scale\_y\_continuous(labels=dollar\_format()) +  
 theme\_few()



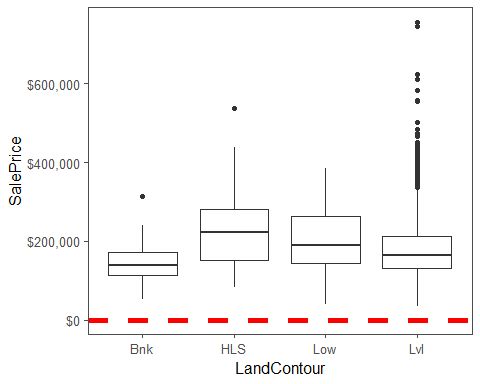
ggplot(training, aes(x = MSZoning, y = SalePrice)) +geom\_boxplot() +  
 geom\_hline(aes(yintercept=80),   
 colour='red', linetype='dashed', lwd=2) +  
 scale\_y\_continuous(labels=dollar\_format()) +  
 theme\_few()



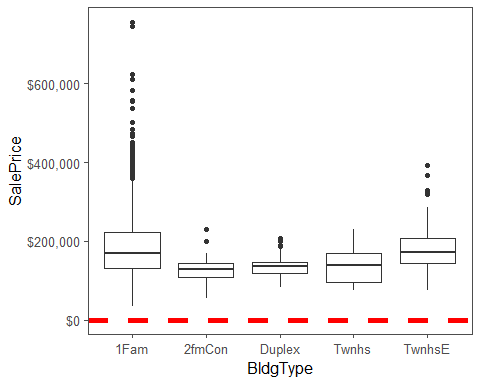
ggplot(training, aes(x = Street, y = SalePrice)) +geom\_boxplot() +  
 geom\_hline(aes(yintercept=80),   
 colour='red', linetype='dashed', lwd=2) +  
 scale\_y\_continuous(labels=dollar\_format()) +  
 theme\_few()



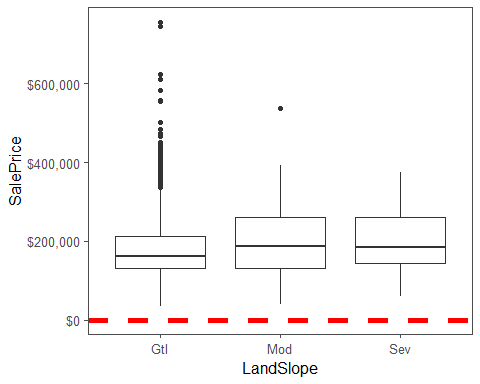
ggplot(training, aes(x = LandContour, y = SalePrice)) +geom\_boxplot() +  
 geom\_hline(aes(yintercept=80),   
 colour='red', linetype='dashed', lwd=2) +  
 scale\_y\_continuous(labels=dollar\_format()) +  
 theme\_few()



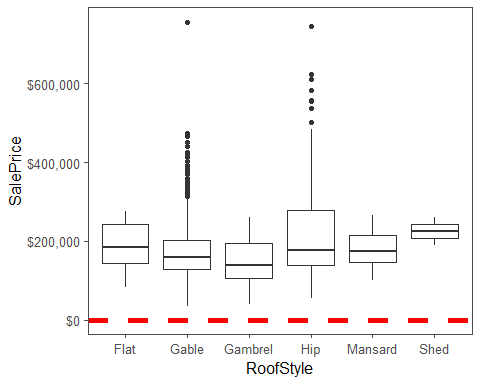
ggplot(training, aes(x = BldgType, y = SalePrice)) +geom\_boxplot() +  
 geom\_hline(aes(yintercept=80),   
 colour='red', linetype='dashed', lwd=2) +  
 scale\_y\_continuous(labels=dollar\_format()) +  
 theme\_few()



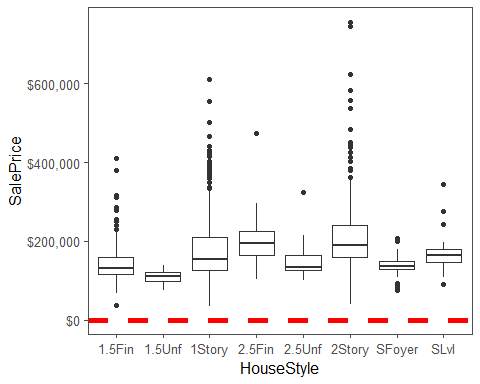
ggplot(training, aes(x = LandSlope, y = SalePrice)) +geom\_boxplot() +  
 geom\_hline(aes(yintercept=80),   
 colour='red', linetype='dashed', lwd=2) +  
 scale\_y\_continuous(labels=dollar\_format()) +  
 theme\_few()



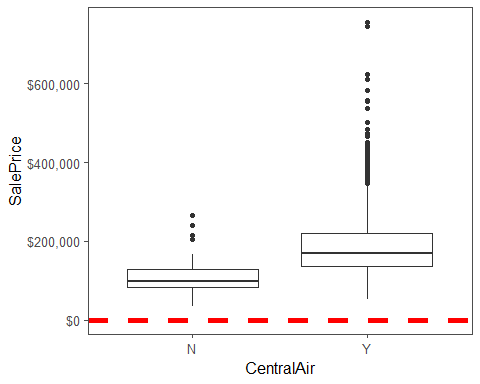
ggplot(training, aes(x = RoofStyle, y = SalePrice)) +geom\_boxplot() +  
 geom\_hline(aes(yintercept=80),   
 colour='red', linetype='dashed', lwd=2) +  
 scale\_y\_continuous(labels=dollar\_format()) +  
 theme\_few()



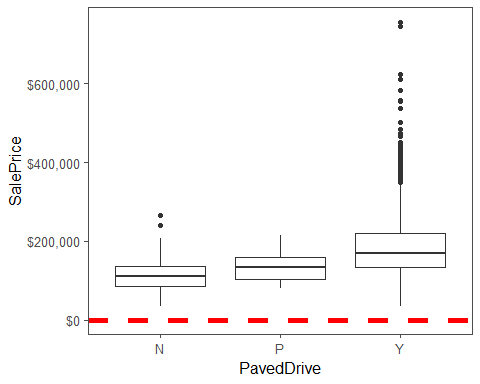
ggplot(training, aes(x = HouseStyle, y = SalePrice)) +geom\_boxplot() +  
 geom\_hline(aes(yintercept=80),   
 colour='red', linetype='dashed', lwd=2) +  
 scale\_y\_continuous(labels=dollar\_format()) +  
 theme\_few()



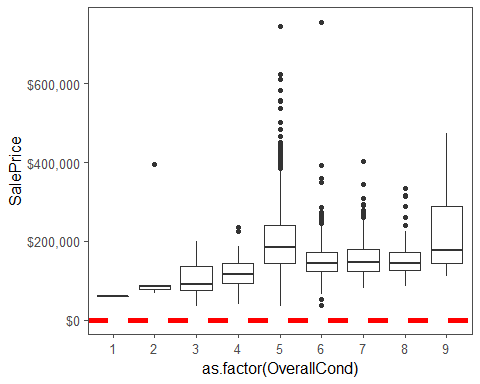
ggplot(training, aes(x = CentralAir, y = SalePrice)) +geom\_boxplot() +  
 geom\_hline(aes(yintercept=80),   
 colour='red', linetype='dashed', lwd=2) +  
 scale\_y\_continuous(labels=dollar\_format()) +  
 theme\_few()



ggplot(training, aes(x = PavedDrive, y = SalePrice)) +geom\_boxplot() +  
 geom\_hline(aes(yintercept=80),   
 colour='red', linetype='dashed', lwd=2) +  
 scale\_y\_continuous(labels=dollar\_format()) +  
 theme\_few()



ggplot(training, aes(x = as.factor(OverallCond), y = SalePrice)) +geom\_boxplot() +  
 geom\_hline(aes(yintercept=80),   
 colour='red', linetype='dashed', lwd=2) +  
 scale\_y\_continuous(labels=dollar\_format()) +  
 theme\_few()



#Check relationship between Sale price and Lot Area  
plot(SalePrice ~ LotArea, data = training, xlab ="Sale Price ", ylab = "Lot Area")  
  
  
# Car Library has a scatterplot matrix  
library(car)

## Warning: package 'car' was built under R version 3.5.2

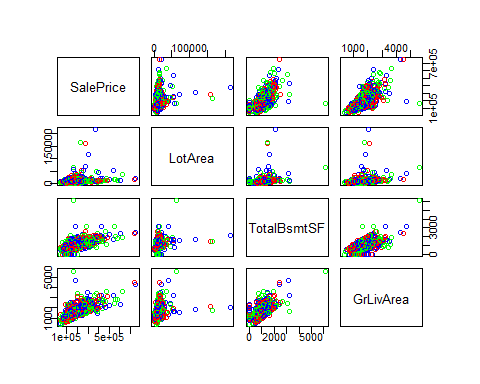
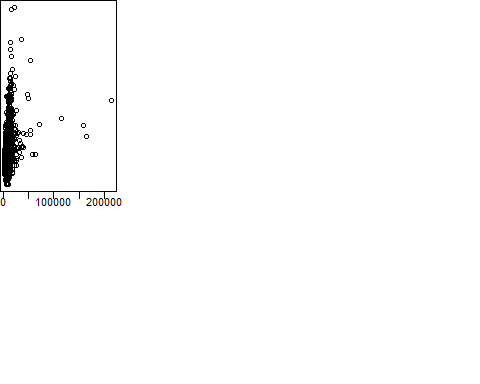
## Loading required package: carData

## Warning: package 'carData' was built under R version 3.5.2

##   
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':  
##   
## recode

#Check how Sale Price varies with plot area, basement area and ground living area  
pairs(~SalePrice+LotArea+TotalBsmtSF+GrLivArea, data=training,col=c('red','blue','green'))



#Check for correlation between Sale Price and other variables  
library(PerformanceAnalytics)

## Warning: package 'PerformanceAnalytics' was built under R version 3.5.2

## Loading required package: xts

## Loading required package: zoo

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

##   
## Attaching package: 'xts'

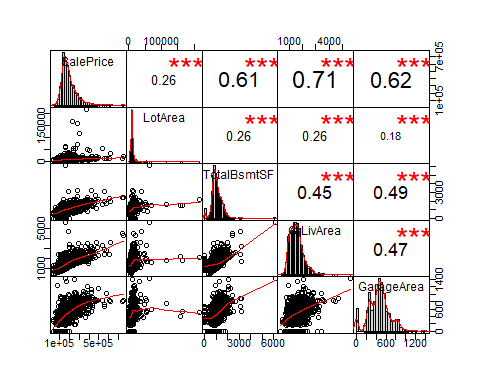
## The following objects are masked from 'package:data.table':  
##   
## first, last

## The following objects are masked from 'package:dplyr':  
##   
## first, last

##   
## Attaching package: 'PerformanceAnalytics'

## The following object is masked from 'package:graphics':  
##   
## legend

my\_data <- training[, c('SalePrice','LotArea','TotalBsmtSF','GrLivArea','GarageArea')]  
chart.Correlation(my\_data, histogram=TRUE, pch=19)



my\_data <- training[, c('SalePrice','BedroomAbvGr','TotRmsAbvGrd','GarageCars','OpenPorchSF','EnclosedPorch','WoodDeckSF')]  
chart.Correlation(my\_data, histogram=TRUE, pch=19)

