Week-2

Viraj Sapre, Sinchana Eshwar, Kartik Nagarajan

February 15, 2019

#Loading the libraries  
library(ggplot2)  
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(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

library(Rmisc)

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

## Loading required package: lattice

## Loading required package: plyr

## -------------------------------------------------------------------------

## You have loaded plyr after dplyr - this is likely to cause problems.  
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:  
## library(plyr); library(dplyr)

## -------------------------------------------------------------------------

##   
## Attaching package: 'plyr'

## The following objects are masked from 'package:dplyr':  
##   
## arrange, count, desc, failwith, id, mutate, rename, summarise,  
## summarize

# loading the dataset  
training <- read.csv("~/MS SEM 2/Multivariate Analysis - Raunak Parikh/MVA Grp Project/train.csv")  
View(training)

UNDERSTANDING THE DATA

dim(training) # checking the dimensions

## [1] 1460 81

str(training)# checking the structure of dataset

## '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 : 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 : 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 : int 0 1 1 1 1 0 1 2 2 2 ...  
## $ 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 : int 2 2 2 3 3 2 2 2 2 1 ...  
## $ 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 ...

summary(training)# checking the summary of dataset

## Id MSSubClass MSZoning LotFrontage   
## Min. : 1.0 Min. : 20.0 C (all): 10 Min. : 21.00   
## 1st Qu.: 365.8 1st Qu.: 20.0 FV : 65 1st Qu.: 59.00   
## Median : 730.5 Median : 50.0 RH : 16 Median : 69.00   
## Mean : 730.5 Mean : 56.9 RL :1151 Mean : 70.05   
## 3rd Qu.:1095.2 3rd Qu.: 70.0 RM : 218 3rd Qu.: 80.00   
## Max. :1460.0 Max. :190.0 Max. :313.00   
## NA's :259   
## LotArea Street Alley LotShape LandContour  
## Min. : 1300 Grvl: 6 Grvl: 50 IR1:484 Bnk: 63   
## 1st Qu.: 7554 Pave:1454 Pave: 41 IR2: 41 HLS: 50   
## Median : 9478 NA's:1369 IR3: 10 Low: 36   
## Mean : 10517 Reg:925 Lvl:1311   
## 3rd Qu.: 11602   
## Max. :215245   
##   
## Utilities LotConfig LandSlope Neighborhood Condition1   
## AllPub:1459 Corner : 263 Gtl:1382 NAmes :225 Norm :1260   
## NoSeWa: 1 CulDSac: 94 Mod: 65 CollgCr:150 Feedr : 81   
## FR2 : 47 Sev: 13 OldTown:113 Artery : 48   
## FR3 : 4 Edwards:100 RRAn : 26   
## Inside :1052 Somerst: 86 PosN : 19   
## Gilbert: 79 RRAe : 11   
## (Other):707 (Other): 15   
## Condition2 BldgType HouseStyle OverallQual   
## Norm :1445 1Fam :1220 1Story :726 Min. : 1.000   
## Feedr : 6 2fmCon: 31 2Story :445 1st Qu.: 5.000   
## Artery : 2 Duplex: 52 1.5Fin :154 Median : 6.000   
## PosN : 2 Twnhs : 43 SLvl : 65 Mean : 6.099   
## RRNn : 2 TwnhsE: 114 SFoyer : 37 3rd Qu.: 7.000   
## PosA : 1 1.5Unf : 14 Max. :10.000   
## (Other): 2 (Other): 19   
## OverallCond YearBuilt YearRemodAdd RoofStyle   
## Min. :1.000 Min. :1872 Min. :1950 Flat : 13   
## 1st Qu.:5.000 1st Qu.:1954 1st Qu.:1967 Gable :1141   
## Median :5.000 Median :1973 Median :1994 Gambrel: 11   
## Mean :5.575 Mean :1971 Mean :1985 Hip : 286   
## 3rd Qu.:6.000 3rd Qu.:2000 3rd Qu.:2004 Mansard: 7   
## Max. :9.000 Max. :2010 Max. :2010 Shed : 2   
##   
## RoofMatl Exterior1st Exterior2nd MasVnrType MasVnrArea   
## CompShg:1434 VinylSd:515 VinylSd:504 BrkCmn : 15 Min. : 0.0   
## Tar&Grv: 11 HdBoard:222 MetalSd:214 BrkFace:445 1st Qu.: 0.0   
## WdShngl: 6 MetalSd:220 HdBoard:207 None :864 Median : 0.0   
## WdShake: 5 Wd Sdng:206 Wd Sdng:197 Stone :128 Mean : 103.7   
## ClyTile: 1 Plywood:108 Plywood:142 NA's : 8 3rd Qu.: 166.0   
## Membran: 1 CemntBd: 61 CmentBd: 60 Max. :1600.0   
## (Other): 2 (Other):128 (Other):136 NA's :8   
## ExterQual ExterCond Foundation BsmtQual BsmtCond BsmtExposure  
## Ex: 52 Ex: 3 BrkTil:146 Ex :121 Fa : 45 Av :221   
## Fa: 14 Fa: 28 CBlock:634 Fa : 35 Gd : 65 Gd :134   
## Gd:488 Gd: 146 PConc :647 Gd :618 Po : 2 Mn :114   
## TA:906 Po: 1 Slab : 24 TA :649 TA :1311 No :953   
## TA:1282 Stone : 6 NA's: 37 NA's: 37 NA's: 38   
## Wood : 3   
##   
## BsmtFinType1 BsmtFinSF1 BsmtFinType2 BsmtFinSF2   
## ALQ :220 Min. : 0.0 ALQ : 19 Min. : 0.00   
## BLQ :148 1st Qu.: 0.0 BLQ : 33 1st Qu.: 0.00   
## GLQ :418 Median : 383.5 GLQ : 14 Median : 0.00   
## LwQ : 74 Mean : 443.6 LwQ : 46 Mean : 46.55   
## Rec :133 3rd Qu.: 712.2 Rec : 54 3rd Qu.: 0.00   
## Unf :430 Max. :5644.0 Unf :1256 Max. :1474.00   
## NA's: 37 NA's: 38   
## BsmtUnfSF TotalBsmtSF Heating HeatingQC CentralAir  
## Min. : 0.0 Min. : 0.0 Floor: 1 Ex:741 N: 95   
## 1st Qu.: 223.0 1st Qu.: 795.8 GasA :1428 Fa: 49 Y:1365   
## Median : 477.5 Median : 991.5 GasW : 18 Gd:241   
## Mean : 567.2 Mean :1057.4 Grav : 7 Po: 1   
## 3rd Qu.: 808.0 3rd Qu.:1298.2 OthW : 2 TA:428   
## Max. :2336.0 Max. :6110.0 Wall : 4   
##   
## Electrical X1stFlrSF X2ndFlrSF LowQualFinSF   
## FuseA: 94 Min. : 334 Min. : 0 Min. : 0.000   
## FuseF: 27 1st Qu.: 882 1st Qu.: 0 1st Qu.: 0.000   
## FuseP: 3 Median :1087 Median : 0 Median : 0.000   
## Mix : 1 Mean :1163 Mean : 347 Mean : 5.845   
## SBrkr:1335 3rd Qu.:1391 3rd Qu.: 728 3rd Qu.: 0.000   
## Max. :4692 Max. :2065 Max. :572.000   
##   
## GrLivArea BsmtFullBath BsmtHalfBath FullBath   
## Min. : 334 Min. :0.0000 Min. :0.00000 Min. :0.000   
## 1st Qu.:1130 1st Qu.:0.0000 1st Qu.:0.00000 1st Qu.:1.000   
## Median :1464 Median :0.0000 Median :0.00000 Median :2.000   
## Mean :1515 Mean :0.4253 Mean :0.05753 Mean :1.565   
## 3rd Qu.:1777 3rd Qu.:1.0000 3rd Qu.:0.00000 3rd Qu.:2.000   
## Max. :5642 Max. :3.0000 Max. :2.00000 Max. :3.000   
##   
## HalfBath BedroomAbvGr KitchenAbvGr KitchenQual  
## Min. :0.0000 Min. :0.000 Min. :0.000 Ex:100   
## 1st Qu.:0.0000 1st Qu.:2.000 1st Qu.:1.000 Fa: 39   
## Median :0.0000 Median :3.000 Median :1.000 Gd:586   
## Mean :0.3829 Mean :2.866 Mean :1.047 TA:735   
## 3rd Qu.:1.0000 3rd Qu.:3.000 3rd Qu.:1.000   
## Max. :2.0000 Max. :8.000 Max. :3.000   
##   
## TotRmsAbvGrd Functional Fireplaces FireplaceQu GarageType   
## Min. : 2.000 Maj1: 14 Min. :0.000 Ex : 24 2Types : 6   
## 1st Qu.: 5.000 Maj2: 5 1st Qu.:0.000 Fa : 33 Attchd :870   
## Median : 6.000 Min1: 31 Median :1.000 Gd :380 Basment: 19   
## Mean : 6.518 Min2: 34 Mean :0.613 Po : 20 BuiltIn: 88   
## 3rd Qu.: 7.000 Mod : 15 3rd Qu.:1.000 TA :313 CarPort: 9   
## Max. :14.000 Sev : 1 Max. :3.000 NA's:690 Detchd :387   
## Typ :1360 NA's : 81   
## GarageYrBlt GarageFinish GarageCars GarageArea GarageQual   
## Min. :1900 Fin :352 Min. :0.000 Min. : 0.0 Ex : 3   
## 1st Qu.:1961 RFn :422 1st Qu.:1.000 1st Qu.: 334.5 Fa : 48   
## Median :1980 Unf :605 Median :2.000 Median : 480.0 Gd : 14   
## Mean :1979 NA's: 81 Mean :1.767 Mean : 473.0 Po : 3   
## 3rd Qu.:2002 3rd Qu.:2.000 3rd Qu.: 576.0 TA :1311   
## Max. :2010 Max. :4.000 Max. :1418.0 NA's: 81   
## NA's :81   
## GarageCond PavedDrive WoodDeckSF OpenPorchSF EnclosedPorch   
## Ex : 2 N: 90 Min. : 0.00 Min. : 0.00 Min. : 0.00   
## Fa : 35 P: 30 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.00   
## Gd : 9 Y:1340 Median : 0.00 Median : 25.00 Median : 0.00   
## Po : 7 Mean : 94.24 Mean : 46.66 Mean : 21.95   
## TA :1326 3rd Qu.:168.00 3rd Qu.: 68.00 3rd Qu.: 0.00   
## NA's: 81 Max. :857.00 Max. :547.00 Max. :552.00   
##   
## X3SsnPorch ScreenPorch PoolArea PoolQC   
## Min. : 0.00 Min. : 0.00 Min. : 0.000 Ex : 2   
## 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.000 Fa : 2   
## Median : 0.00 Median : 0.00 Median : 0.000 Gd : 3   
## Mean : 3.41 Mean : 15.06 Mean : 2.759 NA's:1453   
## 3rd Qu.: 0.00 3rd Qu.: 0.00 3rd Qu.: 0.000   
## Max. :508.00 Max. :480.00 Max. :738.000   
##   
## Fence MiscFeature MiscVal MoSold   
## GdPrv: 59 Gar2: 2 Min. : 0.00 Min. : 1.000   
## GdWo : 54 Othr: 2 1st Qu.: 0.00 1st Qu.: 5.000   
## MnPrv: 157 Shed: 49 Median : 0.00 Median : 6.000   
## MnWw : 11 TenC: 1 Mean : 43.49 Mean : 6.322   
## NA's :1179 NA's:1406 3rd Qu.: 0.00 3rd Qu.: 8.000   
## Max. :15500.00 Max. :12.000   
##   
## YrSold SaleType SaleCondition SalePrice   
## Min. :2006 WD :1267 Abnorml: 101 Min. : 34900   
## 1st Qu.:2007 New : 122 AdjLand: 4 1st Qu.:129975   
## Median :2008 COD : 43 Alloca : 12 Median :163000   
## Mean :2008 ConLD : 9 Family : 20 Mean :180921   
## 3rd Qu.:2009 ConLI : 5 Normal :1198 3rd Qu.:214000   
## Max. :2010 ConLw : 5 Partial: 125 Max. :755000   
## (Other): 9

Checking for MISSING VALUES

#Missing data  
sum(is.na(training)/(nrow(training)\*nrow(training)))# printing percentage of missing data

## [1] 0.003267029

unique(nrow(training)) # printing all the unique values

## [1] 1460

colSums(sapply(training,is.na))# prinitng number of missing values in each column

## Id MSSubClass MSZoning LotFrontage LotArea   
## 0 0 0 259 0   
## Street Alley LotShape LandContour Utilities   
## 0 1369 0 0 0   
## 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 0 0   
## MasVnrType MasVnrArea ExterQual ExterCond Foundation   
## 8 8 0 0 0   
## BsmtQual BsmtCond BsmtExposure BsmtFinType1 BsmtFinSF1   
## 37 37 38 37 0   
## BsmtFinType2 BsmtFinSF2 BsmtUnfSF TotalBsmtSF Heating   
## 38 0 0 0 0   
## HeatingQC CentralAir Electrical X1stFlrSF X2ndFlrSF   
## 0 0 0 0 0   
## LowQualFinSF GrLivArea BsmtFullBath BsmtHalfBath FullBath   
## 0 0 0 0 0   
## HalfBath BedroomAbvGr KitchenAbvGr KitchenQual TotRmsAbvGrd   
## 0 0 0 0 0   
## Functional Fireplaces FireplaceQu GarageType GarageYrBlt   
## 0 0 690 81 81   
## GarageFinish GarageCars GarageArea GarageQual GarageCond   
## 81 0 0 81 81   
## PavedDrive WoodDeckSF OpenPorchSF EnclosedPorch X3SsnPorch   
## 0 0 0 0 0   
## ScreenPorch PoolArea PoolQC Fence MiscFeature   
## 0 0 1453 1179 1406   
## MiscVal MoSold YrSold SaleType SaleCondition   
## 0 0 0 0 0   
## SalePrice   
## 0

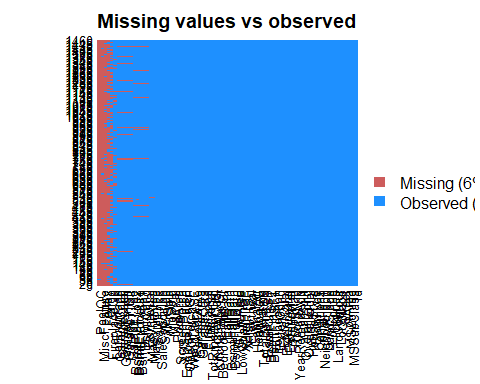
library(Amelia)

## 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")



# creating dataframe of categorical and numerical variables  
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')

unique(nrow(training$SalePrice))

## NULL

Removing columns with NA values

training$Alley = NULL   
training$LotFrontage = NULL  
training$FireplaceQu = NULL  
training$Fence = NULL  
training$PoolQC = NULL   
training$MiscFeature = NULL   
training$BsmtQual = NULL   
training$BsmtCond = NULL   
training$BsmtExposure = NULL   
training$BsmtFinType1 = NULL   
training$BsmtFinType2 = NULL  
training$GarageType = NULL  
training$GarageYrBlt = NULL  
training$MasVnrType = NULL   
training$MasVnrArea = NULL   
training$GarageQual = NULL   
training$GarageFinish = NULL   
training$GarageCond = NULL

training[!complete.cases(training),]

## [1] Id MSSubClass MSZoning LotArea Street   
## [6] LotShape LandContour Utilities LotConfig LandSlope   
## [11] Neighborhood Condition1 Condition2 BldgType HouseStyle   
## [16] OverallQual OverallCond YearBuilt YearRemodAdd RoofStyle   
## [21] RoofMatl Exterior1st Exterior2nd ExterQual ExterCond   
## [26] Foundation BsmtFinSF1 BsmtFinSF2 BsmtUnfSF TotalBsmtSF   
## [31] Heating HeatingQC CentralAir Electrical X1stFlrSF   
## [36] X2ndFlrSF LowQualFinSF GrLivArea BsmtFullBath BsmtHalfBath   
## [41] FullBath HalfBath BedroomAbvGr KitchenAbvGr KitchenQual   
## [46] TotRmsAbvGrd Functional Fireplaces GarageCars GarageArea   
## [51] PavedDrive WoodDeckSF OpenPorchSF EnclosedPorch X3SsnPorch   
## [56] ScreenPorch PoolArea MiscVal MoSold YrSold   
## [61] SaleType SaleCondition SalePrice   
## <0 rows> (or 0-length row.names)

head(training)

## Id MSSubClass MSZoning LotArea Street LotShape LandContour Utilities  
## 1 1 60 RL 8450 Pave Reg Lvl AllPub  
## 2 2 20 RL 9600 Pave Reg Lvl AllPub  
## 3 3 60 RL 11250 Pave IR1 Lvl AllPub  
## 4 4 70 RL 9550 Pave IR1 Lvl AllPub  
## 5 5 60 RL 14260 Pave IR1 Lvl AllPub  
## 6 6 50 RL 14115 Pave IR1 Lvl AllPub  
## LotConfig LandSlope Neighborhood Condition1 Condition2 BldgType  
## 1 Inside Gtl CollgCr Norm Norm 1Fam  
## 2 FR2 Gtl Veenker Feedr Norm 1Fam  
## 3 Inside Gtl CollgCr Norm Norm 1Fam  
## 4 Corner Gtl Crawfor Norm Norm 1Fam  
## 5 FR2 Gtl NoRidge Norm Norm 1Fam  
## 6 Inside Gtl Mitchel Norm Norm 1Fam  
## HouseStyle OverallQual OverallCond YearBuilt YearRemodAdd RoofStyle  
## 1 2Story 7 5 2003 2003 Gable  
## 2 1Story 6 8 1976 1976 Gable  
## 3 2Story 7 5 2001 2002 Gable  
## 4 2Story 7 5 1915 1970 Gable  
## 5 2Story 8 5 2000 2000 Gable  
## 6 1.5Fin 5 5 1993 1995 Gable  
## RoofMatl Exterior1st Exterior2nd ExterQual ExterCond Foundation  
## 1 CompShg VinylSd VinylSd Gd TA PConc  
## 2 CompShg MetalSd MetalSd TA TA CBlock  
## 3 CompShg VinylSd VinylSd Gd TA PConc  
## 4 CompShg Wd Sdng Wd Shng TA TA BrkTil  
## 5 CompShg VinylSd VinylSd Gd TA PConc  
## 6 CompShg VinylSd VinylSd TA TA Wood  
## BsmtFinSF1 BsmtFinSF2 BsmtUnfSF TotalBsmtSF Heating HeatingQC CentralAir  
## 1 706 0 150 856 GasA Ex Y  
## 2 978 0 284 1262 GasA Ex Y  
## 3 486 0 434 920 GasA Ex Y  
## 4 216 0 540 756 GasA Gd Y  
## 5 655 0 490 1145 GasA Ex Y  
## 6 732 0 64 796 GasA Ex Y  
## Electrical X1stFlrSF X2ndFlrSF LowQualFinSF GrLivArea BsmtFullBath  
## 1 SBrkr 856 854 0 1710 1  
## 2 SBrkr 1262 0 0 1262 0  
## 3 SBrkr 920 866 0 1786 1  
## 4 SBrkr 961 756 0 1717 1  
## 5 SBrkr 1145 1053 0 2198 1  
## 6 SBrkr 796 566 0 1362 1  
## BsmtHalfBath FullBath HalfBath BedroomAbvGr KitchenAbvGr KitchenQual  
## 1 0 2 1 3 1 Gd  
## 2 1 2 0 3 1 TA  
## 3 0 2 1 3 1 Gd  
## 4 0 1 0 3 1 Gd  
## 5 0 2 1 4 1 Gd  
## 6 0 1 1 1 1 TA  
## TotRmsAbvGrd Functional Fireplaces GarageCars GarageArea PavedDrive  
## 1 8 Typ 0 2 548 Y  
## 2 6 Typ 1 2 460 Y  
## 3 6 Typ 1 2 608 Y  
## 4 7 Typ 1 3 642 Y  
## 5 9 Typ 1 3 836 Y  
## 6 5 Typ 0 2 480 Y  
## WoodDeckSF OpenPorchSF EnclosedPorch X3SsnPorch ScreenPorch PoolArea  
## 1 0 61 0 0 0 0  
## 2 298 0 0 0 0 0  
## 3 0 42 0 0 0 0  
## 4 0 35 272 0 0 0  
## 5 192 84 0 0 0 0  
## 6 40 30 0 320 0 0  
## MiscVal MoSold YrSold SaleType SaleCondition SalePrice  
## 1 0 2 2008 WD Normal 208500  
## 2 0 5 2007 WD Normal 181500  
## 3 0 9 2008 WD Normal 223500  
## 4 0 2 2006 WD Abnorml 140000  
## 5 0 12 2008 WD Normal 250000  
## 6 700 10 2009 WD Normal 143000

#Missing data  
sum(is.na(training)/(nrow(training)\*nrow(training)))# printing percentage of missing data

## [1] 0

unique(nrow(training)) # printing all the unique values

## [1] 1460

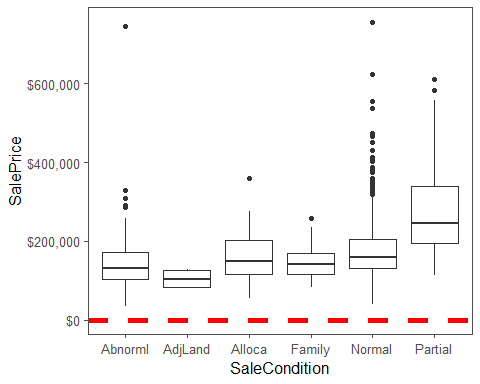
colSums(sapply(training,is.na))# prinitng number of missing values in each column

## Id MSSubClass MSZoning LotArea Street   
## 0 0 0 0 0   
## LotShape LandContour Utilities LotConfig LandSlope   
## 0 0 0 0 0   
## Neighborhood Condition1 Condition2 BldgType HouseStyle   
## 0 0 0 0 0   
## OverallQual OverallCond YearBuilt YearRemodAdd RoofStyle   
## 0 0 0 0 0   
## RoofMatl Exterior1st Exterior2nd ExterQual ExterCond   
## 0 0 0 0 0   
## Foundation BsmtFinSF1 BsmtFinSF2 BsmtUnfSF TotalBsmtSF   
## 0 0 0 0 0   
## Heating HeatingQC CentralAir Electrical X1stFlrSF   
## 0 0 0 0 0   
## X2ndFlrSF LowQualFinSF GrLivArea BsmtFullBath BsmtHalfBath   
## 0 0 0 0 0   
## FullBath HalfBath BedroomAbvGr KitchenAbvGr KitchenQual   
## 0 0 0 0 0   
## TotRmsAbvGrd Functional Fireplaces GarageCars GarageArea   
## 0 0 0 0 0   
## PavedDrive WoodDeckSF OpenPorchSF EnclosedPorch X3SsnPorch   
## 0 0 0 0 0   
## ScreenPorch PoolArea MiscVal MoSold YrSold   
## 0 0 0 0 0   
## SaleType SaleCondition SalePrice   
## 0 0 0

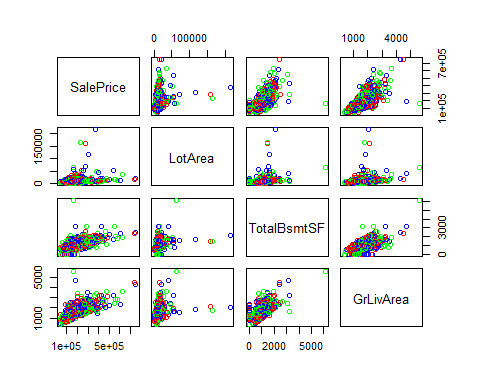
attach(training)   
catdf<-training[,catvar]  
numdf<-training[,numvar]

VISUALIZING THE DATA

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()



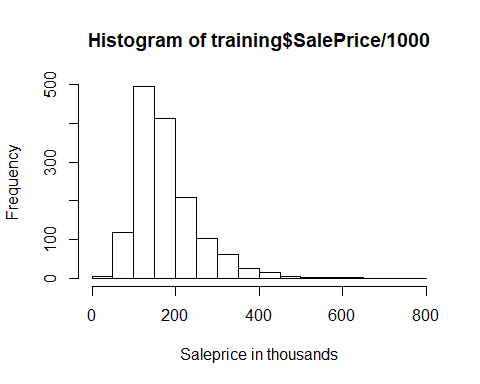
pairs(~SalePrice+LotArea+TotalBsmtSF+GrLivArea, data=training,col=c('red','blue','green'))



as.factor(training$SalePrice)

## [1] 208500 181500 223500 140000 250000 143000 307000 200000 129900  
## [10] 118000 129500 345000 144000 279500 157000 132000 149000 90000   
## [19] 159000 139000 325300 139400 230000 129900 154000 256300 134800  
## [28] 306000 207500 68500 40000 149350 179900 165500 277500 309000  
## [37] 145000 153000 109000 82000 160000 170000 144000 130250 141000  
## [46] 319900 239686 249700 113000 127000 177000 114500 110000 385000  
## [55] 130000 180500 172500 196500 438780 124900 158000 101000 202500  
## [64] 140000 219500 317000 180000 226000 80000 225000 244000 129500  
## [73] 185000 144900 107400 91000 135750 127000 136500 110000 193500  
## [82] 153500 245000 126500 168500 260000 174000 164500 85000 123600  
## [91] 109900 98600 163500 133900 204750 185000 214000 94750 83000   
## [100] 128950 205000 178000 118964 198900 169500 250000 100000 115000  
## [109] 115000 190000 136900 180000 383970 217000 259500 176000 139000  
## [118] 155000 320000 163990 180000 100000 136000 153900 181000 84500   
## [127] 128000 87000 155000 150000 226000 244000 150750 220000 180000  
## [136] 174000 143000 171000 230000 231500 115000 260000 166000 204000  
## [145] 125000 130000 105000 222500 141000 115000 122000 372402 190000  
## [154] 235000 125000 79000 109500 269500 254900 320000 162500 412500  
## [163] 220000 103200 152000 127500 190000 325624 183500 228000 128500  
## [172] 215000 239000 163000 184000 243000 211000 172500 501837 100000  
## [181] 177000 200100 120000 200000 127000 475000 173000 135000 153337  
## [190] 286000 315000 184000 192000 130000 127000 148500 311872 235000  
## [199] 104000 274900 140000 171500 112000 149000 110000 180500 143900  
## [208] 141000 277000 145000 98000 186000 252678 156000 161750 134450  
## [217] 210000 107000 311500 167240 204900 200000 179900 97000 386250  
## [226] 112000 290000 106000 125000 192500 148000 403000 94500 128200  
## [235] 216500 89500 185500 194500 318000 113000 262500 110500 79000   
## [244] 120000 205000 241500 137000 140000 180000 277000 76500 235000  
## [253] 173000 158000 145000 230000 207500 220000 231500 97000 176000  
## [262] 276000 151000 130000 73000 175500 185000 179500 120500 148000  
## [271] 266000 241500 290000 139000 124500 205000 201000 141000 415298  
## [280] 192000 228500 185000 207500 244600 179200 164700 159000 88000   
## [289] 122000 153575 233230 135900 131000 235000 167000 142500 152000  
## [298] 239000 175000 158500 157000 267000 205000 149900 295000 305900  
## [307] 225000 89500 82500 360000 165600 132000 119900 375000 178000  
## [316] 188500 260000 270000 260000 187500 342643 354000 301000 126175  
## [325] 242000 87000 324000 145250 214500 78000 119000 139000 284000  
## [334] 207000 192000 228950 377426 214000 202500 155000 202900 82000   
## [343] 87500 266000 85000 140200 151500 157500 154000 437154 318061  
## [352] 190000 95000 105900 140000 177500 173000 134000 130000 280000  
## [361] 156000 145000 198500 118000 190000 147000 159000 165000 132000  
## [370] 162000 172400 134432 125000 123000 219500 61000 148000 340000  
## [379] 394432 179000 127000 187750 213500 76000 240000 192000 81000   
## [388] 125000 191000 426000 119000 215000 106500 100000 109000 129000  
## [397] 123000 169500 67000 241000 245500 164990 108000 258000 168000  
## [406] 150000 115000 177000 280000 339750 60000 145000 222000 115000  
## [415] 228000 181134 149500 239000 126000 142000 206300 215000 113000  
## [424] 315000 139000 135000 275000 109008 195400 175000 85400 79900   
## [433] 122500 181000 81000 212000 116000 119000 90350 110000 555000  
## [442] 118000 162900 172500 210000 127500 190000 199900 119500 120000  
## [451] 110000 280000 204000 210000 188000 175500 98000 256000 161000  
## [460] 110000 263435 155000 62383 188700 124000 178740 167000 146500  
## [469] 250000 187000 212000 190000 148000 440000 251000 132500 208900  
## [478] 380000 297000 89471 326000 374000 155000 164000 132500 147000  
## [487] 156000 175000 160000 86000 115000 133000 172785 155000 91300   
## [496] 34900 430000 184000 130000 120000 113000 226700 140000 289000  
## [505] 147000 124500 215000 208300 161000 124500 164900 202665 129900  
## [514] 134000 96500 402861 158000 265000 211000 234000 106250 150000  
## [523] 159000 184750 315750 176000 132000 446261 86000 200624 175000  
## [532] 128000 107500 39300 178000 107500 188000 111250 158000 272000  
## [541] 315000 248000 213250 133000 179665 229000 210000 129500 125000  
## [550] 263000 140000 112500 255500 108000 284000 113000 141000 108000  
## [559] 175000 234000 121500 170000 108000 185000 268000 128000 325000  
## [568] 214000 316600 135960 142600 120000 224500 170000 139000 118500  
## [577] 145000 164500 146000 131500 181900 253293 118500 325000 133000  
## [586] 369900 130000 137000 143000 79500 185900 451950 138000 140000  
## [595] 110000 319000 114504 194201 217500 151000 275000 141000 220000  
## [604] 151000 221000 205000 152000 225000 359100 118500 313000 148000  
## [613] 261500 147000 75500 137500 183200 105500 314813 305000 67000   
## [622] 240000 135000 168500 165150 160000 139900 153000 135000 168500  
## [631] 124000 209500 82500 139400 144000 200000 60000 93000 85000   
## [640] 264561 274000 226000 345000 152000 370878 143250 98300 155000  
## [649] 155000 84500 205950 108000 191000 135000 350000 88000 145500  
## [658] 149000 97500 167000 197900 402000 110000 137500 423000 230500  
## [667] 129000 193500 168000 137500 173500 103600 165000 257500 140000  
## [676] 148500 87000 109500 372500 128500 143000 159434 173000 285000  
## [685] 221000 207500 227875 148800 392000 194700 141000 755000 335000  
## [694] 108480 141500 176000 89000 123500 138500 196000 312500 140000  
## [703] 361919 140000 213000 55000 302000 254000 179540 109900 52000   
## [712] 102776 189000 129000 130500 165000 159500 157000 341000 128500  
## [721] 275000 143000 124500 135000 320000 120500 222000 194500 110000  
## [730] 103000 236500 187500 222500 131400 108000 163000 93500 239900  
## [739] 179000 190000 132000 142000 179000 175000 180000 299800 236000  
## [748] 265979 260400 98000 96500 162000 217000 275500 156000 172500  
## [757] 212000 158900 179400 290000 127500 100000 215200 337000 270000  
## [766] 264132 196500 160000 216837 538000 134900 102000 107000 114500  
## [775] 395000 162000 221500 142500 144000 135000 176000 175900 187100  
## [784] 165500 128000 161500 139000 233000 107900 187500 160200 146800  
## [793] 269790 225000 194500 171000 143500 110000 485000 175000 200000  
## [802] 109900 189000 582933 118000 227680 135500 223500 159950 106000  
## [811] 181000 144500 55993 157900 116000 224900 137000 271000 155000  
## [820] 224000 183000 93000 225000 139500 232600 385000 109500 189000  
## [829] 185000 147400 166000 151000 237000 167000 139950 128000 153500  
## [838] 100000 144000 130500 140000 157500 174900 141000 153900 171000  
## [847] 213000 133500 240000 187000 131500 215000 164000 158000 170000  
## [856] 127000 147000 174000 152000 250000 189950 131500 152000 132500  
## [865] 250580 148500 248900 129000 169000 236000 109500 200500 116000  
## [874] 133000 66500 303477 132250 350000 148000 136500 157000 187500  
## [883] 178000 118500 100000 328900 145000 135500 268000 149500 122900  
## [892] 172500 154500 165000 118858 140000 106500 142953 611657 135000  
## [901] 110000 153000 180000 240000 125500 128000 255000 250000 131000  
## [910] 174000 154300 143500 88000 145000 173733 75000 35311 135000  
## [919] 238000 176500 201000 145900 169990 193000 207500 175000 285000  
## [928] 176000 236500 222000 201000 117500 320000 190000 242000 79900   
## [937] 184900 253000 239799 244400 150900 214000 150000 143000 137500  
## [946] 124900 143000 270000 192500 197500 129000 119900 133900 172000  
## [955] 127500 145000 124000 132000 185000 155000 116500 272000 155000  
## [964] 239000 214900 178900 160000 135000 37900 140000 135000 173000  
## [973] 99500 182000 167500 165000 85500 199900 110000 139000 178400  
## [982] 336000 159895 255900 126000 125000 117000 395192 195000 197000  
## [991] 348000 168000 187000 173900 337500 121600 136500 185000 91000   
## [1000] 206000 82000 86000 232000 136905 181000 149900 163500 88000   
## [1009] 240000 102000 135000 100000 165000 85000 119200 227000 203000  
## [1018] 187500 160000 213490 176000 194000 87000 191000 287000 112500  
## [1027] 167500 293077 105000 118000 160000 197000 310000 230000 119750  
## [1036] 84000 315500 287000 97000 80000 155000 173000 196000 262280  
## [1045] 278000 139600 556581 145000 115000 84900 176485 200141 165000  
## [1054] 144500 255000 180000 185850 248000 335000 220000 213500 81000   
## [1063] 90000 110500 154000 328000 178000 167900 151400 135000 135000  
## [1072] 154000 91500 159500 194000 219500 170000 138800 155900 126000  
## [1081] 145000 133000 192000 160000 187500 147000 83500 252000 137500  
## [1090] 197000 92900 160000 136500 146000 129000 176432 127000 170000  
## [1099] 128000 157000 60000 119500 135000 159500 106000 325000 179900  
## [1108] 274725 181000 280000 188000 205000 129900 134500 117000 318000  
## [1117] 184100 130000 140000 133700 118400 212900 112000 118000 163900  
## [1126] 115000 174000 259000 215000 140000 135000 93500 117500 239500  
## [1135] 169000 102000 119000 94000 196000 144000 139000 197500 424870  
## [1144] 80000 80000 149000 180000 174500 116900 143000 124000 149900  
## [1153] 230000 120500 201800 218000 179900 230000 235128 185000 146000  
## [1162] 224000 129000 108959 194000 233170 245350 173000 235000 625000  
## [1171] 171000 163000 171900 200500 239000 285000 119500 115000 154900  
## [1180] 93000 250000 392500 745000 120000 186700 104900 95000 262000  
## [1189] 195000 189000 168000 174000 125000 165000 158000 176000 219210  
## [1198] 144000 178000 148000 116050 197900 117000 213000 153500 271900  
## [1207] 107000 200000 140000 290000 189000 164000 113000 145000 134500  
## [1216] 125000 112000 229456 80500 91500 115000 134000 143000 137900  
## [1225] 184000 145000 214000 147000 367294 127000 190000 132500 101800  
## [1234] 142000 130000 138887 175500 195000 142500 265900 224900 248328  
## [1243] 170000 465000 230000 178000 186500 169900 129500 119000 244000  
## [1252] 171750 130000 294000 165400 127500 301500 99900 190000 151000  
## [1261] 181000 128900 161500 180500 181000 183900 122000 378500 381000  
## [1270] 144000 260000 185750 137000 177000 139000 137000 162000 197900  
## [1279] 237000 68400 227000 180000 150500 139000 169000 132500 143000  
## [1288] 190000 278000 281000 180500 119500 107500 162900 115000 138500  
## [1297] 155000 140000 160000 154000 225000 177500 290000 232000 130000  
## [1306] 325000 202500 138000 147000 179200 335000 203000 302000 333168  
## [1315] 119000 206900 295493 208900 275000 111000 156500 72500 190000  
## [1324] 82500 147000 55000 79000 130500 256000 176500 227000 132500  
## [1333] 100000 125500 125000 167900 135000 52500 200000 128500 123000  
## [1342] 155000 228500 177000 155835 108500 262500 283463 215000 122000  
## [1351] 200000 171000 134900 410000 235000 170000 110000 149900 177500  
## [1360] 315000 189000 260000 104900 156932 144152 216000 193000 127000  
## [1369] 144000 232000 105000 165500 274300 466500 250000 239000 91000   
## [1378] 117000 83000 167500 58500 237500 157000 112000 105000 125500  
## [1387] 250000 136000 377500 131000 235000 124000 123000 163000 246578  
## [1396] 281213 160000 137500 138000 137450 120000 193000 193879 282922  
## [1405] 105000 275000 133000 112000 125500 215000 230000 140000 90000   
## [1414] 257000 207000 175900 122500 340000 124000 223000 179900 127500  
## [1423] 136500 274970 144000 142000 271000 140000 119000 182900 192140  
## [1432] 143750 64500 186500 160000 174000 120500 394617 149700 197000  
## [1441] 191000 149300 310000 121000 179600 129000 157900 240000 112000  
## [1450] 92000 136000 287090 145000 84500 185000 175000 210000 266500  
## [1459] 142125 147500  
## 663 Levels: 34900 35311 37900 39300 40000 52000 52500 55000 55993 ... 755000

hist(training$SalePrice / 1000, xlab = "Saleprice in thousands")



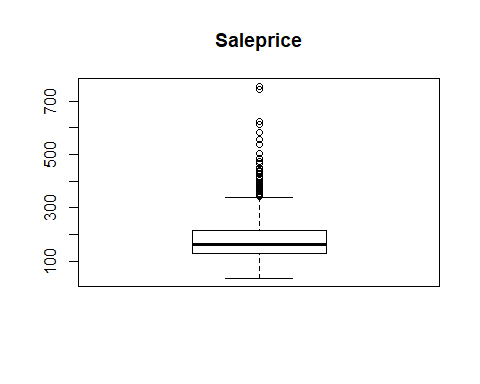
library(moments)

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

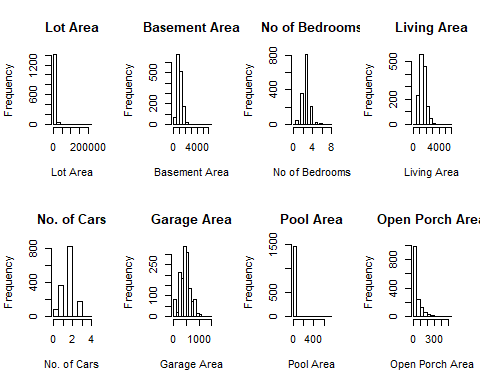
skewness(SalePrice)

## [1] 1.880941

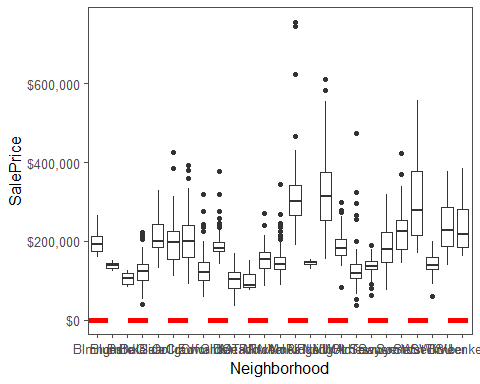
boxplot(training$SalePrice/ 1000, main = "Saleprice")



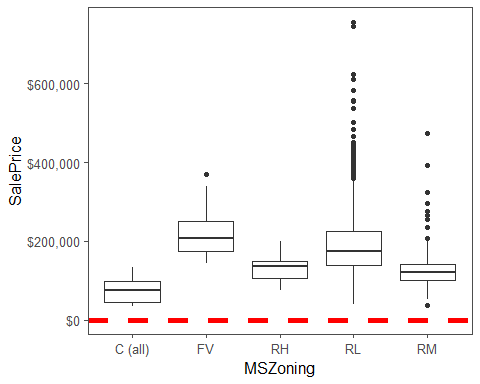
par(mfrow=c(2,4))  
hist(training$LotArea,xlab="Lot Area", main="Lot Area")  
hist(training$TotalBsmtSF, xlab="Basement Area", main="Basement Area")  
hist(training$BedroomAbvGr, xlab="No of Bedrooms", main="No of Bedrooms")  
hist(training$GrLivArea, xlab="Living Area",main="Living Area")  
hist(training$GarageCars, xlab="No. of Cars",main="No. of Cars")  
hist(training$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")



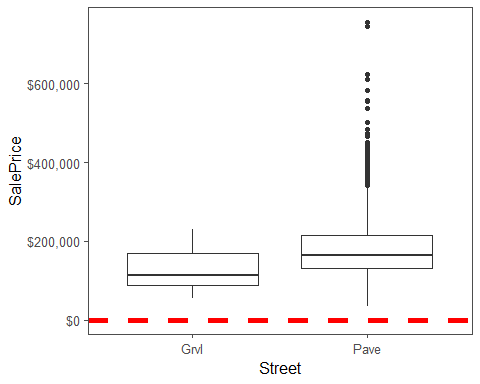
ggplot(training, aes(x = Neighborhood, y = SalePrice)) +  
 geom\_boxplot() +  
 geom\_hline(aes(yintercept=80),   
 colour='red', linetype='dashed', lwd=2) +  
 scale\_y\_continuous(labels=dollar\_format()) +  
 theme\_few()



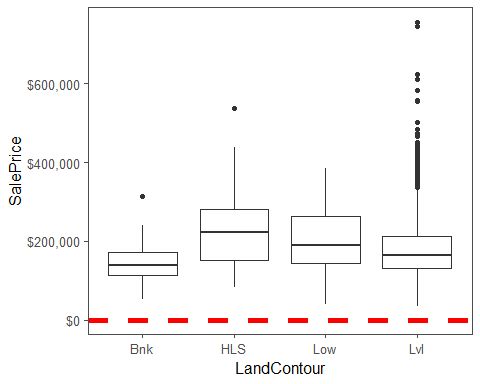
ggplot(training, aes(x = MSZoning, y = SalePrice)) +geom\_boxplot() +  
 geom\_hline(aes(yintercept=80),   
 colour='red', linetype='dashed', lwd=2) +  
 scale\_y\_continuous(labels=dollar\_format()) +  
 theme\_few()



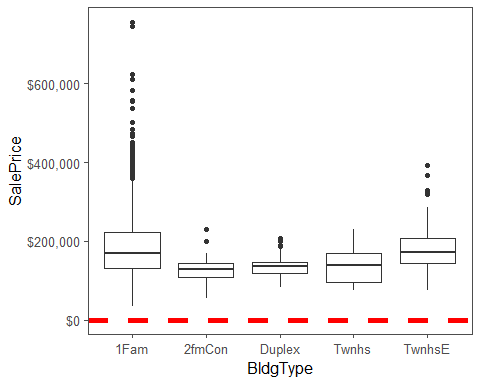
ggplot(training, aes(x = Street, y = SalePrice)) +geom\_boxplot() +  
 geom\_hline(aes(yintercept=80),   
 colour='red', linetype='dashed', lwd=2) +  
 scale\_y\_continuous(labels=dollar\_format()) +  
 theme\_few()



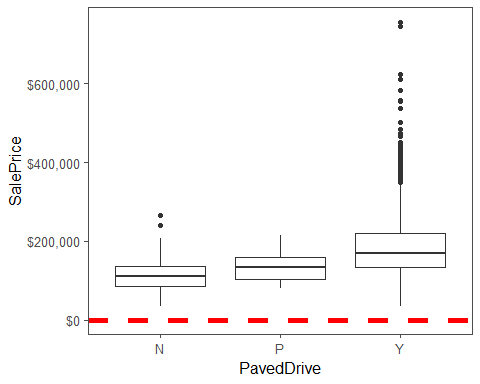
ggplot(training, aes(x = LandContour, y = SalePrice)) +geom\_boxplot() +  
 geom\_hline(aes(yintercept=80),   
 colour='red', linetype='dashed', lwd=2) +  
 scale\_y\_continuous(labels=dollar\_format()) +  
 theme\_few()



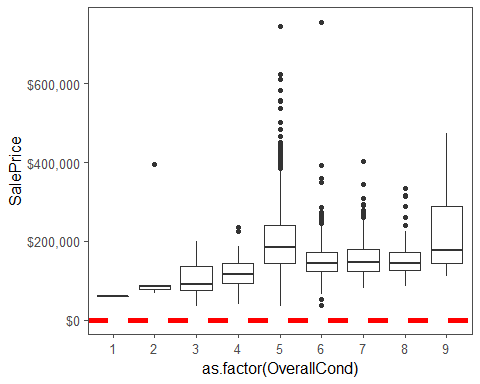
ggplot(training, aes(x = BldgType, y = SalePrice)) +geom\_boxplot() +  
 geom\_hline(aes(yintercept=80),   
 colour='red', linetype='dashed', lwd=2) +  
 scale\_y\_continuous(labels=dollar\_format()) +  
 theme\_few()



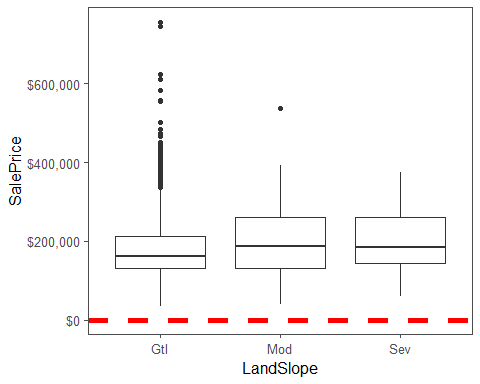
ggplot(training, aes(x = PavedDrive, y = SalePrice)) +geom\_boxplot() +  
 geom\_hline(aes(yintercept=80),   
 colour='red', linetype='dashed', lwd=2) +  
 scale\_y\_continuous(labels=dollar\_format()) +  
 theme\_few()



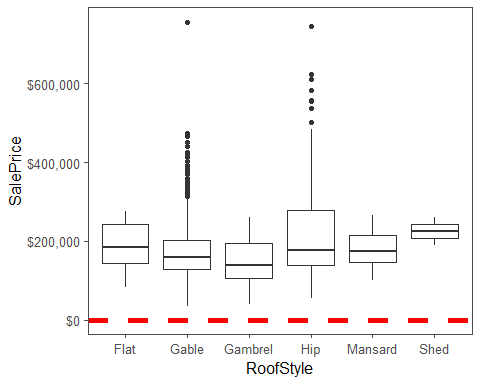
ggplot(training, aes(x = as.factor(OverallCond), y = SalePrice)) +geom\_boxplot() +  
 geom\_hline(aes(yintercept=80),   
 colour='red', linetype='dashed', lwd=2) +  
 scale\_y\_continuous(labels=dollar\_format()) +  
 theme\_few()



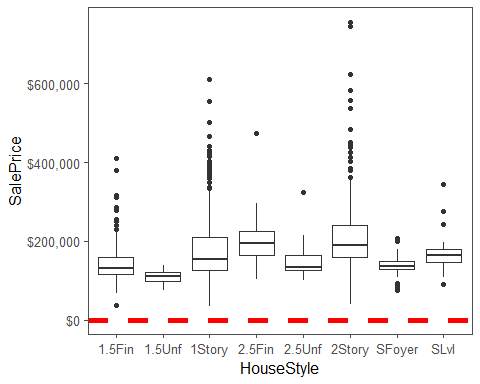
ggplot(training, aes(x = LandSlope, y = SalePrice)) +geom\_boxplot() +  
 geom\_hline(aes(yintercept=80),   
 colour='red', linetype='dashed', lwd=2) +  
 scale\_y\_continuous(labels=dollar\_format()) +  
 theme\_few()



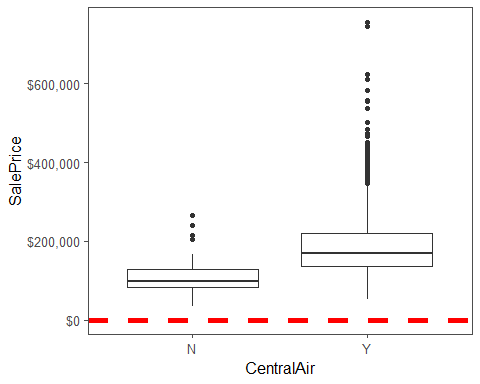
ggplot(training, aes(x = RoofStyle, y = SalePrice)) +geom\_boxplot() +  
 geom\_hline(aes(yintercept=80),   
 colour='red', linetype='dashed', lwd=2) +  
 scale\_y\_continuous(labels=dollar\_format()) +  
 theme\_few()



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()



library(PerformanceAnalytics)

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

## Loading required package: xts

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

## 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

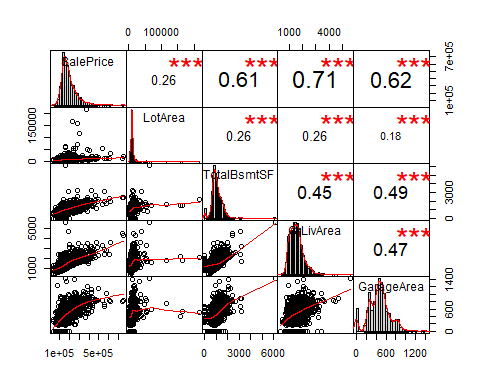
## The following objects are masked from 'package:dplyr':  
##   
## first, last

##   
## Attaching package: 'PerformanceAnalytics'

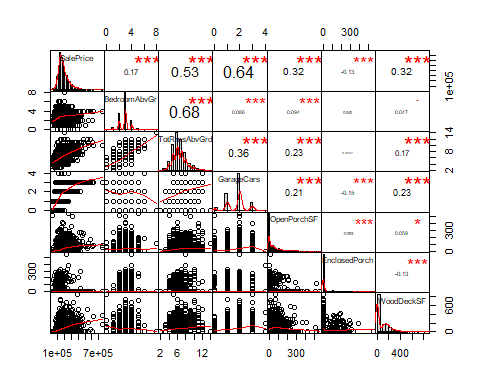
## The following objects are masked from 'package:moments':  
##   
## kurtosis, skewness

## 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)



library(forecast)

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

linear <- lm(SalePrice~., data=training, metric="RMSE", maximize=FALSE)

## Warning: In lm.fit(x, y, offset = offset, singular.ok = singular.ok, ...) :  
## extra arguments 'metric', 'maximize' will be disregarded

summary(linear)

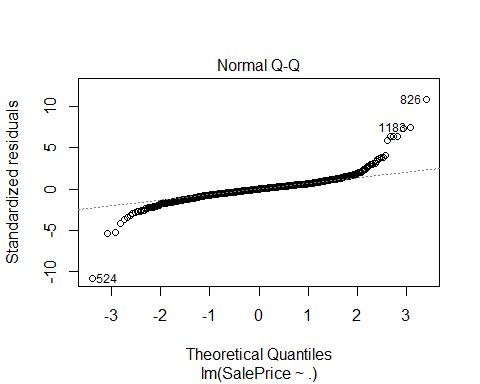
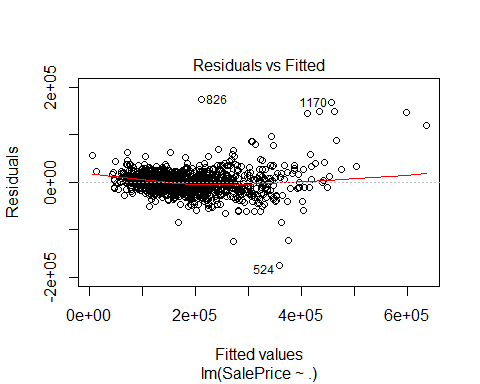
##   
## Call:  
## lm(formula = SalePrice ~ ., data = training, metric = "RMSE",   
## maximize = FALSE)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -174300 -10507 12 9742 174300   
##   
## Coefficients: (3 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.200e+06 1.065e+06 -1.127 0.260067   
## Id 6.458e-01 1.601e+00 0.403 0.686753   
## MSSubClass -7.728e+00 8.550e+01 -0.090 0.927989   
## MSZoningFV 3.098e+04 1.224e+04 2.531 0.011489 \*   
## MSZoningRH 2.371e+04 1.231e+04 1.925 0.054391 .   
## MSZoningRL 2.587e+04 1.050e+04 2.463 0.013911 \*   
## MSZoningRM 2.502e+04 9.848e+03 2.541 0.011180 \*   
## LotArea 7.006e-01 1.083e-01 6.472 1.38e-10 \*\*\*  
## StreetPave 3.870e+04 1.228e+04 3.151 0.001665 \*\*   
## LotShapeIR2 4.625e+03 4.321e+03 1.070 0.284680   
## LotShapeIR3 4.581e+03 9.059e+03 0.506 0.613167   
## LotShapeReg 5.684e+02 1.665e+03 0.341 0.732961   
## LandContourHLS 1.348e+04 5.304e+03 2.542 0.011149 \*   
## LandContourLow -4.236e+03 6.527e+03 -0.649 0.516526   
## LandContourLvl 7.077e+03 3.820e+03 1.853 0.064182 .   
## UtilitiesNoSeWa -3.056e+04 2.659e+04 -1.149 0.250618   
## LotConfigCulDSac 7.638e+03 3.326e+03 2.296 0.021810 \*   
## LotConfigFR2 -5.854e+03 4.158e+03 -1.408 0.159440   
## LotConfigFR3 -1.348e+04 1.308e+04 -1.031 0.302866   
## LotConfigInside -1.237e+03 1.810e+03 -0.684 0.494410   
## LandSlopeMod 1.045e+04 4.043e+03 2.586 0.009823 \*\*   
## LandSlopeSev -2.562e+04 1.110e+04 -2.308 0.021142 \*   
## NeighborhoodBlueste -2.730e+03 1.935e+04 -0.141 0.887809   
## NeighborhoodBrDale 8.394e+03 1.113e+04 0.754 0.450817   
## NeighborhoodBrkSide -2.082e+03 9.505e+03 -0.219 0.826632   
## NeighborhoodClearCr -1.276e+04 9.428e+03 -1.354 0.176057   
## NeighborhoodCollgCr -9.715e+03 7.334e+03 -1.325 0.185536   
## NeighborhoodCrawfor 9.620e+03 8.669e+03 1.110 0.267362   
## NeighborhoodEdwards -1.675e+04 8.082e+03 -2.073 0.038377 \*   
## NeighborhoodGilbert -1.390e+04 7.850e+03 -1.770 0.076956 .   
## NeighborhoodIDOTRR -7.794e+03 1.087e+04 -0.717 0.473358   
## NeighborhoodMeadowV -1.385e+03 1.140e+04 -0.121 0.903359   
## NeighborhoodMitchel -2.037e+04 8.277e+03 -2.461 0.013972 \*   
## NeighborhoodNAmes -1.447e+04 7.902e+03 -1.831 0.067347 .   
## NeighborhoodNoRidge 2.879e+04 8.402e+03 3.426 0.000631 \*\*\*  
## NeighborhoodNPkVill 8.169e+03 1.433e+04 0.570 0.568824   
## NeighborhoodNridgHt 2.461e+04 7.378e+03 3.336 0.000874 \*\*\*  
## NeighborhoodNWAmes -2.053e+04 8.151e+03 -2.518 0.011916 \*   
## NeighborhoodOldTown -1.300e+04 9.677e+03 -1.343 0.179404   
## NeighborhoodSawyer -1.005e+04 8.233e+03 -1.221 0.222375   
## NeighborhoodSawyerW -6.143e+03 7.854e+03 -0.782 0.434301   
## NeighborhoodSomerst 7.906e-01 8.980e+03 0.000 0.999930   
## NeighborhoodStoneBr 3.898e+04 8.385e+03 4.650 3.67e-06 \*\*\*  
## NeighborhoodSWISU -9.586e+03 9.834e+03 -0.975 0.329848   
## NeighborhoodTimber -5.971e+03 8.371e+03 -0.713 0.475794   
## NeighborhoodVeenker 3.148e+03 1.073e+04 0.293 0.769344   
## Condition1Feedr 2.817e+03 5.116e+03 0.551 0.581914   
## Condition1Norm 1.209e+04 4.224e+03 2.861 0.004286 \*\*   
## Condition1PosA 7.405e+03 1.031e+04 0.718 0.472582   
## Condition1PosN 7.856e+03 7.631e+03 1.029 0.303482   
## Condition1RRAe -1.709e+04 9.377e+03 -1.822 0.068677 .   
## Condition1RRAn 6.186e+03 7.038e+03 0.879 0.379552   
## Condition1RRNe -7.318e+03 1.838e+04 -0.398 0.690585   
## Condition1RRNn 3.848e+03 1.312e+04 0.293 0.769315   
## Condition2Feedr -9.514e+03 2.306e+04 -0.413 0.680029   
## Condition2Norm -7.560e+03 1.966e+04 -0.385 0.700612   
## Condition2PosA 2.018e+04 3.802e+04 0.531 0.595576   
## Condition2PosN -2.303e+05 2.763e+04 -8.334 < 2e-16 \*\*\*  
## Condition2RRAe -1.290e+05 4.684e+04 -2.753 0.005982 \*\*   
## Condition2RRAn -1.211e+04 3.196e+04 -0.379 0.704859   
## Condition2RRNn -8.732e+03 2.712e+04 -0.322 0.747487   
## BldgType2fmCon -6.296e+03 1.288e+04 -0.489 0.624929   
## BldgTypeDuplex -1.025e+03 7.463e+03 -0.137 0.890815   
## BldgTypeTwnhs -2.562e+04 1.017e+04 -2.520 0.011847 \*   
## BldgTypeTwnhsE -2.344e+04 9.211e+03 -2.544 0.011063 \*   
## HouseStyle1.5Unf 1.120e+04 7.935e+03 1.411 0.158491   
## HouseStyle1Story 8.956e+03 4.356e+03 2.056 0.039956 \*   
## HouseStyle2.5Fin -1.715e+04 1.232e+04 -1.392 0.164170   
## HouseStyle2.5Unf -1.189e+04 9.392e+03 -1.266 0.205716   
## HouseStyle2Story -6.382e+03 3.552e+03 -1.797 0.072605 .   
## HouseStyleSFoyer 7.644e+03 6.200e+03 1.233 0.217884   
## HouseStyleSLvl 7.343e+03 5.468e+03 1.343 0.179504   
## OverallQual 8.032e+03 1.020e+03 7.874 7.30e-15 \*\*\*  
## OverallCond 5.439e+03 8.755e+02 6.212 7.08e-10 \*\*\*  
## YearBuilt 3.311e+02 7.392e+01 4.479 8.17e-06 \*\*\*  
## YearRemodAdd 1.066e+02 5.569e+01 1.914 0.055835 .   
## RoofStyleGable 1.496e+03 1.876e+04 0.080 0.936444   
## RoofStyleGambrel 4.299e+03 2.051e+04 0.210 0.834017   
## RoofStyleHip 3.087e+03 1.881e+04 0.164 0.869639   
## RoofStyleMansard 1.722e+04 2.185e+04 0.788 0.430629   
## RoofStyleShed 8.762e+04 3.553e+04 2.466 0.013783 \*   
## RoofMatlCompShg 6.502e+05 3.302e+04 19.689 < 2e-16 \*\*\*  
## RoofMatlMembran 7.375e+05 4.779e+04 15.433 < 2e-16 \*\*\*  
## RoofMatlMetal 6.975e+05 4.720e+04 14.778 < 2e-16 \*\*\*  
## RoofMatlRoll 6.493e+05 4.165e+04 15.592 < 2e-16 \*\*\*  
## RoofMatlTar&Grv 6.556e+05 3.797e+04 17.269 < 2e-16 \*\*\*  
## RoofMatlWdShake 6.306e+05 3.677e+04 17.152 < 2e-16 \*\*\*  
## RoofMatlWdShngl 7.280e+05 3.426e+04 21.246 < 2e-16 \*\*\*  
## Exterior1stAsphShn -1.263e+04 3.421e+04 -0.369 0.711971   
## Exterior1stBrkComm -1.328e+04 2.868e+04 -0.463 0.643398   
## Exterior1stBrkFace 5.456e+03 1.287e+04 0.424 0.671596   
## Exterior1stCBlock -2.810e+04 2.760e+04 -1.018 0.308807   
## Exterior1stCemntBd -1.486e+04 1.946e+04 -0.764 0.445236   
## Exterior1stHdBoard -1.378e+04 1.299e+04 -1.061 0.288838   
## Exterior1stImStucc -6.919e+04 2.862e+04 -2.418 0.015762 \*   
## Exterior1stMetalSd -3.221e+03 1.483e+04 -0.217 0.828084   
## Exterior1stPlywood -1.795e+04 1.287e+04 -1.395 0.163304   
## Exterior1stStone -1.509e+04 2.437e+04 -0.619 0.535916   
## Exterior1stStucco -4.988e+03 1.417e+04 -0.352 0.724892   
## Exterior1stVinylSd -1.765e+04 1.347e+04 -1.311 0.190052   
## Exterior1stWd Sdng -1.355e+04 1.242e+04 -1.091 0.275503   
## Exterior1stWdShing -6.407e+03 1.344e+04 -0.477 0.633574   
## Exterior2ndAsphShn 8.073e+03 2.282e+04 0.354 0.723521   
## Exterior2ndBrk Cmn 1.498e+04 2.074e+04 0.722 0.470225   
## Exterior2ndBrkFace -8.318e+02 1.330e+04 -0.063 0.950144   
## Exterior2ndCBlock NA NA NA NA   
## Exterior2ndCmentBd 1.302e+04 1.919e+04 0.678 0.497706   
## Exterior2ndHdBoard 8.077e+03 1.251e+04 0.646 0.518500   
## Exterior2ndImStucc 3.368e+04 1.447e+04 2.328 0.020049 \*   
## Exterior2ndMetalSd 2.920e+03 1.448e+04 0.202 0.840153   
## Exterior2ndOther -6.162e+03 2.821e+04 -0.218 0.827115   
## Exterior2ndPlywood 9.196e+03 1.215e+04 0.757 0.449227   
## Exterior2ndStone -1.017e+04 1.738e+04 -0.585 0.558385   
## Exterior2ndStucco 2.446e+03 1.365e+04 0.179 0.857830   
## Exterior2ndVinylSd 1.649e+04 1.300e+04 1.269 0.204721   
## Exterior2ndWd Sdng 1.052e+04 1.199e+04 0.877 0.380550   
## Exterior2ndWd Shng 3.422e+03 1.250e+04 0.274 0.784385   
## ExterQualFa -8.626e+03 1.089e+04 -0.792 0.428269   
## ExterQualGd -3.081e+04 4.791e+03 -6.432 1.78e-10 \*\*\*  
## ExterQualTA -3.068e+04 5.363e+03 -5.721 1.32e-08 \*\*\*  
## ExterCondFa -2.629e+03 1.888e+04 -0.139 0.889247   
## ExterCondGd -8.011e+03 1.802e+04 -0.445 0.656662   
## ExterCondPo 1.214e+04 3.285e+04 0.370 0.711749   
## ExterCondTA -5.363e+03 1.798e+04 -0.298 0.765562   
## FoundationCBlock 1.753e+03 3.198e+03 0.548 0.583771   
## FoundationPConc 4.836e+03 3.507e+03 1.379 0.168180   
## FoundationSlab 8.476e+03 7.860e+03 1.078 0.281078   
## FoundationStone 2.507e+03 1.117e+04 0.224 0.822555   
## FoundationWood -3.341e+04 1.512e+04 -2.209 0.027332 \*   
## BsmtFinSF1 3.708e+01 4.421e+00 8.387 < 2e-16 \*\*\*  
## BsmtFinSF2 2.457e+01 5.798e+00 4.239 2.41e-05 \*\*\*  
## BsmtUnfSF 1.496e+01 4.069e+00 3.675 0.000248 \*\*\*  
## TotalBsmtSF NA NA NA NA   
## HeatingGasA -7.004e+03 2.546e+04 -0.275 0.783309   
## HeatingGasW -1.557e+04 2.626e+04 -0.593 0.553449   
## HeatingGrav -1.538e+04 2.764e+04 -0.557 0.577934   
## HeatingOthW -4.573e+04 3.172e+04 -1.441 0.149709   
## HeatingWall 8.300e+03 2.950e+04 0.281 0.778505   
## HeatingQCFa -1.577e+03 4.829e+03 -0.326 0.744101   
## HeatingQCGd -3.603e+03 2.147e+03 -1.679 0.093487 .   
## HeatingQCPo 8.348e+03 2.774e+04 0.301 0.763466   
## HeatingQCTA -4.392e+03 2.121e+03 -2.071 0.038588 \*   
## CentralAirY -3.678e+03 3.997e+03 -0.920 0.357636   
## ElectricalFuseF -1.229e+03 5.990e+03 -0.205 0.837468   
## ElectricalFuseP -1.003e+04 1.745e+04 -0.575 0.565337   
## ElectricalMix 3.595e+03 2.892e+04 0.124 0.901084   
## ElectricalSBrkr -1.338e+03 3.026e+03 -0.442 0.658435   
## X1stFlrSF 5.502e+01 5.334e+00 10.314 < 2e-16 \*\*\*  
## X2ndFlrSF 7.004e+01 5.265e+00 13.302 < 2e-16 \*\*\*  
## LowQualFinSF 2.520e+01 1.872e+01 1.346 0.178417   
## GrLivArea NA NA NA NA   
## BsmtFullBath 1.546e+03 1.968e+03 0.786 0.432273   
## BsmtHalfBath 3.461e+02 3.116e+03 0.111 0.911592   
## FullBath 2.608e+03 2.246e+03 1.161 0.245703   
## HalfBath -1.342e+02 2.138e+03 -0.063 0.949952   
## BedroomAbvGr -5.509e+03 1.385e+03 -3.979 7.33e-05 \*\*\*  
## KitchenAbvGr -1.577e+04 5.774e+03 -2.731 0.006396 \*\*   
## KitchenQualFa -2.066e+04 6.413e+03 -3.222 0.001307 \*\*   
## KitchenQualGd -2.775e+04 3.488e+03 -7.955 3.93e-15 \*\*\*  
## KitchenQualTA -2.523e+04 3.997e+03 -6.314 3.76e-10 \*\*\*  
## TotRmsAbvGrd 1.345e+03 9.760e+02 1.378 0.168400   
## FunctionalMaj2 -5.227e+02 1.480e+04 -0.035 0.971825   
## FunctionalMin1 4.455e+03 8.666e+03 0.514 0.607279   
## FunctionalMin2 8.561e+03 8.581e+03 0.998 0.318610   
## FunctionalMod -7.241e+03 1.056e+04 -0.685 0.493155   
## FunctionalSev -5.984e+04 2.758e+04 -2.170 0.030227 \*   
## FunctionalTyp 1.971e+04 7.419e+03 2.656 0.008008 \*\*   
## Fireplaces 2.818e+03 1.374e+03 2.051 0.040467 \*   
## GarageCars 4.264e+03 2.221e+03 1.920 0.055097 .   
## GarageArea 1.333e+01 7.646e+00 1.744 0.081413 .   
## PavedDriveP -3.332e+03 5.574e+03 -0.598 0.550131   
## PavedDriveY -2.108e+03 3.458e+03 -0.610 0.542226   
## WoodDeckSF 1.370e+01 5.956e+00 2.301 0.021579 \*   
## OpenPorchSF 1.209e+01 1.184e+01 1.021 0.307444   
## EnclosedPorch 5.647e+00 1.285e+01 0.440 0.660360   
## X3SsnPorch 2.429e+01 2.315e+01 1.049 0.294148   
## ScreenPorch 3.718e+01 1.259e+01 2.952 0.003216 \*\*   
## PoolArea 7.122e+01 1.835e+01 3.882 0.000109 \*\*\*  
## MiscVal -3.127e-01 1.469e+00 -0.213 0.831474   
## MoSold -6.399e+02 2.540e+02 -2.519 0.011895 \*   
## YrSold -1.754e+02 5.248e+02 -0.334 0.738202   
## SaleTypeCon 3.535e+04 1.838e+04 1.923 0.054648 .   
## SaleTypeConLD 1.672e+04 1.002e+04 1.668 0.095464 .   
## SaleTypeConLI 9.989e+03 1.192e+04 0.838 0.402177   
## SaleTypeConLw -2.450e+03 1.243e+04 -0.197 0.843733   
## SaleTypeCWD 2.314e+04 1.336e+04 1.731 0.083642 .   
## SaleTypeNew 3.436e+04 1.605e+04 2.141 0.032443 \*   
## SaleTypeOth 1.853e+04 1.504e+04 1.233 0.217982   
## SaleTypeWD 4.521e+02 4.342e+03 0.104 0.917077   
## SaleConditionAdjLand 1.047e+04 1.505e+04 0.696 0.486820   
## SaleConditionAlloca 5.066e+03 8.784e+03 0.577 0.564214   
## SaleConditionFamily -1.323e+03 6.327e+03 -0.209 0.834445   
## SaleConditionNormal 6.623e+03 2.993e+03 2.213 0.027082 \*   
## SaleConditionPartial -9.120e+03 1.547e+04 -0.590 0.555508   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 24000 on 1269 degrees of freedom  
## Multiple R-squared: 0.9206, Adjusted R-squared: 0.9087   
## F-statistic: 77.43 on 190 and 1269 DF, p-value: < 2.2e-16

accuracy(linear)

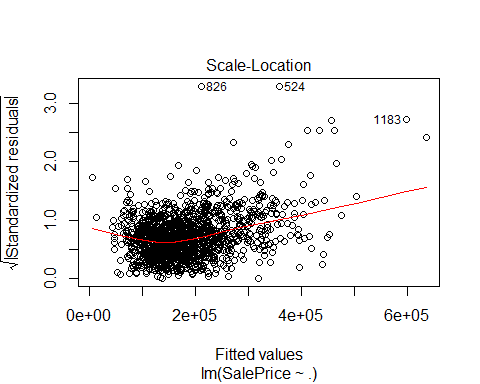
## ME RMSE MAE MPE MAPE MASE  
## Training set -4.809863e-13 22378.52 14525.28 -0.589597 8.490661 0.2529004

plot(linear)

## Warning: not plotting observations with leverage one:  
## 121, 251, 272, 326, 399, 584, 667, 1004, 1012, 1188, 1231, 1271, 1276, 1299, 1322, 1371

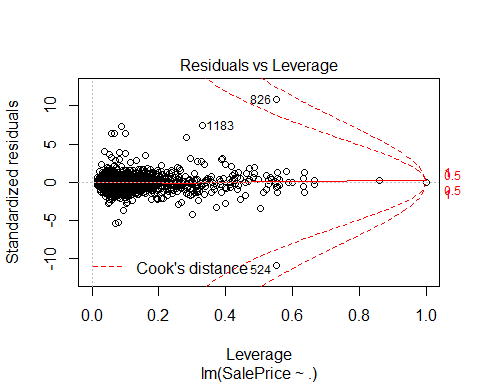


## Warning: not plotting observations with leverage one:  
## 121, 251, 272, 326, 399, 584, 667, 1004, 1012, 1188, 1231, 1271, 1276, 1299, 1322, 1371



## Warning in sqrt(crit \* p \* (1 - hh)/hh): NaNs produced

## Warning in sqrt(crit \* p \* (1 - hh)/hh): NaNs produced



library(car)

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

## Loading required package: carData

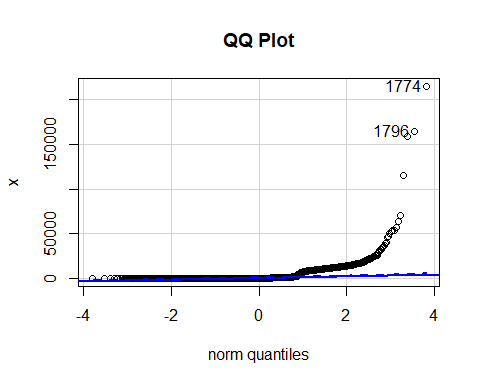
##   
## Attaching package: 'car'

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

outlierTest(linear)

## rstudent unadjusted p-value Bonferonni p  
## 826 11.365873 1.4159e-28 2.0445e-25  
## 524 -11.365873 1.4159e-28 2.0445e-25  
## 1183 7.603303 5.5944e-14 8.0784e-11  
## 1170 7.464835 1.5445e-13 2.2303e-10  
## 899 6.531315 9.4131e-11 1.3592e-07  
## 1047 6.514330 1.0503e-10 1.5167e-07  
## 804 6.509241 1.0854e-10 1.5672e-07  
## 692 5.944277 3.5843e-09 5.1757e-06  
## 1325 -5.460861 5.6961e-08 8.2252e-05  
## 582 -5.337710 1.1142e-07 1.6089e-04

x <- c( BedroomAbvGr, LotArea, PoolArea, TotalBsmtSF, TotRmsAbvGrd)  
qqPlot(x , main="QQ Plot")



## [1] 1774 1796

library(ggpubr)

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

## Loading required package: magrittr

##   
## Attaching package: 'ggpubr'

## The following object is masked from 'package:forecast':  
##   
## gghistogram

## The following object is masked from 'package:plyr':  
##   
## mutate

t.test(SalePrice, x, data = training)

##   
## Welch Two Sample t-test  
##   
## data: SalePrice and x  
## t = 85.854, df = 1462.4, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 174523.2 182684.6  
## sample estimates:  
## mean of x mean of y   
## 180921.20 2317.28