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)  
testing <- read.csv("~/MS SEM 2/Multivariate Analysis - Raunak Parikh/MVA Grp Project/test.csv")  
View(testing)

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:1334 3rd Qu.:1391 3rd Qu.: 728 3rd Qu.: 0.000   
## NA's : 1 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

dim(testing)# checking the dimensions

## [1] 1459 80

str(testing)# checking the structure of dataset

## '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)# checking the summary of dataset

## 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

# combining training and testing datasets  
housedata<-bind\_rows(training,testing)

## Warning in bind\_rows\_(x, .id): Unequal factor levels: coercing to character

## 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): Unequal factor levels: coercing to character

## 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): Unequal factor levels: coercing to character

## 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): Unequal factor levels: coercing to character

## 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): Unequal factor levels: coercing to character

## 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): Unequal factor levels: coercing to character

## 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): Unequal factor levels: coercing to character

## 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): Unequal factor levels: coercing to character

## 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): Unequal factor levels: coercing to character

## 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): Unequal factor levels: coercing to character

## 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): Unequal factor levels: coercing to character

## 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) # printing the first 6 rows of the dataset

## 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) # checking the dimensions of dataset

## [1] 2919 81

str(housedata)# checking the structure of dataset

## '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 : 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 : chr "AllPub" "AllPub" "AllPub" "AllPub" ...  
## $ 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 : chr "Norm" "Norm" "Norm" "Norm" ...  
## $ BldgType : Factor w/ 5 levels "1Fam","2fmCon",..: 1 1 1 1 1 1 1 1 1 2 ...  
## $ 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 : Factor w/ 6 levels "Flat","Gable",..: 2 2 2 2 2 2 2 2 2 2 ...  
## $ RoofMatl : chr "CompShg" "CompShg" "CompShg" "CompShg" ...  
## $ Exterior1st : chr "VinylSd" "MetalSd" "VinylSd" "Wd Sdng" ...  
## $ Exterior2nd : chr "VinylSd" "MetalSd" "VinylSd" "Wd Shng" ...  
## $ 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 : chr "GasA" "GasA" "GasA" "GasA" ...  
## $ 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 : 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 : 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 : chr "TA" "TA" "TA" "TA" ...  
## $ 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 : chr NA NA NA NA ...  
## $ Fence : Factor w/ 4 levels "GdPrv","GdWo",..: NA NA NA NA NA 3 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 : 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(housedata)# checking the summary of dataset

## Id MSSubClass MSZoning LotFrontage   
## Min. : 1.0 Min. : 20.00 C (all): 25 Min. : 21.00   
## 1st Qu.: 730.5 1st Qu.: 20.00 FV : 139 1st Qu.: 59.00   
## Median :1460.0 Median : 50.00 RH : 26 Median : 68.00   
## Mean :1460.0 Mean : 57.14 RL :2265 Mean : 69.31   
## 3rd Qu.:2189.5 3rd Qu.: 70.00 RM : 460 3rd Qu.: 80.00   
## Max. :2919.0 Max. :190.00 NA's : 4 Max. :313.00   
## NA's :486   
## LotArea Street Alley LotShape LandContour  
## Min. : 1300 Grvl: 12 Grvl: 120 IR1: 968 Bnk: 117   
## 1st Qu.: 7478 Pave:2907 Pave: 78 IR2: 76 HLS: 120   
## Median : 9453 NA's:2721 IR3: 16 Low: 60   
## Mean : 10168 Reg:1859 Lvl:2622   
## 3rd Qu.: 11570   
## Max. :215245   
##   
## Utilities LotConfig LandSlope Neighborhood   
## Length:2919 Corner : 511 Gtl:2778 NAmes : 443   
## Class :character CulDSac: 176 Mod: 125 CollgCr: 267   
## Mode :character FR2 : 85 Sev: 16 OldTown: 239   
## FR3 : 14 Edwards: 194   
## Inside :2133 Somerst: 182   
## NridgHt: 166   
## (Other):1428   
## Condition1 Condition2 BldgType HouseStyle   
## Norm :2511 Length:2919 1Fam :2425 Length:2919   
## Feedr : 164 Class :character 2fmCon: 62 Class :character   
## Artery : 92 Mode :character Duplex: 109 Mode :character   
## RRAn : 50 Twnhs : 96   
## PosN : 39 TwnhsE: 227   
## RRAe : 28   
## (Other): 35   
## OverallQual OverallCond YearBuilt YearRemodAdd   
## Min. : 1.000 Min. :1.000 Min. :1872 Min. :1950   
## 1st Qu.: 5.000 1st Qu.:5.000 1st Qu.:1954 1st Qu.:1965   
## Median : 6.000 Median :5.000 Median :1973 Median :1993   
## Mean : 6.089 Mean :5.565 Mean :1971 Mean :1984   
## 3rd Qu.: 7.000 3rd Qu.:6.000 3rd Qu.:2001 3rd Qu.:2004   
## Max. :10.000 Max. :9.000 Max. :2010 Max. :2010   
##   
## RoofStyle RoofMatl Exterior1st Exterior2nd   
## Flat : 20 Length:2919 Length:2919 Length:2919   
## Gable :2310 Class :character Class :character Class :character   
## Gambrel: 22 Mode :character Mode :character Mode :character   
## Hip : 551   
## Mansard: 11   
## Shed : 5   
##   
## MasVnrType MasVnrArea ExterQual ExterCond Foundation   
## BrkCmn : 25 Min. : 0.0 Ex: 107 Ex: 12 BrkTil: 311   
## BrkFace: 879 1st Qu.: 0.0 Fa: 35 Fa: 67 CBlock:1235   
## None :1742 Median : 0.0 Gd: 979 Gd: 299 PConc :1308   
## Stone : 249 Mean : 102.2 TA:1798 Po: 3 Slab : 49   
## NA's : 24 3rd Qu.: 164.0 TA:2538 Stone : 11   
## Max. :1600.0 Wood : 5   
## NA's :23   
## BsmtQual BsmtCond BsmtExposure BsmtFinType1 BsmtFinSF1   
## Ex : 258 Fa : 104 Av : 418 ALQ :429 Min. : 0.0   
## Fa : 88 Gd : 122 Gd : 276 BLQ :269 1st Qu.: 0.0   
## Gd :1209 Po : 5 Mn : 239 GLQ :849 Median : 368.5   
## TA :1283 TA :2606 No :1904 LwQ :154 Mean : 441.4   
## NA's: 81 NA's: 82 NA's: 82 Rec :288 3rd Qu.: 733.0   
## Unf :851 Max. :5644.0   
## NA's: 79 NA's :1   
## BsmtFinType2 BsmtFinSF2 BsmtUnfSF TotalBsmtSF   
## ALQ : 52 Min. : 0.00 Min. : 0.0 Min. : 0.0   
## BLQ : 68 1st Qu.: 0.00 1st Qu.: 220.0 1st Qu.: 793.0   
## GLQ : 34 Median : 0.00 Median : 467.0 Median : 989.5   
## LwQ : 87 Mean : 49.58 Mean : 560.8 Mean :1051.8   
## Rec : 105 3rd Qu.: 0.00 3rd Qu.: 805.5 3rd Qu.:1302.0   
## Unf :2493 Max. :1526.00 Max. :2336.0 Max. :6110.0   
## NA's: 80 NA's :1 NA's :1 NA's :1   
## Heating HeatingQC CentralAir Electrical X1stFlrSF   
## Length:2919 Ex:1493 N: 196 Length:2919 Min. : 334   
## Class :character Fa: 92 Y:2723 Class :character 1st Qu.: 876   
## Mode :character Gd: 474 Mode :character Median :1082   
## Po: 3 Mean :1160   
## TA: 857 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 Ex : 205 Min. : 2.000 Typ :2717   
## 1st Qu.:1.000 Fa : 70 1st Qu.: 5.000 Min2 : 70   
## Median :1.000 Gd :1151 Median : 6.000 Min1 : 65   
## Mean :1.045 TA :1492 Mean : 6.452 Mod : 35   
## 3rd Qu.:1.000 NA's: 1 3rd Qu.: 7.000 Maj1 : 19   
## Max. :3.000 Max. :15.000 (Other): 11   
## NA's : 2   
## Fireplaces FireplaceQu GarageType GarageYrBlt GarageFinish  
## Min. :0.0000 Ex : 43 2Types : 23 Min. :1895 Fin : 719   
## 1st Qu.:0.0000 Fa : 74 Attchd :1723 1st Qu.:1960 RFn : 811   
## Median :1.0000 Gd : 744 Basment: 36 Median :1979 Unf :1230   
## Mean :0.5971 Po : 46 BuiltIn: 186 Mean :1978 NA's: 159   
## 3rd Qu.:1.0000 TA : 592 CarPort: 15 3rd Qu.:2002   
## Max. :4.0000 NA's:1420 Detchd : 779 Max. :2207   
## NA's : 157 NA's :159   
## GarageCars GarageArea GarageQual GarageCond   
## Min. :0.000 Min. : 0.0 Length:2919 Ex : 3   
## 1st Qu.:1.000 1st Qu.: 320.0 Class :character Fa : 74   
## Median :2.000 Median : 480.0 Mode :character Gd : 15   
## Mean :1.767 Mean : 472.9 Po : 14   
## 3rd Qu.:2.000 3rd Qu.: 576.0 TA :2654   
## Max. :5.000 Max. :1488.0 NA's: 159   
## NA's :1 NA's :1   
## PavedDrive WoodDeckSF OpenPorchSF EnclosedPorch   
## N: 216 Min. : 0.00 Min. : 0.00 Min. : 0.0   
## P: 62 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.0   
## Y:2641 Median : 0.00 Median : 26.00 Median : 0.0   
## Mean : 93.71 Mean : 47.49 Mean : 23.1   
## 3rd Qu.: 168.00 3rd Qu.: 70.00 3rd Qu.: 0.0   
## Max. :1424.00 Max. :742.00 Max. :1012.0   
##   
## X3SsnPorch ScreenPorch PoolArea PoolQC   
## Min. : 0.000 Min. : 0.00 Min. : 0.000 Length:2919   
## 1st Qu.: 0.000 1st Qu.: 0.00 1st Qu.: 0.000 Class :character   
## Median : 0.000 Median : 0.00 Median : 0.000 Mode :character   
## Mean : 2.602 Mean : 16.06 Mean : 2.252   
## 3rd Qu.: 0.000 3rd Qu.: 0.00 3rd Qu.: 0.000   
## Max. :508.000 Max. :576.00 Max. :800.000   
##   
## Fence MiscFeature MiscVal MoSold   
## GdPrv: 118 Length:2919 Min. : 0.00 Min. : 1.000   
## GdWo : 112 Class :character 1st Qu.: 0.00 1st Qu.: 4.000   
## MnPrv: 329 Mode :character Median : 0.00 Median : 6.000   
## MnWw : 12 Mean : 50.83 Mean : 6.213   
## NA's :2348 3rd Qu.: 0.00 3rd Qu.: 8.000   
## Max. :17000.00 Max. :12.000   
##   
## YrSold SaleType SaleCondition SalePrice   
## Min. :2006 WD :2525 Abnorml: 190 Min. : 34900   
## 1st Qu.:2007 New : 239 AdjLand: 12 1st Qu.:129975   
## Median :2008 COD : 87 Alloca : 24 Median :163000   
## Mean :2008 ConLD : 26 Family : 46 Mean :180921   
## 3rd Qu.:2009 CWD : 12 Normal :2402 3rd Qu.:214000   
## Max. :2010 (Other): 29 Partial: 245 Max. :755000   
## NA's : 1 NA's :1459

Checking for MISSING VALUES

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

## [1] 0.001810209

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

## [1] 2919

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

## 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

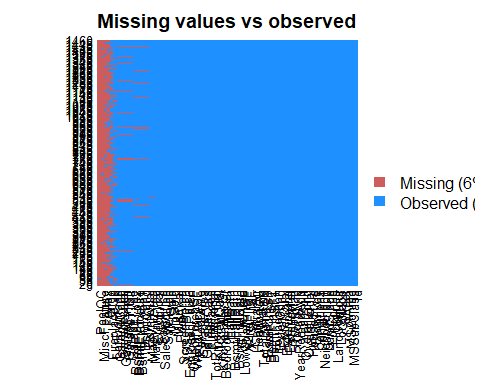
library(Amelia)

## 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")



# 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')

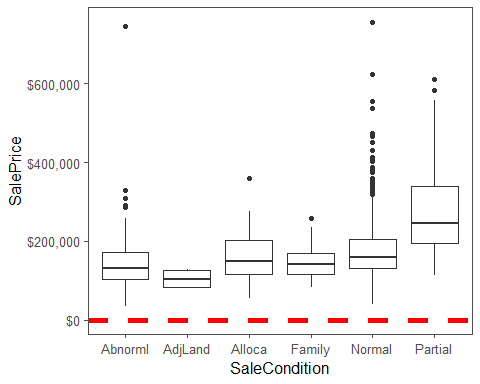
unique(nrow(training$SalePrice))

## NULL

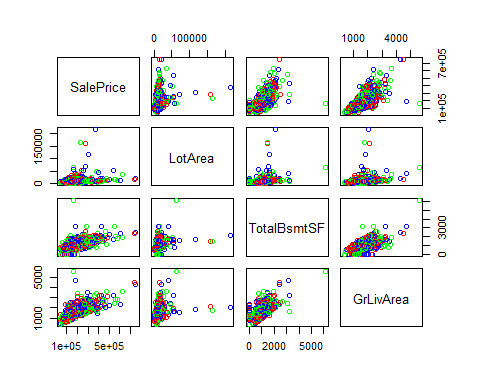
attach(housedata)   
catdf<-housedata[,catvar]  
numdf<-housedata[,numvar]

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()



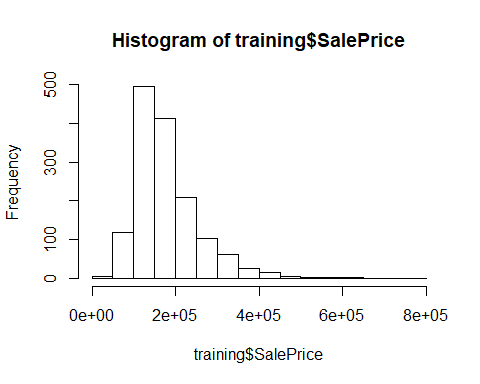
pairs(~SalePrice+LotArea+TotalBsmtSF+GrLivArea, data=training,col=c('red','blue','green'))



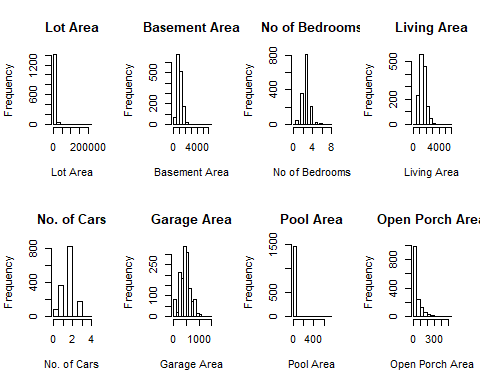
as.factor(training$SalePrice)

## [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

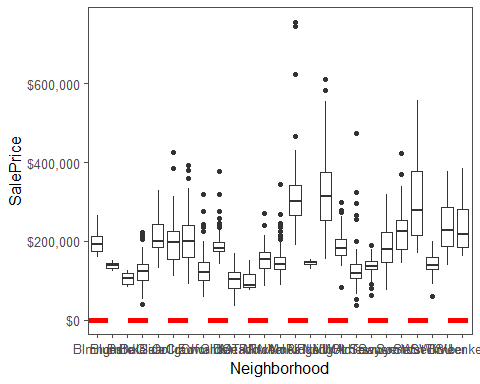
hist(training$SalePrice)



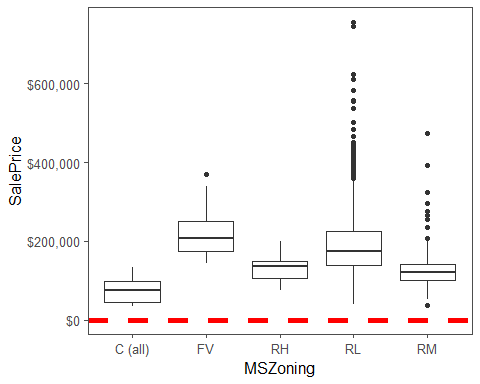
par(mfrow=c(2,4))  
hist(training$LotArea,xlab="Lot Area", main="Lot Area")  
hist(training$TotalBsmtSF, xlab="Basement Area", main="Basement Area")  
hist(training$BedroomAbvGr, xlab="No of Bedrooms", main="No of Bedrooms")  
hist(training$GrLivArea, xlab="Living Area",main="Living Area")  
hist(training$GarageCars,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



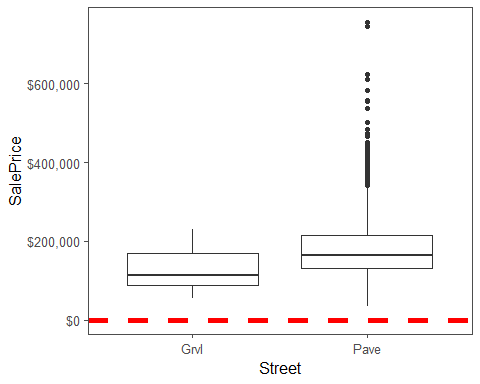
ggplot(training, aes(x = Neighborhood, y = SalePrice)) +  
 geom\_boxplot() +  
 geom\_hline(aes(yintercept=80),   
 colour='red', linetype='dashed', lwd=2) +  
 scale\_y\_continuous(labels=dollar\_format()) +  
 theme\_few()



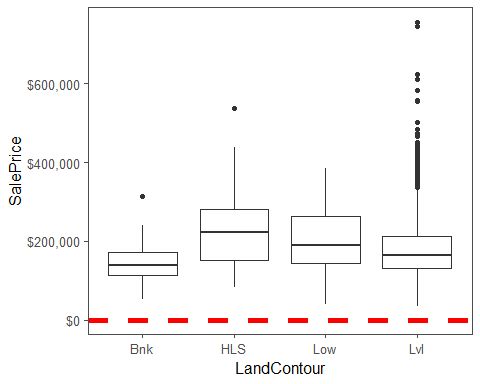
ggplot(training, aes(x = MSZoning, y = SalePrice)) +geom\_boxplot() +  
 geom\_hline(aes(yintercept=80),   
 colour='red', linetype='dashed', lwd=2) +  
 scale\_y\_continuous(labels=dollar\_format()) +  
 theme\_few()



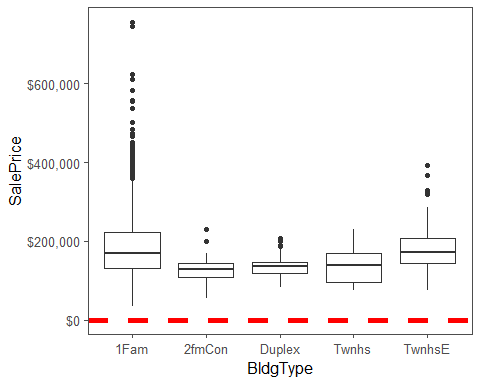
ggplot(training, aes(x = Street, y = SalePrice)) +geom\_boxplot() +  
 geom\_hline(aes(yintercept=80),   
 colour='red', linetype='dashed', lwd=2) +  
 scale\_y\_continuous(labels=dollar\_format()) +  
 theme\_few()



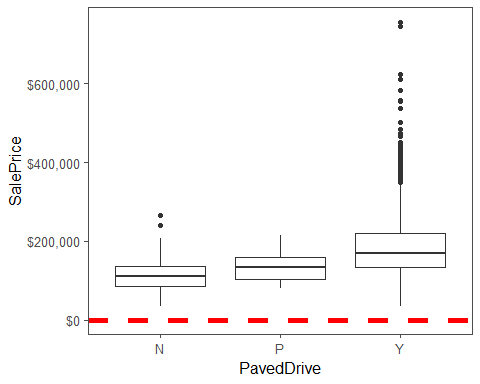
ggplot(training, aes(x = LandContour, y = SalePrice)) +geom\_boxplot() +  
 geom\_hline(aes(yintercept=80),   
 colour='red', linetype='dashed', lwd=2) +  
 scale\_y\_continuous(labels=dollar\_format()) +  
 theme\_few()



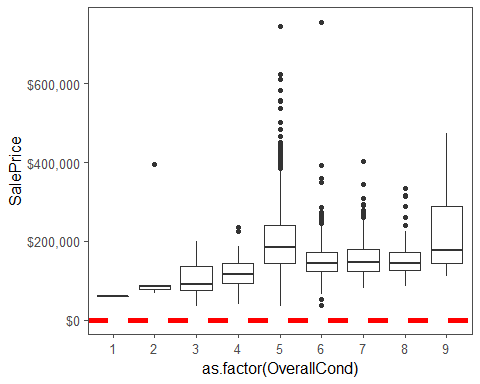
ggplot(training, aes(x = BldgType, y = SalePrice)) +geom\_boxplot() +  
 geom\_hline(aes(yintercept=80),   
 colour='red', linetype='dashed', lwd=2) +  
 scale\_y\_continuous(labels=dollar\_format()) +  
 theme\_few()



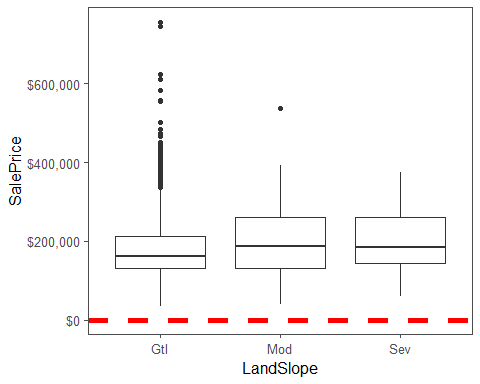
ggplot(training, aes(x = PavedDrive, y = SalePrice)) +geom\_boxplot() +  
 geom\_hline(aes(yintercept=80),   
 colour='red', linetype='dashed', lwd=2) +  
 scale\_y\_continuous(labels=dollar\_format()) +  
 theme\_few()



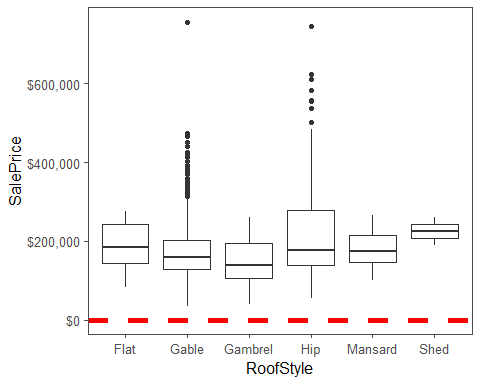
ggplot(training, aes(x = as.factor(OverallCond), y = SalePrice)) +geom\_boxplot() +  
 geom\_hline(aes(yintercept=80),   
 colour='red', linetype='dashed', lwd=2) +  
 scale\_y\_continuous(labels=dollar\_format()) +  
 theme\_few()



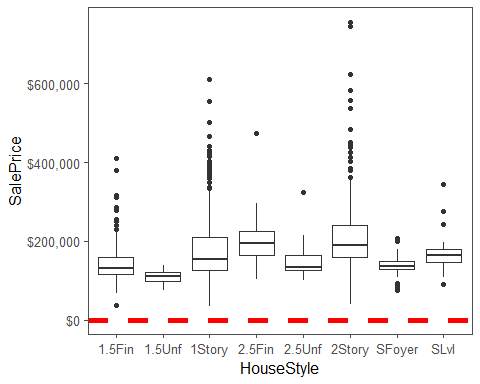
ggplot(training, aes(x = LandSlope, y = SalePrice)) +geom\_boxplot() +  
 geom\_hline(aes(yintercept=80),   
 colour='red', linetype='dashed', lwd=2) +  
 scale\_y\_continuous(labels=dollar\_format()) +  
 theme\_few()



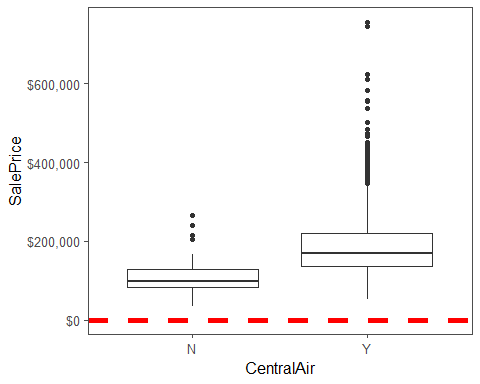
ggplot(training, aes(x = RoofStyle, y = SalePrice)) +geom\_boxplot() +  
 geom\_hline(aes(yintercept=80),   
 colour='red', linetype='dashed', lwd=2) +  
 scale\_y\_continuous(labels=dollar\_format()) +  
 theme\_few()



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'

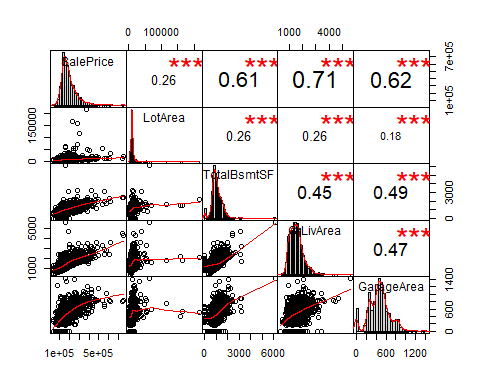
## The following objects are masked from 'package:data.table':  
##   
## first, last

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

##   
## Attaching package: 'PerformanceAnalytics'

## The following 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)

