Prediction of sale prices of house

Problem Statement

To predict the sale prices of houses

Project Team

1. Tejaswini Nutalapati
2. Aditi Bhargava

About the Data

Ask a home buyer to describe their dream house, and they probably won't begin with the height of the basement ceiling or the proximity to an east-west railroad. But this dataset proves that much more influences price negotiations than the number of bedrooms or a white-picket fence. With 79 explanatory variables describing (almost) every aspect of residential homes in Ames, Iowa, this dataset allows us to predict the final price of each home.

The Ames Housing dataset was compiled by Dean De Cock for use in data science education. It's an incredible alternative for data scientists looking for a modernized and expanded version of the often-cited Boston Housing dataset.

Data Souce: https://www.kaggle.com/c/house-prices-advanced-regression-techniques

Principle Component Analysis

```
#loading the libraries

library(reshape2)
library(dplyr)

##
## 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(ggplot2)
library(lattice)
library(caret)
library(scales)
library(dummies)
```

```
## dummies-1.5.6 provided by Decision Patterns
```

```
library(fmsb)
```

```
## Registered S3 methods overwritten by 'fmsb':
##   method    from
##   print.roc pROC
##   plot.roc  pROC
```

```
library(randomForest)
```

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

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

```
##
## Attaching package: 'randomForest'
```

```
## The following object is masked from 'package:ggplot2':
##
##     margin
```

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

```
library(DescTools)
```

```
##
## Attaching package: 'DescTools'
```

```
## The following objects are masked from 'package:fmsb':
##
##     CronbachAlpha, VIF
```

```
## The following objects are masked from 'package:caret':
##
##     MAE, RMSE
```

```
library(outliers)
```

```
##
## Attaching package: 'outliers'
## The following object is masked from 'package:randomForest':
## 
##     outlier
```

```r
library(VIM)
```

```
## Loading required package: colorspace
```

```
## Loading required package: grid
```

```
## Loading required package: data.table
```

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

```
## The following object is masked from 'package:DescTools':
## 
##     %like%
```

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

```
## The following objects are masked from 'package:reshape2':
## 
##     dcast, melt
```

```
## VIM is ready to use. 
##  Since version 4.0.0 the GUI is in its own package VIMGUI.
## 
##           Please use the package to use the new (and old) GUI.
```

```
## Suggestions and bug-reports can be submitted at: https://github.com/alexko
wa/VIM/issues
```

```
## 
## Attaching package: 'VIM'
```

```
## The following object is masked from 'package:datasets':
## 
##     sleep
```

```r
library(GGally)
```

```
## Registered S3 method overwritten by 'GGally':
##   method from   
##   +.gg   ggplot2
```

```r
library(corrplot)
```

```
## corrplot 0.84 loaded
```

```r
library(ggfortify)
```

```
# Loading the dataset
list.files("../input")

## character(0)

Train<-read.csv("C:/Users/aditi/OneDrive/Desktop/MVA/train.csv")
Test<-read.csv("C:/Users/aditi/OneDrive/Desktop/MVA/test.csv")

# Add sale price new column in test dataset
Test["SalePrice"] <- NA

# Let's explore the structure of the data
dim(Train)

## [1] 1460   81

str(Train)

## 'data.frame':    1460 obs. of  81 variables:
##  $ Id           : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ MSSubClass   : int  60 20 60 70 60 50 20 60 50 190 ...
##  $ MSZoning     : Factor w/ 5 levels "C (all)","FV",..: 4 4 4 4 4 4 4 4 5
4 ...
##  $ LotFrontage  : int  65 80 68 60 84 85 75 NA 51 50 ...
##  $ LotArea      : int  8450 9600 11250 9550 14260 14115 10084 10382 6120 7
420 ...
##  $ 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 N
A 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 1
3 13 13 7 4 9 ...
##  $ Exterior2nd  : Factor w/ 16 levels "AsbShng","AsphShn",..: 14 9 14 16 1
4 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 N
A 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 20
0000 129900 118000 ...
```

```
dim(Test)
```

```
## [1] 1459   81
```

```
str(Test)
```

```
## 'data.frame':    1459 obs. of  81 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 8
400 ...
##  $ 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 N
A 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 ...
##  $ SalePrice    : logi  NA NA NA NA NA NA ...
```

```
#The categorical variables are stored as factors in our dataframe.

# Combining the dataset
Test$SalePrice <- -1
df <- rbind(Train,Test)
str(df)
```

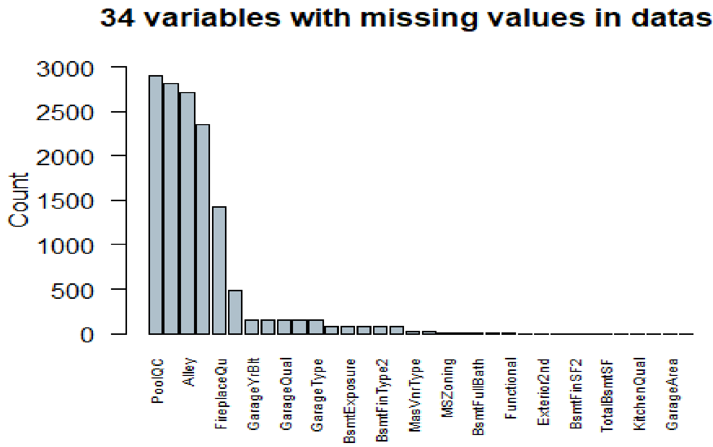
```
## 'data.frame':    2919 obs. of  81 variables:
##  $ Id           : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ MSSubClass   : int  60 20 60 70 60 50 20 60 50 190 ...
##  $ MSZoning     : Factor w/ 5 levels "C (all)","FV",..: 4 4 4 4 4 4 4 4 5
4 ...
##  $ LotFrontage  : int  65 80 68 60 84 85 75 NA 51 50 ...
##  $ LotArea      : int  8450 9600 11250 9550 14260 14115 10084 10382 6120 7
420 ...
##  $ 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 N
A 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 1
3 13 13 7 4 9 ...
##  $ Exterior2nd  : Factor w/ 16 levels "AsbShng","AsphShn",..: 14 9 14 16 1
4 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 N
A 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    : num  208500 181500 223500 140000 250000 ...
```

```
summary(df)
```

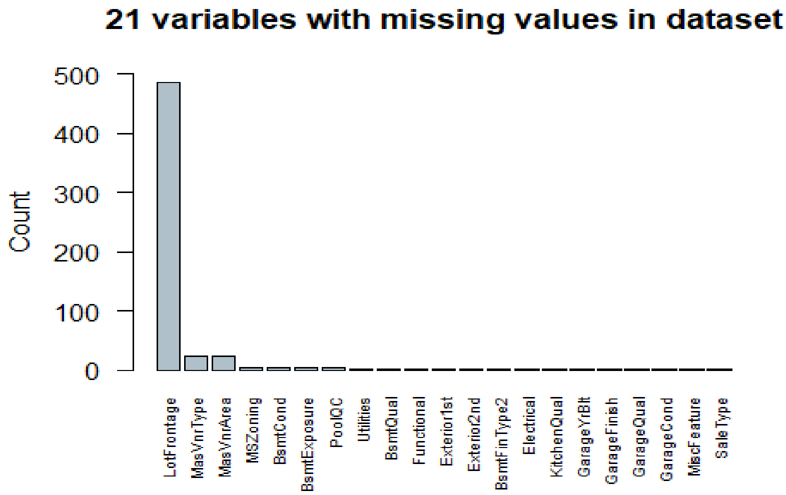
```
##        Id           MSSubClass       MSZoning     LotFrontage
##  Min.   :   1.0   Min.   : 20.00   C (all):  25   Min.   : 21.00
##  1st Qu.: 730.5   1st Qu.: 20.00   FV     : 139   1st Qu.: 59.00
##  Median :1460.0   Median : 50.00   RH     :  26   Median : 68.00
##  Mean   :1460.0   Mean   : 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  Utilitie
s
##  Min.   :  1300   Grvl:  12   Grvl: 120   IR1: 968   Bnk: 117    AllPub:29
16
##  1st Qu.:  7478   Pave:2907   Pave:  78   IR2:  76   HLS: 120    NoSeWa:
1
##  Median :  9453               NA's:2721   IR3:  16   Low:  60    NA's  :
2
##  Mean   : 10168                           Reg:1859   Lvl:2622
##  3rd Qu.: 11570
##  Max.   :215245
##
##    LotConfig    LandSlope   Neighborhood   Condition1     Condition2
##  Corner : 511   Gtl:2778   NAmes  : 443   Norm   :2511   Norm   :2889
##  CulDSac: 176   Mod: 125   CollgCr: 267   Feedr  : 164   Feedr  :  13
##  FR2    :  85   Sev:  16   OldTown: 239   Artery :  92   Artery :   5
##  FR3    :  14              Edwards: 194   RRAn   :  50   PosA   :   4
##  Inside :2133              Somerst: 182   PosN   :  39   PosN   :   4
##                            NridgHt: 166   RRAe   :  28   RRNn   :   2
##                            (Other):1428   (Other):  35   (Other):   2
##    BldgType      HouseStyle    OverallQual      OverallCond      YearBuilt
##  1Fam  :2425   1Story :1471   Min.   : 1.000   Min.   :1.000   Min.   :187
2
##  2fmCon:  62   2Story : 872   1st Qu.: 5.000   1st Qu.:5.000   1st Qu.:195
4
##  Duplex: 109   1.5Fin : 314   Median : 6.000   Median :5.000   Median :197
3
##  Twnhs :  96   SLvl   : 128   Mean   : 6.089   Mean   :5.565   Mean   :197
1
##  TwnhsE: 227   SFoyer :  83   3rd Qu.: 7.000   3rd Qu.:6.000   3rd Qu.:200
1
##                2.5Unf :  24   Max.   :10.000   Max.   :9.000   Max.   :201
0
##                (Other):  27
##   YearRemodAdd    RoofStyle       RoofMatl     Exterior1st    Exterior2nd
##  Min.   :1950   Flat   :  20   CompShg:2876   VinylSd:1025   VinylSd:1014
##  1st Qu.:1965   Gable  :2310   Tar&Grv:  23   MetalSd: 450   MetalSd: 447
##  Median :1993   Gambrel:  22   WdShake:   9   HdBoard: 442   HdBoard: 406
##  Mean   :1984   Hip    : 551   WdShngl:   7   Wd Sdng: 411   Wd Sdng: 391
##  3rd Qu.:2004   Mansard:  11   ClyTile:   1   Plywood: 221   Plywood: 270
##  Max.   :2010   Shed   :   5   Membran:   1   (Other): 369   (Other): 390
##                                (Other):   2   NA's   :   1   NA's   :   1
##    MasVnrType     MasVnrArea     ExterQual ExterCond  Foundation   BsmtQua
l
##  BrkCmn :  25   Min.   :   0.0   Ex: 107   Ex:  12   BrkTil: 311   Ex  : 2
58
##  BrkFace: 879   1st Qu.:   0.0   Fa:  35   Fa:  67   CBlock:1235   Fa  :  
88
##  None   :1742   Median :   0.0   Gd: 979   Gd: 299   PConc :1308   Gd  :12
09
##  Stone  : 249   Mean   : 102.2   TA:1798   Po:   3   Slab  :  49   TA  :12
83
##  NA's   :  24   3rd Qu.: 164.0             TA:2538   Stone :  11   NA's:  
81
##                 Max.   :1600.0                       Wood  :   5
##                 NA's   :23
##  BsmtCond    BsmtExposure BsmtFinType1   BsmtFinSF1     BsmtFinType2
##  Fa  : 104   Av  : 418    ALQ :429     Min.   :   0.0   ALQ :  52
##  Gd  : 122   Gd  : 276    BLQ :269     1st Qu.:   0.0   BLQ :  68
##  Po  :   5   Mn  : 239    GLQ :849     Median : 368.5   GLQ :  34
##  TA  :2606   No  :1904    LwQ :154     Mean   : 441.4   LwQ :  87
##  NA's:  82   NA's:  82    Rec :288     3rd Qu.: 733.0   Rec : 105
##                           Unf :851     Max.   :5644.0   Unf :2493
##                           NA's: 79     NA's   :1        NA's:  80
##    BsmtFinSF2        BsmtUnfSF       TotalBsmtSF      Heating     HeatingQ
C
##  Min.   :   0.00   Min.   :   0.0   Min.   :   0.0   Floor:   1   Ex:1493
##  1st Qu.:   0.00   1st Qu.: 220.0   1st Qu.: 793.0   GasA :2874   Fa:  92
##  Median :   0.00   Median : 467.0   Median : 989.5   GasW :  27   Gd: 474
##  Mean   :  49.58   Mean   : 560.8   Mean   :1051.8   Grav :   9   Po:   3
##  3rd Qu.:   0.00   3rd Qu.: 805.5   3rd Qu.:1302.0   OthW :   2   TA: 857
##  Max.   :1526.00   Max.   :2336.0   Max.   :6110.0   Wall :   6
##  NA's   :1         NA's   :1        NA's   :1
##  CentralAir Electrical     X1stFlrSF      X2ndFlrSF       LowQualFinSF
##  N: 196     FuseA: 188   Min.   : 334   Min.   :   0.0   Min.   :   0.000
##  Y:2723     FuseF:  50   1st Qu.: 876   1st Qu.:   0.0   1st Qu.:   0.000
##             FuseP:   8   Median :1082   Median :   0.0   Median :   0.000
##             Mix  :   1   Mean   :1160   Mean   : 336.5   Mean   :   4.694
##             SBrkr:2671   3rd Qu.:1388   3rd Qu.: 704.0   3rd Qu.:   0.000
##             NA's :   1   Max.   :5095   Max.   :2065.0   Max.   :1064.000
##
##    GrLivArea     BsmtFullBath     BsmtHalfBath       FullBath
##  Min.   : 334   Min.   :0.0000   Min.   :0.00000   Min.   :0.000
##  1st Qu.:1126   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:1.000
##  Median :1444   Median :0.0000   Median :0.00000   Median :2.000
##  Mean   :1501   Mean   :0.4299   Mean   :0.06136   Mean   :1.568
##  3rd Qu.:1744   3rd Qu.:1.0000   3rd Qu.:0.00000   3rd Qu.:2.000
##  Max.   :5642   Max.   :3.0000   Max.   :2.00000   Max.   :4.000
##                 NA's   :2        NA's   :2
##     HalfBath       BedroomAbvGr   KitchenAbvGr   KitchenQual  TotRmsAbvGrd
##  Min.   :0.0000   Min.   :0.00   Min.   :0.000   Ex  : 205   Min.   : 2.00
0
##  1st Qu.:0.0000   1st Qu.:2.00   1st Qu.:1.000   Fa  :  70   1st Qu.: 5.00
0
##  Median :0.0000   Median :3.00   Median :1.000   Gd  :1151   Median : 6.00
0
##  Mean   :0.3803   Mean   :2.86   Mean   :1.045   TA  :1492   Mean   : 6.45
2
##  3rd Qu.:1.0000   3rd Qu.:3.00   3rd Qu.:1.000   NA's:   1   3rd Qu.: 7.00
0
##  Max.   :2.0000   Max.   :8.00   Max.   :3.000               Max.   :15.00
0
##
##    Functional     Fireplaces     FireplaceQu   GarageType    GarageYrBlt
##  Typ    :2717   Min.   :0.0000   Ex  :  43   2Types :  23   Min.   :1895
##  Min2   :  70   1st Qu.:0.0000   Fa  :  74   Attchd :1723   1st Qu.:1960
##  Min1   :  65   Median :1.0000   Gd  : 744   Basment:  36   Median :1979
##  Mod    :  35   Mean   :0.5971   Po  :  46   BuiltIn: 186   Mean   :1978
##  Maj1   :  19   3rd Qu.:1.0000   TA  : 592   CarPort:  15   3rd Qu.:2002
##  (Other):  11   Max.   :4.0000   NA's:1420   Detchd : 779   Max.   :2207
##  NA's   :   2                                NA's   : 157   NA's   :159
##  GarageFinish GarageCars        GarageArea     GarageQual  GarageCond
##  Fin : 719    Min.   :0.000   Min.   :   0.0   Ex  :   3   Ex  :   3
##  RFn : 811    1st Qu.:1.000   1st Qu.: 320.0   Fa  : 124   Fa  :  74
##  Unf :1230    Median :2.000   Median : 480.0   Gd  :  24   Gd  :  15
##  NA's: 159    Mean   :1.767   Mean   : 472.9   Po  :   5   Po  :  14
##               3rd Qu.:2.000   3rd Qu.: 576.0   TA  :2604   TA  :2654
##               Max.   :5.000   Max.   :1488.0   NA's: 159   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       Fence
##  Min.   :  0.000   Min.   :  0.00   Min.   :  0.000   Ex  :   4   GdPrv: 1
18
##  1st Qu.:  0.000   1st Qu.:  0.00   1st Qu.:  0.000   Fa  :   2   GdWo : 1
12
##  Median :  0.000   Median :  0.00   Median :  0.000   Gd  :   4   MnPrv: 3
29
##  Mean   :  2.602   Mean   : 16.06   Mean   :  2.252   NA's:2909   MnWw :
12
##  3rd Qu.:  0.000   3rd Qu.:  0.00   3rd Qu.:  0.000               NA's :23
48
##  Max.   :508.000   Max.   :576.00   Max.   :800.000
##
##  MiscFeature    MiscVal             MoSold          YrSold         SaleTyp
e
##  Gar2:   5   Min.   :    0.00   Min.   : 1.000   Min.   :2006   WD     :25
25
##  Othr:   4   1st Qu.:    0.00   1st Qu.: 4.000   1st Qu.:2007   New    : 2
39
##  Shed:  95   Median :    0.00   Median : 6.000   Median :2008   COD    :
87
##  TenC:   1   Mean   :   50.83   Mean   : 6.213   Mean   :2008   ConLD  :
26
##  NA's:2814   3rd Qu.:    0.00   3rd Qu.: 8.000   3rd Qu.:2009   CWD    :
12
##              Max.   :17000.00   Max.   :12.000   Max.   :2010   (Other):
29
##                                                                 NA's   :
1
##  SaleCondition    SalePrice
##  Abnorml: 190   Min.   :    -1
##  AdjLand:  12   1st Qu.:    -1
##  Alloca :  24   Median : 34900
##  Family :  46   Mean   : 90491
##  Normal :2402   3rd Qu.:163000
##  Partial: 245   Max.   :755000
##

#finding how many variables with missing values are in the dataset
options(repr.plot.width=6, repr.plot.height=5)
cMiss = function(x){sum(is.na(x))}
CM <- sort(apply(df,2,cMiss),decreasing=T);
barplot(CM[CM!=0],
        las=2,
        cex.names=0.6,
        ylab="Count",
        ylim=c(0,3000),
        horiz=F,
        col="#AFC0CB",
        main=paste(toString(sum(CM!=0)), "variables with missing values in da
taset"))
```



```
dfClean <-function(df)
{
  # Pool Variable: If PoolQC = NA and PoolArea = 0 , assign factor NoPool
  df$PoolQC <- as.character(df$PoolQC)
  df$PoolQC[df$PoolArea %in% c(0,NA) & is.na(df$PoolQC)] <- "NoPool"
  df$PoolQC <- as.factor(df$PoolQC)

  # MiscFeature Variable: If MiscFeature = NA and MiscVal = 0, assign factor
None
  df$MiscFeature <- as.character(df$MiscFeature)
  df$MiscFeature[df$MiscVal %in% c(0,NA) & is.na(df$MiscFeature)] <- "None"
  df$MiscFeature <- as.factor(df$MiscFeature)

  # Alley Variable: If Alley = NA, assign factor NoAccess
  df$Alley <- as.character(df$Alley)
  df$Alley[is.na(df$Alley)] <- "NoAccess"
  df$Alley <- as.factor(df$Alley)

  # Fence Variable: If Fence = NA, assign factor NoFence
  df$Fence <- as.character(df$Fence)
  df$Fence[is.na(df$Fence)] <- "NoFence"
  df$Fence <- as.factor(df$Fence)

  # FireplaceQu Variable: If FireplaceQu = NA and Fireplaces = 0 , assign fac
tor NoFirePlace
  df$FireplaceQu <- as.character(df$FireplaceQu)
  df$FireplaceQu[df$Fireplaces %in% c(0,NA) & is.na(df$FireplaceQu)] <- "NoFi
rePlace"
  df$FireplaceQu <- as.factor(df$FireplaceQu)

  # GarageYrBlt Variable: If GarageYrBlt = NA and GarageArea = 0 assign facto
r NoGarage
  df$GarageYrBlt <- as.character(df$GarageYrBlt)
  df$GarageYrBlt[df$GarageArea %in% c(0,NA) & is.na(df$GarageYrBlt)] <- "NoGa
rage"
  df$GarageYrBlt <- as.factor(df$GarageYrBlt)

  # GarageFinish Variable: If GarageFinish = NA and GarageArea = 0 assign fac
tor NoGarage
  df$GarageFinish <- as.character(df$GarageFinish)
  df$GarageFinish[df$GarageArea %in% c(0,NA) & is.na(df$GarageFinish)] <- "No
Garage"
  df$GarageFinish <- as.factor(df$GarageFinish)

  # GarageQual Variable: If GarageQual = NA and GarageArea = 0 assign factor
NoGarage
  df$GarageQual <- as.character(df$GarageQual)
  df$GarageQual[df$GarageArea %in% c(0,NA) & is.na(df$GarageQual)] <- "NoGara
ge"
  df$GarageQual <- as.factor(df$GarageQual)

  # GarageCond Variable: If GarageCond = NA and GarageArea = 0 assign factor
NoGarage
  df$GarageCond <- as.character(df$GarageCond)
  df$GarageCond[df$GarageArea %in% c(0,NA) & is.na(df$GarageCond)] <- "NoGara
ge"
  df$GarageCond <- as.factor(df$GarageCond)

  # GarageType Variable: If GarageType = NA and GarageArea = 0 assign factor
NoGarage
  df$GarageType <- as.character(df$GarageType)
  df$GarageType[df$GarageArea %in% c(0,NA) & is.na(df$GarageType)] <- "NoGara
ge"
  df$GarageType <- as.factor(df$GarageType)
  df$GarageArea[is.na(df$GarageArea) & df$GarageCars %in% c(0,NA)] <- 0
  df$GarageCars[is.na(df$GarageCars) & df$GarageArea %in% c(0,NA)] <- 0

  # BsmtFullBath Variable: If BsmtFullBath = NA and TotalBsmtSF = 0 assign 0
  df$BsmtFullBath[df$TotalBsmtSF %in% c(0,NA) & is.na(df$BsmtFullBath)] <- 0

  # BsmtHalfBath Variable: If BsmtHalfBath = NA and TotalBsmtSF = 0 assign 0
  df$BsmtHalfBath[df$TotalBsmtSF %in% c(0,NA) & is.na(df$BsmtHalfBath)] <- 0

  # BsmtFinSF1 Variable: If BsmtFinSF1 = NA and TotalBsmtSF = 0 assign 0
  df$BsmtFinSF1[df$TotalBsmtSF %in% c(0,NA) & is.na(df$BsmtFinSF1)] <- 0
```

```r
  # BsmtFinSF2 Variable: If BsmtFinSF2 = NA and TotalBsmtSF = 0 assign 0
  df$BsmtFinSF2[df$TotalBsmtSF %in% c(0,NA) & is.na(df$BsmtFinSF2)] <- 0

  # BsmtUnfSF Variable: If BsmtUnfSF = NA and TotalBsmtSF = 0 assign 0
  df$BsmtUnfSF[df$TotalBsmtSF %in% c(0,NA) & is.na(df$BsmtUnfSF)] <- 0

  # TotalBsmtSF Variable: If TotalBsmtSF = NA and TotalBsmtSF = 0 assign 0
  df$TotalBsmtSF[df$TotalBsmtSF %in% c(0,NA) & is.na(df$TotalBsmtSF)] <- 0

  # BsmtQual Variable: If BsmtQual = NA and TotalBsmtSF = 0 assign factor NoBasement
  df$BsmtQual <- as.character(df$BsmtQual)
  df$BsmtQual[df$TotalBsmtSF %in% c(0,NA) & is.na(df$BsmtQual)] <- "NoBasement"
  df$BsmtQual <- as.factor(df$BsmtQual)

  # BsmtFinType1 Variable: If BsmtFinType1 = NA and TotalBsmtSF = 0 assign factor NoBasement
  df$BsmtFinType1 <- as.character(df$BsmtFinType1)
  df$BsmtFinType1[df$TotalBsmtSF %in% c(0,NA) & is.na(df$BsmtFinType1)] <- "NoBasement"
  df$BsmtFinType1 <- as.factor(df$BsmtFinType1)

  # BsmtFinType2 Variable: If BsmtFinType2 = NA and TotalBsmtSF = 0 assign factor NoBasement
  df$BsmtFinType2 <- as.character(df$BsmtFinType2)
  df$BsmtFinType2[df$TotalBsmtSF %in% c(0,NA) & is.na(df$BsmtFinType2)] <- "NoBasement"
  df$BsmtFinType2 <- as.factor(df$BsmtFinType2)

  # BsmtExposure Variable: If BsmtExposure = NA and TotalBsmtSF = 0 assign factor NoBasement
  df$BsmtExposure <- as.character(df$BsmtExposure)
  df$BsmtExposure[df$TotalBsmtSF %in% c(0,NA) & is.na(df$BsmtExposure)] <- "NoBasement"
  df$BsmtExposure <- as.factor(df$BsmtExposure)

  # BsmtCond Variable: If BsmtCond = NA and TotalBsmtSF = 0 assign factor NoBasement
  df$BsmtCond <- as.character(df$BsmtCond)
  df$BsmtCond[df$TotalBsmtSF %in% c(0,NA) & is.na(df$BsmtCond)] <- "NoBasement"
  df$BsmtCond <- as.factor(df$BsmtCond)
  return(df)
}
df <- dfClean(df)

PM <- sort(apply(df,2,cMiss),decreasing=T);
```

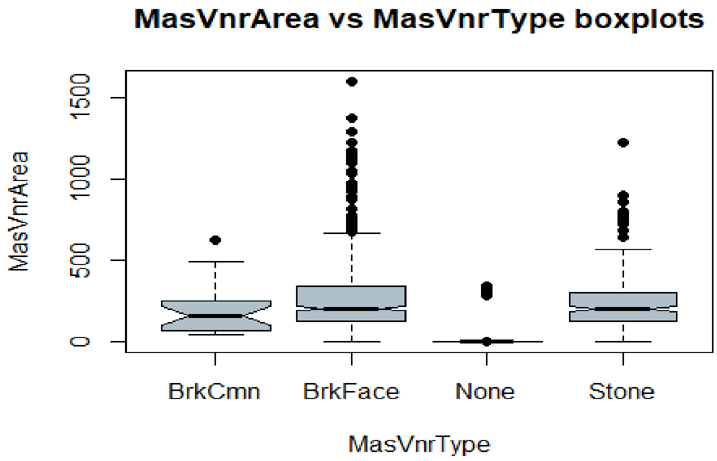
```
barplot(PM[PM!=0],
        las=2,
        cex.names=0.6,
        ylab="Count",
        ylim=c(0,500),
        horiz=F,
        col="#AFC0CB",
        main=paste(toString(sum(PM!=0)), "variables with missing values in da
taset"))
```



```
#That certainly helped a little bit. Let's see if there's a pattern to the re
maining missing data.
data = df[, names(PM[PM!=0])];
aggr_plot <- aggr(data,
                  col=c('navyblue','red'),
                  bars=T,
                  numbers=T,
                  combined = T,
                  labels=names(data),
                  cex.axis=.7,
                  gap=3,
                  ylab=c("Pattern"),
                  cex.numbers=0.74)
```

```
## Warning in plot.aggr(res, ...): not enough horizontal space to display
## frequencies
```

```
#MasVnrType and MasVnrArea
plot(df[,c("MasVnrType","MasVnrArea")],
     pch=16,
     notch=TRUE,
     main="MasVnrArea vs MasVnrType boxplots",
     col="#AFC0CB")
```



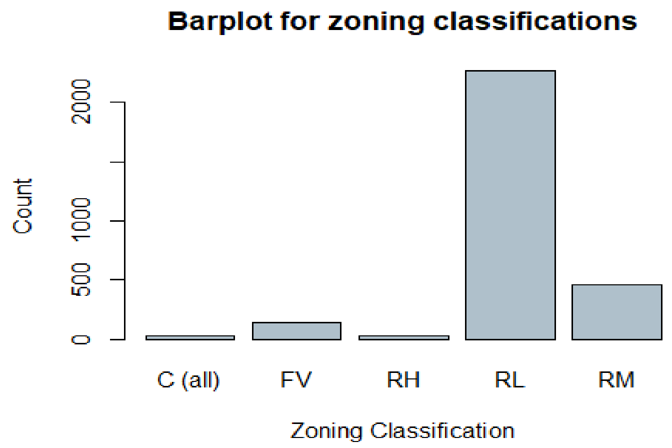
```
df[ (is.na(df$MasVnrType) | is.na(df$MasVnrArea)) ,c("MasVnrType","MasVnrArea")]
```

```
##      MasVnrType MasVnrArea
## 235        <NA>         NA
## 530        <NA>         NA
## 651        <NA>         NA
## 937        <NA>         NA
## 974        <NA>         NA
## 978        <NA>         NA
## 1244       <NA>         NA
## 1279       <NA>         NA
## 1692       <NA>         NA
## 1707       <NA>         NA
## 1883       <NA>         NA
## 1993       <NA>         NA
## 2005       <NA>         NA
## 2042       <NA>         NA
## 2312       <NA>         NA
## 2326       <NA>         NA
## 2341       <NA>         NA
## 2350       <NA>         NA
## 2369       <NA>         NA
## 2593       <NA>         NA
## 2611       <NA>        198
## 2658       <NA>         NA
## 2687       <NA>         NA
## 2863       <NA>         NA
```

```
summary(df[ !(is.na(df$MasVnrType) | is.na(df$MasVnrArea)) ,c("MasVnrType","MasVnrArea")])
```

```
##   MasVnrType     MasVnrArea
##  BrkCmn :  25   Min.   :   0.0
##  BrkFace: 879   1st Qu.:   0.0
##  None   :1742   Median :   0.0  
##  Stone  : 249   Mean   : 102.2  
##                 3rd Qu.: 164.0  
##                 Max.   :1600.0
```

```
df$MasVnrType <- as.character(df$MasVnrType)
df$MasVnrType[is.na(df$MasVnrType)] <- "None"
df$MasVnrType <- as.factor(df$MasVnrType)
df$MasVnrArea[is.na(df$MasVnrArea)] <- 0

#MSZoning
plot(df$MSZoning,
     col="#AFC0CB",
     xlab="Zoning Classification",
     ylab = "Count",
     main = "Barplot for zoning classifications")
```



```
df[ is.na(df$MSZoning) ,c("MSZoning","MSSubClass")]
```

```
##      MSZoning MSSubClass
## 1916     <NA>         30
## 2217     <NA>         20
## 2251     <NA>         70
## 2905     <NA>         20
```

```
ZoneClassTable <- table(df[ ,c("MSZoning","MSSubClass")])
ZoneClassTable
```

```
##          MSSubClass
## MSZoning   20   30   40   45   50   60   70   75   80   85   90  120  150  
160
##   C (all)    3    8    0    0    7    0    4    0    0    0    0    0    0  
0
##   FV        34    0    0    0    0   43    0    0    0    0    0   19    0  
43
##   RH         4    2    0    1    2    0    3    0    0    0    4    6    0
```

```
0
##    RL       1016   61     4     6  159  529   57     9  115   47   92  117     1
21
##    RM         20   67     2    11  119     3   63    14     3     1   13   40     0
64
##           MSSubClass
## MSZoning   180  190
##   C (all)    0    3
##   FV         0    0
##   RH         0    4
##   RL         0   31
##   RM        17   23
```

```
mosaicplot(ZoneClassTable,
           main="Mosaic Plot of MSZoning VS MSSubClass",
           las=1,
           color=T,
           shade=T)
```

```
GTest(ZoneClassTable)
```

```
##
##  Log likelihood ratio (G-test) test of independence without correction
##
## data:  ZoneClassTable
## G = 1321.9, X-squared df = 60, p-value < 2.2e-16
```

```
Table<-table(df[ df$MSSubClass %in% c(30,70) ,c("MSZoning","MSSubClass")])
Table <- Table[ , colSums(Table != 0) > 0 ]
Table
```

```
##           MSSubClass
## MSZoning  30 70
##   C (all)  8  4
##   FV       0  0
##   RH       2  3
##   RL      61 57
##   RM      67 63
```

```
mosaicplot(Table,
           main="Mosaic Plot of MSZoning VS MSSubClass (30,70)",
           las=1,
           color=T,
           shade=T)
```

```
Test1<-GTest(Table)
Test1
```

```
##
##  Log likelihood ratio (G-test) test of independence without correction
##
```

```
## data:  Table
## G = 1.3625, X-squared df = 4, p-value = 0.8507
```

```
paste("At a 95% confidence level, since the p-value =", as.character(round(Test1$p.value,2)),
      "> 0.05, we cannot reject the null hypothesis that MSZoning and MSSubClass are independent when MSSubClass = 30 or 70.")
```

```
## [1] "At a 95% confidence level, since the p-value = 0.85 > 0.05, we cannot reject the null hypothesis that MSZoning and MSSubClass are independent when MSSubClass = 30 or 70."
```

```
df$MSZoning <- as.character(df$MSZoning)
df$MSZoning[is.na(df$MSZoning)] <- "RL"
df$MSZoning <- as.factor(df$MSZoning)

#Basement
MissBsmt = c('BsmtCond','BsmtExposure','BsmtQual','BsmtFinType2')
df[!complete.cases(df[,names(df) %in% MissBsmt]),names(df) %in% names(df)[which(grepl("Bsmt",names(df)))]]
```

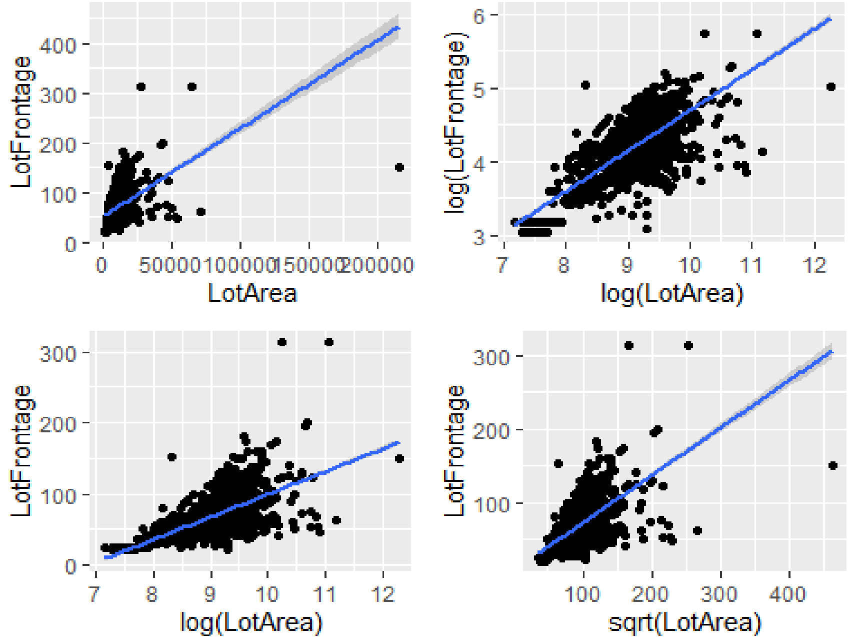
```
##      BsmtQual BsmtCond BsmtExposure BsmtFinType1 BsmtFinSF1 BsmtFinType2
## 333        Gd       TA           No          GLQ       1124         <NA>
## 949        Gd       TA         <NA>          Unf          0          Unf
## 1488       Gd       TA         <NA>          Unf          0          Unf
## 2041       Gd     <NA>           Mn          GLQ       1044          Rec
## 2186       TA     <NA>           No          BLQ       1033          Unf
## 2218     <NA>       Fa           No          Unf          0          Unf
## 2219     <NA>       TA           No          Unf          0          Unf
## 2349       Gd       TA         <NA>          Unf          0          Unf
## 2525       TA     <NA>           Av          ALQ        755          Unf
##      BsmtFinSF2 BsmtUnfSF TotalBsmtSF BsmtFullBath BsmtHalfBath
## 333         479      1603        3206            1            0
## 949           0       936         936            0            0
## 1488          0      1595        1595            0            0
## 2041        382         0        1426            1            0
## 2186          0        94        1127            0            1
## 2218          0       173         173            0            0
## 2219          0       356         356            0            0
## 2349          0       725         725            0            0
## 2525          0       240         995            0            0
```

```
#BsmtExposure
df$BsmtExposure <- as.character(df$BsmtExposure)
df$BsmtExposure[is.na(df$BsmtExposure)]<-"No"
df$BsmtExposure <- as.factor(df$BsmtExposure)

#BsmtFinType2
BsmtFinQuality<-table(df[ !(df$BsmtFinType2 %in% c("NoBasement","Unf") | df$BsmtFinType1 %in% c("NoBasement","Unf")) ,c("BsmtFinType2","BsmtFinType1")])
BsmtFinQuality<-BsmtFinQuality[rowSums(BsmtFinQuality != 0) > 0 , colSums(Bsm
tFinQuality != 0) > 0]
BsmtFinQuality

##             BsmtFinType1
## BsmtFinType2 ALQ BLQ GLQ LwQ Rec
##           ALQ   0   4  15  14  19
##           BLQ  30   1   7  11  19
##           GLQ   3  10   0  14   7
##           LwQ  27  23  17   0  20
##           Rec  36  34  19  16   0

mosaicplot(BsmtFinQuality,
           main="Mosaic Plot of BsmtFinType",
           las=1,
           color=T,
           shade=T)
#BsmtCond
TableBsmtCond<-table(df$HouseStyle,df$BsmtCond)
TableBsmtCond<-TableBsmtCond[rowSums(TableBsmtCond != 0) > 0 , colSums(TableB
smtCond != 0) > 0]
TableBsmtCond

## 
##            Fa   Gd NoBasement   Po   TA
##   1.5Fin   33    9          8    1  263
##   1.5Unf    3    0          0    0   16
##   1Story   31   60         59    3 1316
##   2.5Fin    2    0          0    0    6
##   2.5Unf    3    0          0    0   21
##   2Story   29   41         10    1  791
##   SFoyer    2    5          1    0   75
##   SLvl      1    7          1    0  118

mosaicplot(TableBsmtCond,
           main="Mosaic Plot of Basement Quality",
           las=1,
           color=T,
           shade=T)

TestQ2<-GTest(TableBsmtCond)
TestQ2

## 
##  Log likelihood ratio (G-test) test of independence without correction
## 
## data:  TableBsmtCond
## G = 89.202, X-squared df = 28, p-value = 2.64e-08

df$HouseStyle[is.na(df$BsmtCond)]

## [1] 1Story 1Story SLvl
## Levels: 1.5Fin 1.5Unf 1Story 2.5Fin 2.5Unf 2Story SFoyer SLvl
```

```r
df$BsmtCond <- as.character(df$BsmtCond)
df$BsmtCond[is.na(df$BsmtCond)]<-"TA"
df$BsmtCond <- as.factor(df$BsmtCond)

PM <- sort(apply(df,2,cMiss),decreasing=T);
barplot(PM[PM!=0],
        las=2,
        cex.names=0.6,
        ylab="Count",
        ylim=c(0,500),
        horiz=F,
        col="#AFC0CB",
        main=paste(toString(sum(PM!=0)), "variables with missing values in dataset")

data = df[, names(PM[PM!=0])];
aggr_plot <- aggr(data,
                  col=c('navyblue','red'),
                  bars=T,
                  numbers=T,
                  combined = T,
                  labels=names(data),
                  cex.axis=.7,
                  gap=3,
                  ylab=c("Pattern"),
                  cex.numbers=0.74)

#The rest
fillMiss<- function(x)
{
  ux <- unique(x[!is.na(x)])
  x <- as.character(x)
  mode <- ux[which.max(tabulate(match(x[!is.na(x)], ux)))]
  x[is.na(x)] <- as.character(mode)
  x <- as.factor(x)
  return(x)
}
df[,sapply(df,function(x){!(is.numeric(x))}) ]<-as.data.frame(apply(df[,sapply(df,function(x){!(is.numeric(x))}) ],2,fillMiss))
PM <- sort(apply(df,2,cMiss),decreasing=T);
barplot(PM[PM!=0],
        las=2,
        cex.names=0.6,
        ylab="Count",
        ylim=c(0,500),
        horiz=F,
        col="#AFC0CB",
        main=paste(toString(sum(PM!=0)), "variables with missing values in dataset"))
```

```
data = df[, names(PM[PM!=0])];
aggr_plot <- aggr(data,
                  col=c('navyblue','red'),
                  bars=T,
                  numbers=T,
                  combined = T,
                  labels=names(data),
                  cex.axis=.7,
                  gap=3,
                  ylab=c("Pattern"),
                  cex.numbers=0.74)

#LotFrontage Imputation

multiplot <- function(..., plotlist=NULL, file, cols=1, layout=NULL)
{
  library(grid)
  # Make a list from the ... arguments and plotlist
  plots <- c(list(...), plotlist)
  numPlots = length(plots)
  # If layout is NULL, then use 'cols' to determine layout
  if (is.null(layout))
  {
    # Make the panel
    # ncol: Number of columns of plots
    # nrow: Number of rows needed, calculated from # of cols
    layout <- matrix(seq(1, cols * ceiling(numPlots/cols)),
                     ncol = cols, nrow = ceiling(numPlots/cols))
  }
  if (numPlots==1)
  {
    print(plots[[1]])
  }
  else
  {
    # Set up the page
    grid.newpage()
    pushViewport(viewport(layout = grid.layout(nrow(layout), ncol(layout))))
    # Make each plot, in the correct location
    for (i in 1:numPlots)
    {
      # Get the i,j matrix positions of the regions that contain this subplot
      matchidx <- as.data.frame(which(layout == i, arr.ind = TRUE))
      print(plots[[i]], vp = viewport(layout.pos.row = matchidx$row,
                                      layout.pos.col = matchidx$col))
    }
  }
}
p1<-ggplot(df, aes(LotArea, LotFrontage)) + geom_point() + geom_smooth(method
= "lm", se = T)
```

```
p2<-ggplot(df, aes(log(LotArea), LotFrontage)) + geom_point() + geom_smooth(m
ethod = "lm", se = T)
p3<-ggplot(df, aes(log(LotArea), log(LotFrontage))) + geom_point() + geom_smo
oth(method = "lm", se = T)
p4<-ggplot(df, aes(sqrt(LotArea), LotFrontage)) + geom_point() + geom_smooth(
method = "lm", se = T)
multiplot(p1, p2, p3, p4, cols=2)
```



```
#To check outliers
chisq.out.test(df$LotArea,opposite=F)

##
##  chi-squared test for outlier
##
## data:  df$LotArea
## X-squared = 676.1, p-value < 2.2e-16
## alternative hypothesis: highest value 215245 is an outlier

chisq.out.test(df$LotFrontage,opposite=F)

##
##  chi-squared test for outlier
##
## data:  df$LotFrontage
## X-squared = 108.97, p-value < 2.2e-16
## alternative hypothesis: highest value 313 is an outlier

chisq.out.test(df$LotArea,opposite=T)

##
##  chi-squared test for outlier
```

```
##
## data:  df$LotArea
## X-squared = 1.2643, p-value = 0.2608
## alternative hypothesis: lowest value 1300 is an outlier
```

```
chisq.out.test(df$LotFrontage,opposite=T)
```

```
##
##  chi-squared test for outlier
##
## data:  df$LotFrontage
## X-squared = 4.2817, p-value = 0.03853
## alternative hypothesis: lowest value 21 is an outlier
```

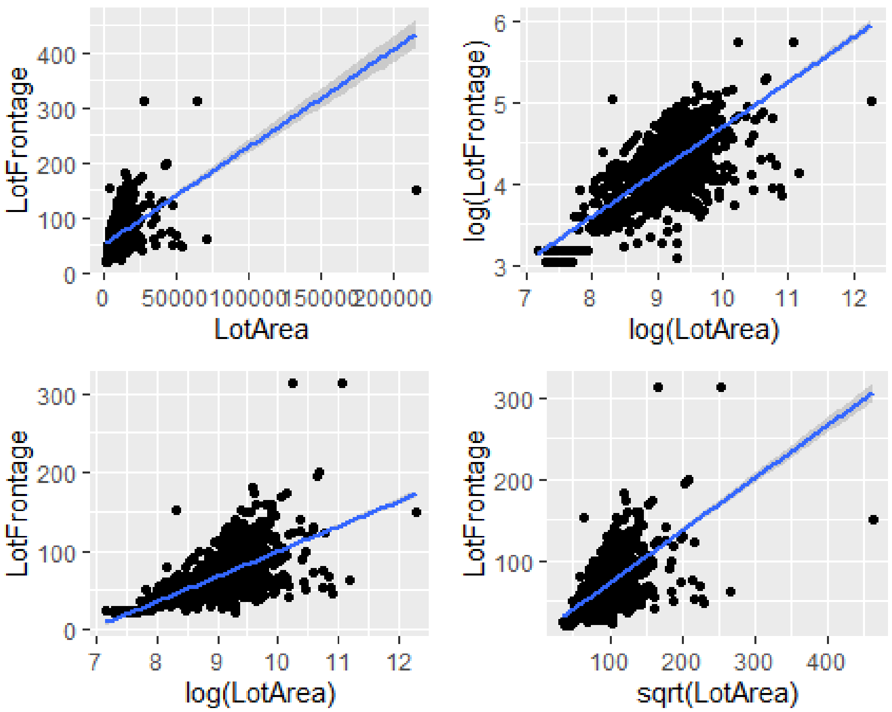
```
grubbs.test(df$LotArea,type=11)
```

```
##
##  Grubbs test for two opposite outliers
##
## data:  df$LotArea
## G = 27.12630, U = 0.76779, p-value < 2.2e-16
## alternative hypothesis: 1300 and 215245 are outliers
```

```
grubbs.test(df$LotFrontage,type=11)
```

```
##
##  Grubbs test for two opposite outliers
##
## data:  df$LotFrontage
## G = 12.50808, U = 0.95342, p-value < 2.2e-16
## alternative hypothesis: 21 and 313 are outliers
```

```
p1<-ggplot(df  , aes(LotArea, LotFrontage)) + geom_point() + geom_smooth(method = "lm", se = T)
p2<-ggplot(df, aes(log(LotArea), LotFrontage)) + geom_point() + geom_smooth(method = "lm", se = T)
p3<-ggplot(df, aes(log(LotArea), log(LotFrontage))) + geom_point() + geom_smooth(method = "lm", se = T)
p4<-ggplot(df, aes(sqrt(LotArea), LotFrontage)) + geom_point() + geom_smooth(method = "lm", se = T)
multiplot(p1, p2, p3, p4, cols=2)
```



```
cor(as.numeric(df$LotArea),as.numeric(df$LotFrontage),use="complete.obs")

## [1] 0.4898956

cor(log(as.numeric(df$LotArea)),log(as.numeric(df$LotFrontage)),use="complete
.obs")

## [1] 0.7662858

cor(log(as.numeric(df$LotArea)),as.numeric(df$LotFrontage),use="complete.obs"
)

## [1] 0.6835123

cor(sqrt(as.numeric(df$LotArea)),as.numeric(df$LotFrontage),use="complete.obs
")

## [1] 0.647658

str(df)

## '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 7
420 ...
##  $ Street       : Factor w/ 2 levels "Grvl","Pave": 2 2 2 2 2 2 2 2 2 2 ..
```

```
.
##  $ Alley        : Factor w/ 3 levels "Grvl","NoAccess",..: 2 2 2 2 2 2 2 2
2 2 ...
##  $ 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 1
3 13 13 7 4 9 ...
##  $ Exterior2nd  : Factor w/ 16 levels "AsbShng","AsphShn",..: 14 9 14 16 1
4 14 14 7 16 9 ...
##  $ MasVnrType   : Factor w/ 4 levels "BrkCmn","BrkFace",..: 2 3 2 3 2 3 4
4 3 3 ...
##  $ MasVnrArea   : num  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/ 5 levels "Ex","Fa","Gd",..: 3 3 3 5 3 3 1 3 5
5 ...
##  $ BsmtCond     : Factor w/ 5 levels "Fa","Gd","NoBasement",..: 5 5 5 2 5
5 5 5 5 5 ...
```

```
##  $ BsmtExposure : Factor w/ 5 levels "Av","Gd","Mn",..: 4 2 3 4 1 4 1 3 4
4 ...
##  $ BsmtFinType1 : Factor w/ 7 levels "ALQ","BLQ","GLQ",..: 3 1 3 1 3 3 3 1
7 3 ...
##  $ BsmtFinSF1   : num  706 978 486 216 655 ...
##  $ BsmtFinType2 : Factor w/ 7 levels "ALQ","BLQ","GLQ",..: 7 7 7 7 7 7 7 2
7 7 ...
##  $ BsmtFinSF2   : num  0 0 0 0 0 0 0 32 0 0 ...
##  $ BsmtUnfSF    : num  150 284 434 540 490 64 317 216 952 140 ...
##  $ TotalBsmtSF  : num  856 1262 920 756 1145 ...
##  $ 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 : num  1 0 1 1 1 1 1 1 0 1 ...
##  $ BsmtHalfBath : num  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/ 6 levels "Ex","Fa","Gd",..: 4 6 6 3 6 4 3 6 6
6 ...
##  $ GarageType   : Factor w/ 7 levels "2Types","Attchd",..: 2 2 2 6 2 2 2 2
6 2 ...
##  $ GarageYrBlt  : Factor w/ 104 levels "1895","1896",..: 95 68 93 90 92 85
96 65 24 32 ...
##  $ GarageFinish : Factor w/ 4 levels "Fin","NoGarage",..: 3 3 3 4 3 4 3 3
4 3 ...
##  $ GarageCars   : num  2 2 2 3 3 2 2 2 2 1 ...
##  $ GarageArea   : num  548 460 608 642 836 480 636 484 468 205 ...
##  $ GarageQual   : Factor w/ 6 levels "Ex","Fa","Gd",..: 6 6 6 6 6 6 6 6 2
3 ...
##  $ GarageCond   : Factor w/ 6 levels "Ex","Fa","Gd",..: 6 6 6 6 6 6 6 6 6
6 ...
##  $ 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/ 4 levels "Ex","Fa","Gd",..: 4 4 4 4 4 4 4 4 4
4 ...
##  $ Fence        : Factor w/ 5 levels "GdPrv","GdWo",..: 5 5 5 5 5 3 5 5 5
5 ...
##  $ MiscFeature  : Factor w/ 5 levels "Gar2","None",..: 2 2 2 2 2 4 2 4 2 2
...
##  $ 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    : num  208500 181500 223500 140000 250000 ...
```

```
#splitting back to Test and Train
Traindata<-df[1:1460,]
Testdata<-df[(1461):nrow(df),]
#Testdata<- testdata[ , -which(names(Testdata) %in% c("SalePrice"))]

# We have cleaned all of the data
```

PCA Preperation

```
response <- Traindata$SalePrice
train_dummy <- dummy.data.frame(Traindata, sep = ".", all = TRUE)

names(train_dummy)

##   [1] "Id"                          "MSSubClass"
##   [3] "MSZoning.C (all)"            "MSZoning.FV"
##   [5] "MSZoning.RH"                 "MSZoning.RL"
##   [7] "MSZoning.RM"                 "LotFrontage"
##   [9] "LotArea"                     "Street.Grvl"
##  [11] "Street.Pave"                 "Alley.Grvl"
##  [13] "Alley.NoAccess"              "Alley.Pave"
##  [15] "LotShape.IR1"                "LotShape.IR2"
##  [17] "LotShape.IR3"                "LotShape.Reg"
##  [19] "LandContour.Bnk"             "LandContour.HLS"
##  [21] "LandContour.Low"             "LandContour.Lvl"
##  [23] "Utilities.AllPub"            "Utilities.NoSeWa"
##  [25] "LotConfig.Corner"            "LotConfig.CulDSac"
##  [27] "LotConfig.FR2"           "LotConfig.FR3"
##  [29] "LotConfig.Inside"        "LandSlope.Gtl"
##  [31] "LandSlope.Mod"           "LandSlope.Sev"
##  [33] "Neighborhood.Blmngtn"    "Neighborhood.Blueste"
##  [35] "Neighborhood.BrDale"     "Neighborhood.BrkSide"
##  [37] "Neighborhood.ClearCr"    "Neighborhood.CollgCr"
##  [39] "Neighborhood.Crawfor"    "Neighborhood.Edwards"
##  [41] "Neighborhood.Gilbert"    "Neighborhood.IDOTRR"
##  [43] "Neighborhood.MeadowV"    "Neighborhood.Mitchel"
##  [45] "Neighborhood.NAmes"      "Neighborhood.NoRidge"
##  [47] "Neighborhood.NPkVill"    "Neighborhood.NridgHt"
##  [49] "Neighborhood.NWAmes"     "Neighborhood.OldTown"
##  [51] "Neighborhood.Sawyer"     "Neighborhood.SawyerW"
##  [53] "Neighborhood.Somerst"    "Neighborhood.StoneBr"
##  [55] "Neighborhood.SWISU"      "Neighborhood.Timber"
##  [57] "Neighborhood.Veenker"    "Condition1.Artery"
##  [59] "Condition1.Feedr"        "Condition1.Norm"
##  [61] "Condition1.PosA"         "Condition1.PosN"
##  [63] "Condition1.RRAe"         "Condition1.RRAn"
##  [65] "Condition1.RRNe"         "Condition1.RRNn"
##  [67] "Condition2.Artery"       "Condition2.Feedr"
##  [69] "Condition2.Norm"         "Condition2.PosA"
##  [71] "Condition2.PosN"         "Condition2.RRAe"
##  [73] "Condition2.RRAn"         "Condition2.RRNn"
##  [75] "BldgType.1Fam"           "BldgType.2fmCon"
##  [77] "BldgType.Duplex"         "BldgType.Twnhs"
##  [79] "BldgType.TwnhsE"         "HouseStyle.1.5Fin"
##  [81] "HouseStyle.1.5Unf"       "HouseStyle.1Story"
##  [83] "HouseStyle.2.5Fin"       "HouseStyle.2.5Unf"
##  [85] "HouseStyle.2Story"       "HouseStyle.SFoyer"
##  [87] "HouseStyle.SLvl"         "OverallQual"
##  [89] "OverallCond"             "YearBuilt"
##  [91] "YearRemodAdd"            "RoofStyle.Flat"
##  [93] "RoofStyle.Gable"         "RoofStyle.Gambrel"
##  [95] "RoofStyle.Hip"           "RoofStyle.Mansard"
##  [97] "RoofStyle.Shed"          "RoofMatl.ClyTile"
##  [99] "RoofMatl.CompShg"        "RoofMatl.Membran"
## [101] "RoofMatl.Metal"          "RoofMatl.Roll"
## [103] "RoofMatl.Tar&Grv"        "RoofMatl.WdShake"
## [105] "RoofMatl.WdShngl"        "Exterior1st.AsbShng"
## [107] "Exterior1st.AsphShn"     "Exterior1st.BrkComm"
## [109] "Exterior1st.BrkFace"     "Exterior1st.CBlock"
## [111] "Exterior1st.CemntBd"     "Exterior1st.HdBoard"
## [113] "Exterior1st.ImStucc"     "Exterior1st.MetalSd"
## [115] "Exterior1st.Plywood"     "Exterior1st.Stone"
## [117] "Exterior1st.Stucco"      "Exterior1st.VinylSd"
## [119] "Exterior1st.Wd Sdng"     "Exterior1st.WdShing"
## [121] "Exterior2nd.AsbShng"     "Exterior2nd.AsphShn"
## [123] "Exterior2nd.Brk Cmn"     "Exterior2nd.BrkFace"
## [125] "Exterior2nd.CBlock"      "Exterior2nd.CmentBd"
## [127] "Exterior2nd.HdBoard"     "Exterior2nd.ImStucc"
## [129] "Exterior2nd.MetalSd"     "Exterior2nd.Other"
## [131] "Exterior2nd.Plywood"     "Exterior2nd.Stone"
## [133] "Exterior2nd.Stucco"      "Exterior2nd.VinylSd"
## [135] "Exterior2nd.Wd Sdng"     "Exterior2nd.Wd Shng"
## [137] "MasVnrType.BrkCmn"       "MasVnrType.BrkFace"
## [139] "MasVnrType.None"         "MasVnrType.Stone"
## [141] "MasVnrArea"              "ExterQual.Ex"
## [143] "ExterQual.Fa"            "ExterQual.Gd"
## [145] "ExterQual.TA"            "ExterCond.Ex"
## [147] "ExterCond.Fa"            "ExterCond.Gd"
## [149] "ExterCond.Po"            "ExterCond.TA"
## [151] "Foundation.BrkTil"       "Foundation.CBlock"
## [153] "Foundation.PConc"        "Foundation.Slab"
## [155] "Foundation.Stone"        "Foundation.Wood"
## [157] "BsmtQual.Ex"             "BsmtQual.Fa"
## [159] "BsmtQual.Gd"             "BsmtQual.NoBasement"
## [161] "BsmtQual.TA"             "BsmtCond.Fa"
## [163] "BsmtCond.Gd"             "BsmtCond.NoBasement"
## [165] "BsmtCond.Po"             "BsmtCond.TA"
## [167] "BsmtExposure.Av"         "BsmtExposure.Gd"
## [169] "BsmtExposure.Mn"         "BsmtExposure.No"
## [171] "BsmtExposure.NoBasement" "BsmtFinType1.ALQ"
## [173] "BsmtFinType1.BLQ"        "BsmtFinType1.GLQ"
## [175] "BsmtFinType1.LwQ"        "BsmtFinType1.NoBasement"
## [177] "BsmtFinType1.Rec"        "BsmtFinType1.Unf"
## [179] "BsmtFinSF1"              "BsmtFinType2.ALQ"
## [181] "BsmtFinType2.BLQ"        "BsmtFinType2.GLQ"
## [183] "BsmtFinType2.LwQ"        "BsmtFinType2.NoBasement"
## [185] "BsmtFinType2.Rec"        "BsmtFinType2.Unf"
## [187] "BsmtFinSF2"              "BsmtUnfSF"
## [189] "TotalBsmtSF"             "Heating.Floor"
## [191] "Heating.GasA"            "Heating.GasW"
## [193] "Heating.Grav"            "Heating.OthW"
## [195] "Heating.Wall"            "HeatingQC.Ex"
## [197] "HeatingQC.Fa"            "HeatingQC.Gd"
## [199] "HeatingQC.Po"            "HeatingQC.TA"
## [201] "CentralAir.N"            "CentralAir.Y"
## [203] "Electrical.FuseA"        "Electrical.FuseF"
## [205] "Electrical.FuseP"        "Electrical.Mix"
## [207] "Electrical.SBrkr"        "X1stFlrSF"
## [209] "X2ndFlrSF"               "LowQualFinSF"
## [211] "GrLivArea"               "BsmtFullBath"
## [213] "BsmtHalfBath"            "FullBath"
## [215] "HalfBath"                "BedroomAbvGr"
## [217] "KitchenAbvGr"            "KitchenQual.Ex"
## [219] "KitchenQual.Fa"          "KitchenQual.Gd"
## [221] "KitchenQual.TA"          "TotRmsAbvGrd"
## [223] "Functional.Maj1"         "Functional.Maj2"
## [225] "Functional.Min1"         "Functional.Min2"
## [227] "Functional.Mod"          "Functional.Sev"
## [229] "Functional.Typ"          "Fireplaces"
## [231] "FireplaceQu.Ex"          "FireplaceQu.Fa"
## [233] "FireplaceQu.Gd"          "FireplaceQu.NoFirePlace"
## [235] "FireplaceQu.Po"          "FireplaceQu.TA"
## [237] "GarageType.2Types"       "GarageType.Attchd"
## [239] "GarageType.Basment"      "GarageType.BuiltIn"
## [241] "GarageType.CarPort"      "GarageType.Detchd"
## [243] "GarageType.NoGarage"     "GarageYrBlt.1900"
## [245] "GarageYrBlt.1906"        "GarageYrBlt.1908"
## [247] "GarageYrBlt.1910"        "GarageYrBlt.1914"
## [249] "GarageYrBlt.1915"        "GarageYrBlt.1916"
## [251] "GarageYrBlt.1918"        "GarageYrBlt.1920"
## [253] "GarageYrBlt.1921"        "GarageYrBlt.1922"
## [255] "GarageYrBlt.1923"        "GarageYrBlt.1924"
## [257] "GarageYrBlt.1925"        "GarageYrBlt.1926"
## [259] "GarageYrBlt.1927"        "GarageYrBlt.1928"
## [261] "GarageYrBlt.1929"        "GarageYrBlt.1930"
## [263] "GarageYrBlt.1931"        "GarageYrBlt.1932"
## [265] "GarageYrBlt.1933"        "GarageYrBlt.1934"
## [267] "GarageYrBlt.1935"        "GarageYrBlt.1936"
## [269] "GarageYrBlt.1937"        "GarageYrBlt.1938"
## [271] "GarageYrBlt.1939"        "GarageYrBlt.1940"
## [273] "GarageYrBlt.1941"        "GarageYrBlt.1942"
## [275] "GarageYrBlt.1945"        "GarageYrBlt.1946"
## [277] "GarageYrBlt.1947"        "GarageYrBlt.1948"
## [279] "GarageYrBlt.1949"        "GarageYrBlt.1950"
## [281] "GarageYrBlt.1951"        "GarageYrBlt.1952"
## [283] "GarageYrBlt.1953"        "GarageYrBlt.1954"
## [285] "GarageYrBlt.1955"        "GarageYrBlt.1956"
## [287] "GarageYrBlt.1957"        "GarageYrBlt.1958"
## [289] "GarageYrBlt.1959"        "GarageYrBlt.1960"
## [291] "GarageYrBlt.1961"        "GarageYrBlt.1962"
## [293] "GarageYrBlt.1963"        "GarageYrBlt.1964"
## [295] "GarageYrBlt.1965"        "GarageYrBlt.1966"
## [297] "GarageYrBlt.1967"        "GarageYrBlt.1968"
## [299] "GarageYrBlt.1969"        "GarageYrBlt.1970"
## [301] "GarageYrBlt.1971"        "GarageYrBlt.1972"
## [303] "GarageYrBlt.1973"        "GarageYrBlt.1974"
## [305] "GarageYrBlt.1975"        "GarageYrBlt.1976"
## [307] "GarageYrBlt.1977"        "GarageYrBlt.1978"
## [309] "GarageYrBlt.1979"        "GarageYrBlt.1980"
## [311] "GarageYrBlt.1981"        "GarageYrBlt.1982"
## [313] "GarageYrBlt.1983"        "GarageYrBlt.1984"
## [315] "GarageYrBlt.1985"        "GarageYrBlt.1986"
## [317] "GarageYrBlt.1987"        "GarageYrBlt.1988"
## [319] "GarageYrBlt.1989"        "GarageYrBlt.1990"
## [321] "GarageYrBlt.1991"        "GarageYrBlt.1992"
## [323] "GarageYrBlt.1993"        "GarageYrBlt.1994"
## [325] "GarageYrBlt.1995"        "GarageYrBlt.1996"
## [327] "GarageYrBlt.1997"        "GarageYrBlt.1998"
## [329] "GarageYrBlt.1999"        "GarageYrBlt.2000"
## [331] "GarageYrBlt.2001"        "GarageYrBlt.2002"
## [333] "GarageYrBlt.2003"        "GarageYrBlt.2004"
## [335] "GarageYrBlt.2005"        "GarageYrBlt.2006"
## [337] "GarageYrBlt.2007"        "GarageYrBlt.2008"
## [339] "GarageYrBlt.2009"        "GarageYrBlt.2010"
## [341] "GarageYrBlt.NoGarage"    "GarageFinish.Fin"
## [343] "GarageFinish.NoGarage"   "GarageFinish.RFn"
## [345] "GarageFinish.Unf"        "GarageCars"
## [347] "GarageArea"              "GarageQual.Ex"
## [349] "GarageQual.Fa"           "GarageQual.Gd"
## [351] "GarageQual.NoGarage"     "GarageQual.Po"
## [353] "GarageQual.TA"           "GarageCond.Ex"
## [355] "GarageCond.Fa"           "GarageCond.Gd"
## [357] "GarageCond.NoGarage"     "GarageCond.Po"
## [359] "GarageCond.TA"           "PavedDrive.N"
## [361] "PavedDrive.P"            "PavedDrive.Y"
## [363] "WoodDeckSF"              "OpenPorchSF"
## [365] "EnclosedPorch"           "X3SsnPorch"
## [367] "ScreenPorch"             "PoolArea"
## [369] "PoolQC.Ex"               "PoolQC.Fa"
## [371] "PoolQC.Gd"               "PoolQC.NoPool"
## [373] "Fence.GdPrv"             "Fence.GdWo"
## [375] "Fence.MnPrv"             "Fence.MnWw"
## [377] "Fence.NoFence"           "MiscFeature.Gar2"
## [379] "MiscFeature.None"        "MiscFeature.Othr"
## [381] "MiscFeature.Shed"        "MiscFeature.TenC"
## [383] "MiscVal"                 "MoSold"
## [385] "YrSold"                  "SaleType.COD"
## [387] "SaleType.Con"            "SaleType.ConLD"
## [389] "SaleType.ConLI"          "SaleType.ConLw"
## [391] "SaleType.CWD"            "SaleType.New"
## [393] "SaleType.Oth"            "SaleType.WD"
## [395] "SaleCondition.Abnorml"   "SaleCondition.AdjLand"
## [397] "SaleCondition.Alloca"    "SaleCondition.Family"
## [399] "SaleCondition.Normal"    "SaleCondition.Partial"
## [401] "SalePrice"
```

```r
split <- createDataPartition(y=response, p=.5, list=F)
training <- train_dummy[split,]
testing <- train_dummy[-split,]
str(training)
```

```
## 'data.frame':    731 obs. of  401 variables:
##  $ Id                     : int  2 3 4 5 6 8 10 11 12 15 ...
##  $ MSSubClass             : int  20 60 70 60 50 60 190 20 60 20 ...
##  $ MSZoning.C (all)       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ MSZoning.FV            : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ MSZoning.RH            : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ MSZoning.RL            : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ MSZoning.RM            : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ LotFrontage            : int  80 68 60 84 85 NA 50 70 85 NA ...
##  $ LotArea                : int  9600 11250 9550 14260 14115 10382 7420 11
200 11924 10920 ...
##  $ Street.Grvl            : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Street.Pave            : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ Alley.Grvl             : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Alley.NoAccess         : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ Alley.Pave             : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ LotShape.IR1           : int  0 1 1 1 1 1 0 0 1 1 ...
##  $ LotShape.IR2           : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ LotShape.IR3           : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ LotShape.Reg           : int  1 0 0 0 0 0 1 1 0 0 ...
##  $ LandContour.Bnk        : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ LandContour.HLS        : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ LandContour.Low        : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ LandContour.Lvl        : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ Utilities.AllPub       : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ Utilities.NoSeWa       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ LotConfig.Corner       : int  0 0 1 0 0 1 1 0 0 1 ...
##  $ LotConfig.CulDSac      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ LotConfig.FR2          : int  1 0 0 1 0 0 0 0 0 0 ...
##  $ LotConfig.FR3          : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ LotConfig.Inside       : int  0 1 0 0 1 0 0 1 1 0 ...
##  $ LandSlope.Gtl          : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ LandSlope.Mod          : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ LandSlope.Sev          : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Neighborhood.Blmngtn   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Neighborhood.Blueste   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Neighborhood.BrDale    : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Neighborhood.BrkSide   : int  0 0 0 0 0 0 1 0 0 0 ...
##  $ Neighborhood.ClearCr   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Neighborhood.CollgCr   : int  0 1 0 0 0 0 0 0 0 0 ...
##  $ Neighborhood.Crawfor   : int  0 0 1 0 0 0 0 0 0 0 ...
##  $ Neighborhood.Edwards   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Neighborhood.Gilbert   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Neighborhood.IDOTRR    : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Neighborhood.MeadowV   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Neighborhood.Mitchel   : int  0 0 0 0 1 0 0 0 0 0 ...
##  $ Neighborhood.NAmes     : int  0 0 0 0 0 0 0 0 0 1 ...
##  $ Neighborhood.NoRidge   : int  0 0 0 1 0 0 0 0 0 0 ...
##  $ Neighborhood.NPkVill   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Neighborhood.NridgHt   : int  0 0 0 0 0 0 0 0 1 0 ...
##  $ Neighborhood.NWAmes    : int  0 0 0 0 0 1 0 0 0 0 ...
##  $ Neighborhood.OldTown   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Neighborhood.Sawyer    : int  0 0 0 0 0 0 0 1 0 0 ...
##  $ Neighborhood.SawyerW   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Neighborhood.Somerst   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Neighborhood.StoneBr   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Neighborhood.SWISU     : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Neighborhood.Timber    : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Neighborhood.Veenker   : int  1 0 0 0 0 0 0 0 0 0 ...
##  $ Condition1.Artery      : int  0 0 0 0 0 0 1 0 0 0 ...
##  $ Condition1.Feedr       : int  1 0 0 0 0 0 0 0 0 0 ...
##  $ Condition1.Norm        : int  0 1 1 1 1 0 0 1 1 1 ...
##  $ Condition1.PosA        : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Condition1.PosN        : int  0 0 0 0 0 1 0 0 0 0 ...
##  $ Condition1.RRAe        : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Condition1.RRAn        : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Condition1.RRNe        : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Condition1.RRNn        : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Condition2.Artery      : int  0 0 0 0 0 0 1 0 0 0 ...
##  $ Condition2.Feedr       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Condition2.Norm        : int  1 1 1 1 1 1 0 1 1 1 ...
##  $ Condition2.PosA        : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Condition2.PosN        : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Condition2.RRAe        : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Condition2.RRAn        : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Condition2.RRNn        : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ BldgType.1Fam          : int  1 1 1 1 1 1 0 1 1 1 ...
##  $ BldgType.2fmCon        : int  0 0 0 0 0 0 1 0 0 0 ...
##  $ BldgType.Duplex        : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ BldgType.Twnhs         : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ BldgType.TwnhsE        : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ HouseStyle.1.5Fin      : int  0 0 0 0 1 0 0 0 0 0 ...
##  $ HouseStyle.1.5Unf      : int  0 0 0 0 0 0 1 0 0 0 ...
##  $ HouseStyle.1Story      : int  1 0 0 0 0 0 0 1 0 1 ...
##  $ HouseStyle.2.5Fin      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ HouseStyle.2.5Unf      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ HouseStyle.2Story      : int  0 1 1 1 0 1 0 0 1 0 ...
##  $ HouseStyle.SFoyer      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ HouseStyle.SLvl        : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ OverallQual            : int  6 7 7 8 5 7 5 5 9 6 ...
##  $ OverallCond            : int  8 5 5 5 5 6 6 5 5 5 ...
##  $ YearBuilt              : int  1976 2001 1915 2000 1993 1973 1939 1965 2
005 1960 ...
##  $ YearRemodAdd           : int  1976 2002 1970 2000 1995 1973 1950 1965 2
006 1960 ...
##  $ RoofStyle.Flat         : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ RoofStyle.Gable        : int  1 1 1 1 1 1 1 0 0 0 ...
##  $ RoofStyle.Gambrel      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ RoofStyle.Hip          : int  0 0 0 0 0 0 0 1 1 1 ...
##  $ RoofStyle.Mansard      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ RoofStyle.Shed         : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ RoofMatl.ClyTile       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ RoofMatl.CompShg       : int  1 1 1 1 1 1 1 1 1 1 ...
##   [list output truncated]
##  - attr(*, "dummies")=List of 44
##   ..$ MSZoning     : int  3 4 5 6 7
##   ..$ Street       : int  10 11
##   ..$ Alley        : int  12 13 14
##   ..$ LotShape     : int  15 16 17 18
##   ..$ LandContour  : int  19 20 21 22
##   ..$ Utilities    : int  23 24
##   ..$ LotConfig    : int  25 26 27 28 29
##   ..$ LandSlope    : int  30 31 32
##   ..$ Neighborhood : int  33 34 35 36 37 38 39 40 41 42 ...
##   ..$ Condition1   : int  58 59 60 61 62 63 64 65 66
##   ..$ Condition2   : int  67 68 69 70 71 72 73 74
##   ..$ BldgType     : int  75 76 77 78 79
##   ..$ HouseStyle   : int  80 81 82 83 84 85 86 87
##   ..$ RoofStyle    : int  92 93 94 95 96 97
##   ..$ RoofMatl     : int  98 99 100 101 102 103 104 105
##   ..$ Exterior1st  : int  106 107 108 109 110 111 112 113 114 115 ...
##   ..$ Exterior2nd  : int  121 122 123 124 125 126 127 128 129 130 ...
##   ..$ MasVnrType   : int  137 138 139 140
##   ..$ ExterQual    : int  142 143 144 145
##   ..$ ExterCond    : int  146 147 148 149 150
##   ..$ Foundation   : int  151 152 153 154 155 156
##   ..$ BsmtQual     : int  157 158 159 160 161
##   ..$ BsmtCond     : int  162 163 164 165 166
##   ..$ BsmtExposure : int  167 168 169 170 171
##   ..$ BsmtFinType1 : int  172 173 174 175 176 177 178
##   ..$ BsmtFinType2 : int  180 181 182 183 184 185 186
##   ..$ Heating      : int  190 191 192 193 194 195
##   ..$ HeatingQC    : int  196 197 198 199 200
##   ..$ CentralAir   : int  201 202
##   ..$ Electrical   : int  203 204 205 206 207
##   ..$ KitchenQual  : int  218 219 220 221
##   ..$ Functional   : int  223 224 225 226 227 228 229
##   ..$ FireplaceQu  : int  231 232 233 234 235 236
##   ..$ GarageType   : int  237 238 239 240 241 242 243
##   ..$ GarageYrBlt  : int  244 245 246 247 248 249 250 251 252 253 ...
##   ..$ GarageFinish : int  342 343 344 345
##   ..$ GarageQual   : int  348 349 350 351 352 353
##   ..$ GarageCond   : int  354 355 356 357 358 359
##   ..$ PavedDrive   : int  360 361 362
##   ..$ PoolQC       : int  369 370 371 372
##   ..$ Fence        : int  373 374 375 376 377
##   ..$ MiscFeature  : int  378 379 380 381 382
##   ..$ SaleType     : int  386 387 388 389 390 391 392 393 394
##   ..$ SaleCondition: int  395 396 397 398 399 400
```

```
# First, we will build a simple linear regression to get a feel for the varia
bles and relationship.
model.lm <- lm(SalePrice ~ ., data = training)
summary(model.lm)
```

```
##
## Call:
## lm(formula = SalePrice ~ ., data = training)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -93630  -8351      0   8217  92935
##
## Coefficients: (84 not defined because of singularities)
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)              2.478e+06  2.110e+06   1.174 0.241295
## Id                       2.507e+00  2.998e+00   0.836 0.403822
## MSSubClass              -6.869e+01  2.308e+02  -0.298 0.766220
## `MSZoning.C (all)`      -4.475e+04  1.842e+04  -2.429 0.015747 *
## MSZoning.FV              2.332e+04  1.438e+04   1.622 0.105968
## MSZoning.RH              3.684e+03  1.555e+04   0.237 0.812952
## MSZoning.RL              6.273e+03  8.044e+03   0.780 0.436145
## MSZoning.RM                     NA         NA      NA       NA
## LotFrontage              5.887e+01  8.924e+01   0.660 0.510013
## LotArea                  9.099e-01  4.374e-01   2.080 0.038401 *
## Street.Grvl             -2.299e+04  5.381e+04  -0.427 0.669471
## Street.Pave                     NA         NA      NA       NA
## Alley.Grvl               9.374e+03  1.180e+04   0.795 0.427553
## Alley.NoAccess           4.033e+03  8.238e+03   0.490 0.624860
## Alley.Pave                      NA         NA      NA       NA
## LotShape.IR1            -5.164e+03  3.513e+03  -1.470 0.142684
## LotShape.IR2             5.207e+03  9.422e+03   0.553 0.580932
## LotShape.IR3             2.071e+04  1.794e+04   1.154 0.249311
## LotShape.Reg                    NA         NA      NA       NA
## LandContour.Bnk         -6.775e+03  9.589e+03  -0.707 0.480420
## LandContour.HLS          8.540e+03  8.121e+03   1.052 0.293878
## LandContour.Low          3.117e+01  1.422e+04   0.002 0.998253
## LandContour.Lvl                 NA         NA      NA       NA
## Utilities.AllPub                NA         NA      NA       NA
## Utilities.NoSeWa                NA         NA      NA       NA
## LotConfig.Corner         3.223e+03  3.850e+03   0.837 0.403288
## LotConfig.CulDSac        1.528e+04  7.298e+03   2.094 0.037117 *
## LotConfig.FR2           -6.217e+03  6.520e+03  -0.953 0.341155
## LotConfig.FR3           -1.364e+04  2.171e+04  -0.628 0.530472
## LotConfig.Inside                NA         NA      NA       NA
## LandSlope.Gtl            1.641e+04  3.370e+04   0.487 0.626602
## LandSlope.Mod            1.406e+04  3.461e+04   0.406 0.684921
## LandSlope.Sev                   NA         NA      NA       NA
## Neighborhood.Blmngtn     4.119e+04  2.335e+04   1.764 0.078835 .
## Neighborhood.Blueste    -2.030e+04  4.033e+04  -0.503 0.615097
## Neighborhood.BrDale      1.498e+04  2.994e+04   0.500 0.617225
## Neighborhood.BrkSide     2.680e+04  2.276e+04   1.178 0.239844
## Neighborhood.ClearCr     9.886e+03  2.631e+04   0.376 0.707365
## Neighborhood.CollgCr     1.719e+04  2.064e+04   0.833 0.405749
## Neighborhood.Crawfor     3.945e+04  2.299e+04   1.716 0.087216 .
## Neighborhood.Edwards     1.686e+04  2.121e+04   0.795 0.427388   
## Neighborhood.Gilbert     2.144e+04  2.187e+04   0.980 0.327824   
## Neighborhood.IDOTRR      4.038e+04  2.535e+04   1.593 0.112283   
## Neighborhood.MeadowV     2.904e+04  3.043e+04   0.954 0.340818   
## Neighborhood.Mitchel     1.137e+04  2.128e+04   0.534 0.593673   
## Neighborhood.NAmes       2.380e+04  2.048e+04   1.162 0.246185   
## Neighborhood.NoRidge     3.856e+04  2.181e+04   1.768 0.078109 . 
## Neighborhood.NPkVill     4.104e+04  3.432e+04   1.196 0.232795   
## Neighborhood.NridgHt     4.243e+04  2.137e+04   1.986 0.048020 * 
## Neighborhood.NWAmes      5.070e+03  2.017e+04   0.251 0.801702   
## Neighborhood.OldTown     1.641e+04  2.258e+04   0.727 0.467966   
## Neighborhood.Sawyer      3.072e+04  2.075e+04   1.481 0.139835   
## Neighborhood.SawyerW     2.950e+04  2.214e+04   1.332 0.183881   
## Neighborhood.Somerst     4.749e+03  2.280e+04   0.208 0.835146   
## Neighborhood.StoneBr     5.202e+04  2.346e+04   2.218 0.027356 * 
## Neighborhood.SWISU       1.934e+04  2.447e+04   0.790 0.429943   
## Neighborhood.Timber      1.386e+04  2.232e+04   0.621 0.535193   
## Neighborhood.Veenker            NA         NA      NA       NA   
## Condition1.Artery       -6.494e+03  2.178e+04  -0.298 0.765772   
## Condition1.Feedr         2.664e+03  2.070e+04   0.129 0.897689   
## Condition1.Norm          5.570e+03  1.979e+04   0.281 0.778604   
## Condition1.PosA          2.766e+04  3.596e+04   0.769 0.442396   
## Condition1.PosN         -9.461e+03  2.734e+04  -0.346 0.729595   
## Condition1.RRAe         -2.413e+04  2.638e+04  -0.915 0.361064   
## Condition1.RRAn         -6.626e+03  2.090e+04  -0.317 0.751472   
## Condition1.RRNe         -9.982e+03  3.252e+04  -0.307 0.759104   
## Condition1.RRNn                 NA         NA      NA       NA   
## Condition2.Artery       -3.956e+04  4.231e+04  -0.935 0.350555   
## Condition2.Feedr        -3.020e+04  3.617e+04  -0.835 0.404400   
## Condition2.Norm         -2.413e+04  2.390e+04  -1.010 0.313481   
## Condition2.PosA                 NA         NA      NA       NA   
## Condition2.PosN                 NA         NA      NA       NA   
## Condition2.RRAe                 NA         NA      NA       NA   
## Condition2.RRAn                 NA         NA      NA       NA   
## Condition2.RRNn                 NA         NA      NA       NA   
## BldgType.1Fam            1.111e+04  2.399e+04   0.463 0.643537   
## BldgType.2fmCon          7.326e+03  1.879e+04   0.390 0.696849   
## BldgType.Duplex         -2.741e+04  2.201e+04  -1.245 0.214052   
## BldgType.Twnhs          -6.773e+03  9.424e+03  -0.719 0.472938   
## BldgType.TwnhsE                 NA         NA      NA       NA   
## HouseStyle.1.5Fin       -1.601e+04  1.352e+04  -1.184 0.237439   
## HouseStyle.1.5Unf       -1.066e+04  1.769e+04  -0.603 0.547305   
## HouseStyle.1Story       -3.894e+03  1.583e+04  -0.246 0.805832   
## HouseStyle.2.5Fin       -2.375e+04  4.208e+04  -0.564 0.572896   
## HouseStyle.2.5Unf       -1.481e+04  2.888e+04  -0.513 0.608354   
## HouseStyle.2Story       -1.563e+04  1.247e+04  -1.254 0.211057   
## HouseStyle.SFoyer        6.339e+02  1.198e+04   0.053 0.957852   
## HouseStyle.SLvl                 NA         NA      NA       NA   
## OverallQual              6.025e+03  2.164e+03   2.784 0.005733 **
## OverallCond              3.809e+03  2.127e+03   1.791 0.074387 . 
## YearBuilt                3.892e+02  1.628e+02   2.390 0.017494 *
## YearRemodAdd             1.863e+02  1.270e+02   1.467 0.143622
## RoofStyle.Flat          -1.054e+05  5.248e+04  -2.009 0.045543 *
## RoofStyle.Gable          1.214e+04  3.208e+04   0.378 0.705361
## RoofStyle.Gambrel        1.254e+04  3.757e+04   0.334 0.738735
## RoofStyle.Hip            9.128e+03  3.234e+04   0.282 0.777941
## RoofStyle.Mansard               NA         NA      NA       NA
## RoofStyle.Shed                  NA         NA      NA       NA
## RoofMatl.ClyTile        -8.675e+05  6.399e+04 -13.558  < 2e-16 ***
## RoofMatl.CompShg        -6.560e+04  2.143e+04  -3.061 0.002416 **
## RoofMatl.Membran         5.604e+04  6.962e+04   0.805 0.421550
## RoofMatl.Metal                  NA         NA      NA       NA
## RoofMatl.Roll                   NA         NA      NA       NA
## `RoofMatl.Tar&Grv`              NA         NA      NA       NA
## RoofMatl.WdShake        -9.160e+04  5.375e+04  -1.704 0.089447 .
## RoofMatl.WdShngl                NA         NA      NA       NA
## Exterior1st.AsbShng      2.690e+03  3.000e+04   0.090 0.928625
## Exterior1st.AsphShn             NA         NA      NA       NA
## Exterior1st.BrkComm     -7.654e+04  5.531e+04  -1.384 0.167483
## Exterior1st.BrkFace     -1.959e+04  2.035e+04  -0.962 0.336637
## Exterior1st.CBlock              NA         NA      NA       NA
## Exterior1st.CemntBd     -3.583e+04  4.000e+04  -0.896 0.371108
## Exterior1st.HdBoard     -3.107e+04  1.902e+04  -1.634 0.103447
## Exterior1st.ImStucc     -7.057e+04  3.592e+04  -1.964 0.050458 .
## Exterior1st.MetalSd     -4.590e+04  2.327e+04  -1.973 0.049526 *
## Exterior1st.Plywood     -3.730e+04  2.015e+04  -1.851 0.065189 .
## Exterior1st.Stone               NA         NA      NA       NA
## Exterior1st.Stucco      -3.327e+04  2.898e+04  -1.148 0.251904
## Exterior1st.VinylSd     -4.438e+04  1.877e+04  -2.365 0.018721 *
## `Exterior1st.Wd Sdng`   -2.933e+04  1.760e+04  -1.667 0.096724 .
## Exterior1st.WdShing             NA         NA      NA       NA
## Exterior2nd.AsbShng      3.012e+04  2.781e+04   1.083 0.279661
## Exterior2nd.AsphShn      3.214e+04  5.079e+04   0.633 0.527411
## `Exterior2nd.Brk Cmn`    4.389e+04  3.829e+04   1.146 0.252712
## Exterior2nd.BrkFace      2.023e+04  2.089e+04   0.969 0.333500
## Exterior2nd.CBlock              NA         NA      NA       NA
## Exterior2nd.CmentBd      4.404e+04  3.915e+04   1.125 0.261518
## Exterior2nd.HdBoard      4.353e+04  1.814e+04   2.399 0.017072 *
## Exterior2nd.ImStucc      4.695e+04  2.101e+04   2.235 0.026215 *
## Exterior2nd.MetalSd      5.459e+04  2.239e+04   2.438 0.015384 *
## Exterior2nd.Other        2.448e+04  3.225e+04   0.759 0.448463
## Exterior2nd.Plywood      2.956e+04  1.799e+04   1.643 0.101478
## Exterior2nd.Stone        3.331e+04  3.252e+04   1.025 0.306461
## Exterior2nd.Stucco       5.133e+04  2.617e+04   1.962 0.050785 .
## Exterior2nd.VinylSd      4.719e+04  1.721e+04   2.742 0.006491 **
## `Exterior2nd.Wd Sdng`    3.251e+04  1.591e+04   2.043 0.042018 *
## `Exterior2nd.Wd Shng`           NA         NA      NA       NA
## MasVnrType.BrkCmn       -2.678e+04  2.166e+04  -1.236 0.217321
## MasVnrType.BrkFace      -1.251e+04  5.776e+03  -2.166 0.031158 *
## MasVnrType.None         -6.467e+03  6.229e+03  -1.038 0.300045
## MasVnrType.Stone                NA         NA      NA       NA    
## MasVnrArea              2.853e+01  1.003e+01   2.845 0.004766 ** 
## ExterQual.Ex            3.545e+04  1.027e+04   3.451 0.000643 ***
## ExterQual.Fa            2.618e+04  4.548e+04   0.576 0.565308    
## ExterQual.Gd            6.873e+03  5.123e+03   1.342 0.180776    
## ExterQual.TA                   NA         NA      NA       NA    
## ExterCond.Ex           -4.049e+03  2.244e+04  -0.180 0.856975    
## ExterCond.Fa           -5.151e+02  1.373e+04  -0.038 0.970092    
## ExterCond.Gd           -3.033e+03  5.435e+03  -0.558 0.577270    
## ExterCond.Po            2.343e+03  4.159e+04   0.056 0.955101    
## ExterCond.TA                   NA         NA      NA       NA    
## Foundation.BrkTil       1.645e+04  3.383e+04   0.486 0.627073    
## Foundation.CBlock       1.290e+04  3.297e+04   0.391 0.695839    
## Foundation.PConc        1.408e+04  3.280e+04   0.429 0.668030    
## Foundation.Slab        -3.522e+03  3.796e+04  -0.093 0.926155    
## Foundation.Stone        7.045e+04  4.705e+04   1.497 0.135423    
## Foundation.Wood                NA         NA      NA       NA    
## BsmtQual.Ex             3.024e+03  8.476e+03   0.357 0.721494    
## BsmtQual.Fa             2.181e+03  1.101e+04   0.198 0.843048    
## BsmtQual.Gd            -8.053e+03  5.982e+03  -1.346 0.179308    
## BsmtQual.NoBasement     1.900e+04  1.844e+04   1.030 0.303904    
## BsmtQual.TA                    NA         NA      NA       NA    
## BsmtCond.Fa            -4.399e+03  9.734e+03  -0.452 0.651670    
## BsmtCond.Gd            -5.667e+02  6.445e+03  -0.088 0.930005    
## BsmtCond.NoBasement            NA         NA      NA       NA    
## BsmtCond.Po            -4.223e+04  1.035e+05  -0.408 0.683580    
## BsmtCond.TA                    NA         NA      NA       NA    
## BsmtExposure.Av        -5.829e+03  3.946e+03  -1.477 0.140776    
## BsmtExposure.Gd         3.170e+04  5.554e+03   5.708 2.90e-08 ***
## BsmtExposure.Mn         1.396e+03  5.059e+03   0.276 0.782830    
## BsmtExposure.No                NA         NA      NA       NA    
## BsmtExposure.NoBasement        NA         NA      NA       NA    
## BsmtFinType1.ALQ        1.609e+03  5.884e+03   0.273 0.784761    
## BsmtFinType1.BLQ       -1.369e+04  6.630e+03  -2.065 0.039802 *  
## BsmtFinType1.GLQ        2.254e+02  5.104e+03   0.044 0.964808    
## BsmtFinType1.LwQ       -1.403e+04  8.338e+03  -1.683 0.093542 .  
## BsmtFinType1.NoBasement        NA         NA      NA       NA    
## BsmtFinType1.Rec       -8.117e+03  7.518e+03  -1.080 0.281191    
## BsmtFinType1.Unf               NA         NA      NA       NA    
## BsmtFinSF1              4.346e+01  1.186e+01   3.665 0.000295 ***
## BsmtFinType2.ALQ        1.031e+04  1.857e+04   0.555 0.579006    
## BsmtFinType2.BLQ       -9.533e+03  1.189e+04  -0.802 0.423177    
## BsmtFinType2.GLQ        5.433e+04  2.939e+04   1.849 0.065541 .  
## BsmtFinType2.LwQ        8.846e+03  1.090e+04   0.812 0.417531    
## BsmtFinType2.NoBasement        NA         NA      NA       NA    
## BsmtFinType2.Rec        1.437e+03  1.083e+04   0.133 0.894472    
## BsmtFinType2.Unf               NA         NA      NA       NA    
## BsmtFinSF2              3.885e+01  2.215e+01   1.754 0.080491 .  
## BsmtUnfSF               2.296e+01  1.176e+01   1.952 0.051869 .  
## TotalBsmtSF                    NA         NA      NA       NA    
## Heating.Floor                   NA        NA      NA       NA    
## Heating.GasA            -1.924e+04 4.276e+04  -0.450 0.653145    
## Heating.GasW            -2.623e+04 4.567e+04  -0.574 0.566183    
## Heating.Grav            -1.257e+04 5.013e+04  -0.251 0.802265    
## Heating.OthW            -6.134e+04 5.202e+04  -1.179 0.239363    
## Heating.Wall                    NA        NA      NA       NA    
## HeatingQC.Ex             6.988e+03 4.793e+03   1.458 0.145953    
## HeatingQC.Fa            -9.876e+03 1.096e+04  -0.901 0.368209    
## HeatingQC.Gd            -1.329e+02 4.728e+03  -0.028 0.977587    
## HeatingQC.Po                    NA        NA      NA       NA    
## HeatingQC.TA                    NA        NA      NA       NA    
## CentralAir.N            -1.031e+03 8.958e+03  -0.115 0.908438    
## CentralAir.Y                    NA        NA      NA       NA    
## Electrical.FuseA         9.731e+03 6.626e+03   1.469 0.143035    
## Electrical.FuseF        -8.697e+02 1.382e+04  -0.063 0.949858    
## Electrical.FuseP         8.497e+03 3.414e+04   0.249 0.803618    
## Electrical.Mix                  NA        NA      NA       NA    
## Electrical.SBrkr                NA        NA      NA       NA    
## X1stFlrSF                6.091e+01 1.409e+01   4.322 2.15e-05 ***
## X2ndFlrSF                8.911e+01 1.166e+01   7.641 3.38e-13 ***
## LowQualFinSF             6.316e+01 5.118e+01   1.234 0.218167    
## GrLivArea                       NA        NA      NA       NA    
## BsmtFullBath             3.844e+03 3.797e+03   1.012 0.312241    
## BsmtHalfBath             4.994e+03 6.962e+03   0.717 0.473800    
## FullBath                 8.213e+03 4.900e+03   1.676 0.094789 .  
## HalfBath                -4.707e+03 4.699e+03  -1.002 0.317328    
## BedroomAbvGr            -4.168e+03 2.999e+03  -1.390 0.165646    
## KitchenAbvGr             1.174e+04 1.775e+04   0.662 0.508791    
## KitchenQual.Ex           8.518e+03 7.237e+03   1.177 0.240178    
## KitchenQual.Fa           4.056e+03 9.989e+03   0.406 0.684979    
## KitchenQual.Gd          -4.528e+03 4.468e+03  -1.014 0.311664    
## KitchenQual.TA                  NA        NA      NA       NA    
## TotRmsAbvGrd            -2.978e+02 1.936e+03  -0.154 0.877870    
## Functional.Maj1         -3.823e+04 1.824e+04  -2.096 0.037008 *  
## Functional.Maj2         -1.772e+04 4.412e+04  -0.402 0.688277    
## Functional.Min1         -1.335e+03 1.188e+04  -0.112 0.910642    
## Functional.Min2         -4.855e+03 8.756e+03  -0.554 0.579713    
## Functional.Mod          -1.824e+04 2.628e+04  -0.694 0.488230    
## Functional.Sev                  NA        NA      NA       NA    
## Functional.Typ                  NA        NA      NA       NA    
## Fireplaces              -4.666e+03 5.658e+03  -0.825 0.410325    
## FireplaceQu.Ex           1.176e+04 9.698e+03   1.212 0.226336    
## FireplaceQu.Fa           9.003e+03 9.941e+03   0.906 0.365895    
## FireplaceQu.Gd           3.867e+03 4.787e+03   0.808 0.419904    
## FireplaceQu.NoFirePlace  3.886e+03 7.419e+03   0.524 0.600829    
## FireplaceQu.Po           1.374e+04 1.338e+04   1.027 0.305349    
## FireplaceQu.TA                  NA        NA      NA       NA    
## GarageType.2Types        2.954e+04 7.061e+04   0.418 0.675963    
## GarageType.Attchd        4.322e+04 3.135e+04   1.378 0.169181    
## GarageType.Basment       4.272e+04 3.525e+04   1.212 0.226548    
## GarageType.BuiltIn       3.934e+04  3.214e+04   1.224 0.222024    
## GarageType.CarPort       4.048e+04  3.696e+04   1.095 0.274355    
## GarageType.Detchd        4.861e+04  3.146e+04   1.545 0.123470    
## GarageType.NoGarage             NA         NA      NA       NA    
## GarageYrBlt.1900                NA         NA      NA       NA    
## GarageYrBlt.1906        -8.884e+04  4.310e+04  -2.061 0.040175 *  
## GarageYrBlt.1908                NA         NA      NA       NA    
## GarageYrBlt.1910        -2.245e+05  7.192e+04  -3.122 0.001986 ** 
## GarageYrBlt.1914                NA         NA      NA       NA    
## GarageYrBlt.1915                NA         NA      NA       NA    
## GarageYrBlt.1916        -1.983e+05  5.773e+04  -3.435 0.000681 ***
## GarageYrBlt.1918        -3.605e+05  6.240e+04  -5.777 2.01e-08 ***
## GarageYrBlt.1920        -5.074e+04  3.264e+04  -1.555 0.121172    
## GarageYrBlt.1921                NA         NA      NA       NA    
## GarageYrBlt.1922        -3.408e+04  3.879e+04  -0.879 0.380287    
## GarageYrBlt.1923        -5.985e+04  3.609e+04  -1.659 0.098321 .  
## GarageYrBlt.1924        -5.842e+04  3.744e+04  -1.560 0.119806    
## GarageYrBlt.1925        -5.211e+04  3.296e+04  -1.581 0.115005    
## GarageYrBlt.1926        -5.723e+04  3.615e+04  -1.583 0.114455    
## GarageYrBlt.1927        -3.513e+04  4.161e+04  -0.844 0.399248    
## GarageYrBlt.1928        -4.583e+04  3.938e+04  -1.164 0.245565    
## GarageYrBlt.1929        -4.506e+04  3.981e+04  -1.132 0.258557    
## GarageYrBlt.1930        -7.831e+04  3.541e+04  -2.212 0.027789 *  
## GarageYrBlt.1931        -5.634e+04  4.053e+04  -1.390 0.165582    
## GarageYrBlt.1932         4.553e+04  4.854e+04   0.938 0.349021    
## GarageYrBlt.1933        -6.255e+04  4.264e+04  -1.467 0.143492    
## GarageYrBlt.1934         1.183e+04  5.096e+04   0.232 0.816618    
## GarageYrBlt.1935        -4.986e+04  3.637e+04  -1.371 0.171479    
## GarageYrBlt.1936        -2.245e+04  4.472e+04  -0.502 0.616097    
## GarageYrBlt.1937        -7.902e+04  4.365e+04  -1.810 0.071308 .  
## GarageYrBlt.1938        -5.447e+04  4.755e+04  -1.146 0.252958    
## GarageYrBlt.1939        -4.190e+04  3.415e+04  -1.227 0.220868    
## GarageYrBlt.1940        -5.173e+04  3.335e+04  -1.551 0.121996    
## GarageYrBlt.1941        -4.382e+04  3.370e+04  -1.300 0.194591    
## GarageYrBlt.1942                NA         NA      NA       NA    
## GarageYrBlt.1945        -2.553e+04  3.903e+04  -0.654 0.513556    
## GarageYrBlt.1946        -7.045e+04  3.637e+04  -1.937 0.053713 .  
## GarageYrBlt.1947                NA         NA      NA       NA    
## GarageYrBlt.1948        -6.189e+04  3.189e+04  -1.941 0.053285 .  
## GarageYrBlt.1949        -5.761e+04  3.461e+04  -1.664 0.097127 .  
## GarageYrBlt.1950        -5.822e+04  3.171e+04  -1.836 0.067426 .  
## GarageYrBlt.1951        -2.329e+04  3.916e+04  -0.595 0.552507    
## GarageYrBlt.1952        -5.641e+04  4.132e+04  -1.365 0.173275    
## GarageYrBlt.1953        -6.088e+04  3.142e+04  -1.938 0.053651 .  
## GarageYrBlt.1954        -5.671e+04  3.065e+04  -1.850 0.065315 .  
## GarageYrBlt.1955        -6.121e+04  3.423e+04  -1.788 0.074857 .  
## GarageYrBlt.1956        -5.344e+04  3.211e+04  -1.664 0.097143 .  
## GarageYrBlt.1957        -6.061e+04  3.137e+04  -1.932 0.054372 .  
## GarageYrBlt.1958        -5.356e+04  3.192e+04  -1.678 0.094522 .  
## GarageYrBlt.1959        -4.272e+04  3.220e+04  -1.327 0.185689    
## GarageYrBlt.1960        -5.509e+04  3.704e+04  -1.487 0.138021
## GarageYrBlt.1961        -4.915e+04  3.607e+04  -1.362 0.174137
## GarageYrBlt.1962        -5.821e+04  3.252e+04  -1.790 0.074564 .
## GarageYrBlt.1963        -7.390e+04  3.244e+04  -2.278 0.023468 *
## GarageYrBlt.1964        -5.320e+04  3.241e+04  -1.641 0.101824
## GarageYrBlt.1965        -8.058e+04  3.277e+04  -2.459 0.014541 *
## GarageYrBlt.1966        -4.424e+04  3.193e+04  -1.386 0.166933
## GarageYrBlt.1967        -5.891e+04  3.438e+04  -1.713 0.087720 .
## GarageYrBlt.1968        -4.056e+04  3.168e+04  -1.280 0.201458
## GarageYrBlt.1969        -6.728e+04  3.212e+04  -2.094 0.037112 *
## GarageYrBlt.1970        -6.800e+04  3.467e+04  -1.961 0.050813 .
## GarageYrBlt.1971        -5.613e+04  3.610e+04  -1.555 0.121154
## GarageYrBlt.1972        -3.946e+04  3.317e+04  -1.190 0.235193
## GarageYrBlt.1973        -3.818e+04  3.301e+04  -1.157 0.248410
## GarageYrBlt.1974        -5.890e+04  3.213e+04  -1.833 0.067801 .
## GarageYrBlt.1975        -5.644e+04  3.249e+04  -1.737 0.083395 .
## GarageYrBlt.1976        -4.927e+04  3.229e+04  -1.526 0.128169
## GarageYrBlt.1977        -6.565e+04  3.090e+04  -2.125 0.034464 *
## GarageYrBlt.1978        -4.607e+04  3.257e+04  -1.415 0.158301
## GarageYrBlt.1979        -4.428e+04  3.627e+04  -1.221 0.223125
## GarageYrBlt.1980        -4.475e+04  3.344e+04  -1.338 0.181905
## GarageYrBlt.1981        -7.894e+04  3.473e+04  -2.273 0.023764 *
## GarageYrBlt.1982        -7.953e+04  3.623e+04  -2.195 0.028972 *
## GarageYrBlt.1983        -5.094e+04  3.635e+04  -1.401 0.162262
## GarageYrBlt.1984        -6.810e+04  3.357e+04  -2.028 0.043458 *
## GarageYrBlt.1985        -4.679e+04  3.687e+04  -1.269 0.205493
## GarageYrBlt.1986        -4.686e+04  3.272e+04  -1.432 0.153193
## GarageYrBlt.1987        -6.119e+04  3.266e+04  -1.874 0.062031 .
## GarageYrBlt.1988        -6.511e+04  3.795e+04  -1.716 0.087302 .
## GarageYrBlt.1989        -3.502e+04  3.550e+04  -0.987 0.324706
## GarageYrBlt.1990        -8.299e+04  3.282e+04  -2.529 0.011994 *
## GarageYrBlt.1991        -9.337e+04  3.816e+04  -2.447 0.015019 *
## GarageYrBlt.1992        -8.282e+04  3.254e+04  -2.545 0.011463 *
## GarageYrBlt.1993        -6.892e+04  3.156e+04  -2.184 0.029796 *
## GarageYrBlt.1994        -5.789e+04  3.075e+04  -1.882 0.060808 .
## GarageYrBlt.1995        -3.701e+04  3.212e+04  -1.152 0.250142
## GarageYrBlt.1996        -5.940e+04  3.045e+04  -1.951 0.052053 .
## GarageYrBlt.1997        -5.760e+04  3.052e+04  -1.887 0.060194 .
## GarageYrBlt.1998        -5.594e+04  3.030e+04  -1.846 0.065956 .
## GarageYrBlt.1999        -5.547e+04  3.052e+04  -1.817 0.070232 .
## GarageYrBlt.2000        -5.944e+04  3.138e+04  -1.895 0.059177 .
## GarageYrBlt.2001        -4.522e+04  3.033e+04  -1.491 0.137109
## GarageYrBlt.2002        -5.389e+04  2.992e+04  -1.801 0.072762 .
## GarageYrBlt.2003        -5.088e+04  2.974e+04  -1.711 0.088181 .
## GarageYrBlt.2004        -5.640e+04  2.960e+04  -1.905 0.057748 .
## GarageYrBlt.2005        -5.910e+04  2.889e+04  -2.046 0.041720 *
## GarageYrBlt.2006        -4.637e+04  2.890e+04  -1.605 0.109676
## GarageYrBlt.2007        -6.521e+04  2.879e+04  -2.265 0.024278 *
## GarageYrBlt.2008        -3.818e+04  2.895e+04  -1.319 0.188292
## GarageYrBlt.2009        -2.756e+04  2.934e+04  -0.939 0.348341
## GarageYrBlt.2010                  NA         NA      NA       NA    
## GarageYrBlt.NoGarage              NA         NA      NA       NA    
## GarageFinish.Fin          -4.663e+03  5.227e+03  -0.892 0.373171    
## GarageFinish.NoGarage             NA         NA      NA       NA    
## GarageFinish.RFn          -3.929e+03  4.914e+03  -0.800 0.424629    
## GarageFinish.Unf                  NA         NA      NA       NA    
## GarageCars                 1.398e+04  4.503e+03   3.104 0.002101 ** 
## GarageArea                -2.763e+01  1.484e+01  -1.862 0.063583 .  
## GarageQual.Ex              2.405e+03  2.157e+04   0.111 0.911304    
## GarageQual.Fa              4.940e+02  9.812e+03   0.050 0.959882    
## GarageQual.Gd              2.895e+03  1.260e+04   0.230 0.818475    
## GarageQual.NoGarage               NA         NA      NA       NA    
## GarageQual.Po             -2.088e+03  7.394e+04  -0.028 0.977493    
## GarageQual.TA                     NA         NA      NA       NA    
## GarageCond.Ex                     NA         NA      NA       NA    
## GarageCond.Fa             -1.943e+03  1.161e+04  -0.167 0.867258    
## GarageCond.Gd             -4.747e+04  3.348e+04  -1.418 0.157292    
## GarageCond.NoGarage               NA         NA      NA       NA    
## GarageCond.Po              3.261e+04  4.355e+04   0.749 0.454500    
## GarageCond.TA                     NA         NA      NA       NA    
## PavedDrive.N              -7.403e+03  8.183e+03  -0.905 0.366406    
## PavedDrive.P               3.617e+03  1.254e+04   0.288 0.773291    
## PavedDrive.Y                      NA         NA      NA       NA    
## WoodDeckSF                 9.848e+00  1.156e+01   0.852 0.395004    
## OpenPorchSF                7.104e+01  2.404e+01   2.956 0.003385 ** 
## EnclosedPorch             -5.696e+00  2.927e+01  -0.195 0.845877    
## X3SsnPorch                 2.525e+01  4.080e+01   0.619 0.536498    
## ScreenPorch                6.407e+01  2.609e+01   2.456 0.014661 *  
## PoolArea                   9.740e+01  5.541e+01   1.758 0.079905 .  
## PoolQC.Ex                  1.476e+05  4.775e+04   3.090 0.002201 ** 
## PoolQC.Fa                 -4.239e+04  4.551e+04  -0.931 0.352454    
## PoolQC.Gd                         NA         NA      NA       NA    
## PoolQC.NoPool                     NA         NA      NA       NA    
## Fence.GdPrv                1.004e+03  8.002e+03   0.125 0.900246    
## Fence.GdWo                -2.848e+03  7.362e+03  -0.387 0.699172    
## Fence.MnPrv               -4.099e+02  4.967e+03  -0.083 0.934294    
## Fence.MnWw                -1.670e+04  1.308e+04  -1.277 0.202778    
## Fence.NoFence                     NA         NA      NA       NA    
## MiscFeature.Gar2                  NA         NA      NA       NA    
## MiscFeature.None          -1.154e+03  1.516e+04  -0.076 0.939381    
## MiscFeature.Othr                  NA         NA      NA       NA    
## MiscFeature.Shed                  NA         NA      NA       NA    
## MiscFeature.TenC                  NA         NA      NA       NA    
## MiscVal                    1.053e+00  2.025e+01   0.052 0.958581    
## MoSold                    -4.242e+02  4.941e+02  -0.858 0.391367    
## YrSold                    -1.805e+03  1.046e+03  -1.726 0.085475 .  
## SaleType.COD               9.209e+03  1.071e+04   0.860 0.390682    
## SaleType.Con               9.246e+04  3.546e+04   2.608 0.009601 ** 
## SaleType.ConLD             1.195e+04  2.375e+04   0.503 0.615297    
## SaleType.ConLI             4.573e+02  2.259e+04   0.020 0.983866    
## SaleType.ConLw                  NA         NA      NA       NA    
## SaleType.CWD             8.937e+04  2.733e+04   3.270 0.001208 ** 
## SaleType.New             9.907e+03  3.488e+04   0.284 0.776599    
## SaleType.Oth            -7.517e+03  2.162e+04  -0.348 0.728308    
## SaleType.WD                     NA         NA      NA       NA    
## SaleCondition.Abnorml   -8.228e+03  3.576e+04  -0.230 0.818201    
## SaleCondition.AdjLand   -1.706e+04  4.840e+04  -0.352 0.724811    
## SaleCondition.Alloca    -7.116e+03  3.865e+04  -0.184 0.854074    
## SaleCondition.Family    -1.603e+03  3.640e+04  -0.044 0.964898    
## SaleCondition.Normal     3.344e+03  3.486e+04   0.096 0.923657    
## SaleCondition.Partial           NA         NA      NA       NA    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 21630 on 282 degrees of freedom
##   (132 observations deleted due to missingness)
## Multiple R-squared:  0.9712, Adjusted R-squared:  0.939 
## F-statistic: 30.11 on 316 and 282 DF,  p-value: < 2.2e-16
```

#Our R-Squared of 0.93 is not bad at all. Looking at the coefficients and their corresponding values, we see there are lots of predictors that we can drop or are not significant. The F-Statistic of 45 shows that there is relationship between the response variable - 'SalePrice' and predictors. Quick side note: Referencing and cross checking, highly correlated variables with SalePrice in our correlation plot above and simple linear regression, we can be assured that the highly correlated variables are indeed significant variables.

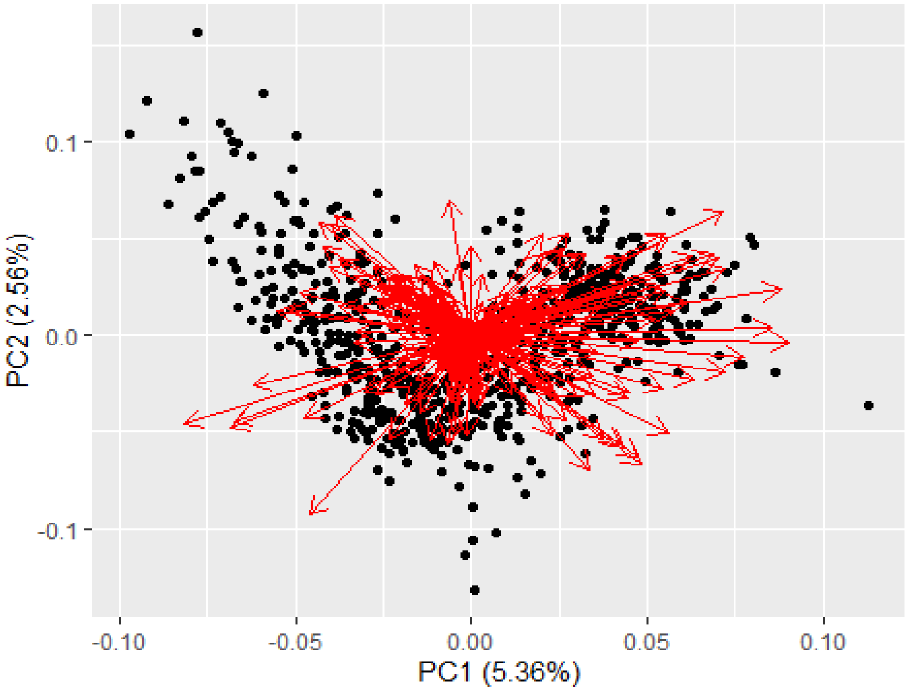
# Principal component analysis

```
# PCA works well on normalized dataset.
# This is because there could be large loadings due to the way variables are measured.
training.scaled <- data.frame(apply(training, 2, scale))
# Remove missing values or NAs
# sum(is.na(training.scaled))
training.scale.na.omit <- data.frame(t(na.omit(t(training.scaled))))
# Run PCA
training_pca <- prcomp(training.scale.na.omit, retx=TRUE)
names(training_pca)
```

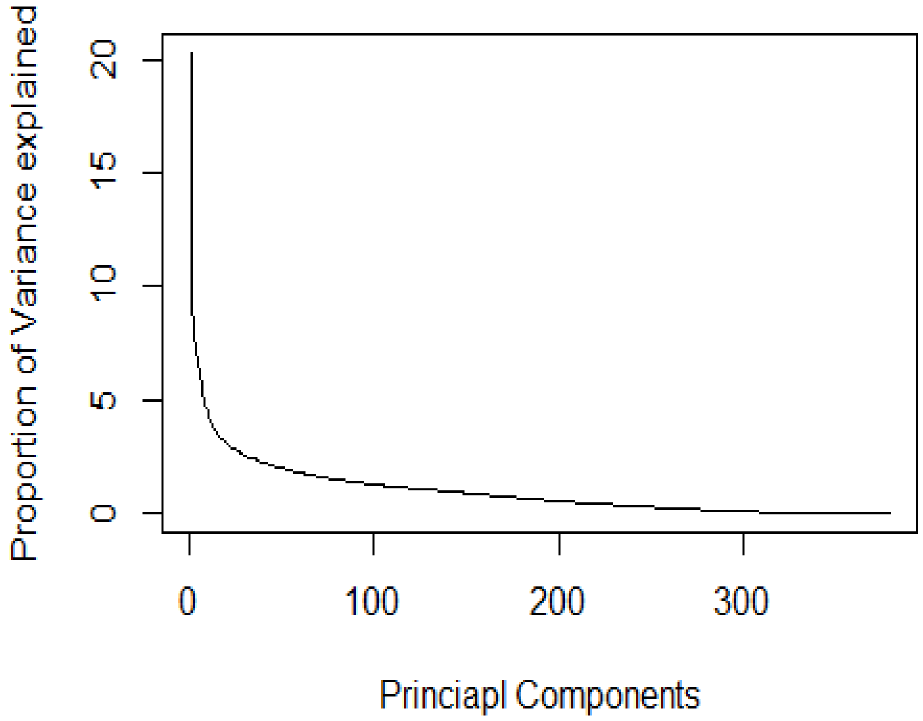
```
## [1] "sdev"     "rotation" "center"   "scale"    "x"
```

```
training_pca$center
```

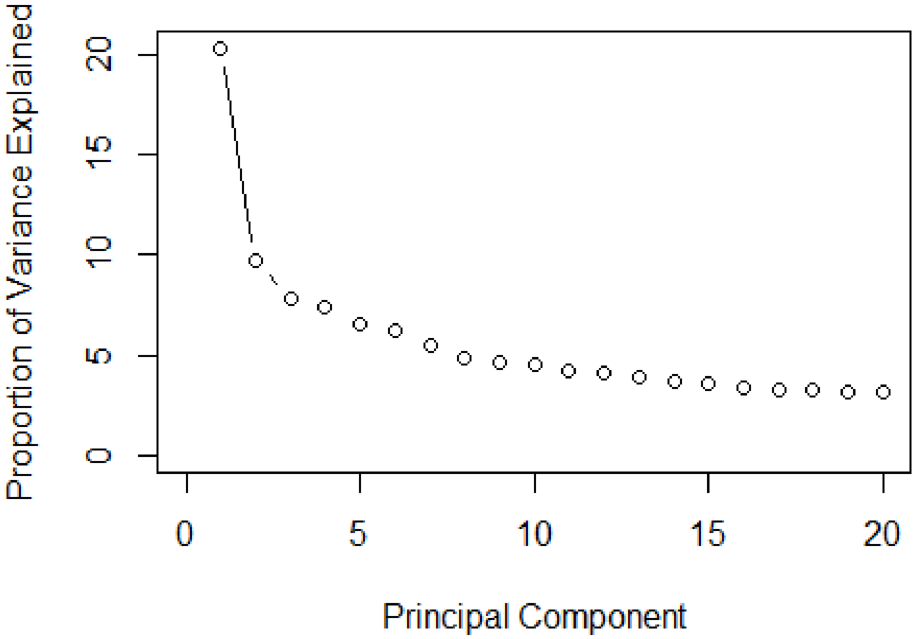
```
##                     Id              MSSubClass        MSZoning.C..all. 
##          -2.278159e-18           -7.783711e-17           -3.037546e-19 
##            MSZoning.FV             MSZoning.RH             MSZoning.RL 
##            3.493178e-17            1.822528e-18            2.688228e-17
##             MSZoning.RM                 LotArea             Street.Grvl
##            5.308111e-17           -9.896408e-17           -8.884822e-18
.

training_pca$scale

## [1] FALSE

training_pca$rotation

##                                     PC1           PC2           PC3
PC4
## Id                         0.0008392510  4.403398e-04 -2.010423e-03  1.48952
7e-02
.
##                                  PC377         PC378         PC379
## Id                         0.000000e+00  0.000000e+00  0.000000e+00
## MSSubClass                -1.863201e-16  2.744116e-16 -4.228727e-16
## MSZoning.C..all.           7.702083e-03  3.279661e-03  1.830014e-02
.
##  [ reached getOption("max.print") -- omitted 116 rows ]

dim(training_pca$x)

## [1] 731 379

# This returns 286 principal component loadings.
# Plot
autoplot(training_pca, loadings = TRUE)
```



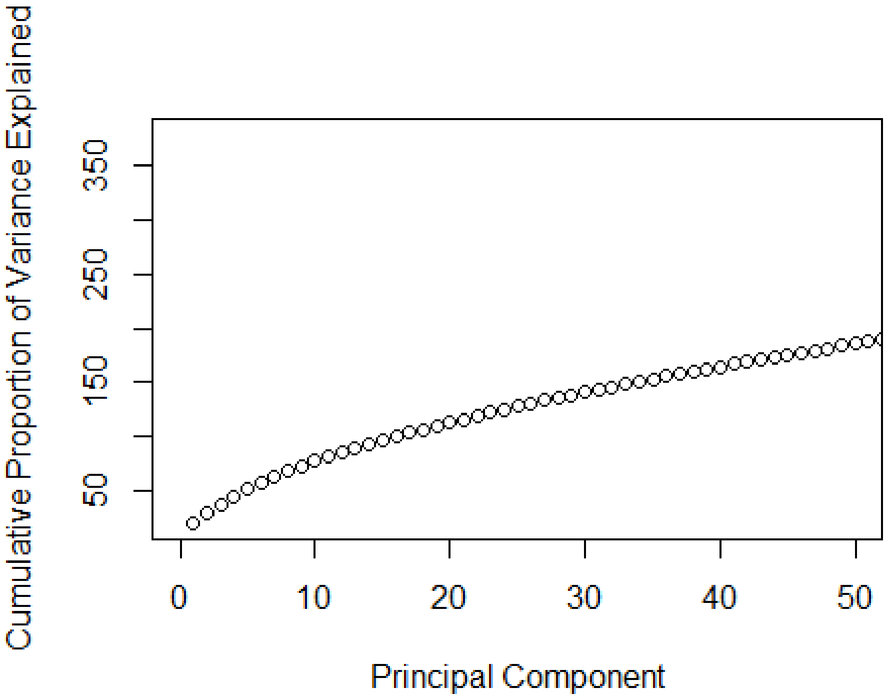
```
summary(training_pca)

## Importance of components:
##                             PC1     PC2     PC3     PC4     PC5     PC6
PC7
## Standard deviation     4.50834 3.11391 2.79687 2.71610 2.55075 2.48698 2.3
4173
## Proportion of Variance 0.05363 0.02558 0.02064 0.01946 0.01717 0.01632 0.0
1447
## Cumulative Proportion  0.05363 0.07921 0.09985 0.11932 0.13648 0.15280 0.1
6727

.
##                             PC379
## Standard deviation     4.515e-17
## Proportion of Variance 0.000e+00
## Cumulative Proportion  1.000e+00

#The 1 PC explains 6.8%, 2 PC explains 3.1% of variance in the data and so on
.
# Calculate Variance
pr_var <- training_pca$sdev^2
plot(pr_var, type = "l", xlab = "Princiapl Components", ylab = "Proportion of
Variance explained")
```



```
#or
plot(pr_var, xlab = "Principal Component", ylab = "Proportion of Variance Explained", type = "b", xlim=c(0, 20))
```



```
#cumulative variance plot
plot(cumsum(pr_var), xlab = "Principal Component", ylab = "Cumulative Proportion of Variance Explained", type = "b", xlim=c(0, 50))
```



```
#The plot method returns a plot of the variances (y-axis) associated with the
PCs (x-axis). The Figure below is useful to decide how many PCs to retain for
further analysis.

# Transformation similar to training set.
#Add a training set with principal components
training.data.pca <- data.frame(training$SalePrice, training_pca$x)
# Extract first 40 Principal Components
training.data.pca <- training.data.pca[,1:40]


# Run a linear regression with PCA transformed data
dim(training.data.pca)
```

```
## [1] 731  40
```

```
l.model <- lm(training.SalePrice ~ ., data = training.data.pca)
summary(l.model)
```

```
## 
## Call:
## lm(formula = training.SalePrice ~ ., data = training.data.pca)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -259741  -16274     270   13475  234901 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 183169.4     1186.8 154.334  < 2e-16 ***
## PC1          14594.5      263.4  55.401  < 2e-16 ***
## PC2            741.8      381.4   1.945 0.052200 .  
## PC3           2139.8      424.6   5.039 5.97e-07 ***
## PC4          10354.8      437.3  23.681  < 2e-16 ***
## PC5           1487.0      465.6   3.194 0.001468 ** 
## PC6            856.8      477.5   1.794 0.073228 .  
## PC7           3859.7      507.2   7.610 8.99e-14 ***
## PC8           2000.5      539.7   3.707 0.000227 ***
## PC9            344.6      551.9   0.624 0.532616    
## PC10         -2123.0      560.0  -3.791 0.000163 ***
## PC11         -1013.6      581.0  -1.744 0.081532 .  
## PC12          3193.0      586.8   5.442 7.34e-08 ***
## PC13         -5857.4      601.5  -9.737  < 2e-16 ***
## PC14          2288.9      616.1   3.715 0.000219 ***
## PC15          1172.5      627.4   1.869 0.062080 .  
## PC16          -965.4      643.6  -1.500 0.134059    
## PC17         -3747.6      655.5  -5.717 1.61e-08 ***
## PC18         -1627.4      657.9  -2.474 0.013618 *  
## PC19          -735.7      667.1  -1.103 0.270526    
## PC20          2561.1      672.8   3.807 0.000153 ***
## PC21          3489.3      681.9   5.117 4.02e-07 ***
## PC22         -1202.4      689.6  -1.744 0.081659 .  
## PC23         -1300.5      697.7  -1.864 0.062737 .  
## PC24          -687.8      701.1  -0.981 0.326911    
## PC25         -4597.6      705.6  -6.516 1.39e-10 ***
## PC26         -1615.8      710.5  -2.274 0.023252 *  
## PC27          -922.1      722.7  -1.276 0.202365    
## PC28         -3601.4      724.7  -4.969 8.48e-07 ***
## PC29         -4174.0      730.5  -5.713 1.65e-08 ***
## PC30         -2522.2      743.1  -3.394 0.000728 ***
## PC31          1106.8      745.7   1.484 0.138193    
## PC32         -2261.0      755.5  -2.993 0.002864 ** 
## PC33          -405.7      757.6  -0.536 0.592476    
## PC34         -1175.0      762.4  -1.541 0.123714    
## PC35          2303.9      766.7   3.005 0.002753 ** 
## PC36          -283.3      769.9  -0.368 0.713040    
## PC37         -4194.4      776.1  -5.404 8.97e-08 ***
## PC38          -843.8      780.7  -1.081 0.280139    
## PC39           361.6      789.5   0.458 0.647107    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 32090 on 691 degrees of freedom
## Multiple R-squared:  0.8577, Adjusted R-squared:  0.8497 
## F-statistic: 106.8 on 39 and 691 DF,  p-value: < 2.2e-16
```