(PREDICTION OF HOUSE PRICE)

1. Introduction

An accurate prediction on the house price is important to prospective homeowners, developers, investors, appraisers, tax assessors and other real estate market participants, such as, mortgage lenders and insurers. Traditional house price prediction is based on cost and sale price comparison lacking an accepted standard and a certification process. Therefore, the availability of a house price pre Over the last two decades there has been a proliferation of empirical studies analyzing residential property values. Each succeeding research has generally improved the predictive power of the mdiction model helps fill up a valuable information gap and improve the efficiency of the real estate market. In New Zealand, most people know the benefit of owning a house, because buying a house is considered the most utilized and profitable investment. New Zealand has one of the highest ratios of people owning their houses in western world with over 70% of its citizens living in their own houses. As house market in New Zealand is thriving, house price becomes a crucial factor for house seekers.odels by emphasizing attributes of property value such as housing site, housing quality, geographical location and the environment. More recent studies have focused on location externalities, transaction costs and factors affecting the future expected cost in homeownership. Reference:https://ageconsearch.umn.edu/bitstream/97781/2/2004-9-house%20price%20prediction.pdf

2. Data Description

In this dataset, there are 1460 observations with 81 explanatory variables describing (almost) every aspect of residential homes in Ames, Iowa. Among explanatory variables, there are 37 integer variables, such as Id, MSSubClass, LotFrontage, and 43 factor variables, such as MSZoning, Street, LotShape.

The description of the data fields are as follows:

. SalePrice - the property’s sale price in dollars. This is the target variable that you’re trying to predict.

. MSSubClass: The building class

. MSZoning: The general zoning classification

. LotFrontage: Linear feet of street connected to property

. LotArea: Lot size in square feet

. Street: Type of road access

. Alley: Type of alley access

. LotShape: General shape of property

. LandContour: Flatness of the property

. Utilities: Type of utilities available

. LotConfig: Lot configuration

. LandSlope: Slope of property

. Neighborhood: Physical locations within Ames city limits

. Condition1: Proximity to main road or railroad

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

. BldgType: Type of dwelling

. HouseStyle: Style of dwelling

. OverallQual: Overall material and finish quality

. OverallCond: Overall condition rating

. YearBuilt: Original construction date

. YearRemodAdd: Remodel date

. RoofStyle: Type of roof

. RoofMatl: Roof material

. Exterior1st: Exterior covering on house

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

. MasVnrType: Masonry veneer type

. MasVnrArea: Masonry veneer area in square feet

. ExterQual: Exterior material quality

. ExterCond: Present condition of the material on the exterior

. Foundation: Type of foundation

. BsmtQual: Height of the basement

. BsmtCond: General condition of the basement

. BsmtExposure: Walkout or garden level basement walls

. BsmtFinType1: Quality of basement finished area

. BsmtFinSF1: Type 1 finished square feet

. BsmtFinType2: Quality of second finished area (if present)

. BsmtFinSF2: Type 2 finished square feet

. BsmtUnfSF: Unfinished square feet of basement area

. TotalBsmtSF: Total square feet of basement area

. Heating: Type of heating

. HeatingQC: Heating quality and condition

. CentralAir: Central air conditioning

. Electrical: Electrical system

. 1stFlrSF: First Floor square feet

. 2ndFlrSF: Second floor square feet

. LowQualFinSF: Low quality finished square feet (all floors)

. GrLivArea: Above grade (ground) living area square feet

. BsmtFullBath: Basement full bathrooms

. BsmtHalfBath: Basement half bathrooms

. FullBath: Full bathrooms above grade

. HalfBath: Half baths above grade

. Bedroom: Number of bedrooms above basement level

. Kitchen: Number of kitchens

. KitchenQual: Kitchen quality

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

. Functional: Home functionality rating

. Fireplaces: Number of fireplaces

. FireplaceQu: Fireplace quality

. GarageType: Garage location

. GarageYrBlt: Year garage was built

. GarageFinish: Interior finish of the garage

. GarageCars: Size of garage in car capacity

. GarageArea: Size of garage in square feet

. GarageQual: Garage quality

. GarageCond: Garage condition

. PavedDrive: Paved driveway

. WoodDeckSF: Wood deck area in square feet

. OpenPorchSF: Open porch area in square feet

. EnclosedPorch: Enclosed porch area in square feet

. 3SsnPorch: Three season porch area in square feet

. ScreenPorch: Screen porch area in square feet

. PoolArea: Pool area in square feet

. PoolQC: Pool quality

. Fence: Fence quality

. MiscFeature: Miscellaneous feature not covered in other categories

. MiscVal: $Value of miscellaneous feature

. MoSold: Month Sold

. YrSold: Year Sold

. SaleType: Type of sale

. SaleCondition: Condition of sale

. SalePrice: Sale price of house

Data Source - I got this data set from https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data

Here, I will import the dataset to R.

```
train <- read.csv("C:/Users/gks/Desktop/train.csv")
```

Now, I will import test data set.

```
test <- read.csv("C:/Users/gks/Desktop/test.csv")
```

Now we will find how many columns and rows are present in train dataset.

```
cat('Train has', dim(train)[1], 'rows and', dim(train)[2], 'columns.')

## Train has 1460 rows and 81 columns.
```

We will find how many columns and rows are present in test data set.

```r
cat("Test has", dim(test)[1], "rows and", dim(test)[2], " columns.")
```

```
## Test has 1459 rows and 80  columns.
```

Percentage of data missing in train

```r
sum(is.na(train)) / (nrow(train) *ncol(train))
```

```
## [1] 0.05889565
```

3. Data Assessment

Summary of this dataset is as follows.

```r
summary(train)
```

```
##        Id           MSSubClass       MSZoning     LotFrontage
##  Min.   :   1.0   Min.   : 20.0   C (all):  10   Min.   : 21.00
##  1st Qu.: 365.8   1st Qu.: 20.0   FV     :  65   1st Qu.: 59.00
##  Median : 730.5   Median : 50.0   RH     :  16   Median : 69.00
##  Mean   : 730.5   Mean   : 56.9   RL     :1151   Mean   : 70.05
##  3rd Qu.:1095.2   3rd Qu.: 70.0   RM     : 218   3rd Qu.: 80.00
##  Max.   :1460.0   Max.   :190.0                  Max.   :313.00
##                                                  NA's   :259
##     LotArea         Street      Alley       LotShape  LandContour
##  Min.   :   1300   Grvl:   6   Grvl:  50   IR1:484   Bnk:  63
##  1st Qu.:   7554   Pave:1454   Pave:  41   IR2: 41   HLS:  50
##  Median :   9478               NA's:1369   IR3: 10   Low:  36
##  Mean   :  10517                           Reg:925   Lvl:1311
##  3rd Qu.:  11602
##  Max.   :215245
##
##   Utilities      LotConfig    LandSlope   Neighborhood   Condition1
##  AllPub:1459   Corner : 263   Gtl:1382   NAmes  :225   Norm    1260
##  NoSeWa:   1   CulDSac:  94   Mod:  65   CollgCr:150   Feedr  :  81
##                FR2    :  47   Sev:  13   OldTown:113   Artery :  48
##                FR3    :   4              Edwards:100   RRAn   :  26
##                Inside :1052              Somerst: 86   PosN   :  19
##                                          Gilbert: 79   RRAe   :  11
##                                          (Other):707   (Other):  15
##    Condition2     BldgType      HouseStyle    OverallQual
##  Norm   :1445   1Fam   :1220   1Story :726   Min.   : 1.000
##  Feedr  :   6   2fmCon:   31   2Story :445   1st Qu.: 5.000
##  Artery :   2   Duplex:   52   1.5Fin :154   Median : 6.000
##  PosN   :   2   Twnhs :   43   SLvl   : 65   Mean   : 6.099
##  RRNn   :   2   TwnhsE:  114   SFoyer : 37   3rd Qu.: 7.000
##  PosA   :   1                  1.5Unf : 14   Max.   :10.000
##  (Other):   2                  (Other): 19
##   OverallCond      YearBuilt      YearRemodAdd     RoofStyle
##  Min.   :1.000   Min.   :1872   Min.   :1950   Flat   :  13
##  1st Qu.:5.000   1st Qu.:1954   1st Qu.:1967   Gable  :1141
##  Median :5.000   Median :1973   Median :1994   Gambrel:  11
```

```
##  Mean   :5.575   Mean   :1971   Mean   :1985   Hip    : 286
##  3rd Qu.:6.000   3rd Qu.:2000   3rd Qu.:2004   Mansard:   7
##  Max.   :9.000   Max.   :2010   Max.   :2010   Shed   :   2
##
##     RoofMatl     Exterior1st    Exterior2nd    MasVnrType    MasVnrArea
##  CompShg:1434   VinylSd:515   VinylSd:504   BrkCmn : 15   Min.   :   0.0
##  Tar&Grv:  11   HdBoard:222   MetalSd:214   BrkFace:445   1st Qu.:   0.0
##  WdShngl:   6   MetalSd:220   HdBoard:207   None   :864   Median :   0.0
##  WdShake:   5   Wd Sdng:206   Wd Sdng:197   Stone  :128   Mean   : 103.7
##  ClyTile:   1   Plywood:108   Plywood:142   NA's   :  8   3rd Qu.: 166.0
##  Membran:   1   CemntBd: 61   CmentBd: 60                 Max.   :1600.0
##  (Other):   2   (Other):128   (Other):136                 NA's    8
##  ExterQual ExterCond  Foundation  BsmtQual   BsmtCond    BsmtExposure
##  Ex: 52    Ex:   3   BrkTil:146   Ex  :121   Fa  :  45   Av   221
##  Fa: 14    Fa:  28   CBlock:634   Fa  : 35   Gd  :  65   Gd   134
##  Gd:488    Gd: 146   PConc :647   Gd  :618   Po  :   2   Mn   114
##  TA:906    Po:   1   Slab  : 24   TA  :649   TA  :1311   No   953
##            TA:1282   Stone :  6   NA's: 37   NA's:  37   NA's: 38
##                      Wood  :  3
##
##  BsmtFinType1   BsmtFinSF1     BsmtFinType2   BsmtFinSF2
##  ALQ :220     Min.   :   0.0   ALQ :  19    Min.   :   0.00
##  BLQ :148     1st Qu.:   0.0   BLQ :  33    1st Qu.:   0.00
##  GLQ :418     Median : 383.5   GLQ :  14    Median :   0.00
##  LwQ : 74     Mean   : 443.6   LwQ :  46    Mean   :  46.55
##  Rec :133     3rd Qu.: 712.2   Rec :  54    3rd Qu.:   0.00
##  Unf :430     Max.   :5644.0   Unf :1256    Max.   :1474.00
##  NA's: 37                      NA's:  38
##    BsmtUnfSF       TotalBsmtSF       Heating      HeatingQC CentralAir
##  Min.   :   0.0   Min.   :   0.0   Floor:   1   Ex:741    N:  95
##  1st Qu.: 223.0   1st Qu.: 795.8   GasA :1428   Fa: 49    Y:1365
##  Median : 477.5   Median : 991.5   GasW :  18   Gd:241
##  Mean   : 567.2   Mean   :1057.4   Grav :   7   Po:  1
##  3rd Qu.: 808.0   3rd Qu.:1298.2   OthW :   2   TA:428
##  Max.   :2336.0   Max.   :6110.0   Wall :   4
##
##  Electrical     X1stFlrSF      X2ndFlrSF     LowQualFinSF
##  FuseA:  94   Min.   : 334   Min.   :   0   Min.   :  0.000
##  FuseF:  27   1st Qu.: 882   1st Qu.:   0   1st Qu.:  0.000
##  FuseP:   3   Median :1087   Median :   0   Median :  0.000
##  Mix  :   1   Mean   :1163   Mean   : 347   Mean   :  5.845
##  SBrkr:1334   3rd Qu.:1391   3rd Qu.: 728   3rd Qu.:  0.000
##  NA's :   1   Max.   :4692   Max.   :2065   Max.   :572.000
##
##    GrLivArea      BsmtFullBath     BsmtHalfBath        FullBath
##  Min.   : 334   Min.   :0.0000   Min.   :0.00000   Min.   :0.000
##  1st Qu.:1130   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:1.000
##  Median :1464   Median :0.0000   Median :0.00000   Median :2.000
##  Mean   :1515   Mean   :0.4253   Mean   :0.05753   Mean   :1.565
##  3rd Qu.:1777   3rd Qu.:1.0000   3rd Qu.:0.00000   3rd Qu.:2.000
```

```
##  Max.   :5642   Max.   :3.0000   Max.   :2.00000   Max.   :3.000  
##                                                                    
##     HalfBath       BedroomAbvGr    KitchenAbvGr   KitchenQual
##  Min.   :0.0000   Min.   :0.000   Min.   :0.000   Ex:100     
##  1st Qu.:0.0000   1st Qu.:2.000   1st Qu.:1.000   Fa: 39     
##  Median :0.0000   Median :3.000   Median :1.000   Gd:586     
##  Mean   :0.3829   Mean   :2.866   Mean   :1.047   TA:735     
##  3rd Qu.:1.0000   3rd Qu.:3.000   3rd Qu.:1.000              
##  Max.   :2.0000   Max.   :8.000   Max.   :3.000              
##                                                              
##   TotRmsAbvGrd    Functional    Fireplaces    FireplaceQu   GarageType 
##  Min.   : 2.000   Maj1:  14   Min.   :0.000   Ex  : 24    2Types :  6  
##  1st Qu.: 5.000   Maj2:   5   1st Qu.:0.000   Fa  : 33    Attchd :870  
##  Median : 6.000   Min1:  31   Median :1.000   Gd  :380    Basment: 19  
##  Mean   : 6.518   Min2:  34   Mean   :0.613   Po  : 20    BuiltIn: 88  
##  3rd Qu.: 7.000   Mod :  15   3rd Qu.:1.000   TA  :313    CarPort:  9  
##  Max.   :14.000   Sev :   1   Max.   :3.000   NA's:690    Detchd :387  
##                   Typ :1360                               NA's   : 81  
##   GarageYrBlt   GarageFinish   GarageCars      GarageArea     GarageQual
##  Min.   :1900   Fin :352     Min.   :0.000   Min.   :   0.0   Ex  :   3  
##  1st Qu.:1961   RFn :422     1st Qu.:1.000   1st Qu.: 334.5   Fa  :  48  
##  Median :1980   Unf :605     Median :2.000   Median : 480.0   Gd  :  14  
##  Mean   :1979   NA's: 81     Mean   :1.767   Mean   : 473.0   Po  :   3  
##  3rd Qu.:2002                3rd Qu.:2.000   3rd Qu.: 576.0   TA  :1311  
##  Max.   :2010                Max.   :4.000   Max.   :1418.0   NA's:  81  
##  NA's   :81                                                              
##  GarageCond  PavedDrive   WoodDeckSF      OpenPorchSF     EnclosedPorch
##  Ex  :   2   N:  90     Min.   :  0.00   Min.   :  0.00   Min.   :  0.00  
##  Fa  :  35   P:  30     1st Qu.:  0.00   1st Qu.:  0.00   1st Qu.:  0.00  
##  Gd  :   9   Y:1340     Median :  0.00   Median : 25.00   Median :  0.00  
##  Po  :   7              Mean   : 94.24   Mean   : 46.66   Mean   : 21.95  
##  TA  :1326              3rd Qu.:168.00   3rd Qu.: 68.00   3rd Qu.:  0.00  
##  NA's:  81              Max.   :857.00   Max.   :547.00   Max.   :552.00  
##                                                                           
##    X3SsnPorch      ScreenPorch        PoolArea        PoolQC    
##  Min.   :  0.00   Min.   :  0.00   Min.   :  0.000   Ex  :   2  
##  1st Qu.:  0.00   1st Qu.:  0.00   1st Qu.:  0.000   Fa  :   2  
##  Median :  0.00   Median :  0.00   Median :  0.000   Gd  :   3  
##  Mean   :  3.41   Mean   : 15.06   Mean   :  2.759   NA's:1453  
##  3rd Qu.:  0.00   3rd Qu.:  0.00   3rd Qu.:  0.000              
##  Max.   :508.00   Max.   :480.00   Max.   :738.000              
##                                                                 
##    Fence      MiscFeature    MiscVal             MoSold      
##  GdPrv:  59   Gar2:   2   Min.   :    0.00   Min.   : 1.000  
##  GdWo :  54   Othr:   2   1st Qu.:    0.00   1st Qu.: 5.000  
##  MnPrv: 157   Shed:  49   Median :    0.00   Median : 6.000  
##  MnWw :  11   TenC:   1   Mean   :   43.49   Mean   : 6.322  
##  NA's :1179   NA's:1406   3rd Qu.:    0.00   3rd Qu.: 8.000  
##                           Max.   :15500.00   Max.   :12.000  
##                                                              
```

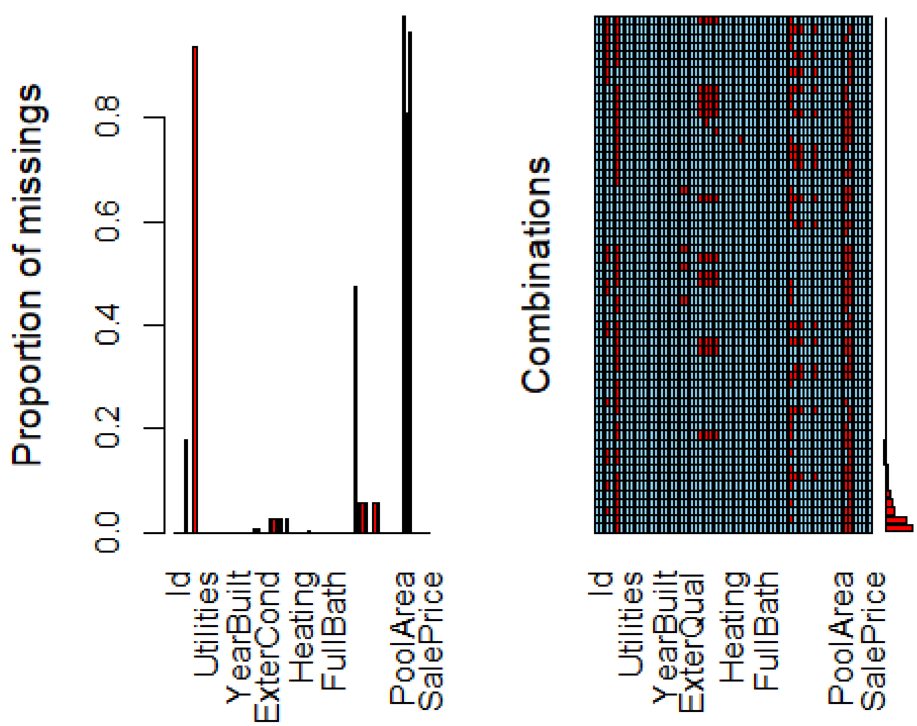
```
##       YrSold          SaleType    SaleCondition    SalePrice
##  Min.    :2006   WD      :1267   Abnorml: 101   Min.   : 34900
##  1st Qu.:2007   New     : 122   AdjLand:   4   1st Qu.:129975
##  Median :2008   COD     :  43   Alloca :  12   Median :163000
##  Mean    :2008   ConLD  :   9   Family :  20   Mean    :180921
##  3rd Qu.:2009   ConLI  :   5   Normal :1198   3rd Qu.:214000
##  Max.    :2010   ConLw  :   5   Partial: 125   Max.    :755000
##                  (Other):   9
```

Here, we can get the information about the mean, median, first quartile, third quartile of all columns and we can also see the number of NAs present in each column. We can see that the houses are sold between 2006 and 2010 and in which the normal conditioned housees are sold the most. The average sale price of houses that are sold in around 180921. Single family houses are sold the most. These houses have an average basement area of around 1057 sq ft.

3.1 Data Wrangling:

Now we will visualize is there any missing data in train data set? Refrence: https://cran.r-project.org/web/packages/VIMGUI/vignettes/VIM-Imputation.pdf

```r
library(VIM)

aggr(train)
```



From, this output we can say that there is some missing value in ID, Street, poolArea , ExterQual and some other columns. So, we have to remove missing values from this dataset.

3.1.1 Check for Missing Data

We can see the number of rows by using the following command.

```
nrow(train)

## [1] 1460
```

From the following output we can say that there are 1460 rows in this dataset.

First, we have to see if there are any missing data in the dataset. For that I used following command.

```
missing_row <- train[!complete.cases(train),]
nrow(missing_row)

## [1] 1460
```

Here, we can see that all rows have NAs. We can not clear NAs from all the rows. The other option is that we can also manually provide values to NA's but for the house price prediction we will not use all columns. So, we will take some columns and build subset of train data set. And try to clean the subset.

We can see all the columns using the following command.

```
colnames(train)

##  [1] "Id"            "MSSubClass"    "MSZoning"      "LotFrontage"
##  [5] "LotArea"       "Street"        "Alley"         "LotShape"
##  [9] "LandContour"   "Utilities"     "LotConfig"     "LandSlope"
## [13] "Neighborhood"  "Condition1"    "Condition2"    "BldgType"
## [17] "HouseStyle"    "OverallQual"   "OverallCond"   "YearBuilt"
## [21] "YearRemodAdd"  "RoofStyle"     "RoofMatl"      "Exterior1st"
## [25] "Exterior2nd"   "MasVnrType"    "MasVnrArea"    "ExterQual"
## [29] "ExterCond"     "Foundation"    "BsmtQual"      "BsmtCond"
## [33] "BsmtExposure"  "BsmtFinType1"  "BsmtFinSF1"    "BsmtFinType2"
## [37] "BsmtFinSF2"    "BsmtUnfSF"     "TotalBsmtSF"   "Heating"
## [41] "HeatingQC"     "CentralAir"    "Electrical"    "X1stFlrSF"
## [45] "X2ndFlrSF"     "LowQualFinSF"  "GrLivArea"     "BsmtFullBath"
## [49] "BsmtHalfBath"  "FullBath"      "HalfBath"      "BedroomAbvGr"
## [53] "KitchenAbvGr"  "KitchenQual"   "TotRmsAbvGrd"  "Functional"
## [57] "Fireplaces"    "FireplaceQu"   "GarageType"    "GarageYrBlt"
## [61] "GarageFinish"  "GarageCars"    "GarageArea"    "GarageQual"
## [65] "GarageCond"    "PavedDrive"    "WoodDeckSF"    "OpenPorchSF"
## [69] "EnclosedPorch" "X3SsnPorch"    "ScreenPorch"   "PoolArea"
## [73] "PoolQC"        "Fence"         "MiscFeature"   "MiscVal"
## [77] "MoSold"        "YrSold"        "SaleType"      "SaleCondition"
## [81] "SalePrice"
```

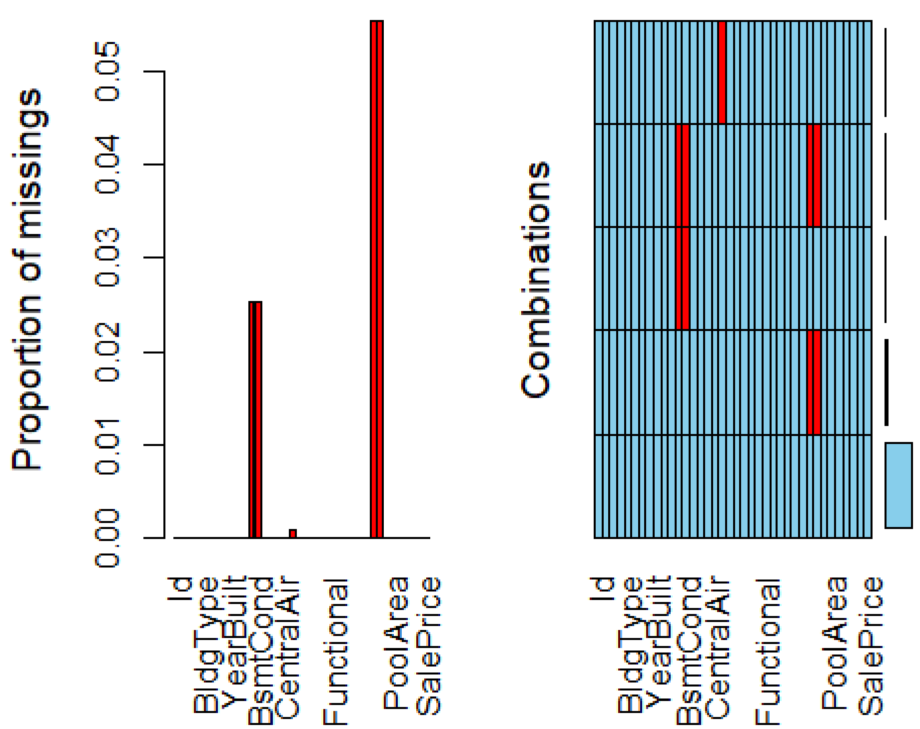
We dont need all columns to predict the house price so we eliminate some columns and build new data set which help us to predict house price.

```
#library(magrittr)   #you have to install and load this package
library(dplyr)

#library(tidyr)
req_var <-  train %>% select('Id',"MSZoning","Utilities",
"Neighborhood","BldgType","HouseStyle",
"OverallQual","OverallCond","YearBuilt", "ExterQual","ExterCond",
"BsmtQual","BsmtCond","TotalBsmtSF","Heating","HeatingQC",
"CentralAir","Electrical","GrLivArea","BedroomAbvGr","KitchenAbvGr",
"KitchenQual","TotRmsAbvGrd","Functional", "BsmtHalfBath" , "FullBath",
"HalfBath" ,
"BsmtFullBath",
"GarageArea","GarageQual","GarageCond","OpenPorchSF","PoolArea","MoSold",'YrSold',"SaleType","SaleCondition","SalePrice")
```

Now we will use this data set for future prediction and visualization.

Now we will see in how many columns, missing values are present. Refrence: https://cran.r-project.org/web/packages/VIMGUI/vignettes/VIM-Imputation.pdf

```
library(VIM)
aggr(req_var)
```



From this output we can say that in ExterCond, BsmtCond and in other columns missing values are present.

Now we will check is there any duplicate rows in this data set?

```
cat("The number of duplicated rows are", nrow(req_var) -
nrow(unique(req_var)))

## The number of duplicated rows are 0
```

There is no duplicate rows.

Now I will check is there any NA present in this data set? If yes, then I have to remove these NA from this data set.

```
library(tidyr)
req_var<- drop_na(req_var)
nrow(req_var)

## [1] 1348
```

After dropping NA, total row is 1348.

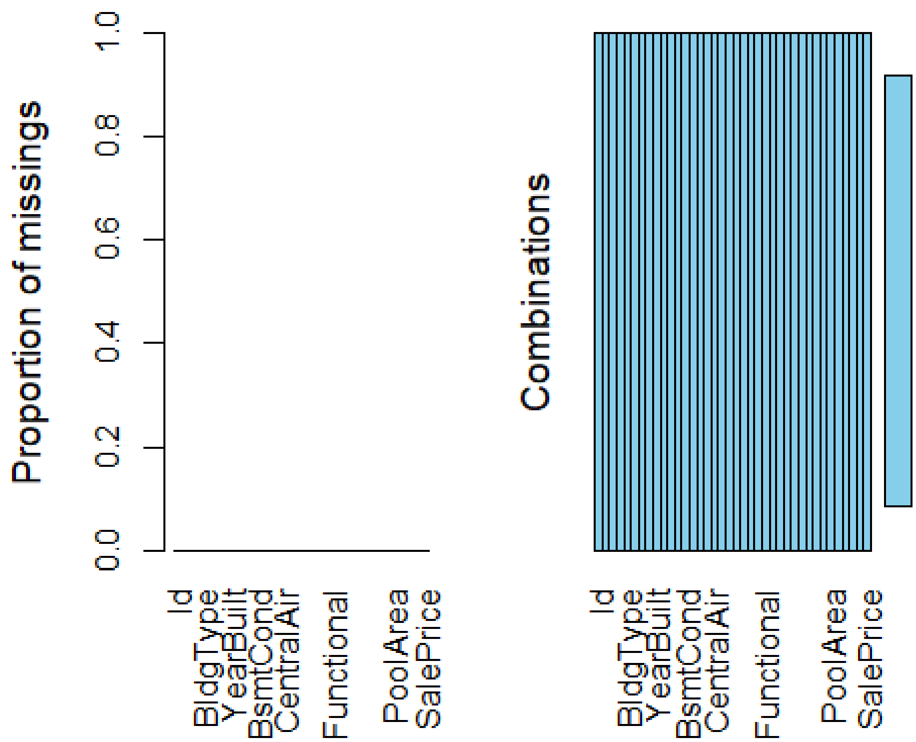
Now we will see is there any duplication in ID column presents in data set?

```
req_var$Id[duplicated(req_var$Id)]     #Extract the duplicate elements from ID
column

## integer(0)

#to remove the duplicate element from id column use following command.
#req_var$Id[!duplicated(req_var$Id)]
```

From this output we can get idea that there is no duplicate element in ID column.

Now we will visualize that still is there any missing value present in this data set? Refrence: https://cran.r-project.org/web/packages/VIMGUI/vignettes/VIM-Imputation.pdf

```
library(VIM)
aggr(req_var)
```

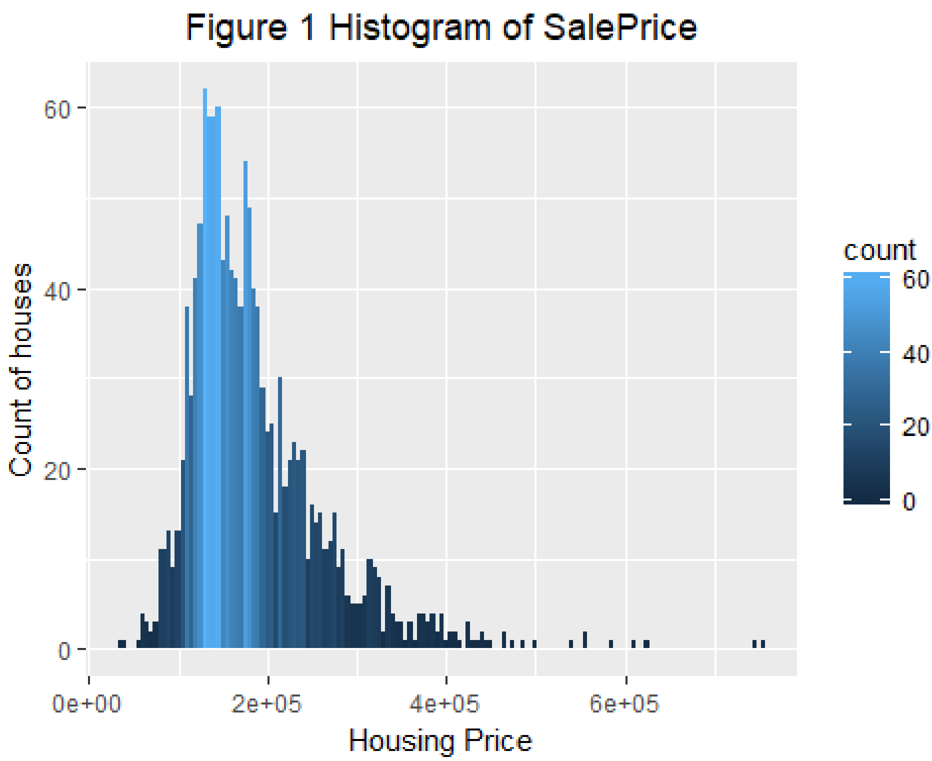


From this output we can say that there is no missing value present in this data set and our data set is clean.

Now we will see some visualization of this data set.

Now, we will see the comparison between the number of house and housing price.

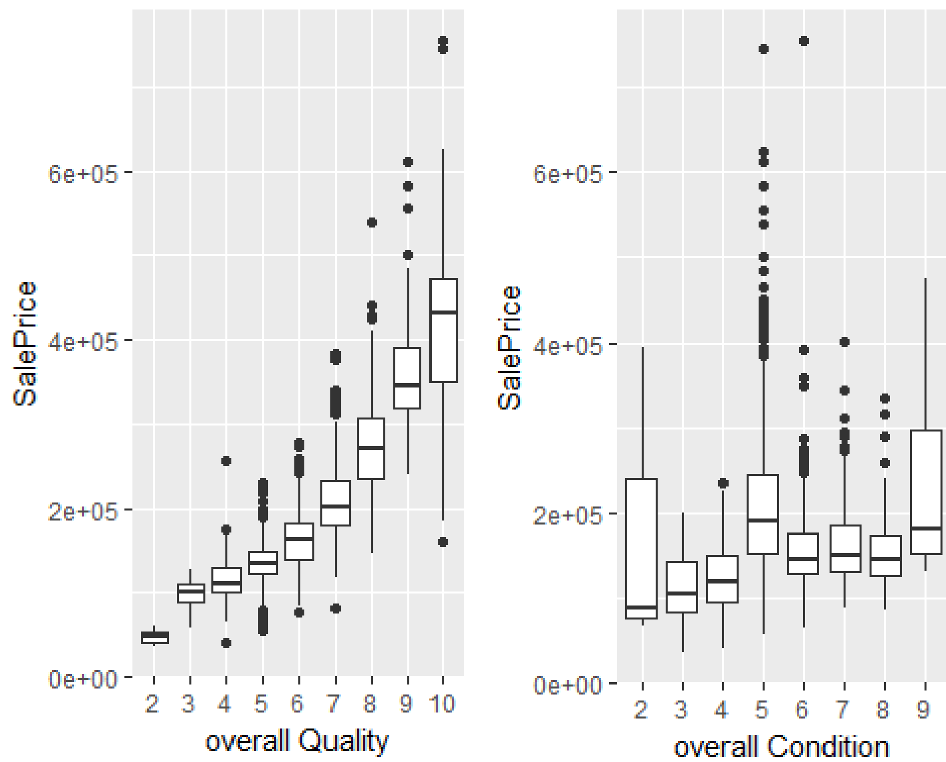
```
library(ggplot2)
ggplot(req_var, aes(x = SalePrice, fill = ..count..)) +
  geom_histogram(binwidth = 5000) +
  ggtitle("Figure 1 Histogram of SalePrice") +
  ylab("Count of houses") +
  xlab("Housing Price") +
  theme(plot.title = element_text(hjust = 0.5))
```



From the above plot, we can say that it is right skewed. We can also say that as the housing price increases, the number of houses decreases. So, there are less number of expensive houses.

Now we will find is there any relationship between price range and overallQuality & price range and OverallCondition.

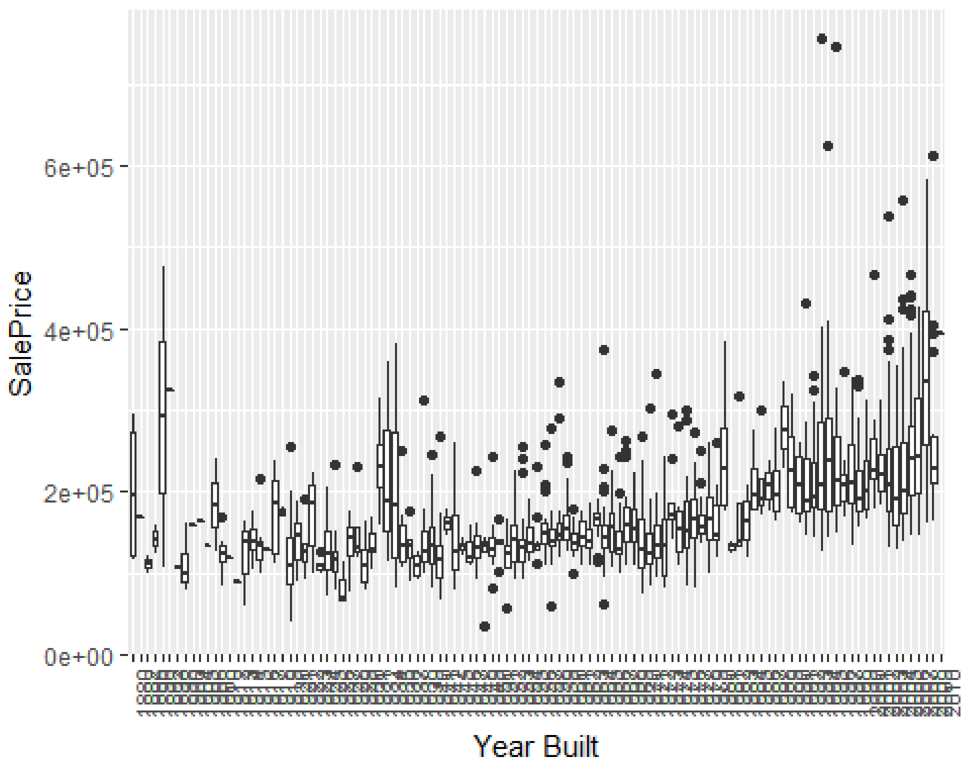
```
library(ggplot2)
library(gridExtra)

b<-ggplot(req_var , aes(x=factor(OverallQual) , y= SalePrice)) +
  geom_boxplot() + labs(x="overall Quality")

library(ggplot2)
 a<-ggplot(req_var , aes(x=factor(OverallCond) , y= SalePrice)) +
  geom_boxplot() + labs(x="overall Condition")

grid.arrange(b,a, widths = c(2,2))
```



Here, we can see that overall quality and sales price are directly proportional i.e. as the overall quality increases, the sale price also increases. However, this is not the case for the overall condition. An overall condition of the house and the sale price shows some strange relation. For condition rating 2, 5 and 9, the prices go high and for the rest of the levels, sale prices are comparatively low.

Now, we will see if there is a relationship between the sale price and the year in which they were built.

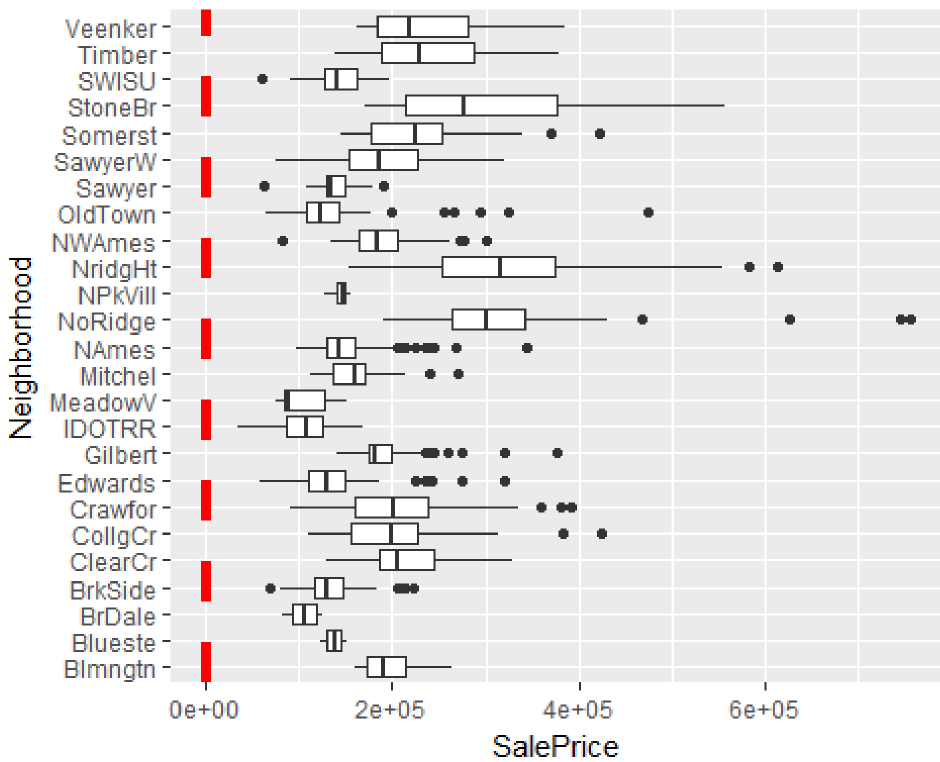
```
library(ggplot2)
ggplot(req_var , aes(x=factor(YearBuilt) , y= SalePrice)) +
  geom_boxplot() + labs(x="Year Built") + theme(axis.text.x = element_text(
size=8, angle=90))
```



From the above plot, we can say that as the time passes, the sale price of houses is increasing. The older houses were cheaper than the newer ones.

Now, we will see if there is a relationship between the sale price and the Neighborhood?

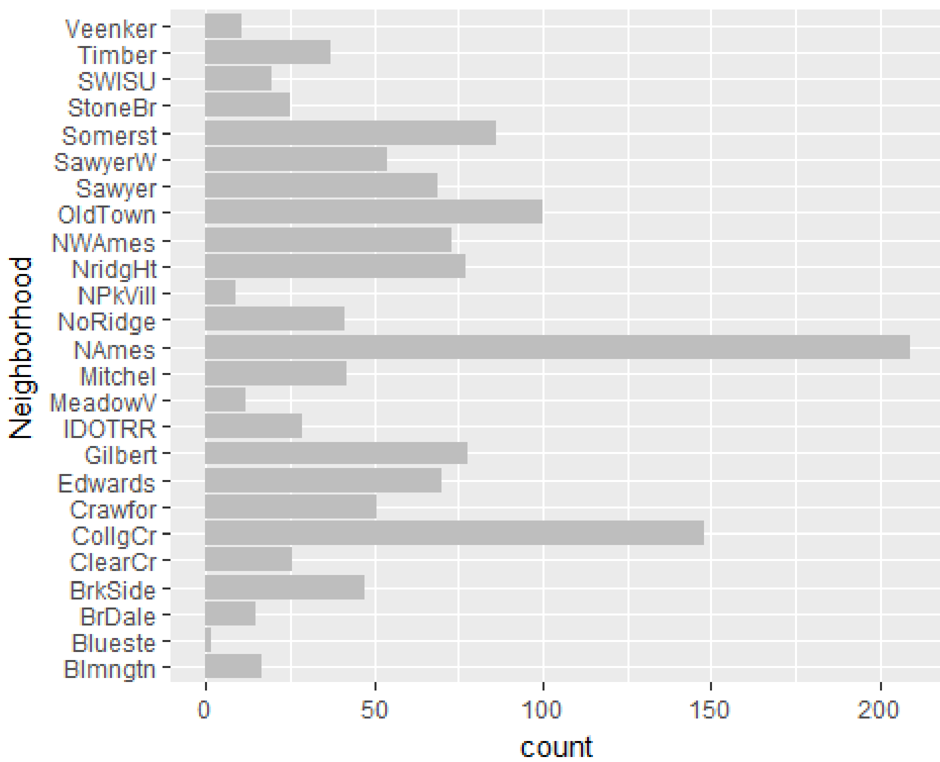
```
ggplot(req_var, aes(x = Neighborhood, y = SalePrice)) +
  geom_boxplot() +
  geom_hline(aes(yintercept=80),
             colour="red", linetype="dashed", lwd=2) +coord_flip()
```



From this plot we can say that the sale price of houses located in Northridge height and North Brook are the highest and the saleprice of houses located in BlueStem neighborhood are the lowest.

Now, we will see comparison among different neighbourhood.

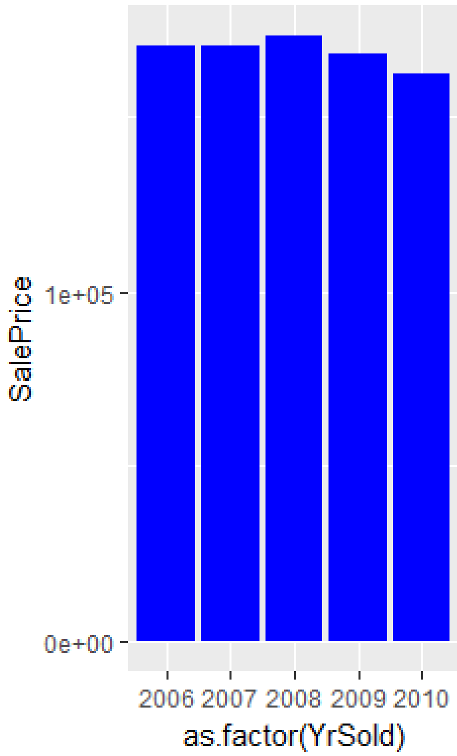
```
library(ggplot2)
ggplot(req_var) +
  geom_bar(aes(x = Neighborhood), fill = "gray") +coord_flip()
```

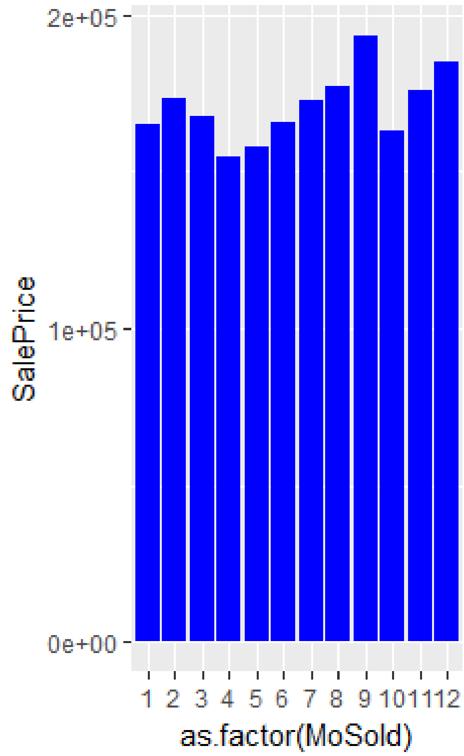


From the above plot, we can say that the maximum number of houses are preffered in the North Ames neighbourhood.

Now, we will see if there is a relationship between the sale price and the year and month in which they were sold.

```
library(gridExtra)
library(ggplot2)
ys<- ggplot(req_var , aes(x= as.factor(YrSold) , y= SalePrice)) +
  geom_bar(stat = "summary" , fun.y= "median" , fill= "blue") +
  scale_y_continuous( breaks = seq(0,200000, 100000))

zs<- ggplot(req_var , aes(x= as.factor(MoSold) , y= SalePrice)) +
  geom_bar(stat = "summary" , fun.y= "median" , fill= "blue") +
  scale_y_continuous( breaks = seq(0,200000, 100000))

grid.arrange(ys,zs, widths= c(2,2))
```



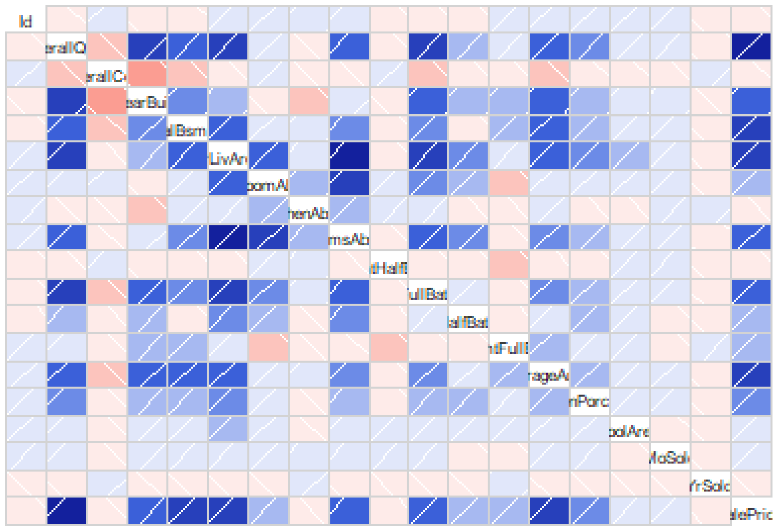


From the above plot, we an say that the sale price does not vary much for the year in which houses were sold but they do differ according to the month in which they were sold. We can see that in september, the the sale price go the highest.

Now, we will find out the correlation of sale price(prediction variable) to the other variables.

```
library("corrgram")
corrgram(req_var, order=NULL, panel=panel.shade, text.panel=panel.txt,
         main="Correlogram")
```

Correlogram



A correlogram shows the relationship among the columns.

From the above plot, we can say that the saleprice has strong correlation with garagearea, overall quality, grlivarea and totalbsmtsf.However, an OverallQual quality has the most effect on the sale price.

Saleprice is highly positively correlated with the overall quality.

```
cor(req_var$OverallQual,req_var$SalePrice)

## [1] 0.784784
```

Saleprice is highly positively correlated with the total basement surface area.

```
cor(req_var$TotalBsmtSF,req_var$SalePrice)

## [1] 0.6009608
```

Saleprice is highly positively correlated with the Above ground living area square feet.

```
cor(req_var$GrLivArea,req_var$SalePrice)

## [1] 0.7097373
```

Saleprice is highly positively correlated with the Garage Area.

```
cor(req_var$GarageArea,req_var$SalePrice)

## [1] 0.6092983
```

Saleprice is medium positively correlated with the Year Built.

```
cor(req_var$YearBuilt,req_var$SalePrice)

## [1] 0.5044485
```

Saleprice is low positively correlated with the open porch surface area.

```
cor(req_var$OpenPorchSF,req_var$SalePrice)

## [1] 0.3264622
```

Saleprice is low positively correlated with the open porch surface area.

```
cor(req_var$OpenPorchSF,req_var$SalePrice)

## [1] 0.3264622
```

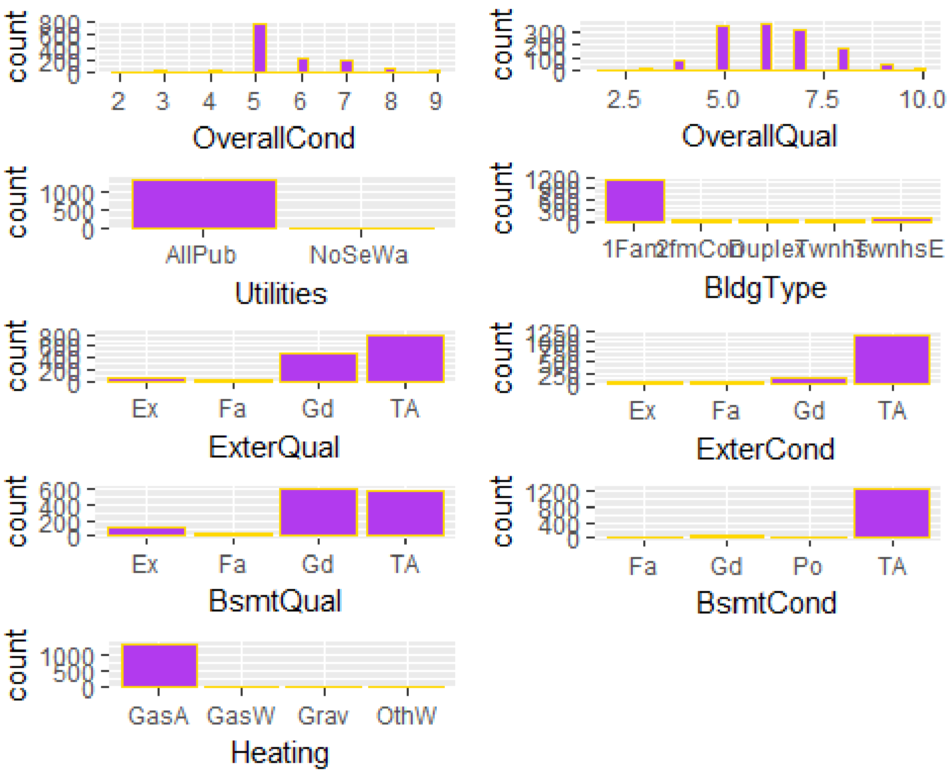
Saleprice is medium negatively correlated with the total rooms above grade (does not include bathrooms) .

```
cor(req_var$TotRmsAbvGrd,req_var$SalePrice)

## [1] 0.5488329
```

Now we will figure out some of informations from below plots.

```
library(ggplot2)
library(gridExtra)
a<-ggplot(req_var) +
  geom_bar(aes(x = Utilities), color="gold", fill="darkorchid2")

b<-ggplot(req_var) +
   geom_histogram(aes(x=OverallQual), color="gold", fill="darkorchid2")

c<-ggplot(req_var) +
   geom_histogram(aes(x=OverallCond), color="gold", fill="darkorchid2")+
scale_x_continuous(breaks=seq(1,10,1))

d<-ggplot(req_var) +
  geom_bar(aes(x = BldgType), color="gold", fill="darkorchid2")

e<-ggplot(req_var) +
  geom_bar(aes(x = ExterQual), color="gold", fill="darkorchid2")

f<-ggplot(req_var) +
  geom_bar(aes(x = ExterCond), color="gold", fill="darkorchid2")

g<-ggplot(req_var) +
  geom_bar(aes(x = BsmtQual), color="gold", fill="darkorchid2")

h<-ggplot(req_var) +
```

```
  geom_bar(aes(x = BsmtCond), color="gold", fill="darkorchid2")

i<-ggplot(req_var) +
  geom_bar(aes(x = Heating), color="gold", fill="darkorchid2")

grid.arrange(c,b,a,d,e,f,g,h,i, widths= c(2,2))

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



From above plots, we can say that houses which have the lowest and the highest overall condition are less as compared with the houses are of medium overall condition.

There are less number of the lowest and the highest overall quality houses than there are medium overall quality houses.

All houses provide all public utilities such as electric, gas, water and septic tank.

A 1 family building types are more than the other building types.

There are more houses with the typical/average External quality while there are less houses which have good external quality.

Most of the houses have typical/average external condition while very few houses which have good external condition.

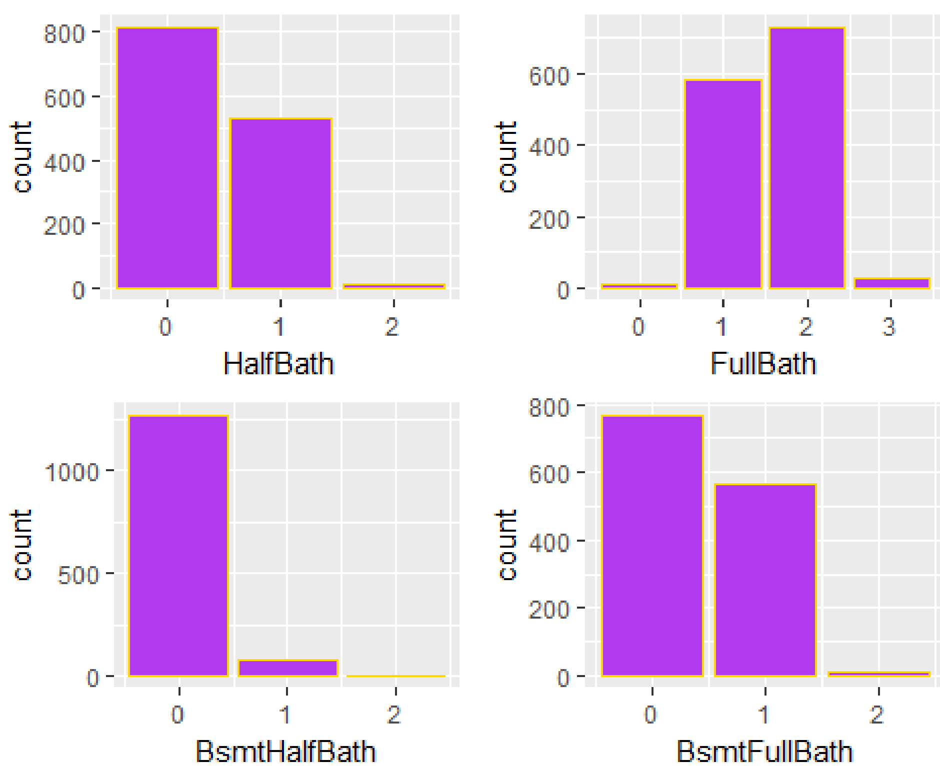
Most of the houses have typical/average and good basement quality while very few houses have excellent basement quality.

All houses have typical/average basement condition.

All houses provide Gas forced warm air furnace(GasA) type of Heating. They did not provide other type of heating in houses.

Now, we will see plots regarding different types of bathrooms.

```
k<-ggplot(req_var) +
  geom_bar(aes(x = HalfBath), color="gold", fill="darkorchid2")

l<-ggplot(req_var) +
  geom_bar(aes(x = FullBath), color="gold", fill="darkorchid2")

m<-ggplot(req_var) +
  geom_bar(aes(x = BsmtHalfBath), color="gold", fill="darkorchid2")

n<-ggplot(req_var) +
  geom_bar(aes(x = BsmtFullBath), color="gold", fill="darkorchid2")


grid.arrange(k,l,m,n, widths= c(2,2))
```



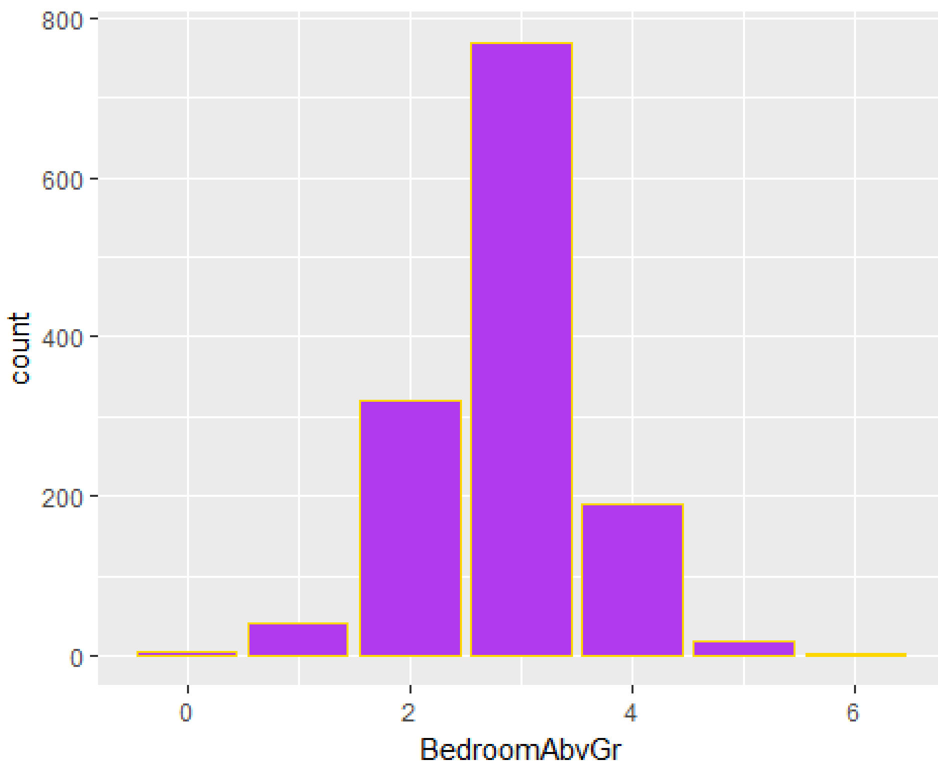
There are less number of houses that have half bath. A half bath includes a toilet and a sink.

Most of the houses include a full bath which includes a toilet, sink, bathtub and a shower.

Most of the houses do not include a basement half bath and comparatively more houses contain basement full bath.

Now we will see the count of bedrooms above basement level

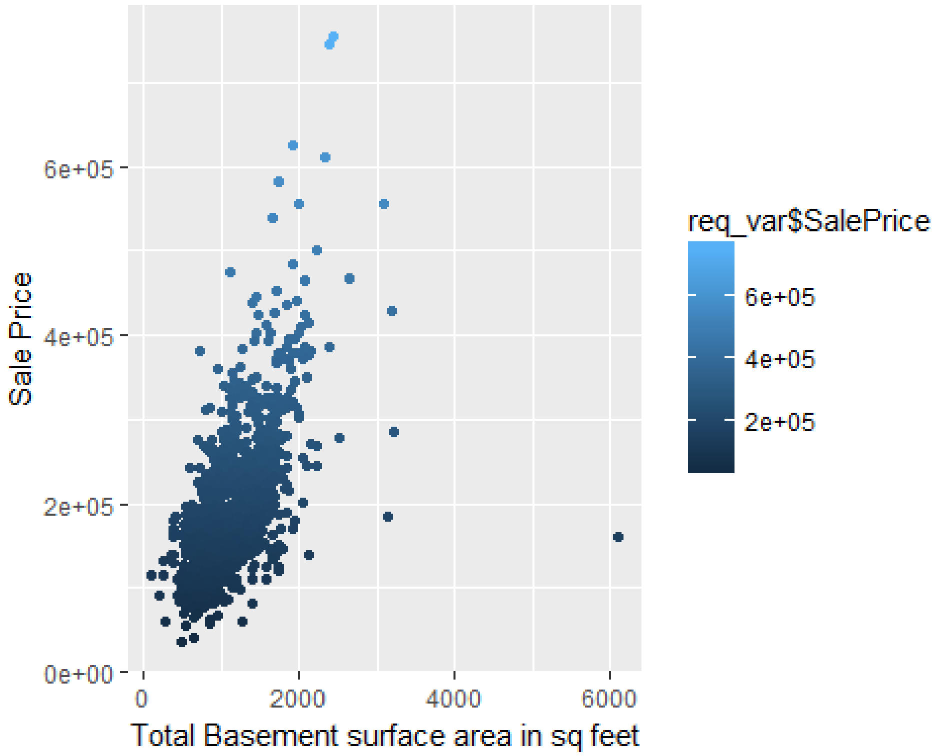
```
i<-ggplot(req_var) +
  geom_bar(aes(x = BedroomAbvGr), color="gold", fill="darkorchid2")
i
```



From this plot we can say that there are more number of houses that have approxinarely 2 to 4 bedrooms above ground.

Now, we will see the relationship between sale price and the total basement surface area.

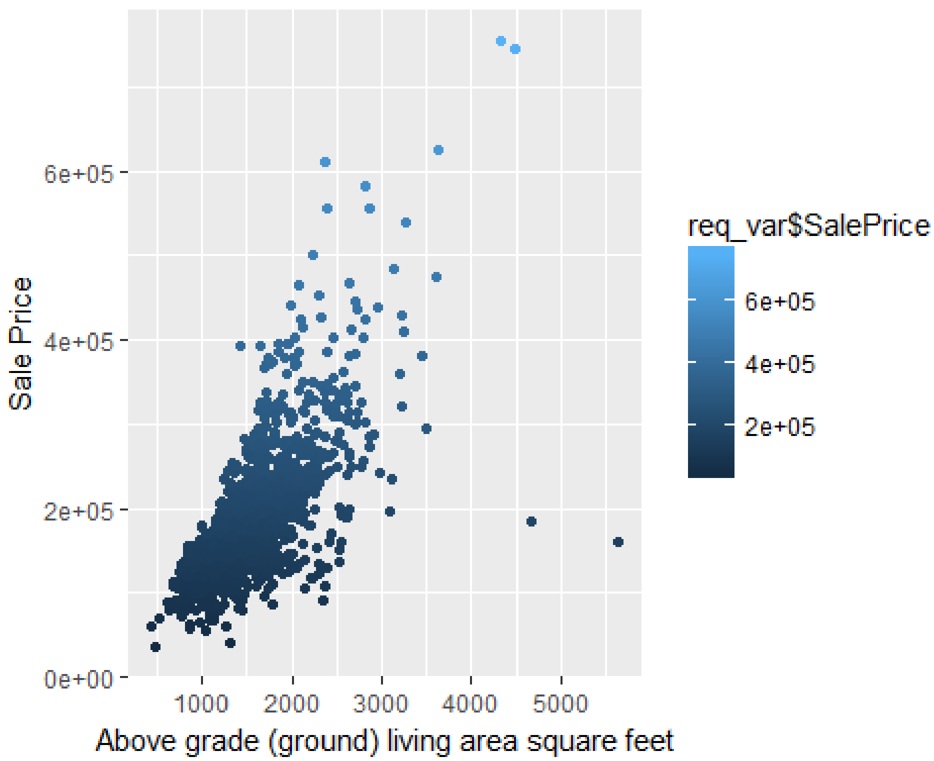
```
library(ggplot2)
ggplot(req_var, aes(req_var$TotalBsmtSF, req_var$SalePrice , color =
req_var$SalePrice)) + geom_point() +    scale_x_continuous(name="Total
Basement surface area in sq feet") + scale_y_continuous( name= "Sale Price")
```



From the aboce plot, we can say that most of the houses have surface area below 2000 sq ft. We can also see that as the surface area increases, the sale price also increases.

Now, we will see the relationship between the sale price and the above ground living area.

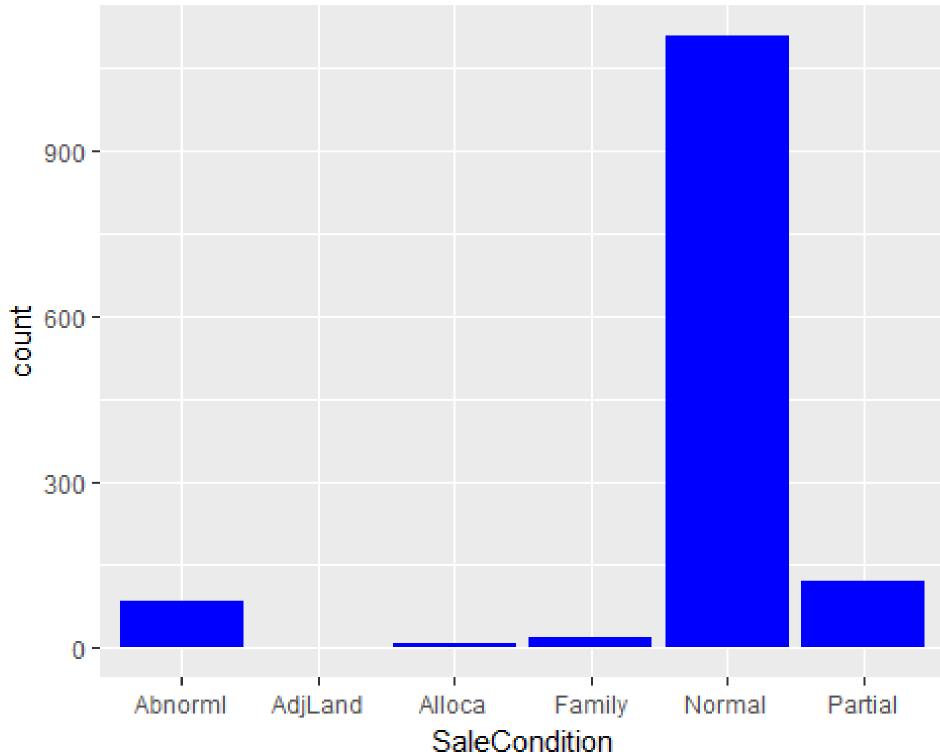
```
library(ggplot2)
ggplot(req_var, aes(req_var$GrLivArea, req_var$SalePrice , color =
req_var$SalePrice)) + geom_point() +   scale_x_continuous(name="Above grade
(ground) living area square feet") + scale_y_continuous( name= "Sale Price")
```



From the above plot, we can say that most of the houses have above ground living area less than approximately 3000 sq ft. As the surface area increases, the sale price also increases. These variables have a linear relationship.

Now we will see the comparison among different sale conditions.

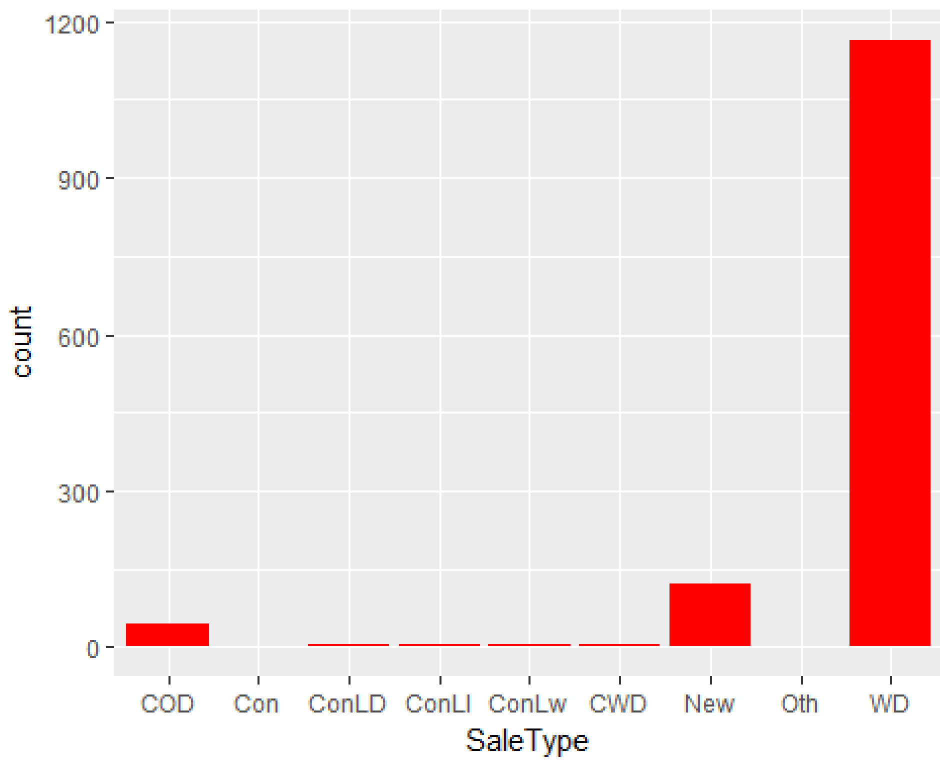
```
library(ggplot2)
ggplot(req_var) +
  geom_bar(aes(x = SaleCondition), fill = "blue")
```



From the above plot, we can say that there are more houses of normal salecondition.

Here, we will see the comparison among different sale types.

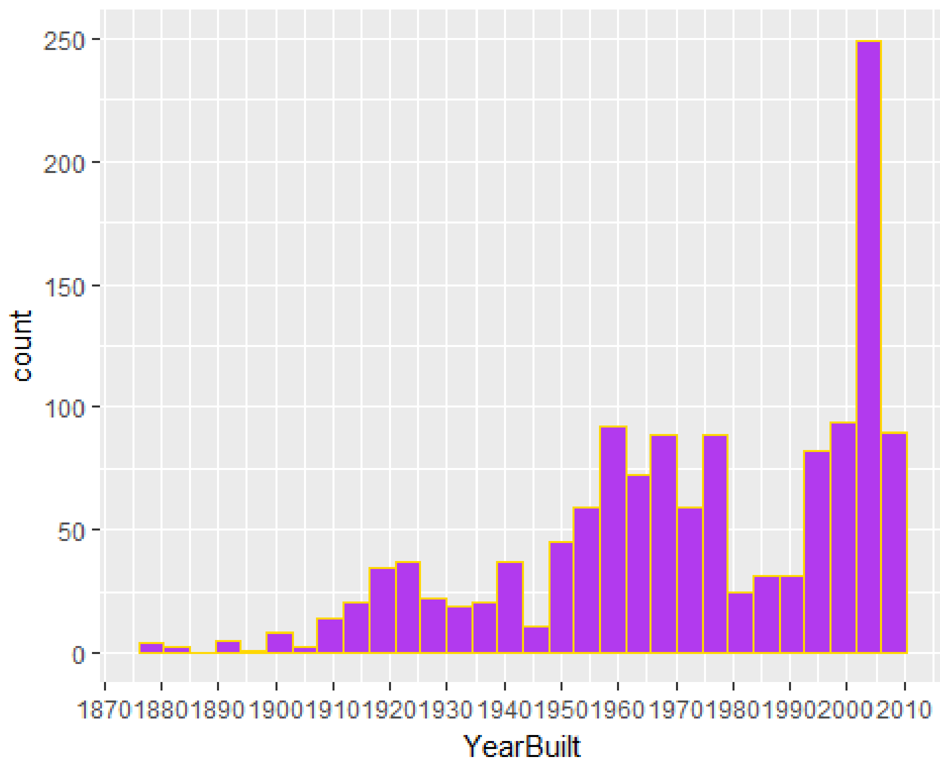
```
library(ggplot2)
ggplot(req_var) +
  geom_bar(aes(x = SaleType), fill = "red")
```



From the above plot, we can see that the warrenty deed houses are more preffered as compared to the other sale types.

Here, we will see that in which year maximum houses were built.

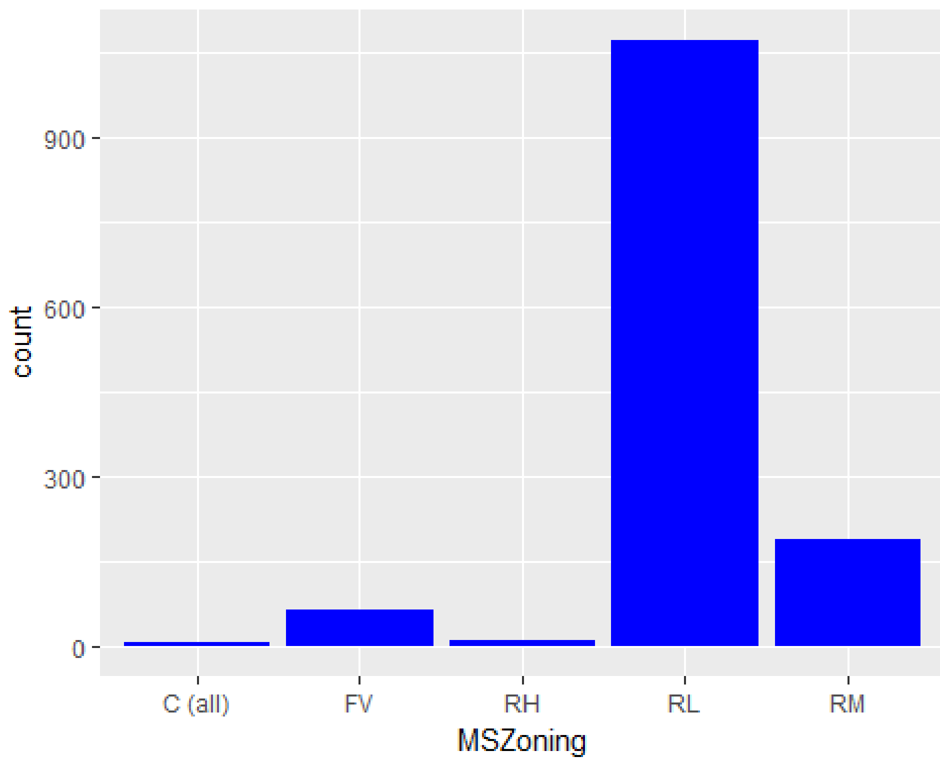
```
ggplot(req_var) +
  geom_histogram(aes(x=YearBuilt), color="gold", fill="darkorchid2") +
scale_x_continuous(breaks=seq(1870,2010,10))

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



From the above plot, as the time went, there were more number of houses built. From approximately 2001 to 2006, there were maximum number of houses built.

Now, we will see comparison among classifications different zoning.

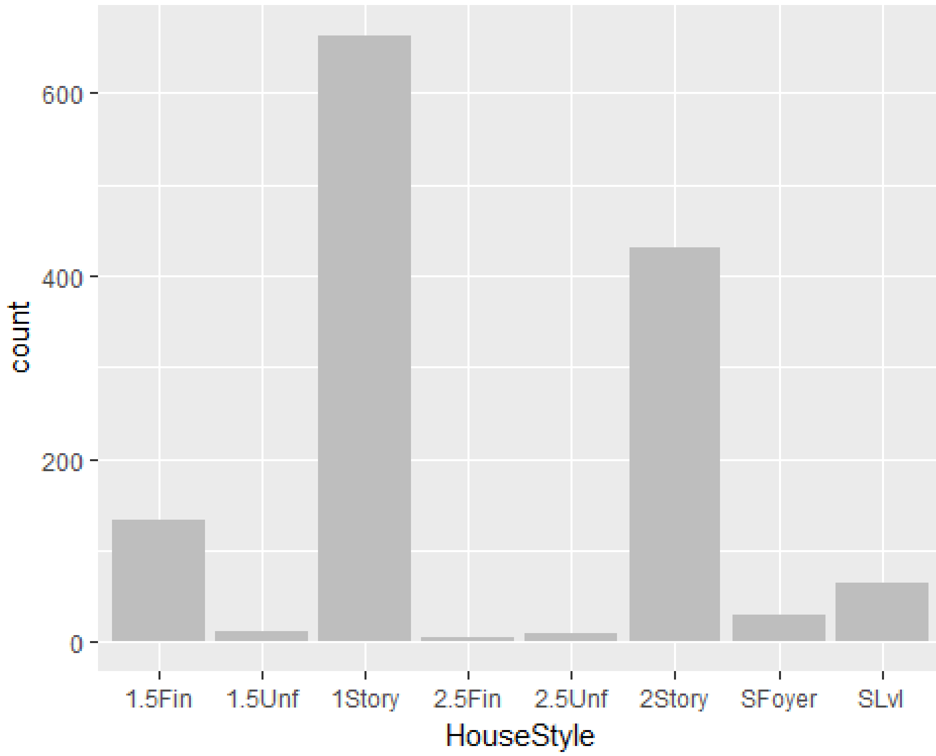
```
library(ggplot2)
ggplot(req_var) +
  geom_bar(aes(x = MSZoning), fill = "blue")
```



From the above plot, we can say that the maximum of the property is of the RL which is the Residential Land.

Here, we will see comparison among different house styles.

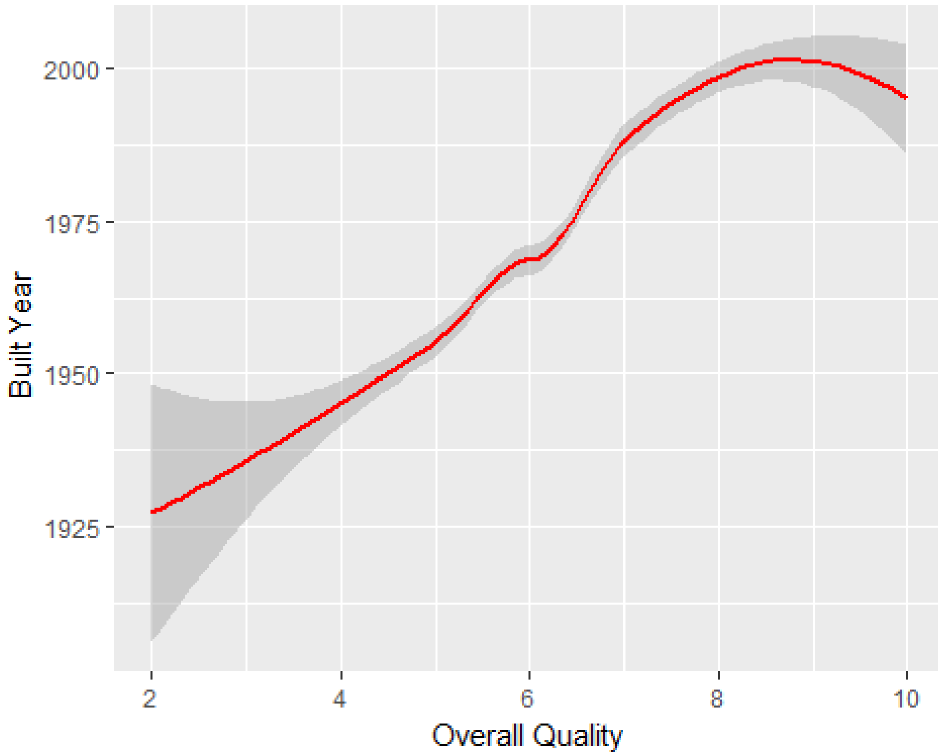
```
library(ggplot2)
ggplot(req_var) +
  geom_bar(aes(x = HouseStyle), fill = "gray")
```



From the above plot, we can say that there are more number of one story buildings than the other house styles.

Here, we will see the comparison between built year and overall quality.

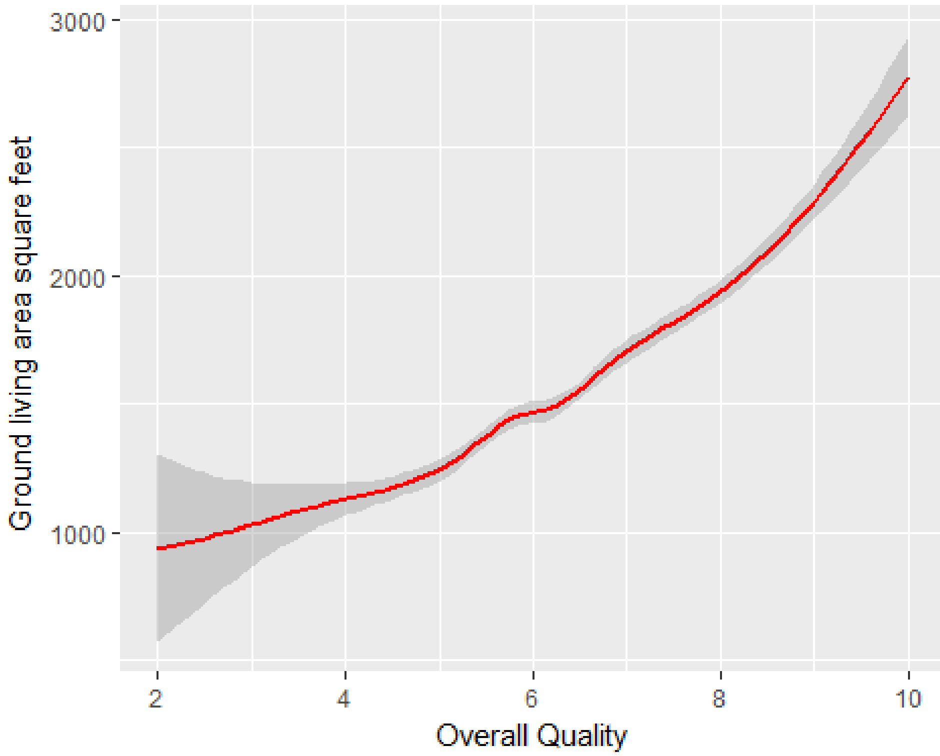
```
library(ggplot2)
a<- ggplot(req_var, aes(x=req_var$OverallQual, y=req_var$YearBuilt)) +
stat_smooth(method = "loess", colour = "red", size = 1) +
   xlab("Overall Quality") + ylab(" Built Year")
a
```



From the above plot, we can see that as the time goes, an overall quality of houses increase.

Here, we will see the comparison between the ground living area square feet and the overall quality.

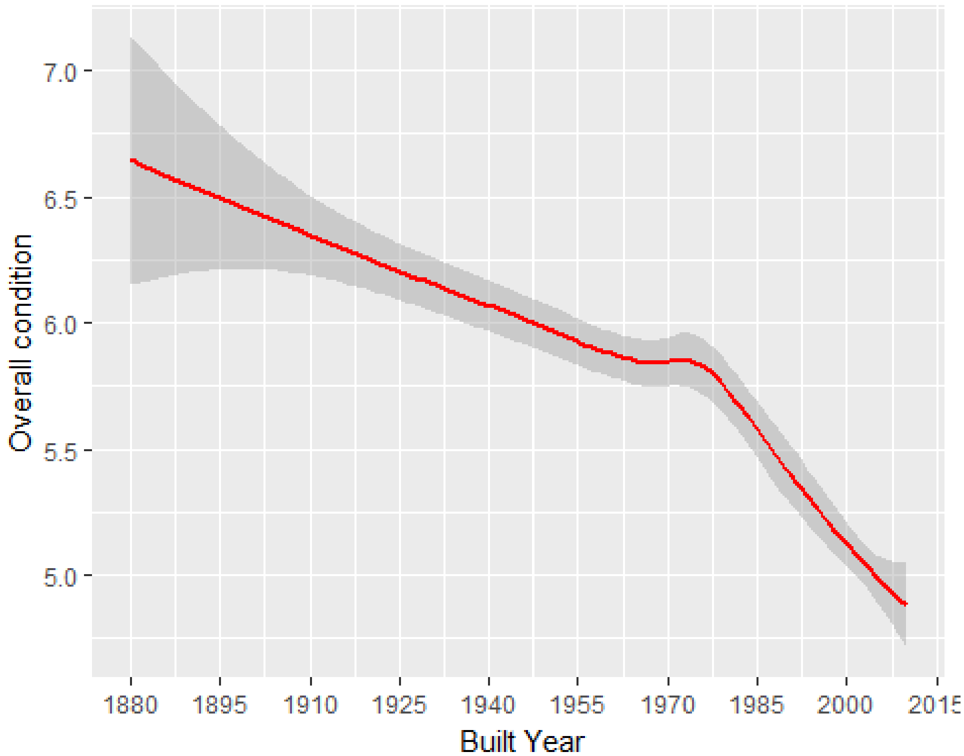
```
library(ggplot2)
a<- ggplot(req_var, aes(x=req_var$OverallQual, y=req_var$GrLivArea)) +
stat_smooth(method = "loess", colour = "red", size = 1) +
   xlab("Overall Quality") + ylab(" Ground living area square feet")
a
```



As, the overall Quality of house is increasing then ground living area square feet also increases.

Here, we will see the comparison between the overall condition and the year built.

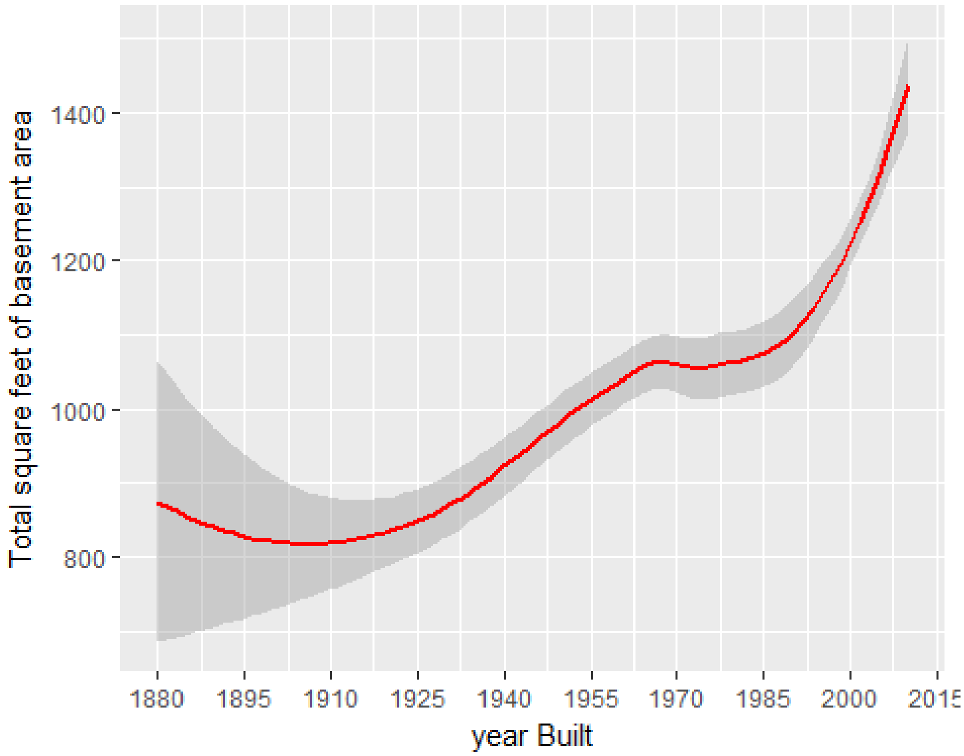
```
library(ggplot2)
a<- ggplot(req_var, aes(y=req_var$OverallCond, x=req_var$YearBuilt)) +
stat_smooth(method = "loess", colour = "red", size = 1) +
   ylab("Overall condition") + xlab(" Built Year") +scale_x_continuous(breaks
= seq(1850, 2015, 15))
a
```



From the above plot, we can see that the condition of the older houses was better than the newly built houses.

Here we will compare the total basement surface area and built area.

```
library(ggplot2)
a<- ggplot(req_var, aes(x=req_var$YearBuilt, y=req_var$TotalBsmtSF)) +
stat_smooth(method = "loess", colour = "red", size = 1) +
   xlab("year Built") + ylab("Total square feet of basement area")
+scale_x_continuous(breaks = seq(1850, 2015, 15))
a
```



As, the time passed, basement surface area increases.

Regression with rpart:

Now we look at using rpart for regression and use the req_var dataset.Our Y variable is SalePrice and we will use YrSold, OverallQual, OverallCond, ExterCond as a our predictors.

Refrence: https://blackboard.umbc.edu/webapps/blackboard/execute/content/file?cmd=view&content_id=_2962653_1&course_id=_37782_1&framesetWrapped=true

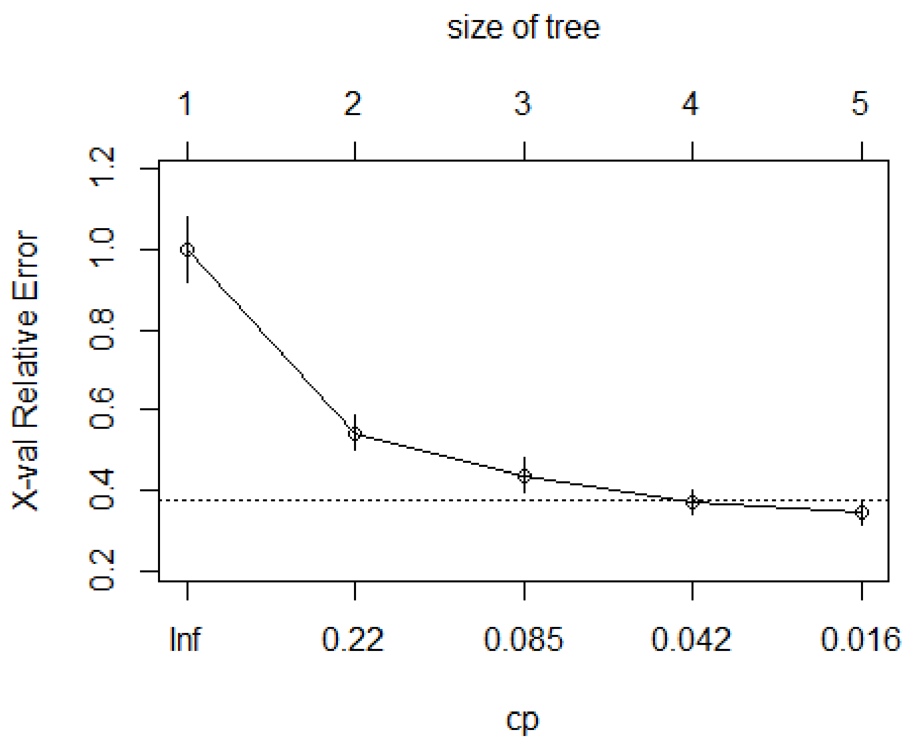
Here, I build regression model

```
library(rpart)
library(rpart.plot)
fit5 <- rpart(SalePrice ~ YrSold + OverallQual +OverallCond+ ExterCond,
method="anova", data=req_var)
printcp(fit5)

##
## Regression tree:
## rpart(formula = SalePrice ~ YrSold + OverallQual + OverallCond +
##     ExterCond, data = req_var, method = "anova")
##
## Variables actually used in tree construction:
## [1] OverallQual
##
## Root node error: 8.4169e+12/1348 = 6.244e+09
##
```

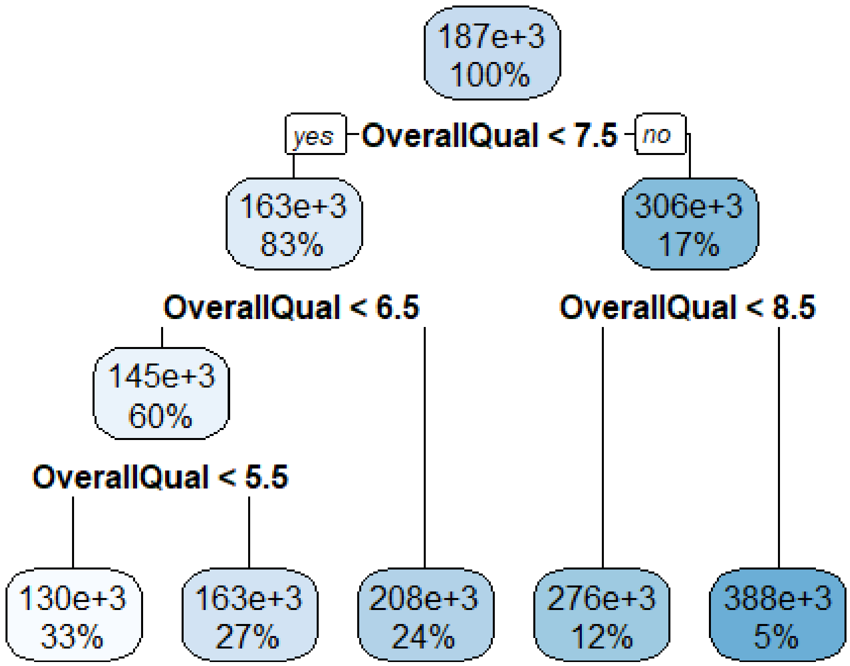
```
## n= 1348
##
##         CP nsplit rel error  xerror     xstd
## 1 0.459404      0   1.00000 1.00126 0.080462
## 2 0.106316      1   0.54060 0.54369 0.044580
## 3 0.067586      2   0.43428 0.43770 0.043289
## 4 0.025565      3   0.36669 0.37109 0.030850
## 5 0.010000      4   0.34113 0.34565 0.030668
```

Visualize cross- validation result

```
plotcp(fit5)
```



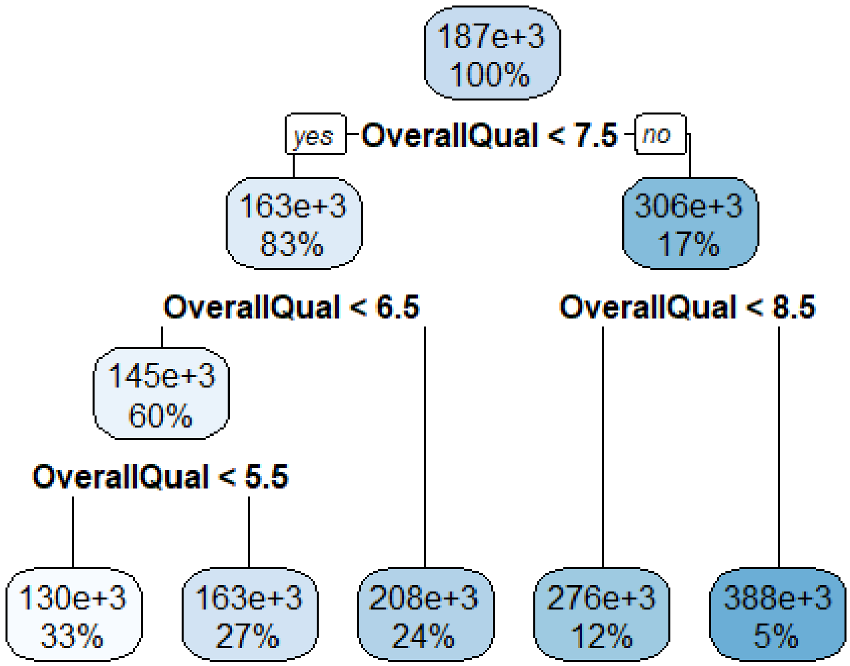

```
rpart.plot(fit5)
```



Here, only 1 predictors used.

Pruning in a regression is similar to that in a classification-we have to pick the Cp corresponding to the smallest xerror.

```
pr.fit5<- prune(fit5, cp =
fit5$cptable[which.min(fit5$cptable[,"xerror"]),"CP"])
rpart.plot(pr.fit5)
```



After Pruning we got the same result which we got before.

```
library(dplyr)
library(magrittr)

req_var11 <- req_var %>% select( YrSold , OverallQual ,OverallCond ,
ExterCond , SalePrice)
```

REGRESSION USING nnet PACKAGE

```
library("Matrix")

library("dplyr")
```

Convert all factors to numerical OR not

```
req_var11 %>% mutate_if(is.factor, as.numeric) -> req_var.num
req_var.num <- req_var11

newdata    <- req_var.num
num.row    <- nrow(newdata)
set.seed(113, "L'Ecuyer")
subData  <- newdata[sample(1:num.row, 1000, replace=FALSE),]

num.row2    <- nrow(subData)
pred.size <- round(0.20 * num.row2)
subData$index <- 1:num.row2
set.seed(113, "L'Ecuyer")
```

```
predData  <- subData[sample(1:num.row2, pred.size, replace=FALSE),]
trainData <- anti_join(subData, predData, by='index')

predData$index
```

```
##   [1] 172 576 621 899 568 677 968 120 921 830 884 439 343 533 368 245 413
##  [18] 959 496 217 704  33 258 390 549 919 284 878 269 322  62 577 350 534
##  [35] 891 908 657 351  27 199 587 887  39 504 692 150 404 904 950 155  37
##  [52] 327 274 144 641 766 798 453 479 485 387 659 733 207 112 791  47 300
##  [69] 332 981 936 963 671 323 802 935 261 913 160 411 361 582 445 658 422
##  [86] 337 752 513 601 713 647 334 295 410 180 821 816 987 651 600 777  97
## [103] 726 608 312 583 338  81 980  29 604 617 788 266 101 618 116 964 736
## [120] 656 393 428  59 131 108 810 461 333 104 541 954  54 575 918 246 939
## [137] 211 934 196 420 646 562 458 609 687 367 942 893 975 206 972 132 239
## [154] 192 473 319 169 863 804 419 873 375 441  22 299 635  45 662 304 307
## [171] 528 154 949 347  12 999 147 668 472 509 220  21 279 349 844 667  19
## [188]  87 696 589 474 508 216 210 885 926 288 491 561 563
```

```
trainData$index
```

```
##   [1]    1    2    3    4    5    6    7    8    9   10   11   13   14
15
##  [15]   16   17   18   20   23   24   25   26   28   30   31   32   34
35
##  [29]   36   38   40   41   42   43   44   46   48   49   50   51   52
53
##  [43]   55   56   57   58   60   61   63   64   65   66   67   68   69
70
##  [57]   71   72   73   74   75   76   77   78   79   80   82   83   84
85
##  [71]   86   88   89   90   91   92   93   94   95   96   98   99  100
102
##  [85]  103  105  106  107  109  110  111  113  114  115  117  118  119
121
##  [99]  122  123  124  125  126  127  128  129  130  133  134  135  136
137
## [113]  138  139  140  141  142  143  145  146  148  149  151  152  153
156
## [127]  157  158  159  161  162  163  164  165  166  167  168  170  171
173
## [141]  174  175  176  177  178  179  181  182  183  184  185  186  187
188
## [155]  189  190  191  193  194  195  197  198  200  201  202  203  204
205
## [169]  208  209  212  213  214  215  218  219  221  222  223  224  225
226
## [183]  227  228  229  230  231  232  233  234  235  236  237  238  240
241
## [197]  242  243  244  247  248  249  250  251  252  253  254  255  256
257
```

```
## [211]   259   260   262   263   264   265   267   268   270   271   272   273   275
276
## [225]   277   278   280   281   282   283   285   286   287   289   290   291   292
293
## [239]   294   296   297   298   301   302   303   305   306   308   309   310   311
313
## [253]   314   315   316   317   318   320   321   324   325   326   328   329   330
331
## [267]   335   336   339   340   341   342   344   345   346   348   352   353   354
355
## [281]   356   357   358   359   360   362   363   364   365   366   369   370   371
372
## [295]   373   374   376   377   378   379   380   381   382   383   384   385   386
388
## [309]   389   391   392   394   395   396   397   398   399   400   401   402   403
405
## [323]   406   407   408   409   412   414   415   416   417   418   421   423   424
425
## [337]   426   427   429   430   431   432   433   434   435   436   437   438   440
442
## [351]   443   444   446   447   448   449   450   451   452   454   455   456   457
459
## [365]   460   462   463   464   465   466   467   468   469   470   471   475   476
477
## [379]   478   480   481   482   483   484   486   487   488   489   490   492   493
494
## [393]   495   497   498   499   500   501   502   503   505   506   507   510   511
512
## [407]   514   515   516   517   518   519   520   521   522   523   524   525   526
527
## [421]   529   530   531   532   535   536   537   538   539   540   542   543   544
545
## [435]   546   547   548   550   551   552   553   554   555   556   557   558   559
560
## [449]   564   565   566   567   569   570   571   572   573   574   578   579   580
581
## [463]   584   585   586   588   590   591   592   593   594   595   596   597   598
599
## [477]   602   603   605   606   607   610   611   612   613   614   615   616   619
620
## [491]   622   623   624   625   626   627   628   629   630   631   632   633   634
636
## [505]   637   638   639   640   642   643   644   645   648   649   650   652   653
654
## [519]   655   660   661   663   664   665   666   669   670   672   673   674   675
676
## [533]   678   679   680   681   682   683   684   685   686   688   689   690   691
693
## [547]   694   695   697   698   699   700   701   702   703   705   706   707   708
709
```

```
## [561]  710  711  712  714  715  716  717  718  719  720  721  722  723
724
## [575]  725  727  728  729  730  731  732  734  735  737  738  739  740
741
## [589]  742  743  744  745  746  747  748  749  750  751  753  754  755
756
## [603]  757  758  759  760  761  762  763  764  765  767  768  769  770
771
## [617]  772  773  774  775  776  778  779  780  781  782  783  784  785
786
## [631]  787  789  790  792  793  794  795  796  797  799  800  801  803
805
## [645]  806  807  808  809  811  812  813  814  815  817  818  819  820
822
## [659]  823  824  825  826  827  828  829  831  832  833  834  835  836
837
## [673]  838  839  840  841  842  843  845  846  847  848  849  850  851
852
## [687]  853  854  855  856  857  858  859  860  861  862  864  865  866
867
## [701]  868  869  870  871  872  874  875  876  877  879  880  881  882
883
## [715]  886  888  889  890  892  894  895  896  897  898  900  901  902
903
## [729]  905  906  907  909  910  911  912  914  915  916  917  920  922
923
## [743]  924  925  927  928  929  930  931  932  933  937  938  940  941
943
## [757]  944  945  946  947  948  951  952  953  955  956  957  958  960
961
## [771]  962  965  966  967  969  970  971  973  974  976  977  978  979
982
## [785]  983  984  985  986  988  989  990  991  992  993  994  995  996
997
## [799]  998 1000
```

```
trueData <- data.frame(predData$SalePrice, predData$index)

# Get rid of index which is the last column
(predData  <-  predData[,-(ncol(predData))])
```

```
##      YrSold OverallQual OverallCond ExterCond SalePrice
## 239    2007           8           5        TA    318000
## 590    2008           5           6        TA     79500
## 222    2009           6           5        TA    200000
## 1134   2009           8           5        TA    239500
## 709    2007           7           5        TA    179540
## 1037   2009           9           5        TA    315500
## 20     2009           5           6        TA    139000
## 990    2006           7           5        TA    197000
```

##	175	2008	6	5	TA	184000
##	414	2010	5	6	TA	115000
##	223	2006	6	6	TA	179900
##	207	2007	5	5	Gd	143900
##	194	2006	7	5	TA	130000
##	894	2008	5	5	TA	165000
##	1437	2007	4	6	TA	120500
##	745	2008	8	5	TA	180000
##	311	2006	6	5	TA	165600
##	565	2006	7	5	TA	268000
##	333	2009	8	5	TA	284000
##	543	2009	7	5	TA	213250
##	837	2007	5	6	TA	153500
##	515	2007	5	5	TA	96500
##	1299	2008	10	5	TA	160000
##	1240	2007	8	5	TA	265900
##	634	2007	5	7	TA	139400
##	73	2009	7	5	TA	185000
##	1205	2006	5	6	TA	153500
##	781	2007	7	5	TA	176000
##	287	2006	6	7	TA	159000
##	1433	2007	4	6	TA	64500
##	978	2007	7	5	TA	199900
##	619	2007	9	5	TA	314813
##	349	2008	7	5	TA	154000
##	483	2009	7	8	TA	155000
##	1215	2006	5	5	TA	134500
##	952	2006	5	5	TA	119900
##	216	2006	5	6	TA	134450
##	377	2006	5	5	TA	148000
##	419	2007	5	6	TA	126000
##	870	2010	7	5	TA	236000
##	1319	2006	8	5	TA	275000
##	58	2006	7	5	TA	196500
##	38	2009	5	6	TA	153000
##	240	2010	6	4	TA	113000
##	644	2007	5	5	TA	152000
##	317	2009	7	7	TA	260000
##	786	2009	6	5	TA	161500
##	1340	2006	5	7	Gd	128500
##	1053	2007	6	6	TA	165000
##	725	2009	9	5	TA	320000
##	965	2010	7	5	TA	214900
##	1060	2007	6	7	TA	220000
##	1320	2007	4	5	TA	111000
##	925	2006	6	6	Gd	207500
##	1420	2006	6	5	Gd	223000
##	1321	2009	6	3	TA	156500
##	1399	2009	5	4	TA	138000
##	1426	2008	6	6	TA	142000

##	108	2008	5	5	TA	115000
##	1086	2010	6	6	TA	147000
##	378	2007	8	5	TA	340000
##	1029	2007	5	5	TA	105000
##	623	2009	5	6	TA	135000
##	1288	2006	5	5	Gd	190000
##	928	2008	7	5	TA	176000
##	500	2007	5	7	TA	120000
##	598	2007	7	5	TA	194201
##	313	2006	5	7	TA	119900
##	992	2009	8	9	Gd	168000
##	956	2007	6	6	TA	145000
##	550	2009	7	5	TA	263000
##	525	2007	7	5	TA	315750
##	1287	2010	6	5	TA	143000
##	1103	2007	5	7	TA	135000
##	258	2009	7	5	TA	220000
##	476	2007	5	6	TA	132500
##	198	2006	8	4	Gd	235000
##	1457	2010	6	6	TA	210000
##	1446	2007	6	5	TA	129000
##	1410	2008	7	6	Gd	215000
##	972	2009	7	5	TA	173000
##	1017	2009	7	5	TA	203000
##	459	2008	8	7	Gd	161000
##	1367	2008	7	5	TA	193000
##	197	2007	7	5	TA	311872
##	708	2009	8	5	TA	254000
##	71	2007	7	6	Gd	244000
##	346	2006	6	5	TA	140200
##	1038	2008	8	5	TA	287000
##	179	2009	9	5	TA	501837
##	1229	2008	9	5	TA	367294
##	1314	2010	9	5	TA	333168
##	879	2010	5	7	TA	148000
##	1318	2007	7	5	TA	208900
##	1424	2006	6	7	Gd	274970
##	284	2009	8	5	TA	244600
##	1439	2010	6	7	TA	149700
##	104	2010	7	5	TA	198900
##	380	2009	6	5	TA	179000
##	866	2009	5	6	TA	148500
##	1304	2006	7	5	TA	232000
##	1225	2008	7	5	TA	184000
##	298	2007	7	5	TA	239000
##	233	2006	6	5	TA	94500
##	315	2006	7	7	TA	178000
##	845	2008	5	8	TA	153900
##	617	2006	6	5	TA	183200
##	1443	2009	10	5	TA	310000

##	705	2010	7	5	TA	213000
##	399	2007	5	2	Fa	67000
##	660	2009	5	7	Gd	167000
##	1236	2006	5	5	TA	138887
##	205	2009	5	7	TA	110000
##	819	2010	6	7	TA	155000
##	1169	2008	6	7	TA	235000
##	1207	2006	4	4	TA	107000
##	930	2006	7	5	TA	222000
##	792	2007	6	5	TA	146800
##	462	2009	7	9	Gd	155000
##	809	2006	5	5	TA	159950
##	386	2010	8	5	TA	192000
##	658	2008	7	6	Fa	149000
##	712	2010	4	7	TA	102776
##	612	2007	6	6	TA	148000
##	122	2007	4	5	TA	100000
##	402	2006	7	5	TA	164990
##	1119	2008	5	6	Gd	140000
##	113	2007	7	5	TA	383970
##	912	2009	5	7	TA	143500
##	835	2008	5	6	TA	139950
##	390	2008	10	5	TA	426000
##	218	2006	4	4	TA	107000
##	21	2006	8	5	TA	325300
##	1027	2010	5	5	TA	167500
##	479	2009	8	5	TA	297000
##	1369	2009	6	5	TA	144000
##	1146	2006	5	6	TA	149000
##	1390	2007	6	6	Gd	131000
##	1246	2007	6	7	TA	178000
##	915	2009	6	5	TA	173733
##	703	2006	8	5	TA	361919
##	42	2007	5	6	Gd	170000
##	1059	2009	9	5	TA	335000
##	98	2007	4	5	TA	94750
##	903	2006	7	5	TA	180000
##	1370	2010	8	5	TA	232000
##	1066	2010	7	5	TA	328000
##	579	2008	7	5	TA	146000
##	1302	2009	6	7	TA	177500
##	913	2006	5	7	TA	88000
##	1081	2008	6	7	Gd	145000
##	78	2008	5	5	TA	127000
##	370	2010	5	7	Gd	162000
##	789	2008	4	7	TA	107900
##	447	2010	6	6	TA	190000
##	1392	2009	5	5	TA	124000
##	469	2007	8	5	TA	250000
##	673	2006	6	6	TA	165000

```
## 1297    2008           5            6        Gd    155000
## 1176    2007           8            5        TA    285000
## 1430    2007           6            7        Gd    182900
## 1168    2006           6            5        TA    173000
## 1269    2008           8            9        Gd    381000
## 46      2010           9            5        TA    319900
## 993     2007           6            8        TA    187000
## 129     2006           6            5        TA    155000
## 1023    2007           5            5        TA     87000
## 278     2010           4            5        TA    141000
## 914     2007           5            6        TA    145000
## 254     2007           6            7        TA    158000
## 407     2008           6            5        TA    115000
## 296     2006           6            6        TA    142500
## 1247    2006           7            5        TA    186500
## 352     2006           7            5        TA    190000
## 646     2007           6            5        TA    143250
## 415     2008           7            5        TA    228000
## 731     2010           8            5        TA    236500
## 1360    2006           9            5        TA    315000
## 829     2009           5            5        TA    185000
## 1441    2008           6            7        TA    191000
## 350     2006           9            5        TA    437154
## 1034    2006           7            5        TA    230000
## 1005    2009           7            5        TA    181000
## 138     2006           7            5        TA    171000
## 417     2006           6            7        TA    149500
## 339     2006           7            7        TA    202500
## 729     2009           5            5        TA    110000
## 1126    2009           4            5        TA    115000
## 416     2007           7            5        TA    181134
## 582     2009           8            5        TA    253293
## 1344    2009           6            6        TA    177000
## 506     2009           5            5        TA    124500
## 593     2008           5            8        Gd    138000
## 388     2009           6            6        TA    125000
## 1015    2007           6            5        TA    119200
## 771     2009           5            5        TA    134900
## 510     2009           5            6        TA    124500
## 441     2009          10            5        TA    555000
## 901     2007           4            6        TA    110000
## 926     2008           5            6        TA    175000
```

```
(trainData <- trainData[,-(ncol(trainData))])
```

```
##      YrSold OverallQual OverallCond ExterCond SalePrice
## 1      2010           5           6        Gd    145000
## 2      2008           6           8        TA    140000
## 3      2007           5           6        TA    125500
## 4      2007           8           5        TA    302000
```

```
## 5     2006      7      5    TA   147400
## 6     2007      6      6    TA   195000
## 7     2008      6      7    TA   207000
## 8     2007      5      6    TA   100000
## 9     2007      7      5    TA   215000
## 10    2006      4      9    Gd   145000
## 11    2006      7      5    TA   162900
## 12    2007      5      7    TA   140000
## 13    2006      6      5    Gd   175900
## 14    2007      8      5    TA   248000
## 15    2006      8      5    TA   280000
## 16    2009      7      6    TA   205000
## 17    2006      3      4    TA   105000
## 18    2006      9      5    TA   342643
## 19    2010      5      6    TA   127000
## 20    2007      5      7    TA   139000
## 21    2009      5      6    TA   223500
## 22    2009      6      9    TA   134900
## 23    2009      5      7    TA   119500
## 24    2007      8      5    TA   251000
## 25    2006      5      3    TA    98600
## 26    2008      6      8    TA   240000
## 27    2010      7      5    TA   233000
## 28    2007      7      5    TA   203000
## 29    2007      4      5    TA   123000
## 30    2009      6      6    TA   158000
## 31    2007      5      5    TA   152000
## 32    2009      5      5    TA   132500
## 33    2009      5      5    Gd   138000
## 34    2008      6      5    TA   180500
## 35    2009      7      5    TA   236000
## 36    2009     10      5    TA   386250
## 37    2006      7      5    TA   200000
## 38    2007      8      5    TA   282922
## 39    2009      8      5    TA   403000
## 40    2010      5      5    TA   110000
## 41    2007      5      7    Gd   132500
## 42    2008      6      7    TA   177000
## 43    2006      5      6    Gd   143000
## 44    2008      6      5    TA   119000
## 45    2007      5      4    TA   186700
## 46    2007      6      7    TA   257500
## 47    2010      6      5    TA   143000
## 48    2008      5      5    TA   118500
## 49    2006      7      5    TA   137500
## 50    2006      9      5    TA   360000
## 51    2007      5      6    TA   128500
## 52    2009      6      8    TA   115000
## 53    2010      7      5    TA   180000
## 54    2006      7      5    TA   199900
```

##	55	2006	6	5	TA	172785
##	56	2010	7	7	Gd	242000
##	57	2006	7	5	TA	235000
##	58	2007	6	6	TA	150000
##	59	2006	3	4	TA	91000
##	60	2006	9	5	TA	374000
##	61	2006	7	5	TA	219210
##	62	2006	7	5	TA	250000
##	63	2008	5	8	TA	119000
##	64	2007	6	5	TA	156932
##	65	2006	5	4	TA	79000
##	66	2010	6	7	Gd	137500
##	67	2007	8	5	TA	248900
##	68	2006	5	6	TA	168000
##	69	2008	7	5	TA	159895
##	70	2007	8	5	TA	209500
##	71	2010	7	8	TA	289000
##	72	2010	8	5	TA	538000
##	73	2006	4	6	TA	106500
##	74	2007	5	7	TA	135000
##	75	2007	6	5	TA	178900
##	76	2007	6	5	TA	113000
##	77	2008	7	5	Gd	210000
##	78	2008	7	5	TA	250000
##	79	2008	7	7	TA	241500
##	80	2006	5	6	TA	132500
##	81	2007	4	7	TA	109900
##	82	2006	7	5	TA	165000
##	83	2010	9	5	TA	611657
##	84	2009	5	7	TA	105000
##	85	2006	5	8	TA	146000
##	86	2010	8	5	TA	212000
##	87	2009	5	6	TA	66500
##	88	2009	6	5	TA	208300
##	89	2006	6	6	Gd	110500
##	90	2007	5	7	TA	144000
##	91	2006	7	7	TA	250000
##	92	2009	7	5	TA	195000
##	93	2006	8	5	TA	258000
##	94	2006	6	5	TA	125000
##	95	2009	8	5	TA	287090
##	96	2006	8	5	TA	348000
##	97	2006	5	7	TA	129900
##	98	2008	5	6	TA	110000
##	99	2009	7	5	TA	191000
##	100	2006	8	5	TA	260000
##	101	2008	6	6	TA	171500
##	102	2007	7	5	TA	183500
##	103	2008	5	8	TA	147000
##	104	2008	7	5	TA	196000

##	105	2009	6	5	TA	147000
##	106	2008	9	5	TA	412500
##	107	2010	7	5	TA	224000
##	108	2009	5	5	TA	109500
##	109	2009	5	8	TA	109500
##	110	2007	7	5	TA	219500
##	111	2010	7	5	TA	206000
##	112	2007	8	5	TA	354000
##	113	2010	4	6	Gd	80000
##	114	2006	5	5	TA	157000
##	115	2009	8	7	TA	345000
##	116	2009	4	6	TA	105000
##	117	2006	5	4	TA	170000
##	118	2006	5	7	TA	123500
##	119	2007	7	5	TA	176485
##	120	2006	5	5	TA	170000
##	121	2008	6	5	TA	153900
##	122	2009	5	5	TA	100000
##	123	2010	7	5	TA	222000
##	124	2009	7	5	TA	180500
##	125	2009	5	7	Gd	147000
##	126	2006	5	4	Fa	160000
##	127	2007	8	5	TA	276000
##	128	2010	7	8	Gd	335000
##	129	2010	5	5	TA	142000
##	130	2008	7	9	Ex	161000
##	131	2006	7	5	TA	246578
##	132	2009	8	5	TA	372500
##	133	2008	5	6	TA	149350
##	134	2009	5	4	TA	110000
##	135	2008	7	5	TA	182000
##	136	2007	8	5	TA	317000
##	137	2009	8	5	TA	187500
##	138	2007	8	5	TA	440000
##	139	2008	6	5	TA	144500
##	140	2007	7	5	TA	179600
##	141	2008	7	5	TA	188000
##	142	2010	6	7	TA	149000
##	143	2006	4	5	TA	118500
##	144	2009	7	5	TA	192500
##	145	2006	6	7	TA	118000
##	146	2009	7	5	TA	194500
##	147	2010	8	5	TA	202900
##	148	2008	4	6	TA	68500
##	149	2007	5	5	TA	122500
##	150	2007	10	5	TA	466500
##	151	2010	5	7	TA	130000
##	152	2009	7	3	TA	197000
##	153	2010	8	5	TA	306000
##	154	2008	7	6	TA	174000

##	155	2006	6	5	TA	155900
##	156	2009	6	5	TA	174000
##	157	2009	5	6	TA	131400
##	158	2008	5	6	TA	228950
##	159	2008	8	5	TA	324000
##	160	2008	5	4	TA	172000
##	161	2006	6	6	TA	119500
##	162	2009	8	5	TA	337000
##	163	2008	5	7	TA	116000
##	164	2007	4	6	TA	119000
##	165	2007	5	5	TA	112500
##	166	2009	7	5	TA	224500
##	167	2007	8	8	TA	315000
##	168	2010	6	5	TA	216500
##	169	2009	7	5	TA	214000
##	170	2010	9	2	TA	394432
##	171	2008	5	6	TA	144000
##	172	2008	5	6	TA	189000
##	173	2009	4	5	TA	82500
##	174	2008	7	5	TA	235000
##	175	2009	5	7	Gd	160000
##	176	2010	5	7	TA	157900
##	177	2007	7	6	TA	228500
##	178	2007	7	5	Gd	167000
##	179	2010	6	5	TA	107000
##	180	2009	8	5	TA	256300
##	181	2009	6	5	TA	174000
##	182	2008	6	5	TA	187000
##	183	2007	6	5	TA	181000
##	184	2008	8	5	TA	237000
##	185	2010	6	5	TA	88000
##	186	2006	6	4	TA	136900
##	187	2008	6	6	TA	165500
##	188	2010	7	5	TA	155000
##	189	2009	6	6	Fa	132500
##	190	2006	7	5	TA	187750
##	191	2008	8	5	TA	286000
##	192	2010	6	5	TA	155000
##	193	2009	5	6	TA	137000
##	194	2009	9	5	TA	315000
##	195	2008	5	6	TA	147500
##	196	2009	7	5	Gd	82500
##	197	2007	7	5	TA	238000
##	198	2008	7	5	TA	183000
##	199	2006	7	5	Gd	229000
##	200	2009	4	7	TA	133000
##	201	2009	7	5	TA	207500
##	202	2006	7	5	TA	183900
##	203	2008	6	5	TA	155000
##	204	2009	6	5	TA	172500

##	205	2006	7	5	TA	320000
##	206	2008	6	8	Gd	142000
##	207	2010	6	7	TA	205000
##	208	2009	6	6	TA	148500
##	209	2006	6	5	TA	158000
##	210	2009	5	6	Fa	141500
##	211	2006	7	6	TA	200100
##	212	2010	5	5	TA	160000
##	213	2009	5	5	TA	159434
##	214	2009	8	5	TA	263435
##	215	2009	7	5	TA	192000
##	216	2009	8	5	TA	278000
##	217	2008	7	9	TA	178400
##	218	2009	8	5	TA	280000
##	219	2009	9	5	TA	582933
##	220	2010	5	8	Gd	138500
##	221	2008	8	5	TA	255900
##	222	2006	7	5	TA	178740
##	223	2008	7	6	TA	205000
##	224	2007	8	5	TA	395000
##	225	2006	5	4	TA	155000
##	226	2006	5	5	TA	160000
##	227	2006	5	5	TA	135960
##	228	2008	5	5	TA	131500
##	229	2007	7	5	TA	200141
##	230	2009	7	5	TA	213000
##	231	2010	8	5	TA	213500
##	232	2006	8	5	TA	264561
##	233	2008	6	7	TA	145000
##	234	2007	5	4	TA	119000
##	235	2006	7	5	TA	167240
##	236	2006	7	5	TA	140000
##	237	2006	7	7	TA	235000
##	238	2007	7	5	TA	185000
##	239	2006	5	5	Gd	172500
##	240	2009	5	3	Fa	60000
##	241	2009	6	5	TA	127500
##	242	2007	8	5	TA	255000
##	243	2009	3	5	TA	95000
##	244	2007	8	5	TA	281213
##	245	2010	6	5	TA	169900
##	246	2010	7	5	TA	196500
##	247	2008	6	7	Gd	256000
##	248	2008	7	5	TA	175000
##	249	2008	6	6	TA	171000
##	250	2006	4	7	TA	121600
##	251	2009	5	5	TA	62383
##	252	2010	6	5	TA	99500
##	253	2008	7	7	TA	116900
##	254	2010	7	7	TA	194500

##	255	2007	5	7	TA	127000
##	256	2009	8	5	TA	274900
##	257	2007	5	6	TA	139000
##	258	2006	5	5	TA	145000
##	259	2007	6	7	Gd	152000
##	260	2006	8	5	TA	233230
##	261	2007	7	6	TA	190000
##	262	2006	7	5	TA	260000
##	263	2008	4	4	TA	108959
##	264	2008	5	5	TA	135000
##	265	2009	5	6	TA	154900
##	266	2009	7	5	TA	188500
##	267	2010	5	7	TA	154000
##	268	2009	4	4	Gd	104900
##	269	2009	8	5	TA	201000
##	270	2006	5	5	TA	123000
##	271	2009	6	5	TA	140000
##	272	2007	6	7	Gd	148000
##	273	2006	6	5	TA	190000
##	274	2009	7	5	TA	174000
##	275	2008	6	5	TA	141000
##	276	2009	5	7	TA	147000
##	277	2008	7	9	Gd	311500
##	278	2007	4	4	TA	108500
##	279	2006	10	5	TA	438780
##	280	2008	6	8	Gd	120500
##	281	2009	8	5	TA	172500
##	282	2006	6	6	TA	181000
##	283	2009	8	5	TA	184100
##	284	2010	5	6	TA	121500
##	285	2007	7	5	TA	160000
##	286	2008	5	5	TA	142600
##	287	2007	7	5	TA	170000
##	288	2009	5	5	TA	112000
##	289	2009	6	6	TA	185000
##	290	2007	5	5	TA	126000
##	291	2009	6	5	TA	215000
##	292	2008	6	5	TA	214000
##	293	2006	6	5	Fa	140000
##	294	2009	5	4	TA	93000
##	295	2009	5	6	TA	125000
##	296	2009	9	5	TA	325000
##	297	2010	7	9	Gd	266500
##	298	2009	6	5	TA	174000
##	299	2006	6	5	TA	130000
##	300	2008	7	5	TA	250580
##	301	2009	7	6	TA	212000
##	302	2007	6	7	Gd	165000
##	303	2006	8	5	TA	255000
##	304	2006	6	5	TA	200000

##	305	2009	7	9	Gd	143000
##	306	2010	5	6	Gd	139000
##	307	2006	8	5	TA	309000
##	308	2006	5	5	TA	133700
##	309	2008	5	5	TA	110000
##	310	2009	5	7	TA	135000
##	311	2007	6	5	Gd	177500
##	312	2007	5	5	TA	144000
##	313	2009	7	7	TA	145000
##	314	2006	8	5	TA	423000
##	315	2006	8	5	TA	281000
##	316	2009	9	5	TA	370878
##	317	2007	8	6	Gd	268000
##	318	2009	6	7	TA	155000
##	319	2008	9	5	TA	446261
##	320	2009	6	5	TA	178000
##	321	2007	7	5	TA	164700
##	322	2007	8	5	TA	285000
##	323	2007	7	5	TA	213500
##	324	2010	6	7	TA	197500
##	325	2007	8	5	TA	147000
##	326	2009	6	6	TA	137500
##	327	2010	5	5	TA	128000
##	328	2006	7	5	TA	241500
##	329	2006	5	6	TA	180000
##	330	2009	4	4	TA	80000
##	331	2008	7	5	TA	191000
##	332	2009	8	5	TA	229456
##	333	2007	5	6	TA	136905
##	334	2007	6	7	TA	277000
##	335	2006	6	8	Fa	137000
##	336	2009	6	6	TA	148500
##	337	2010	6	7	TA	169000
##	338	2007	5	7	TA	119000
##	339	2008	7	5	TA	208500
##	340	2008	8	5	TA	275000
##	341	2009	5	7	Gd	163500
##	342	2006	10	5	Gd	625000
##	343	2008	5	6	TA	95000
##	344	2008	6	9	Gd	174000
##	345	2006	7	7	TA	262500
##	346	2006	7	5	TA	212900
##	347	2009	5	7	Gd	153000
##	348	2008	5	6	TA	130250
##	349	2010	7	6	TA	140000
##	350	2009	6	5	TA	196000
##	351	2006	5	6	TA	217500
##	352	2010	5	7	Gd	119000
##	353	2009	4	6	TA	113000
##	354	2007	5	6	TA	139000

##	355	2006	5	5	TA	188000
##	356	2008	6	6	Gd	151000
##	357	2008	6	5	TA	193000
##	358	2010	5	6	TA	132000
##	359	2008	6	5	TA	157000
##	360	2006	6	5	TA	140000
##	361	2007	10	5	TA	745000
##	362	2009	8	5	TA	340000
##	363	2008	7	5	TA	165400
##	364	2007	6	5	TA	211000
##	365	2009	4	6	TA	120500
##	366	2009	7	9	TA	234000
##	367	2006	7	6	TA	226000
##	368	2006	3	8	Gd	126175
##	369	2007	7	6	TA	197000
##	370	2006	7	5	TA	194000
##	371	2007	8	5	TA	305900
##	372	2008	6	5	TA	206900
##	373	2009	7	5	TA	265000
##	374	2006	6	5	TA	140000
##	375	2009	4	3	TA	112000
##	376	2007	9	5	TA	285000
##	377	2007	7	5	TA	302000
##	378	2007	7	4	TA	122000
##	379	2008	8	5	TA	293077
##	380	2009	5	7	TA	164900
##	381	2007	6	5	Gd	204750
##	382	2010	7	5	TA	222500
##	383	2006	8	5	TA	197000
##	384	2006	8	5	TA	275000
##	385	2009	6	3	TA	76000
##	386	2006	6	5	TA	185000
##	387	2007	5	6	TA	87000
##	388	2007	7	7	TA	197900
##	389	2008	6	7	Gd	136000
##	390	2007	6	5	TA	194700
##	391	2008	5	7	TA	139000
##	392	2006	7	5	TA	239799
##	393	2006	7	5	TA	280000
##	394	2008	6	7	TA	174900
##	395	2009	6	6	TA	120000
##	396	2010	6	5	TA	143000
##	397	2006	8	5	TA	266000
##	398	2010	6	5	TA	146000
##	399	2006	6	8	Gd	117000
##	400	2006	8	5	TA	312500
##	401	2007	5	5	TA	130000
##	402	2008	7	5	TA	179400
##	403	2007	8	5	TA	318000
##	404	2010	6	6	TA	160000

##	405	2009	7	8	Gd	167500
##	406	2007	6	5	Gd	129000
##	407	2010	7	5	TA	189000
##	408	2009	10	5	TA	402861
##	409	2009	6	6	TA	214500
##	410	2007	7	5	TA	301000
##	411	2010	6	6	TA	171000
##	412	2009	5	6	TA	124900
##	413	2006	7	5	TA	192500
##	414	2007	5	5	TA	132000
##	415	2006	8	5	TA	178000
##	416	2006	5	5	TA	132500
##	417	2010	4	5	TA	83500
##	418	2009	7	5	TA	157000
##	419	2009	2	5	TA	60000
##	420	2009	6	5	TA	165000
##	421	2009	5	4	TA	85000
##	422	2006	7	5	TA	230000
##	423	2008	5	6	TA	120500
##	424	2006	5	5	TA	115000
##	425	2008	4	6	TA	135000
##	426	2010	7	6	TA	237000
##	427	2010	7	5	TA	185500
##	428	2008	5	5	TA	73000
##	429	2010	6	6	TA	153000
##	430	2008	8	5	TA	245000
##	431	2008	7	7	TA	242000
##	432	2006	6	5	TA	142500
##	433	2007	7	5	TA	280000
##	434	2008	6	7	TA	235000
##	435	2008	7	5	Gd	206300
##	436	2009	6	5	TA	181000
##	437	2009	6	6	Gd	132000
##	438	2007	7	6	TA	175500
##	439	2008	5	7	TA	147000
##	440	2006	7	7	TA	163000
##	441	2010	5	6	TA	144500
##	442	2006	8	5	TA	275500
##	443	2007	8	5	TA	424870
##	444	2010	5	7	TA	180500
##	445	2009	7	5	TA	219500
##	446	2008	6	5	TA	163000
##	447	2009	5	5	TA	117000
##	448	2009	6	6	TA	137450
##	449	2009	6	5	TA	159000
##	450	2007	6	6	TA	177000
##	451	2010	6	6	TA	190000
##	452	2009	7	5	TA	275000
##	453	2007	6	5	TA	176000
##	454	2007	5	5	TA	171000

##	455	2007	5	7	TA	194000
##	456	2006	5	5	TA	131000
##	457	2008	5	5	TA	110000
##	458	2010	5	4	TA	131500
##	459	2009	5	7	Gd	197500
##	460	2010	6	5	TA	178000
##	461	2009	8	5	TA	220000
##	462	2009	4	5	TA	91000
##	463	2008	5	6	TA	115000
##	464	2007	7	7	TA	139400
##	465	2007	5	7	TA	132000
##	466	2006	6	8	TA	110000
##	467	2009	9	5	TA	377500
##	468	2008	4	5	TA	137500
##	469	2006	8	7	TA	230000
##	470	2007	7	5	TA	179000
##	471	2010	5	6	TA	129000
##	472	2007	6	5	TA	128000
##	473	2008	8	5	TA	232000
##	474	2010	6	5	TA	152000
##	475	2009	8	8	TA	316600
##	476	2009	7	5	TA	260000
##	477	2009	8	5	TA	261500
##	478	2007	5	5	TA	135000
##	479	2008	5	5	TA	129500
##	480	2008	4	6	TA	256000
##	481	2006	5	6	TA	136500
##	482	2008	7	5	TA	222500
##	483	2006	4	5	TA	116050
##	484	2007	6	3	TA	163500
##	485	2008	5	7	TA	127000
##	486	2007	4	6	TA	129500
##	487	2010	5	5	TA	55993
##	488	2006	7	5	TA	216837
##	489	2007	6	5	TA	175000
##	490	2006	8	5	TA	319000
##	491	2008	5	7	TA	162900
##	492	2007	6	5	TA	211000
##	493	2008	4	5	TA	176000
##	494	2008	6	5	TA	175000
##	495	2008	6	5	TA	168500
##	496	2009	7	5	TA	230000
##	497	2009	7	7	Gd	205000
##	498	2007	9	5	TA	415298
##	499	2007	6	7	TA	243000
##	500	2007	5	6	TA	109000
##	501	2007	7	5	TA	207500
##	502	2006	6	5	TA	83000
##	503	2006	6	7	TA	230000
##	504	2006	6	5	TA	180000

##	505	2009	5	6	TA	154000
##	506	2006	6	5	TA	150900
##	507	2008	7	5	TA	341000
##	508	2008	6	5	TA	193500
##	509	2006	6	6	TA	179900
##	510	2009	5	6	TA	117500
##	511	2010	6	5	TA	177500
##	512	2006	5	5	TA	135500
##	513	2006	7	5	TA	202500
##	514	2006	6	5	TA	163990
##	515	2008	9	5	TA	325000
##	516	2006	5	7	TA	115000
##	517	2006	7	5	TA	179665
##	518	2010	8	5	TA	272000
##	519	2007	7	5	TA	133000
##	520	2007	8	5	TA	270000
##	521	2008	7	6	TA	205950
##	522	2006	6	5	Gd	153500
##	523	2006	5	3	TA	91500
##	524	2008	6	5	TA	91000
##	525	2007	7	5	TA	259000
##	526	2008	6	6	TA	162500
##	527	2010	5	5	TA	165500
##	528	2006	7	5	TA	171750
##	529	2008	5	7	TA	109900
##	530	2009	6	8	TA	118000
##	531	2007	7	8	Gd	159500
##	532	2008	5	6	TA	112500
##	533	2009	6	6	TA	124000
##	534	2008	7	6	TA	201800
##	535	2007	7	5	TA	225000
##	536	2006	5	7	TA	150000
##	537	2008	6	5	TA	148000
##	538	2006	10	5	TA	465000
##	539	2008	6	5	TA	180000
##	540	2010	8	5	TA	410000
##	541	2008	7	5	TA	284000
##	542	2009	8	5	TA	257000
##	543	2006	6	5	TA	187500
##	544	2006	7	5	TA	144152
##	545	2007	5	7	TA	124500
##	546	2007	7	5	TA	163900
##	547	2007	6	3	TA	200624
##	548	2006	6	5	TA	165150
##	549	2007	5	8	Gd	218000
##	550	2010	5	6	TA	128200
##	551	2008	5	7	TA	124900
##	552	2008	7	5	TA	262280
##	553	2009	8	6	TA	227000
##	554	2007	5	5	TA	190000

##	555	2010	6	5	TA	186000
##	556	2007	5	8	TA	129000
##	557	2006	8	5	TA	239900
##	558	2009	5	5	TA	190000
##	559	2008	7	5	TA	191000
##	560	2006	7	5	TA	140000
##	561	2010	6	5	TA	158900
##	562	2006	8	5	TA	290000
##	563	2008	7	5	TA	207000
##	564	2010	7	5	TA	215200
##	565	2010	6	4	TA	128000
##	566	2009	7	5	TA	204000
##	567	2007	8	5	TA	267000
##	568	2009	5	6	TA	139000
##	569	2006	5	8	Gd	225000
##	570	2007	5	6	TA	175000
##	571	2008	6	5	TA	149000
##	572	2009	4	5	TA	103000
##	573	2007	7	8	TA	189950
##	574	2009	6	6	TA	121000
##	575	2010	7	5	TA	190000
##	576	2007	5	5	TA	169500
##	577	2009	6	6	TA	143750
##	578	2009	4	8	Gd	86000
##	579	2009	8	5	TA	252000
##	580	2006	5	4	TA	149900
##	581	2009	4	6	TA	128000
##	582	2010	4	4	Fa	80000
##	583	2007	9	5	TA	380000
##	584	2006	7	4	TA	225000
##	585	2008	8	5	TA	372402
##	586	2008	6	8	TA	128000
##	587	2006	7	6	TA	274300
##	588	2008	6	5	TA	85400
##	589	2007	5	6	TA	150750
##	590	2009	4	7	Gd	124500
##	591	2006	4	6	TA	108000
##	592	2006	5	5	TA	138500
##	593	2009	4	4	TA	150000
##	594	2007	8	5	TA	270000
##	595	2007	6	8	Gd	151500
##	596	2006	5	5	TA	141000
##	597	2008	5	5	TA	127000
##	598	2009	8	5	TA	236500
##	599	2006	7	5	TA	202665
##	600	2007	4	5	Fa	118500
##	601	2008	6	5	TA	176000
##	602	2007	7	5	TA	148800
##	603	2007	8	5	TA	303477
##	604	2010	6	6	TA	148000

##	605	2007	5	7	Gd	119750
##	606	2008	5	5	TA	134000
##	607	2007	6	8	TA	259500
##	608	2006	5	6	TA	125500
##	609	2007	6	6	TA	155000
##	610	2008	6	8	TA	151000
##	611	2008	5	8	TA	149900
##	612	2007	8	5	TA	174000
##	613	2009	5	6	TA	110000
##	614	2009	7	7	TA	271900
##	615	2009	7	5	TA	187500
##	616	2006	5	5	TA	129900
##	617	2006	8	5	TA	215000
##	618	2009	10	5	TA	337500
##	619	2009	5	5	TA	106000
##	620	2008	6	7	TA	188700
##	621	2006	4	5	TA	129000
##	622	2007	7	5	TA	226000
##	623	2008	5	7	Gd	125000
##	624	2010	6	7	TA	158000
##	625	2008	10	5	TA	385000
##	626	2007	6	5	TA	179200
##	627	2010	5	6	TA	142125
##	628	2006	7	5	TA	180000
##	629	2009	4	6	TA	135000
##	630	2009	5	6	TA	130000
##	631	2007	5	5	TA	134000
##	632	2006	7	6	TA	170000
##	633	2009	6	5	TA	173500
##	634	2007	7	5	TA	277000
##	635	2008	6	5	TA	193000
##	636	2007	5	6	TA	90350
##	637	2007	9	5	TA	277500
##	638	2007	7	5	TA	226000
##	639	2009	6	6	TA	127500
##	640	2007	7	5	TA	192000
##	641	2006	4	5	TA	102000
##	642	2006	5	6	TA	124000
##	643	2007	5	7	TA	134500
##	644	2007	4	5	TA	75000
##	645	2009	6	7	TA	185750
##	646	2006	4	2	TA	87000
##	647	2009	7	8	Gd	135000
##	648	2010	6	8	TA	130500
##	649	2008	5	7	Gd	135500
##	650	2008	6	5	TA	208900
##	651	2007	7	5	Gd	187500
##	652	2007	6	7	Gd	136000
##	653	2008	5	5	TA	135000
##	654	2009	7	5	TA	189000

##	655	2008	6	5	TA	188000
##	656	2008	6	5	TA	155000
##	657	2006	7	5	TA	221500
##	658	2008	5	5	TA	98000
##	659	2009	7	7	TA	174500
##	660	2007	8	5	TA	290000
##	661	2007	4	4	Gd	168000
##	662	2010	4	8	Gd	125500
##	663	2009	6	5	TA	231500
##	664	2007	5	5	TA	128500
##	665	2008	7	6	Gd	184000
##	666	2006	6	6	TA	159500
##	667	2007	6	6	TA	180500
##	668	2008	5	6	TA	144000
##	669	2007	7	6	TA	227875
##	670	2008	6	6	TA	163000
##	671	2006	6	5	TA	164000
##	672	2007	6	6	TA	175000
##	673	2009	7	5	TA	152000
##	674	2006	5	7	TA	149900
##	675	2007	5	7	TA	146500
##	676	2007	5	5	TA	155000
##	677	2009	6	4	TA	149500
##	678	2008	6	5	TA	129500
##	679	2006	5	5	TA	145000
##	680	2009	8	5	TA	278000
##	681	2009	6	5	TA	175000
##	682	2008	7	5	TA	190000
##	683	2006	9	5	TA	385000
##	684	2010	4	4	Gd	68400
##	685	2007	5	5	TA	106000
##	686	2009	5	5	TA	145000
##	687	2007	5	5	TA	108000
##	688	2009	5	8	TA	91300
##	689	2009	6	8	Gd	185000
##	690	2008	6	6	TA	140000
##	691	2010	6	5	TA	125000
##	692	2010	6	5	TA	189000
##	693	2008	5	5	TA	113000
##	694	2008	5	8	Gd	179900
##	695	2006	7	5	TA	195000
##	696	2009	7	5	TA	253000
##	697	2009	6	6	TA	162000
##	698	2008	7	5	TA	224900
##	699	2008	6	7	TA	224000
##	700	2008	8	9	Gd	299800
##	701	2009	6	5	TA	165000
##	702	2006	7	5	TA	176000
##	703	2007	5	7	Gd	132000
##	704	2007	6	5	Gd	217000

##	705	2006	5	8	TA	170000
##	706	2009	7	5	TA	173000
##	707	2006	4	5	Gd	79900
##	708	2006	8	6	TA	228000
##	709	2008	6	5	TA	139000
##	710	2009	7	5	TA	227000
##	711	2009	5	6	TA	123000
##	712	2007	6	5	TA	131500
##	713	2010	6	7	TA	161750
##	714	2006	7	7	Gd	239000
##	715	2009	5	7	TA	133000
##	716	2009	6	6	TA	159500
##	717	2007	4	7	TA	89471
##	718	2008	7	5	TA	227680
##	719	2006	6	5	TA	145250
##	720	2009	5	7	TA	130000
##	721	2009	7	5	TA	184900
##	722	2006	7	5	TA	130000
##	723	2008	6	5	TA	140000
##	724	2008	10	9	Ex	325000
##	725	2009	7	7	TA	265979
##	726	2010	9	5	TA	395192
##	727	2006	6	7	TA	151000
##	728	2009	7	5	TA	160200
##	729	2010	7	6	TA	170000
##	730	2010	5	5	TA	154300
##	731	2007	6	7	TA	153575
##	732	2008	6	5	TA	149300
##	733	2006	5	4	TA	130000
##	734	2009	8	5	TA	210000
##	735	2010	5	6	TA	142000
##	736	2006	5	4	TA	115000
##	737	2008	6	5	TA	181000
##	738	2008	4	7	TA	135750
##	739	2009	5	7	TA	133000
##	740	2007	9	5	TA	318061
##	741	2008	7	5	TA	200000
##	742	2010	5	8	Gd	127000
##	743	2009	9	5	TA	350000
##	744	2007	5	7	TA	101000
##	745	2008	5	9	Gd	130500
##	746	2008	8	5	TA	301500
##	747	2007	6	6	Gd	141000
##	748	2006	2	3	TA	35311
##	749	2008	6	6	TA	106000
##	750	2010	8	5	TA	262500
##	751	2007	5	8	TA	125000
##	752	2008	8	6	TA	287000
##	753	2009	7	5	TA	252678
##	754	2007	7	5	TA	162000

```
## 755   2007        6        5      TA    168500
## 756   2007        8        5      TA    239000
## 757   2007        5        7      TA    156000
## 758   2009        6        5      Gd    181000
## 759   2006        9        5      TA    326000
## 760   2007        8        5      TA    227000
## 761   2008        6        6      TA    171000
## 762   2009        4        5      TA    108000
## 763   2009        6        8      TA    110000
## 764   2007        9        5      TA    320000
## 765   2007        7        5      TA    194000
## 766   2009        6        5      TA    167000
## 767   2008        6        7      TA    240000
## 768   2008        6        7      TA    163000
## 769   2007        6        5      TA    168000
## 770   2007        5        5      TA    135000
## 771   2006        6        4      TA    151400
## 772   2007        5        5      TA    126500
## 773   2010        6        8      Gd    105900
## 774   2006        5        5      TA    158000
## 775   2007        6        5      TA    240000
## 776   2008        6        5      TA    136500
## 777   2008        7        5      TA    160000
## 778   2009        7        5      TA    190000
## 779   2009        7        5      TA    213000
## 780   2008        8        5      TA    392500
## 781   2010        7        5      TA    220000
## 782   2010        4        5      TA    140000
## 783   2006        8        5      Gd    270000
## 784   2006        6        5      TA    169990
## 785   2008        5        5      TA    145000
## 786   2006        5        6      TA    128500
## 787   2007        6        5      TA    118400
## 788   2007        6        5      TA    143500
## 789   2006        6        8      TA    116000
## 790   2007        6        7      TA    210000
## 791   2010        7        5      TA    214000
## 792   2008        9        5      TA    266000
## 793   2009        7        5      TA    233170
## 794   2007        6        5      TA    124000
## 795   2008        3        4      TA     81000
## 796   2009        6        6      Gd    193500
## 797   2006        7        5      TA    213490
## 798   2008        8        5      TA    271000
## 799   2006        5        6      TA    164500
## 800   2010        3        6      Gd    107400

num.row3 <- nrow(trainData)
tune.size <- round(0.20 * num.row3)
```

```r
set.seed(113, "L'Ecuyer")
(tuneData  <- trainData[sample(1:num.row3, tune.size, replace=FALSE),] )
```

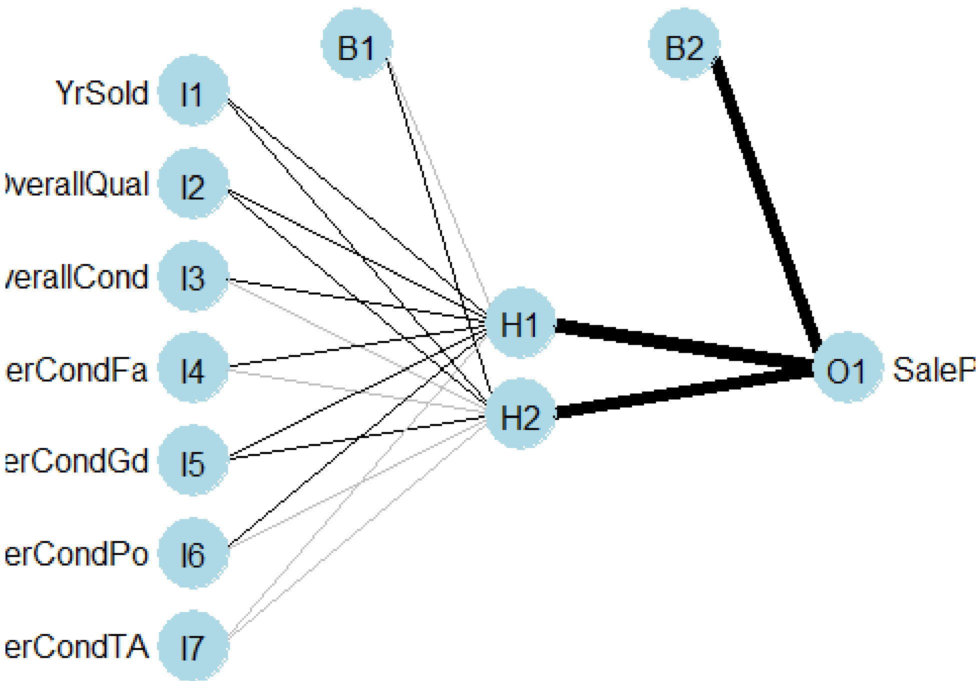
```
##      YrSold OverallQual OverallCond ExterCond SalePrice
## 138    2007           8           5        TA    440000
## 461    2009           8           5        TA    220000
## 497    2009           7           7        Gd    205000
## 719    2006           6           5        TA    145250
## 454    2007           5           5        TA    171000
## 541    2008           7           5        TA    284000
## 773    2010           6           8        Gd    105900
## 96     2006           8           5        TA    348000
## 736    2006           5           4        TA    115000
## 662    2010           4           8        Gd    125500
## 705    2006           5           8        TA    170000
## 350    2009           6           5        TA    196000
## 274    2009           7           5        TA    174000
## 425    2008           4           6        TA    135000
## 294    2009           5           4        TA     93000
## 196    2009           7           5        Gd     82500
## 329    2006           5           6        TA    180000
## 764    2007           9           5        TA    320000
## 395    2009           6           6        TA    120000
## 173    2009           4           5        TA     82500
## 561    2010           6           5        TA    158900
## 26     2008           6           8        TA    240000
## 206    2008           6           8        Gd    142000
## 310    2009           5           7        TA    135000
## 437    2009           6           6        Gd    132000
## 731    2007           6           7        TA    153575
## 226    2006           5           5        TA    160000
## 698    2008           7           5        TA    224900
## 213    2009           5           5        TA    159434
## 256    2009           8           5        TA    274900
## 50     2006           9           5        TA    360000
## 458    2010           5           4        TA    131500
## 278    2007           4           4        TA    108500
## 423    2008           5           6        TA    120500
## 707    2006           4           5        Gd     79900
## 720    2009           5           7        TA    130000
## 521    2008           7           6        TA    205950
## 279    2006          10           5        TA    438780
## 22     2009           6           9        TA    134900
## 158    2008           5           6        TA    228950
## 465    2007           5           7        TA    132000
## 702    2006           7           5        TA    176000
## 31     2007           5           5        TA    152000
## 399    2006           6           8        Gd    117000
## 548    2006           6           5        TA    165150
## 119    2007           7           5        TA    176485
```

##	319	2008	9	5	TA	446261
##	715	2009	5	7	TA	133000
##	750	2010	8	5	TA	262500
##	122	2009	5	5	TA	100000
##	29	2007	4	5	TA	123000
##	258	2006	5	5	TA	145000
##	216	2009	8	5	TA	278000
##	114	2006	5	5	TA	157000
##	506	2006	6	5	TA	150900
##	604	2010	6	6	TA	148000
##	629	2009	4	6	TA	135000
##	357	2008	6	5	TA	193000
##	377	2007	7	5	TA	302000
##	382	2010	7	5	TA	222500
##	305	2009	7	9	Gd	143000
##	518	2010	8	5	TA	272000
##	577	2009	6	6	TA	143750
##	163	2008	5	7	TA	116000
##	88	2009	6	5	TA	208300
##	622	2007	7	5	TA	226000
##	37	2006	7	5	TA	200000
##	236	2006	7	5	TA	140000
##	260	2006	8	5	TA	233230
##	170	2010	9	2	TA	394432
##	792	2008	9	5	TA	266000
##	275	2008	6	5	TA	141000
##	527	2010	5	5	TA	165500
##	253	2008	7	7	TA	116900
##	744	2007	5	7	TA	101000
##	620	2008	6	7	TA	188700
##	205	2006	7	5	TA	320000
##	716	2009	6	6	TA	159500
##	126	2006	5	4	Fa	160000
##	322	2007	8	5	TA	285000
##	283	2009	8	5	TA	184100
##	456	2006	5	5	TA	131000
##	348	2008	5	6	TA	130250
##	515	2008	9	5	TA	325000
##	330	2009	4	4	TA	80000
##	263	2008	4	4	TA	108959
##	588	2008	6	5	TA	85400
##	401	2007	5	5	TA	130000
##	469	2006	8	7	TA	230000
##	556	2007	5	8	TA	129000
##	505	2009	5	6	TA	154000
##	261	2007	7	6	TA	190000
##	230	2009	7	5	TA	213000
##	320	2009	6	5	TA	178000
##	140	2007	7	5	TA	179600
##	640	2007	7	5	TA	192000

##	635	2008	6	5	TA	193000
##	415	2006	8	5	TA	178000
##	507	2008	7	5	TA	341000
##	467	2009	9	5	TA	377500
##	605	2007	5	7	Gd	119750
##	75	2007	6	5	TA	178900
##	565	2010	6	4	TA	128000
##	473	2008	8	5	TA	232000
##	242	2007	8	5	TA	255000
##	452	2009	7	5	TA	275000
##	262	2006	7	5	TA	260000
##	63	2008	5	8	TA	119000
##	547	2007	6	3	TA	200624
##	23	2009	5	7	TA	119500
##	712	2007	6	5	TA	131500
##	478	2007	5	5	TA	135000
##	610	2008	6	8	TA	151000
##	778	2009	7	5	TA	190000
##	79	2008	7	7	TA	241500
##	479	2008	5	5	TA	129500
##	90	2007	5	7	TA	144000
##	508	2008	6	5	TA	193500
##	569	2006	5	8	Gd	225000
##	759	2006	9	5	TA	326000
##	304	2006	6	5	TA	200000
##	331	2008	7	5	TA	191000
##	46	2007	6	7	TA	257500
##	101	2008	6	6	TA	171500
##	84	2009	5	7	TA	105000
##	625	2008	10	5	TA	385000
##	356	2008	6	6	Gd	151000
##	257	2007	5	6	TA	139000
##	81	2007	4	7	TA	109900
##	417	2010	4	5	TA	83500
##	311	2007	6	5	Gd	177500
##	42	2008	6	7	TA	177000
##	443	2007	8	5	TA	424870
##	342	2006	10	5	Gd	625000
##	189	2009	6	6	Fa	132500
##	681	2009	6	5	TA	175000
##	162	2009	8	5	TA	337000
##	36	2009	10	5	TA	386250
##	151	2010	5	7	TA	130000
##	323	2007	7	5	TA	213500
##	496	2009	7	5	TA	230000
##	432	2006	6	5	TA	142500
##	351	2006	5	6	TA	217500
##	701	2009	6	5	TA	165000
##	728	2009	7	5	TA	160200
##	281	2009	8	5	TA	172500

```
## 367   2006          7          6        TA     226000
## 62    2006          7          5        TA     250000
## 445   2009          7          5        TA     219500
## 157   2009          5          6        TA     131400
## 687   2007          5          5        TA     108000
## 677   2009          6          4        TA     149500
## 183   2007          6          5        TA     181000
## 147   2010          8          5        TA     202900
## 362   2009          8          5        TA     340000
## 244   2007          8          5        TA     281213
## 129   2010          5          5        TA     142000
## 663   2009          6          5        TA     231500
## 613   2009          5          6        TA     110000
## 766   2009          6          5        TA     167000
```

```
#Run the full trainData set with the above hyperparameters-this will take a
while .

We are scaling the response varaibale.

library(nnet)
set.seed(113, "L'Ecuyer")
nn.di <- nnet(SalePrice/19000 ~ ., data = trainData, size = 2, decay=0)

## # weights:  19
## initial value 81197.807900
## final  value 75956.952592
## converged

library(NeuralNetTools)
par(mar= numeric(4))
plotnet(nn.di)

#plot a neural network
```




```
pred.nn <- predict(nn.di , newdata = predData) *19000
```

Now we will determine our regression residuals and plot them

```
residuals<- (trueData$predData.SalePrice- pred.nn)

# Calculation of root mean square error - good metric to compare regression models
# since its units are the same as the response which is price
# but it is sensitive to large residuals.
(rmse <- sqrt(mean(residuals^2)) )

## [1] 187058.3

# An alternate metric is mean absolute error which is less sensitive to large residuals
(mae <- mean(abs(residuals)) )

## [1] 170100

(res.df <- data.frame(trueData$predData.SalePrice, as.vector(pred.nn),
residuals) )

##      trueData.predData.SalePrice as.vector.pred.nn. residuals
## 239                       318000              19000    299000
## 590                        79500              19000     60500
## 222                       200000              19000    181000
## 1134                      239500              19000    220500
```

##	709	179540	19000	160540
##	1037	315500	19000	296500
##	20	139000	19000	120000
##	990	197000	19000	178000
##	175	184000	19000	165000
##	414	115000	19000	96000
##	223	179900	19000	160900
##	207	143900	19000	124900
##	194	130000	19000	111000
##	894	165000	19000	146000
##	1437	120500	19000	101500
##	745	180000	19000	161000
##	311	165600	19000	146600
##	565	268000	19000	249000
##	333	284000	19000	265000
##	543	213250	19000	194250
##	837	153500	19000	134500
##	515	96500	19000	77500
##	1299	160000	19000	141000
##	1240	265900	19000	246900
##	634	139400	19000	120400
##	73	185000	19000	166000
##	1205	153500	19000	134500
##	781	176000	19000	157000
##	287	159000	19000	140000
##	1433	64500	19000	45500
##	978	199900	19000	180900
##	619	314813	19000	295813
##	349	154000	19000	135000
##	483	155000	19000	136000
##	1215	134500	19000	115500
##	952	119900	19000	100900
##	216	134450	19000	115450
##	377	148000	19000	129000
##	419	126000	19000	107000
##	870	236000	19000	217000
##	1319	275000	19000	256000
##	58	196500	19000	177500
##	38	153000	19000	134000
##	240	113000	19000	94000
##	644	152000	19000	133000
##	317	260000	19000	241000
##	786	161500	19000	142500
##	1340	128500	19000	109500
##	1053	165000	19000	146000
##	725	320000	19000	301000
##	965	214900	19000	195900
##	1060	220000	19000	201000
##	1320	111000	19000	92000
##	925	207500	19000	188500

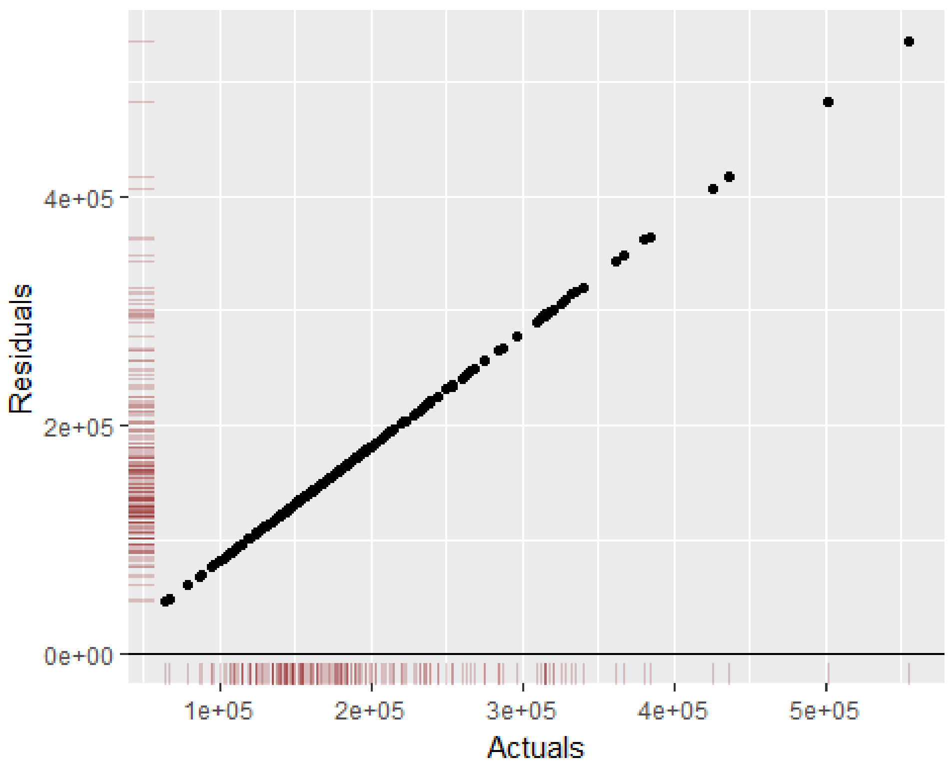
##	1420	223000	19000	204000
##	1321	156500	19000	137500
##	1399	138000	19000	119000
##	1426	142000	19000	123000
##	108	115000	19000	96000
##	1086	147000	19000	128000
##	378	340000	19000	321000
##	1029	105000	19000	86000
##	623	135000	19000	116000
##	1288	190000	19000	171000
##	928	176000	19000	157000
##	500	120000	19000	101000
##	598	194201	19000	175201
##	313	119900	19000	100900
##	992	168000	19000	149000
##	956	145000	19000	126000
##	550	263000	19000	244000
##	525	315750	19000	296750
##	1287	143000	19000	124000
##	1103	135000	19000	116000
##	258	220000	19000	201000
##	476	132500	19000	113500
##	198	235000	19000	216000
##	1457	210000	19000	191000
##	1446	129000	19000	110000
##	1410	215000	19000	196000
##	972	173000	19000	154000
##	1017	203000	19000	184000
##	459	161000	19000	142000
##	1367	193000	19000	174000
##	197	311872	19000	292872
##	708	254000	19000	235000
##	71	244000	19000	225000
##	346	140200	19000	121200
##	1038	287000	19000	268000
##	179	501837	19000	482837
##	1229	367294	19000	348294
##	1314	333168	19000	314168
##	879	148000	19000	129000
##	1318	208900	19000	189900
##	1424	274970	19000	255970
##	284	244600	19000	225600
##	1439	149700	19000	130700
##	104	198900	19000	179900
##	380	179000	19000	160000
##	866	148500	19000	129500
##	1304	232000	19000	213000
##	1225	184000	19000	165000
##	298	239000	19000	220000
##	233	94500	19000	75500

##	315	178000	19000	159000
##	845	153900	19000	134900
##	617	183200	19000	164200
##	1443	310000	19000	291000
##	705	213000	19000	194000
##	399	67000	19000	48000
##	660	167000	19000	148000
##	1236	138887	19000	119887
##	205	110000	19000	91000
##	819	155000	19000	136000
##	1169	235000	19000	216000
##	1207	107000	19000	88000
##	930	222000	19000	203000
##	792	146800	19000	127800
##	462	155000	19000	136000
##	809	159950	19000	140950
##	386	192000	19000	173000
##	658	149000	19000	130000
##	712	102776	19000	83776
##	612	148000	19000	129000
##	122	100000	19000	81000
##	402	164990	19000	145990
##	1119	140000	19000	121000
##	113	383970	19000	364970
##	912	143500	19000	124500
##	835	139950	19000	120950
##	390	426000	19000	407000
##	218	107000	19000	88000
##	21	325300	19000	306300
##	1027	167500	19000	148500
##	479	297000	19000	278000
##	1369	144000	19000	125000
##	1146	149000	19000	130000
##	1390	131000	19000	112000
##	1246	178000	19000	159000
##	915	173733	19000	154733
##	703	361919	19000	342919
##	42	170000	19000	151000
##	1059	335000	19000	316000
##	98	94750	19000	75750
##	903	180000	19000	161000
##	1370	232000	19000	213000
##	1066	328000	19000	309000
##	579	146000	19000	127000
##	1302	177500	19000	158500
##	913	88000	19000	69000
##	1081	145000	19000	126000
##	78	127000	19000	108000
##	370	162000	19000	143000
##	789	107900	19000	88900

```
## 447                       190000        19000   171000
## 1392                      124000        19000   105000
## 469                       250000        19000   231000
## 673                       165000        19000   146000
## 1297                      155000        19000   136000
## 1176                      285000        19000   266000
## 1430                      182900        19000   163900
## 1168                      173000        19000   154000
## 1269                      381000        19000   362000
## 46                        319900        19000   300900
## 993                       187000        19000   168000
## 129                       155000        19000   136000
## 1023                       87000        19000    68000
## 278                       141000        19000   122000
## 914                       145000        19000   126000
## 254                       158000        19000   139000
## 407                       115000        19000    96000
## 296                       142500        19000   123500
## 1247                      186500        19000   167500
## 352                       190000        19000   171000
## 646                       143250        19000   124250
## 415                       228000        19000   209000
## 731                       236500        19000   217500
## 1360                      315000        19000   296000
## 829                       185000        19000   166000
## 1441                      191000        19000   172000
## 350                       437154        19000   418154
## 1034                      230000        19000   211000
## 1005                      181000        19000   162000
## 138                       171000        19000   152000
## 417                       149500        19000   130500
## 339                       202500        19000   183500
## 729                       110000        19000    91000
## 1126                      115000        19000    96000
## 416                       181134        19000   162134
## 582                       253293        19000   234293
## 1344                      177000        19000   158000
## 506                       124500        19000   105500
## 593                       138000        19000   119000
## 388                       125000        19000   106000
## 1015                      119200        19000   100200
## 771                       134900        19000   115900
## 510                       124500        19000   105500
## 441                       555000        19000   536000
## 901                       110000        19000    91000
## 926                       175000        19000   156000
```

```r
names(res.df)[1] <- "Actuals"
names(res.df)[2] <- "Responses"
names(res.df)[3] <- "Residuals"
```

```r
ggplot(res.df, aes(x = Actuals, y = Residuals)) +
     geom_point() +
     geom_hline(aes(yintercept = 0)) +
     geom_rug(col=rgb(.5,0,0,alpha=.2))


#Plot actuval vs residuals
```



A line with 45 degree slop is showing that our neural network is good fit in our dataset.

We used the following libraries in this project.

```r
library(ggplot2)
library(magrittr)
library(dplyr)
library(corrgram)
library(tidyr)
library(VIM)
library(corrplot)

library(tidyverse)

library(ggmap)

library(DT)
library(knitr)
```

```r
library(rpart)
library(rpart.plot)
library(gridExtra)
library(sp)
library(validate)
library(rworldmap)
```

Future Plan:

I will identify is there any relationship between most expensive neighborhood and most prefer neighborhoods.