DATA605: Fundamentals of Computational Mathematics

Final

Donald Butler

05/15/2022

```
library(tidyverse)
```

Problem 1: Playing with PageRank

You'll verify for yourself that PageRank works by performing calculations on a small universe of web pages. Let's use the 6 page universe that we had in the previous discussion. For this directed graph, perform the following calculations in R.

Part 1A: Create Transition Matrix

Form the A matrix. Then, introduce decay and form the B matrix as we did in the course notes.

Each column of the transition matrix, A, is the probability of following a link to the next page. The sum of each column must be 1. Page #2 has no outbound links and is considered a dangling node, so we assume that a user can choose the next site at random, so the value of each node in column 2 is $1/n$.

```
n <- 6
A <- matrix(c(0, 1/2, 1/2, 0, 0, 0,
              rep(1/n,n), # dangling node
              1/3, 1/3, 0, 0, 1/3, 0,
              0, 0, 0, 0, 1/2, 1/2,
              0, 0, 0, 1/2, 0, 1/2,
              0, 0, 0, 1, 0, 0), nrow = n, ncol = n, byrow = FALSE)
A
```

```
##      [,1]      [,2]      [,3] [,4] [,5] [,6]
## [1,]  0.0 0.1666667 0.3333333  0.0  0.0    0
## [2,]  0.5 0.1666667 0.3333333  0.0  0.0    0
## [3,]  0.5 0.1666667 0.0000000  0.0  0.0    0
## [4,]  0.0 0.1666667 0.0000000  0.0  0.5    1
## [5,]  0.0 0.1666667 0.3333333  0.5  0.0    0
## [6,]  0.0 0.1666667 0.0000000  0.5  0.5    0
```

Using a decay factor of .85 indicates an 85% chance that users will follow links on the current page, and a 15% chance of going to a random site.

```
decay <- .85
B <- decay * A + (1-decay)/n
B
```

```
##       [,1]      [,2]      [,3]  [,4]  [,5]  [,6]
## [1,] 0.025 0.1666667 0.3083333 0.025 0.025 0.025
## [2,] 0.450 0.1666667 0.3083333 0.025 0.025 0.025
## [3,] 0.450 0.1666667 0.0250000 0.025 0.025 0.025
## [4,] 0.025 0.1666667 0.0250000 0.025 0.450 0.875
## [5,] 0.025 0.1666667 0.3083333 0.450 0.025 0.025
## [6,] 0.025 0.1666667 0.0250000 0.450 0.450 0.025
```

Part 1B: Compute PageRank

Start with a uniform rank vector, r, and perform power iterations on B until convergence. That is, compute the solution $r = B^n \times r$. Attempt this for a sufficiently large n so that r actually converges.

The PageRank function relies on the property of matrix multiplication that $A^n \times r = A(A^{n-1} \times r)$. This function will iterate until the next calculation is the same as the previous, or max_iterations is reached.

```
PageRank <- function(r, B, max_iterations = 10000) {
  for (i in 1:max_iterations) {
    new_r <- B %*% r
    if (all(r == new_r)) {
      break
    }
    r <- new_r
  }

  # Will print the number of iterations needed for convergence
  print(i-1)
  return(r)
}

r = rep(1/n,n)

pr <- PageRank(r,B)
```

```
## [1] 71
```

```
pr
```

```
##            [,1]
## [1,] 0.05170475
## [2,] 0.07367926
## [3,] 0.05741241
## [4,] 0.34870369
## [5,] 0.19990381
## [6,] 0.26859608
```

From the PageRank output, we see that it took 71 iterations to converge.

Verify that the sum of the PageRank result is 1.

```
sum(pr)
```

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

Part 1C: Compute eigen-decomposition

Computer the eigen-decomposition of B and verify that you indeed get an eigenvalue vector that you obtained in the previous power iteration method. Further, this eigenvector has all positive entries and it sums to 1.

```
decomp <- eigen(B)
decomp
```

```
## eigen() decomposition
## $values
## [1]  1.00000000+0i  0.57619235+0i -0.42500000+0i -0.42500000-0i -0.34991524+0i
## [6] -0.08461044+0i
##
## $vectors
##              [,1]          [,2]                       [,3]
## [1,] 0.1044385+0i  0.2931457+0i  2.486934e-15+0.0000e+00i
## [2,] 0.1488249+0i  0.5093703+0i -8.528385e-16-6.9832e-23i
## [3,] 0.1159674+0i  0.3414619+0i -1.930646e-15-0.0000e+00i
## [4,] 0.7043472+0i -0.5890805+0i -7.071068e-01+0.0000e+00i
## [5,] 0.4037861+0i -0.1413606+0i  7.071068e-01+0.0000e+00i
## [6,] 0.5425377+0i -0.4135367+0i  0.000000e+00-1.7058e-08i
##                            [,4]           [,5]             [,6]
## [1,]  2.486934e-15-0.0000e+00i -0.06471710+0i -0.212296003+0i
## [2,] -8.528385e-16+6.9832e-23i  0.01388698+0i  0.854071294+0i
## [3,] -1.930646e-15+0.0000e+00i  0.07298180+0i -0.363638739+0i
## [4,] -7.071068e-01+0.0000e+00i -0.66058664+0i  0.018399984+0i
## [5,]  7.071068e-01-0.0000e+00i  0.73761812+0i -0.304719509+0i
## [6,]  0.000000e+00+1.7058e-08i -0.09918316+0i  0.008182973+0i
```

The largest eigenvalue is 1 and the associated eigenvector is the first vector. Our pr vector is normalized, so we can normalize the eigenvector to compare.

```
ev <- Re(decomp$vectors[,1]/sum(decomp$vectors[,1]))
ev
```

```
## [1] 0.05170475 0.07367926 0.05741241 0.34870369 0.19990381 0.26859608
```

Verify pr is equal to the normalized ev found using eigen-decomposition.

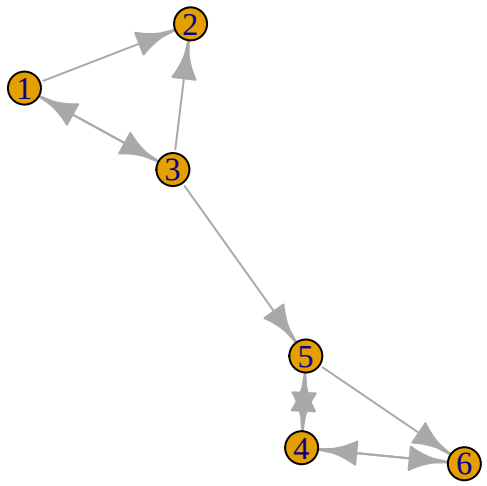
```
all(abs(ev - pr) < 1e-10)
```

```
## [1] TRUE
```

Part 1D: Graph PageRank Method

The adjacency matrix is not quite the same as the transition matrix in Part 1A. Here, the row represents the pages that can be reached from the site. Additionally, we do not need to account for the dangling node for site 2, so we'll leave the row with zeros.

```
library(igraph)
AdjM <- matrix(c(0, 1/2, 1/2, 0, 0, 0,
                 0, 0, 0, 0, 0, 0,
                 1/3, 1/3, 0, 0, 1/3, 0,
                 0, 0, 0, 0, 1/2, 1/2,
                 0, 0, 0, 1/2, 0, 1/2,
                 0, 0, 0, 1, 0, 0), nrow = n, ncol = n, byrow = TRUE)

g <- graph_from_adjacency_matrix(AdjM, weighted = TRUE)
plot(g)
```



Use the page.rank function to calculate the vector

```
gpr <- page.rank(g)
gpr$vector
```

```
## [1] 0.05170475 0.07367926 0.05741241 0.34870369 0.19990381 0.26859608
```

Verify *pr* is equal to the vector found using the igraph package.

```
all(abs(gpr$vector - pr) < 1e-10)
```

```
## [1] TRUE
```

Problem 2: Digit Recognizer

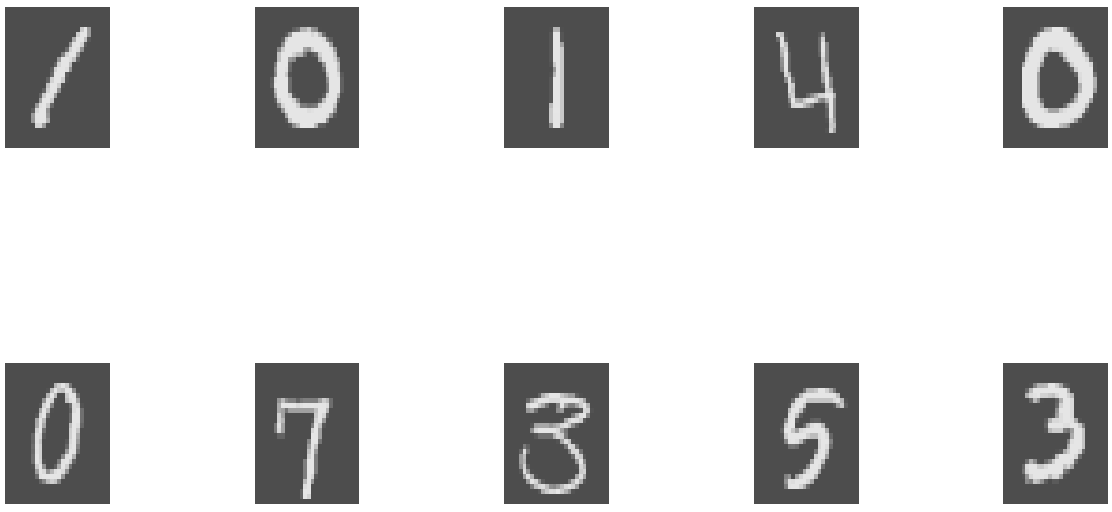
Go to https://www.kaggle.com/c/digit-recognizer/overview, and download the test and train dataset files.

Part 2A: Plot Training Images

Using the train.csv file, plot representations of the first 10 images to understand the data format. Go ahead and divide all pixels by 255 to produce values between 0 and 1.

From the kaggle website, the training dataset contains a label column which has the number represented by the image, and 784 pixel columns (pixel0 to pixel783), representing the 28 x 28 image. The image function displays the matrix with the top-left corner in the 1,n position in the matrix. So to display the image properly, we need to rotate the matrix 90 degrees clockwise.

```
rotate <- function(m) {
  return(t(apply(m,2,rev)))
}

train_digits <- read.csv('train_digits.csv') %>%
  mutate(across(starts_with('pixel'), ~./255))

par(mfrow=c(2,5))
for (i in 1:10) {
  m <- rotate(matrix(as.numeric(train_digits[i,-c(1)]),nrow = 28, byrow = TRUE))
  image(m, axes=FALSE, col = grey.colors(255), useRaster = TRUE)
}
```



View the labels of the first 10 records.

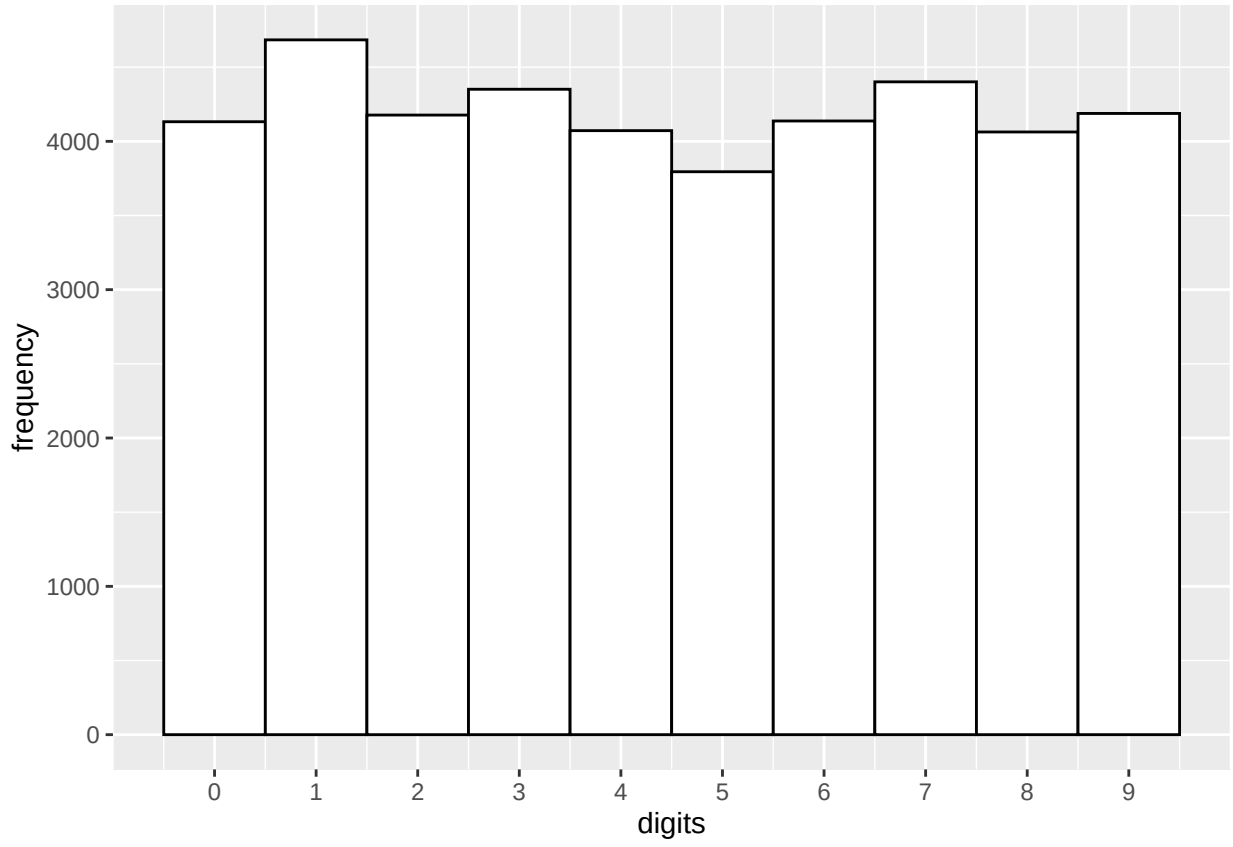
```
train_digits$label[1:10]
```

```
##  [1] 1 0 1 4 0 0 7 3 5 3
```

Part 2B: Training Digit Distribution

What is the frequency distribution of the numbers in the dataset?

```
library(ggplot2)

ggplot(train_digits, aes(x = label)) +
  geom_histogram(bins = 10, color = 'black', fill = 'white') +
  xlab('digits') +
  ylab('frequency') +
  scale_x_continuous(breaks = 0:9)
```



The training dataset contains 42000 records and is nearly uniformly distributed, which should make a good training set.

Part 2C: Mean Pixel Intensity

For each number, provide the mean pixel intensity. What does this tell you?

```r
if (!exists('train_intensity')) {
  train_digits_intensity <- train_digits %>%
    rowwise() %>%
    transmute(label, intensity = sum(c_across(starts_with('pixel')))) %>%
    ungroup()
}

train_digits_intensity %>%
  group_by(label) %>%
  summarise(mean_intensity = mean(intensity))
```

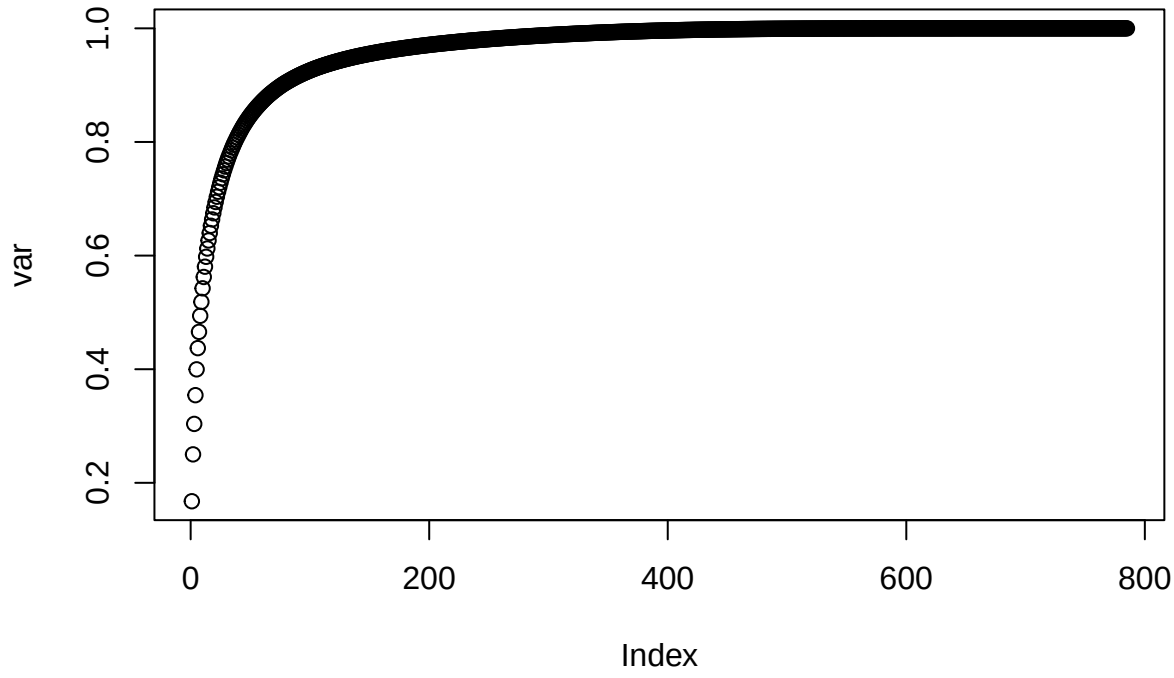
```
## # A tibble: 10 x 2
##    label mean_intensity
##    <int>          <dbl>
##  1     0          136.
##  2     1           59.6
##  3     2          117.
##  4     3          111.
##  5     4           95.0
##  6     5          101.
##  7     6          109.
##  8     7           89.9
##  9     8          118.
## 10     9           96.3
```

I would expect the mean pixel intensity to be highest for digits with the longest total pen stroke. As expected 1 has the lowest value, followed by 4, 7, and 9 which are all relatively close together. I am somewhat surprised about how much higher 0 is than the numbers 2, 3, and 8.

Part 2D: Principal Component Analysis

Reduce the data by using principal components that account for 95% of the variance. How many components did you generate? Use PCA to generate all possible components. How many components are possible?

```r
if (!exists('train_digits.pca')) {
  train_digits.pca <- prcomp(train_digits)
}

sd <- train_digits.pca$sdev
var <- cumsum(sd^2)/sum(sd^2)
plot(var)
```




```
which.max(var >= .95)
```

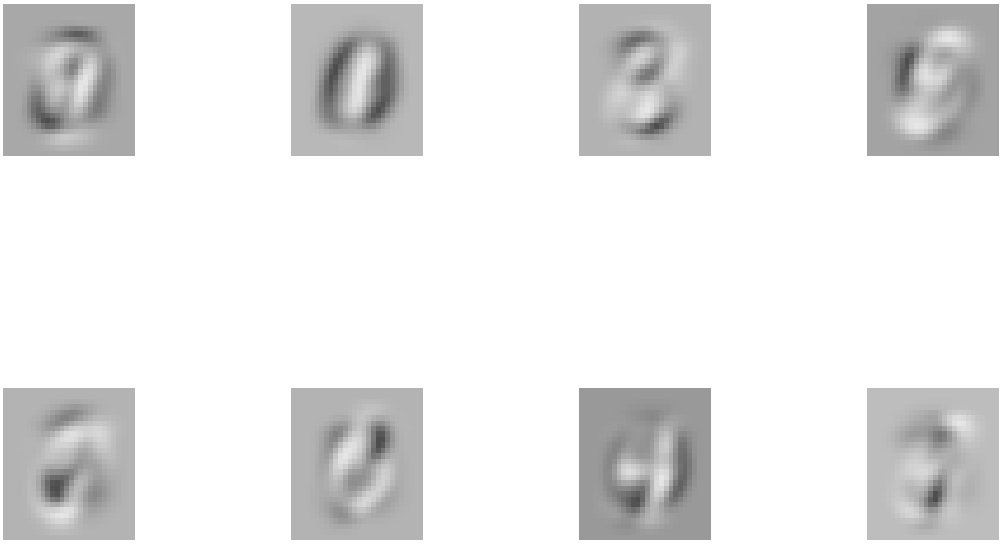
```
## [1] 139
```

To account for 95% of the variance, we needed 139 components. The maximum number of components possible is 785, one for each pixel (28x28), and one for the label column.

Part 2E: Plot PCA

Plot the first 10 images generated by the PCA. They will appear to be noise. Why?

```
par(mfrow=c(2,4))
for (i in 1:8) {
  m <- rotate(matrix(as.numeric(train_digits.pca$rotation[-1,i]),nrow = 28, byrow = TRUE))
  image(m, axes=FALSE, col = grey.colors(255), useRaster = TRUE)
}
```

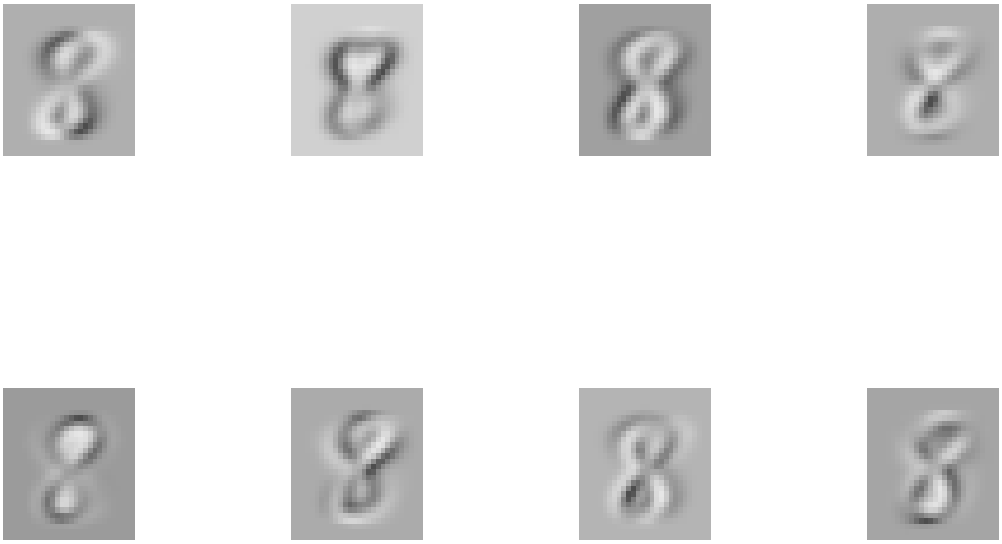


The images appear to be noise because we didn't separate the digits using the label column.

Part 2F: PCA 8

Now, select only those images that have labels that are 8. Re-run PCA that accounts for all of the variance. Plot the first 10 images. What do you see?

```
if (!exists('train_digits_8.pca')) {
  train_digits_8.pca <- prcomp(train_digits %>% filter(label == 8))
}

par(mfrow=c(2,4))
for (i in 1:8) {
  m <- rotate(matrix(as.numeric(train_digits_8.pca$rotation[-1,i]),nrow = 28, byrow = TRUE))
  image(m, axes=FALSE, col = grey.colors(255), useRaster = TRUE)
}
```



The PCA results generate images that can clearly be recognized as an 8.

Part 2G: Multinomial Model

An incorrect approach to predicting the images would be to build a linear regression model with y as the digit values and X as the pixel matrix. Instead, we can build a multinomial model that classifies the digits. Build a multinomial model on the entirety of the training set. Then provide its classification accuracy as well as a matrix of observed versus forecast values. This matrix will be a 10×10, and correct classification will be on the diagonal.

```
library(nnet)
library(caret)

train_digits_X <- train_digits[,-1]
train_digits_y <- as.factor(train_digits[,'label'])

train_digits_model <- multinom(train_digits_y~., train_digits_X, MaxNWts = 10000)
```

```
## # weights:  7860 (7065 variable)
## initial  value 96708.573906
## iter  10 value 25322.714106
## iter  20 value 20402.086316
## iter  30 value 19312.872829
## iter  40 value 18703.256586
## iter  50 value 18197.815143
## iter  60 value 17732.985798
## iter  70 value 16739.962157
## iter  80 value 14961.658448
```

```
## iter  90 value 13446.085942
## iter 100 value 12442.636014
## final  value 12442.636014
## stopped after 100 iterations
```

Use the model to predict against the training dataset and construct the confusion matrix.

```
train_digits_prediction <- predict(train_digits_model, newdata = train_digits_X)
confusionMatrix(train_digits_prediction, train_digits_y)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    0    1    2    3    4    5    6    7    8    9
##          0 3994    0   19   11    4   35   17    5   19   12
##          1    3 4588   59   32   20   39   28   37  134   18
##          2   11   13 3753   88   17   19   17   37   21   10
##          3    8   12   65 3879    9   91    1    9   91   53
##          4   13    6   60   10 3852   55   35   44   38  136
##          5   33   12   20  162    5 3386   45   10  132   33
##          6   35    3   38   15   22   52 3973    2   19    3
##          7    7    8   54   35    7   28    2 4076   18   87
##          8   20   32   85   78   22   51   17    4 3519   25
##          9    8   10   24   41  114   39    2  177   72 3811
##
## Overall Statistics
##
##                Accuracy : 0.9245
##                  95% CI : (0.922, 0.9271)
##     No Information Rate : 0.1115
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.9161
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
## Statistics by Class:
##
##                      Class: 0 Class: 1 Class: 2 Class: 3 Class: 4 Class: 5
## Sensitivity           0.96660   0.9795  0.89849  0.89152  0.94597  0.89223
## Specificity           0.99678   0.9901  0.99384  0.99100  0.98953  0.98817
## Pos Pred Value        0.97036   0.9254  0.94155  0.91963  0.90657  0.88223
## Neg Pred Value        0.99636   0.9974  0.98885  0.98751  0.99417  0.98928
## Prevalence            0.09838   0.1115  0.09945  0.10360  0.09695  0.09036
## Detection Rate        0.09510   0.1092  0.08936  0.09236  0.09171  0.08062
## Detection Prevalence  0.09800   0.1180  0.09490  0.10043  0.10117  0.09138
## Balanced Accuracy     0.98169   0.9848  0.94617  0.94126  0.96775  0.94020
##                      Class: 6 Class: 7 Class: 8 Class: 9
## Sensitivity            0.9604  0.92615  0.86611  0.90998
## Specificity            0.9950  0.99346  0.99120  0.98712
## Pos Pred Value         0.9546  0.94308  0.91331  0.88669
## Neg Pred Value         0.9957  0.99137  0.98574  0.99000
## Prevalence             0.0985  0.10479  0.09674  0.09971
```

```
## Detection Rate         0.0946  0.09705  0.08379  0.09074
## Detection Prevalence   0.0991  0.10290  0.09174  0.10233
## Balanced Accuracy      0.9777  0.95981  0.92865  0.94855
```

The model has an accuracy of 92% against the training set used to build the model.

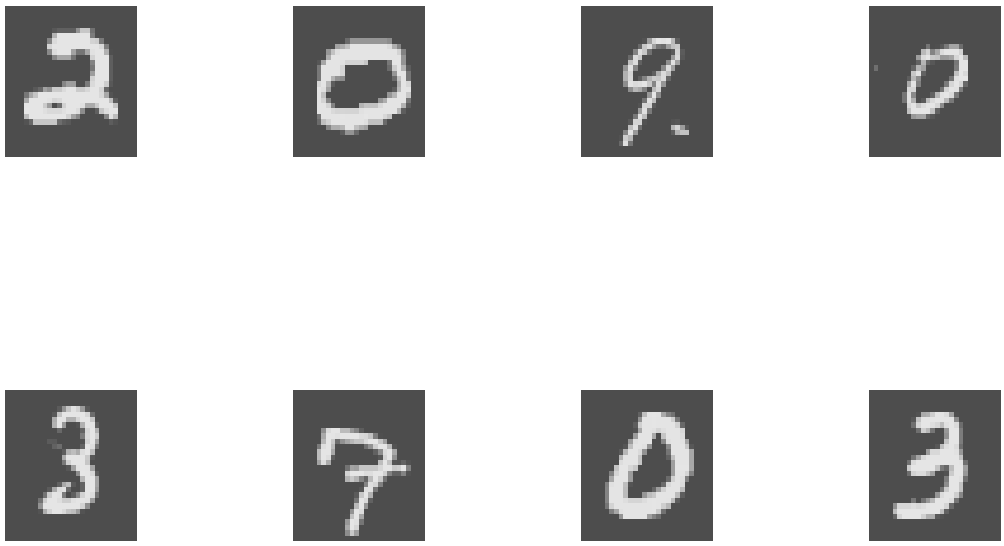
Part 2H: Verify Model with Test Dataset

Load the test dataset, and predict against the model.

```
test_digits <- read.csv('test_digits.csv') %>%
  mutate(across(starts_with('pixel'), ~./255))

test_digits_prediction <- predict(train_digits_model, newdata = test_digits)
```

Graph the first 8 from the test set to compare with the predictions.

```
par(mfrow=c(2,4))
for (i in 1:8) {
  m <- rotate(matrix(as.numeric(test_digits[i,]),nrow = 28, byrow = TRUE))
  image(m, axes=FALSE, col = grey.colors(255), useRaster = TRUE)
}
```



```
test_digits_prediction[1:8]
```

```
## [1] 2 0 9 7 3 7 0 3
## Levels: 0 1 2 3 4 5 6 7 8 9
```

Looks like our model incorrectly predicted the 4th image as a 7. Generate submission.csv file to submit to kaggle to get the results of the model.

```
write.csv(test_digits_prediction,'digits_submission.csv',row.names = TRUE, quote = FALSE)
```

Submitted results to kaggle, https://www.kaggle.com/donaldbutler95/competitions, and was scored at .90021.

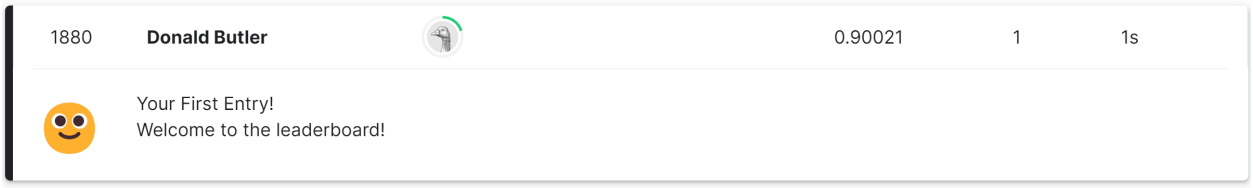


Figure 1: Kaggle Submission

Problem 3: House Prices Advanced Regression Techniques

You are to compete in the House Prices: Advanced Regression Techniques competition, https://www.kaggle.com/c/house-prices-advanced-regression-techniques . I want you to do the following.

```
train_house <- read.csv('train_housing.csv')
test_house <- read.csv('test_housing.csv')

str(train_house)
```

```
## 'data.frame':    1460 obs. of  81 variables:
##  $ Id           : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ MSSubClass   : int  60 20 60 70 60 50 20 60 50 190 ...
##  $ MSZoning     : chr  "RL" "RL" "RL" "RL" ...
##  $ LotFrontage  : int  65 80 68 60 84 85 75 NA 51 50 ...
##  $ LotArea      : int  8450 9600 11250 9550 14260 14115 10084 10382 6120 7420 ...
##  $ Street       : chr  "Pave" "Pave" "Pave" "Pave" ...
##  $ Alley        : chr  NA NA NA NA ...
##  $ LotShape     : chr  "Reg" "Reg" "IR1" "IR1" ...
##  $ LandContour  : chr  "Lvl" "Lvl" "Lvl" "Lvl" ...
##  $ Utilities    : chr  "AllPub" "AllPub" "AllPub" "AllPub" ...
##  $ LotConfig    : chr  "Inside" "FR2" "Inside" "Corner" ...
##  $ LandSlope    : chr  "Gtl" "Gtl" "Gtl" "Gtl" ...
##  $ Neighborhood : chr  "CollgCr" "Veenker" "CollgCr" "Crawfor" ...
##  $ Condition1   : chr  "Norm" "Feedr" "Norm" "Norm" ...
##  $ Condition2   : chr  "Norm" "Norm" "Norm" "Norm" ...
##  $ BldgType     : chr  "1Fam" "1Fam" "1Fam" "1Fam" ...
##  $ HouseStyle   : chr  "2Story" "1Story" "2Story" "2Story" ...
##  $ OverallQual  : int  7 6 7 7 8 5 8 7 7 5 ...
##  $ OverallCond  : int  5 8 5 5 5 5 5 6 5 6 ...
##  $ YearBuilt    : int  2003 1976 2001 1915 2000 1993 2004 1973 1931 1939 ...
##  $ YearRemodAdd : int  2003 1976 2002 1970 2000 1995 2005 1973 1950 1950 ...
##  $ RoofStyle    : chr  "Gable" "Gable" "Gable" "Gable" ...
##  $ RoofMatl     : chr  "CompShg" "CompShg" "CompShg" "CompShg" ...
##  $ Exterior1st  : chr  "VinylSd" "MetalSd" "VinylSd" "Wd Sdng" ...
```

```
##  $ Exterior2nd  : chr  "VinylSd" "MetalSd" "VinylSd" "Wd Shng" ...
##  $ MasVnrType   : chr  "BrkFace" "None" "BrkFace" "None" ...
##  $ MasVnrArea   : int  196 0 162 0 350 0 186 240 0 0 ...
##  $ ExterQual    : chr  "Gd" "TA" "Gd" "TA" ...
##  $ ExterCond    : chr  "TA" "TA" "TA" "TA" ...
##  $ Foundation   : chr  "PConc" "CBlock" "PConc" "BrkTil" ...
##  $ BsmtQual     : chr  "Gd" "Gd" "Gd" "TA" ...
##  $ BsmtCond     : chr  "TA" "TA" "TA" "Gd" ...
##  $ BsmtExposure : chr  "No" "Gd" "Mn" "No" ...
##  $ BsmtFinType1 : chr  "GLQ" "ALQ" "GLQ" "ALQ" ...
##  $ BsmtFinSF1   : int  706 978 486 216 655 732 1369 859 0 851 ...
##  $ BsmtFinType2 : chr  "Unf" "Unf" "Unf" "Unf" ...
##  $ BsmtFinSF2   : int  0 0 0 0 0 0 0 32 0 0 ...
##  $ BsmtUnfSF    : int  150 284 434 540 490 64 317 216 952 140 ...
##  $ TotalBsmtSF  : int  856 1262 920 756 1145 796 1686 1107 952 991 ...
##  $ Heating      : chr  "GasA" "GasA" "GasA" "GasA" ...
##  $ HeatingQC    : chr  "Ex" "Ex" "Ex" "Gd" ...
##  $ CentralAir   : chr  "Y" "Y" "Y" "Y" ...
##  $ Electrical   : chr  "SBrkr" "SBrkr" "SBrkr" "SBrkr" ...
##  $ X1stFlrSF    : int  856 1262 920 961 1145 796 1694 1107 1022 1077 ...
##  $ X2ndFlrSF    : int  854 0 866 756 1053 566 0 983 752 0 ...
##  $ LowQualFinSF : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ GrLivArea    : int  1710 1262 1786 1717 2198 1362 1694 2090 1774 1077 ...
##  $ BsmtFullBath : int  1 0 1 1 1 1 1 1 0 1 ...
##  $ BsmtHalfBath : int  0 1 0 0 0 0 0 0 0 0 ...
##  $ FullBath     : int  2 2 2 1 2 1 2 2 2 1 ...
##  $ HalfBath     : int  1 0 1 0 1 1 0 1 0 0 ...
##  $ BedroomAbvGr : int  3 3 3 3 4 1 3 3 2 2 ...
##  $ KitchenAbvGr : int  1 1 1 1 1 1 1 1 2 2 ...
##  $ KitchenQual  : chr  "Gd" "TA" "Gd" "Gd" ...
##  $ TotRmsAbvGrd : int  8 6 6 7 9 5 7 7 8 5 ...
##  $ Functional   : chr  "Typ" "Typ" "Typ" "Typ" ...
##  $ Fireplaces   : int  0 1 1 1 1 0 1 2 2 2 ...
##  $ FireplaceQu  : chr  NA "TA" "TA" "Gd" ...
##  $ GarageType   : chr  "Attchd" "Attchd" "Attchd" "Detchd" ...
##  $ GarageYrBlt  : int  2003 1976 2001 1998 2000 1993 2004 1973 1931 1939 ...
##  $ GarageFinish : chr  "RFn" "RFn" "RFn" "Unf" ...
##  $ GarageCars   : int  2 2 2 3 3 2 2 2 2 1 ...
##  $ GarageArea   : int  548 460 608 642 836 480 636 484 468 205 ...
##  $ GarageQual   : chr  "TA" "TA" "TA" "TA" ...
##  $ GarageCond   : chr  "TA" "TA" "TA" "TA" ...
##  $ PavedDrive   : chr  "Y" "Y" "Y" "Y" ...
##  $ WoodDeckSF   : int  0 298 0 0 192 40 255 235 90 0 ...
##  $ OpenPorchSF  : int  61 0 42 35 84 30 57 204 0 4 ...
##  $ EnclosedPorch: int  0 0 0 272 0 0 0 228 205 0 ...
##  $ X3SsnPorch   : int  0 0 0 0 0 320 0 0 0 0 ...
##  $ ScreenPorch  : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ PoolArea     : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ PoolQC       : chr  NA NA NA NA ...
##  $ Fence        : chr  NA NA NA NA ...
##  $ MiscFeature  : chr  NA NA NA NA ...
##  $ MiscVal      : int  0 0 0 0 0 700 0 350 0 0 ...
##  $ MoSold       : int  2 5 9 2 12 10 8 11 4 1 ...
##  $ YrSold       : int  2008 2007 2008 2006 2008 2009 2007 2009 2008 2008 ...
```

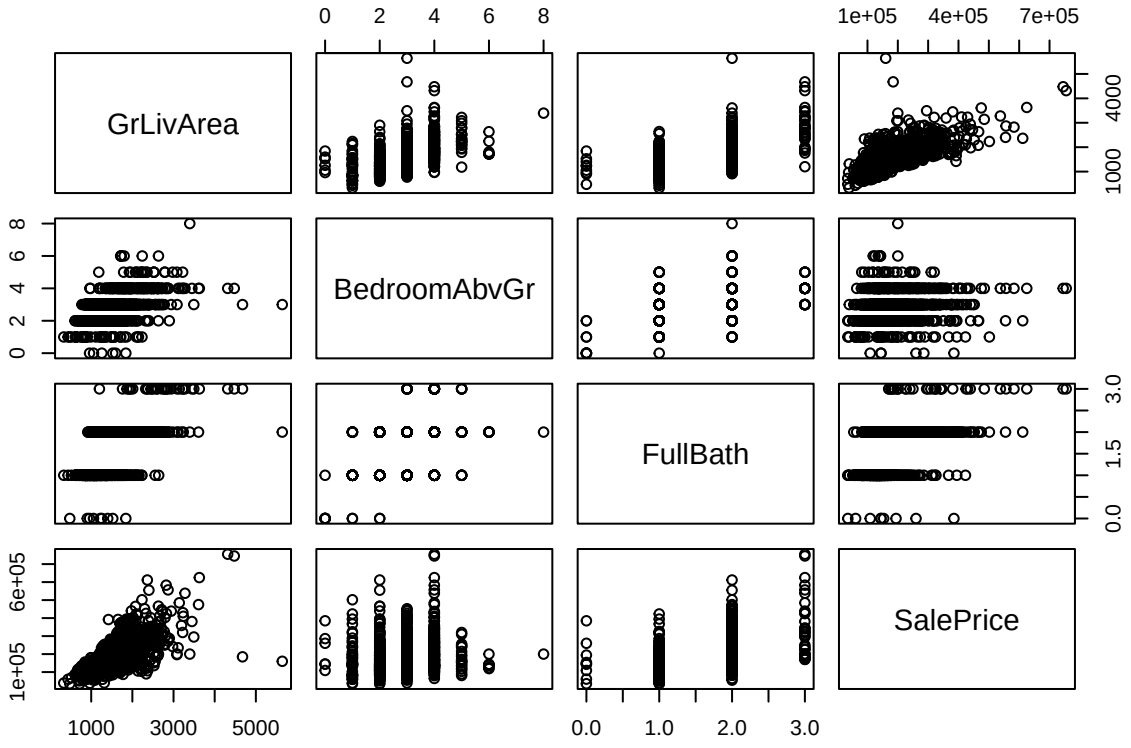
```
##  $ SaleType     : chr  "WD" "WD" "WD" "WD" ...
##  $ SaleCondition: chr  "Normal" "Normal" "Normal" "Abnorml" ...
##  $ SalePrice    : int  208500 181500 223500 140000 250000 143000 307000 200000 129900 118000 ...
```

Part 3A: Descriptive and Inferential Statistics

Provide univariate descriptive statistics and appropriate plots for the training data set. Provide a scatterplot matrix for at least two of the independent variables and the dependent variable. Derive a correlation matrix for any three quantitative variables in the dataset. Test the hypotheses that the correlations between each pairwise set of variables is 0 and provide an 80% confidence interval. Discuss the meaning of your analysis. Would you be worried about familywise error? Why or why not?

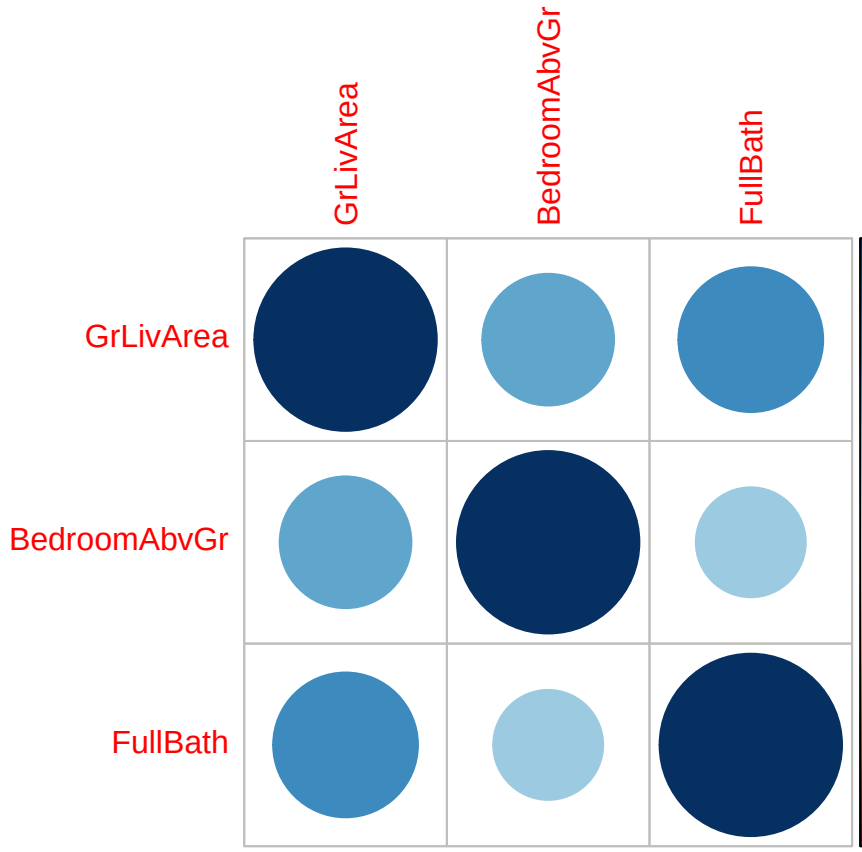
Part 3A1: Scatterplot Matrix The dependent variable is SalePrice, and I'll select the total area (GrLivArea), the number of bedrooms (BedroomAbvGr), and the number of full bathrooms (FullBath).

```
pairs(train_house[,c('GrLivArea','BedroomAbvGr','FullBath','SalePrice')])
```



The number of bedrooms and bathrooms, and the total square footage, correlate with the sale price, which is what we intuitively believe. It's interesting to see that the number of bedrooms doesn't necessarily correlate directly with sale price, but the number of bathrooms does.

```
library(corrplot)
train_house_cor <- cor(train_house[,c('GrLivArea','BedroomAbvGr','FullBath')])
corrplot(train_house_cor)
```



Part 3A2: Correlation Matrix

The three variables selected are vary highly correlated to each other. Perform correlation test with 80% confidence interval to determine if the correleation between these variables is 0.

```
cor.test(train_house$GrLivArea,train_house$BedroomAbvGr,conf.level = .8)
```

```
##
##  Pearson's product-moment correlation
##
## data:  train_house$GrLivArea and train_house$BedroomAbvGr
## t = 23.323, df = 1458, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 80 percent confidence interval:
##  0.4963921 0.5452915
## sample estimates:
##       cor
## 0.5212695
```

```
cor.test(train_house$GrLivArea,train_house$FullBath,conf.level = .8)
```

```
##
```

```
##  Pearson's product-moment correlation
##
## data:  train_house$GrLivArea and train_house$FullBath
## t = 30.977, df = 1458, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 80 percent confidence interval:
##  0.6093339 0.6498331
## sample estimates:
##       cor
## 0.6300116
```

```
cor.test(train_house$BedroomAbvGr,train_house$FullBath,conf.level = .8)
```

```
##
##  Pearson's product-moment correlation
##
## data:  train_house$BedroomAbvGr and train_house$FullBath
## t = 14.887, df = 1458, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 80 percent confidence interval:
##  0.3337593 0.3920342
## sample estimates:
##      cor
## 0.363252
```

These results indicate that the variables are highly related to themselves, which may impact the model using these variables to predict the sale price. We saw how the number of bedrooms may not be an indicator of sale price in the scatterplot matrix, but we see how the number of bedrooms is highly correlated to square footage.

Part 3B: Linear Algebra and Correlation

Invert your correlation matrix from above. This is known as the precision matrix and contains variance inflation factors on the diagonal. Multiply the correlation matrix by the precision matrix, and then multiply the precision matrix by the correlation matrix. Conduct LU decomposition on the matrix.

```
train_house_precision <- solve(train_house_cor)
train_house_precision
```

Part 3B1: Create Precision Matrix

```
##               GrLivArea BedroomAbvGr    FullBath
## GrLivArea     1.9818466  -0.66761829 -1.00607279
## BedroomAbvGr -0.6676183   1.37690845 -0.07955743
## FullBath     -1.0060728  -0.07955743  1.66273697
```

```
train_house_cor %*% train_house_precision
```

Part 3B2: Correlation Matrix × Precision Matrix

```
##                  GrLivArea  BedroomAbvGr      FullBath
## GrLivArea     1.000000e+00 -2.081668e-17  0.000000e+00
## BedroomAbvGr -5.551115e-17  1.000000e+00 -1.110223e-16
## FullBath      2.220446e-16  2.775558e-17  1.000000e+00
```

```
train_house_precision %*% train_house_cor
```

Part 3B3: Precision Matrix × Correlation Matrix

```
##                  GrLivArea  BedroomAbvGr     FullBath
## GrLivArea     1.000000e+00 -1.665335e-16 0.000000e+00
## BedroomAbvGr -2.081668e-17  1.000000e+00 2.775558e-17
## FullBath      0.000000e+00  0.000000e+00 1.000000e+00
```

```
library(matrixcalc)
```

Part 3B4: LU Decomposition

```
##
## Attaching package: 'matrixcalc'

## The following object is masked from 'package:igraph':
##
##     %s%
```

```
train_house_precision_LU <- lu.decomposition(train_house_precision)
train_house_precision_LU
```

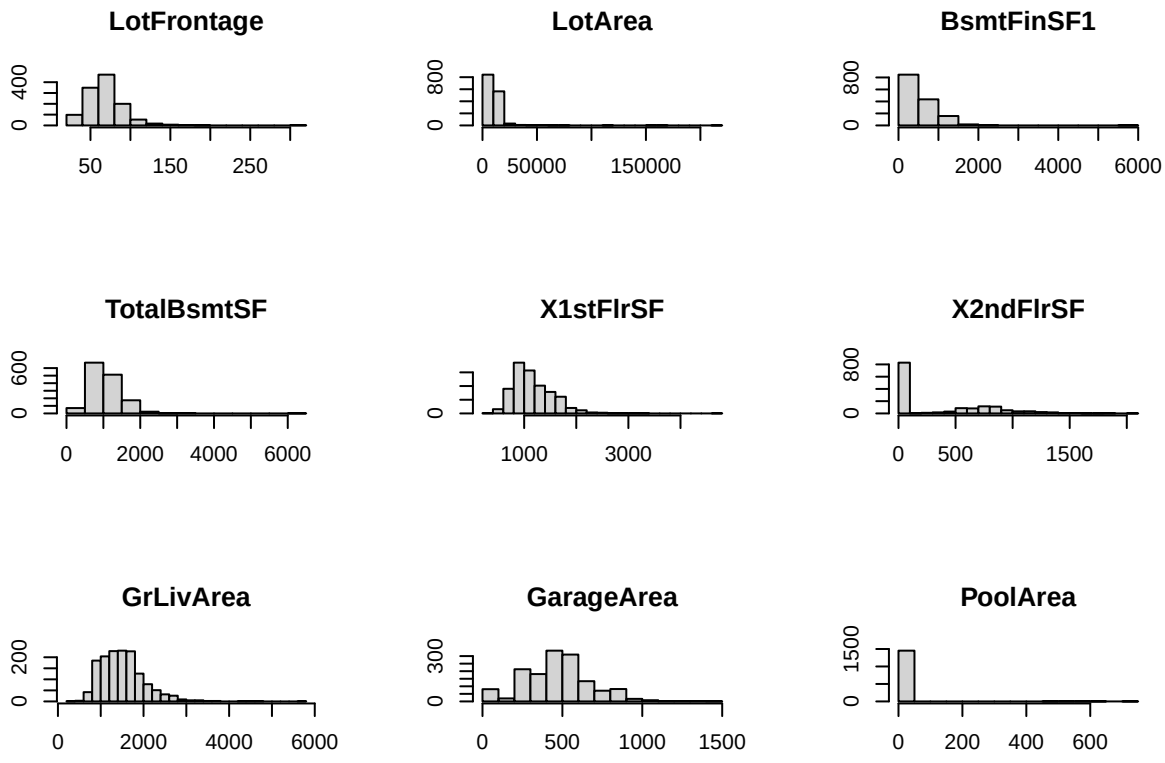
```
## $L
##            [,1]      [,2] [,3]
## [1,]  1.0000000  0.000000    0
## [2,] -0.3368668  1.000000    0
## [3,] -0.5076441 -0.363252    1
##
## $U
##          [,1]       [,2]       [,3]
## [1,] 1.981847 -0.6676183 -1.0060728
## [2,] 0.000000  1.1520100 -0.4184699
## [3,] 0.000000  0.0000000  1.0000000
```

Part 3C: Calculus-Based Probability & Statistics

Many times, it makes sense to fit a closed form distribution to data. Select a variable in the kaggle training dataset that is skewed to the right, shift it so that the minimum value is absolutely above zero if necessary. Then load the MASS package and run fitdistr to fit an exponential probability density function. Find the optimal value of λ for this distribution, and then take 1000 samples from this exponential distribution using this value. Plot a histogram and compare it with a histogram of your original variable. Using the exponential pdf, find the 5^{th} and 95^{th} percentiles using the cumulative distribution function. Also generate a 95% confidence interval from the empirical data, assuming normality. Finally, provide the empirical 5^{th} percentile and 95^{th} percentile of the data.

Part 3C1: Identify Right-Skewed Variable Plot a few of the variables that should contain right-skewed data.

```
par(mfrow=c(3,3))
hist(train_house$LotFrontage, xlab="", ylab = '', main="LotFrontage", breaks = 20)
hist(train_house$LotArea, xlab="", ylab = '', main="LotArea", breaks = 20)
hist(train_house$BsmtFinSF1, xlab="", ylab = '', main="BsmtFinSF1", breaks = 20)
hist(train_house$TotalBsmtSF, xlab="", ylab = '', main="TotalBsmtSF", breaks = 20)
hist(train_house$X1stFlrSF, xlab="", ylab = '', main="X1stFlrSF", breaks = 20)
hist(train_house$X2ndFlrSF, xlab="", ylab = '', main="X2ndFlrSF", breaks = 20)
hist(train_house$GrLivArea, xlab="", ylab = '', main="GrLivArea", breaks = 20)
hist(train_house$GarageArea, xlab="", ylab = '', main="GarageArea", breaks = 20)
hist(train_house$PoolArea, xlab="", ylab = '', main="PoolArea", breaks = 20)
```



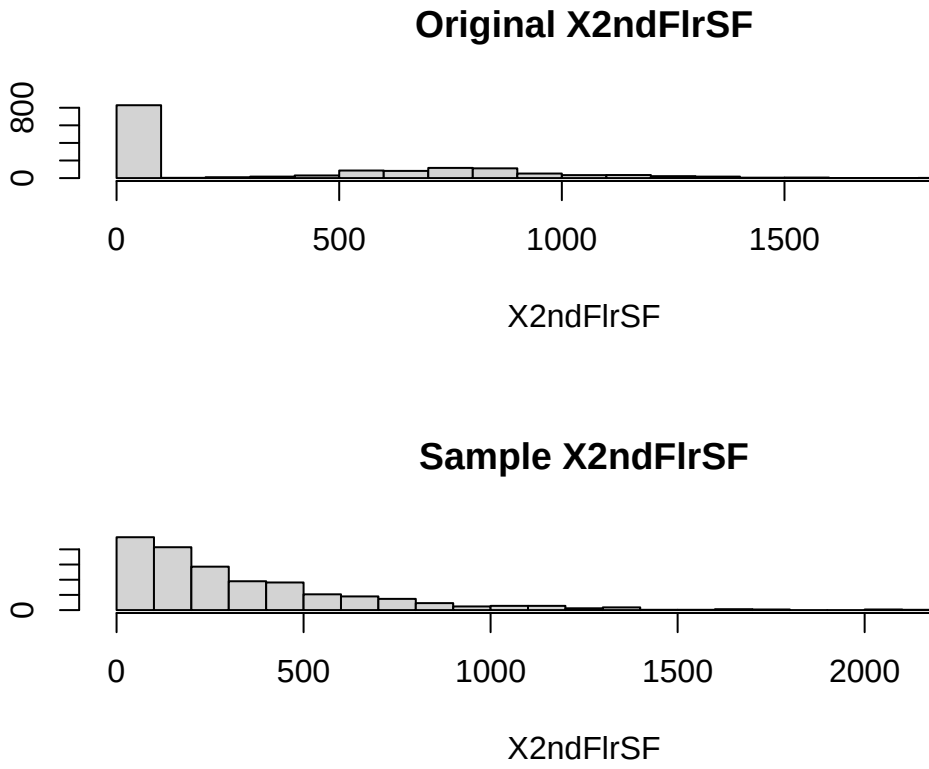
Let's select 2^{nd} floor square footage (X2ndFlrSF), which is strictly greater than or equal to 0.

```
library(MASS)

lambda <- fitdistr(train_house$X2ndFlrSF, densfun = 'exponential')
lambda
```

Part 3C2: Fit Exponential Distribution

```
##        rate
##   2.881907e-03
##  (7.542295e-05)
```

```
sample <- rexp(1000, lambda$estimate)

par(mfrow=c(2,1))
hist(train_house$X2ndFlrSF, xlab="X2ndFlrSF", ylab = '', main="Original X2ndFlrSF", breaks = 20)
hist(sample, xlab="X2ndFlrSF", ylab = '', main="Sample X2ndFlrSF", breaks = 20)
```



Part 3C3: Sample Distribution

Part 3C4: Distribution Percentiles Calculate the 5^{th} and 95^{th} percentiles of the exponential distribution.

```
exp_5th <- qexp(.05, lambda$estimate)
exp_95th <- qexp(.95, lambda$estimate)

print(c(exp_5th,exp_95th))
```

```
## [1]   17.79839 1039.49653
```

Compare with the 5^{th} and 95^{th} percentiles of the empirical data.

```
emp_5th <- quantile(train_house$X2ndFlrSF, .05)
emp_95th <- quantile(train_house$X2ndFlrSF, .95)

print(c(emp_5th,emp_95th))
```

```
##      5%     95%
##    0.00 1141.05
```

Part 3D: Modeling

Build some type of multiple regression model and submit your model to the competition board. Provide your complete model summary and results with analysis.

Part 3D1: Data Cleanup Resolve NA values and convert categorical variables into numeric.

For each of the categorical variables, we'll order the category dummy value in order from lowest average sale price to highest. This function will be used to evaluate the ordering of each category.

```
examine <- function(col) {
  print(train_house %>% group_by_at(col) %>% summarise(AvgSalePrice = mean(SalePrice), n = n()) %>% arra
}

texamine <- function(col) {
  print(test_house %>% group_by_at(col) %>% summarise(n = n()) %>% arrange(n))
}

examine('Neighborhood')
```

```
## # A tibble: 25 x 3
##    Neighborhood AvgSalePrice     n
##    <chr>               <dbl> <int>
##  1 MeadowV            98576.    17
##  2 IDOTRR            100124.    37
##  3 BrDale            104494.    16
##  4 BrkSide           124834.    58
##  5 Edwards           128220.   100
##  6 OldTown           128225.   113
##  7 Sawyer            136793.    74
##  8 Blueste           137500      2
##  9 SWISU             142591.    25
## 10 NPkVill           142694.     9
## # ... with 15 more rows
```

Creating a cleanup function that can be used on both the training and test sets to ensure the are cleaned the same way.

```
house_clean <- function(df) {
  df <- df %>%
    mutate(MSSubClass = case_when(MSSubClass == 30 ~ 0, MSSubClass == 180 ~ 1,
                                  MSSubClass == 45 ~ 2, MSSubClass == 190 ~ 3,
                                  MSSubClass == 90 ~ 4, MSSubClass == 160 ~ 5,
                                  MSSubClass == 50 ~ 6, MSSubClass == 85 ~ 7,
                                  MSSubClass == 40 ~ 8, MSSubClass == 70 ~ 9,
                                  MSSubClass == 80 ~ 10, MSSubClass == 20 ~ 11,
                                  MSSubClass == 75 ~ 12, MSSubClass == 120 ~ 13,
                                  MSSubClass == 150 ~ 14, MSSubClass == 60 ~ 15)) %>%
    mutate(MSZoning = case_when(MSZoning == 'A' ~ 0, MSZoning == 'C' ~ 1,
                                MSZoning == 'C (all)' ~ 1, MSZoning == 'I' ~ 2,
                                MSZoning == 'RM' ~ 3, MSZoning == 'RH' ~ 4,
                                MSZoning == 'RL' ~ 5, MSZoning == 'RP' ~ 6,
                                MSZoning == 'FV' ~ 7, is.na(MSZoning) ~ 0)) %>%
    mutate(Street = case_when(Street == 'Grvl' ~ 0, Street == 'Pave' ~ 1, is.na(Street) ~ 0)) %>%
    mutate(Alley = case_when(Alley == 'Grvl' ~ 0, Alley == 'Pave' ~ 1, is.na(Alley) ~ 2)) %>%
    mutate(LotShape = case_when(LotShape == 'Reg' ~ 0, LotShape == 'IR1' ~ 1,
                                LotShape == 'IR3' ~ 2, LotShape == 'IR2' ~ 3)) %>%
    mutate(LandContour = case_when(LandContour == 'Bnk' ~ 0, LandContour == 'Lvl' ~ 1,
                                   LandContour == 'Low' ~ 2, LandContour == 'HLS' ~ 3)) %>%
    mutate(Utilities = case_when(Utilities == 'ELO' ~ 0, Utilities == 'NoSeWa' ~ 1,
                                 Utilities == 'NoSewr' ~ 2, Utilities == 'AllPub' ~ 3, is.na(Utilities)
    mutate(LotConfig = case_when(LotConfig == 'Inside' ~ 0, LotConfig == 'FR2' ~ 1,
                                 LotConfig == 'Corner' ~ 2, LotConfig == 'FR3' ~ 3,
                                 LotConfig == 'CulDSac' ~ 4)) %>%
    mutate(LandSlope = case_when(LandSlope == 'Gtl' ~ 0, LandSlope == 'Mod' ~ 1,
                                 LandSlope == 'Sev' ~ 2)) %>%
    mutate(Neighborhood = case_when(Neighborhood == 'MeadowV' ~ 0, Neighborhood == 'IDOTRR' ~ 1,
                                    Neighborhood == 'BrDale' ~ 2, Neighborhood == 'BrkSide' ~ 3,
                                    Neighborhood == 'Edwards' ~ 4, Neighborhood == 'OldTown' ~ 5,
                                    Neighborhood == 'Sawyer' ~ 6, Neighborhood == 'Blueste' ~ 7,
                                    Neighborhood == 'SWISU' ~ 8, Neighborhood == 'NPkVill' ~ 9,
                                    Neighborhood == 'NAmes' ~ 10, Neighborhood == 'Mitchel' ~ 11,
                                    Neighborhood == 'SawyerW' ~ 12, Neighborhood == 'NWAmes' ~ 13,
                                    Neighborhood == 'Gilbert' ~ 14, Neighborhood == 'Blmngtn' ~ 15,
                                    Neighborhood == 'CollgCr' ~ 16, Neighborhood == 'Crawfor' ~ 17,
                                    Neighborhood == 'ClearCr' ~ 18, Neighborhood == 'Somerst' ~ 19,
                                    Neighborhood == 'Veenker' ~ 20, Neighborhood == 'Timber' ~ 21,
                                    Neighborhood == 'StoneBr' ~ 22, Neighborhood == 'NridgHt' ~ 23,
                                    Neighborhood == 'NoRidge' ~ 24)) %>%
    mutate(Condition1 = case_when(Condition1 == 'Artery' ~ 0, Condition1 == 'RRAe' ~ 1,
                                  Condition1 == 'Feedr' ~ 2, Condition1 == 'RRAn' ~ 3,
                                  Condition1 == 'Norm' ~ 4, Condition1 == 'RRNe' ~ 5,
                                  Condition1 == 'RRNn' ~ 6, Condition1 == 'PosN' ~ 7,
                                  Condition1 == 'PosA' ~ 8)) %>%
    mutate(Condition2 = case_when(Condition2 == 'Artery' ~ 0, Condition2 == 'RRAe' ~ 1,
                                  Condition2 == 'Feedr' ~ 2, Condition2 == 'RRAn' ~ 3,
                                  Condition2 == 'Norm' ~ 4, Condition2 == 'RRNe' ~ 5,
                                  Condition2 == 'RRNn' ~ 6, Condition2 == 'PosN' ~ 7,
                                  Condition2 == 'PosA' ~ 8)) %>%
```

```
    mutate(BldgType = case_when(BldgType == '2fmCon' ~ 0, BldgType == 'Twnhs' ~ 1,
                                BldgType == 'Duplex' ~ 2, BldgType == 'TwnhsE' ~ 3,
                                BldgType == '1Fam' ~ 4)) %>%
    mutate(HouseStyle = case_when(HouseStyle == '1.5Unf' ~ 0, HouseStyle == 'SFoyer' ~ 1,
                                  HouseStyle == '1.5Fin' ~ 2, HouseStyle == '2.5Unf' ~ 3,
                                  HouseStyle == 'SLvl' ~ 4, HouseStyle == '1Story' ~ 5,
                                  HouseStyle == '2Story' ~ 6, HouseStyle == '2.5Fin' ~ 7)) %>%
    mutate(RoofStyle = case_when(RoofStyle == 'Gambrel' ~ 0, RoofStyle == 'Gable' ~ 1,
                                 RoofStyle == 'Mansard' ~ 2, RoofStyle == 'Flat' ~ 3,
                                 RoofStyle == 'Hip' ~ 4, RoofStyle == 'Shed' ~ 5)) %>%
    mutate(RoofMatl = case_when(RoofMatl == 'Roll' ~ 0, RoofMatl == 'ClyTile' ~ 1,
                                RoofMatl == 'CompShg' ~ 2, RoofMatl == 'Metal' ~ 3,
                                RoofMatl == 'Tar&Grv' ~ 4, RoofMatl == 'WdShake' ~ 5,
                                RoofMatl == 'Membran' ~ 6, RoofMatl == 'WdShngl' ~ 7)) %>%
    mutate(Exterior1st = case_when(Exterior1st == 'BrkComm' ~ 0, Exterior1st == 'AsphShn' ~ 1,
                                   Exterior1st == 'CBlock' ~ 2, Exterior1st == 'AsbShng' ~ 3,
                                   Exterior1st == 'MetalSd' ~ 4, Exterior1st == 'Wd Sdng' ~ 5,
                                   Exterior1st == 'WdShing' ~ 6, Exterior1st == 'PreCast' ~ 7,
                                   Exterior1st == 'Stucco' ~ 8, Exterior1st == 'HdBoard' ~ 9,
                                   Exterior1st == 'Plywood' ~ 10, Exterior1st == 'BrkFace' ~ 11,
                                   Exterior1st == 'VinylSd' ~ 12, Exterior1st == 'CemntBd' ~ 13,
                                   Exterior1st == 'Stone' ~ 14, Exterior1st == 'ImStucc' ~ 15,
                                   Exterior1st == 'Other' ~ 16, is.na(Exterior1st) ~ 0)) %>%
    mutate(Exterior2nd = case_when(Exterior2nd == 'Brk Cmn' ~ 0, Exterior2nd == 'AsphShn' ~ 1,
                                   Exterior2nd == 'CBlock' ~ 2, Exterior2nd == 'AsbShng' ~ 3,
                                   Exterior2nd == 'MetalSd' ~ 4, Exterior2nd == 'Wd Sdng' ~ 5,
                                   Exterior2nd == 'Wd Shng' ~ 6, Exterior2nd == 'PreCast' ~ 7,
                                   Exterior2nd == 'Stucco' ~ 8, Exterior2nd == 'HdBoard' ~ 9,
                                   Exterior2nd == 'Plywood' ~ 10, Exterior2nd == 'BrkFace' ~ 11,
                                   Exterior2nd == 'VinylSd' ~ 12, Exterior2nd == 'CmentBd' ~ 13,
                                   Exterior2nd == 'Stone' ~ 14, Exterior2nd == 'ImStucc' ~ 15,
                                   Exterior2nd == 'Other' ~ 16, is.na(Exterior2nd) ~ 0)) %>%
    mutate(MasVnrType = case_when(MasVnrType == 'CBlock' ~ 0, MasVnrType == 'BrkCmn' ~ 1,
                                  MasVnrType == 'None' ~ 2, MasVnrType == 'BrkFace' ~ 3,
                                  MasVnrType == 'Stone' ~ 4, is.na(MasVnrType) ~ 0)) %>%
    mutate(ExterQual = case_when(ExterQual == 'Po' ~ 0, ExterQual == 'Fa' ~ 1,
                                 ExterQual == 'TA' ~ 2, ExterQual == 'Gd' ~ 3,
                                 ExterQual == 'Ex' ~ 4, is.na(ExterQual) ~ 2)) %>%
    mutate(ExterCond = case_when(ExterCond == 'Po' ~ 0, ExterCond == 'Fa' ~ 1,
                                 ExterCond == 'TA' ~ 2, ExterCond == 'Gd' ~ 3,
                                 ExterCond == 'Ex' ~ 4, is.na(ExterCond) ~ 2)) %>%
    mutate(Foundation = case_when(Foundation == 'Slab' ~ 0, Foundation == 'BrkTil' ~ 1,
                                  Foundation == 'CBlock' ~ 2, Foundation == 'Stone' ~ 3,
                                  Foundation == 'Wood' ~ 4, Foundation == 'PConc' ~ 5)) %>%
    mutate(BsmtQual = case_when(BsmtQual == 'Po' ~ 0, is.na(BsmtQual) ~ 1,
                                BsmtQual == 'Fa' ~ 2, BsmtQual == 'TA' ~ 3,
                                BsmtQual == 'Gd' ~ 4, BsmtQual == 'Ex' ~ 5)) %>%
    mutate(BsmtCond = case_when(BsmtCond == 'Po' ~ 0, is.na(BsmtCond) ~ 1,
                                BsmtCond == 'Fa' ~ 2, BsmtCond == 'TA' ~ 3,
                                BsmtCond == 'Gd' ~ 4, BsmtCond == 'Ex' ~ 5)) %>%
    mutate(BsmtExposure = case_when(is.na(BsmtExposure) ~ 0, BsmtExposure == 'No' ~ 1,
                                BsmtExposure == 'Mn' ~ 2, BsmtExposure == 'Av' ~ 3,
                                BsmtExposure == 'Gd' ~ 4)) %>%
```

```
    mutate(BsmtFinType1 = case_when(is.na(BsmtFinType1) ~ 0, BsmtFinType1 == 'Rec' ~ 1,
                                    BsmtFinType1 == 'BLQ' ~ 2, BsmtFinType1 == 'LwQ' ~ 3,
                                    BsmtFinType1 == 'ALQ' ~ 4, BsmtFinType1 == 'Unf' ~ 5,
                                    BsmtFinType1 == 'GLQ' ~ 6)) %>%
    mutate(BsmtFinType2 = case_when(is.na(BsmtFinType2) ~ 0, BsmtFinType2 == 'Rec' ~ 1,
                                    BsmtFinType2 == 'BLQ' ~ 2, BsmtFinType2 == 'LwQ' ~ 3,
                                    BsmtFinType2 == 'ALQ' ~ 4, BsmtFinType2 == 'Unf' ~ 5,
                                    BsmtFinType2 == 'GLQ' ~ 6)) %>%
    mutate(Heating = case_when(Heating == 'Floor' ~ 0, Heating == 'Grav' ~ 1,
                               Heating == 'Wall' ~ 2, Heating == 'OthW' ~ 3,
                               Heating == 'GasW' ~ 4, Heating == 'GasA' ~ 5)) %>%
    mutate(HeatingQC = case_when(HeatingQC == 'Po' ~ 0, HeatingQC == 'Fa' ~ 1,
                                 HeatingQC == 'TA' ~ 2, HeatingQC == 'Gd' ~ 3,
                                 HeatingQC == 'Ex' ~ 4)) %>%
    mutate(CentralAir = case_when(CentralAir == 'N' ~ 0, CentralAir == 'Y' ~ 1)) %>%
    mutate(Electrical = case_when(Electrical == 'Mix' ~ 0, Electrical == 'FuseP' ~ 1,
                                  Electrical == 'FuseF' ~ 2, Electrical == 'FuseA' ~ 3,
                                  Electrical == 'FuseA' ~ 4, Electrical == 'SBrkr' ~ 4,
                                  is.na(Electrical) ~ 3)) %>%
    mutate(KitchenQual = case_when(KitchenQual == 'Po' ~ 0, KitchenQual == 'Fa' ~ 1,
                                   KitchenQual == 'TA' ~ 2, KitchenQual == 'Gd' ~ 3,
                                   KitchenQual == 'Ex' ~ 4, is.na(KitchenQual) ~ 2)) %>%
    mutate(Functional = case_when(Functional == 'Sal' ~ 0, Functional == 'Sev' ~ 1,
                                  Functional == 'Maj2' ~ 2, Functional == 'Maj1' ~ 3,
                                  Functional == 'Mod' ~ 4, Functional == 'Min2' ~ 5,
                                  Functional == 'Min1' ~ 6, Functional == 'Typ' ~ 7,
                                  is.na(Functional) ~ 7)) %>%
    mutate(FireplaceQu = case_when(FireplaceQu == 'Po' ~ 0, is.na(FireplaceQu) ~ 1,
                                   FireplaceQu == 'Fa' ~ 2, FireplaceQu == 'TA' ~ 3,
                                   FireplaceQu == 'Gd' ~ 4, FireplaceQu == 'Ex' ~ 5)) %>%
    mutate(GarageType = case_when(is.na(GarageType) ~ 0, GarageType == 'CarPort' ~ 1,
                                  GarageType == 'Detchd' ~ 2, GarageType == '2Types' ~ 3,
                                  GarageType == 'Basment' ~ 4, GarageType == 'Attchd' ~ 5,
                                  GarageType == 'BuiltIn' ~ 6)) %>%
    mutate(GarageFinish = case_when(is.na(GarageFinish) ~ 0, GarageFinish == 'Unf' ~ 1,
                                    GarageFinish == 'RFn' ~ 2, GarageFinish == 'Fin' ~ 3)) %>%
    mutate(GarageQual = case_when(GarageQual == 'Po' ~ 0, is.na(GarageQual) ~ 1,
                                  GarageQual == 'Fa' ~ 2, GarageQual == 'TA' ~ 3,
                                  GarageQual == 'Gd' ~ 4, GarageQual == 'Ex' ~ 5)) %>%
    mutate(GarageCond = case_when(GarageCond == 'Po' ~ 0, is.na(GarageCond) ~ 1,
                                  GarageCond == 'Fa' ~ 2, GarageCond == 'TA' ~ 3,
                                  GarageCond == 'Gd' ~ 4, GarageCond == 'Ex' ~ 5)) %>%
    mutate(PavedDrive = case_when(PavedDrive == 'N' ~ 0, PavedDrive == 'P' ~ 1,
                                  PavedDrive == 'Y' ~ 2)) %>%
    mutate(PoolQC = case_when(is.na(PoolQC) ~ 0, PoolQC == 'Fa' ~ 1,
                              PoolQC == 'TA' ~ 2, PoolQC == 'Gd' ~ 3,
                              PoolQC == 'Ex' ~ 4)) %>%
    mutate(Fence = case_when(is.na(Fence) ~ 0, Fence == 'MnWw' ~ 1,
                             Fence == 'GdWo' ~ 2, Fence == 'MnPrv' ~ 3,
                             Fence == 'GdPrv' ~ 4)) %>%
    mutate(MiscFeature = case_when(is.na(MiscFeature) ~ 0, MiscFeature == 'Othr' ~ 1,
                                   MiscFeature == 'Elev' ~ 2, MiscFeature == 'Shed' ~ 3,
                                   MiscFeature == 'Gar2' ~ 4, MiscFeature == 'TenC' ~ 5)) %>%
```

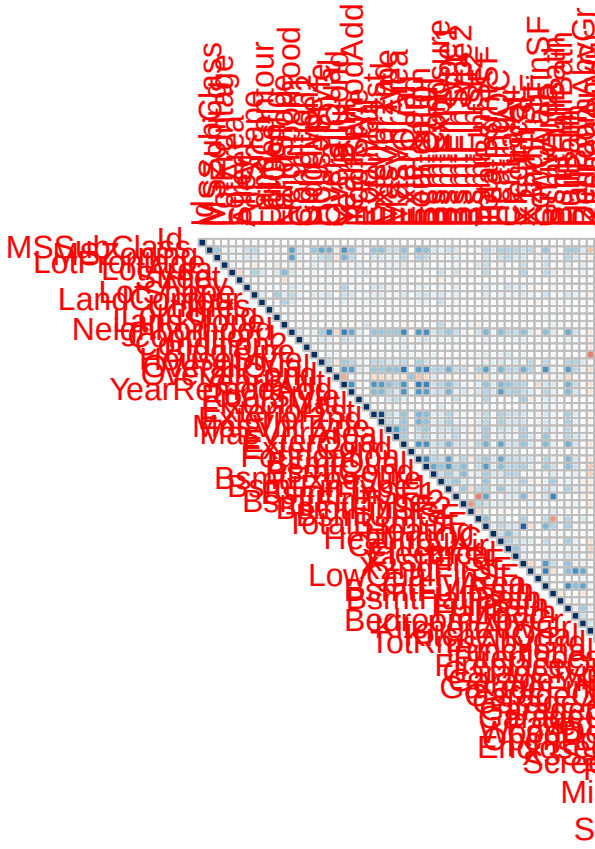
```
    mutate(SaleType = case_when(SaleType == 'Oth' ~ 0, SaleType == 'ConLD' ~ 1,
                                SaleType == 'ConLw' ~ 2, SaleType == 'COD' ~ 3,
                                SaleType == 'WD' ~ 4, SaleType == 'ConLI' ~ 5,
                                SaleType == 'VWD' ~ 6, SaleType == 'CWD' ~ 7,
                                SaleType == 'Con' ~ 8, SaleType == 'New' ~ 9,
                                is.na(SaleType) ~ 4)) %>%
    mutate(SaleCondition = case_when(SaleCondition == 'AdjLand' ~ 0, SaleCondition == 'Alloca' ~ 1,
                                     SaleCondition == 'Family' ~ 2, SaleCondition == 'Abnorml' ~ 3,
                                     SaleCondition == 'Partial' ~ 4, SaleCondition == 'Normal' ~ 5)) %>%
    mutate(MasVnrArea = ifelse(is.na(MasVnrArea),0,MasVnrArea)) %>%
    mutate(LotFrontage = ifelse(is.na(LotFrontage),0,LotFrontage)) %>%
    mutate(GarageYrBlt = ifelse(is.na(GarageYrBlt),0,GarageYrBlt)) %>%
    mutate(GarageCars = ifelse(is.na(GarageCars),0,GarageCars)) %>%
    mutate(BsmtFinSF1 = ifelse(is.na(BsmtFinSF1),0,BsmtFinSF1))

  return(df)
}

my_train_house <- house_clean(train_house)
str(my_train_house)
```

```
## 'data.frame':    1460 obs. of  81 variables:
##  $ Id           : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ MSSubClass   : num  15 11 15 9 15 6 11 15 6 3 ...
##  $ MSZoning     : num  5 5 5 5 5 5 5 5 3 5 ...
##  $ LotFrontage  : num  65 80 68 60 84 85 75 0 51 50 ...
##  $ LotArea      : int  8450 9600 11250 9550 14260 14115 10084 10382 6120 7420 ...
##  $ Street       : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ Alley        : num  2 2 2 2 2 2 2 2 2 2 ...
##  $ LotShape     : num  0 0 1 1 1 1 0 1 0 0 ...
##  $ LandContour  : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ Utilities    : num  3 3 3 3 3 3 3 3 3 3 ...
##  $ LotConfig    : num  0 1 0 2 1 0 0 2 0 2 ...
##  $ LandSlope    : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Neighborhood : num  16 20 16 17 24 11 19 13 5 3 ...
##  $ Condition1   : num  4 2 4 4 4 4 4 7 0 0 ...
##  $ Condition2   : num  4 4 4 4 4 4 4 4 4 0 ...
##  $ BldgType     : num  4 4 4 4 4 4 4 4 4 0 ...
##  $ HouseStyle   : num  6 5 6 6 6 2 5 6 2 0 ...
##  $ OverallQual  : int  7 6 7 7 8 5 8 7 7 5 ...
##  $ OverallCond  : int  5 8 5 5 5 5 5 6 5 6 ...
##  $ YearBuilt    : int  2003 1976 2001 1915 2000 1993 2004 1973 1931 1939 ...
##  $ YearRemodAdd : int  2003 1976 2002 1970 2000 1995 2005 1973 1950 1950 ...
##  $ RoofStyle    : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ RoofMatl     : num  2 2 2 2 2 2 2 2 2 2 ...
##  $ Exterior1st  : num  12 4 12 5 12 12 12 9 11 4 ...
##  $ Exterior2nd  : num  12 4 12 6 12 12 12 9 6 4 ...
##  $ MasVnrType   : num  3 2 3 2 3 2 4 4 2 2 ...
##  $ MasVnrArea   : num  196 0 162 0 350 0 186 240 0 0 ...
##  $ ExterQual    : num  3 2 3 2 3 2 3 2 2 2 ...
##  $ ExterCond    : num  2 2 2 2 2 2 2 2 2 2 ...
##  $ Foundation   : num  5 2 5 1 5 4 5 2 1 1 ...
##  $ BsmtQual     : num  4 4 4 3 4 4 5 4 3 3 ...
```

```
##  $ BsmtCond     : num  3 3 3 4 3 3 3 3 3 3 ...
##  $ BsmtExposure : num  1 4 2 1 3 1 3 2 1 1 ...
##  $ BsmtFinType1 : num  6 4 6 4 6 6 6 4 5 6 ...
##  $ BsmtFinSF1   : int  706 978 486 216 655 732 1369 859 0 851 ...
##  $ BsmtFinType2 : num  5 5 5 5 5 5 5 2 5 5 ...
##  $ BsmtFinSF2   : int  0 0 0 0 0 0 0 32 0 0 ...
##  $ BsmtUnfSF    : int  150 284 434 540 490 64 317 216 952 140 ...
##  $ TotalBsmtSF  : int  856 1262 920 756 1145 796 1686 1107 952 991 ...
##  $ Heating      : num  5 5 5 5 5 5 5 5 5 5 ...
##  $ HeatingQC    : num  4 4 4 3 4 4 4 4 3 4 ...
##  $ CentralAir   : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ Electrical   : num  4 4 4 4 4 4 4 4 2 4 ...
##  $ X1stFlrSF    : int  856 1262 920 961 1145 796 1694 1107 1022 1077 ...
##  $ X2ndFlrSF    : int  854 0 866 756 1053 566 0 983 752 0 ...
##  $ LowQualFinSF : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ GrLivArea    : int  1710 1262 1786 1717 2198 1362 1694 2090 1774 1077 ...
##  $ BsmtFullBath : int  1 0 1 1 1 1 1 1 0 1 ...
##  $ BsmtHalfBath : int  0 1 0 0 0 0 0 0 0 0 ...
##  $ FullBath     : int  2 2 2 1 2 1 2 2 2 1 ...
##  $ HalfBath     : int  1 0 1 0 1 1 0 1 0 0 ...
##  $ BedroomAbvGr : int  3 3 3 3 4 1 3 3 2 2 ...
##  $ KitchenAbvGr : int  1 1 1 1 1 1 1 1 2 2 ...
##  $ KitchenQual  : num  3 2 3 3 3 2 3 2 2 2 ...
##  $ TotRmsAbvGrd : int  8 6 6 7 9 5 7 7 8 5 ...
##  $ Functional   : num  7 7 7 7 7 7 7 7 6 7 ...
##  $ Fireplaces   : int  0 1 1 1 1 0 1 2 2 2 ...
##  $ FireplaceQu  : num  1 3 3 4 3 1 4 3 3 3 ...
##  $ GarageType   : num  5 5 5 2 5 5 5 5 2 5 ...
##  $ GarageYrBlt  : num  2003 1976 2001 1998 2000 ...
##  $ GarageFinish : num  2 2 2 1 2 1 2 2 1 2 ...
##  $ GarageCars   : int  2 2 2 3 3 2 2 2 2 1 ...
##  $ GarageArea   : int  548 460 608 642 836 480 636 484 468 205 ...
##  $ GarageQual   : num  3 3 3 3 3 3 3 3 2 4 ...
##  $ GarageCond   : num  3 3 3 3 3 3 3 3 3 3 ...
##  $ PavedDrive   : num  2 2 2 2 2 2 2 2 2 2 ...
##  $ WoodDeckSF   : int  0 298 0 0 192 40 255 235 90 0 ...
##  $ OpenPorchSF  : int  61 0 42 35 84 30 57 204 0 4 ...
##  $ EnclosedPorch: int  0 0 0 272 0 0 0 228 205 0 ...
##  $ X3SsnPorch   : int  0 0 0 0 0 320 0 0 0 0 ...
##  $ ScreenPorch  : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ PoolArea     : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ PoolQC       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Fence        : num  0 0 0 0 0 3 0 0 0 0 ...
##  $ MiscFeature  : num  0 0 0 0 0 3 0 3 0 0 ...
##  $ MiscVal      : int  0 0 0 0 0 700 0 350 0 0 ...
##  $ MoSold       : int  2 5 9 2 12 10 8 11 4 1 ...
##  $ YrSold       : int  2008 2007 2008 2006 2008 2009 2007 2009 2008 2008 ...
##  $ SaleType     : num  4 4 4 4 4 4 4 4 4 4 ...
##  $ SaleCondition: num  5 5 5 3 5 5 5 5 3 5 ...
##  $ SalePrice    : int  208500 181500 223500 140000 250000 143000 307000 200000 129900 118000 ...
```

```
cor_house <- cor(my_train_house)
```

```
corrplot(cor_house, type = 'upper')
```



Part 3D2: Identify variables to include in model

Filter out variables with less than .25 correlation.

```
cor_house_df <- data.frame(cor_house)['SalePrice'] %>%
  filter(SalePrice > .25) %>%
  arrange(desc(SalePrice))

cor_house_df
```

```
##              SalePrice
## SalePrice    1.0000000
## OverallQual  0.7909816
## GrLivArea    0.7086245
## Neighborhood 0.6968822
## ExterQual    0.6826392
## KitchenQual  0.6595997
## GarageCars   0.6404092
## GarageArea   0.6234314
## BsmtQual     0.6229247
## TotalBsmtSF  0.6135806
## X1stFlrSF    0.6058522
## FullBath     0.5606638
## GarageFinish 0.5492468
## TotRmsAbvGrd 0.5337232
```

```
## FireplaceQu  0.5259324
## YearBuilt    0.5228973
## YearRemodAdd 0.5071010
## Foundation   0.5055032
## GarageType   0.4891300
## MSSubClass   0.4862847
## MasVnrArea   0.4726145
## Fireplaces   0.4669288
## HeatingQC    0.4276487
## MasVnrType   0.3980961
## BsmtFinSF1   0.3864198
## BsmtFinType1 0.3809643
## BsmtExposure 0.3746962
## SaleType     0.3682631
## Exterior1st  0.3649094
## Exterior2nd  0.3598194
## WoodDeckSF   0.3244134
## X2ndFlrSF    0.3193338
## OpenPorchSF  0.3158562
## MSZoning     0.3130421
## HalfBath     0.2841077
## GarageQual   0.2792265
## HouseStyle   0.2730754
## LotShape     0.2656990
## LotArea      0.2638434
## GarageCond   0.2631897
## GarageYrBlt  0.2613664
## CentralAir   0.2513282
```

Part 3D3: Create Linear Model Build the model with all of the variables that have moderate correlation.

```
(lm_cols <- rownames(cor_house_df %>% arrange()))
```

```
##  [1] "SalePrice"    "OverallQual"  "GrLivArea"    "Neighborhood" "ExterQual"
##  [6] "KitchenQual"  "GarageCars"   "GarageArea"   "BsmtQual"     "TotalBsmtSF"
## [11] "X1stFlrSF"    "FullBath"     "GarageFinish" "TotRmsAbvGrd" "FireplaceQu"
## [16] "YearBuilt"    "YearRemodAdd" "Foundation"   "GarageType"   "MSSubClass"
## [21] "MasVnrArea"   "Fireplaces"   "HeatingQC"    "MasVnrType"   "BsmtFinSF1"
## [26] "BsmtFinType1" "BsmtExposure" "SaleType"     "Exterior1st"  "Exterior2nd"
## [31] "WoodDeckSF"   "X2ndFlrSF"    "OpenPorchSF"  "MSZoning"     "HalfBath"
## [36] "GarageQual"   "HouseStyle"   "LotShape"     "LotArea"      "GarageCond"
## [41] "GarageYrBlt"  "CentralAir"
```

```
house_model.lm <- lm(data = my_train_house, formula = SalePrice ~ OverallQual + GrLivArea + Neighborhoo
                       BsmtQual + TotalBsmtSF + X1stFlrSF + FullBath + GarageFinish + TotRmsAbvGrd + Fi
                       GarageType + MSSubClass + MasVnrArea + Fireplaces + HeatingQC + MasVnrType + Bsm
                       Exterior1st + Exterior2nd + WoodDeckSF + X2ndFlrSF + OpenPorchSF + MSZoning + Ha
                       GarageCond + GarageYrBlt + CentralAir)

summary(house_model.lm)
```

```
##
```

```
## Call:
## lm(formula = SalePrice ~ OverallQual + GrLivArea + Neighborhood +
##     ExterQual + KitchenQual + GarageCars + GarageArea + BsmtQual +
##     TotalBsmtSF + X1stFlrSF + FullBath + GarageFinish + TotRmsAbvGrd +
##     FireplaceQu + YearBuilt + YearRemodAdd + Foundation + GarageType +
##     MSSubClass + MasVnrArea + Fireplaces + HeatingQC + MasVnrType +
##     BsmtFinSF1 + BsmtFinType1 + BsmtExposure + SaleType + Exterior1st +
##     Exterior2nd + WoodDeckSF + X2ndFlrSF + OpenPorchSF + MSZoning +
##     HalfBath + GarageQual + HouseStyle + LotShape + LotArea +
##     GarageCond + GarageYrBlt + CentralAir, data = my_train_house)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -436299  -15878    -645   14737  296453
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.567e+05  1.707e+05   1.503 0.132942
## OverallQual   1.147e+04  1.230e+03   9.333  < 2e-16 ***
## GrLivArea     2.863e+00  1.877e+01   0.153 0.878746
## Neighborhood  2.095e+03  2.417e+02   8.668  < 2e-16 ***
## ExterQual     9.938e+03  2.676e+03   3.714 0.000212 ***
## KitchenQual   1.090e+04  2.121e+03   5.141 3.12e-07 ***
## GarageCars    1.063e+04  2.882e+03   3.688 0.000234 ***
## GarageArea    7.638e+00  9.403e+00   0.812 0.416755
## BsmtQual      6.021e+03  2.038e+03   2.954 0.003190 **
## TotalBsmtSF  -5.318e-01  4.426e+00  -0.120 0.904373
## X1stFlrSF     4.292e+01  1.931e+01   2.223 0.026365 *
## FullBath     -2.174e+03  2.657e+03  -0.818 0.413510
## GarageFinish  1.466e+03  1.539e+03   0.952 0.341006
## TotRmsAbvGrd  1.910e+03  1.006e+03   1.899 0.057768 .
## FireplaceQu   1.324e+03  1.207e+03   1.097 0.272986
## YearBuilt    -2.111e+02  6.618e+01  -3.190 0.001452 **
## YearRemodAdd  1.823e+01  6.495e+01   0.281 0.779057
## Foundation   -5.854e+02  8.732e+02  -0.670 0.502674
## GarageType   -1.661e+02  8.892e+02  -0.187 0.851849
## MSSubClass    5.626e+02  3.299e+02   1.706 0.088304 .
## MasVnrArea    3.020e+01  6.420e+00   4.705 2.79e-06 ***
## Fireplaces    2.210e+03  2.518e+03   0.878 0.380331
## HeatingQC     2.063e+03  1.192e+03   1.730 0.083796 .
## MasVnrType   -2.850e+03  1.678e+03  -1.698 0.089656 .
## BsmtFinSF1    1.551e+01  2.445e+00   6.345 2.99e-10 ***
## BsmtFinType1 -5.369e+02  6.649e+02  -0.807 0.419581
## BsmtExposure  4.907e+03  9.877e+02   4.968 7.57e-07 ***
## SaleType      3.500e+03  6.781e+02   5.161 2.80e-07 ***
## Exterior1st   5.907e+02  6.393e+02   0.924 0.355624
## Exterior2nd  -2.107e+02  6.225e+02  -0.338 0.735073
## WoodDeckSF    2.285e+01  7.548e+00   3.027 0.002517 **
## X2ndFlrSF     3.127e+01  1.896e+01   1.649 0.099303 .
## OpenPorchSF  -8.068e-01  1.455e+01  -0.055 0.955798
## MSZoning      1.781e+02  1.235e+03   0.144 0.885325
## HalfBath      7.963e+02  2.592e+03   0.307 0.758685
## GarageQual    9.536e+03  4.502e+03   2.118 0.034320 *
## HouseStyle   -4.952e+02  7.607e+02  -0.651 0.515139
```

```
## LotShape      2.144e+03  1.457e+03   1.471 0.141509
## LotArea       2.559e-01  1.008e-01   2.538 0.011253 *
## GarageCond   -1.413e+03  4.214e+03  -0.335 0.737502
## GarageYrBlt  -1.824e+01  4.646e+00  -3.927 9.02e-05 ***
## CentralAir    3.590e+03  4.149e+03   0.865 0.387012
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32930 on 1418 degrees of freedom
## Multiple R-squared:  0.833,  Adjusted R-squared:  0.8282
## F-statistic: 172.5 on 41 and 1418 DF,  p-value: < 2.2e-16
```

Part 3D4: Backward Elimination I will proceed with backward elimination.

```
# Removing OpenPorchSF
house_model.lm <- lm(data = my_train_house, formula = SalePrice ~ OverallQual + GrLivArea + Neighborhoo
                       BsmtQual + TotalBsmtSF + X1stFlrSF + FullBath + GarageFinish + TotRmsAbvGrd + Fi
                       GarageType + MSSubClass + MasVnrArea + Fireplaces + HeatingQC + MasVnrType + Bsm
                       Exterior1st + Exterior2nd + WoodDeckSF + X2ndFlrSF + MSZoning + HalfBath + Garag
                       GarageCond + GarageYrBlt + CentralAir)

summary(house_model.lm)
```

```
##
## Call:
## lm(formula = SalePrice ~ OverallQual + GrLivArea + Neighborhood +
##     ExterQual + KitchenQual + GarageCars + GarageArea + BsmtQual +
##     TotalBsmtSF + X1stFlrSF + FullBath + GarageFinish + TotRmsAbvGrd +
##     FireplaceQu + YearBuilt + YearRemodAdd + Foundation + GarageType +
##     MSSubClass + MasVnrArea + Fireplaces + HeatingQC + MasVnrType +
##     BsmtFinSF1 + BsmtFinType1 + BsmtExposure + SaleType + Exterior1st +
##     Exterior2nd + WoodDeckSF + X2ndFlrSF + MSZoning + HalfBath +
##     GarageQual + HouseStyle + LotShape + LotArea + GarageCond +
##     GarageYrBlt + CentralAir, data = my_train_house)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -436286  -15859    -626   14756  296532
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.565e+05  1.706e+05   1.503 0.133018
## OverallQual   1.147e+04  1.228e+03   9.339  < 2e-16 ***
## GrLivArea     2.850e+00  1.876e+01   0.152 0.879234
## Neighborhood  2.096e+03  2.408e+02   8.707  < 2e-16 ***
## ExterQual     9.928e+03  2.669e+03   3.720 0.000207 ***
## KitchenQual   1.090e+04  2.120e+03   5.142 3.10e-07 ***
## GarageCars    1.064e+04  2.877e+03   3.698 0.000226 ***
## GarageArea    7.594e+00  9.366e+00   0.811 0.417616
## BsmtQual      6.015e+03  2.034e+03   2.957 0.003161 **
## TotalBsmtSF  -5.436e-01  4.419e+00  -0.123 0.902114
## X1stFlrSF     4.292e+01  1.930e+01   2.224 0.026322 *
## FullBath     -2.181e+03  2.653e+03  -0.822 0.411062
```

```
## GarageFinish  1.465e+03  1.538e+03   0.952 0.341026
## TotRmsAbvGrd  1.912e+03  1.004e+03   1.905 0.057015 .
## FireplaceQu   1.328e+03  1.205e+03   1.102 0.270501
## YearBuilt    -2.109e+02  6.604e+01  -3.194 0.001435 **
## YearRemodAdd  1.814e+01  6.491e+01   0.280 0.779899
## Foundation   -5.864e+02  8.727e+02  -0.672 0.501769
## GarageType   -1.645e+02  8.885e+02  -0.185 0.853116
## MSSubClass    5.606e+02  3.279e+02   1.710 0.087481 .
## MasVnrArea    3.022e+01  6.410e+00   4.714 2.67e-06 ***
## Fireplaces    2.202e+03  2.513e+03   0.876 0.381110
## HeatingQC     2.063e+03  1.192e+03   1.731 0.083745 .
## MasVnrType   -2.851e+03  1.677e+03  -1.700 0.089342 .
## BsmtFinSF1    1.551e+01  2.444e+00   6.348 2.93e-10 ***
## BsmtFinType1 -5.363e+02  6.646e+02  -0.807 0.419857
## BsmtExposure  4.908e+03  9.874e+02   4.970 7.50e-07 ***
## SaleType      3.498e+03  6.772e+02   5.165 2.74e-07 ***
## Exterior1st   5.909e+02  6.390e+02   0.925 0.355260
## Exterior2nd  -2.116e+02  6.221e+02  -0.340 0.733818
## WoodDeckSF    2.287e+01  7.533e+00   3.036 0.002443 **
## X2ndFlrSF     3.127e+01  1.895e+01   1.650 0.099213 .
## MSZoning      1.759e+02  1.234e+03   0.143 0.886655
## HalfBath      7.850e+02  2.583e+03   0.304 0.761215
## GarageQual    9.531e+03  4.499e+03   2.118 0.034318 *
## HouseStyle   -4.962e+02  7.602e+02  -0.653 0.514041
## LotShape      2.144e+03  1.457e+03   1.472 0.141328
## LotArea       2.558e-01  1.008e-01   2.538 0.011242 *
## GarageCond   -1.410e+03  4.212e+03  -0.335 0.737777
## GarageYrBlt  -1.823e+01  4.641e+00  -3.929 8.95e-05 ***
## CentralAir    3.606e+03  4.138e+03   0.871 0.383722
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32920 on 1419 degrees of freedom
## Multiple R-squared:  0.833,  Adjusted R-squared:  0.8283
## F-statistic: 176.9 on 40 and 1419 DF,  p-value: < 2.2e-16
```

```
# Removing TotalBsmtSF
house_model.lm <- lm(data = my_train_house, formula = SalePrice ~ OverallQual + GrLivArea + Neighborhood +
                       BsmtQual + X1stFlrSF + FullBath + GarageFinish + TotRmsAbvGrd + FireplaceQu + YearBuilt +
                       GarageType + MSSubClass + MasVnrArea + Fireplaces + HeatingQC + MasVnrType + BsmtFinSF1 +
                       Exterior1st + Exterior2nd + WoodDeckSF + X2ndFlrSF + MSZoning + HalfBath + GarageQual +
                       GarageCond + GarageYrBlt + CentralAir)

summary(house_model.lm)
```

```
##
## Call:
## lm(formula = SalePrice ~ OverallQual + GrLivArea + Neighborhood +
##     ExterQual + KitchenQual + GarageCars + GarageArea + BsmtQual +
##     X1stFlrSF + FullBath + GarageFinish + TotRmsAbvGrd + FireplaceQu +
##     YearBuilt + YearRemodAdd + Foundation + GarageType + MSSubClass +
##     MasVnrArea + Fireplaces + HeatingQC + MasVnrType + BsmtFinSF1 +
##     BsmtFinType1 + BsmtExposure + SaleType + Exterior1st + Exterior2nd +
##     WoodDeckSF + X2ndFlrSF + MSZoning + HalfBath + GarageQual +
```

```
##     HouseStyle + LotShape + LotArea + GarageCond + GarageYrBlt +
##     CentralAir, data = my_train_house)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -436994  -15907    -669   14710  296440
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.536e+05  1.689e+05   1.501 0.133575
## OverallQual   1.146e+04  1.225e+03   9.355  < 2e-16 ***
## GrLivArea     2.840e+00  1.875e+01   0.151 0.879632
## Neighborhood  2.098e+03  2.405e+02   8.723  < 2e-16 ***
## ExterQual     9.914e+03  2.665e+03   3.719 0.000207 ***
## KitchenQual   1.090e+04  2.119e+03   5.145 3.05e-07 ***
## GarageCars    1.067e+04  2.867e+03   3.721 0.000207 ***
## GarageArea    7.513e+00  9.340e+00   0.804 0.421297
## BsmtQual      5.920e+03  1.881e+03   3.147 0.001686 **
## X1stFlrSF     4.253e+01  1.903e+01   2.235 0.025588 *
## FullBath     -2.171e+03  2.650e+03  -0.819 0.412926
## GarageFinish  1.475e+03  1.535e+03   0.961 0.336809
## TotRmsAbvGrd  1.914e+03  1.004e+03   1.907 0.056749 .
## FireplaceQu   1.333e+03  1.204e+03   1.107 0.268335
## YearBuilt    -2.103e+02  6.584e+01  -3.194 0.001432 **
## YearRemodAdd  1.922e+01  6.430e+01   0.299 0.765056
## Foundation   -5.940e+02  8.702e+02  -0.683 0.494991
## GarageType   -1.630e+02  8.881e+02  -0.184 0.854375
## MSSubClass    5.585e+02  3.273e+02   1.706 0.088144 .
## MasVnrArea    3.017e+01  6.393e+00   4.718 2.61e-06 ***
## Fireplaces    2.206e+03  2.512e+03   0.878 0.379878
## HeatingQC     2.055e+03  1.190e+03   1.727 0.084347 .
## MasVnrType   -2.850e+03  1.676e+03  -1.700 0.089331 .
## BsmtFinSF1    1.545e+01  2.392e+00   6.461 1.42e-10 ***
## BsmtFinType1 -5.434e+02  6.619e+02  -0.821 0.411849
## BsmtExposure  4.905e+03  9.868e+02   4.971 7.48e-07 ***
## SaleType      3.496e+03  6.768e+02   5.166 2.74e-07 ***
## Exterior1st   5.878e+02  6.383e+02   0.921 0.357264
## Exterior2nd  -2.092e+02  6.216e+02  -0.337 0.736480
## WoodDeckSF    2.291e+01  7.524e+00   3.045 0.002370 **
## X2ndFlrSF     3.133e+01  1.894e+01   1.654 0.098338 .
## MSZoning      1.755e+02  1.233e+03   0.142 0.886823
## HalfBath      7.810e+02  2.581e+03   0.303 0.762278
## GarageQual    9.560e+03  4.491e+03   2.129 0.033463 *
## HouseStyle   -5.016e+02  7.586e+02  -0.661 0.508591
## LotShape      2.146e+03  1.456e+03   1.474 0.140722
## LotArea       2.549e-01  1.005e-01   2.537 0.011287 *
## GarageCond   -1.405e+03  4.210e+03  -0.334 0.738702
## GarageYrBlt  -1.828e+01  4.625e+00  -3.952 8.14e-05 ***
## CentralAir    3.580e+03  4.131e+03   0.867 0.386361
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32910 on 1420 degrees of freedom
## Multiple R-squared:  0.833,  Adjusted R-squared:  0.8284
```

```
## F-statistic: 181.6 on 39 and 1420 DF,  p-value: < 2.2e-16
```

```
# Removing MSZoning
house_model.lm <- lm(data = my_train_house, formula = SalePrice ~ OverallQual + GrLivArea + Neighborhood +
                       BsmtQual + X1stFlrSF + FullBath + GarageFinish + TotRmsAbvGrd + FireplaceQu + YearBuilt +
                       GarageType + MSSubClass + MasVnrArea + Fireplaces + HeatingQC + MasVnrType + BsmtFinSF1 +
                       Exterior1st + Exterior2nd + WoodDeckSF + X2ndFlrSF + HalfBath + GarageQual + HouseStyle +
                       GarageCond + GarageYrBlt + CentralAir)

summary(house_model.lm)
```

```
##
## Call:
## lm(formula = SalePrice ~ OverallQual + GrLivArea + Neighborhood +
##     ExterQual + KitchenQual + GarageCars + GarageArea + BsmtQual +
##     X1stFlrSF + FullBath + GarageFinish + TotRmsAbvGrd + FireplaceQu +
##     YearBuilt + YearRemodAdd + Foundation + GarageType + MSSubClass +
##     MasVnrArea + Fireplaces + HeatingQC + MasVnrType + BsmtFinSF1 +
##     BsmtFinType1 + BsmtExposure + SaleType + Exterior1st + Exterior2nd +
##     WoodDeckSF + X2ndFlrSF + HalfBath + GarageQual + HouseStyle +
##     LotShape + LotArea + GarageCond + GarageYrBlt + CentralAir,
##     data = my_train_house)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -436937  -15862    -712   14747  296347
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.515e+05  1.683e+05   1.495 0.135183
## OverallQual   1.146e+04  1.224e+03   9.358  < 2e-16 ***
## GrLivArea     2.724e+00  1.873e+01   0.145 0.884349
## Neighborhood  2.110e+03  2.231e+02   9.460  < 2e-16 ***
## ExterQual     9.912e+03  2.664e+03   3.720 0.000207 ***
## KitchenQual   1.090e+04  2.119e+03   5.146 3.04e-07 ***
## GarageCars    1.063e+04  2.854e+03   3.725 0.000203 ***
## GarageArea    7.587e+00  9.322e+00   0.814 0.415823
## BsmtQual      5.902e+03  1.876e+03   3.145 0.001695 **
## X1stFlrSF     4.263e+01  1.901e+01   2.243 0.025060 *
## FullBath     -2.131e+03  2.635e+03  -0.809 0.418754
## GarageFinish  1.465e+03  1.533e+03   0.956 0.339467
## TotRmsAbvGrd  1.917e+03  1.003e+03   1.911 0.056215 .
## FireplaceQu   1.328e+03  1.203e+03   1.104 0.269649
## YearBuilt    -2.088e+02  6.489e+01  -3.217 0.001324 **
## YearRemodAdd  1.911e+01  6.427e+01   0.297 0.766309
## Foundation   -5.961e+02  8.698e+02  -0.685 0.493203
## GarageType   -1.612e+02  8.877e+02  -0.182 0.855900
## MSSubClass    5.611e+02  3.266e+02   1.718 0.086046 .
## MasVnrArea    3.011e+01  6.380e+00   4.720 2.60e-06 ***
## Fireplaces    2.211e+03  2.511e+03   0.881 0.378542
## HeatingQC     2.058e+03  1.189e+03   1.730 0.083789 .
## MasVnrType   -2.868e+03  1.671e+03  -1.716 0.086443 .
## BsmtFinSF1    1.546e+01  2.389e+00   6.472 1.33e-10 ***
## BsmtFinType1 -5.522e+02  6.588e+02  -0.838 0.402032
```

```
## BsmtExposure  4.895e+03  9.840e+02   4.975 7.32e-07 ***
## SaleType      3.503e+03  6.745e+02   5.194 2.36e-07 ***
## Exterior1st   5.811e+02  6.363e+02   0.913 0.361300
## Exterior2nd  -2.095e+02  6.214e+02  -0.337 0.736031
## WoodDeckSF    2.290e+01  7.521e+00   3.045 0.002368 **
## X2ndFlrSF     3.139e+01  1.893e+01   1.658 0.097532 .
## HalfBath      8.282e+02  2.559e+03   0.324 0.746293
## GarageQual    9.579e+03  4.488e+03   2.135 0.032964 *
## HouseStyle   -5.011e+02  7.584e+02  -0.661 0.508906
## LotShape      2.147e+03  1.456e+03   1.475 0.140536
## LotArea       2.554e-01  1.004e-01   2.544 0.011074 *
## GarageCond   -1.421e+03  4.207e+03  -0.338 0.735658
## GarageYrBlt  -1.828e+01  4.624e+00  -3.954 8.07e-05 ***
## CentralAir    3.633e+03  4.113e+03   0.883 0.377134
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32900 on 1421 degrees of freedom
## Multiple R-squared:  0.833,  Adjusted R-squared:  0.8285
## F-statistic: 186.5 on 38 and 1421 DF,  p-value: < 2.2e-16
```

```
# Removing GrLivArea
house_model.lm <- lm(data = my_train_house, formula = SalePrice ~ OverallQual + Neighborhood + ExterQual
                       BsmtQual + X1stFlrSF + FullBath + GarageFinish + TotRmsAbvGrd + FireplaceQu + Yea
                       GarageType + MSSubClass + MasVnrArea + Fireplaces + HeatingQC + MasVnrType + Bsm
                       Exterior1st + Exterior2nd + WoodDeckSF + X2ndFlrSF + HalfBath + GarageQual + Hous
                       GarageCond + GarageYrBlt + CentralAir)

summary(house_model.lm)
```

```
##
## Call:
## lm(formula = SalePrice ~ OverallQual + Neighborhood + ExterQual +
##     KitchenQual + GarageCars + GarageArea + BsmtQual + X1stFlrSF +
##     FullBath + GarageFinish + TotRmsAbvGrd + FireplaceQu + YearBuilt +
##     YearRemodAdd + Foundation + GarageType + MSSubClass + MasVnrArea +
##     Fireplaces + HeatingQC + MasVnrType + BsmtFinSF1 + BsmtFinType1 +
##     BsmtExposure + SaleType + Exterior1st + Exterior2nd + WoodDeckSF +
##     X2ndFlrSF + HalfBath + GarageQual + HouseStyle + LotShape +
##     LotArea + GarageCond + GarageYrBlt + CentralAir, data = my_train_house)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -436897  -15868    -693   14732  296346
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.540e+05  1.673e+05   1.518 0.129223
## OverallQual   1.146e+04  1.223e+03   9.374  < 2e-16 ***
## Neighborhood  2.111e+03  2.230e+02   9.467  < 2e-16 ***
## ExterQual     9.911e+03  2.664e+03   3.721 0.000206 ***
## KitchenQual   1.090e+04  2.118e+03   5.148 3.00e-07 ***
## GarageCars    1.062e+04  2.852e+03   3.723 0.000204 ***
## GarageArea    7.642e+00  9.311e+00   0.821 0.411951
```

```
## BsmtQual      5.910e+03  1.875e+03   3.152 0.001653 **
## X1stFlrSF     4.532e+01  4.590e+00   9.873  < 2e-16 ***
## FullBath     -2.121e+03  2.633e+03  -0.805 0.420710
## GarageFinish  1.469e+03  1.532e+03   0.959 0.337854
## TotRmsAbvGrd  1.938e+03  9.924e+02   1.952 0.051078 .
## FireplaceQu   1.327e+03  1.202e+03   1.104 0.269780
## YearBuilt    -2.100e+02  6.431e+01  -3.265 0.001121 **
## YearRemodAdd  1.899e+01  6.424e+01   0.296 0.767639
## Foundation   -5.928e+02  8.692e+02  -0.682 0.495350
## GarageType   -1.707e+02  8.850e+02  -0.193 0.847068
## MSSubClass    5.634e+02  3.261e+02   1.728 0.084293 .
## MasVnrArea    3.009e+01  6.376e+00   4.719 2.60e-06 ***
## Fireplaces    2.208e+03  2.510e+03   0.880 0.379105
## HeatingQC     2.060e+03  1.189e+03   1.733 0.083324 .
## MasVnrType   -2.870e+03  1.671e+03  -1.718 0.086036 .
## BsmtFinSF1    1.546e+01  2.389e+00   6.473 1.32e-10 ***
## BsmtFinType1 -5.496e+02  6.583e+02  -0.835 0.403948
## BsmtExposure  4.893e+03  9.836e+02   4.975 7.31e-07 ***
## SaleType      3.502e+03  6.742e+02   5.194 2.36e-07 ***
## Exterior1st   5.797e+02  6.360e+02   0.911 0.362262
## Exterior2nd  -2.093e+02  6.212e+02  -0.337 0.736158
## WoodDeckSF    2.291e+01  7.518e+00   3.047 0.002353 **
## X2ndFlrSF     3.406e+01  4.593e+00   7.415 2.09e-13 ***
## HalfBath      8.289e+02  2.558e+03   0.324 0.745980
## GarageQual    9.624e+03  4.475e+03   2.150 0.031688 *
## HouseStyle   -5.012e+02  7.581e+02  -0.661 0.508620
## LotShape      2.153e+03  1.455e+03   1.480 0.139060
## LotArea       2.553e-01  1.004e-01   2.544 0.011070 *
## GarageCond   -1.416e+03  4.206e+03  -0.337 0.736494
## GarageYrBlt  -1.836e+01  4.592e+00  -3.998 6.72e-05 ***
## CentralAir    3.656e+03  4.108e+03   0.890 0.373732
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32890 on 1422 degrees of freedom
## Multiple R-squared:  0.833,  Adjusted R-squared:  0.8286
## F-statistic: 191.7 on 37 and 1422 DF,  p-value: < 2.2e-16
```

```
# Removing GarageType
house_model.lm <- lm(data = my_train_house, formula = SalePrice ~ OverallQual + Neighborhood + ExterQual
                       BsmtQual + X1stFlrSF + FullBath + GarageFinish + TotRmsAbvGrd + FireplaceQu + Yea
                       MSSubClass + MasVnrArea + Fireplaces + HeatingQC + MasVnrType + BsmtFinSF1 + Bsm
                       Exterior1st + Exterior2nd + WoodDeckSF + X2ndFlrSF + HalfBath + GarageQual + Hous
                       GarageCond + GarageYrBlt + CentralAir)

summary(house_model.lm)
```

```
##
## Call:
## lm(formula = SalePrice ~ OverallQual + Neighborhood + ExterQual +
##     KitchenQual + GarageCars + GarageArea + BsmtQual + X1stFlrSF +
##     FullBath + GarageFinish + TotRmsAbvGrd + FireplaceQu + YearBuilt +
##     YearRemodAdd + Foundation + MSSubClass + MasVnrArea + Fireplaces +
##     HeatingQC + MasVnrType + BsmtFinSF1 + BsmtFinType1 + BsmtExposure +
```

```
##     SaleType + Exterior1st + Exterior2nd + WoodDeckSF + X2ndFlrSF +
##     HalfBath + GarageQual + HouseStyle + LotShape + LotArea +
##     GarageCond + GarageYrBlt + CentralAir, data = my_train_house)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -436636  -15945    -702   14652  296493
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.576e+05  1.662e+05   1.550 0.121369
## OverallQual   1.146e+04  1.223e+03   9.377  < 2e-16 ***
## Neighborhood  2.109e+03  2.227e+02   9.470  < 2e-16 ***
## ExterQual     9.916e+03  2.663e+03   3.724 0.000204 ***
## KitchenQual   1.091e+04  2.117e+03   5.156 2.88e-07 ***
## GarageCars    1.065e+04  2.846e+03   3.742 0.000190 ***
## GarageArea    7.765e+00  9.286e+00   0.836 0.403155
## BsmtQual      5.912e+03  1.874e+03   3.154 0.001642 **
## X1stFlrSF     4.524e+01  4.569e+00   9.901  < 2e-16 ***
## FullBath     -2.140e+03  2.630e+03  -0.814 0.416065
## GarageFinish  1.391e+03  1.477e+03   0.942 0.346581
## TotRmsAbvGrd  1.939e+03  9.920e+02   1.955 0.050829 .
## FireplaceQu   1.320e+03  1.201e+03   1.099 0.272066
## YearBuilt    -2.118e+02  6.359e+01  -3.331 0.000887 ***
## YearRemodAdd  1.909e+01  6.422e+01   0.297 0.766280
## Foundation   -5.898e+02  8.687e+02  -0.679 0.497262
## MSSubClass    5.495e+02  3.179e+02   1.728 0.084156 .
## MasVnrArea    3.014e+01  6.368e+00   4.732 2.44e-06 ***
## Fireplaces    2.191e+03  2.507e+03   0.874 0.382265
## HeatingQC     2.067e+03  1.188e+03   1.740 0.082059 .
## MasVnrType   -2.896e+03  1.665e+03  -1.740 0.082160 .
## BsmtFinSF1    1.547e+01  2.387e+00   6.482 1.25e-10 ***
## BsmtFinType1 -5.475e+02  6.580e+02  -0.832 0.405480
## BsmtExposure  4.882e+03  9.815e+02   4.974 7.35e-07 ***
## SaleType      3.506e+03  6.736e+02   5.206 2.22e-07 ***
## Exterior1st   5.749e+02  6.353e+02   0.905 0.365711
## Exterior2nd  -2.104e+02  6.209e+02  -0.339 0.734807
## WoodDeckSF    2.283e+01  7.505e+00   3.042 0.002392 **
## X2ndFlrSF     3.404e+01  4.591e+00   7.415 2.08e-13 ***
## HalfBath      8.183e+02  2.557e+03   0.320 0.749000
## GarageQual    9.631e+03  4.474e+03   2.153 0.031501 *
## HouseStyle   -4.978e+02  7.576e+02  -0.657 0.511239
## LotShape      2.161e+03  1.453e+03   1.487 0.137319
## LotArea       2.544e-01  1.002e-01   2.538 0.011240 *
## GarageCond   -1.464e+03  4.197e+03  -0.349 0.727337
## GarageYrBlt  -1.857e+01  4.463e+00  -4.159 3.38e-05 ***
## CentralAir    3.619e+03  4.103e+03   0.882 0.377813
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32870 on 1423 degrees of freedom
## Multiple R-squared:  0.833,  Adjusted R-squared:  0.8288
## F-statistic: 197.1 on 36 and 1423 DF,  p-value: < 2.2e-16
```

```r
# Removing YearRemodAdd
house_model.lm <- lm(data = my_train_house, formula = SalePrice ~ OverallQual + Neighborhood + ExterQual
                       BsmtQual + X1stFlrSF + FullBath + GarageFinish + TotRmsAbvGrd + FireplaceQu + Yea
                       GarageType + MSSubClass + MasVnrArea + Fireplaces + HeatingQC + MasVnrType + Bsmt
                       Exterior1st + Exterior2nd + WoodDeckSF + X2ndFlrSF + HalfBath + GarageQual + Hous
                       GarageCond + GarageYrBlt + CentralAir)

summary(house_model.lm)
```

```
##
## Call:
## lm(formula = SalePrice ~ OverallQual + Neighborhood + ExterQual +
##     KitchenQual + GarageCars + GarageArea + BsmtQual + X1stFlrSF +
##     FullBath + GarageFinish + TotRmsAbvGrd + FireplaceQu + YearBuilt +
##     Foundation + GarageType + MSSubClass + MasVnrArea + Fireplaces +
##     HeatingQC + MasVnrType + BsmtFinSF1 + BsmtFinType1 + BsmtExposure +
##     SaleType + Exterior1st + Exterior2nd + WoodDeckSF + X2ndFlrSF +
##     HalfBath + GarageQual + HouseStyle + LotShape + LotArea +
##     GarageCond + GarageYrBlt + CentralAir, data = my_train_house)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -436693  -15897    -717   14622  295997
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.880e+05  1.214e+05   2.373 0.017800 *
## OverallQual   1.148e+04  1.222e+03   9.393  < 2e-16 ***
## Neighborhood  2.111e+03  2.229e+02   9.470  < 2e-16 ***
## ExterQual     9.947e+03  2.660e+03   3.740 0.000192 ***
## KitchenQual   1.106e+04  2.048e+03   5.401 7.77e-08 ***
## GarageCars    1.065e+04  2.848e+03   3.740 0.000191 ***
## GarageArea    7.564e+00  9.305e+00   0.813 0.416360
## BsmtQual      5.937e+03  1.872e+03   3.171 0.001550 **
## X1stFlrSF     4.530e+01  4.588e+00   9.873  < 2e-16 ***
## FullBath     -2.054e+03  2.623e+03  -0.783 0.433592
## GarageFinish  1.480e+03  1.532e+03   0.966 0.334012
## TotRmsAbvGrd  1.932e+03  9.918e+02   1.948 0.051652 .
## FireplaceQu   1.311e+03  1.201e+03   1.092 0.274905
## YearBuilt    -2.088e+02  6.416e+01  -3.254 0.001166 **
## Foundation   -5.892e+02  8.688e+02  -0.678 0.497800
## GarageType   -1.730e+02  8.846e+02  -0.196 0.845013
## MSSubClass    5.635e+02  3.260e+02   1.728 0.084128 .
## MasVnrArea    2.995e+01  6.357e+00   4.711 2.70e-06 ***
## Fireplaces    2.188e+03  2.508e+03   0.873 0.383025
## HeatingQC     2.134e+03  1.162e+03   1.835 0.066653 .
## MasVnrType   -2.870e+03  1.670e+03  -1.719 0.085917 .
## BsmtFinSF1    1.544e+01  2.387e+00   6.469 1.35e-10 ***
## BsmtFinType1 -5.293e+02  6.545e+02  -0.809 0.418826
## BsmtExposure  4.896e+03  9.832e+02   4.979 7.16e-07 ***
## SaleType      3.515e+03  6.726e+02   5.226 1.99e-07 ***
## Exterior1st   5.804e+02  6.358e+02   0.913 0.361491
## Exterior2nd  -2.013e+02  6.204e+02  -0.324 0.745678
```

```
## WoodDeckSF    2.301e+01  7.508e+00   3.065 0.002216 **
## X2ndFlrSF     3.403e+01  4.591e+00   7.413 2.11e-13 ***
## HalfBath      8.484e+02  2.557e+03   0.332 0.740053
## GarageQual    9.672e+03  4.471e+03   2.163 0.030698 *
## HouseStyle   -4.814e+02  7.549e+02  -0.638 0.523747
## LotShape      2.168e+03  1.453e+03   1.492 0.135928
## LotArea       2.549e-01  1.003e-01   2.541 0.011168 *
## GarageCond   -1.447e+03  4.203e+03  -0.344 0.730631
## GarageYrBlt  -1.843e+01  4.584e+00  -4.021 6.10e-05 ***
## CentralAir    3.776e+03  4.087e+03   0.924 0.355669
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32880 on 1423 degrees of freedom
## Multiple R-squared:  0.833,  Adjusted R-squared:  0.8287
## F-statistic: 197.1 on 36 and 1423 DF,  p-value: < 2.2e-16
```

```
# Removing GarageType
house_model.lm <- lm(data = my_train_house, formula = SalePrice ~ OverallQual + Neighborhood + ExterQual
                     BsmtQual + X1stFlrSF + FullBath + GarageFinish + TotRmsAbvGrd + FireplaceQu + Yea
                     MSSubClass + MasVnrArea + Fireplaces + HeatingQC + MasVnrType + BsmtFinSF1 + Bsm
                     Exterior1st + Exterior2nd + WoodDeckSF + X2ndFlrSF + HalfBath + GarageQual + Hous
                     GarageCond + GarageYrBlt + CentralAir)

summary(house_model.lm)
```

```
##
## Call:
## lm(formula = SalePrice ~ OverallQual + Neighborhood + ExterQual +
##     KitchenQual + GarageCars + GarageArea + BsmtQual + X1stFlrSF +
##     FullBath + GarageFinish + TotRmsAbvGrd + FireplaceQu + YearBuilt +
##     Foundation + MSSubClass + MasVnrArea + Fireplaces + HeatingQC +
##     MasVnrType + BsmtFinSF1 + BsmtFinType1 + BsmtExposure + SaleType +
##     Exterior1st + Exterior2nd + WoodDeckSF + X2ndFlrSF + HalfBath +
##     GarageQual + HouseStyle + LotShape + LotArea + GarageCond +
##     GarageYrBlt + CentralAir, data = my_train_house)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -436427  -15901    -779   14594  296144
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.919e+05  1.197e+05   2.438 0.014907 *
## OverallQual   1.148e+04  1.221e+03   9.396  < 2e-16 ***
## Neighborhood  2.109e+03  2.226e+02   9.473  < 2e-16 ***
## ExterQual     9.952e+03  2.659e+03   3.743 0.000189 ***
## KitchenQual   1.107e+04  2.047e+03   5.409 7.41e-08 ***
## GarageCars    1.069e+04  2.843e+03   3.759 0.000177 ***
## GarageArea    7.689e+00  9.279e+00   0.829 0.407448
## BsmtQual      5.939e+03  1.871e+03   3.173 0.001539 **
## X1stFlrSF     4.522e+01  4.567e+00   9.901  < 2e-16 ***
## FullBath     -2.073e+03  2.620e+03  -0.791 0.428900
## GarageFinish  1.401e+03  1.477e+03   0.949 0.342900
```

```
## TotRmsAbvGrd  1.933e+03  9.915e+02   1.950 0.051403 .
## FireplaceQu   1.304e+03  1.200e+03   1.087 0.277293
## YearBuilt    -2.106e+02  6.344e+01  -3.320 0.000923 ***
## Foundation   -5.862e+02  8.684e+02  -0.675 0.499762
## MSSubClass    5.494e+02  3.178e+02   1.729 0.084094 .
## MasVnrArea    3.000e+01  6.350e+00   4.725 2.53e-06 ***
## Fireplaces    2.171e+03  2.506e+03   0.867 0.386280
## HeatingQC     2.141e+03  1.161e+03   1.843 0.065483 .
## MasVnrType   -2.897e+03  1.664e+03  -1.740 0.082009 .
## BsmtFinSF1    1.545e+01  2.385e+00   6.478 1.28e-10 ***
## BsmtFinType1 -5.271e+02  6.542e+02  -0.806 0.420527
## BsmtExposure  4.884e+03  9.812e+02   4.978 7.21e-07 ***
## SaleType      3.520e+03  6.719e+02   5.238 1.87e-07 ***
## Exterior1st   5.756e+02  6.351e+02   0.906 0.364987
## Exterior2nd  -2.023e+02  6.201e+02  -0.326 0.744364
## WoodDeckSF    2.293e+01  7.494e+00   3.060 0.002254 **
## X2ndFlrSF     3.402e+01  4.589e+00   7.413 2.11e-13 ***
## HalfBath      8.377e+02  2.555e+03   0.328 0.743075
## GarageQual    9.679e+03  4.469e+03   2.166 0.030509 *
## HouseStyle   -4.779e+02  7.544e+02  -0.633 0.526556
## LotShape      2.176e+03  1.452e+03   1.499 0.134175
## LotArea       2.539e-01  1.002e-01   2.535 0.011344 *
## GarageCond   -1.496e+03  4.194e+03  -0.357 0.721330
## GarageYrBlt  -1.864e+01  4.455e+00  -4.185 3.03e-05 ***
## CentralAir    3.740e+03  4.081e+03   0.916 0.359614
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32860 on 1424 degrees of freedom
## Multiple R-squared:  0.833,  Adjusted R-squared:  0.8289
## F-statistic: 202.9 on 35 and 1424 DF,  p-value: < 2.2e-16
```

```
# Removing Exterior2nd
house_model.lm <- lm(data = my_train_house, formula = SalePrice ~ OverallQual + Neighborhood + ExterQual
                       BsmtQual + X1stFlrSF + FullBath + GarageFinish + TotRmsAbvGrd + FireplaceQu + Yea
                       MSSubClass + MasVnrArea + Fireplaces + HeatingQC + MasVnrType + BsmtFinSF1 + Bsmt
                       Exterior1st + WoodDeckSF + X2ndFlrSF + HalfBath + GarageQual + HouseStyle + LotSh
                       GarageCond + GarageYrBlt + CentralAir)

summary(house_model.lm)
```

```
##
## Call:
## lm(formula = SalePrice ~ OverallQual + Neighborhood + ExterQual +
##     KitchenQual + GarageCars + GarageArea + BsmtQual + X1stFlrSF +
##     FullBath + GarageFinish + TotRmsAbvGrd + FireplaceQu + YearBuilt +
##     Foundation + MSSubClass + MasVnrArea + Fireplaces + HeatingQC +
##     MasVnrType + BsmtFinSF1 + BsmtFinType1 + BsmtExposure + SaleType +
##     Exterior1st + WoodDeckSF + X2ndFlrSF + HalfBath + GarageQual +
##     HouseStyle + LotShape + LotArea + GarageCond + GarageYrBlt +
##     CentralAir, data = my_train_house)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -436133  -15973    -718   14565  294224
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.921e+05  1.197e+05   2.440 0.014794 *
## OverallQual   1.149e+04  1.221e+03   9.411  < 2e-16 ***
## Neighborhood  2.111e+03  2.225e+02   9.490  < 2e-16 ***
## ExterQual     9.953e+03  2.658e+03   3.745 0.000188 ***
## KitchenQual   1.106e+04  2.046e+03   5.407 7.52e-08 ***
## GarageCars    1.068e+04  2.842e+03   3.760 0.000177 ***
## GarageArea    7.609e+00  9.273e+00   0.820 0.412076
## BsmtQual      5.961e+03  1.870e+03   3.188 0.001462 **
## X1stFlrSF     4.519e+01  4.565e+00   9.900  < 2e-16 ***
## FullBath     -2.057e+03  2.619e+03  -0.786 0.432196
## GarageFinish  1.389e+03  1.476e+03   0.941 0.346763
## TotRmsAbvGrd  1.934e+03  9.912e+02   1.951 0.051285 .
## FireplaceQu   1.312e+03  1.199e+03   1.094 0.274091
## YearBuilt    -2.109e+02  6.341e+01  -3.325 0.000906 ***
## Foundation   -6.047e+02  8.662e+02  -0.698 0.485233
## MSSubClass    5.387e+02  3.160e+02   1.705 0.088484 .
## MasVnrArea    2.991e+01  6.341e+00   4.716 2.64e-06 ***
## Fireplaces    2.176e+03  2.505e+03   0.869 0.385037
## HeatingQC     2.125e+03  1.160e+03   1.832 0.067169 .
## MasVnrType   -2.883e+03  1.663e+03  -1.733 0.083235 .
## BsmtFinSF1    1.547e+01  2.384e+00   6.491 1.18e-10 ***
## BsmtFinType1 -5.365e+02  6.533e+02  -0.821 0.411681
## BsmtExposure  4.868e+03  9.796e+02   4.969 7.53e-07 ***
## SaleType      3.516e+03  6.716e+02   5.235 1.90e-07 ***
## Exterior1st   3.979e+02  3.264e+02   1.219 0.222990
## WoodDeckSF    2.289e+01  7.491e+00   3.056 0.002285 **
## X2ndFlrSF     3.404e+01  4.587e+00   7.421 1.99e-13 ***
## HalfBath      8.566e+02  2.554e+03   0.335 0.737366
## GarageQual    9.688e+03  4.468e+03   2.168 0.030295 *
## HouseStyle   -4.764e+02  7.542e+02  -0.632 0.527724
## LotShape      2.175e+03  1.452e+03   1.499 0.134199
## LotArea       2.525e-01  1.000e-01   2.524 0.011701 *
## GarageCond   -1.492e+03  4.193e+03  -0.356 0.722069
## GarageYrBlt  -1.858e+01  4.449e+00  -4.176 3.14e-05 ***
## CentralAir    3.808e+03  4.075e+03   0.935 0.350171
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32850 on 1425 degrees of freedom
## Multiple R-squared:  0.833,  Adjusted R-squared:  0.829
## F-statistic:   209 on 34 and 1425 DF,  p-value: < 2.2e-16
```

```
# Removing HalfBath
house_model.lm <- lm(data = my_train_house, formula = SalePrice ~ OverallQual + Neighborhood + ExterQual
                       BsmtQual + X1stFlrSF + FullBath + GarageFinish + TotRmsAbvGrd + FireplaceQu + Yea
                       MSSubClass + MasVnrArea + Fireplaces + HeatingQC + MasVnrType + BsmtFinSF1 + Bsmt
                       Exterior1st + WoodDeckSF + X2ndFlrSF + GarageQual + HouseStyle + LotShape + LotAr
                       GarageCond + GarageYrBlt + CentralAir)

summary(house_model.lm)
```

```
##
## Call:
## lm(formula = SalePrice ~ OverallQual + Neighborhood + ExterQual +
##     KitchenQual + GarageCars + GarageArea + BsmtQual + X1stFlrSF +
##     FullBath + GarageFinish + TotRmsAbvGrd + FireplaceQu + YearBuilt +
##     Foundation + MSSubClass + MasVnrArea + Fireplaces + HeatingQC +
##     MasVnrType + BsmtFinSF1 + BsmtFinType1 + BsmtExposure + SaleType +
##     Exterior1st + WoodDeckSF + X2ndFlrSF + GarageQual + HouseStyle +
##     LotShape + LotArea + GarageCond + GarageYrBlt + CentralAir,
##     data = my_train_house)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -436340  -16026    -817   14655  293595
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.808e+05  1.148e+05   2.445 0.014590 *
## OverallQual   1.149e+04  1.220e+03   9.419  < 2e-16 ***
## Neighborhood  2.110e+03  2.224e+02   9.491  < 2e-16 ***
## ExterQual     9.938e+03  2.657e+03   3.741 0.000191 ***
## KitchenQual   1.106e+04  2.045e+03   5.410 7.38e-08 ***
## GarageCars    1.076e+04  2.832e+03   3.799 0.000151 ***
## GarageArea    7.415e+00  9.252e+00   0.801 0.423045
## BsmtQual      5.980e+03  1.868e+03   3.201 0.001400 **
## X1stFlrSF     4.526e+01  4.559e+00   9.926  < 2e-16 ***
## FullBath     -2.395e+03  2.417e+03  -0.991 0.321925
## GarageFinish  1.422e+03  1.472e+03   0.966 0.334337
## TotRmsAbvGrd  1.940e+03  9.907e+02   1.958 0.050403 .
## FireplaceQu   1.295e+03  1.198e+03   1.082 0.279607
## YearBuilt    -2.049e+02  6.083e+01  -3.368 0.000777 ***
## Foundation   -6.232e+02  8.642e+02  -0.721 0.470991
## MSSubClass    5.364e+02  3.159e+02   1.698 0.089708 .
## MasVnrArea    3.002e+01  6.331e+00   4.742 2.33e-06 ***
## Fireplaces    2.246e+03  2.495e+03   0.900 0.368350
## HeatingQC     2.096e+03  1.156e+03   1.812 0.070162 .
## MasVnrType   -2.928e+03  1.657e+03  -1.766 0.077569 .
## BsmtFinSF1    1.547e+01  2.383e+00   6.493 1.16e-10 ***
## BsmtFinType1 -5.349e+02  6.531e+02  -0.819 0.412925
## BsmtExposure  4.844e+03  9.766e+02   4.960 7.91e-07 ***
## SaleType      3.523e+03  6.710e+02   5.250 1.75e-07 ***
## Exterior1st   3.959e+02  3.262e+02   1.214 0.225033
## WoodDeckSF    2.290e+01  7.489e+00   3.058 0.002273 **
## X2ndFlrSF     3.477e+01  4.033e+00   8.621  < 2e-16 ***
## GarageQual    9.626e+03  4.463e+03   2.157 0.031168 *
## HouseStyle   -4.480e+02  7.492e+02  -0.598 0.549950
## LotShape      2.183e+03  1.451e+03   1.504 0.132683
## LotArea       2.516e-01  9.997e-02   2.517 0.011952 *
## GarageCond   -1.428e+03  4.187e+03  -0.341 0.733075
## GarageYrBlt  -1.863e+01  4.445e+00  -4.191 2.95e-05 ***
## CentralAir    3.870e+03  4.069e+03   0.951 0.341803
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 32840 on 1426 degrees of freedom
## Multiple R-squared:  0.8329, Adjusted R-squared:  0.8291
## F-statistic: 215.5 on 33 and 1426 DF,  p-value: < 2.2e-16
```

```
# Removing GarageCond
house_model.lm <- lm(data = my_train_house, formula = SalePrice ~ OverallQual + Neighborhood + ExterQual
                     BsmtQual + X1stFlrSF + FullBath + GarageFinish + TotRmsAbvGrd + FireplaceQu + Yea
                     MSSubClass + MasVnrArea + Fireplaces + HeatingQC + MasVnrType + BsmtFinSF1 + Bsmt
                     Exterior1st + WoodDeckSF + X2ndFlrSF + GarageQual + HouseStyle + LotShape + LotAr
                     GarageYrBlt + CentralAir)

summary(house_model.lm)
```

```
##
## Call:
## lm(formula = SalePrice ~ OverallQual + Neighborhood + ExterQual +
##     KitchenQual + GarageCars + GarageArea + BsmtQual + X1stFlrSF +
##     FullBath + GarageFinish + TotRmsAbvGrd + FireplaceQu + YearBuilt +
##     Foundation + MSSubClass + MasVnrArea + Fireplaces + HeatingQC +
##     MasVnrType + BsmtFinSF1 + BsmtFinType1 + BsmtExposure + SaleType +
##     Exterior1st + WoodDeckSF + X2ndFlrSF + GarageQual + HouseStyle +
##     LotShape + LotArea + GarageYrBlt + CentralAir, data = my_train_house)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -436298  -16035    -920   14621  293604
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.829e+05  1.146e+05   2.468 0.013708 *
## OverallQual   1.149e+04  1.220e+03   9.418  < 2e-16 ***
## Neighborhood  2.110e+03  2.223e+02   9.490  < 2e-16 ***
## ExterQual     9.960e+03  2.655e+03   3.751 0.000183 ***
## KitchenQual   1.108e+04  2.044e+03   5.422 6.92e-08 ***
## GarageCars    1.078e+04  2.831e+03   3.808 0.000146 ***
## GarageArea    7.359e+00  9.248e+00   0.796 0.426310
## BsmtQual      5.940e+03  1.864e+03   3.187 0.001470 **
## X1stFlrSF     4.526e+01  4.558e+00   9.930  < 2e-16 ***
## FullBath     -2.386e+03  2.416e+03  -0.988 0.323456
## GarageFinish  1.417e+03  1.471e+03   0.963 0.335639
## TotRmsAbvGrd  1.944e+03  9.903e+02   1.963 0.049893 *
## FireplaceQu   1.303e+03  1.197e+03   1.089 0.276383
## YearBuilt    -2.062e+02  6.069e+01  -3.397 0.000699 ***
## Foundation   -6.064e+02  8.626e+02  -0.703 0.482160
## MSSubClass    5.384e+02  3.157e+02   1.706 0.088312 .
## MasVnrArea    3.004e+01  6.329e+00   4.746 2.28e-06 ***
## Fireplaces    2.240e+03  2.495e+03   0.898 0.369363
## HeatingQC     2.073e+03  1.154e+03   1.797 0.072626 .
## MasVnrType   -2.931e+03  1.657e+03  -1.769 0.077085 .
## BsmtFinSF1    1.546e+01  2.382e+00   6.491 1.17e-10 ***
## BsmtFinType1 -5.208e+02  6.516e+02  -0.799 0.424234
## BsmtExposure  4.858e+03  9.755e+02   4.980 7.14e-07 ***
## SaleType      3.525e+03  6.708e+02   5.255 1.71e-07 ***
## Exterior1st   3.983e+02  3.260e+02   1.222 0.222017
```

```
## WoodDeckSF    2.274e+01  7.472e+00   3.043 0.002383 **
## X2ndFlrSF     3.478e+01  4.032e+00   8.625  < 2e-16 ***
## GarageQual    8.688e+03  3.512e+03   2.474 0.013488 *
## HouseStyle   -4.456e+02  7.489e+02  -0.595 0.551954
## LotShape      2.186e+03  1.451e+03   1.507 0.131960
## LotArea       2.516e-01  9.994e-02   2.518 0.011925 *
## GarageYrBlt  -1.909e+01  4.238e+00  -4.504 7.20e-06 ***
## CentralAir    3.727e+03  4.046e+03   0.921 0.357225
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32830 on 1427 degrees of freedom
## Multiple R-squared:  0.8329, Adjusted R-squared:  0.8292
## F-statistic: 222.3 on 32 and 1427 DF,  p-value: < 2.2e-16
```

```
# Removing HouseStyle
house_model.lm <- lm(data = my_train_house, formula = SalePrice ~ OverallQual + Neighborhood + ExterQual
                       BsmtQual + X1stFlrSF + FullBath + GarageFinish + TotRmsAbvGrd + FireplaceQu + Yea
                       MSSubClass + MasVnrArea + Fireplaces + HeatingQC + MasVnrType + BsmtFinSF1 + Bsmt
                       Exterior1st + WoodDeckSF + X2ndFlrSF + GarageQual + LotShape + LotArea +
                       GarageYrBlt + CentralAir)

summary(house_model.lm)
```

```
##
## Call:
## lm(formula = SalePrice ~ OverallQual + Neighborhood + ExterQual +
##     KitchenQual + GarageCars + GarageArea + BsmtQual + X1stFlrSF +
##     FullBath + GarageFinish + TotRmsAbvGrd + FireplaceQu + YearBuilt +
##     Foundation + MSSubClass + MasVnrArea + Fireplaces + HeatingQC +
##     MasVnrType + BsmtFinSF1 + BsmtFinType1 + BsmtExposure + SaleType +
##     Exterior1st + WoodDeckSF + X2ndFlrSF + GarageQual + LotShape +
##     LotArea + GarageYrBlt + CentralAir, data = my_train_house)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -436735  -15961    -650   14560  293581
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.829e+05  1.146e+05   2.469 0.013682 *
## OverallQual   1.151e+04  1.218e+03   9.449  < 2e-16 ***
## Neighborhood  2.099e+03  2.216e+02   9.475  < 2e-16 ***
## ExterQual     9.936e+03  2.654e+03   3.743 0.000189 ***
## KitchenQual   1.110e+04  2.043e+03   5.430 6.61e-08 ***
## GarageCars    1.074e+04  2.829e+03   3.796 0.000153 ***
## GarageArea    7.336e+00  9.246e+00   0.793 0.427647
## BsmtQual      5.969e+03  1.863e+03   3.204 0.001385 **
## X1stFlrSF     4.529e+01  4.556e+00   9.940  < 2e-16 ***
## FullBath     -2.435e+03  2.414e+03  -1.009 0.313309
## GarageFinish  1.403e+03  1.471e+03   0.954 0.340300
## TotRmsAbvGrd  1.932e+03  9.899e+02   1.951 0.051216 .
## FireplaceQu   1.355e+03  1.194e+03   1.135 0.256543
## YearBuilt    -2.071e+02  6.066e+01  -3.414 0.000659 ***
```

```
## Foundation   -6.435e+02  8.601e+02  -0.748 0.454444
## MSSubClass    4.788e+02  2.993e+02   1.600 0.109872
## MasVnrArea    2.994e+01  6.325e+00   4.733 2.43e-06 ***
## Fireplaces    2.208e+03  2.493e+03   0.885 0.376061
## HeatingQC     2.092e+03  1.154e+03   1.813 0.069970 .
## MasVnrType   -2.973e+03  1.655e+03  -1.796 0.072683 .
## BsmtFinSF1    1.547e+01  2.381e+00   6.494 1.15e-10 ***
## BsmtFinType1 -4.992e+02  6.504e+02  -0.767 0.442952
## BsmtExposure  4.927e+03  9.682e+02   5.089 4.08e-07 ***
## SaleType      3.521e+03  6.706e+02   5.250 1.75e-07 ***
## Exterior1st   3.895e+02  3.256e+02   1.196 0.231777
## WoodDeckSF    2.263e+01  7.468e+00   3.030 0.002486 **
## X2ndFlrSF     3.464e+01  4.024e+00   8.607  < 2e-16 ***
## GarageQual    8.694e+03  3.511e+03   2.476 0.013403 *
## LotShape      2.184e+03  1.450e+03   1.506 0.132242
## LotArea       2.532e-01  9.988e-02   2.535 0.011357 *
## GarageYrBlt  -1.902e+01  4.235e+00  -4.491 7.65e-06 ***
## CentralAir    3.882e+03  4.037e+03   0.961 0.336472
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32830 on 1428 degrees of freedom
## Multiple R-squared:  0.8329, Adjusted R-squared:  0.8293
## F-statistic: 229.6 on 31 and 1428 DF,  p-value: < 2.2e-16
```

```
# Removing Foundation
house_model.lm <- lm(data = my_train_house, formula = SalePrice ~ OverallQual + Neighborhood + ExterQual
                     BsmtQual + X1stFlrSF + FullBath + GarageFinish + TotRmsAbvGrd + FireplaceQu + Yea
                     MSSubClass + MasVnrArea + Fireplaces + HeatingQC + MasVnrType + BsmtFinSF1 + Bsmt
                     Exterior1st + WoodDeckSF + X2ndFlrSF + GarageQual + LotShape + LotArea +
                     GarageYrBlt + CentralAir)

summary(house_model.lm)
```

```
##
## Call:
## lm(formula = SalePrice ~ OverallQual + Neighborhood + ExterQual +
##     KitchenQual + GarageCars + GarageArea + BsmtQual + X1stFlrSF +
##     FullBath + GarageFinish + TotRmsAbvGrd + FireplaceQu + YearBuilt +
##     MSSubClass + MasVnrArea + Fireplaces + HeatingQC + MasVnrType +
##     BsmtFinSF1 + BsmtFinType1 + BsmtExposure + SaleType + Exterior1st +
##     WoodDeckSF + X2ndFlrSF + GarageQual + LotShape + LotArea +
##     GarageYrBlt + CentralAir, data = my_train_house)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -437477  -16153    -694   14440  293798
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.056e+05  1.105e+05   2.765 0.005765 **
## OverallQual   1.150e+04  1.218e+03   9.443  < 2e-16 ***
## Neighborhood  2.089e+03  2.211e+02   9.448  < 2e-16 ***
## ExterQual     9.788e+03  2.647e+03   3.698 0.000225 ***
```

```
## KitchenQual   1.107e+04  2.043e+03   5.417 7.10e-08 ***
## GarageCars    1.069e+04  2.828e+03   3.779 0.000164 ***
## GarageArea    7.368e+00  9.244e+00   0.797 0.425595
## BsmtQual      5.742e+03  1.838e+03   3.124 0.001817 **
## X1stFlrSF     4.531e+01  4.556e+00   9.946  < 2e-16 ***
## FullBath     -2.463e+03  2.414e+03  -1.021 0.307595
## GarageFinish  1.348e+03  1.469e+03   0.917 0.359079
## TotRmsAbvGrd  1.939e+03  9.897e+02   1.959 0.050328 .
## FireplaceQu   1.348e+03  1.193e+03   1.130 0.258819
## YearBuilt    -2.185e+02  5.868e+01  -3.724 0.000204 ***
## MSSubClass    4.704e+02  2.991e+02   1.573 0.115995
## MasVnrArea    3.017e+01  6.316e+00   4.776 1.97e-06 ***
## Fireplaces    2.308e+03  2.489e+03   0.927 0.354040
## HeatingQC     1.937e+03  1.135e+03   1.707 0.087993 .
## MasVnrType   -2.912e+03  1.653e+03  -1.762 0.078323 .
## BsmtFinSF1    1.552e+01  2.380e+00   6.521 9.67e-11 ***
## BsmtFinType1 -5.962e+02  6.373e+02  -0.935 0.349723
## BsmtExposure  4.930e+03  9.681e+02   5.093 4.00e-07 ***
## SaleType      3.512e+03  6.704e+02   5.238 1.87e-07 ***
## Exterior1st   3.828e+02  3.254e+02   1.176 0.239744
## WoodDeckSF    2.267e+01  7.467e+00   3.036 0.002439 **
## X2ndFlrSF     3.450e+01  4.019e+00   8.583  < 2e-16 ***
## GarageQual    8.650e+03  3.510e+03   2.464 0.013852 *
## LotShape      2.175e+03  1.450e+03   1.500 0.133808
## LotArea       2.574e-01  9.971e-02   2.581 0.009942 **
## GarageYrBlt  -1.888e+01  4.230e+00  -4.463 8.73e-06 ***
## CentralAir    4.013e+03  4.033e+03   0.995 0.319809
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32820 on 1429 degrees of freedom
## Multiple R-squared:  0.8328, Adjusted R-squared:  0.8293
## F-statistic: 237.3 on 30 and 1429 DF,  p-value: < 2.2e-16
```

```
# Removing GarageArea
house_model.lm <- lm(data = my_train_house, formula = SalePrice ~ OverallQual + Neighborhood + ExterQual
                     BsmtQual + X1stFlrSF + FullBath + GarageFinish + TotRmsAbvGrd + FireplaceQu + Yea
                     MSSubClass + MasVnrArea + Fireplaces + HeatingQC + MasVnrType + BsmtFinSF1 + Bsmt
                     Exterior1st + WoodDeckSF + X2ndFlrSF + GarageQual + LotShape + LotArea +
                     GarageYrBlt + CentralAir)

summary(house_model.lm)
```

```
##
## Call:
## lm(formula = SalePrice ~ OverallQual + Neighborhood + ExterQual +
##     KitchenQual + GarageCars + BsmtQual + X1stFlrSF + FullBath +
##     GarageFinish + TotRmsAbvGrd + FireplaceQu + YearBuilt + MSSubClass +
##     MasVnrArea + Fireplaces + HeatingQC + MasVnrType + BsmtFinSF1 +
##     BsmtFinType1 + BsmtExposure + SaleType + Exterior1st + WoodDeckSF +
##     X2ndFlrSF + GarageQual + LotShape + LotArea + GarageYrBlt +
##     CentralAir, data = my_train_house)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -434475  -16132    -633   14462  293140
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.053e+05  1.105e+05   2.763 0.005802 **
## OverallQual   1.151e+04  1.218e+03   9.448  < 2e-16 ***
## Neighborhood  2.092e+03  2.211e+02   9.464  < 2e-16 ***
## ExterQual     9.827e+03  2.646e+03   3.714 0.000212 ***
## KitchenQual   1.112e+04  2.041e+03   5.446 6.05e-08 ***
## GarageCars    1.230e+04  1.972e+03   6.239 5.78e-10 ***
## BsmtQual      5.664e+03  1.835e+03   3.087 0.002064 **
## X1stFlrSF     4.599e+01  4.475e+00  10.278  < 2e-16 ***
## FullBath     -2.602e+03  2.407e+03  -1.081 0.279764
## GarageFinish  1.284e+03  1.466e+03   0.875 0.381572
## TotRmsAbvGrd  1.898e+03  9.883e+02   1.921 0.054966 .
## FireplaceQu   1.341e+03  1.193e+03   1.124 0.261096
## YearBuilt    -2.189e+02  5.868e+01  -3.730 0.000199 ***
## MSSubClass    4.709e+02  2.990e+02   1.575 0.115514
## MasVnrArea    3.041e+01  6.308e+00   4.821 1.58e-06 ***
## Fireplaces    2.148e+03  2.481e+03   0.866 0.386667
## HeatingQC     1.934e+03  1.134e+03   1.705 0.088482 .
## MasVnrType   -2.914e+03  1.653e+03  -1.763 0.078108 .
## BsmtFinSF1    1.574e+01  2.363e+00   6.663 3.81e-11 ***
## BsmtFinType1 -5.798e+02  6.369e+02  -0.910 0.362745
## BsmtExposure  4.929e+03  9.679e+02   5.093 4.00e-07 ***
## SaleType      3.559e+03  6.677e+02   5.330 1.14e-07 ***
## Exterior1st   3.837e+02  3.254e+02   1.179 0.238573
## WoodDeckSF    2.275e+01  7.465e+00   3.048 0.002349 **
## X2ndFlrSF     3.481e+01  3.999e+00   8.707  < 2e-16 ***
## GarageQual    8.996e+03  3.483e+03   2.583 0.009892 **
## LotShape      2.132e+03  1.449e+03   1.472 0.141341
## LotArea       2.607e-01  9.961e-02   2.618 0.008944 **
## GarageYrBlt  -1.892e+01  4.229e+00  -4.474 8.30e-06 ***
## CentralAir    4.052e+03  4.032e+03   1.005 0.315077
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32820 on 1430 degrees of freedom
## Multiple R-squared:  0.8327, Adjusted R-squared:  0.8294
## F-statistic: 245.5 on 29 and 1430 DF,  p-value: < 2.2e-16
```

```
# Removing Fireplaces
house_model.lm <- lm(data = my_train_house, formula = SalePrice ~ OverallQual + Neighborhood + ExterQual
                     BsmtQual + X1stFlrSF + FullBath + GarageFinish + TotRmsAbvGrd + FireplaceQu + Yea
                     MSSubClass + MasVnrArea + HeatingQC + MasVnrType + BsmtFinSF1 + BsmtFinType1 + Bs
                     Exterior1st + WoodDeckSF + X2ndFlrSF + GarageQual + LotShape + LotArea +
                     GarageYrBlt + CentralAir)

summary(house_model.lm)
```

```
##
## Call:
## lm(formula = SalePrice ~ OverallQual + Neighborhood + ExterQual +
```

```
##     KitchenQual + GarageCars + BsmtQual + X1stFlrSF + FullBath +
##     GarageFinish + TotRmsAbvGrd + FireplaceQu + YearBuilt + MSSubClass +
##     MasVnrArea + HeatingQC + MasVnrType + BsmtFinSF1 + BsmtFinType1 +
##     BsmtExposure + SaleType + Exterior1st + WoodDeckSF + X2ndFlrSF +
##     GarageQual + LotShape + LotArea + GarageYrBlt + CentralAir,
##     data = my_train_house)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -433793  -16027    -601   14483  294069
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.079e+05  1.105e+05   2.787 0.005386 **
## OverallQual   1.151e+04  1.218e+03   9.449  < 2e-16 ***
## Neighborhood  2.096e+03  2.210e+02   9.483  < 2e-16 ***
## ExterQual     9.689e+03  2.641e+03   3.669 0.000253 ***
## KitchenQual   1.102e+04  2.038e+03   5.408 7.45e-08 ***
## GarageCars    1.225e+04  1.971e+03   6.214 6.75e-10 ***
## BsmtQual      5.602e+03  1.833e+03   3.056 0.002286 **
## X1stFlrSF     4.652e+01  4.432e+00  10.497  < 2e-16 ***
## FullBath     -2.599e+03  2.407e+03  -1.080 0.280431
## GarageFinish  1.300e+03  1.466e+03   0.887 0.375406
## TotRmsAbvGrd  1.834e+03  9.854e+02   1.861 0.062977 .
## FireplaceQu   2.116e+03  7.900e+02   2.678 0.007485 **
## YearBuilt    -2.205e+02  5.864e+01  -3.760 0.000177 ***
## MSSubClass    4.877e+02  2.984e+02   1.635 0.102354
## MasVnrArea    3.036e+01  6.307e+00   4.814 1.64e-06 ***
## HeatingQC     1.900e+03  1.134e+03   1.676 0.093963 .
## MasVnrType   -2.904e+03  1.652e+03  -1.758 0.079024 .
## BsmtFinSF1    1.600e+01  2.344e+00   6.828 1.27e-11 ***
## BsmtFinType1 -5.939e+02  6.366e+02  -0.933 0.351046
## BsmtExposure  4.975e+03  9.664e+02   5.148 2.99e-07 ***
## SaleType      3.521e+03  6.662e+02   5.285 1.45e-07 ***
## Exterior1st   3.844e+02  3.254e+02   1.181 0.237651
## WoodDeckSF    2.266e+01  7.464e+00   3.035 0.002445 **
## X2ndFlrSF     3.528e+01  3.962e+00   8.905  < 2e-16 ***
## GarageQual    9.110e+03  3.480e+03   2.618 0.008945 **
## LotShape      2.141e+03  1.449e+03   1.478 0.139551
## LotArea       2.707e-01  9.894e-02   2.736 0.006302 **
## GarageYrBlt  -1.898e+01  4.228e+00  -4.490 7.70e-06 ***
## CentralAir    4.333e+03  4.018e+03   1.078 0.281085
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32810 on 1431 degrees of freedom
## Multiple R-squared:  0.8327, Adjusted R-squared:  0.8294
## F-statistic: 254.3 on 28 and 1431 DF,  p-value: < 2.2e-16
```

```
# Removing GarageFinish
house_model.lm <- lm(data = my_train_house, formula = SalePrice ~ OverallQual + Neighborhood + ExterQual
                     BsmtQual + X1stFlrSF + FullBath + GarageFinish + TotRmsAbvGrd + FireplaceQu + Yea
                     MSSubClass + MasVnrArea + HeatingQC + MasVnrType + BsmtFinSF1 + BsmtFinType1 + Bs
                     Exterior1st + WoodDeckSF + X2ndFlrSF + GarageQual + LotShape + LotArea +
```

```
                        GarageYrBlt + CentralAir)

summary(house_model.lm)
```

```
##
## Call:
## lm(formula = SalePrice ~ OverallQual + Neighborhood + ExterQual +
##     KitchenQual + GarageCars + BsmtQual + X1stFlrSF + FullBath +
##     GarageFinish + TotRmsAbvGrd + FireplaceQu + YearBuilt + MSSubClass +
##     MasVnrArea + HeatingQC + MasVnrType + BsmtFinSF1 + BsmtFinType1 +
##     BsmtExposure + SaleType + Exterior1st + WoodDeckSF + X2ndFlrSF +
##     GarageQual + LotShape + LotArea + GarageYrBlt + CentralAir,
##     data = my_train_house)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -433793  -16027    -601   14483  294069
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.079e+05  1.105e+05   2.787 0.005386 **
## OverallQual   1.151e+04  1.218e+03   9.449  < 2e-16 ***
## Neighborhood  2.096e+03  2.210e+02   9.483  < 2e-16 ***
## ExterQual     9.689e+03  2.641e+03   3.669 0.000253 ***
## KitchenQual   1.102e+04  2.038e+03   5.408 7.45e-08 ***
## GarageCars    1.225e+04  1.971e+03   6.214 6.75e-10 ***
## BsmtQual      5.602e+03  1.833e+03   3.056 0.002286 **
## X1stFlrSF     4.652e+01  4.432e+00  10.497  < 2e-16 ***
## FullBath     -2.599e+03  2.407e+03  -1.080 0.280431
## GarageFinish  1.300e+03  1.466e+03   0.887 0.375406
## TotRmsAbvGrd  1.834e+03  9.854e+02   1.861 0.062977 .
## FireplaceQu   2.116e+03  7.900e+02   2.678 0.007485 **
## YearBuilt    -2.205e+02  5.864e+01  -3.760 0.000177 ***
## MSSubClass    4.877e+02  2.984e+02   1.635 0.102354
## MasVnrArea    3.036e+01  6.307e+00   4.814 1.64e-06 ***
## HeatingQC     1.900e+03  1.134e+03   1.676 0.093963 .
## MasVnrType   -2.904e+03  1.652e+03  -1.758 0.079024 .
## BsmtFinSF1    1.600e+01  2.344e+00   6.828 1.27e-11 ***
## BsmtFinType1 -5.939e+02  6.366e+02  -0.933 0.351046
## BsmtExposure  4.975e+03  9.664e+02   5.148 2.99e-07 ***
## SaleType      3.521e+03  6.662e+02   5.285 1.45e-07 ***
## Exterior1st   3.844e+02  3.254e+02   1.181 0.237651
## WoodDeckSF    2.266e+01  7.464e+00   3.035 0.002445 **
## X2ndFlrSF     3.528e+01  3.962e+00   8.905  < 2e-16 ***
## GarageQual    9.110e+03  3.480e+03   2.618 0.008945 **
## LotShape      2.141e+03  1.449e+03   1.478 0.139551
## LotArea       2.707e-01  9.894e-02   2.736 0.006302 **
## GarageYrBlt  -1.898e+01  4.228e+00  -4.490 7.70e-06 ***
## CentralAir    4.333e+03  4.018e+03   1.078 0.281085
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32810 on 1431 degrees of freedom
```

```
## Multiple R-squared:  0.8327, Adjusted R-squared:  0.8294
## F-statistic: 254.3 on 28 and 1431 DF,  p-value: < 2.2e-16
```

```
# Removing BsmtFinType1
house_model.lm <- lm(data = my_train_house, formula = SalePrice ~ OverallQual + Neighborhood + ExterQual
                       BsmtQual + X1stFlrSF + FullBath + GarageFinish + TotRmsAbvGrd + FireplaceQu + Yea
                       MSSubClass + MasVnrArea + HeatingQC + MasVnrType + BsmtFinSF1 + BsmtExposure + Sa
                       Exterior1st + WoodDeckSF + X2ndFlrSF + GarageQual + LotShape + LotArea +
                       GarageYrBlt + CentralAir)

summary(house_model.lm)
```

```
##
## Call:
## lm(formula = SalePrice ~ OverallQual + Neighborhood + ExterQual +
##     KitchenQual + GarageCars + BsmtQual + X1stFlrSF + FullBath +
##     GarageFinish + TotRmsAbvGrd + FireplaceQu + YearBuilt + MSSubClass +
##     MasVnrArea + HeatingQC + MasVnrType + BsmtFinSF1 + BsmtExposure +
##     SaleType + Exterior1st + WoodDeckSF + X2ndFlrSF + GarageQual +
##     LotShape + LotArea + GarageYrBlt + CentralAir, data = my_train_house)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -435111  -15937    -621   14374  294374
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.085e+05  1.105e+05   2.793 0.005284 **
## OverallQual   1.138e+04  1.210e+03   9.405  < 2e-16 ***
## Neighborhood  2.100e+03  2.209e+02   9.503  < 2e-16 ***
## ExterQual     9.628e+03  2.640e+03   3.647 0.000275 ***
## KitchenQual   1.090e+04  2.034e+03   5.359 9.73e-08 ***
## GarageCars    1.213e+04  1.967e+03   6.169 8.94e-10 ***
## BsmtQual      5.102e+03  1.753e+03   2.910 0.003666 **
## X1stFlrSF     4.672e+01  4.427e+00  10.554  < 2e-16 ***
## FullBath     -2.683e+03  2.405e+03  -1.116 0.264689
## GarageFinish  1.306e+03  1.466e+03   0.891 0.373051
## TotRmsAbvGrd  1.870e+03  9.845e+02   1.899 0.057749 .
## FireplaceQu   2.155e+03  7.888e+02   2.732 0.006378 **
## YearBuilt    -2.210e+02  5.864e+01  -3.768 0.000171 ***
## MSSubClass    4.985e+02  2.981e+02   1.672 0.094698 .
## MasVnrArea    3.075e+01  6.293e+00   4.885 1.15e-06 ***
## HeatingQC     1.783e+03  1.127e+03   1.582 0.113793
## MasVnrType   -2.920e+03  1.652e+03  -1.767 0.077372 .
## BsmtFinSF1    1.611e+01  2.341e+00   6.884 8.66e-12 ***
## BsmtExposure  4.872e+03  9.600e+02   5.075 4.38e-07 ***
## SaleType      3.559e+03  6.650e+02   5.352 1.01e-07 ***
## Exterior1st   3.751e+02  3.252e+02   1.154 0.248873
## WoodDeckSF    2.242e+01  7.460e+00   3.006 0.002692 **
## X2ndFlrSF     3.523e+01  3.961e+00   8.893  < 2e-16 ***
## GarageQual    9.500e+03  3.454e+03   2.750 0.006033 **
## LotShape      2.172e+03  1.448e+03   1.500 0.133834
## LotArea       2.748e-01  9.883e-02   2.781 0.005496 **
## GarageYrBlt  -1.919e+01  4.222e+00  -4.544 5.99e-06 ***
```

```
## CentralAir    4.375e+03  4.018e+03   1.089 0.276387
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32810 on 1432 degrees of freedom
## Multiple R-squared:  0.8326, Adjusted R-squared:  0.8294
## F-statistic: 263.7 on 27 and 1432 DF,  p-value: < 2.2e-16
```

```
# Removing GarageFinish
house_model.lm <- lm(data = my_train_house, formula = SalePrice ~ OverallQual + Neighborhood + ExterQual +
                       BsmtQual + X1stFlrSF + FullBath + TotRmsAbvGrd + FireplaceQu + YearBuilt +
                       MSSubClass + MasVnrArea + HeatingQC + MasVnrType + BsmtFinSF1 + BsmtExposure + Sa
                       Exterior1st + WoodDeckSF + X2ndFlrSF + GarageQual + LotShape + LotArea +
                       GarageYrBlt + CentralAir)

summary(house_model.lm)
```

```
##
## Call:
## lm(formula = SalePrice ~ OverallQual + Neighborhood + ExterQual +
##     KitchenQual + GarageCars + BsmtQual + X1stFlrSF + FullBath +
##     TotRmsAbvGrd + FireplaceQu + YearBuilt + MSSubClass + MasVnrArea +
##     HeatingQC + MasVnrType + BsmtFinSF1 + BsmtExposure + SaleType +
##     Exterior1st + WoodDeckSF + X2ndFlrSF + GarageQual + LotShape +
##     LotArea + GarageYrBlt + CentralAir, data = my_train_house)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -435701  -15998    -515   14525  294805
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.923e+05  1.089e+05   2.683 0.007375 **
## OverallQual   1.142e+04  1.209e+03   9.444  < 2e-16 ***
## Neighborhood  2.092e+03  2.207e+02   9.476  < 2e-16 ***
## ExterQual     9.760e+03  2.636e+03   3.703 0.000221 ***
## KitchenQual   1.087e+04  2.033e+03   5.347 1.04e-07 ***
## GarageCars    1.220e+04  1.965e+03   6.205 7.12e-10 ***
## BsmtQual      5.197e+03  1.750e+03   2.970 0.003029 **
## X1stFlrSF     4.667e+01  4.426e+00  10.544  < 2e-16 ***
## FullBath     -2.545e+03  2.400e+03  -1.061 0.288986
## TotRmsAbvGrd  1.875e+03  9.845e+02   1.905 0.057037 .
## FireplaceQu   2.257e+03  7.804e+02   2.892 0.003886 **
## YearBuilt    -2.134e+02  5.801e+01  -3.678 0.000243 ***
## MSSubClass    5.423e+02  2.940e+02   1.844 0.065331 .
## MasVnrArea    3.070e+01  6.293e+00   4.879 1.19e-06 ***
## HeatingQC     1.877e+03  1.122e+03   1.674 0.094385 .
## MasVnrType   -2.890e+03  1.652e+03  -1.750 0.080404 .
## BsmtFinSF1    1.624e+01  2.336e+00   6.952 5.45e-12 ***
## BsmtExposure  4.902e+03  9.593e+02   5.110 3.65e-07 ***
## SaleType      3.568e+03  6.649e+02   5.366 9.36e-08 ***
## Exterior1st   3.900e+02  3.248e+02   1.201 0.229969
## WoodDeckSF    2.273e+01  7.451e+00   3.051 0.002324 **
## X2ndFlrSF     3.517e+01  3.960e+00   8.880  < 2e-16 ***
```

```
## GarageQual    9.512e+03  3.454e+03   2.754 0.005966 **
## LotShape      2.218e+03  1.447e+03   1.533 0.125511
## LotArea       2.772e-01  9.879e-02   2.806 0.005082 **
## GarageYrBlt  -1.847e+01  4.145e+00  -4.456 9.00e-06 ***
## CentralAir    4.221e+03  4.014e+03   1.052 0.293147
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32810 on 1433 degrees of freedom
## Multiple R-squared:  0.8325, Adjusted R-squared:  0.8294
## F-statistic: 273.9 on 26 and 1433 DF,  p-value: < 2.2e-16
```

```
# Removing CentralAir
house_model.lm <- lm(data = my_train_house, formula = SalePrice ~ OverallQual + Neighborhood + ExterQual
                       BsmtQual + X1stFlrSF + FullBath + TotRmsAbvGrd + FireplaceQu + YearBuilt +
                       MSSubClass + MasVnrArea + HeatingQC + MasVnrType + BsmtFinSF1 + BsmtExposure + Sa
                       Exterior1st + WoodDeckSF + X2ndFlrSF + GarageQual + LotShape + LotArea +
                       GarageYrBlt)

summary(house_model.lm)
```

```
##
## Call:
## lm(formula = SalePrice ~ OverallQual + Neighborhood + ExterQual +
##     KitchenQual + GarageCars + BsmtQual + X1stFlrSF + FullBath +
##     TotRmsAbvGrd + FireplaceQu + YearBuilt + MSSubClass + MasVnrArea +
##     HeatingQC + MasVnrType + BsmtFinSF1 + BsmtExposure + SaleType +
##     Exterior1st + WoodDeckSF + X2ndFlrSF + GarageQual + LotShape +
##     LotArea + GarageYrBlt, data = my_train_house)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -436615  -15867    -668   14473  294739
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.731e+05  1.074e+05   2.543 0.011095 *
## OverallQual   1.149e+04  1.207e+03   9.524  < 2e-16 ***
## Neighborhood  2.088e+03  2.207e+02   9.458  < 2e-16 ***
## ExterQual     9.447e+03  2.619e+03   3.607 0.000320 ***
## KitchenQual   1.097e+04  2.031e+03   5.400 7.78e-08 ***
## GarageCars    1.202e+04  1.958e+03   6.137 1.09e-09 ***
## BsmtQual      5.210e+03  1.750e+03   2.977 0.002956 **
## X1stFlrSF     4.665e+01  4.426e+00  10.541  < 2e-16 ***
## FullBath     -2.758e+03  2.391e+03  -1.153 0.249022
## TotRmsAbvGrd  1.901e+03  9.842e+02   1.931 0.053658 .
## FireplaceQu   2.306e+03  7.791e+02   2.960 0.003130 **
## YearBuilt    -2.024e+02  5.707e+01  -3.547 0.000402 ***
## MSSubClass    5.527e+02  2.939e+02   1.881 0.060219 .
## MasVnrArea    3.083e+01  6.292e+00   4.899 1.07e-06 ***
## HeatingQC     2.095e+03  1.102e+03   1.900 0.057591 .
## MasVnrType   -2.930e+03  1.651e+03  -1.774 0.076239 .
## BsmtFinSF1    1.633e+01  2.335e+00   6.997 4.00e-12 ***
## BsmtExposure  4.890e+03  9.593e+02   5.098 3.90e-07 ***
```

```
## SaleType      3.538e+03  6.643e+02   5.326 1.16e-07 ***
## Exterior1st   3.988e+02  3.247e+02   1.228 0.219535
## WoodDeckSF    2.304e+01  7.445e+00   3.095 0.002005 **
## X2ndFlrSF     3.506e+01  3.959e+00   8.855  < 2e-16 ***
## GarageQual    9.835e+03  3.441e+03   2.858 0.004320 **
## LotShape      2.200e+03  1.447e+03   1.520 0.128642
## LotArea       2.803e-01  9.875e-02   2.838 0.004602 **
## GarageYrBlt  -1.832e+01  4.143e+00  -4.422 1.05e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32810 on 1434 degrees of freedom
## Multiple R-squared:  0.8323, Adjusted R-squared:  0.8294
## F-statistic: 284.8 on 25 and 1434 DF,  p-value: < 2.2e-16
```

```
# Removing FullBath
house_model.lm <- lm(data = my_train_house, formula = SalePrice ~ OverallQual + Neighborhood + ExterQual
                       BsmtQual + X1stFlrSF + TotRmsAbvGrd + FireplaceQu + YearBuilt +
                       MSSubClass + MasVnrArea + HeatingQC + MasVnrType + BsmtFinSF1 + BsmtExposure + Sa
                       Exterior1st + WoodDeckSF + X2ndFlrSF + GarageQual + LotShape + LotArea +
                       GarageYrBlt)

summary(house_model.lm)
```

```
##
## Call:
## lm(formula = SalePrice ~ OverallQual + Neighborhood + ExterQual +
##     KitchenQual + GarageCars + BsmtQual + X1stFlrSF + TotRmsAbvGrd +
##     FireplaceQu + YearBuilt + MSSubClass + MasVnrArea + HeatingQC +
##     MasVnrType + BsmtFinSF1 + BsmtExposure + SaleType + Exterior1st +
##     WoodDeckSF + X2ndFlrSF + GarageQual + LotShape + LotArea +
##     GarageYrBlt, data = my_train_house)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -433602  -16210    -161   14639  295366
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.078e+05  1.031e+05   2.985 0.002884 **
## OverallQual   1.143e+04  1.206e+03   9.483  < 2e-16 ***
## Neighborhood  2.067e+03  2.200e+02   9.395  < 2e-16 ***
## ExterQual     9.401e+03  2.619e+03   3.590 0.000342 ***
## KitchenQual   1.098e+04  2.032e+03   5.403 7.65e-08 ***
## GarageCars    1.185e+04  1.953e+03   6.067 1.66e-09 ***
## BsmtQual      5.234e+03  1.750e+03   2.991 0.002828 **
## X1stFlrSF     4.529e+01  4.265e+00  10.619  < 2e-16 ***
## TotRmsAbvGrd  1.778e+03  9.785e+02   1.817 0.069470 .
## FireplaceQu   2.371e+03  7.771e+02   3.052 0.002317 **
## YearBuilt    -2.209e+02  5.479e+01  -4.031 5.84e-05 ***
## MSSubClass    6.004e+02  2.910e+02   2.063 0.039258 *
## MasVnrArea    3.147e+01  6.268e+00   5.021 5.78e-07 ***
## HeatingQC     2.066e+03  1.102e+03   1.875 0.061049 .
## MasVnrType   -2.991e+03  1.651e+03  -1.812 0.070219 .
```

```
## BsmtFinSF1    1.674e+01  2.309e+00   7.249 6.84e-13 ***
## BsmtExposure  4.977e+03  9.565e+02   5.203 2.24e-07 ***
## SaleType      3.548e+03  6.643e+02   5.341 1.07e-07 ***
## Exterior1st   3.873e+02  3.245e+02   1.193 0.232904
## WoodDeckSF    2.301e+01  7.446e+00   3.091 0.002035 **
## X2ndFlrSF     3.367e+01  3.772e+00   8.926  < 2e-16 ***
## GarageQual    9.818e+03  3.441e+03   2.853 0.004389 **
## LotShape      2.170e+03  1.447e+03   1.500 0.133899
## LotArea       2.788e-01  9.875e-02   2.823 0.004825 **
## GarageYrBlt  -1.804e+01  4.137e+00  -4.362 1.38e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32820 on 1435 degrees of freedom
## Multiple R-squared:  0.8322, Adjusted R-squared:  0.8294
## F-statistic: 296.5 on 24 and 1435 DF,  p-value: < 2.2e-16
```

```
# Removing Exterior1st
house_model.lm <- lm(data = my_train_house, formula = SalePrice ~ OverallQual + Neighborhood + ExterQual
                     BsmtQual + X1stFlrSF + TotRmsAbvGrd + FireplaceQu + YearBuilt +
                     MSSubClass + MasVnrArea + HeatingQC + MasVnrType + BsmtFinSF1 + BsmtExposure + Sa
                     WoodDeckSF + X2ndFlrSF + GarageQual + LotShape + LotArea +
                     GarageYrBlt)

summary(house_model.lm)
```

```
##
## Call:
## lm(formula = SalePrice ~ OverallQual + Neighborhood + ExterQual +
##     KitchenQual + GarageCars + BsmtQual + X1stFlrSF + TotRmsAbvGrd +
##     FireplaceQu + YearBuilt + MSSubClass + MasVnrArea + HeatingQC +
##     MasVnrType + BsmtFinSF1 + BsmtExposure + SaleType + WoodDeckSF +
##     X2ndFlrSF + GarageQual + LotShape + LotArea + GarageYrBlt,
##     data = my_train_house)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -434989  -16226    -185   14865  293052
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.687e+05  9.778e+04   2.748 0.006074 **
## OverallQual   1.142e+04  1.206e+03   9.468  < 2e-16 ***
## Neighborhood  2.051e+03  2.196e+02   9.338  < 2e-16 ***
## ExterQual     9.410e+03  2.619e+03   3.593 0.000338 ***
## KitchenQual   1.107e+04  2.031e+03   5.449 5.95e-08 ***
## GarageCars    1.198e+04  1.950e+03   6.145 1.03e-09 ***
## BsmtQual      5.312e+03  1.749e+03   3.037 0.002431 **
## X1stFlrSF     4.532e+01  4.265e+00  10.624  < 2e-16 ***
## TotRmsAbvGrd  1.834e+03  9.775e+02   1.876 0.060828 .
## FireplaceQu   2.372e+03  7.772e+02   3.053 0.002311 **
## YearBuilt    -1.998e+02  5.186e+01  -3.852 0.000122 ***
## MSSubClass    6.301e+02  2.900e+02   2.173 0.029932 *
## MasVnrArea    3.098e+01  6.255e+00   4.953 8.19e-07 ***
```

```
## HeatingQC     2.109e+03  1.102e+03   1.914 0.055836 .
## MasVnrType   -2.934e+03  1.650e+03  -1.778 0.075670 .
## BsmtFinSF1    1.662e+01  2.307e+00   7.205 9.33e-13 ***
## BsmtExposure  4.980e+03  9.566e+02   5.206 2.21e-07 ***
## SaleType      3.579e+03  6.639e+02   5.391 8.16e-08 ***
## WoodDeckSF    2.303e+01  7.447e+00   3.093 0.002022 **
## X2ndFlrSF     3.361e+01  3.772e+00   8.911  < 2e-16 ***
## GarageQual    9.719e+03  3.441e+03   2.825 0.004797 **
## LotShape      2.161e+03  1.447e+03   1.493 0.135581
## LotArea       2.801e-01  9.876e-02   2.837 0.004624 **
## GarageYrBlt  -1.819e+01  4.135e+00  -4.398 1.17e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32820 on 1436 degrees of freedom
## Multiple R-squared:  0.832,  Adjusted R-squared:  0.8293
## F-statistic: 309.2 on 23 and 1436 DF,  p-value: < 2.2e-16
```

```
# Removing LotShape
house_model.lm <- lm(data = my_train_house, formula = SalePrice ~ OverallQual + Neighborhood + ExterQual
                       BsmtQual + X1stFlrSF + TotRmsAbvGrd + FireplaceQu + YearBuilt +
                       MSSubClass + MasVnrArea + HeatingQC + MasVnrType + BsmtFinSF1 + BsmtExposure + Sa
                       WoodDeckSF + X2ndFlrSF + GarageQual + LotArea + GarageYrBlt)

summary(house_model.lm)
```

```
##
## Call:
## lm(formula = SalePrice ~ OverallQual + Neighborhood + ExterQual +
##     KitchenQual + GarageCars + BsmtQual + X1stFlrSF + TotRmsAbvGrd +
##     FireplaceQu + YearBuilt + MSSubClass + MasVnrArea + HeatingQC +
##     MasVnrType + BsmtFinSF1 + BsmtExposure + SaleType + WoodDeckSF +
##     X2ndFlrSF + GarageQual + LotArea + GarageYrBlt, data = my_train_house)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -433629  -16070     -61   14738  293087
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.613e+05  9.770e+04   2.674 0.007578 **
## OverallQual   1.134e+04  1.205e+03   9.412  < 2e-16 ***
## Neighborhood  2.070e+03  2.194e+02   9.438  < 2e-16 ***
## ExterQual     9.549e+03  2.619e+03   3.647 0.000275 ***
## KitchenQual   1.099e+04  2.031e+03   5.412 7.30e-08 ***
## GarageCars    1.201e+04  1.951e+03   6.158 9.56e-10 ***
## BsmtQual      5.413e+03  1.748e+03   3.096 0.002001 **
## X1stFlrSF     4.545e+01  4.266e+00  10.654  < 2e-16 ***
## TotRmsAbvGrd  1.815e+03  9.778e+02   1.856 0.063638 .
## FireplaceQu   2.410e+03  7.771e+02   3.101 0.001964 **
## YearBuilt    -1.960e+02  5.182e+01  -3.782 0.000162 ***
## MSSubClass    6.707e+02  2.888e+02   2.322 0.020346 *
## MasVnrArea    3.084e+01  6.257e+00   4.928 9.25e-07 ***
## HeatingQC     2.112e+03  1.102e+03   1.916 0.055567 .
```

```
## MasVnrType   -3.060e+03  1.649e+03  -1.856 0.063726 .
## BsmtFinSF1    1.658e+01  2.308e+00   7.183 1.09e-12 ***
## BsmtExposure  5.051e+03  9.558e+02   5.285 1.45e-07 ***
## SaleType      3.526e+03  6.632e+02   5.317 1.23e-07 ***
## WoodDeckSF    2.339e+01  7.446e+00   3.142 0.001714 **
## X2ndFlrSF     3.382e+01  3.771e+00   8.966  < 2e-16 ***
## GarageQual    9.928e+03  3.439e+03   2.887 0.003951 **
## LotArea       3.153e-01  9.595e-02   3.286 0.001039 **
## GarageYrBlt  -1.846e+01  4.133e+00  -4.466 8.57e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32830 on 1437 degrees of freedom
## Multiple R-squared:  0.8318, Adjusted R-squared:  0.8292
## F-statistic: 322.9 on 22 and 1437 DF,  p-value: < 2.2e-16
```

```
# Removing MasVnrType
house_model.lm <- lm(data = my_train_house, formula = SalePrice ~ OverallQual + Neighborhood + ExterQual
                       BsmtQual + X1stFlrSF + TotRmsAbvGrd + FireplaceQu + YearBuilt +
                       MSSubClass + MasVnrArea + HeatingQC + BsmtFinSF1 + BsmtExposure + SaleType +
                       WoodDeckSF + X2ndFlrSF + GarageQual + LotArea + GarageYrBlt)

summary(house_model.lm)
```

```
##
## Call:
## lm(formula = SalePrice ~ OverallQual + Neighborhood + ExterQual +
##     KitchenQual + GarageCars + BsmtQual + X1stFlrSF + TotRmsAbvGrd +
##     FireplaceQu + YearBuilt + MSSubClass + MasVnrArea + HeatingQC +
##     BsmtFinSF1 + BsmtExposure + SaleType + WoodDeckSF + X2ndFlrSF +
##     GarageQual + LotArea + GarageYrBlt, data = my_train_house)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -434475  -16370      21   14718  293354
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.792e+05  9.730e+04   2.869 0.004175 **
## OverallQual   1.117e+04  1.203e+03   9.288  < 2e-16 ***
## Neighborhood  2.078e+03  2.195e+02   9.468  < 2e-16 ***
## ExterQual     9.537e+03  2.621e+03   3.639 0.000284 ***
## KitchenQual   1.093e+04  2.032e+03   5.376 8.89e-08 ***
## GarageCars    1.180e+04  1.949e+03   6.056 1.77e-09 ***
## BsmtQual      5.408e+03  1.750e+03   3.091 0.002036 **
## X1stFlrSF     4.609e+01  4.256e+00  10.830  < 2e-16 ***
## TotRmsAbvGrd  1.577e+03  9.702e+02   1.626 0.104259
## FireplaceQu   2.347e+03  7.770e+02   3.020 0.002570 **
## YearBuilt    -2.076e+02  5.149e+01  -4.032 5.83e-05 ***
## MSSubClass    6.911e+02  2.888e+02   2.393 0.016854 *
## MasVnrArea    2.543e+01  5.542e+00   4.589 4.85e-06 ***
## HeatingQC     2.080e+03  1.103e+03   1.886 0.059550 .
## BsmtFinSF1    1.640e+01  2.308e+00   7.107 1.86e-12 ***
## BsmtExposure  4.983e+03  9.559e+02   5.212 2.14e-07 ***
```

```
## SaleType      3.474e+03  6.632e+02   5.238 1.86e-07 ***
## WoodDeckSF    2.338e+01  7.453e+00   3.137 0.001743 **
## X2ndFlrSF     3.503e+01  3.718e+00   9.422  < 2e-16 ***
## GarageQual    9.973e+03  3.442e+03   2.898 0.003818 **
## LotArea       3.314e-01  9.564e-02   3.465 0.000546 ***
## GarageYrBlt  -1.844e+01  4.137e+00  -4.459 8.88e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32860 on 1438 degrees of freedom
## Multiple R-squared:  0.8313, Adjusted R-squared:  0.8289
## F-statistic: 337.5 on 21 and 1438 DF,  p-value: < 2.2e-16
```

```
# Removing TotRmsAbvGrd
house_model.lm <- lm(data = my_train_house, formula = SalePrice ~ OverallQual + Neighborhood + ExterQual
                       BsmtQual + X1stFlrSF + FireplaceQu + YearBuilt +
                       MSSubClass + MasVnrArea + HeatingQC + BsmtFinSF1 + BsmtExposure + SaleType +
                       WoodDeckSF + X2ndFlrSF + GarageQual + LotArea + GarageYrBlt)

summary(house_model.lm)
```

```
##
## Call:
## lm(formula = SalePrice ~ OverallQual + Neighborhood + ExterQual +
##     KitchenQual + GarageCars + BsmtQual + X1stFlrSF + FireplaceQu +
##     YearBuilt + MSSubClass + MasVnrArea + HeatingQC + BsmtFinSF1 +
##     BsmtExposure + SaleType + WoodDeckSF + X2ndFlrSF + GarageQual +
##     LotArea + GarageYrBlt, data = my_train_house)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -439630  -16367    -558   14537  287441
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.906e+05  9.710e+04   2.992 0.002816 **
## OverallQual   1.112e+04  1.203e+03   9.247  < 2e-16 ***
## Neighborhood  2.076e+03  2.196e+02   9.452  < 2e-16 ***
## ExterQual     9.435e+03  2.622e+03   3.599 0.000330 ***
## KitchenQual   1.088e+04  2.033e+03   5.352 1.01e-07 ***
## GarageCars    1.209e+04  1.942e+03   6.222 6.41e-10 ***
## BsmtQual      5.268e+03  1.749e+03   3.012 0.002637 **
## X1stFlrSF     5.032e+01  3.369e+00  14.939  < 2e-16 ***
## FireplaceQu   2.336e+03  7.774e+02   3.004 0.002707 **
## YearBuilt    -2.106e+02  5.148e+01  -4.091 4.54e-05 ***
## MSSubClass    7.106e+02  2.888e+02   2.461 0.013969 *
## MasVnrArea    2.522e+01  5.544e+00   4.549 5.84e-06 ***
## HeatingQC     2.040e+03  1.103e+03   1.849 0.064717 .
## BsmtFinSF1    1.583e+01  2.282e+00   6.937 6.05e-12 ***
## BsmtExposure  4.927e+03  9.558e+02   5.155 2.90e-07 ***
## SaleType      3.541e+03  6.623e+02   5.347 1.04e-07 ***
## WoodDeckSF    2.331e+01  7.457e+00   3.127 0.001804 **
## X2ndFlrSF     3.935e+01  2.598e+00  15.148  < 2e-16 ***
## GarageQual    9.823e+03  3.443e+03   2.853 0.004388 **
```

```
## LotArea       3.287e-01  9.568e-02   3.436 0.000607 ***
## GarageYrBlt  -1.867e+01  4.137e+00  -4.513 6.92e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32880 on 1439 degrees of freedom
## Multiple R-squared:  0.831,  Adjusted R-squared:  0.8287
## F-statistic: 353.9 on 20 and 1439 DF,  p-value: < 2.2e-16
```

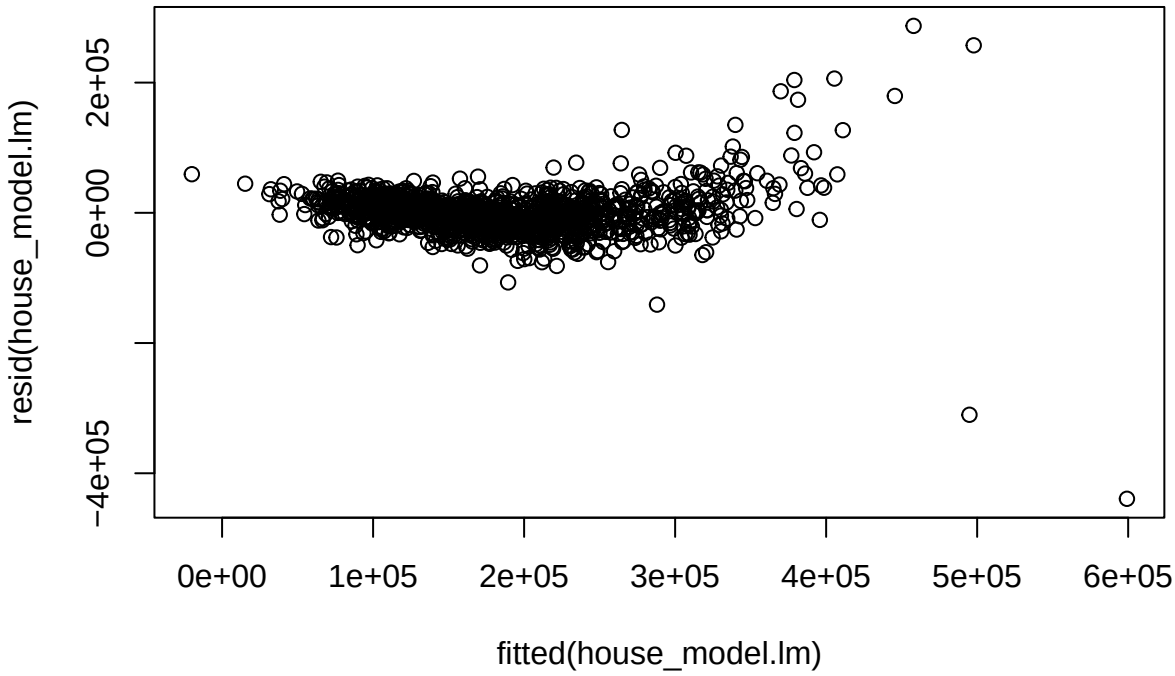
```
# Removing HeatingQC
house_model.lm <- lm(data = my_train_house, formula = SalePrice ~ OverallQual + Neighborhood + ExterQual
                       BsmtQual + X1stFlrSF + FireplaceQu + YearBuilt +
                       MSSubClass + MasVnrArea + BsmtFinSF1 + BsmtExposure + SaleType +
                       WoodDeckSF + X2ndFlrSF + GarageQual + LotArea + GarageYrBlt)

summary(house_model.lm)
```

```
##
## Call:
## lm(formula = SalePrice ~ OverallQual + Neighborhood + ExterQual +
##     KitchenQual + GarageCars + BsmtQual + X1stFlrSF + FireplaceQu +
##     YearBuilt + MSSubClass + MasVnrArea + BsmtFinSF1 + BsmtExposure +
##     SaleType + WoodDeckSF + X2ndFlrSF + GarageQual + LotArea +
##     GarageYrBlt, data = my_train_house)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -439193  -16075    -182   14308  287097
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.719e+05  9.666e+04   2.813 0.004980 **
## OverallQual   1.119e+04  1.204e+03   9.294  < 2e-16 ***
## Neighborhood  2.087e+03  2.197e+02   9.496  < 2e-16 ***
## ExterQual     1.017e+04  2.593e+03   3.922 9.18e-05 ***
## KitchenQual   1.154e+04  2.004e+03   5.757 1.04e-08 ***
## GarageCars    1.198e+04  1.943e+03   6.165 9.13e-10 ***
## BsmtQual      5.381e+03  1.749e+03   3.076 0.002134 **
## X1stFlrSF     5.026e+01  3.371e+00  14.909  < 2e-16 ***
## FireplaceQu   2.334e+03  7.781e+02   3.000 0.002744 **
## YearBuilt    -1.999e+02  5.120e+01  -3.905 9.86e-05 ***
## MSSubClass    7.150e+02  2.890e+02   2.474 0.013465 *
## MasVnrArea    2.461e+01  5.539e+00   4.444 9.52e-06 ***
## BsmtFinSF1    1.564e+01  2.282e+00   6.854 1.06e-11 ***
## BsmtExposure  4.921e+03  9.566e+02   5.144 3.05e-07 ***
## SaleType      3.575e+03  6.626e+02   5.396 7.96e-08 ***
## WoodDeckSF    2.325e+01  7.463e+00   3.115 0.001874 **
## X2ndFlrSF     3.956e+01  2.598e+00  15.227  < 2e-16 ***
## GarageQual    9.758e+03  3.445e+03   2.832 0.004690 **
## LotArea       3.253e-01  9.574e-02   3.398 0.000697 ***
## GarageYrBlt  -1.858e+01  4.140e+00  -4.488 7.77e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

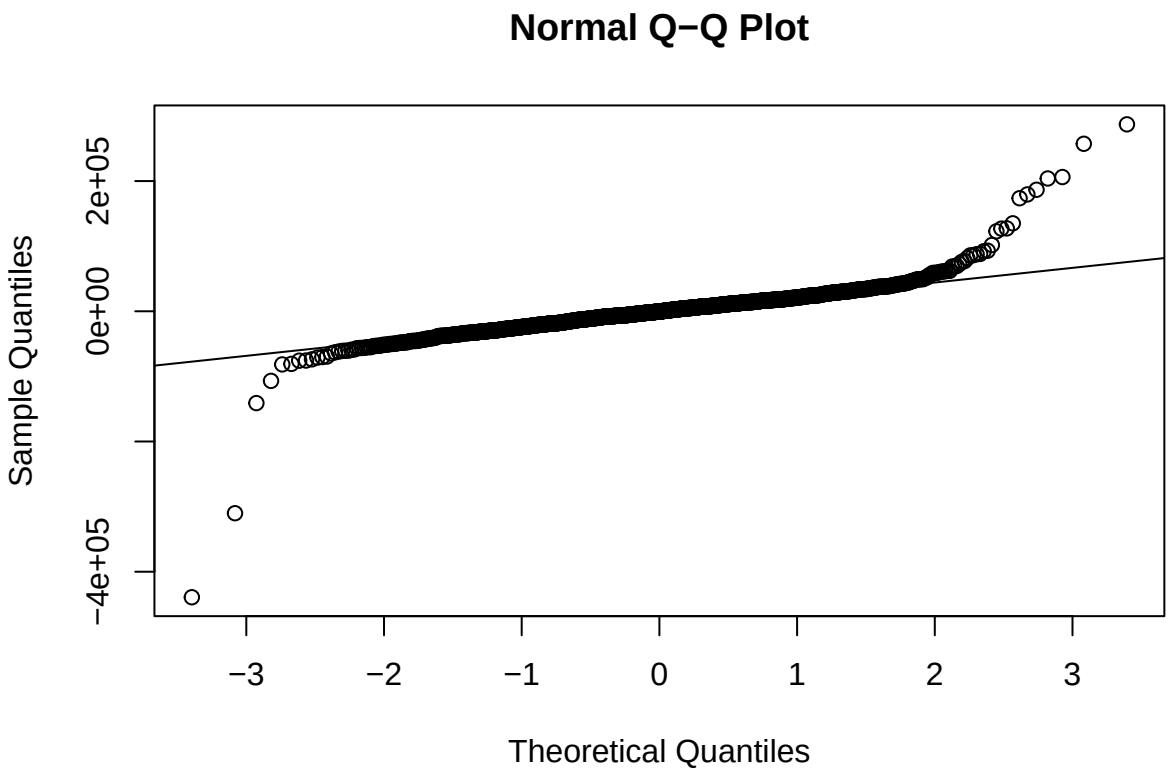
```
## Residual standard error: 32910 on 1440 degrees of freedom
## Multiple R-squared:  0.8306, Adjusted R-squared:  0.8284
## F-statistic: 371.7 on 19 and 1440 DF,  p-value: < 2.2e-16
```

Part 3D5: Evaluate Model All of the remaining variables have p-values less than .05. The Adjusted R^2 value indicates that these variables are accounting for 82% of the variation in Sale Price.

```
plot(fitted(house_model.lm),resid(house_model.lm))
```




```
qqnorm(resid(house_model.lm))
qqline(resid(house_model.lm))
```



The QQ-plot shows that the model is nearly normal within the ±2 standard deviations, only deviating at the ends.

```
my_test_house <- house_clean(test_house)

house_prediction <- predict(house_model.lm, newdata = my_test_house)
summary(house_prediction)
```

Part 3D6: Predict Test Dataset

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   -4771  125433  165231  177745  219812  575687
```

Part 3D7: Submit Predictions Clearly we can't have a home sale price less than 0, so that will be an issue. Lets examine the record.

```
test_house[757,c('OverallQual','Neighborhood','ExterQual','KitchenQual','GarageCars','BsmtQual','X1stFl
```

```
##     OverallQual Neighborhood ExterQual KitchenQual GarageCars BsmtQual
## 757           1       IDOTRR        Fa          Fa          2     <NA>
##     X1stFlrSF FireplaceQu YearBuilt MSSubClass MasVnrArea BsmtFinSF1
## 757       733        <NA>      1952         20          0          0
```

```
##     BsmtExposure SaleType WoodDeckSF X2ndFlrSF GarageQual LotArea GarageYrBlt
## 757         <NA>       WD          0         0         Fa   14584        1952
```

Overall Quality is poor, and it is in a weak neighborhood with a number of fair quality features, these all explain the low price.

Create the csv for submission.

```
house_prediction_df <- as.data.frame(house_prediction)
house_prediction_df$Id <- test_house$Id
write.csv(house_prediction_df[,c(2,1)],'house_submission.csv',row.names = FALSE, quote = FALSE)
```

Submitted results to kaggle, https://www.kaggle.com/donaldbutler95/competitions, and was scored at .31130.

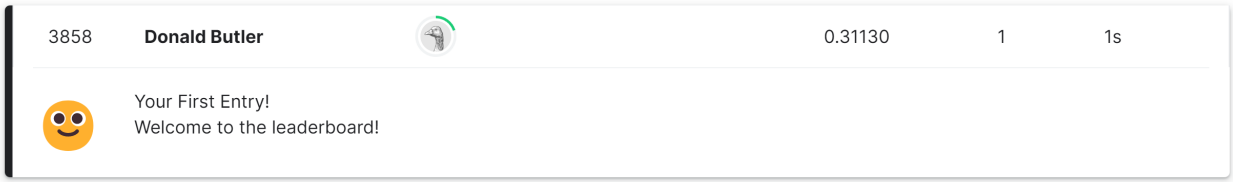


Figure 2: Kaggle Submission