Final_Report

Sai Deepthi Matam, Sri Harsha Samanthula

INTRODUCTION

The primary requirement is a real-time effective model to predict final selling price of houses in the city of Ames, Iowa.

OBJECTIVE

Initial focus of the project is to gain knowledge of the data and understand the relation between each of the variables to the house's sale price. Later, further statistical analysis will be conducted to see how strongly these features effect the house prices the most.

Later, using the training data set, a best-fitting model will be constructed with the some strong variables as predictors of housing prices. Performance of various statistical models will be compared against each other to determine which model fits the best. The results from this projet will finally help in estimating the house prices in Ames, Iowa for given feature set.

ABOUT THE DATA

The data set available on Kaggle contains 80 variables that involve in assessing home values. Out of these, 20 are continuous, 14 are discrete and the remaining 46 are categorical variables. This data has been randomized and then split in to two sets(train and test) of equal size. "SalePrice" is the outcome variable

Exploring the data further we found certain columns have missing values(NAs). Below is the summary of all missing value information.

Train Set ::

	No_of_NAs	% of Missing variables	Numerical(y/n)
LotFrontage	259	17.74	Y
Alley	1369	93.77	N
MasVnrType	8	0.55	N
MasVnrArea	8	0.55	Y
BsmtQual	37	2.53	N
BsmtCond	37	2.53	N
BsmtExposure	38	2.60	N
BsmtFinType1	37	2.53	N
BsmtFinType2	38	2.60	N
Electrical	1	0.07	N
FireplaceQu	690	47.26	N
GarageType	81	5.55	N
GarageYrBlt	81	5.55	Y
GarageFinish	81	5.55	N
GarageQual	81	5.55	N
GarageCond	81	5.55	N
PoolQC	1453	99.52	N
Fence	1179	80.75	N
MiscFeature	1406	96.30	N

Test Set ::

	No_of_NAs	% of Missing variables	Numerical(y/n)
MSZoning	4	0.27	N
LotFrontage	227	15.55	Y
Alley	1352	92.60	N
Utilities	2	0.14	N
Exterior1st	1	0.07	N
Exterior2nd	1	0.07	N
MasVnrType	16	1.10	N
MasVnrArea	15	1.03	Y
BsmtQual	44	3.01	N
BsmtCond	45	3.08	N
BsmtExposure	44	3.01	N
BsmtFinType1	42	2.88	N
BsmtFinSF1	1	0.07	Y
BsmtFinType2	42	2.88	N
BsmtFinSF2	1	0.07	Y
BsmtUnfSF	1	0.07	Y
TotalBsmtSF	1	0.07	Y
BsmtFullBath	2	0.14	Y
BsmtHalfBath	2	0.14	Y
KitchenQual	1	0.07	N
Functional	2	0.14	N
FireplaceQu	730	50.00	N
GarageType	76	5.21	N
GarageYrBlt	78	5.34	Y
GarageFinish	78	5.34	N
GarageCars	1	0.07	Y
GarageArea	1	0.07	Y
GarageQual	78	5.34	N
GarageCond	78	5.34	N
PoolQC	1456	99.73	N
Fence	1169	80.07	N
MiscFeature	1408	96.44	N
SaleType	1	0.07	N
### DATA CLEANI	NG		

Numerical Variables NAs in numeric variables: Since these variables have an impact on the outcome variables, they can not be ignored. Also, the number of missing values for each variable is significantly higher which might introduce a substantial amount of bias or create reductions in efficiency. To avoid this, Imputation has been performed and Include methods on these variables. Imputation is a process of replacing missing data with an estimated value based on other available information.

Here, In train set out of 79 varaibles, there are only 3 variables that has missing values, while in test set there are about 11 variables have missing values. Single imputations works well in this case. So, we used Bagimpute

Charactor/ Categorical Variables NAs in character variables: All character variables contain the category of a certain feature available in the house. As per the data description from Kaggle, NAs in such cases means absence of that feature. Hence, replacing NAs with proper descriptive words.

DATA VISUALIZATION

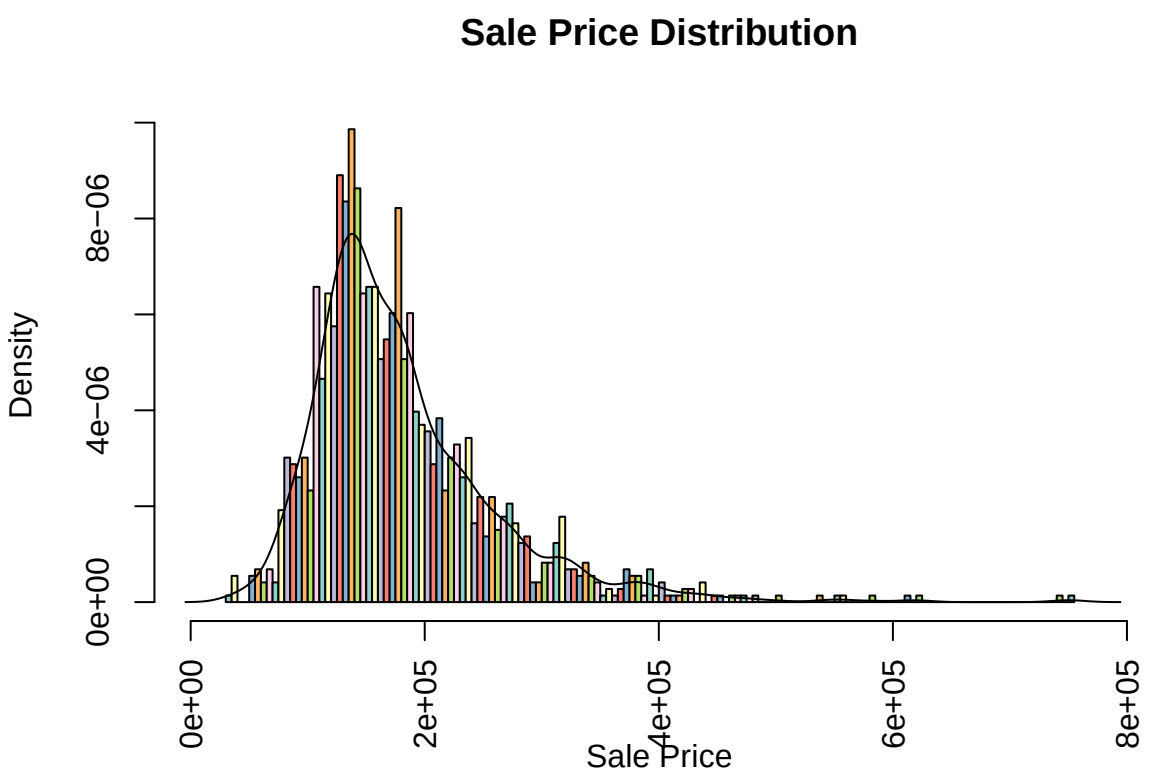
To understand the spread of the Sale Price of houses in Ames.

```
## Mean :  180921.2
```

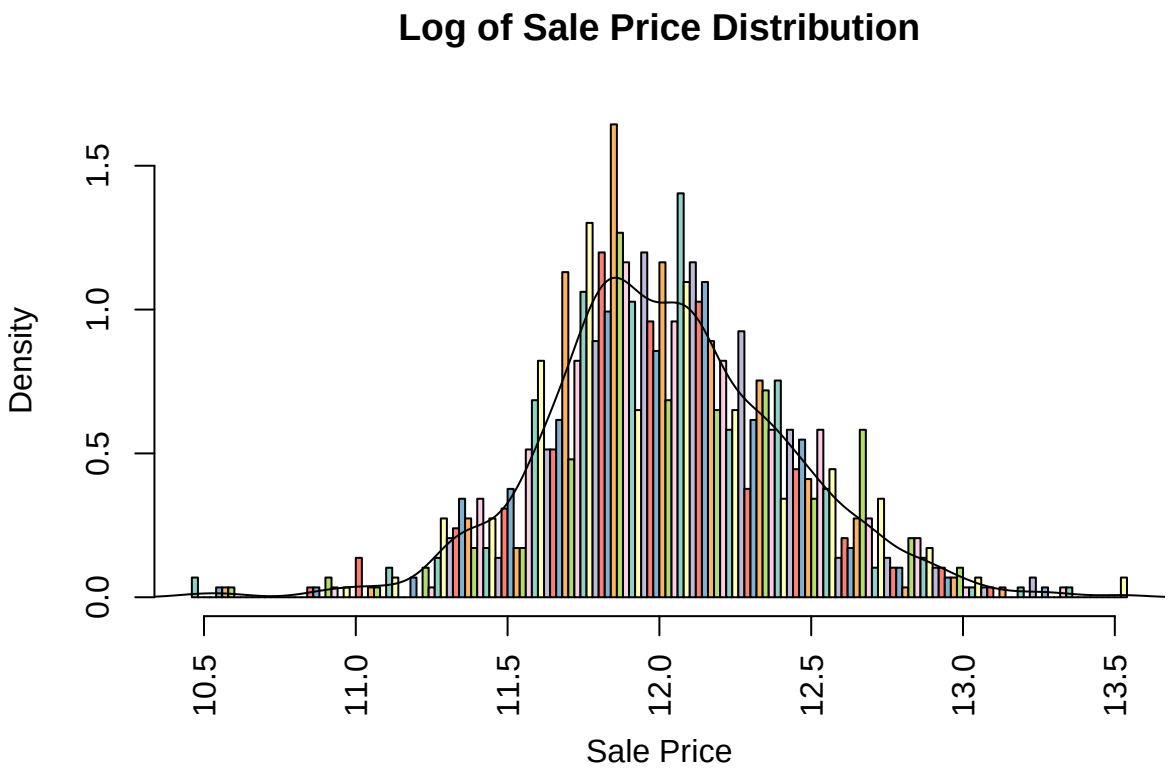
```
## Median :  163000
```

```
## Standard Deviation :  79442.5
```

Here the Mean > Median which indicates a right skew in the data. The same is also plotted below:



This histogram clearly shows that distribution of SalesPrice is Skewed to the right. To rectify this we need to apply log or power functions to SalesPrice variable.

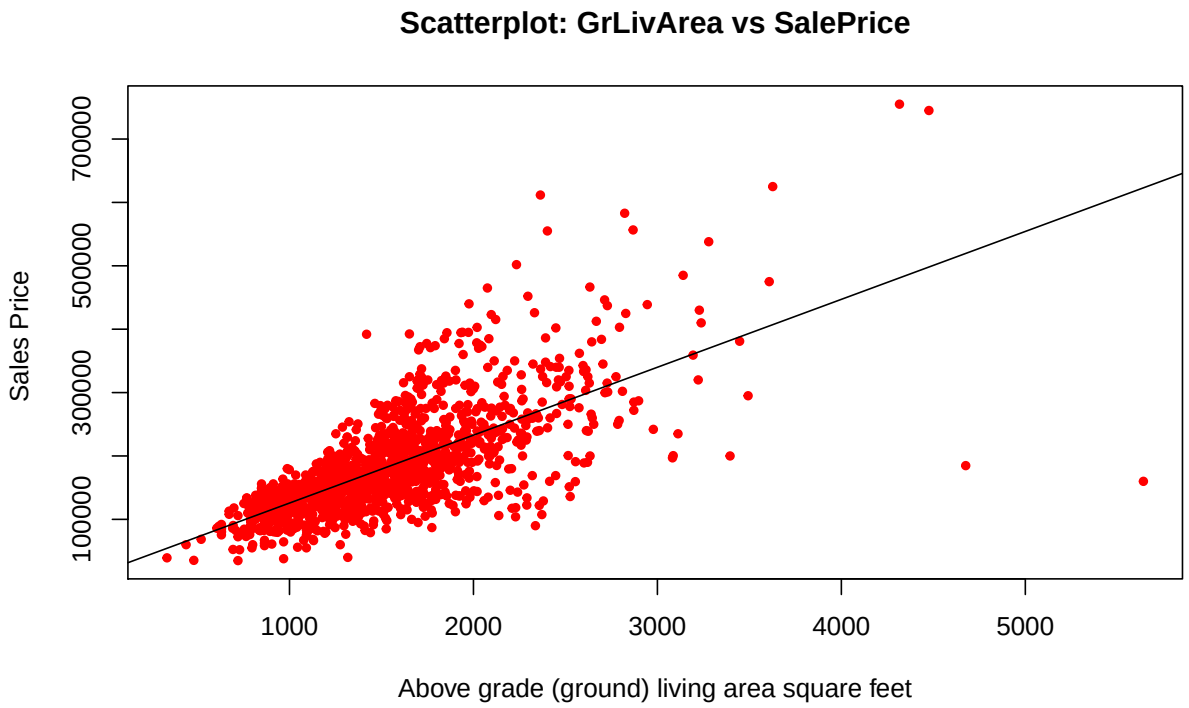


After applying the log function to the SalePrice, the distribution is closer to a normal distribution. Hence we can apply central limit theorm.

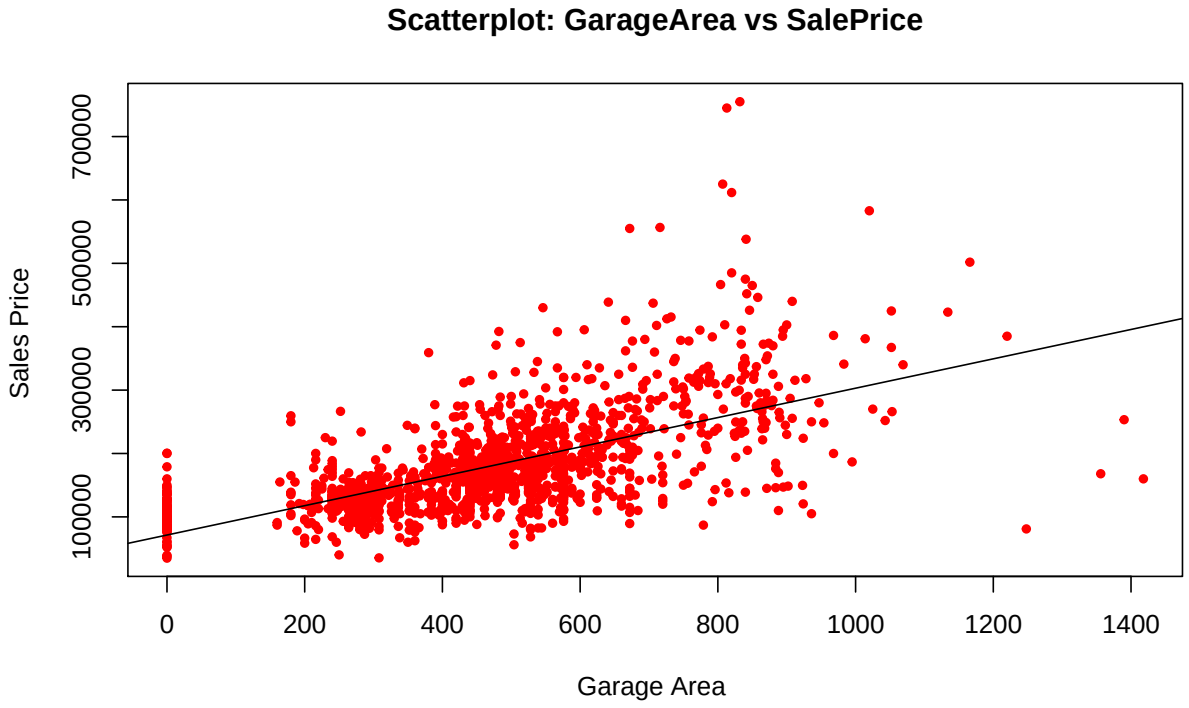
Top 5 Correlation Numerical Variables

Features	Cors
OverallQual	0.7909816
GrLivArea	0.7086245
GarageCars	0.6404092
GarageArea	0.6234314
TotalBsmtSF	0.6135806
X1stFlrSF	0.6058522

Exploring top 5 correlated features using Scatterplots, BoxPlots etc

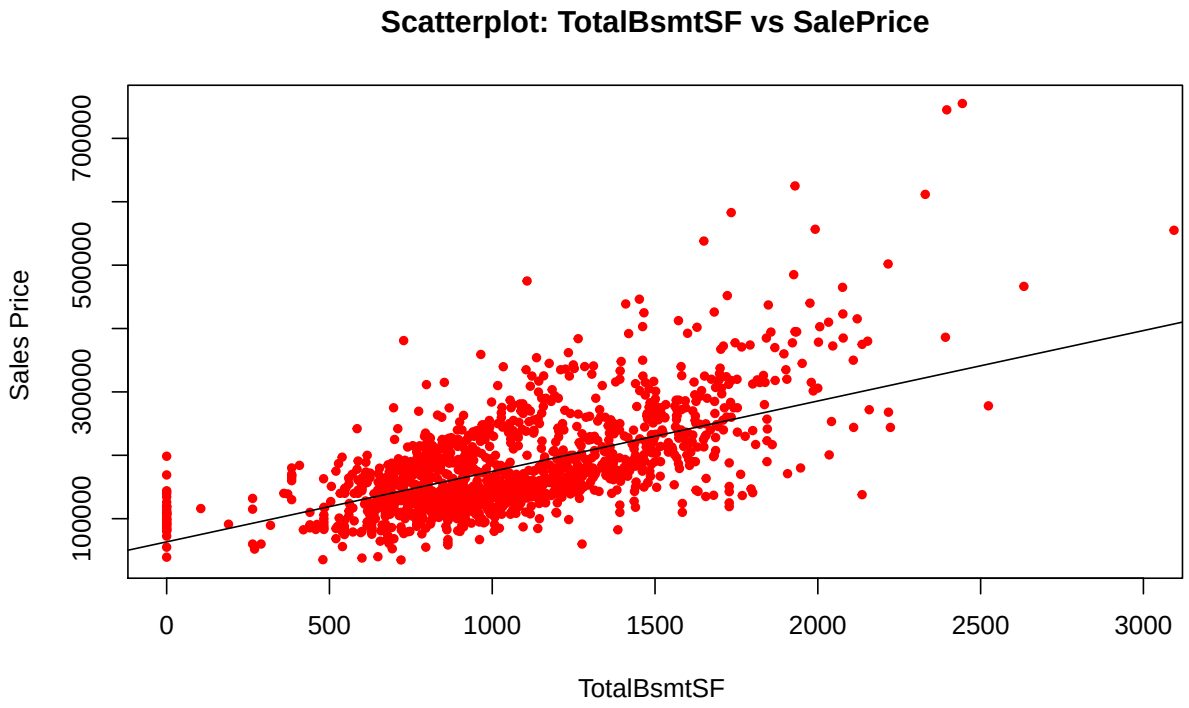


This plot clearly shows that the Living area above grade has a strong positive linear relationship with the Sale price.

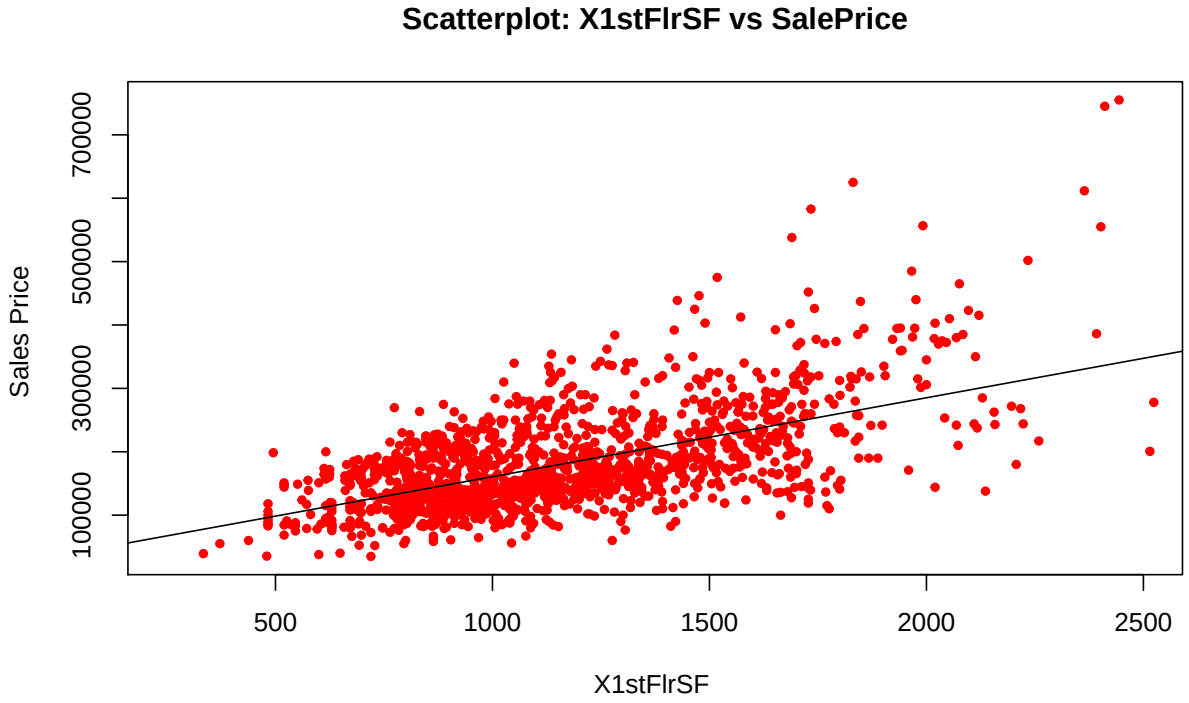


This plot clearly shows that the Garage Area has a strong positive linear relationship with the Sale price.But, this graph has lot of data points concentrated at units '0' which results in an anomaly. There are considerable

amount of houses with no basement at all. That resulted in this anomaly

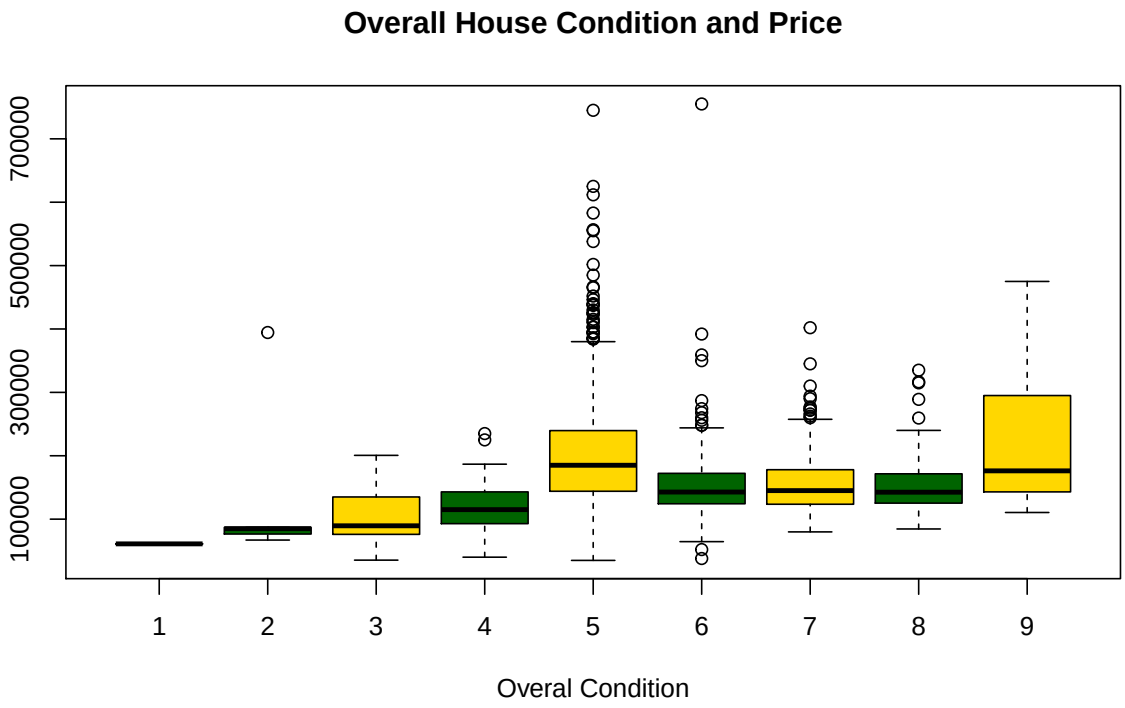


This plot clearly shows that the Total Basement Area has a strong positive linear relationship with the Sale price.But, this graph has lot of data points concentrated at units '0' which results in an anomaly. There are considerable amount of houses with no basement at all. That resulted in this anomaly

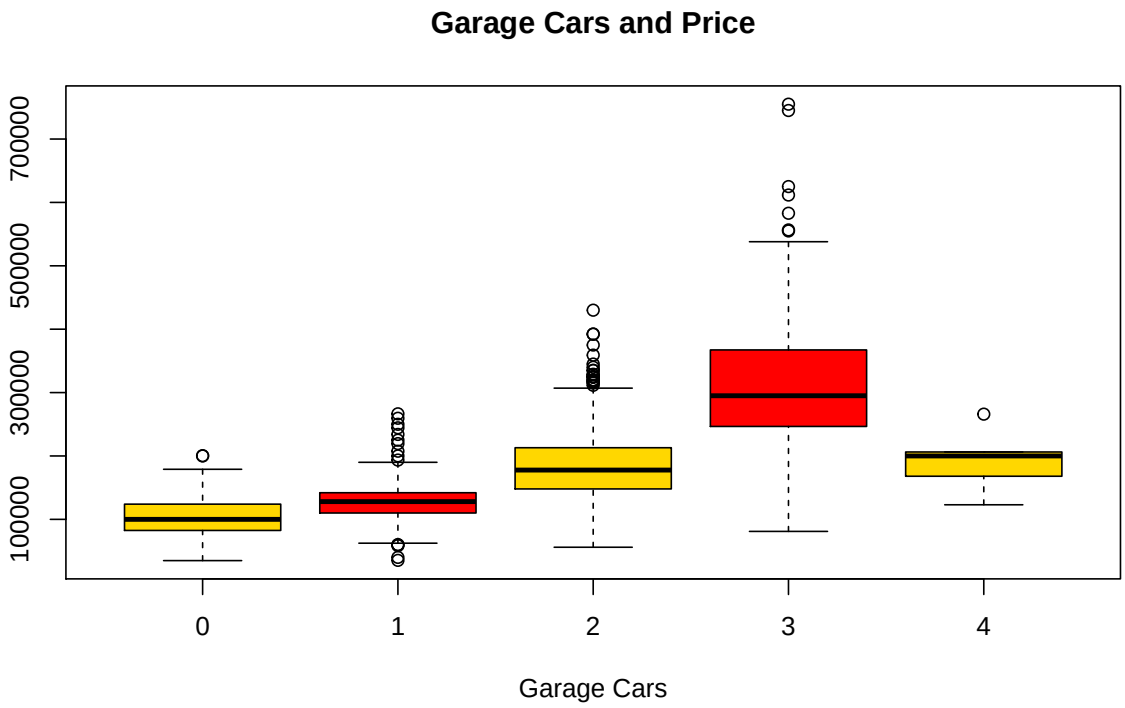


This plot clearly shows that the First Floor area has a strong positive linear relationship with the Sale price.

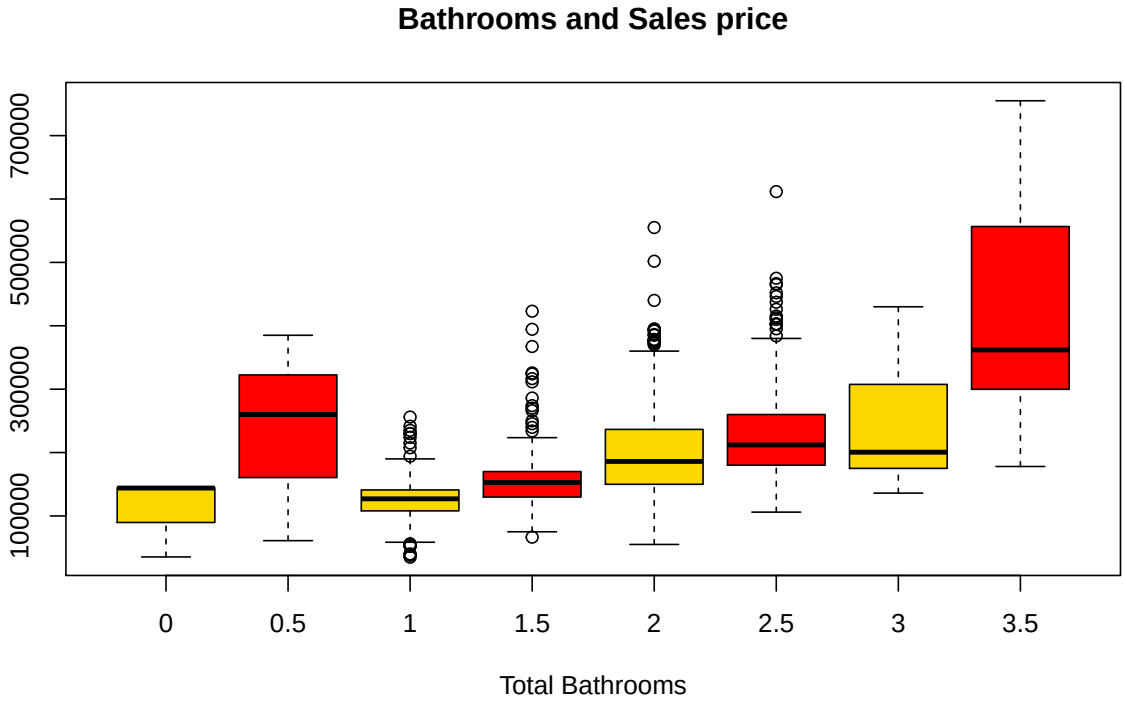
This violin plot shows probability density of the data at different values. For a house with maximum(10) Over all Quality has very high spread and distribution is close to normal where as Over all Quality with 2 has no standard probability and has minimum spread. Rest of the values has close to normal distribution with mean value increasing as the Over all Quality increase



It is quiet evident that OverallCond with 5 units has many outliers and mean sales price of houses with more than 5 rating for Over all condition is similar

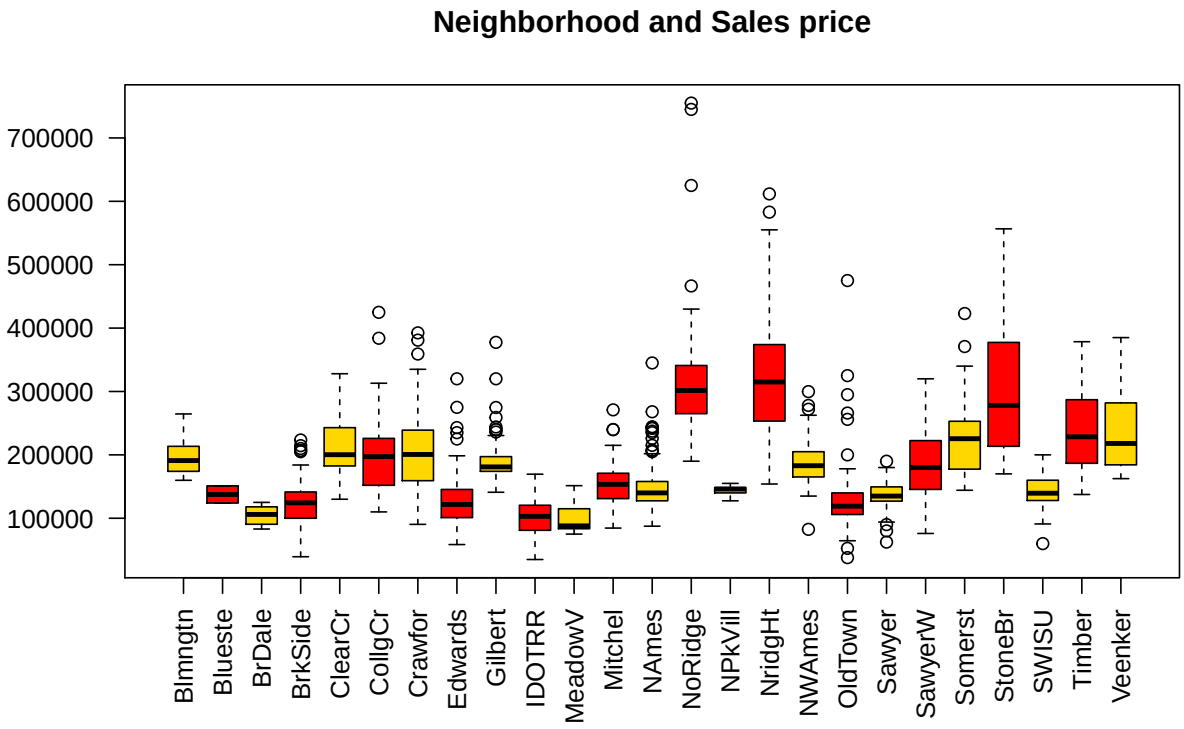


This plot shows that houses with 3 car Garage Space has suprisingly greater mean than the rest of the values

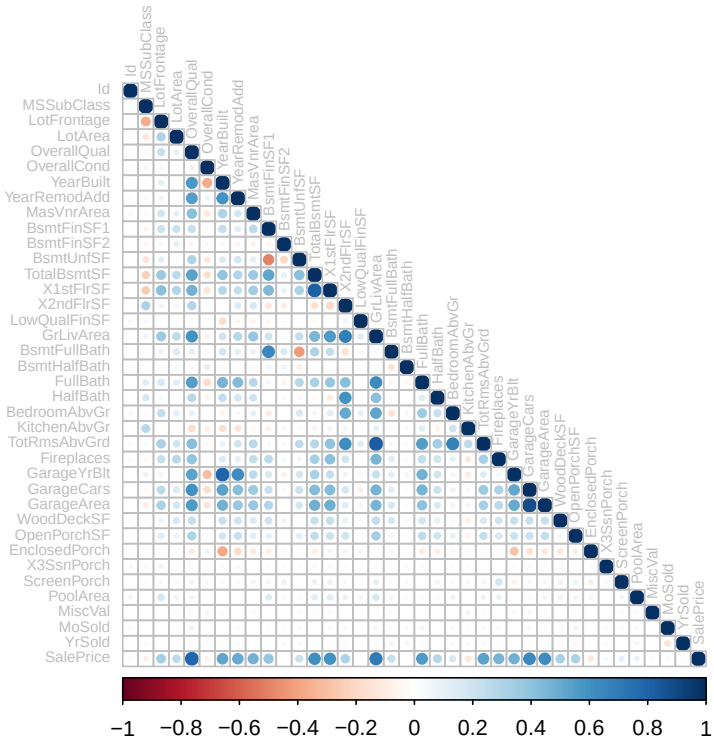


Data given has Full and Half bathrooms. Here, we combined those columns to see data so that both full and half bathroom quantity is quantized in a single value. Box plot clearly shows that prices for each value of

1,1.5, 2 and 2.5 house prices are quite similar to each other as the width of box is short



Viewing the Correlation Plot



Above Correlation heat map helps to visualize correlation between different combinations of variables

Inspecting Multicolinearity between features in order to eliminate highly corelated features.

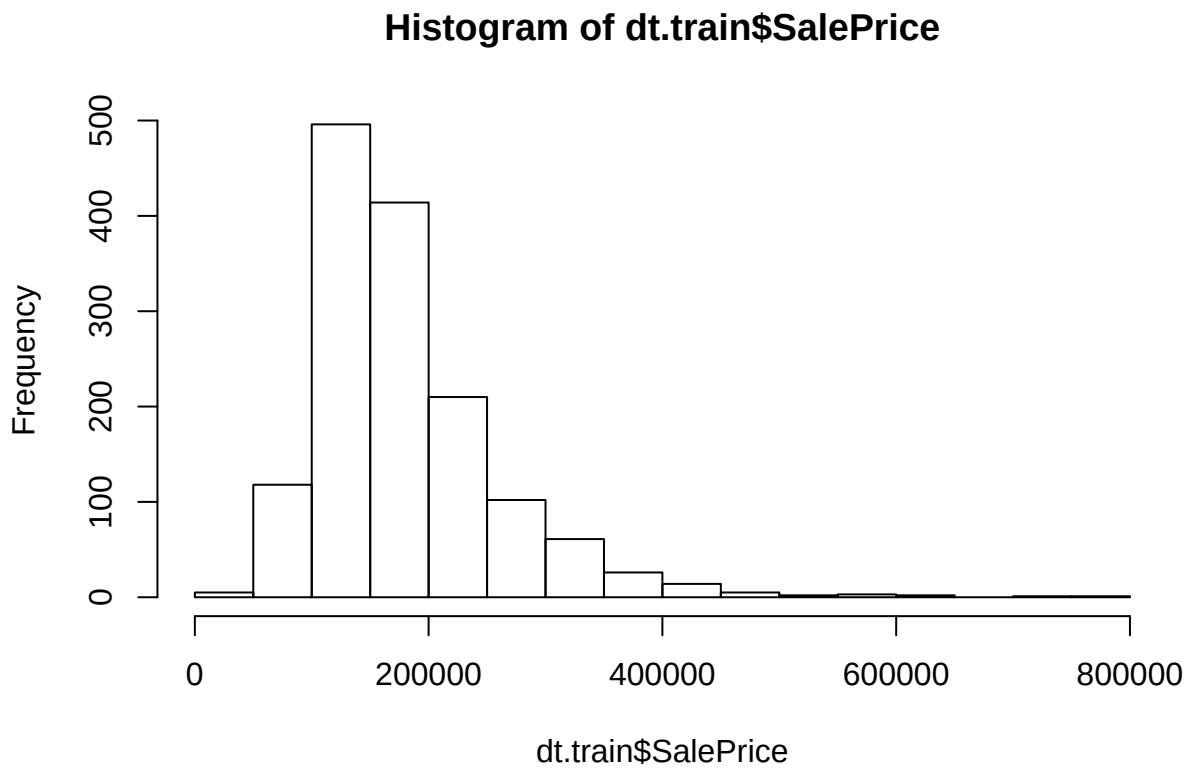
Following table contains the combinations of variables with highest correlation which has a minimum of 0.6 as corelation value. This will identify redundant predictors

name1	name2	cor
X1stFlrSF	TotalBsmtSF	0.81953
GrLivArea	X2ndFlrSF	0.6875011
BsmtFullBath	BsmtFinSF1	0.6492118
FullBath	GrLivArea	0.6300116
HalfBath	X2ndFlrSF	0.6097073
TotRmsAbvGrd	X2ndFlrSF	0.6164226
TotRmsAbvGrd	GrLivArea	0.8254894
TotRmsAbvGrd	BedroomAbvGr	0.6766199
GarageYrBlt	YearBuilt	0.8008755
GarageYrBlt	YearRemodAdd	0.6232214
GarageCars	OverallQual	0.6006707
GarageArea	GarageCars	0.8824754

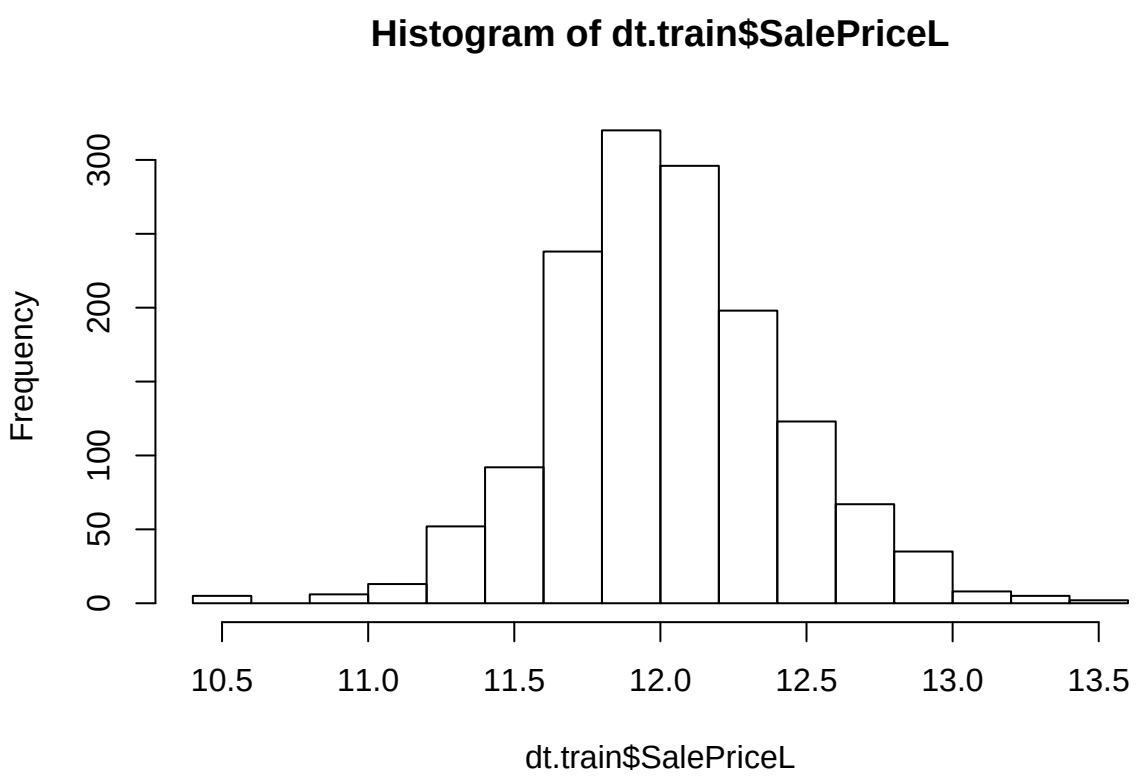
Combining Bath into one variable BsmtFullbath, BsmtHalfBath, FullBath, HalfBath

Feature Selection

```
dt.train1 <- dt.train
dt.test1 <- dt.test

dt.train$SalePriceL <- log(dt.train$SalePrice+1)

hist(dt.train$SalePrice)
```




```
hist(dt.train$SalePriceL)
```




```
dt.train$SalePrice <- dt.train$SalePriceL
dt.train$SalePriceL <- NULL

# train set
skewed_feats <- sapply(numCols.tr,function(x){skewness(dt.train[[x]],na.rm=TRUE)})

skewed_feats <- skewed_feats[skewed_feats > 0.75]

for(x in names(skewed_feats)) {
  dt.train[[x]] <- log(dt.train[[x]] + 1)
}

# test set
skewed_feats <- sapply(numCols.te,function(x){skewness(dt.test[[x]],na.rm=TRUE)})

skewed_feats <- skewed_feats[skewed_feats > 0.75]
for(x in names(skewed_feats)) {
  dt.test[[x]] <- log(dt.test[[x]] + 1)
}


X_train <- dt.train
X_train$Id <- NULL
X_test <- dt.test
y <- dt.train$SalePrice
```

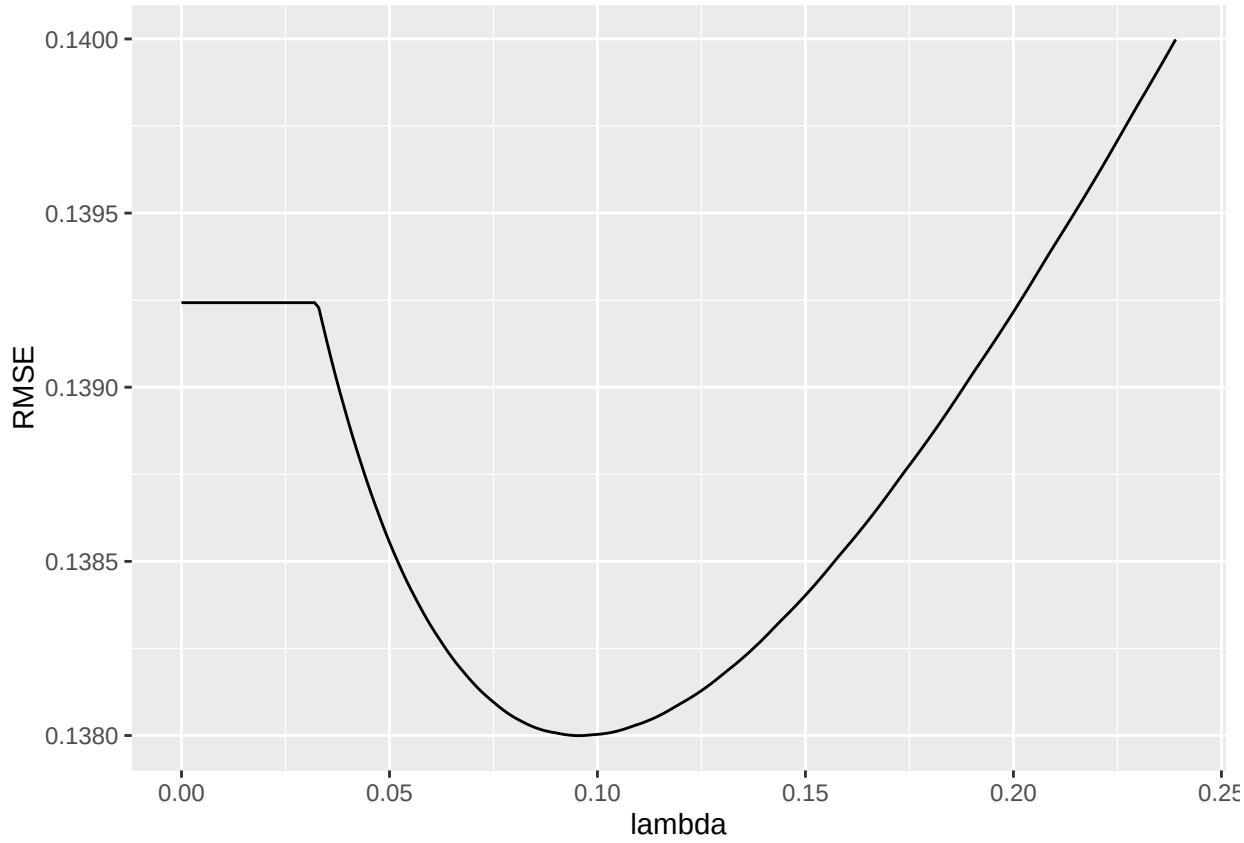
```
lambdas <- seq(1,0,-0.001)

train.cr <- trainControl(method="repeatedcv",
                                 number=5,
                                 repeats=5,
                                 verboseIter=FALSE)
set.seed(123)

model_ridge <- train(SalePrice~.,
                data = X_train,
                  method="glmnet",
                  metric="RMSE",
                preProcess = c("center", "scale"),
                  maximize=FALSE,
                  trControl=train.cr,
                  tuneGrid=expand.grid(alpha=0, # Ridge regression
                                       lambda=lambdas))
```

```
## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: Exterior2ndOther, ElectricalMix

## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: Condition2RRAe, Condition2RRAn,
## RoofMatlMetal, RoofMatlRoll, HeatingOthW, HeatingQCPo, ElectricalSBrKr,
## FunctionalSev

## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: UtilitiesNoSeWa, NeighborhoodBlueste,
## Condition2PosA, Condition2PosN, RoofMatlMembran, Exterior1stAsphShn,
## Exterior1stCBlock, Exterior2ndCBlock

## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: MiscFeatureTenC

## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: Exterior1stImStucc, ExterCondPo

## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: Condition2RRAe, RoofMatlMetal,
## Exterior1stImStucc, MiscFeatureTenC

## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: Condition2RRAn, Exterior1stAsphShn,
## FunctionalSev

## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: UtilitiesNoSeWa, Condition2PosA,
## RoofMatlRoll, Exterior1stCBlock, Exterior2ndCBlock, HeatingQCPo,
## ElectricalSBrKr

## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: RoofMatlMembran, Exterior2ndOther,
## ExterCondPo
```

```
## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: HeatingOthW, ElectricalMix

## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: Exterior1stAsphShn, HeatingQCPo,
## ElectricalSBrKr, FunctionalSev, MiscFeatureTenC

## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: Condition2PosA, Exterior1stStone,
## Exterior2ndOther, ElectricalMix

## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: Condition2PosN, Condition2RRAn,
## RoofMatlMembran, Exterior1stImStucc, ExterCondPo

## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: UtilitiesNoSeWa, Condition2RRAe,
## MiscFeatureOthr

## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: RoofMatlMetal, RoofMatlRoll,
## Exterior1stCBlock, Exterior2ndCBlock

## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: Condition2PosN, Exterior1stImStucc,
## Exterior2ndOther, ExterCondPo, MiscFeatureOthr

## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: UtilitiesNoSeWa, Condition2RRAn,
## Exterior1stAsphShn, FunctionalSev

## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: Exterior1stBrkComm, HeatingQCPo

## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: RoofMatlMembran, Exterior1stCBlock,
## Exterior2ndCBlock, ElectricalSBrKr, MiscFeatureTenC

## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: Condition2PosA, Condition2RRAe,
## RoofMatlMetal, RoofMatlRoll, ElectricalMix

## Warning in preProcess.default(thresh = 0.95, k = 5, method =
## c("center", : These variables have zero variances: Exterior1stAsphShn,
## Exterior1stImStucc, Exterior2ndAsphShn

## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: Condition2RRAe, RoofMatlMembran,
## Exterior2ndOther, HeatingQCPo, MiscFeatureTenC
```

```
## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: Condition2RRAn, RoofMatlMetal,
## FunctionalSev

## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: UtilitiesNoSeWa, ExterCondPo

## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: Condition2PosA, RoofMatlRoll,
## Exterior1stCBlock, Exterior2ndCBlock, ElectricalMix, ElectricalSBrKr
```

```
ggplot(data=filter(model_ridge$result,RMSE<0.14)) +
    geom_line(aes(x=lambda,y=RMSE))
```

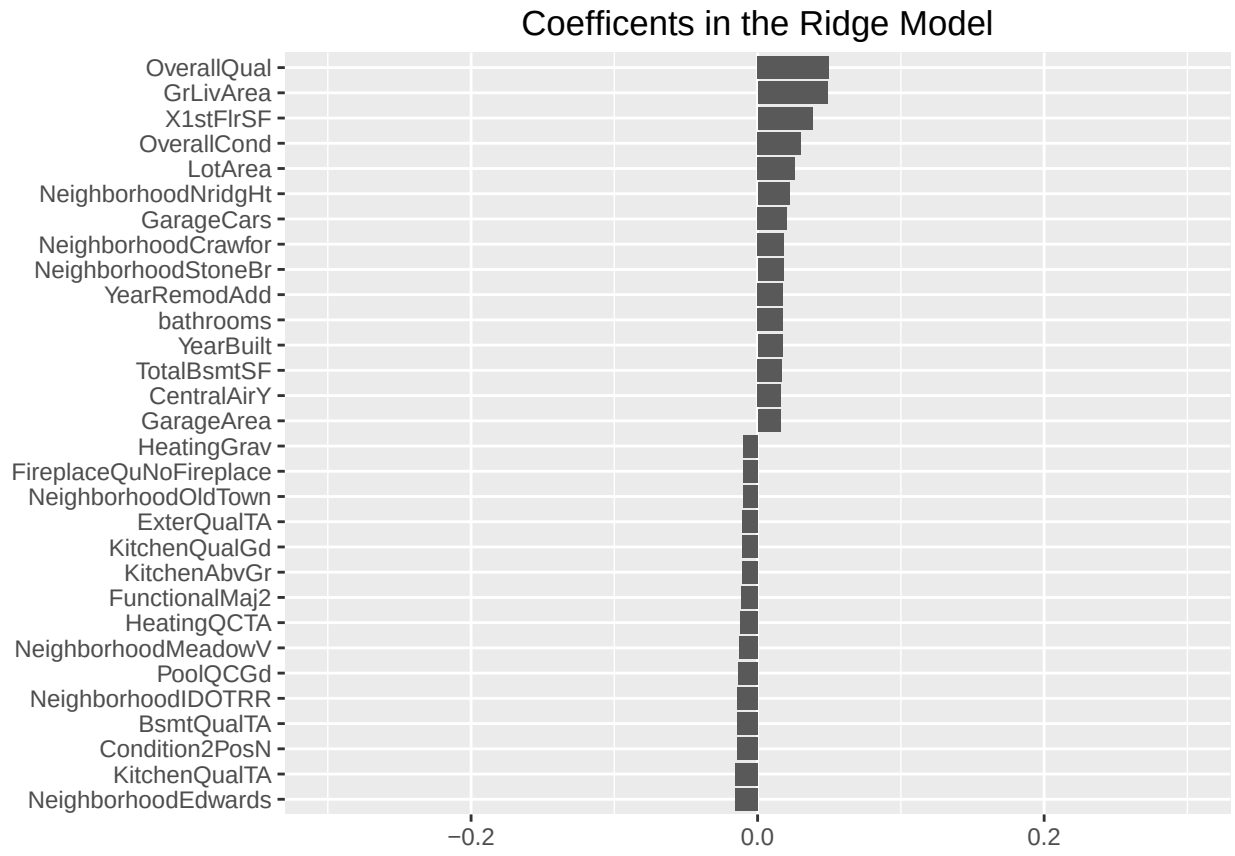



```
coef <- data.frame(coef.name = dimnames(coef(model_ridge$finalModel,s=model_ridge$bestTune$lambda))[[1]]
           coef.value = matrix(coef(model_ridge$finalModel,s=model_ridge$bestTune$lambda)))

# exclude the (Intercept) term
coef <- coef[-1,]

picked_features <- nrow(filter(coef,coef.value!=0))
not_picked_features <- nrow(filter(coef,coef.value==0))

cat("ridge picked",picked_features,"variables and eliminated the other",
    not_picked_features,"variables\n")
```

```
## ridge picked 263 variables and eliminated the other 0 variables
```

```
# sort coefficients in ascending order
coef <- arrange(coef,-coef.value)

# extract the top 10 and bottom 10 features
imp_coef <- rbind(head(coef,15),
                  tail(coef,15))


ggplot(imp_coef) +
    geom_bar(aes(x=reorder(coef.name,coef.value),y=coef.value),
             stat="identity") +
    ylim(-0.3,0.3) +
    coord_flip() +
    ggtitle("Coefficents in the Ridge Model") +
    theme(axis.title=element_blank())
```

```
## Warning: Stacking not well defined when ymin != 0
```




```
set.seed(123)  # for reproducibility
model_lasso <- train(SalePrice~.,
                data = X_train,
                  method="glmnet",
                preProcess = c("center", "scale"),
```

```
                  metric="RMSE",
                  maximize=FALSE,
                  trControl=train.cr,
                  tuneGrid=expand.grid(alpha=1,  # Lasso regression
                                       lambda=c(1,0.1,0.05,0.01,seq(0.009,0.001,-0.001),
                                            0.00075,0.0005,0.0001)))
```

```
## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: Exterior2ndOther, ElectricalMix
```

```
## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: Condition2RRAe, Condition2RRAn,
## RoofMatlMetal, RoofMatlRoll, HeatingOthW, HeatingQCPo, ElectricalSBrKr,
## FunctionalSev
```

```
## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: UtilitiesNoSeWa, NeighborhoodBlueste,
## Condition2PosA, Condition2PosN, RoofMatlMembran, Exterior1stAsphShn,
## Exterior1stCBlock, Exterior2ndCBlock
```

```
## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: MiscFeatureTenC
```

```
## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: Exterior1stImStucc, ExterCondPo
```

```
## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: Condition2RRAe, RoofMatlMetal,
## Exterior1stImStucc, MiscFeatureTenC
```

```
## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: Condition2RRAn, Exterior1stAsphShn,
## FunctionalSev
```

```
## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: UtilitiesNoSeWa, Condition2PosA,
## RoofMatlRoll, Exterior1stCBlock, Exterior2ndCBlock, HeatingQCPo,
## ElectricalSBrKr
```

```
## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: RoofMatlMembran, Exterior2ndOther,
## ExterCondPo
```

```
## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: HeatingOthW, ElectricalMix
```

```
## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: Exterior1stAsphShn, HeatingQCPo,
## ElectricalSBrKr, FunctionalSev, MiscFeatureTenC
```

```
## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: Condition2PosA, Exterior1stStone,
## Exterior2ndOther, ElectricalMix

## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: Condition2PosN, Condition2RRAn,
## RoofMatlMembran, Exterior1stImStucc, ExterCondPo

## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: UtilitiesNoSeWa, Condition2RRAe,
## MiscFeatureOthr

## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: RoofMatlMetal, RoofMatlRoll,
## Exterior1stCBlock, Exterior2ndCBlock

## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: Condition2PosN, Exterior1stImStucc,
## Exterior2ndOther, ExterCondPo, MiscFeatureOthr

## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: UtilitiesNoSeWa, Condition2RRAn,
## Exterior1stAsphShn, FunctionalSev

## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: Exterior1stBrkComm, HeatingQCPo

## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: RoofMatlMembran, Exterior1stCBlock,
## Exterior2ndCBlock, ElectricalSBrKr, MiscFeatureTenC

## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: Condition2PosA, Condition2RRAe,
## RoofMatlMetal, RoofMatlRoll, ElectricalMix

## Warning in preProcess.default(thresh = 0.95, k = 5, method =
## c("center", : These variables have zero variances: Exterior1stAsphShn,
## Exterior1stImStucc, Exterior2ndAsphShn

## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: Condition2RRAe, RoofMatlMembran,
## Exterior2ndOther, HeatingQCPo, MiscFeatureTenC

## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: Condition2RRAn, RoofMatlMetal,
## FunctionalSev

## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: UtilitiesNoSeWa, ExterCondPo
```

```
## Warning in preProcess.default(thresh = 0.95, k = 5, method = c("center", :
## These variables have zero variances: Condition2PosA, RoofMatlRoll,
## Exterior1stCBlock, Exterior2ndCBlock, ElectricalMix, ElectricalSBrKr
```

```
## Warning in nominalTrainWorkflow(x = x, y = y, wts = weights, info =
## trainInfo, : There were missing values in resampled performance measures.
```

```
model_lasso
```

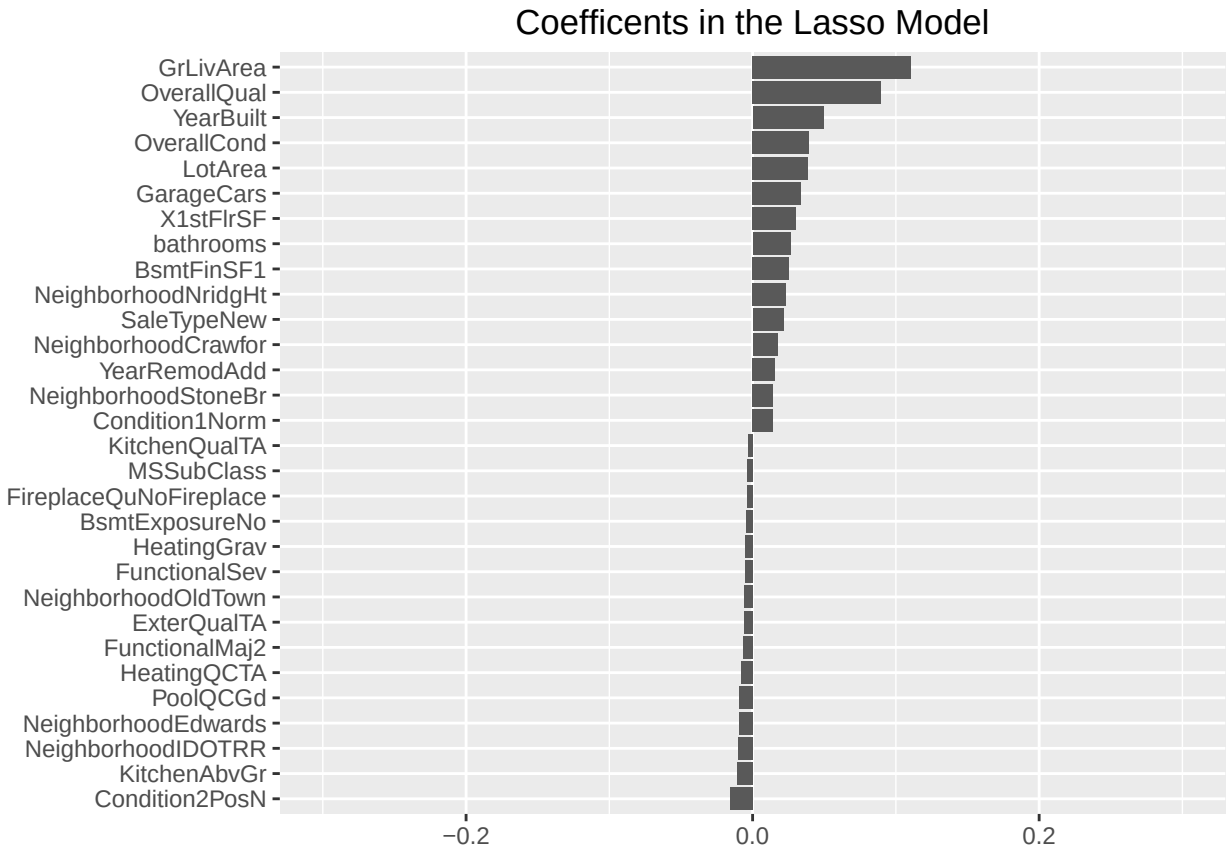
```
## glmnet
##
## 1460 samples
##   82 predictor
##
## Pre-processing: centered (263), scaled (263)
## Resampling: Cross-Validated (5 fold, repeated 5 times)
## Summary of sample sizes: 1169, 1168, 1168, 1167, 1168, 1168, ...
## Resampling results across tuning parameters:
##
##   lambda   RMSE       Rsquared   RMSE SD     Rsquared SD
##   0.00010  0.1543423  0.8539242  0.02905579  0.04843805
##   0.00050  0.1452135  0.8688352  0.02495792  0.04147162
##   0.00075  0.1419580  0.8741623  0.02345801  0.03923135
##   0.00100  0.1398333  0.8776553  0.02258879  0.03798899
##   0.00200  0.1352053  0.8848768  0.02240149  0.03788638
##   0.00300  0.1334883  0.8875431  0.02187979  0.03713974
##   0.00400  0.1327198  0.8888716  0.02115471  0.03605400
##   0.00500  0.1323945  0.8895662  0.02067215  0.03534469
##   0.00600  0.1325017  0.8896417  0.02019582  0.03467087
##   0.00700  0.1329592  0.8891799  0.01982344  0.03424793
##   0.00800  0.1336399  0.8883817  0.01950619  0.03392848
##   0.00900  0.1345144  0.8872738  0.01927401  0.03380675
##   0.01000  0.1354737  0.8860508  0.01902918  0.03365331
##   0.05000  0.1723708  0.8414500  0.01442985  0.02899499
##   0.10000  0.2118968  0.8084506  0.01511017  0.02630537
##   1.00000  0.3990550        NaN  0.01568509          NA
##
## Tuning parameter 'alpha' was held constant at a value of 1
## RMSE was used to select the optimal model using  the smallest value.
## The final values used for the model were alpha = 1 and lambda = 0.005.
```

```
coef <- data.frame(coef.name = dimnames(coef(model_lasso$finalModel,s=model_lasso$bestTune$lambda))[[1]]
           coef.value = matrix(coef(model_lasso$finalModel,s=model_lasso$bestTune$lambda)))

# exclude the (Intercept) term
coef <- coef[-1,]

picked_features <- nrow(filter(coef,coef.value!=0))
not_picked_features <- nrow(filter(coef,coef.value==0))

cat("Lasso picked",picked_features,"variables and eliminated the other",
    not_picked_features,"variables\n")
```

```
## Lasso picked 75 variables and eliminated the other 188 variables
```

```
# sort coefficients in ascending order
coef <- arrange(coef,-coef.value)

# extract the top 10 and bottom 10 features
imp_coef <- rbind(head(coef,15),
                  tail(coef,15))


ggplot(imp_coef) +
    geom_bar(aes(x=reorder(coef.name,coef.value),y=coef.value),
             stat="identity") +
    ylim(-0.3,0.3) +
    coord_flip() +
    ggtitle("Coefficents in the Lasso Model") +
    theme(axis.title=element_blank())
```

```
## Warning: Stacking not well defined when ymin != 0
```




```
coef
```

```
##                  coef.name    coef.value
## 1                GrLivArea  0.1105650351
```

```
## 2              OverallQual  0.0896176990
## 3                YearBuilt  0.0495197096
## 4              OverallCond  0.0395120171
## 5                  LotArea  0.0387951220
## 6               GarageCars  0.0340258725
## 7                X1stFlrSF  0.0304722085
## 8                bathrooms  0.0264672623
## 9               BsmtFinSF1  0.0252244836
## 10     NeighborhoodNridgHt  0.0230836445
## 11             SaleTypeNew  0.0215561730
## 12     NeighborhoodCrawfor  0.0173371444
## 13            YearRemodAdd  0.0156564977
## 14     NeighborhoodStoneBr  0.0142978472
## 15          Condition1Norm  0.0140187567
## 16           FunctionalTyp  0.0140166958
## 17                   porch  0.0131003189
## 18     NeighborhoodSomerst  0.0119197403
## 19     NeighborhoodNoRidge  0.0117302290
## 20         FoundationPConc  0.0113387799
## 21          BsmtExposureGd  0.0112807833
## 22             CentralAirY  0.0111880178
## 23             TotalBsmtSF  0.0110108066
## 24     SaleConditionNormal  0.0109815266
## 25      Exterior1stBrkFace  0.0085204135
## 26              Fireplaces  0.0085133646
## 27              MSZoningRL  0.0068251989
## 28              GarageArea  0.0060650873
## 29         RoofMatlWdShngl  0.0056987174
## 30        LotConfigCulDSac  0.0048634413
## 31             ScreenPorch  0.0040514340
## 32              StreetPave  0.0038202782
## 33     NeighborhoodBrkSide  0.0029888648
## 34           FireplaceQuGd  0.0026797440
## 35             PavedDriveY  0.0024297702
## 36     NeighborhoodClearCr  0.0010734921
## 37             HeatingGasW  0.0010415798
## 38             ExterCondTA  0.0005564458
## 39         BsmtFinType1GLQ  0.0005093489
## 40              MSZoningFV  0.0004697593
## 41            GarageQualGd  0.0004360088
## 42             LotFrontage  0.0001861447
## 43              MSZoningRH  0.0000000000
## 44              MSZoningRM  0.0000000000
## 45           AlleyNo Alley  0.0000000000
## 46               AlleyPave  0.0000000000
## 47             LotShapeIR2  0.0000000000
## 48             LotShapeReg  0.0000000000
## 49          LandContourHLS  0.0000000000
## 50          LandContourLow  0.0000000000
## 51          LandContourLvl  0.0000000000
## 52            LotConfigFR2  0.0000000000
## 53            LotConfigFR3  0.0000000000
## 54         LotConfigInside  0.0000000000
## 55            LandSlopeMod  0.0000000000
```

```
## 56            LandSlopeSev  0.0000000000
## 57     NeighborhoodBlueste  0.0000000000
## 58      NeighborhoodBrDale  0.0000000000
## 59     NeighborhoodCollgCr  0.0000000000
## 60     NeighborhoodGilbert  0.0000000000
## 61     NeighborhoodMitchel  0.0000000000
## 62       NeighborhoodNAmes  0.0000000000
## 63     NeighborhoodNPkVill  0.0000000000
## 64      NeighborhoodSawyer  0.0000000000
## 65     NeighborhoodSawyerW  0.0000000000
## 66       NeighborhoodSWISU  0.0000000000
## 67      NeighborhoodTimber  0.0000000000
## 68     NeighborhoodVeenker  0.0000000000
## 69         Condition1Feedr  0.0000000000
## 70          Condition1PosA  0.0000000000
## 71          Condition1PosN  0.0000000000
## 72          Condition1RRAn  0.0000000000
## 73          Condition1RRNe  0.0000000000
## 74          Condition1RRNn  0.0000000000
## 75         Condition2Feedr  0.0000000000
## 76          Condition2Norm  0.0000000000
## 77          Condition2PosA  0.0000000000
## 78          Condition2RRAe  0.0000000000
## 79          Condition2RRAn  0.0000000000
## 80          Condition2RRNn  0.0000000000
## 81          BldgType2fmCon  0.0000000000
## 82          BldgTypeDuplex  0.0000000000
## 83          BldgTypeTwnhsE  0.0000000000
## 84        HouseStyle1.5Unf  0.0000000000
## 85        HouseStyle1Story  0.0000000000
## 86        HouseStyle2.5Fin  0.0000000000
## 87        HouseStyle2.5Unf  0.0000000000
## 88        HouseStyle2Story  0.0000000000
## 89        HouseStyleSFoyer  0.0000000000
## 90          HouseStyleSLvl  0.0000000000
## 91        RoofStyleGambrel  0.0000000000
## 92            RoofStyleHip  0.0000000000
## 93        RoofStyleMansard  0.0000000000
## 94           RoofStyleShed  0.0000000000
## 95         RoofMatlCompShg  0.0000000000
## 96         RoofMatlMembran  0.0000000000
## 97           RoofMatlMetal  0.0000000000
## 98            RoofMatlRoll  0.0000000000
## 99         RoofMatlTar&Grv  0.0000000000
## 100        RoofMatlWdShake  0.0000000000
## 101     Exterior1stAsphShn  0.0000000000
## 102      Exterior1stCBlock  0.0000000000
## 103     Exterior1stCemntBd  0.0000000000
## 104     Exterior1stImStucc  0.0000000000
## 105     Exterior1stMetalSd  0.0000000000
## 106     Exterior1stPlywood  0.0000000000
## 107       Exterior1stStone  0.0000000000
## 108      Exterior1stStucco  0.0000000000
## 109     Exterior1stVinylSd  0.0000000000
```

```
## 110     Exterior1stWdShing  0.0000000000
## 111     Exterior2ndAsphShn  0.0000000000
## 112     Exterior2ndBrk Cmn  0.0000000000
## 113     Exterior2ndBrkFace  0.0000000000
## 114      Exterior2ndCBlock  0.0000000000
## 115     Exterior2ndCmentBd  0.0000000000
## 116     Exterior2ndHdBoard  0.0000000000
## 117     Exterior2ndImStucc  0.0000000000
## 118     Exterior2ndMetalSd  0.0000000000
## 119       Exterior2ndOther  0.0000000000
## 120     Exterior2ndPlywood  0.0000000000
## 121       Exterior2ndStone  0.0000000000
## 122      Exterior2ndStucco  0.0000000000
## 123     Exterior2ndVinylSd  0.0000000000
## 124     Exterior2ndWd Sdng  0.0000000000
## 125     Exterior2ndWd Shng  0.0000000000
## 126      MasVnrTypeBrkFace  0.0000000000
## 127         MasVnrTypeNone  0.0000000000
## 128        MasVnrTypeStone  0.0000000000
## 129             MasVnrArea  0.0000000000
## 130            ExterQualFa  0.0000000000
## 131            ExterQualGd  0.0000000000
## 132            ExterCondGd  0.0000000000
## 133            ExterCondPo  0.0000000000
## 134       FoundationCBlock  0.0000000000
## 135         FoundationSlab  0.0000000000
## 136        FoundationStone  0.0000000000
## 137             BsmtQualFa  0.0000000000
## 138             BsmtQualGd  0.0000000000
## 139         BsmtQualNoBsmt  0.0000000000
## 140             BsmtQualTA  0.0000000000
## 141             BsmtCondGd  0.0000000000
## 142         BsmtCondNoBsmt  0.0000000000
## 143             BsmtCondPo  0.0000000000
## 144             BsmtCondTA  0.0000000000
## 145         BsmtExposureMn  0.0000000000
## 146     BsmtExposureNoBsmt  0.0000000000
## 147        BsmtFinType1BLQ  0.0000000000
## 148        BsmtFinType1LwQ  0.0000000000
## 149     BsmtFinType1NoBsmt  0.0000000000
## 150        BsmtFinType1Rec  0.0000000000
## 151        BsmtFinType1Unf  0.0000000000
## 152        BsmtFinType2BLQ  0.0000000000
## 153        BsmtFinType2GLQ  0.0000000000
## 154        BsmtFinType2LwQ  0.0000000000
## 155     BsmtFinType2NoBsmt  0.0000000000
## 156        BsmtFinType2Rec  0.0000000000
## 157        BsmtFinType2Unf  0.0000000000
## 158             BsmtFinSF2  0.0000000000
## 159              BsmtUnfSF  0.0000000000
## 160            HeatingGasA  0.0000000000
## 161            HeatingWall  0.0000000000
## 162            HeatingQCFa  0.0000000000
## 163            HeatingQCPo  0.0000000000
```

```
## 164        ElectricalFuseF  0.0000000000
## 165        ElectricalFuseP  0.0000000000
## 166          ElectricalMix  0.0000000000
## 167        ElectricalSBrkr  0.0000000000
## 168        ElectricalSBrKr  0.0000000000
## 169             X2ndFlrSF  0.0000000000
## 170           LowQualFinSF  0.0000000000
## 171           BsmtFullBath  0.0000000000
## 172           BsmtHalfBath  0.0000000000
## 173               FullBath  0.0000000000
## 174               HalfBath  0.0000000000
## 175           BedroomAbvGr  0.0000000000
## 176          KitchenQualFa  0.0000000000
## 177          KitchenQualGd  0.0000000000
## 178           TotRmsAbvGrd  0.0000000000
## 179         FunctionalMin1  0.0000000000
## 180         FunctionalMin2  0.0000000000
## 181          FunctionalMod  0.0000000000
## 182          FireplaceQuFa  0.0000000000
## 183          FireplaceQuPo  0.0000000000
## 184          FireplaceQuTA  0.0000000000
## 185       GarageTypeAttchd  0.0000000000
## 186      GarageTypeBuiltIn  0.0000000000
## 187       GarageTypeDetchd  0.0000000000
## 188     GarageTypeNoGarage  0.0000000000
## 189            GarageYrBlt  0.0000000000
## 190   GarageFinishNoGarage  0.0000000000
## 191        GarageFinishRFn  0.0000000000
## 192        GarageFinishUnf  0.0000000000
## 193           GarageQualFa  0.0000000000
## 194     GarageQualNoGarage  0.0000000000
## 195           GarageQualPo  0.0000000000
## 196           GarageQualTA  0.0000000000
## 197           GarageCondGd  0.0000000000
## 198     GarageCondNoGarage  0.0000000000
## 199           GarageCondPo  0.0000000000
## 200           GarageCondTA  0.0000000000
## 201            PavedDriveP  0.0000000000
## 202             WoodDeckSF  0.0000000000
## 203            OpenPorchSF  0.0000000000
## 204          EnclosedPorch  0.0000000000
## 205             X3SsnPorch  0.0000000000
## 206               PoolArea  0.0000000000
## 207               PoolQCFa  0.0000000000
## 208           PoolQCNoPool  0.0000000000
## 209             FenceMnPrv  0.0000000000
## 210             FenceMnWw  0.0000000000
## 211           FenceNoFence  0.0000000000
## 212     MiscFeatureNoFence  0.0000000000
## 213        MiscFeatureOthr  0.0000000000
## 214        MiscFeatureShed  0.0000000000
## 215        MiscFeatureTenC  0.0000000000
## 216                MiscVal  0.0000000000
## 217                 MoSold  0.0000000000
```

```
## 218                 YrSold  0.0000000000
## 219           SaleTypeCon  0.0000000000
## 220         SaleTypeConLD  0.0000000000
## 221         SaleTypeConLI  0.0000000000
## 222         SaleTypeConLw  0.0000000000
## 223           SaleTypeCWD  0.0000000000
## 224           SaleTypeOth  0.0000000000
## 225            SaleTypeWD  0.0000000000
## 226   SaleConditionAdjLand  0.0000000000
## 227    SaleConditionAlloca  0.0000000000
## 228    SaleConditionFamily  0.0000000000
## 229   SaleConditionPartial  0.0000000000
## 230             totalRoom  0.0000000000
## 231           HeatingOthW -0.0001605583
## 232        FoundationWood -0.0002819546
## 233       UtilitiesNoSeWa -0.0003702506
## 234    Exterior1stHdBoard -0.0006851719
## 235     GarageTypeCarPort -0.0006968637
## 236        RoofStyleGable -0.0008306883
## 237     GarageTypeBasment -0.0009125864
## 238           HeatingQCGd -0.0010680400
## 239    NeighborhoodMeadowV -0.0010935984
## 240             FenceGdWo -0.0013484024
## 241    Exterior1stWd Sdng -0.0014673068
## 242           ExterCondFa -0.0020004057
## 243    NeighborhoodNWAmes -0.0022029429
## 244        Condition1RRAe -0.0022067910
## 245          GarageCondFa -0.0022726765
## 246         BldgTypeTwnhs -0.0026755750
## 247    Exterior1stBrkComm -0.0027055057
## 248           LotShapeIR3 -0.0031862302
## 249         KitchenQualTA -0.0034285742
## 250            MSSubClass -0.0035514233
## 251 FireplaceQuNoFireplace -0.0040921183
## 252         BsmtExposureNo -0.0042511556
## 253           HeatingGrav -0.0051493622
## 254         FunctionalSev -0.0051692884
## 255    NeighborhoodOldTown -0.0059144057
## 256           ExterQualTA -0.0060342737
## 257         FunctionalMaj2 -0.0069479271
## 258           HeatingQCTA -0.0079686484
## 259              PoolQCGd -0.0094190071
## 260    NeighborhoodEdwards -0.0096274007
## 261     NeighborhoodIDOTRR -0.0099236196
## 262          KitchenAbvGr -0.0107378452
## 263        Condition2PosN -0.0154586822
```

```
imp_coef$coef.name
```

```
##  [1] GrLivArea              OverallQual            YearBuilt
##  [4] OverallCond            LotArea                GarageCars
##  [7] X1stFlrSF              bathrooms              BsmtFinSF1
## [10] NeighborhoodNridgHt    SaleTypeNew            NeighborhoodCrawfor
## [13] YearRemodAdd           NeighborhoodStoneBr    Condition1Norm
```

```
## [16] KitchenQualTA          MSSubClass             FireplaceQuNoFireplace
## [19] BsmtExposureNo         HeatingGrav            FunctionalSev
## [22] NeighborhoodOldTown    ExterQualTA            FunctionalMaj2
## [25] HeatingQCTA            PoolQCGd               NeighborhoodEdwards
## [28] NeighborhoodIDOTRR     KitchenAbvGr           Condition2PosN
## 264 Levels: (Intercept) AlleyNo Alley AlleyPave bathrooms ... YrSold
```

```
dt.train$Conditio
```

```
## NULL
```

GrLivArea OverallQual YearBuilt
Neighborhood LotArea GarageCars bathrooms TotalBsmtSF

Functional ExterQual HeatingQC PoolQC
KitchenAbvGr Condition2

OverallCond

Creating model ensemble

```
index <- sample(1:(0.75*nrow(dt.train)), replace = FALSE)

t.train <- dt.train[index,]
nrow(t.train)
```

```
## [1] 1095
```

```
t.test <- dt.train[-index,]
nrow(t.test)
```

```
## [1] 365
```