Step 3 - Feature Selection & Engineering

Hyunkyung Kim

November 1, 2018

#install.packages("glmnet")  
#install.packages("mlbench")  
#install.packages("Boruta")  
library(MASS) # stepwise regression  
library(caret)

## Loading required package: lattice

## Loading required package: ggplot2

library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.2.1 --

## v tibble 1.4.2 v purrr 0.2.4  
## v tidyr 0.8.0 v dplyr 0.7.7  
## v readr 1.1.1 v stringr 1.3.0  
## v tibble 1.4.2 v forcats 0.3.0

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()  
## x purrr::lift() masks caret::lift()  
## x dplyr::select() masks MASS::select()

library(psych)

##   
## Attaching package: 'psych'

## The following objects are masked from 'package:ggplot2':  
##   
## %+%, alpha

library(glmnet)

## Loading required package: Matrix

##   
## Attaching package: 'Matrix'

## The following object is masked from 'package:tidyr':  
##   
## expand

## Loading required package: foreach

##   
## Attaching package: 'foreach'

## The following objects are masked from 'package:purrr':  
##   
## accumulate, when

## Loaded glmnet 2.0-16

library(mlbench)  
library(Boruta)

## Loading required package: ranger

library(leaps) # all subsets regression  
library(randomForest)

## randomForest 4.6-14

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

##   
## Attaching package: 'randomForest'

## The following object is masked from 'package:ranger':  
##   
## importance

## The following object is masked from 'package:psych':  
##   
## outlier

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

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

### Import Clean Data

H\_Clean<-read.csv( file = "C:\\Users\\Hyunkyung Kim\\Desktop\\CKME999\\136\\dataset\\all\\H\_clean.csv")  
  
Train<-H\_Clean[!is.na(H\_Clean$SalePrice),-1]  
Test<-H\_Clean[is.na(H\_Clean$SalePrice),-1]

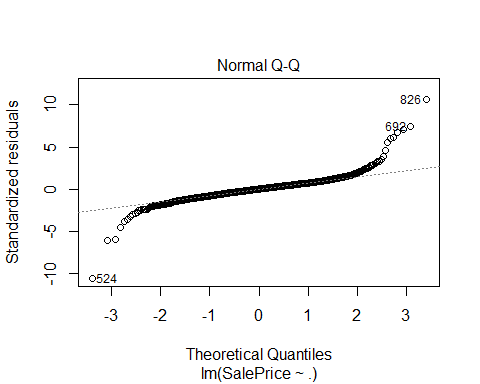
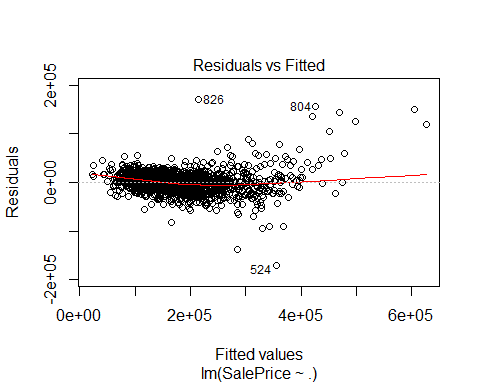
### Check Outlier for Linear Regression - Cook’s Distance.

# Linear Fit Model  
lmfit<-lm(SalePrice~.,data=Train)  
summary(lmfit)

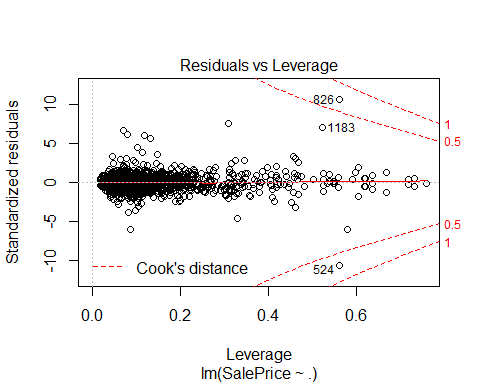
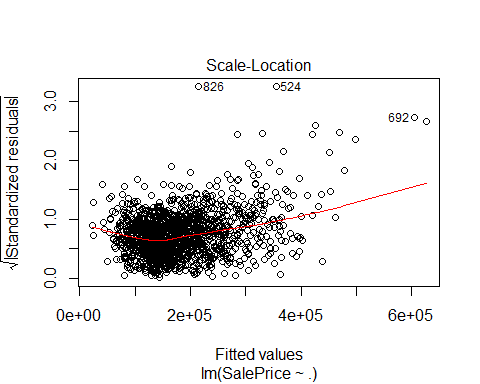
##   
## Call:  
## lm(formula = SalePrice ~ ., data = Train)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -169991 -11292 0 11008 169991   
##   
## Coefficients: (6 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.502e+06 1.095e+06 -1.372 0.170227   
## MSSubClass -7.322e+01 4.445e+01 -1.647 0.099775 .   
## MSZoningFV 2.819e+04 1.254e+04 2.248 0.024760 \*   
## MSZoningRH 2.569e+04 1.253e+04 2.050 0.040560 \*   
## MSZoningRL 2.485e+04 1.063e+04 2.338 0.019530 \*   
## MSZoningRM 2.197e+04 1.002e+04 2.194 0.028435 \*   
## LotFrontage 9.964e+01 4.405e+01 2.262 0.023853 \*   
## LotArea 4.192e-01 9.726e-02 4.310 1.76e-05 \*\*\*  
## StreetPave 2.146e+04 1.209e+04 1.775 0.076120 .   
## AlleyNoAlley -2.126e+03 4.412e+03 -0.482 0.630019   
## AlleyPave -8.441e+01 6.302e+03 -0.013 0.989316   
## LotShapeIR2 8.368e+03 4.423e+03 1.892 0.058760 .   
## LotShapeIR3 4.688e+03 9.309e+03 0.504 0.614594   
## LotShapeReg 2.381e+03 1.684e+03 1.414 0.157594   
## LandContourHLS 9.272e+03 5.362e+03 1.729 0.084010 .   
## LandContourLow -6.186e+03 6.565e+03 -0.942 0.346171   
## LandContourLvl 3.868e+03 3.858e+03 1.003 0.316201   
## Utilities 1.321e+04 9.126e+03 1.448 0.147994   
## LotConfigCulDSac 1.122e+04 3.696e+03 3.036 0.002444 \*\*   
## LotConfigFR2 -5.781e+03 4.254e+03 -1.359 0.174443   
## LotConfigFR3 -2.090e+04 1.327e+04 -1.574 0.115642   
## LotConfigInside -1.397e+02 1.893e+03 -0.074 0.941198   
## LandSlope -1.030e+02 3.638e+03 -0.028 0.977422   
## NeighborhoodBlueste 1.691e+04 1.991e+04 0.849 0.395848   
## NeighborhoodBrDale 1.024e+04 1.090e+04 0.939 0.347750   
## NeighborhoodBrkSide 7.449e+03 9.581e+03 0.777 0.437038   
## NeighborhoodClearCr -7.181e+03 9.499e+03 -0.756 0.449812   
## NeighborhoodCollgCr -7.829e+03 7.493e+03 -1.045 0.296322   
## NeighborhoodCrawfor 1.798e+04 8.786e+03 2.046 0.040932 \*   
## NeighborhoodEdwards -9.799e+03 8.194e+03 -1.196 0.231955   
## NeighborhoodGilbert -8.135e+03 7.881e+03 -1.032 0.302154   
## NeighborhoodIDOTRR -1.611e+03 1.091e+04 -0.148 0.882692   
## NeighborhoodMeadowV -1.644e+03 1.160e+04 -0.142 0.887353   
## NeighborhoodMitchel -1.016e+04 8.361e+03 -1.216 0.224382   
## NeighborhoodNAmes -7.537e+03 8.053e+03 -0.936 0.349496   
## NeighborhoodNoRidge 2.431e+04 8.711e+03 2.790 0.005347 \*\*   
## NeighborhoodNPkVill 2.229e+04 1.474e+04 1.512 0.130729   
## NeighborhoodNridgHt 3.326e+04 7.665e+03 4.340 1.54e-05 \*\*\*  
## NeighborhoodNWAmes -1.099e+04 8.309e+03 -1.322 0.186308   
## NeighborhoodOldTown -4.126e+03 9.856e+03 -0.419 0.675579   
## NeighborhoodSawyer 3.664e+02 8.364e+03 0.044 0.965068   
## NeighborhoodSawyerW -1.009e+03 8.066e+03 -0.125 0.900511   
## NeighborhoodSomerst 6.938e+02 9.381e+03 0.074 0.941055   
## NeighborhoodStoneBr 4.233e+04 8.690e+03 4.871 1.25e-06 \*\*\*  
## NeighborhoodSWISU 1.962e+02 1.003e+04 0.020 0.984399   
## NeighborhoodTimber -6.135e+03 8.422e+03 -0.728 0.466519   
## NeighborhoodVeenker 7.469e+03 1.099e+04 0.680 0.496844   
## Condition1Feedr 5.955e+03 5.271e+03 1.130 0.258747   
## Condition1Norm 1.581e+04 4.381e+03 3.609 0.000320 \*\*\*  
## Condition1PosA 8.704e+03 1.050e+04 0.829 0.407159   
## Condition1PosN 1.152e+04 7.810e+03 1.476 0.140287   
## Condition1RRAe -1.213e+04 9.175e+03 -1.323 0.186228   
## Condition1RRAn 1.580e+04 7.299e+03 2.164 0.030645 \*   
## Condition1RRNe -3.699e+03 1.855e+04 -0.199 0.841951   
## Condition1RRNn 1.747e+04 1.346e+04 1.298 0.194574   
## Condition2Feedr -1.912e+04 2.422e+04 -0.789 0.430024   
## Condition2Norm -2.159e+04 2.095e+04 -1.031 0.302835   
## Condition2PosA 3.900e+04 3.455e+04 1.129 0.259097   
## Condition2PosN -2.435e+05 2.874e+04 -8.474 < 2e-16 \*\*\*  
## Condition2RRAe -6.085e+04 6.744e+04 -0.902 0.367053   
## Condition2RRAn -1.679e+04 3.247e+04 -0.517 0.605143   
## Condition2RRNn -1.048e+04 2.803e+04 -0.374 0.708442   
## BldgType 4.880e+03 1.782e+03 2.739 0.006256 \*\*   
## HouseStyle1.5Unf 9.861e+03 8.093e+03 1.218 0.223275   
## HouseStyle1Story 8.657e+03 4.207e+03 2.058 0.039794 \*   
## HouseStyle2.5Fin -2.028e+04 1.278e+04 -1.586 0.112880   
## HouseStyle2.5Unf -9.504e+03 9.559e+03 -0.994 0.320282   
## HouseStyle2Story -3.935e+03 3.451e+03 -1.140 0.254422   
## HouseStyleSFoyer 8.912e+02 6.080e+03 0.147 0.883479   
## HouseStyleSLvl 7.024e+03 5.241e+03 1.340 0.180413   
## OverallQual 8.241e+03 1.056e+03 7.807 1.23e-14 \*\*\*  
## OverallCond 5.274e+03 8.961e+02 5.885 5.11e-09 \*\*\*  
## YearBuilt 2.884e+02 8.244e+01 3.498 0.000485 \*\*\*  
## YearRemodAdd 1.963e+01 5.712e+01 0.344 0.731111   
## RoofStyleGable 7.055e+02 1.912e+04 0.037 0.970571   
## RoofStyleGambrel 2.510e+03 2.084e+04 0.120 0.904151   
## RoofStyleHip 4.317e+03 1.919e+04 0.225 0.822023   
## RoofStyleMansard 7.420e+03 2.233e+04 0.332 0.739762   
## RoofStyleShed 4.467e+04 3.519e+04 1.269 0.204588   
## RoofMatlCompShg 6.785e+05 3.522e+04 19.264 < 2e-16 \*\*\*  
## RoofMatlMembran 7.293e+05 4.804e+04 15.180 < 2e-16 \*\*\*  
## RoofMatlMetal 6.971e+05 4.820e+04 14.463 < 2e-16 \*\*\*  
## RoofMatlRoll 6.598e+05 4.383e+04 15.055 < 2e-16 \*\*\*  
## RoofMatlTar&Grv 6.733e+05 4.039e+04 16.672 < 2e-16 \*\*\*  
## RoofMatlWdShake 6.570e+05 3.875e+04 16.954 < 2e-16 \*\*\*  
## RoofMatlWdShngl 7.443e+05 3.679e+04 20.233 < 2e-16 \*\*\*  
## Exterior1stAsphShn -1.346e+04 3.468e+04 -0.388 0.698046   
## Exterior1stBrkComm -9.478e+03 2.871e+04 -0.330 0.741366   
## Exterior1stBrkFace 6.959e+03 1.331e+04 0.523 0.601119   
## Exterior1stCBlock 2.403e+03 2.682e+04 0.090 0.928617   
## Exterior1stCemntBd -3.561e+03 1.978e+04 -0.180 0.857158   
## Exterior1stHdBoard -9.955e+03 1.347e+04 -0.739 0.459948   
## Exterior1stImStucc -3.450e+04 2.945e+04 -1.171 0.241655   
## Exterior1stMetalSd -1.612e+02 1.514e+04 -0.011 0.991505   
## Exterior1stPlywood -1.456e+04 1.328e+04 -1.097 0.272914   
## Exterior1stStone -3.239e+03 2.499e+04 -0.130 0.896907   
## Exterior1stStucco 2.492e+03 1.463e+04 0.170 0.864786   
## Exterior1stVinylSd -1.167e+04 1.395e+04 -0.837 0.402758   
## Exterior1stWd Sdng -1.065e+04 1.290e+04 -0.826 0.408976   
## Exterior1stWdShing -3.038e+03 1.394e+04 -0.218 0.827529   
## Exterior2ndAsphShn 7.411e+03 2.321e+04 0.319 0.749582   
## Exterior2ndBrk Cmn 5.875e+03 2.101e+04 0.280 0.779819   
## Exterior2ndBrkFace 3.282e+03 1.367e+04 0.240 0.810389   
## Exterior2ndCBlock NA NA NA NA   
## Exterior2ndCmentBd 1.243e+04 1.947e+04 0.639 0.523196   
## Exterior2ndHdBoard 4.629e+03 1.282e+04 0.361 0.718146   
## Exterior2ndImStucc 2.610e+04 1.485e+04 1.758 0.079041 .   
## Exterior2ndMetalSd 2.376e+03 1.474e+04 0.161 0.871994   
## Exterior2ndOther -2.412e+04 2.869e+04 -0.841 0.400519   
## Exterior2ndPlywood 5.198e+03 1.245e+04 0.418 0.676349   
## Exterior2ndStone -4.761e+03 1.785e+04 -0.267 0.789759   
## Exterior2ndStucco -6.042e+03 1.416e+04 -0.427 0.669602   
## Exterior2ndVinylSd 1.024e+04 1.338e+04 0.765 0.444211   
## Exterior2ndWd Sdng 8.476e+03 1.234e+04 0.687 0.492477   
## Exterior2ndWd Shng -1.065e+03 1.300e+04 -0.082 0.934707   
## MasVnrTypeBrkFace 7.270e+03 6.988e+03 1.040 0.298347   
## MasVnrTypeNone 1.334e+04 7.038e+03 1.896 0.058156 .   
## MasVnrTypeStone 1.319e+04 7.418e+03 1.778 0.075573 .   
## MasVnrArea 3.115e+01 6.001e+00 5.191 2.44e-07 \*\*\*  
## ExterQual 4.020e+03 2.188e+03 1.837 0.066408 .   
## ExterCond -3.769e+03 2.216e+03 -1.701 0.089231 .   
## FoundationCBlock 3.596e+03 3.272e+03 1.099 0.271952   
## FoundationPConc 4.659e+03 3.558e+03 1.309 0.190710   
## FoundationSlab 1.529e+04 9.435e+03 1.620 0.105417   
## FoundationStone 8.531e+03 1.199e+04 0.712 0.476889   
## FoundationWood -2.974e+04 1.562e+04 -1.904 0.057137 .   
## BsmtQual 3.650e+03 1.785e+03 2.044 0.041113 \*   
## BsmtCond -5.723e+03 2.235e+03 -2.561 0.010568 \*   
## BsmtExposure 5.307e+03 8.819e+02 6.018 2.31e-09 \*\*\*  
## BsmtFinType1 -1.561e+02 5.308e+02 -0.294 0.768819   
## BsmtFinSF1 3.607e+01 5.018e+00 7.188 1.13e-12 \*\*\*  
## BsmtFinType2 2.517e+02 1.361e+03 0.185 0.853387   
## BsmtFinSF2 2.216e+01 8.527e+00 2.599 0.009468 \*\*   
## BsmtUnfSF 1.588e+01 4.533e+00 3.503 0.000477 \*\*\*  
## TotalBsmtSF NA NA NA NA   
## HeatingGasA -6.757e+03 2.655e+04 -0.254 0.799157   
## HeatingGasW -7.918e+03 2.733e+04 -0.290 0.772111   
## HeatingGrav -3.215e+03 2.870e+04 -0.112 0.910811   
## HeatingOthW -4.347e+04 3.251e+04 -1.337 0.181401   
## HeatingWall 4.322e+03 3.066e+04 0.141 0.887933   
## HeatingQC 1.165e+03 1.009e+03 1.155 0.248247   
## CentralAirY -8.057e+02 4.021e+03 -0.200 0.841245   
## ElectricalFuseF -2.154e+03 5.961e+03 -0.361 0.717850   
## ElectricalFuseP 3.051e+03 1.924e+04 0.159 0.874029   
## ElectricalMix 1.663e+04 3.464e+04 0.480 0.631305   
## ElectricalSBrkr -3.670e+03 3.082e+03 -1.191 0.234028   
## X1stFlrSF 4.716e+01 5.598e+00 8.423 < 2e-16 \*\*\*  
## X2ndFlrSF 6.206e+01 5.838e+00 10.630 < 2e-16 \*\*\*  
## LowQualFinSF 8.323e+00 1.960e+01 0.425 0.671139   
## GrLivArea NA NA NA NA   
## BsmtFullBath 1.854e+03 2.042e+03 0.908 0.363995   
## BsmtHalfBath 4.201e+02 3.153e+03 0.133 0.894016   
## FullBath 3.356e+03 2.292e+03 1.465 0.143283   
## HalfBath 3.112e+03 2.186e+03 1.423 0.154856   
## BedroomAbvGr -5.846e+03 1.392e+03 -4.198 2.88e-05 \*\*\*  
## KitchenAbvGr -1.063e+04 4.707e+03 -2.259 0.024070 \*   
## KitchenQual 6.656e+03 1.717e+03 3.875 0.000112 \*\*\*  
## TotRmsAbvGrd 2.978e+03 9.880e+02 3.014 0.002630 \*\*   
## Functional 6.007e+03 1.168e+03 5.143 3.13e-07 \*\*\*  
## Fireplaces 5.031e+03 2.320e+03 2.169 0.030293 \*   
## FireplaceQu -1.311e+03 8.393e+02 -1.562 0.118532   
## GarageTypeAttchd 1.381e+04 1.153e+04 1.198 0.231134   
## GarageTypeBasment 2.157e+04 1.337e+04 1.614 0.106748   
## GarageTypeBuiltIn 1.454e+04 1.205e+04 1.206 0.227910   
## GarageTypeCarPort 1.445e+04 1.492e+04 0.969 0.332969   
## GarageTypeDetchd 1.744e+04 1.152e+04 1.515 0.130052   
## GarageTypeNoGarage 2.478e+04 2.192e+04 1.131 0.258380   
## GarageYrBlt -2.453e+01 6.260e+01 -0.392 0.695298   
## GarageFinish1 -1.218e+03 2.525e+03 -0.483 0.629483   
## GarageFinish2 -5.760e+03 2.051e+03 -2.809 0.005048 \*\*   
## GarageFinish3 NA NA NA NA   
## GarageCars 4.355e+03 2.371e+03 1.837 0.066507 .   
## GarageArea 1.798e+01 8.259e+00 2.177 0.029658 \*   
## GarageQual1 -1.197e+05 4.047e+04 -2.959 0.003146 \*\*   
## GarageQual2 -1.083e+05 3.160e+04 -3.427 0.000630 \*\*\*  
## GarageQual3 -1.054e+05 3.125e+04 -3.375 0.000761 \*\*\*  
## GarageQual4 -1.027e+05 3.231e+04 -3.178 0.001517 \*\*   
## GarageQual5 NA NA NA NA   
## GarageCond1 1.030e+05 3.900e+04 2.642 0.008337 \*\*   
## GarageCond2 1.024e+05 3.634e+04 2.819 0.004891 \*\*   
## GarageCond3 1.035e+05 3.607e+04 2.870 0.004169 \*\*   
## GarageCond4 1.042e+05 3.775e+04 2.761 0.005843 \*\*   
## GarageCond5 NA NA NA NA   
## PavedDrive -1.167e+02 1.747e+03 -0.067 0.946747   
## WoodDeckSF 1.384e+01 6.127e+00 2.259 0.024028 \*   
## OpenPorchSF 9.489e+00 1.194e+01 0.794 0.427064   
## EnclosedPorch -1.703e+00 1.302e+01 -0.131 0.895924   
## X3SsnPorch 4.412e+01 2.367e+01 1.864 0.062598 .   
## ScreenPorch 3.383e+01 1.318e+01 2.566 0.010389 \*   
## PoolArea -1.792e+01 5.637e+01 -0.318 0.750553   
## PoolQC 2.017e+04 8.507e+03 2.371 0.017889 \*   
## FenceGdWo 1.087e+04 5.131e+03 2.118 0.034402 \*   
## FenceMnPrv 1.190e+04 4.196e+03 2.837 0.004629 \*\*   
## FenceMnWw 7.750e+03 8.649e+03 0.896 0.370355   
## FenceNoFence 1.043e+04 3.842e+03 2.716 0.006703 \*\*   
## MiscFeatureNoMiscFeature 5.914e+04 9.986e+04 0.592 0.553814   
## MiscFeatureOthr 7.170e+04 9.233e+04 0.777 0.437539   
## MiscFeatureShed 5.899e+04 9.566e+04 0.617 0.537533   
## MiscFeatureTenC -1.227e+04 9.365e+04 -0.131 0.895816   
## MiscVal 3.883e+00 6.257e+00 0.621 0.535002   
## MoSold -6.748e+02 2.552e+02 -2.644 0.008297 \*\*   
## YrSold -2.496e+01 5.346e+02 -0.047 0.962767   
## SaleTypeCon 2.357e+04 1.861e+04 1.266 0.205691   
## SaleTypeConLD 1.765e+04 1.010e+04 1.748 0.080738 .   
## SaleTypeConLI 7.716e+03 1.217e+04 0.634 0.526320   
## SaleTypeConLw 7.008e+02 1.257e+04 0.056 0.955554   
## SaleTypeCWD 1.649e+04 1.349e+04 1.222 0.221906   
## SaleTypeNew 3.565e+04 1.605e+04 2.221 0.026505 \*   
## SaleTypeOth 9.891e+03 1.537e+04 0.643 0.520062   
## SaleTypeWD 6.050e+02 4.386e+03 0.138 0.890304   
## SaleConditionAdjLand 1.483e+04 1.505e+04 0.985 0.324575   
## SaleConditionAlloca 7.013e+03 8.970e+03 0.782 0.434495   
## SaleConditionFamily -7.477e+02 6.429e+03 -0.116 0.907422   
## SaleConditionNormal 7.213e+03 3.021e+03 2.388 0.017111 \*   
## SaleConditionPartial -9.543e+03 1.544e+04 -0.618 0.536660   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 24160 on 1251 degrees of freedom  
## Multiple R-squared: 0.9207, Adjusted R-squared: 0.9075   
## F-statistic: 69.83 on 208 and 1251 DF, p-value: < 2.2e-16

plot(lmfit)

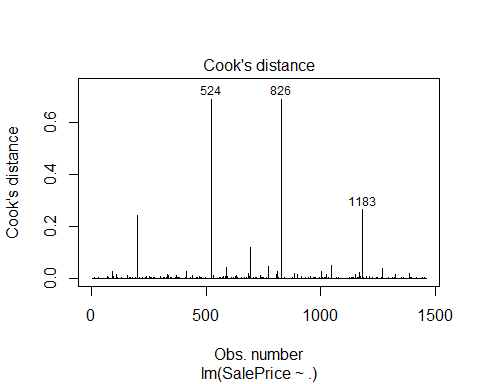
## Warning: not plotting observations with leverage one:  
## 121, 186, 272, 347, 399, 584, 596, 945, 1004, 1012, 1188, 1231, 1271, 1276, 1299, 1322, 1371, 1387



## Warning: not plotting observations with leverage one:  
## 121, 186, 272, 347, 399, 584, 596, 945, 1004, 1012, 1188, 1231, 1271, 1276, 1299, 1322, 1371, 1387



plot(lmfit,which=c(4))



#anova(lmfit)

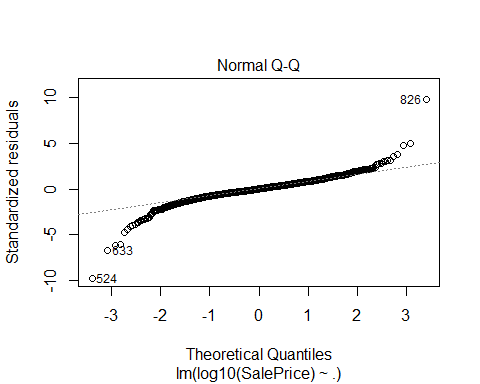
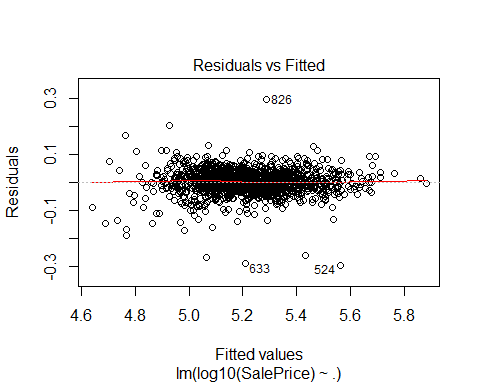
### Another linear model fit - using Log10 of Saleprice because Sale is skewed.

lmfit\_log<-lm(log10(SalePrice)~.,data=Train)  
summary(lmfit)

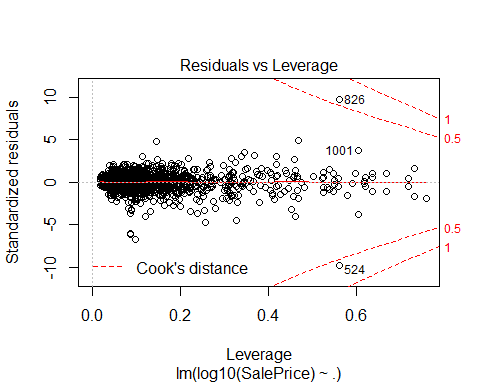
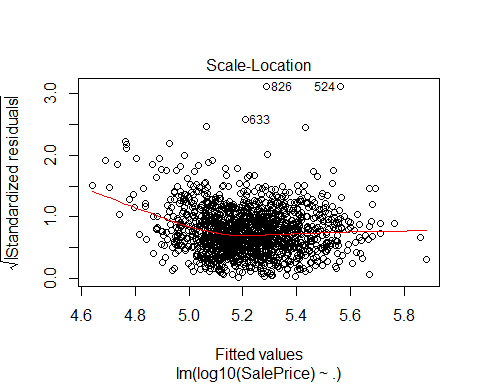
##   
## Call:  
## lm(formula = SalePrice ~ ., data = Train)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -169991 -11292 0 11008 169991   
##   
## Coefficients: (6 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.502e+06 1.095e+06 -1.372 0.170227   
## MSSubClass -7.322e+01 4.445e+01 -1.647 0.099775 .   
## MSZoningFV 2.819e+04 1.254e+04 2.248 0.024760 \*   
## MSZoningRH 2.569e+04 1.253e+04 2.050 0.040560 \*   
## MSZoningRL 2.485e+04 1.063e+04 2.338 0.019530 \*   
## MSZoningRM 2.197e+04 1.002e+04 2.194 0.028435 \*   
## LotFrontage 9.964e+01 4.405e+01 2.262 0.023853 \*   
## LotArea 4.192e-01 9.726e-02 4.310 1.76e-05 \*\*\*  
## StreetPave 2.146e+04 1.209e+04 1.775 0.076120 .   
## AlleyNoAlley -2.126e+03 4.412e+03 -0.482 0.630019   
## AlleyPave -8.441e+01 6.302e+03 -0.013 0.989316   
## LotShapeIR2 8.368e+03 4.423e+03 1.892 0.058760 .   
## LotShapeIR3 4.688e+03 9.309e+03 0.504 0.614594   
## LotShapeReg 2.381e+03 1.684e+03 1.414 0.157594   
## LandContourHLS 9.272e+03 5.362e+03 1.729 0.084010 .   
## LandContourLow -6.186e+03 6.565e+03 -0.942 0.346171   
## LandContourLvl 3.868e+03 3.858e+03 1.003 0.316201   
## Utilities 1.321e+04 9.126e+03 1.448 0.147994   
## LotConfigCulDSac 1.122e+04 3.696e+03 3.036 0.002444 \*\*   
## LotConfigFR2 -5.781e+03 4.254e+03 -1.359 0.174443   
## LotConfigFR3 -2.090e+04 1.327e+04 -1.574 0.115642   
## LotConfigInside -1.397e+02 1.893e+03 -0.074 0.941198   
## LandSlope -1.030e+02 3.638e+03 -0.028 0.977422   
## NeighborhoodBlueste 1.691e+04 1.991e+04 0.849 0.395848   
## NeighborhoodBrDale 1.024e+04 1.090e+04 0.939 0.347750   
## NeighborhoodBrkSide 7.449e+03 9.581e+03 0.777 0.437038   
## NeighborhoodClearCr -7.181e+03 9.499e+03 -0.756 0.449812   
## NeighborhoodCollgCr -7.829e+03 7.493e+03 -1.045 0.296322   
## NeighborhoodCrawfor 1.798e+04 8.786e+03 2.046 0.040932 \*   
## NeighborhoodEdwards -9.799e+03 8.194e+03 -1.196 0.231955   
## NeighborhoodGilbert -8.135e+03 7.881e+03 -1.032 0.302154   
## NeighborhoodIDOTRR -1.611e+03 1.091e+04 -0.148 0.882692   
## NeighborhoodMeadowV -1.644e+03 1.160e+04 -0.142 0.887353   
## NeighborhoodMitchel -1.016e+04 8.361e+03 -1.216 0.224382   
## NeighborhoodNAmes -7.537e+03 8.053e+03 -0.936 0.349496   
## NeighborhoodNoRidge 2.431e+04 8.711e+03 2.790 0.005347 \*\*   
## NeighborhoodNPkVill 2.229e+04 1.474e+04 1.512 0.130729   
## NeighborhoodNridgHt 3.326e+04 7.665e+03 4.340 1.54e-05 \*\*\*  
## NeighborhoodNWAmes -1.099e+04 8.309e+03 -1.322 0.186308   
## NeighborhoodOldTown -4.126e+03 9.856e+03 -0.419 0.675579   
## NeighborhoodSawyer 3.664e+02 8.364e+03 0.044 0.965068   
## NeighborhoodSawyerW -1.009e+03 8.066e+03 -0.125 0.900511   
## NeighborhoodSomerst 6.938e+02 9.381e+03 0.074 0.941055   
## NeighborhoodStoneBr 4.233e+04 8.690e+03 4.871 1.25e-06 \*\*\*  
## NeighborhoodSWISU 1.962e+02 1.003e+04 0.020 0.984399   
## NeighborhoodTimber -6.135e+03 8.422e+03 -0.728 0.466519   
## NeighborhoodVeenker 7.469e+03 1.099e+04 0.680 0.496844   
## Condition1Feedr 5.955e+03 5.271e+03 1.130 0.258747   
## Condition1Norm 1.581e+04 4.381e+03 3.609 0.000320 \*\*\*  
## Condition1PosA 8.704e+03 1.050e+04 0.829 0.407159   
## Condition1PosN 1.152e+04 7.810e+03 1.476 0.140287   
## Condition1RRAe -1.213e+04 9.175e+03 -1.323 0.186228   
## Condition1RRAn 1.580e+04 7.299e+03 2.164 0.030645 \*   
## Condition1RRNe -3.699e+03 1.855e+04 -0.199 0.841951   
## Condition1RRNn 1.747e+04 1.346e+04 1.298 0.194574   
## Condition2Feedr -1.912e+04 2.422e+04 -0.789 0.430024   
## Condition2Norm -2.159e+04 2.095e+04 -1.031 0.302835   
## Condition2PosA 3.900e+04 3.455e+04 1.129 0.259097   
## Condition2PosN -2.435e+05 2.874e+04 -8.474 < 2e-16 \*\*\*  
## Condition2RRAe -6.085e+04 6.744e+04 -0.902 0.367053   
## Condition2RRAn -1.679e+04 3.247e+04 -0.517 0.605143   
## Condition2RRNn -1.048e+04 2.803e+04 -0.374 0.708442   
## BldgType 4.880e+03 1.782e+03 2.739 0.006256 \*\*   
## HouseStyle1.5Unf 9.861e+03 8.093e+03 1.218 0.223275   
## HouseStyle1Story 8.657e+03 4.207e+03 2.058 0.039794 \*   
## HouseStyle2.5Fin -2.028e+04 1.278e+04 -1.586 0.112880   
## HouseStyle2.5Unf -9.504e+03 9.559e+03 -0.994 0.320282   
## HouseStyle2Story -3.935e+03 3.451e+03 -1.140 0.254422   
## HouseStyleSFoyer 8.912e+02 6.080e+03 0.147 0.883479   
## HouseStyleSLvl 7.024e+03 5.241e+03 1.340 0.180413   
## OverallQual 8.241e+03 1.056e+03 7.807 1.23e-14 \*\*\*  
## OverallCond 5.274e+03 8.961e+02 5.885 5.11e-09 \*\*\*  
## YearBuilt 2.884e+02 8.244e+01 3.498 0.000485 \*\*\*  
## YearRemodAdd 1.963e+01 5.712e+01 0.344 0.731111   
## RoofStyleGable 7.055e+02 1.912e+04 0.037 0.970571   
## RoofStyleGambrel 2.510e+03 2.084e+04 0.120 0.904151   
## RoofStyleHip 4.317e+03 1.919e+04 0.225 0.822023   
## RoofStyleMansard 7.420e+03 2.233e+04 0.332 0.739762   
## RoofStyleShed 4.467e+04 3.519e+04 1.269 0.204588   
## RoofMatlCompShg 6.785e+05 3.522e+04 19.264 < 2e-16 \*\*\*  
## RoofMatlMembran 7.293e+05 4.804e+04 15.180 < 2e-16 \*\*\*  
## RoofMatlMetal 6.971e+05 4.820e+04 14.463 < 2e-16 \*\*\*  
## RoofMatlRoll 6.598e+05 4.383e+04 15.055 < 2e-16 \*\*\*  
## RoofMatlTar&Grv 6.733e+05 4.039e+04 16.672 < 2e-16 \*\*\*  
## RoofMatlWdShake 6.570e+05 3.875e+04 16.954 < 2e-16 \*\*\*  
## RoofMatlWdShngl 7.443e+05 3.679e+04 20.233 < 2e-16 \*\*\*  
## Exterior1stAsphShn -1.346e+04 3.468e+04 -0.388 0.698046   
## Exterior1stBrkComm -9.478e+03 2.871e+04 -0.330 0.741366   
## Exterior1stBrkFace 6.959e+03 1.331e+04 0.523 0.601119   
## Exterior1stCBlock 2.403e+03 2.682e+04 0.090 0.928617   
## Exterior1stCemntBd -3.561e+03 1.978e+04 -0.180 0.857158   
## Exterior1stHdBoard -9.955e+03 1.347e+04 -0.739 0.459948   
## Exterior1stImStucc -3.450e+04 2.945e+04 -1.171 0.241655   
## Exterior1stMetalSd -1.612e+02 1.514e+04 -0.011 0.991505   
## Exterior1stPlywood -1.456e+04 1.328e+04 -1.097 0.272914   
## Exterior1stStone -3.239e+03 2.499e+04 -0.130 0.896907   
## Exterior1stStucco 2.492e+03 1.463e+04 0.170 0.864786   
## Exterior1stVinylSd -1.167e+04 1.395e+04 -0.837 0.402758   
## Exterior1stWd Sdng -1.065e+04 1.290e+04 -0.826 0.408976   
## Exterior1stWdShing -3.038e+03 1.394e+04 -0.218 0.827529   
## Exterior2ndAsphShn 7.411e+03 2.321e+04 0.319 0.749582   
## Exterior2ndBrk Cmn 5.875e+03 2.101e+04 0.280 0.779819   
## Exterior2ndBrkFace 3.282e+03 1.367e+04 0.240 0.810389   
## Exterior2ndCBlock NA NA NA NA   
## Exterior2ndCmentBd 1.243e+04 1.947e+04 0.639 0.523196   
## Exterior2ndHdBoard 4.629e+03 1.282e+04 0.361 0.718146   
## Exterior2ndImStucc 2.610e+04 1.485e+04 1.758 0.079041 .   
## Exterior2ndMetalSd 2.376e+03 1.474e+04 0.161 0.871994   
## Exterior2ndOther -2.412e+04 2.869e+04 -0.841 0.400519   
## Exterior2ndPlywood 5.198e+03 1.245e+04 0.418 0.676349   
## Exterior2ndStone -4.761e+03 1.785e+04 -0.267 0.789759   
## Exterior2ndStucco -6.042e+03 1.416e+04 -0.427 0.669602   
## Exterior2ndVinylSd 1.024e+04 1.338e+04 0.765 0.444211   
## Exterior2ndWd Sdng 8.476e+03 1.234e+04 0.687 0.492477   
## Exterior2ndWd Shng -1.065e+03 1.300e+04 -0.082 0.934707   
## MasVnrTypeBrkFace 7.270e+03 6.988e+03 1.040 0.298347   
## MasVnrTypeNone 1.334e+04 7.038e+03 1.896 0.058156 .   
## MasVnrTypeStone 1.319e+04 7.418e+03 1.778 0.075573 .   
## MasVnrArea 3.115e+01 6.001e+00 5.191 2.44e-07 \*\*\*  
## ExterQual 4.020e+03 2.188e+03 1.837 0.066408 .   
## ExterCond -3.769e+03 2.216e+03 -1.701 0.089231 .   
## FoundationCBlock 3.596e+03 3.272e+03 1.099 0.271952   
## FoundationPConc 4.659e+03 3.558e+03 1.309 0.190710   
## FoundationSlab 1.529e+04 9.435e+03 1.620 0.105417   
## FoundationStone 8.531e+03 1.199e+04 0.712 0.476889   
## FoundationWood -2.974e+04 1.562e+04 -1.904 0.057137 .   
## BsmtQual 3.650e+03 1.785e+03 2.044 0.041113 \*   
## BsmtCond -5.723e+03 2.235e+03 -2.561 0.010568 \*   
## BsmtExposure 5.307e+03 8.819e+02 6.018 2.31e-09 \*\*\*  
## BsmtFinType1 -1.561e+02 5.308e+02 -0.294 0.768819   
## BsmtFinSF1 3.607e+01 5.018e+00 7.188 1.13e-12 \*\*\*  
## BsmtFinType2 2.517e+02 1.361e+03 0.185 0.853387   
## BsmtFinSF2 2.216e+01 8.527e+00 2.599 0.009468 \*\*   
## BsmtUnfSF 1.588e+01 4.533e+00 3.503 0.000477 \*\*\*  
## TotalBsmtSF NA NA NA NA   
## HeatingGasA -6.757e+03 2.655e+04 -0.254 0.799157   
## HeatingGasW -7.918e+03 2.733e+04 -0.290 0.772111   
## HeatingGrav -3.215e+03 2.870e+04 -0.112 0.910811   
## HeatingOthW -4.347e+04 3.251e+04 -1.337 0.181401   
## HeatingWall 4.322e+03 3.066e+04 0.141 0.887933   
## HeatingQC 1.165e+03 1.009e+03 1.155 0.248247   
## CentralAirY -8.057e+02 4.021e+03 -0.200 0.841245   
## ElectricalFuseF -2.154e+03 5.961e+03 -0.361 0.717850   
## ElectricalFuseP 3.051e+03 1.924e+04 0.159 0.874029   
## ElectricalMix 1.663e+04 3.464e+04 0.480 0.631305   
## ElectricalSBrkr -3.670e+03 3.082e+03 -1.191 0.234028   
## X1stFlrSF 4.716e+01 5.598e+00 8.423 < 2e-16 \*\*\*  
## X2ndFlrSF 6.206e+01 5.838e+00 10.630 < 2e-16 \*\*\*  
## LowQualFinSF 8.323e+00 1.960e+01 0.425 0.671139   
## GrLivArea NA NA NA NA   
## BsmtFullBath 1.854e+03 2.042e+03 0.908 0.363995   
## BsmtHalfBath 4.201e+02 3.153e+03 0.133 0.894016   
## FullBath 3.356e+03 2.292e+03 1.465 0.143283   
## HalfBath 3.112e+03 2.186e+03 1.423 0.154856   
## BedroomAbvGr -5.846e+03 1.392e+03 -4.198 2.88e-05 \*\*\*  
## KitchenAbvGr -1.063e+04 4.707e+03 -2.259 0.024070 \*   
## KitchenQual 6.656e+03 1.717e+03 3.875 0.000112 \*\*\*  
## TotRmsAbvGrd 2.978e+03 9.880e+02 3.014 0.002630 \*\*   
## Functional 6.007e+03 1.168e+03 5.143 3.13e-07 \*\*\*  
## Fireplaces 5.031e+03 2.320e+03 2.169 0.030293 \*   
## FireplaceQu -1.311e+03 8.393e+02 -1.562 0.118532   
## GarageTypeAttchd 1.381e+04 1.153e+04 1.198 0.231134   
## GarageTypeBasment 2.157e+04 1.337e+04 1.614 0.106748   
## GarageTypeBuiltIn 1.454e+04 1.205e+04 1.206 0.227910   
## GarageTypeCarPort 1.445e+04 1.492e+04 0.969 0.332969   
## GarageTypeDetchd 1.744e+04 1.152e+04 1.515 0.130052   
## GarageTypeNoGarage 2.478e+04 2.192e+04 1.131 0.258380   
## GarageYrBlt -2.453e+01 6.260e+01 -0.392 0.695298   
## GarageFinish1 -1.218e+03 2.525e+03 -0.483 0.629483   
## GarageFinish2 -5.760e+03 2.051e+03 -2.809 0.005048 \*\*   
## GarageFinish3 NA NA NA NA   
## GarageCars 4.355e+03 2.371e+03 1.837 0.066507 .   
## GarageArea 1.798e+01 8.259e+00 2.177 0.029658 \*   
## GarageQual1 -1.197e+05 4.047e+04 -2.959 0.003146 \*\*   
## GarageQual2 -1.083e+05 3.160e+04 -3.427 0.000630 \*\*\*  
## GarageQual3 -1.054e+05 3.125e+04 -3.375 0.000761 \*\*\*  
## GarageQual4 -1.027e+05 3.231e+04 -3.178 0.001517 \*\*   
## GarageQual5 NA NA NA NA   
## GarageCond1 1.030e+05 3.900e+04 2.642 0.008337 \*\*   
## GarageCond2 1.024e+05 3.634e+04 2.819 0.004891 \*\*   
## GarageCond3 1.035e+05 3.607e+04 2.870 0.004169 \*\*   
## GarageCond4 1.042e+05 3.775e+04 2.761 0.005843 \*\*   
## GarageCond5 NA NA NA NA   
## PavedDrive -1.167e+02 1.747e+03 -0.067 0.946747   
## WoodDeckSF 1.384e+01 6.127e+00 2.259 0.024028 \*   
## OpenPorchSF 9.489e+00 1.194e+01 0.794 0.427064   
## EnclosedPorch -1.703e+00 1.302e+01 -0.131 0.895924   
## X3SsnPorch 4.412e+01 2.367e+01 1.864 0.062598 .   
## ScreenPorch 3.383e+01 1.318e+01 2.566 0.010389 \*   
## PoolArea -1.792e+01 5.637e+01 -0.318 0.750553   
## PoolQC 2.017e+04 8.507e+03 2.371 0.017889 \*   
## FenceGdWo 1.087e+04 5.131e+03 2.118 0.034402 \*   
## FenceMnPrv 1.190e+04 4.196e+03 2.837 0.004629 \*\*   
## FenceMnWw 7.750e+03 8.649e+03 0.896 0.370355   
## FenceNoFence 1.043e+04 3.842e+03 2.716 0.006703 \*\*   
## MiscFeatureNoMiscFeature 5.914e+04 9.986e+04 0.592 0.553814   
## MiscFeatureOthr 7.170e+04 9.233e+04 0.777 0.437539   
## MiscFeatureShed 5.899e+04 9.566e+04 0.617 0.537533   
## MiscFeatureTenC -1.227e+04 9.365e+04 -0.131 0.895816   
## MiscVal 3.883e+00 6.257e+00 0.621 0.535002   
## MoSold -6.748e+02 2.552e+02 -2.644 0.008297 \*\*   
## YrSold -2.496e+01 5.346e+02 -0.047 0.962767   
## SaleTypeCon 2.357e+04 1.861e+04 1.266 0.205691   
## SaleTypeConLD 1.765e+04 1.010e+04 1.748 0.080738 .   
## SaleTypeConLI 7.716e+03 1.217e+04 0.634 0.526320   
## SaleTypeConLw 7.008e+02 1.257e+04 0.056 0.955554   
## SaleTypeCWD 1.649e+04 1.349e+04 1.222 0.221906   
## SaleTypeNew 3.565e+04 1.605e+04 2.221 0.026505 \*   
## SaleTypeOth 9.891e+03 1.537e+04 0.643 0.520062   
## SaleTypeWD 6.050e+02 4.386e+03 0.138 0.890304   
## SaleConditionAdjLand 1.483e+04 1.505e+04 0.985 0.324575   
## SaleConditionAlloca 7.013e+03 8.970e+03 0.782 0.434495   
## SaleConditionFamily -7.477e+02 6.429e+03 -0.116 0.907422   
## SaleConditionNormal 7.213e+03 3.021e+03 2.388 0.017111 \*   
## SaleConditionPartial -9.543e+03 1.544e+04 -0.618 0.536660   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 24160 on 1251 degrees of freedom  
## Multiple R-squared: 0.9207, Adjusted R-squared: 0.9075   
## F-statistic: 69.83 on 208 and 1251 DF, p-value: < 2.2e-16

plot(lmfit\_log)

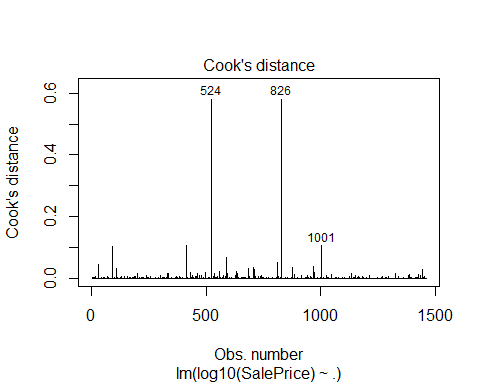
## Warning: not plotting observations with leverage one:  
## 121, 186, 272, 347, 399, 584, 596, 945, 1004, 1012, 1188, 1231, 1271, 1276, 1299, 1322, 1371, 1387



## Warning: not plotting observations with leverage one:  
## 121, 186, 272, 347, 399, 584, 596, 945, 1004, 1012, 1188, 1231, 1271, 1276, 1299, 1322, 1371, 1387



plot(lmfit\_log,which=c(4))



524 is a really big house, with great features but has extremely low sale price. 826 is the opposite. It was sold way over than others when looking at its features.

dishould probably try out removing the 524, 826, and possibly 1001, 1183 to see if they make a significant difference in model.

#### This outlier removal option is a possible option to enhance the performance in the future.

### 1. Forward Selection Results

#full - assigned as lmfit earlier  
null <- lm(SalePrice~1,data=Train)  
stepF <- stepAIC(null, scope=list(lower=null, upper=lmfit), direction= "forward", trace=FALSE) #trace=TRUE  
StepF\_log<- stepAIC(null, scope=list(lower=null, upper=lmfit\_log), direction= "forward", trace=FALSE) #trace=TRUE

#### 2. Linear Model looks like below.

stepF

##   
## Call:  
## lm(formula = SalePrice ~ OverallQual + GrLivArea + Neighborhood +   
## BsmtFinSF1 + RoofMatl + MSSubClass + BsmtExposure + KitchenQual +   
## Condition2 + SaleCondition + LotArea + YearBuilt + OverallCond +   
## MasVnrArea + PoolQC + BedroomAbvGr + GarageCars + GarageFinish +   
## TotalBsmtSF + BldgType + MasVnrType + Functional + ExterQual +   
## BsmtCond + Condition1 + Exterior1st + LandContour + MoSold +   
## ScreenPorch + LowQualFinSF + LotConfig + LotFrontage + WoodDeckSF +   
## Street + GarageArea + KitchenAbvGr + TotRmsAbvGrd + LotShape +   
## ExterCond + PoolArea + BsmtFinSF2 + Fireplaces + FireplaceQu +   
## BsmtQual + X3SsnPorch + Foundation + Utilities, data = Train)  
##   
## Coefficients:  
## (Intercept) OverallQual GrLivArea   
## -1.578e+06 7.954e+03 5.495e+01   
## NeighborhoodBlueste NeighborhoodBrDale NeighborhoodBrkSide   
## 1.097e+03 6.360e+03 7.498e+03   
## NeighborhoodClearCr NeighborhoodCollgCr NeighborhoodCrawfor   
## -5.999e+03 -7.722e+03 1.676e+04   
## NeighborhoodEdwards NeighborhoodGilbert NeighborhoodIDOTRR   
## -7.360e+03 -7.343e+03 -4.989e+03   
## NeighborhoodMeadowV NeighborhoodMitchel NeighborhoodNAmes   
## -7.977e+03 -1.037e+04 -7.045e+03   
## NeighborhoodNoRidge NeighborhoodNPkVill NeighborhoodNridgHt   
## 2.587e+04 2.188e+04 3.440e+04   
## NeighborhoodNWAmes NeighborhoodOldTown NeighborhoodSawyer   
## -1.091e+04 -5.331e+03 -8.190e+00   
## NeighborhoodSawyerW NeighborhoodSomerst NeighborhoodStoneBr   
## -2.301e+03 5.657e+03 4.191e+04   
## NeighborhoodSWISU NeighborhoodTimber NeighborhoodVeenker   
## 4.607e+02 -5.165e+03 7.827e+03   
## BsmtFinSF1 RoofMatlCompShg RoofMatlMembran   
## 2.038e+01 7.020e+05 7.524e+05   
## RoofMatlMetal RoofMatlRoll RoofMatlTar&Grv   
## 7.217e+05 6.924e+05 6.919e+05   
## RoofMatlWdShake RoofMatlWdShngl MSSubClass   
## 6.970e+05 7.801e+05 -8.719e+01   
## BsmtExposure KitchenQual Condition2Feedr   
## 5.357e+03 6.983e+03 -2.290e+04   
## Condition2Norm Condition2PosA Condition2PosN   
## -2.178e+04 4.333e+04 -2.406e+05   
## Condition2RRAe Condition2RRAn Condition2RRNn   
## -3.345e+04 -1.795e+04 -1.047e+04   
## SaleConditionAdjLand SaleConditionAlloca SaleConditionFamily   
## 1.836e+04 5.322e+03 -8.197e+02   
## SaleConditionNormal SaleConditionPartial LotArea   
## 7.499e+03 2.469e+04 3.953e-01   
## YearBuilt OverallCond MasVnrArea   
## 3.455e+02 5.915e+03 3.304e+01   
## PoolQC BedroomAbvGr GarageCars   
## 2.845e+04 -5.772e+03 3.988e+03   
## GarageFinish1 GarageFinish2 GarageFinish3   
## -1.396e+04 -1.926e+04 -1.270e+04   
## TotalBsmtSF BldgType MasVnrTypeBrkFace   
## 1.353e+01 4.524e+03 9.254e+03   
## MasVnrTypeNone MasVnrTypeStone Functional   
## 1.576e+04 1.415e+04 5.614e+03   
## ExterQual BsmtCond Condition1Feedr   
## 5.338e+03 -6.000e+03 4.105e+03   
## Condition1Norm Condition1PosA Condition1PosN   
## 1.358e+04 3.078e+03 7.634e+03   
## Condition1RRAe Condition1RRAn Condition1RRNe   
## -1.238e+04 1.187e+04 -6.803e+03   
## Condition1RRNn Exterior1stAsphShn Exterior1stBrkComm   
## 1.755e+04 -8.073e+03 -1.252e+03   
## Exterior1stBrkFace Exterior1stCBlock Exterior1stCemntBd   
## 1.669e+04 6.685e+03 1.178e+04   
## Exterior1stHdBoard Exterior1stImStucc Exterior1stMetalSd   
## -4.010e+02 -1.658e+04 6.574e+03   
## Exterior1stPlywood Exterior1stStone Exterior1stStucco   
## -5.211e+03 -1.466e+04 6.047e+03   
## Exterior1stVinylSd Exterior1stWd Sdng Exterior1stWdShing   
## 2.308e+03 2.340e+03 1.780e+03   
## LandContourHLS LandContourLow LandContourLvl   
## 1.003e+04 -8.129e+03 4.069e+03   
## MoSold ScreenPorch LowQualFinSF   
## -6.305e+02 3.782e+01 -4.923e+01   
## LotConfigCulDSac LotConfigFR2 LotConfigFR3   
## 1.144e+04 -5.342e+03 -1.985e+04   
## LotConfigInside LotFrontage WoodDeckSF   
## 4.230e+02 1.093e+02 1.280e+01   
## StreetPave GarageArea KitchenAbvGr   
## 2.461e+04 1.825e+01 -1.010e+04   
## TotRmsAbvGrd LotShapeIR2 LotShapeIR3   
## 2.626e+03 9.554e+03 8.335e+03   
## LotShapeReg ExterCond PoolArea   
## 2.338e+03 -3.866e+03 -8.462e+01   
## BsmtFinSF2 Fireplaces FireplaceQu   
## 8.058e+00 6.021e+03 -1.733e+03   
## BsmtQual X3SsnPorch FoundationCBlock   
## 3.428e+03 4.493e+01 2.434e+03   
## FoundationPConc FoundationSlab FoundationStone   
## 4.109e+03 1.409e+04 4.845e+03   
## FoundationWood Utilities   
## -3.305e+04 1.153e+04

### Backward Elimination Results

StepB <- stepAIC(lmfit, direction= "backward", trace=FALSE)  
summary(StepB)

##   
## Call:  
## lm(formula = SalePrice ~ LotFrontage + LotArea + Street + LandContour +   
## LotConfig + Neighborhood + Condition1 + Condition2 + BldgType +   
## HouseStyle + OverallQual + OverallCond + YearBuilt + RoofMatl +   
## Exterior1st + MasVnrType + MasVnrArea + ExterQual + ExterCond +   
## Foundation + BsmtQual + BsmtCond + BsmtExposure + BsmtFinSF1 +   
## BsmtFinSF2 + BsmtUnfSF + X1stFlrSF + X2ndFlrSF + BedroomAbvGr +   
## KitchenAbvGr + KitchenQual + TotRmsAbvGrd + Functional +   
## Fireplaces + FireplaceQu + GarageFinish + GarageCars + GarageArea +   
## GarageQual + GarageCond + WoodDeckSF + X3SsnPorch + ScreenPorch +   
## PoolQC + Fence + MiscFeature + MoSold + SaleCondition, data = Train)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -172977 -11278 0 11014 172977   
##   
## Coefficients: (2 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.569e+06 1.424e+05 -11.013 < 2e-16 \*\*\*  
## LotFrontage 1.037e+02 4.222e+01 2.456 0.014172 \*   
## LotArea 4.377e-01 8.531e-02 5.131 3.31e-07 \*\*\*  
## StreetPave 2.531e+04 1.127e+04 2.246 0.024889 \*   
## LandContourHLS 8.413e+03 5.169e+03 1.628 0.103832   
## LandContourLow -8.735e+03 6.076e+03 -1.438 0.150789   
## LandContourLvl 2.710e+03 3.590e+03 0.755 0.450427   
## LotConfigCulDSac 1.040e+04 3.479e+03 2.988 0.002858 \*\*   
## LotConfigFR2 -5.635e+03 4.085e+03 -1.379 0.168073   
## LotConfigFR3 -2.052e+04 1.295e+04 -1.585 0.113253   
## LotConfigInside -2.043e+02 1.808e+03 -0.113 0.910008   
## NeighborhoodBlueste 1.313e+04 1.907e+04 0.688 0.491419   
## NeighborhoodBrDale 1.162e+04 9.755e+03 1.191 0.233888   
## NeighborhoodBrkSide 9.364e+03 8.526e+03 1.098 0.272294   
## NeighborhoodClearCr -5.717e+03 8.999e+03 -0.635 0.525325   
## NeighborhoodCollgCr -6.544e+03 7.200e+03 -0.909 0.363541   
## NeighborhoodCrawfor 1.789e+04 8.372e+03 2.137 0.032754 \*   
## NeighborhoodEdwards -5.786e+03 7.747e+03 -0.747 0.455289   
## NeighborhoodGilbert -7.172e+03 7.537e+03 -0.952 0.341502   
## NeighborhoodIDOTRR -3.354e+03 8.955e+03 -0.374 0.708117   
## NeighborhoodMeadowV -4.402e+03 1.044e+04 -0.422 0.673359   
## NeighborhoodMitchel -9.081e+03 7.941e+03 -1.144 0.252990   
## NeighborhoodNAmes -5.246e+03 7.645e+03 -0.686 0.492698   
## NeighborhoodNoRidge 2.522e+04 8.390e+03 3.006 0.002694 \*\*   
## NeighborhoodNPkVill 2.286e+04 1.116e+04 2.049 0.040665 \*   
## NeighborhoodNridgHt 3.464e+04 7.407e+03 4.677 3.21e-06 \*\*\*  
## NeighborhoodNWAmes -9.604e+03 7.924e+03 -1.212 0.225732   
## NeighborhoodOldTown -3.221e+03 8.274e+03 -0.389 0.697101   
## NeighborhoodSawyer 1.546e+03 7.987e+03 0.194 0.846589   
## NeighborhoodSawyerW -3.870e+02 7.799e+03 -0.050 0.960431   
## NeighborhoodSomerst 7.133e+03 7.449e+03 0.958 0.338389   
## NeighborhoodStoneBr 4.023e+04 8.375e+03 4.803 1.74e-06 \*\*\*  
## NeighborhoodSWISU 4.451e+03 9.510e+03 0.468 0.639878   
## NeighborhoodTimber -5.936e+03 8.078e+03 -0.735 0.462599   
## NeighborhoodVeenker 8.365e+03 1.041e+04 0.803 0.421981   
## Condition1Feedr 6.216e+03 5.023e+03 1.238 0.216078   
## Condition1Norm 1.497e+04 4.196e+03 3.568 0.000373 \*\*\*  
## Condition1PosA 7.061e+03 1.014e+04 0.697 0.486217   
## Condition1PosN 1.202e+04 7.529e+03 1.596 0.110732   
## Condition1RRAe -1.054e+04 8.839e+03 -1.192 0.233327   
## Condition1RRAn 1.547e+04 6.949e+03 2.227 0.026147 \*   
## Condition1RRNe -6.287e+03 1.834e+04 -0.343 0.731773   
## Condition1RRNn 2.029e+04 1.257e+04 1.614 0.106710   
## Condition2Feedr -1.424e+04 2.230e+04 -0.639 0.523067   
## Condition2Norm -1.377e+04 1.923e+04 -0.716 0.473977   
## Condition2PosA 5.910e+04 3.269e+04 1.808 0.070839 .   
## Condition2PosN -2.399e+05 2.729e+04 -8.789 < 2e-16 \*\*\*  
## Condition2RRAe -3.296e+04 3.995e+04 -0.825 0.409495   
## Condition2RRAn -1.173e+04 3.116e+04 -0.376 0.706699   
## Condition2RRNn -7.288e+03 2.623e+04 -0.278 0.781123   
## BldgType 6.859e+03 1.056e+03 6.498 1.15e-10 \*\*\*  
## HouseStyle1.5Unf 1.042e+04 7.624e+03 1.367 0.171909   
## HouseStyle1Story 1.043e+04 3.565e+03 2.925 0.003499 \*\*   
## HouseStyle2.5Fin -1.782e+04 1.042e+04 -1.710 0.087429 .   
## HouseStyle2.5Unf -1.700e+04 8.633e+03 -1.970 0.049095 \*   
## HouseStyle2Story -4.020e+03 3.156e+03 -1.274 0.203008   
## HouseStyleSFoyer -1.698e+02 5.670e+03 -0.030 0.976117   
## HouseStyleSLvl 4.878e+03 4.657e+03 1.047 0.295104   
## OverallQual 8.402e+03 1.004e+03 8.368 < 2e-16 \*\*\*  
## OverallCond 5.679e+03 7.777e+02 7.302 4.89e-13 \*\*\*  
## YearBuilt 3.484e+02 6.616e+01 5.266 1.62e-07 \*\*\*  
## RoofMatlCompShg 6.965e+05 3.271e+04 21.294 < 2e-16 \*\*\*  
## RoofMatlMembran 7.430e+05 4.226e+04 17.580 < 2e-16 \*\*\*  
## RoofMatlMetal 7.132e+05 4.181e+04 17.055 < 2e-16 \*\*\*  
## RoofMatlRoll 6.864e+05 4.122e+04 16.652 < 2e-16 \*\*\*  
## RoofMatlTar&Grv 6.861e+05 3.363e+04 20.403 < 2e-16 \*\*\*  
## RoofMatlWdShake 6.886e+05 3.498e+04 19.684 < 2e-16 \*\*\*  
## RoofMatlWdShngl 7.578e+05 3.424e+04 22.132 < 2e-16 \*\*\*  
## Exterior1stAsphShn -4.425e+03 2.606e+04 -0.170 0.865184   
## Exterior1stBrkComm -2.939e+03 1.923e+04 -0.153 0.878529   
## Exterior1stBrkFace 1.499e+04 7.123e+03 2.104 0.035540 \*   
## Exterior1stCBlock 5.720e+03 2.580e+04 0.222 0.824562   
## Exterior1stCemntBd 1.060e+04 7.468e+03 1.420 0.155880   
## Exterior1stHdBoard -3.238e+03 6.460e+03 -0.501 0.616249   
## Exterior1stImStucc -7.266e+03 2.576e+04 -0.282 0.777956   
## Exterior1stMetalSd 4.723e+03 6.271e+03 0.753 0.451470   
## Exterior1stPlywood -6.702e+03 6.742e+03 -0.994 0.320365   
## Exterior1stStone -1.249e+04 2.000e+04 -0.624 0.532591   
## Exterior1stStucco 4.036e+03 7.957e+03 0.507 0.612047   
## Exterior1stVinylSd 2.373e+01 6.322e+03 0.004 0.997006   
## Exterior1stWd Sdng 3.090e+02 6.268e+03 0.049 0.960692   
## Exterior1stWdShing 4.409e+02 7.805e+03 0.056 0.954957   
## MasVnrTypeBrkFace 8.054e+03 6.734e+03 1.196 0.231917   
## MasVnrTypeNone 1.480e+04 6.768e+03 2.186 0.028965 \*   
## MasVnrTypeStone 1.243e+04 7.132e+03 1.743 0.081642 .   
## MasVnrArea 3.303e+01 5.738e+00 5.756 1.07e-08 \*\*\*  
## ExterQual 4.938e+03 2.121e+03 2.328 0.020049 \*   
## ExterCond -3.769e+03 2.117e+03 -1.781 0.075220 .   
## FoundationCBlock 2.655e+03 3.086e+03 0.860 0.389845   
## FoundationPConc 4.671e+03 3.422e+03 1.365 0.172433   
## FoundationSlab 1.586e+04 8.418e+03 1.884 0.059782 .   
## FoundationStone 4.068e+03 1.084e+04 0.375 0.707598   
## FoundationWood -3.042e+04 1.513e+04 -2.011 0.044481 \*   
## BsmtQual 3.454e+03 1.692e+03 2.041 0.041453 \*   
## BsmtCond -6.451e+03 2.074e+03 -3.111 0.001907 \*\*   
## BsmtExposure 5.422e+03 8.375e+02 6.474 1.34e-10 \*\*\*  
## BsmtFinSF1 3.741e+01 4.385e+00 8.532 < 2e-16 \*\*\*  
## BsmtFinSF2 2.559e+01 5.752e+00 4.450 9.33e-06 \*\*\*  
## BsmtUnfSF 1.664e+01 4.280e+00 3.887 0.000106 \*\*\*  
## X1stFlrSF 4.938e+01 5.139e+00 9.609 < 2e-16 \*\*\*  
## X2ndFlrSF 6.555e+01 4.755e+00 13.786 < 2e-16 \*\*\*  
## BedroomAbvGr -6.358e+03 1.306e+03 -4.868 1.26e-06 \*\*\*  
## KitchenAbvGr -1.191e+04 4.218e+03 -2.824 0.004811 \*\*   
## KitchenQual 6.727e+03 1.602e+03 4.200 2.85e-05 \*\*\*  
## TotRmsAbvGrd 3.240e+03 9.455e+02 3.427 0.000629 \*\*\*  
## Functional 5.596e+03 1.090e+03 5.135 3.25e-07 \*\*\*  
## Fireplaces 5.450e+03 2.256e+03 2.416 0.015823 \*   
## FireplaceQu -1.629e+03 8.157e+02 -1.997 0.046014 \*   
## GarageFinish1 -6.290e+03 1.800e+04 -0.350 0.726749   
## GarageFinish2 -1.195e+04 1.810e+04 -0.661 0.508950   
## GarageFinish3 -5.459e+03 1.805e+04 -0.302 0.762375   
## GarageCars 4.323e+03 2.289e+03 1.889 0.059166 .   
## GarageArea 1.610e+01 7.632e+00 2.110 0.035048 \*   
## GarageQual1 -1.250e+05 3.619e+04 -3.455 0.000567 \*\*\*  
## GarageQual2 -1.191e+05 3.024e+04 -3.938 8.64e-05 \*\*\*  
## GarageQual3 -1.175e+05 2.993e+04 -3.926 9.08e-05 \*\*\*  
## GarageQual4 -1.146e+05 3.107e+04 -3.690 0.000233 \*\*\*  
## GarageQual5 NA NA NA NA   
## GarageCond1 1.162e+05 3.748e+04 3.102 0.001964 \*\*   
## GarageCond2 1.103e+05 3.505e+04 3.146 0.001692 \*\*   
## GarageCond3 1.113e+05 3.477e+04 3.200 0.001406 \*\*   
## GarageCond4 1.102e+05 3.641e+04 3.027 0.002516 \*\*   
## GarageCond5 NA NA NA NA   
## WoodDeckSF 1.233e+01 5.929e+00 2.080 0.037752 \*   
## X3SsnPorch 4.516e+01 2.317e+01 1.949 0.051501 .   
## ScreenPorch 3.110e+01 1.262e+01 2.463 0.013889 \*   
## PoolQC 1.807e+04 2.912e+03 6.203 7.37e-10 \*\*\*  
## FenceGdWo 9.803e+03 4.954e+03 1.979 0.048037 \*   
## FenceMnPrv 1.197e+04 4.058e+03 2.949 0.003240 \*\*   
## FenceMnWw 6.638e+03 8.359e+03 0.794 0.427249   
## FenceNoFence 9.751e+03 3.714e+03 2.626 0.008746 \*\*   
## MiscFeatureNoMiscFeature 1.499e+03 2.458e+04 0.061 0.951390   
## MiscFeatureOthr 1.167e+04 3.029e+04 0.385 0.700118   
## MiscFeatureShed 4.039e+03 2.483e+04 0.163 0.870795   
## MiscFeatureTenC -6.949e+04 3.624e+04 -1.917 0.055404 .   
## MoSold -5.896e+02 2.454e+02 -2.403 0.016401 \*   
## SaleConditionAdjLand 2.085e+04 1.324e+04 1.575 0.115443   
## SaleConditionAlloca 5.661e+03 8.365e+03 0.677 0.498679   
## SaleConditionFamily 2.350e+02 6.139e+03 0.038 0.969477   
## SaleConditionNormal 7.691e+03 2.691e+03 2.858 0.004325 \*\*   
## SaleConditionPartial 2.473e+04 3.804e+03 6.500 1.14e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 24080 on 1321 degrees of freedom  
## Multiple R-squared: 0.9168, Adjusted R-squared: 0.9081   
## F-statistic: 105.5 on 138 and 1321 DF, p-value: < 2.2e-16

### 3. Boruta Feature Selection

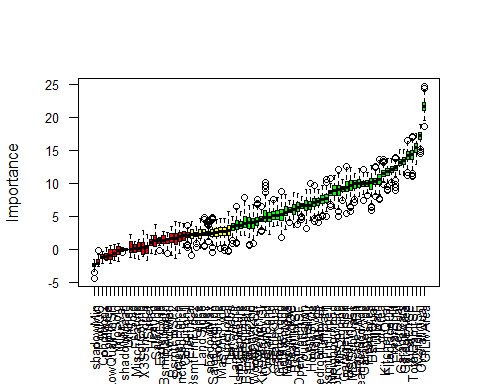
Based on Random forest. Selects attributes better than shadow.

set.seed(99)  
  
FS\_Boruta<-Boruta(SalePrice~.,data=Train, maxRuns=200)

print(FS\_Boruta)

## Boruta performed 199 iterations in 7.483495 mins.  
## 48 attributes confirmed important: BedroomAbvGr, BldgType,  
## BsmtCond, BsmtExposure, BsmtFinSF1 and 43 more;  
## 22 attributes confirmed unimportant: BsmtFinSF2, BsmtHalfBath,  
## Condition1, Condition2, Electrical and 17 more;  
## 9 tentative attributes left: Alley, BsmtFinType2, Fence,  
## Functional, LandSlope and 4 more;

plot(FS\_Boruta,las=2, cex.axis=0.8) # Shows the confirmed feature green, tentative yellow and rejected red.



getConfirmedFormula(FS\_Boruta)

## SalePrice ~ MSSubClass + MSZoning + LotFrontage + LotArea + LandContour +   
## Neighborhood + BldgType + HouseStyle + OverallQual + OverallCond +   
## YearBuilt + YearRemodAdd + Exterior1st + Exterior2nd + MasVnrArea +   
## ExterQual + Foundation + BsmtQual + BsmtCond + BsmtExposure +   
## BsmtFinType1 + BsmtFinSF1 + BsmtUnfSF + TotalBsmtSF + HeatingQC +   
## CentralAir + X1stFlrSF + X2ndFlrSF + GrLivArea + BsmtFullBath +   
## FullBath + HalfBath + BedroomAbvGr + KitchenAbvGr + KitchenQual +   
## TotRmsAbvGrd + Fireplaces + FireplaceQu + GarageType + GarageYrBlt +   
## GarageFinish + GarageCars + GarageArea + GarageQual + GarageCond +   
## PavedDrive + WoodDeckSF + OpenPorchSF  
## <environment: 0x0000000020bcd808>

Graph shows important features. Ground Level Area scored the highest then the overall quality.

borcheck<-TentativeRoughFix(FS\_Boruta)  
print(borcheck)

## Boruta performed 199 iterations in 7.483495 mins.  
## Tentatives roughfixed over the last 199 iterations.  
## 55 attributes confirmed important: Alley, BedroomAbvGr, BldgType,  
## BsmtCond, BsmtExposure and 50 more;  
## 24 attributes confirmed unimportant: BsmtFinSF2, BsmtFinType2,  
## BsmtHalfBath, Condition1, Condition2 and 19 more;

getSelectedAttributes(borcheck, withTentative = F)

## [1] "MSSubClass" "MSZoning" "LotFrontage" "LotArea"   
## [5] "Alley" "LotShape" "LandContour" "LandSlope"   
## [9] "Neighborhood" "BldgType" "HouseStyle" "OverallQual"   
## [13] "OverallCond" "YearBuilt" "YearRemodAdd" "Exterior1st"   
## [17] "Exterior2nd" "MasVnrType" "MasVnrArea" "ExterQual"   
## [21] "Foundation" "BsmtQual" "BsmtCond" "BsmtExposure"   
## [25] "BsmtFinType1" "BsmtFinSF1" "BsmtUnfSF" "TotalBsmtSF"   
## [29] "HeatingQC" "CentralAir" "X1stFlrSF" "X2ndFlrSF"   
## [33] "GrLivArea" "BsmtFullBath" "FullBath" "HalfBath"   
## [37] "BedroomAbvGr" "KitchenAbvGr" "KitchenQual" "TotRmsAbvGrd"   
## [41] "Functional" "Fireplaces" "FireplaceQu" "GarageType"   
## [45] "GarageYrBlt" "GarageFinish" "GarageCars" "GarageArea"   
## [49] "GarageQual" "GarageCond" "PavedDrive" "WoodDeckSF"   
## [53] "OpenPorchSF" "Fence" "SaleCondition"

Three options from three different selections.