IDA HW5 Group 16

Vignesh Murugan and Vivek Talluri

2024-10-05

```r
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.4      v readr     2.1.5
## v forcats   1.0.0      v stringr   1.5.1
## v ggplot2   3.5.1      v tibble    3.2.1
## v lubridate 1.9.3      v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(gridExtra)
```

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

```r
library(VIM)
```

```
## Loading required package: colorspace
## Loading required package: grid
## VIM is ready to use.
##
## Suggestions and bug-reports can be submitted at: https://github.com/statistikat/VIM/issues
##
## Attaching package: 'VIM'
##
## The following object is masked from 'package:datasets':
##
##     sleep
```

```r
library(e1071)
library(caret)
```

```
## Loading required package: lattice
##
## Attaching package: 'caret'
##
## The following object is masked from 'package:purrr':
##
##     lift
```

```r
library(car)
```

```
## Loading required package: carData
##
## Attaching package: 'car'
##
## The following object is masked from 'package:dplyr':
##
##     recode
##
## The following object is masked from 'package:purrr':
##
##     some
```

```r
library(MASS)
```

```
##
## Attaching package: 'MASS'
##
## The following object is masked from 'package:dplyr':
##
##     select
```

```r
library(pls)
```

```
##
## Attaching package: 'pls'
##
## The following object is masked from 'package:caret':
##
##     R2
##
## The following object is masked from 'package:stats':
##
##     loadings
```

```r
library(glmnet)
```

```
## Loading required package: Matrix
##
## Attaching package: 'Matrix'
##
## The following objects are masked from 'package:tidyr':
```

```
## 
##      expand, pack, unpack
## 
## Loaded glmnet 4.1-8
```

```r
library(knitr)
```

Data Is PreProcessed before Modelling

```r
#  Data Quality Report
```

```r
housingData <- read_csv("housingData.csv")
```

```
## Rows: 1000 Columns: 74
## -- Column specification ---------------------------------------------------------
## Delimiter: ","
## chr (38): MSZoning, Alley, LotShape, LandContour, LotConfig, LandSlope, Neig...
## dbl (36): Id, MSSubClass, LotFrontage, LotArea, OverallQual, OverallCond, Ye...
## 
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
is_tibble(housingData)
```

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

```r
# Creating New Variables called age, ageSinceRemodel, ageofGarage
housingData <- housingData %>%
  mutate(age = YrSold - YearBuilt,
         ageSinceRemodel = YrSold - YearRemodAdd,
         ageofGarage = YrSold - GarageYrBlt,
         logSalePrice = log(SalePrice))

# Creating New Tibbles for Numeric and Factor Data
housingNumeric <- housingData %>% select_if(is.numeric)
housingFactor <- housingData %>% transmute_if(is.character, as.factor)

# Funtion for Calculating Q1 Quartile.
Q1<-function(x,na.rm=TRUE) {
  quantile(x,na.rm=na.rm)[2]
}

# Funtion for Calculating Q3 Quartile.
Q3<-function(x,na.rm=TRUE) {
  quantile(x,na.rm=na.rm)[4]
}
```

```r
# Funtion to set Summary Functions for Numeric Data
myNumericSummary <- function(x) {
  c(length(x),
    n_distinct(x),
    sum(is.na(x)),
    mean(x, na.rm = TRUE),
    min(x, na.rm = TRUE),
    Q1(x, na.rm = TRUE),
    median(x, na.rm = TRUE),
    Q3(x, na.rm = TRUE),
    max(x, na.rm = TRUE),
    sd(x, na.rm = TRUE))
}

# Compute Summary of Numeric Data
numericSummary <- housingNumeric %>%
  reframe(across(everything(), myNumericSummary)) %>%
  cbind(stat = c("n", "unique", "missing", "mean", "min", "Q1", "median", "Q3", "max", "sd")) %>%
  pivot_longer("Id":"logSalePrice", names_to = "variable", values_to = "value") %>%
  pivot_wider(names_from = stat, values_from = value) %>%
  mutate(
    n = as.numeric(n),
    unique = as.numeric(unique),
    missing = as.numeric(missing),
    missing_pct = 100 * missing / n,
    unique_pct = 100 * unique / n
  ) %>%
  dplyr::select(variable, n, missing, missing_pct, unique, unique_pct, everything())

# Variables with Missing data Percentage :
# LotFrontage - 20.7%
# MasVnrArea - 0.4%
# GarageYrBlt - 5.3%
# ageofGarage - 5.3%

# ID Variable is removed because it not a valid variable for analysis
housingNumeric_NoID <- housingNumeric %>%  dplyr::select(-Id)

housingNumeric_imputed <- kNN(housingNumeric_NoID, , imp_var=FALSE)

p1 <- ggplot(housingNumeric_NoID, aes(x = LotFrontage)) +
  geom_histogram(bins = 30, fill = "cornflowerblue", color = "black") +
  ggtitle("Distribution of LotFrontage ") +
  xlab("LotFrontage") +
  ylab("Frequency")

p3 <- ggplot(housingNumeric_imputed, aes(x = LotFrontage)) +
  geom_histogram(bins = 30, fill = "salmon", color = "black") +
  ggtitle("LotFrontage after Imputation") +
  xlab("LotFrontage") +
  ylab("Frequency")

# Calculate the maximum y-axis limit based on bin counts directly from ggplot
```
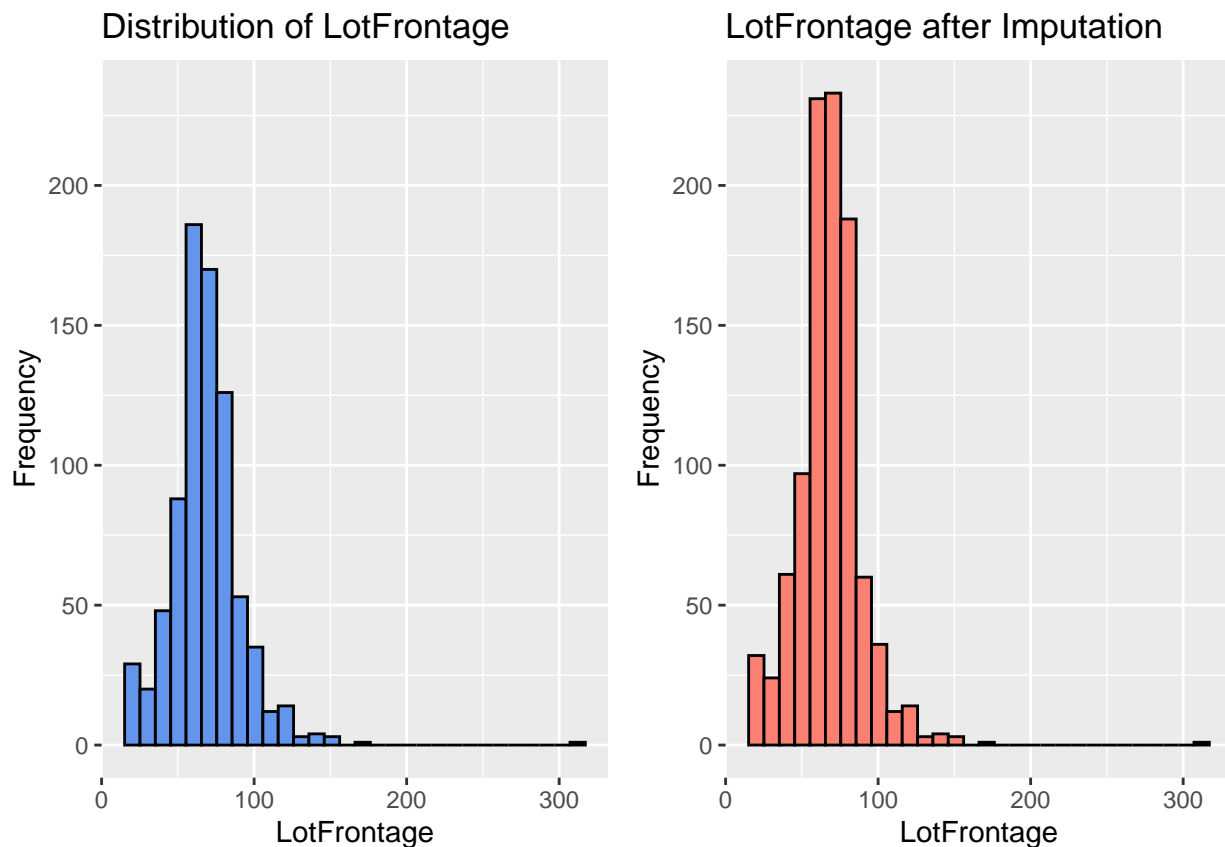
```
max_y <- max(layer_data(p1)$count, layer_data(p3)$count)
```

```
## Warning: Removed 207 rows containing non-finite outside the scale range
## ('stat_bin()').
```

```
# Set the y-axis limit to the same range for both plots
p1 <- p1 + ylim(0, max_y)
p3 <- p3 + ylim(0, max_y)

# Arrange the plots side by side
grid.arrange(p1, p3, ncol = 2)
```

```
## Warning: Removed 207 rows containing non-finite outside the scale range
## ('stat_bin()').
```



```
# Computing the skewness of variables and Sorting them
skewValues_tibble <- housingNumeric_imputed %>%
  summarise(across(everything(), ~ skewness(.x, na.rm = TRUE))) %>%
  pivot_longer(cols = everything(), names_to = "Variable", values_to = "Skewness") %>%
  arrange(Skewness)

Trans <- preProcess(housingNumeric_imputed, method = "BoxCox")

# Apply the Box-Cox transformation to the Imputed Data
```

```r
housingNumeric_transformed <- predict(Trans, housingNumeric_imputed)

skewValues_tibble_Trans <- housingNumeric_transformed %>%
  summarise(across(everything(), ~ skewness(.x, na.rm = TRUE))) %>%
  pivot_longer(cols = everything(), names_to = "Variable", values_to = "Skewness") %>%
  arrange(Skewness)

# There are skewed Variables even after boxcox Transformation , Dont know if we should apply more trans
# I believe in removing them as Linear regression requires normality
# I didnt remove them


# Function to compute mode for Factor data
getmodes <- function(v,type=1) {
  tbl <- table(v)
  m1<-which.max(tbl)
  if (type==1) {
    return (names(m1)) #1st mode
  }
  else if (type==2) {
    return (names(which.max(tbl[-m1]))) #2nd mode
  }
  else if (type==-1) {
    return (names(which.min(tbl))) #least common mode
  }
  else {
    stop("Invalid type selected")
  }
}

# funtion to compute the count of mode for Factor data
getmodesCnt <- function(v,type=1) {
  tbl <- table(v)
  m1<-which.max(tbl)
  if (type==1) {
    return (max(tbl)) #1st mode freq
  }
  else if (type==2) {
    return (max(tbl[-m1])) #2nd mode freq
  }
  else if (type==-1) {
    return (min(tbl)) #least common freq
  }
  else {
    stop("Invalid type selected")
  }
}

# Funtion to set Summary Functions for Factor Data
myFactorSummary <- function(x) {
  c(length(x),
    n_distinct(x),
    sum(is.na(x)),
    getmodes(x, type = 1),
```

```r
    getmodesCnt(x, type = 1),
    getmodes(x, type = 2),
    getmodesCnt(x, type = 2),
    getmodes(x, type = -1),
    getmodesCnt(x, type = -1))
}

# Compute Summary of Factor Data
FactorSummary <- housingFactor %>%
  reframe(across(everything(), myFactorSummary)) %>%
  cbind(stat = c("n", "unique", "missing", "most_common", "most_common_count",
                 "2nd_most_common", "2nd_most_common_count", "least_common",
                 "least_common_count")) %>%
  pivot_longer("MSZoning":"SaleType", names_to = "variable", values_to = "value") %>%
  pivot_wider(names_from = stat, values_from = value) %>%
  mutate(
    n = as.numeric(n),
    unique = as.numeric(unique),
    missing = as.numeric(missing),
    missing_pct = 100 * missing / n,
    unique_pct = 100 * unique / n
  ) %>%
  dplyr::select(variable, n, missing, missing_pct, unique, unique_pct, everything())

# Missing value Percentage Of Varibales with Missing Values
# Alley - 93.8%
# MasVnrType - 0.4%
# BsmtQual - 3.1%
# BsmtCond - 3.1%
# BsmtExposure - 3.2%
# BsmtFinType1 - 3.1%
# BsmtFinType2 - 3.2%
# FireplaceQu - 46.6%
# GarageType - 5.3%
# GarageFinish - 5.3%
# GarageQual - 5.3%
# GarageCond - 5.3%
# PoolQC - 99.8%
# Fence - 80.5%
# MiscFeature - 96.6%
# Electrical - 0.1%

# Dropping variables which have missing values grater than 60%
housingFactor <- housingFactor %>%
  dplyr::select(-c(Alley, PoolQC, MiscFeature, Fence))

# Impute Missing Values in housingFactor
housingFactor_imputed <- kNN(housingFactor, imp_var=FALSE)

# # Calculate frequencies directly from the data
# count_before <- table(housingFactor$FireplaceQu)
# count_after <- table(housingFactor_imputed$FireplaceQu)
#
```

```r
# max_y <- max(c(count_before, count_after))

# Bar plot before imputation
p1 <- ggplot(housingFactor, aes(x = FireplaceQu)) +
  geom_bar(fill = "cornflowerblue", color = "black") +
  ggtitle("Distribution of FireplaceQu (Before Imputation)") +
  xlab("FireplaceQu") +
  ylab("Frequency")

# Bar plot after imputation
p3 <- ggplot(housingFactor_imputed, aes(x = FireplaceQu)) +
  geom_bar(fill = "salmon", color = "black") +
  ggtitle("Distribution of FireplaceQu (After Imputation)") +
  xlab("FireplaceQu") +
  ylab("Frequency")

# Determine the maximum y-axis value from both plots
max_y <- max(layer_data(p1)$count, layer_data(p3)$count)

# Set y-axis limits for both plots
p1 <- p1 + ylim(0, max_y)
p3 <- p3 + ylim(0, max_y)

# Arrange plots side by side
grid.arrange(p1, p3, ncol = 2)
```
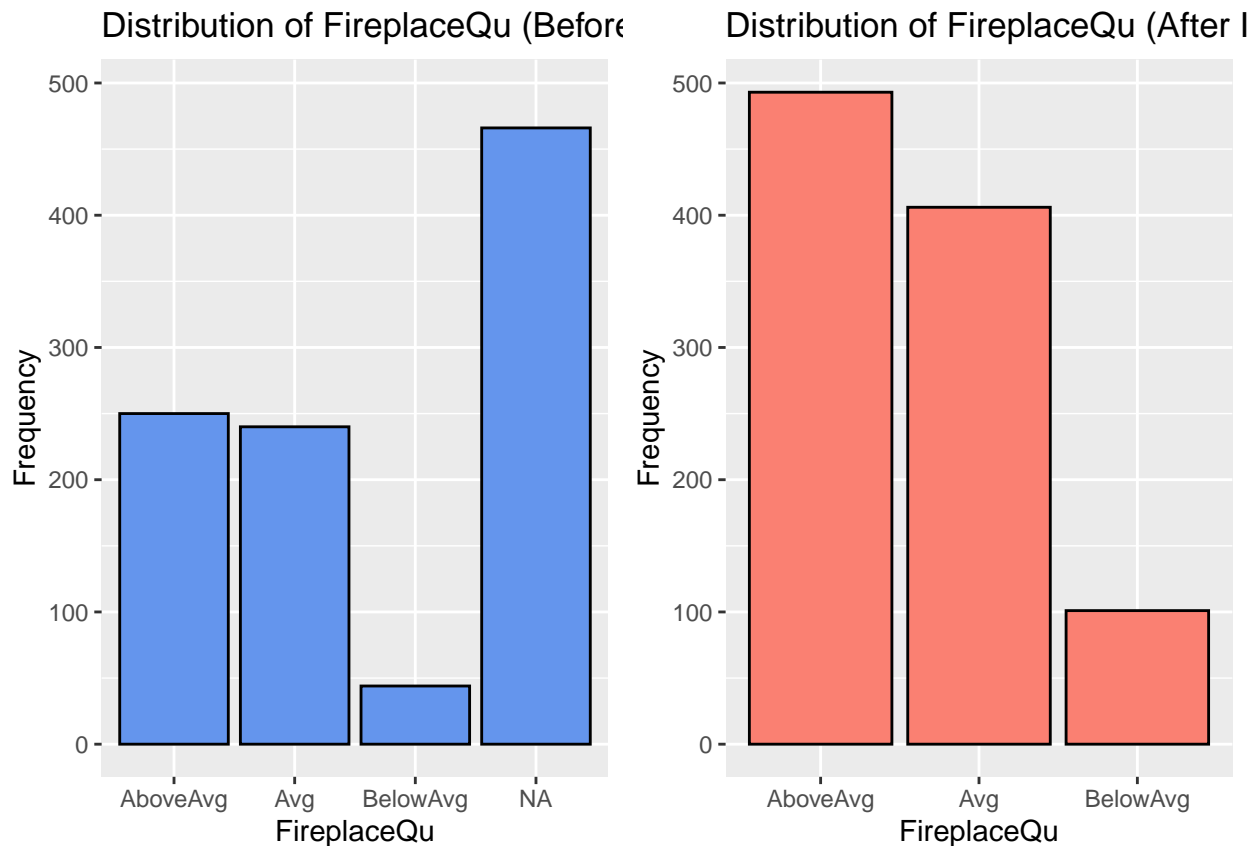
```r
factor_levels <- map(housingFactor_imputed, ~fct_count(.x, sort = TRUE))
# Factors to Collapse :
# MSZoning - collapse to 3 groups
# LotShape - collapse to 3 groups
# HouseStyle - collapse to 4 groups

# Collapse MSZoning to 3 groups
housingFactor_imputed$MSZoning <- fct_lump_n(housingFactor_imputed$MSZoning, n=2)
fct_count(housingFactor_imputed$MSZoning)
```

```
## # A tibble: 3 x 2
##   f         n
##   <fct> <int>
## 1 RL      803
## 2 RM      151
## 3 Other    46
```

```r
# Collapse LotShape tp 3 groups
housingFactor_imputed$LotShape <- fct_lump_n(housingFactor_imputed$LotShape, n=2)
fct_count(housingFactor_imputed$LotShape)
```

```
## # A tibble: 3 x 2
##   f         n
##   <fct> <int>
## 1 IR1     330
## 2 Reg     633
## 3 Other    37
```

```r
# Collapse HouseStyle to 4 groups
housingFactor_imputed$HouseStyle <- fct_lump_n(housingFactor_imputed$HouseStyle, n=3)
fct_count(housingFactor_imputed$HouseStyle)
```

```
## # A tibble: 4 x 2
##   f          n
##   <fct>  <int>
## 1 1.5Fin   109
## 2 1Story   488
## 3 2Story   310
## 4 Other     93
```

```r
# Combining Numeric and Factor variables
housingDataProcessed <- bind_cols(housingNumeric_transformed, housingFactor_imputed)

# Drop SalePrice as its high Correlated with LogSalePrice
housingDataProcessed <- housingDataProcessed %>%
  dplyr::select(-SalePrice)

# Outlier Detection and Removal
housingDataProcessed <-  housingDataProcessed[-c(402, 124), ]

ols_fit <- lm(logSalePrice ~ ., data = housingDataProcessed)
summary(ols_fit)
```

```
##
## Call:
## lm(formula = logSalePrice ~ ., data = housingDataProcessed)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0258614 -0.0035834  0.0002115  0.0040149  0.0211165
##
## Coefficients: (1 not defined because of singularities)
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -1.216e+00  3.107e+00  -0.391 0.695564
## MSSubClass       3.055e-04  1.139e-03   0.268 0.788593
## LotFrontage     -2.262e-04  1.514e-04  -1.494 0.135634
## LotArea          9.787e-03  9.108e-04  10.746  < 2e-16 ***
## OverallQual      3.721e-03  3.382e-04  11.003  < 2e-16 ***
## OverallCond      1.490e-02  1.200e-03  12.420  < 2e-16 ***
## YearBuilt        4.999e-08  2.711e-08   1.844 0.065588 .
## YearRemodAdd    -2.385e-06  2.357e-06  -1.012 0.311968
## MasVnrArea       3.936e-06  1.921e-06   2.049 0.040742 *
## BsmtFinSF1       1.151e-05  1.364e-06   8.437  < 2e-16 ***
## BsmtFinSF2       1.230e-05  3.241e-06   3.795 0.000158 ***
## BsmtUnfSF        7.342e-06  1.182e-06   6.212 8.22e-10 ***
## TotalBsmtSF             NA         NA      NA       NA
## X1stFlrSF        1.269e-02  3.311e-03   3.832 0.000137 ***
## X2ndFlrSF        1.188e-05  2.355e-06   5.043 5.62e-07 ***
## LowQualFinSF     5.290e-06  5.891e-06   0.898 0.369463
## GrLivArea        1.867e-02  3.974e-03   4.698 3.06e-06 ***
## BsmtFullBath     2.025e-03  6.707e-04   3.019 0.002614 **
## BsmtHalfBath     1.279e-03  1.030e-03   1.243 0.214355
## FullBath         1.228e-03  7.942e-04   1.547 0.122320
## HalfBath         2.945e-04  7.414e-04   0.397 0.691310
## BedroomAbvGr    -2.092e-04  4.907e-04  -0.426 0.669917
## KitchenAbvGr    -1.078e-02  6.121e-03  -1.761 0.078530 .
## TotRmsAbvGrd    -1.328e-03  1.238e-03  -1.073 0.283644
## Fireplaces       1.999e-03  4.545e-04   4.397 1.23e-05 ***
## GarageYrBlt      1.413e-07  1.084e-07   1.303 0.192809
## GarageCars       2.505e-03  7.518e-04   3.332 0.000899 ***
## GarageArea       5.248e-06  2.788e-06   1.882 0.060127 .
## WoodDeckSF       4.955e-06  2.015e-06   2.459 0.014120 *
## OpenPorchSF      9.781e-06  3.974e-06   2.461 0.014038 *
## EncPorchSF       1.575e-05  3.001e-06   5.247 1.95e-07 ***
## PoolArea         1.415e-05  8.071e-06   1.754 0.079845 .
## MiscVal          9.465e-07  1.347e-06   0.703 0.482489
## MoSold          -2.967e-05  1.466e-04  -0.202 0.839626
## YrSold           1.010e-03  1.025e-03   0.986 0.324633
## age             -1.249e-03  4.624e-04  -2.702 0.007031 **
## ageSinceRemodel -4.769e-03  4.661e-03  -1.023 0.306504
## ageofGarage      3.089e-04  2.125e-04   1.453 0.146458
## MSZoningRM      -5.532e-03  1.167e-03  -4.740 2.50e-06 ***
## MSZoningOther    6.784e-05  1.871e-03   0.036 0.971081
## LotShapeReg      7.082e-04  5.595e-04   1.266 0.205887
## LotShapeOther   -2.018e-03  1.332e-03  -1.515 0.130148
## LandContourHLS   1.569e-03  1.863e-03   0.842 0.400066
## LandContourLow  -9.305e-04  2.163e-03  -0.430 0.667117
```

```
## LandContourLvl       1.064e-04  1.285e-03   0.083 0.933998
## LotConfigCulDSac      1.207e-03  1.088e-03   1.109 0.267767
## LotConfigInside      -7.005e-04  6.201e-04  -1.130 0.258882
## LotConfigother       -3.298e-03  1.308e-03  -2.522 0.011848 *
## LandSlopeMod         -1.323e-05  1.390e-03  -0.010 0.992407
## LandSlopeSev         -2.337e-03  3.661e-03  -0.638 0.523505
## NeighborhoodClearCr  -4.597e-03  2.274e-03  -2.022 0.043517 *
## NeighborhoodCollgCr  -7.334e-03  1.855e-03  -3.954 8.33e-05 ***
## NeighborhoodCrawfor   5.529e-03  1.912e-03   2.891 0.003938 **
## NeighborhoodEdwards  -1.081e-02  1.639e-03  -6.594 7.53e-11 ***
## NeighborhoodGilbert  -8.471e-03  2.030e-03  -4.173 3.31e-05 ***
## NeighborhoodIDOTRR   -1.843e-03  2.053e-03  -0.897 0.369751
## NeighborhoodMitchel  -9.749e-03  1.956e-03  -4.985 7.50e-07 ***
## NeighborhoodNAmes    -7.704e-03  1.602e-03  -4.810 1.79e-06 ***
## NeighborhoodNoRidge  -5.712e-03  2.267e-03  -2.520 0.011912 *
## NeighborhoodNridgHt  -3.200e-03  2.278e-03  -1.405 0.160476
## NeighborhoodNWAmes   -8.879e-03  1.910e-03  -4.648 3.89e-06 ***
## NeighborhoodOldTown  -5.122e-03  1.596e-03  -3.210 0.001380 **
## Neighborhoodother    -3.878e-03  1.768e-03  -2.193 0.028573 *
## NeighborhoodSawyer   -7.441e-03  1.792e-03  -4.153 3.61e-05 ***
## NeighborhoodSawyerW  -8.837e-03  1.984e-03  -4.455 9.51e-06 ***
## NeighborhoodSomerst  -2.840e-03  2.682e-03  -1.059 0.289940
## NeighborhoodTimber   -6.973e-03  2.372e-03  -2.940 0.003368 **
## Condition1Feedr       2.429e-03  1.679e-03   1.446 0.148472
## Condition1Norm        5.765e-03  1.409e-03   4.090 4.72e-05 ***
## Condition1PosA        1.405e-03  3.115e-03   0.451 0.652010
## Condition1PosN        4.137e-03  2.372e-03   1.744 0.081466 .
## Condition1RR          2.607e-04  2.023e-03   0.129 0.897508
## BldgType2fmCon       -2.172e-03  2.633e-03  -0.825 0.409694
## BldgTypeDuplex       -3.454e-03  2.541e-03  -1.360 0.174338
## BldgTypeTwnhs        -5.407e-04  2.395e-03  -0.226 0.821446
## BldgTypeTwnhsE       -4.476e-05  2.212e-03  -0.020 0.983860
## HouseStyle1Story      1.898e-03  1.512e-03   1.255 0.209685
## HouseStyle2Story      4.147e-04  1.078e-03   0.385 0.700562
## HouseStyleOther       5.554e-04  1.388e-03   0.400 0.689183
## RoofStyleHip          6.617e-04  6.437e-04   1.028 0.304299
## RoofStyleother        4.568e-03  1.736e-03   2.631 0.008664 **
## Exterior1stCemntBd   -1.168e-02  4.681e-03  -2.496 0.012747 *
## Exterior1stHdBoard   -5.473e-03  2.176e-03  -2.515 0.012082 *
## Exterior1stMetalSd   -3.217e-03  3.060e-03  -1.051 0.293395
## Exterior1stother     -3.171e-03  2.159e-03  -1.469 0.142260
## Exterior1stPlywood   -5.440e-03  2.077e-03  -2.620 0.008959 **
## Exterior1stVinylSd   -7.032e-03  2.985e-03  -2.356 0.018699 *
## Exterior1stWd Sdng   -8.822e-03  1.858e-03  -4.749 2.40e-06 ***
## Exterior2ndCmentBd    7.515e-03  4.915e-03   1.529 0.126631
## Exterior2ndHdBoard    1.049e-04  2.521e-03   0.042 0.966803
## Exterior2ndMetalSd   -3.439e-04  3.351e-03  -0.103 0.918272
## Exterior2ndother      6.556e-04  2.490e-03   0.263 0.792367
## Exterior2ndPlywood   -1.126e-03  2.406e-03  -0.468 0.639815
## Exterior2ndVinylSd    2.596e-03  3.263e-03   0.796 0.426457
## Exterior2ndWd Sdng    4.401e-03  2.241e-03   1.964 0.049875 *
## Exterior2ndWd Shng   -5.497e-04  2.767e-03  -0.199 0.842574
## MasVnrTypeBrkFace     3.572e-03  2.712e-03   1.317 0.188241
## MasVnrTypeNone        3.920e-03  2.725e-03   1.438 0.150670
```

```
## MasVnrTypeStone        3.923e-03  2.884e-03   1.360 0.174042
## ExterQualAvg           4.387e-04  8.150e-04   0.538 0.590489
## ExterQualBelowAvg     -5.054e-03  3.325e-03  -1.520 0.128911
## ExterCondAvg           1.589e-03  8.127e-04   1.955 0.050874 .
## ExterCondBelowAvg      2.029e-03  2.244e-03   0.904 0.366366
## FoundationCBlock       1.903e-03  1.096e-03   1.736 0.082858 .
## Foundationother        1.244e-04  1.897e-03   0.066 0.947723
## FoundationPConc        2.461e-03  1.194e-03   2.061 0.039580 *
## BsmtQualAvg            1.051e-03  8.315e-04   1.264 0.206653
## BsmtQualBelowAvg       2.361e-03  1.869e-03   1.263 0.206866
## BsmtCondAvg           -1.332e-03  1.235e-03  -1.079 0.280824
## BsmtCondBelowAvg      -2.531e-03  1.974e-03  -1.282 0.200225
## BsmtExposureGd         2.196e-03  1.061e-03   2.070 0.038798 *
## BsmtExposureMn        -1.581e-03  1.089e-03  -1.452 0.146950
## BsmtExposureNo        -9.782e-04  7.875e-04  -1.242 0.214524
## BsmtFinType1BLQ       -1.938e-03  8.892e-04  -2.180 0.029553 *
## BsmtFinType1GLQ        6.469e-04  8.308e-04   0.779 0.436407
## BsmtFinType1LwQ       -3.472e-03  1.251e-03  -2.775 0.005642 **
## BsmtFinType1Rec       -1.841e-03  1.004e-03  -1.834 0.067008 .
## BsmtFinType1Unf       -2.061e-03  9.851e-04  -2.092 0.036738 *
## BsmtFinType2BLQ       -3.794e-03  2.764e-03  -1.373 0.170196
## BsmtFinType2GLQ       -9.579e-04  3.045e-03  -0.315 0.753197
## BsmtFinType2LwQ       -4.485e-03  2.746e-03  -1.633 0.102811
## BsmtFinType2Rec       -4.408e-03  2.679e-03  -1.645 0.100327
## BsmtFinType2Unf       -2.750e-03  2.914e-03  -0.944 0.345528
## Heatingother           4.261e-03  1.803e-03   2.363 0.018370 *
## HeatingQCAvg          -1.427e-03  6.244e-04  -2.285 0.022552 *
## HeatingQCBelowAvg     -2.831e-03  1.418e-03  -1.996 0.046209 *
## CentralAirY            3.550e-03  1.363e-03   2.604 0.009373 **
## ElectricalFuseF       -4.526e-03  2.054e-03  -2.204 0.027803 *
## ElectricalFuseP       -3.215e-03  5.530e-03  -0.581 0.561188
## ElectricalSBrkr        6.566e-05  9.878e-04   0.066 0.947016
## KitchenQualAvg        -8.185e-04  7.278e-04  -1.125 0.261062
## KitchenQualBelowAvg   -2.274e-03  1.767e-03  -1.287 0.198512
## FunctionalMaj2        -2.067e-02  4.281e-03  -4.828 1.63e-06 ***
## FunctionalMin1         1.339e-03  2.728e-03   0.491 0.623588
## FunctionalMin2         4.221e-03  2.734e-03   1.544 0.122949
## FunctionalMod          1.415e-03  3.161e-03   0.448 0.654620
## FunctionalTyp          7.037e-03  2.322e-03   3.031 0.002511 **
## FireplaceQuAvg        -1.721e-04  5.964e-04  -0.289 0.772976
## FireplaceQuBelowAvg    8.315e-05  8.838e-04   0.094 0.925062
## GarageTypeAttchd       9.164e-03  4.492e-03   2.040 0.041642 *
## GarageTypeBasment      7.763e-03  5.094e-03   1.524 0.127845
## GarageTypeBuiltIn      9.773e-03  4.658e-03   2.098 0.036205 *
## GarageTypeCarPort      6.929e-03  5.814e-03   1.192 0.233680
## GarageTypeDetchd       1.006e-02  4.448e-03   2.262 0.023944 *
## GarageFinishRFn       -9.215e-04  6.755e-04  -1.364 0.172889
## GarageFinishUnf       -1.078e-03  8.003e-04  -1.347 0.178255
## GarageQualAvg         -4.160e-03  2.892e-03  -1.439 0.150648
## GarageQualBelowAvg    -6.224e-03  3.332e-03  -1.868 0.062090 .
## GarageCondAvg          1.285e-03  3.065e-03   0.419 0.675104
## GarageCondBelowAvg    -3.715e-04  3.468e-03  -0.107 0.914721
## PavedDriveP           -1.195e-03  1.743e-03  -0.686 0.492979
## PavedDriveY            1.807e-03  1.134e-03   1.594 0.111404
```

```
## SaleTypeWD              2.634e-06  1.345e-03    0.002 0.998439
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.006552 on 846 degrees of freedom
## Multiple R-squared:  0.9596, Adjusted R-squared:  0.9524
## F-statistic: 133.2 on 151 and 846 DF,  p-value: < 2.2e-16
```

```
par(mfrow=c(2,2))
plot(ols_fit)
```
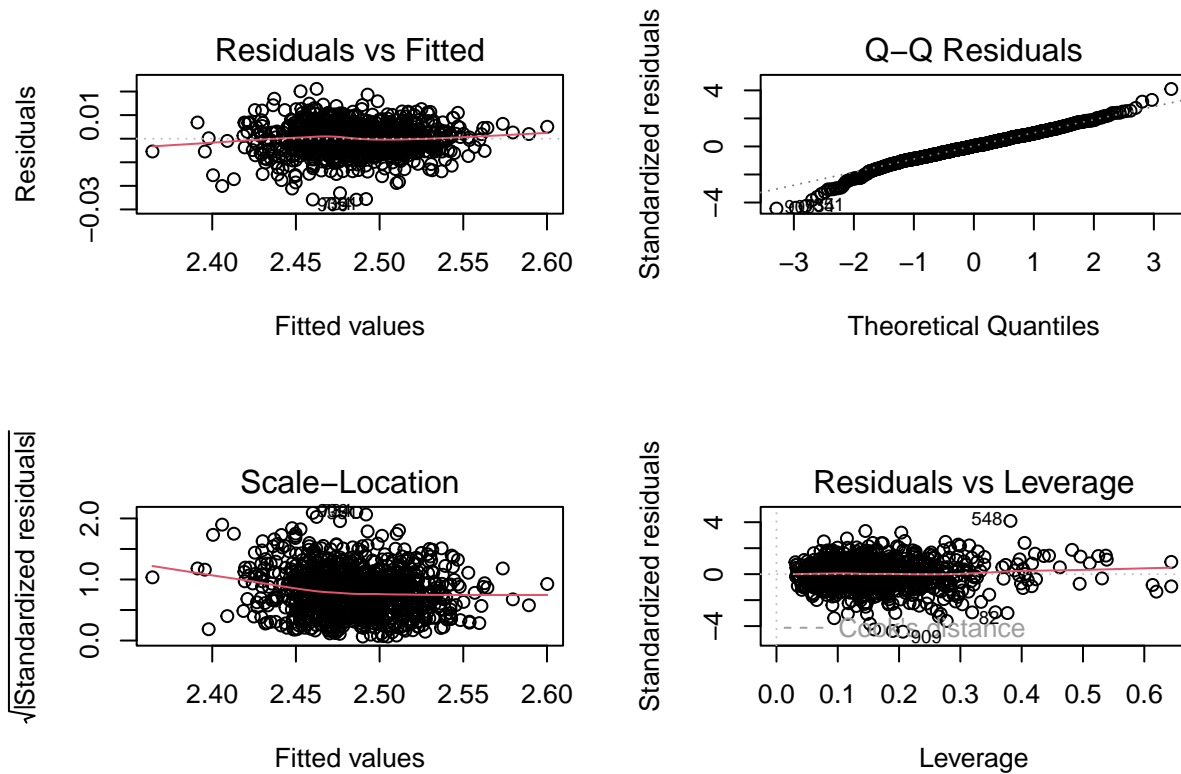


```
par(mfrow=c(1,1))
```

```
qqPlot(ols_fit)
```
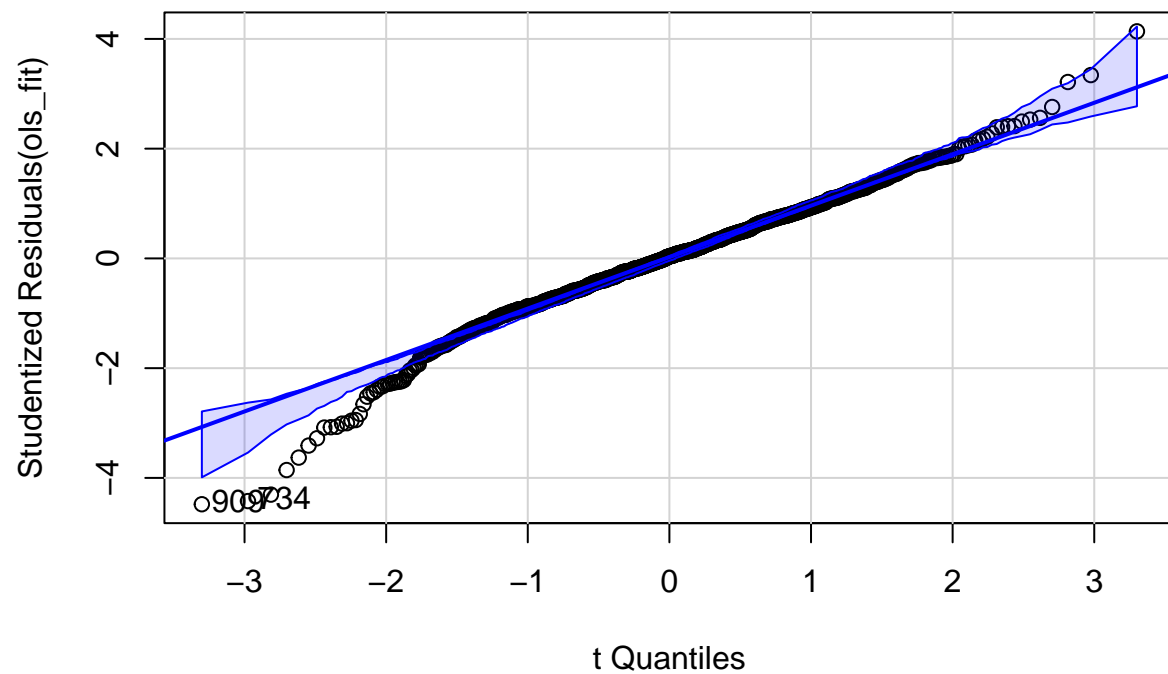
```
## 734 909
## 732 907
```

```
# There is Non Constant error Variance
ncvTest(ols_fit)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 24.51143, Df = 1, p = 7.387e-07
```

```
spreadLevelPlot(ols_fit)
```

Spread–Level Plot for
ols_fit
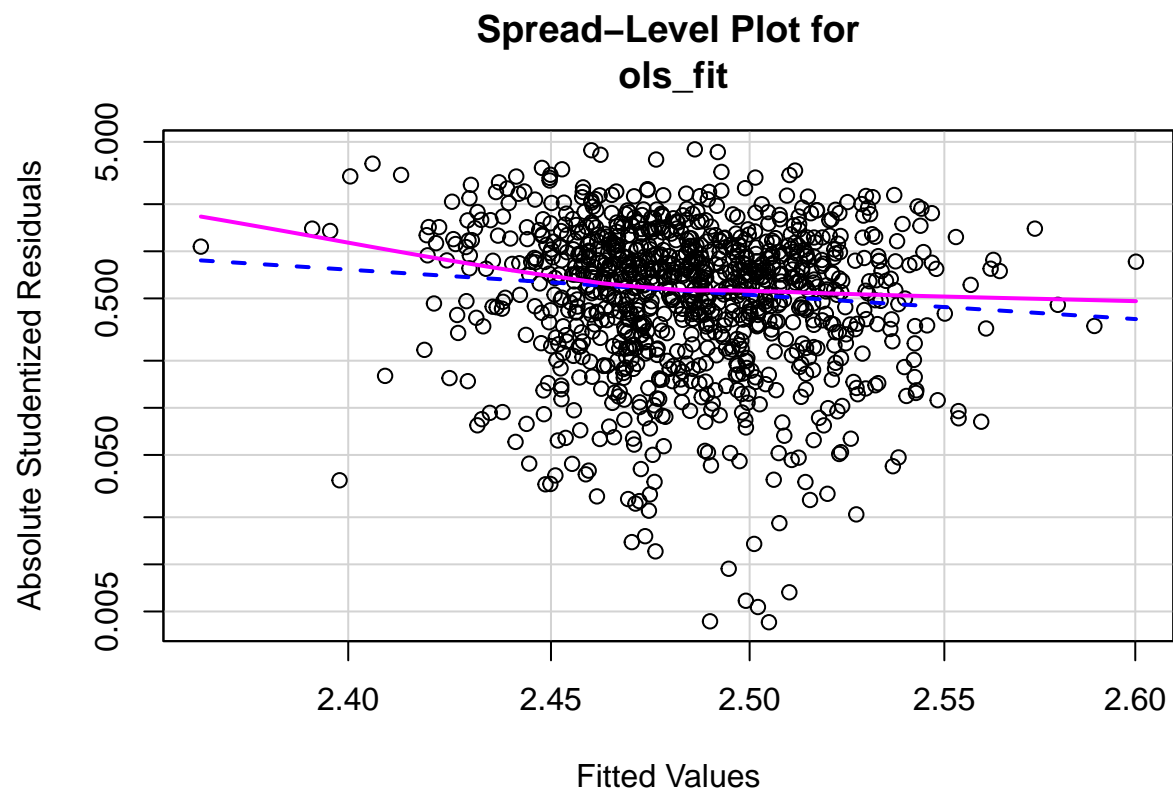

```
##
## Suggested power transformation:  10.07964
```

```
outlierTest(ols_fit)
```

```
##      rstudent unadjusted p-value Bonferroni p
## 909 -4.479838         8.4971e-06    0.0084801
## 734 -4.422021         1.1054e-05    0.0110320
## 351 -4.304796         1.8672e-05    0.0186350
## 548  4.137761         3.8582e-05    0.0385050
```

```
# Many Outliers are present but they don't have leverage on model.
```

1a. OLS MODEL

```r
# OLS with lm
ols_fit <- lm(logSalePrice ~ ., data = housingDataProcessed)
summary(ols_fit)
```

```
##
## Call:
## lm(formula = logSalePrice ~ ., data = housingDataProcessed)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0258614 -0.0035834  0.0002115  0.0040149  0.0211165
##
## Coefficients: (1 not defined because of singularities)
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -1.216e+00  3.107e+00  -0.391 0.695564
## MSSubClass        3.055e-04  1.139e-03   0.268 0.788593
## LotFrontage      -2.262e-04  1.514e-04  -1.494 0.135634
## LotArea           9.787e-03  9.108e-04  10.746  < 2e-16 ***
## OverallQual       3.721e-03  3.382e-04  11.003  < 2e-16 ***
## OverallCond       1.490e-02  1.200e-03  12.420  < 2e-16 ***
## YearBuilt         4.999e-08  2.711e-08   1.844 0.065588 .
## YearRemodAdd     -2.385e-06  2.357e-06  -1.012 0.311968
## MasVnrArea        3.936e-06  1.921e-06   2.049 0.040742 *
## BsmtFinSF1        1.151e-05  1.364e-06   8.437  < 2e-16 ***
## BsmtFinSF2        1.230e-05  3.241e-06   3.795 0.000158 ***
## BsmtUnfSF         7.342e-06  1.182e-06   6.212 8.22e-10 ***
## TotalBsmtSF              NA         NA      NA       NA
## X1stFlrSF         1.269e-02  3.311e-03   3.832 0.000137 ***
## X2ndFlrSF         1.188e-05  2.355e-06   5.043 5.62e-07 ***
## LowQualFinSF      5.290e-06  5.891e-06   0.898 0.369463
## GrLivArea         1.867e-02  3.974e-03   4.698 3.06e-06 ***
## BsmtFullBath      2.025e-03  6.707e-04   3.019 0.002614 **
## BsmtHalfBath      1.279e-03  1.030e-03   1.243 0.214355
## FullBath          1.228e-03  7.942e-04   1.547 0.122320
## HalfBath          2.945e-04  7.414e-04   0.397 0.691310
## BedroomAbvGr     -2.092e-04  4.907e-04  -0.426 0.669917
## KitchenAbvGr     -1.078e-02  6.121e-03  -1.761 0.078530 .
## TotRmsAbvGrd     -1.328e-03  1.238e-03  -1.073 0.283644
## Fireplaces        1.999e-03  4.545e-04   4.397 1.23e-05 ***
## GarageYrBlt       1.413e-07  1.084e-07   1.303 0.192809
## GarageCars        2.505e-03  7.518e-04   3.332 0.000899 ***
## GarageArea        5.248e-06  2.788e-06   1.882 0.060127 .
## WoodDeckSF        4.955e-06  2.015e-06   2.459 0.014120 *
## OpenPorchSF       9.781e-06  3.974e-06   2.461 0.014038 *
## EncPorchSF        1.575e-05  3.001e-06   5.247 1.95e-07 ***
## PoolArea          1.415e-05  8.071e-06   1.754 0.079845 .
## MiscVal           9.465e-07  1.347e-06   0.703 0.482489
## MoSold           -2.967e-05  1.466e-04  -0.202 0.839626
## YrSold            1.010e-03  1.025e-03   0.986 0.324633
## age              -1.249e-03  4.624e-04  -2.702 0.007031 **
## ageSinceRemodel  -4.769e-03  4.661e-03  -1.023 0.306504
## ageofGarage       3.089e-04  2.125e-04   1.453 0.146458
## MSZoningRM       -5.532e-03  1.167e-03  -4.740 2.50e-06 ***
```

```
## MSZoningOther          6.784e-05  1.871e-03   0.036 0.971081
## LotShapeReg            7.082e-04  5.595e-04   1.266 0.205887
## LotShapeOther         -2.018e-03  1.332e-03  -1.515 0.130148
## LandContourHLS         1.569e-03  1.863e-03   0.842 0.400066
## LandContourLow        -9.305e-04  2.163e-03  -0.430 0.667117
## LandContourLvl         1.064e-04  1.285e-03   0.083 0.933998
## LotConfigCulDSac       1.207e-03  1.088e-03   1.109 0.267767
## LotConfigInside       -7.005e-04  6.201e-04  -1.130 0.258882
## LotConfigother        -3.298e-03  1.308e-03  -2.522 0.011848 *
## LandSlopeMod          -1.323e-05  1.390e-03  -0.010 0.992407
## LandSlopeSev          -2.337e-03  3.661e-03  -0.638 0.523505
## NeighborhoodClearCr   -4.597e-03  2.274e-03  -2.022 0.043517 *
## NeighborhoodCollgCr   -7.334e-03  1.855e-03  -3.954 8.33e-05 ***
## NeighborhoodCrawfor    5.529e-03  1.912e-03   2.891 0.003938 **
## NeighborhoodEdwards   -1.081e-02  1.639e-03  -6.594 7.53e-11 ***
## NeighborhoodGilbert   -8.471e-03  2.030e-03  -4.173 3.31e-05 ***
## NeighborhoodIDOTRR    -1.843e-03  2.053e-03  -0.897 0.369751
## NeighborhoodMitchel   -9.749e-03  1.956e-03  -4.985 7.50e-07 ***
## NeighborhoodNAmes     -7.704e-03  1.602e-03  -4.810 1.79e-06 ***
## NeighborhoodNoRidge   -5.712e-03  2.267e-03  -2.520 0.011912 *
## NeighborhoodNridgHt   -3.200e-03  2.278e-03  -1.405 0.160476
## NeighborhoodNWAmes    -8.879e-03  1.910e-03  -4.648 3.89e-06 ***
## NeighborhoodOldTown   -5.122e-03  1.596e-03  -3.210 0.001380 **
## Neighborhoodother     -3.878e-03  1.768e-03  -2.193 0.028573 *
## NeighborhoodSawyer    -7.441e-03  1.792e-03  -4.153 3.61e-05 ***
## NeighborhoodSawyerW   -8.837e-03  1.984e-03  -4.455 9.51e-06 ***
## NeighborhoodSomerst   -2.840e-03  2.682e-03  -1.059 0.289940
## NeighborhoodTimber    -6.973e-03  2.372e-03  -2.940 0.003368 **
## Condition1Feedr        2.429e-03  1.679e-03   1.446 0.148472
## Condition1Norm         5.765e-03  1.409e-03   4.090 4.72e-05 ***
## Condition1PosA         1.405e-03  3.115e-03   0.451 0.652010
## Condition1PosN         4.137e-03  2.372e-03   1.744 0.081466 .
## Condition1RR           2.607e-04  2.023e-03   0.129 0.897508
## BldgType2fmCon        -2.172e-03  2.633e-03  -0.825 0.409694
## BldgTypeDuplex        -3.454e-03  2.541e-03  -1.360 0.174338
## BldgTypeTwnhs         -5.407e-04  2.395e-03  -0.226 0.821446
## BldgTypeTwnhsE        -4.476e-05  2.212e-03  -0.020 0.983860
## HouseStyle1Story       1.898e-03  1.512e-03   1.255 0.209685
## HouseStyle2Story       4.147e-04  1.078e-03   0.385 0.700562
## HouseStyleOther        5.554e-04  1.388e-03   0.400 0.689183
## RoofStyleHip           6.617e-04  6.437e-04   1.028 0.304299
## RoofStyleother         4.568e-03  1.736e-03   2.631 0.008664 **
## Exterior1stCemntBd    -1.168e-02  4.681e-03  -2.496 0.012747 *
## Exterior1stHdBoard    -5.473e-03  2.176e-03  -2.515 0.012082 *
## Exterior1stMetalSd    -3.217e-03  3.060e-03  -1.051 0.293395
## Exterior1stother      -3.171e-03  2.159e-03  -1.469 0.142260
## Exterior1stPlywood    -5.440e-03  2.077e-03  -2.620 0.008959 **
## Exterior1stVinylSd    -7.032e-03  2.985e-03  -2.356 0.018699 *
## Exterior1stWd Sdng    -8.822e-03  1.858e-03  -4.749 2.40e-06 ***
## Exterior2ndCmentBd     7.515e-03  4.915e-03   1.529 0.126631
## Exterior2ndHdBoard     1.049e-04  2.521e-03   0.042 0.966803
## Exterior2ndMetalSd    -3.439e-04  3.351e-03  -0.103 0.918272
## Exterior2ndother       6.556e-04  2.490e-03   0.263 0.792367
## Exterior2ndPlywood    -1.126e-03  2.406e-03  -0.468 0.639815
```

```
## Exterior2ndVinylSd      2.596e-03  3.263e-03   0.796 0.426457
## Exterior2ndWd Sdng      4.401e-03  2.241e-03   1.964 0.049875 *
## Exterior2ndWd Shng     -5.497e-04  2.767e-03  -0.199 0.842574
## MasVnrTypeBrkFace       3.572e-03  2.712e-03   1.317 0.188241
## MasVnrTypeNone          3.920e-03  2.725e-03   1.438 0.150670
## MasVnrTypeStone         3.923e-03  2.884e-03   1.360 0.174042
## ExterQualAvg            4.387e-04  8.150e-04   0.538 0.590489
## ExterQualBelowAvg      -5.054e-03  3.325e-03  -1.520 0.128911
## ExterCondAvg            1.589e-03  8.127e-04   1.955 0.050874 .
## ExterCondBelowAvg       2.029e-03  2.244e-03   0.904 0.366366
## FoundationCBlock        1.903e-03  1.096e-03   1.736 0.082858 .
## Foundationother         1.244e-04  1.897e-03   0.066 0.947723
## FoundationPConc         2.461e-03  1.194e-03   2.061 0.039580 *
## BsmtQualAvg             1.051e-03  8.315e-04   1.264 0.206653
## BsmtQualBelowAvg        2.361e-03  1.869e-03   1.263 0.206866
## BsmtCondAvg            -1.332e-03  1.235e-03  -1.079 0.280824
## BsmtCondBelowAvg       -2.531e-03  1.974e-03  -1.282 0.200225
## BsmtExposureGd          2.196e-03  1.061e-03   2.070 0.038798 *
## BsmtExposureMn         -1.581e-03  1.089e-03  -1.452 0.146950
## BsmtExposureNo         -9.782e-04  7.875e-04  -1.242 0.214524
## BsmtFinType1BLQ        -1.938e-03  8.892e-04  -2.180 0.029553 *
## BsmtFinType1GLQ         6.469e-04  8.308e-04   0.779 0.436407
## BsmtFinType1LwQ        -3.472e-03  1.251e-03  -2.775 0.005642 **
## BsmtFinType1Rec        -1.841e-03  1.004e-03  -1.834 0.067008 .
## BsmtFinType1Unf        -2.061e-03  9.851e-04  -2.092 0.036738 *
## BsmtFinType2BLQ        -3.794e-03  2.764e-03  -1.373 0.170196
## BsmtFinType2GLQ        -9.579e-04  3.045e-03  -0.315 0.753197
## BsmtFinType2LwQ        -4.485e-03  2.746e-03  -1.633 0.102811
## BsmtFinType2Rec        -4.408e-03  2.679e-03  -1.645 0.100327
## BsmtFinType2Unf        -2.750e-03  2.914e-03  -0.944 0.345528
## Heatingother            4.261e-03  1.803e-03   2.363 0.018370 *
## HeatingQCAvg           -1.427e-03  6.244e-04  -2.285 0.022552 *
## HeatingQCBelowAvg      -2.831e-03  1.418e-03  -1.996 0.046209 *
## CentralAirY             3.550e-03  1.363e-03   2.604 0.009373 **
## ElectricalFuseF        -4.526e-03  2.054e-03  -2.204 0.027803 *
## ElectricalFuseP        -3.215e-03  5.530e-03  -0.581 0.561188
## ElectricalSBrkr         6.566e-05  9.878e-04   0.066 0.947016
## KitchenQualAvg         -8.185e-04  7.278e-04  -1.125 0.261062
## KitchenQualBelowAvg    -2.274e-03  1.767e-03  -1.287 0.198512
## FunctionalMaj2         -2.067e-02  4.281e-03  -4.828 1.63e-06 ***
## FunctionalMin1          1.339e-03  2.728e-03   0.491 0.623588
## FunctionalMin2          4.221e-03  2.734e-03   1.544 0.122949
## FunctionalMod           1.415e-03  3.161e-03   0.448 0.654620
## FunctionalTyp           7.037e-03  2.322e-03   3.031 0.002511 **
## FireplaceQuAvg         -1.721e-04  5.964e-04  -0.289 0.772976
## FireplaceQuBelowAvg     8.315e-05  8.838e-04   0.094 0.925062
## GarageTypeAttchd        9.164e-03  4.492e-03   2.040 0.041642 *
## GarageTypeBasment       7.763e-03  5.094e-03   1.524 0.127845
## GarageTypeBuiltIn       9.773e-03  4.658e-03   2.098 0.036205 *
## GarageTypeCarPort       6.929e-03  5.814e-03   1.192 0.233680
## GarageTypeDetchd        1.006e-02  4.448e-03   2.262 0.023944 *
## GarageFinishRFn        -9.215e-04  6.755e-04  -1.364 0.172889
## GarageFinishUnf        -1.078e-03  8.003e-04  -1.347 0.178255
## GarageQualAvg          -4.160e-03  2.892e-03  -1.439 0.150648
```

18

```
## GarageQualBelowAvg   -6.224e-03  3.332e-03   -1.868 0.062090 .
## GarageCondAvg         1.285e-03  3.065e-03    0.419 0.675104
## GarageCondBelowAvg   -3.715e-04  3.468e-03   -0.107 0.914721
## PavedDriveP          -1.195e-03  1.743e-03   -0.686 0.492979
## PavedDriveY           1.807e-03  1.134e-03    1.594 0.111404
## SaleTypeWD            2.634e-06  1.345e-03    0.002 0.998439
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.006552 on 846 degrees of freedom
## Multiple R-squared:  0.9596, Adjusted R-squared:  0.9524
## F-statistic: 133.2 on 151 and 846 DF,  p-value: < 2.2e-16
```

```
AIC(ols_fit)
```

```
## [1] -7062.584
```

```
BIC(ols_fit)
```

```
## [1] -6312.004
```

```
# Stepwise selection based on AIC
stepwise_model <- stepAIC(ols_fit, direction = "both", trace = FALSE)
summary(stepwise_model)
```

```
##
## Call:
## lm(formula = logSalePrice ~ LotArea + OverallQual + OverallCond +
##      YearBuilt + MasVnrArea + BsmtFinSF1 + BsmtFinSF2 + BsmtUnfSF +
##      X1stFlrSF + X2ndFlrSF + GrLivArea + BsmtFullBath + KitchenAbvGr +
##      Fireplaces + GarageYrBlt + GarageCars + GarageArea + WoodDeckSF +
##      OpenPorchSF + EncPorchSF + PoolArea + age + ageSinceRemodel +
##      ageofGarage + MSZoning + LotConfig + Neighborhood + Condition1 +
##      RoofStyle + Exterior1st + ExterQual + BsmtExposure + BsmtFinType1 +
##      Heating + HeatingQC + CentralAir + Functional + GarageQual +
##      PavedDrive, data = housingDataProcessed)
##
## Residuals:
##        Min        1Q    Median        3Q       Max
## -0.0259391 -0.0034657  0.0002856  0.0039441  0.0229499
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.775e+00  1.340e-01  13.243  < 2e-16 ***
## LotArea          8.144e-03  6.783e-04  12.007  < 2e-16 ***
## OverallQual      3.810e-03  3.192e-04  11.935  < 2e-16 ***
## OverallCond      1.576e-02  1.091e-03  14.446  < 2e-16 ***
## YearBuilt        4.260e-08  2.212e-08   1.926 0.054459 .
## MasVnrArea       3.738e-06  1.502e-06   2.488 0.013026 *
## BsmtFinSF1       1.264e-05  1.189e-06  10.627  < 2e-16 ***
## BsmtFinSF2       1.316e-05  1.811e-06   7.265 7.98e-13 ***
## BsmtUnfSF        7.937e-06  9.992e-07   7.943 5.77e-15 ***
```

19

```
## X1stFlrSF              1.297e-02  2.587e-03   5.013 6.44e-07 ***
## X2ndFlrSF              1.141e-05  2.030e-06   5.624 2.48e-08 ***
## GrLivArea              1.716e-02  3.005e-03   5.712 1.51e-08 ***
## BsmtFullBath           1.658e-03  5.950e-04   2.787 0.005432 **
## KitchenAbvGr          -1.724e-02  3.525e-03  -4.890 1.19e-06 ***
## Fireplaces             2.200e-03  4.346e-04   5.063 5.00e-07 ***
## GarageYrBlt            1.348e-07  6.742e-08   1.999 0.045858 *
## GarageCars             2.534e-03  7.137e-04   3.550 0.000405 ***
## GarageArea             4.578e-06  2.595e-06   1.764 0.078075 .
## WoodDeckSF             4.979e-06  1.911e-06   2.605 0.009334 **
## OpenPorchSF            1.175e-05  3.801e-06   3.091 0.002053 **
## EncPorchSF             1.520e-05  2.882e-06   5.275 1.66e-07 ***
## PoolArea               1.424e-05  7.823e-06   1.820 0.069112 .
## age                   -1.506e-03  3.739e-04  -4.028 6.08e-05 ***
## ageSinceRemodel       -5.387e-05  1.756e-05  -3.068 0.002219 **
## ageofGarage            2.794e-04  1.324e-04   2.111 0.035085 *
## MSZoningRM            -5.534e-03  1.054e-03  -5.248 1.91e-07 ***
## MSZoningOther         -7.598e-04  1.768e-03  -0.430 0.667565
## LotConfigCulDSac       1.261e-03  9.862e-04   1.279 0.201326
## LotConfigInside       -2.062e-04  5.809e-04  -0.355 0.722726
## LotConfigother        -2.345e-03  1.235e-03  -1.899 0.057937 .
## NeighborhoodClearCr   -4.260e-03  2.045e-03  -2.083 0.037526 *
## NeighborhoodCollgCr   -7.269e-03  1.666e-03  -4.364 1.43e-05 ***
## NeighborhoodCrawfor    5.303e-03  1.755e-03   3.022 0.002578 **
## NeighborhoodEdwards   -1.023e-02  1.486e-03  -6.885 1.07e-11 ***
## NeighborhoodGilbert   -7.912e-03  1.856e-03  -4.262 2.23e-05 ***
## NeighborhoodIDOTRR    -2.015e-03  1.944e-03  -1.037 0.300176
## NeighborhoodMitchel   -9.333e-03  1.799e-03  -5.187 2.63e-07 ***
## NeighborhoodNAmes     -7.271e-03  1.422e-03  -5.115 3.83e-07 ***
## NeighborhoodNoRidge   -5.658e-03  2.112e-03  -2.679 0.007508 **
## NeighborhoodNridgHt   -3.626e-03  2.029e-03  -1.787 0.074305 .
## NeighborhoodNWAmes    -8.439e-03  1.743e-03  -4.843 1.50e-06 ***
## NeighborhoodOldTown   -4.324e-03  1.488e-03  -2.905 0.003763 **
## Neighborhoodother     -3.417e-03  1.521e-03  -2.247 0.024880 *
## NeighborhoodSawyer    -6.884e-03  1.642e-03  -4.192 3.03e-05 ***
## NeighborhoodSawyerW   -8.360e-03  1.797e-03  -4.651 3.79e-06 ***
## NeighborhoodSomerst   -2.323e-03  2.459e-03  -0.945 0.345047
## NeighborhoodTimber    -6.826e-03  2.162e-03  -3.158 0.001642 **
## Condition1Feedr        3.135e-03  1.606e-03   1.953 0.051155 .
## Condition1Norm         6.297e-03  1.323e-03   4.760 2.24e-06 ***
## Condition1PosA         1.630e-03  2.978e-03   0.547 0.584271
## Condition1PosN         4.379e-03  2.273e-03   1.927 0.054350 .
## Condition1RR           5.099e-04  1.928e-03   0.265 0.791426
## RoofStyleHip           7.493e-04  6.108e-04   1.227 0.220220
## RoofStyleother         5.029e-03  1.609e-03   3.126 0.001827 **
## Exterior1stCemntBd    -6.471e-03  1.706e-03  -3.793 0.000159 ***
## Exterior1stHdBoard    -7.277e-03  1.292e-03  -5.632 2.37e-08 ***
## Exterior1stMetalSd    -5.338e-03  1.284e-03  -4.158 3.51e-05 ***
## Exterior1stother      -4.693e-03  1.485e-03  -3.161 0.001625 **
## Exterior1stPlywood    -8.299e-03  1.420e-03  -5.846 7.02e-09 ***
## Exterior1stVinylSd    -6.669e-03  1.314e-03  -5.074 4.72e-07 ***
## Exterior1stWd Sdng    -7.052e-03  1.272e-03  -5.543 3.89e-08 ***
## ExterQualAvg           1.763e-04  7.420e-04   0.238 0.812258
## ExterQualBelowAvg     -5.941e-03  2.940e-03  -2.021 0.043593 *
```

```
## BsmtExposureGd        2.326e-03  1.008e-03   2.307 0.021285 *
## BsmtExposureMn       -9.636e-04  9.967e-04  -0.967 0.333917
## BsmtExposureNo       -5.628e-04  6.944e-04  -0.810 0.417900
## BsmtFinType1BLQ      -1.969e-03  8.349e-04  -2.359 0.018529 *
## BsmtFinType1GLQ       9.137e-04  7.805e-04   1.171 0.242045
## BsmtFinType1LwQ      -2.526e-03  1.160e-03  -2.178 0.029684 *
## BsmtFinType1Rec      -1.744e-03  9.319e-04  -1.872 0.061546 .
## BsmtFinType1Unf      -1.603e-03  9.315e-04  -1.721 0.085583 .
## Heatingother          2.578e-03  1.640e-03   1.572 0.116330
## HeatingQCAvg         -1.291e-03  5.793e-04  -2.228 0.026112 *
## HeatingQCBelowAvg    -2.630e-03  1.356e-03  -1.940 0.052737 .
## CentralAirY           4.230e-03  1.209e-03   3.499 0.000490 ***
## FunctionalMaj2       -1.884e-02  4.022e-03  -4.684 3.24e-06 ***
## FunctionalMin1        1.953e-03  2.542e-03   0.768 0.442461
## FunctionalMin2        4.400e-03  2.475e-03   1.778 0.075724 .
## FunctionalMod         1.160e-03  2.855e-03   0.406 0.684536
## FunctionalTyp         7.191e-03  2.119e-03   3.394 0.000718 ***
## GarageQualAvg        -5.133e-03  2.672e-03  -1.921 0.055061 .
## GarageQualBelowAvg   -7.561e-03  2.986e-03  -2.533 0.011491 *
## PavedDriveP          -7.546e-04  1.656e-03  -0.456 0.648679
## PavedDriveY           2.048e-03  1.080e-03   1.895 0.058381 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.006562 on 914 degrees of freedom
## Multiple R-squared:  0.9563, Adjusted R-squared:  0.9523
## F-statistic: 240.8 on 83 and 914 DF,  p-value: < 2.2e-16
```

```
AIC(stepwise_model)
```

```
## [1] -7118.51
```

```
BIC(stepwise_model)
```

```
## [1] -6701.521
```

```
vif(stepwise_model)
```

```
##                       GVIF Df GVIF^(1/(2*Df))
## LotArea           2.887969  1        1.699402
## OverallQual       4.045604  1        2.011369
## OverallCond       2.099621  1        1.449007
## YearBuilt        36.944206  1        6.078175
## MasVnrArea        1.640761  1        1.280922
## BsmtFinSF1        5.399695  1        2.323724
## BsmtFinSF2        1.718106  1        1.310765
## BsmtUnfSF         4.039486  1        2.009847
## X1stFlrSF        14.045550  1        3.747739
## X2ndFlrSF        17.355818  1        4.166031
## GrLivArea        22.358425  1        4.728470
## BsmtFullBath      2.123952  1        1.457378
## KitchenAbvGr      1.574074  1        1.254621
```

```
## Fireplaces         1.805127   1      1.343550
## GarageYrBlt      238.633269   1     15.447759
## GarageCars         6.017547   1      2.453069
## GarageArea         6.102967   1      2.470418
## WoodDeckSF         1.368430   1      1.169799
## OpenPorchSF        1.283666   1      1.132990
## EncPorchSF         1.292650   1      1.136948
## PoolArea           1.066237   1      1.032588
## age               46.359573   1      6.808786
## ageSinceRemodel    2.863608   1      1.692220
## ageofGarage      234.367559   1     15.309068
## MSZoning          10.405731   2      1.796049
## LotConfig          1.464079   3      1.065600
## Neighborhood     762.298913  17      1.215538
## Condition1         2.136733   5      1.078885
## RoofStyle          1.596048   2      1.123988
## Exterior1st        9.785575   7      1.176945
## ExterQual          3.631490   2      1.380452
## BsmtExposure       2.123585   3      1.133735
## BsmtFinType1       8.205982   5      1.234278
## Heating            1.581648   1      1.257636
## HeatingQC          2.175085   2      1.214420
## CentralAir         2.033699   1      1.426078
## Functional         2.162783   5      1.080193
## GarageQual         1.550572   2      1.115894
## PavedDrive         1.879470   2      1.170871
```

```r
# Linear regression after variable selection with stepAIC
ols2 = lm(formula = logSalePrice ~ LotArea +
    OverallCond + MasVnrArea + BsmtFinSF2 + BsmtFullBath +
    KitchenAbvGr + Fireplaces + GarageArea +
    WoodDeckSF + OpenPorchSF + EncPorchSF + PoolArea +
    ageSinceRemodel + MSZoning + LotConfig + Neighborhood +
    Condition1 + RoofStyle + Exterior1st + ExterQual + BsmtExposure +
    BsmtFinType1 + Heating + HeatingQC + CentralAir + Functional +
    GarageQual + PavedDrive,
    data = housingDataProcessed)
summary(ols2)
```

```
##
## Call:
## lm(formula = logSalePrice ~ LotArea + OverallCond + MasVnrArea +
##     BsmtFinSF2 + BsmtFullBath + KitchenAbvGr + Fireplaces + GarageArea +
##     WoodDeckSF + OpenPorchSF + EncPorchSF + PoolArea + ageSinceRemodel +
##     MSZoning + LotConfig + Neighborhood + Condition1 + RoofStyle +
##     Exterior1st + ExterQual + BsmtExposure + BsmtFinType1 + Heating +
##     HeatingQC + CentralAir + Functional + GarageQual + PavedDrive,
##     data = housingDataProcessed)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.040979 -0.007496  0.000295  0.006967  0.039450
##
## Coefficients:
```

```
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)            2.321e+00  1.312e-02 176.863  < 2e-16 ***
## LotArea                1.277e-02  1.152e-03  11.087  < 2e-16 ***
## OverallCond            8.771e-03  1.769e-03   4.957 8.53e-07 ***
## MasVnrArea             1.690e-05  2.637e-06   6.409 2.33e-10 ***
## BsmtFinSF2             2.228e-06  2.907e-06   0.766 0.443618
## BsmtFullBath           1.278e-03  9.487e-04   1.347 0.178163
## KitchenAbvGr           1.101e-02  5.915e-03   1.860 0.063138 .
## Fireplaces             7.373e-03  7.427e-04   9.927  < 2e-16 ***
## GarageArea             2.971e-05  2.579e-06  11.520  < 2e-16 ***
## WoodDeckSF             1.479e-05  3.431e-06   4.310 1.81e-05 ***
## OpenPorchSF            3.755e-05  6.807e-06   5.516 4.50e-08 ***
## EncPorchSF             2.366e-05  5.145e-06   4.599 4.84e-06 ***
## PoolArea               1.557e-05  1.421e-05   1.096 0.273315
## ageSinceRemodel       -2.180e-04  2.951e-05  -7.388 3.32e-13 ***
## MSZoningRM            -7.780e-03  1.897e-03  -4.101 4.48e-05 ***
## MSZoningOther         -2.672e-03  3.187e-03  -0.838 0.402022
## LotConfigCulDSac       1.950e-04  1.781e-03   0.109 0.912857
## LotConfigInside       -7.831e-04  1.053e-03  -0.744 0.457109
## LotConfigother        -2.837e-03  2.230e-03  -1.272 0.203536
## NeighborhoodClearCr    1.236e-03  3.596e-03   0.344 0.731066
## NeighborhoodCollgCr    1.137e-03  2.810e-03   0.405 0.685738
## NeighborhoodCrawfor    1.439e-02  3.161e-03   4.552 6.02e-06 ***
## NeighborhoodEdwards   -8.337e-03  2.622e-03  -3.180 0.001522 **
## NeighborhoodGilbert    3.668e-03  3.004e-03   1.221 0.222393
## NeighborhoodIDOTRR     5.066e-04  3.477e-03   0.146 0.884181
## NeighborhoodMitchel   -5.858e-03  3.145e-03  -1.862 0.062861 .
## NeighborhoodNAmes     -5.318e-03  2.504e-03  -2.123 0.033978 *
## NeighborhoodNoRidge    1.344e-02  3.592e-03   3.742 0.000194 ***
## NeighborhoodNridgHt    1.155e-02  3.445e-03   3.352 0.000835 ***
## NeighborhoodNWAmes     2.811e-03  3.021e-03   0.930 0.352504
## NeighborhoodOldTown   -2.107e-03  2.674e-03  -0.788 0.430815
## Neighborhoodother      7.315e-03  2.670e-03   2.740 0.006271 **
## NeighborhoodSawyer    -6.603e-03  2.895e-03  -2.281 0.022777 *
## NeighborhoodSawyerW    6.041e-04  3.078e-03   0.196 0.844464
## NeighborhoodSomerst    1.043e-02  4.206e-03   2.481 0.013273 *
## NeighborhoodTimber     6.147e-03  3.725e-03   1.650 0.099249 .
## Condition1Feedr        7.501e-04  2.906e-03   0.258 0.796351
## Condition1Norm         5.325e-03  2.392e-03   2.226 0.026273 *
## Condition1PosA         7.631e-03  5.378e-03   1.419 0.156255
## Condition1PosN         8.766e-03  4.101e-03   2.137 0.032831 *
## Condition1RR          -3.421e-04  3.485e-03  -0.098 0.921834
## RoofStyleHip           2.835e-03  1.069e-03   2.652 0.008148 **
## RoofStyleother         4.715e-03  2.890e-03   1.631 0.103159
## Exterior1stCemntBd    -6.507e-03  3.046e-03  -2.136 0.032929 *
## Exterior1stHdBoard    -9.419e-03  2.313e-03  -4.072 5.07e-05 ***
## Exterior1stMetalSd    -8.021e-03  2.306e-03  -3.479 0.000527 ***
## Exterior1stother      -7.748e-03  2.666e-03  -2.907 0.003742 **
## Exterior1stPlywood    -1.053e-02  2.568e-03  -4.100 4.49e-05 ***
## Exterior1stVinylSd    -7.476e-03  2.312e-03  -3.234 0.001264 **
## Exterior1stWd Sdng    -1.109e-02  2.279e-03  -4.865 1.35e-06 ***
## ExterQualAvg          -9.085e-03  1.247e-03  -7.284 6.94e-13 ***
## ExterQualBelowAvg     -3.245e-02  5.152e-03  -6.299 4.62e-10 ***
## BsmtExposureGd         6.160e-03  1.797e-03   3.427 0.000637 ***
```

```
## BsmtExposureMn          1.007e-03  1.797e-03   0.560 0.575489
## BsmtExposureNo          7.746e-04  1.232e-03   0.629 0.529688
## BsmtFinType1BLQ        -2.057e-03  1.510e-03  -1.362 0.173491
## BsmtFinType1GLQ         5.558e-03  1.382e-03   4.023 6.23e-05 ***
## BsmtFinType1LwQ         1.842e-03  2.081e-03   0.885 0.376332
## BsmtFinType1Rec         9.683e-04  1.681e-03   0.576 0.564668
## BsmtFinType1Unf         1.064e-03  1.434e-03   0.742 0.458483
## Heatingother            5.559e-03  2.966e-03   1.874 0.061249 .
## HeatingQCAvg           -1.876e-03  1.046e-03  -1.793 0.073231 .
## HeatingQCBelowAvg      -5.295e-03  2.448e-03  -2.163 0.030810 *
## CentralAirY             1.216e-02  2.127e-03   5.719 1.45e-08 ***
## FunctionalMaj2         -4.390e-02  7.162e-03  -6.130 1.30e-09 ***
## FunctionalMin1         -1.290e-02  4.525e-03  -2.851 0.004459 **
## FunctionalMin2         -7.688e-03  4.392e-03  -1.751 0.080360 .
## FunctionalMod          -4.639e-03  5.152e-03  -0.900 0.368155
## FunctionalTyp          -6.390e-03  3.732e-03  -1.712 0.087229 .
## GarageQualAvg           6.158e-03  4.756e-03   1.295 0.195741
## GarageQualBelowAvg      3.399e-03  5.232e-03   0.650 0.516086
## PavedDriveP             3.395e-03  2.983e-03   1.138 0.255309
## PavedDriveY             7.648e-03  1.905e-03   4.016 6.41e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01194 on 925 degrees of freedom
## Multiple R-squared:  0.8536, Adjusted R-squared:  0.8422
## F-statistic: 74.88 on 72 and 925 DF,  p-value: < 2.2e-16
```

`vif(ols2)`

```
##                     GVIF Df GVIF^(1/(2*Df))
## LotArea         2.517501  1        1.586663
## OverallCond     1.668742  1        1.291798
## MasVnrArea      1.527229  1        1.235811
## BsmtFinSF2      1.337952  1        1.156699
## BsmtFullBath    1.631980  1        1.277490
## KitchenAbvGr    1.339710  1        1.157458
## Fireplaces      1.593614  1        1.262384
## GarageArea      1.821166  1        1.349506
## WoodDeckSF      1.333082  1        1.154592
## OpenPorchSF     1.244155  1        1.115417
## EncPorchSF      1.244915  1        1.115758
## PoolArea        1.063111  1        1.031073
## ageSinceRemodel 2.444251  1        1.563410
## MSZoning        9.990296  2        1.777848
## LotConfig       1.406028  3        1.058439
## Neighborhood  265.253964 17        1.178377
## Condition1      1.987959  5        1.071126
## RoofStyle       1.439301  2        1.095312
## Exterior1st     6.833167  7        1.147139
## ExterQual       2.957533  2        1.311392
## BsmtExposure    1.859777  3        1.108945
## BsmtFinType1    4.619024  5        1.165346
## Heating         1.564197  1        1.250679
## HeatingQC       2.114455  2        1.205868
```

```
## CentralAir     1.901653  1        1.379004
## Functional     1.732230  5        1.056478
## GarageQual     1.355215  2        1.078952
## PavedDrive     1.729115  2        1.146716
```

```
par(mfrow=c(2,2))
plot(ols2)
```
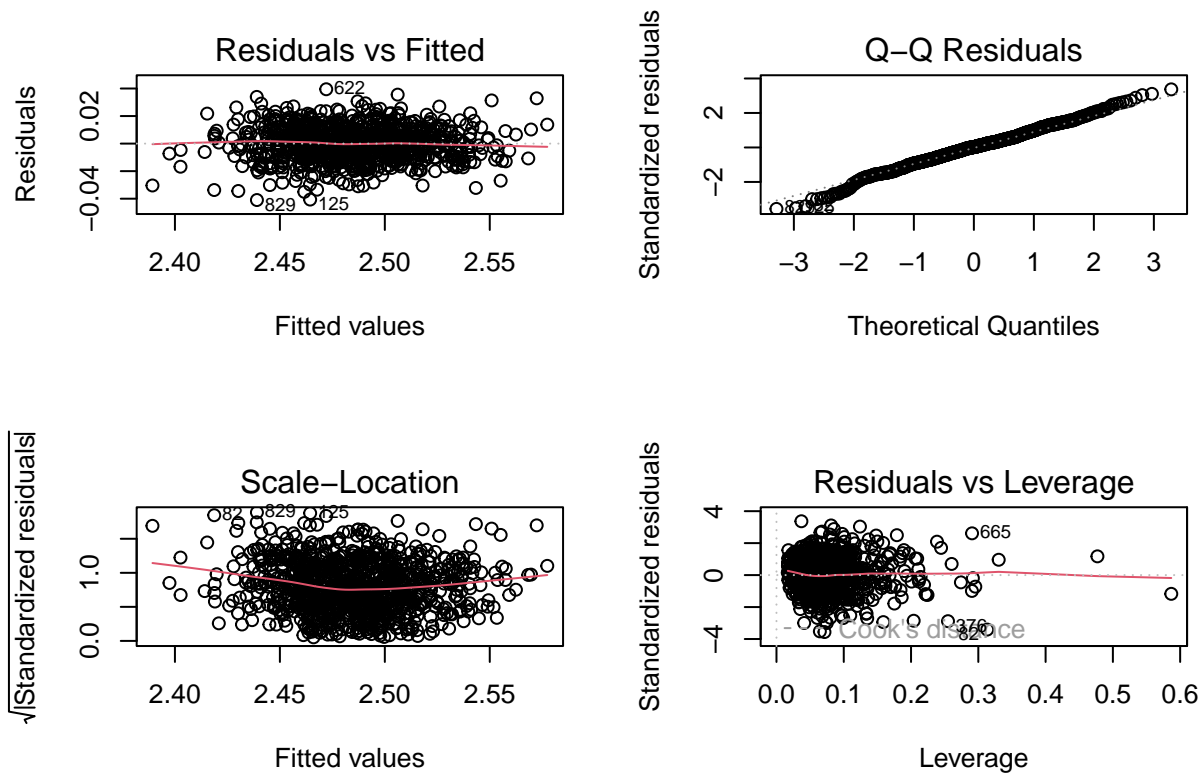


```
par(mfrow=c(1,1))
# Non Constant Error variance is Present
ncvTest(ols2)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 8.788312, Df = 1, p = 0.0030317
```

```
# Build the linear model with interaction terms included
ols_interaction <- lm(formula = logSalePrice ~ LotArea +
                    OverallCond * ExterQual + MasVnrArea + BsmtFinSF2 + BsmtFullBath +
                    KitchenAbvGr + Fireplaces + GarageArea * GarageQual +
                    WoodDeckSF + OpenPorchSF + EncPorchSF + PoolArea +
                    ageSinceRemodel + MSZoning + LotConfig + Neighborhood * GrLivArea +
                    Condition1 + RoofStyle + Exterior1st + BsmtExposure +
                    BsmtFinType1 + Heating + HeatingQC + CentralAir + Functional +
```

```
                        PavedDrive, data = housingDataProcessed)
summary(ols_interaction)
```

```
##
## Call:
## lm(formula = logSalePrice ~ LotArea + OverallCond * ExterQual +
##     MasVnrArea + BsmtFinSF2 + BsmtFullBath + KitchenAbvGr + Fireplaces +
##     GarageArea * GarageQual + WoodDeckSF + OpenPorchSF + EncPorchSF +
##     PoolArea + ageSinceRemodel + MSZoning + LotConfig + Neighborhood *
##     GrLivArea + Condition1 + RoofStyle + Exterior1st + BsmtExposure +
##     BsmtFinType1 + Heating + HeatingQC + CentralAir + Functional +
##     PavedDrive, data = housingDataProcessed)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.032909 -0.004332  0.000252  0.005060  0.023821
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             1.997e+00  3.594e-02  55.566  < 2e-16 ***
## LotArea                 7.585e-03  8.798e-04   8.621  < 2e-16 ***
## OverallCond            -1.088e-03  2.758e-03  -0.394 0.693340
## ExterQualAvg           -3.241e-02  6.168e-03  -5.254 1.86e-07 ***
## ExterQualBelowAvg      -1.587e-02  1.667e-02  -0.952 0.341111
## MasVnrArea              6.930e-06  1.968e-06   3.522 0.000451 ***
## BsmtFinSF2              4.146e-06  2.080e-06   1.994 0.046465 *
## BsmtFullBath            2.681e-03  6.872e-04   3.901 0.000103 ***
## KitchenAbvGr           -2.583e-02  4.571e-03  -5.650 2.15e-08 ***
## Fireplaces              3.366e-03  5.598e-04   6.012 2.66e-09 ***
## GarageArea              3.200e-05  2.192e-05   1.460 0.144561
## GarageQualAvg           8.986e-03  1.432e-02   0.627 0.530569
## GarageQualBelowAvg      5.563e-03  1.461e-02   0.381 0.703495
## WoodDeckSF              8.691e-06  2.477e-06   3.509 0.000473 ***
## OpenPorchSF             1.858e-05  5.059e-06   3.672 0.000255 ***
## EncPorchSF              1.420e-05  3.717e-06   3.819 0.000143 ***
## PoolArea                1.014e-05  1.017e-05   0.997 0.318983
## ageSinceRemodel        -1.570e-04  2.137e-05  -7.346 4.55e-13 ***
## MSZoningRM             -6.070e-03  1.415e-03  -4.290 1.98e-05 ***
## MSZoningOther          -2.488e-03  2.317e-03  -1.074 0.283123
## LotConfigCulDSac        1.255e-03  1.275e-03   0.984 0.325353
## LotConfigInside        -1.192e-03  7.630e-04  -1.562 0.118592
## LotConfigother         -3.325e-03  1.603e-03  -2.073 0.038410 *
## NeighborhoodClearCr     2.158e-01  6.216e-02   3.471 0.000542 ***
## NeighborhoodCollgCr     1.161e-01  3.741e-02   3.103 0.001978 **
## NeighborhoodCrawfor     1.114e-01  4.919e-02   2.265 0.023720 *
## NeighborhoodEdwards     7.097e-02  4.030e-02   1.761 0.078537 .
## NeighborhoodGilbert     2.241e-01  5.720e-02   3.917 9.63e-05 ***
## NeighborhoodIDOTRR      1.268e-01  8.070e-02   1.571 0.116558
## NeighborhoodMitchel     1.867e-01  4.724e-02   3.953 8.32e-05 ***
## NeighborhoodNAmes       1.470e-01  3.578e-02   4.107 4.37e-05 ***
## NeighborhoodNoRidge    -4.673e-02  6.716e-02  -0.696 0.486750
## NeighborhoodNridgHt     4.619e-02  6.033e-02   0.766 0.444076
## NeighborhoodNWAmes      1.852e-01  5.332e-02   3.473 0.000539 ***
```

```
## NeighborhoodOldTown              1.041e-01  3.731e-02   2.792 0.005354 **
## Neighborhoodother                8.865e-02  3.786e-02   2.341 0.019429 *
## NeighborhoodSawyer               2.012e-01  4.752e-02   4.235 2.52e-05 ***
## NeighborhoodSawyerW              1.562e-01  4.303e-02   3.630 0.000299 ***
## NeighborhoodSomerst              1.306e-01  6.831e-02   1.912 0.056174 .
## NeighborhoodTimber               1.172e-01  7.085e-02   1.655 0.098307 .
## GrLivArea                        5.462e-02  4.479e-03  12.194  < 2e-16 ***
## Condition1Feedr                  7.406e-04  2.087e-03   0.355 0.722765
## Condition1Norm                   5.913e-03  1.732e-03   3.414 0.000669 ***
## Condition1PosA                   4.623e-03  3.894e-03   1.187 0.235382
## Condition1PosN                   4.855e-03  2.973e-03   1.633 0.102854
## Condition1RR                     2.046e-03  2.527e-03   0.809 0.418511
## RoofStyleHip                     3.285e-03  7.724e-04   4.253 2.33e-05 ***
## RoofStyleother                   1.465e-03  2.143e-03   0.684 0.494396
## Exterior1stCemntBd              -5.454e-03  2.207e-03  -2.472 0.013630 *
## Exterior1stHdBoard              -7.852e-03  1.662e-03  -4.726 2.66e-06 ***
## Exterior1stMetalSd              -7.493e-03  1.668e-03  -4.492 7.96e-06 ***
## Exterior1stother                -7.083e-03  1.919e-03  -3.691 0.000237 ***
## Exterior1stPlywood              -6.783e-03  1.843e-03  -3.681 0.000246 ***
## Exterior1stVinylSd              -6.045e-03  1.672e-03  -3.616 0.000316 ***
## Exterior1stWd Sdng              -1.007e-02  1.642e-03  -6.130 1.31e-09 ***
## BsmtExposureGd                   5.162e-03  1.301e-03   3.968 7.82e-05 ***
## BsmtExposureMn                  -1.949e-03  1.301e-03  -1.498 0.134567
## BsmtExposureNo                  -2.186e-03  9.029e-04  -2.422 0.015643 *
## BsmtFinType1BLQ                 -2.266e-03  1.080e-03  -2.098 0.036204 *
## BsmtFinType1GLQ                  3.364e-03  9.933e-04   3.386 0.000739 ***
## BsmtFinType1LwQ                 -2.763e-03  1.511e-03  -1.828 0.067878 .
## BsmtFinType1Rec                 -8.302e-04  1.216e-03  -0.683 0.494920
## BsmtFinType1Unf                 -2.203e-03  1.043e-03  -2.112 0.034996 *
## Heatingother                     2.788e-03  2.154e-03   1.294 0.195944
## HeatingQCAvg                    -1.343e-03  7.545e-04  -1.780 0.075443 .
## HeatingQCBelowAvg               -2.952e-03  1.773e-03  -1.665 0.096293 .
## CentralAirY                      7.753e-03  1.550e-03   5.001 6.84e-07 ***
## FunctionalMaj2                  -2.424e-02  5.192e-03  -4.669 3.48e-06 ***
## FunctionalMin1                  -3.316e-03  3.300e-03  -1.005 0.315181
## FunctionalMin2                  -9.604e-04  3.194e-03  -0.301 0.763711
## FunctionalMod                    1.086e-03  3.761e-03   0.289 0.772760
## FunctionalTyp                    7.151e-03  2.767e-03   2.584 0.009909 **
## PavedDriveP                     -2.283e-05  2.164e-03  -0.011 0.991585
## PavedDriveY                      5.829e-03  1.372e-03   4.248 2.38e-05 ***
## OverallCond:ExterQualAvg         1.239e-02  2.913e-03   4.254 2.32e-05 ***
## OverallCond:ExterQualBelowAvg   -2.315e-04  8.955e-03  -0.026 0.979381
## GarageArea:GarageQualAvg        -1.471e-05  2.189e-05  -0.672 0.501875
## GarageArea:GarageQualBelowAvg   -1.438e-05  2.409e-05  -0.597 0.550694
## NeighborhoodClearCr:GrLivArea   -2.991e-02  8.436e-03  -3.545 0.000412 ***
## NeighborhoodCollgCr:GrLivArea   -1.642e-02  5.292e-03  -3.103 0.001976 **
## NeighborhoodCrawfor:GrLivArea   -1.459e-02  6.751e-03  -2.160 0.031002 *
## NeighborhoodEdwards:GrLivArea   -1.128e-02  5.694e-03  -1.982 0.047808 *
## NeighborhoodGilbert:GrLivArea   -3.112e-02  7.860e-03  -3.960 8.10e-05 ***
## NeighborhoodIDOTRR:GrLivArea    -1.846e-02  1.139e-02  -1.621 0.105367
## NeighborhoodMitchel:GrLivArea   -2.676e-02  6.672e-03  -4.012 6.53e-05 ***
## NeighborhoodNAmes:GrLivArea     -2.122e-02  5.063e-03  -4.191 3.05e-05 ***
## NeighborhoodNoRidge:GrLivArea    4.828e-03  8.841e-03   0.546 0.585144
## NeighborhoodNridgHt:GrLivArea   -5.882e-03  8.215e-03  -0.716 0.474131
```

```
## NeighborhoodNWAmes:GrLivArea  -2.592e-02  7.313e-03  -3.545 0.000413 ***
## NeighborhoodOldTown:GrLivArea -1.538e-02  5.253e-03  -2.927 0.003505 **
## Neighborhoodother:GrLivArea   -1.246e-02  5.327e-03  -2.339 0.019558 *
## NeighborhoodSawyer:GrLivArea  -2.897e-02  6.740e-03  -4.299 1.90e-05 ***
## NeighborhoodSawyerW:GrLivArea -2.227e-02  6.012e-03  -3.705 0.000224 ***
## NeighborhoodSomerst:GrLivArea -1.744e-02  9.335e-03  -1.868 0.062039 .
## NeighborhoodTimber:GrLivArea  -1.606e-02  9.655e-03  -1.664 0.096482 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.008479 on 903 degrees of freedom
## Multiple R-squared:  0.9278, Adjusted R-squared:  0.9203
## F-statistic: 123.5 on 94 and 903 DF,  p-value: < 2.2e-16
```

```r
vif(ols_interaction, type = "predictor")
```

```
## GVIFs computed for predictors

##                          GVIF Df GVIF^(1/(2*Df)) Interacts With
## LotArea            2.909279  1        1.705661            --
## OverallCond        9.344269  5        1.250416      ExterQual
## ExterQual          9.344269  5        1.250416    OverallCond
## MasVnrArea         1.685608  1        1.298310            --
## BsmtFinSF2         1.356489  1        1.164684            --
## BsmtFullBath       1.696622  1        1.302545            --
## KitchenAbvGr       1.585304  1        1.259088            --
## Fireplaces         1.793850  1        1.339347            --
## GarageArea         4.488382  5        1.162008      GarageQual
## GarageQual         4.488382  5        1.162008      GarageArea
## WoodDeckSF         1.376732  1        1.173342            --
## OpenPorchSF        1.362081  1        1.167082            --
## EncPorchSF         1.287794  1        1.134810            --
## PoolArea           1.079569  1        1.039023            --
## ageSinceRemodel    2.539422  1        1.593556            --
## MSZoning          11.555979  2        1.843749            --
## LotConfig          1.525870  3        1.072967            --
## Neighborhood    5957.619835 35        1.132217       GrLivArea
## GrLivArea       5957.619835 35        1.132217    Neighborhood
## Condition1         2.428394  5        1.092778            --
## RoofStyle          1.622281  2        1.128578            --
## Exterior1st        8.881867  7        1.168827            --
## BsmtExposure       2.133315  3        1.134599            --
## BsmtFinType1       5.404703  5        1.183797            --
## Heating            1.634780  1        1.278585            --
## HeatingQC          2.286785  2        1.229720            --
## CentralAir         2.001792  1        1.414847            --
## Functional         2.182070  5        1.081152            --
## PavedDrive         1.860027  2        1.167830            --
##
## LotArea         OverallCond, ExterQual, MasVnrArea, BsmtFinSF2, BsmtFullBath, KitchenAbvGr, Fireplace
## OverallCond              LotArea, MasVnrArea, BsmtFinSF2, BsmtFullBath, KitchenAbvGr, Fireplace
## ExterQual                LotArea, MasVnrArea, BsmtFinSF2, BsmtFullBath, KitchenAbvGr, Fireplace
## MasVnrArea       LotArea, OverallCond, ExterQual, BsmtFinSF2, BsmtFullBath, KitchenAbvGr, Fireplace
```

```
## BsmtFinSF2        LotArea, OverallCond, ExterQual, MasVnrArea, BsmtFullBath, KitchenAbvGr, Fireplace
## BsmtFullBath      LotArea, OverallCond, ExterQual, MasVnrArea, BsmtFinSF2, KitchenAbvGr, Fireplace
## KitchenAbvGr      LotArea, OverallCond, ExterQual, MasVnrArea, BsmtFinSF2, BsmtFullBath, Fireplace
## Fireplaces        LotArea, OverallCond, ExterQual, MasVnrArea, BsmtFinSF2, BsmtFullBath, KitchenAbv(
## GarageArea             LotArea, OverallCond, ExterQual, MasVnrArea, BsmtFinSF2, BsmtFullBath
## GarageQual             LotArea, OverallCond, ExterQual, MasVnrArea, BsmtFinSF2, BsmtFullBath
## WoodDeckSF        LotArea, OverallCond, ExterQual, MasVnrArea, BsmtFinSF2, BsmtFullBath, KitchenAbv(
## OpenPorchSF       LotArea, OverallCond, ExterQual, MasVnrArea, BsmtFinSF2, BsmtFullBath, KitchenAbv
## EncPorchSF        LotArea, OverallCond, ExterQual, MasVnrArea, BsmtFinSF2, BsmtFullBath, KitchenAbv
## PoolArea        LotArea, OverallCond, ExterQual, MasVnrArea, BsmtFinSF2, BsmtFullBath, KitchenAbvGr
## ageSinceRemodel        LotArea, OverallCond, ExterQual, MasVnrArea, BsmtFinSF2, BsmtFullBath, Kitch
## MSZoning        LotArea, OverallCond, ExterQual, MasVnrArea, BsmtFinSF2, BsmtFullBath, KitchenAbvGr
## LotConfig       LotArea, OverallCond, ExterQual, MasVnrArea, BsmtFinSF2, BsmtFullBath, KitchenAbv
## Neighborhood          LotArea, OverallCond, ExterQual, MasVnrArea, BsmtFinSF2, BsmtFullBat
## GrLivArea             LotArea, OverallCond, ExterQual, MasVnrArea, BsmtFinSF2, BsmtFullBat
## Condition1      LotArea, OverallCond, ExterQual, MasVnrArea, BsmtFinSF2, BsmtFullBath, KitchenAbv
## RoofStyle       LotArea, OverallCond, ExterQual, MasVnrArea, BsmtFinSF2, BsmtFullBath, KitchenAbv
## Exterior1st     LotArea, OverallCond, ExterQual, MasVnrArea, BsmtFinSF2, BsmtFullBath, KitchenAbv
## BsmtExposure      LotArea, OverallCond, ExterQual, MasVnrArea, BsmtFinSF2, BsmtFullBath, KitchenAl
## BsmtFinType1      LotArea, OverallCond, ExterQual, MasVnrArea, BsmtFinSF2, BsmtFullBath, KitchenAl
## Heating       LotArea, OverallCond, ExterQual, MasVnrArea, BsmtFinSF2, BsmtFullBath, KitchenAbvGr,
## HeatingQC       LotArea, OverallCond, ExterQual, MasVnrArea, BsmtFinSF2, BsmtFullBath, KitchenAbv
## CentralAir      LotArea, OverallCond, ExterQual, MasVnrArea, BsmtFinSF2, BsmtFullBath, KitchenAbv(
## Functional      LotArea, OverallCond, ExterQual, MasVnrArea, BsmtFinSF2, BsmtFullBath, KitchenAbv(
## PavedDrive      LotArea, OverallCond, ExterQual, MasVnrArea, BsmtFinSF2, BsmtFullBath, KitchenAbv(
```

```r
# Perform 5-fold cross-validation using caret
cv_control <- trainControl(method = "cv", number = 10)
ols_cv_interaction_model <- train(logSalePrice ~ LotArea +
                  OverallCond * ExterQual + MasVnrArea + BsmtFinSF2 + BsmtFullBath +
                  KitchenAbvGr + Fireplaces + GarageArea * GarageQual +
                  WoodDeckSF + OpenPorchSF + EncPorchSF + PoolArea +
                  ageSinceRemodel + MSZoning + LotConfig + Neighborhood * GrLivArea +
                  Condition1 + RoofStyle + Exterior1st + BsmtExposure +
                  BsmtFinType1 + Heating + HeatingQC + CentralAir + Functional + PavedDrive,
                  data = housingDataProcessed,
                  method = "lm",
                  trControl = cv_control)
```

```
## Warning in predict.lm(modelFit, newdata): prediction from rank-deficient fit;
## attr(*, "non-estim") has doubtful cases
```

```r
# Display cross-validation results
ols_cv_interaction_model
```

```
## Linear Regression
##
## 998 samples
##   29 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 898, 898, 899, 899, 898, 898, ...
```

```
## Resampling results:
##
##   RMSE         Rsquared    MAE
##   0.009089974  0.9097846   0.006928345
##
## Tuning parameter 'intercept' was held constant at a value of TRUE
```

```
summary(ols_cv_interaction_model)
```

```
##
## Call:
## lm(formula = .outcome ~ ., data = dat)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.032909 -0.004332  0.000252  0.005060  0.023821
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)               1.997e+00  3.594e-02  55.566  < 2e-16 ***
## LotArea                   7.585e-03  8.798e-04   8.621  < 2e-16 ***
## OverallCond              -1.088e-03  2.758e-03  -0.394 0.693340
## ExterQualAvg             -3.241e-02  6.168e-03  -5.254 1.86e-07 ***
## ExterQualBelowAvg        -1.587e-02  1.667e-02  -0.952 0.341111
## MasVnrArea                6.930e-06  1.968e-06   3.522 0.000451 ***
## BsmtFinSF2                4.146e-06  2.080e-06   1.994 0.046465 *
## BsmtFullBath              2.681e-03  6.872e-04   3.901 0.000103 ***
## KitchenAbvGr             -2.583e-02  4.571e-03  -5.650 2.15e-08 ***
## Fireplaces                3.366e-03  5.598e-04   6.012 2.66e-09 ***
## GarageArea                3.200e-05  2.192e-05   1.460 0.144561
## GarageQualAvg             8.986e-03  1.432e-02   0.627 0.530569
## GarageQualBelowAvg        5.563e-03  1.461e-02   0.381 0.703495
## WoodDeckSF                8.691e-06  2.477e-06   3.509 0.000473 ***
## OpenPorchSF               1.858e-05  5.059e-06   3.672 0.000255 ***
## EncPorchSF                1.420e-05  3.717e-06   3.819 0.000143 ***
## PoolArea                  1.014e-05  1.017e-05   0.997 0.318983
## ageSinceRemodel          -1.570e-04  2.137e-05  -7.346 4.55e-13 ***
## MSZoningRM               -6.070e-03  1.415e-03  -4.290 1.98e-05 ***
## MSZoningOther            -2.488e-03  2.317e-03  -1.074 0.283123
## LotConfigCulDSac          1.255e-03  1.275e-03   0.984 0.325353
## LotConfigInside          -1.192e-03  7.630e-04  -1.562 0.118592
## LotConfigother           -3.325e-03  1.603e-03  -2.073 0.038410 *
## NeighborhoodClearCr       2.158e-01  6.216e-02   3.471 0.000542 ***
## NeighborhoodCollgCr       1.161e-01  3.741e-02   3.103 0.001978 **
## NeighborhoodCrawfor       1.114e-01  4.919e-02   2.265 0.023720 *
## NeighborhoodEdwards       7.097e-02  4.030e-02   1.761 0.078537 .
## NeighborhoodGilbert       2.241e-01  5.720e-02   3.917 9.63e-05 ***
## NeighborhoodIDOTRR        1.268e-01  8.070e-02   1.571 0.116558
## NeighborhoodMitchel       1.867e-01  4.724e-02   3.953 8.32e-05 ***
## NeighborhoodNAmes         1.470e-01  3.578e-02   4.107 4.37e-05 ***
## NeighborhoodNoRidge      -4.673e-02  6.716e-02  -0.696 0.486750
## NeighborhoodNridgHt       4.619e-02  6.033e-02   0.766 0.444076
## NeighborhoodNWAmes        1.852e-01  5.332e-02   3.473 0.000539 ***
## NeighborhoodOldTown       1.041e-01  3.731e-02   2.792 0.005354 **
```

```
## Neighborhoodother                       8.865e-02  3.786e-02   2.341 0.019429 *
## NeighborhoodSawyer                       2.012e-01  4.752e-02   4.235 2.52e-05 ***
## NeighborhoodSawyerW                      1.562e-01  4.303e-02   3.630 0.000299 ***
## NeighborhoodSomerst                      1.306e-01  6.831e-02   1.912 0.056174 .
## NeighborhoodTimber                       1.172e-01  7.085e-02   1.655 0.098307 .
## GrLivArea                                5.462e-02  4.479e-03  12.194  < 2e-16 ***
## Condition1Feedr                          7.406e-04  2.087e-03   0.355 0.722765
## Condition1Norm                           5.913e-03  1.732e-03   3.414 0.000669 ***
## Condition1PosA                           4.623e-03  3.894e-03   1.187 0.235382
## Condition1PosN                           4.855e-03  2.973e-03   1.633 0.102854
## Condition1RR                             2.046e-03  2.527e-03   0.809 0.418511
## RoofStyleHip                             3.285e-03  7.724e-04   4.253 2.33e-05 ***
## RoofStyleother                          1.465e-03  2.143e-03   0.684 0.494396
## Exterior1stCemntBd                      -5.454e-03  2.207e-03  -2.472 0.013630 *
## Exterior1stHdBoard                      -7.852e-03  1.662e-03  -4.726 2.66e-06 ***
## Exterior1stMetalSd                      -7.493e-03  1.668e-03  -4.492 7.96e-06 ***
## Exterior1stother                        -7.083e-03  1.919e-03  -3.691 0.000237 ***
## Exterior1stPlywood                      -6.783e-03  1.843e-03  -3.681 0.000246 ***
## Exterior1stVinylSd                      -6.045e-03  1.672e-03  -3.616 0.000316 ***
## `Exterior1stWd Sdng`                    -1.007e-02  1.642e-03  -6.130 1.31e-09 ***
## BsmtExposureGd                           5.162e-03  1.301e-03   3.968 7.82e-05 ***
## BsmtExposureMn                          -1.949e-03  1.301e-03  -1.498 0.134567
## BsmtExposureNo                          -2.186e-03  9.029e-04  -2.422 0.015643 *
## BsmtFinType1BLQ                         -2.266e-03  1.080e-03  -2.098 0.036204 *
## BsmtFinType1GLQ                          3.364e-03  9.933e-04   3.386 0.000739 ***
## BsmtFinType1LwQ                         -2.763e-03  1.511e-03  -1.828 0.067878 .
## BsmtFinType1Rec                         -8.302e-04  1.216e-03  -0.683 0.494920
## BsmtFinType1Unf                         -2.203e-03  1.043e-03  -2.112 0.034996 *
## Heatingother                             2.788e-03  2.154e-03   1.294 0.195944
## HeatingQCAvg                            -1.343e-03  7.545e-04  -1.780 0.075443 .
## HeatingQCBelowAvg                       -2.952e-03  1.773e-03  -1.665 0.096293 .
## CentralAirY                              7.753e-03  1.550e-03   5.001 6.84e-07 ***
## FunctionalMaj2                          -2.424e-02  5.192e-03  -4.669 3.48e-06 ***
## FunctionalMin1                          -3.316e-03  3.300e-03  -1.005 0.315181
## FunctionalMin2                          -9.604e-04  3.194e-03  -0.301 0.763711
## FunctionalMod                            1.086e-03  3.761e-03   0.289 0.772760
## FunctionalTyp                            7.151e-03  2.767e-03   2.584 0.009909 **
## PavedDriveP                             -2.283e-05  2.164e-03  -0.011 0.991585
## PavedDriveY                              5.829e-03  1.372e-03   4.248 2.38e-05 ***
## `OverallCond:ExterQualAvg`               1.239e-02  2.913e-03   4.254 2.32e-05 ***
## `OverallCond:ExterQualBelowAvg` -2.315e-04  8.955e-03  -0.026 0.979381
## `GarageArea:GarageQualAvg`              -1.471e-05  2.189e-05  -0.672 0.501875
## `GarageArea:GarageQualBelowAvg` -1.438e-05  2.409e-05  -0.597 0.550694
## `NeighborhoodClearCr:GrLivArea` -2.991e-02  8.436e-03  -3.545 0.000412 ***
## `NeighborhoodCollgCr:GrLivArea` -1.642e-02  5.292e-03  -3.103 0.001976 **
## `NeighborhoodCrawfor:GrLivArea` -1.459e-02  6.751e-03  -2.160 0.031002 *
## `NeighborhoodEdwards:GrLivArea` -1.128e-02  5.694e-03  -1.982 0.047808 *
## `NeighborhoodGilbert:GrLivArea` -3.112e-02  7.860e-03  -3.960 8.10e-05 ***
## `NeighborhoodIDOTRR:GrLivArea`  -1.846e-02  1.139e-02  -1.621 0.105367
## `NeighborhoodMitchel:GrLivArea` -2.676e-02  6.672e-03  -4.012 6.53e-05 ***
## `NeighborhoodNAmes:GrLivArea`   -2.122e-02  5.063e-03  -4.191 3.05e-05 ***
## `NeighborhoodNoRidge:GrLivArea`  4.828e-03  8.841e-03   0.546 0.585144
## `NeighborhoodNridgHt:GrLivArea` -5.882e-03  8.215e-03  -0.716 0.474131
## `NeighborhoodNWAmes:GrLivArea`  -2.592e-02  7.313e-03  -3.545 0.000413 ***
```

```
## `NeighborhoodOldTown:GrLivArea` -1.538e-02  5.253e-03  -2.927 0.003505 **
## `Neighborhoodother:GrLivArea`   -1.246e-02  5.327e-03  -2.339 0.019558 *
## `NeighborhoodSawyer:GrLivArea`  -2.897e-02  6.740e-03  -4.299 1.90e-05 ***
## `NeighborhoodSawyerW:GrLivArea` -2.227e-02  6.012e-03  -3.705 0.000224 ***
## `NeighborhoodSomerst:GrLivArea` -1.744e-02  9.335e-03  -1.868 0.062039 .
## `NeighborhoodTimber:GrLivArea`  -1.606e-02  9.655e-03  -1.664 0.096482 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.008479 on 903 degrees of freedom
## Multiple R-squared:  0.9278, Adjusted R-squared:  0.9203
## F-statistic: 123.5 on 94 and 903 DF,  p-value: < 2.2e-16
```

```r
ols_cv_interaction_rmse <- ols_cv_interaction_model$results$RMSE
ols_cv_interaction_r2 <- ols_cv_interaction_model$results$Rsquared

cv_results <- ols_cv_interaction_model$resample

# Plot RMSE for each fold
ggplot(cv_results, aes(x = Resample, y = RMSE)) +
  geom_point(color = "blue", size = 3) +
  geom_line(group = 1, color = "blue") +
  labs(title = "Cross-Validation RMSE for OLS Model", x = "Fold", y = "RMSE") +
  theme_minimal()
```
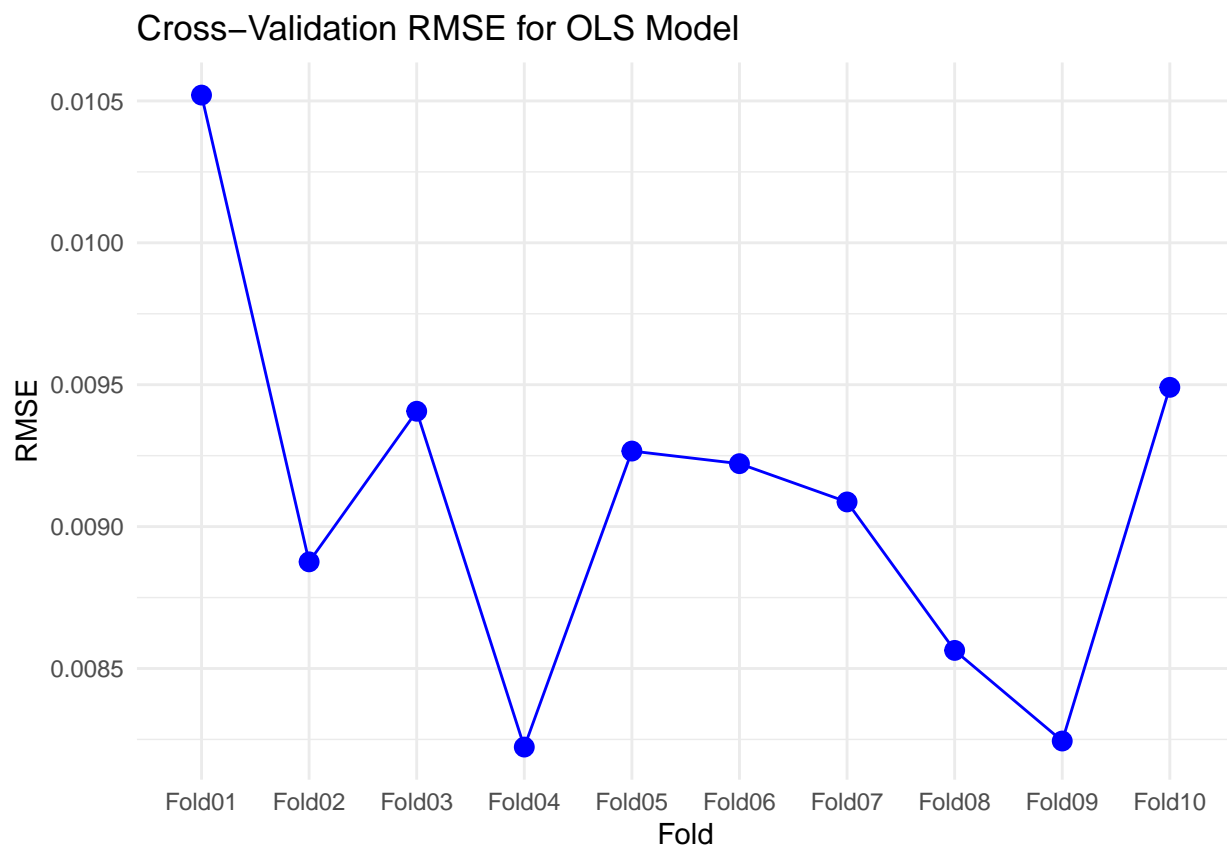
i

The variables, the coefficient estimates, p-values, AIC, BIC, and VIF for the model are shown above.

5 fold cross validation was choosing. Different cross validations like 10,15,20 were tried but since we have less data , lower folds is better.

The Cross Validated rmse and r-square are reported in the summary table below.

ii

The model has Non constant Error Variance. Meaning the assumption of linear regression is not met. Tranformations were tried and it were no avail. Its better to use non-linear models.

There were outliers but they were pf low leverage. Only two outliers were removed.

1b. PLS Model

```r
# PLS
cv_control <- trainControl(method = "cv", number = 5)

tune_grid <- expand.grid(ncomp = seq(25,35, length = 50))

# Train PLS model using caret
pls_model <- train(logSalePrice ~ .,
                   data = housingDataProcessed,
                   method = "pls",
                   trControl = cv_control,
                   tuneGrid = tune_grid
                   )
pls_model
```

```
## Partial Least Squares
##
## 998 samples
##  71 predictor
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 800, 798, 798, 799, 797
## Resampling results across tuning parameters:
##
##   ncomp     RMSE         Rsquared   MAE
##   25.00000  0.007862956  0.9309716  0.005846899
##   25.20408  0.007862956  0.9309716  0.005846899
##   25.40816  0.007862956  0.9309716  0.005846899
##   25.61224  0.007862956  0.9309716  0.005846899
##   25.81633  0.007862956  0.9309716  0.005846899
```

```
##    26.02041  0.007796578  0.9320843  0.005798977
##    26.22449  0.007796578  0.9320843  0.005798977
##    26.42857  0.007796578  0.9320843  0.005798977
##    26.63265  0.007796578  0.9320843  0.005798977
##    26.83673  0.007796578  0.9320843  0.005798977
##    27.04082  0.007759373  0.9326297  0.005762128
##    27.24490  0.007759373  0.9326297  0.005762128
##    27.44898  0.007759373  0.9326297  0.005762128
##    27.65306  0.007759373  0.9326297  0.005762128
##    27.85714  0.007759373  0.9326297  0.005762128
##    28.06122  0.007787499  0.9321348  0.005807988
##    28.26531  0.007787499  0.9321348  0.005807988
##    28.46939  0.007787499  0.9321348  0.005807988
##    28.67347  0.007787499  0.9321348  0.005807988
##    28.87755  0.007787499  0.9321348  0.005807988
##    29.08163  0.007786878  0.9322805  0.005831875
##    29.28571  0.007786878  0.9322805  0.005831875
##    29.48980  0.007786878  0.9322805  0.005831875
##    29.69388  0.007786878  0.9322805  0.005831875
##    29.89796  0.007786878  0.9322805  0.005831875
##    30.10204  0.007814511  0.9318538  0.005865570
##    30.30612  0.007814511  0.9318538  0.005865570
##    30.51020  0.007814511  0.9318538  0.005865570
##    30.71429  0.007814511  0.9318538  0.005865570
##    30.91837  0.007814511  0.9318538  0.005865570
##    31.12245  0.007840446  0.9314295  0.005877551
##    31.32653  0.007840446  0.9314295  0.005877551
##    31.53061  0.007840446  0.9314295  0.005877551
##    31.73469  0.007840446  0.9314295  0.005877551
##    31.93878  0.007840446  0.9314295  0.005877551
##    32.14286  0.007854419  0.9313128  0.005899078
##    32.34694  0.007854419  0.9313128  0.005899078
##    32.55102  0.007854419  0.9313128  0.005899078
##    32.75510  0.007854419  0.9313128  0.005899078
##    32.95918  0.007854419  0.9313128  0.005899078
##    33.16327  0.007876723  0.9310551  0.005897562
##    33.36735  0.007876723  0.9310551  0.005897562
##    33.57143  0.007876723  0.9310551  0.005897562
##    33.77551  0.007876723  0.9310551  0.005897562
##    33.97959  0.007876723  0.9310551  0.005897562
##    34.18367  0.007886526  0.9309626  0.005902107
##    34.38776  0.007886526  0.9309626  0.005902107
##    34.59184  0.007886526  0.9309626  0.005902107
##    34.79592  0.007886526  0.9309626  0.005902107
##    35.00000  0.007908797  0.9304839  0.005915283
##
## RMSE was used to select the optimal model using the smallest value.
## The final value used for the model was ncomp = 27.04082.
```
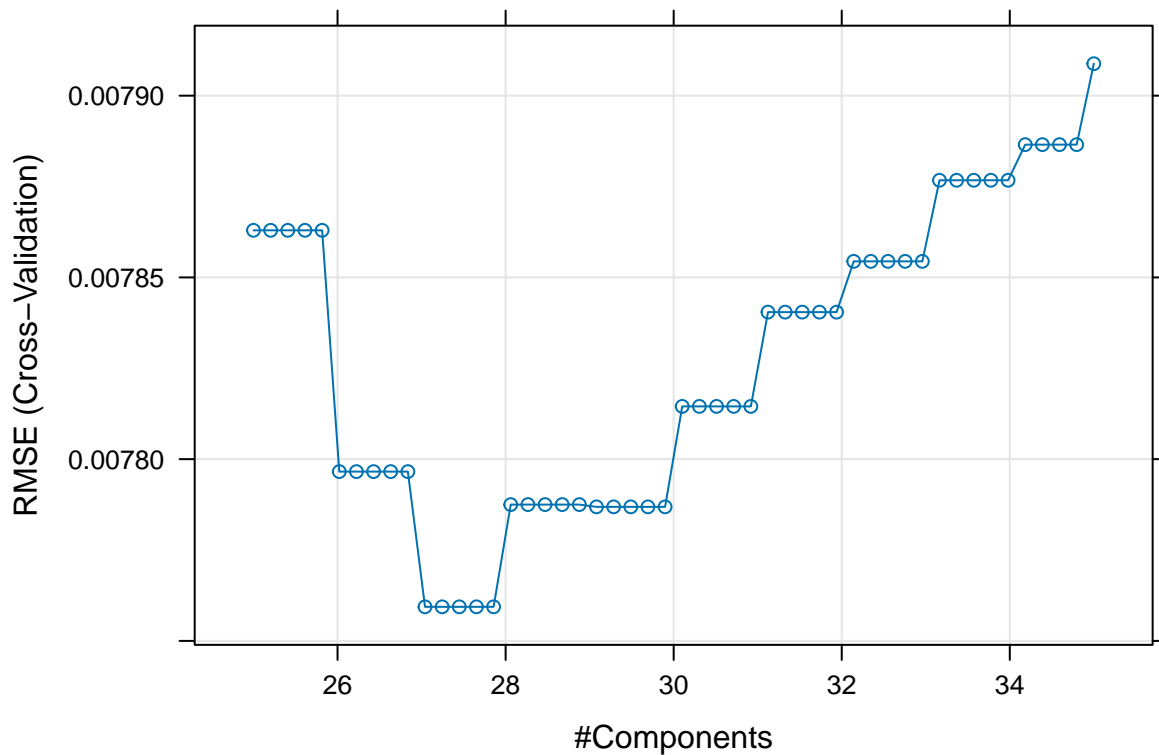
```
summary(pls_model)
```

```
## Data:     X dimension: 998 152
##  Y dimension: 998 1
## Fit method: oscorespls
```

```
## Number of components considered: 27.04082
## TRAINING: % variance explained
##              1 comps   2 comps   3 comps   4 comps   5 comps   6 comps   7 comps
## X             80.71     93.47     99.99     99.99     99.99    100.00    100.00
## .outcome      37.24     38.53     38.66     77.83     79.43     81.15     81.28
##              8 comps   9 comps   10 comps  11 comps  12 comps  13 comps  14 comps
## X            100.00    100.00    100.00    100.00    100.00    100.00    100.00
## .outcome      81.51     81.67     81.69     81.72     81.73     81.74     81.78
##              15 comps  16 comps  17 comps  18 comps  19 comps  20 comps  21 comps
## X            100.00    100.00    100.00    100.00    100.0     100.00    100.00
## .outcome      81.88     84.12     85.53     89.96     91.3      92.19     93.54
##              22 comps  23 comps  24 comps  25 comps  26 comps  27 comps
## X            100.0     100.00    100.00    100.00    100.0     100.0
## .outcome      94.3      94.68     95.06     95.34     95.5      95.6
```

```r
plot(pls_model)
```



```r
pls_rmse <- pls_model$results$RMSE
pls_r2 <- pls_model$results$Rsquared

# Print the number of components chosen
chosen_components <- pls_model$bestTune$ncomp
cat("Number of components chosen:", chosen_components)
```

```
## Number of components chosen: 27.04082
```

```
# Print the chosen RMSE value
chosen_rmse <- pls_model$results$RMSE[pls_model$results$ncomp == chosen_components]
cat("Chosen RMSE:", chosen_rmse)
```

Chosen RMSE: 0.007759373

1c. LASSO Model

```
# LASSO

cv_control <- trainControl(method = "cv", number = 5)

# Define the grid of hyperparameters (lambda values) to test
tune_grid <- expand.grid(alpha = 1,  # LASSO regression is when alpha = 1
                         lambda = seq(0.0001, 0.0004, length = 100))  # Range of lambda values

# Train LASSO model using caret and glmnet
lasso_model <- train(
  logSalePrice ~ .,
  data = housingDataProcessed,
  method = "glmnet",
  trControl = cv_control,
  tuneGrid = tune_grid
)

lasso_model
```

```
## glmnet
##
## 998 samples
##  71 predictor
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 798, 797, 800, 799, 798
## Resampling results across tuning parameters:
##
##    lambda        RMSE         Rsquared     MAE
##    0.0001000000  0.007476398  0.9384725    0.005679405
##    0.0001030303  0.007476235  0.9384802    0.005679619
##    0.0001060606  0.007476233  0.9384854    0.005679950
##    0.0001090909  0.007476498  0.9384862    0.005680371
##    0.0001121212  0.007476597  0.9384902    0.005680464
##    0.0001151515  0.007476719  0.9384939    0.005680410
##    0.0001181818  0.007477080  0.9384937    0.005680567
##    0.0001212121  0.007477488  0.9384928    0.005680957
##    0.0001242424  0.007477599  0.9384966    0.005681175
##    0.0001272727  0.007477905  0.9384972    0.005681436
```

```
##    0.0001303030   0.007478416   0.9384944   0.005681750
##    0.0001333333   0.007478778   0.9384942   0.005681970
##    0.0001363636   0.007478761   0.9385009   0.005681885
##    0.0001393939   0.007478927   0.9385045   0.005681872
##    0.0001424242   0.007479280   0.9385050   0.005682018
##    0.0001454545   0.007479731   0.9385039   0.005682244
##    0.0001484848   0.007479412   0.9385160   0.005681655
##    0.0001515152   0.007479205   0.9385262   0.005681182
##    0.0001545455   0.007479163   0.9385337   0.005680849
##    0.0001575758   0.007479287   0.9385386   0.005680657
##    0.0001606061   0.007479309   0.9385456   0.005680386
##    0.0001636364   0.007479068   0.9385574   0.005679940
##    0.0001666667   0.007478952   0.9385671   0.005679516
##    0.0001696970   0.007478983   0.9385744   0.005679168
##    0.0001727273   0.007479161   0.9385795   0.005679132
##    0.0001757576   0.007479604   0.9385804   0.005679273
##    0.0001787879   0.007480203   0.9385795   0.005679511
##    0.0001818182   0.007480913   0.9385769   0.005679808
##    0.0001848485   0.007481755   0.9385723   0.005680172
##    0.0001878788   0.007482730   0.9385655   0.005680562
##    0.0001909091   0.007483837   0.9385566   0.005680971
##    0.0001939394   0.007484755   0.9385510   0.005681228
##    0.0001969697   0.007485373   0.9385502   0.005681321
##    0.0002000000   0.007486099   0.9385478   0.005681528
##    0.0002030303   0.007486935   0.9385435   0.005681736
##    0.0002060606   0.007487883   0.9385375   0.005681943
##    0.0002090909   0.007488941   0.9385298   0.005682175
##    0.0002121212   0.007489565   0.9385298   0.005682160
##    0.0002151515   0.007489899   0.9385353   0.005682018
##    0.0002181818   0.007490312   0.9385397   0.005682022
##    0.0002212121   0.007490815   0.9385427   0.005682080
##    0.0002242424   0.007491409   0.9385443   0.005682199
##    0.0002272727   0.007492093   0.9385445   0.005682476
##    0.0002303030   0.007492869   0.9385432   0.005682895
##    0.0002333333   0.007493697   0.9385409   0.005683139
##    0.0002363636   0.007494555   0.9385382   0.005683384
##    0.0002393939   0.007495488   0.9385343   0.005683650
##    0.0002424242   0.007496502   0.9385292   0.005684078
##    0.0002454545   0.007497597   0.9385228   0.005684650
##    0.0002484848   0.007498774   0.9385151   0.005685318
##    0.0002515152   0.007500032   0.9385062   0.005686040
##    0.0002545455   0.007501335   0.9384963   0.005686713
##    0.0002575758   0.007502764   0.9384841   0.005687390
##    0.0002606061   0.007504301   0.9384701   0.005688122
##    0.0002636364   0.007505911   0.9384550   0.005688855
##    0.0002666667   0.007507594   0.9384387   0.005689600
##    0.0002696970   0.007509349   0.9384212   0.005690607
##    0.0002727273   0.007511176   0.9384026   0.005691666
##    0.0002757576   0.007513075   0.9383828   0.005692730
##    0.0002787879   0.007515037   0.9383620   0.005693831
##    0.0002818182   0.007517143   0.9383386   0.005695155
##    0.0002848485   0.007519296   0.9383143   0.005696489
##    0.0002878788   0.007521510   0.9382890   0.005697823
##    0.0002909091   0.007523785   0.9382627   0.005699217
```
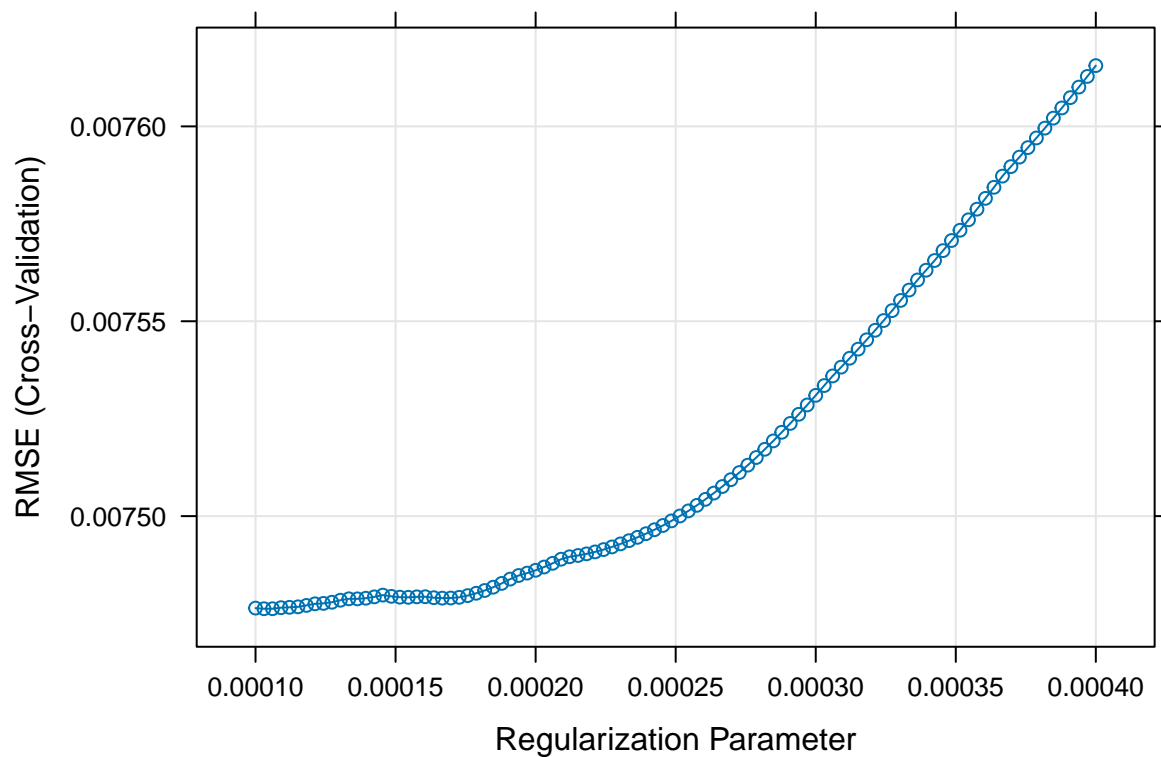
```
##    0.0002939394  0.007526121  0.9382355  0.005700688
##    0.0002969697  0.007528519  0.9382073  0.005702165
##    0.0003000000  0.007530977  0.9381780  0.005703643
##    0.0003030303  0.007533497  0.9381478  0.005705161
##    0.0003060606  0.007535966  0.9381184  0.005706657
##    0.0003090909  0.007538228  0.9380920  0.005708087
##    0.0003121212  0.007540511  0.9380654  0.005709608
##    0.0003151515  0.007542848  0.9380378  0.005711429
##    0.0003181818  0.007545239  0.9380094  0.005713355
##    0.0003212121  0.007547685  0.9379802  0.005715300
##    0.0003242424  0.007550184  0.9379500  0.005717246
##    0.0003272727  0.007552737  0.9379190  0.005719191
##    0.0003303030  0.007555344  0.9378872  0.005721136
##    0.0003333333  0.007558004  0.9378544  0.005723081
##    0.0003363636  0.007560602  0.9378224  0.005724885
##    0.0003393939  0.007563090  0.9377918  0.005726550
##    0.0003424242  0.007565583  0.9377612  0.005728187
##    0.0003454545  0.007568125  0.9377298  0.005729850
##    0.0003484848  0.007570713  0.9376977  0.005731637
##    0.0003515152  0.007573350  0.9376648  0.005733444
##    0.0003545455  0.007576033  0.9376311  0.005735328
##    0.0003575758  0.007578765  0.9375966  0.005737240
##    0.0003606061  0.007581544  0.9375614  0.005739152
##    0.0003636364  0.007584370  0.9375254  0.005741093
##    0.0003666667  0.007587244  0.9374886  0.005743044
##    0.0003696970  0.007589681  0.9374590  0.005744722
##    0.0003727273  0.007592118  0.9374293  0.005746406
##    0.0003757576  0.007594554  0.9373998  0.005748126
##    0.0003787879  0.007597033  0.9373695  0.005749856
##    0.0003818182  0.007599555  0.9373386  0.005751594
##    0.0003848485  0.007602121  0.9373069  0.005753378
##    0.0003878788  0.007604730  0.9372746  0.005755198
##    0.0003909091  0.007607383  0.9372415  0.005757042
##    0.0003939394  0.007610079  0.9372077  0.005758969
##    0.0003969697  0.007612818  0.9371733  0.005761045
##    0.0004000000  0.007615601  0.9371381  0.005763133
##
## 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 = 0.0001060606.
```

```r
summary(lasso_model)
```

```
##            Length Class      Mode
## a0            97  -none-     numeric
## beta       14744  dgCMatrix  S4
## df            97  -none-     numeric
## dim            2  -none-     numeric
## lambda        97  -none-     numeric
## dev.ratio     97  -none-     numeric
## nulldev        1  -none-     numeric
## npasses        1  -none-     numeric
## jerr           1  -none-     numeric
## offset         1  -none-     logical
```
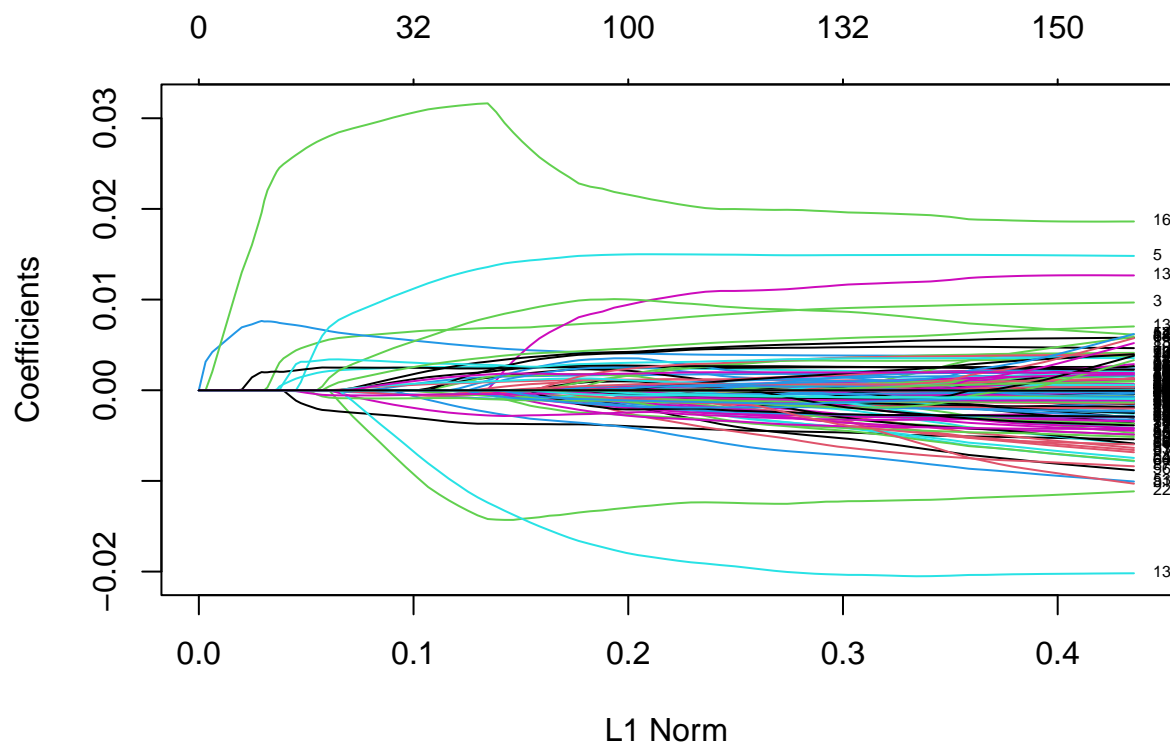
```
## call          5  -none-    call
## nobs          1  -none-    numeric
## lambdaOpt     1  -none-    numeric
## xNames      152  -none-    character
## problemType   1  -none-    character
## tuneValue     2  data.frame list
## obsLevels     1  -none-    logical
## param         0  -none-    list
```

```r
plot(lasso_model)
```



```r
# Visualize coefficient shrinkage as lambda increases
plot(lasso_model$finalModel, label = TRUE)
```

```r
lasso_rmse <- lasso_model$results$RMSE
lasso_r2 <- lasso_model$results$Rsquared

# Extract coefficients at the best lambda value
best_lambda <- lasso_model$bestTune$lambda
lasso_coefficients <- coef(lasso_model$finalModel, s = best_lambda)

# Convert the coefficient matrix to a data frame for easy viewing
lasso_coeff_df <- as.data.frame(as.matrix(lasso_coefficients))
lasso_coeff_df$variable <- rownames(lasso_coeff_df)
colnames(lasso_coeff_df)[1] <- "coefficient"

# Filter out variables with zero coefficients
non_zero_coeff <- subset(lasso_coeff_df, coefficient != 0)

# Display non-zero coefficients and their variable names
print(non_zero_coeff)
```

```
##                  coefficient            variable
## (Intercept)     1.973916e+00         (Intercept)
## LotFrontage    -2.036235e-06         LotFrontage
## LotArea         7.775962e-03             LotArea
## OverallQual     4.004853e-03         OverallQual
## OverallCond     1.499834e-02         OverallCond
## YearBuilt       4.524142e-08           YearBuilt
```

```
## YearRemodAdd         2.345532e-08            YearRemodAdd
## MasVnrArea            2.991958e-06              MasVnrArea
## BsmtFinSF1            4.819566e-06              BsmtFinSF1
## BsmtFinSF2            3.714053e-06              BsmtFinSF2
## TotalBsmtSF           7.700559e-06             TotalBsmtSF
## X1stFlrSF             1.006402e-02               X1stFlrSF
## X2ndFlrSF             7.787775e-06               X2ndFlrSF
## GrLivArea             2.095308e-02                GrLivArea
## BsmtFullBath          1.695339e-03             BsmtFullBath
## BsmtHalfBath          1.847707e-04             BsmtHalfBath
## FullBath              5.208858e-04                 FullBath
## HalfBath              3.035856e-04                 HalfBath
## BedroomAbvGr         -2.707707e-04             BedroomAbvGr
## KitchenAbvGr         -1.266245e-02             KitchenAbvGr
## Fireplaces            2.408956e-03               Fireplaces
## GarageCars            2.527433e-03               GarageCars
## GarageArea            5.366516e-06               GarageArea
## WoodDeckSF            4.130959e-06               WoodDeckSF
## OpenPorchSF           9.386855e-06              OpenPorchSF
## EncPorchSF            1.398712e-05               EncPorchSF
## PoolArea              7.783859e-06                 PoolArea
## MiscVal              -5.287048e-07                  MiscVal
## YrSold                1.467653e-07                   YrSold
## age                  -8.200541e-04                      age
## MSZoningRM           -4.092646e-03               MSZoningRM
## LotShapeOther        -6.555539e-04            LotShapeOther
## LandContourHLS        2.619546e-04            LandContourHLS
## LandContourLow       -1.093244e-05            LandContourLow
## LotConfigCulDSac      1.189646e-03          LotConfigCulDSac
## LotConfigother       -1.676626e-03           LotConfigother
## LandSlopeSev         -2.877442e-04             LandSlopeSev
## NeighborhoodClearCr   1.800921e-03 NeighborhoodClearCr
## NeighborhoodCollgCr  -2.974242e-04 NeighborhoodCollgCr
## NeighborhoodCrawfor   9.821455e-03 NeighborhoodCrawfor
## NeighborhoodEdwards  -4.563440e-03 NeighborhoodEdwards
## NeighborhoodGilbert  -1.303855e-03 NeighborhoodGilbert
## NeighborhoodMitchel  -2.329404e-03 NeighborhoodMitchel
## NeighborhoodNAmes    -9.249347e-04   NeighborhoodNAmes
## NeighborhoodNoRidge   1.217151e-03 NeighborhoodNoRidge
## NeighborhoodNridgHt   2.412644e-03 NeighborhoodNridgHt
## NeighborhoodNWAmes   -1.188896e-03   NeighborhoodNWAmes
## NeighborhoodOldTown  -1.521261e-03 NeighborhoodOldTown
## Neighborhoodother     8.134588e-04    Neighborhoodother
## NeighborhoodSawyer   -6.889400e-04   NeighborhoodSawyer
## NeighborhoodSawyerW  -1.788638e-03 NeighborhoodSawyerW
## NeighborhoodSomerst   3.291691e-03 NeighborhoodSomerst
## Condition1Feedr       3.035996e-04          Condition1Feedr
## Condition1Norm        4.398693e-03           Condition1Norm
## Condition1PosN        2.023385e-03           Condition1PosN
## BldgType2fmCon       -8.039267e-04           BldgType2fmCon
## BldgTypeDuplex       -2.182335e-03           BldgTypeDuplex
## BldgTypeTwnhsE        4.052216e-04           BldgTypeTwnhsE
## RoofStyleHip          6.050976e-04             RoofStyleHip
## RoofStyleother        4.238400e-03           RoofStyleother
```

```
## Exterior1stHdBoard   -1.177376e-03   Exterior1stHdBoard
## Exterior1stPlywood   -1.973621e-03   Exterior1stPlywood
## Exterior1stVinylSd   -3.677874e-04   Exterior1stVinylSd
## Exterior1stWd Sdng   -3.166369e-03   Exterior1stWd Sdng
## Exterior2ndPlywood   -2.716752e-04   Exterior2ndPlywood
## Exterior2ndWd Sdng    2.229592e-03   Exterior2ndWd Sdng
## Exterior2ndWd Shng   -2.020496e-04   Exterior2ndWd Shng
## MasVnrTypeNone         3.229467e-04          MasVnrTypeNone
## ExterQualBelowAvg    -2.889534e-03   ExterQualBelowAvg
## ExterCondAvg          7.015922e-04            ExterCondAvg
## FoundationPConc       1.177146e-03      FoundationPConc
## BsmtCondAvg          -8.422335e-05             BsmtCondAvg
## BsmtCondBelowAvg     -2.437971e-03    BsmtCondBelowAvg
## BsmtExposureGd        2.685726e-03        BsmtExposureGd
## BsmtExposureMn       -5.545865e-04        BsmtExposureMn
## BsmtExposureNo       -3.694468e-04        BsmtExposureNo
## BsmtFinType1BLQ      -9.599672e-04        BsmtFinType1BLQ
## BsmtFinType1GLQ       9.696517e-04        BsmtFinType1GLQ
## BsmtFinType1LwQ      -1.213133e-03        BsmtFinType1LwQ
## BsmtFinType1Rec      -6.505120e-04        BsmtFinType1Rec
## BsmtFinType1Unf      -9.317532e-04        BsmtFinType1Unf
## BsmtFinType2BLQ      -7.852759e-04        BsmtFinType2BLQ
## BsmtFinType2GLQ       3.677316e-04        BsmtFinType2GLQ
## BsmtFinType2Rec      -2.886756e-04        BsmtFinType2Rec
## Heatingother          2.635461e-03            Heatingother
## HeatingQCAvg         -1.115955e-03            HeatingQCAvg
## HeatingQCBelowAvg    -1.955446e-03    HeatingQCBelowAvg
## CentralAirY           3.086237e-03             CentralAirY
## ElectricalFuseF      -3.033244e-03        ElectricalFuseF
## ElectricalFuseP      -2.082504e-03        ElectricalFuseP
## KitchenQualAvg       -1.100734e-03         KitchenQualAvg
## KitchenQualBelowAvg  -1.847124e-03 KitchenQualBelowAvg
## FunctionalMaj2       -1.847284e-02          FunctionalMaj2
## FunctionalMin2        3.723526e-04          FunctionalMin2
## FunctionalTyp         4.858628e-03           FunctionalTyp
## FireplaceQuAvg       -4.731988e-04          FireplaceQuAvg
## GarageTypeBasment    -6.252009e-04    GarageTypeBasment
## GarageTypeBuiltIn     7.040120e-04    GarageTypeBuiltIn
## GarageTypeCarPort    -1.299415e-03    GarageTypeCarPort
## GarageTypeDetchd      3.995535e-04     GarageTypeDetchd
## GarageFinishRFn      -4.123526e-04        GarageFinishRFn
## GarageFinishUnf      -7.402746e-04        GarageFinishUnf
## GarageQualBelowAvg   -1.136449e-03   GarageQualBelowAvg
## GarageCondBelowAvg   -1.197109e-03   GarageCondBelowAvg
## PavedDriveP          -5.764007e-04             PavedDriveP
## PavedDriveY           1.893098e-03             PavedDriveY
```

```r
# Get CV RMSE estimate
cv_rmse <- min(lasso_model$results$RMSE)
print(paste("CV RMSE Estimate:", cv_rmse))
```

```
## [1] "CV RMSE Estimate: 0.00747623279324956"
```

1d. RIDGE, ElasticNet, PCR MODELS

```r
# Ridge

cv_control <- trainControl(method = "cv", number = 5)

# Define the grid of hyperparameters (lambda values) to test
tune_grid <- expand.grid(alpha = 0,  # Ridge regression is when alpha = 0
                         lambda = seq(0.0023,0.0024, length = 500))  # Range of lambda values

# Train Ridge model using caret and glmnet
Ridge_model <- train(
  logSalePrice ~ .,
  data = housingDataProcessed,
  method = "glmnet",
  trControl = cv_control,
  tuneGrid = tune_grid
)

summary(Ridge_model)
```

```
##              Length Class      Mode
## a0             100  -none-     numeric
## beta         15200  dgCMatrix  S4
## df             100  -none-     numeric
## dim              2  -none-     numeric
## lambda         100  -none-     numeric
## dev.ratio      100  -none-     numeric
## nulldev          1  -none-     numeric
## npasses          1  -none-     numeric
## jerr             1  -none-     numeric
## offset           1  -none-     logical
## call             5  -none-     call
## nobs             1  -none-     numeric
## lambdaOpt        1  -none-     numeric
## xNames         152  -none-     character
## problemType      1  -none-     character
## tuneValue        2  data.frame list
## obsLevels        1  -none-     logical
## param            0  -none-     list
```

```r
plot(Ridge_model)
```
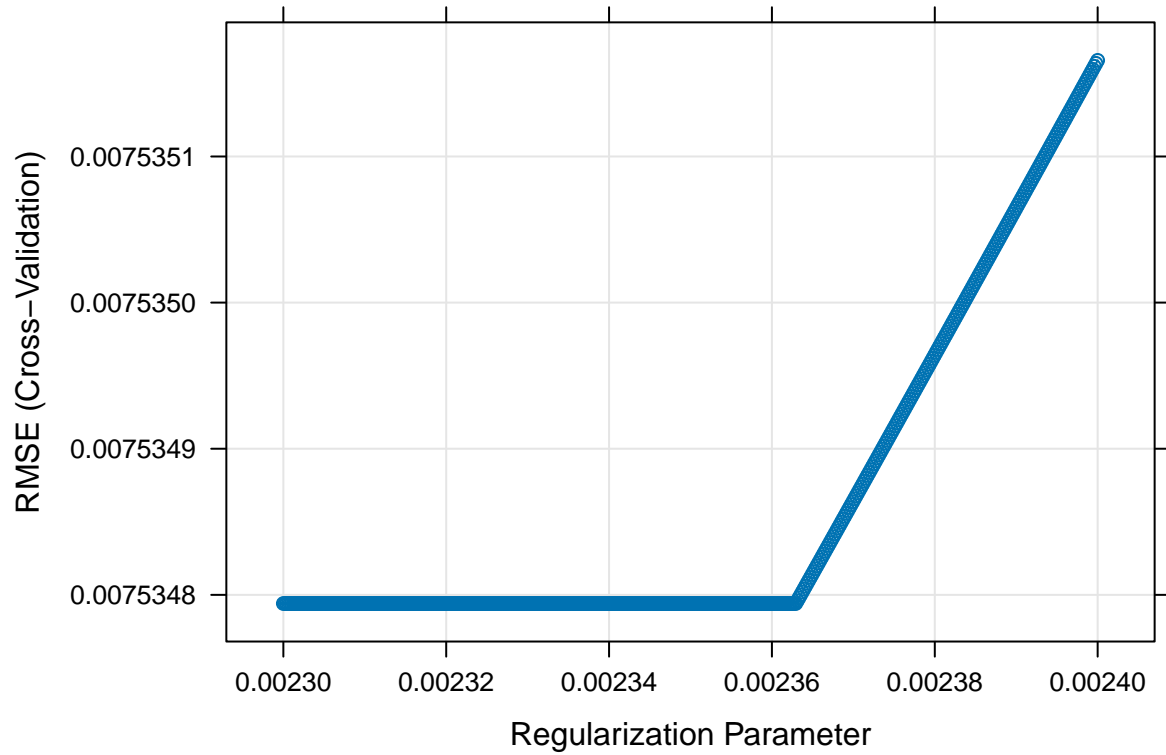
```r
# Visualize coefficient shrinkage as lambda increases
plot(Ridge_model$finalModel, label = TRUE)
```
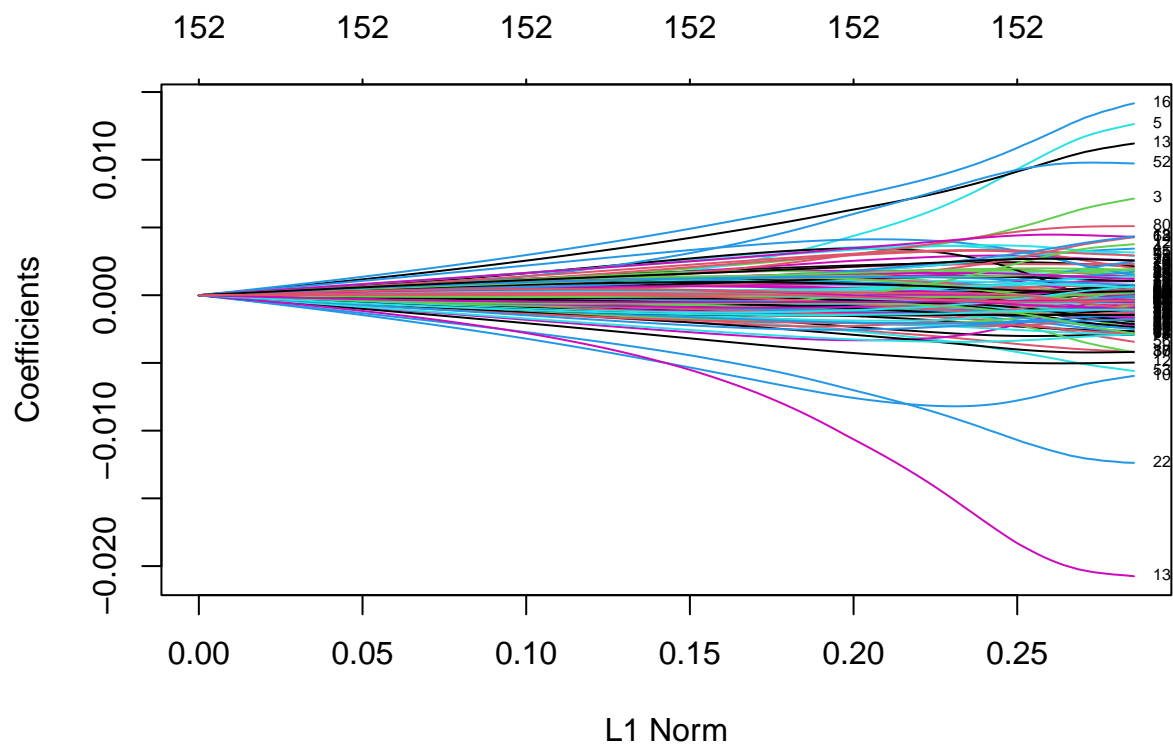
```
best_lambda <- Ridge_model$bestTune$lambda
coef(Ridge_model$finalModel, s = best_lambda)
```

```
## 153 x 1 sparse Matrix of class "dgCMatrix"
##                             s1
## (Intercept)      1.936056e+00
## MSSubClass      -1.890377e-05
## LotFrontage     -1.067970e-05
## LotArea          7.135474e-03
## OverallQual      3.438494e-03
## OverallCond      1.263826e-02
## YearBuilt        3.697488e-08
## YearRemodAdd     1.672190e-08
## MasVnrArea       4.257910e-06
## BsmtFinSF1       5.089038e-06
## BsmtFinSF2       4.943971e-06
## BsmtUnfSF        9.184494e-07
## TotalBsmtSF      6.797031e-06
## X1stFlrSF        1.120172e-02
## X2ndFlrSF        7.077044e-06
## LowQualFinSF    -4.986144e-07
## GrLivArea        1.416535e-02
## BsmtFullBath     1.828217e-03
## BsmtHalfBath     1.032992e-03
## FullBath         2.134128e-03
```

```
## HalfBath              1.361168e-03
## BedroomAbvGr         -1.788481e-04
## KitchenAbvGr         -1.237923e-02
## TotRmsAbvGrd          1.552013e-03
## Fireplaces            2.572099e-03
## GarageYrBlt           1.523361e-09
## GarageCars            2.304265e-03
## GarageArea            7.371082e-06
## WoodDeckSF            5.249605e-06
## OpenPorchSF           1.139922e-05
## EncPorchSF            1.482354e-05
## PoolArea              1.284416e-05
## MiscVal              -4.053925e-07
## MoSold               -6.398538e-06
## YrSold                1.528592e-05
## age                  -5.490733e-04
## ageSinceRemodel      -2.991234e-05
## ageofGarage           6.268602e-06
## MSZoningRM           -4.173626e-03
## MSZoningOther         1.302215e-05
## LotShapeReg           1.874258e-04
## LotShapeOther        -1.221832e-03
## LandContourHLS        1.434565e-03
## LandContourLow       -5.220890e-04
## LandContourLvl        1.051599e-04
## LotConfigCulDSac      1.566208e-03
## LotConfigInside      -4.904752e-04
## LotConfigother       -2.766679e-03
## LandSlopeMod         -6.072658e-05
## LandSlopeSev         -8.808468e-04
## NeighborhoodClearCr   2.000201e-03
## NeighborhoodCollgCr  -1.079950e-03
## NeighborhoodCrawfor   9.728967e-03
## NeighborhoodEdwards  -5.595183e-03
## NeighborhoodGilbert  -2.451792e-03
## NeighborhoodIDOTRR    1.360663e-04
## NeighborhoodMitchel  -3.439061e-03
## NeighborhoodNAmes    -2.146110e-03
## NeighborhoodNoRidge   1.478774e-03
## NeighborhoodNridgHt   2.889677e-03
## NeighborhoodNWAmes   -2.258866e-03
## NeighborhoodOldTown  -2.660634e-03
## Neighborhoodother     5.427601e-04
## NeighborhoodSawyer   -1.957216e-03
## NeighborhoodSawyerW  -2.839480e-03
## NeighborhoodSomerst   3.153485e-03
## NeighborhoodTimber   -3.348366e-04
## Condition1Feedr       6.059330e-04
## Condition1Norm        4.279535e-03
## Condition1PosA        2.016993e-04
## Condition1PosN        2.495287e-03
## Condition1RR         -2.742148e-04
## BldgType2fmCon       -1.315744e-03
## BldgTypeDuplex       -2.916274e-03
```

```
## BldgTypeTwnhs        -1.614060e-03
## BldgTypeTwnhsE       -1.322541e-04
## HouseStyle1Story     -6.749146e-04
## HouseStyle2Story      6.033547e-04
## HouseStyleOther      -9.498762e-04
## RoofStyleHip          1.045457e-03
## RoofStyleother        5.097688e-03
## Exterior1stCemntBd   -2.440584e-03
## Exterior1stHdBoard   -1.677904e-03
## Exterior1stMetalSd   -6.208206e-04
## Exterior1stother     -1.048300e-04
## Exterior1stPlywood   -2.125345e-03
## Exterior1stVinylSd   -1.574698e-03
## Exterior1stWd Sdng   -4.218995e-03
## Exterior2ndCmentBd   -7.484913e-05
## Exterior2ndHdBoard   -1.348863e-03
## Exterior2ndMetalSd   -8.816245e-04
## Exterior2ndother     -7.028466e-04
## Exterior2ndPlywood   -1.819716e-03
## Exterior2ndVinylSd   -5.749212e-04
## Exterior2ndWd Sdng    1.726958e-03
## Exterior2ndWd Shng   -2.486483e-03
## MasVnrTypeBrkFace     1.839200e-04
## MasVnrTypeNone        7.003133e-04
## MasVnrTypeStone       3.204992e-04
## ExterQualAvg         -7.485218e-04
## ExterQualBelowAvg    -5.961850e-03
## ExterCondAvg          1.407985e-03
## ExterCondBelowAvg     2.055790e-05
## FoundationCBlock      6.744192e-04
## Foundationother       4.533792e-04
## FoundationPConc       1.903498e-03
## BsmtQualAvg          -4.188902e-05
## BsmtQualBelowAvg      3.807623e-04
## BsmtCondAvg          -1.658984e-03
## BsmtCondBelowAvg     -4.185467e-03
## BsmtExposureGd        2.941917e-03
## BsmtExposureMn       -1.251118e-03
## BsmtExposureNo       -1.008676e-03
## BsmtFinType1BLQ      -1.481931e-03
## BsmtFinType1GLQ       1.200213e-03
## BsmtFinType1LwQ      -2.361893e-03
## BsmtFinType1Rec      -1.142820e-03
## BsmtFinType1Unf      -1.416910e-03
## BsmtFinType2BLQ      -2.476254e-03
## BsmtFinType2GLQ       5.442562e-04
## BsmtFinType2LwQ      -1.378382e-03
## BsmtFinType2Rec      -1.934195e-03
## BsmtFinType2Unf      -7.200500e-04
## Heatingother          3.768814e-03
## HeatingQCAvg         -1.184991e-03
## HeatingQCBelowAvg    -2.894658e-03
## CentralAirY           4.314198e-03
## ElectricalFuseF      -4.974712e-03
```

```
## ElectricalFuseP      -2.840791e-03
## ElectricalSBrkr      -3.477188e-04
## KitchenQualAvg       -1.436961e-03
## KitchenQualBelowAvg  -2.895430e-03
## FunctionalMaj2       -2.075026e-02
## FunctionalMin1       -1.030774e-03
## FunctionalMin2        4.462280e-04
## FunctionalMod         1.670876e-04
## FunctionalTyp         4.339377e-03
## FireplaceQuAvg       -5.503425e-04
## FireplaceQuBelowAvg  -3.366733e-04
## GarageTypeAttchd      2.913915e-04
## GarageTypeBasment    -1.259594e-03
## GarageTypeBuiltIn     1.897867e-03
## GarageTypeCarPort    -1.801870e-03
## GarageTypeDetchd      6.564434e-04
## GarageFinishRFn      -9.547353e-04
## GarageFinishUnf      -1.460834e-03
## GarageQualAvg        -1.042438e-03
## GarageQualBelowAvg   -2.834860e-03
## GarageCondAvg         7.733132e-04
## GarageCondBelowAvg   -1.079914e-03
## PavedDriveP          -5.239847e-04
## PavedDriveY           2.558404e-03
## SaleTypeWD           -3.966965e-04
```

```r
Ridge_rmse <- Ridge_model$results$RMSE
Ridge_r2 <- Ridge_model$results$Rsquared


# *****************************************************************************

# ElasticNet

cv_control <- trainControl(method = "cv", number = 5)

# Define the grid of hyperparameters (lambda values) to test
tune_grid <- expand.grid(alpha = seq(0,1, length = 100),
                         lambda = seq(0.0023,0.0024, length = 500))  # Range of lambda values


# Train LASSO model using caret and glmnet
elasticNet_model <- train(
  logSalePrice ~ .,
  data = housingDataProcessed,
  method = "glmnet",
  trControl = cv_control,
  tuneGrid = tune_grid
)

summary(elasticNet_model)
```

```
##           Length Class      Mode
## a0           97  -none-     numeric
## beta      14744  dgCMatrix  S4
```

```
## df               97  -none-      numeric
## dim               2  -none-      numeric
## lambda           97  -none-      numeric
## dev.ratio        97  -none-      numeric
## nulldev           1  -none-      numeric
## npasses           1  -none-      numeric
## jerr              1  -none-      numeric
## offset            1  -none-      logical
## call              5  -none-      call
## nobs              1  -none-      numeric
## lambdaOpt         1  -none-      numeric
## xNames          152  -none-      character
## problemType       1  -none-      character
## tuneValue         2  data.frame  list
## obsLevels         1  -none-      logical
## param             0  -none-      list
```
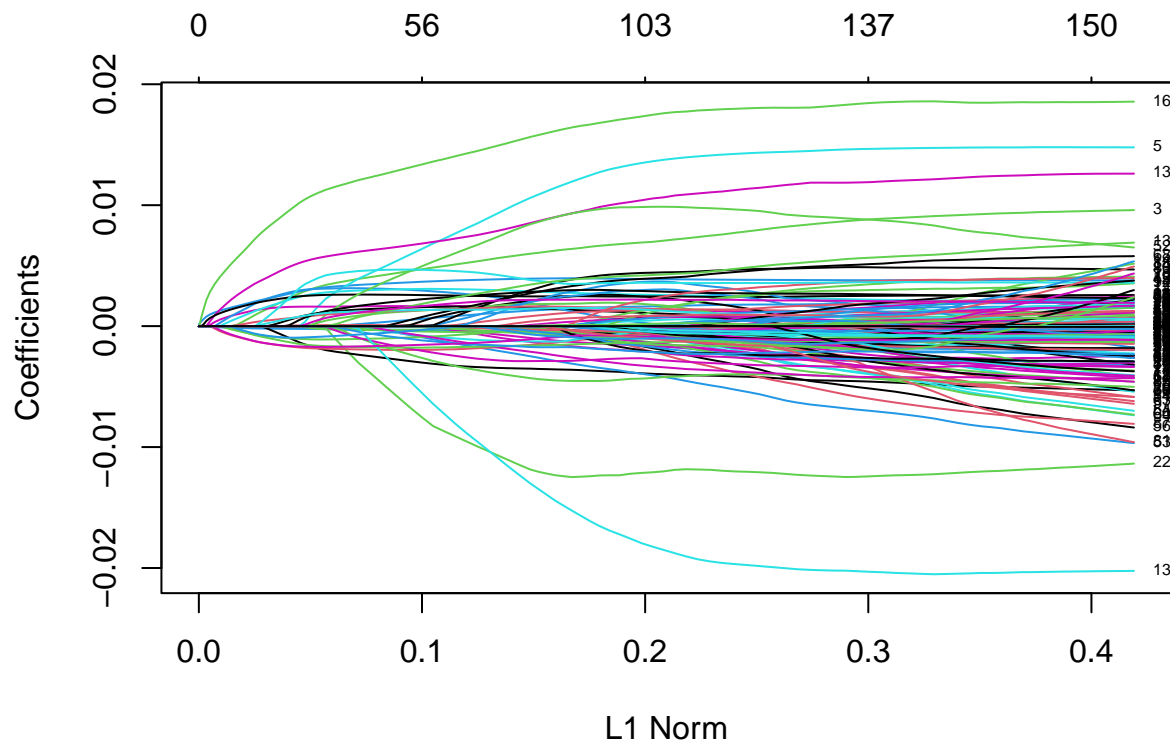
```
plot(elasticNet_model)
```
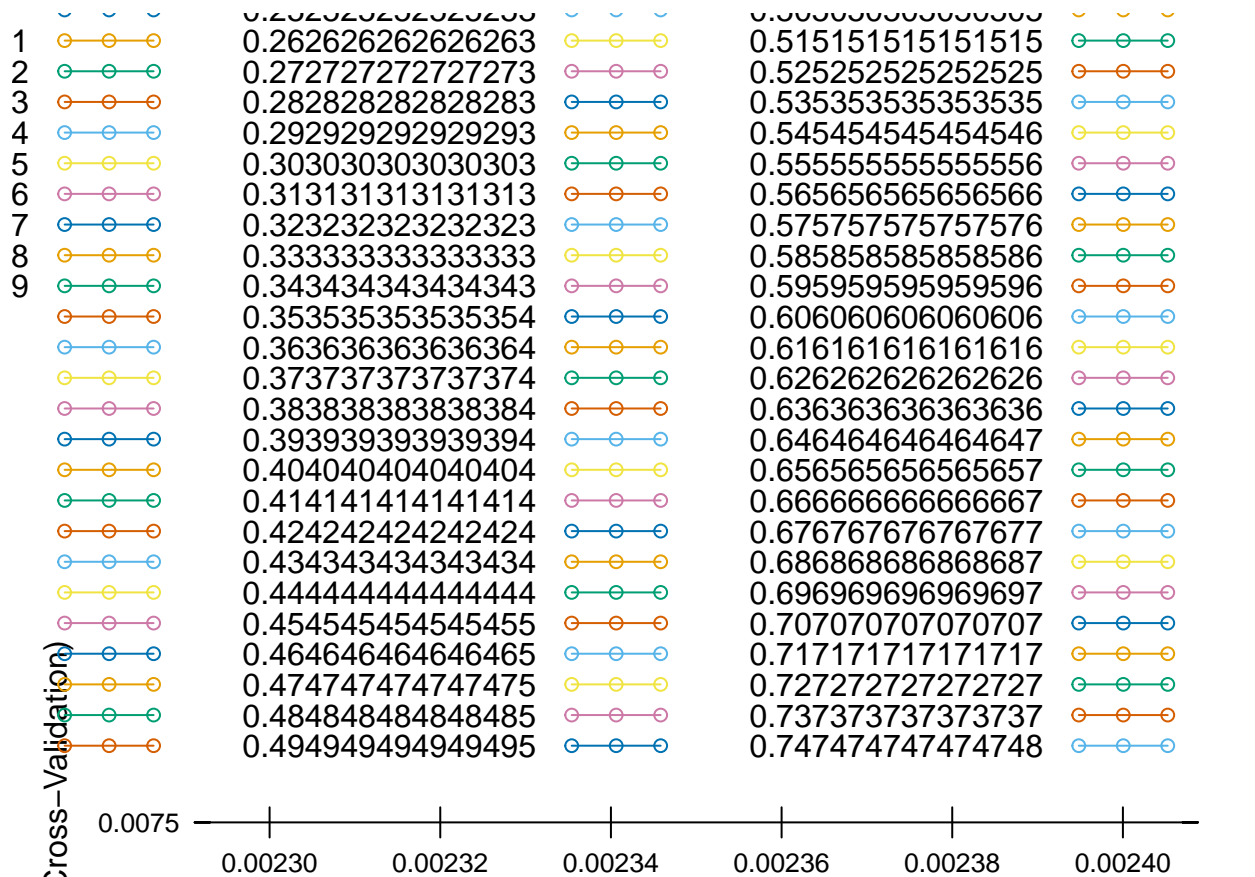


```
# Visualize coefficient shrinkage as lambda increases
plot(elasticNet_model$finalModel, label = TRUE)
```

```
best_alpha <- elasticNet_model$bestTune$alpha
best_lambda <- elasticNet_model$bestTune$lambda

coef(elasticNet_model$finalModel, s = best_lambda)
```

```
## 153 x 1 sparse Matrix of class "dgCMatrix"
##                              s1
## (Intercept)      2.057019e+00
## MSSubClass           .
## LotFrontage          .
## LotArea          6.571669e-03
## OverallQual      3.942457e-03
## OverallCond      1.249925e-02
## YearBuilt        3.335438e-08
## YearRemodAdd     1.582107e-08
## MasVnrArea       2.810545e-06
## BsmtFinSF1       4.824822e-06
## BsmtFinSF2       2.439357e-06
## BsmtUnfSF            .
## TotalBsmtSF      7.922090e-06
## X1stFlrSF        9.676454e-03
## X2ndFlrSF        6.788564e-06
## LowQualFinSF         .
## GrLivArea        1.668247e-02
## BsmtFullBath     1.518002e-03
```

```
## BsmtHalfBath            .
## FullBath                1.461685e-03
## HalfBath                1.269621e-03
## BedroomAbvGr            .
## KitchenAbvGr           -1.240445e-02
## TotRmsAbvGrd            9.405164e-04
## Fireplaces              2.697935e-03
## GarageYrBlt             .
## GarageCars              2.465415e-03
## GarageArea              7.385795e-06
## WoodDeckSF              4.558689e-06
## OpenPorchSF             9.094694e-06
## EncPorchSF              1.172190e-05
## PoolArea                3.218195e-06
## MiscVal                -4.510963e-07
## MoSold                  .
## YrSold                  .
## age                    -5.622730e-04
## ageSinceRemodel        -3.148467e-05
## ageofGarage             .
## MSZoningRM             -3.678567e-03
## MSZoningOther           .
## LotShapeReg             .
## LotShapeOther           .
## LandContourHLS          2.013111e-04
## LandContourLow          .
## LandContourLvl          .
## LotConfigCulDSac        1.202011e-03
## LotConfigInside         .
## LotConfigother         -1.054441e-03
## LandSlopeMod            .
## LandSlopeSev            .
## NeighborhoodClearCr     2.559943e-03
## NeighborhoodCollgCr     .
## NeighborhoodCrawfor     9.579219e-03
## NeighborhoodEdwards    -3.350645e-03
## NeighborhoodGilbert     .
## NeighborhoodIDOTRR      .
## NeighborhoodMitchel    -9.062379e-04
## NeighborhoodNAmes      -1.321586e-04
## NeighborhoodNoRidge     1.906360e-03
## NeighborhoodNridgHt     2.649445e-03
## NeighborhoodNWAmes      .
## NeighborhoodOldTown    -1.399384e-03
## Neighborhoodother       2.657827e-04
## NeighborhoodSawyer      .
## NeighborhoodSawyerW    -9.538624e-04
## NeighborhoodSomerst     3.145108e-03
## NeighborhoodTimber      .
## Condition1Feedr         .
## Condition1Norm          3.534908e-03
## Condition1PosA          .
## Condition1PosN          3.520361e-04
## Condition1RR            .
```

```
## BldgType2fmCon           .
## BldgTypeDuplex        -1.799464e-03
## BldgTypeTwnhs        -2.379447e-05
## BldgTypeTwnhsE          .
## HouseStyle1Story        .
## HouseStyle2Story        .
## HouseStyleOther         .
## RoofStyleHip          4.575394e-04
## RoofStyleother        4.091058e-03
## Exterior1stCemntBd      .
## Exterior1stHdBoard    -5.846692e-04
## Exterior1stMetalSd      .
## Exterior1stother        .
## Exterior1stPlywood    -8.826713e-04
## Exterior1stVinylSd      .
## Exterior1stWd Sdng    -8.808911e-04
## Exterior2ndCmentBd      .
## Exterior2ndHdBoard    -2.049587e-04
## Exterior2ndMetalSd      .
## Exterior2ndother        .
## Exterior2ndPlywood    -4.621205e-04
## Exterior2ndVinylSd      .
## Exterior2ndWd Sdng      .
## Exterior2ndWd Shng      .
## MasVnrTypeBrkFace       .
## MasVnrTypeNone          .
## MasVnrTypeStone         .
## ExterQualAvg         -1.059199e-03
## ExterQualBelowAvg    -4.532421e-03
## ExterCondAvg          2.751095e-04
## ExterCondBelowAvg    -9.287141e-04
## FoundationCBlock        .
## Foundationother         .
## FoundationPConc       1.550876e-03
## BsmtQualAvg             .
## BsmtQualBelowAvg        .
## BsmtCondAvg             .
## BsmtCondBelowAvg     -2.794712e-03
## BsmtExposureGd        2.994647e-03
## BsmtExposureMn       -4.628353e-06
## BsmtExposureNo       -4.221446e-04
## BsmtFinType1BLQ      -1.776154e-04
## BsmtFinType1GLQ       1.502208e-03
## BsmtFinType1LwQ         .
## BsmtFinType1Rec         .
## BsmtFinType1Unf      -3.607046e-04
## BsmtFinType2BLQ         .
## BsmtFinType2GLQ         .
## BsmtFinType2LwQ         .
## BsmtFinType2Rec         .
## BsmtFinType2Unf         .
## Heatingother          8.933450e-04
## HeatingQCAvg         -8.545462e-04
## HeatingQCBelowAvg    -1.096745e-03
```

```
## CentralAirY          3.524465e-03
## ElectricalFuseF     -2.798227e-03
## ElectricalFuseP     -1.138421e-03
## ElectricalSBrkr          .
## KitchenQualAvg       -1.460028e-03
## KitchenQualBelowAvg -1.846226e-03
## FunctionalMaj2       -1.616458e-02
## FunctionalMin1           .
## FunctionalMin2           .
## FunctionalMod            .
## FunctionalTyp         3.963152e-03
## FireplaceQuAvg       -1.813090e-04
## FireplaceQuBelowAvg      .
## GarageTypeAttchd         .
## GarageTypeBasment        .
## GarageTypeBuiltIn     7.297094e-04
## GarageTypeCarPort    -8.109239e-04
## GarageTypeDetchd         .
## GarageFinishRFn      -3.007715e-05
## GarageFinishUnf      -7.595630e-04
## GarageQualAvg            .
## GarageQualBelowAvg   -5.723741e-04
## GarageCondAvg         9.801053e-05
## GarageCondBelowAvg   -1.005734e-03
## PavedDriveP              .
## PavedDriveY           2.185022e-03
## SaleTypeWD               .
```

```r
elasticnet_rmse <- elasticNet_model$results$RMSE
elasticnet_r2 <- elasticNet_model$results$Rsquared


# *******************************************************************************

# PCR
cv_control <- trainControl(method = "cv", number = 5)

tune_grid <- expand.grid(ncomp = seq(25,35, length = 100))

# Train PCR model using caret
pcr_model <- train(logSalePrice ~ .,
                data = housingDataProcessed,
                method = "pcr",
                trControl = cv_control,
                tuneGrid = tune_grid
)

summary(pls_model)
```
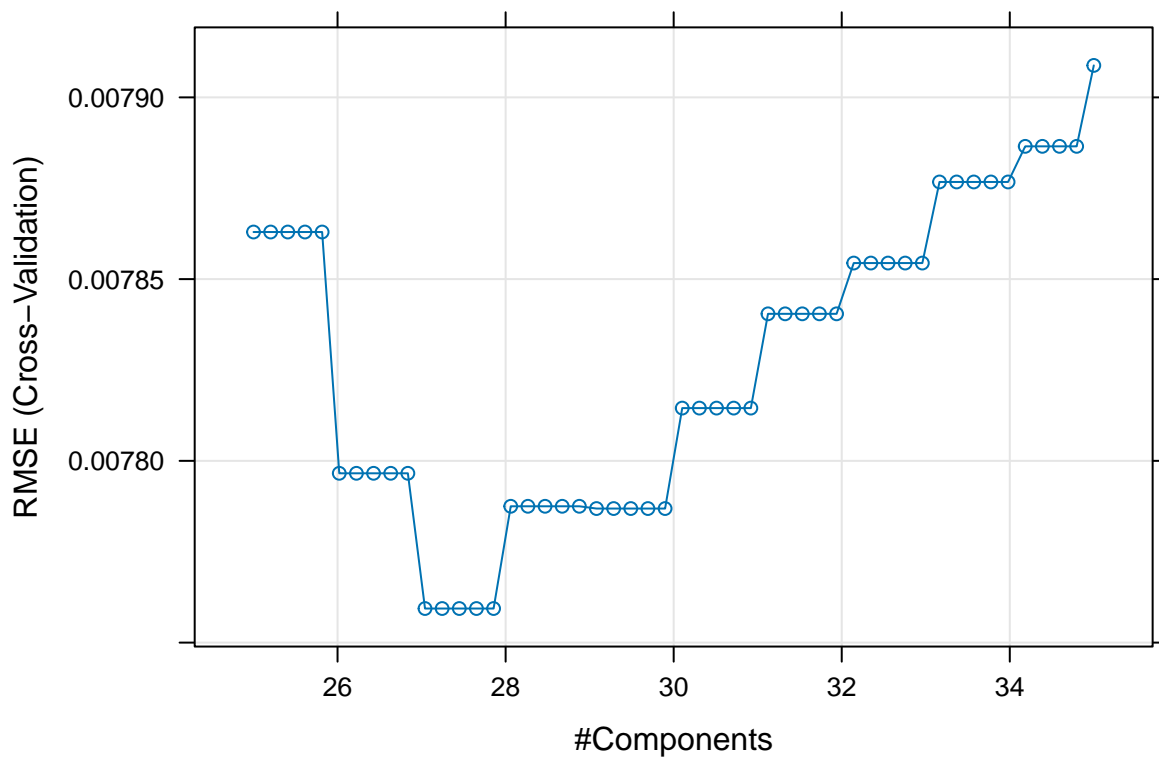
```
## Data:    X dimension: 998 152
##  Y dimension: 998 1
## Fit method: oscorespls
## Number of components considered: 27.04082
## TRAINING: % variance explained
##            1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps
```

```
## X            80.71    93.47    99.99    99.99    99.99   100.00    100.00
## .outcome     37.24    38.53    38.66    77.83    79.43    81.15     81.28
##            8 comps  9 comps  10 comps  11 comps  12 comps  13 comps  14 comps
## X           100.00   100.00    100.00    100.00    100.00    100.00    100.00
## .outcome     81.51    81.67     81.69     81.72     81.73     81.74     81.78
##            15 comps  16 comps  17 comps  18 comps  19 comps  20 comps  21 comps
## X           100.00    100.00    100.00    100.00     100.0    100.00    100.00
## .outcome     81.88     84.12     85.53     89.96      91.3     92.19     93.54
##            22 comps  23 comps  24 comps  25 comps  26 comps  27 comps
## X             100.0    100.00    100.00    100.00     100.0     100.0
## .outcome       94.3     94.68     95.06     95.34      95.5      95.6
```

```r
plot(pls_model)
```



```r
pcr_rmse <- pcr_model$results$RMSE
pcr_r2 <- pcr_model$results$Rsquared

# ****************************************************************************

model_performance <- data.frame(
  Model = c("OLS with Interactions", "PLS", "LASSO", "Ridge", "ElasticNet", "PCR"),
  Hyperparameter = c("None", "ncomp", "alpha = 1, lambda", "alpha = 0, lambda", "alpha and lambda", "nc
  CV_RMSE = c(min(ols_cv_interaction_rmse), min(pls_rmse), min(lasso_rmse), min(Ridge_rmse), min(elasti
  CV_R2 = c(max(ols_cv_interaction_r2), max(pls_r2), max(lasso_r2), max(Ridge_r2), max(elasticnet_r2), 
)
```

```r
options(digits=3)
options(scipen=99)
model_performance %>% kable()
```

Model	Hyperparameter	CV_RMSE	CV_R2
OLS with Interactions	None	0.009	0.910
PLS	ncomp	0.008	0.933
LASSO	alpha = 1, lambda	0.007	0.939
Ridge	alpha = 0, lambda	0.008	0.936
ElasticNet	alpha and lambda	0.007	0.938
PCR	ncomp	0.010	0.893