

12/9/2022

Don Anderson

Shawn Deng

Living Area and House Prices: Predicting House Prices in Ames

When we evaluate the price of a house, there are many factors that could affect it, such as living area and neighborhood. In this project, we are given data from a Kaggle competition regarding houses in Ames, Iowa. We conducted a series of modeling to test if sale price of the house is related to the square footage of the living area of the house in certain neighborhoods, and whether the neighborhood they are located in is a significant factor when determining the sale price of those houses. We would also build the most predictive model for sales prices of homes in all of Ames, Iowa.

Analysis 1: Sale Price Versus Living Area and Neighborhood

In this section, we would like to help the real estate company, Century 21 Ames, derive meaningful insights from their data, which in turn will drive informed decision making. We are taking a close look at 'NAmes,' 'Edwards' and 'BrkSide' neighborhoods and would like to simply get an estimate of how the sale price of the house is related to the square footage of the living area of the house. Particularly, we are interested in modeling the relationship between total square footage of the living area (GrLivArea) of houses in Ames Iowa USA with their Sale prices (SalePrice)

The Real Estate company sold 383 houses in 'BrkSide' 'NAmes' and 'Edwards.' Most of which were from the 'NAmes' Neighborhood. The plot below summarizes the distribution of houses sold across the three neighborhoods. The average general living

area size is about 1,301 square feet. Also, the average sale price of a home in Ames, Iowa, USA is \$138,062. Whereas the house with the largest general living area had a living area of 5,642 square feet, the most expensive house was worth \$345,000.

Descriptive Statistics of SalePrice by Neighborhood						
The MEANS Procedure						
Analysis Variable : SalePrice						
Neighborhood	N Obs	N	Mean	Std Dev	Minimum	Maximum
BrkSide	55	55	128892.27	37284.40	68500.00	223500.00
Edwards	96	96	126130.21	35031.15	61000.00	243000.00
NAmes	221	221	143504.95	27796.96	87500.00	242000.00

Picture 1: Descriptive Statistics of sale price by neighborhood

In order for our analysis to be valid, there are assumptions that we need to make. Most of them can be easily met, because the data appears to be large in size, normal and equal variance. First of all, the samples should be independent of each other. In reality, it is hard to achieve, because the sale of a house might influence each other, especially when they are close in the same neighborhood. In order to proceed, we can assume this is met.

With the data provided, we can easily come up with a linear model for the sale price and living area by using R. We can see the result below:

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 78205.578   4536.054   17.24  <2e-16 ***
## GrLivArea    45.979     3.265    14.08  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 30980 on 381 degrees of freedom
## Multiple R-squared:  0.3423, Adjusted R-squared:  0.3406
## F-statistic: 198.3 on 1 and 381 DF,  p-value: < 2.2e-16
```

With the statistic calculated, we can finalize the linear model to be:

$$\text{SalePrice} = 78205.58 + 45.98 * \text{GrLivArea}$$

The slope of the model 45.98 suggests that for every 100 units increase in living area, the sale price will increase by 4598. For a house with even 0 living area, it is expected to cost 78205.58. However, we might think it is reasonable for the neighborhood the house is located to influence the price. We would also want to test for it statistically. We will add the neighborhood into our model, and see whether there is a statistically significant difference.

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  19971.514  12351.125   1.617  0.10672
## GrLivArea      87.163     9.782    8.911 < 2e-16 ***
## NeighborhoodEdwards  68381.591  13969.511   4.895 1.46e-06 ***
## NeighborhoodNames  54704.888  13882.334   3.941 9.69e-05 ***
## GrLivArea:NeighborhoodEdwards  -57.412    10.718  -5.357 1.48e-07 ***
## GrLivArea:NeighborhoodNames   -32.847    10.815  -3.037 0.00256 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 28550 on 377 degrees of freedom
## Multiple R-squared:  0.4474, Adjusted R-squared:  0.44
## F-statistic: 61.04 on 5 and 377 DF,  p-value: < 2.2e-16

confint(price_fit3)

##              2.5 %      97.5 %
## (Intercept) -4314.21151 44257.23910
## GrLivArea      67.92850  106.39657
## NeighborhoodEdwards  40913.67039 95849.51158
## NeighborhoodNames  27408.38290 82001.39258
## GrLivArea:NeighborhoodEdwards  -78.48612  -36.33834
## GrLivArea:NeighborhoodNames   -54.11269  -11.58065
```

In the model above, we used 'BrkSide' as our control group, meaning the other 2 neighborhoods will be compared to it for possible differences. From the test result above, we can see that the differences, either in intercepts or slope, are indeed

significant, because their P-values are all below alpha. Therefore, in order to predict the housing price to the maximum extent, we need to upgrade our model into:

For BrkSide: SalePrice = 19971.51 + 87.16 * GrLivArea

For Edwards: SalePrice = 88353.1 + 29.75 * GrLivArea

For NAmes: SalePrice = 74676.4 + 54.31 * GrLivArea

This model is indeed better than the previous one, because we can see that R^2 increased from 0.342 to 0.447, which means more variance can now be explained by the model. Furthermore, we can also find a 95% confidence interval, which means we are 95% confident that the value will be between this range.

We also can test our result in SAS as well for verification. Since the tool is designed differently, the result is also different. But it is similar to our R finding.

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	88105	4089.52659	21.65	<.0001
GrLivArea	1	37.67208	2.95373	12.75	<.0001

Analysis 2: Build the Most Predictive Model for Sales Prices of Homes in All of Ames

In this section, we are still building a model for sales prices. However, this time we are no longer limited to just the 3 neighborhoods we discussed in the previous question. We are also no longer restricted to the 'GrLivArea'. We can take any variables into consideration.

First of all, we start building our model by using the forward model selection. Since this process involves some randomization, we set our seed to 123 for repeatable

results. We will choose the best model based on Bayesian Information Criterion(BIC). The model with lowest BIC value will be the best model here. Our model based on the forward subset selection method contains 18 predictors. Adjusted R^2 value of the model is 0.8226 which is pretty good. This means 82% variability of the Sale Price is explained by this model. Since there are too many variables, we are only showing some of the significant variables here. We did residual analysis for this model, and the result can be found in Appendix.

```

Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -8.953e+05  1.726e+05  -5.186 2.47e-07 ***
## LotArea         9.744e-01  1.131e-01   8.612 < 2e-16 ***
## NeighborhoodCollgCr -1.962e+04  9.763e+03  -2.010 0.044647 *
## NeighborhoodCrawfor  2.429e+04  1.142e+04   2.127 0.033613 *

.....

## SaleConditionNormal  8.357e+03  3.717e+03   2.248 0.024719 *
## SaleConditionPartial 2.710e+04  5.273e+03   5.140 3.14e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 34480 on 1374 degrees of freedom
## Multiple R-squared:  0.8226, Adjusted R-squared:  0.8116
## F-statistic: 74.96 on 85 and 1374 DF,  p-value: < 2.2e-16

```

Second, we will use a backward elimination model. The predictors are chosen for the final model based on Bayesian Information Criterion(BIC). Based on the lowest BIC value the final model contains 13 predictors. The Adjusted R^2 of this model is 0.786 which is good but slightly lower than our previous model. Just like the previous model we are only showing some significant results here. We did residual analysis for this model, and the result can be found in Appendix.

```

## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -8.953e+05  1.726e+05  -5.186 2.47e-07 ***

```

```
## MSSubClass          6.082e+01  4.805e+01   1.266 0.205879
## LotArea             9.744e-01  1.131e-01   8.612 < 2e-16 ***
.....

## SaleConditionPartial 2.710e+04  5.273e+03   5.140 3.14e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 34480 on 1374 degrees of freedom
## Multiple R-squared:  0.8226, Adjusted R-squared:  0.8116
## F-statistic: 74.96 on 85 and 1374 DF,  p-value: < 2.2e-16
```

Thirdly, we will do the Stepwise selection using regsubsets. The best coefficients based on lowest BIC values are extracted. The final best model contains 18 variables. The model performs better than the previous model with an adjusted R^2 value of 0.8171. We did residual analysis for this model, and the result can be found in Appendix.

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -8.180e+05  1.705e+05  -4.797 1.79e-06 ***
## LotArea      9.693e-01  1.115e-01   8.693 < 2e-16 ***
...

## KitchenQualTA      -5.895e+04  5.010e+03 -11.766 < 2e-16 ***
## SaleConditionPartial 2.565e+04  5.201e+03   4.931 9.17e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 33980 on 1373 degrees of freedom
## Multiple R-squared:  0.8279, Adjusted R-squared:  0.8171
## F-statistic: 76.78 on 86 and 1373 DF,  p-value: < 2.2e-16
```

The last model is a custom model, we chose 10 predictors that we think will have a close association on the sales price of a house, and put them into the model. They are: 'OverallCond', 'YearBuilt', 'YearRemodAdd', 'BedroomAbvGr', 'GrLivArea', 'KitchenAbvGr', 'TotRmsAbvGrd', 'GarageCars', 'PoolArea' and 'LotArea', The summary

of this model shows that this performed the worst compared to other models with adjusted R^2 value of 0.737. We did residual analysis for this model, and the result can be found in Appendix.

```
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.806e+06  1.219e+05 -14.820  < 2e-16 ***
## OverallCond  7.138e+03  1.148e+03   6.219 6.52e-10 ***
## YearBuilt    7.232e+02  5.603e+01  12.906  < 2e-16 ***
## YearRemodAdd 2.057e+02  7.324e+01   2.809 0.00504 **
## BedroomAbvGr -1.760e+04  1.855e+03  -9.489  < 2e-16 ***
## GrLivArea     8.201e+01  4.001e+00  20.499  < 2e-16 ***
## KitchenAbvGr -3.965e+04  5.269e+03  -7.525 9.24e-14 ***
## TotRmsAbvGrd  5.975e+03  1.419e+03   4.210 2.71e-05 ***
## GarageCars    1.997e+04  1.912e+03  10.444  < 2e-16 ***
## PoolArea     -2.178e+01  2.717e+01  -0.801  0.42301
## LotArea       6.784e-01  1.116e-01   6.078 1.55e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 40720 on 1449 degrees of freedom
## Multiple R-squared:  0.7391, Adjusted R-squared:  0.7373
## F-statistic: 410.5 on 10 and 1449 DF,  p-value: < 2.2e-16
```

After running the 4 tests, we now look back on them to see which model is the best. The table below suggests that the stepwise subset model is the best and results with the highest adjusted r-squared value of 0.817. The forward model results are pretty close to the best subset model with adjusted r-squared value of 0.812. The custom model with 10 predictors performed the worst here.

	R^2	CV Press	Kaggle Score
Forward	0.812		
Backward	0.787		
Stepwise	0.817		
Custom	0.737		

Appendix

GitHub Link:

Shawn Deng: <https://github.com/ShawnYixiaoDeng/DS-6371-Group-Peoject>

Don Anderson:

<https://github.com/DonAnderson1/DS-6371-Real-Estate-Analysis/tree/main>

Code used:

Q1

```
library(ggplot2)
```

```
library(dplyr)
```

```
train = read.csv(file.choose())
```

```
train_set = train %>% filter(Neighborhood == "NAMES" | Neighborhood == "Edwards" |  
  Neighborhood == "BrkSide")
```

```
price_fit1 <- lm(SalePrice ~ GrLivArea, data = train_set)
```

```
summary(price_fit1)
```

```
price_fit2 <- lm(SalePrice ~ GrLivArea*Neighborhood, data = train_set)
```

```
summary(price_fit2)
```

```
confint(price_fit2)
```

```
p1 <- ggplot(data = train_set, aes(GrLivArea, SalePrice)) + geom_point()
```

```
p2 <- ggplot(data = train_set, aes(GrLivArea, SalePrice, color = Neighborhood)) +  
  geom_point()
```



```

p3 <- ggplot(data = train_set, aes(GrLivArea, SalePrice, color = Neighborhood)) +
  geom_point() + geom_smooth(method = 'lm', se = FALSE)
p4 <- ggplot(data = train_set, aes(SalePrice, Neighborhood, fill = Neighborhood)) +
  geom_boxplot()

gridExtra::grid.arrange(p1, p2, p3, p4 , nrow = 2, ncol = 2)

par(mfrow = c(2, 2))

plot(price_fit3)

plot(price_fit3, 4)

```

Q2

```

set.seed(123)

forward_model <- regsubsets(SalePrice ~ ., data = train, nvmax = 20, method =
  "forward")

plot(summary(forward_model)$bic, xlab="Number of Var.", ylab = "BIC")

coef(forward_model, which.min(summary(forward_model)$bic))

fit_forward <- lm(SalePrice ~ MSSubClass + LotArea + LotShape + Neighborhood +
  Condition1 + Condition2 + BldgType + OverallQual +
  OverallCond + YearBuilt + RoofMatl + Exterior2nd + BsmtFinSF2 +
  FullBath + FullBath + KitchenQual + PavedDrive + SaleCondition,
  data = train)

summary(fit_forward)

confint(fit_forward)

```

```

par(mfrow = c(2, 2))

plot(fit_forward)

ncvTest(fit_forward)

vif(fit_forward)

outlierTest(fit_forward)


backward_model <- regsubsets(SalePrice ~ ., data = train, nvmax = 20, method =
  "backward")

plot(summary(backward_model)$bic, xlab="Number of Var.", ylab = "BIC")

coef(backward_model, which.min(summary(backward_model)$bic))

fit_backward <- lm(SalePrice ~ LotArea + Neighborhood + Condition2 + OverallQual +
  RoofMatl + OverallCond + YearBuilt + BsmtFinSF2 +LowQualFinSF+BsmntFullBath
  + KitchenQual + Functional + SaleCondition, data = train)

summary(fit_backward)

confint(fit_backward)

par(mfrow = c(2, 2))

plot(fit_backward)

ncvTest(fit_backward)

vif(fit_backward)

outlierTest(fit_backward)

plot(fit_backward, 4)


best_model <- regsubsets(SalePrice ~ . , data = train, nvmax = 20, method = "seqrep")

```

```

plot(summary(best_model)$bic, xlab="Number of Var.", ylab = "BIC")
coef(best_model, which.min(summary(best_model)$bic))
fit_best <- lm(SalePrice ~ MSSubClass + LotArea + LotShape + Neighborhood +
  Condition1 + Condition2 + BldgType + OverallQual + OverallCond + YearBuilt +
  RoofMatl + Exterior2nd + BsmtFinSF2 + FullBath + HalfBath + KitchenQual +
  PavedDrive + SaleCondition, data = train)
summary(fit_best)
confint(fit_best)
par(mfrow = c(2, 2))
plot(fit_best)
ncvTest(fit_best)
vif(fit_best)
outlierTest(fit_best)
plot(fit_best, 4)

custom_fit <- lm(SalePrice ~ OverallCond + YearBuilt + YearRemodAdd +
  BedroomAbvGr + GrLivArea + KitchenAbvGr + TotRmsAbvGrd + GarageCars +
  PoolArea + LotArea, data = train)
summary(custom_fit)
confint(custom_fit)
par(mfrow = c(2, 2))
plot(custom_fit)
ncvTest(custom_fit)

```

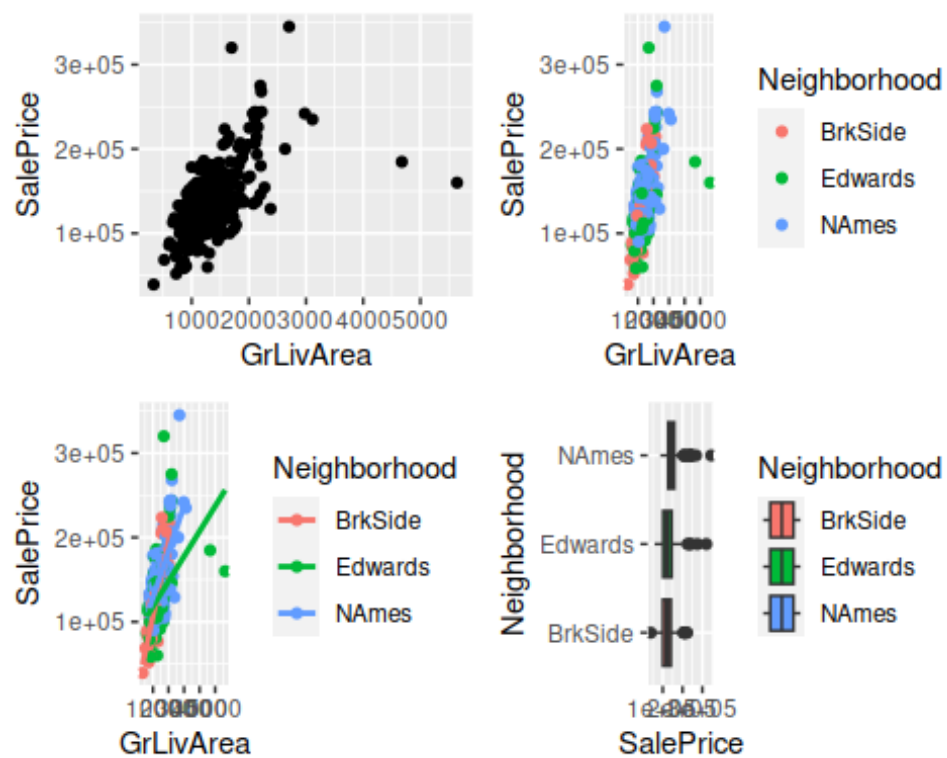
```
vif(custom_fit)
```

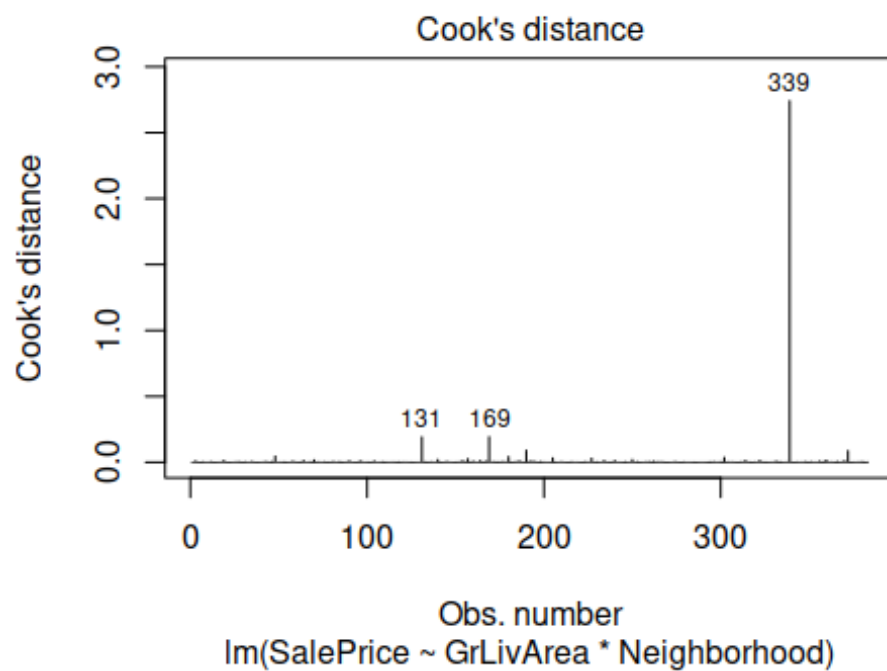
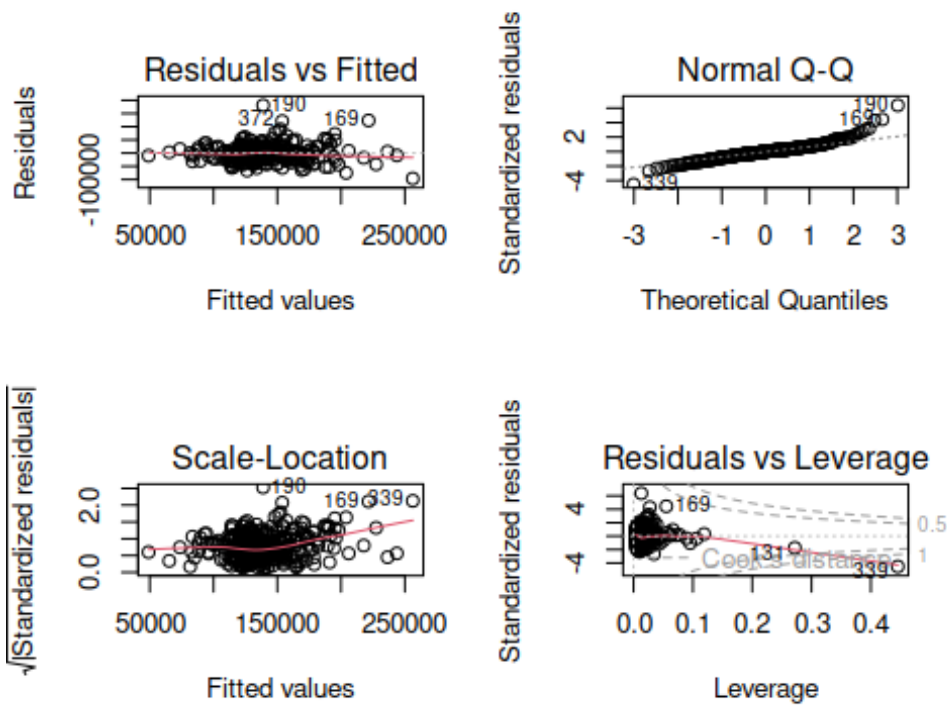
```
outlierTest(custom_fit)
```

```
plot(custom_fit, 4)
```

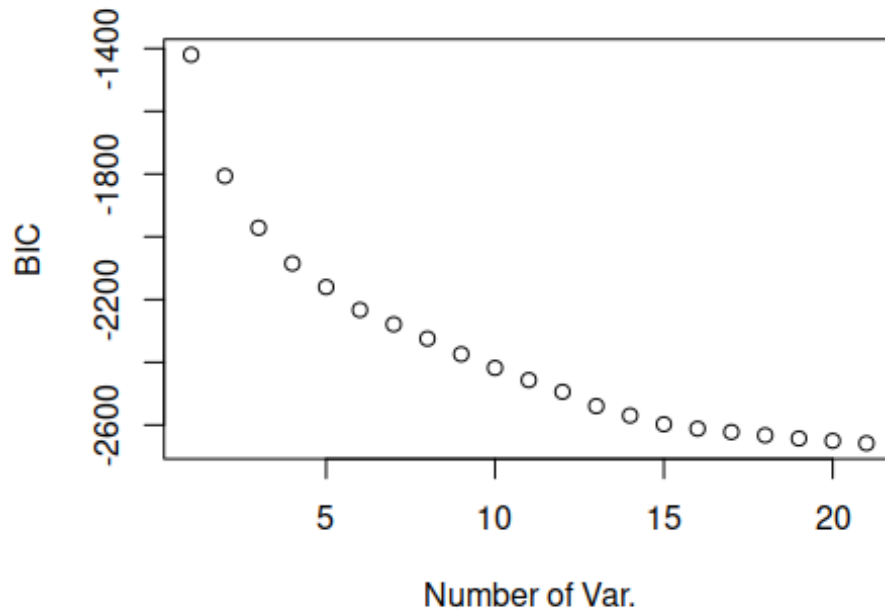
Plots and outcomes of tests for models

Q1





Q2 forward



```
coef(forward_model, which.min(summary(forward_model)$bic))
```

```
##      (Intercept)      MSSubClass      LotArea
##      -7.662740e+05      -3.113764e+02      1.117732e+00
##      LotShapeIR3 NeighborhoodCrawfor NeighborhoodNoRidge
##      -2.689683e+04      3.839663e+04      8.168489e+04
##      NeighborhoodNrldgHt NeighborhoodStoneBr Condition1Norm
##      6.588392e+04      7.825019e+04      3.723630e+03
##      Condition2PosN BldgType2fmCon OverallQual
##      -8.632146e+04      3.977891e+04      2.656121e+04
##      OverallCond      YearBuilt      RoofMatlWdShngl
##      5.228468e+03      3.546448e+02      1.122659e+05
##      Exterior2ndHdBoard BsmtFinSF2      FullBath
##      -9.573793e+03      1.301091e+01      2.754496e+04
##      HalfBath      KitchenQualGd      PavedDriveY
##      1.706036e+04      -8.868306e+03      5.384316e+03
##      SaleConditionAdjLand
##      -1.389608e+04
```

```
fit_forward <- lm(SalePrice ~ MSSubClass + LotArea + LotShape + Neighborhood
+
                    Condition1 + Condition2 + BldgType + OverallQual +
                    OverallCond + YearBuilt + RoofMatl + Exterior2nd +
```

```

BsmtFinSF2 +
          FullBath + FullBath + KitchenQual + PavedDrive +
SaleCondition,
          data = train)

summary(fit_forward)

##
## Call:
## lm(formula = SalePrice ~ MSSubClass + LotArea + LotShape + Neighborhood +
##      Condition1 + Condition2 + BldgType + OverallQual + OverallCond +
##      YearBuilt + RoofMatl + Exterior2nd + BsmtFinSF2 + FullBath +
##      FullBath + KitchenQual + PavedDrive + SaleCondition, data = train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -109998  -18011   -1243    15277   273254
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -8.953e+05  1.726e+05  -5.186 2.47e-07 ***
## MSSubClass      6.082e+01  4.805e+01   1.266 0.205879
## LotArea        9.744e-01  1.131e-01   8.612 < 2e-16 ***
## LotShapeIR2     5.947e+03  5.916e+03   1.005 0.314968
## LotShapeIR3     5.585e+03  1.234e+04   0.452 0.651020
## LotShapeReg    -3.977e+03  2.207e+03  -1.802 0.071737 .
## NeighborhoodBlueste -1.143e+03  2.643e+04  -0.043 0.965508
## NeighborhoodBrDale  -4.812e+03  1.387e+04  -0.347 0.728760
## NeighborhoodBrkSide -1.002e+04  1.166e+04  -0.860 0.390076
## NeighborhoodClearCr  -5.307e+03  1.239e+04  -0.428 0.668536
## NeighborhoodCollgCr -1.962e+04  9.763e+03  -2.010 0.044647 *
## NeighborhoodCrawfor  2.429e+04  1.142e+04   2.127 0.033613 *
## NeighborhoodEdwards -1.814e+04  1.060e+04  -1.711 0.087253 .
## NeighborhoodGilbert  -3.130e+04  1.049e+04  -2.985 0.002886 **
## NeighborhoodIDOTRR  -2.156e+04  1.224e+04  -1.762 0.078331 .
## NeighborhoodMeadowV  -6.161e+03  1.446e+04  -0.426 0.670189
## NeighborhoodMitchel -1.617e+04  1.094e+04  -1.477 0.139876
## NeighborhoodNames   -1.233e+04  1.038e+04  -1.188 0.235122
## NeighborhoodNoRidge  6.890e+04  1.114e+04   6.184 8.21e-10 ***
## NeighborhoodNPkVill  -4.619e+03  1.772e+04  -0.261 0.794422
## NeighborhoodNridgHt  3.347e+04  9.987e+03   3.352 0.000824 ***
## NeighborhoodNWAmes  -1.910e+04  1.081e+04  -1.766 0.077540 .
## NeighborhoodOldTown -2.128e+04  1.125e+04  -1.892 0.058686 .
## NeighborhoodSawyer   -1.315e+04  1.097e+04  -1.199 0.230822
## NeighborhoodSawyerW  -9.949e+03  1.057e+04  -0.941 0.346907
## NeighborhoodSomerst  -1.266e+04  9.850e+03  -1.285 0.198905
## NeighborhoodStoneBr  4.740e+04  1.148e+04   4.129 3.86e-05 ***
## NeighborhoodSWISU    -1.072e+04  1.276e+04  -0.840 0.401030
## NeighborhoodTimber   -1.079e+04  1.118e+04  -0.965 0.334702
## NeighborhoodVeenker  1.533e+04  1.448e+04   1.058 0.290105

```

## Condition1Feedr	6.886e+03	6.926e+03	0.994	0.320300	
## Condition1Norm	8.965e+03	5.679e+03	1.579	0.114649	
## Condition1PosA	1.568e+04	1.400e+04	1.120	0.262798	
## Condition1PosN	2.062e+04	1.034e+04	1.994	0.046362	*
## Condition1RR Ae	-1.030e+04	1.230e+04	-0.837	0.402487	
## Condition1RR An	9.427e+03	9.465e+03	0.996	0.319433	
## Condition1RR Ne	-9.345e+03	2.567e+04	-0.364	0.715881	
## Condition1RR Nn	2.688e+03	1.727e+04	0.156	0.876352	
## Condition2Feedr	-3.362e+04	3.037e+04	-1.107	0.268549	
## Condition2Norm	-1.880e+04	2.623e+04	-0.717	0.473643	
## Condition2PosA	3.949e+04	4.376e+04	0.903	0.366933	
## Condition2PosN	-1.642e+05	3.775e+04	-4.351	1.46e-05	***
## Condition2RR Ae	2.135e+04	4.446e+04	0.480	0.631215	
## Condition2RR An	-2.575e+04	4.414e+04	-0.583	0.559751	
## Condition2RR Nn	-3.863e+04	3.637e+04	-1.062	0.288292	
## BldgType2fmCon	-1.146e+04	9.621e+03	-1.191	0.233776	
## BldgTypeDuplex	-1.727e+04	6.317e+03	-2.733	0.006349	**
## BldgTypeTwnhs	-5.695e+04	8.901e+03	-6.398	2.15e-10	***
## BldgTypeTwnhsE	-4.501e+04	6.431e+03	-6.999	4.01e-12	***
## OverallQual	2.162e+04	1.183e+03	18.274	< 2e-16	***
## OverallCond	3.372e+03	9.919e+02	3.400	0.000694	***
## YearBuilt	3.480e+02	8.241e+01	4.223	2.57e-05	***
## RoofMatlCompShg	2.610e+05	3.855e+04	6.769	1.92e-11	***
## RoofMatlMembran	2.739e+05	5.278e+04	5.190	2.42e-07	***
## RoofMatlMetal	2.626e+05	5.243e+04	5.009	6.18e-07	***
## RoofMatlRoll	2.575e+05	5.246e+04	4.909	1.02e-06	***
## RoofMatlTar&Grv	2.600e+05	4.008e+04	6.488	1.21e-10	***
## RoofMatlWdShake	2.839e+05	4.187e+04	6.780	1.77e-11	***
## RoofMatlWdShngl	3.679e+05	4.091e+04	8.992	< 2e-16	***
## Exterior2ndAsphShn	-1.302e+04	2.176e+04	-0.598	0.549742	
## Exterior2ndBrk Cmn	-4.887e+03	1.921e+04	-0.254	0.799266	
## Exterior2ndBrkFace	-2.875e+03	1.096e+04	-0.262	0.793180	
## Exterior2ndCBlock	2.147e+04	3.595e+04	0.597	0.550342	
## Exterior2ndCmentBd	3.952e+03	1.030e+04	0.384	0.701355	
## Exterior2ndHdBoard	-1.121e+04	8.730e+03	-1.285	0.199171	
## Exterior2ndImStucc	2.330e+04	1.413e+04	1.649	0.099309	.
## Exterior2ndMetalSd	-5.569e+03	8.519e+03	-0.654	0.513362	
## Exterior2ndOther	-1.005e+04	3.647e+04	-0.276	0.782923	
## Exterior2ndPlywood	-7.141e+03	8.948e+03	-0.798	0.424944	
## Exterior2ndStone	-2.889e+04	1.801e+04	-1.603	0.109061	
## Exterior2ndStucco	1.241e+03	1.084e+04	0.114	0.908908	
## Exterior2ndVinylSd	-7.078e+03	8.643e+03	-0.819	0.412997	
## Exterior2ndWd Sdng	-8.695e+03	8.506e+03	-1.022	0.306831	
## Exterior2ndWd Shng	-2.214e+04	9.925e+03	-2.230	0.025878	*
## BsmtFinSF2	1.272e+01	6.041e+00	2.106	0.035417	*
## FullBath	2.348e+04	2.341e+03	10.031	< 2e-16	***
## KitchenQualFa	-6.267e+04	8.018e+03	-7.816	1.08e-14	***
## KitchenQualGd	-5.477e+04	4.474e+03	-12.242	< 2e-16	***
## KitchenQualTA	-6.117e+04	5.072e+03	-12.060	< 2e-16	***
## PavedDriveP	2.114e+03	7.493e+03	0.282	0.777912	


```

## PavedDriveY          6.745e+03  4.537e+03   1.487 0.137364
## SaleConditionAdjLand -3.380e+03  1.852e+04  -0.183 0.855212
## SaleConditionAlloca  2.326e+04  1.134e+04   2.050 0.040526 *
## SaleConditionFamily  6.838e+02  8.651e+03   0.079 0.937014
## SaleConditionNormal  8.357e+03  3.717e+03   2.248 0.024719 *
## SaleConditionPartial 2.710e+04  5.273e+03   5.140 3.14e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 34480 on 1374 degrees of freedom
## Multiple R-squared:  0.8226, Adjusted R-squared:  0.8116
## F-statistic: 74.96 on 85 and 1374 DF,  p-value: < 2.2e-16

```

```
confint(fit_forward)
```

```

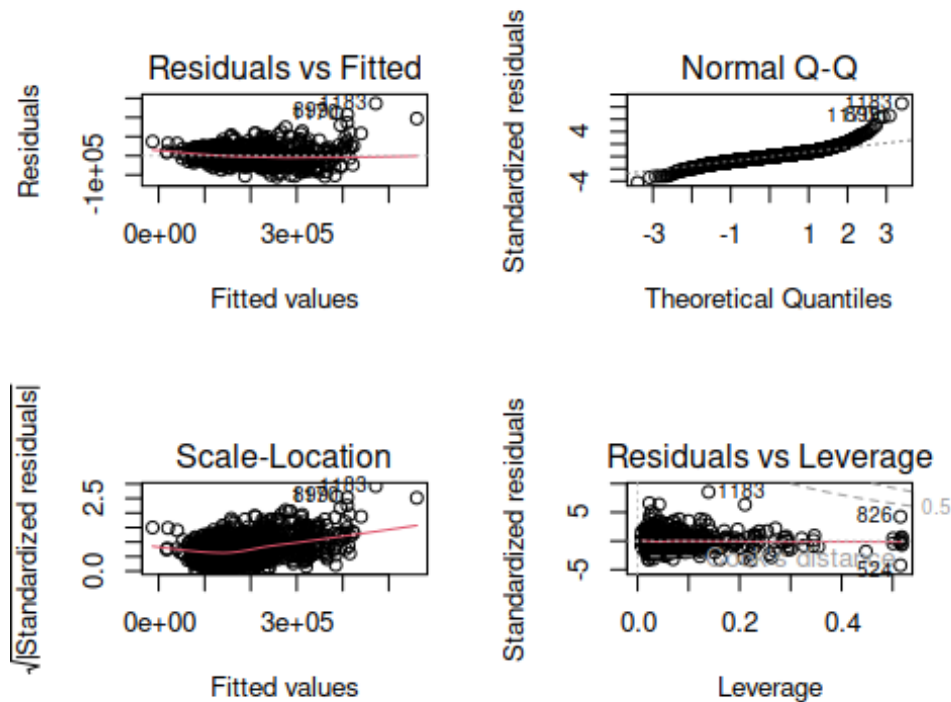
##              2.5 %          97.5 %
## (Intercept) -1.233920e+06 -5.566403e+05
## MSSubClass  -3.345067e+01  1.550813e+02
## LotArea      7.524319e-01  1.196338e+00
## LotShapeIR2  -5.658147e+03  1.755119e+04
## LotShapeIR3  -1.862938e+04  2.979886e+04
## LotShapeReg  -8.306440e+03  3.520365e+02
## NeighborhoodBlueste -5.299256e+04  5.070622e+04
## NeighborhoodBrDale  -3.202663e+04  2.240297e+04
## NeighborhoodBrkSide -3.288681e+04  1.284426e+04
## NeighborhoodClearCr -2.961753e+04  1.900327e+04
## NeighborhoodCollgCr -3.877278e+04 -4.697479e+02
## NeighborhoodCrawfor  1.885631e+03  4.668741e+04
## NeighborhoodEdwards -3.892549e+04  2.653704e+03
## NeighborhoodGilbert -5.187626e+04 -1.073101e+04
## NeighborhoodIDOTRR  -4.556760e+04  2.446649e+03
## NeighborhoodMeadowV -3.453362e+04  2.221146e+04
## NeighborhoodMitchel -3.763339e+04  5.303173e+03
## NeighborhoodNames  -3.268467e+04  8.031399e+03
## NeighborhoodNoRidge  4.704707e+04  9.076085e+04
## NeighborhoodNPkVill -3.938244e+04  3.014524e+04
## NeighborhoodNridgHt  1.388361e+04  5.306521e+04
## NeighborhoodNWAmes  -4.030877e+04  2.110767e+03
## NeighborhoodOldTown -4.334286e+04  7.825476e+02
## NeighborhoodSawyer  -3.468099e+04  8.371777e+03
## NeighborhoodSawyerW -3.069128e+04  1.079320e+04
## NeighborhoodSomerst -3.198236e+04  6.662395e+03
## NeighborhoodStoneBr  2.488456e+04  6.992260e+04
## NeighborhoodSWISU   -3.574003e+04  1.430819e+04
## NeighborhoodTimber  -3.272923e+04  1.114564e+04
## NeighborhoodVeenker -1.308496e+04  4.374235e+04
## Condition1Feedr    -6.701280e+03  2.047388e+04
## Condition1Norm     -2.175304e+03  2.010496e+04
## Condition1PosA     -1.178000e+04  4.314728e+04
## Condition1PosN      3.328211e+02  4.089792e+04

```

## Condition1RR Ae	-3.443715e+04	1.383131e+04
## Condition1RR An	-9.140060e+03	2.799333e+04
## Condition1RR Ne	-5.970057e+04	4.101086e+04
## Condition1RR Nn	-3.118929e+04	3.656429e+04
## Condition2Feedr	-9.319894e+04	2.596327e+04
## Condition2Norm	-7.026024e+04	3.265660e+04
## Condition2PosA	-4.634810e+04	1.253363e+05
## Condition2PosN	-2.383062e+05	-9.018892e+04
## Condition2RR Ae	-6.587171e+04	1.085653e+05
## Condition2RR An	-1.123503e+05	6.084633e+04
## Condition2RR Nn	-1.099676e+05	3.270700e+04
## BldgType2fmCon	-3.033342e+04	7.412609e+03
## BldgTypeDuplex	-2.965846e+04	-4.874771e+03
## BldgTypeTwnhs	-7.441602e+04	-3.949293e+04
## BldgTypeTwnhsE	-5.762596e+04	-3.239593e+04
## OverallQual	1.929698e+04	2.393834e+04
## OverallCond	1.426510e+03	5.318188e+03
## YearBuilt	1.863195e+02	5.096500e+02
## RoofMatlCompShg	1.853438e+05	3.366084e+05
## RoofMatlMembran	1.703856e+05	3.774496e+05
## RoofMatlMetal	1.597781e+05	3.654851e+05
## RoofMatlRoll	1.546093e+05	3.604204e+05
## RoofMatlTar&Grv	1.814099e+05	3.386533e+05
## RoofMatlWdShake	2.017628e+05	3.660382e+05
## RoofMatlWdShngl	2.876079e+05	4.481052e+05
## Exterior2ndAsphShn	-5.569947e+04	2.966443e+04
## Exterior2ndBrk Cmn	-4.257583e+04	3.280253e+04
## Exterior2ndBrkFace	-2.438384e+04	1.863341e+04
## Exterior2ndCBlock	-4.904136e+04	9.198956e+04
## Exterior2ndCmentBd	-1.626067e+04	2.416520e+04
## Exterior2ndHdBoard	-2.834084e+04	5.911846e+03
## Exterior2ndImStucc	-4.412613e+03	5.101127e+04
## Exterior2ndMetalSd	-2.228010e+04	1.114152e+04
## Exterior2ndOther	-8.159801e+04	6.149698e+04
## Exterior2ndPlywood	-2.469412e+04	1.041147e+04
## Exterior2ndStone	-6.422443e+04	6.453168e+03
## Exterior2ndStucco	-2.002962e+04	2.251130e+04
## Exterior2ndVinylSd	-2.403351e+04	9.877759e+03
## Exterior2ndWd Sdng	-2.538125e+04	7.990568e+03
## Exterior2ndWd Shng	-4.160578e+04	-2.667479e+03
## BsmtFinSF2	8.695797e-01	2.457054e+01
## FullBath	1.889100e+04	2.807603e+04
## KitchenQualFa	-7.840199e+04	-4.694312e+04
## KitchenQualGd	-6.354442e+04	-4.599177e+04
## KitchenQualTA	-7.111947e+04	-5.122025e+04
## PavedDriveP	-1.258569e+04	1.681335e+04
## PavedDriveY	-2.155896e+03	1.564583e+04
## SaleConditionAdjLand	-3.971177e+04	3.295152e+04
## SaleConditionAlloca	1.004932e+03	4.551118e+04
## SaleConditionFamily	-1.628713e+04	1.765466e+04

```
## SaleConditionNormal 1.065091e+03 1.564861e+04
## SaleConditionPartial 1.675898e+04 3.744631e+04
```

```
Forward_Adjusted_R2 <- 0.812
```



```
dev.off()

## null device
##          1

ncvTest(fit_forward)

## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 986.1211, Df = 1, p = < 2.22e-16

vif(fit_forward)

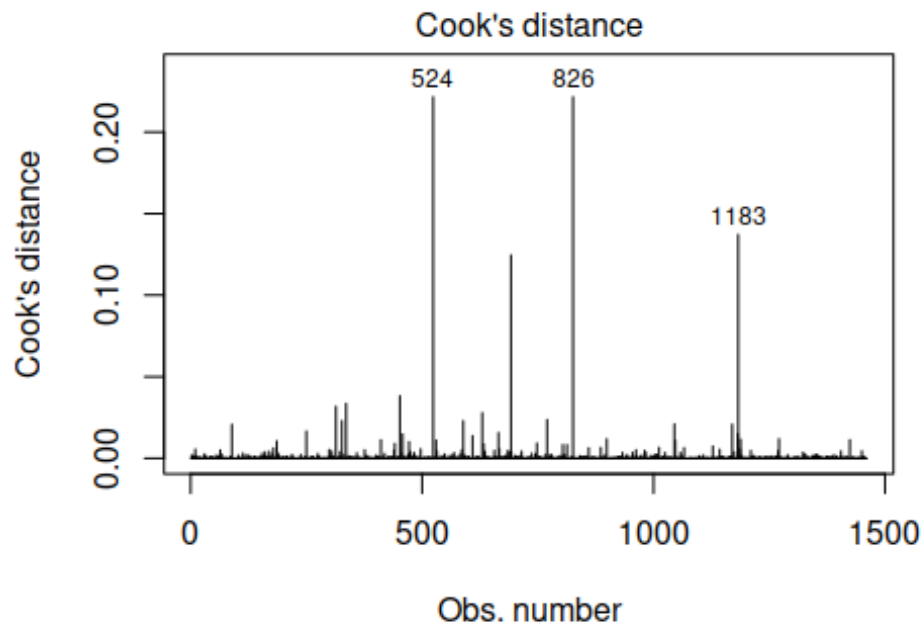
##              GVIF Df GVIF^(1/(2*Df))
## MSSubClass    5.071101  1    2.251911
## LotArea       1.565287  1    1.251114
## LotShape      1.864085  3    1.109373
## Neighborhood 577.952343 24    1.141667
## Condition1    3.156006  8    1.074475
## Condition2    2.305334  7    1.061475
## BldgType     16.150575  4    1.415870
```

```
## OverallQual      3.285278  1      1.812534
## OverallCond      1.495366  1      1.222852
## YearBuilt        7.603801  1      2.757499
## RoofMatl         2.167505  7      1.056811
## Exterior2nd     22.937318 15      1.110073
## BsmtFinSF2       1.165586  1      1.079623
## FullBath         2.041604  1      1.428847
## KitchenQual      3.811692  3      1.249836
## PavedDrive       1.589166  2      1.122774
## SaleCondition     2.143455  5      1.079224
```

```
outlierTest(fit_forward)
```

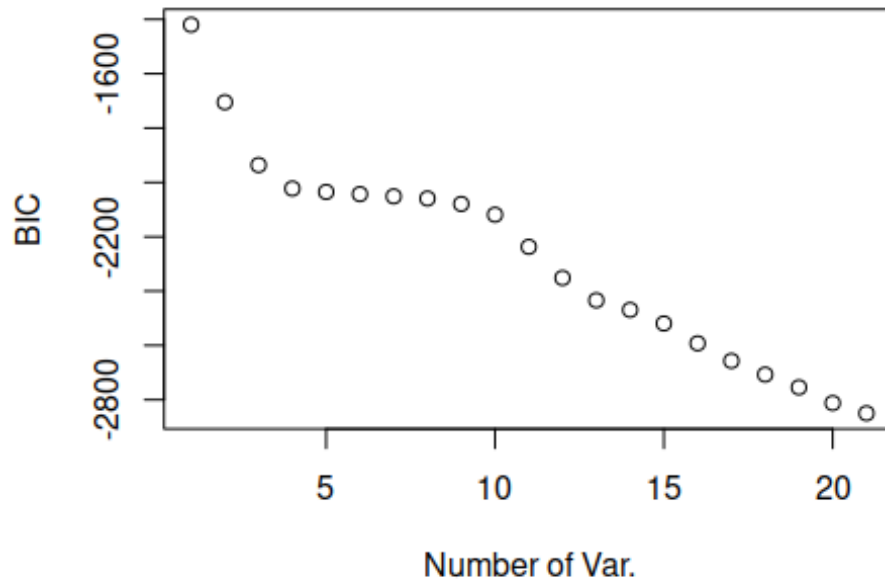
```
##      rstudent unadjusted p-value Bonferroni p
## 1183  8.774990      4.9728e-18  7.2155e-15
## 899   6.684378      3.3630e-11  4.8797e-08
## 1170  6.463254      1.4201e-10  2.0606e-07
## 692   6.428361      1.7754e-10  2.5762e-07
## 804   5.133320      3.2556e-07  4.7239e-04
## 441   4.750599      2.2409e-06  3.2515e-03
## 1143  4.401285      1.1594e-05  1.6823e-02
## 524   -4.259017      2.1930e-05  3.1820e-02
## 826   4.259017      2.1930e-05  3.1820e-02
```

```
plot(fit_forward, 4)
```



```
salePrice ~ MSSubClass + LotArea + LotShape + Neighborhood + Co
```

Q2 backward



```
coef(backward_model, which.min(summary(backward_model)$bic))
```

```
##           (Intercept)           LotArea  NeighborhoodNoRidge
##      -8.061159e+05      1.402461e+00      7.738422e+04
## NeighborhoodNridgHt Condition2PosN      OverallQual
##      5.453875e+04      -9.268434e+04      3.310640e+04
##      OverallCond           YearBuilt      RoofMatlCompShg
##      2.996439e+03      2.500979e+02      2.557807e+05
##      RoofMatlMembran      RoofMatlMetal      RoofMatlRoll
##      2.456086e+05      2.489145e+05      2.725861e+05
##      RoofMatlTar&Grv      RoofMatlWdShake      RoofMatlWdShngl
##      2.631555e+05      2.639996e+05      3.434576e+05
##      BsmtFinSF2           LowQualFinSF      BsmtFullBath
##      4.424551e+00      2.558645e+01      1.413706e+04
##      KitchenQualTA      FunctionalMaj2      FunctionalMin1
##      -1.308605e+04      -3.285878e+04      2.630804e+03
## SaleConditionAdjLand
##      -1.853191e+04
```

```
fit_backward <- lm(SalePrice ~ LotArea + Neighborhood + Condition2 +
OverallQual + RoofMatl +
                    OverallCond + YearBuilt + BsmtFinSF2 +LowQualFinSF+
                    BsmtFullBath + KitchenQual + Functional + SaleCondition,
```

```

data = train)
summary(fit_backward)

##
## Call:
## lm(formula = SalePrice ~ LotArea + Neighborhood + Condition2 +
##     OverallQual + RoofMatl + OverallCond + YearBuilt + BsmtFinSF2 +
##     LowQualFinSF + BsmtFullBath + KitchenQual + Functional +
##     SaleCondition, data = train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -146740  -19944   -1872   16414   311056
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -8.071e+05  1.631e+05  -4.947  8.44e-07 ***
## LotArea       1.192e+00  1.109e-01  10.749  < 2e-16 ***
## NeighborhoodBlueste -6.719e+03  2.771e+04  -0.242  0.808470
## NeighborhoodBrDale  -2.021e+04  1.325e+04  -1.525  0.127559
## NeighborhoodBrkSide  1.566e+04  1.159e+04   1.351  0.176796
## NeighborhoodClearCr  2.535e+04  1.235e+04   2.053  0.040258 *
## NeighborhoodCollgCr  1.377e+04  9.498e+03   1.450  0.147266
## NeighborhoodCrawfor  5.270e+04  1.134e+04   4.649  3.65e-06 ***
## NeighborhoodEdwards  6.089e+03  1.056e+04   0.577  0.564364
## NeighborhoodGilbert  1.519e+04  9.943e+03   1.528  0.126689
## NeighborhoodIDOTRR   2.776e+03  1.218e+04   0.228  0.819743
## NeighborhoodMeadowV -6.680e+03  1.325e+04  -0.504  0.614200
## NeighborhoodMitchel  9.437e+03  1.080e+04   0.874  0.382400
## NeighborhoodNames    1.334e+04  1.003e+04   1.330  0.183710
## NeighborhoodNoRidge  1.088e+05  1.083e+04  10.047  < 2e-16 ***
## NeighborhoodNPkVill  5.121e+03  1.543e+04   0.332  0.740079
## NeighborhoodNridgHt  5.216e+04  1.011e+04   5.160  2.83e-07 ***
## NeighborhoodNWAmes   2.375e+04  1.033e+04   2.299  0.021651 *
## NeighborhoodOldTown  5.699e+03  1.123e+04   0.507  0.611961
## NeighborhoodSawyer   1.087e+04  1.065e+04   1.021  0.307510
## NeighborhoodSawyerW  1.806e+04  1.032e+04   1.750  0.080378 .
## NeighborhoodSomerst  1.394e+04  9.776e+03   1.426  0.154138
## NeighborhoodStoneBr  5.727e+04  1.176e+04   4.871  1.24e-06 ***
## NeighborhoodSWISU    2.182e+04  1.307e+04   1.670  0.095201 .
## NeighborhoodTimber   2.074e+04  1.106e+04   1.874  0.061104 .
## NeighborhoodVeenker  2.923e+04  1.464e+04   1.997  0.046068 *
## Condition2Feedr     -2.282e+04  3.030e+04  -0.753  0.451556
## Condition2Norm      -1.449e+04  2.631e+04  -0.551  0.581891
## Condition2PosA       5.783e+04  4.575e+04   1.264  0.206400
## Condition2PosN      -1.383e+05  3.753e+04  -3.684  0.000238 ***
## Condition2RR Ae     3.572e+04  4.547e+04   0.786  0.432240
## Condition2RRAn      -2.244e+04  4.540e+04  -0.494  0.621123
## Condition2RRNn      -3.057e+04  3.706e+04  -0.825  0.409619
## OverallQual         2.585e+04  1.181e+03  21.889  < 2e-16 ***

```

```

## RoofMatlCompShg      2.807e+05  3.814e+04  7.361 3.10e-13 ***
## RoofMatlMembran      2.535e+05  5.478e+04  4.627 4.06e-06 ***
## RoofMatlMetal        2.700e+05  5.354e+04  5.042 5.20e-07 ***
## RoofMatlRoll         2.941e+05  5.309e+04  5.540 3.61e-08 ***
## RoofMatlTar&Grv      2.888e+05  3.978e+04  7.260 6.41e-13 ***
## RoofMatlWdShake      3.017e+05  4.168e+04  7.238 7.49e-13 ***
## RoofMatlWdShngl      3.715e+05  4.067e+04  9.133 < 2e-16 ***
## OverallCond          2.810e+03  1.038e+03  2.708 0.006843 **
## YearBuilt            2.809e+02  7.670e+01  3.663 0.000259 ***
## BsmtFinSF2           7.246e+00  6.435e+00  1.126 0.260360
## LowQualFinSF         2.440e+01  2.157e+01  1.131 0.258275
## BsmtFullBath         1.137e+04  2.041e+03  5.573 3.00e-08 ***
## KitchenQualFa        -6.703e+04  8.406e+03  -7.975 3.15e-15 ***
## KitchenQualGd        -5.672e+04  4.627e+03  -12.258 < 2e-16 ***
## KitchenQualTA        -6.583e+04  5.285e+03  -12.457 < 2e-16 ***
## FunctionalMaj2       -2.937e+04  1.949e+04  -1.507 0.132051
## FunctionalMin1        6.697e+03  1.221e+04  0.549 0.583349
## FunctionalMin2       1.524e+04  1.206e+04  1.264 0.206370
## FunctionalMod         2.469e+04  1.413e+04  1.748 0.080641 .
## FunctionalSev        -3.416e+04  4.041e+04  -0.845 0.398068
## FunctionalTyp         2.693e+03  1.027e+04  0.262 0.793176
## SaleConditionAdjLand -1.110e+03  1.936e+04  -0.057 0.954284
## SaleConditionAlloca  2.144e+04  1.158e+04  1.851 0.064309 .
## SaleConditionFamily   4.533e+03  9.103e+03  0.498 0.618556
## SaleConditionNormal   1.013e+04  3.929e+03  2.579 0.010020 *
## SaleConditionPartial  3.720e+04  5.495e+03  6.769 1.90e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 36710 on 1400 degrees of freedom
## Multiple R-squared:  0.7951, Adjusted R-squared:  0.7865
## F-statistic: 92.08 on 59 and 1400 DF,  p-value: < 2.2e-16

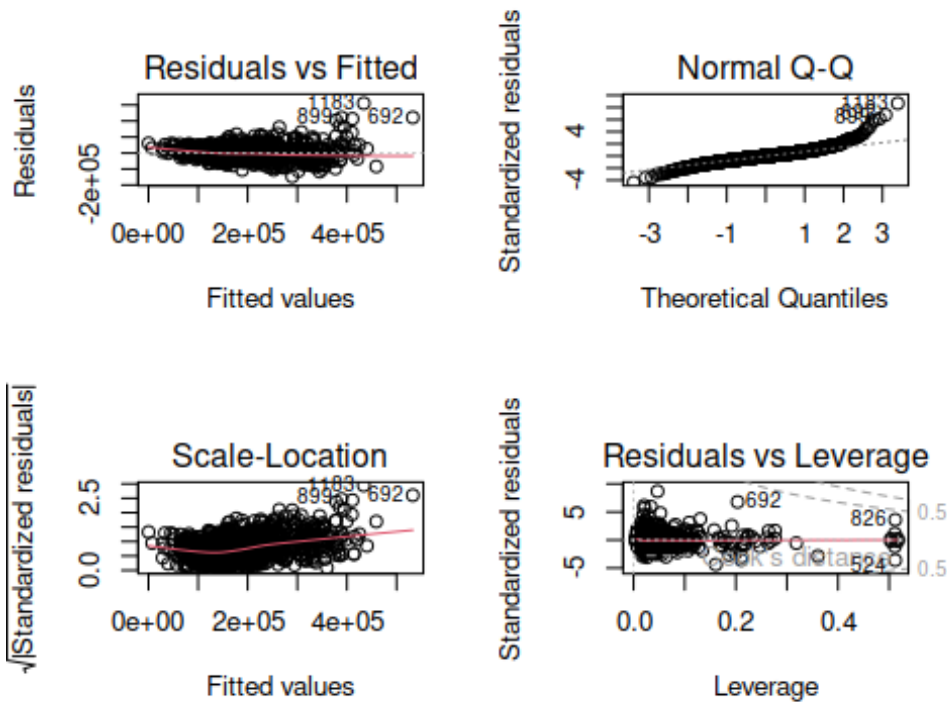
confint(fit_backward)

##              2.5 %          97.5 %
## (Intercept) -1.127131e+06 -4.870680e+05
## LotArea      9.744272e-01  1.409496e+00
## NeighborhoodBlueste -6.108523e+04  4.764681e+04
## NeighborhoodBrDale  -4.620764e+04  5.791373e+03
## NeighborhoodBrkSide -7.074453e+03  3.840423e+04
## NeighborhoodClearCr  1.127614e+03  4.956282e+04
## NeighborhoodCollgCr -4.859188e+03  3.240466e+04
## NeighborhoodCrawfor  3.046443e+04  7.494243e+04
## NeighborhoodEdwards -1.462950e+04  2.680728e+04
## NeighborhoodGilbert -4.309759e+03  3.469840e+04
## NeighborhoodIDOTRR  -2.111375e+04  2.666501e+04
## NeighborhoodMeadowV -3.267058e+04  1.931012e+04
## NeighborhoodMitchel -1.175011e+04  3.062450e+04
## NeighborhoodNames  -6.336747e+03  3.302625e+04

```

## NeighborhoodNoRidge	8.753605e+04	1.300124e+05
## NeighborhoodNPkVill	-2.515209e+04	3.539332e+04
## NeighborhoodNridgHt	3.233046e+04	7.199535e+04
## NeighborhoodNWAmes	3.485312e+03	4.401829e+04
## NeighborhoodOldTown	-1.633479e+04	2.773311e+04
## NeighborhoodSawyer	-1.001862e+04	3.175978e+04
## NeighborhoodSawyerW	-2.187566e+03	3.831600e+04
## NeighborhoodSomerst	-5.238385e+03	3.311740e+04
## NeighborhoodStoneBr	3.420296e+04	8.032816e+04
## NeighborhoodSWISU	-3.814563e+03	4.744642e+04
## NeighborhoodTimber	-9.670997e+02	4.244169e+04
## NeighborhoodVeenker	5.106755e+02	5.794431e+04
## Condition2Feedr	-8.225186e+04	3.661989e+04
## Condition2Norm	-6.609324e+04	3.711630e+04
## Condition2PosA	-3.191294e+04	1.475800e+05
## Condition2PosN	-2.119215e+05	-6.466087e+04
## Condition2RRAE	-5.347912e+04	1.249252e+05
## Condition2RRAN	-1.115067e+05	6.661678e+04
## Condition2RRNn	-1.032770e+05	4.213585e+04
## OverallQual	2.353496e+04	2.816846e+04
## RoofMatlCompShg	2.059160e+05	3.555431e+05
## RoofMatlMembran	1.459956e+05	3.609253e+05
## RoofMatlMetal	1.649397e+05	3.749835e+05
## RoofMatlRoll	1.899857e+05	3.982876e+05
## RoofMatlTar&Grv	2.107547e+05	3.668215e+05
## RoofMatlWdShake	2.199022e+05	3.834192e+05
## RoofMatlWdShngl	2.916685e+05	4.512389e+05
## OverallCond	7.748509e+02	4.845696e+03
## YearBuilt	1.304919e+02	4.314015e+02
## BsmtFinSF2	-5.377733e+00	1.986996e+01
## LowQualFinSF	-1.792223e+01	6.672065e+01
## BsmtFullBath	7.370865e+03	1.537837e+04
## KitchenQualFa	-8.352283e+04	-5.054372e+04
## KitchenQualGd	-6.579762e+04	-4.764291e+04
## KitchenQualTA	-7.620235e+04	-5.546706e+04
## FunctionalMaj2	-6.759676e+04	8.861629e+03
## FunctionalMin1	-1.724929e+04	3.064380e+04
## FunctionalMin2	-8.408607e+03	3.889083e+04
## FunctionalMod	-3.014621e+03	5.240398e+04
## FunctionalSev	-1.134396e+05	4.511418e+04
## FunctionalTyp	-1.745496e+04	2.284193e+04
## SaleConditionAdjLand	-3.909531e+04	3.687474e+04
## SaleConditionAlloca	-1.275476e+03	4.414780e+04
## SaleConditionFamily	-1.332308e+04	2.238947e+04
## SaleConditionNormal	2.423960e+03	1.783781e+04
## SaleConditionPartial	2.641714e+04	4.797498e+04

Backward_Adjusted_R2 <- 0.7865



```
dev.off()

## null device
##          1

ncvTest(fit_backward)

## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 987.3706, Df = 1, p = < 2.22e-16

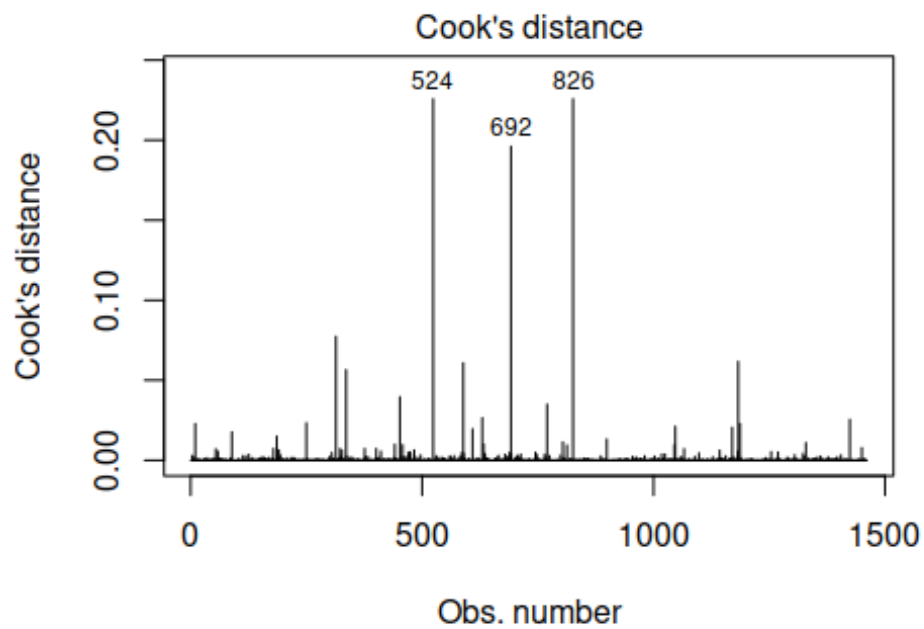
vif(fit_backward)

##              GVIF Df GVIF^(1/(2*Df))
## LotArea        1.326381  1      1.151686
## Neighborhood  28.191307 24      1.072039
## Condition2     1.294373  7      1.018601
## OverallQual    2.888295  1      1.699499
## RoofMatl       1.770769  7      1.041660
## OverallCond    1.443390  1      1.201412
## YearBuilt      5.809656  1      2.410323
## BsmtFinSF2     1.166801  1      1.080186
## LowQualFinSF   1.191370  1      1.091499
## BsmtFullBath   1.214397  1      1.101997
## KitchenQual    3.312330  3      1.220925
## Functional     1.599519  6      1.039918
## SaleCondition  1.721596  5      1.055828
```

```
outlierTest(fit_backward)
```

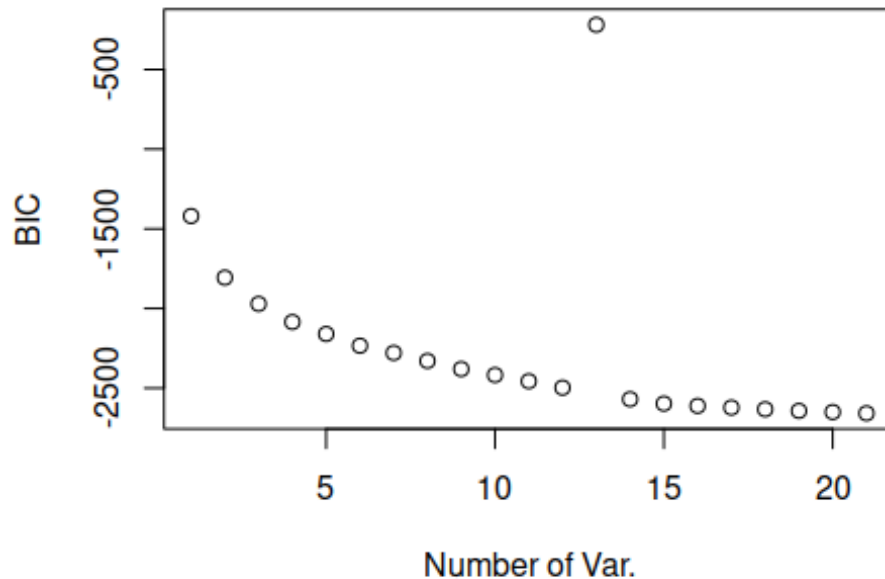
##		rstudent	unadjusted p-value	Bonferroni p
##	1183	8.919143	1.4362e-18	2.0854e-15
##	692	6.896715	8.0345e-12	1.1666e-08
##	899	6.207569	7.0773e-10	1.0276e-06
##	1170	6.002381	2.4734e-09	3.5914e-06
##	804	5.688194	1.5612e-08	2.2668e-05
##	1047	4.827028	1.5379e-06	2.2330e-03
##	441	4.467253	8.5611e-06	1.2431e-02
##	1143	4.459566	8.8695e-06	1.2878e-02
##	589	-4.391850	1.2085e-05	1.7548e-02

```
plot(fit_backward, 4)
```



salePrice ~ LotArea + Neighborhood + Condition2 + OverallQual + R

Q3 best fit



```
coef(best_model, which.min(summary(best_model)$bic))
```

```
##      (Intercept)      MSSubClass      LotArea
##      -7.662740e+05      -3.113764e+02      1.117732e+00
##      LotShapeIR3 NeighborhoodCrawfor NeighborhoodNoRidge
##      -2.689683e+04      3.839663e+04      8.168489e+04
## NeighborhoodNridgHt NeighborhoodStoneBr Condition1Norm
##      6.588392e+04      7.825019e+04      3.723630e+03
## Condition2PosN BldgType2fmCon OverallQual
##      -8.632146e+04      3.977891e+04      2.656121e+04
## OverallCond YearBuilt RoofMatlWdShngl
##      5.228468e+03      3.546448e+02      1.122659e+05
## Exterior2ndHdBoard BsmtFinSF2 FullBath
##      -9.573793e+03      1.301091e+01      2.754496e+04
## HalfBath KitchenQualGd PavedDriveY
##      1.706036e+04      -8.868306e+03      5.384316e+03
## SaleConditionAdjLand
##      -1.389608e+04
```

```
fit_best <- lm(SalePrice ~ MSSubClass + LotArea + LotShape + Neighborhood +
Condition1 +
      Condition2 + BldgType + OverallQual + OverallCond +
YearBuilt +
```

```

RoofMatl + Exterior2nd + BsmtFinSF2 + FullBath + HalfBath +
KitchenQual + PavedDrive + SaleCondition, data = train)
summary(fit_best)

##
## Call:
## lm(formula = SalePrice ~ MSSubClass + LotArea + LotShape + Neighborhood +
##   Condition1 + Condition2 + BldgType + OverallQual + OverallCond +
##   YearBuilt + RoofMatl + Exterior2nd + BsmtFinSF2 + FullBath +
##   HalfBath + KitchenQual + PavedDrive + SaleCondition, data = train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -116889  -18325   -1233   14141  269800
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -8.180e+05  1.705e+05  -4.797 1.79e-06 ***
## MSSubClass     -7.714e+01  5.194e+01  -1.485 0.137710
## LotArea        9.693e-01  1.115e-01   8.693 < 2e-16 ***
## LotShapeIR2     5.264e+03  5.831e+03   0.903 0.366800
## LotShapeIR3     5.891e+03  1.216e+04   0.484 0.628245
## LotShapeReg    -3.757e+03  2.175e+03  -1.727 0.084335 .
## NeighborhoodBlueste -9.779e+03  2.608e+04  -0.375 0.707757
## NeighborhoodBrDale -1.372e+04  1.374e+04  -0.999 0.318210
## NeighborhoodBrkSide -8.491e+03  1.149e+04  -0.739 0.459988
## NeighborhoodClearCr -4.408e+03  1.221e+04  -0.361 0.718190
## NeighborhoodCollgCr -2.030e+04  9.621e+03  -2.110 0.035009 *
## NeighborhoodCrawfor  2.243e+04  1.126e+04   1.993 0.046468 *
## NeighborhoodEdwards -1.737e+04  1.044e+04  -1.663 0.096491 .
## NeighborhoodGilbert -3.601e+04  1.036e+04  -3.476 0.000525 ***
## NeighborhoodIDOTRR  -2.072e+04  1.206e+04  -1.718 0.086040 .
## NeighborhoodMeadowV -8.834e+03  1.426e+04  -0.620 0.535676
## NeighborhoodMitchel -1.434e+04  1.079e+04  -1.329 0.184025
## NeighborhoodNames  -1.324e+04  1.023e+04  -1.294 0.195847
## NeighborhoodNoRidge  6.320e+04  1.102e+04   5.738 1.18e-08 ***
## NeighborhoodNPkVill -7.894e+03  1.747e+04  -0.452 0.651456
## NeighborhoodNridgHt  3.317e+04  9.842e+03   3.370 0.000771 ***
## NeighborhoodNWAmes  -2.365e+04  1.068e+04  -2.215 0.026957 *
## NeighborhoodOldTown -2.108e+04  1.108e+04  -1.902 0.057429 .
## NeighborhoodSawyer  -1.347e+04  1.081e+04  -1.245 0.213269
## NeighborhoodSawyerW -1.153e+04  1.042e+04  -1.106 0.268860
## NeighborhoodSomerst -1.550e+04  9.717e+03  -1.595 0.110928
## NeighborhoodStoneBr  4.655e+04  1.131e+04   4.114 4.11e-05 ***
## NeighborhoodSWISU   -1.082e+04  1.257e+04  -0.861 0.389592
## NeighborhoodTimber  -1.097e+04  1.102e+04  -0.996 0.319512
## NeighborhoodVeenker  1.074e+04  1.429e+04   0.752 0.452386
## Condition1Feedr     4.412e+03  6.836e+03   0.645 0.518844
## Condition1Norm       7.843e+03  5.599e+03   1.401 0.161507
## Condition1PosA      1.349e+04  1.380e+04   0.977 0.328659

```

## Condition1PosN	1.923e+04	1.019e+04	1.887	0.059318	.
## Condition1RR Ae	-1.354e+04	1.213e+04	-1.116	0.264614	
## Condition1RR An	8.640e+03	9.328e+03	0.926	0.354463	
## Condition1RR Ne	-1.606e+04	2.532e+04	-0.635	0.525841	
## Condition1RR Nn	4.758e+03	1.702e+04	0.280	0.779897	
## Condition2Feedr	-2.877e+04	2.994e+04	-0.961	0.336699	
## Condition2Norm	-1.704e+04	2.585e+04	-0.659	0.509834	
## Condition2PosA	2.952e+04	4.315e+04	0.684	0.493956	
## Condition2PosN	-1.634e+05	3.720e+04	-4.392	1.21e-05	***
## Condition2RR Ae	-6.156e+02	4.395e+04	-0.014	0.988825	
## Condition2RR An	-1.557e+04	4.353e+04	-0.358	0.720662	
## Condition2RR Nn	-3.482e+04	3.584e+04	-0.971	0.331527	
## BldgType2fmCon	9.065e+03	9.998e+03	0.907	0.364712	
## BldgTypeDuplex	-1.077e+04	6.306e+03	-1.707	0.087988	.
## BldgTypeTwnhs	-4.226e+04	9.061e+03	-4.663	3.42e-06	***
## BldgTypeTwnhsE	-2.972e+04	6.764e+03	-4.394	1.20e-05	***
## OverallQual	2.095e+04	1.170e+03	17.895	< 2e-16	***
## OverallCond	3.717e+03	9.790e+02	3.797	0.000153	***
## YearBuilt	3.080e+02	8.145e+01	3.782	0.000162	***
## RoofMatlCompShg	2.618e+05	3.799e+04	6.890	8.46e-12	***
## RoofMatlMembran	2.768e+05	5.201e+04	5.322	1.20e-07	***
## RoofMatlMetal	2.742e+05	5.170e+04	5.303	1.32e-07	***
## RoofMatlRoll	2.610e+05	5.170e+04	5.048	5.05e-07	***
## RoofMatlTar&Grv	2.633e+05	3.950e+04	6.667	3.78e-11	***
## RoofMatlWdShake	2.801e+05	4.127e+04	6.788	1.69e-11	***
## RoofMatlWdShngl	3.664e+05	4.031e+04	9.088	< 2e-16	***
## Exterior2ndAsphShn	-1.574e+04	2.145e+04	-0.734	0.463121	
## Exterior2ndBrk Cmn	-1.103e+04	1.896e+04	-0.582	0.560697	
## Exterior2ndBrkFace	-5.298e+03	1.081e+04	-0.490	0.624177	
## Exterior2ndCBlock	2.201e+04	3.542e+04	0.621	0.534494	
## Exterior2ndCmentBd	4.656e+03	1.015e+04	0.459	0.646662	
## Exterior2ndHdBoard	-1.244e+04	8.606e+03	-1.445	0.148564	
## Exterior2ndImStucc	2.462e+04	1.392e+04	1.769	0.077181	.
## Exterior2ndMetalSd	-6.888e+03	8.397e+03	-0.820	0.412223	
## Exterior2ndOther	-6.674e+03	3.595e+04	-0.186	0.852722	
## Exterior2ndPlywood	-8.842e+03	8.822e+03	-1.002	0.316356	
## Exterior2ndStone	-2.987e+04	1.775e+04	-1.682	0.092724	.
## Exterior2ndStucco	-6.874e+02	1.069e+04	-0.064	0.948736	
## Exterior2ndVinylSd	-8.203e+03	8.519e+03	-0.963	0.335770	
## Exterior2ndWd Sdng	-1.013e+04	8.385e+03	-1.208	0.227096	
## Exterior2ndWd Shng	-2.287e+04	9.781e+03	-2.338	0.019544	*
## BsmtFinSF2	1.260e+01	5.953e+00	2.117	0.034428	*
## FullBath	2.611e+04	2.343e+03	11.147	< 2e-16	***
## HalfBath	1.443e+04	2.231e+03	6.468	1.37e-10	***
## KitchenQualFa	-6.161e+04	7.903e+03	-7.796	1.26e-14	***
## KitchenQualGd	-5.313e+04	4.416e+03	-12.032	< 2e-16	***
## KitchenQualTA	-5.895e+04	5.010e+03	-11.766	< 2e-16	***
## PavedDriveP	1.344e+03	7.385e+03	0.182	0.855577	
## PavedDriveY	6.670e+03	4.471e+03	1.492	0.135992	
## SaleConditionAdjLand	-5.219e+03	1.825e+04	-0.286	0.774996	

```
## SaleConditionAlloca 2.065e+04 1.119e+04 1.846 0.065055 .
## SaleConditionFamily 1.713e+03 8.527e+03 0.201 0.840810
## SaleConditionNormal 7.069e+03 3.668e+03 1.927 0.054180 .
## SaleConditionPartial 2.565e+04 5.201e+03 4.931 9.17e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 33980 on 1373 degrees of freedom
## Multiple R-squared:  0.8279, Adjusted R-squared:  0.8171
## F-statistic: 76.78 on 86 and 1373 DF,  p-value: < 2.2e-16
```

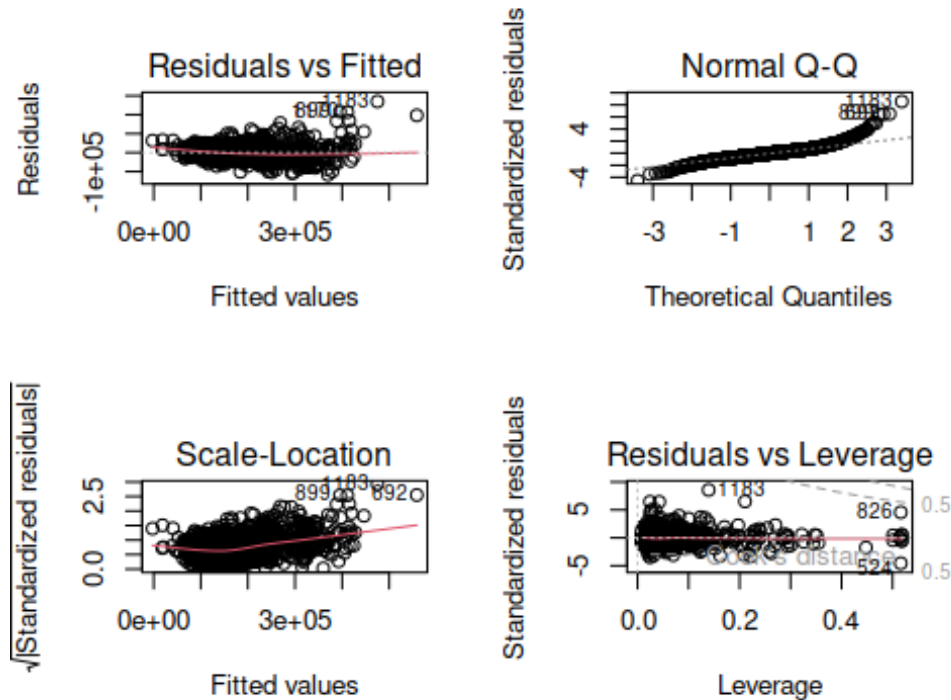
```
confint(fit_best)
```

```
##              2.5 %      97.5 %
## (Intercept) -1.152541e+06 -4.834642e+05
## MSSubClass  -1.790186e+02  2.474500e+01
## LotArea      7.505843e-01  1.188048e+00
## LotShapeIR2  -6.174083e+03  1.670158e+04
## LotShapeIR3  -1.797110e+04  2.975345e+04
## LotShapeReg  -8.023864e+03  5.097792e+02
## NeighborhoodBlueste -6.094166e+04  4.138369e+04
## NeighborhoodBrDale  -4.067498e+04  1.323477e+04
## NeighborhoodBrkSide -3.102912e+04  1.404667e+04
## NeighborhoodClearCr -2.836692e+04  1.955013e+04
## NeighborhoodCollgCr -3.917876e+04 -1.430301e+03
## NeighborhoodCrawfor  3.514644e+02  4.451621e+04
## NeighborhoodEdwards -3.786027e+04  3.117072e+03
## NeighborhoodGilbert -5.633434e+04 -1.568688e+04
## NeighborhoodIDOTRR  -4.437865e+04  2.940316e+03
## NeighborhoodMeadowV -3.680557e+04  1.913804e+04
## NeighborhoodMitchel -3.550258e+04  6.824268e+03
## NeighborhoodNames  -3.329990e+04  6.828032e+03
## NeighborhoodNoRidge  4.159185e+04  8.480881e+04
## NeighborhoodNPkVill -4.216694e+04  2.637874e+04
## NeighborhoodNridgHt  1.386424e+04  5.247666e+04
## NeighborhoodNWAmes  -4.459332e+04 -2.699579e+03
## NeighborhoodOldTown -4.281833e+04  6.657605e+02
## NeighborhoodSawyer  -3.467921e+04  7.748069e+03
## NeighborhoodSawyerW -3.197535e+04  8.917258e+03
## NeighborhoodSomerst -3.455946e+04  3.562379e+03
## NeighborhoodStoneBr  2.435376e+04  6.874008e+04
## NeighborhoodSWISU   -3.547926e+04  1.384139e+04
## NeighborhoodTimber  -3.259291e+04  1.064425e+04
## NeighborhoodVeenker -1.729295e+04  3.877724e+04
## Condition1Feedr     -8.999525e+03  1.782260e+04
## Condition1Norm      -3.140546e+03  1.882635e+04
## Condition1PosA      -1.358709e+04  4.055807e+04
## Condition1PosN      -7.571948e+02  3.922694e+04
## Condition1RR Ae     -3.734552e+04  1.026175e+04
## Condition1RR AN     -9.658141e+03  2.693851e+04
```

## Condition1RRNe	-6.573026e+04	3.360068e+04
## Condition1RRNn	-2.863265e+04	3.814774e+04
## Condition2Feedr	-8.750707e+04	2.995951e+04
## Condition2Norm	-6.775634e+04	3.366989e+04
## Condition2PosA	-5.512405e+04	1.141724e+05
## Condition2PosN	-2.363655e+05	-9.040059e+04
## Condition2RR Ae	-8.682381e+04	8.559256e+04
## Condition2RRAn	-1.009641e+05	6.982630e+04
## Condition2RRNn	-1.051255e+05	3.549389e+04
## BldgType2fmCon	-1.054742e+04	2.867798e+04
## BldgTypeDuplex	-2.313565e+04	1.603969e+03
## BldgTypeTwnhs	-6.003080e+04	-2.447928e+04
## BldgTypeTwnhsE	-4.298975e+04	-1.645333e+04
## OverallQual	1.864916e+04	2.324119e+04
## OverallCond	1.797029e+03	5.637837e+03
## YearBuilt	1.482393e+02	4.677901e+02
## RoofMatlCompShg	1.872571e+05	3.363235e+05
## RoofMatlMembran	1.747599e+05	3.788210e+05
## RoofMatlMetal	1.727667e+05	3.756043e+05
## RoofMatlRoll	1.595799e+05	3.624099e+05
## RoofMatlTar&Grv	1.858420e+05	3.408123e+05
## RoofMatlWdShake	1.991693e+05	3.610726e+05
## RoofMatlWdShngl	2.872796e+05	4.454462e+05
## Exterior2ndAsphShn	-5.780888e+04	2.633019e+04
## Exterior2ndBrk Cmn	-4.822002e+04	2.615595e+04
## Exterior2ndBrkFace	-2.650679e+04	1.591054e+04
## Exterior2ndCBlock	-4.748110e+04	9.149987e+04
## Exterior2ndCmentBd	-1.526428e+04	2.457616e+04
## Exterior2ndHdBoard	-2.932020e+04	4.442688e+03
## Exterior2ndImStucc	-2.688252e+03	5.193578e+04
## Exterior2ndMetalSd	-2.336041e+04	9.585029e+03
## Exterior2ndOther	-7.718924e+04	6.384029e+04
## Exterior2ndPlywood	-2.614756e+04	8.463049e+03
## Exterior2ndStone	-6.469382e+04	4.958804e+03
## Exterior2ndStucco	-2.165677e+04	2.028199e+04
## Exterior2ndVinylSd	-2.491599e+04	8.509247e+03
## Exterior2ndWd Sdng	-2.658185e+04	6.316357e+03
## Exterior2ndWd Shng	-4.205281e+04	-3.678054e+03
## BsmtFinSF2	9.253563e-01	2.428186e+01
## FullBath	2.151741e+04	3.070833e+04
## HalfBath	1.005268e+04	1.880411e+04
## KitchenQualFa	-7.711640e+04	-4.610822e+04
## KitchenQualGd	-6.179607e+04	-4.447019e+04
## KitchenQualTA	-6.877524e+04	-4.911902e+04
## PavedDriveP	-1.314326e+04	1.583213e+04
## PavedDriveY	-2.101256e+03	1.544173e+04
## SaleConditionAdjLand	-4.102651e+04	3.058907e+04
## SaleConditionAlloca	-1.290113e+03	4.259753e+04
## SaleConditionFamily	-1.501408e+04	1.844009e+04

```
## SaleConditionNormal -1.270897e+02 1.426561e+04
## SaleConditionPartial 1.544412e+04 3.584981e+04

best_Ajdusted_R2 <- 0.817
```



```
dev.off()

## null device
##          1

ncvTest(fit_best)

## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 1015.275, Df = 1, p = < 2.22e-16

vif(fit_best)

##           GVIF Df GVIF^(1/(2*Df))
## MSSubClass    6.099664  1      2.469750
## LotArea       1.565365  1      1.251145
## LotShape      1.865408  3      1.109504
## Neighborhood 627.936223 24      1.143642
## Condition1    3.177080  8      1.074922
## Condition2    2.343452  7      1.062719
## BldgType      18.903718  4      1.444004
## OverallQual    3.311406  1      1.819727
## OverallCond   1.499820  1      1.224672
```

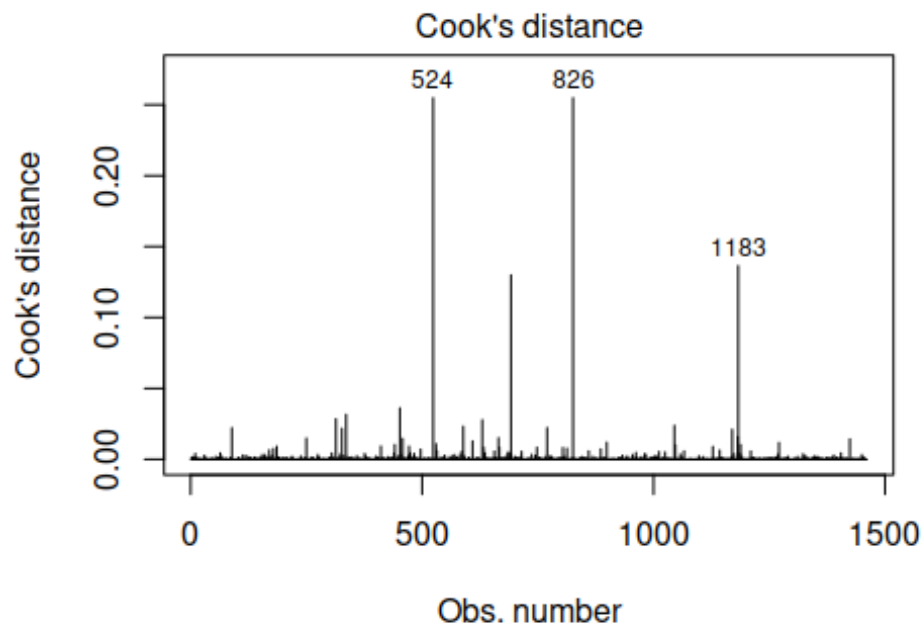


```
## YearBuilt      7.647821  1      2.765469
## RoofMatl      2.181393  7      1.057293
## Exterior2nd   23.126439 15      1.110377
## BsmtFinSF2    1.165597  1      1.079628
## FullBath      2.104983  1      1.450856
## HalfBath      1.590216  1      1.261038
## KitchenQual   3.833219  3      1.251010
## PavedDrive    1.589627  2      1.122855
## SaleCondition  2.155019  5      1.079804
```

```
outlierTest(fit_best)
```

```
##      rstudent unadjusted p-value Bonferroni p
## 1183  8.794335      4.2284e-18  6.1354e-15
## 692   6.606250      5.6238e-11  8.1601e-08
## 899   6.541040      8.6002e-11  1.2479e-07
## 1170  6.524249      9.5884e-11  1.3913e-07
## 804   5.047347      5.0808e-07  7.3723e-04
## 441   5.031627      5.5075e-07  7.9914e-04
## 1143  4.683814      3.0952e-06  4.4911e-03
## 524   -4.591180      4.8112e-06  6.9810e-03
## 826   4.591180      4.8112e-06  6.9810e-03
```

```
plot(fit_best, 4)
```



```
salePrice ~ MSSubClass + LotArea + LotShape + Neighborhood + Co
```

Q4 custom

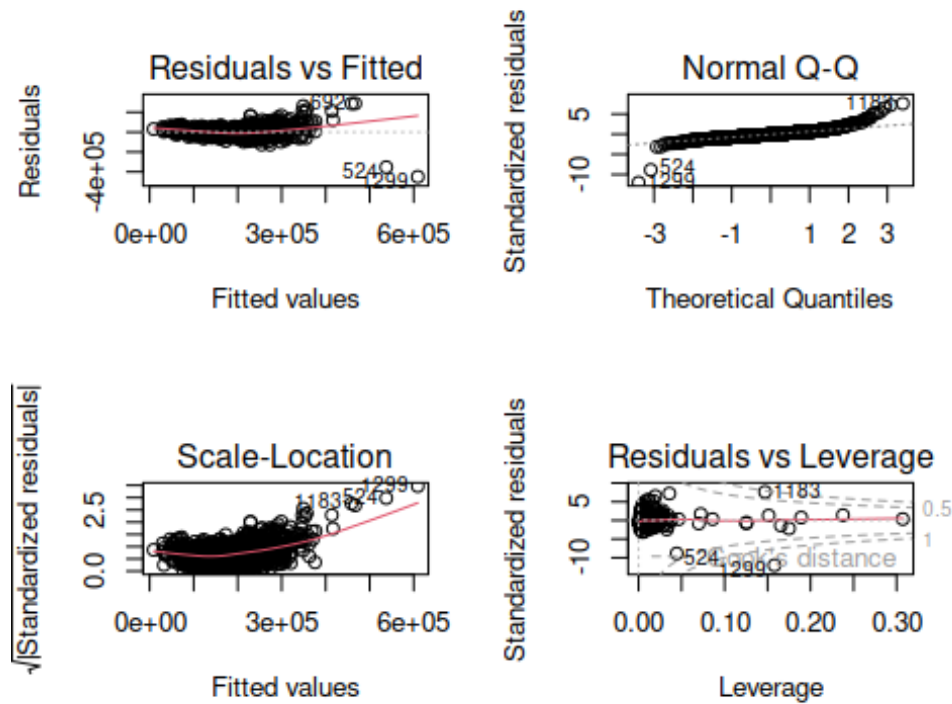
```
summary(custom_fit)

##
## Call:
## lm(formula = SalePrice ~ OverallCond + YearBuilt + YearRemodAdd +
##      BedroomAbvGr + GrLivArea + KitchenAbvGr + TotRmsAbvGrd +
##      GarageCars + PoolArea + LotArea, data = train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -449279  -21631   -3228    17607   287987
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.806e+06  1.219e+05 -14.820 < 2e-16 ***
## OverallCond   7.138e+03  1.148e+03   6.219 6.52e-10 ***
## YearBuilt     7.232e+02  5.603e+01  12.906 < 2e-16 ***
## YearRemodAdd  2.057e+02  7.324e+01   2.809 0.00504 **
## BedroomAbvGr -1.760e+04  1.855e+03  -9.489 < 2e-16 ***
## GrLivArea     8.201e+01  4.001e+00  20.499 < 2e-16 ***
## KitchenAbvGr -3.965e+04  5.269e+03  -7.525 9.24e-14 ***
## TotRmsAbvGrd  5.975e+03  1.419e+03   4.210 2.71e-05 ***
## GarageCars    1.997e+04  1.912e+03  10.444 < 2e-16 ***
## PoolArea     -2.178e+01  2.717e+01  -0.801 0.42301
## LotArea       6.784e-01  1.116e-01   6.078 1.55e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 40720 on 1449 degrees of freedom
## Multiple R-squared:  0.7391, Adjusted R-squared:  0.7373
## F-statistic: 410.5 on 10 and 1449 DF,  p-value: < 2.2e-16

confint(custom_fit)

##              2.5 %       97.5 %
## (Intercept) -2.045540e+06 -1.567328e+06
## OverallCond  4.886535e+03  9.389196e+03
## YearBuilt     6.132901e+02  8.331261e+02
## YearRemodAdd  6.202857e+01  3.493568e+02
## BedroomAbvGr -2.124017e+04 -1.396307e+04
## GrLivArea     7.416379e+01  8.985940e+01
## KitchenAbvGr -4.998605e+04 -2.931343e+04
## TotRmsAbvGrd  3.191226e+03  8.759741e+03
## GarageCars    1.622112e+04  2.372316e+04
## PoolArea     -7.507878e+01  3.152473e+01
## LotArea       4.594348e-01  8.972829e-01
```

```
custom_Ajdsuted_R2 <- 0.737
```



```
dev.off()
```

```
## null device
##          1
```

```
ncvTest(custom_fit)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 2473.03, Df = 1, p = < 2.22e-16
```

```
vif(custom_fit)
```

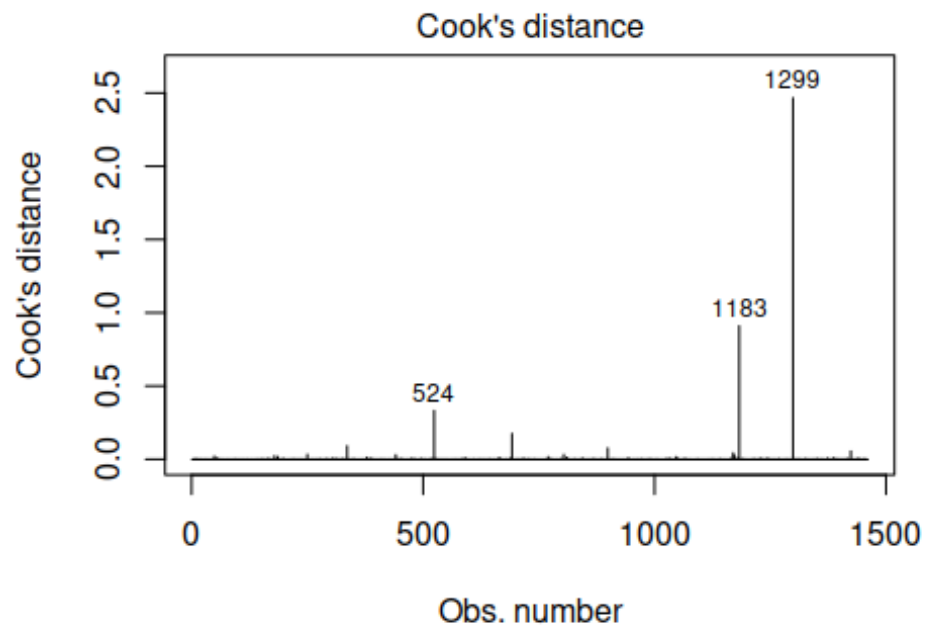
```
## OverallCond    YearBuilt YearRemodAdd BedroomAbvGr    GrLivArea
KitchenAbvGr
##      1.435351      2.520468      2.011821      2.014869      3.889160
1.186196
## TotRmsAbvGrd   GarageCars    PoolArea      LotArea
##      4.683619      1.797015      1.048795      1.091953
```

```
outlierTest(custom_fit)
```

```
##          rstudent unadjusted p-value Bonferroni p
## 1299 -12.668985      5.8379e-35      8.5233e-32
## 524  -9.113197      2.5899e-19      3.7813e-16
## 1183  7.771990      1.4554e-14      2.1249e-11
```

```
## 692    7.332827      3.7349e-13  5.4529e-10
## 899    6.607988      5.4593e-11  7.9706e-08
## 804    5.635761      2.0916e-08  3.0537e-05
## 1170    5.277021      1.5127e-07  2.2086e-04
## 441     5.192279      2.3730e-07  3.4645e-04
## 1047    5.044967      5.1099e-07  7.4604e-04
```

```
plot(custom_fit, 4)
```



```
lmePrice ~ OverallCond + YearBuilt + YearRemodAdd + BedroomAbv
```