Car Price Prediction using Multiple Linear Regression

Gerry Cruz, Yushan Zhao, Kierra Manuel

STAT 510, Fall 2022

The format of the report should follow the report structure, starting from a title page to conclusion paragraph(s). Each section must have a clear title. Make sure that there is some margin between two consecutive sections. Title, variable names, legends, etc in a plot must be clear and easy to read.

Contents

(I)Introduction and Background

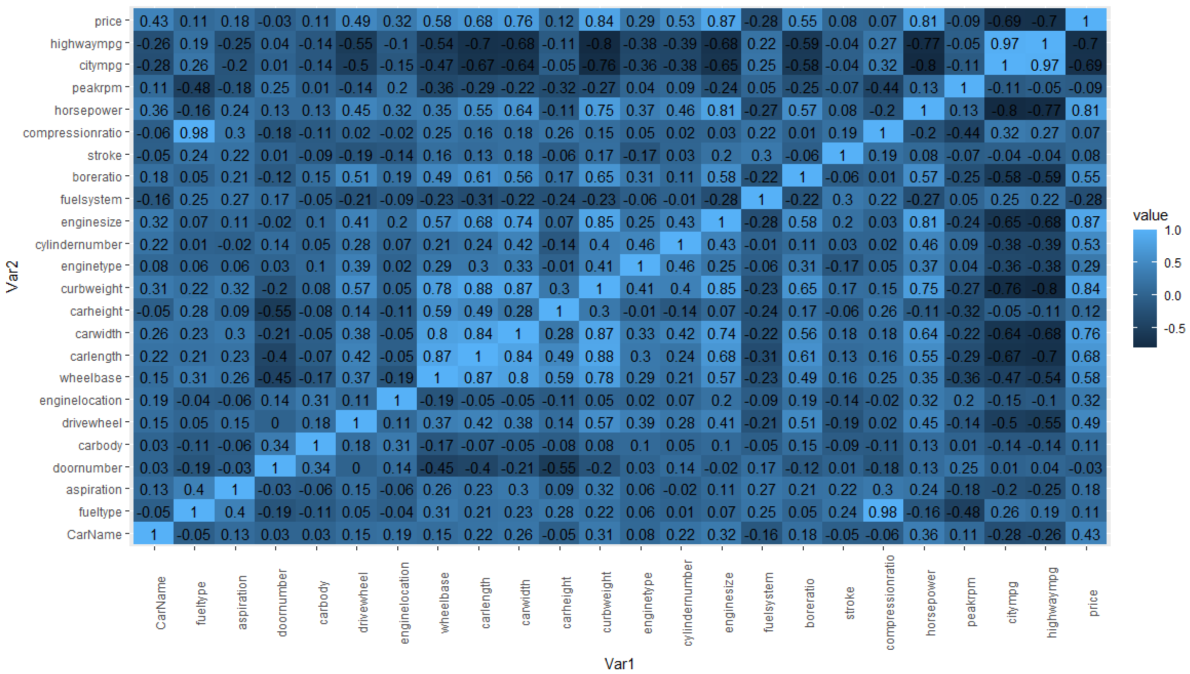
A Chinese automobile company Geely Auto aspires to enter the US market by setting up their manufacturing unit there and producing cars locally to give competition to their US and European counterparts. They have contracted an automobile consulting company to understand the factors on which the pricing of cars depends. Specifically, they want to understand the factors affecting the pricing of cars in the American market, since those may be very different from the Chinese market. The company wants to know: 1) Which variables are significant in predicting the price of a car? 2) How well do those variables describe the price of a car? Based on various market surveys, the consulting firm has gathered a large data set of different types of cars across the America market.

We are required to model the price of cars with the available independent variables. It will be used by the management to understand how exactly the prices vary with the independent variables. They can accordingly manipulate the design of the cars, the business strategy etc. to meet certain price levels. Further, the model will be a good way for management to understand the pricing dynamics of a new market.

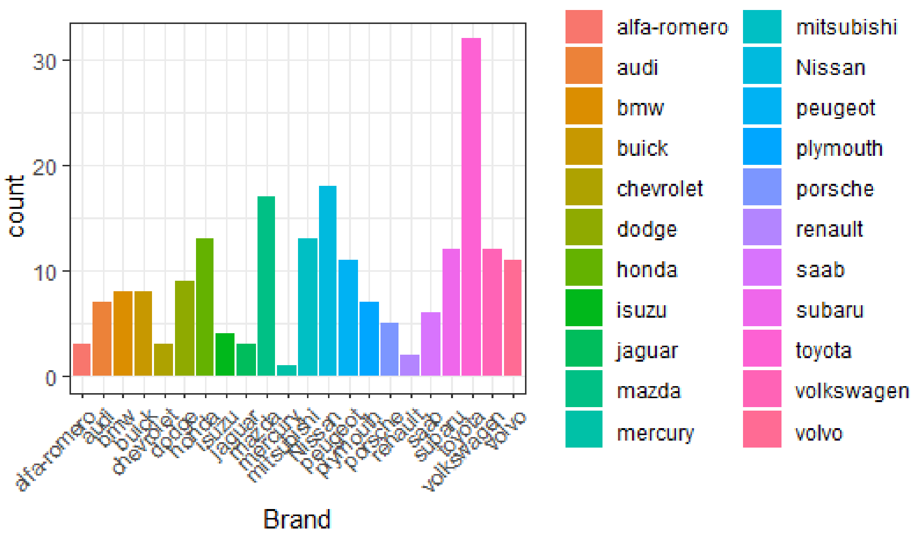
Summary of the Data Set

We have 205 cases, 23 independent variables and 1 dependent variable. The dependent variable is car price. The independent variables are: car name, fuel type, drive wheel, engine location, engine type, cylinder number, engine size, stroke, door number, carbody, fuel system, aspiration, wheel base, car length, car width, car height, curb weight, bore ratio, compression ratio, horsepower, peak rpm, city mpg, highway mpg.

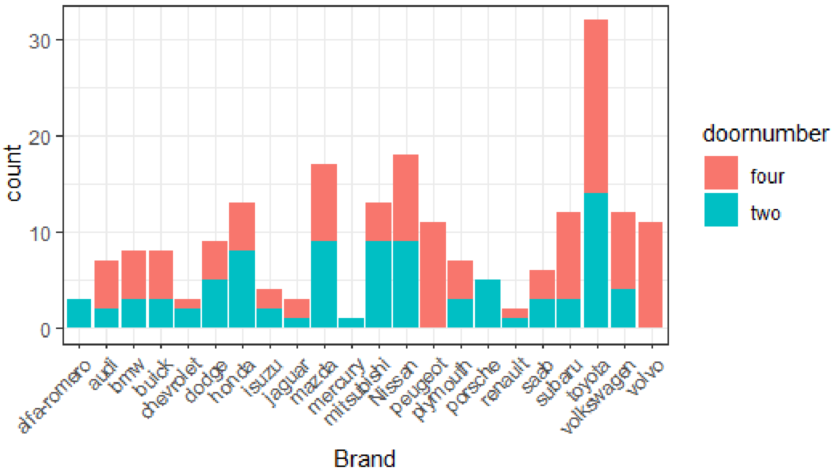
(II)Exploratory Data Analysis



From the correlation heat map, we found that compression ratio and fuel type, city mpg and highway mpg, curve weight and horsepower are highly correlated.

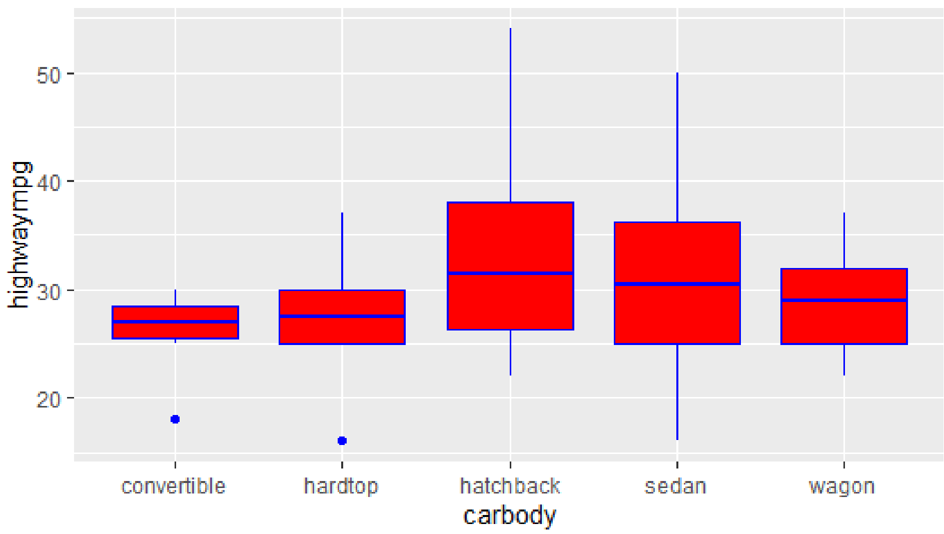


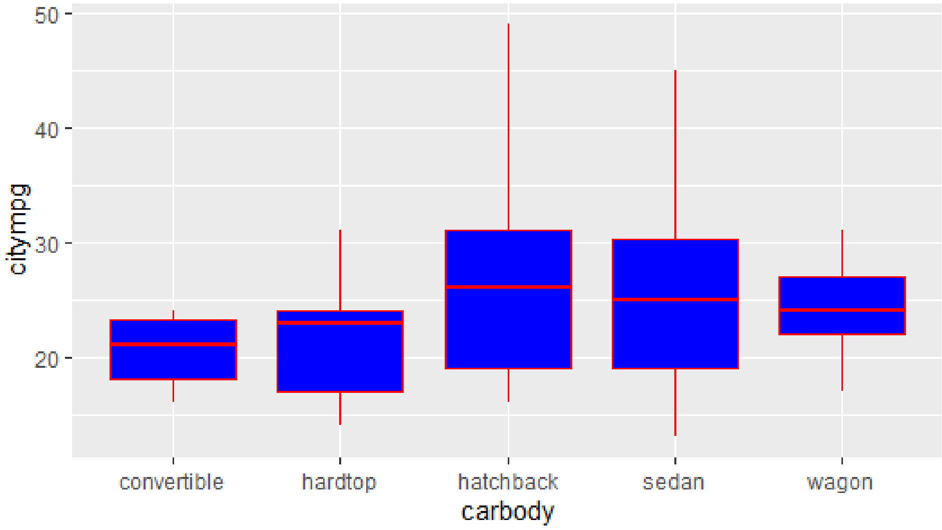
One thing we wanted to know from our dataset is which car brand was represented in our dataset the most. From plotting our histogram we found out that Toyota was represented the most then came Nissan and mazda.



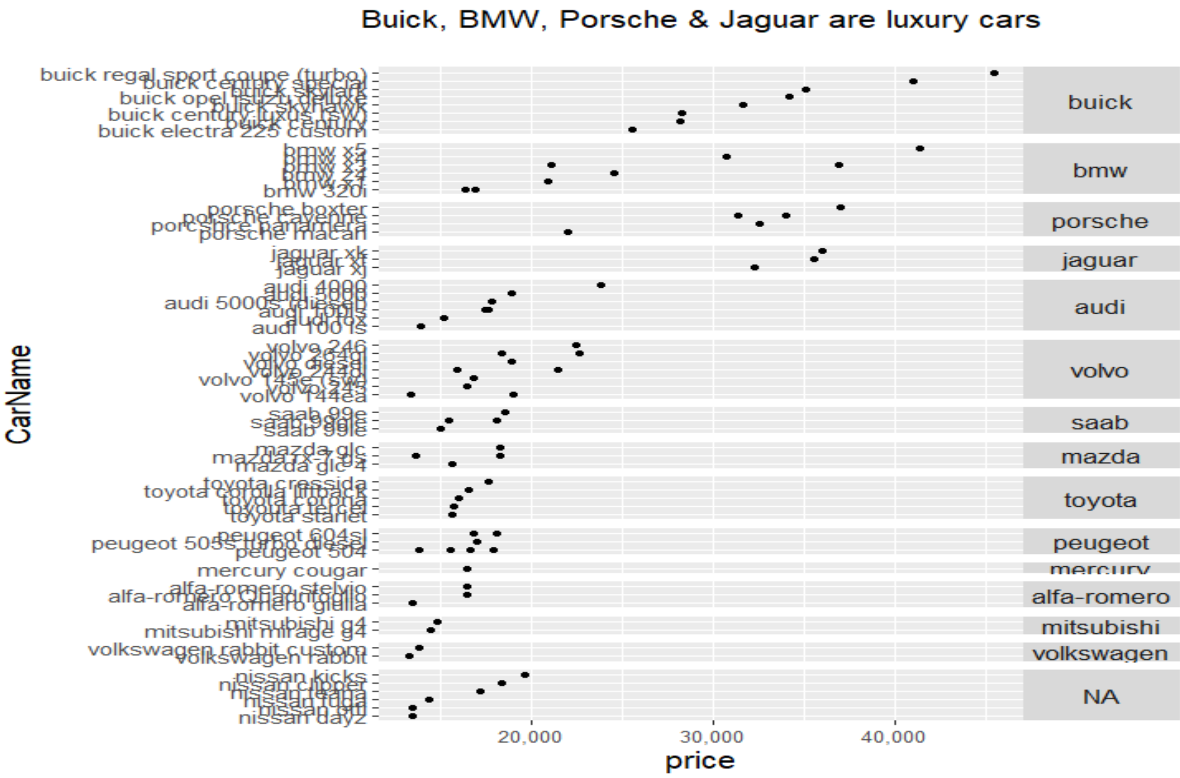
Another thing we wanted to know was what was the proportion of sports cars vs non sports cars in our dataset. We looked at an insurance company's classification of a sports car and it consisted of a car having 2 doors. So we classified a sports car as a car having two doors, and so we made a hist ogram to see the proportion of sports cars vs non sports cars in our datas et with respect to the brand names. We see from histogram that sports cars were evenly distributed among car brands.

Which body type has the worst/best highway gallon per mile?





After further EDA, we create box plots of each car body’s highway mpg and city mpg and the plots show us each car body’s median value. These plots show which body type has the best or worst highway mpg and city mpg. Hatchback has the best highway and city mpg Convertible has the worst highway and city mpg. After our model is complete, we could look into the highway and city mpg significance on car price.



Here we graphed the prices of each brand car. We can notice here that

Buick, BMW, Porsche, and Jaguar car brands seem to be the more expensive car brands, while Volkswagen seems to be the cheapests car brands.

Questions of Interests

1. Which variables are significantly related to a car's price?
2. Given a list of car specifications, would our model predict the requested car price and average price of cars of these specifications?
3. How much variation can our model explain the car price?

Regression Method

Step 1:

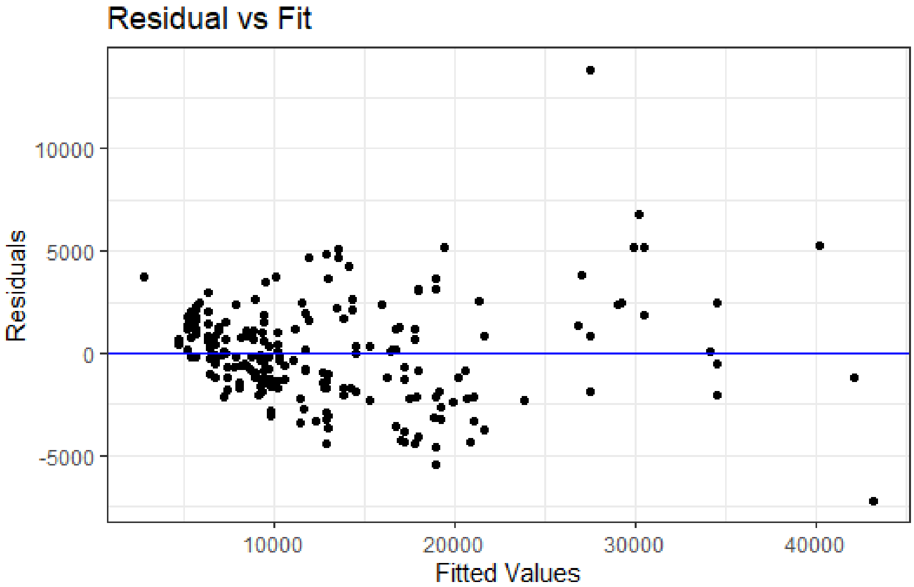
We use the Multiple Linear Regression Method to analyze which variables are significant predictors of a car's price. In order to build this model, we need to identify which of the 23 variables to include in the model. While we learned stepwise using AIC and the best subset methods for variable selection, we chose to use an alternate method, backwards stepwise regression, because our group member, Gerry, learned this method in previous courses and wanted the opportunity to apply it to our regression method. We then checked if there were any significant interaction terms to be included in our model. We were then able to determine a model:

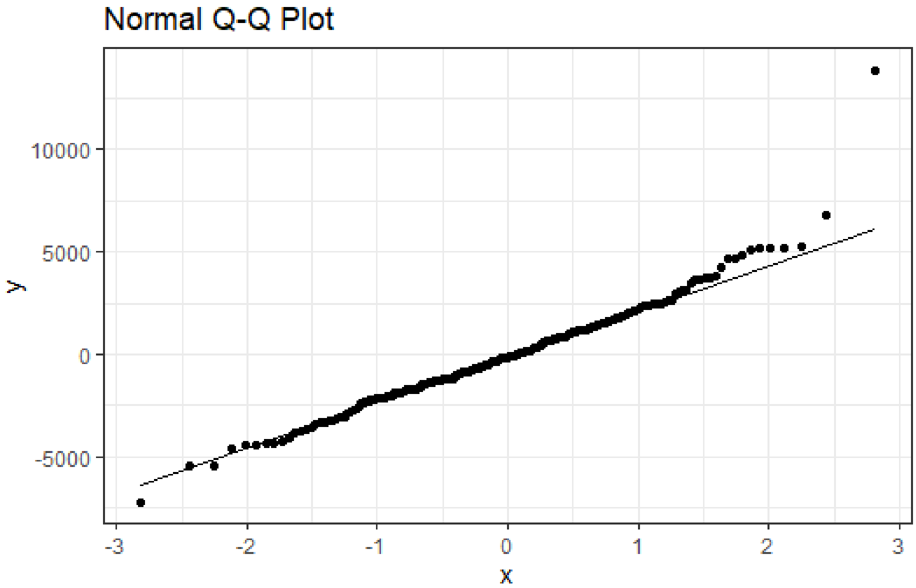
Car price^0.09 = car names x1+fuel type 2+drive wheelx3+engine location x4+ wheel base x5+ car width x6+ engine type x7+ cylinder numberx8+ engine size x9+ stroke x10+ compression ratio 11 horse power x12+ peak xrpmx13 +car name*cylinder number

(III)Regression Analysis

Step 2:

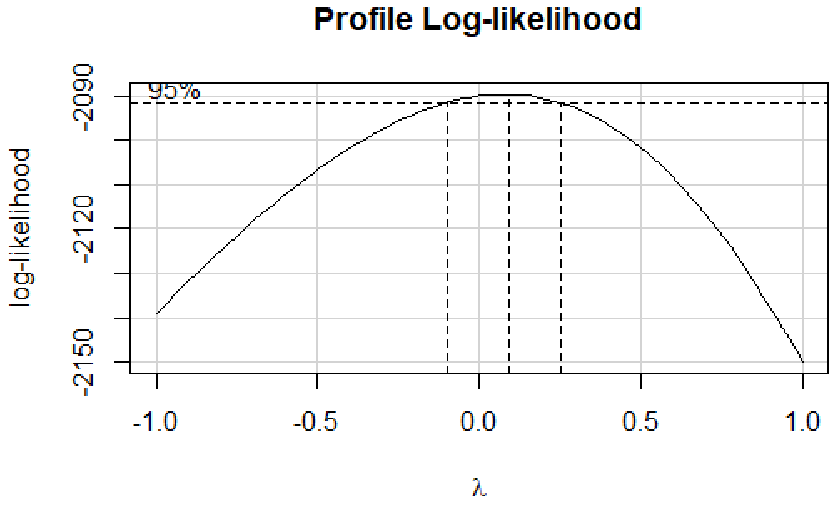
The next steps would be to perform residual analysis on our model. We check for linearity, independence, equal variance, and normality (LINE) of our residuals to make sure our model is good at analyzing our data. We find most LINE conditions are met except equal variance.





Step 3:

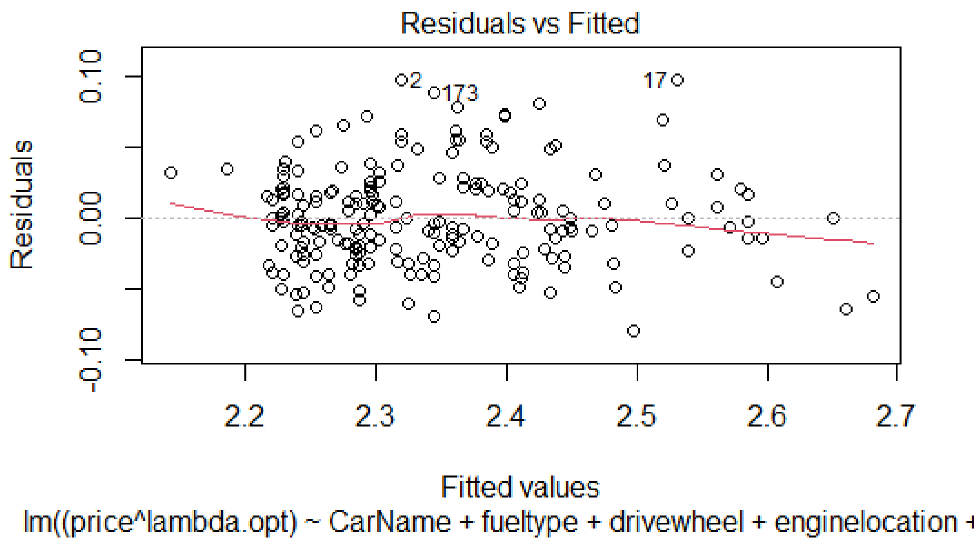
To fix the unequal variance, we chose to apply the Box-Cox Method in order to find the optimal lambda to perform a power transformation on our response variable, car price.

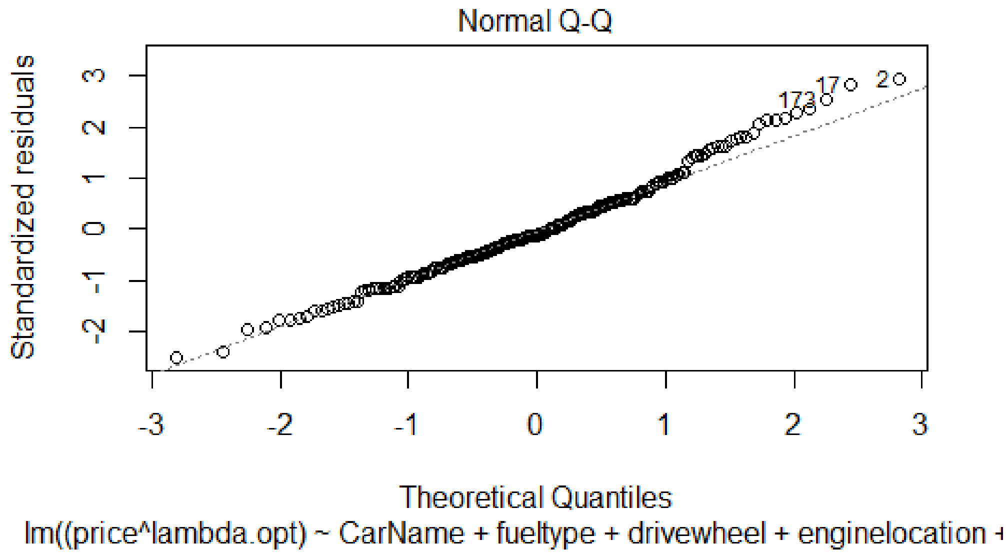


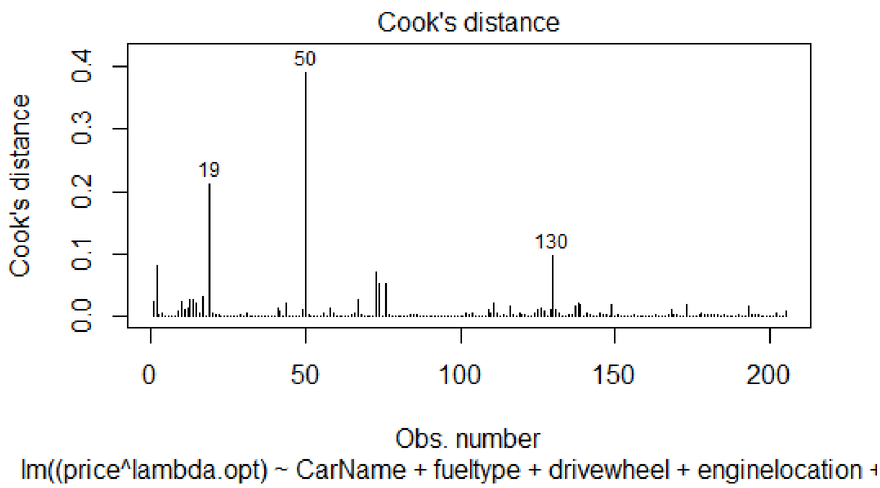
Lambda.opt = 0.09. After the power transformation, our final model is: Car priceˆ10.09 = car name + fuel type + drive wheel + engine location + wheelbase + car width + engine type + cylinder number + engine size + stroke + compression ratio + horsepower + peak rpm + car name:cylinder number

Step 4:

We check if our power transformation affected the LINE conditions and if this final model is good for analyzations and interpretations. From the residual vs. fitted plot below, we can see that the equal variance condition is satisfied now and other LINE conditions still satisfied.







In our check for high leverage points using Cook's distance, we find there are no high leverage points since Cook's distance for our plot is below 0.05

(IV)Results and Interpretation

```
Call:
lm(formula = (price^lambda.opt) ~ CarName + fueltype + drivewheel +
    enginelocation + wheelbase + carwidth + enginetype + cylindernumber +
    enginesize + stroke + compressionratio + horsepower + peakrpm +
    CarName:cylindernumber, data = data)

Residuals:
      Min        1Q    Median        3Q       Max
-0.080081 -0.023134 -0.004182  0.019577  0.098205

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)             1.248e+00  1.494e-01   8.354 1.36e-14 ***
CarName                 2.405e-03  5.379e-04   4.471 1.34e-05 ***
fueltype                9.392e-02  5.975e-02   1.572 0.117648
drivewheel              2.201e-02  5.507e-03   3.997 9.17e-05 ***
enginelocation          1.111e-01  2.526e-02   4.401 1.80e-05 ***
wheelbase               3.479e-03  8.218e-04   4.233 3.59e-05 ***
carwidth                8.973e-03  2.734e-03   3.282 0.001229 **
enginetype             -8.216e-03  2.376e-03  -3.458 0.000672 ***
cylindernumber          2.674e-02  5.604e-03   4.773 3.62e-06 ***
enginesize              5.944e-04  1.566e-04   3.796 0.000198 ***
stroke                 -1.887e-02  1.014e-02  -1.860 0.064373 .
compressionratio       -3.884e-03  4.309e-03  -0.901 0.368572
horsepower              9.514e-04  1.659e-04   5.735 3.79e-08 ***
peakrpm                 1.007e-05  7.880e-06   1.279 0.202611
CarName:cylindernumber -1.660e-03  4.595e-04  -3.612 0.000389 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.03571 on 190 degrees of freedom
Multiple R-squared:    0.9,     Adjusted R-squared:  0.8926
F-statistic: 122.2 on 14 and 190 DF,  p-value: < 2.2e-16
```

Our results indicate car name was a strong predictor of car price^0.09, controlling for fuel type, drive wheel, engine location, wheel base, car width, engine type, cylinder number, engine size, stroke, compression ratio, horse power, peak rpm, and the interaction between car name and cylinder number, $b = 2.405e-03$, $SE = 5.739e-04$, $t = 4.471$, $p < .001$. Also, drive wheel, engine location, wheel base, car width, engine type, cylinder number, engine size, horsepower, and the interaction between car name and cylinder number was a significant predictor of car price^0.09, controlling for remaining variables. However, fuel type, engine size, stroke, compression ratio, and peak rpm were not strong predictors of car price^0.09. The results highlight that car name, drive wheel, engine location, wheel base, car width, engine type, cylinder number, engine size, horsepower, and the interaction between car name and cylinder number are important for car price.

A multiple linear regression was conducted to investigate whether car name, fuel type, drive wheel, engine location, wheel base, car width, engine type, cylinder number, engine size, stroke, compression ratio, horsepower

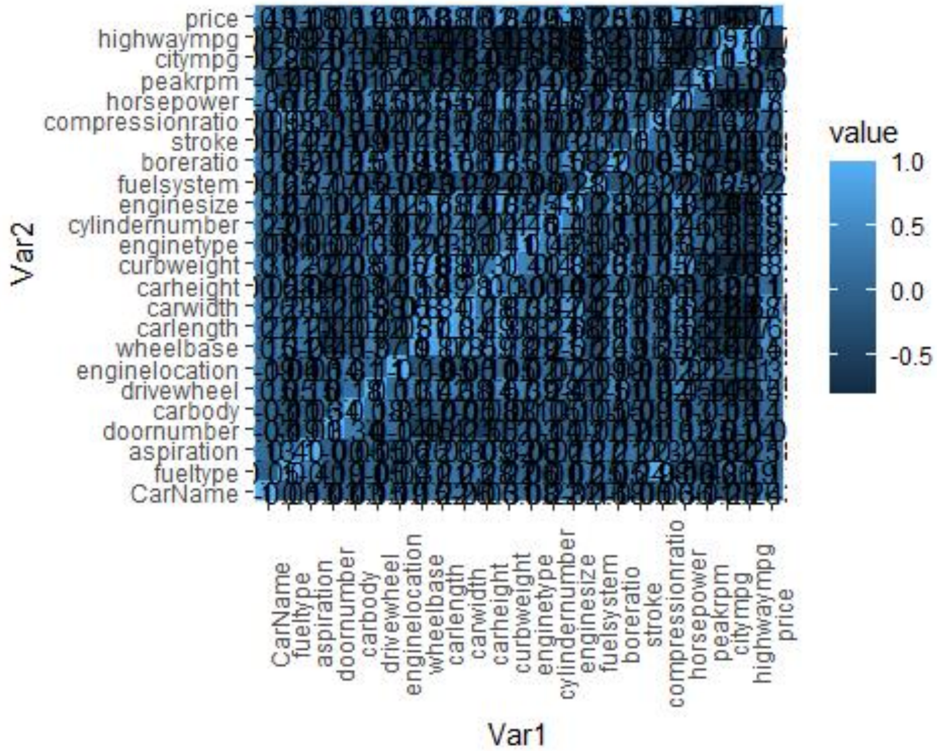
, peak rpm, and the interaction between car name and cylinder number would predict car price. We were able to use our model to make a prediction for a car fanatic with the requested specifics: a larger size, electric Porsche with all wheel drive and higher wheelbase, for increased tire traction, and engine specifics for a larger, powerful, and efficient engine. We found a car with these specifics was (122.403, 7187.106) and the average range for this type of car is \$129.05 to \$6926.84. Our results also indicate that 89% of variation in car price can be explained by our final model which confirms our final model is a good model.

(V)Conclusion

Using a multiple linear regression model, we found that car name, drive wheel, engine location, wheel base, car width, engine type, cylinder number, engine size, horsepower, and the interaction between car name and cylinder number are important for predicting car price in the American market. Thus, as a manager in Geely Auto, he or she should pay more attention to helping his or her company to design and create cars having high quality of those variables such as drive wheels and engine type. Although car name is a strong predictor of car price, and Geely Auto can not change its name to satisfy this requirement, it will be beneficial if Geely Auto can learn from other car brands about how to increase their quality and influential power by investing funding on hiring excellent engineers or advertisement.

(VI)Appendix

```
#load reshape2 package

library(reshape2)

# creating correlation matrix
corr_mat <- round(cor(data),2)

# reduce the size of correlation matrix
melted_corr_mat <- melt(corr_mat)
head(melted_corr_mat)

##          Var1    Var2 value
## 1     CarName CarName  1.00
## 2    fueltype CarName -0.05
## 3  aspiration CarName  0.13
## 4  doornumber CarName  0.03
## 5     carbody CarName  0.03
## 6  drivewheel CarName  0.15

# plotting the correlation heatmap
library(ggplot2)
p.dia=ggplot(data = melted_corr_mat, aes(x=Var1, y=Var2,
                                         fill=value)) +
  geom_tile() +
  geom_text(aes(Var2, Var1, label = value),
            color = "black", size = 4)

p <- p.dia +theme(axis.text.x= element_text(angle=90, size=9))
p
```




```
model <- lm(price ~., data = data)
summary(model)

##
## Call:
## lm(formula = price ~ ., data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8488.4 -1526.8  -116.5  1171.8 13542.6
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -4.796e+04  1.457e+04  -3.291 0.001199 **
## CarName             1.201e+02  3.919e+01   3.064 0.002521 **
## fueltype            1.195e+04  6.035e+03   1.980 0.049168 *
## aspiration         -4.874e+02  9.407e+02  -0.518 0.605033
## doornumber          3.278e+02  5.211e+02   0.629 0.530135
## carbody            -5.358e+00  2.396e+02  -0.022 0.982188
## drivewheel          9.921e+02  5.024e+02   1.975 0.049829 *
## enginelocation      1.144e+04  2.071e+03   5.523 1.14e-07 ***
## wheelbase           2.059e+02  9.060e+01   2.272 0.024244 *
## carlength          -2.582e+01  5.136e+01  -0.503 0.615752
## carwidth            4.611e+02  2.234e+02   2.063 0.040493 *
## carheight           5.239e+00  1.257e+02   0.042 0.966811
## curbweight          2.368e+00  1.854e+00   1.277 0.203228
```

```
## enginetype       -6.767e+02  1.929e+02  -3.508 0.000569 ***
## cylindernumber    1.393e+03  2.711e+02   5.140 7.08e-07 ***
## enginesize        8.158e+01  1.661e+01   4.913 2.00e-06 ***
## fuelsystem       -1.278e+02  1.934e+02  -0.661 0.509483
## boreratio        -7.541e+02  1.104e+03  -0.683 0.495433
## stroke           -2.329e+03  7.884e+02  -2.953 0.003559 **
## compressionratio -7.058e+02  4.241e+02  -1.664 0.097827 .
## horsepower        2.182e+01  1.753e+01   1.245 0.214752
## peakrpm           1.787e+00  6.402e-01   2.791 0.005817 **
## citympg          -1.385e+02  1.620e+02  -0.855 0.393865
## highwaympg        1.645e+02  1.502e+02   1.096 0.274693
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2660 on 181 degrees of freedom
## Multiple R-squared:  0.9017, Adjusted R-squared:  0.8892
## F-statistic: 72.15 on 23 and 181 DF,  p-value: < 2.2e-16
```

```r
#Highly correlated predictors:(compression ratio,fueltype)
#(citympg,highwaympg),(curveweight,horsepower),
#load the car library
library(car)
```

```
## Loading required package: carData
```

```r
#Use stepwiseRegression backwards
step(model,direction = "backward")
```

```
## Start:  AIC=3255.73
## price ~ CarName + fueltype + aspiration + doornumber + carbody +
##     drivewheel + enginelocation + wheelbase + carlength + carwidth +
##     carheight + curbweight + enginetype + cylindernumber + enginesize +
##     fuelsystem + boreratio + stroke + compressionratio + horsepower +
##     peakrpm + citympg + highwaympg
##
##                    Df Sum of Sq        RSS    AIC
## - carbody           1      3536 1280473730 3253.7
## - carheight         1     12281 1280482476 3253.7
## - carlength         1   1788108 1282258303 3254.0
## - aspiration        1   1898851 1282369045 3254.0
## - doornumber        1   2799028 1283269222 3254.2
## - fuelsystem        1   3090515 1283560710 3254.2
## - boreratio         1   3300869 1283771064 3254.3
## - citympg           1   5167553 1285637748 3254.6
## - highwaympg        1   8492290 1288962485 3255.1
## - horsepower        1  10965016 1291435210 3255.5
## - curbweight        1  11536962 1292007157 3255.6
## <none>                          1280470195 3255.7
```

```
## - compressionratio  1  19590167 1300060362 3256.8
## - drivewheel        1  27584885 1308055080 3258.1
## - fueltype          1  27747247 1308217442 3258.1
## - carwidth          1  30122803 1310592998 3258.5
## - wheelbase         1  36527938 1316998133 3259.5
## - peakrpm           1  55110231 1335580426 3262.4
## - stroke            1  61710022 1342180217 3263.4
## - CarName           1  66397154 1346867349 3264.1
## - enginetype        1  87080379 1367550573 3267.2
## - enginesize        1 170756113 1451226308 3279.4
## - cylindernumber    1 186912437 1467382632 3281.7
## - enginelocation    1 215823739 1496293933 3285.7
##
## Step:  AIC=3253.73
## price ~ CarName + fueltype + aspiration + doornumber + drivewheel +
##     enginelocation + wheelbase + carlength + carwidth + carheight +
##     curbweight + enginetype + cylindernumber + enginesize + fuelsystem +
##     boreratio + stroke + compressionratio + horsepower + peakrpm +
##     citympg + highwaympg
##
##                    Df Sum of Sq        RSS    AIC
## - carheight         1     10654 1280484384 3251.7
## - carlength         1   1793964 1282267694 3252.0
## - aspiration        1   1896121 1282369852 3252.0
## - doornumber        1   3084586 1283558316 3252.2
## - fuelsystem        1   3098133 1283571863 3252.2
## - boreratio         1   3396448 1283870179 3252.3
## - citympg           1   5234812 1285708543 3252.6
## - highwaympg        1   8651465 1289125195 3253.1
## - horsepower        1  11420162 1291893893 3253.6
## - curbweight        1  12544265 1293017995 3253.7
## <none>                          1280473730 3253.7
## - compressionratio  1  19628818 1300102549 3254.9
## - drivewheel        1  27829481 1308303212 3256.1
## - fueltype          1  27877331 1308351062 3256.1
## - carwidth          1  30121798 1310595529 3256.5
## - wheelbase         1  38843458 1319317188 3257.9
## - peakrpm           1  55380999 1335854729 3260.4
## - stroke            1  61889293 1342363023 3261.4
## - CarName           1  66561860 1347035591 3262.1
## - enginetype        1  87238958 1367712688 3265.2
## - enginesize        1 171809121 1452282851 3277.5
## - cylindernumber    1 186920878 1467394608 3279.7
## - enginelocation    1 235509823 1515983554 3286.3
##
## Step:  AIC=3251.74
## price ~ CarName + fueltype + aspiration + doornumber + drivewheel +
```

```
##     enginelocation + wheelbase + carlength + carwidth + curbweight +
##     enginetype + cylindernumber + enginesize + fuelsystem + boreratio +
##     stroke + compressionratio + horsepower + peakrpm + citympg +
##     highwaympg
##
##                    Df Sum of Sq        RSS    AIC
## - carlength         1   1785502 1282269887 3250.0
## - aspiration        1   1935069 1282419453 3250.0
## - doornumber        1   3106463 1283590847 3250.2
## - fuelsystem        1   3285551 1283769935 3250.3
## - boreratio         1   3405665 1283890049 3250.3
## - citympg           1   5254355 1285738739 3250.6
## - highwaympg        1   8654055 1289138439 3251.1
## - horsepower        1  11417843 1291902227 3251.6
## <none>                          1280484384 3251.7
## - curbweight        1  13118842 1293603226 3251.8
## - compressionratio  1  19876458 1300360842 3252.9
## - drivewheel        1  27898604 1308382988 3254.2
## - fueltype          1  28245660 1308730044 3254.2
## - carwidth          1  31047105 1311531489 3254.6
## - wheelbase         1  44128106 1324612490 3256.7
## - peakrpm           1  55377297 1335861681 3258.4
## - stroke            1  63168178 1343652562 3259.6
## - CarName           1  67001328 1347485712 3260.2
## - enginetype        1  92010491 1372494875 3264.0
## - enginesize        1 181265637 1461750021 3276.9
## - cylindernumber    1 187150863 1467635247 3277.7
## - enginelocation    1 244430951 1524915335 3285.5
##
## Step:  AIC=3250.02
## price ~ CarName + fueltype + aspiration + doornumber + drivewheel +
##     enginelocation + wheelbase + carwidth + curbweight + enginetype +
##     cylindernumber + enginesize + fuelsystem + boreratio + stroke +
##     compressionratio + horsepower + peakrpm + citympg + highwaympg
##
##                    Df Sum of Sq        RSS    AIC
## - aspiration        1   1588156 1283858043 3248.3
## - fuelsystem        1   2686633 1284956519 3248.5
## - boreratio         1   4190642 1286460528 3248.7
## - doornumber        1   4207794 1286477681 3248.7
## - citympg           1   4385198 1286655085 3248.7
## - highwaympg        1   8465617 1290735504 3249.4
## - curbweight        1  11333389 1293603275 3249.8
## - horsepower        1  11525211 1293795097 3249.9
## <none>                          1282269887 3250.0
## - compressionratio  1  21867706 1304137593 3251.5
## - carwidth          1  29835412 1312105299 3252.7
```

```
## - fueltype          1  30056598 1312326484 3252.8
## - drivewheel        1  33031890 1315301776 3253.2
## - wheelbase         1  46383267 1328653154 3255.3
## - peakrpm           1  56804682 1339074569 3256.9
## - stroke            1  63045800 1345315687 3257.9
## - CarName           1  67132283 1349402170 3258.5
## - enginetype        1  90252314 1372522200 3262.0
## - enginesize        1 190853775 1473123661 3276.5
## - cylindernumber    1 192798443 1475068329 3276.7
## - enginelocation    1 244540892 1526810779 3283.8
##
## Step:  AIC=3248.28
## price ~ CarName + fueltype + doornumber + drivewheel + enginelocation +
##     wheelbase + carwidth + curbweight + enginetype + cylindernumber +
##     enginesize + fuelsystem + boreratio + stroke + compressionratio +
##     horsepower + peakrpm + citympg + highwaympg
##
##                    Df Sum of Sq        RSS    AIC
## - fuelsystem        1   3581877 1287439920 3246.8
## - doornumber        1   3978641 1287836684 3246.9
## - boreratio         1   4160406 1288018449 3246.9
## - citympg           1   7493489 1291351531 3247.5
## - horsepower        1  10766302 1294624344 3248.0
## - curbweight        1  10803608 1294661651 3248.0
## <none>                          1283858043 3248.3
## - highwaympg        1  12653321 1296511363 3248.3
## - compressionratio  1  21932706 1305790749 3249.7
## - carwidth          1  29553526 1313411569 3250.9
## - fueltype          1  33075600 1316933643 3251.5
## - drivewheel        1  35206780 1319064822 3251.8
## - wheelbase         1  45993443 1329851485 3253.5
## - stroke            1  62445215 1346303257 3256.0
## - peakrpm           1  63875892 1347733935 3256.2
## - CarName           1  65685963 1349544005 3256.5
## - enginetype        1  89515594 1373373637 3260.1
## - cylindernumber    1 197253681 1481111724 3275.6
## - enginelocation    1 249732246 1533590289 3282.7
## - enginesize        1 297646668 1581504710 3289.0
##
## Step:  AIC=3246.85
## price ~ CarName + fueltype + doornumber + drivewheel + enginelocation +
##     wheelbase + carwidth + curbweight + enginetype + cylindernumber +
##     enginesize + boreratio + stroke + compressionratio + horsepower +
##     peakrpm + citympg + highwaympg
##
##                    Df Sum of Sq        RSS    AIC
## - doornumber        1   3241195 1290681115 3245.4
```

```
## - boreratio         1   5095184 1292535104 3245.7
## - citympg           1   8823231 1296263152 3246.2
## - curbweight        1   9886829 1297326749 3246.4
## - horsepower        1  12412455 1299852375 3246.8
## <none>                          1287439920 3246.8
## - highwaympg        1  14440544 1301880464 3247.1
## - compressionratio  1  18881537 1306321457 3247.8
## - fueltype          1  29537562 1316977482 3249.5
## - carwidth          1  30979014 1318418934 3249.7
## - drivewheel        1  38423494 1325863414 3250.9
## - wheelbase         1  52410200 1339850120 3253.0
## - peakrpm           1  60778645 1348218565 3254.3
## - CarName           1  68083130 1355523050 3255.4
## - stroke            1  72477744 1359917664 3256.1
## - enginetype        1  89928225 1377368145 3258.7
## - cylindernumber    1 194190618 1481630538 3273.6
## - enginelocation    1 250923005 1538362925 3281.3
## - enginesize        1 306172631 1593612552 3288.6
##
## Step:  AIC=3245.36
## price ~ CarName + fueltype + drivewheel + enginelocation + wheelbase +
##     carwidth + curbweight + enginetype + cylindernumber + enginesize +
##     boreratio + stroke + compressionratio + horsepower + peakrpm +
##     citympg + highwaympg
##
##                    Df Sum of Sq        RSS    AIC
## - boreratio         1   5355090 1296036204 3244.2
## - curbweight        1   8062940 1298744055 3244.6
## - citympg           1   8100099 1298781214 3244.6
## <none>                          1290681115 3245.4
## - highwaympg        1  13179278 1303860393 3245.4
## - horsepower        1  13797813 1304478927 3245.5
## - compressionratio  1  19774283 1310455398 3246.5
## - fueltype          1  31013983 1321695098 3248.2
## - carwidth          1  33587245 1324268359 3248.6
## - drivewheel        1  42922560 1333603675 3250.1
## - wheelbase         1  49578007 1340259122 3251.1
## - peakrpm           1  61759804 1352440919 3252.9
## - CarName           1  69618472 1360299587 3254.1
## - stroke            1  70187794 1360868908 3254.2
## - enginetype        1  86995585 1377676700 3256.7
## - cylindernumber    1 195289786 1485970901 3272.2
## - enginelocation    1 247873863 1538554978 3279.4
## - enginesize        1 324940068 1615621183 3289.4
##
## Step:  AIC=3244.21
## price ~ CarName + fueltype + drivewheel + enginelocation + wheelbase +
```

```
##     carwidth + curbweight + enginetype + cylindernumber + enginesize +
##     stroke + compressionratio + horsepower + peakrpm + citympg +
##     highwaympg
##
##                    Df Sum of Sq        RSS    AIC
## - citympg           1   6595370 1302631574 3243.3
## - curbweight        1   9208443 1305244647 3243.7
## - horsepower        1  11839004 1307875209 3244.1
## - highwaympg        1  12307515 1308343719 3244.1
## <none>                          1296036204 3244.2
## - compressionratio  1  19752771 1315788975 3245.3
## - carwidth          1  30637881 1326674085 3247.0
## - fueltype          1  30907659 1326943863 3247.0
## - drivewheel        1  38372036 1334408241 3248.2
## - wheelbase         1  47303966 1343340170 3249.6
## - stroke            1  67486772 1363522976 3252.6
## - CarName           1  69777346 1365813551 3253.0
## - peakrpm           1  74525984 1370562189 3253.7
## - enginetype        1  95977314 1392013518 3256.9
## - cylindernumber    1 238352457 1534388661 3276.8
## - enginelocation    1 242551421 1538587625 3277.4
## - enginesize        1 320581460 1616617664 3287.5
##
## Step:  AIC=3243.25
## price ~ CarName + fueltype + drivewheel + enginelocation + wheelbase +
##     carwidth + curbweight + enginetype + cylindernumber + enginesize +
##     stroke + compressionratio + horsepower + peakrpm + highwaympg
##
##                    Df Sum of Sq        RSS    AIC
## - highwaympg        1   6769710 1309401284 3242.3
## - curbweight        1  11576889 1314208463 3243.1
## <none>                          1302631574 3243.3
## - compressionratio  1  18943460 1321575034 3244.2
## - horsepower        1  19036783 1321668357 3244.2
## - fueltype          1  28623406 1331254980 3245.7
## - carwidth          1  32791489 1335423063 3246.3
## - drivewheel        1  34716184 1337347758 3246.6
## - wheelbase         1  43752560 1346384134 3248.0
## - stroke            1  64190786 1366822360 3251.1
## - peakrpm           1  71137869 1373769443 3252.2
## - CarName           1  72705256 1375336830 3252.4
## - enginetype        1  97237786 1399869360 3256.0
## - cylindernumber    1 235717985 1538349559 3275.3
## - enginelocation    1 260344860 1562976434 3278.6
## - enginesize        1 319925313 1622556887 3286.3
##
## Step:  AIC=3242.31
```

```
## price ~ CarName + fueltype + drivewheel + enginelocation + wheelbase +
##     carwidth + curbweight + enginetype + cylindernumber + enginesize +
##     stroke + compressionratio + horsepower + peakrpm
##
##                    Df Sum of Sq        RSS    AIC
## - curbweight        1   6305113 1315706397 3241.3
## <none>                          1309401284 3242.3
## - compressionratio  1  13467224 1322868508 3242.4
## - horsepower        1  16561341 1325962625 3242.9
## - fueltype          1  23490972 1332892257 3244.0
## - drivewheel        1  31529586 1340930870 3245.2
## - carwidth          1  32050301 1341451585 3245.3
## - wheelbase         1  43044770 1352446054 3246.9
## - stroke            1  60657845 1370059129 3249.6
## - peakrpm           1  65106516 1374507800 3250.3
## - CarName           1  81425601 1390826885 3252.7
## - enginetype        1  94136483 1403537767 3254.5
## - cylindernumber    1 228993988 1538395272 3273.4
## - enginelocation    1 256270109 1565671393 3277.0
## - enginesize        1 337948289 1647349573 3287.4
##
## Step:  AIC=3241.3
## price ~ CarName + fueltype + drivewheel + enginelocation + wheelbase +
##     carwidth + enginetype + cylindernumber + enginesize + stroke +
##     compressionratio + horsepower + peakrpm
##
##                    Df Sum of Sq        RSS    AIC
## <none>                          1315706397 3241.3
## - compressionratio  1  17628383 1333334780 3242.0
## - horsepower        1  25839112 1341545509 3243.3
## - fueltype          1  30601932 1346308329 3244.0
## - carwidth          1  46974022 1362680419 3246.5
## - drivewheel        1  56610752 1372317149 3247.9
## - stroke            1  58433177 1374139575 3248.2
## - peakrpm           1  63788631 1379495028 3249.0
## - wheelbase         1  70042664 1385749061 3249.9
## - enginetype        1  87913403 1403619800 3252.6
## - CarName           1  88897396 1404603793 3252.7
## - cylindernumber    1 222851923 1538558320 3271.4
## - enginelocation    1 250876662 1566583059 3275.1
## - enginesize        1 453850116 1769556513 3300.1
```

```
##
## Call:
## lm(formula = price ~ CarName + fueltype + drivewheel + enginelocation +
##     wheelbase + carwidth + enginetype + cylindernumber + enginesize +
##     stroke + compressionratio + horsepower + peakrpm, data = data)
```

```
##
## Coefficients:
##      (Intercept)           CarName          fueltype        drivewheel
##       -53534.637           132.029          9242.231          1149.270
##   enginelocation         wheelbase          carwidth        enginetype
##        11194.717           192.613           520.832          -609.407
##   cylindernumber        enginesize            stroke  compressionratio
##         1377.244            91.229         -2147.224          -505.609
##       horsepower           peakrpm
##           23.616             1.749
```

```
#AIC=3241.3

model2 = lm(price ~ CarName + fueltype + drivewheel + enginelocation +
              wheelbase + carwidth + enginetype + cylindernumber + enginesize +
              stroke + compressionratio + horsepower + peakrpm, data = data)

summary(model2)
```

```
##
## Call:
## lm(formula = price ~ CarName + fueltype + drivewheel + enginelocation +
##     wheelbase + carwidth + enginetype + cylindernumber + enginesize +
##     stroke + compressionratio + horsepower + peakrpm, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -9141  -1500   -244   1267  14084
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -5.353e+04  1.083e+04  -4.943 1.68e-06 ***
## CarName           1.320e+02  3.675e+01   3.592 0.000417 ***
## fueltype          9.242e+03  4.385e+03   2.108 0.036360 *
## drivewheel        1.149e+03  4.009e+02   2.867 0.004613 **
## enginelocation    1.119e+04  1.855e+03   6.035 8.11e-09 ***
## wheelbase         1.926e+02  6.040e+01   3.189 0.001670 **
## carwidth          5.208e+02  1.994e+02   2.611 0.009735 **
## enginetype       -6.094e+02  1.706e+02  -3.572 0.000448 ***
## cylindernumber    1.377e+03  2.421e+02   5.688 4.77e-08 ***
## enginesize        9.123e+01  1.124e+01   8.117 5.74e-14 ***
## stroke           -2.147e+03  7.372e+02  -2.913 0.004013 **
## compressionratio -5.056e+02  3.161e+02  -1.600 0.111314
## horsepower        2.362e+01  1.219e+01   1.937 0.054250 .
## peakrpm           1.749e+00  5.748e-01   3.043 0.002672 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 2625 on 191 degrees of freedom
## Multiple R-squared:  0.8989, Adjusted R-squared:  0.8921
## F-statistic: 130.7 on 13 and 191 DF,  p-value: < 2.2e-16

#Adjusted R-squared:  0.8921
#model2 is our best model so far.

# Estimate Std. Error t value Pr(>|t|)
# (Intercept)      -5.353e+04  1.083e+04  -4.943 1.68e-06 ***
#   CarName           1.320e+02  3.675e+01   3.592 0.000417 ***
#   fueltype          9.242e+03  4.385e+03   2.108 0.036360 *
#   drivewheel        1.149e+03  4.009e+02   2.867 0.004613 **
#   enginelocation    1.119e+04  1.855e+03   6.035 8.11e-09 ***
#   wheelbase         1.926e+02  6.040e+01   3.189 0.001670 **
#   carwidth          5.208e+02  1.994e+02   2.611 0.009735 **
#   enginetype       -6.094e+02  1.706e+02  -3.572 0.000448 ***
#   cylindernumber    1.377e+03  2.421e+02   5.688 4.77e-08 ***
#   enginesize        9.123e+01  1.124e+01   8.117 5.74e-14 ***
#   stroke           -2.147e+03  7.372e+02  -2.913 0.004013 **
#   compressionratio -5.056e+02  3.161e+02  -1.600 0.111314
# horsepower        2.362e+01  1.219e+01   1.937 0.054250 .
# peakrpm           1.749e+00  5.748e-01   3.043 0.002672 **

# We want to check interaction terms

library(MASS)

# We want to check if the model2 check the LINE checkmarks.
```

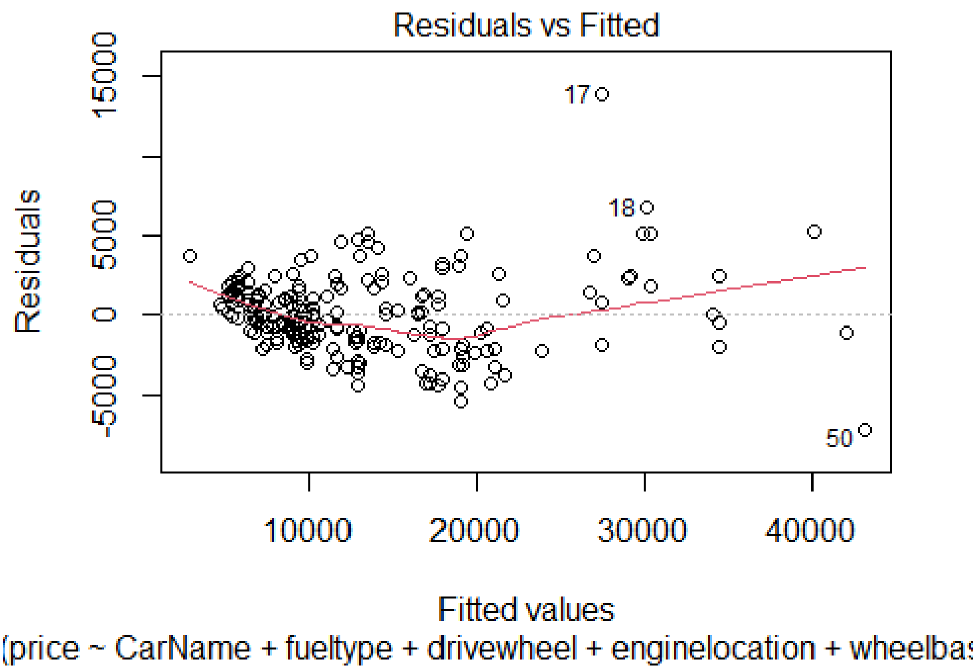
```
#plot(model2)


#we can check for potential significant interaction terms before doing any transfo
rmations.


add1(model2, ~. + fueltype*CarName + drivewheel*CarName + enginelocation*CarName +
       wheelbase*CarName + carwidth*CarName + enginetype*CarName + cylindernumber*
CarName + enginesize*CarName +
       stroke*CarName + compressionratio*CarName + horsepower*CarName + peakrpm*Ca
rName, test = 'F')

## Single term additions
##
## Model:
## price ~ CarName + fueltype + drivewheel + enginelocation + wheelbase +
##     carwidth + enginetype + cylindernumber + enginesize + stroke +
##     compressionratio + horsepower + peakrpm
##                         Df Sum of Sq        RSS    AIC F value   Pr(>F)
## <none>                                1315706397 3241.3
## CarName:fueltype         1    189737 1315516660 3243.3  0.0274 0.868694
## CarName:drivewheel       1  18228257 1297478140 3240.4  2.6693 0.103956
## CarName:enginelocation   0         0 1315706397 3241.3
## CarName:wheelbase        1  44752587 1270953810 3236.2  6.6902 0.010442 *
## CarName:carwidth         1  29892597 1285813800 3238.6  4.4171 0.036898 *
## CarName:enginetype       1  48585730 1267120667 3235.6  7.2852 0.007579 **
## CarName:cylindernumber   1  28371156 1287335241 3238.8  4.1873 0.042104 *
## CarName:enginesize       1  44438659 1271267738 3236.3  6.6417 0.010720 *
## CarName:stroke           1   1436145 1314270252 3243.1  0.2076 0.649160
## CarName:compressionratio 1   1076328 1314630069 3243.1  0.1556 0.693721
## CarName:horsepower       1  38213746 1277492651 3237.3  5.6835 0.018110 *
## CarName:peakrpm          1   1263459 1314442939 3243.1  0.1826 0.669606
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#The results are amazing to see, I would have expected for all interactions to be
significant
#but based on the interaction terms only wheelbase,carwidth,enginetype,cylindernum
ber,enginesize,horsepower are
#significant.

model3 = update(model2, ~.+CarName:cylindernumber)
summary(model3)

##
## Call:
## lm(formula = price ~ CarName + fueltype + drivewheel + enginelocation +
```

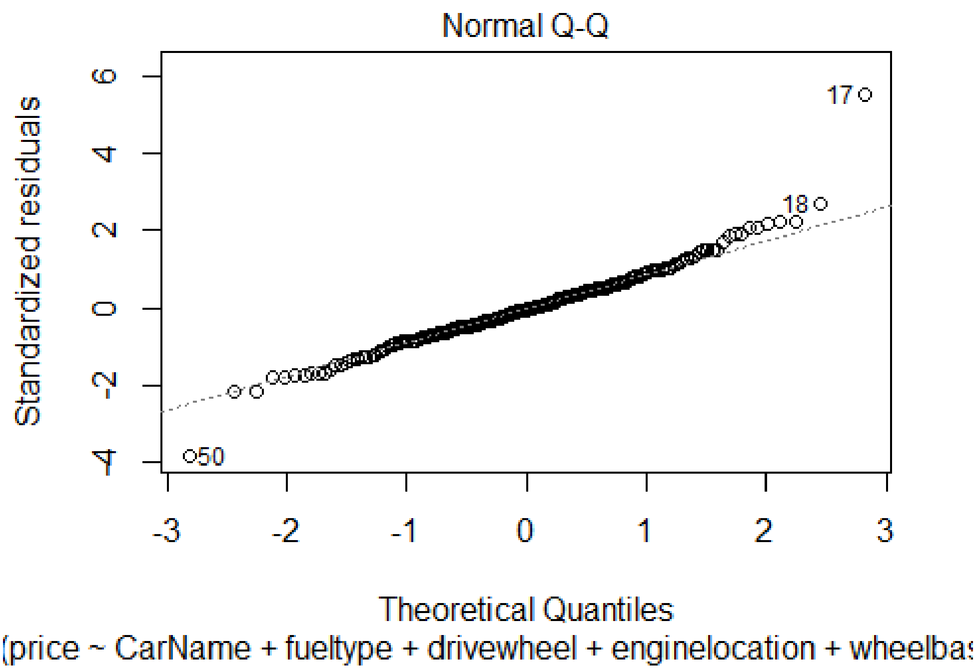
```
##     wheelbase + carwidth + enginetype + cylindernumber + enginesize +
##     stroke + compressionratio + horsepower + peakrpm + CarName:cylindernumber,
##     data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7168.8 -1611.7  -132.5  1380.1 13861.7
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -4.989e+04  1.089e+04  -4.582 8.31e-06 ***
## CarName                  1.616e+02  3.921e+01   4.121 5.62e-05 ***
## fueltype                 8.748e+03  4.356e+03   2.009 0.046003 *
## drivewheel               1.036e+03  4.015e+02   2.579 0.010655 *
## enginelocation           1.105e+04  1.841e+03   6.003 9.66e-09 ***
## wheelbase                1.933e+02  5.991e+01   3.227 0.001475 **
## carwidth                 4.704e+02  1.993e+02   2.360 0.019284 *
## enginetype              -6.854e+02  1.732e+02  -3.957 0.000107 ***
## cylindernumber           2.053e+03  4.085e+02   5.027 1.15e-06 ***
## enginesize               9.628e+01  1.142e+01   8.433 8.33e-15 ***
## stroke                  -2.372e+03  7.393e+02  -3.208 0.001571 **
## compressionratio        -4.644e+02  3.141e+02  -1.479 0.140914
## horsepower               2.334e+01  1.209e+01   1.930 0.055130 .
## peakrpm                  1.605e+00  5.744e-01   2.794 0.005734 **
## CarName:cylindernumber -6.855e+01  3.350e+01  -2.046 0.042104 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2603 on 190 degrees of freedom
## Multiple R-squared:  0.9011, Adjusted R-squared:  0.8938
## F-statistic: 123.7 on 14 and 190 DF,  p-value: < 2.2e-16
```

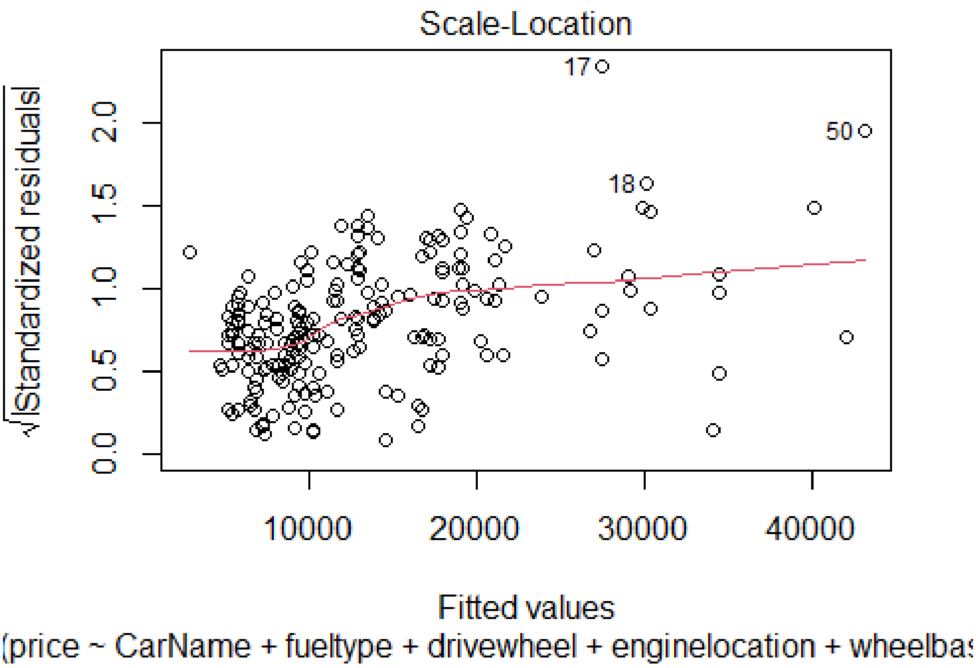
```
anova(model2,model3)
```

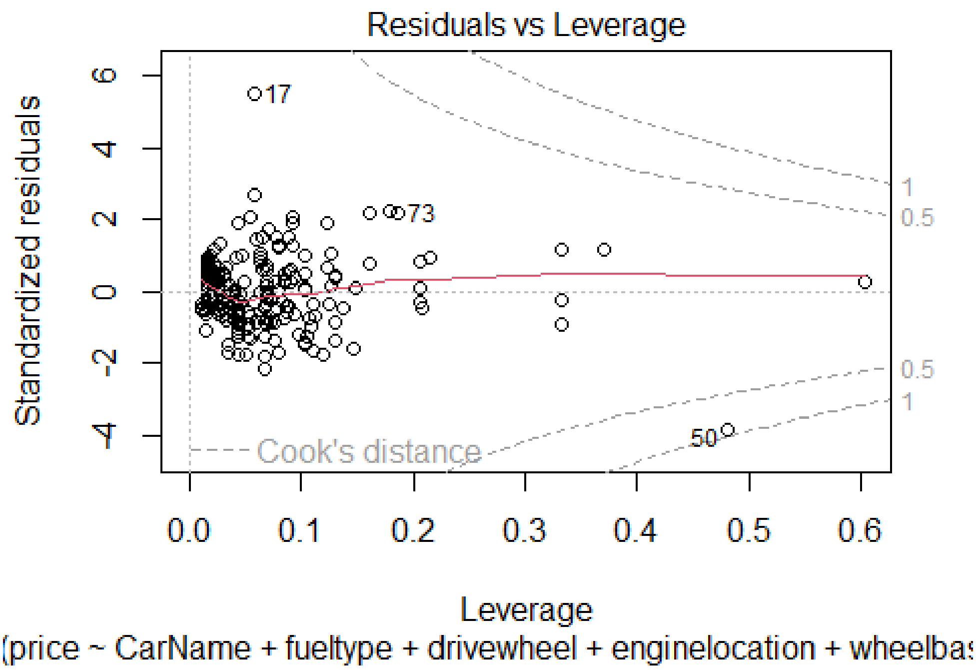
```
## Analysis of Variance Table
##
## Model 1: price ~ CarName + fueltype + drivewheel + enginelocation + wheelbase +
##     carwidth + enginetype + cylindernumber + enginesize + stroke +
##     compressionratio + horsepower + peakrpm
## Model 2: price ~ CarName + fueltype + drivewheel + enginelocation + wheelbase +
##     carwidth + enginetype + cylindernumber + enginesize + stroke +
##     compressionratio + horsepower + peakrpm + CarName:cylindernumber
##   Res.Df        RSS Df Sum of Sq      F Pr(>F)
## 1    191 1315706397
## 2    190 1287335241  1  28371156 4.1873 0.0421 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(model3)
```


Residuals vs Fitted
Residuals
15000
5000
0
-5000
17
18
50
10000
20000
30000
40000
Fitted values
(price ~ CarName + fueltype + drivewheel + enginelocation + wheelbas


Normal Q-Q
Standardized residuals
6
4
2
0
-2
-4
17
18
50
-3
-2
-1
0
1
2
3
Theoretical Quantiles
(price ~ CarName + fueltype + drivewheel + enginelocation + wheelbas



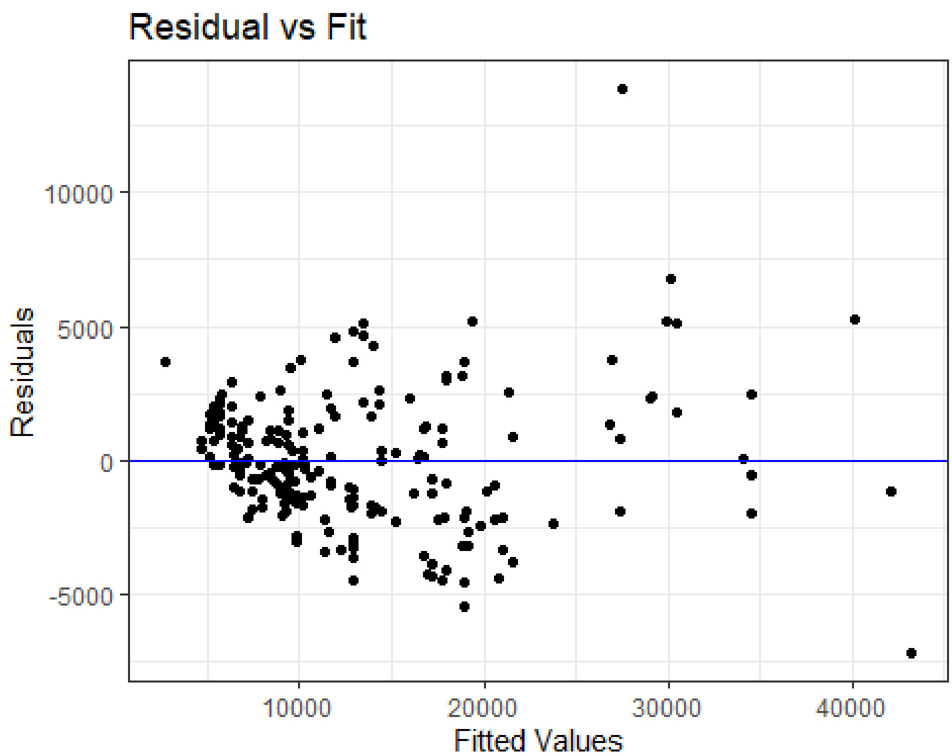



```
#In the residual vs Fitted plot we see a fanning pattern
#In the QQ plot we see a non normality.

library(broom)
#Another way to plot these graphs are:
```

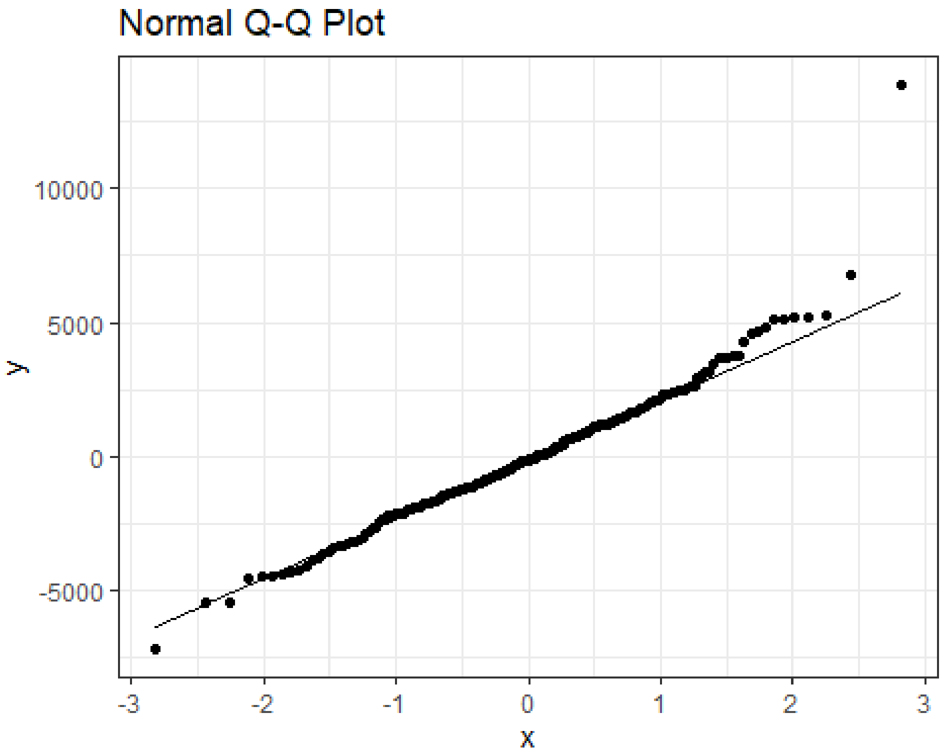
```
mod_table = augment(model3)
names(model3)
```

```
##  [1] "coefficients"  "residuals"     "effects"       "rank"
##  [5] "fitted.values" "assign"        "qr"            "df.residual"
##  [9] "xlevels"       "call"          "terms"         "model"
```

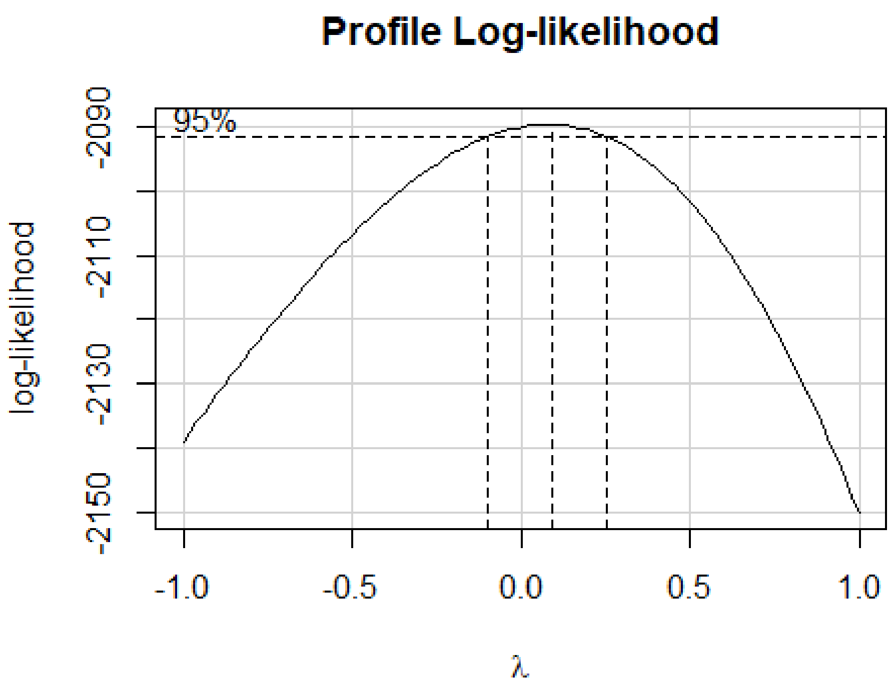
```
ggplot(mod_table, aes(x = .fitted, y = .resid)) +
  geom_point() +
  geom_hline(yintercept = 0, colour = 'blue') +
  labs(x = 'Fitted Values', y = 'Residuals') +
  ggtitle('Residual vs Fit') +
  theme_bw()
```




```
#Here we see a fanning pattern
#We need to transform the y-variable
ggplot(mod_table, aes(sample = .resid)) +
  stat_qq() +
  stat_qq_line() +
  ggtitle('Normal Q-Q Plot') +
  theme_bw()
```



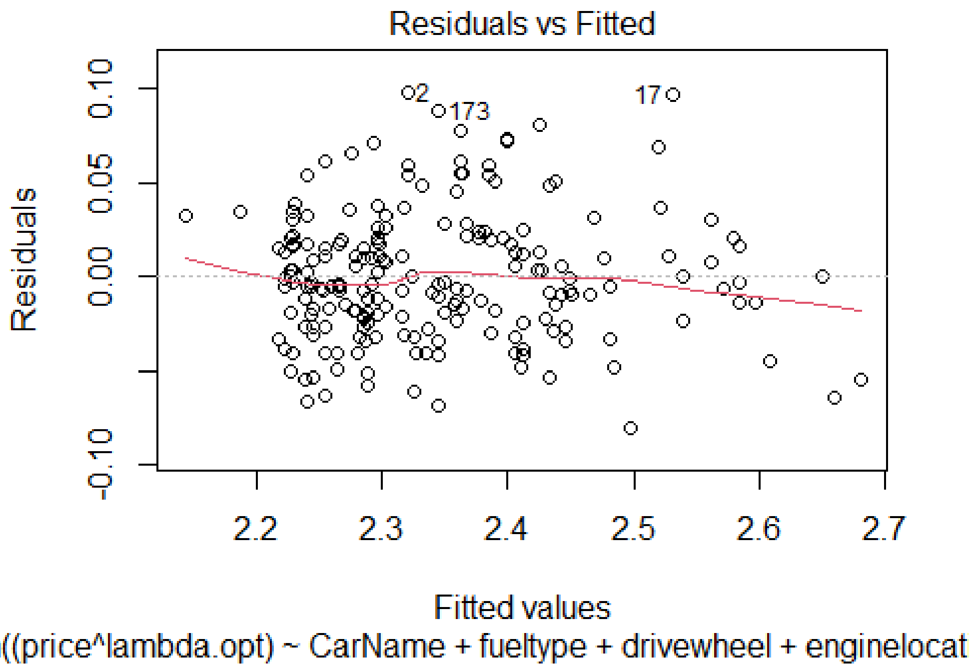

```
shapiro.test(resid(model3))

##
##  Shapiro-Wilk normality test
##
## data:  resid(model3)
## W = 0.95893, p-value = 1.191e-05

#We can try box-cox method

library(car)
mod.boxcox = boxCox(model3, lambda = seq(-1, 1, length=10))
```

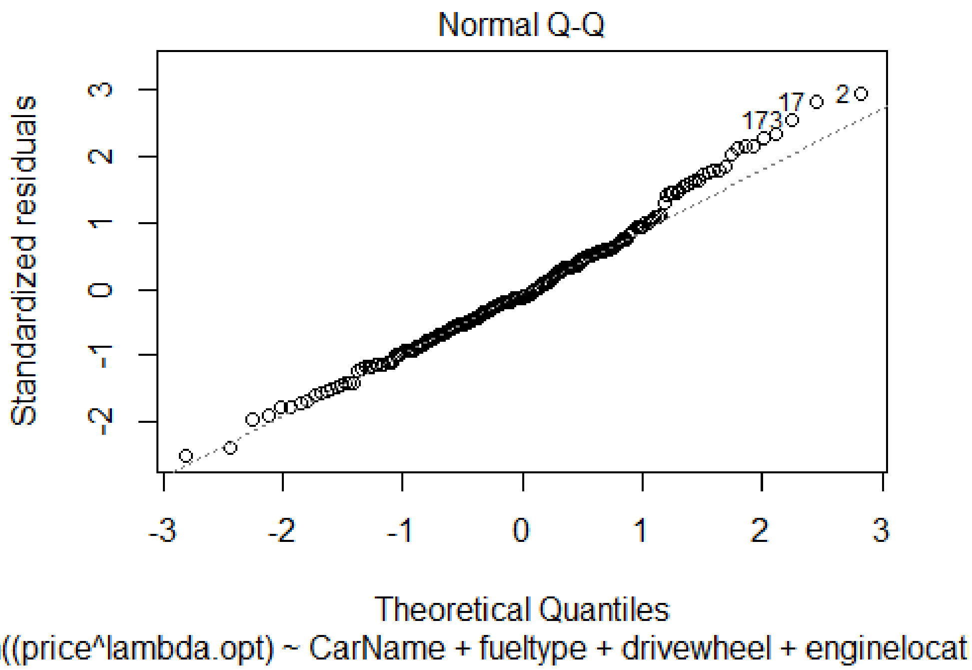


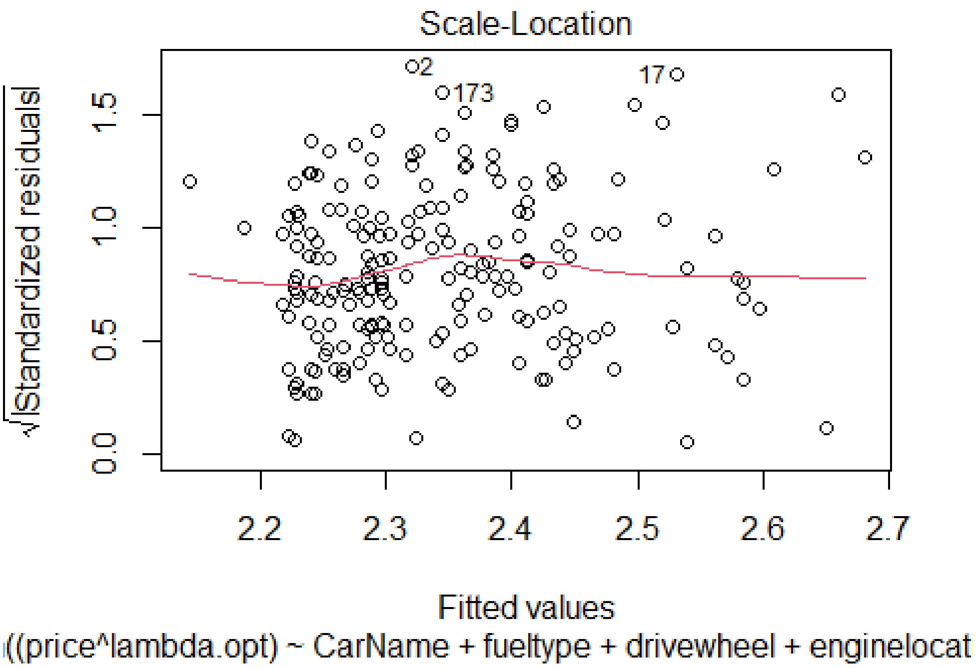

```
lambda.opt = mod.boxcox$x[which.max(mod.boxcox$y)] #find the optimal lambda
lambda.opt

## [1] 0.09090909

model4 = lm((price^lambda.opt) ~ CarName + fueltype + drivewheel + enginelocation
+
              wheelbase + carwidth + enginetype + cylindernumber + enginesize +
              stroke + compressionratio + horsepower + peakrpm + CarName:cylindern
umber,
            data = data)
summary(model4)

##
## Call:
## lm(formula = (price^lambda.opt) ~ CarName + fueltype + drivewheel +
##     enginelocation + wheelbase + carwidth + enginetype + cylindernumber +
##     enginesize + stroke + compressionratio + horsepower + peakrpm +
##     CarName:cylindernumber, data = data)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.080081 -0.023134 -0.004182  0.019577  0.098205
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
```

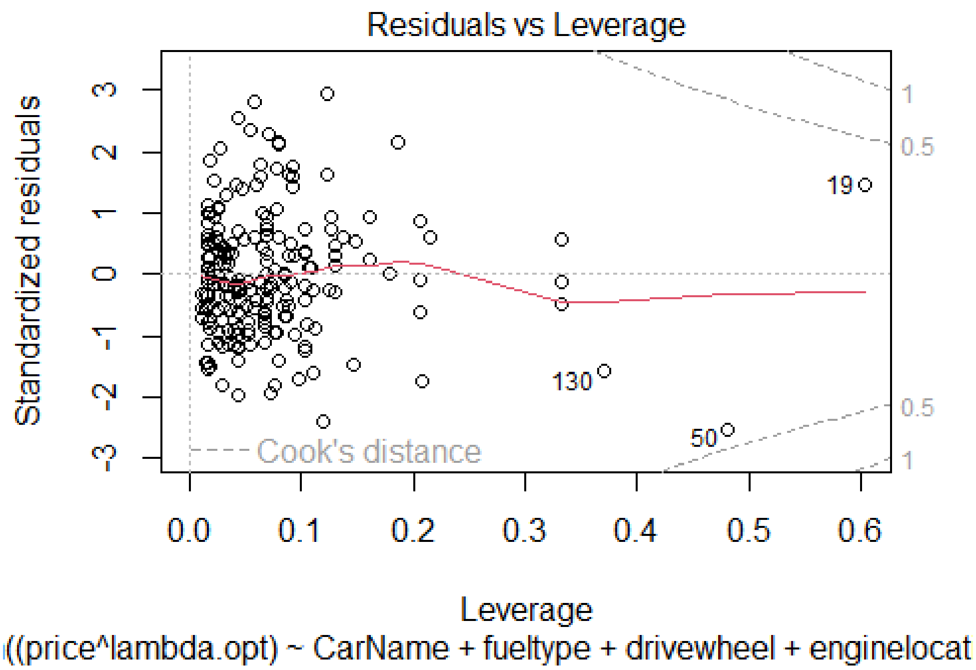
```
## (Intercept)             1.248e+00  1.494e-01   8.354 1.36e-14 ***
## CarName                 2.405e-03  5.379e-04   4.471 1.34e-05 ***
## fueltype                9.392e-02  5.975e-02   1.572 0.117648
## drivewheel              2.201e-02  5.507e-03   3.997 9.17e-05 ***
## enginelocation          1.111e-01  2.526e-02   4.401 1.80e-05 ***
## wheelbase               3.479e-03  8.218e-04   4.233 3.59e-05 ***
## carwidth                8.973e-03  2.734e-03   3.282 0.001229 **
## enginetype             -8.216e-03  2.376e-03  -3.458 0.000672 ***
## cylindernumber          2.674e-02  5.604e-03   4.773 3.62e-06 ***
## enginesize              5.944e-04  1.566e-04   3.796 0.000198 ***
## stroke                 -1.887e-02  1.014e-02  -1.860 0.064373 .
## compressionratio       -3.884e-03  4.309e-03  -0.901 0.368572
## horsepower              9.514e-04  1.659e-04   5.735 3.79e-08 ***
## peakrpm                 1.007e-05  7.880e-06   1.279 0.202611
## CarName:cylindernumber -1.660e-03  4.595e-04  -3.612 0.000389 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03571 on 190 degrees of freedom
## Multiple R-squared:    0.9,  Adjusted R-squared:  0.8926
## F-statistic: 122.2 on 14 and 190 DF,  p-value: < 2.2e-16
```

```
plot(model4)
```

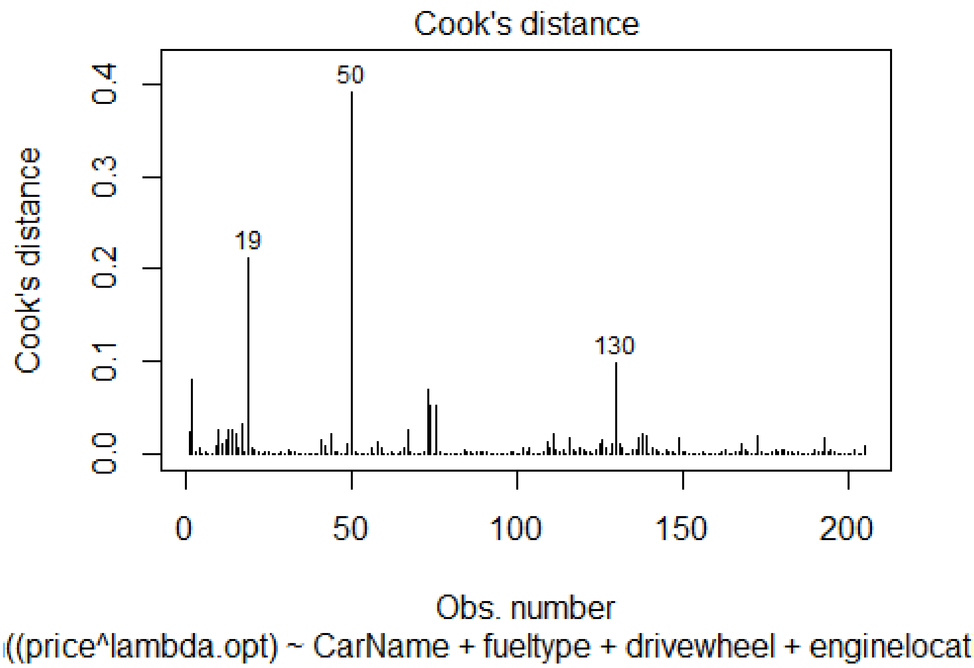

Residuals vs Fitted
Residuals
Fitted values
((price^lambda.opt) ~ CarName + fueltype + drivewheel + enginelocat
-0.10
0.00
0.05
0.10
2.2
2.3
2.4
2.5
2.6
2.7
2
173
17


Normal Q-Q
Standardized residuals
Theoretical Quantiles
((price^lambda.opt) ~ CarName + fueltype + drivewheel + enginelocat
-2
-1
0
1
2
3
-3
173
17
2






```
plot(model4, which = 4)
```




```
rs = rstandard(model4)
which(rs > 3)

## named integer(0)

#We see no influential points

#potential prediction and confidence intervals
library(dplyr)

##
## Attaching package: 'dplyr'

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

## The following object is masked from 'package:car':
##
##     recode

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
new_data = data %>% select (CarName, fueltype, drivewheel , enginelocation ,
                             wheelbase ,carwidth, enginetype,cylindernumber,engine
size ,
                             stroke , compressionratio , horsepower ,peakrpm, pric
e)
new_data
```

```
##     CarName fueltype drivewheel enginelocation wheelbase carwidth enginetype
## 1        18        0          1              0      88.6     64.1          3
## 2        18        0          1              0      88.6     64.1          3
## 3        18        0          1              0      94.5     65.5          2
## 4        12        0          0              0      99.8     66.2          0
## 5        12        0          2              0      99.4     66.4          0
## 6        12        0          0              0      99.8     66.3          0
## 7        12        0          0              0     105.8     71.4          0
## 8        12        0          0              0     105.8     71.4          0
## 9        12        0          0              0     105.8     71.4          0
## 10       12        0          2              0      99.5     67.9          0
## 11       11        0          1              0     101.2     64.8          0
## 12       11        0          1              0     101.2     64.8          0
## 13       11        0          1              0     101.2     64.8          0
## 14       11        0          1              0     101.2     64.8          0
## 15       11        0          1              0     103.5     66.9          0
## 16       11        0          1              0     103.5     66.9          0
## 17       11        0          1              0     103.5     67.9          0
## 18       11        0          1              0     110.0     70.9          0
## 19       16        0          0              0      88.4     60.3          4
## 20       16        0          0              0      94.5     63.6          0
## 21       16        0          0              0      94.5     63.6          0
## 22        9        0          0              0      93.7     63.8          0
## 23        9        0          0              0      93.7     63.8          0
## 24        9        0          0              0      93.7     63.8          0
## 25        9        0          0              0      93.7     63.8          0
## 26        9        0          0              0      93.7     63.8          0
## 27        9        0          0              0      93.7     63.8          0
## 28        9        0          0              0      93.7     63.8          0
## 29        9        0          0              0     103.3     64.6          0
## 30        9        0          0              0      95.9     66.3          0
## 31        3        0          0              0      86.6     63.9          0
## 32        3        0          0              0      86.6     63.9          0
## 33        3        0          0              0      93.7     64.0          0
## 34        3        0          0              0      93.7     64.0          0
## 35        3        0          0              0      93.7     64.0          0
## 36        3        0          0              0      96.5     64.0          0
## 37        3        0          0              0      96.5     63.9          0
## 38        3        0          0              0      96.5     65.2          0
## 39        3        0          0              0      96.5     65.2          0
```

```
## 40          3          0          0               0       96.5       65.2           0
## 41          3          0          0               0       96.5       62.5           0
## 42          3          0          0               0       96.5       65.2           0
## 43          3          0          0               0       96.5       66.0           0
## 44         17          0          1               0       94.3       61.8           0
## 45         17          0          0               0       94.5       63.6           0
## 46         17          0          0               0       94.5       63.6           0
## 47         17          0          1               0       96.0       65.2           0
## 48         19          0          1               0      113.0       69.6           3
## 49         19          0          1               0      113.0       69.6           3
## 50         19          0          1               0      102.0       70.6           2
## 51          2          0          0               0       93.1       64.2           0
## 52          2          0          0               0       93.1       64.2           0
## 53          2          0          0               0       93.1       64.2           0
## 54          2          0          0               0       93.1       64.2           0
## 55          2          0          0               0       93.1       64.2           0
## 56          2          0          1               0       95.3       65.7           5
## 57          2          0          1               0       95.3       65.7           5
## 58          2          0          1               0       95.3       65.7           5
## 59          2          0          1               0       95.3       65.7           5
## 60          2          0          0               0       98.8       66.5           0
## 61          2          0          0               0       98.8       66.5           0
## 62          2          0          0               0       98.8       66.5           0
## 63          2          0          0               0       98.8       66.5           0
## 64          2          1          0               0       98.8       66.5           0
## 65          2          0          0               0       98.8       66.5           0
## 66          2          0          1               0      104.9       66.1           0
## 67          2          1          1               0      104.9       66.1           0
## 68         10          1          1               0      110.0       70.3           0
## 69         10          1          1               0      110.0       70.3           0
## 70         10          1          1               0      106.7       70.3           0
## 71         10          1          1               0      115.6       71.7           0
## 72         10          0          1               0      115.6       71.7           2
## 73         10          0          1               0       96.6       70.5           2
## 74         10          0          1               0      120.9       71.7           2
## 75         10          0          1               0      112.0       72.0           2
## 76         21          0          1               0      102.7       68.0           0
## 77          4          0          0               0       93.7       64.4           0
## 78          4          0          0               0       93.7       64.4           0
## 79          4          0          0               0       93.7       64.4           0
## 80          4          0          0               0       93.0       63.8           0
## 81          4          0          0               0       96.3       65.4           0
## 82          4          0          0               0       96.3       65.4           0
## 83          4          0          0               0       95.9       66.3           0
## 84          4          0          0               0       95.9       66.3           0
## 85          4          0          0               0       95.9       66.3           0
## 86          4          0          0               0       96.3       65.4           0
```

```
## 87     4  0  0  0  96.3  65.4  0
## 88     4  0  0  0  96.3  65.4  0
## 89     4  0  0  0  96.3  65.4  0
## 90     1  0  0  0  94.5  63.8  0
## 91     1  1  0  0  94.5  63.8  0
## 92     1  0  0  0  94.5  63.8  0
## 93     1  0  0  0  94.5  63.8  0
## 94     1  0  0  0  94.5  63.8  0
## 95     1  0  0  0  94.5  63.8  0
## 96     1  0  0  0  94.5  63.8  0
## 97     1  0  0  0  94.5  63.8  0
## 98     1  0  0  0  94.5  63.8  0
## 99     1  0  0  0  95.1  63.8  0
## 100    1  0  0  0  97.2  65.2  0
## 101    1  0  0  0  97.2  65.2  0
## 102    1  0  0  0 100.4  66.5  2
## 103    1  0  0  0 100.4  66.5  2
## 104    1  0  0  0 100.4  66.5  2
## 105    1  0  1  0  91.3  67.9  2
## 106    1  0  1  0  91.3  67.9  2
## 107    1  0  1  0  99.2  67.9  2
## 108    7  0  1  0 107.9  68.4  4
## 109    7  1  1  0 107.9  68.4  4
## 110    7  0  1  0 114.2  68.4  4
## 111    7  1  1  0 114.2  68.4  4
## 112    7  0  1  0 107.9  68.4  4
## 113    7  1  1  0 107.9  68.4  4
## 114    7  0  1  0 114.2  68.4  4
## 115    7  1  1  0 114.2  68.4  4
## 116    7  0  1  0 107.9  68.4  4
## 117    7  1  1  0 107.9  68.4  4
## 118    7  0  1  0 108.0  68.3  4
## 119   13  0  0  0  93.7  63.8  0
## 120   13  0  0  0  93.7  63.8  0
## 121   13  0  0  0  93.7  63.8  0
## 122   13  0  0  0  93.7  63.8  0
## 123   13  0  0  0  93.7  63.8  0
## 124   13  0  0  0 103.3  64.6  0
## 125   13  0  1  0  95.9  66.3  0
## 126   15  0  1  0  94.5  68.3  0
## 127   15  0  1  1  89.5  65.0  1
## 128   15  0  1  1  89.5  65.0  1
## 129   15  0  1  1  89.5  65.0  1
## 130   15  0  1  0  98.4  72.3  6
## 131   20  0  0  0  96.1  66.5  0
## 132   20  0  0  0  96.1  66.6  0
## 133   14  0  0  0  99.1  66.5  0
```

```
## 134    14    0    0    0    99.1   66.5    0
## 135    14    0    0    0    99.1   66.5    0
## 136    14    0    0    0    99.1   66.5    0
## 137    14    0    0    0    99.1   66.5    3
## 138    14    0    0    0    99.1   66.5    3
## 139     5    0    0    0    93.7   63.4    1
## 140     5    0    0    0    93.7   63.6    1
## 141     5    0    2    0    93.3   63.8    1
## 142     5    0    0    0    97.2   65.4    1
## 143     5    0    0    0    97.2   65.4    1
## 144     5    0    0    0    97.2   65.4    1
## 145     5    0    2    0    97.0   65.4    1
## 146     5    0    2    0    97.0   65.4    1
## 147     5    0    0    0    97.0   65.4    1
## 148     5    0    0    0    97.0   65.4    1
## 149     5    0    2    0    96.9   65.4    1
## 150     5    0    2    0    96.9   65.4    1
## 151     0    0    0    0    95.7   63.6    0
## 152     0    0    0    0    95.7   63.6    0
## 153     0    0    0    0    95.7   63.6    0
## 154     0    0    0    0    95.7   63.6    0
## 155     0    0    2    0    95.7   63.6    0
## 156     0    0    2    0    95.7   63.6    0
## 157     0    0    0    0    95.7   64.4    0
## 158     0    0    0    0    95.7   64.4    0
## 159     0    1    0    0    95.7   64.4    0
## 160     0    1    0    0    95.7   64.4    0
## 161     0    0    0    0    95.7   64.4    0
## 162     0    0    0    0    95.7   64.4    0
## 163     0    0    0    0    95.7   64.4    0
## 164     0    0    1    0    94.5   64.0    0
## 165     0    0    1    0    94.5   64.0    0
## 166     0    0    1    0    94.5   64.0    3
## 167     0    0    1    0    94.5   64.0    3
## 168     0    0    1    0    98.4   65.6    0
## 169     0    0    1    0    98.4   65.6    0
## 170     0    0    1    0    98.4   65.6    0
## 171     0    0    1    0    98.4   65.6    0
## 172     0    0    1    0    98.4   65.6    0
## 173     0    0    1    0    98.4   65.6    0
## 174     0    0    0    0   102.4   66.5    0
## 175     0    1    0    0   102.4   66.5    0
## 176     0    0    0    0   102.4   66.5    0
## 177     0    0    0    0   102.4   66.5    0
## 178     0    0    0    0   102.4   66.5    0
## 179     0    0    1    0   102.9   67.7    3
## 180     0    0    1    0   102.9   67.7    3
```

```
## 181       0        0          1              0     104.5     66.5          3
## 182       0        0          1              0     104.5     66.5          3
## 183       6        1          0              0      97.3     65.5          0
## 184       6        0          0              0      97.3     65.5          0
## 185       6        1          0              0      97.3     65.5          0
## 186       6        0          0              0      97.3     65.5          0
## 187       6        0          0              0      97.3     65.5          0
## 188       6        1          0              0      97.3     65.5          0
## 189       6        0          0              0      97.3     65.5          0
## 190       6        0          0              0      94.5     64.2          0
## 191       6        0          0              0      94.5     64.0          0
## 192       6        0          0              0     100.4     66.9          0
## 193       6        1          0              0     100.4     66.9          0
## 194       6        0          0              0     100.4     66.9          0
## 195       8        0          1              0     104.3     67.2          0
## 196       8        0          1              0     104.3     67.2          0
## 197       8        0          1              0     104.3     67.2          0
## 198       8        0          1              0     104.3     67.2          0
## 199       8        0          1              0     104.3     67.2          0
## 200       8        0          1              0     104.3     67.2          0
## 201       8        0          1              0     109.1     68.9          0
## 202       8        0          1              0     109.1     68.8          0
## 203       8        0          1              0     109.1     68.9          2
## 204       8        1          1              0     109.1     68.9          0
## 205       8        0          1              0     109.1     68.9          0
##     cylindernumber enginesize stroke compressionratio horsepower peakrpm
## 1                0        130  2.680             9.00        111    5000
## 2                0        130  2.680             9.00        111    5000
## 3                1        152  3.470             9.00        154    5000
## 4                0        109  3.400            10.00        102    5500
## 5                2        136  3.400             8.00        115    5500
## 6                2        136  3.400             8.50        110    5500
## 7                2        136  3.400             8.50        110    5500
## 8                2        136  3.400             8.50        110    5500
## 9                2        131  3.400             8.30        140    5500
## 10               2        131  3.400             7.00        160    5500
## 11               0        108  2.800             8.80        101    5800
## 12               0        108  2.800             8.80        101    5800
## 13               1        164  3.190             9.00        121    4250
## 14               1        164  3.190             9.00        121    4250
## 15               1        164  3.190             9.00        121    4250
## 16               1        209  3.390             8.00        182    5400
## 17               1        209  3.390             8.00        182    5400
## 18               1        209  3.390             8.00        182    5400
## 19               6         61  3.030             9.50         48    5100
## 20               0         90  3.110             9.60         70    5400
## 21               0         90  3.110             9.60         70    5400
```

```
## 22          0      90  3.230       9.41      68   5500
## 23          0      90  3.230       9.40      68   5500
## 24          0      98  3.390       7.60     102   5500
## 25          0      90  3.230       9.40      68   5500
## 26          0      90  3.230       9.40      68   5500
## 27          0      90  3.230       9.40      68   5500
## 28          0      98  3.390       7.60     102   5500
## 29          0     122  3.460       8.50      88   5000
## 30          0     156  3.900       7.00     145   5000
## 31          0      92  3.410       9.60      58   4800
## 32          0      92  3.410       9.20      76   6000
## 33          0      79  3.070      10.10      60   5500
## 34          0      92  3.410       9.20      76   6000
## 35          0      92  3.410       9.20      76   6000
## 36          0      92  3.410       9.20      76   6000
## 37          0      92  3.410       9.20      76   6000
## 38          0     110  3.580       9.00      86   5800
## 39          0     110  3.580       9.00      86   5800
## 40          0     110  3.580       9.00      86   5800
## 41          0     110  3.580       9.00      86   5800
## 42          0     110  3.580       9.00     101   5800
## 43          0     110  3.580       9.10     100   5500
## 44          0     111  3.230       8.50      78   4800
## 45          0      90  3.110       9.60      70   5400
## 46          0      90  3.110       9.60      70   5400
## 47          0     119  3.230       9.20      90   5000
## 48          1     258  4.170       8.10     176   4750
## 49          1     258  4.170       8.10     176   4750
## 50          5     326  2.760      11.50     262   5000
## 51          0      91  3.150       9.00      68   5000
## 52          0      91  3.150       9.00      68   5000
## 53          0      91  3.150       9.00      68   5000
## 54          0      91  3.150       9.00      68   5000
## 55          0      91  3.150       9.00      68   5000
## 56          4      70  3.255       9.40     101   6000
## 57          4      70  3.255       9.40     101   6000
## 58          4      70  3.255       9.40     101   6000
## 59          4      80  3.255       9.40     135   6000
## 60          0     122  3.390       8.60      84   4800
## 61          0     122  3.390       8.60      84   4800
## 62          0     122  3.390       8.60      84   4800
## 63          0     122  3.390       8.60      84   4800
## 64          0     122  3.390      22.70      64   4650
## 65          0     122  3.390       8.60      84   4800
## 66          0     140  3.160       8.00     120   5000
## 67          0     134  3.640      22.00      72   4200
## 68          2     183  3.640      21.50     123   4350
```

```
## 69                   2         183  3.640             21.50         123     4350
## 70                   2         183  3.640             21.50         123     4350
## 71                   2         183  3.640             21.50         123     4350
## 72                   3         234  3.100              8.30         155     4750
## 73                   3         234  3.100              8.30         155     4750
## 74                   3         308  3.350              8.00         184     4500
## 75                   3         304  3.350              8.00         184     4500
## 76                   0         140  3.120              8.00         175     5000
## 77                   0          92  3.230              9.40          68     5500
## 78                   0          92  3.230              9.40          68     5500
## 79                   0          92  3.230              9.40          68     5500
## 80                   0          98  3.390              7.60         102     5500
## 81                   0         110  3.460              7.50         116     5500
## 82                   0         122  3.460              8.50          88     5000
## 83                   0         156  3.860              7.00         145     5000
## 84                   0         156  3.860              7.00         145     5000
## 85                   0         156  3.860              7.00         145     5000
## 86                   0         122  3.460              8.50          88     5000
## 87                   0         122  3.460              8.50          88     5000
## 88                   0         110  3.460              7.50         116     5500
## 89                   0         110  3.460              7.50         116     5500
## 90                   0          97  3.290              9.40          69     5200
## 91                   0         103  3.470             21.90          55     4800
## 92                   0          97  3.290              9.40          69     5200
## 93                   0          97  3.290              9.40          69     5200
## 94                   0          97  3.290              9.40          69     5200
## 95                   0          97  3.290              9.40          69     5200
## 96                   0          97  3.290              9.40          69     5200
## 97                   0          97  3.290              9.40          69     5200
## 98                   0          97  3.290              9.40          69     5200
## 99                   0          97  3.290              9.40          69     5200
## 100                  0         120  3.470              8.50          97     5200
## 101                  0         120  3.470              8.50          97     5200
## 102                  1         181  3.270              9.00         152     5200
## 103                  1         181  3.270              9.00         152     5200
## 104                  1         181  3.270              9.00         152     5200
## 105                  1         181  3.270              9.00         160     5200
## 106                  1         181  3.270              7.80         200     5200
## 107                  1         181  3.270              9.00         160     5200
## 108                  0         120  3.190              8.40          97     5000
## 109                  0         152  3.520             21.00          95     4150
## 110                  0         120  3.190              8.40          97     5000
## 111                  0         152  3.520             21.00          95     4150
## 112                  0         120  2.190              8.40          95     5000
## 113                  0         152  3.520             21.00          95     4150
## 114                  0         120  2.190              8.40          95     5000
## 115                  0         152  3.520             21.00          95     4150
```

```
## 116                    0         120  3.190             8.40          97     5000
## 117                    0         152  3.520            21.00          95     4150
## 118                    0         134  3.210             7.00         142     5600
## 119                    0          90  3.230             9.40          68     5500
## 120                    0          98  3.390             7.60         102     5500
## 121                    0          90  3.230             9.40          68     5500
## 122                    0          90  3.230             9.40          68     5500
## 123                    0          98  3.230             9.40          68     5500
## 124                    0         122  3.460             8.50          88     5000
## 125                    0         156  3.860             7.00         145     5000
## 126                    0         151  3.110             9.50         143     5500
## 127                    1         194  2.900             9.50         207     5900
## 128                    1         194  2.900             9.50         207     5900
## 129                    1         194  2.900             9.50         207     5900
## 130                    3         203  3.110            10.00         288     5750
## 131                    0         132  3.900             8.70          90     5100
## 132                    0         132  3.900             8.70          90     5100
## 133                    0         121  3.070             9.31         110     5250
## 134                    0         121  3.070             9.30         110     5250
## 135                    0         121  2.070             9.30         110     5250
## 136                    0         121  3.070             9.30         110     5250
## 137                    0         121  3.070             9.00         160     5500
## 138                    0         121  3.070             9.00         160     5500
## 139                    0          97  2.360             9.00          69     4900
## 140                    0         108  2.640             8.70          73     4400
## 141                    0         108  2.640             8.70          73     4400
## 142                    0         108  2.640             9.50          82     4800
## 143                    0         108  2.640             9.50          82     4400
## 144                    0         108  2.640             9.00          94     5200
## 145                    0         108  2.640             9.00          82     4800
## 146                    0         108  2.640             7.70         111     4800
## 147                    0         108  2.640             9.00          82     4800
## 148                    0         108  2.640             9.00          94     5200
## 149                    0         108  2.640             9.00          82     4800
## 150                    0         108  2.640             7.70         111     4800
## 151                    0          92  3.030             9.00          62     4800
## 152                    0          92  3.030             9.00          62     4800
## 153                    0          92  3.030             9.00          62     4800
## 154                    0          92  3.030             9.00          62     4800
## 155                    0          92  3.030             9.00          62     4800
## 156                    0          92  3.030             9.00          62     4800
## 157                    0          98  3.030             9.00          70     4800
## 158                    0          98  3.030             9.00          70     4800
## 159                    0         110  3.350            22.50          56     4500
## 160                    0         110  3.350            22.50          56     4500
## 161                    0          98  3.030             9.00          70     4800
## 162                    0          98  3.030             9.00          70     4800
```

```
## 163        0        98  3.030          9.00       70   4800
## 164        0        98  3.030          9.00       70   4800
## 165        0        98  3.030          9.00       70   4800
## 166        0        98  3.080          9.40      112   6600
## 167        0        98  3.080          9.40      112   6600
## 168        0       146  3.500          9.30      116   4800
## 169        0       146  3.500          9.30      116   4800
## 170        0       146  3.500          9.30      116   4800
## 171        0       146  3.500          9.30      116   4800
## 172        0       146  3.500          9.30      116   4800
## 173        0       146  3.500          9.30      116   4800
## 174        0       122  3.540          8.70       92   4200
## 175        0       110  3.350         22.50       73   4500
## 176        0       122  3.540          8.70       92   4200
## 177        0       122  3.540          8.70       92   4200
## 178        0       122  3.540          8.70       92   4200
## 179        1       171  3.350          9.30      161   5200
## 180        1       171  3.350          9.30      161   5200
## 181        1       171  3.350          9.20      156   5200
## 182        1       161  3.350          9.20      156   5200
## 183        0        97  3.400         23.00       52   4800
## 184        0       109  3.400          9.00       85   5250
## 185        0        97  3.400         23.00       52   4800
## 186        0       109  3.400          9.00       85   5250
## 187        0       109  3.400          9.00       85   5250
## 188        0        97  3.400         23.00       68   4500
## 189        0       109  3.400         10.00      100   5500
## 190        0       109  3.400          8.50       90   5500
## 191        0       109  3.400          8.50       90   5500
## 192        2       136  3.400          8.50      110   5500
## 193        0        97  3.400         23.00       68   4500
## 194        0       109  3.400          9.00       88   5500
## 195        0       141  3.150          9.50      114   5400
## 196        0       141  3.150          9.50      114   5400
## 197        0       141  3.150          9.50      114   5400
## 198        0       141  3.150          9.50      114   5400
## 199        0       130  3.150          7.50      162   5100
## 200        0       130  3.150          7.50      162   5100
## 201        0       141  3.150          9.50      114   5400
## 202        0       141  3.150          8.70      160   5300
## 203        1       173  2.870          8.80      134   5500
## 204        1       145  3.400         23.00      106   4800
## 205        0       141  3.150          9.50      114   5400
##         price
## 1    13495.00
## 2    16500.00
## 3    16500.00
```

```
## 4   13950.00
## 5   17450.00
## 6   15250.00
## 7   17710.00
## 8   18920.00
## 9   23875.00
## 10  17859.17
## 11  16430.00
## 12  16925.00
## 13  20970.00
## 14  21105.00
## 15  24565.00
## 16  30760.00
## 17  41315.00
## 18  36880.00
## 19   5151.00
## 20   6295.00
## 21   6575.00
## 22   5572.00
## 23   6377.00
## 24   7957.00
## 25   6229.00
## 26   6692.00
## 27   7609.00
## 28   8558.00
## 29   8921.00
## 30  12964.00
## 31   6479.00
## 32   6855.00
## 33   5399.00
## 34   6529.00
## 35   7129.00
## 36   7295.00
## 37   7295.00
## 38   7895.00
## 39   9095.00
## 40   8845.00
## 41  10295.00
## 42  12945.00
## 43  10345.00
## 44   6785.00
## 45   8916.50
## 46   8916.50
## 47  11048.00
## 48  32250.00
## 49  35550.00
## 50  36000.00
```

```
## 51   5195.00
## 52   6095.00
## 53   6795.00
## 54   6695.00
## 55   7395.00
## 56  10945.00
## 57  11845.00
## 58  13645.00
## 59  15645.00
## 60   8845.00
## 61   8495.00
## 62  10595.00
## 63  10245.00
## 64  10795.00
## 65  11245.00
## 66  18280.00
## 67  18344.00
## 68  25552.00
## 69  28248.00
## 70  28176.00
## 71  31600.00
## 72  34184.00
## 73  35056.00
## 74  40960.00
## 75  45400.00
## 76  16503.00
## 77   5389.00
## 78   6189.00
## 79   6669.00
## 80   7689.00
## 81   9959.00
## 82   8499.00
## 83  12629.00
## 84  14869.00
## 85  14489.00
## 86   6989.00
## 87   8189.00
## 88   9279.00
## 89   9279.00
## 90   5499.00
## 91   7099.00
## 92   6649.00
## 93   6849.00
## 94   7349.00
## 95   7299.00
## 96   7799.00
## 97   7499.00
```

```
## 98   7999.00
## 99   8249.00
## 100  8949.00
## 101  9549.00
## 102 13499.00
## 103 14399.00
## 104 13499.00
## 105 17199.00
## 106 19699.00
## 107 18399.00
## 108 11900.00
## 109 13200.00
## 110 12440.00
## 111 13860.00
## 112 15580.00
## 113 16900.00
## 114 16695.00
## 115 17075.00
## 116 16630.00
## 117 17950.00
## 118 18150.00
## 119  5572.00
## 120  7957.00
## 121  6229.00
## 122  6692.00
## 123  7609.00
## 124  8921.00
## 125 12764.00
## 126 22018.00
## 127 32528.00
## 128 34028.00
## 129 37028.00
## 130 31400.50
## 131  9295.00
## 132  9895.00
## 133 11850.00
## 134 12170.00
## 135 15040.00
## 136 15510.00
## 137 18150.00
## 138 18620.00
## 139  5118.00
## 140  7053.00
## 141  7603.00
## 142  7126.00
## 143  7775.00
## 144  9960.00
```

```
## 145  9233.00
## 146 11259.00
## 147  7463.00
## 148 10198.00
## 149  8013.00
## 150 11694.00
## 151  5348.00
## 152  6338.00
## 153  6488.00
## 154  6918.00
## 155  7898.00
## 156  8778.00
## 157  6938.00
## 158  7198.00
## 159  7898.00
## 160  7788.00
## 161  7738.00
## 162  8358.00
## 163  9258.00
## 164  8058.00
## 165  8238.00
## 166  9298.00
## 167  9538.00
## 168  8449.00
## 169  9639.00
## 170  9989.00
## 171 11199.00
## 172 11549.00
## 173 17669.00
## 174  8948.00
## 175 10698.00
## 176  9988.00
## 177 10898.00
## 178 11248.00
## 179 16558.00
## 180 15998.00
## 181 15690.00
## 182 15750.00
## 183  7775.00
## 184  7975.00
## 185  7995.00
## 186  8195.00
## 187  8495.00
## 188  9495.00
## 189  9995.00
## 190 11595.00
## 191  9980.00
```

```
## 192 13295.00
## 193 13845.00
## 194 12290.00
## 195 12940.00
## 196 13415.00
## 197 15985.00
## 198 16515.00
## 199 18420.00
## 200 18950.00
## 201 16845.00
## 202 19045.00
## 203 21485.00
## 204 22470.00
## 205 22625.00
```

```
model4 = lm((price^lambda.opt) ~ CarName + fueltype + drivewheel + enginelocation
+
              wheelbase + carwidth + enginetype + cylindernumber + enginesize +
              stroke + compressionratio + horsepower + peakrpm + CarName:cylindern
umber,
            data = new_data)

new = data.frame(CarName=15, fueltype = 1,drivewheel = 0,
                 enginelocation= 0, wheelbase = 100, carwidth = 7,
                 cylindernumber = 5, enginesize = 250, stroke = 3,
                 compressionratio = 17, horsepower = 18, peakrpm = 4150, enginetyp
e=0)
predict(model4, new,interval = 'prediction', level = 0.95)
```

```
##        fit      lwr     upr
## 1 1.882564 1.541358 2.22377
```

```
#(1.541358 2.22377)
predict(model4, new,interval = 'confidence', level = 0.95)
```

```
##        fit      lwr      upr
## 1 1.882564 1.548707 2.216421
```

```
#(1.548707 2.216421)
```