STAT456 Final Project

Michelle Harris, Sri Medicherla, Hailey Lee

2024-04-19

This project aims to solve the business problem of Geely Auto. Through this analysis, we aim to answer the two business questions: Which variables significantly predict the price of a car and how well do those variables explain car prices. The steps taken to approach this analysis include: conduct exploratory data analysis (EDA), split dataset into training and testing, investigate each variable's effect on price through modeling, and finally determine which model is the strongest through testing each model. Our final solution will be based on our conclusions and findings.

#Reading in the dataset

```
car_price_data <- read.csv("C://Users//shell//OneDrive//Documents//STAT463//Final Project//CarPrice_Ass
```

```
car_price_data$symboling <- as.character(car_price_data$symboling)
```

#EDA

```
summary(car_price_data)
```

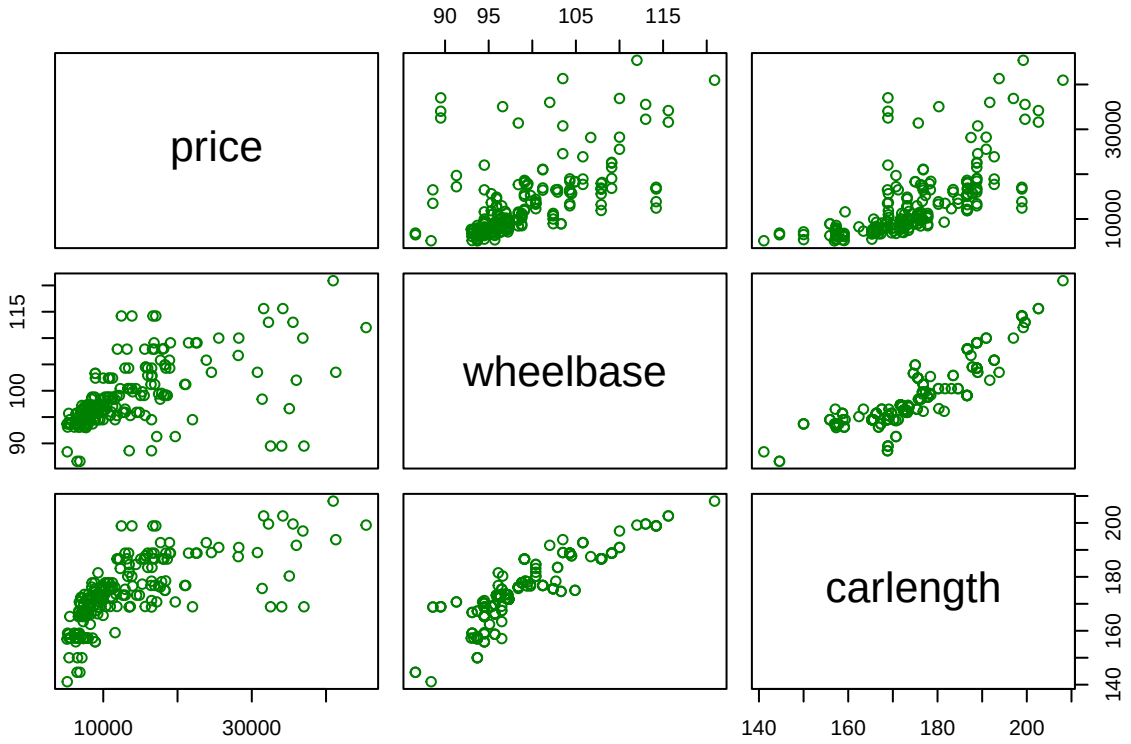
```
##      car_ID     symboling           CarName            fueltype
##  Min.   :  1   Length:205         Length:205         Length:205
##  1st Qu.: 52   Class :character   Class :character   Class :character
##  Median :103   Mode  :character   Mode  :character   Mode  :character
##  Mean   :103
##  3rd Qu.:154
##  Max.   :205
##   aspiration         doornumber          carbody           drivewheel
##  Length:205         Length:205         Length:205         Length:205
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##  enginelocation       wheelbase        carlength        carwidth
##  Length:205         Min.   : 86.60   Min.   :141.1   Min.   :60.30
##  Class :character   1st Qu.: 94.50   1st Qu.:166.3   1st Qu.:64.10
##  Mode  :character   Median : 97.00   Median :173.2   Median :65.50
##                     Mean   : 98.76   Mean   :174.0   Mean   :65.91
##                     3rd Qu.:102.40   3rd Qu.:183.1   3rd Qu.:66.90
##                     Max.   :120.90   Max.   :208.1   Max.   :72.30
##    carheight       curbweight    enginetype        cylindernumber
##  Min.   :47.80   Min.   :1488   Length:205         Length:205
##  1st Qu.:52.00   1st Qu.:2145   Class :character   Class :character
```

```
##  Median :54.10   Median :2414   Mode  :character   Mode  :character
##  Mean   :53.72   Mean   :2556
##  3rd Qu.:55.50   3rd Qu.:2935
##  Max.   :59.80   Max.   :4066
##    enginesize     fuelsystem          boreratio        stroke
##  Min.   : 61.0   Length:205         Min.   :2.54   Min.   :2.070
##  1st Qu.: 97.0   Class :character   1st Qu.:3.15   1st Qu.:3.110
##  Median :120.0   Mode  :character   Median :3.31   Median :3.290
##  Mean   :126.9                      Mean   :3.33   Mean   :3.255
##  3rd Qu.:141.0                      3rd Qu.:3.58   3rd Qu.:3.410
##  Max.   :326.0                      Max.   :3.94   Max.   :4.170
##  compressionratio   horsepower       peakrpm        citympg
##  Min.   : 7.00    Min.   : 48.0   Min.   :4150   Min.   :13.00
##  1st Qu.: 8.60    1st Qu.: 70.0   1st Qu.:4800   1st Qu.:19.00
##  Median : 9.00    Median : 95.0   Median :5200   Median :24.00
##  Mean   :10.14    Mean   :104.1   Mean   :5125   Mean   :25.22
##  3rd Qu.: 9.40    3rd Qu.:116.0   3rd Qu.:5500   3rd Qu.:30.00
##  Max.   :23.00    Max.   :288.0   Max.   :6600   Max.   :49.00
##    highwaympg        price
##  Min.   :16.00   Min.   : 5118
##  1st Qu.:25.00   1st Qu.: 7788
##  Median :30.00   Median :10295
##  Mean   :30.75   Mean   :13277
##  3rd Qu.:34.00   3rd Qu.:16503
##  Max.   :54.00   Max.   :45400
```

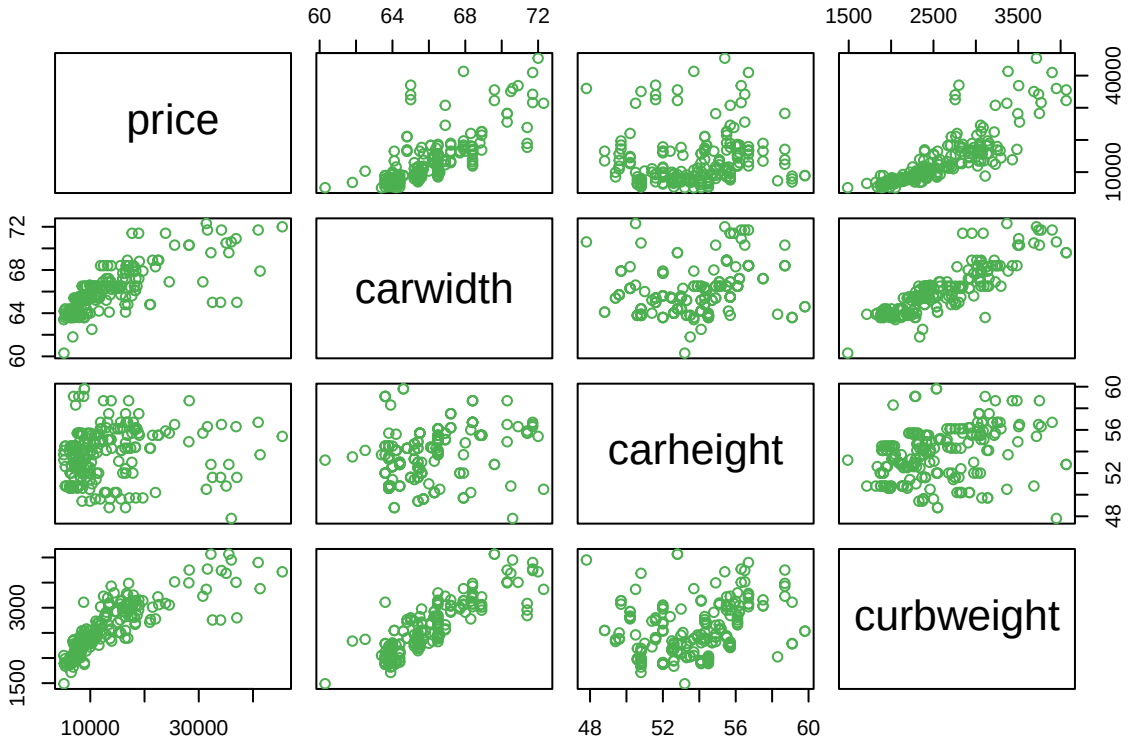
```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

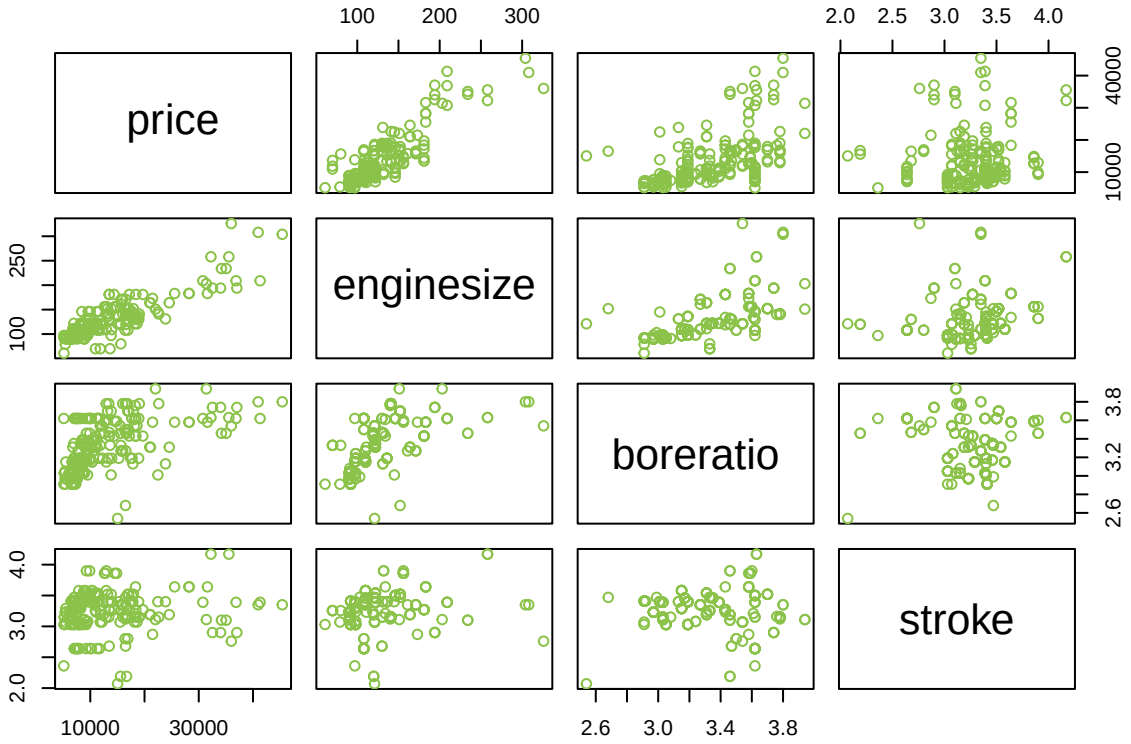
```
pairs(car_price_data[, c('price', 'wheelbase', 'carlength')], col = '#008000')
```



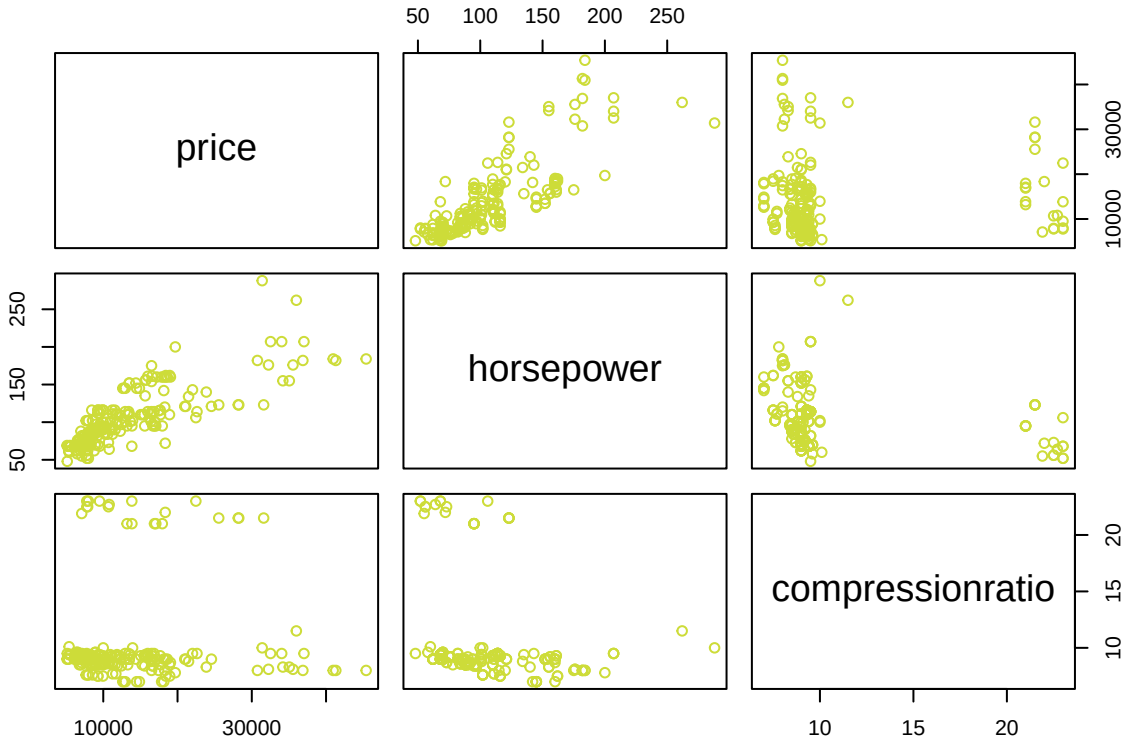

```
pairs(car_price_data[, c('price', 'carwidth', 'carheight', 'curbweight')], col = '#4CAF50')
```



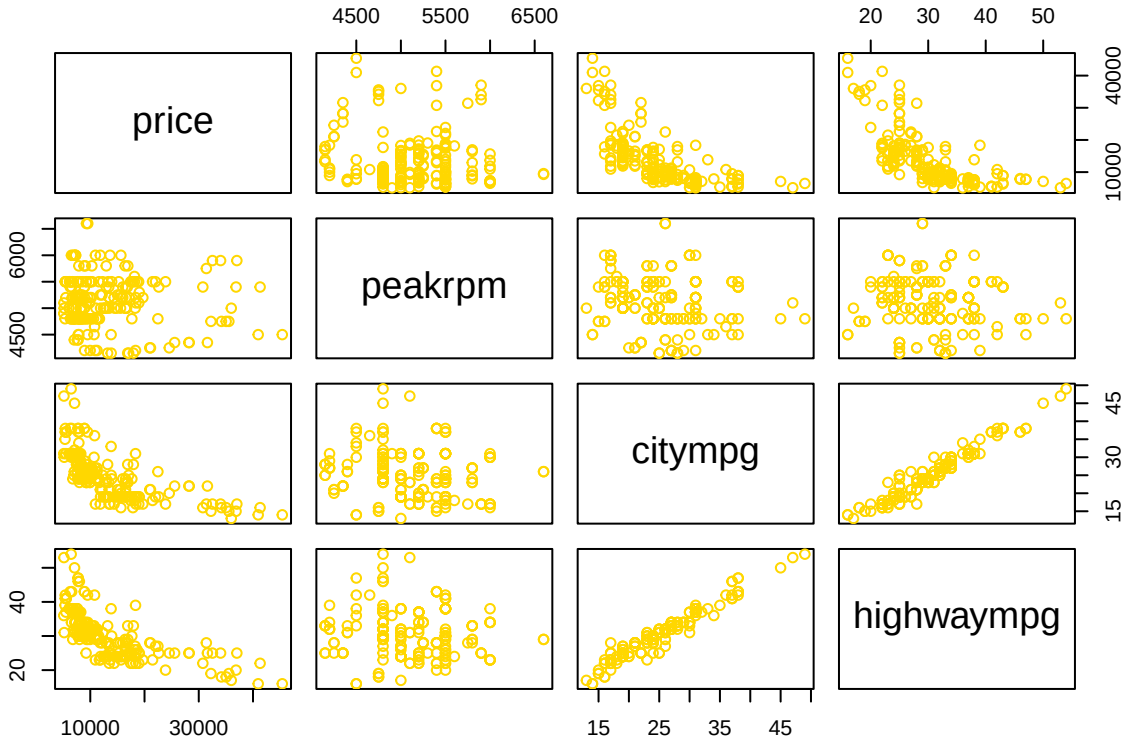

```
pairs(car_price_data[, c('price', 'enginesize', 'boreratio', 'stroke')], col = '#8BC34A')
```




```
pairs(car_price_data[, c('price', 'horsepower', 'compressionratio')], col = '#CDDC39')
```




```
pairs(car_price_data[, c('price', 'peakrpm', 'citympg', 'highwaympg')], col = '#FFD700')
```

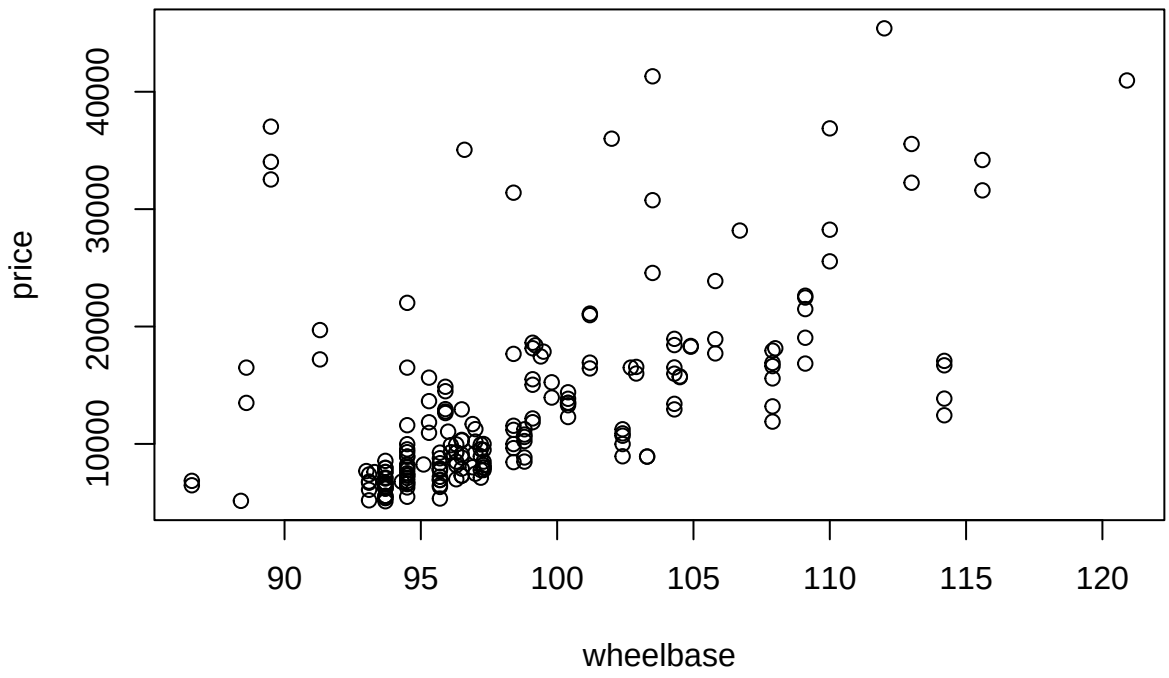


#EDA Findings

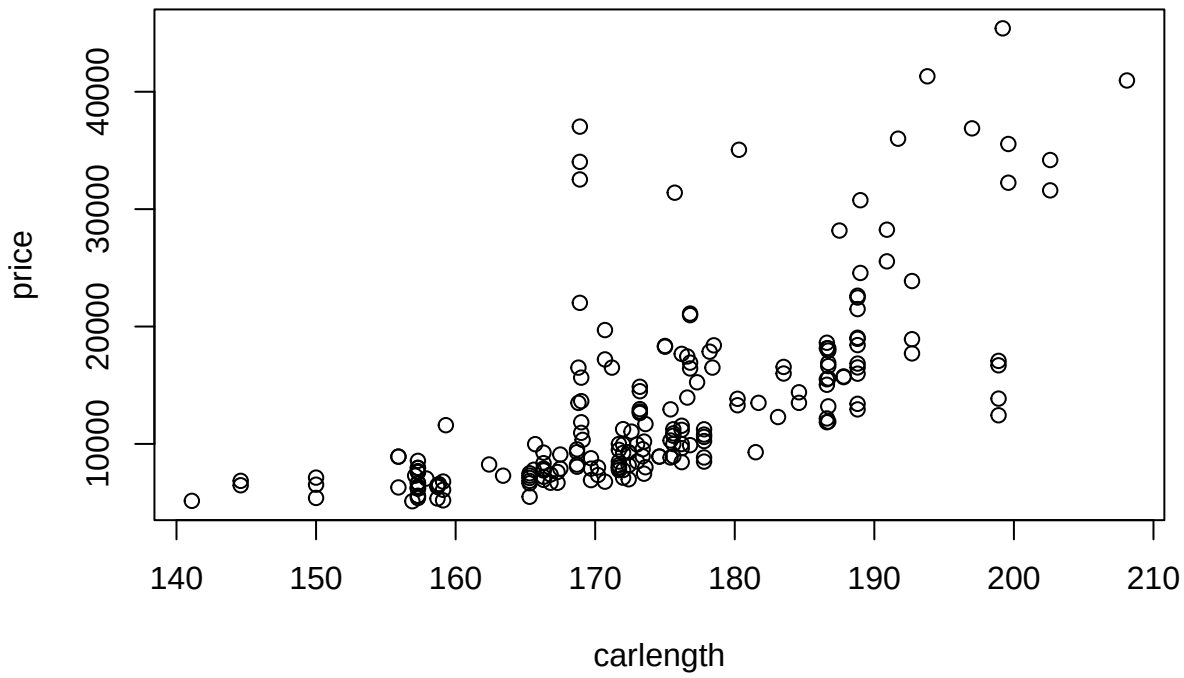
Visually, 'wheelbase', 'carlength', 'carwidth', 'curbweight', 'enginesize', 'horsepower', 'citympg', and 'highwaympg' are associated with 'price.' Some appear to have a linear relationship with price; for example, highway mpg, engine size, and horsepower.

#EDA plots based on findings

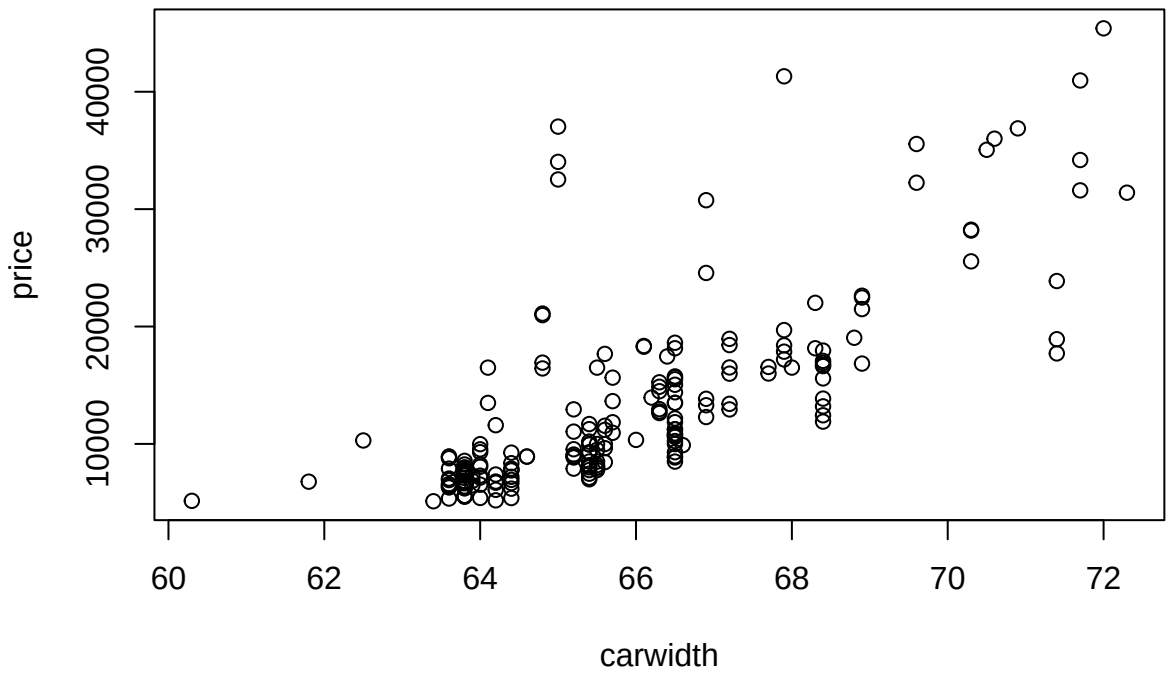
```
plot(price ~ wheelbase, data = car_price_data)
```



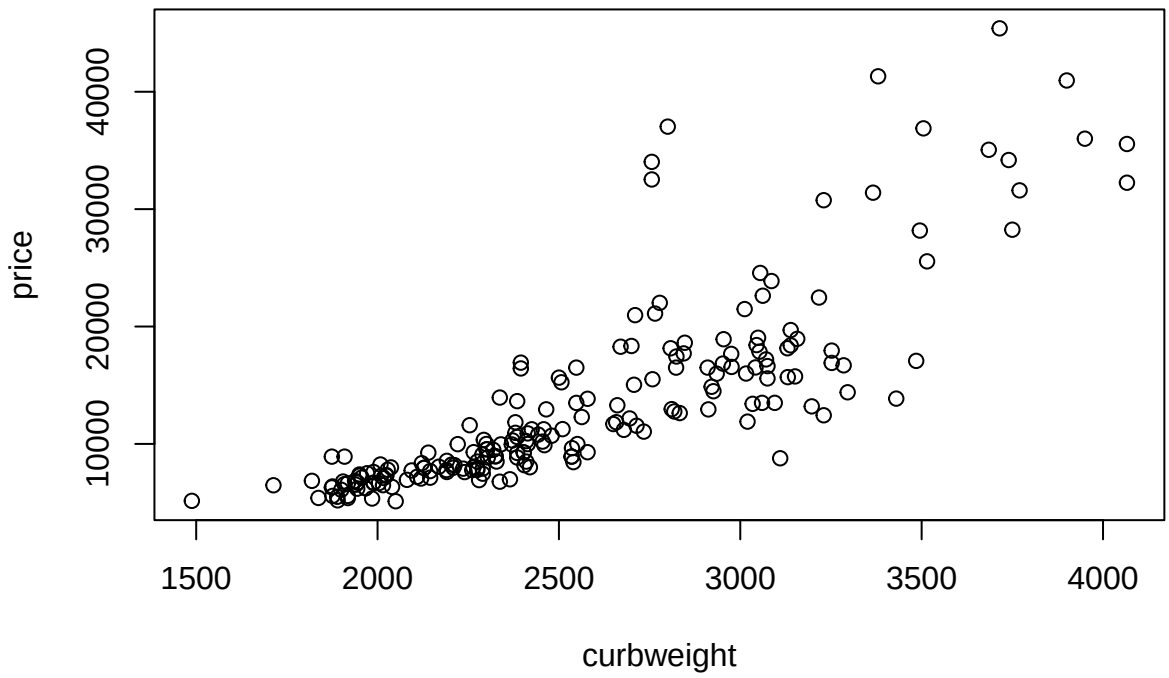

```
plot(price ~ carlength, data = car_price_data)
```



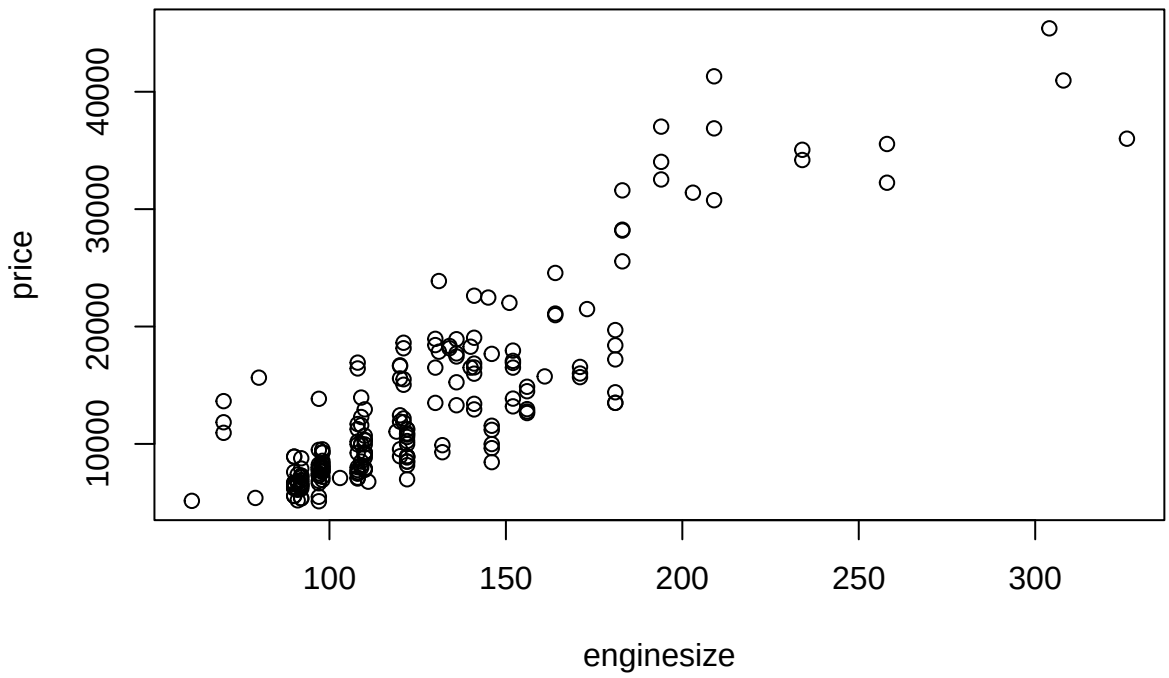

```
plot(price ~ carwidth, data = car_price_data)
```



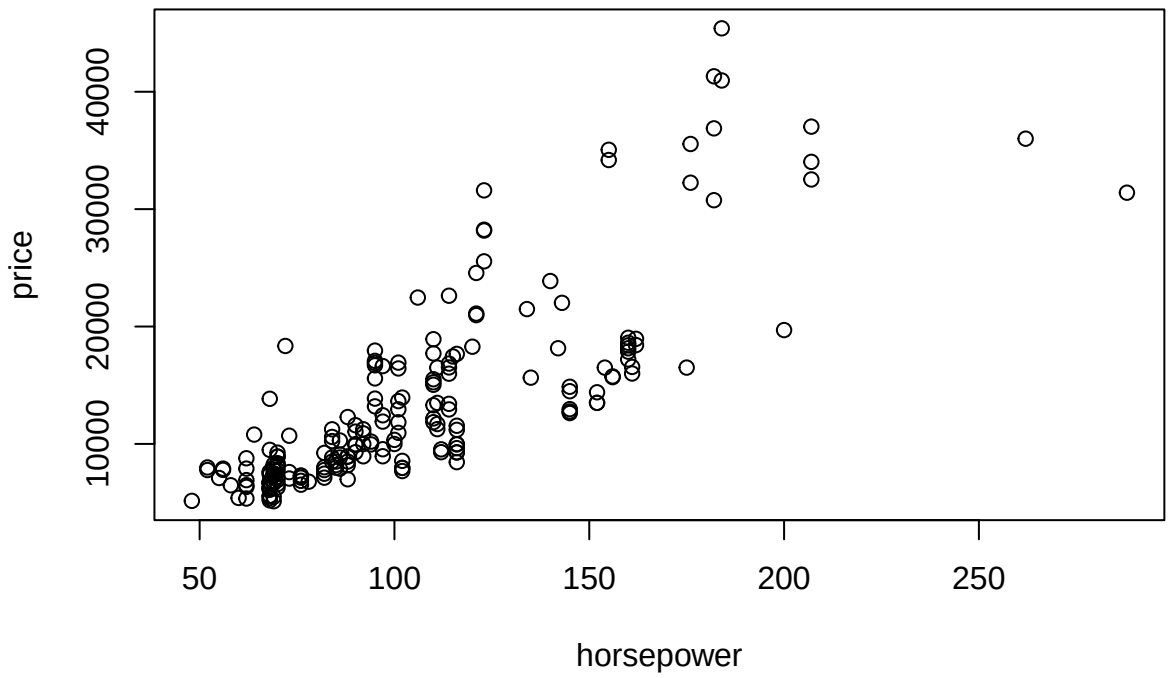

```
plot(price ~ curbweight, data = car_price_data)
```



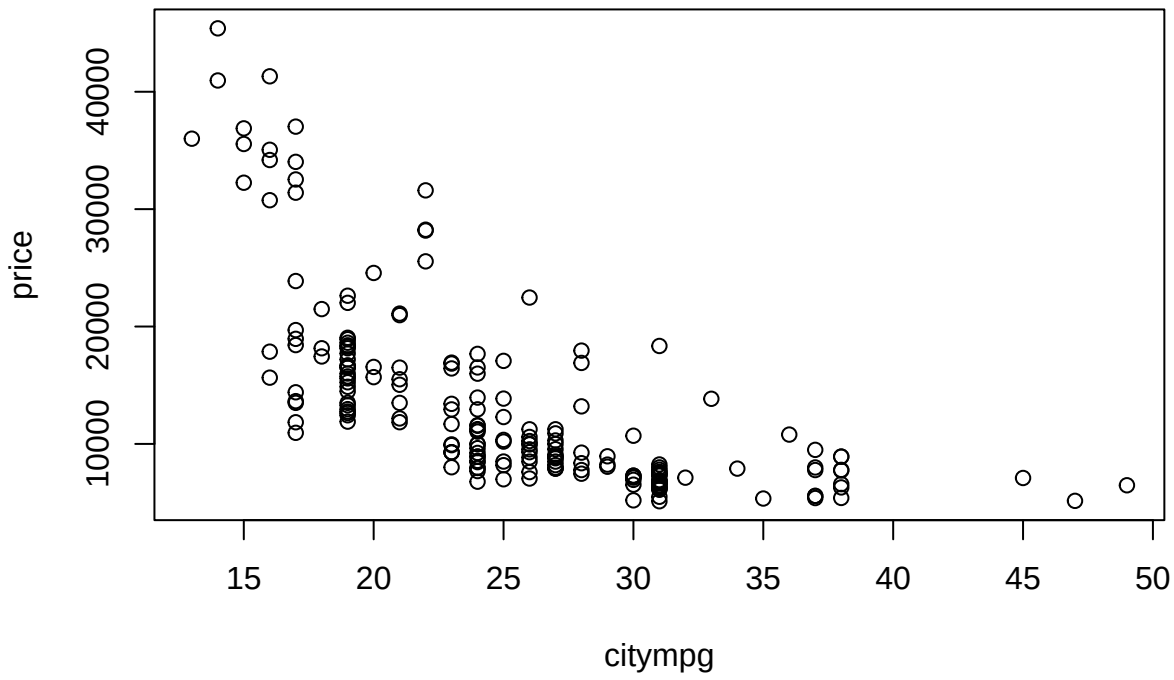

```
plot(price ~ enginesize, data = car_price_data)
```



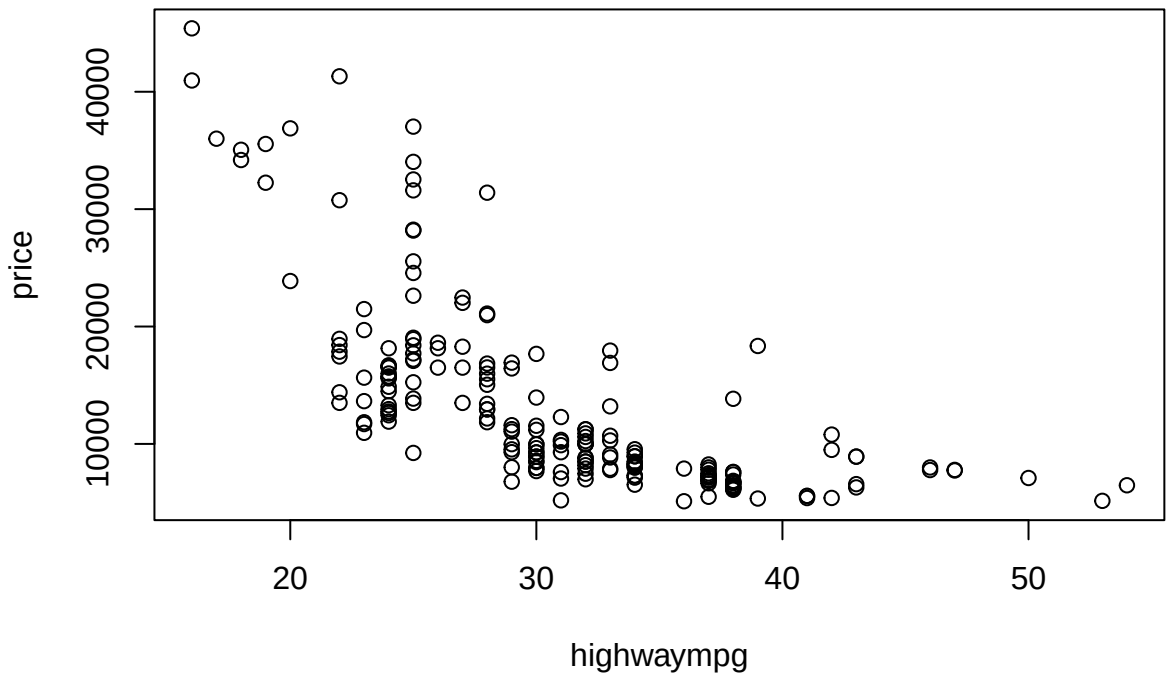

```
plot(price ~ horsepower, data = car_price_data)
```



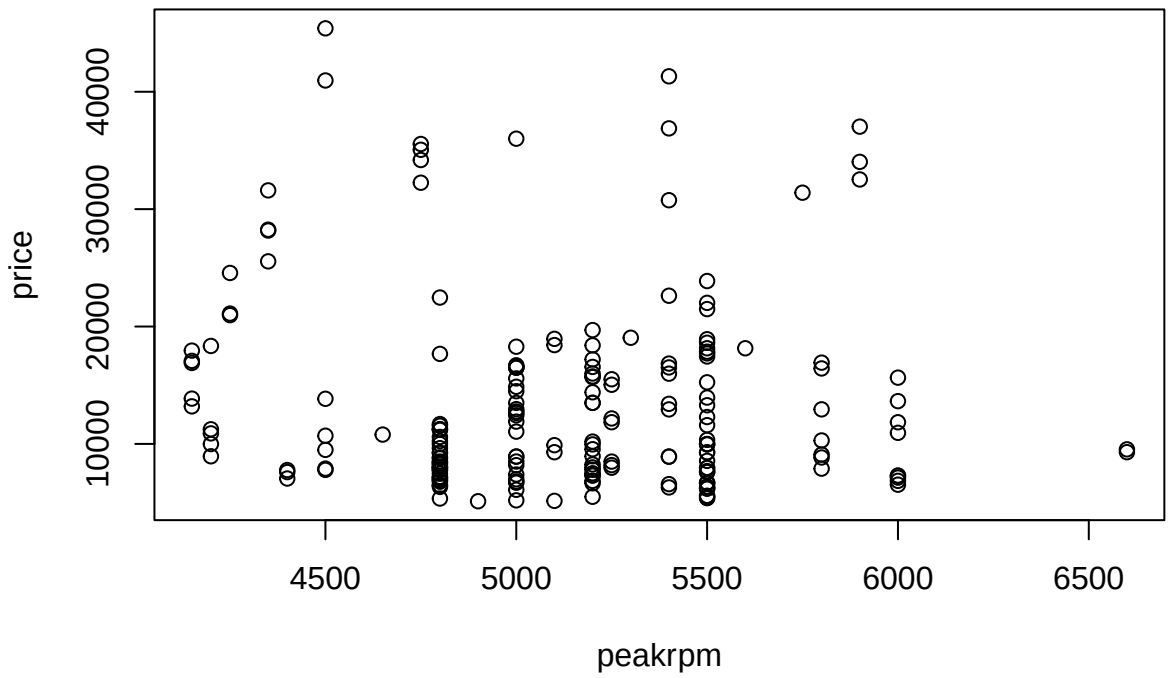

```
plot(price ~ citympg, data = car_price_data)
```




```
plot(price ~ highwaympg, data = car_price_data)
```




```
plot(price ~ peakrpm, data = car_price_data)
```



#Train and Test Data

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.2.3
```

```
##
## Attaching package: 'dplyr'
```

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

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

```
exclude_vars <- c('car_ID','symboling','CarName','fueltype','aspiration','doornumber','carbody','drivew

car_price_data_new <- select(car_price_data, -one_of(exclude_vars))
```

```
train_data <- car_price_data_new[1:165, ]
test_data <- car_price_data_new[166:205, ]
```

#Linear Regression Model

```
linear_model <- lm(price ~ . - car_ID, data = car_price_data)
summary(linear_model)
```

```
##
## Call:
## lm(formula = price ~ . - car_ID, data = car_price_data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -2253      0      0      0   1846
##
## Coefficients: (12 not defined because of singularities)
##                                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)                         -12860.335  29449.024  -0.437  0.66640
## symboling-2                            -209.486   2051.456  -0.102  0.91955
## symboling0                            -1379.028   1495.713  -0.922  0.36611
## symboling1                             -891.809   1638.222  -0.544  0.59142
## symboling2                              472.767   1535.890   0.308  0.76099
## symboling3                             1467.196   1705.228   0.860  0.39844
## CarNamealfa-romero Quadrifoglio        6005.932   4880.173   1.231  0.23088
## CarNamealfa-romero stelvio             3005.000   1744.317   1.723  0.09836
## CarNameaudi 100 ls                     4462.149   3982.918   1.120  0.27414
## CarNameaudi 100ls                      5266.406   4778.496   1.102  0.28182
## CarNameaudi 4000                       9107.930   5803.760   1.569  0.13023
## CarNameaudi 5000                       7539.916   4917.495   1.533  0.13885
## CarNameaudi 5000s (diesel)             5583.834   6030.926   0.926  0.36413
## CarNameaudi fox                        8892.636   4968.007   1.790  0.08663
## CarNamebmw 320i                        4718.626   3491.879   1.351  0.18974
## CarNamebmw x1                         18756.766   5708.379   3.286  0.00324
## CarNamebmw x3                         17314.422   5087.862   3.403  0.00244
## CarNamebmw x4                         18795.582   5703.008   3.296  0.00316
## CarNamebmw x5                         28350.558   5708.695   4.966 5.07e-05
## CarNamebmw z4                         16726.772   5291.918   3.161  0.00437
## CarNamebuick century                  11955.777   5731.039   2.086  0.04825
## CarNamebuick century luxus (sw)        9334.169   5920.713   1.577  0.12856
## CarNamebuick century special           9401.833   6697.633   1.404  0.17375
## CarNamebuick electra 225 custom        5686.443   5312.796   1.070  0.29558
## CarNamebuick opel isuzu deluxe         3547.133   4754.222   0.746  0.46316
## CarNamebuick regal sport coupe (turbo) 18415.571   6004.602   3.067  0.00546
## CarNamebuick skyhawk                   7936.399   5591.516   1.419  0.16920
## CarNamebuick skylark                   7413.701   3852.883   1.924  0.06679
## CarNamechevrolet impala                5664.048   4552.430   1.244  0.22596
## CarNamechevrolet monte carlo           -819.544   3846.907  -0.213  0.83317
## CarNamechevrolet vega 2300            -2484.122   3811.413  -0.652  0.52102
## CarNamedodge challenger se             -702.325   3464.595  -0.203  0.84114
## CarNamedodge colt (sw)                -3482.024   3453.821  -1.008  0.32386
## CarNamedodge colt hardtop             -4399.024   3453.821  -1.274  0.21550
## CarNamedodge coronet custom           -3681.786   3626.974  -1.015  0.32061
## CarNamedodge coronet custom (sw)       -957.774   5209.753  -0.184  0.85575
## CarNamedodge d200                     -2632.687   3628.533  -0.726  0.47543
## CarNamedodge dart custom                324.408   4148.591   0.078  0.93835
## CarNamedodge monaco (sw)              -3797.379   3568.662  -1.064  0.29833
```

```
## CarNamedodge rampage                -2756.815   3794.218  -0.727  0.47481
## CarNamehonda accord                  3525.986   5824.533   0.605  0.55086
## CarNamehonda accord cvcc               97.633   4640.488   0.021  0.98340
## CarNamehonda accord lx                745.195   5243.580   0.142  0.88823
## CarNamehonda civic                   2196.364   3877.941   0.566  0.57663
## CarNamehonda civic (auto)            -329.260   3676.873  -0.090  0.92942
## CarNamehonda civic 1300              2830.264   5601.930   0.505  0.61820
## CarNamehonda civic 1500 gl           4036.994   5423.387   0.744  0.46419
## CarNamehonda civic cvcc               706.167   4510.595   0.157  0.87696
## CarNamehonda prelude                 1818.993   5682.073   0.320  0.75176
## CarNameisuzu D-Max                    989.913   3715.958   0.266  0.79231
## CarNameisuzu D-Max V-Cross           -712.240   3854.534  -0.185  0.85502
## CarNameisuzu MU-X                    -920.331   3577.597  -0.257  0.79927
## CarNamejaguar xf                     5339.507   5300.709   1.007  0.32426
## CarNamejaguar xj                     2039.507   5300.709   0.385  0.70395
## CarNamejaguar xk                    11948.442   7060.743   1.692  0.10411
## CarNamemaxda glc deluxe              2063.490   3482.282   0.593  0.55925
## CarNamemaxda rx3                      885.783   3507.078   0.253  0.80284
## CarNamemazda 626                     1493.118   3455.911   0.432  0.66973
## CarNamemazda glc                     3247.571   3421.232   0.949  0.35237
## CarNamemazda glc 4                   3877.801   4545.551   0.853  0.40240
## CarNamemazda glc custom              2258.390   3656.748   0.618  0.54291
## CarNamemazda glc custom l           -1208.713   3657.781  -0.330  0.74405
## CarNamemazda glc deluxe              1971.294   3363.734   0.586  0.56355
## CarNamemazda rx-4                     800.488   3359.650   0.238  0.81379
## CarNamemazda rx-7 gs                 4171.490   3449.082   1.209  0.23878
## CarNamemazda rx2 coupe               2706.080   3483.042   0.777  0.44512
## CarNamemercury cougar                3416.111   4667.424   0.732  0.47162
## CarNamemitsubishi g4                -3176.366   3237.367  -0.981  0.33672
## CarNamemitsubishi lancer            -3661.663   3548.229  -1.032  0.31282
## CarNamemitsubishi mirage            -5431.248   3860.316  -1.407  0.17281
## CarNamemitsubishi mirage g4         -4174.039   3556.104  -1.174  0.25250
## CarNamemitsubishi montero           -4701.339   3227.382  -1.457  0.15872
## CarNamemitsubishi outlander         -4363.991   3333.529  -1.309  0.20342
## CarNamemitsubishi pajero            -3960.615   3242.152  -1.222  0.23423
## CarNamenissan clipper                1319.206   3828.500   0.345  0.73354
## CarNamenissan dayz                    347.080   4745.918   0.073  0.94233
## CarNamenissan fuga                   2470.626   5025.371   0.492  0.62764
## CarNamenissan gt-r                    725.364   3775.436   0.192  0.84933
## CarNamenissan juke                   3227.186   3627.355   0.890  0.38285
## CarNamenissan kicks                   419.762   4748.881   0.088  0.93033
## CarNamenissan latio                  1475.814   3338.052   0.442  0.66253
## CarNamenissan leaf                   3428.506   3611.582   0.949  0.35233
## CarNamenissan note                   4130.833   3322.485   1.243  0.22628
## CarNamenissan nv200                   463.956   3682.505   0.126  0.90084
## CarNamenissan otti                   1081.668   4824.702   0.224  0.82459
## CarNamenissan rogue                  1790.403   3568.887   0.502  0.62067
## CarNamenissan teana                  -714.020   4263.803  -0.167  0.86847
## CarNamenissan titan                  3630.097   3320.569   1.093  0.28562
## CarNameNissan versa                  2340.383   3619.778   0.647  0.52432
## CarNamepeugeot 304                 -11796.957   4040.895  -2.919  0.00772
## CarNamepeugeot 504                  -8803.870   3988.505  -2.207  0.03754
## CarNamepeugeot 504 (sw)             -9075.305   4567.950  -1.987  0.05899
## CarNamepeugeot 505s turbo diesel    -7346.024   4288.132  -1.713  0.10014
```

```
## CarNamepeugeot 604sl                    -9163.085   3865.834  -2.370  0.02654
## CarNameplymouth cricket                 -2350.530   3650.630  -0.644  0.52603
## CarNameplymouth duster                  -5813.268   3466.188  -1.677  0.10706
## CarNameplymouth fury gran sedan         -4159.825   3512.083  -1.184  0.24834
## CarNameplymouth fury iii                -3527.529   3548.059  -0.994  0.33046
## CarNameplymouth satellite custom (sw)   -2320.127   3401.086  -0.682  0.50194
## CarNameplymouth valiant                   360.742   4147.959   0.087  0.93145
## CarNameporcshce panamera                16573.089   7422.964   2.233  0.03559
## CarNameporsche boxter                   22210.651   7864.899   2.824  0.00962
## CarNameporsche cayenne                  18073.089   7215.106   2.505  0.01978
## CarNameporsche macan                     6374.833   3592.333   1.775  0.08921
## CarNamerenault 12tl                      2683.687   3891.019   0.690  0.49728
## CarNamerenault 5 gtl                     -695.182   3713.211  -0.187  0.85313
## CarNamesaab 99e                          4080.720   4243.609   0.962  0.34624
## CarNamesaab 99gle                        5316.211   4268.016   1.246  0.22545
## CarNamesaab 99le                         2294.588   4386.283   0.523  0.60589
## CarNamesubaru                           -4490.083   3106.365  -1.445  0.16182
## CarNamesubaru baja                      -3059.328   3074.401  -0.995  0.33004
## CarNamesubaru brz                       -1535.164   3261.427  -0.471  0.64229
## CarNamesubaru dl                        -1890.164   3086.947  -0.612  0.54634
## CarNamesubaru r1                        -3687.372   3870.396  -0.953  0.35064
## CarNamesubaru r2                        -3765.075   3825.395  -0.984  0.33524
## CarNamesubaru trezia                    -2502.748   2804.204  -0.892  0.38137
## CarNamesubaru tribeca                    -941.168   2588.125  -0.364  0.71944
## CarNametoyota carina                    -3849.373   5603.213  -0.687  0.49896
## CarNametoyota celica gt                 -1424.868   4029.004  -0.354  0.72682
## CarNametoyota celica gt liftback        -5666.889   3195.947  -1.773  0.08945
## CarNametoyota corolla                     618.705   3192.492   0.194  0.84803
## CarNametoyota corolla 1200               -188.511   3355.559  -0.056  0.95568
## CarNametoyota corolla 1600 (sw)          4685.779   4313.627   1.086  0.28861
## CarNametoyota corolla liftback            759.940   3150.255   0.241  0.81151
## CarNametoyota corolla tercel            -5016.712   3368.093  -1.489  0.14995
## CarNametoyota corona                      294.048   3150.920   0.093  0.92646
## CarNametoyota corona hardtop             3561.232   4415.006   0.807  0.42815
## CarNametoyota corona liftback            -947.211   3586.608  -0.264  0.79406
## CarNametoyota corona mark ii             -541.646   3670.351  -0.148  0.88397
## CarNametoyota cressida                   5494.417   3679.571   1.493  0.14897
## CarNametoyota mark ii                     736.074   3199.345   0.230  0.82007
## CarNametoyota starlet                     -12.212   3421.508  -0.004  0.99718
## CarNametoyota tercel                      206.806   3695.834   0.056  0.95586
## CarNametoyouta tercel                    2465.747   3708.787   0.665  0.51276
## CarNamevokswagen rabbit                   655.692   4272.437   0.153  0.87937
## CarNamevolkswagen 1131 deluxe sedan       590.368   3958.267   0.149  0.88274
## CarNamevolkswagen 411 (sw)              -1434.943   3718.207  -0.386  0.70310
## CarNamevolkswagen dasher                 2091.490   3731.121   0.561  0.58052
## CarNamevolkswagen model 111              -946.259   4057.297  -0.233  0.81765
## CarNamevolkswagen rabbit                 6142.739   5008.301   1.227  0.23241
## CarNamevolkswagen rabbit custom          3419.239   4754.357   0.719  0.47927
## CarNamevolkswagen super beetle            446.212   4548.755   0.098  0.92271
## CarNamevolkswagen type 3                -1011.583   3707.907  -0.273  0.78743
## CarNamevolvo 144ea                       -382.046   4251.393  -0.090  0.92917
## CarNamevolvo 145e (sw)                  -2643.897   3944.066  -0.670  0.50931
## CarNamevolvo 244dl                        315.752   4302.097   0.073  0.94213
## CarNamevolvo 245                         2211.550   4178.137   0.529  0.60166
```

```
## CarNamevolvo 246                5401.122   5973.324   0.904  0.37526
## CarNamevolvo 264gl              1456.677   4309.837   0.338  0.73844
## CarNamevolvo diesel             3848.492   5049.247   0.762  0.45369
## CarNamevw dasher                3392.062   3466.611   0.978  0.33801
## CarNamevw rabbit                -399.629   3174.469  -0.126  0.90091
## fueltypegas                   -26469.533  12164.695  -2.176  0.04009
## aspirationturbo                 1285.811   1224.274   1.050  0.30451
## doornumbertwo                  -1787.505    973.837  -1.836  0.07939
## carbodyhardtop                  1642.765   2055.264   0.799  0.43230
## carbodyhatchback                1576.260   1345.750   1.171  0.25348
## carbodysedan                    2388.303   1155.887   2.066  0.05025
## carbodywagon                          NA         NA      NA       NA
## drivewheelfwd                   -307.330   1453.967  -0.211  0.83446
## drivewheelrwd                   1306.972   1854.734   0.705  0.48809
## enginelocationrear              8188.971   4637.429   1.766  0.09070
## wheelbase                        330.009    177.635   1.858  0.07605
## carlength                       -196.420     98.519  -1.994  0.05817
## carwidth                         861.090    372.882   2.309  0.03025
## carheight                       -573.485    358.399  -1.600  0.12322
## curbweight                        11.482      3.972   2.891  0.00824
## enginetypedohcv                       NA         NA      NA       NA
## enginetypel                           NA         NA      NA       NA
## enginetypeohc                  -4116.337   2361.615  -1.743  0.09469
## enginetypeohcf                        NA         NA      NA       NA
## enginetypeohcv                   935.449   3173.140   0.295  0.77079
## enginetyperotor                 3335.906   5060.643   0.659  0.51632
## cylindernumberfive                    NA         NA      NA       NA
## cylindernumberfour              6855.803   2530.249   2.710  0.01251
## cylindernumbersix                     NA         NA      NA       NA
## cylindernumberthree                   NA         NA      NA       NA
## cylindernumbertwelve                  NA         NA      NA       NA
## cylindernumbertwo                     NA         NA      NA       NA
## enginesize                        54.669     61.148   0.894  0.38055
## fuelsystem2bbl                  3937.890   3490.949   1.128  0.27094
## fuelsystem4bbl                        NA         NA      NA       NA
## fuelsystemidi                         NA         NA      NA       NA
## fuelsystemmfi                         NA         NA      NA       NA
## fuelsystemmpfi                  3159.983   3148.472   1.004  0.32599
## fuelsystemspdi                  2877.826   4486.433   0.641  0.52757
## fuelsystemspfi                  -515.260   5112.075  -0.101  0.92059
## boreratio                      -3633.396   2155.272  -1.686  0.10535
## stroke                         -1166.482   1236.333  -0.944  0.35523
## compressionratio               -1862.412    935.039  -1.992  0.05840
## horsepower                        -65.136     44.535  -1.463  0.15711
## peakrpm                             3.797      1.334   2.846  0.00915
## citympg                           257.805    270.218   0.954  0.34997
## highwaympg                        -92.904    194.846  -0.477  0.63800
##
## (Intercept)
## symboling-2
## symboling0
## symboling1
## symboling2
## symboling3
```

```
## CarNamealfa-romero Quadrifoglio
## CarNamealfa-romero stelvio             .
## CarNameaudi 100 ls
## CarNameaudi 100ls
## CarNameaudi 4000
## CarNameaudi 5000
## CarNameaudi 5000s (diesel)
## CarNameaudi fox                        .
## CarNamebmw 320i
## CarNamebmw x1                          **
## CarNamebmw x3                          **
## CarNamebmw x4                          **
## CarNamebmw x5                          ***
## CarNamebmw z4                          **
## CarNamebuick century                   *
## CarNamebuick century luxus (sw)
## CarNamebuick century special
## CarNamebuick electra 225 custom
## CarNamebuick opel isuzu deluxe
## CarNamebuick regal sport coupe (turbo) **
## CarNamebuick skyhawk
## CarNamebuick skylark                   .
## CarNamechevrolet impala
## CarNamechevrolet monte carlo
## CarNamechevrolet vega 2300
## CarNamedodge challenger se
## CarNamedodge colt (sw)
## CarNamedodge colt hardtop
## CarNamedodge coronet custom
## CarNamedodge coronet custom (sw)
## CarNamedodge d200
## CarNamedodge dart custom
## CarNamedodge monaco (sw)
## CarNamedodge rampage
## CarNamehonda accord
## CarNamehonda accord cvcc
## CarNamehonda accord lx
## CarNamehonda civic
## CarNamehonda civic (auto)
## CarNamehonda civic 1300
## CarNamehonda civic 1500 gl
## CarNamehonda civic cvcc
## CarNamehonda prelude
## CarNameisuzu D-Max
## CarNameisuzu D-Max V-Cross
## CarNameisuzu MU-X
## CarNamejaguar xf
## CarNamejaguar xj
## CarNamejaguar xk
## CarNamemaxda glc deluxe
## CarNamemaxda rx3
## CarNamemazda 626
## CarNamemazda glc
## CarNamemazda glc 4
```

```
## CarNamemazda glc custom
## CarNamemazda glc custom l
## CarNamemazda glc deluxe
## CarNamemazda rx-4
## CarNamemazda rx-7 gs
## CarNamemazda rx2 coupe
## CarNamemercury cougar
## CarNamemitsubishi g4
## CarNamemitsubishi lancer
## CarNamemitsubishi mirage
## CarNamemitsubishi mirage g4
## CarNamemitsubishi montero
## CarNamemitsubishi outlander
## CarNamemitsubishi pajero
## CarNamenissan clipper
## CarNamenissan dayz
## CarNamenissan fuga
## CarNamenissan gt-r
## CarNamenissan juke
## CarNamenissan kicks
## CarNamenissan latio
## CarNamenissan leaf
## CarNamenissan note
## CarNamenissan nv200
## CarNamenissan otti
## CarNamenissan rogue
## CarNamenissan teana
## CarNamenissan titan
## CarNameNissan versa
## CarNamepeugeot 304                     **
## CarNamepeugeot 504                     *
## CarNamepeugeot 504 (sw)                .
## CarNamepeugeot 505s turbo diesel
## CarNamepeugeot 604sl                   *
## CarNameplymouth cricket
## CarNameplymouth duster
## CarNameplymouth fury gran sedan
## CarNameplymouth fury iii
## CarNameplymouth satellite custom (sw)
## CarNameplymouth valiant
## CarNameporcshce panamera               *
## CarNameporsche boxter                  **
## CarNameporsche cayenne                 *
## CarNameporsche macan                   .
## CarNamerenault 12tl
## CarNamerenault 5 gtl
## CarNamesaab 99e
## CarNamesaab 99gle
## CarNamesaab 99le
## CarNamesubaru
## CarNamesubaru baja
## CarNamesubaru brz
## CarNamesubaru dl
## CarNamesubaru r1
```

```
## CarNamesubaru r2
## CarNamesubaru trezia
## CarNamesubaru tribeca
## CarNametoyota carina
## CarNametoyota celica gt
## CarNametoyota celica gt liftback       .
## CarNametoyota corolla
## CarNametoyota corolla 1200
## CarNametoyota corolla 1600 (sw)
## CarNametoyota corolla liftback
## CarNametoyota corolla tercel
## CarNametoyota corona
## CarNametoyota corona hardtop
## CarNametoyota corona liftback
## CarNametoyota corona mark ii
## CarNametoyota cressida
## CarNametoyota mark ii
## CarNametoyota starlet
## CarNametoyota tercel
## CarNametoyouta tercel
## CarNamevokswagen rabbit
## CarNamevolkswagen 1131 deluxe sedan
## CarNamevolkswagen 411 (sw)
## CarNamevolkswagen dasher
## CarNamevolkswagen model 111
## CarNamevolkswagen rabbit
## CarNamevolkswagen rabbit custom
## CarNamevolkswagen super beetle
## CarNamevolkswagen type 3
## CarNamevolvo 144ea
## CarNamevolvo 145e (sw)
## CarNamevolvo 244dl
## CarNamevolvo 245
## CarNamevolvo 246
## CarNamevolvo 264gl
## CarNamevolvo diesel
## CarNamevw dasher
## CarNamevw rabbit
## fueltypegas                            *
## aspirationturbo
## doornumbertwo                          .
## carbodyhardtop
## carbodyhatchback
## carbodysedan                           .
## carbodywagon
## drivewheelfwd
## drivewheelrwd
## enginelocationrear                     .
## wheelbase                              .
## carlength                              .
## carwidth                               *
## carheight
## curbweight                             **
## enginetypedohcv
```

```
## enginetypel
## enginetypeohc                          .
## enginetypeohcf
## enginetypeohcv
## enginetyperotor
## cylindernumberfive
## cylindernumberfour                     *
## cylindernumbersix
## cylindernumberthree
## cylindernumbertwelve
## cylindernumbertwo
## enginesize
## fuelsystem2bbl
## fuelsystem4bbl
## fuelsystemidi
## fuelsystemmfi
## fuelsystemmpfi
## fuelsystemspdi
## fuelsystemspfi
## boreratio
## stroke
## compressionratio                       .
## horsepower
## peakrpm                                **
## citympg
## highwaympg
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1233 on 23 degrees of freedom
## Multiple R-squared:  0.9973, Adjusted R-squared:  0.9762
## F-statistic: 47.16 on 181 and 23 DF,  p-value: 9.666e-16
```

We first started with the linear model with all variables given. We got R-squared value of 0.9762, which indicates that 97.62 of the variability in the car price is explained by all variables. Notice that p-value is 9.666e-16, which is very close to 0. This p-value tells us that this model is statistically significant.

```
#start with linear model containing all continuous variables
linear_model2 <- lm(price ~ ., data = train_data)
predicted <- predict(linear_model2, newdata = test_data)
rmse <- sqrt(mean((test_data$price - predicted)^2))
rmse
```

```
## [1] 3224.741
```

```
summary(linear_model2)
```

```
##
## Call:
## lm(formula = price ~ ., data = train_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -12049.8  -1814.8   -172.7   1408.0  12791.4
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -48703.162  17230.845  -2.827 0.005343 **
## wheelbase           153.510    110.705   1.387 0.167588
## carlength           -35.536     60.793  -0.585 0.559727
## carwidth            255.082    268.948   0.948 0.344418
## carheight           114.007    149.557   0.762 0.447069
## curbweight            1.136      1.970   0.577 0.564857
## enginesize          123.411     15.679   7.871 6.32e-13 ***
## boreratio          -324.821   1468.923  -0.221 0.825291
## stroke            -2975.049    843.628  -3.526 0.000558 ***
## compressionratio    310.901    102.361   3.037 0.002813 **
## horsepower           37.003     19.855   1.864 0.064304 .
## peakrpm               3.233      0.911   3.549 0.000517 ***
## citympg            -216.214    203.923  -1.060 0.290713
## highwaympg          145.037    183.443   0.791 0.430395
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3253 on 151 degrees of freedom
## Multiple R-squared:  0.8701, Adjusted R-squared:  0.8589
## F-statistic: 77.79 on 13 and 151 DF,  p-value: < 2.2e-16
```

We moved onto the second model, where we decided to use only numerical variables to calculate RMSE. We got the adjusted R-squared of 0.8589, which indicates that 85.89% of the variability in car price is explained by all numerical variables. The R-squared value has decreased from the first model, and the p-value has also decreased. Using train and testing data, we got RMSE value of 3224.741.

```
linear_model3 <- lm(price ~ stroke + compressionratio + peakrpm + enginesize + horsepower, data = train
predicted <- predict(linear_model3, newdata = test_data)
rmse <- sqrt(mean((test_data$price - predicted)^2))
rmse
```

```
## [1] 3353.8
```

```
summary(linear_model3)
```

```
##
## Call:
## lm(formula = price ~ stroke + compressionratio + peakrpm + enginesize +
##     horsepower, data = train_data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -15377  -1636   -355   1525  12585
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -1.806e+04  4.897e+03  -3.688 0.000310 ***
## stroke           -2.715e+03  8.556e+02  -3.173 0.001811 **
## compressionratio  3.641e+02  8.633e+01   4.218 4.13e-05 ***
```

```
## peakrpm           2.609e+00  8.338e-01   3.129 0.002086 **
## enginesize        1.419e+02  1.410e+01  10.061  < 2e-16 ***
## horsepower        5.132e+01  1.446e+01   3.550 0.000506 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3420 on 159 degrees of freedom
## Multiple R-squared:  0.8488, Adjusted R-squared:  0.844
## F-statistic: 178.5 on 5 and 159 DF,  p-value: < 2.2e-16
```

For the last model, we selected five most significant variables: stroke, compressionratio, peakrpm, enginesize, and horsepower. These variables had the least p-values from the output of the second model. We got the adjusted R-squared of 0.8488, which indicates that 84.88% of the variability in car price is explained by all numerical variables. The R-squared value has increased from the first model, while the p-value has stayed the same. Using train and testing data, we got RMSE value of 3353.8, which is greater than the RMSE value from model 2.

#Stepwise Regression

Note: we're using train_data here; direction='both' indicates bidirectional; criterion changes for each model to test model using different criterion.

```
initial_model <- lm(price ~ ., data = train_data)
#stepwise using aic
stepwise_model_aic <- step(initial_model, direction = "both", trace = 1, k = log(nrow(train_data)), cri
```

```
## Start:  AIC=2725.72
## price ~ wheelbase + carlength + carwidth + carheight + curbweight +
##     enginesize + boreratio + stroke + compressionratio + horsepower +
##     peakrpm + citympg + highwaympg
##
##                    Df Sum of Sq        RSS    AIC
## - boreratio         1    517577 1598830724 2720.7
## - curbweight        1   3522954 1601836101 2721.0
## - carlength         1   3616746 1601929893 2721.0
## - carheight         1   6150913 1604464060 2721.2
## - highwaympg        1   6616682 1604929829 2721.3
## - carwidth          1   9521592 1607834739 2721.6
## - citympg           1  11899303 1610212449 2721.8
## - wheelbase         1  20352969 1618666116 2722.7
## - horsepower        1  36765715 1635078862 2724.4
## <none>                          1598313147 2725.7
## - compressionratio  1  97647357 1695960504 2730.4
## - stroke            1 131635022 1729948169 2733.7
## - peakrpm           1 133287354 1731600501 2733.8
## - enginesize        1 655765091 2254078238 2777.3
##
## Step:  AIC=2720.67
## price ~ wheelbase + carlength + carwidth + carheight + curbweight +
##     enginesize + stroke + compressionratio + horsepower + peakrpm +
##     citympg + highwaympg
##
##                    Df Sum of Sq        RSS    AIC
## - curbweight        1   3875007 1602705731 2716.0
```

```
## - carlength         1   3945725 1602776449 2716.0
## - carheight         1   6345186 1605175910 2716.2
## - highwaympg        1   6667111 1605497835 2716.2
## - carwidth          1   9596094 1608426818 2716.6
## - citympg           1  11551676 1610382400 2716.8
## - wheelbase         1  20053328 1618884052 2717.6
## - horsepower        1  37245742 1636076466 2719.4
## <none>                           1598830724 2720.7
## - compressionratio  1  98632766 1697463490 2725.4
## + boreratio         1    517577 1598313147 2725.7
## - stroke            1 132455600 1731286325 2728.7
## - peakrpm           1 161775729 1760606454 2731.5
## - enginesize        1 661240480 2260071205 2772.7
##
## Step:  AIC=2715.96
## price ~ wheelbase + carlength + carwidth + carheight + enginesize +
##     stroke + compressionratio + horsepower + peakrpm + citympg +
##     highwaympg
##
##                    Df Sum of Sq        RSS    AIC
## - carlength         1   2550311 1605256043 2711.1
## - highwaympg        1   4686885 1607392617 2711.3
## - carheight         1   6839556 1609545287 2711.6
## - citympg           1  11316774 1614022506 2712.0
## - carwidth          1  12329133 1615034864 2712.1
## - wheelbase         1  25203483 1627909214 2713.4
## <none>                           1602705731 2716.0
## - horsepower        1  50569156 1653274887 2716.0
## + curbweight        1   3875007 1598830724 2720.7
## + boreratio         1    869630 1601836101 2721.0
## - stroke            1 129465717 1732171449 2723.7
## - compressionratio  1 131988862 1734694594 2723.9
## - peakrpm           1 160412292 1763118024 2726.6
## - enginesize        1 736190692 2338896424 2773.2
##
## Step:  AIC=2711.12
## price ~ wheelbase + carwidth + carheight + enginesize + stroke +
##     compressionratio + horsepower + peakrpm + citympg + highwaympg
##
##                    Df Sum of Sq        RSS    AIC
## - highwaympg        1   3611196 1608867239 2706.4
## - carheight         1   5153734 1610409777 2706.5
## - citympg           1   8938210 1614194253 2706.9
## - carwidth          1  10250349 1615506392 2707.1
## - wheelbase         1  23243124 1628499166 2708.4
## <none>                           1605256043 2711.1
## - horsepower        1  53130281 1658386324 2711.4
## + carlength         1   2550311 1602705731 2716.0
## + curbweight        1   2479593 1602776449 2716.0
## + boreratio         1   1119585 1604136458 2716.1
## - stroke            1 128248289 1733504332 2718.7
## - compressionratio  1 129515767 1734771810 2718.8
## - peakrpm           1 167032539 1772288582 2722.3
## - enginesize        1 739801500 2345057543 2768.6
```

```
##
## Step:  AIC=2706.38
## price ~ wheelbase + carwidth + carheight + enginesize + stroke +
##     compressionratio + horsepower + peakrpm + citympg
##
##                    Df Sum of Sq        RSS    AIC
## - carheight         1   5113543 1613980782 2701.8
## - carwidth          1   9722681 1618589920 2702.3
## - citympg           1  10119903 1618987142 2702.3
## - wheelbase         1  21368834 1630236073 2703.4
## <none>                          1608867239 2706.4
## - horsepower        1  61523298 1670390537 2707.5
## + highwaympg        1   3611196 1605256043 2711.1
## + carlength         1   1474622 1607392617 2711.3
## + curbweight        1   1321360 1607545879 2711.3
## + boreratio         1   1001757 1607865482 2711.4
## - stroke            1 125499785 1734367024 2713.7
## - compressionratio  1 128328284 1737195523 2713.9
## - peakrpm           1 166829542 1775696781 2717.6
## - enginesize        1 750296191 2359163430 2764.4
##
## Step:  AIC=2701.8
## price ~ wheelbase + carwidth + enginesize + stroke + compressionratio +
##     horsepower + peakrpm + citympg
##
##                    Df Sum of Sq        RSS    AIC
## - carwidth          1   7462387 1621443169 2697.4
## - citympg           1  10268704 1624249486 2697.7
## <none>                          1613980782 2701.8
## - wheelbase         1  51043846 1665024628 2701.8
## - horsepower        1  61057342 1675038124 2702.8
## + carheight         1   5113543 1608867239 2706.4
## + highwaympg        1   3571006 1610409777 2706.5
## + curbweight        1   1915354 1612065428 2706.7
## + boreratio         1   1161957 1612818825 2706.8
## + carlength         1    332853 1613647930 2706.9
## - compressionratio  1 130433264 1744414047 2709.5
## - stroke            1 136602131 1750582914 2710.1
## - peakrpm           1 162179675 1776160458 2712.5
## - enginesize        1 746050343 2360031125 2759.4
##
## Step:  AIC=2697.45
## price ~ wheelbase + enginesize + stroke + compressionratio +
##     horsepower + peakrpm + citympg
##
##                    Df Sum of Sq        RSS    AIC
## - citympg           1  13442845 1634886014 2693.7
## <none>                          1621443169 2697.4
## - horsepower        1  77310491 1698753660 2700.0
## + carwidth          1   7462387 1613980782 2701.8
## + curbweight        1   3870233 1617572936 2702.2
## + highwaympg        1   3108736 1618334433 2702.2
## + carheight         1   2853248 1618589920 2702.3
## + boreratio         1   1224909 1620218260 2702.4
```

```
## + carlength         1       132 1621443036 2702.6
## - wheelbase         1 123048423 1744491592 2704.4
## - stroke            1 134034050 1755477219 2705.4
## - compressionratio  1 150306405 1771749574 2707.0
## - peakrpm           1 164255863 1785699032 2708.3
## - enginesize        1 761051172 2382494341 2755.8
##
## Step:  AIC=2693.71
## price ~ wheelbase + enginesize + stroke + compressionratio +
##     horsepower + peakrpm
##
##                    Df Sum of Sq        RSS    AIC
## <none>                          1634886014 2693.7
## + curbweight        1  13539421 1621346593 2697.4
## + citympg           1  13442845 1621443169 2697.4
## + carwidth          1  10636527 1624249486 2697.7
## + highwaympg        1   7315398 1627570616 2698.1
## + carlength         1   3016846 1631869168 2698.5
## + carheight         1   2611136 1632274877 2698.6
## + boreratio         1     21746 1634864267 2698.8
## - stroke            1 131940428 1766826441 2701.4
## - compressionratio  1 138351294 1773237307 2702.0
## - peakrpm           1 158520736 1793406750 2703.9
## - horsepower        1 182074941 1816960955 2706.0
## - wheelbase         1 225284392 1860170405 2709.9
## - enginesize        1 749887317 2384773330 2750.9
```

```
#stepwise using bic
stepwise_model_bic <- step(initial_model, direction = "both", trace = 1, k = log(nrow(train_data)), cri
```

```
## Start:  AIC=2725.72
## price ~ wheelbase + carlength + carwidth + carheight + curbweight +
##     enginesize + boreratio + stroke + compressionratio + horsepower +
##     peakrpm + citympg + highwaympg
##
##                    Df Sum of Sq        RSS    AIC
## - boreratio         1    517577 1598830724 2720.7
## - curbweight        1   3522954 1601836101 2721.0
## - carlength         1   3616746 1601929893 2721.0
## - carheight         1   6150913 1604464060 2721.2
## - highwaympg        1   6616682 1604929829 2721.3
## - carwidth          1   9521592 1607834739 2721.6
## - citympg           1  11899303 1610212449 2721.8
## - wheelbase         1  20352969 1618666116 2722.7
## - horsepower        1  36765715 1635078862 2724.4
## <none>                          1598313147 2725.7
## - compressionratio  1  97647357 1695960504 2730.4
## - stroke            1 131635022 1729948169 2733.7
## - peakrpm           1 133287354 1731600501 2733.8
## - enginesize        1 655765091 2254078238 2777.3
##
## Step:  AIC=2720.67
## price ~ wheelbase + carlength + carwidth + carheight + curbweight +
##     enginesize + stroke + compressionratio + horsepower + peakrpm +
```

```
##     citympg + highwaympg
##
##                     Df Sum of Sq        RSS    AIC
## - curbweight         1   3875007 1602705731 2716.0
## - carlength          1   3945725 1602776449 2716.0
## - carheight          1   6345186 1605175910 2716.2
## - highwaympg         1   6667111 1605497835 2716.2
## - carwidth           1   9596094 1608426818 2716.6
## - citympg            1  11551676 1610382400 2716.8
## - wheelbase          1  20053328 1618884052 2717.6
## - horsepower         1  37245742 1636076466 2719.4
## <none>                           1598830724 2720.7
## - compressionratio   1  98632766 1697463490 2725.4
## + boreratio          1    517577 1598313147 2725.7
## - stroke             1 132455600 1731286325 2728.7
## - peakrpm            1 161775729 1760606454 2731.5
## - enginesize         1 661240480 2260071205 2772.7
##
## Step:  AIC=2715.96
## price ~ wheelbase + carlength + carwidth + carheight + enginesize +
##     stroke + compressionratio + horsepower + peakrpm + citympg +
##     highwaympg
##
##                     Df Sum of Sq        RSS    AIC
## - carlength          1   2550311 1605256043 2711.1
## - highwaympg         1   4686885 1607392617 2711.3
## - carheight          1   6839556 1609545287 2711.6
## - citympg            1  11316774 1614022506 2712.0
## - carwidth           1  12329133 1615034864 2712.1
## - wheelbase          1  25203483 1627909214 2713.4
## <none>                           1602705731 2716.0
## - horsepower         1  50569156 1653274887 2716.0
## + curbweight         1   3875007 1598830724 2720.7
## + boreratio          1    869630 1601836101 2721.0
## - stroke             1 129465717 1732171449 2723.7
## - compressionratio   1 131988862 1734694594 2723.9
## - peakrpm            1 160412292 1763118024 2726.6
## - enginesize         1 736190692 2338896424 2773.2
##
## Step:  AIC=2711.12
## price ~ wheelbase + carwidth + carheight + enginesize + stroke +
##     compressionratio + horsepower + peakrpm + citympg + highwaympg
##
##                     Df Sum of Sq        RSS    AIC
## - highwaympg         1   3611196 1608867239 2706.4
## - carheight          1   5153734 1610409777 2706.5
## - citympg            1   8938210 1614194253 2706.9
## - carwidth           1  10250349 1615506392 2707.1
## - wheelbase          1  23243124 1628499166 2708.4
## <none>                           1605256043 2711.1
## - horsepower         1  53130281 1658386324 2711.4
## + carlength          1   2550311 1602705731 2716.0
## + curbweight         1   2479593 1602776449 2716.0
## + boreratio          1   1119585 1604136458 2716.1
```

```
## - stroke            1 128248289 1733504332 2718.7
## - compressionratio  1 129515767 1734771810 2718.8
## - peakrpm           1 167032539 1772288582 2722.3
## - enginesize        1 739801500 2345057543 2768.6
##
## Step:  AIC=2706.38
## price ~ wheelbase + carwidth + carheight + enginesize + stroke +
##     compressionratio + horsepower + peakrpm + citympg
##
##                    Df Sum of Sq        RSS    AIC
## - carheight         1   5113543 1613980782 2701.8
## - carwidth          1   9722681 1618589920 2702.3
## - citympg           1  10119903 1618987142 2702.3
## - wheelbase         1  21368834 1630236073 2703.4
## <none>                          1608867239 2706.4
## - horsepower        1  61523298 1670390537 2707.5
## + highwaympg        1   3611196 1605256043 2711.1
## + carlength         1   1474622 1607392617 2711.3
## + curbweight        1   1321360 1607545879 2711.3
## + boreratio         1   1001757 1607865482 2711.4
## - stroke            1 125499785 1734367024 2713.7
## - compressionratio  1 128328284 1737195523 2713.9
## - peakrpm           1 166829542 1775696781 2717.6
## - enginesize        1 750296191 2359163430 2764.4
##
## Step:  AIC=2701.8
## price ~ wheelbase + carwidth + enginesize + stroke + compressionratio +
##     horsepower + peakrpm + citympg
##
##                    Df Sum of Sq        RSS    AIC
## - carwidth          1   7462387 1621443169 2697.4
## - citympg           1  10268704 1624249486 2697.7
## <none>                          1613980782 2701.8
## - wheelbase         1  51043846 1665024628 2701.8
## - horsepower        1  61057342 1675038124 2702.8
## + carheight         1   5113543 1608867239 2706.4
## + highwaympg        1   3571006 1610409777 2706.5
## + curbweight        1   1915354 1612065428 2706.7
## + boreratio         1   1161957 1612818825 2706.8
## + carlength         1    332853 1613647930 2706.9
## - compressionratio  1 130433264 1744414047 2709.5
## - stroke            1 136602131 1750582914 2710.1
## - peakrpm           1 162179675 1776160458 2712.5
## - enginesize        1 746050343 2360031125 2759.4
##
## Step:  AIC=2697.45
## price ~ wheelbase + enginesize + stroke + compressionratio +
##     horsepower + peakrpm + citympg
##
##                    Df Sum of Sq        RSS    AIC
## - citympg           1  13442845 1634886014 2693.7
## <none>                          1621443169 2697.4
## - horsepower        1  77310491 1698753660 2700.0
## + carwidth          1   7462387 1613980782 2701.8
```

```
## + curbweight        1   3870233 1617572936 2702.2
## + highwaympg        1   3108736 1618334433 2702.2
## + carheight         1   2853248 1618589920 2702.3
## + boreratio         1   1224909 1620218260 2702.4
## + carlength         1       132 1621443036 2702.6
## - wheelbase         1 123048423 1744491592 2704.4
## - stroke            1 134034050 1755477219 2705.4
## - compressionratio  1 150306405 1771749574 2707.0
## - peakrpm           1 164255863 1785699032 2708.3
## - enginesize        1 761051172 2382494341 2755.8
##
## Step:  AIC=2693.71
## price ~ wheelbase + enginesize + stroke + compressionratio +
##     horsepower + peakrpm
##
##                    Df Sum of Sq        RSS    AIC
## <none>                          1634886014 2693.7
## + curbweight        1  13539421 1621346593 2697.4
## + citympg           1  13442845 1621443169 2697.4
## + carwidth          1  10636527 1624249486 2697.7
## + highwaympg        1   7315398 1627570616 2698.1
## + carlength         1   3016846 1631869168 2698.5
## + carheight         1   2611136 1632274877 2698.6
## + boreratio         1     21746 1634864267 2698.8
## - stroke            1 131940428 1766826441 2701.4
## - compressionratio  1 138351294 1773237307 2702.0
## - peakrpm           1 158520736 1793406750 2703.9
## - horsepower        1 182074941 1816960955 2706.0
## - wheelbase         1 225284392 1860170405 2709.9
## - enginesize        1 749887317 2384773330 2750.9
```

```
#stepwise using adjr2
stepwise_model_adjr2 <- step(initial_model, direction = "both", trace = 1, k = log(nrow(train_data)), c
```

```
## Start:  AIC=2725.72
## price ~ wheelbase + carlength + carwidth + carheight + curbweight +
##     enginesize + boreratio + stroke + compressionratio + horsepower +
##     peakrpm + citympg + highwaympg
##
##                    Df Sum of Sq        RSS    AIC
## - boreratio         1    517577 1598830724 2720.7
## - curbweight        1   3522954 1601836101 2721.0
## - carlength         1   3616746 1601929893 2721.0
## - carheight         1   6150913 1604464060 2721.2
## - highwaympg        1   6616682 1604929829 2721.3
## - carwidth          1   9521592 1607834739 2721.6
## - citympg           1  11899303 1610212449 2721.8
## - wheelbase         1  20352969 1618666116 2722.7
## - horsepower        1  36765715 1635078862 2724.4
## <none>                          1598313147 2725.7
## - compressionratio  1  97647357 1695960504 2730.4
## - stroke            1 131635022 1729948169 2733.7
## - peakrpm           1 133287354 1731600501 2733.8
## - enginesize        1 655765091 2254078238 2777.3
```

```
##
## Step:  AIC=2720.67
## price ~ wheelbase + carlength + carwidth + carheight + curbweight +
##     enginesize + stroke + compressionratio + horsepower + peakrpm +
##     citympg + highwaympg
##
##                    Df Sum of Sq        RSS    AIC
## - curbweight        1   3875007 1602705731 2716.0
## - carlength         1   3945725 1602776449 2716.0
## - carheight         1   6345186 1605175910 2716.2
## - highwaympg        1   6667111 1605497835 2716.2
## - carwidth          1   9596094 1608426818 2716.6
## - citympg           1  11551676 1610382400 2716.8
## - wheelbase         1  20053328 1618884052 2717.6
## - horsepower        1  37245742 1636076466 2719.4
## <none>                           1598830724 2720.7
## - compressionratio  1  98632766 1697463490 2725.4
## + boreratio         1    517577 1598313147 2725.7
## - stroke            1 132455600 1731286325 2728.7
## - peakrpm           1 161775729 1760606454 2731.5
## - enginesize        1 661240480 2260071205 2772.7
##
## Step:  AIC=2715.96
## price ~ wheelbase + carlength + carwidth + carheight + enginesize +
##     stroke + compressionratio + horsepower + peakrpm + citympg +
##     highwaympg
##
##                    Df Sum of Sq        RSS    AIC
## - carlength         1   2550311 1605256043 2711.1
## - highwaympg        1   4686885 1607392617 2711.3
## - carheight         1   6839556 1609545287 2711.6
## - citympg           1  11316774 1614022506 2712.0
## - carwidth          1  12329133 1615034864 2712.1
## - wheelbase         1  25203483 1627909214 2713.4
## <none>                           1602705731 2716.0
## - horsepower        1  50569156 1653274887 2716.0
## + curbweight        1   3875007 1598830724 2720.7
## + boreratio         1    869630 1601836101 2721.0
## - stroke            1 129465717 1732171449 2723.7
## - compressionratio  1 131988862 1734694594 2723.9
## - peakrpm           1 160412292 1763118024 2726.6
## - enginesize        1 736190692 2338896424 2773.2
##
## Step:  AIC=2711.12
## price ~ wheelbase + carwidth + carheight + enginesize + stroke +
##     compressionratio + horsepower + peakrpm + citympg + highwaympg
##
##                    Df Sum of Sq        RSS    AIC
## - highwaympg        1   3611196 1608867239 2706.4
## - carheight         1   5153734 1610409777 2706.5
## - citympg           1   8938210 1614194253 2706.9
## - carwidth          1  10250349 1615506392 2707.1
## - wheelbase         1  23243124 1628499166 2708.4
## <none>                           1605256043 2711.1
```

```
## - horsepower        1  53130281 1658386324 2711.4
## + carlength         1   2550311 1602705731 2716.0
## + curbweight        1   2479593 1602776449 2716.0
## + boreratio         1   1119585 1604136458 2716.1
## - stroke            1 128248289 1733504332 2718.7
## - compressionratio  1 129515767 1734771810 2718.8
## - peakrpm           1 167032539 1772288582 2722.3
## - enginesize        1 739801500 2345057543 2768.6
##
## Step:  AIC=2706.38
## price ~ wheelbase + carwidth + carheight + enginesize + stroke +
##     compressionratio + horsepower + peakrpm + citympg
##
##                    Df Sum of Sq        RSS    AIC
## - carheight         1   5113543 1613980782 2701.8
## - carwidth          1   9722681 1618589920 2702.3
## - citympg           1  10119903 1618987142 2702.3
## - wheelbase         1  21368834 1630236073 2703.4
## <none>                          1608867239 2706.4
## - horsepower        1  61523298 1670390537 2707.5
## + highwaympg        1   3611196 1605256043 2711.1
## + carlength         1   1474622 1607392617 2711.3
## + curbweight        1   1321360 1607545879 2711.3
## + boreratio         1   1001757 1607865482 2711.4
## - stroke            1 125499785 1734367024 2713.7
## - compressionratio  1 128328284 1737195523 2713.9
## - peakrpm           1 166829542 1775696781 2717.6
## - enginesize        1 750296191 2359163430 2764.4
##
## Step:  AIC=2701.8
## price ~ wheelbase + carwidth + enginesize + stroke + compressionratio +
##     horsepower + peakrpm + citympg
##
##                    Df Sum of Sq        RSS    AIC
## - carwidth          1   7462387 1621443169 2697.4
## - citympg           1  10268704 1624249486 2697.7
## <none>                          1613980782 2701.8
## - wheelbase         1  51043846 1665024628 2701.8
## - horsepower        1  61057342 1675038124 2702.8
## + carheight         1   5113543 1608867239 2706.4
## + highwaympg        1   3571006 1610409777 2706.5
## + curbweight        1   1915354 1612065428 2706.7
## + boreratio         1   1161957 1612818825 2706.8
## + carlength         1    332853 1613647930 2706.9
## - compressionratio  1 130433264 1744414047 2709.5
## - stroke            1 136602131 1750582914 2710.1
## - peakrpm           1 162179675 1776160458 2712.5
## - enginesize        1 746050343 2360031125 2759.4
##
## Step:  AIC=2697.45
## price ~ wheelbase + enginesize + stroke + compressionratio +
##     horsepower + peakrpm + citympg
##
##                    Df Sum of Sq        RSS    AIC
```

```
## - citympg           1  13442845 1634886014 2693.7
## <none>                          1621443169 2697.4
## - horsepower        1  77310491 1698753660 2700.0
## + carwidth          1   7462387 1613980782 2701.8
## + curbweight        1   3870233 1617572936 2702.2
## + highwaympg        1   3108736 1618334433 2702.2
## + carheight         1   2853248 1618589920 2702.3
## + boreratio         1   1224909 1620218260 2702.4
## + carlength         1       132 1621443036 2702.6
## - wheelbase         1 123048423 1744491592 2704.4
## - stroke            1 134034050 1755477219 2705.4
## - compressionratio  1 150306405 1771749574 2707.0
## - peakrpm           1 164255863 1785699032 2708.3
## - enginesize        1 761051172 2382494341 2755.8
##
## Step:  AIC=2693.71
## price ~ wheelbase + enginesize + stroke + compressionratio +
##     horsepower + peakrpm
##
##                    Df Sum of Sq        RSS    AIC
## <none>                          1634886014 2693.7
## + curbweight        1  13539421 1621346593 2697.4
## + citympg           1  13442845 1621443169 2697.4
## + carwidth          1  10636527 1624249486 2697.7
## + highwaympg        1   7315398 1627570616 2698.1
## + carlength         1   3016846 1631869168 2698.5
## + carheight         1   2611136 1632274877 2698.6
## + boreratio         1     21746 1634864267 2698.8
## - stroke            1 131940428 1766826441 2701.4
## - compressionratio  1 138351294 1773237307 2702.0
## - peakrpm           1 158520736 1793406750 2703.9
## - horsepower        1 182074941 1816960955 2706.0
## - wheelbase         1 225284392 1860170405 2709.9
## - enginesize        1 749887317 2384773330 2750.9
```

```
summary(stepwise_model_aic)
```

```
##
## Call:
## lm(formula = price ~ wheelbase + enginesize + stroke + compressionratio +
##     horsepower + peakrpm, data = train_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12790.1  -1726.3   -244.4   1494.6  12381.9
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -41674.100   6842.603  -6.090 8.25e-09 ***
## wheelbase           248.872     53.337   4.666 6.51e-06 ***
## enginesize          119.839     14.077   8.513 1.24e-14 ***
## stroke            -2875.852    805.365  -3.571 0.000472 ***
## compressionratio    300.989     82.314   3.657 0.000348 ***
## horsepower           57.279     13.655   4.195 4.54e-05 ***
```

```
## peakrpm               3.096      0.791   3.914 0.000135 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3217 on 158 degrees of freedom
## Multiple R-squared:  0.8671, Adjusted R-squared:  0.8621
## F-statistic: 171.8 on 6 and 158 DF,  p-value: < 2.2e-16
```

```
summary(stepwise_model_bic)
```

```
##
## Call:
## lm(formula = price ~ wheelbase + enginesize + stroke + compressionratio +
##     horsepower + peakrpm, data = train_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12790.1  -1726.3   -244.4   1494.6  12381.9
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -41674.100   6842.603  -6.090 8.25e-09 ***
## wheelbase           248.872     53.337   4.666 6.51e-06 ***
## enginesize          119.839     14.077   8.513 1.24e-14 ***
## stroke            -2875.852    805.365  -3.571 0.000472 ***
## compressionratio    300.989     82.314   3.657 0.000348 ***
## horsepower           57.279     13.655   4.195 4.54e-05 ***
## peakrpm               3.096      0.791   3.914 0.000135 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3217 on 158 degrees of freedom
## Multiple R-squared:  0.8671, Adjusted R-squared:  0.8621
## F-statistic: 171.8 on 6 and 158 DF,  p-value: < 2.2e-16
```

```
summary(stepwise_model_adjr2)
```

```
##
## Call:
## lm(formula = price ~ wheelbase + enginesize + stroke + compressionratio +
##     horsepower + peakrpm, data = train_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12790.1  -1726.3   -244.4   1494.6  12381.9
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -41674.100   6842.603  -6.090 8.25e-09 ***
## wheelbase           248.872     53.337   4.666 6.51e-06 ***
## enginesize          119.839     14.077   8.513 1.24e-14 ***
## stroke            -2875.852    805.365  -3.571 0.000472 ***
## compressionratio    300.989     82.314   3.657 0.000348 ***
```

```
## horsepower           57.279     13.655   4.195 4.54e-05 ***
## peakrpm               3.096      0.791   3.914 0.000135 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3217 on 158 degrees of freedom
## Multiple R-squared:  0.8671, Adjusted R-squared:  0.8621
## F-statistic: 171.8 on 6 and 158 DF,  p-value: < 2.2e-16
```

All of these summaries suggest that the best model is:

price = -41674.1 + 248.872 * wheelbase + 119.839 * enginesize - 2875.852 * stroke + 300.989 * compressionratio + 57.279 * horsepower + 3.096 * peakrpm

#Testing the stepwise model

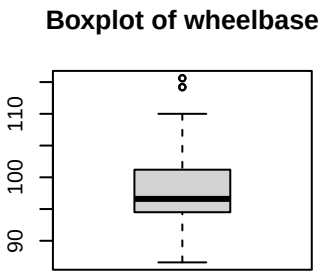
Notice that we are usign the test_data now, and we chose the model using the 'aic' criterion since all three stepwise models above produced the same results.

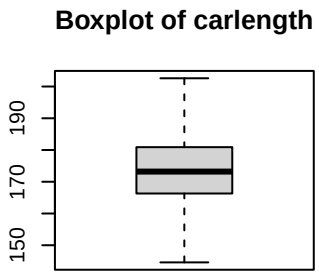
```
stepwise_model_selected <- stepwise_model_aic
predictions <- predict(stepwise_model_selected, newdata = test_data)
rmse <- sqrt(mean((test_data$price - predictions)^2))
rmse
```

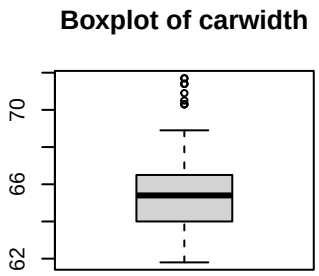
```
## [1] 3493.335
```

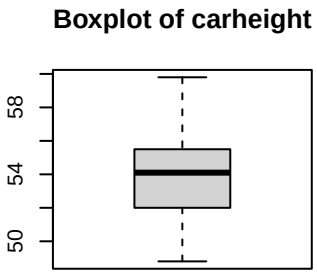
We can see that the RMSE is 3493.335, which is a little large considering the original range of prices is anywhere from 5100 to 45,500.. Perhaps outliers affect this?

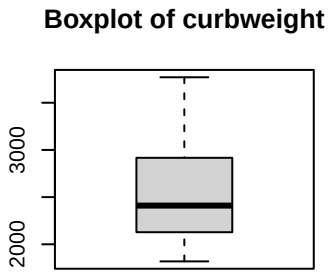
```
continuous_vars <- c('wheelbase', 'carlength', 'carwidth', 'carheight', 'curbweight', 'enginesize', 'bo
remove_outliers_threshold <- function(data, threshold = 3) {
  for (var in continuous_vars) {
    data <- data[abs(data[[var]] - mean(data[[var]])) <= threshold * sd(data[[var]]), ]
  }
  return(data)
}
car_price_data_clean <- remove_outliers_threshold(car_price_data)
par(mfrow = c(2, 3))
for (var in continuous_vars) {
  boxplot(car_price_data_clean[[var]], main = paste("Boxplot of", var))
}
```

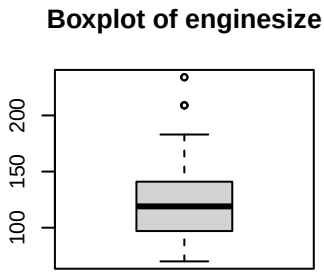

Boxplot of wheelbase
90
100
110

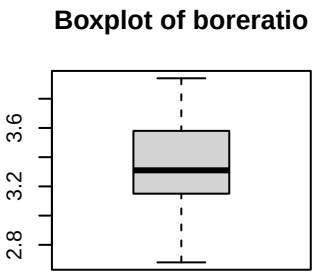

Boxplot of carlength
150
170
190

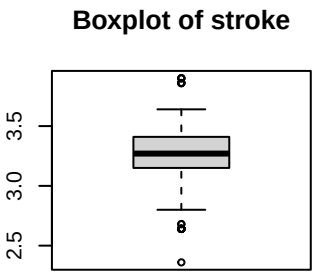

Boxplot of carwidth
62
66
70

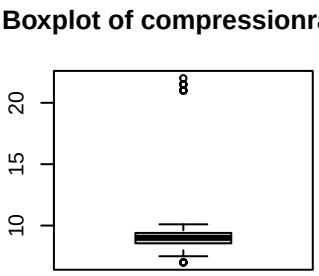

Boxplot of carheight
50
54
58

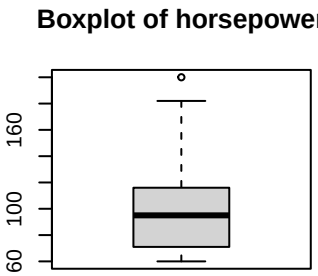

Boxplot of curbweight
2000
3000

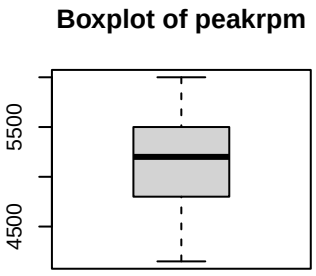

Boxplot of enginesize
100
150
200

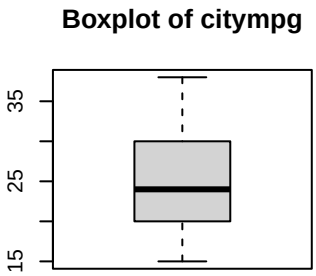

Boxplot of boreratio
2.8
3.2
3.6

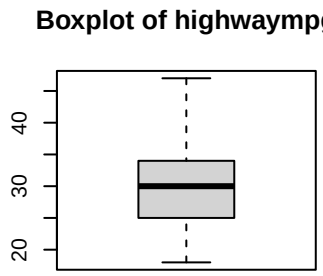

Boxplot of stroke
2.5
3.0
3.5


Boxplot of compressionratio
10
15
20


Boxplot of horsepower
60
100
160


Boxplot of peakrpm
4500
5500


Boxplot of citympg
15
25
35



After outliers were removed, we have 179 observations. Redoing the training and testing data to account for that.

```
#train and test data after outliers were removed
train_data_clean <- car_price_data_clean[1:139, ]
test_data_clean <- car_price_data_clean[140:179, ]
```

We created a new dataset with outliers removed^

```
initial_model <- lm(price ~ wheelbase + carlength + carwidth + carheight + curbweight + enginesize + bo
#stepwise using aic
stepwise_model_aic <- step(initial_model, direction = "both", trace = 1, k = log(nrow(train_data_clean)
```

```
## Start:  AIC=2267.97
## price ~ wheelbase + carlength + carwidth + carheight + curbweight +
##     enginesize + boreratio + stroke + compressionratio + horsepower +
##     peakrpm + citympg + highwaympg
##
##                    Df Sum of Sq        RSS    AIC
## - carlength         1   2001246 1033032315 2263.3
## - highwaympg        1   4028512 1035059581 2263.6
## - curbweight        1   7481411 1038512480 2264.0
## - carheight         1   8579899 1039610969 2264.2
## - carwidth          1  23293176 1054324246 2266.1
## - citympg           1  28867784 1059898853 2266.9
```

```
## - horsepower        1  33608850 1064639920 2267.5
## <none>                          1031031069 2268.0
## - wheelbase         1  45928409 1076959479 2269.1
## - boreratio         1  63603027 1094634097 2271.4
## - peakrpm           1  69509844 1100540913 2272.1
## - compressionratio  1 127943751 1158974820 2279.3
## - stroke            1 225158848 1256189917 2290.5
## - enginesize        1 335100314 1366131384 2302.2
##
## Step:  AIC=2263.31
## price ~ wheelbase + carwidth + carheight + curbweight + enginesize +
##     boreratio + stroke + compressionratio + horsepower + peakrpm +
##     citympg + highwaympg
##
##                    Df Sum of Sq        RSS    AIC
## - highwaympg        1   5969627 1039001942 2259.2
## - curbweight        1   5975562 1039007877 2259.2
## - carheight         1   7028507 1040060822 2259.3
## - carwidth          1  27374797 1060407112 2262.0
## - horsepower        1  31999442 1065031757 2262.6
## - citympg           1  36718124 1069750439 2263.2
## <none>                          1033032315 2263.3
## - wheelbase         1  59080328 1092112643 2266.1
## - boreratio         1  61704947 1094737262 2266.4
## - peakrpm           1  69832972 1102865288 2267.5
## + carlength         1   2001246 1031031069 2268.0
## - compressionratio  1 126540968 1159573283 2274.4
## - stroke            1 223268338 1256300653 2285.6
## - enginesize        1 333506999 1366539314 2297.3
##
## Step:  AIC=2259.17
## price ~ wheelbase + carwidth + carheight + curbweight + enginesize +
##     boreratio + stroke + compressionratio + horsepower + peakrpm +
##     citympg
##
##                    Df Sum of Sq        RSS    AIC
## - carheight         1   7277743 1046279685 2255.2
## - curbweight        1  10842729 1049844671 2255.7
## - carwidth          1  27161972 1066163914 2257.8
## - horsepower        1  34204017 1073205959 2258.7
## <none>                          1039001942 2259.2
## - citympg           1  48472717 1087474659 2260.6
## - boreratio         1  59756327 1098758269 2262.0
## - wheelbase         1  60263831 1099265773 2262.1
## - peakrpm           1  68427656 1107429598 2263.1
## + highwaympg        1   5969627 1033032315 2263.3
## + carlength         1   3942361 1035059581 2263.6
## - compressionratio  1 130704448 1169706390 2270.7
## - stroke            1 217406471 1256408413 2280.7
## - enginesize        1 344856267 1383858209 2294.1
##
## Step:  AIC=2255.21
## price ~ wheelbase + carwidth + curbweight + enginesize + boreratio +
##     stroke + compressionratio + horsepower + peakrpm + citympg
```

```
##
##                    Df Sum of Sq        RSS    AIC
## - curbweight        1  12359425 1058639110 2251.9
## - carwidth          1  35693209 1081972894 2254.9
## <none>                          1046279685 2255.2
## - horsepower        1  38998926 1085278611 2255.4
## - citympg           1  49516770 1095796455 2256.7
## - boreratio         1  56896647 1103176332 2257.6
## - wheelbase         1  57634103 1103913788 2257.7
## + carheight         1   7277743 1039001942 2259.2
## + highwaympg        1   6218863 1040060822 2259.3
## - peakrpm           1  75297279 1121576964 2259.9
## + carlength         1   1559452 1044720232 2259.9
## - compressionratio  1 142966740 1189246425 2268.1
## - stroke            1 210953729 1257233413 2275.8
## - enginesize        1 346994680 1393274364 2290.1
##
## Step:  AIC=2251.91
## price ~ wheelbase + carwidth + enginesize + boreratio + stroke +
##     compressionratio + horsepower + peakrpm + citympg
##
##                    Df Sum of Sq        RSS    AIC
## - carwidth          1  26559512 1085198622 2250.4
## - horsepower        1  29617490 1088256600 2250.8
## - citympg           1  37632632 1096271742 2251.8
## <none>                          1058639110 2251.9
## - wheelbase         1  45342565 1103981675 2252.8
## - boreratio         1  56351529 1114990639 2254.2
## + curbweight        1  12359425 1046279685 2255.2
## + highwaympg        1  11631301 1047007809 2255.3
## + carheight         1   8794440 1049844671 2255.7
## + carlength         1    214880 1058424230 2256.8
## - peakrpm           1  88413490 1147052600 2258.1
## - compressionratio  1 144688401 1203327511 2264.8
## - stroke            1 202739950 1261379060 2271.3
## - enginesize        1 382479575 1441118685 2289.8
##
## Step:  AIC=2250.42
## price ~ wheelbase + enginesize + boreratio + stroke + compressionratio +
##     horsepower + peakrpm + citympg
##
##                    Df Sum of Sq        RSS    AIC
## - horsepower        1  34177215 1119375837 2249.8
## <none>                          1085198622 2250.4
## + carwidth          1  26559512 1058639110 2251.9
## - citympg           1  54120714 1139319336 2252.2
## - boreratio         1  54469574 1139668196 2252.3
## + carheight         1  15937702 1069260921 2253.3
## + highwaympg        1   8405066 1076793556 2254.3
## + curbweight        1   3225729 1081972894 2254.9
## + carlength         1   2360392 1082838231 2255.1
## - peakrpm           1 106510896 1191709518 2258.5
## - wheelbase         1 123324856 1208523479 2260.4
## - compressionratio  1 180228636 1265427258 2266.8
```

```
## - stroke            1 201605625 1286804247 2269.2
## - enginesize        1 456473753 1541672376 2294.3
##
## Step:  AIC=2249.79
## price ~ wheelbase + enginesize + boreratio + stroke + compressionratio +
##     peakrpm + citympg
##
##                    Df  Sum of Sq        RSS    AIC
## <none>                            1119375837 2249.8
## - boreratio         1   44129873 1163505711 2250.2
## + horsepower        1   34177215 1085198622 2250.4
## + carwidth          1   31119238 1088256600 2250.8
## + carheight         1   21911791 1097464047 2252.0
## + highwaympg        1    7636336 1111739502 2253.8
## + carlength         1    1345456 1118030381 2254.6
## + curbweight        1       7358 1119368479 2254.7
## - wheelbase         1  101290981 1220666819 2256.9
## - citympg           1  170649353 1290025191 2264.6
## - compressionratio  1  182629563 1302005400 2265.9
## - peakrpm           1  206198883 1325574720 2268.4
## - stroke            1  214645715 1334021553 2269.2
## - enginesize        1 1328200441 2447576278 2353.6
```

```
#stepwise using bic
stepwise_model_bic <- step(initial_model, direction = "both", trace = 1, k = log(nrow(train_data_clean)))
```

```
## Start:  AIC=2267.97
## price ~ wheelbase + carlength + carwidth + carheight + curbweight +
##     enginesize + boreratio + stroke + compressionratio + horsepower +
##     peakrpm + citympg + highwaympg
##
##                    Df Sum of Sq        RSS    AIC
## - carlength         1   2001246 1033032315 2263.3
## - highwaympg        1   4028512 1035059581 2263.6
## - curbweight        1   7481411 1038512480 2264.0
## - carheight         1   8579899 1039610969 2264.2
## - carwidth          1  23293176 1054324246 2266.1
## - citympg           1  28867784 1059898853 2266.9
## - horsepower        1  33608850 1064639920 2267.5
## <none>                          1031031069 2268.0
## - wheelbase         1  45928409 1076959479 2269.1
## - boreratio         1  63603027 1094634097 2271.4
## - peakrpm           1  69509844 1100540913 2272.1
## - compressionratio  1 127943751 1158974820 2279.3
## - stroke            1 225158848 1256189917 2290.5
## - enginesize        1 335100314 1366131384 2302.2
##
## Step:  AIC=2263.31
## price ~ wheelbase + carwidth + carheight + curbweight + enginesize +
##     boreratio + stroke + compressionratio + horsepower + peakrpm +
##     citympg + highwaympg
##
##                    Df Sum of Sq        RSS    AIC
## - highwaympg        1   5969627 1039001942 2259.2
```

```
## - curbweight        1   5975562 1039007877 2259.2
## - carheight         1   7028507 1040060822 2259.3
## - carwidth          1  27374797 1060407112 2262.0
## - horsepower        1  31999442 1065031757 2262.6
## - citympg           1  36718124 1069750439 2263.2
## <none>                          1033032315 2263.3
## - wheelbase         1  59080328 1092112643 2266.1
## - boreratio         1  61704947 1094737262 2266.4
## - peakrpm           1  69832972 1102865288 2267.5
## + carlength         1   2001246 1031031069 2268.0
## - compressionratio  1 126540968 1159573283 2274.4
## - stroke            1 223268338 1256300653 2285.6
## - enginesize        1 333506999 1366539314 2297.3
##
## Step:  AIC=2259.17
## price ~ wheelbase + carwidth + carheight + curbweight + enginesize +
##     boreratio + stroke + compressionratio + horsepower + peakrpm +
##     citympg
##
##                    Df Sum of Sq        RSS    AIC
## - carheight         1   7277743 1046279685 2255.2
## - curbweight        1  10842729 1049844671 2255.7
## - carwidth          1  27161972 1066163914 2257.8
## - horsepower        1  34204017 1073205959 2258.7
## <none>                          1039001942 2259.2
## - citympg           1  48472717 1087474659 2260.6
## - boreratio         1  59756327 1098758269 2262.0
## - wheelbase         1  60263831 1099265773 2262.1
## - peakrpm           1  68427656 1107429598 2263.1
## + highwaympg        1   5969627 1033032315 2263.3
## + carlength         1   3942361 1035059581 2263.6
## - compressionratio  1 130704448 1169706390 2270.7
## - stroke            1 217406471 1256408413 2280.7
## - enginesize        1 344856267 1383858209 2294.1
##
## Step:  AIC=2255.21
## price ~ wheelbase + carwidth + curbweight + enginesize + boreratio +
##     stroke + compressionratio + horsepower + peakrpm + citympg
##
##                    Df Sum of Sq        RSS    AIC
## - curbweight        1  12359425 1058639110 2251.9
## - carwidth          1  35693209 1081972894 2254.9
## <none>                          1046279685 2255.2
## - horsepower        1  38998926 1085278611 2255.4
## - citympg           1  49516770 1095796455 2256.7
## - boreratio         1  56896647 1103176332 2257.6
## - wheelbase         1  57634103 1103913788 2257.7
## + carheight         1   7277743 1039001942 2259.2
## + highwaympg        1   6218863 1040060822 2259.3
## - peakrpm           1  75297279 1121576964 2259.9
## + carlength         1   1559452 1044720232 2259.9
## - compressionratio  1 142966740 1189246425 2268.1
## - stroke            1 210953729 1257233413 2275.8
## - enginesize        1 346994680 1393274364 2290.1
```

```
##
## Step:  AIC=2251.91
## price ~ wheelbase + carwidth + enginesize + boreratio + stroke +
##     compressionratio + horsepower + peakrpm + citympg
##
##                    Df Sum of Sq        RSS    AIC
## - carwidth          1  26559512 1085198622 2250.4
## - horsepower        1  29617490 1088256600 2250.8
## - citympg           1  37632632 1096271742 2251.8
## <none>                          1058639110 2251.9
## - wheelbase         1  45342565 1103981675 2252.8
## - boreratio         1  56351529 1114990639 2254.2
## + curbweight        1  12359425 1046279685 2255.2
## + highwaympg        1  11631301 1047007809 2255.3
## + carheight         1   8794440 1049844671 2255.7
## + carlength         1    214880 1058424230 2256.8
## - peakrpm           1  88413490 1147052600 2258.1
## - compressionratio  1 144688401 1203327511 2264.8
## - stroke            1 202739950 1261379060 2271.3
## - enginesize        1 382479575 1441118685 2289.8
##
## Step:  AIC=2250.42
## price ~ wheelbase + enginesize + boreratio + stroke + compressionratio +
##     horsepower + peakrpm + citympg
##
##                    Df Sum of Sq        RSS    AIC
## - horsepower        1  34177215 1119375837 2249.8
## <none>                          1085198622 2250.4
## + carwidth          1  26559512 1058639110 2251.9
## - citympg           1  54120714 1139319336 2252.2
## - boreratio         1  54469574 1139668196 2252.3
## + carheight         1  15937702 1069260921 2253.3
## + highwaympg        1   8405066 1076793556 2254.3
## + curbweight        1   3225729 1081972894 2254.9
## + carlength         1   2360392 1082838231 2255.1
## - peakrpm           1 106510896 1191709518 2258.5
## - wheelbase         1 123324856 1208523479 2260.4
## - compressionratio  1 180228636 1265427258 2266.8
## - stroke            1 201605625 1286804247 2269.2
## - enginesize        1 456473753 1541672376 2294.3
##
## Step:  AIC=2249.79
## price ~ wheelbase + enginesize + boreratio + stroke + compressionratio +
##     peakrpm + citympg
##
##                    Df  Sum of Sq        RSS    AIC
## <none>                           1119375837 2249.8
## - boreratio         1   44129873 1163505711 2250.2
## + horsepower        1   34177215 1085198622 2250.4
## + carwidth          1   31119238 1088256600 2250.8
## + carheight         1   21911791 1097464047 2252.0
## + highwaympg        1    7636336 1111739502 2253.8
## + carlength         1    1345456 1118030381 2254.6
## + curbweight        1       7358 1119368479 2254.7
```

```
## - wheelbase         1  101290981 1220666819 2256.9
## - citympg           1  170649353 1290025191 2264.6
## - compressionratio  1  182629563 1302005400 2265.9
## - peakrpm           1  206198883 1325574720 2268.4
## - stroke            1  214645715 1334021553 2269.2
## - enginesize        1 1328200441 2447576278 2353.6
```

```
#stepwise using adjr2
stepwise_model_adjr2 <- step(initial_model, direction = "both", trace = 1, k = log(nrow(train_data_clea
```

```
## Start:  AIC=2267.97
## price ~ wheelbase + carlength + carwidth + carheight + curbweight +
##     enginesize + boreratio + stroke + compressionratio + horsepower +
##     peakrpm + citympg + highwaympg
##
##                    Df Sum of Sq        RSS    AIC
## - carlength         1   2001246 1033032315 2263.3
## - highwaympg        1   4028512 1035059581 2263.6
## - curbweight        1   7481411 1038512480 2264.0
## - carheight         1   8579899 1039610969 2264.2
## - carwidth          1  23293176 1054324246 2266.1
## - citympg           1  28867784 1059898853 2266.9
## - horsepower        1  33608850 1064639920 2267.5
## <none>                          1031031069 2268.0
## - wheelbase         1  45928409 1076959479 2269.1
## - boreratio         1  63603027 1094634097 2271.4
## - peakrpm           1  69509844 1100540913 2272.1
## - compressionratio  1 127943751 1158974820 2279.3
## - stroke            1 225158848 1256189917 2290.5
## - enginesize        1 335100314 1366131384 2302.2
##
## Step:  AIC=2263.31
## price ~ wheelbase + carwidth + carheight + curbweight + enginesize +
##     boreratio + stroke + compressionratio + horsepower + peakrpm +
##     citympg + highwaympg
##
##                    Df Sum of Sq        RSS    AIC
## - highwaympg        1   5969627 1039001942 2259.2
## - curbweight        1   5975562 1039007877 2259.2
## - carheight         1   7028507 1040060822 2259.3
## - carwidth          1  27374797 1060407112 2262.0
## - horsepower        1  31999442 1065031757 2262.6
## - citympg           1  36718124 1069750439 2263.2
## <none>                          1033032315 2263.3
## - wheelbase         1  59080328 1092112643 2266.1
## - boreratio         1  61704947 1094737262 2266.4
## - peakrpm           1  69832972 1102865288 2267.5
## + carlength         1   2001246 1031031069 2268.0
## - compressionratio  1 126540968 1159573283 2274.4
## - stroke            1 223268338 1256300653 2285.6
## - enginesize        1 333506999 1366539314 2297.3
##
## Step:  AIC=2259.17
## price ~ wheelbase + carwidth + carheight + curbweight + enginesize +
```

```
##     boreratio + stroke + compressionratio + horsepower + peakrpm +
##     citympg
##
##                    Df Sum of Sq        RSS    AIC
## - carheight         1   7277743 1046279685 2255.2
## - curbweight        1  10842729 1049844671 2255.7
## - carwidth          1  27161972 1066163914 2257.8
## - horsepower        1  34204017 1073205959 2258.7
## <none>                          1039001942 2259.2
## - citympg           1  48472717 1087474659 2260.6
## - boreratio         1  59756327 1098758269 2262.0
## - wheelbase         1  60263831 1099265773 2262.1
## - peakrpm           1  68427656 1107429598 2263.1
## + highwaympg        1   5969627 1033032315 2263.3
## + carlength         1   3942361 1035059581 2263.6
## - compressionratio  1 130704448 1169706390 2270.7
## - stroke            1 217406471 1256408413 2280.7
## - enginesize        1 344856267 1383858209 2294.1
##
## Step:  AIC=2255.21
## price ~ wheelbase + carwidth + curbweight + enginesize + boreratio +
##     stroke + compressionratio + horsepower + peakrpm + citympg
##
##                    Df Sum of Sq        RSS    AIC
## - curbweight        1  12359425 1058639110 2251.9
## - carwidth          1  35693209 1081972894 2254.9
## <none>                          1046279685 2255.2
## - horsepower        1  38998926 1085278611 2255.4
## - citympg           1  49516770 1095796455 2256.7
## - boreratio         1  56896647 1103176332 2257.6
## - wheelbase         1  57634103 1103913788 2257.7
## + carheight         1   7277743 1039001942 2259.2
## + highwaympg        1   6218863 1040060822 2259.3
## - peakrpm           1  75297279 1121576964 2259.9
## + carlength         1   1559452 1044720232 2259.9
## - compressionratio  1 142966740 1189246425 2268.1
## - stroke            1 210953729 1257233413 2275.8
## - enginesize        1 346994680 1393274364 2290.1
##
## Step:  AIC=2251.91
## price ~ wheelbase + carwidth + enginesize + boreratio + stroke +
##     compressionratio + horsepower + peakrpm + citympg
##
##                    Df Sum of Sq        RSS    AIC
## - carwidth          1  26559512 1085198622 2250.4
## - horsepower        1  29617490 1088256600 2250.8
## - citympg           1  37632632 1096271742 2251.8
## <none>                          1058639110 2251.9
## - wheelbase         1  45342565 1103981675 2252.8
## - boreratio         1  56351529 1114990639 2254.2
## + curbweight        1  12359425 1046279685 2255.2
## + highwaympg        1  11631301 1047007809 2255.3
## + carheight         1   8794440 1049844671 2255.7
## + carlength         1    214880 1058424230 2256.8
```

```
## - peakrpm           1  88413490 1147052600 2258.1
## - compressionratio  1 144688401 1203327511 2264.8
## - stroke            1 202739950 1261379060 2271.3
## - enginesize        1 382479575 1441118685 2289.8
##
## Step:  AIC=2250.42
## price ~ wheelbase + enginesize + boreratio + stroke + compressionratio +
##     horsepower + peakrpm + citympg
##
##                    Df Sum of Sq        RSS    AIC
## - horsepower        1  34177215 1119375837 2249.8
## <none>                         1085198622 2250.4
## + carwidth          1  26559512 1058639110 2251.9
## - citympg           1  54120714 1139319336 2252.2
## - boreratio         1  54469574 1139668196 2252.3
## + carheight         1  15937702 1069260921 2253.3
## + highwaympg        1   8405066 1076793556 2254.3
## + curbweight        1   3225729 1081972894 2254.9
## + carlength         1   2360392 1082838231 2255.1
## - peakrpm           1 106510896 1191709518 2258.5
## - wheelbase         1 123324856 1208523479 2260.4
## - compressionratio  1 180228636 1265427258 2266.8
## - stroke            1 201605625 1286804247 2269.2
## - enginesize        1 456473753 1541672376 2294.3
##
## Step:  AIC=2249.79
## price ~ wheelbase + enginesize + boreratio + stroke + compressionratio +
##     peakrpm + citympg
##
##                    Df  Sum of Sq        RSS    AIC
## <none>                           1119375837 2249.8
## - boreratio         1   44129873 1163505711 2250.2
## + horsepower        1   34177215 1085198622 2250.4
## + carwidth          1   31119238 1088256600 2250.8
## + carheight         1   21911791 1097464047 2252.0
## + highwaympg        1    7636336 1111739502 2253.8
## + carlength         1    1345456 1118030381 2254.6
## + curbweight        1       7358 1119368479 2254.7
## - wheelbase         1  101290981 1220666819 2256.9
## - citympg           1  170649353 1290025191 2264.6
## - compressionratio  1  182629563 1302005400 2265.9
## - peakrpm           1  206198883 1325574720 2268.4
## - stroke            1  214645715 1334021553 2269.2
## - enginesize        1 1328200441 2447576278 2353.6
```

```
summary(stepwise_model_aic)
```

```
##
## Call:
## lm(formula = price ~ wheelbase + enginesize + boreratio + stroke +
##     compressionratio + peakrpm + citympg, data = train_data_clean)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -10711.7  -1305.5    -32.5   1319.7  12747.2
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -1.727e+04  1.158e+04  -1.491 0.138283
## wheelbase         2.311e+02  6.712e+01   3.443 0.000773 ***
## enginesize        1.549e+02  1.242e+01  12.468  < 2e-16 ***
## boreratio        -3.537e+03  1.556e+03  -2.273 0.024682 *
## stroke           -5.117e+03  1.021e+03  -5.012 1.71e-06 ***
## compressionratio  5.177e+02  1.120e+02   4.623 8.95e-06 ***
## peakrpm           3.872e+00  7.881e-01   4.912 2.63e-06 ***
## citympg          -3.339e+02  7.471e+01  -4.469 1.68e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2923 on 131 degrees of freedom
## Multiple R-squared:  0.8369, Adjusted R-squared:  0.8282
## F-statistic: 96.02 on 7 and 131 DF,  p-value: < 2.2e-16
```

```
summary(stepwise_model_bic)
```

```
##
## Call:
## lm(formula = price ~ wheelbase + enginesize + boreratio + stroke +
##     compressionratio + peakrpm + citympg, data = train_data_clean)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10711.7  -1305.5    -32.5   1319.7  12747.2
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -1.727e+04  1.158e+04  -1.491 0.138283
## wheelbase         2.311e+02  6.712e+01   3.443 0.000773 ***
## enginesize        1.549e+02  1.242e+01  12.468  < 2e-16 ***
## boreratio        -3.537e+03  1.556e+03  -2.273 0.024682 *
## stroke           -5.117e+03  1.021e+03  -5.012 1.71e-06 ***
## compressionratio  5.177e+02  1.120e+02   4.623 8.95e-06 ***
## peakrpm           3.872e+00  7.881e-01   4.912 2.63e-06 ***
## citympg          -3.339e+02  7.471e+01  -4.469 1.68e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2923 on 131 degrees of freedom
## Multiple R-squared:  0.8369, Adjusted R-squared:  0.8282
## F-statistic: 96.02 on 7 and 131 DF,  p-value: < 2.2e-16
```

```
summary(stepwise_model_adjr2)
```

```
##
## Call:
## lm(formula = price ~ wheelbase + enginesize + boreratio + stroke +
##     compressionratio + peakrpm + citympg, data = train_data_clean)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10711.7  -1305.5    -32.5   1319.7  12747.2
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -1.727e+04  1.158e+04  -1.491 0.138283
## wheelbase         2.311e+02  6.712e+01   3.443 0.000773 ***
## enginesize        1.549e+02  1.242e+01  12.468  < 2e-16 ***
## boreratio        -3.537e+03  1.556e+03  -2.273 0.024682 *
## stroke           -5.117e+03  1.021e+03  -5.012 1.71e-06 ***
## compressionratio  5.177e+02  1.120e+02   4.623 8.95e-06 ***
## peakrpm           3.872e+00  7.881e-01   4.912 2.63e-06 ***
## citympg          -3.339e+02  7.471e+01  -4.469 1.68e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2923 on 131 degrees of freedom
## Multiple R-squared:  0.8369, Adjusted R-squared:  0.8282
## F-statistic: 96.02 on 7 and 131 DF,  p-value: < 2.2e-16
```

```
stepwise_model_selected <- stepwise_model_aic
predictions <- predict(stepwise_model_selected, newdata = test_data_clean)
rmse <- sqrt(mean((test_data$price - predictions)^2))
rmse
```

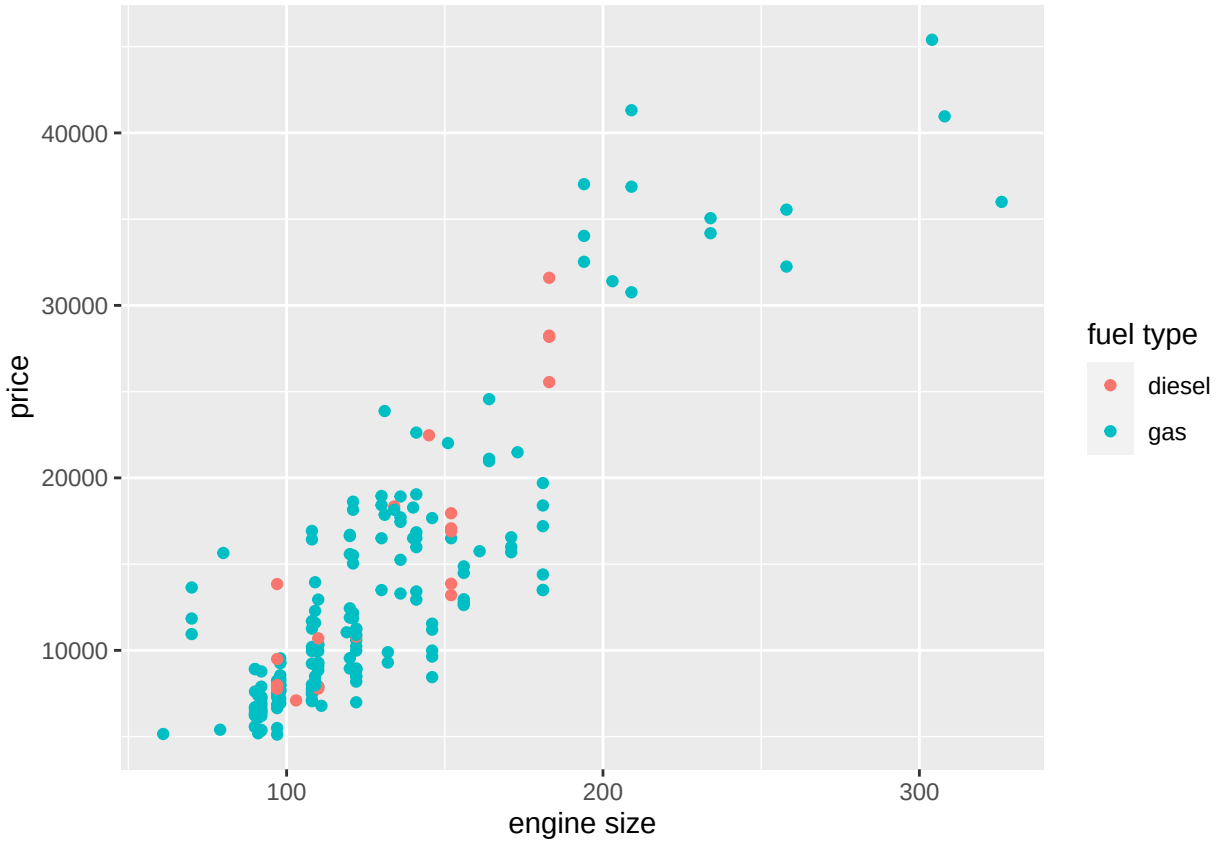
```
## [1] 6028.778
```

RMSE went up after removing outliers, and the adjr2 went down.

#More EDA based on categorical data

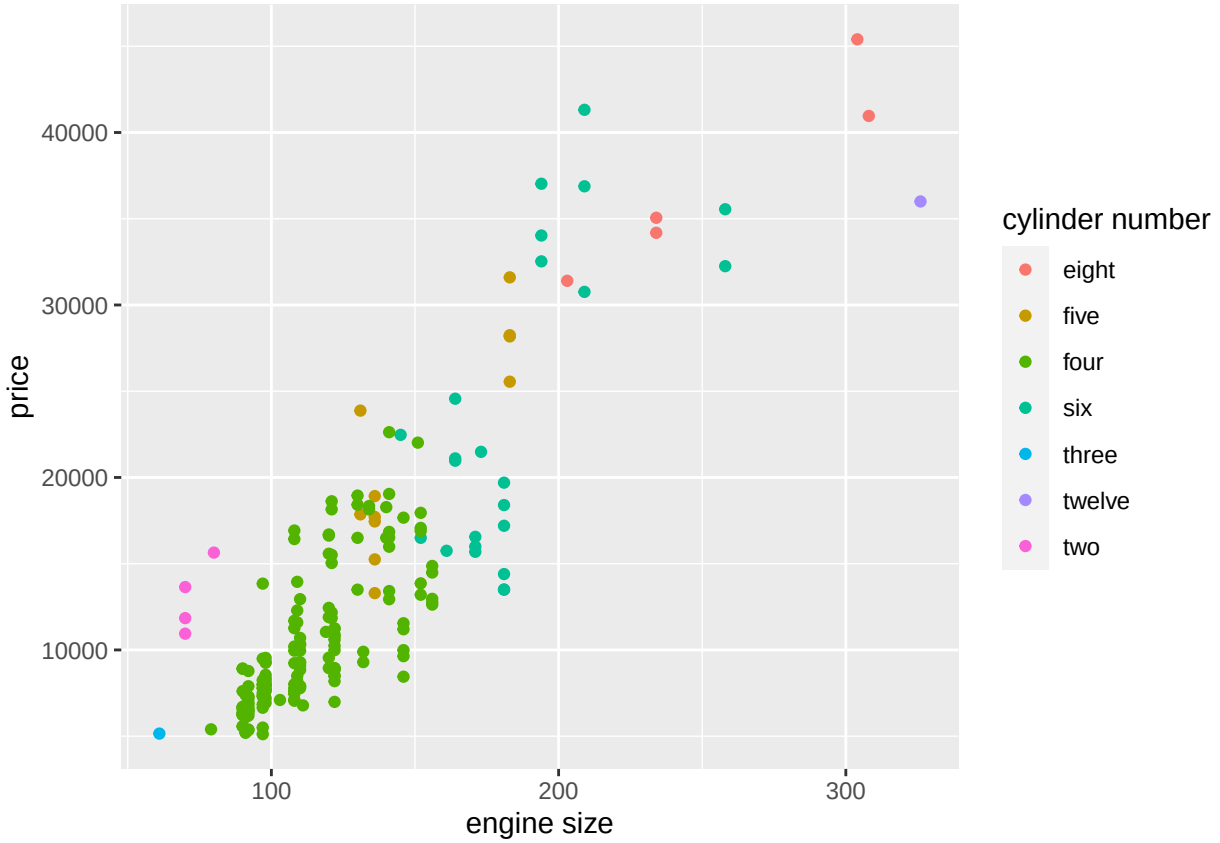
```
library(dplyr)
car_price_data_clean <- car_price_data %>%
  mutate(
    bmwx1 = as.integer(CarName == "bmw x1"),
    bmwx3 = as.integer(CarName == "bmw x3"),
    bmwx4 = as.integer(CarName == "bmw x4"),
    bmwx5 = as.integer(CarName == "bmw x5"),
    bmwz4 = as.integer(CarName == "bmw z4"),
    buick_century = as.integer(CarName == "buick century"),
    buick_regal_sc = as.integer(CarName == "buick regal sport coupe (turbo)"),
    peugeot304 = as.integer(CarName == "peugeot 304"),
    peugeot504 = as.integer(CarName == "peugeot 504"),
    peugeot604_sl = as.integer(CarName == "peugeot 604 sl"),
    porsche_panamera = as.integer(CarName == "porsche panamera"),
    porsce_boxter = as.integer(CarName == "porsche boxter"),
    porsche_cayenne = as.integer(CarName == "porsche cayenne"),
    fuelgas = as.integer(fueltype == "gas"),
    cyl4 = as.integer(cylindernumber == "four")
  )
```

```
library(ggplot2)
ggplot(car_price_data_clean, aes(x = enginesize, y = price, color = fueltype)) +
  geom_point() +
  labs(x = "engine size", y = "price", color = "fuel type")
```



From this chart, we can see that most cars in the dataset use gas.

```
library(ggplot2)
ggplot(car_price_data_clean, aes(x = enginesize, y = price, color = cylindernumber)) +
  geom_point() +
  labs(x = "engine size", y = "price", color = "cylinder number")
```



A little hard to read, but four cylinders seem to be associated with lower price.

#Stepwise Regression Regression vs. Multiple Linear Regression Model Price = -41674.100 + 248.872 * wheelbase + 119.839 * enginesize – 2875.852 * stroke + 300.989 * compressionratio + 57.279 * horsepower + 3.096 * peakrpm

Price = -48703.162 + 153.510 *wheelbase* – *35.536* carlength + 255.082 * carwidth + 114.007 * carheight + 1.136 * curbweight + 123.411 * enginesize – 324.821 * boreratio –2975.049 * stroke + 310.901 * compressionratio + 37.003 * horsepower – 3.233 * peakrpm – 216.214 * citympg + 145.037 * highwaympg

#Conlusion

Q. Which variables significantly predict the price of a car? A. The most significant variables are enginesize, stroke, compressionratio, horsepower, and peakrpm. However, the other variables like wheelbase, carlength, carwidth, carheight, curbweight, boreratio, horsepower, citympg, and highwaympg also contribute to predict the price of a car.

Q. How well do these variables explain car prices? A. We can conclude that from both models, 85.89% of the variability in car price is explained by all of numerical variables.