# CarPricePrediction

Logan Sartain

Please use the Powerpoint along with the R Code as an explanation to the thought process and evaluation of the results from the models built to predict car price.

Read in the Data and some Simple EDA

```
## Car Price Data
library(readr)
carprice <- read.csv("~/Spring Semester Classes 2022/BANA 4080/Final Project/Data/CarPrice_Assig
nment.csv", header = TRUE, stringsAsFactors = TRUE)

####check for missing data###
options(max.print = 175)
is.na(carprice)
```

```
##        car_ID symboling CarName fueltype aspiration doornumber carbody
##   [1,]  FALSE     FALSE   FALSE    FALSE      FALSE      FALSE   FALSE
##   [2,]  FALSE     FALSE   FALSE    FALSE      FALSE      FALSE   FALSE
##   [3,]  FALSE     FALSE   FALSE    FALSE      FALSE      FALSE   FALSE
##   [4,]  FALSE     FALSE   FALSE    FALSE      FALSE      FALSE   FALSE
##   [5,]  FALSE     FALSE   FALSE    FALSE      FALSE      FALSE   FALSE
##   [6,]  FALSE     FALSE   FALSE    FALSE      FALSE      FALSE   FALSE
##        drivewheel enginelocation wheelbase carlength carwidth carheight
##   [1,]      FALSE          FALSE     FALSE     FALSE    FALSE     FALSE
##   [2,]      FALSE          FALSE     FALSE     FALSE    FALSE     FALSE
##   [3,]      FALSE          FALSE     FALSE     FALSE    FALSE     FALSE
##   [4,]      FALSE          FALSE     FALSE     FALSE    FALSE     FALSE
##   [5,]      FALSE          FALSE     FALSE     FALSE    FALSE     FALSE
##   [6,]      FALSE          FALSE     FALSE     FALSE    FALSE     FALSE
##        curbweight enginetype cylindernumber enginesize fuelsystem boreratio
##   [1,]      FALSE      FALSE          FALSE      FALSE      FALSE     FALSE
##   [2,]      FALSE      FALSE          FALSE      FALSE      FALSE     FALSE
##   [3,]      FALSE      FALSE          FALSE      FALSE      FALSE     FALSE
##   [4,]      FALSE      FALSE          FALSE      FALSE      FALSE     FALSE
##   [5,]      FALSE      FALSE          FALSE      FALSE      FALSE     FALSE
##   [6,]      FALSE      FALSE          FALSE      FALSE      FALSE     FALSE
##        stroke compressionratio horsepower peakrpm citympg highwaympg price
##   [1,]  FALSE            FALSE      FALSE   FALSE   FALSE      FALSE FALSE
##   [2,]  FALSE            FALSE      FALSE   FALSE   FALSE      FALSE FALSE
##   [3,]  FALSE            FALSE      FALSE   FALSE   FALSE      FALSE FALSE
##   [4,]  FALSE            FALSE      FALSE   FALSE   FALSE      FALSE FALSE
##   [5,]  FALSE            FALSE      FALSE   FALSE   FALSE      FALSE FALSE
##   [6,]  FALSE            FALSE      FALSE   FALSE   FALSE      FALSE FALSE
##  [ reached getOption("max.print") -- omitted 199 rows ]
```

```
####### summary statistics ###########
summary(carprice)
```

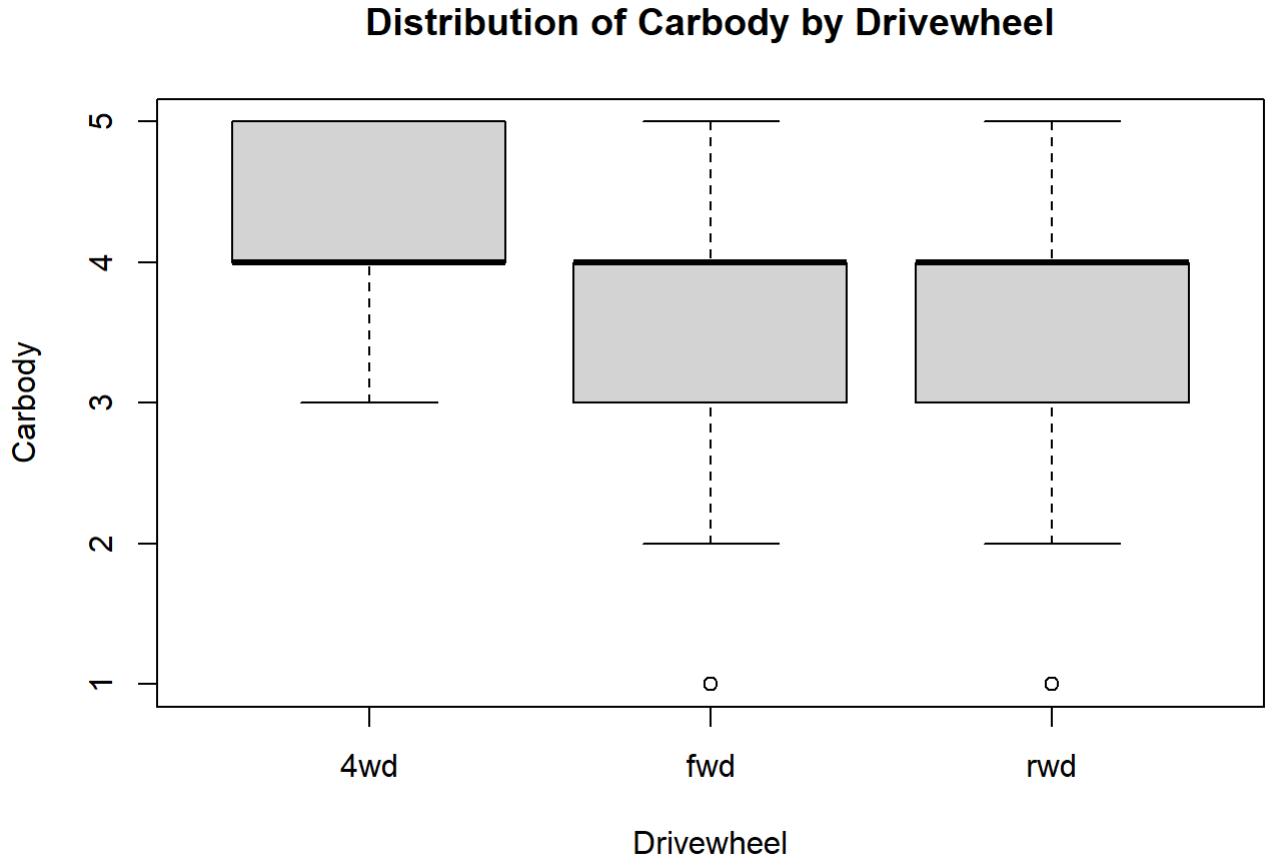
```
##      car_ID      symboling                 CarName      fueltype   aspiration
##  Min.   :  1   Min.   :-2.0000   peugeot 504   :  6   diesel: 20   std  :168
##  1st Qu.: 52   1st Qu.: 0.0000   toyota corolla:  6   gas   :185   turbo: 37
##  Median :103   Median : 1.0000   toyota corona :  6
##  Mean   :103   Mean   : 0.8341   subaru dl     :  4
##  3rd Qu.:154   3rd Qu.: 2.0000   honda civic   :  3
##  Max.   :205   Max.   : 3.0000   mazda 626     :  3
##  doornumber        carbody   drivewheel enginelocation   wheelbase
##  four:115   convertible: 6   4wd:  9    front:202      Min.   : 86.60
##  two : 90   hardtop    : 8   fwd:120    rear :  3      1st Qu.: 94.50
##             hatchback  :70   rwd: 76                   Median : 97.00
##             sedan      :96                             Mean   : 98.76
##             wagon      :25                             3rd Qu.:102.40
##                                                        Max.   :120.90
##    carlength        carwidth       carheight       curbweight   enginetype
##  Min.   :141.1   Min.   :60.30   Min.   :47.80   Min.   :1488   dohc : 12
##  1st Qu.:166.3   1st Qu.:64.10   1st Qu.:52.00   1st Qu.:2145   dohcv:  1
##  Median :173.2   Median :65.50   Median :54.10   Median :2414   l    : 12
##  Mean   :174.0   Mean   :65.91   Mean   :53.72   Mean   :2556   ohc  :148
##  3rd Qu.:183.1   3rd Qu.:66.90   3rd Qu.:55.50   3rd Qu.:2935   ohcf : 15
##  Max.   :208.1   Max.   :72.30   Max.   :59.80   Max.   :4066   ohcv : 13
##  cylindernumber   enginesize      fuelsystem   boreratio        stroke
##  eight :  5     Min.   : 61.0   mpfi   :94   Min.   :2.54   Min.   :2.070
##  five  : 11     1st Qu.: 97.0   2bbl   :66   1st Qu.:3.15   1st Qu.:3.110
##  four  :159     Median :120.0   idi    :20   Median :3.31   Median :3.290
##  six   : 24     Mean   :126.9   1bbl   :11   Mean   :3.33   Mean   :3.255
##  three :  1     3rd Qu.:141.0   spdi   : 9   3rd Qu.:3.58   3rd Qu.:3.410
##  twelve:  1     Max.   :326.0   4bbl   : 3   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
##  [ reached getOption("max.print") -- omitted 1 row ]
```

```
###Summarize Numeric Data####
summary(Filter(is.numeric,carprice))
```

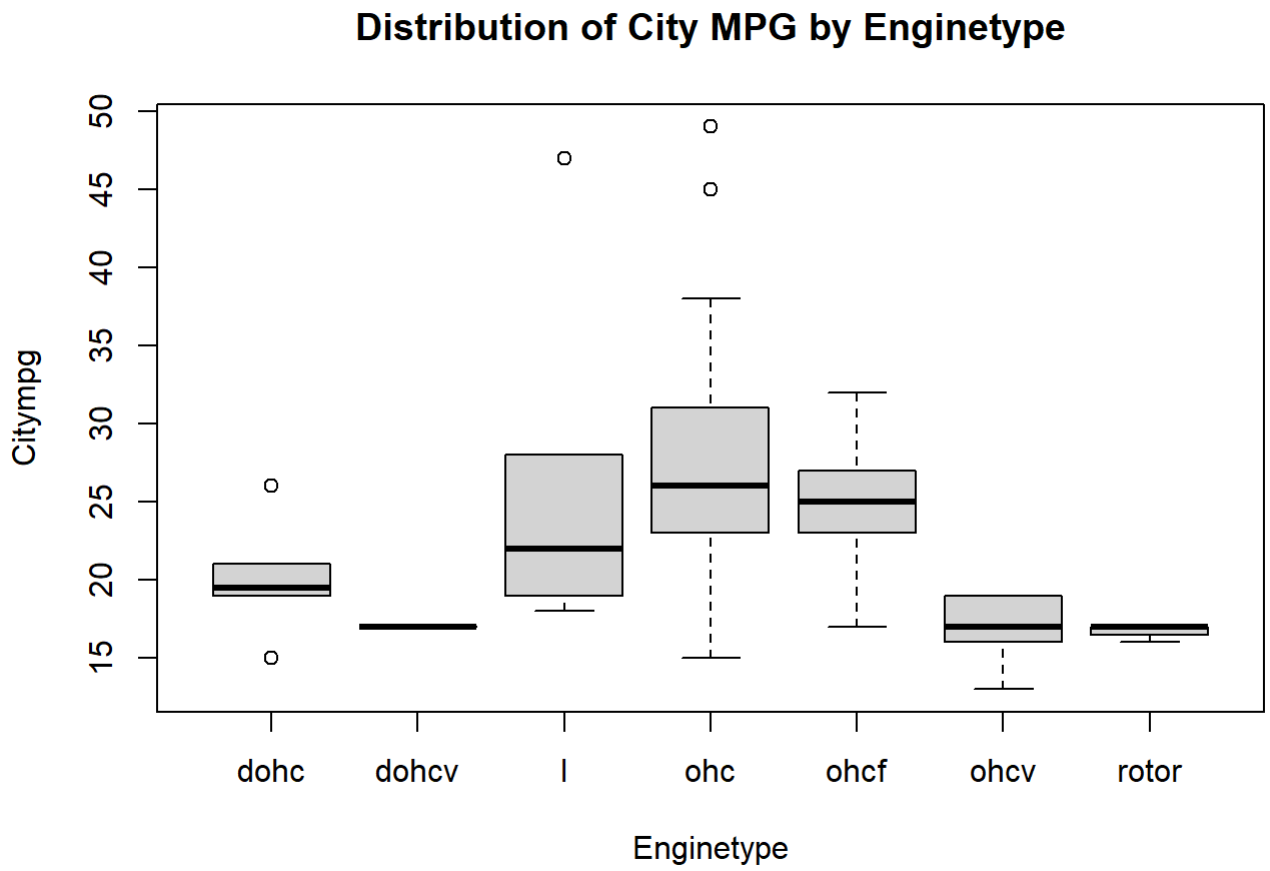
```
##      car_ID      symboling         wheelbase        carlength
##  Min.   :  1   Min.   :-2.0000   Min.   : 86.60   Min.   :141.1
##  1st Qu.: 52   1st Qu.: 0.0000   1st Qu.: 94.50   1st Qu.:166.3
##  Median :103   Median : 1.0000   Median : 97.00   Median :173.2
##  Mean   :103   Mean   : 0.8341   Mean   : 98.76   Mean   :174.0
##  3rd Qu.:154   3rd Qu.: 2.0000   3rd Qu.:102.40   3rd Qu.:183.1
##  Max.   :205   Max.   : 3.0000   Max.   :120.90   Max.   :208.1
##     carwidth       carheight       curbweight     enginesize      boreratio
##  Min.   :60.30   Min.   :47.80   Min.   :1488   Min.   : 61.0   Min.   :2.54
##  1st Qu.:64.10   1st Qu.:52.00   1st Qu.:2145   1st Qu.: 97.0   1st Qu.:3.15
##  Median :65.50   Median :54.10   Median :2414   Median :120.0   Median :3.31
##  Mean   :65.91   Mean   :53.72   Mean   :2556   Mean   :126.9   Mean   :3.33
##  3rd Qu.:66.90   3rd Qu.:55.50   3rd Qu.:2935   3rd Qu.:141.0   3rd Qu.:3.58
##  Max.   :72.30   Max.   :59.80   Max.   :4066   Max.   :326.0   Max.   :3.94
##      stroke      compressionratio   horsepower       peakrpm
##  Min.   :2.070   Min.   : 7.00    Min.   : 48.0   Min.   :4150
##  1st Qu.:3.110   1st Qu.: 8.60    1st Qu.: 70.0   1st Qu.:4800
##  Median :3.290   Median : 9.00    Median : 95.0   Median :5200
##  Mean   :3.255   Mean   :10.14    Mean   :104.1   Mean   :5125
##  3rd Qu.:3.410   3rd Qu.: 9.40    3rd Qu.:116.0   3rd Qu.:5500
##  Max.   :4.170   Max.   :23.00    Max.   :288.0   Max.   :6600
##     citympg        highwaympg        price
##  Min.   :13.00   Min.   :16.00   Min.   : 5118
##  1st Qu.:19.00   1st Qu.:25.00   1st Qu.: 7788
##  Median :24.00   Median :30.00   Median :10295
##  Mean   :25.22   Mean   :30.75   Mean   :13277
##  3rd Qu.:30.00   3rd Qu.:34.00   3rd Qu.:16503
##  Max.   :49.00   Max.   :54.00   Max.   :45400
```

Box Plots

```
# boxplot 1
boxplot(carprice$carbody ~ carprice$drivewheel, color = "blue", xlab = "Drivewheel",
        ylab = "Carbody", main = "Distribution of Carbody by Drivewheel")
```

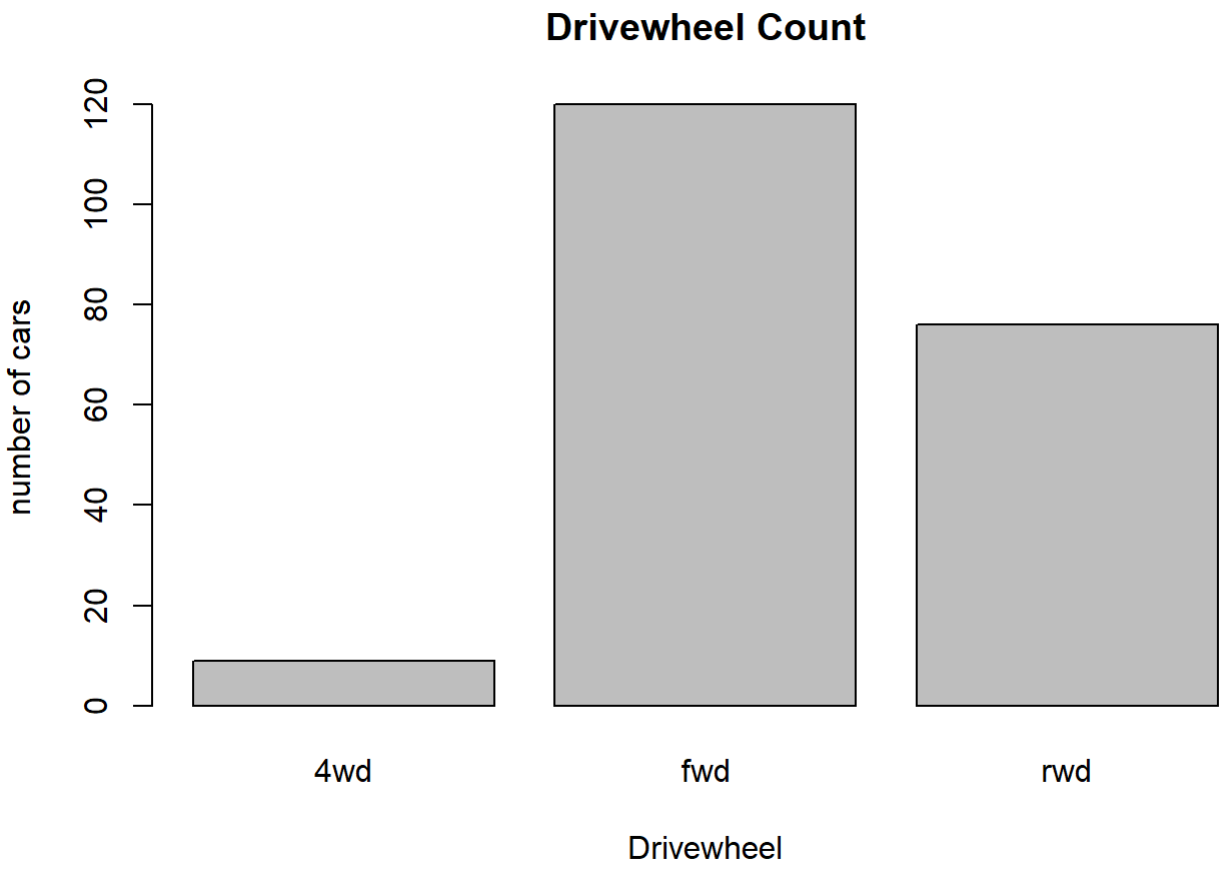



```
# boxplot 2
boxplot(carprice$citympg ~ carprice$enginetype, # numeric.var ~ categorical.var
        xlab = "Enginetype", ylab = "Citympg",
        main = "Distribution of City MPG by Enginetype")
```

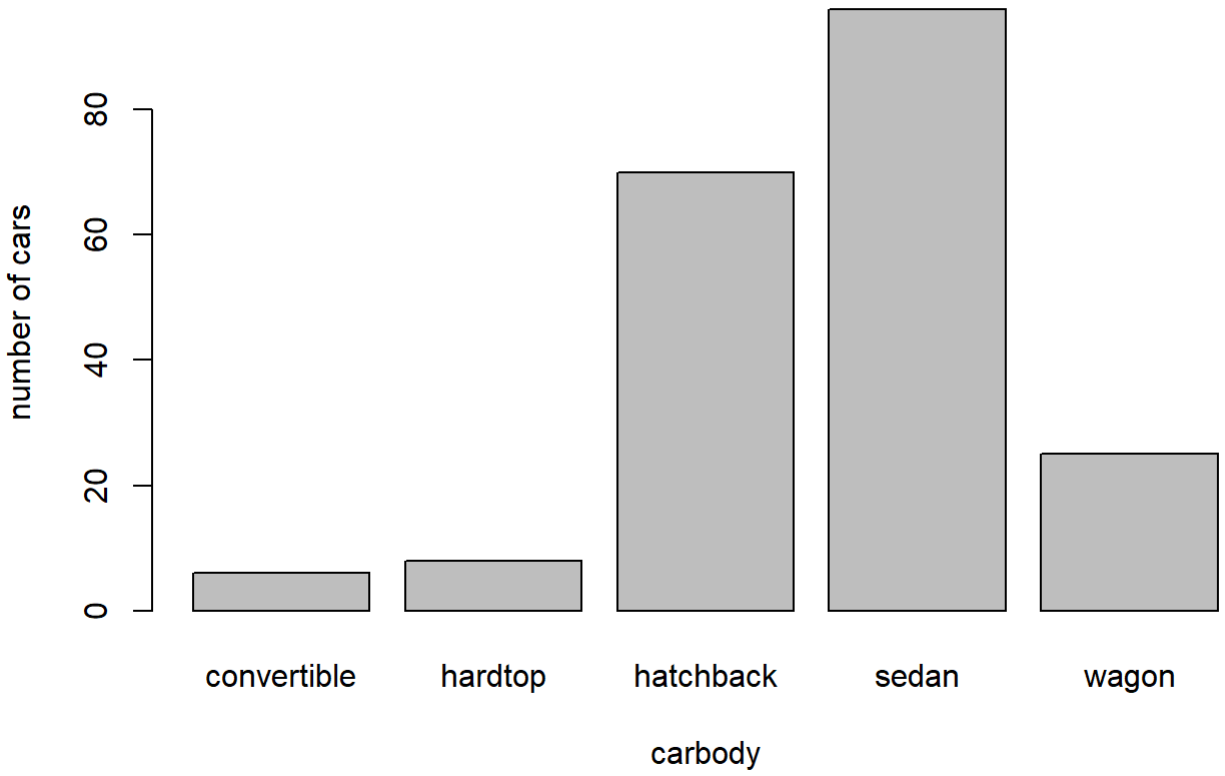



Bar Charts

```
#1: Drivewheel
data.for.plot <- aggregate(carprice$price,                # variable for aggregation
                           by = list(carprice$drivewheel),
                           FUN = length,
                           drop = FALSE)

# rename the variables in data.for.plot for DRIVEWHEEL
names(data.for.plot) <- c("drivewheel","CountofDrivewheels")

# create the bar chart
barplot(height = data.for.plot$CountofDrivewheel, # value for bar height
        names.arg = data.for.plot$drivewheel,       # label the bars
        xlab = "Drivewheel", ylab = "number of cars",  # axis titles
        main = "Drivewheel Count")             # chart title
```




```
##2: Car Body
data.for.plot <- aggregate(carprice$price,                # variable for aggregation
                           by = list(carprice$carbody),
                           FUN = length,
                           drop = FALSE)

# rename the variables in data.for.plot
names(data.for.plot) <- c("CarBody","CountofCarBody")

# create the bar chart
barplot(height = data.for.plot$CountofCarBody, # value for bar height
        names.arg = data.for.plot$CarBody,       # label the bars
        xlab = "carbody", ylab = "number of cars",  # axis titles
        main = "Car Body Count")             # chart title
```



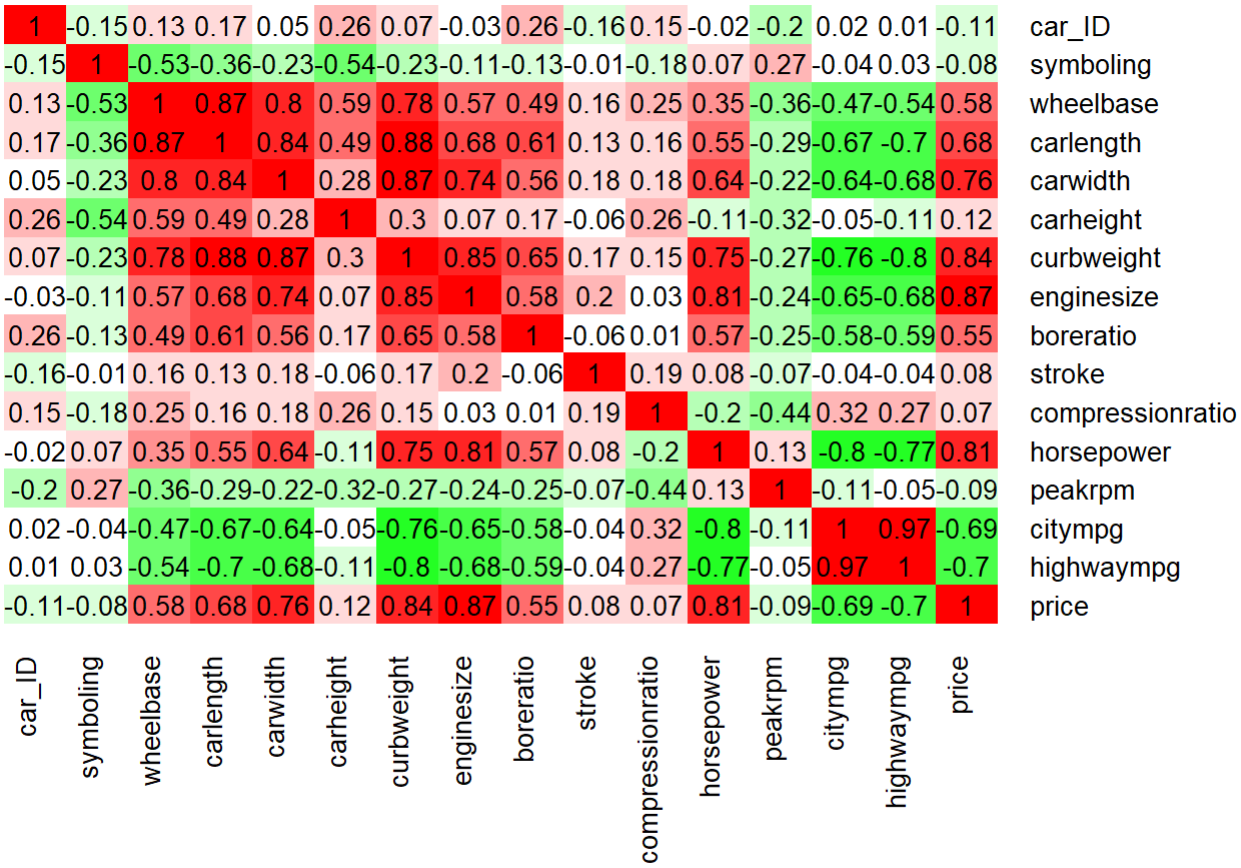

Check for Multicollinearity

```
library(gplots)
```

```
##
## Attaching package: 'gplots'
```

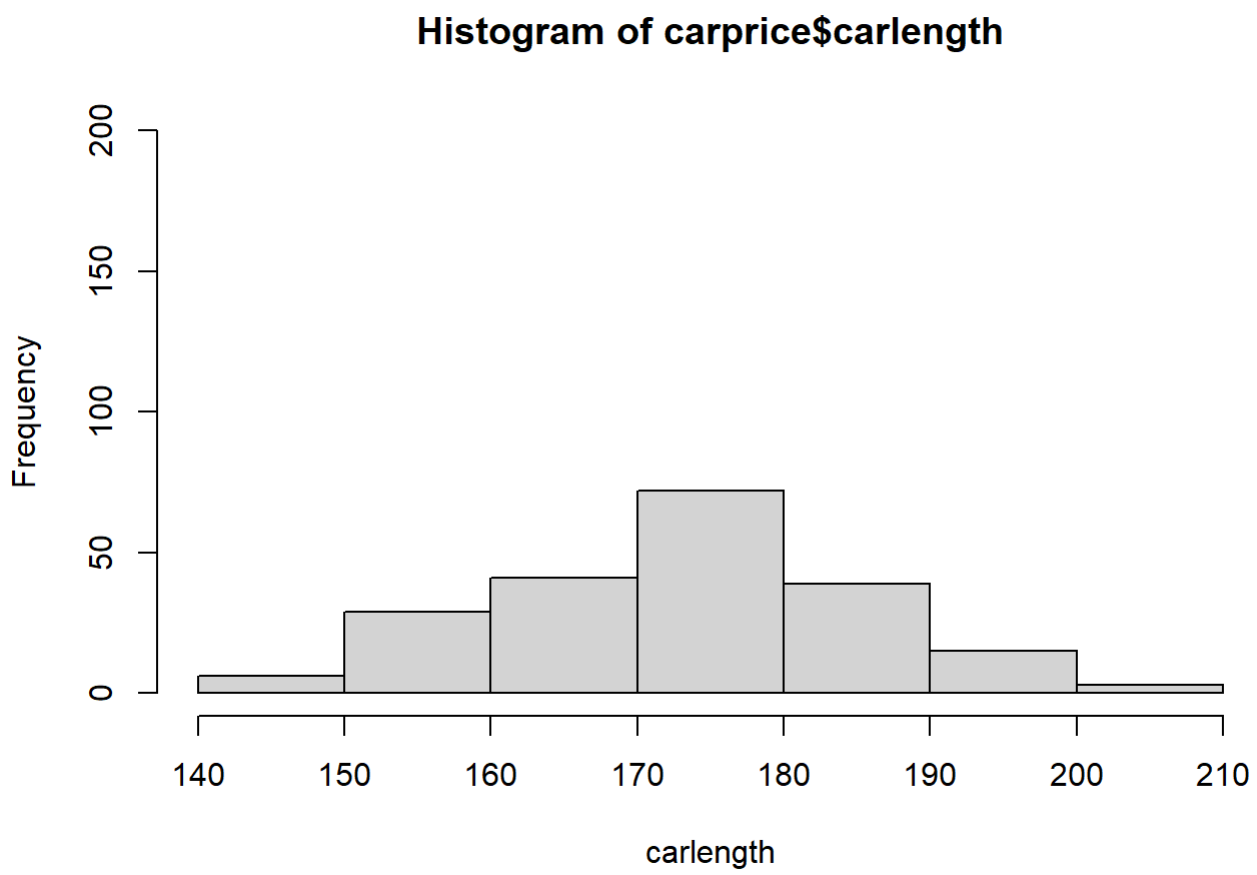
```
## The following object is masked from 'package:stats':
##
##     lowess
```

```
colfunc <- colorRampPalette(c("green","white","red"))
heatmap.2(cor(Filter(is.numeric,carprice), use = "complete.obs"), Rowv = FALSE, Colv = FALSE,
          dendrogram = "none", lwid=c(0.1,4), lhei=c(0.1,4), col = colfunc(15),
          cellnote = round(cor(Filter(is.numeric, carprice), use = "complete.obs"),2),
          notecol = "black", key = FALSE, trace = 'none', margins = c(10,10))
```

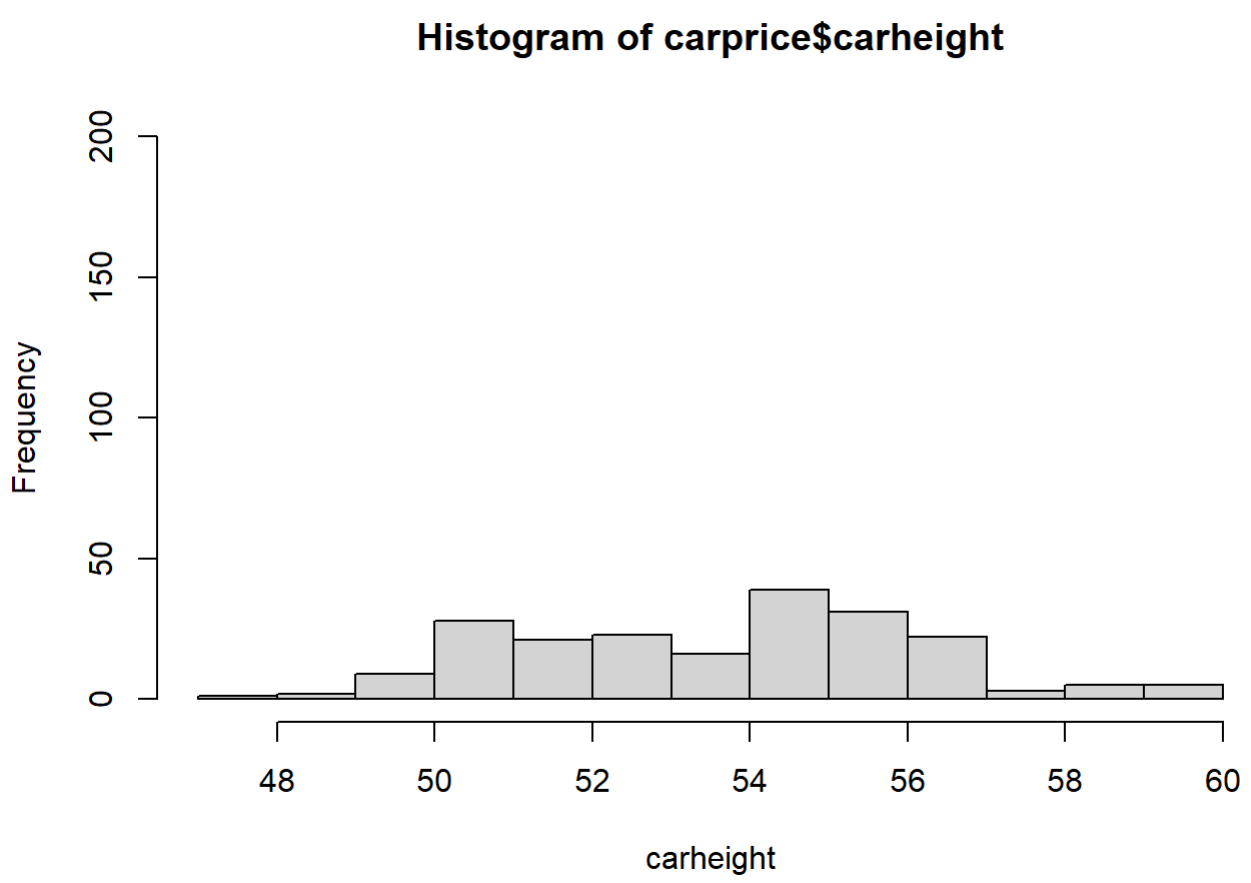



Histograms

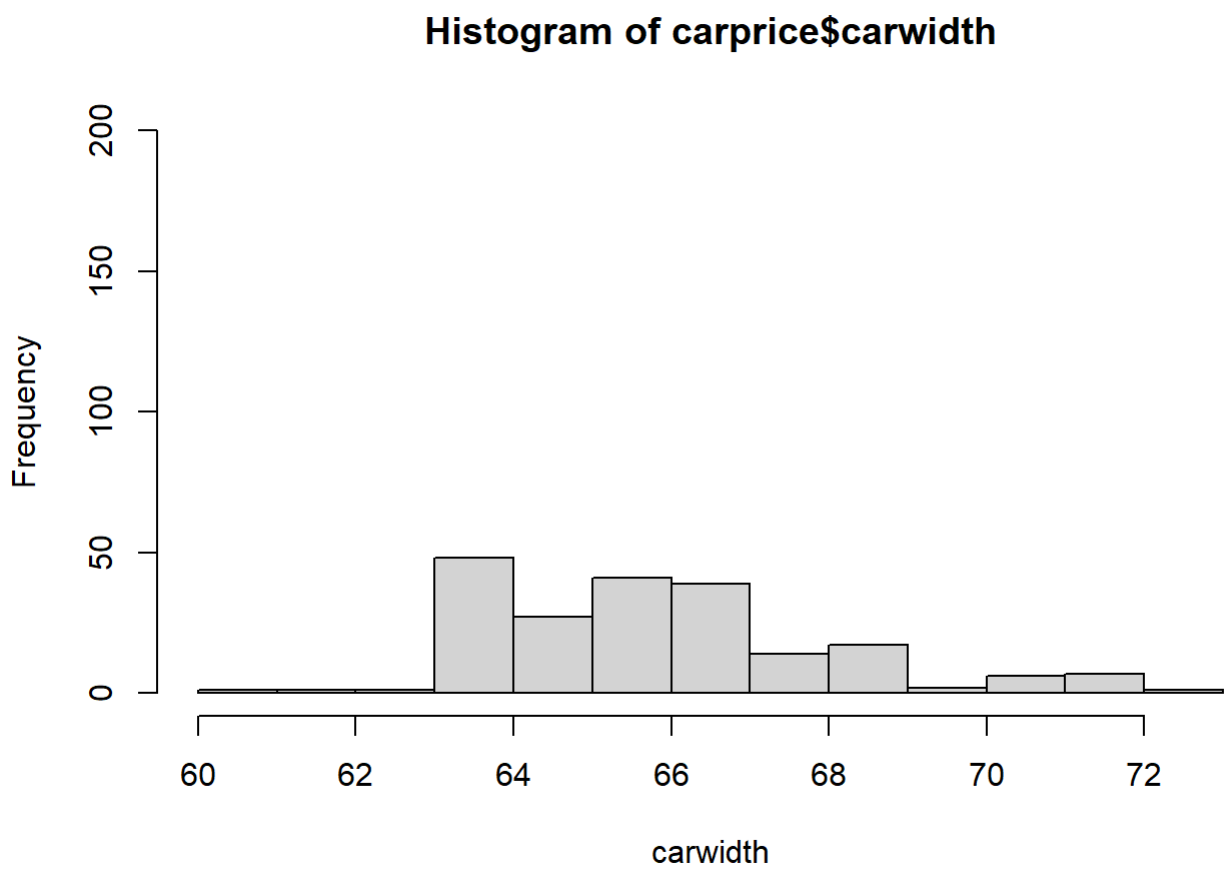
```
####### histogram of carlength#######
hist(carprice$carlength,  # the numeric variable of interest
     ylim = c(0, 200), # defines the y-axis range
     xlab = "carlength")
```



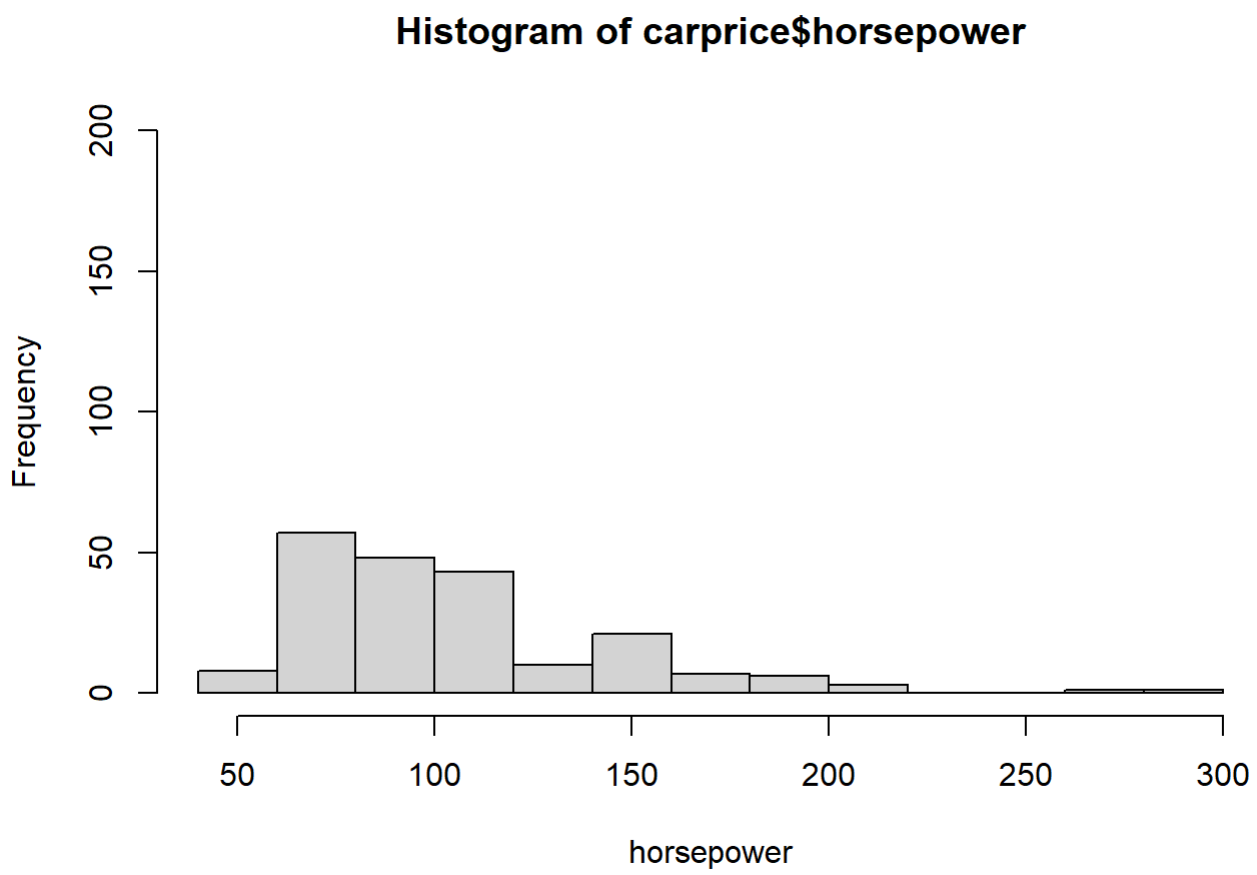

```
####### histogram of carheight#######
hist(carprice$carheight,  # the numeric variable of interest
     ylim = c(0, 200), # defines the y-axis range
     xlab = "carheight")
```




```
####### histogram of carwidth#######
hist(carprice$carwidth,  # the numeric variable of interest
     ylim = c(0, 200), # defines the y-axis range
     xlab = "carwidth")
```

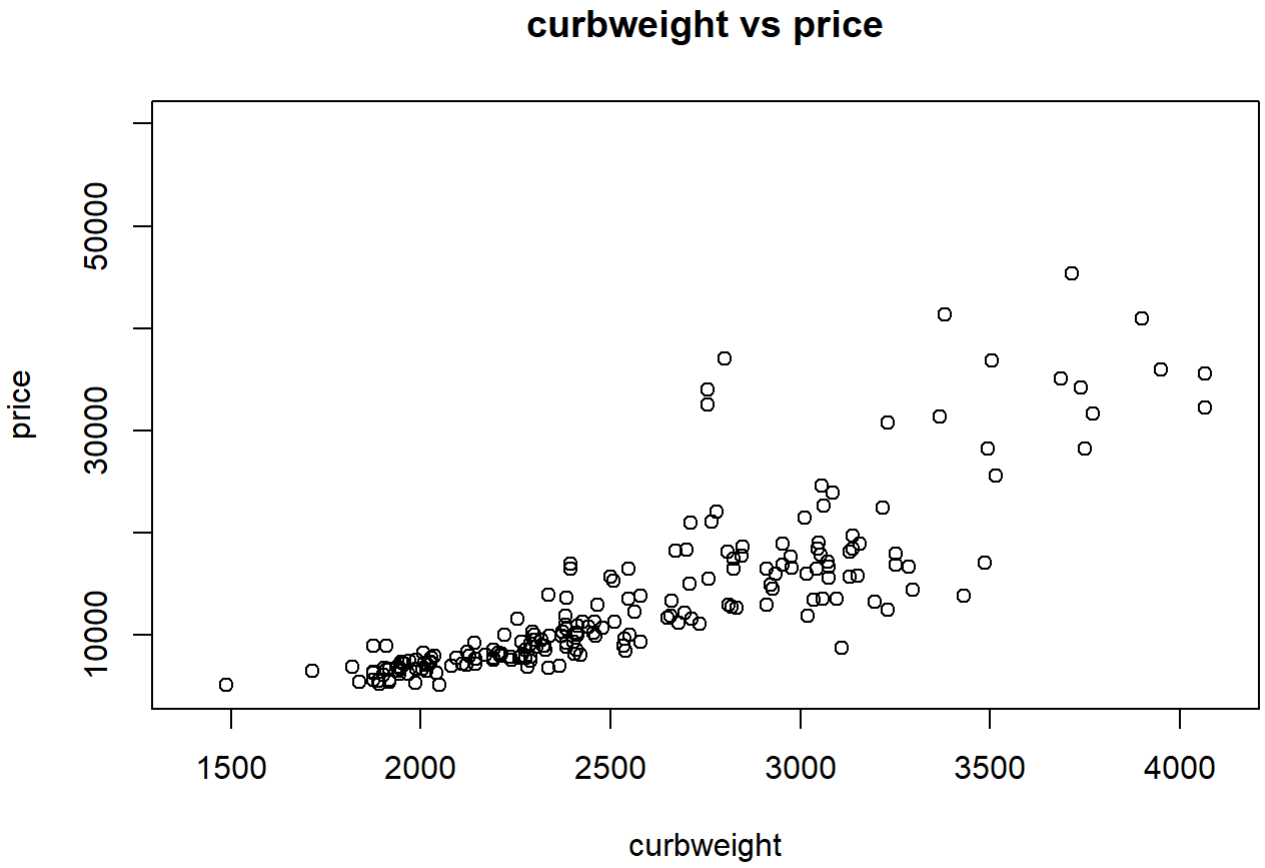



```
####### histogram of horsepower#######
hist(carprice$horsepower,  # the numeric variable of interest
     ylim = c(0, 200), # defines the y-axis range
     xlab = "horsepower")
```

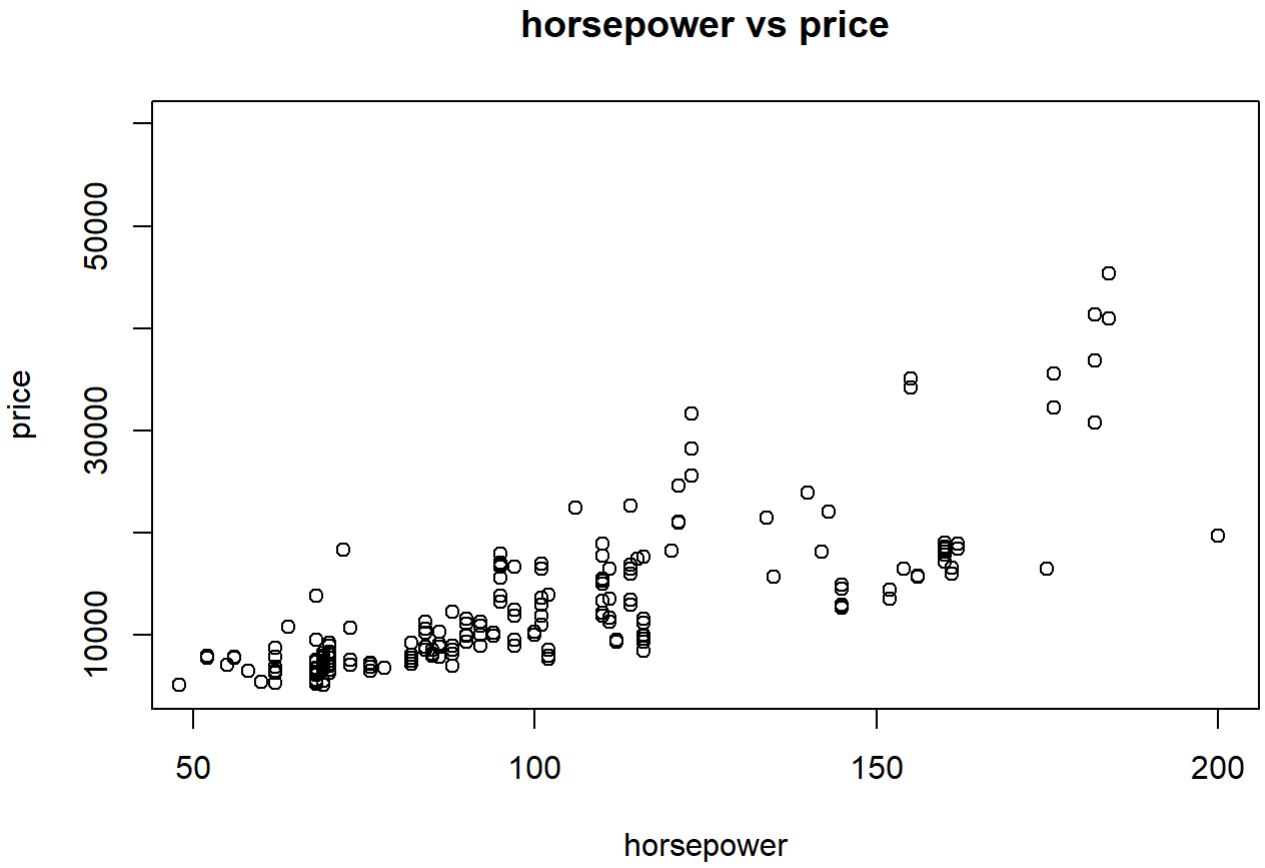



Scatter Plots

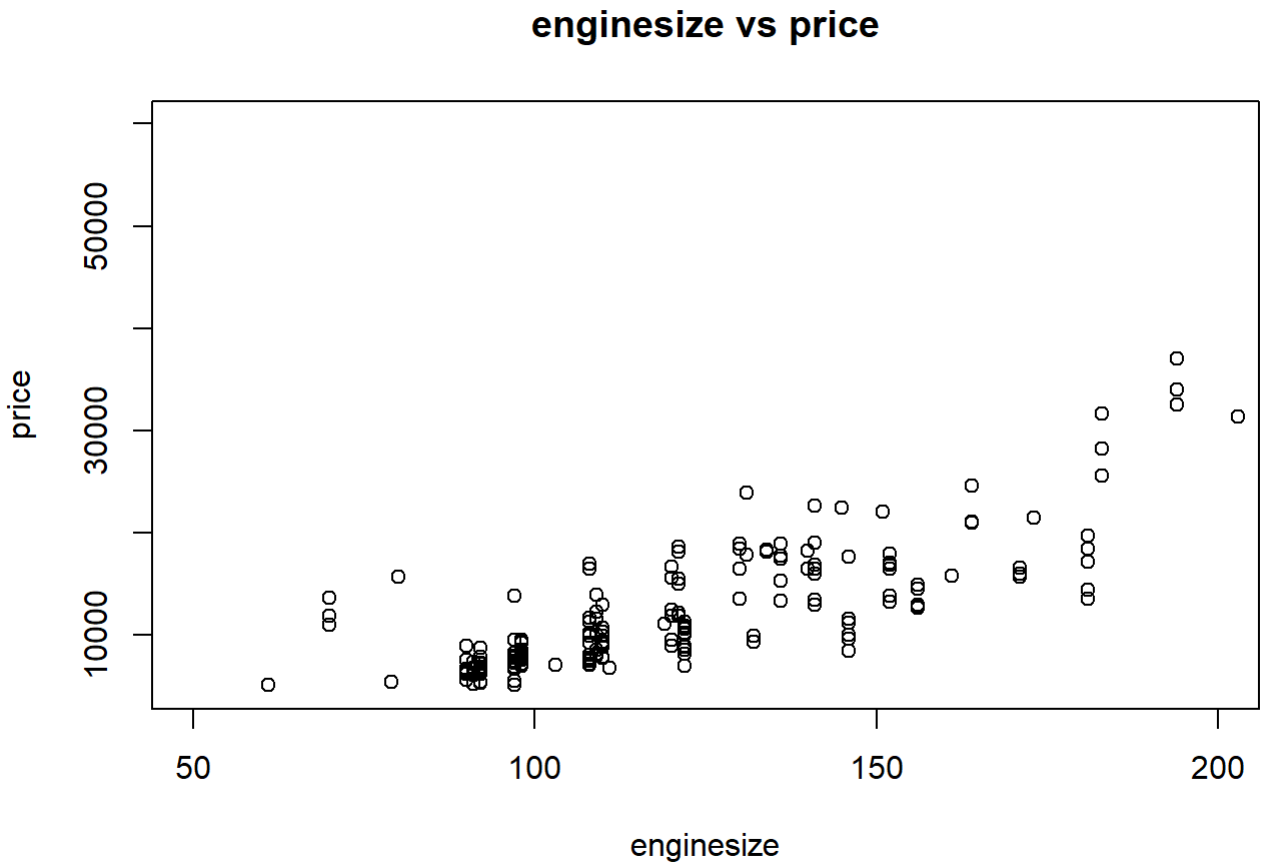
```
####### scatter plot of curbweight vs price########
input <- carprice[,c('curbweight','price')]
plot(x = input$curbweight,y = input$price,
     xlab = "curbweight",
     ylab = "price",
     xlim = c(1400,4100),
     ylim = c(5000,60000),
     main = "curbweight vs price"
)
```



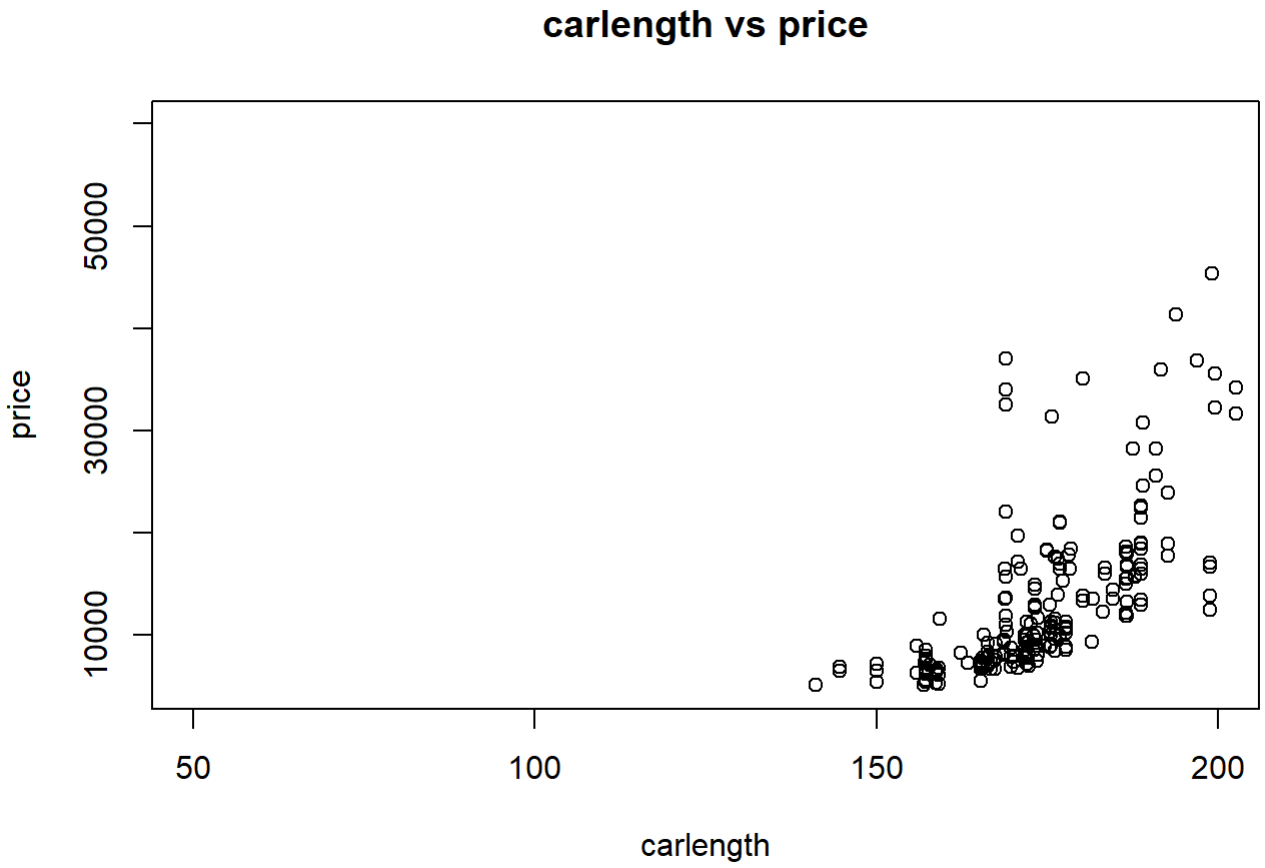

```
####### scatter plot of horsepower vs price#######
input2 <- carprice[,c('horsepower','price')]
plot(x = input2$horsepower,y = input2$price,
     xlab = "horsepower",
     ylab = "price",
     xlim = c(50,200),
     ylim = c(5000,60000),
     main = "horsepower vs price"
)
```




```
####### scatter plot of enginesize vs price#######
input3 <- carprice[,c('enginesize','price')]
plot(x = input3$enginesize,y = input3$price,
     xlab = "enginesize",
     ylab = "price",
     xlim = c(50,200),
     ylim = c(5000,60000),
     main = "enginesize vs price"
)
```




```
####### scatter plot of carlength vs price#######
input4 <- carprice[,c('carlength','price')]
plot(x = input4$carlength,y = input4$price,
     xlab = "carlength",
     ylab = "price",
     xlim = c(50,200),
     ylim = c(5000,60000),
     main = "carlength vs price"
)
```




Regression Tree

```
summary(carprice)
```

```
##      car_ID      symboling                CarName      fueltype   aspiration
##  Min.   :  1   Min.   :-2.0000   peugeot 504   :  6   diesel: 20   std  :168
##  1st Qu.: 52   1st Qu.: 0.0000   toyota corolla:  6   gas   :185   turbo: 37
##  Median :103   Median : 1.0000   toyota corona :  6
##  Mean   :103   Mean   : 0.8341   subaru dl     :  4
##  3rd Qu.:154   3rd Qu.: 2.0000   honda civic   :  3
##  Max.   :205   Max.   : 3.0000   mazda 626     :  3
##  doornumber        carbody   drivewheel enginelocation   wheelbase
##  four:115   convertible: 6   4wd:  9    front:202      Min.   : 86.60
##  two : 90   hardtop    : 8   fwd:120    rear :  3      1st Qu.: 94.50
##             hatchback  :70   rwd: 76                   Median : 97.00
##             sedan      :96                             Mean   : 98.76
##             wagon      :25                             3rd Qu.:102.40
##                                                        Max.   :120.90
##    carlength        carwidth       carheight       curbweight   enginetype
##  Min.   :141.1   Min.   :60.30   Min.   :47.80   Min.   :1488   dohc : 12
##  1st Qu.:166.3   1st Qu.:64.10   1st Qu.:52.00   1st Qu.:2145   dohcv:  1
##  Median :173.2   Median :65.50   Median :54.10   Median :2414   l    : 12
##  Mean   :174.0   Mean   :65.91   Mean   :53.72   Mean   :2556   ohc  :148
##  3rd Qu.:183.1   3rd Qu.:66.90   3rd Qu.:55.50   3rd Qu.:2935   ohcf : 15
##  Max.   :208.1   Max.   :72.30   Max.   :59.80   Max.   :4066   ohcv : 13
##  cylindernumber   enginesize      fuelsystem   boreratio        stroke
##  eight :  5     Min.   : 61.0   mpfi   :94   Min.   :2.54   Min.   :2.070
##  five  : 11     1st Qu.: 97.0   2bbl   :66   1st Qu.:3.15   1st Qu.:3.110
##  four  :159     Median :120.0   idi    :20   Median :3.31   Median :3.290
##  six   : 24     Mean   :126.9   1bbl   :11   Mean   :3.33   Mean   :3.255
##  three :  1     3rd Qu.:141.0   spdi   : 9   3rd Qu.:3.58   3rd Qu.:3.410
##  twelve:  1     Max.   :326.0   4bbl   : 3   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
##  [ reached getOption("max.print") -- omitted 1 row ]
```

```
t(t(names(carprice)))
```

```
##       [,1]
##  [1,] "car_ID"
##  [2,] "symboling"
##  [3,] "CarName"
##  [4,] "fueltype"
##  [5,] "aspiration"
##  [6,] "doornumber"
##  [7,] "carbody"
##  [8,] "drivewheel"
##  [9,] "enginelocation"
## [10,] "wheelbase"
## [11,] "carlength"
## [12,] "carwidth"
## [13,] "carheight"
## [14,] "curbweight"
## [15,] "enginetype"
## [16,] "cylindernumber"
## [17,] "enginesize"
## [18,] "fuelsystem"
## [19,] "boreratio"
## [20,] "stroke"
## [21,] "compressionratio"
## [22,] "horsepower"
## [23,] "peakrpm"
## [24,] "citympg"
## [25,] "highwaympg"
## [26,] "price"
```

```
t.carprice <- carprice[-c(1,3)]
summary(t.carprice)
```

```
##    symboling         fueltype   aspiration  doornumber        carbody
##  Min.   :-2.0000   diesel: 20   std  :168   four:115   convertible: 6
##  1st Qu.: 0.0000   gas   :185   turbo: 37   two : 90   hardtop    : 8
##  Median : 1.0000                                       hatchback  :70
##  Mean   : 0.8341                                       sedan      :96
##  3rd Qu.: 2.0000                                       wagon      :25
##  Max.   : 3.0000
##
##  drivewheel enginelocation   wheelbase        carlength        carwidth
##  4wd:  9    front:202      Min.   : 86.60   Min.   :141.1   Min.   :60.30
##  fwd:120    rear :  3      1st Qu.: 94.50   1st Qu.:166.3   1st Qu.:64.10
##  rwd: 76                   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   enginesize
##  Min.   :47.80   Min.   :1488   dohc : 12   eight :  5     Min.   : 61.0
##  1st Qu.:52.00   1st Qu.:2145   dohcv:  1   five  : 11     1st Qu.: 97.0
##  Median :54.10   Median :2414   l    : 12   four  :159     Median :120.0
##  Mean   :53.72   Mean   :2556   ohc  :148   six   : 24     Mean   :126.9
##  3rd Qu.:55.50   3rd Qu.:2935   ohcf : 15   three :  1     3rd Qu.:141.0
##  Max.   :59.80   Max.   :4066   ohcv : 13   twelve:  1     Max.   :326.0
##                                 rotor:  4   two   :  4
##    fuelsystem   boreratio        stroke      compressionratio   horsepower
##  mpfi   :94   Min.   :2.54   Min.   :2.070   Min.   : 7.00    Min.   : 48.0
##  2bbl   :66   1st Qu.:3.15   1st Qu.:3.110   1st Qu.: 8.60    1st Qu.: 70.0
##  idi    :20   Median :3.31   Median :3.290   Median : 9.00    Median : 95.0
##  1bbl   :11   Mean   :3.33   Mean   :3.255   Mean   :10.14    Mean   :104.1
##  spdi   : 9   3rd Qu.:3.58   3rd Qu.:3.410   3rd Qu.: 9.40    3rd Qu.:116.0
##  4bbl   : 3   Max.   :3.94   Max.   :4.170   Max.   :23.00    Max.   :288.0
##  (Other): 2
##     peakrpm        citympg        highwaympg        price
##  Min.   :4150   Min.   :13.00   Min.   :16.00   Min.   : 5118
##  1st Qu.:4800   1st Qu.:19.00   1st Qu.:25.00   1st Qu.: 7788
##  Median :5200   Median :24.00   Median :30.00   Median :10295
##  Mean   :5125   Mean   :25.22   Mean   :30.75   Mean   :13277
##  3rd Qu.:5500   3rd Qu.:30.00   3rd Qu.:34.00   3rd Qu.:16503
##  Max.   :6600   Max.   :49.00   Max.   :54.00   Max.   :45400
##
```

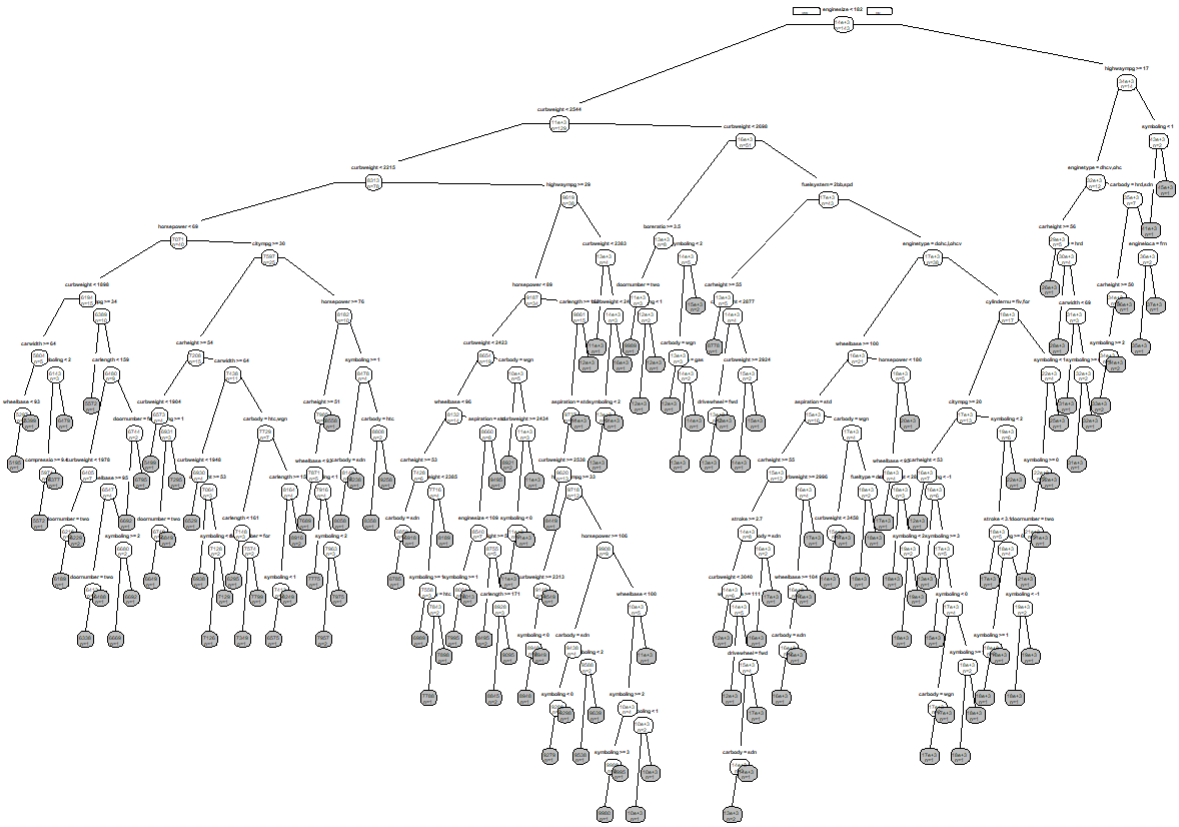
```
# partition the data
set.seed(1)
train.index <- sample(rownames(t.carprice), nrow(t.carprice) * 0.7)
carprice.train <- t.carprice[train.index, ]
valid.index <- setdiff(rownames(t.carprice), train.index)
carprice.valid <- carprice[valid.index, ]

####### full regression tree #######
library(rpart)
library(rpart.plot)
library(caret)
```

```
## Loading required package: ggplot2
```

```
## Loading required package: lattice
```

```
carprice.tree <- rpart(price ~ .,              # quantitative outcome ~ predictors
                       data = carprice.train,    # training data
                       method = "anova",       # regression tree - this is the default
                       cp = 0,                 # complexity parameter
                       minsplit = 2)           # minimum observations required to attempt split

prp(carprice.tree,          # tree model
    type = 1,             # label all the nodes
    extra = 1,            # show # of observations in each node
    varlen = -10,         # truncate variable names to 10 characters
    box.col = ifelse(carprice.tree$frame$var == "<leaf>", 'gray', 'white'))  # make leaves gray,
decision nodes white
```

```
## Warning: labs do not fit even at cex 0.15, there may be some overplotting
```



```
# count number of leaves
length(carprice.tree$frame$var[carprice.tree$frame$var == "<leaf>"])
```

```
## [1] 133
```

```
# assess predictive performance
full.tree.pred <- predict(carprice.tree, carprice.valid)
RMSE(full.tree.pred, carprice.valid$price)
```

```
## [1] 2976.711
```

```
######## best pruned tree #######
cv.carprice.tree <- rpart(price ~ .,
                          data = carprice.train,
                          cp = 0,
                          minsplit = 2,
                          xval = 10)            # 10-fold cross validation
options(scipen = 999, digits = 8)
printcp(cv.carprice.tree)
```

```
##
## Regression tree:
## rpart(formula = price ~ ., data = carprice.train, cp = 0, minsplit = 2,
##     xval = 10)
##
## Variables actually used in tree construction:
##  [1] aspiration       boreratio        carbody          carheight
##  [5] carlength        carwidth         citympg          compressionratio
##  [9] curbweight       cylindernumber   doornumber       drivewheel
## [13] enginelocation   enginesize       enginetype       fuelsystem
## [17] fueltype         highwaympg       horsepower       stroke
## [21] symboling        wheelbase
##
## Root node error: 9457717222/143 = 66137883
##
## n= 143
##
##                       CP nsplit  rel error   xerror      xstd
## 1   0.6871360294994094      0 1.00000000 1.011696 0.1823325
## 2   0.1907471313526041      1 0.31286397 0.317941 0.0341127
## 3   0.0206977832241560      2 0.12211684 0.133398 0.0224087
## 4   0.0133744883908932      3 0.10141906 0.148474 0.0269964
## 5   0.0093412834852147      4 0.08804457 0.142035 0.0263865
## 6   0.0087436296576315      5 0.07870328 0.148771 0.0261935
## 7   0.0081367233966453      6 0.06995965 0.149773 0.0261028
## 8   0.0068232545449018      7 0.06182293 0.148734 0.0246103
## 9   0.0063872941260869      9 0.04817642 0.151233 0.0254019
## 10  0.0037687486944867     10 0.04178913 0.151394 0.0256344
## 11  0.0031140267978450     11 0.03802038 0.150397 0.0274404
## 12  0.0020576514559829     12 0.03490635 0.142824 0.0272132
## 13  0.0020390243857965     13 0.03284870 0.144274 0.0266067
## 14  0.0019820233333174     14 0.03080968 0.145927 0.0272339
## 15  0.0019516012577933     15 0.02882765 0.145927 0.0272339
## 16  0.0015025930591225     16 0.02687605 0.145135 0.0272558
## 17  0.0014127834612213     17 0.02537346 0.147124 0.0275464
## 18  0.0013696129833124     19 0.02254789 0.148621 0.0275395
## 19  0.0013408627616385     20 0.02117828 0.148621 0.0275395
## 20  0.0011118920073899     21 0.01983742 0.148255 0.0276525
## 21  0.0010421965225267     22 0.01872552 0.137478 0.0244755
## 22  0.0008530429324330     23 0.01768333 0.140449 0.0264155
## 23  0.0007785529171947     25 0.01597724 0.140909 0.0270401
## 24  0.0007510538241553     26 0.01519869 0.140054 0.0265310
## 25  0.0007376912246473     27 0.01444763 0.140316 0.0265338
## 26  0.0007051708243462     28 0.01370994 0.140876 0.0265366
## 27  0.0006027078634398     29 0.01300477 0.139688 0.0265301
## 28  0.0005866792692381     30 0.01240207 0.140028 0.0265535
## 29  0.0005501606679046     31 0.01181539 0.140490 0.0265531
## 30  0.0005464439651334     32 0.01126523 0.140278 0.0265604
## 31  0.0005033015072779     33 0.01071878 0.134423 0.0246339
## 32  0.0004916975936955     34 0.01021548 0.134010 0.0246343
## 33  0.0004735904911535     35 0.00972378 0.134768 0.0246816
## 34  0.0004487307402304     38 0.00830301 0.135112 0.0246806
```

```
## 35  0.0004362068283937     39 0.00785428 0.137052 0.0252188
##  [ reached getOption("max.print") -- omitted 84 rows ]
```

```
minerror <- min(cv.carprice.tree$cptable[ ,4 ])                                    # find mini
mum xerror
minerrorstd <- cv.carprice.tree$cptable[cv.carprice.tree$cptable[,4] == minerror, 5]   # and its
corresponding xstd
minerror
```

```
## [1] 0.13339795
```

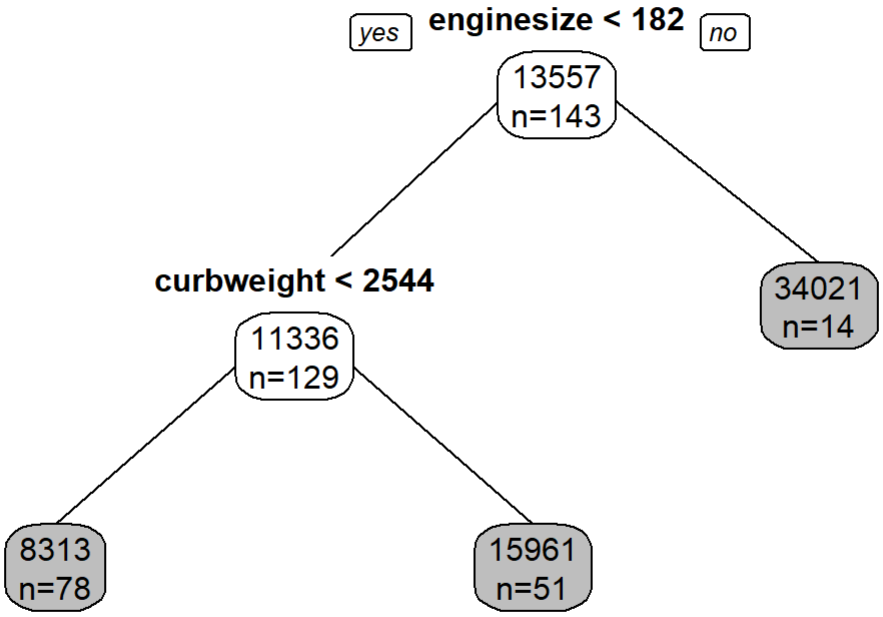
```
minerrorstd
```

```
## [1] 0.022408726
```

```
cv.carprice.tree$cptable[cv.carprice.tree$cptable[,4] == minerror, ]
```

```
##          CP      nsplit   rel error      xerror        xstd
## 0.020697783 2.000000000 0.122116839 0.133397955 0.022408726
```

```
# get list of trees where xerror is less than minerror + minerrorstd
simplertrees <- cv.carprice.tree$cptable[cv.carprice.tree$cptable[,4] < minerror + minerrorstd,
]

# use the cp from the simplest of those trees
bestcp <- simplertrees[1, 1]
carprice.pruned <- prune(cv.carprice.tree, cp = bestcp)
prp(carprice.pruned, type = 1, extra = 1, varlen = -10, digits = -3,
    box.col = ifelse(carprice.pruned$frame$var == "<leaf>", 'gray', 'white'))
```



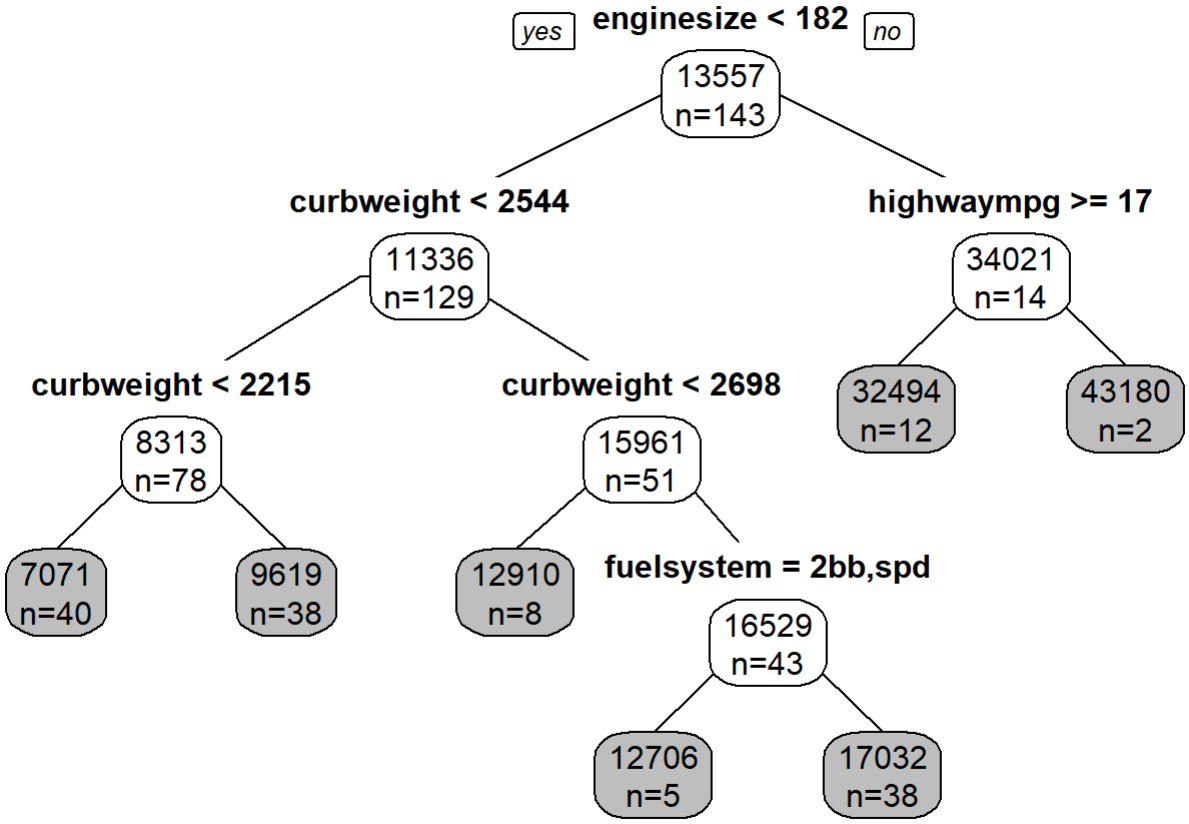

```
length(carprice.pruned$frame$var[carprice.pruned$frame$var == "<leaf>"])
```

```
## [1] 3
```

```
pruned.valid.rt.pred <- predict(carprice.pruned, newdata = carprice.valid)
RMSE(pruned.valid.rt.pred, carprice.valid$price)
```

```
## [1] 3397.3841
```

```
# trying different cps from different trees
carprice.pruned2 <- prune(cv.carprice.tree, cp =  0.00874362965763152)
prp(carprice.pruned2, type = 1, extra = 1, varlen = -10, digits = -3,
    box.col = ifelse(carprice.pruned2$frame$var == "<leaf>", 'gray', 'white'))
```




```
length(carprice.pruned2$frame$var[carprice.pruned$frame$var == "<leaf>"])
```

```
## [1] 7
```

```
pruned.valid.rt.pred2 <- predict(carprice.pruned2, newdata = carprice.valid)
RMSE(pruned.valid.rt.pred2, carprice.valid$price) #smallest rmse
```

```
## [1] 2936.4774
```

Variable Selection Models

```
# Prep work
t2.carprice <- t.carprice[-c(14)]
t(t(names(t2.carprice)))
```

```
##       [,1]
##  [1,] "symboling"
##  [2,] "fueltype"
##  [3,] "aspiration"
##  [4,] "doornumber"
##  [5,] "carbody"
##  [6,] "drivewheel"
##  [7,] "enginelocation"
##  [8,] "wheelbase"
##  [9,] "carlength"
## [10,] "carwidth"
## [11,] "carheight"
## [12,] "curbweight"
## [13,] "enginetype"
## [14,] "enginesize"
## [15,] "fuelsystem"
## [16,] "boreratio"
## [17,] "stroke"
## [18,] "compressionratio"
## [19,] "horsepower"
## [20,] "peakrpm"
## [21,] "citympg"
## [22,] "highwaympg"
## [23,] "price"
```

```
set.seed(1)
train.rows <- sample(rownames(t2.carprice), nrow(t2.carprice)*0.6)
train.data <- t2.carprice[train.rows, ]
valid.rows <- setdiff(rownames(t2.carprice), train.rows)
valid.data <- t2.carprice[valid.rows, ]
```

Exhaustive Search Method

```
library(leaps)
# use regsubsets() to perform exhaustive search
carprice.search <- regsubsets(price ~ .,                  # full model formula
                              data = train.data,          # training dataset
                              nbest = 1,                  # number of subsets of each size
                              nvmax = ncol(train.data),   # maximum number of variables to consid
er
                              method = "exhaustive")      # specify exhaustive search
```

```
## Warning in leaps.setup(x, y, wt = wt, nbest = nbest, nvmax = nvmax, force.in =
## force.in, : 3 linear dependencies found
```

```
## Reordering variables and trying again:
```

```
search.summary <- summary(carprice.search)      # define summary for easy reference
search.summary$which                           # show which variables are included in each best s
ubset
```

```
##    (Intercept) symboling fueltypegas aspirationturbo doornumbertwo
## 1         TRUE     FALSE       FALSE           FALSE         FALSE
## 2         TRUE     FALSE       FALSE           FALSE         FALSE
## 3         TRUE     FALSE       FALSE           FALSE         FALSE
## 4         TRUE     FALSE       FALSE           FALSE         FALSE
##    carbodyhardtop carbodyhatchback carbodysedan carbodywagon drivewheelfwd
## 1           FALSE            FALSE        FALSE        FALSE         FALSE
## 2           FALSE            FALSE        FALSE        FALSE         FALSE
## 3           FALSE            FALSE        FALSE        FALSE         FALSE
## 4           FALSE            FALSE        FALSE        FALSE         FALSE
##    drivewheelrwd enginelocationrear wheelbase carlength carwidth carheight
## 1          FALSE              FALSE     FALSE     FALSE    FALSE     FALSE
## 2          FALSE               TRUE     FALSE     FALSE    FALSE     FALSE
## 3          FALSE               TRUE     FALSE     FALSE     TRUE     FALSE
## 4          FALSE               TRUE     FALSE     FALSE     TRUE     FALSE
##    curbweight enginetypedohcv enginetypel enginetypeohc enginetypeohcf
## 1       FALSE           FALSE       FALSE         FALSE          FALSE
## 2       FALSE           FALSE       FALSE         FALSE          FALSE
## 3       FALSE           FALSE       FALSE         FALSE          FALSE
## 4       FALSE           FALSE       FALSE         FALSE          FALSE
##    enginetypeohcv enginetyperotor enginesize fuelsystem2bbl fuelsystem4bbl
## 1           FALSE           FALSE       TRUE          FALSE          FALSE
## 2           FALSE           FALSE       TRUE          FALSE          FALSE
## 3           FALSE           FALSE       TRUE          FALSE          FALSE
## 4           FALSE            TRUE       TRUE          FALSE          FALSE
##    fuelsystemidi fuelsystemmfi fuelsystemmpfi fuelsystemspdi fuelsystemspfi
## 1          FALSE         FALSE          FALSE          FALSE          FALSE
## 2          FALSE         FALSE          FALSE          FALSE          FALSE
## 3          FALSE         FALSE          FALSE          FALSE          FALSE
## 4          FALSE         FALSE          FALSE          FALSE          FALSE
##    boreratio stroke compressionratio horsepower peakrpm citympg highwaympg
## 1      FALSE  FALSE            FALSE      FALSE   FALSE   FALSE      FALSE
## 2      FALSE  FALSE            FALSE      FALSE   FALSE   FALSE      FALSE
## 3      FALSE  FALSE            FALSE      FALSE   FALSE   FALSE      FALSE
## 4      FALSE  FALSE            FALSE      FALSE   FALSE   FALSE      FALSE
##  [ reached getOption("max.print") -- omitted 20 rows ]
```

```
# compare the models returned by the exhaustive search
options(digits = 8)
t(t(search.summary$rsq))          # shows r-squared values
```

```
##             [,1]
##  [1,] 0.77242710
##  [2,] 0.81183507
##  [3,] 0.86613439
##  [4,] 0.88234978
##  [5,] 0.89223332
##  [6,] 0.90159355
##  [7,] 0.91143096
##  [8,] 0.91813284
##  [9,] 0.92284838
## [10,] 0.92977196
## [11,] 0.93341516
## [12,] 0.93509323
## [13,] 0.93792226
## [14,] 0.93971906
## [15,] 0.94069171
## [16,] 0.94229543
## [17,] 0.94300484
## [18,] 0.94402066
## [19,] 0.94444213
## [20,] 0.94484742
## [21,] 0.94509774
## [22,] 0.94521442
## [23,] 0.94537537
## [24,] 0.94559729
```

```
t(t(search.summary$adjr2))        # shows adjusted r-squared values
```

```
##             [,1]
##  [1,] 0.77054634
##  [2,] 0.80869899
##  [3,] 0.86275963
##  [4,] 0.87836164
##  [5,] 0.88762791
##  [6,] 0.89650356
##  [7,] 0.90603980
##  [8,] 0.91238777
##  [9,] 0.91670357
## [10,] 0.92350160
## [11,] 0.92681666
## [12,] 0.92801249
## [13,] 0.93051850
## [14,] 0.93190487
## [15,] 0.93237747
## [16,] 0.93358530
## [17,] 0.93377705
## [18,] 0.93433192
## [19,] 0.93419360
## [20,] 0.93403319
## [21,] 0.93368241
## [22,] 0.93316160
## [23,] 0.93268480
## [24,] 0.93227418
```

```
t(t(search.summary$cp))          # shows Mallow's Cp values
```

```
##              [,1]
##  [1,] 244.0404837
##  [2,] 183.1740891
##  [3,]  98.5519453
##  [4,]  74.6840057
##  [5,]  60.9170769
##  [6,]  47.9849757
##  [7,]  34.2916447
##  [8,]  25.6003248
##  [9,]  20.0777478
## [10,]  11.0327696
## [11,]   7.2208788
## [12,]   6.5438991
## [13,]   4.0308257
## [14,]   3.1644390
## [15,]   3.6127979
## [16,]   3.0544418
## [17,]   3.9227340
## [18,]   4.3022333
## [19,]   5.6298632
## [20,]   6.9833217
## [21,]   8.5839999
## [22,]  10.3978500
## [23,]  12.1410918
## [24,]  13.7870750
```

```
t(t(search.summary$bic))          # shows BIC values
```

```
##             [,1]
##  [1,] -172.45065
##  [2,] -191.02713
##  [3,] -228.09429
##  [4,] -239.16391
##  [5,] -245.14467
##  [6,] -251.50853
##  [7,] -259.65119
##  [8,] -264.51719
##  [9,] -267.00204
## [10,] -273.75491
## [11,] -275.49502
## [12,] -273.82242
## [13,] -274.49166
## [14,] -273.29218
## [15,] -270.48082
## [16,] -269.04039
## [17,] -265.74972
## [18,] -263.14952
## [19,] -259.26692
## [20,] -255.35529
## [21,] -251.10263
## [22,] -246.55215
## [23,] -242.10184
## [24,] -237.79037
```

```
carprice.best1 <- lm(price ~ enginesize,   # remove 3 predictors not in 8 variable model
                     data = train.data)
options(scipen = 999)
carprice.best1summary <- summary(carprice.best1)
carprice.best1summary
```

```
##
## Call:
## lm(formula = price ~ enginesize, data = train.data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8362.9 -2156.6  -385.0  1503.5 13048.6
##
## Coefficients:
##               Estimate Std. Error t value             Pr(>|t|)
## (Intercept) -7621.2039  1068.5776 -7.1321      0.00000000007859 ***
## enginesize    162.8898     8.0377 20.2657 < 0.0000000000000022 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3679.9 on 121 degrees of freedom
## Multiple R-squared:  0.77243,    Adjusted R-squared:  0.77055
## F-statistic:  410.7 on 1 and 121 DF,  p-value: < 0.000000000000000222
```

Forward Selection Method

```
carprice.lm <- lm(price ~ ., data = train.data)         # full model with all predictors
carprice.lm.null <- lm(price ~ 1, data = train.data)    # intercept-only model

# use step() to run forward selection
carprice.lm.fwd <- step(carprice.lm.null,                                    # initial model
                        scope = list(carprice.lm.null, upper = carprice.lm),    # range of models
                        direction = "forward")                              # forward selection
```

```
## Start:  AIC=2201.87
## price ~ 1
##
##                    Df  Sum of Sq        RSS     AIC
## + enginesize        1 5561412259 1638506332 2021.80
## + curbweight        1 4690880966 2509037624 2074.21
## + horsepower        1 4519164258 2680754332 2082.35
## + carwidth          1 4149121446 3050797144 2098.26
## + highwaympg        1 3448678626 3751239964 2123.68
## + citympg           1 3345191295 3854727295 2127.03
## + carlength         1 2839250295 4360668295 2142.20
## + drivewheel        2 2846712381 4353206209 2143.99
## + boreratio         1 2715463920 4484454670 2145.64
## + fuelsystem        5 2541627437 4658291154 2158.32
## + enginetype        6 2244805091 4955113499 2167.91
## + wheelbase         1 1755202198 5444716392 2169.50
## + carbody           4 1861869651 5338048939 2173.07
## + enginelocation    1 1035114238 6164804352 2184.78
## + aspiration        1  216061385 6983857206 2200.13
## + peakrpm           1  158472939 7041445652 2201.14
## <none>                           7199918590 2201.87
## + symboling         1   54103139 7145815452 2202.95
## + carheight         1   43994324 7155924266 2203.12
## + doornumber        1   37797285 7162121305 2203.23
## + fueltype          1   37236672 7162681919 2203.24
## + stroke            1    6308330 7193610261 2203.77
## + compressionratio  1    6225538 7193693052 2203.77
##
## Step:  AIC=2021.8
## price ~ enginesize
##
##                    Df Sum of Sq        RSS     AIC
## + enginetype        6 478852929 1159653403 1991.28
## + enginelocation    1 283734157 1354772175 2000.41
## + horsepower        1 221804220 1416702111 2005.91
## + drivewheel        2 216132780 1422373552 2008.40
## + fuelsystem        5 280995640 1357510692 2008.66
## + stroke            1 183987692 1454518639 2009.15
## + carwidth          1 176739031 1461767300 2009.76
## + curbweight        1 131010091 1507496240 2013.55
## + citympg           1 115758767 1522747565 2014.79
## + highwaympg        1  93229258 1545277073 2016.59
## + carbody           4 159915964 1478590368 2017.17
## + peakrpm           1  78426596 1560079735 2017.77
## + aspiration        1  65631261 1572875071 2018.77
## + boreratio         1  34987714 1603518618 2021.14
## + carlength         1  34801601 1603704730 2021.16
## <none>                         1638506332 2021.80
## + compressionratio  1  20594506 1617911826 2022.24
## + symboling         1  16582103 1621924228 2022.55
## + fueltype          1  16393755 1622112576 2022.56
## + doornumber        1  12209023 1626297309 2022.88
```

```
## + wheelbase         1   7523265 1630983067 2023.23
## + carheight         1   2569791 1635936540 2023.61
##
## Step:  AIC=1991.28
## price ~ enginesize + enginetype
##
##                    Df Sum of Sq        RSS     AIC
## + stroke            1 200995601  958657802 1969.87
## + enginelocation    1 148262997 1011390406 1976.46
## + horsepower        1  90871061 1068782342 1983.24
## + carwidth          1  81331691 1078321712 1984.34
## + peakrpm           1  72939237 1086714166 1985.29
## + fuelsystem        5 140635114 1019018289 1985.38
## + carbody           4 118892900 1040760503 1985.98
## + curbweight        1  48213350 1111440053 1988.06
## + aspiration        1  44780220 1114873183 1988.44
## + drivewheel        2  50364632 1109288771 1989.82
## + boreratio         1  24746274 1134907129 1990.63
## <none>                         1159653403 1991.28
## + compressionratio  1  14100425 1145552978 1991.78
## + citympg           1  10150928 1149502475 1992.20
## + fueltype          1   9891507 1149761895 1992.23
## + doornumber        1   9364220 1150289183 1992.29
## + carheight         1   9049534 1150603869 1992.32
## + highwaympg        1   8453974 1151199429 1992.38
## + symboling         1   5136317 1154517085 1992.74
## + carlength         1   3558936 1156094467 1992.90
## + wheelbase         1   1154109 1158499294 1993.16
##
## Step:  AIC=1969.87
## price ~ enginesize + enginetype + stroke
##
##                    Df Sum of Sq       RSS     AIC
## + enginelocation    1 137338628 821319174 1952.85
## + aspiration        1 111643600 847014201 1956.64
## + peakrpm           1  90619482 868038320 1959.66
## + horsepower        1  88916556 869741246 1959.90
## + carwidth          1  65535584 893122218 1963.16
## + fuelsystem        5 101837318 856820483 1966.06
## + fueltype          1  37239304 921418498 1967.00
## + compressionratio  1  35297462 923360340 1967.26
## + carbody           4  77657837 880999964 1967.48
## + curbweight        1  32759391 925898411 1967.59
## + boreratio         1  16736853 941920949 1969.70
## <none>                         958657802 1969.87
## + doornumber        1   8538695 950119107 1970.77
## + wheelbase         1   5943356 952714445 1971.11
## + symboling         1   5507584 953150217 1971.16
## + highwaympg        1   3022253 955635549 1971.48
## + citympg           1   1715236 956942565 1971.65
## + carheight         1    555399 958102403 1971.80
## + carlength         1    532082 958125720 1971.80
```

```
## + drivewheel        2   8490015 950167786 1972.78
##
## Step:  AIC=1952.85
## price ~ enginesize + enginetype + stroke + enginelocation
##
##                    Df Sum of Sq       RSS     AIC
## + carwidth          1 183013500 638305674 1923.84
## + aspiration        1 160981718 660337455 1928.02
## + curbweight        1  96107972 725211202 1939.54
## + fuelsystem        5 103727245 717591929 1946.24
## + fueltype          1  43780191 777538982 1948.11
## + horsepower        1  39183186 782135988 1948.84
## + compressionratio  1  34722254 786596919 1949.54
## + peakrpm           1  23191618 798127555 1951.33
## + wheelbase         1  18677970 802641203 1952.02
## + carlength         1  15571773 805747401 1952.50
## + highwaympg        1  14693410 806625763 1952.63
## <none>                          821319174 1952.85
## + carheight         1   7506926 813812247 1953.72
## + drivewheel        2  19628984 801690189 1953.88
## + citympg           1   3802024 817517149 1954.28
## + boreratio         1   3256578 818062595 1954.36
## + symboling         1   1296051 820023122 1954.66
## + doornumber        1     27353 821291821 1954.85
## + carbody           4  26133303 795185870 1956.87
##
## Step:  AIC=1923.84
## price ~ enginesize + enginetype + stroke + enginelocation + carwidth
##
##                    Df Sum of Sq       RSS     AIC
## + aspiration        1  86720892 551584782 1907.88
## + curbweight        1  14578987 623726686 1923.00
## + carlength         1  13207501 625098173 1923.27
## + boreratio         1  12306603 625999071 1923.45
## <none>                          638305674 1923.84
## + fueltype          1   9365518 628940156 1924.03
## + peakrpm           1   7972501 630333173 1924.30
## + drivewheel        2  17423281 620882393 1924.44
## + horsepower        1   6972726 631332948 1924.49
## + compressionratio  1   5907081 632398593 1924.70
## + doornumber        1   2618591 635687083 1925.34
## + wheelbase         1   2457861 635847813 1925.37
## + carbody           4  31812256 606493418 1925.56
## + highwaympg        1   1213763 637091910 1925.61
## + citympg           1    300648 638005025 1925.79
## + symboling         1     64433 638241241 1925.83
## + carheight         1     44521 638261153 1925.84
## + fuelsystem        5  27229588 611076086 1928.48
##
## Step:  AIC=1907.88
## price ~ enginesize + enginetype + stroke + enginelocation + carwidth +
##     aspiration
```

```
##
##                    Df Sum of Sq       RSS     AIC
## + boreratio         1  22631857 528952926 1904.73
## + carbody           4  41477304 510107479 1906.27
## + peakrpm           1  15517718 536067064 1906.37
## + carlength         1  13189256 538395526 1906.91
## + fuelsystem        5  43251732 508333050 1907.84
## + citympg           1   9011132 542573650 1907.86
## <none>                          551584782 1907.88
## + horsepower        1   7542814 544041968 1908.19
## + highwaympg        1   4325627 547259155 1908.91
## + symboling         1   1714374 549870408 1909.50
## + compressionratio  1   1026065 550558717 1909.65
## + carheight         1    549359 551035423 1909.76
## + fueltype          1    416946 551167836 1909.79
## + wheelbase         1    372555 551212227 1909.80
## + curbweight        1     83723 551501059 1909.86
## + doornumber        1      5064 551579718 1909.88
## + drivewheel        2   5438871 546145911 1910.66
##
## Step:  AIC=1904.73
## price ~ enginesize + enginetype + stroke + enginelocation + carwidth +
##     aspiration + boreratio
##
##                    Df Sum of Sq       RSS     AIC
## + carbody           4  51611653 477341273 1900.10
## + peakrpm           1  15925561 513027364 1902.97
## <none>                          528952926 1904.73
## + carlength         1   6229762 522723163 1905.27
## + fuelsystem        5  37662071 491290854 1905.64
## + drivewheel        2  12787012 516165914 1905.72
## + citympg           1   3120533 525832393 1906.00
## + symboling         1   2461875 526491050 1906.16
## + curbweight        1   1747528 527205398 1906.32
## + horsepower        1   1576341 527376584 1906.36
## + carheight         1   1177751 527775175 1906.46
## + highwaympg        1   1056744 527896182 1906.48
## + wheelbase         1    839116 528113809 1906.53
## + compressionratio  1     82783 528870143 1906.71
## + doornumber        1     54185 528898740 1906.72
## + fueltype          1      9225 528943700 1906.73
##
## Step:  AIC=1900.1
## price ~ enginesize + enginetype + stroke + enginelocation + carwidth +
##     aspiration + boreratio + carbody
##
##                    Df Sum of Sq       RSS     AIC
## + peakrpm           1  28596901 448744372 1894.50
## + carlength         1  16852132 460489141 1897.68
## <none>                          477341273 1900.10
## + fuelsystem        5  32361729 444979544 1901.47
## + symboling         1   2230926 475110347 1901.53
```

```
## + drivewheel        2   9719728 467621545 1901.57
## + fueltype          1   2005701 475335572 1901.58
## + citympg           1   1940574 475400699 1901.60
## + doornumber        1   1318022 476023251 1901.76
## + compressionratio  1    994581 476346692 1901.85
## + highwaympg        1    963685 476377588 1901.85
## + horsepower        1    385739 476955534 1902.00
## + carheight         1    344965 476996308 1902.01
## + wheelbase         1    213881 477127392 1902.05
## + curbweight        1      1205 477340068 1902.10
##
## Step:  AIC=1894.5
## price ~ enginesize + enginetype + stroke + enginelocation + carwidth +
##     aspiration + boreratio + carbody + peakrpm
##
##                    Df Sum of Sq       RSS     AIC
## + citympg           1  15306343 433438028 1892.23
## + highwaympg        1  13229431 435514940 1892.82
## + carlength         1  10086898 438657473 1893.71
## <none>                          448744372 1894.50
## + horsepower        1   6751189 441993183 1894.64
## + compressionratio  1   3930019 444814353 1895.42
## + symboling         1   3185860 445558512 1895.63
## + carheight         1   3124334 445620038 1895.64
## + fueltype          1   2427450 446316921 1895.84
## + fuelsystem        5  30329510 418414861 1895.89
## + drivewheel        2   7891481 440852891 1896.32
## + curbweight        1     35127 448709245 1896.49
## + wheelbase         1     18264 448726108 1896.50
## + doornumber        1      1211 448743160 1896.50
##
## Step:  AIC=1892.23
## price ~ enginesize + enginetype + stroke + enginelocation + carwidth +
##     aspiration + boreratio + carbody + peakrpm + citympg
##
##                    Df Sum of Sq       RSS     AIC
## <none>                          433438028 1892.23
## + curbweight        1   4097051 429340977 1893.07
## + carheight         1   3344369 430093660 1893.28
## + carlength         1   3154515 430283514 1893.34
## + symboling         1   2125575 431312453 1893.63
## + drivewheel        2   8930786 424507243 1893.67
## + fueltype          1    221222 433216807 1894.17
## + horsepower        1    111090 433326939 1894.20
## + wheelbase         1     31132 433406897 1894.22
## + compressionratio  1     22244 433415784 1894.23
## + highwaympg        1     16414 433421614 1894.23
## + doornumber        1      2085 433435943 1894.23
## + fuelsystem        5  22955233 410482795 1895.54
```

```
summary(carprice.lm.fwd)
```

```
##
## Call:
## lm(formula = price ~ enginesize + enginetype + stroke + enginelocation +
##     carwidth + aspiration + boreratio + carbody + peakrpm + citympg,
##     data = train.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5665.43  -946.89   -20.97   997.03  5430.49
##
## Coefficients:
##                        Estimate   Std. Error t value             Pr(>|t|)
## (Intercept)        -49748.47065  12302.89155 -4.0436             0.0001012 ***
## enginesize            184.09418     15.01453 12.2611 < 0.00000000000000022 ***
## enginetypedohcv      1873.21193   2396.65316  0.7816             0.4362284
## enginetypel          -665.10967   1286.15401 -0.5171             0.6061630
## enginetypeohc        1806.60496    841.64875  2.1465             0.0341575 *
## enginetypeohcf      -1381.32924   1452.52166 -0.9510             0.3438158
## enginetypeohcv      -4440.73349   1083.62850 -4.0980       0.0000827484492 ***
## enginetyperotor     13414.99350   1935.68089  6.9304       0.0000000003627 ***
## stroke              -6186.59638    863.06321 -7.1682       0.0000000001142 ***
## enginelocationrear  10286.06685   2379.64062  4.3225       0.0000354261866 ***
## carwidth              882.62487    185.02198  4.7704       0.0000060189913 ***
## aspirationturbo      3258.54678    558.54135  5.8340       0.0000000617753 ***
## boreratio           -2604.80165   1164.10937 -2.2376             0.0273803 *
## carbodyhardtop      -2828.86339   1498.33384 -1.8880             0.0618123 .
## carbodyhatchback    -4519.51953   1228.07486 -3.6802             0.0003714 ***
## carbodysedan        -3517.80440   1201.24776 -2.9285             0.0041860 **
## carbodywagon        -3328.31363   1288.46652 -2.5832             0.0111797 *
## peakrpm                 1.80500      0.56884  3.1731             0.0019834 **
## citympg               102.05068     53.25091  1.9164             0.0580578 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2041.5 on 104 degrees of freedom
## Multiple R-squared:  0.9398, Adjusted R-squared:  0.92938
## F-statistic: 90.198 on 18 and 104 DF,  p-value: < 0.000000000000000222
```

Backward Elimination Method

```
carprice.lm.back <- step(carprice.lm,                    # start with full model
                         direction = "backward")       # backward elimination
```

```
## Start:  AIC=1909.22
## price ~ symboling + fueltype + aspiration + doornumber + carbody +
##     drivewheel + enginelocation + wheelbase + carlength + carwidth +
##     carheight + curbweight + enginetype + enginesize + fuelsystem +
##     boreratio + stroke + compressionratio + horsepower + peakrpm +
##     citympg + highwaympg
##
##
## Step:  AIC=1909.22
## price ~ symboling + aspiration + doornumber + carbody + drivewheel +
##     enginelocation + wheelbase + carlength + carwidth + carheight +
##     curbweight + enginetype + enginesize + fuelsystem + boreratio +
##     stroke + compressionratio + horsepower + peakrpm + citympg +
##     highwaympg
##
##                    Df Sum of Sq       RSS     AIC
## - fuelsystem        5  20422785 404052282 1905.60
## - drivewheel        2   3334182 386963680 1906.28
## - doornumber        1     30325 383659822 1907.23
## - citympg           1    199600 383829097 1907.28
## - compressionratio  1   1782181 385411678 1907.79
## - highwaympg        1   1935362 385564859 1907.84
## - symboling         1   3562580 387192078 1908.36
## - curbweight        1   3811536 387441034 1908.43
## - carlength         1   4361751 387991249 1908.61
## - wheelbase         1   5154483 388783980 1908.86
## - carbody           4  25394721 409024218 1909.10
## - carheight         1   6040415 389669912 1909.14
## <none>                          383629497 1909.22
## - horsepower        1   6636874 390266371 1909.33
## - boreratio         1  18887408 402516905 1913.13
## - enginelocation    1  29748361 413377858 1916.41
## - peakrpm           1  33703564 417333061 1917.58
## - carwidth          1  59352461 442981958 1924.91
## - aspiration        1  65563629 449193126 1926.63
## - stroke            1 107333689 490963186 1937.56
## - enginetype        6 223660005 607289502 1953.72
## - enginesize        1 243820492 627449989 1967.73
##
## Step:  AIC=1905.6
## price ~ symboling + aspiration + doornumber + carbody + drivewheel +
##     enginelocation + wheelbase + carlength + carwidth + carheight +
##     curbweight + enginetype + enginesize + boreratio + stroke +
##     compressionratio + horsepower + peakrpm + citympg + highwaympg
##
##                    Df Sum of Sq       RSS     AIC
## - doornumber        1     21152 404073434 1903.61
## - citympg           1    162030 404214312 1903.65
## - symboling         1   1108443 405160725 1903.94
## - wheelbase         1   1138379 405190661 1903.94
## - drivewheel        2   8329903 412382185 1904.11
## - compressionratio  1   1782187 405834469 1904.14
```

```
## - highwaympg        1   2986129 407038411 1904.51
## - curbweight        1   3038120 407090402 1904.52
## - horsepower        1   3293867 407346149 1904.60
## - carlength         1   4996117 409048399 1905.11
## <none>                          404052282 1905.60
## - carheight         1   9785221 413837503 1906.54
## - boreratio         1  19587167 423639449 1909.42
## - carbody           4  40951013 445003295 1909.47
## - peakrpm           1  34904694 438956976 1913.79
## - enginelocation    1  39620871 443673153 1915.11
## - carwidth          1  65579335 469631617 1922.10
## - aspiration        1  67775299 471827581 1922.67
## - stroke            1 138092620 542144902 1939.76
## - enginetype        6 263843035 667895317 1955.42
## - enginesize        1 256763232 660815514 1964.11
##
## Step:  AIC=1903.61
## price ~ symboling + aspiration + carbody + drivewheel + enginelocation +
##     wheelbase + carlength + carwidth + carheight + curbweight +
##     enginetype + enginesize + boreratio + stroke + compressionratio +
##     horsepower + peakrpm + citympg + highwaympg
##
##                    Df Sum of Sq       RSS     AIC
## - citympg           1    147963 404221397 1901.65
## - symboling         1   1123029 405196463 1901.95
## - wheelbase         1   1130810 405204244 1901.95
## - compressionratio  1   1807166 405880600 1902.15
## - drivewheel        2   8968854 413042288 1902.31
## - highwaympg        1   2998897 407072331 1902.52
## - curbweight        1   3041709 407115143 1902.53
## - horsepower        1   3272717 407346151 1902.60
## - carlength         1   5035985 409109419 1903.13
## <none>                          404073434 1903.61
## - carheight         1   9814227 413887661 1904.56
## - boreratio         1  19766303 423839737 1907.48
## - carbody           4  41674322 445747756 1907.68
## - peakrpm           1  35038259 439111693 1911.83
## - enginelocation    1  39610512 443683945 1913.11
## - carwidth          1  65698840 469772274 1920.14
## - aspiration        1  67984456 472057890 1920.73
## - stroke            1 138240851 542314285 1937.80
## - enginetype        6 263826435 667899869 1953.42
## - enginesize        1 256927661 661001095 1962.14
##
## Step:  AIC=1901.65
## price ~ symboling + aspiration + carbody + drivewheel + enginelocation +
##     wheelbase + carlength + carwidth + carheight + curbweight +
##     enginetype + enginesize + boreratio + stroke + compressionratio +
##     horsepower + peakrpm + highwaympg
##
##                    Df Sum of Sq       RSS     AIC
## - symboling         1   1031375 405252772 1899.96
```

```
## - wheelbase         1   1175633 405397030 1900.01
## - compressionratio  1   1984993 406206390 1900.25
## - drivewheel        2   9012394 413233791 1900.36
## - horsepower        1   3167297 407388693 1900.61
## - curbweight        1   3364092 407585489 1900.67
## - carlength         1   4888229 409109626 1901.13
## <none>                          404221397 1901.65
## - highwaympg        1   8084978 412306375 1902.09
## - carheight         1   9674893 413896290 1902.56
## - boreratio         1  19619284 423840681 1905.48
## - carbody           4  41766942 445988338 1905.74
## - peakrpm           1  34964149 439185546 1909.86
## - enginelocation    1  39919138 444140535 1911.23
## - carwidth          1  66134047 470355444 1918.29
## - aspiration        1  73007020 477228416 1920.07
## - stroke            1 140207981 544429378 1936.28
## - enginetype        6 272930127 677151524 1953.11
## - enginesize        1 267447336 671668733 1962.11
##
## Step:  AIC=1899.96
## price ~ aspiration + carbody + drivewheel + enginelocation +
##     wheelbase + carlength + carwidth + carheight + curbweight +
##     enginetype + enginesize + boreratio + stroke + compressionratio +
##     horsepower + peakrpm + highwaympg
##
##                    Df Sum of Sq       RSS     AIC
## - wheelbase         1    353388 405606160 1898.07
## - compressionratio  1   1989681 407242453 1898.57
## - drivewheel        2   9793368 415046140 1898.90
## - horsepower        1   3167425 408420197 1898.92
## - curbweight        1   4052679 409305451 1899.19
## - carlength         1   6605325 411858097 1899.95
## <none>                          405252772 1899.96
## - highwaympg        1   7887728 413140500 1900.34
## - carheight         1   9147368 414400140 1900.71
## - boreratio         1  18760557 424013329 1903.53
## - carbody           4  46365517 451618289 1905.29
## - peakrpm           1  34315282 439568054 1907.96
## - enginelocation    1  42399798 447652570 1910.20
## - carwidth          1  65513331 470766103 1916.40
## - aspiration        1  71995005 477247777 1918.08
## - stroke            1 139188196 544440968 1934.28
## - enginetype        6 273540857 678793629 1951.41
## - enginesize        1 297653296 702906068 1965.70
##
## Step:  AIC=1898.07
## price ~ aspiration + carbody + drivewheel + enginelocation +
##     carlength + carwidth + carheight + curbweight + enginetype +
##     enginesize + boreratio + stroke + compressionratio + horsepower +
##     peakrpm + highwaympg
##
##                    Df Sum of Sq       RSS     AIC
```

```
## - compressionratio  1   1916876 407523036 1896.65
## - drivewheel        2   9441506 415047666 1896.90
## - horsepower        1   2881985 408488145 1896.94
## - curbweight        1   4220240 409826399 1897.34
## <none>                          405606160 1898.07
## - highwaympg        1   7796441 413402601 1898.41
## - carlength         1   8468045 414074205 1898.61
## - carheight         1   9452428 415058587 1898.90
## - boreratio         1  18443029 424049188 1901.54
## - carbody           4  50517481 456123641 1904.51
## - peakrpm           1  34044169 439650329 1905.98
## - enginelocation    1  47101581 452707741 1909.58
## - carwidth          1  68122215 473728375 1915.17
## - aspiration        1  71992056 477598216 1916.17
## - stroke            1 139160490 544766650 1932.35
## - enginetype        6 276306858 681913018 1949.97
## - enginesize        1 380238632 785844791 1977.42
##
## Step:  AIC=1896.65
## price ~ aspiration + carbody + drivewheel + enginelocation +
##     carlength + carwidth + carheight + curbweight + enginetype +
##     enginesize + boreratio + stroke + horsepower + peakrpm +
##     highwaympg
##
##                  Df Sum of Sq       RSS     AIC
## - horsepower      1   1230736 408753772 1895.02
## - drivewheel      2   7953197 415476233 1895.03
## - curbweight      1   2787047 410310082 1895.49
## - highwaympg      1   5957311 413480346 1896.44
## <none>                        407523036 1896.65
## - carheight       1   9168191 416691226 1897.39
## - carlength       1   9327685 416850721 1897.43
## - boreratio       1  17981508 425504544 1899.96
## - carbody         4  51601598 459124634 1903.32
## - peakrpm         1  35202378 442725414 1904.84
## - enginelocation  1  48149765 455672801 1908.39
## - carwidth        1  66234557 473757592 1913.17
## - aspiration      1  86042287 493565323 1918.21
## - stroke          1 149055379 556578415 1932.99
## - enginetype      6 276195067 683718103 1948.30
## - enginesize      1 392342344 799865379 1977.60
##
## Step:  AIC=1895.02
## price ~ aspiration + carbody + drivewheel + enginelocation +
##     carlength + carwidth + carheight + curbweight + enginetype +
##     enginesize + boreratio + stroke + peakrpm + highwaympg
##
##                  Df Sum of Sq       RSS     AIC
## - drivewheel      2   7660170 416413942 1893.31
## - curbweight      1   2374720 411128491 1893.73
## <none>                        408753772 1895.02
## - carlength       1   9153775 417907546 1895.75
```

```
## - highwaympg      1  10203177 418956948 1896.06
## - carheight       1  10545626 419299398 1896.15
## - boreratio       1  23363986 432117758 1899.86
## - peakrpm         1  35017678 443771449 1903.13
## - carbody         4  59432939 468186711 1903.72
## - enginelocation  1  52845078 461598850 1907.98
## - carwidth        1  65933325 474687097 1911.42
## - aspiration      1  97119907 505873679 1919.24
## - stroke          1 149096398 557850170 1931.27
## - enginetype      6 278890026 687643798 1947.00
## - enginesize      1 409293105 818046876 1978.36
##
## Step:  AIC=1893.31
## price ~ aspiration + carbody + enginelocation + carlength + carwidth +
##     carheight + curbweight + enginetype + enginesize + boreratio +
##     stroke + peakrpm + highwaympg
##
##                   Df Sum of Sq       RSS     AIC
## <none>                          416413942 1893.31
## - curbweight      1   7134389 423548331 1893.39
## - carheight       1   8185657 424599599 1893.70
## - highwaympg      1   9919582 426333524 1894.20
## - carlength       1  12381806 428795748 1894.91
## - boreratio       1  18231119 434645061 1896.58
## - peakrpm         1  34776916 451190858 1901.17
## - carbody         4  61384381 477798323 1902.22
## - enginelocation  1  59313832 475727774 1907.68
## - carwidth        1  62653196 479067138 1908.55
## - aspiration      1  92540994 508954936 1915.99
## - stroke          1 192783631 609197573 1938.10
## - enginetype      6 318202050 734615992 1951.13
## - enginesize      1 444907268 861321210 1980.70
```

```
summary(carprice.lm.back)
```

```
##
## Call:
## lm(formula = price ~ aspiration + carbody + enginelocation +
##     carlength + carwidth + carheight + curbweight + enginetype +
##     enginesize + boreratio + stroke + peakrpm + highwaympg, data = train.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5252.45  -962.76   -12.04  1007.90  5422.01
##
## Coefficients:
##                       Estimate  Std. Error t value              Pr(>|t|)
## (Intercept)        -53174.6443  13551.4194 -3.9239             0.0001592 ***
## aspirationturbo      2969.3185    626.7461  4.7377       0.0000070846887 ***
## carbodyhardtop      -2478.7855   1558.1485 -1.5909             0.1147679
## carbodyhatchback    -4291.1497   1264.9442 -3.3924             0.0009906 ***
## carbodysedan        -3187.3058   1332.3909 -2.3922             0.0185985 *
## carbodywagon        -3528.3490   1540.0841 -2.2910             0.0240401 *
## enginelocationrear   9398.8163   2477.9761  3.7929             0.0002538 ***
## carlength             -89.8672     51.8575 -1.7330             0.0861538 .
## carwidth              939.8321    241.0908  3.8982             0.0001745 ***
## carheight             180.2541    127.9264  1.4090             0.1618933
## curbweight              1.9725      1.4994  1.3155             0.1913350
## enginetypedohcv        -7.9498   2726.2530 -0.0029             0.9976791
## enginetypel          -913.1027   1304.2352 -0.7001             0.4854707
## enginetypeohc        1491.1644    893.2973  1.6693             0.0981585 .
## enginetypeohcf      -1236.6977   1655.4446 -0.7470             0.4567698
## enginetypeohcv      -4994.2792   1196.8028 -4.1730       0.0000637623104 ***
## enginetyperotor     13253.2837   1977.9764  6.7004       0.0000000012014 ***
## enginesize            179.7254     17.3012 10.3880 < 0.00000000000000022 ***
## boreratio           -2459.0791   1169.4126 -2.1028             0.0379659 *
## stroke              -6057.2255    885.8101 -6.8381       0.0000000006246 ***
## peakrpm                 1.7818      0.6135  2.9043             0.0045215 **
## highwaympg             98.8964     63.7582  1.5511             0.1240014
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2030.5 on 101 degrees of freedom
## Multiple R-squared:  0.94216,    Adjusted R-squared:  0.93014
## F-statistic: 78.349 on 21 and 101 DF,  p-value: < 0.000000000000000222
```

Stepwise Regression Method

```
carprice.lm.step <- step(carprice.lm.null,                                    # initial model
                         scope = list(carprice.lm.null, upper = carprice.lm),   # range of models
                         direction = "both")                                  # stepwise regression
```

```
## Start:  AIC=2201.87
## price ~ 1
##
##                    Df  Sum of Sq        RSS     AIC
## + enginesize        1 5561412259 1638506332 2021.80
## + curbweight        1 4690880966 2509037624 2074.21
## + horsepower        1 4519164258 2680754332 2082.35
## + carwidth          1 4149121446 3050797144 2098.26
## + highwaympg        1 3448678626 3751239964 2123.68
## + citympg           1 3345191295 3854727295 2127.03
## + carlength         1 2839250295 4360668295 2142.20
## + drivewheel        2 2846712381 4353206209 2143.99
## + boreratio         1 2715463920 4484454670 2145.64
## + fuelsystem        5 2541627437 4658291154 2158.32
## + enginetype        6 2244805091 4955113499 2167.91
## + wheelbase         1 1755202198 5444716392 2169.50
## + carbody           4 1861869651 5338048939 2173.07
## + enginelocation    1 1035114238 6164804352 2184.78
## + aspiration        1  216061385 6983857206 2200.13
## + peakrpm           1  158472939 7041445652 2201.14
## <none>                           7199918590 2201.87
## + symboling         1   54103139 7145815452 2202.95
## + carheight         1   43994324 7155924266 2203.12
## + doornumber        1   37797285 7162121305 2203.23
## + fueltype          1   37236672 7162681919 2203.24
## + stroke            1    6308330 7193610261 2203.77
## + compressionratio  1    6225538 7193693052 2203.77
##
## Step:  AIC=2021.8
## price ~ enginesize
##
##                    Df  Sum of Sq        RSS     AIC
## + enginetype        6  478852929 1159653403 1991.28
## + enginelocation    1  283734157 1354772175 2000.41
## + horsepower        1  221804220 1416702111 2005.91
## + drivewheel        2  216132780 1422373552 2008.40
## + fuelsystem        5  280995640 1357510692 2008.66
## + stroke            1  183987692 1454518639 2009.15
## + carwidth          1  176739031 1461767300 2009.76
## + curbweight        1  131010091 1507496240 2013.55
## + citympg           1  115758767 1522747565 2014.79
## + highwaympg        1   93229258 1545277073 2016.59
## + carbody           4  159915964 1478590368 2017.17
## + peakrpm           1   78426596 1560079735 2017.77
## + aspiration        1   65631261 1572875071 2018.77
## + boreratio         1   34987714 1603518618 2021.14
## + carlength         1   34801601 1603704730 2021.16
## <none>                           1638506332 2021.80
## + compressionratio  1   20594506 1617911826 2022.24
## + symboling         1   16582103 1621924228 2022.55
## + fueltype          1   16393755 1622112576 2022.56
## + doornumber        1   12209023 1626297309 2022.88
```

```
## + wheelbase         1    7523265 1630983067 2023.23
## + carheight         1    2569791 1635936540 2023.61
## - enginesize        1 5561412259 7199918590 2201.87
##
## Step:  AIC=1991.28
## price ~ enginesize + enginetype
##
##                    Df  Sum of Sq        RSS     AIC
## + stroke            1  200995601  958657802 1969.87
## + enginelocation    1  148262997 1011390406 1976.46
## + horsepower        1   90871061 1068782342 1983.24
## + carwidth          1   81331691 1078321712 1984.34
## + peakrpm           1   72939237 1086714166 1985.29
## + fuelsystem        5  140635114 1019018289 1985.38
## + carbody           4  118892900 1040760503 1985.98
## + curbweight        1   48213350 1111440053 1988.06
## + aspiration        1   44780220 1114873183 1988.44
## + drivewheel        2   50364632 1109288771 1989.82
## + boreratio         1   24746274 1134907129 1990.63
## <none>                           1159653403 1991.28
## + compressionratio  1   14100425 1145552978 1991.78
## + citympg           1   10150928 1149502475 1992.20
## + fueltype          1    9891507 1149761895 1992.23
## + doornumber        1    9364220 1150289183 1992.29
## + carheight         1    9049534 1150603869 1992.32
## + highwaympg        1    8453974 1151199429 1992.38
## + symboling         1    5136317 1154517085 1992.74
## + carlength         1    3558936 1156094467 1992.90
## + wheelbase         1    1154109 1158499294 1993.16
## - enginetype        6  478852929 1638506332 2021.80
## - enginesize        1 3795460096 4955113499 2167.91
##
## Step:  AIC=1969.87
## price ~ enginesize + enginetype + stroke
##
##                    Df  Sum of Sq        RSS     AIC
## + enginelocation    1  137338628  821319174 1952.85
## + aspiration        1  111643600  847014201 1956.64
## + peakrpm           1   90619482  868038320 1959.66
## + horsepower        1   88916556  869741246 1959.90
## + carwidth          1   65535584  893122218 1963.16
## + fuelsystem        5  101837318  856820483 1966.06
## + fueltype          1   37239304  921418498 1967.00
## + compressionratio  1   35297462  923360340 1967.26
## + carbody           4   77657837  880999964 1967.48
## + curbweight        1   32759391  925898411 1967.59
## + boreratio         1   16736853  941920949 1969.70
## <none>                           958657802 1969.87
## + doornumber        1    8538695  950119107 1970.77
## + wheelbase         1    5943356  952714445 1971.11
## + symboling         1    5507584  953150217 1971.16
## + highwaympg        1    3022253  955635549 1971.48
```

```
## + citympg           1    1715236  956942565 1971.65
## + carheight         1     555399  958102403 1971.80
## + carlength         1     532082  958125720 1971.80
## + drivewheel        2    8490015  950167786 1972.78
## - stroke            1  200995601 1159653403 1991.28
## - enginetype        6  495860838 1454518639 2009.15
## - enginesize        1 3806743883 4765401685 2165.11
##
## Step:  AIC=1952.85
## price ~ enginesize + enginetype + stroke + enginelocation
##
##                    Df  Sum of Sq        RSS     AIC
## + carwidth          1  183013500  638305674 1923.84
## + aspiration        1  160981718  660337455 1928.02
## + curbweight        1   96107972  725211202 1939.54
## + fuelsystem        5  103727245  717591929 1946.24
## + fueltype          1   43780191  777538982 1948.11
## + horsepower        1   39183186  782135988 1948.84
## + compressionratio  1   34722254  786596919 1949.54
## + peakrpm           1   23191618  798127555 1951.33
## + wheelbase         1   18677970  802641203 1952.02
## + carlength         1   15571773  805747401 1952.50
## + highwaympg        1   14693410  806625763 1952.63
## <none>                           821319174 1952.85
## + carheight         1    7506926  813812247 1953.72
## + drivewheel        2   19628984  801690189 1953.88
## + citympg           1    3802024  817517149 1954.28
## + boreratio         1    3256578  818062595 1954.36
## + symboling         1    1296051  820023122 1954.66
## + doornumber        1      27353  821291821 1954.85
## + carbody           4   26133303  795185870 1956.87
## - enginelocation    1  137338628  958657802 1969.87
## - stroke            1  190071232 1011390406 1976.46
## - enginetype        6  425467032 1246786206 1992.19
## - enginesize        1 3075720522 3897039696 2142.37
##
## Step:  AIC=1923.84
## price ~ enginesize + enginetype + stroke + enginelocation + carwidth
##
##                    Df Sum of Sq        RSS     AIC
## + aspiration        1  86720892  551584782 1907.88
## + curbweight        1  14578987  623726686 1923.00
## + carlength         1  13207501  625098173 1923.27
## + boreratio         1  12306603  625999071 1923.45
## <none>                           638305674 1923.84
## + fueltype          1   9365518  628940156 1924.03
## + peakrpm           1   7972501  630333173 1924.30
## + drivewheel        2  17423281  620882393 1924.44
## + horsepower        1   6972726  631332948 1924.49
## + compressionratio  1   5907081  632398593 1924.70
## + doornumber        1   2618591  635687083 1925.34
## + wheelbase         1   2457861  635847813 1925.37
```

```
## + carbody           4  31812256  606493418 1925.56
## + highwaympg        1   1213763  637091910 1925.61
## + citympg           1    300648  638005025 1925.79
## + symboling         1     64433  638241241 1925.83
## + carheight         1     44521  638261153 1925.84
## + fuelsystem        5  27229588  611076086 1928.48
## - stroke            1 158488431  796794105 1949.12
## - carwidth          1 183013500  821319174 1952.85
## - enginetype        6 266946326  905252000 1954.82
## - enginelocation    1 254816544  893122218 1963.16
## - enginesize        1 738607533 1376913207 2016.40
##
## Step:  AIC=1907.88
## price ~ enginesize + enginetype + stroke + enginelocation + carwidth +
##     aspiration
##
##                    Df Sum of Sq        RSS     AIC
## + boreratio         1  22631857  528952926 1904.73
## + carbody           4  41477304  510107479 1906.27
## + peakrpm           1  15517718  536067064 1906.37
## + carlength         1  13189256  538395526 1906.91
## + fuelsystem        5  43251732  508333050 1907.84
## + citympg           1   9011132  542573650 1907.86
## <none>                           551584782 1907.88
## + horsepower        1   7542814  544041968 1908.19
## + highwaympg        1   4325627  547259155 1908.91
## + symboling         1   1714374  549870408 1909.50
## + compressionratio  1   1026065  550558717 1909.65
## + carheight         1    549359  551035423 1909.76
## + fueltype          1    416946  551167836 1909.79
## + wheelbase         1    372555  551212227 1909.80
## + curbweight        1     83723  551501059 1909.86
## + doornumber        1      5064  551579718 1909.88
## + drivewheel        2   5438871  546145911 1910.66
## - aspiration        1  86720892  638305674 1923.84
## - carwidth          1 108752673  660337455 1928.02
## - stroke            1 215587105  767171888 1946.46
## - enginetype        6 313690966  865275748 1951.26
## - enginelocation    1 270202495  821787277 1954.92
## - enginesize        1 799083094 1350667876 2016.04
##
## Step:  AIC=1904.73
## price ~ enginesize + enginetype + stroke + enginelocation + carwidth +
##     aspiration + boreratio
##
##                    Df Sum of Sq        RSS     AIC
## + carbody           4  51611653  477341273 1900.10
## + peakrpm           1  15925561  513027364 1902.97
## <none>                           528952926 1904.73
## + carlength         1   6229762  522723163 1905.27
## + fuelsystem        5  37662071  491290854 1905.64
## + drivewheel        2  12787012  516165914 1905.72
```

```
## + citympg           1   3120533  525832393 1906.00
## + symboling         1   2461875  526491050 1906.16
## + curbweight        1   1747528  527205398 1906.32
## + horsepower        1   1576341  527376584 1906.36
## + carheight         1   1177751  527775175 1906.46
## + highwaympg        1   1056744  527896182 1906.48
## + wheelbase         1    839116  528113809 1906.53
## + compressionratio  1     82783  528870143 1906.71
## + doornumber        1     54185  528898740 1906.72
## + fueltype          1      9225  528943700 1906.73
## - boreratio         1  22631857  551584782 1907.88
## - aspiration        1  97046145  625999071 1923.45
## - carwidth          1 116497482  645450408 1927.21
## - stroke            1 211239382  740192307 1944.06
## - enginelocation    1 244943695  773896621 1949.54
## - enginetype        6 318404972  847357897 1950.69
## - enginesize        1 767802255 1296755181 2013.03
##
## Step:  AIC=1900.1
## price ~ enginesize + enginetype + stroke + enginelocation + carwidth +
##     aspiration + boreratio + carbody
##
##                    Df Sum of Sq        RSS     AIC
## + peakrpm           1  28596901  448744372 1894.50
## + carlength         1  16852132  460489141 1897.68
## <none>                           477341273 1900.10
## + fuelsystem        5  32361729  444979544 1901.47
## + symboling         1   2230926  475110347 1901.53
## + drivewheel        2   9719728  467621545 1901.57
## + fueltype          1   2005701  475335572 1901.58
## + citympg           1   1940574  475400699 1901.60
## + doornumber        1   1318022  476023251 1901.76
## + compressionratio  1    994581  476346692 1901.85
## + highwaympg        1    963685  476377588 1901.85
## + horsepower        1    385739  476955534 1902.00
## + carheight         1    344965  476996308 1902.01
## + wheelbase         1    213881  477127392 1902.05
## + curbweight        1      1205  477340068 1902.10
## - carbody           4  51611653  528952926 1904.73
## - boreratio         1  32766206  510107479 1906.27
## - aspiration        1 111465329  588806602 1923.91
## - carwidth          1 117823133  595164406 1925.24
## - enginelocation    1 167936231  645277504 1935.18
## - stroke            1 183245729  660587002 1938.06
## - enginetype        6 309280737  786622010 1949.54
## - enginesize        1 685678240 1163019513 2007.64
##
## Step:  AIC=1894.5
## price ~ enginesize + enginetype + stroke + enginelocation + carwidth +
##     aspiration + boreratio + carbody + peakrpm
##
##                    Df Sum of Sq        RSS     AIC
```

```
## + citympg           1  15306343  433438028 1892.23
## + highwaympg        1  13229431  435514940 1892.82
## + carlength         1  10086898  438657473 1893.71
## <none>                           448744372 1894.50
## + horsepower        1   6751189  441993183 1894.64
## + compressionratio  1   3930019  444814353 1895.42
## + symboling         1   3185860  445558512 1895.63
## + carheight         1   3124334  445620038 1895.64
## + fueltype          1   2427450  446316921 1895.84
## + fuelsystem        5  30329510  418414861 1895.89
## + drivewheel        2   7891481  440852891 1896.32
## + curbweight        1     35127  448709245 1896.49
## + wheelbase         1     18264  448726108 1896.50
## + doornumber        1      1211  448743160 1896.50
## - peakrpm           1  28596901  477341273 1900.10
## - boreratio         1  34995811  483740182 1901.74
## - carbody           4  64282993  513027364 1902.97
## - enginelocation    1  83886661  532631032 1913.58
## - carwidth          1  93069496  541813868 1915.68
## - aspiration        1 126633847  575378219 1923.08
## - stroke            1 198940377  647684749 1937.64
## - enginetype        6 307578020  756322392 1946.71
## - enginesize        1 674923817 1123668188 2005.40
##
## Step:  AIC=1892.23
## price ~ enginesize + enginetype + stroke + enginelocation + carwidth +
##     aspiration + boreratio + carbody + peakrpm + citympg
##
##                    Df Sum of Sq        RSS     AIC
## <none>                          433438028 1892.23
## + curbweight        1   4097051  429340977 1893.07
## + carheight         1   3344369  430093660 1893.28
## + carlength         1   3154515  430283514 1893.34
## + symboling         1   2125575  431312453 1893.63
## + drivewheel        2   8930786  424507243 1893.67
## + fueltype          1    221222  433216807 1894.17
## + horsepower        1    111090  433326939 1894.20
## + wheelbase         1     31132  433406897 1894.22
## + compressionratio  1     22244  433415784 1894.23
## + highwaympg        1     16414  433421614 1894.23
## + doornumber        1      2085  433435943 1894.23
## - citympg           1  15306343  448744372 1894.50
## + fuelsystem        5  22955233  410482795 1895.54
## - boreratio         1  20866779  454304807 1896.02
## - peakrpm           1  41962670  475400699 1901.60
## - carbody           4  66033595  499471624 1901.68
## - enginelocation    1  77869899  511307927 1910.56
## - carwidth          1  94841709  528279738 1914.57
## - aspiration        1 141850518  575288546 1925.06
## - stroke            1 214147017  647585046 1939.62
## - enginetype        6 319968413  753406442 1948.24
## - enginesize        1 626541748 1059979777 2000.23
```

```
summary(carprice.lm.step)
```

```
##
## Call:
## lm(formula = price ~ enginesize + enginetype + stroke + enginelocation +
##     carwidth + aspiration + boreratio + carbody + peakrpm + citympg,
##     data = train.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5665.43  -946.89   -20.97   997.03  5430.49
##
## Coefficients:
##                       Estimate   Std. Error t value             Pr(>|t|)
## (Intercept)        -49748.47065  12302.89155 -4.0436             0.0001012 ***
## enginesize            184.09418     15.01453 12.2611 < 0.00000000000000022 ***
## enginetypedohcv      1873.21193   2396.65316  0.7816             0.4362284
## enginetypel          -665.10967   1286.15401 -0.5171             0.6061630
## enginetypeohc        1806.60496    841.64875  2.1465             0.0341575 *
## enginetypeohcf      -1381.32924   1452.52166 -0.9510             0.3438158
## enginetypeohcv      -4440.73349   1083.62850 -4.0980       0.0000827484492 ***
## enginetyperotor     13414.99350   1935.68089  6.9304       0.0000000003627 ***
## stroke              -6186.59638    863.06321 -7.1682       0.0000000001142 ***
## enginelocationrear  10286.06685   2379.64062  4.3225       0.0000354261866 ***
## carwidth              882.62487    185.02198  4.7704       0.0000060189913 ***
## aspirationturbo      3258.54678    558.54135  5.8340       0.0000000617753 ***
## boreratio           -2604.80165   1164.10937 -2.2376             0.0273803 *
## carbodyhardtop      -2828.86339   1498.33384 -1.8880             0.0618123 .
## carbodyhatchback    -4519.51953   1228.07486 -3.6802             0.0003714 ***
## carbodysedan        -3517.80440   1201.24776 -2.9285             0.0041860 **
## carbodywagon        -3328.31363   1288.46652 -2.5832             0.0111797 *
## peakrpm                 1.80500      0.56884  3.1731             0.0019834 **
## citympg               102.05068     53.25091  1.9164             0.0580578 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2041.5 on 104 degrees of freedom
## Multiple R-squared:  0.9398, Adjusted R-squared:  0.92938
## F-statistic: 90.198 on 18 and 104 DF,  p-value: < 0.00000000000000222
```

Prediction Performance

```
library(forecast)
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```
# fwd prediction
valid.fwd.pred <- predict(carprice.lm.fwd, valid.data)
options(digits = 6)
accuracy(valid.fwd.pred, valid.data$price)
```

```
##              ME    RMSE     MAE     MPE    MAPE
## Test set 734.68 3351.24 2266.27 3.62129 16.9072
```

```
# back prediction
valid.back.pred <- predict(carprice.lm.back, valid.data)
options(digits = 6)
accuracy(valid.back.pred, valid.data$price)
```

```
##               ME   RMSE    MAE     MPE    MAPE
## Test set 719.602 3223.3 2169.9 3.45899 16.0479
```

```
# step prediction
valid.step.pred <- predict(carprice.lm.step, valid.data)
options(digits = 6)
accuracy(valid.step.pred, valid.data$price)
```

```
##              ME    RMSE     MAE     MPE    MAPE
## Test set 734.68 3351.24 2266.27 3.62129 16.9072
```

```
# exhaustive prediction
valid.full.pred <- predict(carprice.best1, valid.data)
options(digits = 6)
accuracy(valid.full.pred, valid.data$price)
```

```
##               ME   RMSE     MAE      MPE   MAPE
## Test set 565.003 4206.7 3008.32 -3.73843 22.177
```

Prep for Neural Net

```
summary(carprice)
```

```
##      car_ID      symboling               CarName      fueltype   aspiration
##  Min.   :  1   Min.   :-2.000   peugeot 504   :  6   diesel: 20   std  :168
##  1st Qu.: 52   1st Qu.: 0.000   toyota corolla:  6   gas   :185   turbo: 37
##  Median :103   Median : 1.000   toyota corona :  6
##  Mean   :103   Mean   : 0.834   subaru dl     :  4
##  3rd Qu.:154   3rd Qu.: 2.000   honda civic   :  3
##  Max.   :205   Max.   : 3.000   mazda 626     :  3
##  doornumber        carbody   drivewheel enginelocation   wheelbase
##  four:115   convertible: 6   4wd:  9    front:202      Min.   : 86.6
##  two : 90   hardtop    : 8   fwd:120    rear :  3      1st Qu.: 94.5
##             hatchback  :70   rwd: 76                   Median : 97.0
##             sedan      :96                             Mean   : 98.8
##             wagon      :25                             3rd Qu.:102.4
##                                                        Max.   :120.9
##    carlength      carwidth      carheight      curbweight   enginetype
##  Min.   :141   Min.   :60.3   Min.   :47.8   Min.   :1488   dohc : 12
##  1st Qu.:166   1st Qu.:64.1   1st Qu.:52.0   1st Qu.:2145   dohcv:  1
##  Median :173   Median :65.5   Median :54.1   Median :2414   l    : 12
##  Mean   :174   Mean   :65.9   Mean   :53.7   Mean   :2556   ohc  :148
##  3rd Qu.:183   3rd Qu.:66.9   3rd Qu.:55.5   3rd Qu.:2935   ohcf : 15
##  Max.   :208   Max.   :72.3   Max.   :59.8   Max.   :4066   ohcv : 13
##  cylindernumber   enginesize    fuelsystem   boreratio        stroke
##  eight :  5     Min.   : 61   mpfi   :94   Min.   :2.54   Min.   :2.07
##  five  : 11     1st Qu.: 97   2bbl   :66   1st Qu.:3.15   1st Qu.:3.11
##  four  :159     Median :120   idi    :20   Median :3.31   Median :3.29
##  six   : 24     Mean   :127   1bbl   :11   Mean   :3.33   Mean   :3.26
##  three :  1     3rd Qu.:141   spdi   : 9   3rd Qu.:3.58   3rd Qu.:3.41
##  twelve:  1     Max.   :326   4bbl   : 3   Max.   :3.94   Max.   :4.17
##  compressionratio   horsepower     peakrpm        citympg       highwaympg
##  Min.   : 7.0     Min.   : 48   Min.   :4150   Min.   :13.0   Min.   :16.0
##  1st Qu.: 8.6     1st Qu.: 70   1st Qu.:4800   1st Qu.:19.0   1st Qu.:25.0
##  Median : 9.0     Median : 95   Median :5200   Median :24.0   Median :30.0
##  Mean   :10.1     Mean   :104   Mean   :5125   Mean   :25.2   Mean   :30.8
##  3rd Qu.: 9.4     3rd Qu.:116   3rd Qu.:5500   3rd Qu.:30.0   3rd Qu.:34.0
##  Max.   :23.0     Max.   :288   Max.   :6600   Max.   :49.0   Max.   :54.0
##      price
##  Min.   : 5118
##  1st Qu.: 7788
##  Median :10295
##  Mean   :13277
##  3rd Qu.:16503
##  Max.   :45400
##  [ reached getOption("max.print") -- omitted 1 row ]
```

```
#remove ID, symboling, carname, enginetype, cylindernumber, fuelsystem
carprice.df <- carprice[,-c(1:3,15,16,18)]
summary(carprice.df)
```

```
##    fueltype   aspiration  doornumber        carbody   drivewheel enginelocation
##  diesel: 20   std  :168   four:115   convertible: 6   4wd:  9    front:202
##  gas   :185   turbo: 37   two : 90   hardtop    : 8   fwd:120    rear :  3
##                                      hatchback  :70   rwd: 76
##                                      sedan      :96
##                                      wagon      :25
##
##    wheelbase       carlength      carwidth      carheight      curbweight
##  Min.   : 86.6   Min.   :141   Min.   :60.3   Min.   :47.8   Min.   :1488
##  1st Qu.: 94.5   1st Qu.:166   1st Qu.:64.1   1st Qu.:52.0   1st Qu.:2145
##  Median : 97.0   Median :173   Median :65.5   Median :54.1   Median :2414
##  Mean   : 98.8   Mean   :174   Mean   :65.9   Mean   :53.7   Mean   :2556
##  3rd Qu.:102.4   3rd Qu.:183   3rd Qu.:66.9   3rd Qu.:55.5   3rd Qu.:2935
##  Max.   :120.9   Max.   :208   Max.   :72.3   Max.   :59.8   Max.   :4066
##    enginesize    boreratio        stroke     compressionratio   horsepower
##  Min.   : 61   Min.   :2.54   Min.   :2.07   Min.   : 7.0     Min.   : 48
##  1st Qu.: 97   1st Qu.:3.15   1st Qu.:3.11   1st Qu.: 8.6     1st Qu.: 70
##  Median :120   Median :3.31   Median :3.29   Median : 9.0     Median : 95
##  Mean   :127   Mean   :3.33   Mean   :3.26   Mean   :10.1     Mean   :104
##  3rd Qu.:141   3rd Qu.:3.58   3rd Qu.:3.41   3rd Qu.: 9.4     3rd Qu.:116
##  Max.   :326   Max.   :3.94   Max.   :4.17   Max.   :23.0     Max.   :288
##     peakrpm        citympg       highwaympg       price
##  Min.   :4150   Min.   :13.0   Min.   :16.0   Min.   : 5118
##  1st Qu.:4800   1st Qu.:19.0   1st Qu.:25.0   1st Qu.: 7788
##  Median :5200   Median :24.0   Median :30.0   Median :10295
##  Mean   :5125   Mean   :25.2   Mean   :30.8   Mean   :13277
##  3rd Qu.:5500   3rd Qu.:30.0   3rd Qu.:34.0   3rd Qu.:16503
##  Max.   :6600   Max.   :49.0   Max.   :54.0   Max.   :45400
```

```
#create dummy variable for fueltype, aspiration, doornumber, carbody, drivewheel, enginlocation
library(fastDummies)
carprice.df <- dummy_cols(carprice.df,
                          select_columns = c("fueltype","aspiration"),
                          remove_first_dummy = TRUE,
                          remove_selected_columns = TRUE)
carprice.df <- dummy_cols(carprice.df,
                          select_columns = c("doornumber","carbody"),
                          remove_first_dummy = TRUE,
                          remove_selected_columns = TRUE)
carprice.df <- dummy_cols(carprice.df,
                          select_columns = c("drivewheel","enginelocation"),
                          remove_first_dummy = TRUE,
                          remove_selected_columns = TRUE)
#check skewness for number varible in carprice.df for predictor
library(e1071)
skewness(carprice.df$wheelbase) # >1
```

```
## [1] 1.03489
```

```
skewness(carprice.df$carlength)
```

```
## [1] 0.153679
```

```
skewness(carprice.df$carwidth)
```

```
## [1] 0.890817
```

```
skewness(carprice.df$carheight)
```

```
## [1] 0.062202
```

```
skewness(carprice.df$curbweight)
```

```
## [1] 0.671459
```

```
skewness(carprice.df$enginesize)
```

```
## [1] 1.91925
```

```
skewness(carprice.df$boreratio)
```

```
## [1] 0.0198624
```

```
skewness(carprice.df$stroke)
```

```
## [1] -0.679644
```

```
skewness(carprice.df$compressionratio) #>2
```

```
## [1] 2.57278
```

```
skewness(carprice.df$horsepower) #>1
```

```
## [1] 1.38481
```

```
skewness(carprice.df$peakrpm)
```

```
## [1] 0.0740624
```

```
skewness(carprice.df$citympg)
```

```
## [1] 0.654023
```

```
skewness(carprice.df$highwaympg)
```

```
## [1] 0.53212
```

```
# apply a log transformation to highly skewed predictors.
carprice.df$wheelbase <- log(carprice.df$wheelbase + 1)
carprice.df$compressionratio <- log(carprice.df$compressionratio + 1)
carprice.df$horsepower<- log(carprice.df$horsepower + 1)

# partition the data 60% training and 40% valid
set.seed(1)

train.index <- sample(rownames(carprice.df), nrow(carprice.df) * 0.6)
carprice.train <- carprice.df[train.index, ]
valid.index <- setdiff(rownames(carprice.df), train.index)
carprice.valid <- carprice.df[valid.index, ]

# convert all variables (INCLUDING QUANTITATIVE OUTCOME) to a 0-1 scale
carprice.train.norm <- carprice.train
carprice.valid.norm <- carprice.valid

cols <- colnames(carprice.train)
for (i in cols) {
  carprice.valid.norm[[i]] <-
    (carprice.valid.norm[[i]] - min(carprice.train[[i]])) / (max(carprice.train[[i]]) - min(carp
rice.train[[i]]))
  carprice.train.norm[[i]] <-
    (carprice.train.norm[[i]] - min(carprice.train[[i]])) / (max(carprice.train[[i]]) - min(carp
rice.train[[i]]))
}
summary(carprice.train.norm)
```

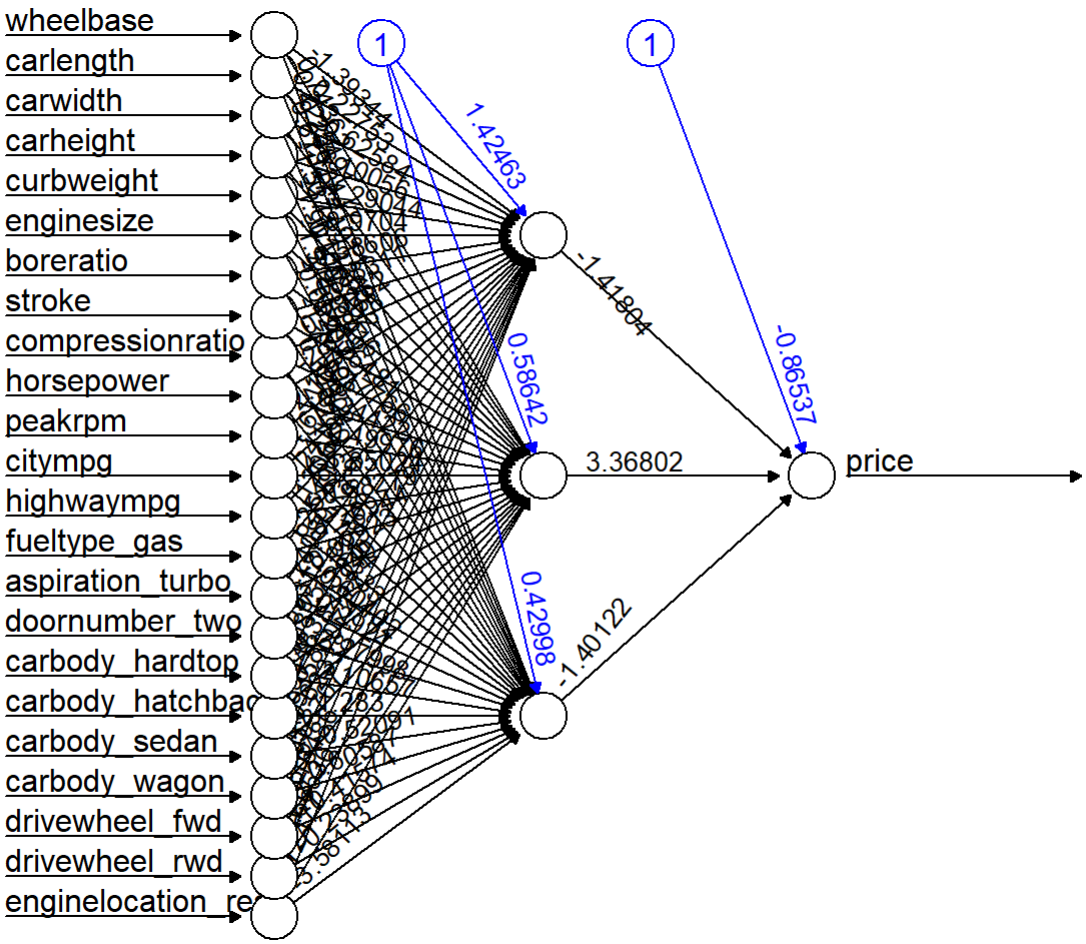
```
##    wheelbase       carlength        carwidth       carheight
##  Min.   :0.000   Min.   :0.000   Min.   :0.000   Min.   :0.000
##  1st Qu.:0.261   1st Qu.:0.342   1st Qu.:0.210   1st Qu.:0.255
##  Median :0.324   Median :0.438   Median :0.352   Median :0.427
##  Mean   :0.382   Mean   :0.455   Mean   :0.386   Mean   :0.442
##  3rd Qu.:0.467   3rd Qu.:0.595   3rd Qu.:0.486   3rd Qu.:0.609
##  Max.   :1.000   Max.   :1.000   Max.   :1.000   Max.   :1.000
##    curbweight      enginesize      boreratio         stroke
##  Min.   :0.000   Min.   :0.000   Min.   :0.000   Min.   :0.000
##  1st Qu.:0.179   1st Qu.:0.113   1st Qu.:0.364   1st Qu.:0.495
##  Median :0.315   Median :0.168   Median :0.521   Median :0.571
##  Mean   :0.360   Mean   :0.237   Mean   :0.549   Mean   :0.561
##  3rd Qu.:0.537   3rd Qu.:0.319   3rd Qu.:0.714   3rd Qu.:0.633
##  Max.   :1.000   Max.   :1.000   Max.   :1.000   Max.   :1.000
##  compressionratio   horsepower       peakrpm         citympg
##  Min.   :0.000    Min.   :0.000   Min.   :0.000   Min.   :0.000
##  1st Qu.:0.161    1st Qu.:0.172   1st Qu.:0.265   1st Qu.:0.143
##  Median :0.203    Median :0.350   Median :0.429   Median :0.286
##  Mean   :0.264    Mean   :0.374   Mean   :0.402   Mean   :0.322
##  3rd Qu.:0.239    3rd Qu.:0.501   3rd Qu.:0.551   3rd Qu.:0.457
##  Max.   :1.000    Max.   :1.000   Max.   :1.000   Max.   :1.000
##    highwaympg        price        fueltype_gas   aspiration_turbo
##  Min.   :0.000   Min.   :0.000   Min.   :0.000   Min.   :0.000
##  1st Qu.:0.237   1st Qu.:0.066   1st Qu.:1.000   1st Qu.:0.000
##  Median :0.368   Median :0.124   Median :1.000   Median :0.000
##  Mean   :0.386   Mean   :0.193   Mean   :0.902   Mean   :0.203
##  3rd Qu.:0.500   3rd Qu.:0.275   3rd Qu.:1.000   3rd Qu.:0.000
##  Max.   :1.000   Max.   :1.000   Max.   :1.000   Max.   :1.000
##  doornumber_two  carbody_hardtop  carbody_hatchback carbody_sedan
##  Min.   :0.000   Min.   :0.0000   Min.   :0.000     Min.   :0.000
##  1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:0.000     1st Qu.:0.000
##  Median :0.000   Median :0.0000   Median :0.000     Median :0.000
##  Mean   :0.455   Mean   :0.0325   Mean   :0.358     Mean   :0.431
##  3rd Qu.:1.000   3rd Qu.:0.0000   3rd Qu.:1.000     3rd Qu.:1.000
##  Max.   :1.000   Max.   :1.0000   Max.   :1.000     Max.   :1.000
##  carbody_wagon   drivewheel_fwd  drivewheel_rwd  enginelocation_rear
##  Min.   :0.000   Min.   :0.000   Min.   :0.000   Min.   :0.0000
##  1st Qu.:0.000   1st Qu.:0.000   1st Qu.:0.000   1st Qu.:0.0000
##  Median :0.000   Median :1.000   Median :0.000   Median :0.0000
##  Mean   :0.146   Mean   :0.585   Mean   :0.374   Mean   :0.0163
##  3rd Qu.:0.000   3rd Qu.:1.000   3rd Qu.:1.000   3rd Qu.:0.0000
##  Max.   :1.000   Max.   :1.000   Max.   :1.000   Max.   :1.0000
```

```
summary(carprice.valid.norm)
```

```
##    wheelbase       carlength          carwidth        carheight
##  Min.   :0.000   Min.   :-0.0551   Min.   :-0.143   Min.   :-0.0909
##  1st Qu.:0.282   1st Qu.: 0.3583   1st Qu.: 0.229   1st Qu.: 0.3364
##  Median :0.346   Median : 0.4543   Median : 0.352   Median : 0.4909
##  Mean   :0.396   Mean   : 0.4764   Mean   : 0.399   Mean   : 0.4569
##  3rd Qu.:0.502   3rd Qu.: 0.6413   3rd Qu.: 0.455   3rd Qu.: 0.6091
##  Max.   :0.865   Max.   : 0.9134   Max.   : 0.943   Max.   : 0.9364
##    curbweight        enginesize        boreratio         stroke
##  Min.   :-0.0956   Min.   :-0.0378   Min.   :0.100   Min.   :0.138
##  1st Qu.: 0.2054   1st Qu.: 0.1176   1st Qu.:0.436   1st Qu.:0.496
##  Median : 0.2915   Median : 0.2101   Median :0.607   Median :0.629
##  Mean   : 0.3551   Mean   : 0.2425   Mean   :0.587   Mean   :0.570
##  3rd Qu.: 0.4705   3rd Qu.: 0.2983   3rd Qu.:0.771   3rd Qu.:0.665
##  Max.   : 1.0000   Max.   : 1.0756   Max.   :0.886   Max.   :1.000
##  compressionratio   horsepower         peakrpm         citympg
##  Min.   :0.000    Min.   :-0.0463   Min.   :0.000   Min.   :-0.0286
##  1st Qu.:0.166    1st Qu.: 0.2026   1st Qu.:0.265   1st Qu.: 0.1571
##  Median :0.203    Median : 0.3502   Median :0.388   Median : 0.2857
##  Mean   :0.263    Mean   : 0.3590   Mean   :0.393   Mean   : 0.3185
##  3rd Qu.:0.239    3rd Qu.: 0.4605   3rd Qu.:0.551   3rd Qu.: 0.4429
##  Max.   :1.000    Max.   : 0.9444   Max.   :0.755   Max.   : 0.9429
##    highwaympg         price           fueltype_gas   aspiration_turbo
##  Min.   :0.0263   Min.   :-0.00192   Min.   :0.000   Min.   :0.000
##  1st Qu.:0.2368   1st Qu.: 0.06348   1st Qu.:1.000   1st Qu.:0.000
##  Median :0.3947   Median : 0.13058   Median :1.000   Median :0.000
##  Mean   :0.3918   Mean   : 0.21269   Mean   :0.902   Mean   :0.146
##  3rd Qu.:0.4737   3rd Qu.: 0.28361   3rd Qu.:1.000   3rd Qu.:0.000
##  Max.   :0.9737   Max.   : 0.89840   Max.   :1.000   Max.   :1.000
##  doornumber_two  carbody_hardtop  carbody_hatchback carbody_sedan
##  Min.   :0.000   Min.   :0.0000   Min.   :0.000     Min.   :0.000
##  1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:0.000     1st Qu.:0.000
##  Median :0.000   Median :0.0000   Median :0.000     Median :1.000
##  Mean   :0.415   Mean   :0.0488   Mean   :0.317     Mean   :0.524
##  3rd Qu.:1.000   3rd Qu.:0.0000   3rd Qu.:1.000     3rd Qu.:1.000
##  Max.   :1.000   Max.   :1.0000   Max.   :1.000     Max.   :1.000
##  carbody_wagon    drivewheel_fwd  drivewheel_rwd  enginelocation_rear
##  Min.   :0.0000   Min.   :0.000   Min.   :0.000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.000   1st Qu.:0.0000
##  Median :0.0000   Median :1.000   Median :0.000   Median :0.0000
##  Mean   :0.0854   Mean   :0.585   Mean   :0.366   Mean   :0.0122
##  3rd Qu.:0.0000   3rd Qu.:1.000   3rd Qu.:1.000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.000   Max.   :1.000   Max.   :1.0000
```

Neural Net with 1 Hidden Layer of 3 Nodes

```
library(neuralnet)
carprice.nn.3 <- neuralnet(price ~ .,                  # categorical outcome ~ predictors
                           data = carprice.train.norm,   # data for training model
                           linear.output = FALSE,      # assume relationship is nonlinear
                           hidden = 3)                 # a single hidden layer containing 3 node
s
plot(carprice.nn.3, rep = "best")
```




```
predict.nn.3 <- predict(carprice.nn.3, carprice.valid.norm)

head(predict.nn.3)
```

```
##       [,1]
## 2 0.254333
## 3 0.203549
## 4 0.121604
## 5 0.246711
## 6 0.269884
## 8 0.190275
```

```
# convert back to original scale
minprice <- min(carprice.train$price)
maxprice <- max(carprice.train$price)
actpred <- data.frame(actual = carprice.valid$price,
                      predicted = minprice + predict.nn.3*(maxprice - minprice))
head(actpred)
```

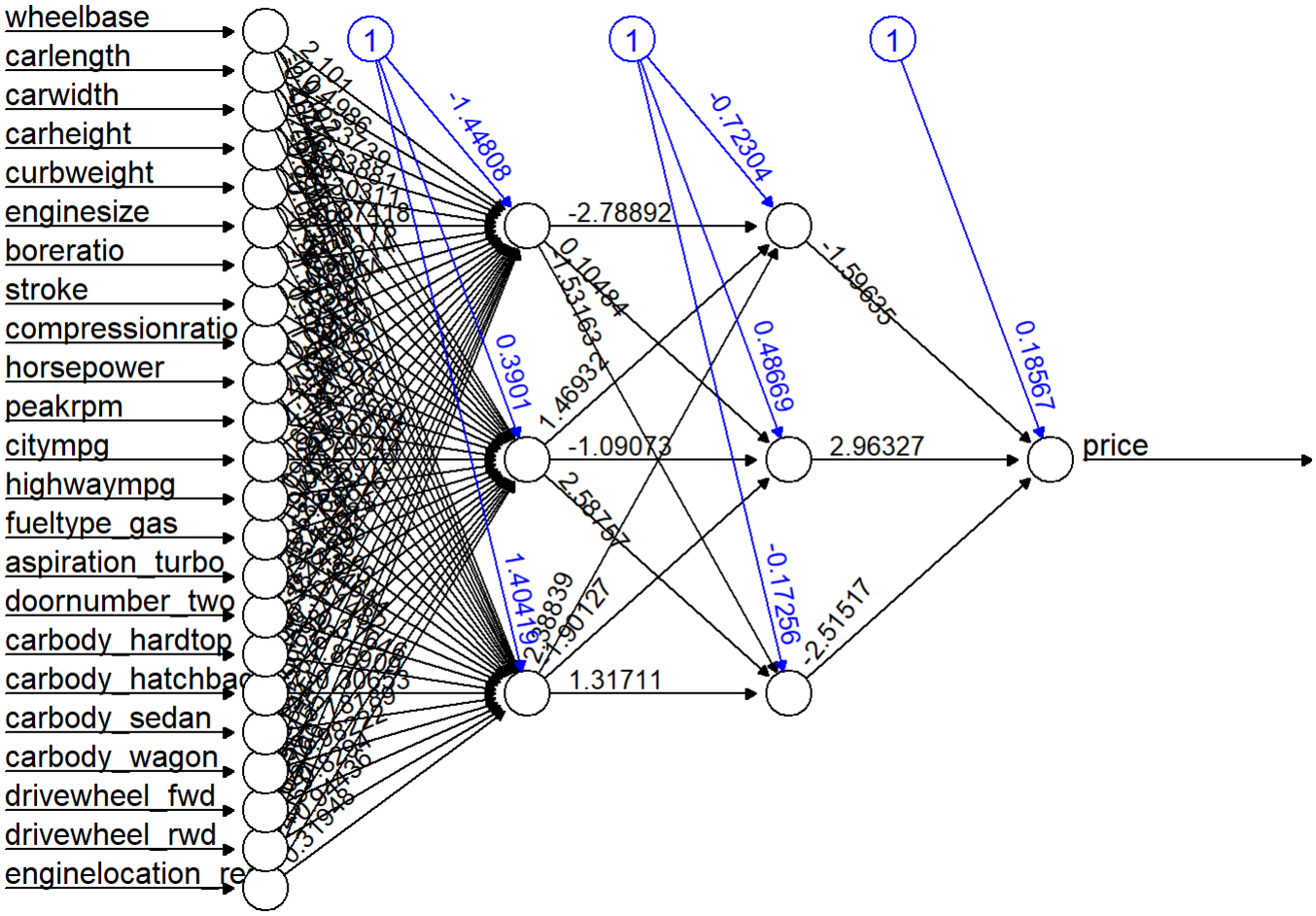
```
##   actual predicted
## 2  16500   15420.5
## 3  16500   13378.7
## 4  13950   10084.1
## 5  17450   15114.0
## 6  15250   16045.7
## 8  18920   12845.0
```

```
library(caret)
RMSE(actpred$predicted, actpred$actual) #3173.13
```

```
## [1] 3173.13
```

Neural Net with 2 Hidden Layers of 3 Nodes

```
carprice.nn.3.3 <- neuralnet(price ~ .,                  # categorical outcome ~ predictors
                             data = carprice.train.norm,   # data for training model
                             linear.output = FALSE,      # assume relationship is nonlinear
                             hidden = c(3,3))            # 2 hidden layers of 3 nodes each
plot(carprice.nn.3.3, rep = "best")
```

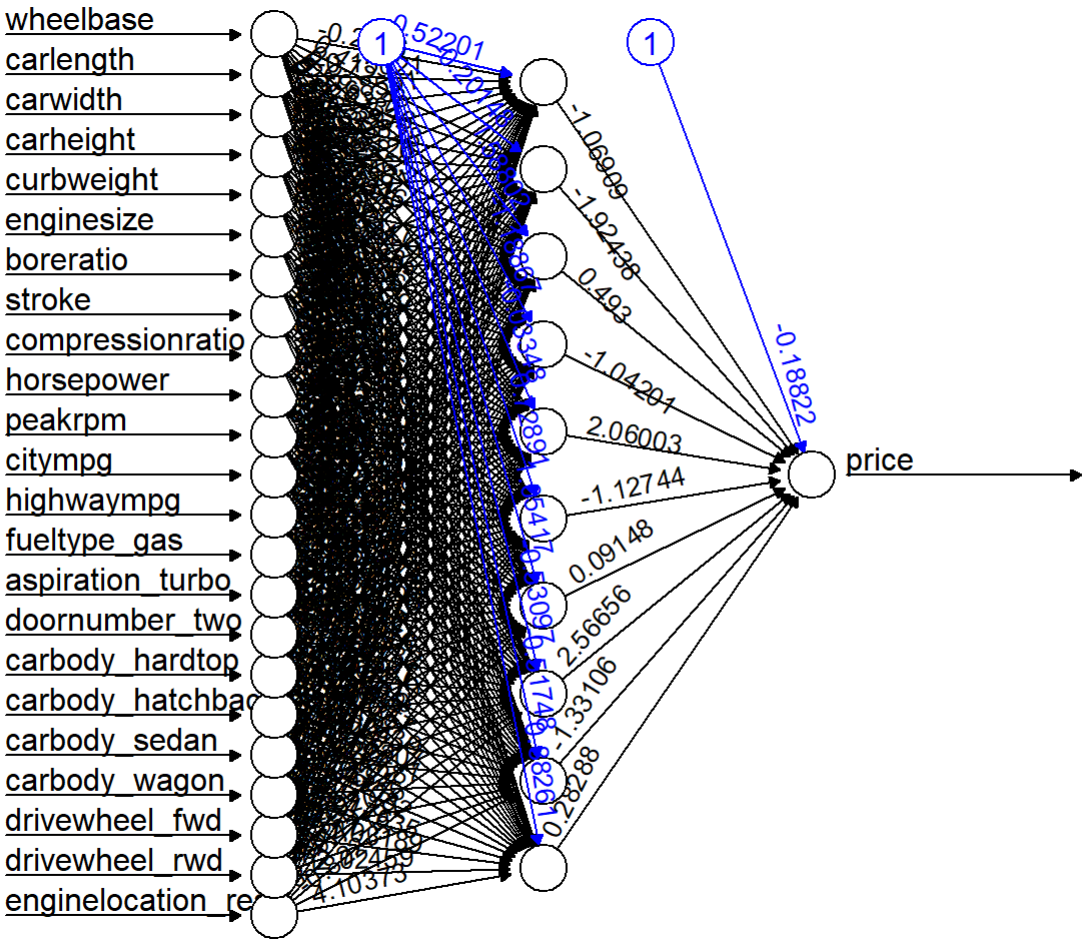



```
predict.nn.3.3 <- predict(carprice.nn.3.3, carprice.valid.norm)

# convert back to original scale
actpred <- data.frame(actual = carprice.valid$price,
                      predicted = minprice + predict.nn.3.3*(maxprice - minprice))
RMSE(actpred$predicted, actpred$actual) #4329.759
```

```
## [1] 3681.66
```

Neural Net with 1 Hidden Layer of 10 Nodes

```
carprice.nn.10 <- neuralnet(price ~ .,                  # categorical outcome ~ predictors
                            data = carprice.train.norm,   # data for training model
                            linear.output = FALSE,      # assume relationship is nonlinear
                            hidden = 10)            # 1 hidden layer of 10 nodes
plot(carprice.nn.10, rep = "best")
```




```
predict.nn.10 <- predict(carprice.nn.10, carprice.valid.norm)

# convert back to original scale
actpred <- data.frame(actual = carprice.valid$price,
                      predicted = minprice + predict.nn.10*(maxprice - minprice))
RMSE(actpred$predicted, actpred$actual) #3163.62
```

```
## [1] 3163.62
```

Loop Through Neural Nets with Different Layer Sizes

```
RMSE.df <- data.frame(n = seq(1, 20, 1), RMSE.k = rep(0, 20))

for (i in 1:20) {
  carprice.nn <- neuralnet(price ~ .,                  # categorical outcome ~ predictors
                           data = carprice.train.norm,   # data for training model
                           linear.output = FALSE,      # assume relationship is nonlinear
                           hidden = i)            # 1 hidden layer of 10 nodes
  predict.nn <- predict(carprice.nn, carprice.valid.norm)

  # convert back to original scale
  actpred <- data.frame(actual = carprice.valid$price,
                        predicted = minprice + predict.nn*(maxprice - minprice))
  RMSE.df[i,2] <- RMSE(actpred$predicted, actpred$actual)
}

RMSE.df #for i from 1-20, n= 3 is the lowest
```

```
##     n  RMSE.k
## 1   1 8446.67
## 2   2 3714.71
## 3   3 2783.91
## 4   4 3663.40
## 5   5 3800.94
## 6   6 3718.80
## 7   7 3557.68
## 8   8 3272.64
## 9   9 2898.99
## 10 10 3602.05
## 11 11 3305.40
## 12 12 3188.97
## 13 13 3374.23
## 14 14 3351.68
## 15 15 3222.20
## 16 16 3228.49
## 17 17 3095.46
## 18 18 3178.31
## 19 19 3917.29
## 20 20 3751.85
```

Through this process we find that the best Neural Net (having the lowest RMSE) is 1 hidden layer with 3 nodes.