Final Assignment

ISHIKA NISHA

12/12/2021

#Importing libraries

```
library('ggplot2')
library('ggthemes')
library('scales')
library('dplyr')

library('mice')

library('randomForest')

library('data.table')

library('gridExtra')

library('corrplot')

library('GGally')

library('e1071')
library('ggpubr')
library('tidyverse')

library('broom')
library('AICcmodavg')
```

#Checking the path

```
getwd()

## [1] "C:/Users/Nikesh/Downloads"
```

#Importing Dataset

```
Data <- read.csv("CarPrice_Assignment.csv")
```

#Description of Data

```
head(Data)
```

```
##   car_ID symboling wheelbase carlength carwidth carheight curbweight
enginesize
## 1      1         3      88.6     168.8     64.1      48.8       2548
130
## 2      2         3      88.6     168.8     64.1      48.8       2548
130
## 3      3         1      94.5     171.2     65.5      52.4       2823
152
## 4      4         2      99.8     176.6     66.2      54.3       2337
109
## 5      5         2      99.4     176.6     66.4      54.3       2824
136
## 6      6         2      99.8     177.3     66.3      53.1       2507
136
##   boreratio stroke compressionratio horsepower peakrpm citympg highwaympg
## 1      3.47   2.68              9.0        111    5000      21         27
## 2      3.47   2.68              9.0        111    5000      21         27
## 3      2.68   3.47              9.0        154    5000      19         26
## 4      3.19   3.40             10.0        102    5500      24         30
## 5      3.19   3.40              8.0        115    5500      18         22
## 6      3.19   3.40              8.5        110    5500      19         25
##   YearBuilt price                  CarName fueltype aspiration doornumber
## 1      2010 13495       alfa-romero giulia      gas        std        two
## 2      2011 16500      alfa-romero stelvio      gas        std        two
## 3      2010 16500 alfa-romero Quadrifoglio      gas        std        two
## 4      2006 13950              audi 100 ls      gas        std       four
## 5      2006 17450               audi 100ls      gas        std       four
## 6      2010 15250                 audi fox      gas        std        two
##       carbody drivewheel enginelocation enginetype cylindernumber
fuelsystem
## 1 convertible        rwd          front       dohc           four
mpfi
## 2 convertible        rwd          front       dohc           four
mpfi
## 3   hatchback        rwd          front       ohcv            six
mpfi
## 4       sedan        fwd          front        ohc           four
mpfi
## 5       sedan        4wd          front        ohc           five
mpfi
## 6       sedan        fwd          front        ohc           five
mpfi

tail(data)

##
## 182             }
## 183             rm(tmp_env)
## 184         }
## 185     }
```

```
## 186     invisible(names)
## 187 }
```

#Details about the Number of Rows and Columns

```
print(paste("Number of records: ", nrow(Data)))
```

```
## [1] "Number of records:  205"
```

```
print(paste("Number of features: ", ncol(Data)))
```

```
## [1] "Number of features:  27"
```

#Summary of the data

```
summary(Data)
```

```
##      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      YearBuilt        price
##  Min.   :13.00   Min.   :16.00   Min.   :1996   Min.   : 5118
##  1st Qu.:19.00   1st Qu.:25.00   1st Qu.:2006   1st Qu.: 7788
##  Median :24.00   Median :30.00   Median :2010   Median :10295
##  Mean   :25.22   Mean   :30.75   Mean   :2010   Mean   :13277
##  3rd Qu.:30.00   3rd Qu.:34.00   3rd Qu.:2015   3rd Qu.:16503
##  Max.   :49.00   Max.   :54.00   Max.   :2021   Max.   :45400
##    CarName            fueltype          aspiration         doornumber
##  Length:205         Length:205         Length:205         Length:205
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##    carbody           drivewheel        enginelocation      enginetype
##  Length:205         Length:205         Length:205         Length:205
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##  cylindernumber      fuelsystem
##  Length:205         Length:205
##  Class :character   Class :character
##  Mode  :character   Mode  :character
##
##
##
```

#Columns names

```
colnames(Data)
```

```
##  [1] "car_ID"          "symboling"        "wheelbase"        "carlength"
##  [5] "carwidth"        "carheight"        "curbweight"       "enginesize"
##  [9] "boreratio"       "stroke"           "compressionratio" "horsepower"
## [13] "peakrpm"         "citympg"          "highwaympg"       "YearBuilt"
## [17] "price"           "CarName"          "fueltype"         "aspiration"
## [21] "doornumber"      "carbody"          "drivewheel"
"enginelocation"
## [25] "enginetype"      "cylindernumber"   "fuelsystem"
```

#Unique Carname

```
unique(Data$CarName)
```

```
##   [1] "alfa-romero giulia"             "alfa-romero stelvio"
##   [3] "alfa-romero Quadrifoglio"       "audi 100 ls"
##   [5] "audi 100ls"                     "audi fox"
##   [7] "audi 5000"                      "audi 4000"
##   [9] "audi 5000s (diesel)"            "bmw 320i"
##  [11] "bmw x1"                         "bmw x3"
##  [13] "bmw z4"                         "bmw x4"
##  [15] "bmw x5"                         "chevrolet impala"
##  [17] "chevrolet monte carlo"          "chevrolet vega 2300"
##  [19] "dodge rampage"                  "dodge challenger se"
##  [21] "dodge d200"                     "dodge monaco (sw)"
##  [23] "dodge colt hardtop"             "dodge colt (sw)"
##  [25] "dodge coronet custom"           "dodge dart custom"
##  [27] "dodge coronet custom (#sw)"      "honda civic"
##  [29] "honda civic cvcc"               "honda accord cvcc"
##  [31] "honda accord lx"                "honda civic 1500 gl"
##  [33] "honda accord"                   "honda civic 1300"
##  [35] "honda prelude"                  "honda civic (auto)"
##  [37] "isuzu MU-X"                     "isuzu D-Max "
##  [39] "isuzu D-Max V-Cross"            "jaguar xj"
##  [41] "jaguar xf"                      "jaguar xk"
##  [43] "maxda rx3"                      "maxda glc deluxe"
##  [45] "mazda rx2 coupe"                "mazda rx-4"
##  [47] "mazda glc deluxe"               "mazda 626"
##  [49] "mazda glc"                      "mazda rx-7 gs"
##  [51] "mazda glc 4"                    "mazda glc custom l"
##  [53] "mazda glc custom"               "buick electra 225 custom"
##  [55] "buick century luxus (sw)"       "buick century"
##  [57] "buick skyhawk"                  "buick opel isuzu deluxe"
##  [59] "buick skylark"                  "buick century special"
##  [61] "buick regal sport coupe (turbo)" "mercury cougar"
##  [63] "mitsubishi mirage"              "mitsubishi lancer"
##  [65] "mitsubishi outlander"           "mitsubishi g4"
##  [67] "mitsubishi mirage g4"           "mitsubishi montero"
##  [69] "mitsubishi pajero"              "Nissan versa"
##  [71] "nissan gt-r"                    "nissan rogue"
##  [73] "nissan latio"                   "nissan titan"
##  [75] "nissan leaf"                    "nissan juke"
##  [77] "nissan note"                    "nissan clipper"
##  [79] "nissan nv200"                   "nissan dayz"
##  [81] "nissan fuga"                    "nissan otti"
##  [83] "nissan teana"                   "nissan kicks"
##  [85] "peugeot 504"                    "peugeot 304"
##  [87] "peugeot 504 (sw)"               "peugeot 604sl"
```

```
##  [89] "peugeot 505s turbo diesel"       "plymouth fury iii"
##  [91] "plymouth cricket"                "plymouth satellite custom (sw)"
##  [93] "plymouth fury gran sedan"        "plymouth valiant"
##  [95] "plymouth duster"                 "porsche macan"
##  [97] "porcshce panamera"               "porsche cayenne"
##  [99] "porsche boxter"                  "renault 12tl"
## [101] "renault 5 gtl"                   "saab 99e"
## [103] "saab 99le"                       "saab 99gle"
## [105] "subaru"                          "subaru dl"
## [107] "subaru brz"                      "subaru baja"
## [109] "subaru r1"                       "subaru r2"
## [111] "subaru trezia"                   "subaru tribeca"
## [113] "toyota corona mark ii"           "toyota corona"
## [115] "toyota corolla 1200"             "toyota corona hardtop"
## [117] "toyota corolla 1600 (sw)"        "toyota carina"
## [119] "toyota mark ii"                  "toyota corolla"
## [121] "toyota corolla liftback"         "toyota celica gt liftback"
## [123] "toyota corolla tercel"           "toyota corona liftback"
## [125] "toyota starlet"                  "toyota tercel"
## [127] "toyota cressida"                 "toyota celica gt"
## [129] "toyouta tercel"                  "vokswagen rabbit"
## [131] "volkswagen 1131 deluxe sedan"    "volkswagen model 111"
## [133] "volkswagen type 3"               "volkswagen 411 (sw)"
## [135] "volkswagen super beetle"         "volkswagen dasher"
## [137] "vw dasher"                       "vw rabbit"
## [139] "volkswagen rabbit"               "volkswagen rabbit custom"
## [141] "volvo 145e (sw)"                 "volvo 144ea"
## [143] "volvo 244dl"                     "volvo 245"
## [145] "volvo 264gl"                     "volvo diesel"
## [147] "volvo 246"
```

#Feature Selection (Numeric columns are taken inro consideration)

```
maindf <- Data[,c("car_ID","symboling","wheelbase",
"carlength","carwidth","carheight","curbweight","enginesize","boreratio",
"stroke","compressionratio","horsepower",
"peakrpm","citympg","highwaympg",
"price")]
```

#Feature Selection (Non-Numeric columns are taken inro consideration)

```
maindf1<- Data[,c("CarName","fueltype","aspiration",
"carbody","drivewheel","enginelocation",
"price")]
```

#Details of selection

```
head(maindf)

##   car_ID symboling wheelbase carlength carwidth carheight curbweight
enginesize
## 1      1         3      88.6     168.8     64.1      48.8       2548
130
## 2      2         3      88.6     168.8     64.1      48.8       2548
130
## 3      3         1      94.5     171.2     65.5      52.4       2823
152
## 4      4         2      99.8     176.6     66.2      54.3       2337
109
## 5      5         2      99.4     176.6     66.4      54.3       2824
136
## 6      6         2      99.8     177.3     66.3      53.1       2507
136
##   boreratio stroke compressionratio horsepower peakrpm citympg highwaympg
price
## 1      3.47   2.68              9.0        111    5000      21         27
13495
## 2      3.47   2.68              9.0        111    5000      21         27
16500
## 3      2.68   3.47              9.0        154    5000      19         26
16500
## 4      3.19   3.40             10.0        102    5500      24         30
13950
## 5      3.19   3.40              8.0        115    5500      18         22
17450
## 6      3.19   3.40              8.5        110    5500      19         25
15250

head(maindf1)

##                    CarName fueltype aspiration     carbody drivewheel
## 1       alfa-romero giulia      gas        std convertible        rwd
## 2      alfa-romero stelvio      gas        std convertible        rwd
## 3 alfa-romero Quadrifoglio      gas        std   hatchback        rwd
## 4              audi 100 ls      gas        std       sedan        fwd
## 5               audi 100ls      gas        std       sedan        4wd
## 6                 audi fox      gas        std       sedan        fwd
##   enginelocation price
## 1          front 13495
## 2          front 16500
## 3          front 16500
## 4          front 13950
## 5          front 17450
## 6          front 15250
```

#Checking for null values

```
sum(is.na(maindf1))
```

```
## [1] 0
```

```
sum(is.na(maindf))
```

```
## [1] 0
```

#Correlation Matrix

```
#Plot Correlation matrix
cor(maindf)
```

```
##                      car_ID    symboling  wheelbase  carlength
carwidth
## car_ID            1.00000000 -0.151621137  0.1297288  0.1706364
0.05238661
## symboling        -0.15162114  1.000000000 -0.5319537 -0.3576115 -
0.23291906
## wheelbase         0.12972878 -0.531953682  1.0000000  0.8745875
0.79514364
## carlength         0.17063639 -0.357611523  0.8745875  1.0000000
0.84111827
## carwidth          0.05238661 -0.232919061  0.7951436  0.8411183
1.00000000
## carheight         0.25596004 -0.541038200  0.5894348  0.4910295
0.27921032
## curbweight        0.07196156 -0.227690588  0.7763863  0.8777285
0.86703246
## enginesize       -0.03392984 -0.105789709  0.5693287  0.6833599
0.73543340
## boreratio         0.26006368 -0.130051360  0.4887499  0.6064544
0.55914991
## stroke           -0.16082362 -0.008735141  0.1609590  0.1295326
0.18294169
## compressionratio  0.15027591 -0.178515084  0.2497858  0.1584137
0.18112863
## horsepower       -0.01500557  0.070872724  0.3532945  0.5526230
0.64073208
## peakrpm          -0.20378920  0.273606245 -0.3604687 -0.2872422 -
0.22001230
## citympg           0.01594004 -0.035822628 -0.4704136 -0.6709087 -
0.64270434
## highwaympg        0.01125532  0.034606001 -0.5440819 -0.7046616 -
0.67721792
## price            -0.10909334 -0.079978225  0.5778156  0.6829200
0.75932530
##                    carheight  curbweight  enginesize    boreratio
```

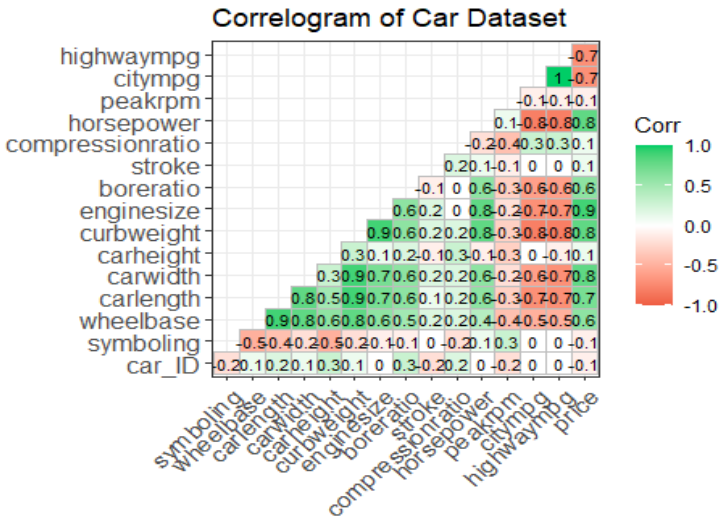
```
stroke
## car_ID            0.25596004  0.07196156 -0.03392984  0.260063680 -
0.160823619
## symboling        -0.54103820 -0.22769059 -0.10578971 -0.130051360 -
0.008735141
## wheelbase         0.58943476  0.77638633  0.56932868  0.488749875
0.160959047
## carlength         0.49102946  0.87772846  0.68335987  0.606454358
0.129532611
## carwidth          0.27921032  0.86703246  0.73543340  0.559149909
0.182941693
## carheight         1.00000000  0.29557173  0.06714874  0.171070922 -
0.055306674
## curbweight        0.29557173  1.00000000  0.85059407  0.648479749
0.168790035
## enginesize        0.06714874  0.85059407  1.00000000  0.583774327
0.203128588
## boreratio         0.17107092  0.64847975  0.58377433  1.000000000 -
0.055908983
## stroke           -0.05530667  0.16879004  0.20312859 -0.055908983
1.000000000
## compressionratio  0.26121423  0.15136174  0.02897136  0.005197339
0.186110110
## horsepower       -0.10880206  0.75073925  0.80976865  0.573676823
0.080939536
## peakrpm          -0.32041072 -0.26624318 -0.24465983 -0.254975528 -
0.067963753
## citympg          -0.04863963 -0.75741378 -0.65365792 -0.584531716 -
0.042144754
## highwaympg       -0.10735763 -0.79746479 -0.67746991 -0.587011784 -
0.043930930
## price             0.11933623  0.83530488  0.87414480  0.553173237
0.079443084
##                  compressionratio  horsepower     peakrpm     citympg
## car_ID               0.150275906 -0.01500557 -0.20378920  0.01594004
## symboling           -0.178515084  0.07087272  0.27360625 -0.03582263
## wheelbase            0.249785845  0.35329448 -0.36046875 -0.47041361
## carlength            0.158413706  0.55262297 -0.28724220 -0.67090866
## carwidth             0.181128627  0.64073208 -0.22001230 -0.64270434
## carheight            0.261214226 -0.10880206 -0.32041072 -0.04863963
## curbweight           0.151361740  0.75073925 -0.26624318 -0.75741378
## enginesize           0.028971360  0.80976865 -0.24465983 -0.65365792
## boreratio            0.005197339  0.57367682 -0.25497553 -0.58453172
## stroke               0.186110110  0.08093954 -0.06796375 -0.04214475
## compressionratio     1.000000000 -0.20432623 -0.43574051  0.32470142
## horsepower          -0.204326226  1.00000000  0.13107251 -0.80145618
## peakrpm             -0.435740514  0.13107251  1.00000000 -0.11354438
## citympg              0.324701425 -0.80145618 -0.11354438  1.00000000
## highwaympg           0.265201389 -0.77054389 -0.05427481  0.97133704
## price                0.067983506  0.80813882 -0.08526715 -0.68575134
```

```
##                   highwaympg       price
## car_ID             0.01125532 -0.10909334
## symboling          0.03460600 -0.07997822
## wheelbase         -0.54408192  0.57781560
## carlength         -0.70466160  0.68292002
## carwidth          -0.67721792  0.75932530
## carheight         -0.10735763  0.11933623
## curbweight        -0.79746479  0.83530488
## enginesize        -0.67746991  0.87414480
## boreratio         -0.58701178  0.55317324
## stroke            -0.04393093  0.07944308
## compressionratio   0.26520139  0.06798351
## horsepower        -0.77054389  0.80813882
## peakrpm           -0.05427481 -0.08526715
## citympg            0.97133704 -0.68575134
## highwaympg         1.00000000 -0.69759909
## price             -0.69759909  1.00000000
```

```
library(ggcorrplot)
corr <- round(cor(maindf), 1)
```

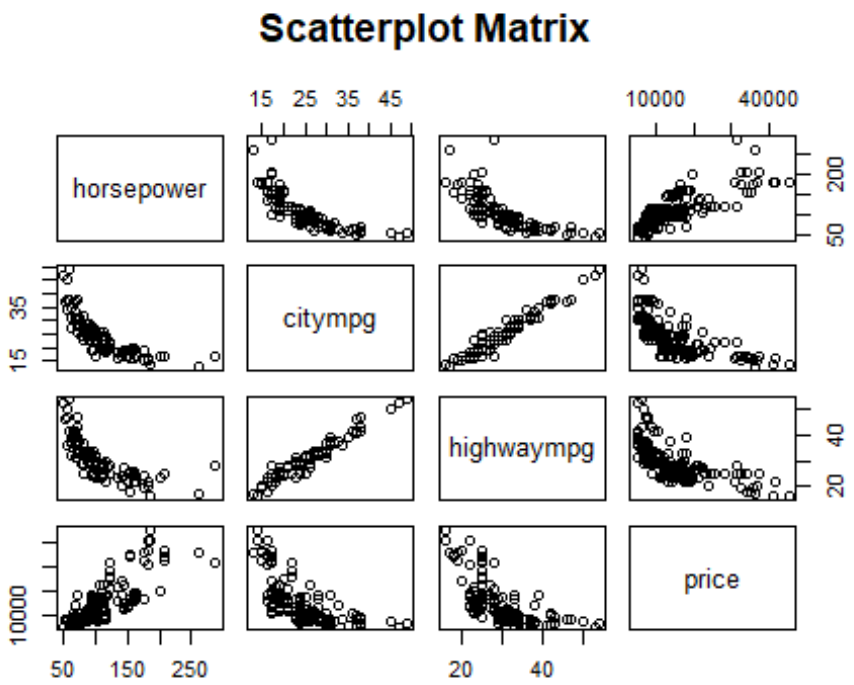
#Plot

```
ggcorrplot(corr,
           type = "lower",
           lab = TRUE,
           lab_size = 3,
           colors = c("tomato2", "white", "springgreen3"),
           title="Correlogram of Car Dataset",
           ggtheme=theme_bw)
```



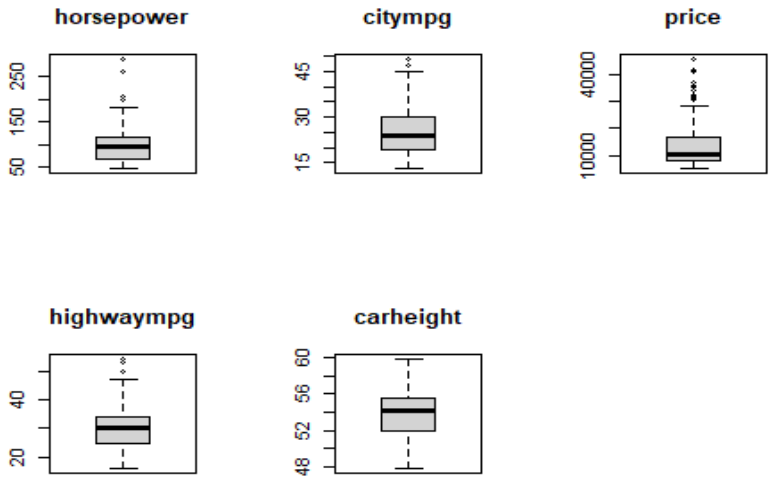
#Plot scatterplot matrix

```
pairs(~ horsepower + citympg + highwaympg + price, data = maindf,
       main = "Scatterplot Matrix")
```



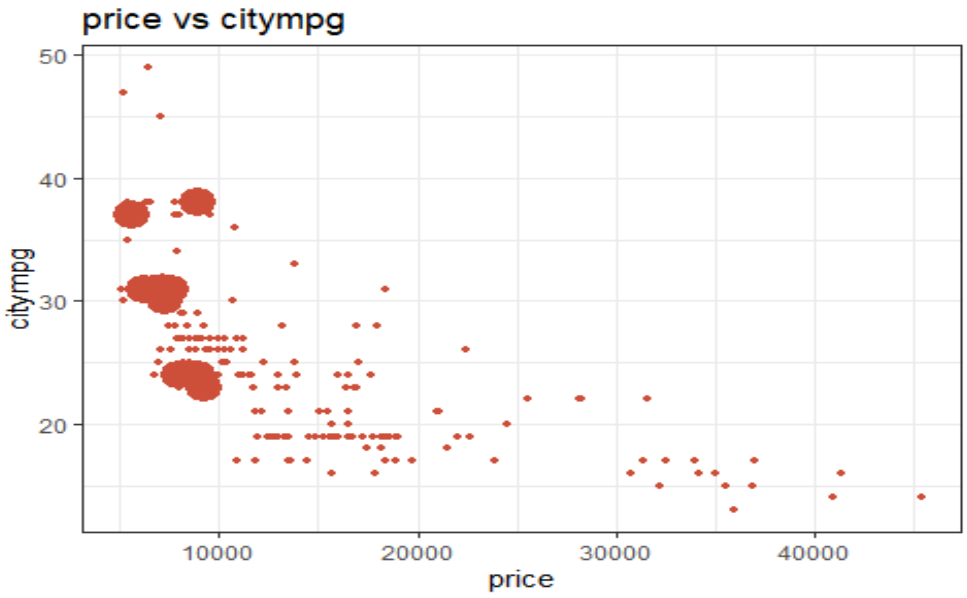
#Plot boxplot for checking outliers

```
par(mfrow=c(2, 3))  # divide graph area in 2 columns
boxplot(maindf$horsepower, main="horsepower")
boxplot(maindf$citympg, main="citympg")
boxplot(maindf$price, main="price")
boxplot(maindf$highwaympg, main="highwaympg")
boxplot(maindf$carheight, main="carheight")
```



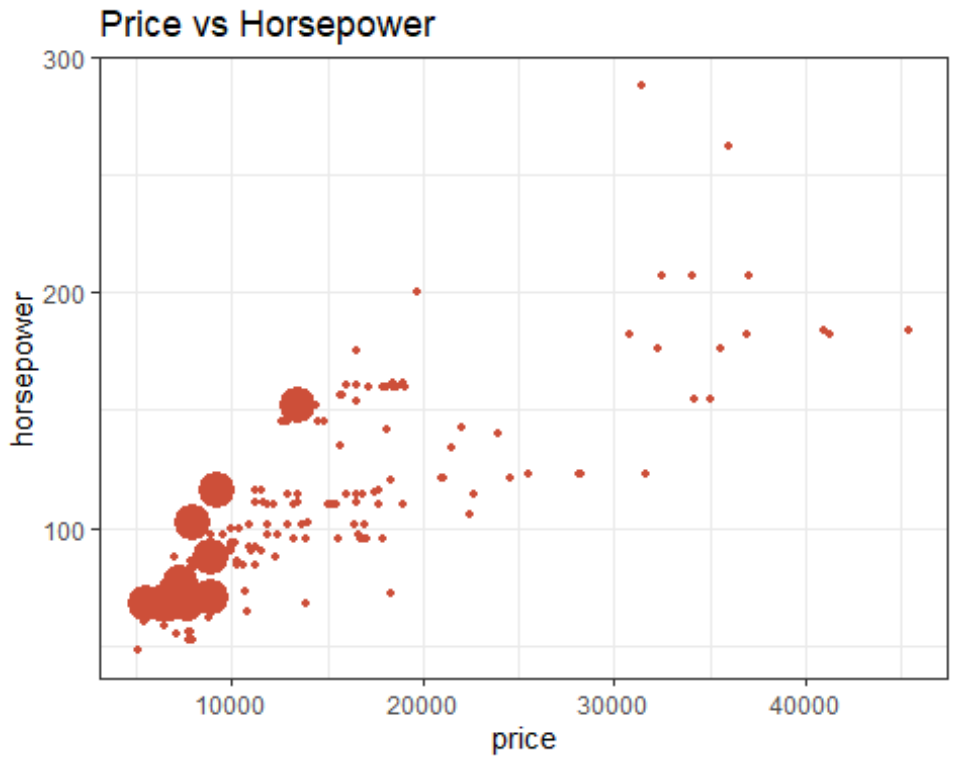
#Scatterplot(price vs citympg)

```
theme_set(theme_bw())
g <- ggplot(maindf, aes(price,citympg))
g + geom_count(col="tomato3", show.legend=F) +
  labs(y="citympg",
       x="price",
       title="price vs citympg")
```



#Scatterplot(price vs horsepower)

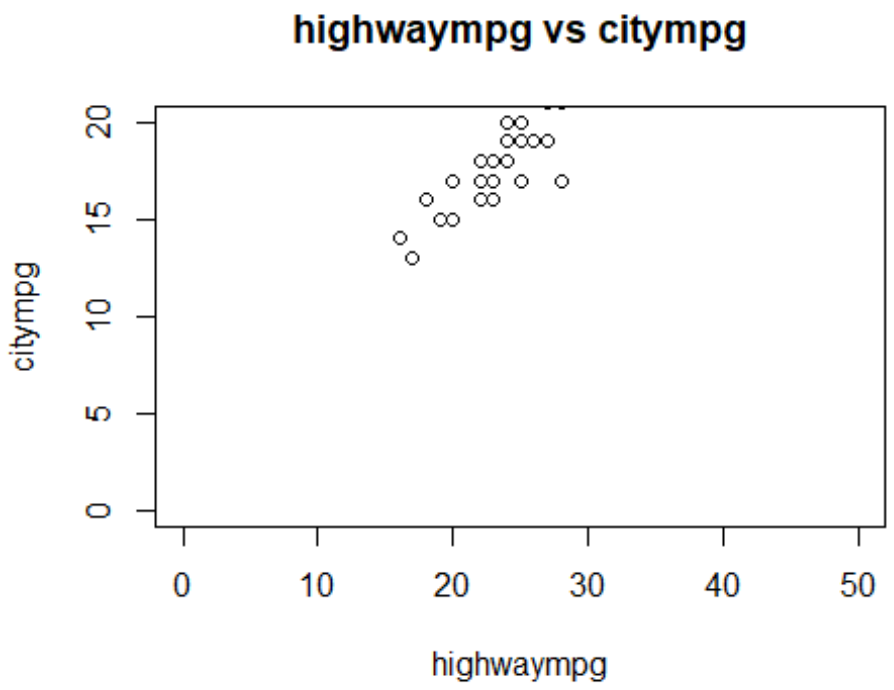
```
theme_set(theme_bw())
g <- ggplot(Data, aes(price, horsepower))
g + geom_count(col="tomato3", show.legend=F) +
  labs(y="horsepower",
       x="price",
       title="Price vs Horsepower")
```



In a similar way we can show the relationship of various factors with Price

#Scatterplot(highwaympg vs citympg)

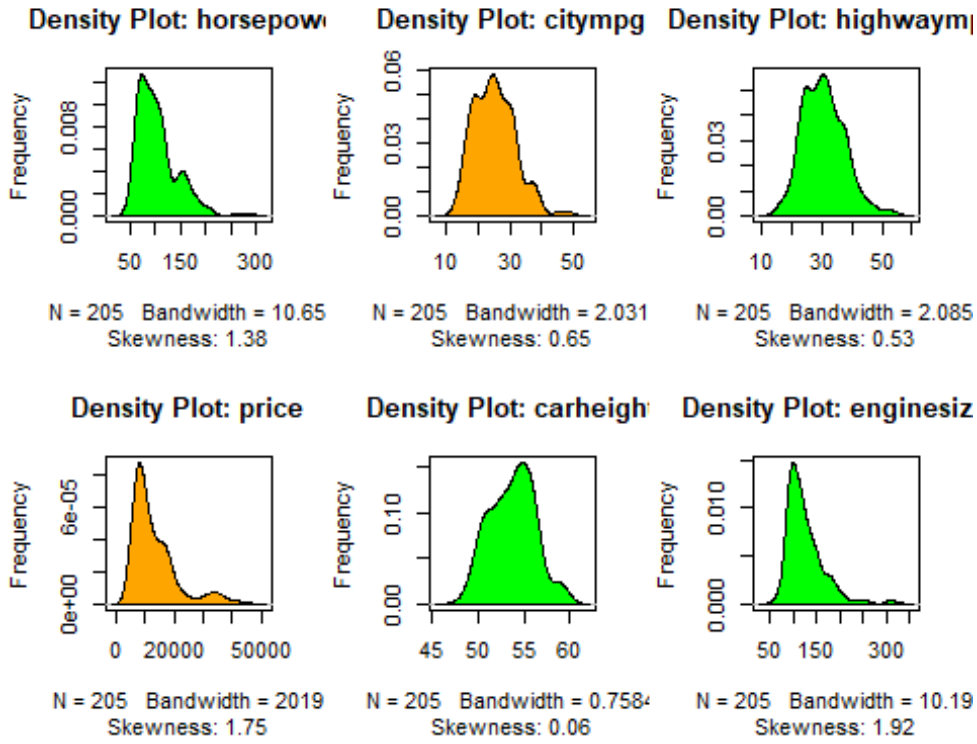
```
plot(x = maindf$highwaympg, y = maindf$citympg,
     xlab = "highwaympg",
     ylab = "citympg",
     xlim = c(0, 50),
     ylim = c(0, 20),
     main = "highwaympg vs citympg",
)
```



#Plot density plot to check normality

```
library(e1071)

par(mfrow=c(2, 3))

plot(density(maindf$horsepower), main="Density Plot: horsepower",
ylab="Frequency",
     sub=paste("Skewness:", round(e1071::skewness(maindf$horsepower), 2)))
polygon(density(maindf$horsepower), col="green")

plot(density(maindf$citympg), main="Density Plot: citympg", ylab="Frequency",
     sub=paste("Skewness:", round(e1071::skewness(maindf$citympg), 2)))
polygon(density(maindf$citympg), col="orange")

plot(density(maindf$highwaympg), main="Density Plot: highwaympg",
ylab="Frequency",
     sub=paste("Skewness:", round(e1071::skewness(maindf$highwaympg), 2)))
polygon(density(maindf$highwaympg), col="green")

plot(density(maindf$price), main="Density Plot: price", ylab="Frequency",
```

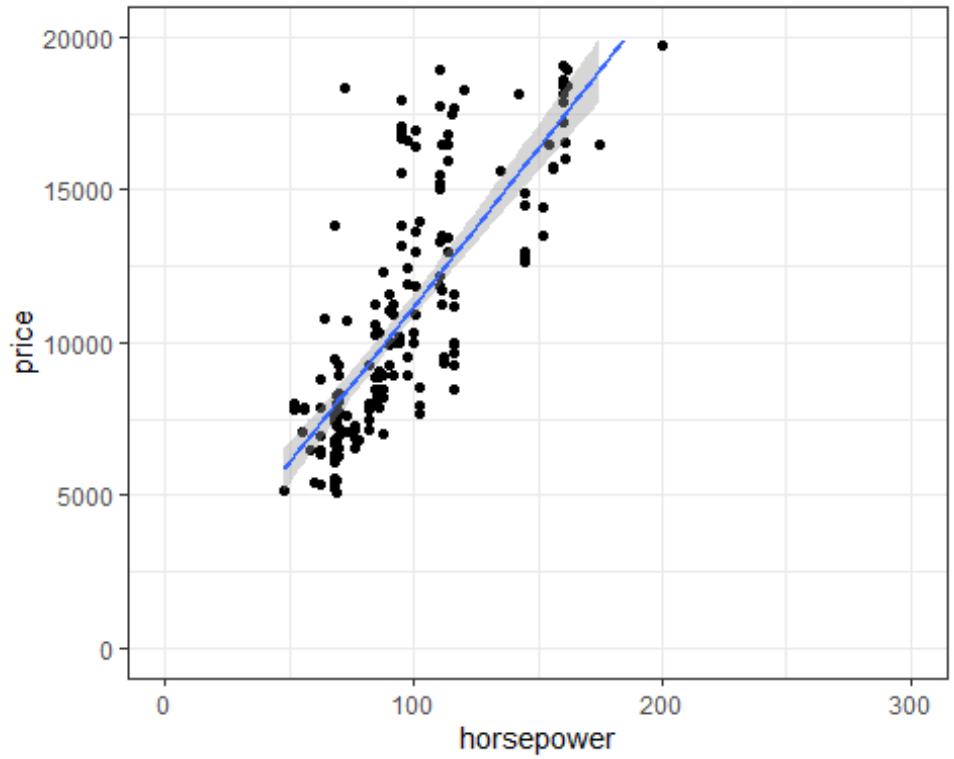
```
     sub=paste("Skewness:", round(e1071::skewness(maindf$price), 2)))
polygon(density(maindf$price), col="orange")

plot(density(maindf$carheight), main="Density Plot: carheight",
ylab="Frequency",
     sub=paste("Skewness:", round(e1071::skewness(maindf$carheight), 2)))
polygon(density(maindf$carheight), col="green")

plot(density(maindf$enginesize), main="Density Plot: enginesize",
ylab="Frequency",
     sub=paste("Skewness:", round(e1071::skewness(maindf$enginesize), 2)))
polygon(density(maindf$eng), col="green")
```



#Plot univariate linear regression between horsepower and price

```
ggplot(maindf,aes(y=price,x=horsepower)) +
       geom_point() +
        xlim(0, 300) +
```

```
        ylim(0, 20000) +
        geom_smooth(formula = y ~ x,method="lm")
```



#In a similar way other can also be done

#Factorizing

```
Data$CarName<- factor(Data$CarName)
Data$fueltype<- factor(Data$fueltype)
Data$aspiration<- factor(Data$aspiration)
Data$doornumber<- factor(Data$doornumber)
Data$carbody<- factor(Data$carbody)
Data$drivewheel<- factor(Data$drivewheel)
Data$enginelocation<- factor(Data$enginelocation)
Data$enginetype<- factor(Data$enginetype)
Data$cylinernumber<- factor(Data$cylindernumber)
Data$fuelsystem<- factor(Data$fuelsystem)
```

#SINGLE MODEL REGRESSION (Non-Numeric) #When we take in consideration fuel type the R2 is only 1%

```
linearmodel_S1 = lm(price~fueltype,
                 data = maindf1)
summary(linearmodel_S1)
```

```
##
## Call:
## lm(formula = price ~ fueltype, data = maindf1)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -8739  -5391  -2802   2998  32400
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    15838       1781   8.894 3.25e-16 ***
## fueltypegas    -2838       1874  -1.514    0.132
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7964 on 203 degrees of freedom
## Multiple R-squared:  0.01117,    Adjusted R-squared:  0.006297
## F-statistic: 2.293 on 1 and 203 DF,  p-value: 0.1315
```

#MULTIPLE MODEL REGRESSION #Try to improve your model by additionally including the other aspects. Does this improve the model accuracy?

```
linearmodel_M1 = lm(price~fueltype+CarName,
                 data = maindf1)
summary(linearmodel_M1)
```

```
##
## Call:
## lm(formula = price ~ fueltype + CarName, data = maindf1)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -7888      0      0      0   7888
##
## Coefficients:
##                                       Estimate Std. Error t value
Pr(>|t|)
## (Intercept)                          13751.11    3545.29   3.879
0.000275 ***
## fueltypegas                           -256.11    1511.72  -0.169
0.866068
## CarNamealfa-romero Quadrifoglio        3005.00    4535.15   0.663
```

```
0.510256
## CarNamealfa-romero stelvio              3005.00    4535.15   0.663
0.510256
## CarNameaudi 100 ls                       455.00    4535.15   0.100
0.920436
## CarNameaudi 100ls                       4085.00    3927.55   1.040
0.302691
## CarNameaudi 4000                       10380.00    4535.15   2.289
0.025818 *
## CarNameaudi 5000                        5425.00    4535.15   1.196
0.236566
## CarNameaudi 5000s (diesel)              4364.17    4535.15   0.962
0.339964
## CarNameaudi fox                         1755.00    4535.15   0.387
0.700214
## CarNamebmw 320i                         3182.50    3927.55   0.810
0.421137
## CarNamebmw x1                           7475.00    4535.15   1.648
0.104806
## CarNamebmw x3                          15497.50    3927.55   3.946
0.000221 ***
## CarNamebmw x4                          17265.00    4535.15   3.807
0.000346 ***
## CarNamebmw x5                          27820.00    4535.15   6.134 8.66e-
08 ***
## CarNamebmw z4                          11070.00    4535.15   2.441
0.017779 *
## CarNamebuick century                   14424.89    4780.47   3.017
0.003806 **
## CarNamebuick century luxus (sw)        14496.89    4780.47   3.033
0.003646 **
## CarNamebuick century special           27465.00    4535.15   6.056 1.16e-
07 ***
## CarNamebuick electra 225 custom        11800.89    4780.47   2.469
0.016588 *
## CarNamebuick opel isuzu deluxe         20689.00    4535.15   4.562 2.74e-
05 ***
## CarNamebuick regal sport coupe (turbo) 31905.00    4535.15   7.035 2.78e-
09 ***
## CarNamebuick skyhawk                   17848.89    4780.47   3.734
0.000438 ***
## CarNamebuick skylark                   21561.00    4535.15   4.754 1.39e-
05 ***
## CarNamechevrolet impala                -8344.00    4535.15  -1.840
0.070999 .
## CarNamechevrolet monte carlo           -7200.00    4535.15  -1.588
0.117909
## CarNamechevrolet vega 2300             -6920.00    4535.15  -1.526
0.132575
## CarNamedodge challenger se             -7118.00    4535.15  -1.570
```

```
0.122062
## CarNamedodge colt (sw)                  -5886.00    4535.15  -1.298
0.199562
## CarNamedodge colt hardtop               -6803.00    4535.15  -1.500
0.139118
## CarNamedodge coronet custom             -4937.00    4535.15  -1.089
0.280908
## CarNamedodge coronet custom (sw)         -531.00    4535.15  -0.117
0.907204
## CarNamedodge d200                       -5538.00    4535.15  -1.221
0.227064
## CarNamedodge dart custom                -4574.00    4535.15  -1.009
0.317445
## CarNamedodge monaco (sw)                -7266.00    4535.15  -1.602
0.114650
## CarNamedodge rampage                    -7923.00    4535.15  -1.747
0.086020 .
## CarNamehonda accord                     -4400.00    3927.55  -1.120
0.267288
## CarNamehonda accord cvcc                -6966.00    4535.15  -1.536
0.130070
## CarNamehonda accord lx                  -6200.00    4535.15  -1.367
0.176962
## CarNamehonda civic                      -5220.67    3702.93  -1.410
0.164010
## CarNamehonda civic (auto)               -3150.00    4535.15  -0.695
0.490143
## CarNamehonda civic 1300                 -4400.00    4535.15  -0.970
0.336046
## CarNamehonda civic 1500 gl              -6200.00    4535.15  -1.367
0.176962
## CarNamehonda civic cvcc                 -6503.00    3927.55  -1.656
0.103271
## CarNamehonda prelude                    -4650.00    4535.15  -1.025
0.309542
## CarNameisuzu D-Max                      -3512.75    3927.55  -0.894
0.374876
## CarNameisuzu D-Max V-Cross              -4578.50    4535.15  -1.010
0.316973
## CarNameisuzu MU-X                       -6710.00    4535.15  -1.480
0.144498
## CarNamejaguar xf                        22055.00    4535.15   4.863 9.47e-
06 ***
## CarNamejaguar xj                        18755.00    4535.15   4.135
0.000118 ***
## CarNamejaguar xk                        22505.00    4535.15   4.962 6.64e-
06 ***
## CarNamemaxda glc deluxe                 -7400.00    4535.15  -1.632
0.108256
## CarNamemaxda rx3                        -8300.00    4535.15  -1.830
```

```
0.072458 .
## CarNamemazda 626                       -3150.00    3702.93  -0.851
0.398508
## CarNamemazda glc                        1567.50    3927.55   0.399
0.691308
## CarNamemazda glc 4                      2150.00    4535.15   0.474
0.637258
## CarNamemazda glc custom                -2900.00    4535.15  -0.639
0.525092
## CarNamemazda glc custom l              -5000.00    4535.15  -1.102
0.274878
## CarNamemazda glc deluxe                -4528.06    3999.62  -1.132
0.262324
## CarNamemazda rx-4                      -5025.00    3927.55  -1.279
0.205932
## CarNamemazda rx-7 gs                    2371.44    3999.62   0.593
0.555582
## CarNamemazda rx2 coupe                 -6700.00    4535.15  -1.477
0.145086
## CarNamemercury cougar                   3008.00    4535.15   0.663
0.509835
## CarNamemitsubishi g4                   -3142.67    3702.93  -0.849
0.399600
## CarNamemitsubishi lancer               -7306.00    4535.15  -1.611
0.112710
## CarNamemitsubishi mirage               -8106.00    4535.15  -1.787
0.079194 .
## CarNamemitsubishi mirage g4            -2252.67    3702.93  -0.608
0.545374
## CarNamemitsubishi montero              -6506.00    4535.15  -1.435
0.156873
## CarNamemitsubishi outlander            -3969.33    3702.93  -1.072
0.288263
## CarNamemitsubishi pajero               -5306.00    4535.15  -1.170
0.246881
## CarNamenissan clipper                   -171.00    3927.55  -0.044
0.965424
## CarNamenissan dayz                         4.00    4535.15   0.001
0.999299
## CarNamenissan fuga                       904.00    4535.15   0.199
0.842712
## CarNamenissan gt-r                     -6652.11    4780.47  -1.392
0.169475
## CarNamenissan juke                     -5696.00    4535.15  -1.256
0.214251
## CarNamenissan kicks                     6204.00    4535.15   1.368
0.176687
## CarNamenissan latio                    -6321.00    3927.55  -1.609
0.113054
## CarNamenissan leaf                     -6196.00    4535.15  -1.366
```

```
0.177237
## CarNamenissan note                      -5496.00    4535.15  -1.212
0.230563
## CarNamenissan nv200                     -3946.00    4535.15  -0.870
0.387899
## CarNamenissan otti                         4.00    4535.15   0.001
0.999299
## CarNamenissan rogue                     -5696.00    3927.55  -1.450
0.152465
## CarNamenissan teana                      3704.00    4535.15   0.817
0.417482
## CarNamenissan titan                     -6146.00    4535.15  -1.355
0.180702
## CarNameNissan versa                     -7996.00    4535.15  -1.763
0.083241 .
## CarNamepeugeot 304                       -551.11    4780.47  -0.115
0.908625
## CarNamepeugeot 504                       1855.46    3500.24   0.530
0.598104
## CarNamepeugeot 504 (sw)                 -1055.00    4535.15  -0.233
0.816884
## CarNamepeugeot 505s turbo diesel         3323.89    4780.47   0.695
0.489688
## CarNamepeugeot 604sl                     3901.94    3999.62   0.976
0.333396
## CarNameplymouth cricket                 -5538.00    4535.15  -1.221
0.227064
## CarNameplymouth duster                   -731.00    4535.15  -0.161
0.872517
## CarNameplymouth fury gran sedan         -5886.00    4535.15  -1.298
0.199562
## CarNameplymouth fury iii                -7594.50    3927.55  -1.934
0.058128 .
## CarNameplymouth satellite custom (sw)   -6803.00    4535.15  -1.500
0.139118
## CarNameplymouth valiant                 -4574.00    4535.15  -1.009
0.317445
## CarNameporcshce panamera                19033.00    4535.15   4.197 9.58e-
05 ***
## CarNameporsche boxter                   23533.00    4535.15   5.189 2.92e-
06 ***
## CarNameporsche cayenne                  19219.25    3927.55   4.893 8.50e-
06 ***
## CarNameporsche macan                     8523.00    4535.15   1.879
0.065316 .
## CarNamerenault 12tl                     -4200.00    4535.15  -0.926
0.358299
## CarNamerenault 5 gtl                    -3600.00    4535.15  -0.794
0.430604
## CarNamesaab 99e                          1740.00    3927.55   0.443
```

```
0.659425
## CarNamesaab 99gle                         3335.00    3927.55   0.849
0.399362
## CarNamesaab 99le                           110.00    3927.55   0.028
0.977754
## CarNamesubaru                            -7373.00    3927.55  -1.877
0.065605 .
## CarNamesubaru baja                       -3535.00    4535.15  -0.779
0.438929
## CarNamesubaru brz                        -5720.00    4535.15  -1.261
0.212353
## CarNamesubaru dl                         -4904.25    3585.35  -1.368
0.176725
## CarNamesubaru r1                         -4262.00    4535.15  -0.940
0.351301
## CarNamesubaru r2                         -2236.00    4535.15  -0.493
0.623880
## CarNamesubaru trezia                     -6032.00    4535.15  -1.330
0.188798
## CarNamesubaru tribeca                    -3297.00    4535.15  -0.727
0.470208
## CarNametoyota carina                     -4717.00    4535.15  -1.040
0.302686
## CarNametoyota celica gt                  -3053.11    4780.47  -0.639
0.525600
## CarNametoyota celica gt liftback         -4197.00    4535.15  -0.925
0.358640
## CarNametoyota corolla                    -4007.68    3472.93  -1.154
0.253325
## CarNametoyota corolla 1200               -6652.00    3927.55  -1.694
0.095786 .
## CarNametoyota corolla 1600 (sw)          -5597.00    4535.15  -1.234
0.222216
## CarNametoyota corolla liftback           -1187.00    3927.55  -0.302
0.763582
## CarNametoyota corolla tercel             -3957.00    4535.15  -0.873
0.386586
## CarNametoyota corona                     -4171.35    3472.93  -1.201
0.234678
## CarNametoyota corona hardtop             -6577.00    4535.15  -1.450
0.152476
## CarNametoyota corona liftback            -5046.00    4535.15  -1.113
0.270532
## CarNametoyota corona mark ii             -8147.00    4535.15  -1.796
0.077729 .
## CarNametoyota cressida                    4174.00    4535.15   0.920
0.361260
## CarNametoyota mark ii                    -4347.00    3702.93  -1.174
0.245303
## CarNametoyota starlet                     -655.50    3927.55  -0.167
```

```
0.868041
## CarNametoyota tercel                    -2296.00    4535.15  -0.506
0.614621
## CarNametoyouta tercel                    2255.00    4535.15   0.497
0.620941
## CarNamevokswagen rabbit                 -5976.11    4780.47  -1.250
0.216367
## CarNamevolkswagen 1131 deluxe sedan     -5520.00    4535.15  -1.217
0.228559
## CarNamevolkswagen 411 (sw)              -5000.00    4535.15  -1.102
0.274878
## CarNamevolkswagen dasher                -2352.50    3927.55  -0.599
0.551564
## CarNamevolkswagen model 111             -5756.11    4780.47  -1.204
0.233531
## CarNamevolkswagen rabbit                 -200.00    4535.15  -0.044
0.964979
## CarNamevolkswagen rabbit custom            93.89    4780.47   0.020
0.984399
## CarNamevolkswagen super beetle          -4256.11    4780.47  -0.890
0.377040
## CarNamevolkswagen type 3                -5300.00    4535.15  -1.169
0.247409
## CarNamevolvo 144ea                       2735.00    3927.55   0.696
0.489032
## CarNamevolvo 145e (sw)                   1397.50    3927.55   0.356
0.723289
## CarNamevolvo 244dl                       5240.00    3927.55   1.334
0.187456
## CarNamevolvo 245                         3020.00    4535.15   0.666
0.508155
## CarNamevolvo 246                         8718.89    4780.47   1.824
0.073418 .
## CarNamevolvo 264gl                       7027.50    3927.55   1.789
0.078883 .
## CarNamevolvo diesel                      5455.00    4535.15   1.203
0.234016
## CarNamevw dasher                        -1900.00    4535.15  -0.419
0.676827
## CarNamevw rabbit                        -3515.00    4535.15  -0.775
0.441509
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3207 on 57 degrees of freedom
## Multiple R-squared:  0.955,  Adjusted R-squared:  0.8389
## F-statistic: 8.225 on 147 and 57 DF,  p-value: 1.855e-15
```

#As we can see this features are best considerating the above dataset as the accuracy is 95%

#SINGLE MODEL REGRESSION (Numeric) #When we take in consideration carheight type the R2 is only 1%

```
linearmodel_S2 = lm(price~carheight,
                 data = maindf)
summary(linearmodel_S2)
```

```
##
## Call:
## lm(formula = price ~ carheight, data = maindf)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -8456  -5466  -2778   2876  31470
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -7684.5    12252.6  -0.627   0.5313
## carheight      390.2      227.8   1.713   0.0883 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7951 on 203 degrees of freedom
## Multiple R-squared:  0.01424,    Adjusted R-squared:  0.009385
## F-statistic: 2.933 on 1 and 203 DF,  p-value: 0.08833
```

#MULTIPLE MODEL REGRESSION #Try to improve your model by additionally including the other aspects. Does this improve the model accuracy? #Building the Model

```
linearmodel_M2 = lm(price~carheight + carwidth + wheelbase,
                 data = maindf)
summary(linearmodel_M2)
```

```
##
## Call:
## lm(formula = price ~ carheight + carwidth + wheelbase, data = maindf)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -10664  -2811  -1086   1013  26072
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -157032.42   15993.28  -9.819   <2e-16 ***
## carheight      -393.20     199.08  -1.975   0.0496 *
```

```
## carwidth       2806.02     302.07   9.289   <2e-16 ***
## wheelbase        65.77     127.91   0.514   0.6077
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5176 on 201 degrees of freedom
## Multiple R-squared:  0.5864, Adjusted R-squared:  0.5803
## F-statistic: 95.01 on 3 and 201 DF,  p-value: < 2.2e-16
```

#here the R2 is 58%

#Further Improvement

```
linearmodel_M3 = lm(price~horsepower + citympg + highwaympg + carheight +
enginesize,
                data = maindf)
summary(linearmodel_M3)

##
## Call:
## lm(formula = price ~ horsepower + citympg + highwaympg + carheight +
##     enginesize, data = maindf)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9826.5 -1986.5   -75.6  1401.9 13504.4
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -24136.12    7204.46  -3.350 0.000966 ***
## horsepower      61.20      14.83   4.127 5.39e-05 ***
## citympg        -13.76     180.89  -0.076 0.939444
## highwaympg     -54.77     164.83  -0.332 0.740000
## carheight      351.72     110.79   3.175 0.001738 **
## enginesize     111.70      11.10  10.066  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3551 on 199 degrees of freedom
## Multiple R-squared:  0.8073, Adjusted R-squared:  0.8025
## F-statistic: 166.7 on 5 and 199 DF,  p-value: < 2.2e-16
```

#As we can see this features are best considerating the above dataset as the accuracy is 80%

#CROSS CHEACKING

```
aov(linearmodel_M1)
```

```
## Call:
##    aov(formula = linearmodel_M1)
##
## Terms:
##                     fueltype     CarName   Residuals
## Sum of Squares    145405324 12288058297   586175740
## Deg. of Freedom           1         146          57
##
## Residual standard error: 3206.834
## Estimated effects may be unbalanced
```

```
aov(linearmodel_M2)
```

```
## Call:
##    aov(formula = linearmodel_M2)
##
## Terms:
##                   carheight   carwidth  wheelbase  Residuals
## Sum of Squares   185414441 7442659231    7082610 5384483080
## Deg. of Freedom          1          1          1        201
##
## Residual standard error: 5175.758
## Estimated effects may be unbalanced
```

```
aov(linearmodel_M3)
```

```
## Call:
##    aov(formula = linearmodel_M3)
##
## Terms:
##                  horsepower    citympg highwaympg  carheight enginesize
## Sum of Squares  8502974873   52739498  320599014  356824677 1277574992
## Deg. of Freedom          1          1          1          1          1
##                   Residuals
## Sum of Squares   2508926308
## Deg. of Freedom         199
##
## Residual standard error: 3550.728
## Estimated effects may be unbalanced
```

#After using ANOVA we can easily identify which factors influence the most (horsepower, enginesize,carwidth,carheight,highwaympg)

#Building the best Model(The accuracy here is 82% which is relatively good)

```
linearmodel_M4 = lm(price~horsepower + carwidth + carheight + highwaympg +
enginesize,
                data = maindf)
summary(linearmodel_M4)
```

```
##
## Call:
## lm(formula = price ~ horsepower + carwidth + carheight + highwaympg +
##     enginesize, data = maindf)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8445.5 -2082.5   -57.5  1309.7 14262.1
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -62948.629  12401.335  -5.076 8.83e-07 ***
## horsepower      59.590     13.039   4.570 8.55e-06 ***
## carwidth       710.471    187.883   3.781 0.000206 ***
## carheight      212.564    112.828   1.884 0.061027 .
## highwaympg      -3.765     59.858  -0.063 0.949906
## enginesize      93.700     11.246   8.332 1.28e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3430 on 199 degrees of freedom
## Multiple R-squared:  0.8202, Adjusted R-squared:  0.8157
## F-statistic: 181.6 on 5 and 199 DF,  p-value: < 2.2e-16
```