# Code Explanation

**Code :**

head(mydata)

summary(mydata)

sapply(mydata, function(x) sum(is.na(x)))

**Description:**

Above lines of code show the data’s 1st 6 row, Summary of data which include number of different values repeated of categorical variables, mix/max values of numeric variable with quartile and means. Sapply shows how many missing values we have in dataset.

**Code:**

xtabs(~price + make , data = mydata)

xtabs(~price + fuel.type, data = mydata)

xtabs(~price + aspiration , data = mydata)

xtabs(~price + num.of.doors, data = mydata)

xtabs(~price + body.style, data = mydata)

xtabs(~price + drive.wheels, data = mydata)

xtabs(~price + engine.location, data = mydata)

xtabs(~price + engine.type, data = mydata)

xtabs(~price +num.of.cylinders, data = mydata)

xtabs(~price +fuel.system, data = mydata)

mydata$make <- factor(mydata$make)

mydata$fuel.type <- factor(mydata$fuel.type)

mydata$aspiration <- factor(mydata$aspiration)

mydata$num.of.doors <- factor(mydata$num.of.doors)

mydata$body.style <- factor(mydata$body.style)

mydata$drive.wheels <- factor(mydata$drive.wheels)

mydata$engine.location <- factor(mydata$engine.location)

mydata$engine.type <- factor(mydata$engine.type)

mydata$num.of.cylinders <- factor(mydata$num.of.cylinders)

mydata$fuel.system <- factor(mydata$fuel.system)

**Description:**

The above line of “xtabs” code shows relationship two categorical variables i.e. independent variable relationship with dependent variable “price”

The “factor” code converts categorical variables into factors.

**Code:**

mylogit <- glm(price ~ make + fuel.type + aspiration + num.of.doors + body.style + drive.wheels + engine.location + engine.type + num.of.cylinders + fuel.system+ wheel.base+ length + height +curb.weight+engine.size+ bore+ stroke+ compression.ratio+ horsepower+ peak.rpm+city.mpg+ highway.mpg, data = mydata, family = "binomial")

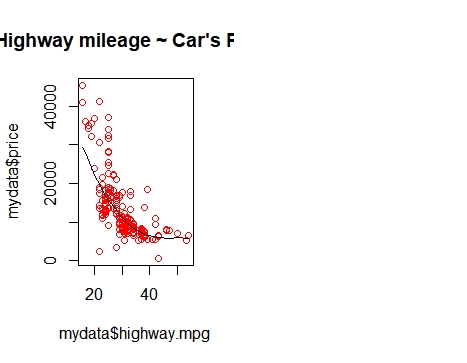
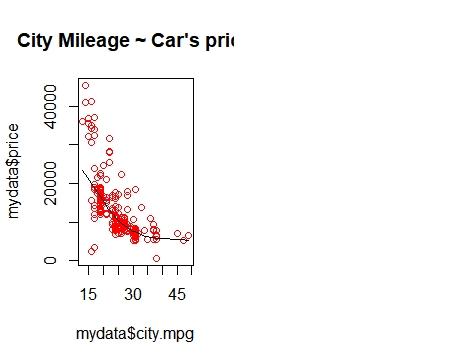
summary(mylogit)

**Description:**

Linear regression analysis has performed by above code “glm” and regression analysis results can be seen by “Summary” code.

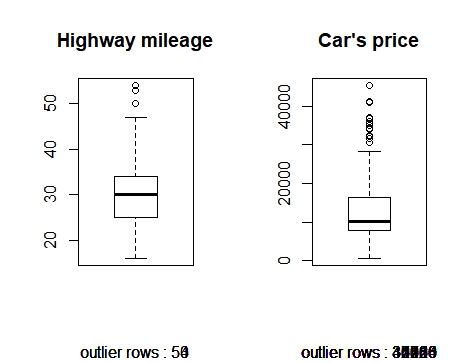
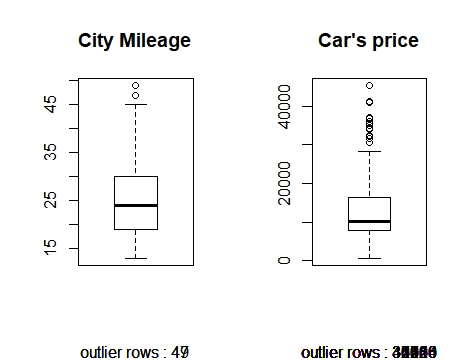
From summary you can see which variable (column) is strongly correlated with dependent variable which is “Price”.

# Graph



**Description:**

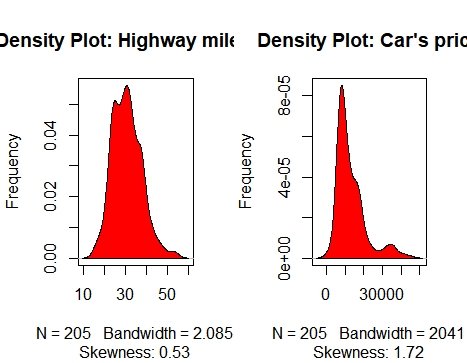
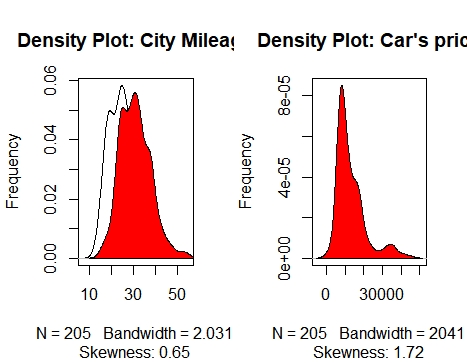
The above graphs are showing relationship of car’s mileage with price of the car. It shows that the cars which give low mileage either on highway or have fall in high price category.



**Description:**

Boxplot shows the outlier in your data. City mileage and highway mileage have outliers over 45 mileages.

Car’s price graph shows that there are many outliers on price variable over 4000.



**Description:**

Density plot shows if response variable is to normality.

