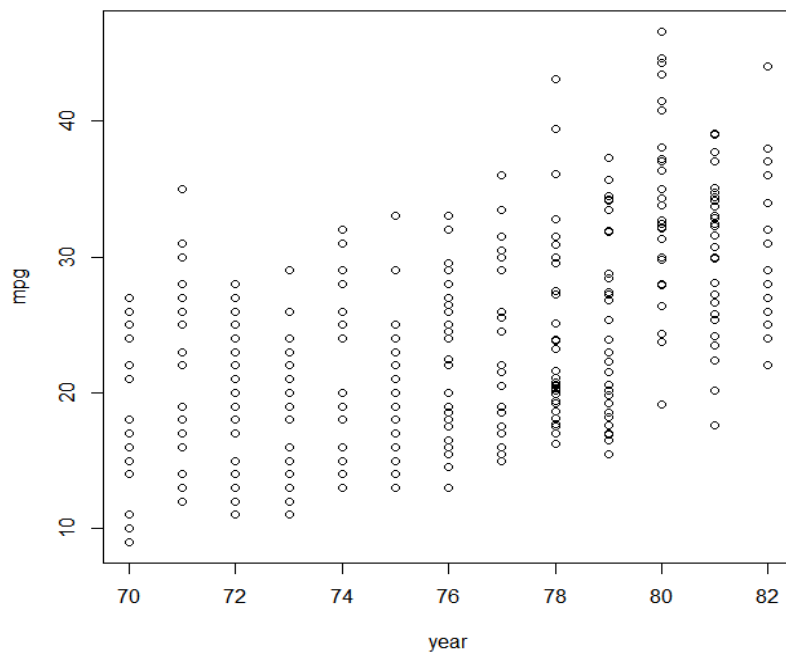# HOMEWORK 1

Question 1:

1.

The figure below shows scatterplot b/w mpg vs year. It shows that miles per gallon increased with the year. Quality of machine went increasing per year gradually.
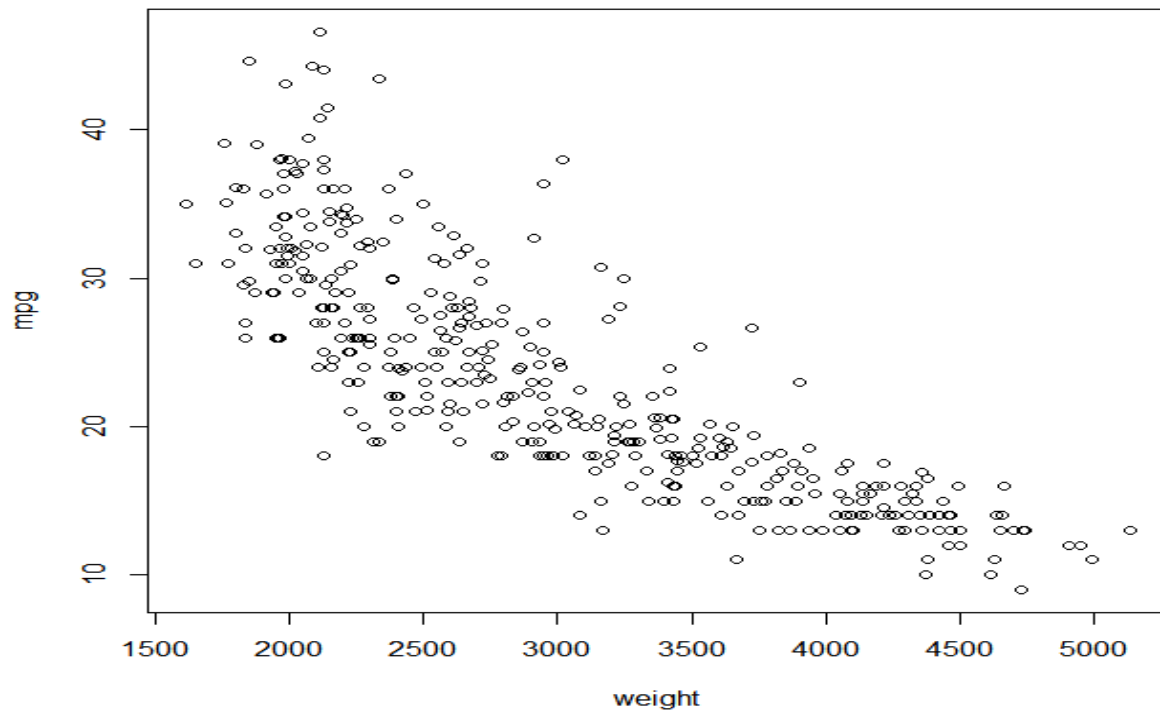
2. The figure below shows scatterplot b/w mpg vs acceleration. It shows that miles per gallon increases with the exponentially with the time to accelerate from 0 to 60. Still for most of the values of mpg, acceleration variable remained in b/w 10 to 20.
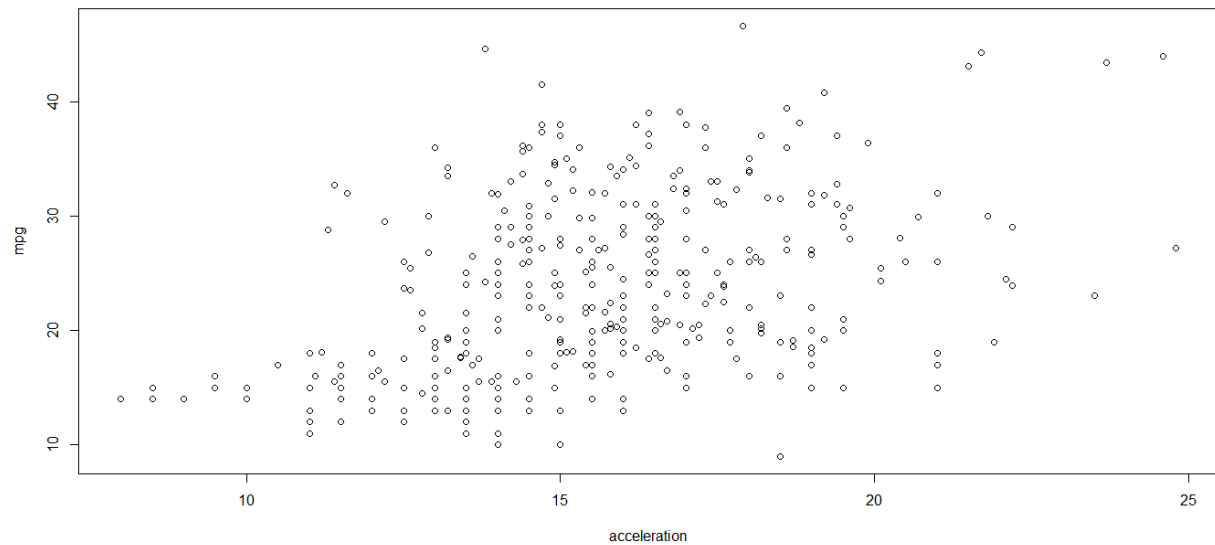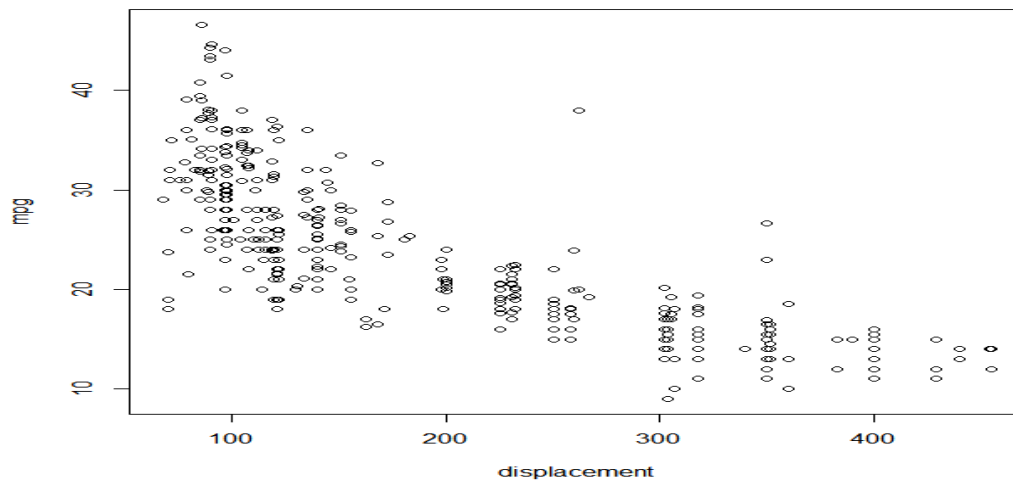
3. The figure below shows scatterplot b/w mpg vs horsepower. It shows that miles per gallon decreases exponentially with the horsepower.
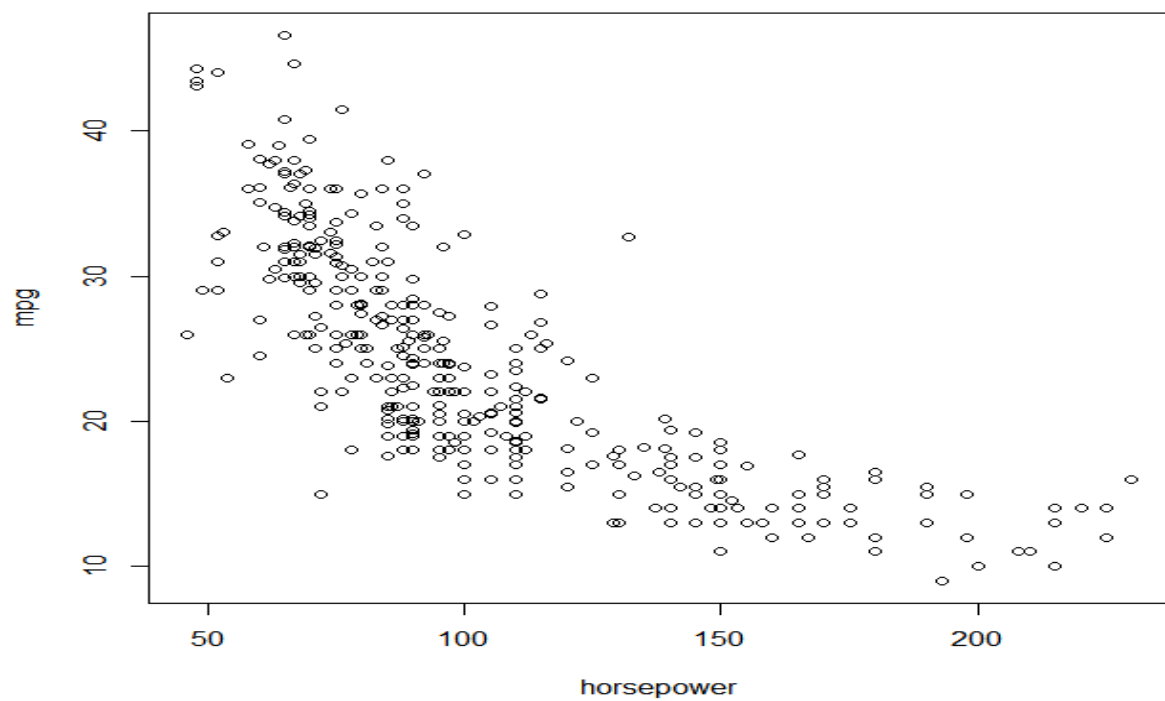

3. The figure below shows scatterplot b/w mpg vs weight. It shows that miles per gallon decreases exponentially with the weight.

4. The figure below shows scatterplot b/w mpg vs displacement. It shows that miles per gallon decreases exponentially with the displacement


5. The figure below shows scatterplot b/w mpg vs cylinder values. It shows that miles per gallon decreases with each value of cylinders.
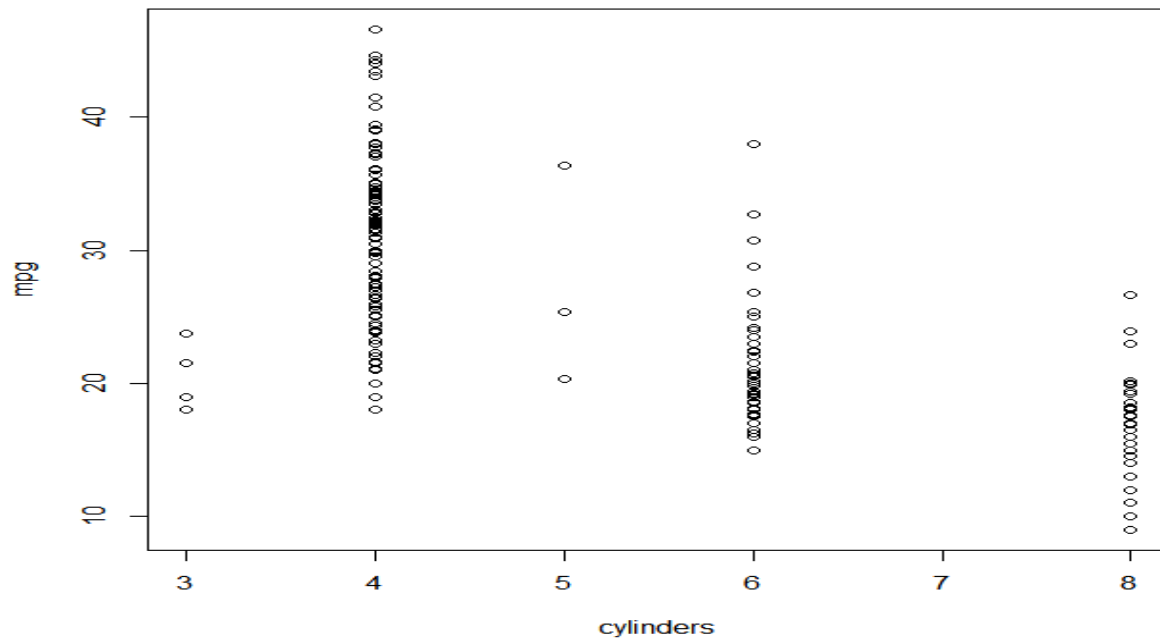
6. Below fig shows box plot for mpg. It shows median, min,max values along with first and third quartile. The median value for mpg lies around 23. All the values lies between 9 and 46.
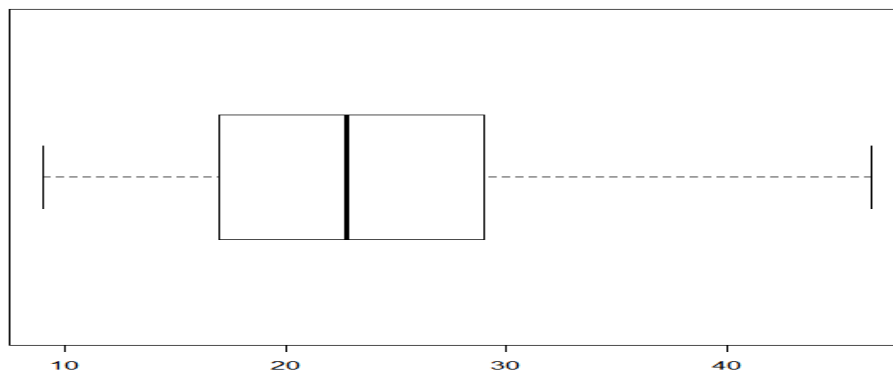
7. figure below shows plot between every variable with all another variables. From the graph we can compare and infer relations between the variables.
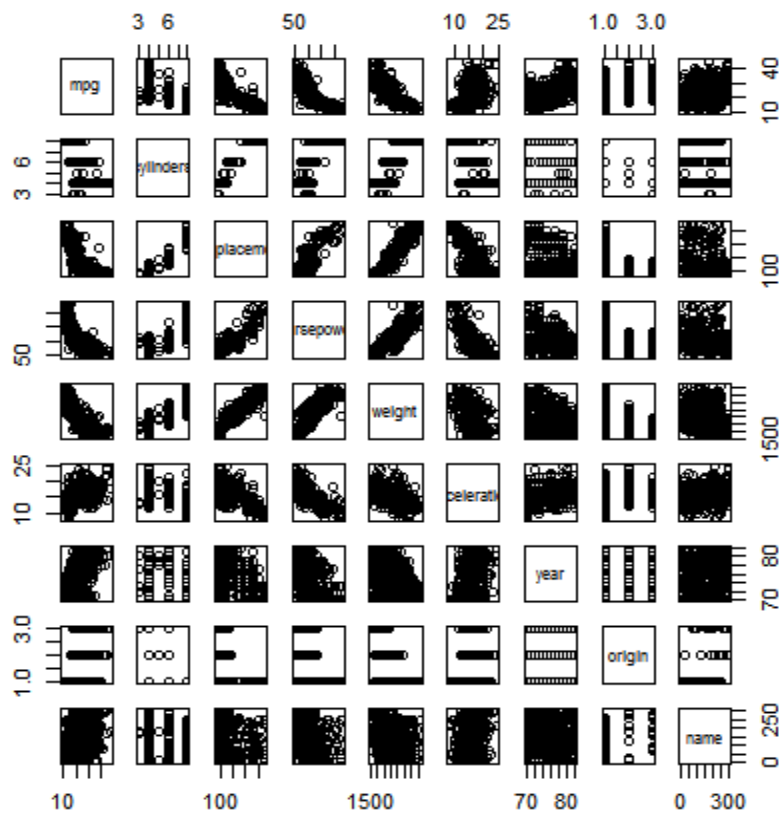
Question 2:

1) Function lm() is used to perform linear regression. We are predicting value of mpg by every variable. If we apply linear regression on mpg corresponding to combination of each function value of RSS is 0.9152709 which shows that they are unrelated.
For acceleration, value of RSS is least. We can infer that acceleration is more related to mpg variable than all other variables.
2) Coefficient variable gives coefficient of variance of the variable against other. This value is 0.3353279 for year against mpg. Which shoes slight variance from mean value.
3)*symbol is used in between multiple variables to get linear regression gives LR with these variables as well as their interaction. This value lies between individual RSS values of these two variables. In the given code we have taken *of weight and acceleration variables. Their RSS value is almost 0.40, whereas RSS values of acceleration and weight are 0.17 and 0.69 respectively.
: symbol gives only the interaction and not the individual values. This value is less than the one we get by using * symbol.

Question 4:

Diagram below shows pairwise scatterplot for every variable against all from boston dataset.
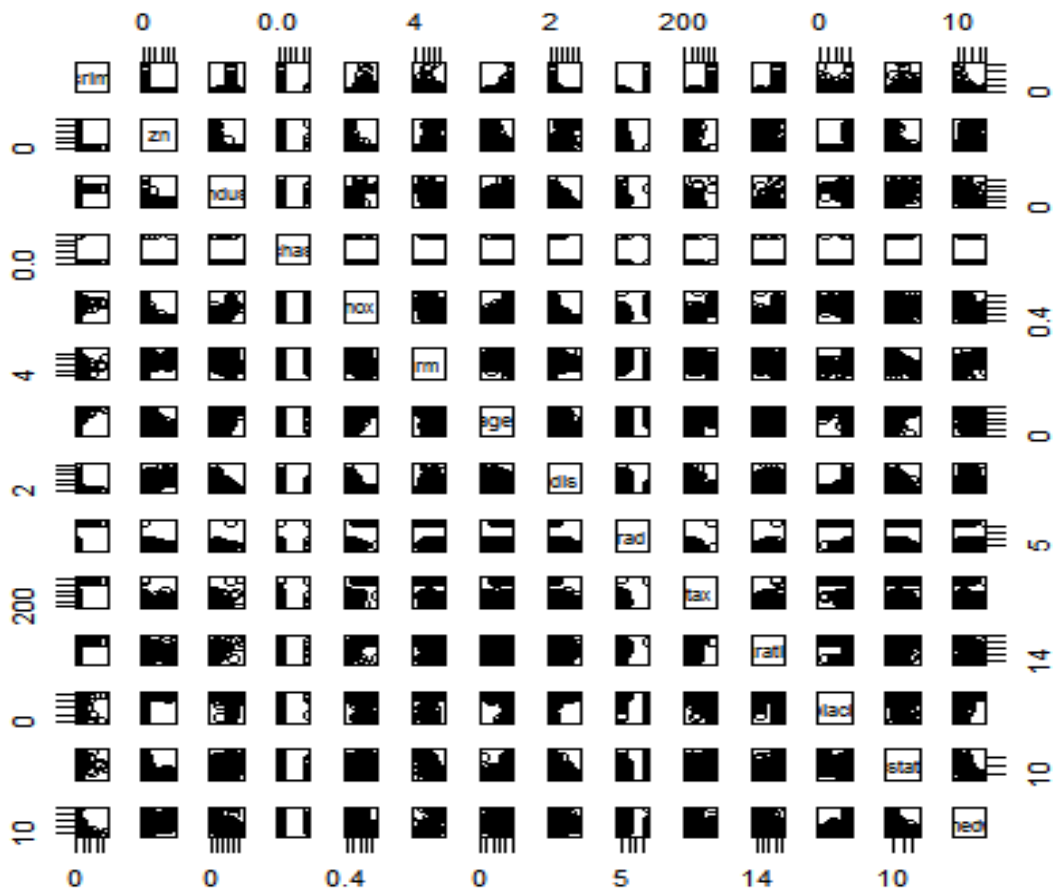
Crime per capita is related to percentage of black in the area as std error value for regression with all the variables individually is least for that which is 0.003.

No.of the suburbs average more than seven rooms per dwelling:64

No.of the suburbs average more than seven rooms per dwelling:13