**2) Tasks:**

* **Perform multiple regression with response variable mpg (miles per gallon)**

a) Which predictors appear to have a significant relationship to the response.

b) What does the coefficient variable for “year” suggest?

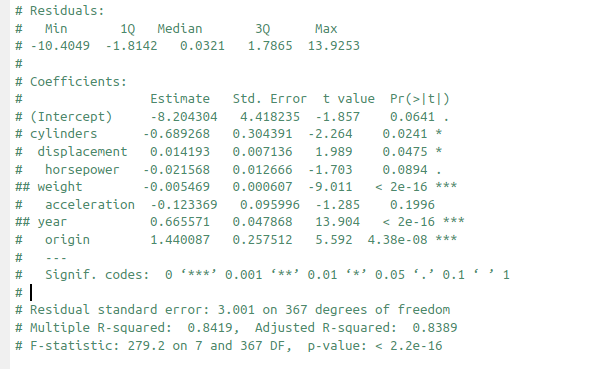
c) Use the \* and : symbols to fit models with interactions. Are there any interactions that are significant?

**ANSWERS:**

**A)**  The cleaned data is saved in the .Rdata format which was loaded for this task of performing Multiple Linear Regression.

After performing the multiple linear regression on the cleanData we obtained from the Task 1, we can see the relationship among different predictors and their significance in fitting the curve for the model.

linear\_model1 <- lm(mpg ~ ., data =df);

summary <- summary(linear\_model1);

We have already excluded the Name predictor from the data as it is irrelevant for framing the model.

From the summary of the Linear Model we can conclude following points:

* The Multiple R-squared = 0.8419 which implies 84% of the variance has been explained by this model.
* The RSE is 3.001.
* The Pr(>|t|) value is maximum for the Acceleration variable which means it is the least significant predictor and has very less or zero impact on the model.
* Below we have found the least and most significant predictors:

coef <- summary$coefficients

########## MAXIMUM SIGNIFICANCE

minPr <- which.min(coef[,"Pr(>|t|)"])

### SO WEIGHT AND YEAR ARE HAVING MAXIMUM SIGNIFICANT RELATIONHIP TO MPG

########## MINIMUM SIGNIFICANCE

maxPr <- which.max(coef[,"Pr(>|t|)"])

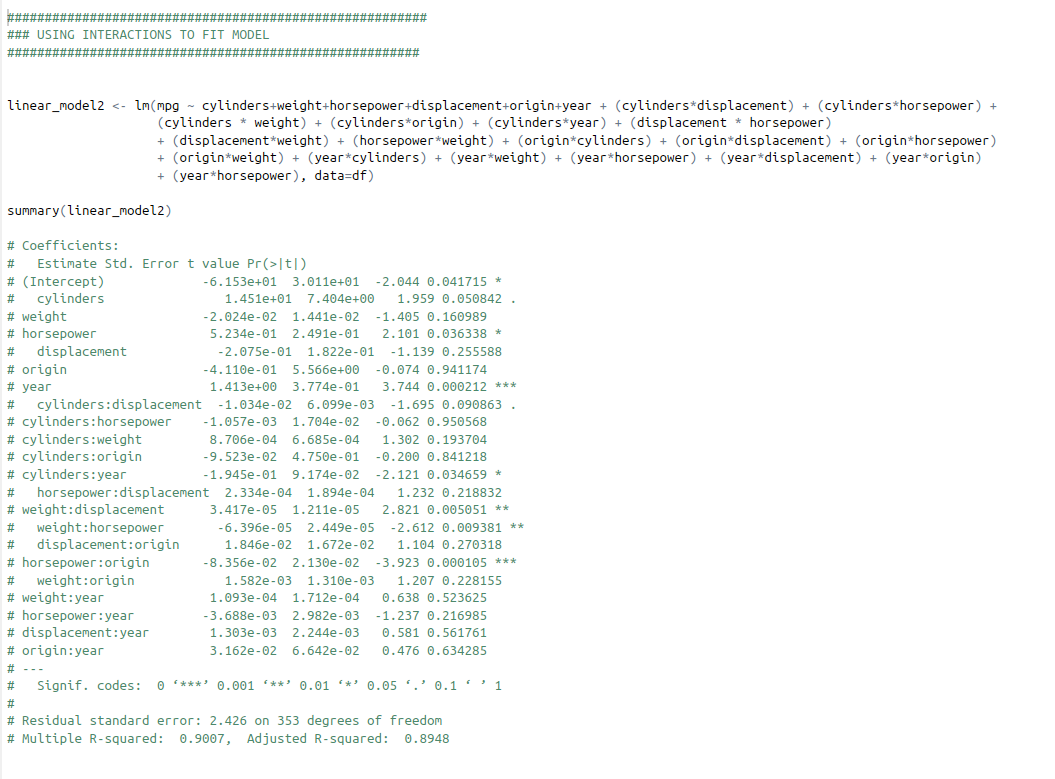
###So, Acceleration is the least significant

* So, Weight and Year has significant relationship with the response MPG and, Acceleration has the least one.

**B)**  The Coefficients for Year, ‘Estimate’ shows for every increase in an Year the MPG increases by 0.66. The Pr(>|t|) value of Year variable is least which implies it is most significant variable in constructing the model for prediction.

In case, we remove the Year variable the RSE **increases** to 3.70 which again proves it is important factor.

**C) Interactions:**



When I tried to fit model by adding the interactions for all the predictors I got the above output for the model summary. From this summary we can conclude below points:

* + - As we can see after adding interactions the RSE has been reduced and Multiple R-squared has increased which is a good sign.
    - Among all the interactions, the **Horsepower:Origin** is the most significant pair after which comes **Weight:Horsepower** and **Weight:Displacement**.
    - On the other hand, **Cylinders:Horsepower** and **Cylinders:Origin** are the least significant interactions and thus such interactions should not be considered.