Multiple Linear Regression

1. Cars Data

Y is Continuous =MPG

X1=HP

X2=VOL

X3=SP

X4=WT

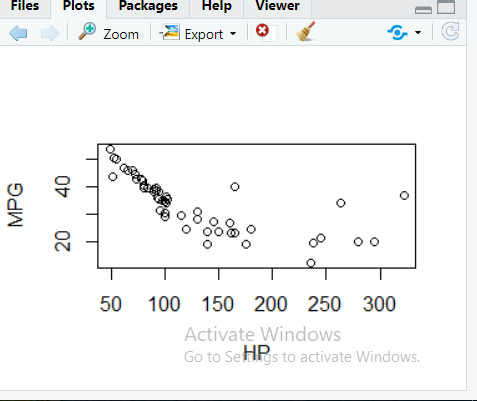
Summary of Cars

I found there is huge difference for Median and Mean. we can do box plot graphical representation- because we can get the outlier through box plot analysis. Probably we can contact to SME to check with this is correct outlier or any typo error or data entry error.

When we build a prediction model, we should give a note -prediction model work well when it is within the range. (Y=B0+B1X1…. +Xn).

We need to check the correlation between output (MPG) and Inputs (HP, VOL, SP,WT). plot(x, y).

Scatter diagram Analysis: plot (HP, MPG)

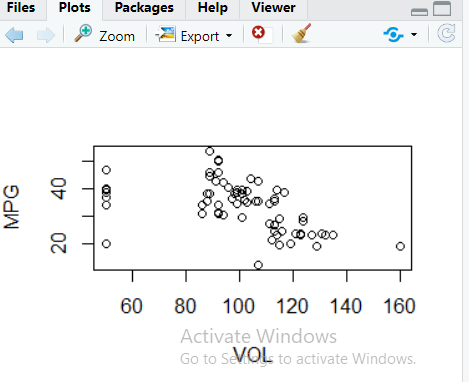


1)Direction- Negative

2)Strength-Strong & Moderate

3)Linear.

Scatter Diagram Analysis of plot(VOL,MPG)

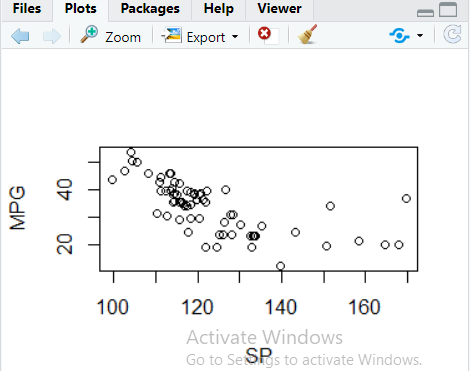


1) Direction- Negative

2) Strength- Weak

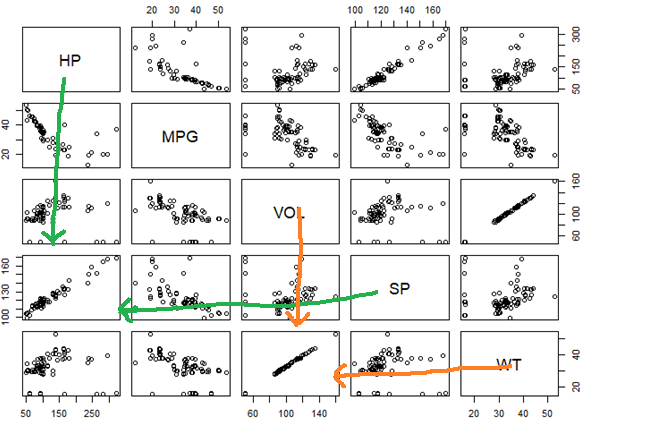
3)Linear.

Scatter Diagram Analysis of plot (SP,MPG)



If we have lot of data then we cant do analysis like this so we can use the function “pairs(Cars)”.

1. As part of Scatter diagram analysis -Collinearity between (HP, SP) and (VOL,WT).



1. Correlation coefficient matrix has given very strong ‘r’ value for correlation between (HP,SP) & (VOL,WT), further strengthening the inference that these are involved in collinearity problem.
2. Even partial correlation coefficient shows presence of Collinearity between (HP,SP) & (VOL,WT)
3. Parameters/Coefficient value pertaining to VOL & WT are insignificant.

Scenario Toyota\_Corolla

