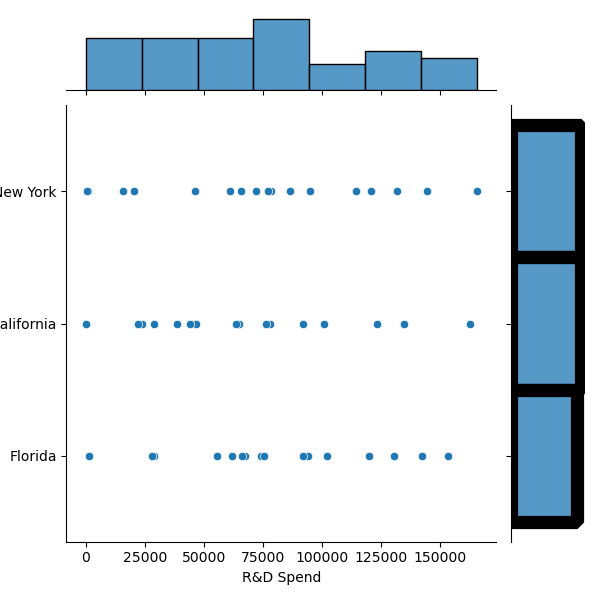
MLR INFERENCES :

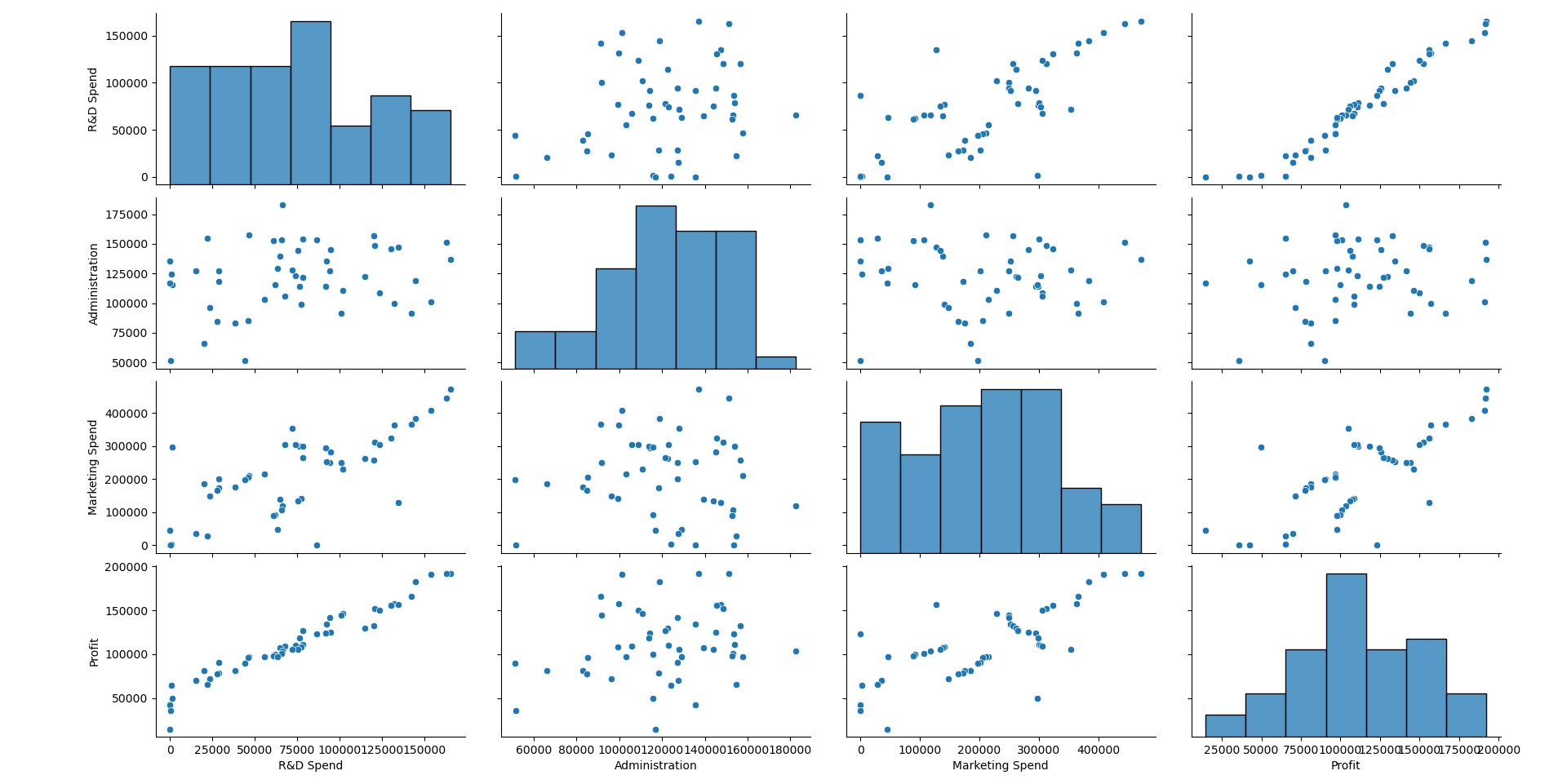
1.)50 STARTUPS DATA:

New York has the highest plot of all countries in terms of Marketing and RnD. 



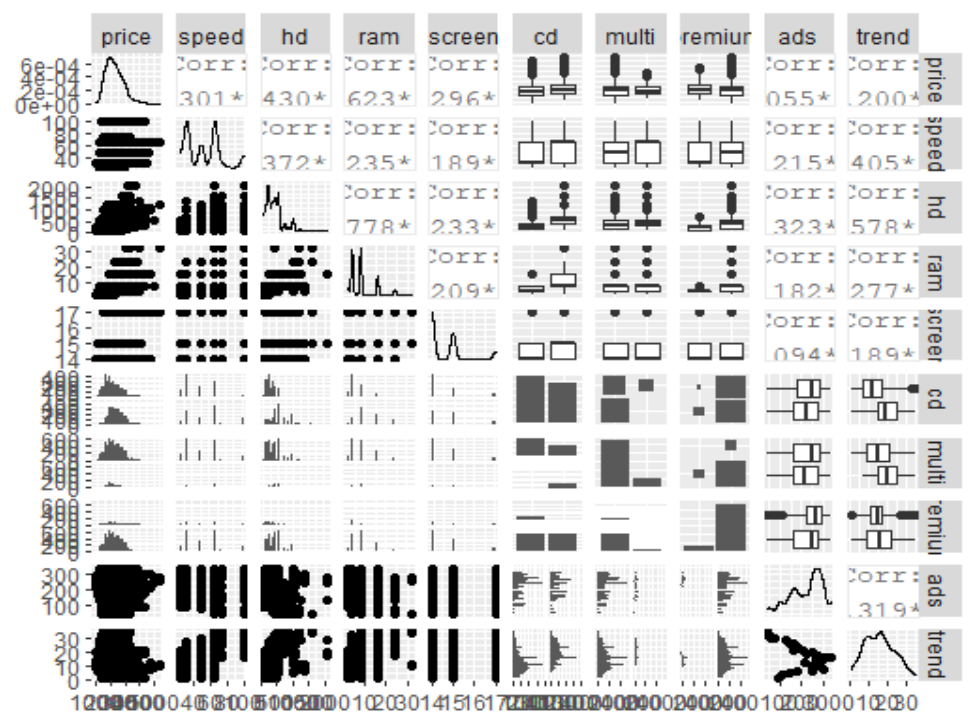
No correlation between Rnd Spend for each state.

Good Correlation between RnD Spent and Profit earned



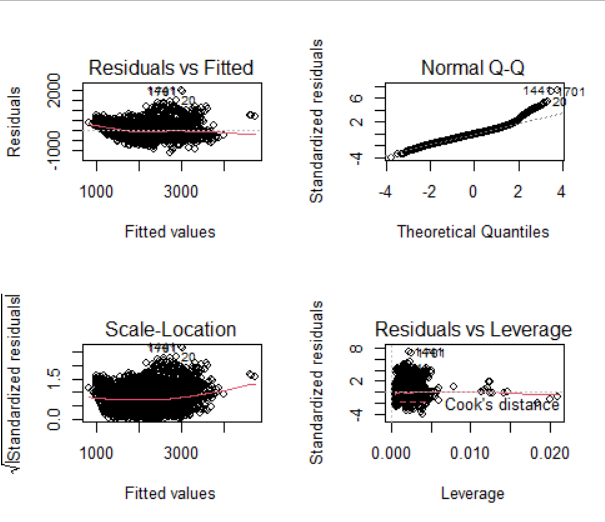
2,) Sales of Computer

No columns have a good collinearity in the dataset



Except 0.7778 corr of HD and Ram since they are storage related they should be brought together in this case.

In Plot(model) we see variance is increasing in X and y . from Resdiual v Fitted plot



The linearity assumption is not met at the end as the line and points differ Normal QQ plot

The Scale location gives the Non linearity because of the downward cuve towards the 1500 in X-axis

After removing the influential points we model our data and get

Multiple R-squared: 0.7756, Adjusted R-squared: 0.7752

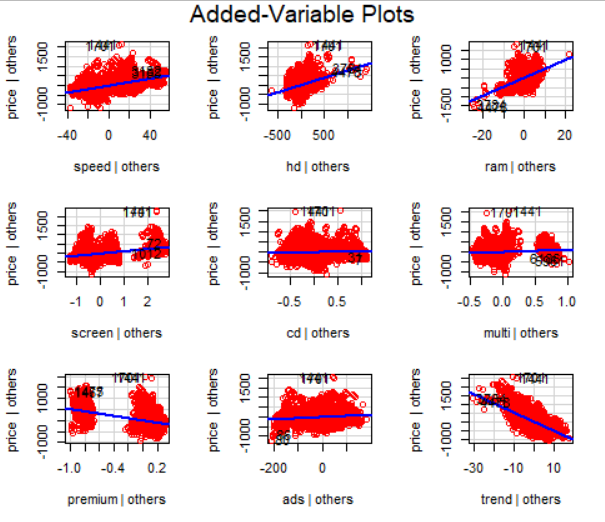
F-statistic: 2399 on 9 and 6249 DF, p-value: < 2.2e-16

To check on variance influence factors should not be more than 10 as it will show collinearity.

speed hd ram screen cd multi premium ads trend

1.265364 4.207395 2.974628 1.081644 1.859370 1.290568 1.109388 1.217218 2.022790

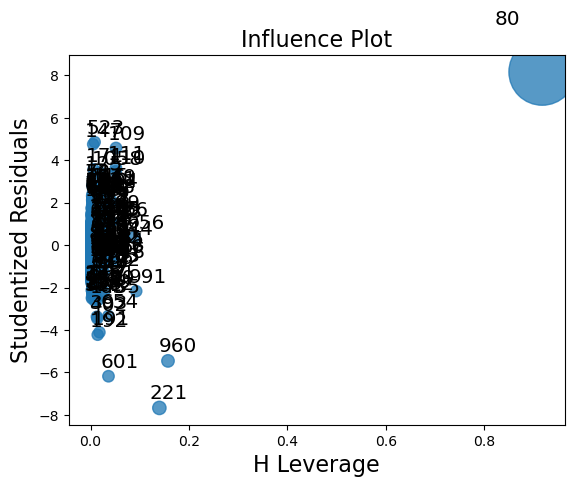
Added Variable plot



3.) Corrola Dataset

Outliers detected by boxplot in Price column.

Influence plot shows 80th row should be removed as shown as Studentized Residual



ProbabPlot for the final Model after removing the influential variables

