Prepare a prediction model for profit of 50\_startups data.

Do transformations for getting better predictions of profit and

make a table containing R^2 value for each prepared model.

Exploratory data analysis:

R&D Spend Administration Marketing Spend State Profit

Min. : 0 Min. : 51283 Min. : 0 Length:50 Min. : 14681

1st Qu.: 39936 1st Qu.:103731 1st Qu.:129300 Class: character 1st Qu.: 90139

Median: 73051 Median :122700 Median :212716 Mode :character Median :107978

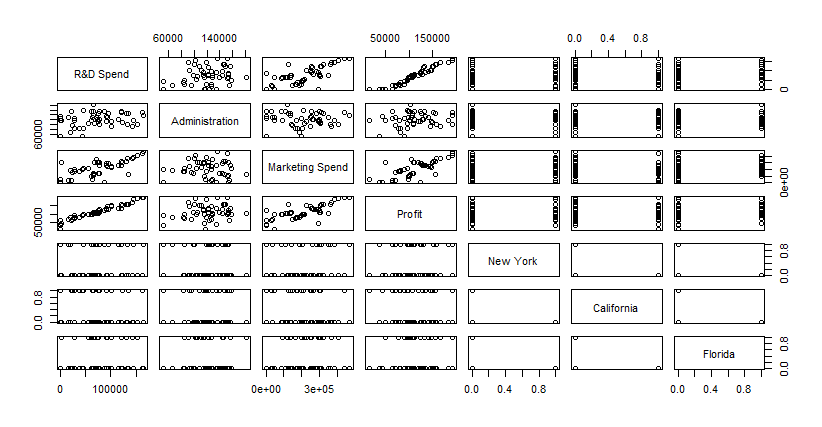
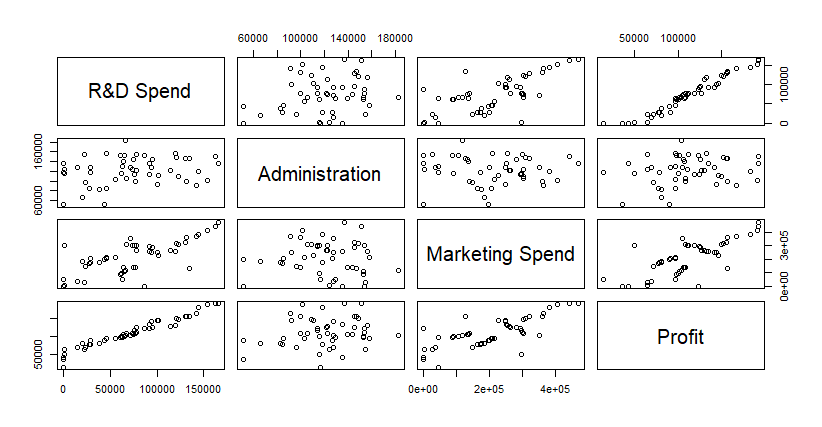
Mean : 73722 Mean :121345 Mean :211025 Mean :112013

3rd Qu.:101603 3rd Qu.:144842 3rd Qu.:299469 3rd Qu.:139766

Max. :165349 Max. :182646 Max. :471784 Max. :192262

-Create dummy Variables

Plots:

Plotting includes dummy variables :plot without dummy variables:

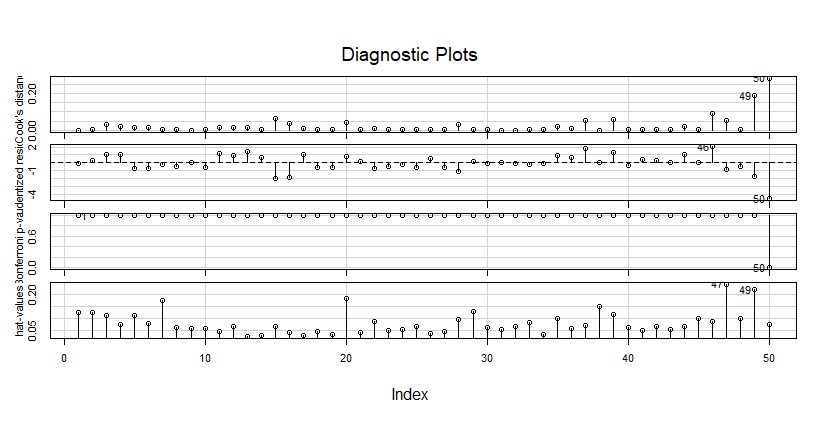
`R&D Spend` 8.057e-01 4.515e-02 17.846 < 2e-16 \*\*\*

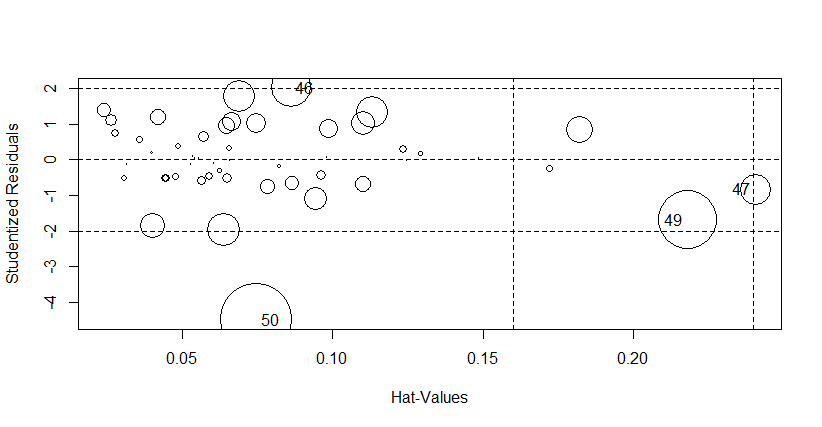
Administration -2.682e-02 5.103e-02 -0.526 0.602

`Marketing Spend` 2.723e-02 1.645e-02 1.655 0.105

P value is greater than 0.05 so now checking influencing records.

Influence index plot:





StudRes Hat CookD

46 2.0220730 0.08617007 0.09032342

47 -0.8268684 0.24060165 0.05453034

49 -1.6861241 0.21801940 0.19052744

50 -4.4961657 0.07477116 0.28808229

Variation influence factor to check collinearity between the variables.

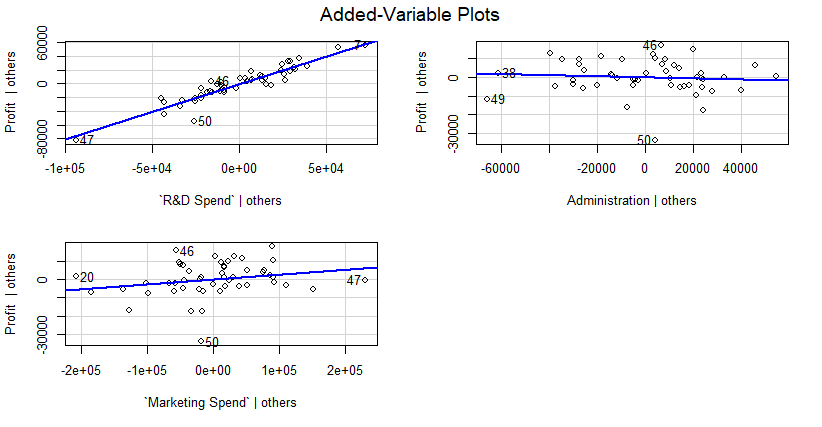
Vif (Profit\_Model)

`R&D Spend` Administration `Marketing Spend`

2.468903 1.175091 2.326773

#vif>10=>there exists collinearity among all variables

added variable plots to check correlation b/w o/p variables:



Creating a final model after removing the data which consists of higher p value when compared to our significance level.

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 4.903e+04 2.538e+03 19.32 <2e-16 \*\*\*

`R&D Spend` 8.543e-01 2.931e-02 29.15 <2e-16 \*\*\*

=>data that can predict profits with the highest impact is composed of only one category:

R&D spending.

Multiple R-squared: 0.9465

Adjusted R-squared: 0.9454

p-value: < 2.2e-16<0.05=>overall the model is good.

Plots of final models:

