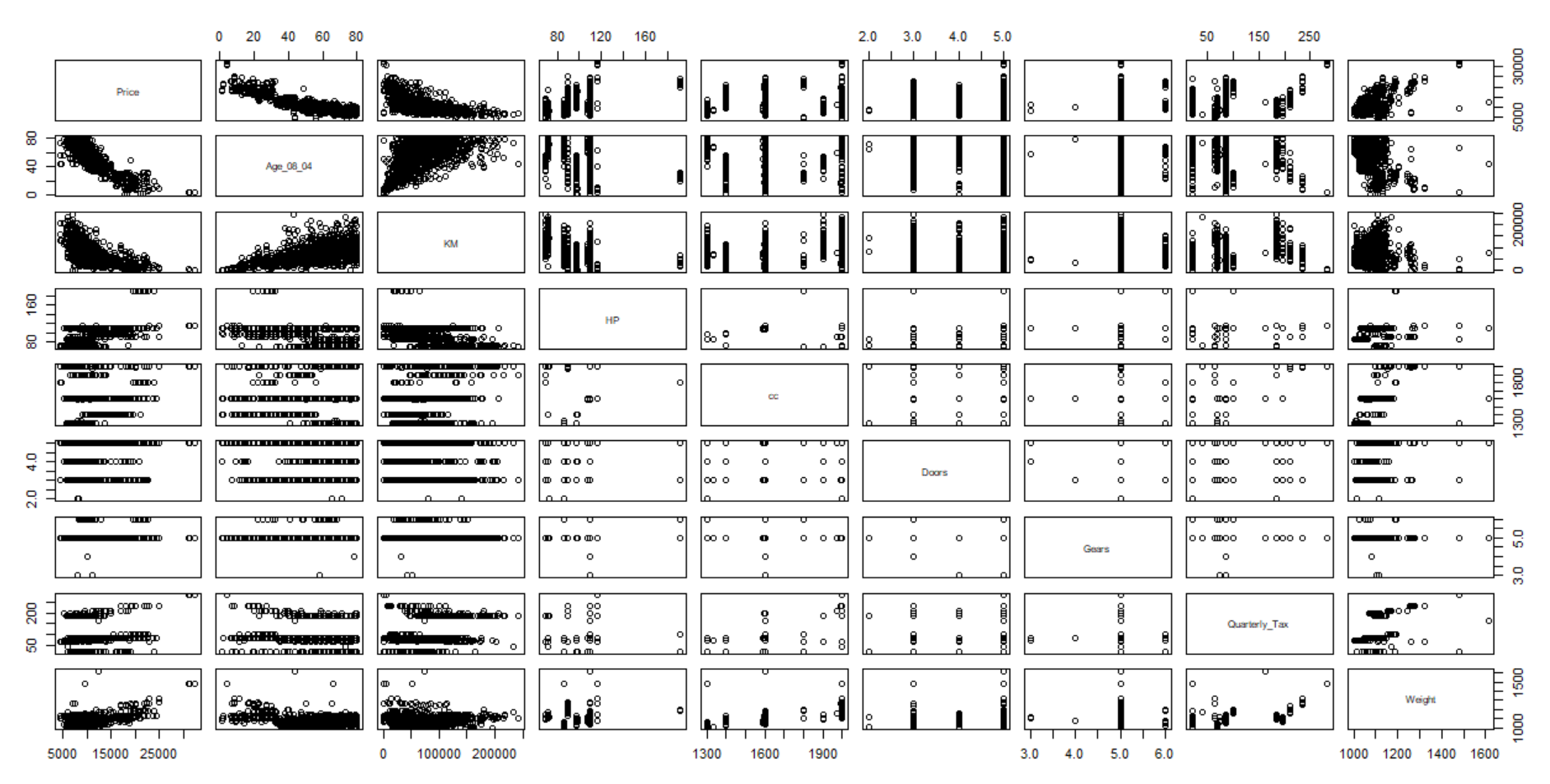
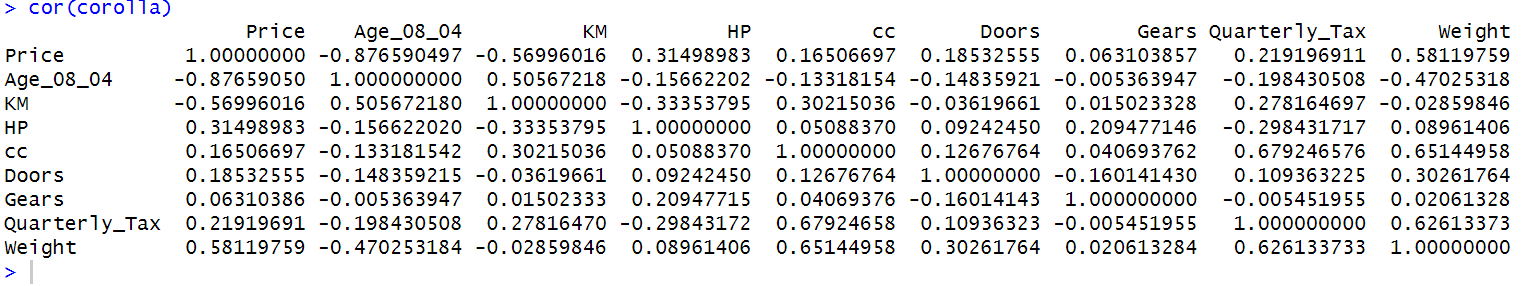
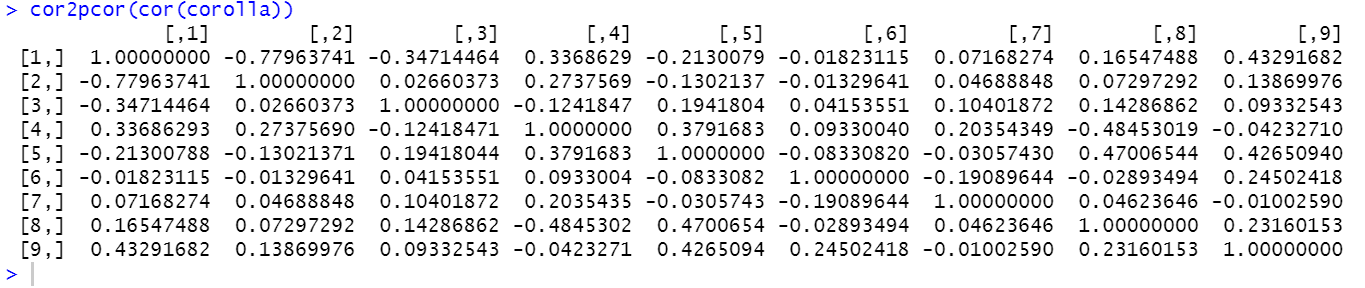
**Multi Linear Regression Solution**

**Corolla Price Prediction**

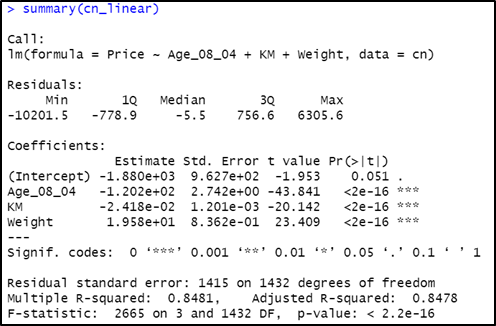
* On conducting EDA, found that tha function gave some outlier, which to me seemed to be not an outlier and for some will have to consult the SME.
* One of which in the “cc” column at the 81st observation was a typo, which I corrected.
* After plotting scatter plot and correlations of the variables, achieved following results:

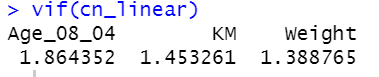




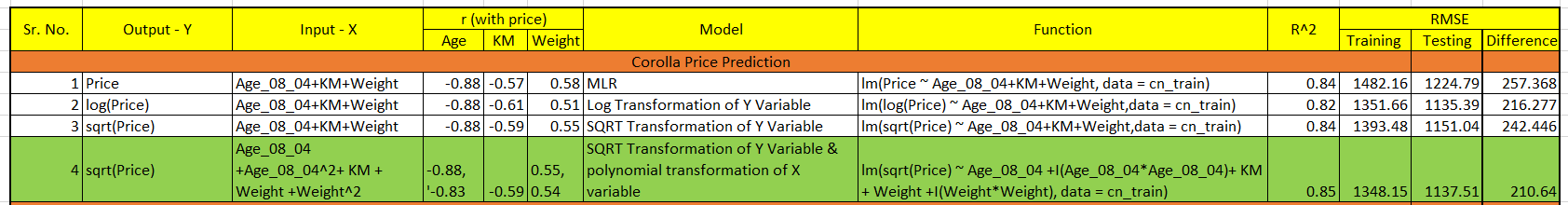


* From observing above plots and values we can infer that variables: cc, Doors, Gears, Quarterly\_Tax & HP have a very less correlation with Price. So, I have decided to drop those variables.
* After forming a linear model, I understood that all the estimate co-efficient have a significant p-value, the model has a good R squared vale. Moreover, the vif values are also less than 10, so there’s no co-linearity. We can proceed with the following variables now.





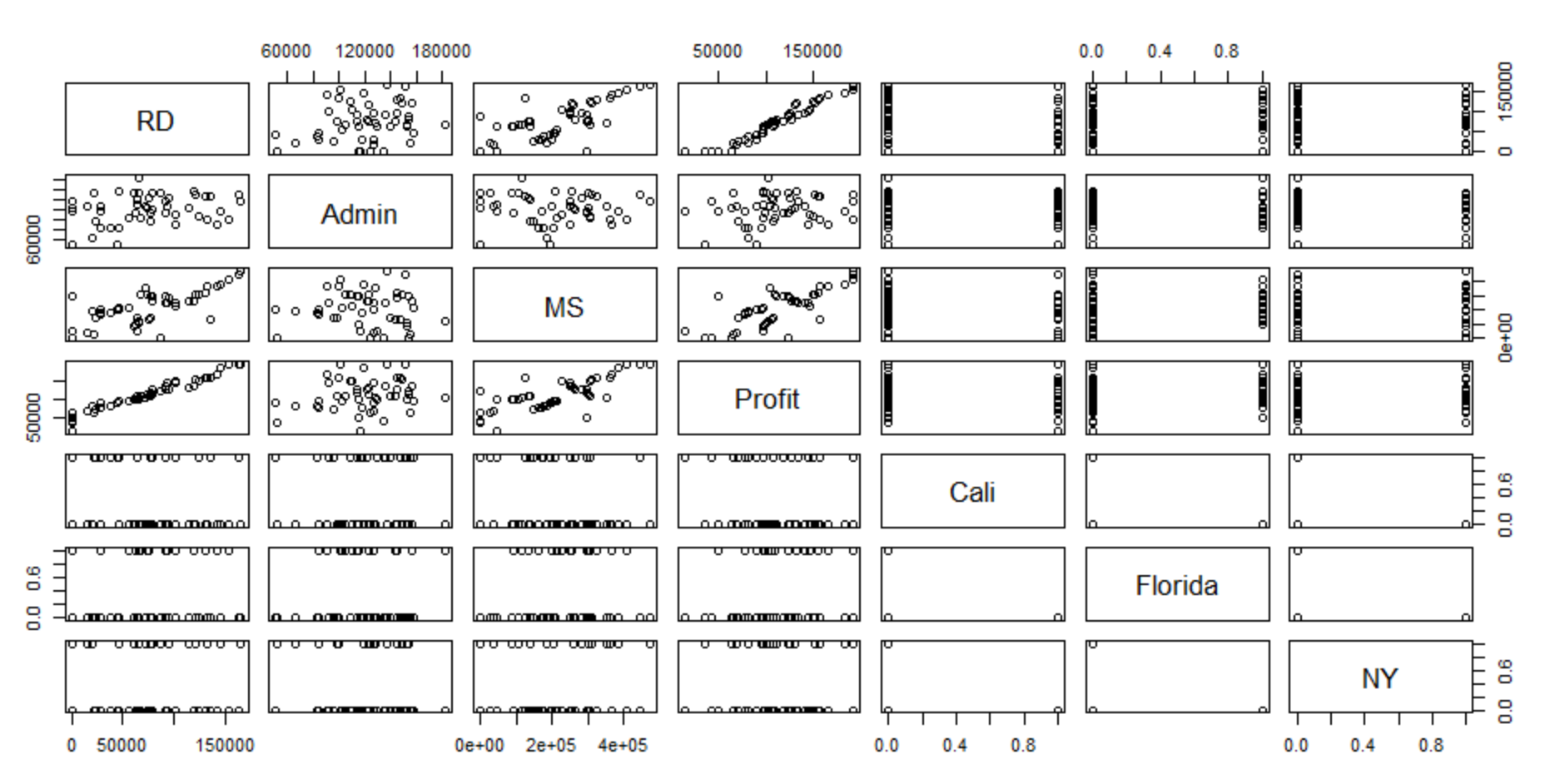
* After making various model by using different transformations I have achieved following result:

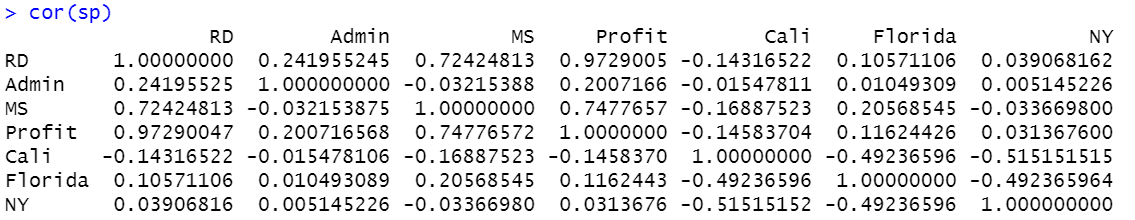


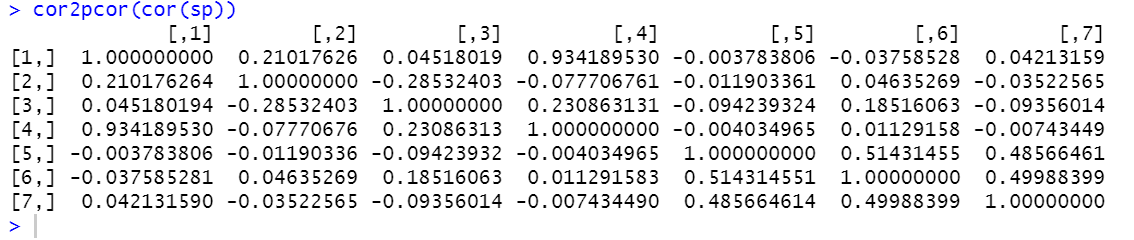
* Although, the transformations don’t stop here, we can keep optimizing the model by using different transformations. Out of the above 4 transformations, I have finalized on the polynomial transformation because it gives the highest R squared value, lowest training & testing RMSE & lowest difference in between RMSE’s. Also, the model follows the LINE assumptions.

**Startup Profit Prediction**

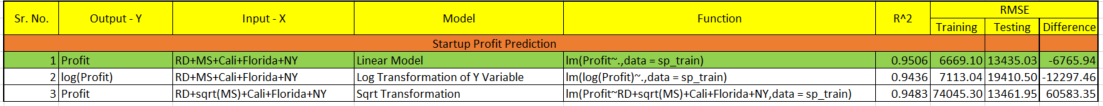
* After plotting scatter plot and correlations of the variables, achieved following results:







* From observing above plots and values we can infer that variable Admin cost has a very less correlation with Profit. So, I have decided to drop the variable Admin.
* After making various model by using different transformations I have achieved following result:



* Although, the transformations don’t stop here, we can keep optimizing the model by using different transformations. Out of the above 4 transformations, I have finalized on the linear model because it gives the highest R squared value, lowest training & testing RMSE & lowest difference in between RMSE’s. Also, the model follows the LINE assumptions.