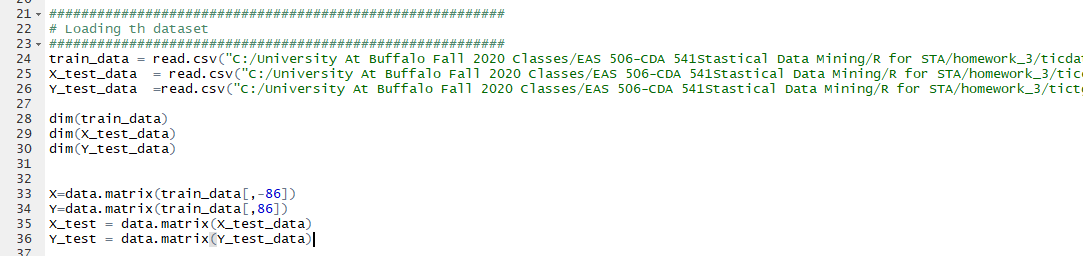
**Question 1 HW3**

* The question asks to predict who will be interested in buying a caravan insurance policy and give an explanation why?

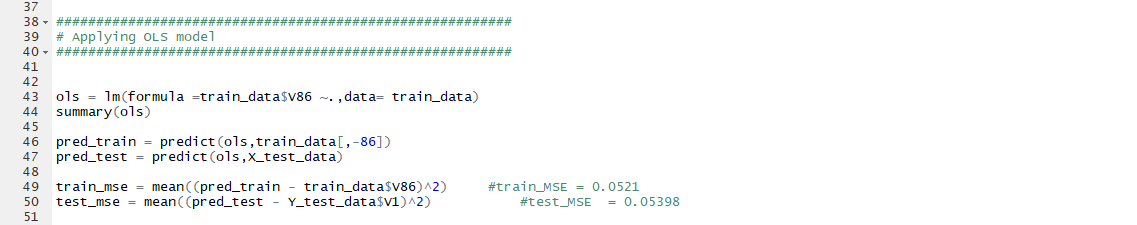
**Step 1 Importing the dataset**

The dataset is taken from uci repository and imported into the model.

****

**Step 2 Applying OLS method to the dataset**

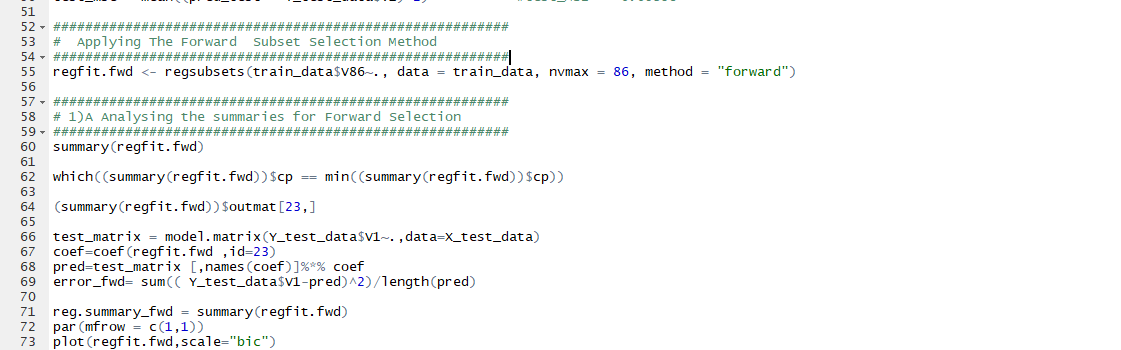
On applying linear model to the dataset againest all the variables it appears that the **test MSE = 0.05398 which is pretty less**. But it takes 85 variables to get to this error rate which is pretty much costly. Lets see the next methods for any improvement.

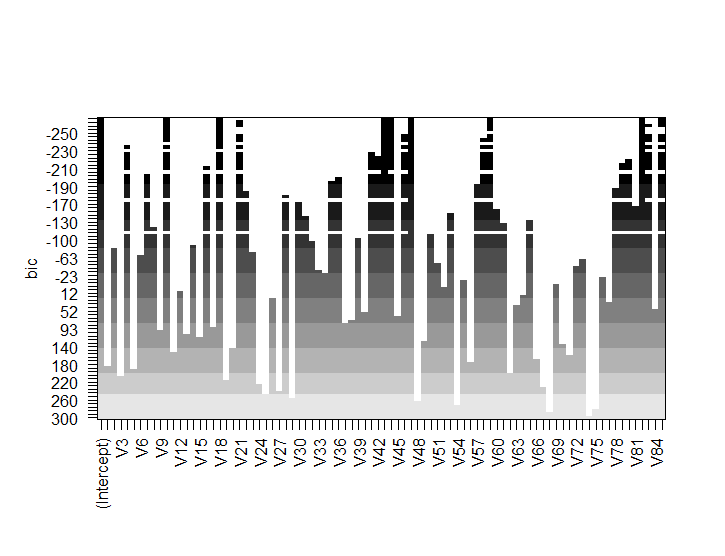


**Step 3 Applying Forward subset selction**

I used **regsubsets** function to fit and apply forward subset selction and kept **nvmax to 86 i.**e no.of variables = 86. And I used Cp score to find the best fit model to the dataset and it appears that having 23 variables will fit the model best to the dataset.

I fitted the model with the 23 variables as shown below. And **the MSE = 0.05393412 using 23 Varaibles**

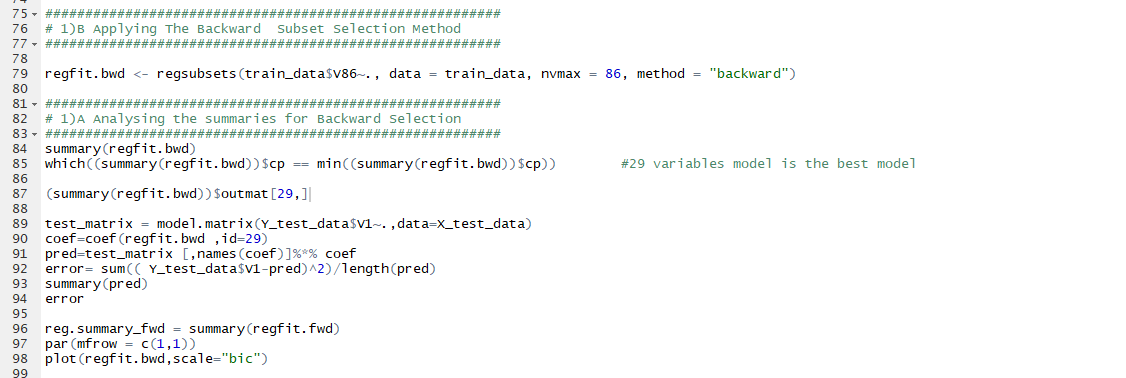


* the above picture shows the graph of BIC score for each variable

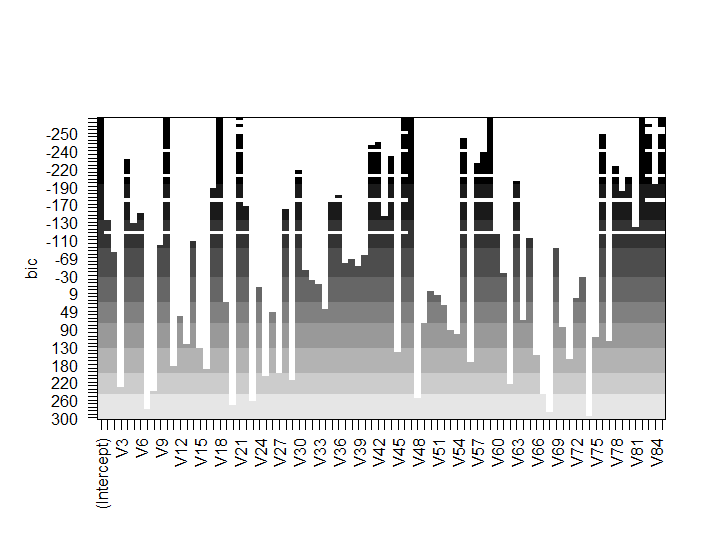
**Step 4 Applying Backward subset selction**

I used **regsubsets** function to fit and apply forward subset selction and kept **nvmax to 86 i.**e no.of variables = 86. And I used Cp score to find the best fit model to the dataset and it appears that having 29 variables will fit the model best to the dataset.

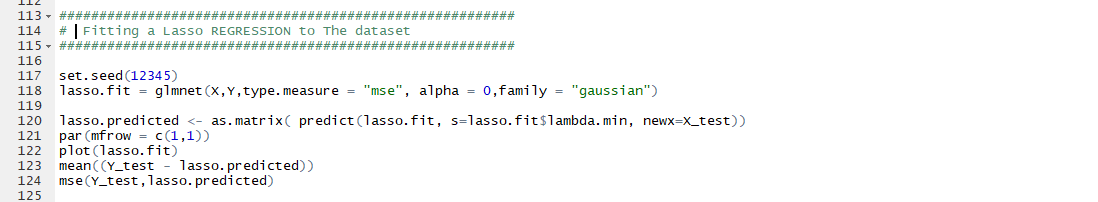
I fitted the model with the **29** variables as shown below. . And **the MSE = 0.05401232 using 29 Varaibles more compared to the pevious model.**



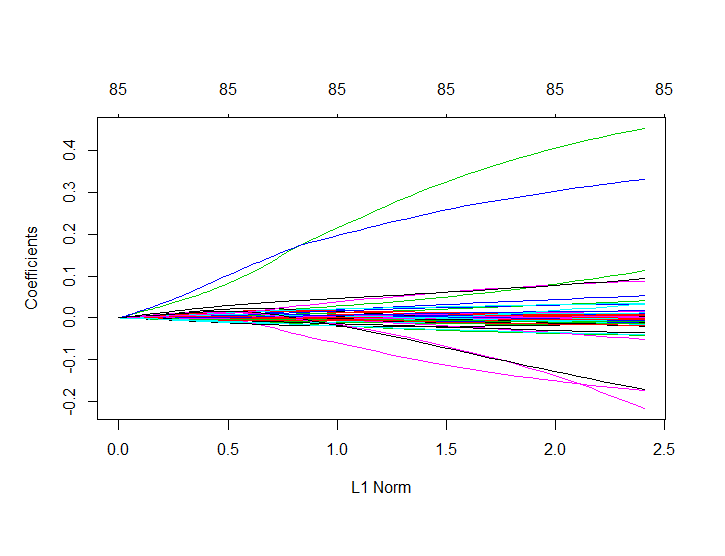
* below picture shows the graph of BIC score for each variable



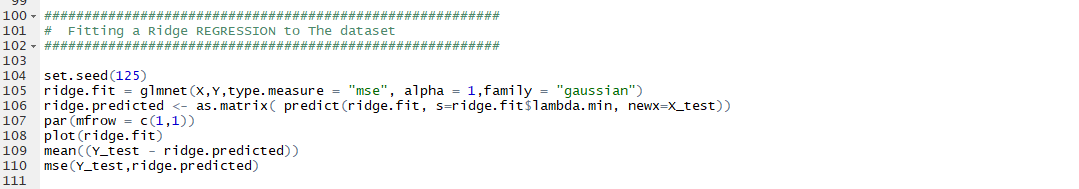
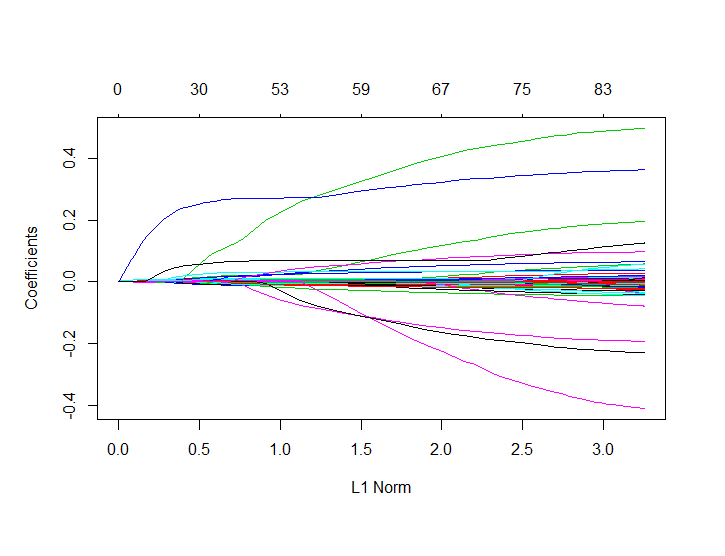
Step 5  **applying lasso regression to the model**



Below is the graph of the model.



Step 6  **applying Ridge regression to the model**



Step 7  **Analysing the results**

It appears that the **forward subset selection appears** to be the best model out of all with very **less test Mse as** low as **0.053** using **23** variables to predict the model.