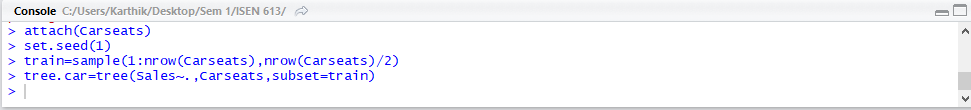
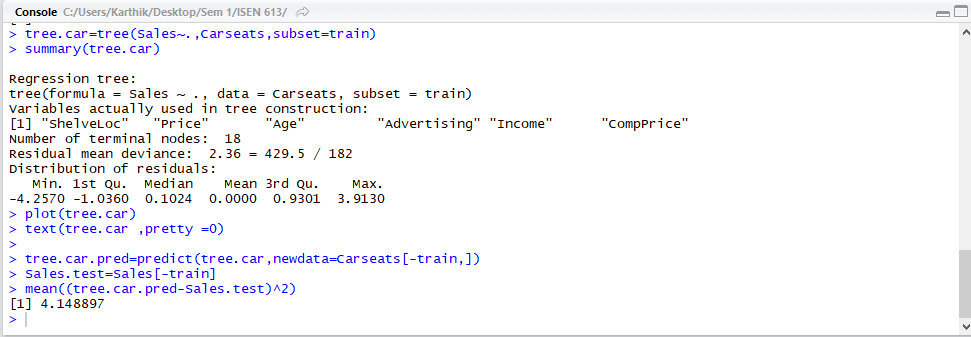
1)

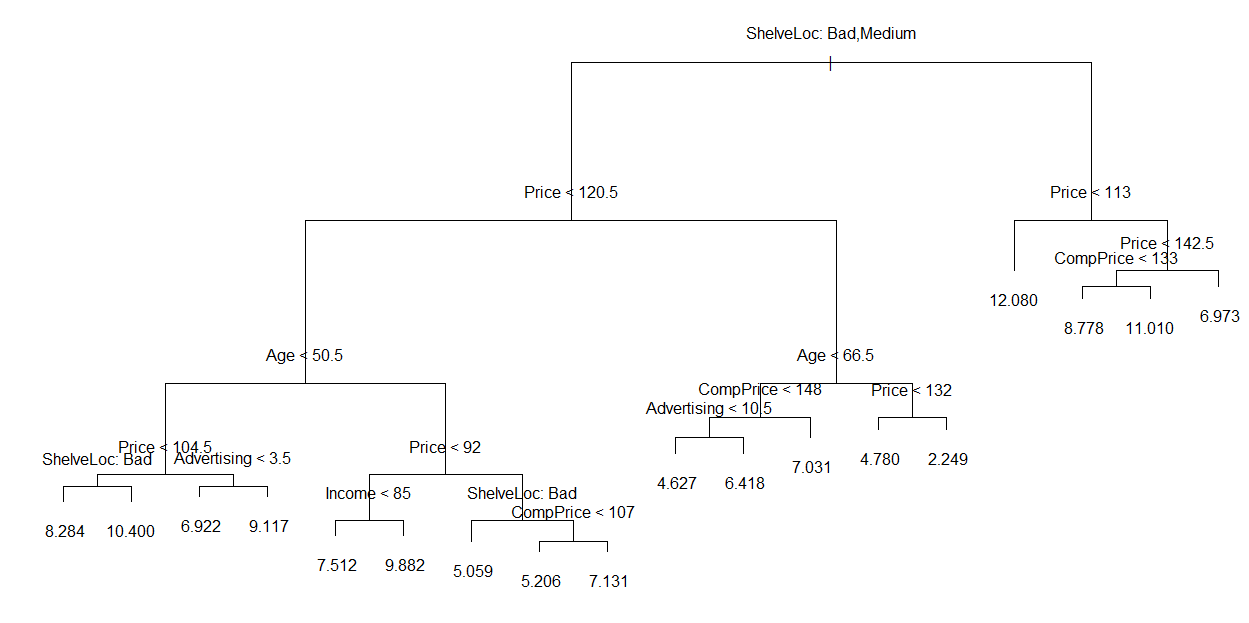
a)



b)



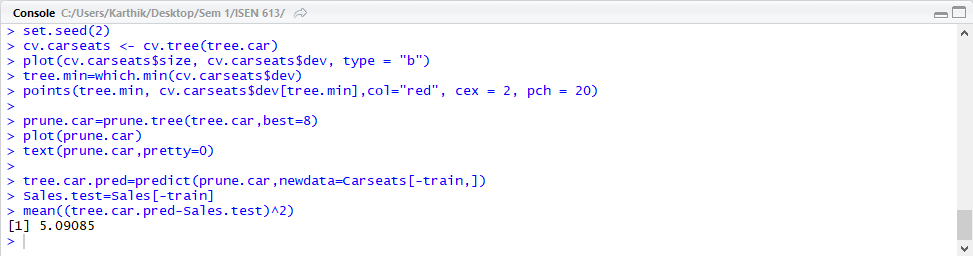
Test MSE:4.15

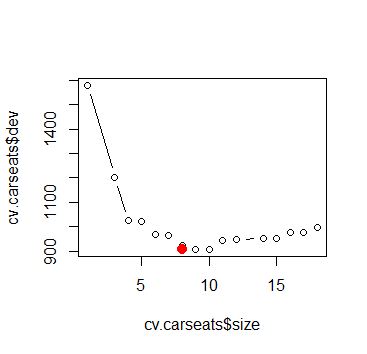


There are a total of 18 terminal nodes.

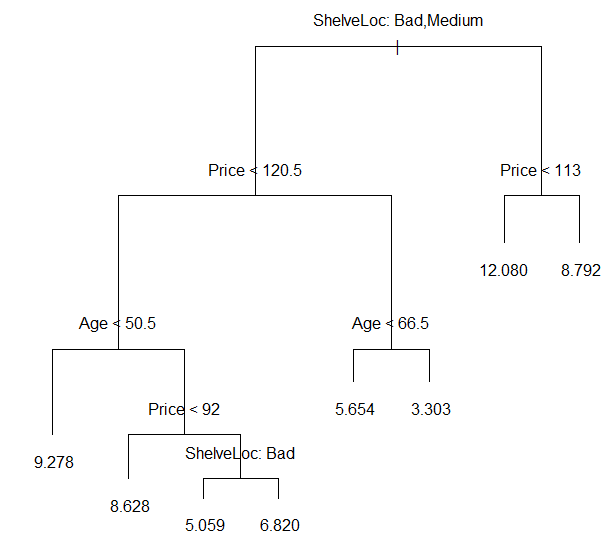
It has been found that the test MSE is 4.15. Certain variables like “Age” is used to find the prediction in certain scenarios only like of Sales if Price is less than 120. The sales is highest for the case where ShelveLoc is Bad or medium and price is lesser than 113. The tree has a large number of branches.

c)





According to cross validation, 8 variable model turns out to be the best with least test MSE.



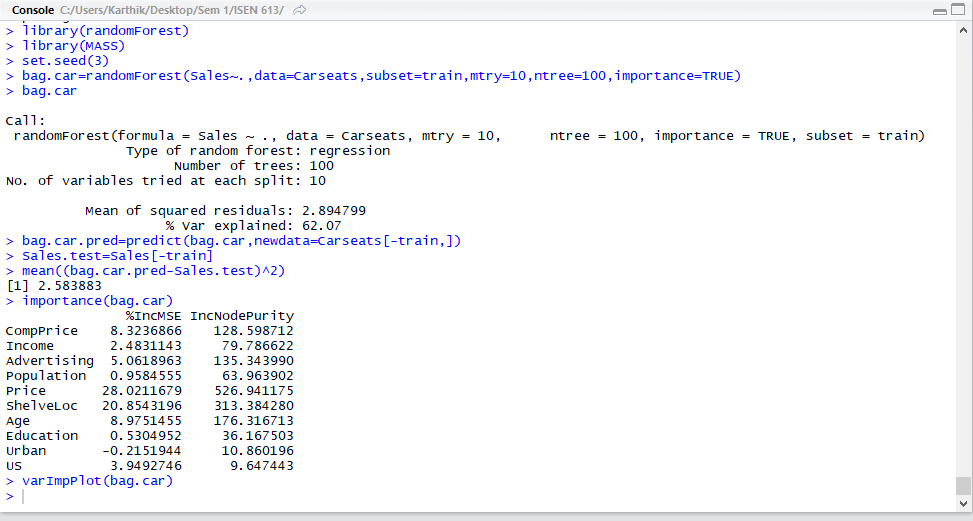
There are a total of 8 terminal nodes.

It has been found that the test MSE is 5.09.

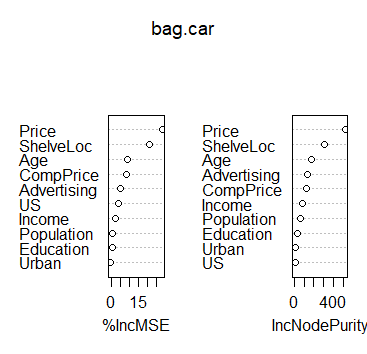
The test MSE has increased on pruning.

Certain variables like “Age” again as in the tree before pruning gives the first level of division is used to find the prediction in certain scenarios only like of Sales if Price is less than 120. The sales is highest for the case where ShelveLoc is Bad or medium and price is lesser than 113. The tree lesser number of branches compared to the tree before pruning.

d)



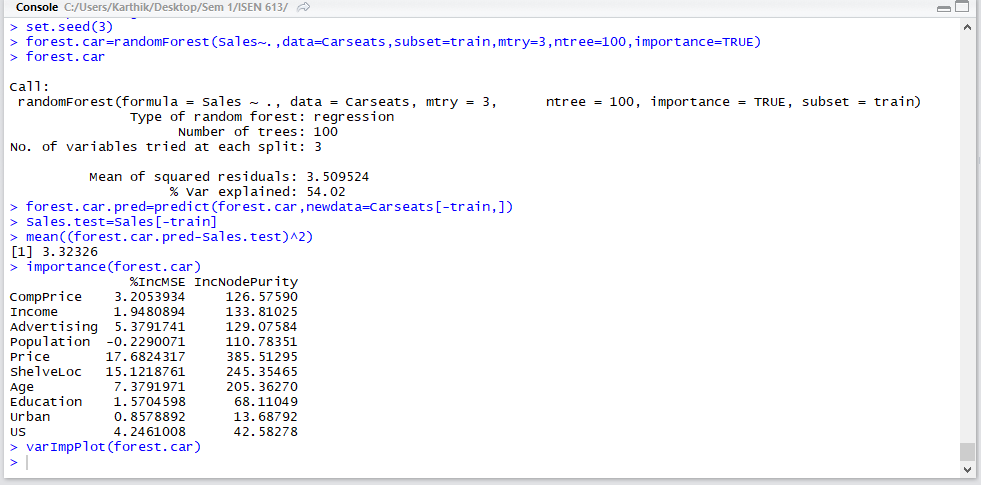
Test MSE: 2.58



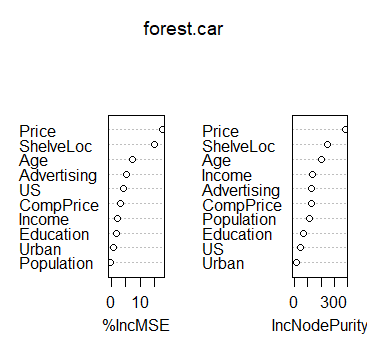
Price is the most important variable. ShelveLoc and Age seem to be the next 2 important variables.

According to the mean decrease in RSS/ impurity averaged over all the trees, the above inference is made.

e)



Test MSE: 3.32

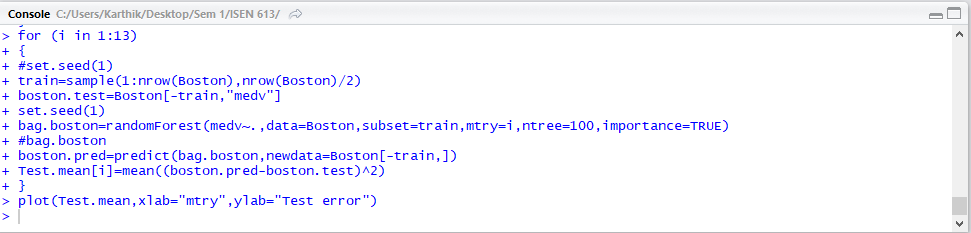


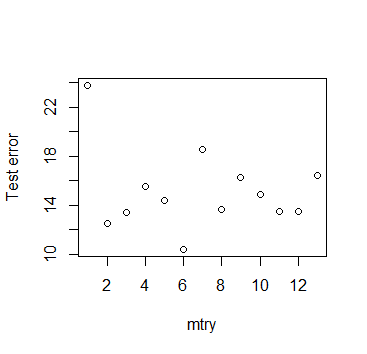
Price is the most important variable. ShelveLoc and Age seem to be the next 2 important variables.

According to the mean decrease in RSS/ impurity averaged over all the trees, the above inference is made.

2)

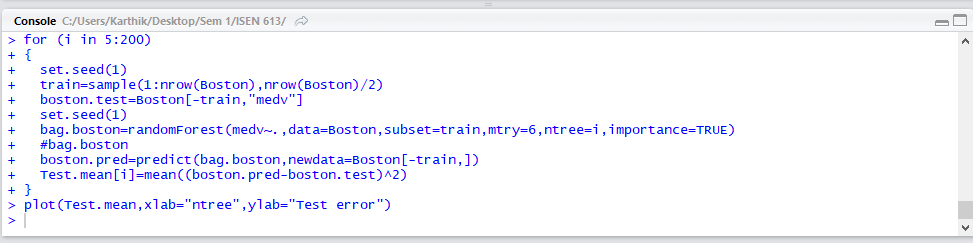
a)

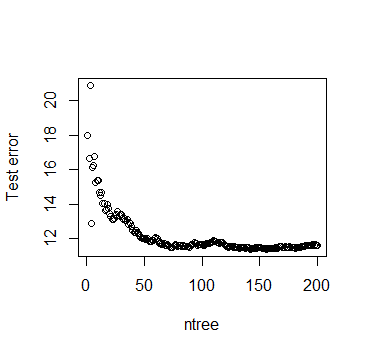




Test error is the maximum when 1 variable alone is considered. Also, when 6 variables are considered, Test error seems to be the least. Moreover there does not seem to be any constant trend between the test error and ‘mtry’.

b)





With an increase in the tree length, the test error decreases.

After a certain increase in the tree length, the test error does not decrease significantly, the increase of which only results in increased computational time.