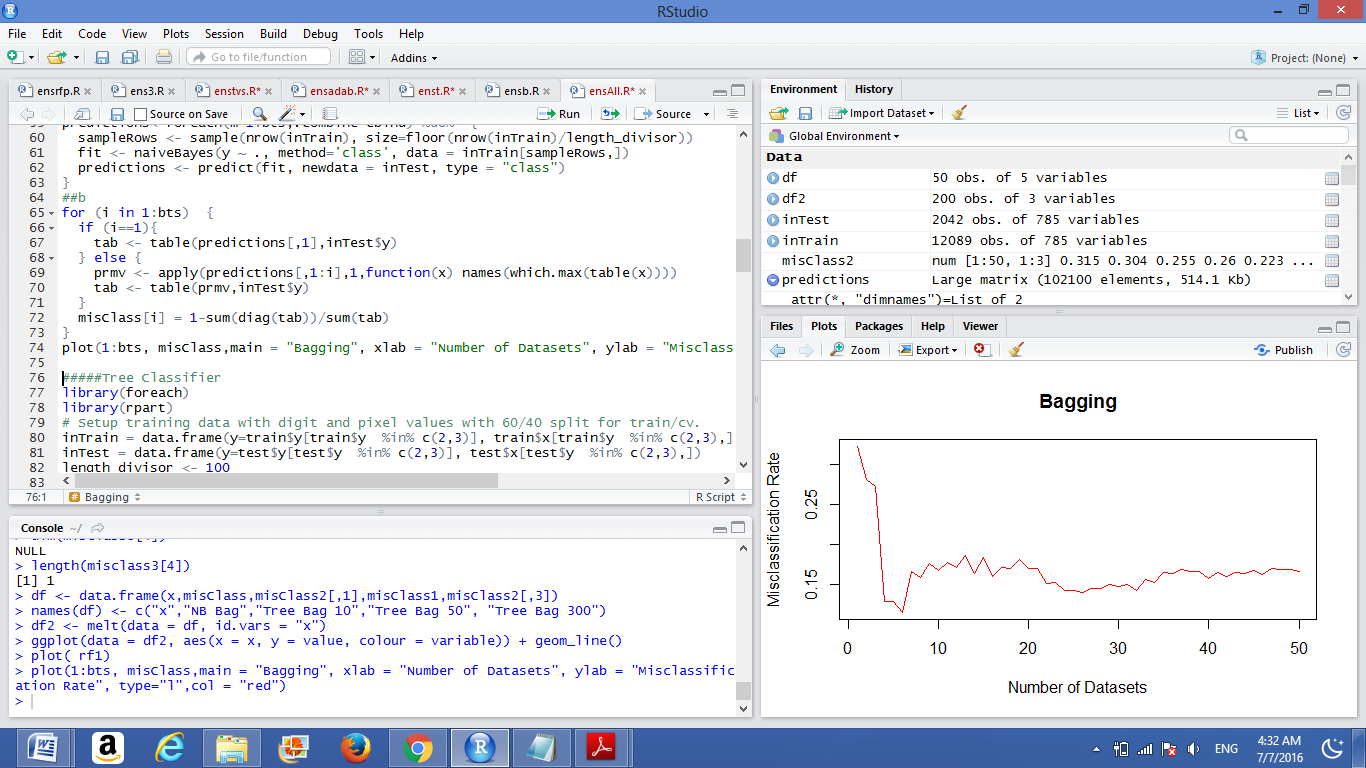
**Lab 3.C Applying Ensemble Learning**

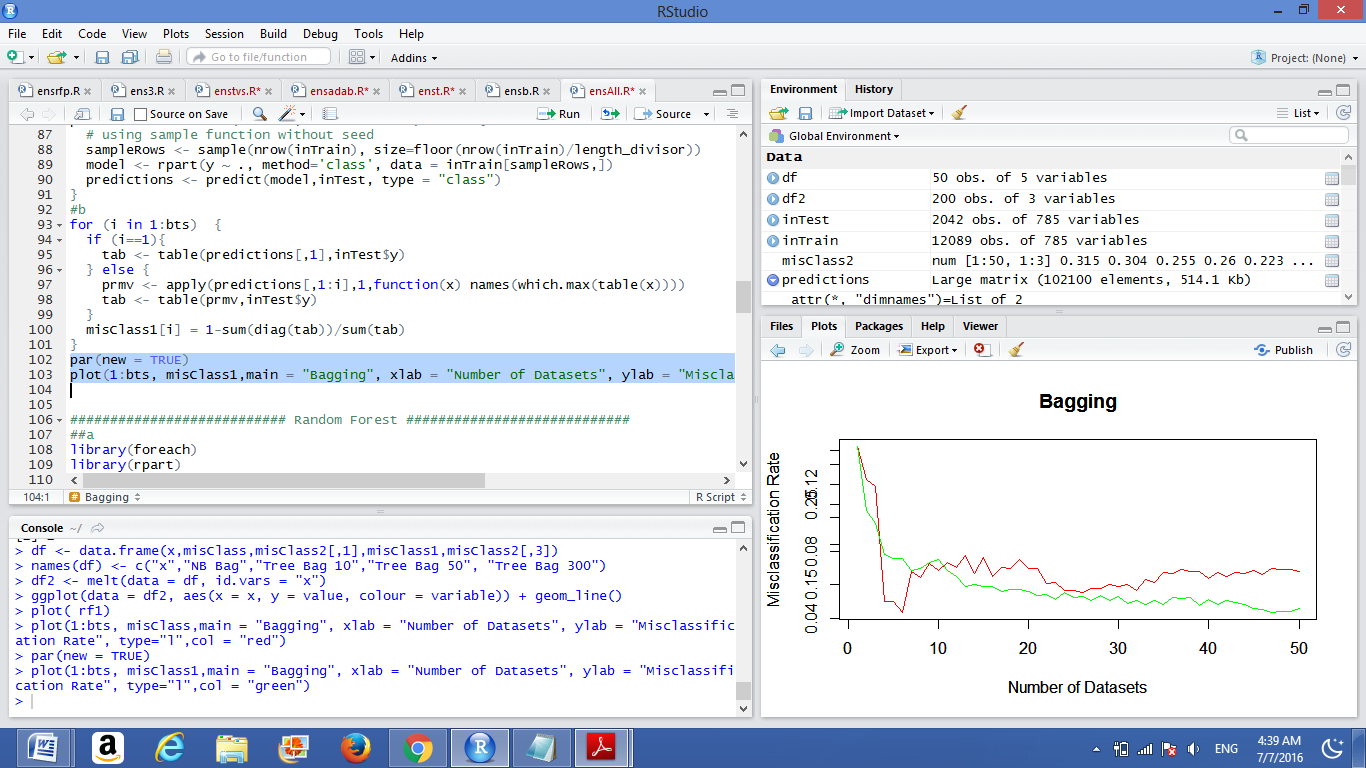
**Bagging**

**b.** Naive Bayes (red)

bootstraps <- 50



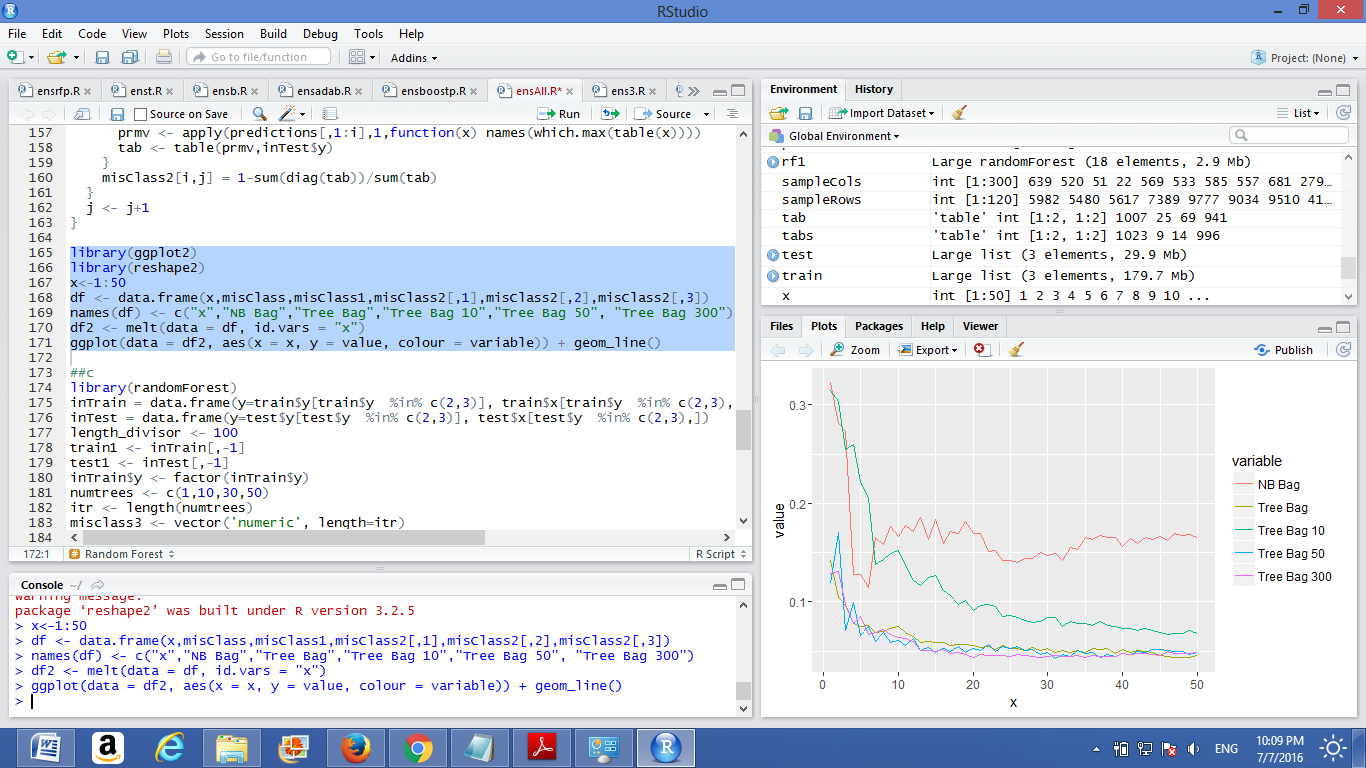
**b.** Tree (green)



Trees has less misclassification rates than Naive Bayes when we increase bootstraps samples (> 10) beacause the main idea of Bagging that it decreases error by decreasing the variancein the results due to *unstable learners*, algorithms (like decision trees and neural networks) whose output can change dramatically when the training data is slightly changed.

**Random Forest**

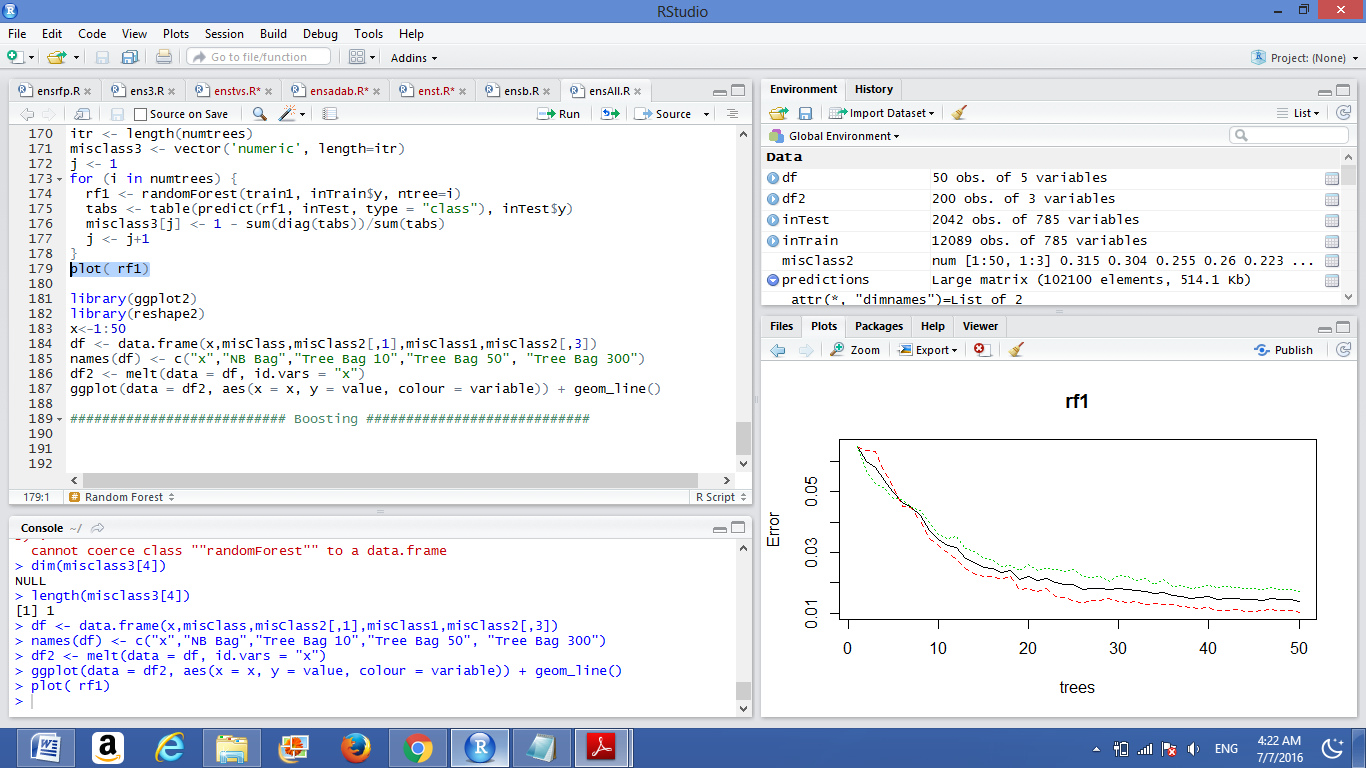
**b.**



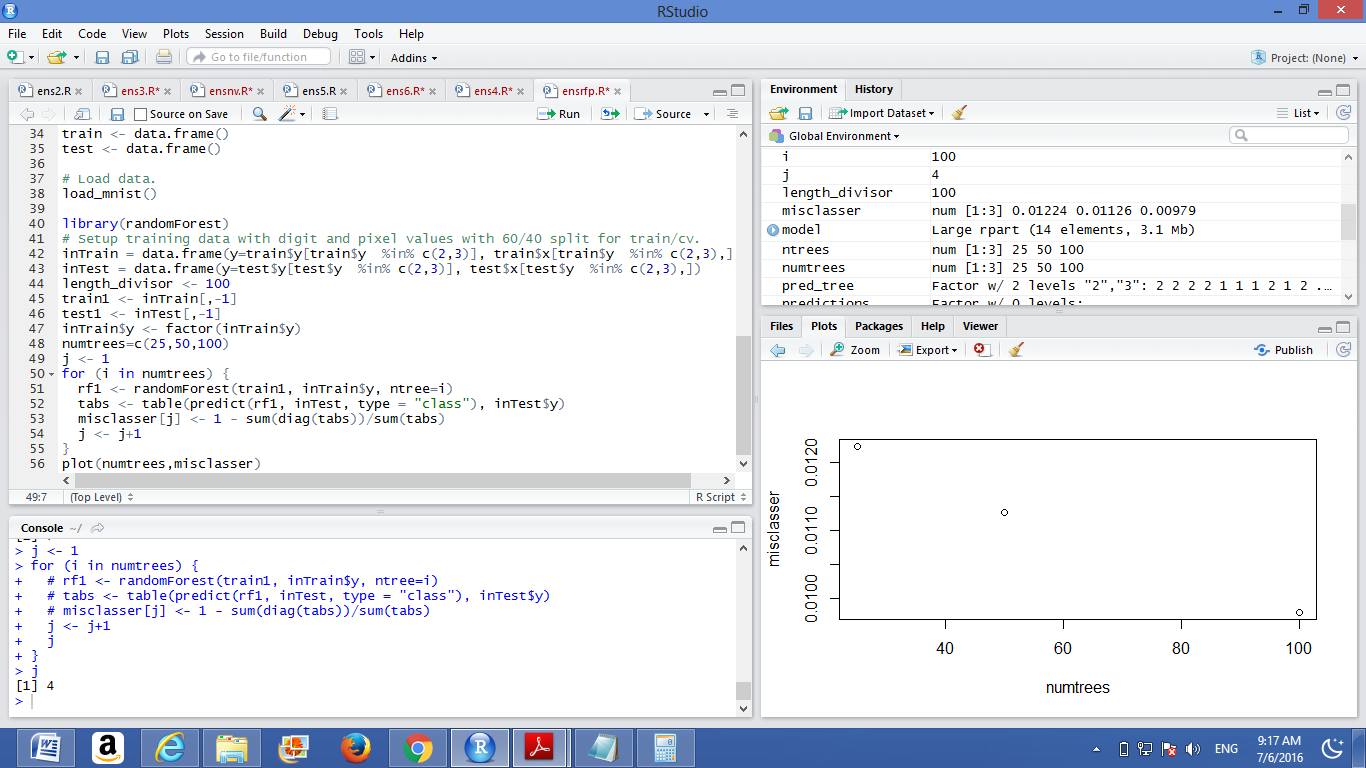
Misclassification error decreases as we increase number of feature selected to a certain number(nearly 50) which when we increase above no tangible improvement was achieved in error

**c.**

plot( rf1)



numtrees=c(25,50,100)



As we increase number of trees the misclassification error decreases (nearly linear)

Boosting

**a**.

Call:

ada(y ~ ., data = inTrain)

Loss: exponential Method: discrete Iteration: 50

Final Confusion Matrix for Data:

Final Prediction

True value 2 3

2 5930 28

3 48 6083

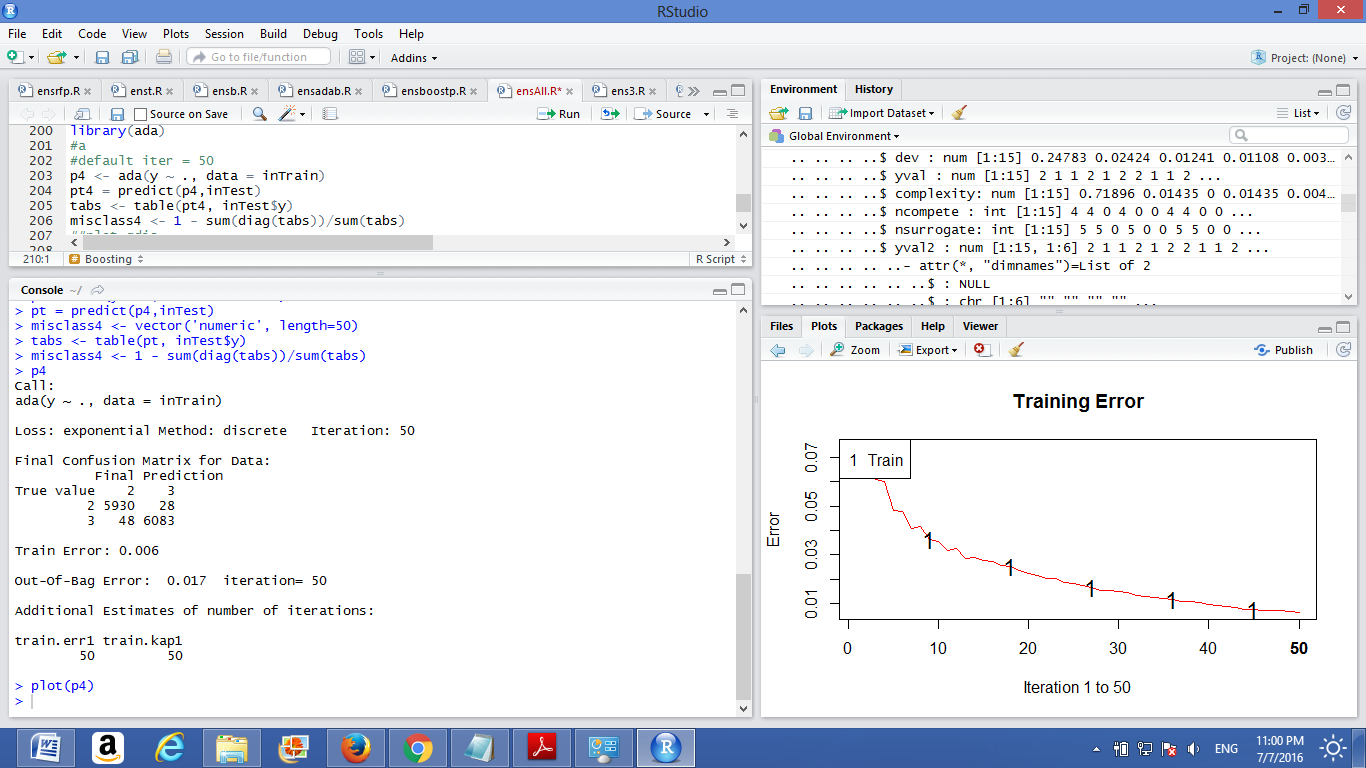
Train Error: 0.006

Out-Of-Bag Error: 0.017 iteration= 50

Additional Estimates of number of iterations:

train.err1 train.kap1

50 50



**b.**

Call:

ada(y ~ ., data = inTrain, control = rpart.control(maxdepth = 2))

Loss: exponential Method: discrete Iteration: 50

Final Confusion Matrix for Data:

Final Prediction

True value 2 3

2 5779 179

3 255 5876

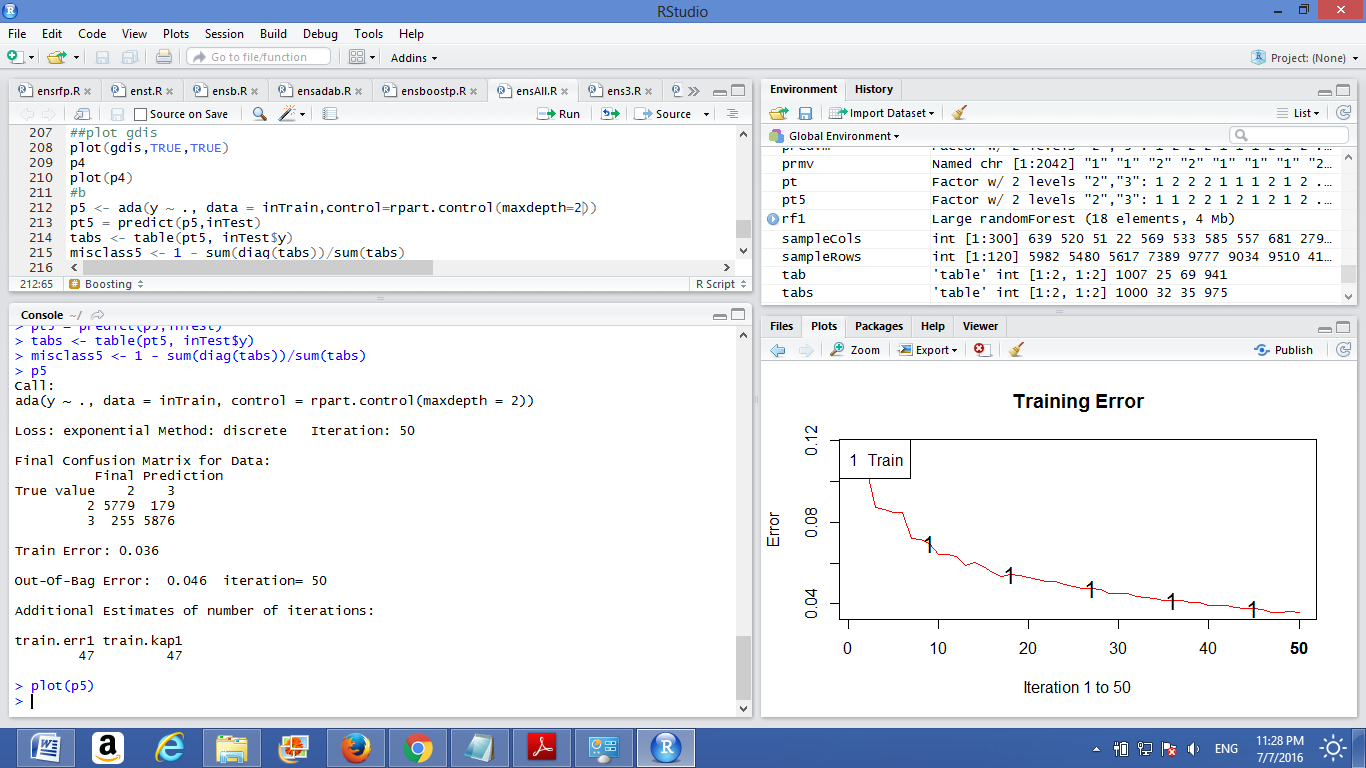
Train Error: 0.036

Out-Of-Bag Error: 0.046 iteration= 50

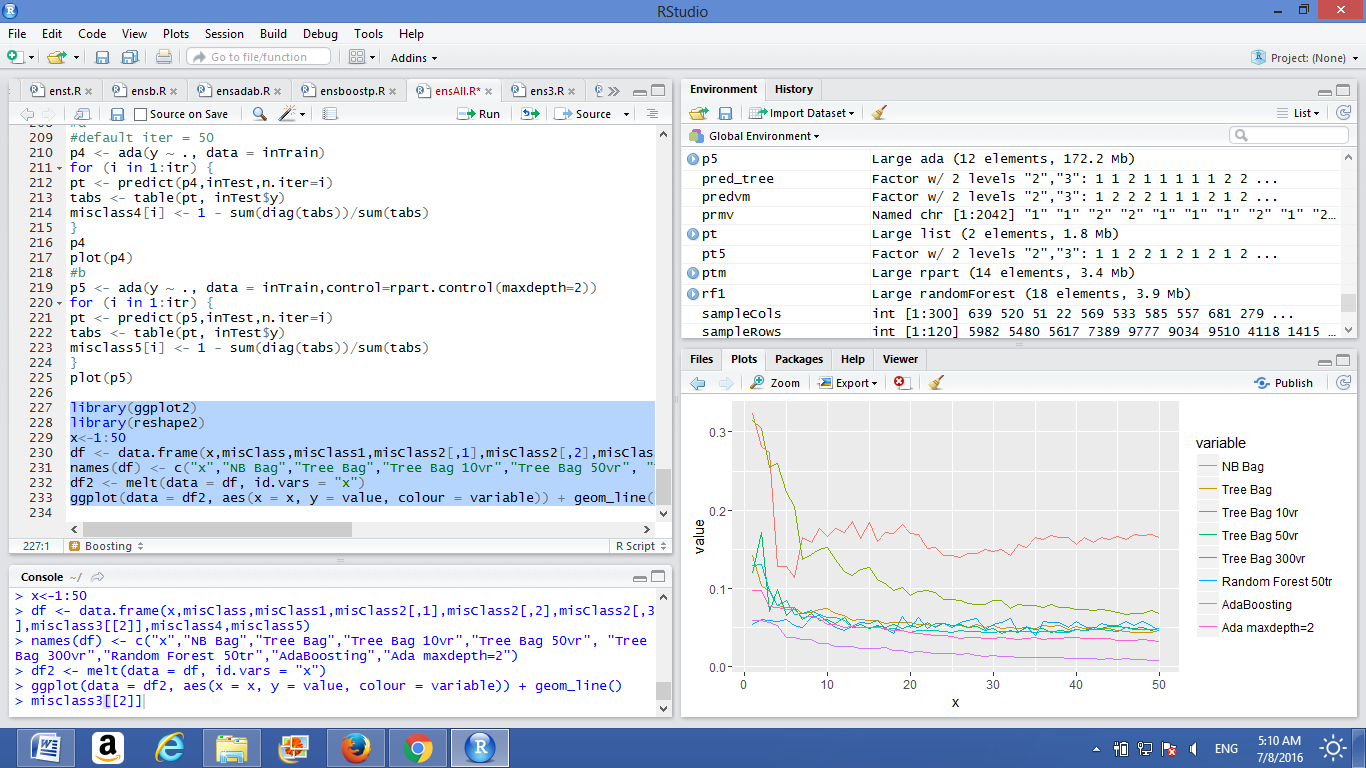
Additional Estimates of number of iterations:

train.err1 train.kap1

47 47

****

**c.**

****

Bagging reduces variance by bootstrapping the samples. Boosting is complex and reduces bias. random Forest is complex which reduces bias and reduces variance because it uses bagging.