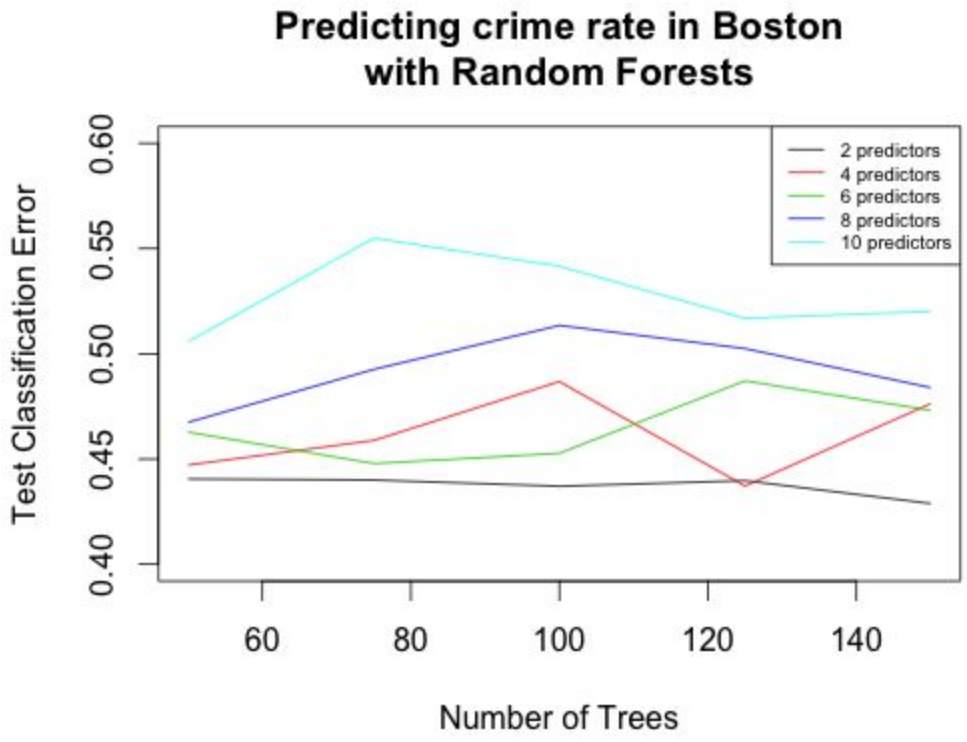
Tatiana Romanchishina
ISLR Ch 8
Ex 7 & 12 (using the Boston data set)

7. In the lab, we applied random forests to the Boston data using mtry=6 and using ntree=25 and ntree=500 . Create a plot displaying the test error resulting from random forests on this data set for a more comprehensive range of values for mtry and ntree . You can model your plot after Figure 8.10. Describe the results obtained.
**** I will try to predict **crim** instead of medv:

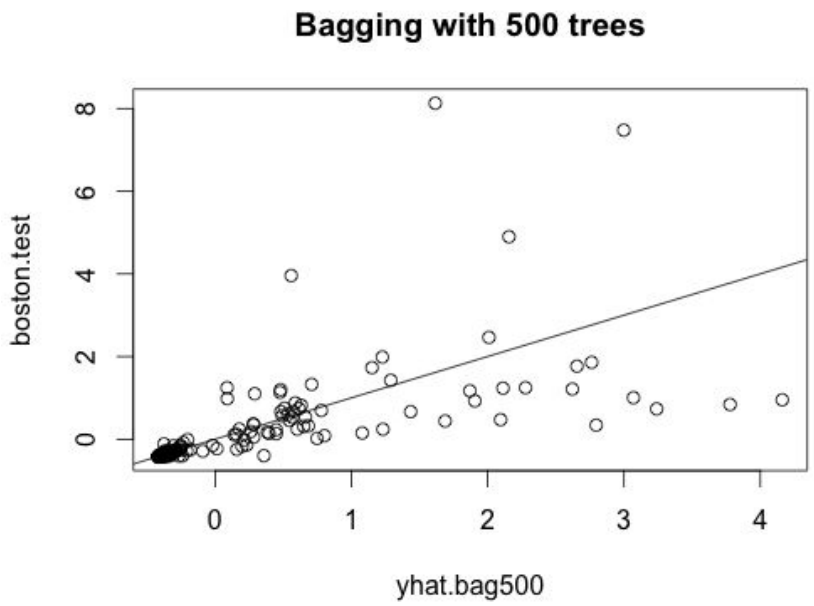


It seems that models with only 2 predictors have the lowest error in general. Within models with two predictors it seems that the prediction accuracy improves with the number of trees.

12. Apply boosting, bagging, and random forests to a data set of your choice. Be sure to fit the models on a training set and to evaluate their performance on a test set. How accurate are the results compared to simple methods like linear or logistic regression? Which of these approaches yields the best performance?

*** Random forests applied in the previous exercise.

- Bagging:
    - 500 trees

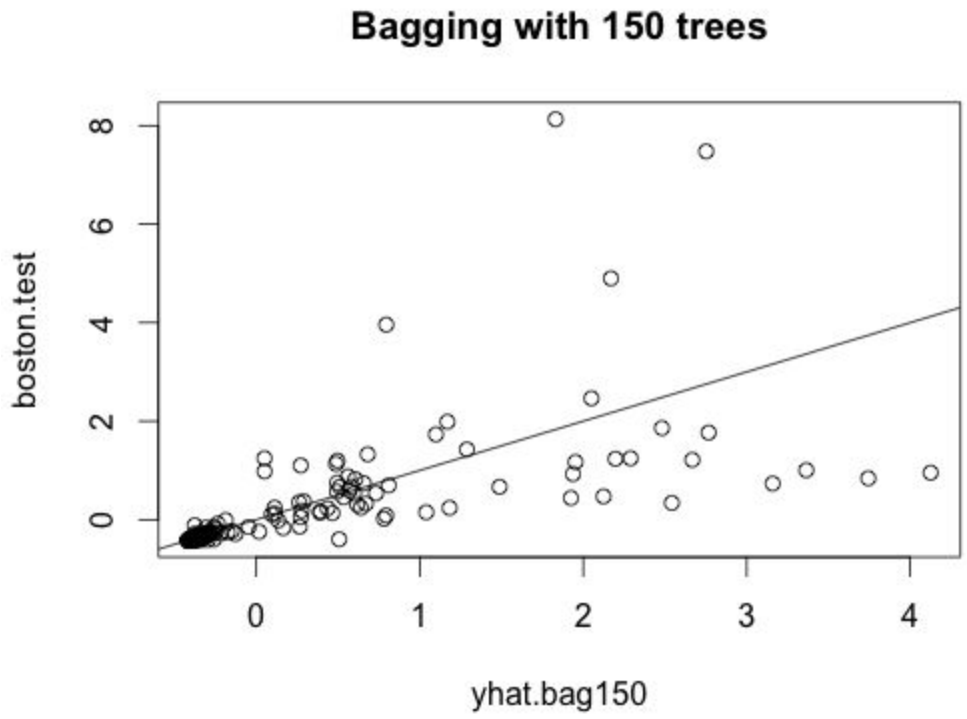


MSE = 0.5506461

    - 50 trees

MSE = 0.556215

- 150 trees



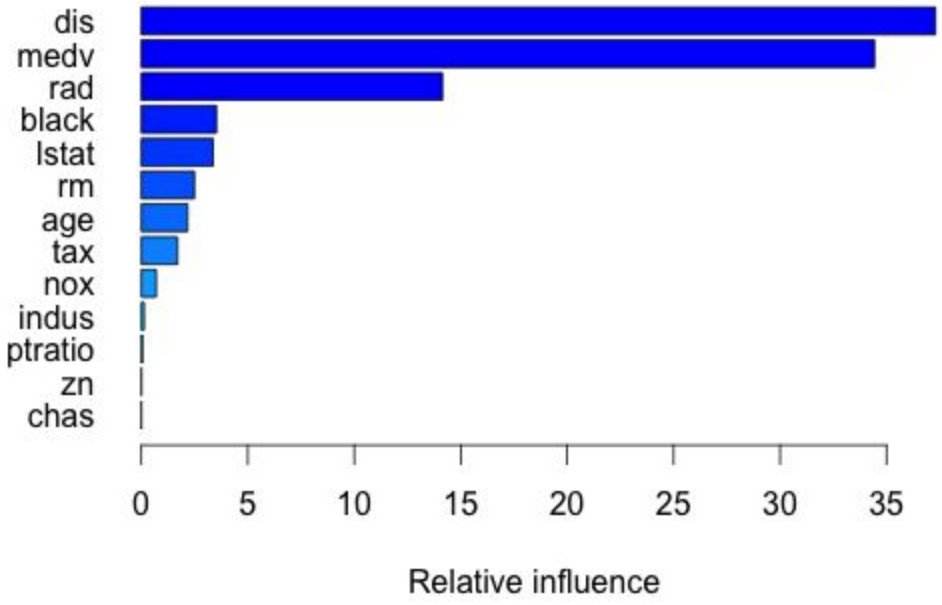
MSE = 0.5449557

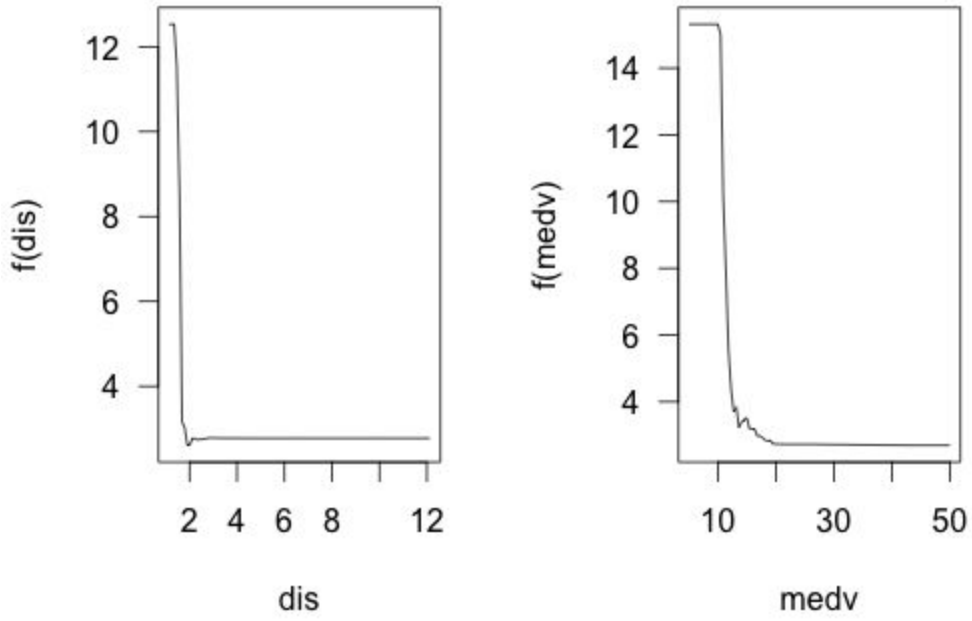
The test error rates were similar, the lowest achieved with 150 trees.

- Boosting:

```
> summary(boost.boston)
          var      rel.inf
dis         dis 3.800600e+01
medv       medv 3.398642e+01
rad         rad 1.364080e+01
black     black 3.570884e+00
lstat     lstat 3.503151e+00
rm           rm 2.529948e+00
age         age 2.311788e+00
tax         tax 1.457335e+00
nox         nox 7.855065e-01
indus     indus 1.564297e-01
ptratio ptratio 5.162057e-02
chas       chas 9.086349e-05
```

zn zn 1.487957e-05





Test error = 30.73486.

Conclusion: boosting produces the lowest testing error rate of 30.73486, while bagging error rate was >54 and the random forests error rate was between 31 and 45.