Exercise 5 – Bagged trees (Boston Housing Data)

```
##############   Bagging ##########################

# We apply Bagging and Random Forests on Boston data set
  # Here we apply bagging and random forests to the Boston data, using the randomForest package in R.
  # The exact results obtained in this section may depend on the version of R and the version of the randomForest package installed on your computer.
  # Recall that bagging is just a special case of random forests with m = p.

library(MASS)
library(randomForest)

# We will use the Boston data set included in the MASS library.
# It records Housing Values in Suburbs of Boston
?Boston # read the data set description

# Setup a training and test set for the Boston data set.
set.seed (1)
train = sample(1:nrow(Boston), nrow(Boston)/2)
boston.test=Boston[-train ,"medv"]

# Apply bagging using the randomForest package in R.
bag.boston=randomForest(medv~.,data=Boston,subset=train, mtry=13,importance =TRUE)
    # mtry = 13 means that we should use all 13 predictors for each split of the tree,
    # hence, do bagging.

# How well does the bagged model perform on the test set?
yhat.bag = predict(bag.boston,newdata=Boston[-train,])
plot(yhat.bag, boston.test)
abline(0,1)
mean((yhat.bag-boston.test)^2)
  # The test set MSE associated with the bagged regression tree is 13.16,
  # That's almost half that obtained using an optimally-pruned single tree
  # (investigate this on your own).

# Exercise: Change the number of trees grown by randomForest()
# using the ntree argument.
# For example, what happens to the MSE when we grow the tree from 13 to 25?
```