# HW8

Runye Hu A13410385
Yijie Fan A13485989
Xinyi He A13561164
Yifan Wu A14060535
Yiwen Cai A13530685
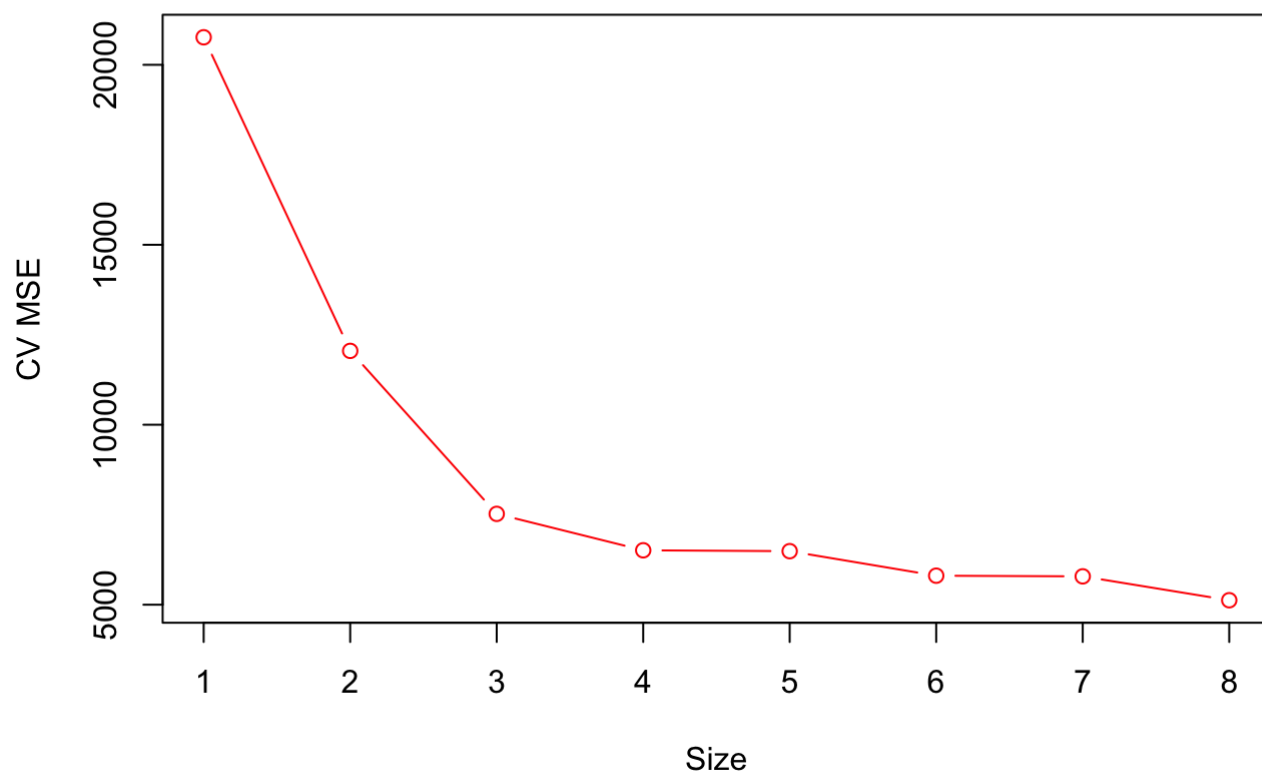
```
set.seed(123)
data <- read.csv("Boston.csv")
```

## Q1

```
size = nrow(data)
train_index = sort(sample.int(size,size/2))
train_set = data[train_index,]
test_set = data[-train_index,]
```

## Q2

```
library(tree)
tree_train = tree(medv~.,train_set)
cv_tree = cv.tree(tree_train, K = 6)
plot(cv_tree$size, cv_tree$dev, type = "b", xlab = "Size",ylab = "CV MSE", col = "red")
```
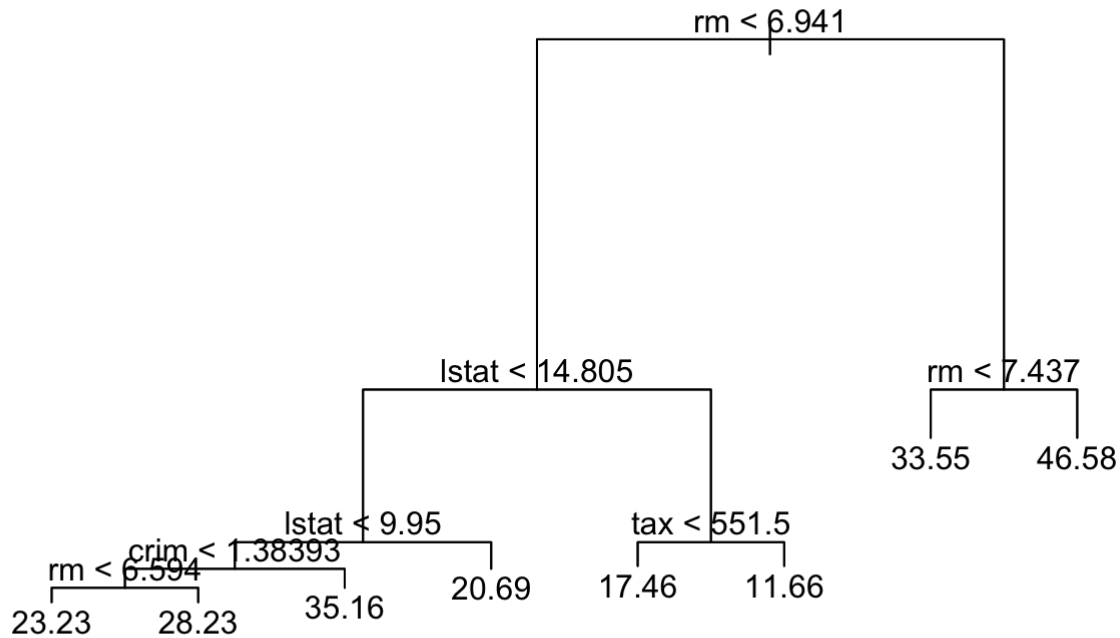
```
cv_size=cv_tree$size[which.min(cv_tree$dev)]
prune_tree = prune.tree(tree_train, best=cv_size)
plot(tree_train)
text(tree_train, pretty=0)
```
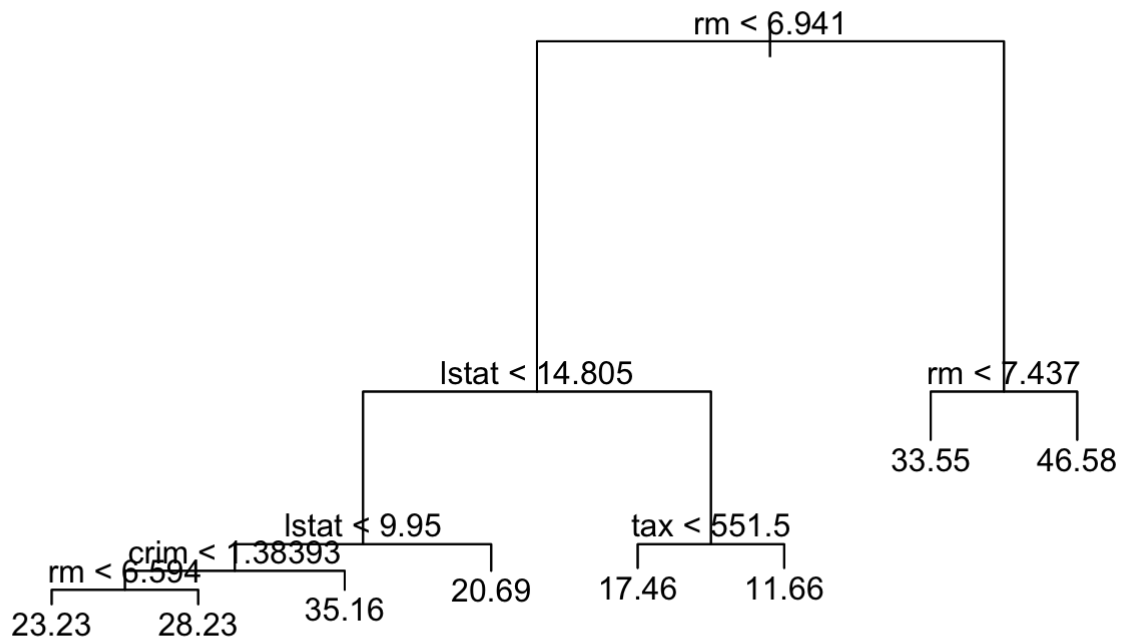
```
plot(prune_tree)
text(prune_tree, pretty = 0)
```

```
y_tree = predict(tree_train, newdata = test_set)
y_prune = predict(prune_tree, newdata = test_set)
y = test_set$medv
MSE_tree = mean((y_tree - y)^2)
MSE_tree
```

```
## [1] 23.51253
```

```
MSE_prune = mean((y_prune - y)^2)
MSE_prune
```

```
## [1] 23.51253
```

# Q3

```
library(ipred)
bagging <- bagging(medv~ .,train_set, nbagg=100 , method = "standard",coob=TRUE)
y_bagging = predict(bagging, newdata = test_set)
MSE_bagging = mean((y_bagging - y)^2)
MSE_bagging
```

```
## [1] 17.55542
```

```
library(randomForest)
```

```
## randomForest 4.6-14
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
random = randomForest(medv~., train_set, mtry = 4, ntree = 100)
y_random = predict(random, test_set)
MSE_random = mean((y_random - y)^2)
MSE_random
```

```
## [1] 12.70533
```

MSEs for both bagging and random forest are smaller than the MSE of single regression tree, and random forest gives out a better result.