TASK 2 : Fitting Regression Trees 20 points

library (MASS)  
require(tree)

## Loading required package: tree

## Warning: package 'tree' was built under R version 3.4.4

require(devtools)

## Loading required package: devtools

require(ISLR)

## Loading required package: ISLR

require(MASS)   
require(randomForest)

## Loading required package: randomForest

## Warning: package 'randomForest' was built under R version 3.4.4

## randomForest 4.6-14

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

require(gbm)

## Loading required package: gbm

## Loading required package: survival

## Loading required package: lattice

## Loading required package: splines

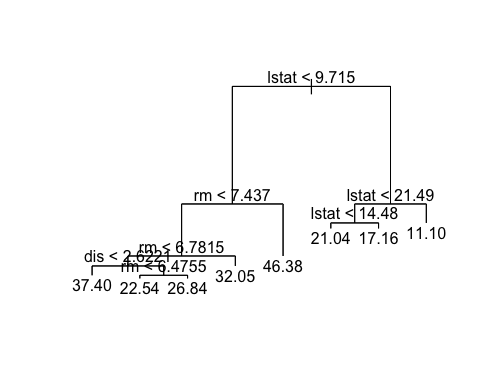
## Loading required package: parallel

## Loaded gbm 2.1.3

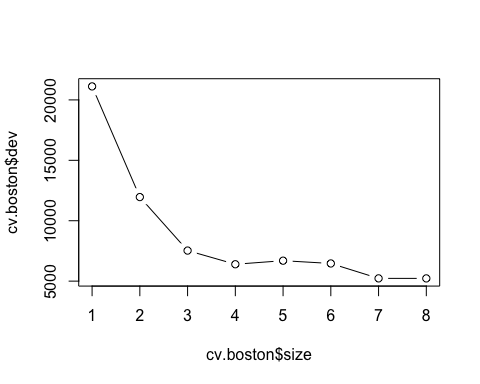
set.seed (1)  
train = sample (1: nrow ( Boston ), nrow( Boston )/2)  
tree.boston = tree(medv~.,Boston , subset = train )  
summary (tree.boston )

##   
## Regression tree:  
## tree(formula = medv ~ ., data = Boston, subset = train)  
## Variables actually used in tree construction:  
## [1] "lstat" "rm" "dis"   
## Number of terminal nodes: 8   
## Residual mean deviance: 12.65 = 3099 / 245   
## Distribution of residuals:  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -14.10000 -2.04200 -0.05357 0.00000 1.96000 12.60000

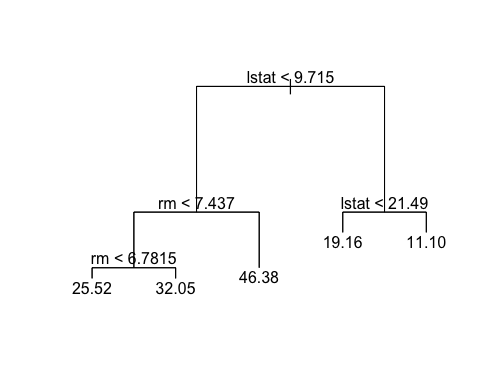
plot(tree.boston )  
text(tree.boston , pretty =0)



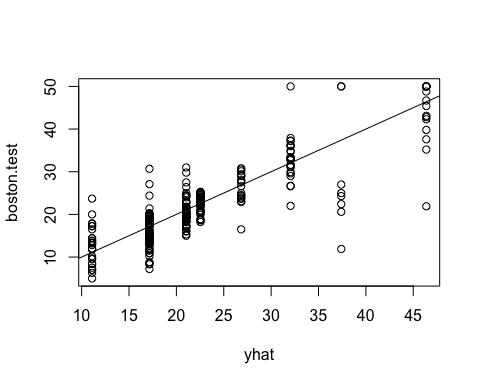
cv.boston =cv.tree( tree.boston )  
plot(cv.boston$size ,cv.boston$dev ,type="b")



prune.boston =prune.tree( tree.boston, best =5)  
plot( prune.boston )  
text( prune.boston , pretty =0)



yhat= predict (tree.boston , newdata = Boston [-train ,])  
boston.test= Boston [-train ,"medv"]  
plot(yhat ,boston.test )  
abline (0 ,1)



mean (( yhat - boston.test)^2)

## [1] 25.04559