Due date November 17, 2017.

1. (30 pts.) In class we considered the `Boston` data set from library `MASS`. It was found that the best tree includes predictors `lstat, rm` and `dis`. Fit a multiple linear regression model with these three predictors and find its MSPE using the same train and test sets. Which model has smallest MSPE?

2. (40 pts.) Consider the linear regression model that uses `horsepower` to predict `mpg`, using the `Auto` data set from library `ISLR` To obtain the bootstrap estimate of the $R^2$ fit the model using `m1=lm(...)` function. Assign the summary to a variable using `aux=summary(m1)`. The model $R^2$ is given by `aux$r.squared`. Use `set.seed(1)` before each time `boot()` is used.

   a) Write a bfunction to get the $R^2$ from a bootstrapped sample. Use the `boot()` function to find the standard errors of 1000 bootstrap estimates for the $R^2$.

   b) Use the `boot.ci()` function to find the bootstrap confidence interval for the $R^2$.

3. (30 pts.) Consider the `Hitters` data set from library `ISLR`. This data set has 322 observations of major league players on 20 variables. The response is `Salary` (make sure the rows with missing values are removed from the data set). Use `set.seed(1)` to divide the data set into a training and a test set (50%). It is of interest to predict the salary of the baseball players.

   a) Fit a regression tree on the training set. Use cross validation to find the best number of terminal nodes (use `set.seed(1)`).

   b) Which predictors are found most important? Report the test MSPE.