1. Consider the Boston housing data set, from the ISLR2 library.

(a) Based on this data set, provide an estimate for the population mean of "medv". Call this estimate $\hat{\mu}$.
(b) Provide an estimate of the standard error of $\hat{\mu}$. Recall, we can compute the standard error of the sample mean by dividing the sample standard deviation by the square root of the number of observations.
(c) Now estimate the standard error of $\hat{\mu}$ using the bootstrap. How does this compare to your answer from (b)?
(d) Based on your bootstrap estimate from (c), provide a 95 % normal confidence interval for the mean of "medv". Compare it to the results obtained using t.test(Boston$medv).
(e) Use sample median to estimate $\hat{m}$ for the median value of medv in the population.
(f) We now would like to estimate the standard error of $\hat{m}$. Unfortunately, there is no simple formula for computing the standard error of the median. Instead, estimate the standard error of the median using the bootstrap.

2. This problem involves the OJ data set which is part of the ISLR2 package. The data set contains sales information for Citrus Hill and Minute Maid orange juice. You may see the detail description of the data using ?OJ in R.
First create a training set containing a random sample of 800 observations, and a test set containing the remaining observations.

(1) Fit a tree to the training data, with Purchase as the label and the other variables except as features. Use the summary() function to produce summary statistics about the tree, and describe the results obtained. What is the training error rate? How many terminal nodes does the tree have?
(2) Create a plot of the tree. Pick one of the terminal nodes, and interpret the information displayed.
(3) Predict the labels on the test data, and produce a confusion matrix comparing the test labels to the predicted test labels. What is the test error rate?
(4) Apply the cv.tree() function to the training set in order to determine the optimal tree size. Produce a plot with tree size on the x-axis and cross-validated classification error rate on the y-axis. Which tree size corresponds to the lowest cross-validated classification error rate?
(5) Produce a pruned tree corresponding to the optimal tree size obtained using cross-validation. If cross-validation does not lead to selection of a pruned tree, then create a pruned tree with five terminal nodes.
(6) Compare the training and test error rates between the pruned and unpruned trees. Which is higher?
(7) Perform random forest on the training set with 1,000 trees for a chosen values of the "mtry". You may experiment with a range of values of the parameter.
(8) Which variables appear to be the most important predictors in the RF model?
(9) Use the RF model to predict the response on the test data. Form a confusion matrix. How does this compare with the result obtained using a single tree?