DS 636 Lab 7

1. The data set **npdb** (UsingR) contains information on malpractice awards in the United States. Attach the data set and make a table of the state variable. Which state had the most awards? (Using sort () on your table is useful here.)

2. The data set **MLBattend** (UsingR) contains attendance information for major league baseball between 1969 and 2000. The following commands will extract just the wins for the New York Yankees, in chronological order.

   ```
   > attach(MLBattend)
   > wins[franchise == "NYA"]
   [1] 80 93 82 79 80 89 83 97 100 100 89 103
   59 79 91
   …
   > detach(MLBattend) # tidy up
   ```

   Add the names 1969:2000 to your variable. Then make a barplot and dot chart showing this data in chronological order.

3. The data set **npdb** (UsingR) contains malpractice-award information. The variable amount is the size of malpractice awards in dollars. Find the mean and median award amount.

4. For the data sets **bumpers** (UsingR), **firstchi** (UsingR), and **math** (UsingR), make histograms. Try to predict the mean, median, and standard deviation. Check your guesses with the appropriate R commands.

5. The data set **DDT** (MASS) contains independent measurements of the pesticide DDT on kale. Make a histogram and a boxplot of the data. From these, estimate the mean and standard deviation. Check your answers with the appropriate functions.

6. There are several built-in data sets on the 50 United States. For instance, state.area (,) showing the area of each U.S. state, and state.abb (,) showing a common abbreviation. First, use state.abb to give names to the state.area variable, then find the percent of states with area less than New Jersey (NJ). What percent have area less than New York (NY)? Make a histogram of all the data. Can you identify the outlier?