1.5 Problems

1. Use the pennies data set for this problem. As always, be sure to read the description of the data set.

   (a) Compute the sample mean, sample median of the mass of these pennies.

   (b) Compute the sample standard deviation and sample variance of the mass of these pennies.

   (c) Compute the five-number summary of the mass of these pennies.

2. Use the pennies data set for this problem. As always, be sure to read the description of the data set if you have not already done so.

   (a) Create a histogram of all the penny masses in this data set, with a reasonable bin width. Don’t forget to label the axes and include a title for the histogram.

   (b) Describe the shape of this histogram.

   (c) Explain the shape of this histogram by looking also at the other variable in the data set, the year of the penny.

   (d) If you have done Problem 1, indicate why the numerical summaries of penny masses computed in that problem are not particularly useful.

3. Use the introductory statistics surveys data set for this problem. As always, be sure to read the description of the data set if you have not already done so.

   (a) Make a histogram of the shoe sizes of female students in the data set. Also make a histogram of the shoe sizes of male students in the data set. Include all relevant axis labels, and make the plots have the same axis limits to help in comparing the two histograms.

   (b) Make side-by-side boxplots of female and male shoe sizes in the data set, including all relevant axis labels.
(c) Interpret what you see in these plots, remembering that the data came from self-reporting in response to the question “What is your shoe size?” (Hint: the context of the data is quite important here. The obvious interpretation needs some additional thought to arrive at a correct interpretation.)

(d) Describe the distribution of the number of states visited by students in this data set, making any relevant plot(s).

4. Use the quizzes and finals data set to answer the question: do the distributions of quiz averages appear similar for the three sections in this data set? More specifically, compare and contrast these distributions by computing the relevant numerical and graphical summaries (preferably several of both) that we have studied so far.