

553.283 Introduction to R

Homework 2

Note 1: If a question asks you for a numerical answer, your submission for that question must consist of the R command that produces that answer followed immediately by the output.

Note 2: Please label all axes on any plots you create.

1. The dataset *npdb* (*UsingR*) contains information on malpractice awards in the United States.
 - a. Attach the data set and make a table of the *state* variable. Which state had the most awards?
 - b. The variable *amount* is the size of malpractice awards in dollars. Find the mean and median award amount. What percentile is the mean? Can you explain why this might be the case?
2. The dataset *pi2000* (*UsingR*) contains the first 2,000 digits of π . What percentage are 3 or less? What percentage of the digits are 5 or more?
3. The *time* variable of the *nym.2002* dataset contains the time to finish the 2002 New York City Marathon for a random sample of runners. Make a histogram and describe the shape of the data.
4. The use of a cell phone while driving is often thought to increase the chance of an accident. The dataset *reaction.time* (*UsingR*) is simulated data on the time it takes to react to an external event while driving. Subjects with *control* == 'C' are not using a cell phone, and those with *control* == 'T' are. Their time to respond to some external event is recorded in seconds. Create side-by-side boxplots of the variable *reaction.time* for the two values of *control*. Compare the centers and spread.
5. The data frame *x77* contains data from each of the fifty United States. First coerce the *state.x77* object (built-in) into a data frame with

$$x77 = \text{data.frame}(\text{state.x77})$$

For each of the following models, make a scatterplot and add the regression line.

- a. The model of illiteracy rate (*Illiteracy*) modeled by high school graduation rate *HS.Grad*.
- b. The model of life expectancy (*Life.Exp*) modeled by murder rate (*Murder*).
- c. The model of income (*Income*) modeled by *Illiteracy*.

6. The dataset *Cars93* (*MASS*) contains information on numerous cars. Make a scatter-plot of the variables *MPG.city* and *Price*, using the argument *pch = as.numeric(Type)* to change the shape of the points according to the variable *Type*.