Introduction to Data Organization and R Studio

Learning Objectives

- 1. Learn/review basics of data organization
- 2. Gain basic familiarity with R:
 - (a) Log in to R Studio server.
 - (b) Perform basic arithmetic in R.
 - (c) Perform object assignment
 - (d) Gain familiarity with help functionality in R.

Data Collection/Organization Activity

The objectives of this part of the activity are to think about how to organize data that you have collected in a spreadsheet and to get to know your fellow classmates. By the end of this section of the activity, you should have:

- 1. a spreadsheet in which you have organized data you collect about your classmates (details follow) and
- 2. *at least two* other people that you have identified to be your points of contact in this class for study/homework partners, notes/announcements if you miss class, etc.

Instructions:

- 1. For this activity, you will be collecting data on your classmates (and yourself). Specifically, you will be collecting information on the following *variables*: Name, Age, Class Year, Hometown, Preferred Pronouns, College Major, and whether they have taken a previous Statistics class (including AP Statistics).
- 2. Determine an efficient way to collect data from your classmates. This will take classwide coordination, so let's spend a minute or two discussing possible strategies.
- 3. Open an Excel spreadsheet (you also can use a Google sheet if you do not have Excel on your computer). Use this to record the data from your classmates. Organize it as you see fit later we will use this to discuss how computers read information. It doesn't matter if you are "right" in how you organize it. Make sure to include yourself, too! Save this file as: first_name_last_name_STAT_140_Lab0.csv.
- 4. Once you have finished your data collection (before the end of class), upload your data set to Moodle under the assignment "Lab 0 Part 1". Upload in a separate file (Word or Notepad or something similar) your name and the two people you have identified as points of contact in this class.

Getting Started with R Studio

- 1. In this class, we will be using R Studio through the Mount Holyoke server. You can access R by pointing a browser to https://rstudio.mtholyoke.edu/. Please log in to the R Studio server now.
- 2. You will notice that there are four panes within your R Studio session.
 - The top left pane is where your R script will go. This is where you write your programs and save them for execution in R.
 - The bottom left pane has several tabs. For now, we will only concern ourselved with the Console this is where R code gets excuted.
 - The top right pane also has several tabs. For now, we will focus on the Environment tab this keeps track of the items (which I may refer to as objects) that are stored in your R session.
 - The bottom right pane has several tabs we will want to use in the future, but for this lab, we will only be using the Help tab.
- 3. In the top left corner, there is a icon that looks like a blank sheet of paper with a white plus sign in a green circle in the upper left part of the paper click on this icon and select R Script. This will open a new R script in the top left pane. The following steps will be completed in this R script, which you should save for reference later.
- 4. **Basic arithmetic in R:** At its most basic, R is just a calculator that can perform arithmetic just as your handheld calculator can. While this might seem obvious, this is good place to start to get comfortable with R. For each of the following, start a new line in your script.
 - (a) Type a line in your R script to add 10 and 8 (like you would on a calculator) Keeping your cursor on this line of code, hit the following: CTRL+ENTER (PC) or COMMAND+RETURN (Mac). This will execute this line of code, and the line, 10+8, along with the result, 18, will show up in the console below.
 - (b) Subtract 10 from 8 using the same process.
 - (c) Multiple 10 by 8.
 - (d) Divide 10 by 8.
 - (e) Divide 10 by 0. What does this return?
 - (f) Calculate 10 to the 8th power. Use the caret (above your 6 key) symbol to accomplish this, just as you would on a calculator. Notice that R returns your answer in scientific notation: 1e+08. This is the same thing as 1×10^8 .
- 5. Help function: As you learn R, there will be many times when you need to consult documentation for a piece of code to find out what it needs to have to work properly and what the result of executing it is supposed to be. Here we will have a basic introduction to the help functionality available in R.
 - (a) To "get help" in R, use the ? followed by the name of the thing you are querying. Let's try this with log(). Type ?log in your script. Then

execute the line to pull up the help documentation in the lower right pane.

- (b) What specific logarithm does log calculate?
- (c) Under usage, the first line is: log(x, base = exp(1)).
 - i. In general terms, what is x?
 - ii. In general terms, what is base?
 - iii. What does log(exp(1)), base = exp(1)) return?
 - iv. What does log(exp(1)) return? Why do you think it is the same as what is returned for the code in (iii)?
 - v. What does log(exp(1), base = 10) return? Why is this different from what is returned for the code in (iii) and (iv)?
- (d) Try ?exp. This will pull up the same documentation as that for ?log, but if you scroll down, there is some information about the exp() function.
- *** Get comfortable with reading the help documentation in R this will help considerably with programming in the future.
- 6. **Object assignment:** If there is a value that we will be using over and over again in our code, it is useful to store it as an object. We use a left arrow key, followed by a dash to make an arrow for object assignment.
 - (a) Type x <- 10 in your R script and execute this line. Now on a separate line, type x and execute. This should return 10. If so, you have successfully assigned an object. You'll notice that an x shows up in your environment pane in the upper right of your R Studio.
 - (b) Type x <- x + 1 and execute this line. Now return x as before. What is the value of x now? You can overwrite previous object assignments. Sometimes you want to do this, but other times you don't, so you'll want to be careful.</p>
 - (c) Type y in your script and execute this line. Notice that you get an error message: Error: object 'y' not found. You haven't assigned a value to y, so R doesn't know what you are talking about. No worries - we can fix this.
 - (d) Type y <- x + 1 and execute this line. Notice that this preserves the assignment of x while also creating a new object, y.</p>
 - (e) Try typing X in your script and executing this line. Do you get the same error as in (c)? Why is this? What do you conclude about R with regard to case?
 - (f) You can assign an object to almost any combination of letters and numbers to create an object. There are some rules, however. Some of the details are included in this blog post:
 https://www.r-bloggers.com/rules-for-naming-objects-in-r/.

You may want to make a note of valid naming conventions if you are not already familiar with this topic from other programming languages. 7. Saving your work: You will need to save your work and submit it to Moodle. Details will be provided in class.