**Lab #1: R Basics**

Due Sunday, April 18

**To receive credit for this assignment, as a workgroup, submit a summary of your experience conducting the following commands, along with the RScript file you will create.** The summary does not have to be exhaustive or lengthy but do note any frustrations that arose and corrections you made when things went wrong. If everything went completely smooth – bravo! (To clarify, each group only needs to submit a single summary and RScript file, rather than individual submissions).

You will need to find a data set in order to accomplish this assignment. Some groups got a head start on this last week. If you come up empty handed after an enthusiastic search on several research sites reach out to me and I will help you locate something interesting to you. Please include a link to the data set you chose.

**Downloading R and RStudio**

First, install R on your computer at <https://cloud.r-project.org/>. Make sure you follow the instructions and download the most recent version (should be R 4.0.5 as of 4/12/2021).

While R is the language we are using, almost everyone who uses R does so within the free RStudio integrated development environment (IDE), so we will need to download that as well. Go to this website <https://www.rstudio.com/products/rstudio/download/> and download the free desktop version of RStudio. If you run into trouble, there are many resources online that should be able to help.

Next, we are going to briefly explore RStudio. It can be overwhelming at first, but don’t worry, you’ll get the hang of it quickly! There are tons of YouTube videos with help for getting started in RStudio, so look those up if you need more help.

**The R Studio Screen**

The R Studio screen is split into 4 areas.

1. The Console (Bottom Left) – where you input commands and where the output is usually displayed.
2. The File Editor and Data Browser Window (Top Left) – when open, this window allows you to look at a two-dimensional, cell-based display of your data and edit special R files.
3. The Environment & History Window (Top Right) – provides a list of objects (like datasets) currently loaded and usable in the Environment tab, and a history of all the commands you typed in the History tab.
4. The Utility Tab (Bottom Right) – provides a list of files you can open in the Files tab, displays plots and graphs in the Plots tab, shows a list of currently installed packages in the Packages tab, and finally shows help documentation in the Help tab.

Graphical user interface, application

Description automatically generated

**Opening a New File**

There are a few ways to run code in RStudio. You can run it in the console (bottom left pane), but this is not recommended because it is not very reproducible and is difficult to edit your code if it doesn’t run. RMarkdown is one great way to write reproducible code that can easily become a professional report, and we will get into RMarkdown into the next lab.

For now, we will focus on RScripts, which allow you to easily write and edit code. When you run lines of an RScript, the output will be generated in the console and plots will be generated in the plot window (bottom right pane). To open a new RScript, go to top left and select “R Script”. Use your new file to execute the commands in the rest of this lab.

If you run into issues, get into the habit of searching for help on google. It is likely someone else has encountered the same error as you and asked for help online!

**Loading Data**

The first step to doing any data analysis in R is loading the data into R. Click the “Import Dataset” button in the upper left pane, which is shown below.

Graphical user interface, text, application, email

Description automatically generated

|  |
| --- |
| Loading Datasets with RStudio |
| For R Data (.rdata files):  File > Open > Select the .rdata File  For Other Data Files, (e.g., .xlsx, .csv., .dta):  File > Import > from file type  Browse and select file  Make sure to give your data set a name! |

Keep in mind that using this point-and-click method will not always be the easiest way to load in data. For now, use this method and we will explore other options in the next labs.

**Viewing and Sorting Data**

|  |
| --- |
| Basic Data Viewing and Editing |
| To browse your dataset (make sure to use a capital V, as R is always case sensitive):  View (*datasetName*)  Typically, when referring to a variable in a command you must refer to the dataset the variable is in as follows:  *datasetName$variableName*  \*Note: in a “tidy” data frame, a variable should correspond to a column in the dataset  To sort your data by a variable:  sort (*variableName*) |

To run commands, either click the “Run” button in the upper righthand corner or use Cmd+Enter on Mac or Cntrl+Enter on Windows to run the current line or multiple highlighted lines.

**Basic Statistical Attributes**

|  |
| --- |
| Viewing Data Summaries |
| To view a list of all variables in your dataset:  names (*datasetName*)  To view the basic structure of all variables in your dataset:  str (*datasetName*)  To view a basic summary of a variable, (e.g., mean, median, min, max):  summary (*variableName*)  To view a more detailed summary of a variable, e.g., N, SD, kurtosis:  library (*psych*)  describe (*variableName*) |

Congratulations!!! You have now familiarized yourself with the basics of R!!! You are now almost ready to learn how to analyze the data you’ve prepared for analysis.

**Remember to submit a summary of your experience and the RScript you used!**

\*This lab was adopted from Dr. Jessica Feezell and John K. Wagner of the University of New Mexico. I am deeply appreciative for their generosity.