**Take home exam 1 – Due March 7th at 11:59 pm**

This Take-Home Exam 1 consists of 10 extended-response type questions worth 10 points each (100 points total). Answer each question with supporting material as necessary (e.g., code, a solid paragraph of lucid text, supported by figures, tables, etc.). The expectation is that most questions will be answered with a solid paragraph worth of text or code. For each question, be sure to address all the items that are included in that question.

The Take Home part is “open-book”, meaning that you can consult library and online resources. Please do not discuss questions with classmates and be aware that your responses to each question must be uniquely yours and in your own words. Also remember, there may be some self-learning involved in answering some questions. The questions are designed so you synthesize information from lecture content and coding assignments.

**Question 1.** You are looking at a collaborator’s R code on github, and download the repository, and start exploring the code. The first line of the script is

setwd(“C:/Users/…”)

* What is the author of this code trying to do with the function setwd()?
* Please discuss what is wrong with this approach in terms of reproducibility.
* Where is the working directory of an **R project**?
* Explain the concept of *relative* file paths. Is the author of this code using *relative* file paths?

**Question 2.** What does the acronym FAIR stand for in the context of this class? Explain how R, GitHub, and other lecture concepts introduced in this course specifically help complete FAIR data principles.

**Question 3.** Explain the concept of R packages. What are R packages? Who writes R packages? What is the difference between installing and loading a package? Explain two ways to install and load packages into R.

**Question 4.** Explain the following concepts of ggplot and give examples of each concept using code and figures generated with ggplot using the data of your choosing.

* Layering
* Scales
* Themes
* Facets

**Question 5.** Explain the differences and similarities between a vector, matrix, and dataframe. Demonstrate you know how to subset a dataframe in two ways using the built in dataset ‘ToothGrowth’ with the prompts below:

* Subset ToothGrowth to include rows such that supp is equal to VC
* Subset ToothGrowth to include rows such that supp is equal to VC and dose is equal to 0.5
* Subset ToothGrowth to include the values of len such that supp is equal to VC and dose is equal to 0.5

**Question 6.** Create an R markdown version of your answer to question 4 and 5. Save the .Rmd file to your computer and render it as a word document (.docx), .html, and a .md file. Push these files to your github and paste your github url here.

**Question 7.** What is the correct order of events to get your code on github through R studio? Explain each step from creation of a repository to pushing.

**Question 8.** After you have worked on a project for a while, you mistakenly delete a file on your github, while it still exists in your local repository (on your computer). Now when you try to push your code to github the push is rejected and gives the following error, “***Updates were rejected because the remote contains work that you do not have locally***.” How do you solve this error?

**Question 9.** Explain the purpose of a Data Management Plan.

**Question 10.** A colleague gives you data in an .xlsx file that looks like this:

A picture containing treemap chart

Description automatically generated

Please discuss at least five things wrong with how these data are formatted that make it not reproducibility friendly. Then describe/show your colleague how the data should be formatted.