**Data wrangling – 25 pts**

PLEASE READ THIS BEFORE CONTINUING

This assignment will help you practice integrating some of the tidyverse functions into your R scripts. It will also involve some more practice with GitHub. You may collaborate with a partner to enhance your learning experience. Please ensure the following:

* **Collaboration**: If you work with a partner, include both names on the final submission by editing the YAML header.
* **Submission**: Only one person should submit the assignment to Canvas *in a Word document or .pdf file generated through R markdown*. Additionally, you should provide a link to your GitHub, where the assignment should be viewable by rendering it as a GitHub-flavored markdown file.
* **Setup**: It is also assumed you already have a GitHub repository for this class.
* **Time**: This should take you no longer than the class period to complete.

1. 3 pts. Download two .csv files from Canvas called DiversityData.csv and Metadata.csv, and read them into R using relative file paths.
2. 4 pts. Join the two dataframes together by the common column ‘Code’. Name the resulting dataframe *alpha*.
3. 4 pts. Calculate Pielou’s evenness index: Pielou’s evenness is an ecological parameter calculated by the Shannon diversity index (column Shannon) divided by the log of the richness column.
   1. Using mutate, create a new column to calculate Pielou’s evenness index.
   2. Name the resulting dataframe *alpha\_even*.
4. 4. Pts. Using tidyverse language of functions and the pipe, use the summarise function and tell me the mean and standard error evenness grouped by crop over time.
   1. Start with the *alpha\_even* dataframe
   2. Group the data: group the data by Crop and Time\_Point.
   3. Summarize the data: Calculate the mean, count, standard deviation, and standard error for the even variable within each group.
   4. Name the resulting dataframe *alpha\_average*
5. 4. Pts. Calculate the difference between the soybean column, the soil column, and the difference between the cotton column and the soil column
   1. Start with the *alpha\_average* dataframe
   2. Select relevant columns: select the columns Time\_Point, Crop, and mean.even.
   3. Reshape the data: Use the pivot\_wider function to transform the data from long to wide format, creating new columns for each Crop with values from mean.even.
   4. Calculate differences: Create new columns named diff.cotton.even and diff.soybean.even by calculating the difference between Soil and Cotton, and Soil and Soybean, respectively.
   5. Name the resulting dataframe *alpha\_average2*
6. 4 pts. Connecting it to plots
   1. Start with the *alpha\_average2* dataframe
   2. Select relevant columns: select the columns Time\_Point, diff.cotton.even, and diff.soybean.even.
   3. Reshape the data: Use the pivot\_longer function to transform the data from wide to long format, creating a new column named diff that contains the values from diff.cotton.even and diff.soybean.even.
      1. This might be challenging, so I’ll give you a break. The code is below.

A graph of a line

AI-generated content may be incorrect.pivot\_longer(c(diff.cotton.even, diff.soybean.even), names\_to = "diff")

* 1. Create the plot: Use ggplot and geom\_line() with ‘Time\_Point’ on the x-axis, the column ‘values’ on the y-axis, and different colors for each ‘diff’ category. The column named ‘values’ come from the pivot\_longer. The resulting plot should look like the one to the right.

1. 2 pts. Commit and push a gfm .md file to GitHub inside a directory called Coding Challenge 5. Provide me a link to your github written as a clickable link in your .pdf or .docx