CodingChallenge\_DataWrangling

Vaibhav B. Shelar

2025-03-23

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.1 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.4 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.4   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

### Q2  
  
# load csv file  
  
diversity\_data <- read.csv("DiversityData.csv")  
  
meta\_data <- read.csv("Metadata.csv")  
  
# join the dataframe by the common column  
  
alpha <- left\_join(diversity\_data, meta\_data, by = "Code")

### Q3  
  
# a)  
  
library(dplyr)  
  
# calculating Pielou's evenness index  
  
alpha$Pielou\_evenness <- log(alpha$richness)  
  
# Create a new column called Pielou\_evenness and name the dataframe as alpha\_even  
alpha\_even <- alpha %>%  
 mutate(Pielou\_evenness = shannon / log(richness))

### Q4

# summarizing Pielou's evenness by Crop and Time\_Point and Summarize the data  
  
alpha\_average <- alpha\_even %>%  
 group\_by(Crop, Time\_Point) %>%  
 summarise(mean\_even = mean(Pielou\_evenness),  
 count\_even = n(),  
 sd.dev = sd(Pielou\_evenness),  
 sd.err = sd(Pielou\_evenness)/sqrt(n()))

## `summarise()` has grouped output by 'Crop'. You can override using the  
## `.groups` argument.

### Q5

# Calculate the difference between the soybean column, the soil column, and the difference between the cotton column and the soil column

alpha\_average2 <- alpha\_average %>%  
 select(Time\_Point, Crop, mean\_even) %>%  
 pivot\_wider(names\_from = Crop, values\_from = mean\_even) %>%  
 mutate(diff.cotton.even = Soil - Cotton,  
 diff.soybean.even = Soil - Soybean)

### Q6

# Connecting it to plots

alpha\_average2 %>%  
 select(Time\_Point, diff.cotton.even, diff.soybean.even) %>% # Select relevant columns  
 pivot\_longer(cols = c(diff.cotton.even, diff.soybean.even), names\_to = "diff") %>%  
 ggplot(aes(x = Time\_Point, y = value, color = diff)) +  
 geom\_line(size = 1) + # Line plot  
 theme\_minimal() + # Clean theme  
 labs(  
 x = "Time (hrs)",  
 y = "Difference from soil in Pielou’s evenness",  
 color = "diff"  
 ) +   
 theme(legend.position = "right")

## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.  
## ℹ Please use `linewidth` instead.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was  
## generated.

