Coding Challenge 5 - Data Wrangling

Olivia Brown

2025-03-20

[Link to Github](https://github.com/olixiabrown/PLPA5820)

#### Q1 - load in required packages and dataframe

library(tidyverse)

## Warning: package 'tidyr' was built under R version 4.4.1

## Warning: package 'purrr' was built under R version 4.4.1

## Warning: package 'lubridate' was built under R version 4.4.1

## ── 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

getwd()

## [1] "/Users/oliviabrown/Desktop/R directories and code/PLPA5820"

metadata <- read.csv("CodingChallenge5/Metadata.csv")  
diversitydata <- read.csv("CodingChallenge5/DiversityData.csv")

#### Q2 - join dataframes by code column

head(metadata)

## Code Crop Time\_Point Replicate Water\_Imbibed  
## 1 S01\_13 Soil 0 1 na  
## 2 S02\_16 Soil 0 2 na  
## 3 S03\_19 Soil 0 3 na  
## 4 S04\_22 Soil 0 4 na  
## 5 S05\_25 Soil 0 5 na  
## 6 S06\_28 Soil 0 6 na

head(diversitydata)

## Code shannon invsimpson simpson richness  
## 1 S01\_13 6.624921 210.7279 0.9952545 3319  
## 2 S02\_16 6.612413 206.8666 0.9951660 3079  
## 3 S03\_19 6.660853 213.0184 0.9953056 3935  
## 4 S04\_22 6.660671 204.6908 0.9951146 3922  
## 5 S05\_25 6.610965 200.2552 0.9950064 3196  
## 6 S06\_28 6.650812 199.3211 0.9949830 3481

alpha <- full\_join(metadata, diversitydata, by = "Code")

#### Q3 - create alpha\_even dataframe that includes pielou’s evenness index

#calculate log richness  
alpha <- mutate(alpha, logRich = log(richness))  
head(alpha)

## Code Crop Time\_Point Replicate Water\_Imbibed shannon invsimpson simpson  
## 1 S01\_13 Soil 0 1 na 6.624921 210.7279 0.9952545  
## 2 S02\_16 Soil 0 2 na 6.612413 206.8666 0.9951660  
## 3 S03\_19 Soil 0 3 na 6.660853 213.0184 0.9953056  
## 4 S04\_22 Soil 0 4 na 6.660671 204.6908 0.9951146  
## 5 S05\_25 Soil 0 5 na 6.610965 200.2552 0.9950064  
## 6 S06\_28 Soil 0 6 na 6.650812 199.3211 0.9949830  
## richness logRich  
## 1 3319 8.107419  
## 2 3079 8.032360  
## 3 3935 8.277666  
## 4 3922 8.274357  
## 5 3196 8.069655  
## 6 3481 8.155075

#calculate PEI and create a new df with PEI  
alphaeven <- mutate(alpha, PEI = (shannon/logRich))

#### Q4 - find mean and SE using pipes

#fix this   
alpha\_average <- alphaeven %>%  
 group\_by(Crop, Time\_Point) %>%   
 summarise(Mean.PEI = mean(PEI),  
 n = n(),   
 sd.dev = sd(PEI),  
 std.err = sd.dev/sqrt(n))

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

#### Q5 - reshape data by crop

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

#### Q6 - reshape to long and make a ggplot

alpha\_average2 %>%  
 select(Time\_Point, diff.cotton.even, diff.soybean.even) %>%  
 pivot\_longer(c(diff.cotton.even, diff.soybean.even), names\_to="diff")%>%  
 ggplot(aes(x=Time\_Point, y=value, color=diff))+  
 geom\_line()+  
 xlab("Time (hrs)")+ylab("Difference from soil in Pielou's evenness")

