DataWrang

Deekshya

2025-03-20

```r
Diversity.data <- read.csv("DiversityData.csv")
Metadata <- read.csv("Metadata.csv", na.strings="na")
```

2. 4 pts. Join the two dataframes together by the common column 'Code'. Name the resulting dataframe alpha

```r
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.3.3
```

```
## Warning: package 'ggplot2' was built under R version 4.3.3
```

```
## Warning: package 'tidyr' was built under R version 4.3.3
```

```
## Warning: package 'readr' was built under R version 4.3.2
```

```
## Warning: package 'dplyr' was built under R version 4.3.3
```

```
## Warning: package 'stringr' was built under R version 4.3.2
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
head(Diversity.data)
```

```
##     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
```

```r
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
```

```r
alpha = (left_join(Diversity.data, Metadata, by = "Code"))
```

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.

   a. Using mutate, create a new column to calculate Pielou's evenness index.
   b. Name the resulting dataframe alpha_even.

```r
alpha_even = alpha %>%
  mutate(Pielou_evenness = shannon / log(richness))
```

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.

   a. Start with the alpha_even dataframe
   b. Group the data: group the data by Crop and Time_Point.
   c. Summarize the data: Calculate the mean, count, standard deviation, and standard error for the even variable within each group.
   d. Name the resulting dataframe alpha_average

```r
alpha_average = alpha_even %>%
  group_by(Crop, Time_Point) %>%
  mutate(log_evenness = log(Pielou_evenness)) %>%
  summarize(
    mean_evenness = mean(Pielou_evenness),
    n = n(),
    sd_evenness = sd(Pielou_evenness)
  ) %>%
  mutate(se_evenness = sd_evenness / sqrt(n))
```

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

5. 4. Pts. Calculate the difference between the soybean column, the soil column, and the difference between the cotton column and the soil column

   a. Start with the alpha_average dataframe
   b. Select relevant columns: select the columns Time_Point, Crop, and mean.even.
   c. 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.

d. 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.

e. Name the resulting dataframe alpha_average2

```
alpha_average2 = alpha_average %>%
  select(Time_Point, Crop, mean_evenness) %>%   # Select relevant columns
  pivot_wider(names_from = Crop, values_from = mean_evenness) %>%   # Transform data
  mutate(
    diff.soybean.even = Soil - Soybean,   # Difference between Soybean and Soil
    diff.cotton.even = Soil - Cotton      # Difference between Cotton and Soil
  )
```

6. 4 pts. Connecting it to plots

a. Start with the alpha_average2 dataframe

b. Select relevant columns: select the columns Time_Point, diff.cotton.even, and diff.soybean.even.

c. 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.

d. This might be challenging, so I'll give you a break. The code is below.

pivot_longer(c(diff.cotton.even, diff.soybean.even), names_to = "diff")

d. 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.

```
colnames(alpha_average2)
```

```
## [1] "Time_Point"        "Cotton"             "Soil"
## [4] "Soybean"           "diff.soybean.even" "diff.cotton.even"
```

```
alpha_average2 %>%
select(Time_Point, diff.cotton.even, diff.soybean.even) %>%
pivot_longer(cols = c(diff.cotton.even, diff.soybean.even),
             names_to = "diff",
             values_to = "values") %>%
  ggplot(aes(x = Time_Point, y = values, color = diff)) +
  geom_line(size = 0.5) +   # Line plot for trends over time
  theme_classic() +   # Clean theme
  xlab("") +   # X-axis label
  ylab("Difference from Soil in Pielous's Evenness ") +   # Y-axis label
  labs(color = "diff")   # Legend title for color
```

```
## 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.
```
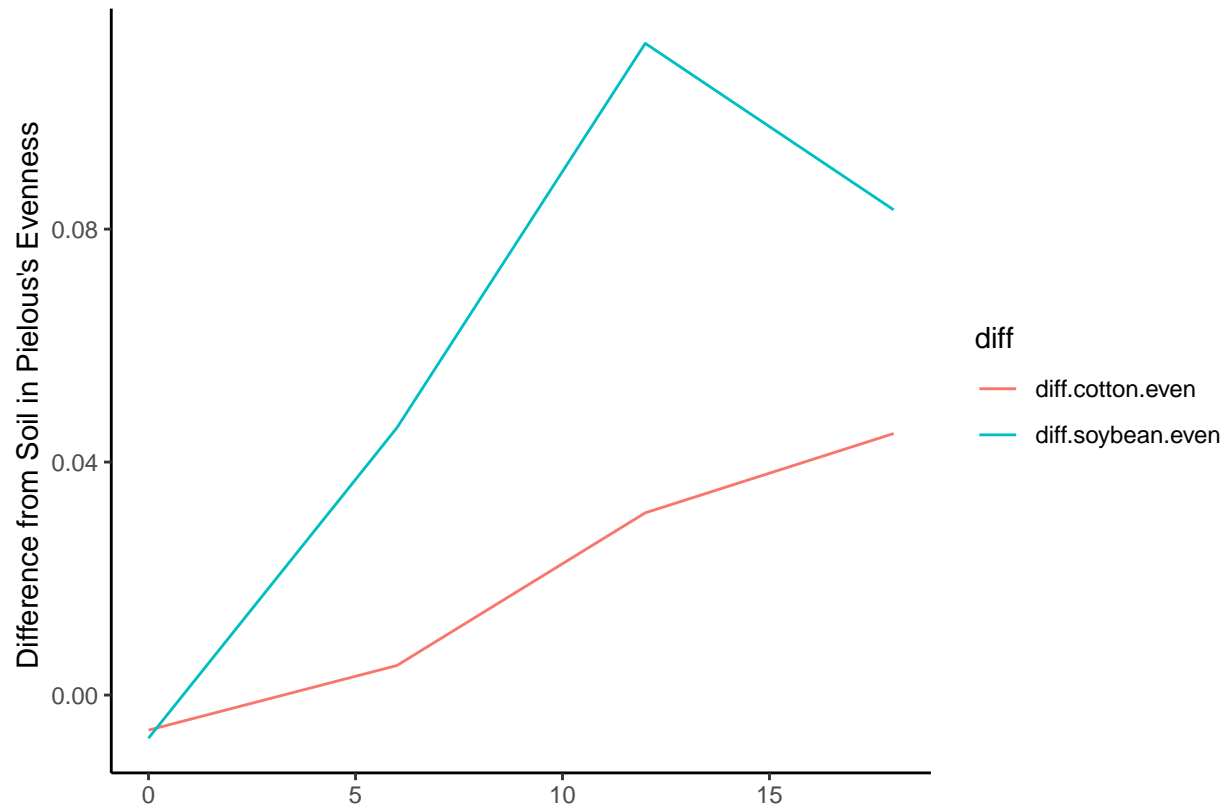
7. 2 pts. Commit and push a gfm .md file to GitHub inside a directory called Coding Challenge

8. Provide me a link to your github written as a clickable link in your .pdf or .docx