

2025_3_27_DataWrangling_CodeChallenge5_mer0127

Madeline Redd

2025 - 03 - 24

Contents

Question 1	1
Question 2	1
Question 3	2
Question 4	2
Question 5	3
Question 6	3
Question 7	4

Question 1

Download two .csv files from Canvas called DiversityData.csv and Metadata.csv, and read them into R using relative file paths.

```
library(tidyverse)
```

```
## -- 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  1.9.4      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
```

```
diversitydata <- read.csv("DiversityData.csv", na.strings= "na")
metadata <- read.csv("Metadata.csv", na.strings= "na")
```

Question 2

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

```
alpha <-left_join(diversitydata, metadata, by= "Code")
head(alpha)
```

```
##      Code shannon invsimpson  simpson richness Crop Time_Point Replicate
## 1 S01_13 6.624921  210.7279 0.9952545   3319 Soil         0          1
## 2 S02_16 6.612413  206.8666 0.9951660   3079 Soil         0          2
## 3 S03_19 6.660853  213.0184 0.9953056   3935 Soil         0          3
## 4 S04_22 6.660671  204.6908 0.9951146   3922 Soil         0          4
## 5 S05_25 6.610965  200.2552 0.9950064   3196 Soil         0          5
## 6 S06_28 6.650812  199.3211 0.9949830   3481 Soil         0          6
## Water_Imbibed
## 1           NA
## 2           NA
## 3           NA
## 4           NA
## 5           NA
## 6           NA
```

Question 3

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.

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

```
alpha_even<-alpha%>%
  mutate(even=shannon/log(richness))
head(alpha_even)
```

```
##      Code shannon invsimpson  simpson richness Crop Time_Point Replicate
## 1 S01_13 6.624921  210.7279 0.9952545   3319 Soil         0          1
## 2 S02_16 6.612413  206.8666 0.9951660   3079 Soil         0          2
## 3 S03_19 6.660853  213.0184 0.9953056   3935 Soil         0          3
## 4 S04_22 6.660671  204.6908 0.9951146   3922 Soil         0          4
## 5 S05_25 6.610965  200.2552 0.9950064   3196 Soil         0          5
## 6 S06_28 6.650812  199.3211 0.9949830   3481 Soil         0          6
## Water_Imbibed even
## 1           NA 0.8171431
## 2           NA 0.8232216
## 3           NA 0.8046776
## 4           NA 0.8049774
## 5           NA 0.8192376
## 6           NA 0.8155427
```

Question 4

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.

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

```
alpha_average<- alpha_even%>%
  group_by(Crop, Time_Point)%>%
  summarise(
    mean.even=mean(even),
    n=n(),
    std.dev=sd(even),
    std.err = (std.dev/sqrt(n))
  )
```

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

Question 5

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

- Start with the `alpha_average` dataframe
- Select relevant columns: select the columns `Time_Point`, `Crop`, and `mean.even`.
- 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`.
- 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.
- Name the resulting dataframe `alpha_average2`

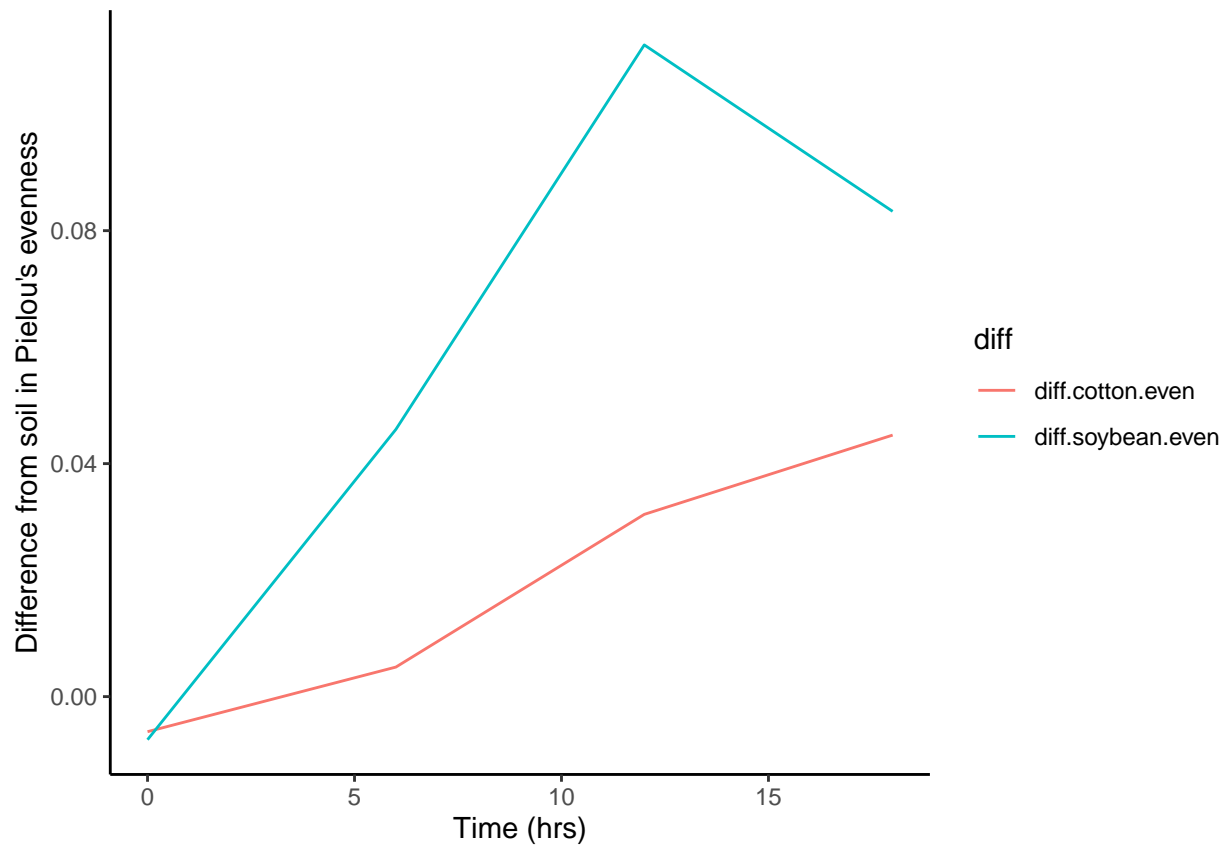
```
alpha_average2<- alpha_average%>%
  select(Time_Point, Crop, mean.even)%>%
  pivot_wider(names_from = Crop, values_from= mean.even)%>%
  mutate(diff.cotton.even = Soil - Cotton)%>%
  mutate(diff.soybean.even = Soil - Soybean)
```

Question 6

Connecting it to plots.

- Start with the `alpha_average2` dataframe
- Select relevant columns: select the columns `Time_Point`, `diff.cotton.even`, and `diff.soybean.even`
- 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`.
- This might be challenging, so I'll give you a break. The code is below
- 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 name '`values`' come from the `pivot_longer`. The resulting plot should look like the one to the right

```
alpha_average2%>% #part 6a
  select(Time_Point,diff.cotton.even, diff.soybean.even)%>% #6b
  pivot_longer(c(diff.cotton.even, diff.soybean.even), names_to = "diff")%>%#6c
  ggplot( aes(x = Time_Point, y = value, color = diff, group = diff))+
  geom_line()+
  xlab("Time (hrs)")+
  ylab ("Difference from soil in Pielou's evenness")+
  theme_classic()+
  theme(strip.background=element_blank())
```



Question 7

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

Coding Challenge Five