Coding challenge 5-Data Wrangling

Bibek

2025-03-20

Contents

#Question 1

#Loading dataset

```
Diversity <- read.csv("Sample_data/DiversityData.csv")
Meta <- read.csv("Sample_data/Metadata.csv")
```

#Question 2 #Loading library #Join them together via common column 'code'.

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.4.2
```

```
## Warning: package 'lubridate' was built under R version 4.4.2
```

```
## -- 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 1.9.4     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
```

```
alpha <- left_join(Diversity, Meta, by = "Code")
```

#Question 3 # Calculate Pielou's evennes index. This equals to the Shannon diversity index (column Shannon) divided by the log of the richness column.

```
aplha.even <- mutate(alpha, Pielou_Eveness_Index= shannon/log(richness)) # creating a new column Pielou
```

#Question 4 #summarise function for Mean and SE of grouped by crop over time.

```
alpha_average <- aplha.even %>%
  group_by(Crop, Time_Point) %>% # grouping by crop and time point
  summarise(Mean.even = mean(Pielou_Eveness_Index), # calculating the mean even, count, stdeviation, an
            count = n(),
            sd.even = sd(Pielou_Eveness_Index),
            std.err = sd.even/sqrt(count))
```

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

```
alpha_average
```

```
## # A tibble: 12 x 6
## # Groups:   Crop [3]
##    Crop    Time_Point Mean.even count sd.even std.err
##    <chr>        <int>     <dbl> <int>   <dbl>   <dbl>
##  1 Cotton           0     0.820     6 0.00556 0.00227
##  2 Cotton           6     0.805     6 0.00920 0.00376
##  3 Cotton          12     0.767     6 0.0157  0.00640
##  4 Cotton          18     0.755     5 0.0169  0.00755
##  5 Soil             0     0.814     6 0.00765 0.00312
##  6 Soil             6     0.810     6 0.00587 0.00240
##  7 Soil            12     0.798     6 0.00782 0.00319
##  8 Soil            18     0.800     5 0.0104  0.00465
##  9 Soybean          0     0.822     6 0.00270 0.00110
## 10 Soybean          6     0.764     6 0.0400  0.0163
## 11 Soybean         12     0.687     6 0.0643  0.0263
## 12 Soybean         18     0.716     6 0.0153  0.00626
```

#**Question 5** # Calulate the differences between soybean column, 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) %>% # pivot to wide format
  mutate(diff.cotton.even = Soil - Cotton,
         diff.soybean.even = Soil - Soybean) ##  calculate the difference between the mean between the
```

#**Question 6** # Connecting plot

```
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()+
  theme_classic()+
  xlab("Time (hrs)") +
  ylab("Difference from soil in Pielou's eveness")
```
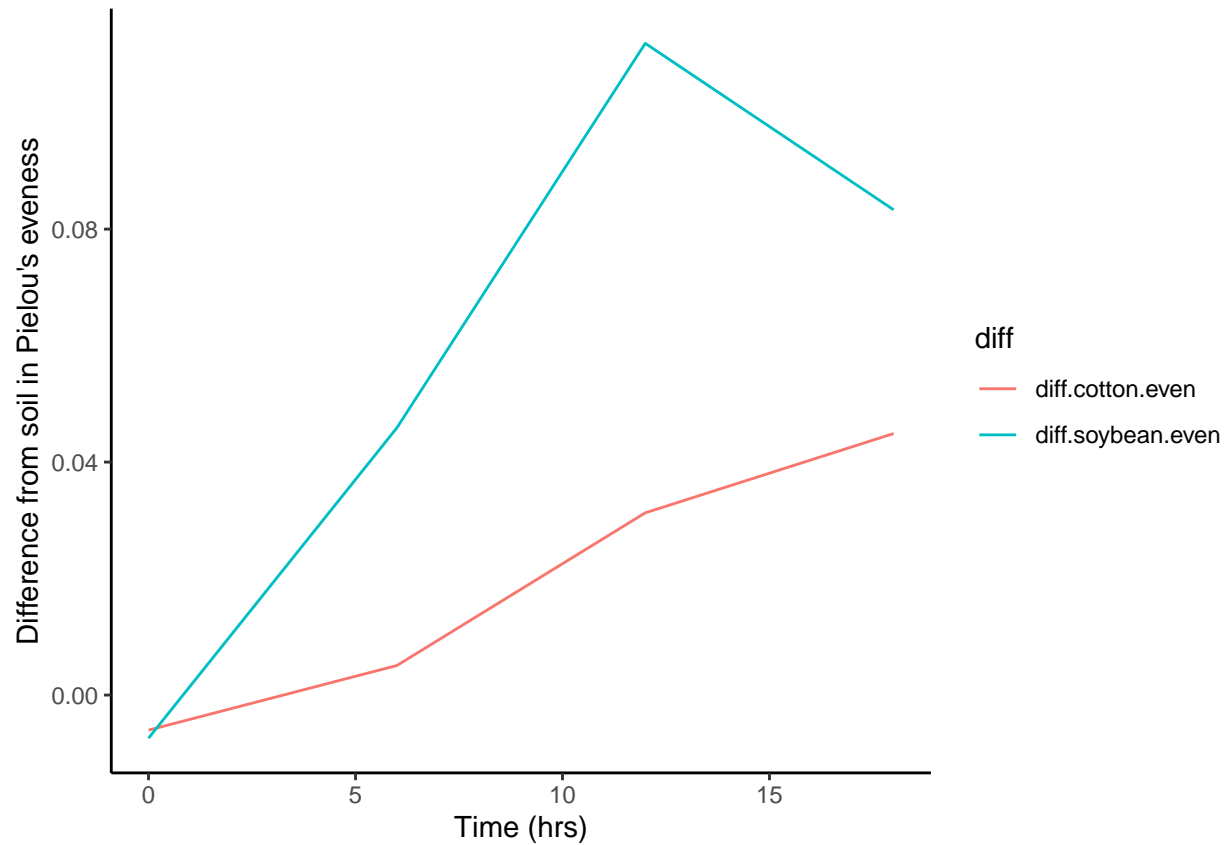
#**Question 7** #Make .md file, push to GitHub isnide directory called Coding challenge 5 and provide link in you submission file. #See Coding challenge 5 folder.

Peasle click this: **Link to my GitHub**