Data Wrangling

Maryam Saeed

2025-03-14

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 1.0.2   
## ── 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

library(readr)  
microbiome.fungi <- read.csv("Bull\_richness.csv")  
str(microbiome.fungi)

## 'data.frame': 287 obs. of 16 variables:  
## $ SampleID : chr "Corn2017LeafObjective2Collection1T1R1CAH2" "Corn2017LeafObjective2Collection1T1R1CBA3" "Corn2017LeafObjective2Collection1T1R1CCB3" "Corn2017LeafObjective2Collection1T1R1FAC3" ...  
## $ Crop : chr "Corn" "Corn" "Corn" "Corn" ...  
## $ Objective : chr "Objective 2" "Objective 2" "Objective 2" "Objective 2" ...  
## $ Collection : int 1 1 1 1 1 1 1 1 1 1 ...  
## $ Compartment : chr "Leaf" "Leaf" "Leaf" "Leaf" ...  
## $ DateSampled : chr "6/26/17" "6/26/17" "6/26/17" "6/26/17" ...  
## $ GrowthStage : chr "V6" "V6" "V6" "V6" ...  
## $ Treatment : chr "Conv." "Conv." "Conv." "Conv." ...  
## $ Rep : chr "R1" "R1" "R1" "R1" ...  
## $ Sample : chr "A" "B" "C" "A" ...  
## $ Fungicide : chr "C" "C" "C" "F" ...  
## $ Target\_organism: chr "Fungi" "Fungi" "Fungi" "Fungi" ...  
## $ Location : chr "Kellogg Biological Station" "Kellogg Biological Station" "Kellogg Biological Station" "Kellogg Biological Station" ...  
## $ Experiment : chr "LTER" "LTER" "LTER" "LTER" ...  
## $ Year : int 2017 2017 2017 2017 2017 2017 2017 2017 2017 2017 ...  
## $ richness : int 9 6 5 7 4 2 3 8 4 4 ...

####‘select()’

microbiome.fungi2<- select(microbiome.fungi, SampleID, Crop, Compartment: Fungicide, richness)

####‘filter()’

head(filter(microbiome.fungi2, Treatment =="Conv."))

## SampleID Crop Compartment DateSampled  
## 1 Corn2017LeafObjective2Collection1T1R1CAH2 Corn Leaf 6/26/17  
## 2 Corn2017LeafObjective2Collection1T1R1CBA3 Corn Leaf 6/26/17  
## 3 Corn2017LeafObjective2Collection1T1R1CCB3 Corn Leaf 6/26/17  
## 4 Corn2017LeafObjective2Collection1T1R1FAC3 Corn Leaf 6/26/17  
## 5 Corn2017LeafObjective2Collection1T1R1FBD3 Corn Leaf 6/26/17  
## 6 Corn2017LeafObjective2Collection1T1R1FCE3 Corn Leaf 6/26/17  
## GrowthStage Treatment Rep Sample Fungicide richness  
## 1 V6 Conv. R1 A C 9  
## 2 V6 Conv. R1 B C 6  
## 3 V6 Conv. R1 C C 5  
## 4 V6 Conv. R1 A F 7  
## 5 V6 Conv. R1 B F 4  
## 6 V6 Conv. R1 C F 2

# A more complex using &  
head(filter(microbiome.fungi2, Treatment == "Conv." & Fungicide == "C"))

## SampleID Crop Compartment DateSampled  
## 1 Corn2017LeafObjective2Collection1T1R1CAH2 Corn Leaf 6/26/17  
## 2 Corn2017LeafObjective2Collection1T1R1CBA3 Corn Leaf 6/26/17  
## 3 Corn2017LeafObjective2Collection1T1R1CCB3 Corn Leaf 6/26/17  
## 4 Corn2017LeafObjective2Collection1T1R2CAF3 Corn Leaf 6/26/17  
## 5 Corn2017LeafObjective2Collection1T1R2CBG3 Corn Leaf 6/26/17  
## 6 Corn2017LeafObjective2Collection1T1R2CCH3 Corn Leaf 6/26/17  
## GrowthStage Treatment Rep Sample Fungicide richness  
## 1 V6 Conv. R1 A C 9  
## 2 V6 Conv. R1 B C 6  
## 3 V6 Conv. R1 C C 5  
## 4 V6 Conv. R2 A C 3  
## 5 V6 Conv. R2 B C 8  
## 6 V6 Conv. R2 C C 4

#Another more complex example using or |  
head(filter(microbiome.fungi2, Sample == "A" | Sample == "B")) # Samples A or B

## SampleID Crop Compartment DateSampled  
## 1 Corn2017LeafObjective2Collection1T1R1CAH2 Corn Leaf 6/26/17  
## 2 Corn2017LeafObjective2Collection1T1R1CBA3 Corn Leaf 6/26/17  
## 3 Corn2017LeafObjective2Collection1T1R1FAC3 Corn Leaf 6/26/17  
## 4 Corn2017LeafObjective2Collection1T1R1FBD3 Corn Leaf 6/26/17  
## 5 Corn2017LeafObjective2Collection1T1R2CAF3 Corn Leaf 6/26/17  
## 6 Corn2017LeafObjective2Collection1T1R2CBG3 Corn Leaf 6/26/17  
## GrowthStage Treatment Rep Sample Fungicide richness  
## 1 V6 Conv. R1 A C 9  
## 2 V6 Conv. R1 B C 6  
## 3 V6 Conv. R1 A F 7  
## 4 V6 Conv. R1 B F 4  
## 5 V6 Conv. R2 A C 3  
## 6 V6 Conv. R2 B C 8

#### ‘mutate()’

Mutate allow us to quickly create new columns

microbiome.fungi2$logRich<- log(microbiome.fungi2$richness)  
#Create a new column called logRich  
head(mutate(microbiome.fungi2, logRich = log(richness)))

## SampleID Crop Compartment DateSampled  
## 1 Corn2017LeafObjective2Collection1T1R1CAH2 Corn Leaf 6/26/17  
## 2 Corn2017LeafObjective2Collection1T1R1CBA3 Corn Leaf 6/26/17  
## 3 Corn2017LeafObjective2Collection1T1R1CCB3 Corn Leaf 6/26/17  
## 4 Corn2017LeafObjective2Collection1T1R1FAC3 Corn Leaf 6/26/17  
## 5 Corn2017LeafObjective2Collection1T1R1FBD3 Corn Leaf 6/26/17  
## 6 Corn2017LeafObjective2Collection1T1R1FCE3 Corn Leaf 6/26/17  
## GrowthStage Treatment Rep Sample Fungicide richness logRich  
## 1 V6 Conv. R1 A C 9 2.1972246  
## 2 V6 Conv. R1 B C 6 1.7917595  
## 3 V6 Conv. R1 C C 5 1.6094379  
## 4 V6 Conv. R1 A F 7 1.9459101  
## 5 V6 Conv. R1 B F 4 1.3862944  
## 6 V6 Conv. R1 C F 2 0.6931472

#Create a new Column which combine crop and treatment  
head(mutate(microbiome.fungi2, Crop\_Treatment = paste(Crop, Treatment)))

## SampleID Crop Compartment DateSampled  
## 1 Corn2017LeafObjective2Collection1T1R1CAH2 Corn Leaf 6/26/17  
## 2 Corn2017LeafObjective2Collection1T1R1CBA3 Corn Leaf 6/26/17  
## 3 Corn2017LeafObjective2Collection1T1R1CCB3 Corn Leaf 6/26/17  
## 4 Corn2017LeafObjective2Collection1T1R1FAC3 Corn Leaf 6/26/17  
## 5 Corn2017LeafObjective2Collection1T1R1FBD3 Corn Leaf 6/26/17  
## 6 Corn2017LeafObjective2Collection1T1R1FCE3 Corn Leaf 6/26/17  
## GrowthStage Treatment Rep Sample Fungicide richness logRich Crop\_Treatment  
## 1 V6 Conv. R1 A C 9 2.1972246 Corn Conv.  
## 2 V6 Conv. R1 B C 6 1.7917595 Corn Conv.  
## 3 V6 Conv. R1 C C 5 1.6094379 Corn Conv.  
## 4 V6 Conv. R1 A F 7 1.9459101 Corn Conv.  
## 5 V6 Conv. R1 B F 4 1.3862944 Corn Conv.  
## 6 V6 Conv. R1 C F 2 0.6931472 Corn Conv.

#### ‘pipe()’

we will combine all the previous steps into one large string of functions. The data from the previous step is transfered to the next step.

microbiome.fungi %>%  
 select(SampleID, Crop, Compartment: Fungicide, richness) %>% #selecting columns  
 filter(Treatment =="Conv.")%>% #subsetting to only include the conventional treatment  
 mutate(logRich = log(richness)) %>% #creating a new column of the log richness  
 head() #displaying the first six rows

## SampleID Crop Compartment DateSampled  
## 1 Corn2017LeafObjective2Collection1T1R1CAH2 Corn Leaf 6/26/17  
## 2 Corn2017LeafObjective2Collection1T1R1CBA3 Corn Leaf 6/26/17  
## 3 Corn2017LeafObjective2Collection1T1R1CCB3 Corn Leaf 6/26/17  
## 4 Corn2017LeafObjective2Collection1T1R1FAC3 Corn Leaf 6/26/17  
## 5 Corn2017LeafObjective2Collection1T1R1FBD3 Corn Leaf 6/26/17  
## 6 Corn2017LeafObjective2Collection1T1R1FCE3 Corn Leaf 6/26/17  
## GrowthStage Treatment Rep Sample Fungicide richness logRich  
## 1 V6 Conv. R1 A C 9 2.1972246  
## 2 V6 Conv. R1 B C 6 1.7917595  
## 3 V6 Conv. R1 C C 5 1.6094379  
## 4 V6 Conv. R1 A F 7 1.9459101  
## 5 V6 Conv. R1 B F 4 1.3862944  
## 6 V6 Conv. R1 C F 2 0.6931472