

R-reproducibility Assignment

Project Info

January 26, 2022

Diego Sancho Kruppa 20101377

DSanchoKruppa on GitHub

<https://github.com/DSanchoKruppa/R-reproducibilityAssignment>

This project demonstrates some basic data wrangling techniques using dplyr by manipulating 'Fallopia-Data.csv' and writing the data to a new file. Additionally, a custom function is defined and used to calculate descriptive statistics of the dataset. Finally, the entire project is linked to a public Git repository **here** to demonstrate version control using Git.

Set up

```
library(tidyverse)
library(ggplot2)
fDat <- read.csv("./InputData/FallopiaData.csv")
```

Trimming data

```
fDat <- fDat %>%
  filter(Total > 60) %>% # removing entries with total < 60
  select(Total, Taxon, Scenario, Nutrients) %>% # reordering columns
  mutate(TotalG = Total * 0.001) %>% # converting Total from mg to g
  select(TotalG, Taxon, Scenario, Nutrients) # replacing Total with TotalG
```

Defining a custom function

```
myFunc <- function(x, op) {
  if (op == "Average") {
    return(mean(x))
  }
  if (op == "Sum") {
    return(sum(x))
  }
  if (op == "Observations") {
```

```

    return(length(x))
  }
  else {
    message("Unknown string; try 'Average', 'Sum' or 'Observations'")
    stop()
  }
}

```

Using the custom function

```

# Total number of observations in the 'Taxon' column
myFunc(fDat$Taxon, "Observations")

```

```
## [1] 45
```

```

# Average TotalG for 'low' nutrient
lowDat <- fDat %>%
  filter(Nutrients == "low")
myFunc(lowDat$TotalG, "Average")

```

```
## [1] 0.06407
```

```

# Average TotalG for 'high' nutrient
highDat <- fDat %>%
  filter(Nutrients == "high")
myFunc(highDat$TotalG, "Average")

```

```
## [1] 0.06646674
```

Writing the wrangled data to a new .csv

```

write.csv(fDat, "./Output/WrangledData.csv")

```