# STA 445 HW3

Paige Hawkinson

03-05-2024

```
library(tidyverse)
library(readxl)
```

## Problem 1

Download from GitHub the data file Example_5.xls. Open it in Excel and figure out which sheet of data we should import into R. At the same time figure out how many initial rows need to be skipped. Import the data set into a data frame and show the structure of the imported data using the `str()` command. Make sure that your data has $n = 31$ observations and the three columns are appropriately named. If you make any modifications to the data file, comment on those modifications.

```
Example.5 <- read_excel("Example_5_HW_3.xlsx", sheet = 'RawData', range = 'A5:C36')
str(Example.5)
```

```
## tibble [31 x 3] (S3: tbl_df/tbl/data.frame)
##  $ Girth : num [1:31] 8.3 8.6 8.8 10.5 10.7 10.8 11 11 11.1 11.2 ...
##  $ Height: num [1:31] 70 65 63 72 81 83 66 75 80 75 ...
##  $ Volume: num [1:31] 10.3 10.3 10.2 16.4 18.8 19.7 15.6 18.2 22.6 19.9 ...
```

I had to include the range of the cells in order to avoid cells that had writing that I did not want to include.

## Problem 2

Download from GitHub the data file Example_3.xls. Import the data set into a data frame and show the structure of the imported data using the `tail()` command which shows the last few rows of a data table. Make sure the Tesla values are `NA` where appropriate and that both `-9999` and `NA` are imported as NA values. If you make any modifications to the data file, comment on those modifications.

```
Example.3 <- read_excel("Example_3.xlsx", sheet = 'data', range = 'A1:L34')
tail(Example.3)
```

```
## # A tibble: 6 x 12
##   model        mpg cyl   disp     hp    drat    wt  qsec vs       am  gear carb
##   <chr>      <dbl> <chr> <chr> <dbl>   <dbl> <dbl> <dbl> <chr> <dbl> <dbl> <chr>
## 1 Lotus Eur~  30.4 4     95.1    113  3.77e0  1.51  16.9 1         1     5 2
## 2 Ford Pant~  15.8 8     351     264  4.22e0  3.17  14.5 0         1     5 4
## 3 Ferrari D~  19.7 6     145     175  3.62e0  2.77  15.5 0         1     5 6
## 4 Maserati ~  15   8     301     335  3.54e0  3.57  14.6 0         1     5 8
## 5 Volvo 142E  21.4 4     121     109  4.11e0  2.78  18.6 1         1     4 2
## 6 Tesla Mod~  98   NA    NA      778 -1.00e4  4.94  10.4 NA        0     1 NA
```

## Problem 3

Download all of the files from GitHub `data-raw/InsectSurveys` directory here. Each month's file contains a sheet contains site level information about each of the sites that was surveyed. The second sheet contains information about the number of each species that was observed at each site. Import the data for each month and create a single `site` data frame with information from each month. Do the same for the `observations`. Document any modifications you make to the data files. Comment on the importance of consistency of your data input sheets.

```
May <- read_excel("May.xlsx", sheet = 'Sites', range = 'A1:F10')
June <- read_excel("June.xlsx", sheet = 'Sites', range = 'A1:F10')
July <-read_excel("July.xlsx", sheet = 'Sites', range = 'A1:F10')
August <- read_excel("August.xlsx", sheet = 'Sites', range = 'A1:F10')
September <- read_excel("September.xlsx", sheet = 'Sites', range = 'A1:F10')
October <- read_excel("October.xlsx", sheet = 'Sites', range = 'A1:F10')
```

```
MayJune <- rbind(May, June)
JulyAugust <- rbind(July, August)
SeptOct <- rbind(September, October)
MayJuneAugJuly <- rbind(MayJune, JulyAugust)
SixMonths.sites <- rbind(MayJuneAugJuly, SeptOct)
str(SixMonths.sites)
```

```
## tibble [54 x 6] (S3: tbl_df/tbl/data.frame)
##  $ Site Name  : chr [1:54] "Araphahoe Road" "Bridger Valley" "Calculus Vector" "Deer Valley" ...
##  $ Pond Area  : num [1:54] 34 240 321 74 28 62 489 398 126 34 ...
##  $ Water Depth: num [1:54] 3 6 13 4.4 2 3.6 4 10 9 3 ...
##  $ ph         : num [1:54] 6.2 6.5 6.4 6.9 7.1 7 7.1 6.8 6.75 6.2 ...
##  $ Date       : POSIXct[1:54], format: "2020-05-15" "2020-05-16" ...
##  $ Observer   : chr [1:54] "Bob" "Bob" "Bob" "Bob" ...
```

```
May.Obs <- read_excel("May.xlsx", sheet = 'Observations', range = 'A1:C37')
June.Obs <- read_excel("June.xlsx", sheet = 'Observations', range = 'A1:C37')
July.Obs <-read_excel("July.xlsx", sheet = 'Observations', range = 'A1:C37')
August.Obs <- read_excel("August.xlsx", sheet = 'Observations', range = 'A1:C37')
September.Obs <- read_excel("September.xlsx", sheet = 'Observations', range = 'A1:C37')
October.Obs <- read_excel("October.xlsx", sheet = 'Observations', range = 'A1:C37')
```

```
MayJune.Obs <- rbind(May.Obs, June.Obs)
JulyAugust.Obs <- rbind(July.Obs, August.Obs)
SeptOct.Obs <- rbind(September.Obs, October.Obs)
MayJuneAugJuly.Obs <- rbind(MayJune.Obs, JulyAugust.Obs)
SixMonths.Obs <- rbind(MayJuneAugJuly.Obs, SeptOct.Obs)
str(SixMonths.Obs)
```

```
## tibble [216 x 3] (S3: tbl_df/tbl/data.frame)
##  $ Site   : chr [1:216] "Araphahoe Road" NA NA NA ...
##  $ Species: chr [1:216] "Caddis Fly" "May Fly" "Stone Fly" "Dragon Fly" ...
##  $ Count  : num [1:216] 2 4 8 7 2 4 8 7 2 4 ...
```

The consistency of data import sheets is extremely important. There were issues binding sheets together due to capitalization and order. Small details such as capitalization should be kept the same when recording data as this will prevent a huge headache when trying to import and bind in R.