

# STA 445 HW3

Nicholas Larson

03/05/2024

```
library(tidyverse)
library(readr)
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.

```
Ex5 <- read_excel("Z:/STAT445/Example_5.xls", sheet='RawData', range='A5:C36')
str(Ex5)
```

```
## 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 ...
```

## 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.

```
Ex3 <- read_excel("Z:/STAT445/Example_3.xls", sheet='data', range='A1:l34', na = c('-9999', 'NA'))
tail(Ex3)
```

```
## # A tibble: 6 x 12
##   model      mpg   cyl  disp    hp  drat    wt   qsec    vs  am  gear  carb
##   <chr>    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 Lotus Europa  30.4     4  95.1   113   3.77  1.51  16.9     1     1     5     2
## 2 Ford Panter~  15.8     8  351    264   4.22  3.17  14.5     0     1     5     4
## 3 Ferrari Dino  19.7     6  145    175   3.62  2.77  15.5     0     1     5     6
## 4 Maserati Bo~  15       8  301    335   3.54  3.57  14.6     0     1     5     8
## 5 Volvo 142E   21.4     4  121    109   4.11  2.78  18.6     1     1     4     2
## 6 Tesla Model~  98      NA   NA     778   NA     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_ob <- read_excel("Z:/STAT445/May.xlsx", sheet='Observations', range='A1:C37')
June_ob <- read_excel("Z:/STAT445/June.xlsx", sheet='Observations', range='A1:C37')
July_ob <- read_excel("Z:/STAT445/July.xlsx", sheet='Observations', range='A1:C37')
August_ob <- read_excel("Z:/STAT445/August.xlsx", sheet='Observations', range='A1:C37')
September_ob <- read_excel("Z:/STAT445/September.xlsx", sheet='Observations', range='A1:C37')
October_ob <- read_excel("Z:/STAT445/October.xlsx", sheet='Observations', range='A1:C37')
```

```
May_S <- read_xlsx("Z:/STAT445/May.xlsx", sheet='Sites', range='A1:F10')
June_S <- read_xlsx("Z:/STAT445/June.xlsx", sheet='Sites', range='A1:F10')
July_S <- read_xlsx("Z:/STAT445/July.xlsx", sheet='Sites', range='A1:F10')
August_S <- read_xlsx("Z:/STAT445/August.xlsx", sheet='Sites', range='A1:F10')
September_S <- read_xlsx("Z:/STAT445/September.xlsx", sheet='Sites', range='A1:F10')
October_S <- read_xlsx("Z:/STAT445/October.xlsx", sheet='Sites', range='A1:F10')
```

```
data_ob <- rbind(May_ob, June_ob, July_ob, August_ob, September_ob, October_ob)
head(data_ob)
```

```
## # A tibble: 6 x 3
##   Site           Species    Count
##   <chr>         <chr>    <dbl>
## 1 Araphahoe Road Caddis Fly     2
## 2 <NA>           May Fly       4
## 3 <NA>           Stone Fly     8
## 4 <NA>           Dragon Fly    7
## 5 Bridger Valley Caddis Fly     2
## 6 <NA>           May Fly       4
```

```
data_S <- rbind(May_S, June_S, July_S, August_S, September_S, October_S)
head(data_S)
```

```
## # A tibble: 6 x 6
##   Site           `Pond Area` `Water Depth`    ph Date              Observer
##   <chr>         <dbl>    <dbl> <dbl> <dtm>              <chr>
## 1 Araphahoe Road      34         3    6.2 2020-05-15 00:00:00 Bob
## 2 Bridger Valley     240         6    6.5 2020-05-16 00:00:00 Bob
## 3 Calculus Vector    321        13    6.4 2020-05-17 00:00:00 Bob
## 4 Deer Valley        74         4.4  6.9 2020-05-18 00:00:00 Bob
## 5 Ephemeral Stream   28          2    7.1 2020-05-15 00:00:00 Charlie
## 6 Fennel Gardens     62         3.6    7   2020-05-16 00:00:00 Charlie
```