STA 445 HW3

Ryan Joel

March 5, 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.

```
tree <- read_excel('Example_5.xls', sheet='RawData', range='A5:C36')
str(tree)
```

```
## 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.

```
car.ex <- read_excel("Example_3.xls", sheet='data', range='A1:L34', na= c('NA', '-9999'))
tail(car.ex)
```

```
## # 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.

Sites

```
site <-
October.new <- read_excel("October.xlsx", sheet= 1, range = 'A1:F10')
September.new <- read_excel("September.xlsx", sheet= 1, range = 'A1:F10')
August.new <- read_excel("August.xlsx", sheet= 1, range='A1:F10')
```

```
## New names:
## * `` -> `...4`
## * `` -> `...5`
## * `` -> `...6`
```

```
July.new <- read_excel("July.xlsx", sheet=1, range = 'A1:F10')
June.new <- read_excel("June.xlsx", sheet=1, range = 'A1:F10')
May.new <- read_excel("May.xlsx", sheet=1, range = 'A1:F10')
#For this problem I had to change the date format for the excels for each month.
#I made sure the order of the sheets was consistent in each one to conveniently type 'sheet=1'
#in the code.
#I fixed inconsistent capitalization in the column names.
#I deleted the "did not observe" from the October dataset, so that R could read the files in correctly.
#Overall, it was important to keep all sheets in cosistent order and format so that R could correctly
#synthesize the existing excels into a new dataframe, like we did in this exercise.
site
```

```
## # A tibble: 9 x 6
##   `Site Name`      `Pond Area` `Water Depth`    ph Date         Observer
##   <chr>                  <dbl>         <dbl> <dbl> <chr>        <chr>
## 1 Araphahoe Road            34             3   6.2  44119        Bob
## 2 Bridger Valley           240             6   6.5  44120        Bob
## 3 Calculus Vector          321            13   6.4  44121        Bob
## 4 Deer Valley               74             4.4 6.9  44122        Bob
## 5 Ephemeral Stream          28             2   7.1  44119        Charlie
## 6 Fennel Gardens            62             3.6 7    Did not vist Charlie
## 7 Gigantic Pain            489             4   7.1  44121        Charlie
## 8 Happy Feet               398            10   6.8  44122        Charlie
## 9 Indigo Flats             126             9   6.75 44123        Charlie
```

Observations

```
observations <-
October.new2 <- read_excel("October.xlsx", sheet= 2, range = 'A1:C37')
September.new2 <- read_excel("September.xlsx", sheet= 2, range = 'A1:C37')
August.new2 <- read_excel("August.xlsx", sheet= 2, range='A1:C37')
July.new2 <- read_excel("July.xlsx", sheet=2, range = 'A1:C37')
June.new2 <- read_excel("June.xlsx", sheet=2, range = 'A1:C37')
May.new2 <- read_excel("May.xlsx", sheet=2, range = 'A1:C37')
#I made sure the order of the sheets was consistent in each one to conveniently type 'sheet=2'
#in the code.
#I fixed inconsistent capitalization in the column names, which only occured once in the July dataset
#in the 'Species' column.
#All items/ values that were N/A were already had blank cells, so I did not need to edit that this time
#Overall, it was important to keep all sheets in cosistent order and format so that R could correctly
```

```
#synthesize the existing excels into a new dataframe, like we did in this exercise.
observations
```

```
## # A tibble: 36 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
##  7 <NA>            Stone Fly      8
##  8 <NA>            Dragon Fly     7
##  9 Calculus Vector Caddis Fly     2
## 10 <NA>            May Fly        4
## # i 26 more rows
```