

STA 445 HW3

Connor Phillips

2/29/2024

```
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.

```
cherryTrees = data.frame(read_excel("Example_5.xls", sheet='RawData', range='A5:C36'))
str(cherryTrees)
```

```
## 'data.frame':    31 obs. of  3 variables:
## $ Girth : num  8.3 8.6 8.8 10.5 10.7 10.8 11 11 11.1 11.2 ...
## $ Height: num  70 65 63 72 81 83 66 75 80 75 ...
## $ Volume: num  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.

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

```
##           model  mpg cyl  disp  hp drat   wt  qsec vs am gear carb
## 28      Lotus Europa 30.4   4  95.1 113 3.77 1.513 16.90  1  1    5    2
## 29   Ford Pantera L 15.8   8 351.0 264 4.22 3.170 14.50  0  1    5    4
## 30   Ferrari Dino 19.7   6 145.0 175 3.62 2.770 15.50  0  1    5    6
## 31  Maserati Bora 15.0   8 301.0 335 3.54 3.570 14.60  0  1    5    8
## 32      Volvo 142E 21.4   4 121.0 109 4.11 2.780 18.60  1  1    4    2
## 33 Tesla ModelS P100D 98.0  NA    NA  778   NA 4.941 10.41 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. //Modifications:// Sites: -June - Capitalized 'Site Name', 'Pond Area', 'Water Depth' -Formatted all dates using a formula in excel, then converted to values Observations: -Filled missing site name fields for all months -July - Capitalized 'Species' Sheets: -Capitalized sheet names for August and September

It's incredibly important that data is in unambiguous formats, columns, and sheets match up exactly and are spelled the same, otherwise the sheets won't bind correctly.

```
maySites = read_excel("May.xlsx", sheet='Sites', range='A1:F10')
juneSites = read_excel("June.xlsx", sheet='Sites', range='A1:F10')
julSites = read_excel("July.xlsx", sheet='Sites', range='A1:F10')
augSites = read_excel("August.xlsx", sheet='Sites', range='A1:F10')
septSites = read_excel("September.xlsx", sheet='Sites', range='A1:F10')
octSites = read_excel("October.xlsx", sheet='Sites', range='A1:F10')

seasonSites = data.frame(rbind(maySites, juneSites, julSites, augSites, septSites, octSites))
head(seasonSites)
```

```
##           Site.Name Pond.Area Water.Depth  ph Observer      Date
## 1  Araphahoe Road      34         3.0 6.2      Bob May 15, 2020
## 2  Bridger Valley     240         6.0 6.5      Bob May 16, 2020
## 3  Calculus Vector    321        13.0 6.4      Bob May 17, 2020
## 4   Deer Valley      74         4.4 6.9      Bob May 18, 2020
## 5 Ephemeral Stream    28         2.0 7.1  Charlie May 15, 2020
## 6  Fennel Gardens     62         3.6 7.0  Charlie May 16, 2020
```

```
mayObserv = read_excel("May.xlsx", sheet='Observations', range='A1:C37')
juneObserv = read_excel("June.xlsx", sheet='Observations', range='A1:C37')
julObserv = read_excel("July.xlsx", sheet='Observations', range='A1:C37')
augObserv = read_excel("August.xlsx", sheet='Observations', range='A1:C37')
septObserv = read_excel("September.xlsx", sheet='Observations', range='A1:C37')
octObserv = read_excel("October.xlsx", sheet='Observations', range='A1:C37')

seasonObserv = data.frame(rbind(mayObserv, juneObserv, julObserv, augObserv, septObserv, octObserv))
head(seasonObserv)
```

```
##           Site      Species Count
## 1 Araphahoe Road Caddis Fly     2
## 2 Araphahoe Road  May Fly     4
## 3 Araphahoe Road  Stone Fly    8
## 4 Araphahoe Road Dragon Fly    7
## 5 Bridger Valley Caddis Fly     2
## 6 Bridger Valley  May Fly     4
```