M3 Problem Set

2024-02-05

Loading libraries

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.3.2

## Warning: package 'ggplot2' was built under R version 4.3.2

## Warning: package 'tibble' was built under R version 4.3.1

## Warning: package 'tidyr' was built under R version 4.3.1

## Warning: package 'readr' was built under R version 4.3.2

## Warning: package 'purrr' was built under R version 4.3.1

## Warning: package 'dplyr' was built under R version 4.3.1

## Warning: package 'stringr' was built under R version 4.3.1

## Warning: package 'forcats' was built under R version 4.3.2

## Warning: package 'lubridate' was built under R version 4.3.2

## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(palmerpenguins)
```

```
## Warning: package 'palmerpenguins' was built under R version 4.3.2
```

```
library(ggthemes)
```

```
## Warning: package 'ggthemes' was built under R version 4.3.2
```

Loading

```
data(trees)
data(palmerpenguins)
```

```
## Warning in data(palmerpenguins): data set 'palmerpenguins' not found
```

Print the first 6 rows

```
head(trees)
```

```
##   Girth Height Volume
## 1   8.3     70   10.3
## 2   8.6     65   10.3
## 3   8.8     63   10.2
## 4  10.5     72   16.4
## 5  10.7     81   18.8
## 6  10.8     83   19.7
```

```
head(penguins)
```

```
## # A tibble: 6 x 8
##   species island    bill_length_mm bill_depth_mm flipper_length_mm body_mass_g
##   <fct>   <fct>              <dbl>         <dbl>             <int>       <int>
## 1 Adelie  Torgersen           39.1          18.7               181        3750
## 2 Adelie  Torgersen           39.5          17.4               186        3800
## 3 Adelie  Torgersen           40.3          18                 195        3250
## 4 Adelie  Torgersen           NA            NA                  NA          NA
## 5 Adelie  Torgersen           36.7          19.3               193        3450
## 6 Adelie  Torgersen           39.3          20.6               190        3650
## # i 2 more variables: sex <fct>, year <int>
```

1. Using the ***nrow()*** command, how many rows does the data.frame have?
2. Using the ***ncol()*** command, how many columns are there in the trees dataset?
3. What information does the ***str(trees)*** command provide about the structure of the trees data frame in R?
4. Data Visualization ***link to section here***. Go through section 1.2.
   This section involves walking through building a ggplot from the ground up, using a dataset which is built into R already. There are a number of datasets which come with R, which you can access through the datasets library. (Try running library(help = "datasets")in R to see a list). Make sure to do this section step by step to avoid encountering errors! (you still may encounter errors, which again, is normal and is a prompt for you to understand what is happening).
   Your submission for this question should be the graph of the penguin data.

a. Which of our fish species seems to have a higher length for a given weight compared to the other fish? How do you know? Do a google image search for the species to see if your answer makes sense!