

M3 Problem Set

2024-02-05

Loading libraries

```
library(tidyverse)
library(palmerpenguins)
library(ggthemes)
```

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>
```

Questions

1. Using the `nrow()` command, how many rows does the data.frame *trees* have?

```
nrow(trees)
```

```
## [1] 31
```

2. Using the `ncol()` command, how many columns are there in the *trees* dataset?

```
ncol(trees)
```

```
## [1] 3
```

3. What information does the `length(trees)` command provide about the structure of the *trees* data frame in R?

```
length(trees)
```

```
## [1] 3
```

4. Create a vector with elements *1 2 3 4 5 6 7 8 9* and call it *x*. create another vector with elements *10 20 30 40 50* and call it *y*. What happens if you try to add *x* and *y* together? why?

```
x <- c(1, 2, 3, 4, 5, 6, 7, 8, 9)
y <- c(10, 20, 30, 40, 50, 60, 70, 80)
x+y
```

```
## Warning in x + y: longer object length is not a multiple of shorter object
## length
```

```
## [1] 11 22 33 44 55 66 77 88 19
```

5. 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!