

# M3 Problem Set

2024-02-05

## Loading libraries

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

## Loading

```
data(trees)
data(penguins)
```

## 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? (10pts)

```
nrow(trees)
```

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

2. Using the `ncol()` command, how many columns are there in the **trees** dataset? (10pts)

```
ncol(trees)
```

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

3. What information does the `length(trees)` command provide about the structure of the trees data frame in R? (10pts)

```
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? (20pts)

```
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
```

When attempting to add **x** and **y** together using the **+** operator like **x + y**, you will encounter an error. This is because the length of **x** is not a multiple of the length of **y**, and R cannot perform arithmetic operations on vectors of different lengths without recycling elements or other operations like broadcasting.

5. Data Visualization ***link to section here***. Navigate through section 1.2. This section guides you through constructing a ggplot from scratch, utilizing a dataset already integrated into R. Numerous datasets are bundled with R, accessible through the datasets library. (To view the list, execute `library(help = "datasets")` in R). Ensure to proceed through this section meticulously to avoid potential errors, which are expected and serve as prompts for comprehension.

Your submission for this question should consist of the graph displaying the penguin data.

- a. Among our penguin species, which appears to possess a greater length relative to weight compared to the other fish? How did you ascertain this? Conduct a Google image search for the species to validate if your conclusion aligns with reality. (50pts)

```
ggplot(
  data = penguins,
  mapping = aes(x = flipper_length_mm, y = body_mass_g, color = species)) +
  geom_point()
```

## Warning: Removed 2 rows containing missing values (`geom\_point()`).

