STAC32 Assignment 1

Packages

```
library(tidyverse)
```

You are expected to complete this assignment on your own: that is, you may discuss general ideas with others, but the writeup of the work must be entirely your own. If your assignment is unreasonably similar to that of another student, you can expect to be asked to explain yourself.

If you run into problems on this assignment, it is up to you to figure out what to do. The only exception is if it is *impossible* for you to complete this assignment, for example a data file cannot be read. (There is a difference between you not knowing how to do something, which you have to figure out, and something being impossible, which you are allowed to contact me about.)

You must hand in a rendered document that shows your code, the output that the code produces, and your answers to the questions. This should be a file with `.html` on the end of its name. There is no credit for handing in your unrendered document (ending in `.qmd`), because the grader cannot then see whether the code in it runs properly. After you have handed in your file, you should be able to see (in Attempts) what file you handed in, and you should make a habit of checking that you did indeed hand in what you intended to, and that it displays as you expect.

Hint: render your document frequently, and solve any problems as they come up, rather than trying to do so at the end (when you may be close to the due date). If your document will not successfully render, it is because of an error in your code that you will have to find and fix. The error message will tell you *where* the problem is, but it is up to you to sort out *what* the problem is.

Reading in data

The questions below ask you to read in and display the data described:

1. (3 points) The data in http://ritsokiguess.site/datafiles/bellpepper.csv come from a study of soil water content of two fields A and B where bell peppers are grown. Read the data into a dataframe and display at least some of it.

2. (3 points) The file http://ritsokiguess.site/datafiles/d1.txt contains values for two quantitative variables `x` and `y`. Read the data into a dataframe `d1`, and display the dataframe. (Hint: look at the file with your web browser first.)

3. (3 points) The file http://ritsokiguess.site/datafiles/d2.txt contains values for a categorical variable `group` and a quantitative variable `y`. Read the data into a dataframe `d2`, and display the dataframe.

Island birds

The data in http://ritsokiguess.site/datafiles/Sleuth2_ex1027.csv are information about 62 species of European birds. For each of 16 islands over many years, observers noted whether what species were present on that island, and for each species present, how many breeding pairs of that species there were. The data in that file are averages over the sixteen islands for that species. In addition, a "time to extinction" was also calculated for each species, based on the length of time the species had been seen on each island. A larger time to extinction is better, and it is believed that a species with more breeding pairs will have a larger time to extinction.

As well as the columns `Pairs` and `Time` described above, the dataset contains these columns:

- `Species` : the name of the bird species
- the `Size` of birds in that species: large (`L`) or small (`S`)
- the `Status` of the bird species: resident (`R`), or migratory (`M`). Resident birds live on the same island all year, while migratory birds spend part of the year in some other part of the world.

4. (1 point) Read in and display (some of) the data.

5. (1 point) Below, you will be drawing some graphs. Why is it that the column `Species` does not feature in any of the graphs?

For each of the questions below, draw an appropriate graph of (only) the variables named, explaining (very) briefly why it is appropriate, and give a one-sentence interpretation of the most important feature of the graph, in the context of the data. Two points for the graph in each case (justified choice of graph and drawing the graph), and one for interpretation.

6. (3 points) Status

7. (3 points) Pairs

8. (3 points) Pairs and Size

9. (3 points) Pairs and extinction time

10. (3 points) Size and status

11. (3 points) Extinction time, pairs, and size