Assignment 1: Data Exploration

# Overview

Students will investigate a dataset (or datasets) of their choice using the tools they learned in Chapter 2. Datasets chosen can (and ideally should!) be used for future assignments as well, and should be suitable for asking multiple questions that are biologically interesting. In other words, the datasets chosen should have several variables of interest.

In this assignment, students will need to:

1. Select a dataset, provide a brief description (a few sentences at most), and note where they got it (e.g., provide a citation or a URL).
2. Read in their dataset.
3. Note the class of the dataset itself (e.g., matrix, list, dataframe, etc.)
4. List the variables of interest and their classes (e.g., numeric, character, date, etc.)
5. Examine the structure of the dataset and correct any formatting mistakes.
6. Provide summary statistics for variables of interest.
7. Produce at least 3 different figures that convey different information.
8. For the figures in 6, briefly describe what the figures show (2 sentences or less each).

## Rationale

The goals of this assignment are to provide you with a chance to practice loading new data into R, checking your data for issues, and conducting exploratory data analyses. These skills are critical for any scientist to develop and are necessary when conducting any data-driven analyses. Further, providing documentation on what you did (and perhaps why) is an increasingly important skill as Open and transparent Science becomes the norm.

## Assignment Format

This assignment can be submitted in one of two formats:

1. The output of an .Rmd file
2. See an example .Rmd file at: <https://github.com/bmaitner/Statistical_ecology_course/blob/main/example_R_markdown/Example.Rmd>
3. See example output at:

<https://github.com/bmaitner/Statistical_ecology_course/blob/main/example_R_markdown/Example.pdf>

<https://github.com/bmaitner/Statistical_ecology_course/raw/refs/heads/main/example_R_markdown/Example.docx>

<https://github.com/bmaitner/Statistical_ecology_course/raw/refs/heads/main/example_R_markdown/Example.html>

1. A fully-reproducible R script that can be run on any computer

a. See an example .R file at:

<https://github.com/bmaitner/Statistical_ecology_course/blob/main/example_R_commented/Example.R>

Assignments can be submitted in 2 ways:

1. Provide the file on canvas (e.g., .R or .html; undergrads only)
2. Upload the file to your own Github repository and then provide a link via Canvas (mandatory for graduate students, undergrads get extra credit).

## Grading

Students will be graded on both meeting the criteria outlined in the overview as well as the quality of their work.

*Grade Breakdown:*

* Reproducibility 20%
* Documentation quality 20%
* At least 3 figures included (conveying different information) 20%
* Description of dataset 10%
* Figure explanations 20%
* Summary statistics present, classes identified 10%
* Uploaded to Github (Undergrads) +10%
* Not uploaded to Github (Grads) -10%