Assignment 2.2 – Reading, Writing, and Cleaning Data

Create an R Markdown document that outputs either an HTML or DOC file. Put this file in its own, unique directory.

Download the robo-mussel.csv data file associated with this assignment. This data is from the Helmuth Lab's Robomussel Database, a collection of intertidal temperatures recorded from model mussels, meant to approximate the body temperatures of mussels, a locations along the west coast of North America. Be sure you put this file in the same directory as the Rmd file so that it runs correctly.

a. First, open the CSV file using a spreadsheet program (e.g., Excel).

   a.1. What types of measurements does the file contain? How many columns and rows?

   a.2. What data classes should each column be imported as?

   a.3. Are there any file headers that should be skipped?

b. Using `read.table()`, load the data from this CSV file into R. If there are any lines that should be skipped, include this argument.

   b.1. Did R load the correct number of column and rows based on a.1? Do the column names reflect the measurements you identified in a.1? If not, how can you fix this using the `read.table()` function?

   b.2. Are the data in each column the appropriate data class? If not, how can you fix that using the `read.table()` function?

c. Clean the data up using the tidy pipe operator and however many commands you need to make the data set easy to work with. Things to consider cleaning would be: removing observations `NA` values, standardizing column names, and removing empty rows and columns. Explain each command you are using and why in the Markdown text before or after the cleaning R chunk of code.

d. Make a simple plot of two columns against each other. Be sure to label everything appropriately (make it look nice!)

e. Save the newly cleaned data set in R as a CSV file using `write.csv()`.

Turn in your Rmd file.

Data reference:

Data are from Robomussel Database: https://helmuthlab.cos.northeastern.edu/databases/robomussel/.