Coding Challenge 1: Getting set up with RStudio and the tidyverse

Public Health 460

Due: 6pm, Friday February 10th, 2023

You will be a better coder if you start off with good, organized habits when using Rstudio. Here is the process we ask you to follow to set up your RStudio development environment for this course.

- Make sure you have the latest version of R installed (should be 4.0 or later).
- Make sure you have an updated version of RStudio.
- Open RStudio and create a "new project" in a new directory on your computer in a relevant directory that is named clearly for this class, e.g. `Home/Documents/classes/PH460`.
- You should open and use this project for every assignment this semester.
- Create new subdirectories in the `PH460` directory you just created. They should be called `coding-challenges` and `data`.
- Open a new RMarkdown file and save it in the `coding-challenges` directory you just made.
- If you haven't already, install the `tidyverse` R package (this is a bundle of a lot of different packages).
- Download the `women.csv` from the Moodle assignment and save it in the `data` folder you created above in the project.
- If you are new to RMarkdown, we recommend reviewing this overview on using RMarkdown files in RStudio.

Answer the questions below for credit

By the deadline listed above, you must hand in both the .Rmd file and the resulting, "knitted" HTML file on Moodle. The HTML file must show all code and output (2 pts). Hint: this involves making sure the `echo` option is set to `TRUE` in your RMarkdown file, which it should be by default.

1) Load the `tidyverse` package. (2pts)

2) Read in the `women.csv` file using `read_csv()` with a relative path to identify your dataset. Save the dataset as a named object. This data set gives the average heights and weights of American women aged 30-39. The height is given in inches and the weight is given in pounds. This data set is believed to have came from the Build and Blood Pressure Study by the American Society of Actuaries. (5pts)

3) Using the `dplyr::mutate()` function to add a the following new variables to your dataset: `height_m` which should represent height in meters, `mass_kg` for weight in kilograms, and `bmi` which will be the body mass index (BMI = mass in kg divided by height in meters squared). (6 pts)

4) Using the `dplyr` function `arrange()`, print the data sorted from largest (at the top) to smallest by BMI. (5pts)