Problem Set 1: Introduction to R and RStudio

Due 11:59 PM, Thursday, April 10

The purpose of this problem set is to get you set up for the quarter. You will install R and RStudio, and then you will write an R Markdown file that loads a dataset, summarizes a single variable, and outputs the file that you will hand in.

1. If you do not already have R and RStudio installed on your computer, go to the RStudio website and complete the steps. When you open RStudio for the first time, type `install.packages("tidyverse")` at the command line in the Console window (bottom left).
2. Download the template for Problem Set 1 from the course GitHub repository and open it in RStudio.
3. Modify the name in the header to your own.
4. The template already has code to load the dataset we analyzed in class in Week 1. It has population, births, deaths, and net migration, all measured in 1000s, for every country and year since 1950. Write code to calculate and report the mean, mimimum, and maximum of the population variable.
5. What is the maximum population in the data?
6. Now write additional code that filters the dataset to only 1950 and 2023 and then produces a table with the mean, minimum, and maximum for each year.
7. What was the maximum population in 1950? In 2023?
8. Tell us any accommodations you will need for the midterm or final exam. If you do not tell us now, we may not be able to accommodate you later.
9. Render the R Markdown file by clicking *knit*. The code is currently set up to produce an HTML file, which you should open in a web browser and save to PDF. If you have LaTeX on your computer, you can skip this step by changing `html_document` to `pdf_document` in the header. Then R Markdown will directly produce a PDF file when you click *knit*.
10. Your PDF file should have your name, summary statistics on population, and your interpretation of the summary statistics. If it does, upload your PDF to Gradescope.