SPEC Lab R Resources: Data Management IIA- Homework

Ben Graham

Summer 2020

Manipulating Data using the Tidyverse

Complete the following assignment and then save your R script and merged dataset into your homework submission folder. The R script should be titled HW_DM2_[YOUR INTIALS]

Exercise 1:

Download the following variables for all countries and all years from the World Bank Ed Stats database: Barro and Lee's measures of the number of 15-19 year-olds that have completed secondary school for both the whole population and for females only. https://databank.worldbank.org/reports.aspx?source=Education%20Statistics

A Reshape the data to a tidy dataset with a country-year unit of analysis

B Rename the two education variables with concise, informative variable names and an appropriate suffix

Exercise 2:

Download information on literacy rates and child (under age 5) mortality rates from the World Development Indicators. **A** If these data are not already tidy, make them so.

B Rename the literacy rate and child mortality variables with concise informative variable names and an appropriate suffix.

Exercise 3:

Merge these two datasets together using `full_join()`.

Submit your well-annotated R script and your merged data file via your homework submission folder.

OPTIONAL BONUS FOR SPEC-AFFILIATED FOLKS Complete Module IIA. Then append gwno numbers to both the Barro & Lee and the WDI data, check carefully for duplicate observations, evaluate which observations, if any, did not receive gwno numbers, and then merge these two datasets together on the basis of gwno and year.