

SPEC Lab R Resources: Data Management 2 - Homework

Ben Graham & Jasmine Chu

February 2022

Manipulating Data using the Tidyverse

This homework assignment requires you to download data from the World Bank Ed Stats dataset and tidy it into country year format. In addition to `pivot_wider()` and `pivot_longer()` functions from this module, you will also need to use some dplyr commands to rename variables, make character variables numeric and to filter out excess rows from the raw data. So this homework requires all your chops from both Data Management 1 and Data Management 2. If you can complete this assignment, you are well on your way to being able to wrangle data effectively for social science research.

If you are a SPEC Lab students, please 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

Helpful Hint: When working with variables that have more than one word, and are separated by spaces, use backticks or back quotes (```).

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 on the basis of country and year, using `full_join()`. Note: SPEC Lab best-practice is to merge by Gleditsch-Ward Country Code and year, but because these two datasets both come from the World Bank, we can get away in this homework with merging by country and year. Please complete our Data Management 2A module to learn how to append Gleditsch-Ward numbers.

OPTIONAL BONUS Complete Data Management 2A. 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.