

SPEC Lab R Resources: Data Management 3- Group Work

Alix Ziff and Miriam Barnum

Summer 2021

Data Management for Visualization

We're continuing to work with our IDC powersharing data to hone our data management skillset. We will use the packages `dplyr` and `countrycode` to work with country-year data. For these exercises, you will follow similar protocol to the Walk-Through-Work but apply it to new variables.

After creating a detailed header, set your working director and use `readRDS()` to load in "IDC_analysis_master_MB_20210414".

Exercise 1 Say we want to look at global averages of subnational data. Collapse our country-year data to get global averages by year for the variable `subpolice`.

Exercise 1.2 Take a look at your results, what is the average for 2007?

Exercise 2 Now, incorporate the other subnational variables in the IDC dataset on subnational education and taxes (`subed`, `subtax`).

Exercise 3 For our next exercise, let's look at regional averages for another variable. You choose.

Helpful Hint: You'll need to add a new variable, either by region or continent. *Helpful Hint:* Avoid categorical variables or those with several NA values. . . review Data Management I for how to explore your data.

Exercise 4 Let's practice simplifying and summarizing data. Choose a single variable and find either the decade averages, minima, maxima, or median values.

Exercise 5 Do the same thing as above, but incorporate multiple variables.

Exercise 6 Pick a year so that we have a single value per country. *Helpful Hint:* Remember you can either do this in base R or `dplyr`.

Exercises 7-10 DIY: apply all the above steps to make a data set which includes annual values for some individual countries within one region, and the global average.