

# SPEC Lab R Resources: Data Management I - Homework

Ben Graham, Alix Ziff, and Jasmine Chu

Summer 2022

## Transforming Data using the Tidyverse

Complete the following assignment. Save your R script and the mini dataset from step 2 into your personal subfolder in the Homework Submission google drive folder. The R script should be titled HW\_DM1\_[YOUR INITIALS]. Please also make sure you save your R script to your own computer for future reference.

Don't forget to annotate your script thoroughly! And save a copy to your personal R script library. These scripts are a resource for future you!

### Exercise 1:

**A** Load in the .rds file of the IDC 2021 powersharing data: "Training Data Summer 2022/IDC\_training\_2021.rds"

**B** Complete step B in a single command, piped together. Create a subset of the data that includes only the following countries, years, and variables:

1. Countries: U.S., China, Russia, France
2. Variables: country, gwno, year, subed\_IDC, subtax\_IDC, subpolice\_IDC, auton\_IDC, stconst\_IDC
3. Years: 2015-2018

**Exercise 2:** Save this smaller dataset as "Minipowersharing\_YOURNAME.rds". Ideally, specify a filepath in your save command so that it saves straight to your homework submission folder in google drive.

**Exercise 3:** Using the full dataset again, create a new variable, "subpower\_additive" that is the sum of subed\_IDC, subtax\_IDC, and subpolice\_IDC. This index should take a value of N/A if any of the three component indicators is missing.

**Bonus:** Create a new version of the index, "subpower\_additive\_nm", that assumes subed\_IDC, subtax\_IDC, and subpolice\_IDC take a value of 0 if they are missing. This version of the index should have no missing values.

**Exercise 4:** Use summarise() or summarise\_at() to answer the following:

1. What is the mean value of your first subpower index (`subpower_additive`) in the entire sample of countries, across the years 2010 through 2019? *Hint:* This should be one value.
2. What about in the year 2019 only?
3. **DOUBLE BONUS:** How about the mean value of your second version of the index (`subpower_additive_nm`) for the entire sample?
4. **TRIPLE BONUS:** How many countries in 2019 have a value for the `_nm` version but not for the original version?