

SPEC Lab R Resources: Data Management I-Homework

Ben Graham and Alix Ziff

Summer 2020

Transforming Data using the Tidyverse

Complete the following assignment and then save your R script and the mini dataset from step 2 into the associated google drive folder. The R script should be titled HW_DM1_[YOUR INTIALS]

Exercise 1:

Complete steps A and B in a single command, piped together.

A Load in the .csv file of the IDC 2019 powersharing data. The file path and file name are: “Data Science Modules/Training Data/idc2019_merged_for_training_May2020.csv”

B Create a subset of the data that includes only the following countries and variables:

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

Exercise 2:

Save this smaller dataset as “Minipowersharing_YOURNAME.Rdata”

Exercise 3:

Using the full dataset again, create a new variable, subpower_additive that is the sum of subed, subtax, and subpolice. 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, subtax, and subpolice 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 index in the entire sample of countries, 2010-2019?
2. What about in the year 2019 only?
3. **DOUBLE BONUS:** How about the mean value of your __nm version 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?