SPEC Lab R Resources: Powersharing Groupwork for Data Management I

Alix Ziff and Benjamin A.T. Graham

April 2021

Data Investigations: Reserved Government Positions

This groupwork uses the "idc2019_merged_for_traning_May2020.csv" data file. This is country-year data gathered in the SPEC Lab. The data is already tidy. Each row is a unique observation identified by country code (gwno) and year; each column is a variable with information about each country's political institutions.

Getting Started

Write the header for your R script and save the script to your personal R script folder.

Note: Annotate the heck out of your code as you complete these exercises. LOTS of notes to yourself. # at the start of a line tells R that what you are writing is a comment, not a line of code.

Set your working directory to the Training Data folder you have downloaded.

Use the `library()` function to load in the dplyr and foreign packages.

Exercise 1: Load and view the dataset. This is an .csv file so you will want to use the `read.csv()` function.

Question 1: What is the second variable in the dataset, reading from left to right? What is the eighth?

FYI

When you use the `select()` function, you will keep only the variables you mention. You can also use `rename()` to rename a variable without dropping all other variables.

There are several additional helper functions you can use inside `select()` such as `starts_with()`, `ends_with()`, `num_range()`, `matches()`. You can use the `filter()` command to find rows/ cases based on certain conditions. The `mutate()` function adds new variables while preserving old ones. When you use `mutate()`, make sure to make a unique variable name so you do not overwrite existing variables. You can use transmute() to add new variables and drop existing ones. In this problem set you will use piping with the notation `%>%` to make your code more efficient.

We will be focusing on variables relating to religion in the power sharing dataset. These variables have the prefix "rel" and include state establishment of religion, state restriction of religion, protection of religion, and protection against discrimination on the basis of religion.

DIY

selecting variables and observations

Exercise 2: select only the variables related to government positions reserved for minority groups and those related to subnational policy authority: "resseats", "resman" "resimp", "subed", "subtax", "subpolice" and then also retain the three variables "gwno", "country" and "year".

*Hint: Try using the `contains()` or `starts_with()` helper function to grab the reserved positions and subnational policy authority variables if you want to be fancy.

Exercise 3: using the `filter()` function, create a smaller dataframe that keeps only cases where subnational governments have either subnational tax authority or subnational education authority (i.e. subtax or subed takes a value of 1).

Question 2: Name a country-year in which more than 20% of the legislative seats are reserved for members of minority groups (i.e. resseats is greater than 0.2).

State Protection of Religion

Exercise 4 : Use the `mutate()` function to create a binary variable named "resseats_10" capturing whether at least 10% of legislative seats are reserved for minority groups.

Exercise 5 We want to know whether reserved executive positions (resimp) have been getting more common over time. Use the `group_by()` and `summarise()` functions together to calculate the global proportion of countries that have this type of protection (i.e. the global average of resimp) by year. Is this type of provision present in a higher percentage of countries in 2010 or in 2019?

Last thing: Put it all together. Use piping to combine exercises 1, 2, 4, and 5 into a single, elegant command.