Your Challenge

Nick Eubank

August 10, 2018

Today we will be shamelessly stealing an exercise from Simon Ejdemyr, the writer of the tutorials we have been using.

Our Goal

A prominent economic theory predicts that higher income inequality should be associated with more redistribution from the rich to the poor (Meltzer and Richard 1981). Lets create a dataset that will allow us to test this prediction using U.S. state-level data. Our goal is to look at the correlation between income inequality and state taxes per capita (i.e. the amount of taxes each state collects from each resident on average). I am happy if we just look at this for one year (say 2010).

Our Data

We have four datasets:

1. State-level measure of inequality (Gini coefficient, and Share of Income Going to Top 1% and 10%s).
2. Total state taxes and total state income taxes.
3. State populations.
4. A dataset that has state names (written out) and state codes (called FIPS codes).

Question 1: What Variables Do We Need?

(Just think about this generally, I don't need specific variable names)

Question 2: Where Can We Find These Variables?

(Now I want specific variable names)

Question 3: What Variables Will Allow Us To Merge the Datasets We Need?

(Draw me a picture where each box is a dataset, and the arrows between tell us what variables we can use to link them.)