Introduction to R

Exercise A

The purpose of this exercise is to "get your hands dirty" measuring, collecting, coding, and assembling data in the R language. In particular, you'll be building four semi-original datasets, each with somewhat different characteristics. This exercise has four parts; specific instructions are below. Feel free to skip Part IV if it seems difficult. Also, note that hot links appear in OSU orange.

I. Cross-Sectional Data

1. Build a database where the unit of analysis is the faculty member – specifically, members of the School of Public Policy at OSU.

2. Operationalize, measure, and code four variables:

   - The faculty member's *name*,
   - The faculty member's *rank* (instructor, assistant, associate, or full professor),
   - The *discipline* of the faculty member (political science, economics, sociology, etc.), and
   - The gender self-identification (male or female) of that faculty member.

Obtain this information from any available sources. Be sure to make your decisions about operationalization, etc. as transparent as possible, and document your sources.

II. Time Series Data

1. Count the number of letters in your first name (k), and pick the kth state from the alphabetical list of U.S. states (so, if your name was "Christopher," you'd select the 11th state, which happens to be Georgia).

2. Go here, grab the Excel file at the bottom of the page, and use the information in that file to create a time-series dataset of your state's values on the following variables:

   - Population, 2000-2010,
   - Annual SNAP participant counts, 1997-2012, and
   - Annual SNAP participant benefits, 1969-2012.

Note that the documentation for the spreadsheet is here.

III. Time-Series Cross-Sectional ("Panel") Data

1. Return to the state you used in the time-series exercise, above, and build a panel database of "your" state's current members of Congress. Record one observation for each such member for each Congressional session (2-year term) they have been in office, as of 2022.

2. For each such member in each session, collect and code data on:

- that member's political party identification in that term,

- the number of terms s/he has served, and

- the party of the state's sitting governor in that term/session.

Once again, obtain this information from any sources you find available, and note your sources in the write-up of the exercise.

IV. Relational Data

1. Go here, roll the die once, and add 2000; that's your year.

2. Navigate to the website of the amazing Correlates of State Policy project. Use the information / data there to build a relational dataset of the 50 U.S. states. In this dataset, each row of the data should be one dyad / pair of states. Thus, your dataset should have a total of $\frac{50 \times 49}{2} = 1225$ rows.

3. In your dataset, code / include the following data; data should be from the year indicated by step (1):

   - Whether (=1) or not (=0) that pair of states both have the same value on the *Income Taxes* variable,

   - Whether (=1) or not (=0) that pair of states both have the same value on the *Electronic Waste Recycling Program* variable, and

   - The absolute value of the difference between the two states' values on the *Top 1% Share of Income* indicator.

Note that this part of the exercise is best done via computation, not "by hand."

General Matters

1. For each part listed above, the data you create should take the form of comma-separated value files (`.csv` files), *not* Excel spreadsheets.

2. Be sure to include unit identifiers in each dataset.

3. In addition to the datasets, write up a PDF file briefly outlining what you did in each of the four sections above. In the latter, include (a) the coding scheme for each variable you coded, (b) its level of measurement, and (c) the number of non-missing observations you collected/assembled data on.

4. This is an ungraded exercise.