PLSC 502: "Statistical Methods for Political Research"

Exercise One

The purpose of this exercise is to "get your hands dirty" measuring, collecting, coding, and assembling data. In particular, you'll be building three original datasets, each with somewhat different characteristics. Specific instructions are below.

Cross-Sectional Data

1. Build a database where the unit of analysis is the faculty office – specifically, offices held by tenured or tenure-track political science faculty at PSU (that is, in Pond Lab).
2. Operationalize, measure, and code three variables:
   - Whether the office in question is on the second or third floor of Pond Lab,
   - The rank (fixed-term instructor, assistant professor, associate professor, or full professor) of each office's inhabitant, and
   - The gender (male or female) of the office's occupant.

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

Time Series Data

1. Go here.
2. Count the number of letters in your first name (k), and pick the kth state from the pull-down menu (so, if your name was "Christopher," you'd select the 11th state on the list, which happens to be Georgia).
3. Use the pull down menus to build an annual time-series dataset of the following three variables:
   - Per capita personal income, 1969-2000,
   - Per capita federal government expenditures, 1983-2005.
   - Infant mortality (infant deaths per 1000 live births), 1990-2004.

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 that 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 2016.
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.

Data should take the form of comma-separated value files (`.csv`), *not* Excel spreadsheets. Submit all three datasets in electronic format – via e-mail attachment – to Nick (dietrich.nicholas@gmail.com) and/or to me (zorn@psu.edu) no later than **5:00 p.m. EST** on **Thursday, September 8, 2016**. In addition, include a PDF file briefly outlining what you did in each of the three sections above. In the latter, include (a) the coding scheme for each variable you coded, (b) it's level of measurement, and (c) the number of non-missing observations you collected/assembled data on. This exercise is worth 50 possible points.