Code Structure

Everyone develops their own approach to organizing analysis files and differences across investigators can pose a challenge for collaborators co-authoring code. Nonetheless, each project will have a similar set of scripts. These will likely include scripts to a) extract variables, b) create new variables, and c) analyze data and d) produce tables. It is useful to have a main do file that calls and executes each set of scripts in order.

To facilitate collaboration and archival of analysis for replication by non-project members, we have developed a set of conventions.

    1) README.txt. Within our main code directory, we have a file called README.txt with basic information anyone needs to know to get started on the project.

    2) Setup files that each person can identify the location of each type of directory in their file structure. ADD MORE SPECIFICS.

    3) A main script that begins with loading original data and ends by producing publication-ready tables. Note that you should nest scripts within other scripts. For example, you might have one stata .do file that executes all of the code to create your analysis files and another .do to execute all the scripts to do the data analysis. The main do file would call and execute both.

    4) A system of logging the state of the computer when it ran the analysis, the set of steps to produce the results, and the results.

We are also developing techniques to automate the production of tables for publication. See automate for more.