Stata Workshop 2: Data Management

Good File Management is important for good research. Data files get messy and complicated and can feel overwhelming. Organizing files – and developing a system that you can use for different projects – will go a long way to organizing your thinking and improving your productivity.

Concrete Tips:

1. Organization folders on your computer and create a similar folder structure across projects.
    a. Create subfolders.
    b. Don't be afraid to reorganize as a project grows in complexity.
    c. Use variables for path names to allow for flexibility.

2. Keep "raw data" pristine. Have a separate folder for it and do not overwrite it.

3. Put lots of comments in your code. There are three ways to do this in Stata.
    a. // comments. For example,

       //   Select sample based on age and gender.

    b. You can add /* and */ comments. For example,

       /*   Selection sample based on age and gender   */

    c. You can add comments as dividers. For example,

       ****
       ***.   Select sample based on age and gender   ***
       ****
4. Use alignment, spacing, and indentation to make things readable.

5. Create a common structure across all projects. I use a:
    a. Raw Data Folder: Folder that contains subfolders for each of the original data I'm using.
    b. Build Data Folder: Folder contains subfolders each with cleaned data corresponding to do files that clean each dataset.
    c. Programs Folder: Folder that contains programs that manipulate the data. Older programs that are no longer needed are in an "archives" subfolder
    d. Output Folder: Same names as the corresponding do files in the Programs folder;
    e. output that is no longer needed also goes in an "archives" subfolder

6. Label Programs and Log Files Similarly
    a. Good things to have at the beginning of a Do File:

    version 15.0
    prog drop _all
    clear all
    capture log close
    set linesize 80
    display "$S_DATE"

    What do these mean?

version 15.0	Tells Stata what version you are using
prog drop_all	Drops any pre-existing programs you are using
clear all	Clears any preexisting data
capture log close	Closes any open log files
set linesize 80	Generates the output that is formatted the same way
display "$_DATE"	Displays current data in Stata