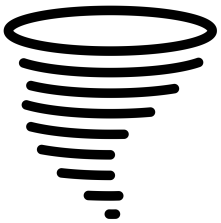
Directory Organizational Tips

Based on *The Workflow of Data Analysis Using Stata* by Scott Long

Common problems of organization



- Multiple versions of a dataset exist and you don't know which is which
- You have trouble finding a file, think you may have deleted it
- When collaborating on writing projects, you find there are multiple "latest" version of the manuscript draft.
- You want to find the final version of a questionnaire and are unsure which version it is because two have the word "final" in the file name.

A carefully designed file naming and file organization structure can help this.

Tips for organizing your directory

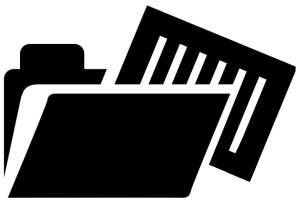
1. Choose a mnemonic for each project (avoiding lengthy or common words like "THE"), and use as a project identifier in directories and file names associated with the project.
2. Create separate folders for "working" and "posted" files:
    - **Working files** are the ones that contain scripts, texts, and files that are actively being worked on.
    - **Posted files** are "frozen" versions of scripts, datafiles, and texts that do not change. For example, once a manuscript is sent to colleagues, it gets moved into "Posted" and is no longer changed - rather a new version is saved to the "Work" folder in which the changes will be made.
3. Create folders for specialized tasks. Beginning these directories with a dash ensures they will show up at the top when the files are sorted by name.
    - "-To Do" (where things that need to be done, read, or revised should go)
    - "-To Clean" (for data that need to be treated or de-identified)
    - "-Hold then delete" (for files that can be deleted but need to hold for easy recovery in case a mistake is made)
    - "-Readme" (file or folder than contains the naming directory and describes what that folder contains)
4. If working with collaborators, you may want to set up mailbox folders for files that are exchanged between people. For example:
    - Eliza to Fong (files from Eliza to Fong)
    - Fong to Eliza (files from Fong to Eliza)

Example directory structure

Top level directory with project name/mnemonic

- Sub-directories for:
    - Administration (correspondence, IRB, budgets, etc)
        - Budget (spreadsheets, billing info)
        - Correspondence (emails, letters)
        - Proposal (grant proposals, related materials)
    - Documentation (research log, codebooks, other documentation)
    - Posted (completed texts, datafiles, syntax files, etc)
        - DataClean (clean data and construct variables)
        - Datasets
            - Derived (datasets constructed from the source datasets)
            - Source (original, raw, unchanged data files)
        - DescStats (descriptive statistics)
        - Figures (programs to create graphs)
        - Panel Models (panel models of discrimination)
        - Text (drafts of paper)
    - Readings (PDF of literature related to the project)
    - Work (text and analyses that are being worked on)

How to make the switch and keep directory organization?



Created by Alex Auda Samora
from Noun Project

- First decide on a directory structure to use. Should have short names for the folders.
- Use a spreadsheet to layout and document the directory structure ahead of time - this will help future you and collaborators get on the same page.
- Use the same structure for and folder naming schemes for all projects.
- Create a text file in the main directory with a description of the folders and the content that should go in each.

- Plan file names as well. Some tips:
    - Save articles as Author's last name_year_journal_keyword (e.g., Smith_2005_JASA_missingdata.pdf).
    - Use year-month-day date formats with leading zeros: "Data_20170118" rather than "Data_Jan1817" so files will sort chronologically.
    - Use leading zeros in version numbers for the same reason: "version01", rather than "version1" so that "version10" will sort in order.