

# Data and Syntax Organization

- **File Content:** Below describes what should be accomplished in each syntax file. You may wish to have one file per section, more, or fewer. In each file, open the same-lettered data file at the top, and save it at the bottom.
- **Syntax:** Kept syntax should not be a record of steps taken (i.e., in order, including mistakes and fixes), it need only allow replication of the result. Put syntax in the appropriate file, regardless of when it was thought of or added.
- **Backups:** Only your syntax files and the original data are important, neither of which shrinks over time. Thus, although it is important to have current backups, saving older versions of the syntax and data files is unnecessary.
- **File Names:** Prefix Syntax and working datafile names with the letter of the relevant step. If you have multiple raw datafiles, give each a nickname and use it in all related file names. No version numbers or dates should be needed.

---

## Group

**Before starting...** put all original data files in a separate directory and never save there

---

**A** These files are for data cleaning and making preparations for merges. Make only the changes necessary.

- Aquire
  - Open, import, or convert each raw data file into your statistical software
  - Convert storage types (numeric vs string) and add display formats (date, %, etc.)
- Adjust  
*as needed*
  - Name or rename variables to simplify, add unique identifiers
  - Create a filter variable to exclude cases with no data or obviously not in the population
  - Prepare datasets for merging by synchronizing names, values, and types
- In Master
  - Aggregate, merge, restructure, or append if little to no preparation is needed
  - Re-save data files as **B\_**'s, deleting filtered cases and dropping useless variables or constants

---

**B** These files prepare the data for sharing or future use, primarily by adding labels and additional variables.

- LaBel
  - Label or re-label variables and values; label as many values as reasonable
  - Identify values considered missing and indicate (for SPSS) or recode to missing
- Blend  
*as needed*
  - Export and code or categorize text or open-ended responses, import
  - Export as needed to perform external calculations (ex. Epi-Info), import
- In Master
  - Merge prepared (and filtered) data into the main datafile
  - Re-save data file as **C\_working**, keeping or dropping variables as desired

---

**C** These files refine the data to address your specific research question through computing, recoding, and filtering.

- Create
  - Perform planned or obvious calculations or conversions
  - Create groupings for subgroup analysis
  - Create scores, scales, or indexes (reversing values as needed, checking reliability)
  - Recode to have fewer groups or avoid small groups (label values immediately)
- ExClude
  - Create filter variables based on sufficient valid data and inclusion criteria (as specified)
  - Address missing values and outliers by recoding or imputing, ignoring, or adding filter criteria
- In Master
  - Re-save data file as **D\_analysis**, deleting filtered cases and keeping only needed variables

---

**D** These files contain the necessary syntax to **reproduce** all numbers, figures, and tables in your paper

- Describe
  - Frequency tables and crosstabulations
  - Charts and Graphs (use comments to indicate Figure #)
- Data  
Analysis
  - Also create analysis-specific variables where needed (ex. dummy, transformed, or centered)
  - Also check for analysis assumptions and model appropriateness where needed