TIDY DATA PRINCIPLES:

Designing spreadsheets that are “tidy, consistent, and as resistant to mistakes as possible” (2)

Be Consistent:

- Use consistent codes for categorical variables
- Use a consistent fixed code for any missing values
- Use consistent variable names
- Use consistent subject identifiers
- Use a consistent data layout in multiple files
- Use consistent file names
- Use a consistent format for all dates
- Use consistent phrases in your notes
- Be careful about extra spaces within cells

Choose Good Names for Things:

- Avoid spaces
- Avoid special characters
- Be short but meaningful

Write Dates as YYYY-MM-DD

No Empty Cells

Put Just One Thing in a Cell

Make it a Rectangle

- Single first row with variable names

Create a Data Dictionary

- “This is part of the *metadata* that you will want to prepare: information *about* the data” (6)
- You might also find this information in a codebook that goes with a dataset
- Things to include:
    - The exact variable name as in the data file
    - A version of the variable name that might be used in data visualizations
    - A longer explanation of what the variable means
    - The measurement units
    - Expected minimum and maximum values

No Calculations in the Raw Data Files

Do Not Use Font Color or Highlighting as Data

Make Backups

- Multiple locations (OneDrive, local computer, etc.)
- Version control program (i.e. Git)
- Write protect the file when not entering data

Use Data Validation to Avoid Errors

Save a Copy of the Data in Plain Text Files

- File formats can include comma-separated values (CSV) or plain-text (TXT)