Foundations - best practices

1. Data standardisation
    a. The process of making sure that your data set can be compared to other data sets.
    b. Key part of research.
    c. Should consider this before you even collect, clean, or analyse your data.
    d. Why? Because data is most valuable when you have something to compare it to.
    e. However, comparisons aren't helpful if the data is bad or irrelevant.
    f. Standardised data is essential for accurate data analysis.
    g. What steps should be taken to standardise your data?
        i. Decide on data standards (formats; baseline measures).
        ii. Figure out where your data will be coming from (internet, questionnaires, etc.).
        iii. Set up a great survey; do data validations.
        iv. Collect data in common formats.
        v. Collect data based on pre-set standards (e.g. International System of Units).
        vi. Transform data to a common format.
        vii. Understand and clean your data - data should be correct, clean, complete, formatted and verified.
2. Plain text formats are your friend
    a. Computers can process them - they are machine-readable.
        i. Tip: if you can't find it by Ctrl + F, it isn't machine-readable!
    b. Use .txt for notes and .csv/.tsv for tabular data.
    c. Proprietary formats (e.g. Microsoft Word) may become obsolete in future.
    d. Use markdown for formatting - markdown files are machine-readable, easily searchable, and human-readable. These files can be converted to PDF, HTML, formatted Word document, etc.
3. Automated/computational approach
    a. Computational skills improve your efficiency and effectiveness.
    b. Keyboard shortcuts save time
        i. Save: Ctrl + S
        ii. Cut: Ctrl + X
        iii. Copy: Ctrl + C
        iv. Paste: Ctrl + V
        v. Undo: Ctrl + Z
        vi. Redo: Ctrl + Y
        vii. Switch applications: Alt + Tab
4. Working directory and naming files

a. Easier to work with your data if it is structured and organised.
b. Directories = folders.
c. File naming convention is crucial.
    i. Choose file names that identify them, create associations between data elements, and to assist with long term readability and comprehension of your data structures.
    ii. With automation, this will come in very handy, especially when you want to use regular expressions.