The Workflow of Data Analysis

Quantitative **M**ethods in **E**ducational **R**esearch

1. What is a Workflow?

Planning Project	Documenting Choices	Cleaning Data	Creating Variables
Producing Analyses	Reproduce Results	Presenting Results	Archiving

Reproducibility

2. Workflow Criteria

Accurate

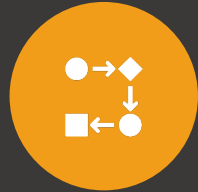Efficient

Simple

Standardized

Automated

Usable

Scalable

3. Steps in Workflow

01
Planning

02
Organization

03
Documentation

04
Execution

Planning

3.1. Planning

a. *Main Idea: Set goals and work backward, Know data and work forward*

b. Naming Conventions
   a. - Prefix & Suffix
   b. - Types of variables needed
   c. - Missing Data

c. Documentation
   a. What types of documentation do you need and where?

d. Backup
   a. Where are files stored, what if the program you use goes out?

Organization

3.2. Organization

a. *Main Idea: Could someone else figure out what this is?*

b. Pick a mnemonic

c. Create a directory structure

   a. Main Subfolders: Working and Posted

   b. E.g. Writing, Documentation, Syntax, Data, Outputs (Figures/Tables)

Documentation

3.3. Documentation

a. *Main Idea: Breadcrumb everything*

b. Project Log

   a. Tasks, Thoughts, Decisions

   b. Daily Choices

c. Syntax Log

   a. Decisions

   b. Syntax Notes

d. Variable Codebook

Execution

3.4. Execution

a. *Main Idea: Reproducibility*
b. What are your processes for executing tasks?
c. Syntax, references, file locations
d. How are outputs generated?

Now you know everything!