Typical data analyses

**Data processing:** may include selecting a subset of data for analysis, merging multiple data sets, manipulating data for usability, or data transformation.

**Graphical analysis:** makes it easier to see patterns and can aid in the identification of outliers.

**Statistical analysis:** conventional statistics are used to analyze experimental data; descriptive statistics are used to analyze observational or descriptive data.

Science is iterative: the process that results in the final product can be complex.

Reproducibility...

...is at the core of the scientific process. If results are not reproducible, they lose credibility.

**Good documentation of the data and the analysis are essential!**

Workflows

**Definition:** Precise description of the procedures used in a project. Can be formal or informal.

**Informal workflow**

No special software is needed to create workflow diagrams. Workflow diagrams include:
- Inputs and outputs
- Transformation rules or analytical processes
- Decision points
- Arrows indicating direction of process flow

**Best practices for data analysis**

Formally or informally document the workflows used to create results. Include:
- Data provenance
- Analyses and parameters used
- Connections between analyses via inputs and outputs

Document the code you write for analyses.
- Well-documented code is easier to review and share and enables repeated analyses
- Include project level information; script dependencies, inputs, and outputs; parameters; and what happens in individual sections

Construct end-to-end scripts that run the entire process from start to finish without intervention.