Structures: Project Folders

- Admin
- Background
- Code
- DataRaw
- Data
- Dissemination
- Reports

Structures: Project Folders

- Admin
  - SOW, Letters of support, biosketch, etc.
- Background
- Code
- DataRaw
- Data
- Dissemination
- Reports

Structures: Project Folders

- Admin
- Background
  - Study docs, prior publications, the grant, and other background info
- Code
- DataRaw
- Data
- Dissemination
- Reports

Structures: Project Folders

- Admin
- Background
- Code
  - ALL code, and ONLY code
- DataRaw
- Data
- Dissemination
- Reports

Structures: Project Folders

- Admin
- Background
- Code
- DataRaw
  - ONLY the raw data provided by investigator, saved as READ ONLY
- Data
- Dissemination
- Reports

Structures: Project Folders

- Admin
- Background
- Code
- DataRaw
- Data
  - Merged, cleaned, or otherwise manipulated datasets, output data, etc.
- Dissemination
- Reports

Structures: Project Folders

- Admin
- Background
- Code
- DataRaw
- Data
- Dissemination
  - Abstract, poster, and manuscript files (including figs and tables submitted for the manuscript)
- Reports

Structures: Project Folders

- Admin
- Background
- Code
- DataRaw
- Data
- Dissemination
- Reports
  - Reports and results sent to investigator

Structures: Program Naming Conventions

- 01_DataImport
- 02_Exploratory
- 03_Descriptives
- 04_Analysis
- 05_Manuscript (OPTIONAL)

Structures: Program Naming Conventions

- 01_DataImport
  - Contains all data import, cleaning, data manipulation, data build stuff, variable creation
- 02_Exploratory
- 03_Descriptives
- 04_Analysis
- 05_Manuscript (OPTIONAL)

Structures: Program Naming Conventions

- 01_DataImport
- 02_Exploratory
  - This file doesn't make it into the final report / project build (if anything of value comes out of this, it goes into another file)
- 03_Descriptives
- 04_Analysis
- 05_Manuscript (OPTIONAL)

Structures: Program Naming Conventions

- 01_DataImport
- 02_Exploratory
- 03_Descriptives
  - Table 1 and other summary plots of the data
- 04_Analysis
- 05_Manuscript (OPTIONAL)

Structures: Program Naming Conventions

- 01_DataImport
- 02_Exploratory
- 03_Descriptives
- 04_Analysis
  - If there are multiple hypotheses or multiple research questions, this needs to be broken up into Research Question (e.g. H1, H2, H3, etc or some other name)
- 05_Manuscript (OPTIONAL)

Structures: Program Naming Conventions

- 01_DataImport
- 02_Exploratory
- 03_Descriptives
- 04_Analysis
- 05_Manuscript (OPTIONAL)
  - If there were multiple analyses which didn't make it into a manuscript, this file can be created to reproduce all numbers in the manuscript (and nothing else)

Structures: Program Headers

```
################################################################################
### file:
### authors:
### creation date:
### description:
################################################################################
```

Structures: Coding

- “You are writing a program that demonstrates the thought process along with the analysis you went through” ~ MM
- All code files should be executable without modification, and sequentially so
  - 01_DataImport file is to create all datasets used for subsequent analyses (with separate names); running the same analysis on different datasets / subsets requires different chunks of code, with explanation as to what / why / final choice
- Use comments to explain “why”, not just “what”

Structures: Coding

- Limit line length (80-100 characters)
- Limit file length (~ 100 lines)
- Minimize extraneous operations or move to “appendix” or “helper function” script
- Be consistent (indenting, capitalization, spacing, etc.)
- Variables are not to be created in the middle of analysis code – this happens during data build

Structures: Coding

- Set the working directory once
- Packages are always loaded at the top of the script
- Things we'll discuss more / later:
  - Helper functions / analysis function files
  - Build scripts and sourcing function files