Best Practices Guidelines for Projects 2 and 3

A goal of science is to produce work that is reproducible and replicable. Data analysis and coding is not exempt from these ideals! One should be able to successfully replicate all the results from code that you create and understand what the code is doing.

The following Best Practices Guidelines exist to help you in these efforts. These guidelines are derived from the two papers that we read at the beginning of Unit 3. Following these guidelines are a requirement for Projects 2 and 3. Treat this list as a checklist before turning in these projects!

- **Organization**: All elements of project are in a single folder and follows standard organization.
  - A subfolder labeled **src** that contains all scripts. Any functions or data cleaning should exist in at least one script in this folder.
  - A subfolder labeled **doc** that contains documentation.
  - A subfolder labeled **data** that contains all raw data files.
  - A subfolder labeled **results** that contains all files produced by the code.
- **Documentation**: All projects should contain basic documentation.
  - **README file** which describes the project, the organization of the project, and how to run any files that produce results and their locations.
  - **A Project Rmd file** which reproduces the visualization(s) in the doc/ folder.
  - Documentation describing individual functions in the project, including a description of inputs and outputs.
- **Style**: All project code should follow style guidelines in http://adv-r.had.co.nz/Style.html. The most important are:
  - Use `<-` and not = to assign value to objects. Use = inside functions with arguments.
  - File names and object names should be lowercase and use a period or underscore (_) to separate words within a name.
  - File names and object names should be short and meaningful.
  - If file names need to be run in a sequence, prefix them with numbers (e.g. 0, 1, 2, ...).
  - Space should be used to separate all operators (=, +, −, `<-`) and after commas.
  - Lines should be about 80 characters long, and longer lines should be broken up with indentation for readability.
- **Version Control** (Optional for independent projects)
  - All project elements are tracked by git in a single folder (including raw data files).
  - The project exists as a single repository on Github.
  - Each member of the team updates the code via commits that are pushed to Github.
  - A release is created and submitted as the final version.