Draft Title: A Code-Based Data Workflow Using R

This webinar provides a brief introduction into using R to develop a code-based workflow for ecological survey data. Participants will learn about the advantages of using a scripting language like R for moving data along the data life cycle. We will walk through an example of using R to manage scientific data.

1 Webinar Information

Presenters: McCrea Cobb and Adam Smith

When:

Location: Webinar link

Additional resources: GitHub repository

2 Outline

1. Data life cycle review
2. The manual data workflow
    - Example
    - Limitations
3. The code-based data workflow
    - Advantages
        - Documented
        - Reproducible
        - Replicable
        - More efficient
        - Less error-prone
4. A data workflow with R
    - **Planning** an RStudio project (*McCrea*)
        - Make an R project self contained and portable
            - File directory structure
            - Relative paths
        - Dependency management
            - packrat
            - containers (docker)
        - Standardize file naming convention
        - Organizing R files (Numeric preface in the names of ordered scripts)
        - Recommended RStudio settings
            - E.g., Uncheck "restore .RData into workspace at startup"
        - Version control

        * Storing versions
        * Collaboration
    - **Documenting** data and scripts with R (*Adam*)
        - rOxygen
        - R documentation file
        - Code commenting
    - **Acquiring** data (*Adam*)
        - local and remote
        - querying data
            * AGOL
            * iNaturalist
            * PRIMR web services
            * SQL query: IRIS warehouse
    - **Processing and Analyzing** (*Adam*)
        - Getting data into R
        - QC
        - Tidying data
        - Visualizing
        - EDA
    - **Sharing** (*McCrea*)
        - Reporting
            * RMarkdown
                · Bat reporting for mobile aucistics
            * Shiny apps
                · collarviewer
                · power analysis for butterfly surveys
    - **Archiving** (*McCrea*)
        - Saving results to ServCat or some other data repository
5. An example R project / Live demo (*Both*)

3 *(Topics Ideas)*

The list below is taken from https://learn.datacamp.com/courses/working-with-data-in-the-tidyverse:

1. Explore your data
    - Import
    - Dealing with missing values
    - Exploratory data analysis (glimpse and skim)
2. Tame your data
    - Cast column types
    - Recode values
    - Select variables
    - Tame variable names
    - Rename variables to convention (janitor package)
3. Tidy your data
4. Archiving

4 Resources (Links)

Resources for Teaching R

- DataCamp's tidyverse course
- learnr package
- RStudio teaching resources
- Data Wrangling, Exploration and Analysis with R "STAT 545"

R Resources

- Why learn R
- What they forgot to teach you about R
- R cheatsheets
- Project-oriented workflow

R Packages

- Packaging your reproducible analysis
- R packages
- Packaging data analytical work reproducibly using R (and friends)

Project management

- Stop working directory insanity!
- A minimal project tree in R
- Organizing the project directory
- Designing projects
- Project management with RStudio
- File structure for data management
- Organizing files for data analysis
- A meaningful file structure for R projects

General Coding Best Practices

- What's in a name? The concepts and language of replication and reproducibility

Version Control

- Happy Git with R

Other

- How to share your data with a statistician
- Tools for reproducible research
- Reproducibility vs. replicability: a brief history of a confused terminology