

R tools for a code-based data workflow

Contents

Webinar Information	1
Description	1
Presenters	1
When	2
Location	2
Additional resources	2
Outline	2
Introduction (<i>McCrea, 5 min</i>)	2
Planning (<i>McCrea, 10 min</i>)	2
Documenting (<i>Adam, 10 min</i>)	3
Acquiring (<i>Adam, 5 min</i>)	3
Processing and Analyzing (<i>Adam, 15 min</i>)	3
Sharing (<i>McCrea, 10 min</i>)	3
Archiving (<i>McCrea, 10 min</i>)	3
An example R project / Live demo (<i>10 min</i>)	3
Questions (<i>10 min</i>)	3
Resources (Links)	3

Webinar Information

Description

After a brief review of the advantages of a code-based workflow for ecological survey data, we introduce participants to some useful tools available via the R programming language for moving data along the data life cycle. We suggest some accessible tools in R for each step of the life cycle, and conclude with a walk through of how the functionality available in R can increase the reliability, efficiency, and transparency of scientific data management.

Presenters

- McCrea Cobb (Refuge Inventory and Monitoring Program, Alaska) and Adam Smith (Refuge Inventory and Monitoring Program, Legacy FWS Region)

When

June 24, 2020 (3:00-4:30 EST)

Location

Webinar link

Additional resources

GitHub repository

Outline

Introduction (*McCrea, 5 min*)

- Data life cycle review
- Manual versus code-based workflow
 - The manual data workflow
 - * Example
 - * Limitations
 - The code-based data workflow
 - * Advantages
 - Documented
 - Reproducible
 - Replicable
 - More efficient
 - Less error-prone

Planning (*McCrea, 10 min*)

- 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 (*Adam, 10 min*)

- rOxygen
- R documentation file
- Code commenting

Acquiring (*Adam, 5 min*)

- local and remote
- querying data
 - AGOL
 - iNaturalist
 - PRIMR web services
 - SQL query: IRIS warehouse

Processing and Analyzing (*Adam, 15 min*)

- Getting data into R
- QC
- Tidying data
- Visualizing
- EDA

Sharing (*McCrea, 10 min*)

- Reporting
 - RMarkdown
 - * Bat reporting for mobile acoustics
 - Dashboards
 - COVID 19 example
 - Shiny apps
 - * collarviewer
 - * power analysis for butterfly surveys

Archiving (*McCrea, 10 min*)

- Saving results to ServCat or some other data repository

An example R project / Live demo (*10 min*)

Questions (*10 min*)

Resources (Links)

Introduction to R

- An Introduction to R book

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 Programming for Research e-book

Style Guides - DataNovia R style guide

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
- An introduction to Docker for R users

Project Directory Templates

- MakeProject package
- rrtools package
- prodigenr package

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