Introduction to R for Ecological Data Analysis

2016 September 22-23

Welcome to Introduction to R for Ecological Data Analysis, a 2-day workshop at the University of New Brunswick taught by Dr. Lindsay Brin. This workshop is part of the Watershed and Aquatics Training in Environmental Research (WATER) program at the Canadian Rivers Institute, funded by an NSERC Collaborative Research and Training Experience (CREATE) grant.

The in-person workshop covers most of the lessons below, roughly in the order in which they are listed. Additional content is provided on this site as a reference. (For example: in the control structures lesson, we will cover if statements in class, but lesson materials for for loops are also provided here; the same is true for simple versus multiple regression, the latter of which we will refer to only briefly.)

The workshop builds off the Data Carpentry lessons on R for data analysis for Ecology, which are distributed under the CC-BY license.

Course details can be found below. Data for the workshop can be downloaded from github. This information can be accessed online at http://lindsaydbrin.github.io/CREATE_R_Workshop.

Workshop lessons outline

Introduction to R

- Before we start *(Data Carpentry)*
- Introduction to R *(Data Carpentry)*

Working with Data

- Starting with data *(Data Carpentry)*
- The `data.frame` class *(Data Carpentry)*
- The `data.frame` class - part 2 (Substitute for second half of Data Carpentry lesson)

Data manipulation

- Data manipulation using `dplyr` *(Data Carpentry)*
- Reshaping data frames: `dplyr` reshape
- Joining data frames: `dplyr` join

Pull it all together

- Day 1 Scripting Challenge
  - Day 1 Scripting Challenge Solution - *Don't peek until you've tried it yourself!*

Computing

- Control structures
- Writing functions

Data visualization

- Plotting with the base package
- Plotting with ggplot2 *(Data Carpentry)*

Statistical tests in R

- Linear regression: Simple and multiple
- Analysis of variance: One-way and two-way

Pull it all together

- Data Workflow Challenge
  - Data Workflow Challenge: One possible solution

Course details

Course overview

R is a powerful, free, open-source, and widely used tool for working with data that can allow you to streamline your analysis process, increase reproducibility, catch and avoid mistakes, and access newly developed analyses and statistical approaches.

This workshop will provide participants with the tools for reading in, manipulating, analyzing, and visualizing data in R. For example, we will cover topics such as cleaning and organizing data, joining data sets, working with factors, writing functions, using for loops and if. . . else statements, etc.

A primary goal is to help participants become comfortable working in a coding environment. To accomplish this, the workshop format will include a combination of explanation, demonstration, and extensive hands-on practice using a variety of sample datasets. Participants will leave with basic skills that can be put into use immediately, as well as a platform from which to learn more complex analyses.

Learning objectives

At the end of this 2-day workshop, participants will:

- Have a conceptual understanding of how R works
- Be aware of the availability of relevant packages to expand R's functionality
- Be able to use R as a calculator and basic data manipulation tool
- Be able to write simple functions in R to automate analyses
- Be able to produce useful graphical and numeric summaries of data
- Be able to implement basic statistical tests (ANOVA, linear regression)