

Introduction to Reproducible Workflows with R, Markdown, and GitHub

Date: Friday March 9, 2018

Location: Yale University, Center for Industrial Ecology

Details: In this one day workshop we will introduce the R programming language with a focus on tidy data, reproducibility, and version control with Git and Github. Attendees should leave with some basic skills in R, be able to create a reproducible research document in markdown, and know the basics of how to use git and GitHub. Our examples will focus on data visualization and examples from industrial ecology.

Set Up

For this workshop, we will be using R, RStudio, and Git and all attendees should have these installed and working prior to the workshop. The best instructions I know of for R and Git are the Software Carpentry workshop set up instructions so I will rely on those.

- Git: Software Carpentry Git Installation Directions
- R: Software Carpentry R Installation Directions

If you follow these directions you should have a working install of both Git and R. They do also point you to RStudio in the R set up instructions. If you followed along there you might have also installed RStudio. If you did not or need some more guidance here are some additional details.

- Go to the RStudio page for the Desktop downloads
- Select from the “Installers for Supported Platforms” for your operating system. The two most likely will be:
 - Windows 7/8/10 64-bit RStudio 1.1.423
 - Mac OS X 10.6+ 64-bit
- Using the installer, install RStudio how you normally would install any application on your computer.

Last thing to take care of is to make sure you have a GitHub account set up. The sign up is available at <https://github.com/>

Agenda (subject to change!)

Time	Subject
9:30 AM - 10:00 AM	Introductions and Checking Set Ups
9:45 AM - 10:00 AM	Motivationals
10:00 AM - 10:30 AM	The Punch Line
10:30 AM - 11:30 AM	R and RStudio Basics
11:30 AM - Noon	Markdown and R Markdown
Noon - 1:00 PM	LUNCH
1:00 PM - 2:00 PM	Tidy Data in R
2:00 PM - 3:00 PM	Data Visualization with <code>ggplot2</code>
3:00 PM - 4:00 PM	Git and GitHub
4:00 PM - 4:30 PM	Miscellaneous

Additional links

- Inspiration on how to use GitHub for a lab/research group
- Jeff Leek: Good stuff on a number of topics, all with an open science bent.
 - R Packages
 - Talks
 - Data Sharing
 - Genomics Papers
- Earth Lab at University of Colorado
- Related Links
- Industrial Ecology
- World Input/Output Datasets
- Industrial Ecology 2.0