Reproducible Research with R

Matthew K. Lau

June 7, 14 and 21 (Total = 4.5 hrs)

Why make your work reproducible and why R?

Science is driven by the exchange of information and knowledge. A recent study demonstrated that only 26% of studies published in the journal *Science* could be reproduced (see Stodden *et al.* 2018). This was even more striking given that the study was conducted after *Science* had instituted its open data policy. Luckily, advances in open-source computer languages, such as **R**, provide a way to produce computations that can more easily document scientific research in a transparent, easily shared way.

In this course, we will cover how to conduct **reproducible** scientific research using the **R** programming language and supporting software that will enable researchers to more clearly and easily document projects. Participants will gain experience coding in **R** using the *RStudio* IDE and using other software for reproducible research, such as the *git* version control system.

Course Outline

The course will be a mix of demos, discussions and activities:

Day 1

- Reproducibility Framework (10 min)
- *RStudio* tour (10 min)
- ACTIVITY: Example Project (15 min)
- Getting help with R (5 min)
- BREAK (10 min)
- Basic plotting and function anatomy (10 min)
- ACTIVITY: Write your own plot code (10 min)
- Q/A and tips

Day 2

- Project Architecture (10 min)
- ACTIVITY: Ecological data project (20 min)
- BREAK (10 min)
- Data wrangling and testing (15 min)
- ACTIVITY: Test those data (20 min)
- Q/A and tips

Day 3

- Style: Best Practices (5 min)
- "Backing Up" Version Control (20 min)

- ACTIVITY: Initiate git for your project (30 min)
- Linear models (5 min)
- ACTIVITY: Add a linear model and version your code (30 min)
- BREAK (10 min)
- Dependencies: R Packages, CRAN and *packrat* (10 min)
- ACTIVITY: Use packrat
- Q/A and tips

Before Class (15-20 min)

Download and install:

1. **R** https://cran.r-project.org/
2. *RStudio* https://www.rstudio.com/products/rstudio/download/#download/

The course notes can be found in the docs directory of my course repo: https://github.com/MKLau/reprosci/archive/master.zip

I also highly recommend the following cheat sheets for reference: https://www.rstudio.com/resources/cheatsheets/

Also, review the code of conduct.

Code of Conduct (aka. how to be a good community member)

Like science, open-source software development is empowered by community. Everyone participating in this course will follow the Code of Conduct outline by the folks at ROpenSci (https://ropensci.org/coc).

Be considerate and respectiful of each other in speech and actions

Contribute a safe and effective learning experience for everyone

We all get out of this class what we put into it

Advanced Topic Discussions via Slack (Depending on Interests)

RMarkdown, *dplyr*, *ggplot*, *github*, *Shiny Apps*, *R packages*, *Code performance with profVis*, *Data Provenance in R*

References

Stodden, V, Seiler, J and Z, Ma (2018) An empirical analysis of journal policy effectiveness for computational reproducibility. PNAS 115: 2584–2589. http://www.pnas.org/content/115/11/2584