Reproducible Research in RStudio
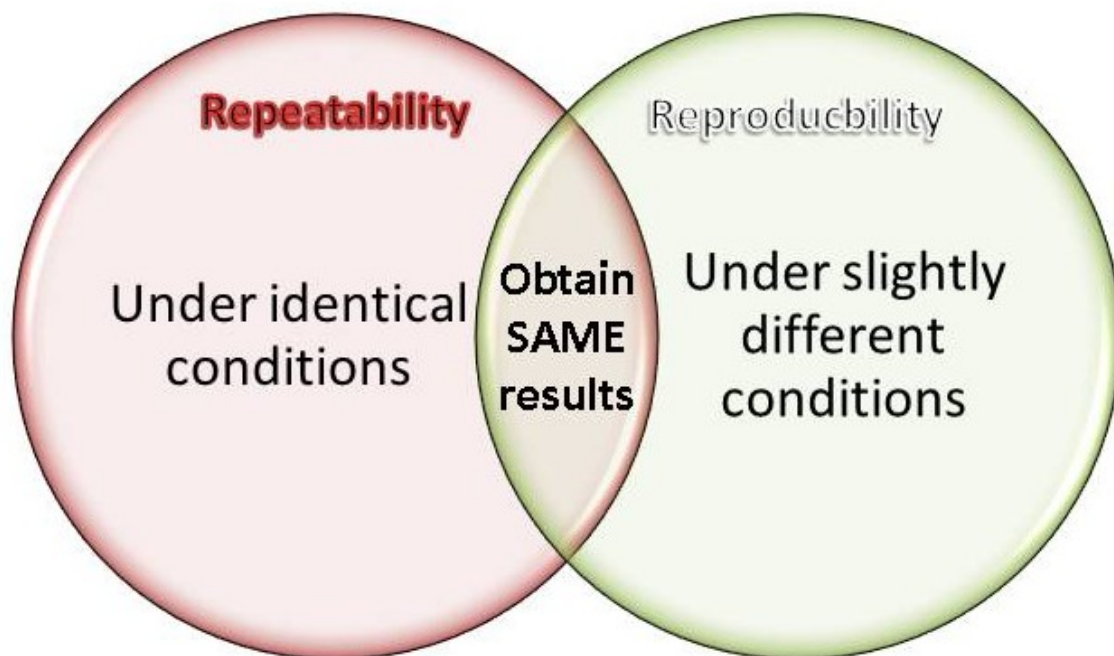
author: Caleb Kibet date: 10 August, 2018 autosize: true

Introduction to Reproducibility

"*Really reproducible research*" in computational sciences means:

> the data and code used to make a finding are available and they are sufficient for an independent researcher to recreate the finding

Reproducibility and replication

Why Should Research Be Reproducible?

1. It provides a standard to judge scientific claims
2. Reproducibility enhances replicability.
3. Helps avoid effort duplication & encourages cumulative knowledge development
4. Higher research impact for the researcher
5. Instils better work habits and teamwork

Future collaborator

[collaborator] (https://image.slidesharecdn.com/layton-repro-research-talk-2015-05-06-150507202403-lva1-app6892/95/reproducible-research-first-steps-6-638.jpg?cb=1431031632)

Tips for Reproducible Research

1. Document everything!
2. Everything is a (text) file.
3. All files should be human readable.
4. Explicitly tie your files together.
5. Have a plan to organize, store, and make your files available.
6. Report your research transparently

Keep track of things

1. Use version Control
2. Use proper documentation: README
3. Use Literate programming: RMarkdown, Jupyter Notebooks

Share and license your work

- Data: Adhere to FAIR data principles

- Software: Github and other repos

-

Project Folders

- Choose a file structure that works for you
- Use relative paths when possible and organize your files
    - Makes paths less dependent on particular File or System structure.
- Avoid putting spaces in your file and directory names
- Include a README.md that describes the purpose and structure of your project

Before you start

- Create project within a folder in your computer
- Create folder for your code
- Create folder for Data
- Raw: Downloaded or gathered from the field
- Derived: processed through your analysis

Continued

- Create a folder for figures generated from your analysis
- NB: Ensure separation of information

Literate Programming

Literate programming is a crucial part of a reproducible quantitative research. Being able to directly link your analyses, your results, and the code you used to produce the results makes tracing your steps much easier.

Quick Demo

A demo on: * Quick tour of RStudio * Creating a project * setting working directory * Creating folders

Version control in RStudio

- RStudio has inbuilt version control support
- Learn more here

RMarkdown

> R Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents.

- Uses Knitr to execute the embedded code and pandoc to convert to output
- The flow is:

RMarkdown (https://d33wubrfki0l68.cloudfront.net/61d189fd9cdf955058415d3e1b28dd60e1bd7c9b/9791d/images/rmarkdownflow.png)

RMarkdown Output

[Output formats] (https://d33wubrfki0l68.cloudfront.net/00ed9c32053cbc805efa51b66be570558480a4c8/7a292/images/rmarkdownoutputformats.png)

Let's hop into RStudio's tutorial to learn more

RMarkdown Tutorial

RMarkdown CheatSheet

RNotebooks

An R Notebook is an R Markdown document with chunks that can be executed independently and interactively, with output visible immediately beneath the input

- RNotebook Tutorial

RShiny

Shiny is an R package that makes it easy to build interactive web applications (apps) straight from R.

Learn more at you own time

Reference

1. Reproducible Research with R and RStudio Second Edition is a great reference text.
2. https://rmarkdown.rstudio.com/lesson-11.html
3. https://rmarkdown.rstudio.com/articles_intro.html
4. Organizing Projects: http://kbroman.org/Tools4RR/assets/lectures/06_org_eda.pdf
5. Research reproducibility: http://stm.sciencemag.org/content/8/341/341ps12.full