

## **GIT & GITHUB**

Version Control as a tool for Data Management

John R Little

Duke University

January 13, 2023

CC BY 4.0 John R Little



### **GIT**

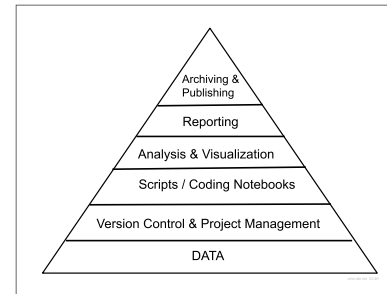
- Git is one of the most popular. There are other version control systems.
- Free open-source application
- Scalable - works with any size project
  - Linux is a world-wide project, spanning time-zones and languages
- Git is a file repository, or repo - repositories exist for many reasons
- Git and GitHub are not the same thing. Symbiotic cousins

CC BY 4.0 John R Little



## **VERSION CONTROL**

- Manage how computers files change over time



CC BY 4.0 John R Little



### **TRACK PROJECT DIRECTORIES**

- Tell the story of your project
- Who, What, and When files change. (and Why, How)

CC BY 4.0 John R Little



## JARGON

**Commit** - ascribe the WWWW values to the storyline

- In this way we
  - Can tell the story of our project
  - Time-travel – rolling-back to an earlier state of the entire project

See Also: Push; Pull; Pull-Request; Remote; Branch; et. al.

CC BY 4.0 John R Little



## FILE TYPES

- Any computer file
- In practice: text documents and scripts work best
  - HTML, CSS, Markdown, SVG files, CSV files,
- Voluminous files
  - Can be managed with adjacent processes (databases, or other)
  - Administrative overhead is always a concern: KISS
- Final documents, posters etc
  - Although, reproducible best practice implies deriving reports from code.

CC BY 4.0 John R Little



## EXPERIMENS

- Through a feature called a branch
- Make replica copies of the git repo
- Alter and experiment
- Merge (or delete) the branch
- dev/prod

CC BY 4.0 John R Little



## PROFILE AND SHARE

- Remotes are shared/distributed via Hubs (GitHub, GitLab, BitBucket)
  - Distribution, Collaboration, Project Management, Pull-Request, etc.
- Other hubs already mentioned
  - but also Duke's instance: [gitlab.oit.duke.edu](https://gitlab.oit.duke.edu)

CC BY 4.0 John R Little



## GITHUB

- A place to store and host git repositories
- More than a repo host
  - VMs
  - Project management
    - Access control, task lists, bug tracking, wiki documentation, Kanban planning

## GITHUB - MINI CASE STUDY

- Clone my repo (or Fork) [difference? permissions!]
- My repo is public
- Code collaboration via Pull-Requests
- Repo editor resolves PRs
- Via version (or releases). Can synchronize **file repo** with **archival repo** for milestones such as an accepted article

CC BY 4.0 John R Little



CC BY 4.0 John R Little



## FIN

John R Little  
Data Science Librarian  
Center for Data & Visualization Sciences  
Duke University Libraries

<https://JohnLittle.info> • <https://Rfun.library.duke.edu> • <https://Library.duke.edu/data>

Creative Commons: Attribution 4.0 • <https://creativecommons.org/licenses/by-nc/4.0>

## Error

×

CC BY 4.0 John R Little

