

# Reproducible Workflows

12 February 2020

*Modern Research Methods*

# Assignments

Assignment 2: Grades posted today

Assignment 3: Due tomorrow at noon

Office Hours:

Jaeah 1:00-3:00pm Monday

Molly 4:30-6:30pm **today** (Porter 223A)

# Last Time: Reproducibility

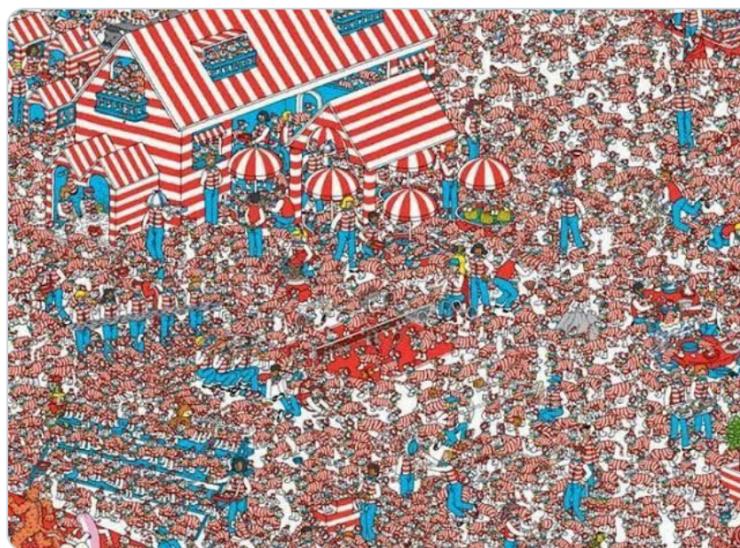


Guy Prochilo 🏳️‍🌈  
@GuyProchilo

"Thank you for your interest in our paper. Please find attached the code for reproducing our findings"

The code

#phdchat #rstats



2:26 PM · Feb 5, 2020 · Buffer

5 Retweets 40 Likes



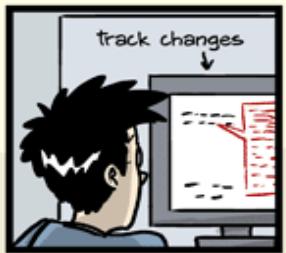
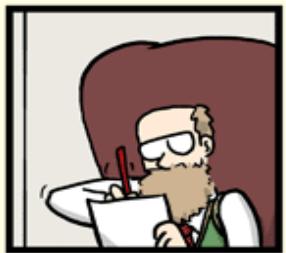
# The problem

- You should be able to hand your data to **someone else** and they should be able to reproduce your analyses.
- You should be able to reproduce the same analysis you produced last week again.
  - When you're working a paper, or analysis, or collecting data, it's nice to be able to go back to see what you did at previous point in time.
- Two goals:
  - Sharing: Make data/code/analysis publicly available
  - Version Control: Way to keep track of versions

# Possible solutions to sharing

- Email
- Texting
- whatsapp
- Google docs
- Google drive
- Usb drives
- Paper
- Verbally??

# "FINAL".doc



Version control allows you to:

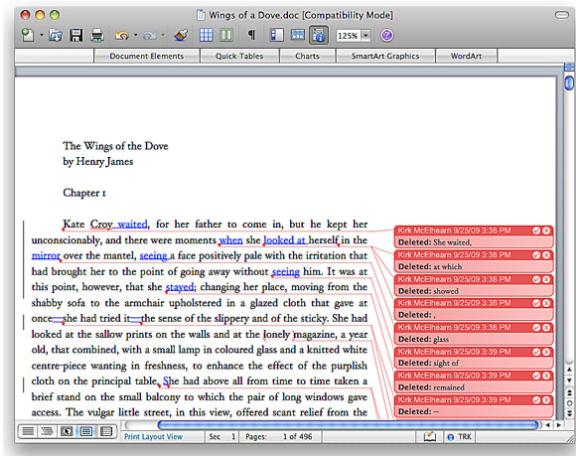
- See history of changes
- Go back to old version

# One Solution:

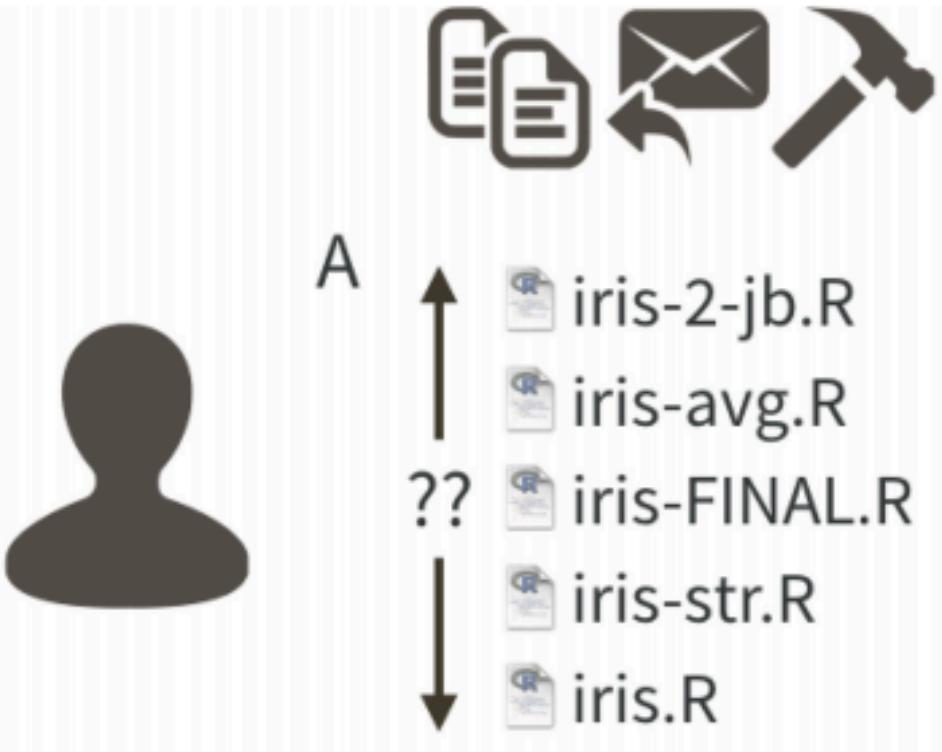


- <https://osf.io/>
- Free and open source project management tool that supports researchers throughout their entire project lifecycle.
- This solves the sharing problem and sort of solves the version control problem
- Easy to use
- Interface clunky, not great version control

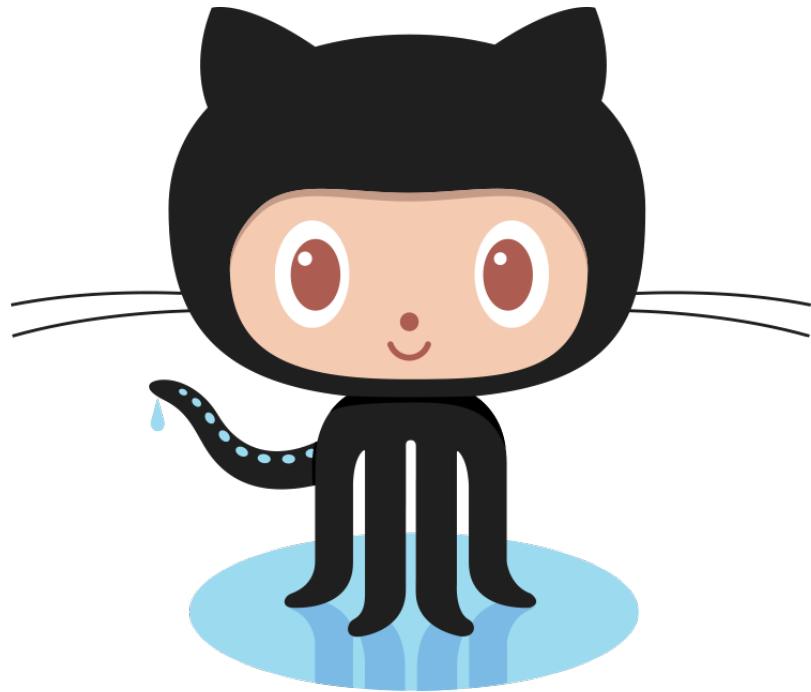
# A more sophisticated solution: Git



- Git is a free and open source software
- Version control system
- Original purpose was to help groups of developers work collaboratively on big software projects.
- Like track changes in Microsoft Word/Google Docs
- Repurposed by data science/science community



# GitHub != Git (but they're similar)



- GitHub is a **web-based hosting service** for version control using Git (like Dropbox or Google Drive).
- Has many extra features designed to improve how people collaborate.

# Repository as the unit of storage

- “repo”
- In our context, one repository = one project
- Repo contains **all** the files associated with a project.





= class project



= software  
package

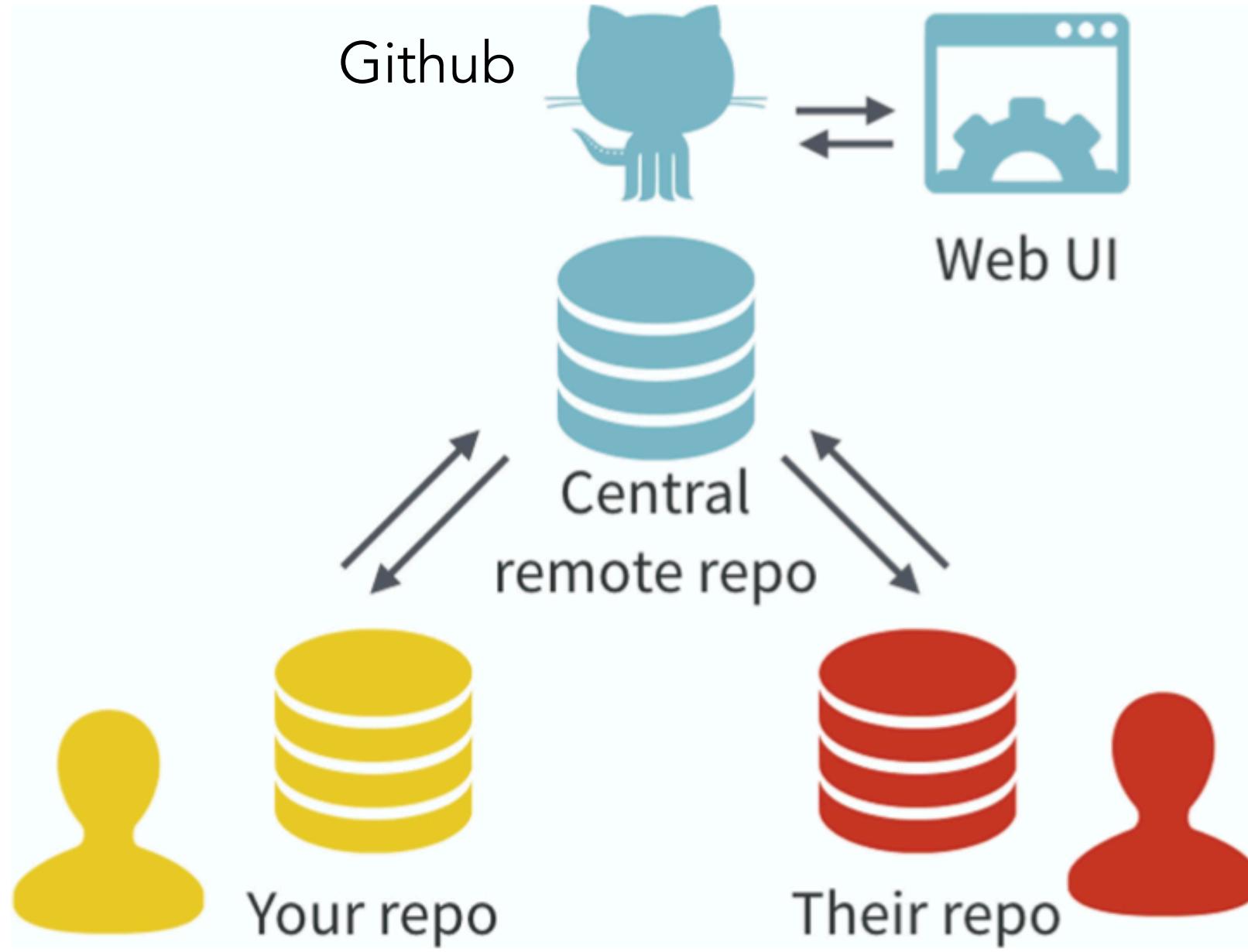


= paper



= code for a  
website

e.g., <https://github.com/mllewis/cumulative-science>



# The workflow

Your Working  
Directory



Staging Area



Local Repo



Central Repo on  
Github



**add**



**commit**



**push**



**rm**



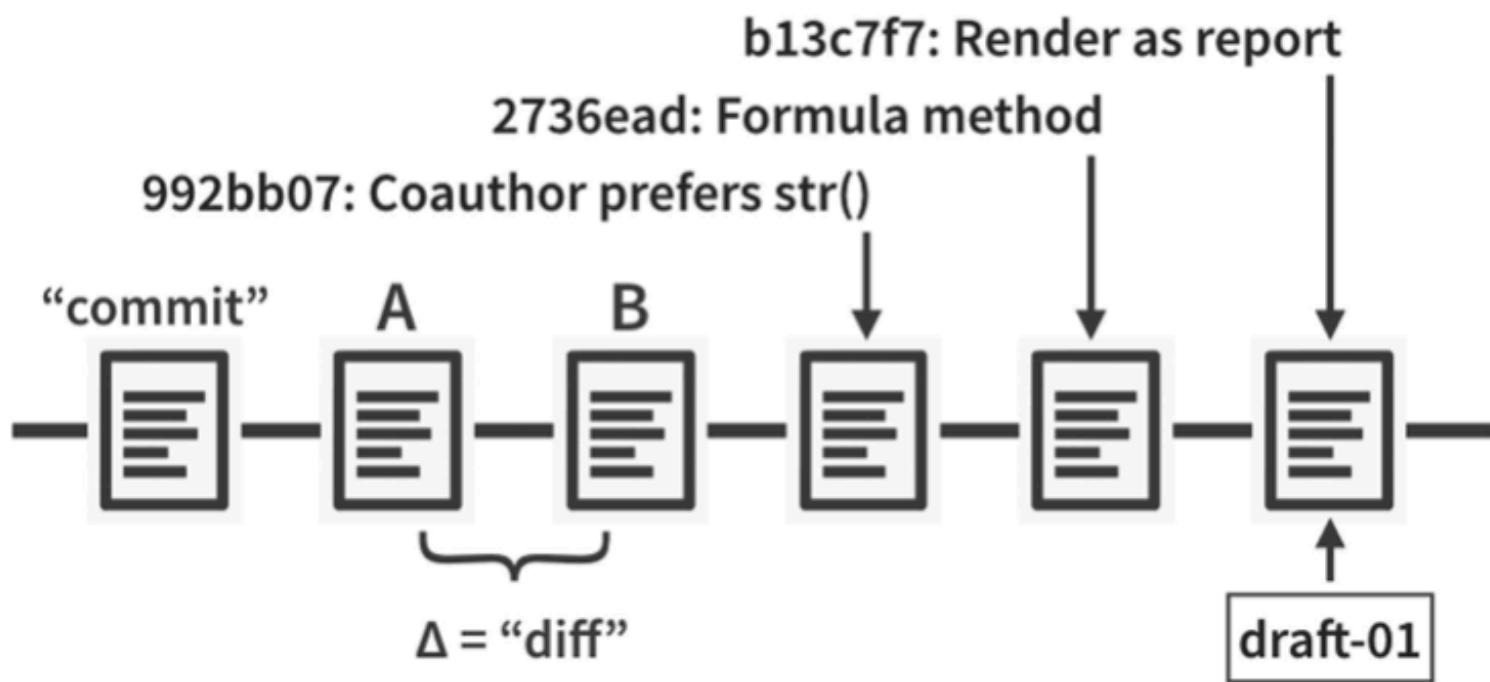
**pull**



**clone (first time)**



# Commit is like a “snapshot” of your file at a point in time



You should include an informative message when you make a commit.

You can always go back and see what your project looked like at different points in your commit history.

# How do you interact with git?

- At the command line in Terminal:

**git <VERB> <ARGS>**

- Already installed on Rstudio cloud



File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins R 3.6.0

01\_MRM\_assignment\_template.Rmd

Knit Insert Run

```
1 ---  
2 title: "Assignment 1: Cumulative Science and Intro to dplyr"  
3 subtitle: "Modern Research Methods"  
4 author: "Molly Lewis"  
5 date: `r format(Sys.time(), '%d %B %Y')`  
6 output:  
7   html_document:  
8     highlight: kate  
9     theme: cosmo  
10 ---  
11
```

155:32 Chunk 21 R Markdown

Console Terminal Jobs

Terminal 1 /cloud/project

```
rstudio-user@application-1697021-deployment-4826933-s56lv:/cloud/project$ git add my_file.rmd
```

Environment History Import Dataset Global Environment Data lf\_da... 600 obs. of... Files Plots Package New Folder Upload Cloud > project ... .Rhistory 01\_MRM\_assignmen... data project.Rproj

# Explore the complexity project repo on Github

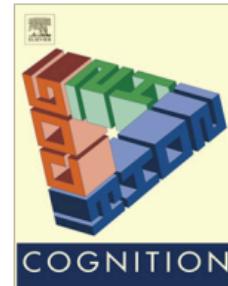


ELSEVIER

Contents lists available at [ScienceDirect](#)

Cognition

journal homepage: [www.elsevier.com/locate/COGNIT](http://www.elsevier.com/locate/COGNIT)



Original Articles

The length of words reflects their conceptual complexity

Molly L. Lewis\*, Michael C. Frank

*Department of Psychology, Stanford University, United States*



CrossMark

<https://github.com/mllewis/RC>

Pair up in groups of 2-3 with someone who has a laptop

<https://github.com/mllewis/RC>

Try to answer the following questions:

1. What is the directory structure of the repo?
2. How many commits are there?
3. What was the first commit message on Aug 21, 2015
4. Can you find the object to the right? (obj\_15.jpg)
5. How many collaborators were there on this project?

