



GitHub 101

Nathalie Sommer

PhD Student

Yale School of Forestry and Environmental Studies

October 25th 2019

Today:



1. Why GitHub?
2. What is GitHub? What is Git?
3. GitHub 101: Definitions
4. GitHub & RStudio
5. GitHub, RStudio, and your master's thesis
6. Tutorial via Our Coding Club

Why GitHub?



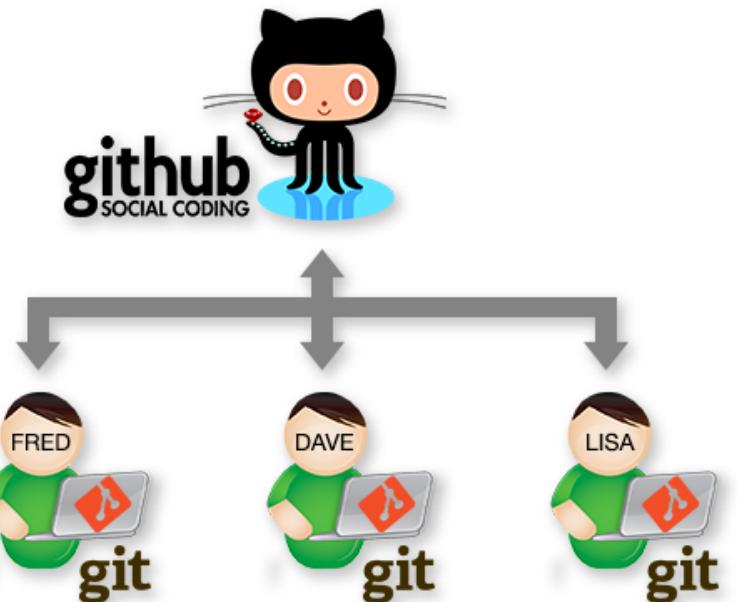
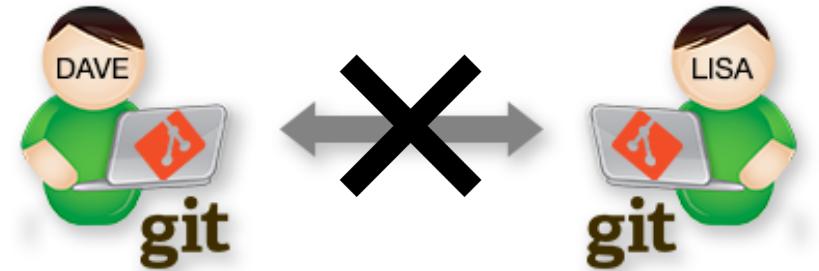
- Version control
- Back-up for your code and data
- Collaborative coding
- Open science

→ Data management
→ Reproducibility

Draft_31Oct
Draft_edit
Draft_FINAL
Draft_FINALFINAL
Draft_TheREALfinal
Draft_FinalButSendHelp

What is GitHub?

- Git = local version control
- GitHub = Version control + repository + collaboration
 - Analogy: the edit history of a google doc within a Google Drive
- You don't need to know the language of git to use GitHub
- GitHub's version control works for any programming platform / file types



Overview

Repositories 225

Projects 0

Stars 114

Followers 17k

Following 6



Hadley Wickham

hadley

Follow

★ PRO

Chief Scientist at @rstudio

@rstudio

Houston, TX

hadley@rstudio.com

<http://hadley.nz>

Block or report user

Organizations



Pinned

tidyverse/ggplot2

An implementation of the Grammar of Graphics in R

• R ★ 4.1k ⚡ 1.5k

tidyverse/dplyr

dplyr: A grammar of data manipulation

• R ★ 3.1k ⚡ 1.1k

tidyverse/tidyverse

Easily install and load packages from the tidyverse

• R ★ 766 ⚡ 161

r4ds

R for data science: a book

• R ★ 2.1k ⚡ 2.7k

r-lib/devtools

Tools to make an R developer's life easier

• R ★ 1.9k ⚡ 665

adv-r

Advanced R: a book

• TeX ★ 1.6k ⚡ 1.5k

4,224 contributions in the last year

2019

2018

2017

2016

2015

2014

2013

2012

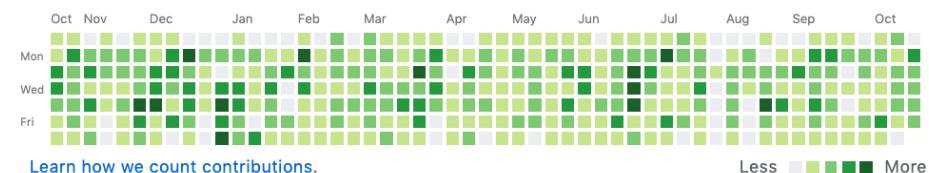
2011

2010

2009

2008

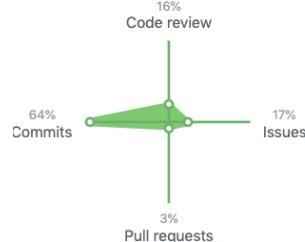
2007



@rstudio @tidyverse @r-lib More

Activity overview

Contributed to hadley/adv-r, tidyverse/tidyr, r-lib/vctrs and 5 other repositories

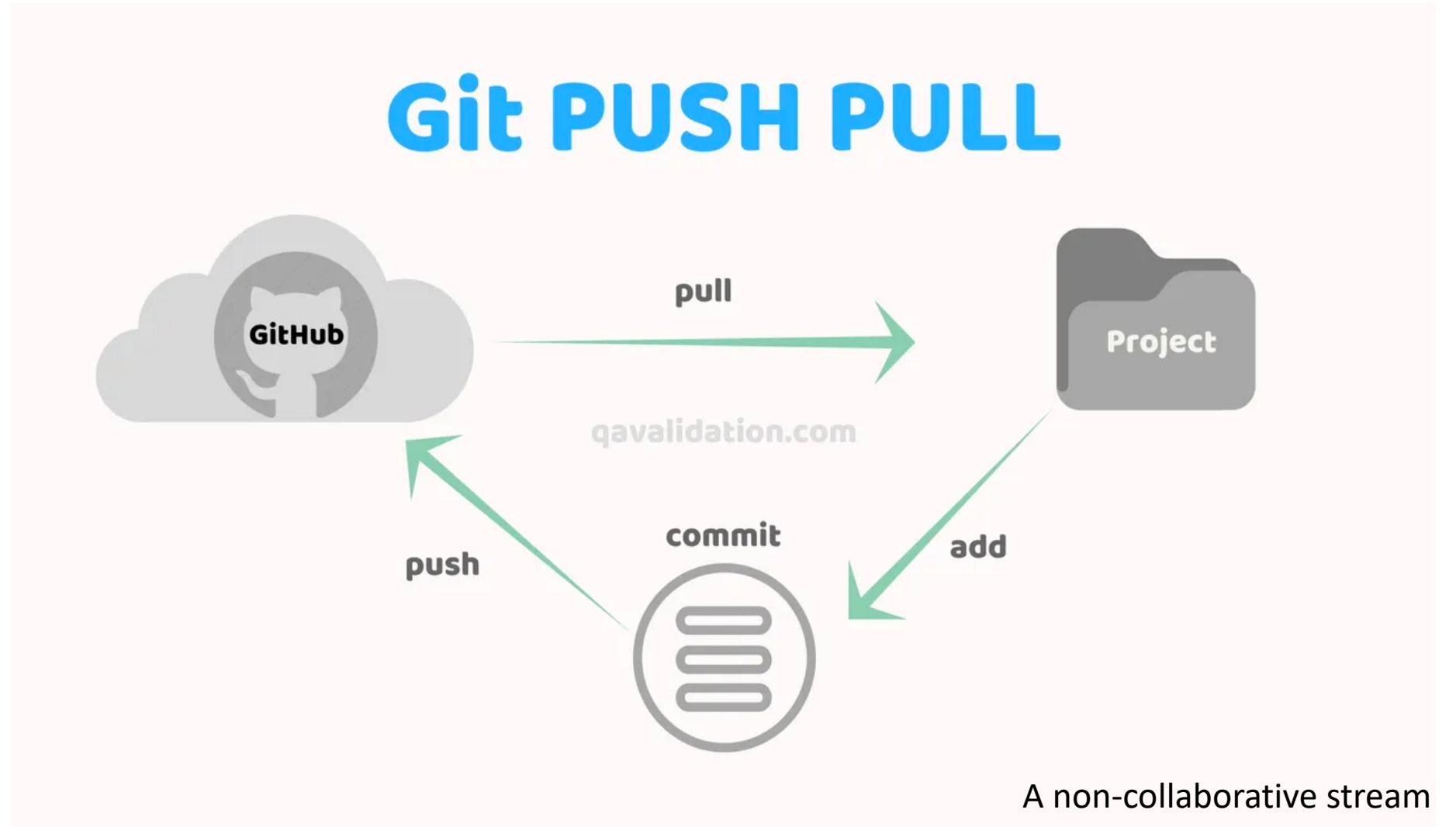


Contribution activity

October 2019

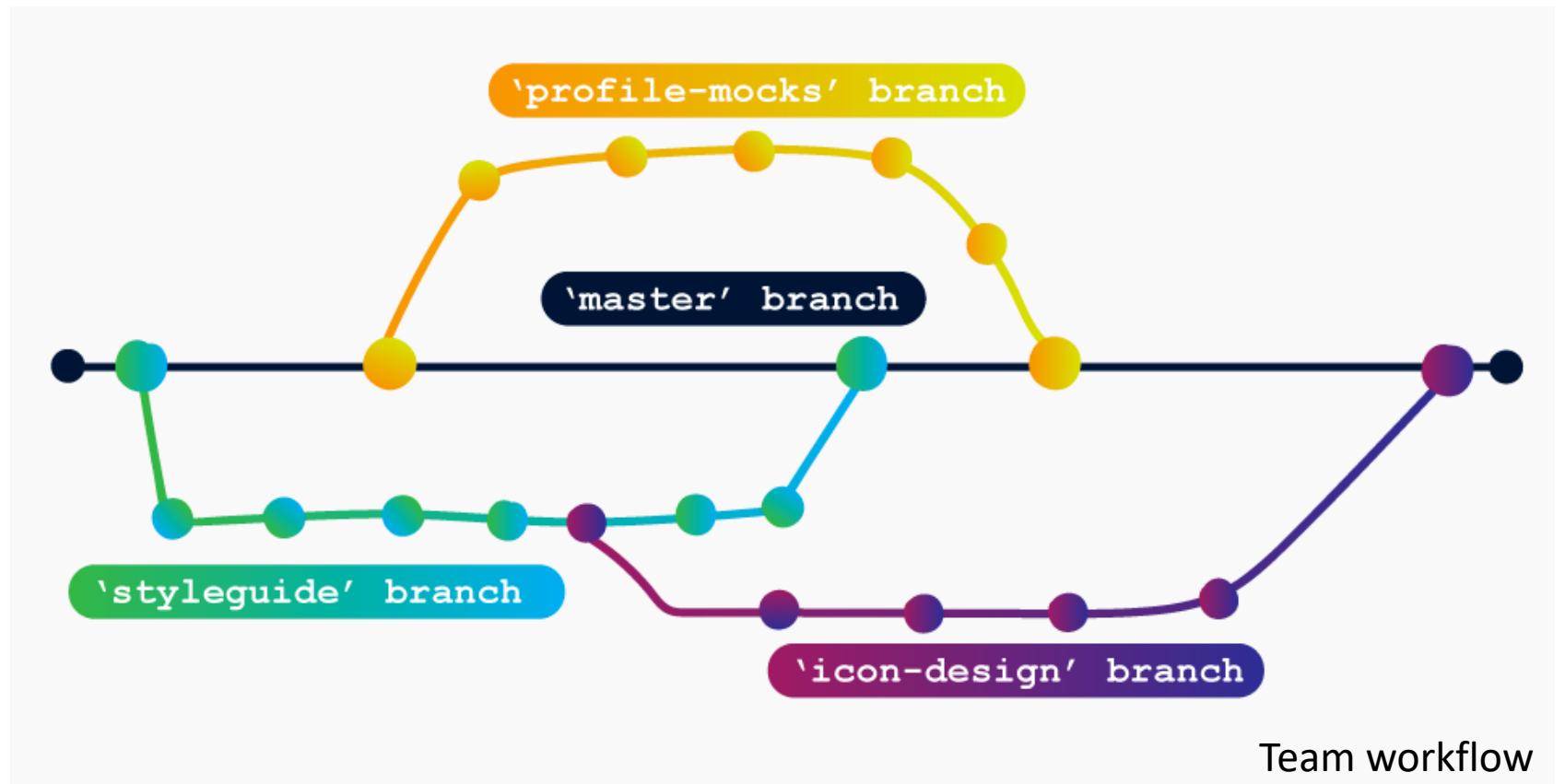
GitHub 101

- Pull
- Commit
- Push



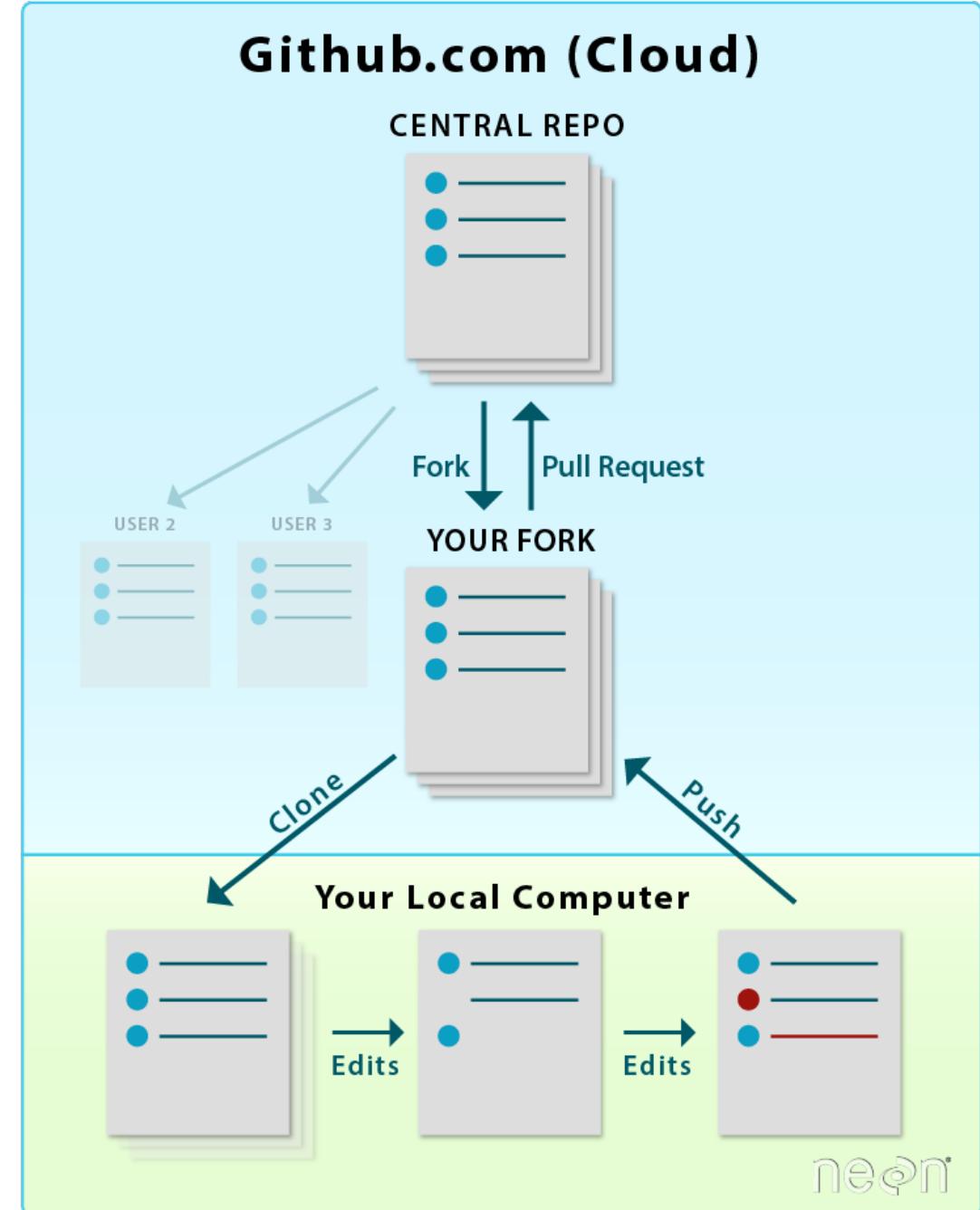
GitHub 101

- Branches
 - Master
 - Fork
 - → others
- Pull request



GitHub 101

- Pull is to pull from upstream
 - To catch yourself up with any changes that may have been made
- To fork and clone and then make a pull request is to “branch” from the master and continue working on something, even as other commits are being made to the master
- Changes are three steps: Commit, Stage, Push
- If you ever get confused, go back to the idea of this as a stream.

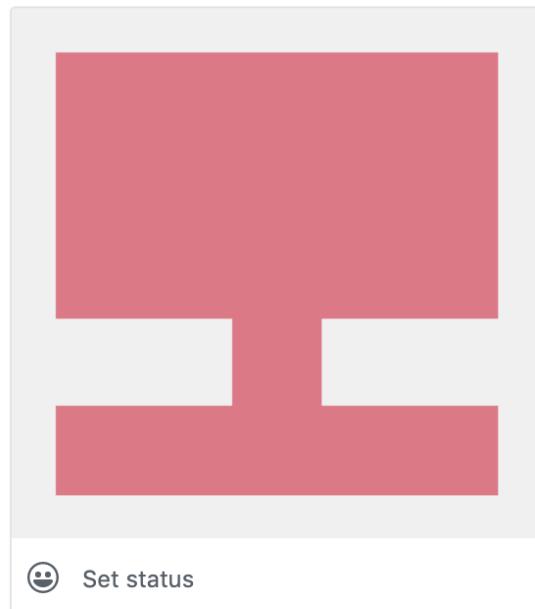


GitHub and your master's thesis

- For each thesis question* (read: publishable unit)
 - Establish a Project Folder
 - Establish a GitHub repo
- Make commits early and often
- Use effective commit messages
- Practice!
 - Use dummy repos and datasets until you get more comfortable

* If you are working with human subjects or sensitive data, you'll need to have a data management plan that may not include GitHub.

Establishing your repo



Overview **Repositories 1** Projects 0 Packages 0 Stars 0 Followers 0 Following 0

Find a repository... Type: All ▾ Language: All ▾ **New**

Personality-and-trophic-cascades ★ Star

Repo for ms: Herbivore prey personality mediates trophic cascades by NR Sommer and OJ Schmitz

● R Updated 18 days ago

nathaliesommer

Edit profile

Create a new repository

A repository contains all project files, including the revision history. Already have a project repository elsewhere? [Import a repository](#).

Name your repo

Owner



Repository name *

nathaliesommer / Best-ever-repo ✓



Great repository names are short and memorable. Need inspiration? How about [expert-winner](#)?

Description (optional)

Public

Anyone can see this repository. You choose who can commit.

Private

You choose who can see and commit to this repository.

Skip this step if you're importing an existing repository.

Initialize this repository with a README

This will let you immediately clone the repository to your computer.

Add .gitignore: None ▾

Add a license: None ▾



Create repository

Create a new repository

A repository contains all project files, including the revision history. Already have a project repository elsewhere? [Import a repository](#).

Owner

 nathaliesommer ▾ / Best-ever-repo ✓

Great repository names are short and memorable. Need inspiration? How about [expert-winner](#)?

Description (optional)

 Public

Anyone can see this repository. You choose who can commit.

 Private

You choose who can see and commit to this repository.

Make it private



Skip this step if you're importing an existing repository.

Initialize this repository with a README

This will let you immediately clone the repository to your computer.

Add .gitignore: None ▾

Add a license: None ▾



Create repository

Create a new repository

A repository contains all project files, including the revision history. Already have a project repository elsewhere? [Import a repository](#).

Owner

 nathaliesommer ▾ / Best-ever-repo ✓

Great repository names are short and memorable. Need inspiration? How about [expert-winner](#)?

Description (optional)

 **Public**

Anyone can see this repository. You choose who can commit.

 **Private**

You choose who can see and commit to this repository.

Skip this step if you're importing an existing repository.

Initialize this repository with a README

This will let you immediately clone the repository to your computer.

Initialize with a README

Add .gitignore: None ▾

Add a license: None ▾



Create repository

Create a new repository

A repository contains all project files, including the revision history. Already have a project repository elsewhere? [Import a repository](#).

Owner

 nathaliesommer ▾ / Best-ever-repo ✓

Great repository names are short and memorable. Need inspiration? How about [expert-winner](#)?

Description (optional)

 **Public**

Anyone can see this repository. You choose who can commit.

 **Private**

You choose who can see and commit to this repository.

Skip this step if you're importing an existing repository.

Initialize this repository with a README

This will let you immediately clone the repository to your computer.

Add .gitignore: None ▾

Add a license: None ▾ 

Create repository

Add a .gitignore

Collecting data

- As your data come pouring in...
 - 1. Consistently back up your computer
 - 2. Upload files to GitHub
- Be descriptive
 - With every commit, GitHub requires a comment / description
- Merge data using code, not in Excel
 - Lowers likelihood of user error
 - Problem emerge quickly when working with multiple collaborators in Excel

The screenshot shows a GitHub repository page for 'robertwbuchkowski / fear_of_predation'. The repository is private and contains a single branch named 'master'. The 'Data' folder is selected. The list of files shows numerous CSV files uploaded between July and September 2019. Most files are named with dates and descriptions, such as '10Sep2019.csv', '19July2019_crickets.csv', and 'IDtable_10Sep2019.csv'. The commits are mostly from a user named 'nathaliesommer'. The repository has 0 stars, 0 forks, and 0 issues. The last commit was made on Sep 13, 2019.

File	Upload Date	Last Commit
10Sep2019.csv	10 Sep 2019	13 Sep 2019
11Sep2019.csv	11 Sep 2019	13 Sep 2019
12Sep2019.csv	12 Sep 2019	13 Sep 2019
19July2019_crickets.csv	Cricket respiration from 19 July 2019.	13 Sep 2019
1Aug2019.csv	Data from 1 Aug. Includes ONAS G-J	13 Sep 2019
20Aug2019.csv	20 Aug 2019	13 Sep 2019
20July2019_crickets.csv	Cricket respiration from 20 July 2019	13 Sep 2019
21Aug2019.csv	21 Aug 2019	13 Sep 2019
21July2019_crickets.csv	Cricket respiration from 21 July 2019	13 Sep 2019
22July2019_isopod_fear.csv	Final analysis of cricket and isopod runs.	13 Sep 2019
23Aug2019.csv	23 Aug 2019	13 Sep 2019
24Aug2019.csv	24Aug2019	13 Sep 2019
26July2019_spiders.csv	Add files via upload	13 Sep 2019
28Aug2019.csv	28Aug2019	13 Sep 2019
29Aug2019.csv	29 Aug 2019	13 Sep 2019
29July2019.csv	Cricket respiration from 29 July	13 Sep 2019
30July2019.csv	Data from 30 July 2019. Includes crickets and TRRA	13 Sep 2019
31July2019.csv	Data from 31 July. Includes cricket and ONAS	13 Sep 2019
3Sep2019.csv	3 Sep 2019	13 Sep 2019
4Sep2019.csv	4 Sep 2019	13 Sep 2019
5Sep2019.csv	5 Sep 2019	13 Sep 2019
7Sep2019.csv	7 Sep 2019	13 Sep 2019
8Sep2019.csv	8 Sep 2019	13 Sep 2019
9Sep2019.csv	9 Sep 2019	13 Sep 2019
IDtable_10Sep2019.csv	10 Sep 2019	13 Sep 2019
IDtable_11Sep2019.csv	11 Sep 2019	13 Sep 2019
IDtable_12Sep2019.csv	12 Sep 2019	13 Sep 2019
IDtable_13Sep2019.csv	13 Sep 2019	13 Sep 2019
IDtable_19July2019_crickets.csv	Cricket respiration from 19 July 2019.	13 Sep 2019
IDtable_1Aug2019.csv	Data from 1 Aug. Includes ONAS G-J	13 Sep 2019
IDtable_20Aug2019.csv	20 Aug 2019	13 Sep 2019
IDtable_20July2019_crickets.csv	Cricket respiration from 20 July 2019	13 Sep 2019
IDtable_21Aug2019.csv	21 Aug 2019	13 Sep 2019
IDtable_21July2019_crickets.csv	Cricket respiration from 21 July 2019	13 Sep 2019
IDtable_22July2019.csv	New data on spiders from July 27 2019.	13 Sep 2019
IDtable_23Aug2019.csv	23 Aug 2019	13 Sep 2019
IDtable_24Aug2019.csv	24Aug2019	13 Sep 2019
IDtable_26July2019.csv	Rename IDtable26July2019.csv to IDtable_26July2019.csv	13 Sep 2019
IDtable_28Aug2019.csv	28Aug2019	13 Sep 2019

GitHub and RStudio

A screenshot of a GitHub profile page for a user named nathaliesommer. The page shows one repository: 'Personality-and-trophic-cascades'. The repository description is: 'Repo for ms: Herbivore prey personality mediates trophic cascades by NR Sommer and OJ Schmitz'. It was updated 18 days ago. The page includes navigation links for Overview, Repositories (1), Projects (0), Packages (0), Stars (0), Followers (0), and Following (0). There are also search and filter options.

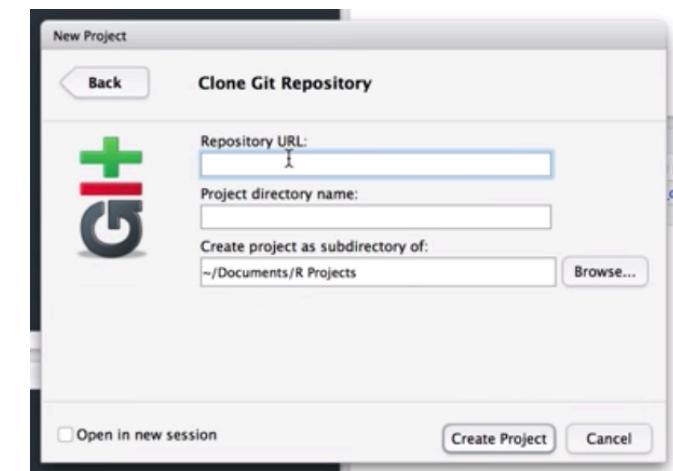
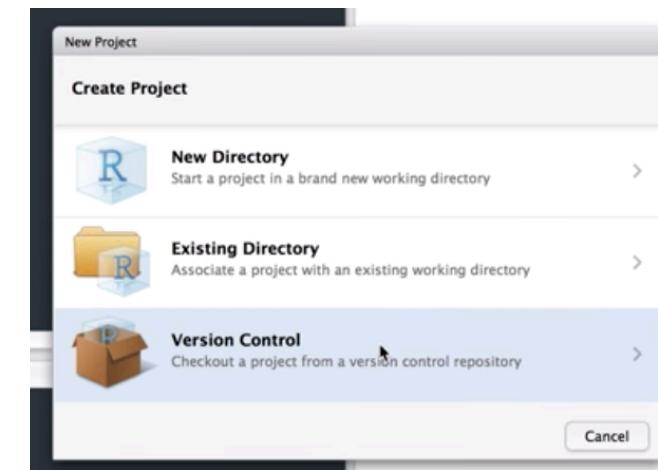
Create your repo



Add some data



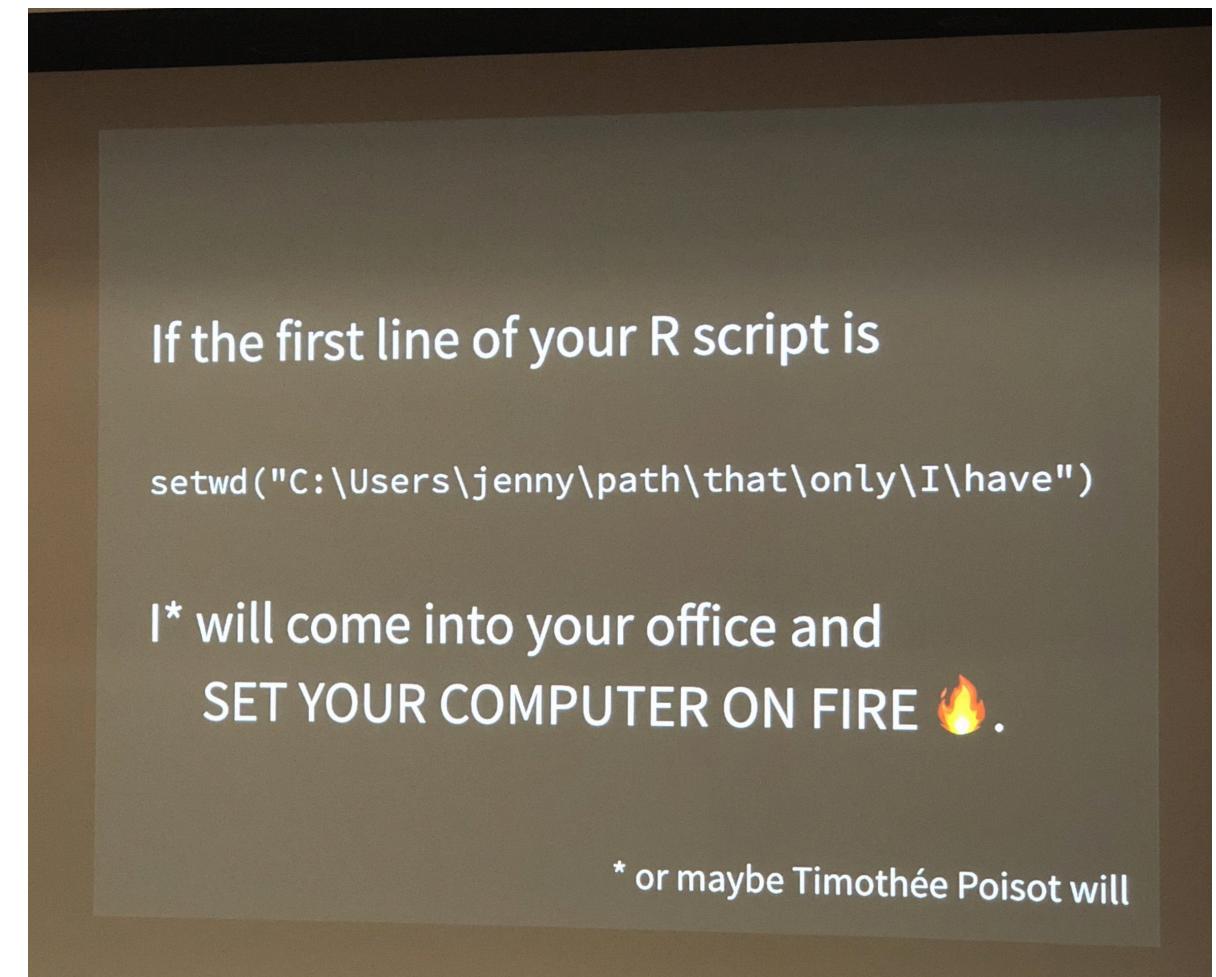
Create an RProject



Link your repository

An aside on Rprojs and your workflow

- Always *always* assume people will be interested in your data and code
- Use Projects!!
 - (even if you aren't using GitHub)
- Projects keep all your files together and they are designed for version control (Git)
- A lot of R packages are developed on GitHub

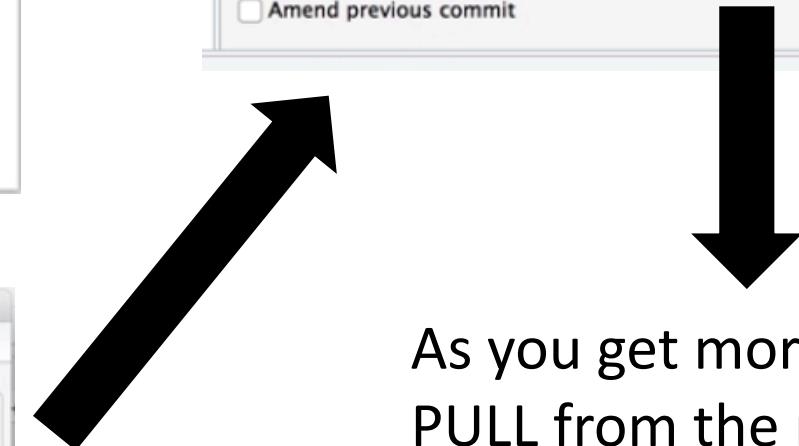
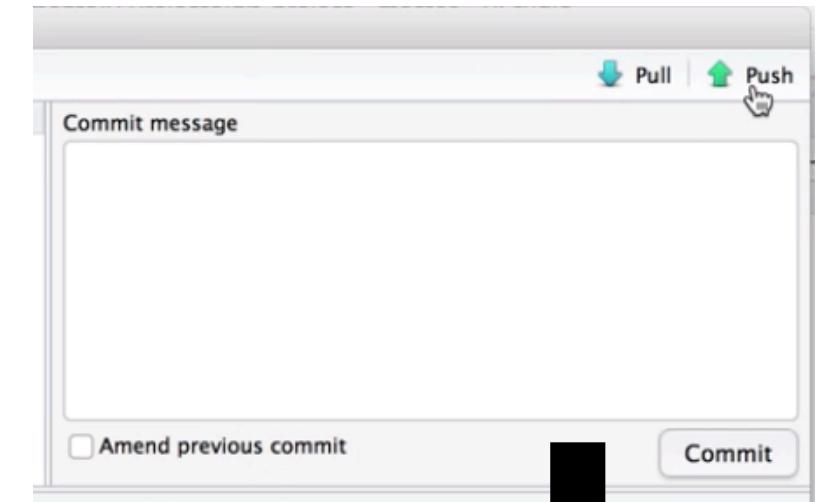
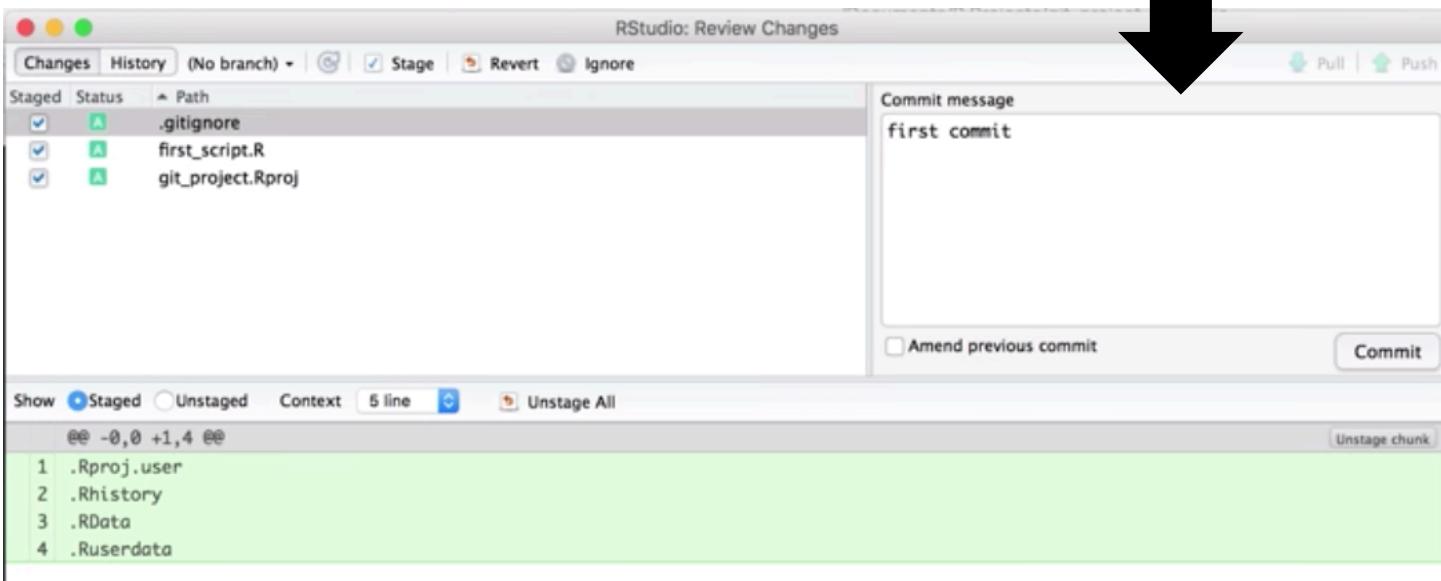
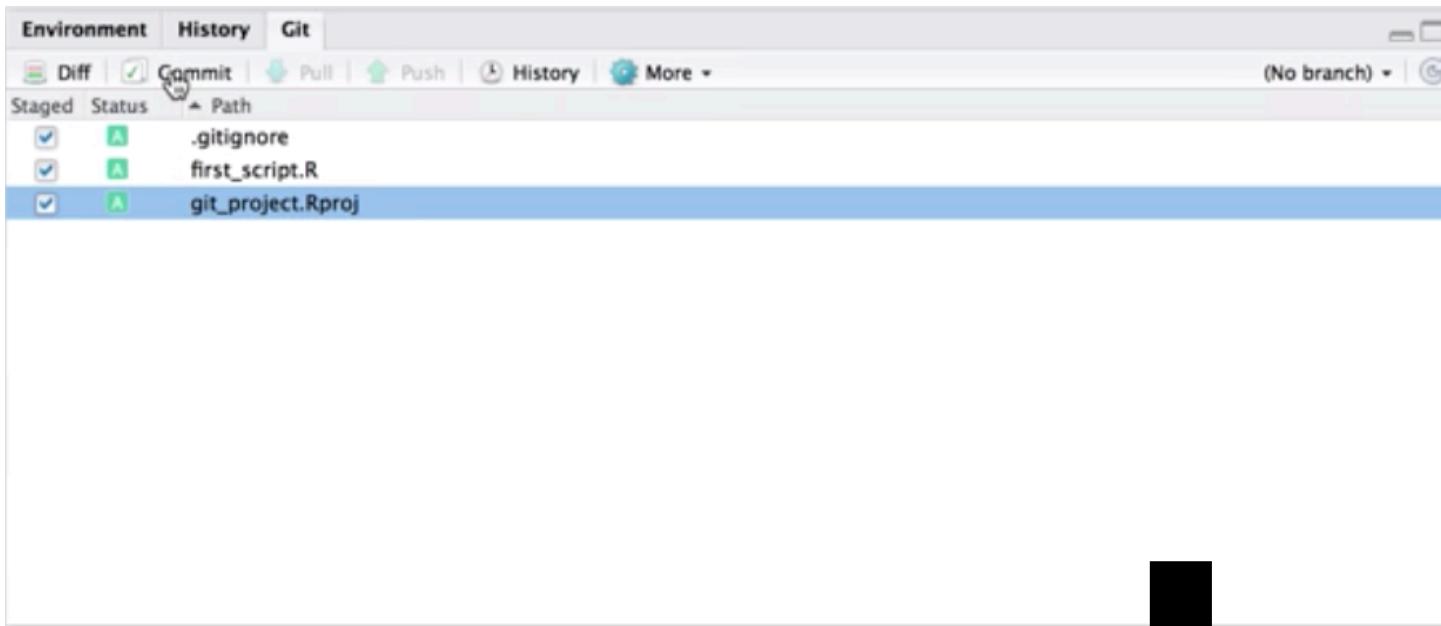


The screenshot shows the RStudio interface with a project titled "git_project".

- Top Bar:** Includes icons for file operations (New, Open, Save, Print), Go to file/function, Addins, and a project dropdown labeled "git_project".
- Environment Tab:** Shows the Global Environment pane, which is currently empty.
- Git Tab:** The "Git" tab is highlighted with a red circle.
- File Explorer:** Shows the project structure:
 - New Folder
 - Delete
 - Rename
 - More

Home > Documents > R Projects > git_project

Name	Size	Modified
..		
.gitignore	40 B	Jan 16, 2017, 1:39 PM
git_project.Rproj	205 B	Jan 16, 2017, 1:39 PM
first_script.R	0 B	Jan 16, 2017, 1:39 PM
- Console:** Displays the R startup message and license information.
- R Script:** Shows the content of the "first_script.R" file.



As you get more data,
PULL from the master
branch to “update”
your local folder

Submitting your paper(s)

- You will make mistakes during your analysis or even change directions with the question.
- Before sending your repo to reviewers, “clean” your commits
- Either:
 - Make an entirely new repo with only the initial commit
 - Or, create an orphan branch
- Make your repo public

Acceptance



zenodo

The Zenodo logo, consisting of the word "zenodo" in a white, lowercase, sans-serif font, centered on a solid blue rectangular background.

- Archive the repository to Zenodo (using your GitHub account)
 - Get DOI for your data + code
 - Make your data + code open access

Why GitHub!



- Data and code stay ~safe~
 - GitHub as your back-up backup
- Ease of open science
 - Journals require data are uploaded to Dryad, at a minimum
 - Why not go ahead and give them your code as well!
- Ease of workflow
- Collaboration
 - Way simpler to ask others for help!
- Even more reasons!
 - You can make a website in GitHub
 - It's free
 - You can link to your GitHub as proof of your coding skills / commitment to open science

Tutorial – Our Coding Club

- **Install**
 - Rstudio (rstudio.com/products/rstudio/download)
 - Base R
 - Git <https://git-scm.com/downloads>
- **Make GitHub account**
 - Use a permanent email address, not an institutional one