

# Workflow basics, Markdown, git/Github

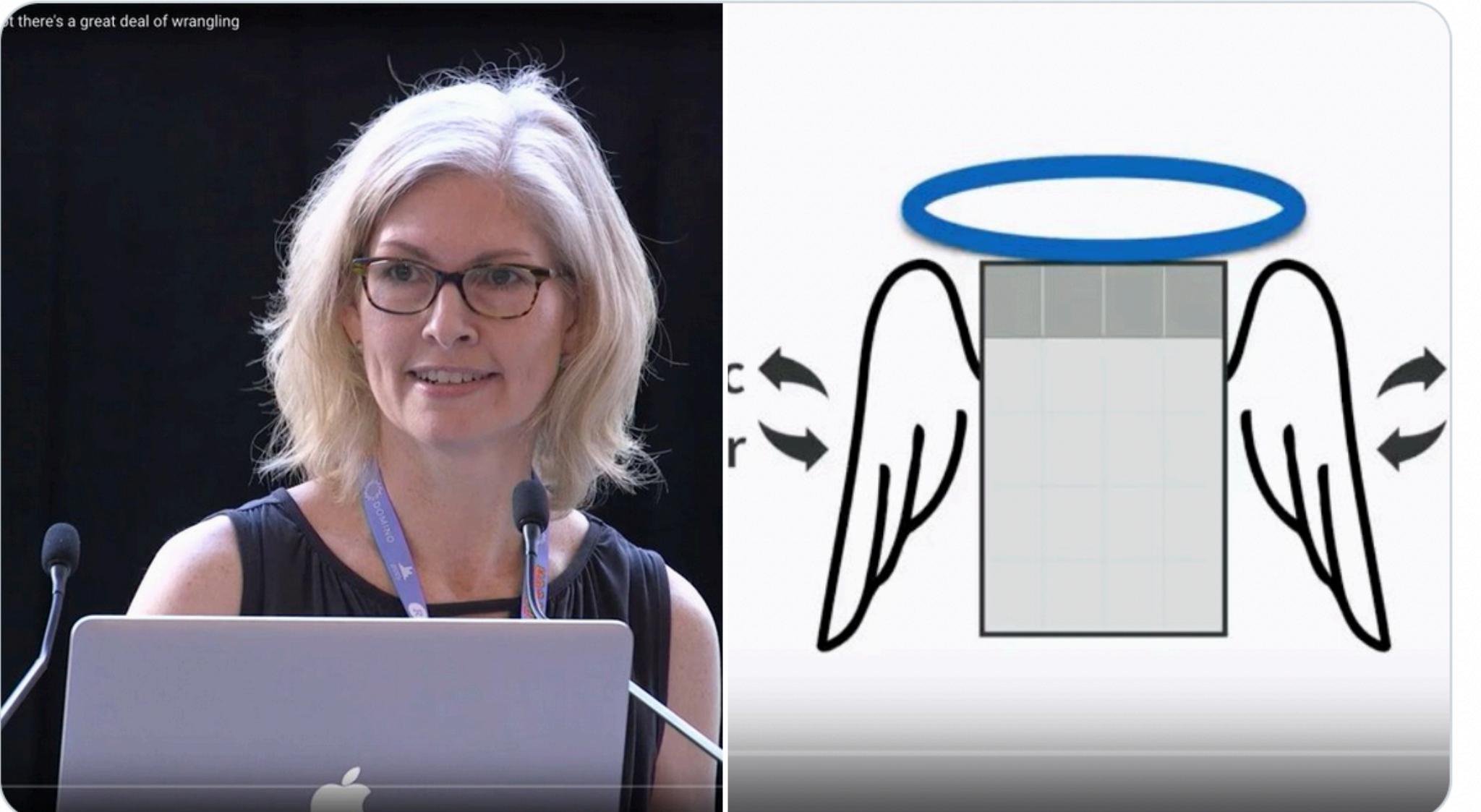
Dr. Amelia McNamara, Department of Computer & Data Sciences

# “Everything I know is from Jenny Bryan”

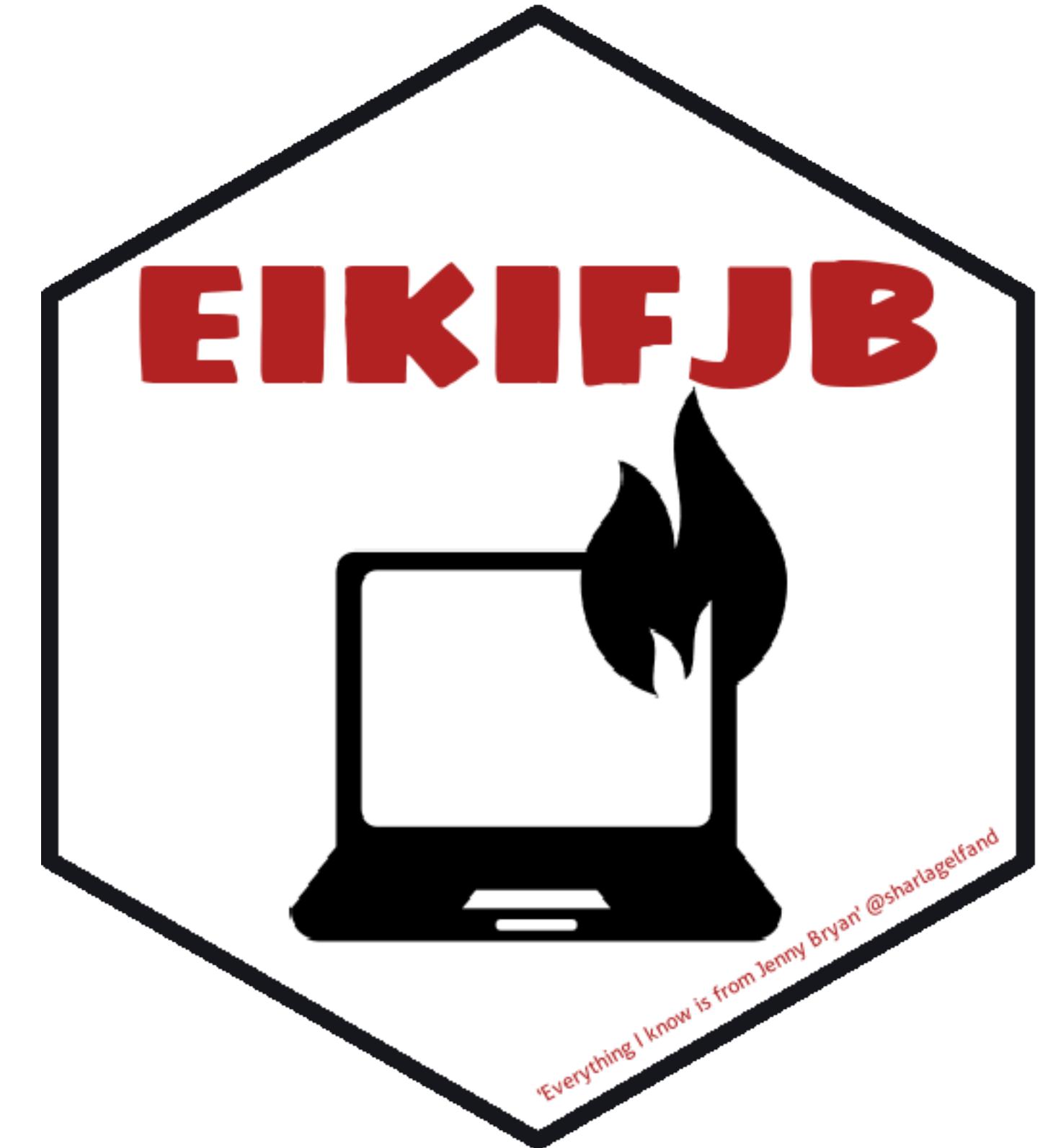
Amelia McNamara  
@AmeliaMN

Data angel @JennyBryan on purrr, list-columns and the importance of data wrangling. #PlotCon

[youtube.com/watch?v=4MfUCX...](https://www.youtube.com/watch?v=4MfUCX...)



11:54 AM · Dec 5, 2016 · TweetDeck



<https://github.com/MonkmanMH/EIKIFJB>

# Reproducibility

# Reproducibility

- We're data analysts, so it is important to work reproducibly
  - Reproducibility means "the code and data are assembled in a way so that another group can re-create all of the results (e.g., the figures in a paper)." ([Karl Broman](#))
- Many people think about reproducibility in science, but it is just as important in business and other domains. **Why?**
  - People need to check your work
  - Your boss wants the same thing done, but with point X removed
  - You need to redo the same analysis next year, on the new data
  - ...you get the idea

# Reproducibility is language-agnostic

A lot of what I have to say is R-flavored,  
because I am primarily an R programmer.  
But, it applies to Python, Julia, SQL, and  
other languages as well.

# Organization

# The first step to reproducibility is organization

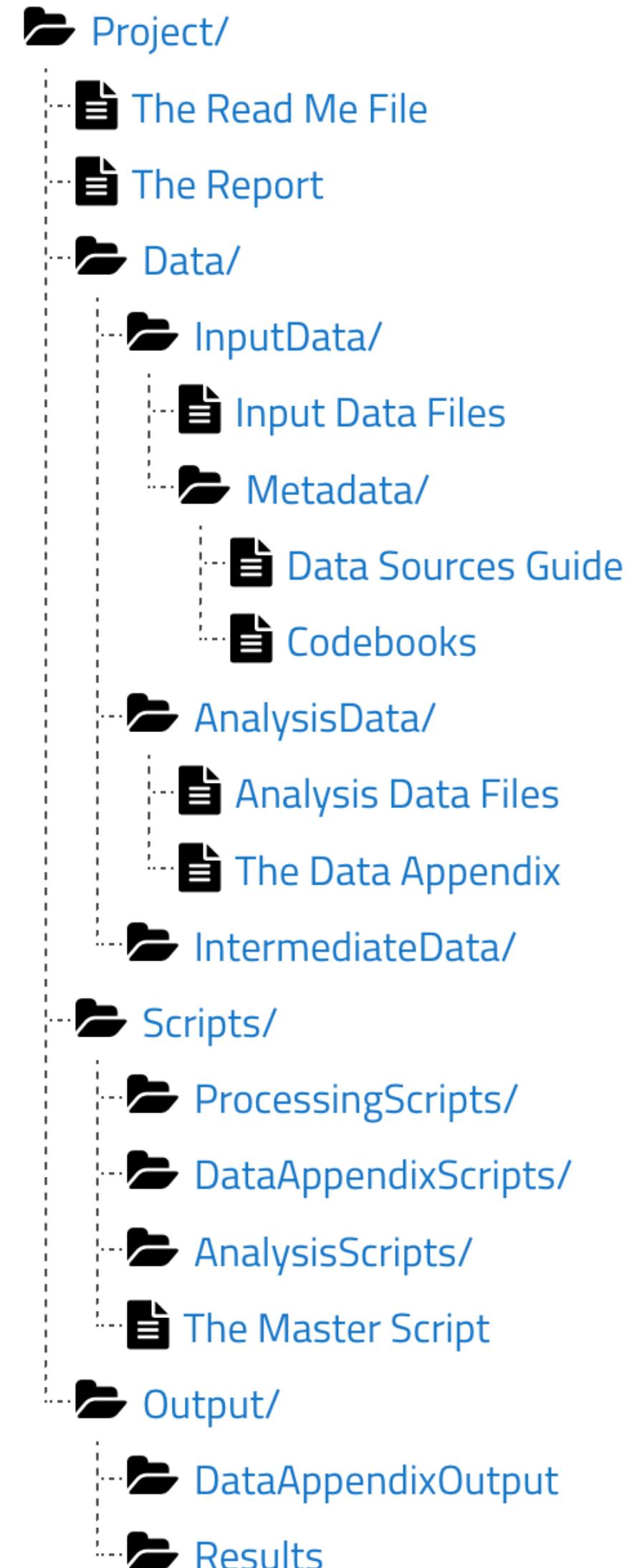
- As analysts, we work with lots of files
  - Data files (.xls, .xlsx, .csv, .json, .xml, etc)
  - Analysis files (.R, .Rmd, .Qmd, .py, .ipynb, etc)
  - Documentation files (.doc, .docx, .html, .pdf, etc)
- Files should be stored in folders (also called directories) in some intelligible way, and any references to files should be “relative” rather than “absolute” paths. (e.g.,  

```
.. /Data /InputData /some_file.csv,
```

rather than  

```
/Users/myUserName/Documents/Project/Data/
InputData /some_file.csv)
```
- The TIER Protocol is one suggested organizational method

## > TIER PROTOCOL 4.0



You need to name your files.

Names should strive to be:

- machine readable
- human readable
- useful with default ordering

## **NO**

my abstract.docx

Joe's Filenames Use Spaces and Punctuation.xlsx

figure I.png

fig 2.png

JW&^(2sl@deletethisandyourcareerisoverWx2\*.txt

## **YES**

2014-06-08\_abstract-for-sla.docx

joes-filenames-are-getting-better.xlsx

fig01\_scatterplot-talk-length-vs-interest.png

fig02\_histogram-talk-attendance.png

1986-01-28\_raw-data-from-challenger-o-rings.txt

# “machine readable”

avoid:

- spaces
- punctuation
- accented characters
- case sensitivity

**NO**

my abstract.docx  
Joe's Filenames Use Spaces and Punctuation.xlsx  
figure I.png  
fig 2.png  
JW&^(2sl@deletethisandyourcareerisoverWx2\*.txt

**YES**

2014-06-08\_abstract-for-sla.docx  
joes-filenames-are-getting-better.xlsx  
fig01\_scatterplot-talk-length-vs-interest.png  
fig02\_histogram-talk-attendance.png  
1986-01-28\_raw-data-from-challenger-o-rings.txt

# “human readable”

name means something to people

## NO

my abstract.docx

Joe's Filenames Use Spaces and Punctuation.xlsx

figure 1.png

fig 2.png

JW&^(2sl@deletethisandyourcareerisoverWx2\*.txt

## YES

2014-06-08\_abstract-for-sla.docx

joes-filenames-are-getting-better.xlsx

fig01\_scatterplot-talk-length-vs-interest.png

fig02\_histogram-talk-attendance.png

1986-01-28\_raw-data-from-challenger-o-rings.txt

# “plays well with default ordering”

start with a “slug” and/or a number

use the ISO 8601 standard for dates

left pad numbers with zeros

## NO

my abstract.docx

Joe's Filenames Use Spaces and Punctuation.xlsx

figure I.png

fig 2.png

JW&^(2sl@deletethisandyourcareerisoverWx2\*.txt

## YES

2014-06-08\_abstract-for-sla.docx

joes-filenames-are-getting-better.xlsx

fig01\_scatterplot-talk-length-vs-interest.png

fig02\_histogram-talk-attendance.png

1986-01-28\_raw-data-from-challenger-o-rings.txt

what you  
need to write



what people  
like to read

foo.R  
foo.Rmd  
foo.Qmd  
foo.ipynb

foo.md  
foo.html  
foo.pdf



# Quarto

## An authoring format for Data Science.

The screenshot displays the Quarto application interface, which includes a top navigation bar with tabs for Source and Visual, and various rendering and editing tools. Below the interface, there are two main sections: a code editor and a rendered document preview.

**Code Editor:**

```
---  
title: "My document"  
author: "Amelia McNamara"  
format: html  
editor: visual  
execute:  
  echo: true  
---
```

**Running Code:**

When you click the **Render** button a document will be generated that includes both content and the output of embedded code. You can embed code like this:

```
{r}  
1 + 1
```

You can add options to executable code like this

**R** is a collaborative project with many contributors. Type 'contributors()' for more information and 'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or 'help.start()' for an HTML browser interface to help. Type 'q()' to quit R.

[Workspace loaded from ~/STAT320/.RData]

**Quarto**

Quarto enables you to weave together content and executable code into a finished document. To learn more about Quarto see <https://quarto.org>.

**Running Code**

When you click the **Render** button a document will be generated that includes both content and the output of embedded code. You can embed code like this:

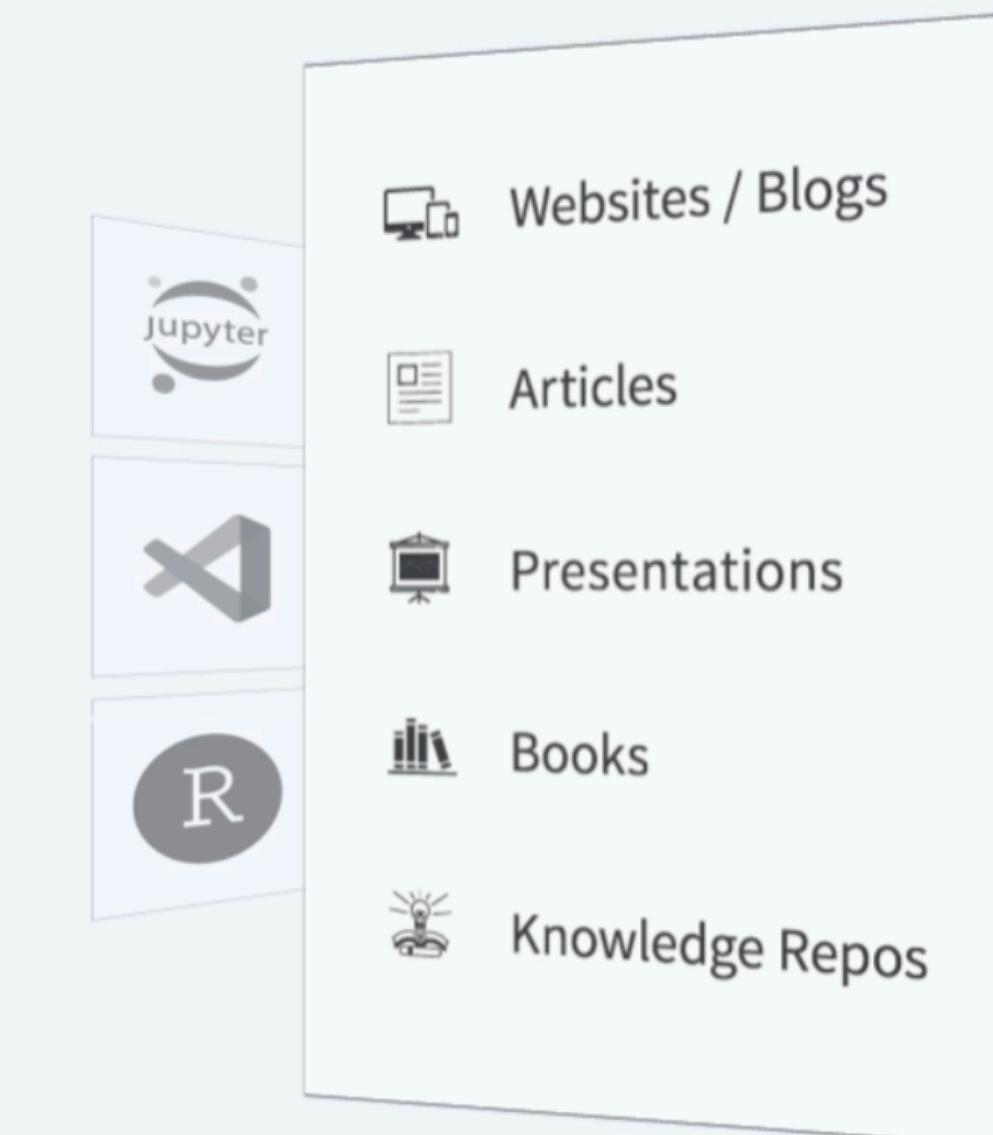
```
{r}  
1 + 1
```

You can add options to executable code like this

# Welcome to Quarto

An open-source scientific and technical publishing system

- Author using [Jupyter](#) notebooks or with plain text markdown in your favorite editor.
- Create dynamic content with [Python](#), [R](#), [Julia](#), and [Observable](#).
- Publish reproducible, production quality articles, presentations, websites, blogs, and books in HTML, PDF, MS Word, ePub, and more.
- Share knowledge and insights organization-wide by publishing to [Posit Connect](#), [Confluence](#), or other publishing systems.
- Write using [Pandoc](#) markdown, including equations, citations, crossrefs, figure panels, callouts, advanced layout, and more.



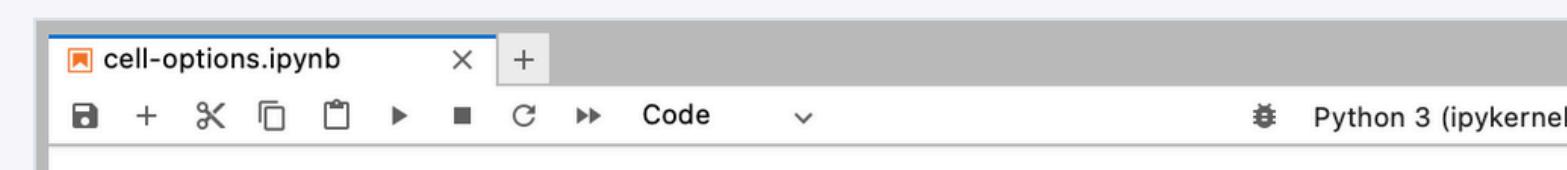
**Analyze. Share. Reproduce. You have a story to tell with data—tell it with Quarto.**

[Get Started](#)[Guide](#)

## Hello, Quarto

[Python](#)[R](#)[Julia](#)[Observable](#)

Combine Jupyter notebooks with flexible options to produce production quality output in a wide variety of formats. Author using traditional notebook UIs or with a plain text markdown representation of notebooks.



# Quarto

A way to mix code (R, Python, Julia, Observable, and more) with text (Markdown).

An example of “literate programming” (Don Knuth).

Next generation of files like RMarkdown and Jupyter notebooks.

# Palmer Penguins

```
---
author: Norah Jones
format:
  html:
    code-tools: true
    code-fold: true
---
```

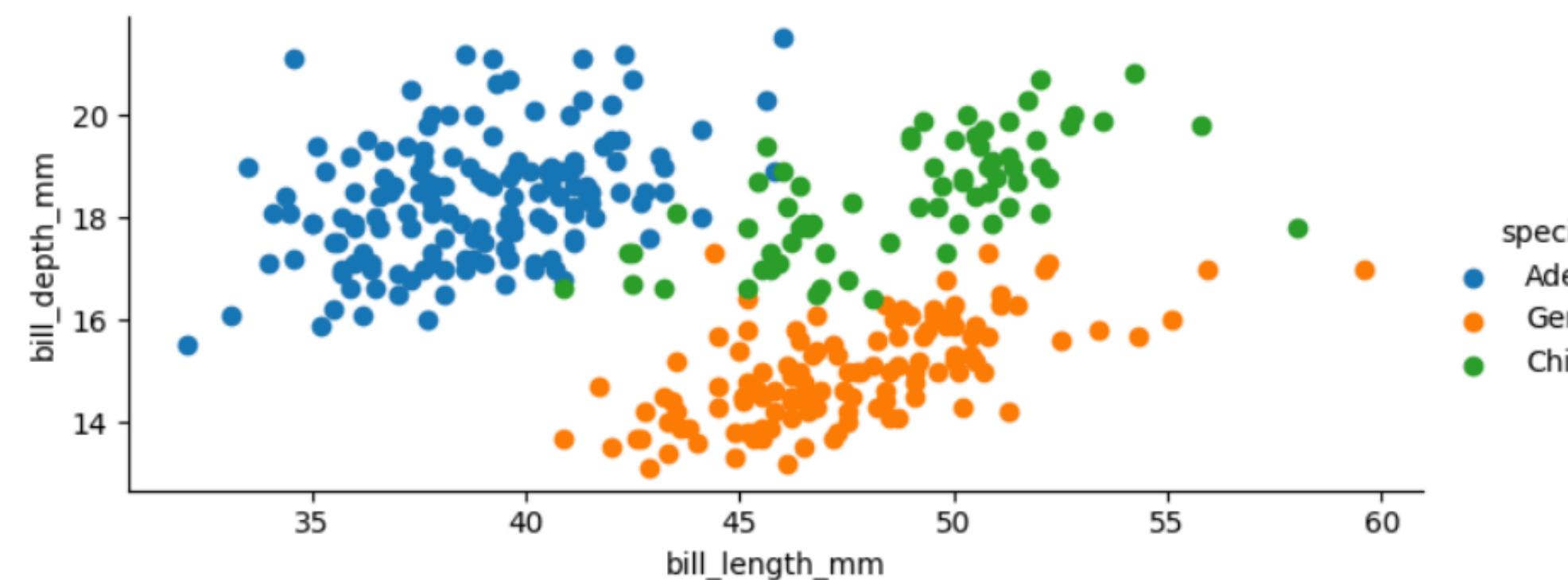
```
[3]: #| echo: false
import pandas as pd
df = pd.read_csv("palmer-penguins.csv")
```

## Exploring the Data

See @fig-bill-sizes for an exploration of bill sizes by species.

```
[6]: #| label: fig-bill-sizes
#| fig-cap: Bill Sizes by Species
import matplotlib.pyplot as plt
import seaborn as sns
g = sns.FacetGrid(df, hue="species", height=3, aspect=3.5/1.5)
g.map(plt.scatter, "bill_length_mm", "bill_depth_mm").add_legend()
```

```
[6]: <seaborn.axisgrid.FacetGrid at 0x2946720e0>
```



# Palmer Penguins

AUTHOR

Norah Jones

PUBLISHED

March 12, 2023

## Exploring the Data

See [Figure 1](#) for an exploration of bill sizes by species.

▼ Code

```
import matplotlib.pyplot as plt
import seaborn as sns
g = sns.FacetGrid(df, hue="species", height=3, aspect=3.5/2)
g.map(plt.scatter, "bill_length_mm", "bill_depth_mm").add_legend()
```

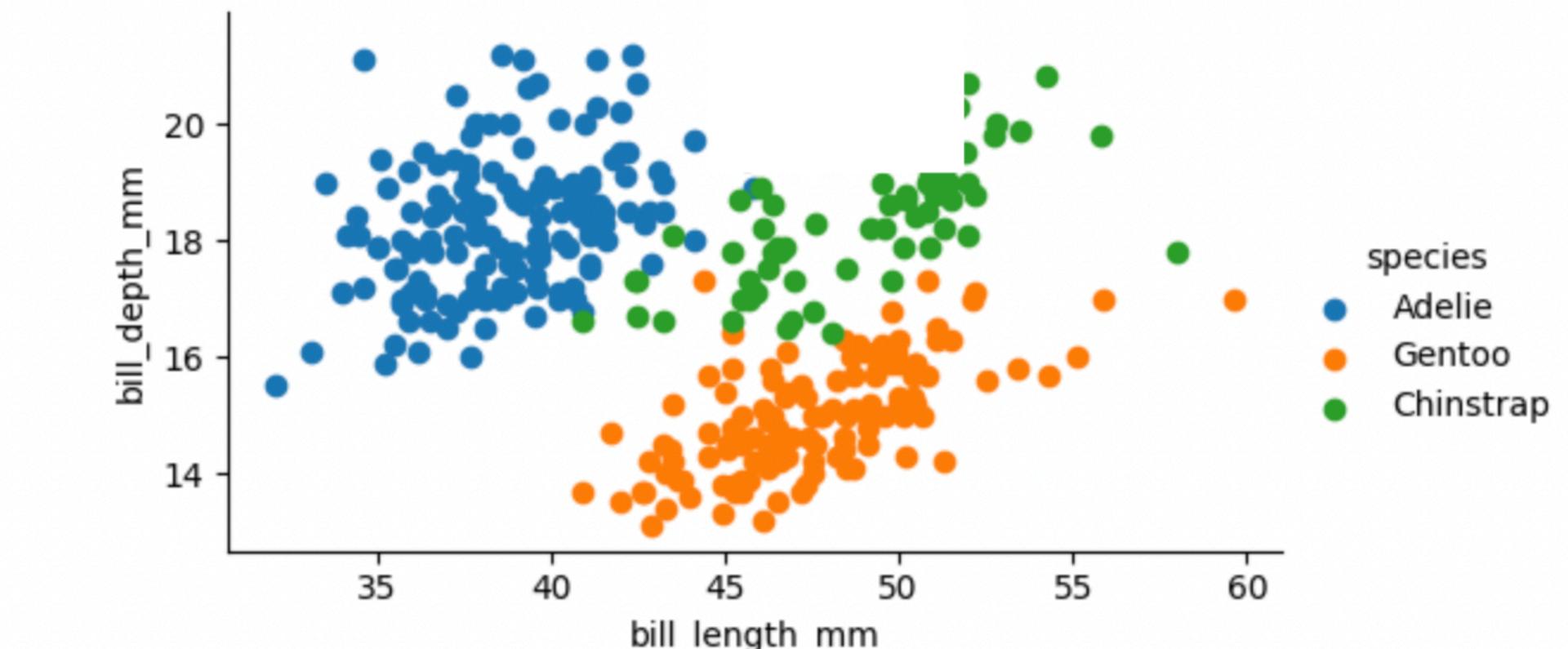


Figure 1: Bill Sizes by Species

Show All Code

Hide All Code

# ggplot2 demo

Norah Jones

May 22nd, 2021

## Air Quality

[Figure 1](#) further explores the impact of temperature on ozone level.

### ► Code

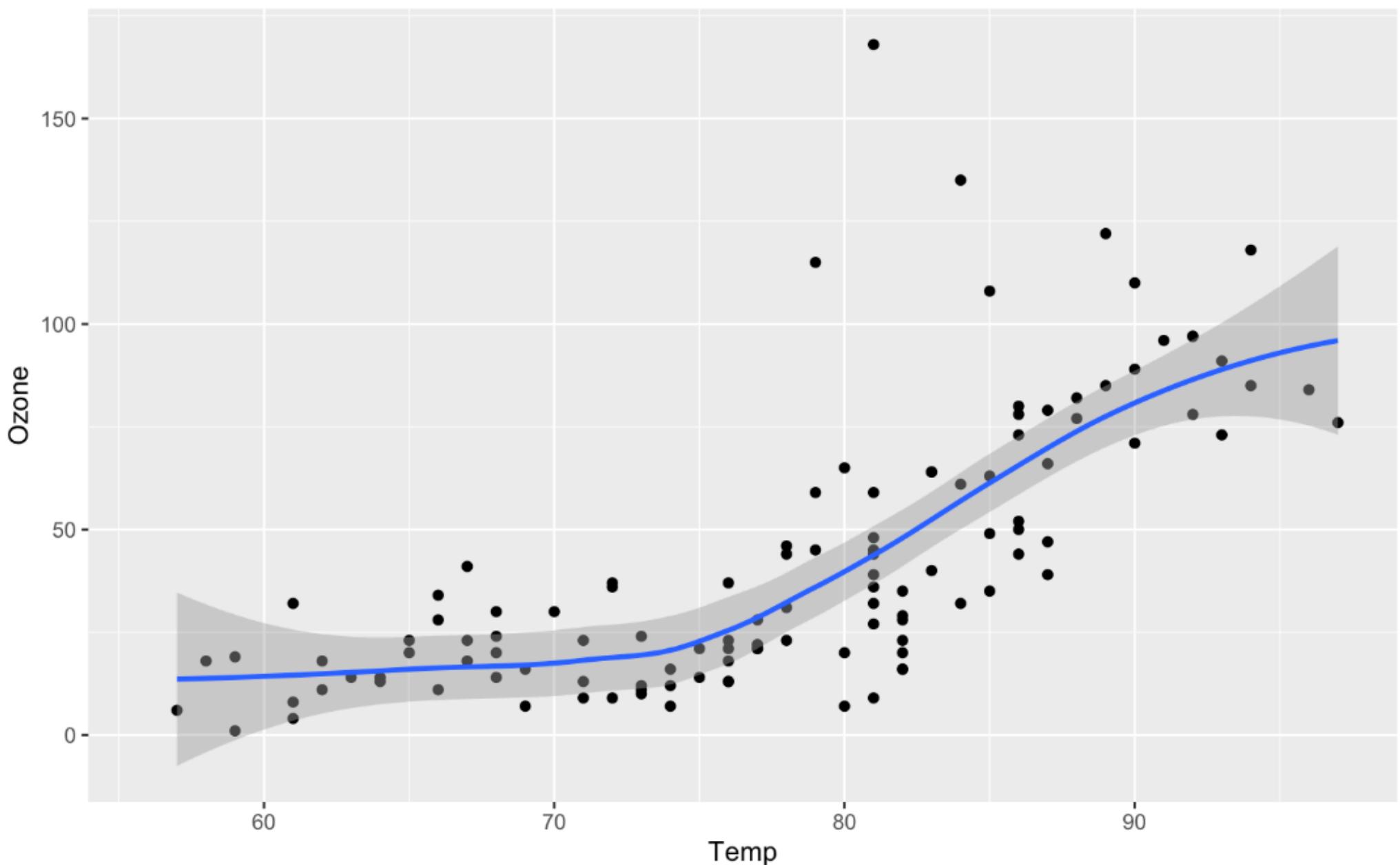


Figure 1: Temperature and ozone level.

Source Visual

```
1 ---  
2 title: "ggplot2 demo"  
3 author: "Norah Jones"  
4 date: "5/22/2021"  
5 format:  
6   html:  
7     fig-width: 8  
8     fig-height: 4  
9     code-fold: true  
10 ---  
11  
12 ## Air Quality  
13  
14 @fig-airquality further explores the impact of temperature on ozone level.  
15  
16 ```{r}  
17 #| label: fig-airquality  
18 #| fig-cap: "Temperature and ozone level."  
19 #| warning: false  
20  
21 library(ggplot2)  
22  
23 ggplot(airquality, aes(Temp, Ozone)) +  
24   geom_point() +  
25   geom_smooth(method = "loess")  
26 )  
27 ````  
28 |
```

# ggplot2 demo

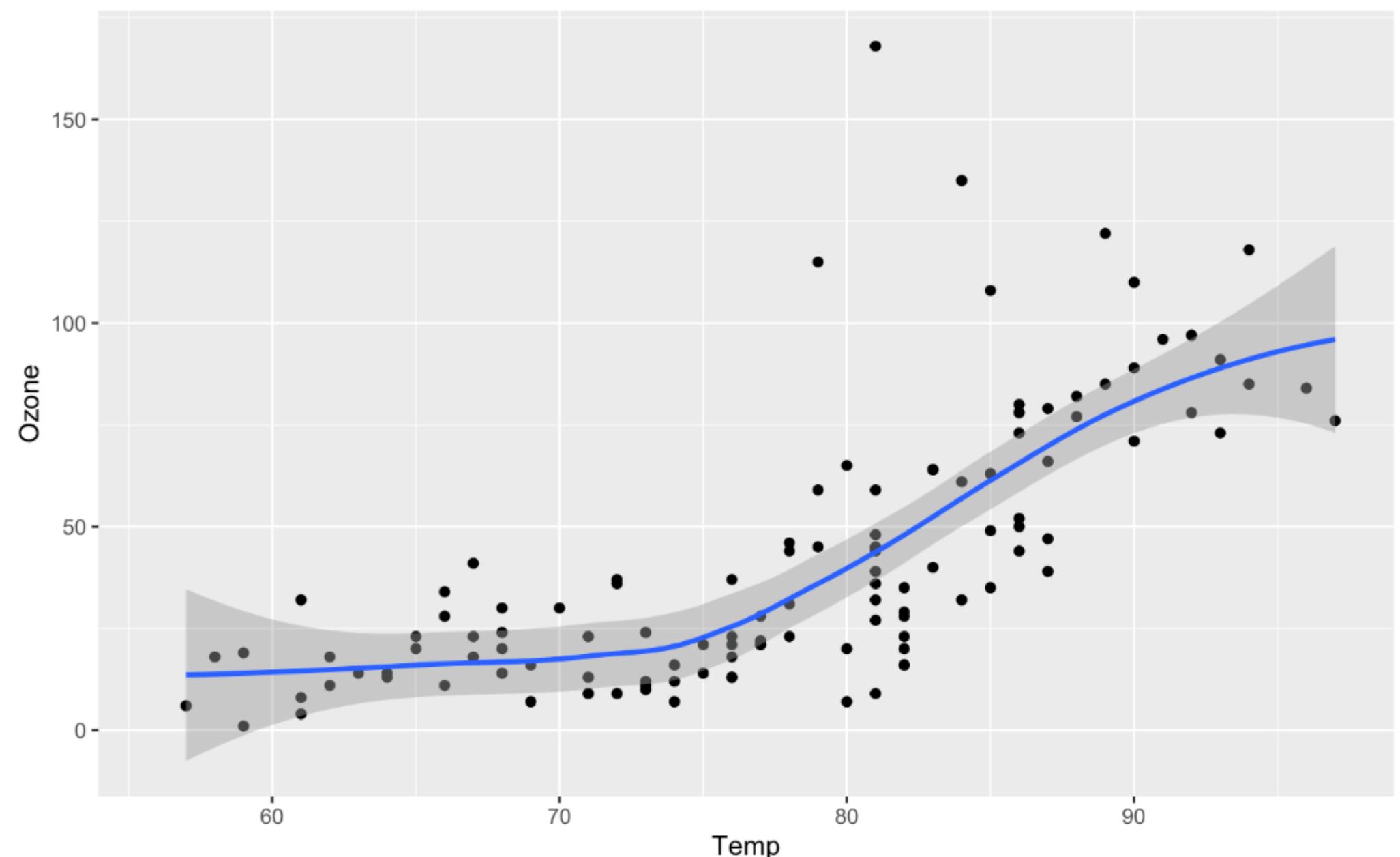
Norah Jones

May 22nd, 2021

## Air Quality

[Figure 1](#) further explores the impact of temperature on ozone level.

### ► Code



## Air Quality

@fig-airquality further explores the impact of temperature on ozone level.

```
{r}
#| label: fig-airquality
#| fig-cap: "Temperature and ozone level."
#| warning: false

library(ggplot2)

ggplot(airquality, aes(Temp, Ozone)) +
  geom_point() +
  geom_smooth(method = "loess")
)
```

# Markdown



# Markdown Complete Reference

By JOHN GRUBER

## Markdown: Syntax

Main   Basics   **Syntax**   License   Dingus

### ARCHIVE

THE TALK SHOW

DITHERING

PROJECTS

CONTACT

COLOPHON

RSS FEED

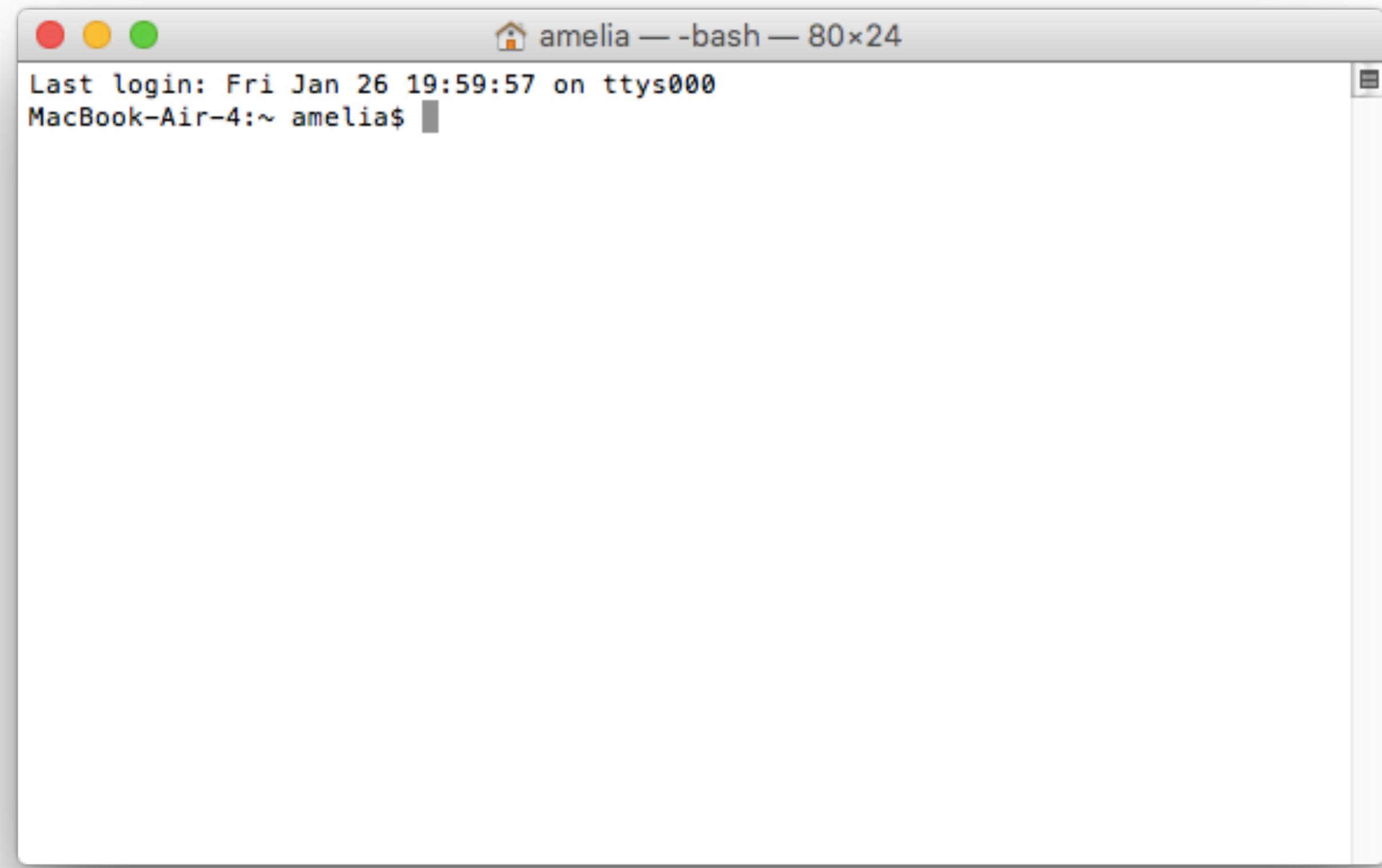
TWITTER

SPONSORSHIP

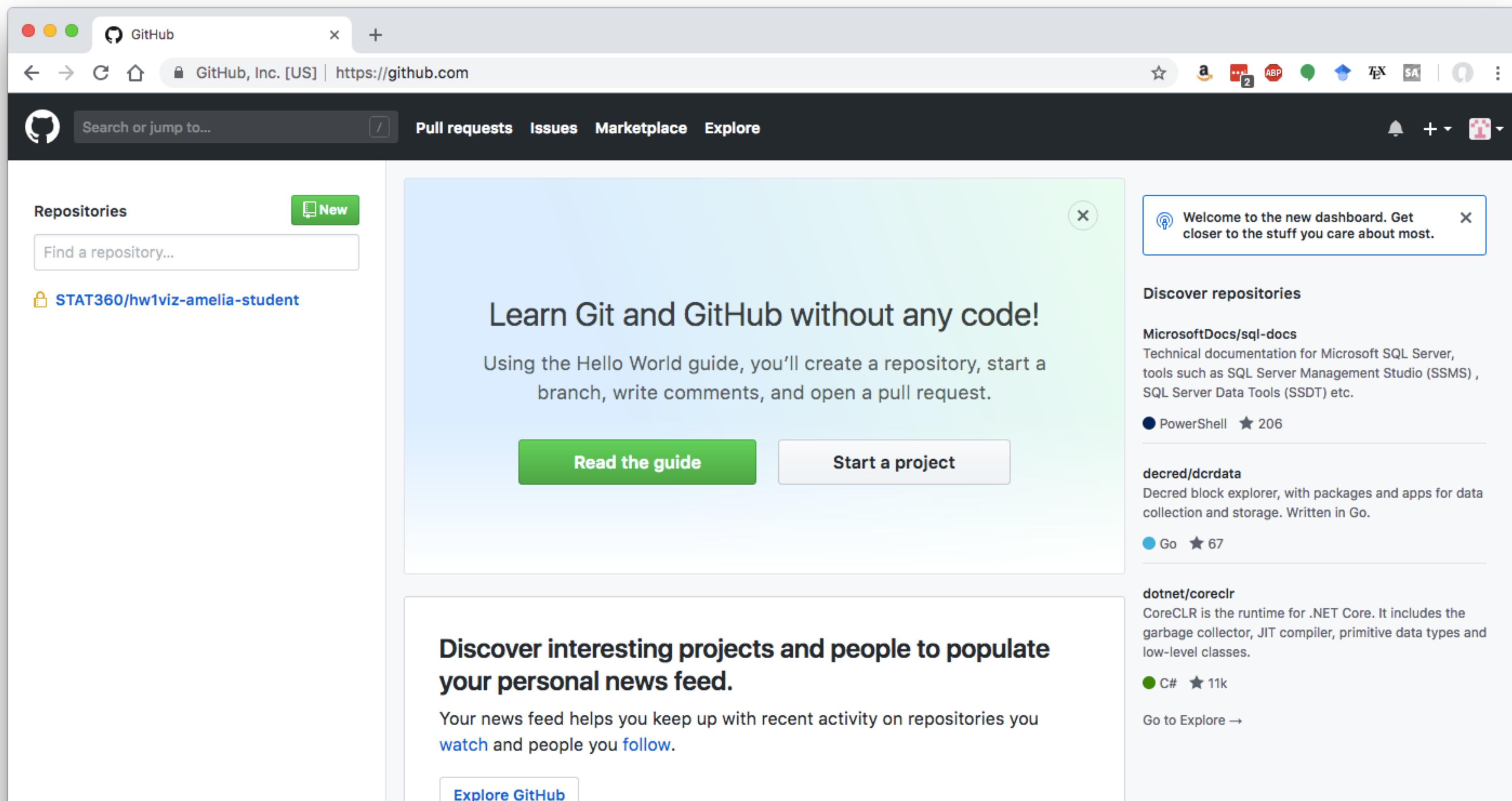
- [Overview](#)
  - [Philosophy](#)
  - [Inline HTML](#)
  - [Automatic Escaping for Special Characters](#)
- [Block Elements](#)
  - [Paragraphs and Line Breaks](#)
  - [Headers](#)
  - [Blockquotes](#)
  - [Lists](#)
  - [Code Blocks](#)
  - [Horizontal Rules](#)
- [Span Elements](#)
  - [Links](#)
  - [Emphasis](#)
  - [Code](#)
  - [Images](#)
- [Miscellaneous](#)
  - [Backslash Escapes](#)

# git and Github

# git: a version control system (almost like a programming language)



# Github: a site to host code (kind of like Dropbox, but better)



# version control

# "FINAL".doc



FINAL.doc!



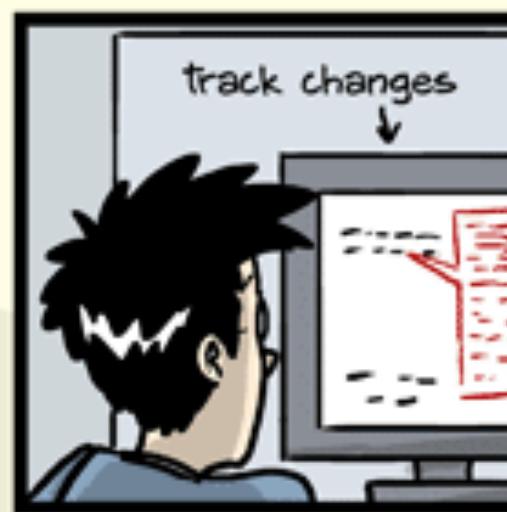
FINAL\_rev.2.doc



FINAL\_rev.6.COMMENTS.doc



FINAL\_rev.8.comments5.  
CORRECTIONS.doc



FINAL\_rev.18.comments7.  
corrections9.MORE.30.doc

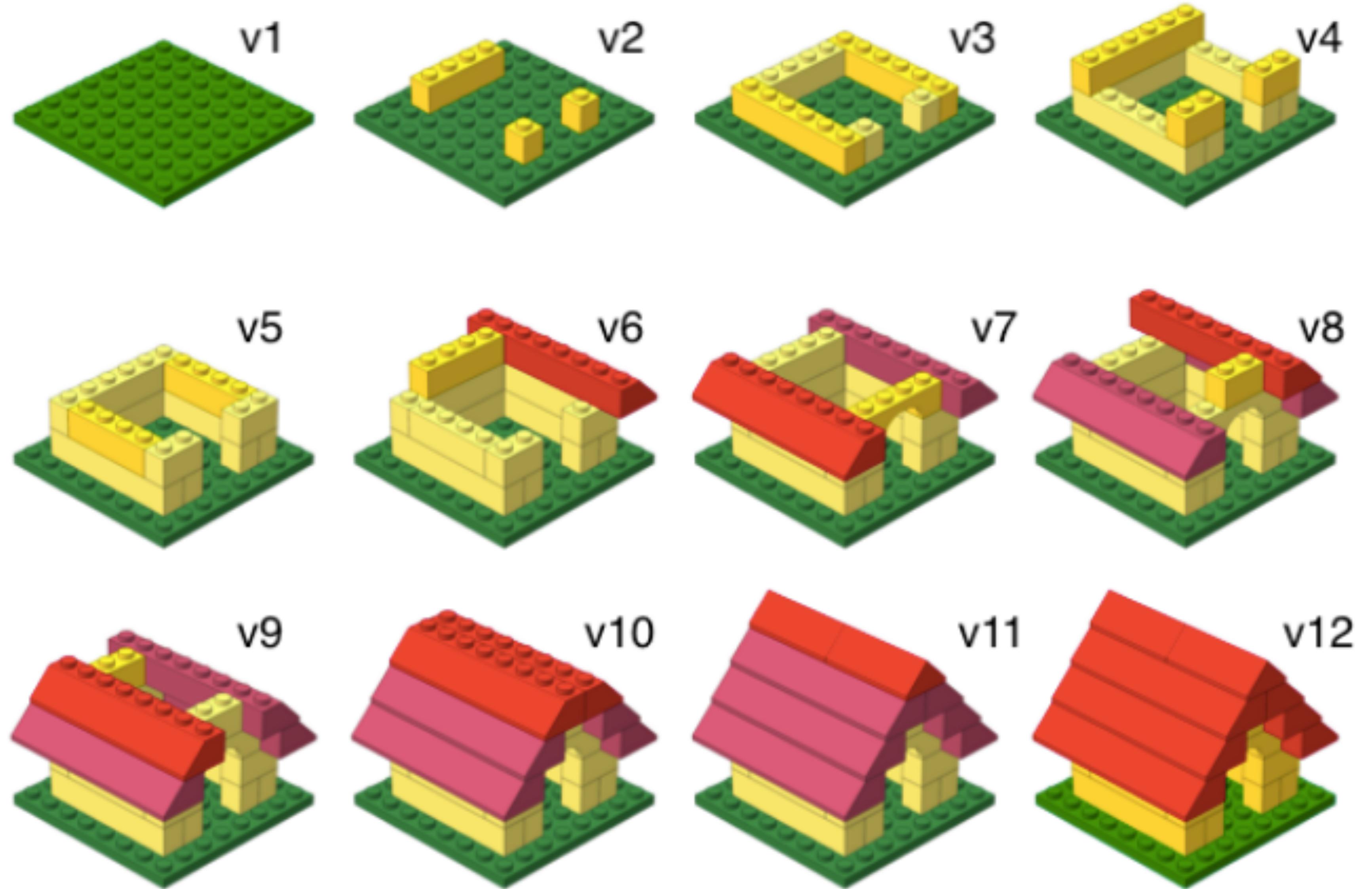


FINAL\_rev.22.comments49.  
corrections.10.#@\$%WHYDID  
ICOMETOGRAD SCHOOL????.doc

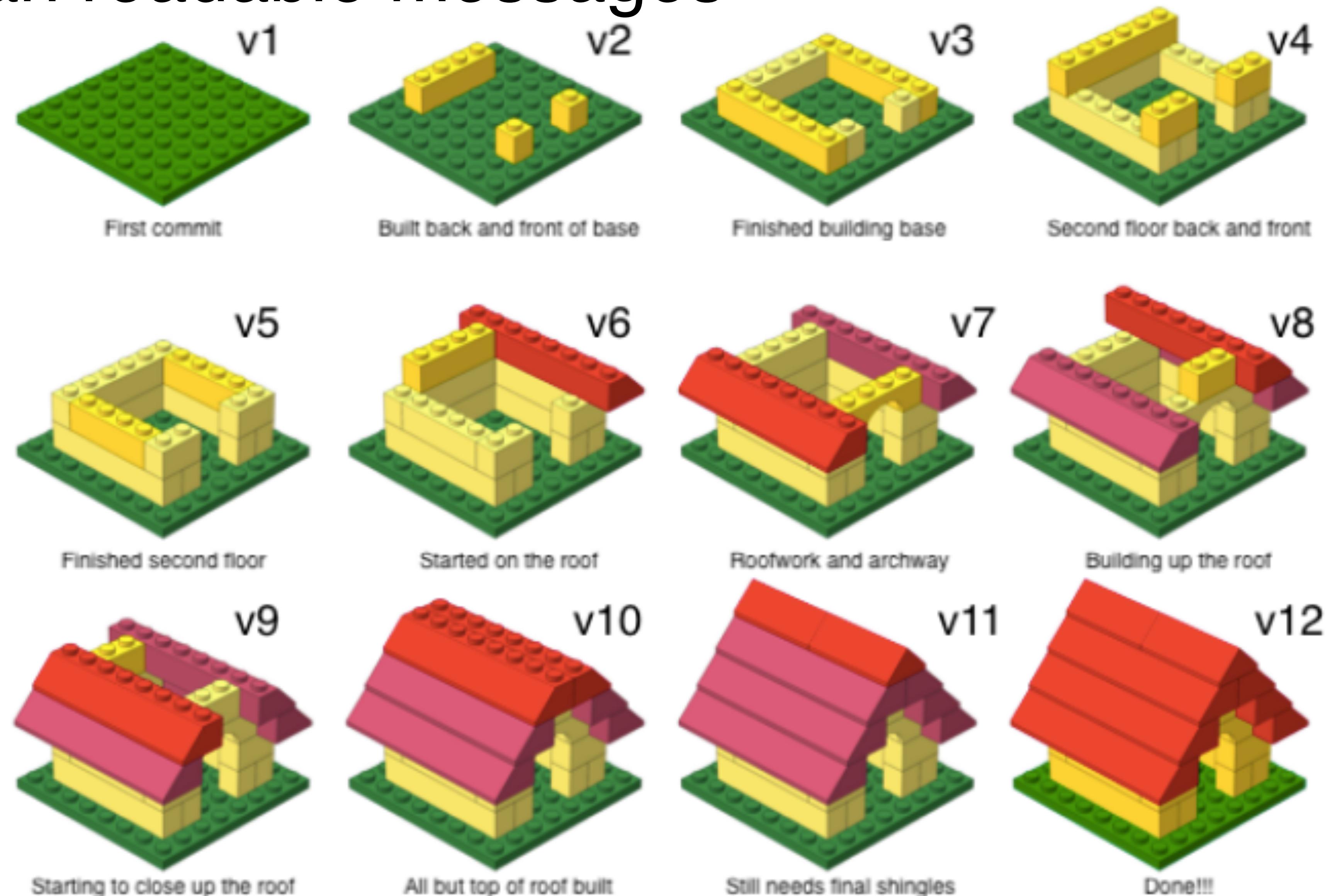


JORGE CHAM © 2012

# Versioning

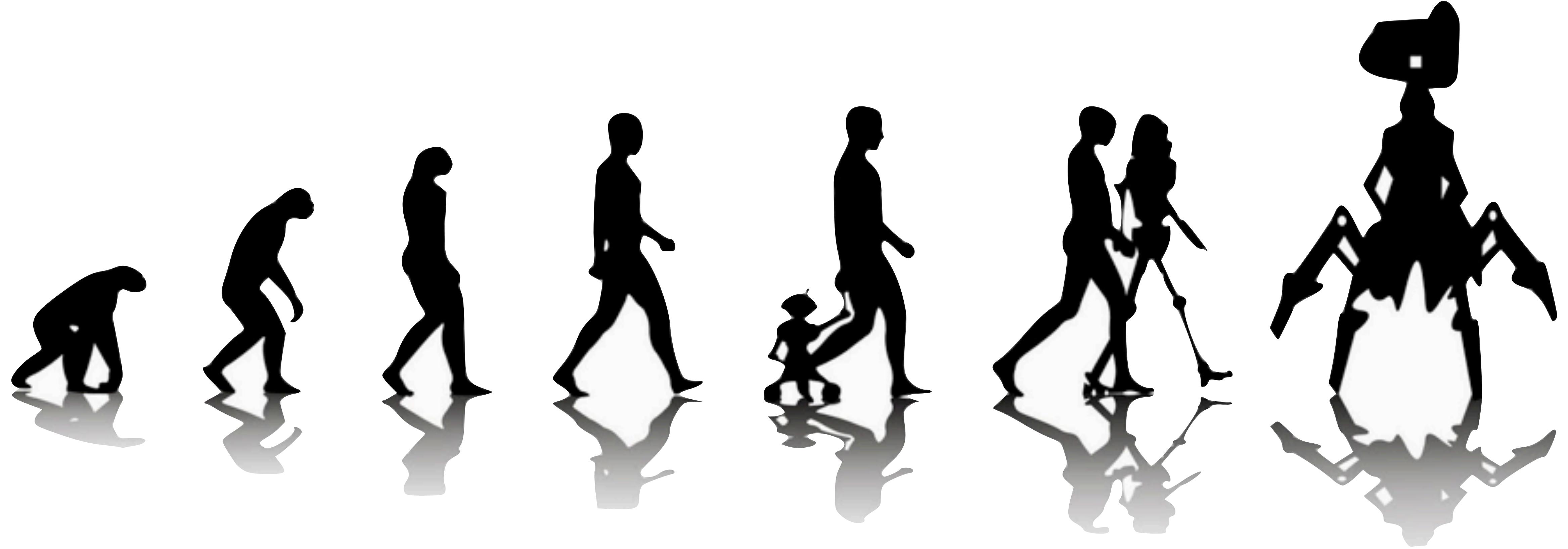


# Versioning with human readable messages

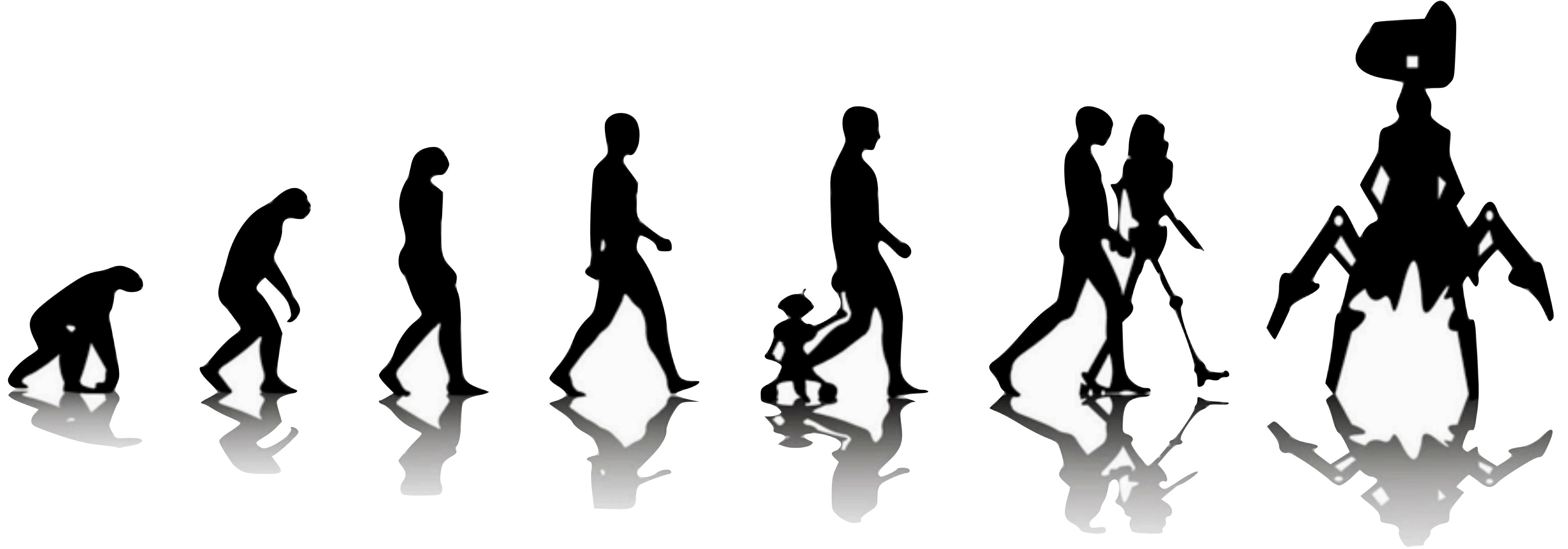


# *Deep Thoughts*

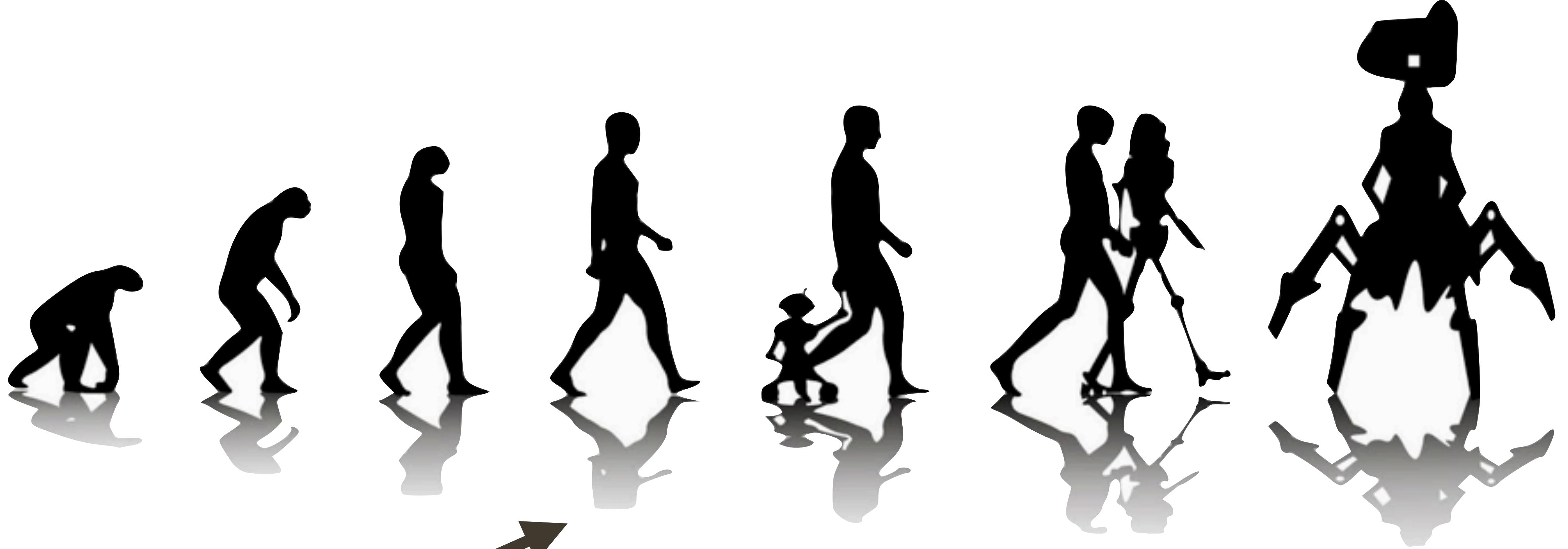




use version control

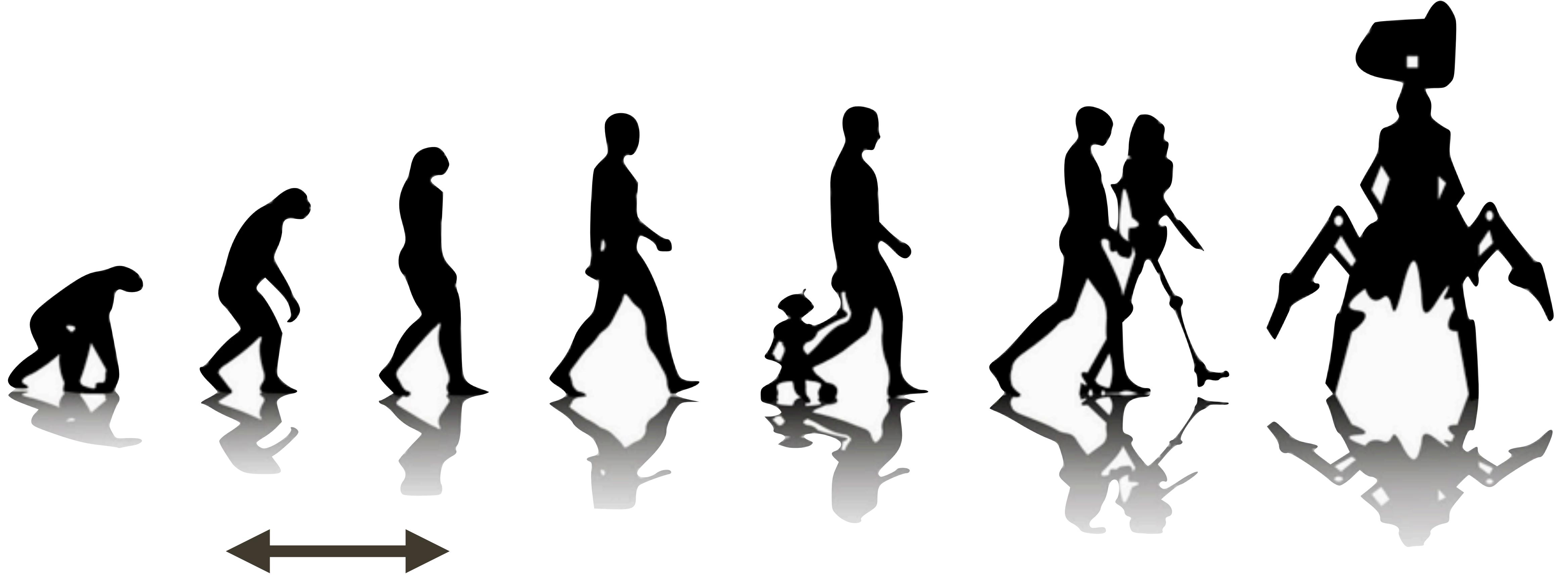


**we teach Git + GitHub**



"commit"

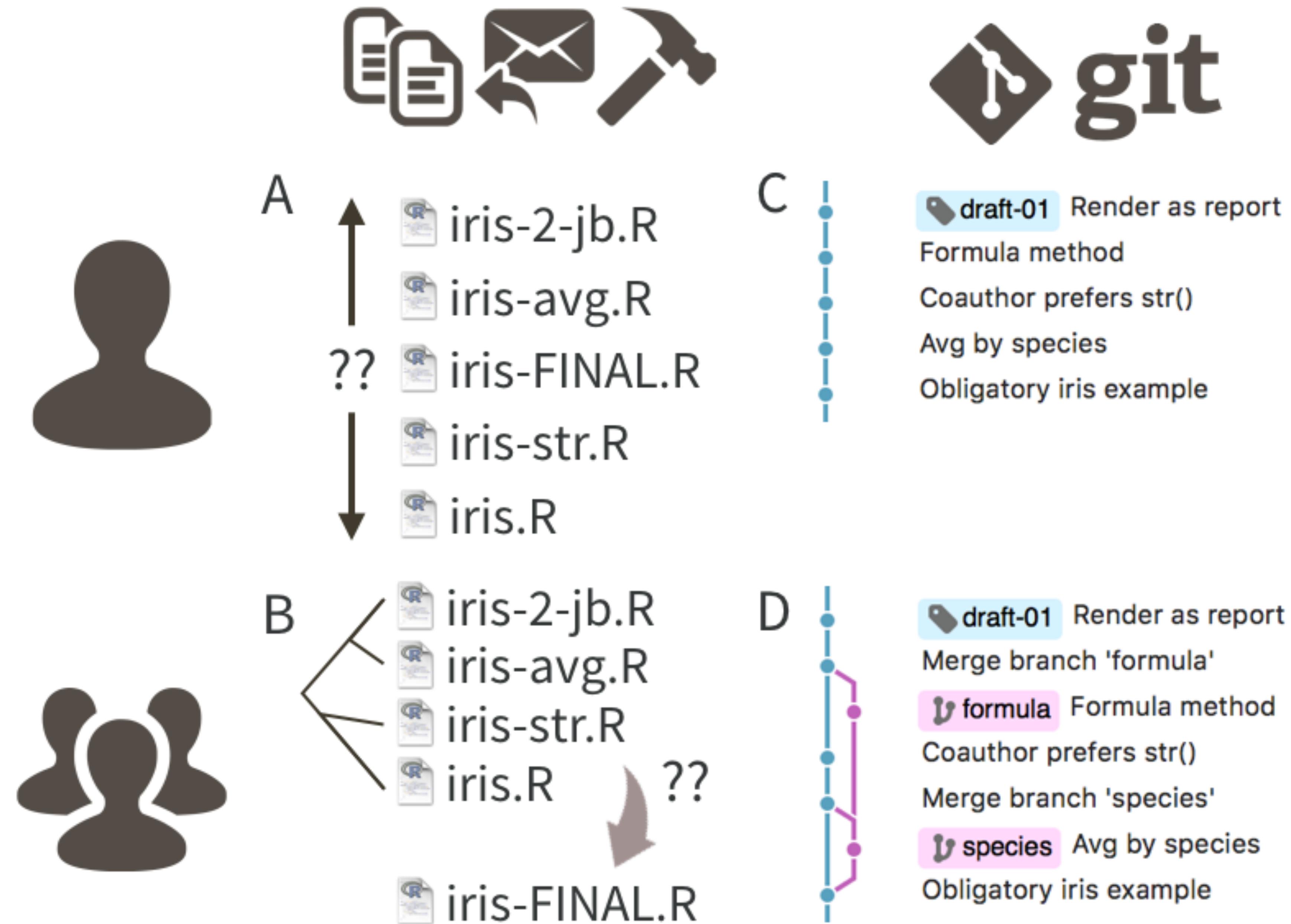
a file or project state that is **meaningful to you**  
for inspection, comparison, restoration



Δ

"diff"

What changed here?  
Why?



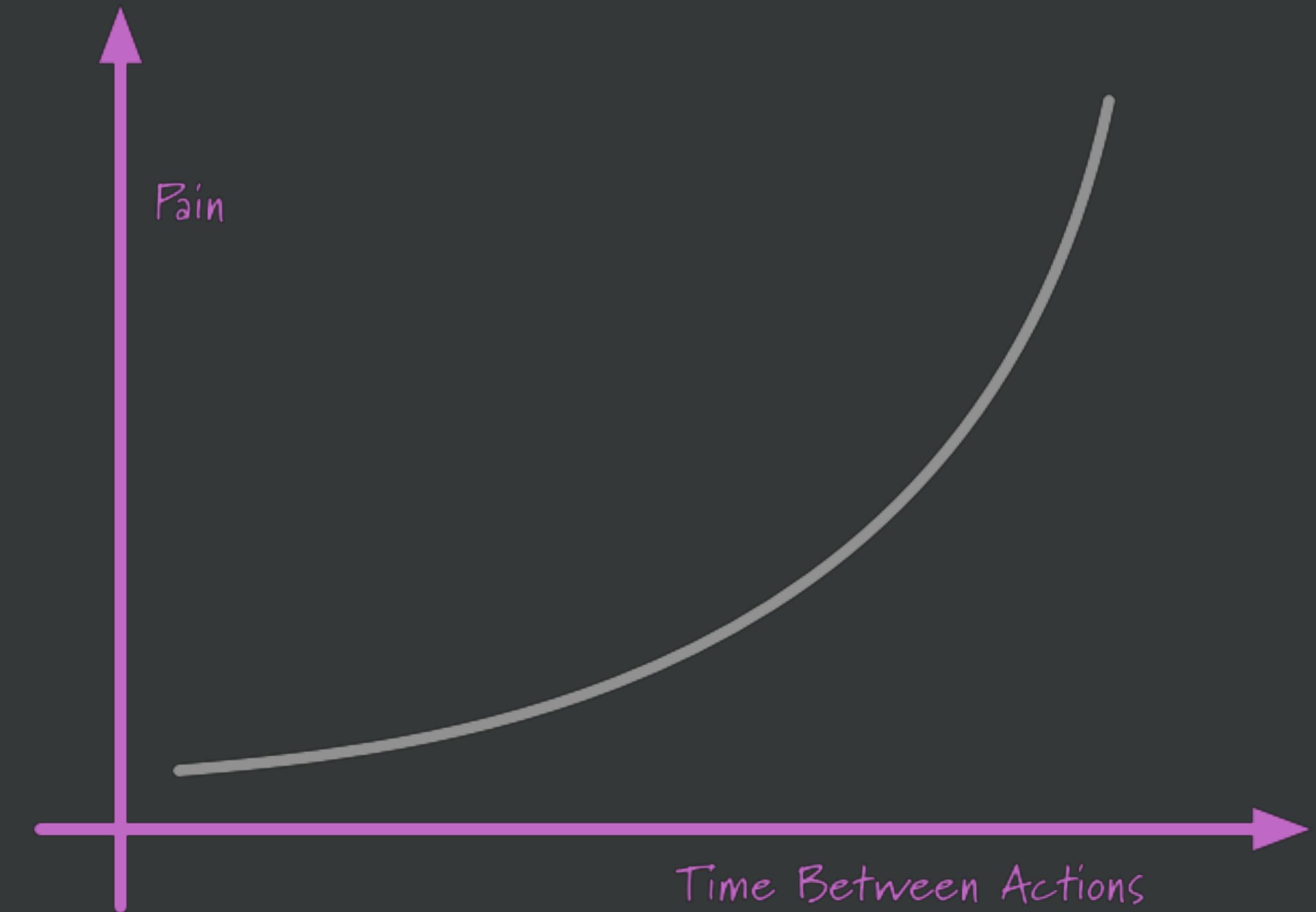
**“In every project you have at least one other collaborator; future-you. You don’t want future-you to curse past-you.”**

**- Hadley Wickham**

# *Deep Thoughts*



# "If it hurts, do it more often."



<https://martinfowler.com/bliki/FrequencyReducesDifficulty.html>

"If it hurts, do it more often."

Apply this to git commit, pull, merge, push.  
(and restarting R, re-running your scripts)

Why?

Take your pain in smaller pieces.

Tight feedback loop can reduce absolute pain.

Practice changes what you find painful.

where we want to “git”

# Github: a place to host a portfolio of work

The screenshot shows a GitHub profile page for a user named kurtispykes. The profile features a circular profile picture of a man with a beard and glasses. The user's name, "Kurtis Pykes", and handle, "kurtispykes", are displayed. Below the profile picture, there are statistics: 277 followers and 8 following. The user is located in Croydon, London. The "Achievements" section shows four icons: a cowboy hat, a brain, a shark, and a crown. A "Follow" button is present.

**Overview**   **Repositories 17**   **Projects**   **Packages**   **Stars 14**

**kurtispykes / README.md**

**Hi, I'm Kurtis Pykes**

I'm a self-taught machine learning engineer working as a freelancer. I have a keen interest in designing, building, and deploying machine learning applications, as well as writing. Outside of work, I enjoy...

- ⚽ Playing football and watching Chelsea FC (The best club in the world.)
- 🕋 Calisthenics - my preferred style of exercise
- 📖 Reading.

**Writing**

- [Building Reproducible Machine Learning Pipelines](#) - 17th June 2021
- [Fighting Overfitting with L1 or L2 Regularization-Which one is Better?](#) - 19th July 2021
- [Best Practices for Data Science Project Workflows and File Organization](#) - 20th July 2021
- [How to Design a Machine Learning System](#) - 5th Sept 2021

**Connect with Me!**

- [Twitter](#)
- [LinkedIn](#)
- [Medium](#)

**Pinned**

**Machine-Learning** Public

All content related to machine learning from my blog

Jupyter Notebook   ★ 118   26

**Python** Public

All curated content and code about Python

Jupyter Notebook   ★ 87   17

# Github: a place to host a portfolio of work

The screenshot shows a GitHub user profile for **Mitchell Nelson**. The profile picture is a circular photo of a young man with glasses and a beard, wearing a blue beanie and a denim jacket. Below the picture, the name **Mitchell Nelson** is displayed, followed by the handle **MitchellNelson**. A **Follow** button is present. The bio indicates he is a **Software Engineer and Musician**. The profile has **8 followers** and **8 following**. Location is listed as **Minneapolis, MN**, and the GitHub URL is **mitchellnelson.github.io**.

The main content area shows the **Overview** tab selected, with **Repositories 16**, **Projects**, **Packages**, and **Stars 3** also visible. The **Pinned** section displays five repositories:

- RileyTschumper/MinneHack2020** (Public) - Python, Stars 2
- drumCV** (Public) - Computer Vision Drum Simulator, Python, Stars 6, Forks 1
- beat-collab** (Public) - Multi user approach to real-time music creation, JavaScript
- AutoGarconAndroid** (Public) - Forked from auto-garcon/android, Java, Stars 2, Forks 1
- airqualitymap** (Public) - CISC 375 Project 2, JavaScript
- nels4929.github.io** (Public) - JavaScript

A note at the bottom of the pinned section says **Single sign-on** to see contributions within the **UniversityOfSaintThomas** organization.

The **Achievements** section shows two icons: a blue globe and a blue gear. A **Beta** button and a **Send feedback** link are also present.

The **Contributions** section shows a grid of activity for the last year, with most days having contributions. The grid spans from April to March of the following year.

Tools and Palettes for Bivariate

biscale 1.1.0.9000 Get started Reference Articles Changelog

# biscale

biscale implements a set of functions for bivariate thematic mapping based on the [tutorial](#) written by Timo Grossenbacher and Angelo Zehr as well as a set of bivariate mapping palettes, including Joshua Stevens' [classic color schemes](#). In addition to support for two-by-two, three-by-three, and four-by-four maps, the package also supports a range of methods for calculating breaks for bivariate maps.



## What's New in v1.0.0?

### New Features

- `bi_class()` now accepts factors for one or both of the `x` and `y` variables, allowing more flexibility for how breaks are calculated. If you want finer grained control over your categories, calculate them ahead of time and then pass the factors on to `bi_class()`.
- `bi_pal()`, `bi_legend()`, `bi_scale_fill()`, and `bi_scale_color()` functions all support four-by-four bivariate maps when `dim = 4`. Note that the original five palettes do not support four-by-four mapping, but very close approximations (e.g. `DkBlue2` instead of `DkBlue`) are now provided in their place. The legacy palettes are all still included in the package.
- The number of built-in palettes has been greatly expanded!
- Palettes can now be flipped and rotated (or both!), so that each built-in palette can be displayed in four different configurations. This includes the built-in palettes and any custom palettes that are four-by-four or smaller. If you want to flip or rotate larger palettes, you should make those decisions while creating the palette itself.
- The workflow for allowing custom palettes has been overhauled to simply the process - users can provide a named vector for the `pal` arguments in the `bi_pal()`, `bi_legend()`, `bi_scale_fill()`, and `bi_scale_color()` functions. All of these functions will validate your input to ensure that it maps correctly.
- `bi_class()` can be used to calculate bivariate breaks for maps larger than four-by-four, though it will return a warning reminding you that these maps are hard to read and that `biscale` does not provide palettes for larger maps. Instead, you should provide a custom palette.
- `bi_class_breaks()` can be used with `bi_legend()` to facilitate optionally adding break values to your legends. Like `bi_class()`, this new function accepts both continuous and pre-made factors.

### Breaking Changes

- R version 3.4 is no longer supported - please use at least R version 3.5
- There is no default supplied for `bi_class()`'s `style` argument since `bi_class()` now accepts factors as well. Users that relied on the default behavior of `bi_class()` will now receive an error asking you to specify a `style` for calculating breaks.

### Deprecated Functions

- `bi_pal_manual()` now returns a warning that it has been deprecated and will be removed in a later release of `biscale` (planned for the end of 2022). Please update your workflows to use the new approach to generating custom palettes.

### Links

[View on CRAN](#)  
[Browse source code](#)  
[Report a bug](#)

### License

[GPL-3](#)

### Community

[Contributing guide](#)  
[Code of conduct](#)

### Citation

[Citing biscale](#)

### Developers

Christopher Prener  
Author, maintainer 

Timo Grossenbacher  
Author

Angelo Zehr  
Author

[More about authors...](#)

### Dev status

 [codecov](#) 95%

 [CRAN](#) 1.0.0

[cran checks](#)

 [downloads](#) 631/month

 [downloads](#) 28K

 [DOI](#) 10.5281/zenodo.6604998

Image Scrubber

https://everestpipkin.github.io/image-scrubber/

Open Image

Save Image

Rotate Image

Paint

Brush

Blur

Undo

Rectangle

Tap

Brush Size

Blur Radius

## Image Scrubber

This is a tool for anonymizing photographs taken at protests.

It will remove identifying metadata ([Exif data](#)) from photographs, and also allow you to selectively blur parts of the image to cover faces and other identifiable information.

Click or drag onto the open button to open a photograph. The program will display the data it is removing.

Click okay, and you can then save the scrubbed image by hitting save or right clicking on it and saving it. Maximum size is 2500x2500 pixels – larger images will be scaled down.

You can select between painting over the image or blurring it out. Dragging on the image will paint on or blur it. You can change your brush size via the slider. The blur function runs on top of shuffled, noisy, upscaled pixels and is fairly secure but sensitive information should be covered with the paint tool.

This tool works offline: on a phone you can load the page or add it to the homescreen, then turn on airplane mode (or turn off wifi/data) before opening any pictures. On a computer, [download](#) the zipped code, open the folder, and open index.html in a browser with the internet turned off.

**All processing happens directly in the browser- no information is stored or sent anywhere.**

Bug reports or questions to: [everest.pipkin@gmail.com](mailto:everest.pipkin@gmail.com)

Github repo at [github.com/everestpipkin/image-scrubber](https://github.com/everestpipkin/image-scrubber)

If you want other ways to cover your digital footprint, I've assembled a list of resources: via [Google Doc](#) or [Pastebin](#).

About

Final Project - Mac Miller Analysis | f X +

https://stat490.github.io/final-BrentAPearson/

# final-BrentAPearson

Final Project Website

[View the Project on GitHub](#)  
STAT490/final-BrentAPearson

## Final Project - Mac Miller Analysis

Brent Pearson

### Background

This project will include an analysis of two Mac Miller albums, Faces and Circles, two albums that are sentimentally similar and entirely different. For an in depth explanation of why I chose to analyze Mac Miller albums, refer to the [Meta-Document](#). To look at my entire R code for this project, refer to the [README RMD Document](#)

A final version of my website can be accessed by [clicking here](#)

### Data

My data was obtained from LyricsOnDemand.com. I was unable to find an api that could scrape this website, so I copied and pasted the words into two separate .txt files, one for each album. The text file for [Faces](#) can be found here. The text file for [Circles](#)

### Data cleaning

For data cleaning, I needed to replace all line breaks with spaces, convert the text to lowercase, remove stop words, punctuation and other words that were skewing my results. One thing that caused problems was removing all punctuation because there was punctuation in most of the curse words to censor them. This produced an abbreviated version of each of the curse words and caused them to not show up in the

### Top 10 word frequency

One of the first things you will notice is what I mentioned above, that most of the curse words are abbreviated to fck, sht, or btch. This occurred because each of the curse words included a "\*" to censor the word and it was removed after taking away the punctuation from each sentence. Just based on the most frequent words, you can tell that Faces is more aggressive and Circles is more tamed, which will be a common narrative throughout this project.

FACES

kurtispykes (Kurtis Pykes) +

https://github.com/kurtispykes

Search or jump to... / Pull requests Issues Codespaces Marketplace Explore

Overview Repositories 17 Projects Packages Stars 14



**Kurtis Pykes**  
kurtispykes

Follow

277 followers · 8 following

Croydon, London

Achievements



Beta Send feedback

Block or Report

**Hi, I'm Kurtis Pykes**

I'm a self-taught machine learning engineer working as a freelancer. I have a keen interest in designing, building, and deploying machine learning applications, as well as writing. Outside of work, I enjoy...

- Playing football and watching Chelsea FC (The best club in the world.)
- Calisthenics - my preferred style of exercise
- Reading.

**Writing**

- [Building Reproducible Machine Learning Pipelines](#) - 17th June 2021
- [Fighting Overfitting with L1 or L2 Regularization-Which one is Better?](#) - 19th July 2021
- [Best Practices for Data Science Project Workflows and File Organization](#) - 20th July 2021
- [How to Design a Machine Learning System](#) - 5th Sept 2021

**Connect with Me**

- [Twitter](#)
- [LinkedIn](#)
- [Medium](#)

Pinned

**Machine-Learning** Public

All content related to machine learning from my blog

Jupyter Notebook 118 ⚡ 26

**Python** Public

All curated content and code about Python

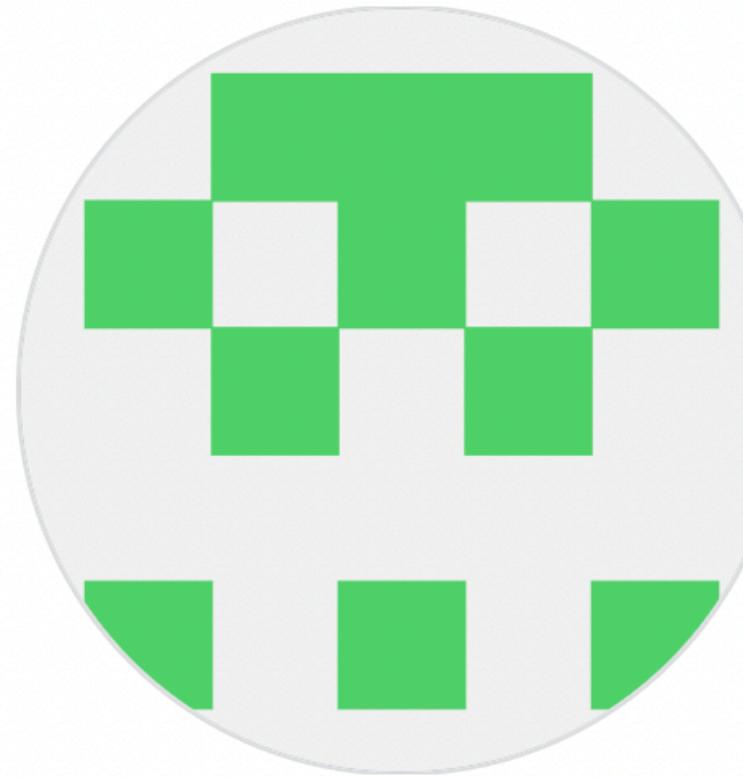
Jupyter Notebook 87 ⚡ 17

cassandrahamilton

https://github.com/cassandrahamilton

80% 

Overview Repositories 4 Projects Packages Stars 2



cassandrahamilton

Follow

1 follower · 1 following

Block or Report

cassandrahamilton / README.md

## Cassandra Hamiton @cassandrahamilton

Hi I'm an aspiring data scientist! Some of my favorite topics include: city planning, transit, and American history.

How to reach me:  
[hamilton.cassandra.m@gmail.com](mailto:hamilton.cassandra.m@gmail.com)

### Research

#### Library Placement and Inequality

Used 2020 US census data to see if there is a correlation between neighborhood racial demographics and library placement in Minneapolis, MN.  
Data work was done in R and ArcGIS -> you can read about it [here](#)

#### Madrid Digital Humanities Project

I'm currently working with the Spanish department to investigate the connotations of different monikers for the city of Madrid. We are creating a corpus from data scraped off Twiter and other online forums.

#### MinneMUDAC 2023- Minnesota Twins MLB Forecasting

Challenge: Forecast game attendance for all games in the 2023 MLB season. Provide business insights for each stadium.  
Used publicly available data, including weather data for each historical game. PCA was used for exploratory analysis and XGBoost for predictions.

#### DataCom 2022- Effect of Exercise on Mental Health

Used R to clean and analyze CDC BRFSS survey data. Specifically looked at a cohort of people with a college degree and without children to limit additional factors. -> read the poster [here](#)

#### Resume

let's “git” started

# Steps

1. Open RStudio
2. Introduce yourself to git
3. Check the `git_sitrep()` and fix things
  - a. Name and email
  - b. `git_vaccinate()`
  - c. Personal access token (PAT)
4. Create a GitHub repo for our work today
5. Make a new RStudio project and associate it with the repo
6. Create a README file
7. Commit and push changes

# 1. Open RStudio

The screenshot shows the RStudio interface with several panels:

- editor**: The leftmost panel displays an R Markdown file named "lab-1-ikea.qmd". It contains the following YAML front matter:

```
---
```

```
title: "Lab 1 - Meet the toolkit"
subtitle: "STA 210 - Spring 2022"
author: "Mine Çetinkaya-Rundel"
date: "January 10, 2022"
format: pdf
editor: visual
---
```

Below the front matter, there's a "Setup" section with the text "Load packages and data:" and a "(Top Level)" dropdown.
- environment**: The top-right panel shows the RStudio environment pane with tabs for Environment, History, Connections, Git, and Tutorial. It displays the message "Environment is empty". A red box highlights this pane.
- files**: The bottom-right panel shows the Files pane with a list of files in the directory "lab-1-mine-cetinkaya-rundel":

Name	Size	Modified
..	40 B	Jan 18, 2022, 10:37 PM
.gitignore	106.4 KB	Jan 18, 2022, 10:37 PM
data	619 B	Jan 18, 2022, 10:38 PM
lab-1-ikea.pdf	235 B	Jan 18, 2022, 10:37 PM
lab-1-ikea.qmd	83 B	Jan 18, 2022, 10:37 PM
lab-1.Rproj		
README.md		

A green box highlights this pane.
- console**: The bottom-left panel shows the RStudio console pane with the R environment information and the R help text for the "R" command.

Annotations in the image:

- "editor" is written in purple above the editor pane.
- "environment + git" is written in red above the environment pane.
- "files + plots + viewer" is written in green above the files pane.
- "console + terminal if you need it" is written in blue above the console pane.

# 2. Introduce yourself to git

This is R code.

Type it in the Console.

Everything we're doing today is through R



```
## install if needed (do this exactly once):
## install.packages("usethis")

library(usethis)
use_git_config(user.name = "Jane Doe",
               user.email = "jane@example.org")
git_vaccinate()
```

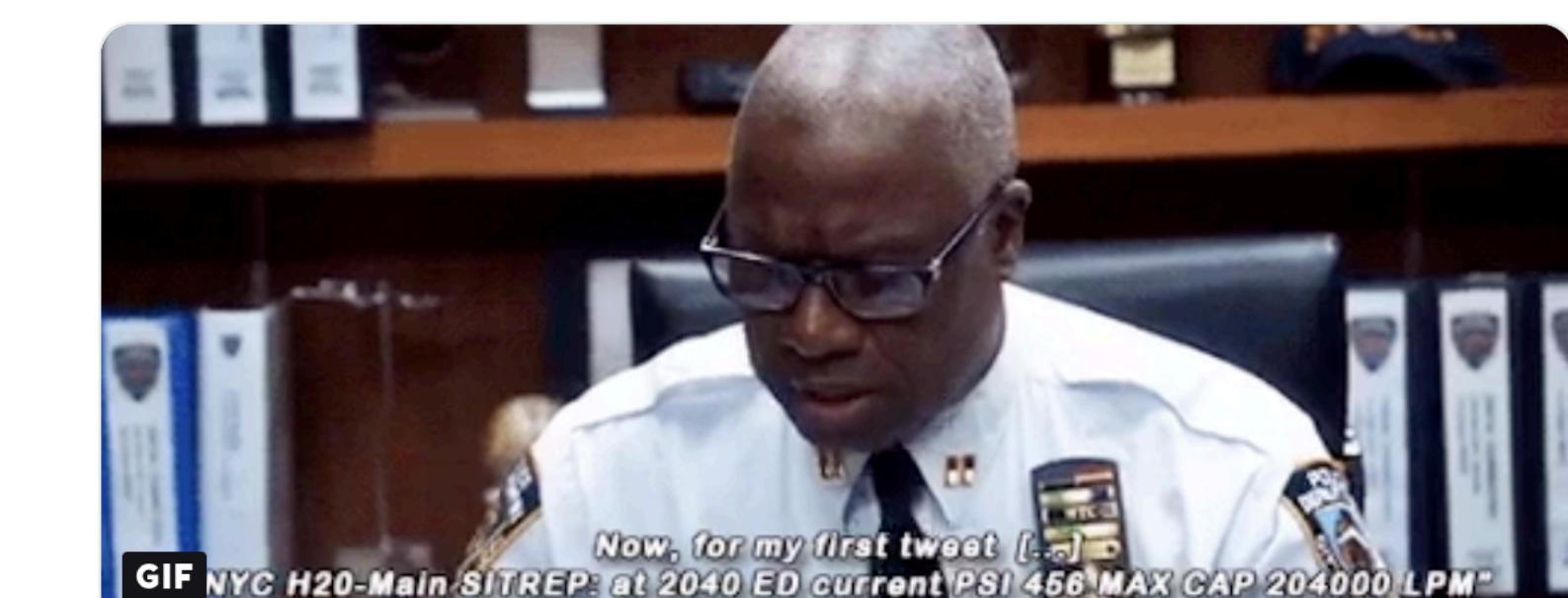
# 3.a-b. git\_sitrep()

```
> library(usethis)
> git_sitrep()
Git config (global)
• Name: 'Amelia McNamara'
• Email: 'amelia.a.mcnamara@gmail.com'
• Global (user-level) gitignore file: '~/.gitignore'
• Vaccinated: TRUE
• Default Git protocol: 'https'
• Default initial branch name: <unset>
GitHub
• Default GitHub host: 'https://github.com'
• Personal access token for 'https://github.com': '<disc
• GitHub user: 'AmeliaMN'
• Token scopes: 'gist, repo, user, workflow'
• Email(s): 'amelia.a.mcnamara@gmail.com (primary)', 'am
Git repo for current project
i No active usethis project
```

If your name, email,  
and vaccinated  
aren't set, go back  
to step 2

Mara Averick  
@dataandme

Hello people of twitter (especially the {usethis}-lovers among you).  
\*sitrep\* — noun sit·rep | \ 'sit·rep\ informal A report on the current military situation in a particular area.  
Origin — 1940s: from sit(uation) rep(ort).



6:27 AM - 18 Apr 2019

3 Retweets 20 Likes

GIF NYC H20-Main SITREP: at 2040 ED current PSI 458 MAX CAP 204000 LPM

JT >

# 3.c. git\_sitrep()

```
> library(usethis)
> git_sitrep()
Git config (global)
• Name: 'Amelia McNamara'
• Email: 'amelia.a.mcnamara@gmail.com'
• Global (user-level) gitignore file: '~/.gitignore'
• Vaccinated: TRUE
• Default Git protocol: 'https'
• Default initial branch name: <unset>
GitHub
• Default GitHub host: 'https://github.com'
• Personal access token for 'https://github.com': '<disc
• GitHub user: 'AmeliaMN'
• Token scopes: 'gist, repo, user, workflow'
• Email(s): 'amelia.a.mcnamara@gmail.com'
Git repo for current project
i No active usethis project
```

Mara Averick  
@dataandme

Hello people of twitter (especially the {usethis}-lovers among you).  
\*sitrep\* — noun sit·rep | \ 'sit·rep\ informal A report on the current military situation in a particular area.  
Origin — 1940s: from sit(uation) rep(ort).



Now, for my first tweet [...]  
GIF NYC H20-Main SITREP: at 2040 ED current PSI 458 MAX CAP 204000 LPM\*

6:27 AM - 18 Apr 2019

3 Retweets 20 Likes

I have a Personal access token (PAT) but you probably don't

### 3.c. Create a personal access token (PAT) using usethis::create\_github\_token()

A browser window will pop up and take you to GitHub

The screenshot shows the GitHub 'Personal access tokens' creation interface. On the left, a sidebar lists 'GitHub Apps', 'OAuth Apps', 'Personal access tokens' (selected), 'Fine-grained tokens', and 'Tokens (classic)'. A 'Beta' badge is visible next to 'Personal access tokens'. The main area is titled 'New personal access token (classic)' and contains a note about the function of PATs. It includes fields for 'Name' (with 'april\_mbp' entered), 'Expiration' (set to '30 days'), and a 'Select scopes' section. A large blue callout bubble points to the 'Name' input field, containing the text: 'Give it a name, leave the default check boxes, scroll down and generate token'.

Settings / Developer settings

GitHub Apps

OAuth Apps

Personal access tokens

Fine-grained tokens

Tokens (classic)

Beta

New personal access token (classic)

Personal access tokens (classic) function like ordinary OAuth access tokens. They can be used instead of a password for Git over HTTPS, or can be used to [authenticate to the API over Basic Authentication](#).

Note

april\_mbp

What's this token for?

Expiration \*

30 days

The token will expire on Sat, May 6 2023

Select scopes

Scopes define the access for personal tokens. [Read more about OAuth scopes.](#)

<input checked="" type="checkbox"/> repo	Full control of private repositories
<input checked="" type="checkbox"/> repo:status	Access commit status
<input checked="" type="checkbox"/> repo_deployment	Access deployment status
<input checked="" type="checkbox"/> public_repo	Access public repositories
<input checked="" type="checkbox"/> repo:invite	Access repository invitations
<input checked="" type="checkbox"/> security_events	Read and write security events

### 3.c. A browser window will pop up and take you to GitHub. Leave this window open

The screenshot shows the GitHub 'Personal access tokens (classic)' page under 'Developer settings'. The left sidebar has 'Personal access tokens' selected. The main area lists several tokens:

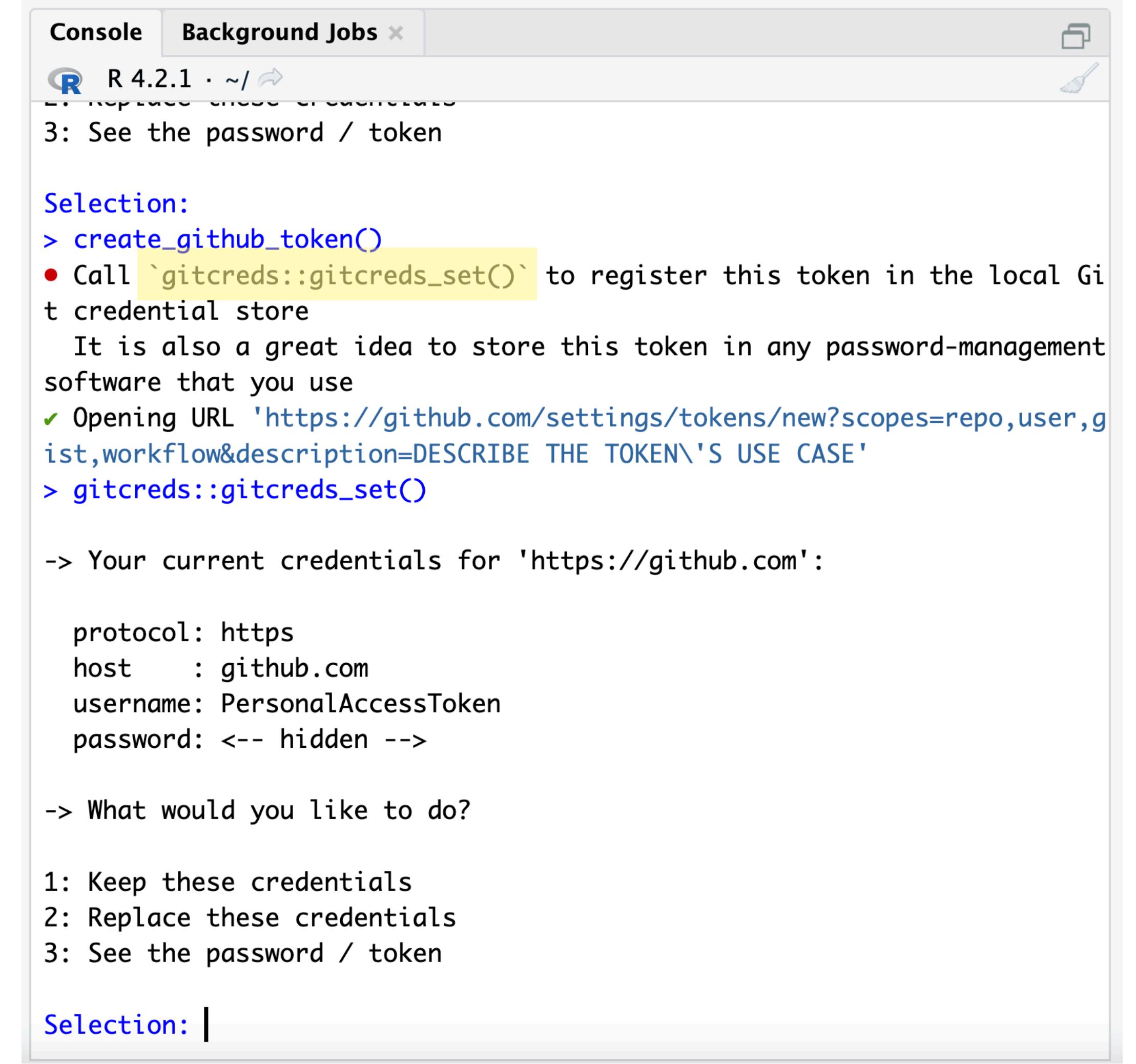
- ghp\_PZ8ZdJd9BgWTtqb2Rt1Kdm6Ar58ym10pWVyt** (highlighted with a blue oval):
  - work\_imac\_march — gist, repo, user, workflow
  - Last used within the last 2 weeks
  - Expires on Fri, Apr 28 2023.
- mbp\_march** — gist, repo, user, workflow
  - Last used within the last week
  - Expires on Fri, Apr 7 2023.
- work\_imac\_feb** — gist, repo, user, workflow
  - Last used within the last 3 weeks
  - Expired on Wed, Mar 22 2023.
- teaching\_mbp** — gist, repo, user, workflow
  - Last used within the last 2 months
  - Expired on Wed, Mar 1 2023.
- mbp\_january** — gist, repo, user, workflow
  - Last used within the last 2 months
  - Expired on Sat, Feb 25 2023.

A large blue diagonal banner across the top right of the page says 'LEAVE THIS WINDOW OPEN'. A blue callout bubble on the left side says '(I deleted this token after I took the screenshot. Don't share tokens with people!)' pointing to the first token listed.

Token Name	Scopes	Last Used	Action
ghp_PZ8ZdJd9BgWTtqb2Rt1Kdm6Ar58ym10pWVyt	gist, repo, user, workflow	Within last 2 weeks	Configure SSO   Delete
work_imac_march	gist, repo, user, workflow	Within last week	Configure SSO   Delete
mbp_march	gist, repo, user, workflow	Within last week	Configure SSO   Delete
work_imac_feb	gist, repo, user, workflow	Within last 3 weeks	Configure SSO   Delete
teaching_mbp	gist, repo, user, workflow	Within last 2 months	Configure SSO   Delete
mbp_january	gist, repo, user, workflow	Within last 2 months	Configure SSO   Delete

### 3.c. Go back to RStudio and enter the PAT

- Run `gitcreds::gitcreds_set()`
- Follow the instructions
  - Likely, copy-paste the PAT from GitHub into your Console
- Restart R (Session -> Restart R)



The screenshot shows the RStudio Console window with the following interaction:

```
R 4.2.1 · ~/Documents
3: See the password / token

Selection:
> create_github_token()
• Call `gitcreds::gitcreds_set()` to register this token in the local Git credential store
  It is also a great idea to store this token in any password-management software that you use
✓ Opening URL 'https://github.com/settings/tokens/new?scopes=repo,user,gist,workflow&description=DESCRIBE THE TOKEN\`S USE CASE'
> gitcreds::gitcreds_set()

-> Your current credentials for 'https://github.com':
  protocol: https
  host    : github.com
  username: PersonalAccessToken
  password: <-- hidden -->

-> What would you like to do?
1: Keep these credentials
2: Replace these credentials
3: See the password / token

Selection: |
```

The console output includes a selection menu at the bottom:

- 1: Keep these credentials
- 2: Replace these credentials
- 3: See the password / token

# 3. git\_sitrep()

```
> library(usethis)
> git_sitrep()
Git config (global)
• Name: 'Amelia McNamara'
• Email: 'amelia.a.mcnamara@gmail.com'
• Global (user-level) gitignore file: '~/.gitignore'
• Vaccinated: TRUE
• Default Git protocol: 'https'
• Default initial branch name: <unset>
GitHub
• Default GitHub host: 'https://github.com'
• Personal access token for 'https://github.com': '<disc
• GitHub user: 'AmeliaMN'
• Token scopes: 'gist, repo, user, workflow'
• Email(s): 'amelia.a.mcnamara@gmail.com (primary)', 'am
Git repo for current project
i No active usethis project
```

Mara Averick  
@dataandme

Following

Hello people of twitter (especially the {usethis}-lovers among you).  
\*sitrep\* — noun sit·rep | \ 'sit·rep\ informal A report on the current military situation in a particular area.  
Origin — 1940s: from sit(uation) rep(ort).

Now, for my first tweet [...]

GIF NYC H20-Main SITREP: at 2040 ED current PSI 458 MAX CAP 204000 LPM

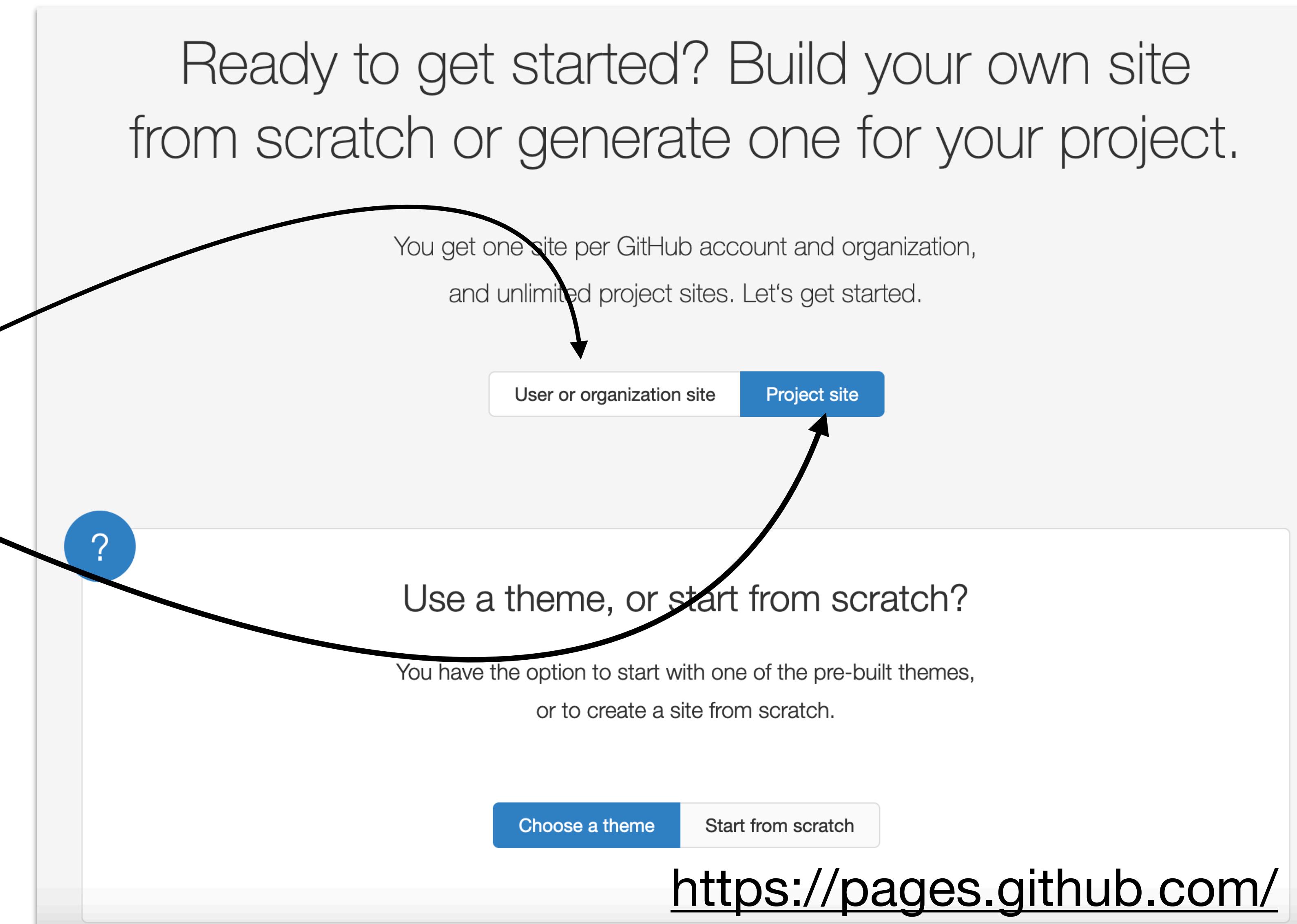
6:27 AM - 18 Apr 2019

3 Retweets 20 Likes

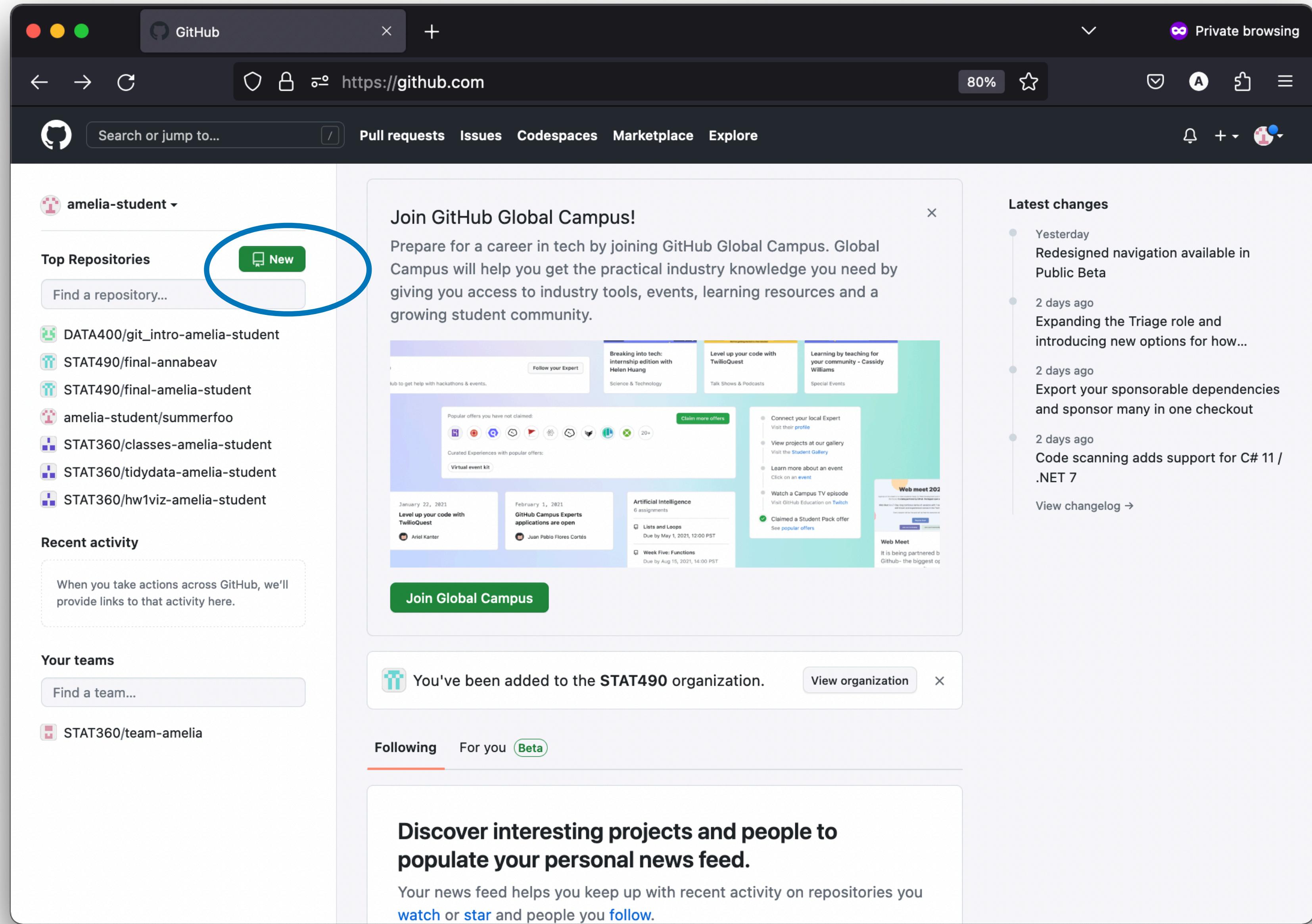
Reply 3 Retweet 20 Like 10 More

# 4. Make a GitHub repo for our work today

- Two choices:
  - Make a repo/website for you
  - Make a repo/website for a project you have done
- I'm going to demo the project site option, but the user site option is very similar. Either way we need a repo



# 4. Make a GitHub repo for our work today



The screenshot shows a GitHub user profile for "amelia-student". On the left, there's a sidebar with "Top Repositories" containing links to various repos like "DATA400/git\_intro-amelia-student" and "STAT360/tidydata-amelia-student". Below that is a "Recent activity" section and a "Your teams" section. A prominent green "New" button is circled in blue at the top of the sidebar. The main content area features a "Join GitHub Global Campus!" modal with text about the program and a "Join Global Campus" button. Below it is a notification about being added to the "STAT490" organization. At the bottom, there's a "Discover interesting projects and people to populate your personal news feed." section.

Top Repositories

- DATA400/git\_intro-amelia-student
- STAT490/final-annabeav
- STAT490/final-amelia-student
- amelia-student/summerfoo
- STAT360/classes-amelia-student
- STAT360/tidydata-amelia-student
- STAT360/hw1viz-amelia-student

Recent activity

When you take actions across GitHub, we'll provide links to that activity here.

Your teams

Find a team...

STAT360/team-amelia

Join GitHub Global Campus!

Prepare for a career in tech by joining GitHub Global Campus. Global Campus will help you get the practical industry knowledge you need by giving you access to industry tools, events, learning resources and a growing student community.

Join Global Campus

You've been added to the STAT490 organization.

View organization

Following For you Beta

Discover interesting projects and people to populate your personal news feed.

Your news feed helps you keep up with recent activity on repositories you watch or star and people you follow.

# 4. Make a GitHub repo for our work today

Repos can be public (anyone can see them) or private. Public is better for what we're doing, but you can change the visibility later

Create a new repository

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

Owner \*      Repository name \*

amelia-student / mn\_psas

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

Description (optional)

A data analysis project exploring the most common PSAs shown on Minnesota highway signs

Public      Anyone on the internet can see this repository. You choose who can commit.

Private      You choose who can see and commit to this repository.

Initialize this repository with:

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

Add a README file      This is where you can write a long description for your project. [Learn more](#).

Add .gitignore

Choose which files not to track from a list of templates. [Learn more](#)

.gitignore template: R

Choose a license

A license tells others what they can and can't do with your code. [Learn more](#)

License: None

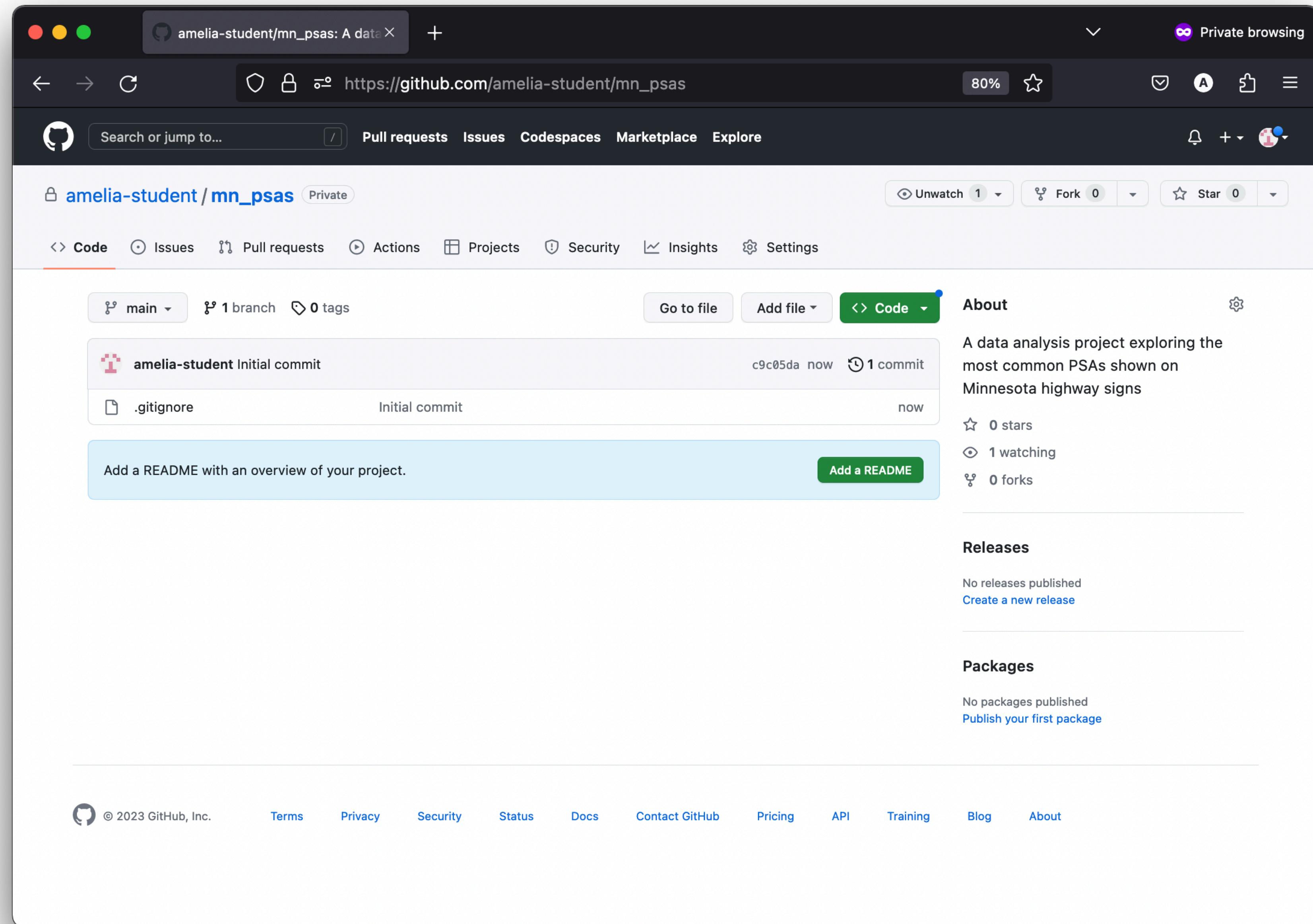
(i) You are creating a private repository in your personal account.

The name will become part of your URL

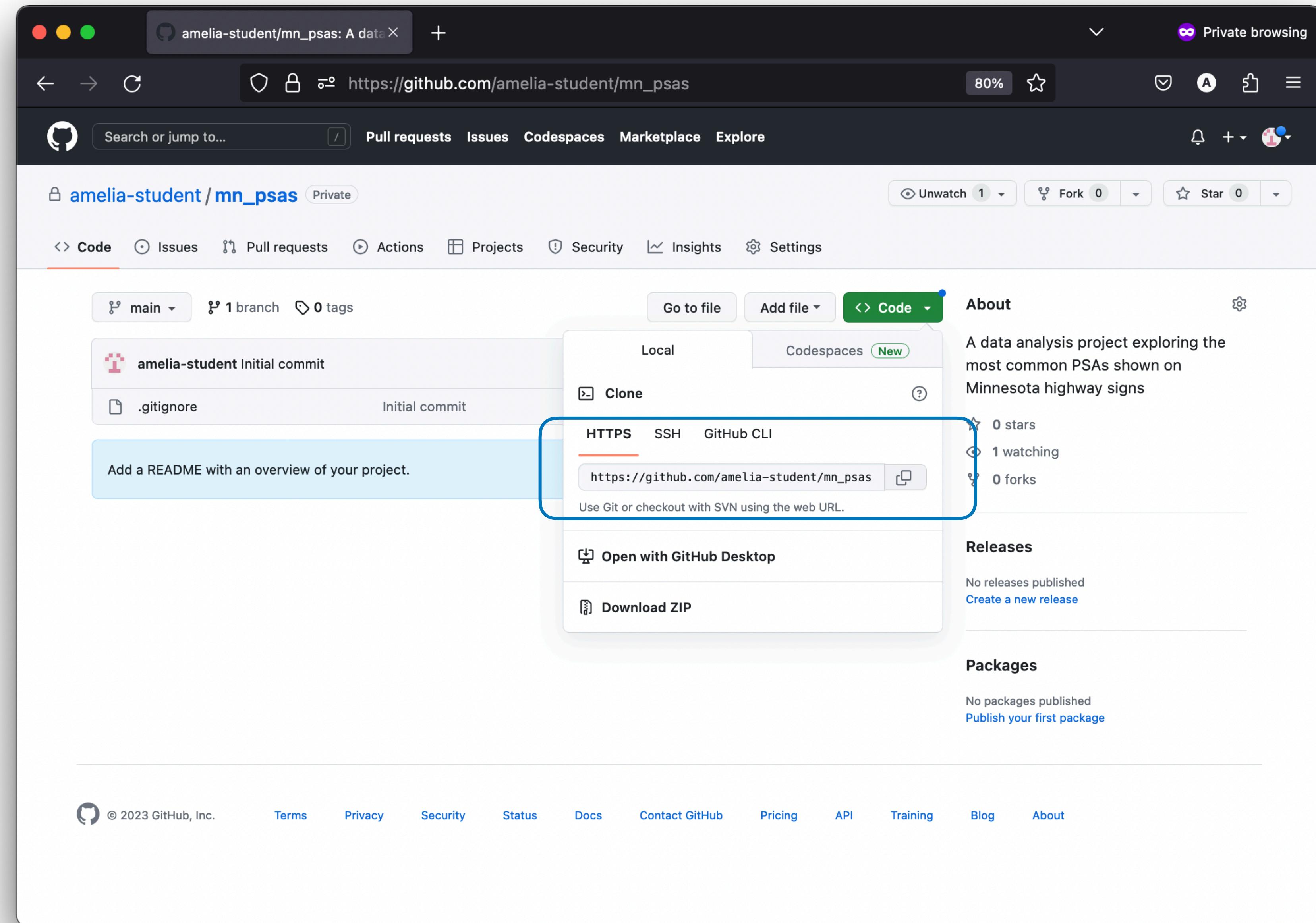
It's nice to have a description

I added an R .gitignore template, if you have a Python project you might want a Python .gitignore template, etc

# 4. Make a GitHub repo for our work today

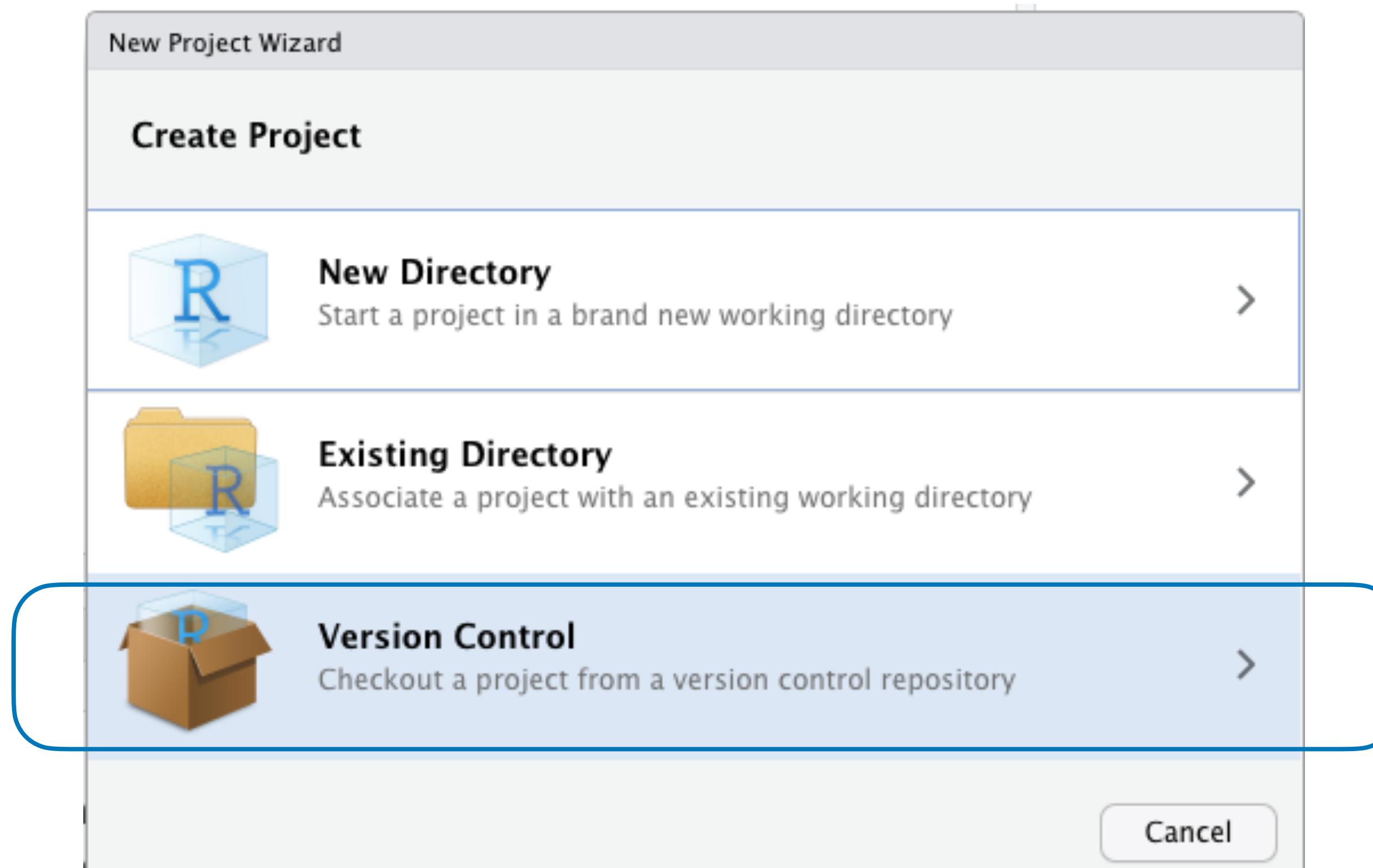


# 4.a. Copy the URL from your repo

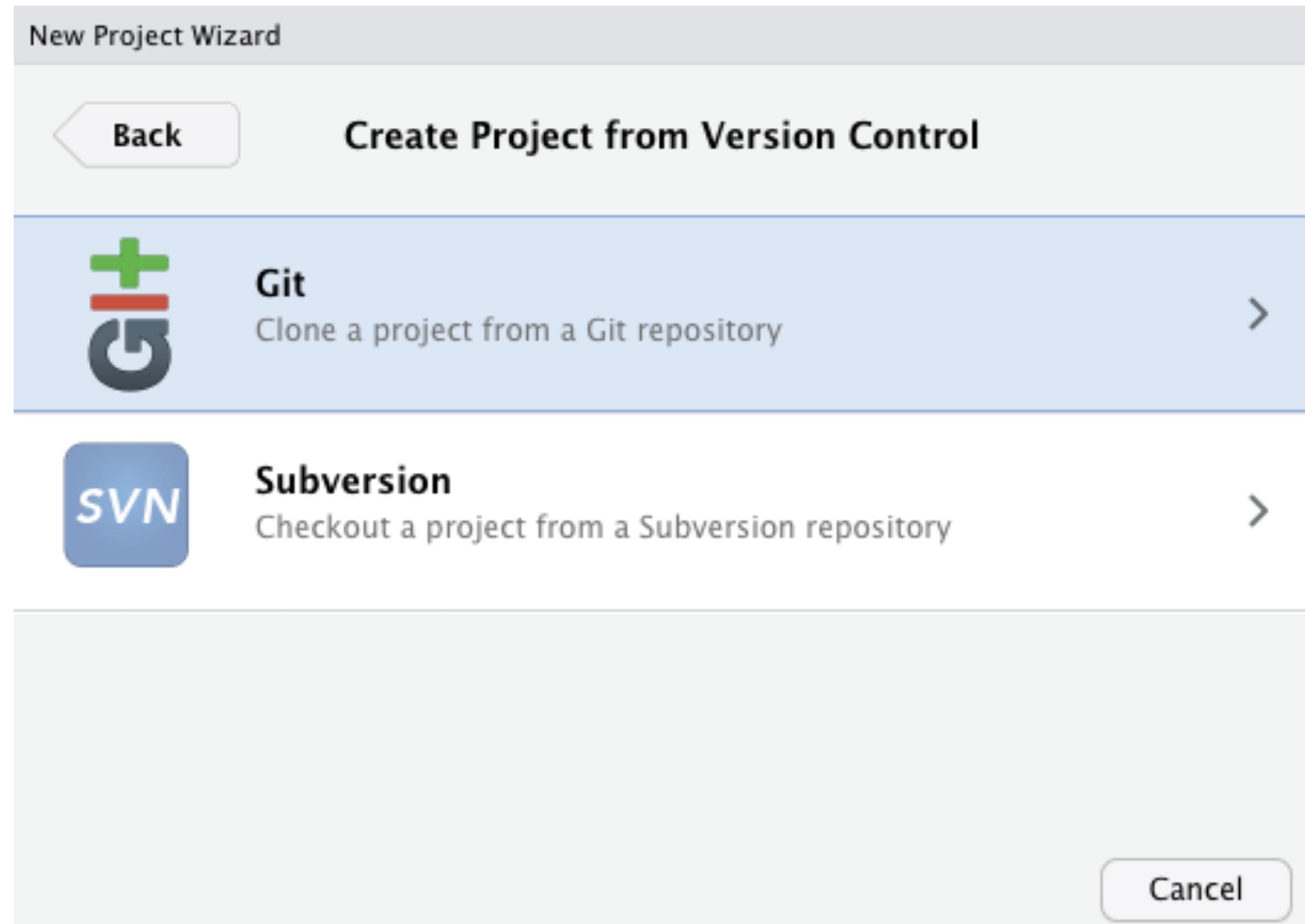


# 5. Make a new RStudio Project with Version Control

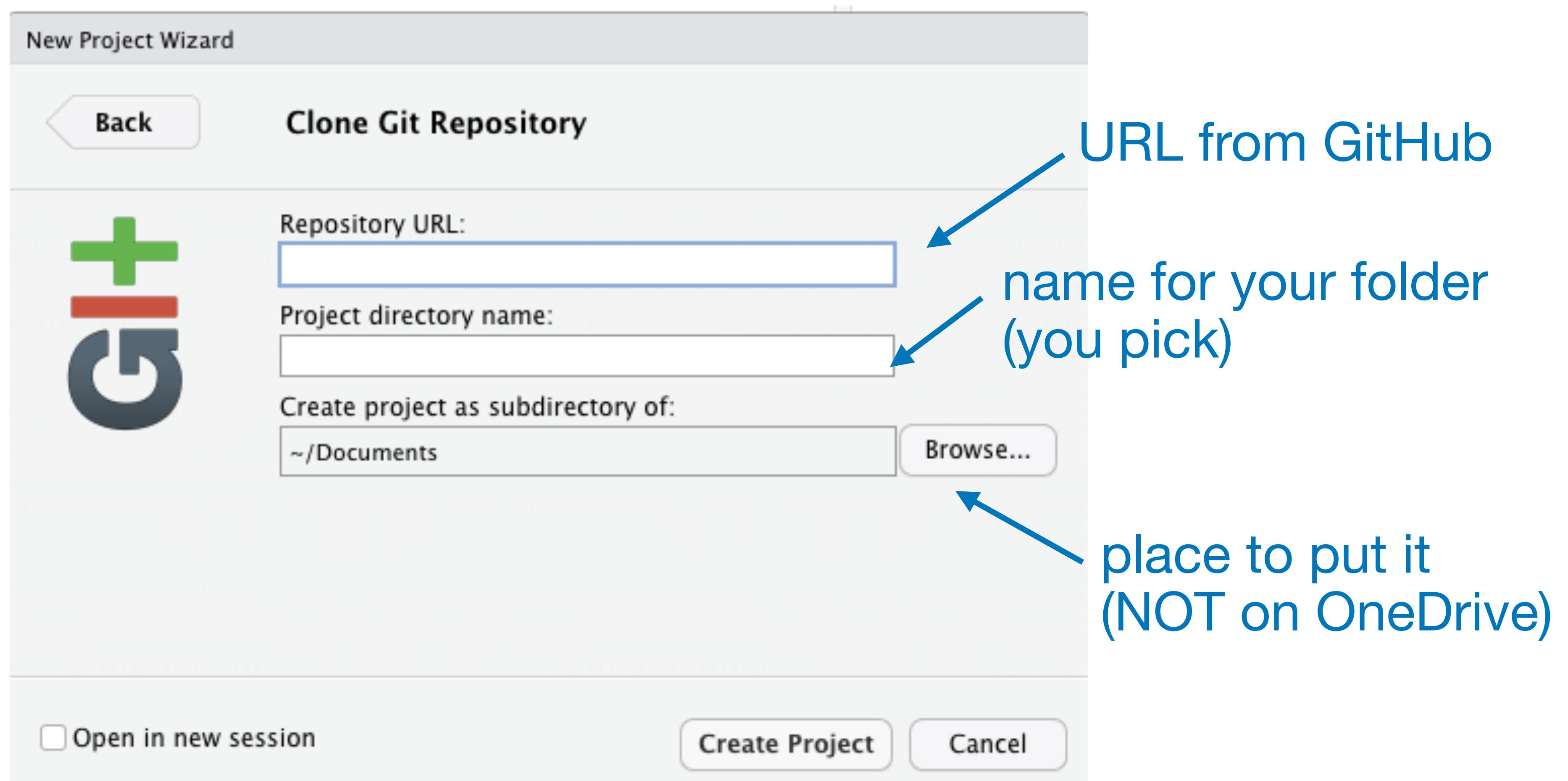
File -> New Project -> Version Control



# 5. Make a new RStudio Project with Version Control

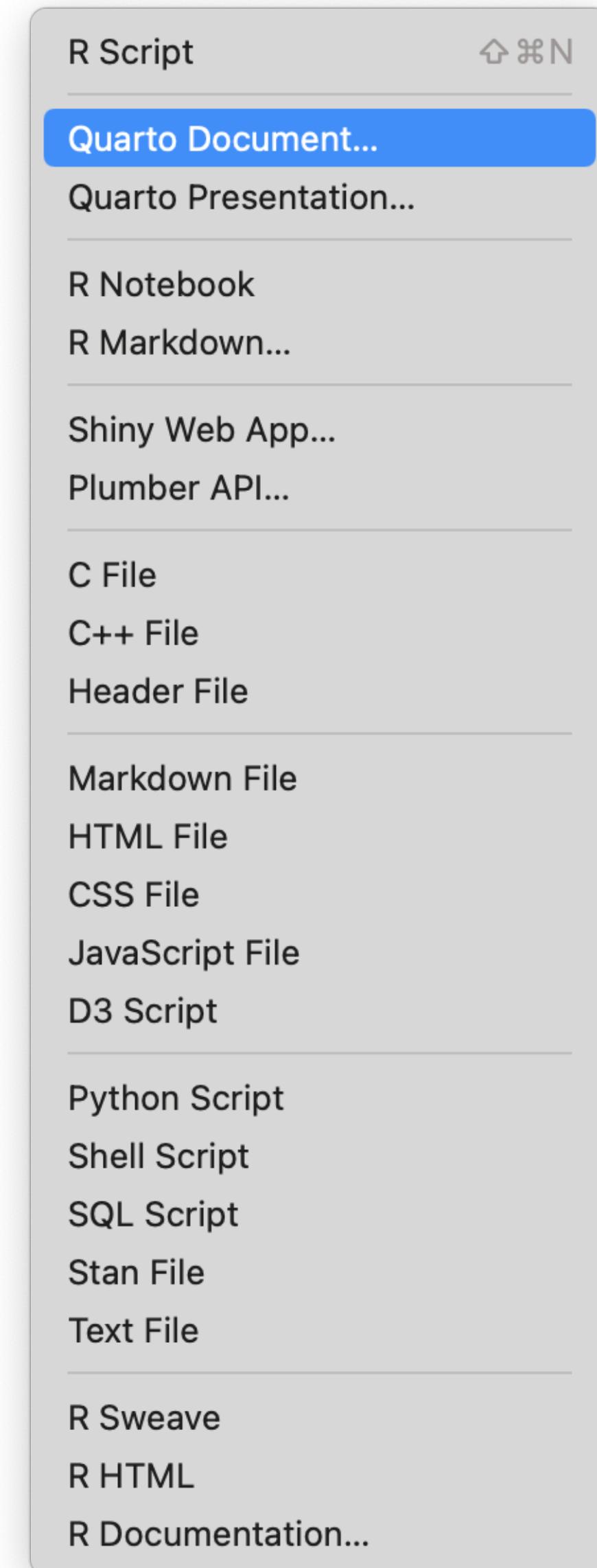


# 5. Make a new RStudio Project with Version Control

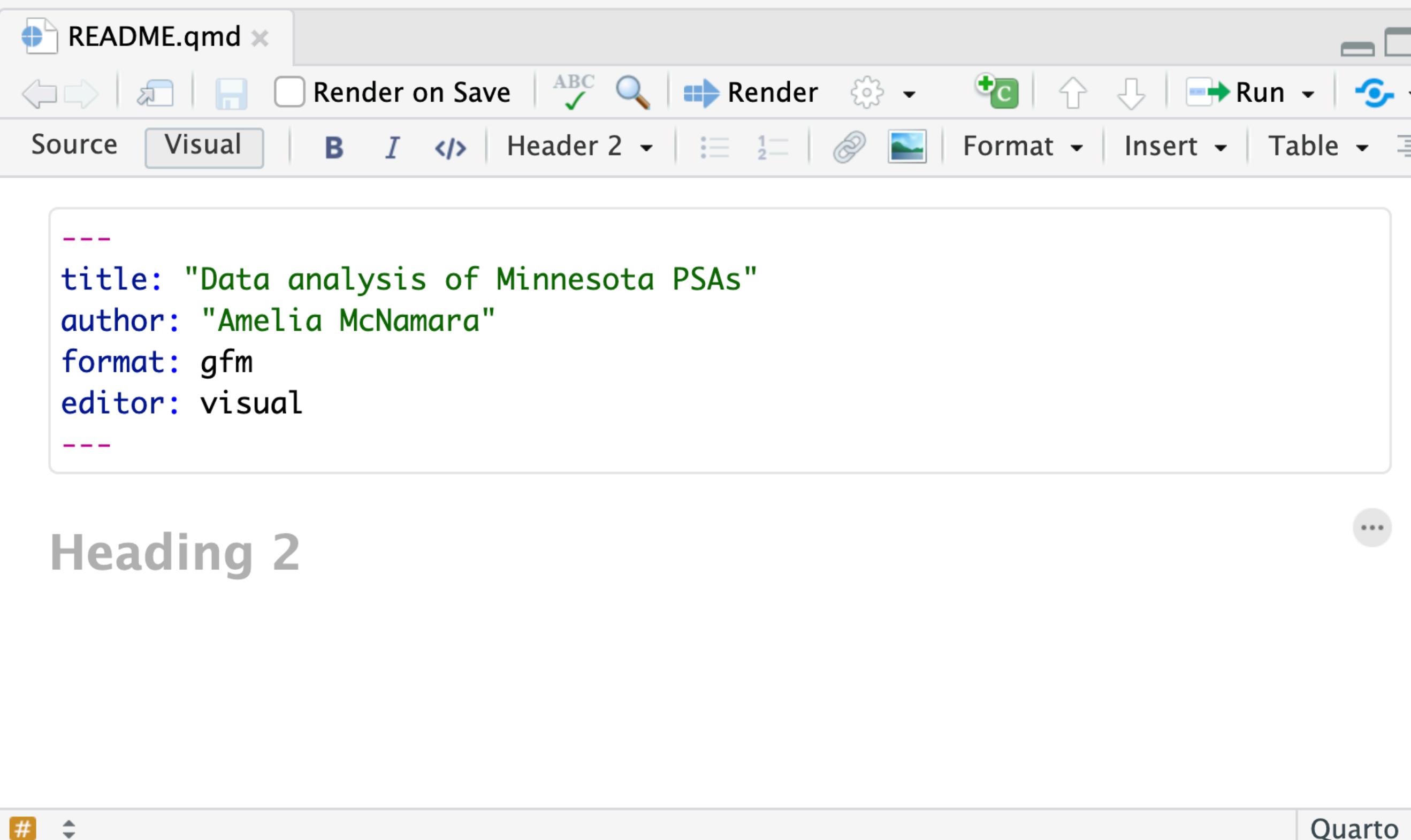


# 6. Create a file called README

- File -> New File
- Pick a good filetype for your project:
  - Quarto will work for either R or Python, so I suggest that
  - But, RMarkdown or Markdown will work as well
- Save the file as README (all caps) with the appropriate file extension



# 6. Use format: gfm



The screenshot shows the Quarto interface with a file named "README.qmd". The front matter contains:

```
---
title: "Data analysis of Minnesota PSAs"
author: "Amelia McNamara"
format: gfm
editor: visual
---
```

A "Heading 2" is visible below the front matter.

## GFM Format

Use the `gfm` format to create GitHub Flavored Markdown from Quarto. For example:

```
---
title: "My Project"
format: gfm
---
```

See the GFM [format reference](#) for a complete list of all options available for GFM output.

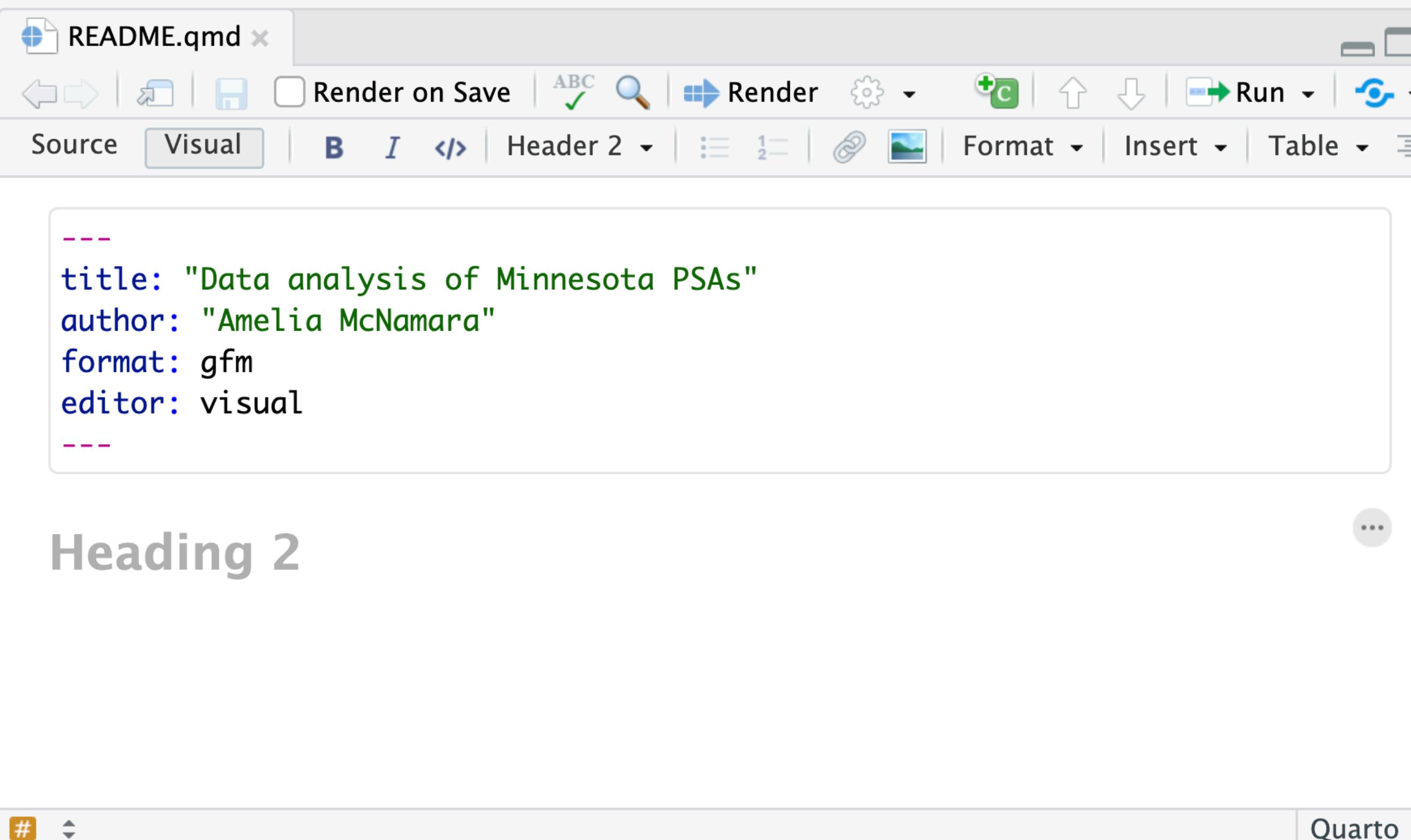
To create a `README.md` using Quarto, start with a notebook (`.ipynb`) or computational markdown file (`.qmd`) that has `README` as its file name stem, for example:

```
README.qmd
---
title: "My Project"
format: gfm
jupyter: python3
---
```

This is a GitHub README that has content dynamically generated from Python:

```
```{python}
1 + 1
...``
```

# 6. Render document and look for .md file



The screenshot shows the Quarto interface with a document titled "README.qmd". The document content includes:

```
---  
title: "Data analysis of Minnesota PSAs"  
author: "Amelia McNamara"  
format: gfm  
editor: visual  
---  


## Heading 2


```

The interface has a toolbar at the top with various icons for file operations, rendering, and styling. The status bar at the bottom right says "Quarto".

## GFM Format

Use the `gfm` format to create GitHub Flavored Markdown from Quarto. For example:

```
---  
title: "My Project"  
format: gfm  
---
```

See the GFM [format reference](#) for a complete list of all options available for GFM output.

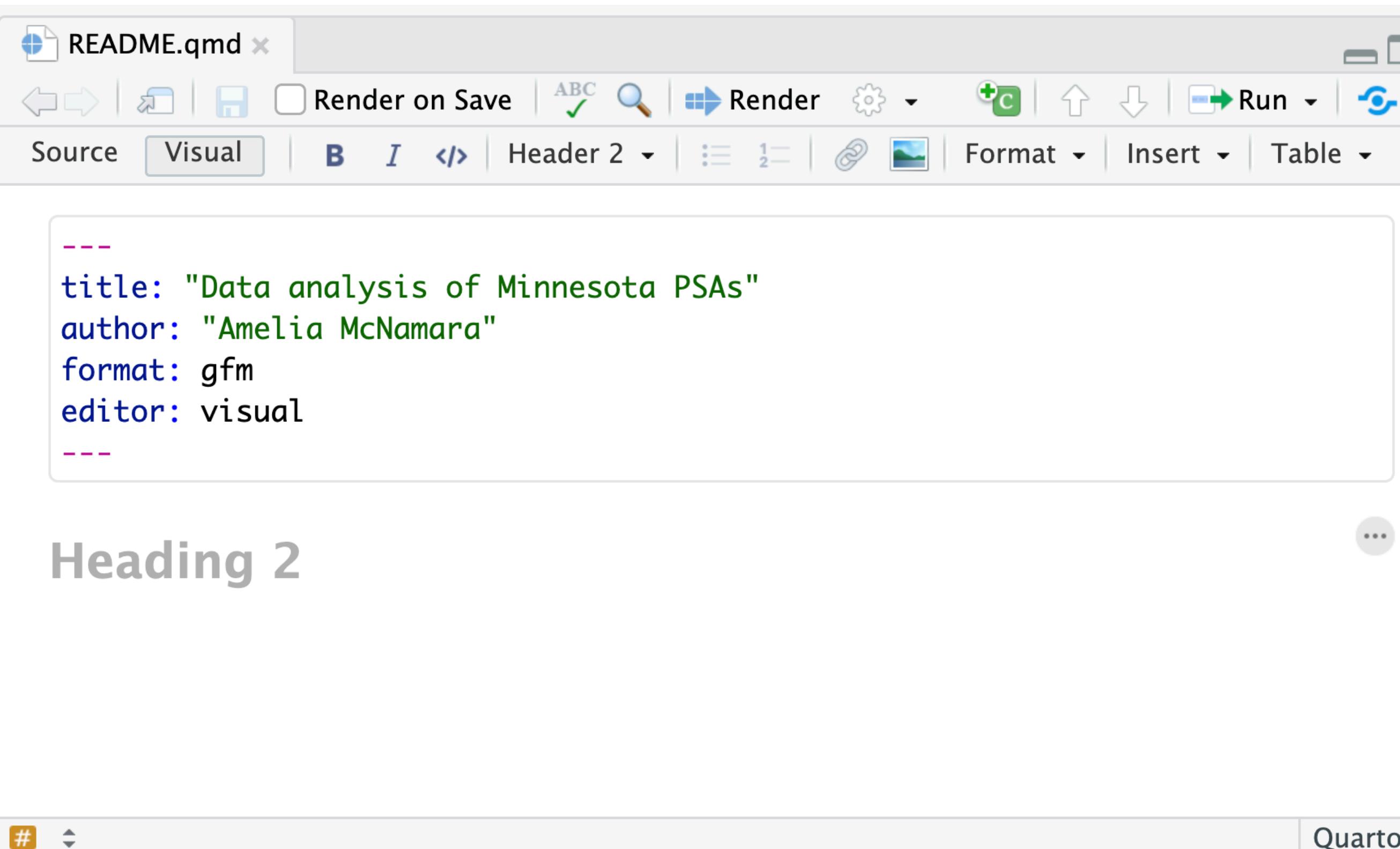
To create a `README.md` using Quarto, start with a notebook (`.ipynb`) or computational markdown file (`.qmd`) that has `README` as its file name stem, for example:

```
README.qmd  
---  
title: "My Project"  
format: gfm  
jupyter: python3  
---
```

This is a GitHub README that has content dynamically generated from Python:

```
```{python}  
1 + 1  
...``
```

# 6. Optional— stick some data analysis in there!



The screenshot shows the Quarto interface with a file named "README.qmd" open. The top bar includes standard file operations like "Render on Save" and "Render". Below the bar, the toolbar has buttons for Source, Visual, bold, italic, header levels, and other document controls. The main content area contains the following YAML front matter:

```
---  
title: "Data analysis of Minnesota PSAs"  
author: "Amelia McNamara"  
format: gfm  
editor: visual  
---
```

Below the front matter is a "Heading 2" section.

## GFM Format

Use the `gfm` format to create GitHub Flavored Markdown from Quarto. For example:

```
---  
title: "My Project"  
format: gfm  
---
```

See the GFM [format reference](#) for a complete list of all options available for GFM output.

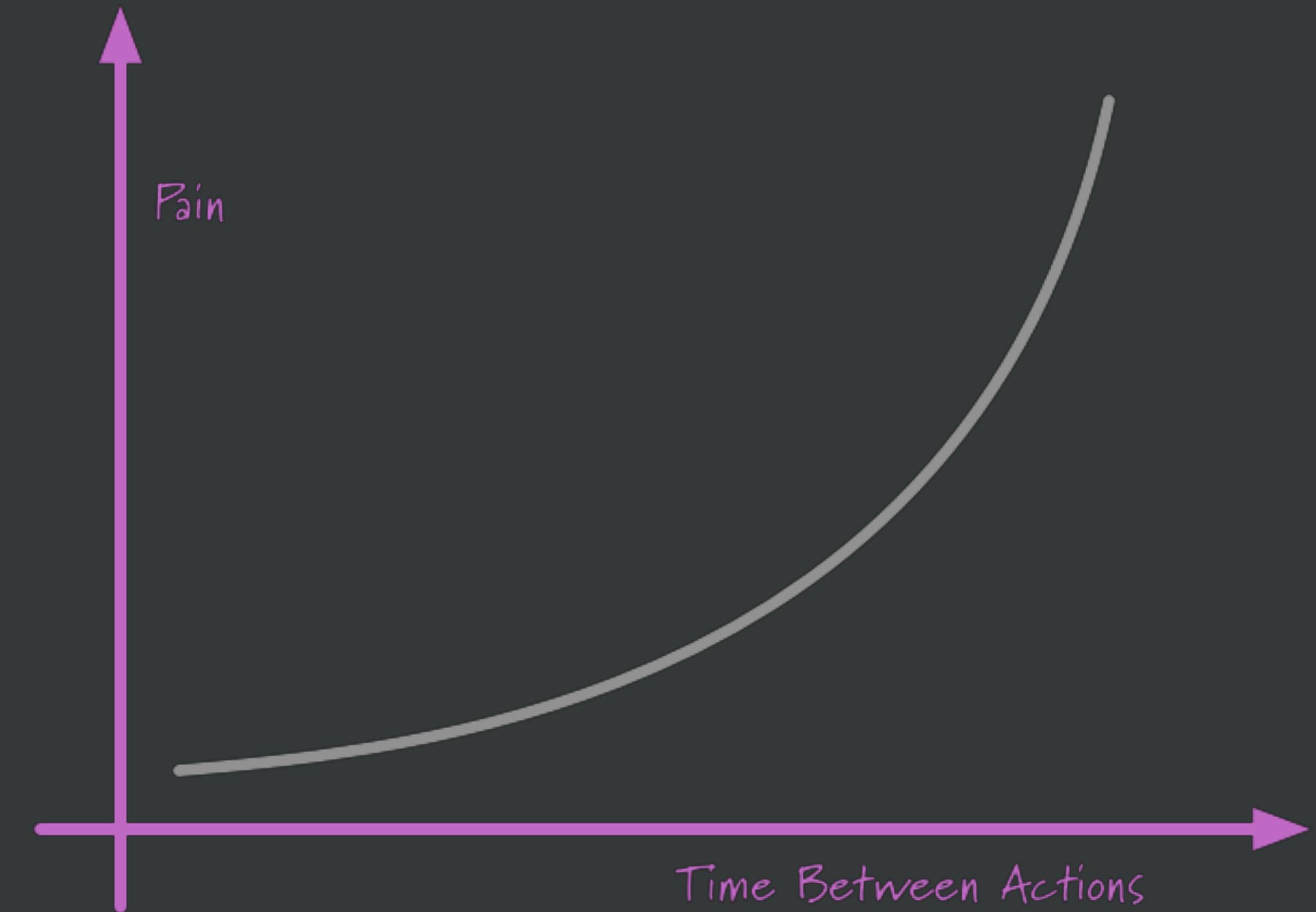
To create a `README.md` using Quarto, start with a notebook (`.ipynb`) or computational markdown file (`.qmd`) that has `README` as its file name stem, for example:

```
README.qmd  
---  
title: "My Project"  
format: gfm  
jupyter: python3  
---
```

This is a GitHub README that has content dynamically generated from Python:

```
```{python}  
1 + 1  
...``
```

# "If it hurts, do it more often."



<https://martinfowler.com/bliki/FrequencyReducesDifficulty.html>

"If it hurts, do it more often."

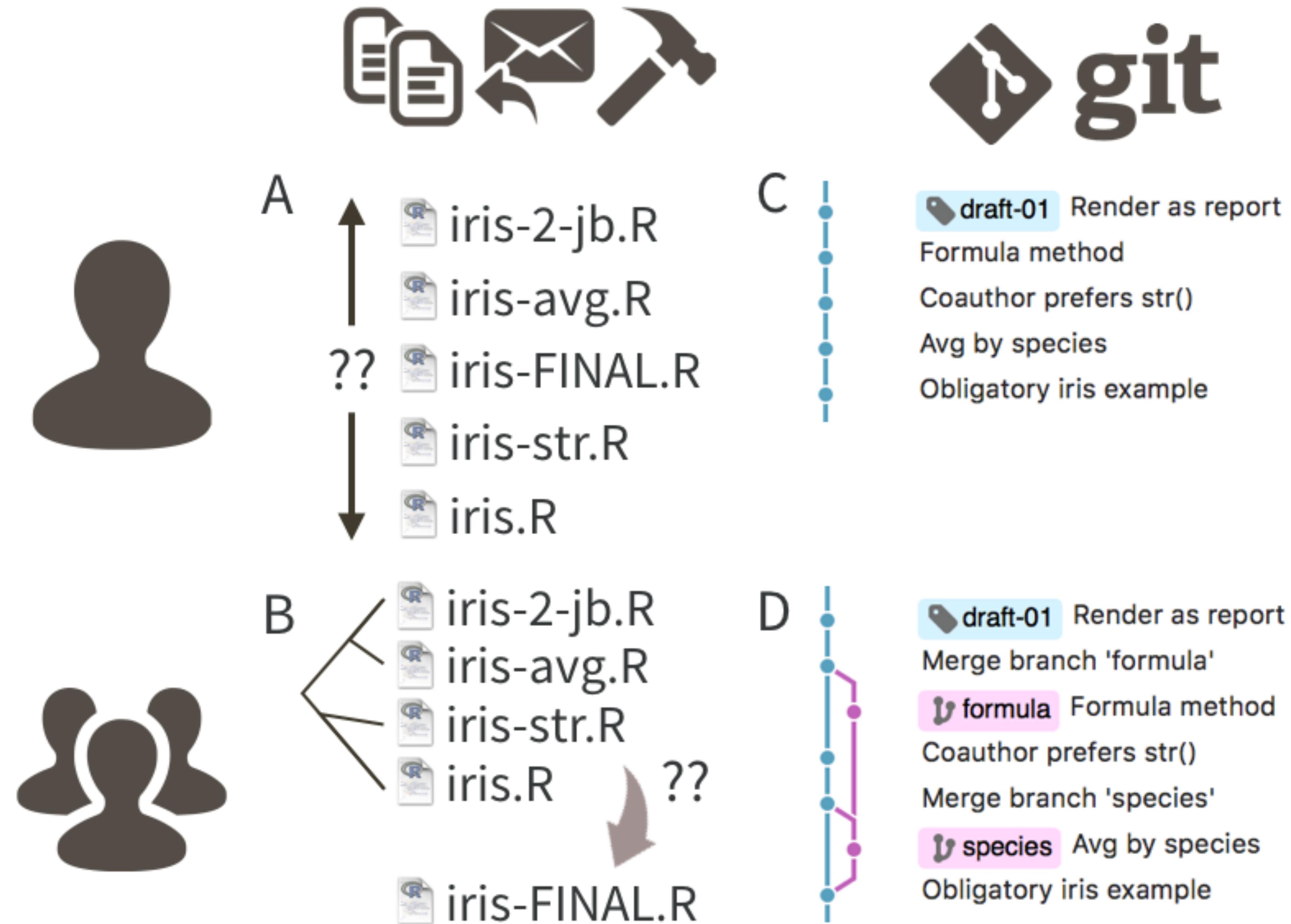
Apply this to git commit, pull, merge, push.  
(and restarting R, re-running your scripts)

Why?

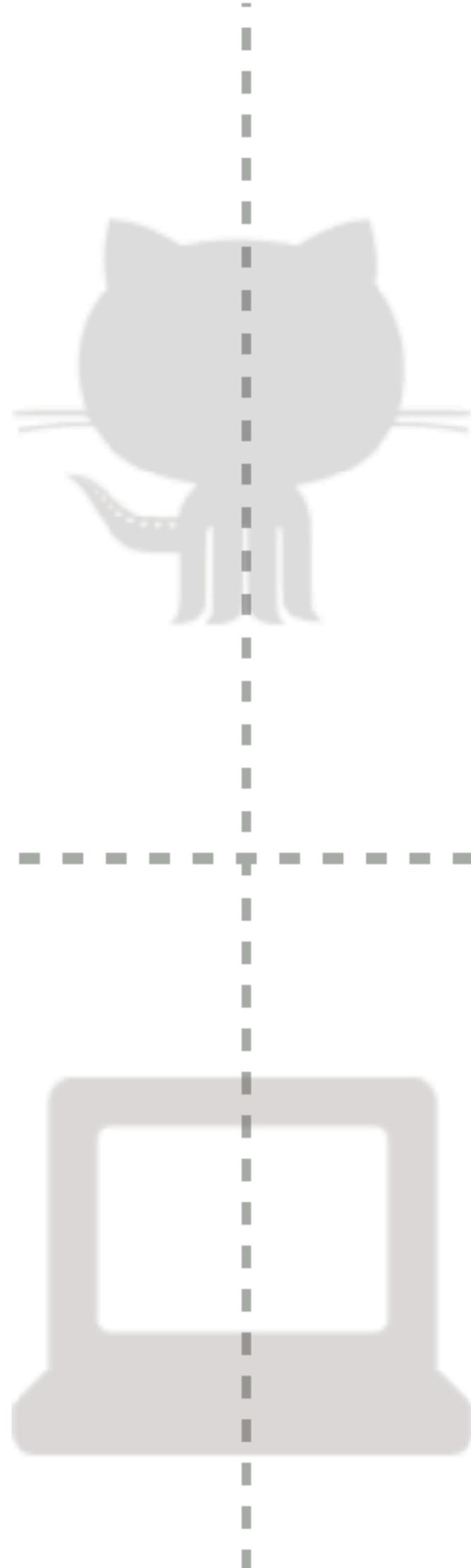
Take your pain in smaller pieces.

Tight feedback loop can reduce absolute pain.

Practice changes what you find painful.



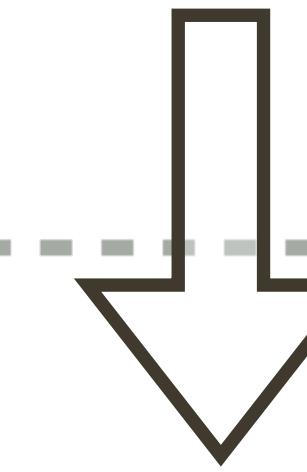
Them



You



origin

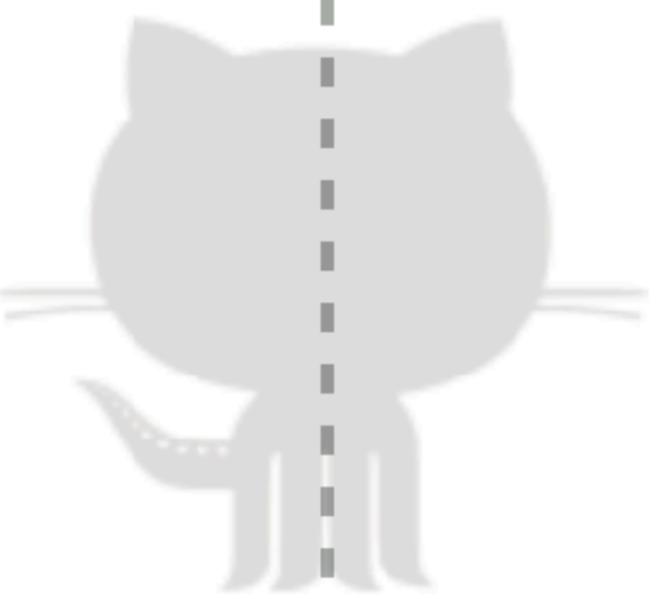


not your  
problem

We've done this by using the New Project -> From git repo workflow

Technically, this is called “cloning” a repo

Them



You

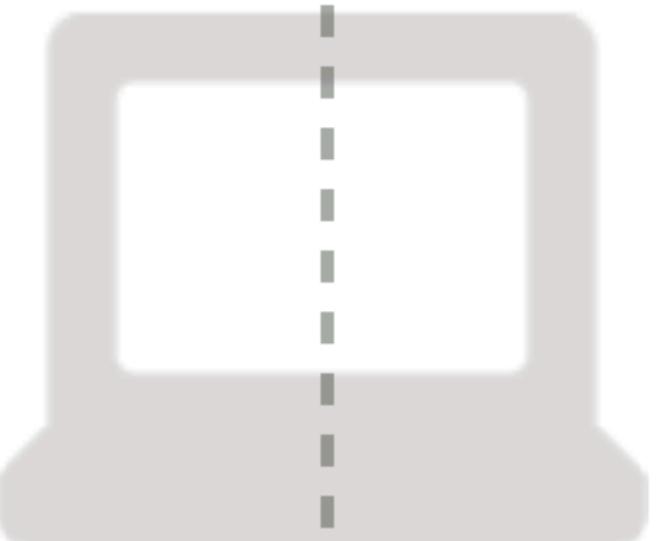


origin

push ↑ ↓ pull



not your  
problem



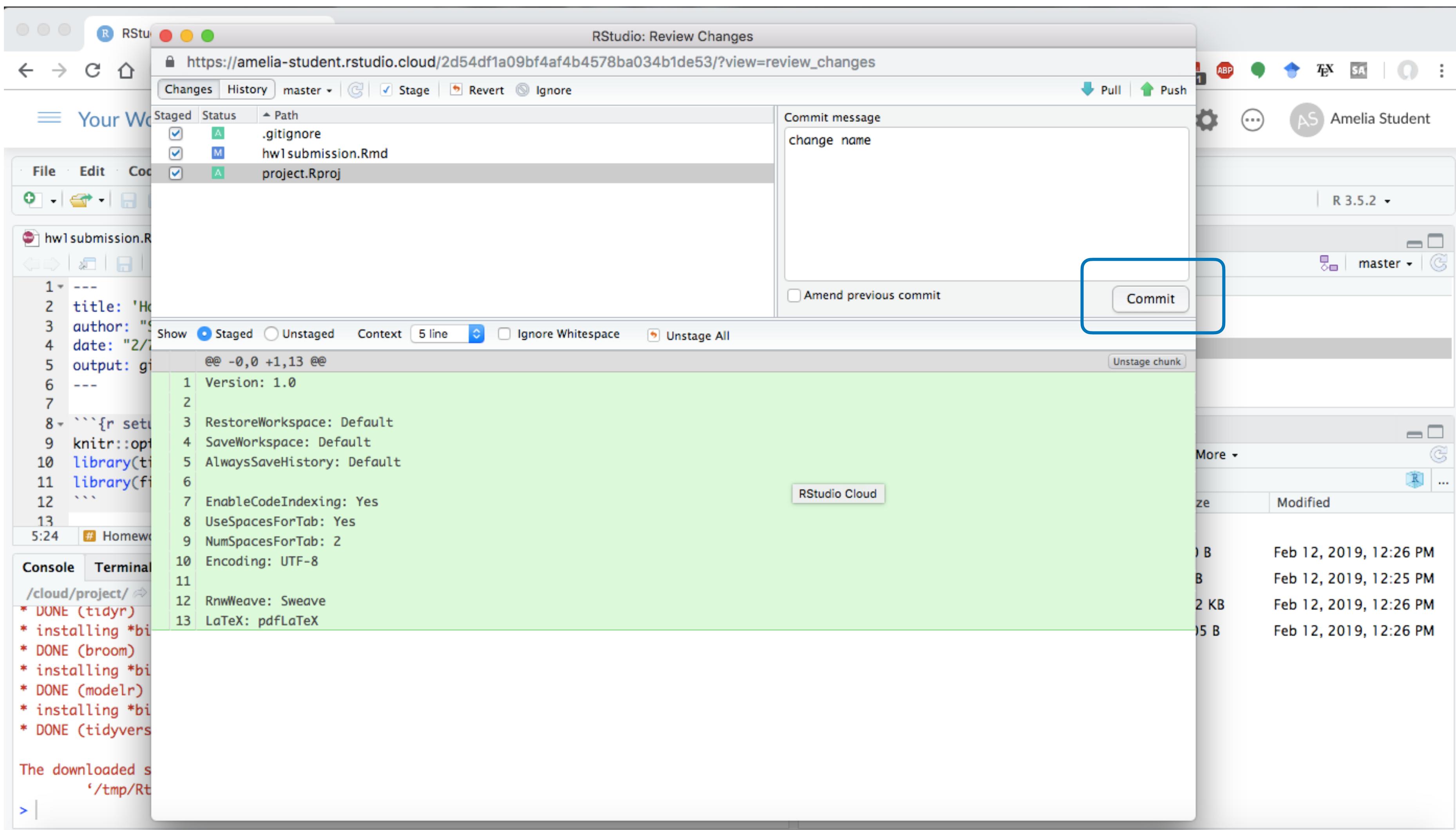
daily work, your stuff

# 7. Commit and push changes

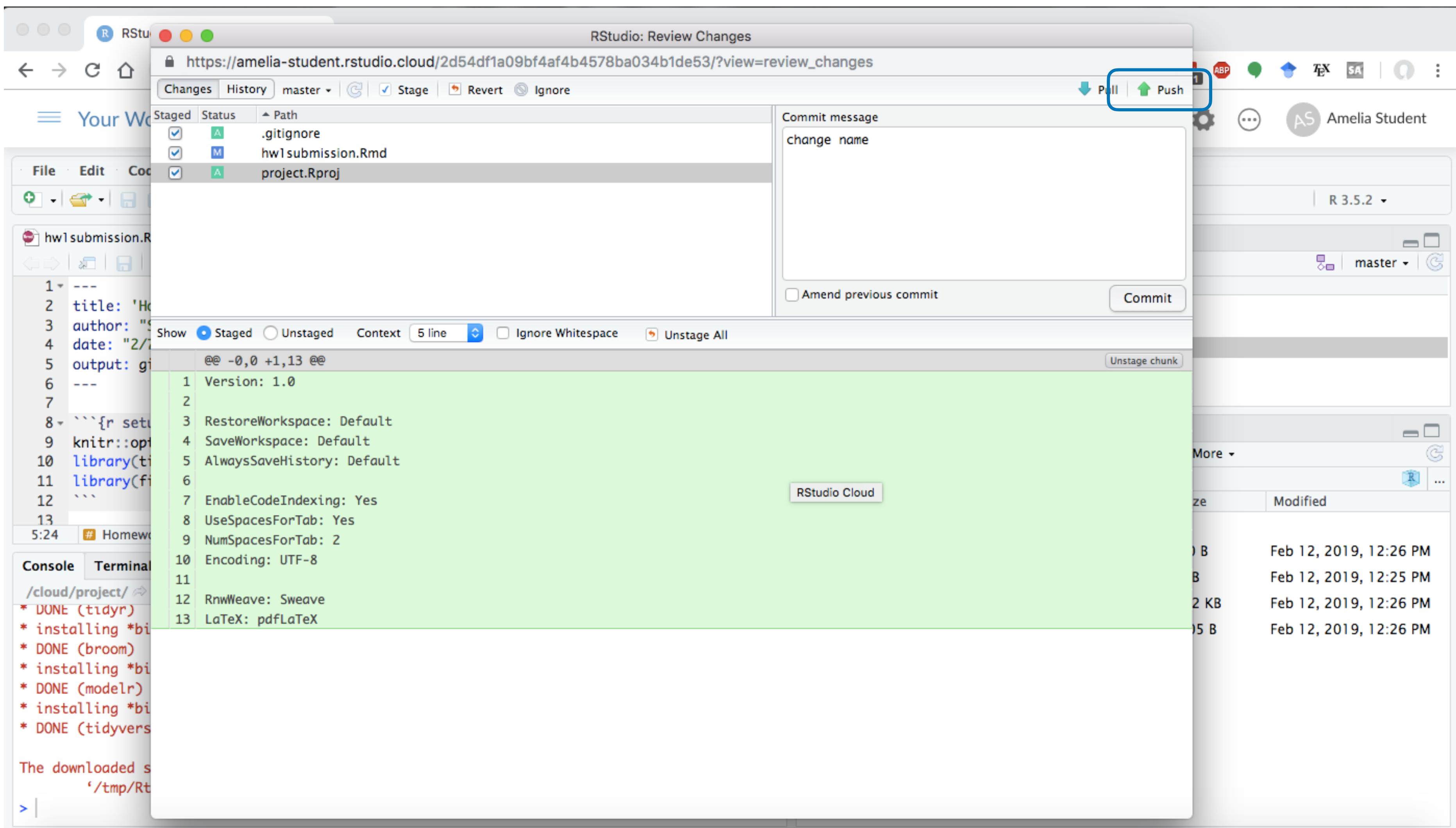
The screenshot shows the RStudio interface with the following components:

- File Menu:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Toolbar:** Go to file/function, Addins, R 3.5.2.
- Code Editor:** hw1submission.Rmd, showing R Markdown code.
- Console:** Output of R session showing project setup and R's collaborative nature.
- Git Panel:** Environment, History, Connections, Git tabs. The Git tab is selected and highlighted with a blue box. It shows a list of staged files: .gitignore (A), hw1submission.Rmd (M), and project.Rproj (A). The "Commit" button is also highlighted.
- Cloud Panel:** Shows a "Cloud > project" section with a list of files: .., .gitignore, .Rhistory, hw1submission.Rmd, and project.Rproj, each with its size and last modified date.

# 7. Commit— write a message



# 7. Push



# 7. Go check online

The screenshot shows a GitHub repository page for a private repository named `STAT360/hw1viz-amelia-student`. The repository was created by GitHub Classroom. It contains 3 commits, 1 branch, 0 releases, and 1 contributor. The latest commit was made 32 minutes ago. The repository includes files like `hw1submission_files/figure-gfm`, `.gitignore`, `hw1submission.Rmd`, `hw1submission.md`, and `project.Rproj`. A message encourages adding a README file.

STAT360 / hw1viz-amelia-student Private

3 commits 1 branch 0 releases 1 contributor

Branch: master New pull request Create new file Upload files Find file Clone or download

amelia-student first stab at a bar chart	Latest commit 53417a9 32 minutes ago
hw1submission_files/figure-gfm	first stab at a bar chart
.gitignore	changed name and date
hw1submission.Rmd	first stab at a bar chart
hw1submission.md	first stab at a bar chart
project.Rproj	changed name and date

Add a README

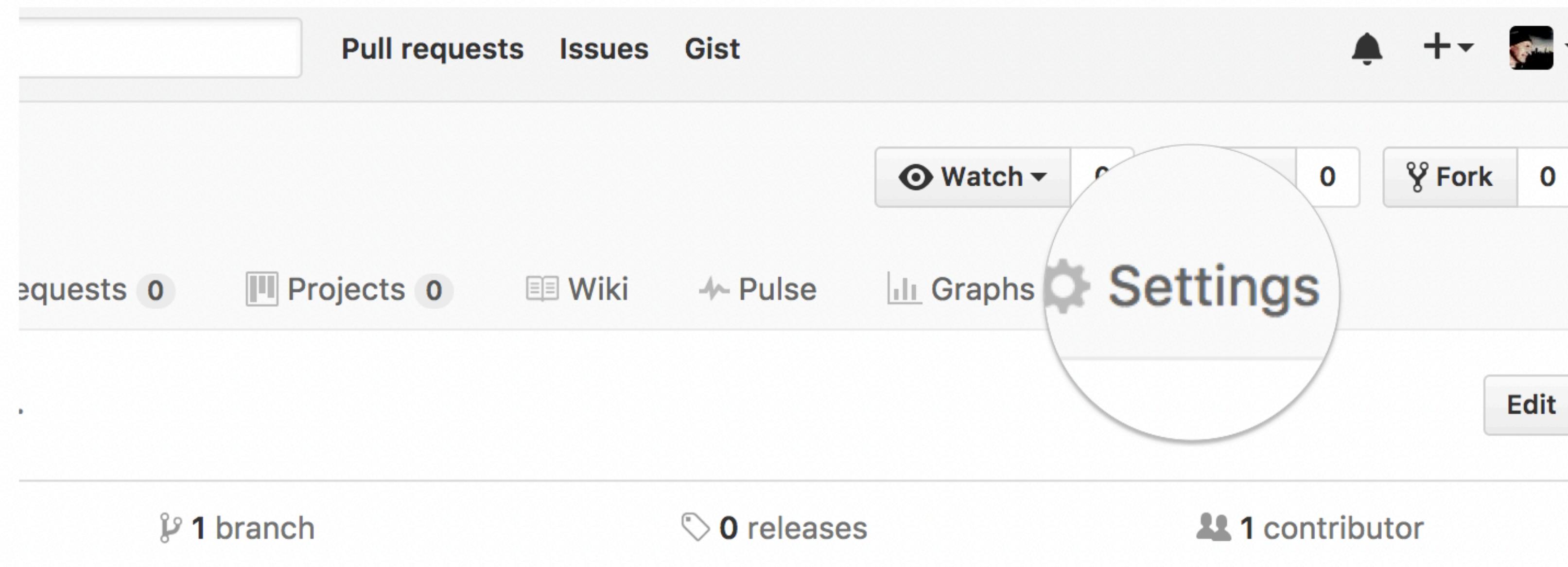
© 2019 GitHub, Inc. Terms Privacy Security Status Help Contact GitHub Pricing API Training Blog About

1

## Repository Settings

Head over to [GitHub.com](#) and create a new repository, or go to an existing one.

**Click on the Settings tab.**



2

## Theme chooser

Scroll down to the **GitHub Pages** section. Press **Choose a theme**.

### GitHub Pages

GitHub Pages is designed to host your personal, organization, or project pages from a GitHub repository.

#### Source

GitHub Pages is currently disabled. Select a source below to enable GitHub Pages for this repository. [Learn more](#).

None ▾

Save

#### Theme chooser

Select a theme to build your site with a Jekyll theme using the `master` branch. [Learn more](#).

Choose a theme

3

## Pick a theme

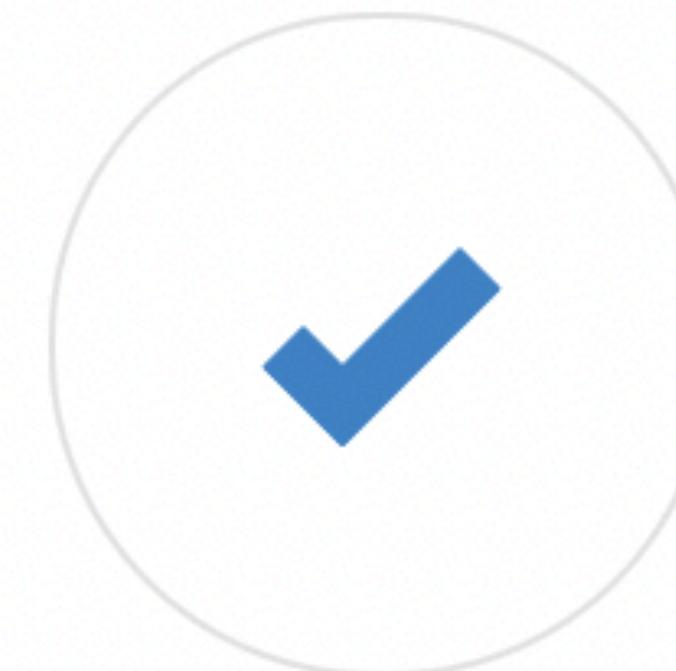
Choose one of the themes from the carousel at the top.

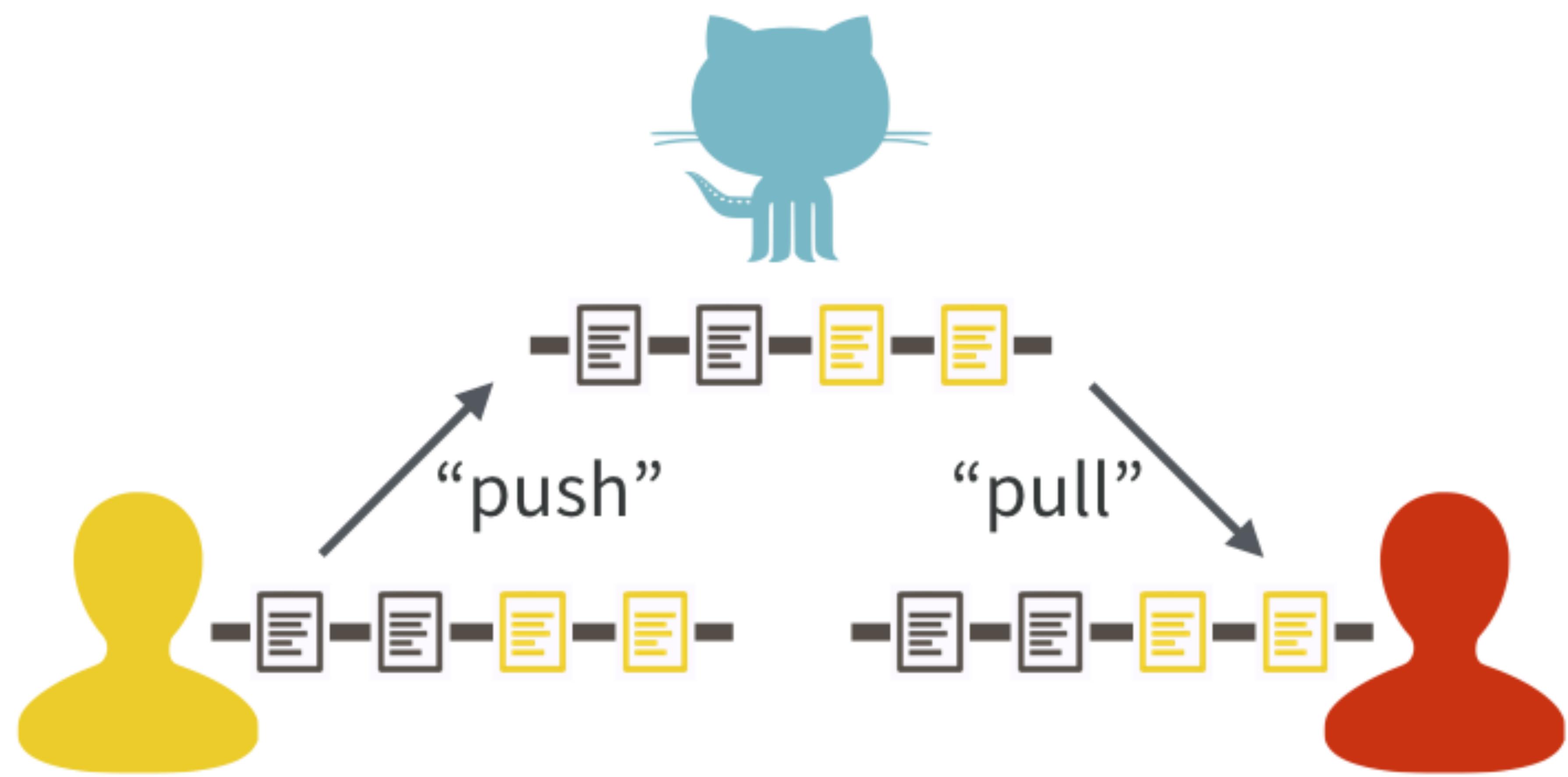
When you're done, click **Select theme** on the right.

The screenshot shows the GitHub interface for selecting a theme. At the top, there's a navigation bar with icons for repository, search, pull requests, issues, and gist, along with user notifications and profile. Below the navigation is a horizontal carousel of theme thumbnails. The 'Slate' theme is highlighted with a blue border, while other themes like 'Cayman', 'MERLOT', 'Time Machine', 'MINIMAL', and 'Leap Day' are shown in grayscale. Below the carousel are two buttons: 'Hide thumbnails' and 'Select theme'. A large dark overlay box covers the bottom half of the screen, featuring the text 'Slate theme' in large white font and 'Slate is a theme for GitHub Pages.' in smaller white font. At the bottom of this overlay are download links for 'View on GitHub' and file formats 'tar.gz' and '.zip'.

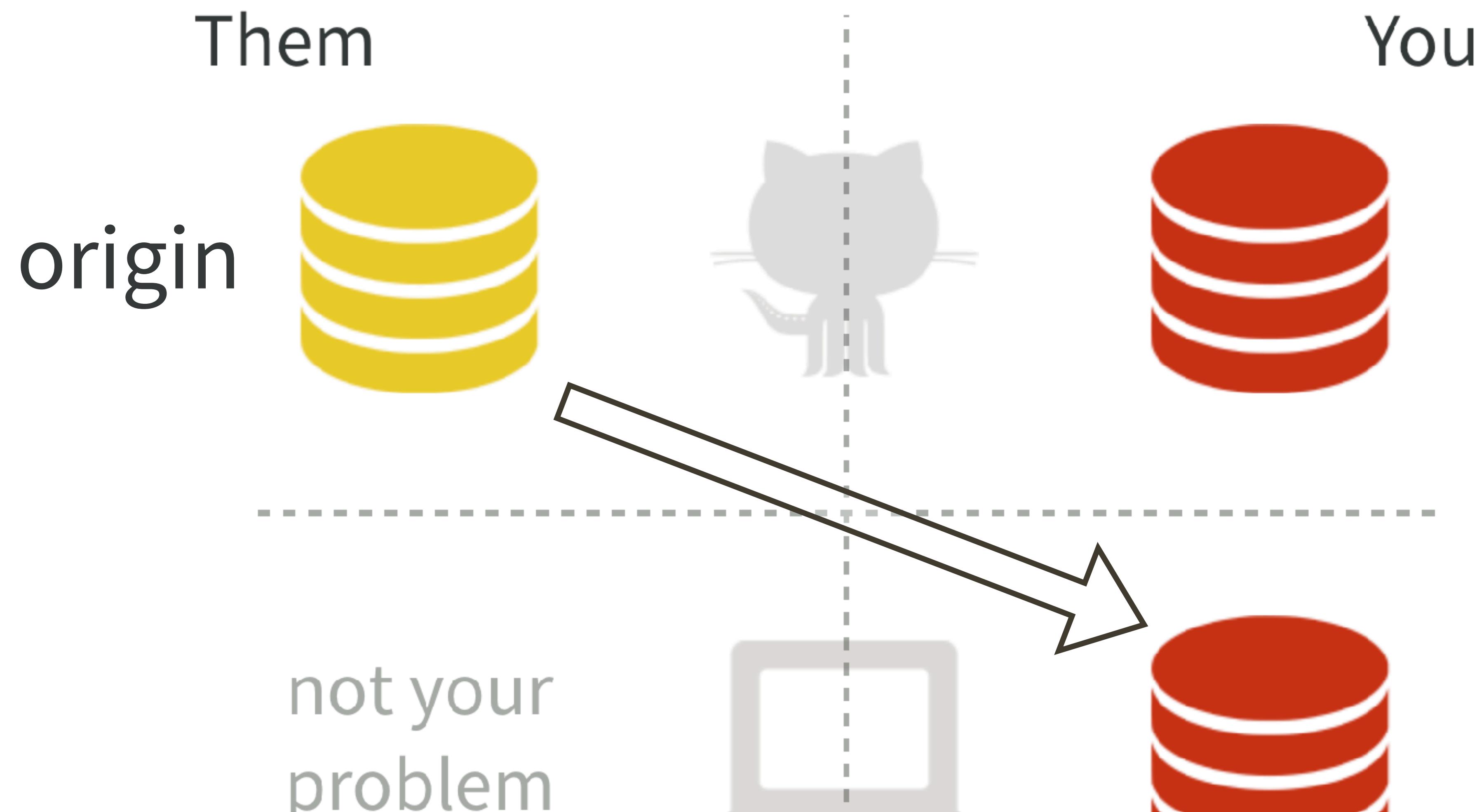
...and you're done!

Fire up a browser and go to **<http://username.github.io/repository>**.





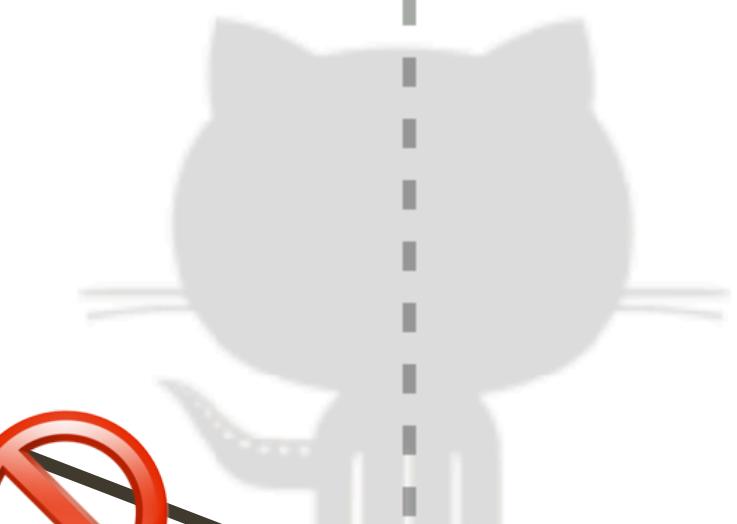
# collaboration



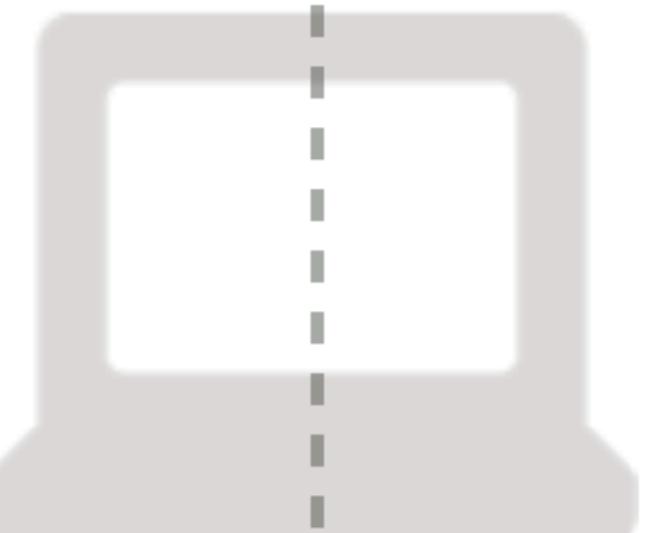
"clone"

\*not as useful as you might think  
because you can never send a PR

Them  
origin



not your  
problem



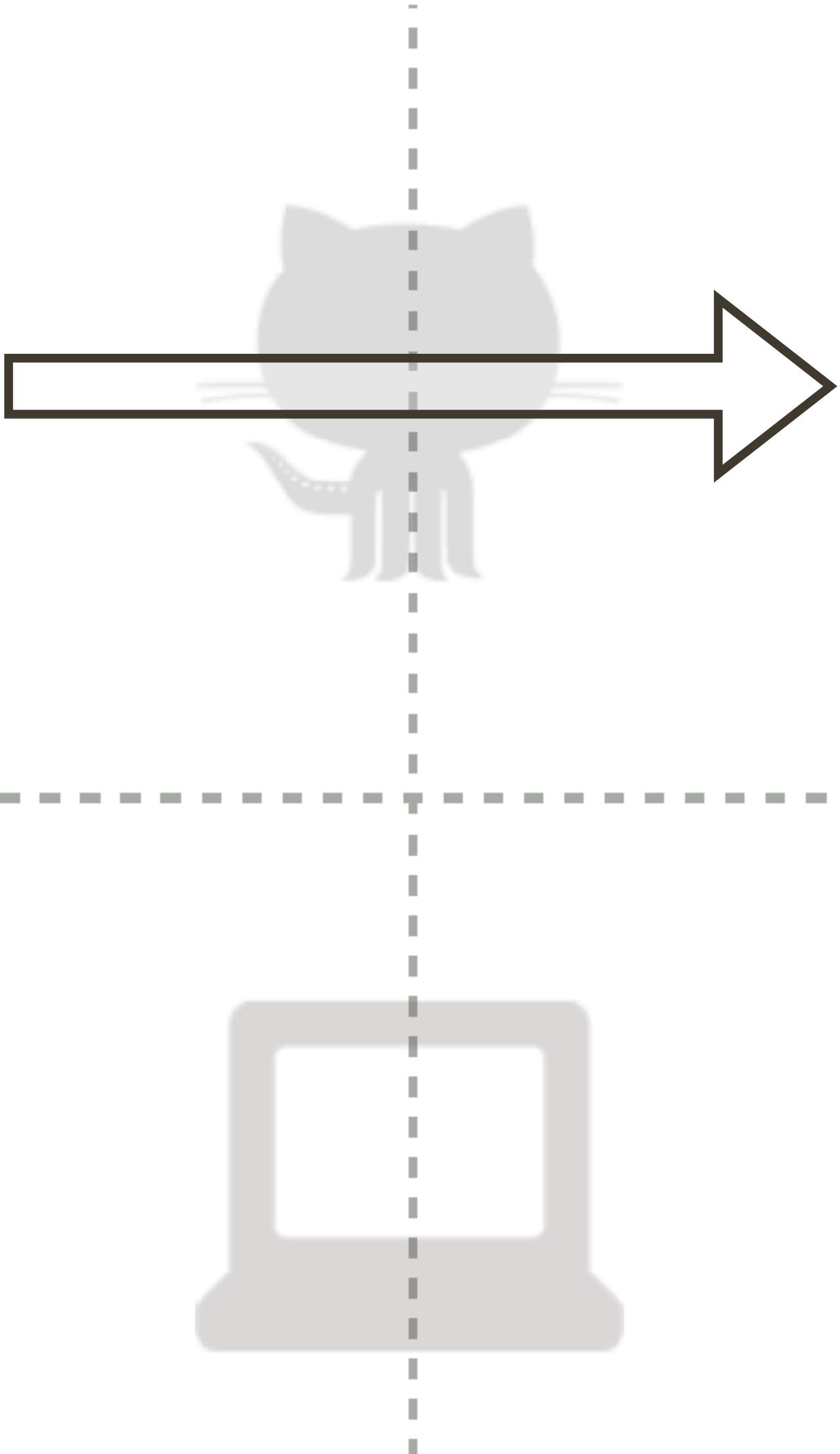
"clone"

\*not as useful as you might think  
because you can never send a PR

Them



You



not your  
problem



"fork"

Them

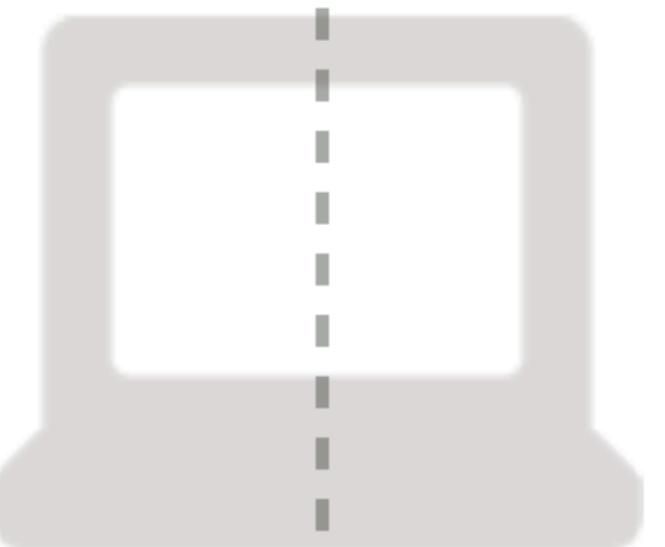


You



origin

not your  
problem



**"fork and clone"**

Them



You



pull request

pull

push

not your  
problem

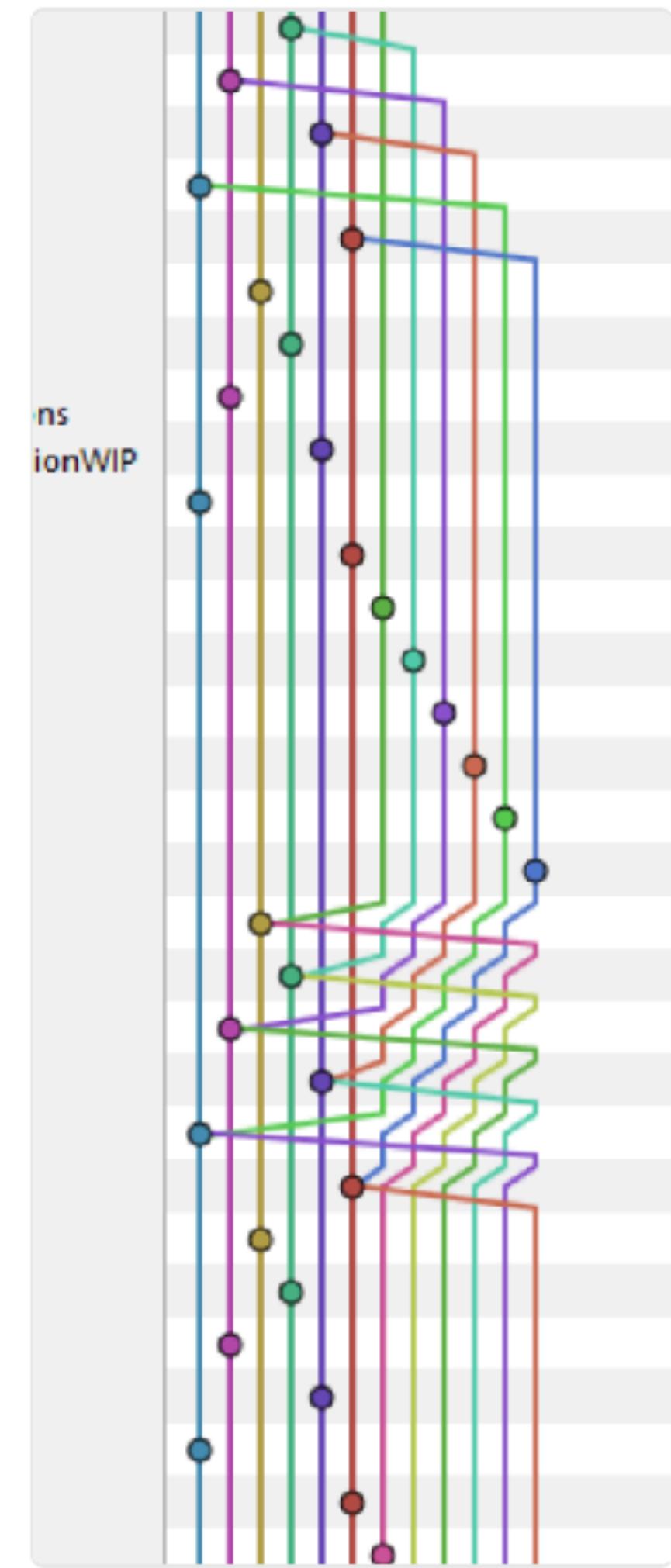
contribute to other people's stuff



Huenry Hueffman  
@HenryHoffman

Follow

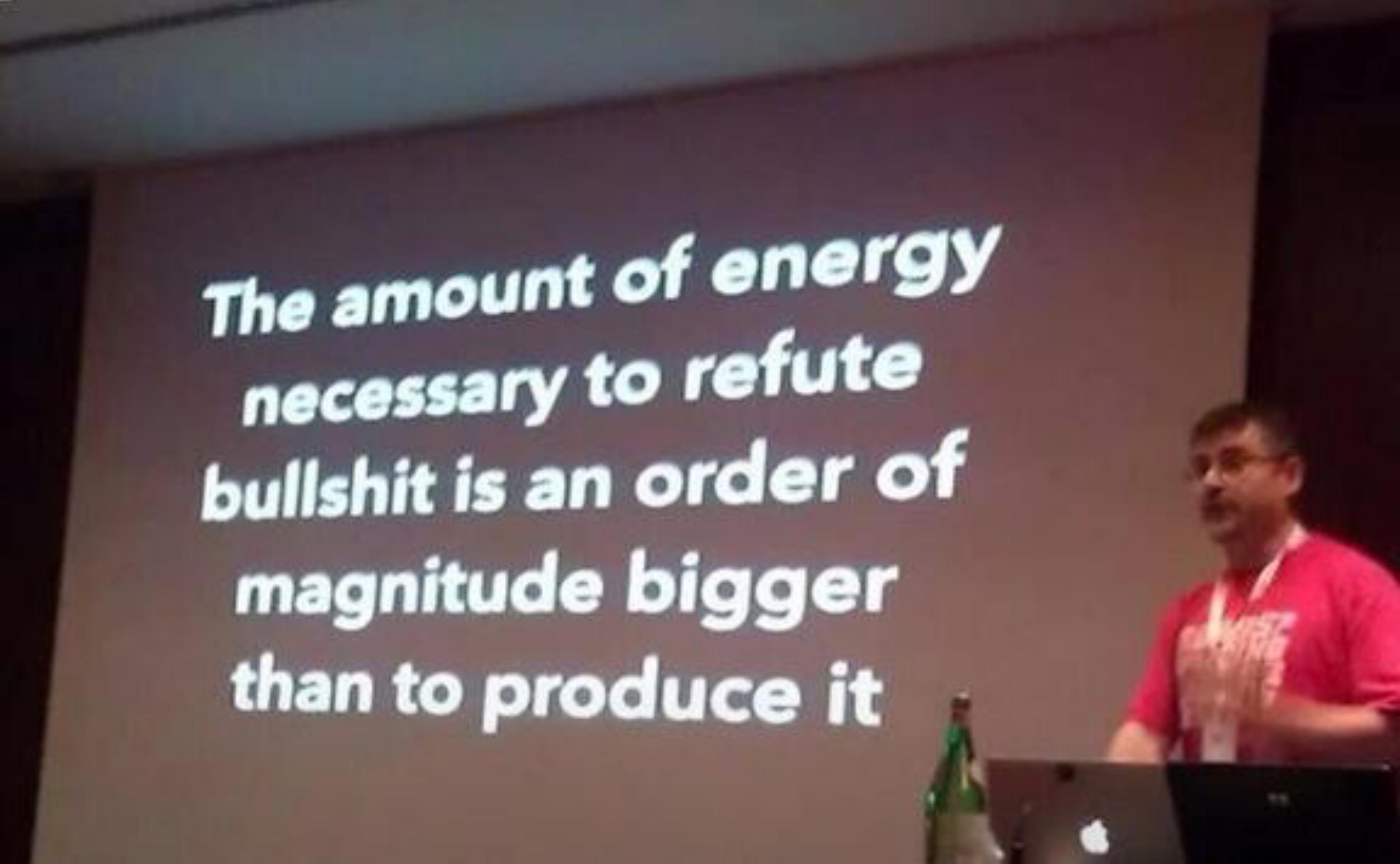
I fucked up Git so bad it turned into Guitar Hero



You do NOT want "Guitar Hero" git history.

The longer you wait to integrate, the harder it will be.

<https://twitter.com/henryhoffman/status/694184106440200192>



***The amount of energy  
necessary to refute  
bullshit is an order of  
magnitude bigger  
than to produce it***

- Alberto Brandolini

The amount of Git skilz  
necessary to fix a borked up  
repo is an order of magnitude  
bigger than to bork it.

- Jenny Bryan

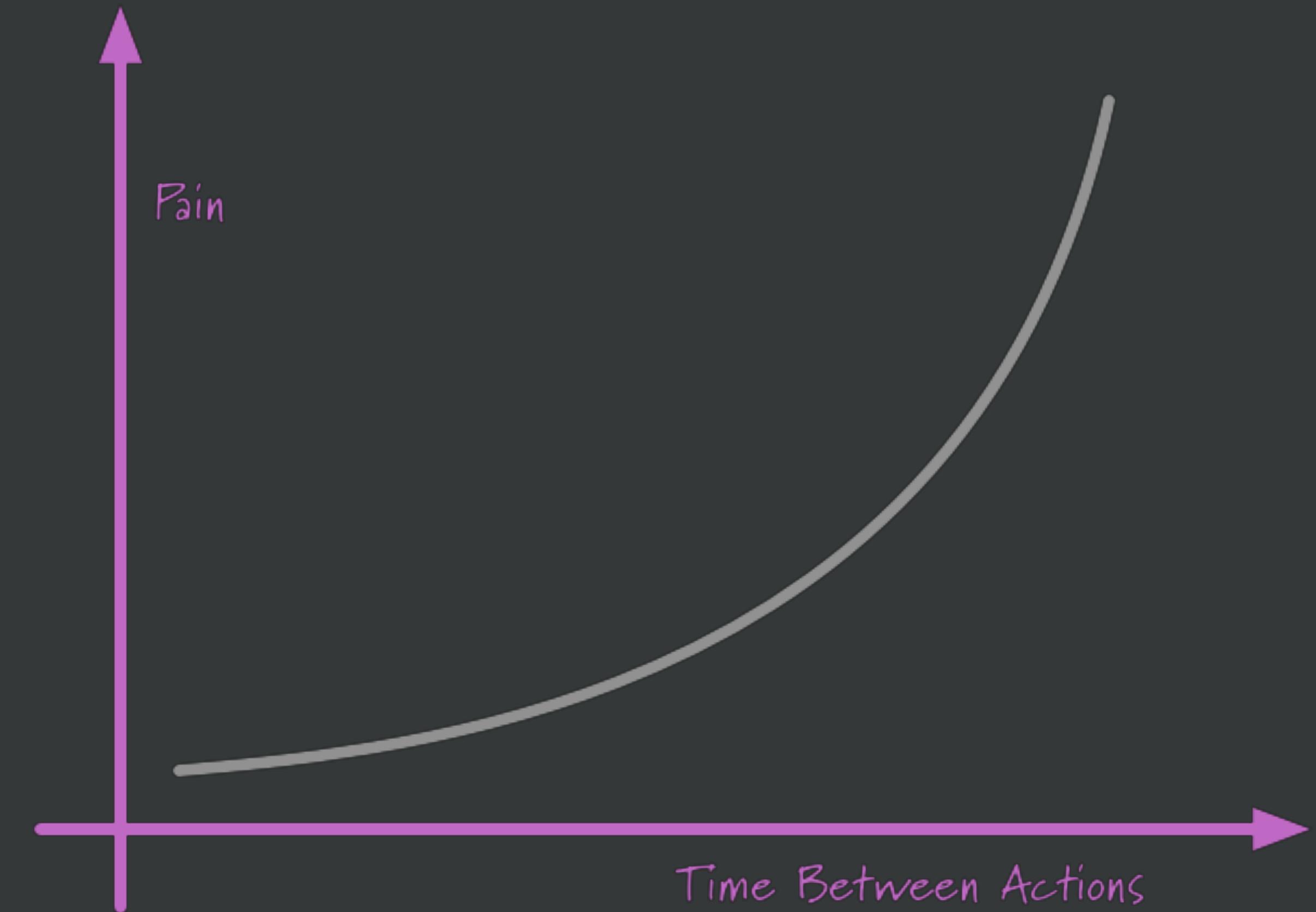
THIS IS GIT. IT TRACKS COLLABORATIVE WORK  
ON PROJECTS THROUGH A BEAUTIFUL  
DISTRIBUTED GRAPH THEORY TREE MODEL.

COOL. HOW DO WE USE IT?

NO IDEA. JUST MEMORIZIZE THESE SHELL  
COMMANDS AND TYPE THEM TO SYNC UP.  
IF YOU GET ERRORS, SAVE YOUR WORK  
ELSEWHERE, DELETE THE PROJECT,  
AND DOWNLOAD A FRESH COPY.



# "If it hurts, do it more often."



<https://martinfowler.com/bliki/FrequencyReducesDifficulty.html>

# Resources

- Initial steps toward reproducible research. Karl Broman.
- TIER Protocol.
- What They Forgot to Teach You About R, Jennifer Bryan and Jim Hester.
- Naming Things. Jennifer Bryan.
- quarto.org
- Introduction to Data Science. Mine Çetinkaya-Rundel
- Happy git and GitHub for the useR. Jenny Bryan, the STAT 545 TAs, Jim Hester.