

Using git

✕

Tales of Peril, Pain and Protection

Julia Piaskowski

2022/11/08

What Do I Use git & GitHub For?

1. Share and access data
2. Collaborate
3. Build things



GitHub Octocat

Share & Access Data/Software/Demos

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

rfordatascience/tidyuesday Public

Watch 593 Fork 2k Star 5.1k

<> Code Issues 183 Pull requests 1 Actions Projects Security Insights

master 5 branches 54 tags

Go to file Add file Code

About

Official repo for the #tidytuesday project

- Readme
- CC0-1.0 license
- Code of conduct
- 5.1k stars
- 593 watching
- 2k forks

Releases 54


tidyuesday_twee... Latest
2 hours ago


+ 53 releases

jthomasmock update tidyuesday_tweets ccf31f85 2 hours ago 1,117 commits		
community_resources	Grammar Corrections :)	2 months ago
data	Update readme.md	6 days ago
static	add horror movie data, closes #467	6 days ago
tidyuesday_tweets	update tidyuesday_tweets	2 hours ago
.gitignore	add data	4 years ago
CODE_OF_CONDUCT...	Create CODE_OF_CONDUCT.md	4 years ago
LICENSE	Update LICENSE	3 years ago
README.md	add horror movie data, closes #467	6 days ago
data.txt	update tidyuesday_tweets	2 hours ago

Tidy Tuesday datasets

Share & Access Data/Software/Demos

 [Pull requests](#) [Issues](#) [Marketplace](#) [Explore](#)




Hugging Face

The AI community building the future.

👥 4.4k followers 📍 NYC + Paris 🔗 <https://huggingface.co/> Verified


[Overview](#) [Repositories 100](#) [Projects](#) [Packages](#) [People 55](#) [Sponsoring](#)

Pinned

 **transformers** Public


🤗 Transformers: State-of-the-art Machine Learning for Pytorch, TensorFlow, and JAX.

● Python ☆ 73.7k 🍴 16.8k

 **datasets** Public


🤗 The largest hub of ready-to-use datasets for ML models with fast, easy-to-use and efficient data manipulation tools

● Python ☆ 14.7k 🍴 1.9k

 **diffusers** Public

🤗 Diffusers: State-of-the-art diffusion models for image and audio generation in PyTorch

● Python ☆ 7.3k 🍴 1k

 **accelerate** Public

🚀 A simple way to train and use PyTorch models with multi-GPU, TPU, mixed-precision

● Python ☆ 3.2k 🍴 258

Hugging Face Deep Learning Tools

Fork (Copy) a Repository

[jpiaskowski](#) / [sta210-s22-template](#) Public

[Pin](#)

[Unwatch](#)

generated from [sta210-s22/website](#)

[Code](#) [Issues](#) [Pull requests](#) [Actions](#) [Projects](#) [Wiki](#) [Security](#) [Insights](#)

[main](#)







[2 branches](#) [0 tags](#)

[Go to file](#)

[Add file](#)

[Code](#)

 **jpiaskowski** Initial commit ✓ d26c8a6 on Sep 13 [1 commit](#)

 .github/workflows	Initial commit	2 months ago
 _freeze	Initial commit	2 months ago
 ae	Initial commit	2 months ago
 documents	Initial commit	2 months ago
 exams	Initial commit	2 months ago
 hw	Initial commit	2 months ago

Course Template

Collaborate: File Issues

- Find out if someone else has experience a problem similar to yours
- Notify developers of a software bug
- Request a new feature

r-lib/actions Public

Edit Pins Watch 15 Fork 166 Starred 698

<> Code **Issues 18** Pull requests 1 Actions Projects Security Insights

XML has troubles installing on R 4.2
#559 by schloerke was closed 16 days ago
Closed 13

unable to find version master
#639 by bschilder was closed 18 days ago
Closed 8

Filters is:issue is:open Labels 12 Milestones 0 **New issue**

18 Open 381 Closed Author Label Projects Milestones Assignee Sort

Consider github pages deployment without commit feature 3
#623 opened on Sep 22 by hadley

setup-renv should allow for specifying working-directory feature 2
help wanted
#618 opened on Sep 8 by mavam

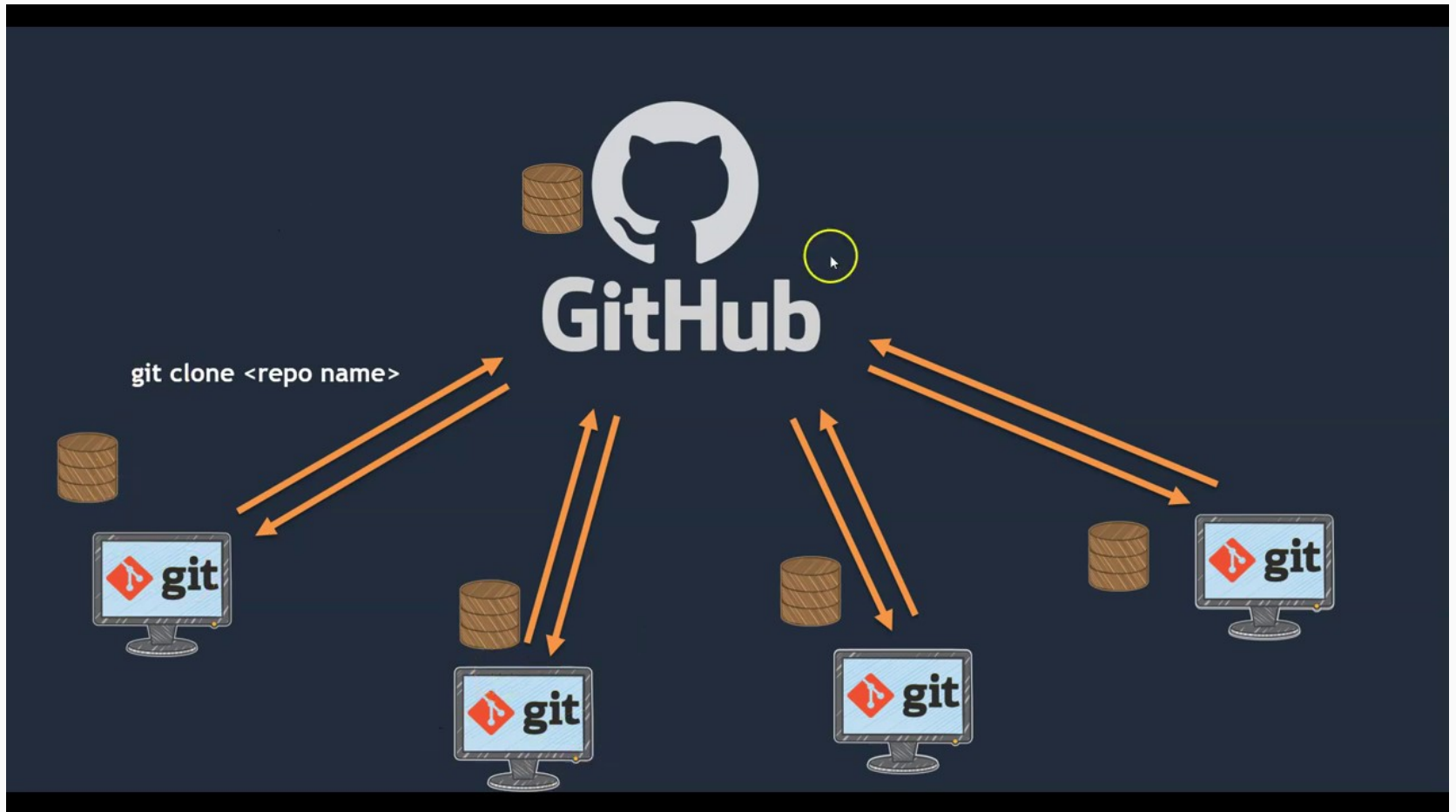
Issue tab for r-lib/actions

Collaborate With Others

Collaboration is the most compelling reason to manage a project with Git and GitHub. My definition of collaboration includes hands-on participation by multiple people, including your past and future self, as well as an asymmetric model, in which some people are active makers and others only read or review.

J. Bryan, Peer J (2017)

git vs. GitHub



Version Control

Version control captures changes across files, describing the difference between each and maintaining a linear history that can be rewound if needed.



Why Might You Need This?

To clarify your data curation and analysis process for *research reproducibility*

- Files change over time
- Results may be shared
- Many people are contributing to a project


...reproducibility is obtaining consistent results using the same input data; computational steps, methods, and code; and conditions of analysis.

- *Reproducibility and Replicability in Science* (2018)
National Academies of Sciences


Reference

Why Might You Need This?


You may be already implementing ad hoc version control.

 tidy_apple_phenotypic_data.csv

 tidy_apple_phenotypic_data2.csv

 tidy_apple_phenotypic_data3.csv


 tidy_apple_phenotypic_data4.csv

 tidy_apple_phenotypic_data5.csv

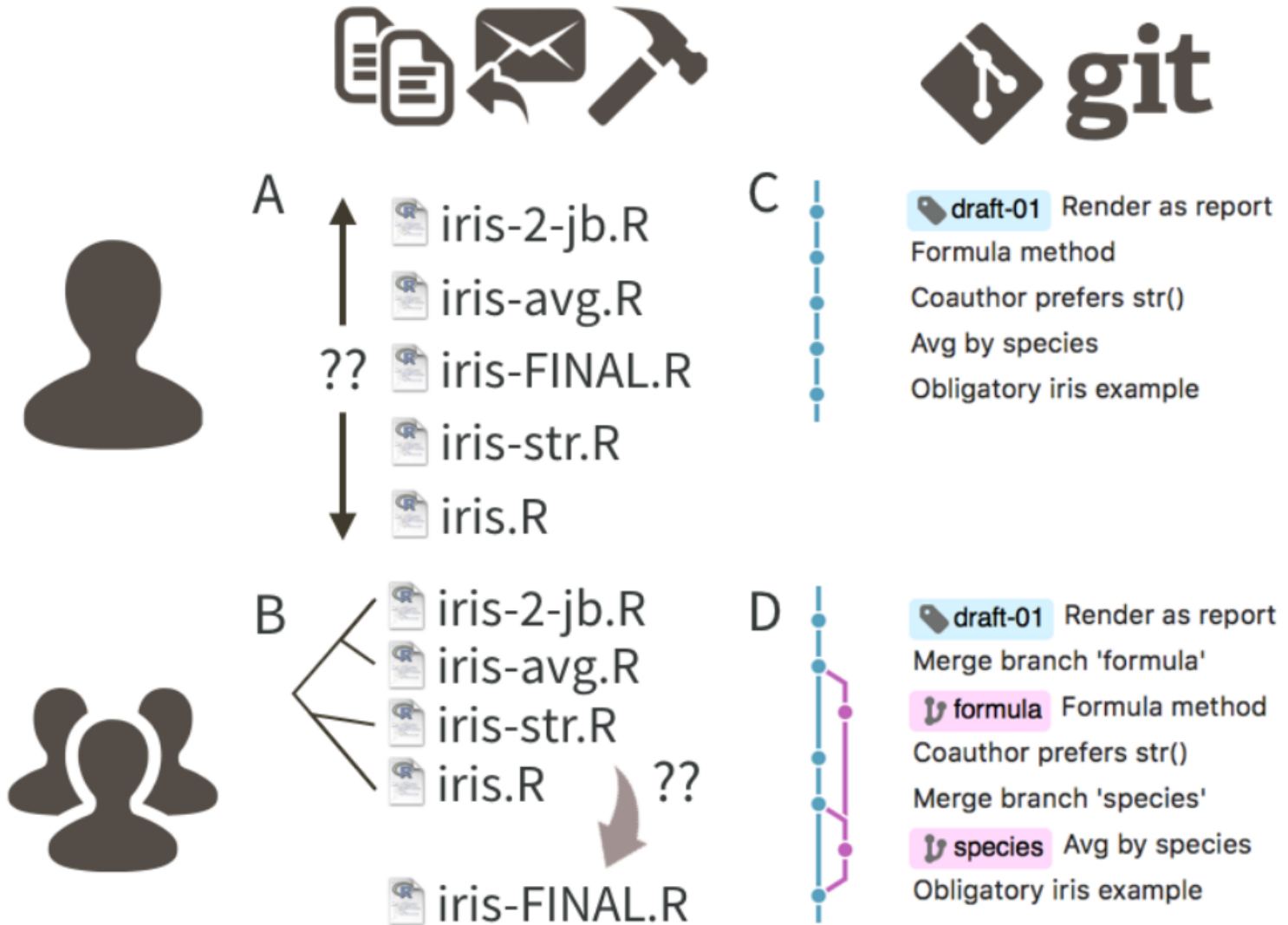
 tidy_apple_phenotypice_data5_SV.CSV

 name_conversion_2017Mar7.csv

 tidy_apple_phenotypice_data6.CSV

 tidy_apple_phenotypice_data7.CSV

What It Could Look Like



Version Control Advantages

- Sharing for asynchronous work - no more "data_final_KC_EP_JLP_BP.csv" or wondering if you have the most recent version of files
- Abundantly clear what is the central or main version of files
- Very advanced version of "track changes"
- Commits create safety points in case of disaster - you can recover previous version
- Branching creates space for experimentation

Generate a Web Presence

Idaho Ag Stats

Posts

Tutorials

Workshops

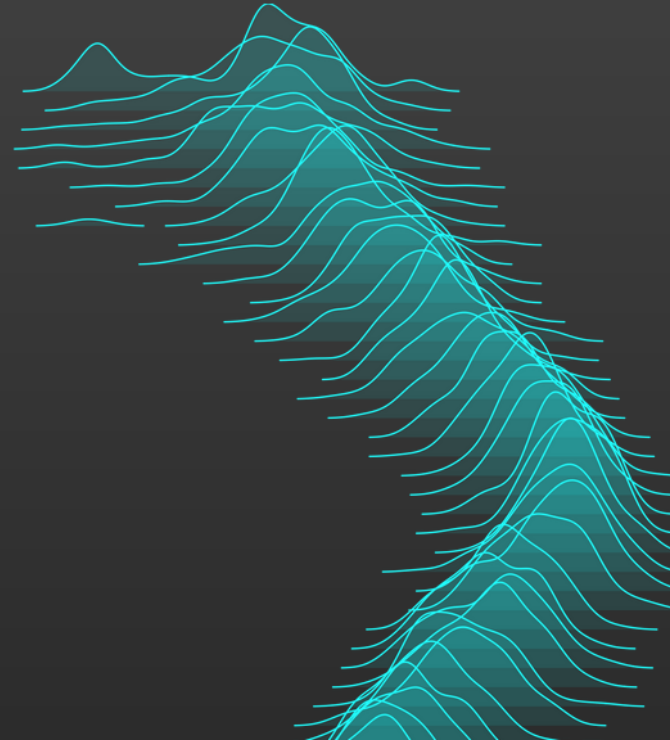
Presentations

contact



Idaho [[Agricultural]] Statistical Programs

Data management and analytical
resources for agricultural science



agstats.io | [GitHub](#)

Post a Tutorial as an Online Book

Happy Git and GitHub for
the useR

Table of contents

Let's Git started

1 Why Git? Why GitHub?

2 Contributors

3 Workshops

Installation

Half the battle

4 Register a GitHub account

5 Install or upgrade R and
RStudio

6 Install Git

7 Introduce yourself to Git



Still from Heaven King video

Happy Git provides opinionated instructions on how to:

- Install Git and get it working smoothly with GitHub, in the shell and in the RStudio IDE.

On this page

Let's Git started

License

View source 

Edit this page 

Happy Git with R | GitHub

Deploy Course Website

bayesf22 Notebook

Main bayesf22 class Statistical Rethinking Bayes Rules! 🔍

Linear models & causal inference

Video #3 code
Video #4 code

Week 3: Causes, confounds, and colliders

Video #5 code
Video #6 code

Week 4: Overfitting & interactions

Video #7 code

Week 5: MCMC & generalized linear models

Week 6: Integers & other monsters

Video #11 code

Week 7: Multilevel models I

Week 8: Multilevel models II

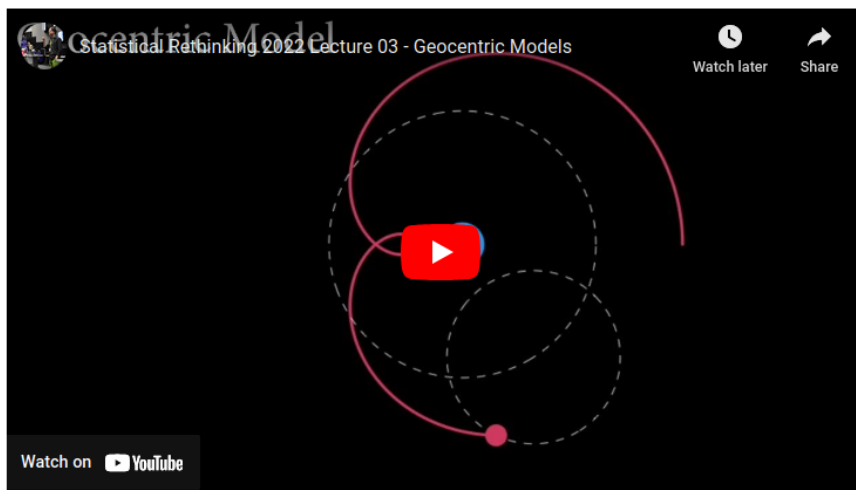
Week 9: Measurement & missingness

Week 10: Generalized linear

Geocentric models / linear regression

PUBLISHED

September 7, 2022



```
library(bayesrules)
library(tidyverse)
library(broom)
library(brms)
library(tidybayes)
library(ggdist)
library(patchwork)
```

On this page


Linear regression

1. Question/goal/estimand
2. Scientific model
3. Statistical model
- 4 & 5. Validate model and analyze data

Edit this page
Report an issue


Statistical Rethinking | GitHub

Documentation for Software

 **rusda** 1.0.8 [Reference](#) [Articles](#) [Changelog](#)

rusda

build failing codecov unknown



Links

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

License

GPL (>= 2)


Citation

[Citing rusda](#)

Developers

Franz-Sebastian Krahl
Maintainer

Software Peer-Rev

 Peer Reviewed

Interface to USDA databases

Description

An interface to the web service methods provided by the United States Department of Agriculture (USDA). The Agricultural Research Service (ARS) provides a large set of databases. The current version of the package holds interfaces to the Systematic Mycology and Microbiology Laboratory (SMML), which consists of four databases: Fungus-Host Distributions, Specimens, Literature and the Nomenclature database. It provides functions for querying these databases. The main function is `associations\(\)`, which allows searching for fungus-host combinations.

Get rusda

From CRAN (not on CRAN yet)

```
install.packages\("rusda"\)
```

Automate Tasks

- Automate action: `R cmd check` or `blogdown::build_site()`
- Run a GitHub action to do CI/CD (continuous integration/continuous deployment)
- GitHub actions can be challenging:

The screenshot displays the GitHub Actions interface. At the top, navigation tabs include Code, Issues (5), Pull requests, Discussions, Actions (selected), Projects, Wiki, Security, Insights, and Settings. The left sidebar shows 'Actions' with a 'New workflow' button and a list of categories: All workflows (selected), Management, Caches, and Deployments. The main area is titled 'All workflows' and shows 'Showing runs from all workflows'. A search bar for 'Filter workflow runs' is present. Below this, a table lists 28 workflow runs, each with a title, commit information, branch, and timing details.




Event	Status	Branch	Actor
Modified SAS code to store png in ./img renderbook #28: Commit ceb27ed pushed by pdiff	Failed	master	15 months ago 1m 23s
Updated SAS chapter 05 renderbook #27: Commit 8c733fc pushed by pdiff	Failed	master	15 months ago 1m 20s
initial incorporations of SAS section 05 renderbook #26: Commit 66316c3 pushed by pdiff	Failed	master	15 months ago 4m 47s
book reorganization for SAS rcbd file renderbook #25: Commit 2998d1b pushed by jpiaskowski	Failed	master	15 months ago 2m 43s
initial addition of SAS files renderbook #24: Commit fb1fc22 pushed by pdiff	Failed	master	15 months ago 2m 49s
renv added renderbook #23: Commit 623c938 pushed by jpiaskowski	Failed	master	2 years ago 1m 46s
Update README.md renderbook #22: Commit b9adfd1 pushed by jpiaskowski	Failed	master	2 years ago 3m 39s

What Git and GitHub Are *Not* Suitable For

- Making publication data sets available - this is not an appropriate long-term repository (just ask the National Academy of Sciences!)
- Tracking changes in binary files (.doc, .docx, .xls, .xlsx, .pdf)

There are special tools to work with large files - make sure you use those if your files are large!

Things Can Go Wrong With Version Control

Sith Lord		Wipe out someone else's commits by force pushing to the main branch.
Procrastinator		Created a repository with a single "init" commit with just a README.md. Then never touch it for 5 years.
Secret Santa		Accidentally commit a secret API key in a public repository.

Lesson 1: Take Learning git Seriously



Decent resources for self study:

- [Happy Git with R](#)
- [Software Carpentry git workshop](#)
- [The book of git](#)

xkcd cartoon

Lesson 2: Use a git Client

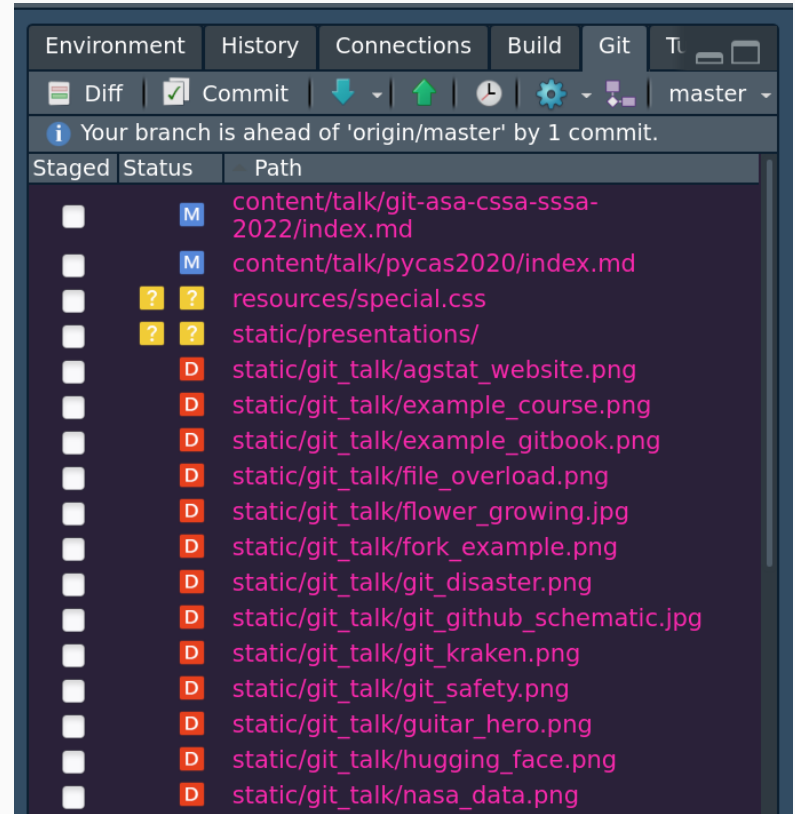
(A graphical user interface)

The screenshot displays a graphical user interface for a Git client. The main window shows a commit history for a repository named 'libgit2' with a working directory of 'ignore-bug'. The commit history is visualized as a vertical timeline of colored bars, each representing a commit. The commits are ordered from top to bottom: 'test for real' (1 hour ago), 'a test', 'Merge pull request #...' (3 hours ago), 'Merge pull request #...' (12 hours ago), 'Merge pull request #3...' (2 days ago), 'Fix unused variable 'm...' (3 days ago), 'Fix initial commit test' (4 days ago), 'Merge pull request #3769 from li...', 'maybe done', 'Add tests for creating ...' (5 days ago), 'Fix `git_commit_create` for an init...', 'rebase: handle no common ances...', 'rebase: test rebase (merge) w/ no ...', 'diff: simplify code for handling em...', 'almost there', 'Merge pull request #3...' (6 days ago), 'delta-apply: fix sign extension', 'odb_loose: fix undefined behavior...', 'checkout: set ignorecase=0 when ...', 'merge_file: do not unnecessarily c...', 'odb: avoid inflating the full delta t...', 'wip', 'index: fix memory leak on error ca...', and 'Merge pull request #...' (a week ago). The commit 'almost there' is highlighted in green. On the right side, a panel titled 'Viewing merged diff of 5 commits' shows a list of commits: 'test for real' (f878f0), 'a test' (271d84), 'maybe done' (0b41a2), 'almost there' (fbb74b), and 'wip' (1959ce). Below this list, a summary indicates '1 added', '0 deleted', and '2 modified' files. The modified files are 'include/git2/tree.h', 'src/tree.c', and 'tests/object/tree/update.c'.

Some Graphical User Interfaces

- GitHub GUI
- Git Kraken
- SourceTree
-**tons more**

The goal is to become accustomed to using git regularly - use the tools that help you reach that



Lesson 3: Be Patient When an Error Occurs

When the inevitable error happens:

- a merge error
- can't pull or push!
- a `git revert` gone horribly wrong

Proceed **cautiously** and diagnose the problem



Lesson 4: Don't Expose Secrets

- Learn about and use the `.gitignore` file
- Consider private repositories when appropriate
- There are some guidelines regarding legal compliance (e.g. [HIPAA](#))



git is a Humbling Experience



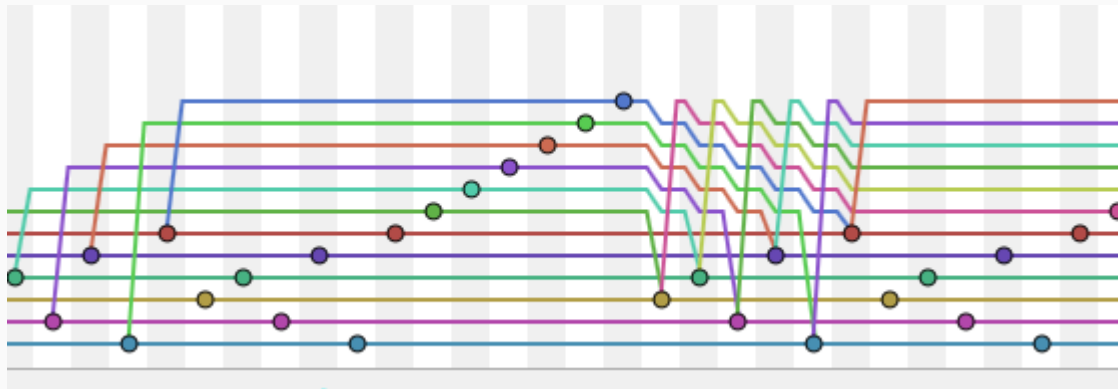
ag stats stan

@SeedsAndBreeds



If you you have not experienced a git disaster, have you really used git?

(I still think git is worth the trade-offs)



original tweet