

The Benefits of Implementing Git & GitHub in a Data Science Curriculum

Eric Burnheimer

Massachusetts Bay CC

eburnheimer@massbay.edu

ericburn.github.io



git

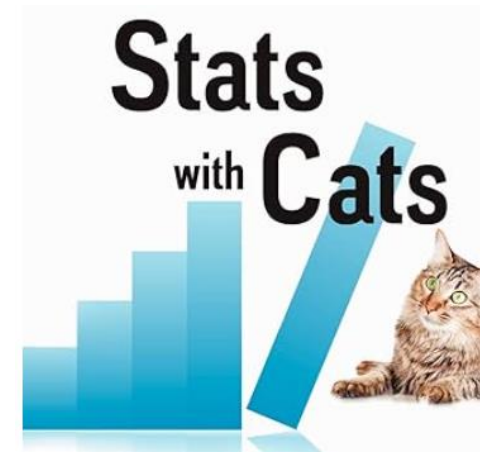


GitHub

Hypothetical Student “Octo”*



- Octo had a keen interest in analyzing cat data.
 - Job Satisfaction in Data Science (DS) work is high (Baumer & Horton, 2023)
- Octo began to explore data science programs
 - Two-year colleges positioned to play large role in DS education (Kotz, 2020; Amstat, 2022; Glantz, et.al. 2023)



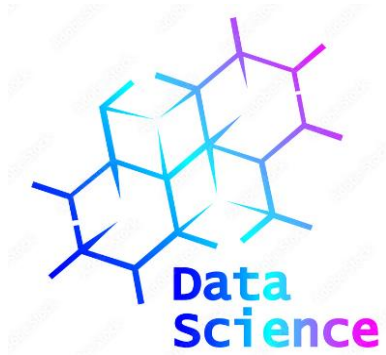
Credit: Charles Kufs



*Adapted from Baumer & Horton (2023)

Data Science Programs in Two-Year Colleges

- Octo's local two-year college had just started a DS certificate!
 - Important curricular decisions to be made for A.S. degree program
- One important curricular decision involves when and how to teach reproducibility and workflow (Baumer & Horton, 2023; Çetinkaya-Rundel & Rundel, 2018)
 - Version Control Systems (e.g., Git)
 - Collaboration and Project Management Venues (e.g., GitHub)





Credit: Zoom

“It’s on my
GitHub”

Two things
happen to Octo
during winter
break...

“Work
collaboratively
through **GitHub**”



“Most of the folks that we interview have their GitHub pages listed on their resume,” says CiBo Technologies talent acquisition manager Jamieson Vasquez. “I think that is important.”

Companies That Use Git & GitHub



[Home](#)
[PUBLIC](#)
[Questions](#)
[Tags](#)
[Users](#)
[Companies](#)
[COLLECTIVES](#)
[Explore Collectives](#)
[LABS](#)
[Discussions](#)
[TEAMS](#)

Stack Overflow for Teams – Start collaborating and sharing organizational knowledge.


[Create a free Team](#)
[Why Teams?](#)

All Questions

[Ask Question](#)


New search experience powered by AI

Stack Overflow is leveraging AI to summarize the most relevant questions and answers from the community, with the option to ask follow-up questions in a conversational format.
[Get early access](#)

23,970,501 questions

[Newest](#)
[Active](#)
[Bountied](#)

179

[Unanswered](#)
[More](#)
[Filter](#)

27128 votes

✓ 25 answers

1.8m views

Why is processing a sorted array faster than processing an unsorted array?

In this C++ code, sorting the data (before the timed region) makes the primary loop ~6x faster:
`#include <algorithm> #include <ctime> #include <iostream> int main() { // ...`

[java](#)
[c++](#)
[performance](#)
[cpu-architecture](#)
[branch-prediction](#)

GManNickG 496k asked Jun 27, 2012 at 13:51

25978 votes

✓ 105 answers

13.3m views

How do I undo the most recent local commits in Git?

I accidentally committed the wrong files to Git, but didn't push the commit to the server yet. How do I undo those commits from the local repository?

[git](#)
[version-control](#)
[git-commit](#)
[undo](#)

Community wiki 93 revs, 64 users 11% Peter Mortensen

20334 votes

✓ 41 answers

11.2m views

How do I delete a Git branch locally and remotely?

Failed Attempts to Delete a Remote Branch: `$ git branch -d remotes/origin/bugfix error: branch 'remotes/origin/bugfix' not found. $ git branch -d origin/bugfix error: branch 'origin/bugfix' not...`

[git](#)
[version-control](#)
[git-branch](#)
[git-push](#)
[git-remote](#)

Matthew Rankin 459k asked Jan 5, 2010 at 1:12

13708 votes

✓ 38 answers

3.4m views

What is the difference between 'git pull' and 'git fetch'?

What are the differences between git pull and git fetch?

[git](#)
[version-control](#)
[git-pull](#)
[git-fetch](#)

Pablo Fernandez 281k asked Nov 15, 2008 at 9:51

3 of top 4
questions
on Stack
Overflow
relate to
Git.

“Whoa, I’ve just read this quick tutorial about git and oh my god it is cool. I feel now super comfortable using it, and I’m not afraid at all to break something.”

— said no one ever ([de Wulf](#))

What is Version Control and What Are Git & GitHub?

What is Version Control?

- “A system that records changes to a file or set of files over time...” (Chacon & Straub, 2014)



What is Git?



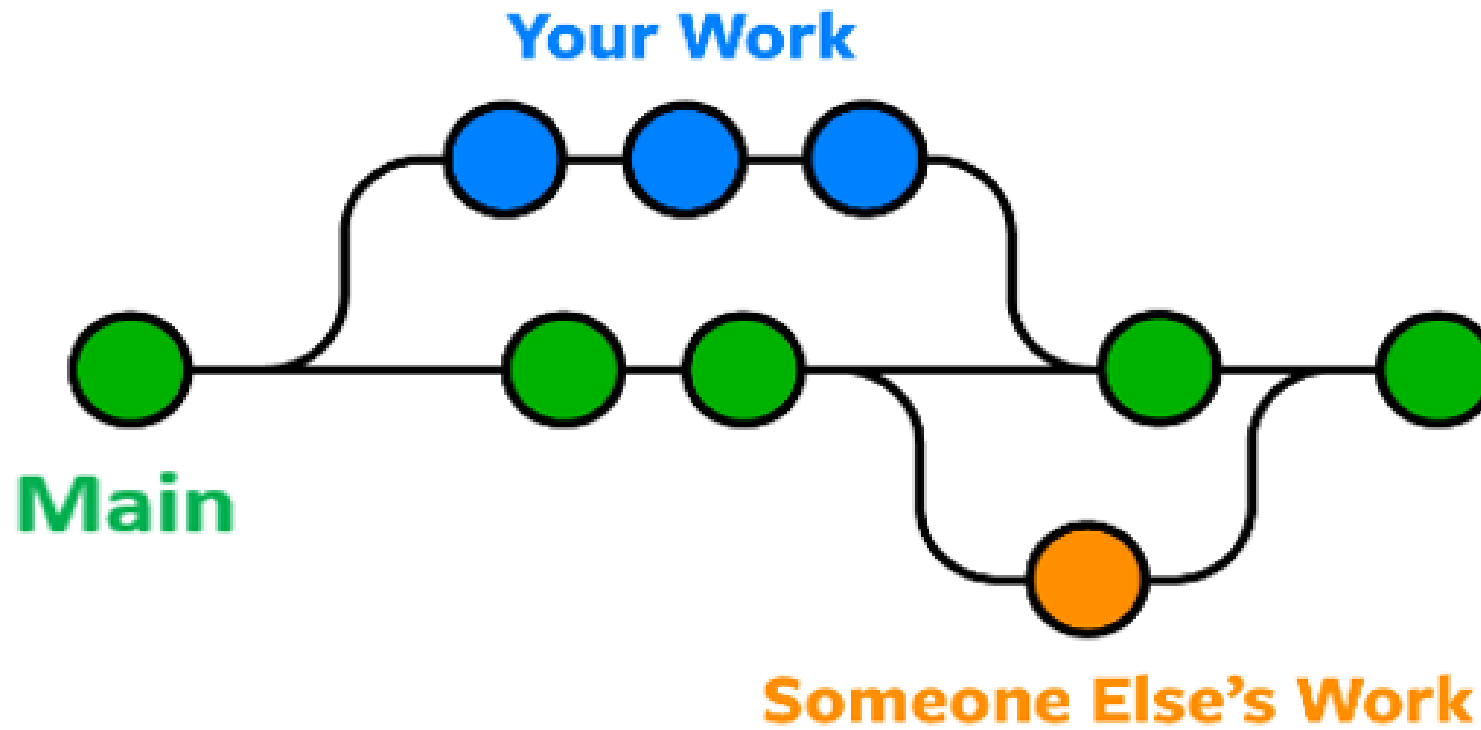
- **Git** is a free version control system
 - Various commands to track and manage changes to files over time
 - ❖ Revisit earlier versions of files
 - ❖ Compare changes between versions
 - ❖ Undo changes
- **Git** use is enabled through a repository (repo)
 - Place that tracks & manages files within a folder



What is GitHub?



- **GitHub** hosts git repositories in the cloud
 - Allows teams (e.g., students) to work on files together (e.g., R scripts)

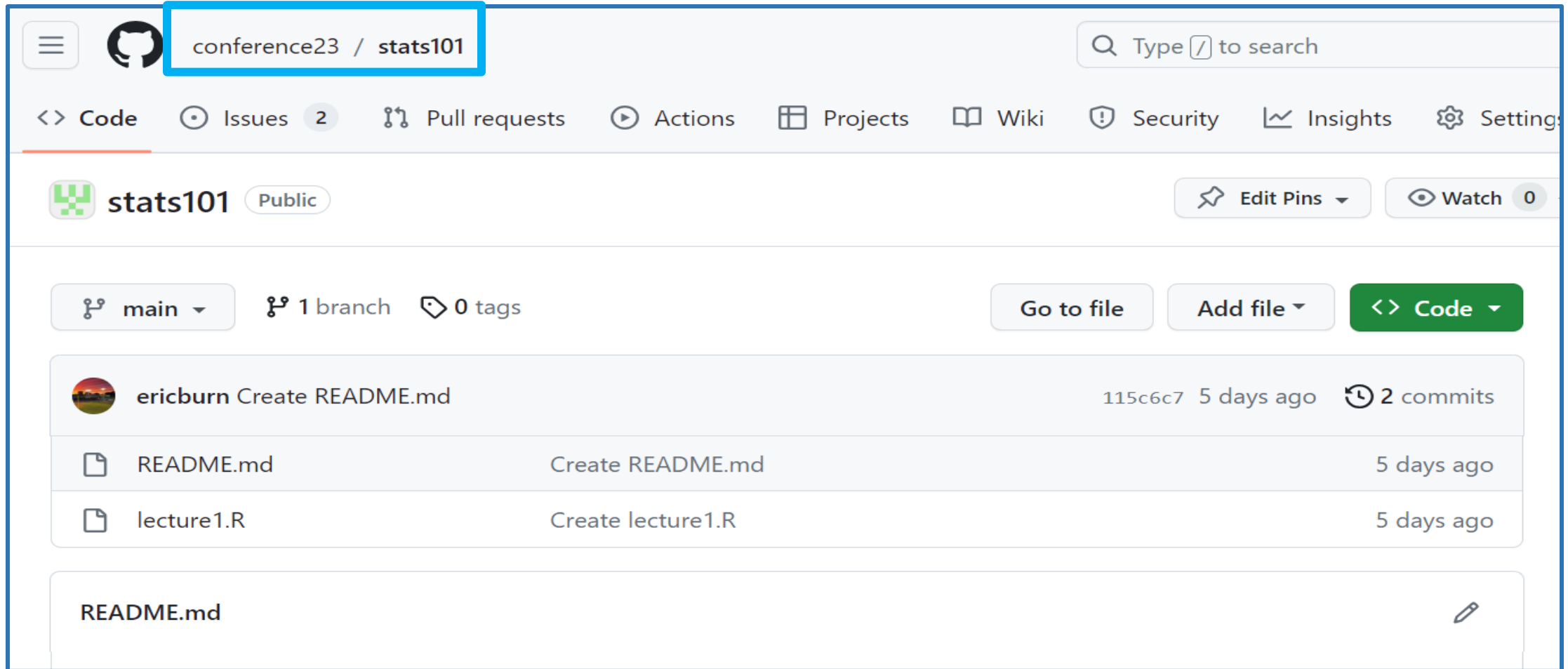


A Pathway for Implementing **Git** and **GitHub** in a DS Curriculum

Semester I: Use GitHub Issues to Collaborate in Introductory Statistics

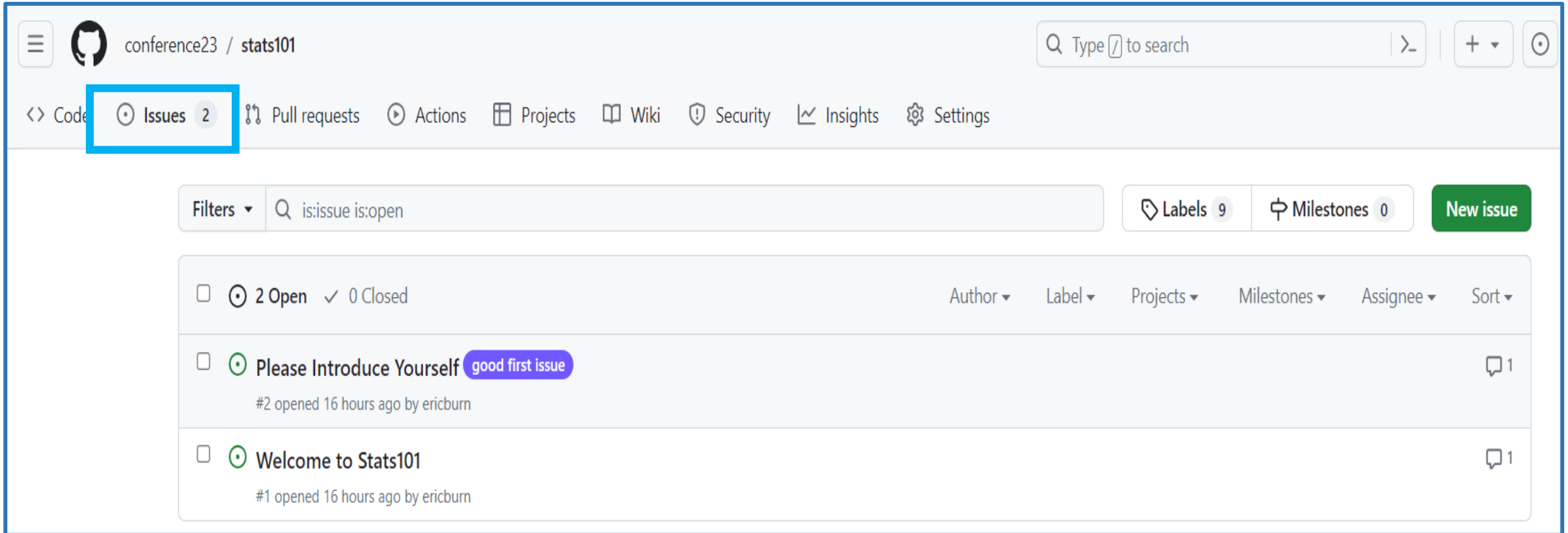
Part I: Create An Organization & Class Repo

- Invite students as collaborators to your repo (e.g., stats101)



Part II: Introduce GitHub (& Git) through GitHub Issues

- **GitHub Issues** acts a message board



The screenshot displays the GitHub repository page for 'conference23 / stats101'. The 'Issues' tab is selected and highlighted with a blue box. The page shows two open issues:

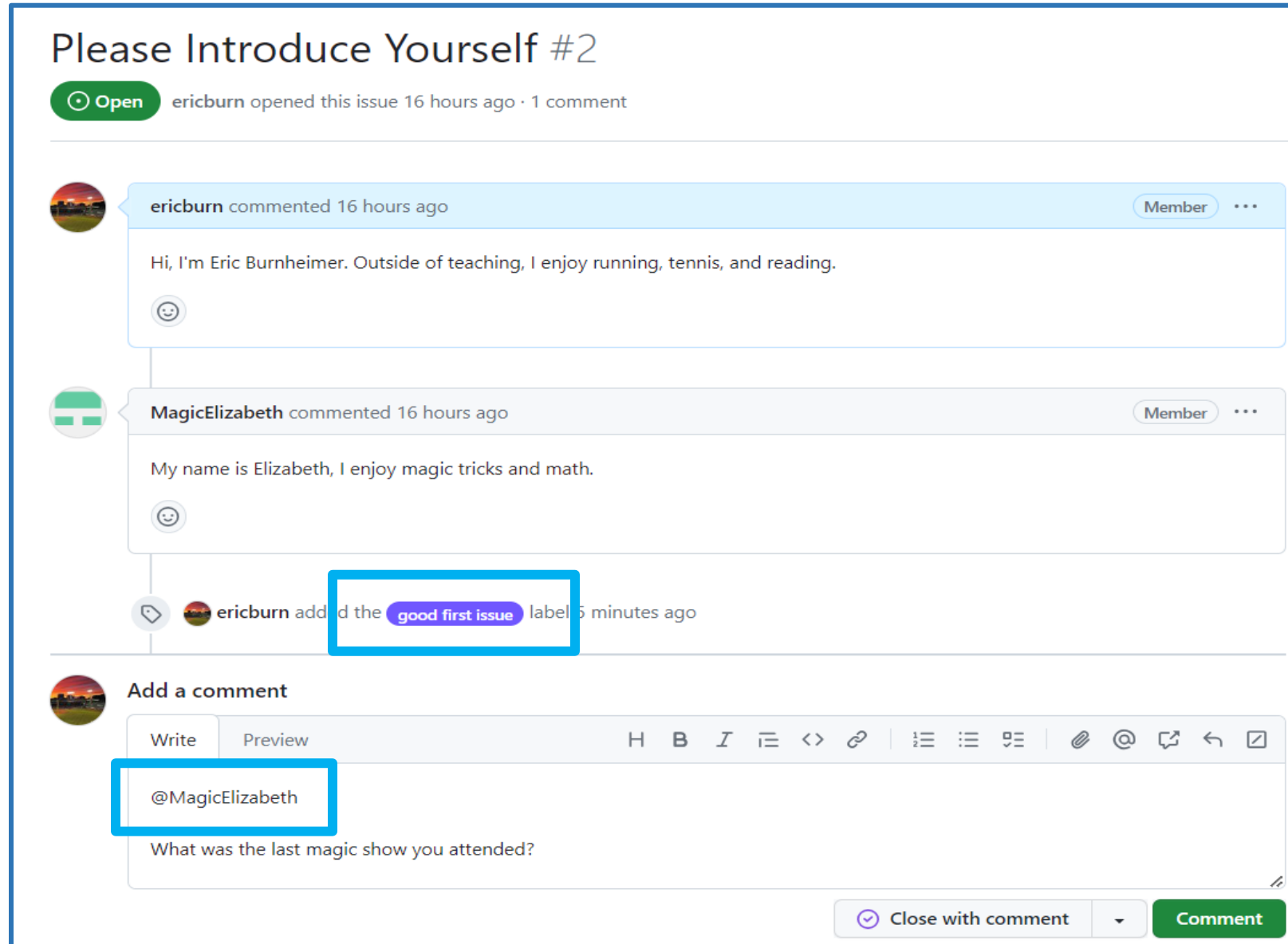
- Please Introduce Yourself** (labeled 'good first issue')
#2 opened 16 hours ago by ericburn
- Welcome to Stats101**
#1 opened 16 hours ago by ericburn

The interface includes a search bar, filters, and a 'New issue' button.

- Students can ask questions, give feedback, share ideas, etc.

Part II: Introduce GitHub (& Git) through GitHub Issues

- **Labels** can be added to posts; students can be mentioned
- Git (I mean get 😊) students used to the user interface right away



The screenshot shows a GitHub issue titled "Please Introduce Yourself #2" with a status of "Open". It was opened by "ericburn" 16 hours ago and has 1 comment. The issue contains two comments:

- A comment by "ericburn" (Member) from 16 hours ago: "Hi, I'm Eric Burnheimer. Outside of teaching, I enjoy running, tennis, and reading."
- A comment by "MagicElizabeth" (Member) from 16 hours ago: "My name is Elizabeth, I enjoy magic tricks and math."

Below the comments, a notification states: "ericburn added the **good first issue** label 5 minutes ago". The "good first issue" label is highlighted with a blue box.

At the bottom, there is a comment input area with the text "@MagicElizabeth" and "What was the last magic show you attended?". The "@MagicElizabeth" text is highlighted with a blue box. The input area includes a "Write" tab, a "Preview" tab, and a rich text editor toolbar. At the bottom right of the input area are buttons for "Close with comment" and "Comment".

Mini-Course After Semester I:

Introduction to Git & GitHub

- Create small groups of students to work together during the course
 - Create files for students to collaborate on in group repos within the organization

Week 1: Core Concepts

- Git Installation
 - Command Line
 - git init
- Basics
 - git add/commit/push
- Correcting Mistakes
 - git reset/revert
- Observing
 - git status/log/diff

Week 2: Branching

- Creating Branches
 - git branch/checkout
- Merging Branches
 - Fast forward/3-way
- Resolving Conflicts
 - git rebase
- Retrieving
 - git pull/fetch

Week 3: Other

- Workflow
 - Pull Requests
- Create Personal Site
 - GitHub Pages
- GUIs
 - GitKraken

Other Topics to Explore

- **GitHub Pages** enables and eases the creation of a personal site

Octo Cat



octo@octocat.edu

[View the Project on GitHub](#)
ericburn/octo

Download
ZIP File

Download
TAR Ball

View On
GitHub

This project is maintained by [ericburn](#)

Hosted on GitHub Pages — Theme by [orderedlist](#)

Data Science Student @ DS CC

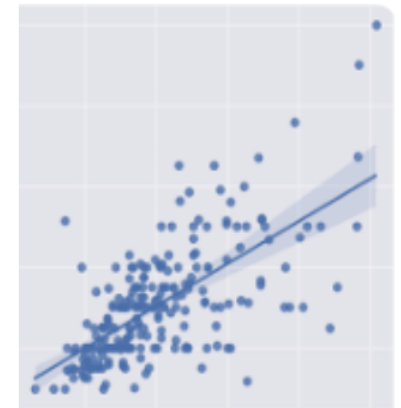
Technical Skills: R, Git & GitHub

Education

- A.A., Data Science (*exp. May 2025*)

Projects

- The Relationship Between X and Y



-The Relationship Between Y and Z (in progress)

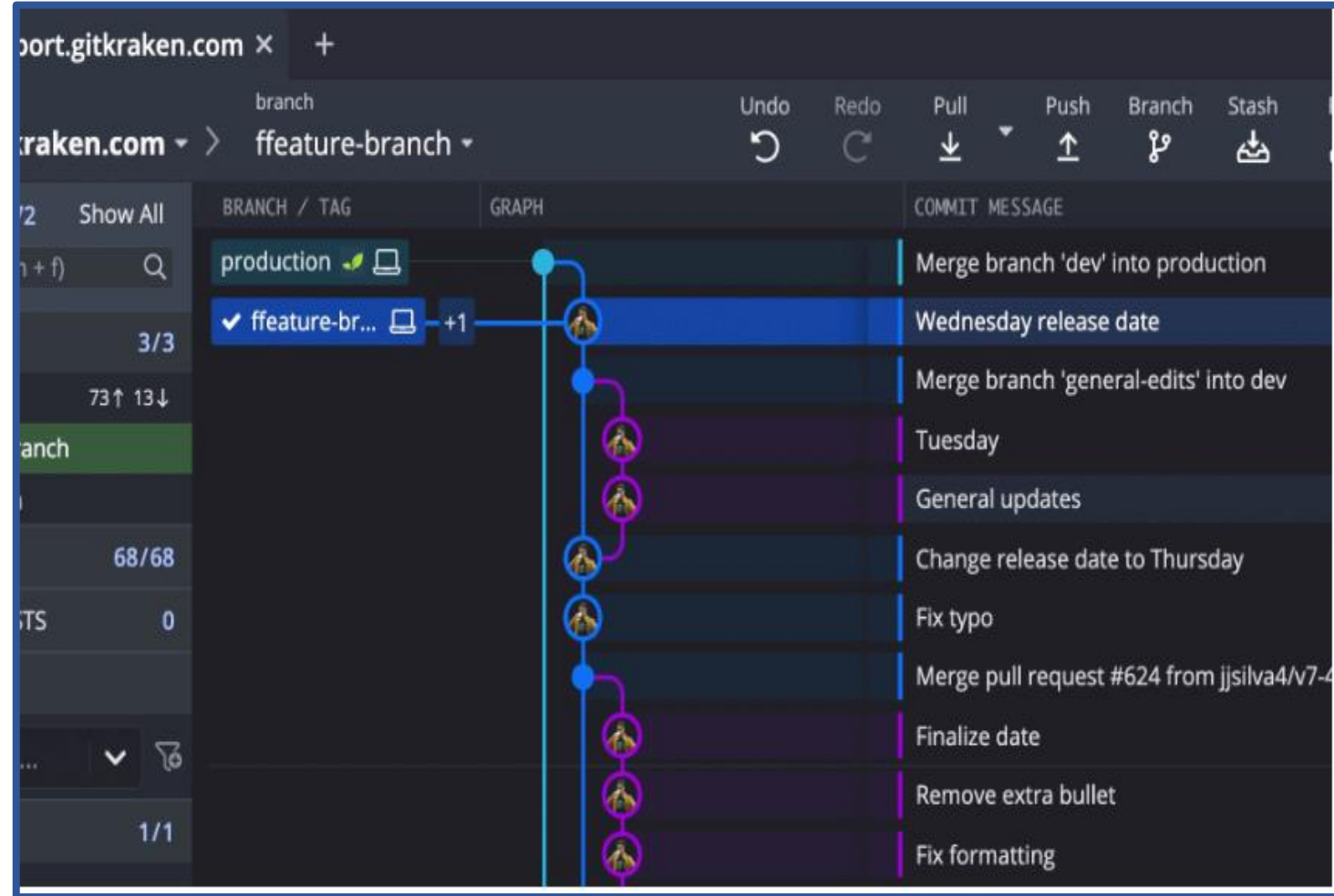
Work Experience

Summer Intern @ TBD Company (*May 2024 - August 2024*)

Graphical User Interface



- **GUIs** such as GitKraken
 - Nice visualization of branches

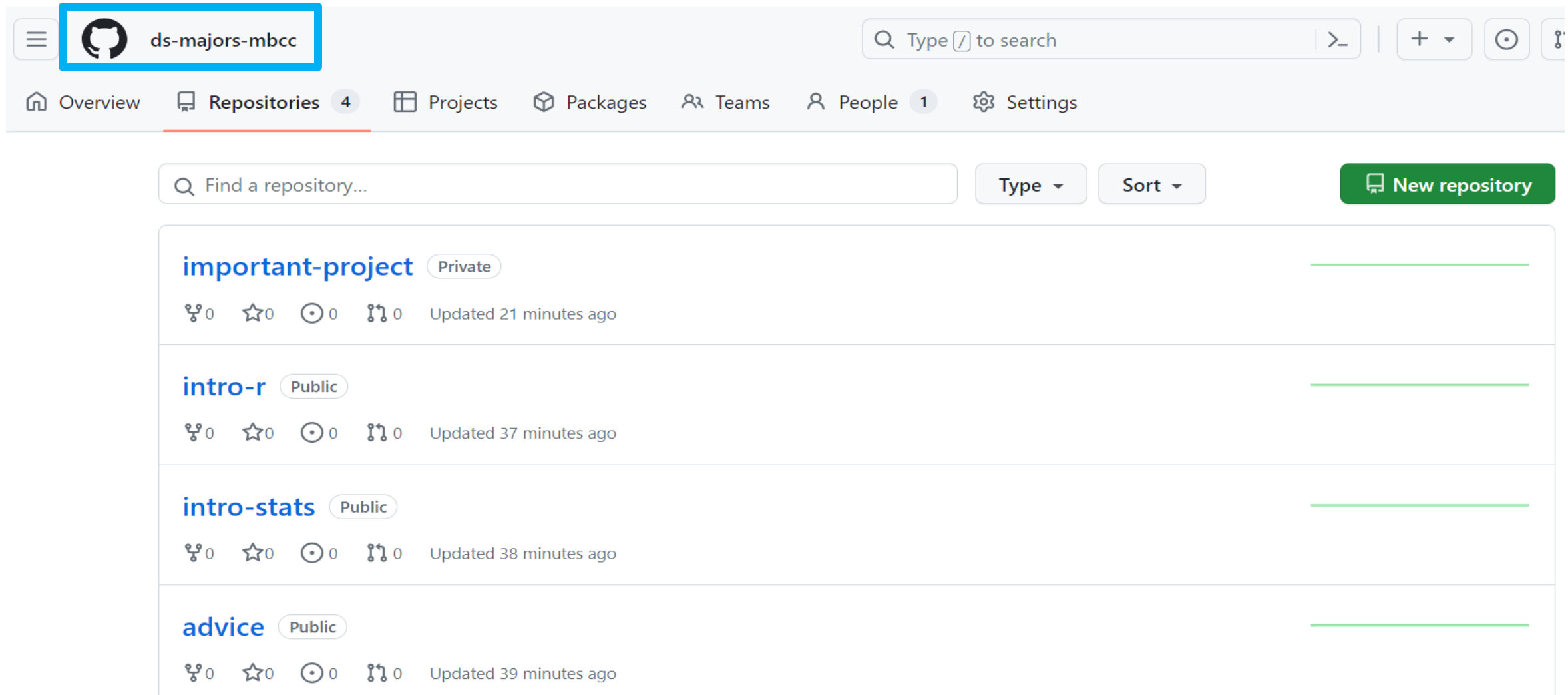


Credit: Medium

After Mini-Course: Access to Potential Data Science Majors Organization and Repos

Data Science Majors Organization

- Upon completing the course, students obtain access to the Data Science Majors Organization



The screenshot shows the GitHub interface for the 'ds-majors-mbcc' organization. The organization name is highlighted with a blue box. The navigation bar includes links for Overview, Repositories (4), Projects, Packages, Teams, People (1), and Settings. A search bar is present with the text 'Type / to search'. Below the navigation bar, there is a search input field labeled 'Find a repository...', a 'Type' dropdown, a 'Sort' dropdown, and a green 'New repository' button. The main content area displays a list of repositories:

Repository Name	Visibility	Forks	Stars	Issues	Pull Requests	Last Updated
important-project	Private	0	0	0	0	Updated 21 minutes ago
intro-r	Public	0	0	0	0	Updated 37 minutes ago
intro-stats	Public	0	0	0	0	Updated 38 minutes ago
advice	Public	0	0	0	0	Updated 39 minutes ago

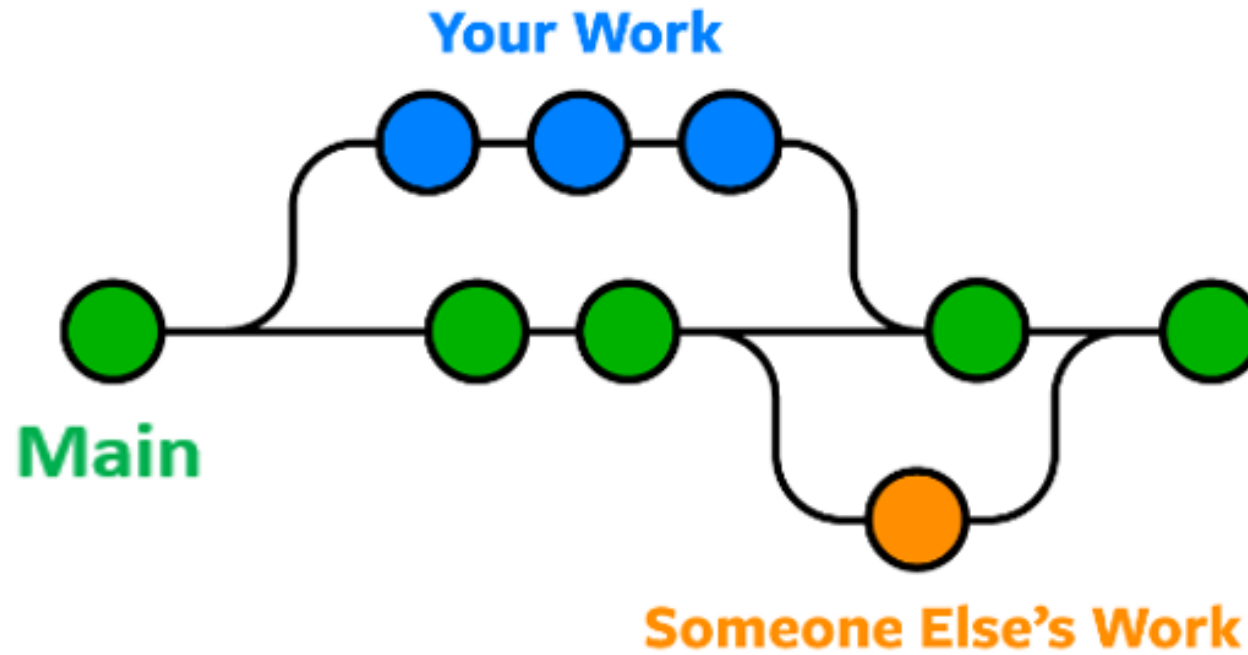
The Benefits of Implementing Git & GitHub in a DS Program

- Potential for Increased Collaboration
- Version Control and Project Management Skills
- Workforce / Transfer Readiness

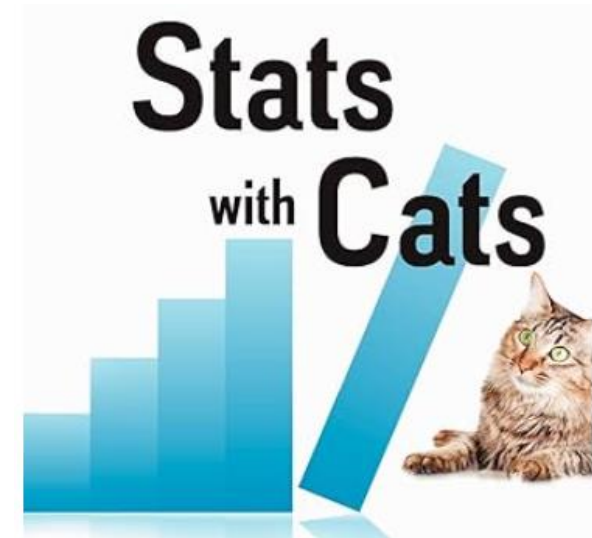
A Pathway for Implementing Git & GitHub in a DS Program

- Semester I: Introduce GitHub (and preview Git) by communicating via Issues
- Mini-Course / Workshop after Semester I: Introduction to Git & GitHub
- Access to Data Science Majors Organization and Repos

Octo Will Be Ready to Collaborate!



Credit: Noble Desktop



Credit: Charles Kufs

Free Resources

- git-scm.com
 - This is the official git site and includes full documentation on all commands
- *Pro Git* by Scott Chacon and Ben Straub
 - This book is affiliated with the official git site

Low-Cost Resources

- *The Git & GitHub Bootcamp* by Colt Steele on Udemy
 - This can be purchased for roughly \$10-\$15 during one of Udemy's many sales, and it provides great exercises for learning both Git & GitHub

Baumer, B. S., & Horton, N. J. (2023). Data Science Transfer Pathways From Associate's to Bachelor's Programs.

Harvard Data Science Review, 5(1). <https://doi.org/10.1162/99608f92.e2720e81>

Çetinkaya-Rundel, M., & Rundel, C. (2018). Infrastructure and tools for teaching computing throughout the statistical curriculum. *The American Statistician*, 72(1), 58–65. <https://doi.org/10.1080/00031305.2017.1397549>

Chacon, S., & Straub, B. (2014). Pro Git (2nd edition). Apress.

Glantz, M., Johnson, J., Macy, M., Nunez, J. J., Saidi, R., & Velez, C. (2023). Students' Experience and Perspective of a Data Science Program in a Two-Year College. *Journal of Statistics and Data Science Education*, 1–10. <https://doi.org/10.1080/26939169.2023.2208185>

Kotz, B. (2020). The Opportunities of Two-Year College Data Science. *Harvard Data Science Review*, 2(4). <https://doi.org/10.1162/99608f92.75aed58b>

New Two-Year College Data Science, Analytics Programs on the Rise. (2022, August 1). *Amstat News*. <https://magazine.amstat.org/blog/2022/08/01/new-two-year-programs/>

Steele, C. (2021). *The Git & GitHub Bootcamp*. Udemy.

Thank you!

Eric Burnheimer
Massachusetts Bay CC
eburnheimer@massbay.edu
ericburn.github.io

GitHub Classroom



- **GitHub Classroom**


- Manage a virtual classroom
 - ❖ Make Assignments
 - ❖ Connect to some LMSs

conference23-classroom-fall23

conference23

☆ Assignments 0 👤 Students 0 👥 TAs and Admins 1 ⚙️ Settings


Assignments



Create an assignment to get started.

[Create an individual assignment](#) to generate an assignment repository for each student to work from. Or, [create a group assignment](#) and have students work collaboratively in groups from team repositories.

Create an assignment



Need to teach Git & GitHub fundamentals?

The Classroom team has created an assignment for you to use to teach your students the fundamentals of Git & GitHub.

Use starter assignment

[Learn more about the GitHub Starter Assignment](#)