

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

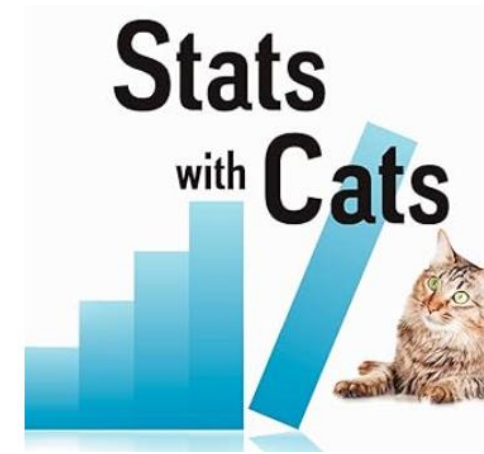
Eric Burnheimer
Massachusetts Bay Community College
eburnheimer@massbay.edu



Hypothetical Student “Octo”*



- Octo had a keen interest in analyzing cat data.
 - Job Satisfaction in Data Science (DS) work is high (Baumer & Horton, 2023)
 - ❖ Passion for the work, strong job growth, and high salaries
- Octo began to explore data science options at nearby two-year colleges
 - Two-year colleges positioned to play large role in DS education
 - ❖ Affordability and Accessibility (Kotz, 2020; Amstat, 2022; Glantz, et.al. 2023)



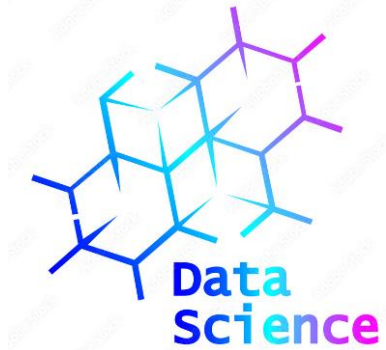
Credit: Charles Kufs



*Adapted from Baumer & Horton (2023)

Data Science Programs in Two-Year Colleges

- Octo's local two-year college had just started a DS certificate!
 - Dozens of programs recently created; many others likely to emerge soon (Kotz, 2020; Amstat, 2022)
 - Important curricular decisions to be made
- 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.”

“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](#))

[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!

What is Version Control?

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



What is Git?



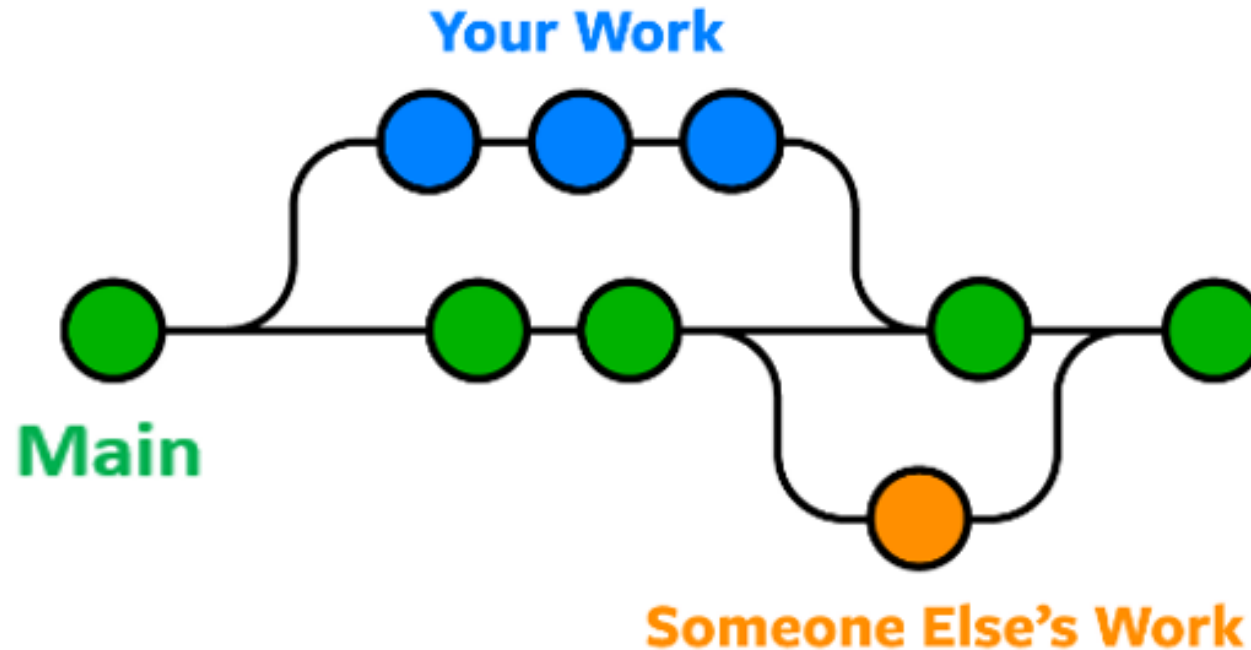
- **Git** is a free version control system
 - Tracks and manages changes to files over time
 - ❖ Revisit earlier versions of files
 - ❖ Compare changes between versions
 - ❖ Undo changes
- **Git** use is enabled through a repository (repo)
 - Place that tracks & manages files within a folder



What is GitHub?



- **GitHub** hosts git repositories in the cloud
 - Provides a venue for people to collaborate on repos
 - Allows teams (e.g., students) to work on files together (such as R scripts)

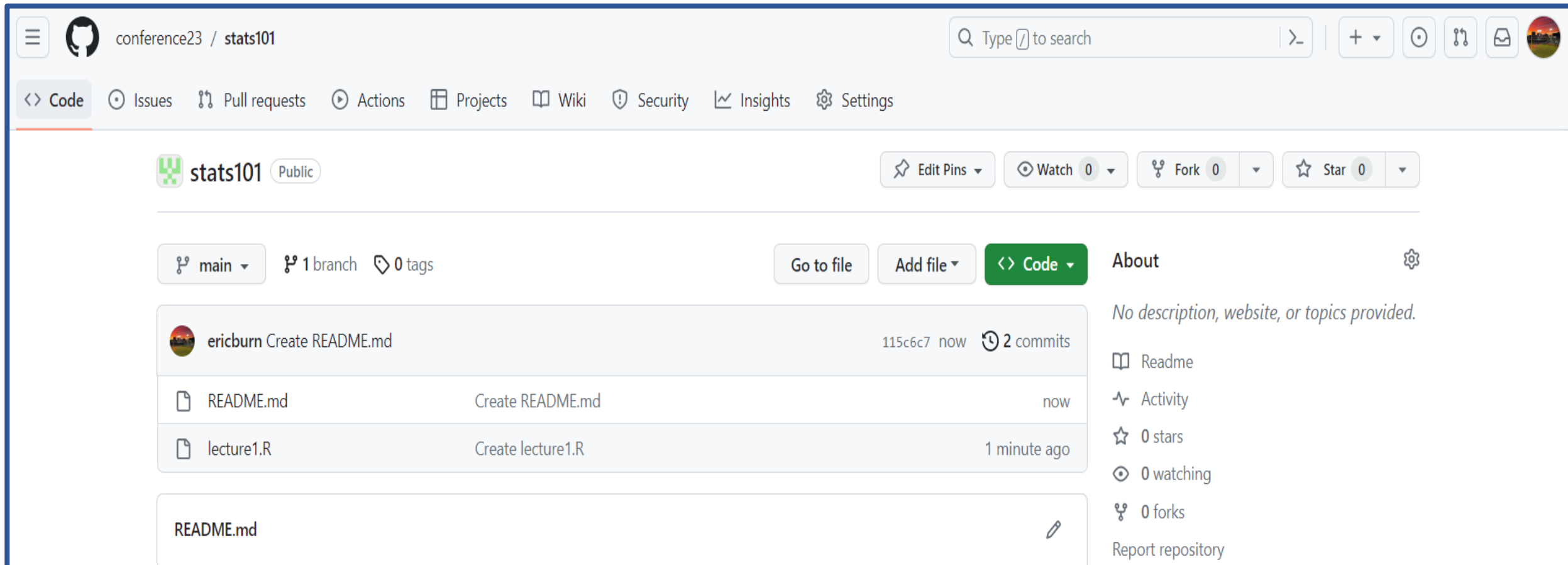


A Pathway for Implementing **Git** and **GitHub** in a DS Curriculum

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)



The screenshot shows the GitHub interface for a repository named 'stats101' owned by 'conference23'. The repository is public. The top navigation bar includes links for Code, Issues, Pull requests, Actions, Projects, Wiki, Security, Insights, and Settings. The repository header shows the name 'stats101', its public status, and buttons for Edit Pins, Watch (0), Fork (0), and Star (0). Below the header, there are buttons for 'main' branch, '1 branch', and '0 tags', along with 'Go to file', 'Add file', and 'Code' buttons. The commit history shows a single commit by 'ericburn' titled 'Create README.md' with a commit hash of '115c6c7' and '2 commits' ago. The file list shows 'README.md' and 'lecture1.R', both created recently. The right sidebar contains an 'About' section with a note that no description, website, or topics are provided, and a list of repository statistics: 0 stars, 0 watching, and 0 forks. A 'Report repository' link is also present.

conference23 / stats101

Type / to search

<> Code Issues Pull requests Actions Projects Wiki Security Insights Settings

stats101 Public

Edit Pins Watch 0 Fork 0 Star 0

main 1 branch 0 tags

Go to file Add file <> Code

About

No description, website, or topics provided.

Readme Activity 0 stars 0 watching 0 forks Report repository

ericburn Create README.md 115c6c7 now 2 commits

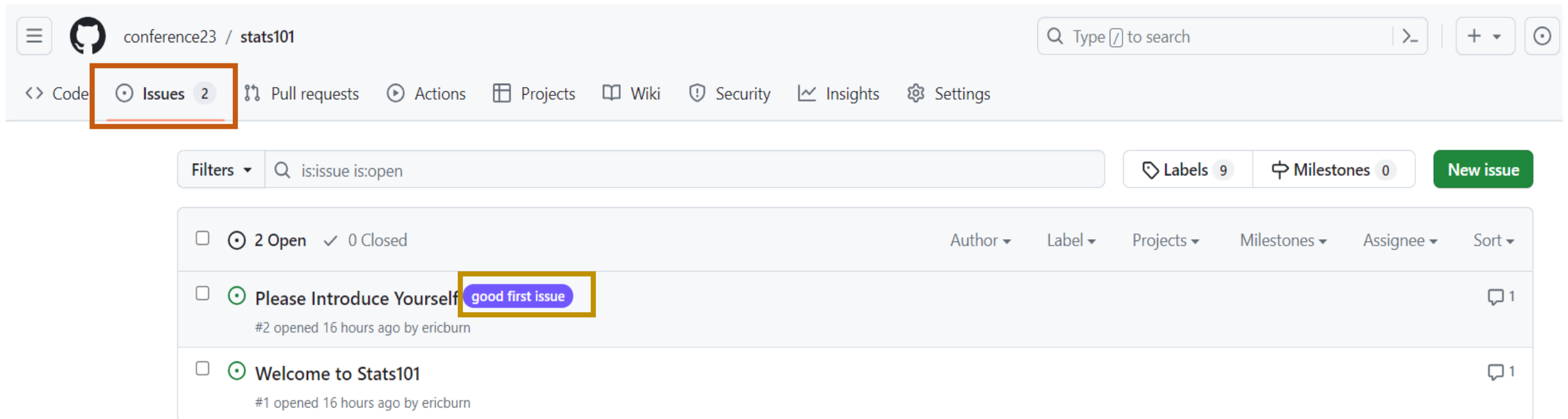
README.md Create README.md now

lecture1.R Create lecture1.R 1 minute ago

README.md

Part II: Introduce Git & GitHub through GitHub Issues

- **GitHub Issues** acts a message board
 - Students can ask questions, give feedback, share ideas, etc.
 - Students can be grouped into teams to increase small group collaboration
 - Students can “mention” others & add labels



conference23 / stats101

Code Issues 2 Pull requests Actions Projects Wiki Security Insights Settings

Filters is:issue is:open Labels 9 Milestones 0 New issue

<input type="checkbox"/>	2 Open ✓ 0 Closed	Author	Label	Projects	Milestones	Assignee	Sort
<input type="checkbox"/>	<div>• Please Introduce Yourself</div> <div>#2 opened 16 hours ago by ericburn</div>		good first issue				1
<input type="checkbox"/>	<div>• Welcome to Stats101</div> <div>#1 opened 16 hours ago by ericburn</div>						1

Part II: Introduce Git & GitHub through GitHub Issues

Please Introduce Yourself #2

 Open ericburn opened this issue 16 hours ago · 1 comment



ericburn commented 16 hours ago

Member



Hi, I'm Eric Burnheimer. Outside of teaching, I enjoy running, tennis, and reading.



MagicElizabeth commented 16 hours ago

Member



My name is Elizabeth, I enjoy magic tricks and math.



ericburn added the **good first issue** label 5 minutes ago



Add a comment

Write

Preview

H B I    |    |  @   

@MagicElizabeth

What was the last magic show you attended?

 Close with comment

Comment

Mini-Session After Semester I:

Introduction to Git & GitHub

Git & GitHub Course (1 credit course?)

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

Week 1: Core Concepts

- Git Installation
 - Command Line
- Basic Workflow
 - 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

- Create personal site
 - GitHub Pages
- Workflow
 - Pull Requests
- Final “Mini-Project”
 - Students decide when to use git tools

Other Topics to Explore

- **GitHub Pages** enables and eases the creation of a personal site
 - Provides aspirational template for students to fill as they progress

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

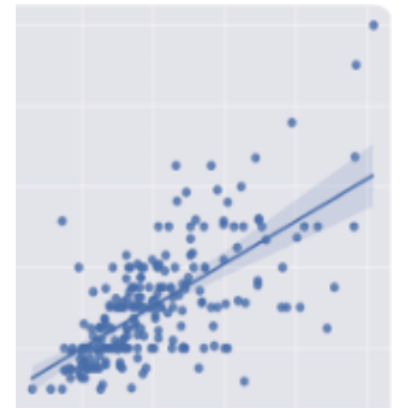
Technical Skills: R, Git & GitHub

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*)

GitHub Classroom




- **GitHub Classroom**
 - Manage a virtual classroom
 - ❖ Make Assignments
 - ❖ Auto-grading
 - ❖ 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](#)

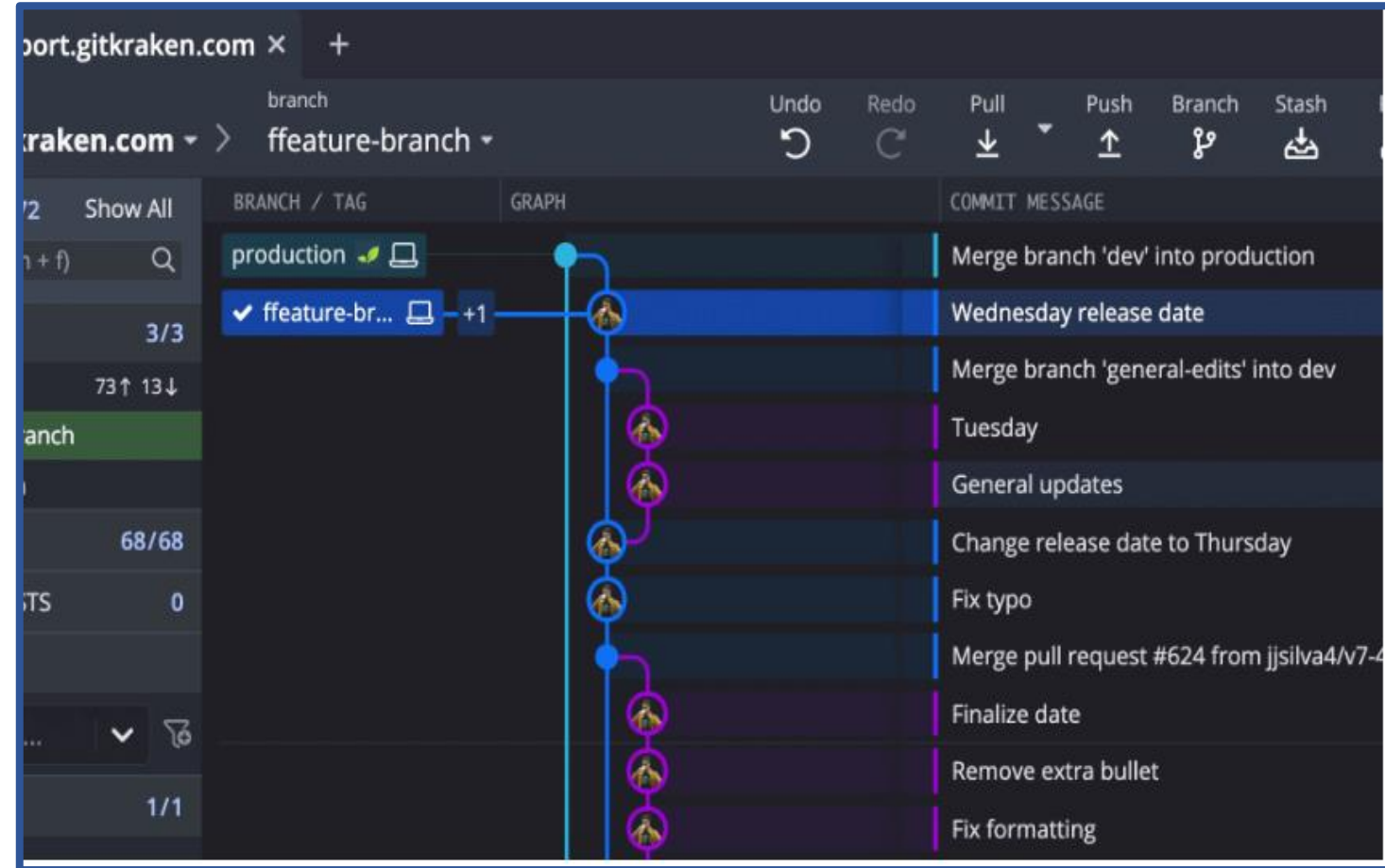
[Learn more about the GitHub Starter Assignment](#)

GUIs such as GitKraken



- **GUIs such as GitKraken**

- Point-and-Click
- Nice visualization of branches



Credit: Medium

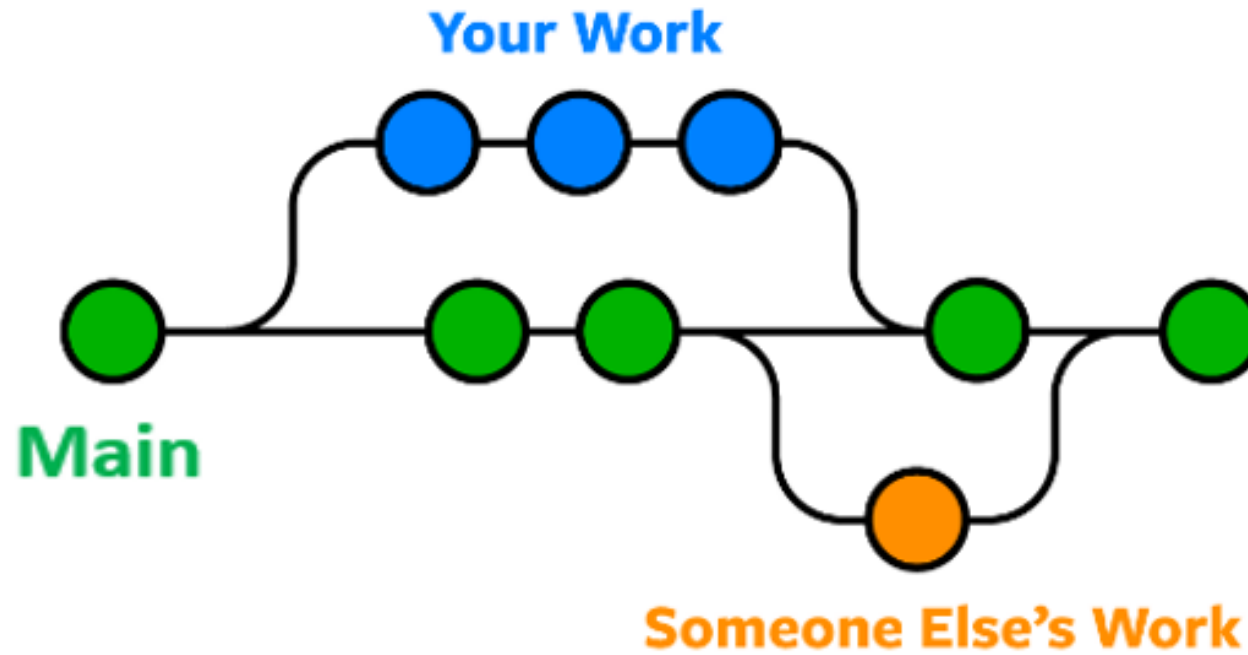
The Benefits of Implementing Git & GitHub in a DS Program

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

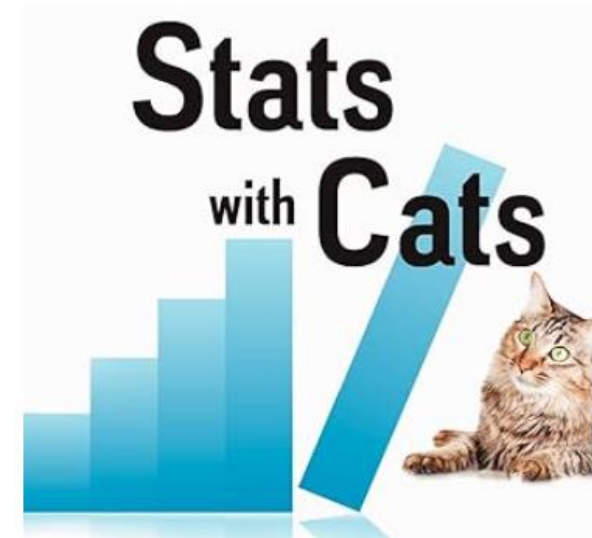
A Pathway for Implementing Git & GitHub in a DS Program

- ❖ Semester I: Introduce GitHub (and preview git) by communicating via Issues
- ❖ Mini-Session after Semester I: Git & GitHub focused course

Octo is Ready to Collaborate!



Credit: Noble Desktop



Credit: Charles Kufs

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/>

Thank you!

Eric Burnheimer
Massachusetts Bay Community College
eburnheimer@massbay.edu