GitHub and intro stats: Technology to support group and project-based learning

Nicholas J. Horton and Maria Tackett

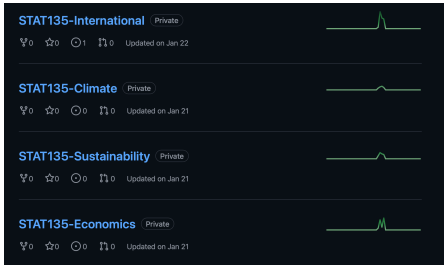
USCOTS, June 2021

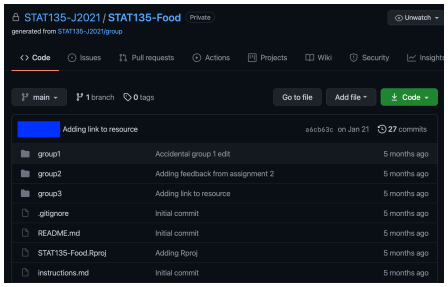
Motivation

GitHub is a powerful version control tool that supports collaboration and reproducibility

Beckman et al., 2021 and Fiksel et al., 2019 demonstrate the feasibility and value of including Git and GitHub in statistics and data science courses

- Free access available for educational purposes
- Focus on teaching basic functionality





Teaching GitHub to expand opportunities

By teaching GitHub early...

- we're preparing students for future opportunities and helping to make them more competitive when they apply to internships and jobs.
- we can help make the pool of undergraduates with these technical skills more broad and diverse.
- we expose them to tools for reproducible analysis and workflow.
- they can leverage more sophisticated use in later courses.

Why GitHub in Intro Stats?

- version control is an important component of reproducibility
- revised GAISE College report (2016) notes the importance of teaching the entire data analysis cycle
- outline a "Minimal GitHub" to support a group and project-based introductory statistics class
- focuses on supporting collaboration and interactions between students and with instructor
- goal: not to cognitively overload students

Methods

- results from a January (intensive online) introductory statistics course (n=21 students)
- used a teaching method due to Katie Kinnaird (personal communication)
- GitHub used to facilitate group and individual project deliverables and foster discussions
- focus on student experience
- data included mid- and end-of-semester assessments and GitHub records
- project approved by Amherst College IRB

Methods (cont.)

- GitHub introduced on day one of class (only web interface)
- clone repo and make commits using GitHub (day two)
- a little bit more (day three)
- for group work, roles rotated with one person responsible for commits each assignment (to minimize merge conflicts)
- for individual projects, instructor provided file structure and used issues for all deliverables
- the **`ghclass`** package used to interact with repositories

Methods (cont.)

- authors handcoded the text responses to the survey questions:
  - "What has your experience been like with Github" (mid-semester)
  - "What did you think about GitHub" (end-of-semester)
  - unable to link the responses at each time point
  - create three indicators of **`comfort`** with GitHub (yes, neutral, or no), **`useful`** (yes, neutral, or no), and **`sentiment`** (positive, neutral, negative)
- high level analysis of GitHub commits, issues, comments, and merge conflicts

Results

There were **29 total repositories** created during the semester. Each student had access to

- 1 group repository (about 3 members per group)
- 1 individual repository

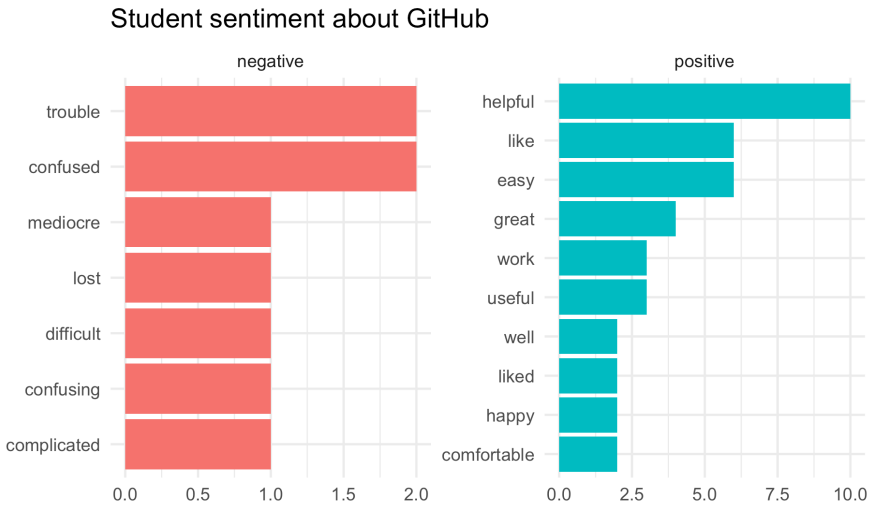
Commits per student

Repo type	Mean	Median	SD
Group	11.5	7	10.0
Individual	12.4	11	5.9

Issues per student

Type	Mean	Median	SD
Initiated	2.7	2.5	1.3
Comments	8.6	7.0	5.4

Student sentiment



Get student buy-in on the usefulness of GitHub early in the semester

Conclusions and Takehomes

- only teach **basic GitHub functionality** in intro courses
  - reinforce "pull/commit/push" (very important for group work)
- **merge conflicts** happened but imposed a relatively minor challenge
- explain the **value of learning GitHub** to prepare for internships, research opportunities, and future careers
- though there is a relatively steep learning curve, **student feedback was generally positive** by the end of the semester

Resources

- https://nicholasjhorton.github.io/Minimal-GitHub/ (other resources)
- https://www.projecttier.org/fellowships-and-workshops/2021-spring-symposium/ (TIER Symposium on teaching reproducibility)
- https://nhorton.people.amherst.edu/call_reproducibility.pdf (Call for papers, teaching reproducibility and responsible workflow)

Appendix

Student sentiment

Mid-semester (n = 21)

- all **9** students who had positive sentiment about GitHub wrote positive or neutral comments about its usefulness
  - true even if they were not yet comfortable using it

End-of-semester (n = 15)

- **9** students had positive sentiment about GitHub, saw its usefulness, and were comfortable using it
- **2** students had positive sentiment about GitHub, were comfortable using it, but felt neutral about its usefulness