Teaching Version Control in Introductory and Advanced Courses With GitHub Classroom: Experiences and Resources

Jacob Fiksel, BA
Leah Jager, PhD
Margaret Taub, PhD
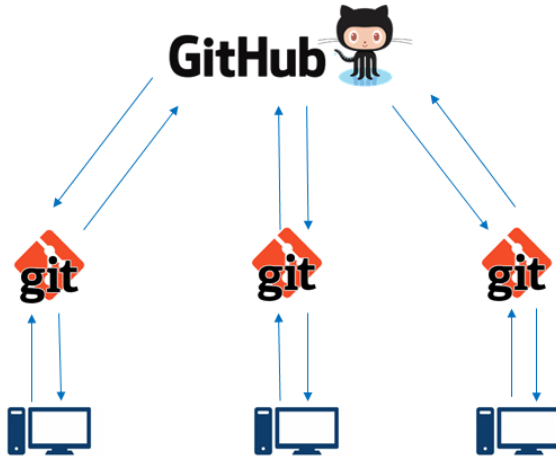Johanna Hardin (Pomona College), PhD

May 22, 2018

JOHNS HOPKINS

BLOOMBERG SCHOOL
of PUBLIC HEALTH

Data analysis is a team effort

- Businesses, governments, and researchers make data-driven decisions

- Data projects are collaborative–data, code, and results must constantly be updated and shared with the analysis team

- Need a method for managing large analysis projects
  - Git and GitHub!

Git and GitHub allow for easy management and sharing of data analytic content

GitHub allows you to track changes made during the course of a project

Why teach Git and GitHub in class?

1. Exposure to commonly used tool

   - Kaggle industry-wide survey: 58.4% of the 16,000 respondents said that Git was the main tool used for sharing code in their workplace

Why teach Git and GitHub in class?

2. Group projects

- Students get to practice common industry work sharing pipelines

- Much easier than Google docs for updating and sharing code

- Teachers see contribution of each student

Why teach Git and GitHub in class?

3. Providing feedback during assignments

- Can give feedback directly through GitHub, rather than emailing the file back and forth

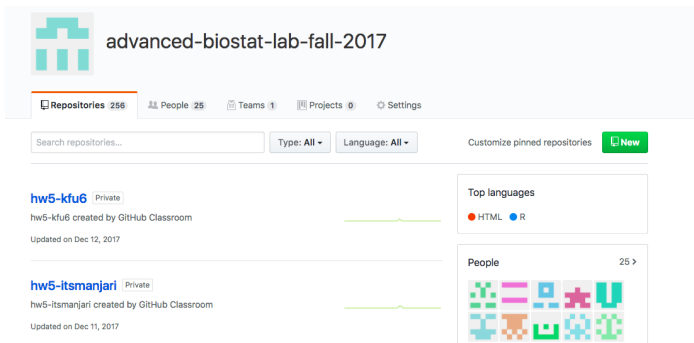- Teachers can easily grab all pieces needed to run the student's code

GitHub Classroom allows for integration of Git and GitHub into a course

- https://classroom.github.com/

- Can organize starter code for assignments

- Automated assignment (repository) creation for students

  - All students start off with same scripts, data, etc...

GitHub Classroom allows for integration of Git and GitHub into a course

GitHub Classroom can be introduced in both introductory and advanced courses

- In the Fall 2017 semester, we introduced GitHub Classroom in:

  - An introduction to statistics lab at JHU–students had little-to-no previous computing or statistics experience
    - Taught by: Leah Jager and Margaret Taub
    - Teaching assistant: Jacob Fiksel
    - ~ 20 students

  - An advanced computational statistics course at Pomona College–students had experience with statistical computing, but not Git or GitHub
    - Taught by: Johanna Hardin
    - ~ 70 students

Lessons from using GitHub Classroom in an introductory course

- Challenge: introduce Git and GitHub to students with no computational experience
  - Git and GitHub "bootcamp" before first class
  - High start-up cost

- Student feedback
  - "I did not enjoy setting [Git and GitHub] up and the instructions seemed really tedious"
  - "I did not like that [Git] was quite hard to install and that issues occur when you [update] your computer"

Lessons from using GitHub Classroom in an introductory course

- Students appreciated learning version control (eventually)

  - "I really enjoyed using GitHub because it's applicable to the other things that I do in my lab"

  - "I liked that we can save our past work on GitHub and that it taught me how to take advantage of it for other projects"

Recommendation: tailor use of GitHub Classroom to student background

- For advanced courses: use GitHub Classroom to start and have students work on team projects

- For introductory courses: introduce Git and GitHub as its own unit after students have gotten more comfortable with computing

Resources for implementing GitHub Classroom

- We have created a detailed guide for teachers to start their own GitHub Classroom
  - https://github.com/jfiksel/github-classroom-for-teachers

  - Step-by-step instructions on everything from assignment creation to grading workflow

- We also have a guide and videos for beginner students who have never used Git or GitHub
  - https://github.com/jfiksel/github-classroom-for-students
  - Covers set-up on both Windows and Apple

- Questions, comments, and collaborations on these guides are more than welcome!