Industry & Academic Peparedness: Version Control and Reproducible Research Outside of the Classroom

Adam J Sullivan
Takeda Pharmaceuticals Company

August 4, 2020
JSM Virtual Conference

Industry: Version Control

- Very important in both highly regulated and other industries.
- Why version control in Industry?
  - Maintains accurate records and history of changes.
  - Allows for regulatory bodies to see full story.
  - Ease for onboarding new members on team

Version Control Example: Basic Documents

- Version control is not just for code.
- In pharma we have version control systems to maintain all regulatory documents on a study.
- This allows for full transparency of the entire clinical trial planning process.
- In addition to this, trials from pre-planning to full completion can take 5-10 years, this leads to turn over in key positions.
- Version control allows an onboarding member to understand history an path of the project.

Industry: Reproducible Research

- It is important for businesses to be able to reproduce results.
- This can come in the form of monthly reports or even interim analysis.
- Tables, figures and listings all need to be created and updated for new data.
- Many of these span accross projects as well.
- Time and Money can be saved by having code and work able to be reproduced within a project or across project teams.

Academic Preparedness

- Learning version control and reproducible research can benefit students outside of your statistics classroom.
- Experience in these skills can benefit:
    - Future math and statistics courses.
    - Future academic research.
    - skills needed for internships.

Academic Preparedness: Future math and stats courses

- Students benefit from being able to keep and track neat code and research.
- Statistics and mathematics are highly detail oriented and it is important to document all steps.
- Learning this skills early creates students with better workflows for more advanced courses.

Academic Preparedness: Future academic research

- Working on research is important for students skillset but can be very time consuming.
- Many students have no real organization skills and cannot track versions of their work well.
- I have personally read and commented on close to 100 pages only to have the student come back and say, " I sent the wrong version can you read my updated one."
- Students who learn to version control are much easier to work with and their work is better prepared.
- In addition, they are taught in a manner that creates excellent organization and attention to detail.

Academic Preparedness: Internship Skills

- Many students in stats courses are interested in data science.
- Computing and coding is an easy skill to have at an entry level internship.
- Learning a programming language as well as verison control makes a student attractive to many data science internships.
- My students have had great success in large banking and data science companies using these skills when they did not have a lot of statistical or modeling skills yet.

Git

- We mostly discuss Git and Github as you will see coming up next.
- This is mainly for the ease of inclusion into the academic environment.
- skills I learned while teaching this stuff has helped when the company I work for now does not use github.