

Utilizing Open Source Resources to Teach Introductory Data Science

Tyler George

tgeorge@cornellcollege.edu

 @stats-tgeorge

 stats_tgeorge

Slides: bit.ly/3ddpp8V



About Cornell College

- Small liberal arts college of about 1000 students
- One-Course-at-A-Time block schedule
 - Each class occurs over 3.5 weeks including 18 days of 4 hours of instructional time and three weekends
- **New majors in Applied Statistics and Data Science**

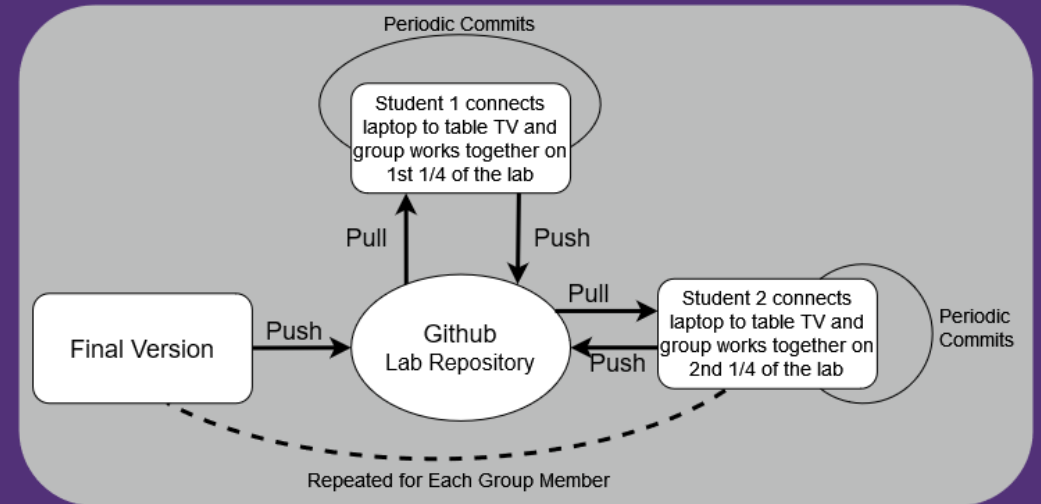


Utilized Resources

- [Data Science in a Box](#) by Mine Çetinkaya-Rundel (2021)
- Data science ethics assignments from the [Quantitative Analysis Institute](#) at Wellesley College (Pattanayak, Gan, Li, Liang, and Wong, 2022)
- Workshops by Julia Silge available at <https://juliasilge.github.io/tidytext-tutorial/site/>
- Many R packages were used throughout the class but [ghclass](#) was particularly useful for teaching/running the course

Teaching Pedagogies

- Collaborative learning (Roseth, Garfield, and Ben-Zvi, 2008)
- Using tidyverse (Çetinkaya-Rundel, Hardin, Baumer, McNamara, Horton, and Rundel, 2022)
- Version control to lead to reproducible work (Beckman, Çetinkaya-Rundel, Horton, Rundel, Sullivan, and Tackett, 2021)



Learning the Bare Necessities

I was new to Git, Github, RStudio Server (Linux), and a handful of the tidyverse packages and functions.

- What did I **need** to learn to teach this class in this way?
 - Basic Git functions such as push, pulling, and handling git merge conflicts in RStudio
 - *ghclass* R package made using Git/Github easier. See my course materials.
 - Some new packages and functions in the *tidyverse* such as *forcats*
- What are some extra things I learned?
 - RStudio server implementation and management in Ubuntu (Linux) [Use Examples](#)
 - Learning the Xaringan R package to modify slides

All course materials can be found at bit.ly/3Qq99Qb



Future Updates

- Expanding the ethics integration in the course
- Trying out the new [gradetools](#) R package
- Less statistical inference and more data wrangling
- Higher project and assessment expectations.
- Github actions
- Quarto



References

See the bibliography at bit.ly/3bCuh6L

Acknowledgment

Dr. Ajit Chavan, Assistant Professor of Computer Science at Cornell College, for setting up the cluster used to run the RStudio Server and answering many of my questions about its use.

Questions?

Thank You!

JMS 2022



Cornell College