

DATA SCIENCE PRACTICUM

SPRING 2021 WRAP-UP LECTURE



OUTLINE

1. Some fun stats
2. Course feedback
3. Wrap-up

COMMITTS

- **511** commits
- **2,771,822** additions, **186,181** deletions
- **2,585,641** new lines of code

* as of Wednesday, April 21 at 11:20am

TAKEAWAYS

- **Reuseability is critical** in academia and industry frameworks
 - Documentation
 - READMEs
 - Comments
 - Development process (Issues, Pull Requests, Gitter/Slack/Listservs)
- The real world operates in **teams**
- Reusability + teamwork = **effective communication is absolutely essential**
 - **More important** than Kaggle leaderboard position (what's the point if only you can understand it)
 - **More important** than raw coding talent (raw coding talent is dime/dozen)

FEEDBACK

- **Activity time!**
- Three categories of feedback:
 1. **STOP:** What would you *remove* or *eliminate* from the course?
 2. **START:** What would you *add* to the course?
 3. **CONTINUE:** What would you *keep* or *retain* that is already present in the course?
- **Individually, write down one item for each category (5 minutes)**
- **Get in groups and develop a consensus list (10 minutes)**

FEEDBACK

MY THOUGHTS

- STOP
 - GCP (maybe Azure?)
- START
 - Self-proposed / self-organized final project (existed in Spring 2019, eliminated this semester due to COVID)
- CONTINUE
 - Project peer feedback forms


YOUR THOUGHTS

FEEDBACK

- START
- STOP
- CONTINUE


FEEDBACK FROM SPRING 2019

- STOP

- Two computer vision projects is too much 


- Did both Neuron Segmentation (didn't do this year) and Ciliary Segmentation (our P3) in Spring 2019

- START

- New category of projects like unsupervised clustering, matrix factorization, reinforcement learning 

- *Almost* got a StarCraft RL project working; ran out of time (but should be ready for next 8360)

- Project peer feedback forms 

- Add Dask as an alternative to Spark 

- CONTINUE

- Project 0 (maybe extend in length a tad) 

- From 1 week to 2 weeks

OTHER THOUGHTS

- Teams aren't going anywhere
- Lectures will be tweaked for additional background knowledge
 - Also, 3360 Data Science I is proposed to become a 4000/6000 course to allow graduate enrollment
 - When that happens, DSI will be a **required prerequisite** to 8360 (no more “toughing it out”)
- Submit *programs* to AutoLab, rather than just predictions
- **EXTRA CREDIT: By midnight, Friday April 30, propose a new project idea.** Needs:
 - A *clear, unambiguous ground-truth* (or evaluation metric) to put in AutoLab
 - An openly available dataset (or one that can be acquired, e.g., Faces or web documents)
 - Can achieve a reasonable solution in 2-3 weeks

OTHER THOUGHTS

- Course evaluations!

<https://webapps.franklin.uga.edu/evaluation/>

- Open until **May 3!**

FINAL NOTES

- Raw scores aren't everything
- ...but they're a reasonable initial proxy
- Balance exploration (examining the data, testing multiple approaches) with exploitation (designing, testing, and documenting a complete pipeline)
 - Corollary: *start early!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!*
- Communicate. Communicate. Communicate.

Questions?

THANK YOU!

