PS239T: Computational Tools and Techniques for Social Science

Instructor(s): Jae Yeon Kim (jaeyeonkim@berkeley.edu),
Nick Kuipers (nkuipers@berkeley.edu)

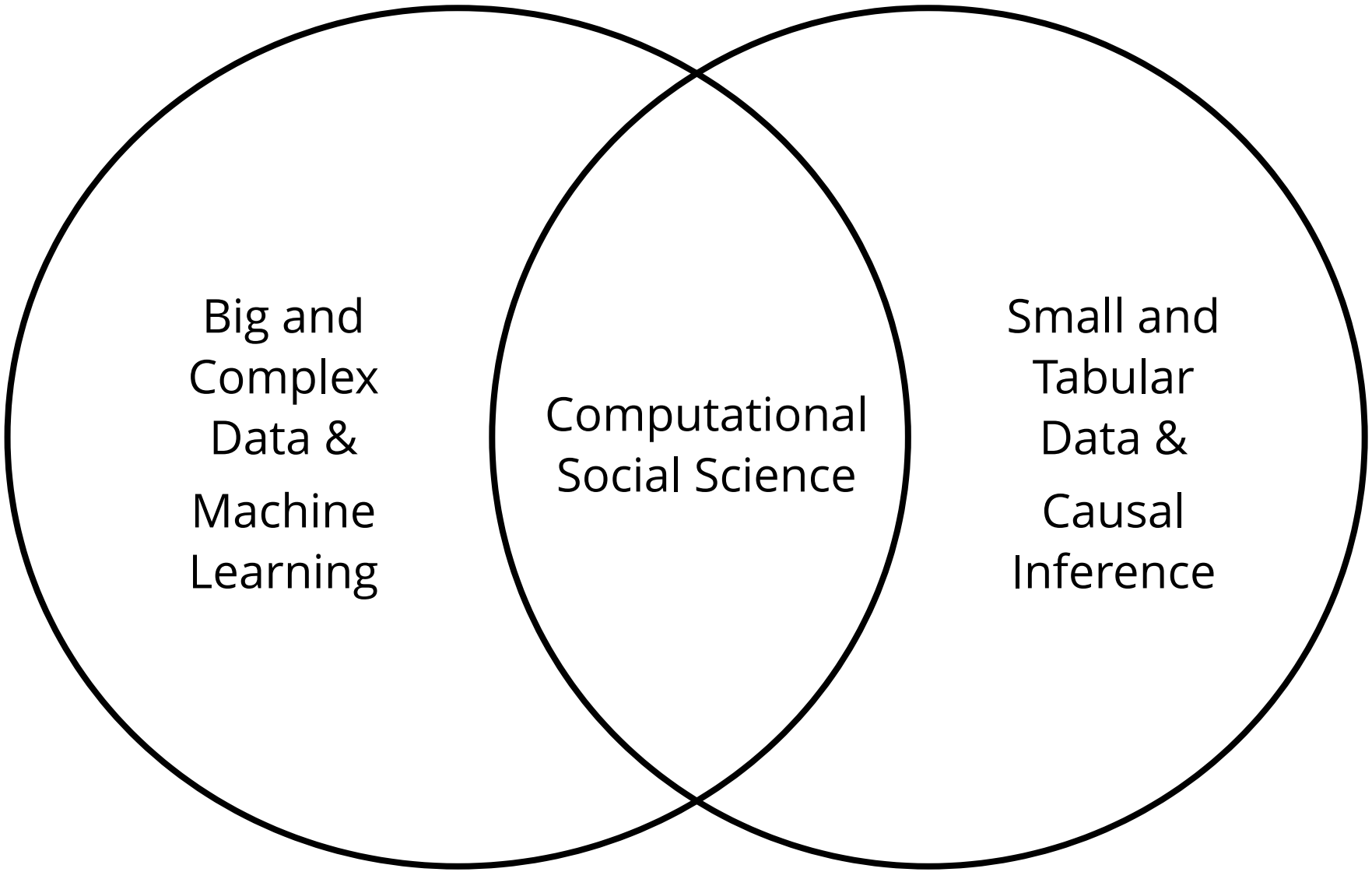
UC Berkeley, Spring 2021

Seriously, how are you doing?

Ice Breaker

Your name, field, and the most memorable data analysis mistake you've made

Exciting time to be in computational social science


Big and Complex Data & Machine Learning
Computational Social Science
Small and Tabular Data & Causal Inference

ML becomes mainstream

https://www.youtube.com/embed/nZdQQcR-wls?enablejsapi=1

https://www.youtube.com/embed/Yx6qXM_rfKQ?enablejsapi=1

- Collecting and analyzing large-scale data
- Utilizing complex data (text, image, speech, audio, video, etc)
- Finding optimal ways of doing surveys and experiments

Why Data Scence Boom?

1. Data: Scale and Complexity

2. Tools: Poweful and Human-friendly Programming Languages + So Many Awesome Packages (Did I tell you they are open source?)

What Does It Mean to Apply Computational Tools and Technieus to Social Science Research?

Decompose your research design and find which part needs computational solutions (=robotic assistants)

Doing this is going to be your FINAL PROJECT and you will receive ample feedback and support to succeed.

10 mins break to discuss your potential projects with your classmates

Teaching Philosophy & Code of Conduct

Arvind Narayanan
@random_walker

An amazing benefit of my privilege is being able to say "I didn't understand that. Could you explain it again?" as many times as necessary without having to worry that people will think I'm stupid.

9:59 AM · Aug 26, 2020 · Twitter Web App

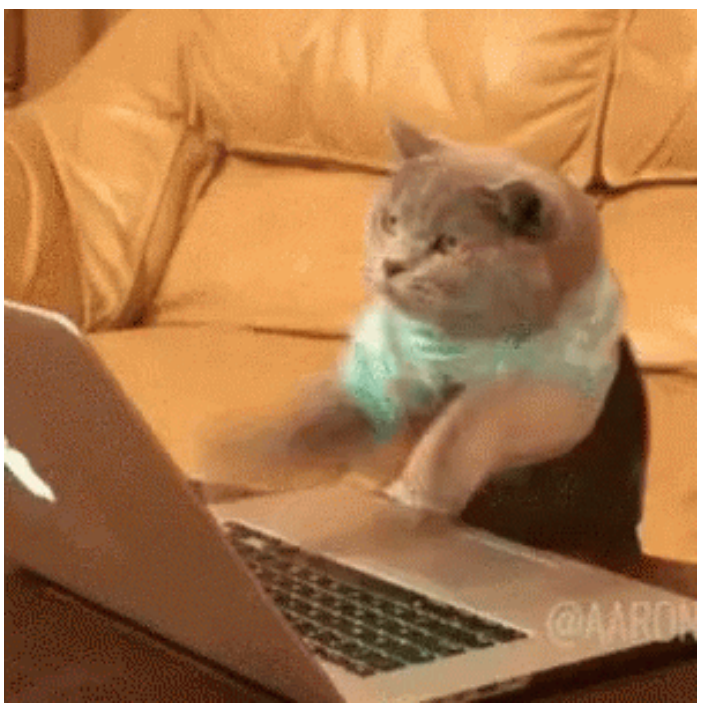
126 Retweets **33** Quote Tweets **1.4K** Likes

It's Okay Not to Know (D-Lab's motto)

Ask us any questions and ask us many times!

One more thing

It's not enough to get it done



You should know what you're doing and how you can do better

- We focus on helping you to write efficient, readable, and reusable code.
- Don't Repeat Yourself (DRY)
- Don't work hard, work smart.

Part 1 Fundamentlas

1. Data and code management (Bash + Git)

2. How to wrangle, model, and visualise data easier and faster

3. How to automate repeated things and develop data products

Part 2 Applications

1. Digital data collection (PDF, websites, social media posts)

2. Machine learning (supervised and unsupervised)

3. Big data and SQL

- Syllabus: https://github.com/PS239T/spring_2021/blob/main/A_Syllabus.pdf
- Course GitHub repo: https://github.com/PS239T/spring_2021
- Open textbook: https://jaeyk.github.io/PS239T/

These are all subject to change during the semester. Please feel create issues if you have questions/comments/suggestions.

- All conversations -> Slack
  - If you haven't, you will get a Slack workspace invitation shortly.
- All course materials + assignments (except final project proposals) + questions/comments -> GitHub
  - Work in a group but submit assignments individually.

- Slack tips:
  - `\` finding features
  - `@` tagging people
  - `#` using hashtags
- Git and GitHub
  - No worries! We'll cover how to use Bash and Git in the first two weeks.
- But it's very important to install everything required following this guideline: https://github.com/PS239T/spring_2021/blob/main/B_Install.md

Questions/ Comments?