CS 102: Working with Data
Tools and Techniques

Spring 2020

Course Staff

Instructor

Jennifer Widom

Course Assistants

Leo Mehr (head)

Kyle D'Souza

Tara Iyer

Aamir Rasheed

Zoom Lecture Protocol

All students on mute, prefer video on

1) For general questions: “Everyone” chat, Prof. Widom will keep an eye on it

2) For private questions: Chat to one of the four TAs, preferably Kyle or Aamir

3) For Prof. Widom’s questions to class: Use “raise hand” feature, will be called on and unmuted

What's This Course About?

"Aimed at non-CS undergraduate and graduate students who want to learn a variety of tools and techniques for working with data. Many of the world's biggest discoveries and decisions in science, technology, business, medicine, politics, and society as a whole, are now being made on the basis of analyzing data sets. This course provides a broad and practical introduction to working with data: data analysis techniques including databases, data mining, machine learning, and data visualization; data analysis tools including spreadsheets, Tableau, relational databases and SQL, Python, and R; introduction to network analysis and unstructured data. Tools and techniques are hands-on but at a cursory level, providing a basis for future exploration and application. Prerequisites: comfort with basic logic and mathematical concepts, along with high school AP computer science, CS106A, or other equivalent programming experience."

What's This Course About?

"Aimed at non-CS undergraduate and graduate students who want to learn a variety of tools and techniques for working with data. Many of the world's biggest discoveries and decisions in science, technology, business, medicine, politics, and society as a whole, are now being made on the basis of analyzing data sets. This course provides **a broad and practical introduction to working with data: data analysis techniques including databases, data mining, machine learning, and data visualization; data analysis tools including spreadsheets, Tableau, relational databases and SQL, Python, and R; introduction to network analysis and unstructured data**. Tools and techniques are hands-on but at a cursory level, providing a basis for future exploration and application. Prerequisites: comfort with basic logic and mathematical concepts, along with high school AP computer science, CS106A, or other equivalent programming experience."

What's This Course About?

"Aimed at non-CS undergraduate and graduate students who want to learn a variety of tools and techniques for working with data. Many of the world's biggest discoveries and decisions in science, technology, business, medicine, politics, and society as a whole, are now being made on the basis of analyzing data sets. This course provides a broad and practical introduction to working with data: data analysis techniques including databases, data mining, machine learning, and data visualization; data analysis tools including spreadsheets, Tableau, relational databases and SQL, Python, and R; introduction to network analysis and unstructured data. **Tools and techniques are hands-on but at a cursory level, providing a basis for future exploration and application**. Prerequisites: comfort with basic logic and mathematical concepts, along with high school AP computer science, CS106A, or other equivalent programming experience."

Who Should Take It?

"Aimed at non-CS undergraduate and graduate students who want to learn a variety of tools and techniques for working with data. Many of the world's biggest discoveries and decisions in science, technology, business, medicine, politics, and society as a whole, are now being made on the basis of analyzing data sets. This course provides a broad and practical introduction to working with data: data analysis techniques including databases, data mining, machine learning, and data visualization; data analysis tools including spreadsheets, Tableau, relational databases and SQL, Python, and R; introduction to network analysis and unstructured data. Tools and techniques are hands-on but at a cursory level, providing a basis for future exploration and application. Prerequisites: comfort with basic logic and mathematical concepts, along with high school AP computer science, CS106A, or other equivalent programming experience."

Who Should Take It?

"**Aimed at non-CS undergraduate and graduate students who want to learn a variety of tools and techniques for working with data.** Many of the world's biggest discoveries and decisions in science, technology, business, medicine, politics, and society as a whole, are now being made on the basis of analyzing data sets. This course provides a broad and practical introduction to working with data: data analysis techniques including databases, data mining, machine learning, and data visualization; data analysis tools including spreadsheets, Tableau, relational databases and SQL, Python, and R; introduction to network analysis and unstructured data. Tools and techniques are hands-on but at a cursory level, providing a basis for future exploration and application. Prerequisites: comfort with basic logic and mathematical concepts, along with high school AP computer science, CS106A, or other equivalent programming experience."

Who Should Take It?

"Aimed at non-CS undergraduate and graduate students who want to learn a variety of tools and techniques for working with data. Many of the world's biggest discoveries and decisions in science, technology, business, medicine, politics, and society as a whole, are now being made on the basis of analyzing data sets. This course provides a broad and practical introduction to working with data: data analysis techniques including databases, data mining, machine learning, and data visualization; data analysis tools including spreadsheets, Tableau, relational databases and SQL, Python, and R; introduction to network analysis and unstructured data. Tools and techniques are hands-on but at a cursory level, providing a basis for future exploration and application. **Prerequisites: comfort with basic logic and mathematical concepts, along with high school AP computer science, CS106A, or other equivalent programming experience.**"

Who Shouldn't Take It?

Computer Science or MCS students

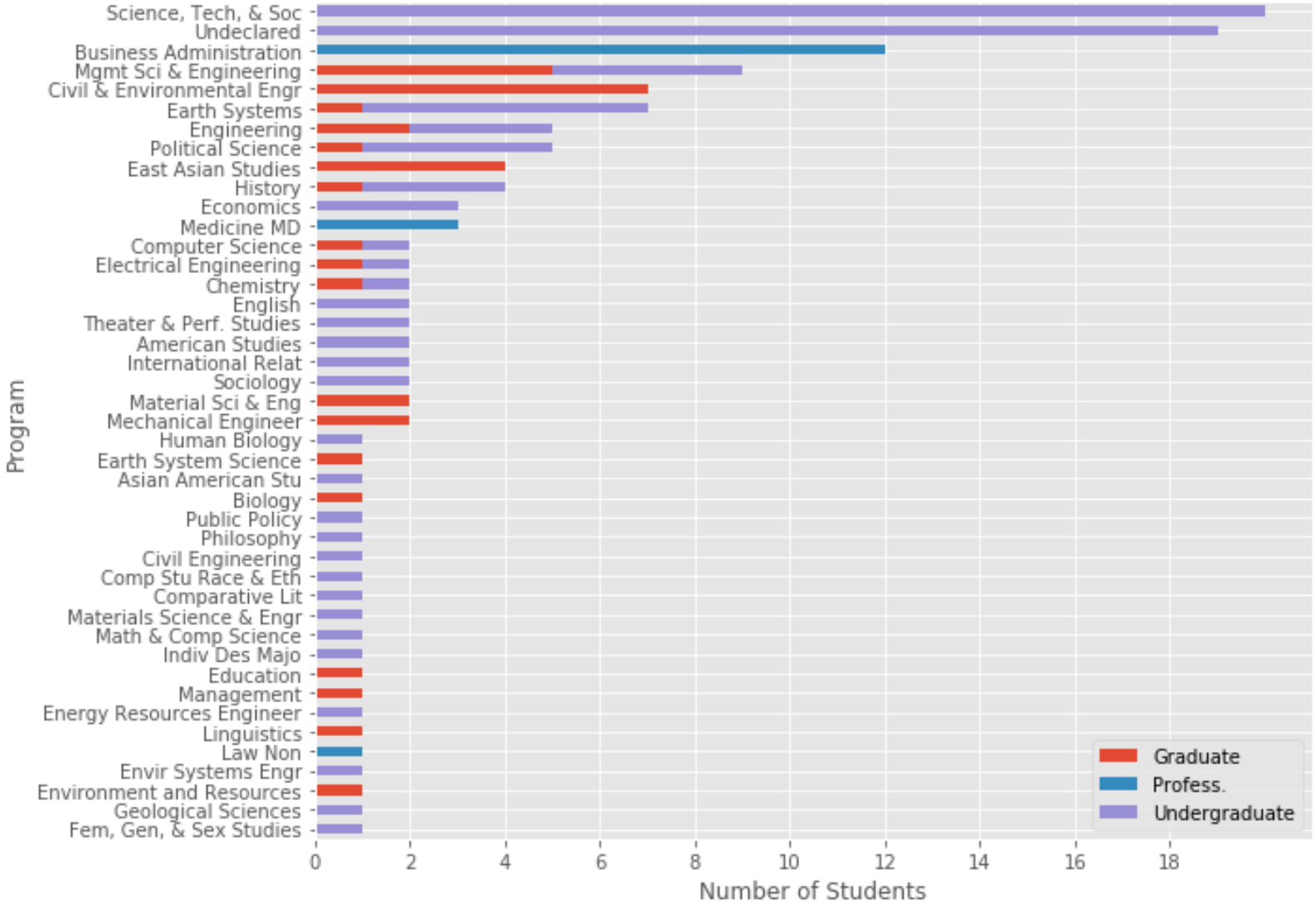
(except by petition)

Who's Taking It - Spring 2020

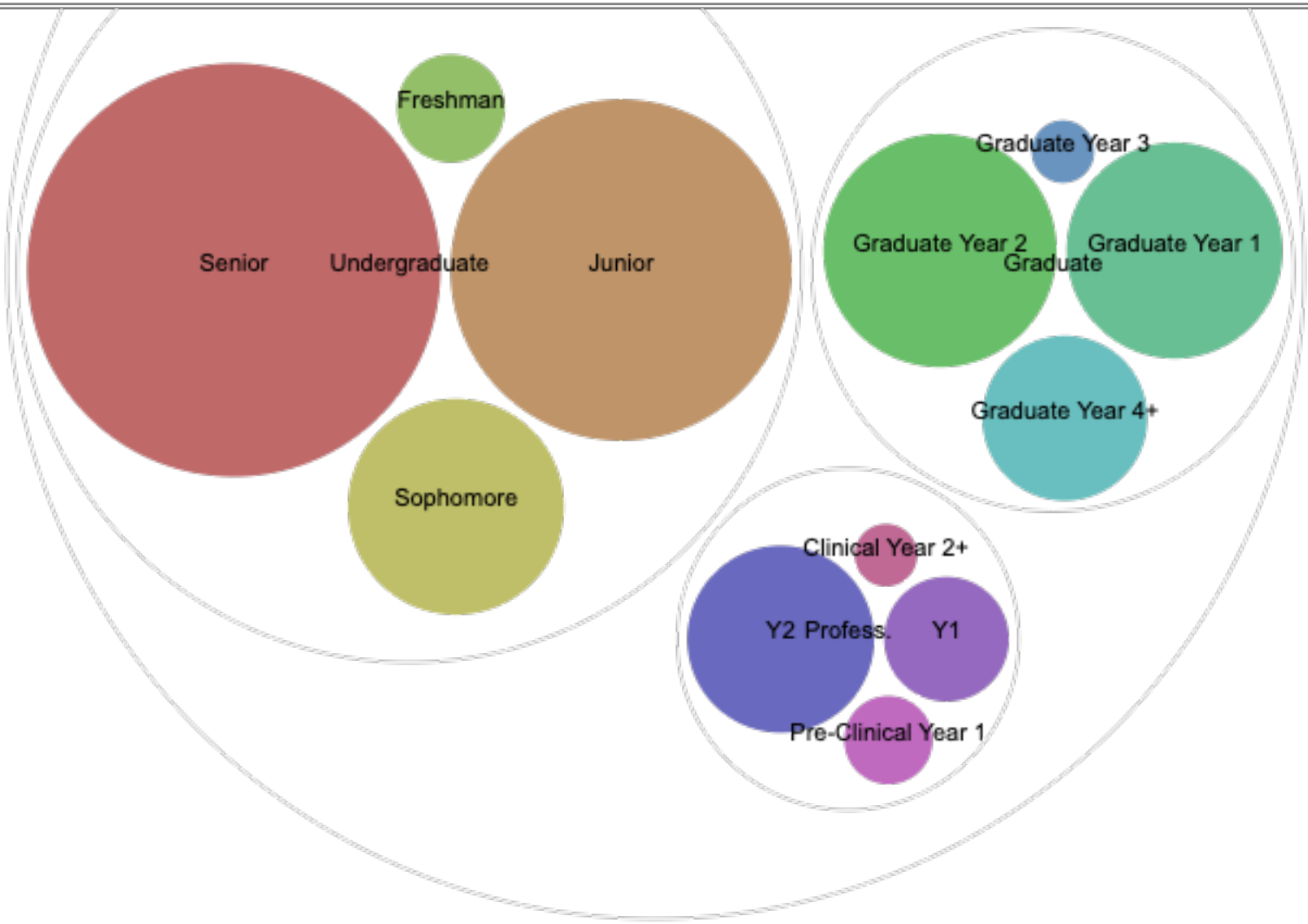
Undergraduate, Masters, MBA, MD, PhD
All seven of Stanford's schools, 42 different majors

American Studies
Asian American Studies
Biology
Business Administration
Chemistry
Civil Engineering
Civil & Environmental Engineering
Comparative Studies in Race & Ethnicity
Comparative Literature
Computer Science
Earth System Science
Earth Systems
East Asian Studies
Economics
Education
Electrical Engineering
Engineering
Energy Resources Engineering
English
Environment and Resources
Environmental Systems Engineering
Feminist, Gender, & Sexuality Studies
Geological Sciences
History
Human Biology
Individually Designed Major
International Relations
Law
Linguistics
Management
Management Science & Engineering
Materials Science & Engineering
Math & Computational Science
Mechanical Engineering
Medicine
Philosophy
Political Science
Public Policy
Science, Technology, & Society
Sociology
Theater and Performance Studies
Undeclared

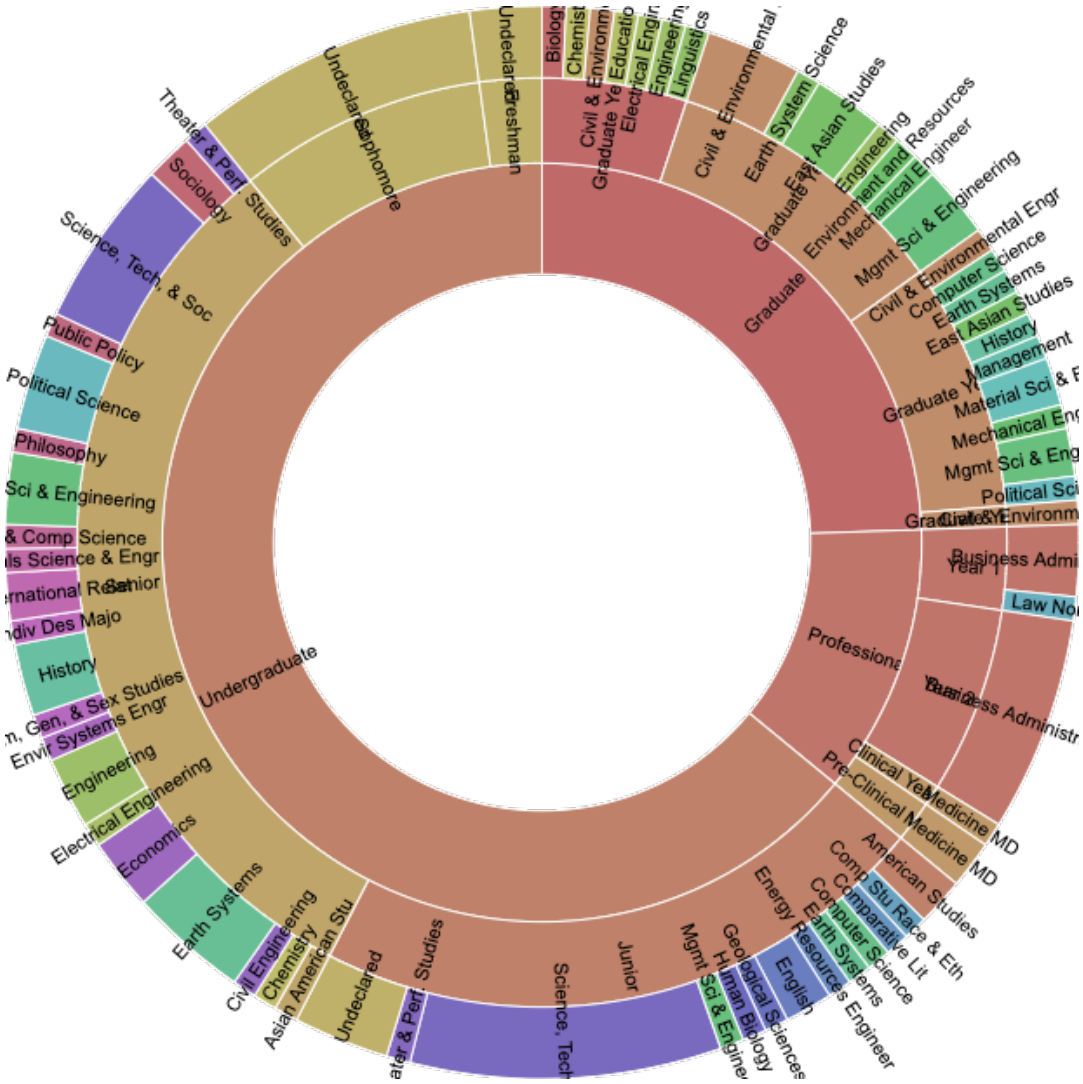
Who's Taking It



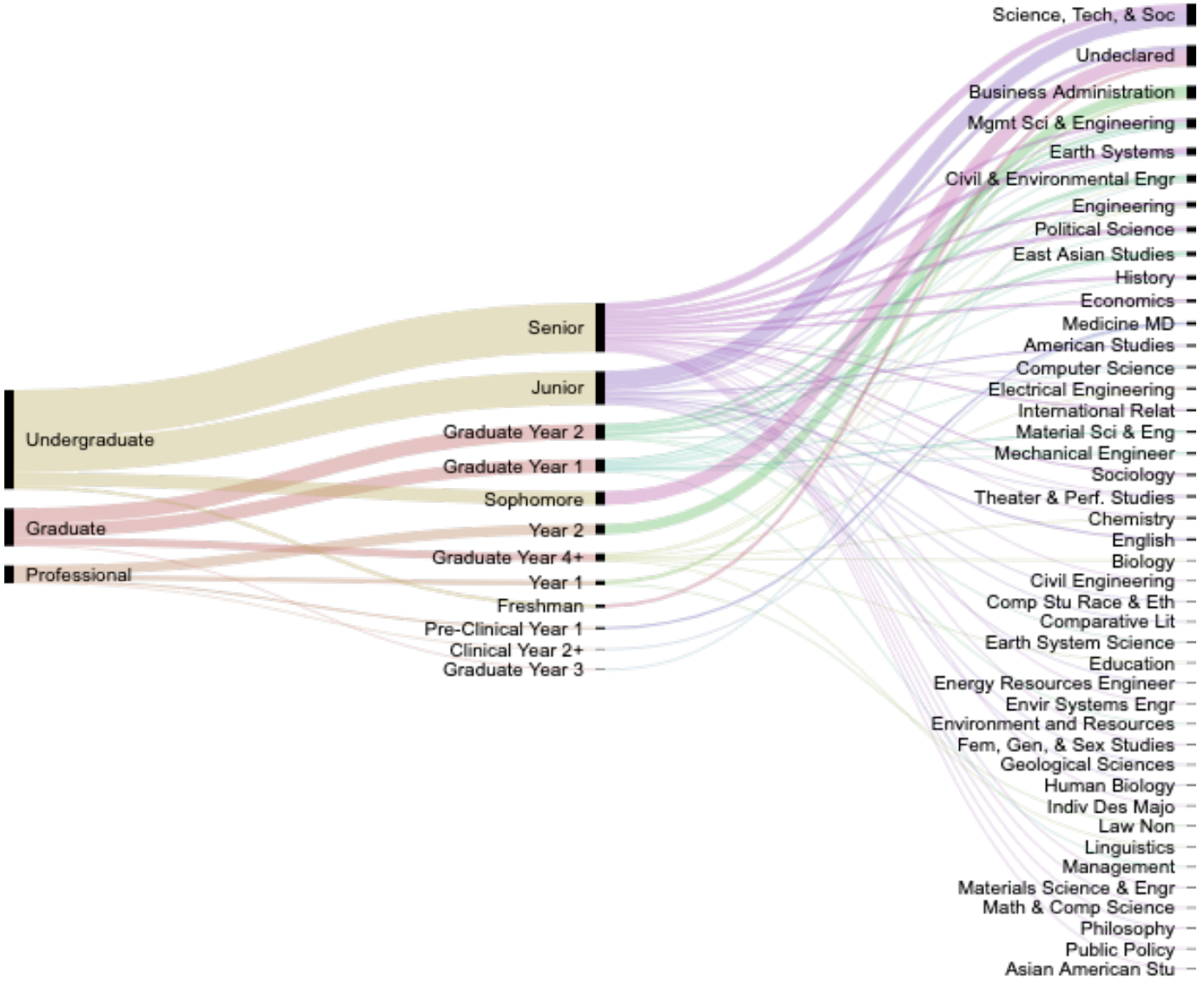
Who's Taking It



Who's Taking It



Who's Taking It



Ordering of Course Topics

- Data Analysis & Visualization Using Spreadsheets
- Advanced Data Visualization Using Tableau
- Relational Databases and SQL
- Python for Data Analysis & Visualization
- Machine Learning – Regression, Classification, Clustering
- Using Python for Machine Learning
- The R Language
- Data Mining Algorithms
- Data Mining Using Python (and SQL)
- Network Analysis
- Unstructured Data
- Correlation and Causation

Assigned Work

Assignment/Project	Assigned	Due
Assignment #1 Spreadsheets for Data Analysis and Visualization	April 13	April 20
Project #1 Personal Data Analysis	April 13	April 27 May 18
Assignment #2 Data Visualization Using Tableau, SQL	April 20	April 30
Assignment #3 Python for Data Analysis and Visualization	April 30	May 9
Assignment #4 Machine Learning, R Language	May 18	May 25
Project #2 Movie-Rating Predictions	May 18	June 1
Assignment #5 Data Mining, Network Analysis	May 28	June 5

Exams

Exam	Date
Exam #1 During class time*	May 12
Exam #2 During class time*	June 9

*Alternate times (but not dates) available by petition

Honor Code

Under the Honor Code at Stanford, **you are expected to submit your own original work for assignments, projects, and exams.** On many occasions when working on assignments or projects (but never exams!) it is useful to ask others – the instructor, the TAs, or other students – for hints, or to talk generally about aspects of the assignment. Such activity is both acceptable and encouraged, but **you must indicate on all submitted work any assistance that you received.** Any assistance received that is not given proper citation will be considered a violation of the Honor Code. In any event, **you are responsible for understanding, writing up, and being able to explain all work that you submit.** The course staff will pursue aggressively all suspected cases of Honor Code violations, and they will be handled through official University channels.

Logistics

- Units - 4 for undergraduates, 3-4 for graduates
- WAYS requirement - Applied Quantitative Reasoning (WAY-AQR)
- Textbook? No Readings? Recommended
- Class "attendance" – Expected
  - Hand-on activities
  - Only cursory notes
  - All class material game for exams

Logistics

- Grading - Letter grades calculated, C- or above for S, otherwise NC
- Grade weighting - 1/3 each assignments, projects, exams
- Graded on a curve? Not really
- Late policy - 10%/30% for 24/48 hours late, four free late days

Office Hours

TA office hours – via Zoom

- ~15 hours per week
- Times can vary

Always check the course calendar for times and links!

Prof. Widom office hours – via Zoom

- Wednesdays 4:00-5:00 PM (usually)

Online

Website - http://cs102.stanford.edu

Canvas - Zoom lectures and help sessions, recordings posted afterward

Piazza

- Announcements
- Q&A (private and public)
- Discussion

Gradescope - Assignment submission & grading

For Thursday's Class

1) Get set up on Google Drive if you're not already
2) Download Europe city temperatures data from course website (three files)
3) Copy data files into Google Drive, make sure you can open with Google Sheets
4) Be prepared to work on your computer alongside the video

Set-up help session On Wednesday

CS 102: Working with Data
Tools and Techniques

Questions?