Computational Structures in Data Science

Lecture 1:
Welcome to C88C!

Berkeley
UNIVERSITY OF CALIFORNIA

Michael Ball | UC Berkeley |

Useful Links

- Website: https://c88c.org
- Ed Thread: https://go.c88c.org/qa1
- Self-Check: https://go.c88c.org/1

Goals today

- Introduce you to
  - the field
  - the course
  - the team
- Answer your questions

- Big Ideas:
  - Expressions

CS88 Team –Michael

- Michael Ball
  - ball@Berkeley.edu – You're best off by using Ed! ☺
  - 784 Soda Hall
  - http://michaelball.co – I don't update this much…
    - Websites are great for procrastination + learning
  - Office "Tea" hours: tentatively Monday afternoon. "coffee chat"
    - My office hours should be conceptual, higher level advice. ☺
- Things I do (aside from Teaching)
  - Intro CS Research (Tools, curriculum)
  - Training TAs
  - Building Educational Software (Flextensions, Snap!, frmr Gradescope)
  - Tools for web accessibility

C88C TAs

Isabelle Ng she/her/hers

@ DSP isabelle.ng@berkeley.edu

Hi there! I am a senior CS/DS/Music Major from San Jose, CA and this is my 4th semester being a TA for c88c/cs61a. I love writing songs and learning more about HCI and NLP. Excited to meet you all and have an awesome semester!

Rebecca Dang she/her/hers

@ DSP rdang@berkeley.edu

Hey there, I'm a 5th year MS EECS student and this is my fifth semester teaching C88C! Happy to chat about this course, classes and clubs at Berkeley, professional development, guitar, books, movies, TV, music, and more :D

Dhruv Syngol he/him/his

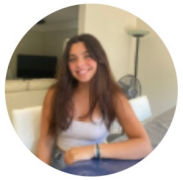
@ dhruvsyngol@berkeley.edu

Hey everyone, I'm a junior studying Data Science and Economics, originally from the Chicago Suburbs! I love to play pickleball, explore cafes and restaurants, and go on hikes! Super excited for this semester!

Grace Xie she/her/hers

gracexie@berkeley.edu

Hello! My name is Grace. I'm a third-year majoring in MCB and Data Science :0 I love reading sci-fi and baking in my free time.



Mira Wagner she/her/hers

@ DSP mirawagner@berkeley.edu

Hi! I am a junior majoring in data science and linguistics. I love reading, especially mysteries, swimming and baking! Excited for this semester :)

Alicia Wang she/her/hers

@ awwang629@berkeley.edu

Hi! I am a junior studying Data Science and Cognitive Science. I love playing badminton and traveling! Excited to meet everyone!

Grace Baek she/her/hers

gracebaek@berkeley.edu

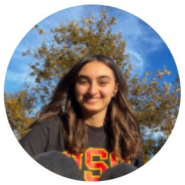
Hi! I'm Grace, a senior majoring in Computer Science and Economics and this is my 2nd semester teaching C88C. In my free time, I like baking, trying new cafes, and watching kdramas :) Super excited to meet everyone!

C88C Tutors

Cynthia Shao she/her/hers

cynthia_shao@berkeley.edu

Hey! I'm Cynthia, a sophomore in data science. You can find me dancing in The 510, jogging around campus, and finding/making/eating food at 1AM. See you in session :)



Maryam Akelyan she/her/hers

maryam.akelyan@berkeley.edu

Hi! I'm Maryam, a junior majoring in Data Science and MCB, and I'm from La Crescenta, CA! In my free time I like to watch sitcoms and sci-fi shows. Super excited to meet you all! :)

Orazaly Kabdrakhmanov he/him/his

kabdrakhmanov@berkeley.edu

Hi there! I'm Orazaly and I am a senior majoring in Data Science. I like playing tennis and video games(CS2). Hope you enjoy the class!

Reema Rafifar she/her/hers

reemarafifar@berkeley.edu

Hi everyone! I'm Reema, a junior majoring in Neuroscience & Data Science. I absolutely love movies so come talk to me about your favorite films! I can't wait to get through C88C with you!

Thompson Zhou he/him/his

chuanjunzhou@berkeley.edu

Hi! I'm Thompson, a third year double majoring in Applied Math and Data Science. During my free time, I like to read, play video games, and follow sports.

In The News

AI-Driven Misinformation 'Biggest Short-Term Threat to Global Economy'

The World Economic Forum's annual risks report, based on a survey of 1,300 experts, revealed that respondents believe the biggest short-term threat to the global economy will come from AI-driven misinformation and disinformation. This is a major concern, given that elections will be held this year in countries accounting for 60% of global gross domestic product. Other short-term risks cited by respondents include extreme weather events, societal polarization, cyber insecurity, and interstate armed conflict.

The Guardian; Larry Elliott (January 10, 2024)

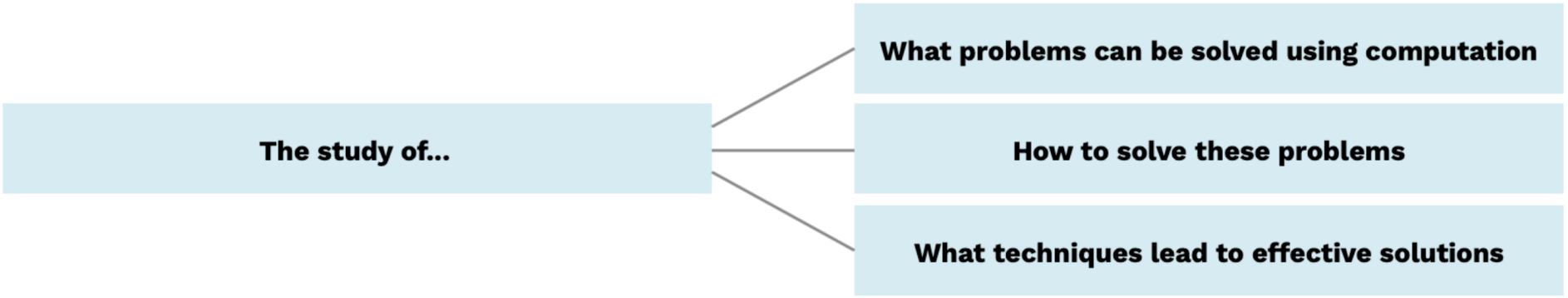
Computational Structures in Data Science

Computer Science & Data Science

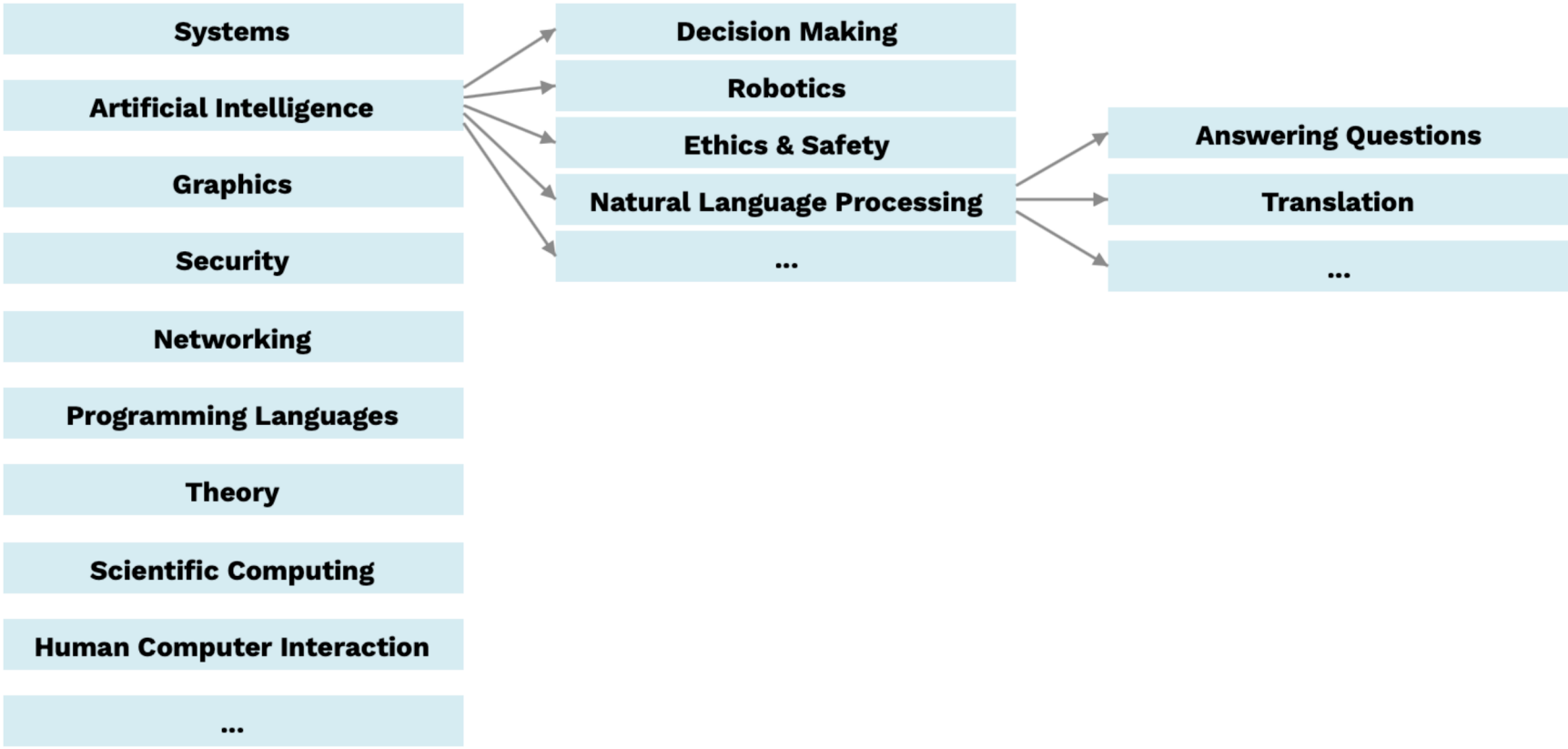
Berkeley
UNIVERSITY OF CALIFORNIA

Michael Ball | UC Berkeley |

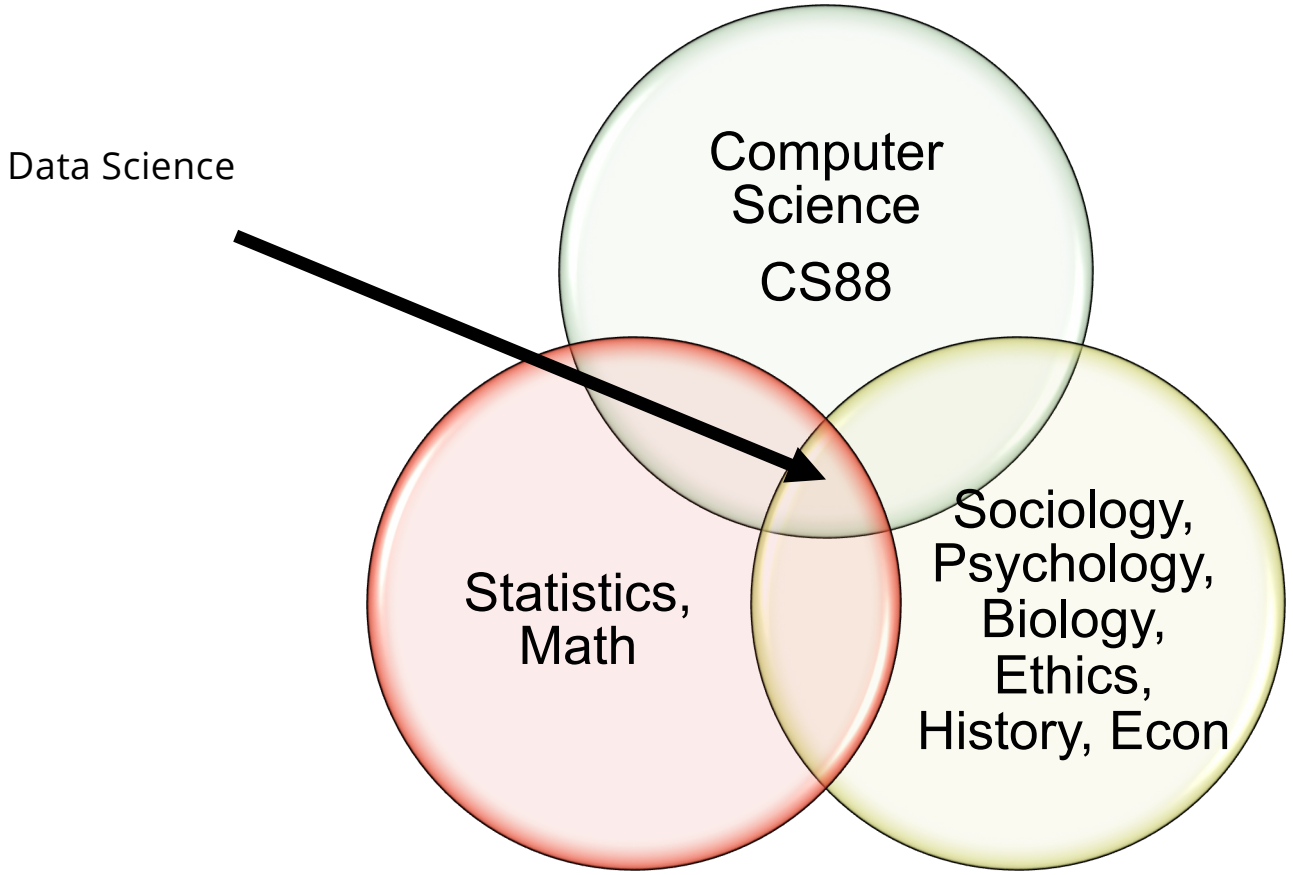
Computer Science



Computer Science, Some Ideas...Definitely Not Exhaustive!

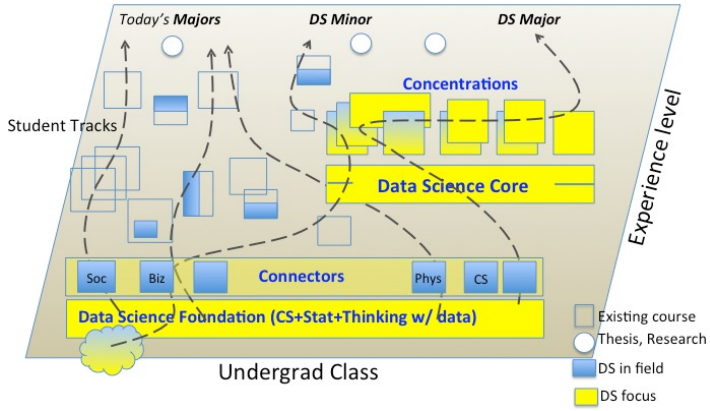


Computer Science & Data Science (One View)



Data 8 – Foundations of Data Science

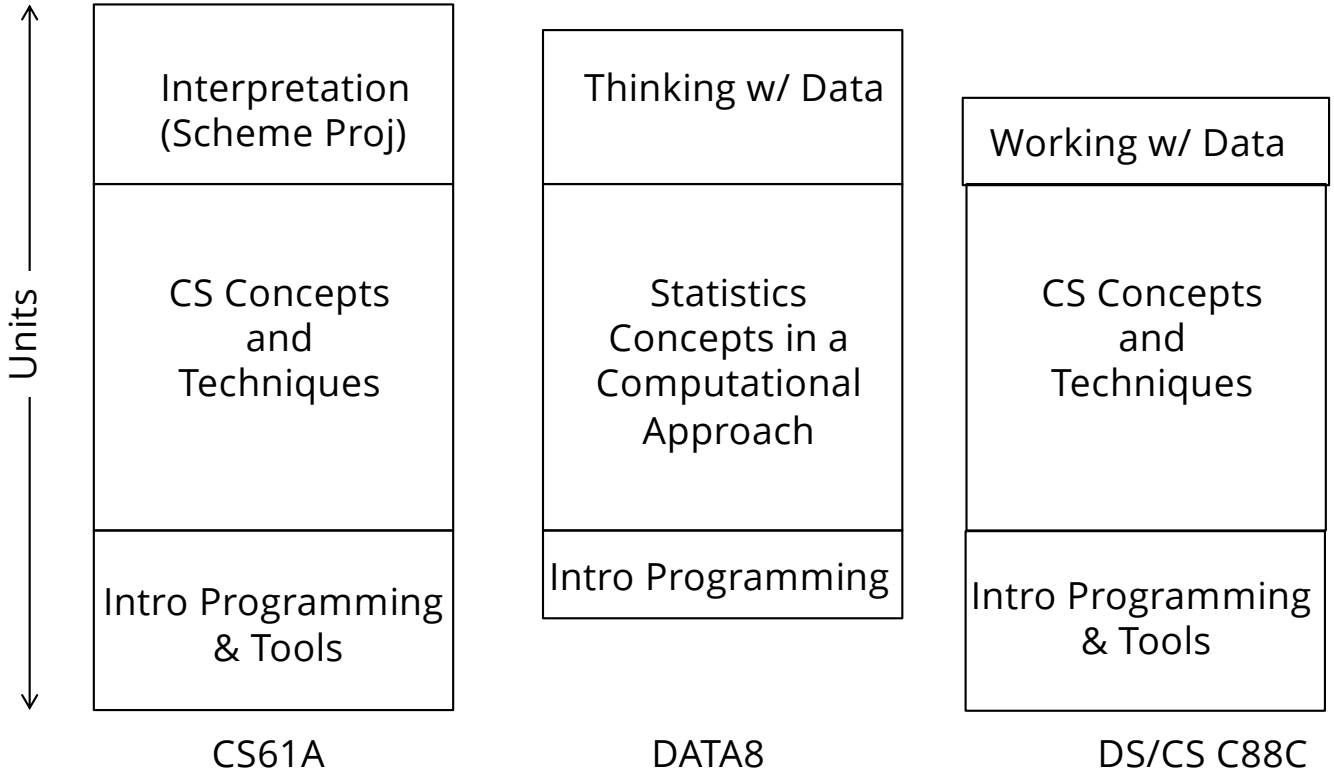
- Computational Thinking + Inferential Thinking in the context of working with real world data
- Introduce you to several computational concepts in a simple data-centered setting
  - Authoring computational documents
  - Tables
  - A LOT of statistics



CS88 – Computational Structures in Data Science

- Deeper understanding of the computing concepts introduced in DATA8
  - Hands-on experience => Foundational Concept
  - How would you create what you use in DATA 8?
- Extend your understanding of the structure of computation
  - What is involved in interpreting the code you write?
  - Deeper CS Concepts: Recursion, Objects, Classes, Higher-order Functions, Declarative programming, …
  - Managing complexity in creating larger software systems through composition
- Create complete (and fun) applications
- In a data-centric approach

How does C88C relate to CS61A?



Computational Structures in Data Science

Success In C88C

Berkeley
UNIVERSITY OF CALIFORNIA

Michael Ball | UC Berkeley |

Course Culture

- Learning
- Community
  - **Collaboration**
  - Peer Instruction
- Respect
- A supportive course staff & environment
  - Lots of outside community, CS Mentors, HKN, others.

Collaboration

- Asking questions, discussing topics, helping each other is always encouraged!
  - When you're working with a partner, you are expected to share in the work.
- Collaboration has limits
  - Please don't read someone else's code
    - except if you have already turned in the assignment, or a TA/staff member is present.
  - You can help others, but not give the solutions.
- We have a very particular set of skills and we will use them.

So... ChatGPT...

Or, why even bother learning anything anymore?

Why are you here?

What are your goals?

Guidelines for AI Use

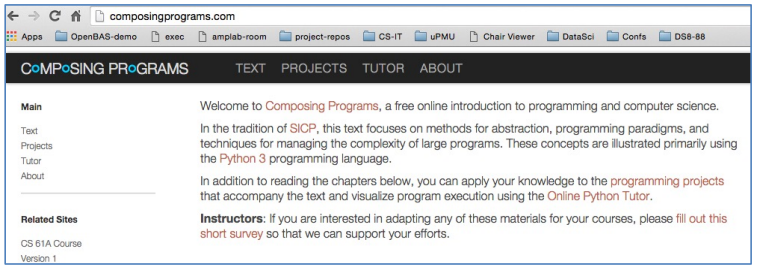
- You *may* use ChatGPT (and other tools) like a **tutor**
- **You may not ask about specific questions assigned to you!**
- Remember: It's a bot.
  - Bots are fallible!
  - Check its advice.
- Good Example:
  - "When would I prefer a for-loop over a while loop in Python?"
  - "Show me two examples of using a HOF to filter data in a list"
- **Bad Example:**
  - "Provide a solution to the Fibonacci sequence using recursion"

Do not lose the human connection in school.

(C88C, or otherwise.)

Course Structure

- 2 lectures, 1 lab each week
- Lecture introduces concepts (quickly!), answers why questions.
- Lab provides concrete detail hands-on
- Homework (11) cements your understanding
- Projects (2) put your understanding to work in building complete applications
  - Maps
  - Ants vs Some Bees



Class Format

- Mon and Weds Lectures:
  - Each lecture has a series of short self-check questions
  - Lectures go quickly
- Labs are paced throughout the week. See the Ed post to pick a time.
- Labs are HANDS ON – Get help as you're trying the lab.
  - Labs are active
  - Find and make friends!
  - Do not ask for more lectures
    - Ask *why* something doesn't make sense.

Class Format: Assignments

- Lecture Quizzes, 1 point, max 20.
  - 1 per lecture, “due” in ~4 days. (Partial credit after)
  - https://go.c88c.org/1
- Lab Work: 4 points, 11 labs, 1 drop
  - Start them during lab! You can probably finish some labs in 2 hours. Will be Python + some interactive questions.
- Homework: 8 points, 11 HW, 1 drop
  - Start early!
- Projects: 100 points between 2 projects
  - Start early! ”Checkpoint” assignments

Lab Attendance & Credit

- Labs are graded on correctness
- Most labs will have 2-5 questions.
  - Solving the questions shouldn't take 2 hours in most cases.
- **Graded attendance will an *option*.**
- There are no "make ups" or excused absences, but if you commit to 10 sections, it will reduce the weight of exams by a bit.

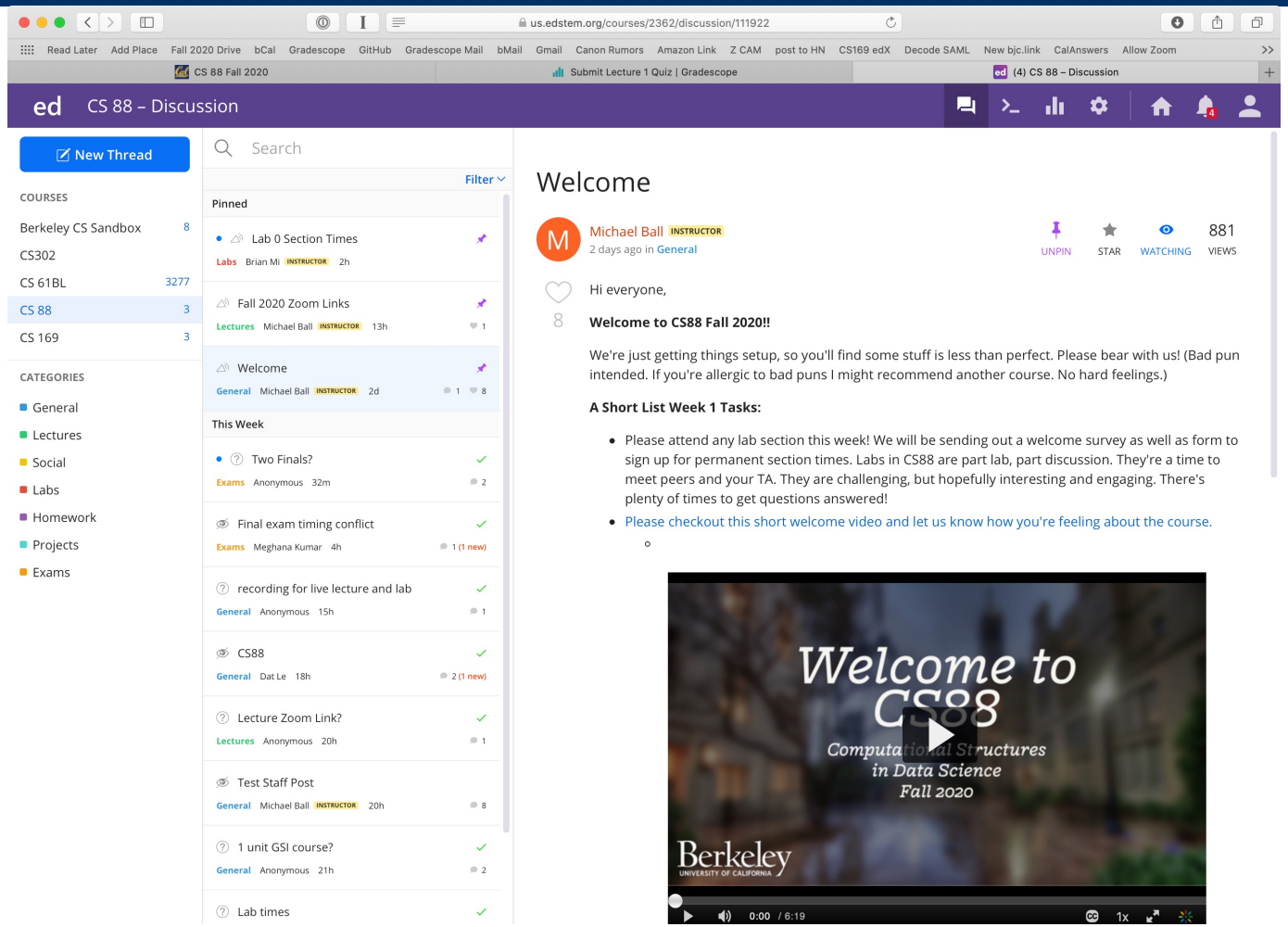
Extensions & Extenuating Circumstances

- https://go.c88c.org/extensions **(up soon)**
- **Contact us early!!**
- Our goal is to have you do the work, but we can't manage things at 11:50PM
- If you need <= 3 days, **do not** submit the form.
- If you know you will be travelling, etc. Let us know ASAP.

Class Format: Exams

- 1 midterm and 1 final exam, **in person**
- **No remote exam options offered.**
  - Required verification for alternate exams.
- Midterm 2 hours, October 15
- Exam will be during the slot assigned by campus.
- 5 handwritten cheat sheets double-sided.
- **You don’t actually need 5 sheets!**

Ed For Class Discussion: Try it!



Let's Stop Harassment & Discrimination

- Disparaging remarks targeting a particular gender, race, or ethnicity are not acceptable.
- From the Berkeley Principles of Community:
  - "We affirm the dignity of all individuals and strive to uphold a just community in which discrimination and hate are not tolerated."
- From the EECS department mission:
  - "Diversity, equity, and inclusion are core values in the Department of Electrical Engineering and Computer Sciences. Our excellence can only be fully realized by faculty, students, and staff who share our commitment to these values."

- **EECS Student Climate & Incident Reporting Form**: Informs the EECS department of any issues. You can also contact Antoine Davis (CS Student Affairs Director) directly.

Who's in the class?

- Hum your answer.

Computational Structures in Data Science

Python & Expressions

Berkeley
UNIVERSITY OF CALIFORNIA

Michael Ball | UC Berkeley |

Types of expressions

An expression describes a computation and evaluates to a value

$$\frac{6}{23} \qquad \sin \pi \qquad \log_2 1024$$

$$2^{100} \qquad f(x) \qquad \sqrt{3493161}$$

$$7 \bmod 2 \qquad \sum_{i=1}^{100} i \qquad \lim_{x \to \infty} \frac{1}{x}$$

$$|-1869| \qquad \binom{69}{18}$$

Call Expressions in Python

(Demo)

Terminals and Interpreters

- **Terminals:** A “shell” — a program that lets you interact with your computer by typing commands
- Getting *good* at this is useful **but not expected!**
  - You will need to build some basic command line skills to do assignments.
- **Python Interpreter:**
- A program which *evaluates* Python code–as you write it–and immediately shows you the results.
- SUPER useful! Use it liberally.
- Tip: Up arrow key to recall previous statements.
- **python3 -** this is often the exact name of the python interpreter program. It will also run other problems for us.

Take Things 1 Step at a Time

- We interact with Python via the *Terminal*
- We type *programs* into files and into other programs.
- Everything you do in this class is safe!
- **Try and experiment!**

Your Tasks

- Lecture 1 "self-check" on Gradescope
- Attend Lab 0
- Attend OH

Welcome, and Good luck!