Computational Structures in Data Science

Lecture 1:
Welcome to Data C88C!

(Summer 2024)

Berkeley
UNIVERSITY OF CALIFORNIA

Eric Kim | UC Berkeley | © CC BY-NC-SA

Revision 01 (2024-06-17 7:28 pm pst)

Lecture 1 Overview

- Course overview
- Online course setup
- Teaching staff introduction
- Course policies, administrivia

Announcements

- Office hours do not start until Thursday (June 20th).
- Holiday: this Wednesday (June 19th, Juneteenth) is an academic holiday, so no lecture/OHs on Wednesday!
- Important Ed posts
  - Week 1 announcements
  - Course index (will be updated throughout the course)
  - Please, please read these!

(tldr reference) Important Links

- **Course website** (where to find assignments): https://c88c.org/su24
  - Tip: The names “CS 88C”, “Data C88C” are basically synonyms for each other
- **Gradescope** (where to submit assignments): https://www.gradescope.com/courses/786589
- **Ed** (where to ask questions, announcements): https://edstem.org/us/courses/59252/discussion/
- **bCourses** (where to find Zoom links and recorded lectures/labs): https://bcourses.berkeley.edu/courses/1534781
- **Tip**: for the above links, use your @berkeley.edu email address (“CalNet ID”) to log in (or via “CalNet” SSO)

Online course setup

- Data C88C su24 is taught **“fully online”**
- What does this mean for you (the student)?
  - Lectures, labs, office hours, and exams done through **Zoom**
  - **You do not need to physically be present in Berkeley, CA for this course**
    - Note: this includes teaching staff
- Where to find the Zoom links?
  - bCourses+Zoom page has ALL course Zoom links

Online course setup

- Where to find the Zoom links?
  - bCourses+Zoom page has ALL course Zoom links
  - Lectures, Labs, OH
- Where to find/submit assignments?
  - Course website will have all assignments
  - Submit ALL assignments (lecture quizzes, labs, hws, projects) to Gradescope
  - Tip: Gradescope has assignment due dates as well
- Where to ask questions?
  - Ed is an online course forum.

Online course setup

- Important: our **primary means of communication to you** (the students) will be through:
  - Ed: https://edstem.org/us/courses/59252/discussion/
  - Course website: https://c88c.org/su24/
- It's **your responsibility** to follow the above pages to keep up to date with things like: assignment releases/deadlines, exam details, etc.
  - I will be very sad if you miss an announcement and, say, miss a project deadline

Online course setup

- Asking questions during lecture
  - Our original intent is for you (the students) to ask questions on the Ed page specific for each Lecture
  - Example: Lecture01 Ed page
- However: if it's more organic to ask it via the Zoom text chat, feel free to do that as well.
  - Note: Zoom text chats will be SAVED after each lecture (along with author names). So, please keep the text chat respectful and professional (eg don't write anything you wouldn't want your parents to see)

“Berkeley time”

- I’ll try to respect “Berkeley time”, aka lectures/labs start 10 minutes AFTER the posted start time.
  - Example: Lecture posted start time is 3pm, but I’ll actually **start the lecture at 3:10pm**.
- I’ll still start the Zoom lecture at 3pm. Feel free to join early and treat it as an informal office hours, eg asking questions, chatting about random things, etc.
- Note: **exam times** will start promptly at the posted start time.

(Quick Poll): Lecture Zoom registration

- For this lecture Zoom link, I enabled “Registration required”
- Poll question: is registering for lectures too much trouble?
  - If it is, I’ll consider disabling the “Register” feature and just have a simple Zoom meeting link for lectures

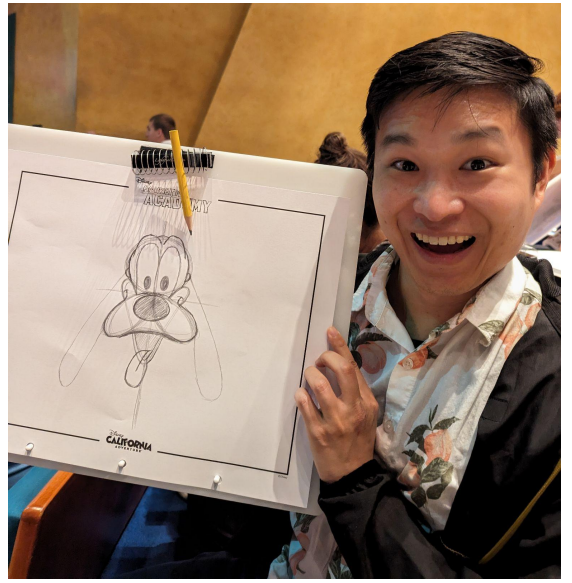
Goals today

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

- Big Ideas:
  - Abstraction
  - Data Type

Data C88C Team – Eric

- Eric Kim (Instructor)
  - ekim555@Berkeley.edu – You're best off by using Ed! ☺️
  - Teaching from Southern California (Fullerton!)
  - https://eric-kim.net
  - Office hours: Mondays 4pm-5pm PST. "coffee chat"
- About me
  - I work as an applied machine learning engineer, working on things like: model training/serving (deep learning, computer vision, representation learning), retrieval/recommendation systems, and big-data pipelines.
  - Graduated UC Berkeley (2011, L&S CS), UCLA (2016, Master's in CS with focus in AI/ML and computer vision)
  - For fun, I enjoy playing music (piano, bass, guitar, vocals. Pop-punk/rock/classical/blues/jazz/...), playing games, and other random hobbies (like Pickleball!)



Data C88C Team – Head TA : cs88@berkeley.edu

Head Teaching Assistant

Lab 103L

John Teng **[he/him]**

Office Hours: Mon 5pm - 6pm, Wed 4pm - 5pm

johnteng9@berkeley.edu

Hi, I'm John, a third year CS major from Pennsylvania. I like playing video games, soccer, and working out. Looking forward to this semester!

Tip: You can also find this info on the course Staff Page

Data C88C Team -- TAs

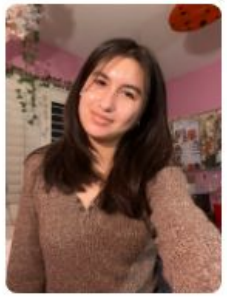
Lab 102L

Satleen Gill [she/her]

Office Hours: Mon 11am - 1pm

satleen@berkeley.edu

Hello! My name is Satleen. I'm a third-year majoring in Data Science and Computer Science from Sonoma County, CA. I love listening to music, reading, day trips, and being with friends. I'm excited for an amazing summer! <3



Labs 101L, 104L

Mia Lopez

Office Hours: Tues 11am - 12pm, Thurs 11am - 12pm

mglopez@berkeley.edu

❀ hello! My name's Mia and I'm a 3rd year studying CS! Some of my favorite artists are Childish Gambino, Chappell Roan, and Sabrina Carpenter. My favorite movies are Arrival and Lego Batman, but I also have a thing for Wes Anderson movies! This is my 3rd semester teaching, so feel free to reach out whenever! ❀

Naveen Nathan [he/him]

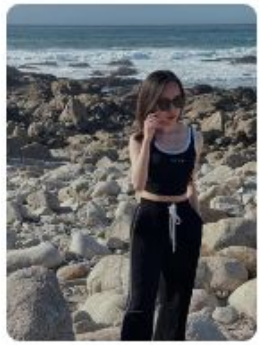
Office Hours:
Thurs 4:30pm - 5:30pm

naveen.nathan@berkeley.edu

Hi everyone! I am Naveen, a rising 4th year CS major. I love to hike and bike in my free time. I look forward to working with you all this summer!

Tip: You can also find this info on the course Staff Page

Data C88C Team -- Tutors



Jiayi Li [she/her]

Office Hours: Tues 5pm - 6pm, Wed 5pm - 6pm

jiayi.li@berkeley.edu

Hi! I'm Jiayi, a senior studying Statistics and Data Science. I love photography, traveling, and road trips! Looking forward to meeting you all and enjoying this summer!

Mira Wagner

Office Hours: Wed 11am - 1pm

mirawagner@berkeley.edu

Hi! I am a sophomore planning to major in data science/statistics and linguistics. I love reading, especially mysteries, swimming and baking! Excited for this semester :)

Tip: You can also find this info on the course Staff Page

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.

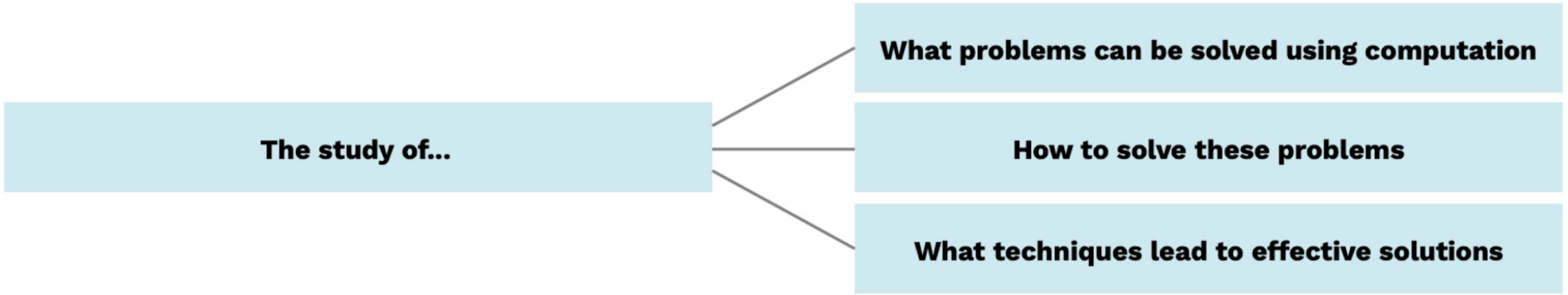
The Guardian; Larry Elliott (January 10, 2024)

Computational Structures in Data Science

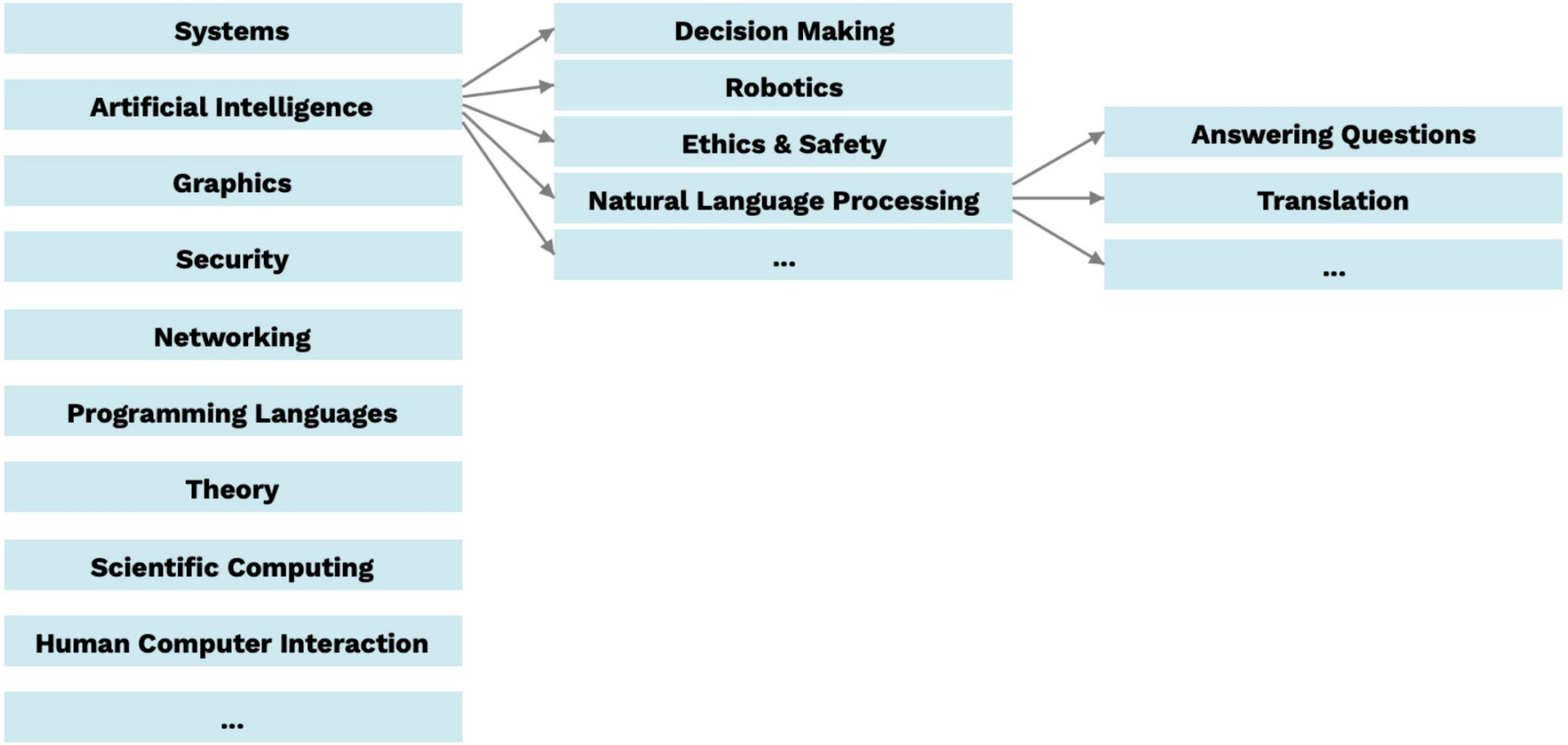
Computer Science and Data Science

Berkeley
UNIVERSITY OF CALIFORNIA

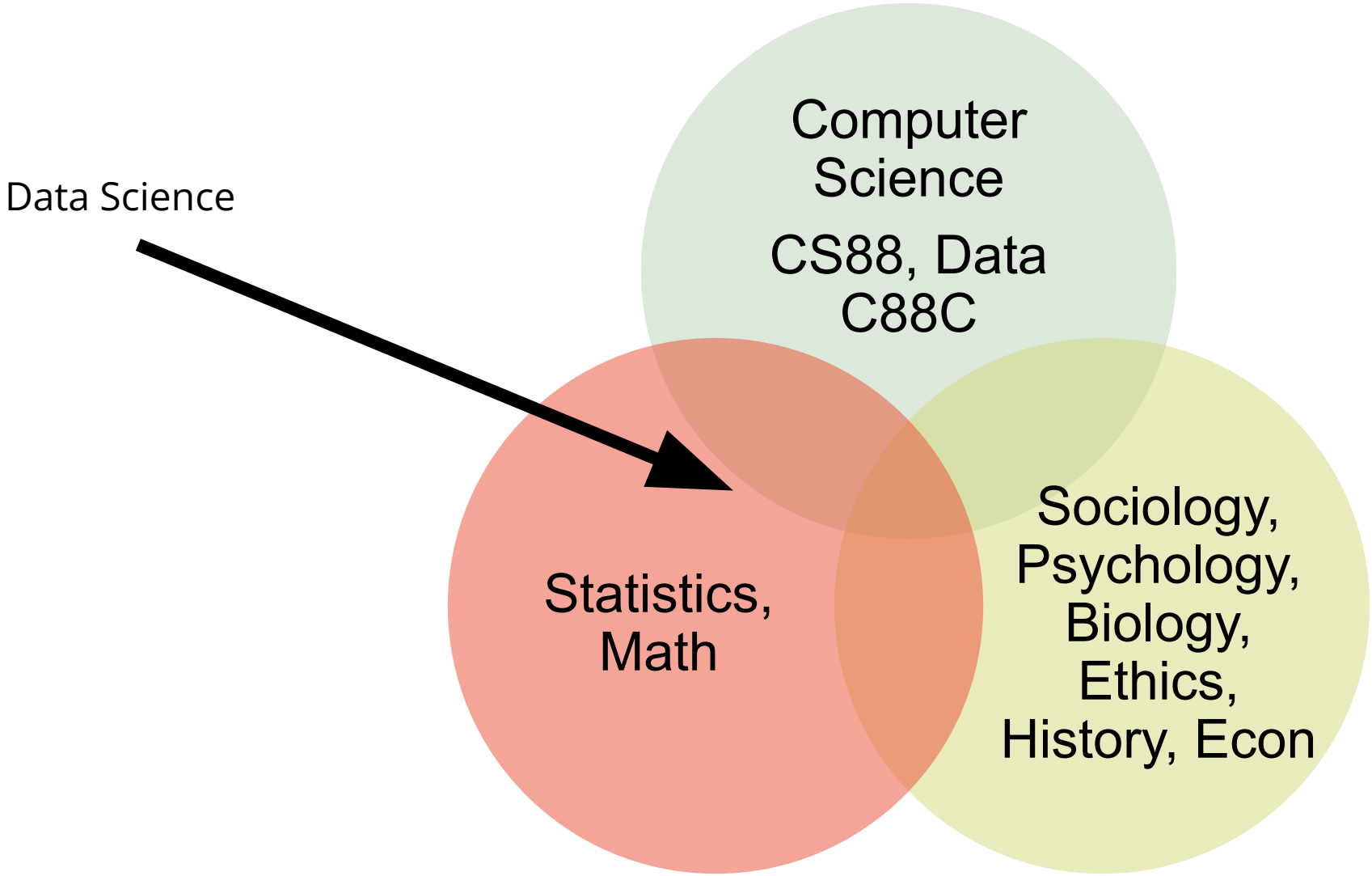
Computer Science



Computer Science, Some Ideas…Definitely Not Exhaustive!

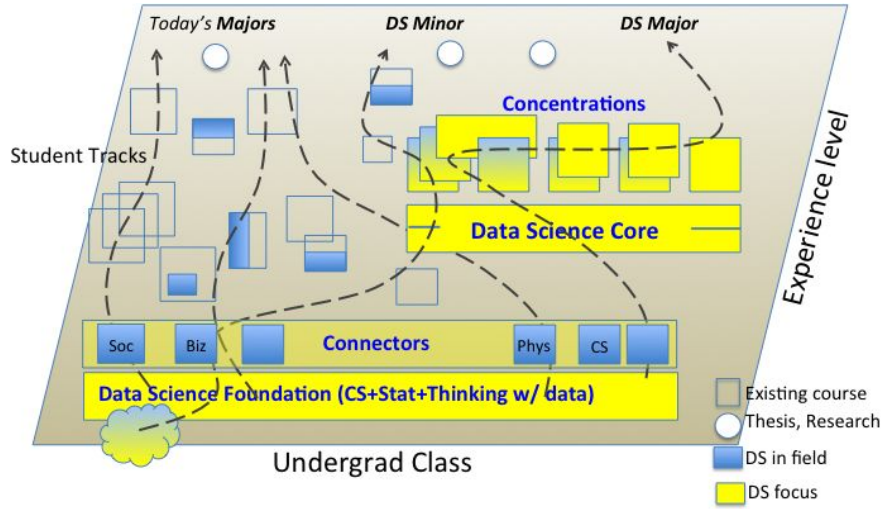


Computer Science & Data Science (One View)



Data 8 – Foundations of Data Science

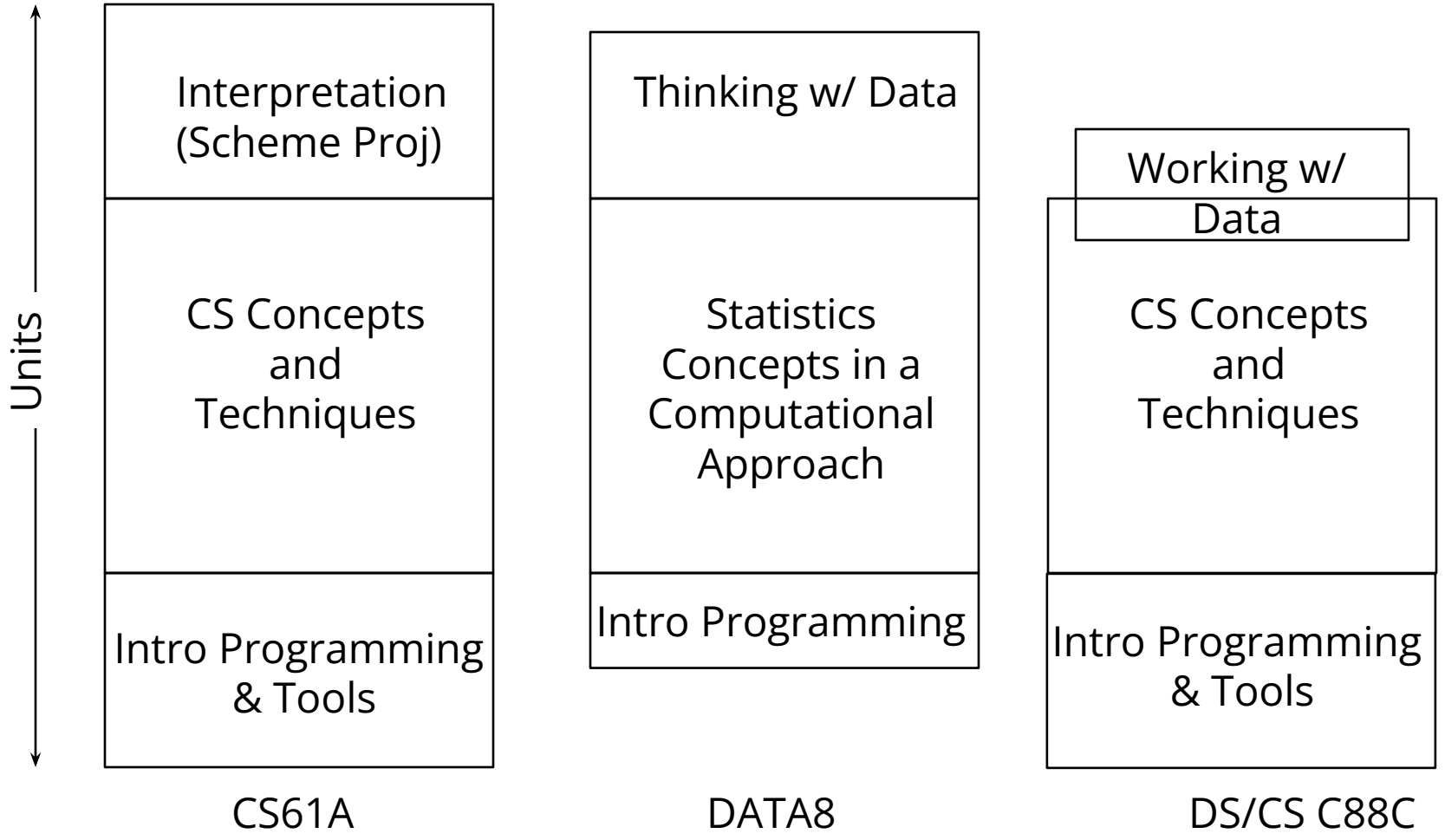
- Computational Thinking + Inferential Thinking in the context of working with real world data
- Introduce you to several computational concepts in a simple data-centered setting
  - Authoring computational documents
  - Tables
  - A LOT of statistics



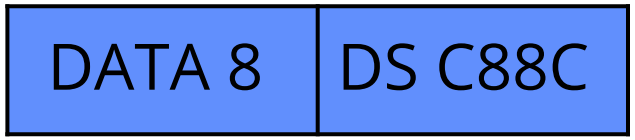
Data C88C – Computational Structures in Data Science

- 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?

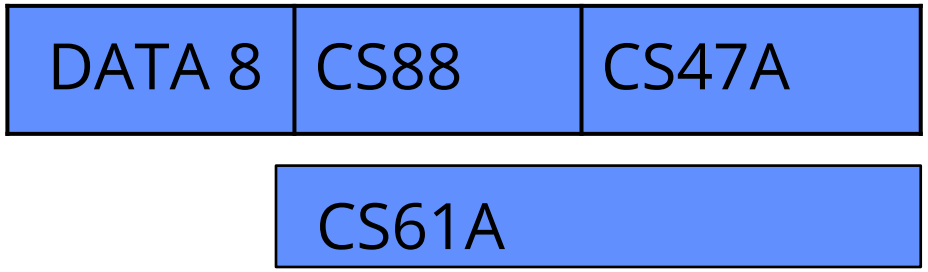


Opportunities for students



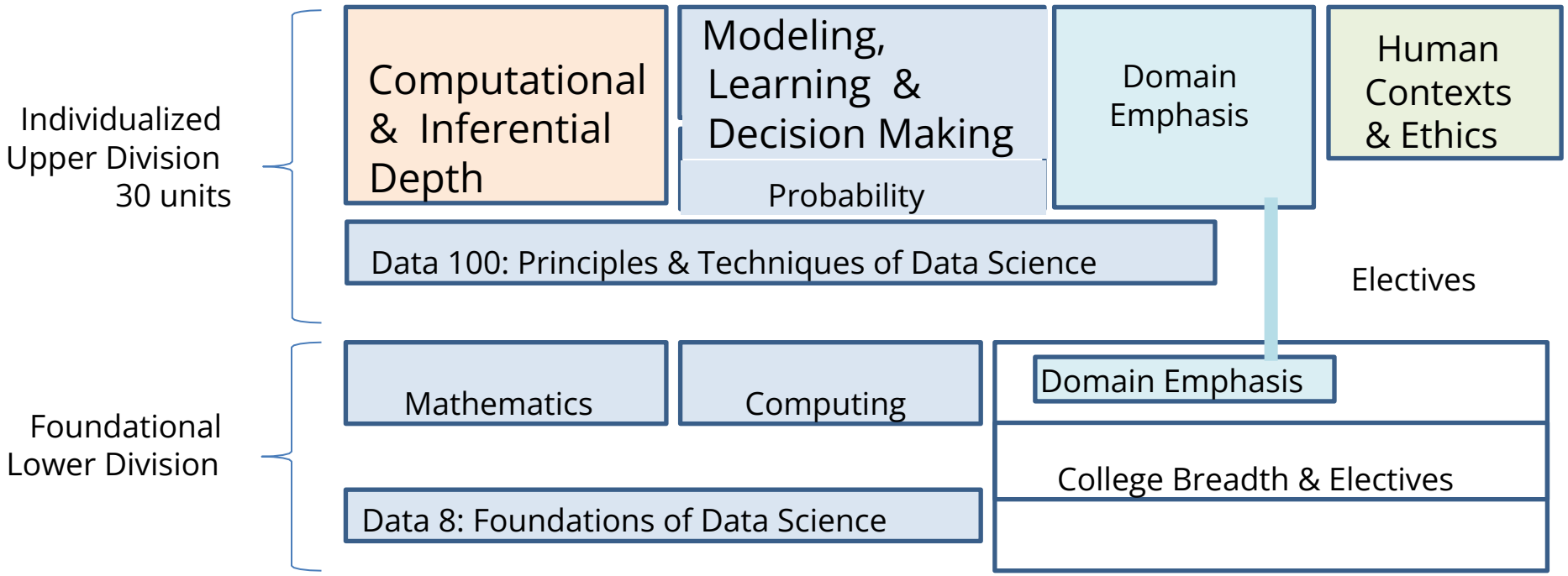
DS Minor

CS minor & DS Major



CS major:
CS47A path is not recommended, but there if you change your mind.

The Data Science Major



Computational Structures in Data Science

Success In C88C

Berkeley
UNIVERSITY OF CALIFORNIA

Course Culture

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

Collaboration

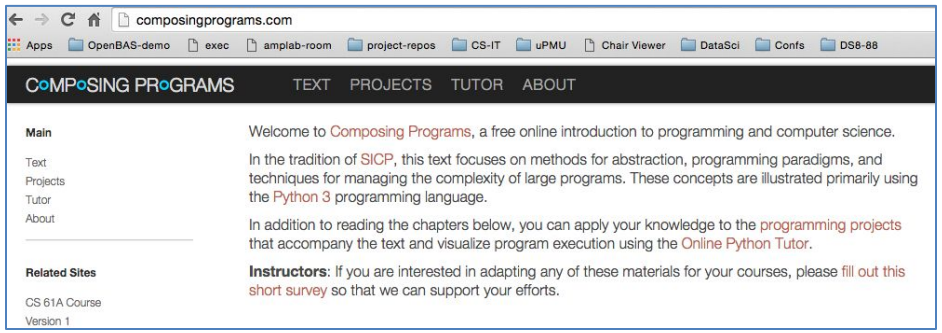
- Asking questions, discussing topics, helping each other is always encouraged!
  - When you're working with a partner, you are expected to share in the work.
- Collaboration has limits
  - Please don't read someone else's code
    - except if you have already turned in the assignment, or a TA/staff 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…We Know About ChatGPT

- 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?"
  - "What is a higher order function?"
- **Bad Example:**
  - "Provide a solution to the Fibonacci sequence using recursion"

Course Structure

- 4 lectures, 2 labs 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

- 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 ~48 hours after lecture. (Partial credit after)
  - Lecture 01 Quiz Link (Gradescope)
- 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

Labs

- Labs are graded on correctness (out of 4 points)
- **There are no "make ups" or excused absences.**

Extensions & Extenuating Circumstances

- If something **truly exception** comes up and you need to reschedule an assignment date, please e-mail us at: cs88@berkeley.edu or create a private Ed post
- Example: If you know you will be travelling, etc. Let us know ASAP.
-

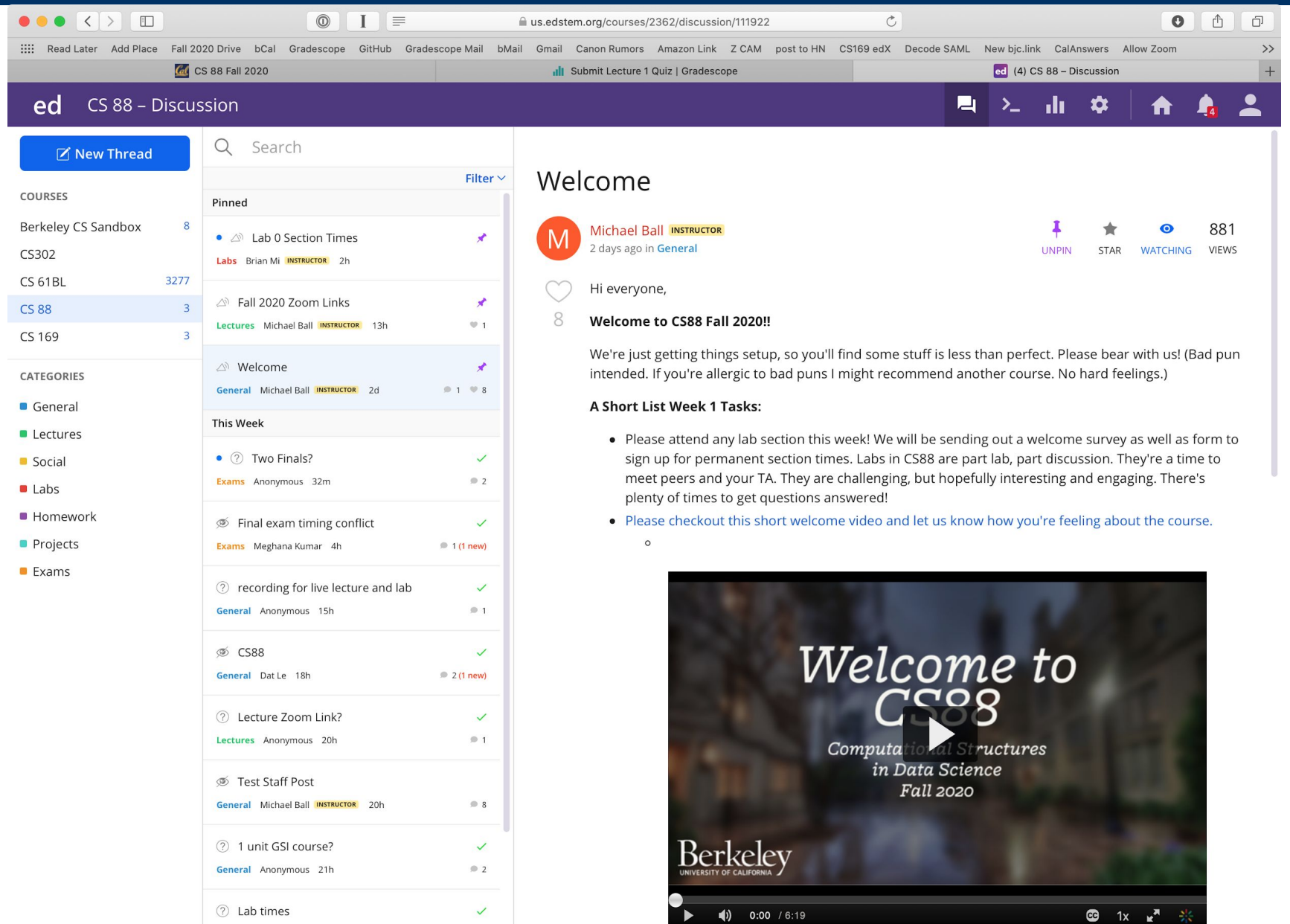
Class Format: Assignments

- Slip Days: 14 total
  - Use up to 2 on any assignment
  - We (auto) apply them in the order that's most beneficial to you!
    - i.e. use them on projects if you need!
  - Can be used for homework, labs, projects, but not project checkpoints.
- Slip Days take care of almost all special circumstances!
- What if you go over slip days?
  - 25% deduction for each day over. Mathematically you can still earn 25% if you turn in something 3 days late.

Class Format: Exams

- 1 midterm and 1 final exam
  - Remote exams proctored via Zoom
  - We will support alternate exam times
  - Required verification (aka student ID)
- **Exact time + details for midterm + final will be announced within 1-2 weeks**
- 5 handwritten cheat sheets double-sided.
- Pro tip: **You don’t actually need 5 sheets!**

Ed For Class Discussion: Try it!



Where will we work?

- Your laptop (using an editor and a terminal)
- Course website: https://c88c.org/su24/
- Gradescope – HW, Lab, Lecture Self-Checks
  - This is where you submit ALL assignments.
  - Note: has a built-in autograder, quick grade feedback
- Ed Discussion: edstem.org
  - Recommended way to ask questions.
  - Course staff (and students!) will help answer questions.
- bCourses page. We won't really use this, but lecture/lab recordings are uploaded here.

Computational Structures in Data Science

Demo?

Berkeley
UNIVERSITY OF CALIFORNIA

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
  - Due date: June 19th, 11:59 PM PST
- Attend Lab 0
  - Due date: June 22th, 11:59 PM PST
- Attend Office Hours (OH)
  - **Important: OH do not start until Thursday 6/20**

Welcome, and Good luck!

Questions?