UC Berkeley EECS
Lecturer
Michael Ball

Computational Structures in Data Science

Wrap Up

Announcements

- Final Exam, Tuesday 7-10pm slot
  - Alternate 8am Weds
  - We will email those with alternates/accommodations soon.
- One last HW, and final time to use slip days on Ants.
  - Also, make sure to address composition feedback from the tutors.
- The grading info is all based on points you earned
  - Sum everything up! (Everything except lecture self-checks are in howamidoing.)
    - » Basically everyone is getting 20/20 on the lecture self-checks. That's the point. ☺
    - » There were 27 self-checks which means you could have skipped 7 lectures
- http://howamidoing.cs88.not.cs61a.org/

UC Berkeley EECS
Lecturer
Michael Ball

Computational Structures in Data Science

SQL: Joins

Joining tables

- Two tables are joined by a comma to yield all combinations of a row from each
  - `select * from sales, cones;`

```
create table sales as
    select "Baskin" as Cashier, 1 as TID union
    select "Baskin", 3 union
    select "Baskin", 4 union
    select "Robin", 2 union
    select "Robin", 5 union
    select "Robin", 6;
```

Cashier	TID
Baskin	1
Robin	2
Baskin	3
Baskin	4
Robin	5
Robin	6

```
sales.join('TID', cones, 'ID')
```

TID	Cashier	Flavor	Color	Price
1	Baskin	strawberry	pink	3.55
2	Robin	chocolate	light brown	4.75
3	Baskin	chocolate	dark brown	5.25
4	Baskin	strawberry	pink	5.25
5	Robin	bubblegum	pink	4.75
6	Robin	chocolate	dark brown	5.25

```
sqlite> select * from sales, cones;
Baskin|1|1|strawberry|pink|3.55
Baskin|1|2|chocolate|light brown|4.75
Baskin|1|3|chocolate|dark brown|5.25
Baskin|1|4|strawberry|pink|5.25
Baskin|1|5|bubblegum|pink|4.75
Baskin|1|6|chocolate|dark brown|5.25
Baskin|3|1|strawberry|pink|3.55
Baskin|3|2|chocolate|light brown|4.75
Baskin|3|3|chocolate|dark brown|5.25
Baskin|3|4|strawberry|pink|5.25
Baskin|3|5|bubblegum|pink|4.75
Baskin|3|6|chocolate|dark brown|5.25
Baskin|4|1|strawberry|pink|3.55
Baskin|4|2|chocolate|light brown|4.75
Baskin|4|3|chocolate|dark brown|5.25
Baskin|4|4|strawberry|pink|5.25
Baskin|4|5|bubblegum|pink|4.75
Baskin|4|6|chocolate|dark brown|5.25
Robin|2|1|strawberry|pink|3.55
Robin|2|2|chocolate|light brown|4.75
Robin|2|3|chocolate|dark brown|5.25
Robin|2|4|strawberry|pink|5.25
Robin|2|5|bubblegum|pink|4.75
Robin|2|6|chocolate|dark brown|5.25
Robin|5|1|strawberry|pink|3.55
Robin|5|2|chocolate|light brown|4.75
Robin|5|3|chocolate|dark brown|5.25
Robin|5|4|strawberry|pink|5.25
Robin|5|5|bubblegum|pink|4.75
Robin|5|6|chocolate|dark brown|5.25
Robin|6|1|strawberry|pink|3.55
Robin|6|2|chocolate|light brown|4.75
Robin|6|3|chocolate|dark brown|5.25
Robin|6|4|strawberry|pink|5.25
Robin|6|5|bubblegum|pink|4.75
Robin|6|6|chocolate|dark brown|5.25
```

Inner Join

```
select * from sales, cones where TID=ID;
```

```
sales.join('TID', cones, 'ID')
```

TID	Cashier	Flavor	Color	Price
1	Baskin	strawberry	pink	3.55
2	Robin	chocolate	light brown	4.75
3	Baskin	chocolate	dark brown	5.25
4	Baskin	strawberry	pink	5.25
5	Robin	bubblegum	pink	4.75
6	Robin	chocolate	dark brown	5.25

```
sqlite> select * from sales, cones where TID=ID;
Baskin|1|1|strawberry|pink|3.55
Baskin|3|3|chocolate|dark brown|5.25
Baskin|4|4|strawberry|pink|5.25
Robin|2|2|chocolate|light brown|4.75
Robin|5|5|bubblegum|pink|4.75
Robin|6|6|chocolate|dark brown|5.25
sqlite>
```

SQL: using named tables - `from`

```
select "delicious" as Taste, Flavor, Color from cones
       where Flavor is "chocolate" union
select "other", Flavor, Color from cones
       where Flavor is not "chocolate";
```

```
sqlite> select "delicious" as Taste, Flavor, Color from cones where Flavor is "chocolate" union
   ...> select "other", Flavor, Color from cones where Flavor is not "chocolate";
Taste|Flavor|Color
delicious|chocolate|dark brown
delicious|chocolate|light brown
other|bubblegum|pink
other|strawberry|pink
sqlite>
```

Queries within queries

- Any place that a table is named within a select statement, a table could be computed
  - As a sub-query

```
select TID from sales where Cashier is "Baskin";

select * from cones
    where ID in (select TID from sales where Cashier is "Baskin");

sqlite> select * from cones
   ...>     where ID in (select TID from sales where Cashier is "Baskin");
ID|Flavor|Color|Price
1|strawberry|pink|3.55
3|chocolate|dark brown|5.25
4|strawberry|pink|5.25
```

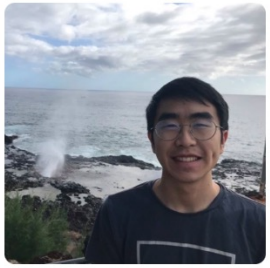
UC Berkeley EECS
Lecturer
Michael Ball

Computational Structures in Data Science

Review and Wrap Up

CS88 Team

Head Teaching Assistants



Nicholas Ng [he/him]

Office Hours: Tuesday 10am - 11am

nickng589@berkeley.edu

Hi! I'm Nicholas, and I'm a fourth year Computer Science Major from the San Gabriel Valley. I like playing video games (I probably spent too much of my summer playing Valorant), and every once in a while I'll go outside for a run. I look forward to meeting everyone!

Shreya Kannan [she/her]

Office Hours: Tuesday 10am - 11am

shreyakannan@berkeley.edu

Hi everyone, I'm super thrilled to meet y'all this semester! In no particular order, I prefer musicals over concerts, cold over warm weather, dancing over singing, and sushi over tacos. Feel free to talk to me about great restaurants, cool study spots, and of course, CS 88 :)

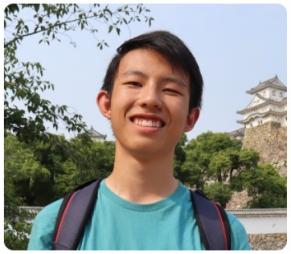
Teaching Assistants

Chi Tsai [she/her]

Office Hours: Monday 7pm - 8pm

chitsai@berkeley.edu

Hi! I'm Chi, and I'm from the LA area. I enjoy cooking and trying new food, playing minecraft, planning upcoming vacations, and watching Brooklyn 99 (Terry loves yogurt). Come talk to me about anything, happy to meet you!



Lukas Chang [he/him]

Office Hours: Wednesday 12pm - 1pm

lukasc@berkeley.edu

Hi everyone I'm Lukas, a 3rd year CS major from the bay area. This is my second semester TAing for CS88 and I'm excited to meet you all! A little about me—in my free time I love making music though I'm not very good playing piano, and I also like going on nightly walks. I hope I can share my love for CS with all of you!

Matt Au [he/him]

Office Hours: Tuesday 1pm - 2pm

mattau@berkeley.edu

Hi! I'm Matt, a 3rd-year CS major from Southern California. I think water engineering is cool. I've been really into the album brent ii by Jeremy Zucker and Chelsea Cutler lately. I also recently discovered Strawberry Milk Tea :D. Let me know if you have any song/boba recommendations! I look forward to meeting you all!

Tommy Joseph [he/him]

Office Hours: Monday 6pm - 7pm

tommy11jo@berkeley.edu

Hi I'm a third-year CS major from Southern California, and this is my second time TAing for CS 88. Outside of CS, I spend my time attempting recipes I find on youtube, watching tv, and learning piano (I'm a beginner, sorry neighbors). Happy to help you with anything I can!

CS88 Team

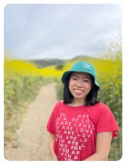
Tutors

Anjali Gurajapu [she/her]

Office Hours: Friday 10am - 11am, Friday 6pm - 7pm

agura@berkeley.edu

Hi! I'm Anjali, and I'm a third year studying Chemistry and Data Science. In my free time, I've been drawing, baking, and trying to befriend the stray cat that sleeps on my porch. I'm excited to be a tutor this semester, and looking forward to meeting you all! :)



Jessica Lin [she/her]

Office Hours: Monday 11am - 1pm

jessical02@berkeley.edu

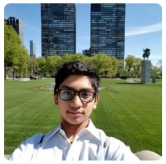
Hi friends! I'm Jessica, a second year majoring in CS from Southern California. I enjoy dancing, reading, doing crosswords, and making random peace signs. This is my first semester as a tutor for this course, but my second semester teaching the material. Feel free to reach out to me for anything (:

Kevin Gu [he/him]

Office Hours: Tuesday 2pm - 3pm

kevinjgu@berkeley.edu

Hi everyone! I'm Kevin, a recent grad and now a student in the UC Berkeley School of Information, and doing the 5th Year Master of Information and Data Science (MIDS) program. I majored in Data Science and minored in CS in my undergrad, and I'm really excited to be working with y'all this semester!! Outside of academics, I really love the humanities; I like learning languages and am also really into classical music, so feel free to hit me up if you ever want to have conversations about any of those things. Happy coding!



Rahul Shah [he/him]

Office Hours: Tuesday 7pm - 9pm

rsha256@berkeley.edu

Hiya!! I'm Rahul, a sophomore studying CS and Data Science. In the past, I've been a tutor for CSM 88 so I'm looking forward to bringing y'all awesome content! I like hiking, reading, and trying new food. Feel free to talk to me about anything at anytime :))

Simon Tsui [he/him]

Office Hours: Friday 11am - 1pm

stsui2@berkeley.edu

Hello! I am a junior studying Computer Science and Statistics. Feel free to reach out to me if you want to talk about classes at Berkeley, food recs, or anything else. Looking forward to a great CS 88 experience this semester!

THANK YOU!

The One Big Thing...

Abstraction

- Detail removal

  "The act of leaving out of consideration one or more properties of a complex object so as to attend to others."

- Generalization

  "The process of formulating general concepts by abstracting common properties of instances"

- Technical terms: Compression, Quantization, Clustering, Unsupervised Learning

Henri Matisse "Naked Blue IV"

The Power of Abstraction, Everywhere!

- **Examples:**
  - **Functions (e.g., sin x)**
  - **Hiring contractors**
  - **Application Programming Interfaces (APIs)**
  - **Technology (e.g., cars)**
- **Amazing things are built when these layer**
  - **And the abstraction layers are getting deeper by the day!**

We only need to worry about the interface, or specification, or contract NOT how (or by whom) it's built

Above the abstraction line

Abstraction Barrier (Interface)
(the interface, or specification, or contract)

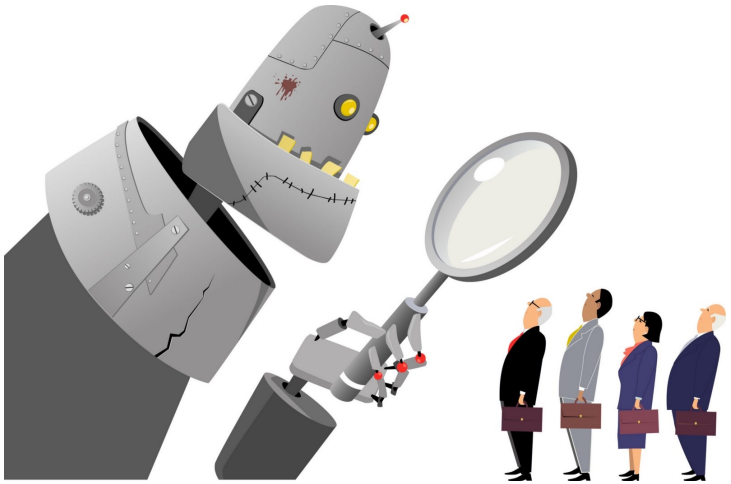
Below the abstraction line

This is where / how / when / by whom it is actually built, which is done according to the interface, specification, or contract.

Abstraction: Pitfalls

- **Abstraction is not universal without loss of information (mathematically provable). This means, in the end, the complexity can only be "moved around"**

- **Abstraction makes us forget how things actually work and can therefore hide bias. Example: AI and hiring decisions.**



- **Abstraction makes things special and that creates dependencies. Dependencies grow longer and longer over time and can become unmanageable.**

Is Your Brain Full Yet?

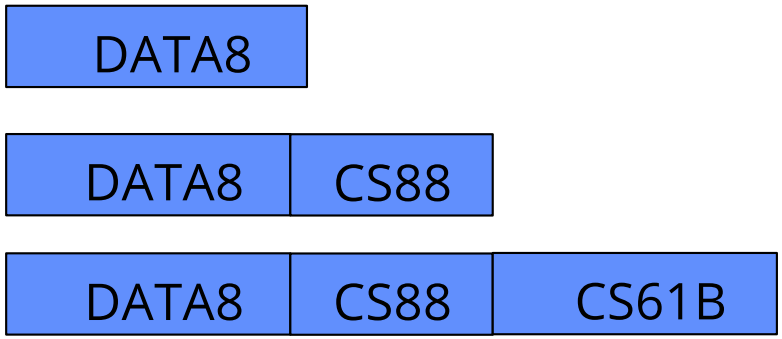
Python Tools:

- Data type: values, literals, operations,
- Expressions, Call expression
- Variables
- Assignment Statement, Tuple assignment
- Sequences: tuple, list
- Dictionaries
- Function Definition Statement
- Conditional Statement
- Iteration: list comp, for, while
- Lambda function expr.

- Higher Order Functions
  - Functions as Values
  - Functions with functions as argument
  - Assignment of function values
  - Function factories – create and return functions
- Higher order function patterns
  - Map, Filter, Reduce
- Recursion
- Abstract Data Types
- Mutation
- Class & Inheritance
- Exceptions
- Iterators & Generators
- SQL / Declarative Programming

Keep on Programming

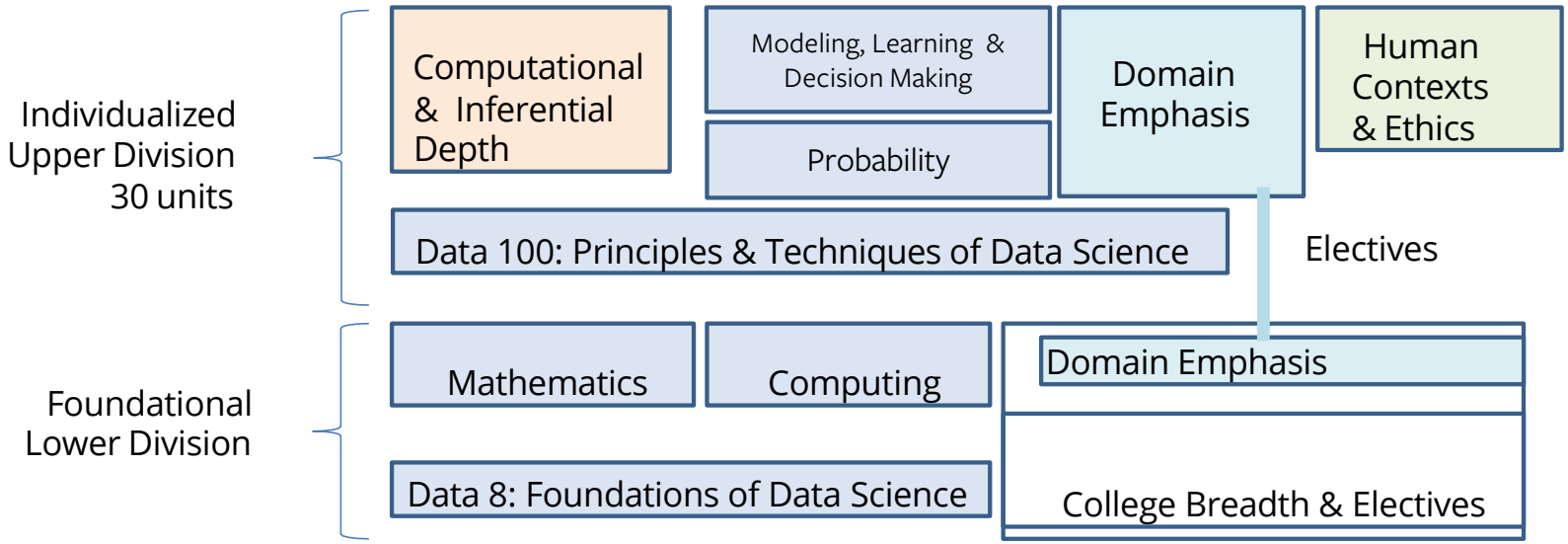
Where to go from here?



CS Major Options:

This is the best path to declare CS!

The Data Science Major



Data Science Info: https://data.berkeley.edu/2020-cal-week (Recordings)
There's also a minor!

Ask Is Anything!

UC Berkeley EECS
Lecturer
Michael Ball

Computational Structures in Data Science

THANK YOU!

(Again!)