

# Data 8 Lab 01

January 22, 2025



# outline

## **intro & logistics** **1**

---

about me

intro to data 8

course policies

resources & other important stuff

## **discussion worksheet** **2**

---

## **lab notebook** **3**

# about me

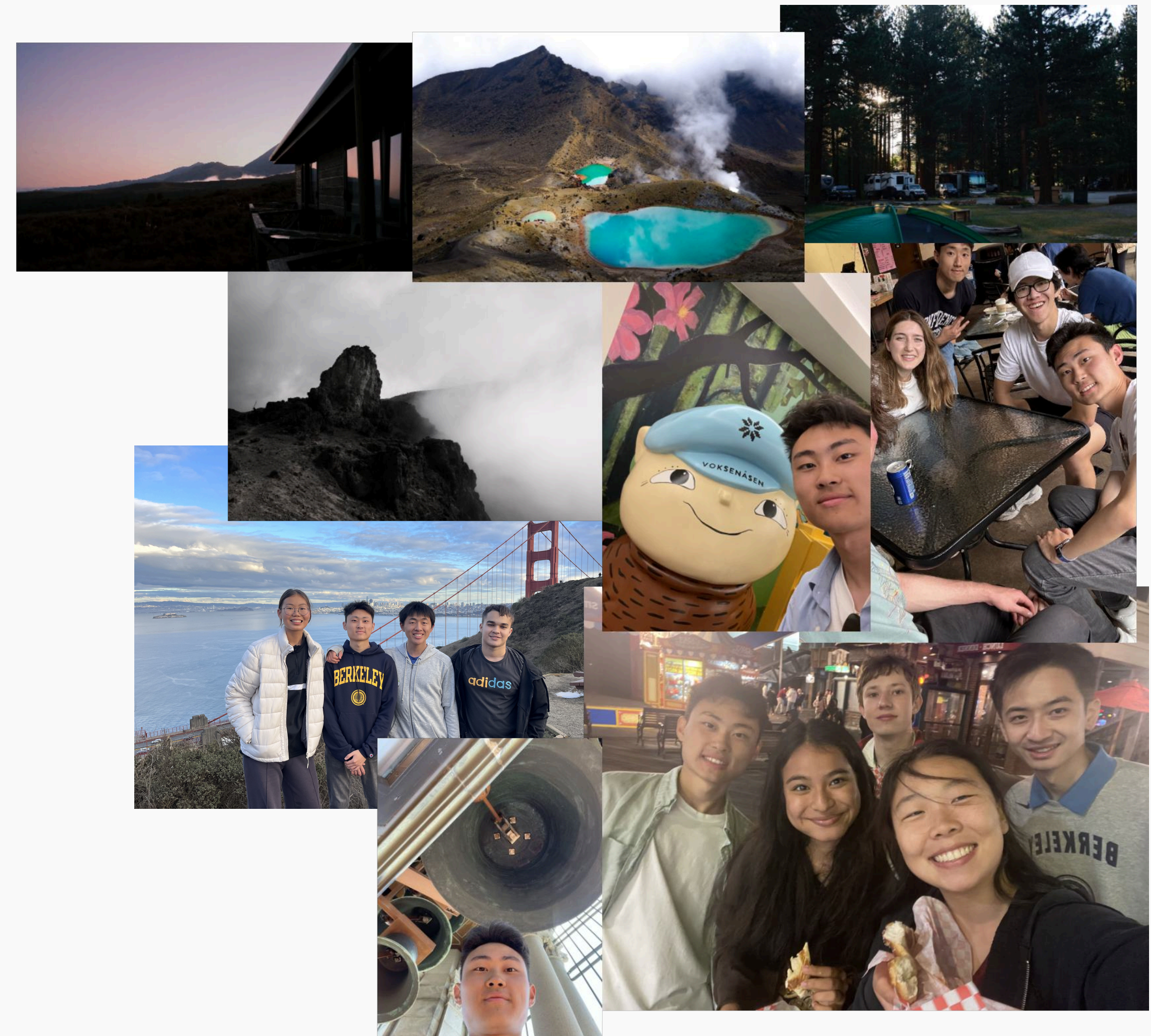
sophomore, studying cs & applied maths

born and raised in new zealand

i love hiking, camping, travelling!

some hobbies: speedcubing, photography, design,  
building custom PCs & keyboards, table tennis

on campus: ML@B, NT@B, Cal Cube Club,  
ml research @ BAIR



---

Feel free to reach out to me via [xie@berkeley.edu](mailto:xie@berkeley.edu) any time if you have questions about Data 8!

# visiting TAs & tutors

just for today, Aileen, Alex & Leo are joining us for this lab!

Aileen is Data 8's Grading Lead, and Alex & Leo are tutors this semester

# what is Data 8?

the best class on campus!

we learn what data science is & how to use data science as an important tool for virtually every field in our current world

you'll learn three crucial parts of data science:

1. programming (with Python) and computational thinking
2. statistics & probability
3. working with data and using computation for statistical processing

data science is done through **exploration** (collection & visualisation), **inference** (drawing conclusions from data with statistics) and **prediction** (making guesses about unseen data with machine learning)

# icebreaker time!

in groups of around 5:

introduce yourselves (name, where you're from, major, ...)

**what's the coolest place you've travelled to?**

and what's the place you want to travel to most?

# lab policy

## regular lab

**80% attendance, 20% lab notebook test cases**

submit notebook to Gradescope by Friday **5pm** deadline

2 lab drops for extenuating circumstances

## self-service lab

**100% lab notebook test cases**

submit assignment to Gradescope by Friday **5pm** deadline

2 lab drops

---

You can only switch option twice. Either during first week of lab sections, or after midterm scores have been released.

# homework policy

released on **Thursday**, due following **Wednesday at 5pm**

early submission bonus of **5 points** if made **24 hours before deadline**

2 homework drops for extenuating circumstances



# project policy

3 projects throughout the semester, each having a checkpoint

you can have a project partner, from your lab section

early submission bonus of **5 points** if made **24 hours before deadline**

no project drops, you have 2 weeks for each project, use your time wisely!

# grading breakdown

course components are labs, homeworks, projects and exams



# DSP accommodations

please submit your DSP letters as soon as you can, as assignments will have their deadlines automatically extended for you!

reach out to [data8@berkeley.edu](mailto:data8@berkeley.edu) if you have any questions

# submitting work

make sure you submit work correctly, or you will not get the correct credit

[submission walkthrough](#)

# so many resources

office hours, tutoring sections, and more...

tutoring sections start week 3

OH calendar can be found on the website

everything is on [data8.org](https://data8.org)

# discussion worksheet

---

discussion worksheets can be found on [data8.org](https://data8.org)

---

# 1 class sizes

Which of the following numbers is the closest to the average size of all undergraduate Berkeley classes in the 2023-2024 academic year?

30 40 50 60 70 80 90

# 1 class sizes

Which of the following numbers is the closest to the average size of all undergraduate Berkeley classes in the 2023-2024 academic year?

30 40 50 60 70 80 90



## 2 popular degrees

In the 2023-2024 academic year, close to half of all undergraduate degrees at Berkeley were awarded to students in the 10 largest major programs.

The top five (in decreasing order) are CDSS Computer Science, CDSS Data Science, Economics, MCB, and EECS. What are the next five?

## 2 popular degrees

In the 2023-2024 academic year, close to half of all undergraduate degrees at Berkeley were awarded to students in the 10 largest major programs.

The top five (in decreasing order) are CDSS Computer Science, CDSS Data Science, Economics, MCB, and EECS. What are the next five?

Business Administration (425), Political Science (403), Psychology (359), Sociology (273), and Media Studies (258).

For those curious, the numbers for the top five were CDSS Computer Science (891), Data Science (846), Economics (678), MCB (637), and EECS (510).

# 3 healthy living

“Life expectancy at birth” (LEB) measures roughly how many years a newborn baby is expected to live, and is a commonly used metric to quantify the health of a population.

What two states would you expect to have the highest LEBs among all states in the US?

# 3 healthy living

“Life expectancy at birth” (LEB) measures roughly how many years a newborn baby is expected to live, and is a commonly used metric to quantify the health of a population.

What two states would you expect to have the highest LEBs among all states in the US?

Hawaii (81.3) and California (80.9), but there is some uncertainty in these predictions. Connecticut and Minnesota are at 80.8.

# 4 worldwide

Picking from the list of numbers below, fill in the numbers that you expect to be the closest to the 2019 LEB of the United States as well as the world.

69 71 73 75 77 79

# 4 worldwide

Picking from the list of numbers below, fill in the numbers that you expect to be the closest to the 2019 LEB of the United States as well as the world.

69 71 73 75 77 79

United States: 78.5, World: 73

my website, [xie.nz/d8](http://xie.nz/d8)

---

all my resources (such as these slides) can be found on my website, [xie.nz/d8](http://xie.nz/d8)

---

# lab notebook

---

lab notebooks can also be found on [data8.org](https://data8.org)

---



that's it for Lab 1! feel free to leave once finished

# thanks for coming

---

all my resources (such as these slides) can be found on my website, [xie.nz/d8](http://xie.nz/d8)

---