CS109: Probability for Computer Scientists

Oishi Banerjee and Cooper Raterink
Based on slides by Lisa Yan
June 22, 2020

Quick slide reference

3	Introduction + Intro to counting	LIVE
65	Counting II	01b_counting_ii
73	Pigeonhole Principle	01c_pigeonhole
79	Permutations I	01d_permutations

Today's discussion thread: https://us.edstem.org/courses/667/discussion/79610

(If you haven't joined Ed yet, use this first: https://us.edstem.org/join/nhECh5)

Welcome to CS109!

Lecture with zoom

- Turn on your camera if you are able, mute your mic in the big room
- Virtual backgrounds are encouraged (classroom-appropriate)

Oishi Banerjee

Stanford Co-term
- B.A. in Classics (Latin and Greek)
- M.S. in Computer Science (Artificial Intelligence)
- Currently conducting medical AI research
- Fun fact: I sing opera in my spare time!

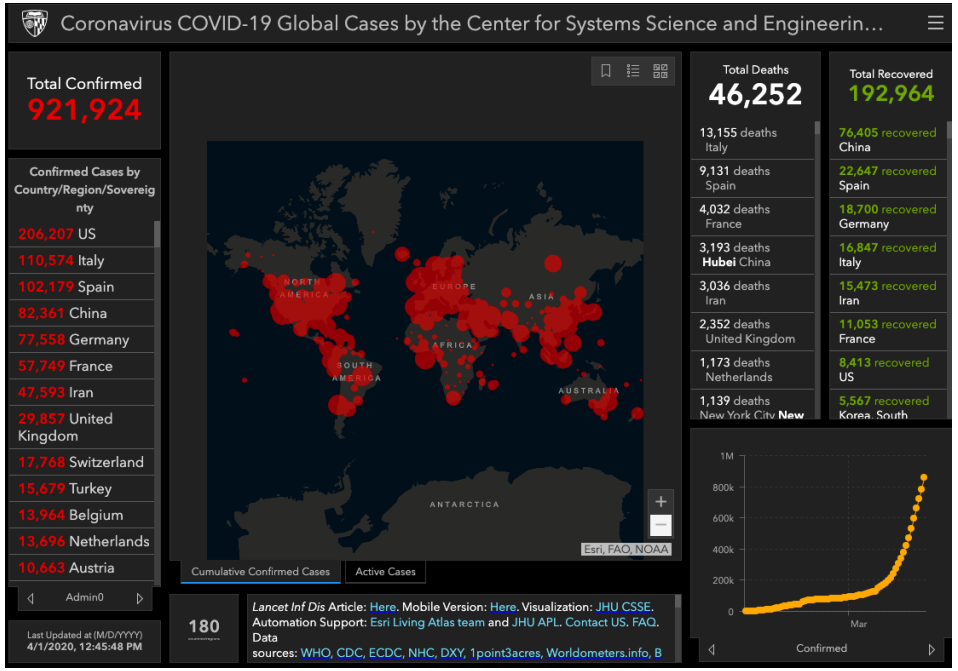
Cooper Raterink

Stanford Master's Student

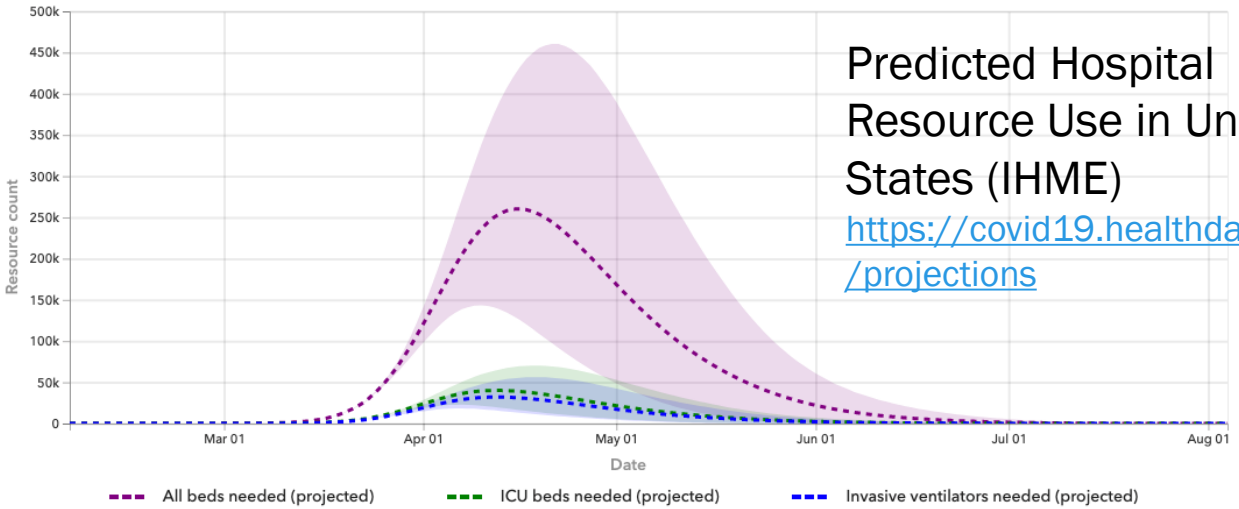
- B.S. in Electrical and Computer Engineering at UT Austin
- M.S. in Computer Science (Artificial Intelligence)
- I've done research on AI & Sustainability
- Interested in Humane AI
- Fun fact

What makes this quarter important

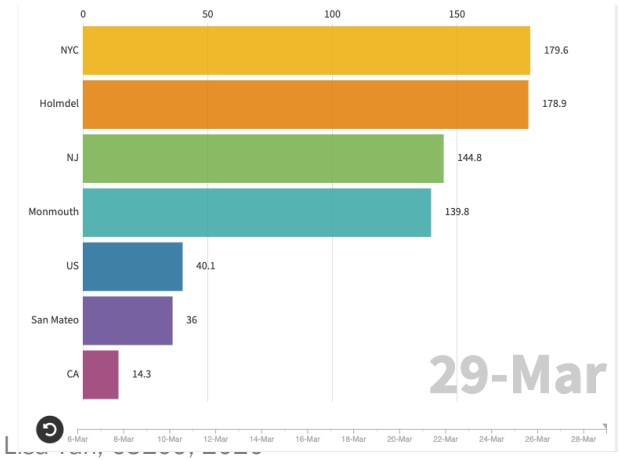
We are seeing a huge surge in **statistics, predictions, and probabilistic models** shared through global news, governing bodies, and social media.



Global cases of COVID-19 as of April 1st (JHU)
https://coronavirus.jhu.edu/map.html



Predicted Hospital Resource Use in United States (IHME)
https://covid19.healthdata.org/projections



Cases per 100K in NY, NJ, and CA counties (my dad)
https://app.flourish.studio/login

What makes this quarter important

We are seeing a huge surge in **statistics, predictions, and probabilistic models** shared through global news, governing bodies, and social media.

The challenge of delivering Stanford-class education online reflects our university's commitment to fostering a **diverse body of students**.

What makes this quarter important

We are seeing a huge surge in **statistics, predictions, and probabilistic models** shared through global news, governing bodies, and social media.

The challenge of delivering Stanford-class education online reflects our university's commitment to fostering a **diverse body of students**.

The **technological and social innovation** we develop during this time will strongly impact how we approach ***truly*** **world-class education**.

What makes this quarter important

We are seeing a huge surge in **statistics, predictions, and probabilistic models** shared through global news, governing bodies, and social media.

The challenge of delivering Stanford-class education online reflects our university's commitment to fostering a **diverse body of students**.

The **technological and social innovation** we develop during this time will strongly impact how we approach ***truly*** **world-class education**.

Our goals this quarter (at minimum)

To teach you how probability applies to real life

To help you foster and maintain human connections throughout this course

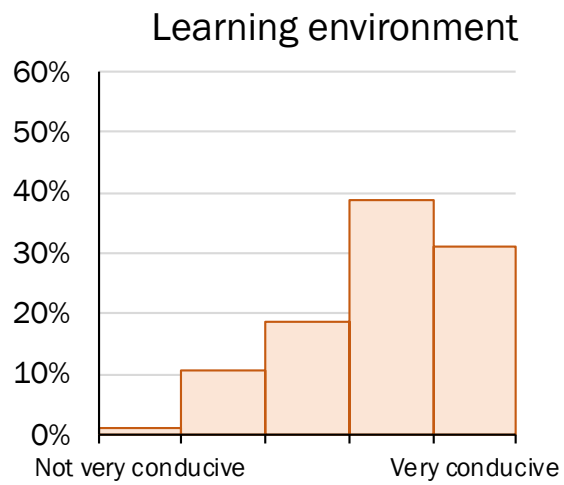
that being said…

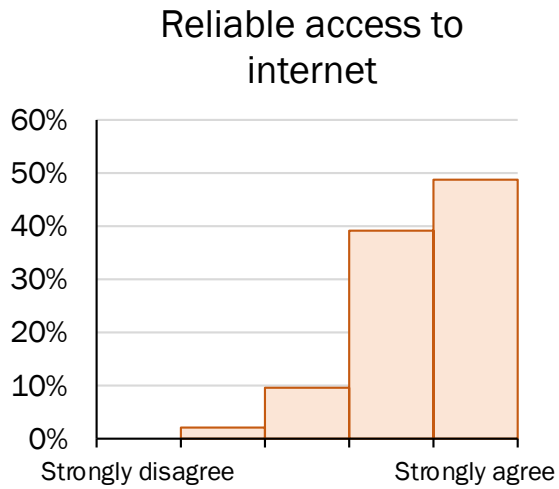
What makes this quarter important

These are extraordinary circumstances.

The teaching staff and I realize that this quarter cannot replace an in-person, on-campus experience. Your diverse backgrounds **amplify this difference**.

All our situations may change.





We are committed to working through this version of this course together and adapting as a class and as a community. We welcome your thoughts.

Thank you in advance for being patient with necessary changes to make this educational experience **fulfilling, meaningful, and equitable**.

What about you?

...first, some Breakout Room guidelines...

Lecture with Zoom

- Turn on your camera if you are able, mute your mic in the big room
- Virtual backgrounds are encouraged (classroom-appropriate)

Breakout Rooms for meeting your classmates

- Just like sitting next to someone new

We will use Ed instead of Zoom chat

- Like raising your hand in the classroom, except with a **lower barrier to entry**
- You can **upvote** your classmates' posts
- Persistent copy: Teaching staff and I can answer questions *during* and *after* lecture
- Better threading/reply support, copy/paste, LaTeX math mode, emojis

Join discussion forum here: https://us.edstem.org/join/nhECh5

Today's discussion thread: https://us.edstem.org/courses/667/discussion/79610

By yourself

Post or upvote some thoughts on Ed:

- What is something you hope to get out of this quarter?
- What are you worried about this quarter?
- What are your hopes for CS109, given that it is online?

Join discussion forum here:
https://us.edstem.org/join/nhECh5

Today's discussion thread:
https://us.edstem.org/courses/667/discussion/79610

Breakout Rooms

Introduce yourself! (name, major, year)

Then check out the responses your classmates wrote, and comment/discuss!

- What is something you hope to get out of this quarter?
- What are you worried about this quarter?
- What are your hopes for CS109, given that it is online?

Join discussion forum here:
https://us.edstem.org/join/nhECh5

Today's discussion thread:
https://us.edstem.org/courses/667/discussion/79610

Course mechanics

Course mechanics (light version)

- For more info, read the Administrivia handout and FAQ
- Course website:

http://cs109.stanford.edu/

- Canvas (only for posting videos/recordings)

Prerequisites

CS106B/X

Programming
Recursion
Hash tables
Binary trees

MATH 51/CME 100

Multivariate differentiation
Multivariate integration
Basic facility with linear algebra (vectors)

CS103
(co-requisite OK)

Proofs (induction)
Set theory
Math maturity

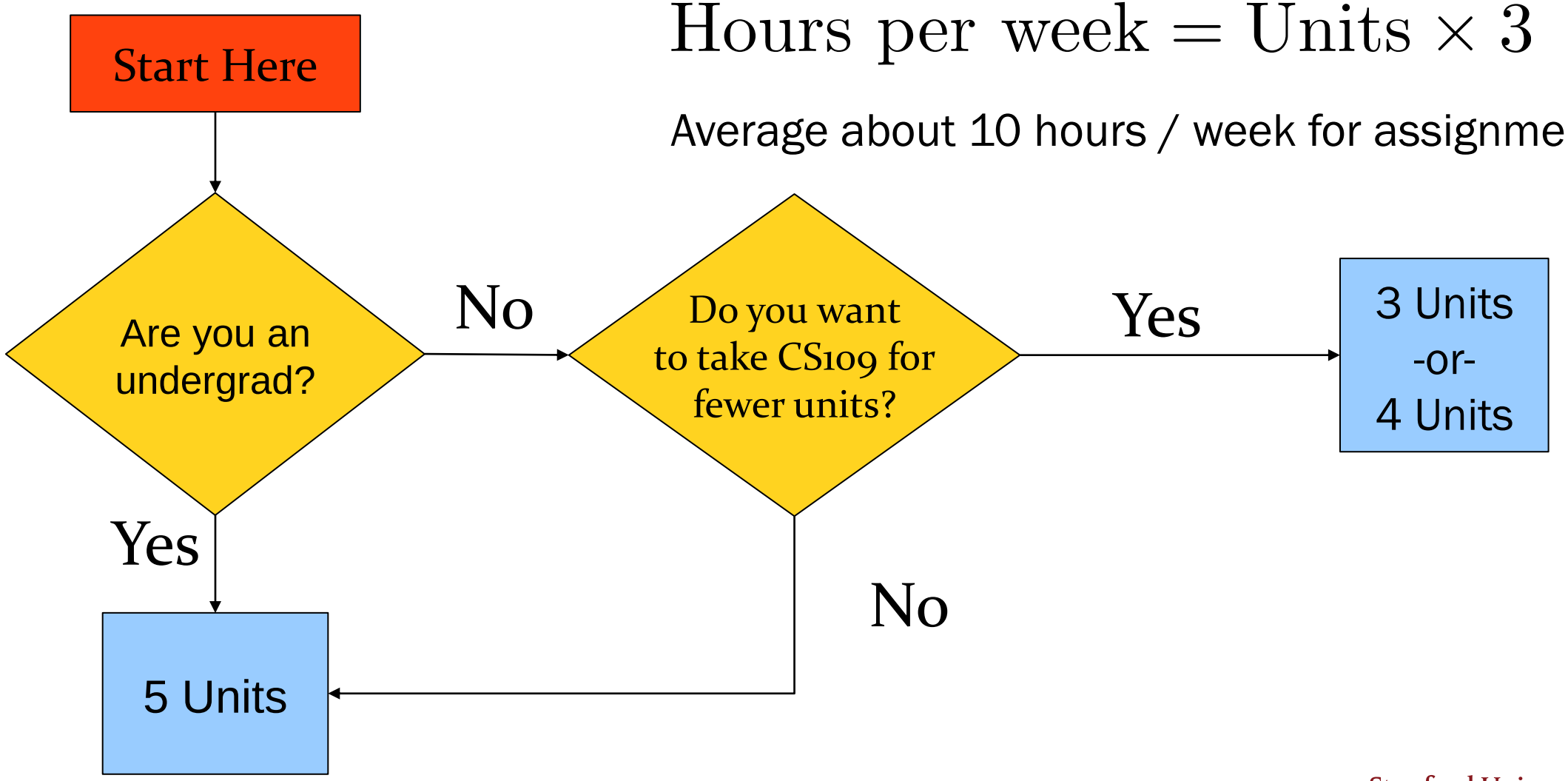
Important!

How many units should I take?

15 5

$$\text{Hours per week} = \text{Units} \times 3$$

Average about 10 hours / week for assignments



Staff contact

- Discussion forum: https://us.edstem.org/courses/667/discussion/
- Staff email: cs109-sum1920-staff@mailman.stanford.edu
- Office Hours start Tuesday
  - Find the schedule on the website
- Contact mailing list for course level issues, extensions, etc.

Lecture format

”Probability is a number between 0 and 1”

”What is the definition of probability? (select one)”

“What is the probability that you get exactly 3 heads in 5 coin flips?”

Short **pre-recorded** lecture
(several 5-10 min videos)

Concept check quiz on Gradescope
(submit infinitely many times,
maybe on-time bonus)

In-person, discussion-oriented lecture
MWF 1:30pm PT
(<110min)

Where you learn

Pre-recorded lectures

Live lectures recordings posted to Canvas

Optional Discussion Section starting Week 1

Lecture notes on website

Textbook readings optional

Problem Sets

Quizzes

Class breakdown

60% **6 Problem Sets**

25% **Quizzes**

- Take-home format, more details later
- Monday, July 20
- Friday, August 14

15% **Participation**

- Concept checks on pre-recorded material

60% Problem Sets

Late Policy

+5% for on-time submission

+0% bonus for 1 **class day** late

-20% for 2 **class days** late

-40% for 3 **class days** (1 week) late

Optional but encouraged, tutorial online

More information coming soon

Quizzes, Participation

25% Quizzes

- 12.5% each
- Around 2 hours of individual work
- 24-hour take-home window

15% Participation

- (15%) Concept checks: based on pre-lecture recordings
- We recommend you complete concept checks before lecture
- Unlimited submissions/autograded until last day of classes, August 13

Stanford Honor Code

Permitted

- Talk to the course staff
- Talk with classmates (cite collaboration)
- Look up general material online

NOT permitted:

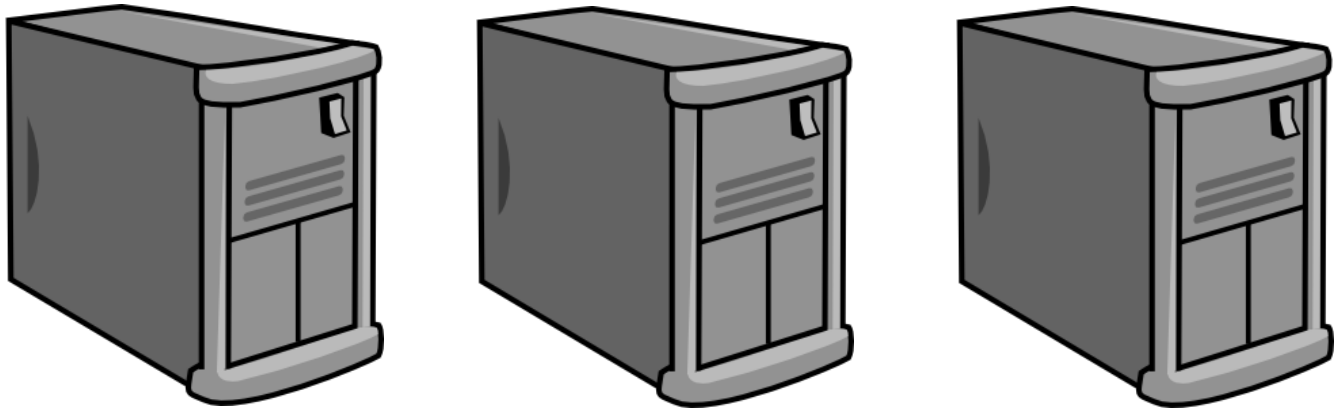
- Copy answers:
  - from classmates
  - from former students
  - from previous quarters
- Copy answers from the internet
  - Besides, these are usually incorrect

Why you should take CS109

Traditional View of Probability

CS view of probability

http://www.site.com



Machine Learning

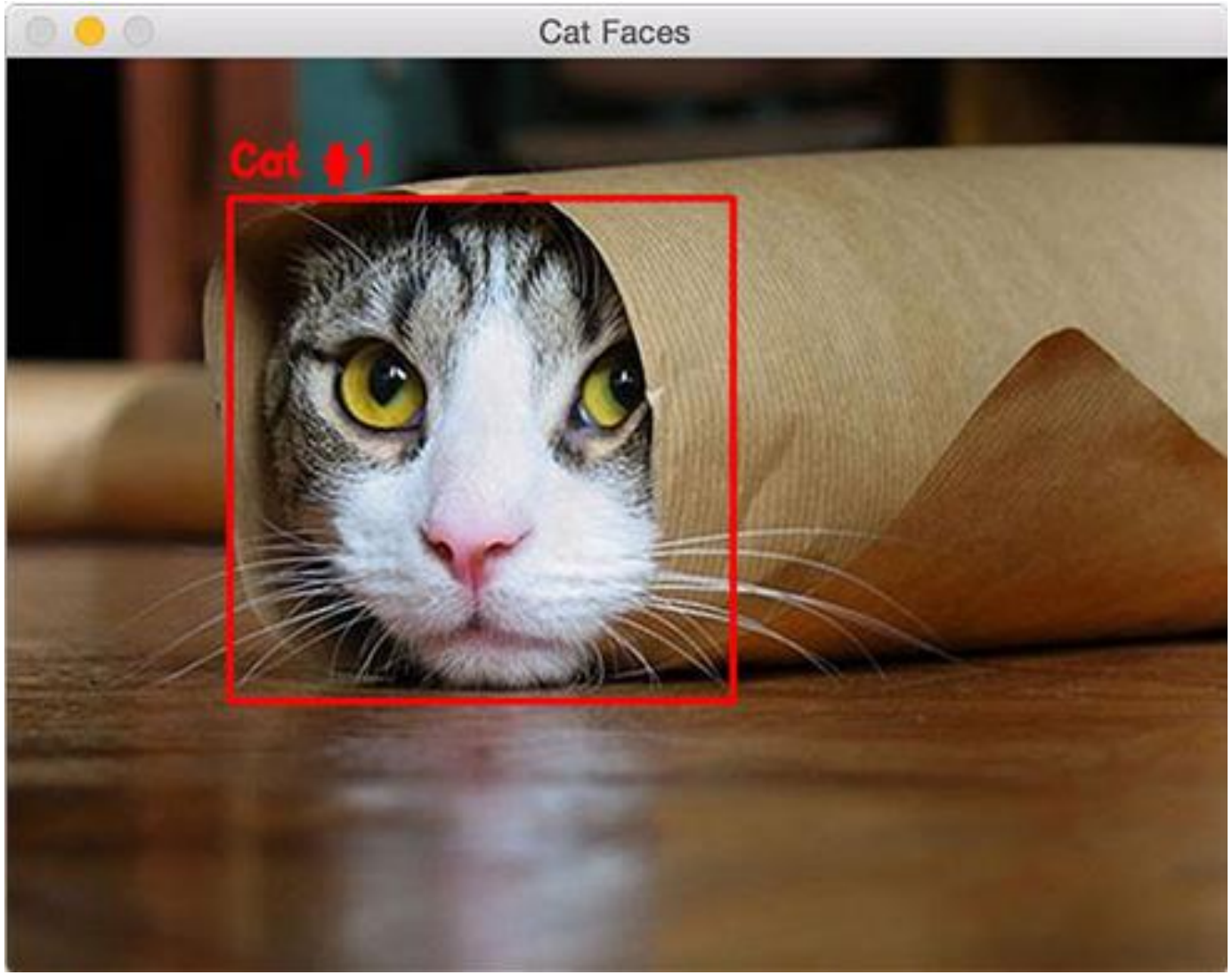
= Machine (compute power)

+ Probability

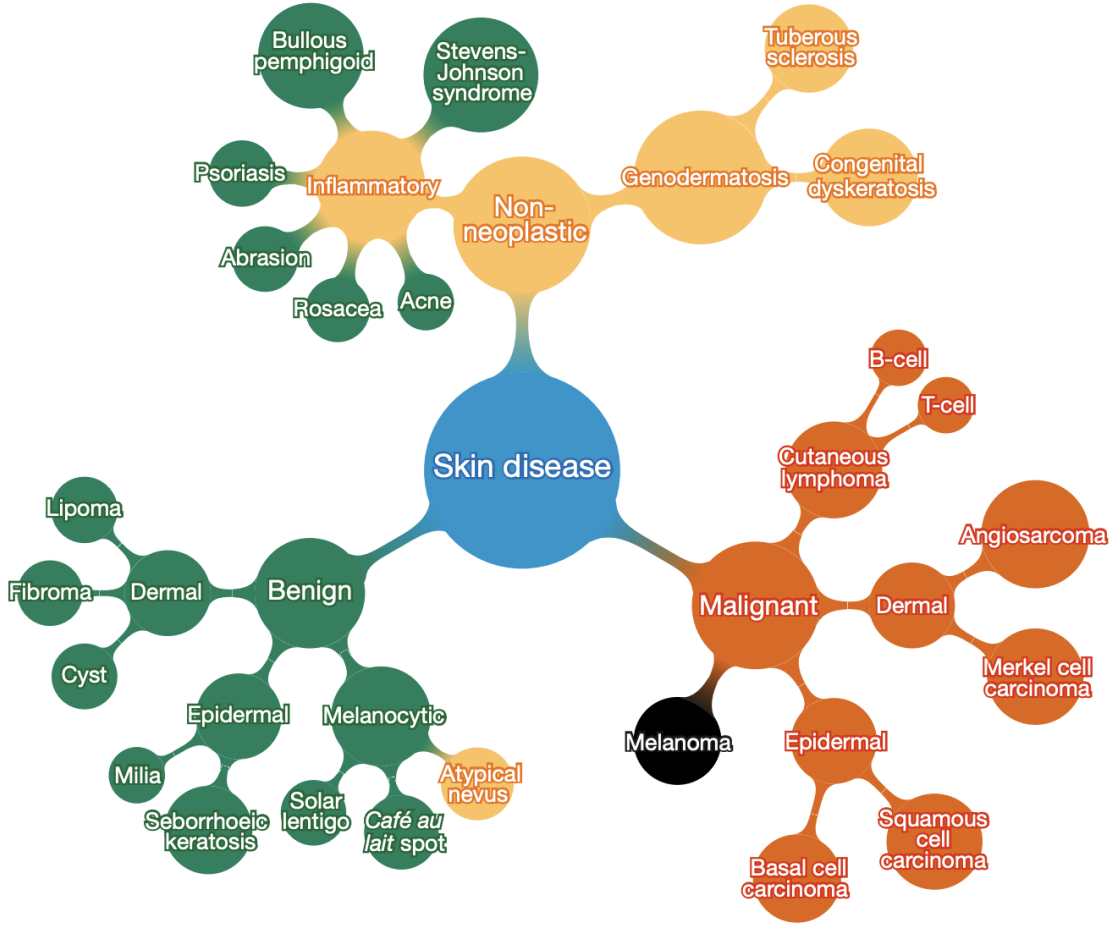
+ Data

Machine Learning Algorithm

Classification



Where is this useful?

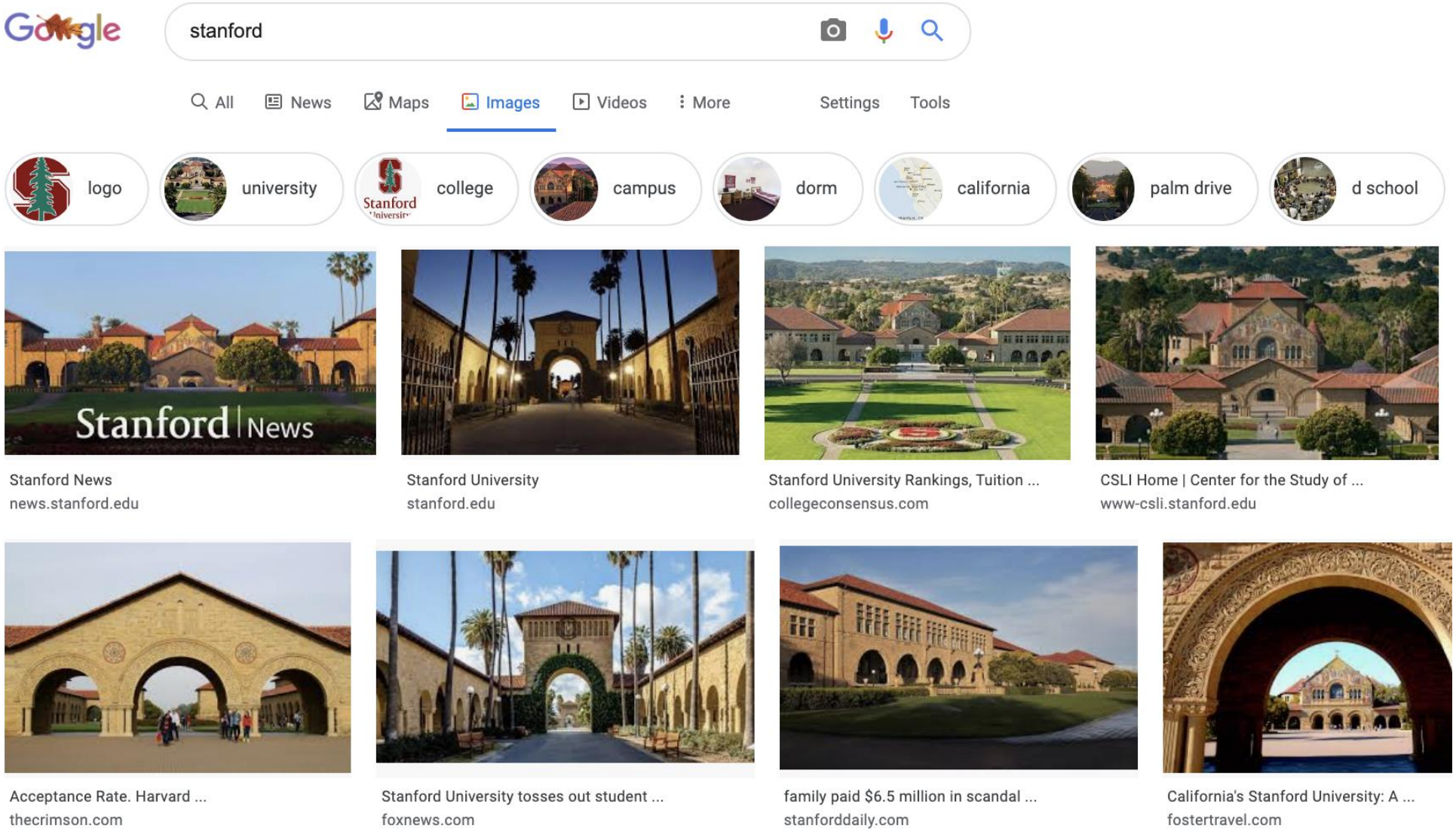


A machine learning algorithm performs **better than** the best dermatologists.

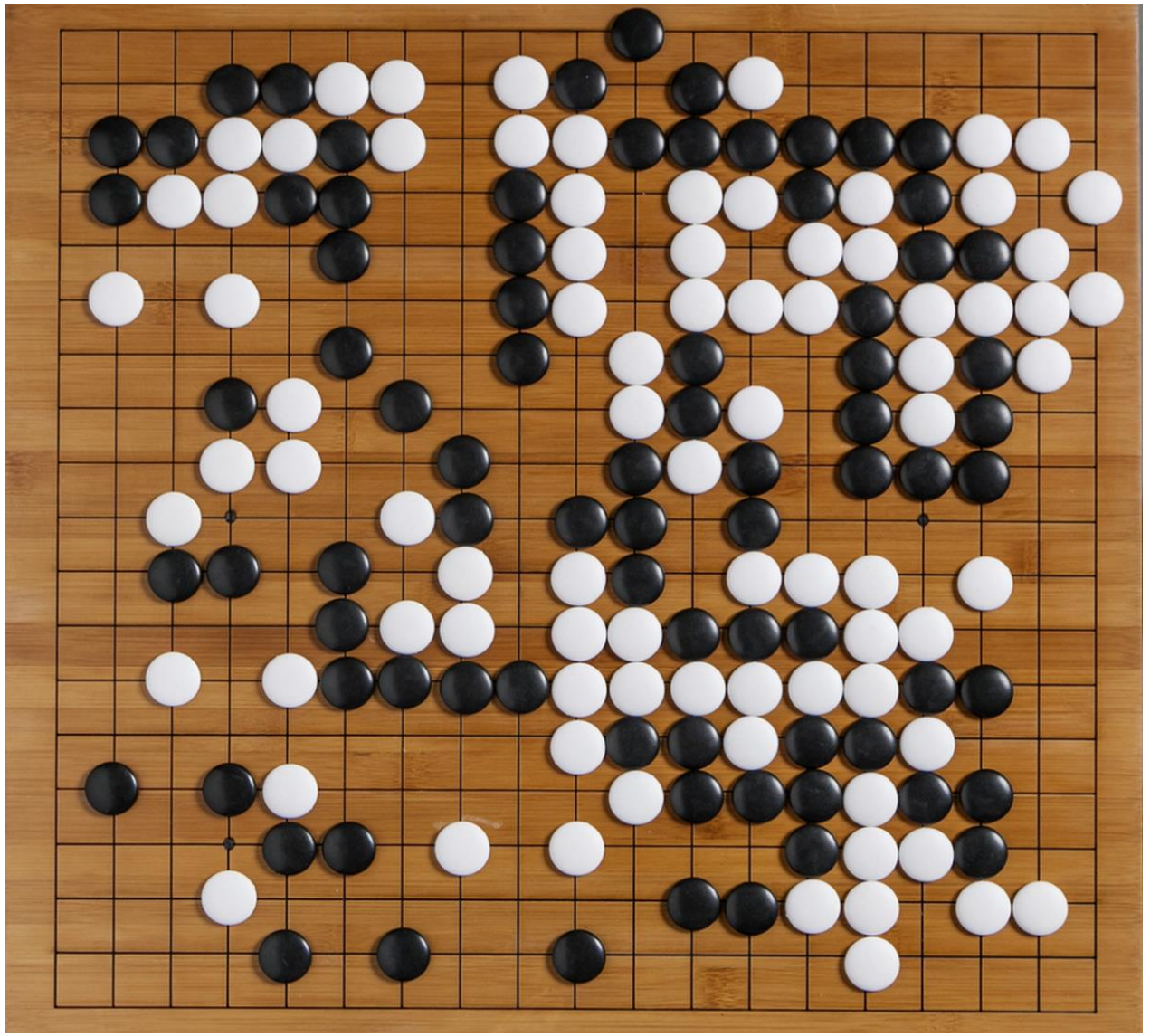
Developed in 2017 at Stanford.

Esteva, Andre, et al. "Dermatologist-level classification of skin cancer with deep neural networks." *Nature* 542.7639 (2017): 115-118.

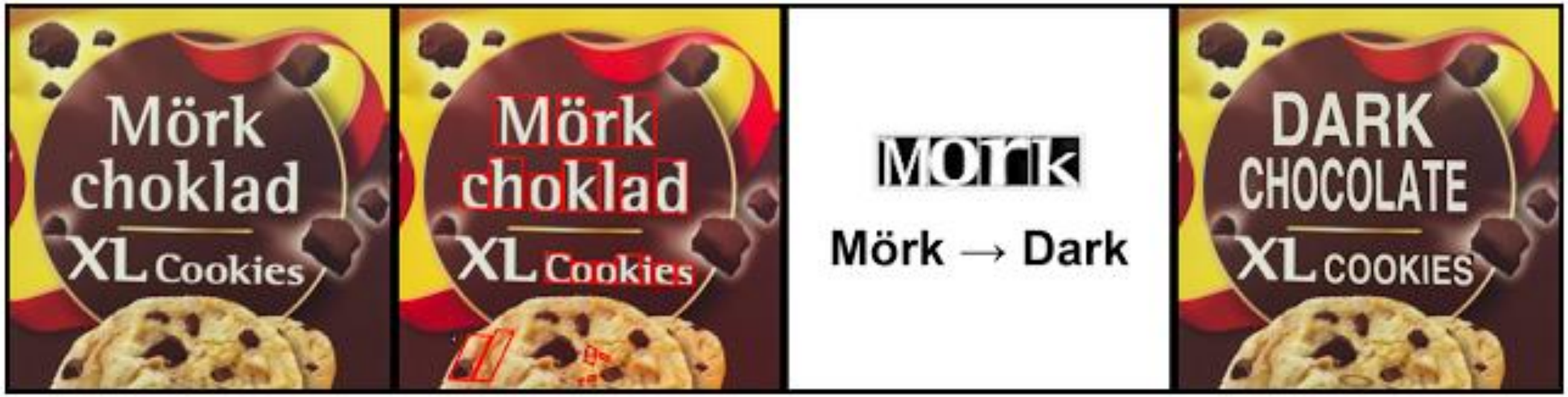
Image tagging



Decision-making: The last remaining board game

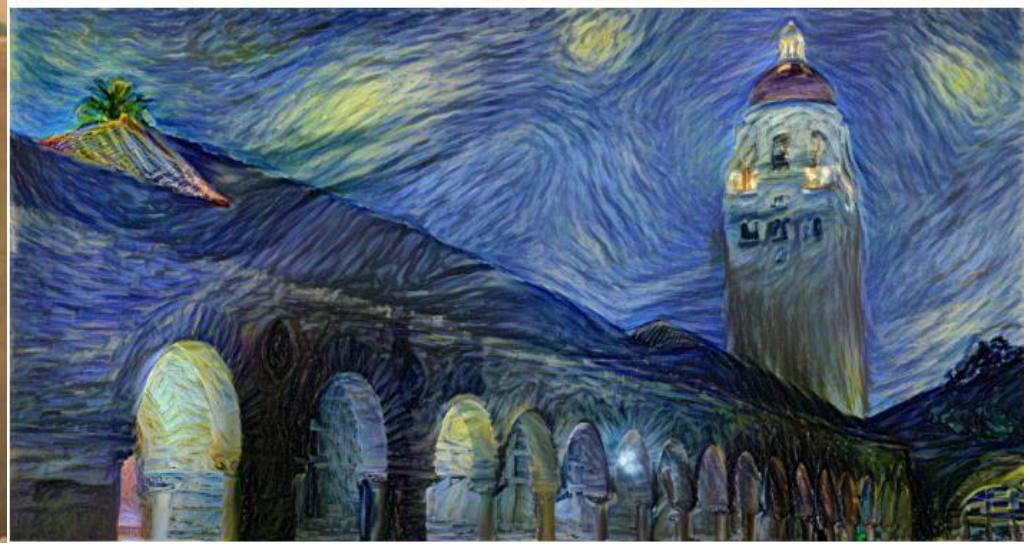


Augmented Reality Machine Translation

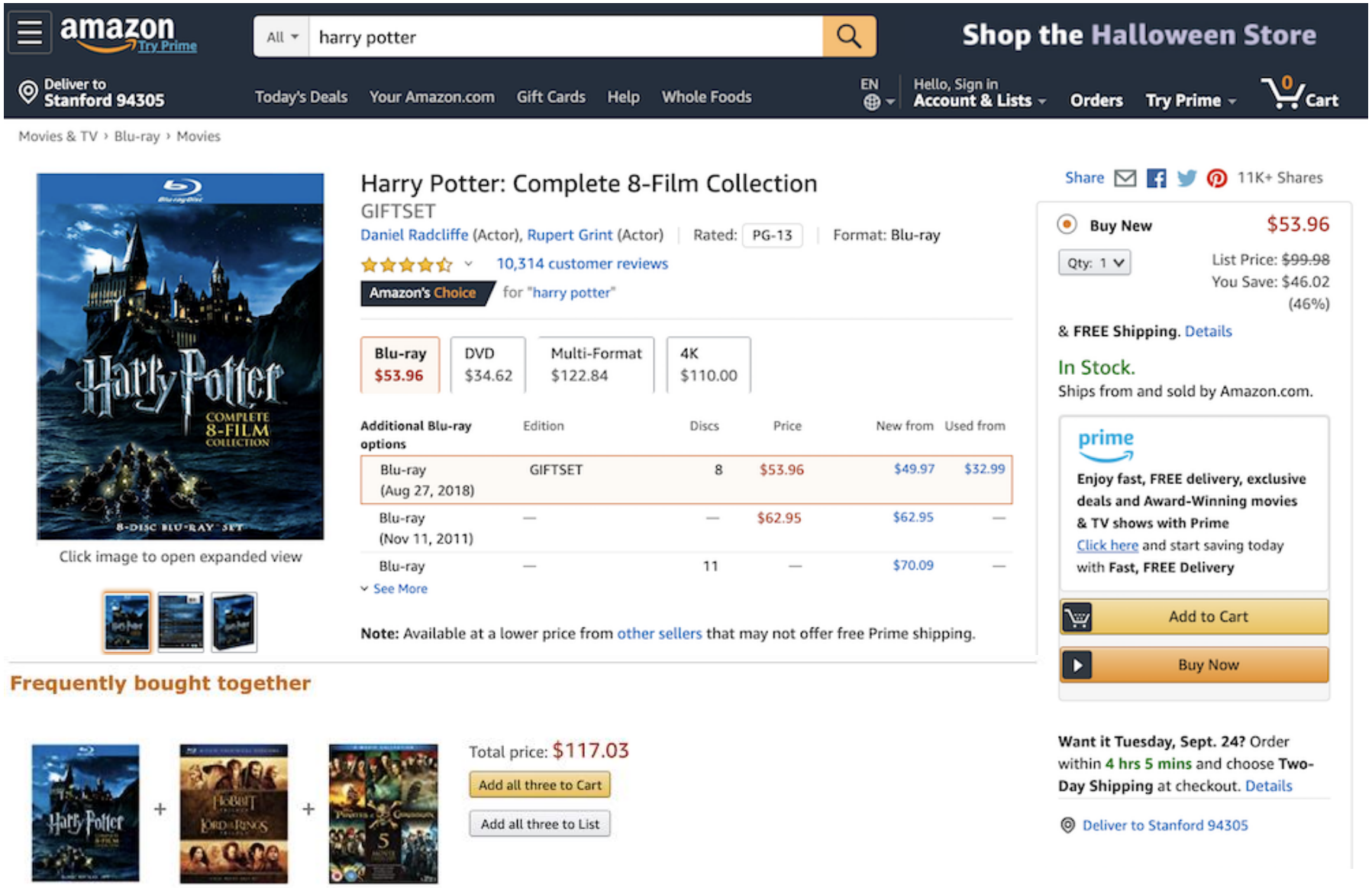


Automatic machine translation on Google Translate

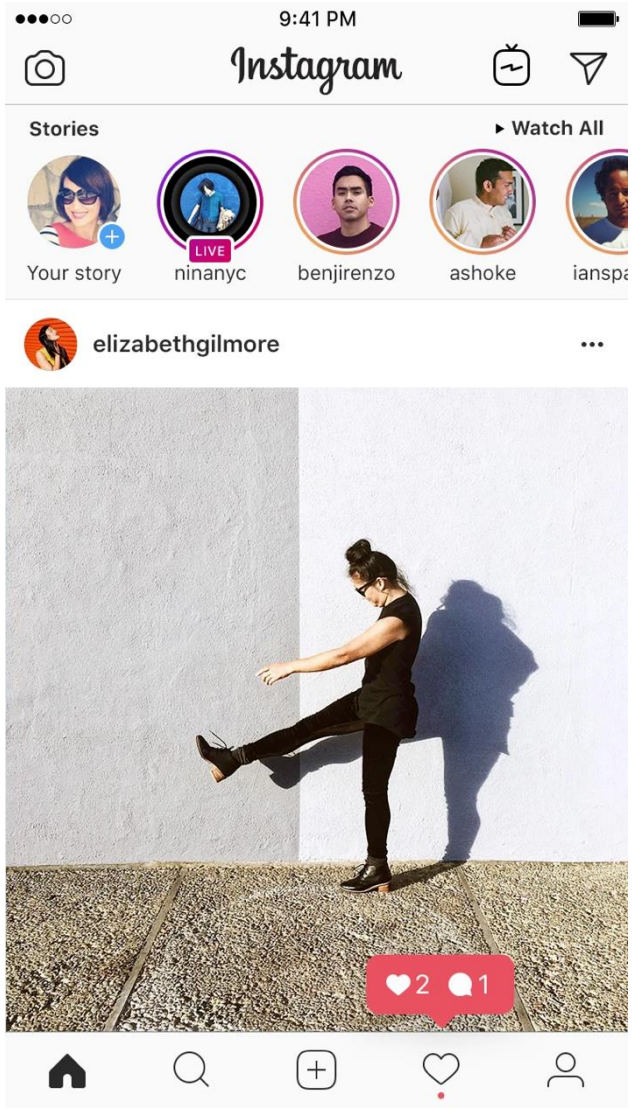
Style transfer



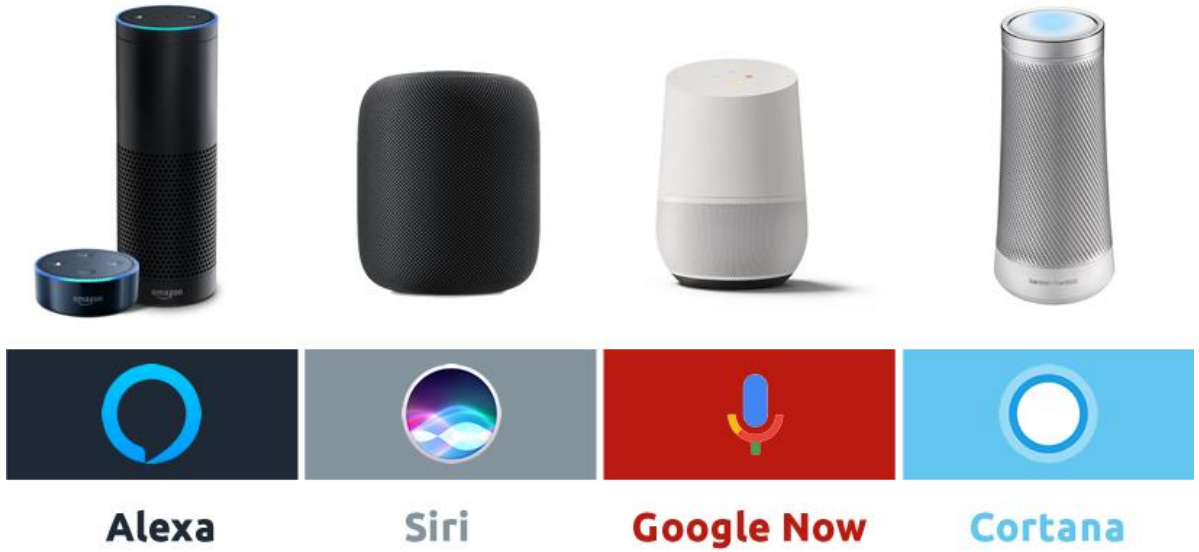
Content ranking and grouping



Probability at your fingertips

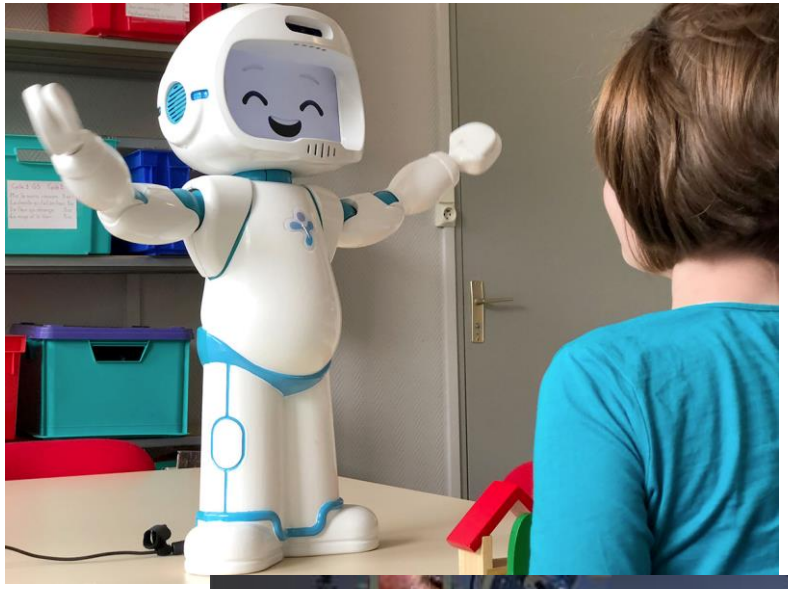


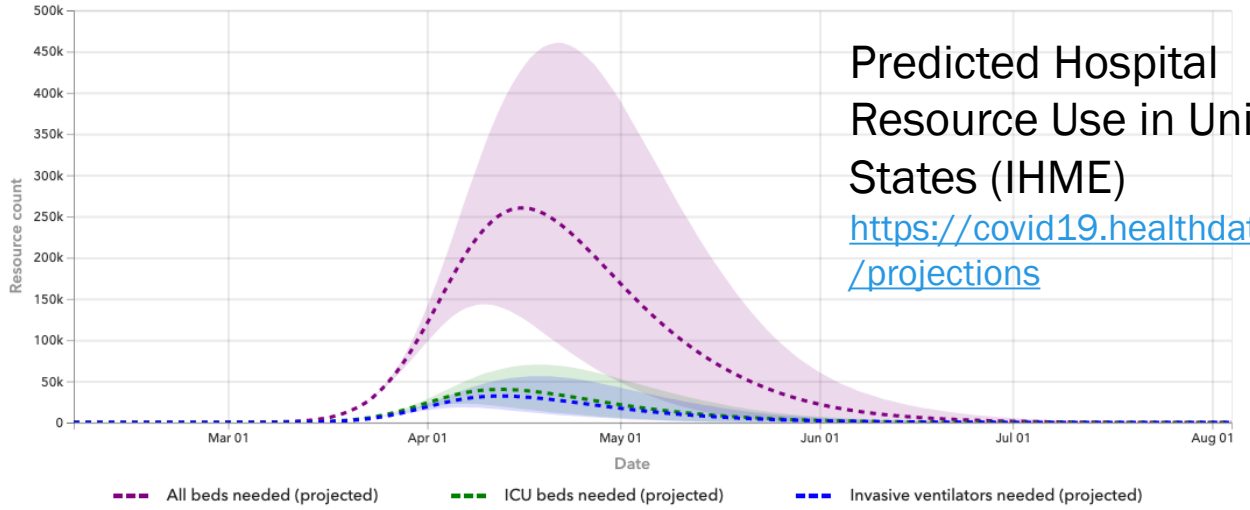
Voice assistants



Probability is *more* than just machine learning.

Probability and medicine

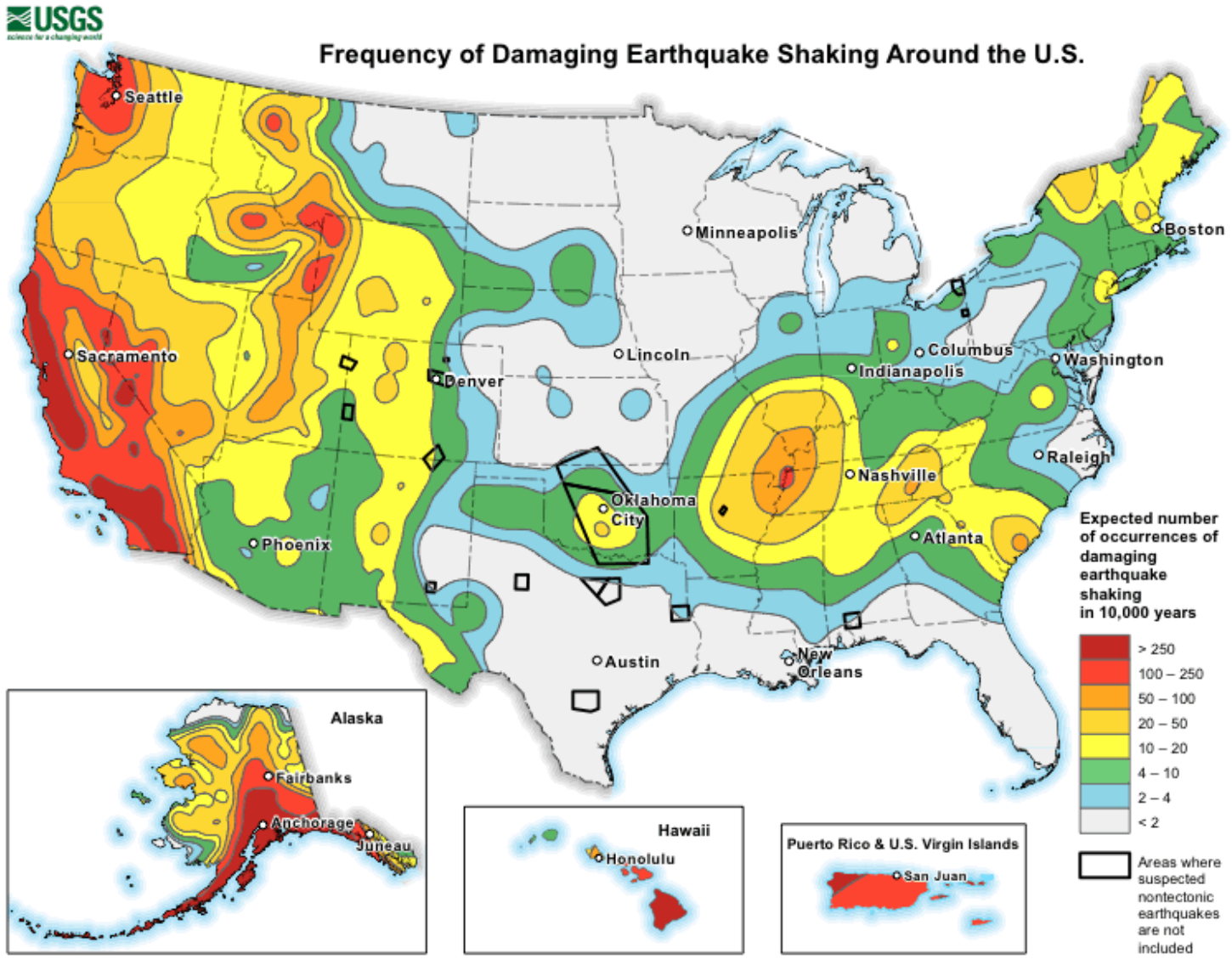




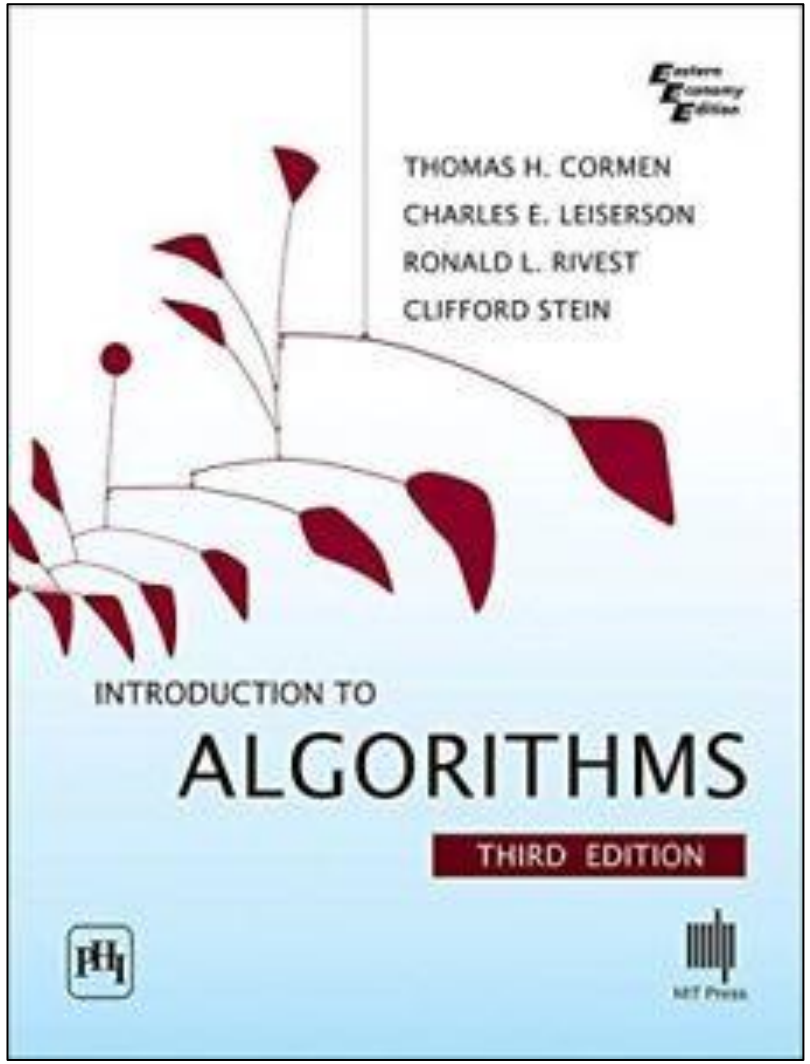
Predicted Hospital Resource Use in United States (IHME) https://covid19.healthdata.org/projections

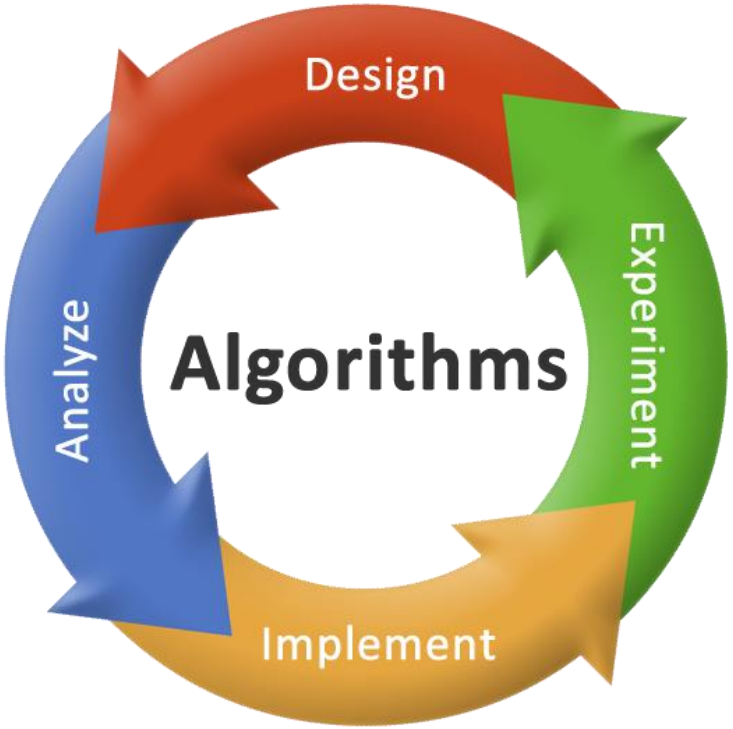
How do COVID-19 testing rates in a region correlate with the actual spread of the disease?

Probability and climate



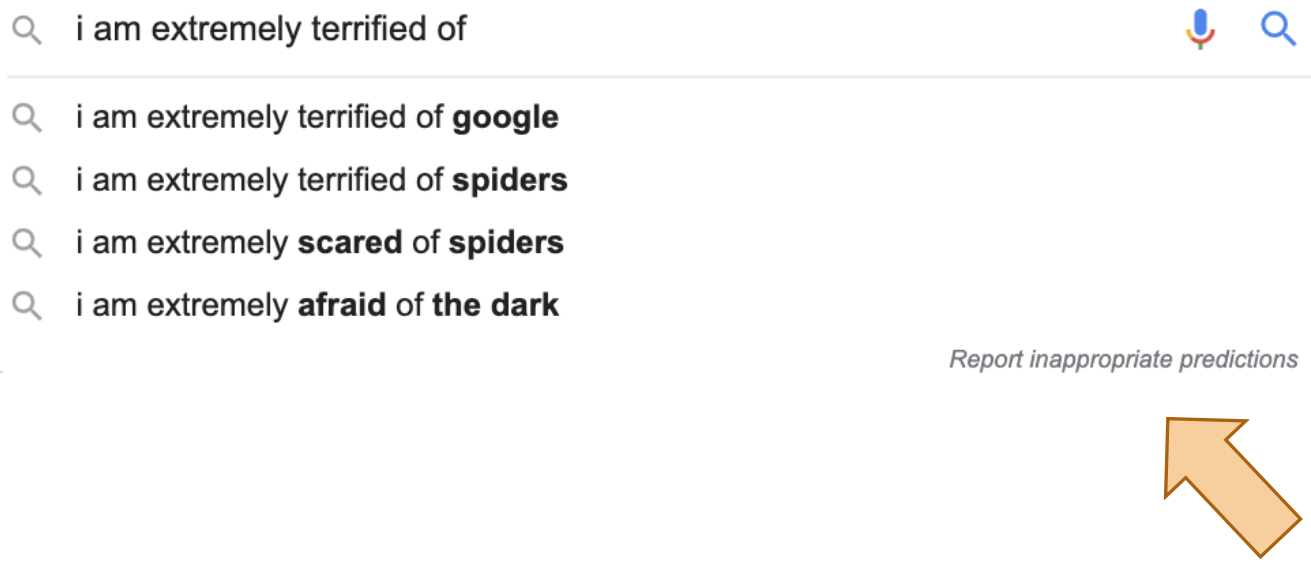
Probabilistic analysis of algorithms





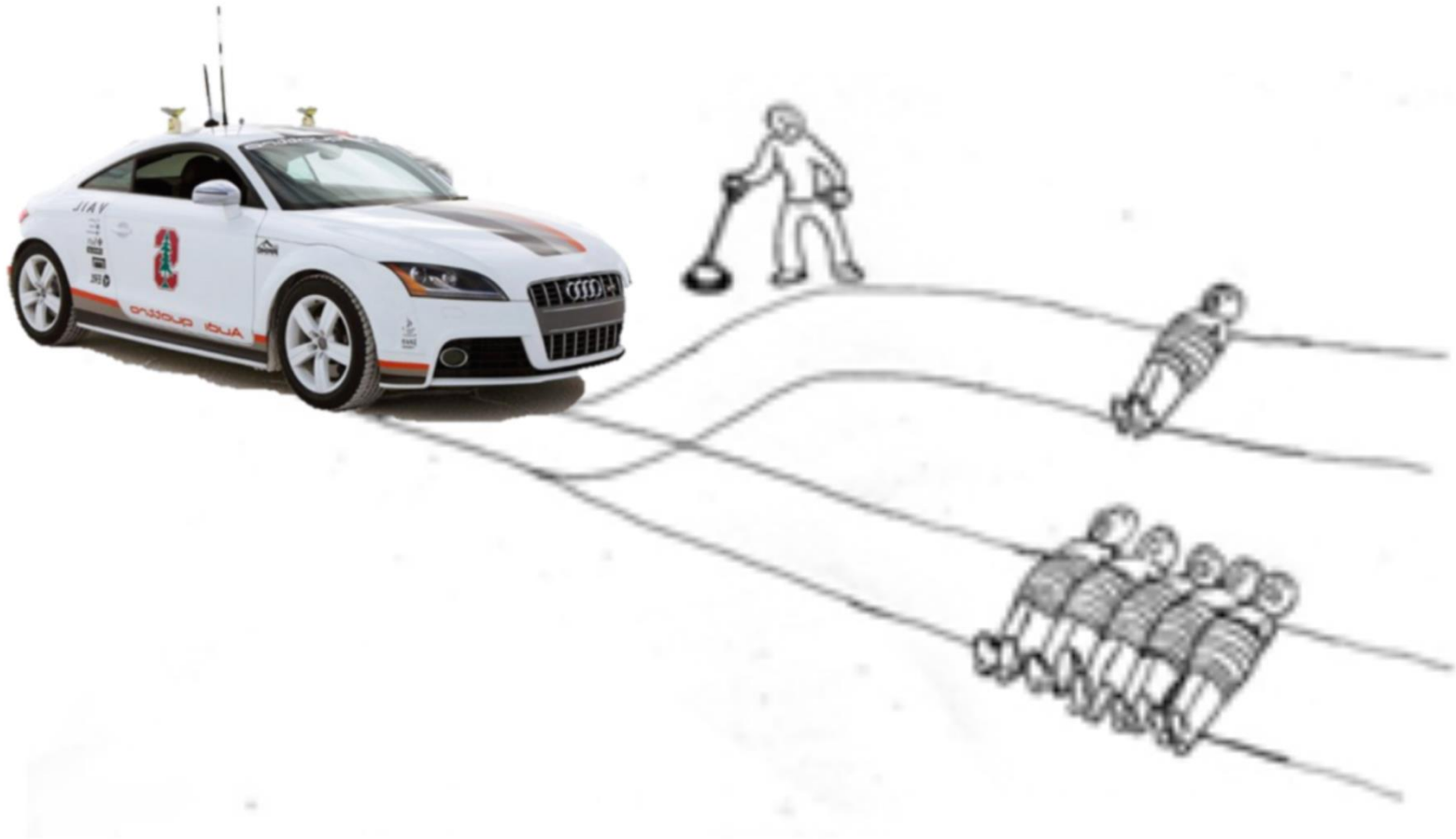
Probability for good

Algorithms of Oppression,
Safiya Umoja Noble. 2018



How do we identify systemic biases in our data and incorporate human judgment into our probabilistic models?

Probability and philosophy



We'll get there!

Probability is not always intuitive.

Disease testing

A patient takes a virus test that returns positive.

What is the probability that they have the virus?

- 0.03% of people have the virus
- Test has 99% positive rate for people with the virus
- Test has 7% positive rate for people without the virus

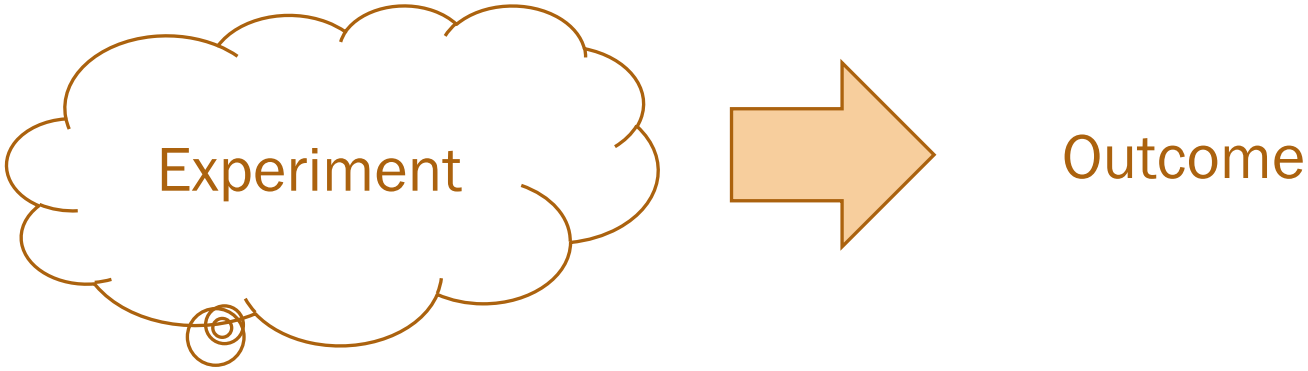
Correct answer: 42/10000 (0.42%)

Probability = Important
+ Needs Studying

Counting I

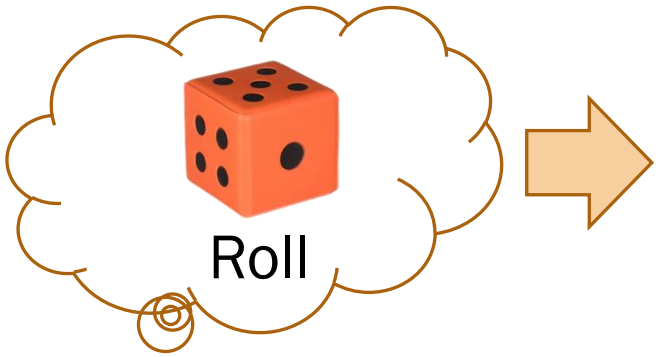
What is Counting?

An experiment in probability:



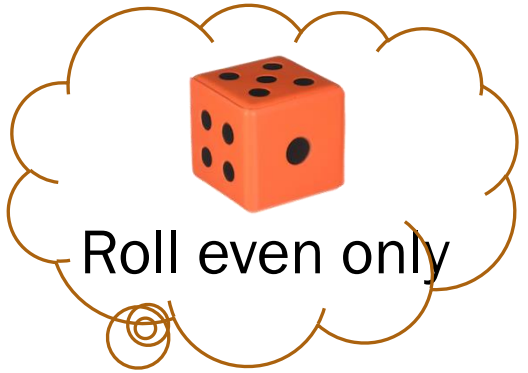
Counting: How many possible outcomes can occur from performing this experiment?

What is Counting?



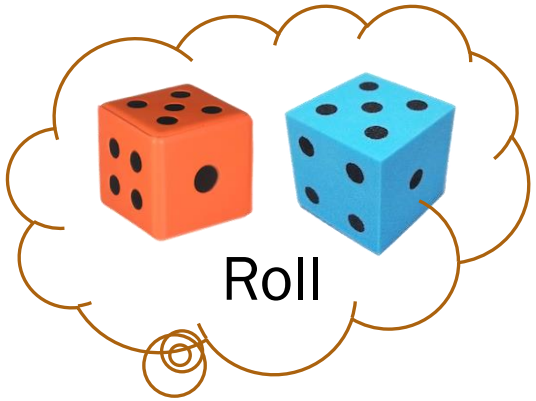
6

{1, 2, 3, 4, 5, 6}



3

{2, 4, 6}



36

{(1, 1) , (1, 2), (1, 3), (1, 4), (1, 5), (1, 6),
(2, 1) , (2, 2), (2, 3), (2, 4), (2, 5), (2, 6),
(3, 1) , (3, 2), (3, 3), (3, 4), (3, 5), (3, 6),
(4, 1) , (4, 2), (4, 3), (4, 4), (4, 5), (4, 6),
(5, 1) , (5, 2), (5, 3), (5, 4), (5, 5), (5, 6),
(6, 1) , (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)}

Sum Rule of Counting

If the outcome of an experiment can be either from

Set A, where $|A| = m$,

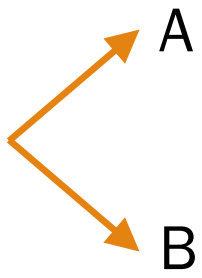
or Set B, where $|B| = n$,

where $A \cap B = \emptyset$,

Then the number of outcomes of the experiment is

$$|A| + |B| = m + n.$$

One experiment



Product Rule of Counting

If an experiment has two parts, where

> The first part's outcomes are from Set A, where $|A| = m$,
> and the second part's outcomes are from Set B, where $|B| = n$
> regardless of part one's outcomes,

Then the number of outcomes of the experiment is

$$|A||B| = mn.$$

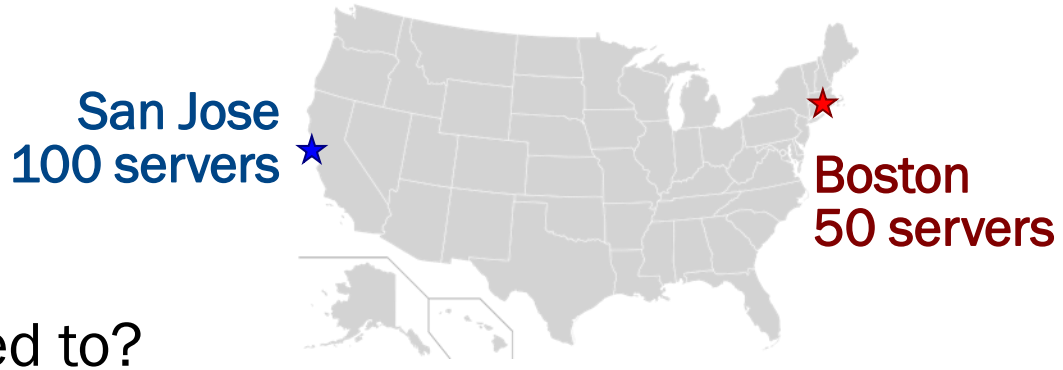
Two-step experiment

Let's try it out

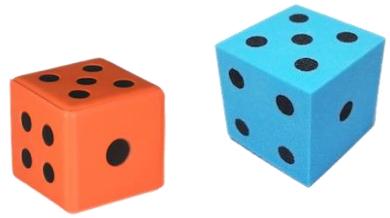
Think, pair, and we'll come back as a group.
Post any questions here:
https://us.edstem.org/courses/109/discussion/24490

Sum Rule, Product Rule, or something else? How many outcomes?

1. Video streaming application
   - Your application has distributed servers in 2 locations (SJ: 100, Boston: 50).
   - If a web request is routed to a server, how large is the set of servers it can get routed to?



2. Dice
   - How many possible outcomes are there from rolling two 6-sided dice?



3. Strings
   - How many *different* orderings of letters are possible for the string BOBA?

BOBA, ABOB, OBBA...

Let's try it out

Sum Rule, Product Rule, or something else? How many outcomes?

1. Video streaming application
   - Your application has distributed servers in 2 locations (SJ: 100, Boston: 50).
   - If a web request is routed to a server, how large is the set of servers it can get routed to?

   A = {100 servers in San Jose}
   B = {50 servers in Boston}
   $|A| + |B| =$ **150**

2. Dice
   - How many possible outcomes are there from rolling two 6-sided dice?

   A = {1, 2, 3, 4, 5, 6 on 1st die}
   B = {1, 2, 3, 4, 5, 6 on 2nd die}
   $|A||B| = 6 \cdot 6 =$ **36**

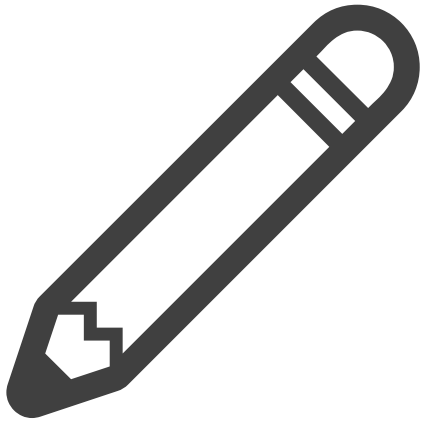
3. Strings
   - How many *different* orderings of letters are possible for the string BOBA?

   First letter's options = {B, O, A}
   Second letter's options = ???
   Final answer is **12.**
   See the recorded videos for why…

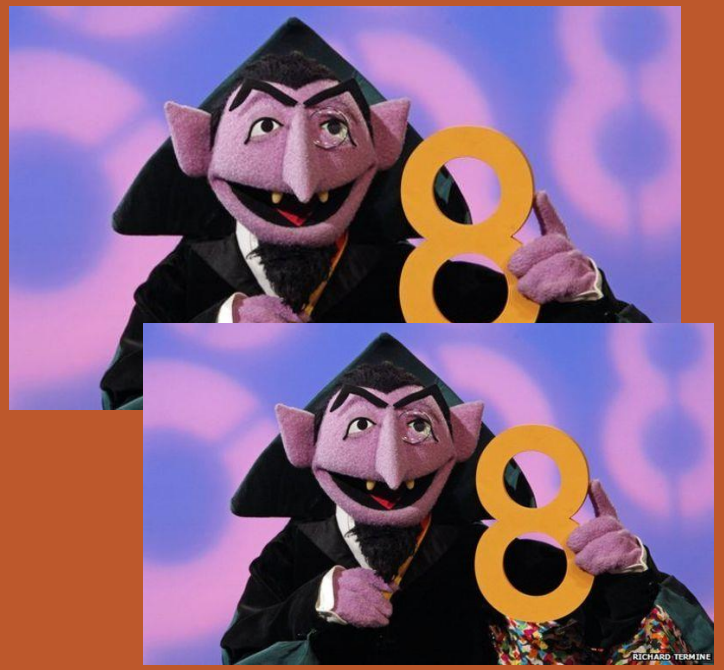
For next time

- Watch pre-recorded lectures for Wednesday 6/24 posted on the website schedule
  - You'll see something like: "Watch: 1_all, 2_all," indicating to watch videos from the 1st and 2nd series on Canvas
- Complete one concept check that covers both lectures to be posted this afternoon PT

http://cs109.stanford.edu/



Thanks for listening!



Counting II

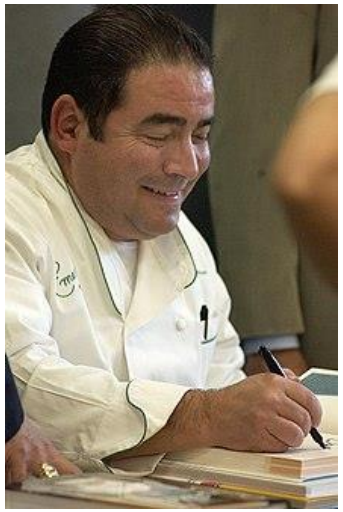
Gradescope quiz, blank slide deck, etc.
(Available Monday 4/6 evening PT)
http://cs109.stanford.edu/

TOP DEFINITION

kick it up a notch

To make things more intense, exciting, or interesting.

(introduced by **chef Emeril Lagasse** in reference to **spicing** up his recipes)



Inclusion-Exclusion Principle

If the outcome of an experiment can be either from

Set A or set B,

where A and B may overlap,

Then the total number of outcomes of the experiment is

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

One experiment

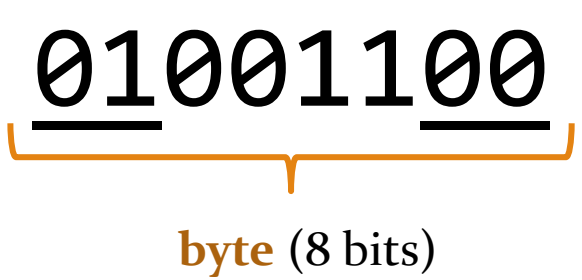
Sum Rule of Counting:
A special case

Transmitting bytes over a network

An 8-bit string is sent over a network.

- The receiver only accepts strings that either start with 01 or end with 10.

How many 8-bit strings will the receiver accept?



Define

A : 8-bit strings starting with 01

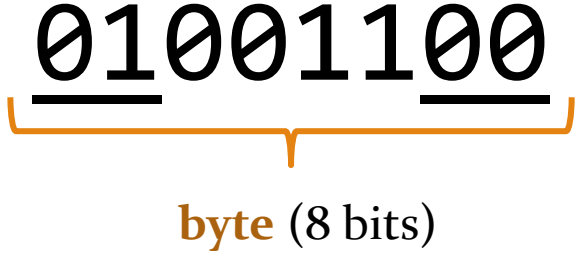
B : 8-bit strings ending with 10

Transmitting bytes over a network

An 8-bit string is sent over a network.

- The receiver only accepts strings that either start with 01 or end with 10.

How many 8-bit strings will the receiver accept?



Define

A : 8-bit strings starting with 01

B : 8-bit strings ending with 10

General Principle of Counting

If an experiment has r steps, such that

Step i has n_i outcomes for all $i = 1, \ldots, r$,

Then the number of outcomes of the experiment is

$$n_1 \times n_2 \times \cdots \times n_r = \prod_{i=1}^{r} n_i .$$

Multi-step experiment

→ 1 → 2 → ...

Product Rule of Counting: A special case

License plates

How many CA license plates are possible if...

(pre-1982)

(present day)

License plates

How many CA license plates are possible if...

(pre-1982)

(present day)

Pigeonhole Principle

Gradescope quiz, blank slide deck, etc.
http://cs109.stanford.edu/

Floors and ceilings

Floor function

$\lfloor x \rfloor$

The largest integer $\leq x$

Ceiling function

$\lceil x \rceil$

The smallest integer $\geq x$

Check it out:

$\lfloor 1/2 \rfloor$ $\lfloor 2.9 \rfloor$ $\lfloor 8.0 \rfloor$ $\lfloor -1/2 \rfloor$

$\lceil 1/2 \rceil$ $\lceil 2.9 \rceil$ $\lceil 8.0 \rceil$ $\lceil -1/2 \rceil$

Pigeonhole Principle

For positive integers m and n,

if m objects are placed in n buckets,

then at least one bucket must contain at least $\lceil m/n \rceil$ objects.

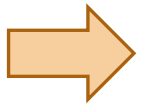
Pigeons in holes

21st century pigeons

Example:

m objects = 10 pigeons

n buckets = 9 pigeonholes



At least one pigeonhole must contain $\lceil m/n \rceil = 2$ pigeons.

Bounds: an important part of CS109

Balls and urns

n balls

r urns
(buckets)

~~Balls and urns~~ Hash Tables and strings

Consider a hash table with 100 buckets.

950 strings are hashed and added to the table.

1. Is it guaranteed that at least one bucket contains ***at least*** 10 entries?

2. Is it guaranteed that at least one bucket contains **at least** 11 entries?

3. Is it possible to have a bucket with ***no entries***?

~~Balls and urns~~ Hash Tables and strings

Consider a hash table with 100 buckets.

950 strings are hashed and added to the table.

$n = 100$

$m = 950$

1. Is it guaranteed that at least one bucket contains ***at least*** 10 entries?

   Yes

2. Is it guaranteed that at least one bucket contains **at least** 11 entries?

   No

3. Is it possible to have a bucket with ***no entries***?

   Sure

Permutations I

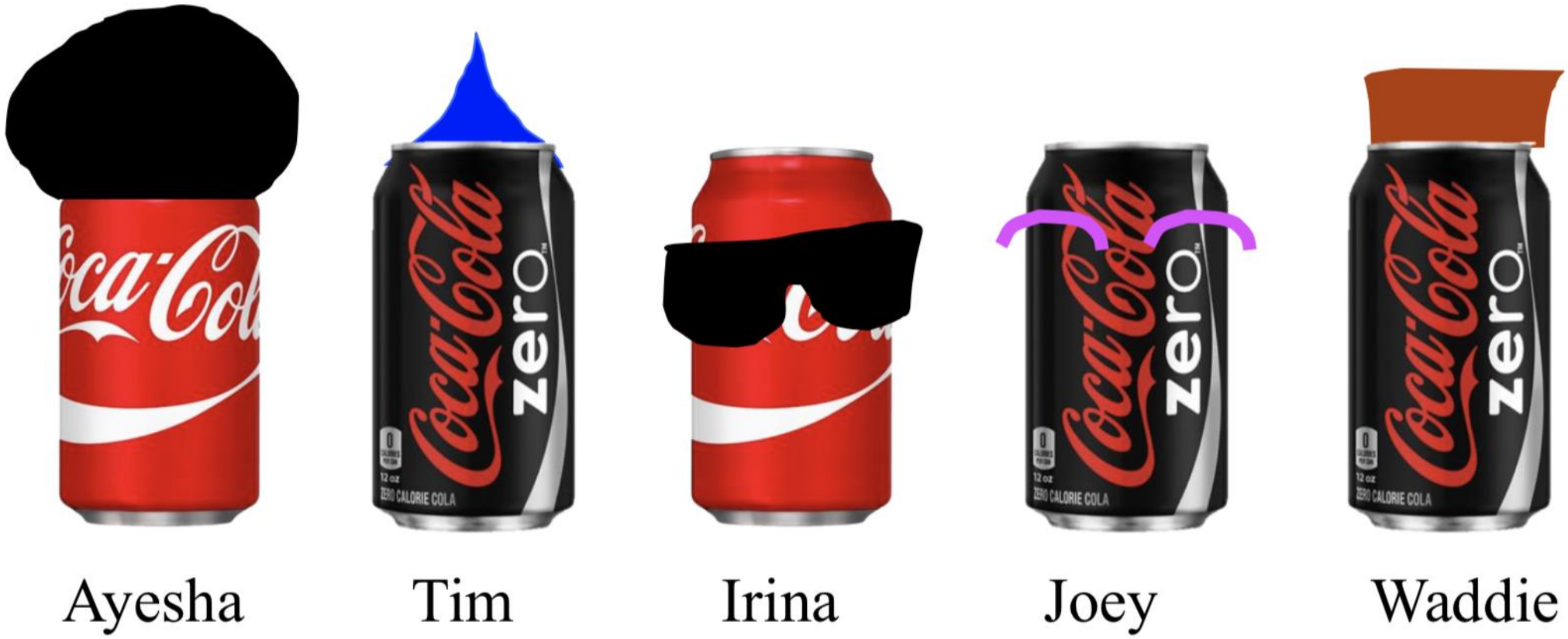
Gradescope quiz, blank slide deck, etc.
http://cs109.stanford.edu/

Unique 6-digit passcodes with six smudges

How many unique 6-digit passcodes are possible if a phone password uses each of **six** distinct numbers?

Sort n indistinct objects

Sort n distinct objects



Sort n distinct objects

Steps:

1. Choose 1st can — 5 options
2. Choose 2nd can — 4 options

...

5. Choose 5th can — 1 option

$$\text{Total} = 5 \times 4 \times 3 \times 2 \times 1$$
$$= 120$$

Permutations

A permutation is an ordered arrangement of objects.

The number of unique orderings (permutations) of n distinct objects is

$$n! = n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1.$$

Unique 6-digit passcodes with six smudges

How many unique 6-digit passcodes are possible if a phone password uses each of **six** distinct numbers?

Total $= 6!$

$= 720$ passcodes

Unique 6-digit passcodes with five smudges

How many unique 6-digit passcodes are possible if a phone password uses each of **five** distinct numbers?