CS109: Probability for Computer Scientists

Lisa Yan
April 6, 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/109/discussion/24490

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)

Lisa Yan

Yes, my undergrad was here…

…But now I'm here!!!

Received PhD 2019
Now:
- Stanford's newest CS lecturer

My interests over time

Networks, Data Science	Create technology	
Teaching	Help people	
Education Tools	Create technology to help people	

Why I like probability

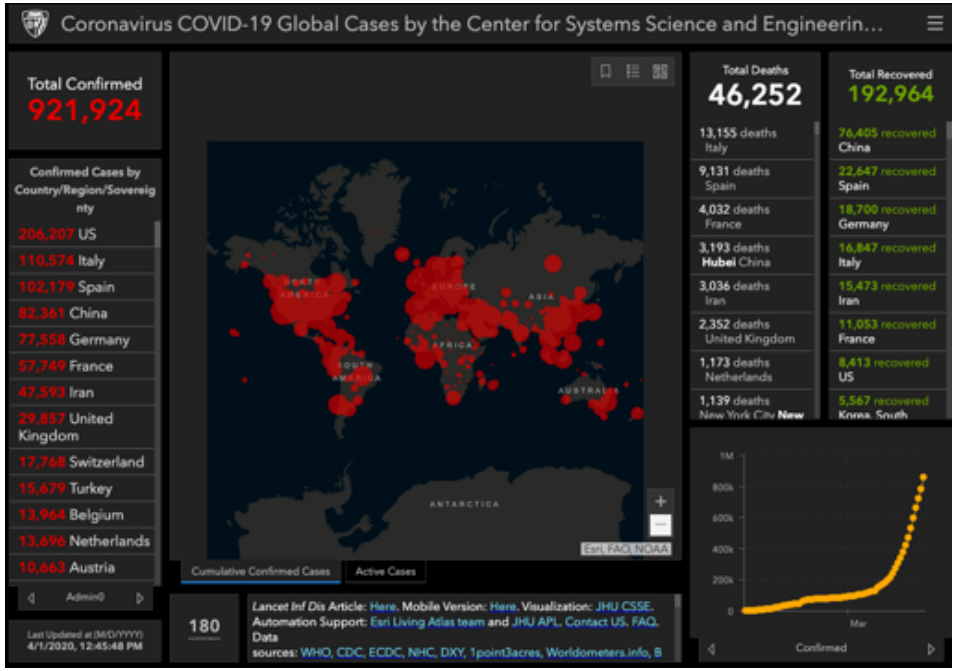
- I like data
- I want to help people
- Probability helps me help people with data
- Also Pokemon

Me, circa 2003

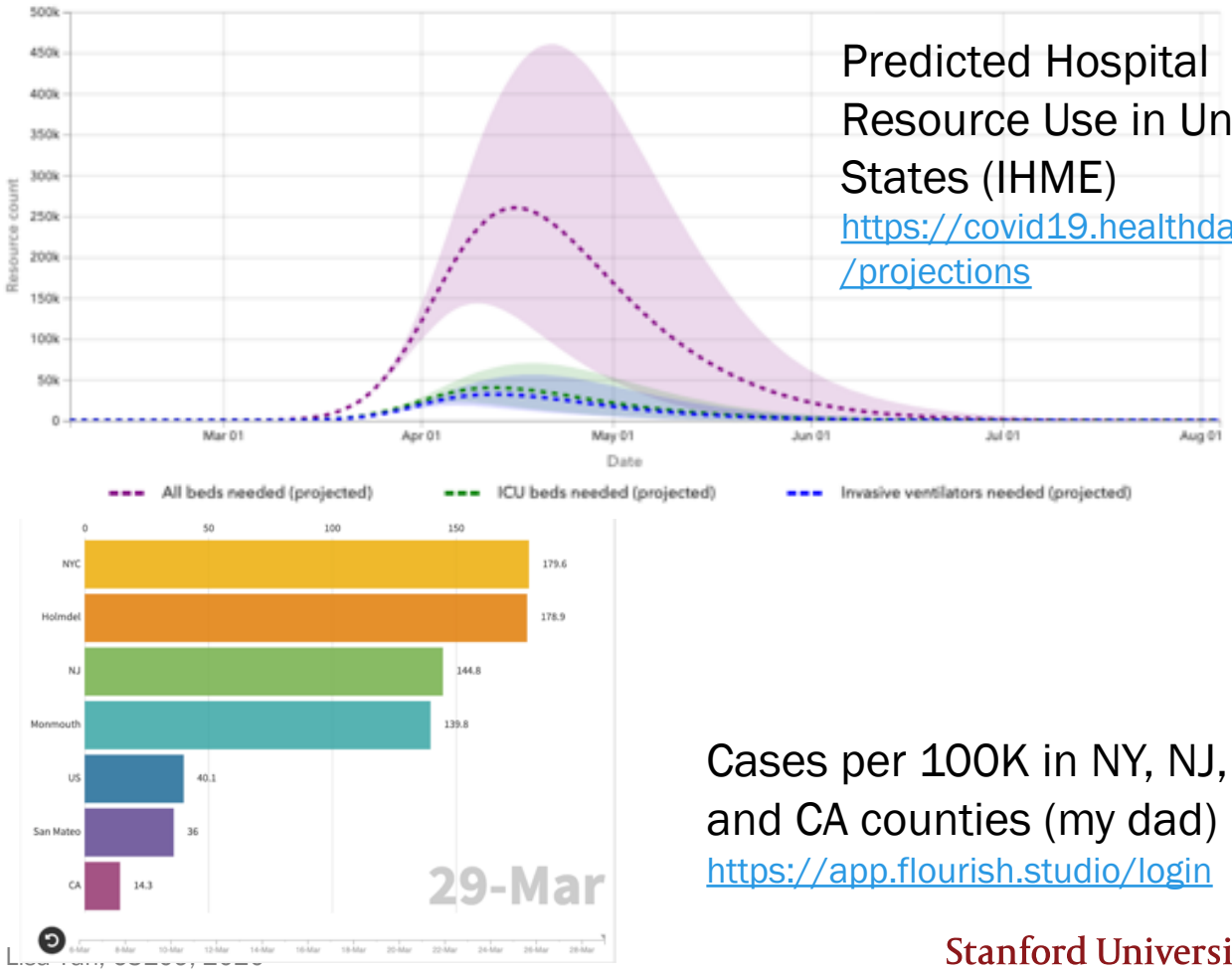
$$a = \frac{(3 \times \mathrm{HP}_{\mathrm{max}} - 2 \times \mathrm{HP}_{\mathrm{current}}) \times \mathrm{rate} \times \mathrm{bonus}_{\mathrm{ball}}}{3 \times \mathrm{HP}_{\mathrm{max}}} \times \mathrm{bonus}_{\mathrm{status}}$$

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**.

126 survey responses

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**.

The S/NC grading guidelines means that you have the freedom to set your own learning goals and **learn for the sake of learning**.

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**.

The S/NC grading guidelines means that you have the freedom to set your own learning goals and **learn for the sake of learning**.

My 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

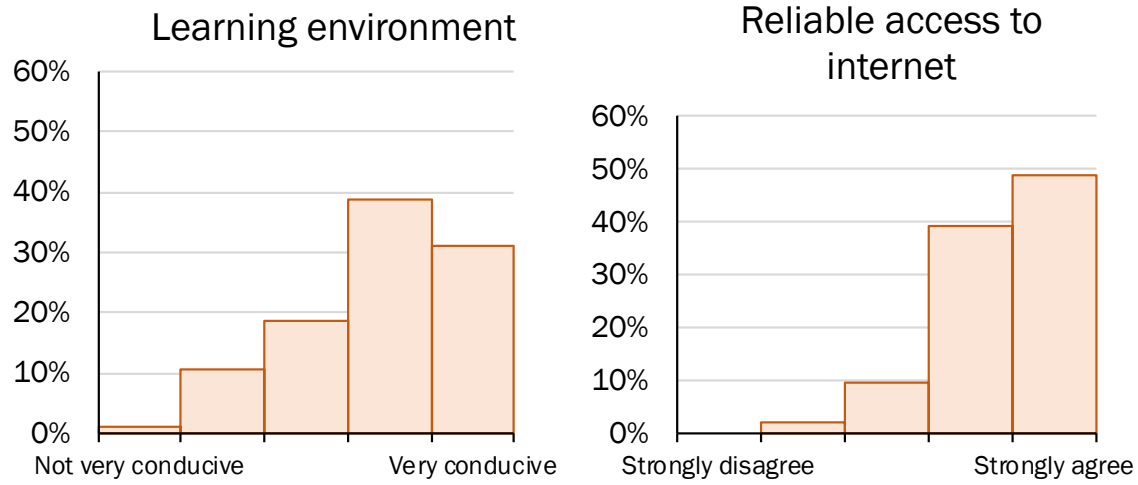
that being said…

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.



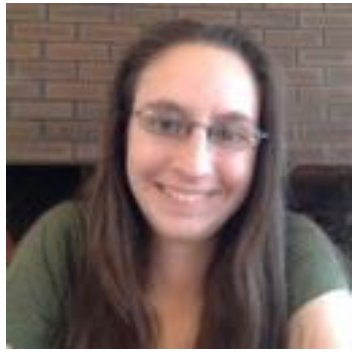
We are committed to working through this version of this course together and adapting as a class and as a community. We welcome your thoughts.

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

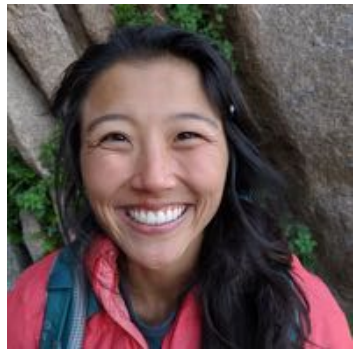
The CS109 teaching team

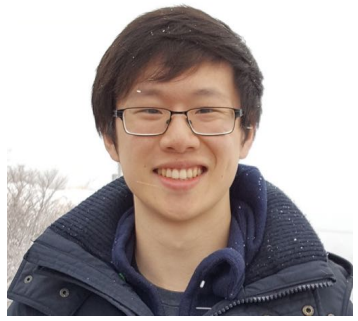











What about you?

...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/BmUE24

Today's discussion thread: https://us.edstem.org/courses/109/discussion/24490

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 and S/NC?

Join discussion forum here:
https://us.edstem.org/join/BmUE24

Today's discussion thread:
https://us.edstem.org/courses/109/discussion/24490

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 and S/NC?

Join discussion forum here:
https://us.edstem.org/join/BmUE24

Today's discussion thread:
https://us.edstem.org/courses/109/discussion/24490

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)

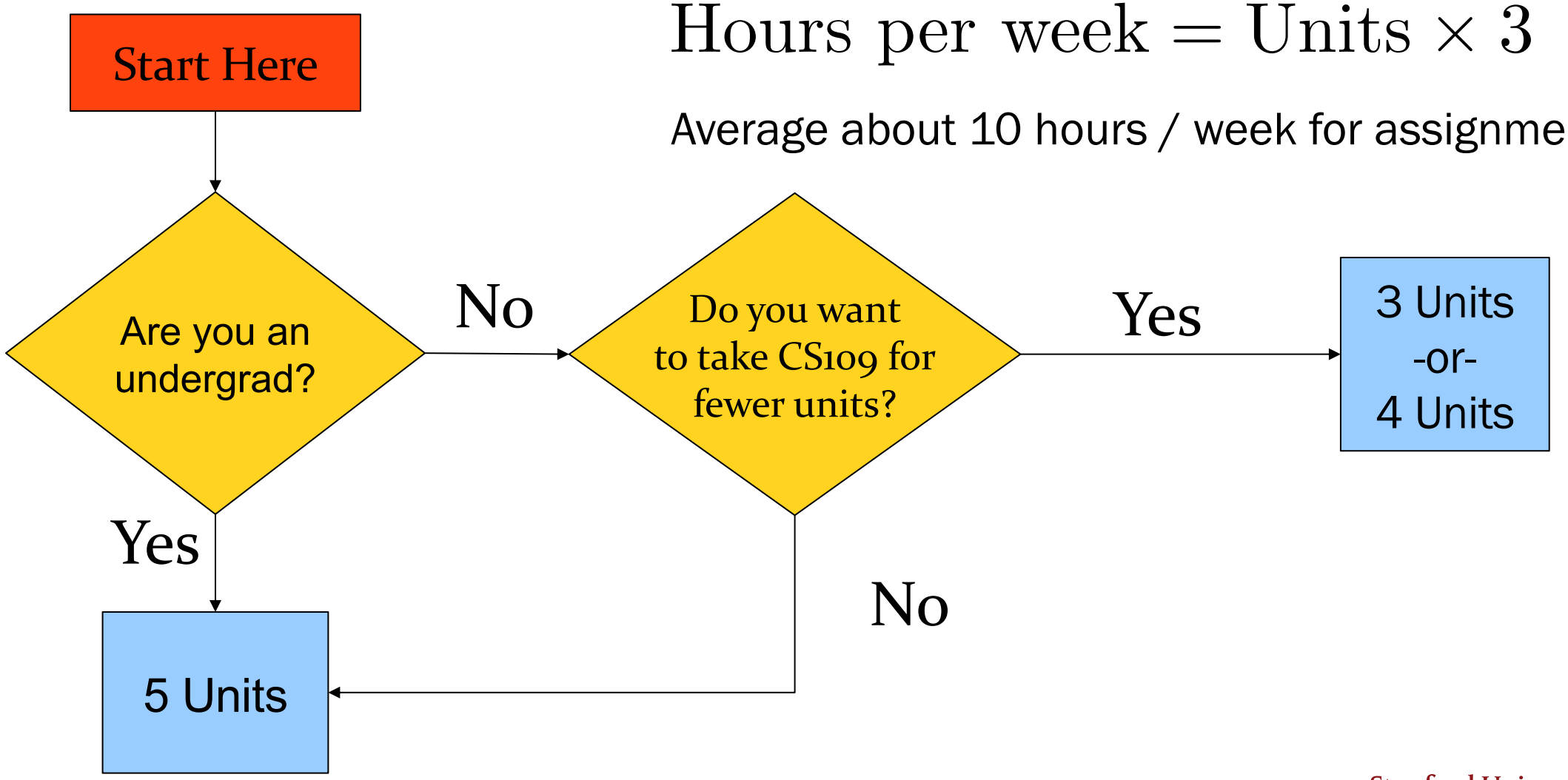
Proofs (induction)
Set theory
Math maturity

Important!

How many units should I take?

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

Average about 10 hours / week for assignments



Will this class count towards my CS degree?

Yes.

"For CS-MS, CS-BS, and CS-Minor students:

All classes taken Spring quarter will satisfy requirements as if taken for a letter grade. This applies to CS-MS requirements, CS-BS requirements, CS-Minor requirements, and the SoE requirements for the CS major."

If you are an undergraduate, you still must take this course for 5 units.

Staff contact

- Discussion forum: https://us.edstem.org/courses/109/discussion/
- Staff email cs109@cs.stanford.edu
- Working office hours: For all timezones (starting later this week)
- 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
(part of grade, submit infinitely many times)

50-min in-person, discussion-oriented lecture
MWF 10:30am-**11:20am** PT
(note 50 min, not 80 min)

Where you learn

Pre-recorded lectures

Live lectures recordings posted to Canvas

Discussion Section starting Week 2

Lecture notes on website

Textbook readings optional

⇒ Administrivia

Problem Sets 6

Quizzes 2

Optional, open-ended contest 1

S/NC Class breakdown

60% 6 Problem Sets

25% Quizzes

- Thursday, April 30
- Thursday, May 20

15% Participation

- Concept checks on pre-recorded material
- Section participation (alternatives provided)

60% Problem Sets

"Passing work" 60% on **each** problem set

95%

Late Policy

+5% grade	for on-time submission	F 99.75%
+0%	bonus for 1 **class day** late	M 95%
cap 80%	for 2 **class days** late	W 80%
cap 60%	for 3 **class days** (1 week) late	F 60%

Review session #1 this Friday 4/10 (time TBA)

Optional but encouraged, tutorial online TBD

Quizzes, Participation

Quizzes

- Ideally, 1-2 hours of individual work
- 24-hour take-home window

2:

Participation (full policy on website)

1. (10%) Concept checks: Submit for pre-lecture recording, unlimited submissions/autograder before each lecture
2. (5%) Section participation

CS109 Contest

- Announced mid-quarter
- A meaningful submission will replace your section grade, stronger submissions replace problem sets for passing work

Your baseline is CS109, and the sky is the limit.

Previous winning submissions:

- Recidivism Risk: Algorithmic Prediction and Racial Bias
- A Better Way to Reform the Electoral College
- Monte Carlo Tree Search for Tic Tac Toe

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

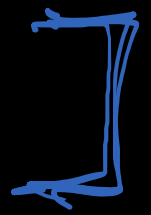
http://www.site.com

Machine Learning

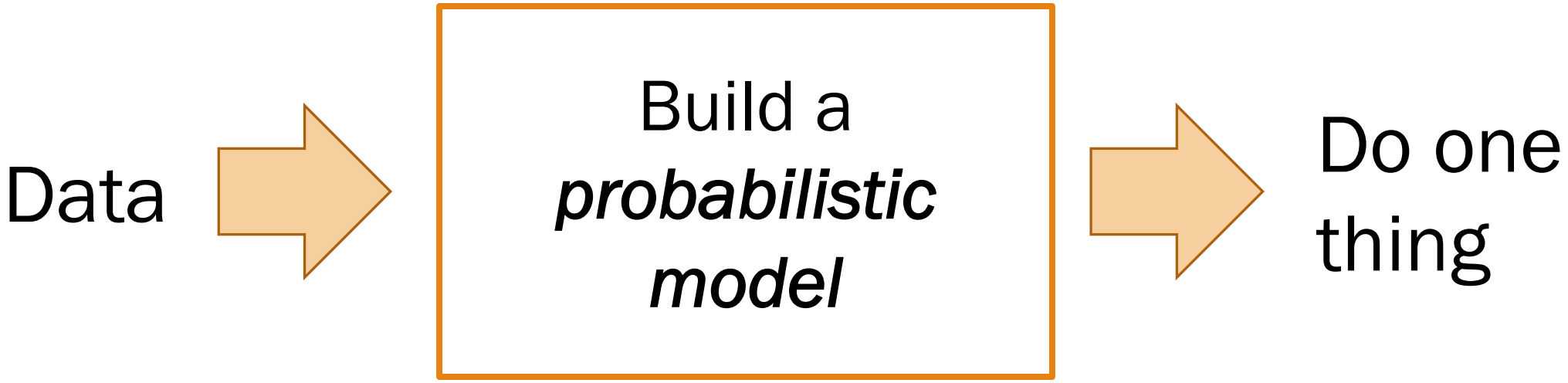
= Machine (compute power)

+ Probability

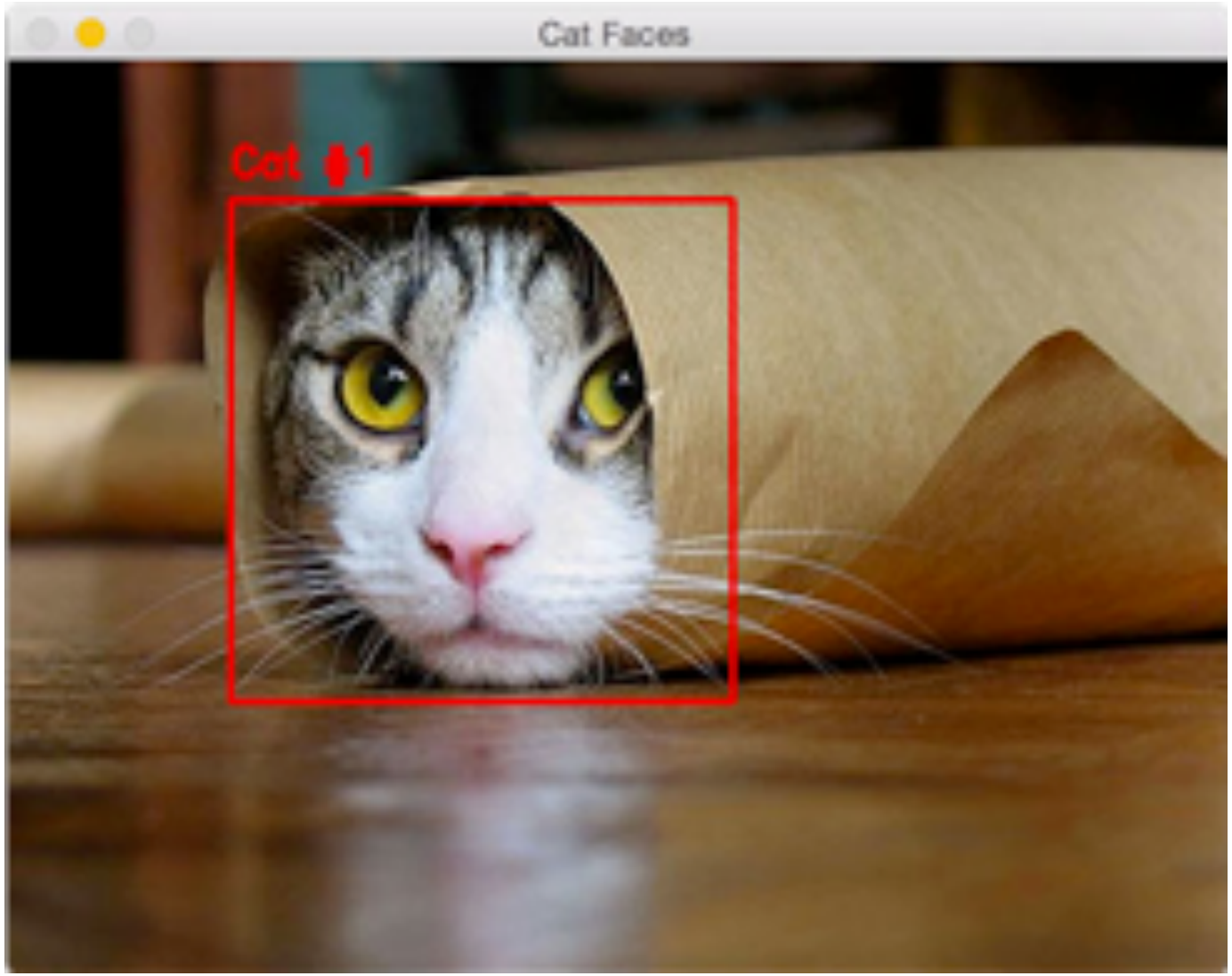
+ Data



Machine Learning Algorithm



Classification



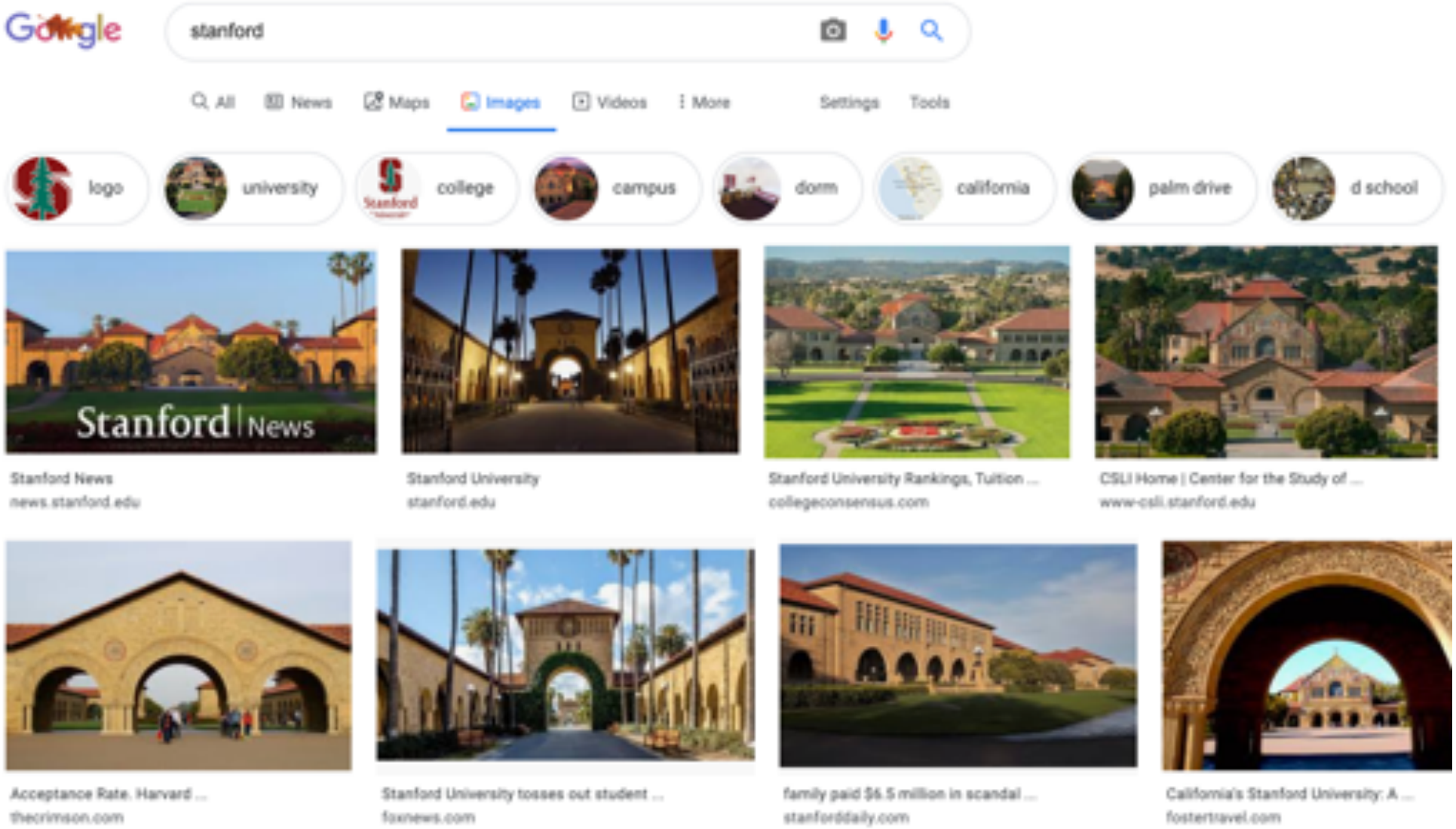
Where is this useful?

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

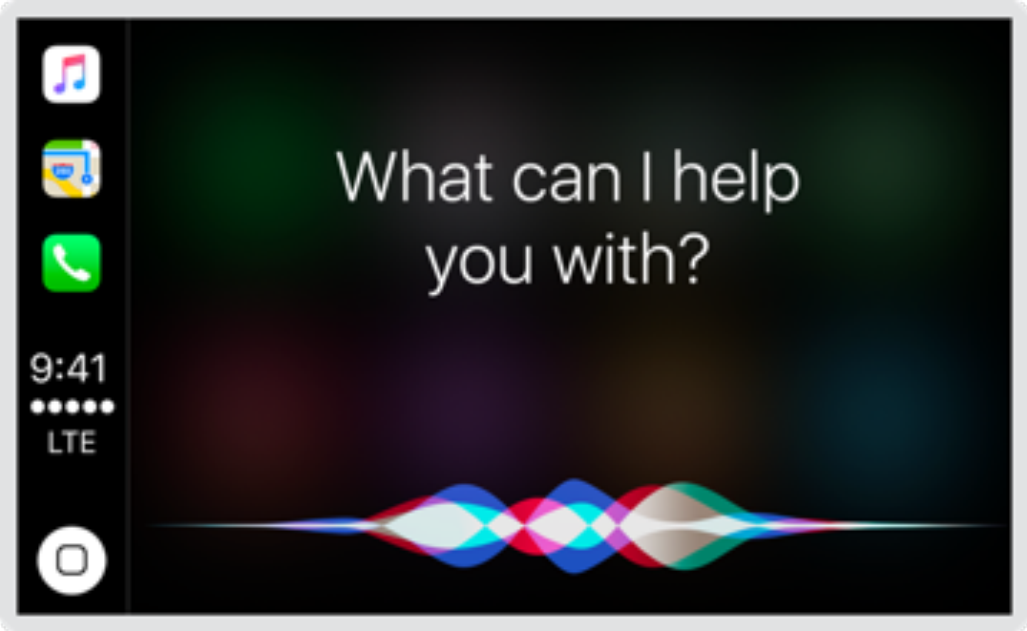


Decision-making: The last remaining board game

Augmented Reality Machine Translation

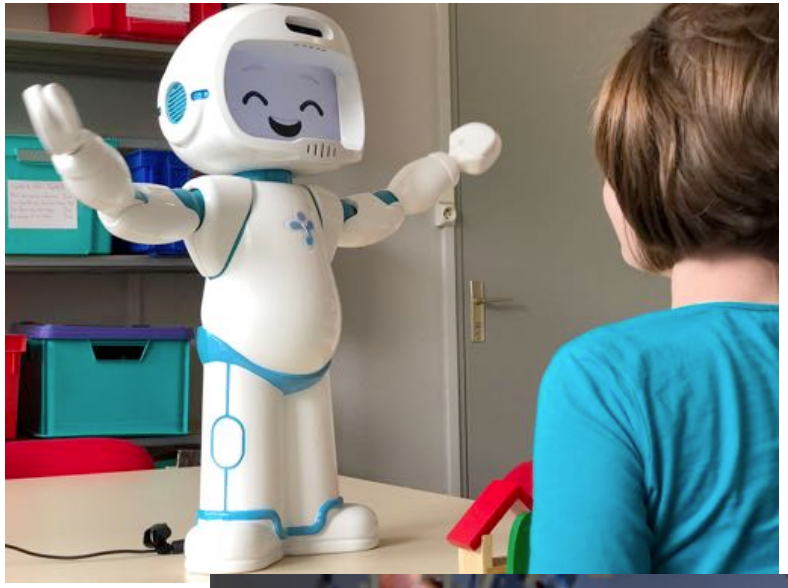
Automatic machine translation on Google Translate

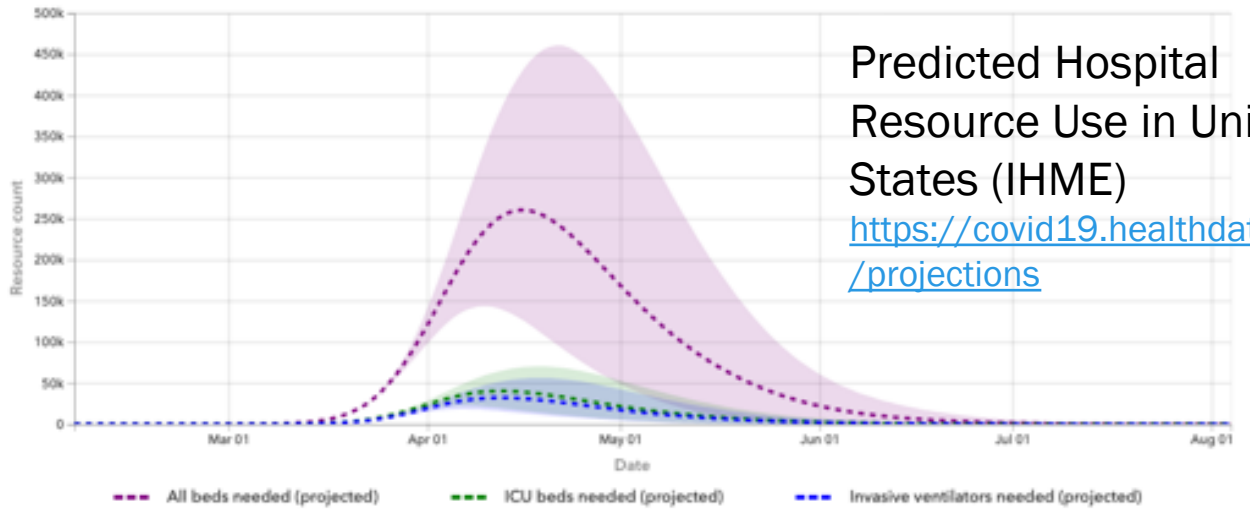
Voice assistants



Probability is *more* than just machine learning.

Probability and medicine



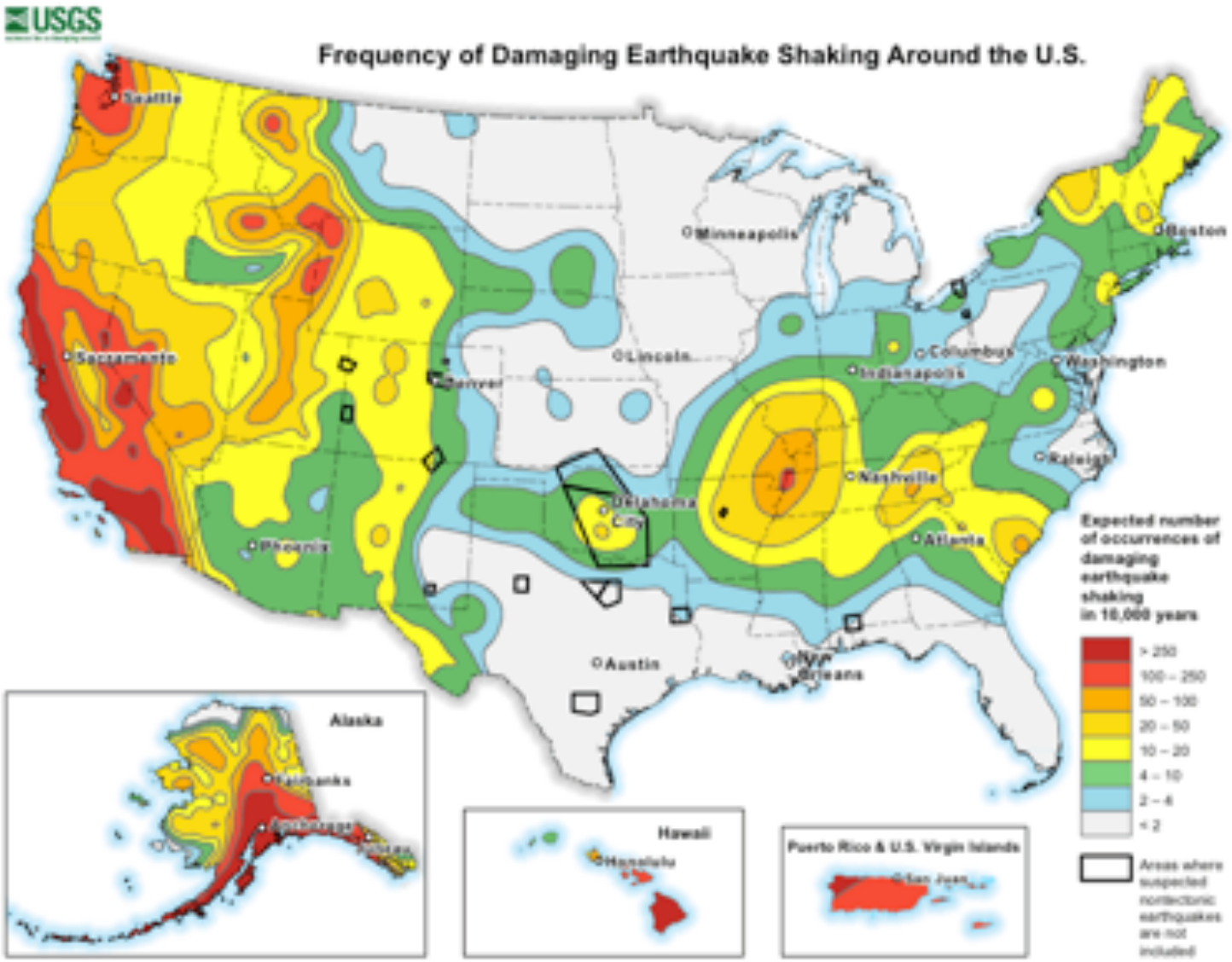


Predicted Hospital Resource Use in United States (IHME) https://covid19.healthdata.org/projections

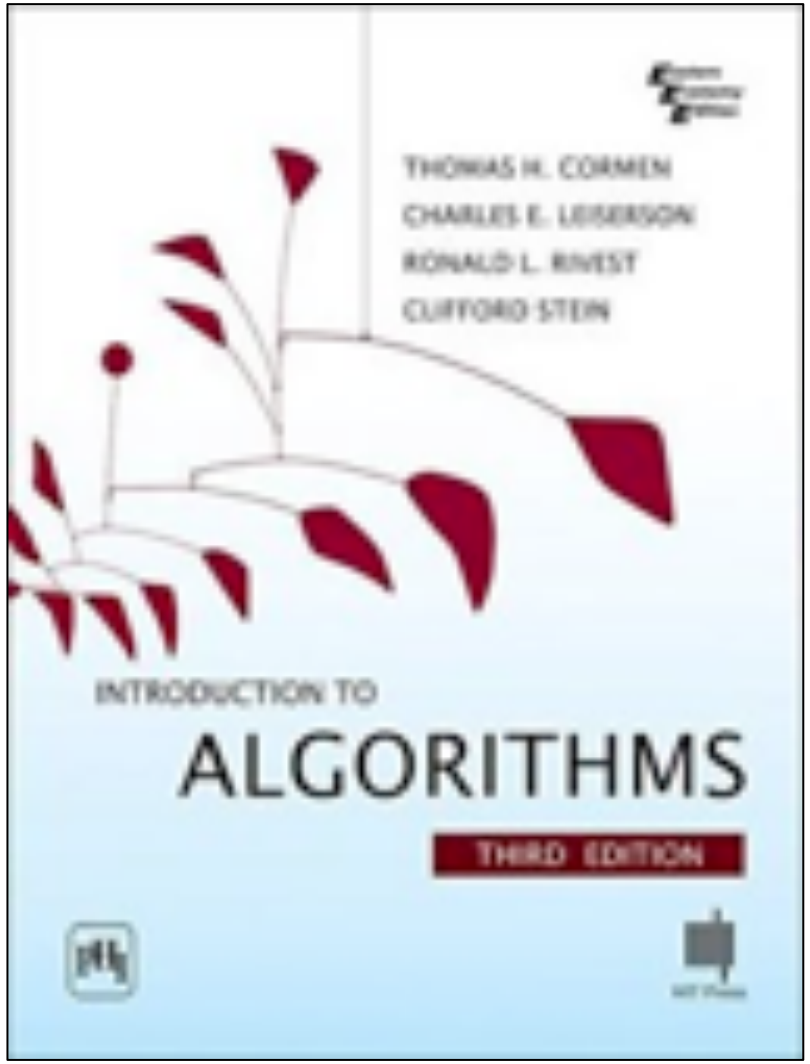
How do COVID-19 testing rates in a region correlate with the actual spread of the disease?

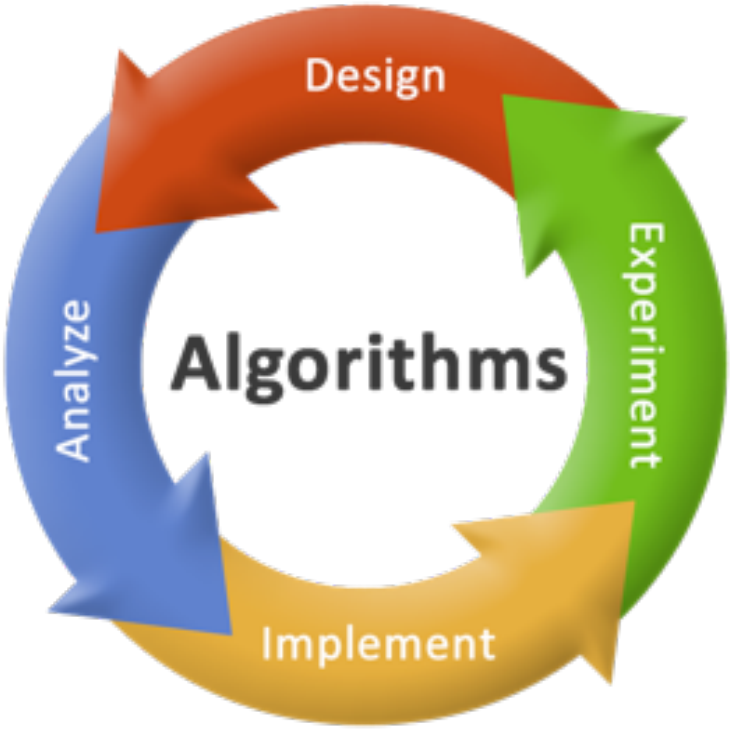
Probability and art

Probability and climate

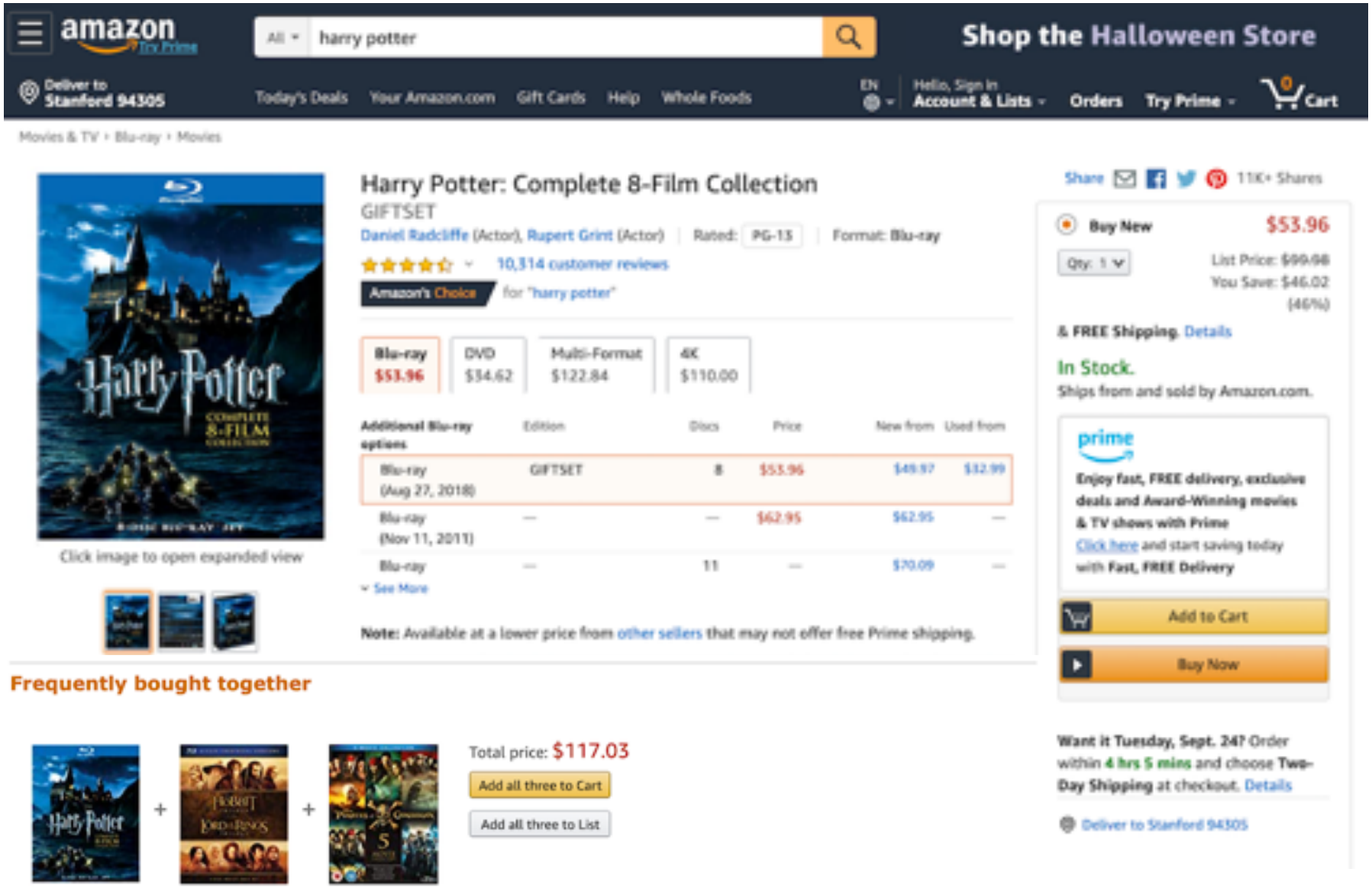


Probabilistic analysis of algorithms

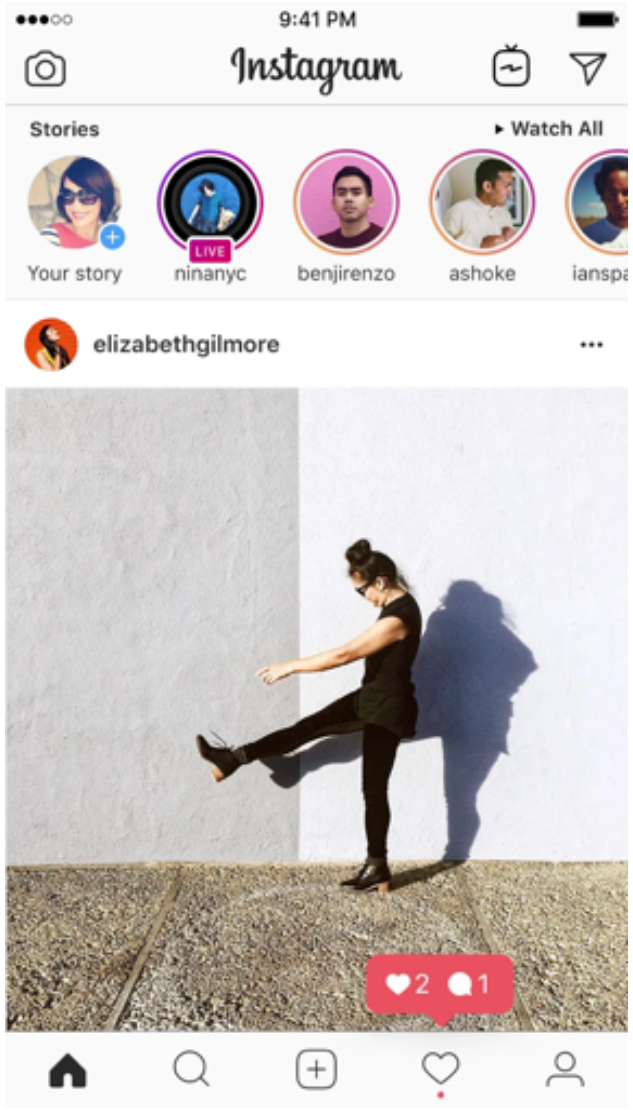




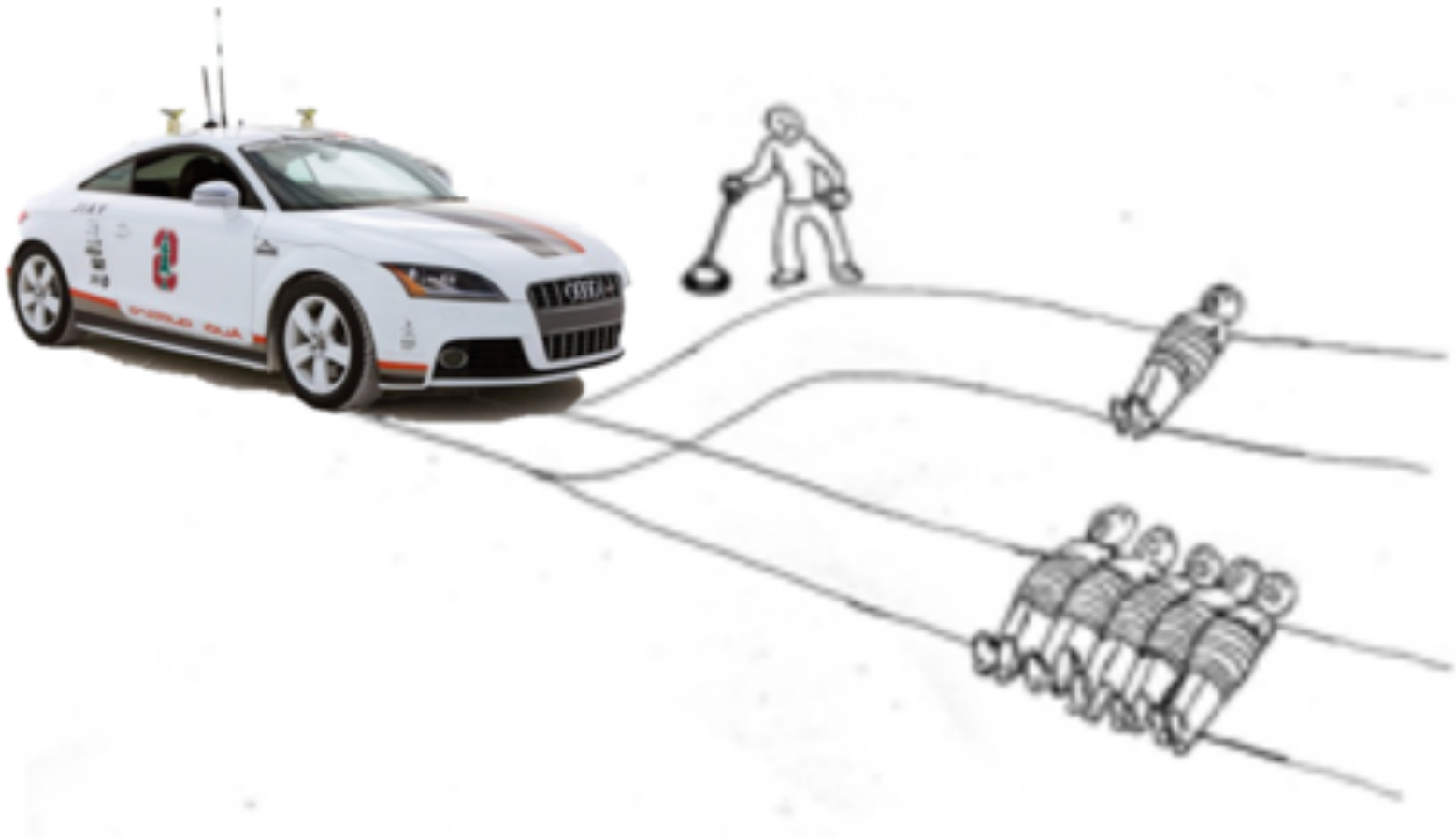
Probability in practice



Probability at your fingertips

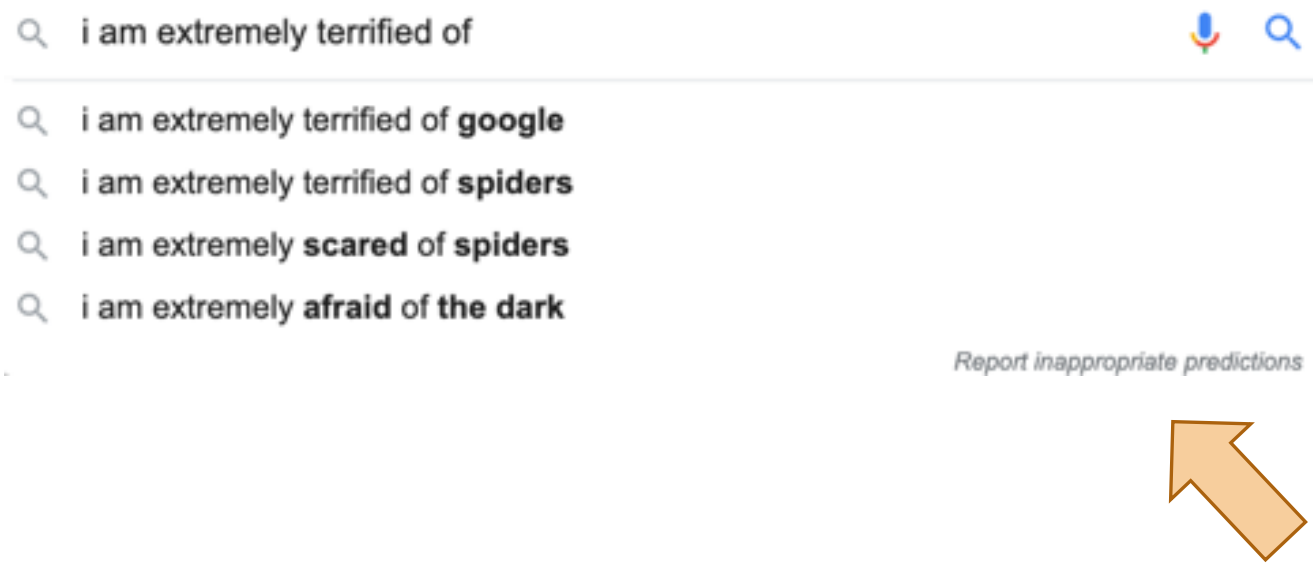


Probability and philosophy



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?

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

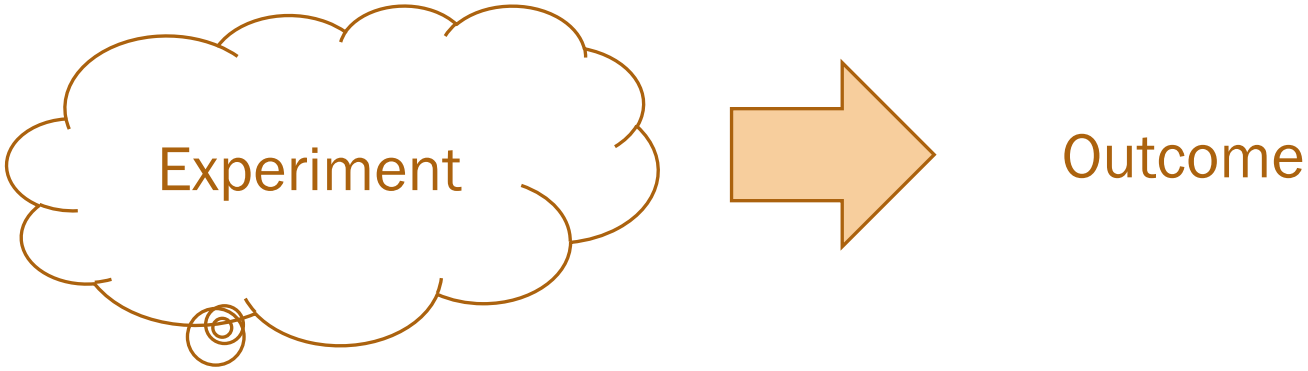
Correct answer: 0.42%

Probability = Important
+ Needs Studying

Counting I

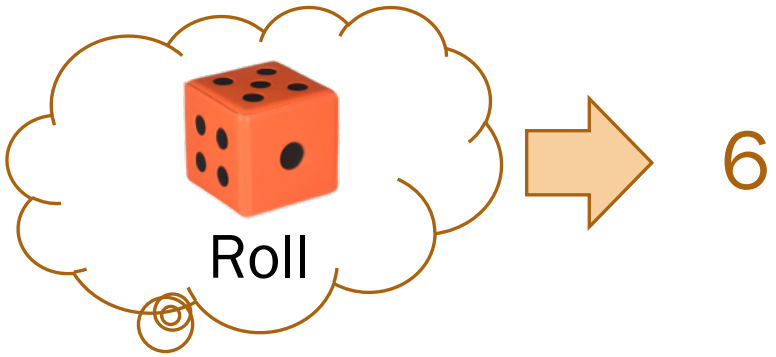
What is Counting?

An experiment in probability:



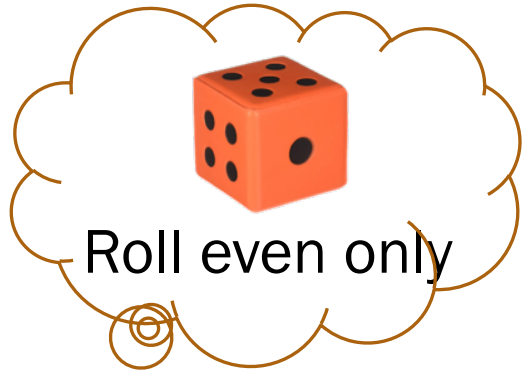
Counting: How many possible outcomes can occur from performing this experiment?

What is Counting?



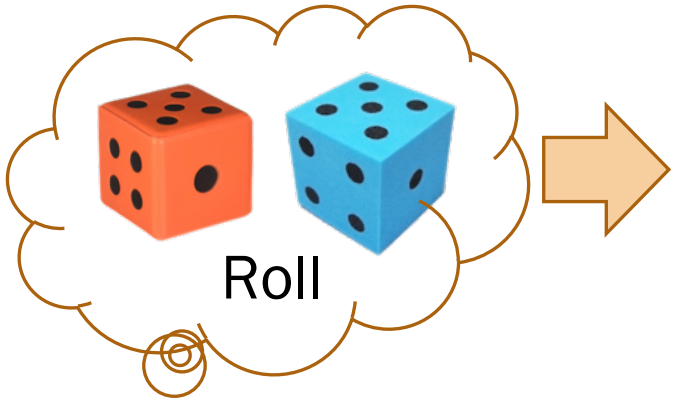
6

{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 $A \cap B = \emptyset$,

$A = \{1, 3, 5\}$

$B = \{2, 4, 6\}$

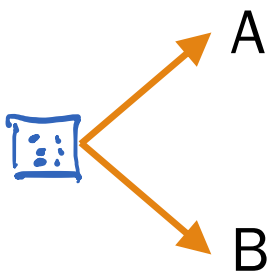
$\cap$ ↑ intersection

Then the number of outcomes of the experiment is

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

$3 + 3 = 6$

One experiment



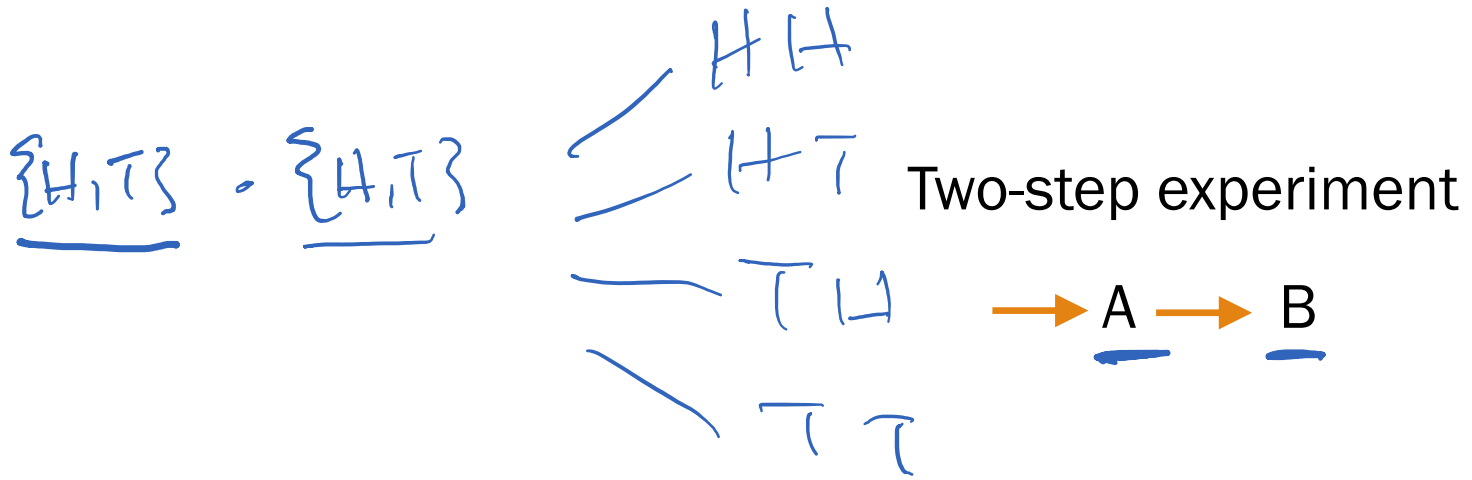
Product Rule of Counting

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 $|B| = n$,

Then the number of outcomes of the experiment is

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

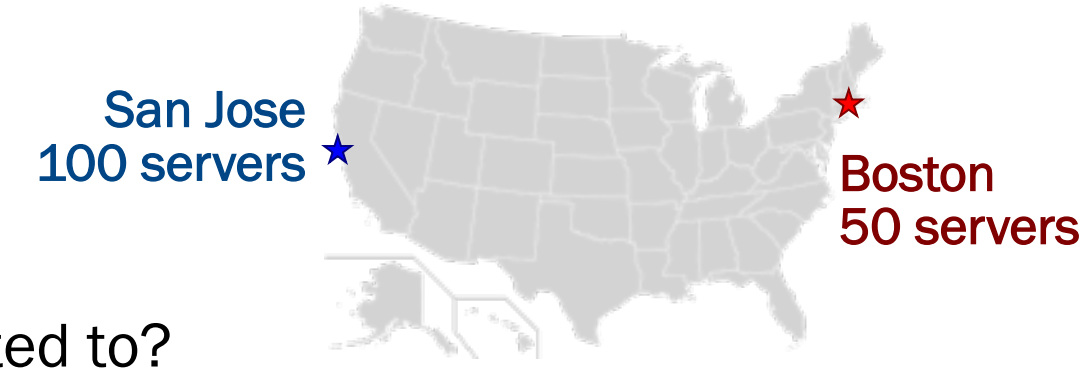


Let's try it out

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?

request → SJ, A, $|A| = 100$

request → BOS, B, $|B| = 50$

$A \cap B = \phi$

Sum Rule: $|A| + |B| = \boxed{150}$

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

$\boxed{36} = \underset{A}{6} \cdot \underset{B}{6}$ Product Rule

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

$\frac{4!}{2!} = \boxed{12 \text{ outcomes}}$

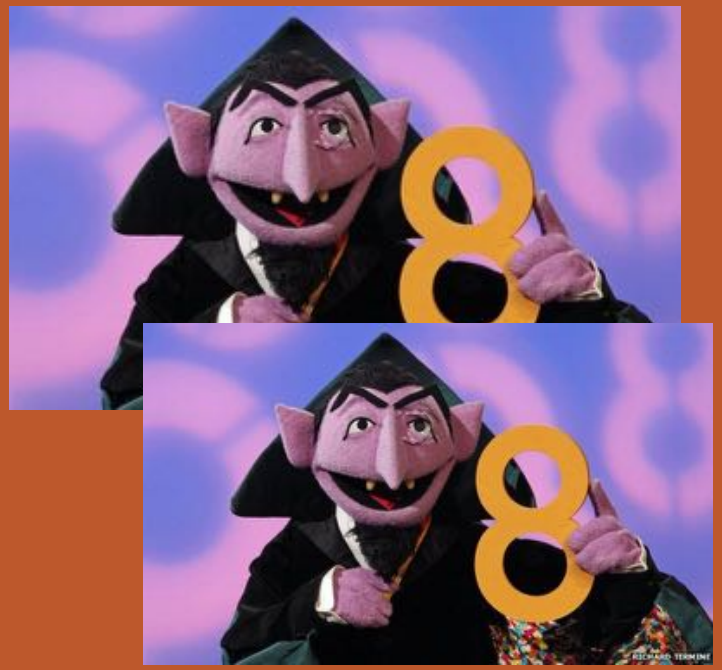
For next time

- Watch pre-recorded lectures for today (Monday 4/6) and Wednesday 4/8 to be posted this afternoon PT
- Complete one concept check that covers both lectures to be posted this afternoon PT

http://cs109.stanford.edu/

Lisa's office hours

Questions?



Counting II

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

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

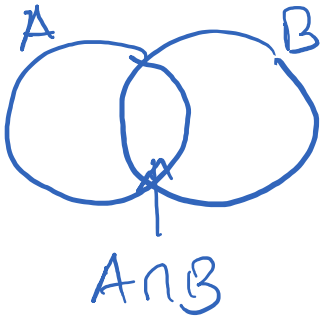
If the outcome of an experiment can be either from

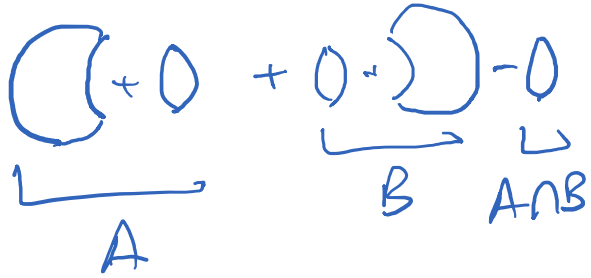
Set A or set B,

where A and B may overlap,

Then the total number of outcomes of the experiment is

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

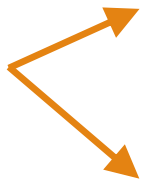




if $A \cap B = \phi$, $|\phi| = 0$

$|A \cup B| = |A| + |B| - 0$

One experiment



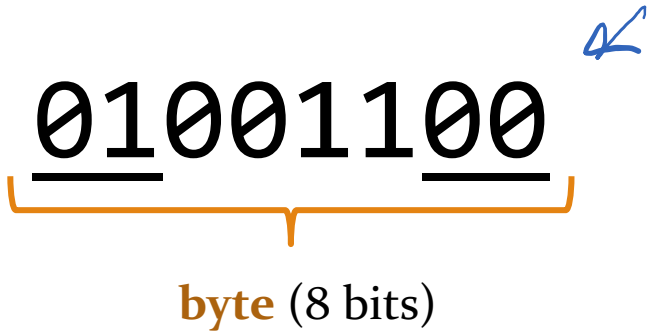
Sum Rule of Counting:
A special case

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

$0\ 1\ *\ *\ *\ *\ *\ *$ $\quad |A| = 2^6$

$*\ *\ *\ *\ *\ *\ 1\ 0$ $\quad |B| = 2^6$

$A \cap B$: $0\ 1\ *\ *\ *\ *\ 1\ 0$ $\quad |A \cap B| = 2^4$

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

$$= 2^6 + 2^6 - 2^4$$

$$= 112$$

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

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

Summation

$$x_1 + x_2 + \ldots + x_n = \sum_{i=1}^{n} x_i$$

Multi-step experiment

→ 1 → 2 → ...

Product Rule of Counting: A special case

product notation

step 1 2 3 ... r

n_1 n_2 ...

r = 2-part expt

step 1's outcomes are in A

step 2's outcomes are in B

$|A| |B|$,

n_1 n_2

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)

$$\underbrace{26}_{A\text{-}Z} \cdot \underbrace{26}_{A\text{-}Z} \cdot \underbrace{26}_{A\text{-}Z} \cdot \underbrace{10}_{0\text{-}9} \cdot \underbrace{10}_{0\text{-}9} \cdot \underbrace{10}_{0\text{-}9} = 17{,}576{,}000$$

(present day)

Soln 1 : $10 \cdot 26 \cdot 26 \cdot 26 \cdot 10 \cdot 10 \cdot 10$

$175{,}760{,}000$

Soln 2 : $\underbrace{10}_{0\text{-}9} \cdot \underbrace{17{,}576{,}000}_{\text{6-place plate expt}}$

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:

$\lfloor 1/2 \rfloor = 0$ $\lfloor 2.9 \rfloor = 2$ $\lfloor 8.0 \rfloor = 8$ $\lfloor -1/2 \rfloor = -1$

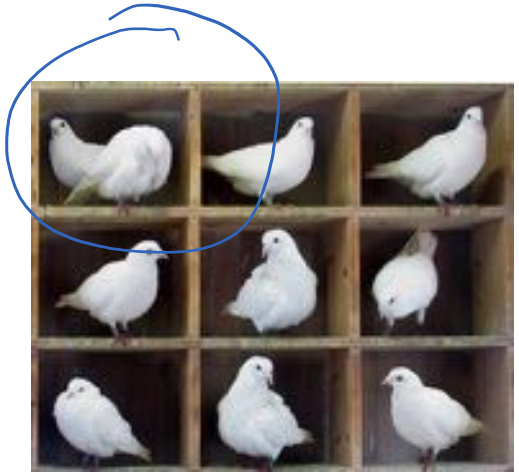
$\lceil 1/2 \rceil = 1$ $\lceil 2.9 \rceil = 3$ $\lceil 8.0 \rceil = 8$ $\lceil -1/2 \rceil = 0$

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.



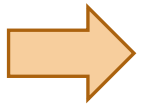
Pigeons in holes

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.

$n = 100$ buckets

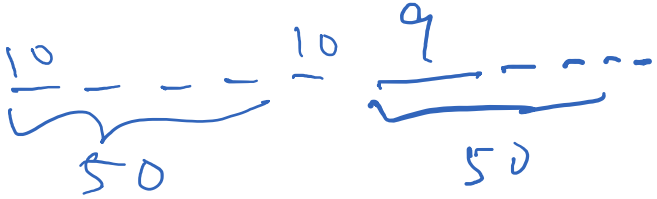
$m = 950$ strings

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

Yes $\lceil \frac{950}{100} \rceil = \lceil 9.5 \rceil = 10$ entries

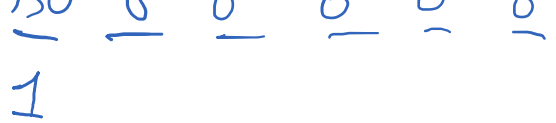
2. Is it guaranteed that at least one bucket contains **at least** 11 entries?

No



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

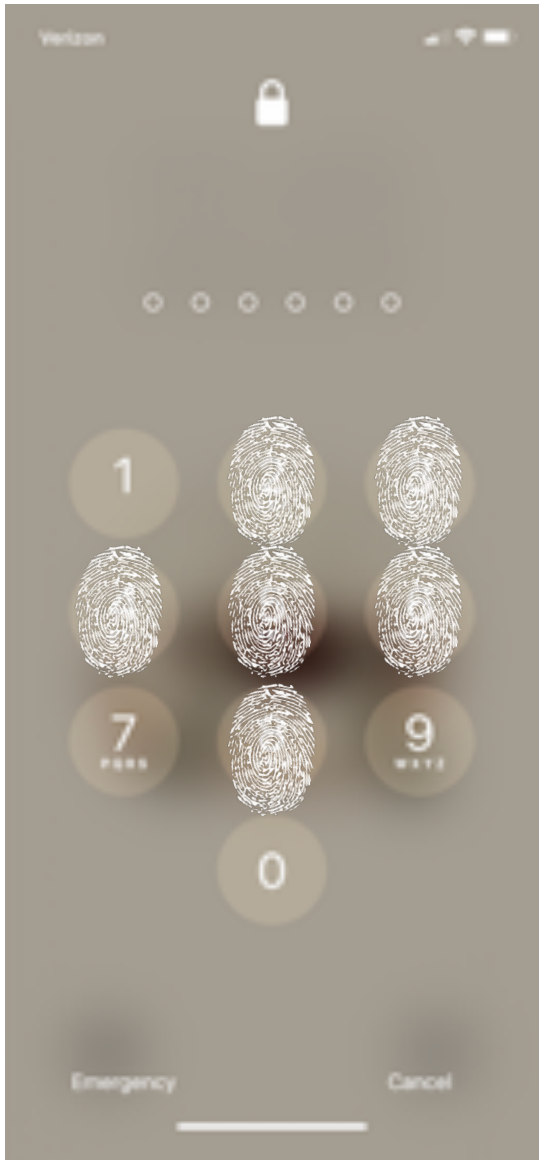
Sure



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

1st 2nd 3rd 4th 5th

options 5 · 4 · 3 · 2 · 1

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.

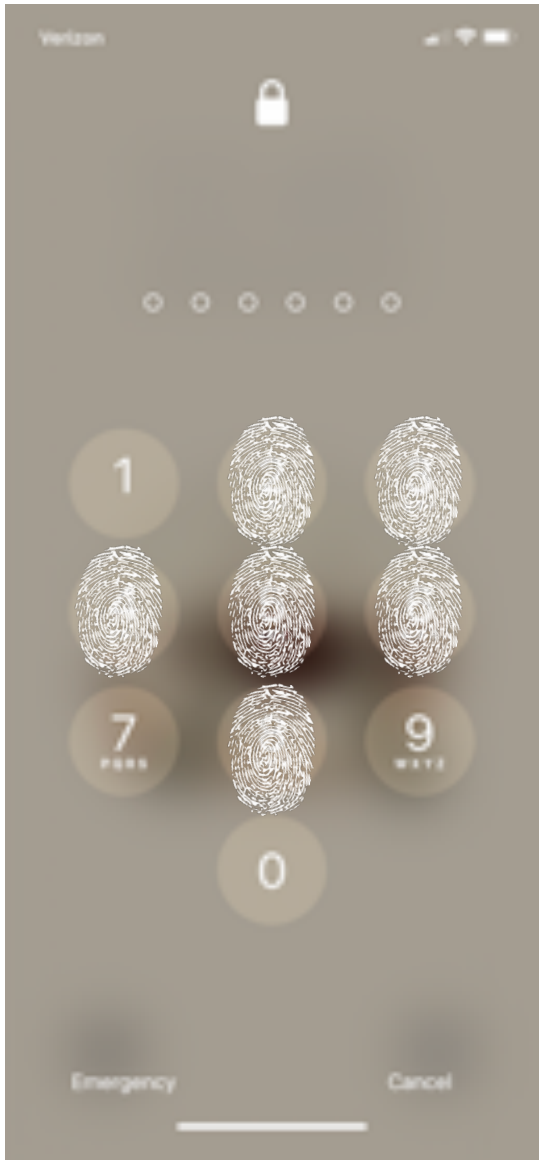
Sorted

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

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

$$= \prod_{i=1}^{n} i$$

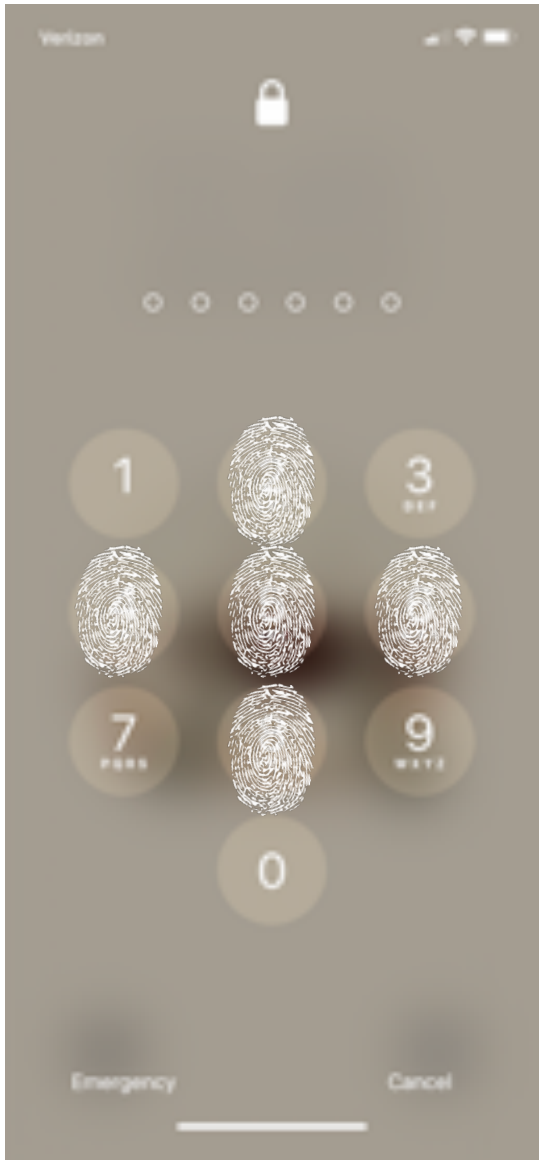
Unique 6-digit passcodes with six smudges



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

$$\text{Total} = 6!$$
$$= 720 \text{ passcodes}$$

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?