

Bradley Voytek, Ph.D.
UC San Diego

Department of Cognitive Science
Halıcıoğlu Data Science Institute
Neurosciences Graduate Program

bvoytek@ucsd.edu
voyteklab.com

UC San Diego

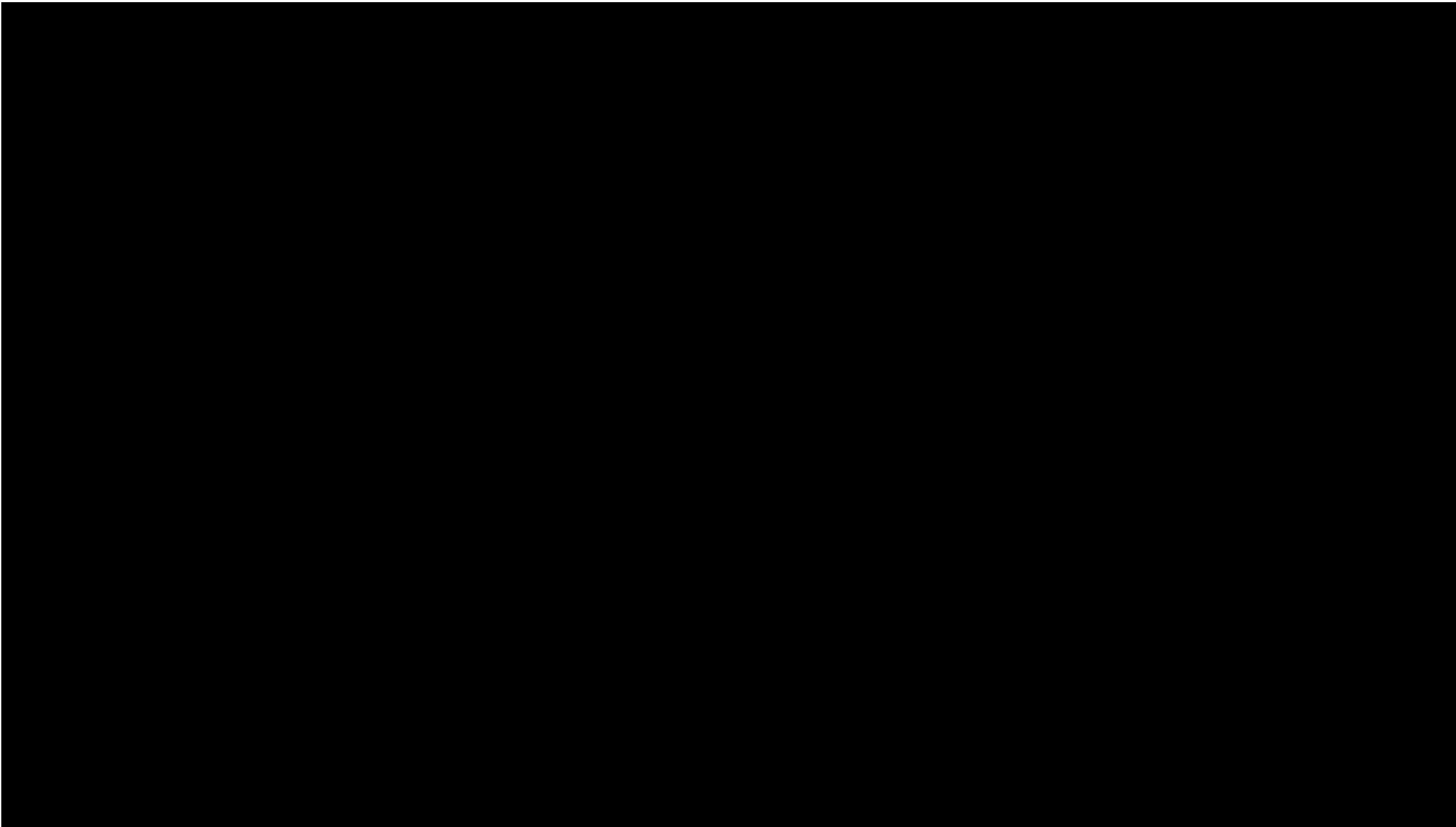
COGS 9
Introduction to Data Science

Data Storytelling

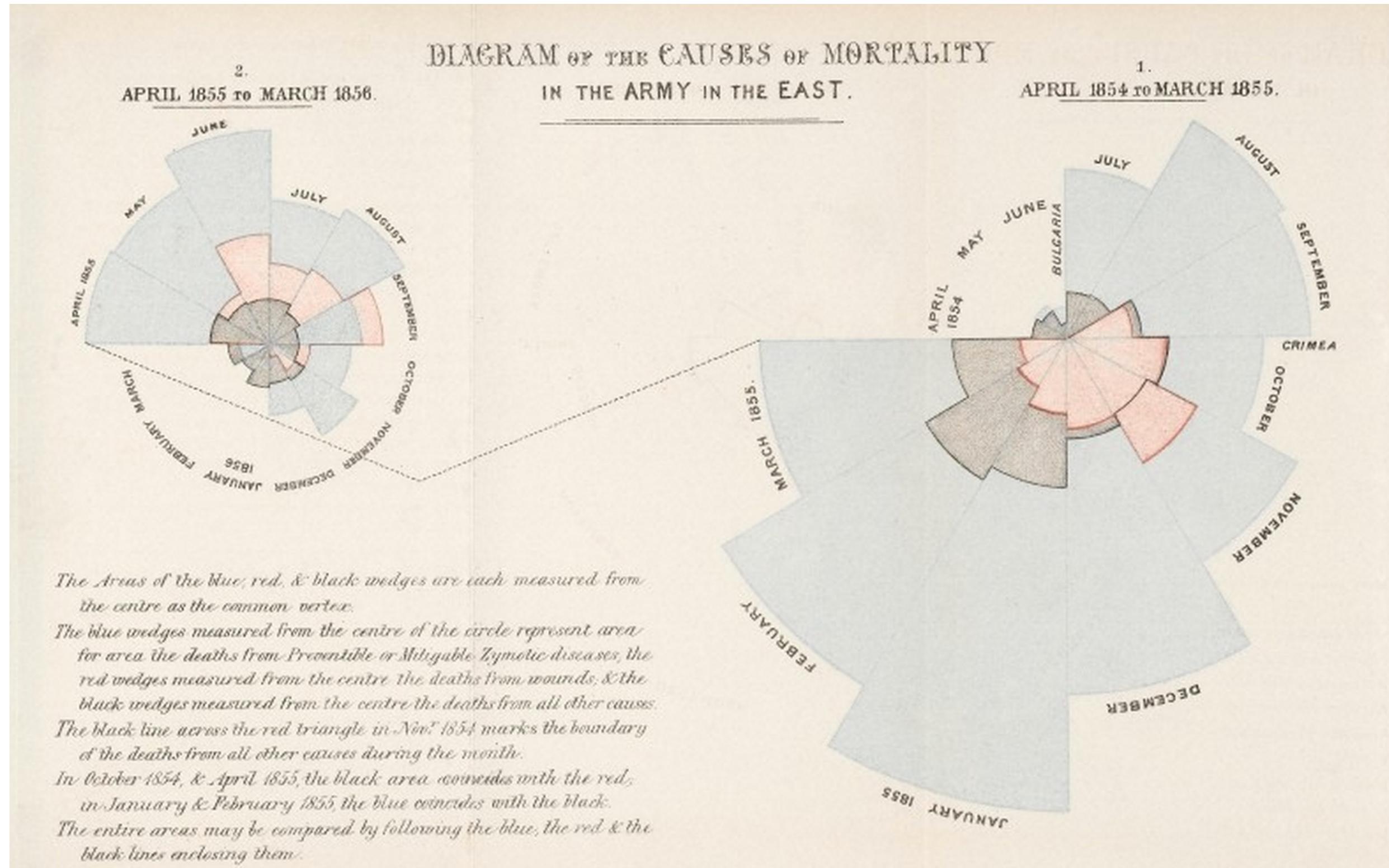
Today's Learning Objective

Why and how do we tell effective stories using data?

Intro

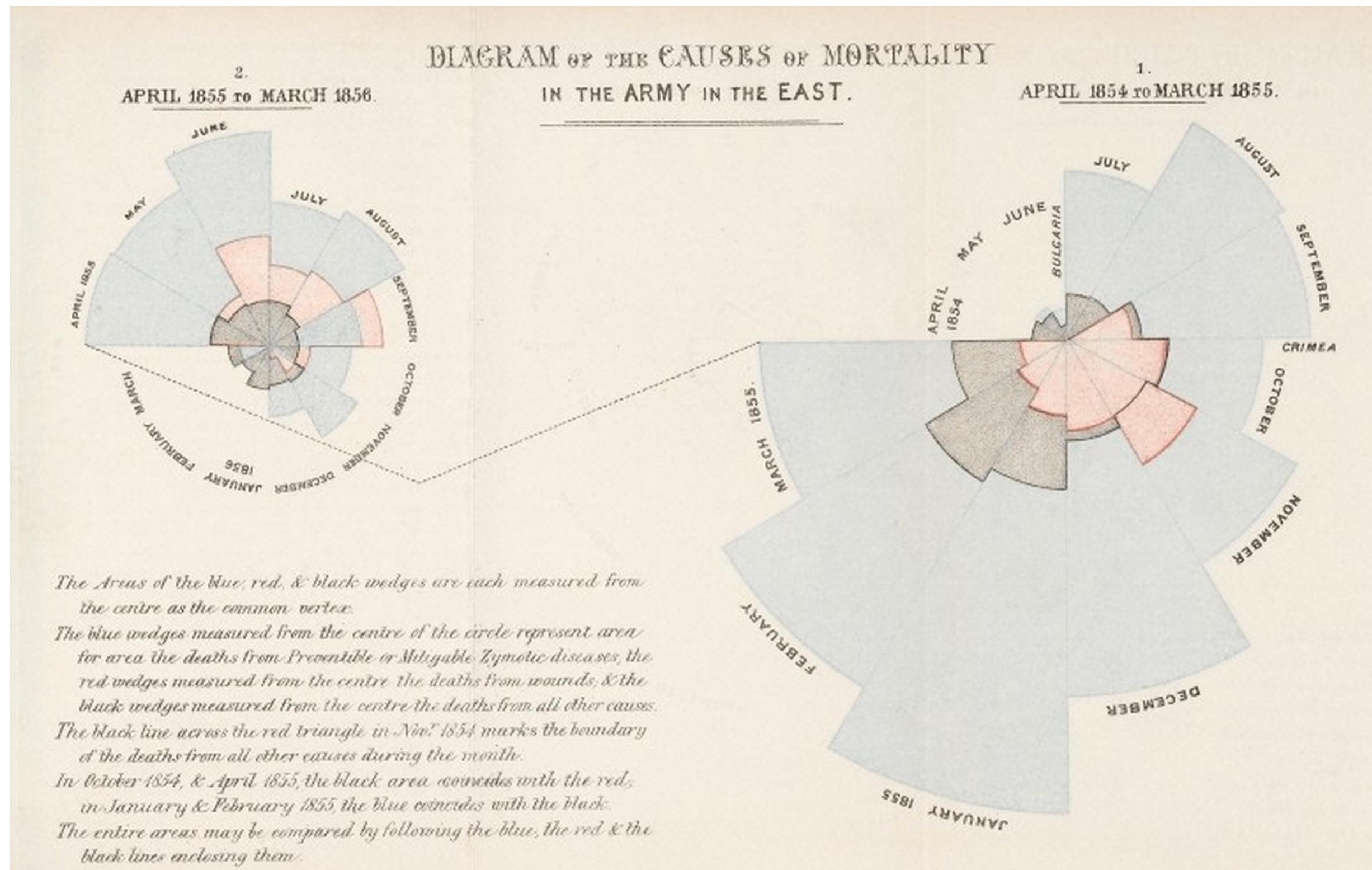


Florence Nightingale



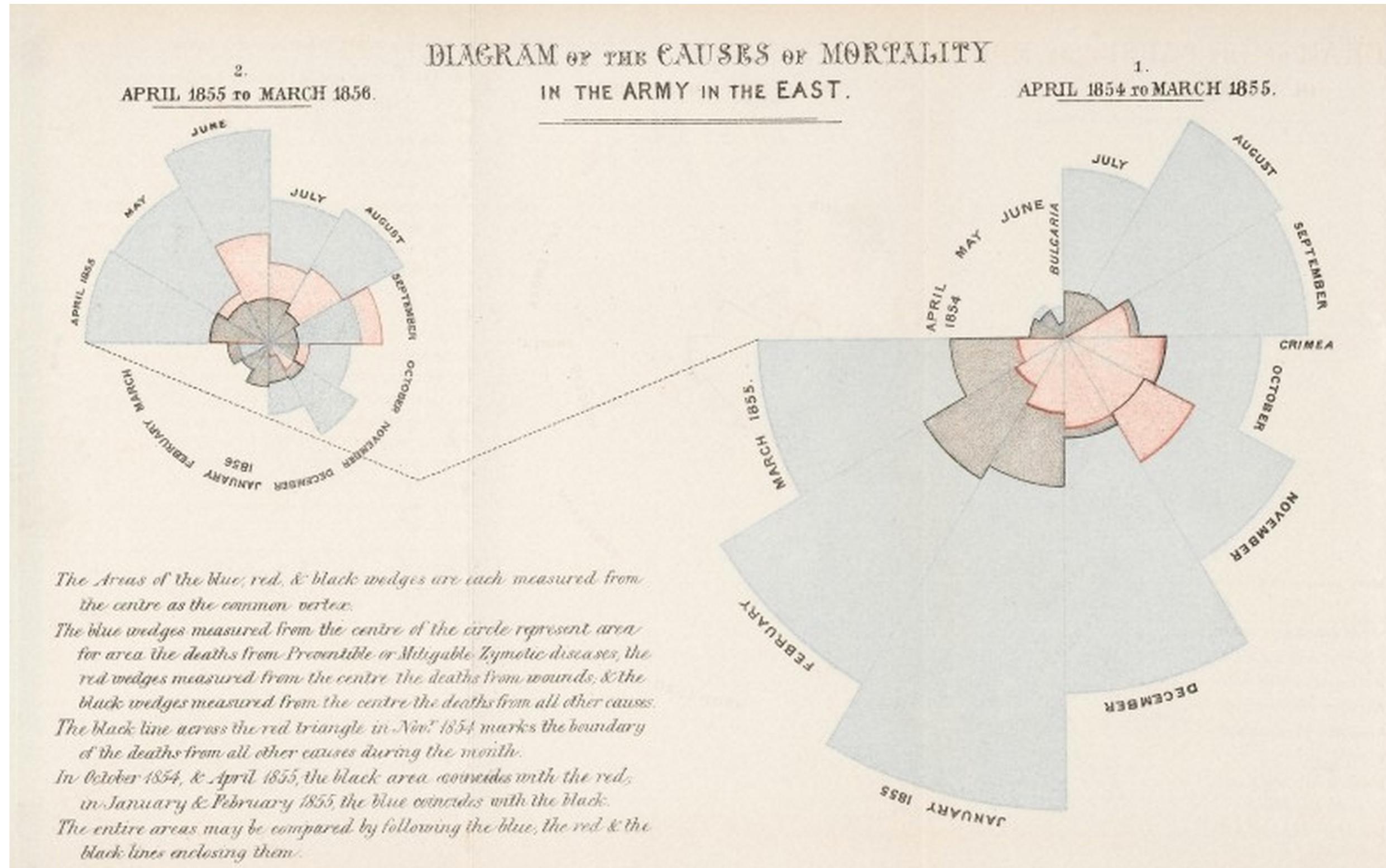
- founded nursing
- treated soldiers in Crimean War

Florence Nightingale



red: death from wounds

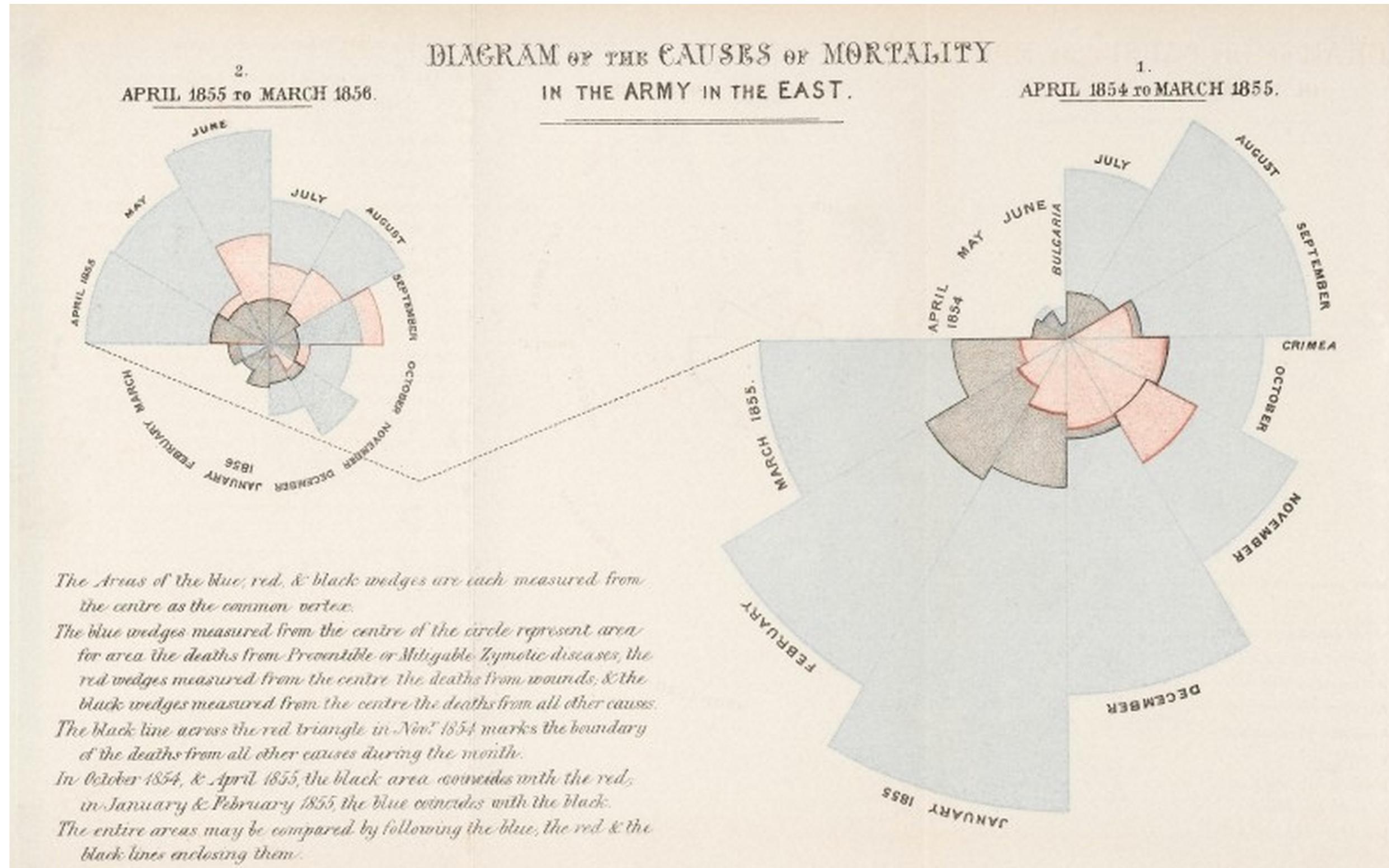
Florence Nightingale



red: death from wounds

black: death from “other” causes

Florence Nightingale



red: death from wounds

black: death from “other” causes

blue: death from preventable causes

Flights

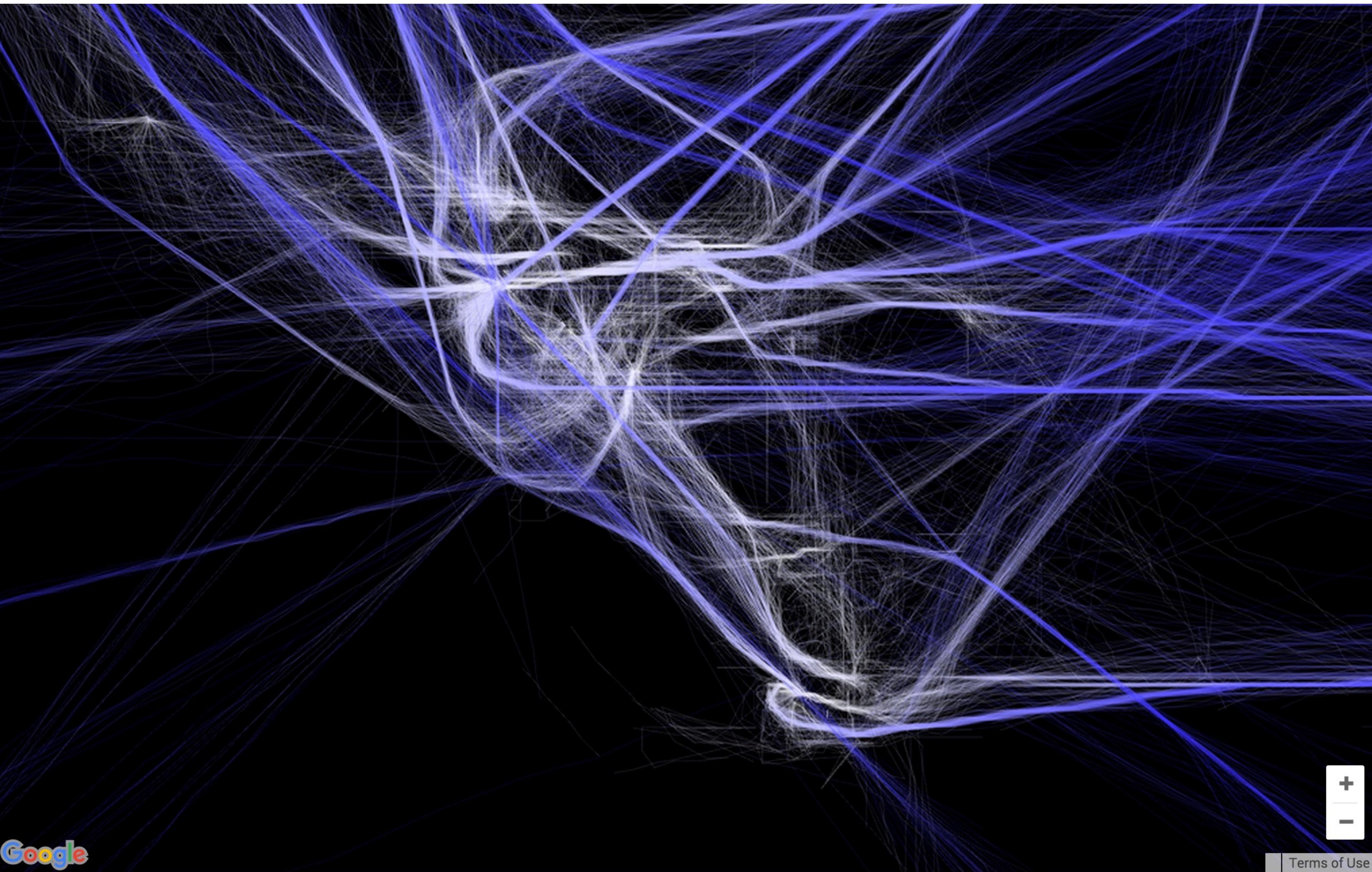


Google

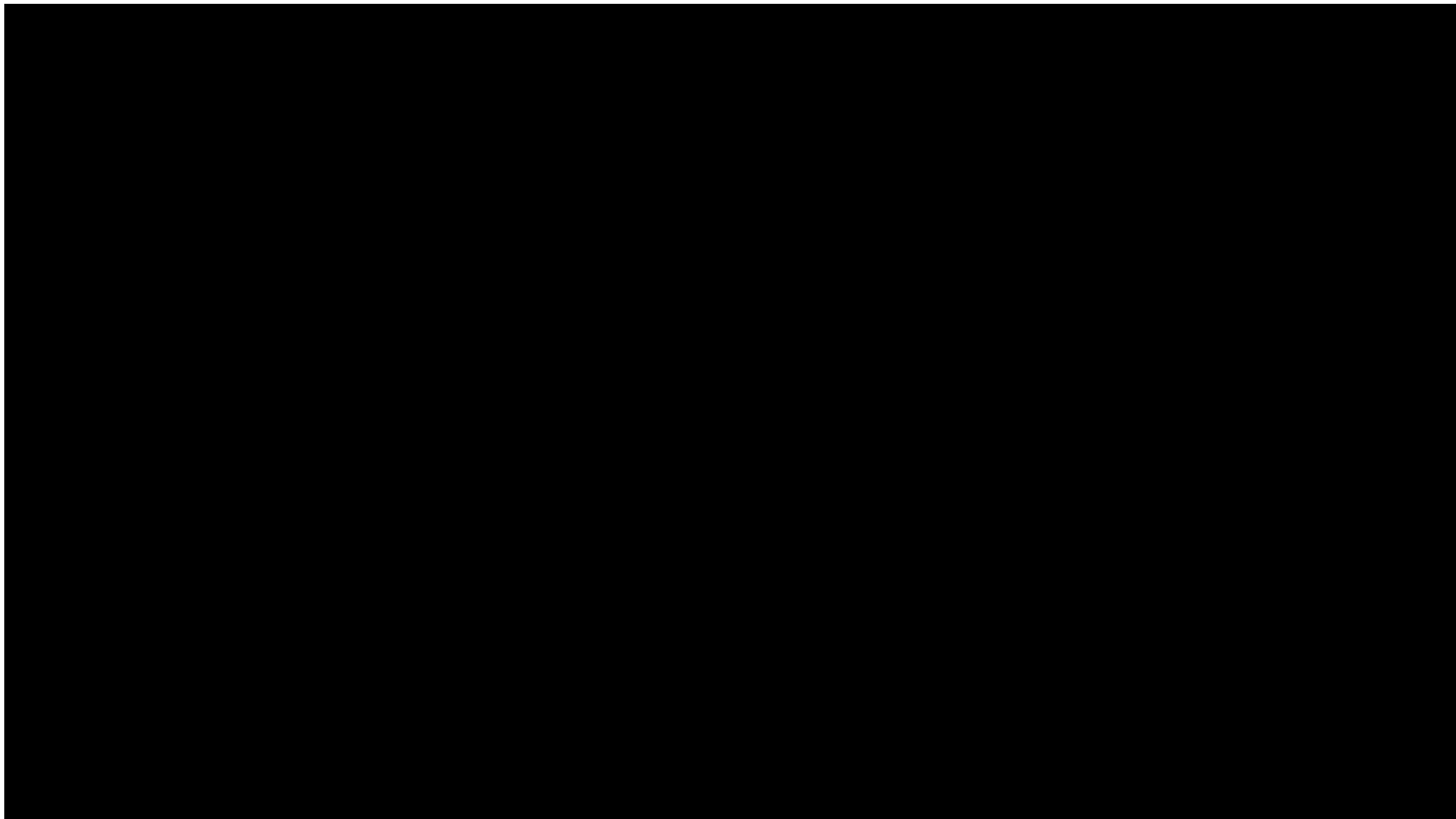


Terms of Use

Flights

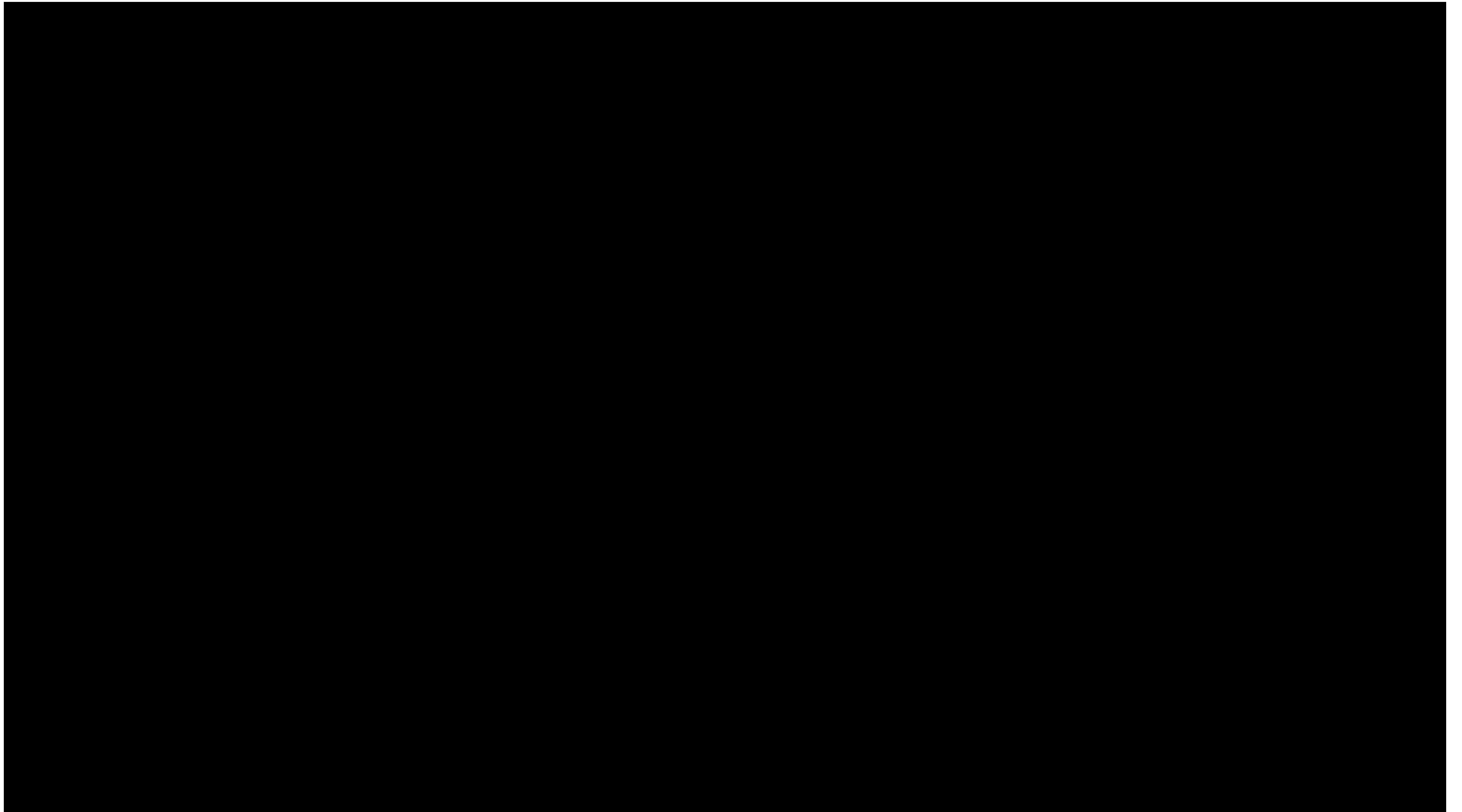


Flights



BBC FOUR





source: Google News Labs - https://www.youtube.com/watch?v=TA_tNh0LMEs

COGS 9
Introduction to Data Science

Algorithmic inference

Today's Learning Objective

How can algorithmic inference inform decision making?

Algorithms

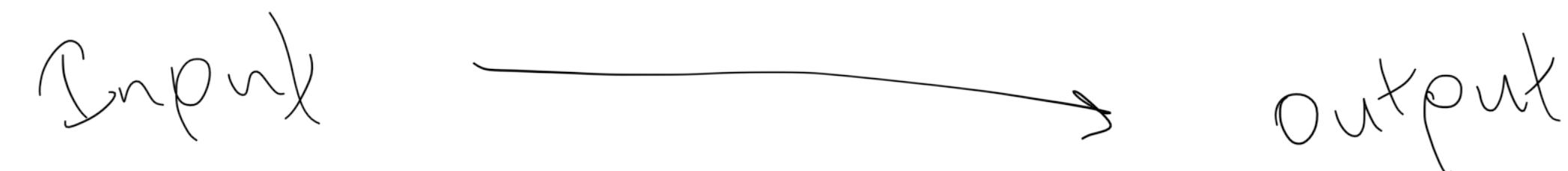
- A finite set of rules that, if followed precisely, defines a sequence of operations that solves the problem in a finite number of steps.

Algorithm - Step by step process to
solve a problem

Algorithms

- ...or: a series of defined steps for solving a problem.

Algorithms



- Takes an input, produces an output. Both must be well-defined!
- How will you store the data? It should be a format that makes it easier to do the operations you want.

Algorithms

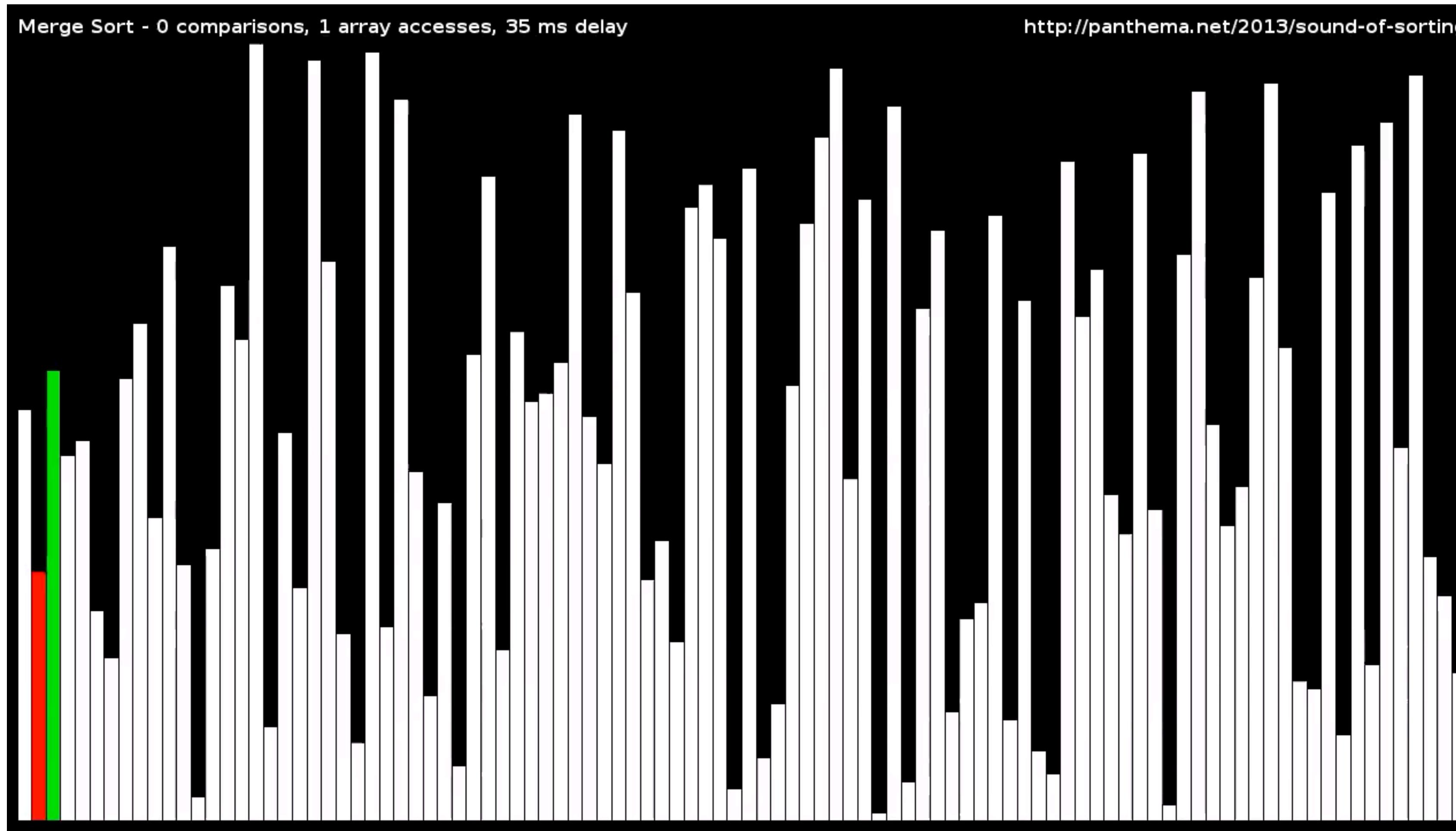
- What happens if you're not precise enough?

SANDWICH ALGORITHM

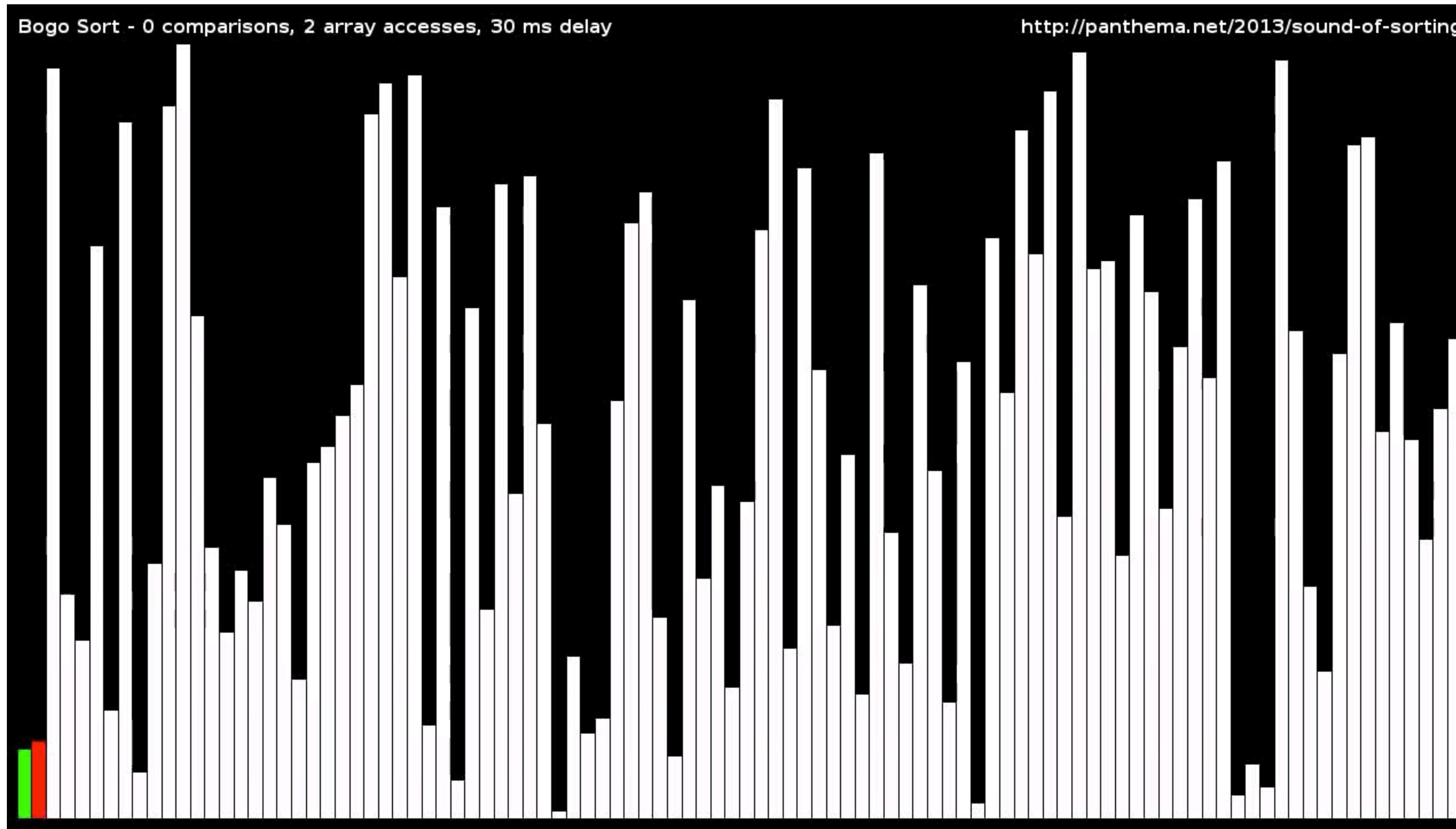
Algorithms take many forms

6 5 3 1 8 7 2 4

Algorithms take many forms



Algorithms take many forms



Algorithms take many forms

- Bogosort: 52!
 - 80658175170943878571660636856403766975289505440883277824000000000000
 - 8.066×10^{67}

Algorithms take many forms

- Bogosort: 52!
 - 80658175170943878571660636856403766975289505440883277824000000000000
 - 8.066×10^{67}
- That's a big number

Algorithms take many forms

- Start a timer that will count down the number of seconds from 52! to 0. What can you do in that time?

Algorithms take many forms

- Starting at the equator; take one step. Every 1B years.

Algorithms take many forms

- Starting at the equator, take one step. Every 1B years.
- Every time you circle the earth, remove a drop of water from the Pacific.

Algorithms take many forms

- Starting at the equator; take one step. Every 1B years.
- Every time you circle the earth, remove a drop of water from the Pacific.
- Circle the Earth again, taking one step every 1B years.

Algorithms take many forms

- Starting at the equator; take one step. Every 1B years.
- Every time you circle the earth, remove a drop of water from the Pacific.
- Circle the Earth again, taking one step every 1B years.
- Continue until the Pacific Ocean is empty.

Algorithms take many forms

- Starting at the equator, take one step. Every 1B years.
- Every time you circle the earth, remove a drop of water from the Pacific.
- Circle the Earth again, taking one step every 1B years.
- Continue until the Pacific Ocean is empty.
- When empty, place a piece of paper on the ground.

Algorithms take many forms

- Starting at the equator; take one step. Every 1B years.
- Every time you circle the earth, remove a drop of water from the Pacific.
- Circle the Earth again, taking one step every 1B years.
- Continue until the Pacific Ocean is empty.
- When empty, place a piece of paper on the ground.
- Fill the Pacific back up and start over: 1B year steps, drop, paper.

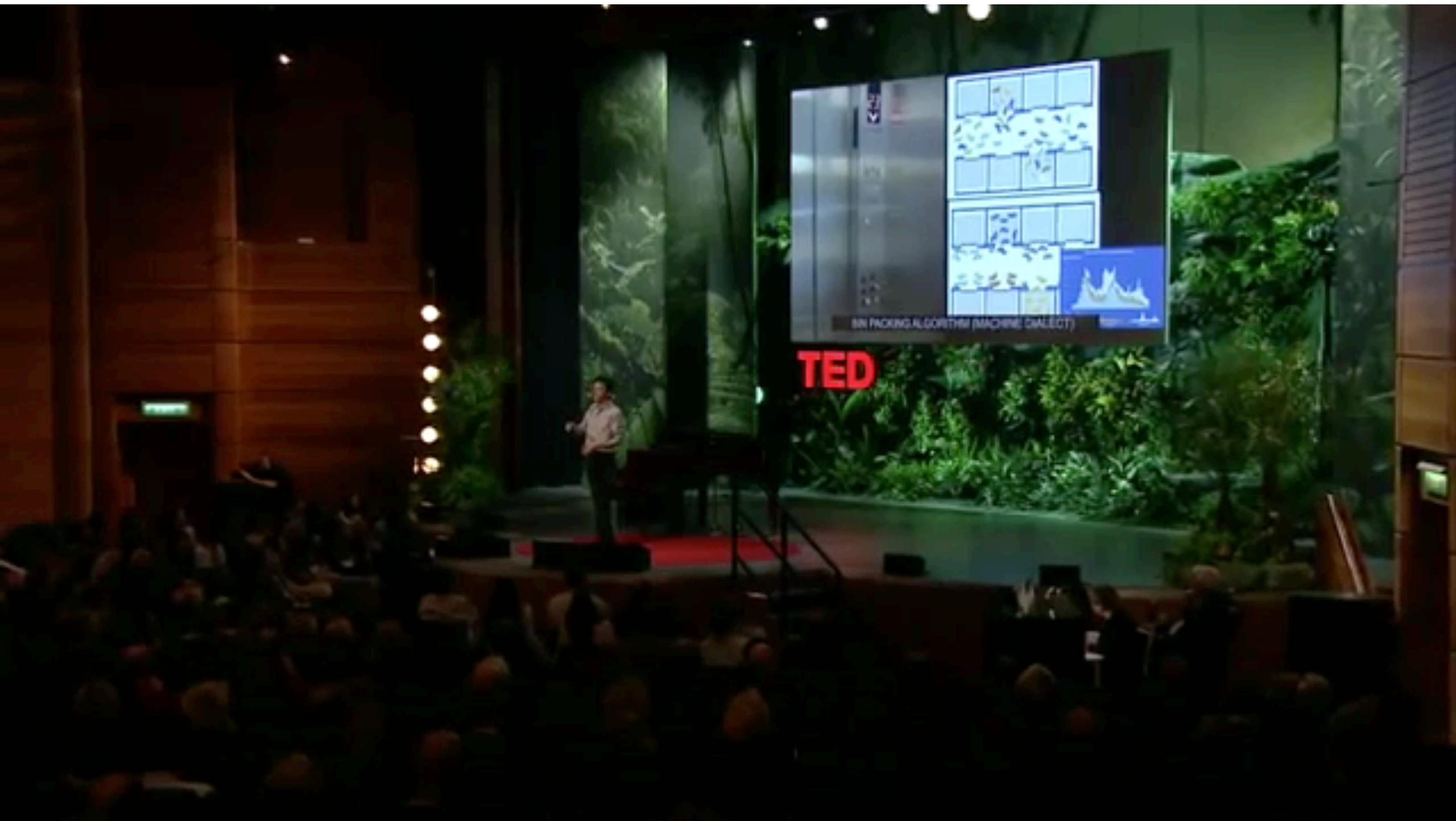
Algorithms take many forms

- Starting at the equator, take one step. Every 1B years.
- Every time you circle the earth, remove a drop of water from the Pacific.
- Circle the Earth again, taking one step every 1B years.
- Continue until the Pacific Ocean is empty.
- When empty, place a piece of paper on the ground.
- Fill the Pacific back up and start over: 1B year steps, drop, paper.
- Do this until the stack of paper reaches from the Earth to the Sun.

Algorithms take many forms

- You still have 8.063e67 more seconds to go.

Algorithms



Algorithms

Trading program sparked May 'flash crash'



Government regulators say a trading program was behind the massive stock slide on May 6.

Automatic computerized traders on the stock market shut down as they detected the sharp rise in buying and selling. Altogether, this led to the abrupt drop in prices of individual stocks and other financial instruments like [exchange-traded funds](#), and caused shares of some prominent companies like Procter & Gamble and Accenture to trade down as low as a penny or as high as \$100,000.

Algorithm design

Algorithm → 1) solve problem

2) be adaptive (by human or computer)

- solve a problem
 - efficiently
 - as simply as possible
 - correctly & ethically
 - ...in a way that doesn't make humans freak out
- be adapted
 - by humans or systems
 - ...with minimal struggle

FAT algorithms

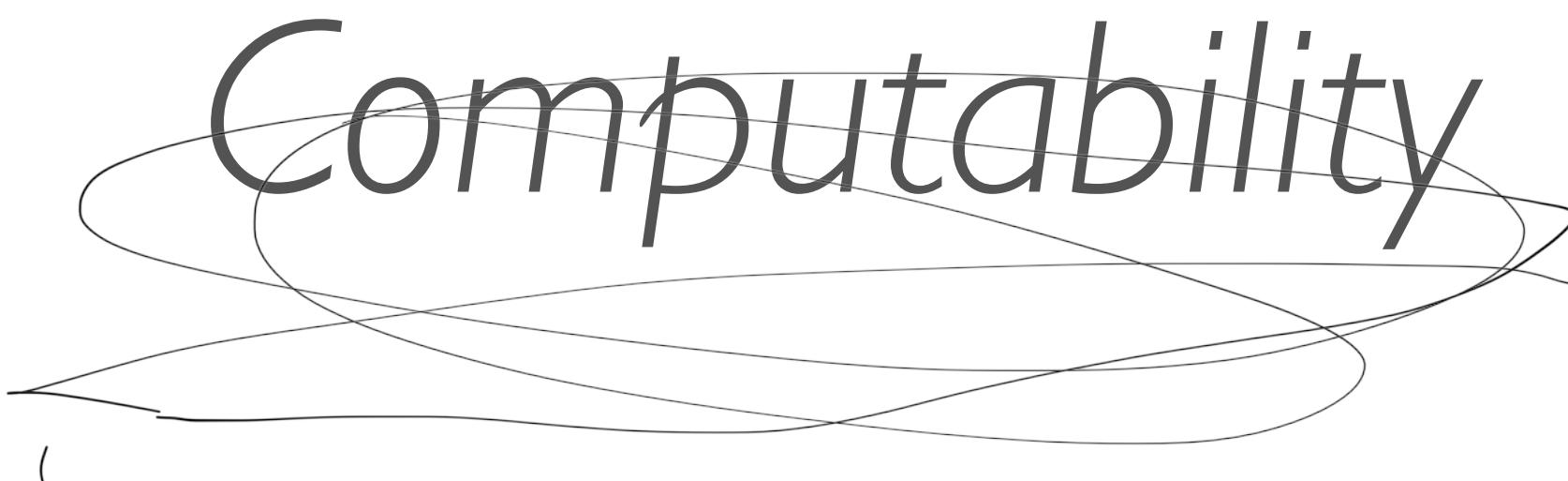
- **F**air: lacking biases which create unfair and discriminatory outcomes
 - For whom does this algorithm fail?
 - Steps to take:
 - 1. Verify data about individual is correct
 - 2. Carry out “sensitivity test”
- **A**ccountable: answerable to the people subject(ed) to them
 - Requires access to algorithm
- **T**ransparent: open about how, and why, decisions were made
 - Think carefully about what transparency is (Handing over source code likely isn't the answer)

Fair - discrimination and bias
Accountable - answerable to the people subject to them
Transparent
Access to the algorithm
how and why decisions are made

COGS 9

Introduction to Data Science

Computability



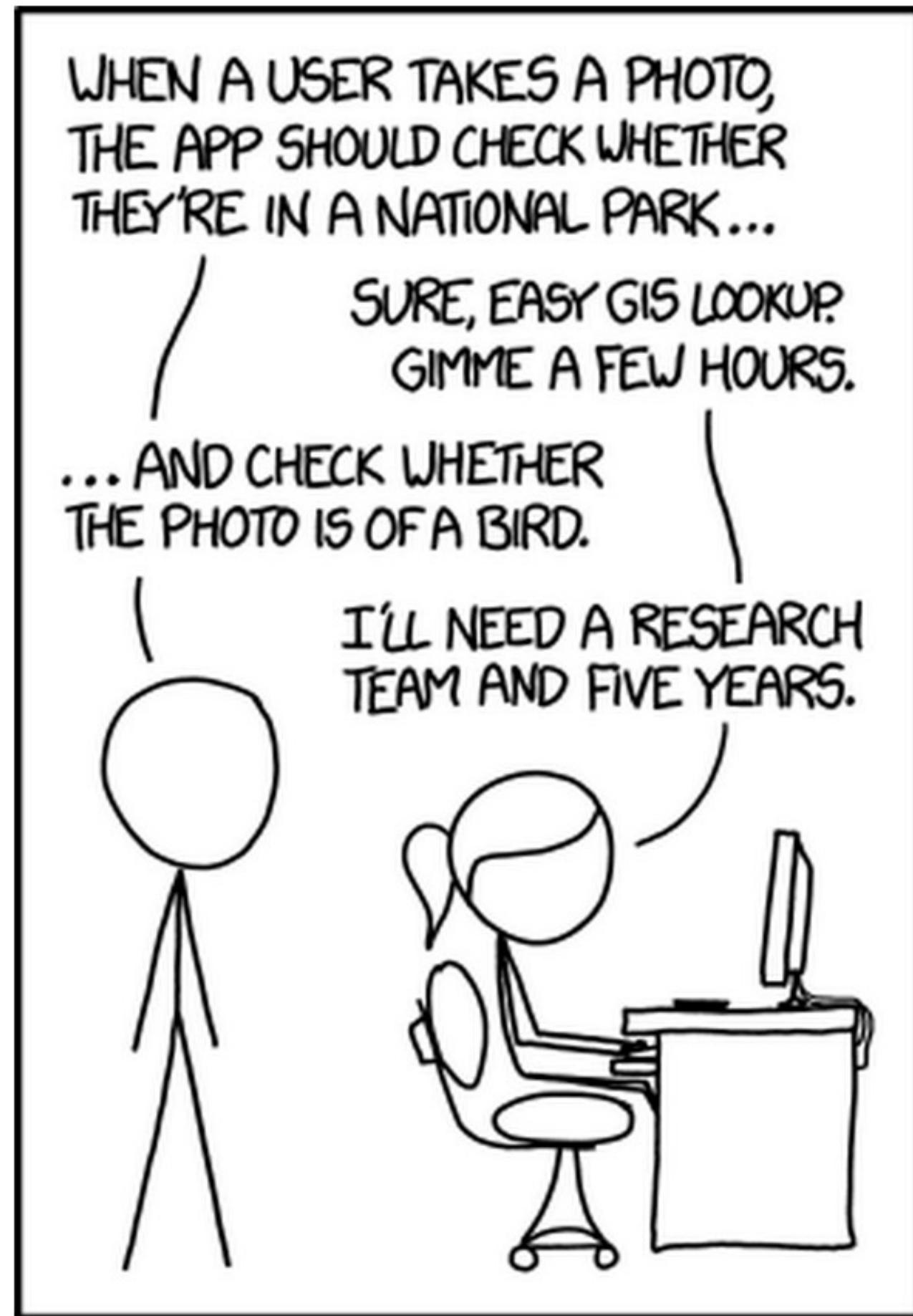
Today's learning objective

What is computability, and how is it measured?

What's computable?

- A problem is *computable* if it can be solved by some algorithm.
- A problem is *noncomputable* if it cannot be solved by any algorithm.

Computability is not intuitive!



IN CS, IT CAN BE HARD TO EXPLAIN
THE DIFFERENCE BETWEEN THE EASY
AND THE VIRTUALLY IMPOSSIBLE.

Computability is not intuitive!

- A surprising number of problems are noncomputable!
- These are precisely-defined problems for which—no matter how smart we are and how much computing power and memory we have—there is provably no algorithm that can solve them.

Efficiency

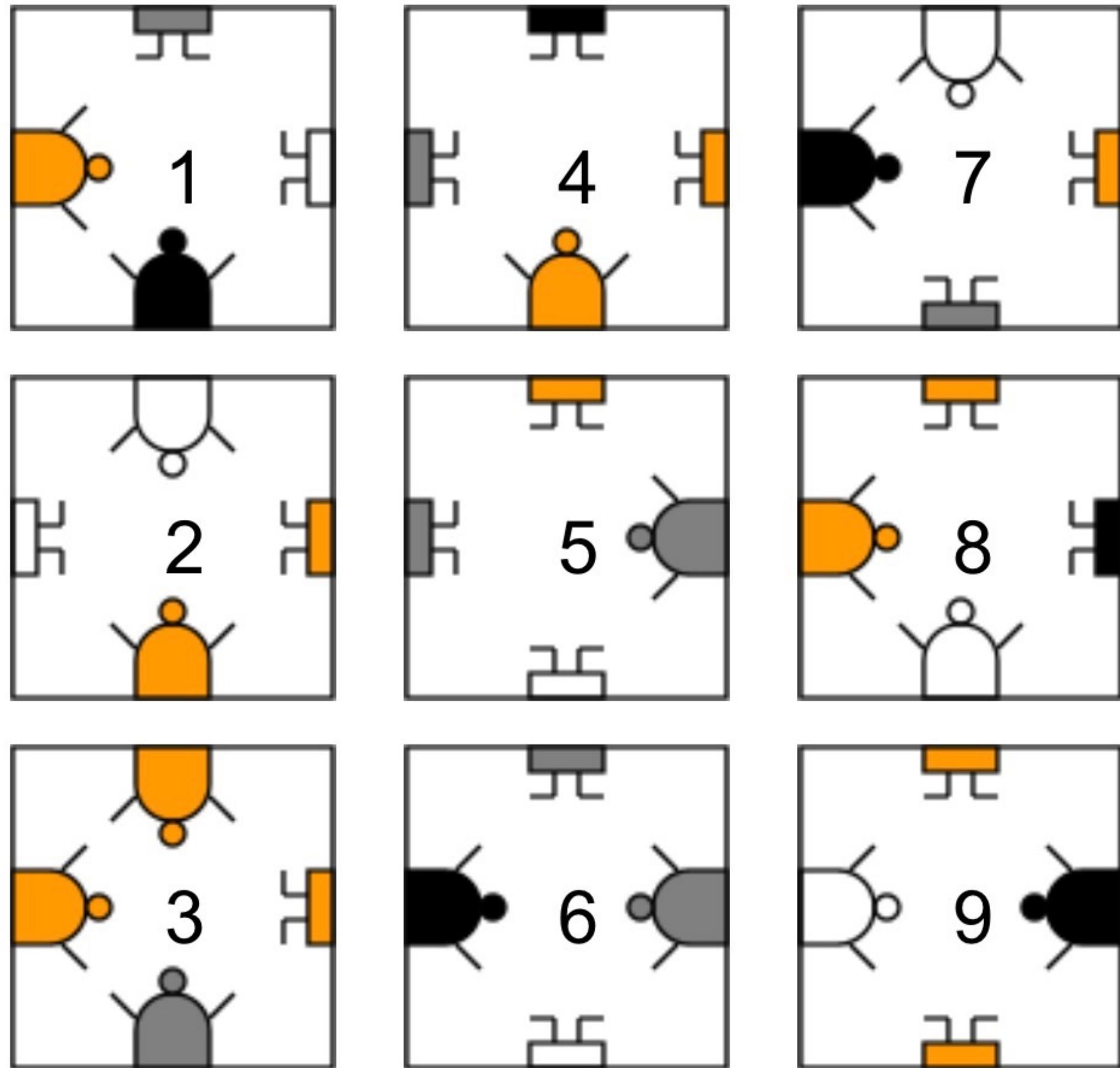
Efficiency in computer programming
time efficiency & storage efficiency

- A computer program should be totally correct, but it should also:
 - Execute as quickly as possible (time-efficiency)
 - Use memory wisely (storage-efficiency)

Efficiency

- How do we compare programs (or algorithms in general) with respect to execution time?
 - Various computers run at different speeds due to different processors.
 - The same algorithm can be written differently depending on the programming language used, etc.

Monkey Puzzle Problem



- Determine if an arrangement of the n cards in an $m \times m$ grid exists such that each adjacent pair of cards displays the upper and lower half of a monkey of the same color.
- Cards cannot be rotated.

Monkey Puzzle Problem

- For n cards, the number of arrangements to examine is $n!$
- If we can analyze one arrangement in a microsecond:
 - $m = 3; n = 9: 0.36288$ seconds
 - $m = 4; n = 16: \sim 242$ days
 - $m = 5; n = 25: \sim 4.9$ billion years

So what do we do for
difficult-to-compute
problems?

COGS 9
Introduction to Data Science

Crowdsourcing & Human-based computation

Today's Learning Objective

What is crowdsourcing, and how is it used?

Today's Learning Objective

What is human-based computation, and how is it taken advantage of?

The Wisdom of Crowds

TO COMPLETE YOUR REGISTRATION, PLEASE TELL US WHETHER OR NOT THIS IMAGE CONTAINS A STOP SIGN:



NO YES

ANSWER QUICKLY—OUR SELF-DRIVING CAR IS ALMOST AT THE INTERSECTION.

SO MUCH OF "AI" IS JUST FIGURING OUT WAYS TO OFFLOAD WORK ONTO RANDOM STRANGERS.

Captchas





Re-captcha



Mechanical Turk

amazon mechanical turk
Artificial Artificial Intelligence

Your Account HITs Qualifications

Introduction | Dashboard | Status | Account Settings

Mechanical Turk is a marketplace for work.
We give businesses and developers access to an on-demand, scalable workforce.
Workers select from thousands of tasks and work whenever it's convenient.

526,579 HITs available. [View them now.](#)

Make Money
by working on HITs

HITs - *Human Intelligence Tasks* - are individual tasks that you work on. [Find HITs now.](#)

As a Mechanical Turk Worker you:

- Can work from home
- Choose your own work hours
- Get paid for doing good work

Find an interesting task **Work** **Earn money**



[Find HITs Now](#)

Get Results
from Mechanical Turk Workers

Ask workers to complete HITs - *Human Intelligence Tasks* - and get results using Mechanical Turk. [Get Started.](#)

As a Mechanical Turk Requester you:

- Have access to a global, on-demand, 24 x 7 workforce
- Get thousands of HITs completed in minutes
- Pay only when you're satisfied with the results

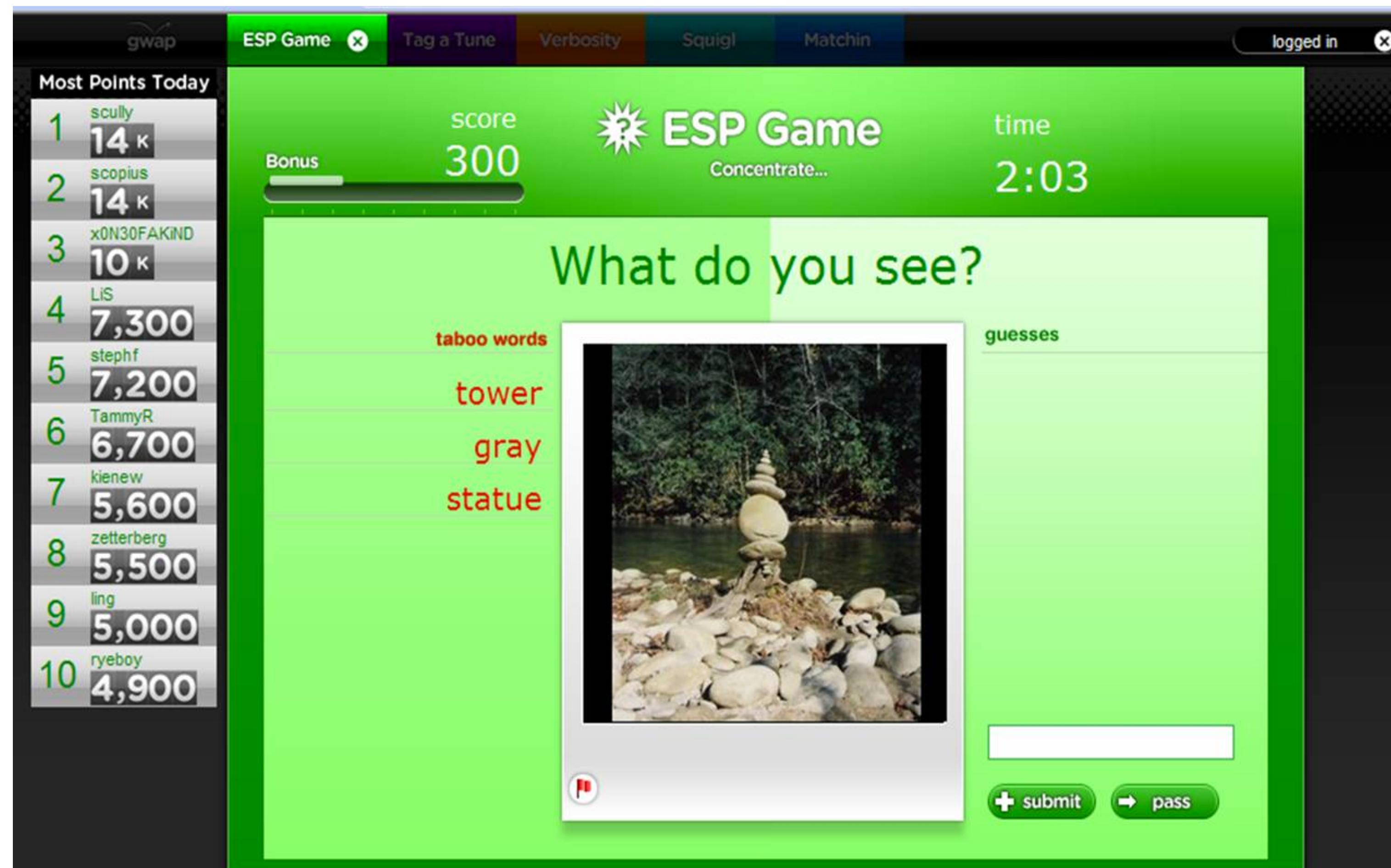
Fund your account **Load your tasks** **Get results**



[Get Started](#)

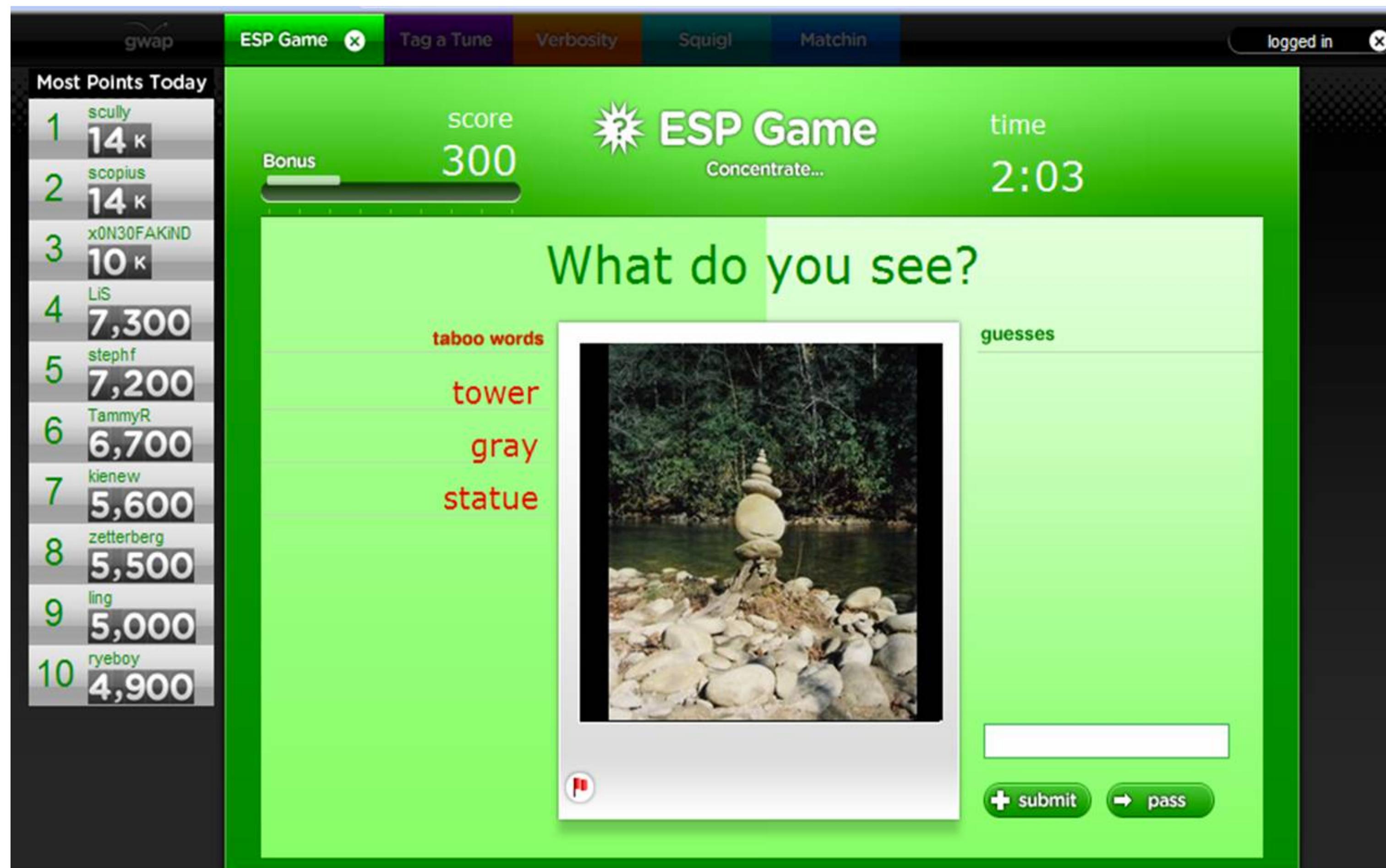
Already have an account?
Sign in as a [Worker](#) | [Requester](#)

ESP Game



So how can the ESP game work?

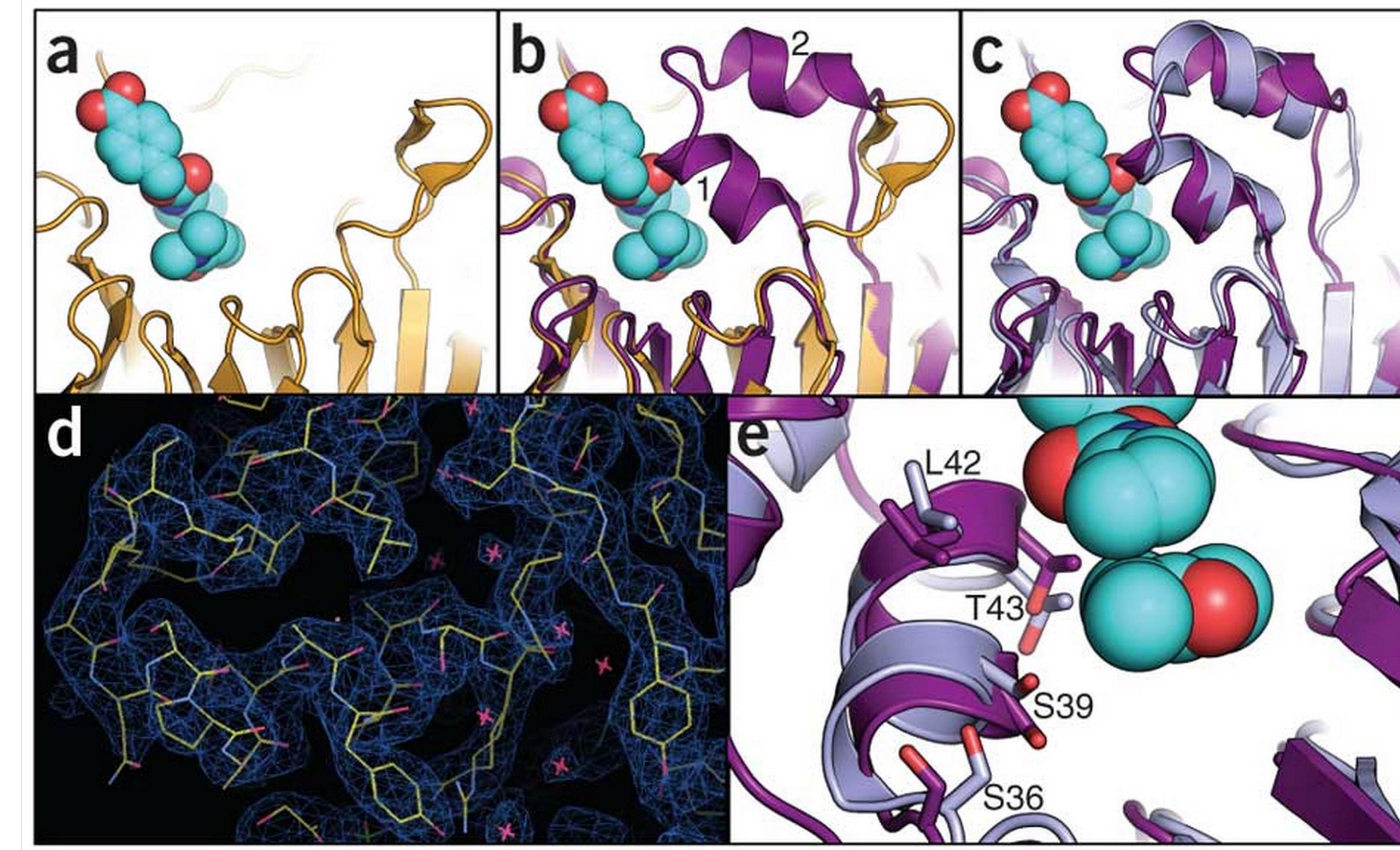
ESP Game



Citizen Science

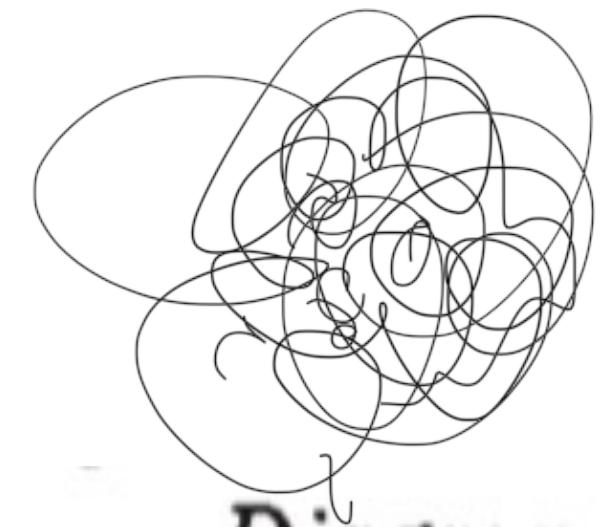


FoldIt



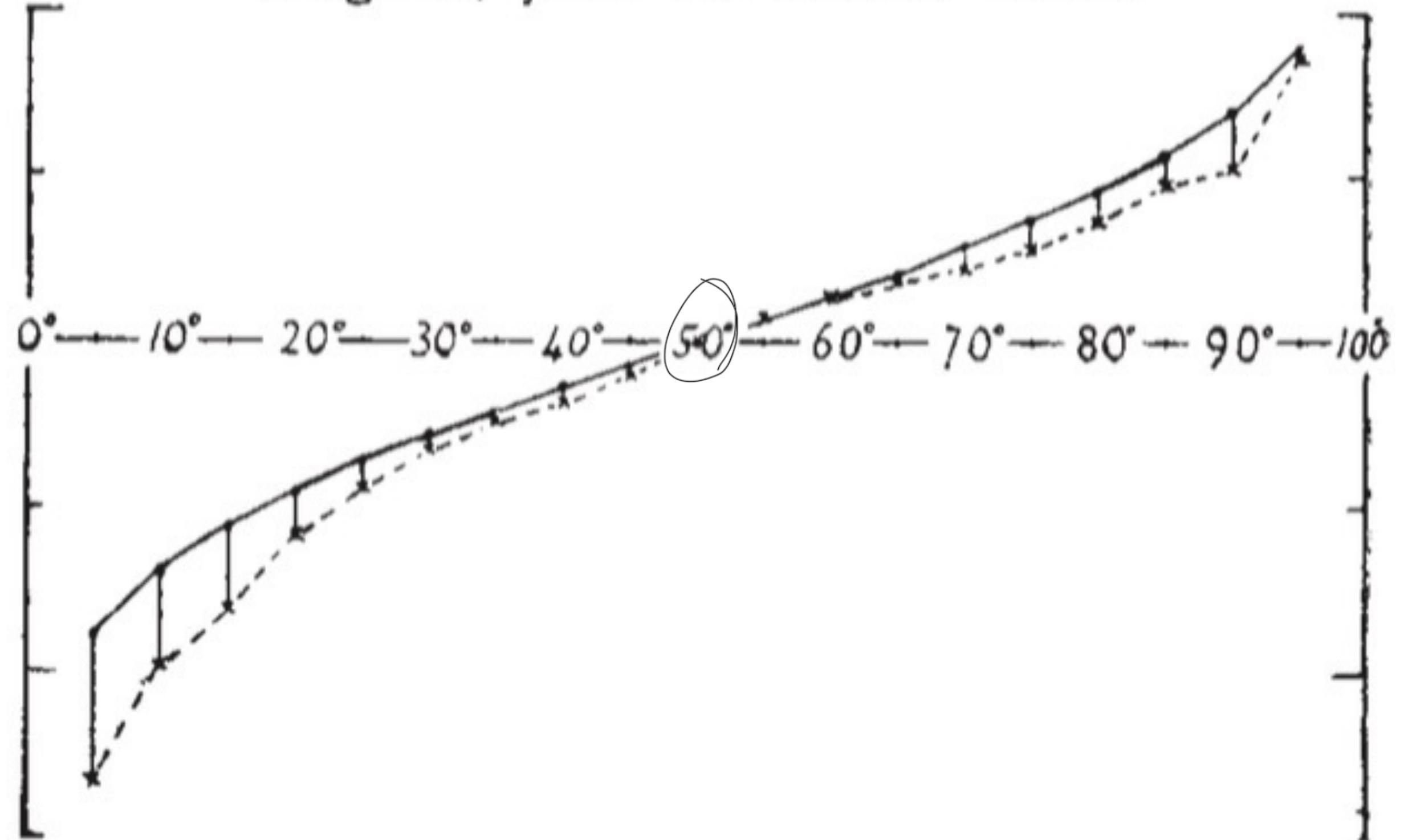
Galaxy Zoo

The screenshot shows the Galaxy Zoo website interface. At the top, there is a navigation bar with links for CLASSIFY, STORY, SCIENCE, DISCUSS, PROFILE, and LANGUAGE. Social media icons for Facebook, Twitter, Google+, and RSS feed are also present. The main content area features a large image of a spiral galaxy on the left. On the right, there is a classification interface with the word "Classify" and two buttons: "DECals" and "Invert". Below this, a note says "Note: Please always classify the galaxy in the center of the image." A "SHAPE" question asks if the galaxy is smooth and rounded. Three options are shown: "Smooth" (a solid white circle), "Features or disk" (a white circle with a spiral or ring-like feature), and "Star or artifact" (a starburst shape). Buttons for "Examples" and "Restart" are located above the classification area.



Vox Populi

Diagram, from the tabular values.



Middlemost estimate: 1207 lbs
True weight: 1198 lbs (0.8% under)

787 estimates

Vox populi
"voice of the people"

The Wisdom of Crowds

Diversity of opinion
Independence

- Diversity of opinion: Each person should have private information even if it's just an eccentric interpretation of the known facts.
- Independence: People's opinions aren't determined by the opinions of those around them.
- Decentralization: People are able to specialize and draw on local knowledge.
- Aggregation: Some mechanism exists for turning private judgments into a collective decision.

Aggregation