Course Introduction:
17880 Algorithms for Private Data Analysis

Instructor: Steven Wu

https://dpcmu.github.io/

Video On if Possible

Introduction: Steven Wu

- CMU SCS faculty (ISR/MLD/HCII)
- Interests: machine learning & algorithms
  - Privacy/Fairness
  - Algorithmic game theory/economics
  - Human-AI interactions
- Outside of work:
  - Basketball/rock climbing/hiking
  - Corgi

Introduction: yourself

- Who you are?
- Your interests (inside and outside of work)
- Why you are interested in this course?

This Course

https://dpcmu.github.io/

- Intro to research on privacy in ML and Statistics
  - Formal models: *differential privacy*
  - Algorithmic techniques (beyond privacy)
- Skills you will work on
  - Formal reasoning
  - Research skills (in CS/ML/Stats)
  - Optional: programming
- Pre-requisites
  - Comfort with reading/writing proofs about basic probability and linear algebra

Every lecture

- Ahead of lecture
  - Finish assigned reading (video/lecture note/papers)
- In class
  - Participate and *try to* be on video

- Lecture format
  - Live lectures with slides/iPad demo.
  - Potential experiments: flip classroom
    - Watch recorded lectures ahead of time
    - Do in-class activities or exercises

Coursework

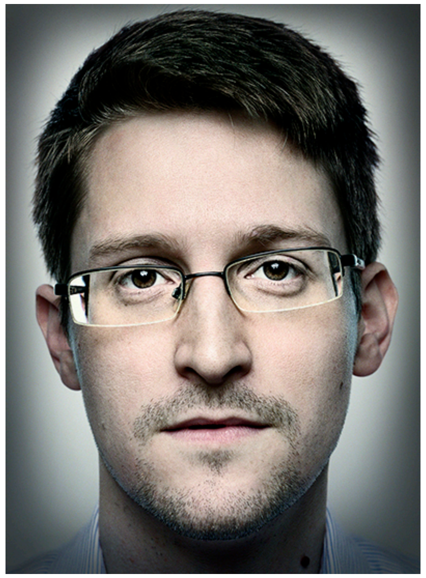
- Lecture prep and in-class work
- Homework (4 assignments)
  - Collaboration allowed
  - Write up your solutions and acknowledge collaborators
- Project (details TBA)
  - *Research project:* work on research related to this course
  - *Reading project:* summarize 2-3 papers
  - Presentation in the last week of class

Grading

- In-class participation: 20%
  - *Soft* rule of thumb: speak up at least 10 times during the whole course
- 4 homework assignments: 50%
  - 5 late days allowed
- Final project: 30%

Questions?

What this course is *not* about



GDPR

Cambridge Analytica

Privacy-Preserving Data Analysis

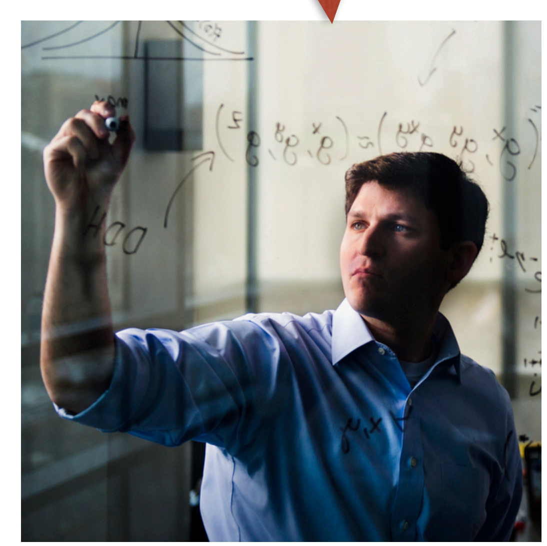
- Epidemic detection
- Analysis of loan application data for evidence of discrimination
- Training of ML model to predict user behavior

Anonymization?

Original Dataset

De-identification

“De-identified” Dataset

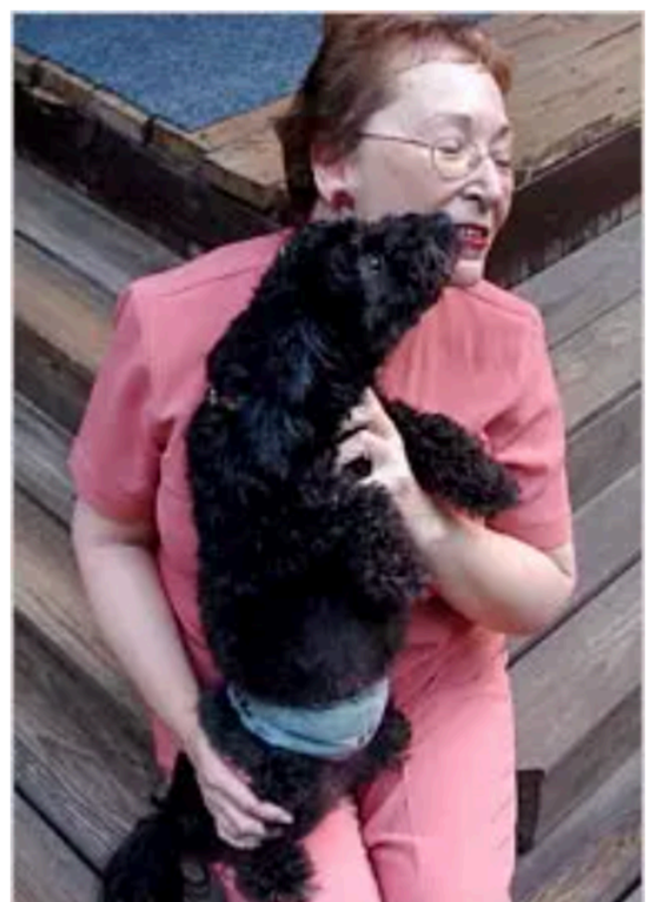


The New York Times

A Face Is Exposed for AOL Searcher No. 4417749

By Michael Barbaro and Tom Zeller Jr.

Aug. 9, 2006



Thelma Arnold's identity was betrayed by AOL records of her Web searches, like ones for her dog, Dudley, who clearly has a problem.
Erik S. Lesser for The New York Times

The New York Times

Netflix Cancels Contest After Concerns Are Raised About Privacy

By Steve Lohr

March 12, 2010

Robust De-anonymization of Large Datasets
(How to Break Anonymity of the Netflix Prize Dataset)

Arvind Narayanan and Vitaly Shmatikov

The University of Texas at Austin

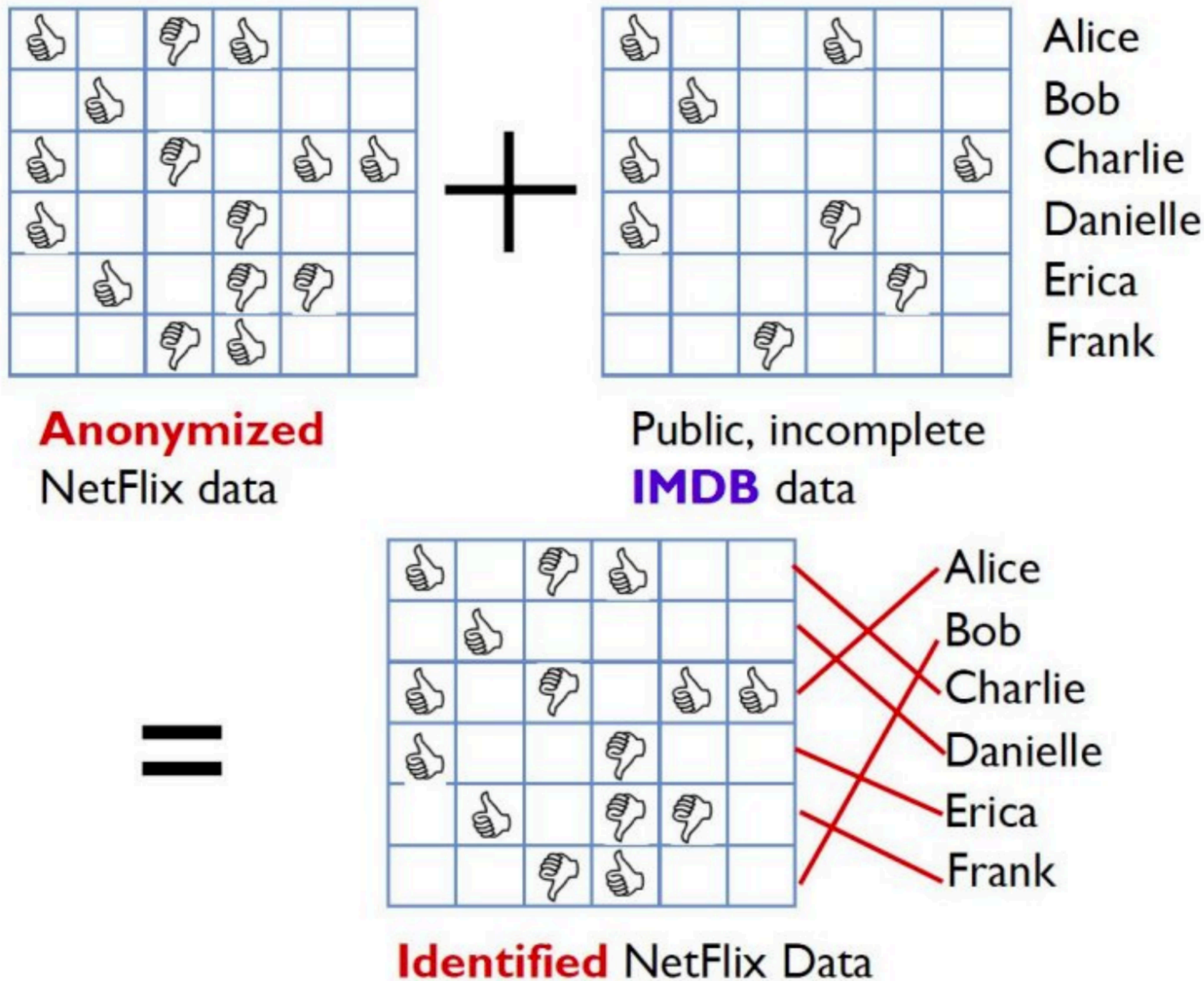
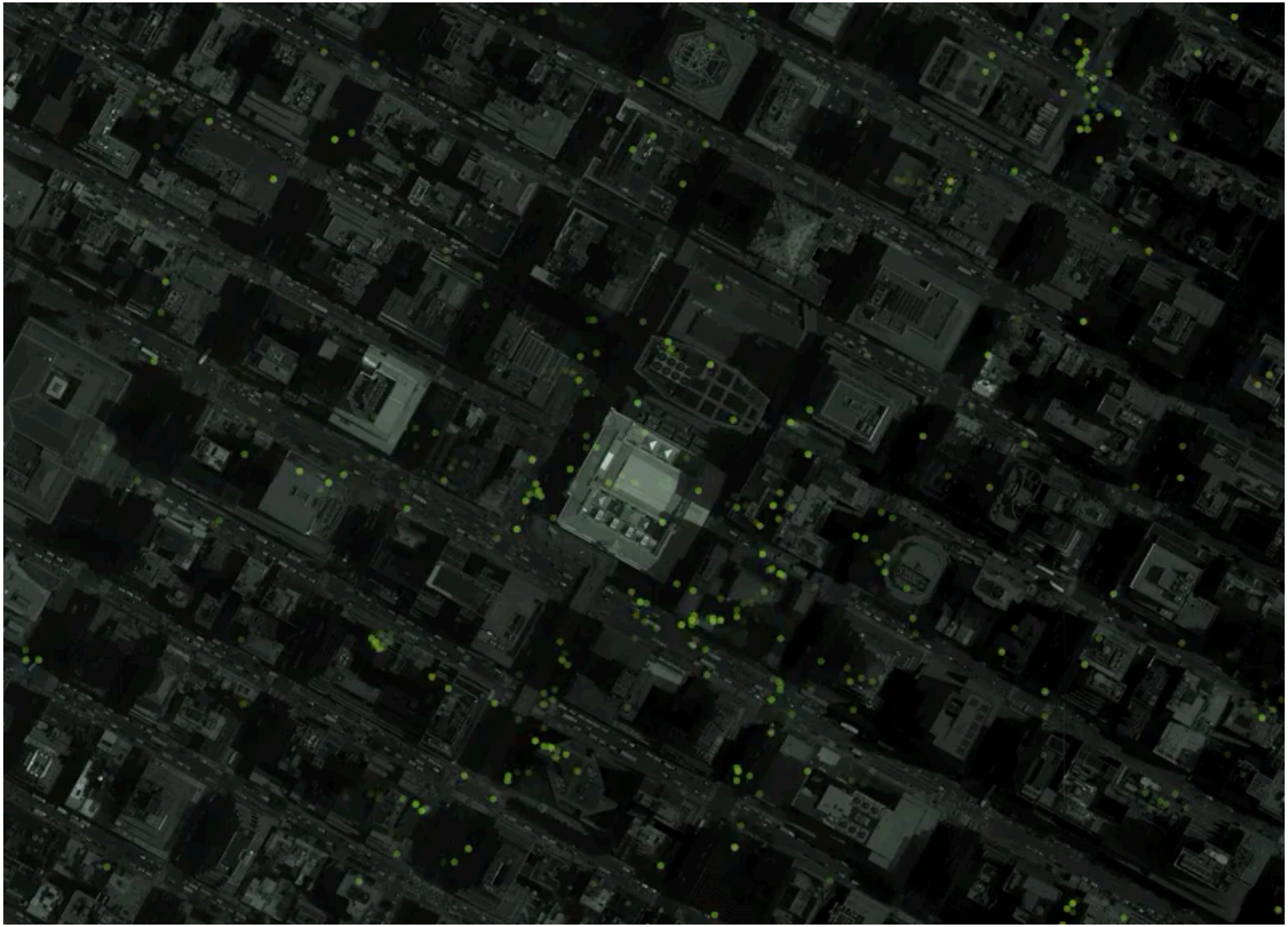


Image credit: Arvind Narayanan

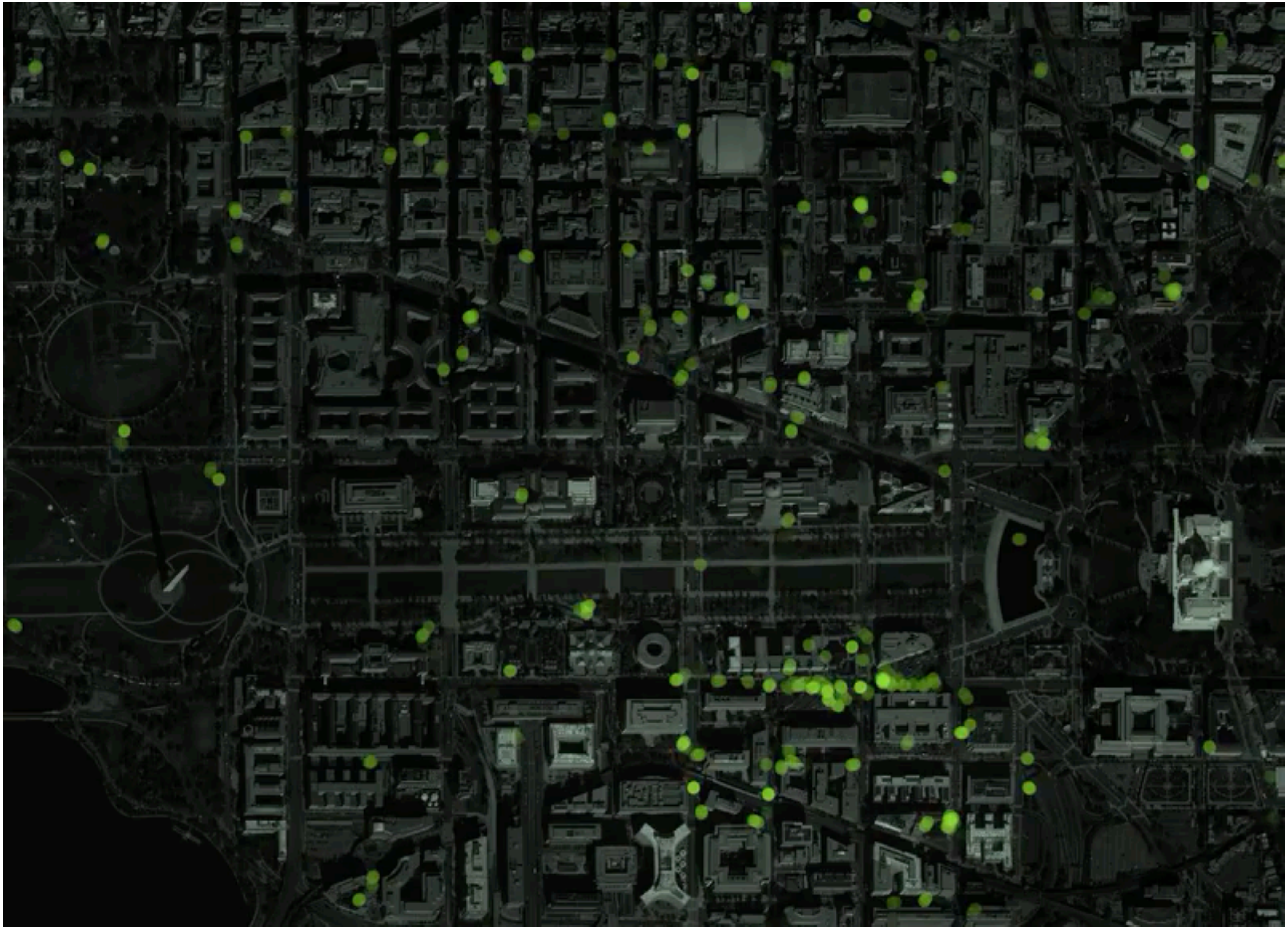
ONE NATION, TRACKED

Twelve Million Phones, One Dataset, Zero Privacy

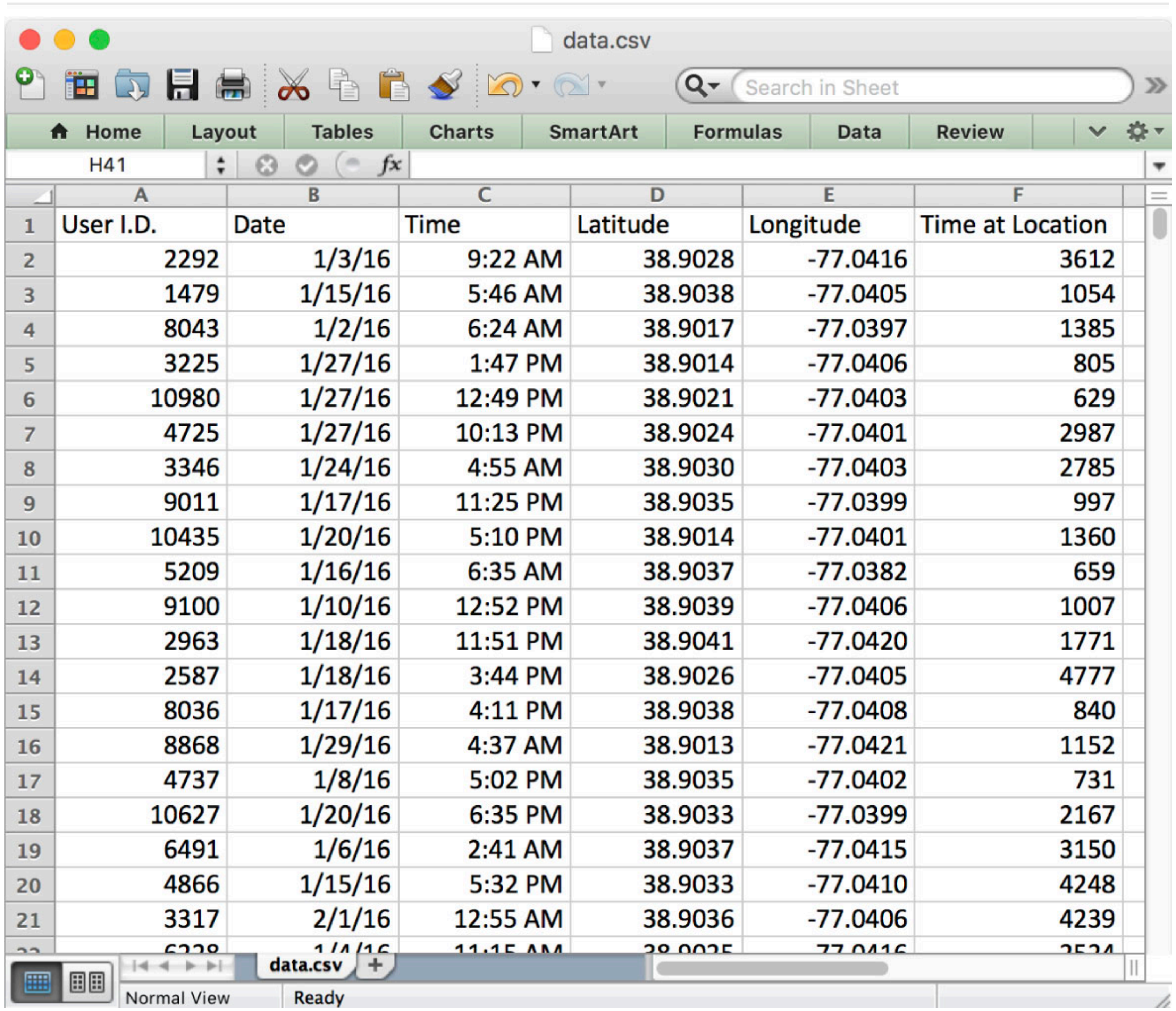
https://www.nytimes.com/interactive/2019/12/19/opinion/location-tracking-cell-phone.html



A typical day at Grand Central Terminal in New York City



Senior Defense Department official and his wife identified at the Women's March



User I.D.	Date	Time	Latitude	Longitude	Time at Location
2292	1/3/16	9:22 AM	38.9028	-77.0416	3612
1479	1/15/16	5:46 AM	38.9038	-77.0405	1054
8043	1/2/16	6:24 AM	38.9017	-77.0397	1385
3225	1/27/16	1:47 PM	38.9014	-77.0406	805
10980	1/27/16	12:49 PM	38.9021	-77.0403	629
4725	1/27/16	10:13 PM	38.9024	-77.0401	2987
3346	1/24/16	4:55 AM	38.9030	-77.0403	2785
9011	1/17/16	11:25 PM	38.9035	-77.0399	997
10435	1/20/16	5:10 PM	38.9014	-77.0401	1360
5209	1/16/16	6:35 AM	38.9037	-77.0382	659
9100	1/10/16	12:52 PM	38.9039	-77.0406	1007
2963	1/18/16	11:51 PM	38.9041	-77.0420	1771
2587	1/18/16	3:44 PM	38.9026	-77.0405	4777
8036	1/17/16	4:11 PM	38.9038	-77.0408	840
8868	1/29/16	4:37 AM	38.9013	-77.0421	1152
4737	1/8/16	5:02 PM	38.9035	-77.0402	731
10627	1/20/16	6:35 PM	38.9033	-77.0399	2167
6491	1/6/16	2:41 AM	38.9037	-77.0415	3150
4866	1/15/16	5:32 PM	38.9033	-77.0410	4248
3317	2/1/16	12:55 AM	38.9036	-77.0406	4239

"De-identified data isn't."

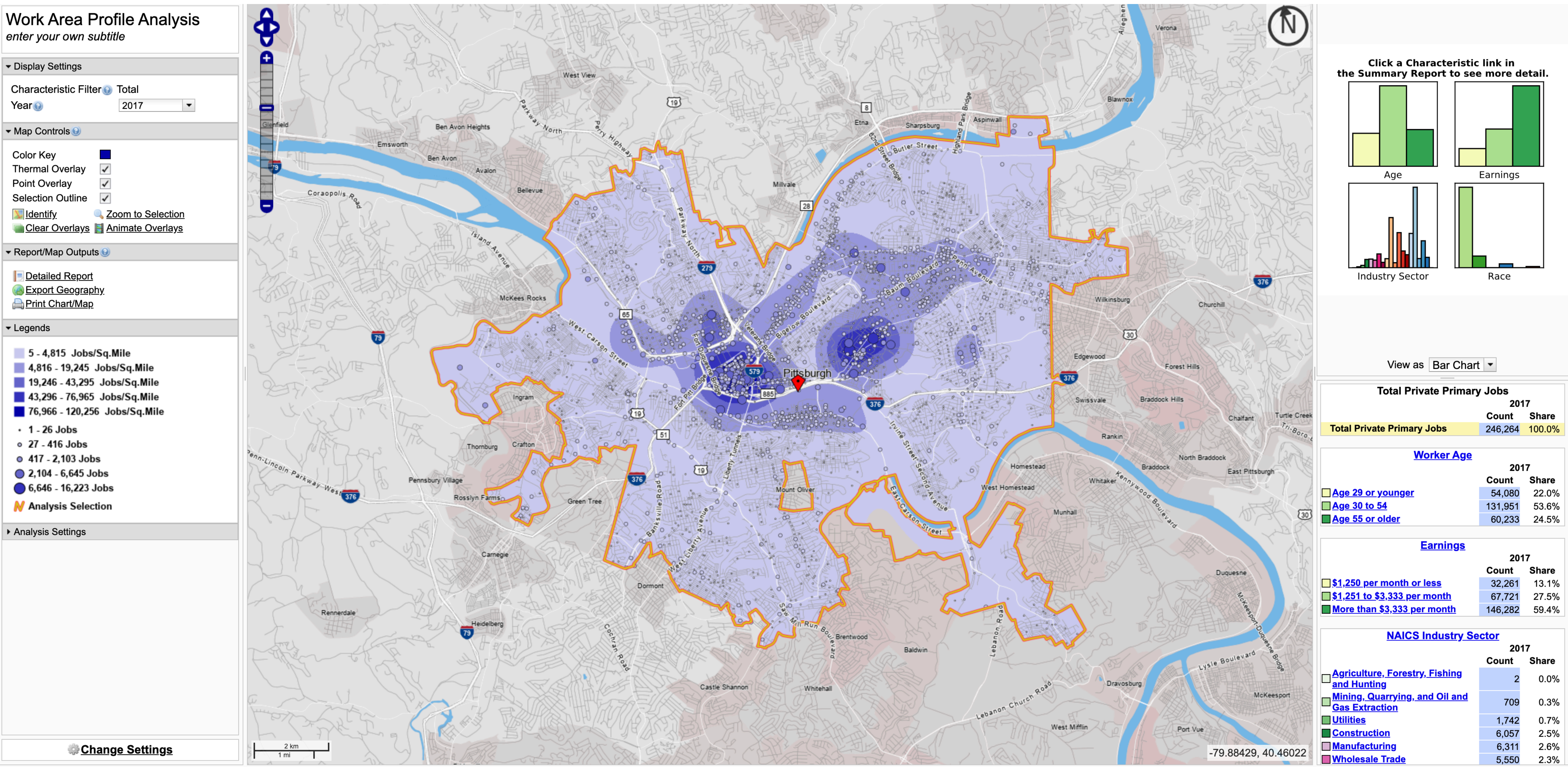
— Cynthia Dwork

How about *just* releasing some statistics?

Differencing Attacks

- *How many people in this Zoom call are wearing socks?*
- *How many people in this Zoom call, except the host, are wearing socks?*

US Census Bureau



Data Collected in 2010 Decennial Census

308,745,538 people $\times$ 6 variables = 1,852,473,228 measurements

Variable	Range
Block	6,207,027 inhabited blocks
Sex	2 (Female/Male)
Age	103 (0-99 single age year categories, 100-104, 105-109, 110+)
Race	63 allowable race combinations
Ethnicity	2 (Hispanic/Not)
Relationship	17 values

Table from Simson L. Garfinkel's slides

Summary of Publications

Publication	Released counts
PL94-171 Redistricting	2,771,998,263
Balance of Summary File 1	2,806,899,669
Total Statistics in PL94-171 and Balance of SF1:	5,578,897,932
Published Statistics/person	18
Recall: Collected variables/person:	6
Published Statistics/collected variable	**18 ÷ 6 ffi 3**

You can create 5.5 billion simultaneous equations and solve for 1.8 billion unknown integers.

US Census Bureau Reconstruction Attack

- "Reconstruction attack" by the Census Bureau researchers on the 2010 Census

- Database reconstruction for 308,745,538 people using census block and tract summary tables from the 2010 Decennial census

Fundamental law of information [Dinur & Nissim]:

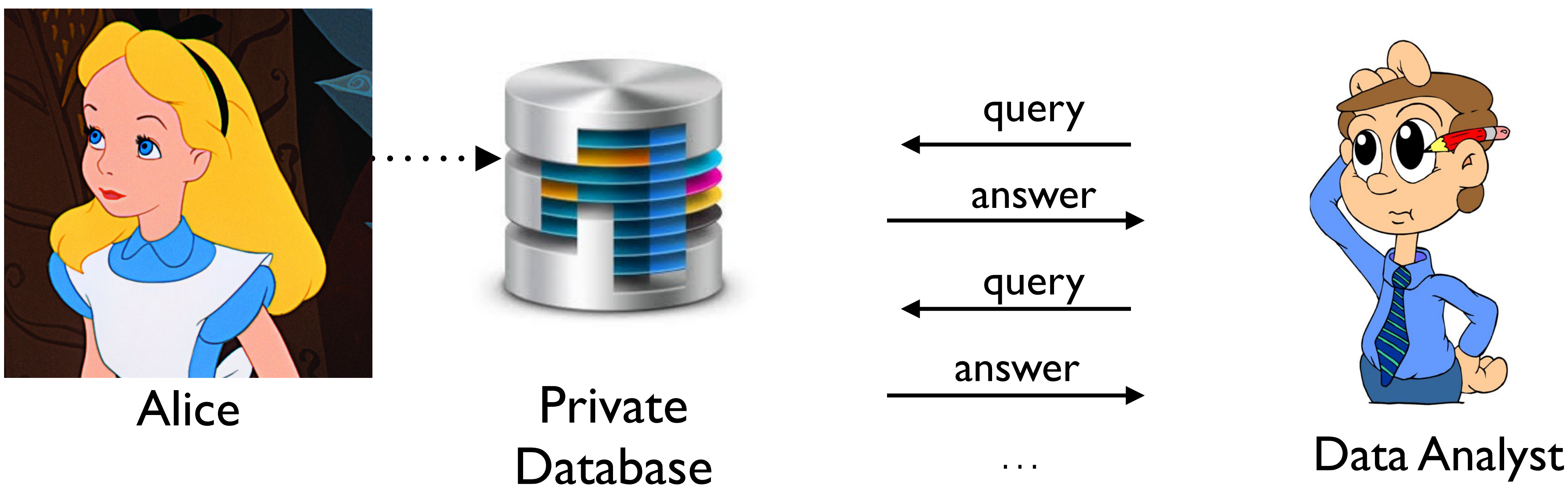
"Overly accurate" estimates of "too many" statistics is non-private.

Lesson Learned

- Ad-hoc privacy measure like de-identification most often fails
- Publishing too many queries on a private database with too much accuracy reveals the contents of the database
- Need for a rigorous and mathematical privacy notion

But what does privacy mean in data analysis?

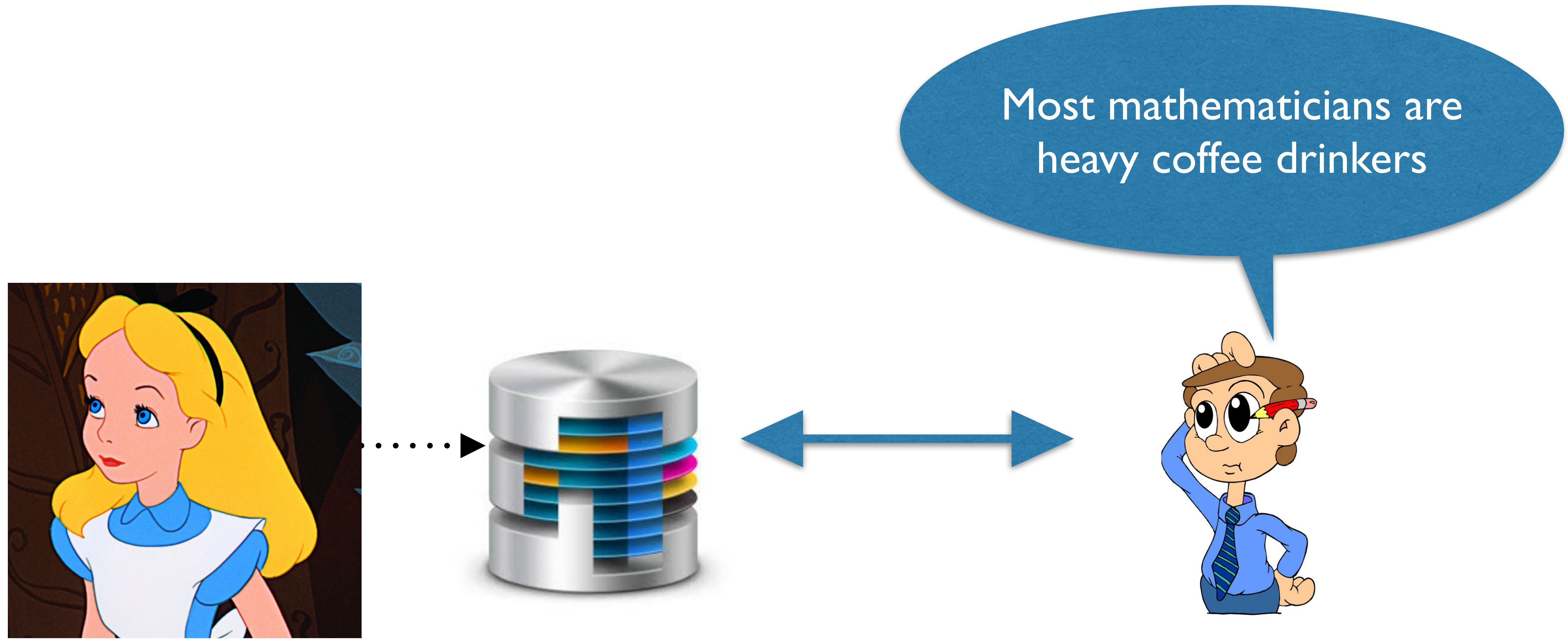
How to formulate privacy?



Privacy Attempt 1:

data analyst can't learn *anything* about Alice??

Hypothetical Scenario



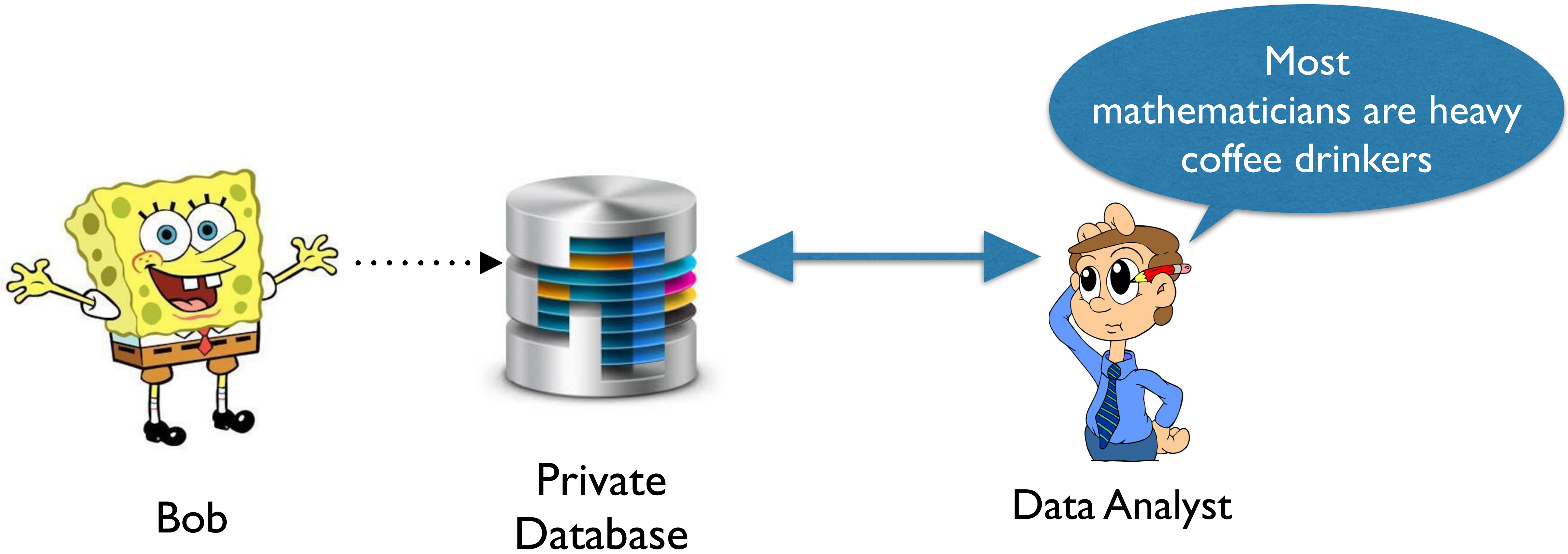
Alice
is a mathematician

Private
Database

Data Analyst

Was Alice's privacy violated?

Replace Alice by Another Random Person



We will learn the same thing if Alice is replaced by any person in the population!

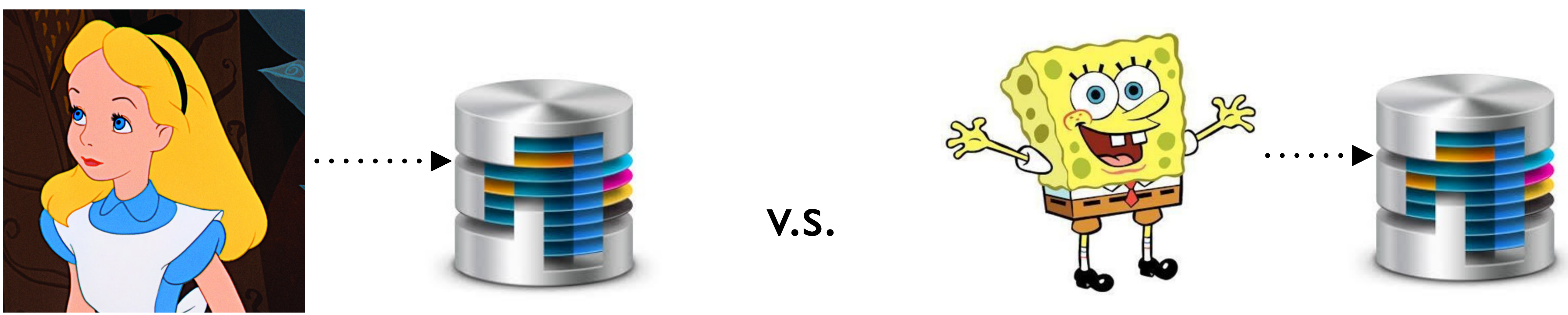
Hypothetical Scenario

- Suppose a study release based on a private database that "most mathematicians are heavy coffee drinkers."
- Knowing Alice is a mathematician, the data analyst infers that Alice is likely a heavy coffee drinker and may have certain health risks

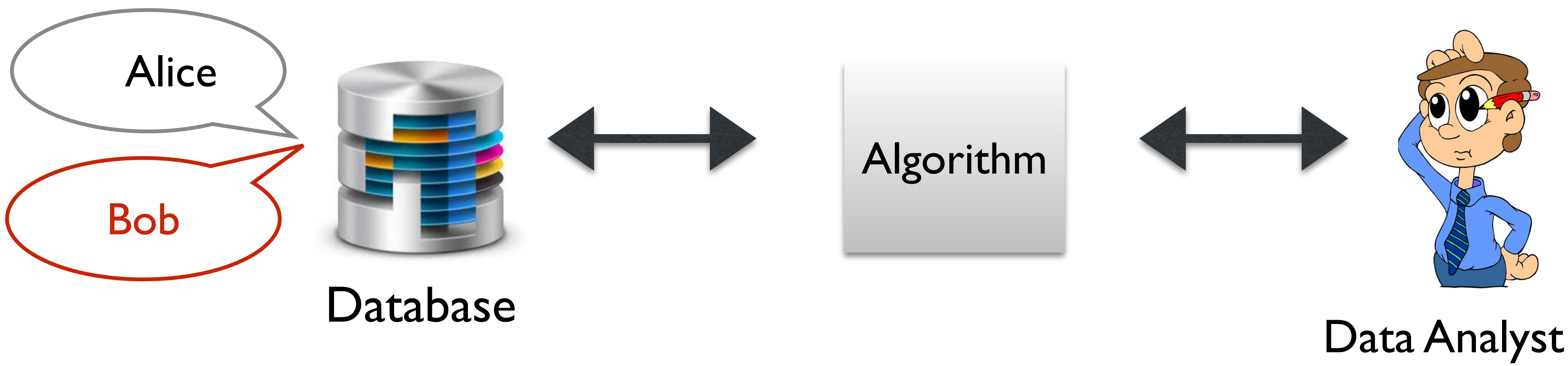
Do you consider this study as a privacy violation on Alice?

Privacy (Attempt 2)

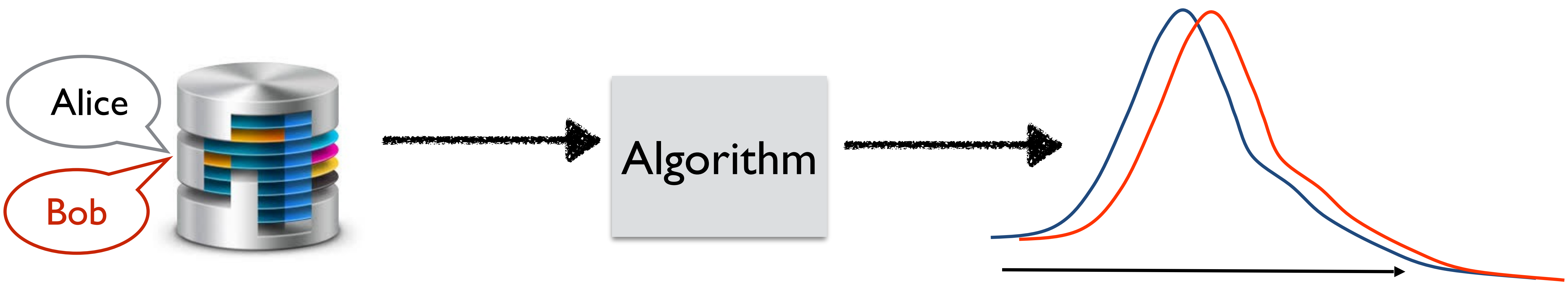
"An analysis is private if the data analyst knows almost no more about Alice after the analysis than analyst would have known had he conducted the same analysis on an identical database with Alice's data replaced."



Differential Privacy as a Stability Notion



Stability: the data analyst learns (approximately) same information if any row is replaced by another person of the population



Sensitive Data Set
(e.g., medical records)

Output Distribution
(e.g., noisy statistics)

"An algorithm is *differentially private* if changing a single record does not alter its output distribution by much."
[DN03, DMNS06]

Differential Privacy

[DN03, DMNS06]

$$D = \begin{array}{|c|}\hline D_1 \\ \hline D_2 \\ \hline D_3 \\ \hline \dots \\ \hline D_n \\ \hline \end{array} \qquad D' = \begin{array}{|c|}\hline D_1 \\ \hline D_2 \\ \hline \boxed{D'_3} \\ \hline \dots \\ \hline D_n \\ \hline \end{array}$$

D and D' are *neighbors* if they differ by at most one row

Definition: A (randomized) algorithm A is ε-differentially private if for all neighbors D, D' and every event $S \subseteq \text{Range}(A)$

$$\Pr[A(D) \in S] \leq \exp(\varepsilon) \Pr[A(D') \in S]$$

"If a bad event is very unlikely when I'm not in the database (D), then it is still very unlikely when I am in the database (D')."

Nice Properties of Differential Privacy

- Privacy loss measure (ε)
  - Bounds the cumulative privacy losses across different computations and databases
- Resilience to arbitrary post-processing
  - Adversary's background knowledge is irrelevant
  - Immune to re-identification attacks
- Compositional reasoning
  - Programmability: construct complicated private analyses from simple private building blocks

Practical Deployment

Topics we will cover

Basic Definitions and Techniques

- Reconstruction attacks
- Laplace/Exponential/Gaussian mechanisms
- Composition

Private synthetic data

- DP GAN
- Private Multiplicative Weights

Practical deployment

- Local DP
- Distributed/Shuffling Models

Machine Learning

- (Non)-convex opt
- Deep learning with DP

Techniques beyond DP

- Statistical Validity
- Game theory
- …

Basic Techniques:
introducing randomness

- Laplace mechanism
- Randomized Response

"When I pour cream in my coffee,
I see randomness with intention."

—Costis Daskalakis

Answer a Counting Query

	Smoke	Lung Cancer	Diabetes	OCD
patient_1	1	1	1	1
patient_2	1	0	0	1
patient_3	1	1	0	1
patient_4	0	0	1	0
…	…	…	…	…
patient_n	1	1	1	0

Counting query: *How many people that satisfy some specified property?*

For example: what is the fraction of people that "Smoke" and have "Lung Cancer"?

Change of One Single Person

	Smoke	Lung Cancer	Diabetes	OCD
patient_1	1	1	1	1
patient_2	1	0	0	1
patient_3	1	1	0	1
patient_4	0	0	1	0
…	…	…	…	…
patient_n	1	1	1	0

Suppose we change any single one person's data (Moving from D to D')

$$| f(D) - f(D') | \leq 1$$

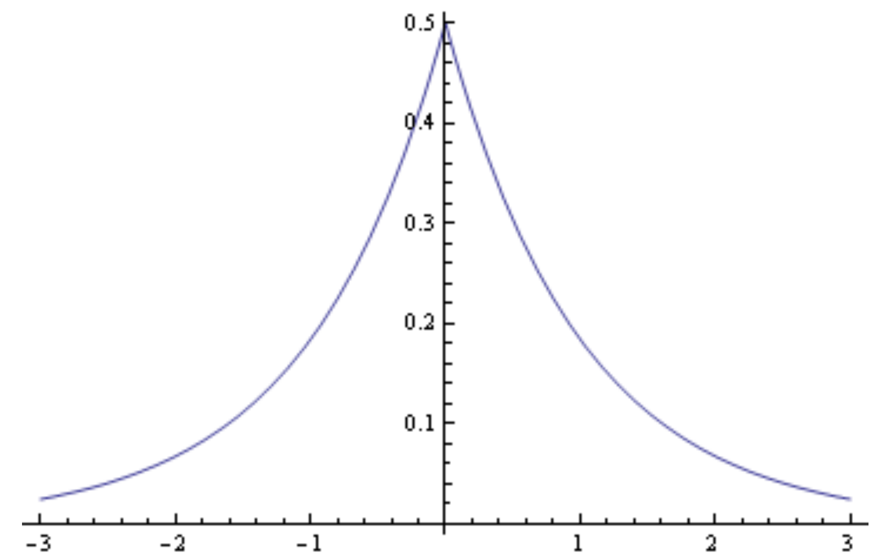
Intuition: hide the influence of any single individual through noise addition

Laplace Mechanism

- Laplace distribution: $X \sim Lap(b)$

$$p(x|b) = \frac{1}{2b} \exp\left(-\frac{|x|}{b}\right)$$

$$\mathbb{E}[|X|] = b$$



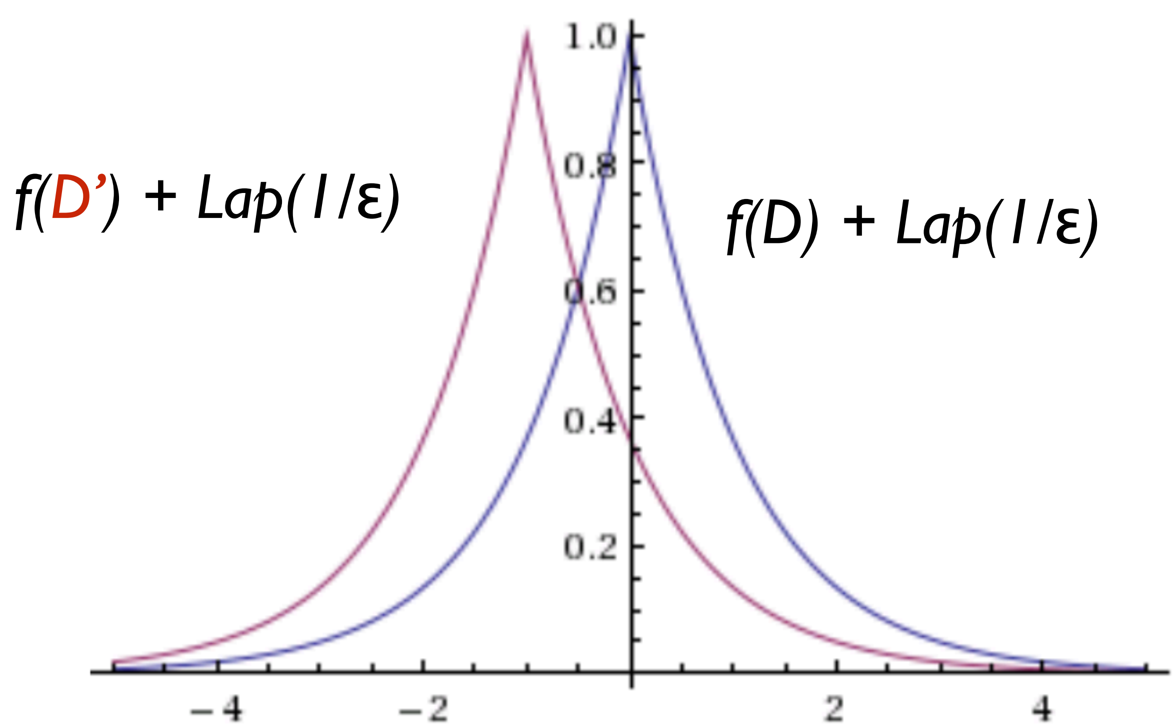
- Laplace mechanism with input dataset D, privacy parameter ε

$$f(D) + Lap(1 / \varepsilon)$$

Privacy Guarantee

Theorem: The Laplace Mechanism satisfies
ε-differential privacy.

- Proof by picture



Randomized Response [Warner 65]

- Data may not be readily available; Need to conduct survey
- Data subjects may be privacy sensitive
- Goal: collect accurate aggregate statistics
  (not about any single individual)

Have you ever done XYZ?

Randomized Response

- Flip a coin
  - If heads, answer truthfully;
  - If tails, then flip another coin: answer "Yes" if heads, "No" otherwise

Plausible Deniability: if your answer is "yes", there is no way of knowing your true status.

In-class activity

- We will follow the steps of randomized response to collect noisy answers of the question *“have you ever cheated in an exam?”*

First step: random seed

- Get a piece of paper or open up a text file in your computer
- Recall a phone number you have remembered since your childhood; write it down.
- We will use last two digits of the phone number
(if your number is 762-2341, the last two digits "41")

Second step: compute your report

Question: have you ever cheated in an exam?

- If the first digit is an even number: then report truthfully
- If the first digit is an odd number: look at the second digit
  - If the second digit is even, report "yes"
  - If the second digit is odd, report "no"

  - If your answer is "yes", indicate yes
  - Also, place your answer in the Zoom poll.

Final: how to compute an estimate?

- For any person i:
- X_i in $\{0,1\}$: true answer
- Y_i in $\{0,1\}$: reported answer

- $\Pr[Y_i = X_i] = 3/4$
- $\Pr[Y_i = 1 - X_i] = 1/4$

The expected value of person i's reported answer

$$\mathbf{E}[Y_i] = (3/4)\,X_i + (1/4)\,(1 - X_i) = \frac{X_i}{2} + 1/4$$

- $\hat{Y}$: fraction of reported "yes" = 60%
- Estimate for true fraction of "cheating"
  $2(\hat{Y} - 1/4)$ = 70%

- Flip a coin
  - If heads, answer truthfully;
  - If tails, then flip another coin: answer “Yes” if heads, “No” otherwise

Pr[say “yes” | truth = “yes”] / Pr[say “yes” | truth = “no”] = 3

- If truth is yes, will say yes with probability 3/4.
- If truth is no, will say yes with probability 1/4.

Pr[say “no” | truth = “no”] / Pr[say “no” | truth = “yes”] = 3

Local Differential Privacy

Definition: A (randomized) algorithm A is ε-locally differentially private if for any two individuals x and y, and every S ⊆ Range(A)

$$\Pr[A(x) \in S] \leq e^{\varepsilon} \Pr[A(y) \in S]$$

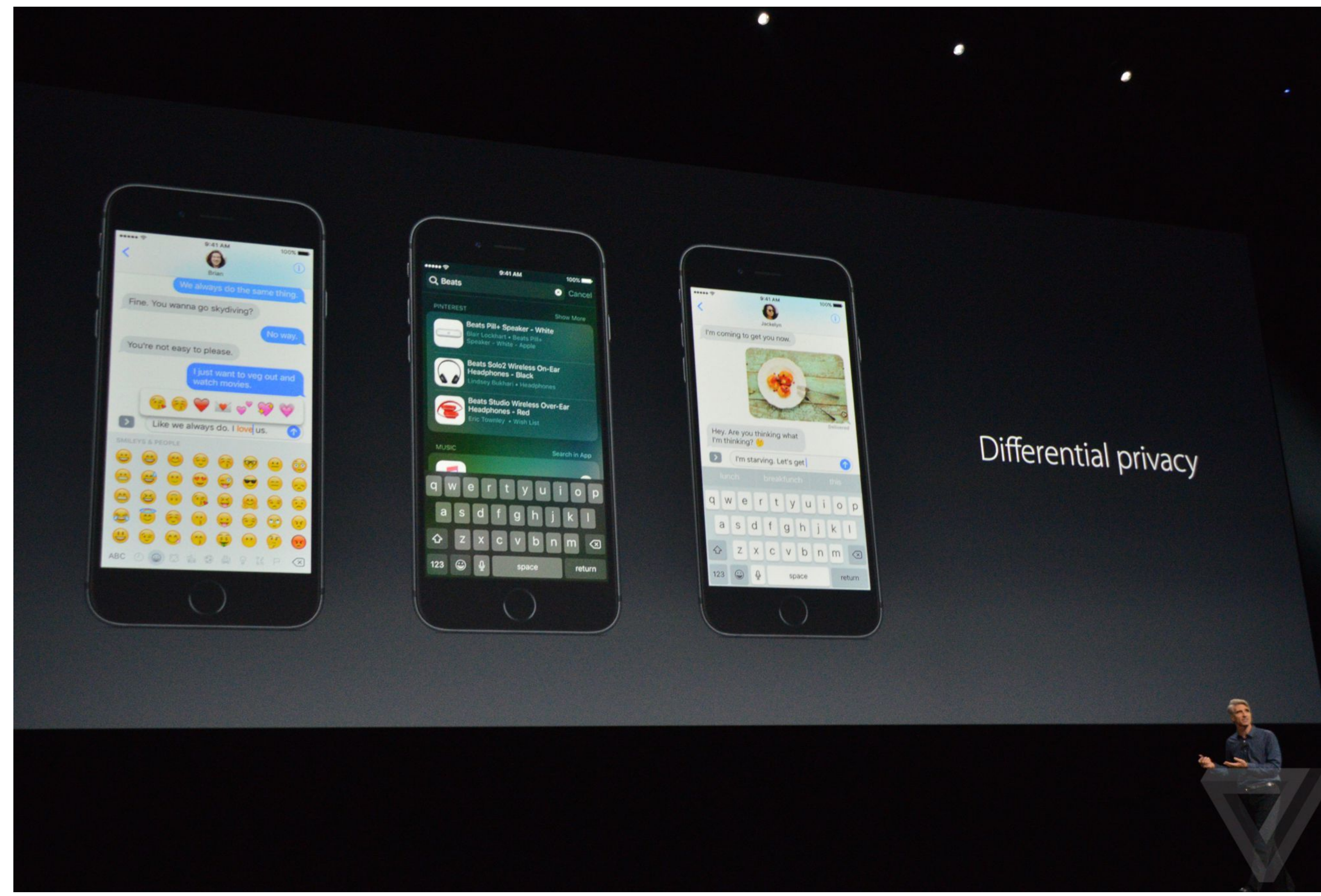
Privacy loss in randomized response: $\varepsilon = \ln(3) \approx 1.098$

Utility: when computing the fraction of n people who has "XYZ",

the error $\approx 1/\sqrt{n}$

Applications

Google chrome



See you on Weds

Reading assignment will be posted today