Course Introduction:
Foundations of Privacy

Instructor: Steven Wu

https://foundpriv.github.io/

Introduction: Steven Wu

- CMU SCS faculty (ISR/MLD/HCII)
- Interests: machine learning & algorithms
  - Privacy/Fairness
  - Social and economic aspects of machine learning
- Outside of work:
  - Basketball, rock bouldering/climbing, biking, snowboarding
  - Other sports I am pretty bad at:
    golf, squash, (beach) volleyball…
- Personal website: zstevenwu.com
  - *I am a world-leading procrastinator and always behind my emails.*

Communication

Canvas (Preferred)

- We will use Canvas for all assignments and grades.
- Please also post all questions on Canvas as discussions

Email

- If you email me (the instructor), please put [FoundPriv Course] in your email title.
- You will probably get a better/faster response if you email the TA, especially for questions regarding grading.

Who are the TA's?

TAs

- Justin Whitehouse
    - Email: jwhiteho@andrew.cmu.edu
- Another mysterious TA??
    - TBA later this week

Course Website

https://foundpriv.github.io/

This Course
https://foundpriv.github.io/

- Topics
  - Formal models on privacy, fairness, and cryptograhy
  - Algorithmic techniques
- Skills you will work on
  - Formal reasoning about privacy and algorithms
  - (Lightweight) 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)

- Lecture format
  - Live lectures with slides/iPad
  - Lecture will be recorded and become available on zoom

Coursework

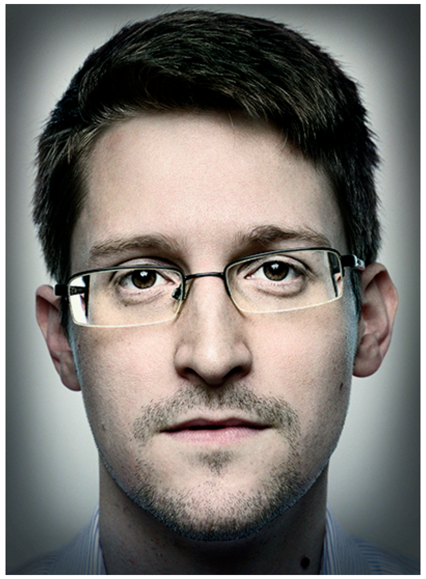
- Lecture prep and in-class work
- Homework (4 assignments)
  - Collaboration allowed
  - Write up your solutions and acknowledge collaborators
- Final exam (details TBA)
  - *Most likely take-home*

Grading

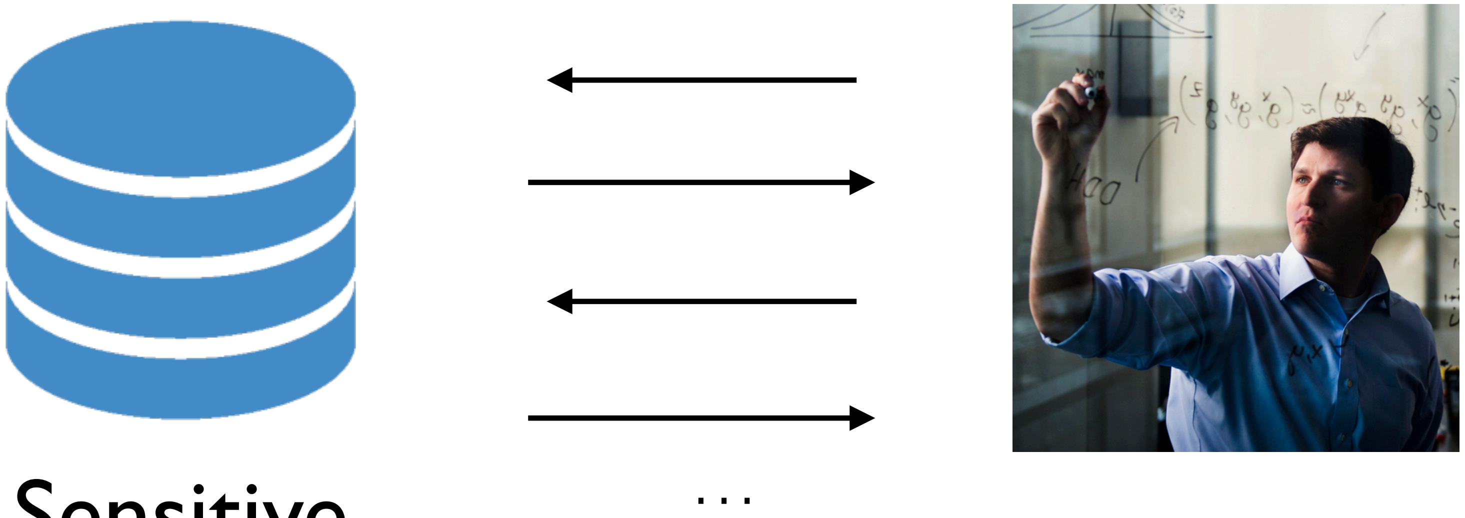
- In-class participation: 20%
  - *Soft* rule of thumb: speak up at least 10 times during the whole course
- 4 homework assignments: 60%
  - 5 late days allowed
- Final: 20%

Questions?

What this course is *not* about



Privacy-Preserving Data Analysis



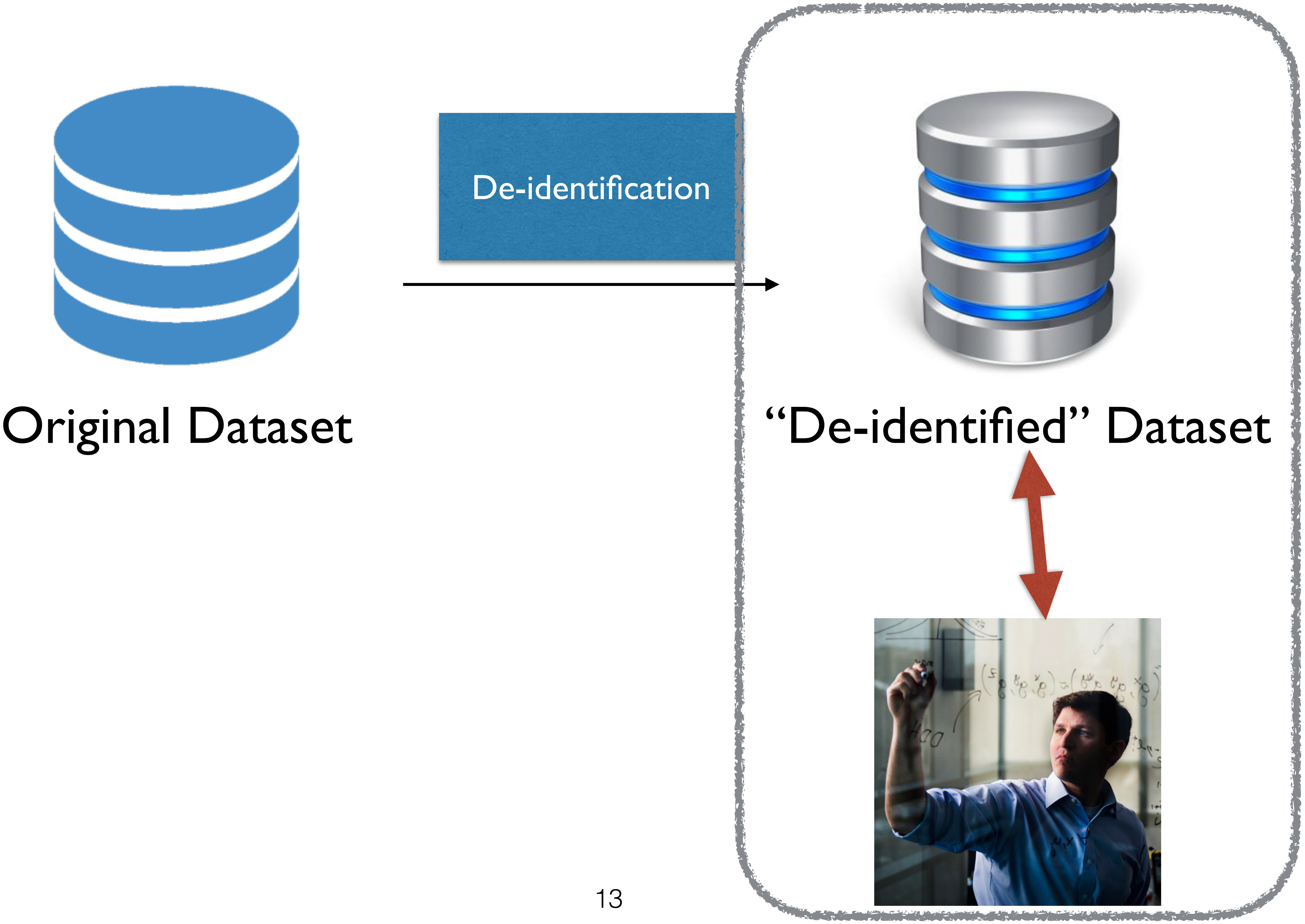
Sensitive Database

...

Data Scientist

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

Anonymization?



Original Dataset

De-identification

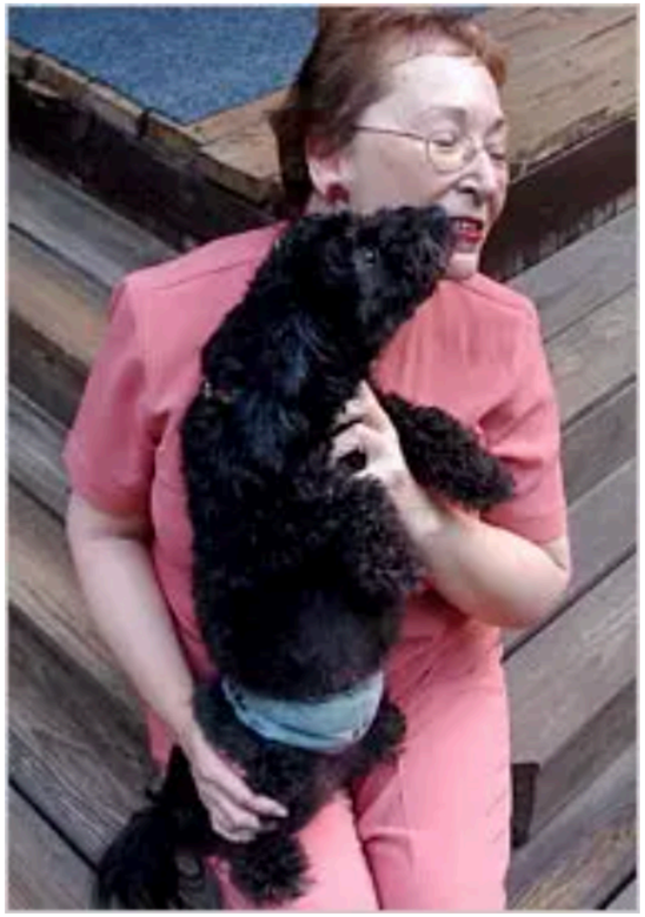
“De-identified” Dataset

The New York Times

A Face Is Exposed for AOL Searcher No. 4417749

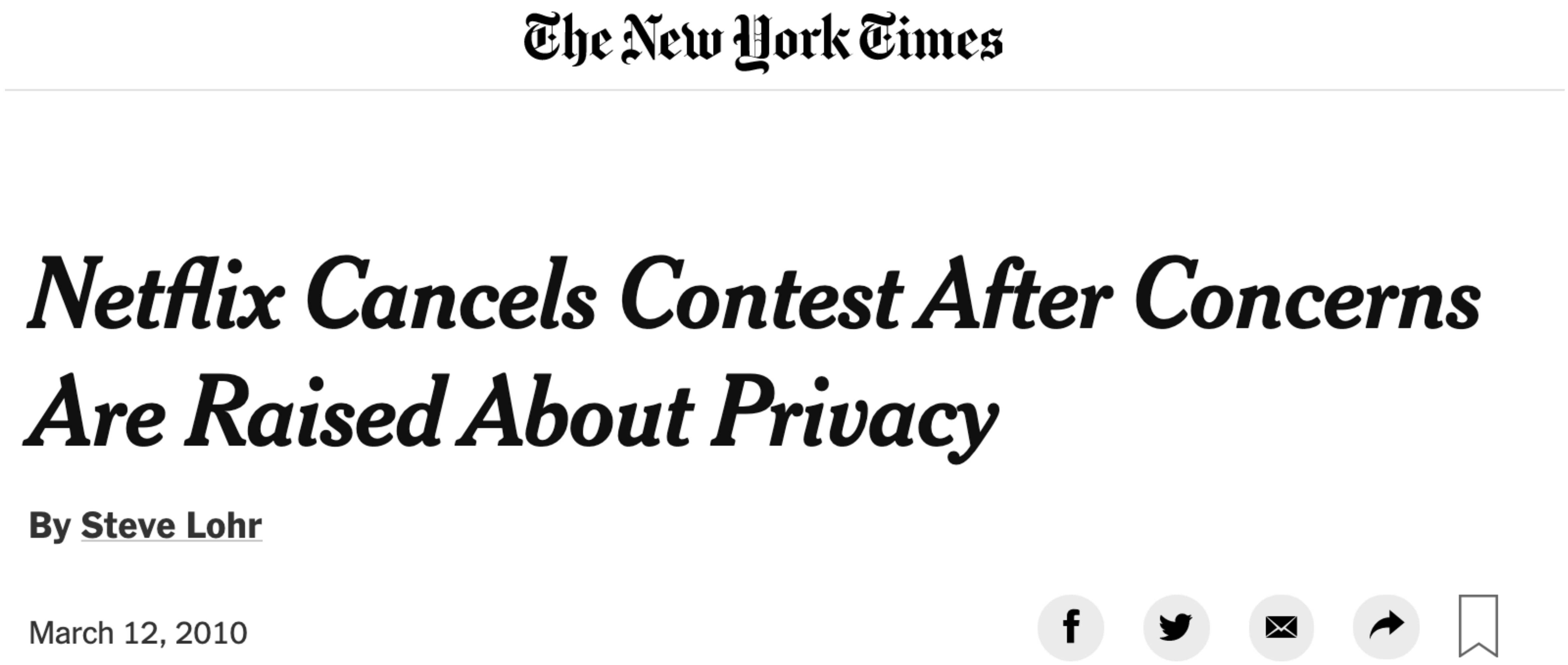
By Michael Barbaro and Tom Zeller Jr.

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

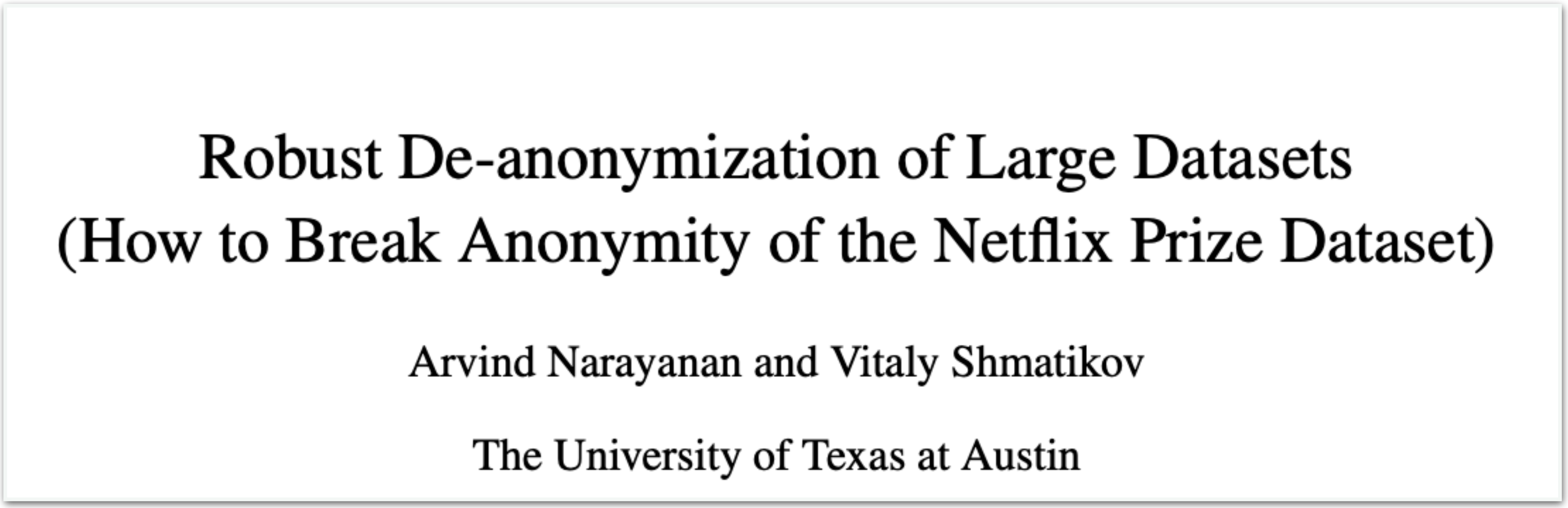


The New York Times

Netflix Cancels Contest After Concerns Are Raised About Privacy

By Steve Lohr

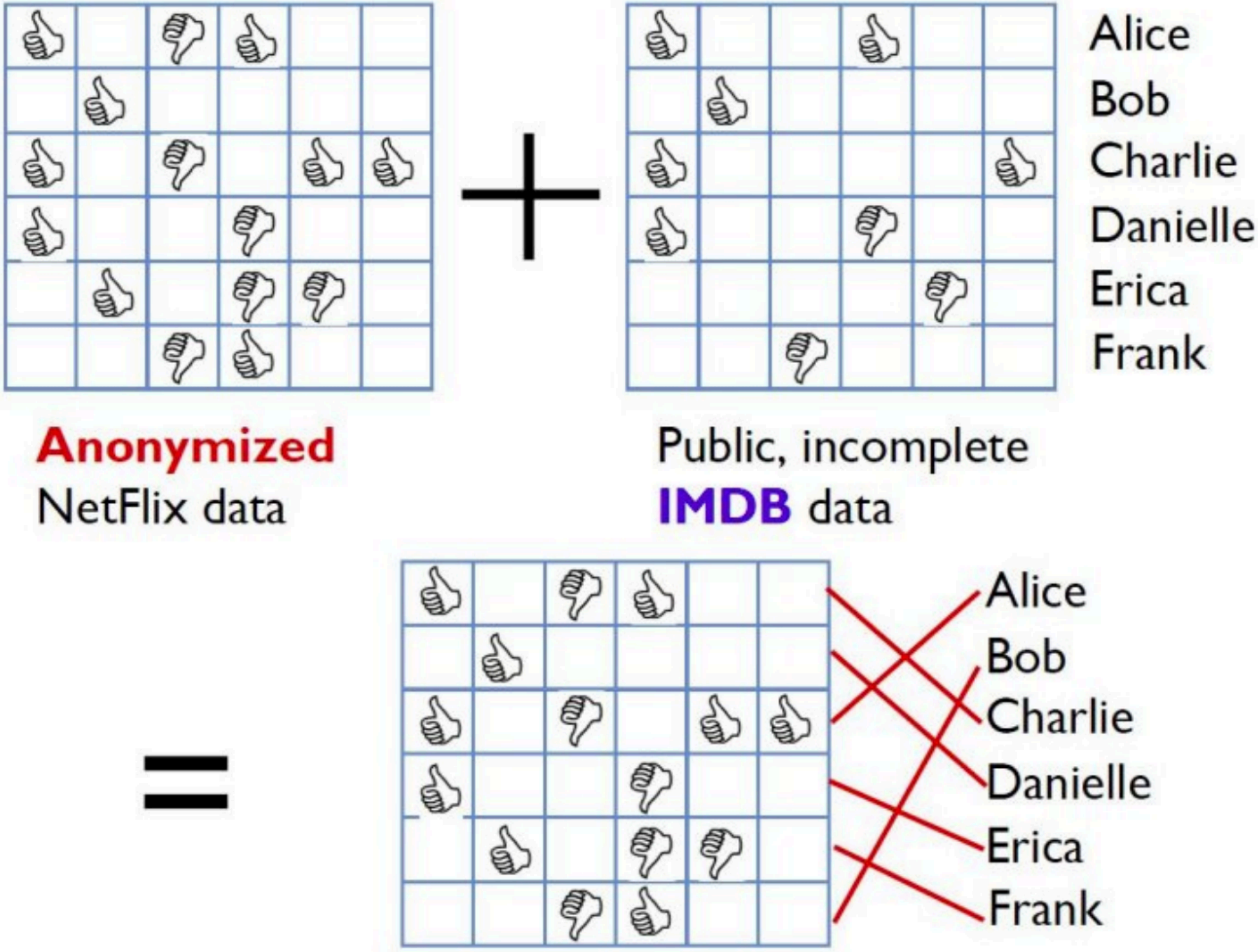
March 12, 2010



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



👍		👎	👍		
	👍				
👍		👎		👍	👍
👍			👎		
	👍		👎	👎	
		👎	👍		

Anonymized NetFlix data

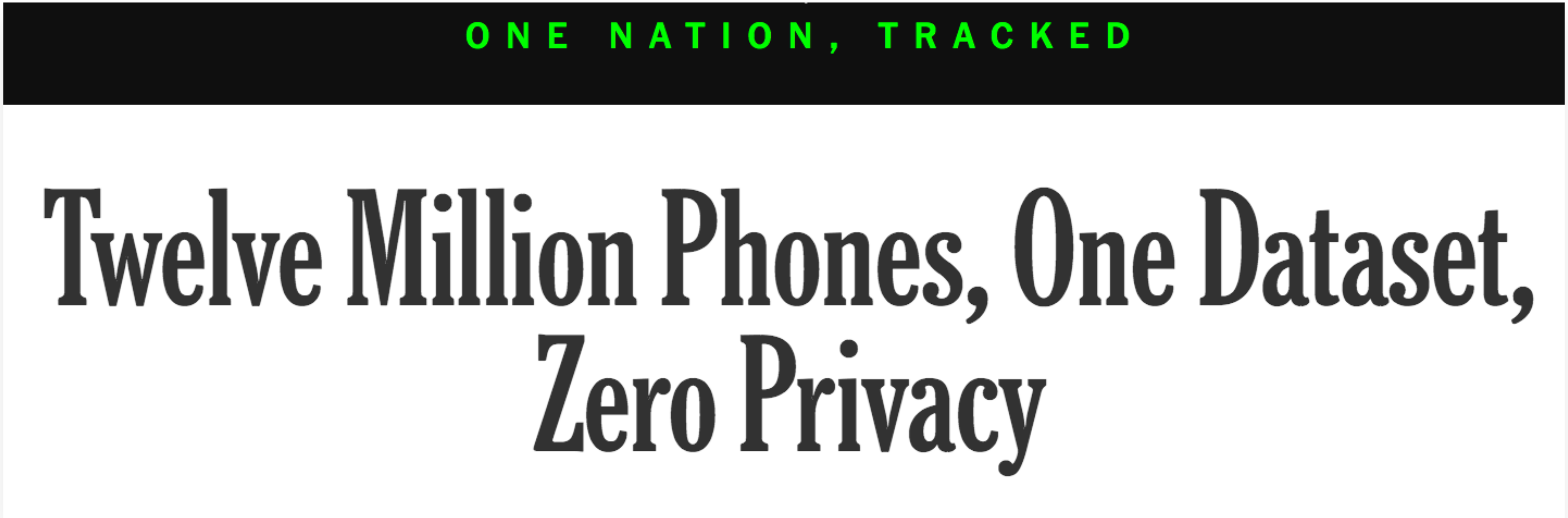
👍			👍			Alice
	👍					Bob
👍					👍	Charlie
👍			👎			Danielle
				👎		Erica
		👎				Frank

Public, incomplete **IMDB** data

👍		👎	👍		
	👍				
👍		👎		👍	👍
👍			👎		
	👍		👎	👎	
		👎	👍		

Identified NetFlix Data

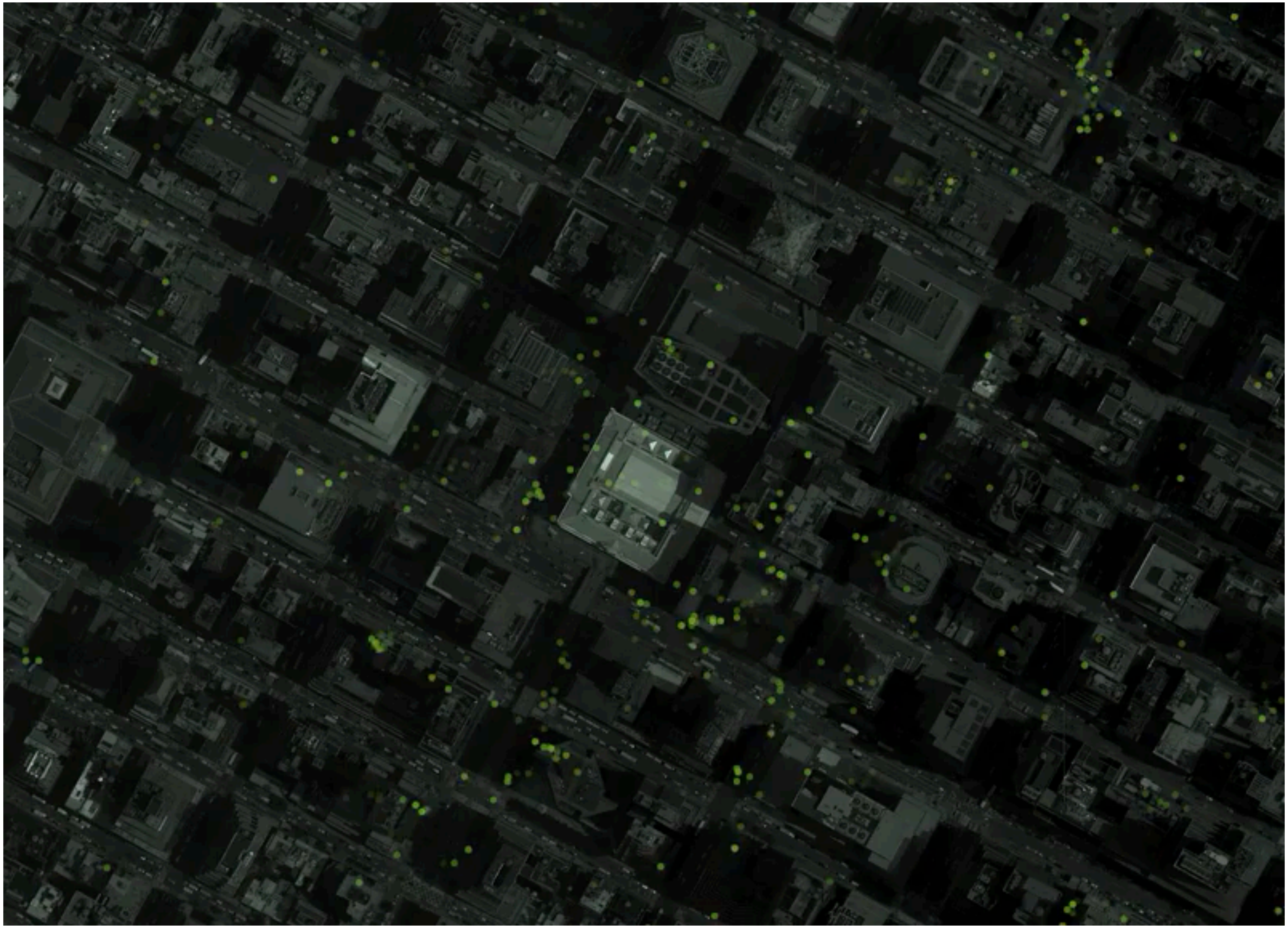
Image credit: Arvind Narayanan



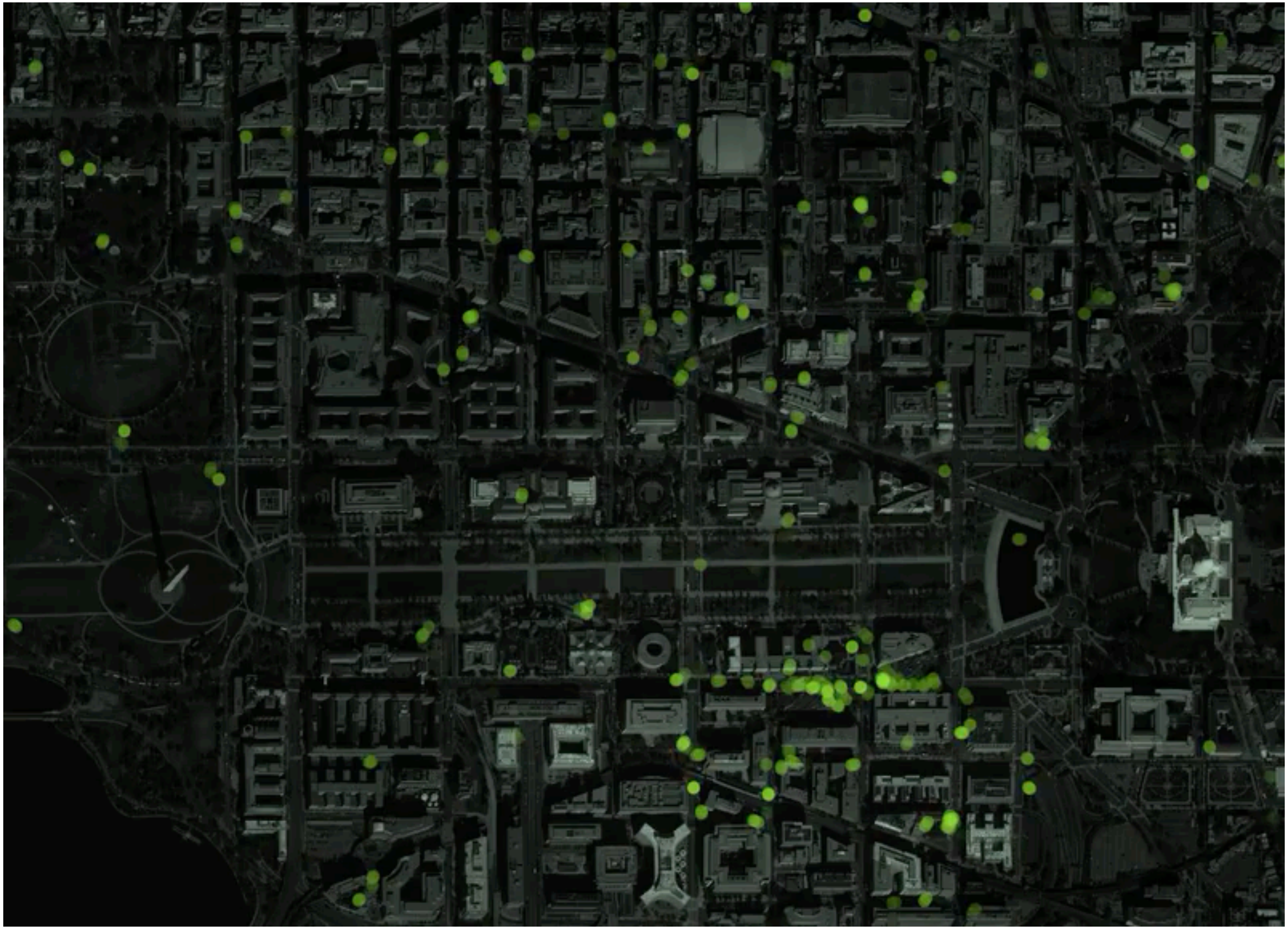
ONE NATION, TRACKED

Twelve Million Phones, One Dataset, Zero Privacy

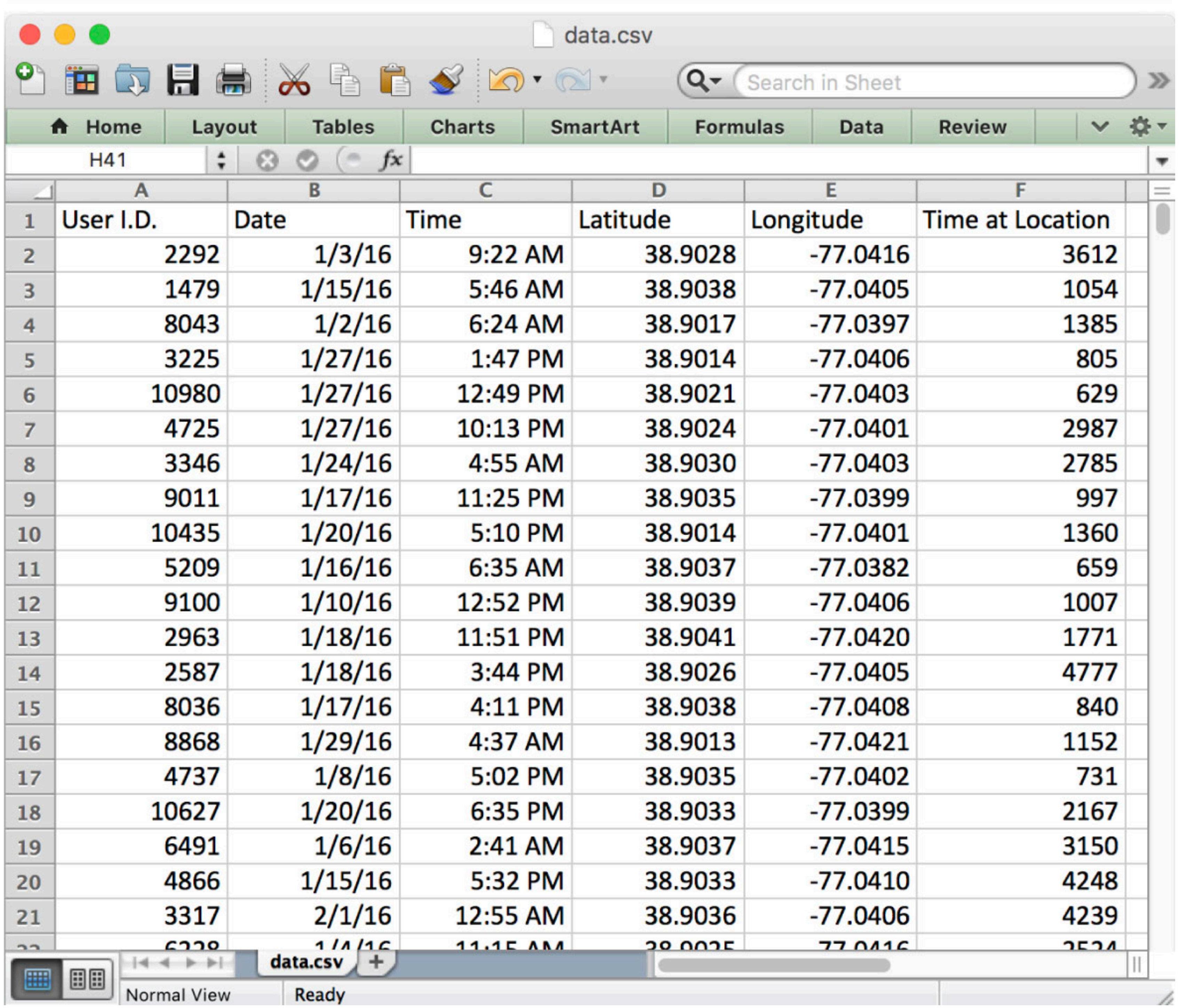
https://www.nytimes.com/interactive/2019/12/19/opinion/location-tracking-cell-phone.html



A typical day at Grand Central Terminal in New York City



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



	A	B	C	D	E	F
1	User I.D.	Date	Time	Latitude	Longitude	Time at Location
2	2292	1/3/16	9:22 AM	38.9028	-77.0416	3612
3	1479	1/15/16	5:46 AM	38.9038	-77.0405	1054
4	8043	1/2/16	6:24 AM	38.9017	-77.0397	1385
5	3225	1/27/16	1:47 PM	38.9014	-77.0406	805
6	10980	1/27/16	12:49 PM	38.9021	-77.0403	629
7	4725	1/27/16	10:13 PM	38.9024	-77.0401	2987
8	3346	1/24/16	4:55 AM	38.9030	-77.0403	2785
9	9011	1/17/16	11:25 PM	38.9035	-77.0399	997
10	10435	1/20/16	5:10 PM	38.9014	-77.0401	1360
11	5209	1/16/16	6:35 AM	38.9037	-77.0382	659
12	9100	1/10/16	12:52 PM	38.9039	-77.0406	1007
13	2963	1/18/16	11:51 PM	38.9041	-77.0420	1771
14	2587	1/18/16	3:44 PM	38.9026	-77.0405	4777
15	8036	1/17/16	4:11 PM	38.9038	-77.0408	840
16	8868	1/29/16	4:37 AM	38.9013	-77.0421	1152
17	4737	1/8/16	5:02 PM	38.9035	-77.0402	731
18	10627	1/20/16	6:35 PM	38.9033	-77.0399	2167
19	6491	1/6/16	2:41 AM	38.9037	-77.0415	3150
20	4866	1/15/16	5:32 PM	38.9033	-77.0410	4248
21	3317	2/1/16	12:55 AM	38.9036	-77.0406	4239
22	6228	1/4/16	11:15 AM	38.9025	-77.0416	2524

"De-identified data isn't."

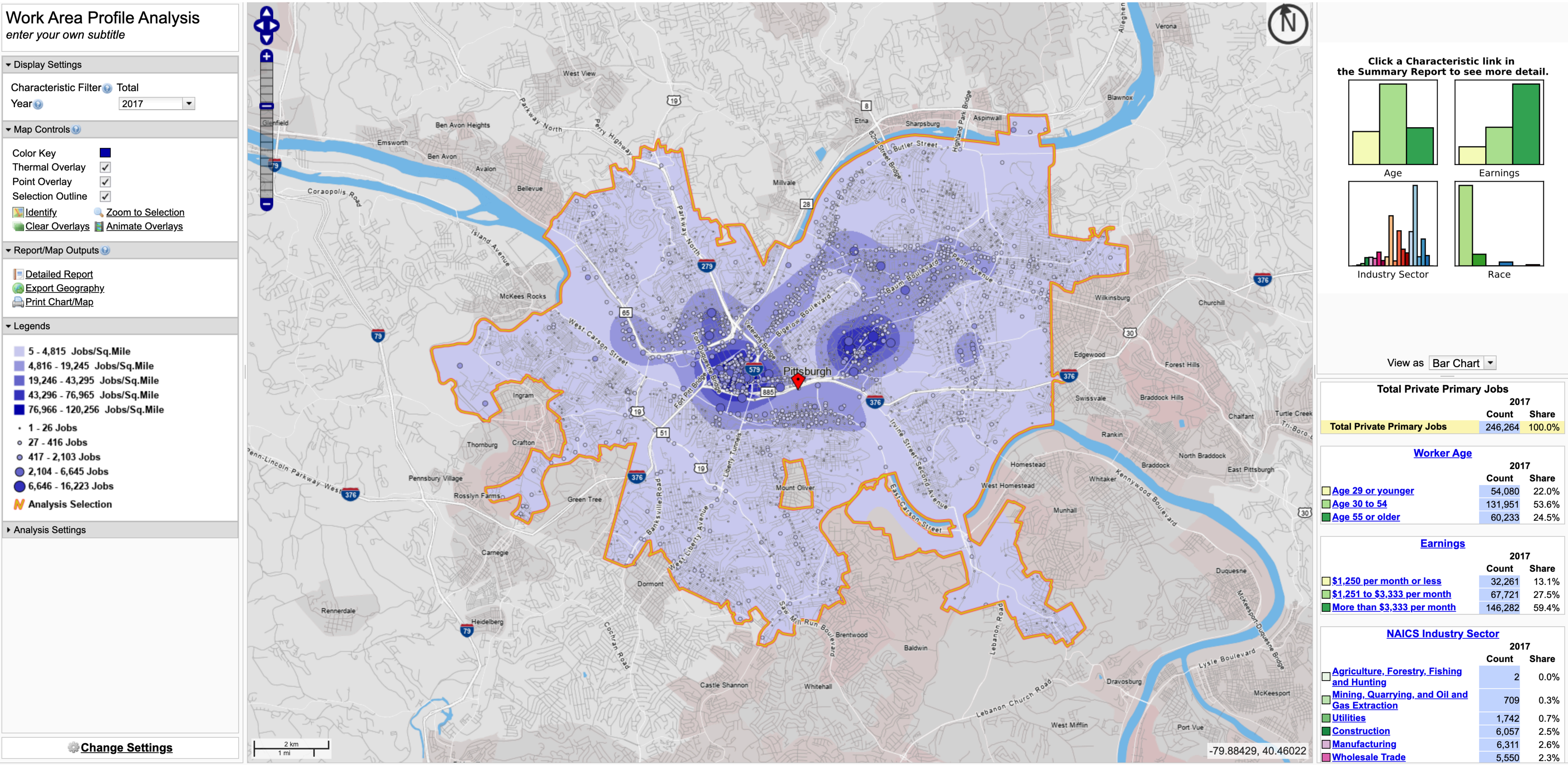
— Cynthia Dwork

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



Work Area Profile Analysis

enter your own subtitle

▾ Display Settings

Characteristic Filter: Total
Year: 2017

▾ Map Controls

Color Key ■
Thermal Overlay ☑
Point Overlay ☑
Selection Outline ☑

Identify · Zoom to Selection
Clear Overlays · Animate Overlays

▾ Report/Map Outputs

Detailed Report
Export Geography
Print Chart/Map

▾ Legends

- 5 - 4,815 Jobs/Sq.Mile
- 4,816 - 19,245 Jobs/Sq.Mile
- 19,246 - 43,295 Jobs/Sq.Mile
- 43,296 - 76,965 Jobs/Sq.Mile
- 76,966 - 120,256 Jobs/Sq.Mile
- 1 - 26 Jobs
- 27 - 416 Jobs
- 417 - 2,103 Jobs
- 2,104 - 6,645 Jobs
- 6,646 - 16,223 Jobs
- Analysis Selection

▸ Analysis Settings

Change Settings

Click a Characteristic link in the Summary Report to see more detail.

View as: Bar Chart

Total Private Primary Jobs

	2017	
	Count	Share
Total Private Primary Jobs	246,264	100.0%

Worker Age

	2017	
	Count	Share
Age 29 or younger	54,080	22.0%
Age 30 to 54	131,951	53.6%
Age 55 or older	60,233	24.5%

Earnings

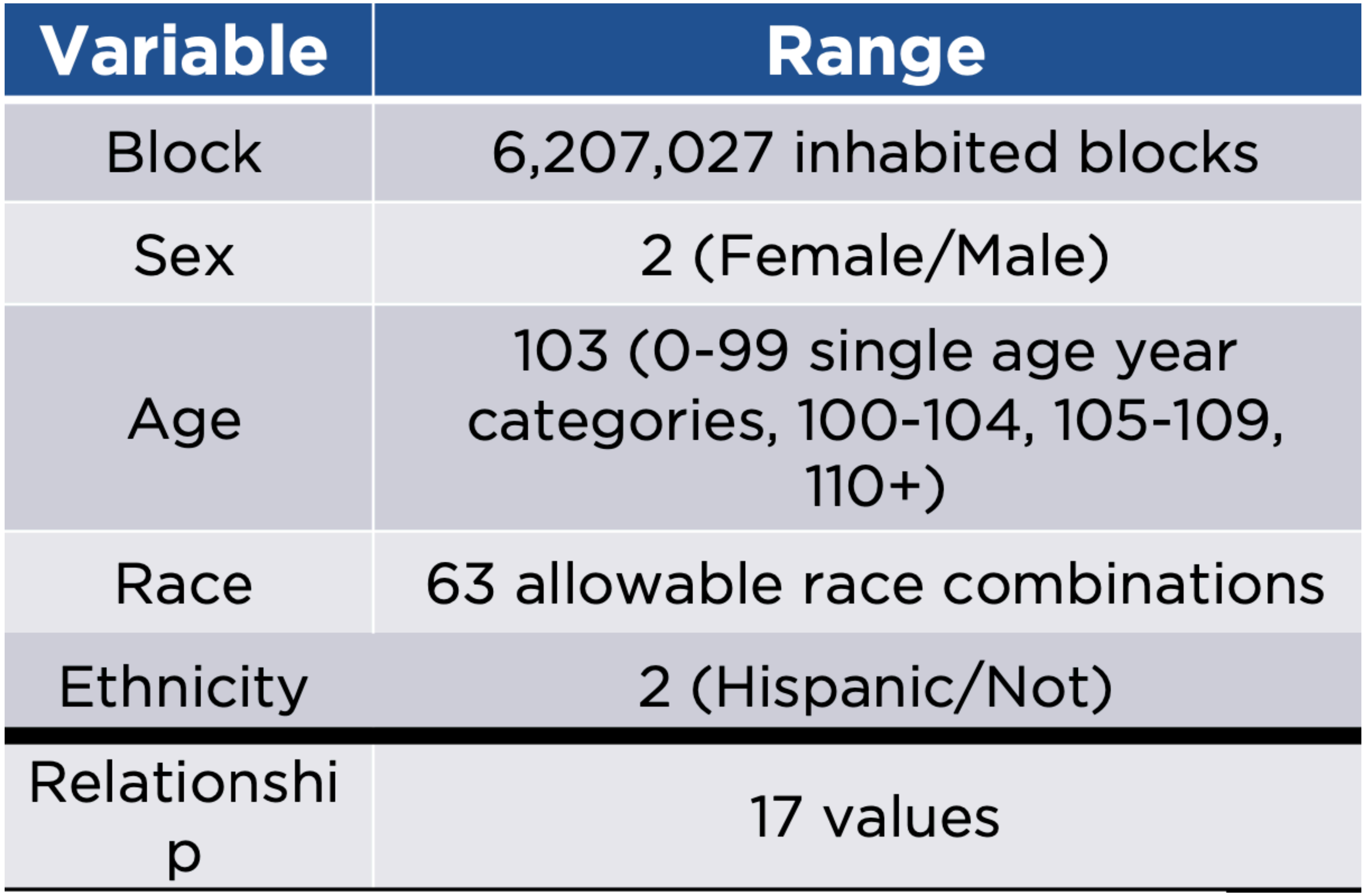
	2017	
	Count	Share
$1,250 per month or less	32,261	13.1%
$1,251 to $3,333 per month	67,721	27.5%
More than $3,333 per month	146,282	59.4%

NAICS Industry Sector

	2017	
	Count	Share
Agriculture, Forestry, Fishing and Hunting	2	0.0%
Mining, Quarrying, and Oil and Gas Extraction	709	0.3%
Utilities	1,742	0.7%
Construction	6,057	2.5%
Manufacturing	6,311	2.6%
Wholesale Trade	5,550	2.3%

Data Collected in 2010 Decennial Census

308,745,538 people × 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

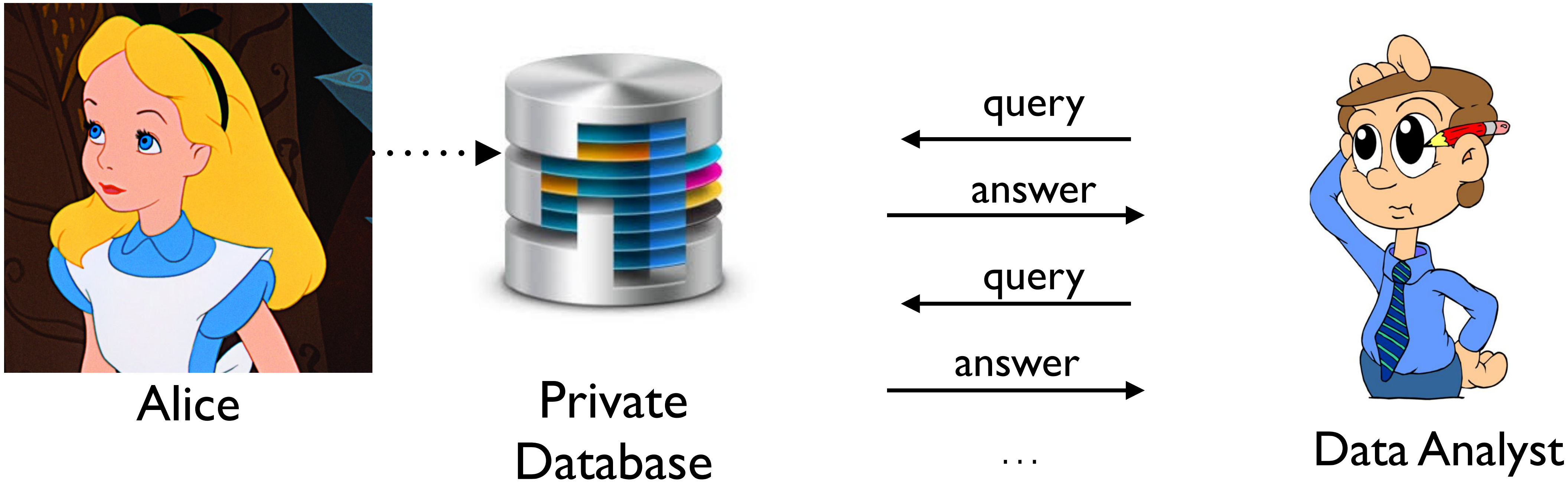
Fundamental law of information [Dinur & Nissim]:
"Overly accurate" estimates of "too many" statistics is non-private.

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?



Alice

Private Database

query

answer

query

answer

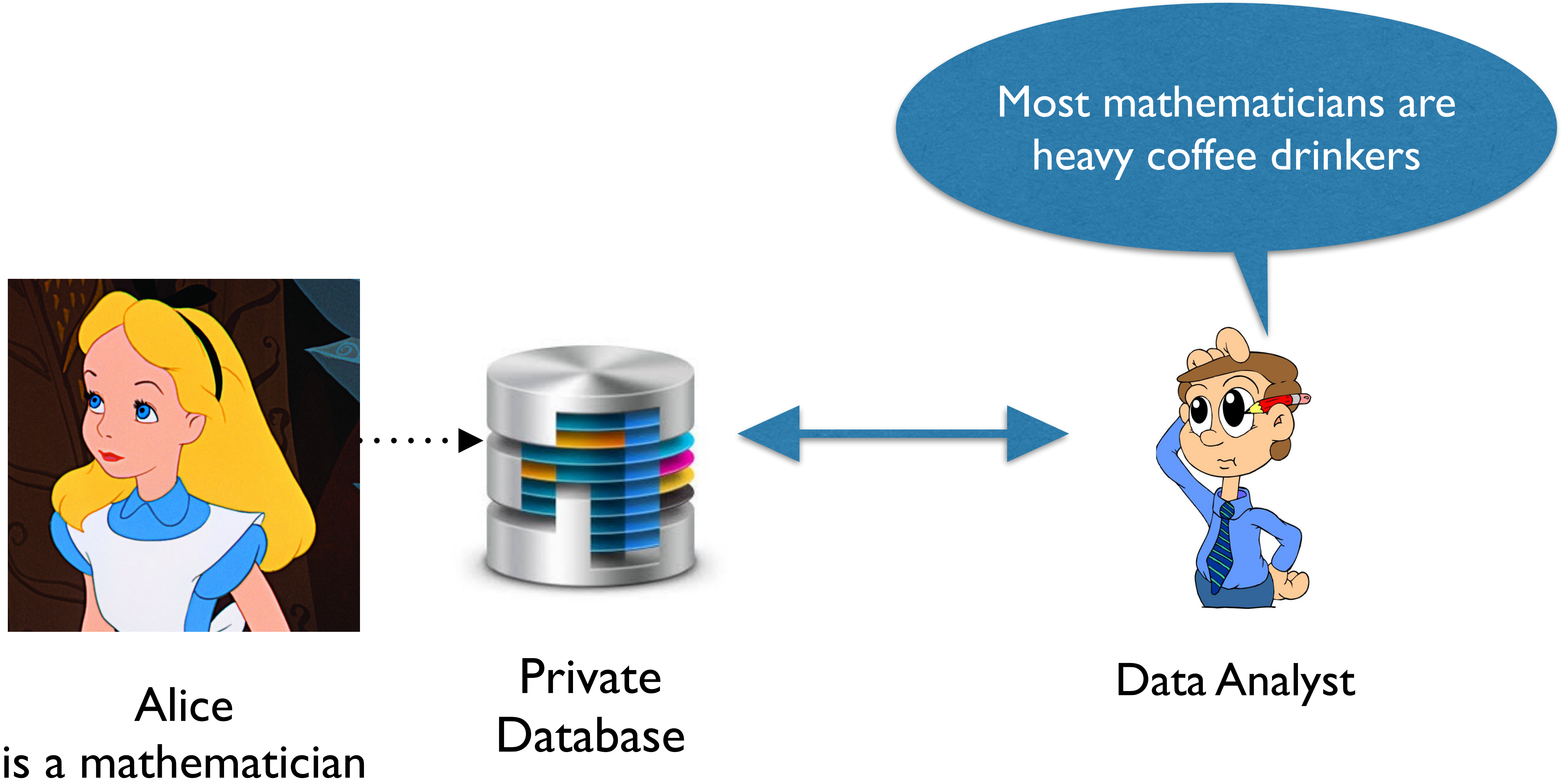
. . .

Data Analyst

Privacy Attempt 1:

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

Hypothetical Scenario



Most mathematicians are heavy coffee drinkers

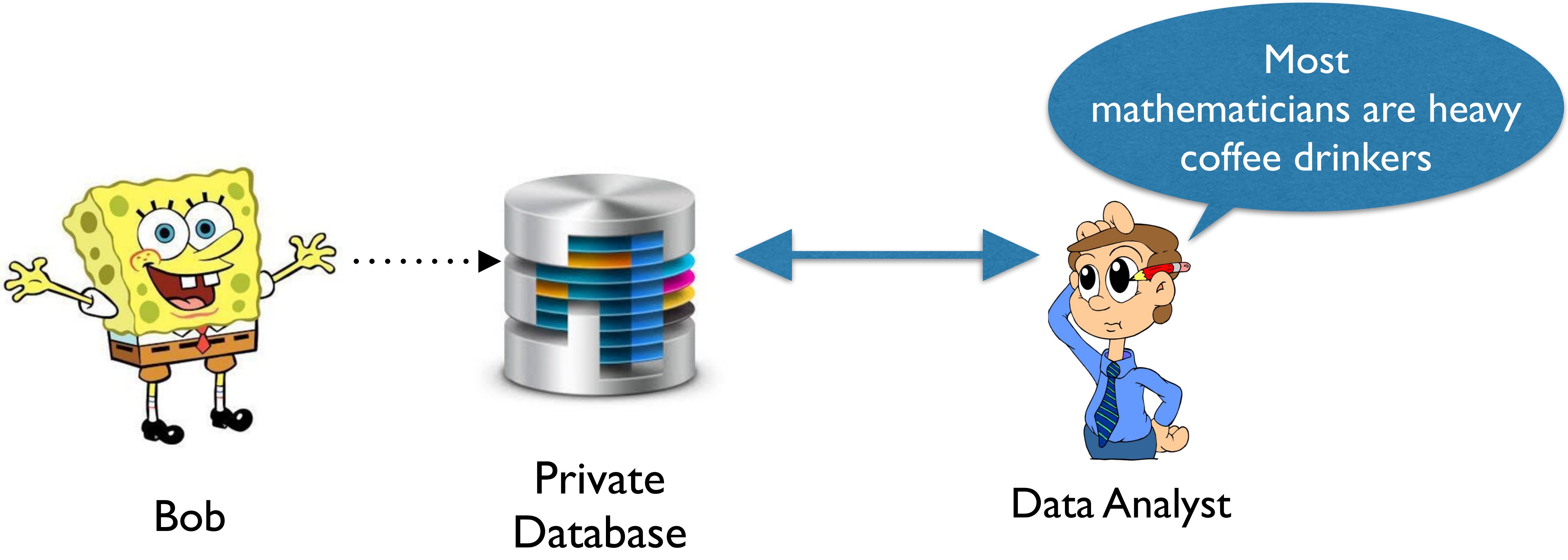
Alice
is a mathematician

Private
Database

Data Analyst

Was Alice's privacy violated?

Replace Alice by Another Random Person



Most mathematicians are heavy coffee drinkers

Bob

Private Database

Data Analyst

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

Hypothetical Scenario

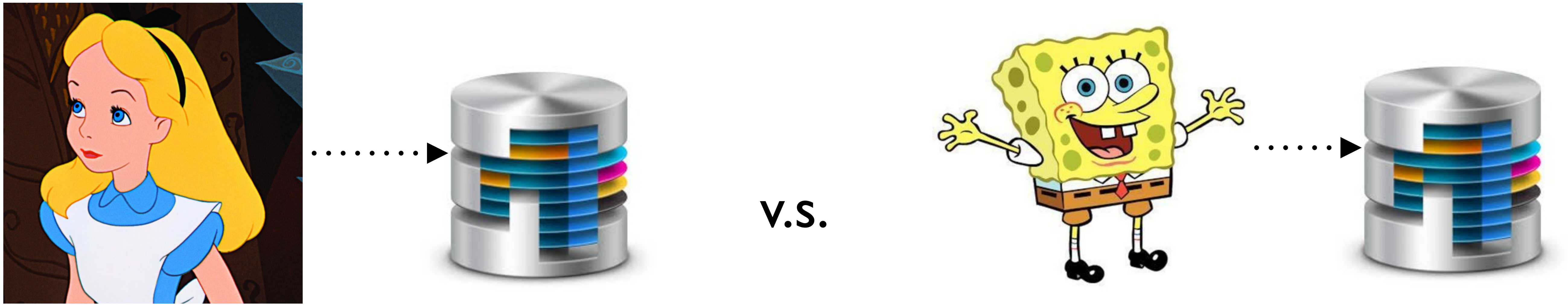
- Suppose a study release based on a private database that "most mathematicians are heavy coffee drinkers."

- 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?

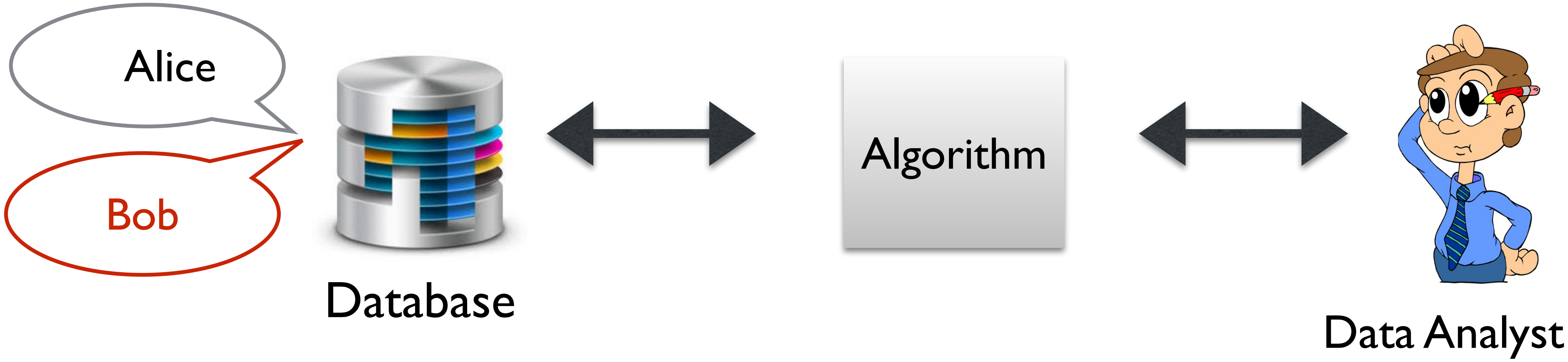
Privacy (Attempt 2)

"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."

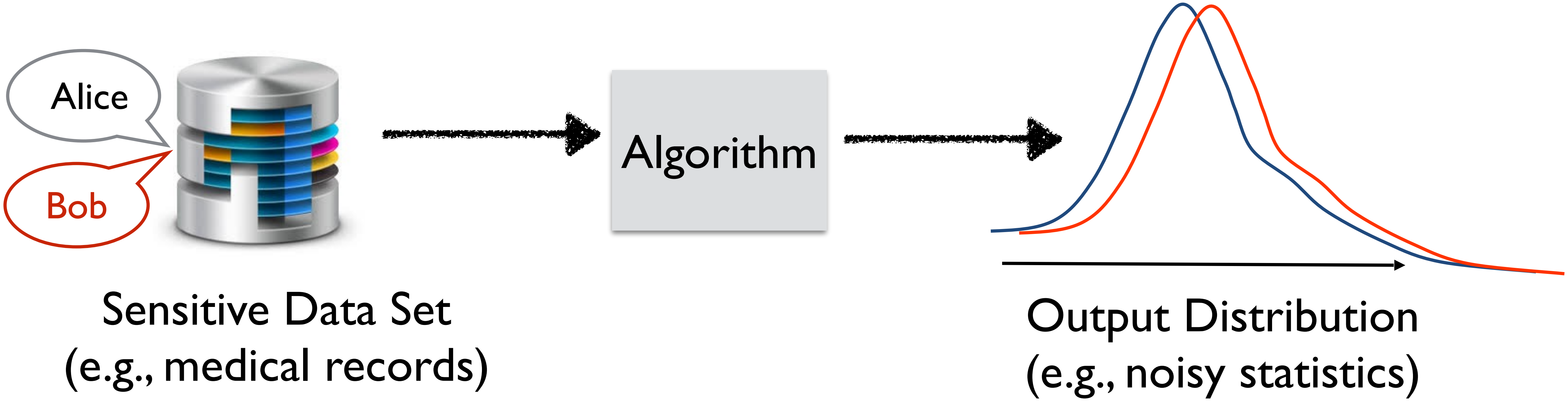


v.s.

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



Alice

Bob

Algorithm

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]

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

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 $S \subseteq \mathrm{Range}(A)$

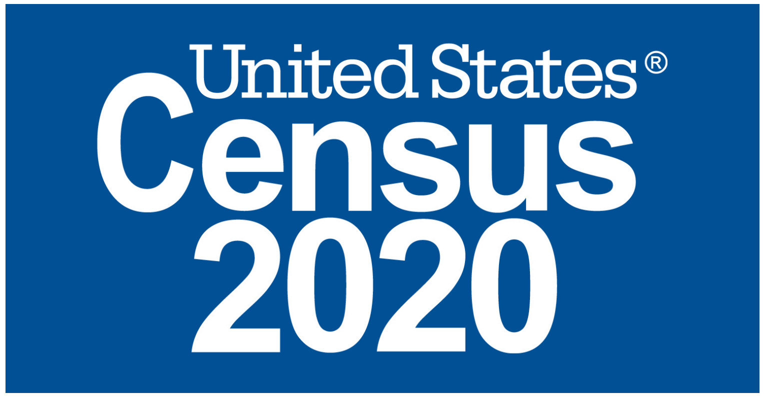
$$\Pr[A(D) \in S] \leq \exp(\varepsilon) \Pr[A(D') \in S]$$

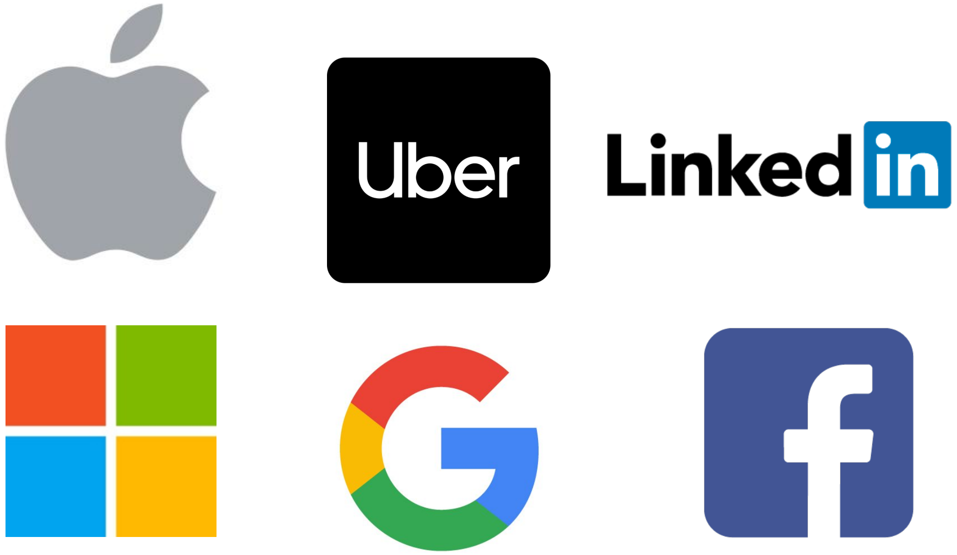
“If a bad event is very unlikely when I’m not in the database (D), then it is still very unlikely when I am in the database (D').”

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

Machine Learning

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

Private synthetic data

- DP GAN

Algorithmic fairness

- Fairness in machine learning
- Definitions and mitigation

Cryptographic approaches

- Secret sharing scheme

Basic Techniques:
introducing randomness

- Randomized Response

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

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

—Costis Daskalakis

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.

For Students over Zoom

- If your randomized response is “yes”, indicate yes with the emoji
- Otherwise use the “no” emoji

Final: how to compute an estimate?

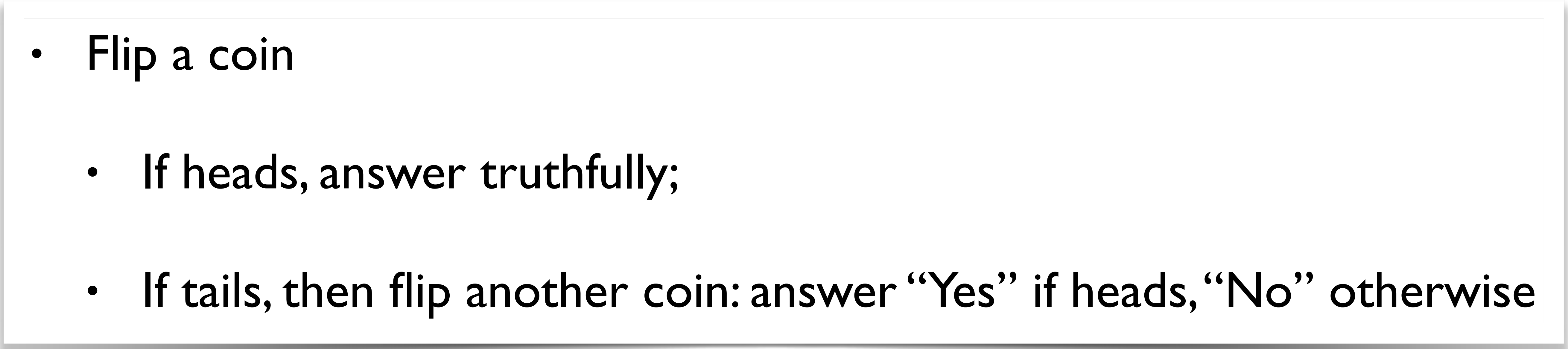
- For any person i:
- X_i in $\{0,1\}$: true answer
- Y_i in $\{0,1\}$: reported answer

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

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" = 30/58
- Estimate for true fraction of "cheating" $2(\hat{Y} - 1/4)$ = 52%



- 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

Applications

Google chrome

See you on Weds

TODO:

- Finish Reading Assignment before Class (Posted on course calendar page: https://foundpriv.github.io/calendar/)