Privacy for Pandemics

An Introduction to Differential Privacy for Human Mobility and Interaction Data

Katrina Ligett - Computer Science - Hebrew University

According to the ministry, the soldier visited a seamstress at the Or Yehuda Mall on February 26, between 12 and 12:30 p.m.

In addition, she traveled on the Kavim bus line 59 from the Jasmine/Hatsav bus stop in Or Yehuda to the bus stop at the David Shipman Bridge Station in Tel Aviv, on February 26 and 27 between 7:50 a.m. and 8:30 a.m.

On February 26 she made the return journey from Tel Aviv to Or Yehuda between 2:15 p.m. and 3 p.m. and on February 27 between 2 p.m. and 2:40 p.m.

In addition, she was in the Red Pirate toy store in Or Yehuda on February 24 from 3 p.m. until 10 p.m.; February 25 from 12:30 p.m. until 12:30 a.m.; and February 26 from 3 p.m. until 9:30 p.m.

The ministry said the other two diagnosed Sunday were members of one family who had returned to Israel from Italy on Thursday and live in a community in southern Israel, where they have been in quarantine since their return.

Work life

arXiv > cs > arXiv:1911.10137

Search... Help | Advance

Computer Science > Data Structures and Algorithms

[Submitted on 22 Nov 2019]

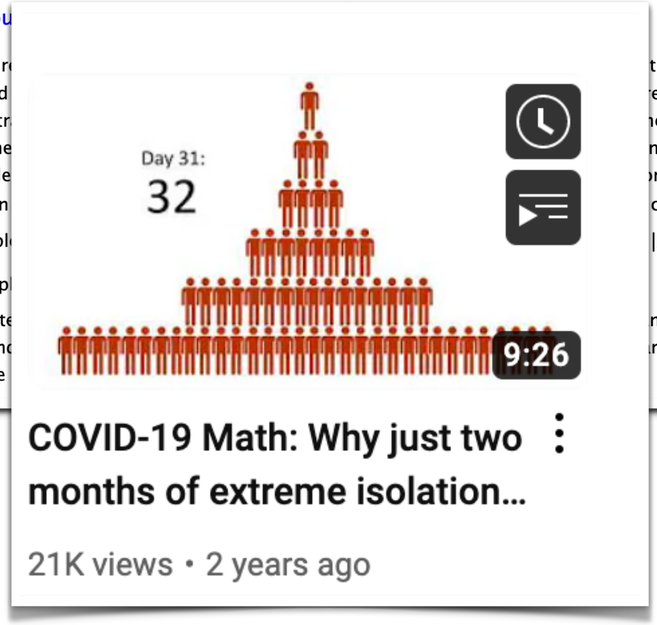
Privately Learning Thresholds: Closing the Exponential Gap

Haim Kaplan, Katrina Ligett, Yishay Mansou

We study the sample complexity of learning thr… that each labe… the information of one individual and we would …ential privacy… means that any single labeled example in the tr…nesis. This pro… recently; unlike the non-private case, where the…n the desired… private learning the sample complexity must de…on et al. (STOC… $\Omega(\log^* |X|)$ on the sample complexity and Bun… complexity O… this gap significantly, almost settling the sampl… $|X|)^2)$ on th…

present an improved version with sample comp…

Our algorithm is constructed for the related inte…nd smallest i… selecting an input-dependent hash function and…rithmically; th… interior point of which can be used to generate …



Parameters

Israel population	9,140,469	
Default cap vaccine coverage for each subpopulatic	0.80	0.176 <- for the 0-19 bin, corresponding to .8 coverage of the 16-19 group
Fraction of the population immune as of Jan 3	0.102	
Israel detects 1 infection for every	2.0	
initial number of infectious cases	130,000	* R0 and initial number number of infections should probably be set to induce initial approximate caseload doubling in two weeks, t
1st shot reduces transmission after 11 days by	0.000	
2nd shot reduces transmission after 7 days by	0.000	
1st shot prevents infection after 11 days by	0.85	
2nd shot prevents infection after 7 days by	0.95	

Instructions

Parameters you are most likely to change are highlighted in green. One thing to be sensitive to is that you probably want the combination of parameters chosen to induce caseload doubling every two weeks
To simulate an intervention or the effects of the UK mutation, you should change the R0 level per day.
If you get data or have different predictions of the way the doses will be distributed by decile, you can replace the values in the blue cells accordingly. (See distribution assumptions below.)

Assumptions

We assume people who will die do so two weeks after infection
The number of doses given in a day is a parameter. The model then distributes those doses as follows: First, all needed second doses are administered. Then, so long as no decile of the population has reached 80% coverage
If some deciles have reached 80% coverage (a parameter above), the remaining does are spread among the remaining deciles with the same relative distribution as Jan 6.
I assume the vaccination takes 11 days after the first dose to get to 85% effectiveness, and 7 days after the second dose to get to 95% effectiveness. I don't phase in those thresholds
I assume all people are equally likely to meet all other people—no structure to the social network or differentiation in centrality by age group
I am using a mu of .125—every day, .125 of the infectious become removed. This, combined with the R0, induces the beta (the beta—the daily risky encounter rate; if you are happy with mu, you can ignore this)
No randomness—all events happening according to their expectation
The ability of the vaccine to reduce infectiousness is modeled as follows:
I = prevI - standard recoveries + beta * prevI * prevS * (1 - percent of pop with effective first dose * effectiveness of first dose in reducing infection)
R = prevR + standard recoveries + beta * prevI * prevS * (percent of pop with effective first dose * effectiveness of first dose in reducing infection)
This is based on a standard SIR model. Patients move straight from Susceptible to Infectious without a delay of a couple of days
I'm assuming age deciles are exposed to the disease in ratios proportional to their presence in the population, not to their historical percentage of cases. This is a somewhat conservative approach.

Days	Total doses given today	Replication	R0 * S	Effective F	Governme	Beta	Mu	Susceptible Fr	Infectious F	Removed	S+I+R	New Case	Susceptible	Infectious	Removed	Ne
20-Dec	7600										0					
21-Dec	23700				1.22						0					
22-Dec	42600				1.235						0					
23-Dec	60600				1.25						0					
24-Dec	74200				1.25						0					
25-Dec	38000				1.27						0					
26-Dec	32000				1.26						0					
27-Dec	97000				1.25						0					
28-Dec	115000				1.23						0					
29-Dec	152000				1.22						0					
30-Dec	151000				1.21						0					

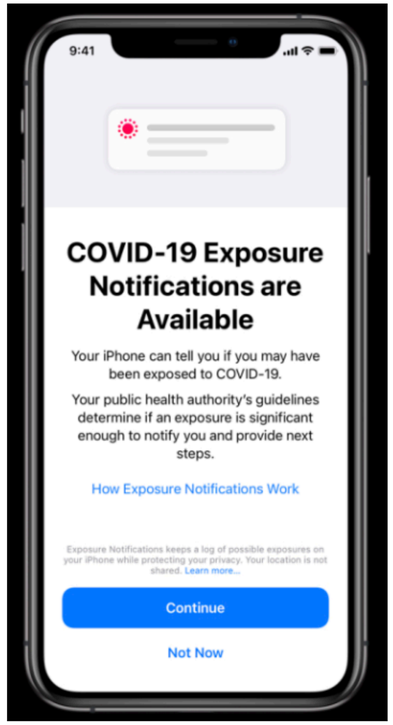
I’m a privacy nerd

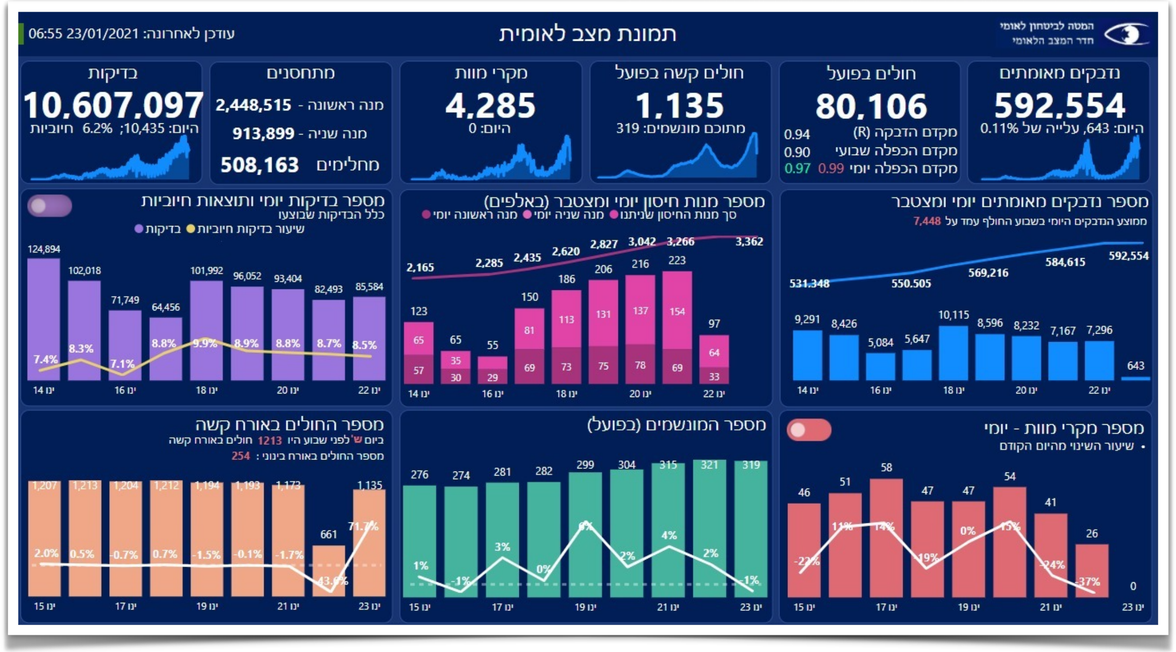
(but I get it)

Israel to track mobile phones of suspected coronavirus cases

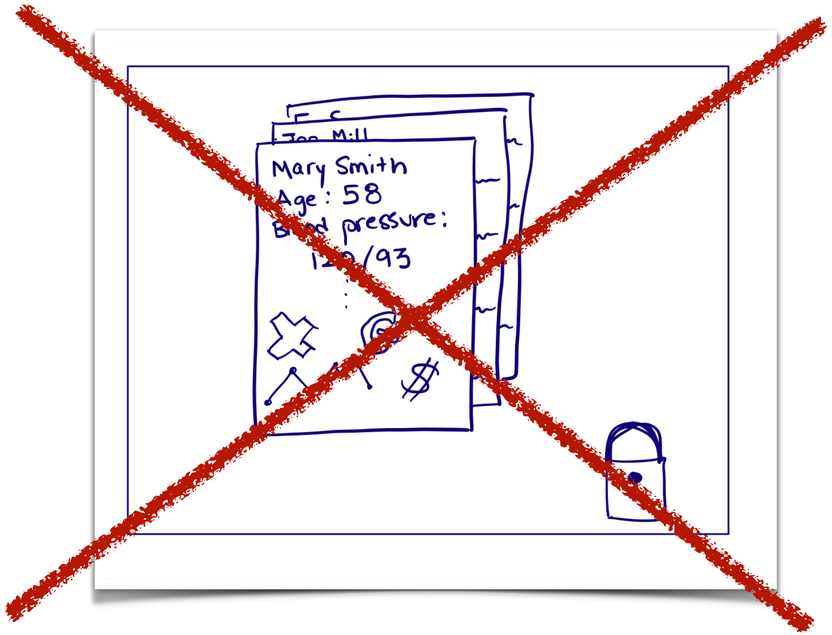
Measure to use counterterrorism technology passed by cabinet, leading to outcry over privacy

- Coronavirus – latest updates
- See all our coronavirus coverage



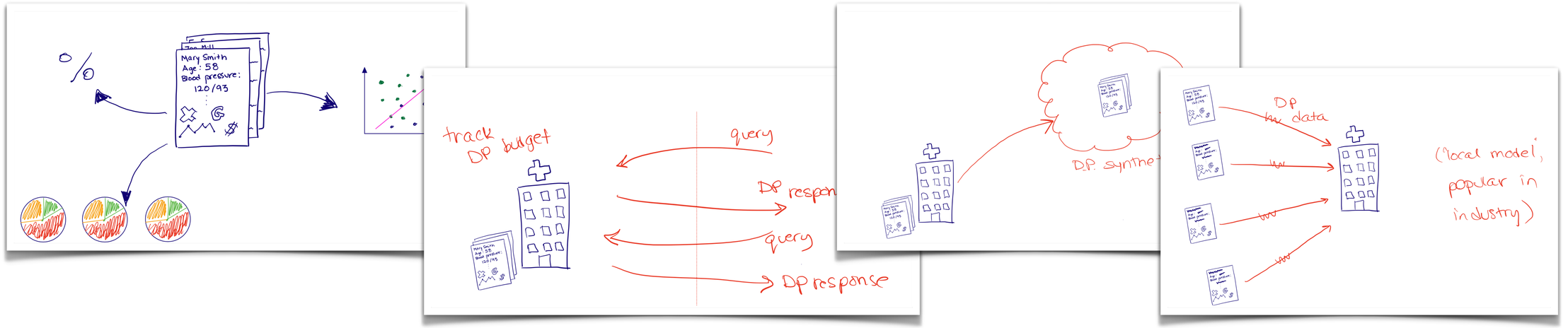


This talk: differential privacy (DP)



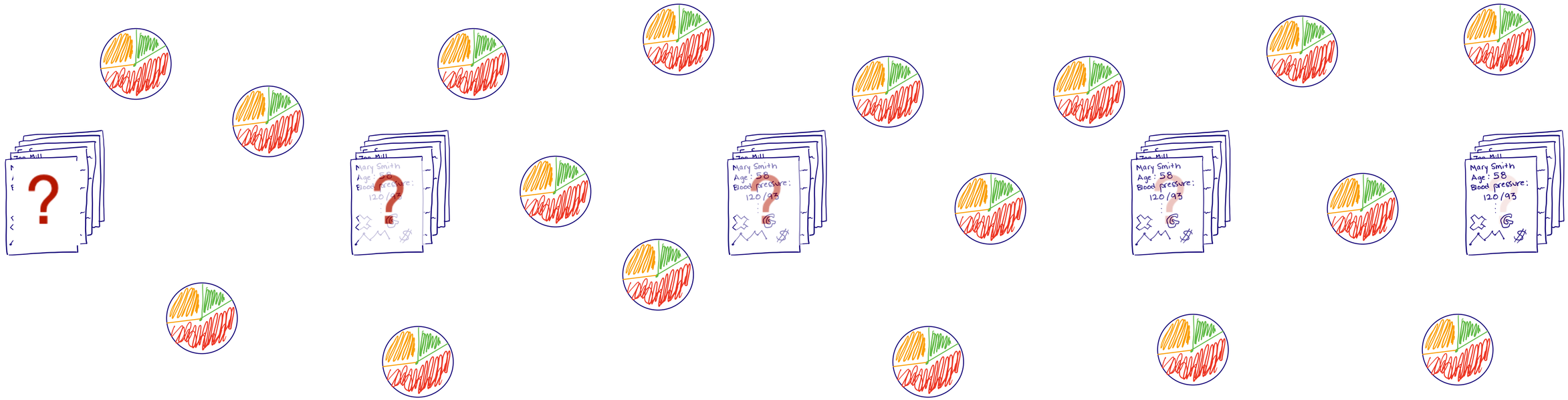
Various models, many possible computations

- Release many aggregate statistics, train machine learning models
- Answer requests formulated over time by various actors
- Create synthetic data that respects key aggregates
- Individual data providers can provide their own privacy



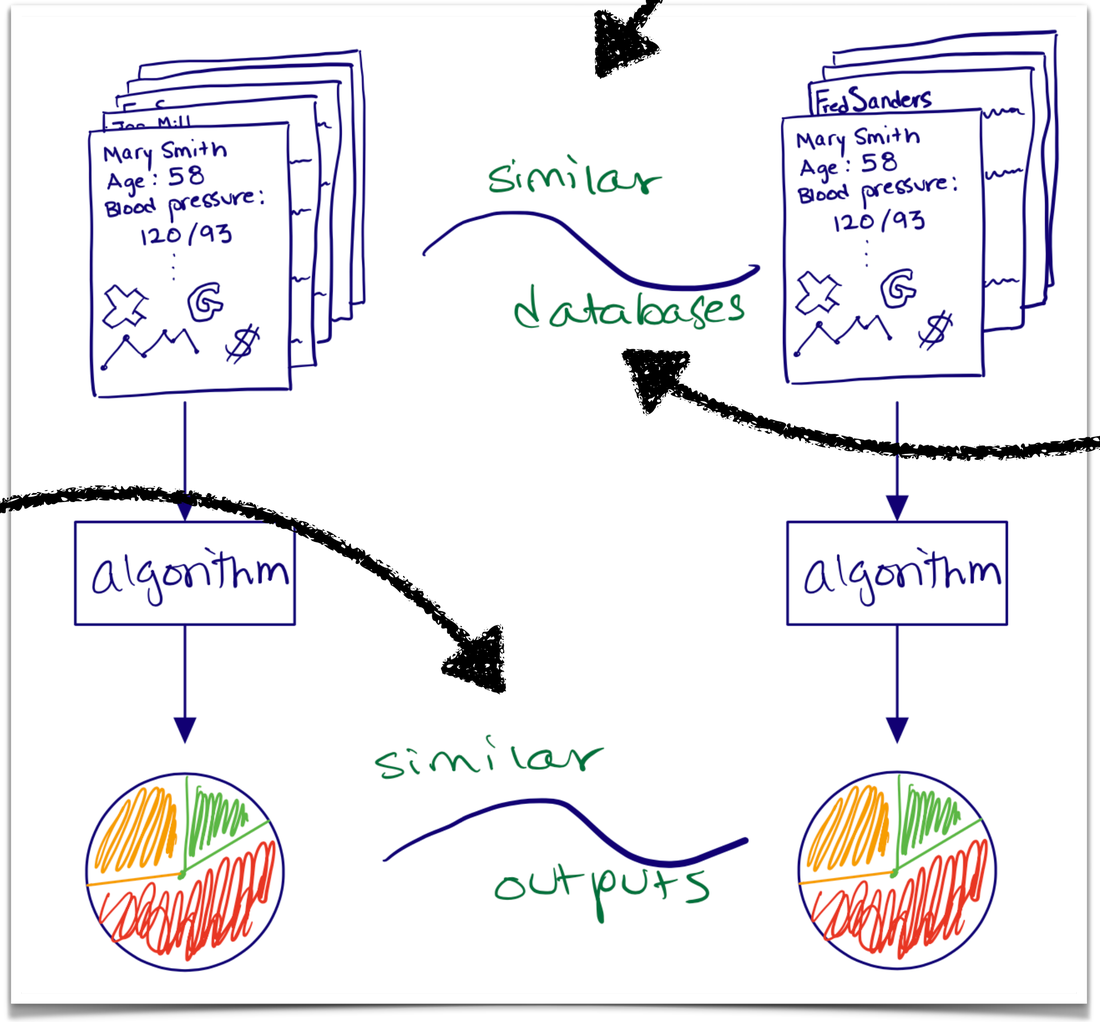
If all we're getting is aggregates, why worry at all?

- Every aggregate statistic is a constraint on the underlying data that could have generated it
- Without any noise on these constraints, about *n* statistics allow substantial reconstruction of a database of size *n*



Differential Privacy (DP)

(Dwork, McSherry, Nissim, Smith ’06)



differ by addition/ removal of one person’s data

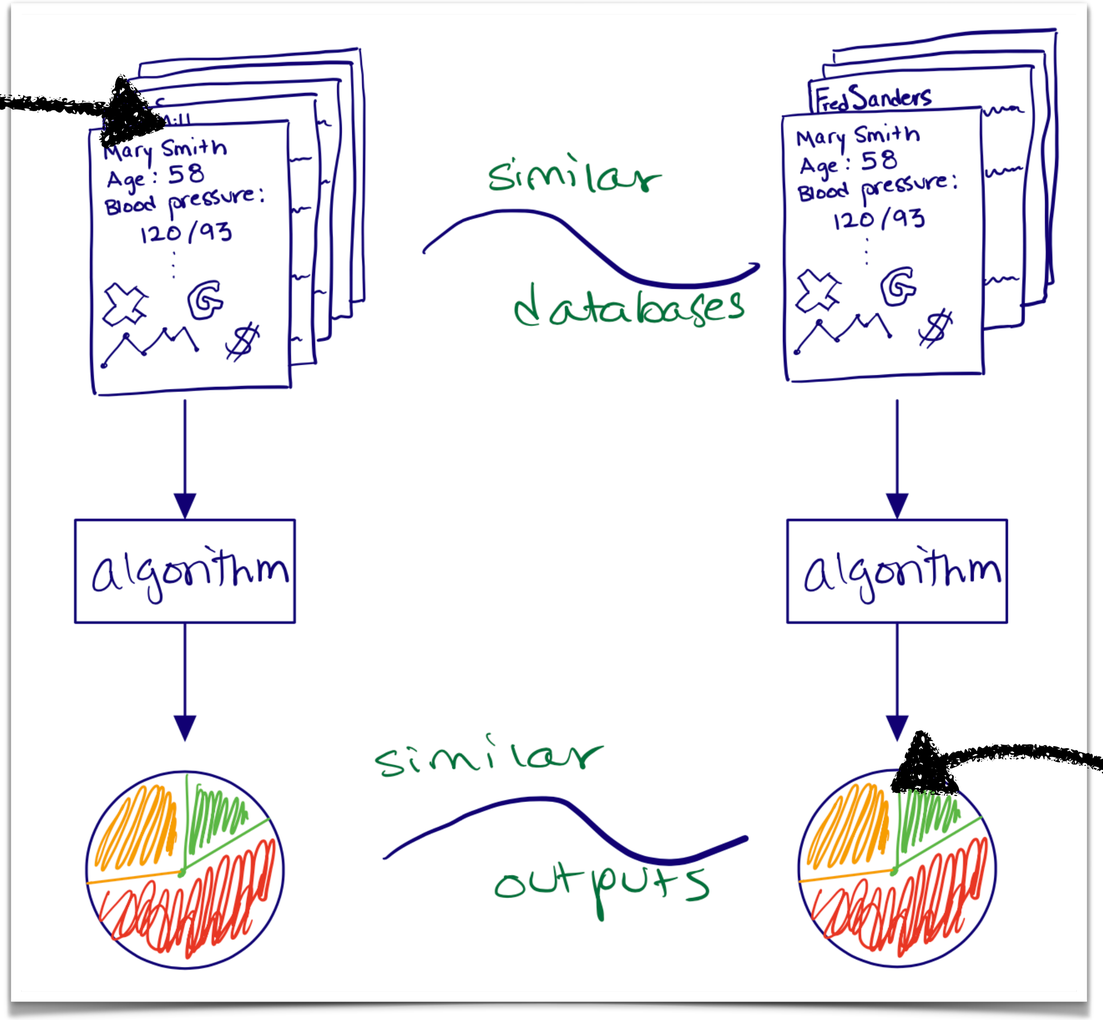
must hold for all possible neighboring databases

meaningful for all output types… statistics, machine learning models, synthetic data, etc.

Differential Privacy (DP)

(Dwork, McSherry, Nissim, Smith ’06)

Intuitively, *your* data couldn’t have affected the outcome by much



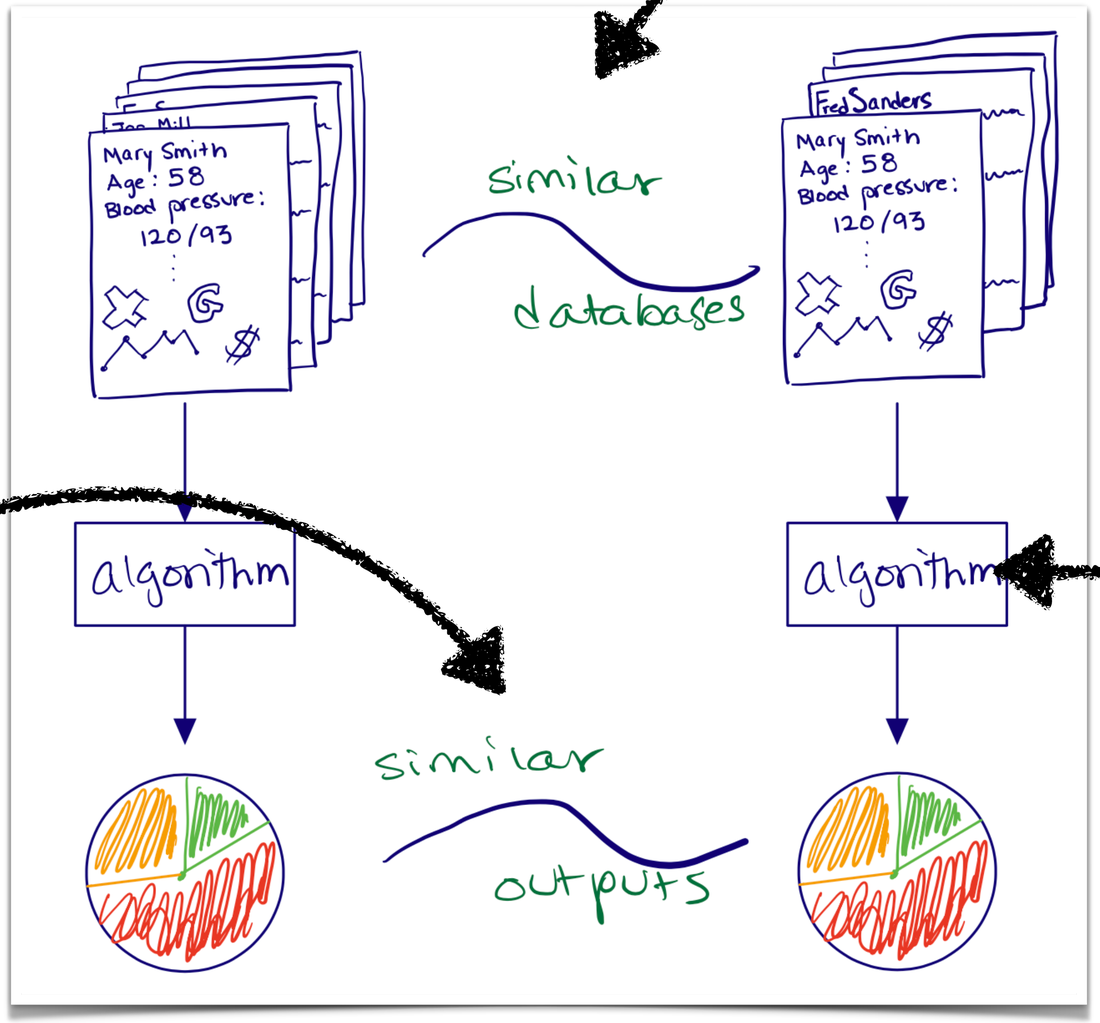
Goal: get this while still retaining correlations, trends, aggregates

Intuitively, by looking at the output, get very little info about whether *you* are in the data

ε- Differential Privacy (DP)

(Dwork, McSherry, Nissim, Smith ’06)

formally, for any *D* and *D’* that are identical sets of rows, except for addition/removal of one person (row)

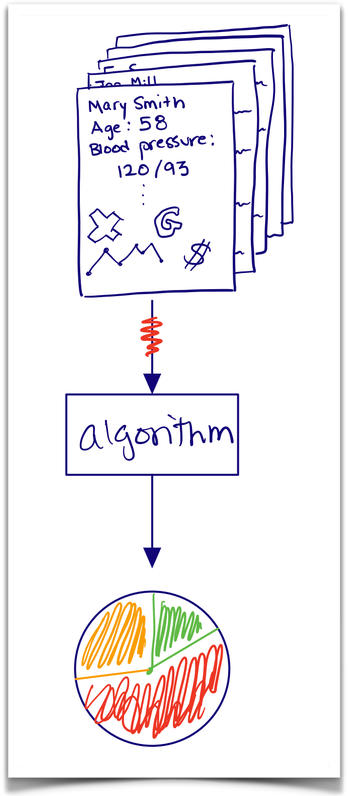


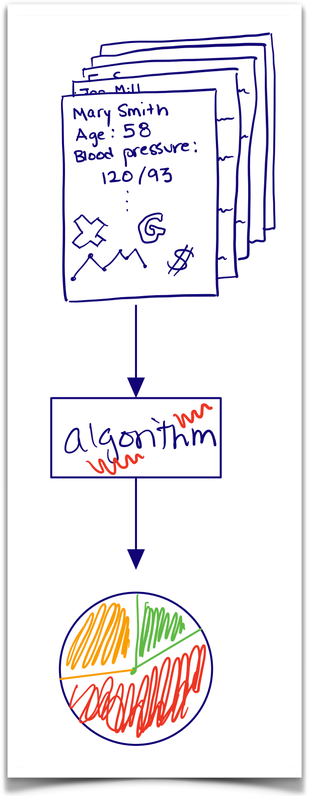
for any possible output *o* of the algorithm, the probability that *A* returns *o* on *D* is almost exactly the same as the probability that it does so on *D’*

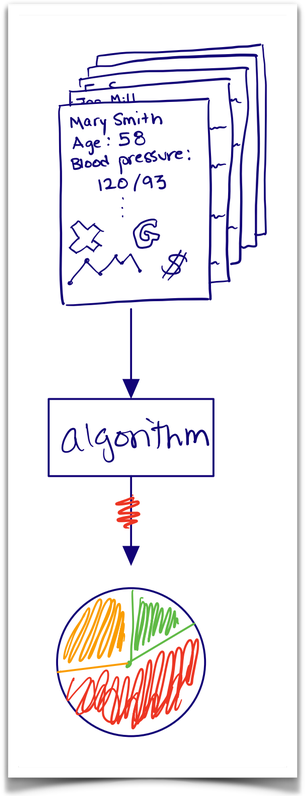
$\Pr[A(D) = o] = \exp(\varepsilon) \Pr[A(D') = o]$

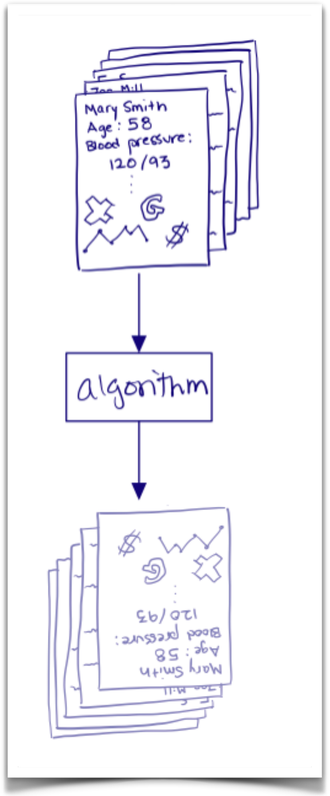
constrain algorithm *A*’s behavior

How to achieve DP?

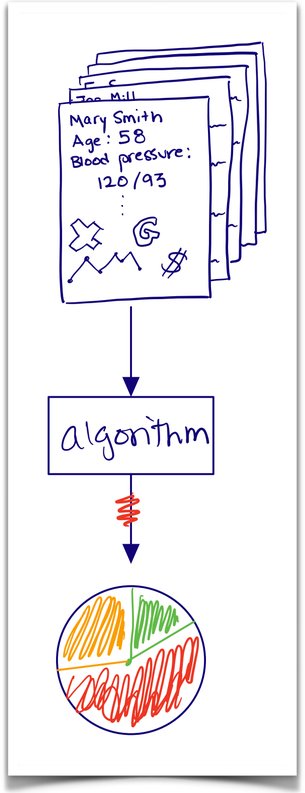






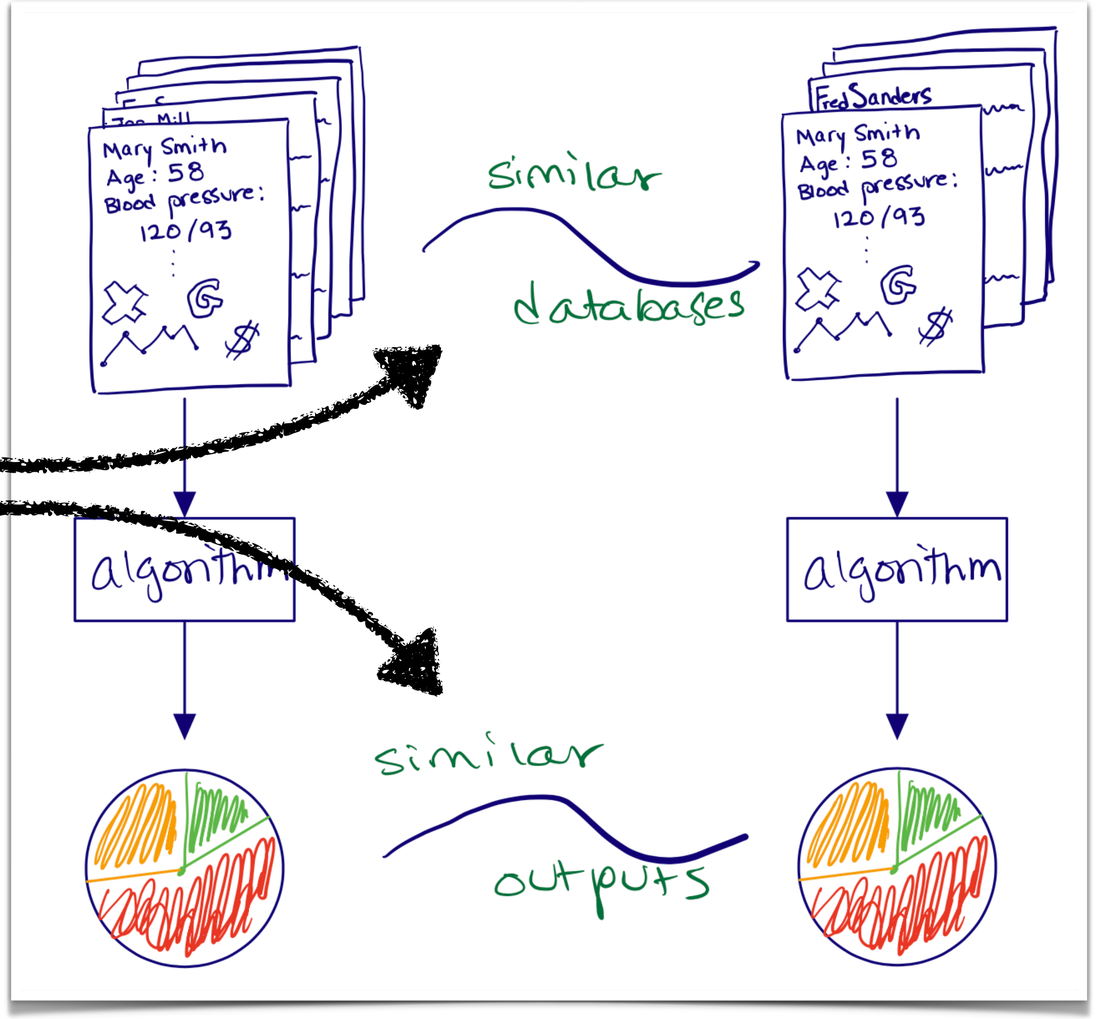


Concretely, ε-DP for a numeric output

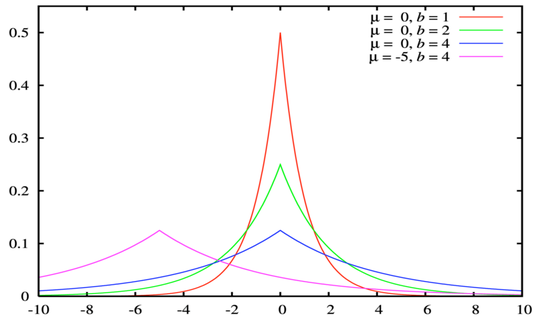


We'll call this the Global Sensitivity (GS) of the query. This is what we need to hide!

Key question: in the worst case, how much could two neighboring databases differ in their true answer?

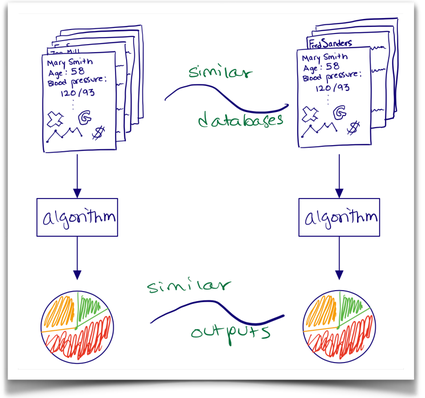


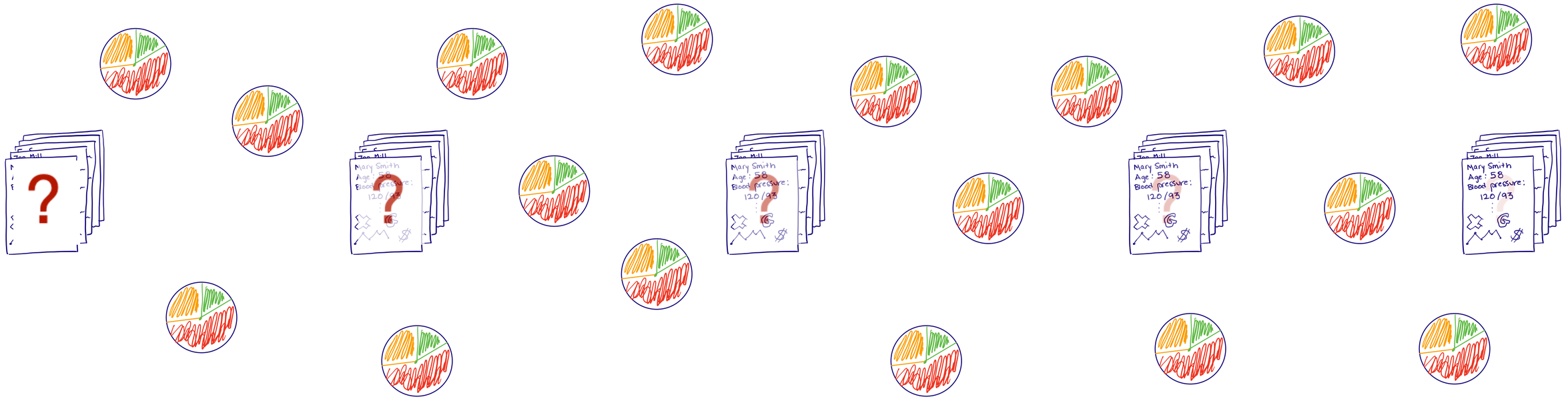
Can achieve ε-DP by adding noise ~Lap(GS/ε)



The “mindset” of differential privacy

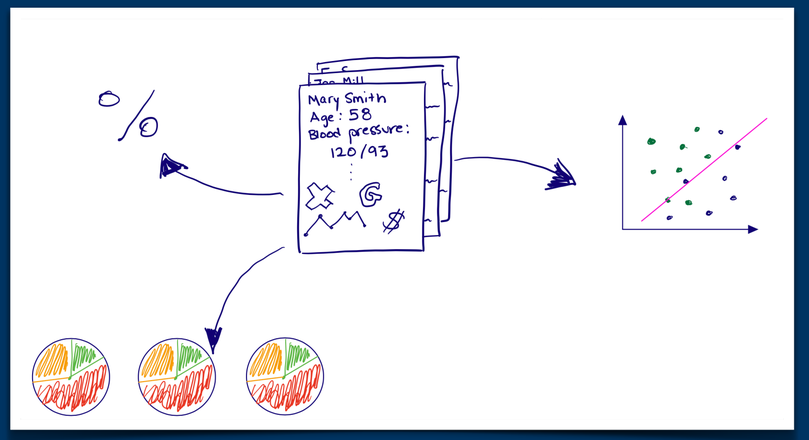
- Privacy harms “add up”
- Privacy budget
- DP actually compatible with doing good science





Privacy for Pandemics

An Introduction to Differential Privacy for Human Mobility and Interaction Data



Katrina Ligett - Computer Science - Hebrew University