6.S060
Lecture 24

Introduction to Differential Privacy

Outline

- Motivation
- Part I:
  - Differential Privacy (DP) Basics
  - DP pros and cons, deployment, challenges
- Part II:
  - DP for Statistics

Material from Harvard class: CS208: Applied Privacy for Data Science Course Overview, by James Honaker & Salil Vadhan.

Material from NIPS 2017 Tutorial by K. Chaudhuri and A. Sarwate.

Motivation

Data Privacy: The Problem

- Given a dataset with sensitive information, such as:
  - Census data
  - Health records
  - Social network activity
  - Telecommunications data
- How can we:
  - enable desirable uses of the data
  - while protecting the privacy of the data subjects?

> • Informing policy
> • Identifying subjects for drug trial
> • Searching for terrorists
> • Market analysis

Approach 1: Encrypt the Data

Name	Sex	Blood	⋯	HIV?
Chen	F	B	⋯	Y
Jones	M	A	⋯	N
Smith	M	O	⋯	N
Ross	M	O	⋯	Y
Lu	F	A	⋯	N
Shah	M	B	⋯	Y

Name	Sex	Blood	⋯	HIV?
100101	001001	110101	⋯	110111
101010	111010	111111	⋯	001001
001010	100100	011001	⋯	110101
001110	010010	110101	⋯	100001
110101	000000	111001	⋯	010010
111110	110010	000101	⋯	110101

Problems: How to search over data or compute statistics? Who has the encryption key?

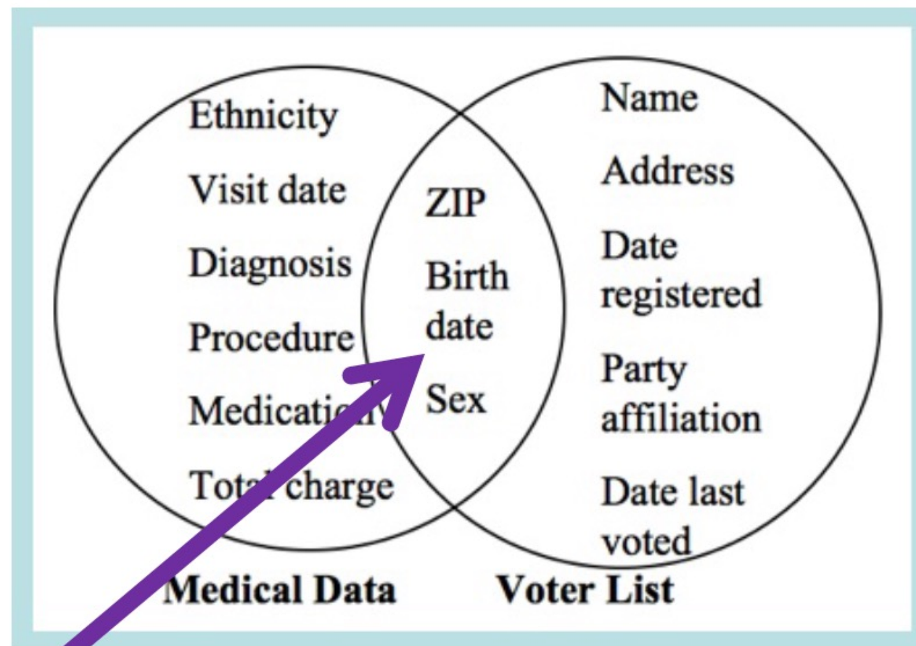
Approach 2: Anonymize the Data

~~Name~~	Sex	Blood	···	HIV?
~~Chen~~	F	B	···	Y
~~Jones~~	M	A	···	N
~~Smith~~	M	O	···	N
~~Ross~~	M	O	···	Y
~~Lu~~	F	A	···	N
~~Shah~~	M	B	···	Y

Problems?

Reidentification via Linkage

~~Name~~	Sex	Blood	…	HIV?
~~Chen~~	F	B	…	Y
~~Jones~~	M	A	…	N
~~Smith~~	M	O	…	N
~~Ross~~	M	O	…	Y
~~Lu~~	F	A	…	N
~~Shah~~	M	B	…	Y



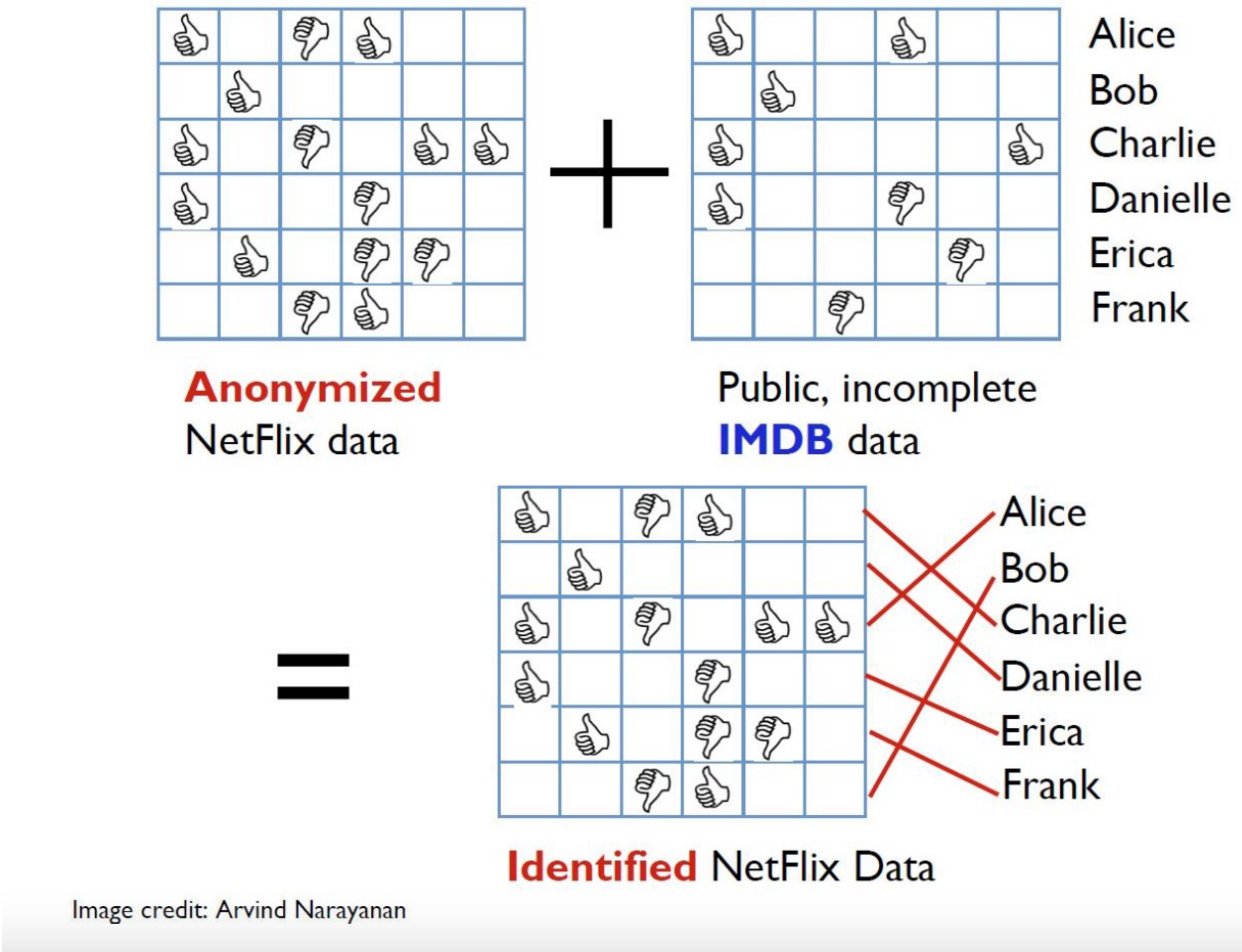
[Sweeney `97]

Uniquely identify > 60% of the US population [Sweeney `00, Golle `06]

All it takes is a knowledge of a small number of attributes to identify/name the person!

Netflix Challenge Re-Identification

[Narayanan-Shmatikov `08]



How many movies required on average to uniquely identify a user?

Four!

Narayanan-Shmatikov Set-Up

- Dataset: x = set of records (e.g., Netflix ratings)
- Adversary's inputs:
  - x' = subset of records from x, distorted slightly
  - aux = auxiliary information about a record $r \in D$ (e.g., a particular identifiable user's IMDB ratings)
- Adversary's goal: output either
  - $r' \in x'$ = record that is "close" to r, or
  - $\perp$ = failed to find a match

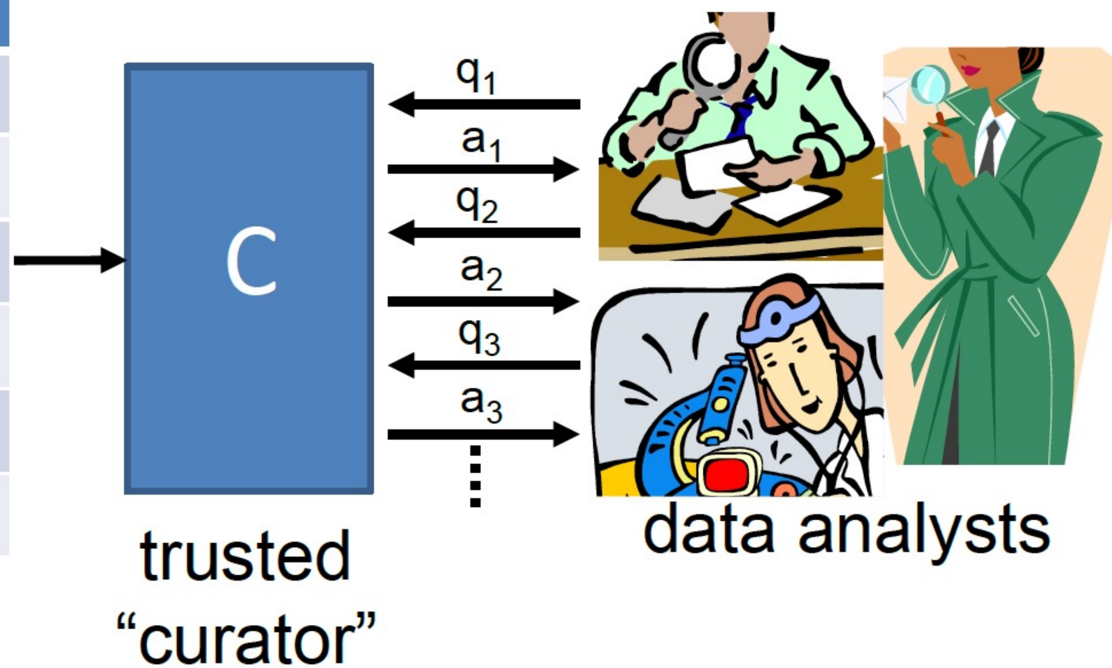
Narayanan-Shmatikov Results

- For the $1m Netflix Challenge, a dataset of 5,00,000 subscribers' ratings (less than 1/10 of all subscribers) was released (total of 100m ratings over 6 years).
- Out of 50 sampled IMDB users, two standouts were found, with eccentricities of 28 and 15.
- Reveals all movies watched from only those publicly rated on IMDB.
- Class action lawsuit, cancelling of Netflix Challenge II.

Message: Any attribute can be a "quasi-identifier"

Approach 3: Mediate Access

Name	Sex	Blood	⋯	HIV?
Chen	F	B	⋯	Y
Jones	M	A	⋯	N
Smith	M	O	⋯	N
Ross	M	O	⋯	Y
Lu	F	A	⋯	N
Shah	M	B	⋯	Y

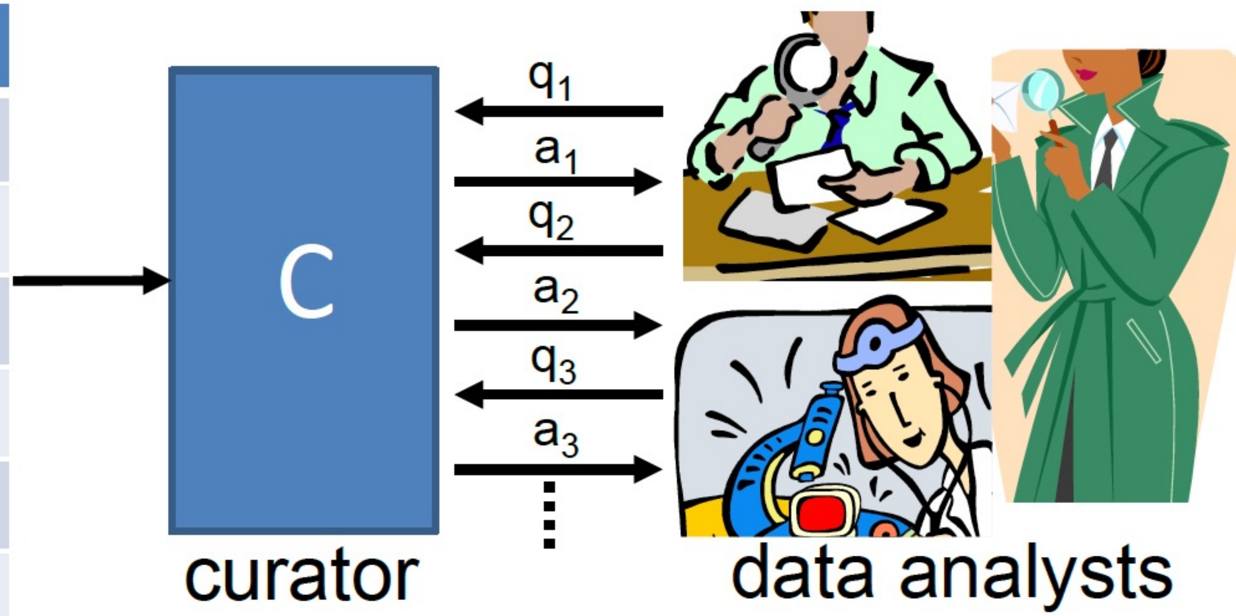


Problems: Curator sees all the data. What queries are allowed? How much do they leak?

Part I

Differential privacy

Sex	Blood	⋯	HIV?
F	B	⋯	Y
M	A	⋯	N
M	O	⋯	N
M	O	⋯	Y
F	A	⋯	N
M	B	⋯	Y

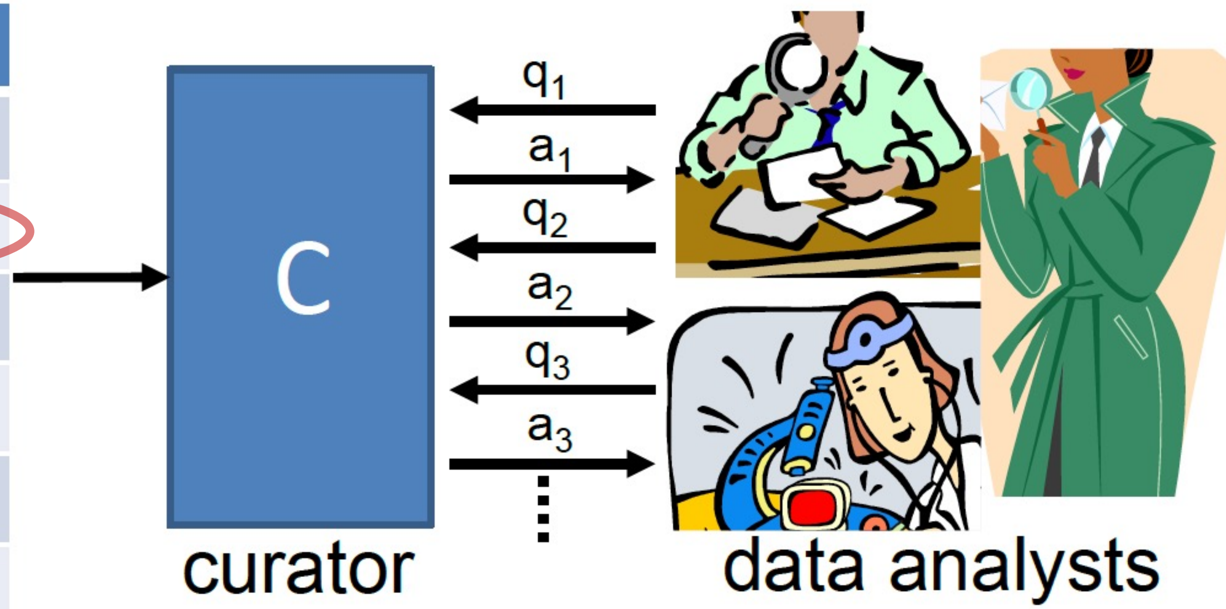


- Requirement: effect of each individual should be "hidden"

[Dinur-Nissim '03+Dwork, Dwork-Nissim '04, Blum-Dwork-McSherry-Nissim '05, Dwork-McSherry-Nissim-Smith '06]

Differential privacy

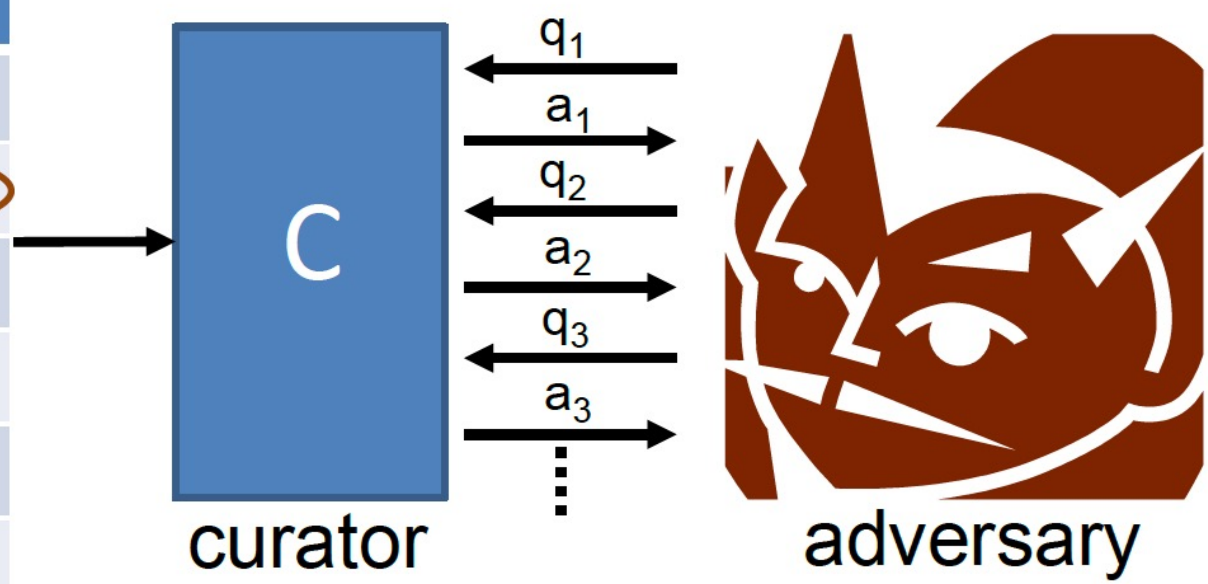
Sex	Blood	⋯	HIV?
F	B	⋯	Y
M	A	⋯	N
M	O	⋯	N
M	O	⋯	Y
F	A	⋯	N
M	B	⋯	Y



- Requirement: Adversary should not be able to tell if any one person's data were changed arbitrarily

Differential privacy

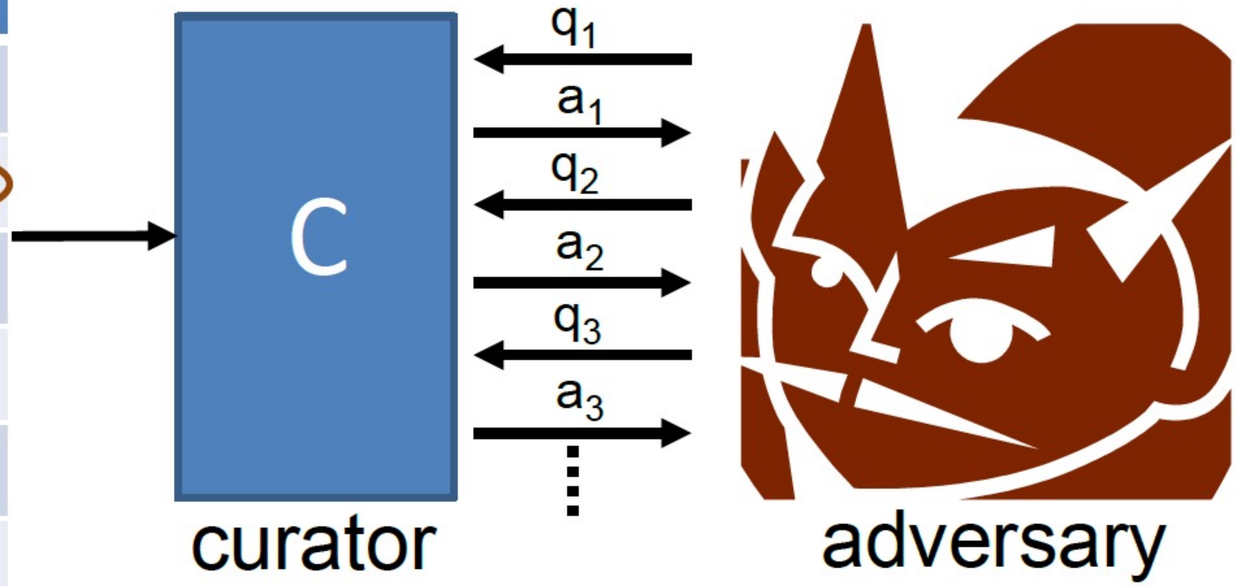
Sex	Blood	…	HIV?
F	B	…	Y
M	O	…	N
M	O	…	Y
F	A	…	N
M	B	…	Y



- Requirement: Adversary should not be able to tell if any one person’s data were changed arbitrarily

Differential privacy

Sex	Blood	⋯	HIV?
F	B	⋯	Y
F	A		Y
M	O	⋯	N
M	O	⋯	Y
F	A	⋯	N
M	B	⋯	Y

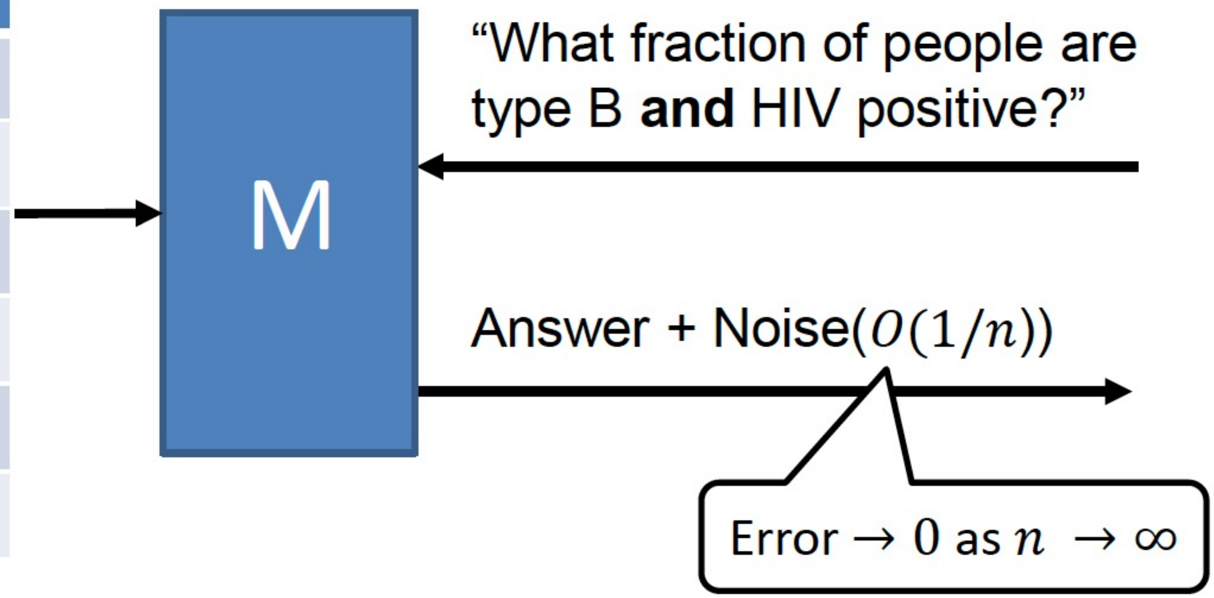


- Requirement: Adversary should not be able to tell if any one person’s data were changed arbitrarily

Simple approach: random noise

Sex	Blood	⋯	HIV?
F	B	⋯	Y
M	A	⋯	N
M	O	⋯	N
M	O	⋯	Y
F	A	⋯	N
M	B	⋯	Y

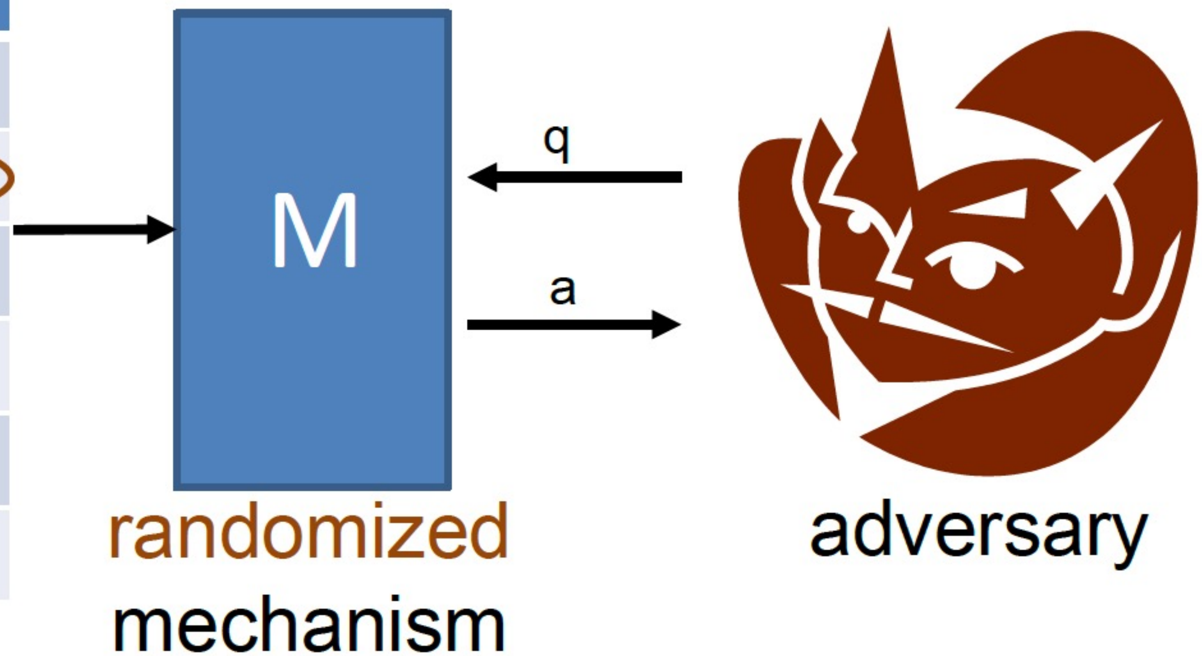
n



- Very little noise needed to hide each person as $n \to \infty$
- This is just for one query

DP for one query/release

Sex	Blood	…	HIV?
F	B	…	Y
F	A		Y
M	O	…	N
M	O	…	Y
F	A	…	N
M	B	…	Y

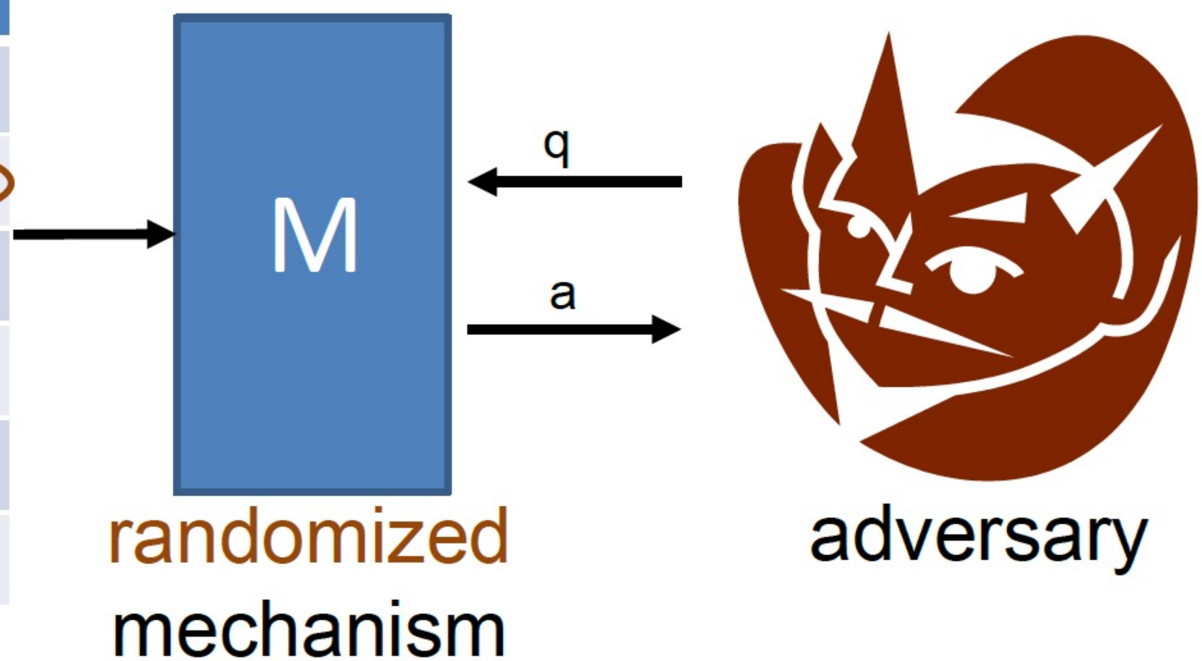


- Requirement: for all D, D' differing on one row, and all q

Distribution of M(D,q) $\approx_{\varepsilon}$ Distribution of M(D',q)

DP for one query/release

Sex	Blood	⋯	HIV?
F	B	⋯	Y
F	A		Y
M	O	⋯	N
M	O	⋯	Y
F	A	⋯	N
M	B	⋯	Y



- Requirement: M is ε-DP if for all D, D' differing on one row, and all q

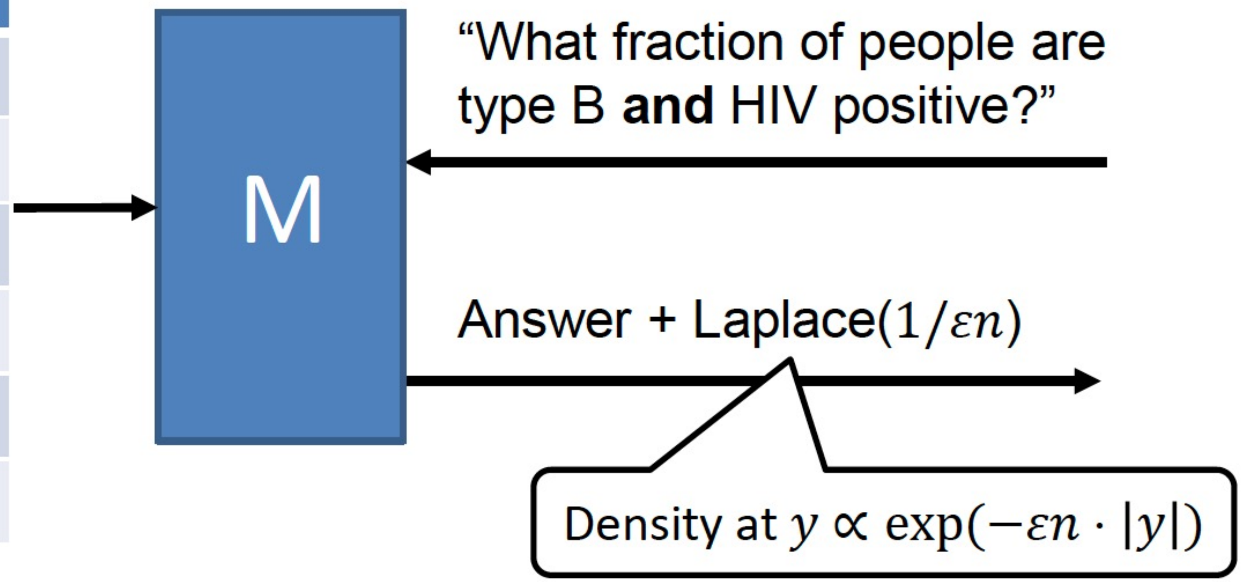
$$\forall \text{ sets } T,\ \Pr[M(D, q) \in T] \le e^{\varepsilon} \cdot \Pr[M(D', q) \in T]$$

The Laplace Mechanism

[Dwork-McSherry-Nissim-Smith ’06]

n

Sex	Blood	⋯	HIV?
F	B	⋯	Y
M	A	⋯	N
M	O	⋯	N
M	O	⋯	Y
F	A	⋯	N
M	B	⋯	Y

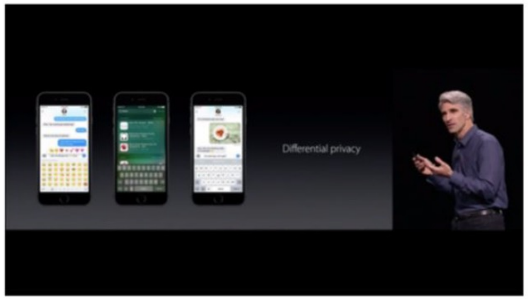


- Very little noise needed to hide each person as $n \to \infty$
- **Theorem: The Laplace Mechanism is Differentially Private**

Differential Privacy: Pros and Cons

+ Whatever an adversary learns about me, it could have learned from everyone else's data
+ Mechanism cannot leak "individual-specific" information
+ Above interpretations hold regardless of adversary's auxiliary information
+ Composes: k repetitions is kε differentially private
- No protection for information that is not localized to a few rows.
- No guarantee that subjects won't be "harmed" by results of analysis

Differential Privacy Deployed



Apple


Google Security Blog

The latest news and insights from Google on security and safety on the Internet

Learning statistics with privacy, aided by the flip of a coin

October 30, 2014

Cross-posted on the Research Blog and the Chromium Blog

At Google, we are constantly trying to improve the techniques we use to protect our users' security and privacy. One such project, RAPPOR (Randomized Aggregatable Privacy-Preserving Ordinal Response), provides a new state-of-the-art, privacy-

Google

United States™ Census Bureau

Census Bureau

Uber

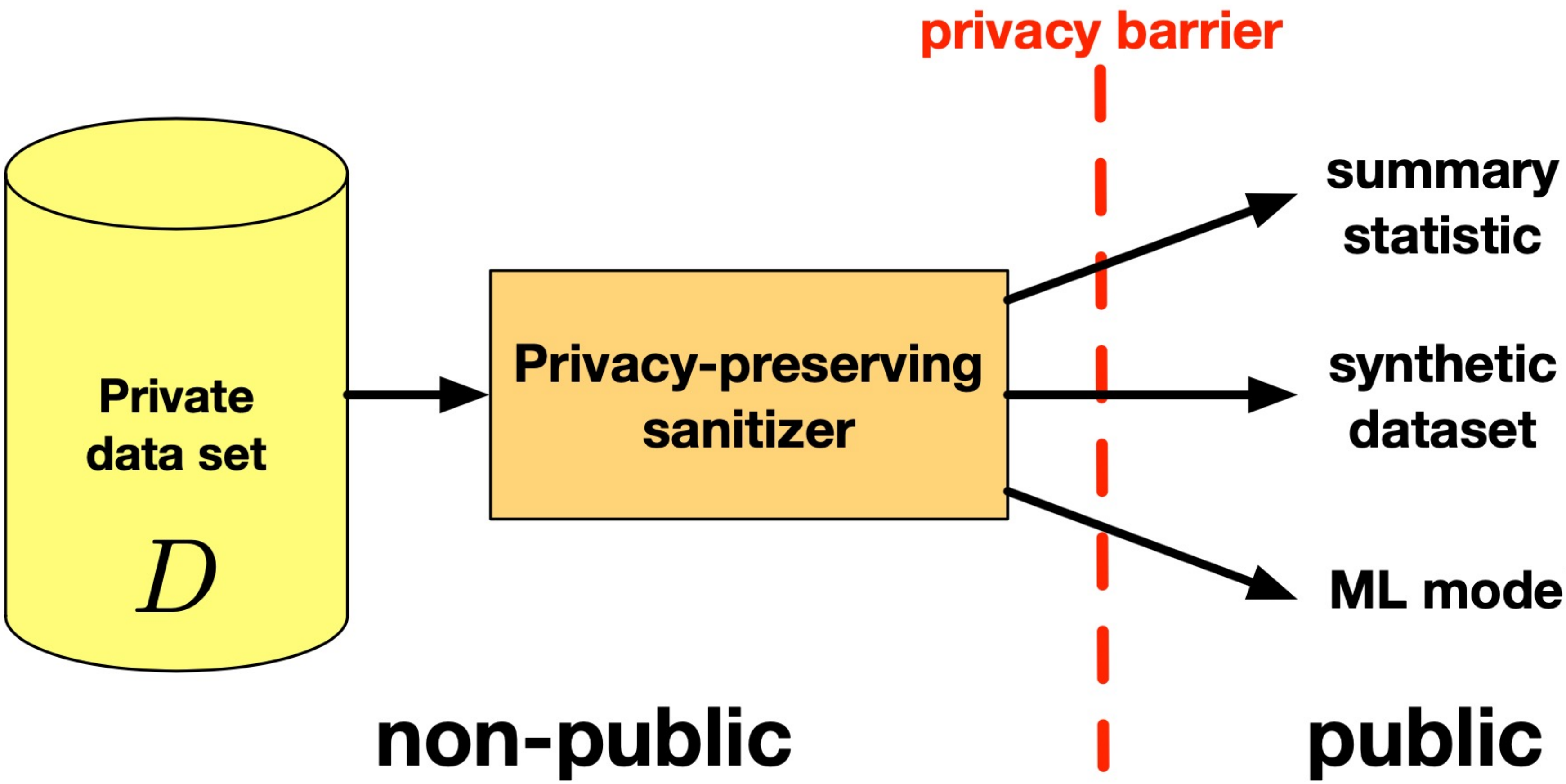
mostly focused on count and average statistics

Challenges for DP in Practice

- Accuracy for “small data” (small n)
- Modeling and managing privacy loss over time
- Analysts are used to working with raw data, not querying (slightly) noisy data
- Matching guarantees with privacy law and regulation
- …

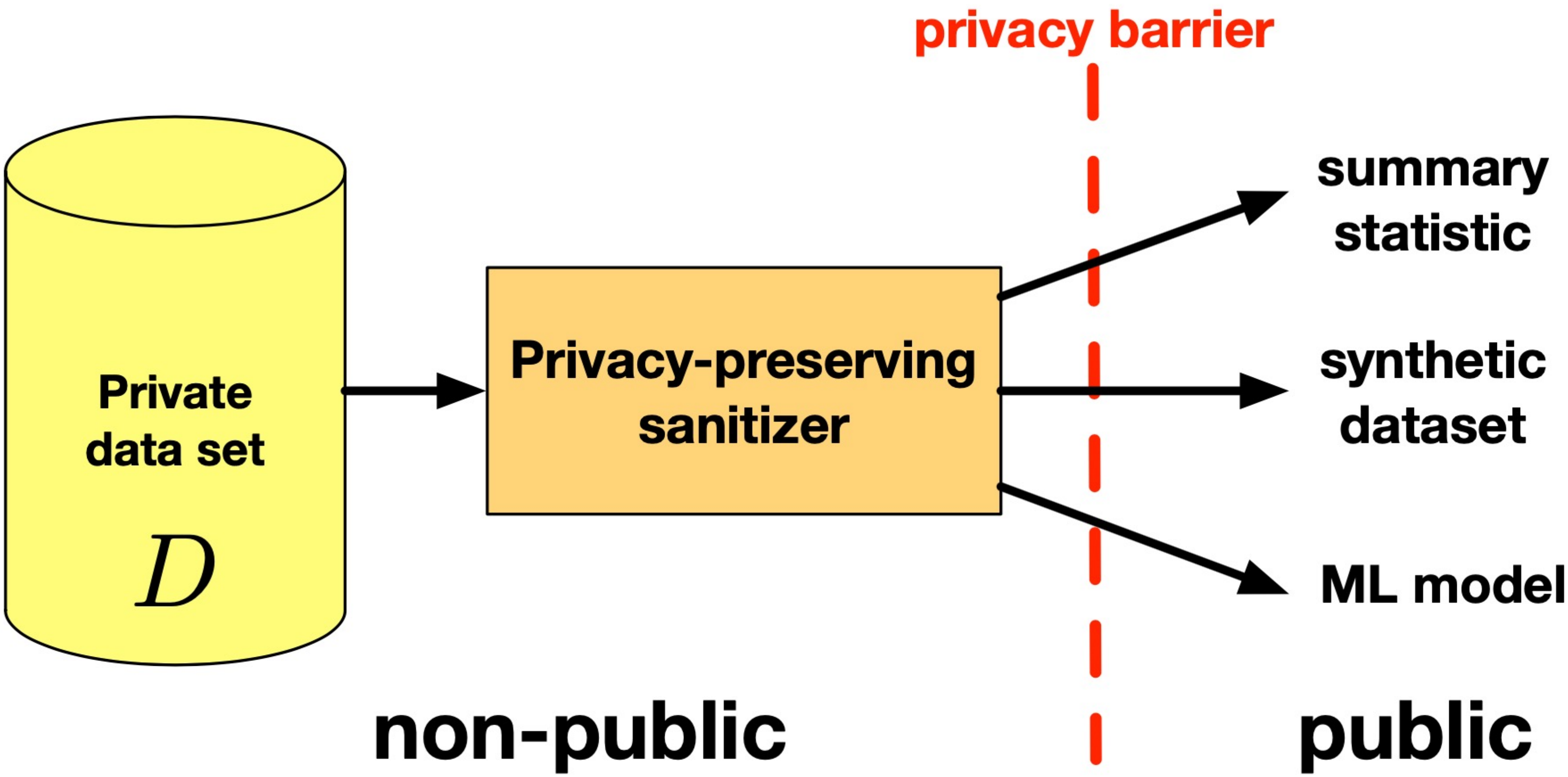
Part II

Setting



privacy barrier
Private
data set
D
Privacy-preserving
sanitizer
summary
statistic
synthetic
dataset
ML model
non-public
public

Property of Sanitizer



- Aggregate information computable
- Individual information protected

Differentially Private Algorithm Design

- Global Sensitivity Method: statistics
- Exponential Method: optimization

- **Problem:**
- Given function f, sensitive dataset D
  Find a differentially private approximation to f(D)
- **Example:** f(D) = mean of data points in D

The Global Sensitivity Method

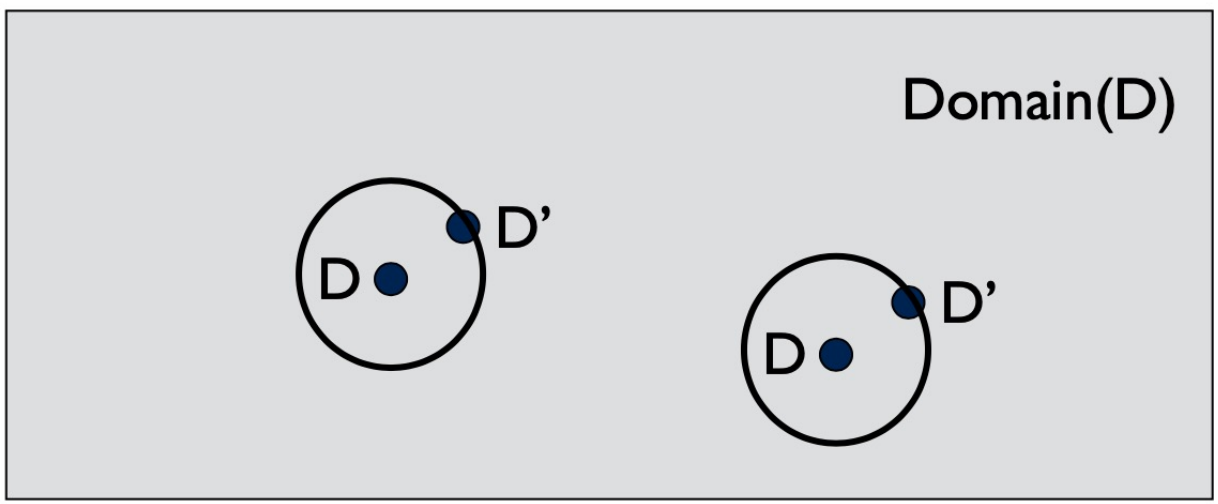
Given: A function f, sensitive dataset D

Define: dist(D, D') = #records that D, D' differ by 1

Add or remove a record from D to get D'

Global Sensitivity of f:

$$S(f) = \max_{dist(D, D') = 1} | f(D) - f(D')|$$



The Laplace Mechanism

Global Sensitivity of f is $S(f) = \max_{\text{dist}(D, D') = 1} | f(D) - f(D')|$

Output $f(D) + Z$, where

$$Z \sim \frac{S(f)}{\epsilon} \text{Lap}(0, 1)$$

ϵ-differentially private

Laplace distribution:

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

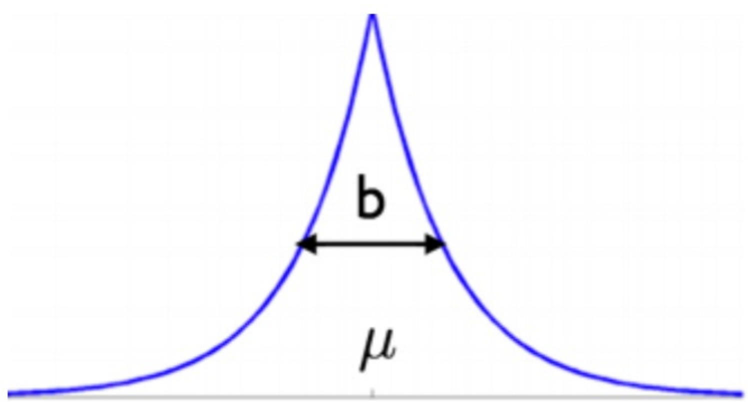
Example: Mean

f(D) = Mean(D), where each record is a scalar in [0,1]

Global Sensitivity of f = 1/n

Laplace Mechanism:

Output $f(D) + Z,$ where $Z \sim \frac{1}{n\epsilon}\text{Lap}(0, 1)$



Exponential Mechanism

Problem:

Given function f(w, D), Sensitive Data D

Find differentially private approximation to

$$w^* = \underset{w}{\operatorname{argmax}}\, f(w, D)$$

Example: f(w, D) = accuracy of classifier w on dataset D

Exponential Mechanism

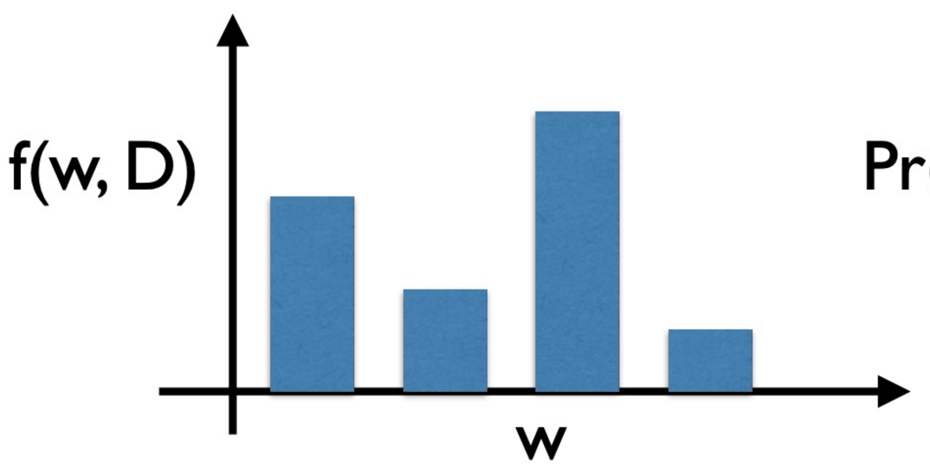
Suppose for any w,

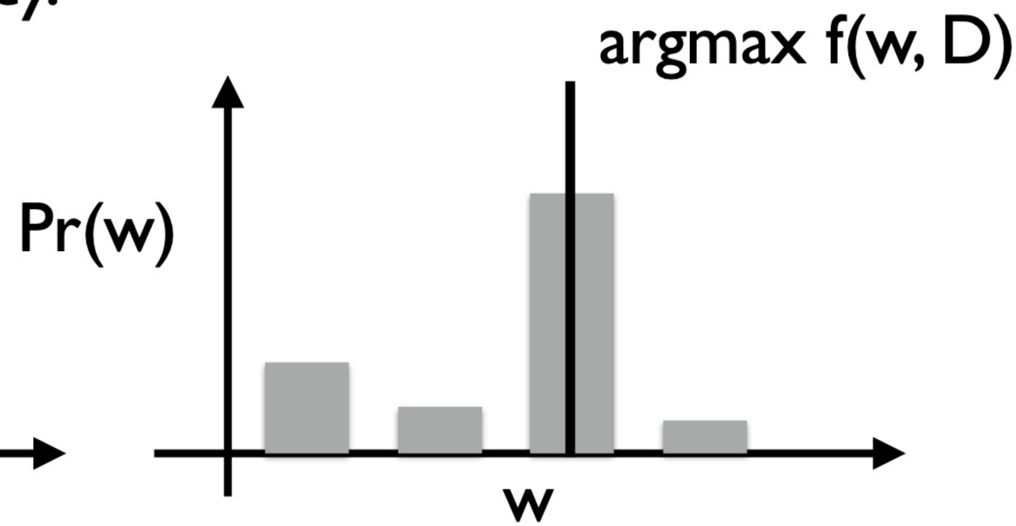
$$|f(w, D) - f(w, D')| \leq S$$

when D and D' differ in 1 record. Sample w from:

$$p(w) \propto e^{\epsilon f(w,D)/2S}$$

for ϵ-differential privacy.





Example: Parameter Tuning

Given validation data D, k classifiers $w_1, .., w_k$

(privately) find the classifier with highest accuracy on D

Here, f(w, D) = classification accuracy of w on D

For any w, any D and D' that differ by one record,

$$|f(w,D) - f(w,D')| \leq \frac{1}{|D|}$$

So, the exponential mechanism outputs w_i with prob:

$$\Pr(w_i) \propto e^{\epsilon|D|f(w_i,D)/2}$$

Conclusion

- Differential privacy can help companies to learn more about a group of users without compromising the privacy of an individual within that group.
- Many of the world's governments now have strict policies about how tech companies collect and share user data.
  - Companies need users' data to provide high-quality services that benefit users, such as personalized recommendations.
  - Companies may face charges if they collect too much user data