Building Privacy-Aware Database Systems

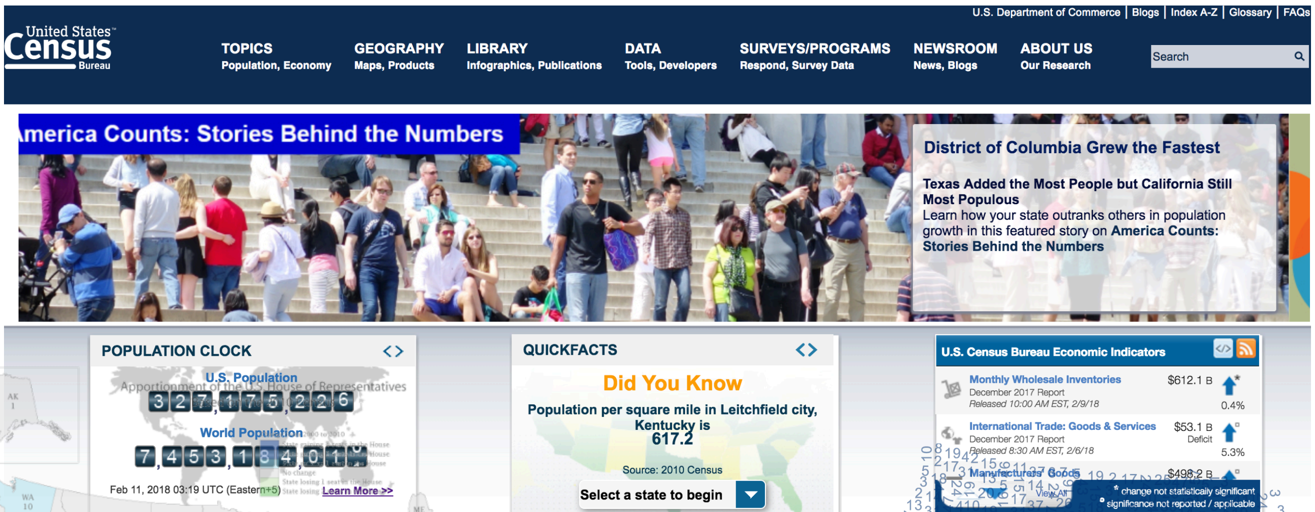
CS848 Winter 2021

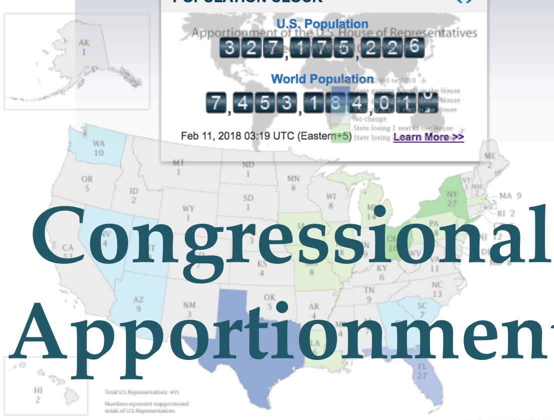
UNIVERSITY OF WATERLOO | DSg Data Systems Group

Logistics

- **CS848, Winter 2021**
  - Option 1: Tue 10am – noon (lecture/paper presentation + discussion)
  - Option 2: Wed 10am – 11am (discussion)
    - Students who attend Wed session need to watch recorded lecture/paper presentation from Tue
- More details at the end of this lecture

An Old Problem to US Census





Congressional Apportionment

IPUMS USA

An Old Problem to US Census

Title 13, U.S. Code

By law, no one – neither the census takers nor any other Census Bureau employee – is permitted to reveal identifiable information about any person, household, or business

If anyone violates this law, it is a federal crime; they will face severe penalties, including a federal prison sentence of up to five years, a fine of up to $250,000, or both.

New Attack on 2010 Decennial Census

"how many people of the age 10-20 live in New York City"

"how many people live in 4 person households"

An internal team was able to

(a) correctly reconstruct records of address (by census block), age, gender, race and ethnicity for 142 million people (about **46% of the US population**),

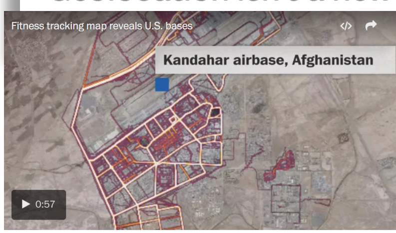
(b) correctly match these data to commercial datasets circa 2010 to associate PII like name for 52 million persons (**17% of the population**).

Fundamental Law of Info Reconstruction [DN03]
"overly accurate" estimates of "too many" statistics is blatantly non-private.

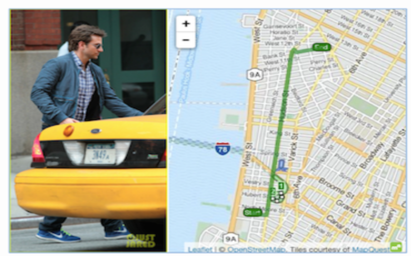
Getting Worse …

Strava's fitness tracker heat map reveals the location of military bases

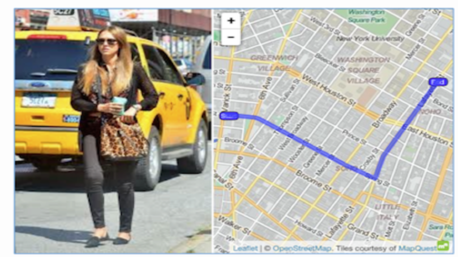
Geolocation isn't a new problem for the military



Jan 28, 2018,



Bradley Cooper (Click to Explore)

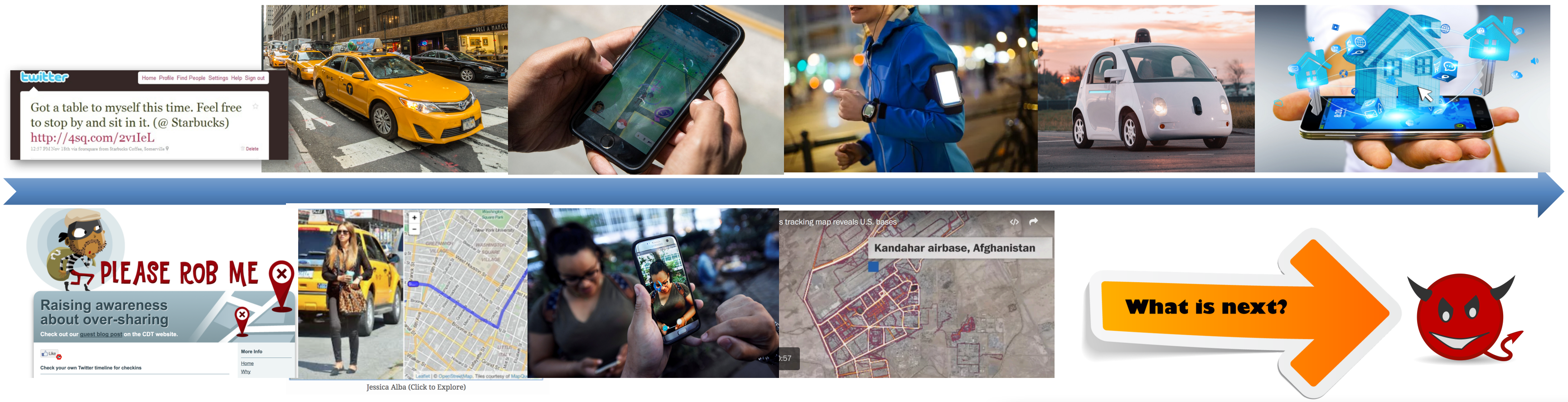


Jessica Alba (Click to Explore)

Riding with the Stars: Passenger Privacy in the NYC Taxicab Dataset

SEPTEMBER 15, 2014 BY ATOCKAR LEAVE A COMMENT

More Real-time Data Collection



The Seven Sins of Personal-Data Processing Systems under GDPR

Supreeth Shastri, *Computer Science*, *University of Texas at Austin*
Melissa Wasserman, *School of Law*, *University of Texas at Austin*
Vijay Chidambaram, *Computer Science*, *University of Texas at Austin*

Privacy Changes Everything

Jennie Rogers[1], Johes Bater[1], Xi He[2], Ashwin Machanavajjhala[3], Madhav Suresh[1], and Xiao Wang[1]

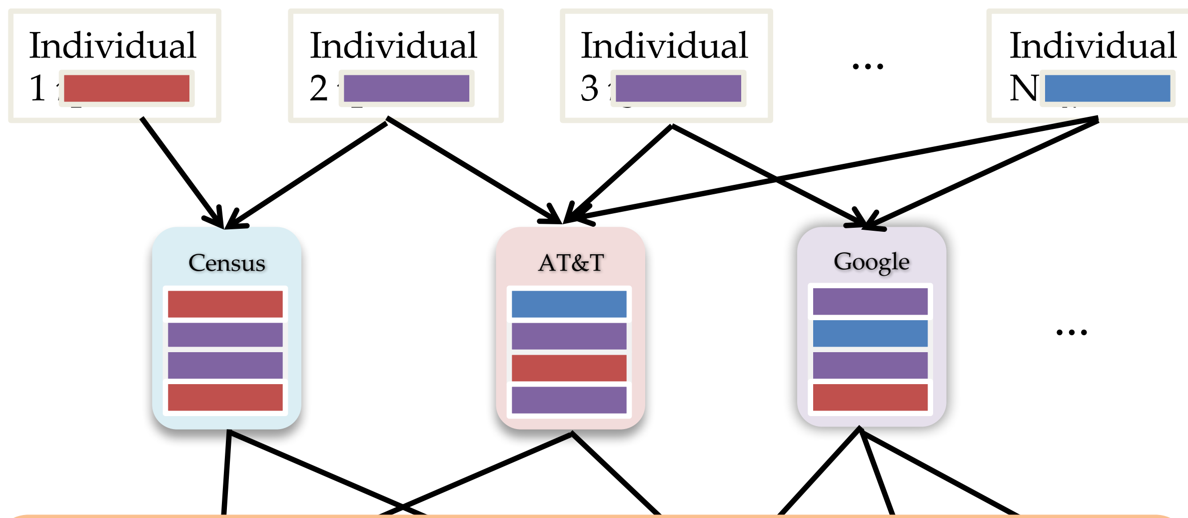
Northwestern University [2] University of Waterloo [3] Duke University

Problem Setting

Individuals with sensitive data

Data Collectors

Data Publication & Analysis

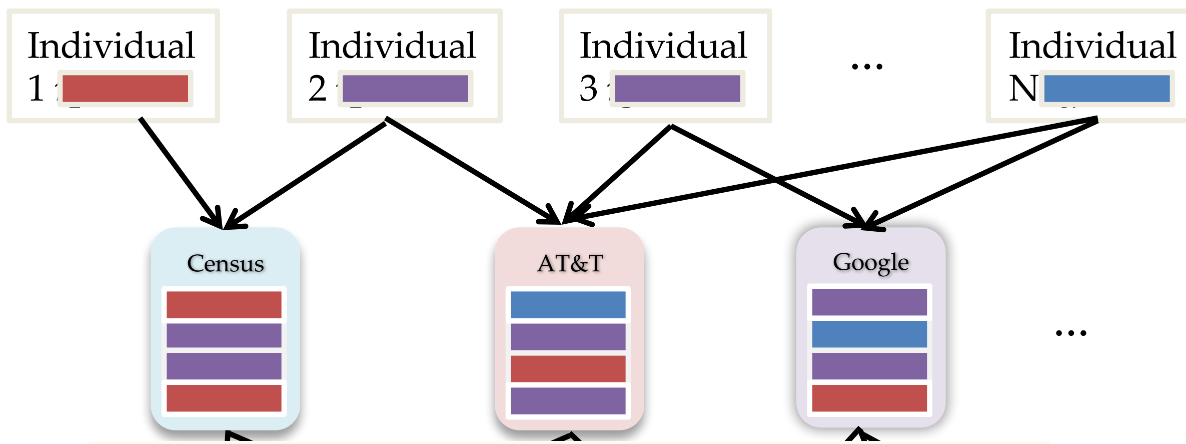


Leaks information about individual records
by the output of the computation!!

A Strong Privacy Promise

Individuals with sensitive data

Data Collectors



Data Publication & Analysis

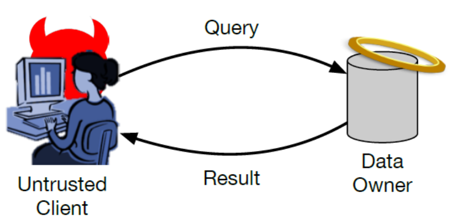
Provable privacy guarantees

Quantify and bound the amount of information disclosed about individual records by the output of the computation.

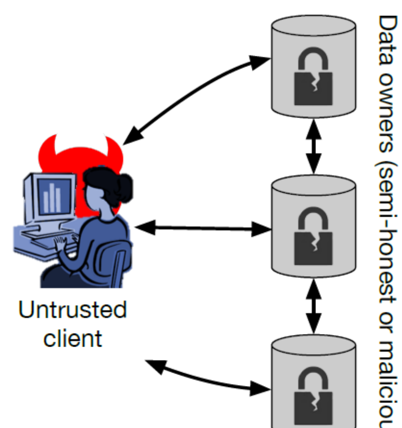
This course will explore …

- How to define a good privacy promise?
- How to design a privacy-preserving algorithms?
- How to build a privacy-aware database systems?

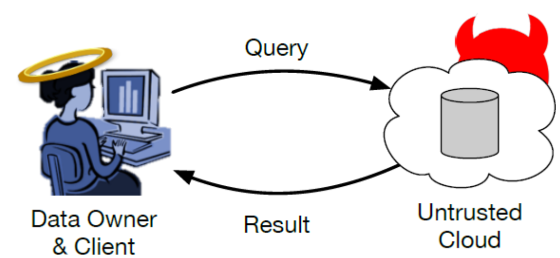
Greatly depend on
the architecture setup and trust assumptions



Client-server with
trusted data curator



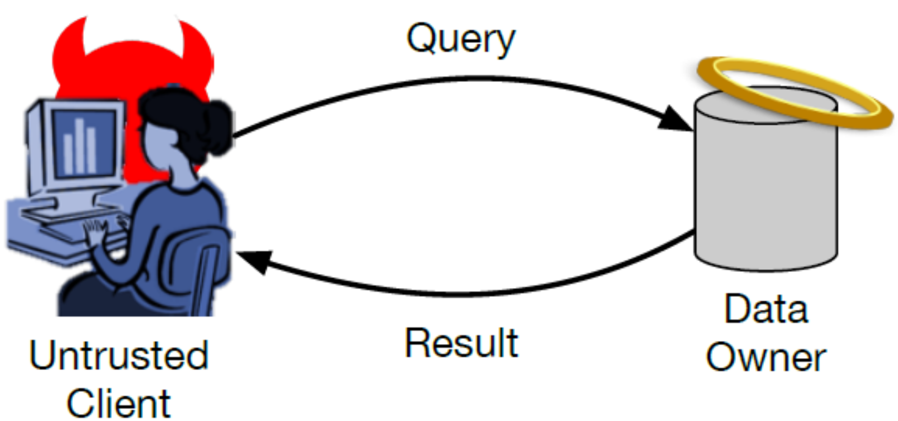
Data federation



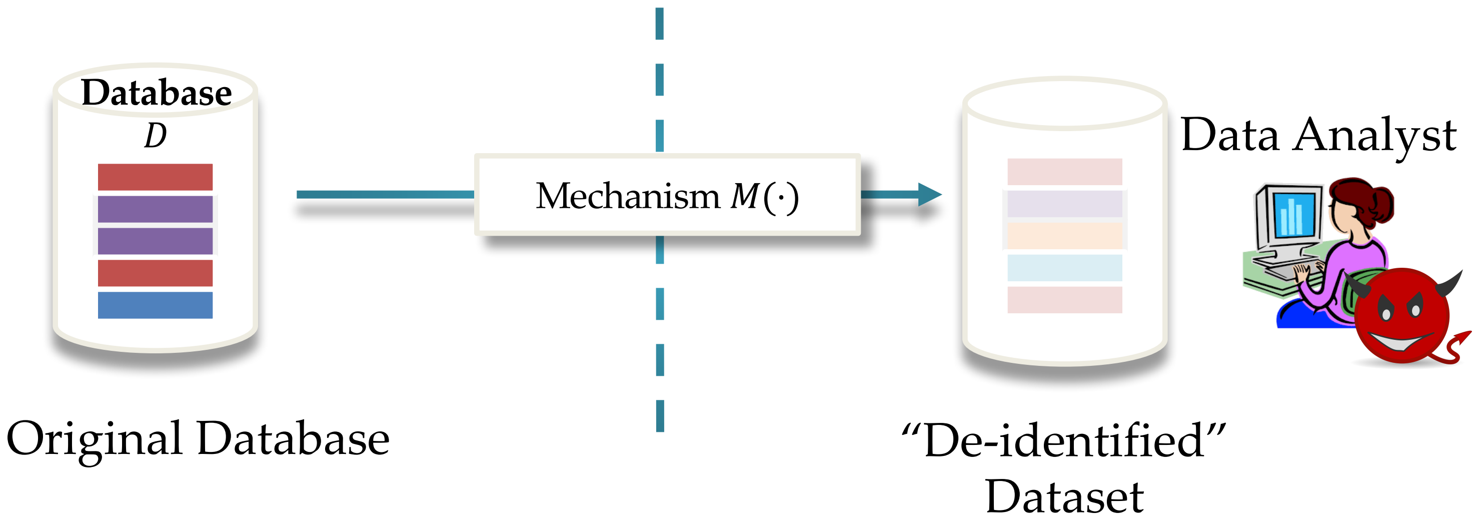
Cloud service provider

Trusted Data Curator

- Centralized setting
  - Data owners trust the data curator and have their true and plaintext data stored on a central server.
  - Client (e.g. data analyst) may infer sensitive information about individuals based on the released data from the trusted data curator



“De-Identification”?

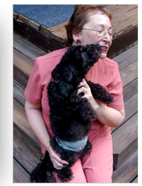


De-identified data ISN’T

A Face Is Exposed for AOL Searcher No. 4417749

By MICHAEL BARBARO and TOM ZELLER Jr.
Published: August 9, 2006

SIGN IN TO E-
THIS



Why 'Anonymous' Data Sometimes Isn't

hallenge
using.

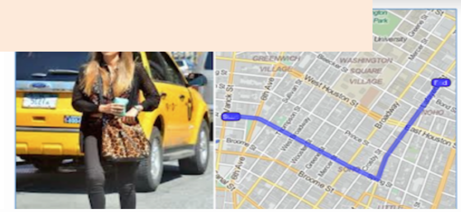
Uniqueness of personal data, side information, …

The Scientis

“Anonymous” Genomes Identified

The names and addresses of people participating in the Personal Genome Project can be easily tracked down despite such data being left off their online profiles.

By Dan Cossins | May 3, 2013



Riding with the Stars: Passenger Privacy in the NYC Taxicab Dataset

SEPTEMBER 15, 2014 BY ATOCKAR   LEAVE A COMMENT

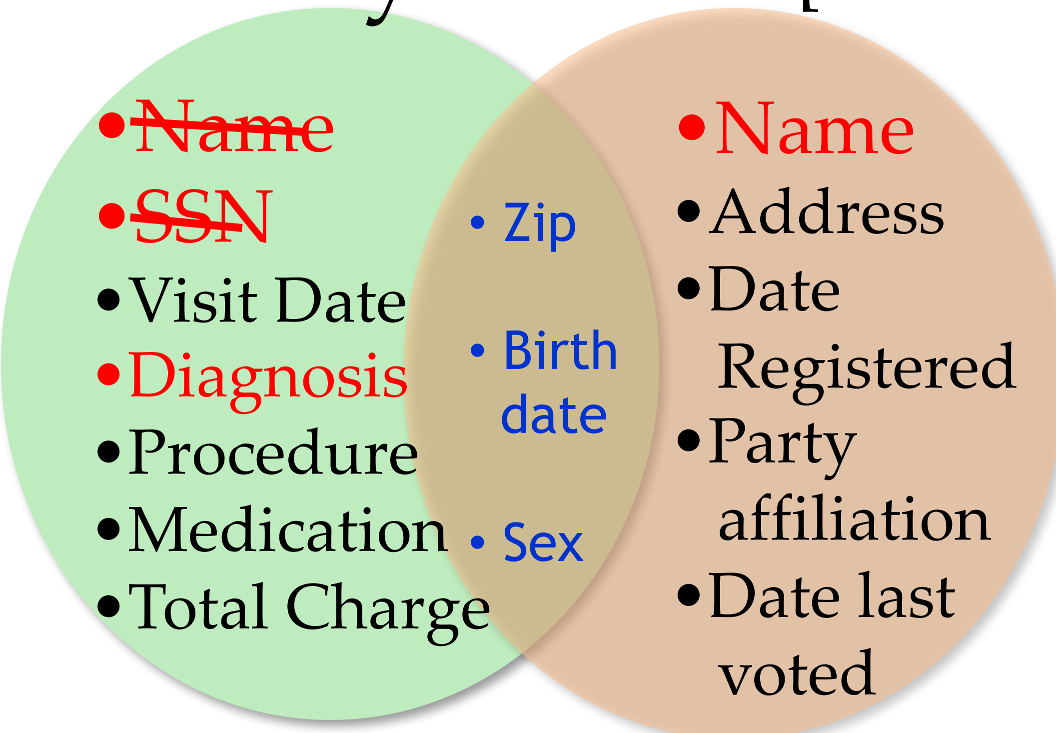
The Massachusetts Governor Privacy Breach [Sweeney IJUFKS 2002]

- ~~Name~~
- ~~SSN~~
- Visit Date
- Diagnosis
- Procedure
- Medication
- Total Charge

- Zip
- Birth date
- Sex

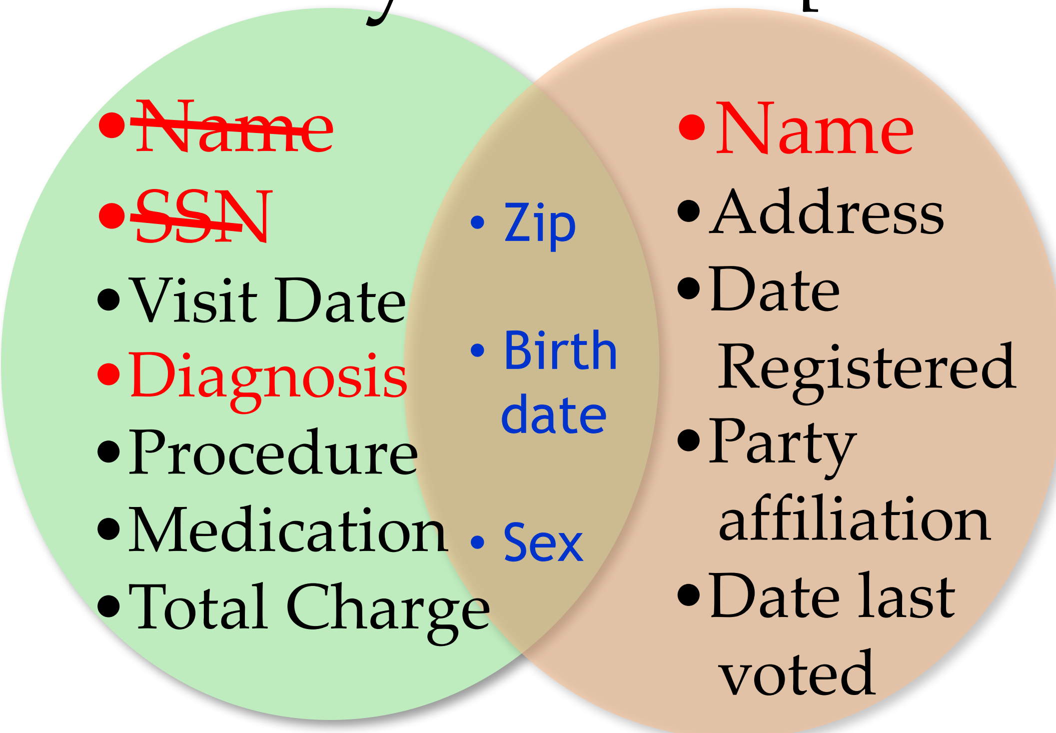
Medical Data

The Massachusetts Governor Privacy Breach [Sweeney IJUFKS 2002]



Medical Data **Voter List**

The Massachusetts Governor Privacy Breach [Sweeney IJUFKS 2002]

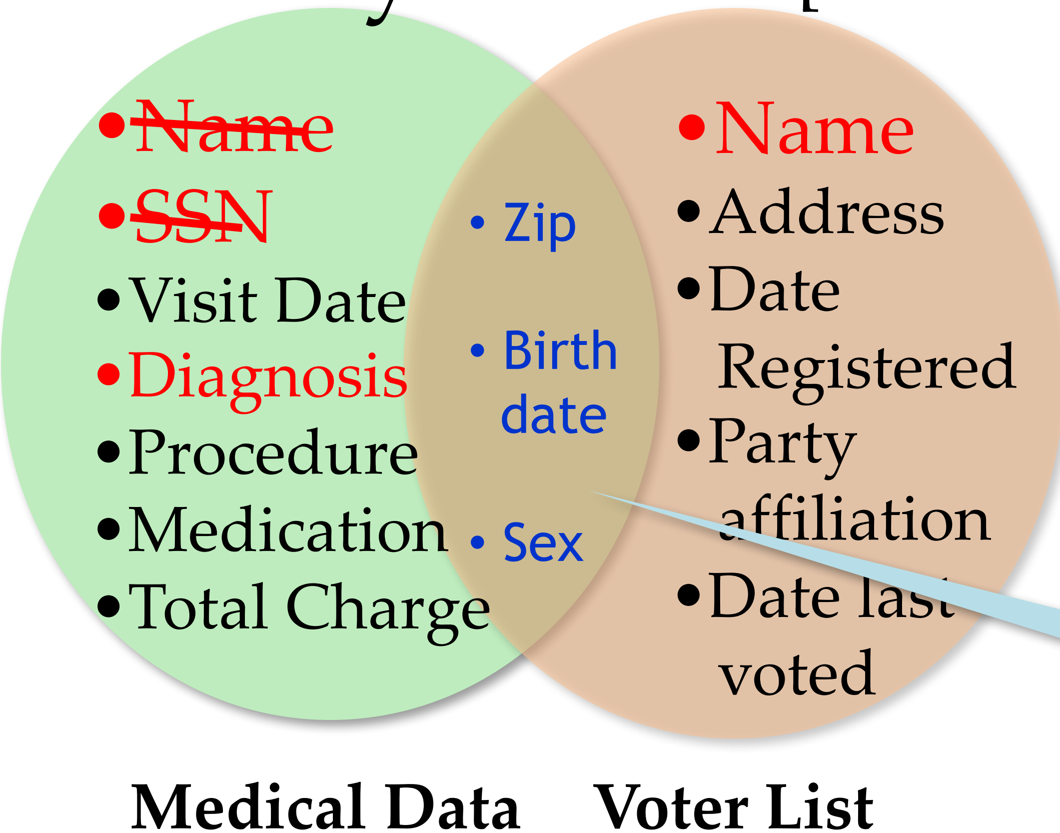


Medical Data **Voter List**

- Governor of MA **uniquely identified** using ZipCode, Birth Date, and Sex.

Name linked to Diagnosis

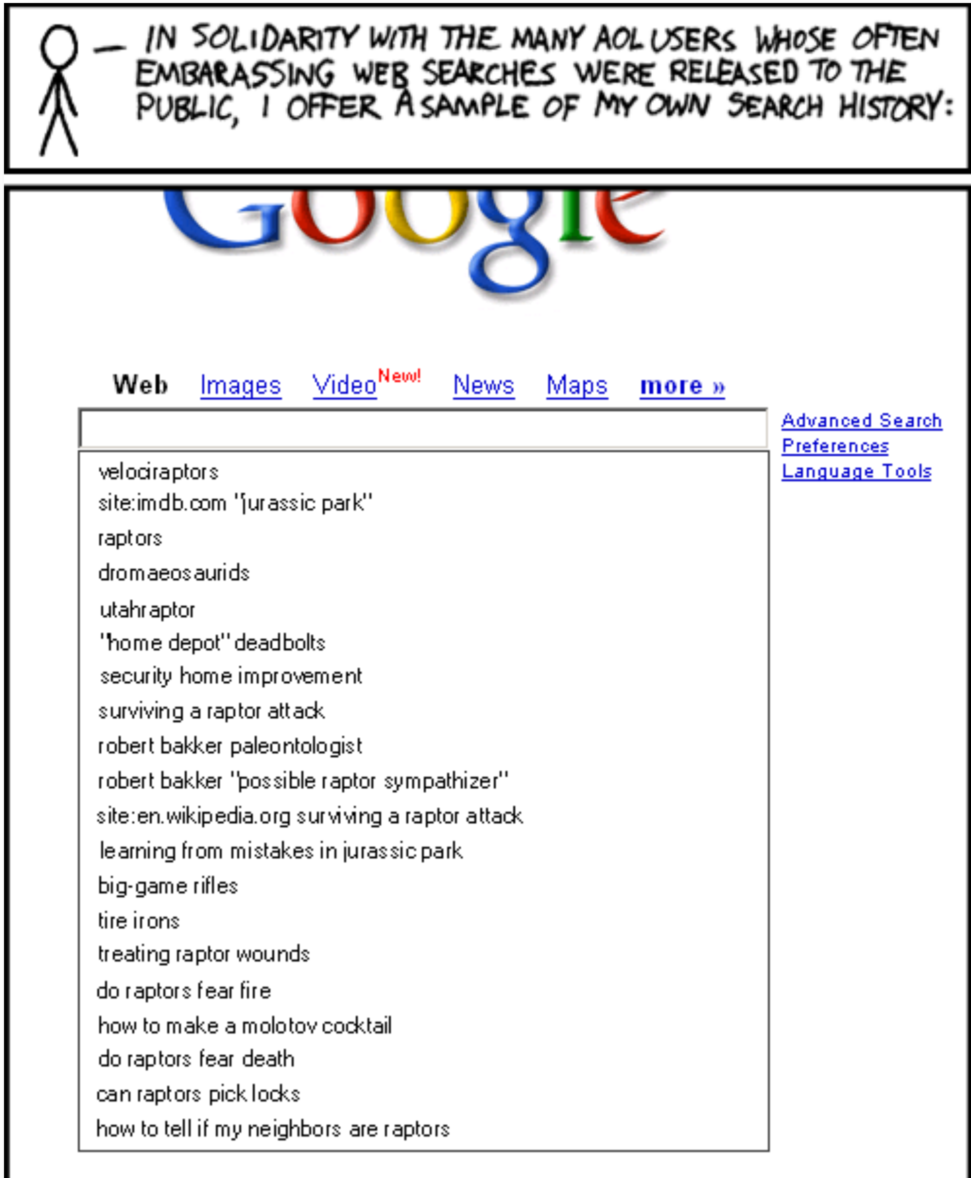
The Massachusetts Governor Privacy Breach [Sweeney IJUFKS 2002]



- 87 % of US population **uniquely identified** using ZipCode, Birth Date, and Sex.

Quasi Identifier

AOL data publishing fiasco



AOL data publishing fiasco …

Xi222	Uefa cup
Xi222	Uefa champions league
Xi222	Champions league final
Xi222	Champions league final 2013
Abel156	exchangeability
Abel156	Proof of deFinitti's theorem
Jane12345	Zombie games
Jane12345	Warcraft
Jane12345	Beatles anthology
Jane12345	Ubuntu breeze
Bob222	Python in thought
Bob222	Enthought Canopy

User IDs replaced with random numbers

865712345	Uefa cup
865712345	Uefa champions league
865712345	Champions league final
865712345	Champions league final 2013
236712909	exchangeability
236712909	Proof of deFinitti's theorem
112765410	Zombie games
112765410	Warcraft
112765410	Beatles anthology
112765410	Ubuntu breeze
865712345	Python in thought
865712345	Enthought Canopy

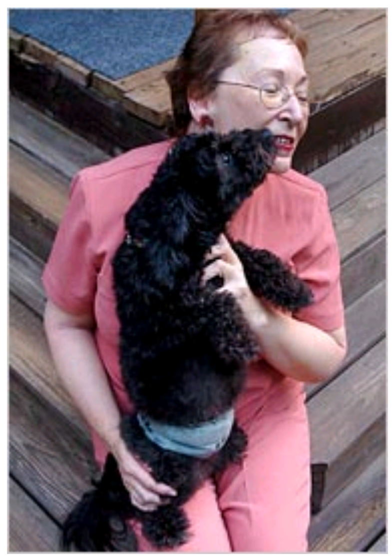
Privacy Breach

[NYTimes 2006]

A Face Is Exposed for AOL Searcher No. 4417749

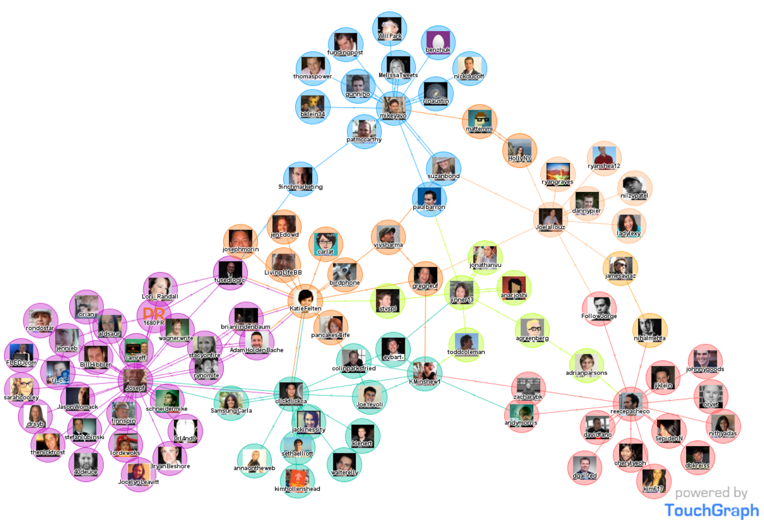
By MICHAEL BARBARO and TOM ZELLER Jr.
Published: August 9, 2006

SIGN IN TO E-
THIS



Problem 1: Naïve Anonymization

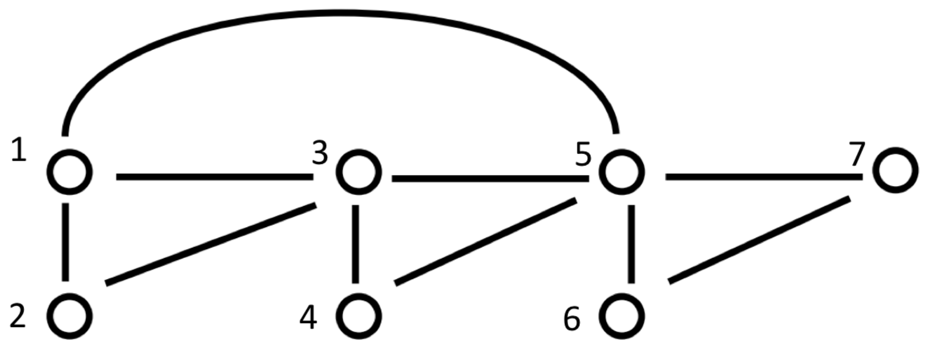
- Social networks: graphs where each node represents a social entity, and each edge represents certain relationship between two entities



- Example: email communication graphs, social interactions like in Facebook, Yahoo! Messenger, etc.

Problem 1: Naïve Anonymization

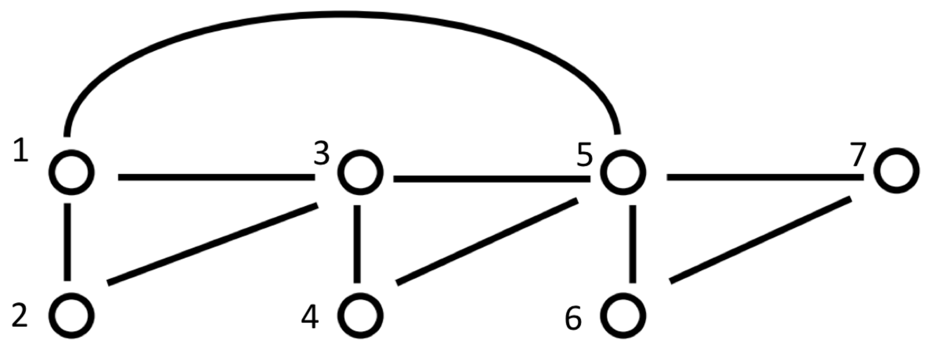
- Anonymized email communication graph



- Unfortunately for the email service providers, investigative journalists Alice and Cathy are part of this graph. What can they deduce?

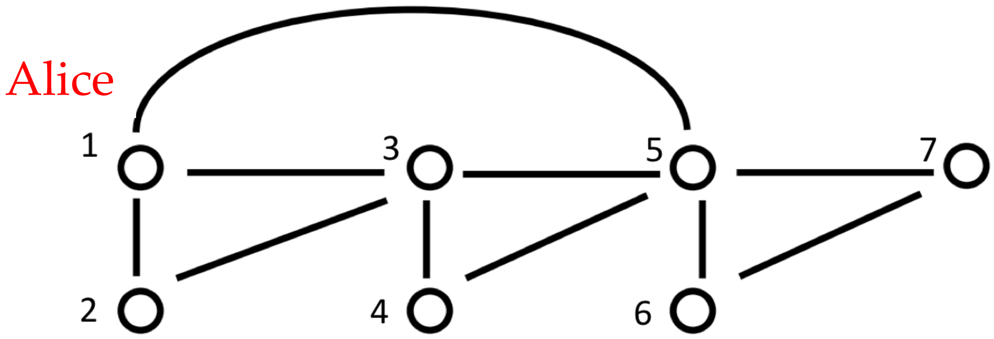
Problem 1: Naïve Anonymization

- Auxiliary knowledge:
  - Alice has sent emails to Bob, Cathy, and Ed
  - Cathy has sent emails to everyone, except Ed



Problem 1: Naïve Anonymization

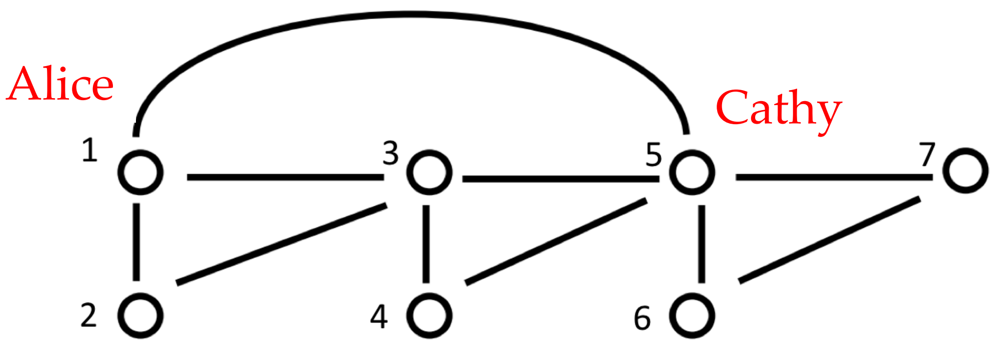
- Auxiliary knowledge:
  - Alice has sent emails to Bob, Cathy, and Ed
  - Cathy has sent emails to everyone, except Ed



- Only one node has a degree 3 → node 1: Alice

Problem 1: Naïve Anonymization

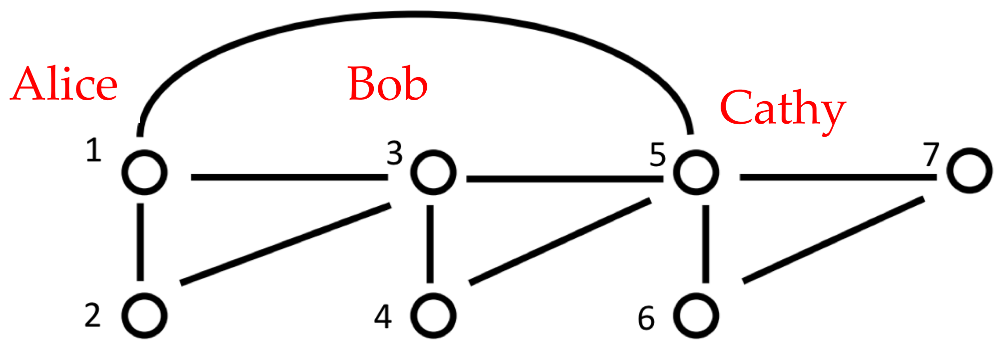
- Auxiliary knowledge:
  - Alice has sent emails to Bob, Cathy, and Ed
  - Cathy has sent emails to everyone, except Ed



- Only one node has a degree 5 → node 5: Cathy

Problem 1: Naïve Anonymization

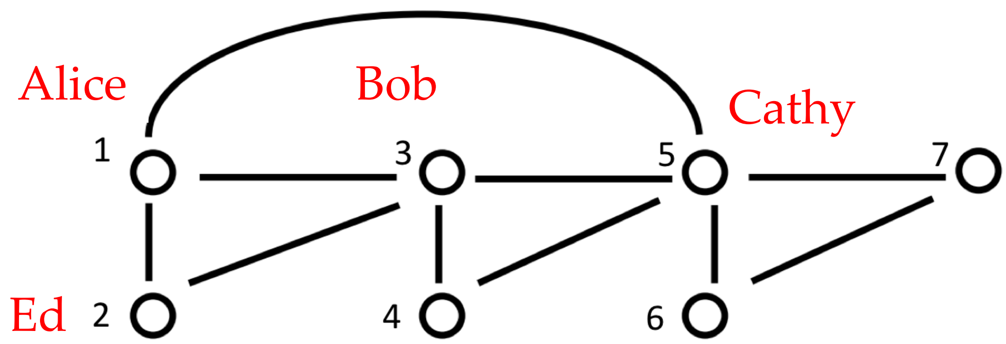
- Auxiliary knowledge:
  - Alice has sent emails to Bob, Cathy, and Ed
  - Cathy has sent emails to everyone, except Ed



- Alice and Cathy know that only Bob has sent emails to both of them → node 3: Bob

Problem 1: Naïve Anonymization

- Auxiliary knowledge:
  - Alice has sent emails to Bob, Cathy, and Ed
  - Cathy has sent emails to everyone, except Ed



- Alice has sent emails to Bob, Cathy, and Ed only
  → node 2: Ed

Attacks using Background Knowledge

- Degrees of nodes [Liu and Terzi, SIGMOD 2008]
- The network structure, e.g., a subgraph of the network. [Zhou and Pei, ICDE 2008, Hay et al., VLDB 2008]
- Anonymized graph with labeled nodes [Pang et al., SIGCOMM CCR 2006]

Desiderata for a Privacy Definition

1. Resilience to background knowledge
   – A privacy mechanism must be able to protect individuals' privacy from attackers who may possess background knowledge

Problem 2: Privacy by Obscurity

- Many organization think their data are private because they perturb the data and make the parameters of perturbation secret.

Problem 2: Privacy by Obscurity

- The email service provider also released perturbed records as per a linear function, but with *secret* parameters. What can Alice and Cathy deduce now?

Node ID	Age (perturbed)	True Age
1 (Alice)	40	25
2 (Ed)	34	
3 (Bob)	52	
4	28	
5 (Cathy)	48	29
6	22	
7	92	

Problem 2: Privacy by Obscurity

Node ID	**Name**	**Age ($\alpha x + \beta$)**	**True Age**
1	Alice	40	25
2	Ed	34	
3	Bob	52	
4		28	
5	Cathy	48	29
6		22	
7		92	

$\boldsymbol{\alpha = 2, \beta = -10}$

Problem 2: Privacy by Obscurity

Node ID	Name	Age ($\alpha x + \beta$)	True Age
1	Alice	40	25
2	Ed	34	22
3	Bob	52	31
4		28	19
5	Cathy	48	29
6		22	16
7		92	51

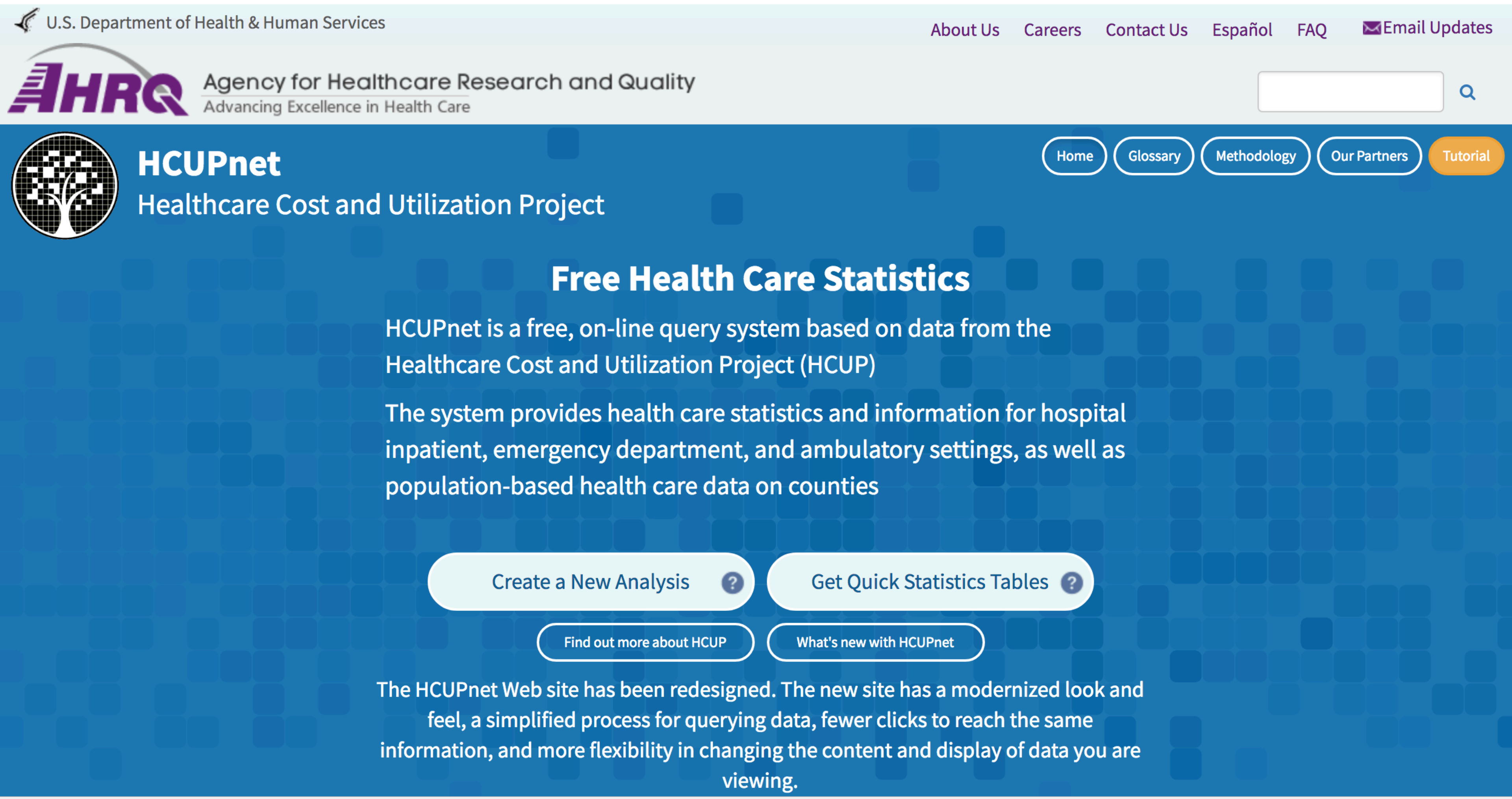
$\boldsymbol{\alpha = 2, \beta = -10}$

Desiderata for a Privacy Definition

1. Resilience to background knowledge
   – A privacy mechanism must be able to protect individuals' privacy from attackers who may possess background knowledge

2. Privacy without obscurity
   – Attacker must be assumed to know the algorithm used as well as all parameters [MK15]

Problem 3: Post-processing



Problem 3: Post-processing

- Publishes tables of counts, for counts that are less than 10, they are suppressed as *

Manage Analysis ?

Analysis Type: Descriptive Statistics **Setting of Care:** Hospital Inpatient **Geographic Settings:** State **Years:** 2009
Categorization Type: Diagnoses--Clinical Classification Software (CCS)
Diagnoses--Clinical Classification Software (CCS): Cancer of ovary **Principal or All-Listed:** Principal
Outcome and Measures: Number
Patient Characteristics: Age groups | Sex | Race/ethnicity | Payer | Location of patient's residence **State:** New Jersey

- Can you tell their values?

Problem 3: Post-processing

Age	#discharges	White	Black	Hispanic	Asian/ Pcf Hlnder	Native American	Other	Missing
#discharges	735	535	82	58	18	*	19	22
1-17	*	*	*	*	*	*	*	*
18-44	70	40	13	*	*	*	*	*
45-64	330	236	31	32	*	*	11	*
65-84	298	229	35	13	*	*	*	*
85+	34	29	*	*	*	*	*	*

Problem 3: Post-processing

Age	#discharges	White	Black	Hispanic	Asian/ Pcf Hlnder	Native American	Other	Missing
#discharges	735	535	82	58	18	**1**	19	22
1-17	**3**	**1**	*	*	*	*	*	*
18-44	70	40	13	*	[illegible]	[illegible]	[illegible]	*
45-64	330	236	31	32	[illegible]	[illegible]	[illegible]	*
65-84	298	229	35	13	*	*	*	*
85+	34	29	*	*	*	*	*	*

$$= 535 - (40+236+229+29)$$

Problem 3: Post-processing

Age	#discharges	White	Black	Hispanic	Asian/ Pcf Hlnder	Native American	Other	Missing
#discharges	735	535	82	58	18	**1**	19	22
1-17	**3**	**1**	[0-2]	[0-2]	[0-2]	[0-2]	[0-2]	[0-2]
18-44	70	40	13	*	*	*	*	*
45-64	330	236	31	32	*	*	11	*
65-84	298	229	35	13	*	*	*	*
85+	34	29	*	*	*	*	*	*

Problem 3: Post-processing

Age	#discharges	White	Black	Hispanic	Asian/ Pcf Hlnder	Native American	Other	Missing
#discharges	735	535	82	58	18	**1**	19	22
1-17	**3**	**1**	[0-2]	[0-2]	[0-2]	[0-2]	[0-2]	[0-2]
18-44	70	40	13	*	*	*	*	*
45-64	330	236	31	32	*	*	11	*
65-84	298	229	35	13	*	*	*	*
85+	34	29	[1-3]	*	*	*	*	*

Can Construct Tight Bounds on Rest of Data

[VSJO 13]

Age	#discharges	White	Black	Hispanic	Asian/ Pcf Hlnder	Native American	Other	Missing
#discharges	735	535	82	58	18	1	19	22
1-17	3	1	[0-2]	[0-2]	[0-1]	[0]	[0-1]	[0-1]
18-44	70	40	13	[9-10]	[0-6]	[0]	[0-6]	[1-8]
45-64	330	236	31	32	[10]	[0]	11	[10]
65-84	298	229	35	13	[2-8]	[1]	[2-8]	[4-10]
85+	34	29	[1-3]	[1-4]	[0-1]	[0]	[0-1]	[0-1]

Desiderata for a Privacy Definition

1. Resilience to background knowledge
   - A privacy mechanism must be able to protect individuals' privacy from attackers who may possess background knowledge

2. Privacy without obscurity
   - Attacker must be assumed to know the algorithm used as well as all parameters [MK15]

3. Post-processing
   - Post-processing the output of a privacy mechanism must not change the privacy guarantee [KL10, MK15]

Problem 4

- Releasing tables that achieve k-anonymity
  - At least k records share the same quasi-identifier
  - E.g. 4-anonymous table by generalization

	Non-Sensitive			Sensitive
	Zip code	**Age**	**Nationality**	**Condition**
1	130**	<30	*	AIDS
2	130**	<30	*	Heart Disease
3	130**	<30	*	Viral Infection
4	130**	<30	*	Viral Infection
5	130**	≥ 40	*	Cancer
6	130**	≥ 40	*	Heart Disease
7	130**	≥ 40	*	Viral Infection
8	130**	≥ 40	*	Viral Infection
9	130**	3*	*	Cancer
10	130**	3*	*	Cancer
11	130**	3*	*	Cancer
12	130**	3*	*	Cancer

(a)

Problem 4: Multiple Releases

- 2 tables of k-anonymous patient records

	Non-Sensitive			Sensitive
	Zip code	Age	Nationality	Condition
1	130**	<30	*	AIDS
2	130**	<30	*	Heart Disease
3	130**	<30	*	Viral Infection
4	130**	<30	*	Viral Infection
5	130**	≥40	*	Cancer
6	130**	≥40	*	Heart Disease
7	130**	≥40	*	Viral Infection
8	130**	≥40	*	Viral Infection
9	130**	3*	*	Cancer
10	130**	3*	*	Cancer
11	130**	3*	*	Cancer
12	130**	3*	*	Cancer

Hospital A (4-anonymous)

	Non-Sensitive			Sensitive
	Zip code	Age	Nationality	Condition
1	130**	<35	*	AIDS
2	130**	<35	*	Tuberculosis
3	130**	<35	*	Flu
4	130**	<35	*	Tuberculosis
5	130**	<35	*	Cancer
6	130**	<35	*	Cancer
7	130**	≥35	*	Cancer
8	130**	≥35	*	Cancer
9	130**	≥35	*	Cancer
10	130**	≥35	*	Tuberculosis
11	130**	≥35	*	Viral Infection
12	130**	≥35	*	Viral Infection

Hospital B (6-anonymous)

- If Alice visited both hospitals and she is 28, can you deduce Alice’s medical condition?

Problem 4: Multiple Releases

- 2 tables of k-anonymous patient records [GKS08]

	Non-Sensitive			Sensitive
	Zip code	Age	Nationality	Condition
1	130**	<30	*	AIDS
2	130**	<30	*	Heart Disease
3	130**	<30	*	Viral Infection
4	130**	<30	*	Viral Infection
5	130**	≥40	*	Cancer
6	130**	≥40	*	Heart Disease
7	130**	≥40	*	Viral Infection
8	130**	≥40	*	Viral Infection
9	130**	3*	*	Cancer
10	130**	3*	*	Cancer
11	130**	3*	*	Cancer
12	130**	3*	*	Cancer

Hospital A (4-anonymous)

	Non-Sensitive			Sensitive
	Zip code	Age	Nationality	Condition
1	130**	<35	*	AIDS
2	130**	<35	*	Tuberculosis
3	130**	<35	*	Flu
4	130**	<35	*	Tuberculosis
5	130**	<35	*	Cancer
6	130**	<35	*	Cancer
7	130**	≥35	*	Cancer
8	130**	≥35	*	Cancer
9	130**	≥35	*	Cancer
10	130**	≥35	*	Tuberculosis
11	130**	≥35	*	Viral Infection
12	130**	≥35	*	Viral Infection

Hospital B (6-anonymous)

- Alice is 28 and she visits both hospitals

Problem 4: Multiple Releases

- 2 tables of k-anonymous patient records [GKS08]

	Non-Sensitive			Sensitive
	Zip code	Age	Nationality	Condition
1	130**	<30	*	AIDS
2	130**	<30	*	Heart Disease
3	130**	<30	*	Viral Infection
4	130**	<30	*	Viral Infection
5	130**	≥40	*	Cancer
6	130**	≥40	*	Heart Disease
7	130**	≥40	*	Viral Infection
8	130**	≥40	*	Viral Infection
9	130**	3*	*	Cancer
10	130**	3*	*	Cancer
11	130**	3*	*	Cancer
12	130**	3*	*	Cancer

Hospital A (4-anonymous)

	Non-Sensitive			Sensitive
	Zip code	Age	Nationality	Condition
1	130**	<35	*	AIDS
2	130**	<35	*	Tuberculosis
3	130**	<35	*	Flu
4	130**	<35	*	Tuberculosis
5	130**	<35	*	Cancer
6	130**	<35	*	Cancer
7	130**	≥35	*	Cancer
8	130**	≥35	*	Cancer
9	130**	≥35	*	Cancer
10	130**	≥35	*	Tuberculosis
11	130**	≥35	*	Viral Infection
12	130**	≥35	*	Viral Infection

Hospital B (6-anonymous)

- 4-anonymity + 6-anonymity $\not\Rightarrow$ k-anonymity , for any k

Desiderata for a Privacy Definition

1. Resilience to background knowledge
   – A privacy mechanism must be able to protect individuals' privacy from attackers who may possess background knowledge

2. Privacy without obscurity
   – Attacker must be assumed to know the algorithm used as well as all parameters [MK15]

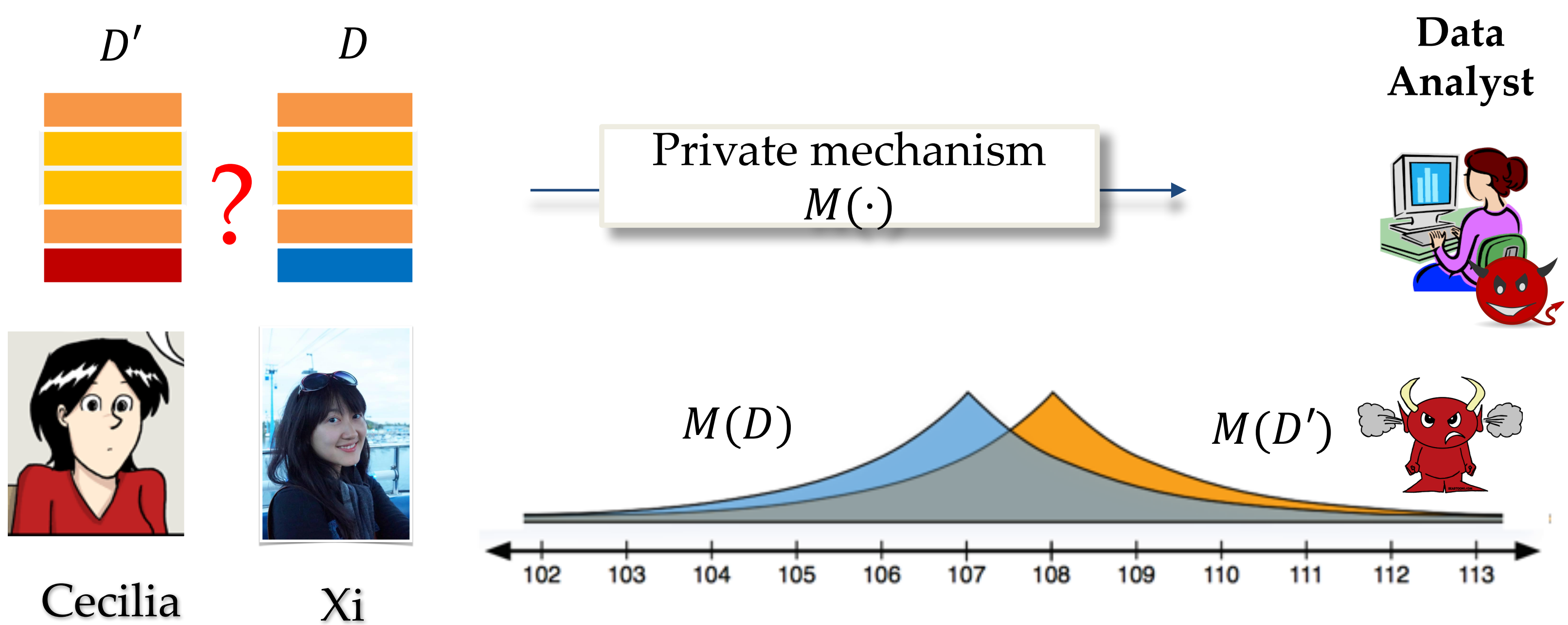
3. Post-processing
   – Post-processing the output of a privacy mechanism must not change the privacy guarantee [KL10, MK15]

4. Composition over multiple releases
   – Allow a graceful degradation of privacy with multiple invocations on the same data [DN03, GKS08]

Differential Privacy

[Dwork06]

- *"An algorithm satisfies differential privacy (DP) if its output is insensitive to adding, removing or changing one record in its input database"*



Differential Privacy

[Dwork ICALP 2006]

For every pair of inputs that differ in one row

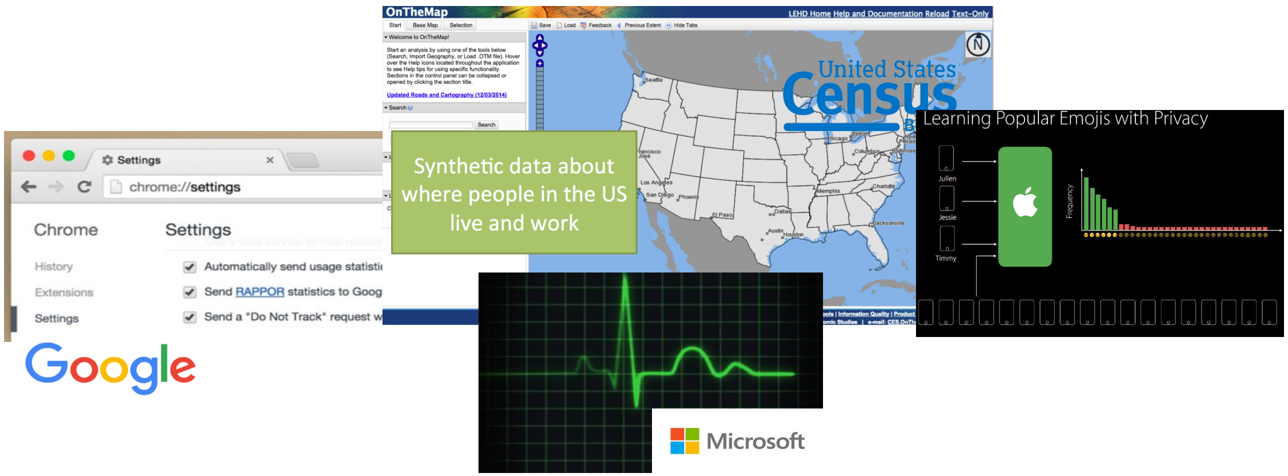
For every output …

D₁ ***D₂*** ***O***

Adversary should not be able to distinguish between any D_1 and D_2 based on any O

$$\ln\left(\frac{\Pr[A(D_1) = o]}{\Pr[A(D_2) = o]}\right) \leq \varepsilon, \qquad \varepsilon > 0$$

Differential Privacy in Practice



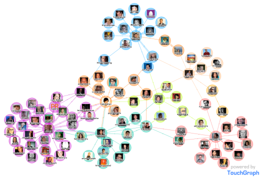
What are the challenges in
building practical systems that ensure DP?

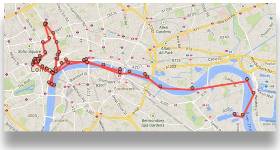
From Theory to Practice

DP mechanisms for
answering linear counting queries
on tabular data

"so many mechanisms, which one to pick?"

"so many definitions, which one to pick?"





"I have my own application,
how to design my own provable privacy guarantee
and how to design mechanism for this guarantee?"

Patient Table

PatientID	Name	Age	...
...	...	...	...
p_4	Alice	60	
...	...	...	...

Foreign Key

Diagnosis Table

PatientID	DoctorID	Timestamp	...
...	...	...	...
p_4	d_2	2017.12.31.9am	
...	...	...	...

Medication Table

PatientID	Medication	Timestamp	...
...	...	...	...
p_4	Aspirin	2017.12.31.10am	
...	...	...	...

Person

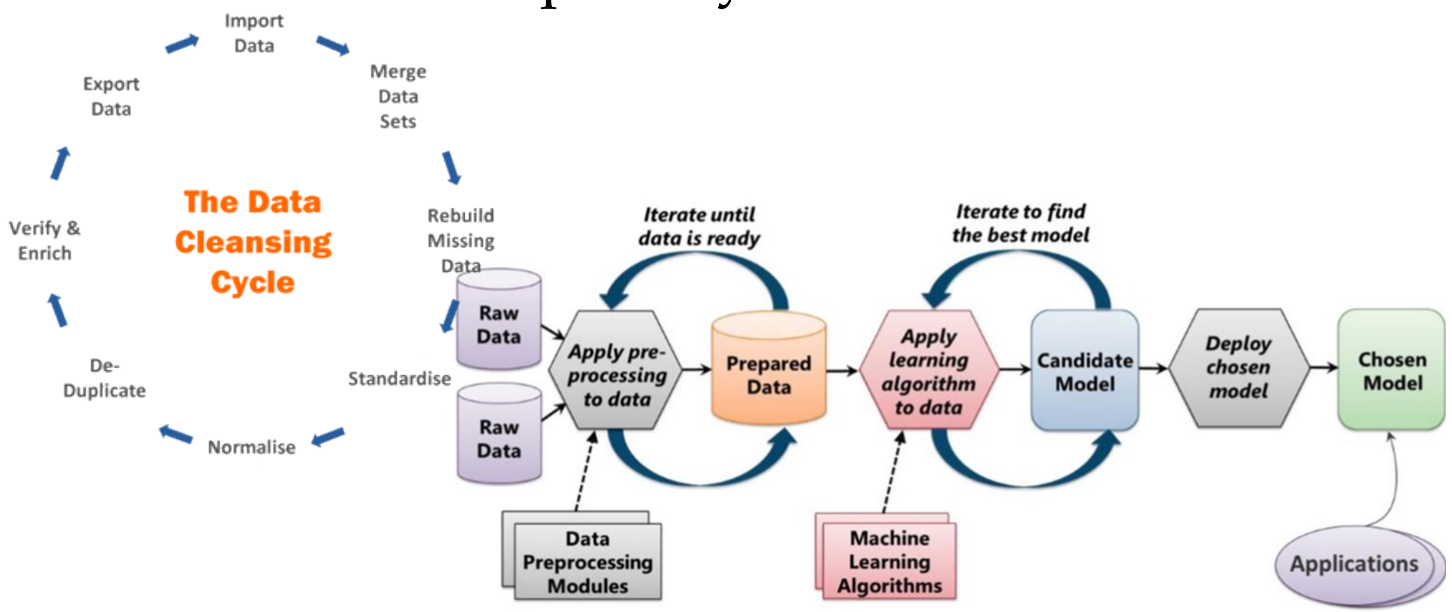
ID	Sex	Age	…	HID
122	M	40	…	H6
123	F	12	…	H6
124	M	23	…	H7
125	M	26	…	H8
126	F	30	…	H8

Household

HID	…	Geo
H6	…	CA
H7	…	FL
H8	…	NC

From Theory to Practice

- Complex data processing workflow
  - Data transformation, repairing, integration, etc.
  - How to track privacy loss?



https://analyticsindiamag.com/get-started-preparing-data-machine-learning/
https://devblogs.microsoft.com/premier-developer/yes-or-no-classification-practical-logistic-regression /

Engineering DP into DB Systems

- Existing DP database systems:
  - PINQ, Airavat, Flex(Uber DP), Google DP, PrivateSQL
  - Rule-based sensitivity analysis of a query plan followed by noise addition
  - Handle more types of data and queries
- But face issues:
  - Inflexible and limited privacy semantics
  - Poor utility guarantee for highly sensitive queries (e.g. involving joins)
  - Unbounded privacy loss
  - Inconsistency between answers etc.

More Questions

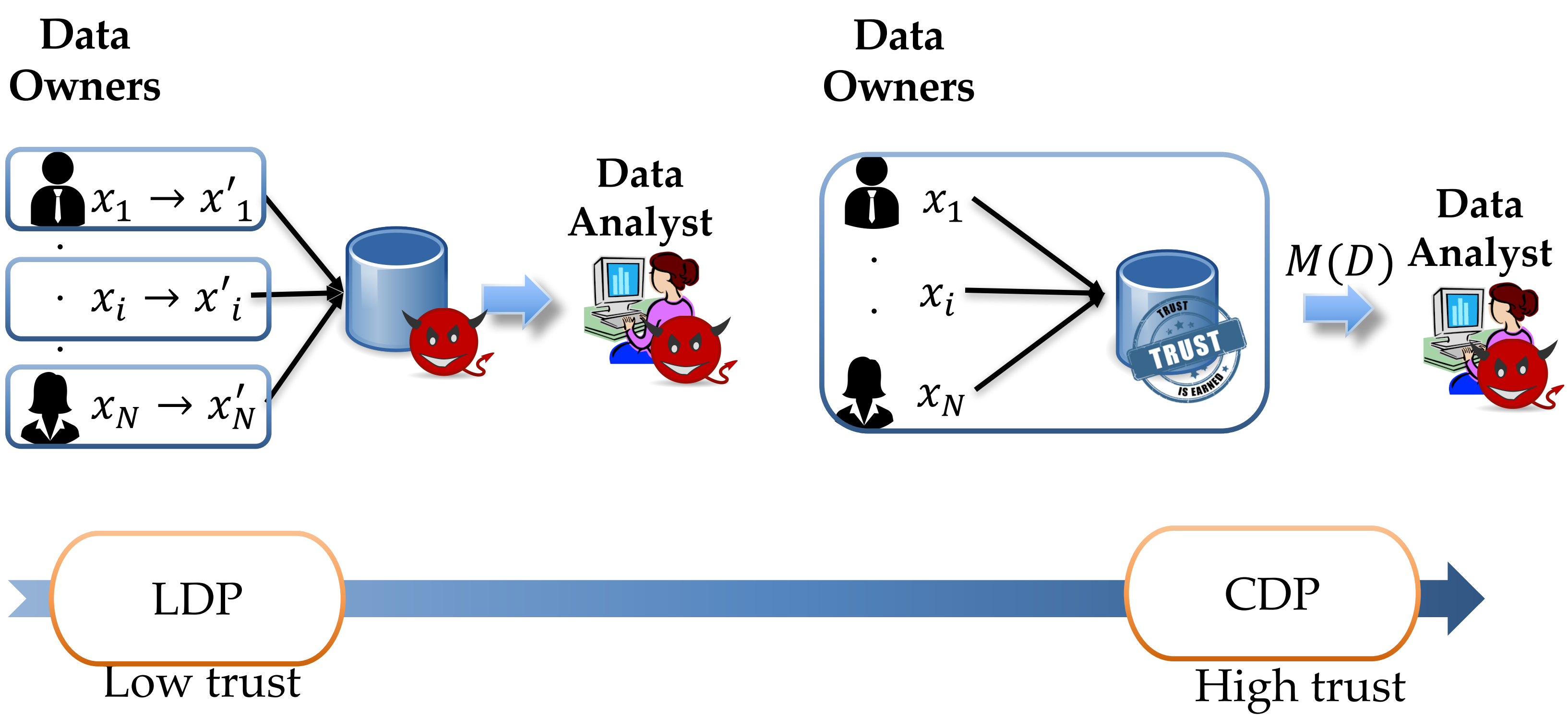
- How to integrate DP into different DB systems?
  - DP program compiler
  - Logical layer vs. physical layer
  - Static vs. dynamic data

- How to verify the correctness of DP implementations?
  - Side channel attacks [HPN11]
  - Floating point issue [Ilvento20,Mironov12]
  - CheckDP [WDKZ20]

- How to support of other privacy requirements?
  - “Rights to be forgotten” by GDPR

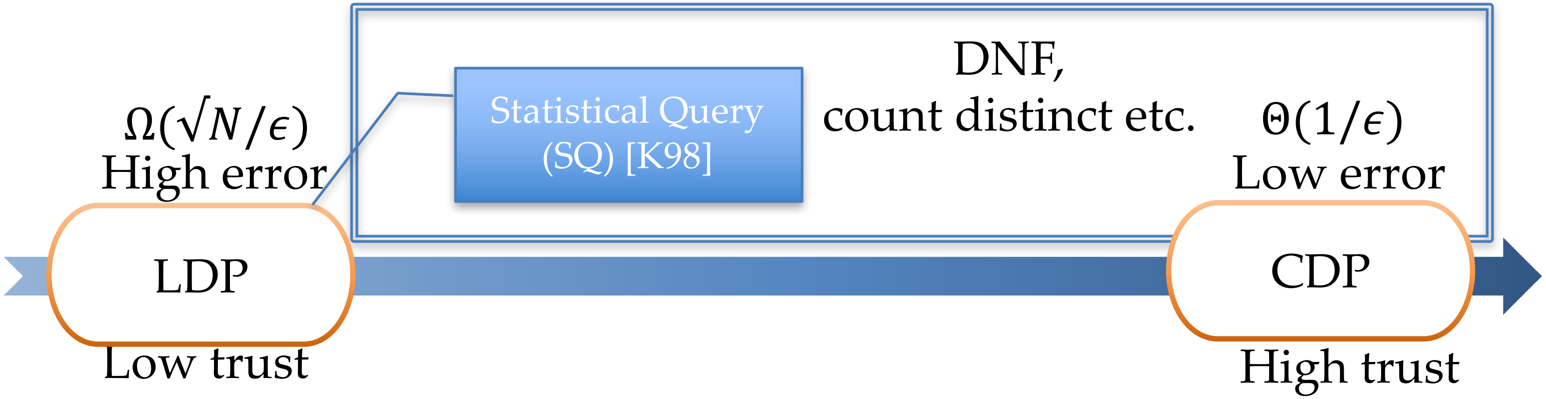
No Trusted Data Curator

- Local DP
  - No trusted data curator
- Centralized DP
  - Trusted data curator



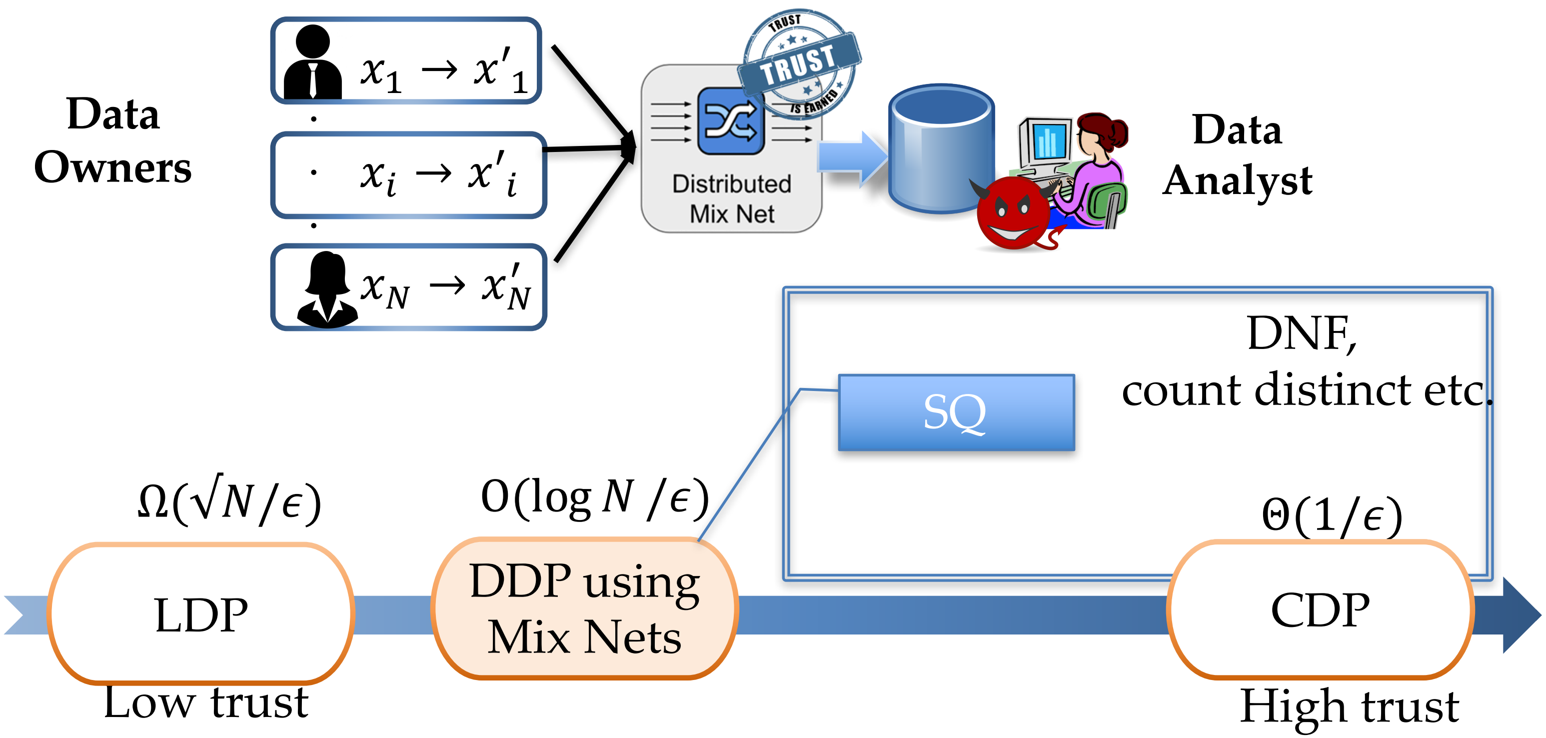
No Trusted Data Curator

- Local DP: Less accurate/expressive
  - $\Omega(\sqrt{N}/\epsilon)$ for statistical counting queries, where N is datasize
  - Separation results between the accuracy and sample complexity of LDP and CDP [KLNRS08]
    - E.g. disjunctive normal form (DNF) queries



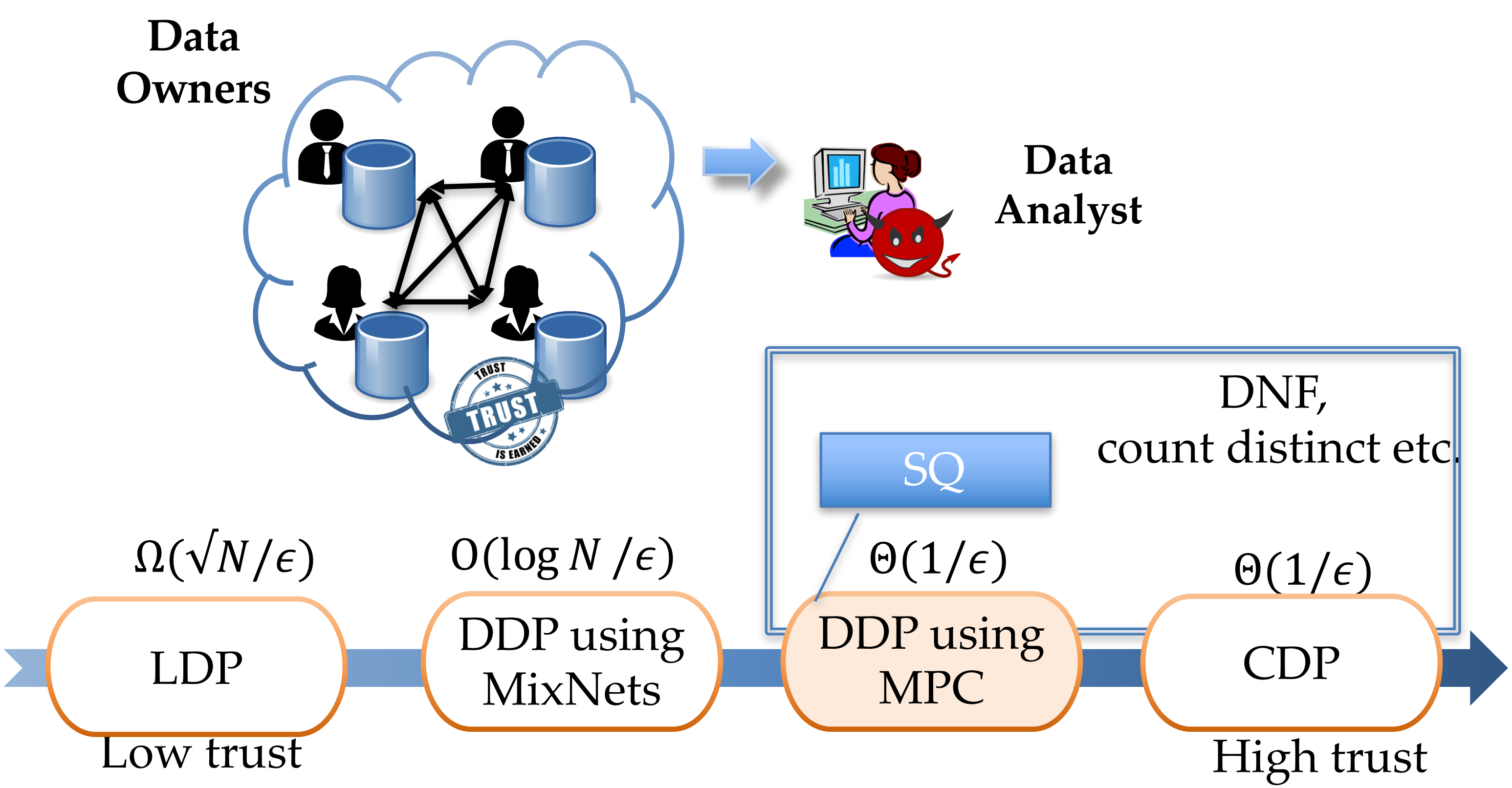
No Trusted Data Curator

- Trusted anonymous communication channels
  [BEMMRLRKTS17, CSUZZ18, EFMRTT19, BBGN19]



No Trusted Data Curator

- Trusted multi-party secure computation (MPC)
  [NH12, BEEGKR17, AHKM18]

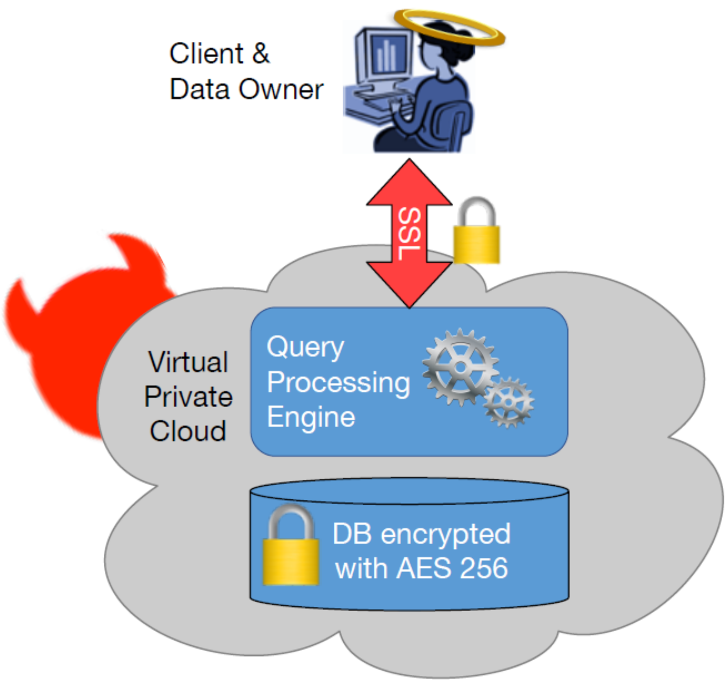


No Trusted Data Curator

- The issues in the centralized setting all remain
- Optimization becomes more complex
  - Privacy, computation and communication cost, query expressiveness and accuracy
  - Hard-coded compiler
- Security/privacy proofs becomes even tricker
  - Even for stand-alone crypto/DP mechanisms [EUROCRYPT 2006, VLDB17]
  - Hybrid approach is vulnerable to faulty proofs [CCS17]
- Additional integrity concerns (storage, query evaluation)
  - Malicious participants who do not follow the protocol

Cloud Service Provider

- Simply encrypted data may do?



What could go wrong?

- Storage: National Security Letter compels service provider to decrypt data
- Query processing: insider threat sees data-dependent query traces and result sizes
- Client side: rogue user systematically queries DB to deduce its private contents

Approaches & Issues

- Improve performance:
  - Property-preserving Encryption
  - Use of secure hardware (TEE)

- Be careful …
  - Improper use of these techniques still leak info [GSBNR16, WCPZWBTG17]
  - How can DP help in this setting?

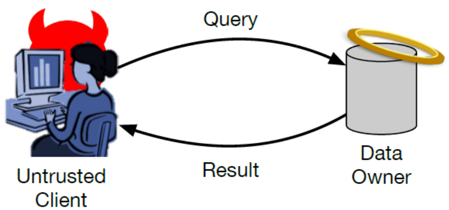
Summary

Privacy Guarantees	Centralized Setting (Client-server)	Federated Setting (Data federation)	Cloud Setting (Cloud service provider)
Input Data	Differential privacy		N/A
Query Evaluation	N/A	Local DP, Secure communication, computation, Encryption, TEE	
Queries	N/A	Private function evaluation	Private information retrieval

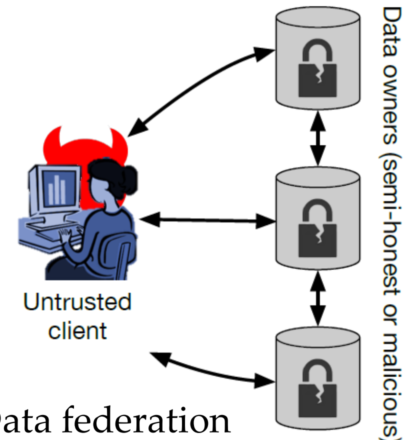
- Existing S&P solutions are piecemeal – they addresses specific steps in the DBMS workflow
- Usually require multiple PhD-level experts to deploy them
- When deployed, their apps are almost always hard-coded
- Composing these techniques is non-trivial

Course Format

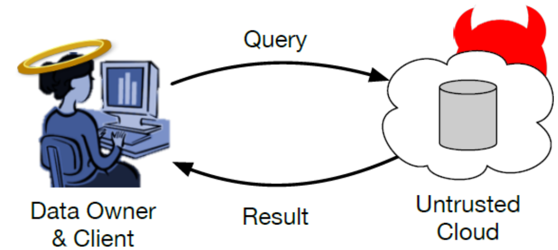
- 3 Modules



Client-server with trusted data curator



Data federation



Cloud service provider

- Each module consists of
  - 1 live/video lecture by the instructor on foundations, classic systems, and related work
  - 1 mini-assignment based on the content of the lecture (offline)
  - 6 paper readings
  - 2-3 live sessions for lecture discussion and student paper presentations

Misc. course info

- **Grading**
  - 3 mini-assignments (individual) 15%
  - 10 paper reviews 10%
  - 1 paper presentation 15%
  - Class participation 10%
  - Project: 50%
- **Website**: https://cs.uwaterloo.ca/~xihe/cs848
  - Schedule (with links to lecture slides, readings, projects, etc.)
- **LEARN** for recorded videos, submission and grades
  - https://learn.uwaterloo.ca/d2l/home/633169
- **Piazza** for questions and discussion

Announcement

- Paper reading assignment survey will be sent soon, please fill it asap, so that students who are presenting in Week 3 will have sufficient time to prepare

- The first round of paper reviews is due before the first paper presentation (Jan 25^{th}, Monday)

Academic Integrity

- See course website
- Mini-assignments and paper reviews are individual work and submission
- Group discussion okay (and encouraged), but
  - Acknowledge help you receive from others
  - Make sure you “own” your solution
- All suspected cases of violation will be aggressively pursued

Next Lecture: Centralized Setting

- We will focus on
  - PINQ
    - Laplace mechanism
    - Global sensitivity analysis of a query plan
  - Flex
    - Sensitivity analysis of query plans with joins
    - Smooth sensitivity mechanism
  - DP under the fire
    - Timing attacks
  - Other related work

Discussion Time