Differential Privacy in Practice

Overview – Data Is the New Oil

- Governments, companies, medical institutions, etc. are collecting & processing lots of sensitive data for legitimate purposes.
- Research institutions often use and share datasets containing private information of individuals with each other.
- Various recent breakthroughs in AI/ML came from techniques that rely on training with large amounts of data.
- Data-driven engineering / decision-making / products has genuinely improved quality of life for everyone and will continue to.
- Unsecure data-handling practices, insufficient security measures and granting readers with improper access to sensitive datasets creates a risk of undesired (and potentially public) disclosure of private information.

Undesired Exposure of Sensitive Information

- Unsecure data-handling – Lax Access Control, unencrypted storage of sensitive data, unencrypted data transfers over network, etc.
- Insufficient Security – Increases likelihood of data breaches, ransomware, etc.
- Data Leakage – Unintentional exposure of sensitive information
    - Leads to a ‘Privacy Breach’
    - Consequence of granting well-meaning users with improper access to sensitive datasets for legitimate reasons but, inadvertently leaking private information inside the dataset
    - Exposing individual’s personally identifiable information (PII) can lead to privacy invasion, discrimination, potential bodily harm, etc.

Context of a Privacy Breach

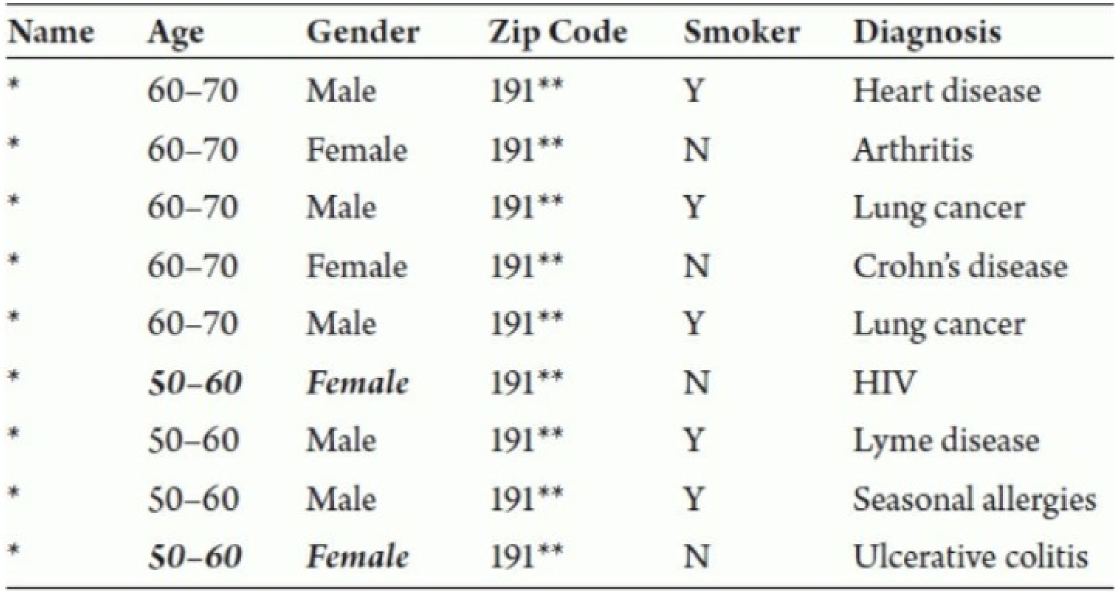
- Sensitive Data – There is a database which potentially contains sensitive information about individuals.
- Data Curator – The database curator has access to the full database and would like to release some statistics from this data to the public.
- Adversaries – An adversary in this case is a party with the intent to reveal, or to learn, at least some of our sensitive data.

Privacy Goal: Ensure that it’s impossible for an adversary to reverse-engineer the sensitive data from what we’ve released.

Could we avoid a Privacy Breach if we anonymized sensitive data?

Why is anonymization hard?

What if we erased unique identifiers of individuals? Would that make our datasets anonymous?

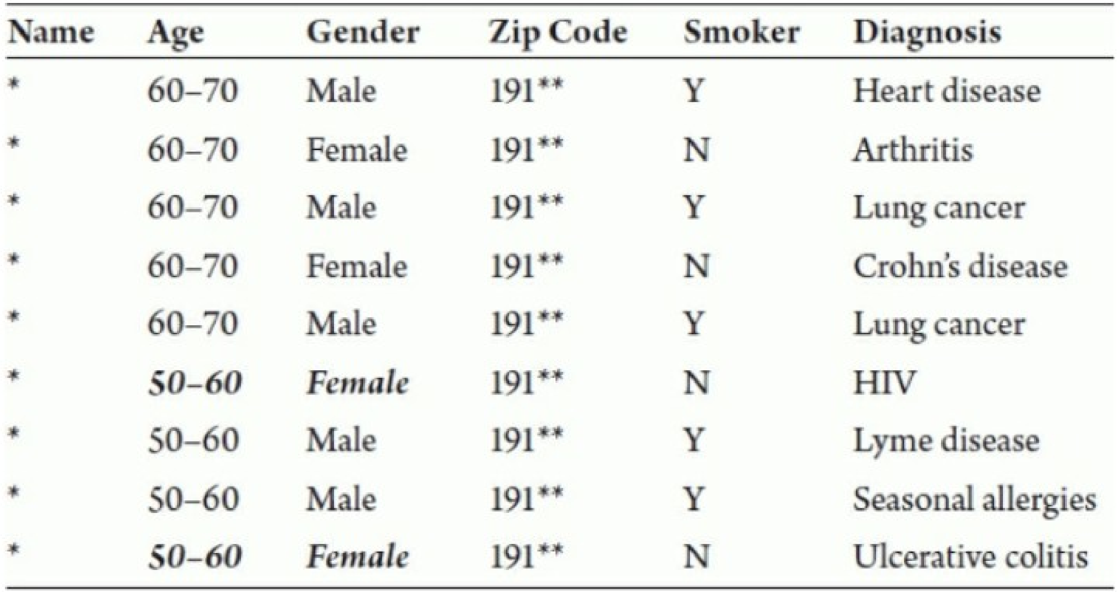


Name	Age	Gender	Zip Code	Smoker	Diagnosis
*	60–70	Male	191**	Y	Heart disease
*	60–70	Female	191**	N	Arthritis
*	60–70	Male	191**	Y	Lung cancer
*	60–70	Female	191**	N	Crohn's disease
*	60–70	Male	191**	Y	Lung cancer
*	***50–60***	***Female***	191**	N	HIV
*	50–60	Male	191**	Y	Lyme disease
*	50–60	Male	191**	Y	Seasonal allergies
*	***50–60***	***Female***	191**	N	Ulcerative colitis

Kearns & Roth, *The Ethical Algorithm*

Why is anonymization hard?

Removing identifiers like names, SSNs, cell phone numbers, etc. might not be enough!



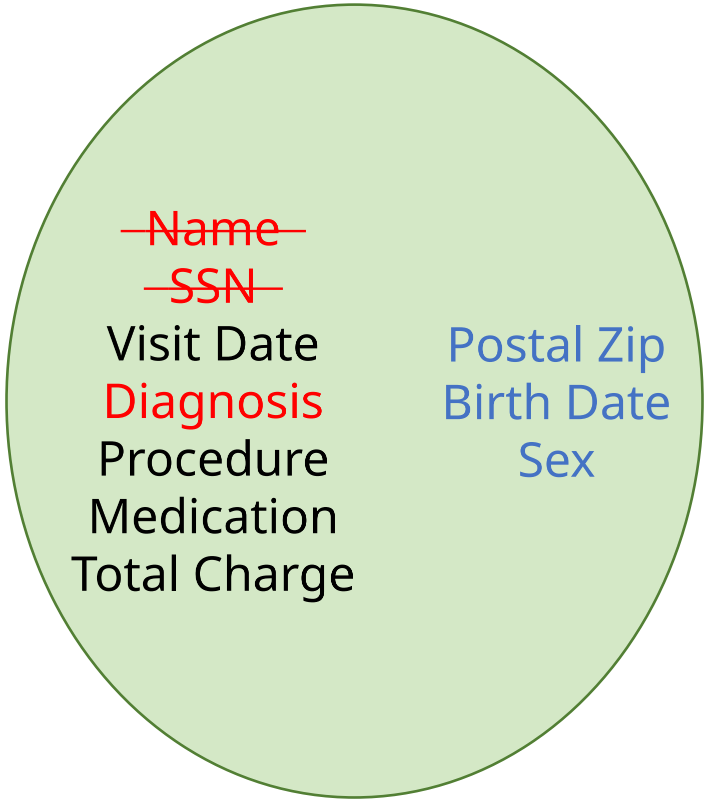
Name	Age	Gender	Zip Code	Smoker	Diagnosis
*	60–70	Male	191**	Y	Heart disease
*	60–70	Female	191**	N	Arthritis
*	60–70	Male	191**	Y	Lung cancer
*	60–70	Female	191**	N	Crohn's disease
*	60–70	Male	191**	Y	Lung cancer
*	***50–60***	***Female***	191**	N	HIV
*	50–60	Male	191**	Y	Lyme disease
*	50–60	Male	191**	Y	Seasonal allergies
*	***50–60***	***Female***	191**	N	Ulcerative colitis

Kearns & Roth, *The Ethical Algorithm*

From this (fictional) hospital database, if we know Rebecca is 55 years ol
and in this database, then we know she has 1 of 2 diseases.

Why is anonymization hard?

The Massachusetts Governor Privacy Breach



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

Postal Zip
Birth Date
Sex

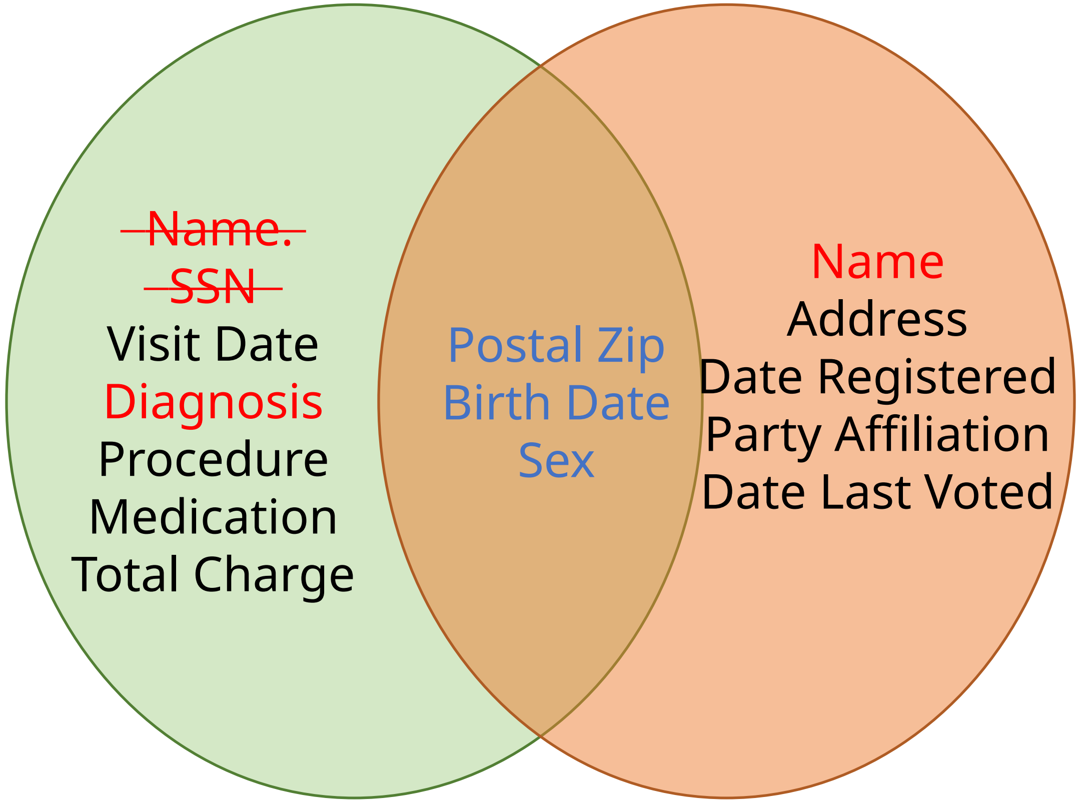
Sensitive data – Hospital visit data on state employees containing PII, medical conditions, etc.

Data Curator – The Massachusetts Group Insurance Commission, which wanted to release anonymized statistics with the goal to help researchers.

William Weld, then Governor of Massachusetts, assured the public that GIC had protected patient privacy by deleting identifiers.

Why is anonymization hard?

The Massachusetts Governor Privacy Breach



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

Postal Zip
Birth Date
Sex

Name
Address
Date Registered
Party Affiliation
Date Last Voted

Adversary – Latanya Sweeney, a graduate student in computer science saw a chance to make a point about the limits of anonymization.

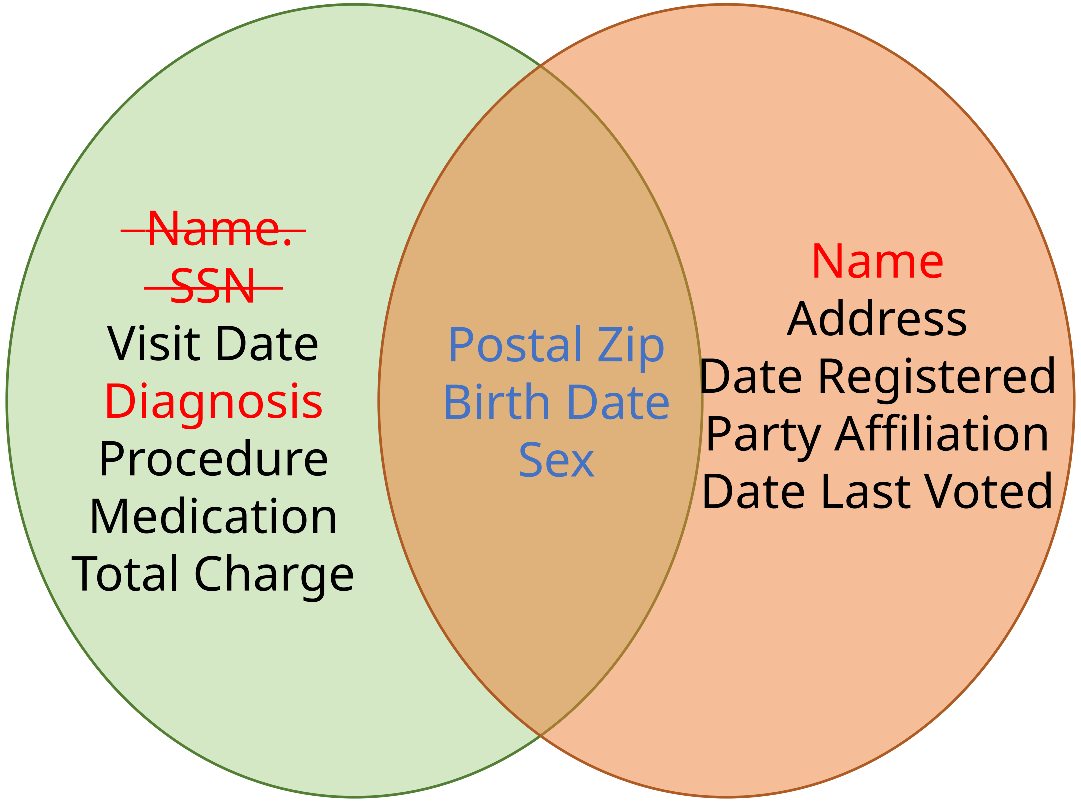
Auxiliary Information –

- The Governor lived in Cambridge, MA
- The Governor’s DoB, Sex, Zip Code from Voter Records

Why is anonymization hard?

The Massachusetts Governor Privacy Breach

Linkage Attack: Governor’s Name linked to Diagnosis

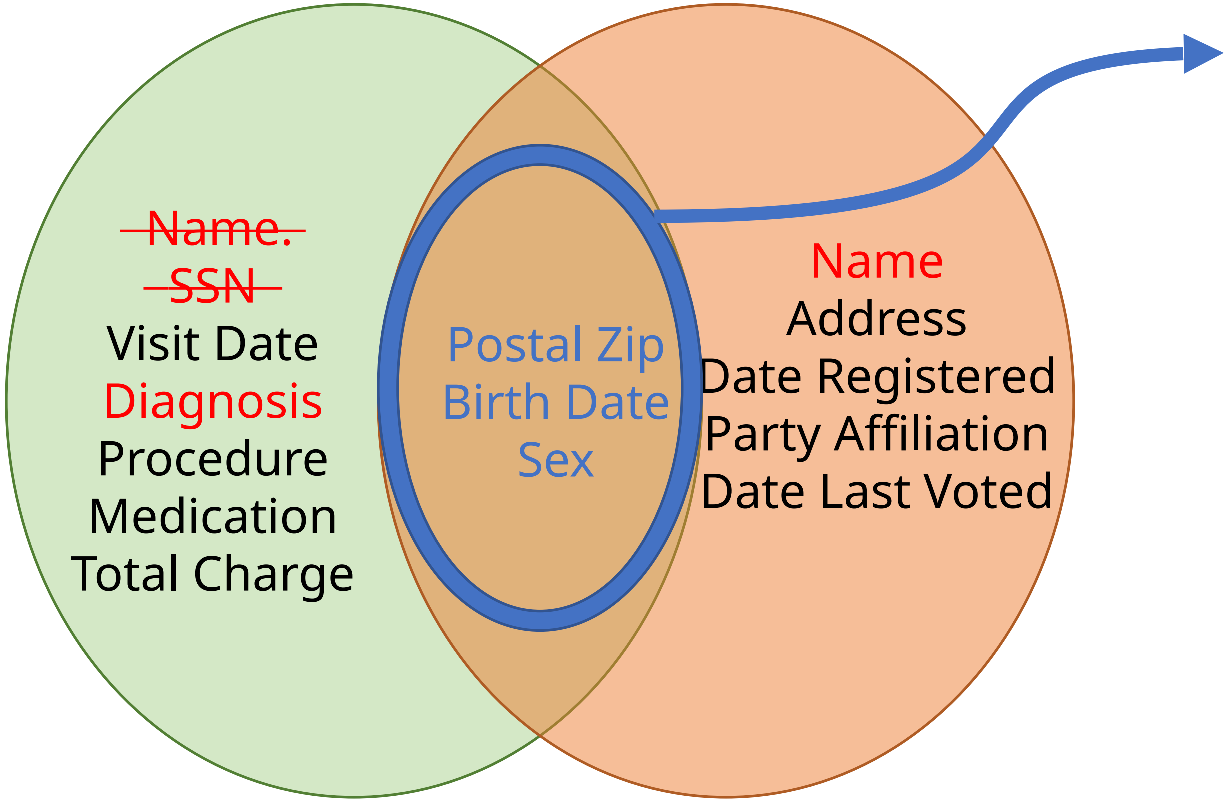


- Cambridge, MA has 54,000 residents and 7 ZIP codes.
- In the anonymized data
  - 6 people shared the governor’s birth date
  - 3 of them were males
  - Only 1, lived in the governor’s zip code

Why is anonymization hard?

The Massachusetts Governor Privacy Breach

Linkage Attack: Achieved By Using Quasi-Identifiers



Using Quasi-Identifiers, the Governor of MA **uniquely identified** using ZipCode, Birth Date, and Sex.

87% of Americans can be **uniquely identified** using only ZipCode, Birth Date, and Sex.

Why is anonymization hard?

Examples of anonymization failures

- AOL researchers released a massive dataset of twenty million search queries for over 650,000 users over a 3-month period. They first "anonymized" the data by scrubbing user IDs and IP addresses
  - But they gave a unique number identifier to each user
  - The search queries were so specific, at times vivid, a lot of users could be identified
- Netflix Challenge (2006), a Kaggle-style competition to improve their movie recommendations, with a $1 million prize
  - They released a dataset consisting of 100 million movie ratings (by “anonymized” numeric user ID), with dates
  - Researchers found they could identify 99% of users who rated 6 or more movies by cross-referencing with IMDB, where people posted reviews publicly with their real names

Could we avoid a Privacy Breach if we anonymized sensitive data?

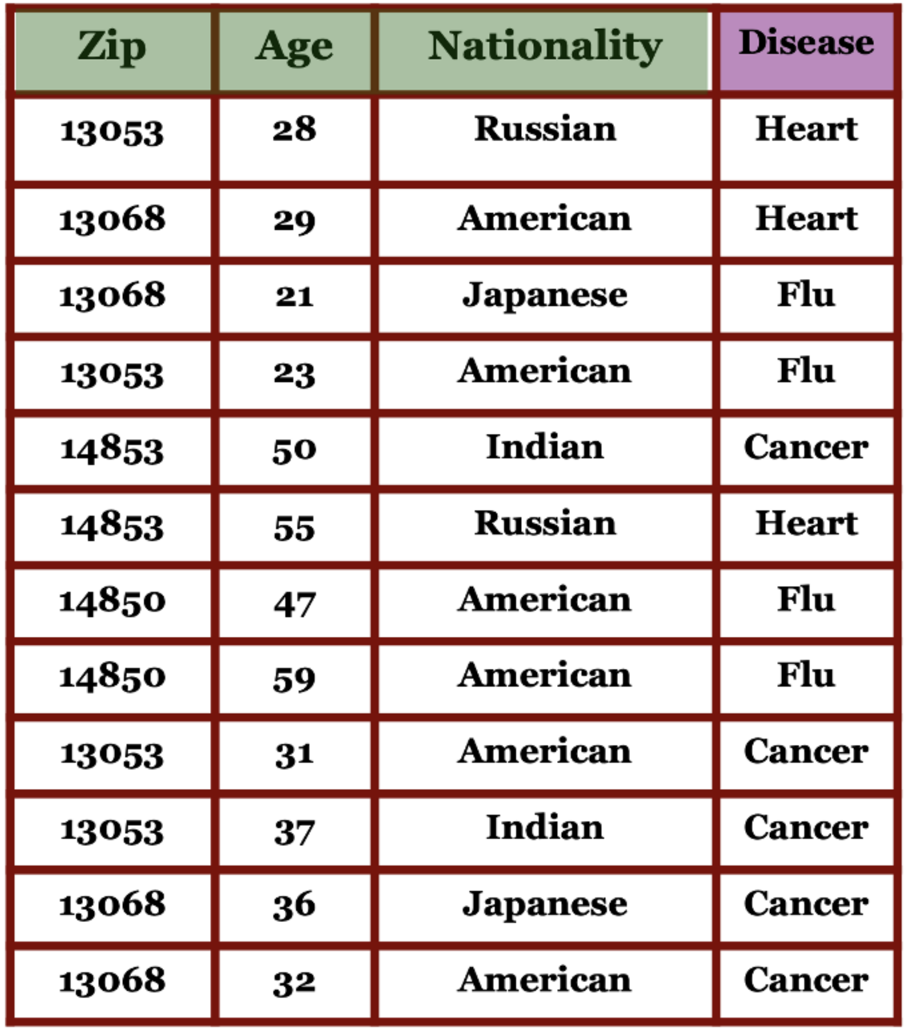
Not if the adversary has auxiliary information from other data sources!

K-Anonymity: Avoiding Linkage Attacks

- If every row corresponds to one individual …

… every row should look like k-1 other rows based on the *quasi-identifier* attributes

K-Anonymity: Sensitive Medical Data

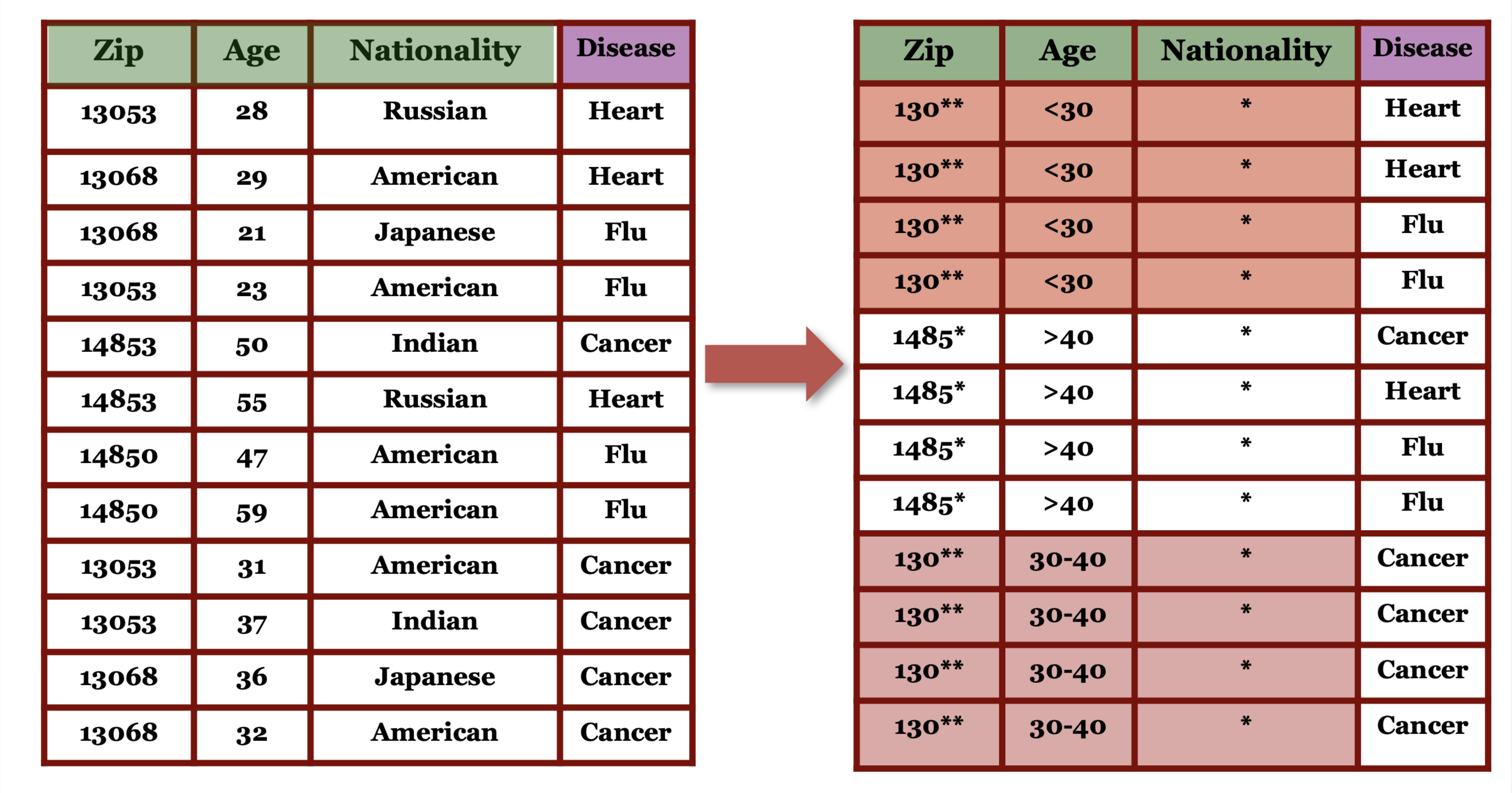


Zip	Age	Nationality	Disease
13053	28	Russian	Heart
13068	29	American	Heart
13068	21	Japanese	Flu
13053	23	American	Flu
14853	50	Indian	Cancer
14853	55	Russian	Heart
14850	47	American	Flu
14850	59	American	Flu
13053	31	American	Cancer
13053	37	Indian	Cancer
13068	36	Japanese	Cancer
13068	32	American	Cancer

There are 3 attributes and 12 records in this data. There are two common methods for achieving *k*-anonymity for some value of *k*:

1. **Suppression**. Certain values of the attributes are replaced by an asterisk "*". All or some values of a column may be replaced by "*". In the anonymized table, we have replaced all the values in the *Nationality* attribute with a "*".
2. **Generalization**. Individual values of attributes are replaced with a broader category. For example, the value ″28" of the attribute *Age* may be replaced by "⩽ 30", the value ″37" by ″30 < Age ⩽ 40", etc.

K-Anonymity: Sensitive Medical Data



Zip	Age	Nationality	Disease
13053	28	Russian	Heart
13068	29	American	Heart
13068	21	Japanese	Flu
13053	23	American	Flu
14853	50	Indian	Cancer
14853	55	Russian	Heart
14850	47	American	Flu
14850	59	American	Flu
13053	31	American	Cancer
13053	37	Indian	Cancer
13068	36	Japanese	Cancer
13068	32	American	Cancer

Zip	Age	Nationality	Disease
130**	<30	*	Heart
130**	<30	*	Heart
130**	<30	*	Flu
130**	<30	*	Flu
1485*	>40	*	Cancer
1485*	>40	*	Heart
1485*	>40	*	Flu
1485*	>40	*	Flu
130**	30-40	*	Cancer
130**	30-40	*	Cancer
130**	30-40	*	Cancer
130**	30-40	*	Cancer

Problem: Background Knowledge breaks K-Anonymity

Adversary knows prior knowledge about Umeko

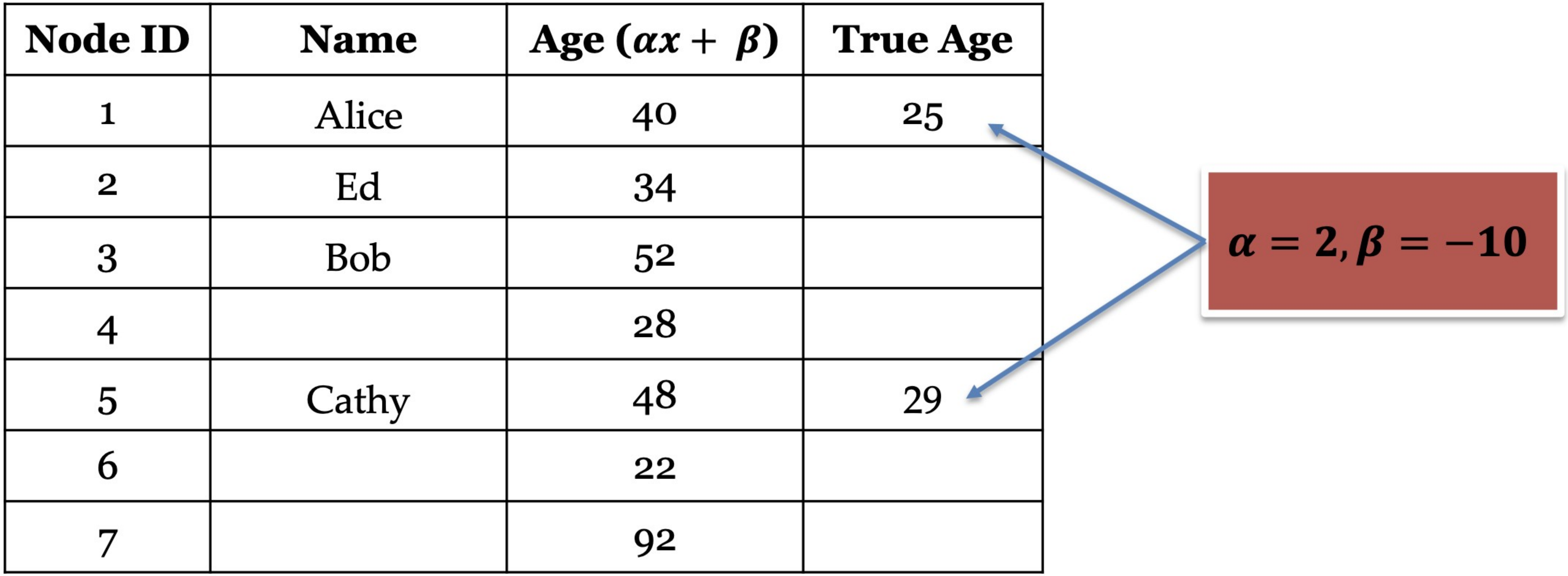
Adversary learns Umeko has Cancer

Name	Zip	Age	Nat.
Umeko	13053	25	Japan

Zip	Age	Nationality	Disease
130**	<30	*	~~Heart~~
130**	<30	*	~~Heart~~
130**	<30	*	Cancer
130**	<30	*	Cancer
1485*	>40	*	Cancer
1485*	>40	*	Heart
1485*	>40	*	Flu
1485*	>40	*	Flu
130**	30-40	*	Cancer
130**	30-40	*	Cancer
130**	30-40	*	Cancer
130**	30-40	*	Cancer

A privacy mechanism must be able to protect individuals’ privacy from attackers who may possess background knowledge

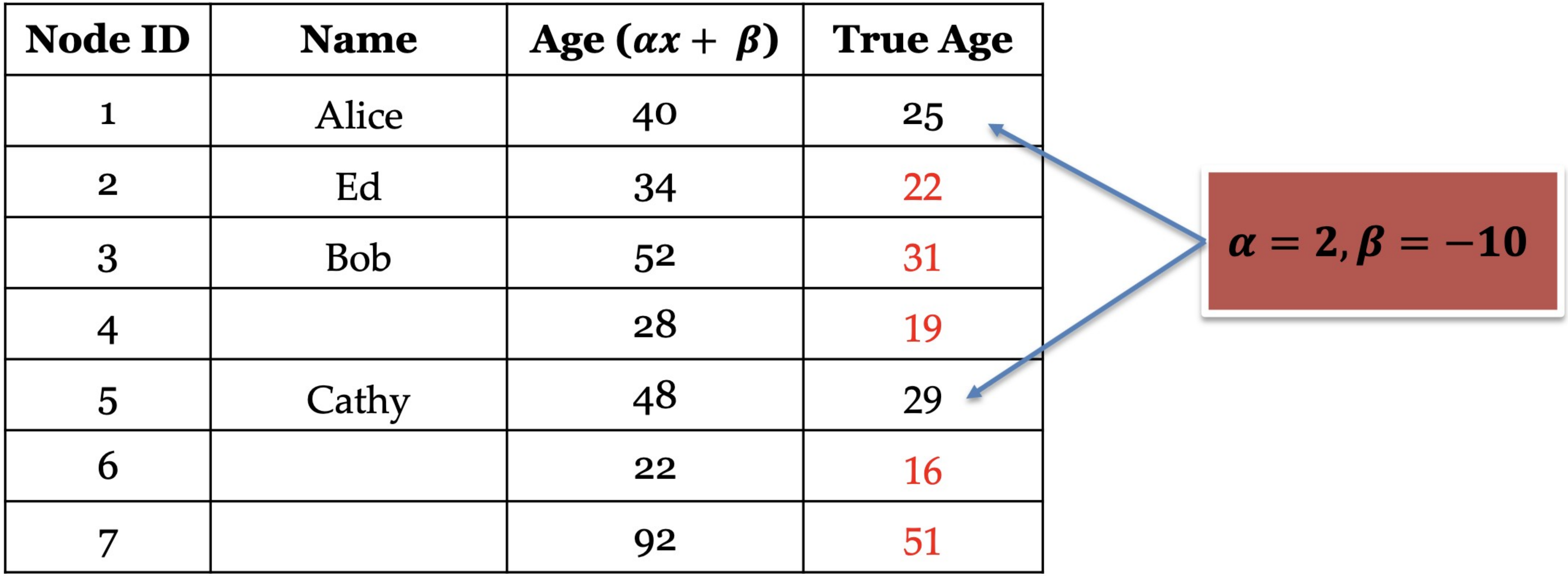
Privacy by Obscuring Sensitive Data



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	

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

Privacy by Obscuring Sensitive Data



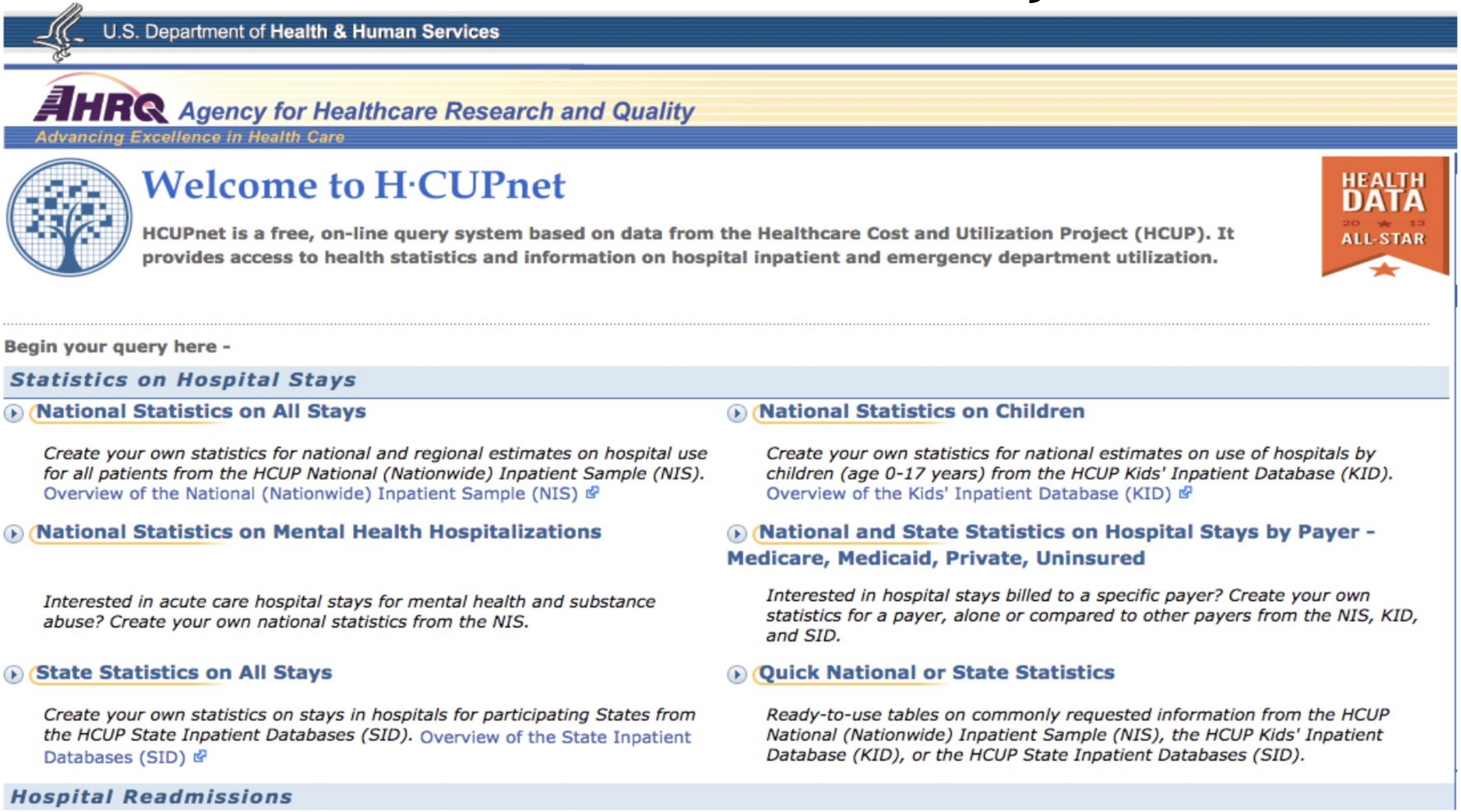
Node ID	Name	Age ($\alpha x + \beta$)	True Age
1	Alice	40	25
2	Ed	34	22
3	Bob	52	31
4		28	19
5	Cathy	48	29
6		22	16
7		92	51

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

Attacker must be assumed to know the algorithm used as well as all parameters

https://dl.acm.org/doi

Healthcare Cost and Utilization Project



U.S. Department of Health & Human Services
AHRQ Agency for Healthcare Research and Quality
Advancing Excellence in Health Care
HEALTH DATA 2013 ALL-STAR

Welcome to H·CUPnet

HCUPnet is a free, on-line query system based on data from the Healthcare Cost and Utilization Project (HCUP). It provides access to health statistics and information on hospital inpatient and emergency department utilization.

Begin your query here -

Statistics on Hospital Stays

National Statistics on All Stays

Create your own statistics for national and regional estimates on hospital use for all patients from the HCUP National (Nationwide) Inpatient Sample (NIS).
Overview of the National (Nationwide) Inpatient Sample (NIS)

National Statistics on Mental Health Hospitalizations

Interested in acute care hospital stays for mental health and substance abuse? Create your own national statistics from the NIS.

State Statistics on All Stays

Create your own statistics on stays in hospitals for participating States from the HCUP State Inpatient Databases (SID). Overview of the State Inpatient Databases (SID)

National Statistics on Children

Create your own statistics for national estimates on use of hospitals by children (age 0-17 years) from the HCUP Kids' Inpatient Database (KID).
Overview of the Kids' Inpatient Database (KID)

National and State Statistics on Hospital Stays by Payer - Medicare, Medicaid, Private, Uninsured

Interested in hospital stays billed to a specific payer? Create your own statistics for a payer, alone or compared to other payers from the NIS, KID, and SID.

Quick National or State Statistics

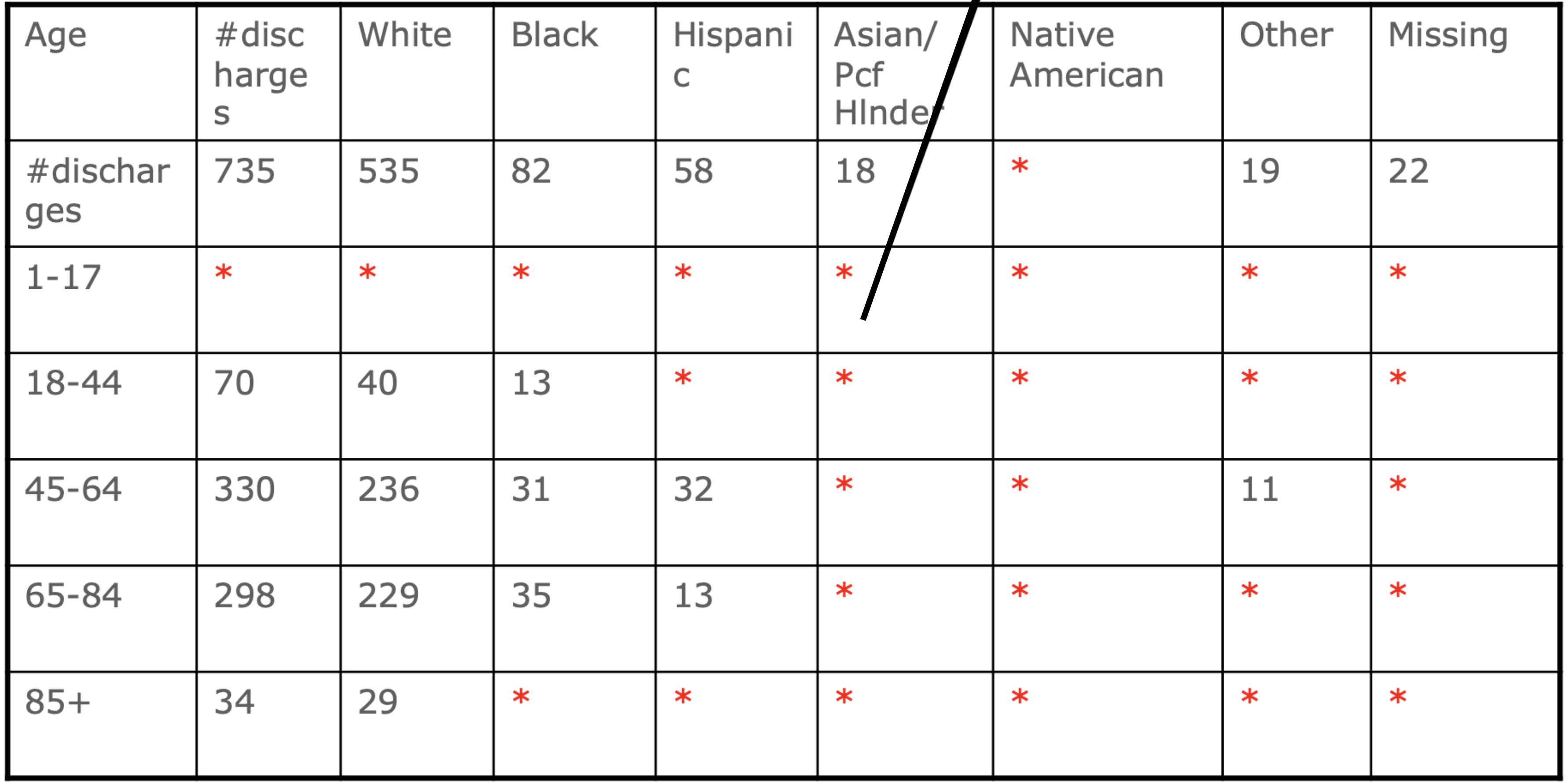
Ready-to-use tables on commonly requested information from the HCUP National (Nationwide) Inpatient Sample (NIS), the HCUP Kids' Inpatient Database (KID), or the HCUP State Inpatient Databases (SID).

Hospital Readmissions

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC384579

#Hospital discharges in NJ of ovarian cancer patients, 2009

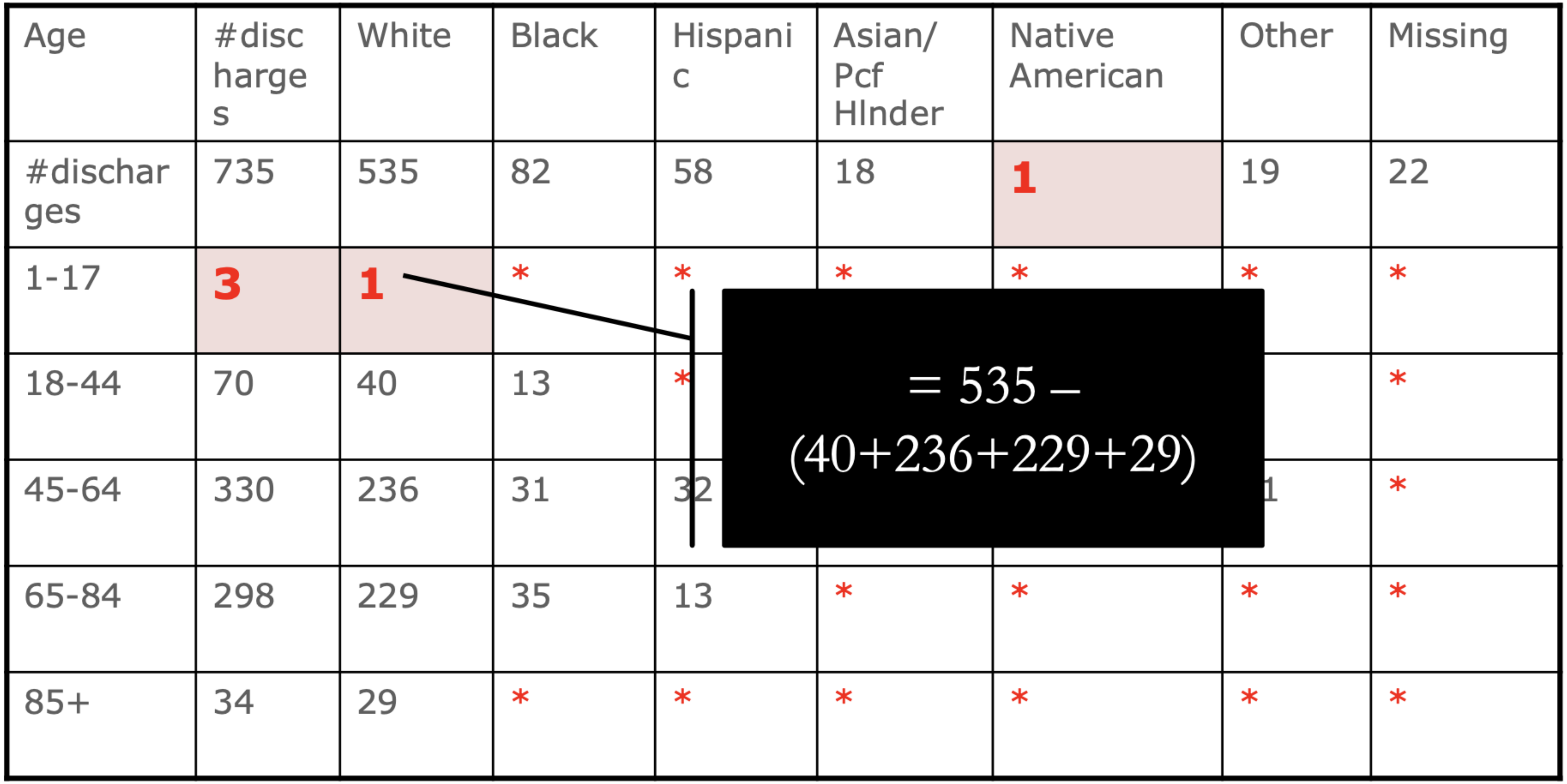
Counts less than k are suppressed achieving k-anonymity



Age	#discharges	White	Black	Hispanic	Asian/ Pcf HInder	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	*	*	*	*	*	*

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC384579

#Hospital discharges in NJ of ovarian cancer patients, 2009

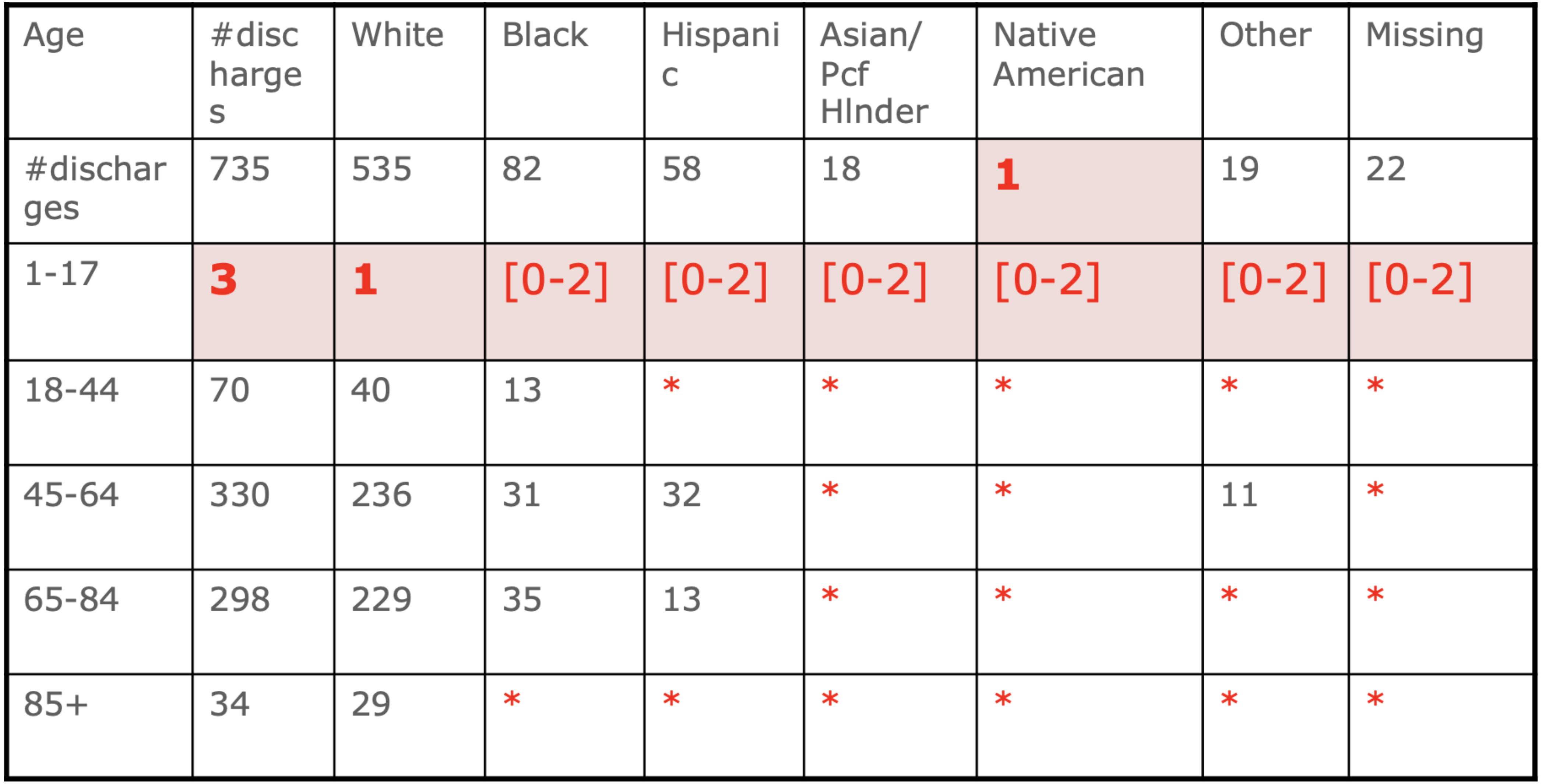


Age	#discharges	White	Black	Hispanic	Asian/ Pcf Hlnder	Native American	Other	Missing
#discharges	735	535	82	58	18	1	19	22
1-17	3	1	*	*	*	*	*	*
18-44	70	40	13	*	[illegible]	[illegible]	[illegible]	*
45-64	330	236	31	32	[illegible]	[illegible]	[illegible]	*
65-84	298	229	35	13	*	*	*	*
85+	34	29	*	*	*	*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

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

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC384579

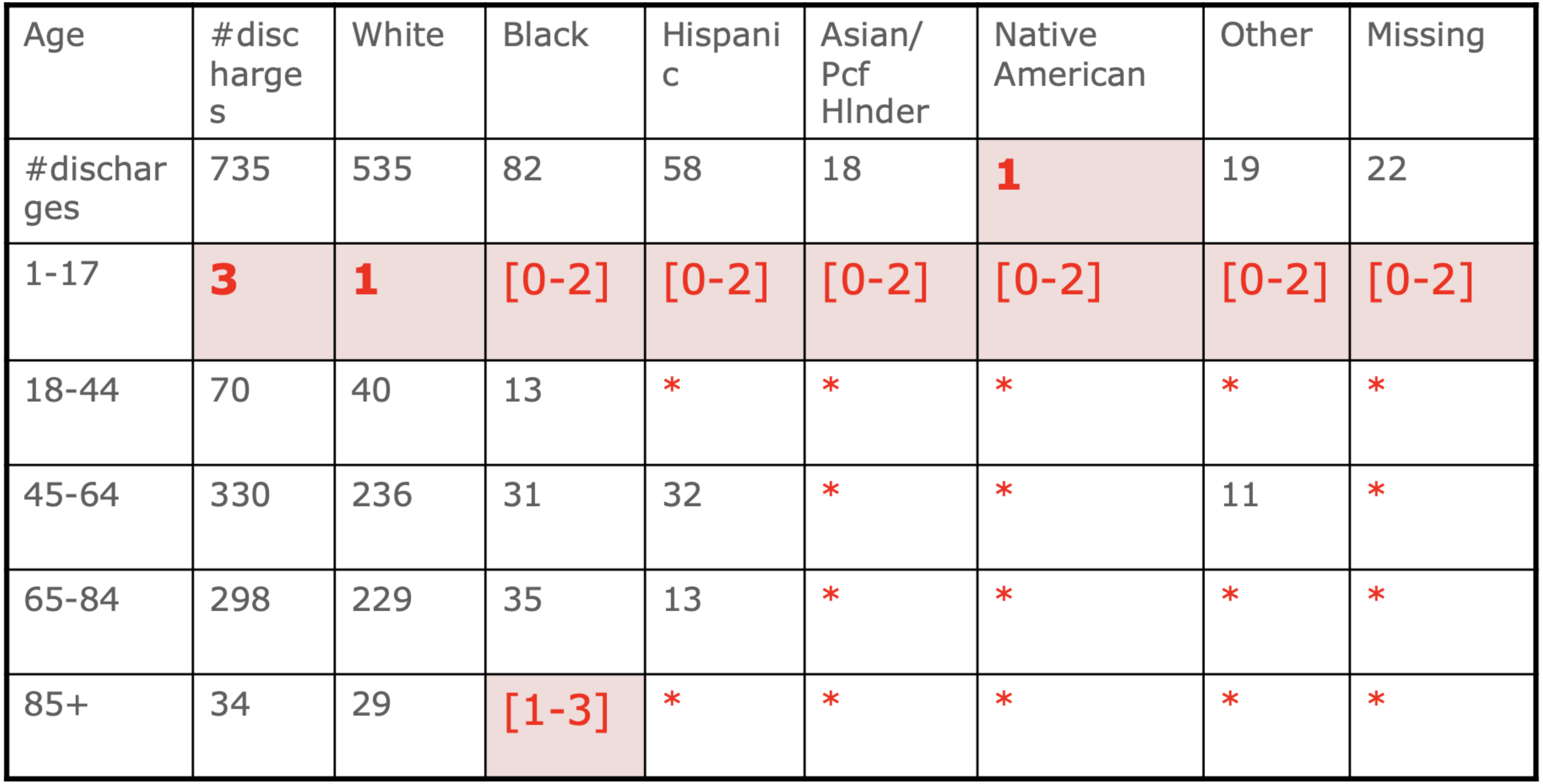
#Hospital discharges in NJ of ovarian cancer patients, 2009



Age	#discharges	White	Black	Hispanic	Asian/ Pcf HInder	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	*	*	*	*	*	*

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC384579

#Hospital discharges in NJ of ovarian cancer patients, 2009



Age	#discharges	White	Black	Hispanic	Asian/Pcf HInder	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]	*	*	*	*	*

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC384579

#Hospital discharges in NJ of ovarian cancer patients, 2009

Based on several additional queries, we can figure out

- Exactly 1 Native American woman diagnosed with ovarian cancer went to a privately owned, not for profit, teaching, hospital with more than 435 beds in 2009.
- Furthermore, the woman did not pay by private insurance, had a routine discharge, with a stay in the hospital of 33.5 days, with her home residence being in a county with 1 million plus residents (large fringe metro, suburbs)
- Using a separate query that tabulates the mean age of the patients with respect to the categorization by ethnicity, we can infer the age of the Native American woman was 75.
- We can even figure out that the woman's hospital costs (which are also suppressed) are exactly $ 32,970.

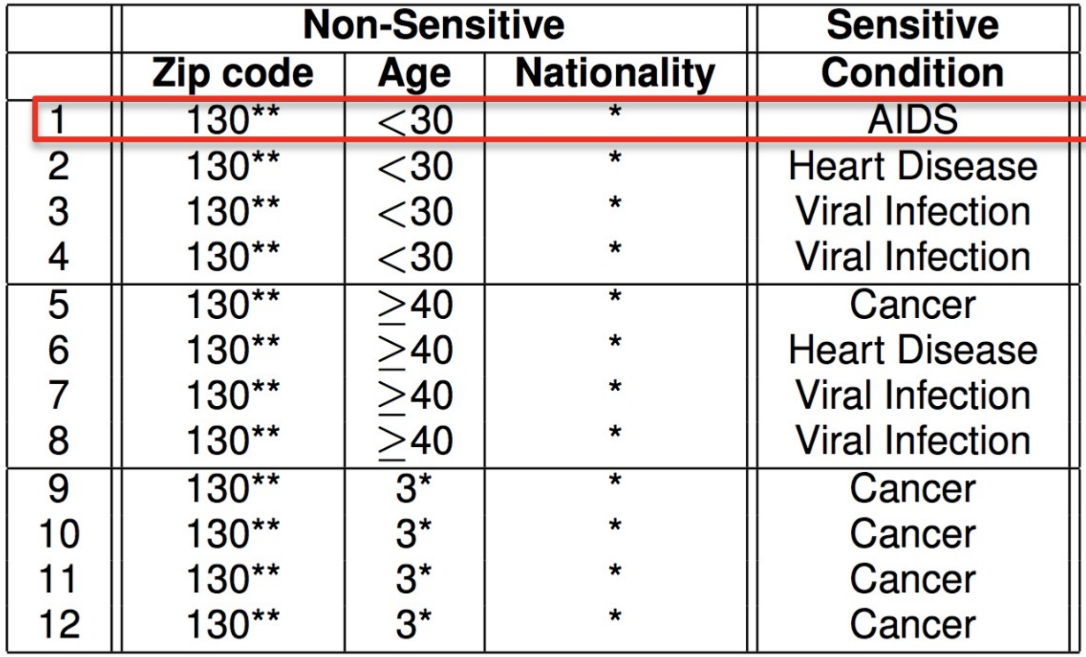
In this case, we have the gender, the age, and an indication of the zip-code based on the NJ state counties with more than 1 million residents. Given that zip-code, age and gender can uniquely identify 87% of all Americans, there

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC384579

Post-processing the output of a privacy mechanism must not change the privacy guarantee

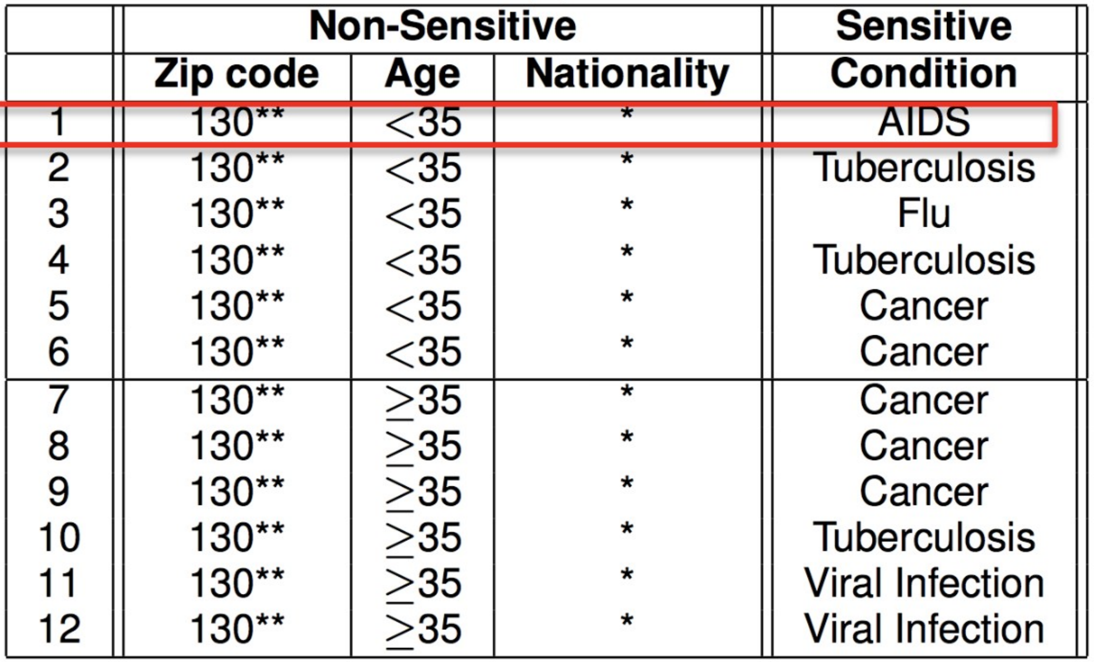
Multiple Releases of Records

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)



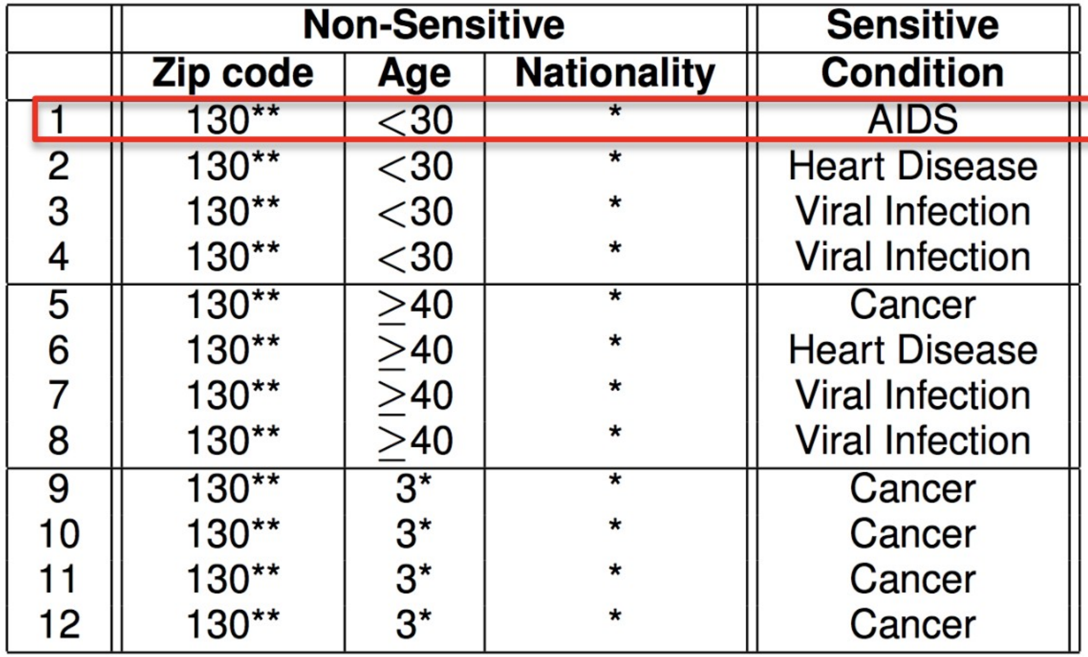
	Non-Sensitive			Sensitive
	Zip code	Age	Nationality	Condition
1	130**	<35	*	AIDS
2	130**	<35	*	Tuberculosis
3	130**	<35	*	Flu
4	130**	<35	*	Tuberculosis
5	130**	<35	*	Cancer
6	130**	<35	*	Cancer
7	130**	≥35	*	Cancer
8	130**	≥35	*	Cancer
9	130**	≥35	*	Cancer
10	130**	≥35	*	Tuberculosis
11	130**	≥35	*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

Hospital B (6-anonymous)

Alice is 28 and she visits both hospitals

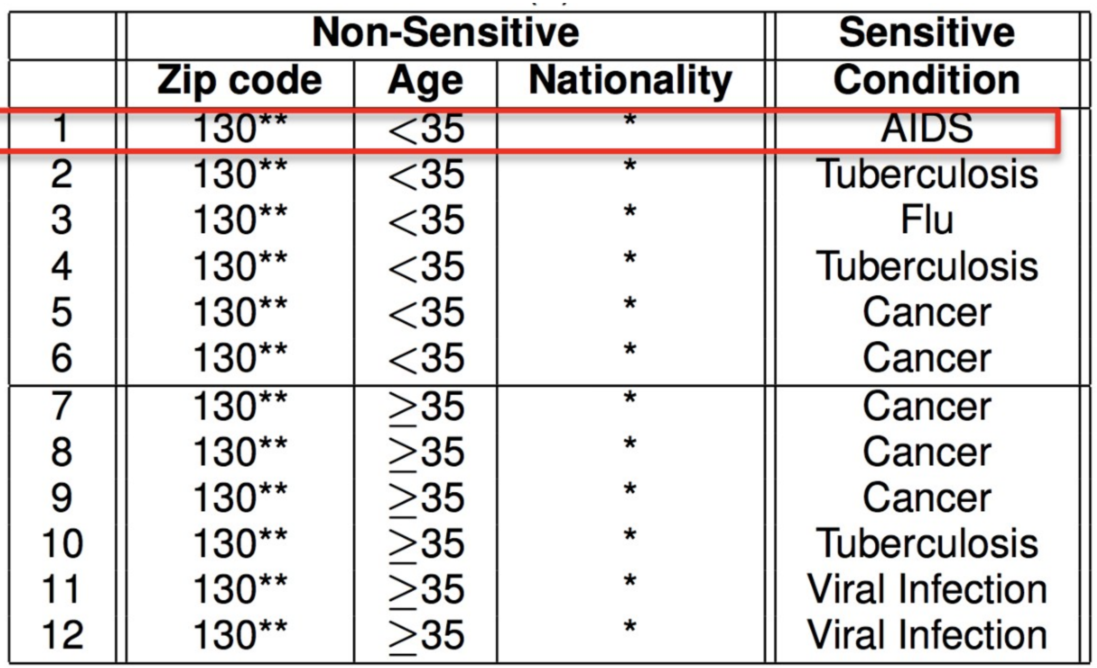
Multiple Releases of Records

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)

4-anonymity + 6-anonymity ⇏ k-anonymity ,for any k

Allow a graceful degradation of privacy with multiple invocations on the same data

Desiderata for Definition of Privacy

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
3. Post-processing
    - Post-processing the output of a privacy mechanism must not change the privacy guarantee
4. Composition over multiple releases
    - Allow a graceful degradation of privacy with multiple invocations on the same data

Differential privacy makes it possible for Curators (Governments, Companies, etc.) to collect Sensitive Data and share aggregate information about user habits, while maintaining the privacy of individual users from Adversaries!

What is Differential Privacy in Practice?

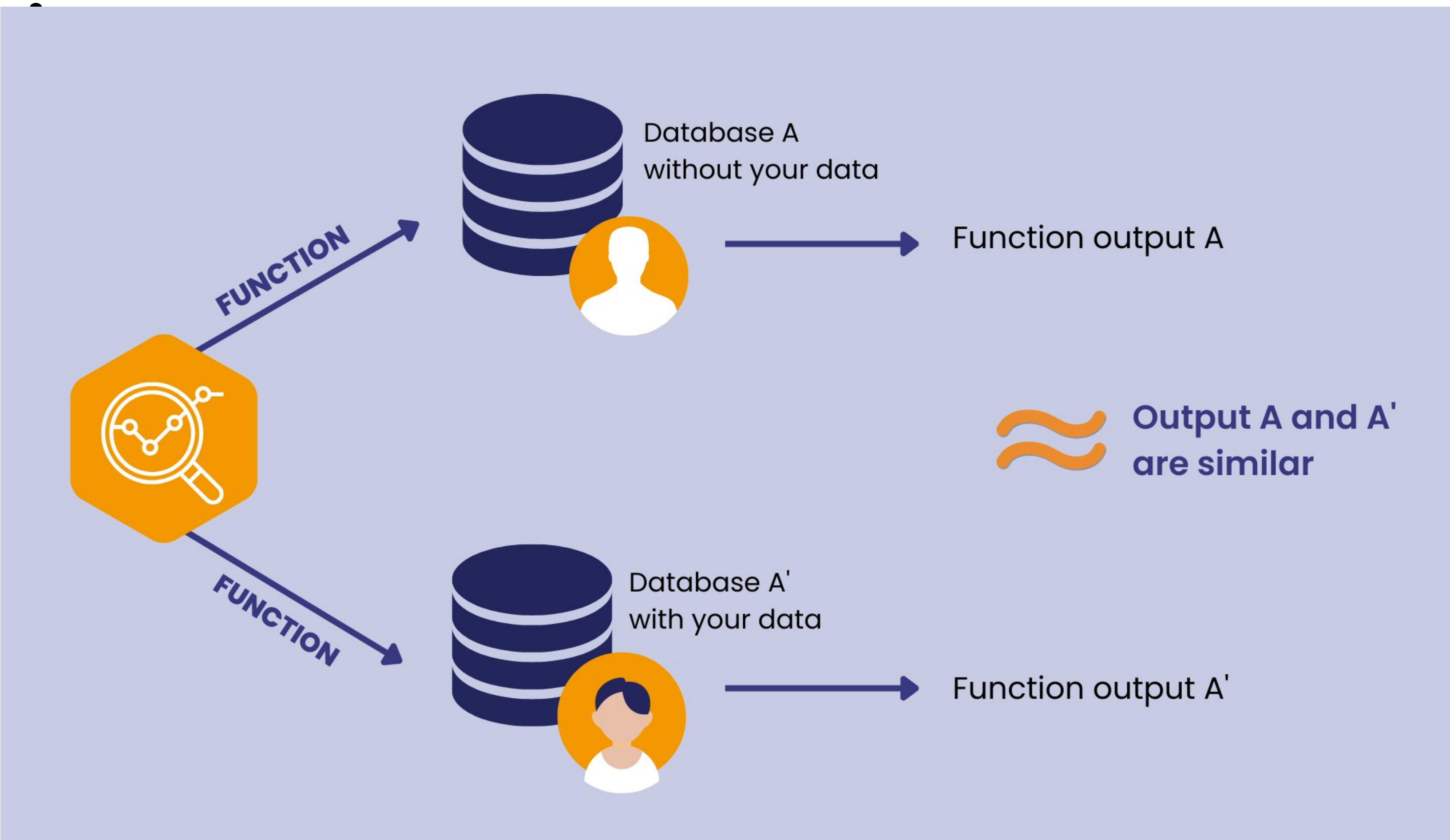
Mathematical Framework for ensuring privacy of individuals in a dataset while maintaining its usefulness

- Framework, Not an algorithm – Approaches to constrain algorithms operating on statistical datasets from leaking private information.
- Strong guarantee of privacy – Allows data analysis without revealing sensitive information even if the adversary has unlimited computing power and complete knowledge of the algorithm and system used to collect and analyze the data.
- Maintaining usefulness of dataset – Retain the ability to provide useful answers to descriptive questions about the dataset
- Provable guarantee – Provides a measure of privacy guarantee which can be mathematically proven
- Futureproof - Even if the adversary were to develop new & sophisticated methods to learn sensitive information from the data, or if new additional information becomes available, the exact same privacy guarantee is maintained.

How does the Industry define Differential Privacy?

- The concept was first introduced in **2006 by Cynthia Dwork and Frank McSherry, et al.** in two papers titled " Calibrating Noise to Sensitivity in Private Data Analysis" and "Differential Privacy".
- Per their definition, the presence or absence of any individual record in the dataset should not significantly affect the outcome of the 'mechanism'
- A 'mechanism' is a randomized computation or an analytical process whose output changes probabilistically every-time for a given input.

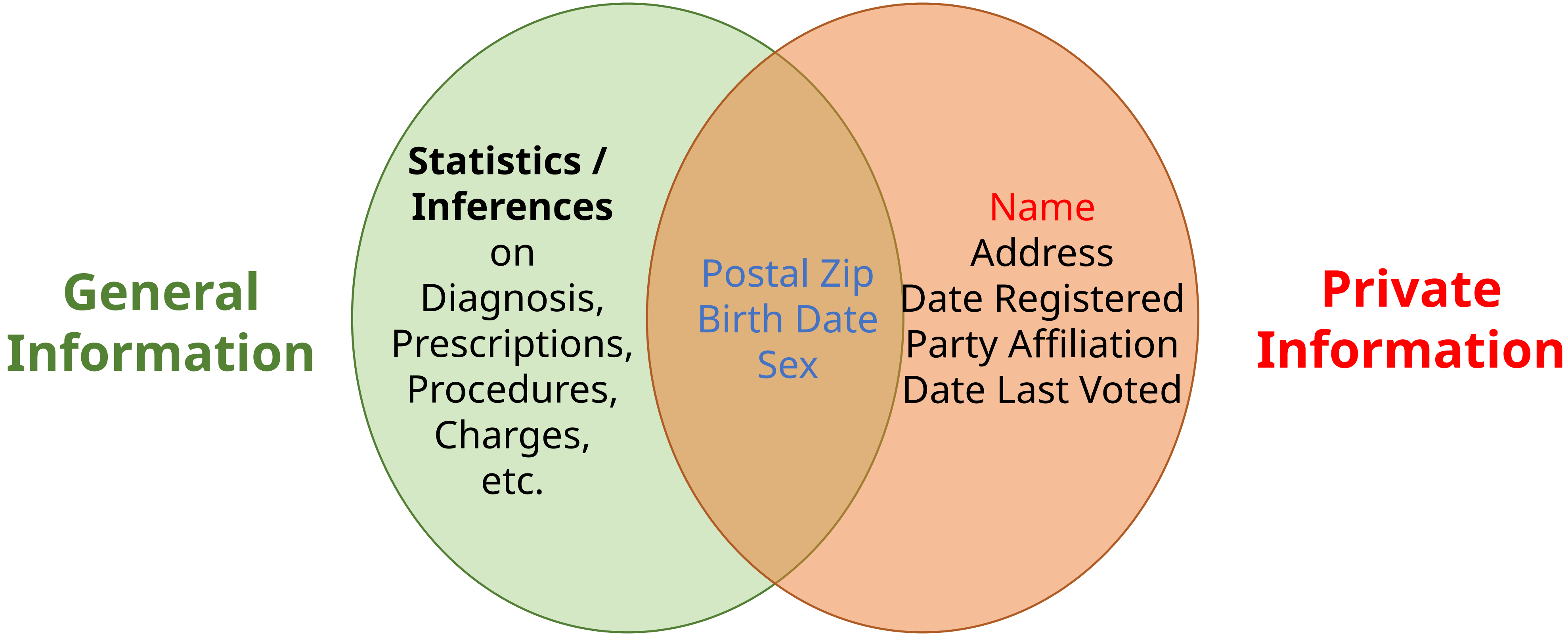
How does the Industry define Differential Privacy?



A computation or analytical process differentially private if its output is almost the same when applied to two datasets that differ only in a single record.

https://www.statice.ai

General Information vs. Private Informatio



General Information

Statistics / Inferences on Diagnosis, Prescriptions, Procedures, Charges, etc.

Postal Zip
Birth Date
Sex

Name
Address
Date Registered
Party Affiliation
Date Last Voted

Private Information

Differential Privacy Information Guarantees

What does it guarantee?

- DP mathematically guarantees that anyone seeing the result of a differentially private analysis will essentially make the same inference about any individual's private information, whether that individual's private information is included in the input to the analysis.
- DP provides a mathematically provable guarantee of privacy protection against a wide range of privacy attacks (e.g., *differencing attack*, *linkage attacks,* and *reconstruction attacks).*

What does it NOT guarantee?

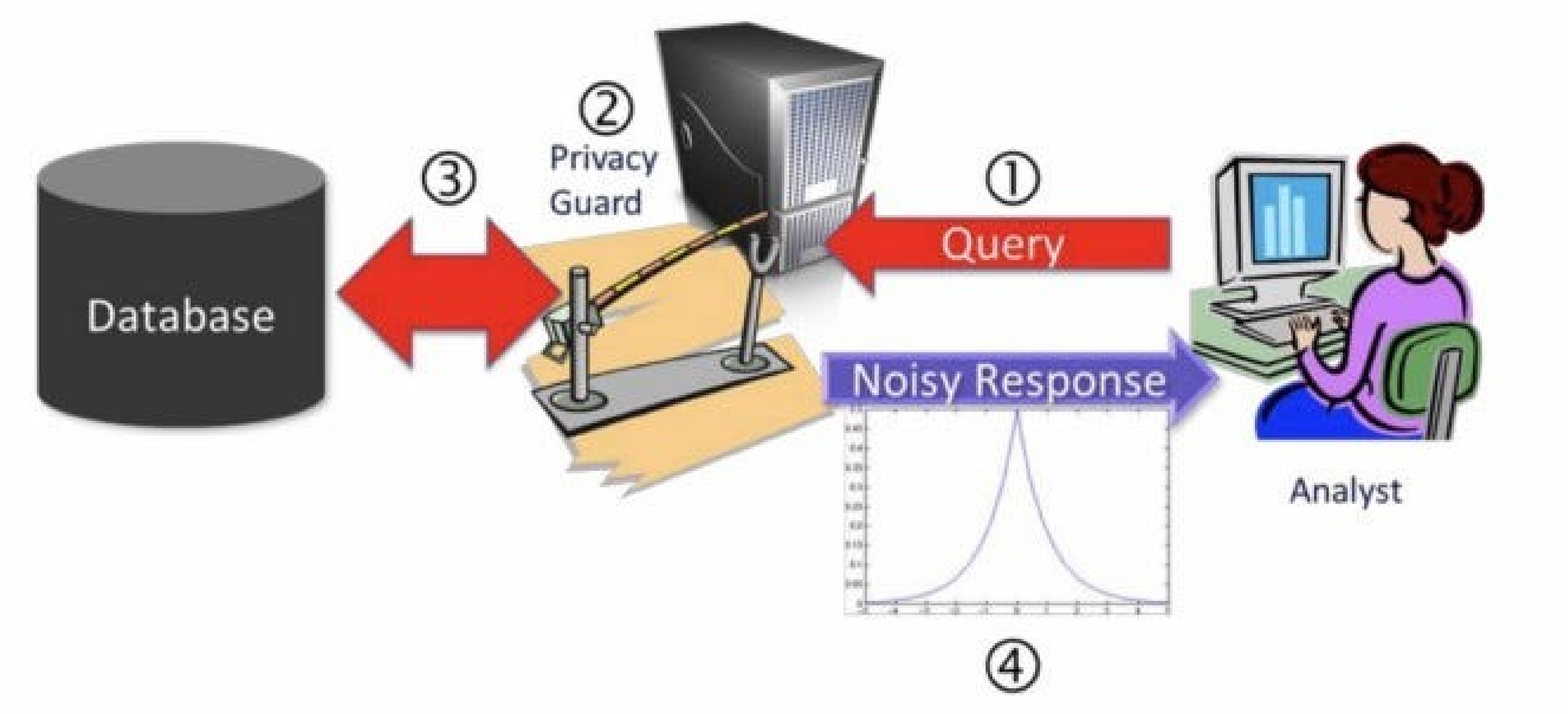
- DP does not guarantee what one "believes" to be their secrets will remain secret. It's important to identify which is general information and which is private information to get benefits from DP umbrella and reduce harm. DP guarantees to protect only private information. If one's secret is general information, it will be not be protected.
- Differential privacy does not prevent statistics and machine learning.

Differential Privacy Information Guarantees

To understand this, let’s consider a scenario when you, a smoker, decided to be included in a survey. Then, analysis on the survey data reveals that smoking causes cancer. Will you, as a smoker, be harmed by the analysis? Perhaps — Based on the fact that you’re a smoker, one may guess at your health status. It is certainly the case that he knows more about you after the study than was known before (this is also the reason behind saying it is “general information”, not “public information”), but was your information leaked? Differential privacy will take the view that it was not, with the rationale that the impact on the smoker is the same independent of whether he was present or absent in the study. It is the conclusions reached in the study that affect the smoker, not his presence or absence in the data set.

The algorithmic foundations of differential p

How does Differential Privacy work?



Differential Privacy (DP) in action: ① Analyst sends a query to an intermediate piece of software, the DP guard. ② The guard assesses the privacy impact of the query using a special algorithm. ③ The guard sends the query to the database, and gets back a clean answer based on data that has not been distorted in any way. ④ The guard then adds the appropriate amount of "noise," scaled to the privacy impact, thus making the answer (hopefully slightly) imprecise in order to protect the confidentiality of the individuals whose information is in the database, and sends the modified response back to the analyst.

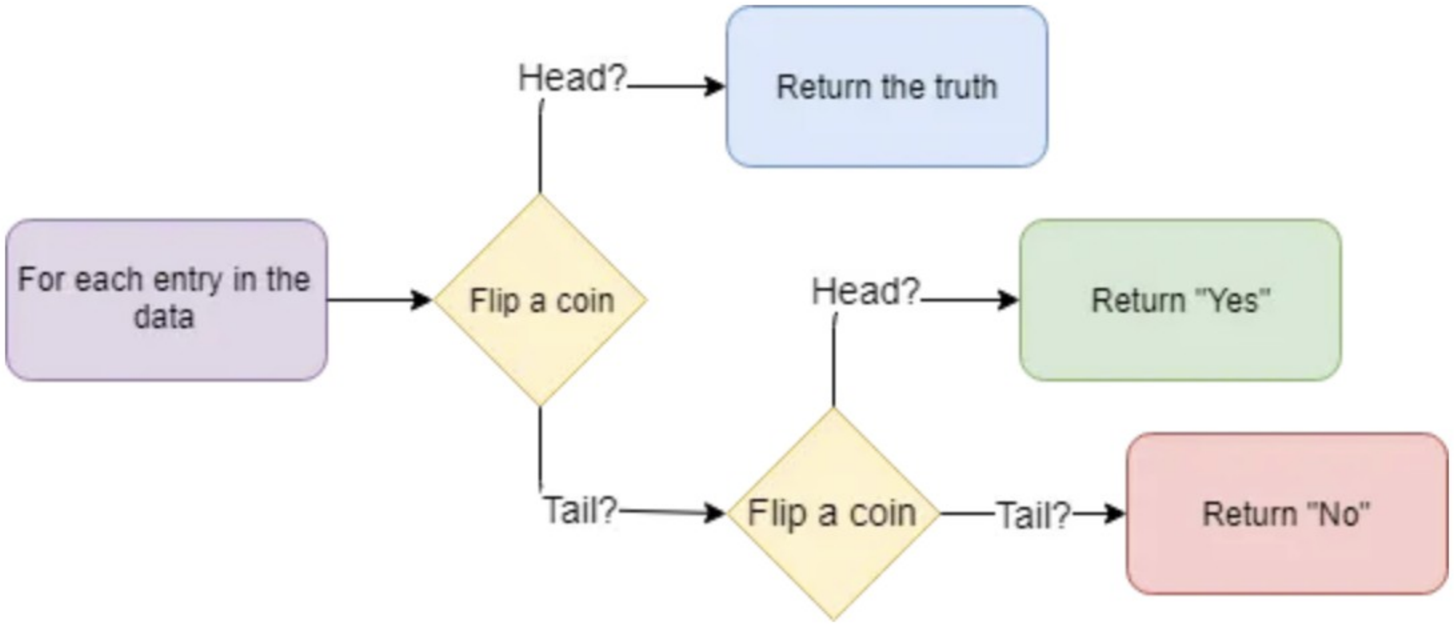
Differential Privacy for Everyone

Common Mechanisms

- Random Responses & Perturbations
- Laplace Mechanism
- Composition of Differentially Private methods

Random Responses & Perturbations

Ask individuals to respond to a "yes" or "no" question in a randomized manner, with a certain probability of giving a truthful answer and a certain probability of giving a random response

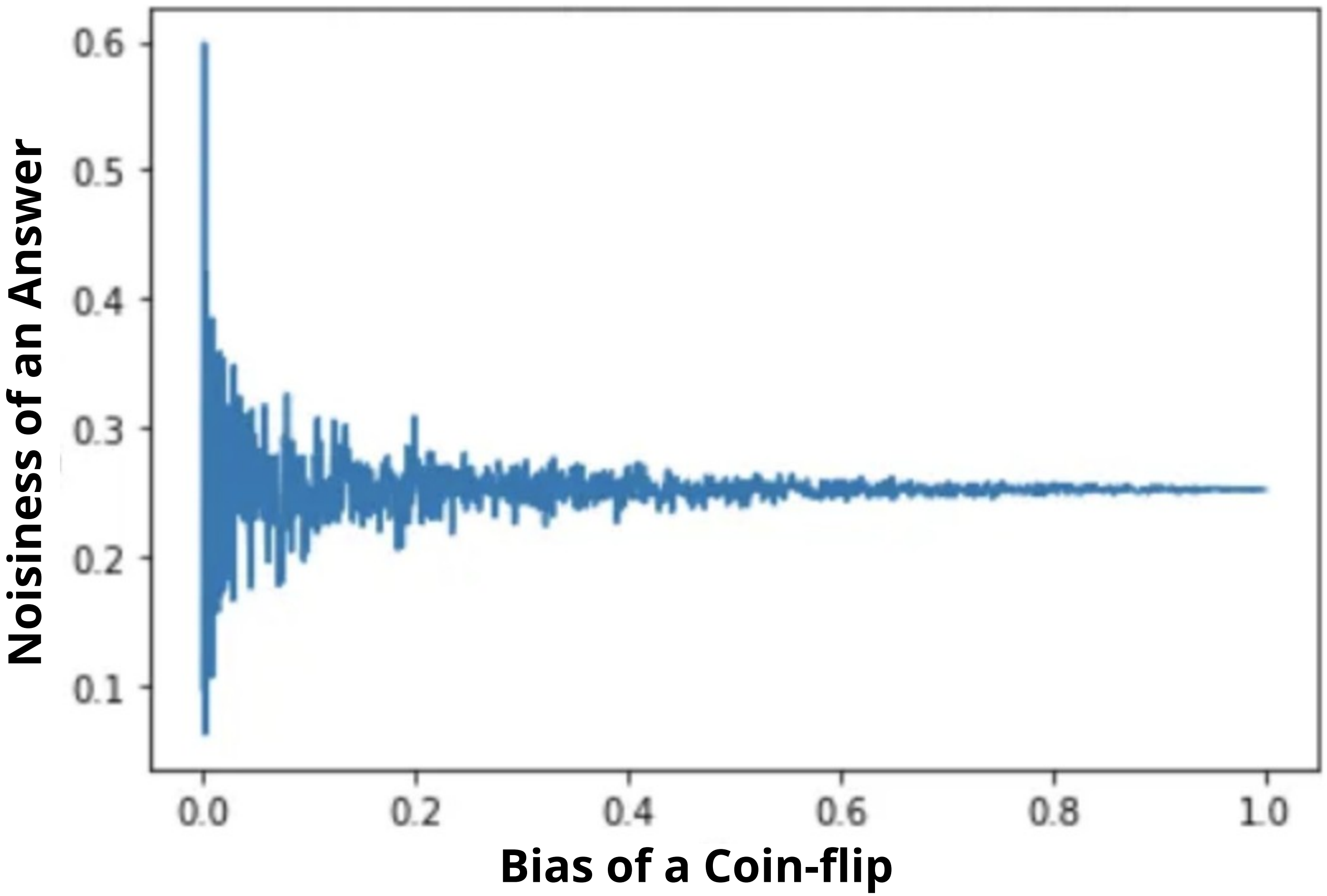


Random Responses & Perturbations

- Useful for research / surveys, especially on datasets containing controversial behavior or taboo topics
- Gives individuals the protection of ‘plausible deniability’ to protect their private information
- Provides valuable general information to researchers about the surveyed population to make inferences about the controversial or taboo topics.
- The curator can adjust the probability of giving a truthful answer (bias of the coin-flip) to drive a balance between privacy loss and utility of the data

Random Responses & Perturbations

Privacy vs. Utility



Where is this used?

- The 2020 census questionnaire asked questions about race, Hispanic background, sex, age, household relationships and whether a home is owned or rented during the head count of very U.S. resident.
- Tech companies use this to study their customers’ preferences without compromising user privacy by accessing individual behavior.

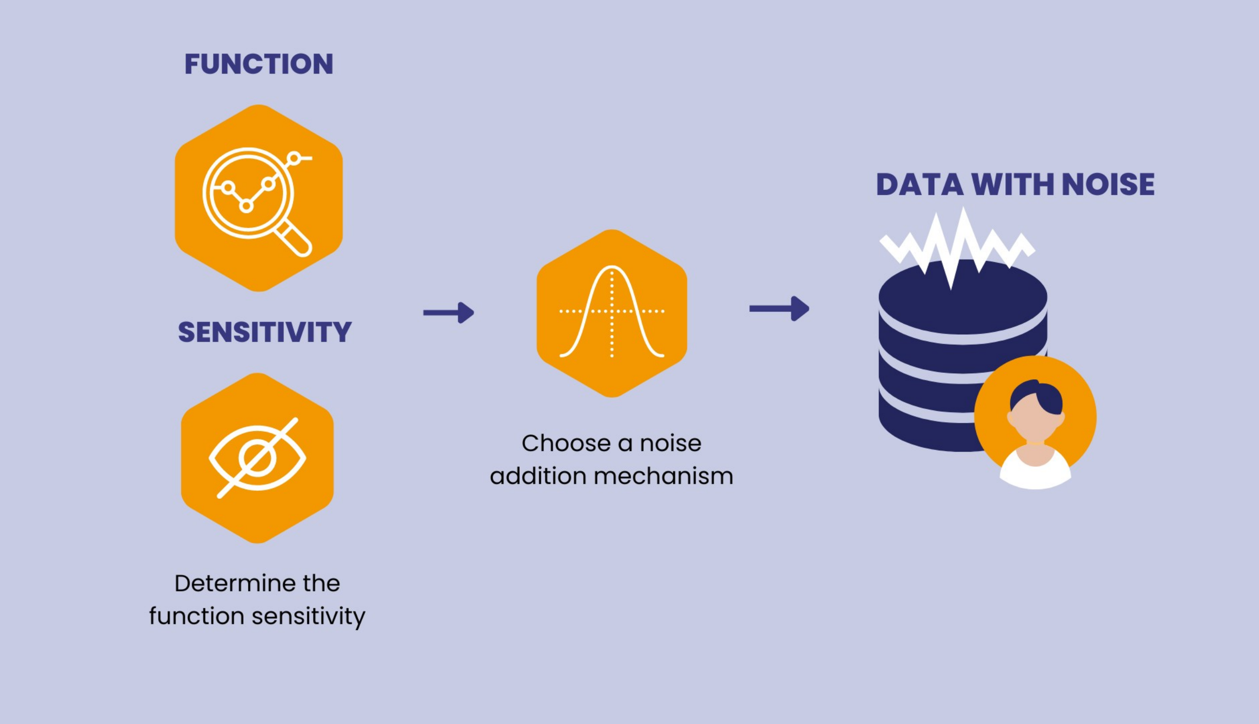
But…

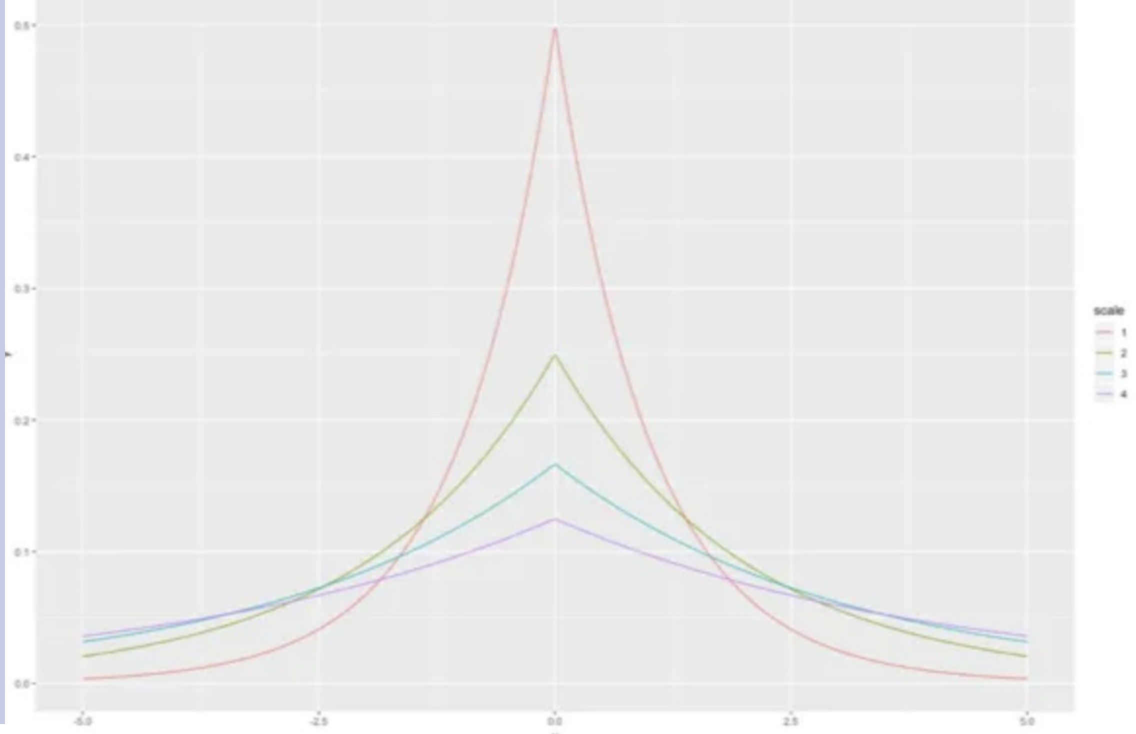
…Prominent researchers and demographers are pushing back against the US Census Bureau’s intent to use DP on the American Community Survey that samples 3.5 million homes and covers almost four dozen topics. They argue this will compromise the utility of the statistics

https://apnews.com
/article/census-2020-us-bureau-government-and-politics-20e683c71

Laplace Mechanism

Draw noise from the Laplace Distribution and add it to the function’s output


FUNCTION
SENSITIVITY
Determine the function sensitivity
Choose a noise addition mechanism
DATA WITH NOISE



Laplace Mechanism

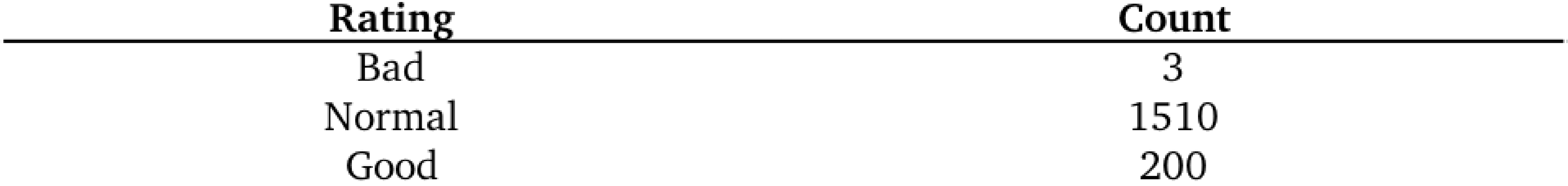
- We can mathematically prove that adding noise drawn from 0-centered Laplace Distribution with an established upper bound on privacy loss.
- The amount of noise we add depends on the sensitivity of a function, i.e. the amount of change in the function’s output when two datasets differ by a single record.

To limit privacy loss to ‘ϵ’, for a function with sensitivity ‘S’, we draw noise from a Laplace Distribution with variance,

$$2 * \left(\frac{S}{\epsilon}\right)^2$$

Laplace Mechanism

Dataset of Credit Ratings of Users

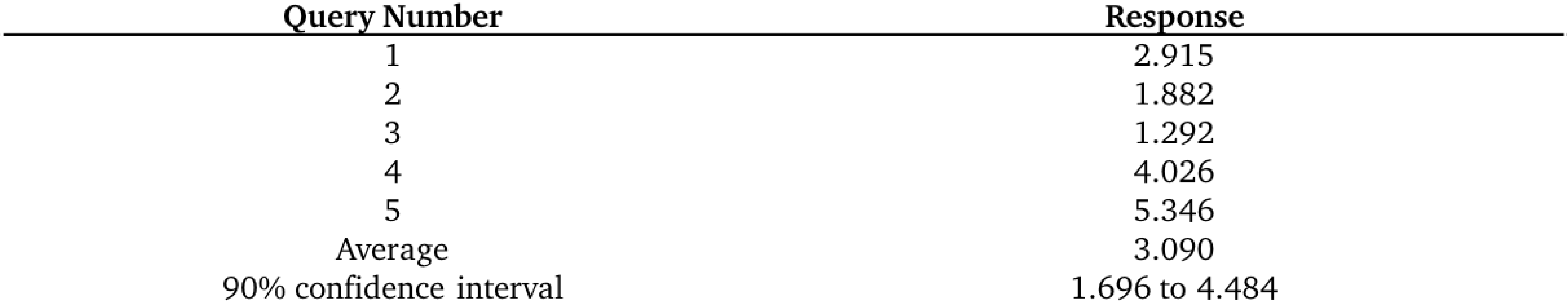


Rating	Count
Bad	3
Normal	1510
Good	200

What if I want to query for the No. of users with ‘Bad’ credit?

Laplace Mechanism

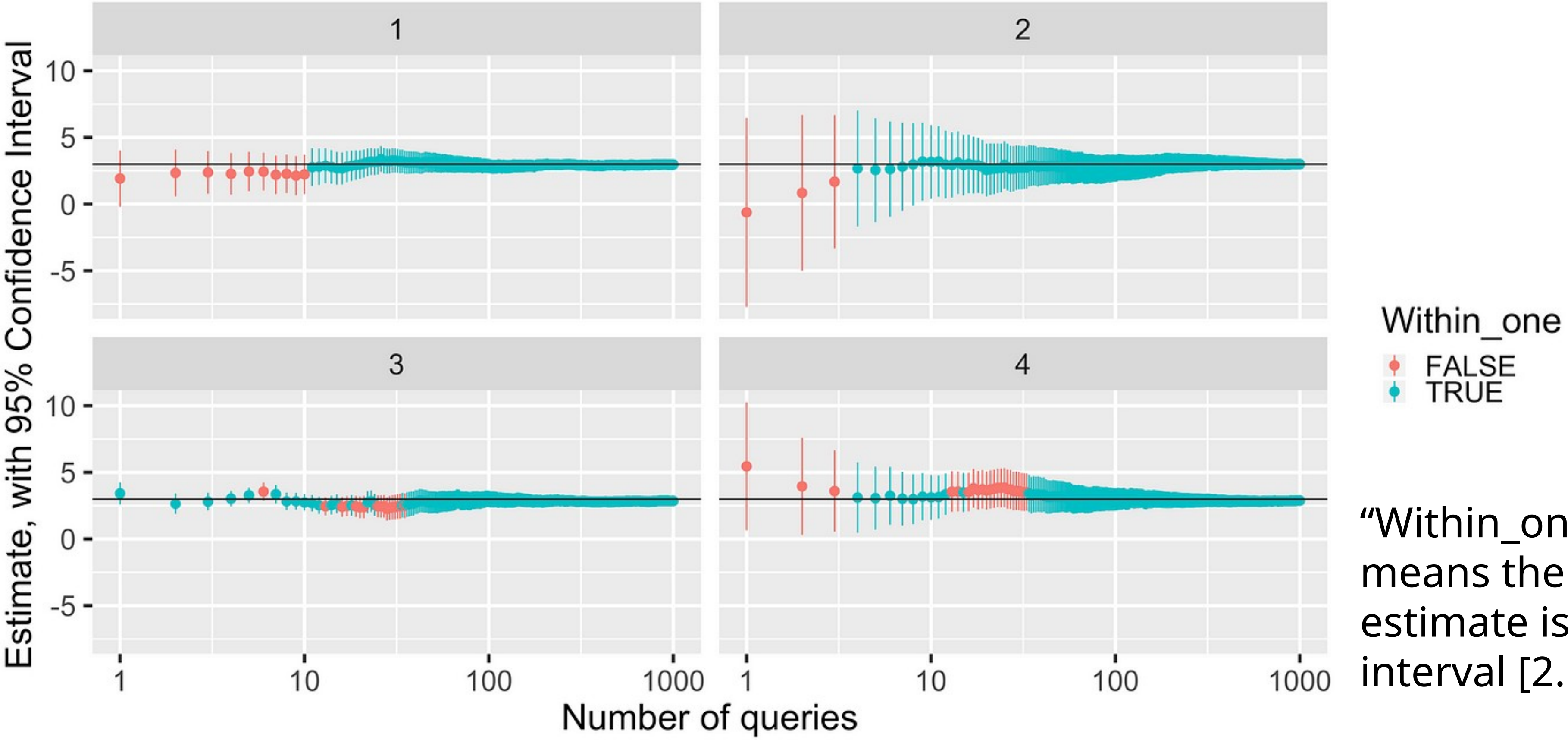
Add noise from a Laplace Distribution within 2 Standard Deviations



Query Number	Response
1	2.915
2	1.882
3	1.292
4	4.026
5	5.346
Average	3.090
90% confidence interval	1.696 to 4.484

But.. Averaging responses to multiple queries gets us really close to the true value

Laplace Mechanism



"Within_one" means the estimate is in the interval [2.5, 3.5]

"Estimation from repeated queries" is one of the fundamental limitations of differential privacy

Privacy Budget

- Privacy Losses accumulate with additional queries, as a consequence of the Composition Property
- For strong privacy guarantee, we need to limit the number of queries to limit privacy loss
- Privacy Budget is the maximum allowable privacy loss that curators can impose on researchers / potential adversaries.

Where is all of this used?

- The 2020 census to answer population statistics by location blocks.
- Apple uses this for QuickType suggestions, Emoji suggestions, finding energy draining domains, etc. More info here.
- Google Privacy Sandbox extensively uses noisy counting for reporting clicks and conversions for advertisers. It limits privacy loss by implementing a privacy budget.
- Uber uses Differential Privacy techniques to protect user privacy while allowing engineers to improve their systems.
- Microsoft uses DP extensively to study workplace analytics, market trends, online advertising and web search, etc. and has open-sourced tools for others to implement DP.