CS459/698
Privacy, Cryptography, Network and Data Security

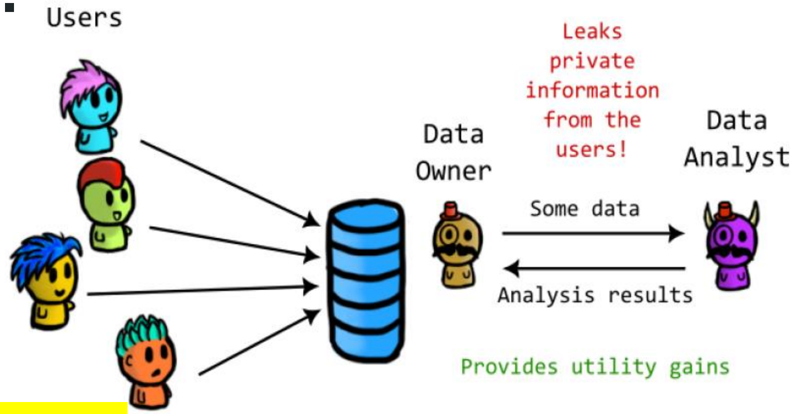
Statistical approaches to de-identification

Fall 2024, Tuesday/Thursday 02:30pm-03:50pm

Syntactic Notions of Privacy

Moving towards Defenses

- We saw many attacks.
- Now, we're going to see some defenses.
- How do we measure privacy?
    - **Empirically**:
        - by measuring the performance of an attack
    - **Theoretically**:
        - **Syntactic** notions: measuring a property on the released data / leakage.
        - **Semantic** notions: ensuring the data release mechanism itself has a property (independent of its inputs/outputs)



Syntactic Notions of Privacy

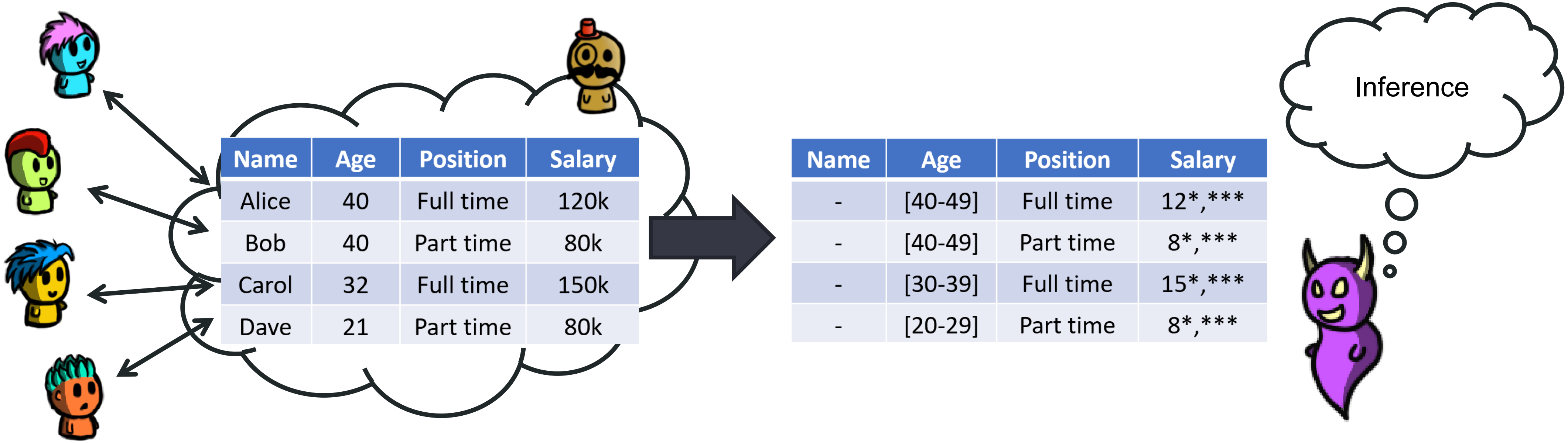
- Syntactic notions of privacy ensure that the released data satisfies a certain property.
- The data to be protected is typically a table, and the set of attributes can be classified into:
    - Identifiers: uniquely identify a participant
    - **Quasi-identifiers**: Indirect identifiers that can lead to identification when combined with other QI in the dataset or external information. These are often demographic variables (ZIP, DOB, Gender, etc.), but could also be timestamps, physical characteristics etc.
    - **Confidential attributes**: attributes (columns) that contains privacy-sensitive information.
    - Non-confidential attributes: are not considered sensitive
- We are going to see three syntactic notions of privacy:
    - k-anonymity
    - l-diversity
    - t-closeness

Syntactic Notions of Privacy

- We are going to see three syntactic notions of privacy:
  - k-anonymity
  - l-diversity
  - t-closeness
- For each syntactic notion of privacy, you will learn (and need to know):
  - What it **is**
  - Why it provides **privacy**
  - How to **compute** it
  - How to **provide** it(e.g., by publishing data in a privacy-preserving way by following certain – given – utility rules)

System Model

- Each user contributes to a row in a database
- A data collector releases a sanitized version of the database
- The adversary/analyst sees the sanitized database

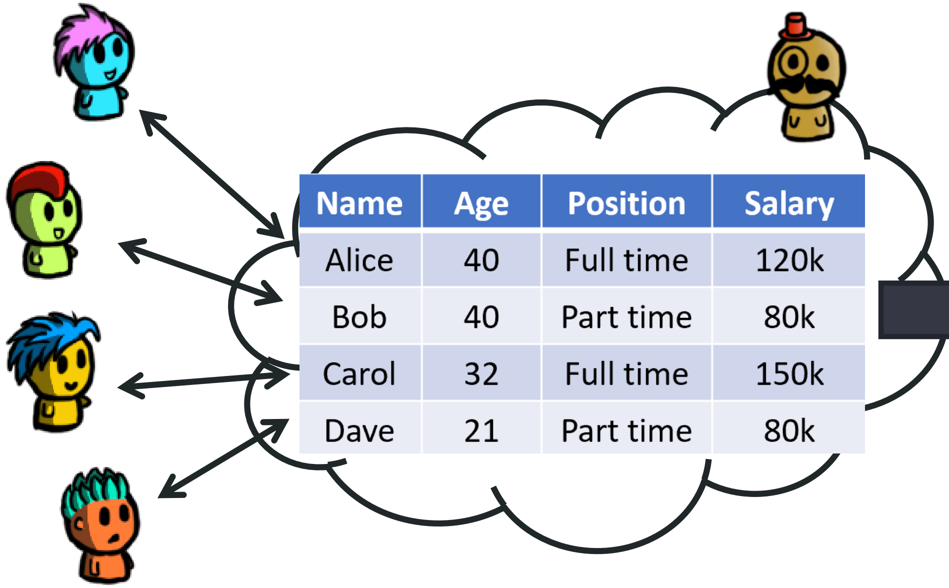


Name	Age	Position	Salary
Alice	40	Full time	120k
Bob	40	Part time	80k
Carol	32	Full time	150k
Dave	21	Part time	80k

Name	Age	Position	Salary
-	[40-49]	Full time	12*,***
-	[40-49]	Part time	8*,***
-	[30-39]	Full time	15*,***
-	[20-29]	Part time	8*,***

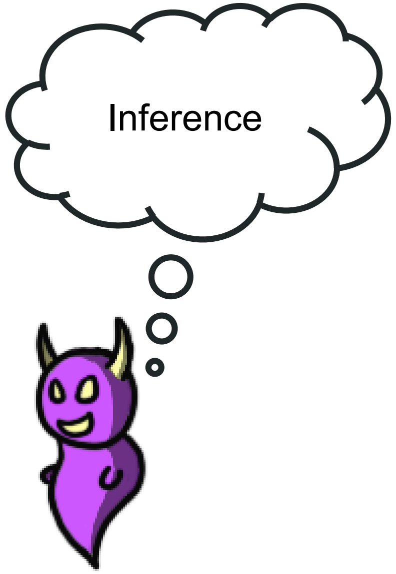
System Model

Q: What are the properties the sanitized database should have to preserve some level of privacy to its users?



Name	Age	Position	Salary
Alice	40	Full time	120k
Bob	40	Part time	80k
Carol	32	Full time	150k
Dave	21	Part time	80k

Name	Age	Position	Salary
-	[40-49]	Full time	12*,***
-	[40-49]	Part time	8*,***
-	[30-39]	Full time	15*,***
-	[20-29]	Part time	8*,***



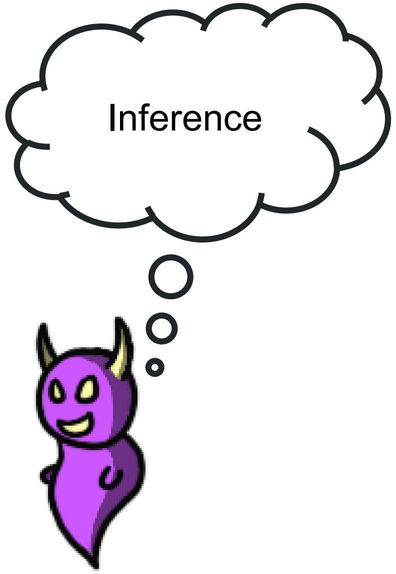
System Model

Q: What are the properties the sanitized database should have to preserve some level of privacy to its users?

A:
- k-anonymity
- ℓ-diversity
- t-closeness

Name	Age	Position	Salary
Alice	40	Full time	120k
Bob	40	Part time	80k
Carol	32	Full time	150k
Dave	21	Part time	80k

Name	Age	Position	Salary
-	[40-49]	Full time	12*,***
-	[40-49]	Part time	8*,***
-	[30-39]	Full time	15*,***
-	[20-29]	Part time	8*,***



k-anonymity

k-anonymity
For each published record, there exists at least $k - 1$ other records with the same quasi-identifiers

k-anonymity ensures that each individual in a dataset cannot be distinguished from at least k-1 other individuals with respect to the quasi-identifiers in the dataset.

This is done through generalization, suppression and sometimes top- and bottom- coding.

Applying k-anonymity:
- ✓ Makes it more difficult for an attacker to re-identify specific individuals in the dataset.
- ✓ It protects against singling out and, to some extent, the Mosaic effect.

k-anonymity

k-anonymity
For each published record, there exists at least $k-1$ other records with the same quasi-identifiers

- To **compute** k-anonymity:
    - Group the rows with the same quasi-identifier(s).
        - These rows form an *equivalence class* or equi-class.
    - **Count:** what is the smallest size of a group? That will be the level of k-anonymity
- To **provide** k-anonymity:
    - Remove a quasi-identifier(e.g., gender)
    - Reduce the granularity of a quasi-identifier (e.g., hiding the last characters of a ZIP code)
    - Group quasi-identifiers (e.g., report age ranges instead of actual ages)

k-anonymity: example

ZIP (QI)	Party affiliation
N1CFFA	Green Party
G0ANFA	Liberal Party
N1C5YN	Green Party
N2J0HJ	Conservative Party
N1C4KH	Green Party
G0A3G4	Conservative Party
G0A3GN	Liberal Party
N2JWBV	New Democratic Party
N2JWBV	Liberal Party

⇨

ZIP	Party affiliation
N1C***	Green Party
G0A***	Liberal Party
N1C***	Green Party
N2J***	Conservative Party
N1C***	Green Party
G0A***	Conservative Party
G0A***	Liberal Party
N2J***	New Democratic Party
N2J***	Liberal Party

Q: what is the k-anonymity level?

k-anonymity: example

ZIP (QI)	**Party affiliation**
N1CFFA	Green Party
G0ANFA	Liberal Party
N1C5YN	Green Party
N2J0HJ	Conservative Party
N1C4KH	Green Party
G0A3G4	Conservative Party
G0A3GN	Liberal Party
N2JWBV	New Democratic Party
N2JWBV	Liberal Party

ZIP	**Party affiliation**
N1C***	Green Party
G0A***	Liberal Party
N1C***	Green Party
N2J***	Conservative Party
N1C***	Green Party
G0A***	Conservative Party
G0A***	Liberal Party
N2J***	New Democratic Party
N2J***	Liberal Party

Q: what is the k-anonymity level?

A: the table is 3-anonymous

k-anonymity: example (II)

ZIP (QI)	DOB (QI)	Party affiliation
N1CFF	1962-01-24	Green Party
G0ANF	1975-12-30	Liberal Party
N1C5YN	1966-10-17	Green Party
N2J0HJ	1996-08-14	Conservative Party
N1C4KH	1963-04-06	Green Party
G0A3G4	1977-07-09	Conservative Party
G0A3GN	1973-08-14	Liberal Party
N2JWBV	1990-11-02	New Democratic Party
N2JWBV	1990-01-25	Liberal Party

⇨

ZIP	DOB	Party affiliation
N1C***	196*-**-**	Green Party
G0A***	197*-**-**	Liberal Party
N1C***	196*-**-**	Green Party
N2J***	199*-**-**	Conservative Party
N1C***	196*-**-**	Green Party
G0A***	197*-**-**	Conservative Party
G0A***	197*-**-**	Liberal Party
N2J***	199*-**-**	New Democratic Party
N2J***	199*-**-**	Liberal Party

Q: what is the k-anonymity level?

k-anonymity: example (II)

ZIP (QI)	DOB (QI)	Party affiliation
N1CFF	1962-01-24	Green Party
G0ANF	1975-12-30	Liberal Party
N1C5YN	1966-10-17	Green Party
N2J0HJ	1996-08-14	Conservative Party
N1C4KH	1963-04-06	Green Party
G0A3G4	1977-07-09	Conservative Party
G0A3GN	1973-08-14	Liberal Party
N2JWBV	1990-11-02	New Democratic Party
N2JWBV	1990-01-25	Liberal Party

⇨

ZIP	DOB	Party affiliation
N1C***	196*-**-**	Green Party
G0A***	197*-**-**	Liberal Party
N1C***	196*-**-**	Green Party
N2J***	199*-**-**	Conservative Party
N1C***	196*-**-**	Green Party
G0A***	197*-**-**	Conservative Party
G0A***	197*-**-**	Liberal Party
N2J***	199*-**-**	New Democratic Party
N2J***	199*-**-**	Liberal Party

Q: what is the k-anonymity level?

A: the table is 3-anonymous

k-anonymity: practice

- Both age and gender are QI.

Age	Gender	...
23	F	
25	F	
33	F	
35	F	
27	M	
30	M	
32	M	
21	NB	
25	NB	

Q: What is the k-anonymity if…

1. We hide the Age
2. We hide the Gender (but not the age)
3. We report the most significant digit of Age, plus the Gender
4. We only report the most significant digit of Age, but not the Gender

k-anonymity: practice

- Both age and gender are QI.

Age	Gender	...
23	F	
25	F	
33	F	
35	F	
27	M	
30	M	
32	M	
21	NB	
25	NB	

Q: What is the k-anonymity if…

1. We hide the Age
2. We hide the Gender (but not the age)
3. We report the most significant digit of Age, plus the Gender
4. We only report the most significant digit of Age, but not the Gender

A: 2, 1, 1, 4

k-anonymity: practice (II)

- Both age and DOB are QI.

Gender	DOB	Party affiliation
M	1968-**-**	Green Party
F	1975-**-**	Liberal Party
O	1966-**-**	Green Party
M	1962-**-**	Green Party
M	1962-**-**	Conservative Party
O	1966-**-**	Conservative Party
F	1973-**-**	Liberal Party
F	1973-**-**	Liberal Party
O	1968-**-**	Green Party
F	1975-**-**	Liberal Party

Q: What is the k-anonymity if…

1. We publish the table as shown
2. We hide the least-significant digit of year
3. We hide the Gender column
4. We hide the least-significant digit of year and hide the Gender column

k-anonymity: practice (II)

- Both age and DOB are QI.

Gender	DOB	Party affiliation
M	1968-**-**	Green Party
F	1975-**-**	Liberal Party
O	1966-**-**	Green Party
M	1962-**-**	Green Party
M	1962-**-**	Conservative Party
O	1966-**-**	Conservative Party
F	1973-**-**	Liberal Party
F	1973-**-**	Liberal Party
O	1968-**-**	Green Party
F	1975-**-**	Liberal Party

Q: What is the k-anonymity if…

1. We publish the table as shown
2. We hide the least-significant digit of year
3. We hide the Gender column
4. We hide the least-significant digit of year and hide the Gender column

A: 1, 3, 2, 4

k-anonymity and privacy

ZIP (QI)	DOB (QI)	Party affiliation
N1C***	196*-**-**	Green Party
N1C***	196*-**-**	Green Party
N1C***	196*-**-**	Green Party
G0A***	197*-**-**	Liberal Party
G0A***	197*-**-**	Liberal Party
G0A***	197*-**-**	Conservative Party
N2J***	199*-**-**	Conservative Party
N2J***	199*-**-**	New Democratic Party
N2J***	199*-**-**	Liberal Party

- This table is 3-anonymous.

Q: This provides some resistance against linking attacks, why?

k-anonymity and privacy

ZIP (QI)	DOB (QI)	Party affiliation
N1C***	196*-**-**	Green Party
N1C***	196*-**-**	Green Party
N1C***	196*-**-**	Green Party
G0A***	197*-**-**	Liberal Party
G0A***	197*-**-**	Liberal Party
G0A***	197*-**-**	Conservative Party
N2J***	199*-**-**	Conservative Party
N2J***	199*-**-**	New Democratic Party
N2J***	199*-**-**	Liberal Party

- This table is 3-anonymous.

Q: This provides some resistance against linking attacks, why?

A: We cannot identify the actual record of a user (that provided a record) based on the QI. This makes it hard to guess the user's confidential attributes.

k-anonymity and privacy

ZIP (QI)	DOB (QI)	Party affiliation
N1C***	196*-**-**	Green Party
N1C***	196*-**-**	Green Party
N1C***	196*-**-**	Green Party
G0A***	197*-**-**	Liberal Party
G0A***	197*-**-**	Liberal Party
G0A***	197*-**-**	Conservative Party
N2J***	199*-**-**	Conservative Party
N2J***	199*-**-**	New Democratic Party
N2J***	199*-**-**	Liberal Party

- This table is 3-anonymous.

Q: Is k-anonymity enough? Can you see any issues with it?

k-anonymity and privacy

ZIP (QI)	DOB (QI)	Party affiliation
N1C***	196*-**-**	Green Party
N1C***	196*-**-**	Green Party
N1C***	196*-**-**	Green Party
G0A***	197*-**-**	Liberal Party
G0A***	197*-**-**	Liberal Party
G0A***	197*-**-**	Conservative Party
N2J***	199*-**-**	Conservative Party
N2J***	199*-**-**	New Democratic Party
N2J***	199*-**-**	Liberal Party

- This table is 3-anonymous.

Q: Is k-anonymity enough? Can you see any issues with it?

Attack 1: if you know Alice has ZIP code N1C***, what can you learn from her?

Attack 2: if you know Bob has ZIP code G0A*** and does not like Liberal Party, what can you learn from him?

Homogeneity attacks happen when sensitive values lack diversity. It filters out infeasible values and, in the worst case, narrows the inference down to a single value.

ℓ-diversity

ℓ-diversity
For each quasi-identifier value, there should be at least ℓ distinct values of the sensitive attributes

ℓ-diversity is an extension to k-anonymity that ensures that there is sufficient variation in a sensitive attribute.

This is important, because if all individuals in a (subset of a) dataset have the same value for the sensitive attribute, there is still a risk of inference.

ℓ-diversity

ℓ-diversity

For each quasi-identifier value, there should be at least ℓ distinct values of the sensitive attributes

- To **compute** ℓ-diversity:
  - Group the rows by quasi-identifiers into equi-classes.
  - For each equi-class, compute how many distinct sensitive values there are
  - The equi-class with the smallest number of distinct sensitive values is the level of ℓ-diversity.
- To **provide** ℓ-diversity:
  - Similar to k-anonymity:
  - ➢ Try to make the equi-classes as large as possible, while making sure there is enough variety of sensitive attributes per class.

ℓ-diversity: example

Gender	DOB	Party affiliation
M	196*-**-**	Green Party
M	196*-**-**	Liberal Party
M	196*-**-**	Conservative Party
O	196*-**-**	Green Party
O	196*-**-**	Green Party
O	196*-**-**	Conservative Party
F	197*-**-**	Liberal Party
F	197*-**-**	Green Party
F	197*-**-**	Conservative Party
F	197*-**-**	Liberal Party

- Gender and DOB are QI, Party affiliation is the sensitive attribute.
- The table is 3-Anonymous

Q: what is the level of ℓ-diversity?

ℓ-diversity: example

Gender	DOB	Party affiliation
M	196*-**-**	Green Party
M	196*-**-**	Liberal Party
M	196*-**-**	Conservative Party
O	196*-**-**	Green Party
O	196*-**-**	Green Party
O	196*-**-**	Conservative Party
F	197*-**-**	Liberal Party
F	197*-**-**	Green Party
F	197*-**-**	Conservative Party
F	197*-**-**	Liberal Party

- Gender and DOB are QI, Party affiliation is the sensitive attribute.
- The table is 3-Anonymous

Q: what is the level of ℓ-diversity?

A: the table is 2-diversified

ℓ-diversity and privacy

ZIP	DOB	Salary
N3P***	199*-**-**	20K
N3P***	199*-**-**	15K
N3P***	199*-**-**	25K
H1A***	196*-**-**	100K
H1A***	196*-**-**	90K
H1A***	196*-**-**	120K
S4N***	197*-**-**	50K
S4N***	197*-**-**	60K
S4N***	197*-**-**	65K

Q: what is the level of k-anonymity and ℓ-diversity?

ℓ-diversity and privacy

ZIP	DOB	Salary
N3P***	199*-**-**	20K
N3P***	199*-**-**	15K
N3P***	199*-**-**	25K
H1A***	196*-**-**	100K
H1A***	196*-**-**	90K
H1A***	196*-**-**	120K
S4N***	197*-**-**	50K
S4N***	197*-**-**	60K
S4N***	197*-**-**	65K

Q: what is the level of k-anonymity and ℓ-diversity?

A: 3 and 3

Q: why does this provide privacy?

ℓ-diversity and privacy

ZIP	DOB	Salary
N3P***	199*-**-**	20K
N3P***	199*-**-**	15K
N3P***	199*-**-**	25K
H1A***	196*-**-**	100K
H1A***	196*-**-**	90K
H1A***	196*-**-**	120K
S4N***	197*-**-**	50K
S4N***	197*-**-**	60K
S4N***	197*-**-**	65K

Q: what is the level of k-anonymity and ℓ-diversity?

A: 3 and 3

Q: why does this provide privacy?

A: it alleviates the problem of k-anonymity that we saw above when all values are the same.

ℓ-diversity and privacy

ZIP	DOB	Salary	Disease
N3P***	199*-**-**	20K	gastric ulcer
N3P***	199*-**-**	15K	gastritis
N3P***	199*-**-**	25K	stomach cancer
H1A***	196*-**-**	100K	heart attack
H1A***	196*-**-**	90K	flu
H1A***	196*-**-**	120K	bronchitis
S4N***	197*-**-**	50K	COVID
S4N***	197*-**-**	60K	kidney stone
S4N***	197*-**-**	65K	pneumonia

Q: if you know Charles, who earns a low salary, is in this table: what else can you learn?

ℓ-diversity and privacy

ZIP	DOB	Salary	Disease
N3P***	199*-**-**	20K	gastric ulcer
N3P***	199*-**-**	15K	gastritis
N3P***	199*-**-**	25K	stomach cancer
H1A***	196*-**-**	100K	heart attack
H1A***	196*-**-**	90K	flu
H1A***	196*-**-**	120K	bronchitis
S4N***	197*-**-**	50K	COVID
S4N***	197*-**-**	60K	kidney stone
S4N***	197*-**-**	65K	pneumonia

Q: if you know Charles, who earns a low salary, is in this table: what else can you learn?

A: Charles has a stomach disease

Similarity Attack: If the sensitive values of an equi-class are different but have the same (or similar) semantic meaning, l-diversity does not prevent the adversary from learning this.

ℓ-diversity and privacy

ZIP	DOB	Virus X Test
N3P***	199*-**-**	Positive
N3P***	199*-**-**	Positive
N3P***	199*-**-**	Positive
... 45 more positive cases ...		
N3P***	199*-**-**	Negative
H1A***	196*-**-**	Negative
H1A***	196*-**-**	Negative
H1A***	196*-**-**	Negative
... 945 more negative cases ...		
H1A***	196*-**-**	Positive

Q: if you know David, who is in his 20s, is in this table: what else did you learn?

ℓ-diversity and privacy

ZIP	DOB	Virus X Test
N3P***	199*-**-**	Positive
N3P***	199*-**-**	Positive
N3P***	199*-**-**	Positive
… 45 more positive cases …		
N3P***	199*-**-**	Negative
H1A***	196*-**-**	Negative
H1A***	196*-**-**	Negative
H1A***	196*-**-**	Negative
… 945 more negative cases …		
H1A***	196*-**-**	Positive

Q: if you know David, who is in his 20s, is in this table: what else did you learn?

A: David probably has the virus

Skewness Attack: If the The distribution of sensitive values matters. Highly-skewed distributions leak (statistically speaking) more information about an individual's sensitive value.

What went wrong?

ZIP	DOB	Virus X Test
N3P***	199*-**-**	Positive
N3P***	199*-**-**	Positive
N3P***	199*-**-**	Positive
… 45 more positive cases …		
N3P***	199*-**-**	Negative
H1A***	196*-**-**	Negative
H1A***	196*-**-**	Negative
H1A***	196*-**-**	Negative
… 945 more negative cases …		
H1A***	196*-**-**	Positive

- The data in each equi-class (i.e., records that share the same quasi-identifier) is unexpectedly skewed.
- This means that learning the equi-class of a person can leak a lot of statistical information about the sensitive attributes of that person.

t-closeness

t-closeness
The distribution of sensitive values in each equi-class is no further than a threshold t from the overall distribution of the sensitive values in the whole table

Equi-class: each set of identical quasi-identifiers is an equi-class.

t-closeness ensures that the distribution of a sensitive attribute within a generalisation of a quasi-identifier is close to the distribution of the sensitive attribute in the entire dataset.

Example:
A dataset contains information on Age (quasi-identifier), Sex (quasi-identifier), and Income (sensitive attribute), and t-closeness is applied with a value of t = 0.1, then for each combination of Age and Sex, the distribution of income must be within 10% of the distribution of income in the entire dataset.

t-closeness

t-closeness
The distribution of sensitive values in each equi-class is no further than a threshold t from the overall distribution of the sensitive values in the whole table

Equi-class: each set of identical quasi-identifiers is an equi-class.

- To **compute** t-closeness:
  - Organize rows by equi-class
  - Compute the distribution of sensitive attributes per equi-class and for the whole table.
  - Compute the maximum difference between a class distribution and the whole table's distribution on a sensitive value.
    → That's the value of t.
- To **provide** t-closeness:
  - Similar to k-anonymity: try to make the equi-classes as large as possible, while trying to maintain a uniform distribution.
  - Could add dummy records to help smooth the distribution.

t-closeness

t-closeness
The distribution of sensitive values in each equi-class is no further than a threshold t from the overall distribution of the sensitive values in the whole table

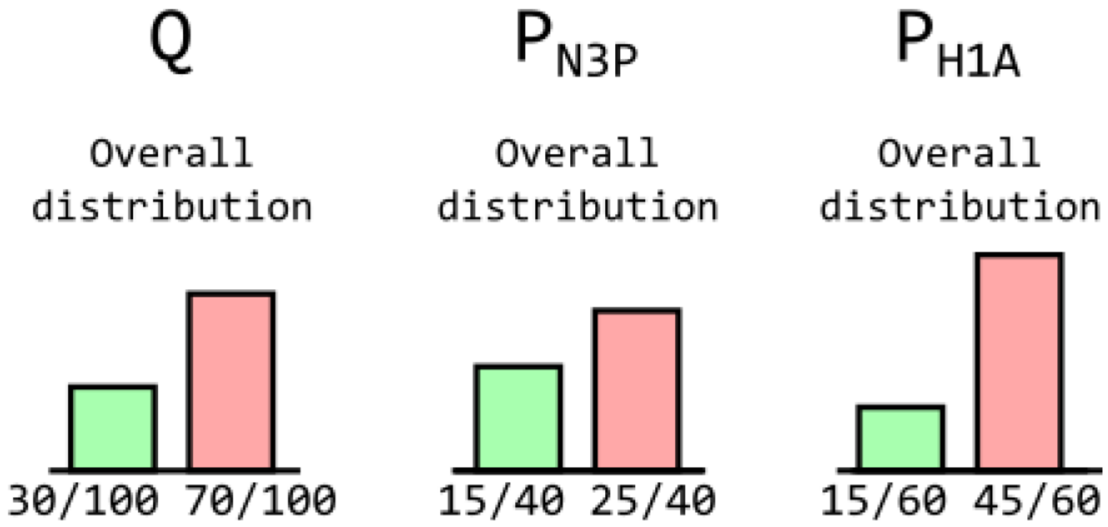
- To **compute** t-closeness we need to define a notion of distance between distributions. See the original paper that proposes t-closeness on ICDE'07
- We will only cover one distance:

Variational distance (or EMD Categorical Distance – using Equal Distance)
For two distributions over m values $P = (p_1, p_2, \dots, p_m)$ and $Q = (q_1, q_2, \dots, q_m)$:

$$D[P, Q] \doteq \frac{1}{2} \sum_{i=1}^{m} |p_i - q_i|$$

t-closeness example

ZIP (QI)	Virus (Sens)	
N3P***	Pos	x15
N3P***	Neg	x25
H1A***	Pos	x15
H1A***	Neg	x45



$$D[\mathbf{P}_{N3P}, \mathbf{Q}] = \frac{1}{2}\left(\left|\frac{15}{40} - \frac{30}{100}\right| + \left|\frac{25}{40} - \frac{70}{100}\right|\right) = 0.075$$

$$D[\mathbf{P}_{H1A}, \mathbf{Q}] = \frac{1}{2}\left(\left|\frac{15}{60} - \frac{30}{100}\right| + \left|\frac{45}{60} - \frac{70}{100}\right|\right) = 0.05$$

Variational distance:

$$D[P,Q] \doteq \frac{1}{2}\sum_{i=1}^{m} |p_i - q_i|$$

t-close with t=0.075 (the maximum of these values)

t-closeness example: more sensitive values

ZIP (QI)	Virus (Sens)	
N3P***	Pos	x5
N3P***	Neg	x22
N3P***	Inc	x3
H1A***	Pos	x12
H1A***	Neg	x47
H1A***	Inc	x1

Variational distance:

$$D[P,Q] \doteq \frac{1}{2}\sum_{i=1}^{m} |p_i - q_i|$$

Q: what is the k-anonymity, ℓ-diversity and t-closeness level of this published dataset?

t-closeness example: more sensitive values

ZIP (QI)	Virus (Sens)	
N3P***	Pos	x5
N3P***	Neg	x22
N3P***	Inc	x3
H1A***	Pos	x12
H1A***	Neg	x47
H1A***	Inc	x1

Variational distance:

$$D[P,Q] \doteq \frac{1}{2}\sum_{i=1}^{m} |p_i - q_i|$$

Q: what is the k-anonymity, ℓ-diversity and t-closeness level of this published dataset?

A: 30-anonymous and 3-diversified.

$$D[P_{N3P},Q] = \frac{1}{2}\left(\left|\frac{5}{30}-\frac{17}{90}\right| + \left|\frac{22}{30}-\frac{69}{90}\right| + \left|\frac{3}{30}-\frac{4}{90}\right|\right) = \frac{1}{18}$$

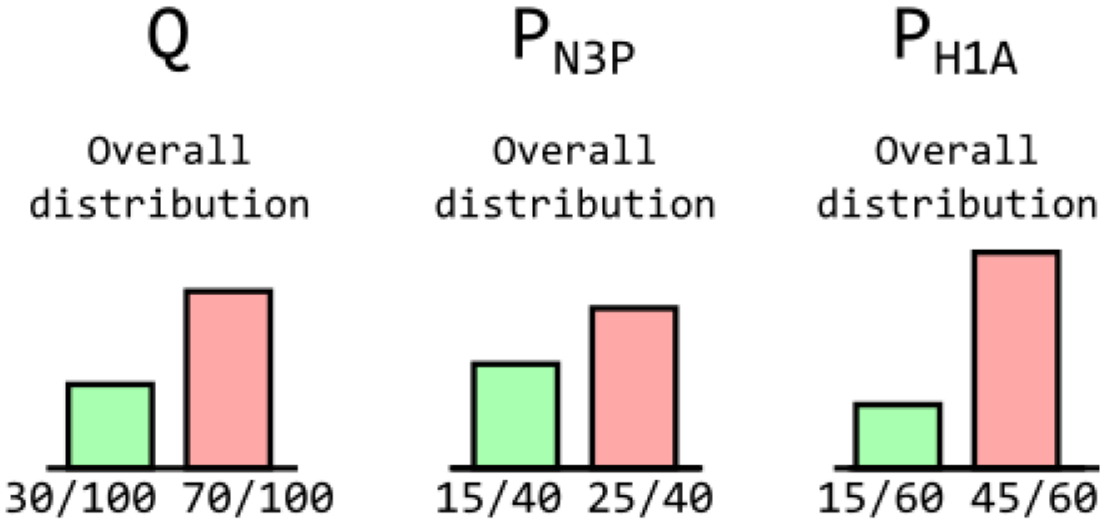
$$D[P_{H1A},Q] = \frac{1}{2}\left(\left|\frac{12}{60}-\frac{17}{90}\right| + \left|\frac{47}{60}-\frac{69}{90}\right| + \left|\frac{1}{60}-\frac{4}{90}\right|\right) = \frac{1}{36}$$

Therefore, the table is $\frac{1}{18}$ -close with respect to Virus

Notes on computing t-closeness

- If you have k equi-classes, you would have to compute k distances and take the maximum of those distances as the value of t.
- If you have m distinct sensitive values, the histograms would have m bars and you would have to add m absolute value terms to compute each distance.

ZIP (QI)	Virus (Sens)	
N3P***	Pos	x15
N3P***	Neg	x25
H1A***	Pos	x15
H1A***	Neg	x45



Notes on computing t-closeness

- If you have more than one sensitive attribute (column), you can compute the t-closeness for each sensitive attribute independently (e.g., a table can be t_1-close with respect to Salary and t_2-close with respect to Virus).
- Check the original paper by Li et al. for other distance metrics and more examples.

Limitations

- t-closeness is overall a reasonable syntactic notion of privacy. It prevents the attacks that we have seen. However:

1. These privacy notions require a clear distinction between quasi-identifiers and sensitive values, which is not always possible (and is subjective)
2. Expensive to compute:
   - Computing the optimal k-anonymous dataset is NP-hard
3. These notions of privacy do not provide guarantees against an adversary with (arbitrary) background knowledge

Limitations Example

Hospital A

	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 B

	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

Q: We know that Dave just had his 35th birthday! He told us on his way to the hospital A. What did we learn?

Q: We know a 28 year old visited hospitals A and B. What can we infer?

Source: Ganta et al. 2008 Composition attacks and auxiliary information in data privacy

Limitations Example

Hospital A

	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 B

	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

Q: We know that Dave just had his 35th birthday! He told us on his way to the hospital A. What did we learn?

A: Dave has Cancer

Q: We know a 28 year old visited hospitals A and B. What can we infer?

A: They likely have AIDS

Source: Ganta et al. 2008 Composition attacks and auxiliary information in data privacy

Issues with syntactic notions of privacy

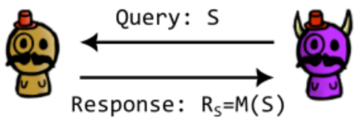
- Syntactic notions of privacy have some issues:
  - Defining which attributes are quasi-identifiers and which are sensitive attributes is hard
  - Mostly apply to relational databases; what about general data releases like machine learning?
  - What if the adversary has arbitrary auxiliary information?

- We need a privacy notion that is adversary-agnostic… a ***semantic*** notion of privacy, that only depends on the mechanism
  - But how do we achieve this?

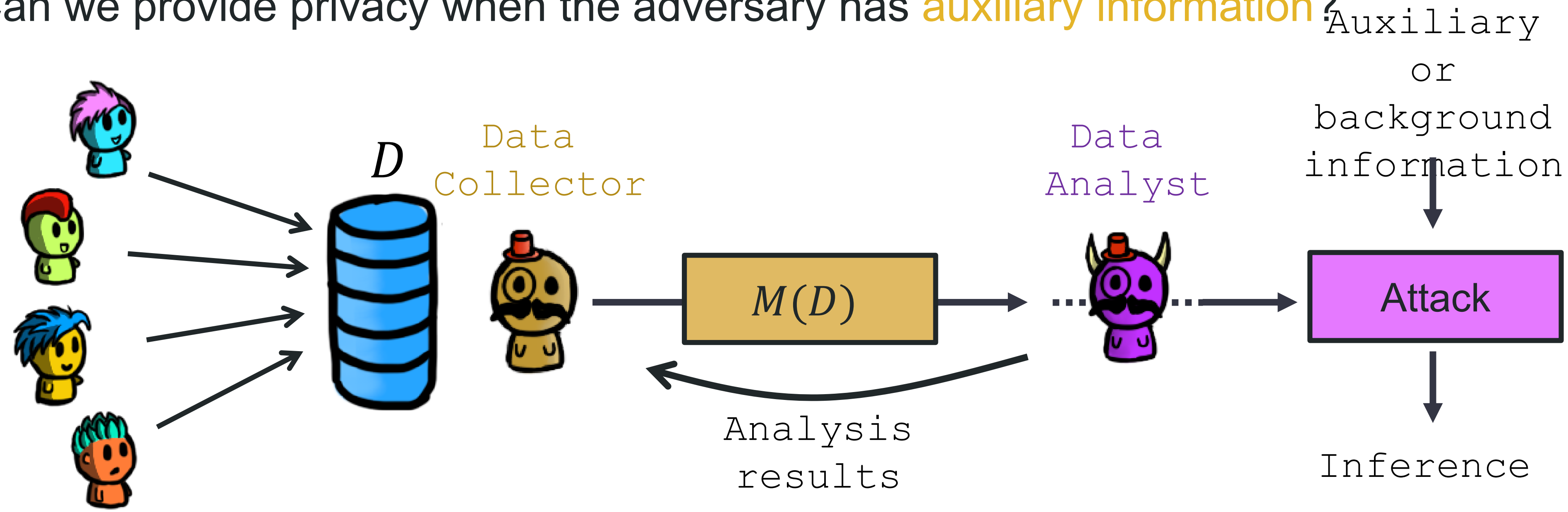
Introduction to Differential Privacy

Can we protect against auxiliary information?

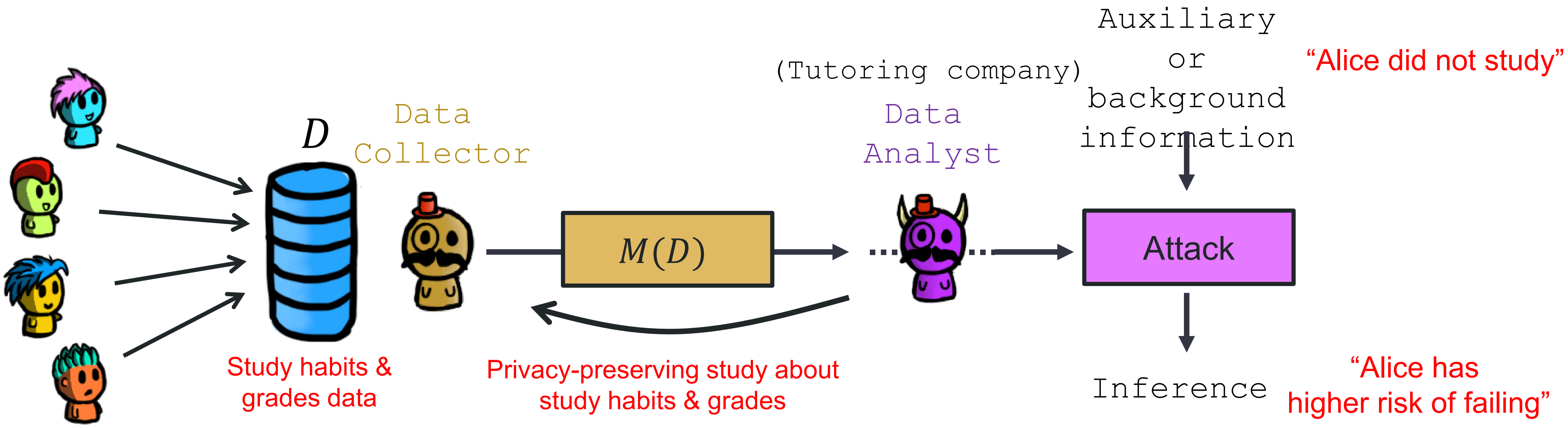
- Each user contributes to one entry (row) of a database D.
- The release mechanism M publishes some data $R = M(D)$.
  - Formally, M : S → R_S , where the Collector provides a response to query S with R_S . The analyst may be honest or malicious.



- Can we provide privacy when the adversary has auxiliary information?

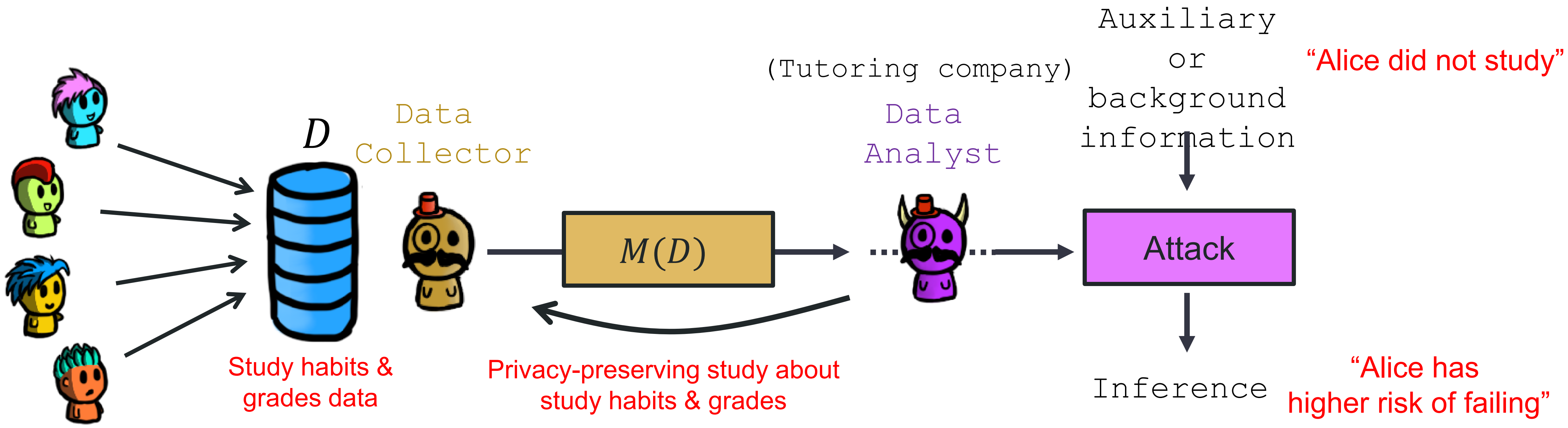


Example: strong auxiliary information



Q: Is this a violation of Alice's privacy ? Is this the study's fault?
Should we design a mechanism M to prevent this?

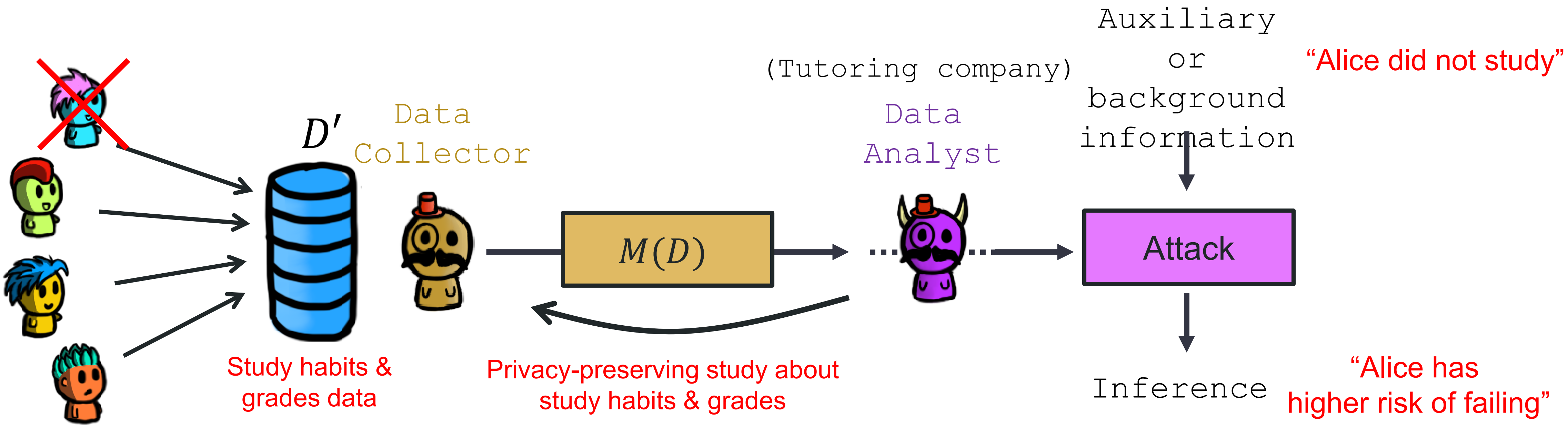
Example: strong auxiliary information



Q: Is this a violation of Alice's privacy ? Is this the study's fault?
Should we design a mechanism M to prevent this?

A: The adversary would've reached the same conclusion even if Alice hadn't participated in the study! We cannot prevent this unless we destroy utility (e.g., not doing the study)

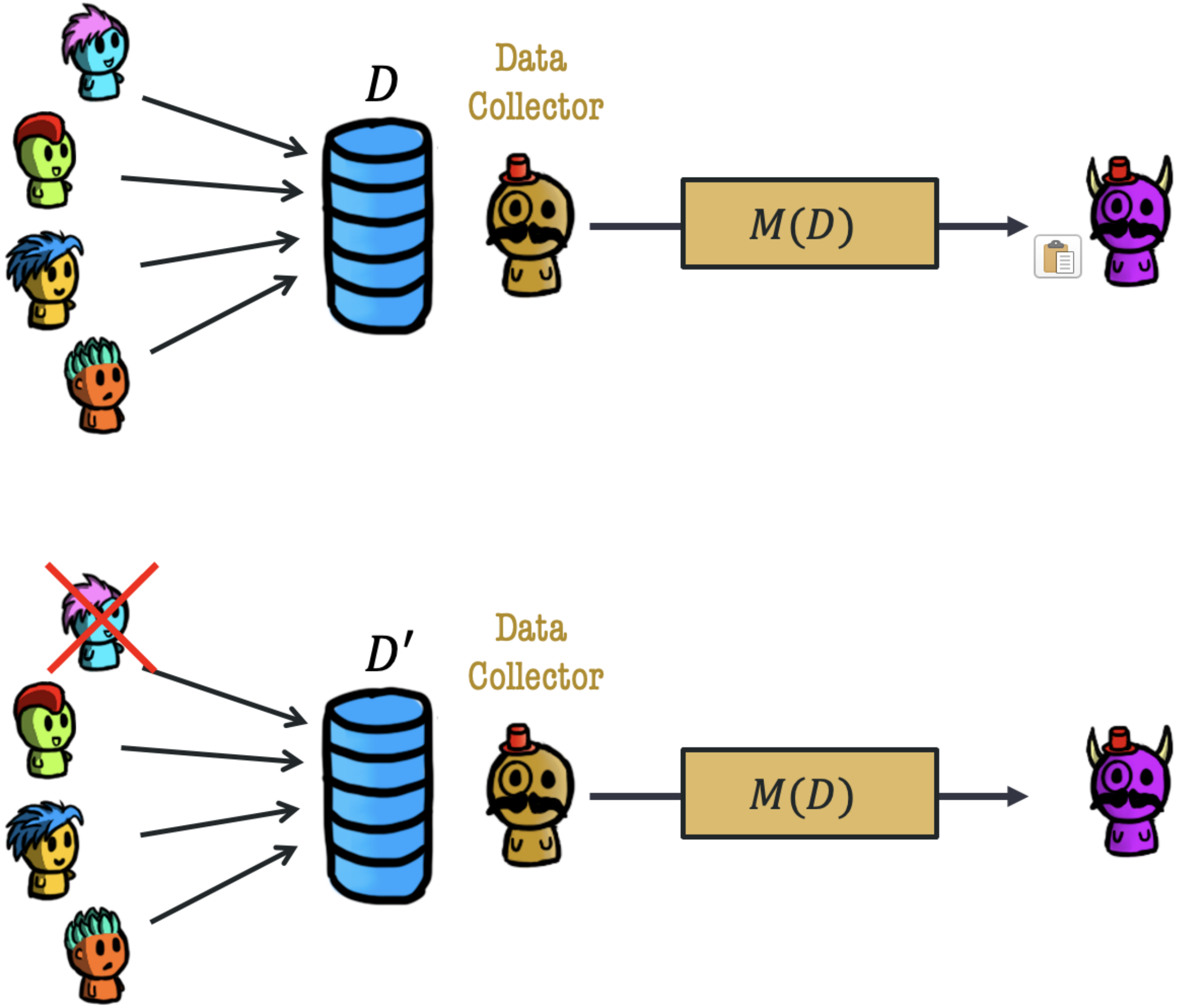
Example: strong auxiliary information



- Note that the adversary reaches the same conclusion in this case, even though Alice has not participated! → We cannot guarantee absolute privacy.

Q: Any ideas of how we could define privacy taking this into account?

Possible Idea:



- If the analyst learns similar things in these two cases about Alice, then M provides enough privacy
- If the adversary learns "a lot" about Alice in both cases, then we cannot prevent this anyway
- Given $R = M(D)$, the adversary should not be able to distinguish whether or not Alice was in the dataset!
- Note that this means that $M(D)$ has to be randomized (or always report the same value, but this makes R constant – independent of D – which is not useful.)

An example from the attacker's perspective

- **Background knowledge 1:** You know that Alice is a top-performer and always gets ≥ 90 in course scores.

- **Background knowledge 2:** CS459 is super-challenging and historical records show that most students score in the range of [45, 55].

An example from the attacker's perspective

- **Background knowledge 1:** You know that Alice is a top-performer and always gets ≥ 90 in course scores.
- **Background knowledge 2:** CS459 is super-challenging and historical records show that most students score in the range of [45, 55].
- **Algorithm:** You are given an algorithm that
  - Allows you to make 5 queries
  - Each query returns the average score of 3 randomly selected students (out of 30 scores in total).

An example from the attacker’s perspective

- **Background knowledge 1:** You know that Alice is a top-performer and always gets ≥ 90 in course scores.

- **Background knowledge 2:** CS459 is super-challenging and historical records show that most students score in the range of [45, 55].

- **Algorithm:** You are given an algorithm that
  - Allows you to make 5 queries
  - Each query returns the average score of 3 randomly selected students (out of 30 scores in total).

Q: How can you infer whether Alice is enrolled in CS459 or not?

The attack

Just send 5 queries and observe what is returned by the database.

- **D** with Alice enrolled:
  - Alice: 90
  - Everyone else (29 of them): 50
- **D’** with Alice not enrolled:
  - Everyone (30 of them): 50

The attack

Just send 5 queries and observe what is returned by the database.

- **D** with Alice enrolled:
  - Alice: 90
  - Everyone else (29 of them): 50
- **D’** with Alice not enrolled:
  - Everyone (30 of them): 50

Q: What will happen if Alice IS NOT enrolled (i.e., D’)?

The attack

Just send 5 queries and observe what is returned by the database.

- **D** with Alice enrolled:
  - Alice: 90
  - Everyone else (29 of them): 50
- **D’** with Alice not enrolled:
  - Everyone (30 of them): 50

Q: What will happen if Alice IS NOT enrolled (i.e., D’)?

A: Expect [50, 50, 50, 50, 50] in response.

The attack

Just send 5 queries and observe what is returned by the database.

- **D** with Alice enrolled:
  - Alice: 90
  - Everyone else (29 of them): 50
- **D'** with Alice not enrolled:
  - Everyone (30 of them): 50

Q: What will happen if Alice IS NOT enrolled (i.e., D')?

A: Expect [50, 50, 50, 50, 50] in response.

Q: What will happen if Alice IS enrolled (i.e., D)?

The attack

Just send 5 queries and observe what is returned by the database.

- **D** with Alice enrolled:
  - Alice: 90
  - Everyone else (29 of them): 50
- **D'** with Alice not enrolled:
  - Everyone (30 of them): 50

Q: What will happen if Alice IS NOT enrolled (i.e., D')?

A: Expect [50, 50, 50, 50, 50] in response.

Q: What will happen if Alice IS enrolled (i.e., D)?

A: For a single response, we either get:

$63 \leftarrow \frac{C_{30}^{2}}{C_{30}^{3}} = 10.7\ \%$

$50 \leftarrow$ otherwise

The attack

Just send 5 queries and observe what is returned by the database.

- **D** with Alice enrolled:
  - Alice: 90
  - Everyone else (29 of them): 50
- **D'** with Alice not enrolled:
  - Everyone (30 of them): 50

Q: What will happen if Alice IS NOT enrolled (i.e., D')?

A: Expect [50, 50, 50, 50, 50] in response.

Q: What will happen if Alice IS enrolled (i.e., D)?

A: For a single response, we either get:

Avg=63 ← $\frac{C_{30}^{2}}{C_{30}^{3}}$ = 10.7 %

Avg=50 ← otherwise

A (cont.): For all 5 responses, the chance of getting at least one 63 is: $1 - (1 - \frac{C_{30}^{2}}{C_{30}^{3}})^5$ = 43.26%

What went wrong?

- Alice's score has **too much impact** on the output! As a result, seeing the output of the algorithm allows the attacker to **differentiate** which database is the underlying database representing the class score.

- This is exactly what **Differential Privacy (DP)** tries to capture!
    - Informally, the DP notion requires **any single** element in a dataset to have only a limited impact on the output.

The strawman defense

- **Background knowledge 1:** You know that Alice is a top-performer and always gets ≥ 90 in course scores.

- **Background knowledge 2:** CS459 is super-challenging and historical records show that most students score in the range of [45, 55].

- **Algorithm:** You are given an algorithm that
  - Allows you to make 5 queries
  - Each query returns the average score of 3 randomly selected students (out of 30 scores in total).

The strawman defense

- **Background knowledge 1:** You know that Alice is a top-performer and always gets ≥ 90 in course scores.

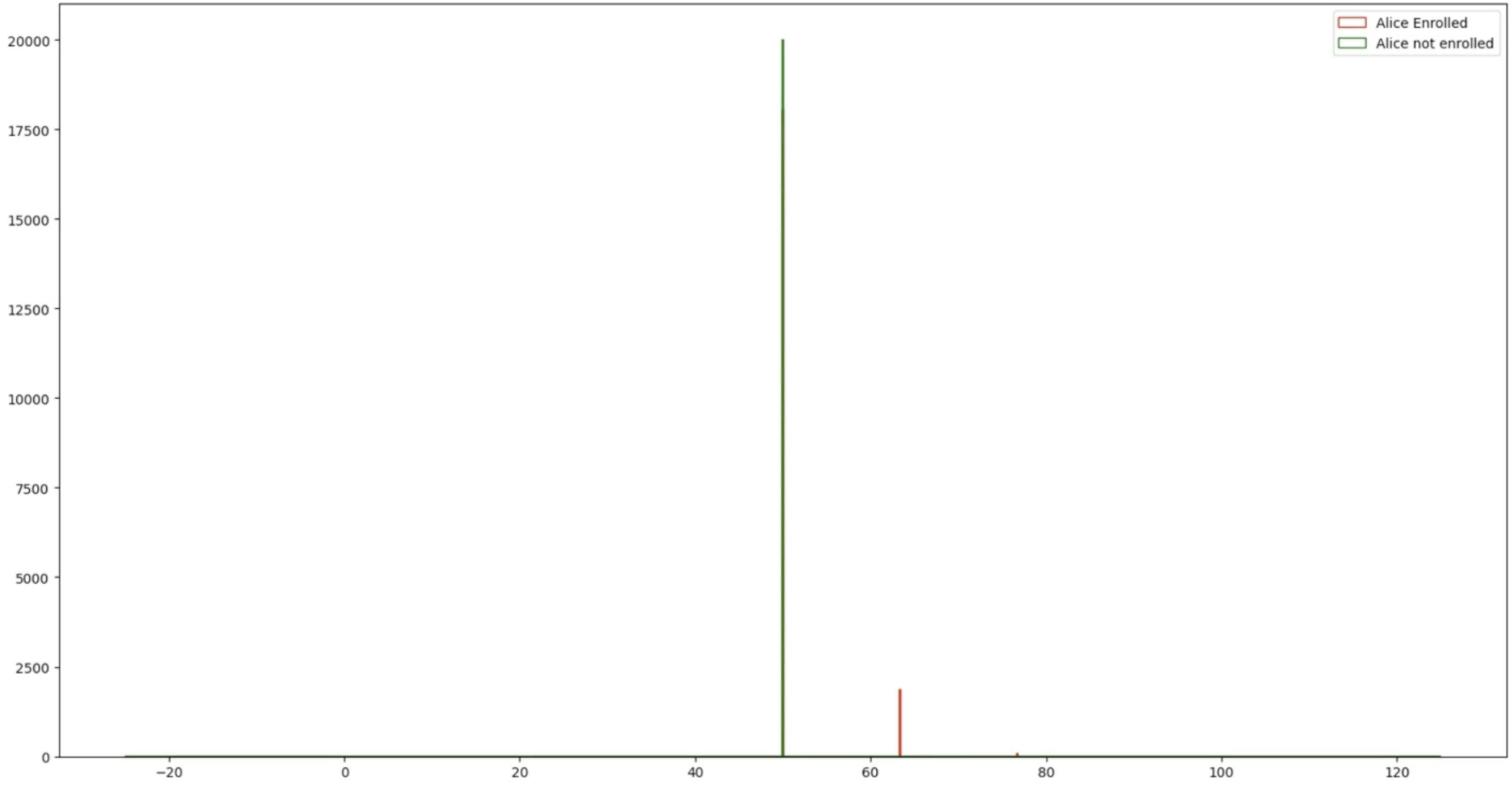
- **Background knowledge 2:** CS459 is super-challenging and historical records show that most students score in the range of [45, 55].

- **Algorithm:** You are given an algorithm that
  - Allows you to make 5 queries
  - Each query returns the average score of 3 randomly selected students (out of 30 scores in total) **plus a random value (i.e., noise).**

Intuition: No noise

When Alice IS in the database: **Noticeable!**

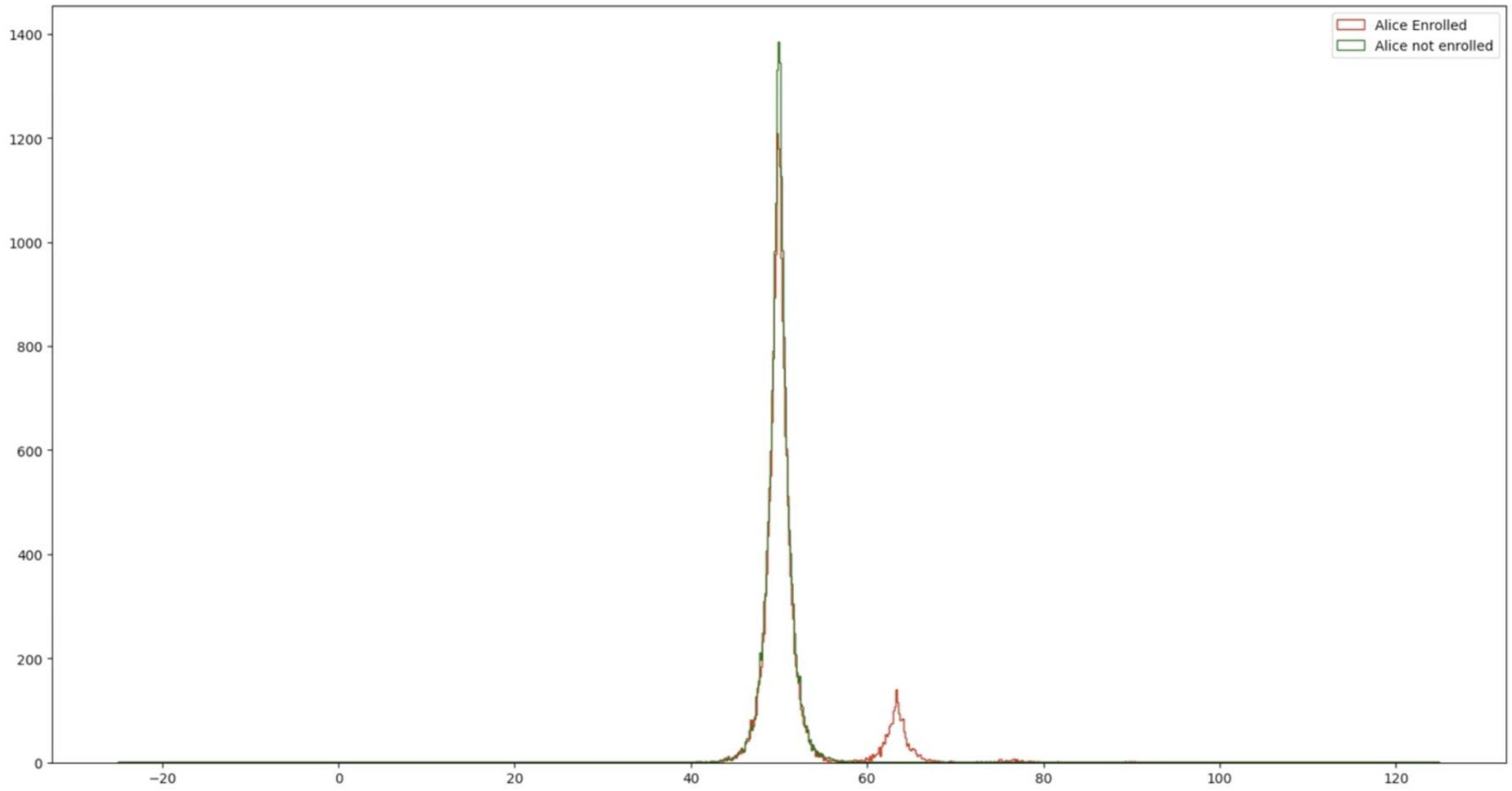
- For a given query, most times it will return 50
- Sometimes (≈ 10%) it will return 63



Intuition: Small noise

When Alice IS in the database: **Still noticeable!**

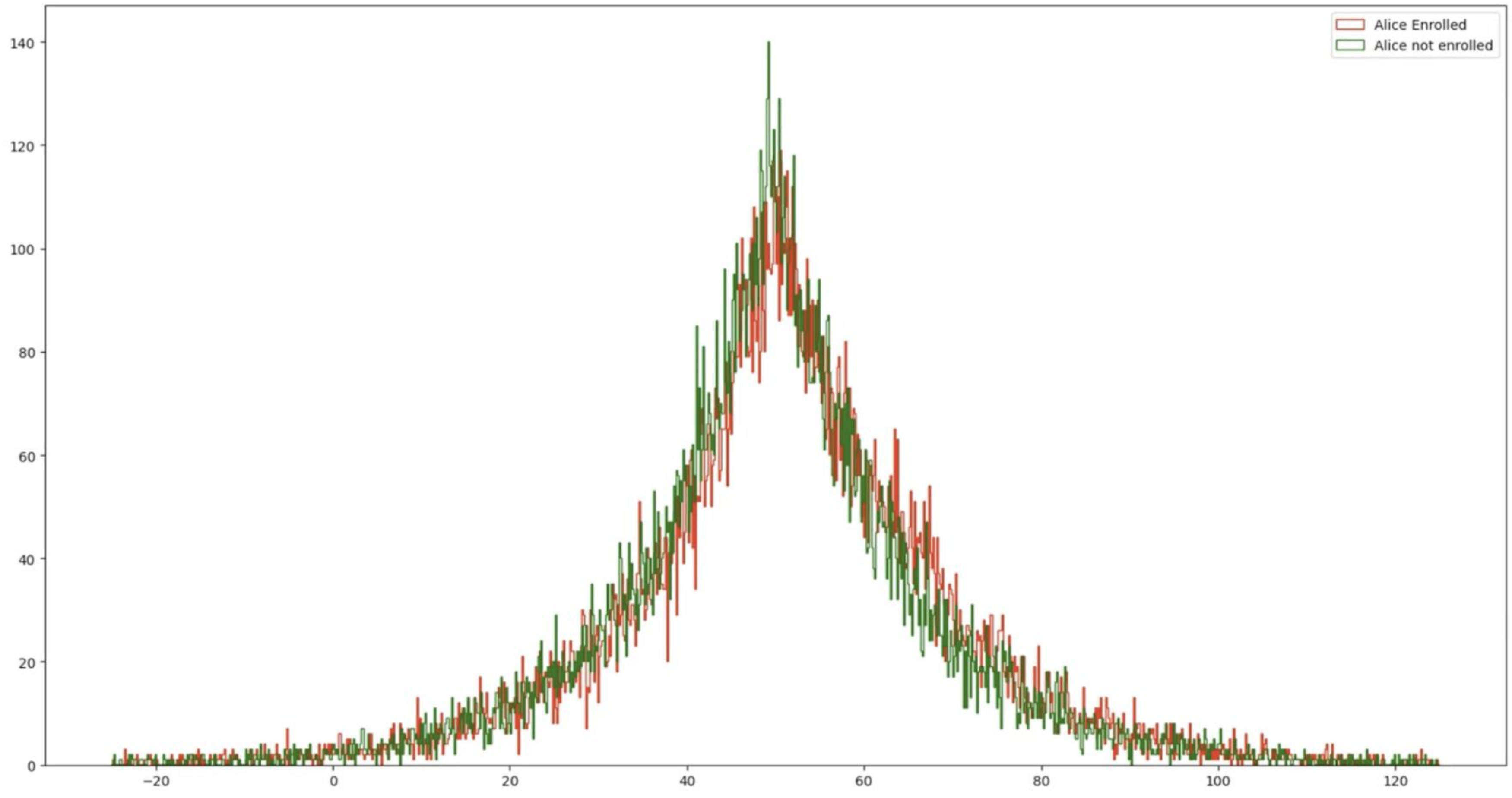
- For a given query, most times it will return ~50
- Sometimes it will return ~63



Intuition: Large noise

When Alice IS in the database: Hardly noticeable!

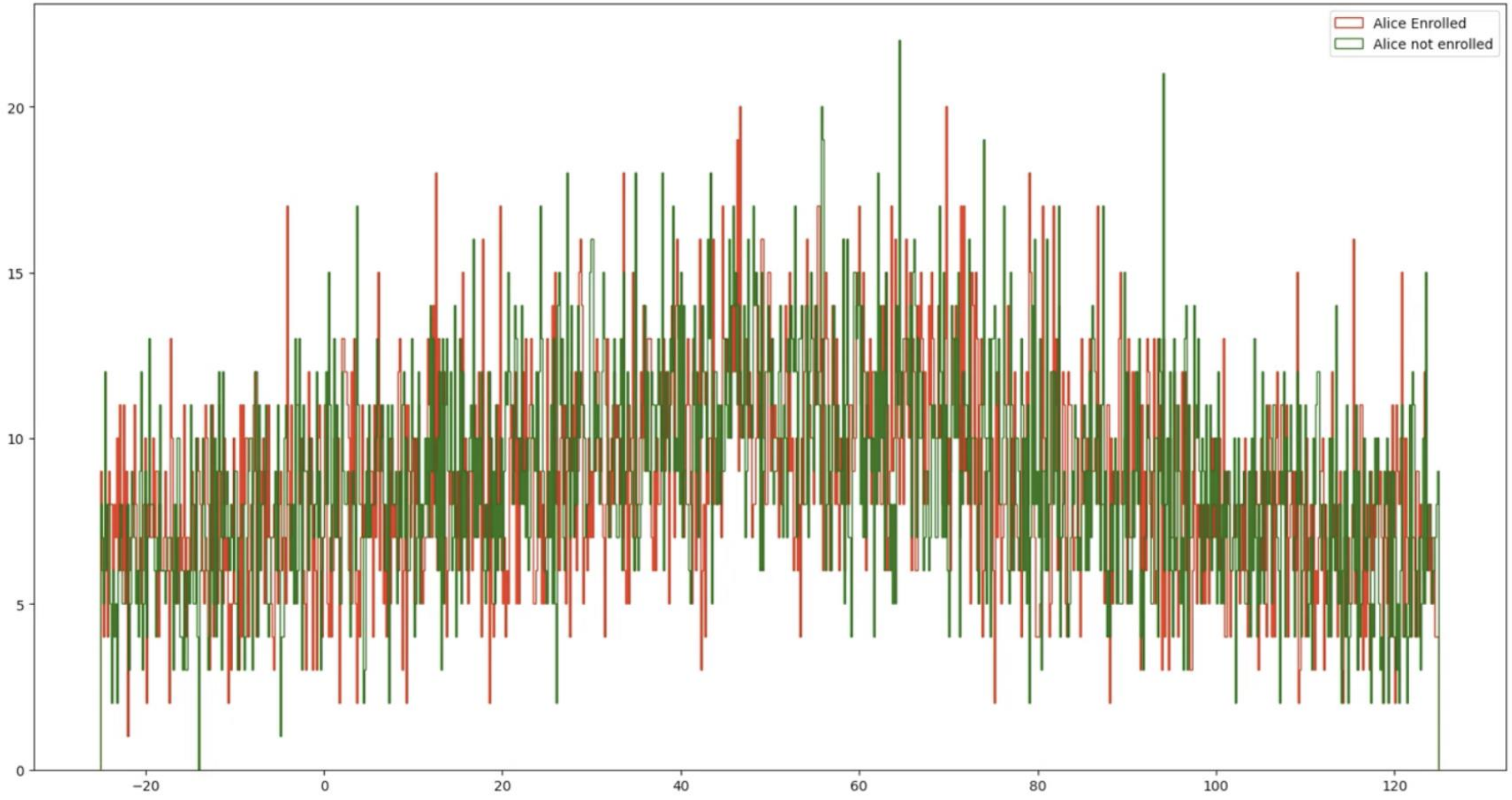
- Query results have a ~ probability whether Alice is in the database or not (with reasonable utility)



When Alice **IS** in the database: **Unnoticeable!**

Intuition: Very large noise

- We can't tell if Alice is in the database
- But we completely destroy utility



Takeaway

- One should set an **appropriate amount of noise** depending on each particular use case.
  - We want to preserve data privacy
  - We don't want to destroy utility

The data collectors' argument

... on trying to persuade you to join a differentially private survey:

- *You will not be affected, adversely or otherwise, by allowing your data to be used in any study or analysis, no matter what other studies, data sets, or information sources, are available. (bla bla… differential privacy … bla bla)*

The data collectors' argument

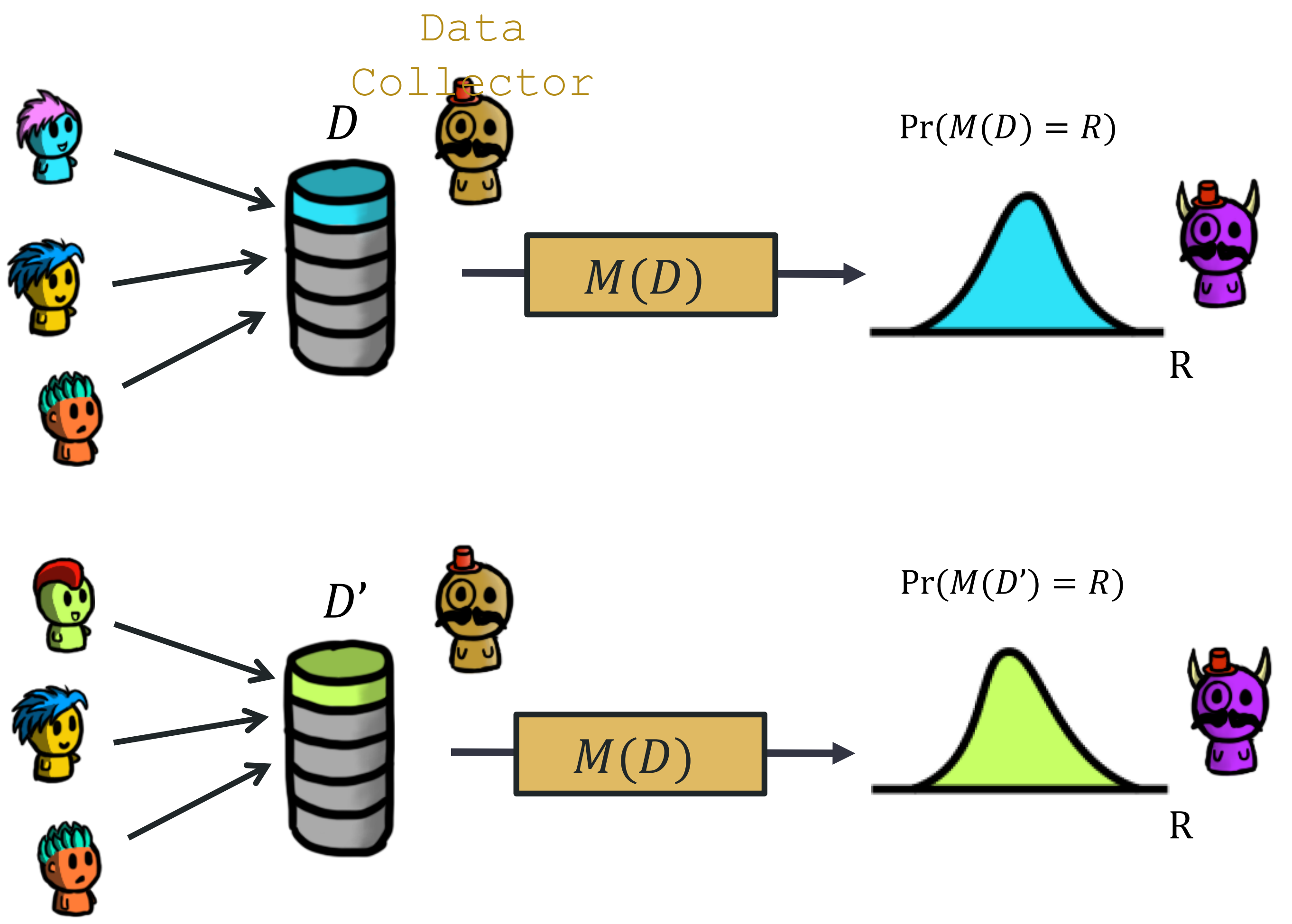
... on trying to persuade you to join a differentially private survey:

- *You will not be affected, adversely or otherwise, by allowing your data to be used in any study or analysis, no matter what other studies, data sets, or information sources, are available. (bla bla… differential privacy … bla bla)*

- But this is only true if they tell you WHAT algorithm they use to release your data and you have verified that their algorithm is indeed differentially private.

Back on topic: We want similar output distributions!

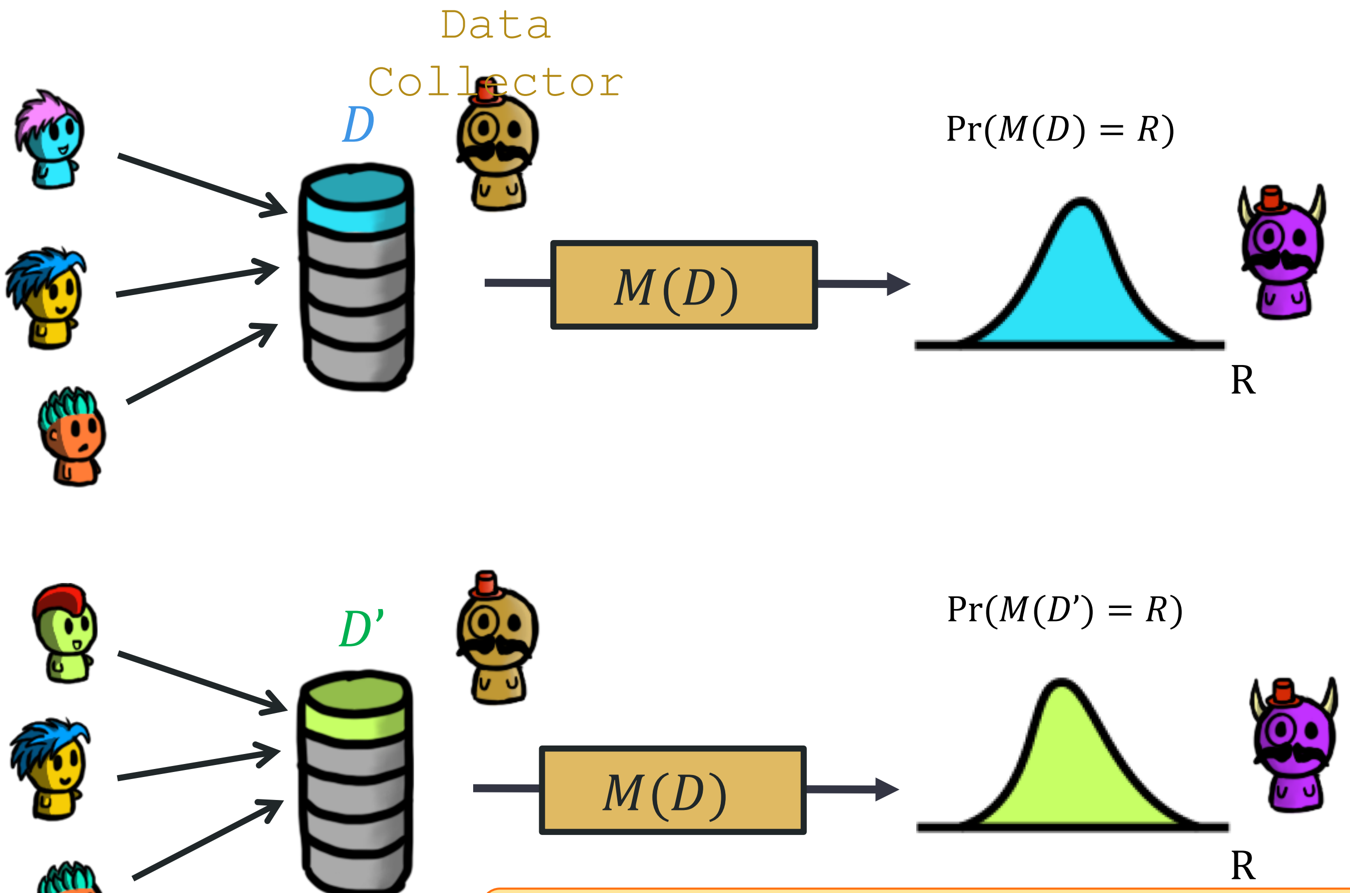
(assume for now that the databases differ on one single record)



- These datasets are usually called neighboring datasets (and usually denoted by D and D')
- We want these distributions to be "similar" (for all R)
- If the mechanism M behaves nearly identically for D and D', then an attacker can't tell whether D or D' was used (and hence can't learn much about the individual).

Back on topic: We want similar output distributions!

(assume for now that the databases differ on one single record)



- These datasets are usually called neighboring datasets (and usually denoted by D and D')
- We want these distributions to be "similar" (for all R)
- If the mechanism M behaves nearly identically for D and D', then an attacker can't tell whether D or D' was used (and hence can't learn much about the individual).

Q: How do we quantify this similarity?

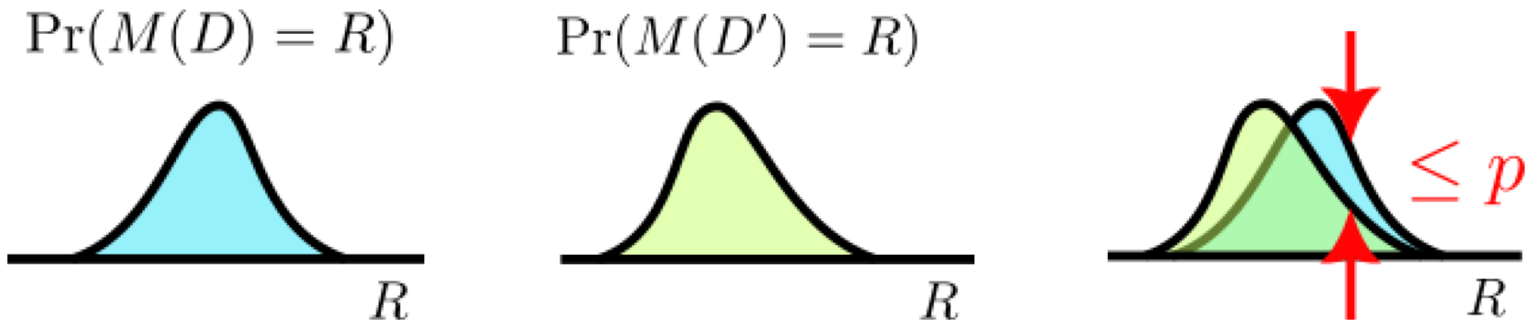
How do we define "similar" distributions?

Tentative privacy definition (with privacy parameter p)

A mechanism M is p-private if the following holds for all possible outputs R and all pairs of neighboring datasets (D, D'):

$$\Pr(M(D') = R) - p \leq \Pr(M(D) = R) \leq \Pr(M(D') = R) + p$$

- This would mean that:



Q: What gives more privacy, small or large p?

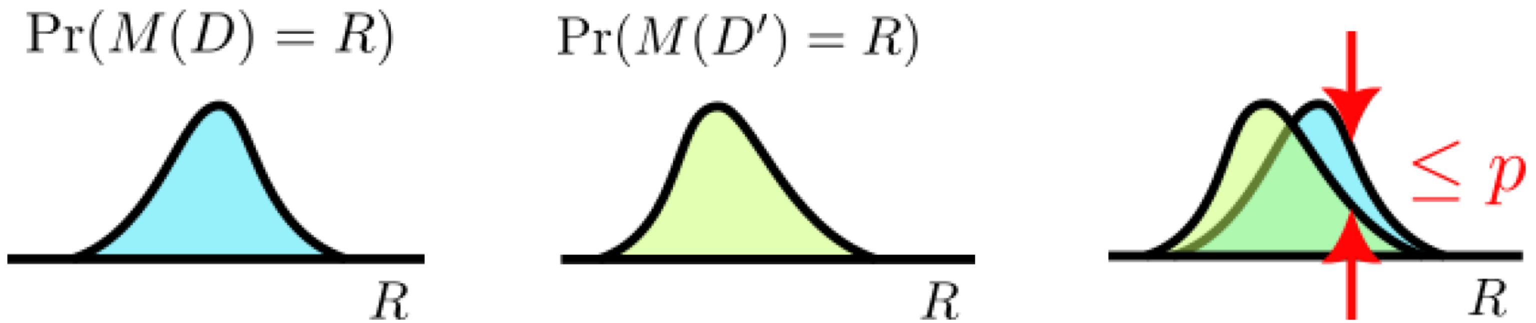
How do we define "similar" distributions?

Tentative privacy definition (with privacy parameter p)

A mechanism M is p-private if the following holds for all possible outputs R and all pairs of neighboring datasets (D, D'):

$$\Pr(M(D') = R) - p \leq \Pr(M(D) = R) \leq \Pr(M(D') = R) + p$$

- This would mean that:



Q: What gives more privacy, small or large p?

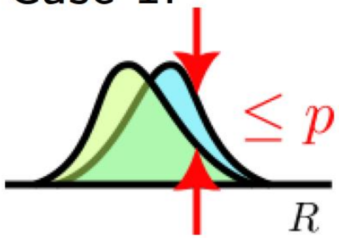
A: Small p, the distributions are more alike

Does this really work?

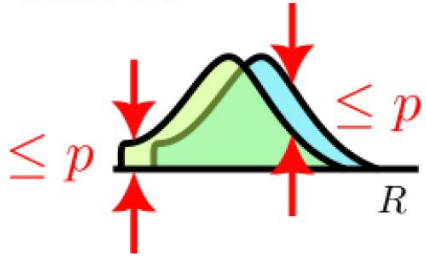
<u>Tentative privacy definition</u> <u>(with privacy parameter p)</u>
A mechanism M is p-private if the following holds for all possible outputs R and all pairs of neighboring datasets (D, D'):

$$\Pr(M(D') = R) - p \le \Pr(M(D) = R) \le \Pr(M(D') = R) + p$$

Case 1:



Case 2:



Q: Case 1 seems fine. What is the issue with case 2?

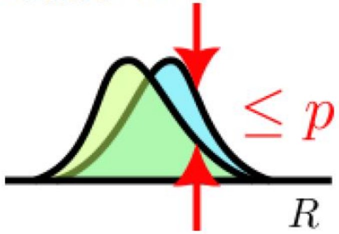
Does this really work?

Tentative privacy definition (with privacy parameter p)

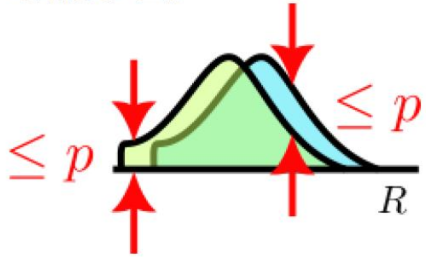
A mechanism M is p-private if the following holds for all possible outputs R and all pairs of neighboring datasets (D, D'):

$$\Pr(M(D') = R) - p \leq \Pr(M(D) = R) \leq \Pr(M(D') = R) + p$$

Case 1:



Case 2:



Q: Case 1 seems fine. What is the issue with case 2?

A: There are some outputs R that can only happen if the input was D (e.g., if Alice was not in the dataset). This allows the adversary to distinguish between D and D' with 100% certainty.

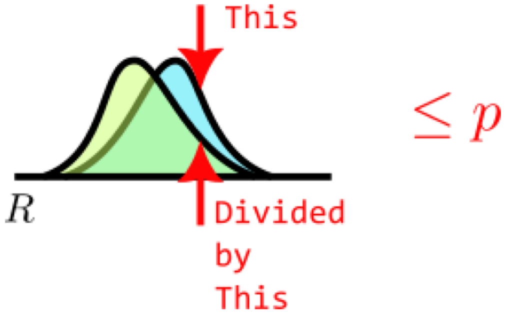
In other words, the attacker can find a **perspective** through which the two databases behave differently.

What if we make the distance multiplicative?

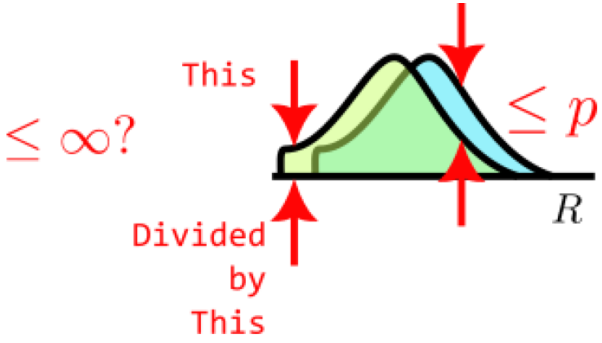
<u>Tentative privacy definition II</u> <u>(with privacy parameter p)</u>
A mechanism M is p-private if the following holds for all possible outputs R and all pairs of neighboring datasets (D, D'):

$$\Pr(M(D') = R) \cdot \frac{1}{p} \leq \Pr(M(D) = R) \leq \Pr(M(D') = R) \cdot p$$

- Again, smaller p (but p ∈ [1, ∞)) means more privacy. This would mean that:



$p = \infty$ means
"NO Privacy"



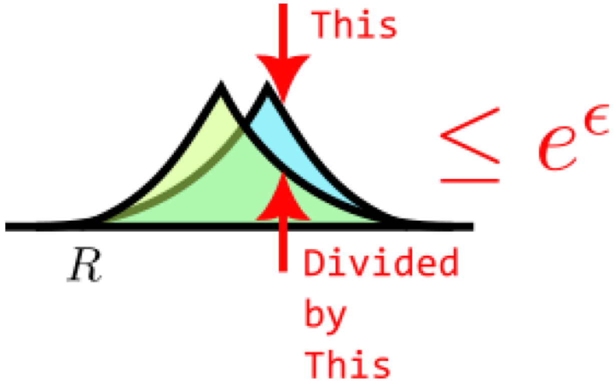
Finally: Differential Privacy

- Same definition, but instead of "p" we use e^{ϵ}

Differential Privacy
A mechanism $M: \mathcal{D} \rightarrow \mathcal{R}$ is ϵ-differentially private (ϵ-DP) if the following holds for all possible outputs $R \in \mathcal{R}$ and all pairs of neighboring datasets $D, D' \in \mathcal{D}$:

$$\Pr(M(D) = R) \leq \Pr(M(D') = R)\, e^{\epsilon}$$

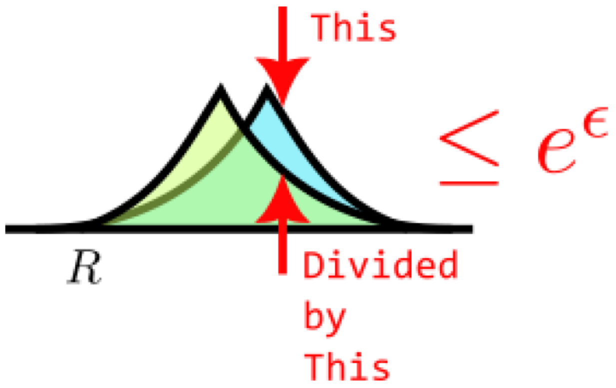
- Some notes:
  - We use e^{ϵ}, instead of just ϵ, because this makes it easier to formulate some useful theorems
  - We do not need the $e^{-\epsilon}$ on the left, since this must hold for all pairs (D, D'). This includes (D', D).
  - $\epsilon \in [0, \infty)$; this ensures that $e^{\epsilon} \in [1, \infty)$



Differential privacy: some questions

<u>Differential Privacy</u>
A mechanism $M : \mathcal{D} \rightarrow \mathcal{R}$ is ϵ-differentially private (ϵ-DP) if the following holds for all possible outputs $R \in \mathcal{R}$ and all pairs of neighboring datasets $D, D' \in \mathcal{D}$:

$$\Pr(M(D) = R) \leq \Pr(M(D') = R)\, e^{\epsilon}$$



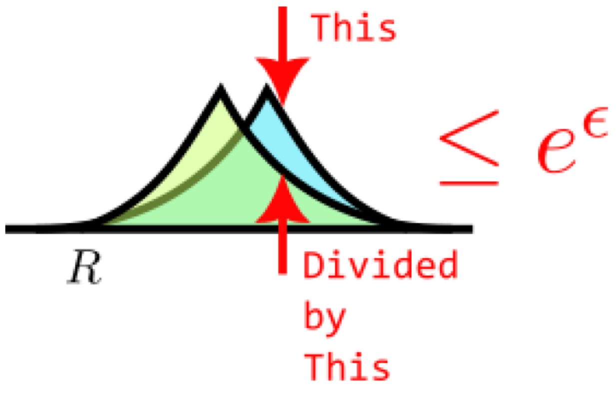
Q: which provides more privacy? $\epsilon = 1$ or $\epsilon = 2$?

Differential privacy: some questions

Differential Privacy

A mechanism $M: \mathcal{D} \rightarrow \mathcal{R}$ is ϵ-differentially private (ϵ-DP) if the following holds for all possible outputs $R \in \mathcal{R}$ and all pairs of neighboring datasets $D, D' \in \mathcal{D}$:

$$\Pr(M(D) = R) \leq \Pr(M(D') = R)\, e^{\epsilon}$$



Q: which provides more privacy? $\epsilon = 1$ or $\epsilon = 2$?

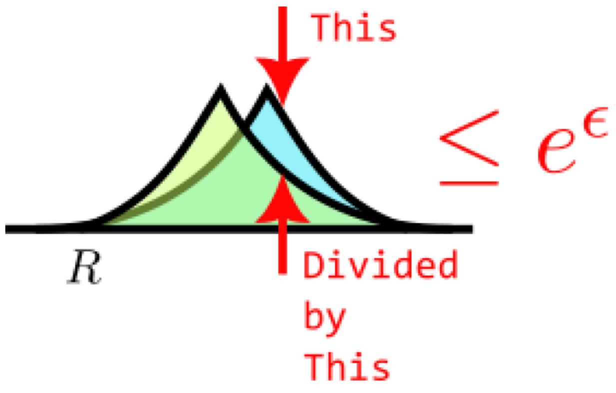
A: Smaller ϵ means more privacy; larger means less privacy

Q: What does $\epsilon = 0$ mean?

Differential privacy: some questions

Differential Privacy
A mechanism $M: \mathcal{D} \rightarrow \mathcal{R}$ is ϵ-differentially private (ϵ-DP) if the following holds for all possible outputs $R \in \mathcal{R}$ and all pairs of neighboring datasets $D, D' \in \mathcal{D}$:

$$\Pr(M(D) = R) \leq \Pr(M(D') = R)\, e^{\epsilon}$$



Q: which provides more privacy? $\epsilon = 1$ or $\epsilon = 2$?

A: Smaller ϵ means more privacy; larger means less privacy

Q: What does $\epsilon = 0$ mean?

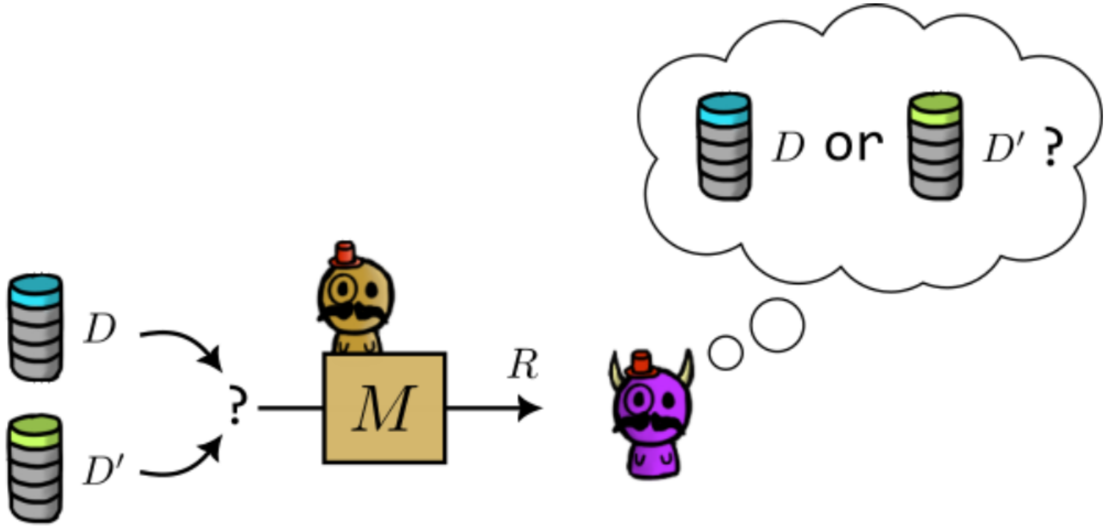
A: Perfect privacy! The output is independent of the dataset! Utility will be very bad.

Some notes on Differential Privacy

- DP was proposed in 2006 by Cynthia Dwork et al. [DMNS06]
- The authors won the Test-of-Time Award in 2016 and the Godel Price in 2017.
- Adopted by big tech like Apple, Google, Microsoft, Facebook, LinkedIn, and by the US Census Bureau for the 2020 US Census
- There is no consensus on how small ϵ should be. “Roughly”
  - $\epsilon < 0.1$ ε is high privacy ($e^{0.1} \approx 1.1$)
  - $0.1 < \epsilon < 1$ is good privacy ($e^{1} \approx 2.7$)
  - $\epsilon > 5$ starts getting too big ($e^{5} \approx 148$)
  - $\epsilon > 100{,}000$ is crazy... yet some works use this

DP interpretation as a game

What does $\Pr(M(D) = R) \leq \Pr(M(D') = R)\, e^{\epsilon}$ even mean ?



We choose the input to be D or D' (at random)

The adversary sees R, and we assume it knows M and knows that the input was either D or D'.

These assumptions are many times unrealistic, but we want privacy even in this worst-case scenario

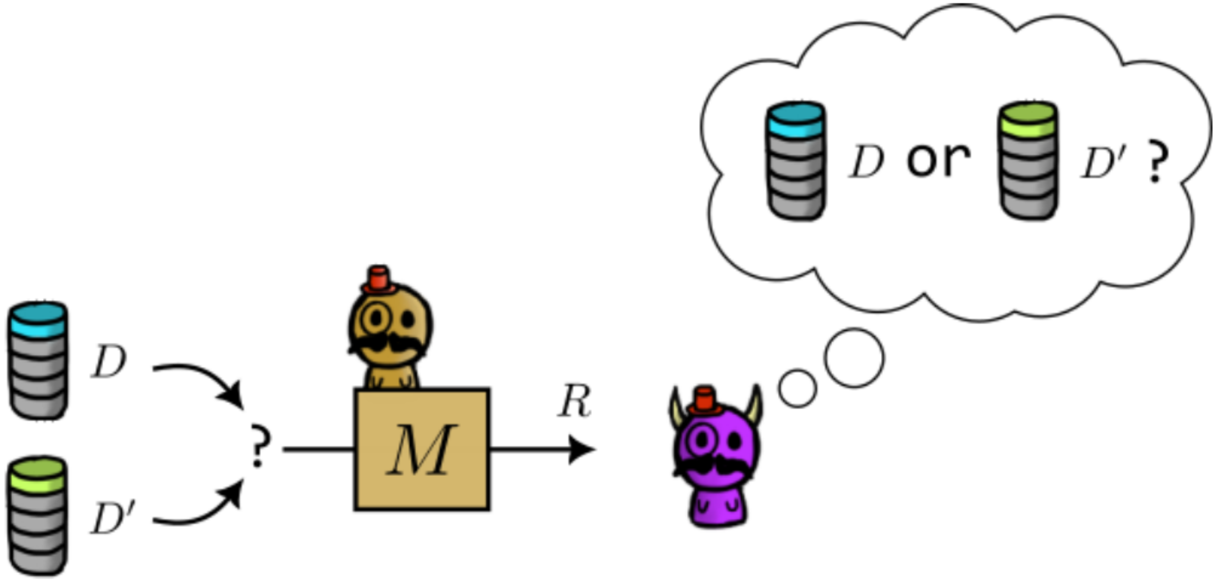
The adversary computes
$p_D = \Pr(M(\mathrm{D}) = R)$ and $p_{D'} = \Pr(M(\mathrm{D}') = R)$

Optimal guess:
The input was D if $p_D \geq p_{D'}$

If M is ϵ-DP, the adversary's probability of error is:

$$\frac{1}{e^{\epsilon}+1} \leq p_{\text{error}} \leq 0.5$$

DP interpretation as a game



What does this mean ?

ϵ	p_{err} range	Privacy
0	$0.5 \leq p_{err} \leq 0.5$	Perfect!
0.1	$0.47 \leq p_{err} \leq 0.5$	Very high
1	$0.26 \leq p_{err} \leq 0.5$	OK?
5	$0.006 \leq p_{err} \leq 0.5$	Bad
10	$0.00004 \leq p_{err} \leq 0.5$	Meaningless?
100 000	$10^{-43\,430} \leq p_{err} \leq 0.5$	

If M is ϵ-DP, the adversary's probability of error is:

$$\frac{1}{e^{\epsilon}+1} \leq \mathrm{p}_{\mathrm{error}} \leq 0.5$$