Anonymity and Privacy

Kostas Papagiannopoulos
University of Amsterdam
k.papagiannopoulos@uva.nl

Contents

Introduction

- Introduction to security vs. Introduction to privacy & anonymity

Introduction

- Introduction to security vs. Introduction to privacy & anonymity
- Security $\neq$ Privacy
- Security tends to be clearly defined, privacy can be contextual and links often to societal and legal aspects

Introduction

- Introduction to security vs. Introduction to privacy & anonymity
- Security $\neq$ Privacy
- Security tends to be clearly defined, privacy can be contextual and links often to societal and legal aspects

- Since 2016, the General Data Protection Regulation (GDPR) specifies our privacy rights within the EU

  `https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02016R0679-20160504`

Introduction

- Introduction to security vs. Introduction to privacy & anonymity
- Security $\neq$ Privacy
- Security tends to be clearly defined, privacy can be contextual and links often to societal and legal aspects

- Since 2016, the General Data Protection Regulation (GDPR) specifies our privacy rights within the EU

  `https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02016R0679-20160504`

- GDPR is a fundamental human right

Introduction

General Data Protection Regulation

- Specifies 'personal data' as information leading **directly or indirectly** to identification of a person

  e.g. linking a name/surname to a medical record

Introduction

General Data Protection Regulation

- Specifies 'personal data' as information leading **directly or indirectly** to identification of a person

  e.g. linking a name/surname to a medical record
- Specifies 'data processing' and 'profiling'

  e.g. performing analysis on your Belastingdienst record to analyse or predict your fraud behavior

Introduction

General Data Protection Regulation

- Specifies 'personal data' as information leading **directly or indirectly** to identification of a person

  e.g. linking a name/surname to a medical record
- Specifies 'data processing' and 'profiling'

  e.g. performing analysis on your Belastingdienst record to analyse or predict your fraud behavior
- Specifies 'anonymization'

  e.g. removing identifiers from your UvA student record so that data like 'BSc start date', 'duration of studies' become public

Introduction

General Data Protection Regulation

- Specifies 'personal data' as information leading **directly or indirectly** to identification of a person

  e.g. linking a name/surname to a medical record
- Specifies 'data processing' and 'profiling'

  e.g. performing analysis on your Belastingdienst record to analyse or predict your fraud behavior
- Specifies 'anonymization'

  e.g. removing identifiers from your UvA student record so that data like 'BSc start date', 'duration of studies' become public
- Specifies 'consent'

  e.g. confirming that you agree to your data being stored and processed by the Amsterdam Gemeente

Introduction

GDPR data-processing principles

1. lawful, fairness, transparency
2. purpose limitation
3. **data minimization**
4. storage minimization
5. accuracy
6. integrity & confidentiality

Introduction

GDPR data-processing principles

1. lawful, fairness, transparency
2. purpose limitation
3. **data minimization**
4. storage minimization
5. accuracy
6. integrity & confidentiality

- The UvA (data controller) shall process your student record, since processing is necessary for the performance of the educational tasks (lawful)

Introduction

GDPR data-processing principles

1. lawful, fairness, transparency
2. purpose limitation
3. **data minimization**
4. storage minimization
5. accuracy
6. integrity & confidentiality

- The UvA (data controller) shall process your student record, since processing is necessary for the performance of the educational tasks (lawful)
- UvA has considered how the processing may affect the student (fair) and the student was sufficiently informed (transparent)

Introduction

GDPR data-processing principles

1. lawful, fairness, transparency
2. purpose limitation
3. **data minimization**
4. storage minimization
5. accuracy
6. integrity & confidentiality

- The UvA (data controller) shall process your student record, since processing is necessary for the performance of the educational tasks (lawful)
- UvA has considered how the processing may affect the student (fair) and the student was sufficiently informed (transparent)
- The only data stored will be your name/surname, start date and grades (data minimization) and UvA will verify it (accuracy)

Introduction

GDPR data-processing principles

1. lawful, fairness, transparency
2. purpose limitation
3. **data minimization**
4. storage minimization
5. accuracy
6. integrity & confidentiality

- The UvA (data controller) shall process your student record, since processing is necessary for the performance of the educational tasks (lawful)
- UvA has considered how the processing may affect the student (fair) and the student was sufficiently informed (transparent)
- The only data stored will be your name/surname, start date and grades (data minimization) and UvA will verify it (accuracy)
- The student record will only be used to decide if you can graduate (purpose limitation)

Introduction

GDPR data-processing principles

1. lawful, fairness, transparency
2. purpose limitation
3. **data minimization**
4. storage minimization
5. accuracy
6. integrity & confidentiality

- The UvA (data controller) shall process your student record, since processing is necessary for the performance of the educational tasks (lawful)
- UvA has considered how the processing may affect the student (fair) and the student was sufficiently informed (transparent)
- The only data stored will be your name/surname, start date and grades (data minimization) and UvA will verify it (accuracy)
- The student record will only be used to decide if you can graduate (purpose limitation)
- After graduation the data will be stored for the next 10 years (storage minimization) in encrypted servers (integrity & confidentiality)

Introduction

GDPR rights

- Right of access: you can access your UvA record and information about how it is being processed

Introduction

GDPR rights

- Right of access: you can access your UvA record and information about how it is being processed
- Right to data portability: you can transfer your UvA record to TU/e

Introduction

GDPR rights

- Right of access: you can access your UvA record and information about how it is being processed
- Right to data portability: you can transfer your UvA record to TU/e
- Right to be forgotten: you can request that UvA erases your record

Introduction

GDPR rights

- Right of access: you can access your UvA record and information about how it is being processed
- Right to data portability: you can transfer your UvA record to TU/e
- Right to be forgotten: you can request that UvA erases your record
- Right to object: you can object to UvA using your data for non-service related purposes such as e.g. marketing

Introduction

GDPR rights

- Right of access: you can access your UvA record and information about how it is being processed
- Right to data portability: you can transfer your UvA record to TU/e
- Right to be forgotten: you can request that UvA erases your record
- Right to object: you can object to UvA using your data for non-service related purposes such as e.g. marketing
- Right to compensation: should UvA violate the GDPR and leak data you are entitled to compensation

k-Anonymity

k-Anonymity

Re-identification by linking

- Databases contain sensitive and non-sensitive information

k-Anonymity

Re-identification by linking

- Databases contain sensitive and non-sensitive information
- **Sensitive** attributes is information that we don't want to be linked to a person

  e.g. medical records, genetic/biometric data, political views, union membership, etc.

k-Anonymity

Re-identification by linking

- Databases contain sensitive and non-sensitive information
- **Sensitive** attributes is information that we don't want to be linked to a person

  e.g. medical records, genetic/biometric data, political views, union membership, etc.
- **Non-sensitive** attributes is information that can be linked to a person

  e.g. country of residence, US sex offender registry

k-Anonymity

Re-identification by linking

- Databases contain sensitive and non-sensitive information
- **Sensitive** attributes is information that we don't want to be linked to a person

  e.g. medical records, genetic/biometric data, political views, union membership, etc.
- **Non-sensitive** attributes is information that can be linked to a person

  e.g. country of residence, US sex offender registry
- Very often we want to publish sensitive information, while maintaining privacy

  e.g. publishing the number of people affected by medical conditions helps health experts analyze and battle diseases

  e.g. publishing a population census helps the government to allocate financial/productive resources

k-Anonymity

- In Massachusetts, the Group Insurance Commission (GIC) collected patient-specific data

k-Anonymity

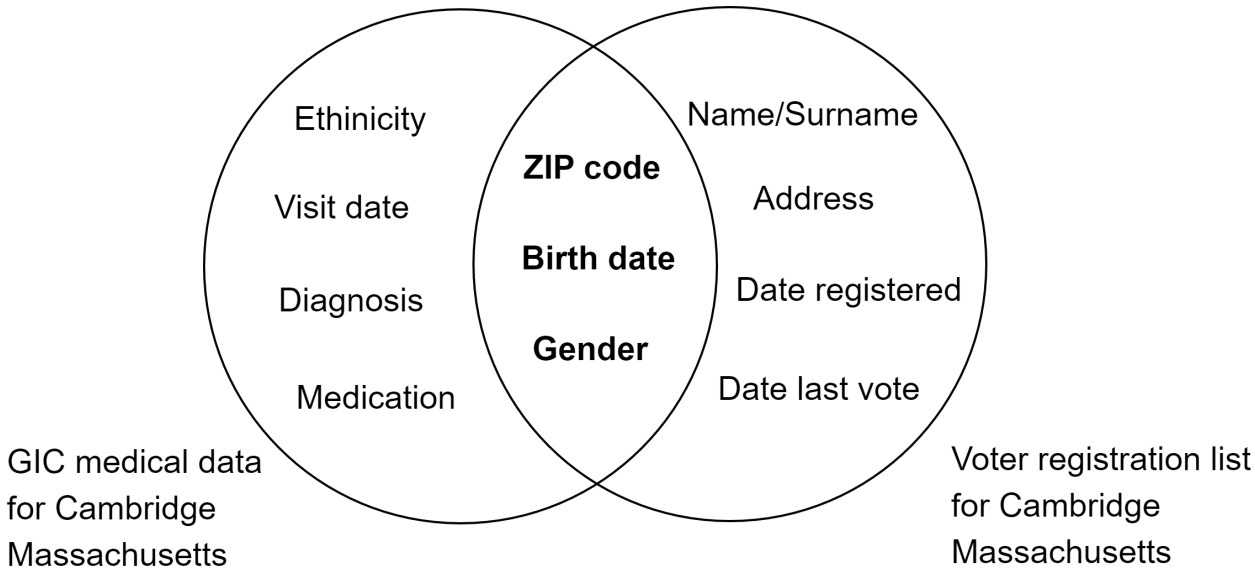
- In Massachusetts, the Group Insurance Commission (GIC) collected patient-specific data
- The GIC dataset was ‘anonymized’ and published so that insurance companies could perform analytics

k-Anonymity

- In Massachusetts, the Group Insurance Commission (GIC) collected patient-specific data
- The GIC dataset was ‘anonymized’ and published so that insurance companies could perform analytics
- In Massachusetts, the voter registration list for Cambridge is public

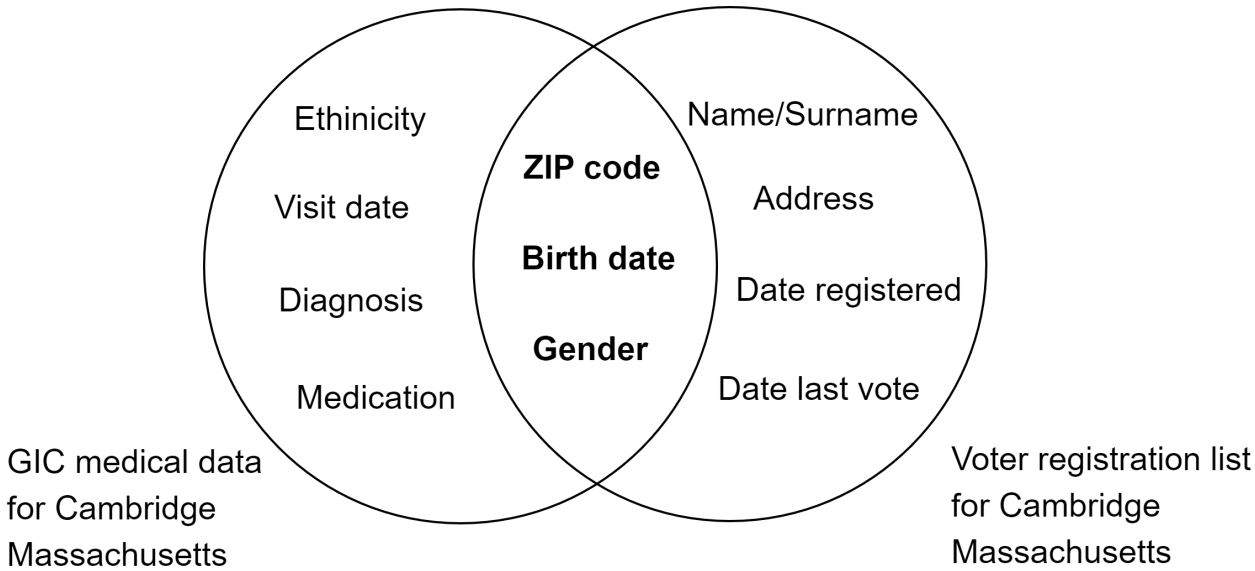
k-Anonymity

- In Massachusetts, the Group Insurance Commission (GIC) collected patient-specific data
- The GIC dataset was 'anonymized' and published so that insurance companies could perform analytics
- In Massachusetts, the voter registration list for Cambridge is public
- The GIC dataset and the voter registration list have overlapping attributes and you could link a patient record (diagnosis, medication, etc.) to a person!



k-Anonymity

- In Massachusetts, the Group Insurance Commission (GIC) collected patient-specific data
- The GIC dataset was 'anonymized' and published so that insurance companies could perform analytics
- In Massachusetts, the voter registration list for Cambridge is public
- The GIC dataset and the voter registration list have overlapping attributes and you could link a patient record (diagnosis, medication, etc.) to a person!



- It was discovered that 87% of the US population can be identified by the attribute set Q

$$Q = \{\text{5-digit ZIP code, Birth date, Gender}\}$$

k-Anonymity

Identifiers

- **Explicit-identifiers.** Attributes that uniquely (or almost uniquely) identify people, causing a disclosure

k-Anonymity

Identifiers

- **Explicit-identifiers.** Attributes that uniquely (or almost uniquely) identify people, causing a disclosure

  e.g. BSN or other social security number, combined name and surname, cellphone number

k-Anonymity

Identifiers

- **Explicit-identifiers.** Attributes that uniquely (or almost uniquely) identify people, causing a disclosure

  e.g. BSN or other social security number, combined name and surname, cellphone number

- The data holder must find explicit-identifiers in their database

k-Anonymity

Identifiers

- **Explicit-identifiers.** Attributes that uniquely (or almost uniquely) identify people, causing a disclosure

  e.g. BSN or other social security number, combined name and surname, cellphone number

- The data holder must find explicit-identifiers in their database
- The explicit-identifiers must be removed from the published dataset

k-Anonymity

Identifiers

- **Explicit-identifiers.** Attributes that uniquely (or almost uniquely) identify people, causing a disclosure

  e.g. BSN or other social security number, combined name and surname, cellphone number

- The data holder must find explicit-identifiers in their database
- The explicit-identifiers must be removed from the published dataset

SSN	Race	Birth year	Gender	ZIP	Medical condition (S)
120-11-1244	Black	1965	m	0214*	short breath
121-37-1499	Black	1965	f	0213*	hypertension
890-55-2209	Black	1964	f	0213*	chest pain
421-41-4412	White	1964	m	0213*	chest pain
889-98-4578	White	1965	m	0213*	chest pain

Extract from the GIC medical data [Sweeney03]

k-Anonymity

Identifiers

- **Explicit-identifiers.** Uniquely (or almost uniquely) identifying information that causes a disclosure

  e.g. BSN or other social security number, combined name and surname, cellphone number

- The data holder must find explicit-identifiers in their database
- The explicit-identifiers must be removed from the published dataset

Race	Birth year	Gender	ZIP	medical condition (S)
Black	1965	m	0214*	short breath
Black	1965	f	0213*	hypertension
Black	1964	f	0213*	chest pain
White	1964	m	0213*	chest pain
White	1965	m	0213*	chest pain

Extract from the GIC medical data [Sweeney03]

k-Anonymity

Identifiers

- **Quasi-identifier** Q**.** A set of attributes that, in combination, can uniquely characterize individuals

k-Anonymity

Identifiers

- **Quasi-identifier** Q**.** A set of attributes that, in combination, can uniquely characterize individuals

  e.g. in the table consider the attribute subset {Race, Birth year, Gender}

Race	Birth year	Gender	ZIP	Medical condition (S)
Black	1965	m	0214*	short breath
Black	1965	f	0213*	hypertension
Black	1964	f	0213*	chest pain
White	1964	m	0213*	chest pain
White	1965	m	0213*	chest pain

k-Anonymity

Identifiers

- **Quasi-identifier** Q**.** A set of attributes that, in combination, can uniquely characterize individuals

  e.g. in the table consider the attribute subset {Race, Birth year, Gender}

Race	Birth year	Gender	ZIP	Medical condition (S)
Black	1965	m	0214*	short breath
Black	1965	f	0213*	hypertension
Black	1964	f	0213*	chest pain
White	1964	m	0213*	chest pain
White	1965	m	0213*	chest pain

- The quasi-identifier Q ={Race, Birth year, Gender} uniquely characterizes all individuals in the table

k-Anonymity

Identifiers

- **Quasi-identifier** Q**.** A set of attributes that, in combination, can uniquely characterize individuals

  e.g. in the table consider the attribute subset {Race, Birth year, Gender}

Race	Birth year	Gender	ZIP	Medical condition (S)
Black	1965	m	0214*	short breath
Black	1965	f	0213*	hypertension
Black	1964	f	0213*	chest pain
White	1964	m	0213*	chest pain
White	1965	m	0213*	chest pain

- The quasi-identifier Q ={Race, Birth year, Gender} uniquely characterizes all individuals in the table
- Notice that none of these attributes are standalone explicit-identifiers

k-Anonymity

Identifiers

- **Quasi-identifier** Q**.** A set of attributes that, in combination, can uniquely characterize individuals

  e.g. in the table consider the attribute subset {Race, Birth year, Gender}

Race	Birth year	Gender	ZIP	Medical condition (S)
Black	1965	m	0214*	short breath
Black	1965	f	0213*	hypertension
Black	1964	f	0213*	chest pain
White	1964	m	0213*	chest pain
White	1965	m	0213*	chest pain

- The quasi-identifier Q ={Race, Birth year, Gender} uniquely characterizes all individuals in the table
- Notice that none of these attributes are standalone explicit-identifiers
- The quasi-identifier Q ={Race, Birth year} does not uniquely characterize all individuals in the table

k-Anonymity

- Quasi-identifiers can be linked to external data to uniquely identify individuals and cause a disclosure

k-Anonymity

- Quasi-identifiers can be linked to external data to uniquely identify individuals and cause a disclosure
- It suffices to find an external database (like the voter registration database) that includes some explicit-identifier and link it to the quasi-identifier found in our database

k-Anonymity

- Quasi-identifiers can be linked to external data to uniquely identify individuals and cause a disclosure
- It suffices to find an external database (like the voter registration database) that includes some explicit-identifier and link it to the quasi-identifier found in our database

Race	Birth year	Gender	ZIP	Medical condition (S)
Black	1965	m	0214*	short breath
Black	1965	f	0213*	hypertension
Black	1964	f	0213*	chest pain
White	1964	m	0213*	chest pain
White	1965	m	0213*	chest pain

Extract from the GIC medical data [Sweeney03]

Name/Surname	Race	Birth year	Gender	Registration date
John Smith	Black	1965	m	7-3-2000
Helen Larsen	Black	1965	f	12-12-1990
Albert Young	Black	1964	f	4-10-2001
Mary Davis	White	1964	m	5-11-1987
Mary Miller	White	1965	m	17-9-1985

Extract from the voter registration database [Sweeney03]

k-Anonymity

Definition of k-Anonymity

- Let a database $\mathcal{D}$ with non-sensitive attributes $\{A_1, A_2, \ldots, A_n\}$

k-Anonymity

Definition of k-Anonymity

- Let a database $\mathcal{D}$ with non-sensitive attributes $\{A_1, A_2, \ldots, A_n\}$
- Let a quasi-identifier $Q = \{A_i, \ldots, A_j\} \subset \{A_1, A_2, \ldots, A_n\}$

k-Anonymity

Definition of k-Anonymity

- Let a database $\mathcal{D}$ with non-sensitive attributes $\{A_1, A_2, \ldots, A_n\}$
- Let a quasi-identifier $Q = \{A_i, \ldots, A_j\} \subset \{A_1, A_2, \ldots, A_n\}$
- Let the database projection $\mathcal{D}[Q]$
  i.e. discard all attributes in $\mathcal{D}$ that are not in the Q set

k-Anonymity

Definition of k-Anonymity

- Let a database $\mathcal{D}$ with non-sensitive attributes $\{A_1, A_2, \ldots, A_n\}$
- Let a quasi-identifier $Q = \{A_i, \ldots, A_j\} \subset \{A_1, A_2, \ldots, A_n\}$
- Let the database projection $\mathcal{D}[Q]$
  i.e. discard all attributes in $\mathcal{D}$ that are not in the Q set

- k-**Anonymity.** We say that the database $\mathcal{D}$ satisfies k-anonymity with respect to quasi-identifier Q iff every sequence of values in the projection $\mathcal{D}[Q]$ appears with at least k occurrences

k-Anonymity

Example with $k = 2$

- The table satisfies 2-anonymity with respect to the quasi-identifier Q ={Race, Birth year, Gender, ZIP}

Race	Birth year	Gender	ZIP	Medical condition (S)
Black	1965	m	0214*	short breath
Black	1965	m	0214*	chest pain
Black	1965	f	0213*	hypertension
Black	1965	f	0213*	hypertension
Black	1964	f	0213*	obesity
Black	1964	f	0213*	chest pain
White	1964	m	0213*	chest pain
White	1964	m	0213*	obesity
White	1964	m	0213*	short breath
White	1967	m	0213*	chest pain
White	1967	m	0213*	chest pain

k-Anonymity

Example with $k = 2$

Race	Birth year	Gender	ZIP	Medical condition (S)
Black	1965	m	0214*	short breath
Black	1965	m	0214*	chest pain
Black	1965	f	0213*	hypertension
Black	1965	f	0213*	hypertension
Black	1964	f	0213*	obesity
Black	1964	f	0213*	chest pain
White	1964	m	0213*	chest pain
White	1964	m	0213*	obesity
White	1964	m	0213*	short breath
White	1967	m	0213*	chest pain
White	1967	m	0213*	chest pain

k-Anonymity

Example with $k = 2$

Race	Birth year	Gender	ZIP	Medical condition (S)
Black	1965	m	0214*	short breath
Black	1965	m	0214*	chest pain
Black	1965	f	0213*	hypertension
Black	1965	f	0213*	hypertension
Black	1964	f	0213*	obesity
Black	1964	f	0213*	chest pain
White	1964	m	0213*	chest pain
White	1964	m	0213*	obesity
White	1964	m	0213*	short breath
White	1967	m	0213*	chest pain
White	1967	m	0213*	chest pain

- There are at least 2 occurrences of each possible attribute value within the database projection $\mathcal{D}(QI)$

$$|\mathcal{D}(\text{Race=Black})| = 6, |\mathcal{D}(\text{Race=White})| = 5$$

$$|\mathcal{D}(\text{Birth year=1964})| = 5, |\mathcal{D}(\text{Birth year=1965})| = 4, \mathcal{D}(\text{Birth year=1967})| = 2$$

$$|\mathcal{D}(\text{Gender=m})| = 7, |\mathcal{D}(\text{Gender=f})| = 4, \text{ etc.}$$

k-Anonymity

Attacks on k-Anonymity

- Ensuring k-anonymity prevents the database linkage attacks with a bound of k

k-Anonymity

Attacks on k-Anonymity

- Ensuring k-anonymity prevents the database linkage attacks with a bound of k
- Linking the datasets to external sources will not yield less than k individuals

k-Anonymity

Attacks on k-Anonymity

- Ensuring k-anonymity prevents the database linkage attacks with a bound of k
- Linking the datasets to external sources will not yield less than k individuals

Unsorted matching attack

- Consider the following database $\mathcal{D}$
- Notice that $\mathcal{D}$ does not satisfy k-anonymity w.r.t. $Q = \{\text{Race, ZIP}\}$

Race	ZIP	Medical condition (S)
Asian	02138	hypertension
Asian	02139	short breath
Asian	02141	hypertension
Asian	02141	chest pain
Black	02138	chest pain
Black	02139	obesity
White	02138	short breath
White	02139	chest pain
White	02138	chest pain

k-Anonymity

Unsorted matchhing attack

- We suppress the last digit of the ZIP attribute to satisfy 2-anonymity, publishing database $\mathcal{D}_1^*$

Race	ZIP	Medical condition (S)
Asian	0213*	hypertension
Asian	0213*	short breath
Asian	0214*	hypertension
Asian	0214*	chest pain
Black	0213*	chest pain
Black	0213*	obesity
White	0213*	short breath
White	0213*	chest pain
White	0213*	chest pain

k-Anonymity

Unsorted matchhing attack

- We suppress the last digit of the ZIP attribute to satisfy 2-anonymity, publishing database $\mathcal{D}_1^*$

Race	ZIP	Medical condition (S)
Asian	0213*	hypertension
Asian	0213*	short breath
Asian	0214*	hypertension
Asian	0214*	chest pain
Black	0213*	chest pain
Black	0213*	obesity
White	0213*	short breath
White	0213*	chest pain
White	0213*	chest pain

- This anonymization technique is known as **domain generalization**

$$Domain = \{02138, 02139, 02141\}, \quad Domain^* = \{0213^*, 0214^*\}$$

$$\phi(02138) = 0213^*,\ \phi(02139) = 0213^*,\ \phi(02141) = 0214^*$$

k-Anonymity

Unsorted matching attack

- We suppress the Race attribute to satisfy 2-anonymity, publishing database $\mathcal{D}_2^*$

Race	ZIP	Medical condition (S)
*	02138	hypertension
*	02139	short breath
*	02141	hypertension
*	02141	chest pain
*	02138	chest pain
*	02139	obesity
*	02138	short breath
*	02139	chest pain
*	02138	chest pain

k-Anonymity

Unsorted matching attack

- We suppress the Race attribute to satisfy 2-anonymity, publishing database $\mathcal{D}_2^*$

Race	ZIP	Medical condition (S)
*	02138	hypertension
*	02139	short breath
*	02141	hypertension
*	02141	chest pain
*	02138	chest pain
*	02139	obesity
*	02138	short breath
*	02139	chest pain
*	02138	chest pain

- Full domain generalization has been applied

$$Domain = \{\text{Asian, Black, White}\}, \quad Domain^* = \{^*\}$$

$$\phi(\text{Black}) = \phi(\text{Asian}) = \phi(\text{White}) = ^*$$

k-Anonymity

Unsorted matching attack

- What happens if the anonymized databases $\mathcal{D}_1^*$ and $\mathcal{D}_2^*$ follow the exact same row order?

Race	ZIP	Race	ZIP	Medical condition (S)
Asian	0213*	*	02138	hypertension
Asian	0213*	*	02139	short breath
Asian	0214*	*	02141	hypertension
Asian	0214*	*	02141	chest pain
Black	0213*	*	02138	chest pain
Black	0213*	*	02139	obesity
White	0213*	*	02138	short breath
White	0213*	*	02139	chest pain
White	0213*	*	02138	chest pain

k-Anonymity

Unsorted matching attack

- What happens if the anonymized databases $\mathcal{D}_1^*$ and $\mathcal{D}_2^*$ follow the exact same row order?

Race	ZIP	Race	ZIP	Medical condition (S)
Asian	0213*	*	02138	hypertension
Asian	0213*	*	02139	short breath
Asian	0214*	*	02141	hypertension
Asian	0214*	*	02141	chest pain
Black	0213*	*	02138	chest pain
Black	0213*	*	02139	obesity
White	0213*	*	02138	short breath
White	0213*	*	02139	chest pain
White	0213*	*	02138	chest pain

- You can de-anonymize the database by placing $\mathcal{D}_1^*$ and $\mathcal{D}_2^*$ side-by-side

k-Anonymity

Unsorted matching attack

- What happens if the anonymized databases $\mathcal{D}_1^*$ and $\mathcal{D}_2^*$ follow the exact same row order?

Race	ZIP	Race	ZIP	Medical condition (S)
Asian	0213*	*	02138	hypertension
Asian	0213*	*	02139	short breath
Asian	0214*	*	02141	hypertension
Asian	0214*	*	02141	chest pain
Black	0213*	*	02138	chest pain
Black	0213*	*	02139	obesity
White	0213*	*	02138	short breath
White	0213*	*	02139	chest pain
White	0213*	*	02138	chest pain

- You can de-anonymize the database by placing $\mathcal{D}_1^*$ and $\mathcal{D}_2^*$ side-by-side
- What would be a countermeasure?

k-Anonymity

Unsorted matching attack

- What happens if the anonymized databases $\mathcal{D}_1^*$ and $\mathcal{D}_2^*$ follow the exact same row order?

Race	ZIP	Race	ZIP	Medical condition (S)
Asian	0213*	*	02138	hypertension
Asian	0213*	*	02139	short breath
Asian	0214*	*	02141	hypertension
Asian	0214*	*	02141	chest pain
Black	0213*	*	02138	chest pain
Black	0213*	*	02139	obesity
White	0213*	*	02138	short breath
White	0213*	*	02139	chest pain
White	0213*	*	02138	chest pain

- You can de-anonymize the database by placing $\mathcal{D}_1^*$ and $\mathcal{D}_2^*$ side-by-side
- What would be a countermeasure?
  Randomize the order of rows.

k-Anonymity

Unsorted matching attack

- What happens if the anonymized databases $\mathcal{D}_1^*$ and $\mathcal{D}_2^*$ follow the exact same row order?

Race	ZIP	Race	ZIP	Medical condition (S)
Asian	0213*	*	02138	hypertension
Asian	0213*	*	02139	short breath
Asian	0214*	*	02141	hypertension
Asian	0214*	*	02141	chest pain
Black	0213*	*	02138	chest pain
Black	0213*	*	02139	obesity
White	0213*	*	02138	short breath
White	0213*	*	02139	chest pain
White	0213*	*	02138	chest pain

- You can de-anonymize the database by placing $\mathcal{D}_1^*$ and $\mathcal{D}_2^*$ side-by-side
- What would be a countermeasure?
  Randomize the order of rows.
- This anonymization technique is an example of **data swapping and randomization**

k-Anonymity

Complementary release attack

- To avoid the unsorted matching attack we randomized the order of rows in databases $\mathcal{D}_1^*$ and $\mathcal{D}_2^*$

k-Anonymity

Complementary release attack

- To avoid the unsorted matching attack we randomized the order of rows in databases $\mathcal{D}_1^*$ and $\mathcal{D}_2^*$
- After randomizing, we publish databases $\mathcal{D}_1^*$ and $\mathcal{D}_2^*$ and place them side-by-side

Race	ZIP	Condition (S)	Race	ZIP	Condition (S)
Asian	0213*	hypertension	*	02139	short breath
Asian	0213*	short breath	*	02138	hypertension
Asian	0214*	hypertension	*	02141	chest pain
Asian	0214*	chest pain	*	02141	hypertension
Black	0213*	chest pain	*	02138	short breath
Black	0213*	obesity	*	02138	chest pain
White	0213*	short breath	*	02138	chest pain
White	0213*	chest pain	*	02139	chest pain
White	0213*	chest pain	*	02139	obesity

k-Anonymity

Complementary release attack

- To avoid the unsorted matching attack we randomized the order of rows in databases $\mathcal{D}_1^*$ and $\mathcal{D}_2^*$
- After randomizing, we publish databases $\mathcal{D}_1^*$ and $\mathcal{D}_2^*$ and place them side-by-side

Race	ZIP	Condition (S)
Asian	0213*	hypertension
Asian	0213*	short breath
Asian	0214*	hypertension
Asian	0214*	chest pain
Black	0213*	chest pain
Black	0213*	obesity
White	0213*	short breath
White	0213*	chest pain
White	0213*	chest pain

Race	ZIP	Condition (S)
*	02139	short breath
*	02138	hypertension
*	02141	chest pain
*	02141	hypertension
*	02138	short breath
*	02138	chest pain
*	02138	chest pain
*	02139	chest pain
*	02139	obesity

- Notice that the quasi-identifier $Q = \{\text{Race, ZIP}\}$ i.e. Medical condition $\notin Q$

k-Anonymity

Complementary release attack

- Linking the randomized $\mathcal{D}_1^*$ and $\mathcal{D}_2^*$ on the Medical Condition violates the 2-anonymity on $Q = \{\text{Race, ZIP}\}$
- E.g. the rows `[Black, 02139, obesity]` and `[Asian, 02138, short breath]` are unique

Race	ZIP	Condition (S)	Race	ZIP	Condition (S)
Asian	0213*	hypertension	*	02139	short breath
Asian	0213*	short breath	*	02138	hypertension
Asian	0214*	hypertension	*	02141	chest pain
Asian	0214*	chest pain	*	02141	hypertension
Black	0213*	chest pain	*	02138	short breath
Black	0213*	obesity	*	02138	chest pain
White	0213*	short breath	*	02138	chest pain
White	0213*	chest pain	*	02139	chest pain
White	0213*	chest pain	*	02139	obesity

k-Anonymity

Complementary release attack

- Linking the randomized $\mathcal{D}_1^*$ and $\mathcal{D}_2^*$ on the Medical Condition violates the 2-anonymity on $Q = \{\text{Race, ZIP}\}$
- E.g. the rows `[Black, 02139, obesity]` and `[Asian, 02138, short breath]` are unique

Race	ZIP	Condition (S)	Race	ZIP	Condition (S)
Asian	0213*	hypertension	*	02139	short breath
Asian	0213*	short breath	*	02138	hypertension
Asian	0214*	hypertension	*	02141	chest pain
Asian	0214*	chest pain	*	02141	hypertension
Black	0213*	chest pain	*	02138	short breath
Black	0213*	obesity	*	02138	chest pain
White	0213*	short breath	*	02138	chest pain
White	0213*	chest pain	*	02139	chest pain
White	0213*	chest pain	*	02139	obesity

- What would be a countermeasure?

k-Anonymity

Complementary release attack

- Linking the randomized $\mathcal{D}_1^*$ and $\mathcal{D}_2^*$ on the Medical Condition violates the 2-anonymity on $Q = \{\text{Race, ZIP}\}$
- E.g. the rows `[Black, 02139, obesity]` and `[Asian, 02138, short breath]` are unique

Race	ZIP	Condition (S)	Race	ZIP	Condition (S)
Asian	0213*	hypertension	*	02139	short breath
Asian	0213*	short breath	*	02138	hypertension
Asian	0214*	hypertension	*	02141	chest pain
Asian	0214*	chest pain	*	02141	hypertension
Black	0213*	chest pain	*	02138	short breath
Black	0213*	obesity	*	02138	chest pain
White	0213*	short breath	*	02138	chest pain
White	0213*	chest pain	*	02139	chest pain
White	0213*	chest pain	*	02139	obesity

- What would be a countermeasure?
  Making sure that *QI* contains all possible attributes, including the Medical Condition

k-Anonymity

Complementary release attack

- Linking the randomized $\mathcal{D}_1^*$ and $\mathcal{D}_2^*$ on the Medical Condition violates the 2-anonymity on Q ={Race, ZIP}
- E.g. the rows `[Black, 02139, obesity]` and `[Asian, 02138, short breath]` are unique

Race	ZIP	Condition (S)	Race	ZIP	Condition (S)
Asian	0213*	hypertension	*	02139	short breath
Asian	0213*	short breath	*	02138	hypertension
Asian	0214*	hypertension	*	02141	chest pain
Asian	0214*	chest pain	*	02141	hypertension
Black	0213*	chest pain	*	02138	short breath
Black	0213*	obesity	*	02138	chest pain
White	0213*	short breath	*	02138	chest pain
White	0213*	chest pain	*	02139	chest pain
White	0213*	chest pain	*	02139	obesity

- What would be a countermeasure?
  Making sure that *QI* contains all possible attributes, including the Medical Condition
  i.e. for these databases Q ={Race, ZIP, Medical Condition}

k-Anonymity

Temporal attack

- The Complementary release attack was caused by releasing 2 anonymized versions of the same database $(\mathcal{D}_1^*, \mathcal{D}_2^*)$ and linking attributes outside the Q

k-Anonymity

Temporal attack

- The Complementary release attack was caused by releasing 2 anonymized versions of the same database $(\mathcal{D}_1^*, \mathcal{D}_2^*)$ and linking attributes outside the Q
- The database $\mathcal{D}$ may change overtime, with people registering/de-registering

k-Anonymity

Temporal attack

- The Complementary release attack was caused by releasing 2 anonymized versions of the same database $(\mathcal{D}_1^*, \mathcal{D}_2^*)$ and linking attributes outside the Q
- The database $\mathcal{D}$ may change overtime, with people registering/de-registering
- This again results in releasing 2 anonymized versions of database, with small differences across time

k-Anonymity

Temporal attack

- The Complementary release attack was caused by releasing 2 anonymized versions of the same database $(\mathcal{D}_1^*, \mathcal{D}_2^*)$ and linking attributes outside the Q
- The database $\mathcal{D}$ may change overtime, with people registering/de-registering
- This again results in releasing 2 anonymized versions of database, with small differences across time

$$\text{non-anonymized database at time } t_1 : \mathcal{D}_{t_1}$$

$$\text{non-anonymized database at time } t_2 : \mathcal{D}_{t_2} = \mathcal{D}_{t_1} \cup \{\text{new registrations}\}$$

k-Anonymity

Temporal attack

- The Complementary release attack was caused by releasing 2 anonymized versions of the same database $(\mathcal{D}_1^*, \mathcal{D}_2^*)$ and linking attributes outside the Q
- The database $\mathcal{D}$ may change overtime, with people registering/de-registering
- This again results in releasing 2 anonymized versions of database, with small differences across time

$$\text{non-anonymized database at time } t_1 : \mathcal{D}_{t_1}$$

$$\text{non-anonymized database at time } t_2 : \mathcal{D}_{t_2} = \mathcal{D}_{t_1} \cup \{\text{new registrations}\}$$

- At time t_1 you release the anonymized $\mathcal{D}_{t_1}^*$ and at time t_2 you release the anonymized $\mathcal{D}_{t_2}^*$

k-Anonymity

Temporal attack

- The Complementary release attack was caused by releasing 2 anonymized versions of the same database $(\mathcal{D}_1^*, \mathcal{D}_2^*)$ and linking attributes outside the Q
- The database $\mathcal{D}$ may change overtime, with people registering/de-registering
- This again results in releasing 2 anonymized versions of database, with small differences across time

$$\text{non-anonymized database at time } t_1 : \mathcal{D}_{t_1}$$

$$\text{non-anonymized database at time } t_2 : \mathcal{D}_{t_2} = \mathcal{D}_{t_1} \cup \{\text{new registrations}\}$$

- At time t_1 you release the anonymized $\mathcal{D}_{t_1}^*$ and at time t_2 you release the anonymized $\mathcal{D}_{t_2}^*$
- If the Q does not include all attributes, you are vulnerable to the Complementary release attack on $\mathcal{D}_{t_1}^*, \mathcal{D}_{t_2}^*$ (with small differences)

l-Diversity

l-Diversity

Homegeneity attack

- Let the following non-anonymized database $\mathcal{D}$ w.r.t. $Q = \{\text{ZIP, Age, Nationality}\}$

ZIP	Age	Nationality	Condition (S)
13053	52	American	Heart
13068	58	Korean	Heart
13068	53	American	Heart
13050	51	Mexican	Heart
14053	34	American	Cancer
14068	45	Japanese	Cancer
14053	31	Mexican	Heart
13050	30	American	Heart
13052	33	American	Heart

l-Diversity

Homegeneity attack

- Let the following non-anonymized database $\mathcal{D}$ w.r.t. $Q = \{\text{ZIP, Age, Nationality}\}$

ZIP	Age	Nationality	Condition (S)
13053	52	American	Heart
13068	58	Korean	Heart
13068	53	American	Heart
13050	51	Mexican	Heart
14053	34	American	Cancer
14068	45	Japanese	Cancer
14053	31	Mexican	Heart
13050	30	American	Heart
13052	33	American	Heart

- The first row of the database is Bob

$$\text{Bob} = \texttt{[13053, 52, American, Heart]}$$

l-Diversity

Homegeneity attack

- We construct database $\mathcal{D}^*$ that satisfies 2-anonymity w.r.t. $Q = \{$ZIP, Age, Nationality$\}$

ZIP	Age	Nationality	Condition (S)
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart

l-Diversity

Homegeneity attack

- We construct database $\mathcal{D}^*$ that satisfies 2-anonymity w.r.t. $Q = \{\text{ZIP, Age, Nationality}\}$

ZIP	Age	Nationality	Condition (S)
130**	≥ 50	*	Heart
130**	≥ 50	*	Heart
130**	≥ 50	*	Heart
130**	≥ 50	*	Heart
140**	< 50	*	Cancer
140**	< 50	*	Cancer
140**	< 50	*	Heart
130**	< 50	*	Heart
130**	< 50	*	Heart

- Where is Bob placed?

l-Diversity

Homegeneity attack

- We construct database $\mathcal{D}^*$ that satisfies 2-anonymity w.r.t. $Q = \{$ZIP, Age, Nationality$\}$

ZIP	Age	Nationality	Condition (S)
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart

- Where is Bob placed? In the first block

l-Diversity

Homegeneity attack

- We construct database $\mathcal{D}^*$ that satisfies 2-anonymity w.r.t. $Q = \{$ZIP, Age, Nationality$\}$

ZIP	Age	Nationality	Condition (S)
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart

- Where is Bob placed? In the first block
- Alice is Bob's neighbour and she knows the precise value q for his quasi-identifier Q

l-Diversity

Homegeneity attack

- We construct database $\mathcal{D}^*$ that satisfies 2-anonymity w.r.t. $Q = \{$ZIP, Age, Nationality$\}$

ZIP	Age	Nationality	Condition (S)
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart

- Where is Bob placed? In the first block
- Alice is Bob's neighbour and she knows the precise value q for his quasi-identifier Q

Alice knows that Bob=`[13053, 52, American, ?]`

l-Diversity

Homegeneity attack

ZIP	Age	Nationality	Condition (S)
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart

- Can Alice find the (sensitive) medical condition of Bob?

l-Diversity

Homegeneity attack

ZIP	Age	Nationality	Condition (S)
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart

- Can Alice find the (sensitive) medical condition of Bob?
  Yes! Bob suffers from a Heart condition

l-Diversity

Homegeneity attack

ZIP	Age	Nationality	Condition (S)
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart

- Can Alice find the (sensitive) medical condition of Bob?
  Yes! Bob suffers from a Heart condition
- Bob is placed in block 1, among 3 more people, he should be 4-anonymous!

l-Diversity

Homegeneity attack

ZIP	Age	Nationality	Condition (S)
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart

- Can Alice find the (sensitive) medical condition of Bob?
  Yes! Bob suffers from a Heart condition
- Bob is placed in block 1, among 3 more people, he should be 4-anonymous!
- However everyone in this block has the same condition! The homogeneity of the sensitive information breaks k-anonymity

l-Diversity

Background knowledge attack

- Let the again the same non-anonymized database $\mathcal{D}$ w.r.t. $Q = \{\text{ZIP, Age, Nationality}\}$

ZIP	Age	Nationality	Condition (S)
13053	52	American	Heart
13068	58	Korean	Heart
13068	53	American	Heart
13050	51	Mexican	Heart
14053	34	American	Cancer
14068	45	Japanese	Cancer
14053	31	Mexican	Heart
13050	30	American	Heart
13052	33	American	Heart

- The 6th row of the database is Trudy

$$\text{Trudy} = \texttt{[14068, 45, Japanese, Cancer]}$$

l-Diversity

Background knowledge attack

- We construct again database $\mathcal{D}^*$ that satisfies 2-anonymity

ZIP	Age	Nationality	Condition (S)
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart

- Where is Trudy placed?

l-Diversity

Background knowledge attack

- We construct again database $\mathcal{D}^*$ that satisfies 2-anonymity

ZIP	Age	Nationality	Condition (S)
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart

- Where is Trudy placed? In the 2nd block

l-Diversity

Background knowledge attack

- We construct again database $\mathcal{D}^*$ that satisfies 2-anonymity

ZIP	Age	Nationality	Condition (S)
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart

- Where is Trudy placed? In the 2nd block
- Is the Homogeneity attack applicable?

l-Diversity

Background knowledge attack

- We construct again database $\mathcal{D}^*$ that satisfies 2-anonymity

ZIP	Age	Nationality	Condition (S)
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart

- Where is Trudy placed? In the 2nd block
- Is the Homogeneity attack applicable?
  No! The sensitive information can be either Cancer or Heart

l-Diversity

Background knowledge attack

- We construct again database $\mathcal{D}^*$ that satisfies 2-anonymity

ZIP	Age	Nationality	Condition (S)
130**	≥ 50	*	Heart
130**	≥ 50	*	Heart
130**	≥ 50	*	Heart
130**	≥ 50	*	Heart
140**	< 50	*	Cancer
140**	< 50	*	Cancer
140**	< 50	*	Heart
130**	< 50	*	Heart
130**	< 50	*	Heart

- Where is Trudy placed? In the 2nd block
- Is the Homogeneity attack applicable?
  No! The sensitive information can be either Cancer or Heart
- Based on $\mathcal{D}^*$, which condition is more likely?

l-Diversity

Background knowledge attack

- We construct again database $\mathcal{D}^*$ that satisfies 2-anonymity

ZIP	Age	Nationality	Condition (S)
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart

- Where is Trudy placed? In the 2nd block
- Is the Homogeneity attack applicable?
  No! The sensitive information can be either Cancer or Heart
- Based on $\mathcal{D}^*$, which condition is more likely?
  Two out of three people in block 2 have Cancer

l-Diversity

Background knowledge attack

- We construct again database $\mathcal{D}^*$ that satisfies 2-anonymity

ZIP	Age	Nationality	Condition (S)
130**	≥ 50	*	Heart
130**	≥ 50	*	Heart
130**	≥ 50	*	Heart
130**	≥ 50	*	Heart
140**	< 50	*	Cancer
140**	< 50	*	Cancer
140**	< 50	*	Heart
130**	< 50	*	Heart
130**	< 50	*	Heart

- Where is Trudy placed? In the 2nd block
- Is the Homogeneity attack applicable?
  No! The sensitive information can be either Cancer or Heart
- Based on $\mathcal{D}^*$, which condition is more likely?
  Two out of three people in block 2 have Cancer
- Alice has background knowledge about Trudy: she knows that Trudy's family has a history of cancer-related diseases

l-Diversity

Background knowledge attack

- We construct again database $\mathcal{D}^*$ that satisfies 2-anonymity

ZIP	Age	Nationality	Condition (S)
130**	≥ 50	*	Heart
130**	≥ 50	*	Heart
130**	≥ 50	*	Heart
130**	≥ 50	*	Heart
140**	< 50	*	Cancer
140**	< 50	*	Cancer
140**	< 50	*	Heart
130**	< 50	*	Heart
130**	< 50	*	Heart

- Where is Trudy placed? In the 2nd block
- Is the Homogeneity attack applicable?
  No! The sensitive information can be either Cancer or Heart
- Based on $\mathcal{D}^*$, which condition is more likely?
  Two out of three people in block 2 have Cancer
- Alice has background knowledge about Trudy: she knows that Trudy's family has a history of cancer-related diseases
- The database $\mathcal{D}^*$ and the background knowledge make Alice very certain that Trudy's condition is Cancer

l-Diversity

Towards the concept of diversity

- Both the homogeneity attack and the background knowledge attack can break k-anonymity

l-Diversity

Towards the concept of diversity

- Both the homogeneity attack and the background knowledge attack can break k-anonymity
- The culprit is the lack of diversity in the sensitive attribute (Medical condition)

l-Diversity

Towards the concept of diversity

- Both the homogeneity attack and the background knowledge attack can break k-anonymity
- The culprit is the lack of diversity in the sensitive attribute (Medical condition)

- Both attacks are very plausible in a real-life scenario, thus we need to revisit our privacy definition

l-Diversity

Threat model

- Alice knows the complete joint distribution of the non-sensitive and the sensitive attributes

l-Diversity

Threat model

- Alice knows the complete joint distribution of the non-sensitive and the sensitive attributes

  E.g. she knows $P(\text{ZIP}, \text{Age}, \text{Nationality}, \text{Condition})$ for the entire US population

l-Diversity

Threat model

- Alice knows the complete joint distribution of the non-sensitive and the sensitive attributes

  E.g. she knows $P(\text{ZIP, Age, Nationality, Condition})$ for the entire US population
  Thus she can compute any conditional on the sensitive value

$$P(\text{Condition}|\ \text{ZIP, Age, Nationality}) = \frac{P(\text{ZIP, Age, Nationality, Condition})}{P(\text{ZIP, Age, Nationality})}$$

l-Diversity

Threat model

- Alice knows the complete joint distribution of the non-sensitive and the sensitive attributes

  E.g. she knows $P(\text{ZIP, Age, Nationality, Condition})$ for the entire US population
  Thus she can compute any conditional on the sensitive value

$$P(\text{Condition}|\ \text{ZIP, Age, Nationality}) = \frac{P(\text{ZIP, Age, Nationality, Condition})}{P(\text{ZIP, Age, Nationality})}$$

- Alice knows that Trudy is on database $\mathcal{D}^*$ and she knows the non-sensitive part of Trudy's record t

$$t = \texttt{[14068, 45, Japanese, ?]}$$

l-Diversity

Threat model

- Alice knows the complete joint distribution of the non-sensitive and the sensitive attributes

  E.g. she knows $P(\text{ZIP, Age, Nationality, Condition})$ for the entire US population
  Thus she can compute any conditional on the sensitive value

$$P(\text{Condition}|\ \text{ZIP, Age, Nationality}) = \frac{P(\text{ZIP, Age, Nationality, Condition})}{P(\text{ZIP, Age, Nationality})}$$

- Alice knows that Trudy is on database $\mathcal{D}^*$ and she knows the non-sensitive part of Trudy's record t

$$t = \texttt{[14068, 45, Japanese, ?]}$$

  Thus she can generalize t to the anonymized version t^*

$$t^* = \textit{anonymize}(t) = \texttt{[140**, <50, *, ?]}$$

l-Diversity

Threat model

- Alice knows the complete joint distribution of the non-sensitive and the sensitive attributes

  E.g. she knows $P(\text{ZIP, Age, Nationality, Condition})$ for the entire US population
  Thus she can compute any conditional on the sensitive value

$$P(\text{Condition}|\ \text{ZIP, Age, Nationality}) = \frac{P(\text{ZIP, Age, Nationality, Condition})}{P(\text{ZIP, Age, Nationality})}$$

- Alice knows that Trudy is on database $\mathcal{D}^*$ and she knows the non-sensitive part of Trudy's record t

$$t = \texttt{[14068, 45, Japanese, ?]}$$

  Thus she can generalize t to the anonymized version t^*

$$t^* = \textit{anonymize}(t) = \texttt{[140**, <50, *, ?]}$$

- Alice doesn't know the sensitive part of Trudy's record
  i.e. she doesn't know Trudy's Medical Condition

l-Diversity

Prior belief

- The database $\mathcal{D}$ has been anonymized into $\mathcal{D}^*$ w.r.t. quasi-identifier Q

$$Q = \{\text{ZIP, Age, Nationality}\}$$

l-Diversity

Prior belief

- The database $\mathcal{D}$ has been anonymized into $\mathcal{D}^*$ w.r.t. quasi-identifier Q

$$Q = \{\text{ZIP, Age, Nationality}\}$$

- Alice knows the non-sensitive part of record t of Trudy thus she can find the projection w.r.t. Q

$$t[Q] = q = \texttt{[14068, 45, Japanese]}$$

l-Diversity

Prior belief

- The database $\mathcal{D}$ has been anonymized into $\mathcal{D}^*$ w.r.t. quasi-identifier Q

$$Q = \{\text{ZIP, Age, Nationality}\}$$

- Alice knows the non-sensitive part of record t of Trudy thus she can find the projection w.r.t. Q

$$t[Q] = q = \texttt{[14068, 45, Japanese]}$$

- Alice knows the joint distribution thus she can compute α

$$\alpha_{(q,s)} = P(t[S] = s |\ t[Q] = q)$$

l-Diversity

Prior belief

- The database $\mathcal{D}$ has been anonymized into $\mathcal{D}^*$ w.r.t. quasi-identifier Q

$$Q = \{\text{ZIP, Age, Nationality}\}$$

- Alice knows the non-sensitive part of record t of Trudy thus she can find the projection w.r.t. Q

$$t[Q] = q = \texttt{[14068, 45, Japanese]}$$

- Alice knows the joint distribution thus she can compute α

$$\alpha_{(q,s)} = P(t[S] = s|\ t[Q] = q)$$

Setting $Q = \{\text{ZIP, Age, Nationality}\}$, $q =$ `[14068, 45, Japanese]` and $S =$Medical condition, we get:

$$\alpha_{(q,Cancer)} = P(\text{Cancer}|\ q) = 0.95$$

$$\alpha_{(q,Heart)} = P(\text{Heart}|\ q) = 0.05$$

l-Diversity

Prior belief

- The database $\mathcal{D}$ has been anonymized into $\mathcal{D}^*$ w.r.t. quasi-identifier Q

$$Q = \{\text{ZIP, Age, Nationality}\}$$

- Alice knows the non-sensitive part of record t of Trudy thus she can find the projection w.r.t. Q

$$t[Q] = q = \texttt{[14068, 45, Japanese]}$$

- Alice knows the joint distribution thus she can compute α

$$\alpha_{(q,s)} = P(t[S] = s|\ t[Q] = q)$$

Setting $Q = \{\text{ZIP, Age, Nationality}\}$, $q =$ `[14068, 45, Japanese]` and $S =$Medical condition, we get:

$$\alpha_{(q,Cancer)} = P(\text{Cancer}|\ q) = 0.95$$

$$\alpha_{(q,Heart)} = P(\text{Heart}|\ q) = 0.05$$

- Alice now has established the prior beliefs α on the 2 diseases w.r.t. Trudy

l-Diversity

Posterior belief

- Alice observes the anonymized database $\mathcal{D}^*$ and computes the anonymized projection of the record of Trudy w.r.t. quasi-identifier Q

$$t = \texttt{[14068, 45, Japanese, ?]}$$

$$t[Q] = q = \texttt{[14068, 45, Japanese]}$$

$$t^*[Q] = q^* = \texttt{[140**, <50, *]}$$

l-Diversity

Posterior belief

- Alice observes the anonymized database $\mathcal{D}^*$ and computes the anonymized projection of the record of Trudy w.r.t. quasi-identifier Q

$$t = \texttt{[14068, 45, Japanese, ?]}$$

$$t[Q] = q = \texttt{[14068, 45, Japanese]}$$

$$t^*[Q] = q^* = \texttt{[140**, <50, *]}$$

- Alice knows the joint distribution $P(t[\mathrm{S}] = s |\ t[Q] = q)$
- She can compute the anonymized joint distribution $P(t[\mathrm{S}] = s |\ t^*[Q] = q^*)$

l-Diversity

Posterior belief

- Alice observes the anonymized database $\mathcal{D}^*$ and computes the anonymized projection of the record of Trudy w.r.t. quasi-identifier Q

$$t = \texttt{[14068, 45, Japanese, ?]}$$

$$t[Q] = q = \texttt{[14068, 45, Japanese]}$$

$$t^*[Q] = q^* = \texttt{[140**, <50, *]}$$

- Alice knows the joint distribution $P(t[\mathrm{S}] = s |\ t[Q] = q)$
- She can compute the anonymized joint distribution $P(t[\mathrm{S}] = s |\ t^*[Q] = q^*)$
- Setting $Q = \{$ZIP, Age, Nationality$\}$, $q^* =$ `[140**, <50, *]` and $S =$Medical Condition, we get:

$$P(\text{Cancer} |\ q^*) = 0.6$$

$$P(\text{Heart} |\ q^*) = 0.4$$

l-Diversity

Posterior belief

- Alice observes the anonymized database $\mathcal{D}^*$ and computes the anonymized projection of the record of Trudy w.r.t. quasi-identifier Q

$$t = \texttt{[14068, 45, Japanese, ?]}$$

$$t[Q] = q = \texttt{[14068, 45, Japanese]}$$

$$t^*[Q] = q^* = \texttt{[140**, <50, *]}$$

- Alice knows the joint distribution $P(t[\mathrm{S}] = s|\ t[Q] = q)$
- She can compute the anonymized joint distribution $P(t[\mathrm{S}] = s|\ t^*[Q] = q^*)$
- Setting $Q = \{$ZIP, Age, Nationality$\}$, $q^* = $ `[140**, <50, *]` and $S =$Medical Condition, we get:

$$P(\text{Cancer}|\ q^*) = 0.6$$

$$P(\text{Heart}|\ q^*) = 0.4$$

- Note that the anonymization process directly affects these probabilities

l-Diversity

Posterior belief

- Alice computes the quantity $n_{(q^*,s)}$ on the anonymized table $\mathcal{D}^*$

$n_{(q^*,s)}$ = no of dataset rows with projection q^* and sensitive attribute s

ZIP	Age	Nationality	Condition (S)
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart

l-Diversity

Posterior belief

- Alice computes the quantity $n_{(q^*,s)}$ on the anonymized table $\mathcal{D}^*$

$$n_{(q^*,s)} = \text{no of dataset rows with projection } q^* \text{ and sensitive attribute } s$$

ZIP	Age	Nationality	Condition (S)
130**	≥ 50	*	Heart
130**	≥ 50	*	Heart
130**	≥ 50	*	Heart
130**	≥ 50	*	Heart
140**	< 50	*	Cancer
140**	< 50	*	Cancer
140**	< 50	*	Heart
130**	< 50	*	Heart
130**	< 50	*	Heart

- Setting $q^* =$ `[140**, <50, *]` takes us to the 2nd block

$$n_{(q^*,\text{Cancer})} = 2, \quad n_{(q^*,\text{Heart})} = 1$$

l-Diversity

Posterior belief

- Alice computes the posterior belief β w.r.t. the anonymized database $\mathcal{D}^*$, for $q = [14068, \text{Japanese}]$ and $s \in \{\text{Cancer, Heart}\}$

$$\beta_{(q,s,\mathcal{D}^*)} = P(t[S] = s |\ t[Q] = q \ \text{ and } \ \exists\, t^* \in \mathcal{D}^* \text{ such that } t \rightarrow t^*)$$

l-Diversity

Posterior belief

- Alice computes the posterior belief β w.r.t. the anonymized database $\mathcal{D}^*$, for $q = [14068, \text{Japanese}]$ and $s \in \{\text{Cancer, Heart}\}$

$$\beta_{(q,s,\mathcal{D}^*)} = P(t[S] = s |\ t[Q] = q \ \text{ and } \ \exists\, t^* \in \mathcal{D}^* \text{ such that } t \to t^*)$$

$$\beta_{(q,s,\mathcal{D}^*)} = \frac{n_{(q^*,s)} \dfrac{P(s|q)}{P(s|q^*)}}{\sum_{s' \in S} n_{(q^*,s')} \dfrac{P(s'|q)}{P(s'|q^*)}}$$

l-Diversity

Posterior belief

- E.g. computing the posterior β for $s =$Cancer yields:

$$\beta_{(q,\text{Cancer},\mathcal{D}^*)} = \frac{n_{(q^*,\text{Cancer})} \dfrac{P(\text{Cancer}|q)}{P(\text{Cancer}|q^*)}}{n_{(q^*,\text{Cancer})} \dfrac{P(\text{Cancer}|q)}{P(\text{Cancer}|q^*)} + n_{(q^*,\text{Heart})} \dfrac{P(\text{Heart}|q)}{P(\text{Heart}|q^*)}} =$$

l-Diversity

Posterior belief

- E.g. computing the posterior β for s =Cancer yields:

$$\beta_{(q,\text{Cancer},\mathcal{D}^*)} = \frac{n_{(q^*,\text{Cancer})}\frac{P(\text{Cancer}|q)}{P(\text{Cancer}|q^*)}}{n_{(q^*,\text{Cancer})}\frac{P(\text{Cancer}|q)}{P(\text{Cancer}|q^*)} + n_{(q^*,\text{Heart})}\frac{P(\text{Heart}|q)}{P(\text{Heart}|q^*)}} =$$

$$\frac{2 * \frac{0.95}{0.6}}{1 * \frac{0.05}{0.4} + 2 * \frac{0.95}{0.6}} \approx 0.96 \approx \text{ prior } \alpha_{(q,\text{Cancer})} = 0.95$$

l-Diversity

Posterior belief

- E.g. computing the posterior β for s =Cancer yields:

$$\beta_{(q,\text{Cancer},\mathcal{D}^*)} = \frac{n_{(q^*,\text{Cancer})}\frac{P(\text{Cancer}|q)}{P(\text{Cancer}|q^*)}}{n_{(q^*,\text{Cancer})}\frac{P(\text{Cancer}|q)}{P(\text{Cancer}|q^*)} + n_{(q^*,\text{Heart})}\frac{P(\text{Heart}|q)}{P(\text{Heart}|q^*)}} =$$

$$\frac{2 * \frac{0.95}{0.6}}{1 * \frac{0.05}{0.4} + 2 * \frac{0.95}{0.6}} \approx 0.96 \approx \text{ prior } \alpha_{(q,\text{Cancer})} = 0.95$$

- Computing the posterior β for s =Heart yields:

$$\beta_{(q,\text{Heart},\mathcal{D}^*)} \approx 0.04 < \text{ prior } \alpha_{(q,\text{Heart})} = 0.05$$

l-Diversity

Posterior belief

- E.g. computing the posterior β for $s =$Cancer yields:

$$\beta_{(q,\text{Cancer},\mathcal{D}^*)} = \frac{n_{(q^*,\text{Cancer})} \frac{P(\text{Cancer}|q)}{P(\text{Cancer}|q^*)}}{n_{(q^*,\text{Cancer})} \frac{P(\text{Cancer}|q)}{P(\text{Cancer}|q^*)} + n_{(q^*,\text{Heart})} \frac{P(\text{Heart}|q)}{P(\text{Heart}|q^*)}} =$$

$$\frac{2 * \frac{0.95}{0.6}}{1 * \frac{0.05}{0.4} + 2 * \frac{0.95}{0.6}} \approx 0.96 \approx \text{ prior } \alpha_{(q,\text{Cancer})} = 0.95$$

- Computing the posterior β for $s =$Heart yields:

$$\beta_{(q,\text{Heart},\mathcal{D}^*)} \approx 0.04 < \text{ prior } \alpha_{(q,\text{Heart})} = 0.05$$

- Observing the published database $\mathcal{D}^*$ can adjust the priors via the $n_{(q^*,\text{Cancer})}$ and $n_{(q^*,\text{Heart})}$

l-Diversity

Posterior belief

- E.g. computing the posterior β for $s =$Cancer yields:

$$\beta_{(q,\text{Cancer},\mathcal{D}^*)} = \frac{n_{(q^*,\text{Cancer})} \frac{P(\text{Cancer}|q)}{P(\text{Cancer}|q^*)}}{n_{(q^*,\text{Cancer})} \frac{P(\text{Cancer}|q)}{P(\text{Cancer}|q^*)} + n_{(q^*,\text{Heart})} \frac{P(\text{Heart}|q)}{P(\text{Heart}|q^*)}} =$$

$$\frac{2 * \frac{0.95}{0.6}}{1 * \frac{0.05}{0.4} + 2 * \frac{0.95}{0.6}} \approx 0.96 \approx \text{ prior } \alpha_{(q,\text{Cancer})} = 0.95$$

- Computing the posterior β for $s =$Heart yields:

$$\beta_{(q,\text{Heart},\mathcal{D}^*)} \approx 0.04 < \text{ prior } \alpha_{(q,\text{Heart})} = 0.05$$

- Observing the published database $\mathcal{D}^*$ can adjust the priors via the $n_{(q^*,\text{Cancer})}$ and $n_{(q^*,\text{Heart})}$
- A strong prior belief α can lead to large posterior β, regardless of anonymization

l-Diversity

Homogeneity attack

- What happens if $n_{(q^*,s')} << n_{(q^*,s)}$, for $s \neq s'$?
- E.g. in the following $\mathcal{D}^*$ let q the projection of Bob i.e. $q =$`[13053, 52, American]` and $q^* =$`[130**, >=50, *]`
- In block 1 we have $n_{(q^*,\text{Cancer})} = 0 << n_{(q^*,\text{Heart})} = 4$

ZIP	Age	Nationality	Condition (S)
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart

l-Diversity

Homogeneity attack

- What happens if $n_{(q^*,s')} << n_{(q^*,s)}$, for $s \neq s'$?
- E.g. in the following $\mathcal{D}^*$ let q the projection of Bob i.e. $q =$[13053, 52, American] and $q^* =$[130**, >=50, *]
- In block 1 we have $n_{(q^*,\text{Cancer})} = 0 << n_{(q^*,\text{Heart})} = 4$

ZIP	Age	Nationality	Condition (S)
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart

$$\beta_{(q,\text{Heart},\mathcal{D}^*)} = \frac{n_{(q^*,s)} \dfrac{P(s|q)}{P(s|q^*)}}{\sum_{s' \in S} n_{(q^*,s')} \dfrac{P(s'|q)}{P(s'|q^*)}} \approx \frac{n_{(q^*,\text{Heart})} \dfrac{P(\text{Heart}|q)}{P(\text{Heart}|q^*)}}{n_{(q^*,\text{Heart})} \dfrac{P(\text{Heart}|q)}{P(\text{Heart}|q^*)}} = 1$$

l-Diversity

Background knowledge attack

- What happens if $\frac{P(s'|q)}{P(s'|q^*)} = 0$, for some s'?
- E.g. in the following $\mathcal{D}^*$ let q the projection of Trudy i.e. $q =$[14068, 45, Japanese] and $q^* =$[140**, <50, *]
- Block 2 has both sensitive values, with $n_{(q^*,\text{Cancer})} = 2, n_{(q^*,\text{Heart})} = 1$

ZIP	Age	Nationality	Condition (S)
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart

l-Diversity

Background knowledge attack

- What happens if $\frac{P(s'|q)}{P(s'|q^*)} = 0$, for some s'?
- E.g. in the following $\mathcal{D}^*$ let q the projection of Trudy i.e. $q =$[14068, 45, Japanese] and $q^* =$[140**, <50, *]
- Block 2 has both sensitive values, with $n_{(q^*,\text{Cancer})} = 2, n_{(q^*,\text{Heart})} = 1$

ZIP	Age	Nationality	Condition (S)
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart

- However the prior information suggests that Trudy is very unlikely to have a Heart condition i.e. $P(\text{Heart}|q)/P(\text{Heart}|q^*) = 0$

$$\beta_{(q,\text{Cancer},\mathcal{D}^*)} = \frac{n_{(q^*,s)}\frac{P(s|q)}{P(s|q^*)}}{\sum_{s'\in S} n_{(q^*,s')}\frac{P(s'|q)}{P(s'|q^*)}} \approx \frac{n_{(q^*,\text{Cancer})}\frac{P(\text{Cancer}|q)}{P(\text{Cancer}|q^*)}}{n_{(q^*,\text{Cancer})}\frac{P(\text{Cancer}|q)}{P(\text{Cancer}|q^*)}} = 1$$

l-Diversity

Principle of l-Diversity

- Let a q^*-block in anonymized database $\mathcal{D}^*$
- The block is l-diverse if it contains at least l well-represented values of the sensitive attribute S

l-Diversity

Principle of l-Diversity

- Let a q^*-block in anonymized database $\mathcal{D}^*$
- The block is l-diverse if it contains at least l well-represented values of the sensitive attribute S

ZIP	Age	Nationality	Condition (S)
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart

- Block 1, with q^* =[130**, >=50] has l-diversity of 1

l-Diversity

Principle of l-Diversity

- Let a q^*-block in anonymized database $\mathcal{D}^*$
- The block is l-diverse if it contains at least l well-represented values of the sensitive attribute S

ZIP	Age	Nationality	Condition (S)
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart

- Block 1, with q^* =[130**, >=50] has l-diversity of 1
- Block 2, with q^* =[140**, <50] has l-diversity of 2

l-Diversity

Principle of l-Diversity

- Let a q^*-block in anonymized database $\mathcal{D}^*$
- The block is l-diverse if it contains at least l well-represented values of the sensitive attribute S

ZIP	Age	Nationality	Condition (S)
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
130**	$\geq$50	*	Heart
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Cancer
140**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart
130**	$<$ 50	*	Heart

- Block 1, with q^* =[130**, >=50] has l-diversity of 1
- Block 2, with q^* =[140**, <50] has l-diversity of 2
- Block 3, with q^* =[130**, <50] has l-diversity of 1

l-Diversity

Principle of l-Diversity

- Large values of diversity will ensure that the homogeneity and background knowledge attacks do not work
- Following the *l*-diversity principle you try to ensure:

$$\beta_{(q,s,\mathcal{D}^*)} \approx \alpha_{(q,s)}$$

l-Diversity

Principle of l-Diversity

- Large values of diversity will ensure that the homogeneity and background knowledge attacks do not work
- Following the *l*-diversity principle you try to ensure:

$$\beta_{(q,s,\mathcal{D}^*)} \approx \alpha_{(q,s)}$$

Entropy *l*-diversity

- A database is entropy *l*-diverse if for every q^*-block it holds that:

$$\sum_{s \in S} p_{(q^*,s)} \log_2(p_{(q^*,s)}) \geq \log_2(l)$$

$$\text{where } p_{(q^*,s)} = \frac{n_{(q^*,s)}}{\sum_{s' \in S} n_{(q^*,s')}}$$

Reconstruction Attacks

Reconstruction Attacks

- The US Census Bureau conducts the Census of Population and Housing, collecting sensitive data

Reconstruction Attacks

- The US Census Bureau conducts the Census of Population and Housing, collecting sensitive data
- So far we have seen how anonymity breaks when publishing the full database, but aggregate statistics should be fine, no?

Reconstruction Attacks

- The US Census Bureau conducts the Census of Population and Housing, collecting sensitive data
- So far we have seen how anonymity breaks when publishing the full database, but aggregate statistics should be fine, no?

		Age		
Statistic	**Group**	**Count**	**Median**	**Mean**
1A	Total Population	7	30	38
2A	Female	4	30	33.5
2B	Male	3	30	44
2C	Black or African American	4	51	48.5
2D	White	3	24	24
3A	Single Adults	(D)	(D)	(D)
3B	Married Adults	4	51	54
4A	Black or African American Female	3	36	36.7
4B	Black or African American Male	(D)	(D)	(D)
4C	White Male	(D)	(D)	(D)
4D	White Female	(D)	(D)	(D)
5A	Persons Under 5 Years	(D)	(D)	(D)
5B	Persons Under 18 Years	(D)	(D)	(D)
5C	Persons 64 Years or Over	(D)	(D)	(D)

Note: Married persons must be 15 or over

Reconstruction Attacks

- The US Census Bureau conducts the Census of Population and Housing, collecting sensitive data
- So far we have seen how anonymity breaks when publishing the full database, but aggregate statistics should be fine, no?

		Age		
Statistic	**Group**	**Count**	**Median**	**Mean**
1A	Total Population	7	30	38
2A	Female	4	30	33.5
2B	Male	3	30	44
2C	Black or African American	4	51	48.5
2D	White	3	24	24
3A	Single Adults	(D)	(D)	(D)
3B	Married Adults	4	51	54
4A	Black or African American Female	3	36	36.7
4B	Black or African American Male	(D)	(D)	(D)
4C	White Male	(D)	(D)	(D)
4D	White Female	(D)	(D)	(D)
5A	Persons Under 5 Years	(D)	(D)	(D)
5B	Persons Under 18 Years	(D)	(D)	(D)
5C	Persons 64 Years or Over	(D)	(D)	(D)

Note: Married persons must be 15 or over

- Statistical disclosure control demands that we don't publish statistics based on 2 people. Why?

Reconstruction Attacks

- The US Census Bureau conducts the Census of Population and Housing, collecting sensitive data
- So far we have seen how anonymity breaks when publishing the full database, but aggregate statistics should be fine, no?

		Age		
Statistic	**Group**	**Count**	**Median**	**Mean**
1A	Total Population	7	30	38
2A	Female	4	30	33.5
2B	Male	3	30	44
2C	Black or African American	4	51	48.5
2D	White	3	24	24
3A	Single Adults	(D)	(D)	(D)
3B	Married Adults	4	51	54
4A	Black or African American Female	3	36	36.7
4B	Black or African American Male	(D)	(D)	(D)
4C	White Male	(D)	(D)	(D)
4D	White Female	(D)	(D)	(D)
5A	Persons Under 5 Years	(D)	(D)	(D)
5B	Persons Under 18 Years	(D)	(D)	(D)
5C	Persons 64 Years or Over	(D)	(D)	(D)

Note: Married persons must be 15 or over

- Statistical disclosure control demands that we don't publish statistics based on 2 people. Why?
- Thus statistics with Count<3 are suppressed (D)

Reconstruction Attacks

- The US Census Bureau conducts the Census of Population and Housing, collecting sensitive data
- So far we have seen how anonymity breaks when publishing the full database, but aggregate statistics should be fine, no?

		Age		
Statistic	**Group**	**Count**	**Median**	**Mean**
1A	Total Population	7	30	38
2A	Female	4	30	33.5
2B	Male	3	30	44
2C	Black or African American	4	51	48.5
2D	White	3	24	24
3A	Single Adults	(D)	(D)	(D)
3B	Married Adults	4	51	54
4A	Black or African American Female	3	36	36.7
4B	Black or African American Male	(D)	(D)	(D)
4C	White Male	(D)	(D)	(D)
4D	White Female	(D)	(D)	(D)
5A	Persons Under 5 Years	(D)	(D)	(D)
5B	Persons Under 18 Years	(D)	(D)	(D)
5C	Persons 64 Years or Over	(D)	(D)	(D)

Note: Married persons must be 15 or over

- Statistical disclosure control demands that we don't publish statistics based on 2 people. Why?
- Thus statistics with Count<3 are suppressed (D)
- Can we publish statistic 2B, the mean and median age of 3 males?

Reconstruction Attacks

- Let the ages of 3 males be x_1, x_2, x_3

$$\text{Average } \bar{x} = 44 = \frac{1}{3}(x_1 + x_2 + x_3)$$

Reconstruction Attacks

- Let the ages of 3 males be x_1, x_2, x_3

$$\text{Average } \bar{x} = 44 = \frac{1}{3}(x_1 + x_2 + x_3)$$

- The oldest person in the world was 125 years old, thus:

$$0 \leq x_1 \leq x_2 \leq x_3 \leq 125$$

Reconstruction Attacks

- Let the ages of 3 males be x_1, x_2, x_3

$$\text{Average } \bar{x} = 44 = \frac{1}{3}(x_1 + x_2 + x_3)$$

- The oldest person in the world was 125 years old, thus:

$$0 \leq x_1 \leq x_2 \leq x_3 \leq 125$$

- The median is equal to 30, thus:

$$x_2 = 30$$

Reconstruction Attacks

- Let the ages of 3 males be x_1, x_2, x_3

$$\text{Average } \bar{x} = 44 = \frac{1}{3}(x_1 + x_2 + x_3)$$

- The oldest person in the world was 125 years old, thus:

$$0 \leq x_1 \leq x_2 \leq x_3 \leq 125$$

- The median is equal to 30, thus:

$$x_2 = 30$$

- Given these constraints we can reduce the search space a lot!

A	B	C	A	B	C	A	B	C
1	30	101	11	30	91	21	30	81
2	30	100	12	30	90	22	30	80
3	30	99	13	30	89	23	30	79
4	30	98	14	30	88	24	30	78
5	30	97	15	30	87	25	30	77
6	30	96	16	30	86	26	30	76
7	30	95	17	30	85	27	30	75
8	30	94	18	30	84	28	30	74
9	30	93	19	30	83	29	30	73
10	30	92	20	30	82	30	30	72

Reconstruction Attack

- The exist many statistical queries, increasing the number of constraints

Reconstruction Attack

- The exist many statistical queries, increasing the number of constraints
- You can convert statistical queries to Linear Programming (LP) constraints

$$\text{Find } \mathbf{x} = [x_1, x_2, \ldots, x_n] \text{ such that } A\mathbf{x} \leq \mathbf{b}$$

Reconstruction Attack

- The exist many statistical queries, increasing the number of constraints
- You can convert statistical queries to Linear Programming (LP) constraints

$$\text{Find } \mathbf{x} = [x_1, x_2, \ldots, x_n] \text{ such that } A\mathbf{x} \leq \mathbf{b}$$

$$\text{e.g. } x_1 + x_2 + x_3 = 44 \iff$$

$$\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} * \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 44$$

Reconstruction Attack

- The exist many statistical queries, increasing the number of constraints
- You can convert statistical queries to Linear Programming (LP) constraints

$$\text{Find } \mathbf{x} = [x_1, x_2, \ldots, x_n] \text{ such that } A\mathbf{x} \leq \mathbf{b}$$

$$\text{e.g. } x_1 + x_2 + x_3 = 44 \iff$$

$$\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} * \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 44$$

$$\text{e.g. } x_1, x_2, x_3 \leq 125 \iff$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \leq \begin{bmatrix} 125 \\ 125 \\ 125 \end{bmatrix}$$

Reconstruction Attack

- The exist many statistical queries, increasing the number of constraints
- You can convert statistical queries to Linear Programming (LP) constraints

$$\text{Find } \mathbf{x} = [x_1, x_2, \ldots, x_n] \text{ such that } A\mathbf{x} \leq \mathbf{b}$$

$$\text{e.g. } x_1 + x_2 + x_3 = 44 \iff$$

$$\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} * \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 44$$

$$\text{e.g. } x_1, x_2, x_3 \leq 125 \iff$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \leq \begin{bmatrix} 125 \\ 125 \\ 125 \end{bmatrix}$$

- LP solvers are potent and reconstruction attacks have been successful at reconstructing the data exactly for 46% of the US population, and allowing errors in age of ± 1 year, for 71% of the population

Differential Privacy

Differential Privacy

- Reconstruction attacks are able to recover datasets, even from noisy statistics

Differential Privacy

- Reconstruction attacks are able to recover datasets, even from noisy statistics
- We need a frame that can **guarantee** that releasing statistical information maintains privacy

Differential Privacy

- Reconstruction attacks are able to recover datasets, even from noisy statistics
- We need a frame that can **guarantee** that releasing statistical information maintains privacy

Differential Privacy

- Ensures protection against adversaries with arbitrary external information, including being intimately familiar with the individuals they are targeting

Differential Privacy

- Reconstruction attacks are able to recover datasets, even from noisy statistics
- We need a frame that can **guarantee** that releasing statistical information maintains privacy

Differential Privacy

- Ensures protection against adversaries with arbitrary external information, including being intimately familiar with the individuals they are targeting
- Does not restrict the computational strategy used by the adversary

Differential Privacy

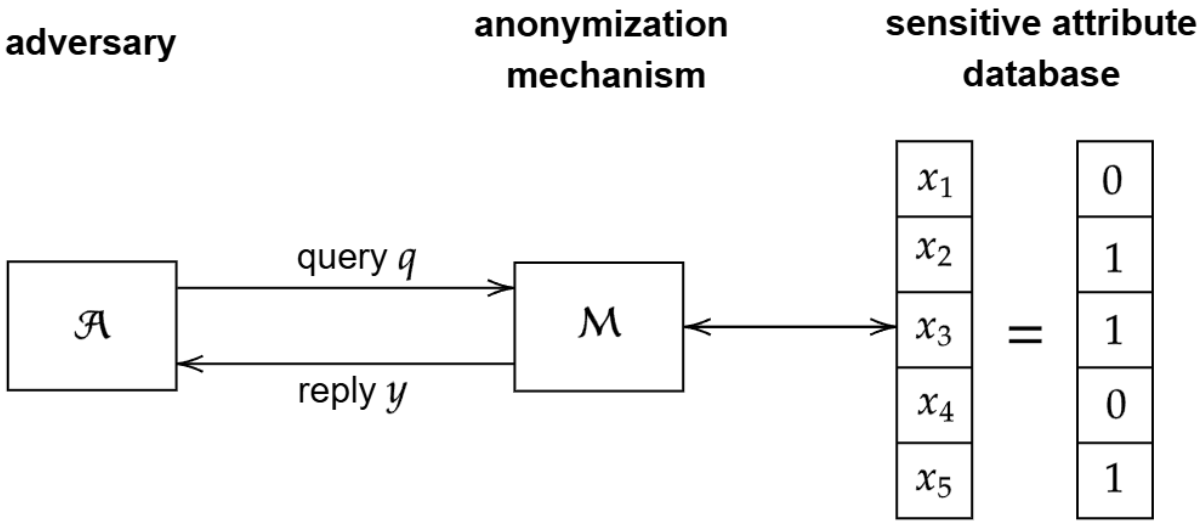
- Reconstruction attacks are able to recover datasets, even from noisy statistics
- We need a frame that can **guarantee** that releasing statistical information maintains privacy

Differential Privacy

- Ensures protection against adversaries with arbitrary external information, including being intimately familiar with the individuals they are targeting
- Does not restrict the computational strategy used by the adversary
- Provides a framework that quantifies how much statistical information is safe to release and with what accuracy

Differential Privacy

Setting



Differential Privacy

Setting

- A database $\mathcal{D}$ of n people stores a binary sensitive attribute

Differential Privacy

Setting

- A database $\mathcal{D}$ of n people stores a binary sensitive attribute
- The adversary $\mathcal{A}$ is a data analyst that submits query q to the dataset $\mathbf{x} = [x_1, x_2, \ldots, x_n]$

Differential Privacy

Setting

- A database $\mathcal{D}$ of n people stores a binary sensitive attribute
- The adversary $\mathcal{A}$ is a data analyst that submits query q to the dataset $\mathbf{x} = [x_1, x_2, \ldots, x_n]$
- The anonymization mechanism $\mathcal{M}$ adds noise to the accurate answer to query q and returns a distorted reply y

$$\mathcal{M}: \text{ distorted reply } y = \text{accurate answer to } q + \text{noise}$$

Differential Privacy

Setting

- A database $\mathcal{D}$ of n people stores a binary sensitive attribute
- The adversary $\mathcal{A}$ is a data analyst that submits query q to the dataset $\mathbf{x} = [x_1, x_2, \ldots, x_n]$
- The anonymization mechanism $\mathcal{M}$ adds noise to the accurate answer to query q and returns a distorted reply y

$$\mathcal{M}: \text{ distorted reply } y = \text{accurate answer to } q + \text{noise}$$

- We will consider a simple **average/proportion query** q

$$\text{accurate answer to } q = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\text{distorted reply } y = \mathcal{M}(x_1, x_2, \ldots, x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i + \text{noise}$$

Differential Privacy

Neighbouring datasets

- Let the following two datasets of $n = 5$ people

$$\mathbf{x} = [x_1, x_2, x_3, x_4, x_5], \quad \mathbf{x}' = [x_1, x_2', x_3, x_4, x_5]$$

$$\text{with } x_2 \neq x_2'$$

Differential Privacy

Neighbouring datasets

- Let the following two datasets of $n = 5$ people

$$\mathbf{x} = [x_1, x_2, x_3, x_4, x_5], \quad \mathbf{x}' = [x_1, x_2', x_3, x_4, x_5]$$

$$\text{with } x_2 \neq x_2'$$

- We say that two datasets are **neighbouring** if they differ only in a single datapoint x_j and write it as $\mathbf{x} \sim \mathbf{x}'$

Differential Privacy

Neighbouring datasets

- Let the following two datasets of $n = 5$ people

$$\mathbf{x} = [x_1, x_2, x_3, x_4, x_5], \quad \mathbf{x}' = [x_1, x_2', x_3, x_4, x_5]$$

$$\text{with } x_2 \neq x_2'$$

- We say that two datasets are **neighbouring** if they differ only in a single datapoint x_j and write it as $\mathbf{x} \sim \mathbf{x}'$

Differential Privacy

Neighbouring datasets

- Let the following two datasets of $n = 5$ people

$$\mathbf{x} = [x_1, x_2, x_3, x_4, x_5], \quad \mathbf{x}' = [x_1, x_2', x_3, x_4, x_5]$$

$$\text{with } x_2 \neq x_2'$$

- We say that two datasets are **neighbouring** if they differ only in a single datapoint x_j and write it as $\mathbf{x} \sim \mathbf{x}'$

Principle of differential privacy

- The goal of differential privacy is to ensure that the distorted replies y for neighbouring datasets $\mathbf{x}$ and $\mathbf{x}'$ are 'close enough'

  e.g. the distorted average of dataset $\mathbf{x}$ should be 'close enough' to the distorted average of dataset $\mathbf{x}'$

Differential Privacy

Neighbouring datasets

- Let the following two datasets of $n = 5$ people

$$\mathbf{x} = [x_1, x_2, x_3, x_4, x_5], \quad \mathbf{x}' = [x_1, x_2', x_3, x_4, x_5]$$

$$\text{with } x_2 \neq x_2'$$

- We say that two datasets are **neighbouring** if they differ only in a single datapoint x_j and write it as $\mathbf{x} \sim \mathbf{x}'$

Principle of differential privacy

- The goal of differential privacy is to ensure that the distorted replies y for neighbouring datasets $\mathbf{x}$ and $\mathbf{x}'$ are 'close enough'

  e.g. the distorted average of dataset $\mathbf{x}$ should be 'close enough' to the distorted average of dataset $\mathbf{x}'$

- A well-made anonymization mechanism $\mathcal{M}$ can achieve this property

Differential Privacy

Differential Privacy definition

- Let privacy parameter $\epsilon \geq 0$ and a query q to the database
- Then for all possible replies y and for all possible neighbouring datasets $(\mathbf{x}, \mathbf{x}')$ the following holds:

$$P(\mathcal{M}(\mathbf{x}, q) = y) \leq e^{\epsilon} P(\mathcal{M}(\mathbf{x}', q) = y)$$

Differential Privacy

Differential Privacy definition

- Let privacy parameter $\epsilon \geq 0$ and a query q to the database
- Then for all possible replies y and for all possible neighbouring datasets $(\mathbf{x}, \mathbf{x}')$ the following holds:

$$P(\mathcal{M}(\mathbf{x}, q) = y) \leq e^{\epsilon} P(\mathcal{M}(\mathbf{x}', q) = y)$$

- Small values of ϵ imply equal probabilities and strong privacy

Differential Privacy

Differential Privacy definition

- Let privacy parameter $\epsilon \geq 0$ and a query q to the database
- Then for all possible replies y and for all possible neighbouring datasets $(\mathbf{x}, \mathbf{x}')$ the following holds:

$$P(\mathcal{M}(\mathbf{x}, q) = y) \leq \ e^{\epsilon} \ P(\mathcal{M}(\mathbf{x}', q) = y)$$

- Small values of ϵ imply equal probabilities and strong privacy

Degenerate example

- Let $\epsilon = 0.1$ and let a naive mechanism $\mathcal{M}$ that outputs the same reply y, regardless of dataset $\mathbf{x}$ and query q

$$\mathcal{M}(\mathbf{x}, q) = 42 \ \text{ always}$$

Differential Privacy

Differential Privacy definition

- Let privacy parameter $\epsilon \geq 0$ and a query q to the database
- Then for all possible replies y and for all possible neighbouring datasets $(\mathbf{x}, \mathbf{x}')$ the following holds:

$$P(\mathcal{M}(\mathbf{x}, q) = y) \leq \ e^{\epsilon} \ P(\mathcal{M}(\mathbf{x}', q) = y)$$

- Small values of ϵ imply equal probabilities and strong privacy

Degenerate example

- Let $\epsilon = 0.1$ and let a naive mechanism $\mathcal{M}$ that outputs the same reply y, regardless of dataset $\mathbf{x}$ and query q

$$\mathcal{M}(\mathbf{x}, q) = 42 \ \text{ always}$$

- Now for any choice of datasets $(\mathbf{x}, \mathbf{x}')$ we have that:

$P(\mathcal{M}(\mathbf{x}, q) = 42) = P(\mathcal{M}(\mathbf{x}', q) = 42) = 1,$ and $y = 42$ is the only possible reply

Differential Privacy

Differential Privacy definition

- Let privacy parameter $\epsilon \geq 0$ and a query q to the database
- Then for all possible replies y and for all possible neighbouring datasets $(\mathbf{x}, \mathbf{x}')$ the following holds:

$$P(\mathcal{M}(\mathbf{x}, q) = y) \leq \ e^{\epsilon} \ P(\mathcal{M}(\mathbf{x}', q) = y)$$

- Small values of ϵ imply equal probabilities and strong privacy

Degenerate example

- Let $\epsilon = 0.1$ and let a naive mechanism $\mathcal{M}$ that outputs the same reply y, regardless of dataset $\mathbf{x}$ and query q

$$\mathcal{M}(\mathbf{x}, q) = 42 \ \text{ always}$$

- Now for any choice of datasets $(\mathbf{x}, \mathbf{x}')$ we have that:

$P(\mathcal{M}(\mathbf{x}, q) = 42) = P(\mathcal{M}(\mathbf{x}', q) = 42) = 1$, and $y = 42$ is the only possible reply

$$\frac{P(\mathcal{M}(\mathbf{x}, q) = 42)}{P(\mathcal{M}(\mathbf{x}', q) = 42)} = 1 \leq e^{0.1} = 1.1$$

Differential Privacy

Differential Privacy definition

- Let privacy parameter $\epsilon \geq 0$ and a query q to the database
- Then for all possible replies y and for all possible neighbouring datasets $(\mathbf{x}, \mathbf{x}')$ the following holds:

$$P(\mathcal{M}(\mathbf{x}, q) = y) \leq \ e^{\epsilon} \ P(\mathcal{M}(\mathbf{x}', q) = y)$$

- Small values of ϵ imply equal probabilities and strong privacy

Degenerate example

- Let $\epsilon = 0.1$ and let a naive mechanism $\mathcal{M}$ that outputs the same reply y, regardless of dataset $\mathbf{x}$ and query q

$$\mathcal{M}(\mathbf{x}, q) = 42 \ \text{ always}$$

- Now for any choice of datasets $(\mathbf{x}, \mathbf{x}')$ we have that:

  $P(\mathcal{M}(\mathbf{x}, q) = 42) = P(\mathcal{M}(\mathbf{x}', q) = 42) = 1$, and $y = 42$ is the only possible reply

$$\frac{P(\mathcal{M}(\mathbf{x}, q) = 42)}{P(\mathcal{M}(\mathbf{x}', q) = 42)} = 1 \leq e^{0.1} = 1.1$$

- Thus this naive $\mathcal{M}$ is differentially private (DP)

Differential Privacy

Randomized response

- *"Given an IntroSec exam, find out the percentage of students cheating"*
- Every student has a sensitive binary value x_i

$$x_i = \begin{cases} 0, & \text{no cheating} \\ 1, & \text{cheating} \end{cases}$$

Differential Privacy

Randomized response

- *"Given an IntroSec exam, find out the percentage of students cheating"*
- Every student has a sensitive binary value x_i

$$x_i = \begin{cases} 0, & \text{no cheating} \\ 1, & \text{cheating} \end{cases}$$

- The adversary/analyst wants to compute the proportion p

$$p = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Differential Privacy

Randomized response

- *"Given an IntroSec exam, find out the percentage of students cheating"*
- Every student has a sensitive binary value x_i

$$x_i = \begin{cases} 0, & \text{no cheating} \\ 1, & \text{cheating} \end{cases}$$

- The adversary/analyst wants to compute the proportion p

$$p = \frac{1}{n} \sum_{i=1}^{n} x_i$$

- We want to construct a mechanism $\mathcal{M}$ that converts the datapoint x_i of each student to a distorted reply y_i and check if it is DP

Differential Privacy

Randomized response

- Consider the following naive mechanism $\mathcal{M}$

$$y_i = \mathcal{M}(x_i) = \begin{cases} x_i, & \text{with probability } 1 \\ 1 - x_i, & \text{with probability } 0 \end{cases}$$

Differential Privacy

Randomized response

- Consider the following naive mechanism $\mathcal{M}$

$$y_i = \mathcal{M}(x_i) = \begin{cases} x_i, & \text{with probability } 1 \\ 1 - x_i, & \text{with probability } 0 \end{cases}$$

- This mechanism gives a perfectly accurate estimation $\hat{p}$ of the accurate proportion p

$$\hat{p} = \frac{1}{n} \sum_{i=1}^{n} y_i = \frac{1}{n} \sum_{i=1}^{n} x_i = p$$

Differential Privacy

Randomized response

- Consider the following naive mechanism $\mathcal{M}$

$$y_i = \mathcal{M}(x_i) = \begin{cases} x_i, & \text{with probability } 1 \\ 1 - x_i, & \text{with probability } 0 \end{cases}$$

- This mechanism gives a perfectly accurate estimation $\hat{p}$ of the accurate proportion p

$$\hat{p} = \frac{1}{n} \sum_{i=1}^{n} y_i = \frac{1}{n} \sum_{i=1}^{n} x_i = p$$

- However $y_i = x_i$ implies no privacy at all!

Differential Privacy

Randomized response

- Consider another mechanism $\mathcal{M}$

$$y_i = \mathcal{M}(x_i) = \begin{cases} x_i, & \text{with probability } 1/2 \\ 1 - x_i, & \text{with probability } 1/2 \end{cases}$$

Differential Privacy

Randomized response

- Consider another mechanism $\mathcal{M}$

$$y_i = \mathcal{M}(x_i) = \begin{cases} x_i, & \text{with probability } 1/2 \\ 1 - x_i, & \text{with probability } 1/2 \end{cases}$$

- This $\mathcal{M}$ is perfectly private since the distorted reply y_i is independent of x_i

$$P(y_i|x_i) = p(y_i) = 1/2$$

Differential Privacy

Randomized response

- Consider another mechanism $\mathcal{M}$

$$y_i = \mathcal{M}(x_i) = \begin{cases} x_i, & \text{with probability } 1/2 \\ 1 - x_i, & \text{with probability } 1/2 \end{cases}$$

- This $\mathcal{M}$ is perfectly private since the distorted reply y_i is independent of x_i

$$P(y_i|x_i) = p(y_i) = 1/2$$

- However, the estimate $\hat{p}$ of proportion p is entirely inaccurate!

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} y_i \text{ follows } \mathit{Binomial}(n, 1/2)$$

Differential Privacy

Randomized response

- Finally, consider an **in-between** mechanism $\mathcal{M}$ with $\epsilon \in (0, 1/2)$

$$y_i = \mathcal{M}(x_i) = \begin{cases} x_i, & \text{with probability } 1/2 + \epsilon \\ 1 - x_i, & \text{with probability } 1/2 - \epsilon \end{cases}$$

Differential Privacy

Randomized response

- Finally, consider an **in-between** mechanism $\mathcal{M}$ with $\epsilon \in (0, 1/2)$

$$y_i = \mathcal{M}(x_i) = \begin{cases} x_i, & \text{with probability } 1/2 + \epsilon \\ 1 - x_i, & \text{with probability } 1/2 - \epsilon \end{cases}$$

- This $\mathcal{M}$ strikes a tradeoff between privacy and accuracy

Differential Privacy

Randomized response

- Finally, consider an **in-between** mechanism $\mathcal{M}$ with $\epsilon \in (0, 1/2)$

$$y_i = \mathcal{M}(x_i) = \begin{cases} x_i, & \text{with probability } 1/2 + \epsilon \\ 1 - x_i, & \text{with probability } 1/2 - \epsilon \end{cases}$$

- This $\mathcal{M}$ strikes a tradeoff between privacy and accuracy
- The estimation error $|p - \hat{p}|$ when using randomized response is the following:

$$\text{error} = |p - \hat{p}| \leq O\left(\frac{1}{\epsilon\sqrt{n}}\right)$$

Differential Privacy

Randomized response

- Finally, consider an **in-between** mechanism $\mathcal{M}$ with $\epsilon \in (0, 1/2)$

$$y_i = \mathcal{M}(x_i) = \begin{cases} x_i, & \text{with probability } 1/2 + \epsilon \\ 1 - x_i, & \text{with probability } 1/2 - \epsilon \end{cases}$$

- This $\mathcal{M}$ strikes a tradeoff between privacy and accuracy
- The estimation error $|p - \hat{p}|$ when using randomized response is the following:

$$\text{error} = |p - \hat{p}| \leq O\left(\frac{1}{\epsilon\sqrt{n}}\right)$$

- As the population n grows large, the error goes to 0
- If the privacy parameter ϵ goes to 0, then we have stronger privacy
- If the privacy parameter ϵ goes to 0, then we have lower accuracy

Differential Privacy

Randomized response

- One can show that randomized response $\mathcal{M}$ is ϵ-DP for a query q that asks the for the average/proportion

Differential Privacy

Randomized response

- One can show that randomized response $\mathcal{M}$ is ϵ-DP for a query q that asks the for the average/proportion
- Let $\mathbf{y}$ the vector of distorted replies that mechanism $\mathcal{M}$ outputs when querying dataset $\mathbf{x}$ with $q =$ "average"

$$\mathbf{y} = [y_1, y_2, \ldots, y_n] = \mathcal{M}(x_1, x_2, \ldots, x_n, \quad q) = \mathcal{M}(\mathbf{x}, q)$$

Differential Privacy

Randomized response

- One can show that randomized response $\mathcal{M}$ is ϵ-DP for a query q that asks the for the average/proportion
- Let $\mathbf{y}$ the vector of distorted replies that mechanism $\mathcal{M}$ outputs when querying dataset $\mathbf{x}$ with $q =$ "average"

$$\mathbf{y} = [y_1, y_2, \ldots, y_n] = \mathcal{M}(x_1, x_2, \ldots, x_n, \quad q) = \mathcal{M}(\mathbf{x}, q)$$

- Let $\mathbf{y}'$ the vector of distorted replies that mechanism $\mathcal{M}$ outputs when querying dataset $\mathbf{x}'$ with $q =$ "average" and $\mathbf{x} \sim \mathbf{x}'$

$$\mathbf{y} = [y'_1, y'_2, \ldots, y'_n] = \mathcal{M}(x'_1, x'_2, \ldots, x'_n, \quad q) = \mathcal{M}(\mathbf{x}', q)$$

Differential Privacy

Randomized response

- One can show that randomized response $\mathcal{M}$ is ϵ-DP for a query q that asks the for the average/proportion
- Let $\mathbf{y}$ the vector of distorted replies that mechanism $\mathcal{M}$ outputs when querying dataset $\mathbf{x}$ with $q =$ "average"

$$\mathbf{y} = [y_1, y_2, \ldots, y_n] = \mathcal{M}(x_1, x_2, \ldots, x_n, \quad q) = \mathcal{M}(\mathbf{x}, q)$$

- Let $\mathbf{y}'$ the vector of distorted replies that mechanism $\mathcal{M}$ outputs when querying dataset $\mathbf{x}'$ with $q =$ "average" and $\mathbf{x} \sim \mathbf{x}'$

$$\mathbf{y} = [y_1', y_2', \ldots, y_n'] = \mathcal{M}(x_1', x_2', \ldots, x_n', \quad q) = \mathcal{M}(\mathbf{x}', q)$$

- Since $\mathbf{x}$ and $\mathbf{x}'$ are neighbouring datasets, there exists a single datapoint x_j such that $x_j \neq x_j'$

$$\frac{P(\mathcal{M}(\mathbf{x}, q) = \mathbf{y})}{P(\mathcal{M}(\mathbf{x}', q) = \mathbf{y})} = \frac{\prod_{i=1}^{n} P(y_i)}{\prod_{i=1}^{n} P(y_i)} = \frac{P(x_j)}{P(x_j')} \leq \frac{1/2 + \epsilon}{1/2 - \epsilon} \leq e^{O(\epsilon)}$$

Differential Privacy

Randomized response

- We constructed bounds on the estimation error and on privacy for the randomized response mechanism, w.r.t. averaging queries

$$\text{error(randomized response)} \leq O\left(\frac{1}{\epsilon\sqrt{n}}\right)$$

$$\frac{P(\mathcal{M}(\mathbf{x}, q) = \mathbf{y})}{P(\mathcal{M}(\mathbf{x}', q) = \mathbf{y})} \leq e^{O(\epsilon)}$$

Differential Privacy

Randomized response

- We constructed bounds on the estimation error and on privacy for the randomized response mechanism, w.r.t. averaging queries

$$\text{error(randomized response)} \leq O\left(\frac{1}{\epsilon\sqrt{n}}\right)$$

$$\frac{P(\mathcal{M}(\mathbf{x}, q) = \mathbf{y})}{P(\mathcal{M}(\mathbf{x}', q) = \mathbf{y})} \leq e^{O(\epsilon)}$$

- Now we can set the privacy parameter ϵ in a way that it controls the accuracy-privacy tradeoff

Differential Privacy

Laplace mechanism

- The randomized response is limited to binary/categorical data and we want a more general mechanism to achieve DP

Differential Privacy

Laplace mechanism

- The randomized response is limited to binary/categorical data and we want a more general mechanism to achieve DP
- The Laplace mechanism adds specific random noise to the answer in order to distort the reply

Differential Privacy

Laplace mechanism

- The randomized response is limited to binary/categorical data and we want a more general mechanism to achieve DP
- The Laplace mechanism adds specific random noise to the answer in order to distort the reply

Query function f

- Let function $f : \mathcal{X}^n \rightarrow \mathbb{R}^k$ that answers the database query
- E.g. when query is an average across dataset $\mathbf{x} = [x_1, x_2, \ldots, x_n]$ we have:

$$f(\mathbf{x}) = f(x_1, x_2, \ldots, x_n) = \frac{1}{n} \sum_{i=1}^{n} x_i$$

with $\mathcal{X}^n$ the set of all possible data values and $k = 1$

Differential Privacy

l_1-sensitivity

- Given a function $f : \mathcal{X}^n \to \mathbb{R}^k$, the l_1-sensitivity of f is defined as:

$$l_1\text{-sensitivity}(f) = \Delta^f = max_{\mathbf{x},\mathbf{x}'} ||f(\mathbf{x}) - f(\mathbf{x}')||_1, \qquad \text{with } \mathbf{x} \sim \mathbf{x}'$$

Differential Privacy

l_1-sensitivity

- Given a function $f : \mathcal{X}^n \to \mathbb{R}^k$, the l_1-sensitivity of f is defined as:

$$l_1\text{-sensitivity}(f) = \Delta^f = max_{\mathbf{x},\mathbf{x}'} ||f(\mathbf{x}) - f(\mathbf{x}')||_1, \qquad \text{with } \mathbf{x} \sim \mathbf{x}'$$

- Compute Δ^f when f corresponds to the "average" query

Differential Privacy

l_1-sensitivity

- Given a function $f : \mathcal{X}^n \to \mathbb{R}^k$, the l_1-sensitivity of f is defined as:

$$l_1\text{-sensitivity}(f) = \Delta^f = max_{\mathbf{x},\mathbf{x}'} ||f(\mathbf{x}) - f(\mathbf{x}')||_1, \qquad \text{with } \mathbf{x} \sim \mathbf{x}'$$

- Compute Δ^f when f corresponds to the "average" query

  Since $\mathbf{x} \sim \mathbf{x}'$ they differ at most at between a single datapoint x_j and x'_j

Differential Privacy

l_1-sensitivity

- Given a function $f : \mathcal{X}^n \to \mathbb{R}^k$, the l_1-sensitivity of f is defined as:

$$l_1\text{-sensitivity}(f) = \Delta^f = max_{\mathbf{x},\mathbf{x}'} ||f(\mathbf{x}) - f(\mathbf{x}')||_1, \qquad \text{with } \mathbf{x} \sim \mathbf{x}'$$

- Compute Δ^f when f corresponds to the "average" query

  Since $\mathbf{x} \sim \mathbf{x}'$ they differ at most at between a single datapoint x_j and x'_j

  Since $x_j, x'_j \in \{0, 1\}$ we have:

$$f(\mathbf{x}) - f(\mathbf{x}') = \frac{1}{n}\sum_{i=1}^{n} x_i - \frac{1}{n}\sum_{i=1}^{n} x'_i = \begin{cases} 0, & \text{if } x_j = x'_j \\ -1, & \text{if } x_j = 0, x'_j = 1 \\ 1, & \text{if } x_j = 1, x'_j = 0 \end{cases}$$

Differential Privacy

l_1-sensitivity

- Given a function $f : \mathcal{X}^n \to \mathbb{R}^k$, the l_1-sensitivity of f is defined as:

$$l_1\text{-sensitivity}(f) = \Delta^f = max_{\mathbf{x},\mathbf{x}'} ||f(\mathbf{x}) - f(\mathbf{x}')||_1, \qquad \text{with } \mathbf{x} \sim \mathbf{x}'$$

- Compute Δ^f when f corresponds to the "average" query

  Since $\mathbf{x} \sim \mathbf{x}'$ they differ at most at between a single datapoint x_j and x'_j

  Since $x_j, x'_j \in \{0, 1\}$ we have:

$$f(\mathbf{x}) - f(\mathbf{x}') = \frac{1}{n}\sum_{i=1}^{n} x_i - \frac{1}{n}\sum_{i=1}^{n} x'_i = \begin{cases} 0, & \text{if } x_j = x'_j \\ -1, & \text{if } x_j = 0, x'_j = 1 \\ 1, & \text{if } x_j = 1, x'_j = 0 \end{cases}$$

$$\Delta^f = max(\{|0|, |-1|, |1|\}) = 1$$

Differential Privacy

l_1-sensitivity

- Given a function $f : \mathcal{X}^n \to \mathbb{R}^k$, the l_1-sensitivity of f is defined as:

$$l_1\text{-sensitivity}(f) = \Delta^f = max_{\mathbf{x},\mathbf{x}'} ||f(\mathbf{x}) - f(\mathbf{x}')||_1, \qquad \text{with } \mathbf{x} \sim \mathbf{x}'$$

- Compute Δ^f when f corresponds to the "average" query

  Since $\mathbf{x} \sim \mathbf{x}'$ they differ at most at between a single datapoint x_j and x'_j

  Since $x_j, x'_j \in \{0, 1\}$ we have:

$$f(\mathbf{x}) - f(\mathbf{x}') = \frac{1}{n}\sum_{i=1}^{n} x_i - \frac{1}{n}\sum_{i=1}^{n} x'_i = \begin{cases} 0, & \text{if } x_j = x'_j \\ -1, & \text{if } x_j = 0, x'_j = 1 \\ 1, & \text{if } x_j = 1, x'_j = 0 \end{cases}$$

$$\Delta^f = max(\{|0|, |-1|, |1|\}) = 1$$

- The sensitivity Δ^f shows how much does a function change when we alter the dataset $\mathbf{x}$ by one datapoint

Differential Privacy

Laplace distribution

- The Laplace distribution with location μ equal to 0 and scale equal to b is a probability distribution with the following density function

$$pdf(x) = \frac{1}{2b} e^{(-|x|/b)}$$

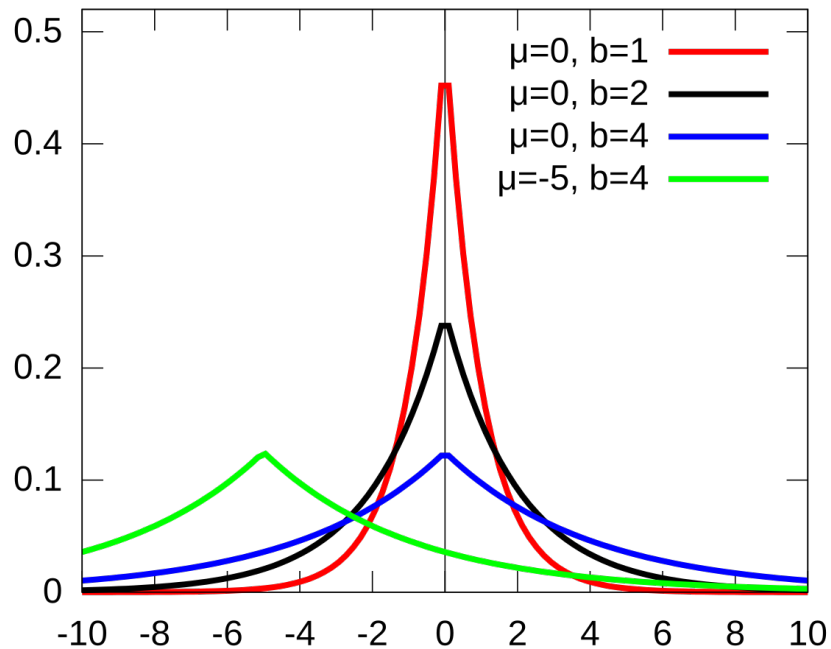
Differential Privacy

Laplace distribution

- The Laplace distribution with location μ equal to 0 and scale equal to b is a probability distribution with the following density function

$$pdf(x) = \frac{1}{2b} e^{(-|x|/b)}$$

- We will sample "noise" from this distribution to distort the query answer



Differential Privacy

Laplace mechanism

- Given a function $f : \mathcal{X}^n \to \mathbb{R}^k$, the Laplace mechanism described as:

$$\mathbf{y} = \mathcal{M}(\mathbf{x}) = f(\mathbf{x}) + [S_1, S_2, \ldots, S_k]$$

where S_i sampled from distribution Laplace$(\mu = 0, b = \Delta^f/\epsilon)$

Differential Privacy

Laplace mechanism

- Given a function $f : \mathcal{X}^n \to \mathbb{R}^k$, the Laplace mechanism described as:

$$\mathbf{y} = \mathcal{M}(\mathbf{x}) = f(\mathbf{x}) + [S_1, S_2, \ldots, S_k]$$

where S_i sampled from distribution Laplace$(\mu = 0, b = \Delta^f/\epsilon)$

- When f is the average function we have:

$$y = \mathcal{M}(\mathbf{x}) = f(\mathbf{x}) + \text{Laplace}(\mu = 0, b = \Delta^f/\epsilon) =$$

$$\frac{1}{n}\sum_{i=1}^{n} x_i + \text{Laplace}(\mu = 0, b = 1/\epsilon)$$

Differential Privacy

Laplace mechanism

- Given a function $f : \mathcal{X}^n \to \mathbb{R}^k$, the Laplace mechanism described as:

$$\mathbf{y} = \mathcal{M}(\mathbf{x}) = f(\mathbf{x}) + [S_1, S_2, \ldots, S_k]$$

where S_i sampled from distribution Laplace$(\mu = 0, b = \Delta^f/\epsilon)$

- When f is the average function we have:

$$y = \mathcal{M}(\mathbf{x}) = f(\mathbf{x}) + \text{Laplace}(\mu = 0, b = \Delta^f/\epsilon) =$$

$$\frac{1}{n}\sum_{i=1}^{n} x_i + \text{Laplace}(\mu = 0, b = 1/\epsilon)$$

- The estimation error $|p - \hat{p}|$ when using the Laplace mechanism and f is the "average" is the following:

$$\text{error} = |p - \hat{p}| \leq O\left(\frac{1}{\epsilon n}\right)$$

Differential Privacy

Laplace mechanism

- Given a function $f : \mathcal{X}^n \to \mathbb{R}^k$, the Laplace mechanism described as:

$$\mathbf{y} = \mathcal{M}(\mathbf{x}) = f(\mathbf{x}) + [S_1, S_2, \ldots, S_k]$$

where S_i sampled from distribution Laplace$(\mu = 0, b = \Delta^f/\epsilon)$

- When f is the average function we have:

$$y = \mathcal{M}(\mathbf{x}) = f(\mathbf{x}) + \text{Laplace}(\mu = 0, b = \Delta^f/\epsilon) =$$

$$\frac{1}{n}\sum_{i=1}^{n} x_i + \text{Laplace}(\mu = 0, b = 1/\epsilon)$$

- The estimation error $|p - \hat{p}|$ when using the Laplace mechanism and f is the "average" is the following:

$$\text{error} = |p - \hat{p}| \leq O\left(\frac{1}{\epsilon n}\right)$$

- Notice that the error of the Laplace mechanism is smaller than the error of the randomized response

$$\text{error(Laplace, } f) \leq O\left(\frac{1}{\epsilon n}\right) \leq \text{error(randomized response, } f) \leq O\left(\frac{1}{\epsilon\sqrt{n}}\right)$$

Differential Privacy

Various private queries

- So far we anonymized the average/proportion query
- *"Given an IntroSec exam, find out the percentage of students cheating"*

Differential Privacy

Various private queries

- So far we anonymized the average/proportion query
- *"Given an IntroSec exam, find out the percentage of students cheating"*

$$f(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \Delta^f = 1$$

Differential Privacy

Various private queries

- So far we anonymized the average/proportion query
- *"Given an IntroSec exam, find out the percentage of students cheating"*

$$f(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \Delta^f = 1$$

Laplace mechanism: $\mathcal{M}(\mathbf{x}) = f(\mathbf{x}) + \text{Laplace}(\mu = 0, b = 1/\epsilon)$

Differential Privacy

Various private queries

- So far we anonymized the average/proportion query
- *"Given an IntroSec exam, find out the percentage of students cheating"*

$$f(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \Delta^f = 1$$

$$\text{Laplace mechanism:} \quad \mathcal{M}(\mathbf{x}) = f(\mathbf{x}) + \text{Laplace}(\mu = 0, b = 1/\epsilon)$$

- *How many people in the dataset have a certain property?*

$$f(\mathbf{x}) = \sum_{i=1}^{n} x_i, \quad \Delta^f = 1$$

Differential Privacy

Various private queries

- So far we anonymized the average/proportion query
- *"Given an IntroSec exam, find out the percentage of students cheating"*

$$f(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \Delta^f = 1$$

$$\text{Laplace mechanism:} \quad \mathcal{M}(\mathbf{x}) = f(\mathbf{x}) + \text{Laplace}(\mu = 0, b = 1/\epsilon)$$

- *How many people in the dataset have a certain property?*

$$f(\mathbf{x}) = \sum_{i=1}^{n} x_i, \quad \Delta^f = 1$$

$$\text{Laplace mechanism:} \quad \mathcal{M}(\mathbf{x}) = f(\mathbf{x}) + \text{Laplace}(\mu = 0, b = 1/\epsilon)$$

Differential Privacy

Various private queries

- Ask k different queries of the form:
  How many people in the dataset have a certain property?

$$f(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), , \ldots, f_k(\mathbf{x})]$$

where f_i counts people with the property and $\Delta^f = k$

Differential Privacy

Various private queries

- Ask k different queries of the form:
  How many people in the dataset have a certain property?

$$f(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), , \ldots, f_k(\mathbf{x})]$$

where f_i counts people with the property and $\Delta^f = k$

Laplace mechanism: $\mathcal{M}(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), , \ldots, f_k(\mathbf{x})] + [S_1, S_2, \ldots, S_k]$

where S_i is sampled from Laplace$(\mu = 0, b = k/\epsilon)$, $1 \leq i \leq k$

Differential Privacy

Various private queries

- Ask k different queries of the form:
  How many people in the dataset have a certain property?

$$f(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), , \ldots, f_k(\mathbf{x})]$$

where f_i counts people with the property and $\Delta^f = k$

Laplace mechanism: $\mathcal{M}(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), , \ldots, f_k(\mathbf{x})] + [S_1, S_2, \ldots, S_k]$

where S_i is sampled from Laplace$(\mu = 0, b = k/\epsilon)$, $1 \leq i \leq k$

- Ask a histogram question:
  "Create the histogram of all ages in the dataset"

$$f(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), , \ldots, f_k(\mathbf{x})]$$

where f_i counts people with age $1 \leq i \leq k$ and $\Delta^f = 2$

Differential Privacy

Various private queries

- Ask k different queries of the form:
  How many people in the dataset have a certain property?

$$f(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), , \ldots, f_k(\mathbf{x})]$$

where f_i counts people with the property and $\Delta^f = k$

$$\text{Laplace mechanism:} \quad \mathcal{M}(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), , \ldots, f_k(\mathbf{x})] + [S_1, S_2, \ldots, S_k]$$

where S_i is sampled from Laplace$(\mu = 0, b = k/\epsilon)$, $\quad 1 \leq i \leq k$

- Ask a histogram question:
  "Create the histogram of all ages in the dataset"

$$f(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), , \ldots, f_k(\mathbf{x})]$$

where f_i counts people with age $1 \leq i \leq k$ and $\Delta^f = 2$

$$\text{Laplace mechanism:} \quad \mathcal{M}(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), , \ldots, f_k(\mathbf{x})] + [S_1, S_2, \ldots, S_k]$$

where S_i is sampled from Laplace$(\mu = 0, b = 2/\epsilon)$

Differential Privacy

How should you set the privacy parameter ϵ?

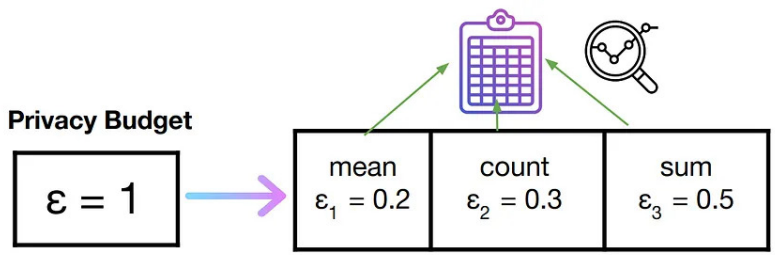
- Empirically, depending on the statistical accuracy needed
  e.g. given an estimation error, find the minimum ϵ that achieves it

Differential Privacy

How should you set the privacy parameter ϵ?

- Empirically, depending on the statistical accuracy needed

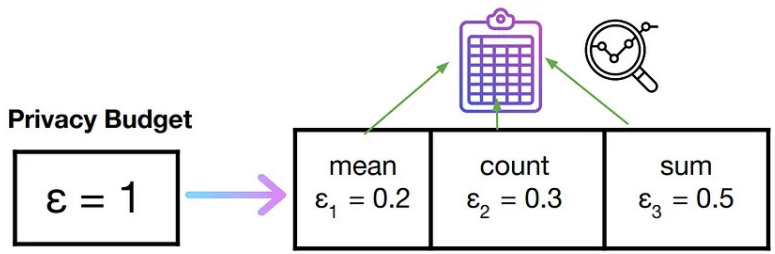
  e.g. given an estimation error, find the minimum ϵ that achieves it
- Privacy 'budget' for statistical operations



Differential Privacy

How should you set the privacy parameter ϵ?

- Empirically, depending on the statistical accuracy needed
  e.g. given an estimation error, find the minimum ϵ that achieves it
- Privacy 'budget' for statistical operations



- Assign ϵ based on the application context: find below many use-cases
  `https://desfontain.es/blog/real-world-differential-privacy.html`