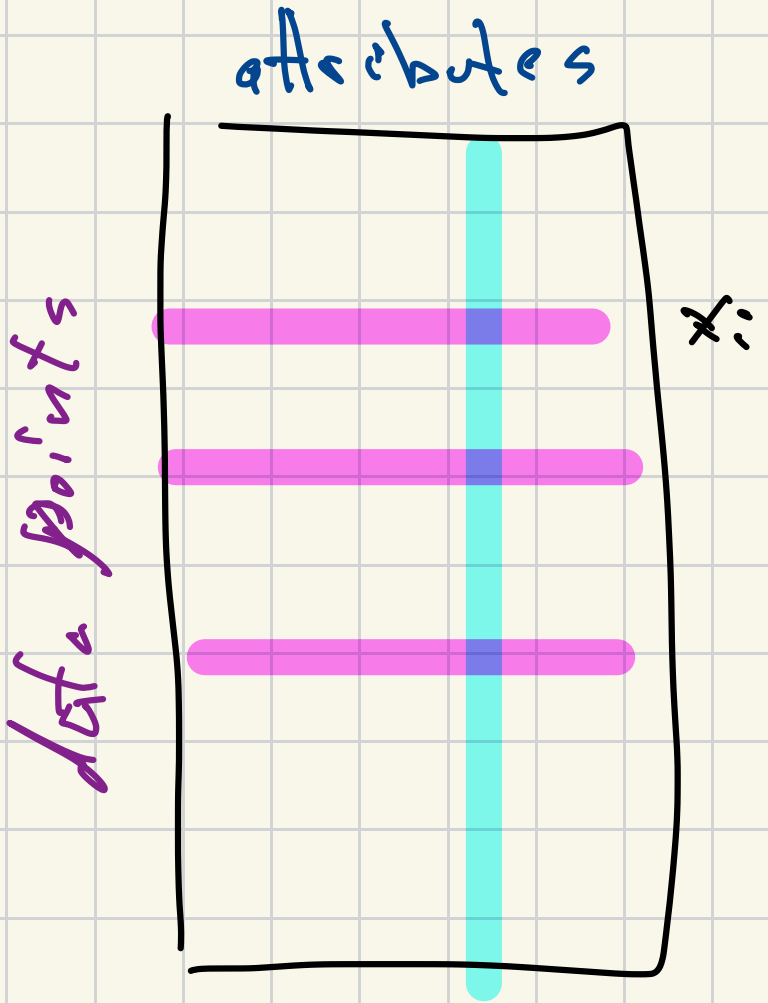
L22: (Differential) Privacy

Jeff M. Phillips

November 17, 2025

Data X



What if this was you?

Ethics ≈ Empathy

Privacy and Data
- who gets access?
- to what?

Dawn of Data Science

→ Early 2000s

big companies

collect lots data about users

(thru web)

↳ Release sample data online

→ let academia figure out.

Late 2000s this stopped.

Example: Heath Records

Story Time:

Example: Heath Records

Story Time:

- In 2000, Massachusetts released all State employee's medical records in an effort for researchers to be able to study them.

Example: Heath Records

Story Time:

- In 2000, Massachusetts released all State employee's medical records in an effort for researchers to be able to study them.
- They wiped all ids, but kept zip codes, birthday, gender. Was declared anonymized by the government.

Example: Heath Records

Story Time:

- In 2000, Massachusetts released all State employee's medical records in an effort for researchers to be able to study them.
- They wiped all ids, but kept zip codes, birthday, gender. Was declared anonymized by the government.
- In Massachusetts, it was possible to buy voter data for $20. It has names, zip codes, and gender of all voters.

Example: Heath Records

Story Time:

- In 2000, Massachusetts released all State employee's medical records in an effort for researchers to be able to study them.
- They wiped all ids, but kept zip codes, birthday, gender. Was declared anonymized by the government.
- In Massachusetts, it was possible to buy voter data for $20. It has names, zip codes, and gender of all voters.
- A (then) grad student, Latanya Sweeney combined the two to identify the governor of Massachusetts. Story is, she mailed him his own health records!

Example: Heath Records

Story Time:

- In 2000, Massachusetts released all State employee's medical records in an effort for researchers to be able to study them.
- They wiped all ids, but kept zip codes, birthday, gender. Was declared anonymized by the government.
- In Massachusetts, it was possible to buy voter data for $20. It has names, zip codes, and gender of all voters.
- A (then) grad student, Latanya Sweeney combined the two to identify the governor of Massachusetts. Story is, she mailed him his own health records!
- **Dr.** Sweeney now teaches at Harvard.

How to release data anonymously?

- to keep anonymity individuals
- be able to make inference

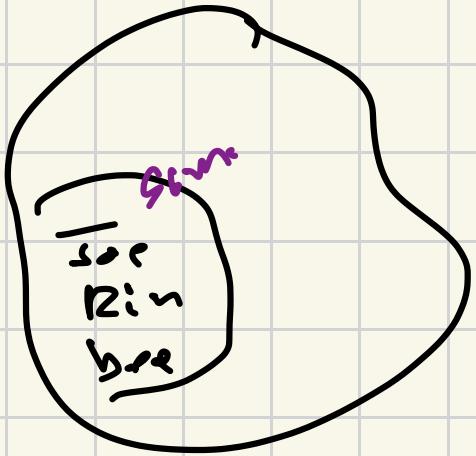
---

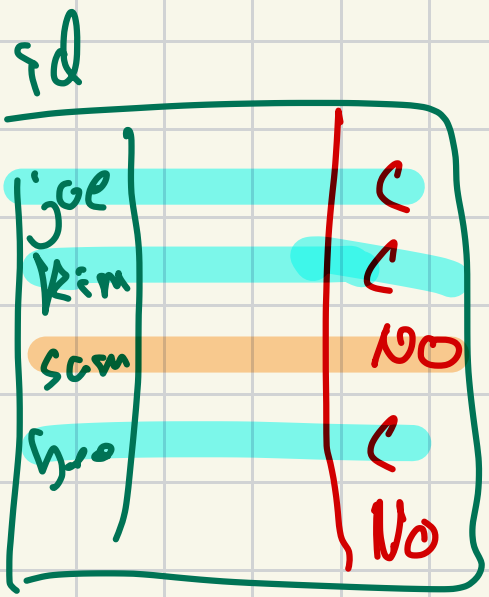
k-anonymity: Remove information from records, until each combination of attributes matches at least k individuals.

Dividing Data

public attributes

- Zip code
- age
- gender





private attributes

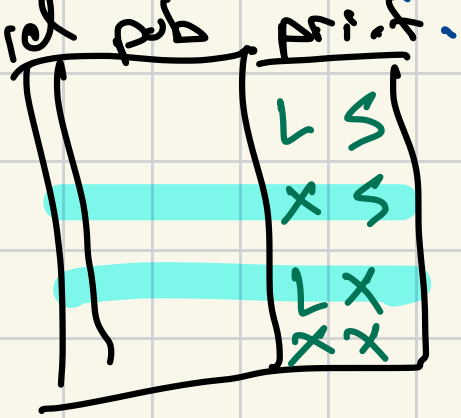
- has cancer
- has diabetes
- search history

l-diversity: k-anonymity ⊕

each group has diversity in private values

What if multiple private attributes?

- lung cancer
- skin cancer



---

t-closeness ⊕ l-diversity ⊕

the private distribution values close full data set.

Example: Netflix Prize

Story Time:

Example: Netflix Prize

Story Time:

- In 2006, Netflix (e.g., DVDs) released awesome data sets $D_1 = \{\langle \text{user-id, movie, date of grade, grade} \rangle\}$.
  And another set $D_2 = \{\langle \text{user-id, movie, date of grade} \rangle\}$.
  Wants researchers to predict grade on D_2.
  (Had another similar private data D_3 to evaluate grades : cross validation.)

Example: Netflix Prize

Story Time:

- In 2006, Netflix (e.g., DVDs) released awesome data sets $D_1 = \{\langle \text{user-id, movie, date of grade, grade} \rangle\}$.
  And another set $D_2 = \{\langle \text{user-id, movie, date of grade} \rangle\}$.
  Wants researchers to predict grade on D_2.
  (Had another similar private data D_3 to evaluate grades : cross validation.)
- If certain improvement over Netflix's algorithm, get \$1 million!

Example: Netflix Prize

Story Time:

- In 2006, Netflix (e.g., DVDs) released awesome data sets $D_1 = \{\langle \text{user-id, movie, date of grade, grade} \rangle\}$.
  And another set $D_2 = \{\langle \text{user-id, movie, date of grade} \rangle\}$.
  Wants researchers to predict grade on D_2.
  (Had another similar private data D_3 to evaluate grades : cross validation.)
- If certain improvement over Netflix's algorithm, get \$1 million!
- Led to lots of cool research!

Example: Netflix Prize

Story Time:

- In 2006, Netflix (e.g., DVDs) released awesome data sets
  $D_1 = \{\langle$user-id, movie, date of grade, grade$\rangle\}$.
  And another set $D_2 = \{\langle$user-id, movie, date of grade$\rangle\}$.
  Wants researchers to predict grade on D_2.
  (Had another similar private data D_3 to evaluate grades : cross validation.)
- If certain improvement over Netflix's algorithm, get \$1 million!
- Led to lots of cool research!
- Raters of movies also rated on IMDB (w/ user id, time stamp)

Example: Netflix Prize

Story Time:

- In 2006, Netflix (e.g., DVDs) released awesome data sets $D_1 = \{\langle \text{user-id, movie, date of grade, grade} \rangle\}$.
  And another set $D_2 = \{\langle \text{user-id, movie, date of grade} \rangle\}$.
  Wants researchers to predict grade on D_2.
  (Had another similar private data D_3 to evaluate grades : cross validation.)
- If certain improvement over Netflix's algorithm, get \$1 million!
- Led to lots of cool research!
- Raters of movies also rated on IMDB (w/ user id, time stamp)
- Researchers showed that by linking who rated similar sets of movies, with similar scores and times, they could identify many people.

Example: Netflix Prize

Story Time:

- In 2006, Netflix (e.g., DVDs) released awesome data sets $D_1 = \{\langle$user-id, movie, date of grade, grade$\rangle\}$.
  And another set $D_2 = \{\langle$user-id, movie, date of grade$\rangle\}$.
  Wants researchers to predict grade on D_2.
  (Had another similar private data D_3 to evaluate grades : cross validation.)
- If certain improvement over Netflix's algorithm, get \$1 million!
- Led to lots of cool research!
- Raters of movies also rated on IMDB (w/ user id, time stamp)
- Researchers showed that by linking who rated similar sets of movies, with similar scores and times, they could identify many people.
- (maybe watched embarrassing films on Netflix, not listed on IMDB)

Example: Netflix Prize

Story Time:

- In 2006, Netflix (e.g., DVDs) released awesome data sets $D_1 = \{\langle$user-id, movie, date of grade, grade$\rangle\}$.
  And another set $D_2 = \{\langle$user-id, movie, date of grade$\rangle\}$.
  Wants researchers to predict grade on D_2.
  (Had another similar private data D_3 to evaluate grades : cross validation.)
- If certain improvement over Netflix's algorithm, get \$1 million!
- Led to lots of cool research!
- Raters of movies also rated on IMDB (w/ user id, time stamp)
- Researchers showed that by linking who rated similar sets of movies, with similar scores and times, they could identify many people.
- (maybe watched embarrassing films on Netflix, not listed on IMDB)
- Class action lawsuit filed (lated dropped) against Netflix.

Example: Netflix Prize

Story Time:

- In 2006, Netflix (e.g., DVDs) released awesome data sets $D_1 = \{\langle$user-id, movie, date of grade, grade$\rangle\}$.
  And another set $D_2 = \{\langle$user-id, movie, date of grade$\rangle\}$.
  Wants researchers to predict grade on D_2.
  (Had another similar private data D_3 to evaluate grades : cross validation.)
- If certain improvement over Netflix's algorithm, get \$1 million!
- Led to lots of cool research!
- Raters of movies also rated on IMDB (w/ user id, time stamp)
- Researchers showed that by linking who rated similar sets of movies, with similar scores and times, they could identify many people.
- (maybe watched embarrassing films on Netflix, not listed on IMDB)
- Class action lawsuit filed (lated dropped) against Netflix.
- Netflix Prize had proposed sequel, dropped in 2010 for more privacy concerns.

Differential Privacy (DP)

Guarantee on release data X_1

so if another dataset X_2

$$\text{diff}(X_1, X_2) = |X_1 \Delta X_2| = 1$$

$$\frac{\Pr[q(X_1) = v]}{\Pr[q(X_2) = v]} \leq \exp(\varepsilon) \approx 1 + \varepsilon$$

$$= e^{\varepsilon}$$

$$\varepsilon = 0.10$$

$$\Rightarrow \varepsilon\text{-DP}$$

Want to hold for class of queries $\mathcal{Q}$

$q \in \mathcal{Q}$

$$\frac{\Pr[g(X_1)=r]}{\Pr[g(X_2)=r]} = 2$$

$1+\varepsilon$

$\Rightarrow \varepsilon = 1$

$\Rightarrow$ twice as likely to predict X_1 than X_2

$$\frac{\Pr(g(X_1)=r]}{\Pr(g(X_2)=r]} = 99$$

Use $\varepsilon \in (0, \infty)$

usually $\varepsilon \approx 1$ OR

Data $X \in \mathbb{R}$ (ex. height of Sly Stallone)

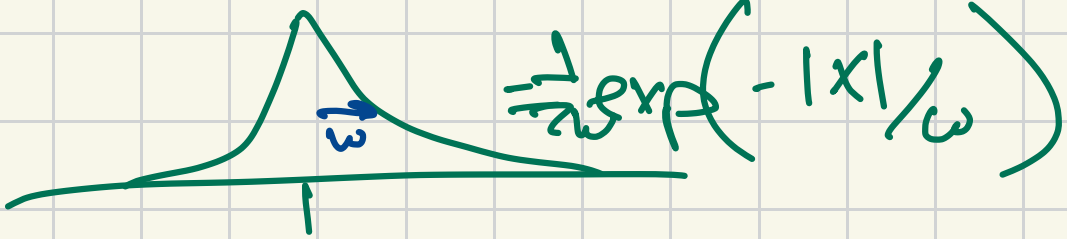
X'

$x_1 = 68$ inches

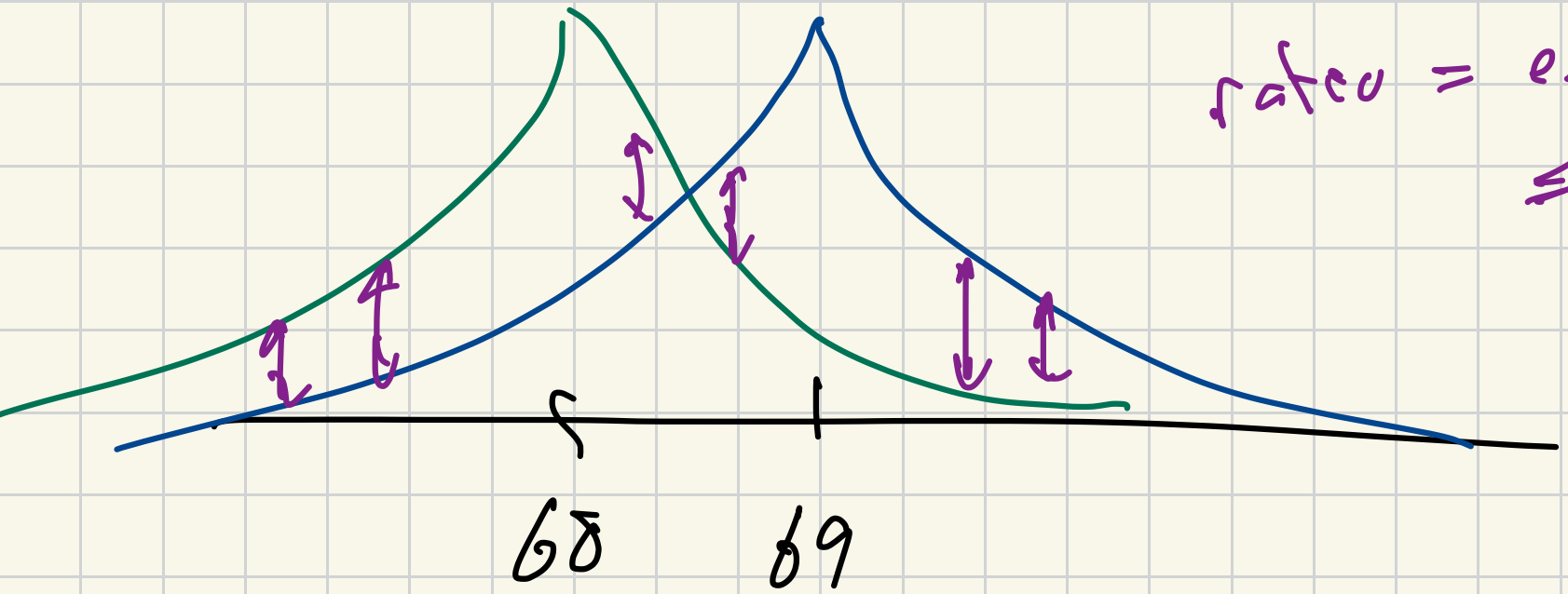
$x_2 = 69$ inches

$\text{diff}(x_1, x_2) = 1$

$s = q(x_1) = x_1 + \text{Lap}(0, \omega)$ $\quad \omega = 1/\varepsilon$



$$\frac{\Pr[s_1 = 70]}{\Pr[s_2 = 70]} = \frac{\frac{1}{2\omega}\exp(|68-70|/\omega)}{\frac{1}{2\omega}\exp(|69-70|/\omega)} = \exp\left(\tfrac{1}{2}|68-70| - \tfrac{1}{2}|69-70|\right) = \exp\left(\tfrac{1}{\omega}\right) = \exp(\varepsilon)$$



$$\text{ratio} = \exp(\epsilon)$$
$$\approx 1+\epsilon$$

Binary Example $X_1 = 1$ $X_2 = 0$

$$q(x_1) = x_1 + Lap(1/\varepsilon)$$

$$\frac{Pr[q(x_1) = 1.2]}{Pr[q(x_2) = 1.2]} = \frac{\exp(|1.2-1|/(1/\varepsilon))}{\exp(|1.2-0|/(1/\varepsilon))}$$

$$= \exp(\varepsilon(|1.2-1| - |1.2-0|))$$

$$= \exp(\varepsilon)$$

Database:

$X_1 = [1, 0, 1, 0, 0, 0]$

$X_2 = [1, 0, 0, 0, 0, 0]$

↑ mg data

$q_R(x_i) = |R \cap X_i| + Lap(1/\varepsilon)$

↑ range