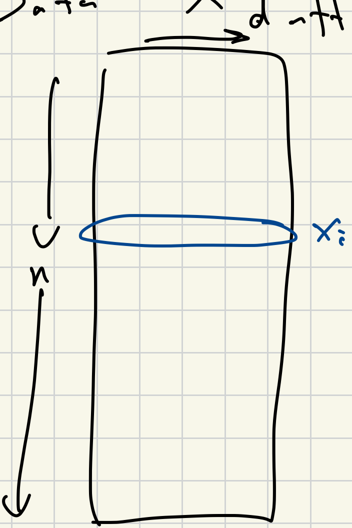
L21: Privacy

Jeff M. Phillips

April 2, 2025

Data X



What if this was you?

Ethics $\approx$ Empathy

Early 2000s

big companies

collect lots of data
usually on customers

make data public (sometimes)

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 can we release data anonymously while preserving information?

<u>Goal</u>: Represent info so we can draw generalization.

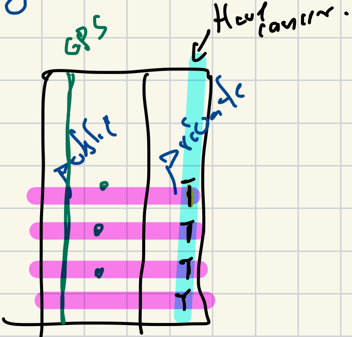
<u>k-anonymity</u>: Remove attributes (from subset) until each combination of attributes available maps to at least k different records.

Divide data

public attributes / private attributes

- zipcode
- age
- gender

- has cancer
- has diabetes
- search for Britney Spears



l-diversity : k-anonymity +

each group, had private attribute diverse values.

What if all in group had either cancer or diabetes

t-closeness : l-diversity +

the private traits were close (in distribution) to all data.

Example: Netflix Prize

Story Time:

Example: Netflix Prize

1, 2, 3, 4, 5

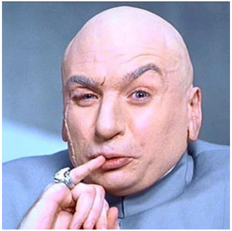
Story Time:

- 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
  $D_1 = \{\langle \text{user-id, movie, date of grade, grade} \rangle\}$.
  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)

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 $D_1 = \{\langle \text{user-id, movie, date of grade, grade} \rangle\}$.
  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.
- (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 $D_1 = \{\langle \text{user-id, movie, date of grade, grade} \rangle\}$.
  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.
- (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 $D_1 = \{\langle \text{user-id, movie, date of grade, grade} \rangle\}$.
  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.
- (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

- Gaurantee on released D_1 (ε-DP)

  so if dataset D_2 so $\text{diff}(D_1, D_2) = 1$

$$\frac{\Pr[q(D_1)]}{\Pr[q(D_2)]} \leq \approx 1+\varepsilon$$

Goal release D. w query w/ noise

so all query $\text{err}(q(D))$

ustore $\leq f(\varepsilon)$

$$\frac{\Pr[q(s) = r]}{\Pr[q(s') = r]} = 2 \implies \text{twice as likely to predict } s \text{ not } s'$$

$x_i = 1 \implies$ cancer

$$\frac{\Pr(s)}{\Pr(s')} = 99 \implies 99\% \text{ sure } X \text{ not } X'$$

$\varepsilon = 0.01$

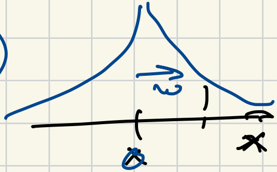
$$\frac{\Pr(s)}{\Pr(s')} \leq 1.01 \implies \text{not sure which } X \text{ or } X'$$

if ε small $\implies$ more private

Laplacian Mechanism

adds Laplace noise $Lap(x; w)$

$$Lap(x; w) = \frac{1}{2w} \exp\left(\frac{-|x|}{w}\right)$$



$S = \{s_i\}$ each

$$s_i = x_i + Lap(1/\epsilon = w)$$

$X = \{x_1, x_2, \ldots x_m\}$

$\downarrow \oplus Lap(1/\epsilon)$

$S = \{s_1, s_2, \ldots s_m\}$

Height

$x = 66$ inches $\qquad X = \{x\}$

$x' = 67$ inches $\qquad X' = \{x'\}$

$$s \leftarrow M(x) = x + \mathrm{Lap}(1/\varepsilon) \qquad w = \frac{1}{\varepsilon}$$

$$\frac{\Pr[s = 70]}{\Pr[s' = 70]} = \frac{\cancel{\frac{1}{2w}} \exp(|66-70|/w)}{\cancel{\frac{1}{2w}} \exp(|67-70|/w)}$$

$$= \exp\left(\frac{1}{w}\left(\overset{4}{|66-70|} \overset{-}{-} \overset{3}{|67-70|}\right)\right)$$

$$= \exp\left(\frac{1}{w}(1)\right) = \exp(\varepsilon)$$

Binary Example

$X = \{x\}$ $\quad x \in \{0,1\}$

$\cancel{X} = \{x = 1\}$ $\quad X' = \{x' = 0\}$

$S = \{s\}$ $\quad s = x + \mathrm{Lap}(1/\varepsilon)$

$$\frac{\Pr[g(s) = 1]}{\Pr[g(s') = 1]} = \frac{\exp\left(|1-1| / (1/\varepsilon)\right)}{\exp\left(|0-1| / (1/\varepsilon)\right)}$$

$$= \exp\left(\frac{1}{(1/\varepsilon)} \left(|1-1| - |0-1|\right)\right)$$

$$= \exp(\varepsilon \cdot (1)) = \exp(\varepsilon)$$

Binary database

$X = \{x_1, x_2, \dots x_n\}$ $\quad x_i \in \{0,1\}$

$X' = \{x_1', x_2', \dots x_n'\}$

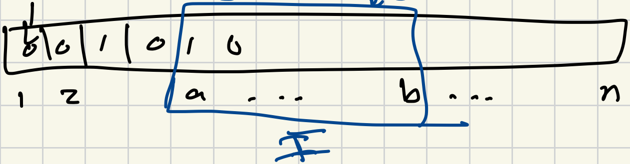
some $x_i \neq x_i'$

R = Interval

$I = [a, b]$

$I \cap X = \{x_a, x_{a+1}, x_{a+2} \dots x_b\}$

$g_I(X) = \sum_{i \in I} x_i = |X \cap I|$



Laplacian Mechanism $\quad S \leftarrow M(X)$

$\{s_1 \dots s_n\}$ $\quad s_j = x_j + \mathrm{Lap}(1/\epsilon)$

How about $g_I(S)$

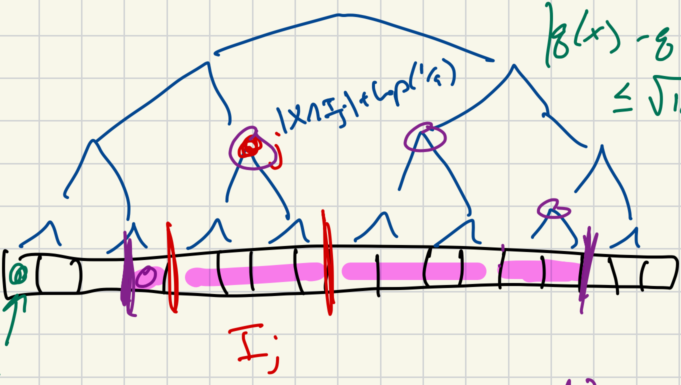
<u>option 1</u> $\quad g_I(S) = |X \cap I| + \mathrm{Lap}(1/\varepsilon)$

<u>option 2</u> $\quad g_I(S) = \sum_{i \in I} s_i = \sum_{i \in I} \left(x_i + \mathrm{Lap}(1/\varepsilon)\right)$

$$= |X \cap I| + |I| \cdot \mathrm{Lap}(1/\varepsilon)$$

$$|g_I(S) - g_I(X)| \leq \sqrt{|I|} \cdot \mathrm{Lap}(1/\varepsilon)$$

$$\leq \sqrt{n} \cdot (1/\varepsilon)$$



$$|g(x) - g(s)| \le \sqrt{\log n} \cdot 1/\varepsilon$$

$s_1 + Lap(1/\varepsilon)$ g_{I_j}

any $I = \bigcup_{j \in J} I_j$

$|J| \le 2 \log n$

$$g(s) = \sum_{j \in J}{}' g_{I_j}(s) = \sum_{j \in J}{}' |X \cap I_j| + Lap(1/\varepsilon)$$

$$= |X \cap I| + \sqrt{|J|}\, Lap(1/\varepsilon)$$