Differential Privacy Tutorial

Simons Institute Workshop on Privacy and Big Data

Katrina Ligett
Caltech

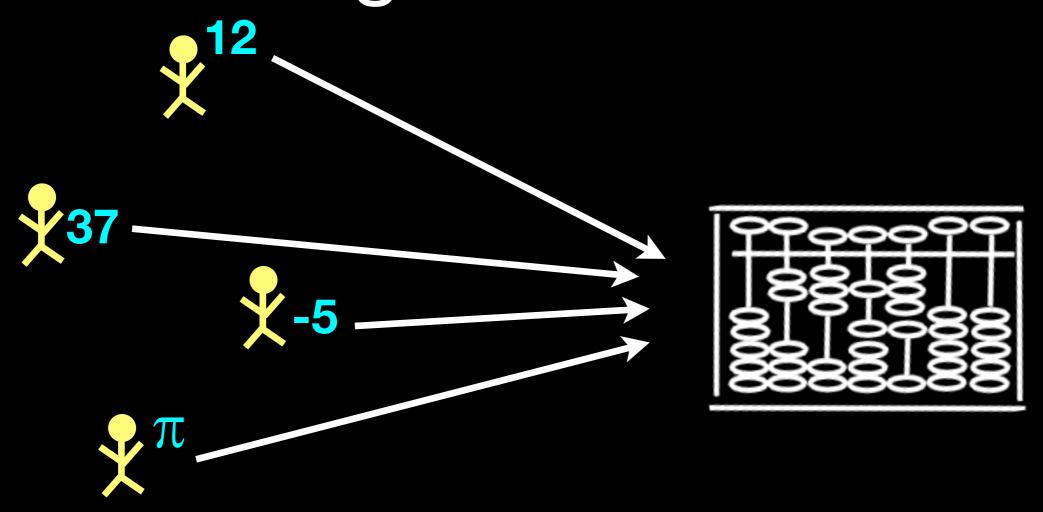
individuals have lots of interesting data...







individuals have lots of interesting data...



...we want to compute on it















1040

Department of the Treasury-Internal Revenue Service









1040

Department of the Treasury-Internal Revenue Service











E 1040

Department of the Treasury-Internal Revenue Service















1040

Department of the Treasury-Internal Revenue Service















Google

data privacy

data privacy day
data privacy laws
data privacy act
data privacy policy
data privacy safe harbor
data privacy breaches
data privacy legislation
data privacy audit
data privacy through optimal k-anonymization
data privacy laws us

75,400,000 results 17,600,000 results 11,100,000 results 60,400,000 results

332,000 results 1,320,000 results

980,000 results

684,000 results 4,200 results

4,200 result

71,900,000 results

close

1040

Department of the Treasury-Internal Revenue Service

U.S. Individual Income Tax Return

2008





PatientCare Portal







facebook

Google

data privacy

data privacy day
data privacy laws
data privacy act
data privacy policy
data privacy safe harbor
data privacy breaches
data privacy legislation
data privacy audit
data privacy through optimal k-anonymization
data privacy laws us

75,400,000 results 17,600,000 results 11,100,000 results 60,400,000 results 332,000 results

1,320,000 results 980,000 results

684,000 results 4,200 results

4,200 results

71,900,000 results

close

1040

Department of the Treasury-Internal Revenue Service

U.S. Individual Income Tax Return

2008



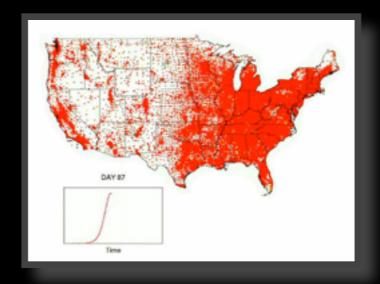


PatientCare Portal

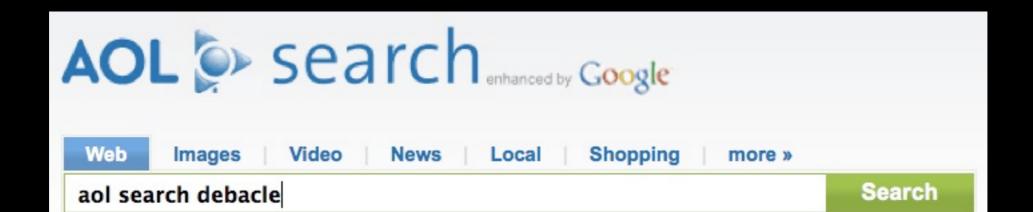




- finding statistical correlations
 - genotype/phenotype associations
 - correlating medical outcomes with risk factors or events
- publishing aggregate statistics
- noticing events/outliers
 - intrusion detection
 - disease outbreaks



- datamining/learning tasks
 - use customer data to update strategies





Web Images Video News Local Shopping more »

aol search debacle

Search

Stats: Who's to blame for AOL's search debacle?

Let the fingerpointing begin A friend of ousted **AOL** advertising executive Mike Kelly takes issue with our assignment of blame.

gawker.com/302054/whos-to-blame-for-aols-search-debacle - Similar pages

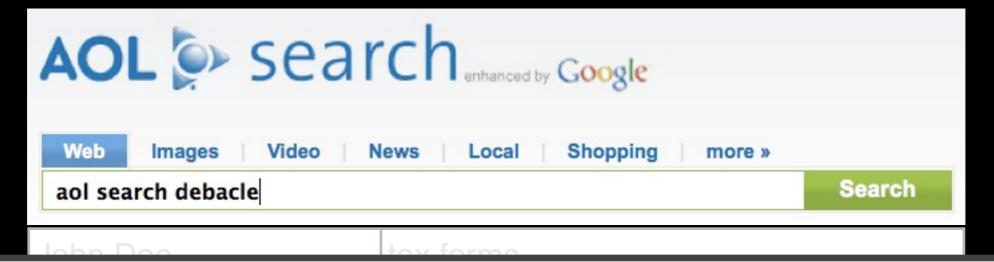
AOL Proudly Releases Massive Amounts of Private Data

Yet Another Update: AOL: This was a screw up Further Update: Sometime after 7 pm the download link went down as well, but ...

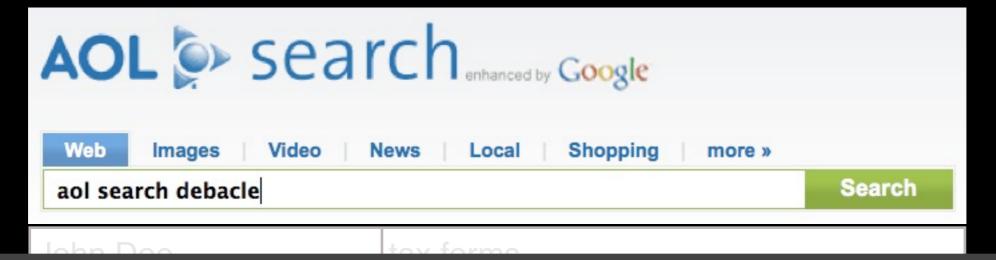
www.techcrunch.com/2006/08/06/aol-proudly-releas... - 123k - Similar pages

AOL search data scandal - Wikipedia, the free encyclopedia

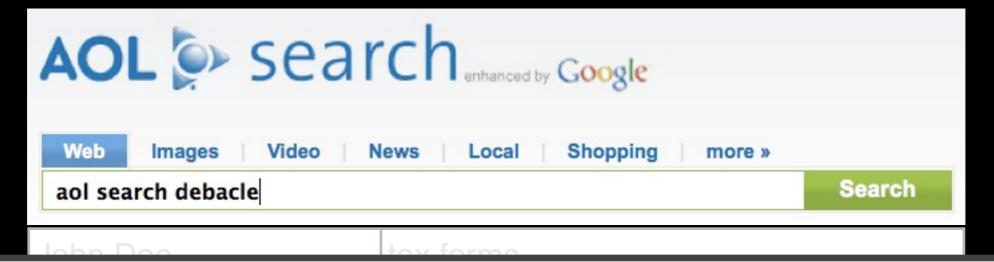
The AOL search data scandal was the result of a research project by AOL. AOL apologizes for release of user search data | CNET News.com; ^ AOL search ... en.wikipedia.org/wiki/AOL_search_data_scandal - 45k - Similar pages



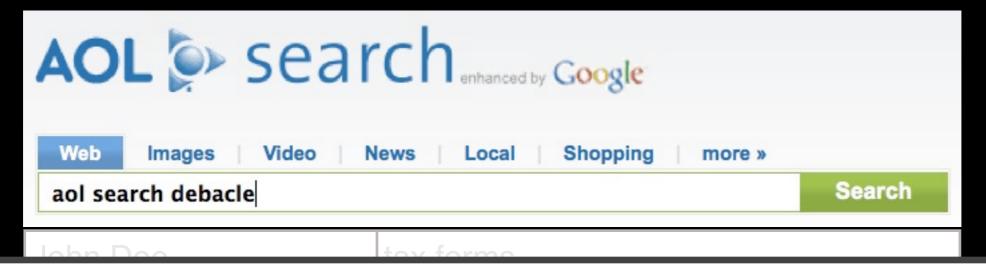
| John Doe | error in form 1099 |
|----------------|-----------------------|
| Katrina Ligett | data privacy |
| Katrina Ligett | aol search debacle |
| Katrina Ligett | Ligett DBLP |
| Katrina Ligett | computer science news |
| Katrina Ligett | Caltech rankings |
| Katrina Ligett | weather Pasadena |
| Jane Smith | youtube |
| Jane Smith | free tv download |
| Jane Smith | streaming tv |
| Chris Jones | childrens books |
| Chris Jones | dr seuss |
| Chris Jones | "the cat and the hat" |
| Chris Jones | gifts for children |
| | |



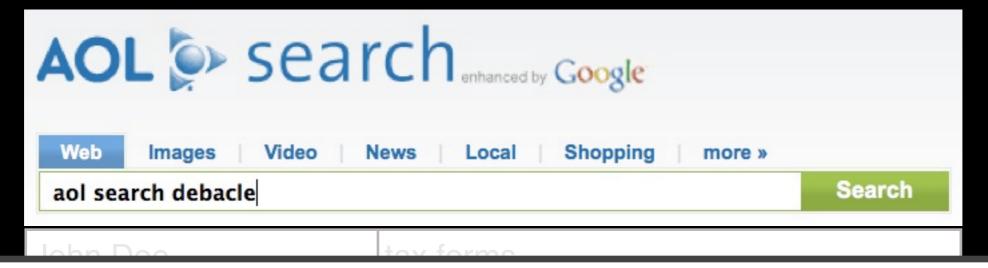
| John Doe | error in form 1099 |
|-------------|-----------------------|
| user195023 | data privacy |
| user195023 | aol search debacle |
| user195023 | Ligett DBLP |
| user195023 | computer science news |
| user195023 | Caltech rankings |
| user195023 | weather Pasadena |
| Jane Smith | youtube |
| Jane Smith | free tv download |
| Jane Smith | streaming tv |
| Chris Jones | childrens books |
| Chris Jones | dr seuss |
| Chris Jones | "the cat and the hat" |
| Chris Jones | gifts for children |
| 01 1 | |



| John Doe | error in form 1099 |
|-------------|-----------------------|
| user195023 | data privacy |
| user195023 | aol search debacle |
| user195023 | Ligett DBLP |
| user195023 | computer science news |
| user195023 | Caltech rankings |
| user195023 | weather Pasadena |
| Jane Smith | youtube |
| Jane Smith | free tv download |
| Jane Smith | streaming tv |
| Chris Jones | childrens books |
| Chris Jones | dr seuss |
| Chris Jones | "the cat and the hat" |
| Chris Jones | gifts for children |
| | |



| John Doe | error in form 1099 | | | | |
|------------|-----------------------|--|--|--|--|
| user195023 | data privacy | | | | |
| user195023 | aol search debacle | | | | |
| user195023 | Ligett DBLP | | | | |
| user195023 | computer science news | | | | |
| user195023 | Caltech rankings | | | | |
| user195023 | weather Pasadena | | | | |
| | youtube | | | | |
| | free tv download | | | | |
| d hoc | streaming tv | | | | |
| | childrens books | | | | |
| ions are | dr seuss | | | | |
| sky! | "the cat and the hat" | | | | |
| SKY: | gifts for children | | | | |



| John Doe | error in form 1099 |
|------------|-----------------------|
| user195023 | data privacy |
| user195023 | aol search debacle |
| user195023 | Ligett DBLP |
| user195023 | computer science news |
| user195023 | Caltech rankings |
| user195023 | weather Pasadena |
| | voutubo |

Ad hoc solutions are risky!

free tv download
streaming tv
childrens books
dr seuss
"the cat and the ha
gifts for children

Huge opportunity for formalism.

| name | DOB | sex | weight | smoker | lung cancer |
|---------------|---------|-----|--------|--------|----------------|
| John Doe | 12/1/51 | М | 185 | Υ | N |
| Jane Smith | 3/3/46 | F | 140 | Z | N |
| Ellen Jones | 4/24/59 | F | 160 | Υ | Υ |
| Jennifer Kim | 3/1/70 | F | 135 | Z | N |
| Rachel Waters | 9/5/43 | F | 140 | Ν | N |

| name | DOB | sex | weight | smoker | lung cancer |
|---------------|---------|-----|--------|--------|----------------|
| John Doe | 12/1/51 | М | 185 | Υ | Ν |
| Jane Smith | 3/3/46 | F | 140 | Ν | Ν |
| Ellen Jones | 4/24/59 | F | 160 | Υ | Υ |
| Jennifer Kim | 3/1/70 | F | 135 | Ν | N |
| Rachel Waters | 9/5/43 | F | 140 | N | N |

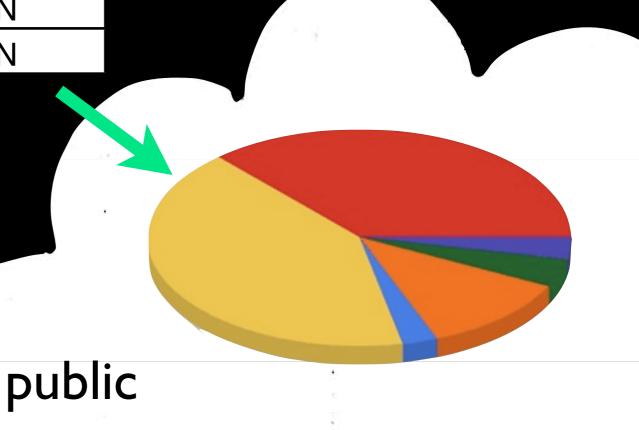
public

| name | DOB | sex | weight | smoker | lung cancer |
|---------------|---------|-----|--------|--------|----------------|
| John Doe | 12/1/51 | М | 185 | Υ | Ν |
| Jane Smith | 3/3/46 | F | 140 | Ν | Ν |
| Ellen Jones | 4/24/59 | F | 160 | Υ | Υ |
| Jennifer Kim | 3/1/70 | F | 135 | N | N |
| Rachel Waters | 9/5/43 | F | 140 | N | N |

| пата | DOB | SDK | waigitt | amaka: | tung |
|--------------|-----------|-----|---------|--------|------|
| Japhin Date | 电影性性的 | W | 1985 | W. | 16 |
| Jame Smith | 3/3/44 | F | 140 | 16 | 16 |
| Ellen Jones | 14/24/50 | F | 1980) | W | W |
| Jannifar Kim | 3/1/70 | F | 1(385) | 16 | 16 |
| Placified. | 13/5/4(3) | F | 140 | 16 | 16 |

| name | DOB | sex | weight | smoker | lung cancer |
|---------------|---------|-----|--------|--------|----------------|
| John Doe | 12/1/51 | М | 185 | Υ | N |
| Jane Smith | 3/3/46 | F | 140 | Ν | N |
| Ellen Jones | 4/24/59 | F | 160 | Υ | Υ |
| Jennifer Kim | 3/1/70 | F | 135 | N | N |
| Rachel Waters | 9/5/43 | F | 140 | Ν | N |

| name | DOB | sex | weight | smoker | lung cancer |
|---------------|---------|-----|--------|--------|----------------|
| John Doe | 12/1/51 | М | 185 | Υ | Ν |
| Jane Smith | 3/3/46 | F | 140 | Ν | Ν |
| Ellen Jones | 4/24/59 | F | 160 | Υ | Υ |
| Jennifer Kim | 3/1/70 | F | 135 | Ν | N |
| Rachel Waters | 9/5/43 | F | 140 | Ν | N |



| name | DOB | sex | weight | smoker | lung cancer |
|---------------|---------|-----|--------|--------|----------------|
| John Doe | 12/1/51 | М | 185 | Υ | N |
| Jane Smith | 3/3/46 | F | 140 | Ν | N |
| Ellen Jones | 4/24/59 | F | 160 | Υ | Υ |
| Jennifer Kim | 3/1/70 | F | 135 | N | N |
| Rachel Waters | 9/5/43 | F | 140 | Ν | N |

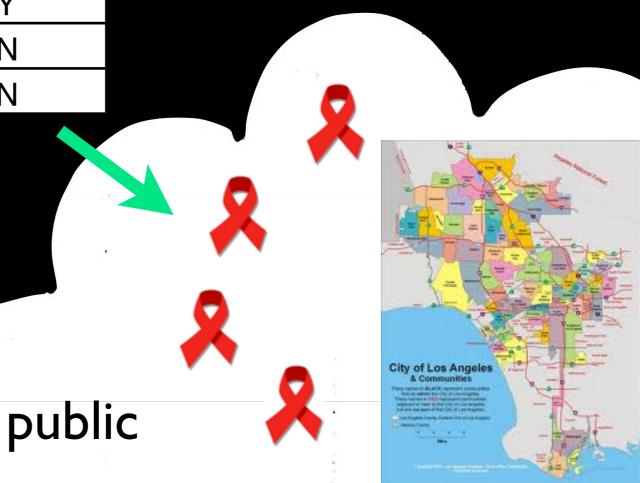
| name | DOB | sex | weight | smoker | lung cancer |
|---------------|---------|-----|--------|--------|----------------|
| John Doe | 12/1/51 | М | 185 | Υ | N |
| Jane Smith | 3/3/46 | F | 140 | Ν | N |
| Ellen Jones | 4/24/59 | F | 160 | Υ | Υ |
| Jennifer Kim | 3/1/70 | F | 135 | Ν | N |
| Rachel Waters | 9/5/43 | F | 140 | Z | Ν |

18%

public

| name | DOB | sex | weight | smoker | lung cancer |
|---------------|---------|-----|--------|--------|----------------|
| John Doe | 12/1/51 | М | 185 | Υ | N |
| Jane Smith | 3/3/46 | F | 140 | Ν | N |
| Ellen Jones | 4/24/59 | F | 160 | Υ | Υ |
| Jennifer Kim | 3/1/70 | F | 135 | N | N |
| Rachel Waters | 9/5/43 | F | 140 | Ν | N |

| name | DOB | sex | weight | smoker | lung cancer |
|---------------|---------|-----|--------|--------|----------------|
| John Doe | 12/1/51 | М | 185 | Υ | N |
| Jane Smith | 3/3/46 | F | 140 | N | N |
| Ellen Jones | 4/24/59 | F | 160 | Υ | Υ |
| Jennifer Kim | 3/1/70 | F | 135 | Ν | N |
| Rachel Waters | 9/5/43 | F | 140 | N | N |



not so fast...

not so fast...

see, e.g., Korolova 2011's Facebook microtargeting attack



not so fast...

see, e.g., Korolova 2011's Facebook microtargeting attack

... must pay attention to all uses of sensitive data



what to promise about output?

access to the output should not enable one to learn anything about an individual that could not be learned without access

access to the output should not enable one to learn anything about an individual that could not be learned without access

is this possible?

access to the output should not enable one to learn anything about an individual that could not be learned without access hint:

is this possible?

either privacy or utility separately is easy

| name | DOB | sex | weight | smoker | lung cancer |
|---------------|---------|-----|--------|--------|----------------|
| John Doe | 12/1/51 | М | 185 | Υ | N |
| Jane Smith | 3/3/46 | F | 140 | N | N |
| Ellen Jones | 4/24/59 | F | 160 | Υ | Υ |
| Jennifer Kim | 3/1/70 | F | 135 | N | N |
| Rachel Waters | 9/5/43 | F | 140 | N | N |

| name | DOB | sex | weight | smoker | lung cancer |
|---------------|---------|-----|--------|--------|----------------|
| John Doe | 12/1/51 | М | 185 | Υ | Z |
| Jane Smith | 3/3/46 | F | 140 | Ν | Ν |
| Ellen Jones | 4/24/59 | F | 160 | Υ | Υ |
| Jennifer Kim | 3/1/70 | F | 135 | Ν | Ν |
| Rachel Waters | 9/5/43 | F | 140 | N | N |

| name | DOB | sex | weight | smoker | lung cancer |
|---------------|---------|-----|--------|--------|----------------|
| John Doe | 12/1/51 | М | 185 | Υ | Ν |
| Jane Smith | 3/3/46 | F | 140 | Ν | Ν |
| Ellen Jones | 4/24/59 | F | 160 | Υ | Υ |
| Jennifer Kim | 3/1/70 | F | 135 | N | N |
| Rachel Waters | 9/5/43 | F | 140 | N | N |

there is a correlation of xxx

public

Ν

public

| name | DOB | sex | weight | smoker | lung cancer |
|----------|---------|-----|--------|--------|----------------|
| John Doe | 12/1/51 | М | 185 | Υ | N |
| Jane Smi | | | 140 | N | N |
| E" | | | | Υ | Υ |
| | | | | 1 | N |

but what if someone knew Alice is a smoker?

there is a correlation of xxx

access to the output should not enable one to learn anything about an individual that could not be learned without access

access to the output should not enable one to learn anything about an individual that could not be learned without access

not possible!

| name | DOB sex weight smoker | | lung cancer | | |
|---------------|-----------------------|---|----------------|---|---|
| John Doe | 12/1/51 | М | 185 | Υ | N |
| Jane Smith | 3/3/46 | F | 140 | Ν | Ν |
| Ellen Jones | 4/24/59 | F | 160 | Υ | Υ |
| Jennifer Kim | 3/1/70 | F | 135 | Ν | Ν |
| Rachel Waters | 9/5/43 | F | 140 | N | Ν |

| name | DOB | sex | weight | smoker | lung cancer |
|---------------|---------|-----|--------|--------|----------------|
| John Doe | 12/1/51 | М | 185 | Υ | N |
| Jane Smith | 3/3/46 | F | 140 | N | N |
| Ellen Jones | 4/24/59 | F | 160 | Υ | Υ |
| Jennifer Kim | 3/1/70 | IL | 135 | Z | N |
| Rachel Waters | 9/5/43 | F | 140 | N | N |

think of output as randomized

| name | DOB | SEX | weight | smoker | lung |
|---------------|---------|-----|--------|--------|--------|
| Harrio | | | | omore | cancer |
| John Doe | 12/1/51 | М | 185 | Υ | N |
| Jane Smith | 3/3/46 | F | 140 | Ν | N |
| Ellen Jones | 4/24/59 | F | 160 | Υ | Υ |
| Jennifer Kim | 3/1/70 | F | 135 | N | N |
| Rachel Waters | 9/5/43 | F | 140 | N | N |

18%

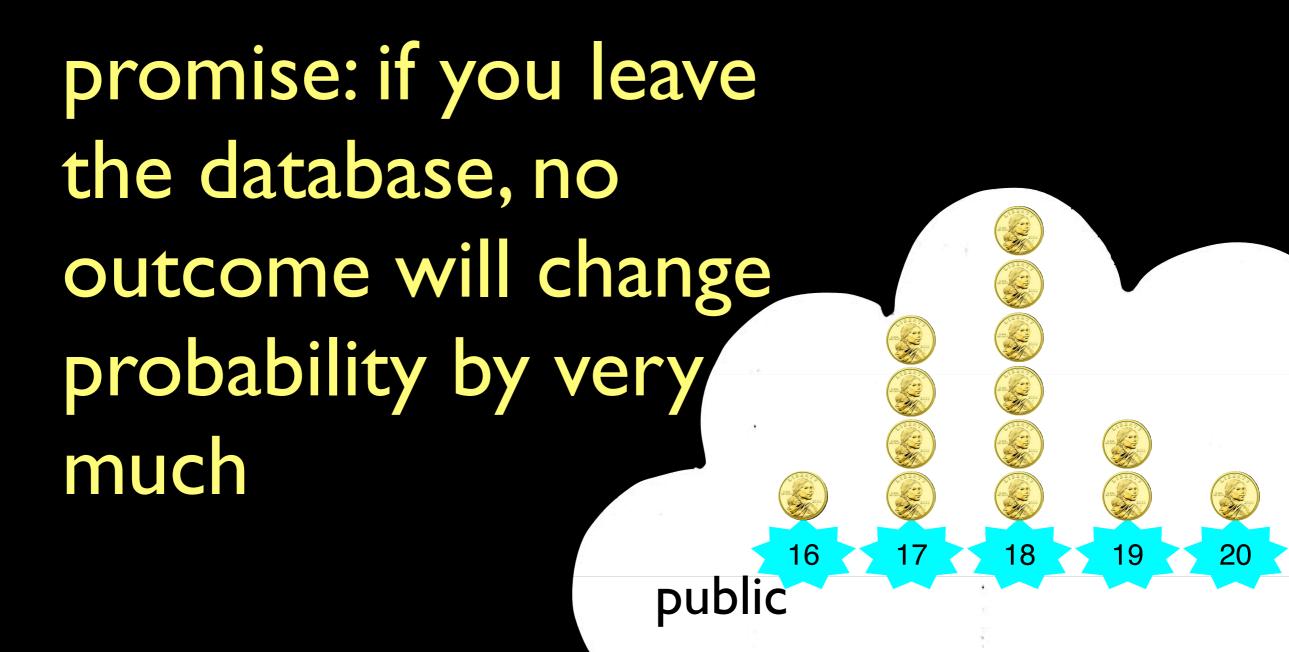
public

| name | DOB | sex | weight | smoker | lung cancer |
|---------------|---------|-----|--------|--------|----------------|
| John Doe | 12/1/51 | М | 185 | Υ | N |
| Jane Smith | 3/3/46 | F | 140 | N | N |
| Ellen Jones | 4/24/59 | F | 160 | Υ | Υ |
| Jennifer Kim | 3/1/70 | IL | 135 | Z | N |
| Rachel Waters | 9/5/43 | F | 140 | N | N |

| DOB | Sex | weight | smoker | lung |
|---------|-----------------------------|--|--|---|
| ם | | woigitt | Siriokci | cancer |
| 12/1/51 | М | 185 | Υ | N |
| 3/3/46 | F | 140 | Ν | N |
| 4/24/59 | F | 160 | Υ | Υ |
| 3/1/70 | F | 135 | N | N |
| 9/5/43 | F | 140 | N | N |
| | 3/3/46 4/24/59 3/1/70 | 12/1/51 M 3/3/46 F 4/24/59 F 3/1/70 F | 12/1/51 M 185 3/3/46 F 140 4/24/59 F 160 3/1/70 F 135 | 12/1/51 M 185 Y 3/3/46 F 140 N 4/24/59 F 160 Y 3/1/70 F 135 N |

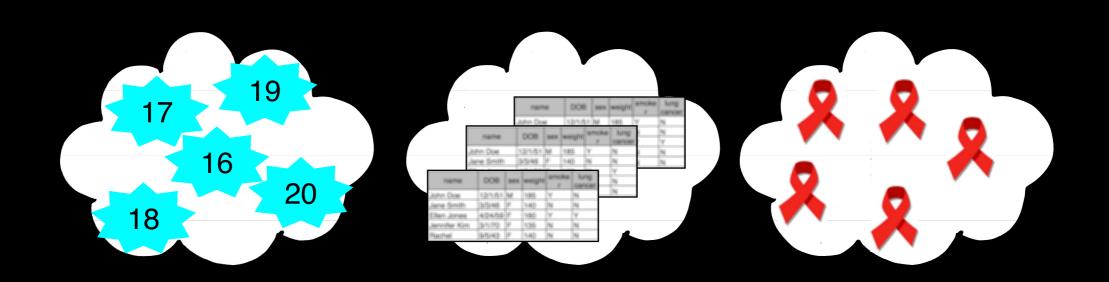


| name | DOB | sex | weight | smoker | lung cancer | | | | | |
|---------------|---------|-----|--------|--------------------|----------------|-------|----|--------|----|----|
| John Doe | 12/1/51 | М | 185 | Υ | N | | | | | |
| Jane Smith | 3/3/46 | F | 140 | Z | Z | | | | | |
| Ellen Jones | 4/24/59 | F | 160 | Υ | Υ | | | | | |
| Jennifer Kim | 3/1/70 | F | 135 | Z | N | | | 3 ± 27 | | |
| Rachel Waters | 9/5/43 | F | 140 | Z | N | | | | | |
| | | | | IA 500 WE TRUST | | 16 | | | 10 | |
| | | | | | 2 | 16 | 17 | 18 | 19 | 20 |
| | | | | | P | ublic | | | | • |



more formally...

- database D a set of rows, one per person
- sanitizing algorithm M probabilistically maps
 D to event or object in outcome space



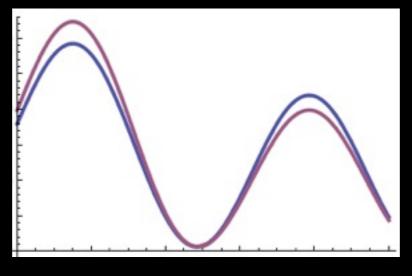
[DinurNissim03, DworkNissimMcSherrySmith06]

ε-Differential Privacy for mechanism M:

for any two neighboring data sets D_1 , D_2 ,

any $C \in range(M)$,

 $\Pr[\mathsf{M}(\mathsf{D}_1) \in \mathsf{C}] \le e^\epsilon \Pr[\mathsf{M}(\mathsf{D}_2) \in \mathsf{C}]$



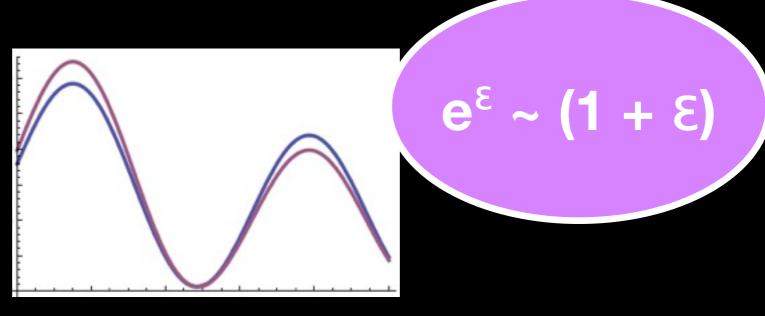
[DinurNissim03, DworkNissimMcSherrySmith06]

ε-Differential Privacy for mechanism M:

for any two neighboring data sets D_1 , D_2 ,

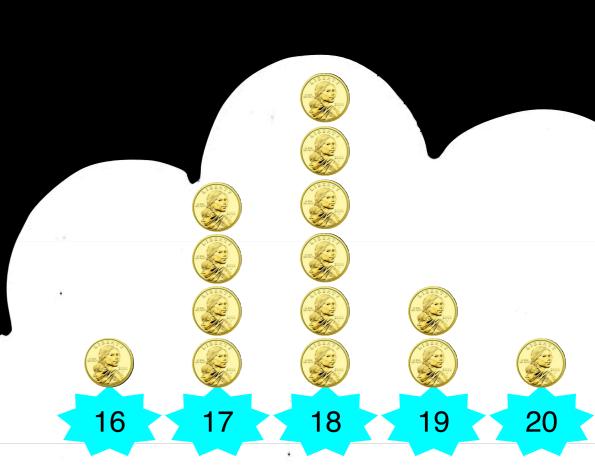
any $C \in range(M)$,

 $Pr[M(D_1) \in C] \leq e^{\epsilon} Pr[M(D_2) \in C]$



 $Pr[M(D_1) \in C] \le e^{\epsilon} Pr[M(D_2) \in C]$

| name | DOB | sex | weight | smoker | lung cancer |
|---------------|---------|-----|--------|--------|----------------|
| John Doe | 12/1/51 | М | 185 | Υ | N |
| Jane Smith | 3/3/46 | F | 140 | Z | N |
| Ellen Jones | 4/24/59 | F | 160 | Υ | Υ |
| Jennifer Kim | 3/1/70 | F | 135 | N | N |
| Rachel Waters | 9/5/43 | F | 140 | N | N |



 $Pr[M(D_1) \in C] \le e^{\epsilon} Pr[M(D_2) \in C]$

| name | DOB | sex | weight | smoker | lung cancer |
|---------------|---------|-----|--------|--------|----------------|
| John Doe | 12/1/51 | М | 185 | Υ | N |
| Jane Smith | 3/3/46 | F | 140 | N | Ν |
| Ellen Jones | 1/2-1/J | | 160 | > | Υ |
| Jennifer Kim | 3/1/70 | F | 135 | Z | Ν |
| Rachel Waters | 9/5/43 | F | 140 | Ν | N |

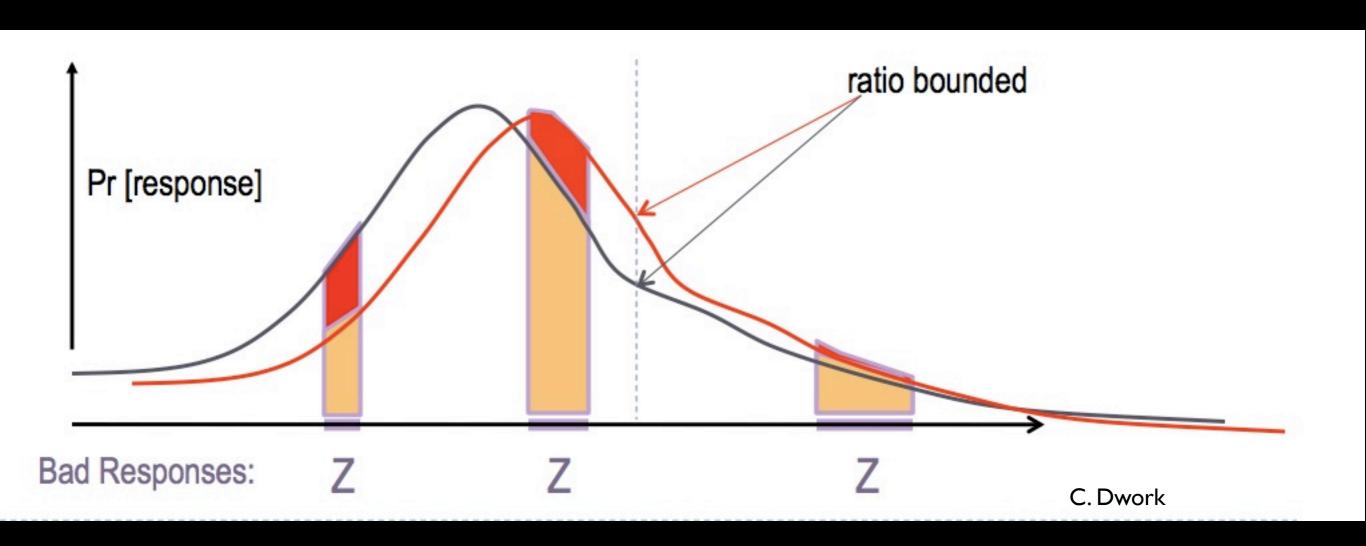


 $Pr[M(D_1) \in C] \leq e^{\epsilon} Pr[M(D_2) \in C]$

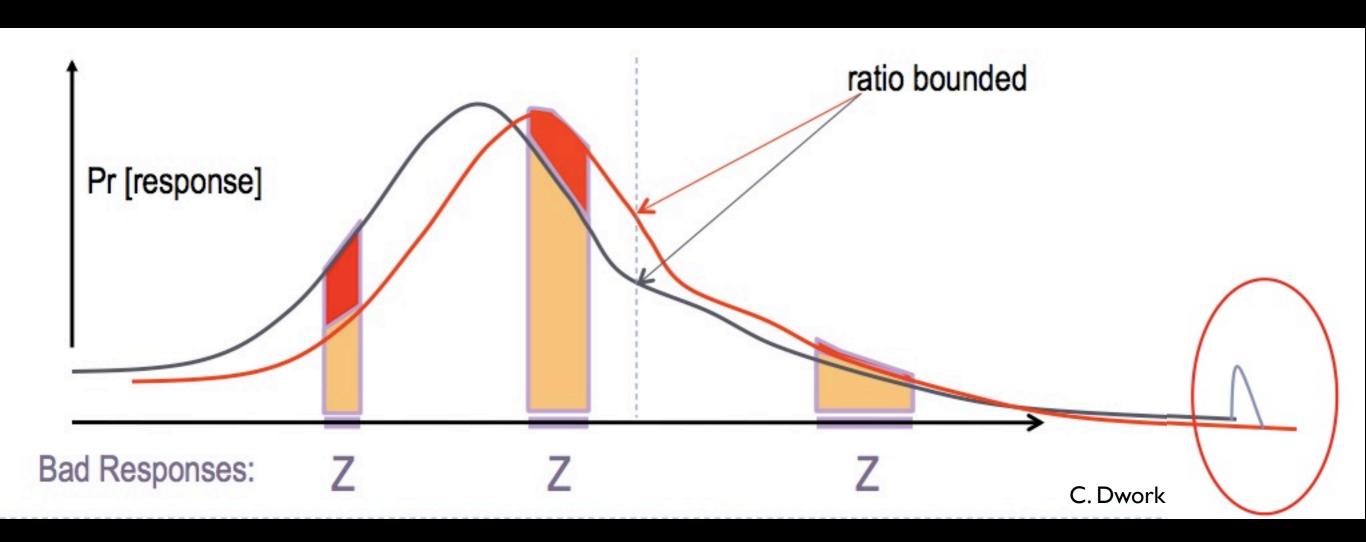
| name | DOB | sex | weight | smoker | lung cancer |
|---------------|---------|-----|--------|--------|----------------|
| John Doe | 12/1/51 | М | 185 | Υ | N |
| Jane Smith | 3/3/46 | F | 140 | N | N |
| Ellen Jones | 4/24/33 | | 160 | V | Υ |
| Jennifer Kim | 3/1/70 | F | 135 | N | N |
| Rachel Waters | 9/5/43 | F | 140 | N | N |



 $\Pr[\mathsf{M}(\mathsf{D}_1) \in \mathsf{C}] \leq e^\epsilon \Pr[\mathsf{M}(\mathsf{D}_2) \in \mathsf{C}]$



(ϵ, δ) -differential privacy $Pr[M(D_1) \in C] \le e^{\epsilon} Pr[M(D_2) \in C] + \delta$



$$\Pr[\mathsf{M}(\mathsf{D}_1) \in \mathsf{C}] \leq e^\epsilon \Pr[\mathsf{M}(\mathsf{D}_2) \in \mathsf{C}]$$

Is a statistical property of mechanism behavior

- unaffected by auxiliary information
- independent of adversary's computational power

 $\Pr[M(D_1) \in C] \le e^{\epsilon} \Pr[M(D_2) \in C]$

promise: if you leave the database, no outcome will change probability by very much

is this achievable?

yes!

if your output is a number...

| name | DOB | sex | weight | smoker | lung cancer |
|---------------|---------|-----|--------|--------|----------------|
| John Doe | 12/1/51 | М | 185 | Υ | N |
| Jane Smith | 3/3/46 | F | 140 | Ν | N |
| Ellen Jones | 4/24/59 | F | 160 | Υ | Υ |
| Jennifer Kim | 3/1/70 | F | 135 | Ν | N |
| Rachel Waters | 9/5/43 | F | 140 | N | N |

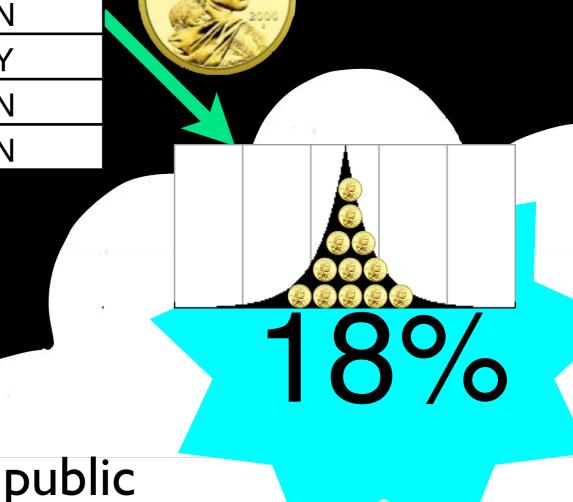
18%

public

if your output is a number...

| name | DOB | sex | weight | smoker | lung cancer |
|---------------|---------|-----|--------|--------|----------------|
| John Doe | 12/1/51 | М | 185 | Υ | N |
| Jane Smith | 3/3/46 | F | 140 | Z | N |
| Ellen Jones | 4/24/59 | F | 160 | Υ | Υ |
| Jennifer Kim | 3/1/70 | F | 135 | Z | N |
| Rachel Waters | 9/5/43 | F | 140 | Ν | N |

add noise with particular shape



sensitivity of a function f

 $\Delta f = \max_{D_1, D_2} |f(D_1) - f(D_2)|$

for neighboring data sets D_1 , D_2

sensitivity of a function f

$$\Delta f = \max_{D_1, D_2} |f(D_1) - f(D_2)|$$

for neighboring data sets D_1, D_2

measures how much one person can affect output

sensitivity of a function f

$$\Delta f = \max_{D_1, D_2} |f(D_1) - f(D_2)|$$

for neighboring data sets D_1, D_2

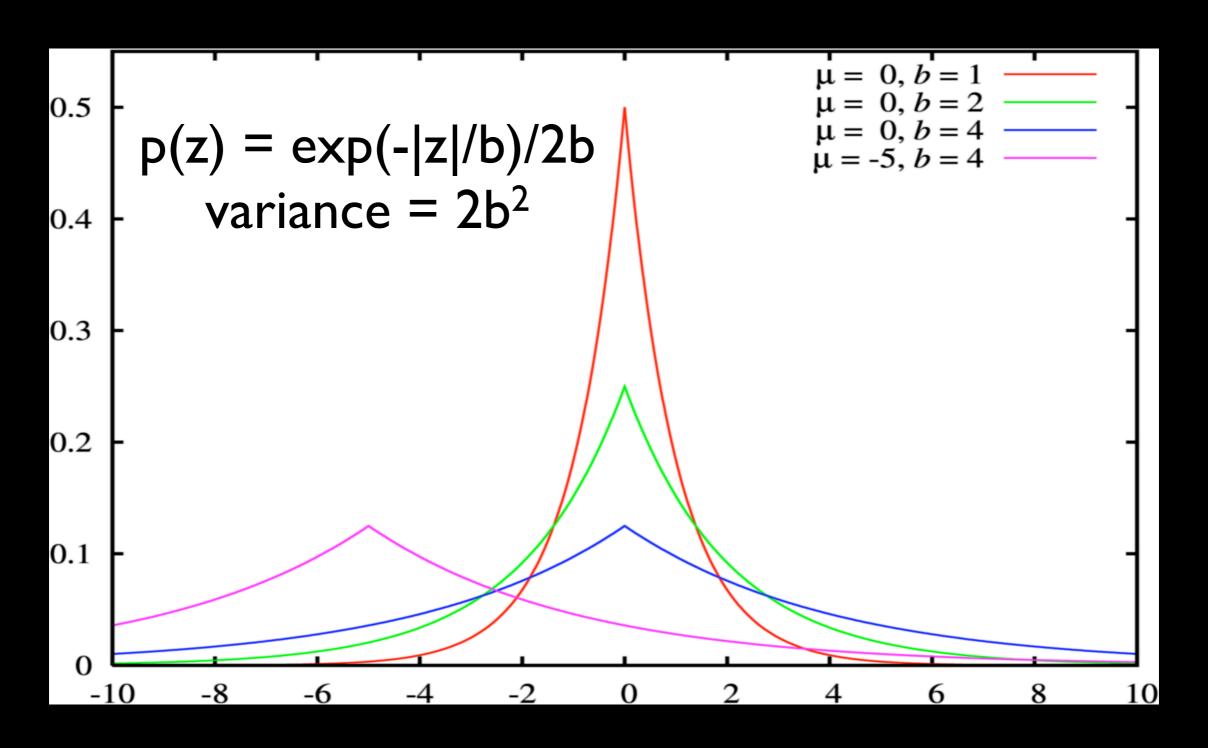
- measures how much one person can affect output
- sensitivity is 1 for counting queries that count number of rows satisfying a predicate

scale noise with sensitivity

$$\Delta f = \max_{D_1, D_2} |f(D_1) - f(D_2)|$$

[DMNS06]: on query f, can add scaled symmetric noise Lap(b) with $b = \Delta f/\epsilon$, to achieve ϵ -differential privacy.

Laplace distribution Lap(b)



applying the Laplace mechanism

applying the Laplace mechanism

single counting query: how many people in the database satisfy predicate P?

- sensitivity = I
- can add noise Lap(1/ε)

what if want more than one query? ...composition

• an ϵ_1 -DP mechanism, followed by an ϵ_2 -DP mechanism, is $(\epsilon_1 + \epsilon_2)$ -DP

what if want more than one query? ...composition

- an ϵ_1 -DP mechanism, followed by an ϵ_2 -DP mechanism, is $(\epsilon_1 + \epsilon_2)$ -DP
- can also sum both the epsilons and the deltas for (ϵ, δ) -DP

what if want more than one query? ...composition

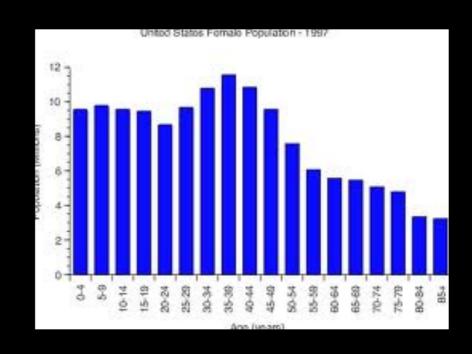
- an ϵ_1 -DP mechanism, followed by an ϵ_2 -DP mechanism, is $(\epsilon_1 + \epsilon_2)$ -DP
- can also sum both the epsilons and the deltas for (ϵ, δ) -DP
- more sophisticated argument: k runs of (ϵ, δ) -DP mechanisms gives $(\epsilon', k\delta + \delta')$ -DP for $\epsilon' = (2 \text{ k ln}(1/\delta'))^{1/2}\epsilon + \text{k }\epsilon \text{ (e}^{\epsilon} 1)$

applying the Laplace mechanism

vector-valued queries of dimension d

• can apply composition and add noise $Lap(d\Delta f/\epsilon)$ in each component of output, where Δf is the sensitivity of each component

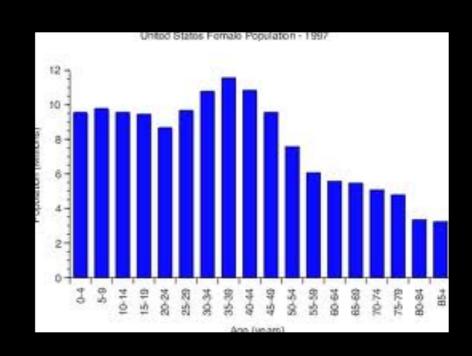
applying the Laplace mechanism



histogram queries

could again use noise Lap(d/ε)

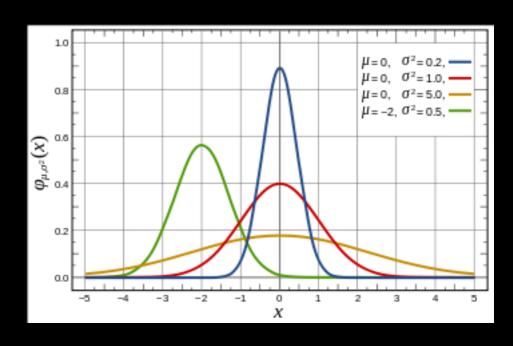
applying the Laplace mechanism



histogram queries

- could again use noise Lap(d/ε)
- but actually only need \sim Lap(I/ ϵ), since sensitivity generalizes as max L_I distance

Gaussian mechanism



[DKMMN06]: Gaussian noise gives (ϵ, δ) -DP with

 $\sigma \geq (2 \ln(2/\delta))^{1/2} / \epsilon * (max L_2 distance)$

Ok, but I wanted to use my database for more than a handful of statistics...

Ok, but I wanted to use my database for more than a handful of statistics...

Data can be "big" in two dimensions: more rows makes privacy easier (lower sensitivity); more columns makes it harder (more queries to preserve)

handling an exponential number of queries

| name | DOB | sex | weight | smoker | lung cancer |
|---------------|---------|-----|--------|--------|----------------|
| John Doe | 12/1/51 | М | 185 | Υ | Ν |
| Jane Smith | 3/3/46 | F | 140 | Z | Ν |
| Ellen Jones | 4/24/59 | F | 160 | Υ | Υ |
| Jennifer Kim | 3/1/70 | F | 135 | Ν | N |
| Rachel Waters | 9/5/43 | F | 140 | N | N |

| пата | DOB | SDK | waigitt | amaka: | tung |
|--------------|-----------|-----|---------|--------|------|
| Japhin Date | 电影性性电 | W | 1985 | W | 16 |
| Jame Smith | 3/3/46 | F | 140 | W | 16 |
| Ellen Jones | 相位相性致 | F | 1980) | W | W |
| Jannifar Kim | 3/1/00 | F | 1(282) | W | 16 |
| FlascHeal | 11/5/4(3) | F | 140 | W | 16 |

public

handling an expg number of que

sex

135

140

N

N

N

N

M

DOB

12/1/51

4/24/59

3/3/46

3/1/70

name

John Doe

Jane Smith

Ellen Jones

Jennifer Kim

Rachel Waters 9/5/43

| | sm | have lung cancer? what |
|-----|----|------------------------|
| 185 | Υ | fraction of males |
| 140 | Ν | |
| 160 | Υ | over 150 lbs? |

what

fraction over

age 50? what

fraction smoke and

. . .

THE MANAGEMENT

5460

11,385

NEW REPORT

知知相執

知为用

THESTINE Hirris gerigh. Japiniffap Sterr public

BLR08: ε-DP, error log^{1/3} |Q| n^{2/3}

BLR08: ε-DP, error log 1/3 |Q| n^{2/3}

DNRRV09: (ϵ, δ) -DP, error $|Q|^{o(1)}n^{1/2}$

BLR08: ε-DP, error log^{1/3} |Q| n^{2/3}

DNRRV09: (ϵ, δ) -DP, error $|Q|^{o(1)}n^{1/2}$

DRV10: (ϵ, δ) -DP, error polylog |Q| n^{1/2}

BLR08: ε-DP, error log^{1/3} |Q| n^{2/3}

DNRRV09: (ϵ, δ) -DP, error $|Q|^{o(1)}n^{1/2}$

DRV10: (ϵ, δ) -DP, error polylog |Q| n^{1/2}

HR10: (ϵ, δ) -DP, error $\log |Q| n^{1/2}$

BLR08: ϵ -DP, error $\log^{1/3} |Q| n^{2/3}$

DNRRV09: (ϵ, δ) -DP, error $|Q|^{o(1)}n^{1/2}$

DRV10: (ϵ, δ) -DP, error polylog |Q| $n^{1/2}$

HRI0: (ϵ, δ) -DP, error $\log |Q| n^{1/2}$

HLM12: simple & matches best bounds

BLR08: ϵ -DP, error $\log^{1/3} |Q| n^{2/3}$

DNRRV09: (ϵ, δ) -DP, error $|Q|^{o(1)}n^{1/2}$

DRV10: (ϵ, δ) -DP, error polylog |Q| $n^{1/2}$

HRI0: (ϵ, δ) -DP, error $\log |Q| n^{1/2}$

HLM12: simple & matches best bounds

BLR08: ϵ -DP, error $\log^{1/3} |Q| n^{2/3}$

DNRRV09: (ϵ, δ) -DP, error $|Q|^{o(1)}n^{1/2}$

DRV10: (ϵ, δ) -DP, error polylog |Q| n^{1/2}

HRI0: (ϵ, δ) -DP, error $\log |Q| n^{1/2}$

HLM12: simple & matches best bounds

Can (sometimes) do much better than naive noise addition, with much more sophisticated techniques

exponential mechanism [MT07]

select an element $C \in \text{range}(M)$ with probability $\sim \exp(\epsilon \ u(D,C)/(2\ \Delta u))$ where u is a "scoring function"

exponential mechanism [MT07]

select an element $C \in \text{range}(M)$ with probability $\sim \exp(\epsilon u(D, C)/(2 \Delta u))$ where u is a "scoring function"

privacy obvious

exponential mechanism [MT07]

select an element $C \in \text{range}(M)$ with probability $\sim \exp(\epsilon \ u(D,C)/(2\ \Delta u))$ where u is a "scoring function"

privacy obvious utility... depends

combines

- exponential mechanism [MT07] for sampling complex output space
- sample complexity bounds from learning theory to guarantee existence of good output

| name | DOB | sex | weight | smoker | lung cancer |
|---------------|---------|-----|--------|--------|----------------|
| John Doe | 12/1/51 | М | 185 | Υ | N |
| Jane Smith | 3/3/46 | F | 140 | N | N |
| Ellen Jones | 4/24/59 | F | 160 | Υ | Υ |
| Michael Ray | 3/2/81 | М | 200 | Υ | N |
| Fran Michaels | 9/9/54 | F | 155 | N | N |
| Rachel Kim | 1/21/77 | F | 130 | Υ | Υ |
| Michelle Lo | 2/2/83 | F | 135 | N | N |
| Nira Waters | 9/5/43 | F | 140 | N | Ν |
| Jennifer Kim | 3/1/70 | F | 135 | Ν | Ν |
| Lisa Smith | 9/5/43 | F | 140 | N | N |
| Tony Miller | 12/1/51 | М | 210 | Υ | Ν |
| Eve Casey | 3/3/46 | F | 140 | Ν | Ν |
| Paul Larson | 4/24/59 | F | 160 | Υ | Υ |
| Noelle Mason | 3/1/70 | F | 130 | Ν | Ν |
| Rachel Waters | 9/5/43 | IL | 140 | Υ | Z |
| Shirley Wu | 3/1/70 | F | 150 | N | N |
| Rachel Waters | 9/5/43 | F | 140 | N | Υ |
| Lawrence Vay | 12/1/51 | М | 185 | Υ | N |
| Laura Rich | 3/3/46 | F | 140 | N | N |

cancer

lung cancer

lung

lung cancer

cancer

lung cancer DOB

4/13/48

4/22/61

1/11/74

10/5/44

DOB

4/13/48

4/22/61

name

www

XXX

YYY

ZZZ

name

WWW

XXX

YYY

ZZZ

name

www

XXX

ZZZ

WWW

XXX

YYY

ZZZ

name

www

XXX

YYY

ZZZ

www

XXX

YYY

ZZZ

name

www

XXX

smoker

N

N

N

N

N

N

N

N

N

N

smoker

smoker

smoker

smoker

cancer

Ν

Ν

Ν

Ν

Ν

cancer

lung cancer

Ν

cancer

Ν

lung cancer

weight

135

165

130

150

135

165

130

150

weight

135

165

130

150

135

165

130

150

weight

135

165

130

150

135

165

130

150

weight

135

165

sex

sex

| name | DOB | sex | weight | smoker | lung | name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | |
|---------------|---------|-----|----------|--------|--------|------------|--------------------|----------------|------------|--------|----------------|------|--------------------|----------------|------------|--------|---|
| 110,1110 | | OOX | 11019111 | on one | cancer | www | 4/13/48 | F | 135 | N | N | www | 4/13/48 | F | 135 | N | |
| John Doe | 12/1/51 | М | 185 | Υ | N | YYY | 4/22/61 1/11/74 | M F | 165 130 | Y N | Y | YYY | 4/22/61 1/11/74 | M F | 165 130 | Y N | |
| | | | | | | ZZZ | 10/5/44 | F | 150 | N | N | ZZZ | 10/5/44 | F | 150 | N | j |
| Jane Smith | 3/3/46 | F | 140 | N | N | name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | |
| Ellen Jones | 4/24/59 | F | 160 | Υ | Y | www | 4/13/48 | F | 135 | N | N | www | 4/13/48 | F | 135 | N | |
| Michael Ray | 3/2/81 | М | 200 | Υ | Ν | YYY | 4/22/61 1/11/74 | M F | 165 130 | Y N | Y | YYY | 4/22/61 1/11/74 | M F | 165 130 | Y N | |
| Fran Michaels | 9/9/54 | F | 155 | N | N | ZZZ | 10/5/44 | F | 150 | N | N | ZZZ | 10/5/44 | F | 150 | N | j |
| Rachel Kim | 1/21/77 | F | 130 | Υ | Υ | name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | |
| Hadrier Killi | 1/21/17 | ' | 130 | _ | 1 | www | 4/13/48 | F | 135 | N | N | www | 4/13/48 | F | 135 | N | 4 |
| Michelle Lo | 2/2/83 | F | 135 | Ν | Ν | YYY | 4/22/61 1/11/74 | M F | 165 130 | Y N | Y | YYY | 4/22/61 1/11/74 | M F | 165 130 | Y N | ┨ |
| Nira Waters | 9/5/43 | F | 140 | N | N | ZZZ | 10/5/44 | F | 150 | N | N | ZZZ | 10/5/44 | F | 150 | N | İ |
| Jennifer Kim | 3/1/70 | F | 135 | N | N | name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | |
| | | - | | | | www | 4/13/48 | F | 135 | N | N | www | 4/13/48 | F | 135 | N | 4 |
| Lisa Smith | 9/5/43 | F | 140 | N | N | YYY | 4/22/61 1/11/74 | M F | 165 130 | Y N | Y | YYY | 4/22/61 1/11/74 | M F | 165 130 | Y N | ┨ |
| Tony Miller | 12/1/51 | М | 210 | Υ | Ν | ZZZ | 10/5/44 | F | 150 | N | N | ZZZ | 10/5/44 | F | 150 | N | İ |
| Eve Casey | 3/3/46 | F | 140 | N | N | name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | |
| Paul Larson | 4/24/59 | F | 160 | Υ | Υ | xxx | 4/13/48 4/22/61 | F M | 135 165 | N Y | N Y | xxx | 4/13/48 4/22/61 | F M | 135 165 | N Y | |
| | | _ | | | N | YYY | 1/11/74 | F | 130 | N | Υ | YYY | 1/11/74 | F | 130 | N | j |
| Noelle Mason | 3/1/70 | ۲ | 130 | N | N | ZZZ | 10/5/44 | F | 150 | N | N | ZZZ | 10/5/44 | F | 150 | N | |
| Rachel Waters | 9/5/43 | F | 140 | Υ | N | name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | |
| Shirley Wu | 3/1/70 | F | 150 | Ν | N | www xxx | 4/13/48 4/22/61 | F M | 135 165 | N Y | N Y | www | 4/13/48 4/22/61 | F M | 135 165 | N Y | |
| Rachel Waters | 9/5/43 | F | 140 | N | Υ | YYY | 1/11/74 | F | 130 | N | Υ | YYY | 1/11/74 | F | 130 | N | |
| | | | | Y | | ZZZ | 10/5/44 | ļ ^F | 150 | N | N | ZZZ | 10/5/44 | l [⊦] | 150 | N | |
| Lawrence Vay | 12/1/51 | М | 185 | T | N | name | DOB | sex | weight | smoker | cancer | name | DOB | sex | ŭ | smoker | |
| Laura Rich | 3/3/46 | F | 140 | N | N | xxx | 4/13/48 4/22/61 | F M | 135 165 | N Y | N Y | xxx | 4/13/48 4/22/61 | F M | 135 165 | N Y | |
| | | | | | | | | | | | | | | | | | i |

| name | | ~ ~ | | | Lluna |
|---------------|---------|-------|-------|-------|---------------------------|
| | Size | $O(\$ | /CDII | M(Q)/ | $\mathbf{\epsilon}^2$) [|
| John Doe | 12/1/51 | М | 185 | Υ | N |
| Jane Smith | 3/3/46 | F | 140 | N | N |
| Ellen Jones | 4/24/59 | F | 160 | Υ | Υ |
| Michael Ray | 3/2/81 | М | 200 | Υ | N |
| Fran Michaels | 9/9/54 | F | 155 | N | N |
| Rachel Kim | 1/21/77 | F | 130 | Υ | Υ |
| Michelle Lo | 2/2/83 | F | 135 | N | N |
| Nira Waters | 9/5/43 | F | 140 | N | N |
| Jennifer Kim | 3/1/70 | F | 135 | N | N |
| Lisa Smith | 9/5/43 | F | 140 | N | N |
| Tony Miller | 12/1/51 | M | 210 | Υ | N |
| Eve Casey | 3/3/46 | F | 140 | N | N |
| Paul Larson | 4/24/59 | F | 160 | Υ | Υ |
| Noelle Mason | 3/1/70 | F | 130 | N | N |
| Rachel Waters | 9/5/43 | F | 140 | Υ | N |
| Shirley Wu | 3/1/70 | F | 150 | N | N |
| Rachel Waters | 9/5/43 | F | 140 | N | Υ |
| Lawrence Vay | 12/1/51 | М | 185 | Υ | N |
| Laura Rich | 3/3/46 | F | 140 | N | N |

| | | | | | lung | | | | | | lung | | | | | | lung |
|------|------------|---------------|---------------|---------------|----------------|------|---------|-----|--------|--------|----------------|------|---------|-----|--------|--------|----------------|
| name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | lung cancer |
| www | 4/13/48 | F | 135 | N | N | www | 4/13/48 | F | 135 | N | N | www | 4/13/48 | F | 135 | N | N |
| XXX | 4/22/61 | М | 165 | Υ | Υ | xxx | 4/22/61 | М | 165 | Υ | Υ | xxx | 4/22/61 | М | 165 | Υ | Υ |
| YYY | 1/11/74 | F | 130 | N | Υ | YYY | 1/11/74 | F | 130 | N | Υ | YYY | 1/11/74 | F | 130 | N | Υ |
| ZZZ | 10/5/44 | F | 150 | N | N | ZZZ | 10/5/44 | F | 150 | N | N | ZZZ | 10/5/44 | F | 150 | N | N |
| name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | lung cancer |
| www | 4/13/48 | F | 135 | N | N | www | 4/13/48 | F | 135 | N | N | www | 4/13/48 | F | 135 | N | N |
| XXX | 4/22/61 | М | 165 | Υ | Υ | XXX | 4/22/61 | М | 165 | Υ | Υ | XXX | 4/22/61 | М | 165 | Υ | Υ |
| YYY | 1/11/74 | F | 130 | N | Υ | YYY | 1/11/74 | F | 130 | N | Υ | YYY | 1/11/74 | F | 130 | N | Υ |
| ZZZ | 10/5/44 | F | 150 | N | N | ZZZ | 10/5/44 | F | 150 | N | N | ZZZ | 10/5/44 | F | 150 | N | N |
| name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | lung cancer |
| www | 4/13/48 | F | 135 | N | N | www | 4/13/48 | F | 135 | N | N | www | 4/13/48 | F | 135 | N | N |
| XXX | 4/22/61 | М | 165 | Υ | Υ | XXX | 4/22/61 | М | 165 | Υ | Υ | XXX | 4/22/61 | М | 165 | Υ | Υ |
| YYY | 1/11/74 | F | 130 | N | Υ | YYY | 1/11/74 | F | 130 | N | Υ | YYY | 1/11/74 | F | 130 | N | Υ |
| ZZZ | 10/5/44 | F | 150 | N | N | ZZZ | 10/5/44 | F_ | 150 | N | N | ZZZ | 10/5/44 | F | 150 | N | N |
| name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | lung cancer |
| www | 4/13/48 | F | 135 | N | N | www | 4/13/48 | F_ | 135 | N | N | www | 4/13/48 | F | 135 | N | N |
| XXX | 4/22/61 | М | 165 | Υ | Υ | XXX | 4/22/61 | М | 165 | Υ | Υ | XXX | 4/22/61 | М | 165 | Υ | Υ |
| YYY | 1/11/74 | F | 130 | N | Υ | YYY | 1/11/74 | F_ | 130 | N | Υ | YYY | 1/11/74 | F | 130 | N | Υ |
| ZZZ | 10/5/44 | F | 150 | N | N | ZZZ | 10/5/44 | F | 150 | N | N | ZZZ | 10/5/44 | F | 150 | N | N |
| name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | lung cancer |
| www | 4/13/48 | F | 135 | N | N | www | 4/13/48 | F | 135 | N | N | www | 4/13/48 | F | 135 | N | N |
| XXX | 4/22/61 | М | 165 | Υ | Υ | XXX | 4/22/61 | М | 165 | Υ | Υ | XXX | 4/22/61 | М | 165 | Υ | Υ |
| YYY | 1/11/74 | F | 130 | N | Υ | YYY | 1/11/74 | F | 130 | N | Υ | YYY | 1/11/74 | F | 130 | N | Υ |
| ZZZ | 10/5/44 | F | 150 | N | N | ZZZ | 10/5/44 | F | 150 | N | N | ZZZ | 10/5/44 | F | 150 | N | N |
| name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | lung cancer |
| www | 4/13/48 | F_ | 135 | N | N | www | 4/13/48 | F | 135 | N | N | www | 4/13/48 | F | 135 | N | N |
| XXX | 4/22/61 | М | 165 | Υ | Υ | xxx | 4/22/61 | М | 165 | Υ | Υ | XXX | 4/22/61 | М | 165 | Υ | Υ |
| YYY | 1/11/74 | F | 130 | N | Υ | YYY | 1/11/74 | F_ | 130 | N | Υ | YYY | 1/11/74 | F | 130 | N | Υ |
| ZZZ | 10/5/44 | F | 150 | N | N | ZZZ | 10/5/44 | F_ | 150 | N | N | ZZZ | 10/5/44 | F | 150 | N | N |
| name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | lung cancer |
| www | 4/13/48 | F / | 135 | N | N | www | 4/13/48 | F | 135 | N | N | www | 4/13/48 | F | 135 | N | N |
| | lacksquare | ightharpoonup | $\overline{}$ | $\overline{}$ | = | = | | | | | | | | | | | |

| name | | ~ | | | Lluna |
|---------------|---------|-------|-------|-------|-------------------|
| Hamo | Size | $O(\$ | /CDII | M(Q)/ | ε^2) |
| John Doe | 12/1/51 | М | 185 | Υ | N |
| Jane Smith | 3/3/46 | F | 140 | N | N |
| Ellen Jones | 4/24/59 | F | 160 | Υ | Υ |
| Michael Ray | 3/2/81 | М | 200 | Υ | N |
| Fran Michaels | 9/9/54 | F | 155 | Ν | N |
| Rachel Kim | 1/21/77 | F | 130 | Υ | Υ |
| Michelle Lo | 2/2/83 | F | 135 | N | N |
| Nira Waters | 9/5/43 | F | 140 | N | N |
| Jennifer Kim | 3/1/70 | F | 135 | N | N |
| Lisa Smith | 9/5/43 | F | 140 | N | N |
| Tony Miller | 12/1/51 | М | 210 | Υ | N |
| Eve Casey | 3/3/46 | F | 140 | N | N |
| Paul Larson | 4/24/59 | F | 160 | Υ | Υ |
| Noelle Mason | 3/1/70 | F | 130 | N | N |
| Rachel Waters | 9/5/43 | F | 140 | Υ | N |
| Shirley Wu | 3/1/70 | F | 150 | N | N |
| Rachel Waters | 9/5/43 | F | 140 | N | Υ |
| Lawrence Vay | 12/1/51 | М | 185 | Υ | N |
| Laura Rich | 3/3/46 | F | 140 | N | N |

| name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | lung cancer |
|------|---------|-----|---------|--------|----------------|------|---------|-----|----------|--------|----------------|------|---------|-----|--------|--------|----------------|
| www | 4/ | | | N | N | www | 4/1 | | | ٧ | N | www | 4/13/48 | | | | N |
| XXX | 4/ | -4 | 4 | Υ | Υ | xxx | 4/2 | _(| 6 | 1 | Υ | xxx | 4/22/61 | | -2 | | Υ |
| YYY | 1/ | | | N | Υ | YYY | 1/1 | _ | | ٧ | Υ | YYY | 1/11/74 | | | | Υ |
| ZZZ | 10/5/44 | F | 150 | N | N | ZZZ | 10/5/44 | F | 150 | N | N | ZZZ | 10/5/44 | F | 150 | N | N |
| name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | lung cancer |
| www | 4/ | | _ | N | N | www | 4/1 | | _ | ٧ | N | www | 4/13/48 | | | | N |
| XXX | 4/ | -: | 3 | Υ | Υ | xxx | 4/2 | _ | 1 | 1 | Υ | xxx | 4/22/61 | | -6 | | Υ |
| YYY | 1/ | _ | <i></i> | N | Υ | YYY | 1/1 | | | ٧ | Υ | YYY | 1/11/74 | | | | Υ |
| ZZZ | 10/5/44 | F | 150 | N | N | ZZZ | 10/5/44 | F | 150 | N | N | ZZZ | 10/5/44 | F | 150 | N | N |
| name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | lung cancer |
| www | 4/ | | | N | N | www | 4/1 | | _ | ٧ | N | www | 4/13/48 | | | | N |
| XXX | 4/ | | 1 | Υ | Υ | xxx | 4/2 | _ | R | 1 | Υ | xxx | 4/22/61 | | -6 | | Υ |
| YYY | 1/ | | ' | N | Υ | YYY | 1/1 | | | ٧ | Υ | YYY | 1/11/74 | | | | Υ |
| ZZZ | 10/5/44 | F | 150 | N | N | ZZZ | 10/5/44 | F | 150 | N | N | ZZZ | 10/5/44 | F | 150 | N | N |
| name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | lung cancer |
| www | 4/ | | | N | N | www | 4/1 | | | ٧ | N | www | 4/13/48 | | | | N |
| XXX | 4/ | | 7 | Υ | Υ | XXX | 4/2 | _(| 9 | 1 | Υ | xxx | 4/22/61 | | -7 | | Υ |
| YYY | 1/ | | 1 | N | Υ | YYY | 1/1 | | <u> </u> | ٧ | Υ | YYY | 1/11/74 | | | | Υ |
| ZZZ | 10/5/44 | F | 150 | N | N | ZZZ | 10/5/44 | F | 150 | N | N | ZZZ | 10/5/44 | F | 150 | N | N |
| name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | lung cancer |
| www | 4/ | | | N | N | www | 4/1 | | | N | N | www | 4/13/48 | | | | N |
| XXX | 4/ | _4 | 1 | Υ | Υ | XXX | 4/2 | _ | 6 | 1 | Υ | XXX | 4/22/61 | | -8 | | Υ |
| YYY | 1/ | | Т | N | Υ | YYY | 1/1 | · · | | ٧ | Υ | YYY | 1/11/74 | | | | Υ |
| ZZZ | 10/5/44 | F | 150 | N | N | ZZZ | 10/5/44 | F | 150 | N | N | ZZZ | 10/5/44 | F | 150 | N | N |
| name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | lung cancer |
| www | 4/10/10 | | 105 | N | N | www | 4/10/40 | | 105 | N | N | www | 4/13/48 | | 105 | | N |
| XXX | 4/ | _5 | 2 | Υ | Υ | XXX | 4/2 | _ | Q | 1 | Υ | XXX | 4/22/61 | | _7 | | Υ |
| YYY | 1/ | _(| J | N | Υ | YYY | 1/1 | | U | ٧ | Υ | YYY | 1/11/74 | | _ / | | Υ |
| ZZZ | 10/5/44 | F | 150 | N | N | ZZZ | 10/5/44 | F | 150 | N | N | ZZZ | 10/5/44 | F | 150 | N | N |
| name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | lung cancer |
| www | 4/10/40 | _ | 105 | N | N | www | 4/10/49 | F | 105 | 7 | N | www | 4/13/48 | | 105 | N. | N |
| XXX | 4/ | _6 | 3 | Υ | Υ | XXX | 4/2 | _ ' | 5 | 1 | Υ | XXX | 4/22/61 | | _Q | | Υ |

smoker

150

Ν

lung cancer

N

N

Ν

lung cancer

N

_Q

smoker

| | | | | | luus er | | DOD | | 2.10 | | lung | | DOD | | 25.14 | | lung | | DOD |
|---------------|-----------------------|------------------|---------------|--------|---------|-------|---------|-----|----------|--------|----------------|------|------------|-----|----------|----------|----------------|------|--------------------|
| name | 0: | Ã۸ | | 1/(0)/ | -2) | name | DOB A/ | sex | weight | smoker | cancer | name | DOB 4/1 | sex | weight | smoker | cancer | name | DOB 4/13/48 |
| | Size | U(1 | /UUII | V(Q)/Q | E^) | XXX | 4/ | -4 | 4 | Y | Y | XXX | 4/2 | _ | 6 | · | Y | XXX | 4/22/61 |
| John Doe | 12/1/51 | M, | 185 | Υ | N | YYY | 1/ | | _ | N | Υ | YYY | 1/1 | _ | | ٧ | Υ | YYY | 1/11/74 |
| Jane Smith | 3/3/46 | | 140 | N | N | ZZZ | 10/5/44 | F | 150 | N | N | ZZZ | 10/5/44 | F | 150 | N | N | ZZZ | 10/5/44 |
| | | / \ | <u> </u> | IN A | IN | name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | lung cancer | name | DOB |
| Ellen Jones | 4/24/59 | | 460 | | Υ | www | 4/ | | <u> </u> | N | N | www | 4/1 | | 4 | ٧ | N | www | 4/13/48 |
| Michael Ray | | | | | N | YYY | 1/ | -, | 3 | Y | Y | YYY | 4/2 1/1 | _ | ı | <u>/</u> | Y | YYY | 4/22/61 1/11/74 |
| | | | emokor lung | | NI | 777 | 10/5/44 | F | 150 | N | N | ZZZ | 10/5/44 | F | 150 | N | N | ZZZ | 10/5/44 |
| Fran Michaele | name DOB WWW 4/13/48 | sex weight F 135 | smoker cancer | | | name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | lung cancer | name | DOB |
| Rachel | XXX 4/22/61 | M 165 | Y Y | | | MAAAA | 4/ | | | N | N | www | 4/1 | | | N | N | www | 4/13/48 |
| Michelle Lo | YYY 1/11/74 | F 130 | N Y | | N | XXX | 4/ | | 1 | Υ | Υ | XXX | 4/2 | - | 8 | 1 | Υ | XXX | 4/22/61 |
| | ZZZ 10/5/44 | F 150 | N N | | 1 4 | YYY | 1/ | - | | N | Y | YYY | 1/1 | _ | | ٧ | Y | YYY | 1/11/74 |
| Nira Wa | | | | | | ZZZ | 10/5/44 | F | 150 | N | N | ZZZ | 10/5/44 | F | 150 | N | N | ZZZ | 10/5/44 |
| Jennifer Kim | 3/1 | | | N | N | name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | lung cancer | name | DOB |
| | | F | | NI | NI | xxx | 4/ | - | 7 | N | N Y | XXX | 4/1 4/2 | | 9 | ٧ | N | xxx | 4/13/48 4/22/61 |
| Lisa Smith | 9/5 | Г | 146 | IN | IN | YYY | 1/ | - 1 | | N | Y | YYY | 1/1 | ; | 9 | ١ | Υ | YYY | 1/11/74 |
| Tony Miller | 12/1/51 | М | 210 | Υ | Ν | ZZZ | 10/5/44 | F | 150 | N | N | ZZZ | 10/5/44 | F | 150 | N | N | ZZZ | 10/5/44 |
| Eve Casey | 3/3/46 | F | 140 | N | N | name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | lung cancer | name | DOB |
| Paul Larson | 4/24/59 | Е | 160 | V | V | www | 4/ | | 1 | N | N Y | www | 4/1 4/2 | | <u> </u> | ٧ | N | xxx | 4/13/48 4/22/61 |
| T aut Latsott | 4/24/33 | | 100 | 1 | 1 | YYY | 1/ | -4 | 4 | N | Y | YYY | 1/1 | _ | O | ٧ | Y | YYY | 1/11/74 |
| Noelle Mason | 3/1/70 | F | 130 | Ν | N | ZZZ | 10/5/44 | F | 150 | N | N | ZZZ | 10/5/44 | F | 150 | N | N | ZZZ | 10/5/44 |
| Rachel Waters | 9/5/43 | F | 140 | Υ | N | name | DOB | sex | weight | smoker | lung cancer | name | DOB | sex | weight | smoker | lung cancer | name | DOB |
| Shirley Wu | 3/1/70 | F | 150 | N | N | www | 4/10/10 | | | N | N | | 4/10/40 | | | 7 | N | | _ |
| Office VVa | 3/1/70 | ' | 130 | 1 4 | 1 4 | YYY | 4/ | -{ | 3 | Y | Y | YYY | 4/2 | - | 8 | 1 | Y | YYY | 4/22/61 1/11/74 |
| Rachel Waters | 9/5/43 | F | 140 | N | Υ | ZZZ | 10/5/44 | | | N | N | ZZZ | 10/5/44 | | | N | N | ZZZ | 10/5/44 |
| Lawrence Vay | 12/1/51 | M | 185 | Υ | N | name | | | weight | smoker | lung cancer | name | | | | smoker | lung cancer | name | DOB |
| Laura Rich | 3/3/46 | F | 140 | N | Ν | www | 4/10/40 | _ | 105 | N | N | www | 4/10/40 | - | 105 | N | N | www | 4/13/48 |
| Laura Hilli | 0/0/40 | 1 | 170 | 1 4 | TV | XXX | 4/ | _6 | 3 | Υ | Υ | XXX | 4/2 | _ | 5 | 1 | Υ | XXX | 4/22/61 |

[HLM12]

- simple to describe and to implement
- actually implemented and tested it
- state of the art in theory, performs well in practice (and quickly, despite bad worstcase news)

[HLM12]

- simple to describe and to implement
- actually implemented and tested it
- state of the art in theory, performs well in practice (and quickly, despite bad worstcase news)

...will hear more about multiplicative weightsbased techniques in Jon's talk I have to know all my queries in advance?!

interactive mechanisms

- so far, have discussed creating synthetic data where must know query set in advance
- tools exist to answer similar number of queries on the fly (correlating randomness across queries)

It seems like DP would add too much noise for my application.

It seems like DP would add too much noise for my application.

... stop and think about what this means

DP connected to *robustness* of computation to presence or absence of individuals

DP connected to *robustness* of computation to presence or absence of individuals

computation not robust? (should worry!)

DP connected to *robustness* of computation to presence or absence of individuals

- computation not robust? (should worry!)
- need more data (individuals) to get desired privacy-utility tradeoff (should think)

DP connected to *robustness* of computation to presence or absence of individuals

- computation not robust? (should worry!)
- need more data (individuals) to get desired privacy-utility tradeoff (should think)
- expect on "real" data will be robust (we can do something about this!)

robustness (an aside)

- robustness in DP sense not identical to statistical robustness---DP is worst-case rather than wrt to distribution
- there are connections (will mention shortly)

expect study robust on actual data

• idea I: bootstrapping

bootstrap aggregation

- given training dataset, create many new training sets of smaller size by sampling uniformly with replacement
- fit your model (estimate your statistic) on each
- combine (e.g., voting, averaging)

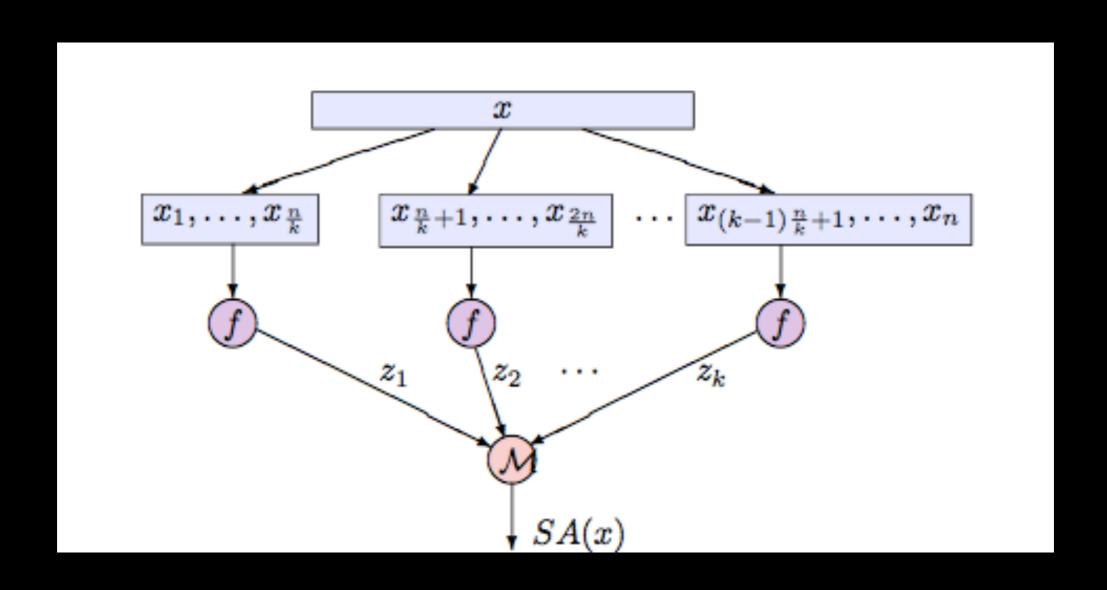
bootstrap intuition

- a "robust" statistic should be stable on most reasonbly sized subsets of the data
- if statistic is somewhat unstable, aggregation "smooths" result
- if statistic was very stable, loss in precision should be small

bootstrap for privacy

- if function is not low-sensitivity but suspect it's usually stable, not clear how to guarantee DP
- if aggregation preserves privacy, get DP guarantee even when aggregating non-DP estimates

bootstrap for privacy = subsample and aggregate [NRS07]



subsample and aggregate: good news

- can use any DP aggregation function (as long as choice doesn't depend on data)
- private aggregation just requires adding noise scaled to sensitivity of the aggregation function
- privacy is immediate!

subsample and aggregate: bad news

- may be difficult to bound worst-case sensitivity of aggregation function
- default bound is max of its range (quite bad)
- may be difficult to get good utility guarantees

subsample and aggregate: applications

 underlying function might be selecting best model from among set of m options; could aggregate with a noisy max

subsample and aggregate: applications

- underlying function might be selecting best model from among set of m options; could aggregate with a noisy max
- similarly, could output a set of significant features (as for LASSO)

expect study robust on actual data

- idea I: bootstrapping
- idea 2: check robustness before proceeding

check robustness

- would like to be able to test in DP manner whether computation "should" proceed
- "should": e.g., whether desired function is robust (low-sensitivity) on actual data
- if not, halt

local sensitivity of function f on database D

$$\max_{D'} |f(D) - f(D')|_1$$

for D' neighboring data set of D

local sensitivity of function f on database D

$$\max_{D'} |f(D) - f(D')|_1$$

for D' neighboring data set of D

 measures how much one person can affect output on this data

• propose a bound on local sensitivity

- propose a bound on local sensitivity
- test in DP manner whether actual data satisfies bound

- propose a bound on local sensitivity
- test in DP manner whether actual data satisfies bound
- if fails, halt

- propose a bound on local sensitivity
- test in DP manner whether actual data satisfies bound
- if fails, halt
- if passes, release function with noise tailored to local sensitivity

 test could be "what is L_I distance to closest database that fails local sensitivity bound?"

- test could be "what is L_I distance to closest database that fails local sensitivity bound?"
- test only has sensitivity I

- test could be "what is L_I distance to closest database that fails local sensitivity bound?"
- test only has sensitivity I
- can use conservative local sensitivity threshold

- test could be "what is L_I distance to closest database that fails local sensitivity bound?"
- test only has sensitivity I
- can use conservative local sensitivity threshold
- could still sometimes fail; can't get ε-DP

DP output need not be noisy!

DP output need not be noisy!

PTR can be used to privately check whether distance to nearest unstable data set is far, and if so release the *true* f(x)

robustness, revisited

robustness, revisited

 robustness wrt adding/removing a few points from dataset

robustness, revisited

- robustness wrt adding/removing a few points from dataset
- robustness wrt subsampling

subsample & aggregate + propose-test-release

[ST13]: can modify subsample & aggregate so that outputs true f(x) with high probability when f is subsampling stable on x

subsample & aggregate + propose-test-release

[STI3]: can modify subsample & aggregate so that outputs true f(x) with high probability when f is subsampling stable on x

shows that DP model selection only increases sample complexity of model selection by $O(\log{(1/\delta)/\epsilon})$

 connections to robustness; interquartile distance, median, linear regression [DworkLei09]

- connections to robustness; interquartile distance, median, linear regression [DworkLei09]
- M-estimators [Leil I, NekipelovYakovlev I I]

- connections to robustness; interquartile distance, median, linear regression [DworkLei09]
- M-estimators [Leil I, NekipelovYakovlev I I]
- for almost any estimator that is asymptotically normal on i.i.d. samples, DP adds asymptotically no additional perturbation [Smith I I]

- connections to robustness; interquartile distance, median, linear regression [DworkLei09]
- M-estimators [Leill, NekipelovYakovlevll]
- for almost any estimator that is asymptotically normal on i.i.d. samples, DP adds asymptotically no additional perturbation [Smith I I]
- convergence rate of DP estimators tied to gross error sensitivity
 [ChaudhuriHsul2]

- connections to robustness; interquartile distance, median, linear regression [DworkLei09]
- M-estimators [Leil I, NekipelovYakovlev I I]
- for almost any estimator that is asymptotically normal on i.i.d. samples, DP adds asymptotically no additional perturbation [Smith I I]
- convergence rate of DP estimators tied to gross error sensitivity
 [ChaudhuriHsul2]
- minimizing convex loss functions [ChaudhuriMonteleoniSarwate II, Rubinstein et al. 2012, KiferSmithThakurta I2]

DP statistics

- connections to robustness; interquartile distance, median, linear regression [DworkLei09]
- M-estimators [Leil I, NekipelovYakovlev I I]
- for almost any estimator that is asymptotically normal on i.i.d.
 samples, DP adds asymptotically no additional perturbation [Smith I I]
- convergence rate of DP estimators tied to gross error sensitivity
 [ChaudhuriHsul2]
- minimizing convex loss functions [ChaudhuriMonteleoniSarwate II, Rubinstein et al. 2012, KiferSmithThakurta I2]
- model selection [SmithThakurta 13]

DP statistics

- connections to robustness; interquartile distance, median, linear regression [DworkLei09]
- M-estimators [Leil], NekipelovYakovlev[]
- for almost any estimator that is asymptotically normal on i.i.d. samples, DP adds asymptotically no additional perturbation [Smith I I]
- convergence rate of DP estimators tied to gross error sensitivity
 [ChaudhuriHsul2]
- minimizing convex loss functions [ChaudhuriMonteleoniSarwate II, Rubinstein et al. 2012, KiferSmithThakurta I2]
- model selection [SmithThakurta 13]
- empirical investigations [VuSlavkovic09, ChaudhuriMonteleoniSarwate I I, AbowdSchneider Vilhuber I 3]

not DP

not DP

interactive (or hybrid interactive/ noninteractive) mechanisms?

not DP

interactive (or hybrid interactive/ noninteractive) mechanisms?

big data force us to formalize "looking at the data"

 privacy easy to get wrong; DP provides compelling definition and useful dose of paranoia

- privacy easy to get wrong; DP provides compelling definition and useful dose of paranoia
- powerful tools exist (some with no cost of privacy, and some with no noise!)

- privacy easy to get wrong; DP provides compelling definition and useful dose of paranoia
- powerful tools exist (some with no cost of privacy, and some with no noise!)
- powerful intuition from notions of robustness

- privacy easy to get wrong; DP provides compelling definition and useful dose of paranoia
- powerful tools exist (some with no cost of privacy, and some with no noise!)
- powerful intuition from notions of robustness
- many nearly ready (and quite relevant) to common big data applications

- privacy easy to get wrong; DP provides compelling definition and useful dose of paranoia
- powerful tools exist (some with no cost of privacy, and some with no noise!)
- powerful intuition from notions of robustness
- many nearly ready (and quite relevant) to common big data applications
- no ready-to-use, commercial- grade applications:
 need demand!

Differential Privacy Tutorial

Katrina Ligett katrina@caltech.edu

"Privacy and Data-Based Research," with Ori Heffetz.

Available on SSRN.