

# DIFFERENTIAL PRIVACY: A CONCISE TUTORIAL

**Kobbi Nissim**

Georgetown University



**Workshop on Algorithmic Challenges in Protecting Privacy for Biomedical Data**  
IPAM, UCLA, January 10, 2018

# THIS “CONCISE” TUTORIAL

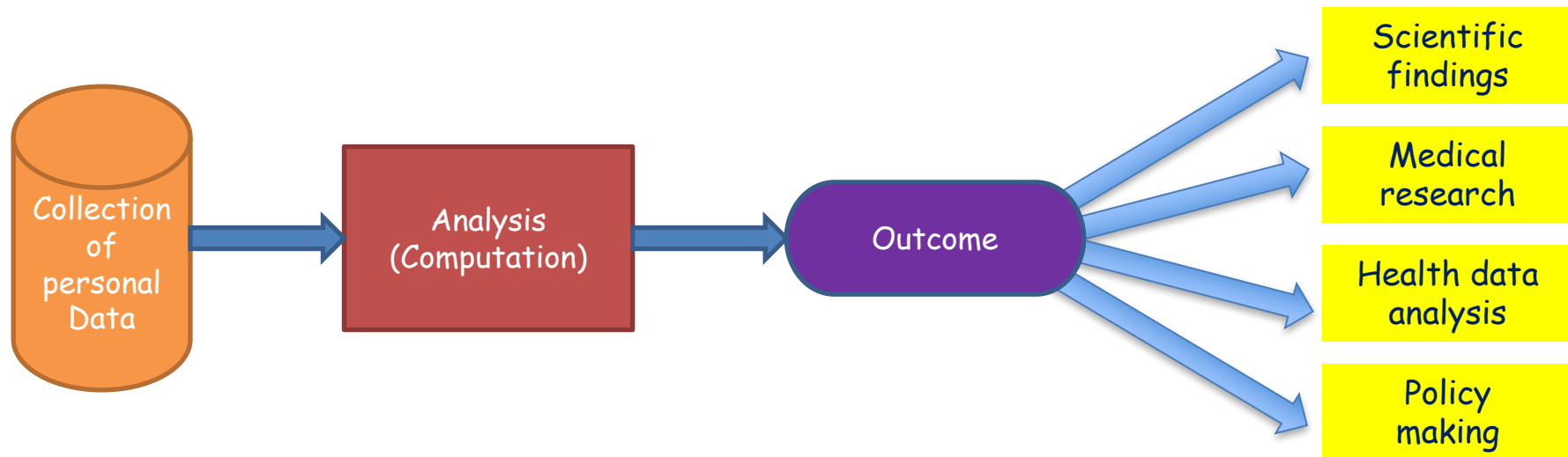
- *Will not* make you a differential privacy expert



# THIS “CONCISE” TUTORIAL

- *Will not* make you a differential privacy expert
- Will provide background for our discussions and breakout sessions
- Main Topics:
  - How privacy concepts can fail
  - Differential privacy:
    - Understanding the concept, basic properties
  - What tasks can be performed with differential privacy?
  - Real-world implementations
  - Challenges in bringing differential privacy to practice
  - Statistical validity (if time permits)
- List of resources

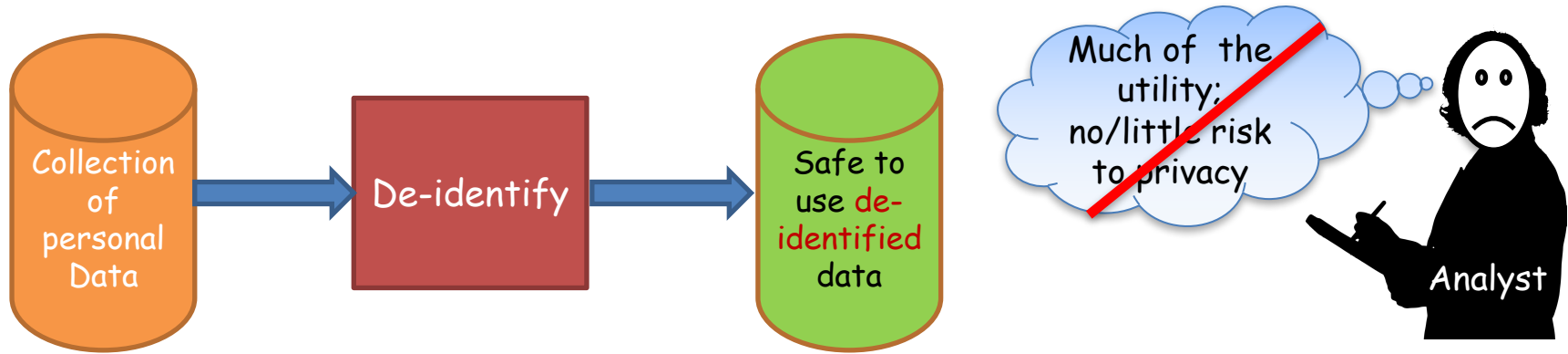
# DATA PRIVACY: THE PROBLEM



Given a dataset with sensitive personal information, how can we compute and release functions of the dataset while protecting individual privacy?

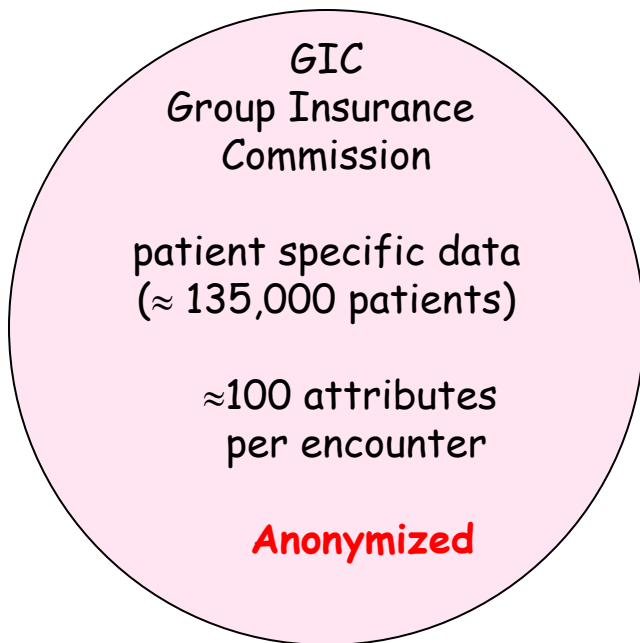
# THE ANONYMIZATION DREAM

A common intuitive idea: **anonymization/de-identification**



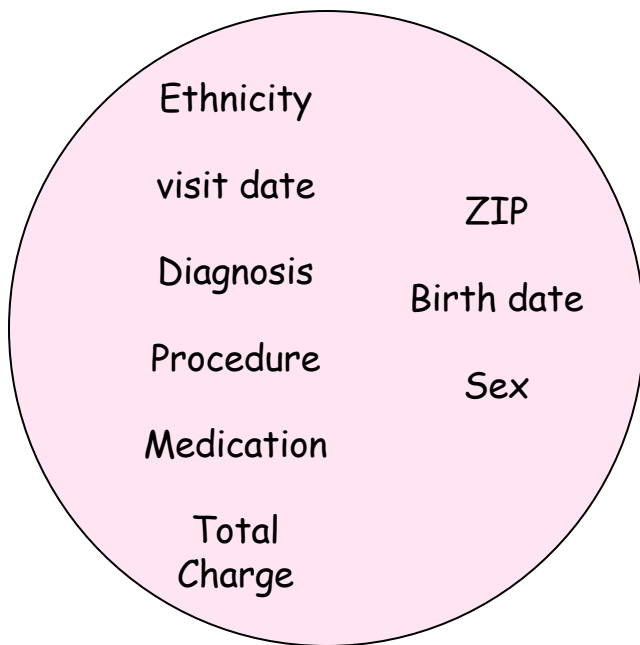
- A trusted data curator removes “identifiers” (SSNs, names, addresses, ...) to get anonymity and hence privacy.
- **Science and practice have shown this to be often wrong.**

# GIC LINKAGE ATTACK



Linkage attacks [Sweeney '00]

# GIC LINKAGE ATTACK



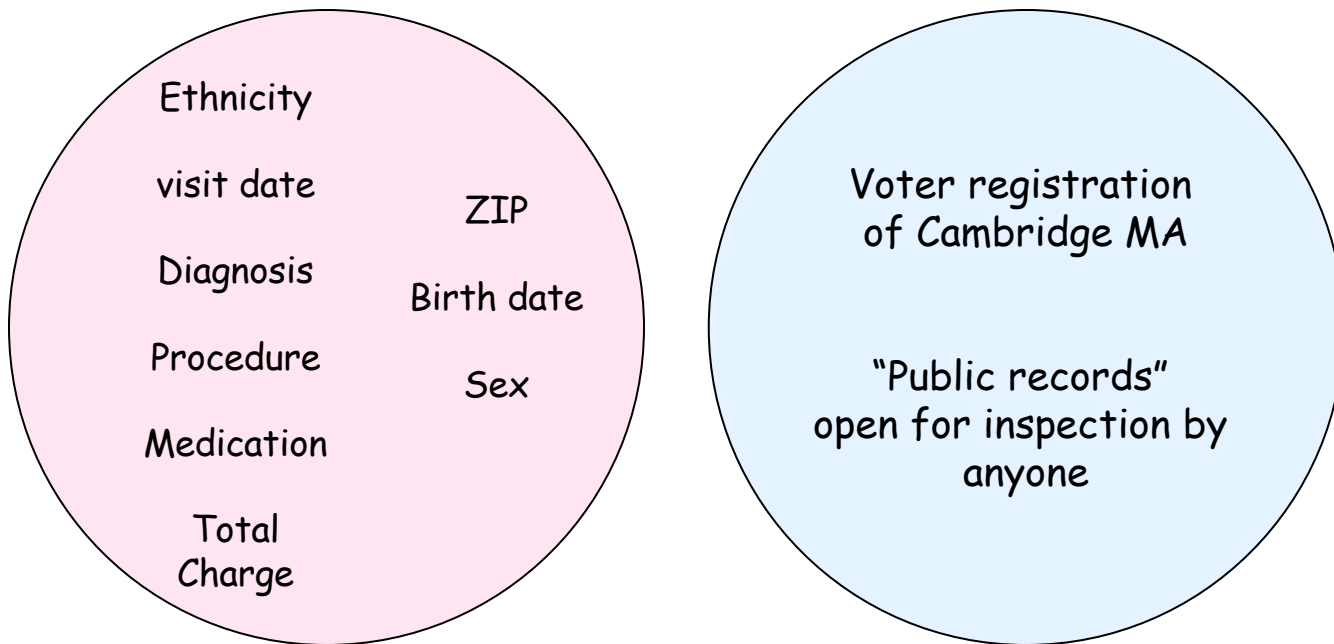
Health related data.  
Can be useful for researchers,  
policy makers, etc.



Linkage attacks [Sweeney '00]

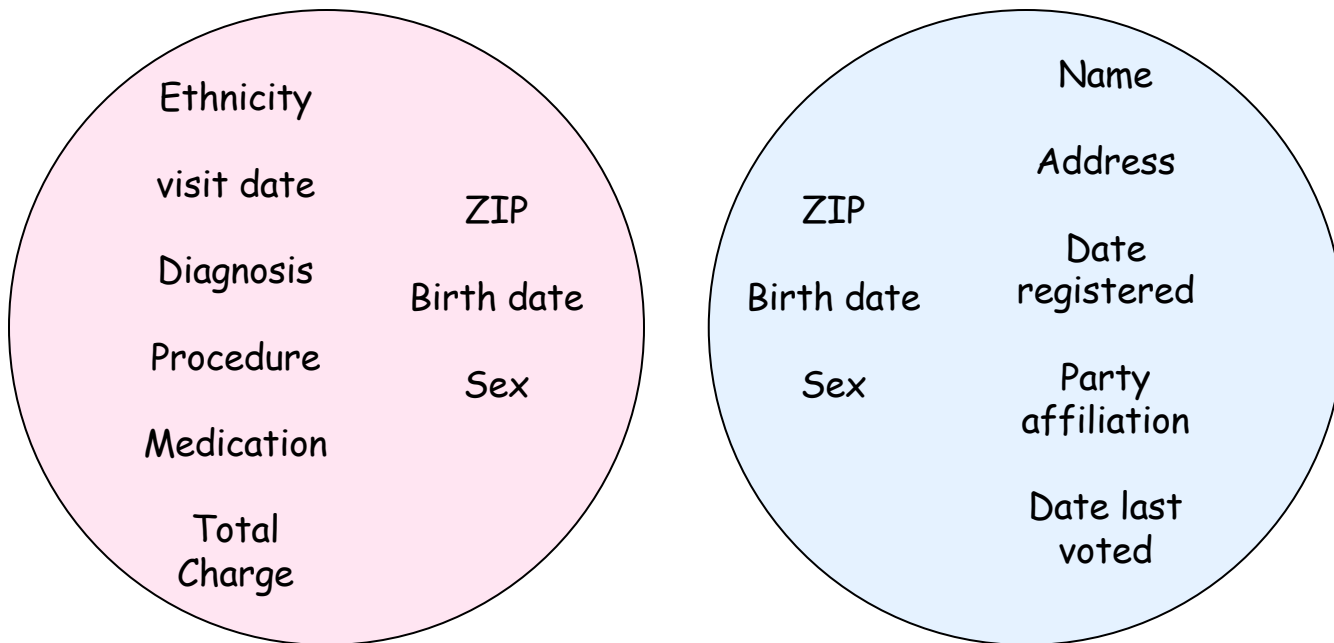


# GIC LINKAGE ATTACK



Linkage attacks [Sweeney '00]

# GIC LINKAGE ATTACK



Linkage attacks [Sweeney '00]

# GIC LINKAGE ATTACK

- As an example, Sweeney re-identified the medical records of William Weld (governor of Massachusetts at the time)
  - According to the Cambridge Voter list:
    - Six people had his particular birth date
    - Of which three were men
    - He was the only one in his 5-digit ZIP code!
- A common phenomenon (1):
  - dob+5zip → re-identify 69% of Americans
  - dob+9zip → re-identify 97% of Americans
- A common phenomenon (2):
  - Health data, clinical trial data, DNA, Pharmacy data, text data, registry information, ...



# LESSONS LEARNED

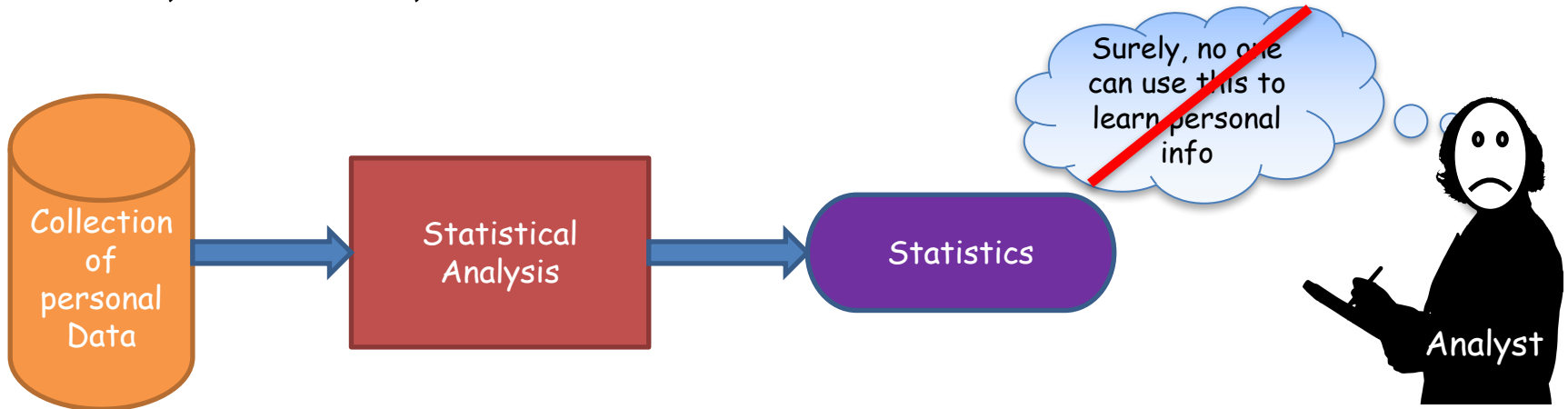
Supposedly de-identified data often contain alternative ways of identification (a.k.a. quasi identifiers)

Access to the appropriate auxiliary information can then result in *re-identification*

This is not 'purely theoretical' but has been demonstrated with real-world de-identified datasets

# THE AGGREGATION DREAM

A common intuitive idea: However, counts, averages, statistical models, classifiers, ... are safe



- Science and practice have shown this to be often wrong

# GWAS MEMBERSHIP ATTACK

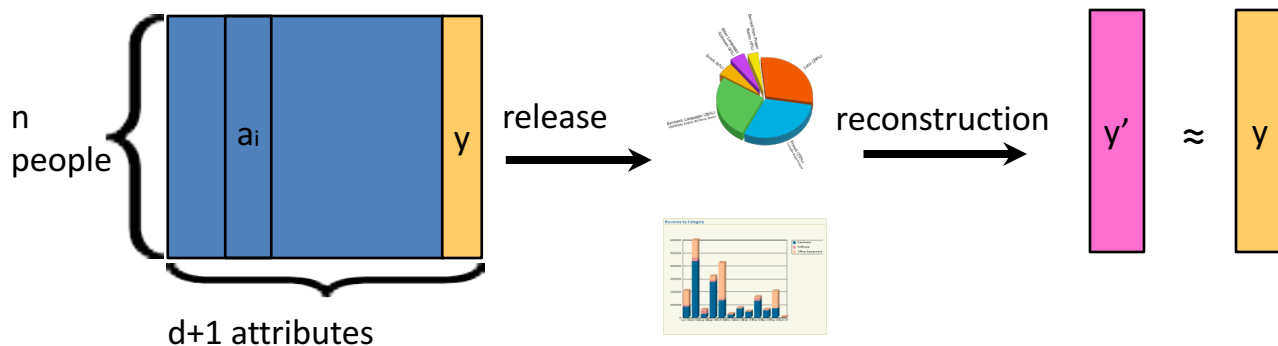
- NIH collects DNA of individuals with a disease; Publishes minor allele frequencies at 100,000 positions (SNPs).
  - Release is Innocuously looking; HIPAA compliant; NIH is trusted
- **Given an individual's DNA can infer whether her data was used in study**

Aggregate info, test group:	5%	3%	12%							3%
Victim's genome:	1	0	0							1
Aggregate info, population:	3%	3%	11%							3%

- **Result:** NIH took down some datasets.

[Homer et al. Dwork et al.]

# RECONSTRUCTION ATTACKS



- Data set:  $d$  “public” attributes per person, one “sensitive” attributes.
- Suppose release reveals rough correlations between attributes.
  - Correlations reveal inner products  $\langle a_i, y \rangle + error$ .
- **Theorem [DiNi03]:** If  $error = o(\sqrt{n})$ ,  $d \gg n$ , and  $a_i$  are uniformly random then attacker can reconstruct all but  $o(n)$  bits of  $y$ .
- Reconstruction attacks have been devised for other settings.
- These teach us about what cannot be performed under *any reasonable notion of privacy*.



# LESSONS LEARNED

- New attack modes emerge as research progresses.
- Redaction of identifiers, release of aggregates, etc. is insufficient.
- Must take auxiliary information into consideration.
- Lack of rigor leads to unanticipated privacy failures.
- Any useful analysis of personal data must leak some information about individuals.
- Leakages accumulate with multiple analyses/releases.

Mathematical  
facts, not  
matters of  
policy



# **DIFFERENTIAL PRIVACY**

Differential privacy is a **definition** (i.e., standard)  
of privacy

Not a specific technique or algorithm!

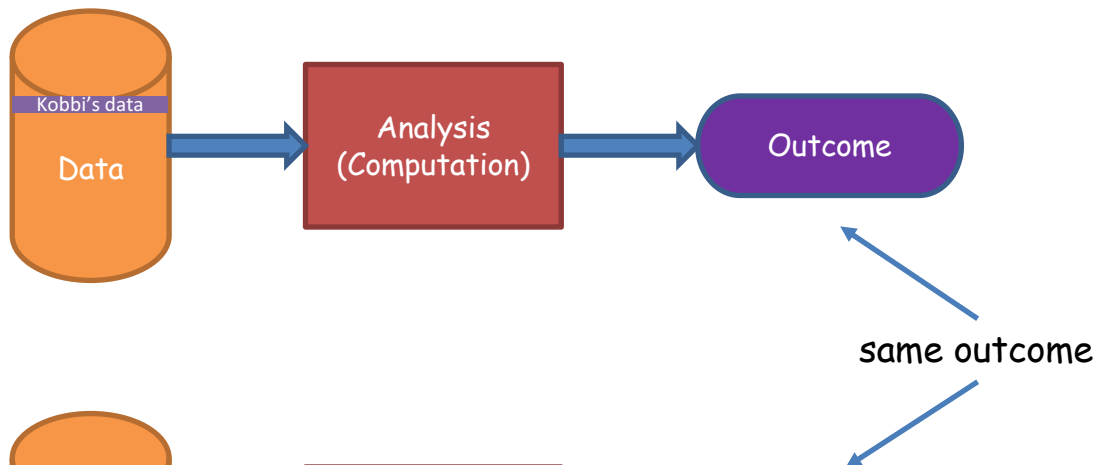
Differential privacy is a **definition** (i.e., standard) of privacy

It expresses a specific desiderata of an *analysis*:

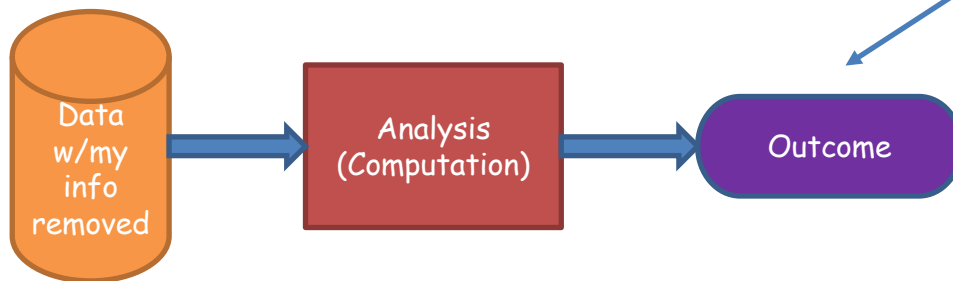
Any information-related risk to a person should not change significantly as a result of that person's information being included, or not, in the analysis.

# A PRIVACY DESIDERATA

Real world:

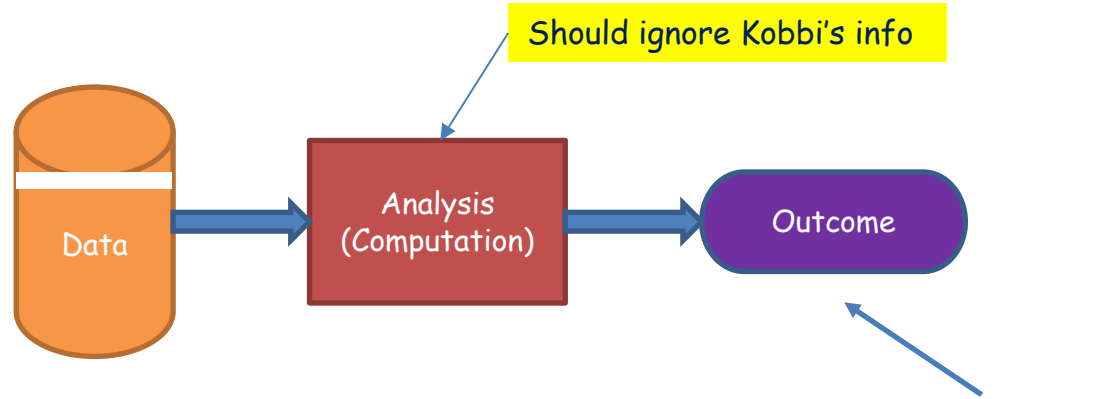


My ideal world:

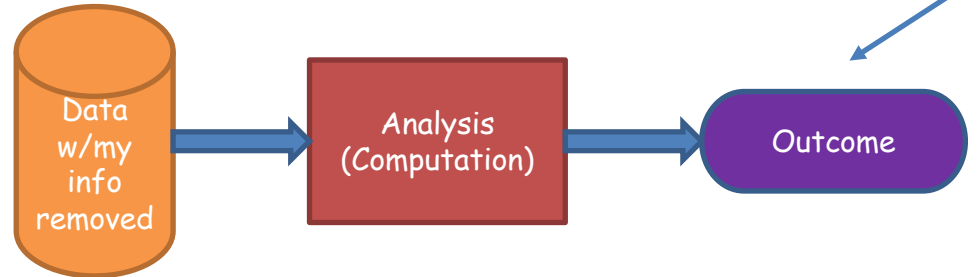


# A PRIVACY DESIDERATA

Real world:

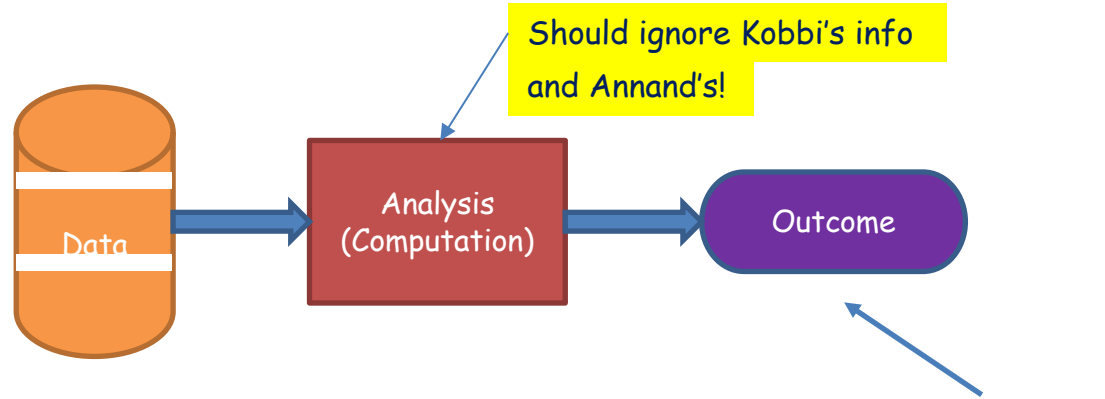


My ideal world:

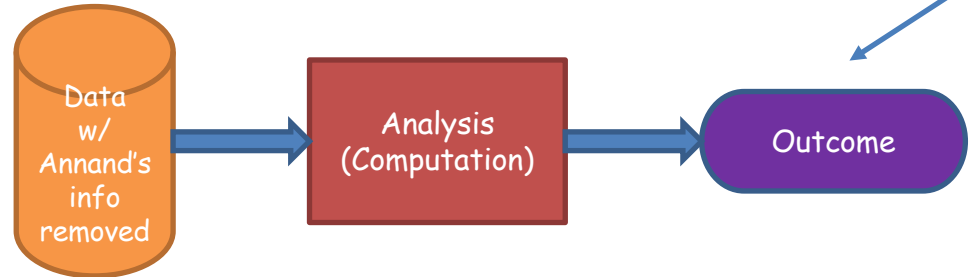


# A PRIVACY DESIDERATA

Real world:



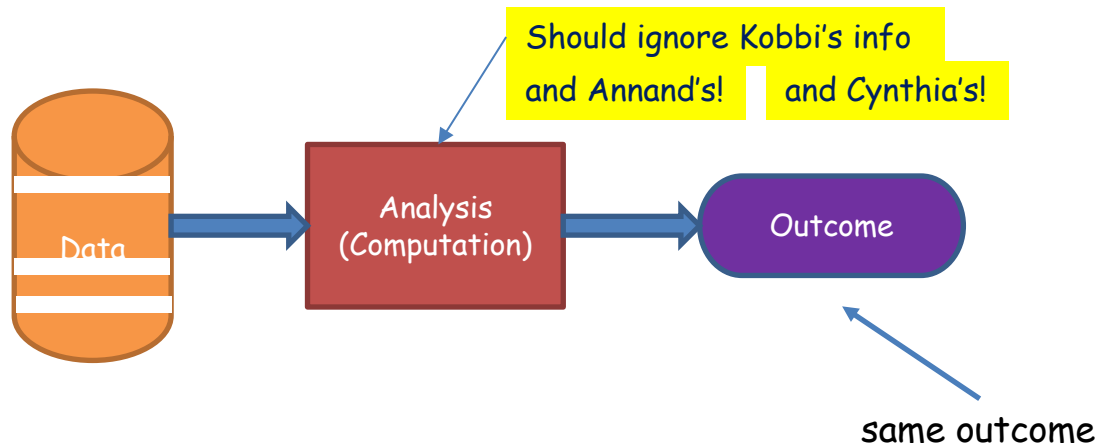
Annand's ideal world:



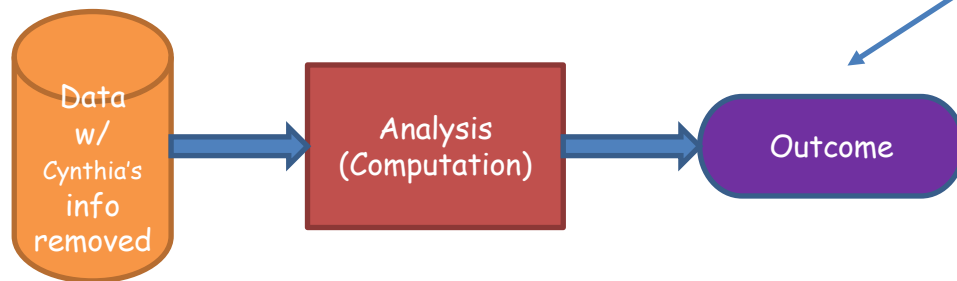
same outcome

# A PRIVACY DESIDERATA

Real world:

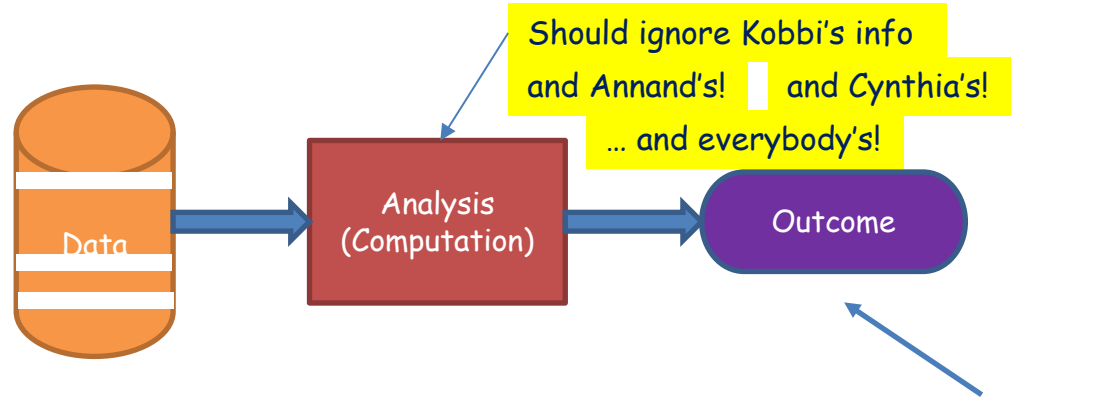


Cynthia's ideal world:

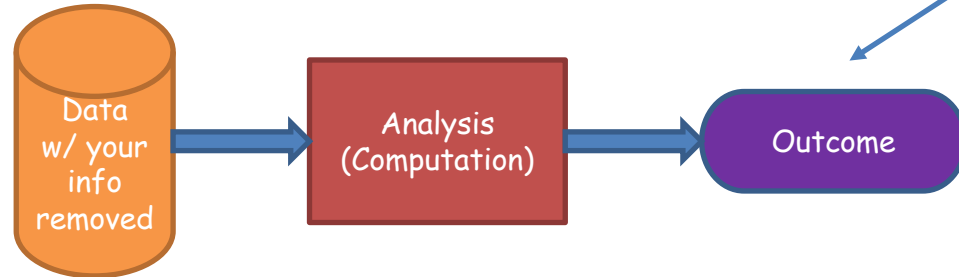


# A PRIVACY DESIDERATA

Real world:



Your ideal world:

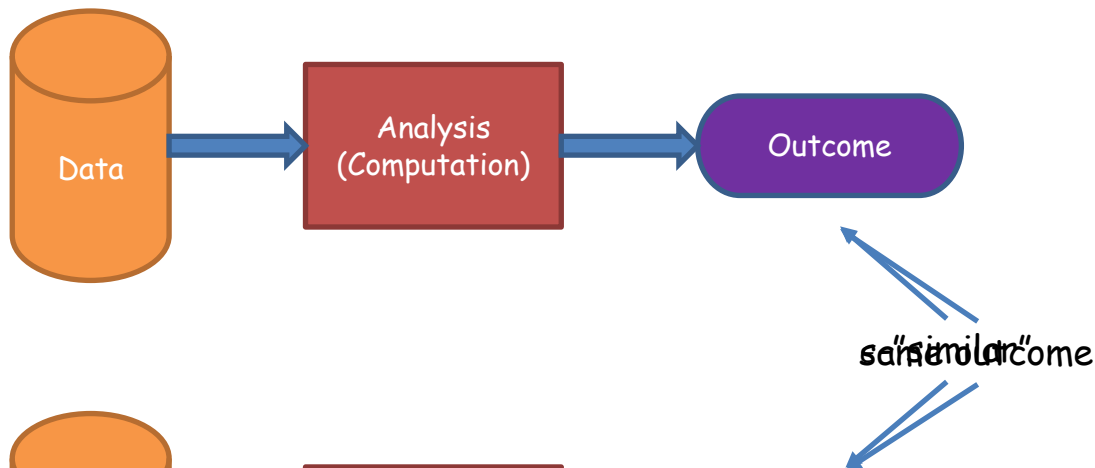


same outcome

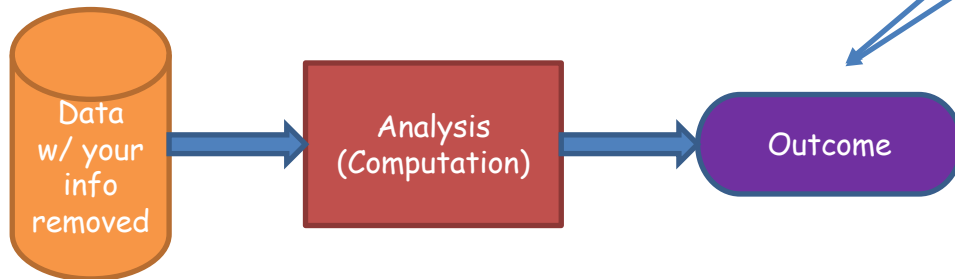


# A MORE REALISTIC PRIVACY DESIDERATA

Real world:

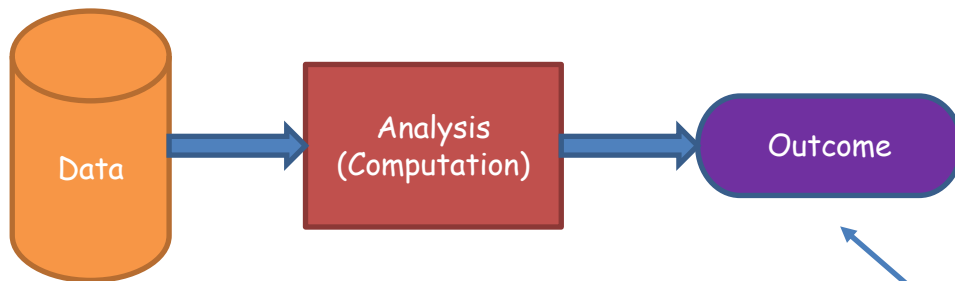


Your ideal world:

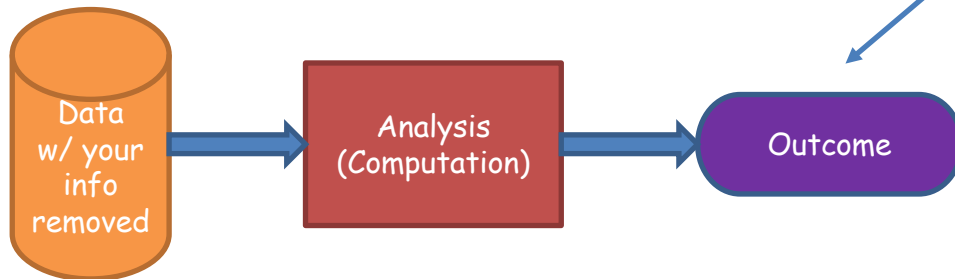


# DIFFERENTIAL PRIVACY [DWORK MCSHERRY NISSIM SMITH '06]

Real world:



Your ideal world:



$\epsilon$ -“similar”

Chance of bad event almost the same in everybody's ideal and real worlds

# DIFFERENTIAL PRIVACY [DWORK MCSHERRY NISSIM SMITH '06, '16]

A (randomized) algorithm  $M: X^n \rightarrow T$  satisfies  $\epsilon$ -differential privacy if  $\forall x, x' \in X^n$  that differ on one entry,

$$M(x) \approx_{\epsilon} M(x')$$

# DIFFERENTIAL PRIVACY [DWORK MCSHERRY NISSIM SMITH '06, '16]

A (randomized) algorithm  $M: X^n \rightarrow T$  satisfies  $\epsilon$  -differential privacy if

$\forall x, x' \in X^n$  that differ on one entry,

$\forall S$  subset of the outcome space  $T$ ,

$$\Pr_M[M(x) \in S] \leq e^\epsilon \Pr_M[M(x') \in S]$$

- The parameter  $\epsilon$  measures ‘leakage’ or ‘harm’
- For small  $\epsilon$ :  $e^\epsilon \approx 1 + \epsilon \approx 1$
- Think  $\epsilon \approx \frac{1}{100}$  or  $\epsilon \approx \frac{1}{10}$  not  $\epsilon \approx 2^{-80}$

Choice of “distance” measure (max log ratio) not accidental!

# DIFFERENTIAL PRIVACY [DWORK MCSHERRY NISSIM SMITH '06, '16]

A (randomized) algorithm  $M: X^n \rightarrow T$  satisfies  $(\epsilon, \delta)$ -differential privacy if  
 $\forall x, x' \in X^n$  that differ on one entry,  
 $\forall S$  subset of the outcome space  $T$ ,

$$\Pr_M[M(x) \in S] \leq e^\epsilon \Pr_M[M(x') \in S] + \delta$$

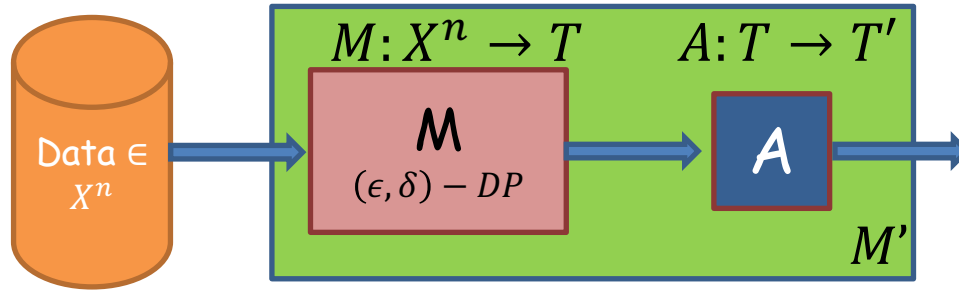
- The parameter  $\epsilon$  measures ‘leakage’ or ‘harm’
- For small  $\epsilon$ :  $e^\epsilon \approx 1 + \epsilon \approx 1$
- Think  $\epsilon \approx \frac{1}{100}$  or  $\epsilon \approx \frac{1}{10}$  not  $\epsilon \approx 2^{-80}$
- Think  $0 \leq \delta \ll \frac{1}{n}$  (often, cryptographically small)

Choice of “distance” measure (max log ratio) not accidental!

# UNDERSTANDING DIFFERENTIAL PRIVACY

- **“Automatic” opt-out:** I am protected (almost) as if my info is not used at all.
- **Plausible deniability:** I can claim any value for my information as outcome is (almost) as likely with that value.
- **I incur limited risk:** Contributing my real info can increase the probability I will be denied insurance by at most 1%.
  - When compared with not participating, or contributing fake info.

# PROPERTIES OF DIFFERENTIAL PRIVACY: POST PROCESSING



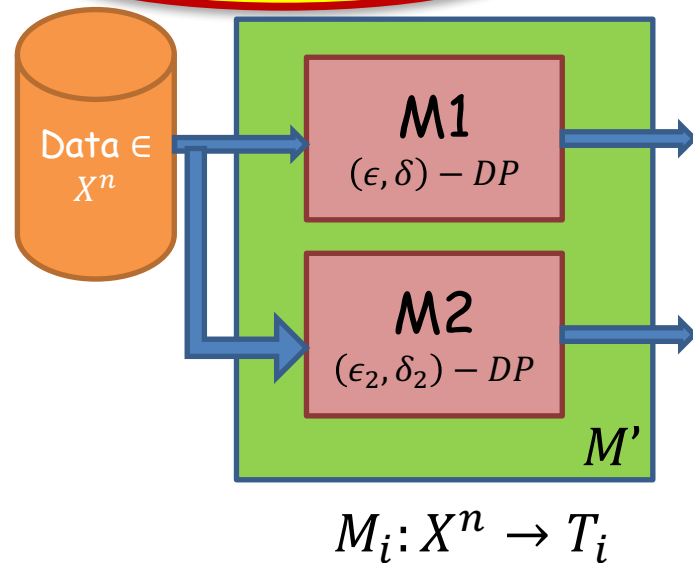
- **Claim:**  $M'$  is  $(\epsilon, \delta)$ -differentially private
- **Proof:**
  - Let  $x, x'$  be neighboring databases and  $S'$  a subset of  $T'$
  - Let  $S = \{z \in T: A(z) \in S'\}$  be the preimage of  $S'$  under  $A$   
$$\Pr[M'(x) \in S'] = \Pr[M(x) \in S]$$
$$\leq e^\epsilon \Pr[M(x') \in S] + \delta = e^\epsilon \Pr[M'(x') \in S] + \delta$$

# PROPERTIES OF DIFFERENTIAL PRIVACY:

## COMPOSITION [DMNS06, DKMMN06, DL09, KOV15, MV16]

- **Claim:**  $M'$  is  $(\epsilon_1 + \epsilon_2, \delta_1 + \delta_2)$ -differentially private
- **Proof** (for the case  $\delta_1 = \delta_2 = 0$ )
  - Let  $x, x'$  be neighboring databases and  $S$  a subset of  $(T_1 \times T_2)$

$$\begin{aligned}\Pr[M'(x) \in S] &= \sum_{(z_1, z_2) \in S} \Pr[M_1(x) = z_1 \wedge M_2(x) = z_2] \\ &= \sum_{(z_1, z_2) \in S} \Pr[M_1(x) = z_1] \Pr[M_2(x) = z_2] \\ &\leq \sum_{(z_1, z_2) \in S} e^{\epsilon_1} \Pr[M_1(x') = z_1] e^{\epsilon_2} \Pr[M_2(x') = z_2] = e^{\epsilon_1 + \epsilon_2} \Pr[M'(x') \in S]\end{aligned}$$





# POST PROCESSING, COMPOSITION

## WHY DO WE CARE?



- **For privacy:** A definition that does not allow post processing/ composing of analyses is (to the least) problematic
- **For DP algorithm design:** Allows a modular design of an analysis from simpler analyses
- Many (all?) other currently known definitions of privacy lack these properties

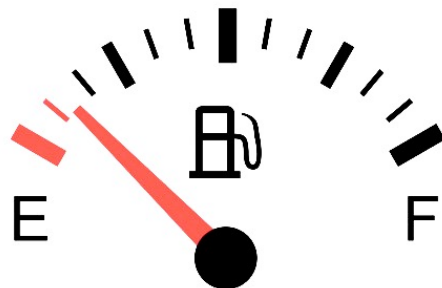
# THE “PRIVACY BUDGET”

**Privacy is a consumable resource:** The parameter  $\epsilon$  measures leakage and can be treated as a “privacy budget” which is consumed as analyses are performed.

Composition theorems help manage the budget by providing a bound on the overall use of the privacy budget.


**This is a feature, not a bug!**

Consider how removing the fuel gauge *would not* make your car run indefinitely without refueling.



# PROPERTIES OF DP:

## GROUP PRIVACY

- Let  $M$  be  $(\epsilon, \delta)$ -differentially private:
  - For all datasets  $x, x' \in X^n$  that **differ on one entry**, for all subsets  $S$  of the outcome space  $T$ :
$$\Pr_M[M(x) \in S] \leq e^\epsilon \Pr_M[M(x') \in S] + \delta.$$
- **Claim:** for all databases  $x, x' \in X^n$  that **differ on  $t$  entries**, for all subsets  $S$  of the outcome space  $T$ :
$$\Pr_M[M(x) \in S] \leq e^{t\epsilon} \Pr_M[M(x') \in S] + t\delta e^{t\epsilon}.$$


# REASONING ABOUT RISK

## GERTRUDE'S LIFE INSURANCE



- Gertrude:
  - Age: 65
  - She has a \$100,000 life insurance policy.
  - She is considering participating in a medical study but is concerned it may affect her insurance premium.

# REASONING ABOUT RISK

## GERTRUDE'S LIFE INSURANCE



- Based on her age and sex, she has a 1% chance of dying next year. Her life insurance premium is set at  $0.01 \times \$100,000 = \$1,000$ .
- Gertrude is a coffee drinker. If the medical study finds that 65-year-old female coffee drinkers have a 2% chance of dying next year, her premium would be set at \$2,000.
  - This would be her **baseline risk**: Her premium would be set at \$2,000 even if she were not to participate in the study.
- **Can Gertrude's premium increase beyond her baseline risk?**
  - She is worried that the study may reveal more about her, such as that she *specifically* has a 50% chance of dying next year. This can increase her premium from \$2,000 to \$50,000!

# REASONING ABOUT RISK

## GERTRUDE'S LIFE INSURANCE



- Reasoning about Gertrude's risk
  - Imagine instead the study is performed using differential privacy with  $\epsilon = 0.01$ .
  - The insurance company's estimate of Gertrude's risk of dying in the next year can increase to at most
$$(1 + \epsilon) \cdot 2\% = 2.02\%.$$
  - Her premium would increase to at most \$2,020. Therefore, Gertrude's risk would be  $\leq \$2020 - \$2000 = \$20$ .

# REASONING ABOUT RISK

## GERTRUDE'S LIFE INSURANCE



- Generally, calculating one's baseline is very complex (if possible at all).
  - In particular, in our example the 2% baseline depends on the potential outcome of the study.
  - The baseline may also depend on many other factors Gertrude does not know.
- However, differential privacy provides simultaneous guarantees for every possible baseline value.
  - The guarantee covers not only changes in Gertrude's life insurance premiums, but also her health insurance and more.

# **HOW IS DIFFERENTIAL PRIVACY ACHIEVED?**

**Answer (part 1):**

**Addition of carefully crafted random noise**



# RANDOMIZED RESPONSE [WARNER 65]

- $x \in \{0,1\}$
- $RR_\alpha(x) = f(x) = \begin{cases} x & w.p. \frac{1}{2} + \alpha \\ \neg x & w.p. \frac{1}{2} - \alpha \end{cases}$
- **Claim:** setting  $\alpha = \frac{1}{2} \frac{e^\epsilon - 1}{e^\epsilon + 1}$ ,  $RR_\alpha(x)$  is  $\epsilon$ -differentially private
- **Proof:**
  - Neighboring databases:  $x = 0; x' = 1$
  - $\frac{\Pr[RR(0)=0]}{\Pr[RR(1)=0]} = \frac{\frac{1}{2}(1+\frac{e^\epsilon-1}{e^\epsilon+1})}{\frac{1}{2}(1-\frac{e^\epsilon-1}{e^\epsilon+1})} = e^\epsilon$

small  $\epsilon$ :  $e^\epsilon \approx 1 + \epsilon$ .  
Get  $\alpha \approx \frac{\epsilon}{4}$

# CAN WE MAKE USE OF RANDOMIZED RESPONSE?

- $(x_1, x_2, \dots, x_n) \in \{0,1\}^n$ ; want to estimate  $\sum x_i$
- $Y_i = RR_\alpha(x_i)$
- $E[Y_i] = x_i \left(\frac{1}{2} + \alpha\right) + (1 - x_i) \left(\frac{1}{2} - \alpha\right) = \frac{1}{2} + \alpha(2x_i - 1)$
- $x_i = \frac{E[Y_i] - \frac{1}{2} + \alpha}{2\alpha}$  suggesting estimate  $\hat{x}_i = \frac{y_i - \frac{1}{2} + \alpha}{2\alpha}$
- $E[\hat{x}_i] = x_i$  by construction but  $Var[\hat{x}_i] = \frac{\frac{1}{4} - \alpha^2}{4\alpha^2} \approx \frac{1}{\epsilon^2}$  high!
- $E[\sum \hat{x}_i] = \sum x_i$  and  $Var[\sum \hat{x}_i] = n \frac{\frac{1}{4} - \alpha^2}{4\alpha^2} \approx \frac{n}{\epsilon^2}$ ; stdev  $\approx \frac{\sqrt{n}}{\epsilon}$
- Useful when  $\frac{\sqrt{n}}{\epsilon} \ll n$  (i.e.  $n \gg \frac{1}{\epsilon^2}$ )

Lots of noise?

Compare with sampling noise  $\approx \sqrt{n}$

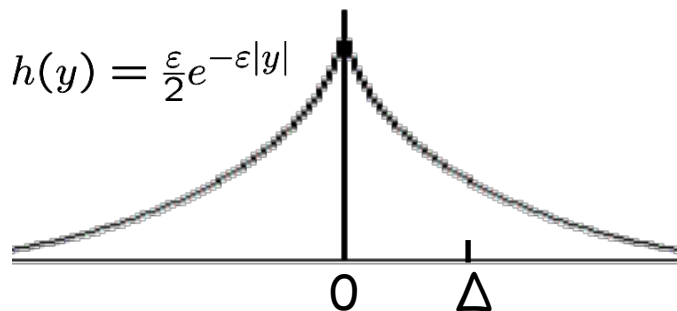
# LAPLACE MECHANISM

- **Database:**  $x = (x_1, \dots, x_n)$  where  $x_i \in \{0,1\}$ .
  - $x_i = 1$  : individual  $i$  is Diabetic.
- **Query:**  $f(x) = \sum x_i$ 
  - For neighboring  $x, x'$ ,  $|f(x) - f(x')| = |\sum_i x_i - \sum_i x'_i| \leq 1$ .
- **Noisy answer:**
  - Return  $f(x) + Y$ , where  $Y \sim \text{Lap}(\frac{1}{\epsilon})$ .

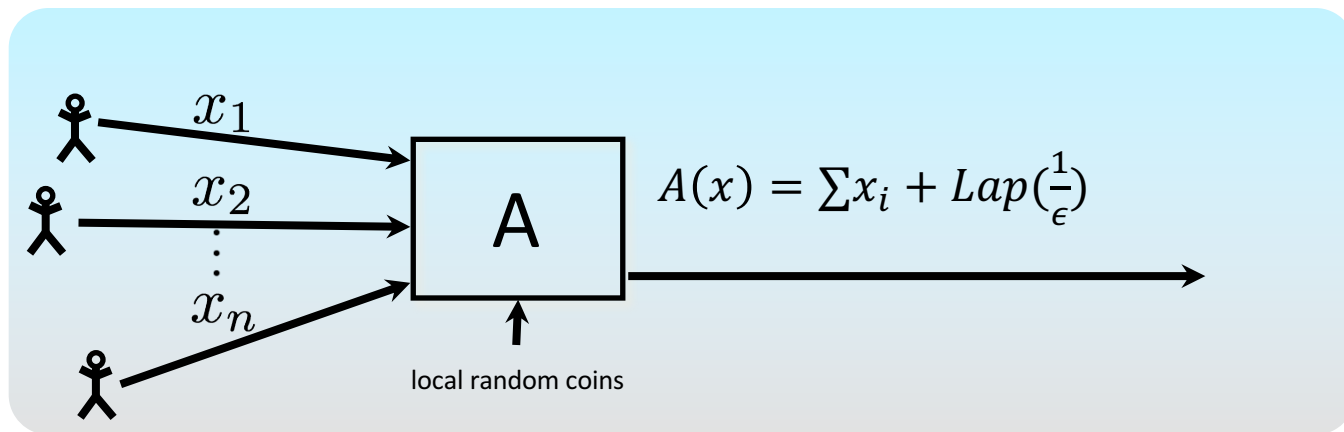
- **Laplace Distribution:**

- $E[Y] = 0; \sigma[Y] = \sqrt{2}/\epsilon$
- Sliding property:

- $\forall y, \Delta: \frac{h(y)}{h(y + \Delta)} \leq e^{\epsilon \Delta}$

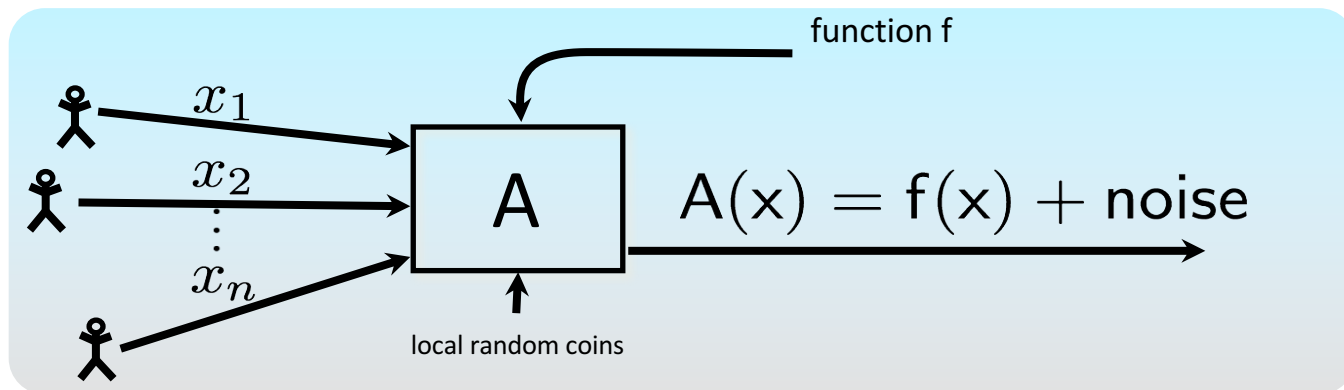


# IS THIS A LOT OF NOISE?



- If  $x$  is a random sample from a large underlying population, then **sampling noise**  $\approx \sqrt{n}$ .
- If  $\frac{1}{\epsilon} \ll \sqrt{n}$  then  $A(x)$  “as good as” if computed over the sample.

# FRAMEWORK OF GLOBAL SENSITIVITY [DMNS06]



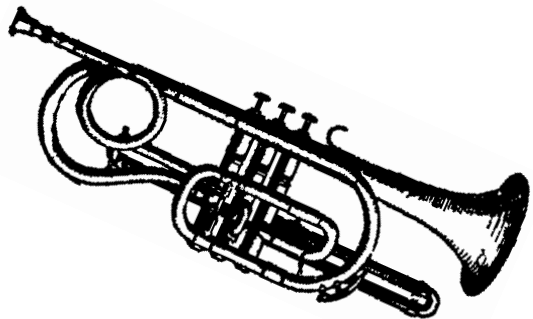
- $GS_f = \max_{X, X'} ||f(X) - f(X')||_1$
- **Theorem [DMNS06]:**
  - $A(X) = f(X) + \text{Lap}(GS_f/\epsilon)$  is  $\epsilon$ -differential private.

# USING GLOBAL SENSITIVITY

$$GS_f = \max_{\text{neighbors } x, x'} \|f(x) - f(x')\|_1$$

- Many natural functions have low sensitivity
  - e.g., histogram, mean, covariance matrix, distance to a function, estimators with bounded “sensitivity curve”, strongly convex optimization problems.
- Laplace mechanism can be a programming interface [BDMN '05].
  - Implemented in several systems [McSherry '09, Roy et al. '10, Haeberlen et al. '11, Moharan et al. '12].

# Many ways of making *less and less* noise



Randomized Response [W65]  
Framework of global sensitivity [DMNS06]  
Framework of smooth sensitivity [NRS07]  
Sample and aggregate [NRS07]  
Exponential mechanism [MT07]  
Secrecy of the sample [KLNRS08]  
Propose test release [DL09]  
Sparse vector technique [DNRRV09]  
Private multiplicative weights [HR10]  
Matrix mechanism [LHRMM10]  
Choosing mechanism [BNS13]  
Large margin mechanism [CHS14]  
Dual query mechanism [GGHRW14]  
+ other algorithmic techniques



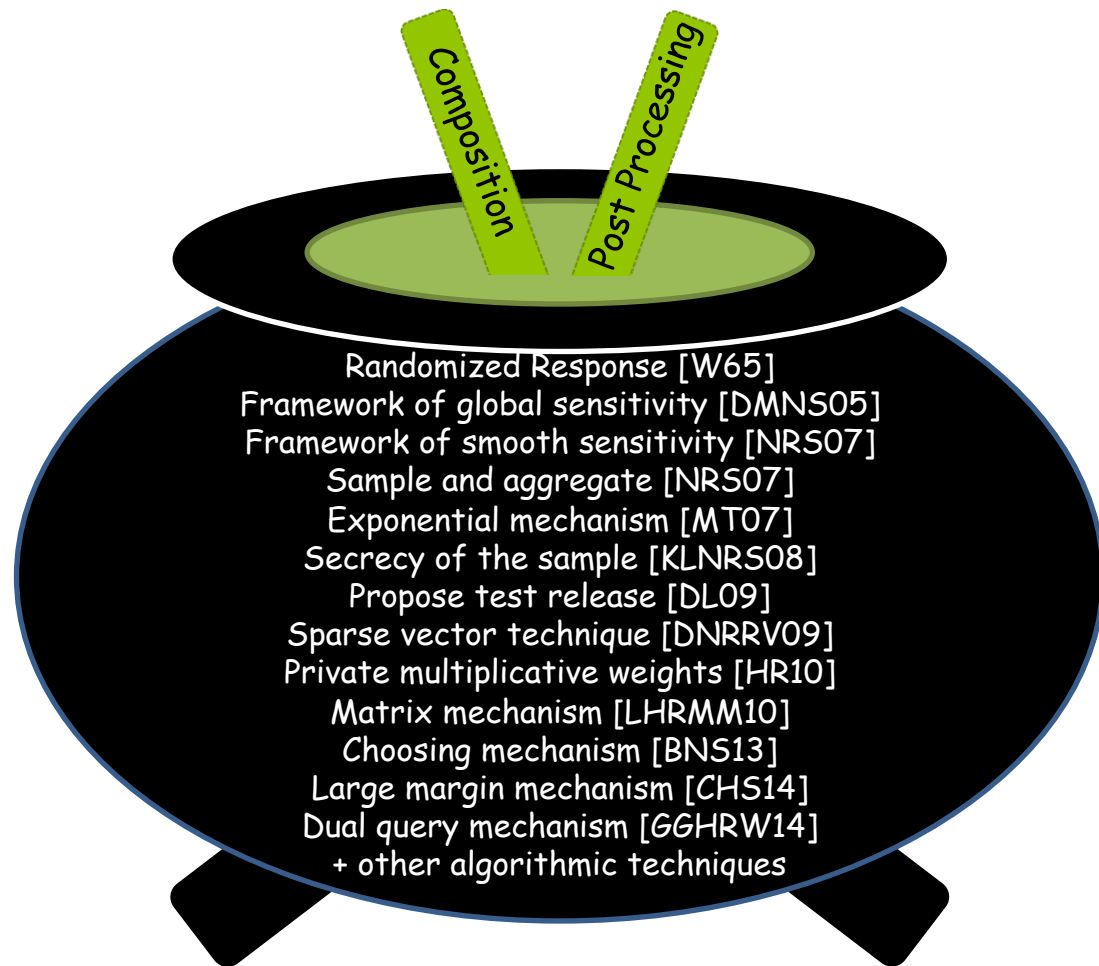
# HOW IS DIFFERENTIAL PRIVACY ACHIEVED?

**Answer (part 2):**  
**Composition of differentially private sub-  
computations and post-processing**



Randomized Response [W65]  
Framework of global sensitivity [DMNS05]  
Framework of smooth sensitivity [NRS07]  
Sample and aggregate [NRS07]  
Exponential mechanism [MT07]  
Secrecy of the sample [KLNRS08]  
Propose test release [DL09]  
Sparse vector technique [DNRRV09]  
Private multiplicative weights [HR10]  
Matrix mechanism [LHRMM10]  
Choosing mechanism [BNS13]  
Large margin mechanism [CHS14]  
Dual query mechanism [GGHRW14]  
+ other algorithmic techniques

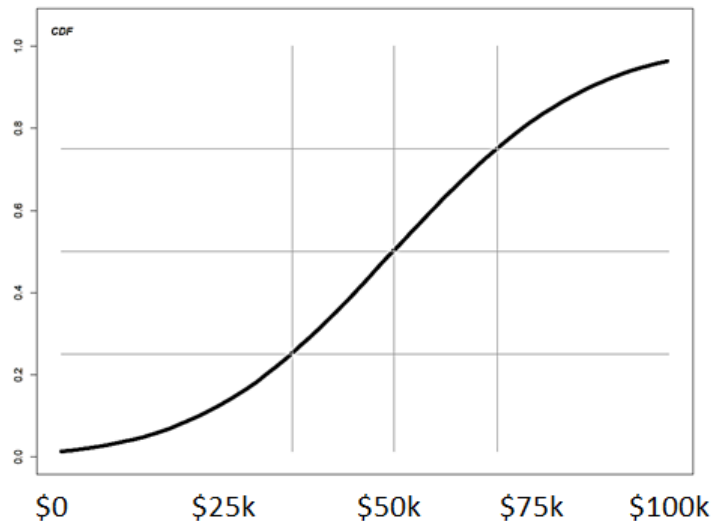
## A programmable framework



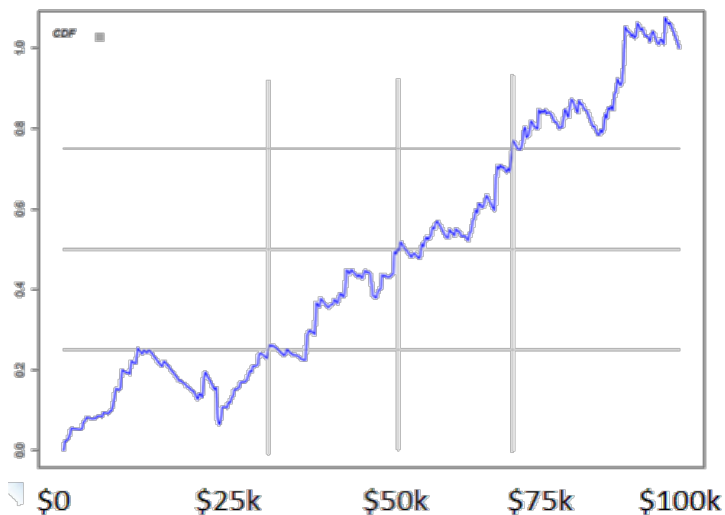
# DIFFERENTIALLY PRIVATE COMPUTATIONS

Algorithms maintain differential privacy via the introduction of *carefully crafted random noise* into the computation.

Income in District Q



Income in District Q



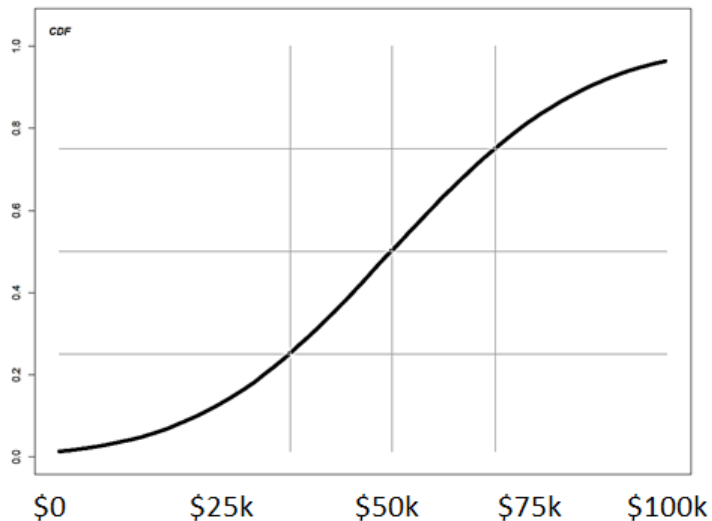
$$\epsilon = 0.005$$

(District Q and its data are stylized examples.)

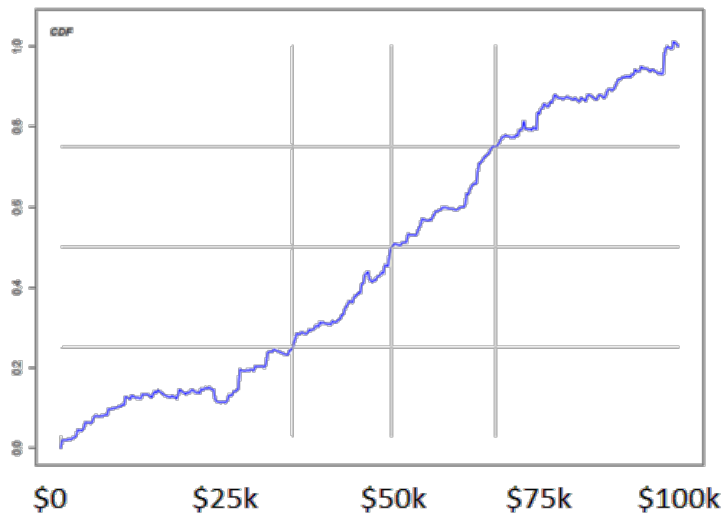
# DIFFERENTIALLY PRIVATE COMPUTATIONS

Algorithms maintain differential privacy via the introduction of *carefully crafted random noise* into the computation.

Income in District Q



Income in District Q



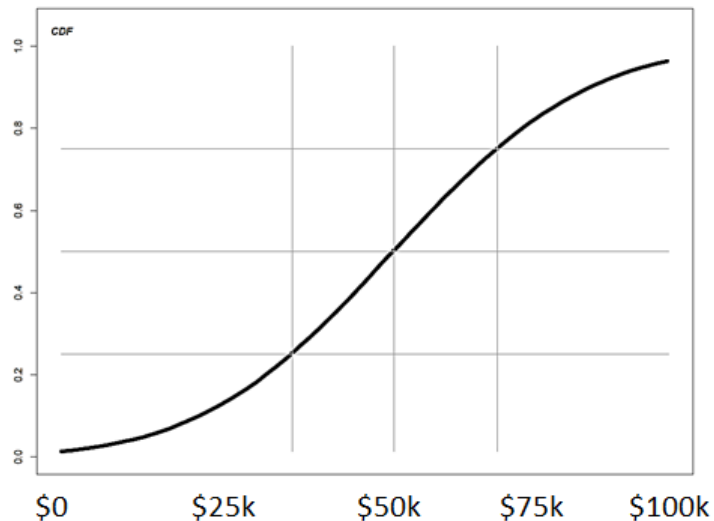
$$\epsilon = 0.01$$

(District Q and its data are stylized examples.)

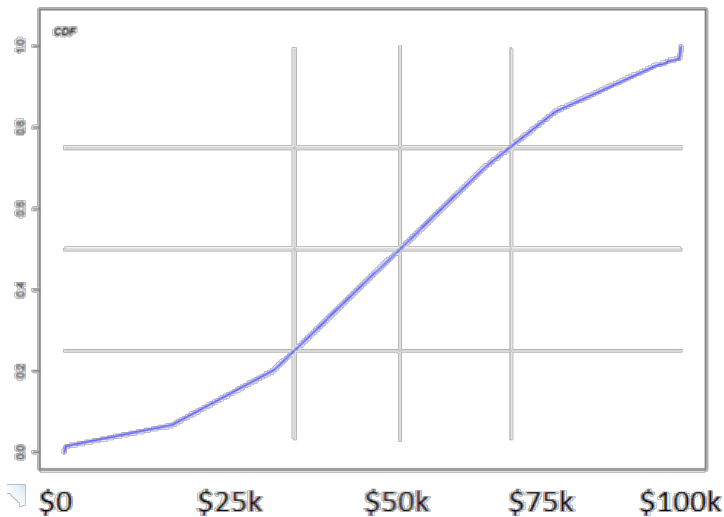
# DIFFERENTIALLY PRIVATE COMPUTATIONS

Algorithms maintain differential privacy via the introduction of *carefully crafted random noise* into the computation.

Income in District Q



Income in District Q



$$\epsilon = 0.1$$

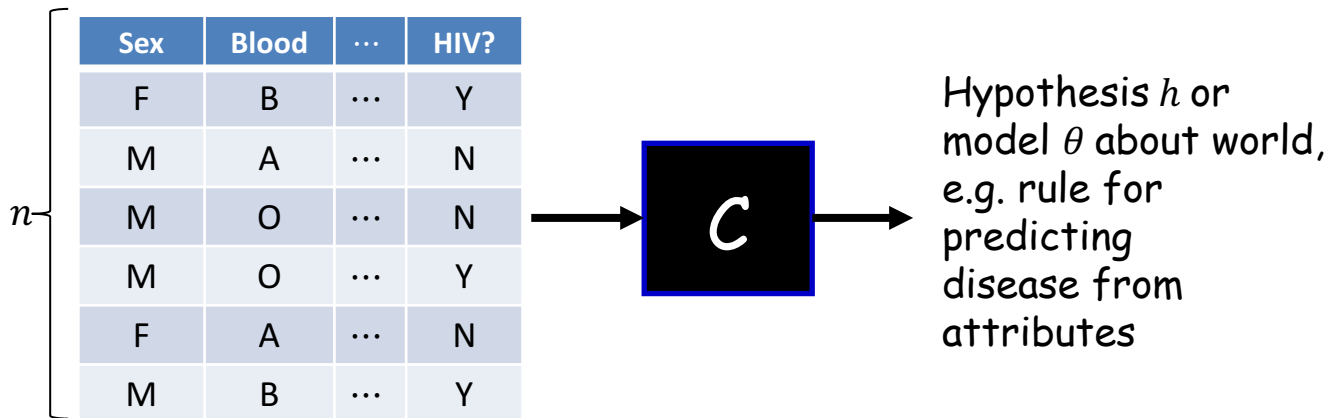
(District Q and its data are stylized examples.)

# WHAT CAN BE COMPUTED WITH DIFFERENTIAL PRIVACY?

- **Descriptive statistics:** counts, mean, median, histograms, boxplots, etc.
- **Supervised and unsupervised ML tasks:** classification, regression, clustering, distribution learning, etc.
- **Generation of synthetic data**

Because of noise addition, differentially private algorithms work best when the number of data records is large.

# STATISTICAL INFERENCE & MACHINE LEARNING



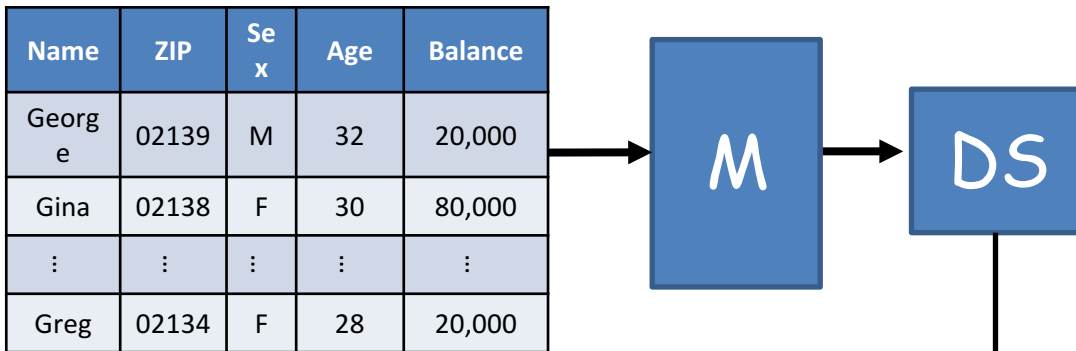
- **Utility:** vast array of machine learning and statistical estimation problems with little loss in convergence rate as  $n \rightarrow \infty$ 
  - Optimizations & practical implementations for logistic regression, ERM, LASSO, SVMs in [RBHT09,CMS11,ST13,JT14, ...]
- **Problem:** Sample complexity higher than in non-private learning [BKN10, CH11, BNS13, BNS13a, BNS15, BNSV15,...]

# DATA SANITIZATION

[BLUM-LIGETT-ROTH '08, HARDT-ROTHBLUM '10]

- Q: A collection of statistical queries

- Sanitization:



- Utility:

For all  $q \in Q$ :

$$q(x) \approx q(DS)$$

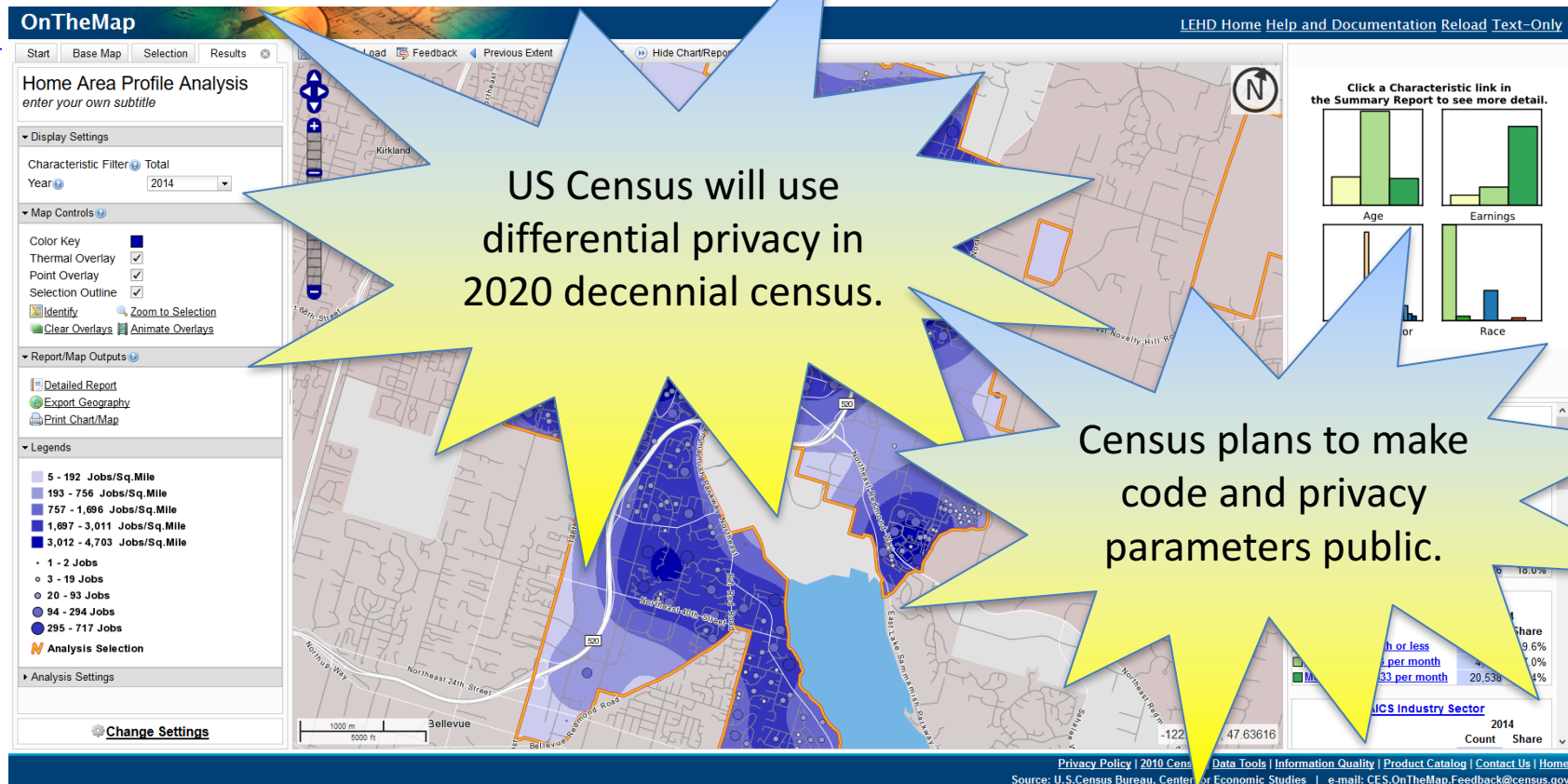
- Problem: uses computation time exponential in  $d$  [UV11, ...]



# **REAL-LIFE APPLICATIONS**

# U.S. CENSUS BUREAU

<http://onthemap.ces.census.gov>



2008 AD



## RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response

Úlfar Erlingsson  
Google, Inc.  
ulfar@google.com

Vasyl Pihur  
Google, Inc.  
vpihur@google.com

Aleksandra Korolova  
University of Southern California  
korolova@usc.edu

### ABSTRACT

Randomized Aggregatable Privacy-Preserving Ordinal Response, or RAPPOR, is a technology for crowdsourcing statistics from end-user client software, anonymously, with strong privacy guarantees. In short, RAPPORs allow the forest of client data to be studied, without permitting the possibility of looking at individual trees. By applying randomized response in a novel manner, RAPPOR provides the mechanisms for such collection as well as for efficient, high-utility analysis of the collected data. In particular, RAPPOR permits statistics to be collected on the population of client-side strings with strong privacy guarantees for each client, and without linkability of their reports.

This paper describes and motivates RAPPOR, details its **differential-privacy** and utility guarantees, discusses its practical deployment and properties in the face of different attack models, and, finally, gives results of its application to both synthetic and real-world data.

### 1 Introduction

Crowdsourcing data to make better, more informed decisions is becoming increasingly commonplace. For any such crowdsourcing, privacy-preservation mechanisms should be

asked to flip a fair coin, in secret, and answer “Yes” if it comes up heads, but tell the truth otherwise (if the coin comes up tails). Using this procedure, each respondent retains very strong deniability for any “Yes” answers, since such answers are most likely attributable to the coin coming up heads; as a refinement, respondents can also choose the untruthful answer by flipping another coin in secret, and get strong deniability for both “Yes” and “No” answers.

Surveys relying on randomized response enable easy computations of accurate population statistics while preserving the privacy of the individuals. Assuming absolute compliance with the randomization protocol (an assumption that may not hold for human subjects, and can even be non-trivial for algorithmic implementations [23]), it is easy to see that in a case where both “Yes” and “No” answers can be denied (flipping two fair coins), the true number of “Yes” answers can be accurately estimated by  $2(Y - 0.25)$ , where  $Y$  is the proportion of “Yes” responses. In expectation, respondents will provide the true answer 75% of the time, as is easy to see by a case analysis of the two fair coin flips.

Importantly, for one-time collection, the above randomized survey mechanism will protect the privacy of any specific respondent, irrespective of any attacker’s prior knowl-

# APPLE

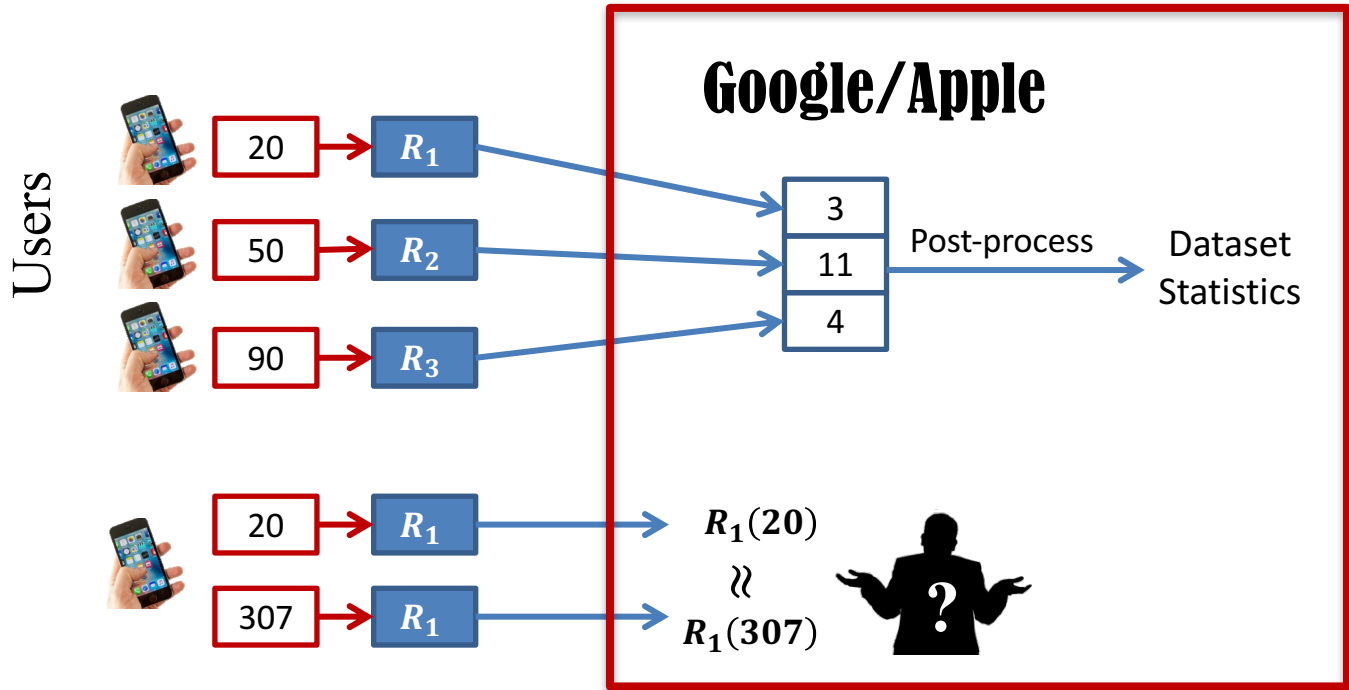
ANDY GREENBERG SECURITY 06.13.16 07:02 PM

**APPLE'S 'DIFFERENTIAL  
PRIVACY' IS ABOUT COLLECTING  
YOUR DATA—BUT NOT *YOUR*  
DATA**

Apple will not  
see your data

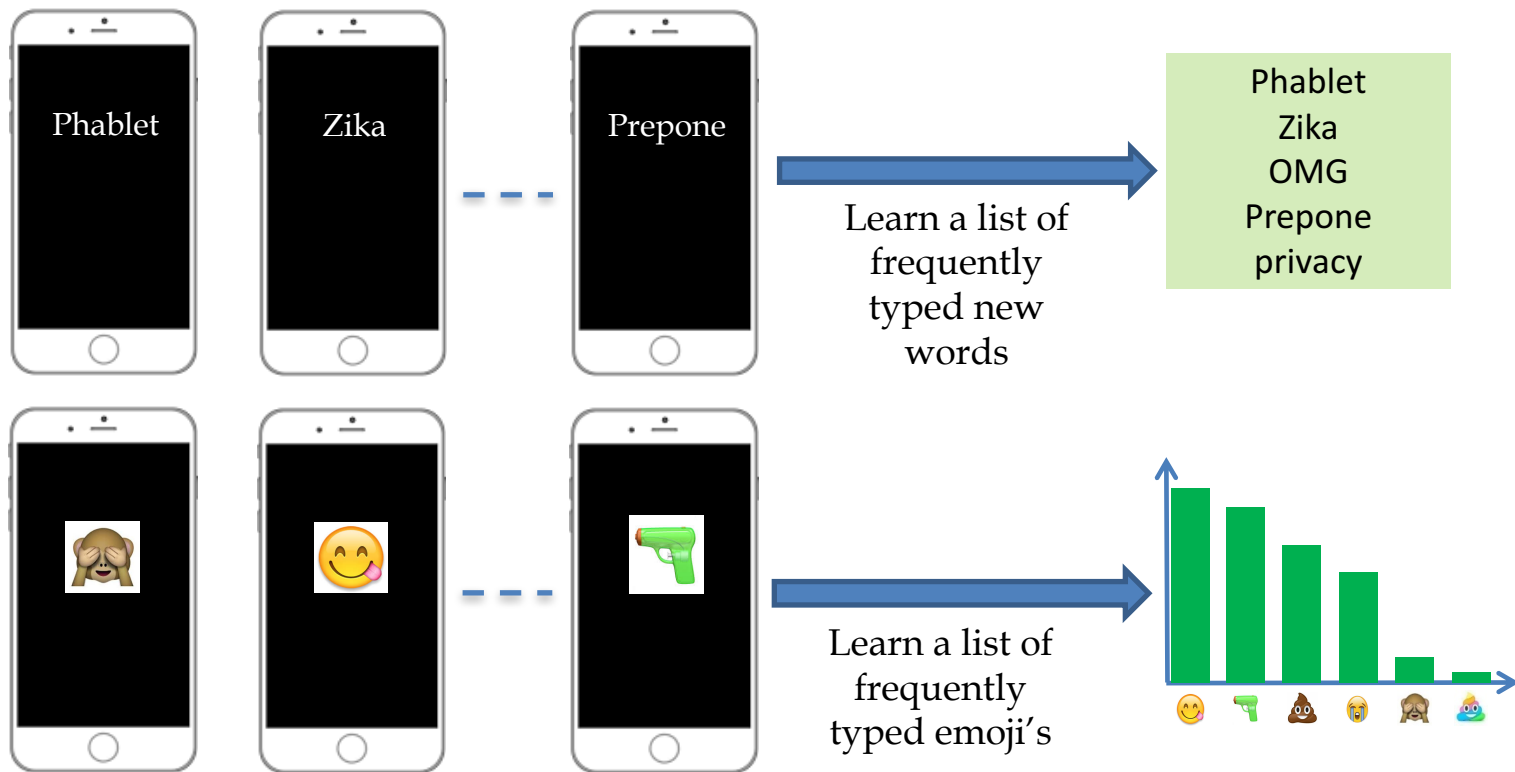


# LOCAL DIFFERENTIAL PRIVACY



Users retain their data and only send the server randomizations which preserve differential privacy even if made public

# LEARNING HEAVY HITTERS



# Harvard University Privacy Tools Project

[Home](#)
[Research ▾](#)
[News](#)
[People ▾](#)
[Publications](#)
[Software ▾](#)
[Outreach ▾](#)



The Privacy Tools Project is a broad effort to advance a multidisciplinary understanding of data privacy issues and build computational, statistical, legal, and policy tools to help address these issues in a variety of contexts. It is a collaborative effort between Harvard's [Center for Research on Computation and](#)

## LATEST NEWS & BLOG POSTS

Graduate Student Michael Bar-Sinai Presented at the 8th Annual ESPAnet Israel 2017

PI Salil Vadhan, PI Kobbi Nissim, and Senior Researcher Marco Gaboardi Presented at the Third Biennial Secure and Trustworthy CyberSpace Principal Investigators' Meeting (SaTC PI Meeting '17)

Berkman Klein Center Seeks Applications for 2017 Summer Internship Program

Harvard Magazine Highlights Privacy Tools Project in Article on Privacy and Security

George Kellaris Featured on CRCS Blog

Privacy Tools Project Featured in Harvard Law Review

Berkman Klein Center Seeks Fellow for Privacy





This prototype system will allow researchers with sensitive datasets to make **differentially private** statistics about their data available through data repositories using the Dataverse platform.



Our prototype system will allow researchers to: [1] upload private data to a secured Dataverse archive, [2] decide what statistics they would like to release about that data, and [3] release privacy preserving versions of those statistics to the repository, [4] that can be explored through a curator interface without releasing the raw data, including [5] interactive queries.

A paper [describing our system can be found here](#). This system was created by the [Privacy Tools for Sharing Research Data project](#). Differential privacy is a mathematical framework for enabling statistical analysis of sensitive datasets while ensuring that individual-level information cannot be leaked. The project website contains resources for [learning more about differential privacy](#).

## Budget Tool

The first part of this system is a tool that helps both data depositors and data analysts distribute a global privacy budget across many statistics. Users select which

## Curator Interface

When the data depositor has distributed their privacy budget, the second portion of our tool system draws differentially private versions of those statistical

## Interactive Queries

Our system will allow some of the privacy budget to be reserved for future data analysts to choose their own differentially private statistics to calculate (selected from





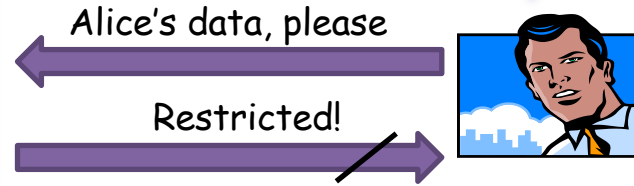
Other researchers may find it useful...

Contains student info protected by FERPA

Dataverse Network

Should I apply for access???

IRB, terms of use... is it worth the trouble?



$\Psi$

Private data Sharing Interface

Access Alice's data w/differential privacy

\* <http://dataverse.org/>

FERPA: Family educational rights and privacy act, 1974

IRB: Institutional Review Board

PSI ( $\Psi$ )

## Private data Sharing Interface

A National Science Foundation Secure and Trustworthy Cyberspace Project

This prototype system will allow researchers with sensitive datasets to make differentially private statistics about their data available through data repositories using the Dataverse platform.



Upload



Budget



Release



Explore



Query

Our prototype system will allow researchers to: [1] upload private data to a secured Dataverse archive, [2] decide what statistics they would like to release about that data, and [3] release privacy preserving versions of those statistics to the repository, [4] that can be explored through a curator interface without releasing the raw data, including [5] interactive queries.

A paper [describing our system can be found here](#). This system was created by the [Privacy Tools for Sharing Research Data project](#). Differential privacy is a mathematical framework for enabling statistical analysis of sensitive datasets while ensuring that individual-level information cannot be leaked. The project website contains resources for [learning more about differential privacy](#).

### Budget Tool

The first part of this system is a tool that helps both data depositors and data analysts distribute a global privacy budget across many statistics. Users select which

### Curator Interface

When the data depositor has distributed their privacy budget, the second portion of our tool system draws differentially private versions of those statistical

### Interactive Queries

Our system will allow some of the privacy budget to be reserved for future data analysts to choose their own differentially private statistics to calculate (selected from

# **BRINGING DIFFERENTIAL PRIVACY TO PRACTICE**

# BRINGING DP TO PRACTICE - CHALLENGES

- **A new, complex privacy concept:**
  - How to communicate its strengths and limitations to data analyzers and individuals contributing their personal information?
  - Risk, baseline risk, accumulation of privacy risk.
- **Analyzers' access to data:**
  - Via a mechanism.
  - Noise added. Very significant when data is scarce.
  - Overall use limited by the "privacy budget".
- **Matching guarantees with privacy law & regulation:**
  - Existing regulations often see privacy risks as binary.
  - Existing regulations refer to concepts such as PII, de-identification, linkage, inference, consent, ...
    - These lack rigorous technical definitions.
- **Choosing privacy parameters, managing privacy loss over time.**

# HIPAA'S *EXPERT DETERMINATION* ME

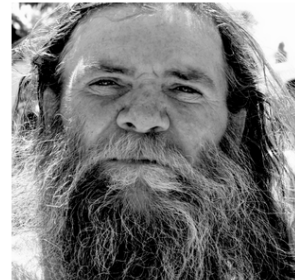
- Obtain confirmation from a qualified statistician that the risk of identification is very small
- Who is an expert?
  - U.S. Dept. of Health & Human Services guidance for HIPAA: *“There is no specific professional degree or certification program for designating who is an expert at rendering health information de-identified.”*
- How would the expert determine that the risk is small?



☐ Prof  
☐ Hobo



☐ Prof  
☐ Hobo



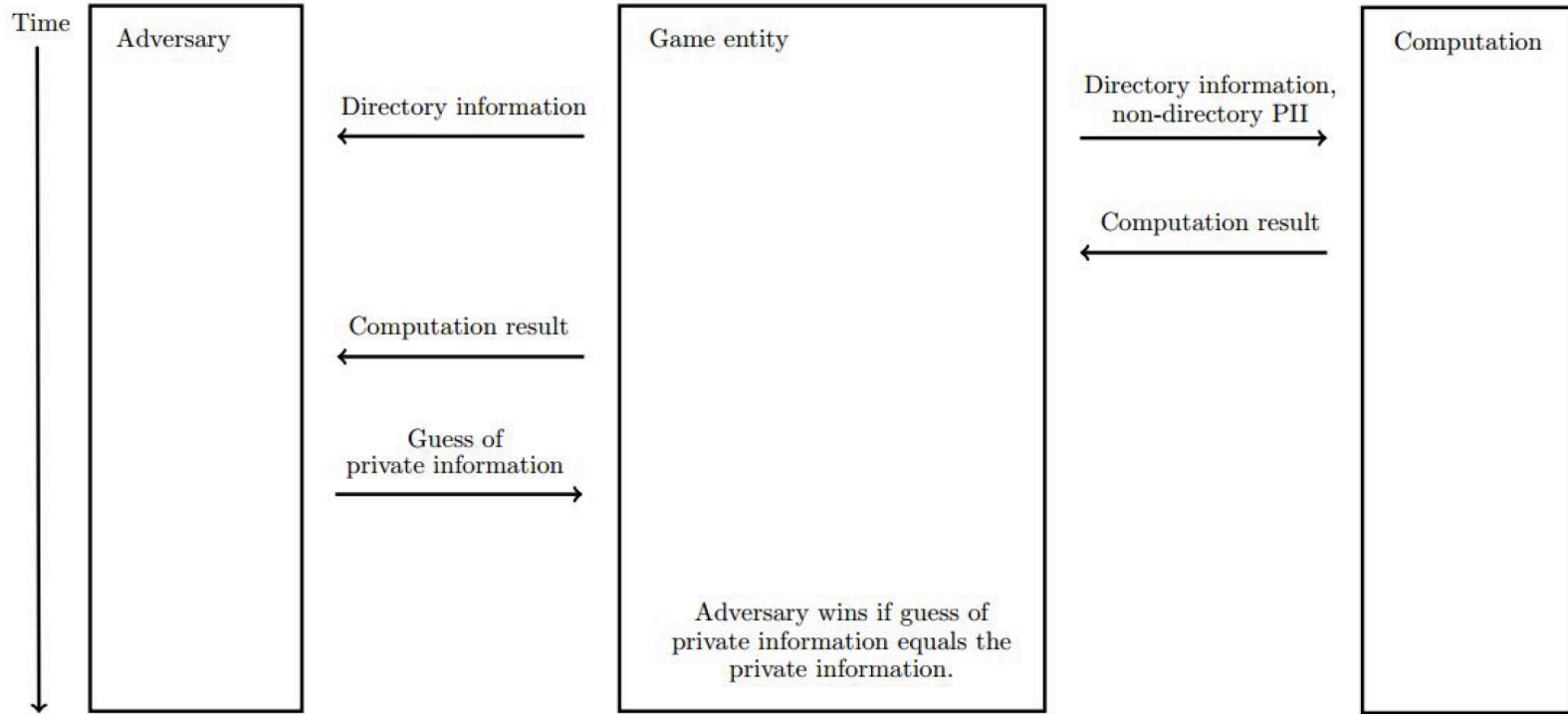
\* HIPAA: Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule  
45 C.F.R. Part 160 and Subparts A and E of Part 164.

# BRIDGING PRIVACY DEFINITIONS

- How can we claim that new technologies like differential privacy satisfy existing regulatory requirements?
- **Formal Modeling\***: A game based modeling approach for the privacy requirements of Family Educational Rights & Privacy Act (FERPA)
  - Concise and intuitive abstraction of the requirements in FERPA, taking care of potential ambiguities in the law

\*See: [Bridging the Gap between Computer Science and Legal Approaches to Privacy](#), Privacy Tools Project

# COMPONENTS OF A GAME-BASED MODELING OF FERPA



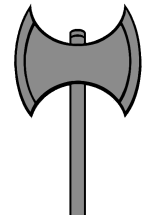
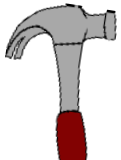
# BRIDGING PRIVACY DEFINITIONS

- How can we claim that new technologies like differential privacy satisfy existing regulatory requirements?
- **Formal Modeling\***: A game based modeling approach for the privacy requirements of Family Educational Rights & Privacy Act (FERPA)
  - Concise and intuitive abstraction of the requirements in FERPA, taking care of potential ambiguities in the law
- **Interpreting the differential privacy guarantee:**
  - With respect to concepts appearing in privacy law:
    - PII, de-identification, linkage, inference, consent

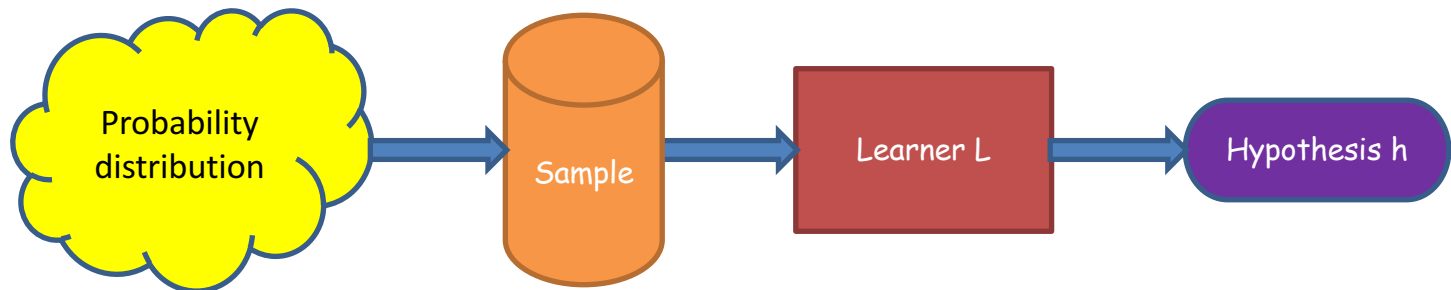
\*See: **Bridging the Gap between Computer Science and Legal Approaches to Privacy**, Privacy Tools Project



# PRIVACY AND THE ANALYST'S TOOLKIT



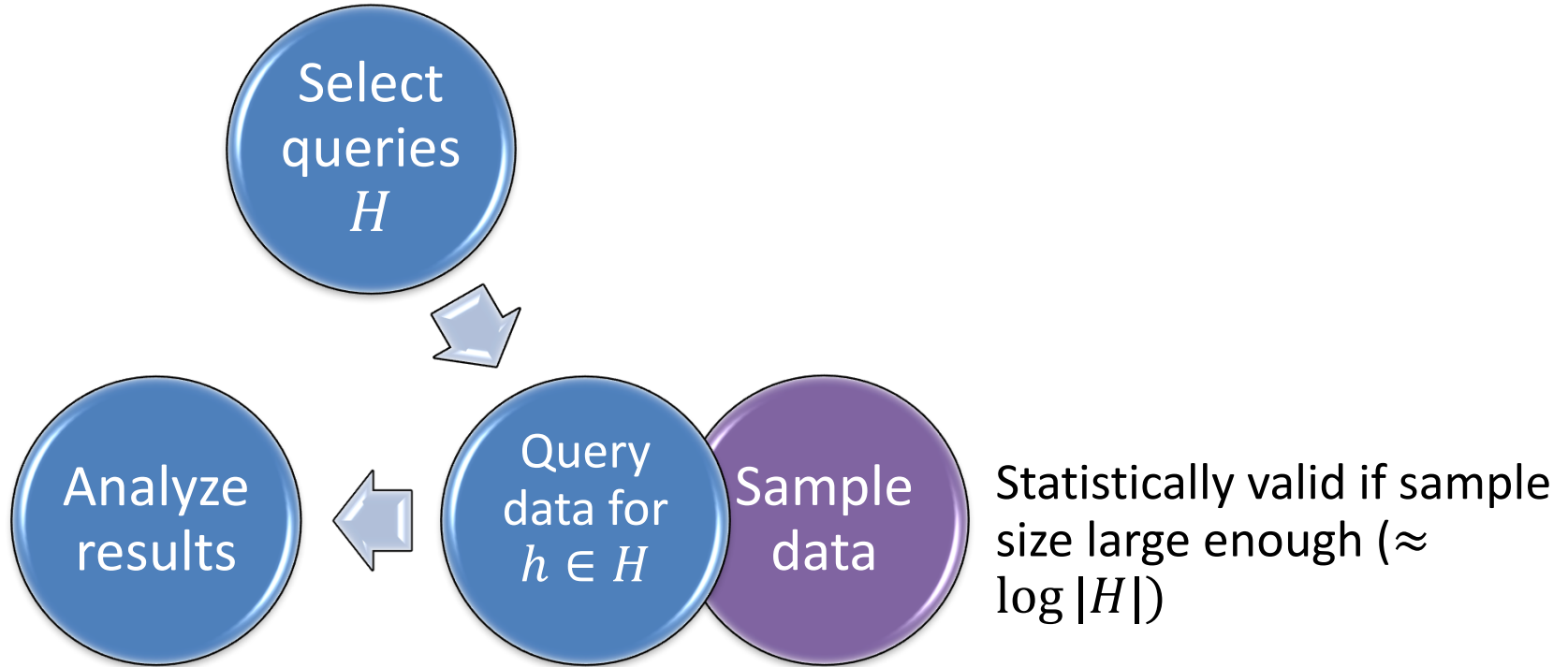
# LEARNING OVER THE SAMPLE VS. OVER THE DISTRIBUTION



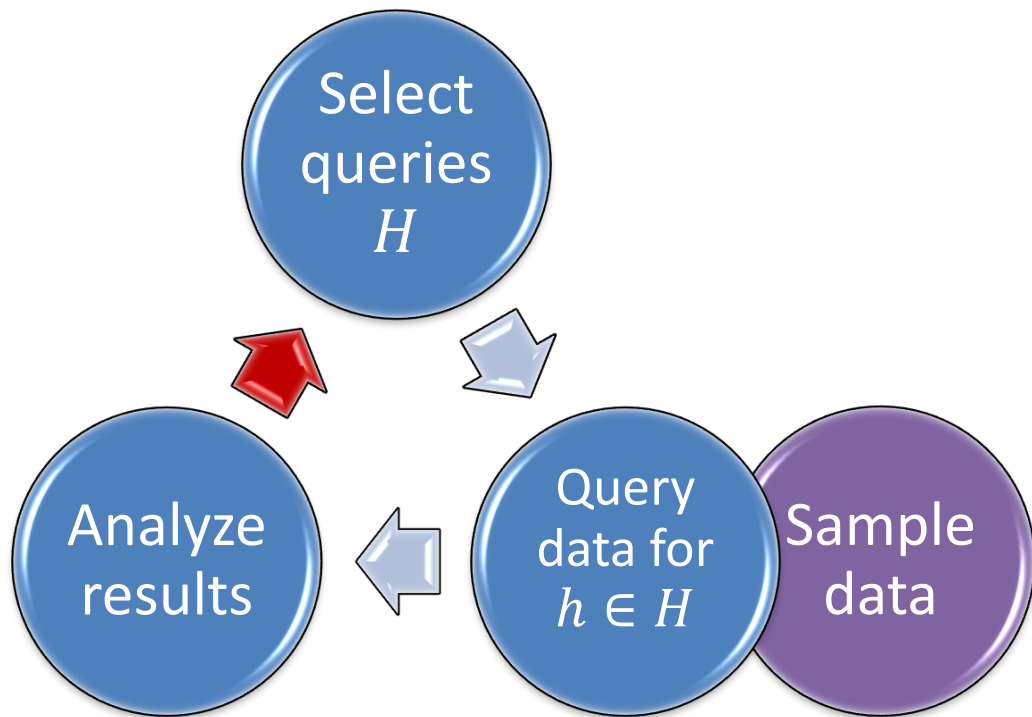
- **The worry:**  $h$  does not generalize, i.e.,
  - $h$  good predictor for the sample
  - $h$  not a good predictor on fresh samples taken from the distribution
- Luckily, for a predetermined  $H = \{h_1, h_2, \dots\}$ , if sample contains  $O(\log|H|)$  examples, then all  $h \in H$  generalize w.h.p.

a.k.a.  
Overfitting

# IN THEORY ...



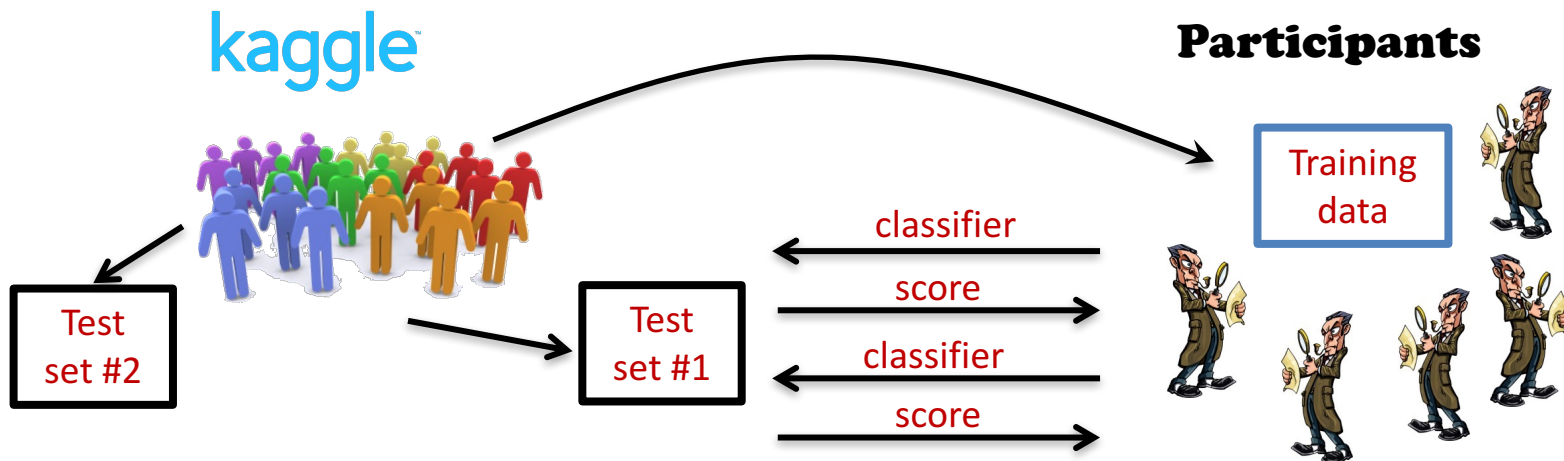
# IN PRACTICE



Analysts makes adaptive decisions:

- Queries selected based on **the results of previous analyses**
- Risk of false discoveries!
- Almost all existing approaches to ensuring generalization assume the entire data-analysis procedure is fixed ahead of time

# EXAMPLE: KAGGLE'S ML COMPETITION

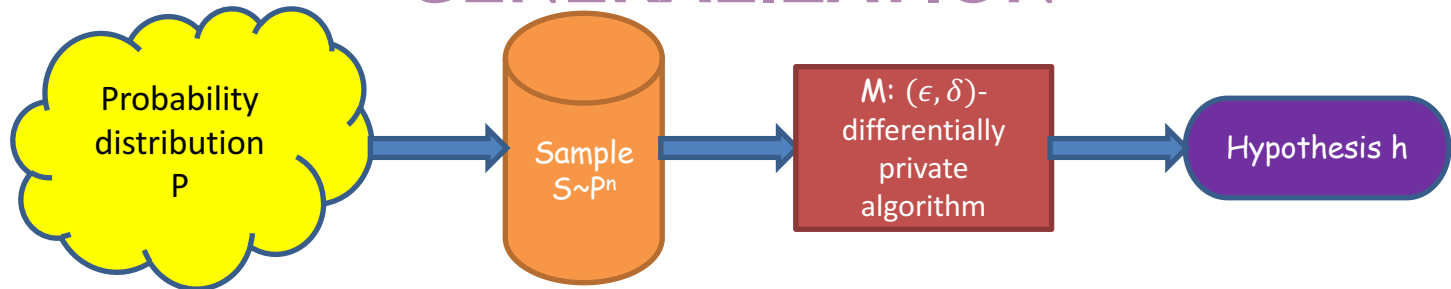


- Kaggle samples training data (made public) + 2 test sets (hidden)
- Participants can submit (multiple) classifiers during competition, these are evaluated on set #1.
- At the end, winner is determined by evaluating classifiers on training set #2.
- **Problem:** Best classifiers for set #1 are not best classifiers for set #2.
- Competitors are overfitting test set #1!

# PRIVACY AND GENERALIZATION??

Overfitting distinguishes who is in the dataset; If being in dataset is sensitive personal information  
→ a privacy issue

# DIFFERENTIAL PRIVACY → GENERALIZATION



- **Define:**  $h(S) = \frac{1}{n} \sum h(s_i)$  and  $h(P) = \Pr_{s \sim P}[h(s)]$
- **Our goal:** show that for  $h \leftarrow M(S)$ , w.h.p.  $h(S) \approx h(P)$

Theorem [McSherry, folklore]:

$$\mathbb{E}_{\substack{S \sim P \\ h \leftarrow M(S)}} [h(S)] \approx \mathbb{E}_{\substack{S \sim P \\ h \leftarrow M(S)}} [h(P)]$$

**Intuition:**  
consider two  
experiments:

$s_i$  : a random  
element of  $S$

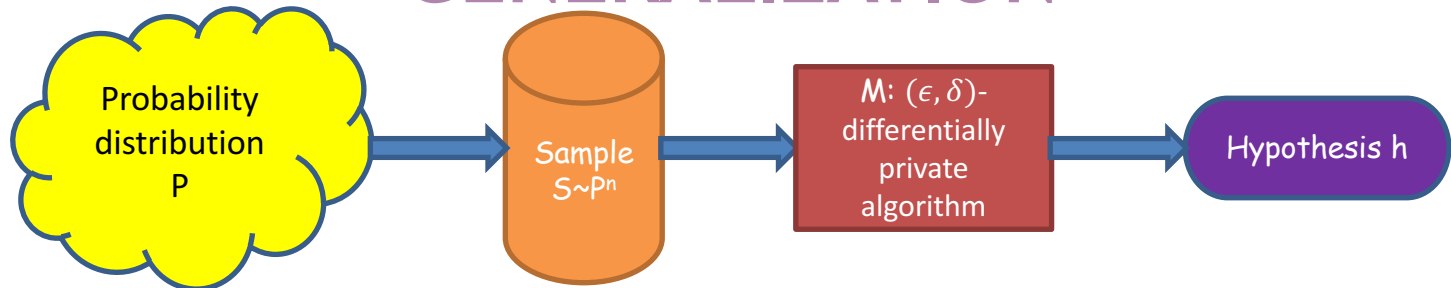
- $S = (s_1, \dots, s_n) \sim P$
- $i \in_R [n]$
- $h \leftarrow M(S)$
- Return  $h(s_i)$

$\approx$   
**DP**

- $S = (s_1, \dots, s_n) \sim P$
- $i \in_R [n]$
- $h \leftarrow M(S \setminus \{s_i\})$
- Return  $h(s_i)$

$s_i$  : a random  
element of  $P$

# DIFFERENTIAL PRIVACY → GENERALIZATION



- **Define:**  $h(S) = \frac{1}{n} \sum h(s_i)$  and  $h(P) = \Pr_{s \sim P}[h(s)]$
- **Our goal:** show that for  $h \leftarrow M(S)$ , w.h.p.  $h(S) \approx h(P)$

Theorem [McSherry, folklore]:	$\mathbb{E}_{\substack{S \sim P \\ h \leftarrow M(S)}} [h(S)] \approx \mathbb{E}_{\substack{S \sim P \\ h \leftarrow M(S)}} [h(P)]$	} Expectation  } High prob.
Theorem [DFHPRR' 15, BNSSSU'16]:	$\Pr_{\substack{S \sim P \\ h \leftarrow M(S)}} [ h(S) - h(P)  > \epsilon] \leq \delta/\epsilon$	

**Differential privacy closed under adaptive composition:** Even adaptive querying with differential privacy would not lead to a non-generalizing hypothesis.



# APPLICATION TO ADAPTIVE QUERYING

- Can import tools developed for answering queries adaptively with differential privacy!
- In particular, differential privacy allows approximating  $h(S) = \frac{1}{n} \sum h(s_i)$  for  $k \approx n^2$  adaptively selected predicates  $h_1, \dots, h_k$

## Upper bounds:

- [DFHPRR'15, BNSSSU'16]: Efficient mechanism that w.h.p. answers any  $k$  adaptively chosen queries  $h_1, \dots, h_k$  within accuracy  $\alpha$  given  $n = \tilde{O}(\sqrt{k}/\alpha^2)$  samples

## Lower bound:

- [Hardt Ullman 14, Steinke Ullman 15]: Any efficient mechanism that answers  $k$  adaptive queries within accuracy  $\alpha$  requires  $n = \Omega(\sqrt{k}/\alpha)$


**CONCLUSION**

# MAIN TAKEAWAYS

- **Accumulating failures:** anonymization & traditional SDL techniques.
- **Differential privacy:**
  - Not an algorithm; A standard providing a rigorous framework for developing privacy technologies with provable quantifiable guarantees.
  - Rich theoretical work, now transitioning to practice.
    - First real-world applications and use, including by US Census, Google, Apple.
  - Very strong protection for cases where data flows across trust boundaries.
  - Legal landscape needs to be taken into account; DP to be combined (wisely!) with other technical and policy tools.
- **Differential privacy:**
  - Leads towards strong tools for guaranteeing statistical validity

# DIFFERENTIAL PRIVACY AND YOU

- Research sensitive data that otherwise would not be available at all
- Collect good quality data
- Share data with other researchers/the public
- Provide respondent with strong quantifiable privacy that can be interpreted, rigorously, as bounding their risk from participation



Price in utility

- Ensuring statistical validity



A tool for ensuring utility

# RESOURCES

# LEARNING MORE ABOUT DIFFERENTIAL PRIVACY

- [Nissim et al, 2017] **Differential Privacy: A Primer for a Non-technical Audience**, Privacy Tools project.
- [Dwork 2011] **A Firm Foundation for Private Data Analysis**, CACM January 2011.
- [Heffetz & Ligett, 2014] **Privacy and Data-Based Research**, Journal of Economic Perspectives.
- [Dwork & Roth, 2014] **The Algorithmic Foundations of Differential Privacy**, Now publishers.
- [Vadhan'16] **The complexity of differential privacy**, Privacy Tools Project.
- + Online course material, lectures and tutorials.

Non-technical

technical

# PROJECTS, SOFTWARE TOOLS [PARTIAL LIST]

- [Microsoft Research] [PINQ](#)
- [UT Austin] [Airavat: Security & Privacy for MapReduce](#)
- [UC Berkeley] [GUPT](#)
- [CMU-Cornell-PennState] [Integrating Statistical and Computational Approaches to Privacy](#)
- [US Census] [OnTheMap](#)
- [Google] [Rappor](#)
- [UCSD] [Integrating Data for Analysis, Anonymization, and Sharing \(iDash\)](#)
- [Upenn] [Putting Differential Privacy to Work](#)
- [Stanford-Berkeley-Microsoft] [Towards Practicing Privacy](#)
- [Duke-NISS] [Triangle Census Research Network](#)
- [Harvard] [Privacy Tools](#)
- [Harvard, Georgetown, Buffalo] [Computing Over Distributed Sensitive Data](#)
- [Georgetown, Harvard, BU] [Formal Privacy Models and Title 13](#)