Responsible Data Science

Differential privacy

March 7 & 9, 2022

Prof. George Wood

Center for Data Science
New York University

NYU Center for Data Science

r/ai

Week 7 reading

Foundations and Trends® in
Theoretical Computer Science
Vol. 9, Nos. 3–4 (2014) 211–407

DOI: 10.1561/0400000042

now
the essence of knowledge

The Algorithmic Foundations of Differential Privacy

Cynthia Dwork
Microsoft Research, USA
dwork@microsoft.com

Aaron Roth
University of Pennsylvania, USA
aaroth@cis.upenn.edu

DOI:10.1145/1866739.1866758

What does it mean to preserve privacy?

BY CYNTHIA DWORK

A Firm Foundation for Private Data Analysis

IN THE INFORMATION realm, loss of privacy is usually associated with failure to control access to information, to control the flow of information, or to control the purposes for which information is employed. Differential privacy arose in a context in which ensuring privacy is a challenge even if all these control problems are solved: privacy-preserving statistical analysis of data.

The problem of *statistical disclosure control*—revealing accurate statistics about a set of respondents while preserving the privacy of individuals—has a venerable history, with an extensive literature spanning statistics, theoretical computer science, security, databases, and cryptography (see, for example, the excellent survey of Adam and Wortmann,[1] the discussion of related work in Blum et al.,[2] and the *Journal of Official Statistics* dedicated to confidentiality and disclosure control).

This long history is a testament to the importance of the problem. Statistical databases can be of enormous social value; they are used for apportioning resources, evaluating medical therapies, understanding the spread of disease, improving economic utility, and informing us about ourselves as a species.

The data may be obtained in diverse ways. Some data, such as census, tax, and other sorts of official data, is compelled; other data is collected opportunistically, for example, from traffic on the Internet, transactions on Amazon, and search engine query logs; other data is provided altruistically, by respondents who hope that sharing their information will help others to avoid a specific misfortune, or more generally, to increase the public good. Altruistic data donors are typically promised their individual data will be kept confidential—in short, they are promised "privacy." Similarly, medical data and legally compelled data, such as census data and tax return data, have legal privacy

» key insights

- **In analyzing private data, only by focusing on rigorous privacy guarantees can we convert the cycle of "propose-break-propose again" into a path of progress.**
- **A natural approach to defining privacy is to require that accessing the database teaches the analyst nothing about any individual. But this is problematic: the whole point of a statistical database is to teach general truths, for example, that smoking causes cancer. Learning this fact teaches the data analyst something about the likelihood with which certain individuals, not necessarily in the database, will develop cancer. We therefore need a definition that separates the utility of the database (learning that smoking causes cancer) from the increased risk of harm due to joining the database. This is the intuition behind *differential privacy*.**
- **This can be achieved, often with low distortion. The key idea is to randomize responses so as to effectively hide the presence or absence of the data of any individual *over the course of the lifetime of the database*.**

r/ai

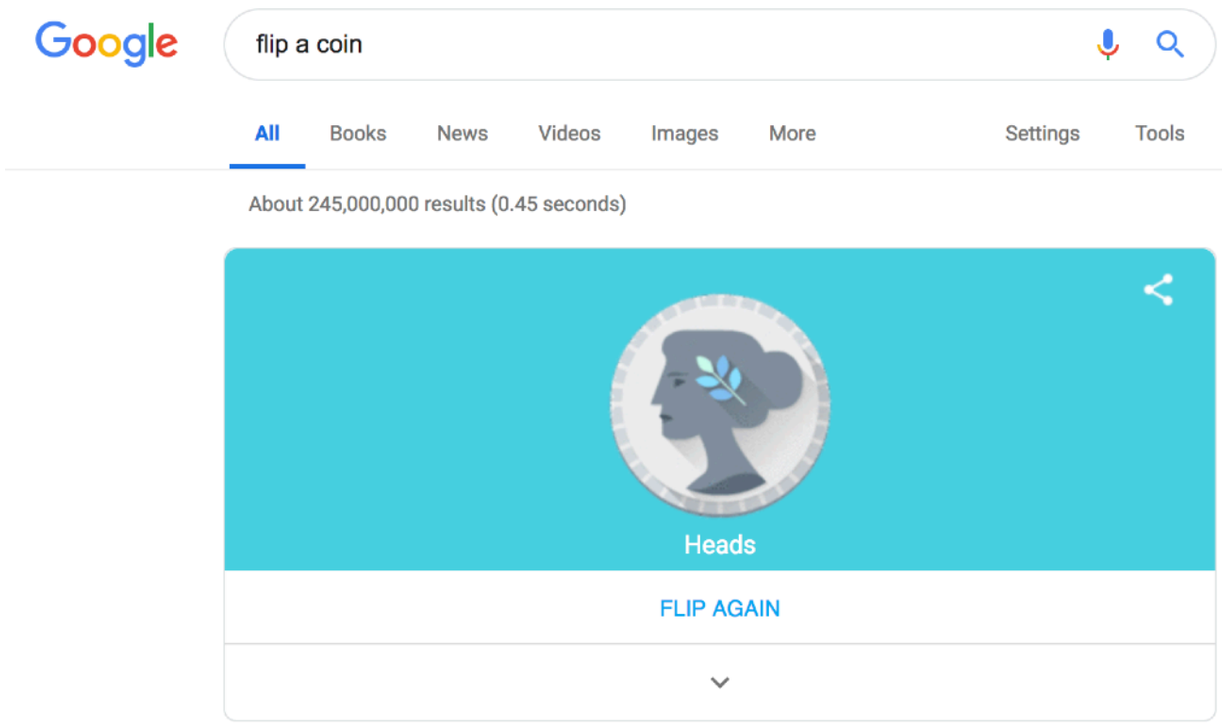
motivation

Truth or dare?

Did you go out drinking over the weekend?

let's call this property **P** (Truth=Yes) and estimate **p**, the fraction of the class for whom **P** holds

1. flip a coin **C1**
    1. if **C1** is tails, then **respond truthfully**
    2. if **C1** is heads, then flip another coin **C2**
        1. if **C2** is heads then **Yes**
        2. else **C2** is tails then respond **No**



the expected number of **Yes** answers is:

$$A = \frac{3}{4}p + \frac{1}{4}(1-p) = \frac{1}{4} + \frac{p}{2}$$

thus, we estimate **p** as:

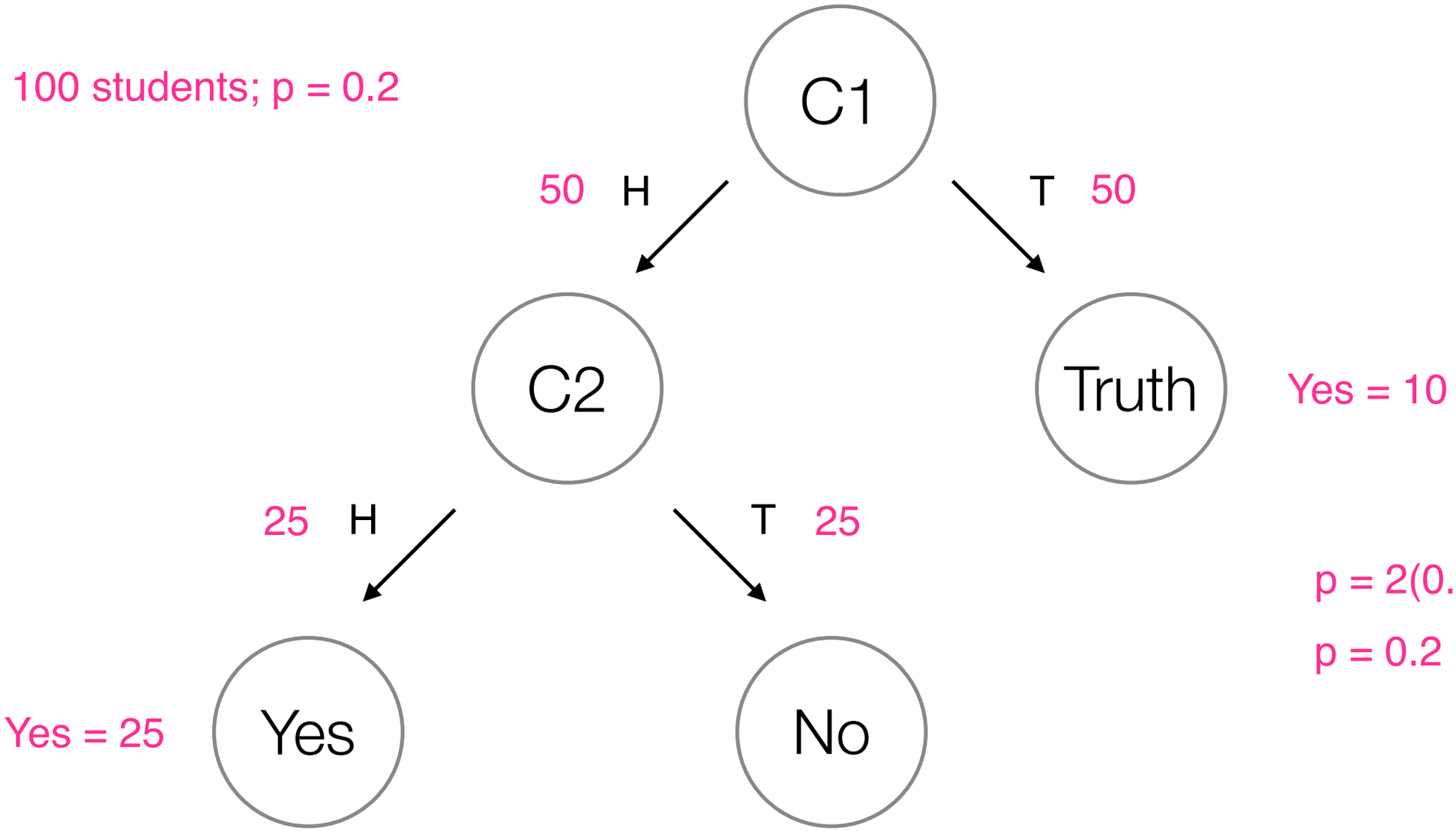
$$\tilde{p} = 2A - \frac{1}{2}$$

r/ai

Truth or dare?

Did you go out drinking over the weekend?

100 students; p = 0.2



p = 2(0.1 + 0.25) - 0.5

p = 0.2

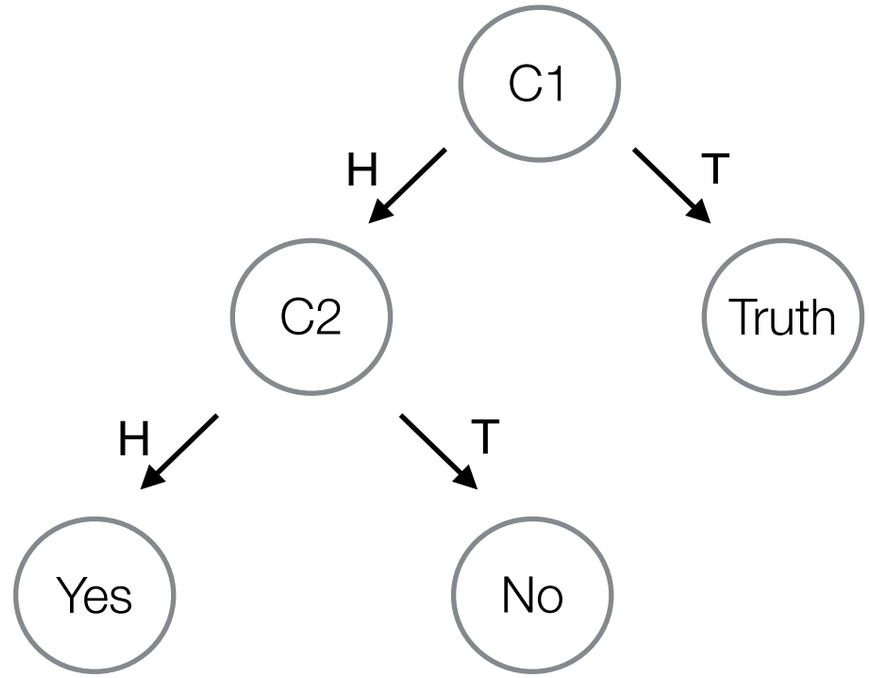
$$\tilde{p} = 2A - \frac{1}{2}$$

Truth or dare?

Did you go out drinking over the weekend?

Student	Response
A	no
B	yes
C	no
D	no
E	no
F	yes
G	no
H	no
I	no
…	…

T or HH?



T or HH?

Randomized response

Did you go out drinking over the weekend?

let's call this property **P** (Truth=Yes) and estimate **p**, the fraction of the class for whom **P** holds

1. flip a coin **C1**
   1. if **C1** is tails, then **respond truthfully**
   2. if **C1** is heads, then flip another coin **C2**
      1. if **C2** is heads then **Yes**
      2. else **C2** is tails then respond **No**

randomization - adding noise - is what gives plausible deniability a process privacy method

the expected number of **Yes** answers is:

$$A = \frac{3}{4}p + \frac{1}{4}(1-p) = \frac{1}{4} + \frac{p}{2}$$

privacy comes from plausible deniability

r/ai

Privacy: two sides of the coin

protecting an individual

plausible deniability

learning about the population

noisy estimates

do we really need randomization?

r/ai

Some other options

- Data release approaches that fail to protect privacy (these are prominent classes of methods, there are others):
  - **sampling** (“**just a few**”) - release a small subset of the database
  - **aggregation** (e.g., **k-anonymity** - each record in the release is indistinguishable from at least k-1 other records)
  - **de-identification** - mask or drop personal identifiers
  - **query auditing** - stop answering queries when they become unsafe

Sampling (“just a few”)

- Suppose that we take a random small sample ***D’*** of ***D*** and release it without any modification
- If ***D’*** is much smaller than ***D***, then every respondent is unlikely to appear in ***D’***
- This technique provides protection for “the typical” (or for “most”) members of the dataset
- But it may be argued that **atypical** individuals are the ones needing stronger protection!
- In any case, this method is problematic because a respondent who does appear has **no plausible deniability!**
- Suppose next that appearing in the sample ***D’*** has terrible consequences. Then, every time subsampling occurs - some individual suffers horribly!

Aggregation without randomization

- Alice and Bob are professors at State University.
- In March, Alice publishes an article: "…. the current freshman class at State U is **3,005** students, **202** of whom are from families earning over $1M per year."
- In April, Bob publishes an article: "… **201** families in State U's freshman class of **3,004** have household incomes exceeding $1M per year."
- Neither statement discloses the income of the family of any one student. But, taken together, they state that **John, a student who dropped out at the end of March**, comes from a family that earns $1M. Anyone who has this **auxiliary information** — that John dropped out at the end of March — will be able to learn about the income of John's family.

this is known as a problem of **composition**, and can be seen as a kind of a **differencing attack**

A basic differencing attack

- **X**: count the number of HIV-positive people in ***D***
- **Y**: count the number of HIV-positive people in ***D*** not named *Freddie*;
- **X - Y** tells you whether *Freddie* is HIV-positive

what if X-Y > 1, do we still have a problem?

Reconstruction: death by a 1000 cuts

- Another serious issue for aggregation without randomization, or with an insufficient amount of randomization: **reconstruction attacks**

- **The Fundamental Law of Information Recovery** (starting with the seminal results by Irit Dinur & Kobbi Nissim, PODS 2003): overly accurate estimates of too many statistics can completely destroy privacy

- Under what conditions can an adversary reconstruct a candidate database ***D'*** that agrees with the real database ***D*** in **99%** of the entries?

- Suppose that ***D*** has ***n*** tuples, and that noise is bounded by some quantity ***E***. Then there exists an adversary that can reconstruct ***D*** to within ***4E*** positions, issuing all possible $\mathbf{2^n}$ queries

$$4E = \frac{4n}{401} < \frac{n}{100}$$

- Put another way: if the magnitude of the noise is less than n/401, then 99% of ***D*** can be reconstructed by the adversary. Really, any number higher than 401 will work

- **There are also reconstruction results under a limited number of queries**

Reconstruction: death by a 1000 cuts

Privacy-Preserving Data Analysis for the Federal Statistical Agencies

January 2017

John Abowd, Lorenzo Alvisi, Cynthia Dwork, Sampath Kannan, Ashwin Machanavajjhala, and Jerome Reiter

CCC
Computing Community Consortium
Catalyst

we'll discuss the use of differential privacy by the 2020 US Census later today

The Fundamental Law of Information Recovery has troubling implications for the publication of large numbers of statistics by a statistical agency: it says that the confidential data may be vulnerable to database reconstruction attacks based entirely on the data published by the agency itself. **Left unattended, such risks threaten to undermine, or even eliminate, the societal benefits inherent in the rich data collected by the nation's statistical agencies.** The most pressing immediate problem for any statistical agency is how to modernize its disclosure limitation methods in light of the Fundamental Law.

De-identification

- Also known as **anonymization**
- Mask or drop identifying attribute or attributes, such as social security number (SSN), name, mailing address
- Turns out that this also doesn't work because **auxiliary information** is available
- Fundamentally, this is due to **the curse of dimensionality**: high-dimensional data is sparse, the more you know about individuals, the less likely it is that two individuals will look alike

de-identified data can be re-identified with a linkage attack

A linkage attack: Governor Weld

In 1997, Massachusetts Group Insurance Commission released "anonymized" data on state employees that showed every single hospital visit!

Latanya Sweeney, a grad student, sought to show the ineffectiveness of this “anonymization.”

She knew that Governor Weld resided in Cambridge, Massachusetts, a city of 54,000 residents and seven ZIP codes.

For twenty dollars, she purchased the complete voter rolls from the city of Cambridge, a database containing, among other things, the name, address, ZIP code, birth date, and sex of every voter.

Only six people in Cambridge shared his birth date, only three of them men, and of them, only he lived in his ZIP code.

Follow up: ZIP code, birthdate, and sex sufficient to identify 87% of Americans!

https://arstechnica.com/tech-policy/2009/09/your-secrets-live-online-in-databases-of-ruin/

The Netflix prize linkage attack

- In 2006, Netflix released a dataset containing ~100M **movie ratings** by ~500K users (about 1/8 of the Nexflix user base at the time)

- **FAQ**: "Is there any customer information in the dataset that should be kept private?"

*"No, all customer identifying information has been removed; all that remains are ratings and dates. This follows our privacy policy, which you can review here. Even if, for example, you knew all your own ratings and their dates you probably couldn't identify them reliably in the data because only **a small sample** was included (less than one-tenth of our complete dataset) and that **data was subject to perturbation**. Of course, since you know all your own ratings that really isn't a privacy problem is it?"*

The real question: How much does the adversary need to know about a Netflix subscriber to identify her record in the dataset, and thus learn her complete movie viewing history?

[Narayanan and Shmatikov, *IEEE S&P 2008*]

r/ai

The Netflix prize linkage attack

- Very little auxiliary information is needed to de-anonymize an average subscriber record from the Netflix Prize dataset
- **Perturbation, you say?** With 8 movie ratings (of which 2 may be completely wrong) and dates that may have a 14-day error, 99% of records be uniquely identified in the dataset
- For 68%, two ratings and dates (with a 3-day error) are sufficient
- **Even without any dates, a substantial privacy breach occurs, especially when the auxiliary information consists of movies that are not blockbusters:** Two movies are no longer sufficient, but 84% of subscribers can be uniquely identified if the adversary knows 6 out of 8 moves outside the top 500

We cannot assume a priori that any data is harmless!

[Narayanan and Shmatikov, *IEEE S&P 2008*]

r/ai

The Netflix prize linkage attack

WIRED

RYAN SINGEL SECURITY 12.17.09 04:29 PM

NETFLIX SPILLED YOUR BROKEBACK MOUNTAIN SECRET, LAWSUIT CLAIMS

An in-the-closet lesbian mother is suing Netflix for privacy invasion, alleging the movie rental company made it possible for her to be outed when it disclosed insufficiently anonymous information about nearly half-a-million customers as part of its $1 million contest to improve its recommendation system.

The suit known as Doe v. Netflix (.pdf) was filed in federal court in California on Thursday, alleging that Netflix violated fair-trade laws and a federal privacy law protecting video rental records, when it launched its popular contest in September 2006.

The suit seeks more than $2,500 in damages for each of more than 2 million Netflix customers.

r/ai

The Netflix prize linkage attack

WIRED

RYAN SINGEL SECURITY 03.12.10 02:48 PM

NETFLIX CANCELS RECOMMENDATION CONTEST AFTER PRIVACY LAWSUIT

Netflix is canceling its second $1 million Netflix Prize to settle a legal challenge that it breached customer privacy as part of the first contest's race for a better movie-recommendation engine.

r/ai

Query auditing

- Monitor queries: each query is granted or denied depending on what other queries were answered in the past
- If this method were to work, it could be used to detect that a differencing attack is about to take place
- Unfortunately, it doesn't work:
- **Query auditing is computationally infeasible**
- Refusal to respond to a query may itself be disclosive
- We refuse to execute a query, then what? No information access at all?

Query auditing

- We have a set of (secret) Boolean variables ***X*** and the result of some *statistical queries* over this set
- A *statistical query* ***Q*** specifies a subset ***S*** of the variables in ***X***, and returns the sum of the values of all variables in ***S***

Example:

Relation `Employees (name, age, salary)`

Query `select sum(salary) from Employees where age > 35`

Suppose that `Employees (name, age)` is public, but `salary` is confidential

[Kleinberg, Papadimitriou, Raghavan, *PODS 2000*]

Query auditing

- We have a set of (secret) Boolean variables ***X*** and the result of some *statistical queries* over this set
- A *statistical query* ***Q*** specifies a subset ***S*** of the variables in ***X***, and returns the sum of the values of all variables in ***S***
- **The auditing problem:** Decide whether the value of any Boolean variable is determined by the results of the queries
- **Main result:** The Boolean auditing problem is **coNP-complete**
  - coNP-complete is the hardest class of problems in coNP: all coNP problems can be formulated as a special case of any coNP-complete problem

[Kleinberg, Papadimitriou, Raghavan, *PODS 2000*]

r/ai

privacy-preserving data analysis

r/ai

Privacy: two sides of the coin

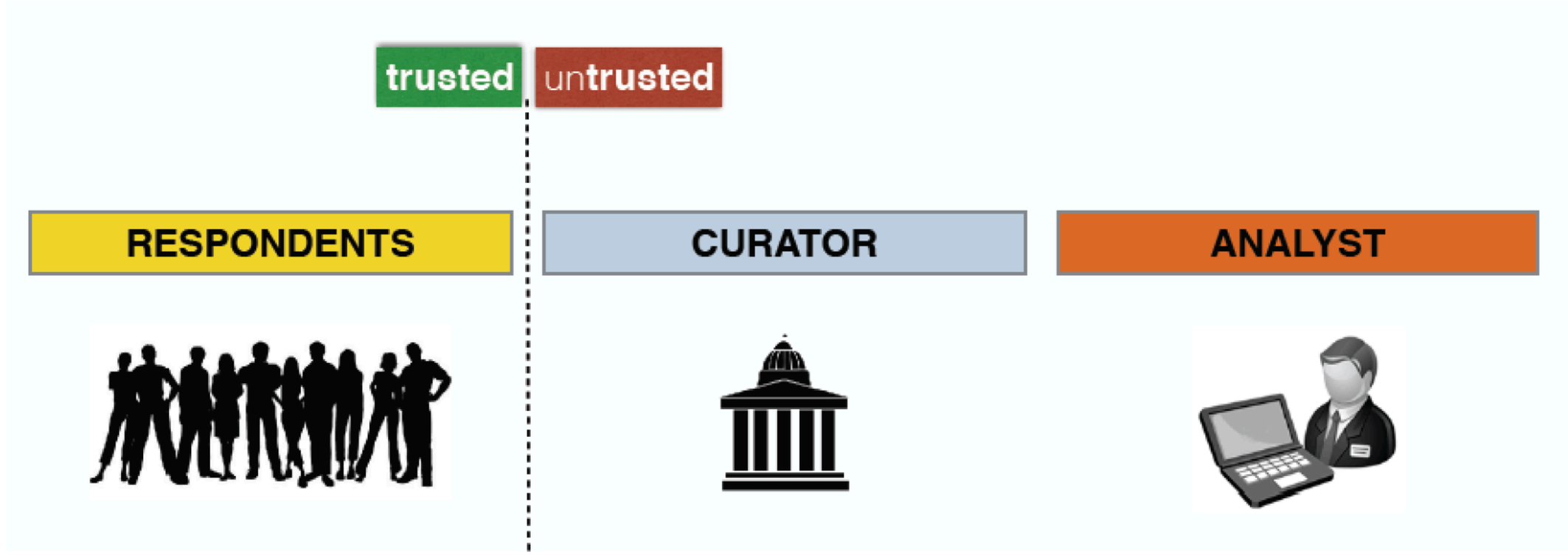
protecting an individual

plausible deniability

learning about the population

noisy estimates

Privacy-preserving data analysis

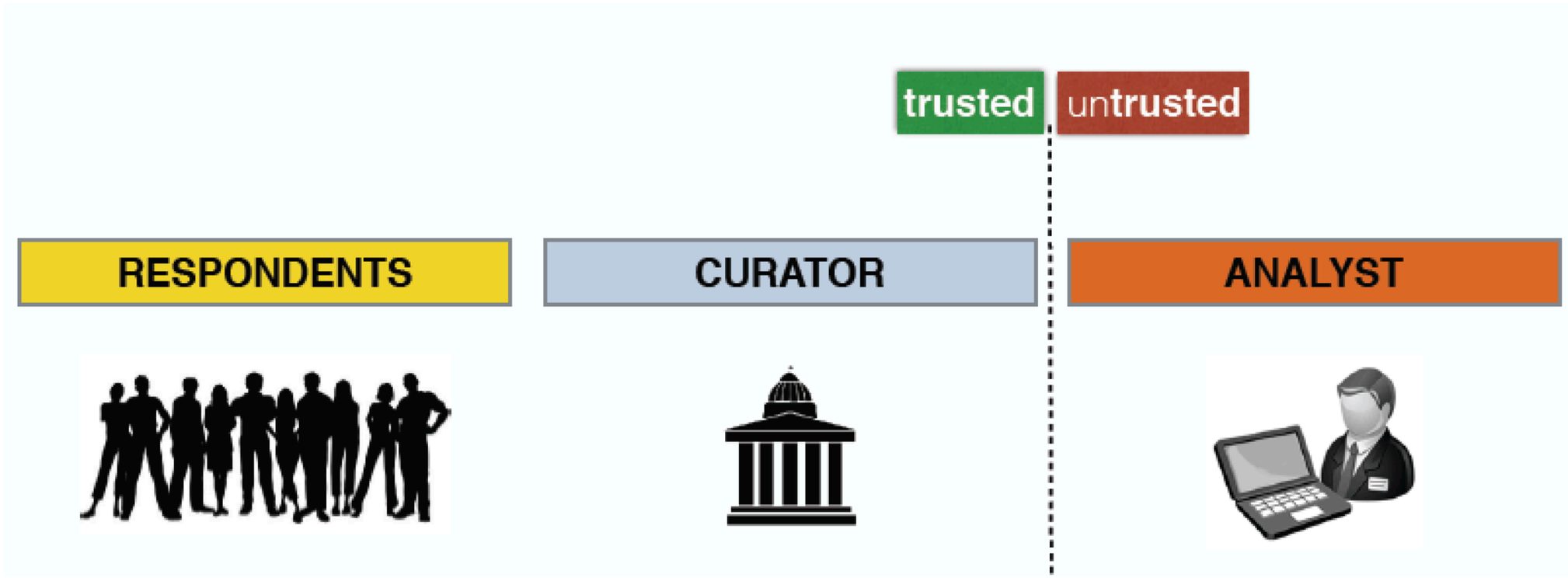


respondents contribute their personal data

the **curator** is **untrusted**, collects data, releases it to analysts

the **analyst** is **untrusted**, extracts value from data

Privacy-preserving data analysis

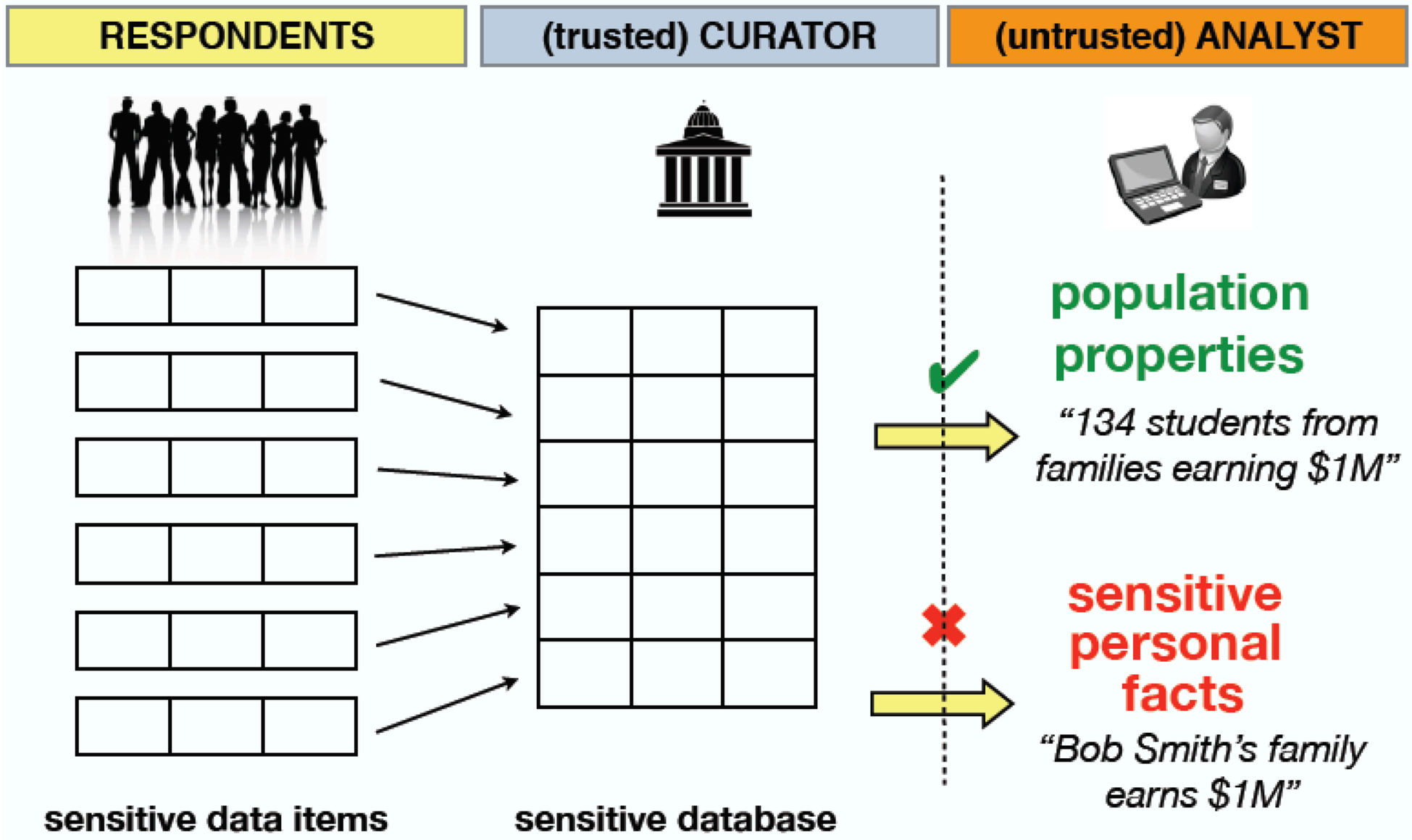


respondents in the population seek protection of their personal data

the **curator** is **trusted** to collect data and is responsible for safely releasing it

the **analyst** is **untrusted** and wants to gain the most accurate insights into the population

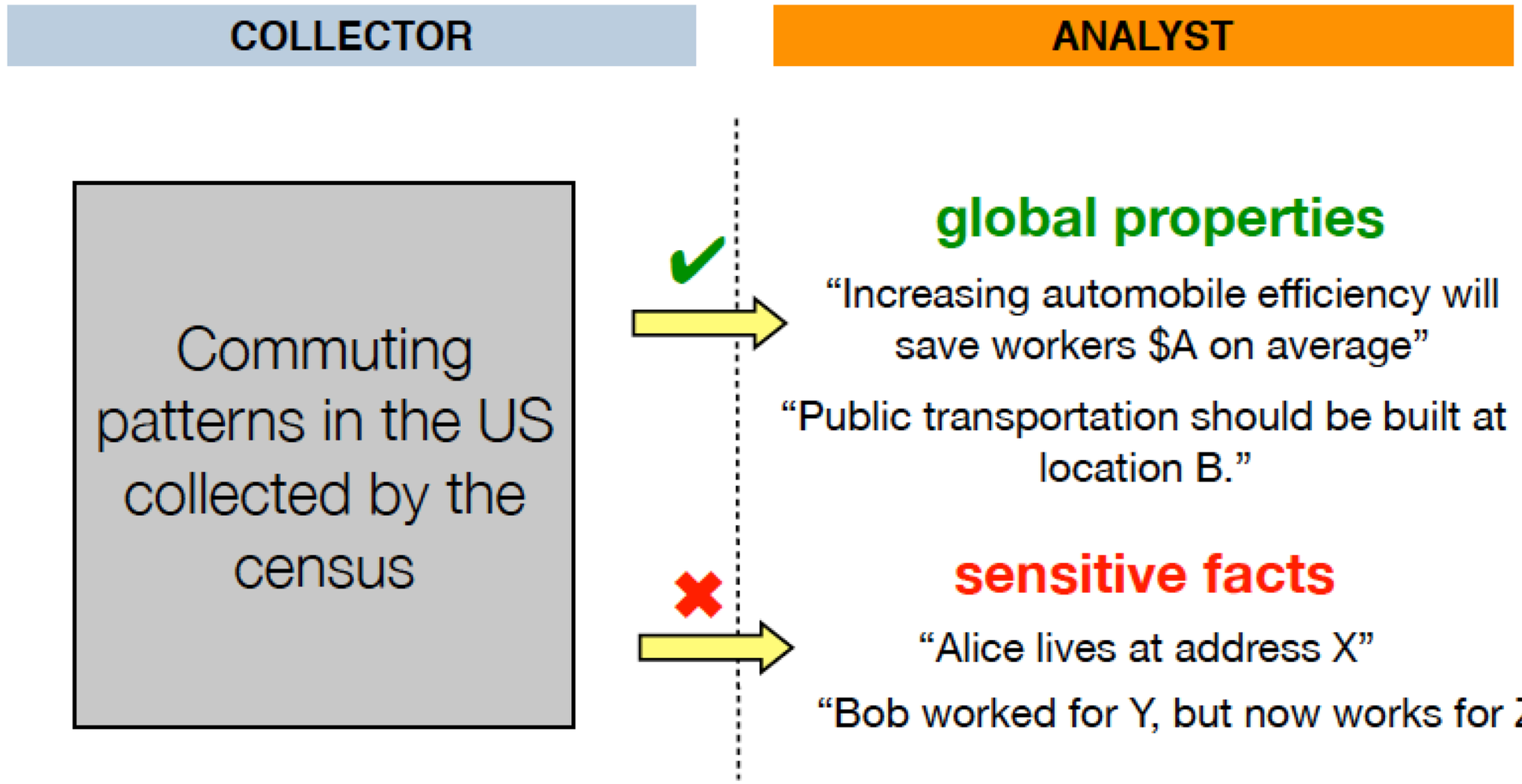
Privacy-preserving data analysis



slide by Gerome Miklau

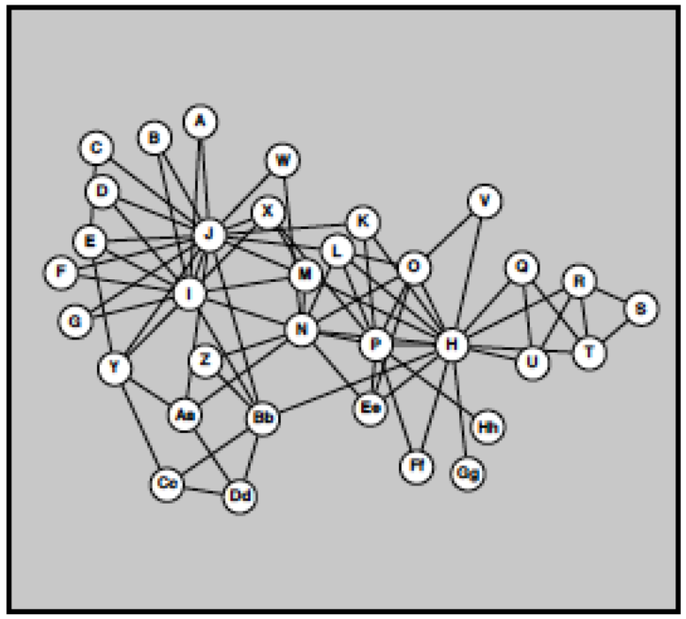
r/ai

Example: US Census



Example: Social networks

COLLECTOR



sensitive data set

ANALYST

✔ **global properties**

"How rapidly do rumors spread in this network?"

"Are people most likely to form friendships with those who share their attributes?"

✖ **sensitive facts**

"Alice is present in this network"

"Alice and Bob are connected"

Defining private data analysis

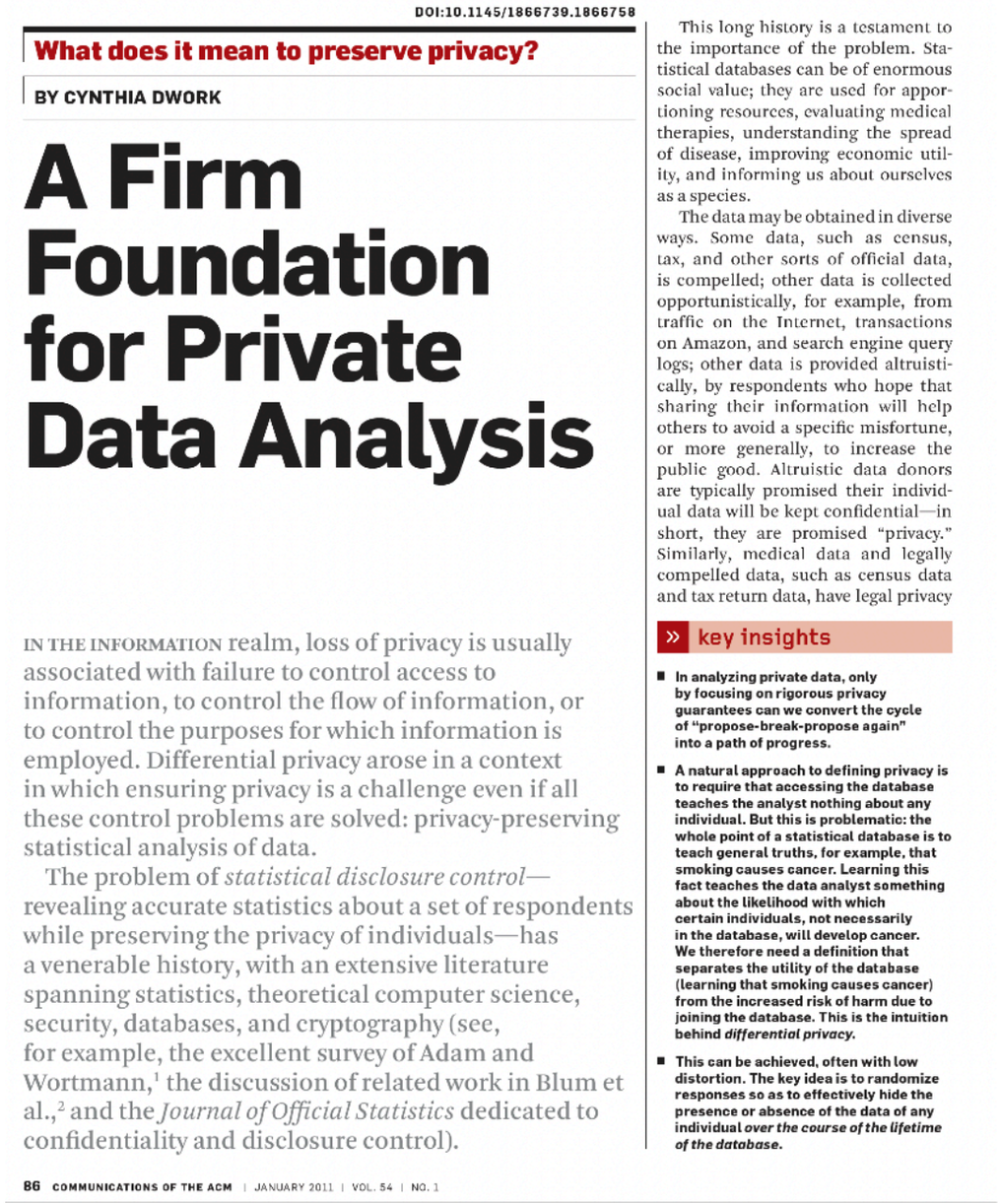
- Take 1: If **nothing is learned** about any individual in the dataset, then no individual can be harmed by analysis.
  - **Dalenius' Desideratum**: an *ad omnia* (Latin: "for all") privacy goal for statistical databases, as opposed to *ad hoc* (Latin: "for this"). Anything that can be learned about a respondent from the statistical database should be learnable without access to the database.
  - Put another way, the adversary's prior and posterior views about an individual should not be different.
  - This objective is **unachievable** because of auxiliary information.
  - **Example**: Alice knows that John smokes. She read a medical research study that found a causal relationship between smoking and lung cancer. Alice concludes, based on study results and her prior knowledge about John, that he has a heightened risk of developing lung cancer.
- Further, the risk is to everyone in a particular group (smokers, in this example), **irrespective of whether they participated in the study**.

Defining private data analysis

- Take 1: If **nothing is learned** about any individual in the dataset, then no individual can be harmed by analysis.
    - **Dalenius' Desideratum**: an "*ad omnia*" (opposed to *ad hoc*) privacy goal for statistical databases: Anything that can be learned about a respondent from the statistical database should be learnable without access to the database.
    - Put another way, the adversary's prior and posterior views about an individual should not be different.

- Take 2: The information released about the sensitive dataset is virtually indistinguishable **whether or not a respondent's data is in the dataset**. This is an informal statement of **differential privacy**: that no information **specific to an individual** is revealed.

Defining private data analysis



DOI:10.1145/1866739.1866758

What does it mean to preserve privacy?

BY CYNTHIA DWORK

A Firm Foundation for Private Data Analysis

IN THE INFORMATION realm, loss of privacy is usually associated with failure to control access to information, to control the flow of information, or to control the purposes for which information is employed. Differential privacy arose in a context in which ensuring privacy is a challenge even if all these control problems are solved: privacy-preserving statistical analysis of data.

The problem of *statistical disclosure control*—revealing accurate statistics about a set of respondents while preserving the privacy of individuals—has a venerable history, with an extensive literature spanning statistics, theoretical computer science, security, databases, and cryptography (see, for example, the excellent survey of Adam and Wortmann,[1] the discussion of related work in Blum et al.,[2] and the *Journal of Official Statistics* dedicated to confidentiality and disclosure control).

This long history is a testament to the importance of the problem. Statistical databases can be of enormous social value; they are used for apportioning resources, evaluating medical therapies, understanding the spread of disease, improving economic utility, and informing us about ourselves as a species.

The data may be obtained in diverse ways. Some data, such as census, tax, and other sorts of official data, is compelled; other data is collected opportunistically, for example, from traffic on the Internet, transactions on Amazon, and search engine query logs; other data is provided altruistically, by respondents who hope that sharing their information will help others to avoid a specific misfortune, or more generally, to increase the public good. Altruistic data donors are typically promised their individual data will be kept confidential—in short, they are promised "privacy." Similarly, medical data and legally compelled data, such as census data and tax return data, have legal privacy

» key insights

- In analyzing private data, only by focusing on rigorous privacy guarantees can we convert the cycle of "propose-break-propose again" into a path of progress.
- A natural approach to defining privacy is to require that accessing the database teaches the analyst nothing about any individual. But this is problematic: the whole point of a statistical database is to teach general truths, for example, that smoking causes cancer. Learning this fact teaches the data analyst something about the likelihood with which certain individuals, not necessarily in the database, will develop cancer. We therefore need a definition that separates the utility of the database (learning that smoking causes cancer) from the increased risk of harm due to joining the database. This is the intuition behind *differential privacy*.
- This can be achieved, often with low distortion. The key idea is to randomize responses so as to effectively hide the presence or absence of the data of any individual *over the course of the lifetime of the database*.

"A natural approach to defining privacy is to require that accessing the database teaches the analyst nothing about any individual. But this is problematic: **the whole point of a statistical database is to teach general truths**, for example, that smoking causes cancer. Learning this fact teaches the data analyst something about the likelihood with which certain individuals, not necessarily in the database, will develop cancer. We therefore **need a definition that separates the utility of the database** (learning that smoking causes cancer) **from the increased risk of harm due to joining the database. This is the intuition behind differential privacy.** "

r/ai

differential privacy (DP)

r/ai

Differential privacy: the formalism

We will define privacy with respect to a database ***D*** that is made up of rows (equivalently, tuples) representing individuals. Tuples come from some universe of datatypes (the set of all possible tuples).

The ℓ_1 norm of a database ***D***, denoted $\|D\|_1$ is the number of tuples in ***D***.

The ℓ_1 distance between databases ***D₁*** *and* ***D₂*** represents the number of tuples on which they differ. $\|D_1 - D_2\|_1$

We refer to a pair of databases that differ in at most 1 tuple as **neighboring databases** $\|D_1 - D_2\|_1 \leq 1$

Of these ***D₁*** and ***D₂***, one, say ***D₁,*** is a subset of the other, and, when a proper subset, the larger database ***D₂*** contains 1 extra tuple.

Differential privacy: the formalism

The information released about the sensitive dataset is virtually indistinguishable **whether or not a respondent's data is in the dataset**. This is an informal statement of **differential privacy**. That is, no information **specific to an individual** is revealed.

A randomized algorithm ***M*** provides **ε-differential privacy** if, for all neighboring databases D_1 and D_2, and for any set of outputs ***S***:

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

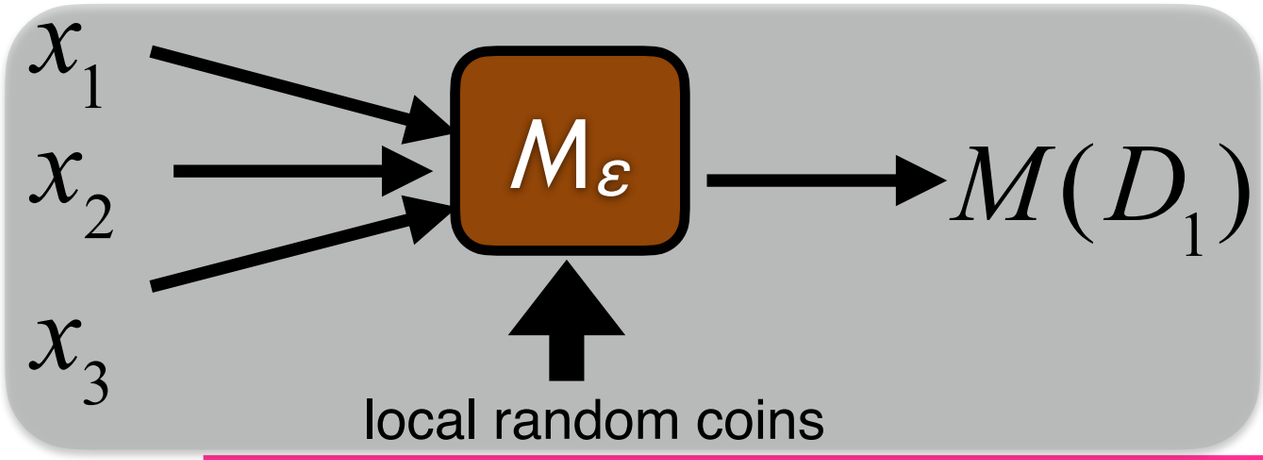
ε (epsilon) is a privacy parameter

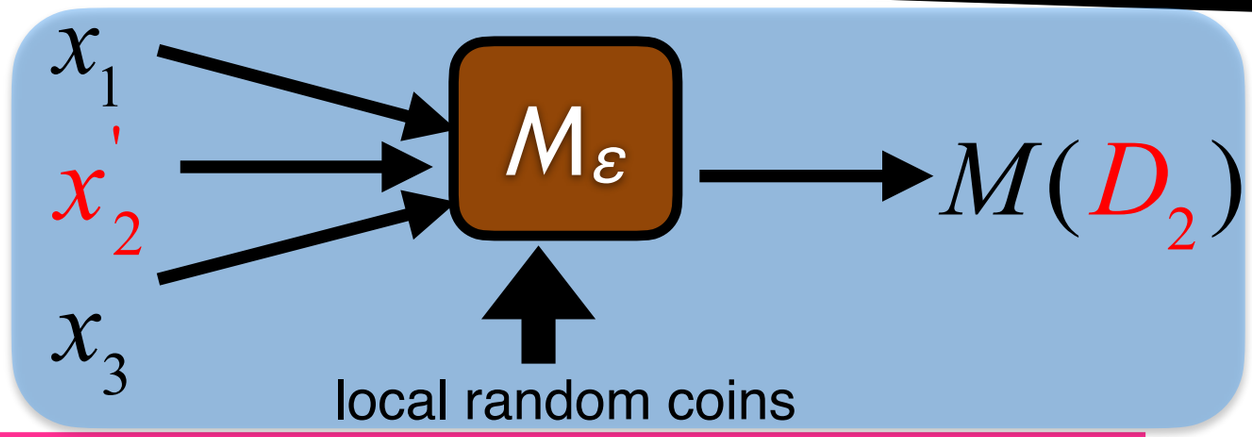
lower ε = stronger privacy

The notion of **neighboring databases** is integral to plausible deniability: D_1 can represent a database with a particular respondent's data, D_2 can represent a neighboring database but without that respondent's data

r/ai

Differential privacy: the formalism





A randomized algorithm ***M*** provides **ε-differential privacy** if, for all neighboring databases ***D₁*** and ***D₂***, and for any set of outputs ***S***:

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

Think of database of respondents ***D=(x₁, .., xₙ)*** as **fixed** (not random), ***M(D)*** is a random variable distributed over possible outputs

Neighboring databases induce **close distributions** on outputs

based on a slide by Adam Smith

r/ai

Back to randomized response

Did you go out drinking over the weekend?

1. flip a coin **C1**
    1. if **C1** is tails, then **respond truthfully**
    2. if **C1** is heads, then flip another coin **C2**
        1. if **C2** is heads then **Yes**
        2. else **C2** is tails then respond **No**

Denote:

- Truth=Yes by **P**
- Response=Yes by **A**
- **C1**=tails by **T**
- **C1**=heads and **C2**=tails by **HT**
- **C1**=heads and **C2**=heads by **HH**

A randomized algorithm ***M*** provides **ε-differential privacy** if, for all neighboring databases D_1 and D_2, and for any set of outputs ***S***:

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

$$\Pr[A \mid P] = \Pr[T] + \Pr[HH] = \frac{3}{4}$$

$$\Pr[A \mid \neg P] = \Pr[HH] = \frac{1}{4}$$

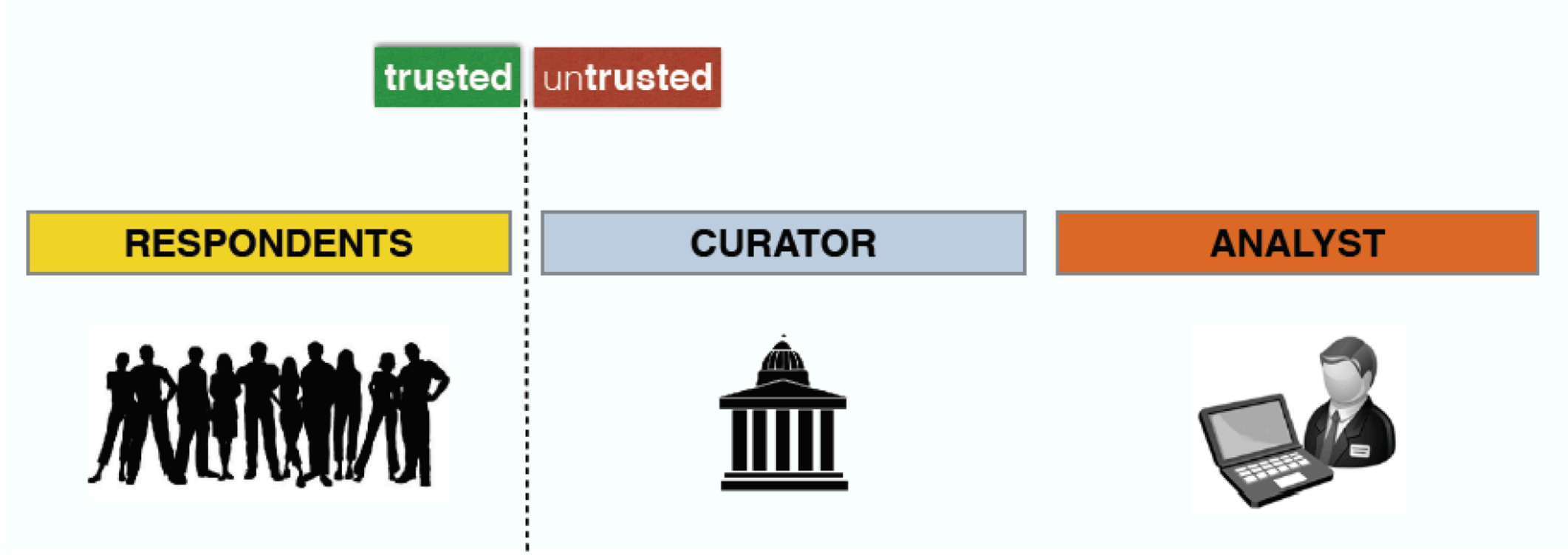
$$\Pr[A \mid P] = 3 \Pr[A \mid \neg P]$$

$$\Rightarrow \varepsilon = \ln 3$$

our version of randomized response is $(\ln 3)$-differentially private

r/ai

Local differential privacy



respondents contribute their personal data

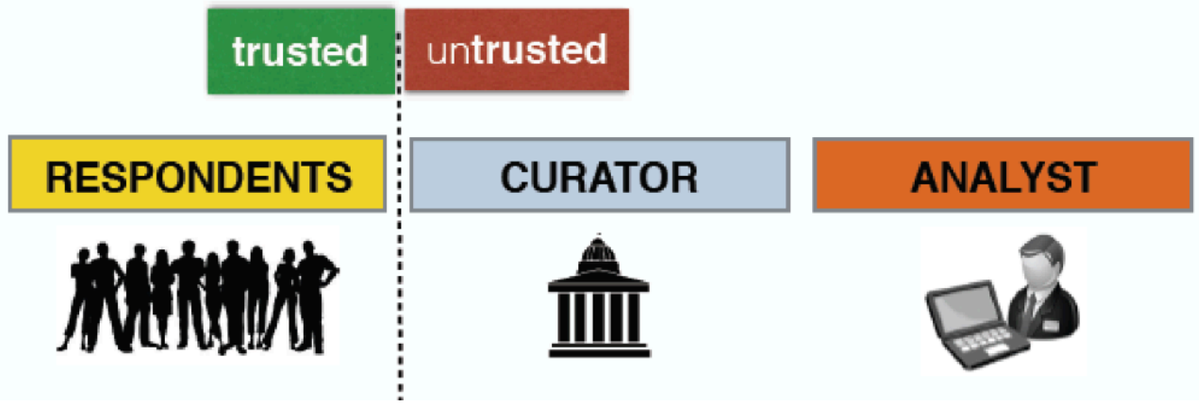
the **curator** is **untrusted**, collects data, releases it to analysts

the **analyst** is **untrusted**, extracts value from data

Differential privacy in the field

Apple

Google



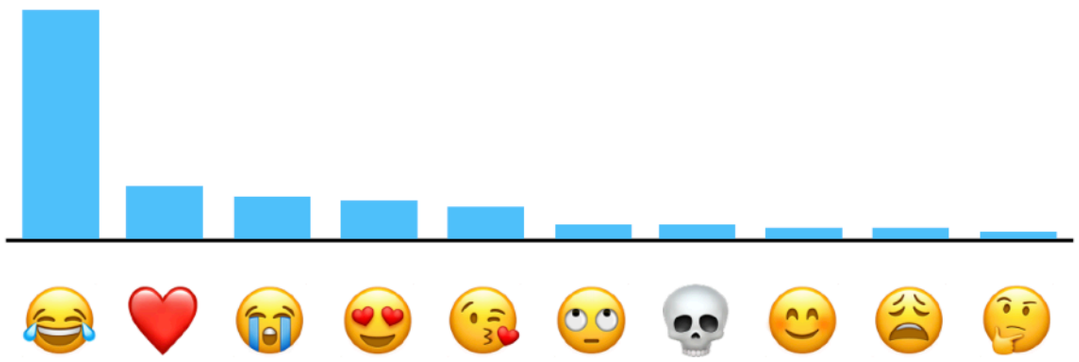
slide by Gerome Miklau

r/ai

Example: What's your favorite emoji?

A privacy-preserving system

Apple has adopted and further developed a technique known in the academic world as *local differential privacy* to do something really exciting: gain insight into what many Apple users are doing, while helping to preserve the privacy of individual users. It is a technique that enables Apple to learn about the user community without learning about individuals in the community. Differential privacy transforms the information shared with Apple before it ever leaves the user's device such that Apple can never reproduce the true data.



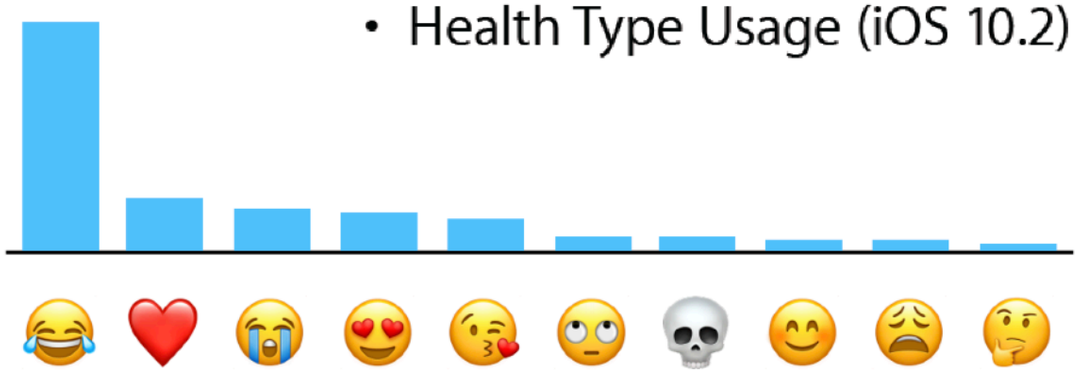
https://www.apple.com/privacy/docs/Differential_Privacy_Overview.pdf

r/ai

Example: What's your favorite emoji?

Apple uses local differential privacy to help protect the privacy of user activity in a given time period, while still gaining insight that improves the intelligence and usability of such features as:

- QuickType suggestions
- Emoji suggestions
- Lookup Hints
- Safari Energy Draining Domains
- Safari Autoplay Intent Detection (macOS High Sierra)
- Safari Crashing Domains (iOS 11)
- Health Type Usage (iOS 10.2)



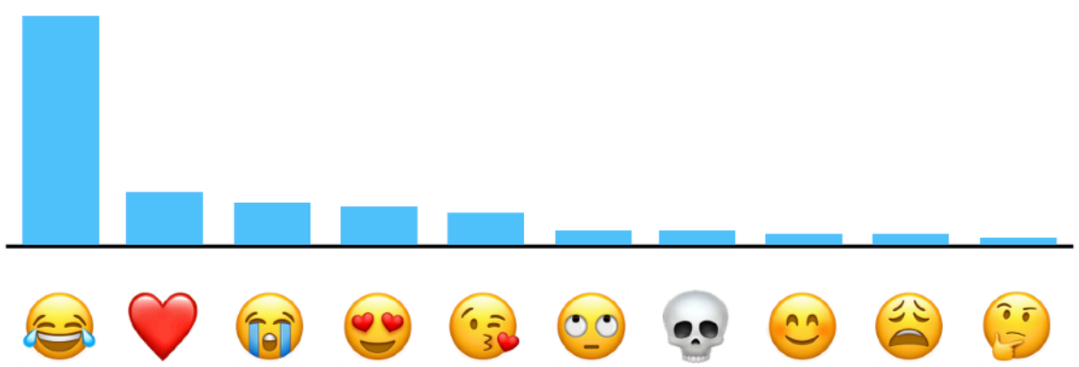
https://www.apple.com/privacy/docs/Differential_Privacy_Overview.pdf

r/ai

Example: What's your favorite emoji?

Privacy budget

The Apple differential privacy implementation incorporates the concept of a per-donation *privacy budget* (quantified by the parameter epsilon), and sets a strict limit on the number of contributions from a user in order to preserve their privacy. The reason is that the slightly-biased noise used in differential privacy tends to average out over a large numbers of contributions, making it theoretically possible to determine information about a user's activity over a large number of observations from a single user (though it's important to note that Apple doesn't associate any identifiers with information collected using differential privacy).



https://www.apple.com/privacy/docs/Differential_Privacy_Overview.pdf

r/ai

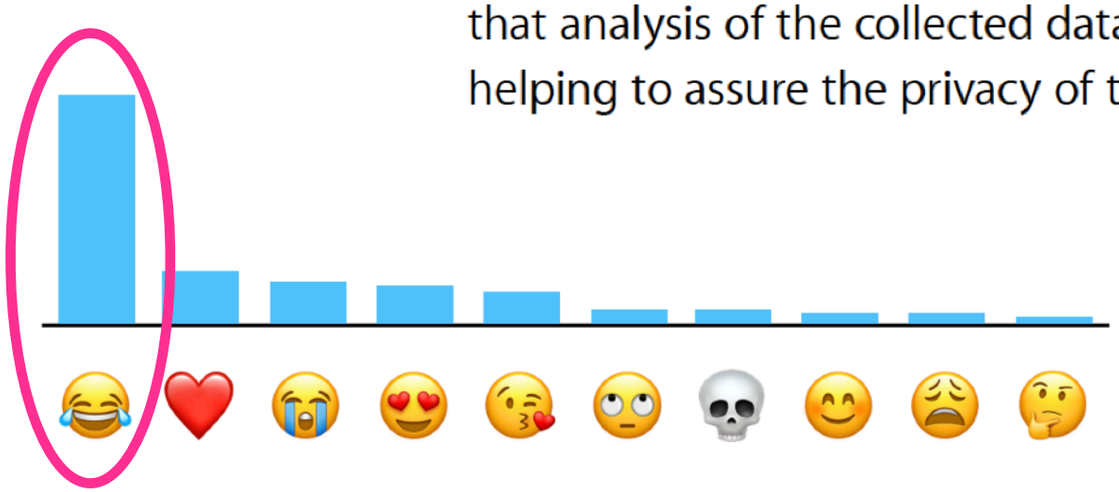
Example: What's your favorite emoji?

Count Mean Sketch

In our use of the Count Mean Sketch technique for differential privacy, the original information being processed for sharing with Apple is encoded using a series of mathematical functions known as *hash functions*, making it easy to represent data of varying sizes in a matrix of fixed size.

The data is encoded using variations of a SHA-256 hash followed by a privatization step and then written into the sketch matrix with its values initialized to zero.

The noise injection step works as follows: After encoding the input as a vector using a hash function, each coordinate of the vector is then flipped (written as an incorrect value) with a probability of $1/(1 + e^{\varepsilon/2})$, where ε is the privacy parameter. This assures that analysis of the collected data cannot distinguish actual values from flipped values, helping to assure the privacy of the shared information.



https://www.apple.com/privacy/docs/Differential_Privacy_Overview.pdf

r/ai

Transparency is important!

ANDY GREENBERG SECURITY 09.15.2017 09:28 AM

How One of Apple's Key Privacy Safeguards Falls Short

Apple has boasted of its use of a cutting-edge data science known as "differential privacy." Researchers say they're doing it wrong.

WIRED

Epsilon, Epsilon

"…[Researchers] examined how Apple's software injects random noise into personal information—ranging from emoji usage to your browsing history to HealthKit data to search queries—before your iPhone or MacBook upload that data to Apple's servers.

Ideally, that obfuscation helps protect your private data from any hacker or government agency that accesses Apple's databases, advertisers Apple might someday sell it to, or even Apple's own staff. But **differential privacy's effectiveness depends on a variable known as the "privacy loss parameter," or "epsilon,"** which determines just how much specificity a data collector is willing to sacrifice for the sake of protecting its users' secrets. By taking apart Apple's software to determine the epsilon the company chose, the researchers found that **MacOS uploads significantly more specific data than the typical differential privacy researcher might consider private.** iOS 10 uploads even more. And perhaps most troubling, according to the study's authors, is that **Apple keeps both its code and epsilon values secret**, allowing the company to potentially change those critical variables and erode their privacy protections with little oversight…."

https://www.wired.com/story/apple-differential-privacy-shortcomings/

r/ai

A closer look at differential privacy

A randomized algorithm ***M*** provides **ε-differential privacy** if, for all neighboring databases ***D₁*** and ***D₂***, and for any set of outputs ***S***:

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

lower ε = stronger privacy

- The state-of-the-art in privacy technology, first proposed in 2006
- Has precise mathematical properties, captures cumulative privacy loss over multiple uses of a particular dataset with the concept of a **privacy budget**
- Privacy guarantee encourages participation by respondents
- Robust against strong adversaries, with auxiliary information, including also **future auxiliary information**!
- Precise error bounds that can be made public

r/ai

A closer look at differential privacy

A randomized algorithm ***M*** provides **ε-differential privacy** if, for all neighboring databases ***D₁*** and ***D₂***, and for any set of outputs ***S***:

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

lower ε = stronger privacy

ε (epsilon) cannot be too small: think 1/10, not $1/2^{50}$

Differential privacy is a condition on the **algorithm M** (process privacy). Saying simply that “the output is safe” does not take into account how it was computed, and is insufficient.

r/ai