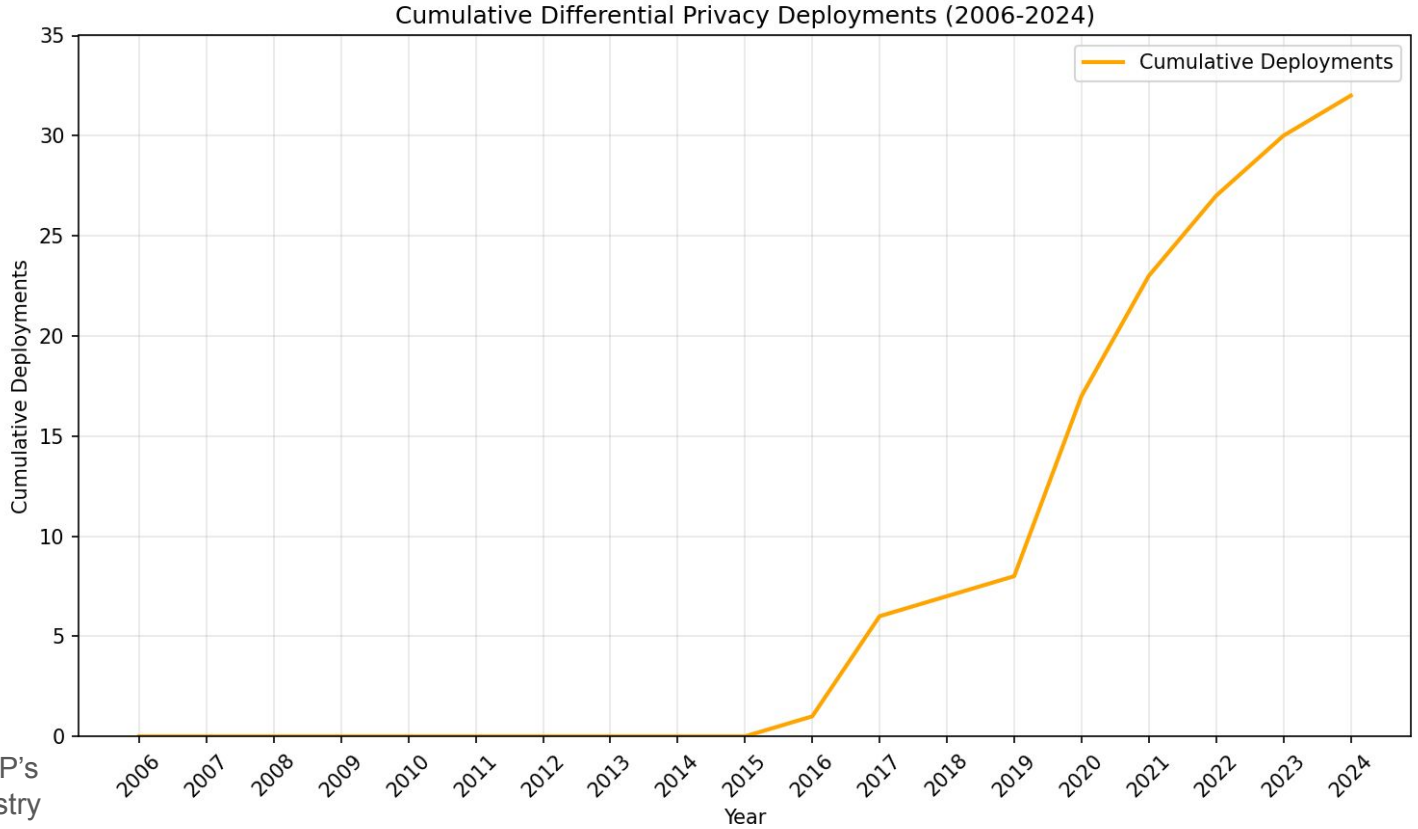
“Having Confidence In My Confidence Intervals”

How Data Users Engage with Privacy-Protected Wikipedia Data

Harold Triedman, Jayshree Sarathy, Priyanka Nanayakkara, Rachel Cummings, Gabriel Kaptchuk, Sean Kross, and Elissa Redmiles

9 February 2026 • PPPC

DP is increasingly being deployed in the real world!



Data from OpenDP's Deployment Registry

UNITED STATES DISTRICT COURT FOR THE MIDDLE DISTRICT OF ALABAMA EASTERN DIVISION

THE STATE OF ALABAMA; ROBERT ADERHOLT, Representative for Alabama's 4th Congressional District, in his official and individual capacities; WILLIAM GREEN; and CAMARAN WILLIAMS,

Plaintiffs,

v.

UNITED STATES DEPARTMENT OF COMMERCE; GINA RAIMONDO, in her official capacity as Secretary of Commerce; UNITED STATES BUREAU OF THE CENSUS, an agency within the United States Department of Commerce; and RON JARMIN, in his official capacity as Acting Director of the U.S. Census Bureau,

Defendants.

CIVIL ACTION NO. 3:21-cv-211-RAH

COMPLAINT FOR DECLARATORY AND INJUNCTIVE RELIEF

THREE-JUDGE COURT REQUESTED PURSUANT TO 28 U.S.C. § 2284

UNITED STATES DISTRICT COURT FOR THE MIDDLE DISTRICT OF ALABAMA EASTERN DIVISION

2021 MAR 10 P 12:38

THE STATE OF ALABAMA; ROBERT ADERHOLT, Representative for Alabama's 4th Congressional District, in his official and individual capacities; WILLIAM GREEN; and CAMARAN WILLIAMS,

Plaintiffs,

v.

UNITED STATES DEPARTMENT OF COMMERCE; GINA RAIMONDO, in her official capacity as Secretary of Commerce; UNITED STATES BUREAU OF THE CENSUS, an agency within the United States Department of Commerce; and RON JARMIN, in his official capacity as Acting Director of the U.S. Census Bureau,

Defendants.

CIVIL ACTION NO. 3:21-cv-211-RAH

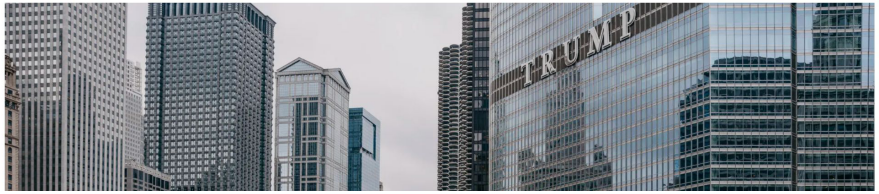
COMPLAINT FOR DECLARATORY AND INJUNCTIVE RELIEF

THREE-JUDGE COURT REQUESTED PURSUANT TO 28 U.S.C. § 2284

The 2020 Census Suggests That People Live Underwater. There's a Reason.

Technology advances forced the Census Bureau to use sweeping measures to ensure privacy for respondents. The ensuing debate goes to the heart of what a census is.

Share full article



UNITED STATES DISTRICT COURT FOR THE
MIDDLE DISTRICT OF ALABAMA
EASTERN DIVISION

2021 MAR 10 P 12:38

THE STATE OF ALABAMA; ROBERT ADERHOLT, Representative for Alabama's 4th Congressional District, in his official and individual capacities; WILLIAM GREEN;

The 2020 Census Suggests That People Live Underwater. There's a Reason.

Technology advances forced the Census Bureau to use sweeping measures to ensure privacy for respondents. The ensuing debate goes to the heart of what a census is.

Share full article

U.S. NEWS

Young Republicans challenge 2020 census results as part of wider GOP attack on head count

UNITED STATES DISTRICT COURT FOR THE
MIDDLE DISTRICT OF ALABAMA
EASTERN DIVISION

2021 MAR 10 P 12:38

THE STATE OF ALABAMA; ROBERT ADERHOLT, Representative for Alabama's 4th Congressional District, in his official and individual capacities; WILLIAM GREEN;

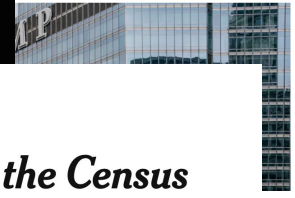
The 2020 Census Suggests That People Live Underwater. There's a Reason.

Technology advances forced the Census Bureau to use sweeping measures to ensure privacy for respondents. The ensuing debate goes to the heart of what a census is.

Share full article

U.S. NEWS

Young Republicans challeng 2020 census results as part of wi attack on head count

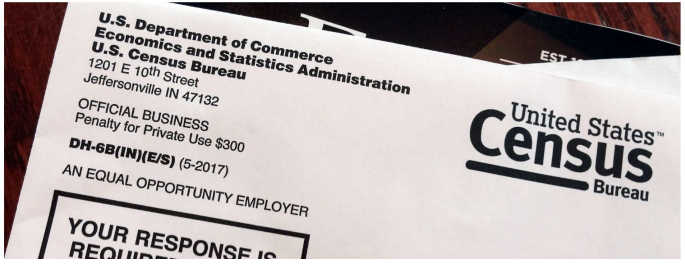


TheUpshot

To Reduce Privacy Risks, the Census Plans to Report Less Accurate Data

Guaranteeing people's confidentiality has become more of a challenge, but some scholars worry that the new system will impede research.

Share full article 37



UNITED STATES DISTRICT COURT FOR THE
MIDDLE DISTRICT OF ALABAMA
EASTERN DIVISION

2021 MAR 10 P 12:38

THE STATE OF ALABAMA; ROBERT ADERHOLT, Representative for Alabama's 4th Congressional District, in his official and individual capacities; WILLIAM GREEN;

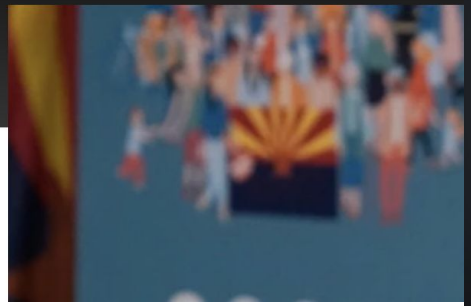
The 2020 Census Suggests That People Live Underwater. There's a Reason.

Technology advances forced the Census Bureau to use sweeping measures to ensure privacy for respondents. The ensuing debate goes to the heart of what a census is.

Share full article

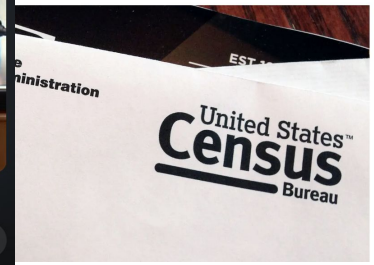
U.S. NEWS

Young Republicans challenge 2020 census resu… attack on he…



…Risks, the Census …ss Accurate Data

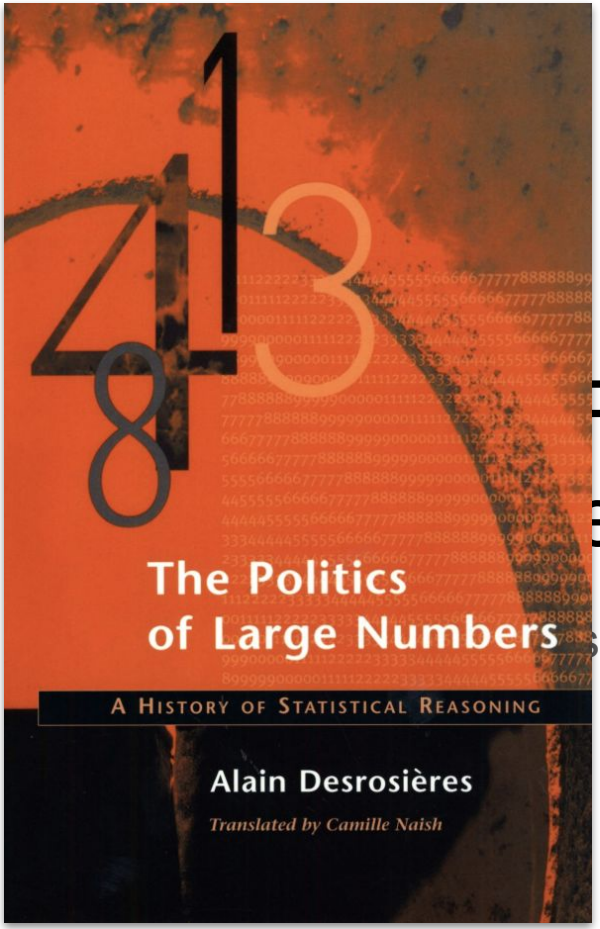
…tiality has become more of a …orry that the new system will



Deploying DP means intervening in a sociotechnical system…

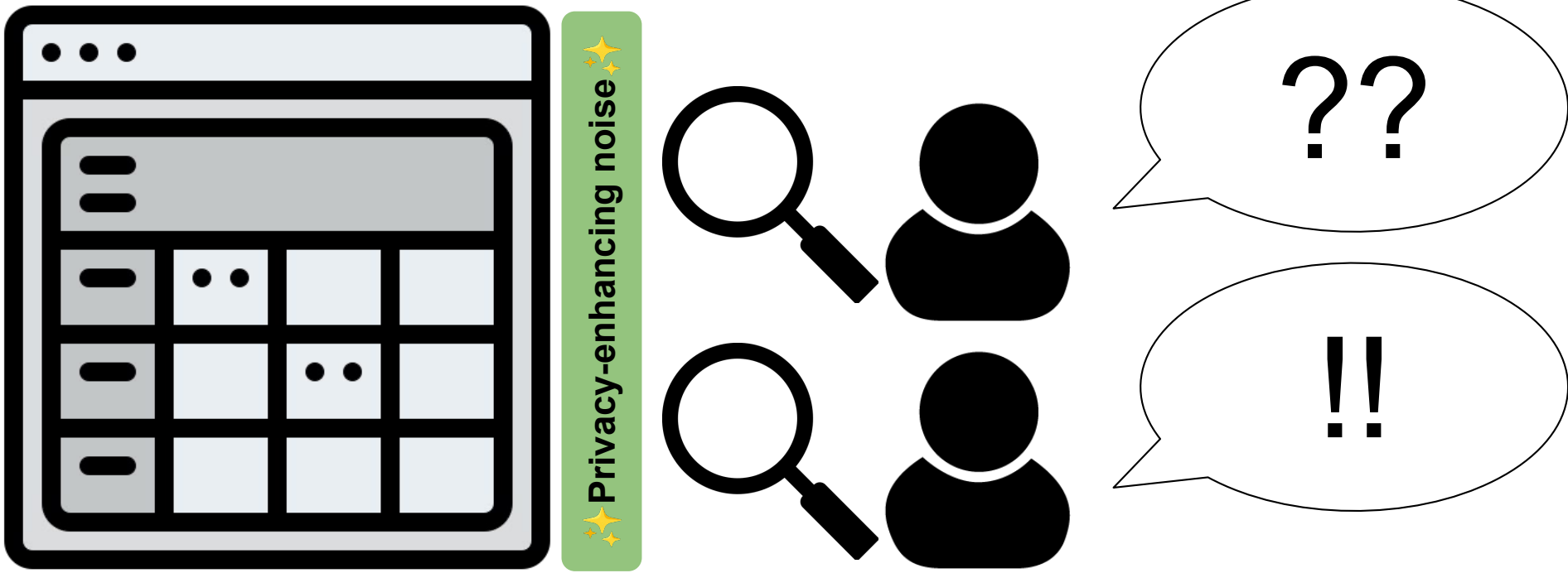
Deploying DP means intervening in a sociotechnical system…

…which changes system users' *statistical imaginaries*.



P means intervening in a
echnical system…

s system users' *statistical imaginaries*.


✨Privacy-enhancing noise✨
??
!!

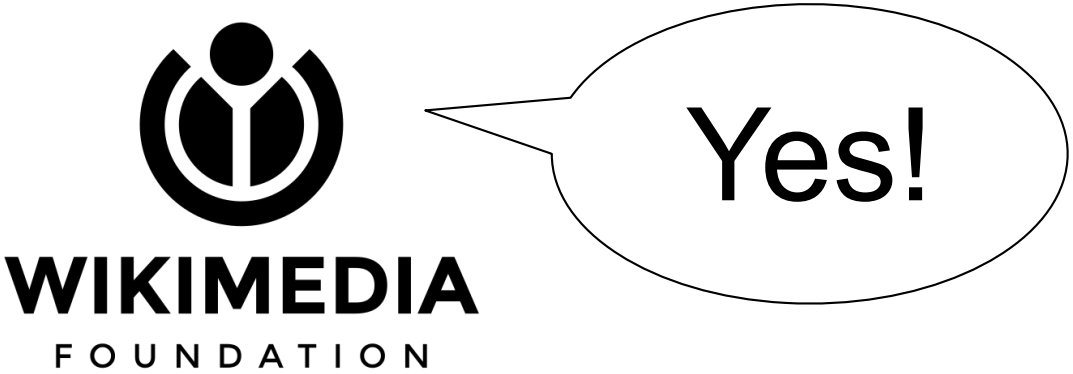
This is not just a question of politics…

This is not just a question of politics…

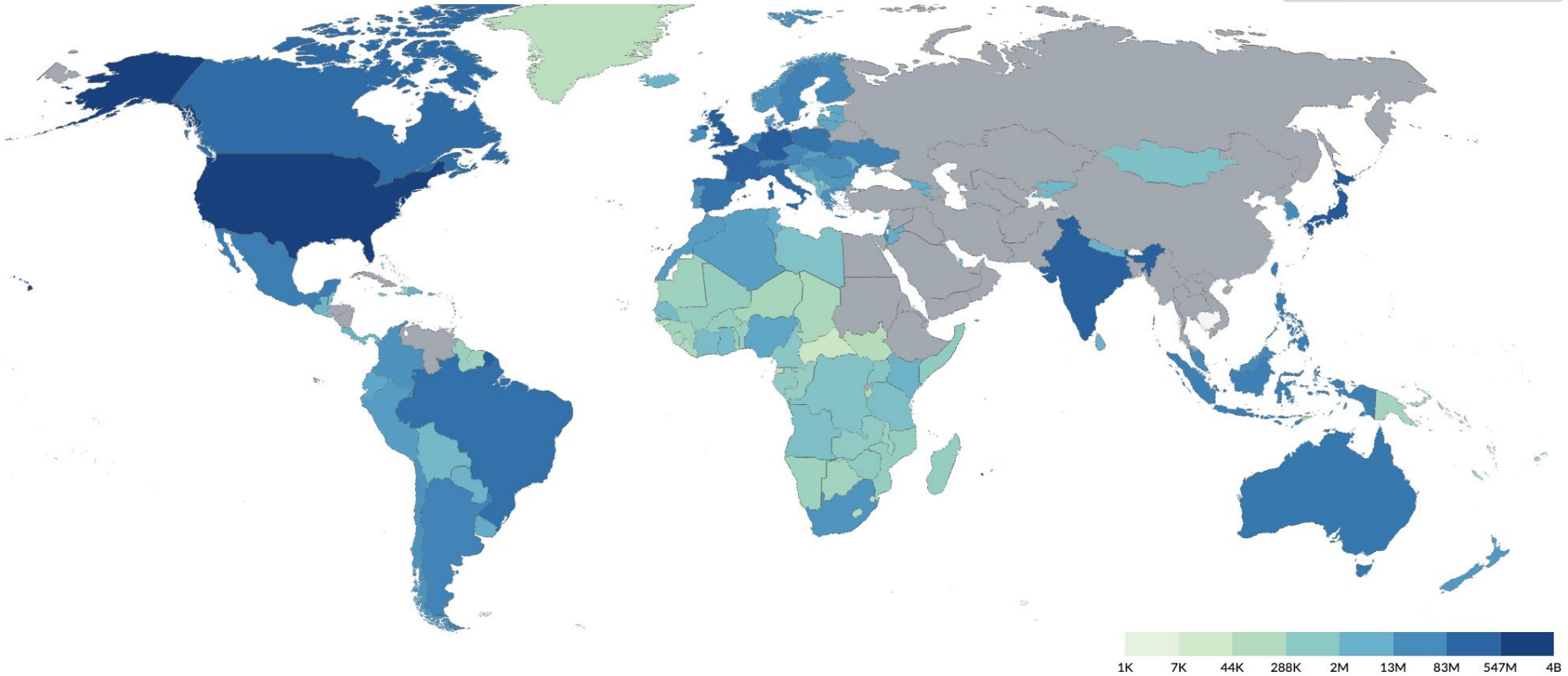
…it's also a question about how *changing the noise mechanism changes the data user experience*.

If we change the context, can we study the broader phenomenon?

If we change the context, can we study the broader phenomenon?



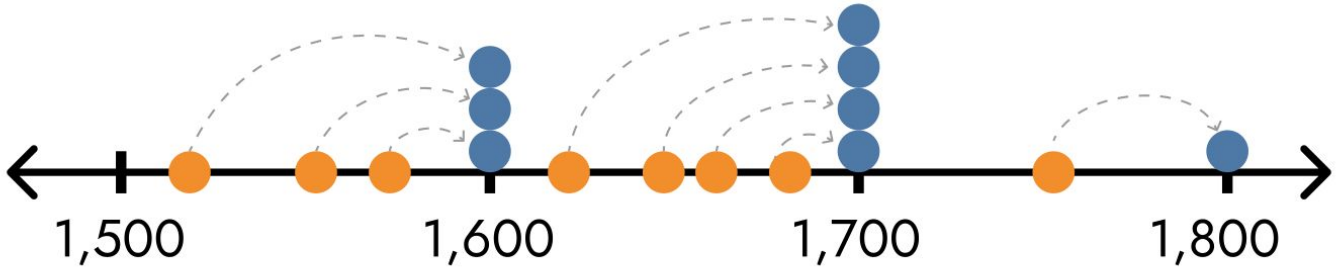
Page views by country



Graphic from stats.wikimedia.org

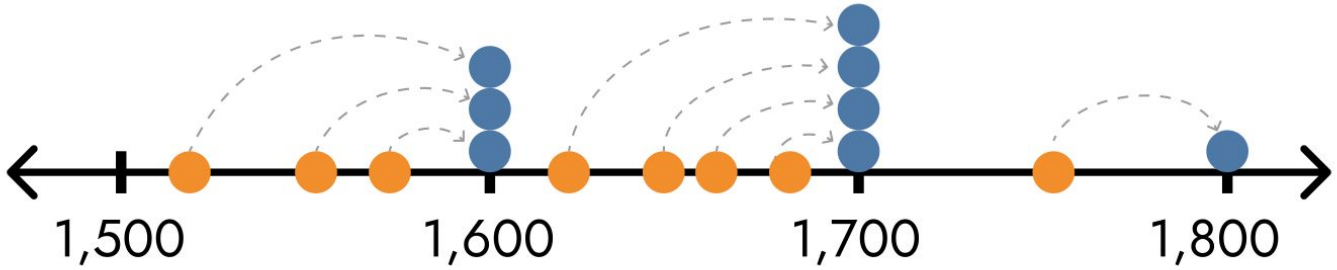
Pageviews by country

Rounded

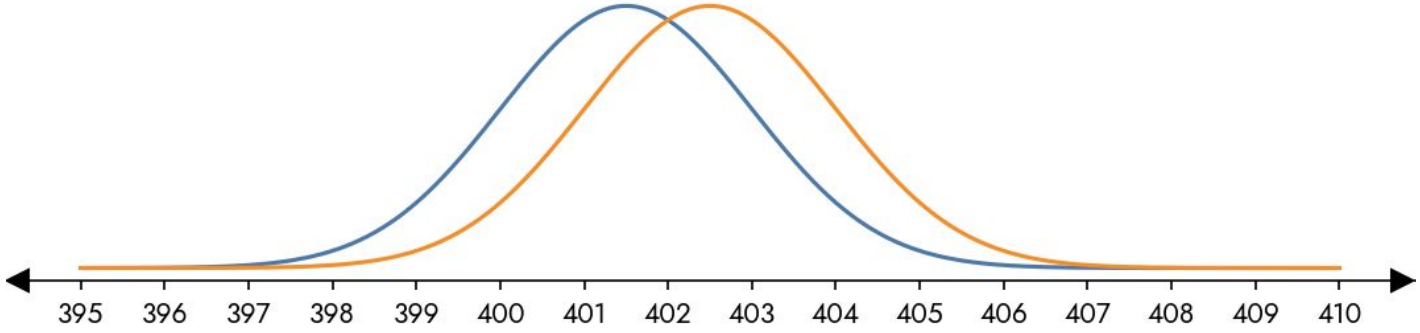


Pageviews by country

Rounded

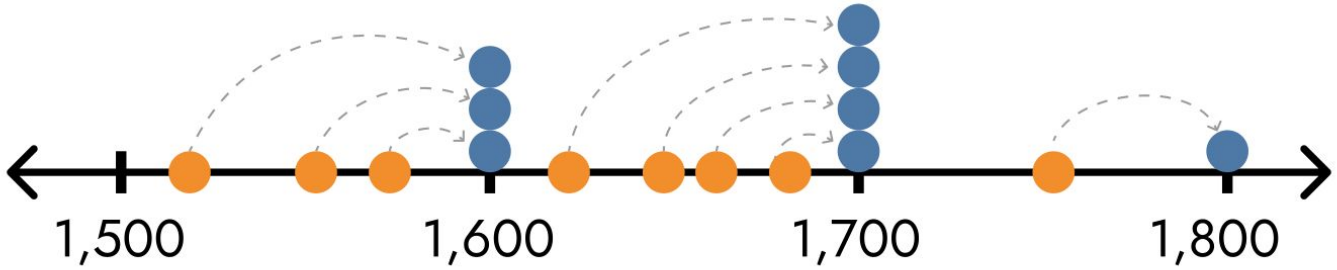


DP



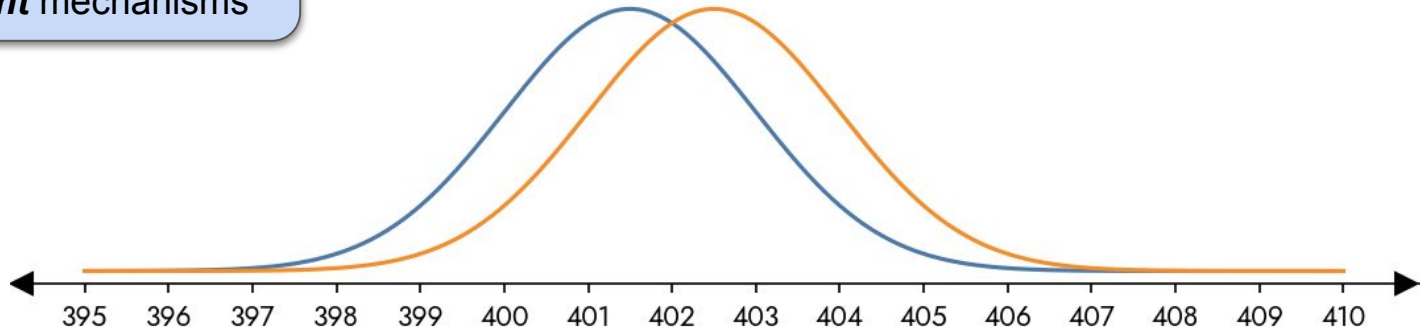
Pageviews by country

Rounded



Same underlying data
Same time range
Different mechanisms

DP



Our methodology

- With DP expert guidance, design ***data user-focused*** DP documentation
  - N = 5
  - Iteration on documentation design throughout

Our methodology

- With DP expert guidance, design ***data user-focused*** DP documentation
  - N = 5
  - Iteration on documentation design throughout
- Create a set of basic data analysis activities about pageview trends
  - Maxima, sums, means, equality checks

Our methodology

- With DP expert guidance, design ***data user-focused*** DP documentation
  - N = 5
  - Iteration on documentation design throughout
- Create a set of basic data analysis activities about pageview trends
  - Maxima, sums, means, equality checks
- Conduct structured interviews with non-DP expert data users on ***both noise mechanism datasets***
  - N = 15
  - Explicitly probe interviewee understandings of how the noise mechanism affects their perceptions of error, reliability, privacy guarantees, and more

Documentation

- Introduction
- Dataset schema
- How were the data pre-processed?
- How was privacy protection applied?
- How does privacy protection affect the accuracy of the published data?
- How can you use the published data?

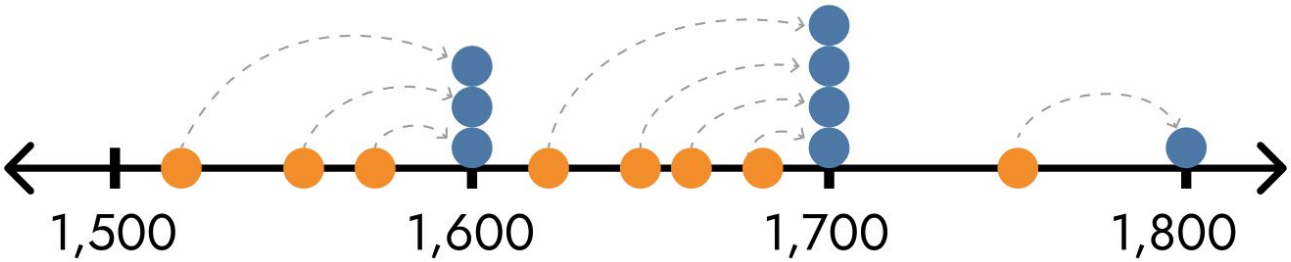
Study materials

Documentation: Privacy Mechanism

These counts are NOT included in the dataset for privacy reasons.

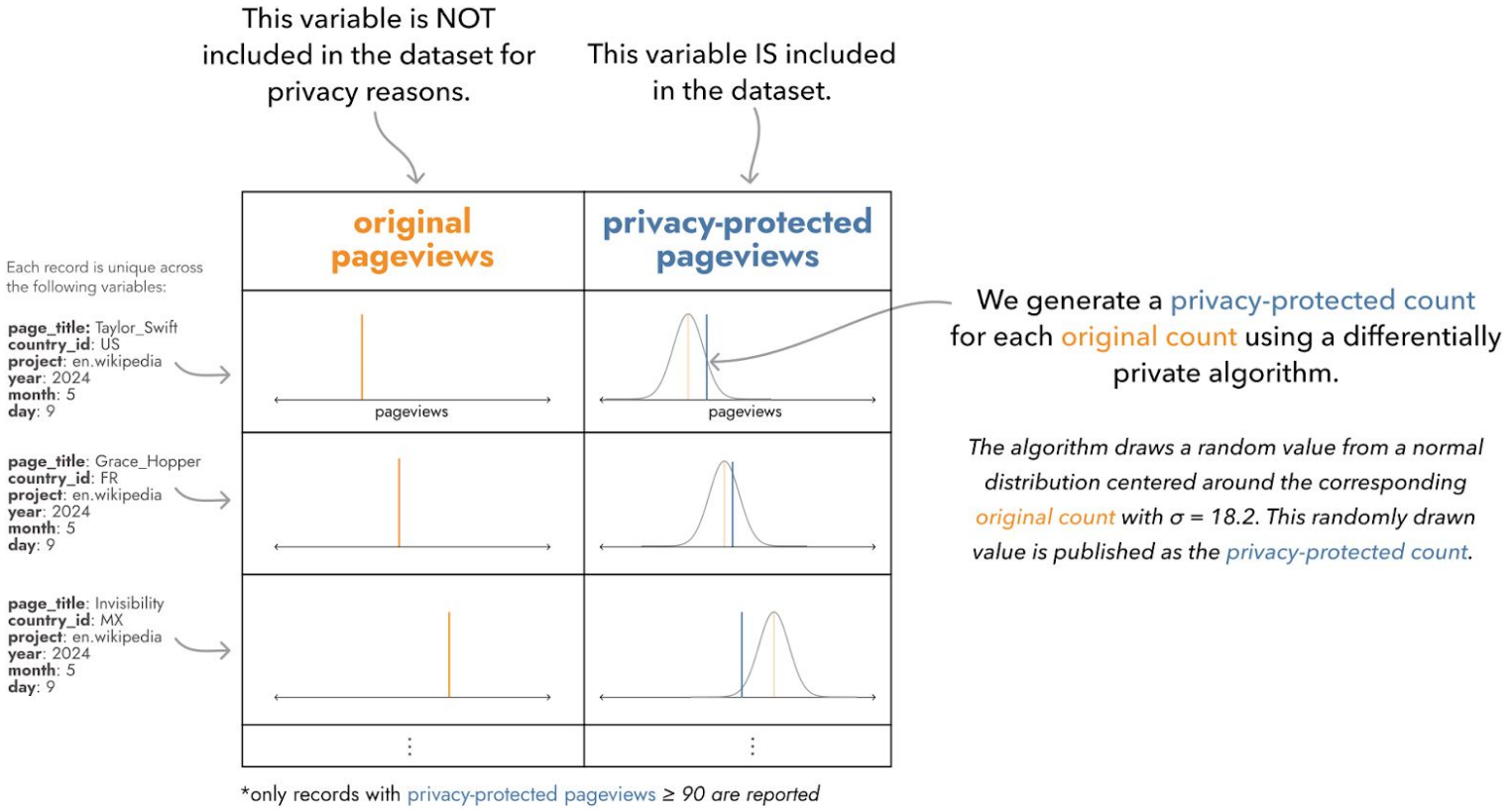
These counts ARE included in the dataset.

original pageviews rounded up to the next greatest multiple of 100 = **privacy-protected pageviews***

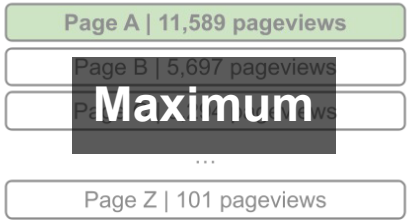


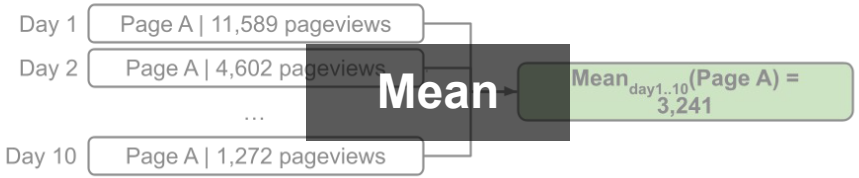
*only records with privacy-protected pageviews ≥ *1,100 are reported*

Documentation: Privacy Mechanism



Tasks





Study materials

Interview Questions

- For each task:
  - Provide a range that you would be 95% sure the *non-noised* value falls within.

Interview Questions

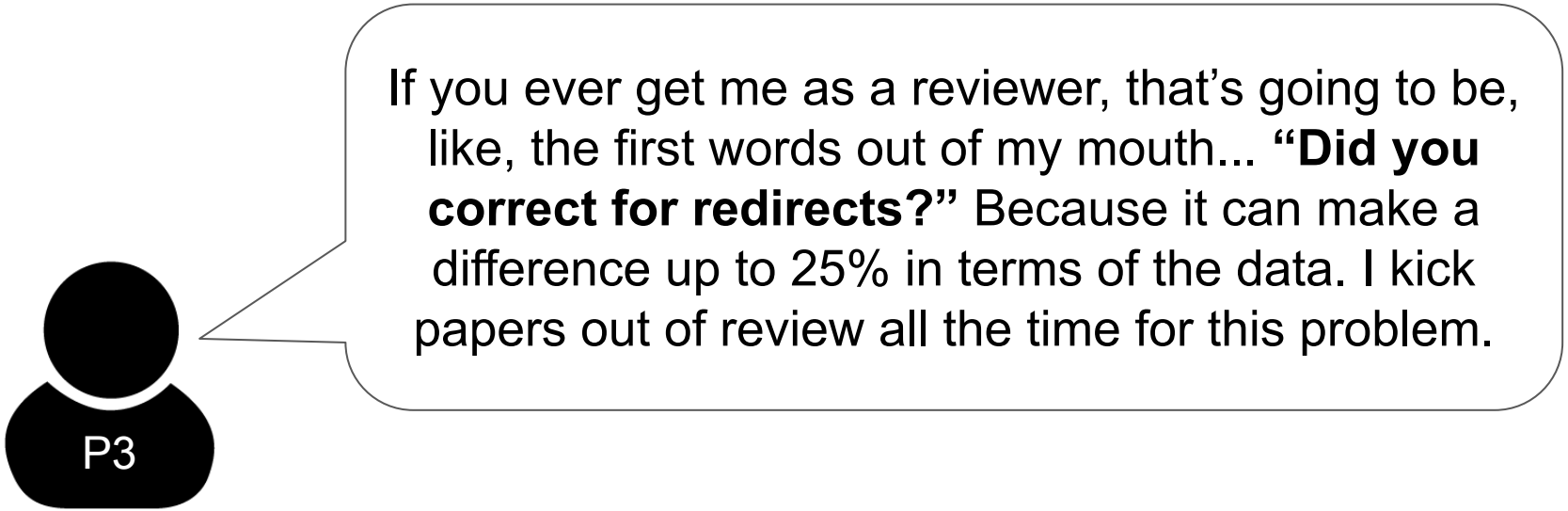
- For each task:
  - Provide a range that you would be 95% sure the *non-noised* value falls within.
- At the end:
  - Which analyses would you want to conduct with each dataset?
  - Does one dataset seem more privacy preserving to you?
  - Do you feel more comfortable analyzing one dataset?
  - Do you feel more comfortable communicating an analysis for one dataset?

Findings

Baseline understanding of (non-privacy) sources of noise = better reasoning about the impact of additional privacy noise

Findings

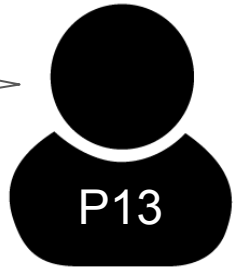
Baseline understanding of (non-privacy) sources of noise = better reasoning about the impact of additional privacy noise



Findings

Baseline understanding of (non-privacy) sources of noise = better reasoning about the impact of additional privacy noise

Internet data is super messy to begin with... **messifying it more for privacy** would not take away from the kind of work that people are already doing

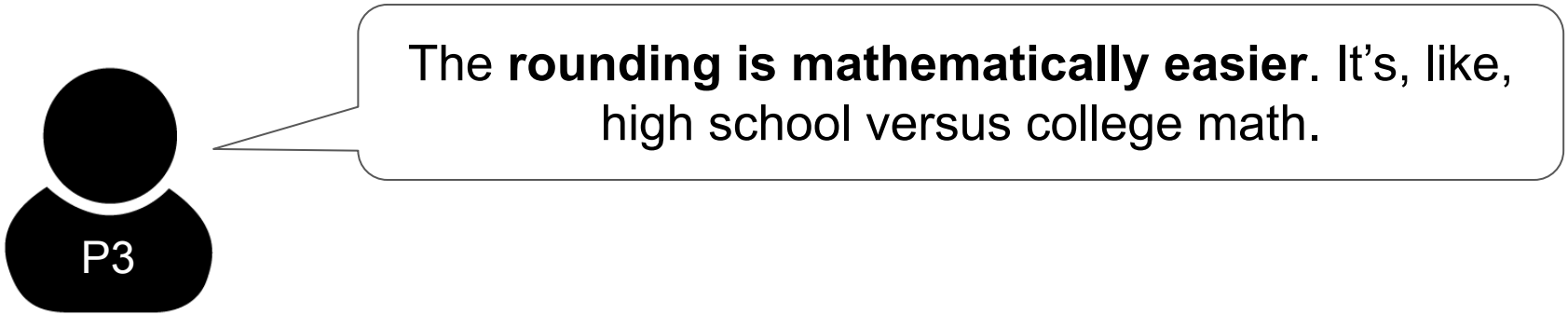


Findings

Rounding = easier to understand the process, DP = easier to reason with/about

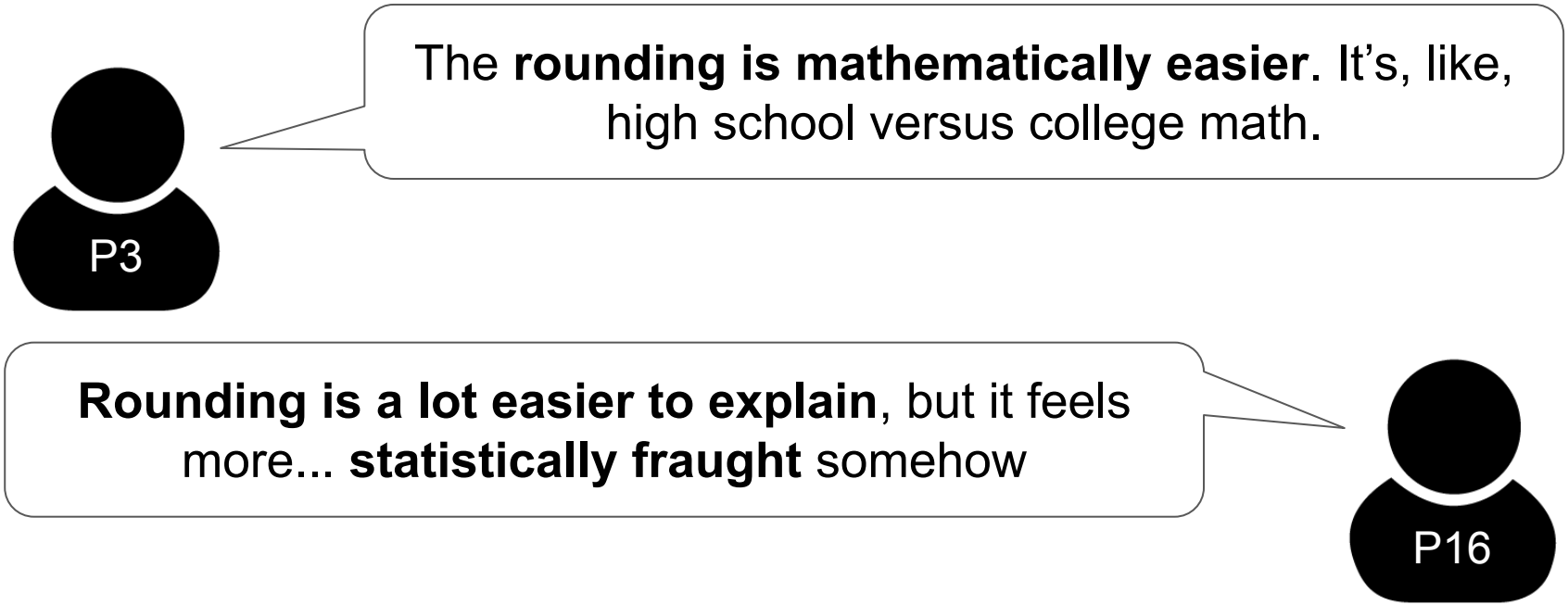
Findings

Rounding = easier to understand the process, DP = easier to reason with/about



Findings

Rounding = easier to understand the process, DP = easier to reason with/about

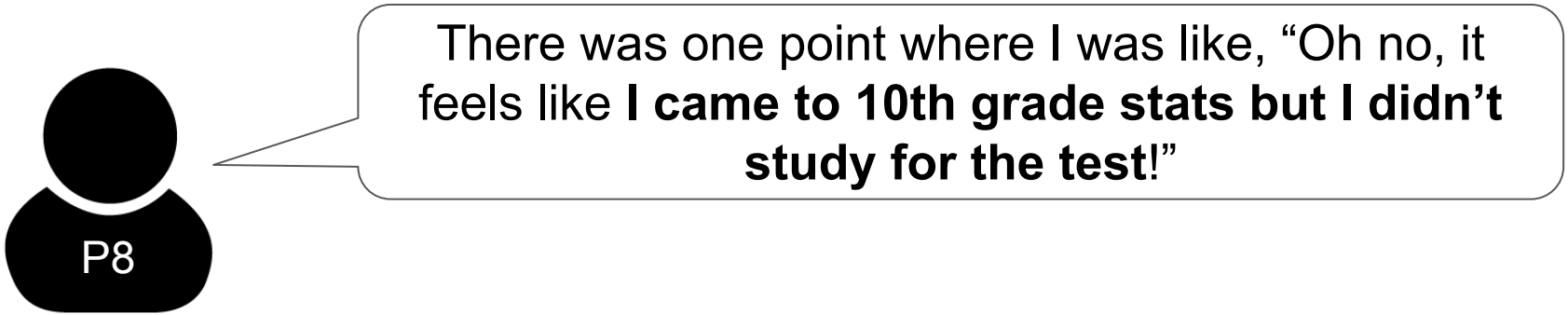


Findings

Multiple perturbed data points = struggle to compute confidence intervals

Findings

Multiple perturbed data points = struggle to compute confidence intervals



Findings

Multiple perturbed data points = struggle to compute confidence intervals

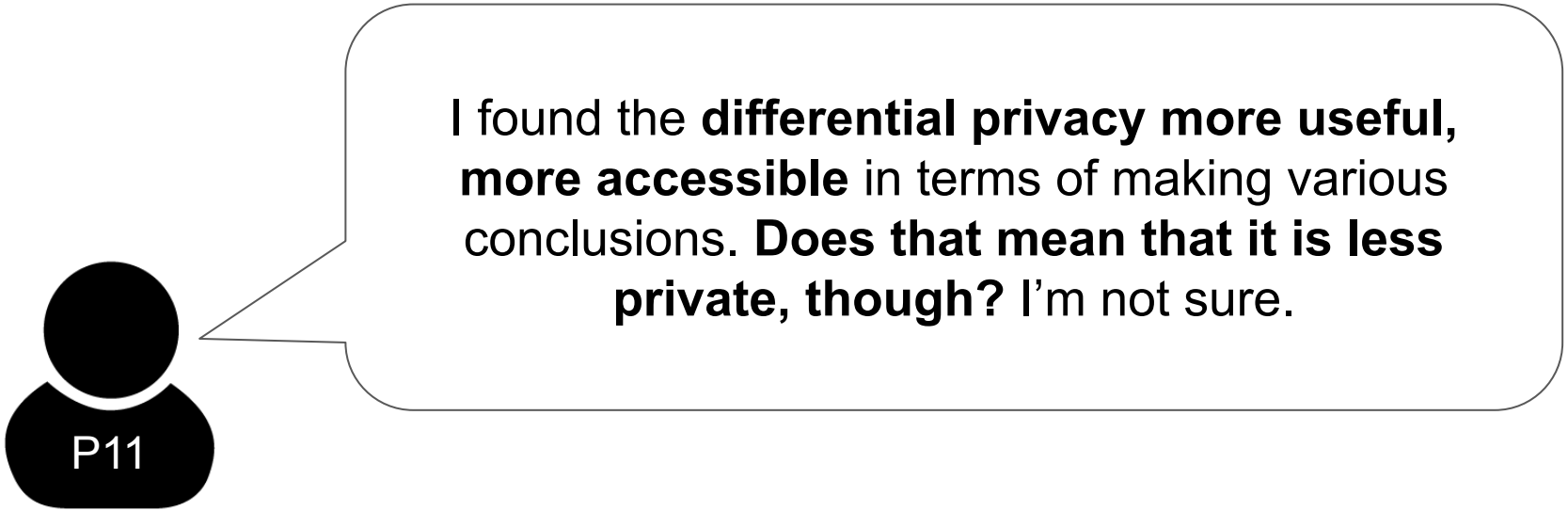
- Only **3 of 15** (20%) of participants do so successfully
  - And 2 of 3 had recently *taught* statistics courses

Findings

Mixed understandings of different methods’ effectiveness at protecting privacy

Findings

Mixed understandings of different methods' effectiveness at protecting privacy

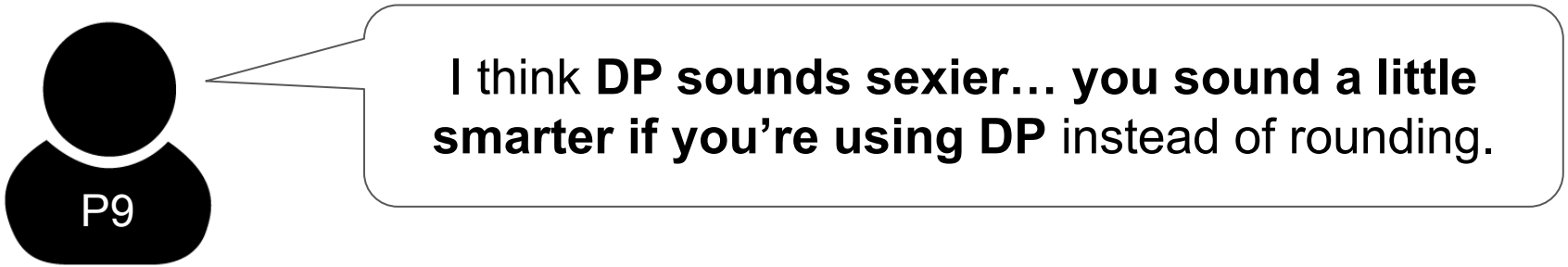


Findings

Preferences for which dataset to use when communicating their results are contextually-dependent

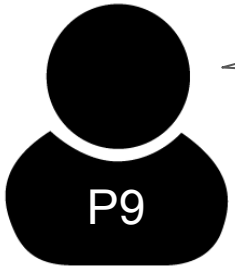
Findings

Preferences for which dataset to use when communicating their results are contextually-dependent



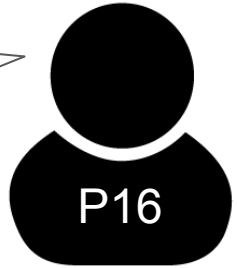
Findings

Preferences for which dataset to use when communicating their results are contextually-dependent



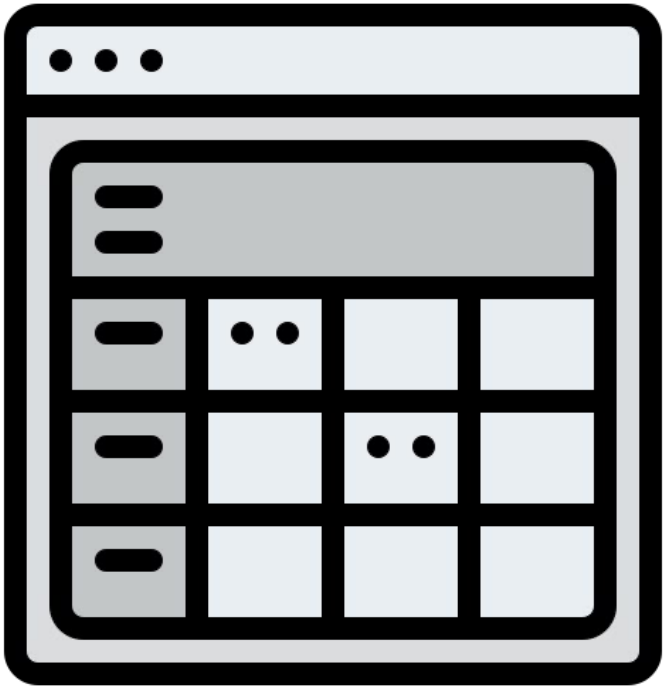
I think **DP sounds sexier… you sound a little smarter if you're using DP** instead of rounding.

What you lose with easy explainability you gain with the ability to cite, like, hundreds of papers that ostensibly say that DP is a reasonable thing to do.

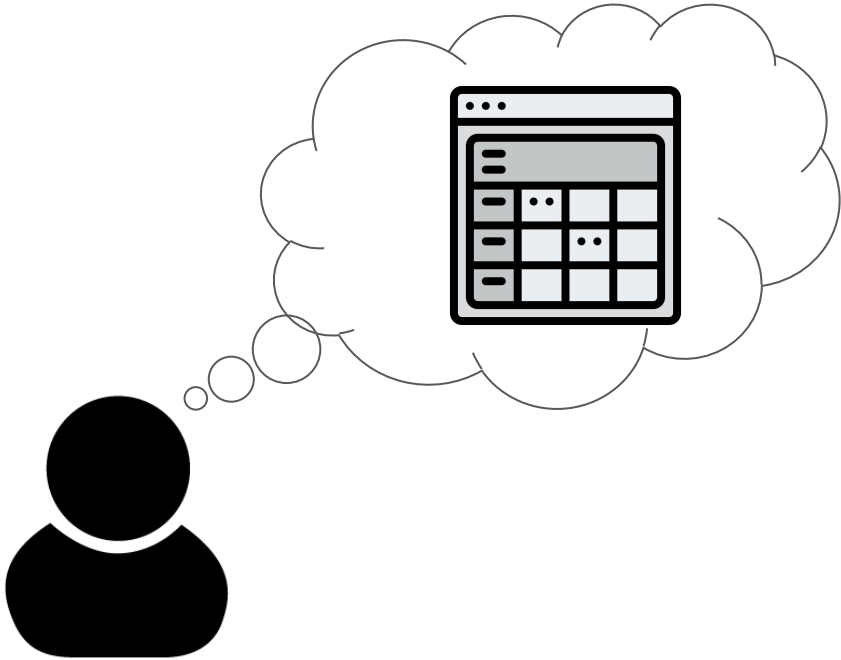


So… what does all of this mean?

Actual statistics

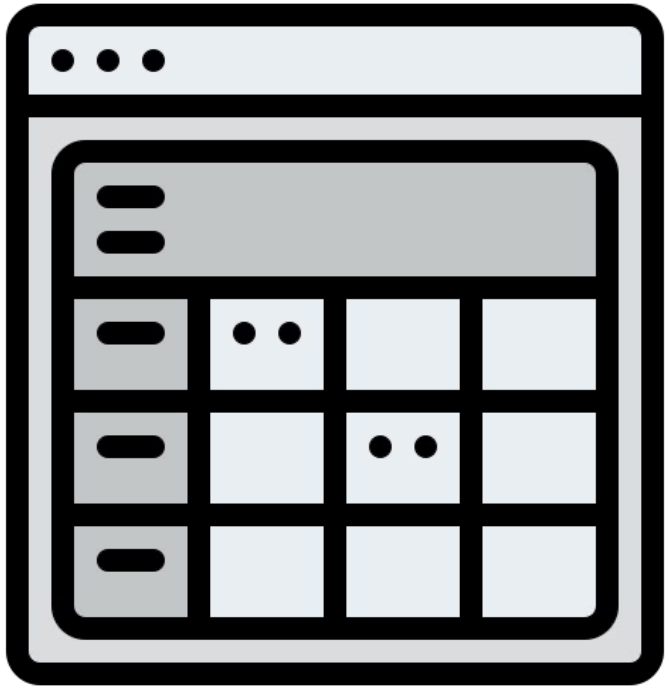


The data user's ~statistical imaginary~



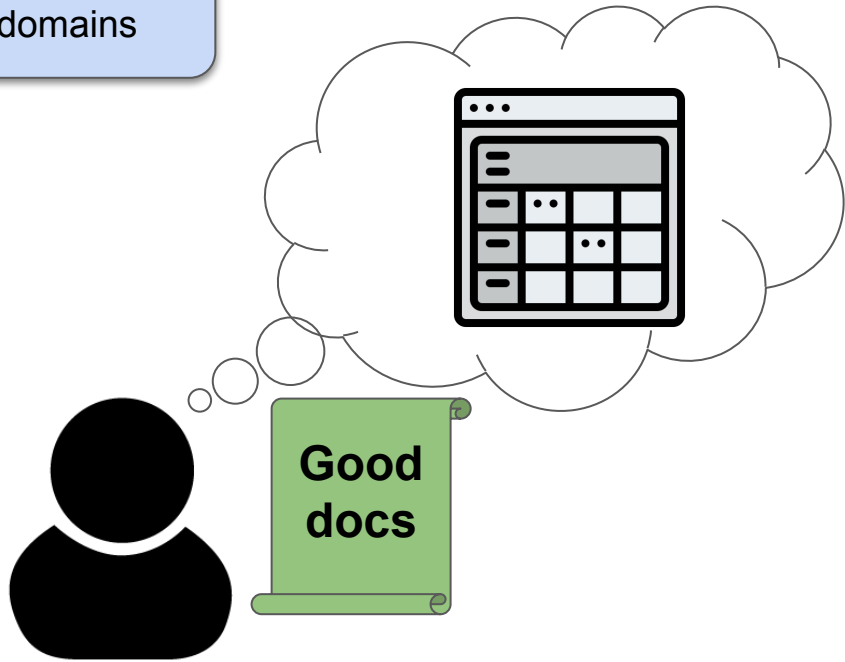
Actual statistics

✨Privacy-enhancing noise✨



The data user's ~statistical imaginary~

Effective interventions span both domains



What seems to work?

- Context-aware documentation
- Present privacy harms and accuracy metrics *together*
- Examples that clearly illustrate what analysts can and cannot deduce from noisy data
- Consider perceptions of “face privacy”

Thank you! Questions?

Paper (arxiv)

Study materials