DATA PRIVACY AND ANONYMIZATION IN R

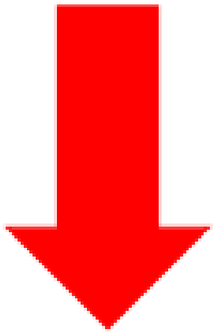
Differential Privacy

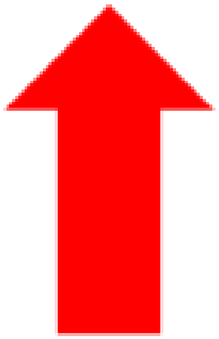
Claire McKay Bowen
Postdoctoral Researcher, Los Alamos National Laboratory

Why Differential Privacy

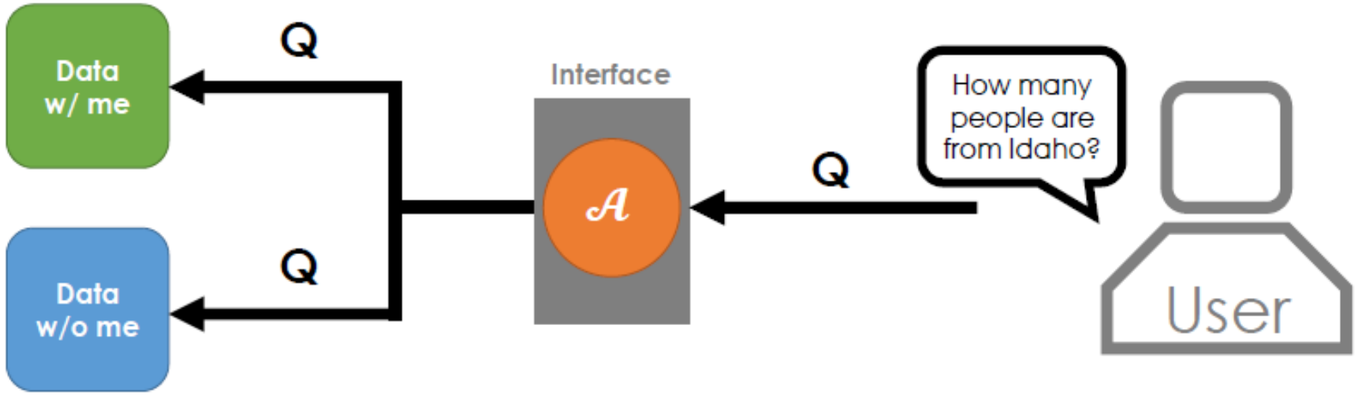
- Quantifies privacy loss via a privacy budget
- Assumes worst-case scenario; no assumptions about the data intruder

Epsilon, the Privacy Budget

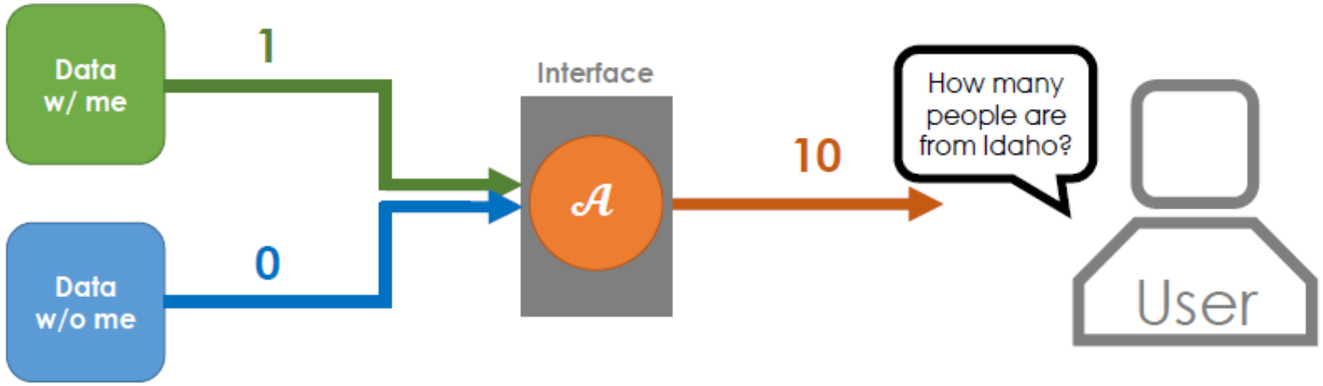
ϵ  Leak **less** information!

ϵ  Leak **more** information!

Differential Privacy: General Concept

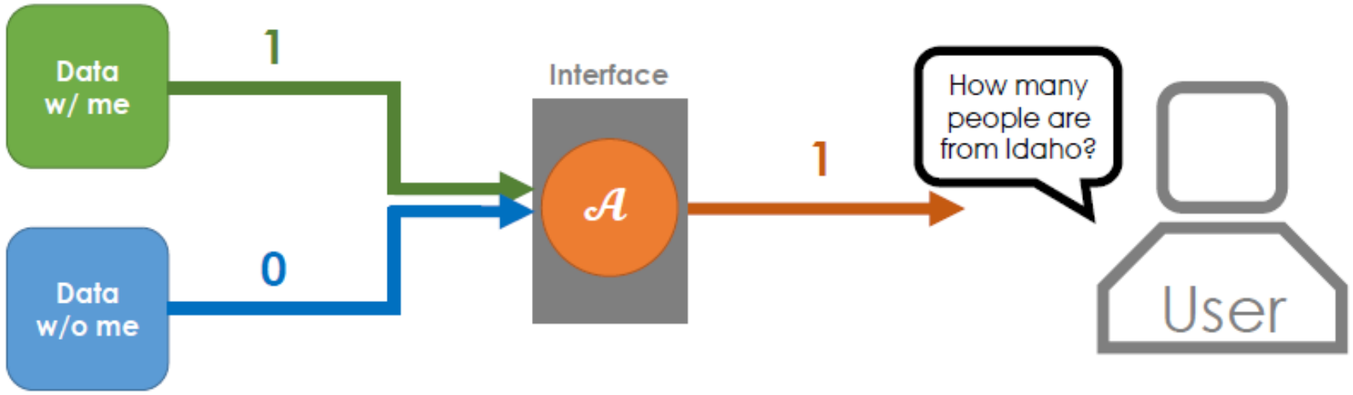


Differential Privacy: Small Privacy Budget



- Smaller privacy budget means less information or a noiser answer.

Differential Privacy: Large Privacy Budget



- Larger privacy budget means more information or a more accurate answer.

DATA PRIVACY AND ANONYMIZATION IN R

Let's practice!

DATA PRIVACY AND ANONYMIZATION IN R

Global Sensitivity

Claire McKay Bowen
Postdoctoral Researcher, Los Alamos National Laboratory

Global Sensitivity of Counting Queries

Global Sensitivity of Other Queries

- n is total number of observations
- a is the lower bound of the data
- b is the upper bound of the data

- **Counting:** 1
- **Proportion:** 1 / n
- **Mean:** (b - a) / n
- **Variance:** (b - a)^2 / n

Global Sensitivity and Noise

- **small** global sensitivity results in **less** noise
- **large** global sensitivity results in **more** noise

DATA PRIVACY AND ANONYMIZATION IN R

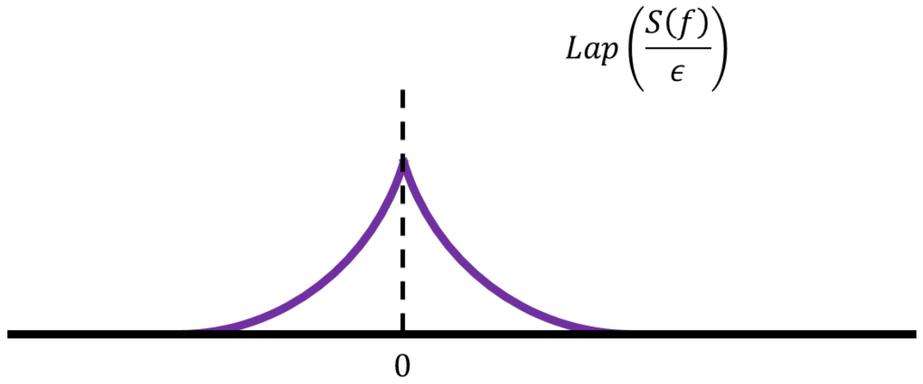
Let's practice!

DATA PRIVACY AND ANONYMIZATION IN R

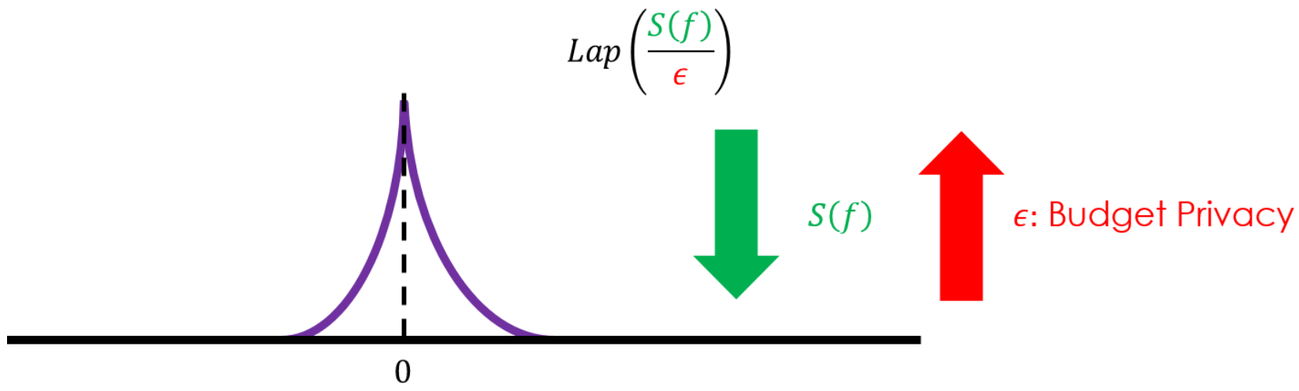
Laplace mechanism

Claire McKay Bowen
Postdoctoral Researcher, Los Alamos National Laboratory

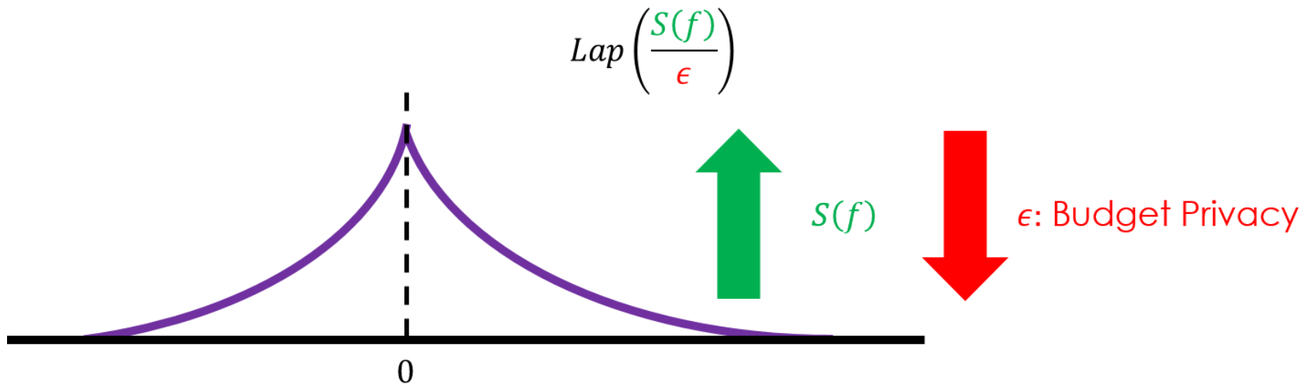
Laplace mechanism Part I



Laplace mechanism Part II



Laplace mechanism Part III



Coding the Laplace mechanism

```
> library(dplyr)
> fertility %>%
    summarise_at(vars(Child_Disease), sum)

# A tibble: 1 x 1
  Child_Disease
          <dbl>
1            87

> library(smoothmest)

# rdoublex(draws, mean, shaping)

> set.seed(42)
> rdoublex(1, 87, 1 / 10)
[1] 87.01983

> set.seed(42)
> rdoublex(1, 87, 1 / 0.1)
[1] 88.98337
```

DATA PRIVACY AND ANONYMIZATION IN R

Let's practice!