Differential privacy

DATA PRIVACY AND ANONYMIZATION IN R

Claire McKay Bowen

Postdoctoral Researcher, Los Alamos National Laboratory

Why differential privacy

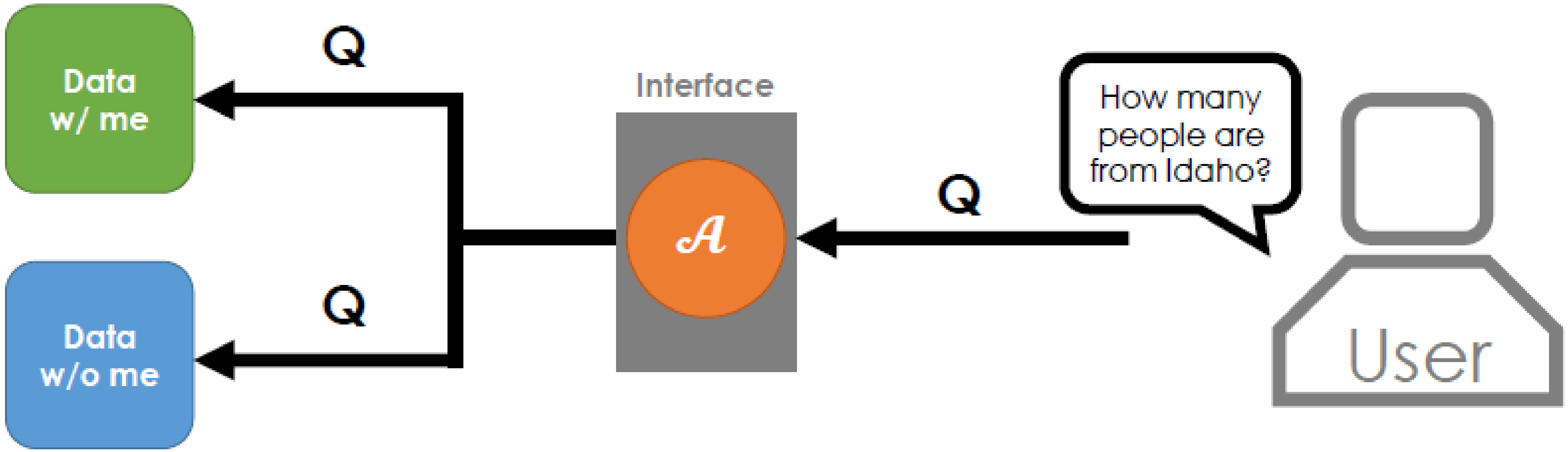
- Quantifies privacy loss via a privacy budget
- Assumes worst-case scenario; no assumptions about the data intruder

Epsilon, the privacy budget

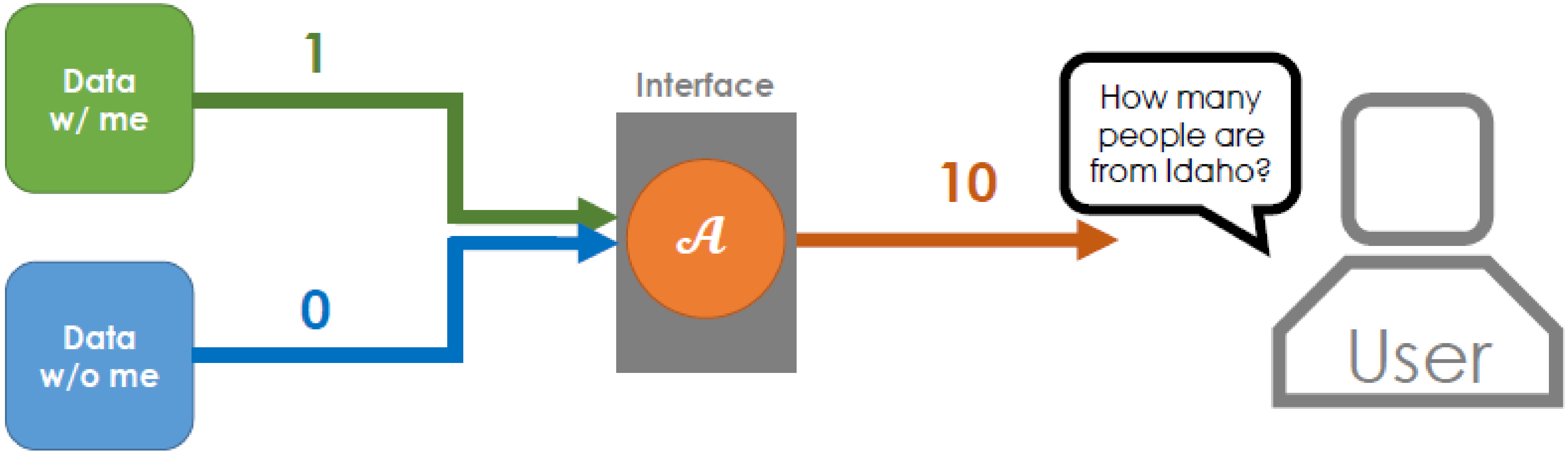
ϵ ↓ Leak **less** information!

ϵ ↑ Leak **more** information!

Differential privacy: general concept

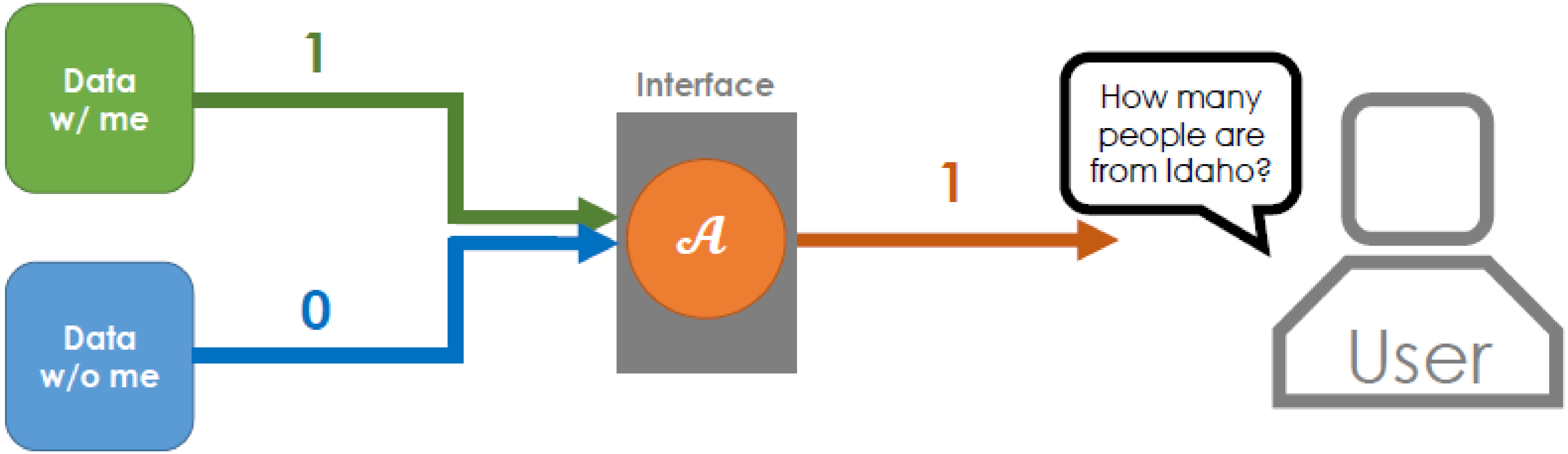


Differential privacy: small privacy budget



- 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.

Let's practice!

DATA PRIVACY AND ANONYMIZATION IN R

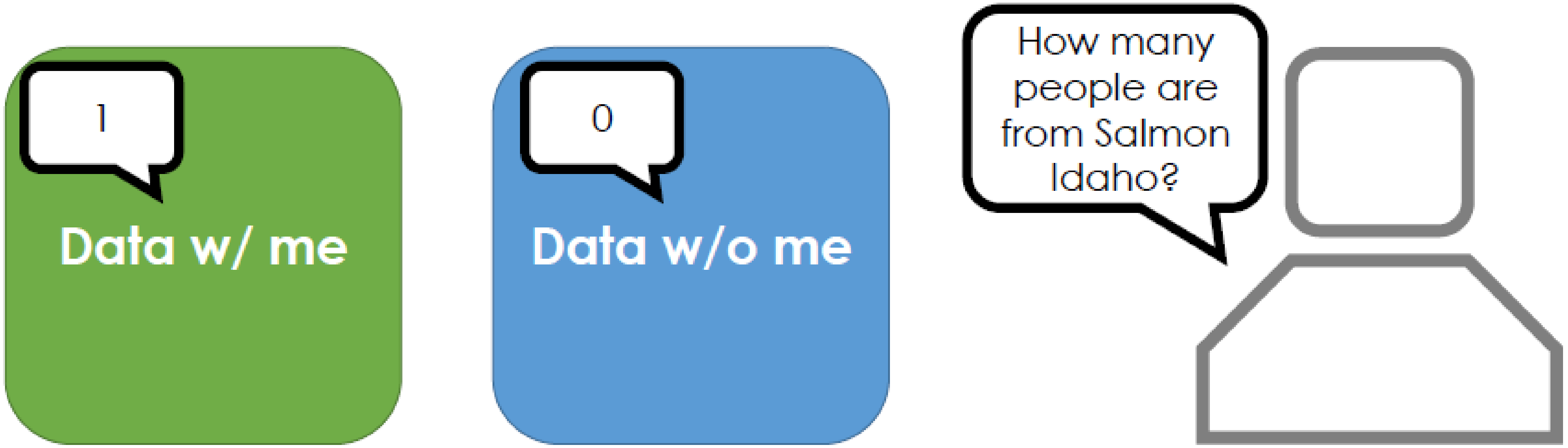
Global sensitivity

DATA PRIVACY AND ANONYMIZATION IN R

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

Let's practice!

DATA PRIVACY AND ANONYMIZATION IN R

Laplace mechanism

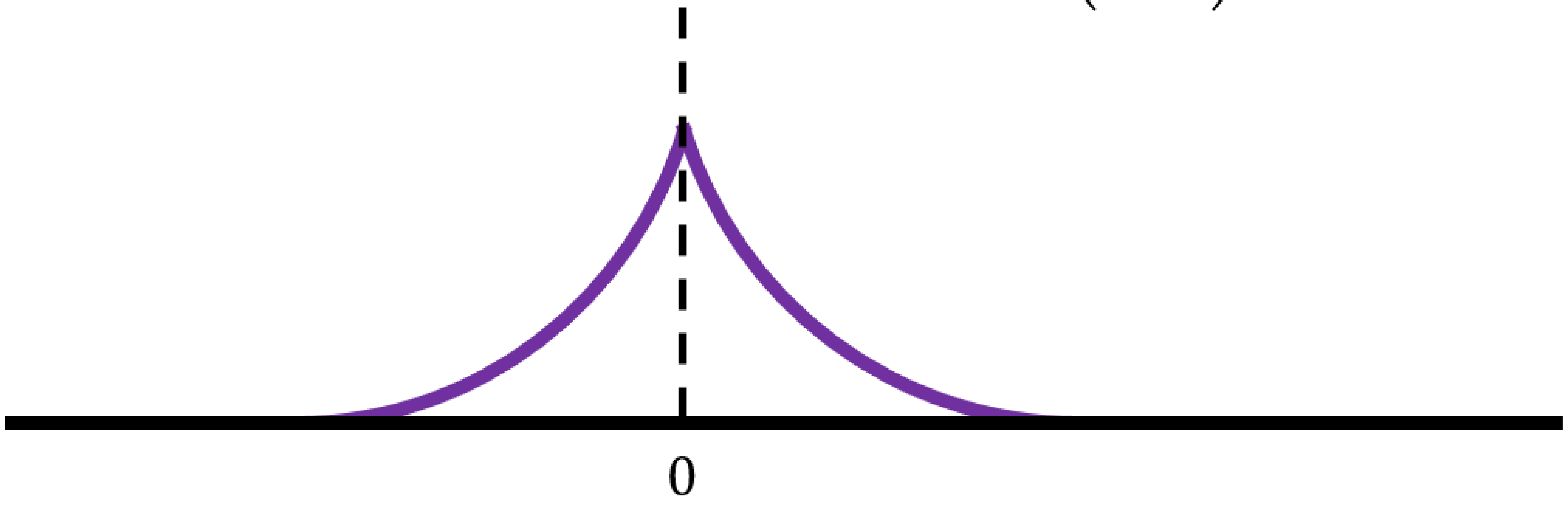
DATA PRIVACY AND ANONYMIZATION IN R

Claire McKay Bowen

Postdoctoral Researcher, Los Alamos National Laboratory

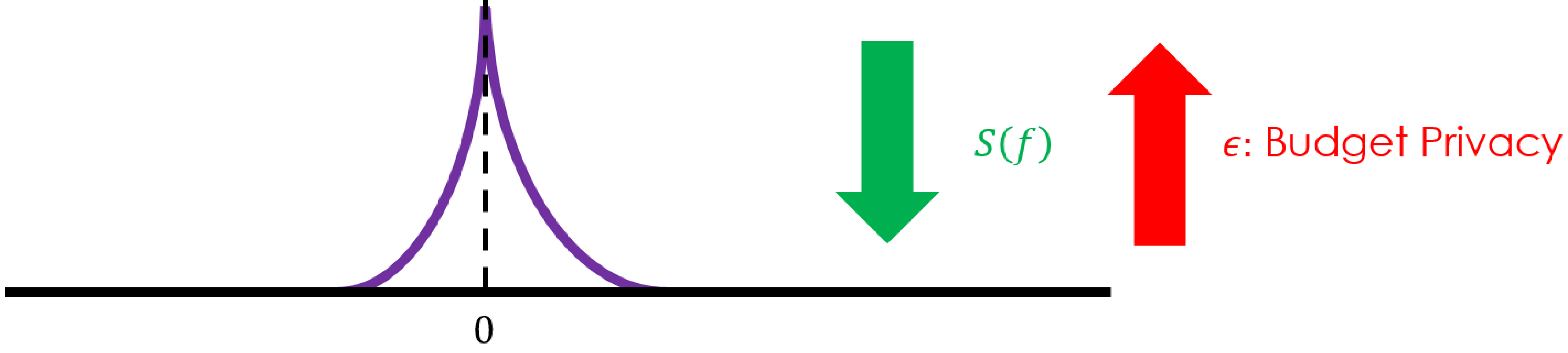
Laplace mechanism Part I

$$Lap\left(\frac{S(f)}{\epsilon}\right)$$



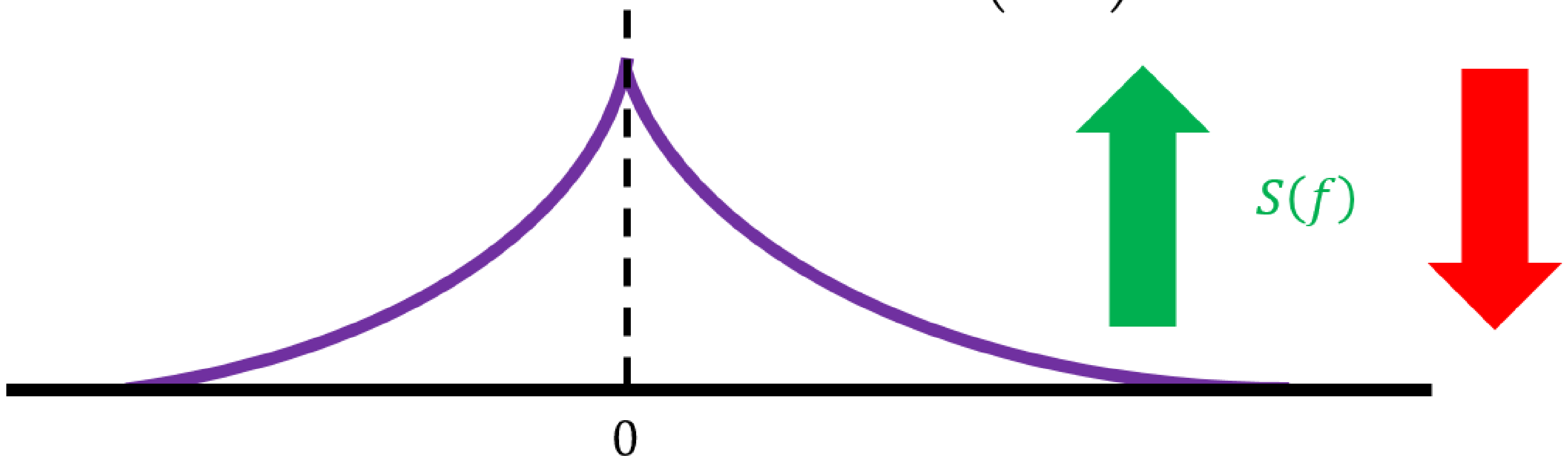
Laplace mechanism Part II

$$Lap\left(\frac{S(f)}{\epsilon}\right)$$



Laplace mechanism Part III

$$Lap\left(\frac{S(f)}{\epsilon}\right)$$



$S(f)$

ϵ: Budget Privacy

Coding the Laplace mechanism

```
library(dplyr)
fertility %>%
    summarize_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)
```

```
87.01983
```

```
set.seed(42)
rdoublex(1, 87, 1 / 0.1)
```

```
88.98337
```

Let's practice!

DATA PRIVACY AND ANONYMIZATION IN R