dataset → locked down (crypto)

Unwanted inferences "breaks" →

Algorithm →

Outputs:
- Modified dataset
- Statistics
- Predictive model
Outline:

- Consider 4 distinct concepts of privacy
  - First three flawed
  - #4; good (DP)
Privacy Concept #1: "Anonymization"

- **basic idea**: take a sensitive dataset, and modify it in order to protect private identity, while hopefully still being useful.

- **two operations**:
  - redaction/deleting entire column
  - coarsening: reducing resolution of cols.
One possible goal:

K-anonymization

Fatal flaw: reverse engineering via combining K-anon DBs + other info.
Concept #2: Aggregation

Example: Computing the average of a set of numbers.

E.g., \( x_1, x_2, x_3, \ldots, x_n \) representing the salaries of employees.

\[ a = \frac{1}{n} \sum_{i=1}^{n} x_i \]

\[ b = \frac{1}{n+1} \sum_{i=1}^{n+1} x_i \]

\[ (n+1)b - na = \frac{1}{n+1} \sum_{i=1}^{n+1} x_i \]

\[ a \pm \frac{E}{\sqrt{n}}, \quad \frac{1}{\sqrt{n}} \]