Poison and Cure: Non-Convex Optimization Techniques for Private Synthetic Data and Reconstruction Attacks

Michael Kearns

University of Pennsylvania and AWS AI/ML

“Differentially Private Query Release Through Adaptive Projection”
S. Aydore, W. Brown, M. Kearns, K. Kenthapadi, L. Melis, A. Roth, A. Siva
In *ICML* 2021

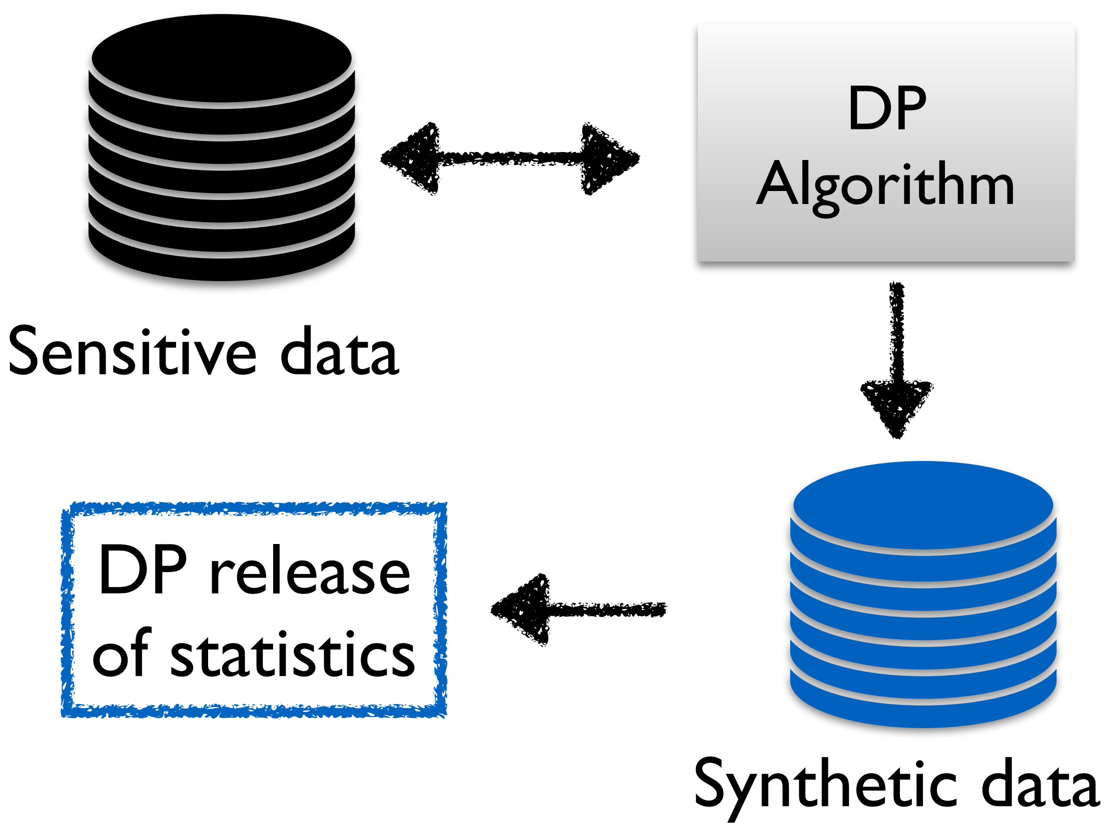
“Confidence-Ranked Reconstruction of Census Microdata from Published Statistics”
T. Dick, C. Dwork, M. Kearns, T. Liu, A. Roth, G. Vietri, Z. S. Wu
In *PNAS* 2023

1. Leveraging non-convex optimization to build *efficient* algorithms for *differentially private synthetic data generation*

2. The same algorithmic ideas enable *efficient* algorithms for large-scale *reconstruction attacks* (on Census data)

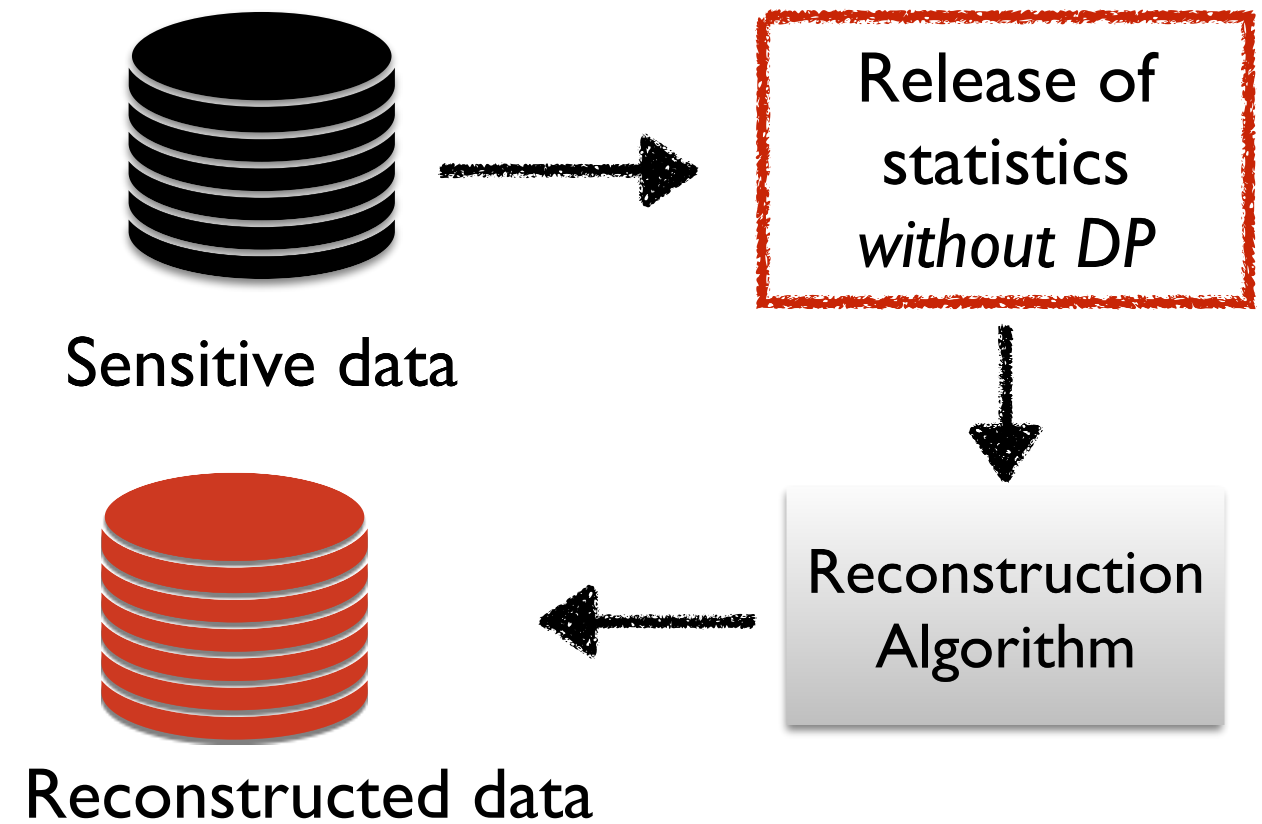
The Duality

Private Synthetic Data



Goal: release *approximation to* a large collection of statistics, downstream ML tasks

Reconstruction Attacks



Goal: reveal privacy risks of existing systems, auditing

Sensitive Data
(e.g., medical records)

Output Distribution
(e.g., noisy statistics)

"An algorithm is *differentially private* if changing a single record does not alter its output distribution by much."
[DN03, DMNS06]

Definition: A (randomized) algorithm A is (ε, δ)-differentially private if for all neighbors D, D' and every $S \subseteq$ Range(A)

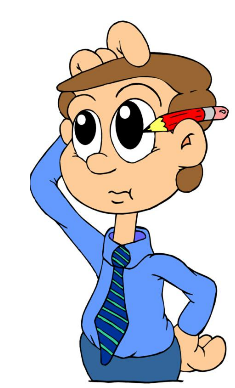
$$\Pr[A(D) \in S] \leq e^{\varepsilon} \Pr[A(D') \in S] + \delta$$

Differentially Private Synthetic Data

Sensitive data set
(e.g. medical records)

Synthetic data set
"*Fake*" data records that preserve important *statistical properties*

Allow arbitrary usage



Data Scientist

What statistical properties?

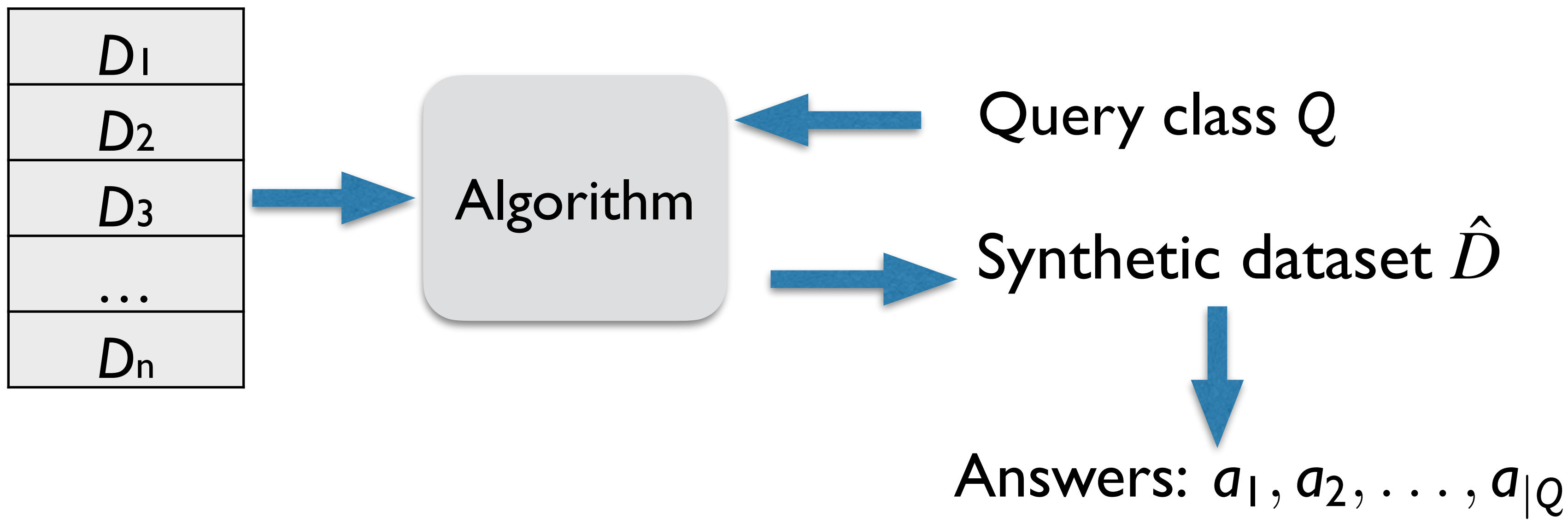
Moment Matching:
(aka Query Release)

	Smoke	Lung Cancer	Diabetes	Age	
patient_id1	1	1	1	35	$q(x) = 1$
patient_id2	1	0	0	40	$q(x) = 0$
patient_id3	1	1	0	43	$q(x) = 1$
patient_id4	0	0	1	21	$q(x) = 0$
					$q(D) = 1/2$

Example:
what is the fraction of people that satisfy some specified property q?

e.g. $q(x)$ = has "Smoke", "Lung Cancer" & "Age ≥ 30"
(3-way Marginals)

Moment Matching:
(aka Query Release)



α-*accurate* if

$|q(D) - a_q| \leq \alpha$ for every $q \in Q$

Algorithm 1 Relaxed Projection (RP)

Input: A vector of differentiable queries $q : \mathcal{X}^r \to \mathbb{R}^{m'}$, a vector of target answers $\hat{a} \in \mathbb{R}^{m'}$, and an initial dataset $D' \in (\mathcal{X}^r)^{n'}$.

Use any differentiable optimization technique (Stochastic Gradient Descent, Adam, etc.) to attempt to find:

$$D_S = \arg \min_{D' \in (\mathcal{X}^r)^{n'}} ||q(D') - \hat{a}||_2^2$$

Output D_S.

Algorithm 2 Relaxed Adaptive Projection (RAP)

Input: A dataset D, a collection of m statistical queries Q, a "queries per round" parameter $K \leq m$, a "number of iterations" parameter $T \leq m/K$, a synthetic dataset size n', and differential privacy parameters ϵ, δ.

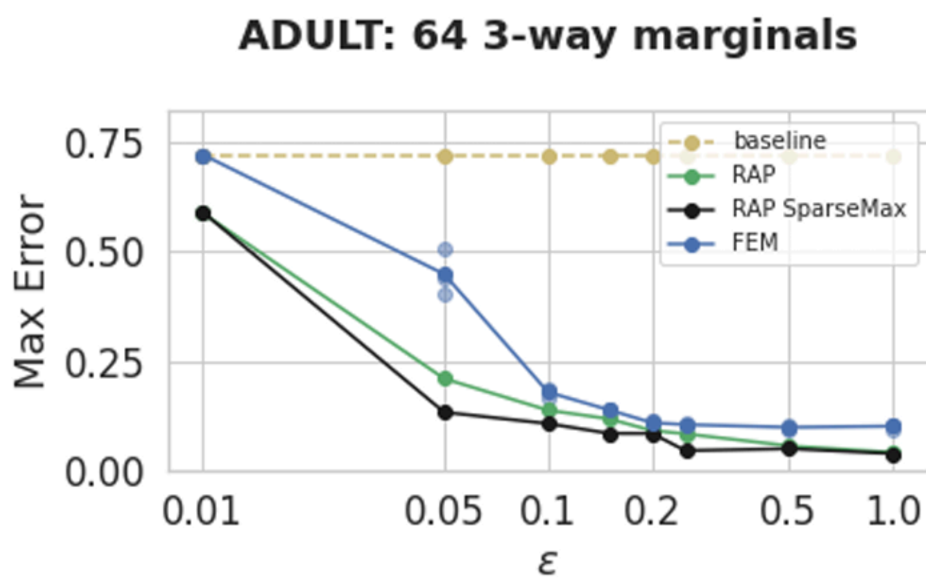
Let ρ be such that:

$$\epsilon = \rho + 2\sqrt{\rho \log(1/\delta)}$$

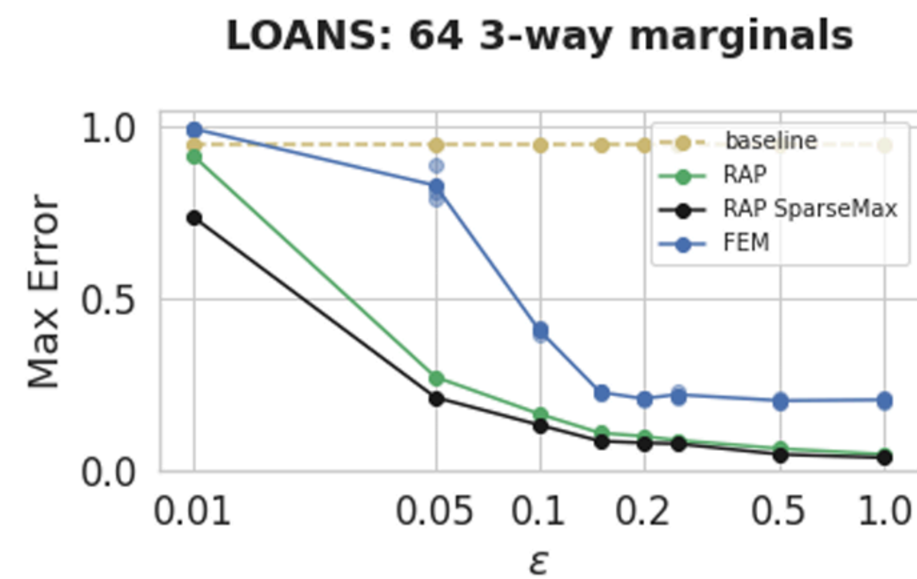
if $T = 1$ **then**
  for $i = 1$ to m **do**
    Let $\hat{a}_i = G(D, q_i, \rho/m)$.
  end for
  Randomly initialize $D' \in (\mathcal{X}^r)^{n'}$.
  Output $D' = RP(q, \hat{a}, D')$.
else
  Let $Q_S = \emptyset$ and $D'_0 \in (\mathcal{X}^r)^{n'}$ be an arbitrary initialization.
  for $t = 1$ to T **do**
    for $k = 1$ to K **do**
      Define $\hat{q}^{Q \setminus Q_S}(x) = (\hat{q}_i(x) : q_i \in Q \setminus Q_S)$ where $\hat{q}_i$ is an equivalent extended differentiable query for q_i.
      Let $q_i = RNM(D, \hat{q}^{Q \setminus Q_S}, \hat{q}^{Q \setminus Q_S}(D'_{t-1}), \frac{\rho}{2T \cdot K})$.
      Let $Q_S = Q_S \cup \{q_i\}$.
      Let $\hat{a}_i = G(D, q_i, \frac{\rho}{2T \cdot K})$.
    end for
    Define $q^{Q_S}(x) = (q_i(x) : q_i \in Q_S)$ and $\hat{a} = \{\hat{a}_i : q_i \in Q_S\}$ where $\hat{q}_i$ is an equivalent extended differentiable query for q_i. Let $D'_t = RP(q^{Q_S}, \hat{a}, D'_{t-1})$.
  end for
  Output D'_T.
end if

Dataset	Records	Features	Transformed Binary Features
ADULT	48842	15	588
LOANS	42535	48	4427

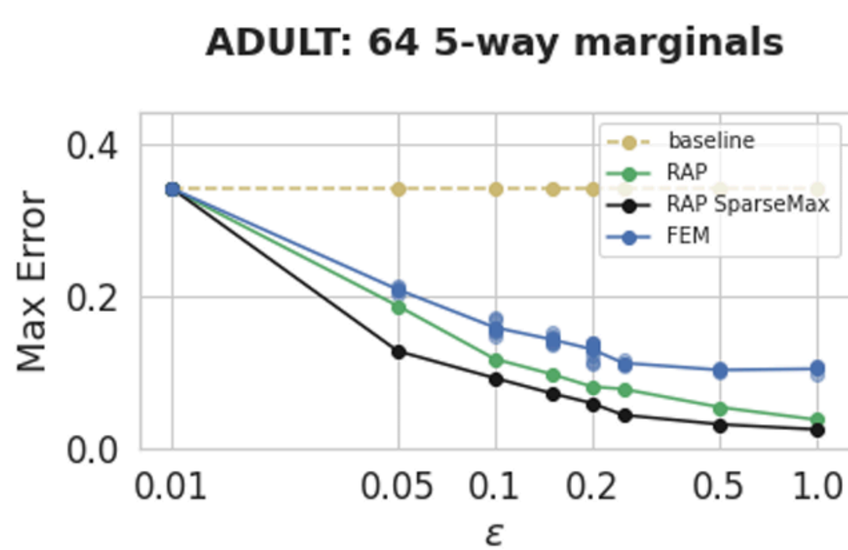
Table 1: Datasets. Each dataset starts with the given number of original (categorical and real valued) features. After our transformation, it is encoded as a dataset with a larger number of binary features.



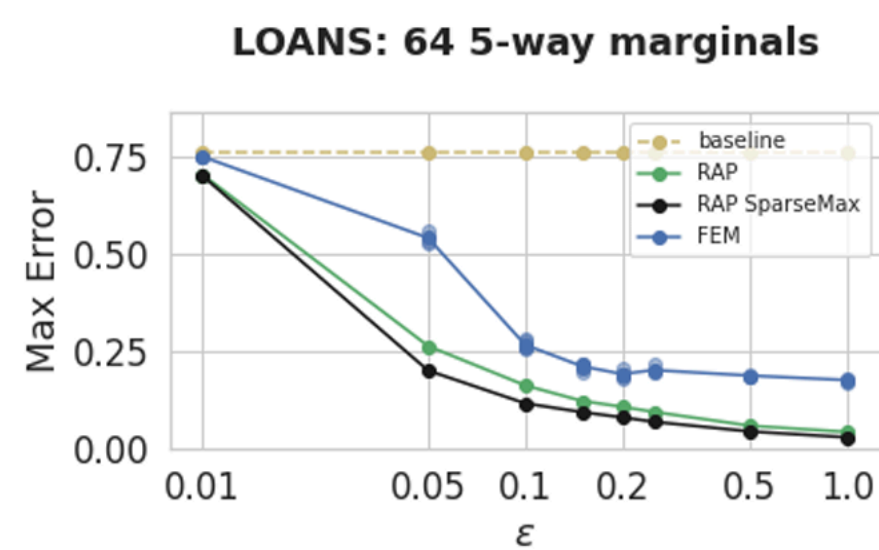
(a) ADULT dataset on 3-way marginals



(b) LOANS dataset on 3-way marginals



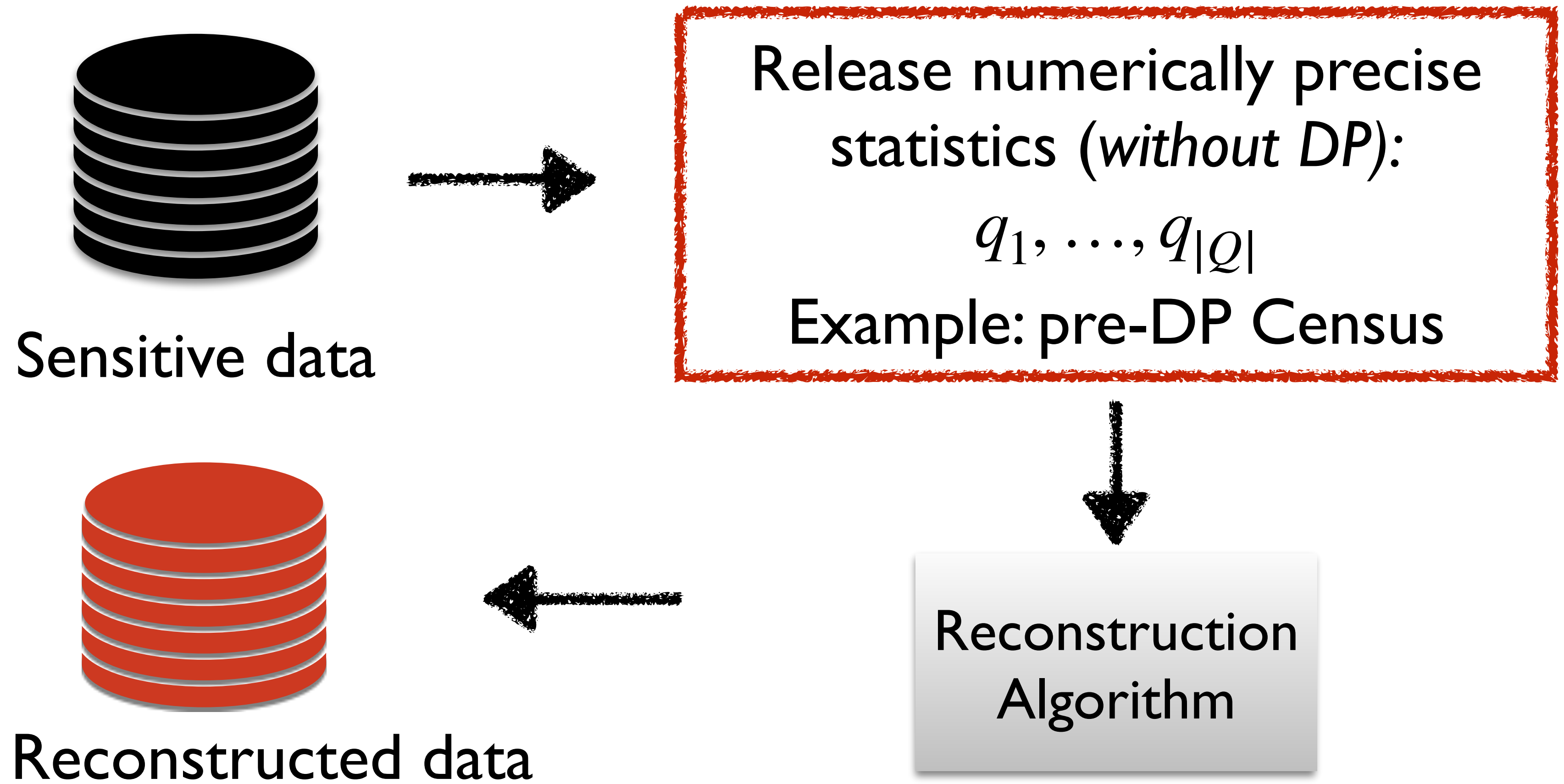
(c) ADULT dataset on 5-way marginals



(d) LOANS dataset on 5-way marginals

Reconstruction Attacks

[DN03]



Empirical attacks:

- Census Bureau's attack on 2010 decennial census
  - Leveraged powerful integer program solvers
- Aircloak Challenge [CN18, JSD20]

Reconstruction as Projection

Given *answers* $a = (a_1, \ldots, a_m)$ to queries $q = (q_1, \ldots, q_m)$
Reconstruct a dataset $\hat{D}$ by minimizing

$$\|q(\hat{D}) - a\|^2 = \sum_j (q_j(\hat{D}) - a_j)^2$$

Leverage the *computational efficiency* and *randomization* of synthetic data methods

RAP-Rank: Confidence-Ranked Reconstruction

Use a *randomized, non-private* synthetic data method to sample solutions to the projection problem

Answers $a_1, \ldots, a_m$ to queries $q_1, \ldots, q_m$ evaluated on dataset D

Relaxed Adaptive Projection (RAP)

Run k times

$\hat{D}_1$

$\hat{D}_2$

$\vdots$

$\hat{D}_k$

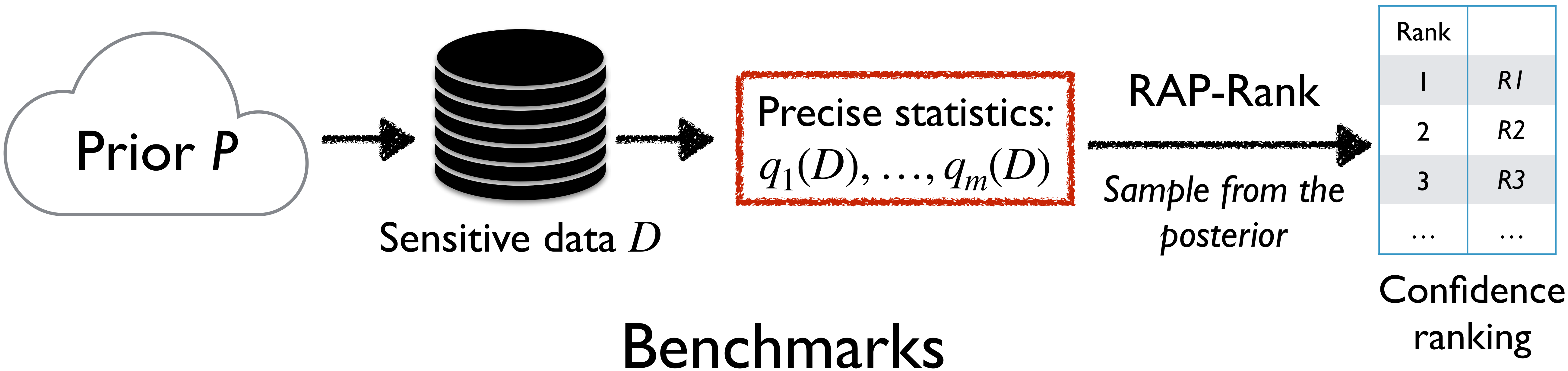
$\hat{D}$

Rank by #times each record appears in $\hat{D}$

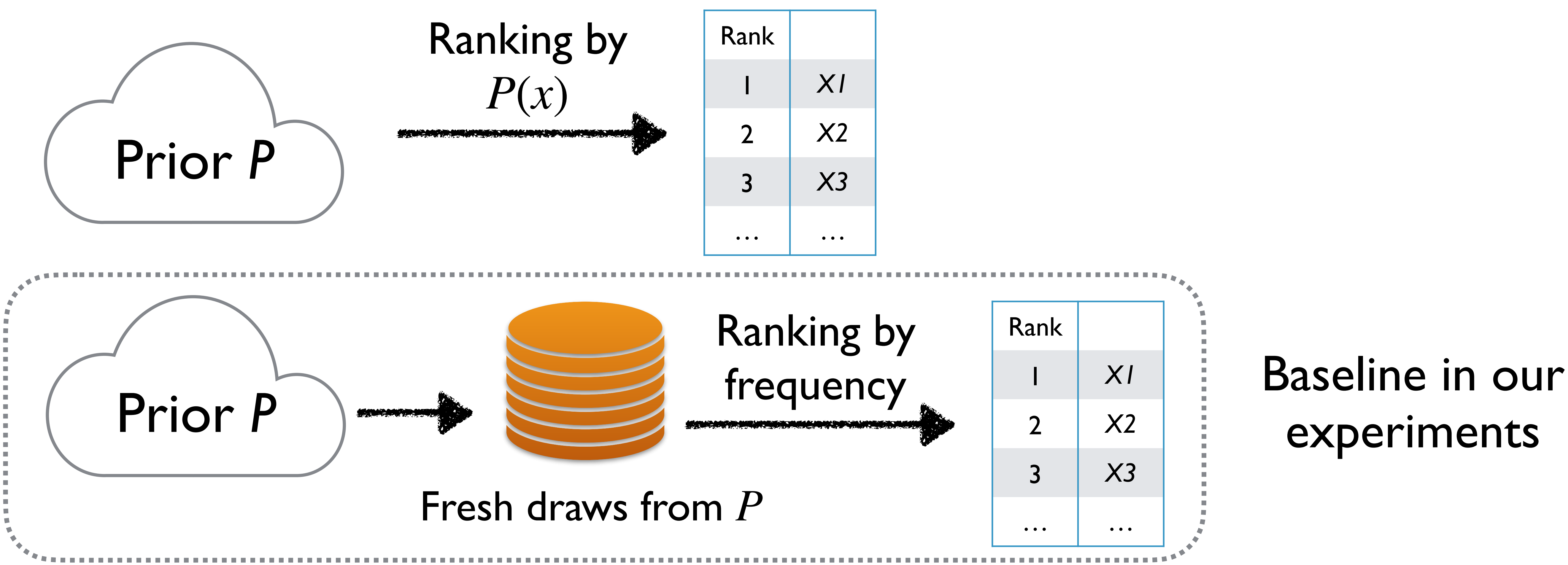
Rank	Record
1	R_1
2	R_2
3	R_3
…	…

$$\text{Match-Rate}_k = \frac{1}{k} \sum_{i \leq k} \mathbf{1}[R_i \in D]$$

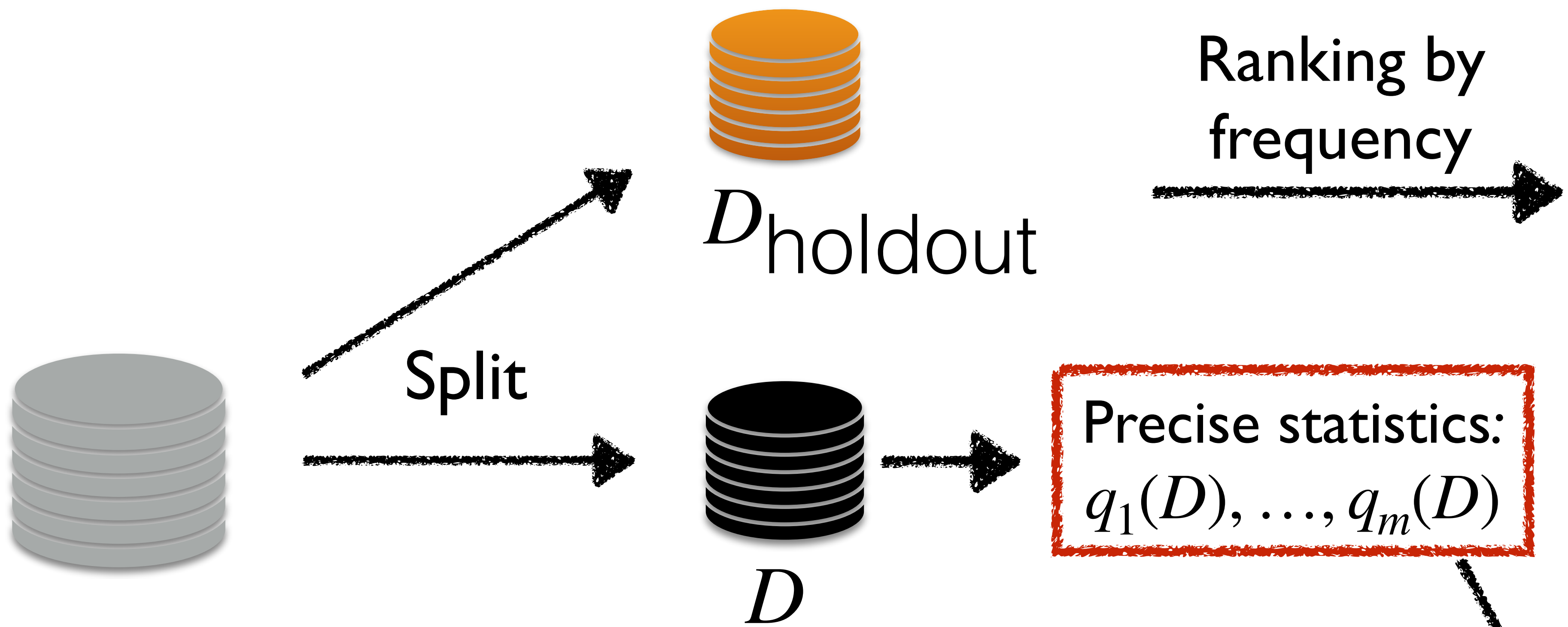
Bayesian Intuition



Benchmarks



Experiments Set Up



Rank	
1	*X1*
2	*X2*
3	*X3*
…	…

Rank	
1	*R1*
2	*R2*
3	*R3*
…	…

Census Data

- American Community Survey (ACS)
- Privacy-protected Microdata File (PPMF)
  - Simulate 2010 Decennial Census

Compare

$$\text{Match-Rate}_k = \frac{1}{k} \sum_{i \le k} \mathbf{1}[R_i \in D]$$

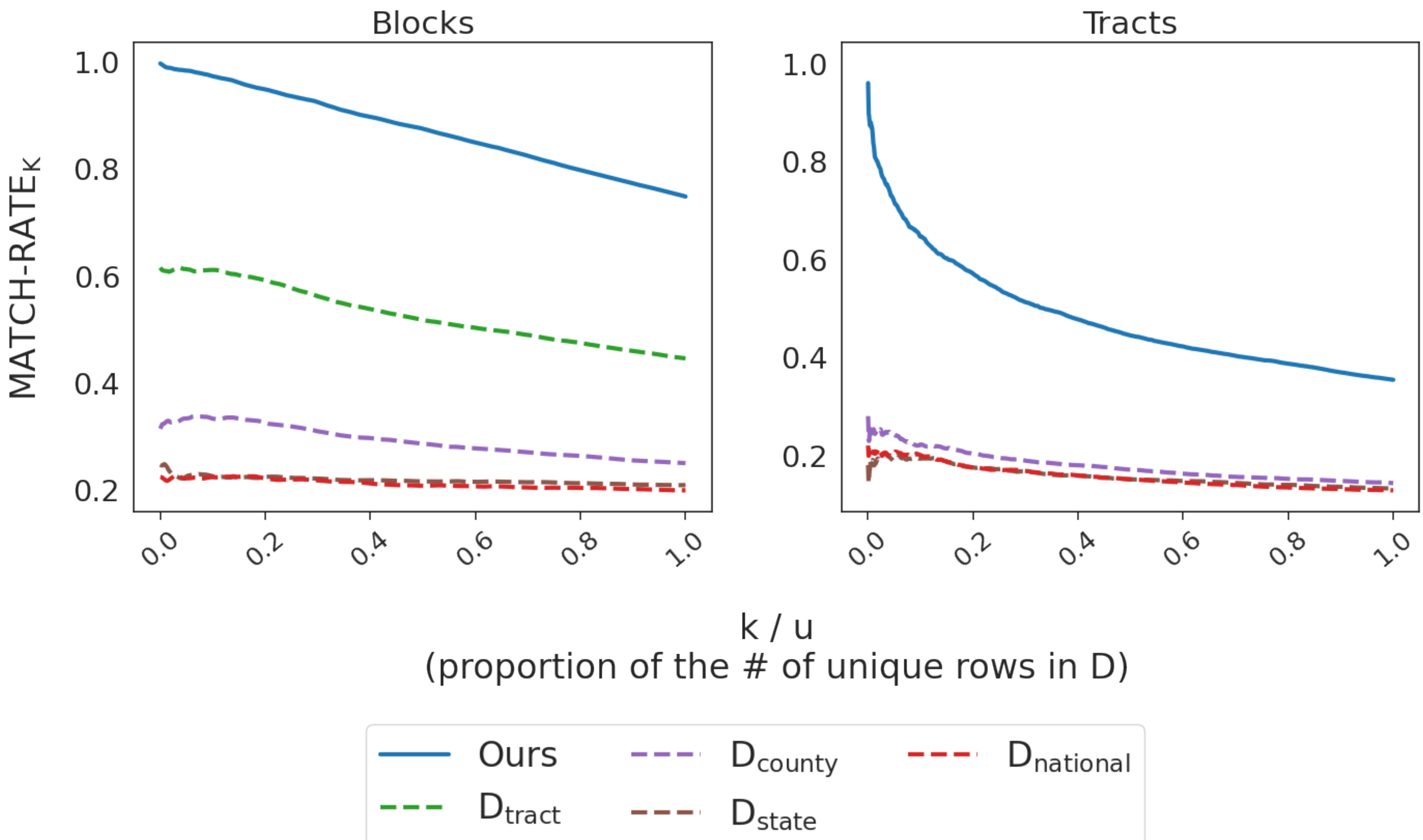
2010 Demonstration Privacy-Protected Microdata Files (PPMF)

- Hierarchy of geographic entities
  - national → state → county → tract → block
  - Hierarchy of prior information

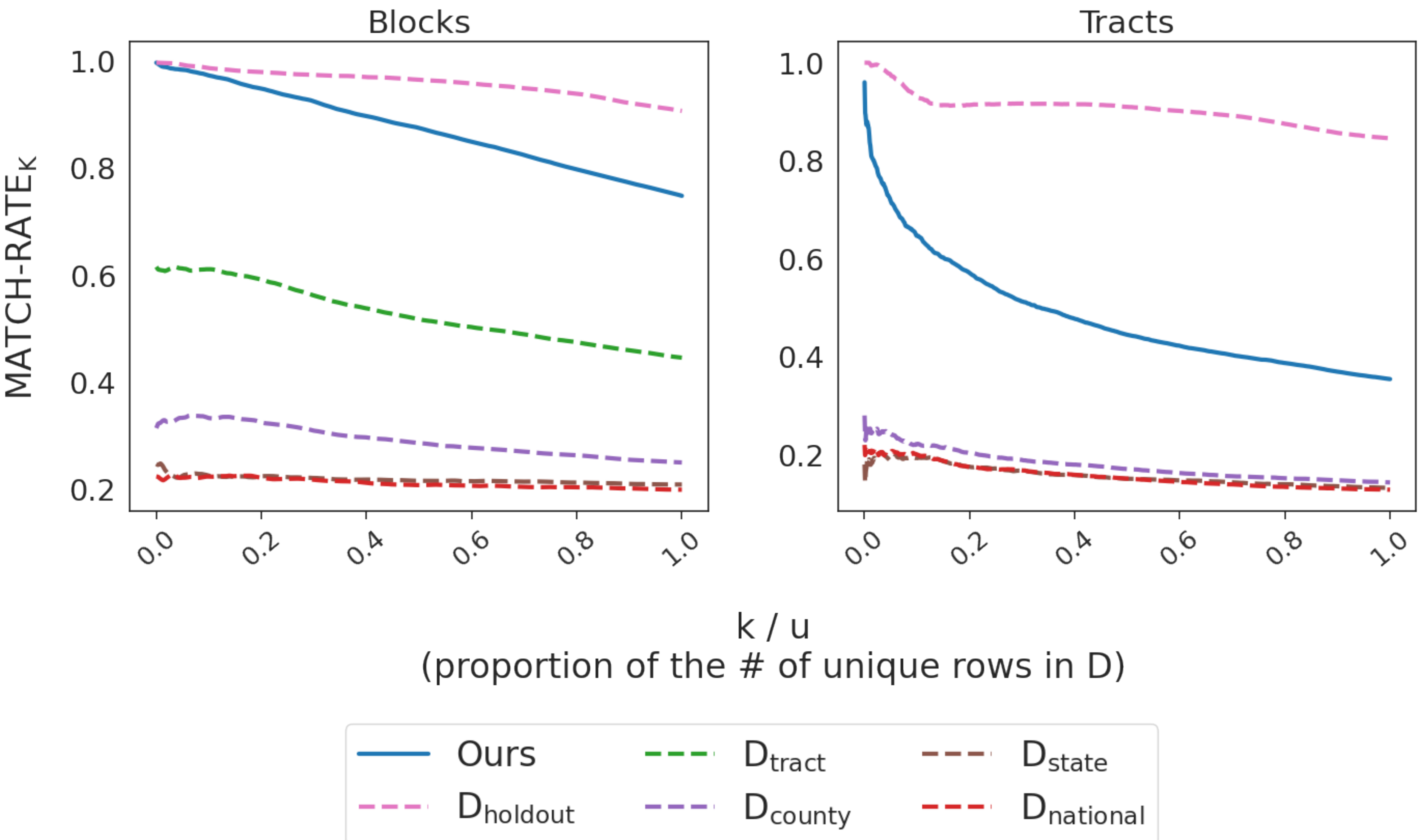
- Reconstruct data at two levels
  - Block: 620 queries
  - Tract: 10-50k queries

P001	Total population by block,
P006	Total races tallied by block,
P007	Hispanic or Latino origin by race by block,
P009	Hispanic or Latino and not Hispanic or Latino by race by block,
P011	Hispanic or Latino and not Hispanic or Latino by race by age (≥ 18) by block,
P012	Sex by age by block,
P012A-I	Sex by age by block iterated by race,
P014	Sex by age (< 20) by block,
PCT012012A-N	Sex by age by tract iterated by major race alone.

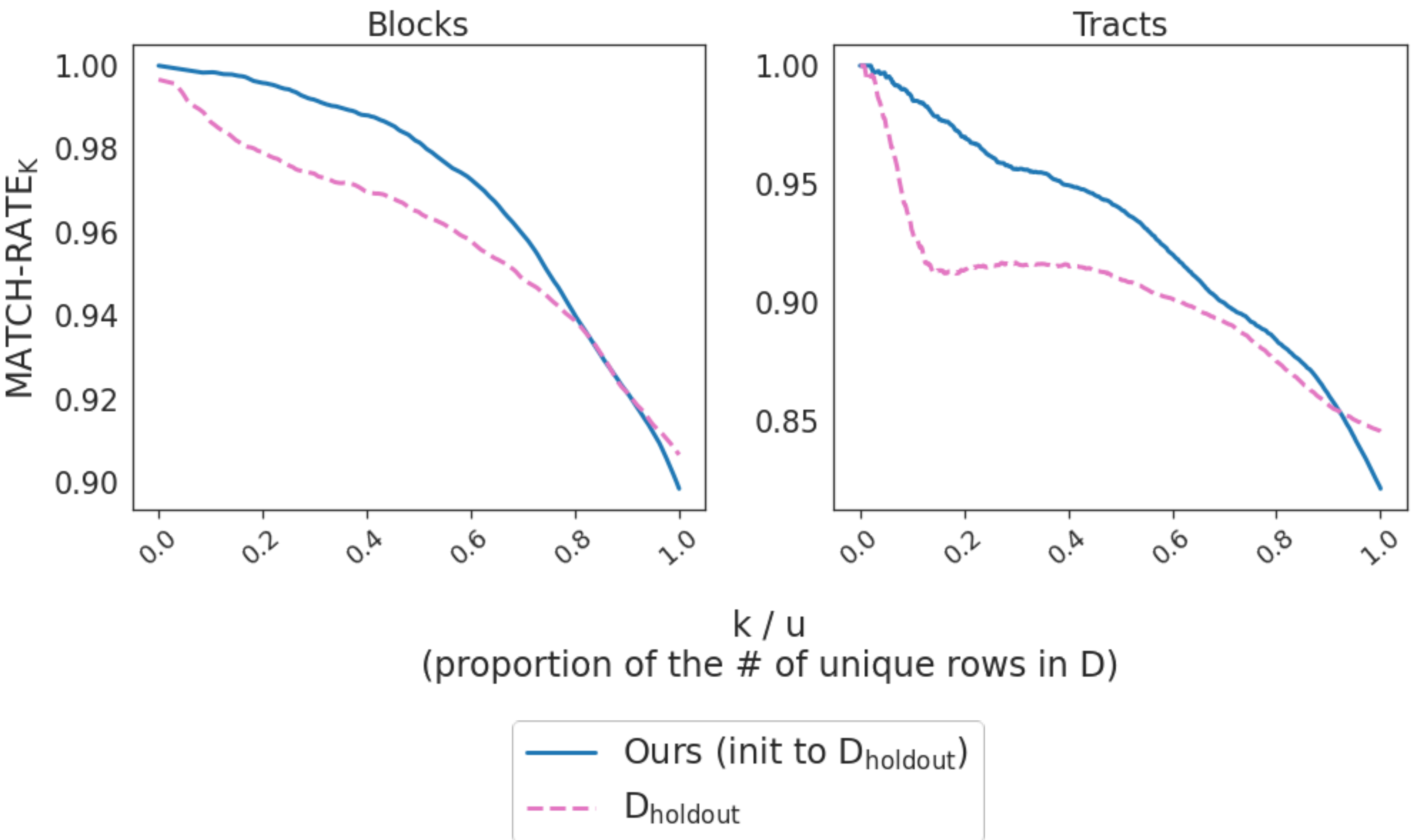
2010 Demonstration Privacy-Protected Microdata Files



2010 Demonstration Privacy-Protected Microdata Files



2010 Demonstration Privacy-Protected Microdata Files



Thanks!