Impacts of Data Privacy and Equity on Public Policy

Ferdinando Fioretto
Syracuse U.

Claire McKay Bowen
Urabn Institute

FAccT 2022

URBAN INSTITUTE · ELEVATE · THE · DEBATE

Syracuse University

Part I

What is data privacy and confidentiality and why is it important?

BRIEFING ROOM

Executive Order On Advancing Racial Equity and Support for Underserved Communities Through the Federal Government

JANUARY 20, 2021 • PRESIDENTIAL ACTIONS

"...advanc[e] equity for all, including people of color and others who have been historically underserved, marginalized, and adversely affected by persistent poverty and inequality."


ONE NATION, TRACKED
AN INVESTIGATION INTO THE SMARTPHONE TRACKING INDUSTRY FROM TIMES OPINION

Lack of Public Data Hampers COVID-19 Fight

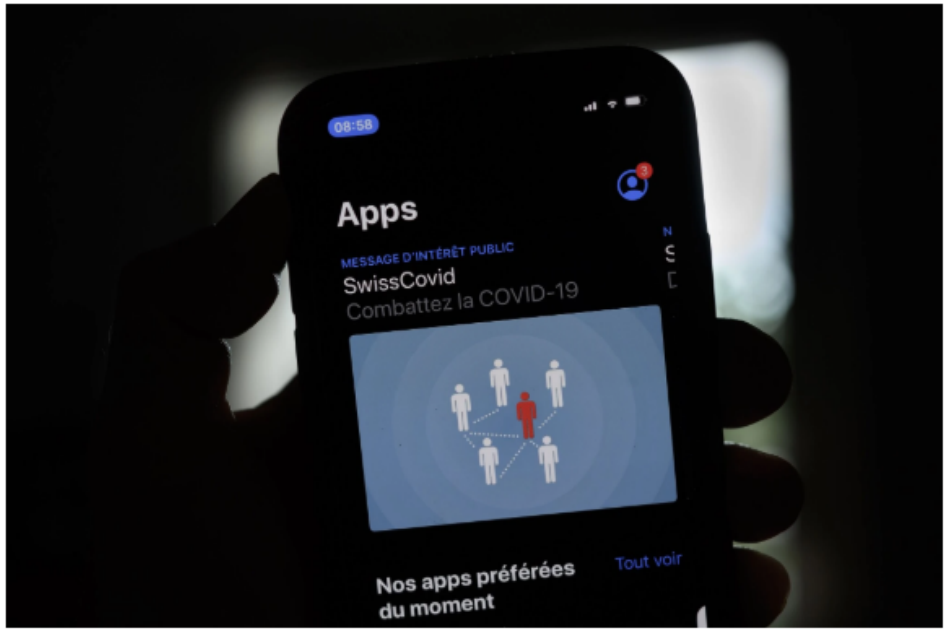
STATELINE ARTICLE August 3, 2020 By: Christine Vestal Topics: Health Read time: 7 min

A ventilator helps a COVID-19 patient breathe at a Houston hospital. Hospital data related to the coronavirus pandemic will now be collected by a private technology firm, rather than the Centers for Disease Control and Prevention. Epidemiologists say better COVID-19 data is needed to improve the nation's response.

David J. Phillip/The Associated Press

Google Promises Privacy With Virus App but Can Still Collect Location Data

Some government agencies that use the software said they were surprised that Google may pick up the locations of certain app users. Others said they had unsuccessfully pushed Google to make a change.



Switzerland has asked Google to decouple the location setting requirement on Android phones from Bluetooth, which the country's virus alert app uses to detect nearby smartphones. Fabrice Coffrini/Agence France-Presse — Getty Images

By Natasha Singer

July 20, 2020

HEALTH

'It's not a pretty picture': Why the lack of racial data around COVID vaccines is 'massive barrier' to better distribution

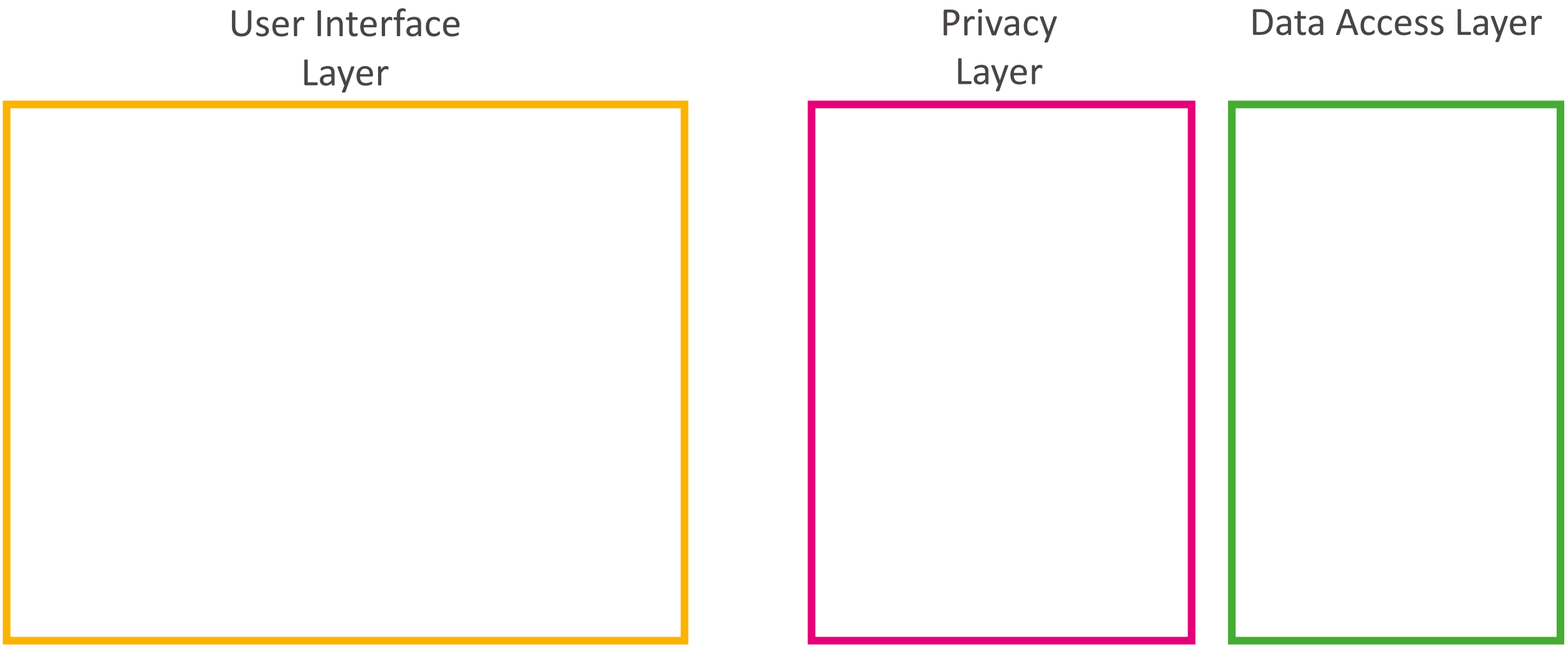
Nada Hassanein USA TODAY

Published 5:30 a.m. ET Feb. 1, 2021 | **Updated 2:10 p.m. ET Feb. 1, 2021**

Abigail Echo-Hawk, chief research officer with Seattle Indian Health Board and a member of the Pawnee Tribe, gets a shot of the Moderna COVID-19 vaccine on Dec. 21. A colleague used a black pen to inscribe "For the (Heart) love of Native People" over the injection spot. *Karen Ducey, Getty Images*

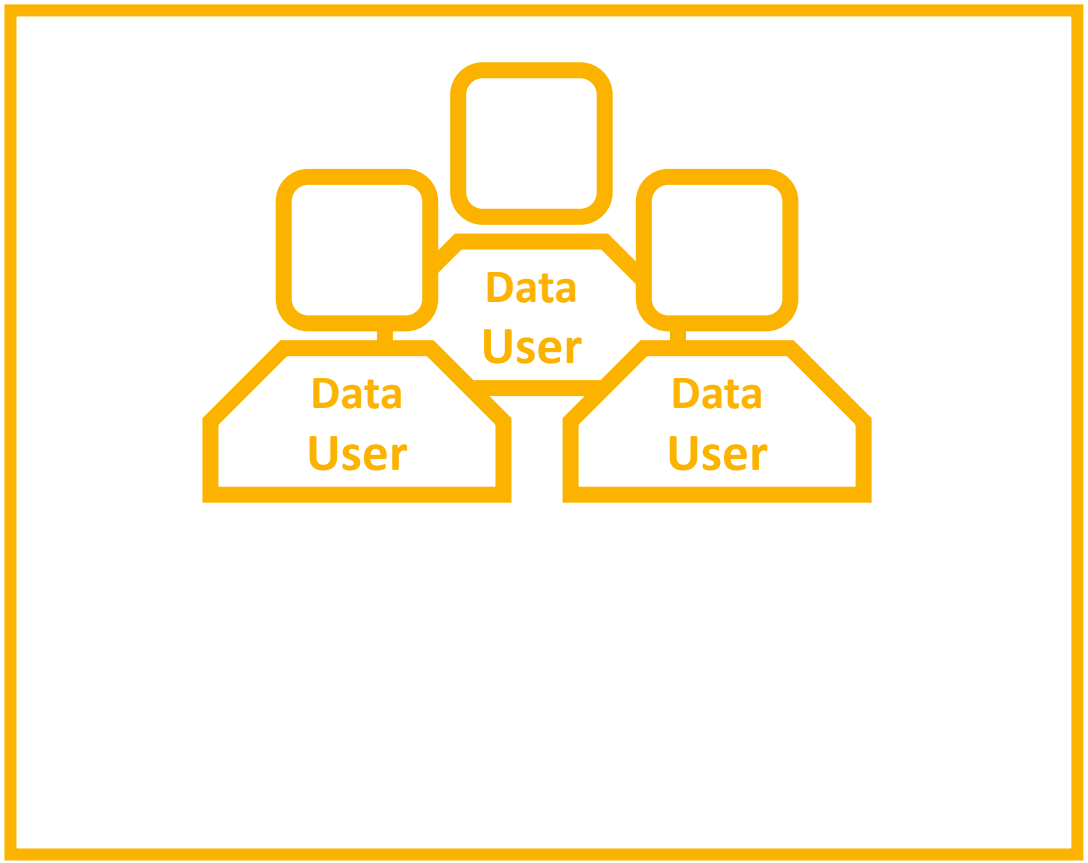
How do public policymakers gain access to confidential data?

How do we access confidential data?



How do we access confidential data?

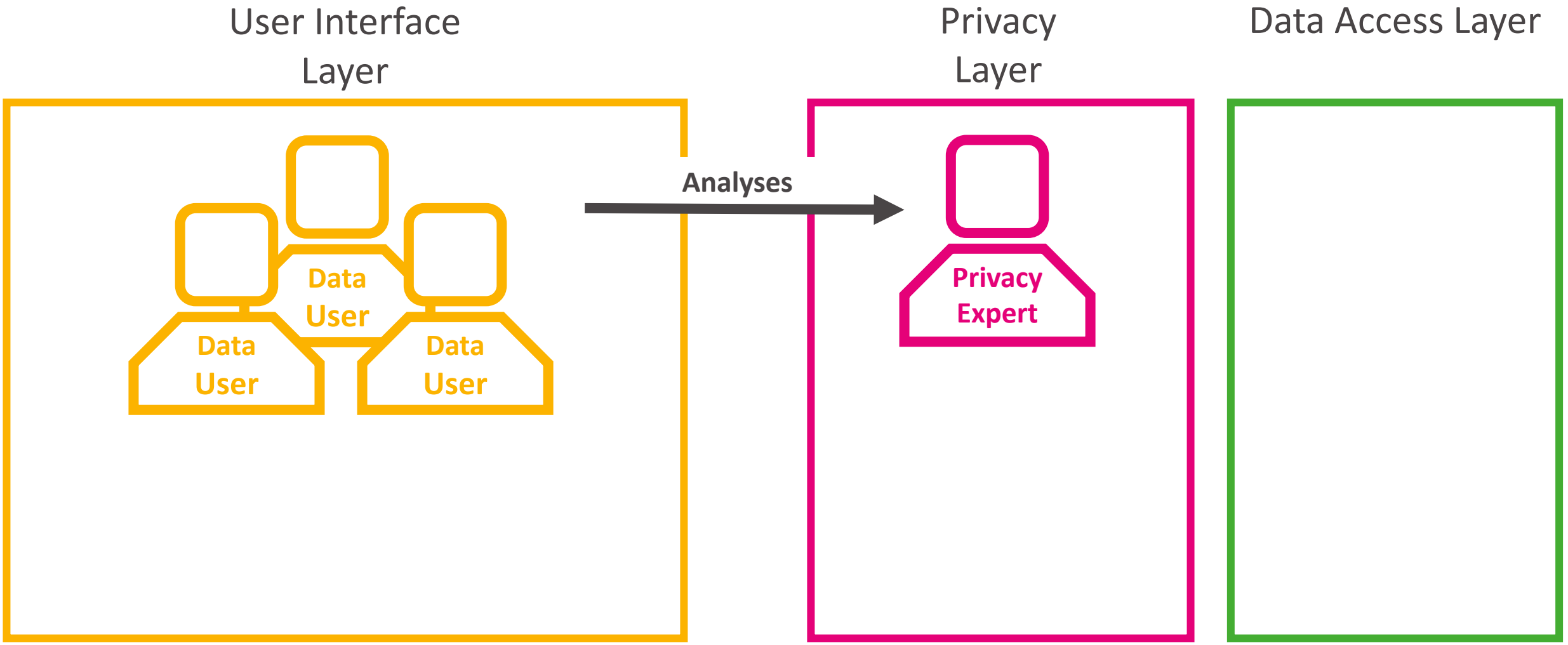
User Interface Layer



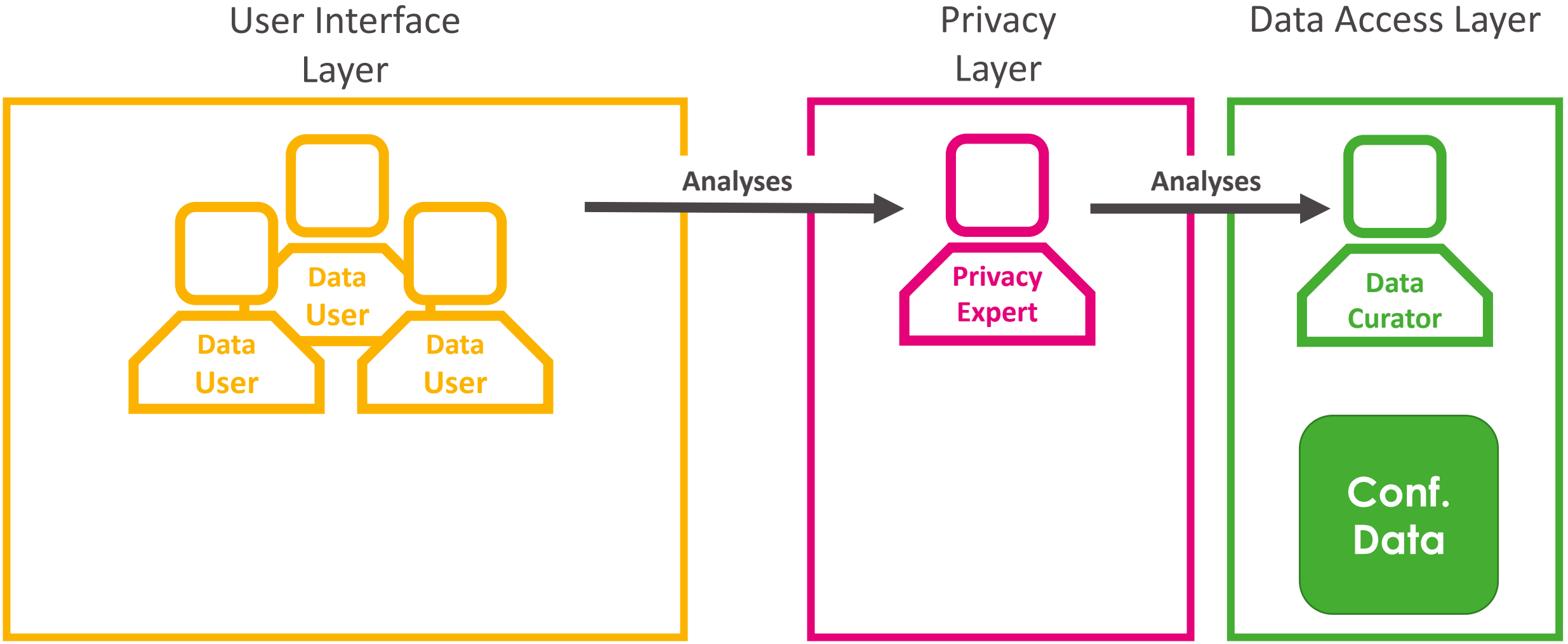
Privacy Layer

Data Access Layer

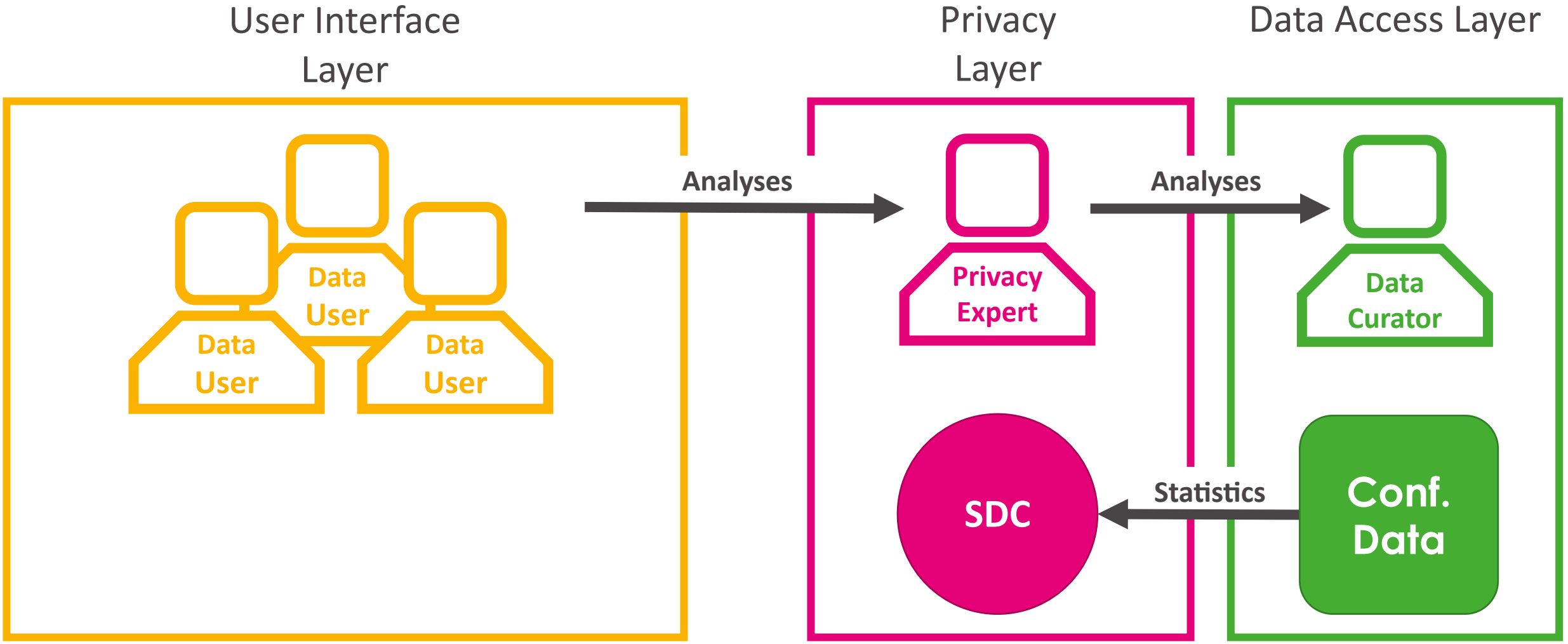
How do we access confidential data?



How do we access confidential data?

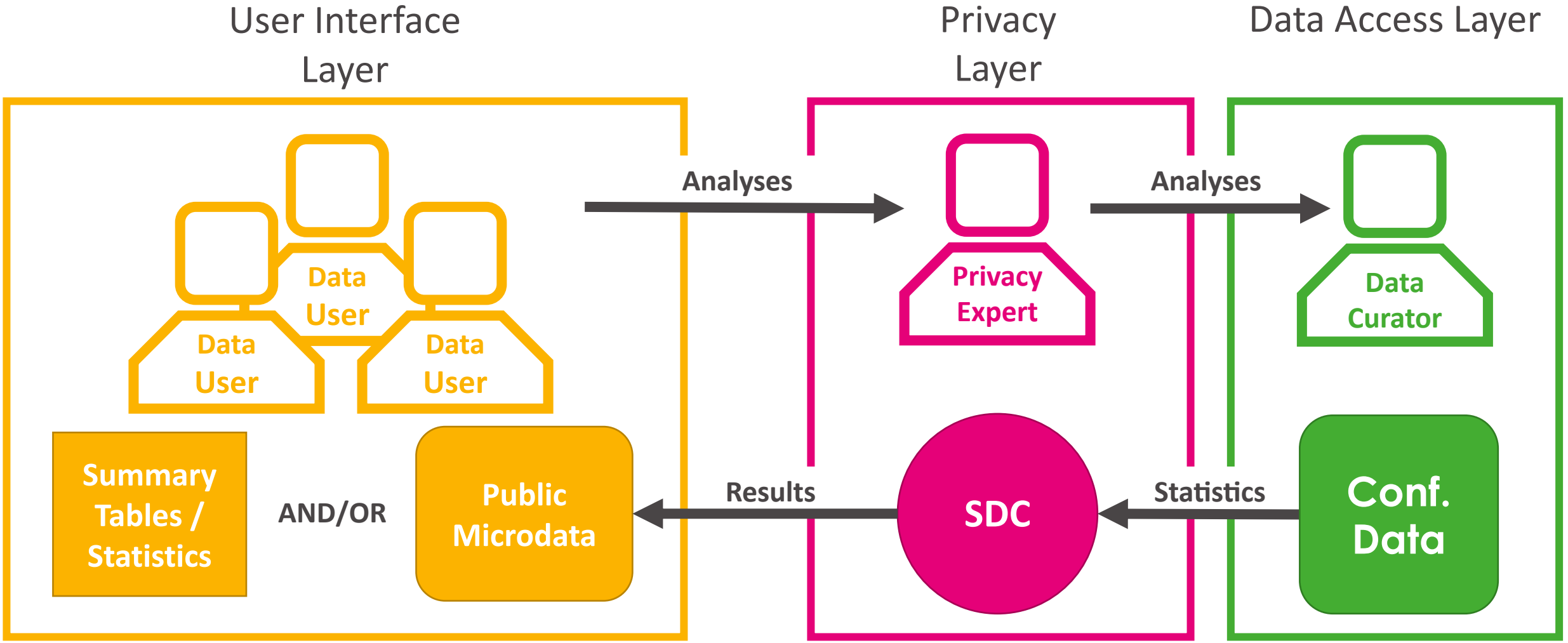


How do we access confidential data?



SDC / SDL = Statistical Disclosure Control or Limitation

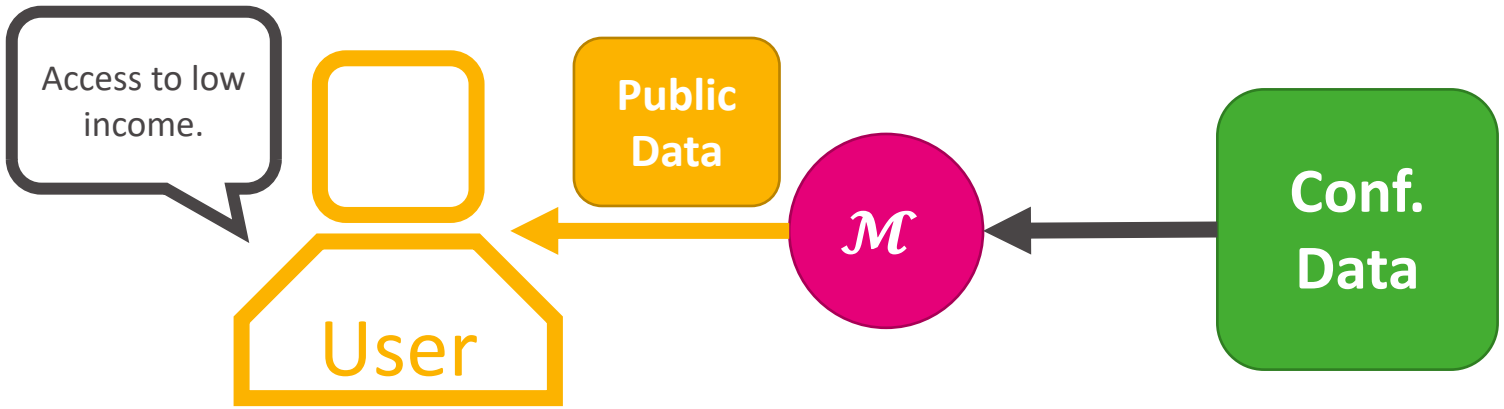
How do we access confidential data?



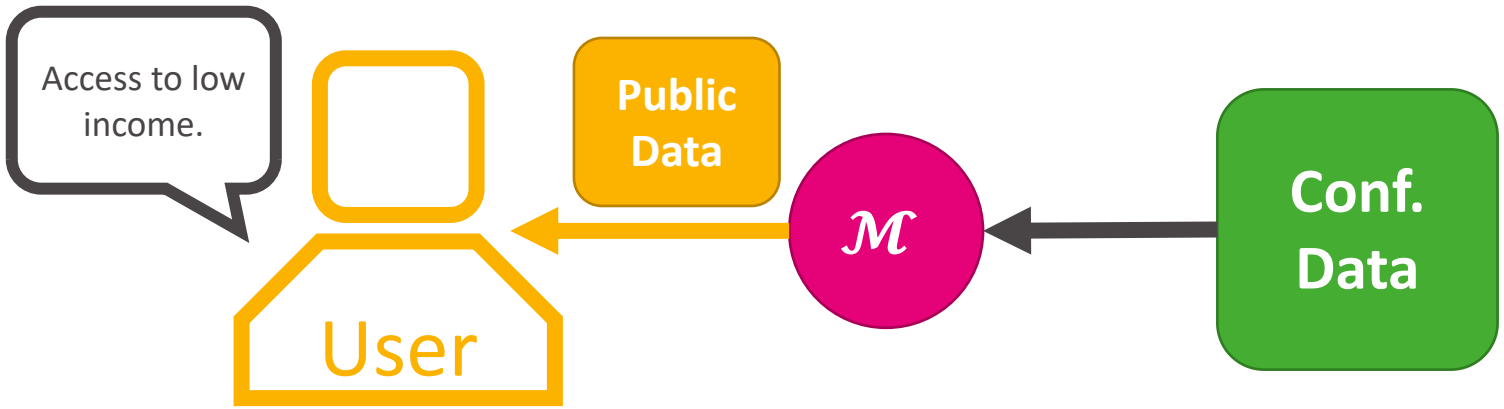
SDC / SDL = Statistical Disclosure Control or Limitation

Why is it hard to implementing this framework?

Why is it hard to implementing this framework?

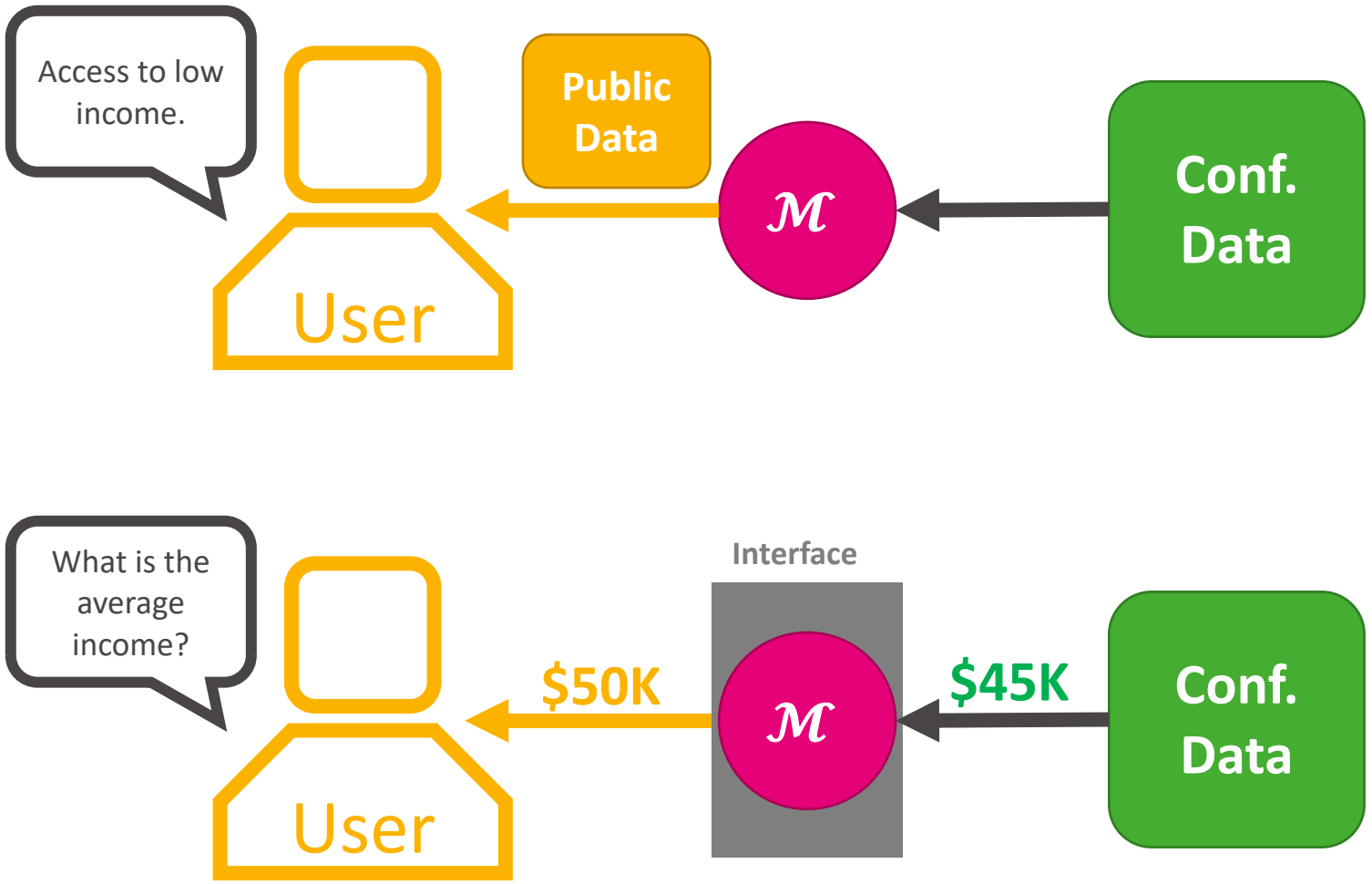


Why is it hard to implementing this framework?



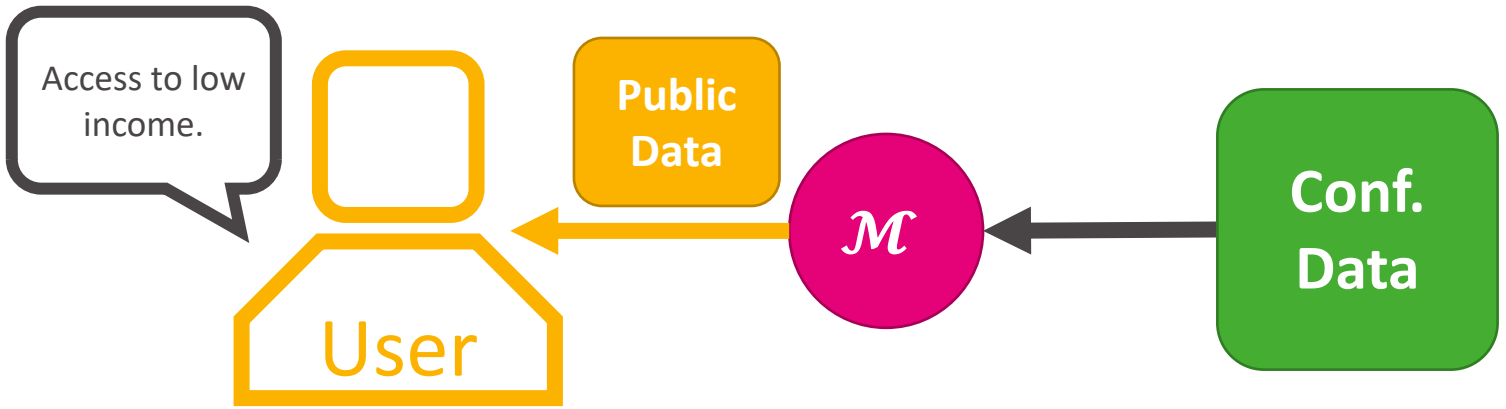
How do we **measure utility** (usefulness) and **disclosure risk** of the data?

Why is it hard to implementing this framework?

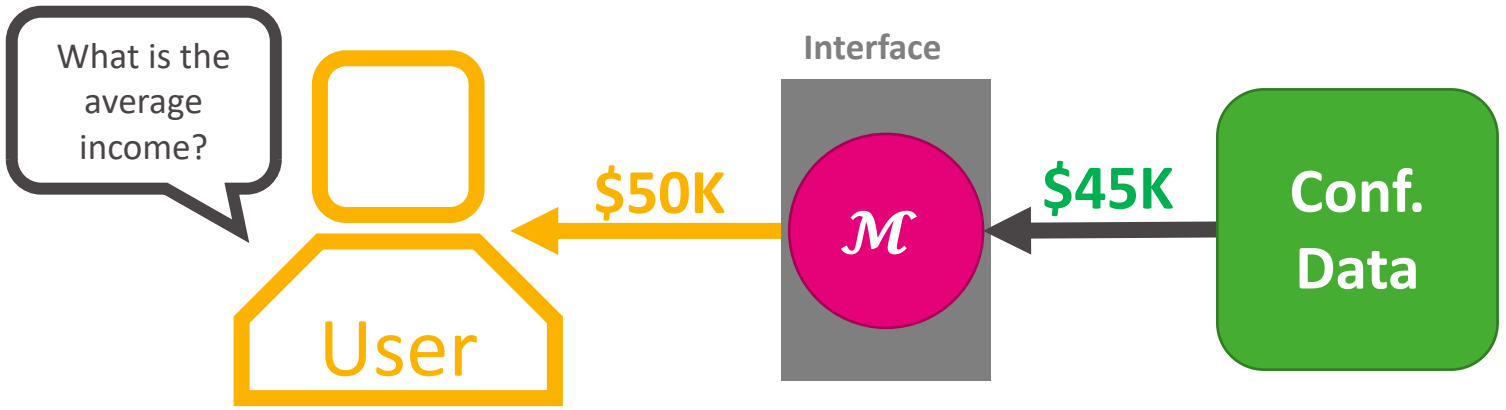


How do we **measure utility** (usefulness) and **disclosure risk** of the data?

Why is it hard to implementing this framework?



How do we **measure utility** (usefulness) and **disclosure risk** of the data?



How **much noise should be added** and how do you **limit the number of queries**?

What is the workflow?

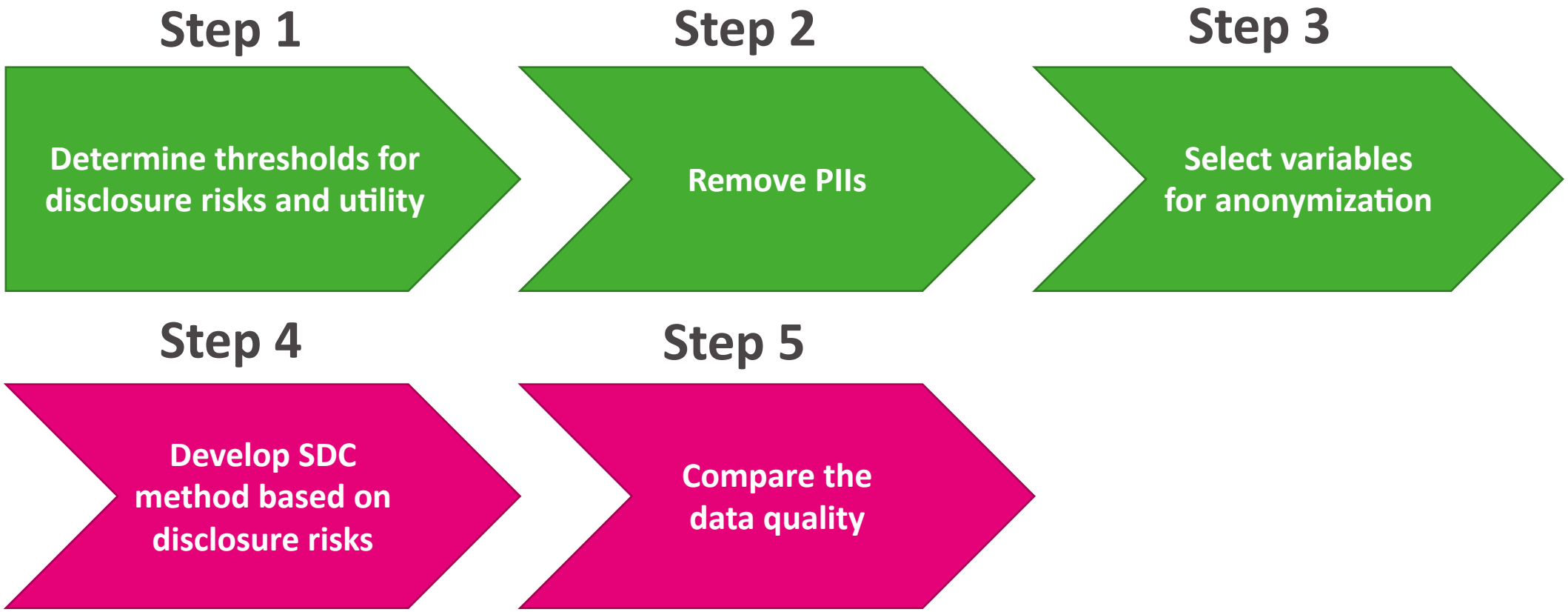
What is the workflow?

Step 1

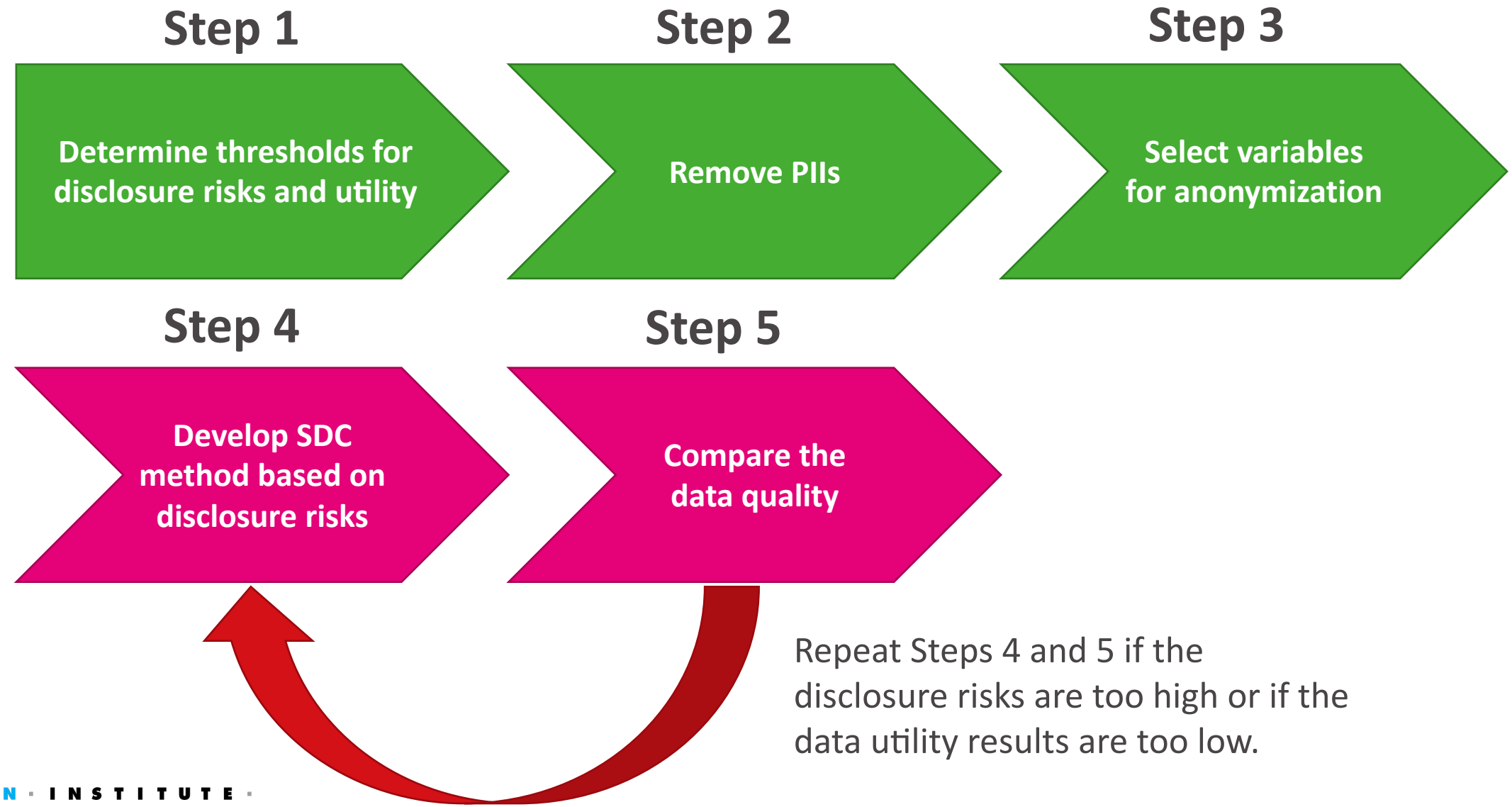
What is the workflow?

What is the workflow?

What is the workflow?

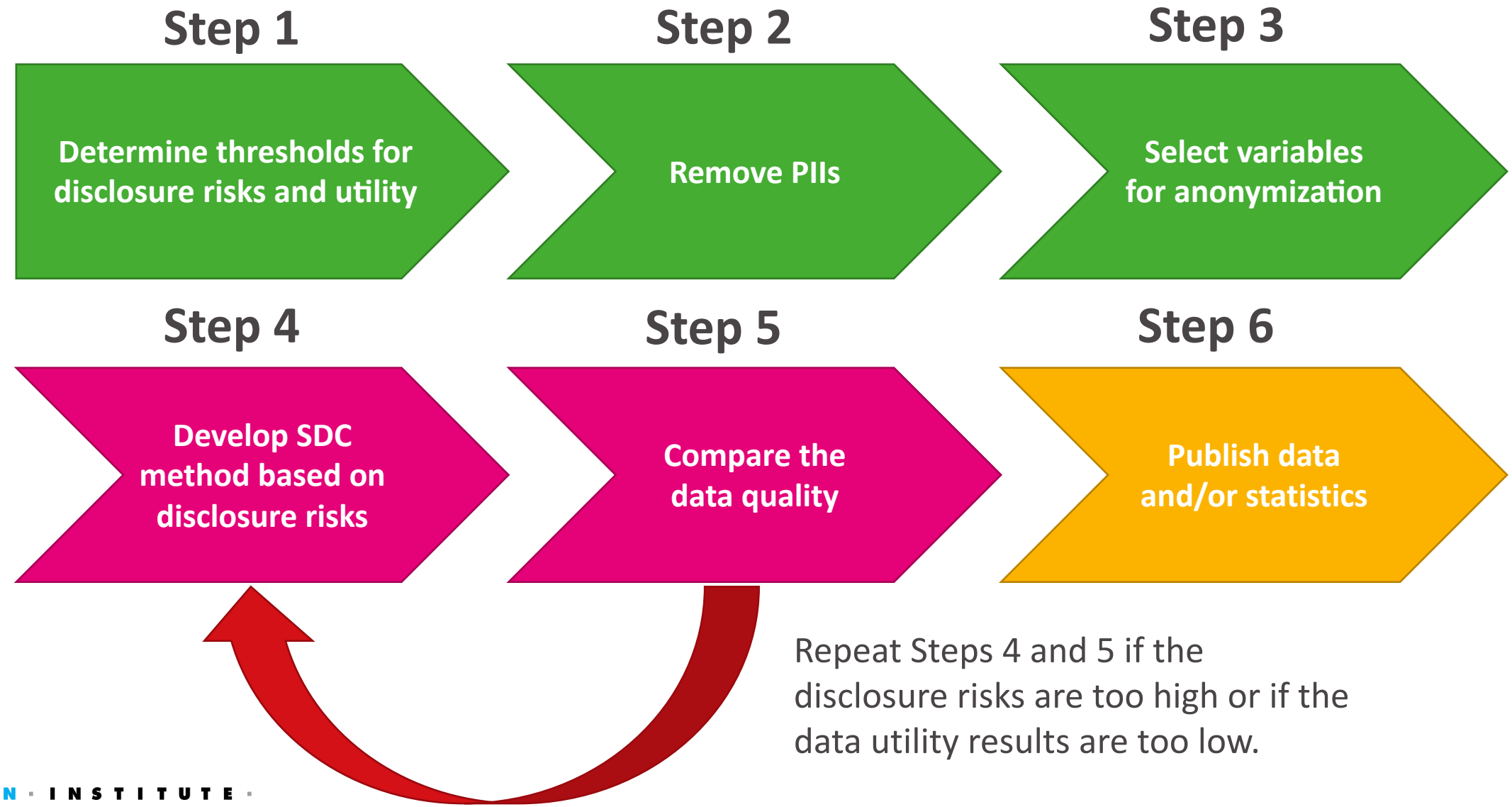


What is the workflow?



Repeat Steps 4 and 5 if the disclosure risks are too high or if the data utility results are too low.

What is the workflow?



Repeat Steps 4 and 5 if the disclosure risks are too high or if the data utility results are too low.

How are these methods structured?

How are these methods structured?

1. **Pre-Processing Step:** determining priorities for which statistics to preserve

How are these methods structured?

1. **Pre-Processing Step:** determining priorities for which statistics to preserve
2. **Privacy Step:** applying a sanitizer to the desired statistic

How are these methods structured?

1. **Pre-Processing Step:** determining priorities for which statistics to preserve
2. **Privacy Step:** applying a sanitizer to the desired statistic
3. **Post-Processing Step:** ensuring the results of the statistics are consistent with realistic constraints, such as negative population counts

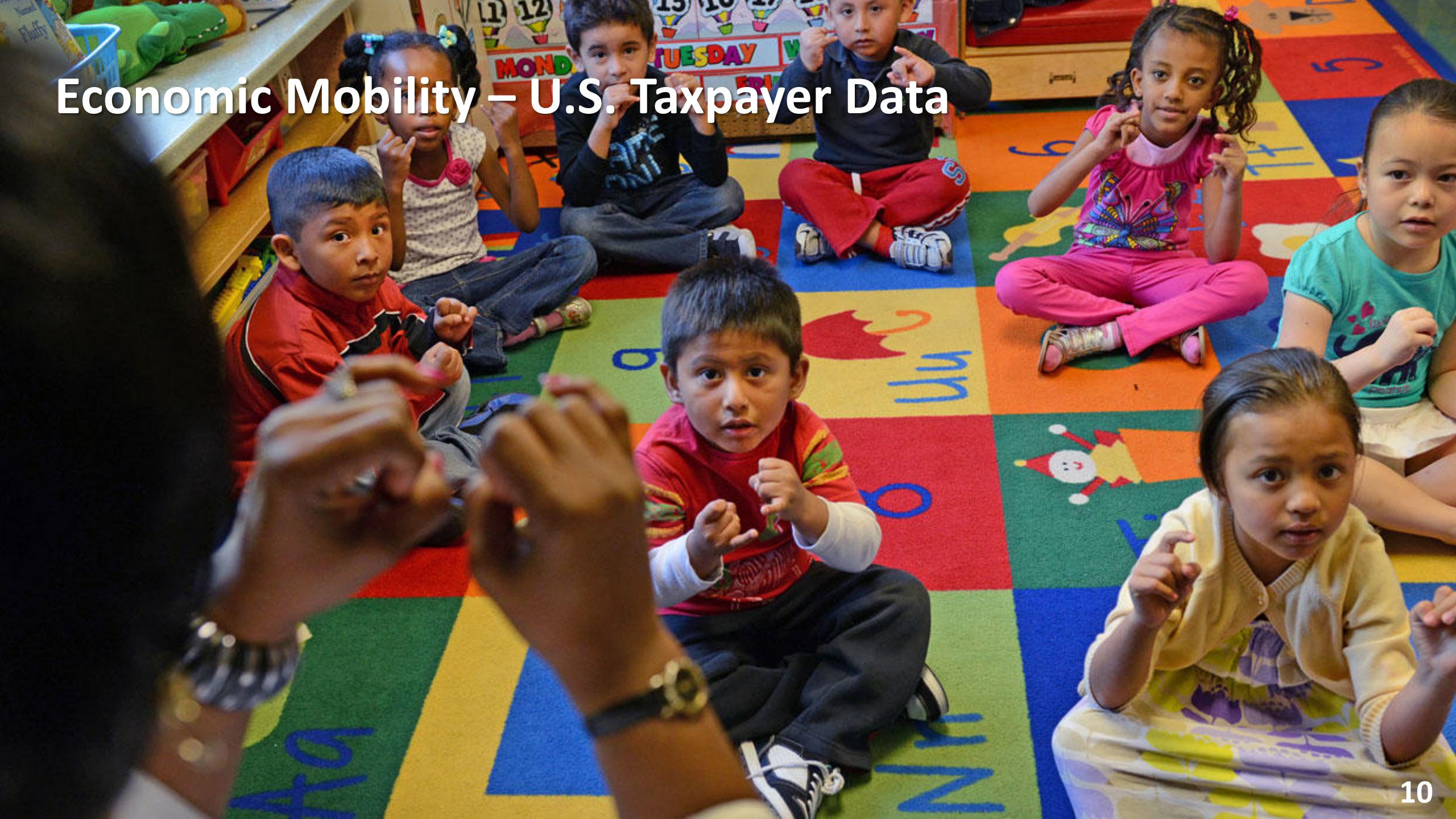
How are these methods structured?

1. **Pre-Processing Step:** determining priorities for which statistics to preserve
2. **Privacy Step:** applying a sanitizer to the desired statistic
3. **Post-Processing Step:** ensuring the results of the statistics are consistent with realistic constraints, such as negative population counts

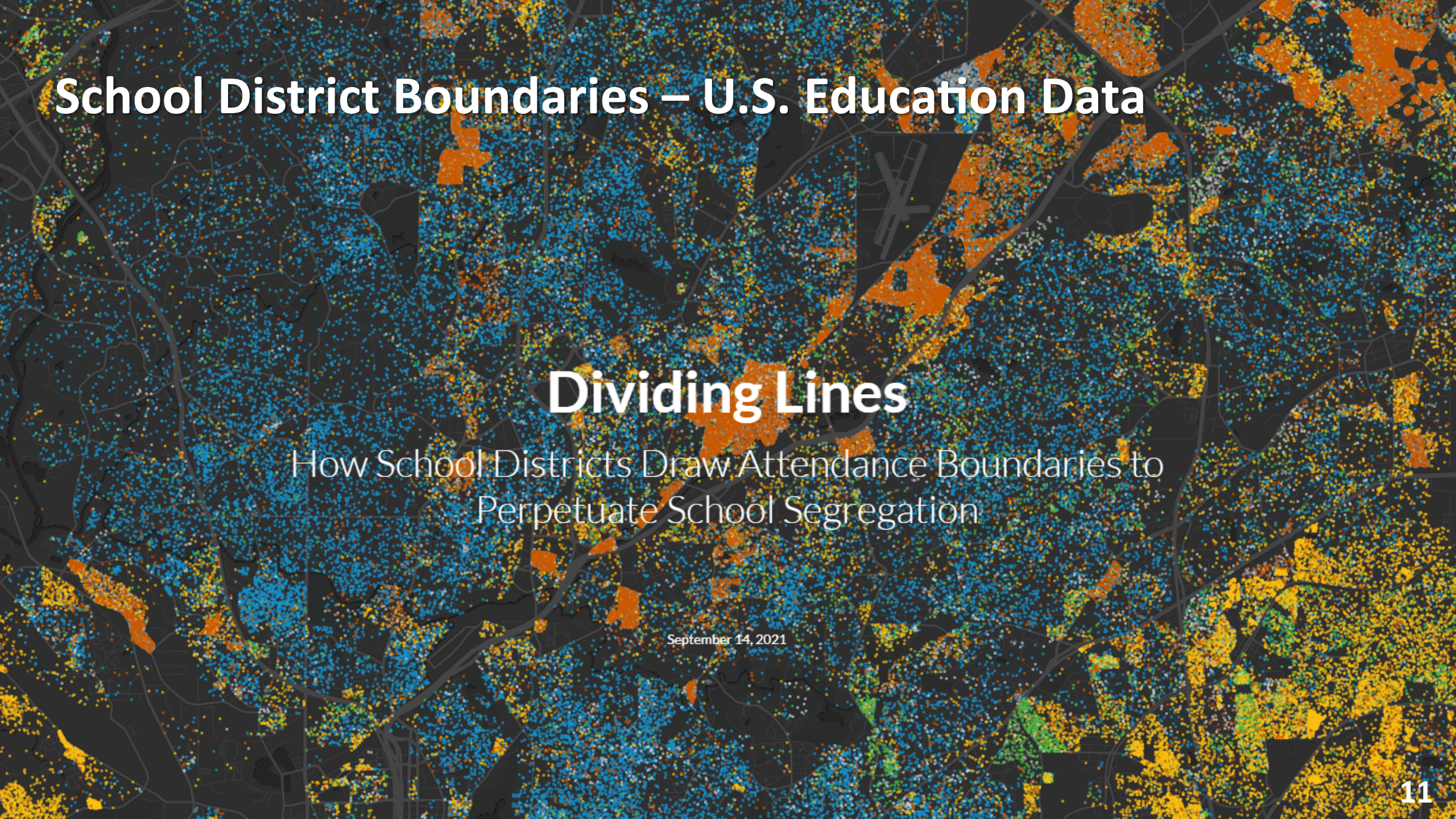
SDC Method

How are data used for public policy decision making?

Economic Mobility – U.S. Taxpayer Data



School District Boundaries – U.S. Education Data



2020 Census Data

- Allocating $1.5 trillion budget
- Apportioning 435 congressional seats
- Planning for natural disasters
- Understanding economic well-being
- Determining the number of restaurant permits

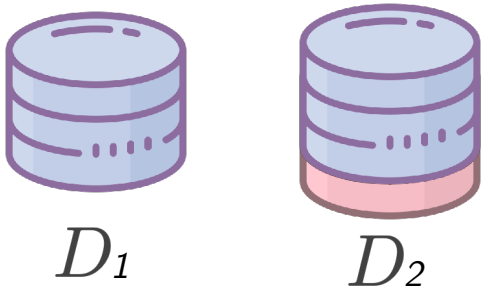
Part II

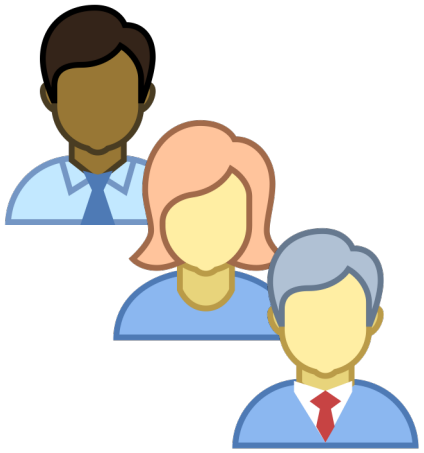
Dwork:06

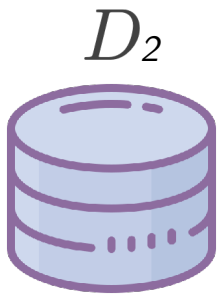
Differential Privacy

A randomized algorithm $\mathcal{A}$ is ε-differentially private if, for all pairs of inputs D1, D2, differing in one entry, and for any output O:

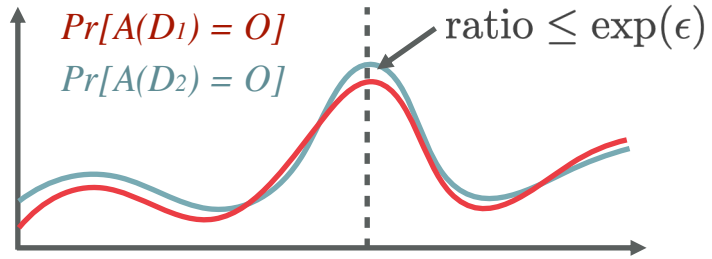
$$\frac{\Pr[\mathcal{A}(D_1) = O]}{\Pr[\mathcal{A}(D_2) = O]} \leq \exp(\epsilon)$$







$\mathcal{A}(D_2)$



Intuition: An adversary should not be able to use output O to distinguish between any D_1 and D_2

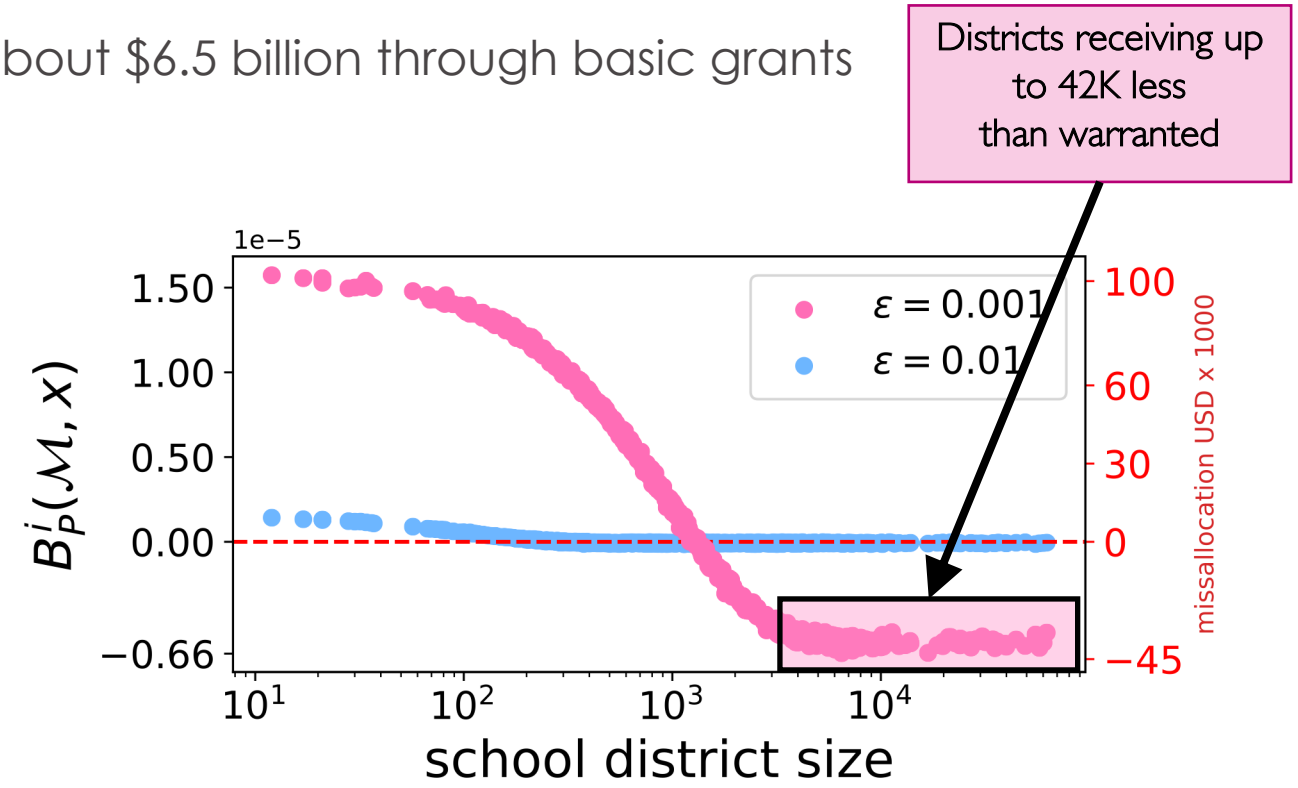
Disproportionate Effects in Title 1 Allotment

- Title 1 of the Elementary and Secondary Education Act is one of the largest U.S. program offering educational assistance to disadvantaged children
- In the fiscal year 2015 alone, it distributed about $6.5 billion through basic grants
- Allotment:

count of children 5 to 17 in district i

$$P_i^F(\boldsymbol{x}) \stackrel{\text{def}}{=} \left(\frac{x_i \cdot a_i}{\sum_{i \in [n]} x_i \cdot a_i} \right)$$

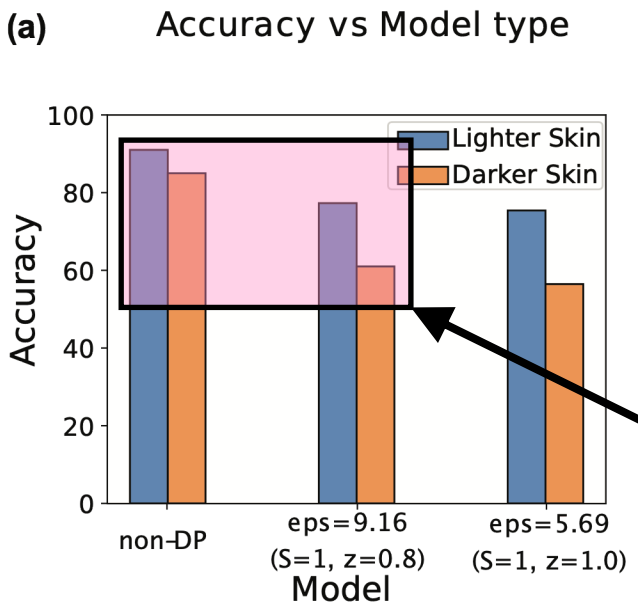
student expenditures in district i

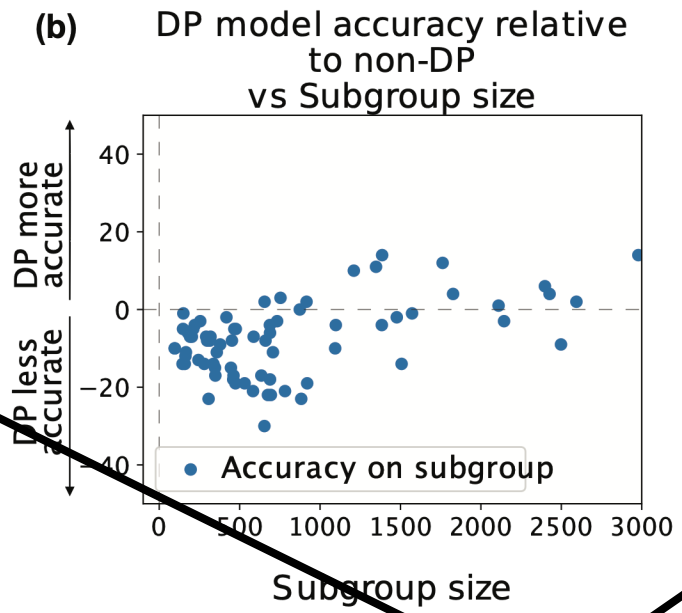


Puyol et al: FAT*20 / Tran et al. IJCAI-21

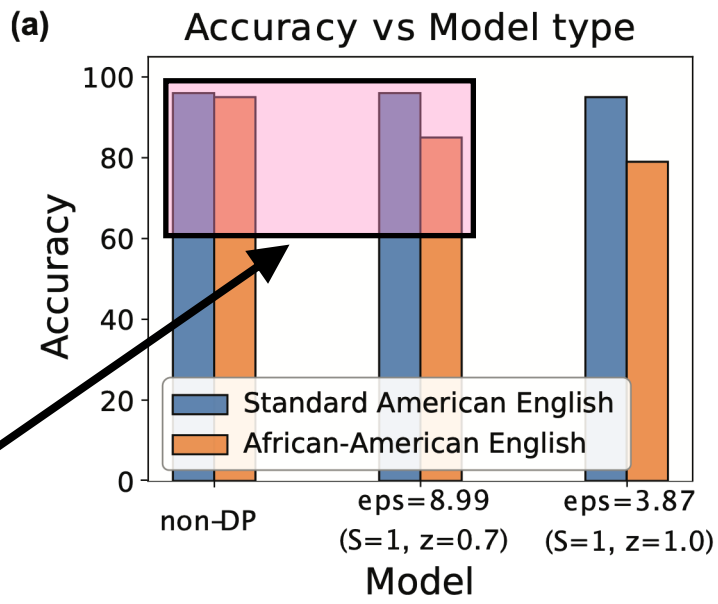
Disproportionate Effects in ML

Gender and age classification on facial images





Sentiment analysis of tweets



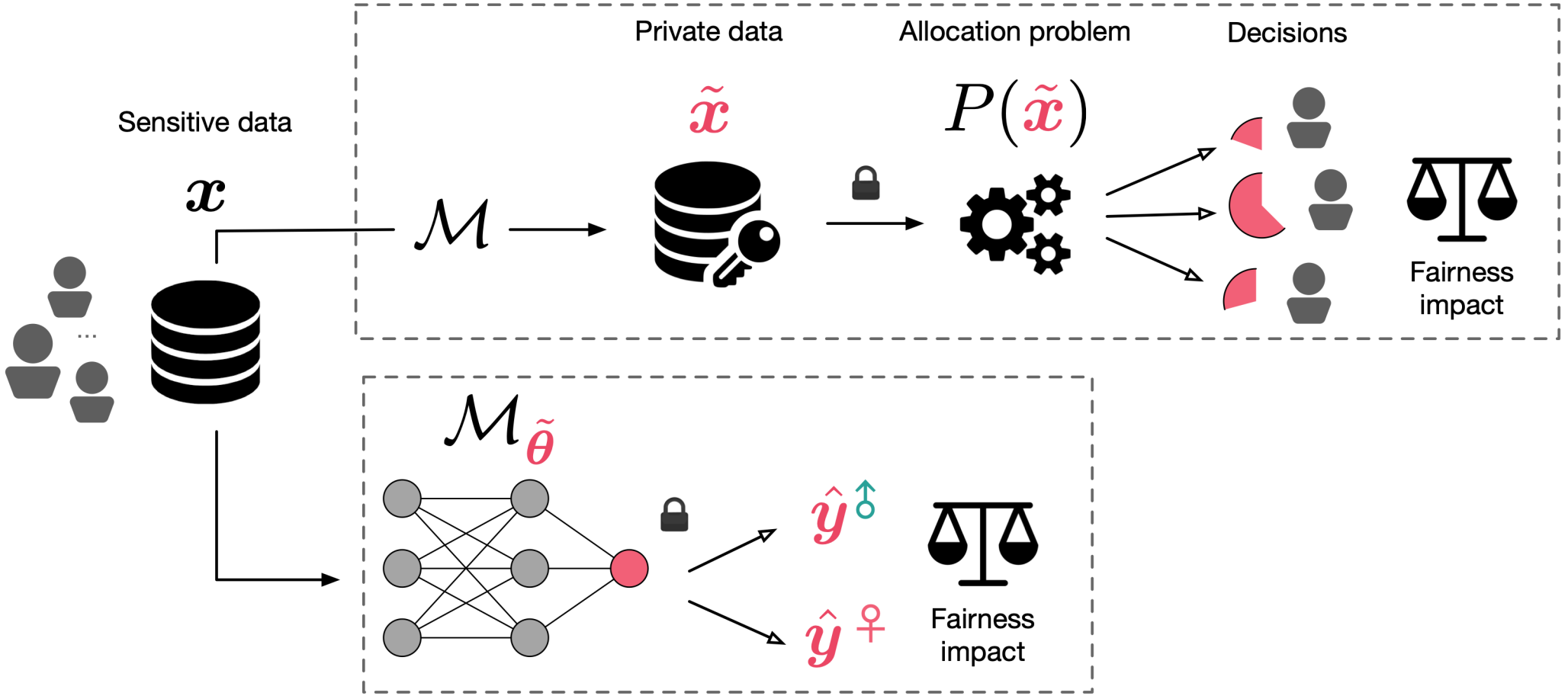
As the privacy increases the accuracy disparity of the learning task increases

Bagdasaryan:19

Societal Impact

- The resulting outcomes can have significant societal and economic impacts on the involved individuals:
- Classification errors may penalize some groups over others in important determinations including criminal assessment, hiring or landing
- Biased decisions can result in disparities regarding the allocation of critical funds, benefits, and therapeutics
- While these observations are increasingly apparent, their causes have only recently started to receive attention.

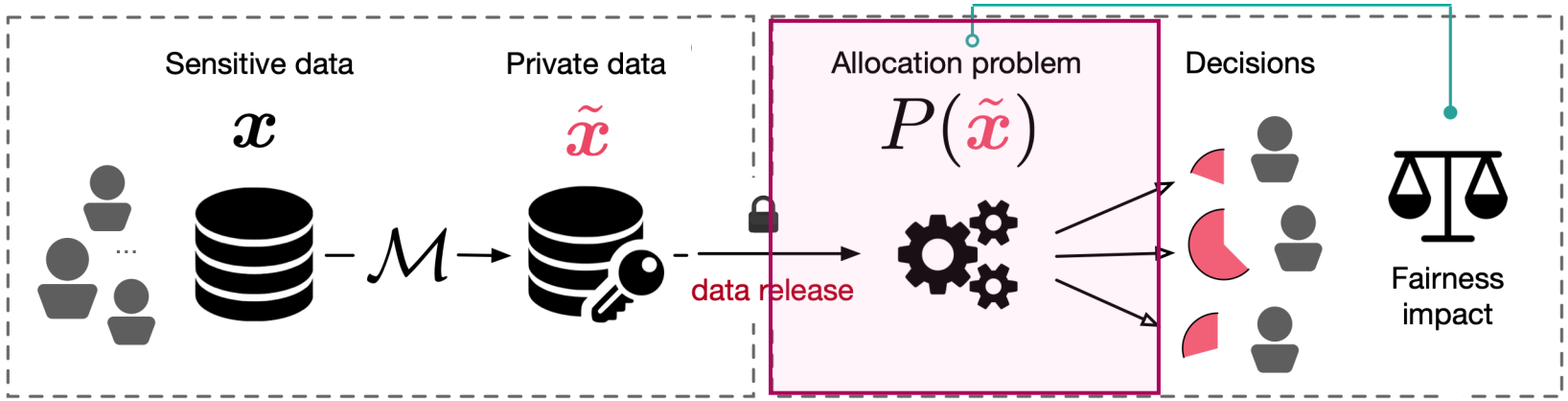
Setting and Outline of Part II



Why does disparity arise in decision making?

Census data-release perspective

Fairness in DP Downstream Decisions



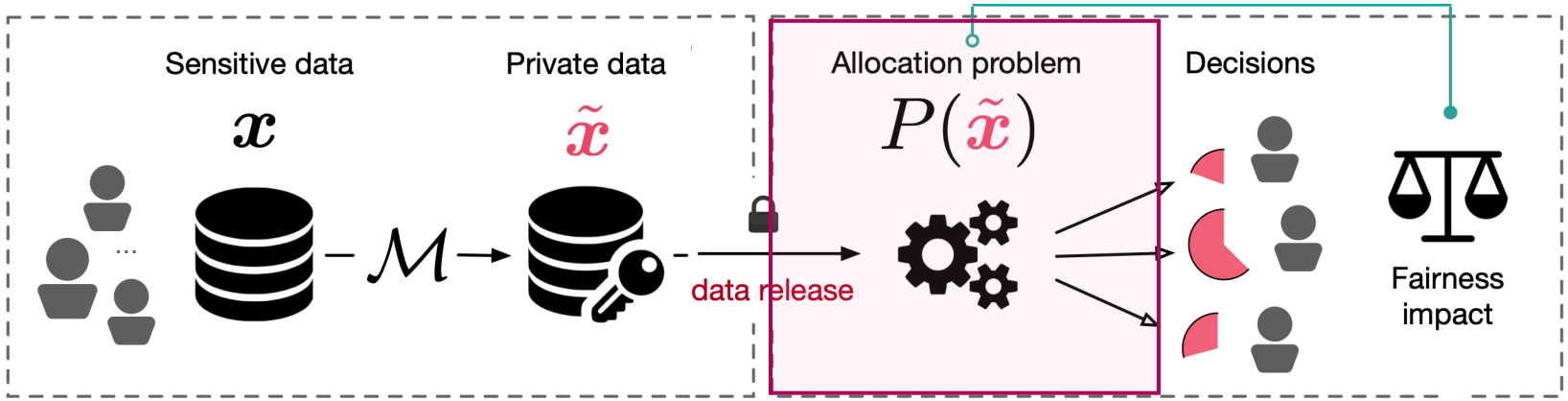
Given a dataset $x \in \mathcal{X} \subseteq \mathbb{R}^k$ of n entities, whose elements x_i describe some measurable quantity of entity $i \in [n]$ we consider two classes of decision problems:

1. Allotment problems: $P\colon \mathcal{X} \times [n] \to \mathbb{R}$,
 which distributes a finite set of resources to some entity.

2. Decision rules: $P\colon \mathcal{X} \times [n] \to \{0,1\}$,
 which determines whether an entity qualifies for some benefit.

Tran et al. IJCAI-21

Fairness in DP Downstream Decisions



Bias: $$B_P^i(\mathcal{M}, x) = \mathbb{E}_{\tilde{x}\sim\mathcal{M}(x)}\left[P_i(\tilde{x})\right] - P_i(x).$$

Definition (α-Fairness). A data-release mechanism M is said α-fair w.r.t. a problem P if, for all datasets $x \in \mathcal{X}$ and all $i \in [n]$

$$\zeta_B^i(P, \mathcal{M}, x) = \max_{j\in[n]} \left| B_P^i(\mathcal{M}, x) - B_P^j(\mathcal{M}, x) \right| \leq \alpha,$$

Tran et al. IJCAI-21

Fair Allotments

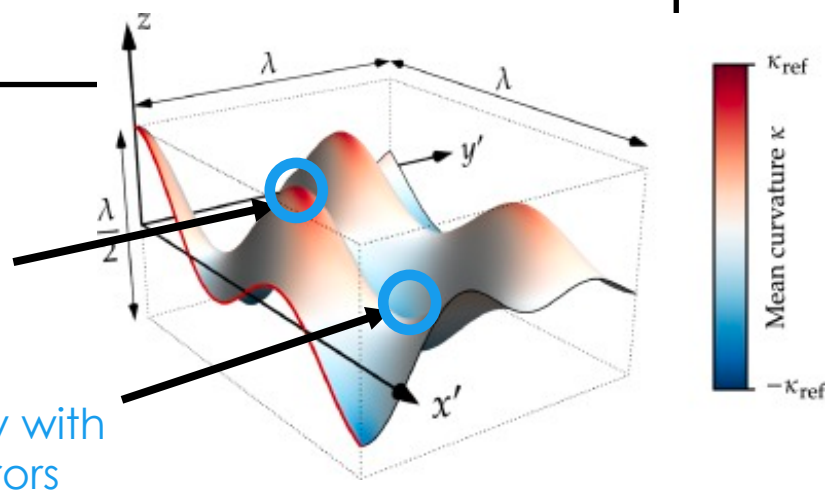
- First (surprising) result:
  Even with an unbiased DP mechanism, the "shape" of the decision problem characterizes the unfairness of the outcomes.

Theorem 3. *Let P be an allotment problem that is at least twice differentiable. A data-release mechanism $\mathcal{M}$ is α-fair w.r.t. P, for some finite α, if for all datasets $\boldsymbol{x} \in \mathcal{X}$ the entries of the Hessian $\boldsymbol{HP}_i$ of problem P_i are a constant function, that is, if there exists $c^i_{jl} \in \mathbb{R}$ $(i \in [n], j, l \in [k])$ such that,*

Hessian of problem Pi → $(\boldsymbol{HP}_i)_{j,l}(\boldsymbol{x}) = c^i_{j,l} \quad (i \in [n]\ j, l \in [k]).$

Corollary 1 (informal). (Perfect)-fairness cannot be achieved if P is any non-convex function, as in the case of the allocations considered.

county with high errors



county with low errors

Adding Laplace noise to the inputs will necessarily introduce fairness issues, despite the noise being unbiased!

Tran et al. IJCAI-21

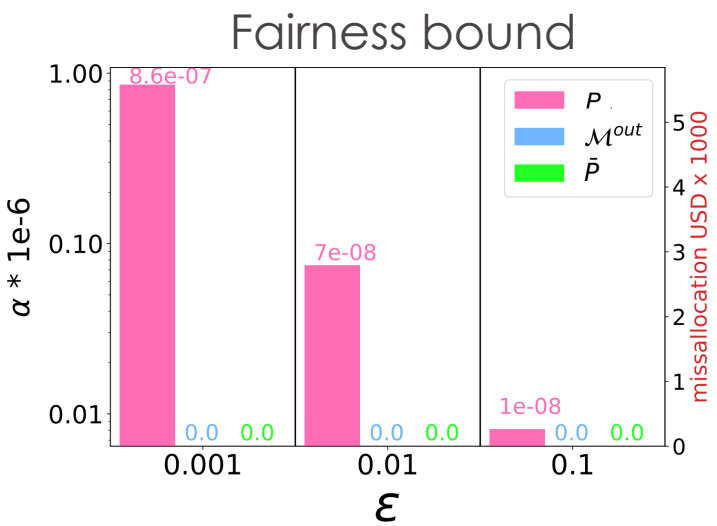
Fair Allotments (mitigation)

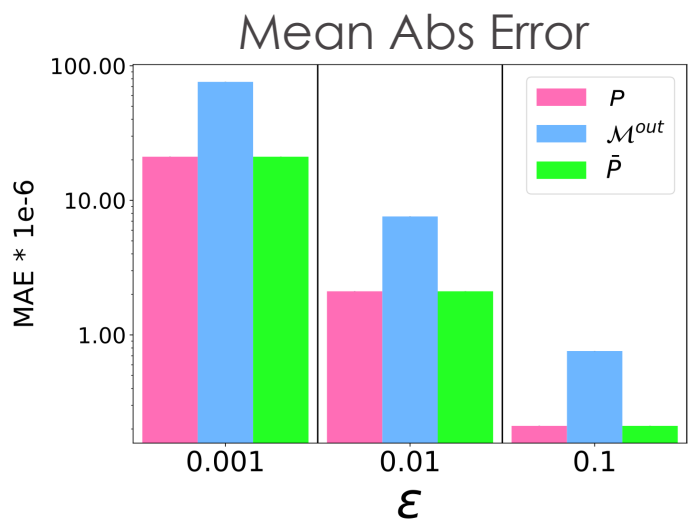
Corollary 2 (informal). If P is a linear function, then mechanism M is fair w.r.t. P

- **Note**: The observed issues are not data-driven, but problem-driven.
- Linearizing the allotment problem — General idea: Given a problem P_i derive a linear approximation $\tilde{P}_i$ of P_i .For example, using a redundant data release:

$$P_i^F(\boldsymbol{x}) \stackrel{\text{def}}{=} \left(\frac{x_i \cdot a_i}{\sum_{i \in [n]} x_i \cdot a_i} \right)$$

release its (noisy) version as a constant





Tran et al. IJCAI-21

Disproportionate Effects in Minority Language Voting Rights

- The *Voting Rights Act* of 1965 provides a body of protections for racial and language minorities.
- Section 203 describes the conditions under which local jurisdictions must provide minority language voting assistance during an election.
- Jurisdiction *i* must provide language assistance (including voter registration, ballots, and instructions) *iff* decision rule $P(\mathbf{x})$ returns *true* with

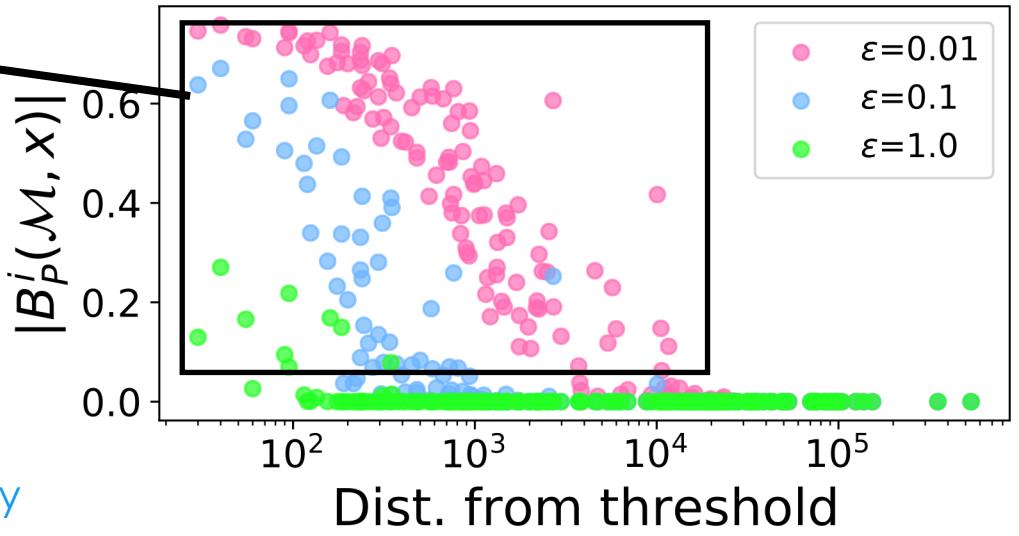
+ < 5th grade education

Misclassification implies potentially disenfranchising

$$P_i^M(\boldsymbol{x}) \stackrel{\text{def}}{=} \left(\frac{x_i^{sp}}{x_i^s} > 0.05 \vee x_i^{sp} > 10^4 \right) \wedge \frac{x_i^{spe}}{x_i^{sp}} > 0.0131.$$

no. of ppl in i speaking minority language s

+ limited English proficiency



Fioretto:IJCAI-21

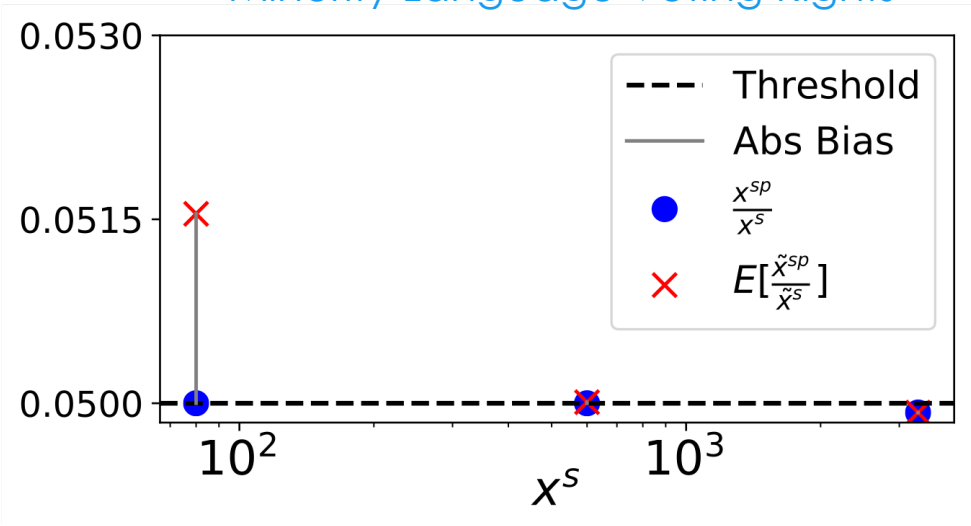
Fair Decision Rules

Ratio Functions

$$P_i^M(\boldsymbol{x}) \stackrel{\text{def}}{=} \left(\frac{x_i^{sp}}{x_i^s} > 0.05 \vee x_i^{sp} > 10^4 \right) \wedge \frac{x_i^{spe}}{x_i^{sp}} > 0.0131.$$

- *Loving county, TX, where* $\frac{\mathrm{x^{sp}}}{\mathrm{x^s}} = 0.05$
- *Terrell county, TX, where* $\frac{\mathrm{x^{sp}}}{\mathrm{x^s}} = 0.05$
- *Union county, NM, where* $\frac{\mathrm{x^{sp}}}{\mathrm{x^s}} = 0.0484$

Minority Language Voting Rights



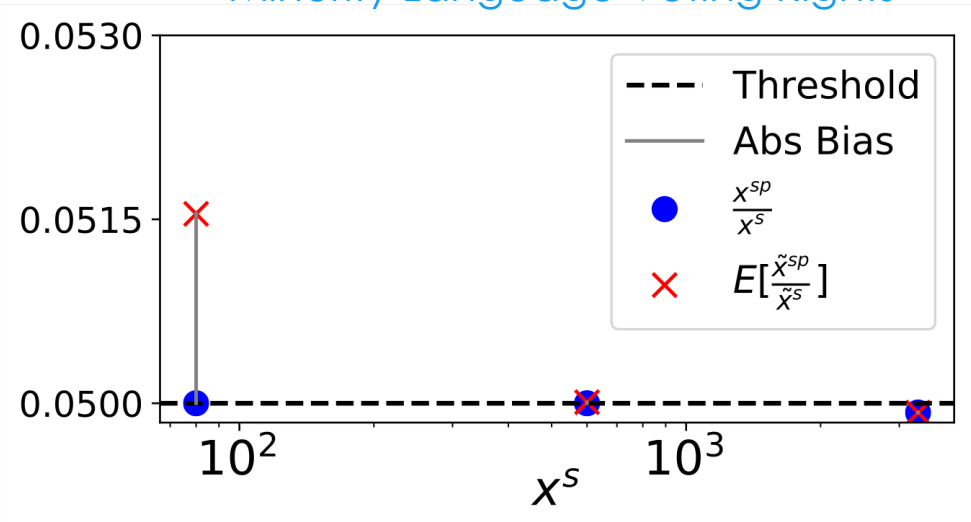
Fair Decision Rules

Ratio Functions

$$P_i^M(\boldsymbol{x}) \stackrel{\text{def}}{=} \left(\frac{x_i^{sp}}{x_i^s} > 0.05 \vee x_i^{sp} > 10^4 \right) \wedge \frac{x_i^{spe}}{x_i^{sp}} > 0.0131.$$

- *Loving county, TX, where* $\frac{\mathrm{x^{sp}}}{\mathrm{x^s}} = \frac{4}{80} = 0.05$
- *Terrell county, TX, where* $\frac{\mathrm{x^{sp}}}{\mathrm{x^s}} = \frac{30}{600} = 0.05$
- *Union county, NM, where* $\frac{\mathrm{x^{sp}}}{\mathrm{x^s}} = \frac{160}{3305} = 0.0484$

Minority Language Voting Rights



Theorem (informal). The perturbation induced by the DP mechanism affects more the county with lower numerator / denominator.

Fair Decisions Rules (cont.)

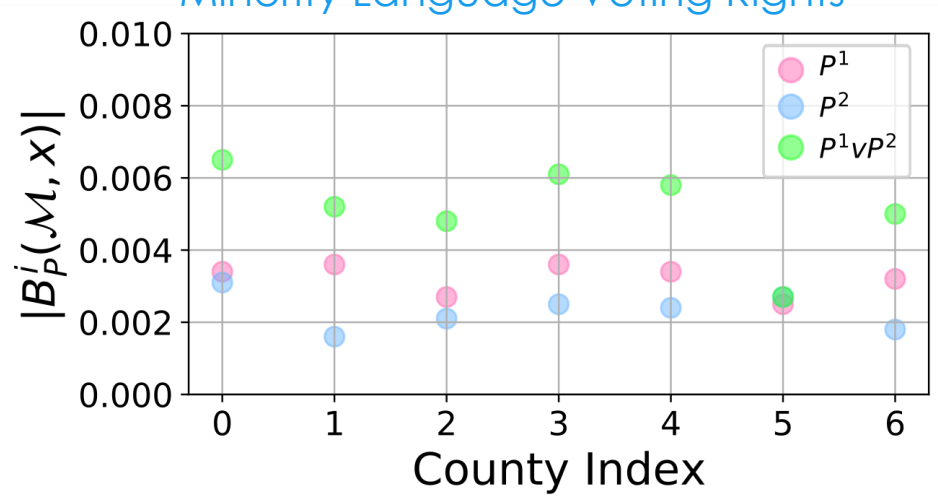
- Second First (surprising) result:
  The unfairness induced by "composing" predicates is larger than that of the individual components.

$$P_i^M(\boldsymbol{x}) \stackrel{\text{def}}{=} \left(\frac{x_i^{sp}}{x_i^s} > 0.05 \vee x_i^{sp} > 10^4 \right) \wedge \frac{x_i^{spe}}{x_i^{sp}} > 0.0131.$$

$$P^1(x^{sp}) = \mathbb{1}\{x^{sp} \geq 10^4\}$$

$$P^2(x^{sp}, x^{spe}) = \mathbb{1}\{\tfrac{x^{spe}}{x^{sp}} > 0.0131\}$$

Minority Language Voting Rights



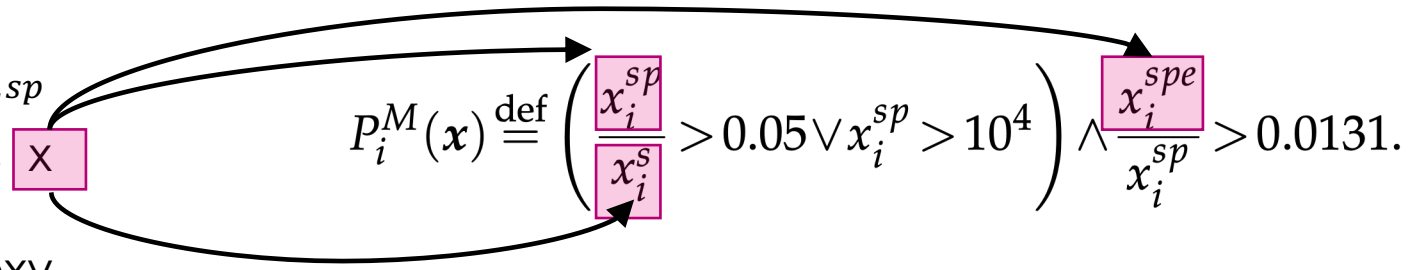
Theorem (informal). The logical composition of two α_1- and α_2-fair mechanisms is α-fair with $\alpha > max(\alpha_1, \alpha_2)$.

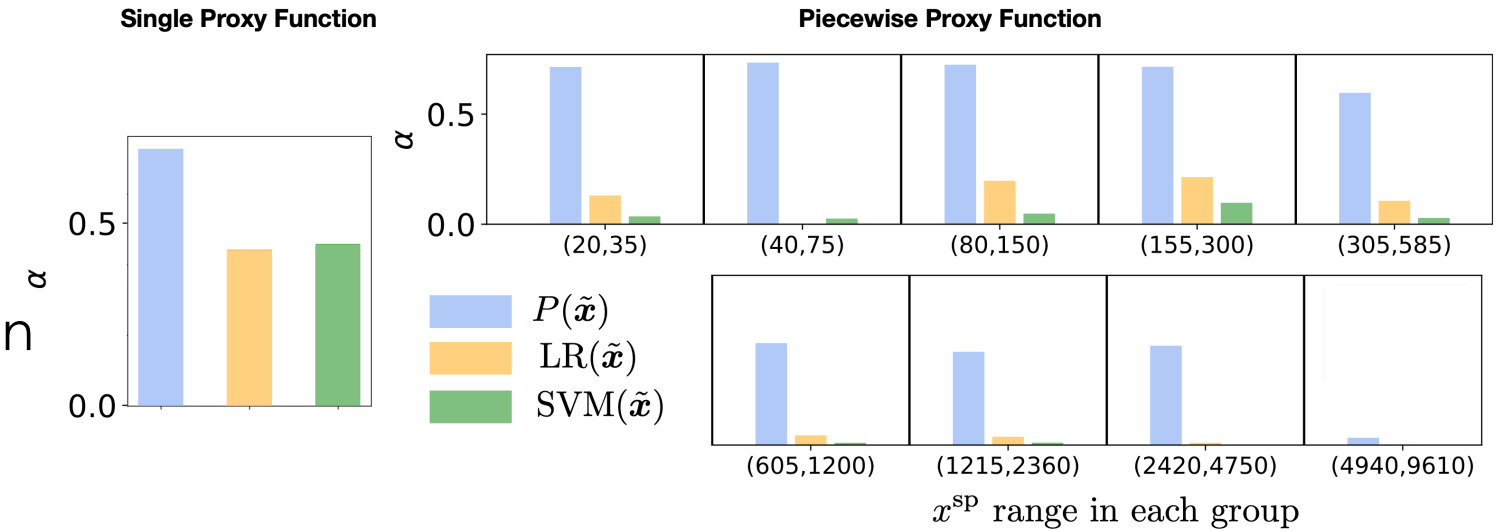
- Small bias when considered individually
- However, when combined using logical connector ∧, the resulting absolute bias increases substantially (green circles).

Fair Decisions (mitigation)

- Much more difficult scenario. But we could resort to the linear approximation trick again

1. Partition dataset into groups x^{sp}
2. Train subgroups using features using a linear classifier
3. Use the parameters of the proxy linear model LR(x) or SVM(x) to make a decision i.e., to approximate P_i^M





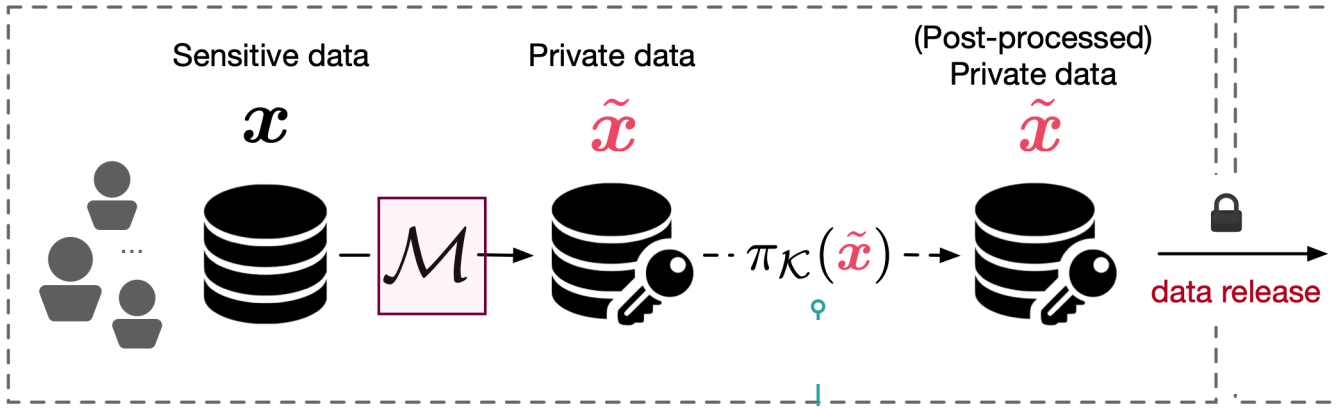
Result summary: Fairness violation decreases substantially, albeit within each subgroup

Why does disparity arise in decision making?

So far:

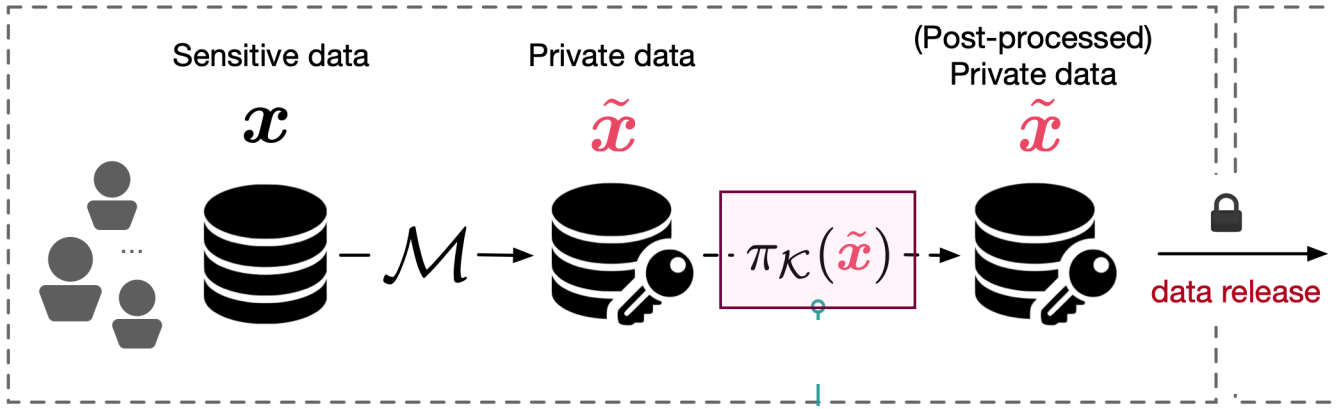
The “shape” of the downstream decision problem

Fairness in DP Downstream Decisions



1. Apply noise with appropriate parameter $\tilde{x} = x + \text{Noise}$

Fairness in DP Downstream Decisions



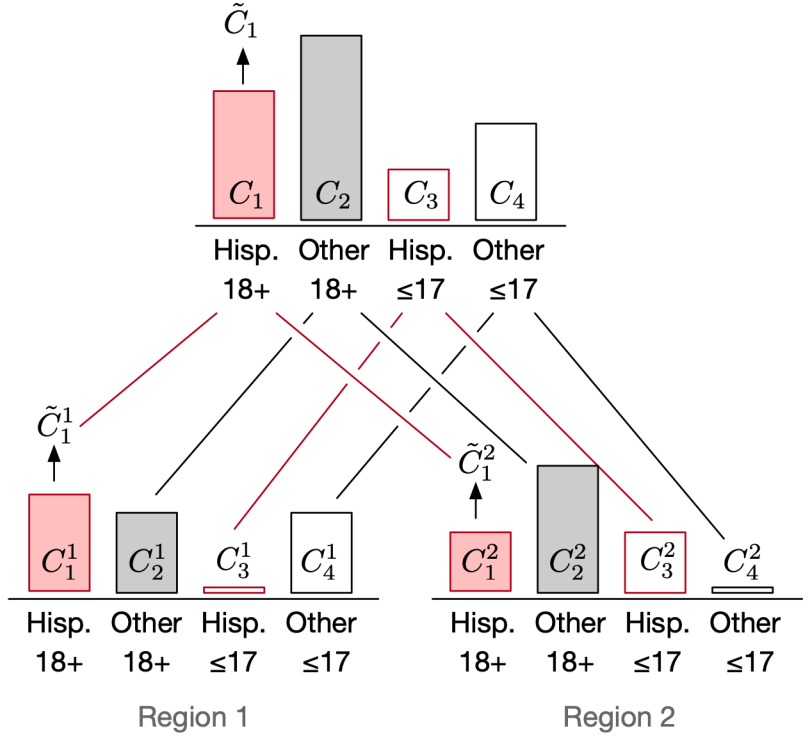
1. Apply noise with appropriate parameter $\tilde{\boldsymbol{x}} = \boldsymbol{x} + \text{Noise}$

2. Postprocess output $\tilde{\boldsymbol{x}}$ to enforce consistency

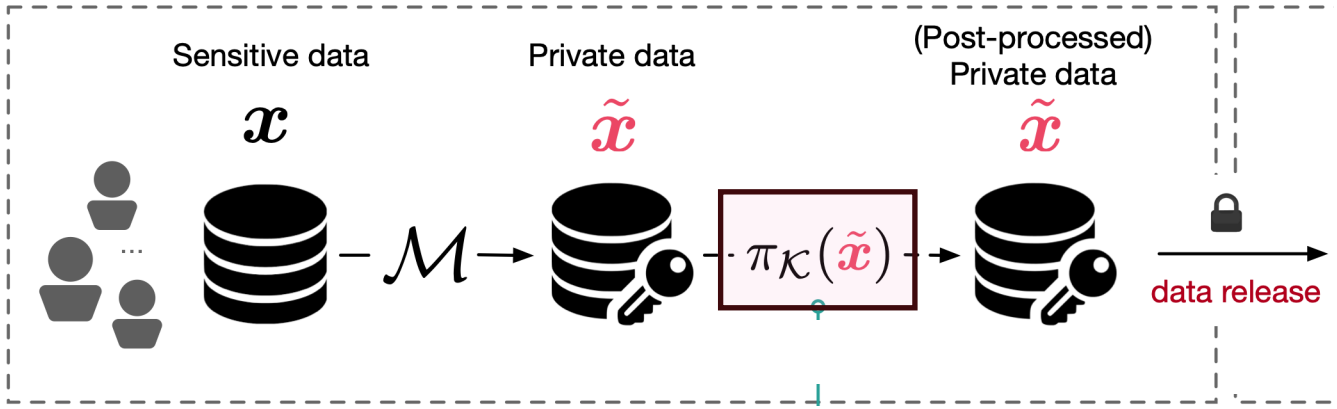
$$\pi_{S+}(\tilde{\boldsymbol{x}}) := \underset{\boldsymbol{v} \in \mathcal{K}_{S+}}{\arg\min} \|\boldsymbol{v} - \tilde{\boldsymbol{x}}\|_2 \,,$$

with feasible region defined as

$$\mathcal{K}_{S+} = \left\{ \boldsymbol{v} \mid \sum_{i=1}^{n} v_i = C \,,\ \boldsymbol{v} \geqslant \boldsymbol{0} \right\}.$$



Fairness in DP Downstream Decisions



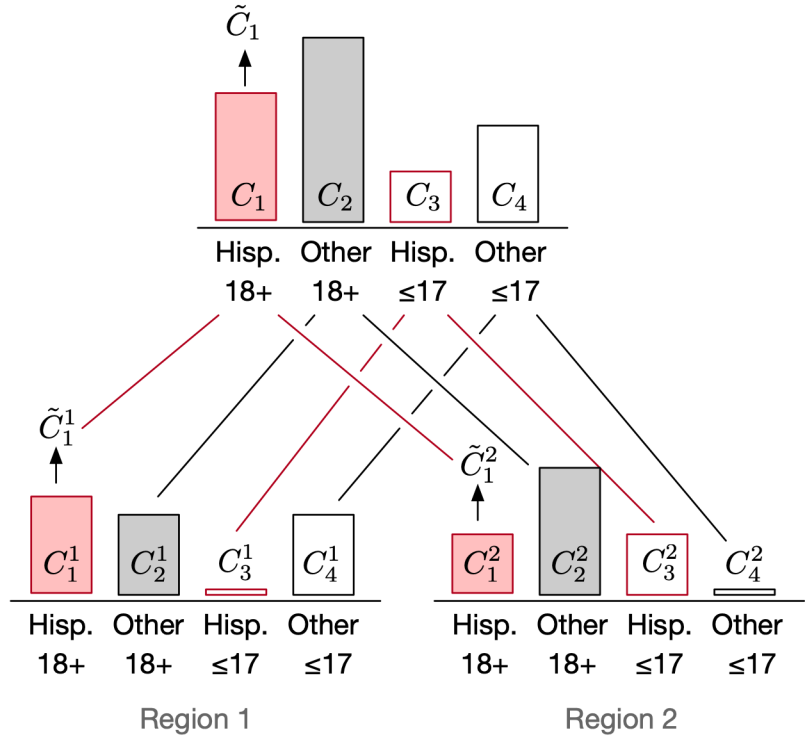
1. Apply noise with appropriate parameter $\tilde{x} = x + \text{Noise}$

2. Postprocess output $\tilde{x}$ to enforce consistency

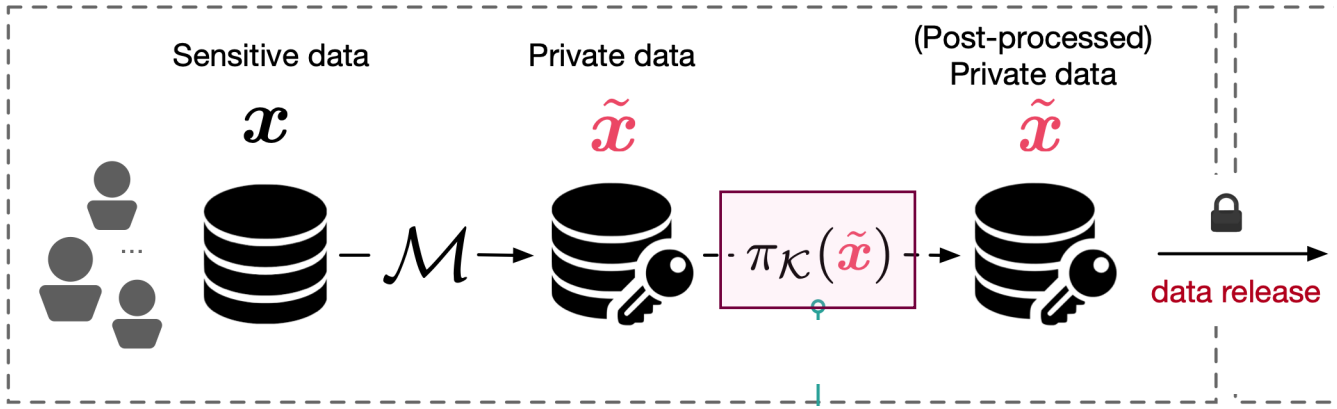
$$\pi_{S+}(\tilde{x}) := \underset{v \in \mathcal{K}_{S+}}{\arg\min} \|v - \tilde{x}\|_2 ,$$

with feasible region defined as

$$\mathcal{K}_{S+} = \left\{ v \mid \sum_{i=1}^{n} v_i = C , \; v \geqslant 0 \right\}.$$



Fairness in DP Downstream Decisions



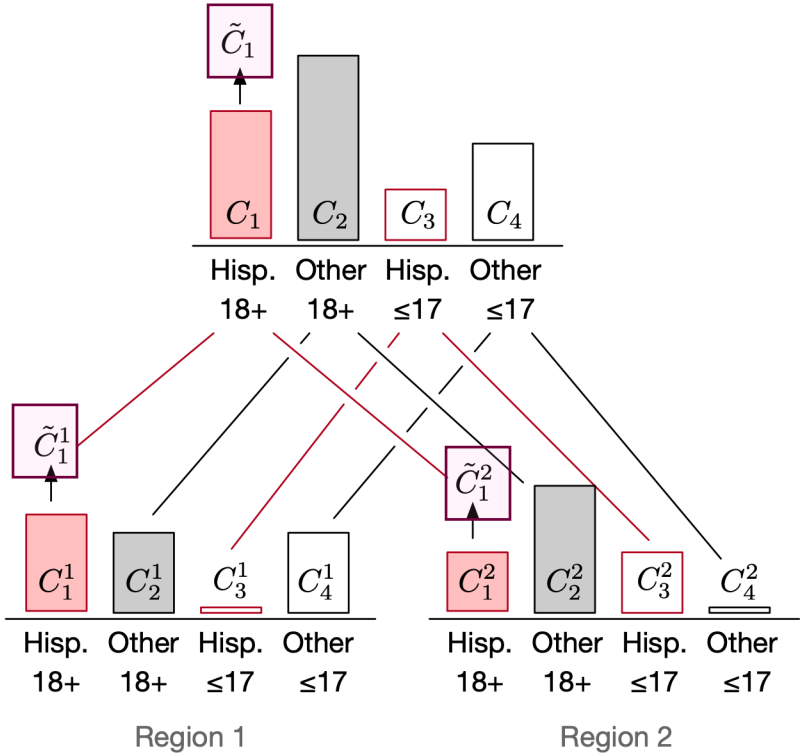
1. Apply noise with appropriate parameter $\tilde{\boldsymbol{x}} = \boldsymbol{x} + \text{Noise}$

2. Postprocess output $\tilde{\boldsymbol{x}}$ to enforce consistency

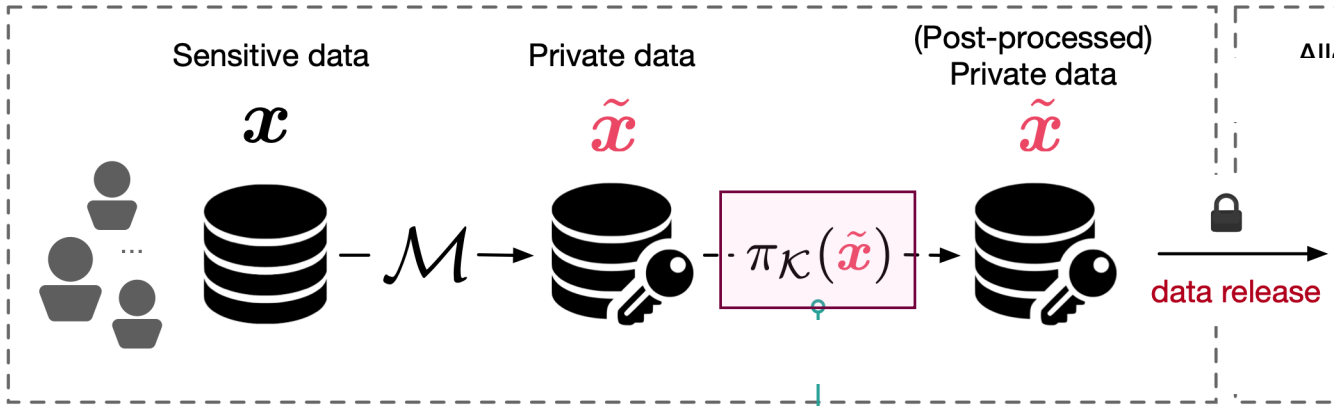
$$\pi_{S+}(\tilde{\boldsymbol{x}}) := \underset{\boldsymbol{v} \in \mathcal{K}_{S+}}{\arg\min} \ \|\boldsymbol{v} - \tilde{\boldsymbol{x}}\|_2 \ ,$$

with feasible region defined as

$$\mathcal{K}_{S+} = \left\{ \boldsymbol{v} \mid \sum_{i=1}^{n} v_i = C \, , \ \boldsymbol{v} \geqslant \mathbf{0} \right\}.$$



Fairness in DP Downstream Decisions



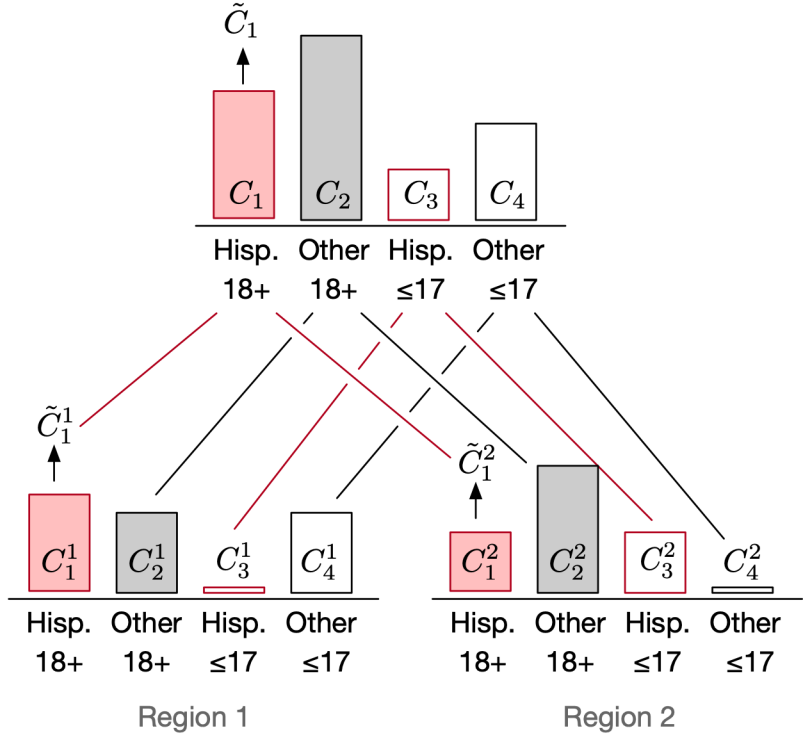
1. Apply noise with appropriate parameter $\tilde{\boldsymbol{x}} = \boldsymbol{x} + \text{Noise}$

2. Postprocess output $\tilde{\boldsymbol{x}}$ to enforce consistency

$$\pi_{S+}(\tilde{\boldsymbol{x}}) := \underset{\boldsymbol{v} \in \mathcal{K}_{S+}}{\arg\min} \; \|\boldsymbol{v} - \tilde{\boldsymbol{x}}\|_2 \;,$$

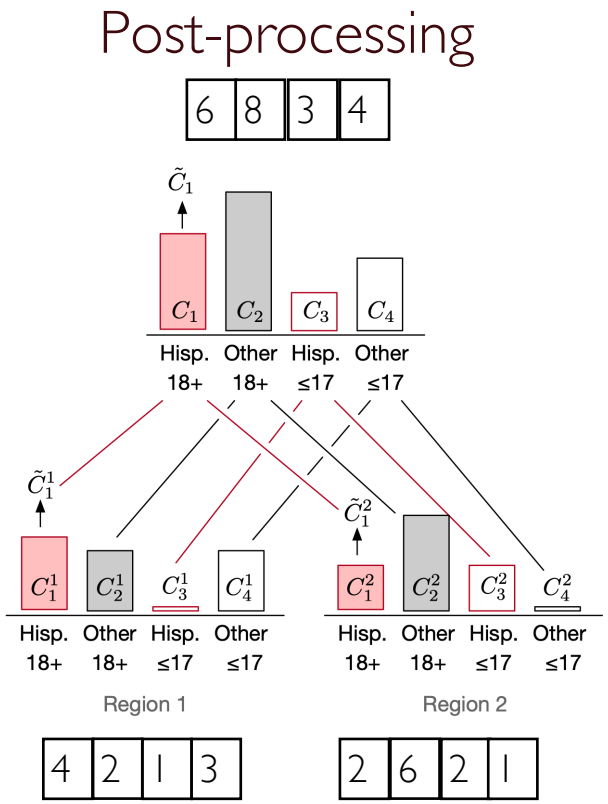
with feasible region defined as

$$\mathcal{K}_{S+} = \left\{ \boldsymbol{v} \mid \sum_{i=1}^{n} v_i = C \,, \; \boldsymbol{v} \geqslant \mathbf{0} \right\}.$$



Satisfies DP due to post-processing immunity

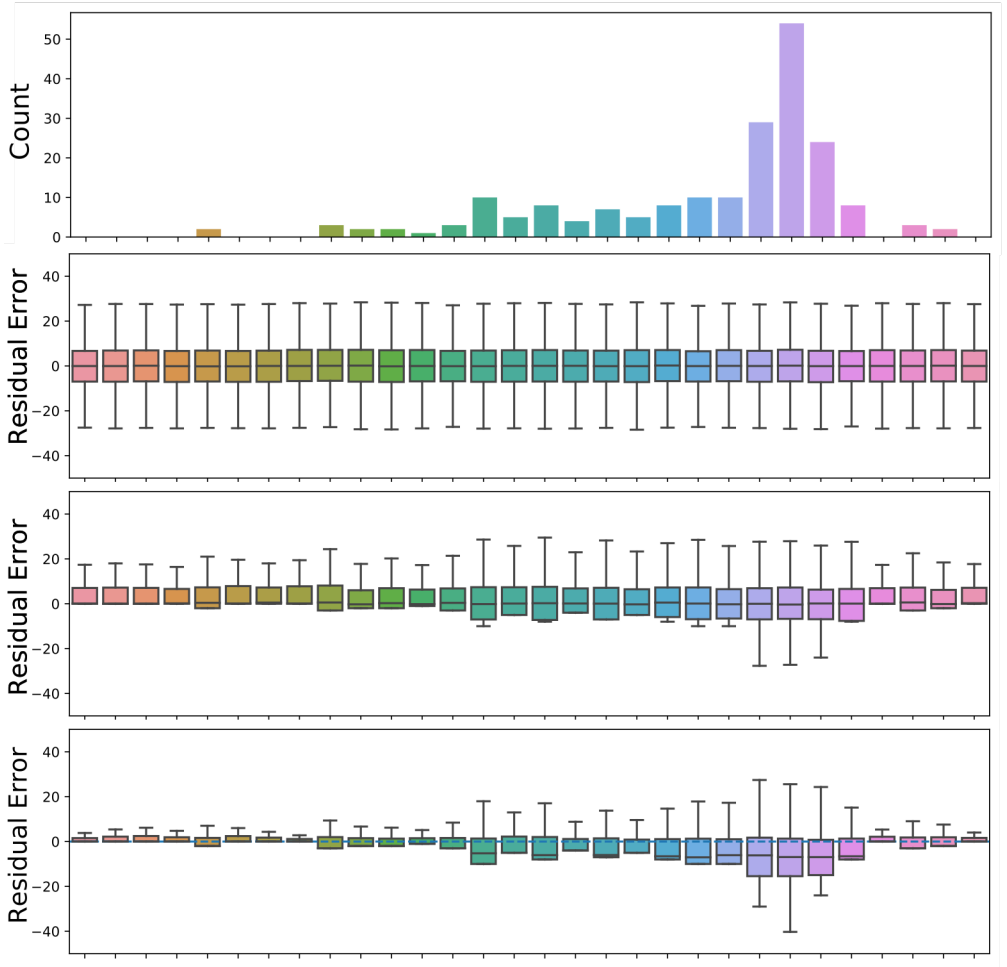
DP Post-Processing



Laplace mechanism

$$\pi_{\geq 0} := \underset{v \geq 0}{\operatorname{argmin}} \|v - \tilde{x}\|_2$$

$$\pi_{\mathcal{K}_S} := \underset{v \in \mathcal{K}_S}{\operatorname{argmin}} \|v - \tilde{x}\|_2 \,,\ \mathcal{K}_S = \{v \in \mathbb{R}^n \mid \sum_i v_i = \tilde{S}, v_i \geq 0\},$$



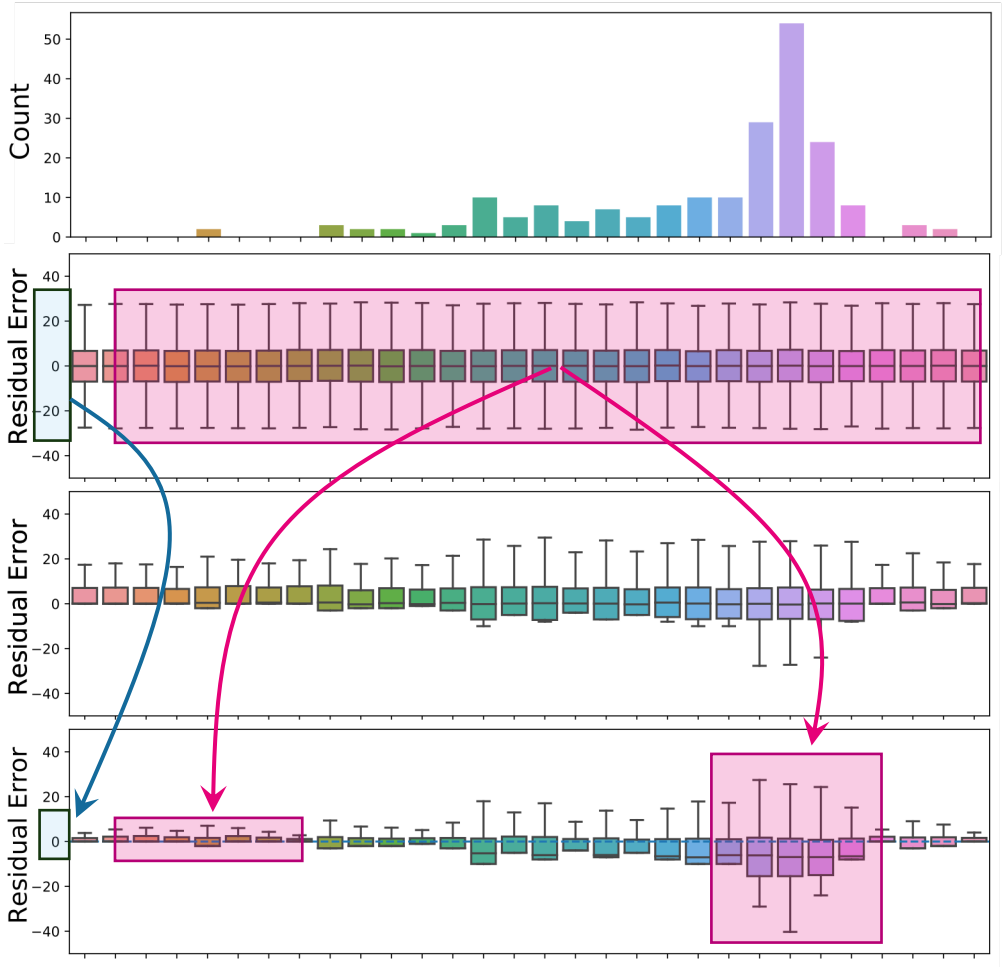
DP Post-Processing

Third result: Observe that post-processing reduce the errors;

However, it increases unfairness! Laplace mechanism

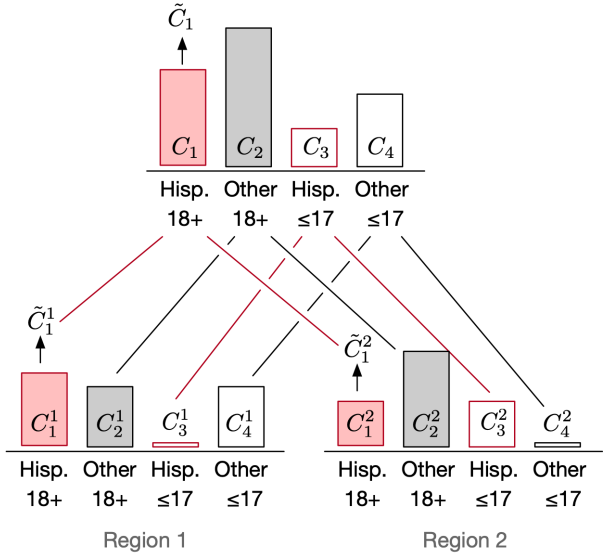
$$\pi_{\geq 0} := \underset{v \geq 0}{\operatorname{argmin}} \|v - \tilde{x}\|_2$$

$$\pi_{\mathcal{K}_S} := \underset{v \in \mathcal{K}_S}{\operatorname{argmin}} \|v - \tilde{x}\|_2 \,,\ \mathcal{K}_S = \{v \in \mathbb{R}^n \mid \sum_i v_i = \tilde{S}, v_i \geq 0\},$$



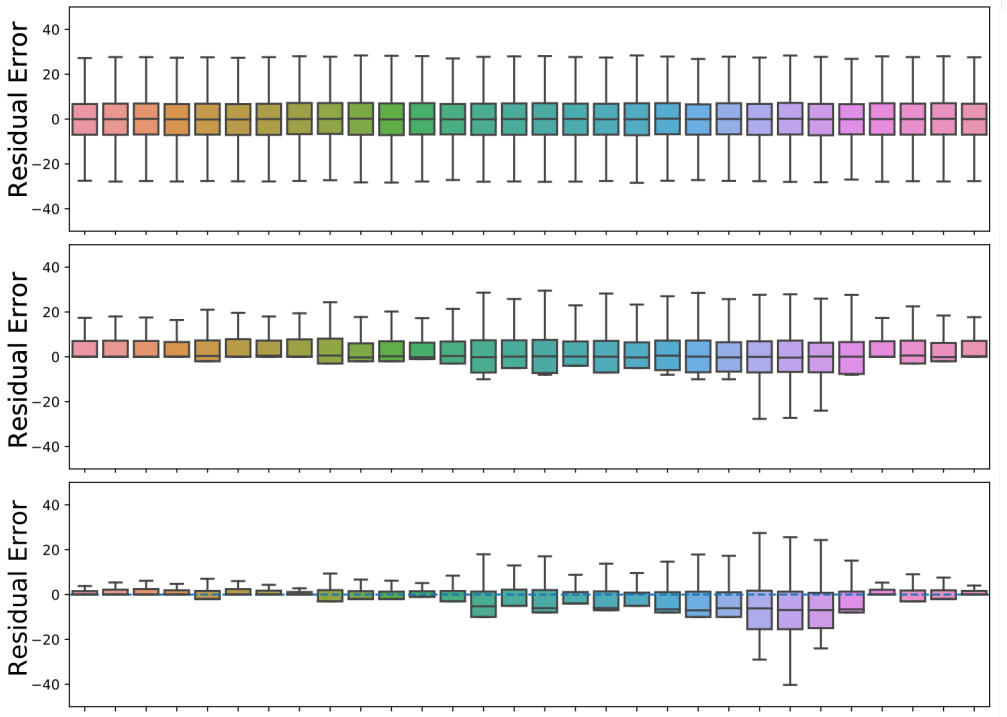
DP Post-Processing

Why is this happening? The bias comes from the non-negativity constraints!



$$\pi_{\geq 0} := \underset{v \geq 0}{\operatorname{argmin}} \|v - \tilde{x}\|_2$$

$$\pi_{\mathcal{K}_S} := \underset{v \in \mathcal{K}_S}{\operatorname{argmin}} \|v - \tilde{x}\|_2 \,,\; \mathcal{K}_S = \{v \in \mathbb{R}^n \mid \sum_i v_i = \tilde{S}, v_i \geq 0\},$$

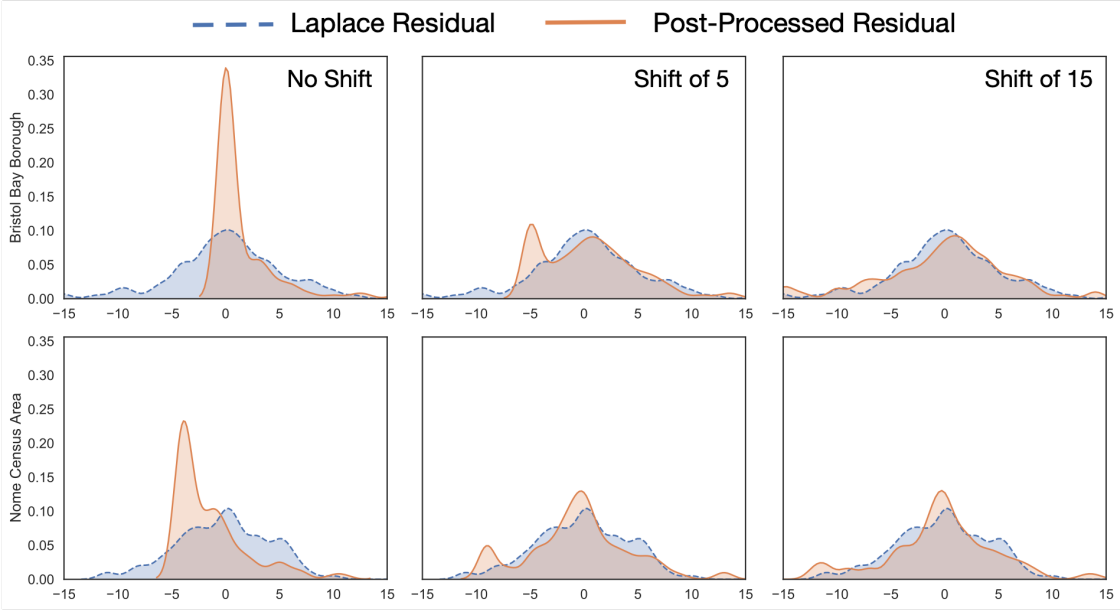


Quantifying Bias in DP Post-Processing

Theorem 6. *Suppose that the noisy data $\tilde{\boldsymbol{x}}$ is the output of the Laplace mechanism with scale λ. The bias of the post-processed solution $\pi_{\mathcal{K}^+}$ of program (L^+) is bounded, in l_∞ norm, by*

$$\|B_{L^+}(\mathcal{M}, \boldsymbol{x})\|_\infty = \left\|\mathbb{E}_{\tilde{\boldsymbol{x}}\sim\mathcal{M}(\boldsymbol{x})}\left[\pi_{L^+}(\tilde{\boldsymbol{x}}) - \boldsymbol{x}\right]\right\|_\infty \leq C' \cdot \exp\left(\frac{-r_m}{\lambda}\right) \cdot \sum_{i=0}^{n-1} \frac{(r_m)^i}{i! \cdot \lambda^i},$$

where C' represents the value $\sup_{v\in\mathcal{K}^+} \|v - \boldsymbol{x}\|_\infty$, which is finite due to the boundedness of the feasible region $\mathcal{K}^+$.



There is a L1-ball of radius $r_m = \min_i x_i$ and centered in x which is a feasible subspace where there is no bias

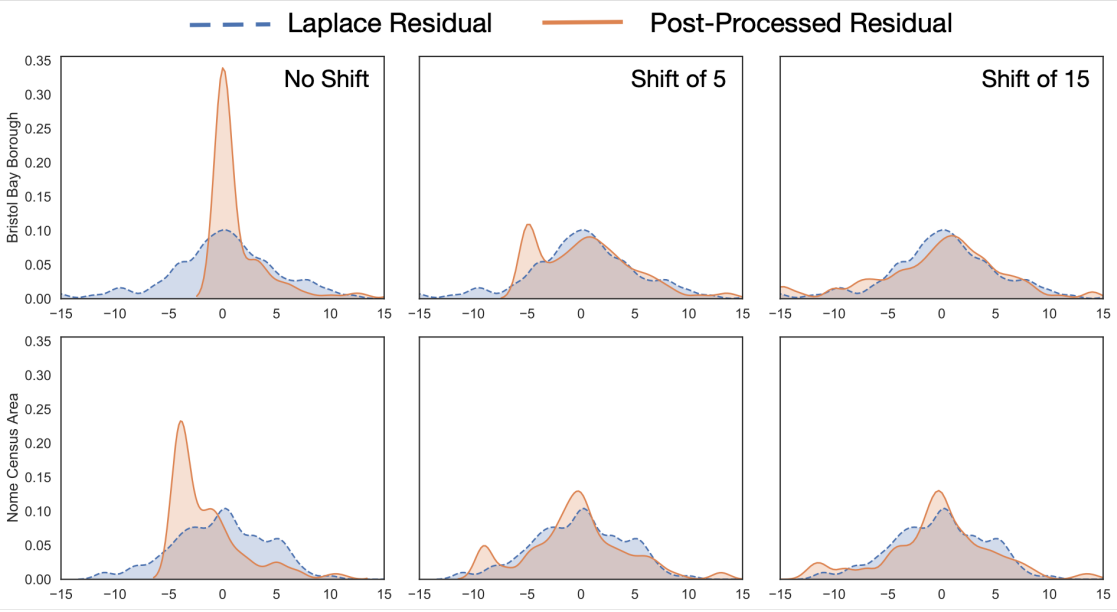
Shifting increases the value of r_m and the bias progressively disappear

Zhu et al. AAAI-21

Quantifying Bias in DP Post-Processing

- What does this means practically? Post-processing reduces the variance of the noise differently in different "regions". Regions with many subregions (e.g., counties, census blocks, etc.) will have more variance than regions with few subregions.

- It creates situations where counties will be treated fundamentally differently in decision processes.



Aggregating the counts for	Variance
Arizona (pop: 2.37ML in15 counties)	186.67
Texas (pop: 8.89ML in 254 counties)	200.01

~6.5% difference
which may affect allocations!

Zhu et al. AAAI-21

Possible Mitigating Solution

Definition 4 (Projection onto Simplex Mechanism (PoS)). *The* projection onto simplex mechanism *outputs the allocation as follows.*

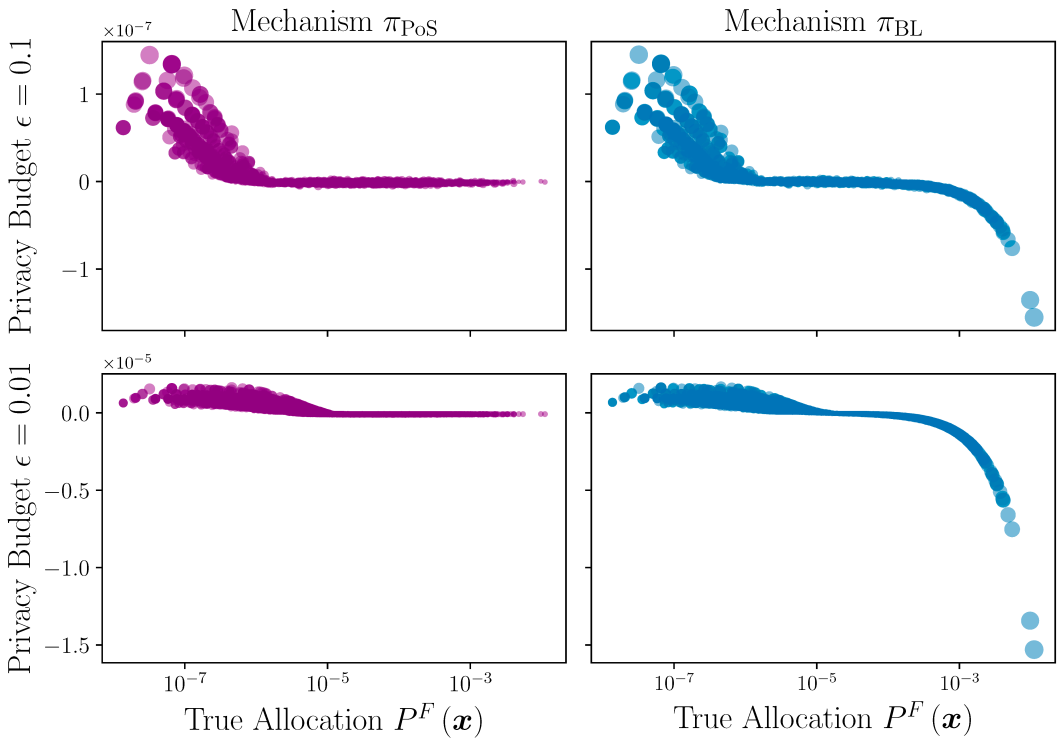
$$\pi_{\mathrm{PoS}}\left(\tilde{\boldsymbol{x}}\right) := \underset{\boldsymbol{v}\in\Delta^n}{\arg\min} \left\|\boldsymbol{v} - P^F\left(\tilde{\boldsymbol{x}}\right)\right\|_2 \qquad (P_{\mathrm{PoS}})$$

Theorem (informal). For any DP dataset $\tilde{\boldsymbol{x}}$ the PoS mechanism generates the unique optimal solution to program

$$\pi^*_\alpha(\tilde{\boldsymbol{x}}) := \underset{\boldsymbol{v}\in\Delta_n}{\arg\min} \left\|\boldsymbol{v} - P^F\left(\tilde{\boldsymbol{x}}\right)\right\|_{\rightleftharpoons} \qquad (P_\alpha)$$

which closely approximate the optimal post-processing mechanism

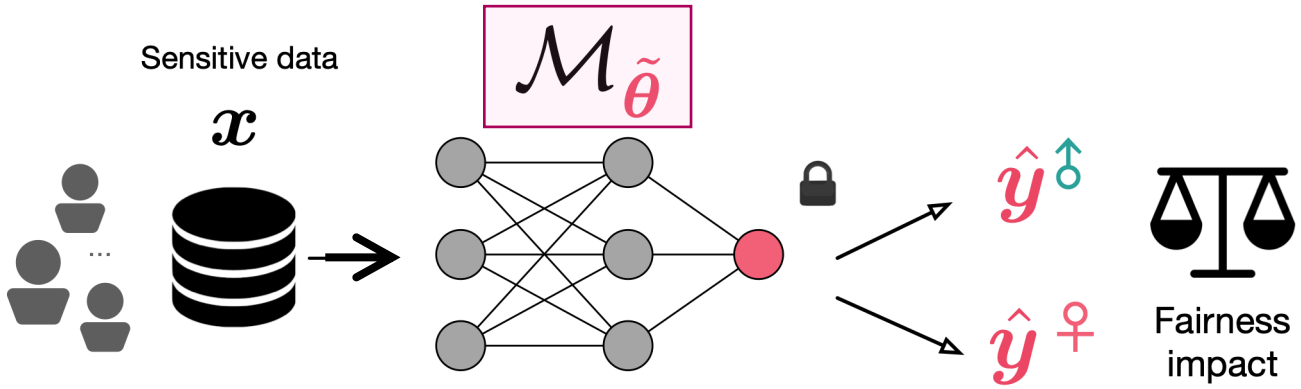
$$\pi^* := \min_{\pi\in\Pi_{\Delta_n}} \left\|\mathbb{E}_{\tilde{\boldsymbol{x}}}\left[\pi(\tilde{\boldsymbol{x}}) - P^F\left(\boldsymbol{x}\right)\right]\right\|_{\rightleftharpoons}, \qquad (4)$$



Zhu et al. AAAI-22

Why disparity arise in learning tasks?

Fairness in DP Learning Tasks



Given a dataset consisting of data points (X_i, A_i, Y_i) the goal is to learn a classifier f_θ that guarantees privacy of the individual data points and the learning task minimizes

$$\min_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}; D) = \frac{1}{n} \sum_{i=1}^{n} \ell(f_{\boldsymbol{\theta}}(X_i), Y_i).$$

Fairness focuses on the notion of excessive risk: $R(\boldsymbol{\theta}, D) = \mathbb{E}_{\tilde{\boldsymbol{\theta}}} \left[\mathcal{L}(\tilde{\boldsymbol{\theta}}; D) \right] - \mathcal{L}(\boldsymbol{\theta}^*; D)$, and is measured with respect to the excessive risk gap

$$\xi_a = |R_a(\boldsymbol{\theta}) - R(\boldsymbol{\theta})|.$$

group-level ER

Population-level ER

Tran et al. NeurIPS-21, ArXiv-22

Warm up: Output Perturbation

Adds Gaussian noise $\mathcal{N}(0, \Delta_\ell^2 \sigma^2)$ to the optimal model parameters θ^*.

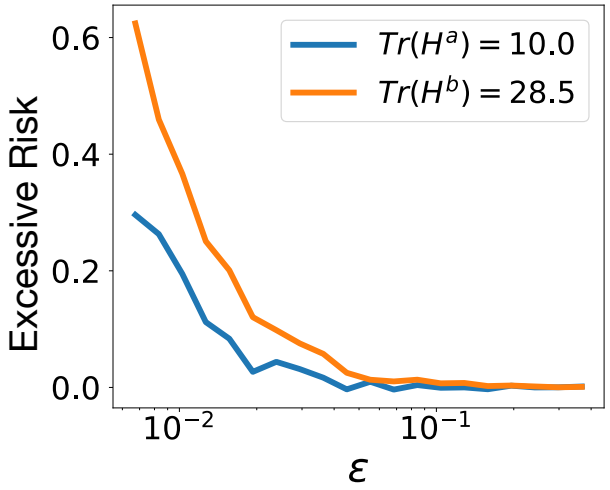
Theorem 1. *Let ℓ be a twice differentiable and convex loss function and consider the output perturbation mechanism described above. Then, the excessive risk gap for group $a \in \mathcal{A}$ is approximated by:*

$$\xi_a \approx \frac{1}{2}\Delta_\ell^2 \sigma^2 \left|\mathrm{Tr}(\boldsymbol{H}_\ell^a) - \mathrm{Tr}(\boldsymbol{H}_\ell)\right|, \tag{3}$$

where $\boldsymbol{H}_\ell^a = \nabla^2_{\boldsymbol{\theta}^*} \sum_{(X,A,Y)\in D_a} \ell(f_{\boldsymbol{\theta}^*}(X), Y)$ is the Hessian matrix of the loss function, at the optimal parameters vector $\boldsymbol{\theta}^*$, computed using the group data D_a, $\boldsymbol{H}_\ell$ is the analogous Hessian computed using the population data D, and $\mathrm{Tr}(\cdot)$ denotes the trace of a matrix.

Corollary 1. Consider the ERM problem for a linear model $f_\theta(X) = \theta^T X$, with L_2 loss function. Then, output perturbation does not guarantee pure fairness.

$$Tr(\boldsymbol{H}_\ell^{\boldsymbol{a}}) = \mathbb{E}_{X\sim D_a} Tr(XX^T) = \mathbb{E}_{X\sim D_a} ||X||^2$$



Tran et al. NeurIPS-21

Warm up: Output Perturbation

Adds Gaussian noise $\mathcal{N}(0, \Delta_\ell^2 \sigma^2)$ to the optimal model parameters θ^*.

Theorem 1. *Let ℓ be a twice differentiable and convex loss function and consider the output perturbation mechanism described above. Then, the excessive risk gap for group $a \in \mathcal{A}$ is approximated by:*

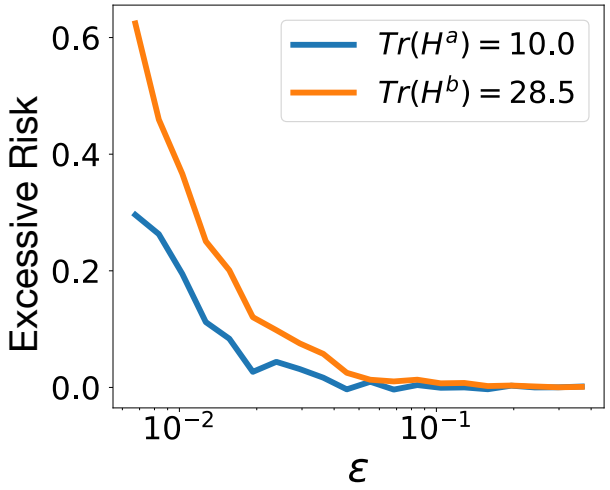
$$\xi_a \approx \frac{1}{2}\Delta_\ell^2 \sigma^2 \left|\mathrm{Tr}(\boldsymbol{H}_\ell^a) - \mathrm{Tr}(\boldsymbol{H}_\ell)\right|, \tag{3}$$

where $\boldsymbol{H}_\ell^a = \nabla^2_{\boldsymbol{\theta}^*} \sum_{(X,A,Y)\in D_a} \ell(f_{\boldsymbol{\theta}^*}(X), Y)$ is the Hessian matrix of the loss function, at the optimal parameters vector $\boldsymbol{\theta}^*$, computed using the group data D_a, $\boldsymbol{H}_\ell$ is the analogous Hessian computed using the population data D, and $\mathrm{Tr}(\cdot)$ denotes the trace of a matrix.

Corollary 2. For groups a and b, if their average group norms

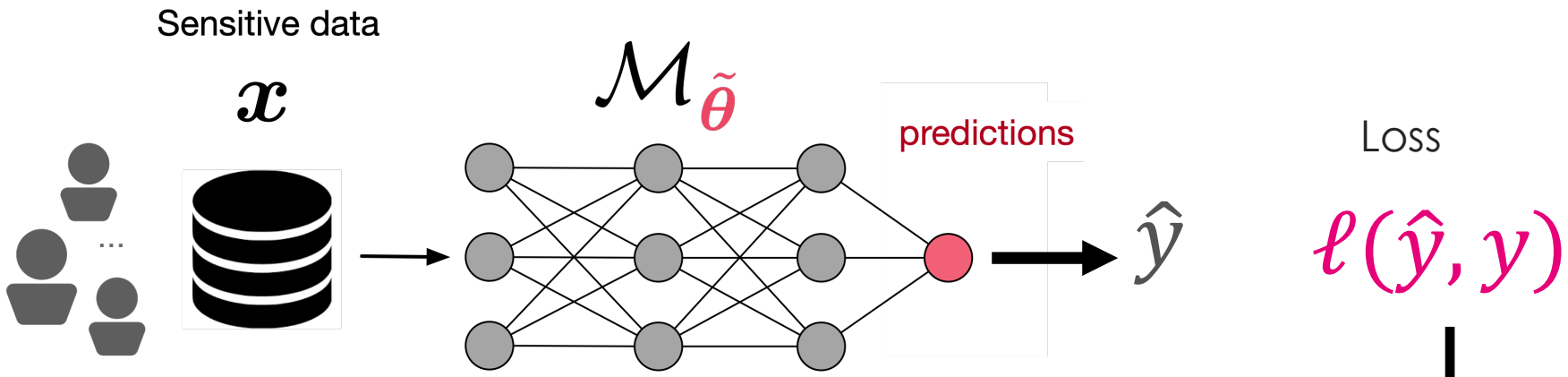
$$\mathbb{E}_{X_a \sim D_a} \| X_a \| = \mathbb{E}_{X_b \sim D_b} \| X_b \|$$

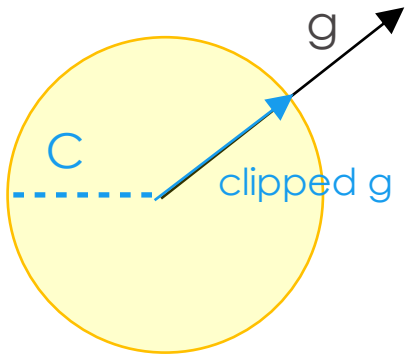
have identical values, then output perturbation with L_2 loss function achieves pure fairness.



Tran et al. NeurIPS-21

DP-Stochastic Gradient Descent





1. Clip g in L2 ball of size C
2. Add noise from $\mathcal{N}(0, IC^2\sigma^2)$
3. Average gradients $\mathbf{g}$'s in a mini-batch
4. Update parameters θ

Fairness in DP-SGD

Theorem 2. *Consider the ERM problem* (L) *with loss ℓ twice differentiable w.r.t. the model parameters. The expected loss $\mathbb{E}[\mathcal{L}(\boldsymbol{\theta}_{t+1}; D_a)]$ of group $a \in \mathcal{A}$ at iteration $t+1$, is approximated as:*

$$\mathbb{E}\left[\mathcal{L}(\boldsymbol{\theta}_{t+1}; D_a)\right] = \underbrace{\mathcal{L}(\boldsymbol{\theta}_t; D_a) - \eta \left\langle \boldsymbol{g}_{D_a}, \boldsymbol{g}_D \right\rangle + \frac{\eta^2}{2} \mathbb{E}\left[\boldsymbol{g}_B^T \boldsymbol{H}_\ell^a \boldsymbol{g}_B\right]}_{\textit{non-private term}} \tag{4}$$

$$+ \underbrace{\eta \left(\left\langle \boldsymbol{g}_{D_a}, \boldsymbol{g}_D \right\rangle - \left\langle \boldsymbol{g}_{D_a}, \bar{\boldsymbol{g}}_D \right\rangle\right) + \frac{\eta^2}{2} \left(\mathbb{E}\left[\bar{\boldsymbol{g}}_B^T \boldsymbol{H}_\ell^a \bar{\boldsymbol{g}}_B\right] - \mathbb{E}\left[\boldsymbol{g}_B^T \boldsymbol{H}_\ell^a \boldsymbol{g}_B\right]\right)}_{\textit{private term due to clipping}} \tag{R_a^{clip}}$$

$$+ \underbrace{\frac{\eta^2}{2} \mathrm{Tr}(\boldsymbol{H}_\ell^a) C^2 \sigma^2}_{\textit{private term due to noise}} \tag{R_a^{noise}}$$

$$+ \; O(\|\boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}_t\|^3),$$

where the expectation is taken over the randomness of the private noise and the mini-batch selection, and the terms $\boldsymbol{g}_Z$ and $\bar{\boldsymbol{g}}_Z$ denote, respectively, the average non-private and private gradients over subset Z of D at iteration t (the iteration number is dropped for ease of notation).

Tran et al. NeurIPS-21

Why clipping causes unfairness in DP-SGD

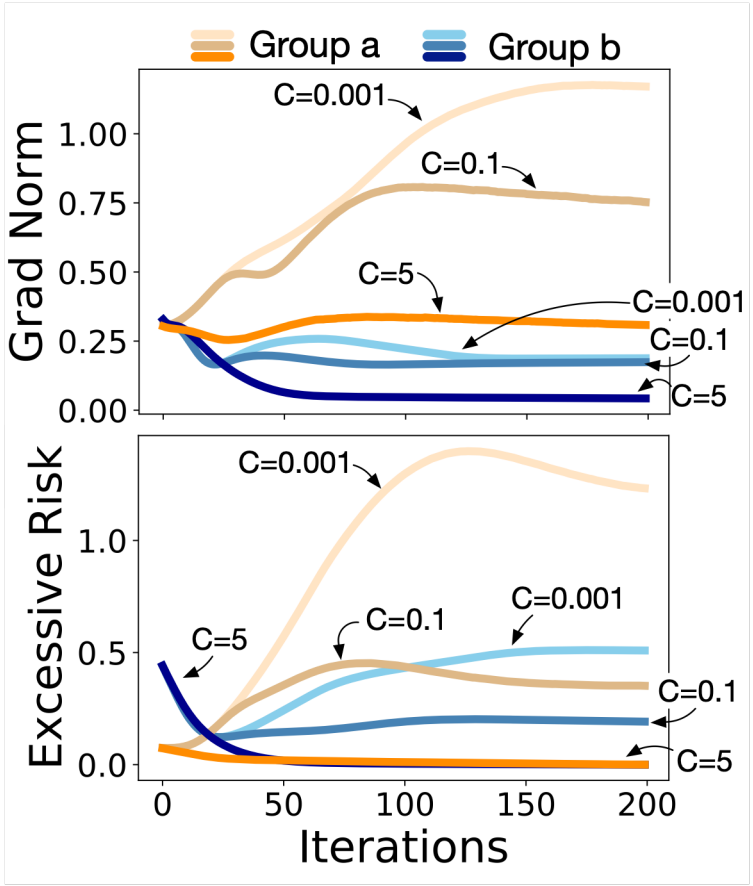
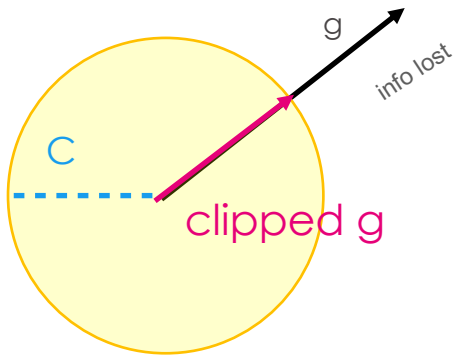
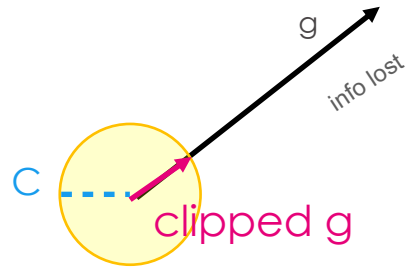


Figure 3: Impact of gradient clipping on gradient norms for different clipping bounds. Bank dataset.

When clipping, the smaller C, the higher
Is the information loss of the average gradients that are backpropagated.



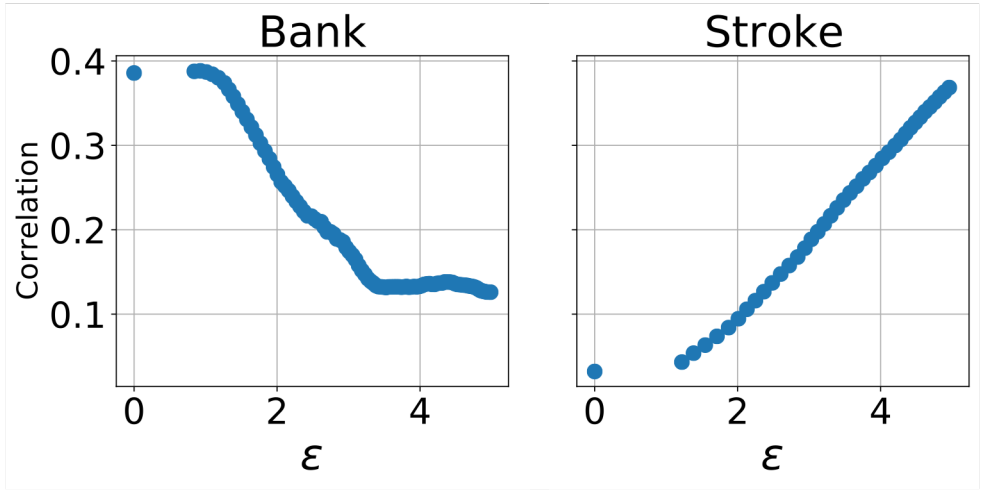


Theorem 3. *Let $p_z = |D_z|/|D|$ be the fraction of training samples in group $z \in \mathcal{A}$. For groups $a, b \in \mathcal{A}$, $R_a^{clip} > R_b^{clip}$ whenever:*

$$\|g_{D_a}\|\left(p_a - \frac{p_a^2}{2}\right) \geq \frac{5}{2}C + \|g_{D_b}\|\left(1 + p_b + \frac{p_b^2}{2}\right). \quad (5)$$

Tran et al. NeurIPS-21

Why clipping causes unfairness in DP-SGD



Correlation between input norms and excessive risk

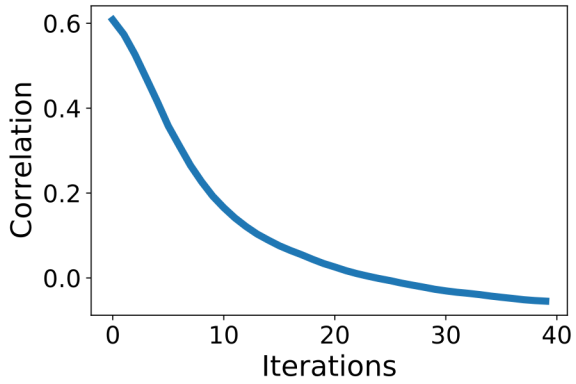


Figure 4: Correlation between inputs and gradients norms.

Crucial proxy to unfairness (due to clipping)

Theorem (informal) Gradients and Input norms greatly affect the excessive risk (unfairness) of the individuals and groups

Individuals with larger features (e.g., salary, age, height) may have larger disparate impacts in classification!

Tran et al. NeurIPS-21

Why noise addition causes unfairness in DP-SGD?

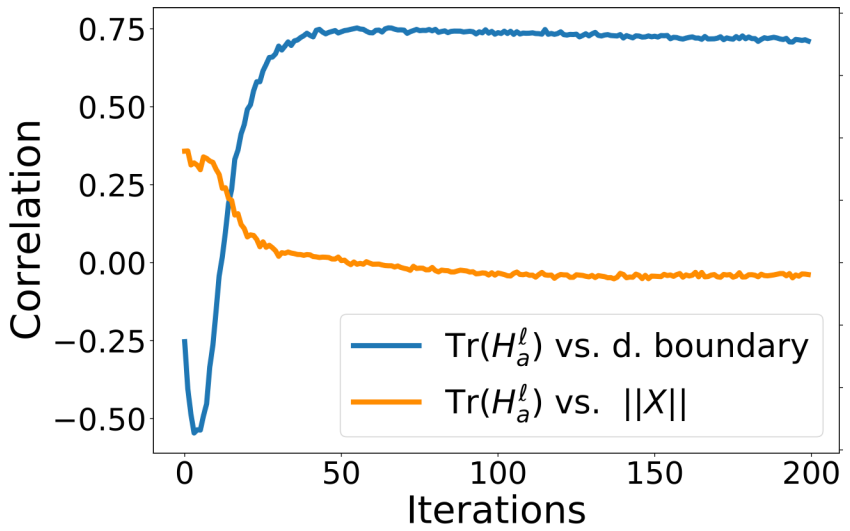


Figure 5: Correlation between trace of Hessian with closeness to boundary (dark color) and input norm (light color).

Crucial proxy to unfairness (due to clipping)

Closeness to the decision boundary!

Theorem (informal). Individuals whose outputs are close to the decision boundary will have higher Hessian Traces (high local curvatures of the loss).

Intuitively, the model decisions for samples which are close to the decision boundary are less robust to the presence of noise w.r.t. samples which are farther away from the boundary

Tran et al. NeurIPS-21

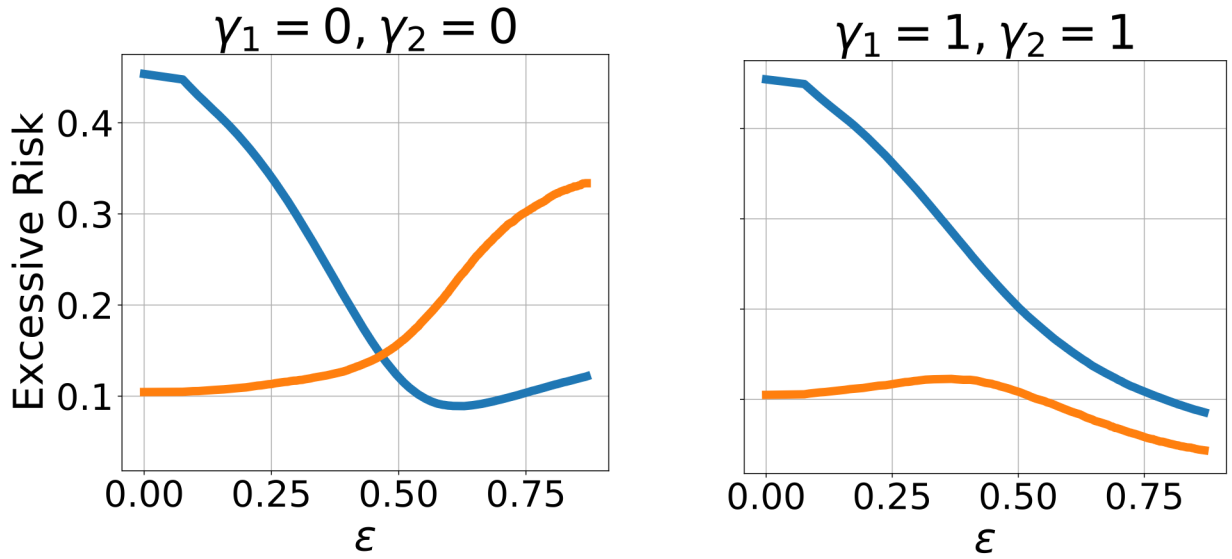
Possible mitigating solution

1. Modify Training so to equalize the factors affecting the excessive risk due to noise and that due to clipping:

$$\min_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}; D) + \sum_{a \in \mathcal{A}} \left(\gamma_1 \left| \langle \boldsymbol{g}_{D_a} - \boldsymbol{g}_D, \boldsymbol{g}_D - \bar{\boldsymbol{g}}_D \rangle \right| + \gamma_2 \left| \mathrm{Tr}(\boldsymbol{H}_\ell^a) - \mathrm{Tr}(\boldsymbol{H}_\ell) \right| \right),$$

Minority group

Majority group



Tran et al. NeurIPS-21

Part III

What are the open scientific challenges?

(Some) Open Questions

- Devise mitigation strategies for data-release mechanisms which account for both fairness and utility.
- Derive more general tools for analyzing fairness for larger classes of decision problems.
- Derive better compositional rules for fairness analysis.
- Analyze the fairness impact of more complex post-processing steps.
- Study the link between Differential Privacy and Robustness under a Fairness Lens.

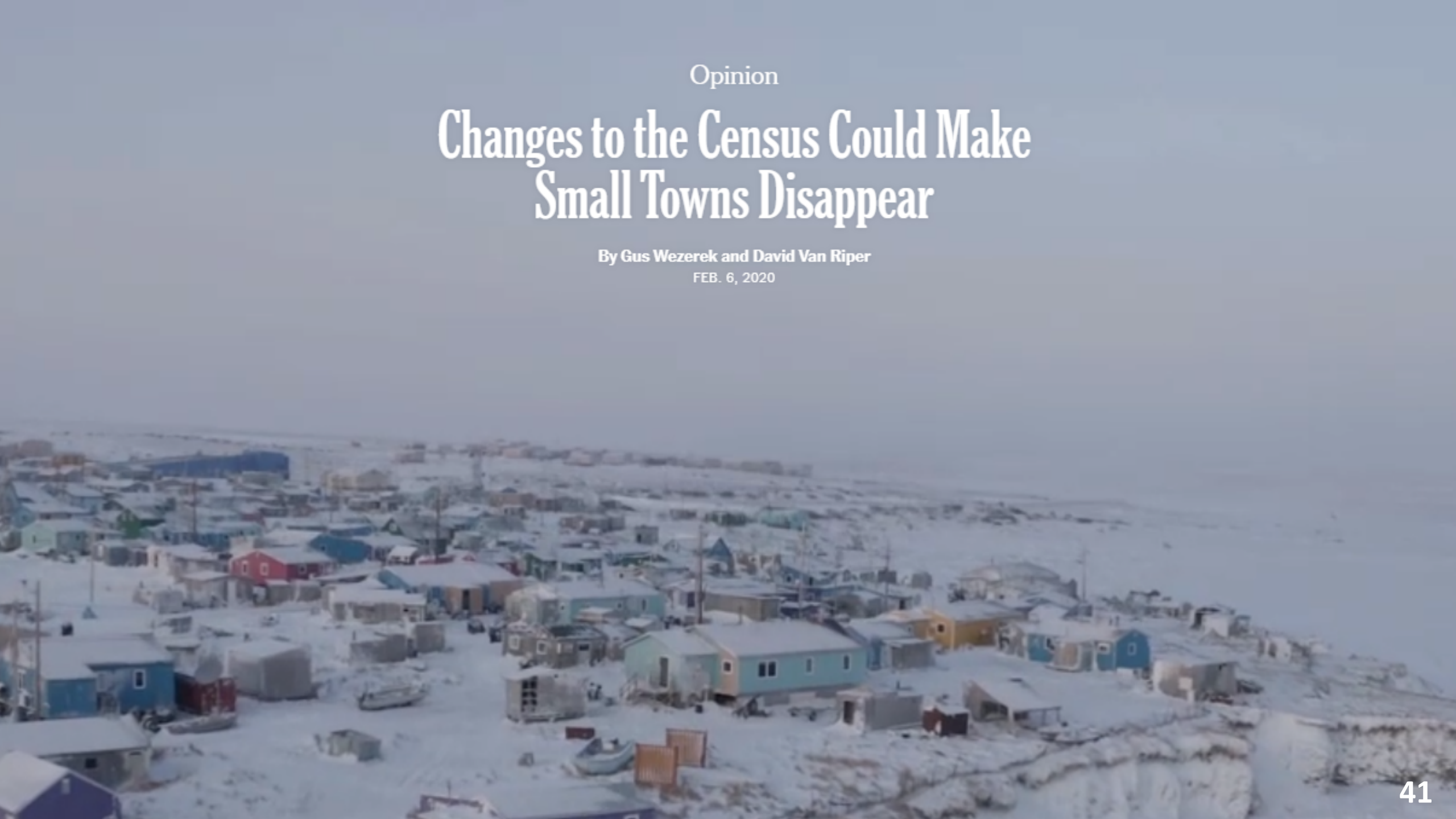
What are the social and equity risks?

Opinion

Changes to the Census Could Make Small Towns Disappear

By Gus Wezerek and David Van Riper

FEB. 6, 2020




WHY AM I ALWAYS BEING RESEARCHED?
A GUIDEBOOK FOR COMMUNITY ORGANIZATIONS, RESEARCHERS, AND FUNDERS TO HELP US GET FROM INSUFFICIENT UNDERSTANDING TO MORE AUTHENTIC TRUTH

Contact Us

nandofioretto@gmail.com  cbowen@urban.org

nandofioretto.com  clairemckaybowen.com

@nandofioretto  @ClaireMKBowen

References

Cynthia Dwork, Frank McSherry, Kobbi Nissim, Adam D. Smith: Calibrating Noise to Sensitivity in Private Data Analysis. TCC 2006: 265-284

Ferdinando Fioretto, Pascal Van Hentenryck, Keyu Zhu: Differential privacy of hierarchical Census data: An optimization approach. Artif. Intell. 296: 103475 (2021)

Keyu Zhu, Ferdinando Fioretto, Pascal Van Hentenryck:Post-processing of Differentially Private Data: A Fairness Perspective. CoRR abs/2201.09425 (2022)

David Pujol, Ryan McKenna, Satya Kuppam, Michael Hay, Ashwin Machanavajjhala, Gerome Miklau: Fair decision making using privacy-protected data. FAT* 2020: 189-199

Cuong Tran, Ferdinando Fioretto, Pascal Van Hentenryck: Differentially Private and Fair Deep Learning: A Lagrangian Dual Approach. AAAI 2021: 9932-9939

Keyu Zhu, Pascal Van Hentenryck, Ferdinando Fioretto: Bias and Variance of Post-processing in Differential Privacy. AAAI 2021: 11177-11184

Cuong Tran, Ferdinando Fioretto, Pascal Van Hentenryck, Zhiyan Yao: Decision Making with Differential Privacy under a Fairness Lens. IJCAI 2021: 560-566

Cuong Tran, My H. Dinh, Ferdinando Fioretto: Differentially Private Deep Learning under the Fairness Lens. NeurIPS 2021

Cuong Tran, My H. Dinh, Kyle Beiter, Ferdinando Fioretto: A Fairness Analysis on Private Aggregation of Teacher Ensembles. CoRR abs/2109.08630 (2021)