# CS2080: Applied Privacy for Data Science
# Attacks on Utility

School of Engineering & Applied Sciences
Harvard University

March 12, 2025

# Housekeeping

- Second call: Please do filling out the mid-semester feedback form (see Henry's post on Ed) **by tonight**
  - Preliminary responses to midterm feedback on next slide
- Add your project ideas to the spreadsheet if you haven't already done so, and express interest in ≥2 other ideas by **Friday night**
- Final project poster session **9am-12pm on Thursday 5/8** and revision of project papers due that night
- Solutions to HWs 1&2 posted on Canvas; others are coming

# Housekeeping

- Preliminary responses to midterm feedback
  - Perusall readings should all be downloadable (let us know if we miss any)
  - Section is the place to get practice problem-solving for HW prep. "probably the most helpful part of the course in terms of understanding"
  - For more depth in theory, see the annotated bibliography and/or take CS2260 in the Fall.
  - We've added HW deadlines to the course Google calendar. We generally are not making use of Canvas.
  - We will more systematically monitor Ed for questions.
  - For HoDP readings, feel free to comment about what you've found clear/unclear/interesting or requests for things for us to go over in lecture.
  - We have been trying release & polish the HWs sooner, to minimize frustrating updates.

# Census DAS Process

"Disclosure Avoidance System"

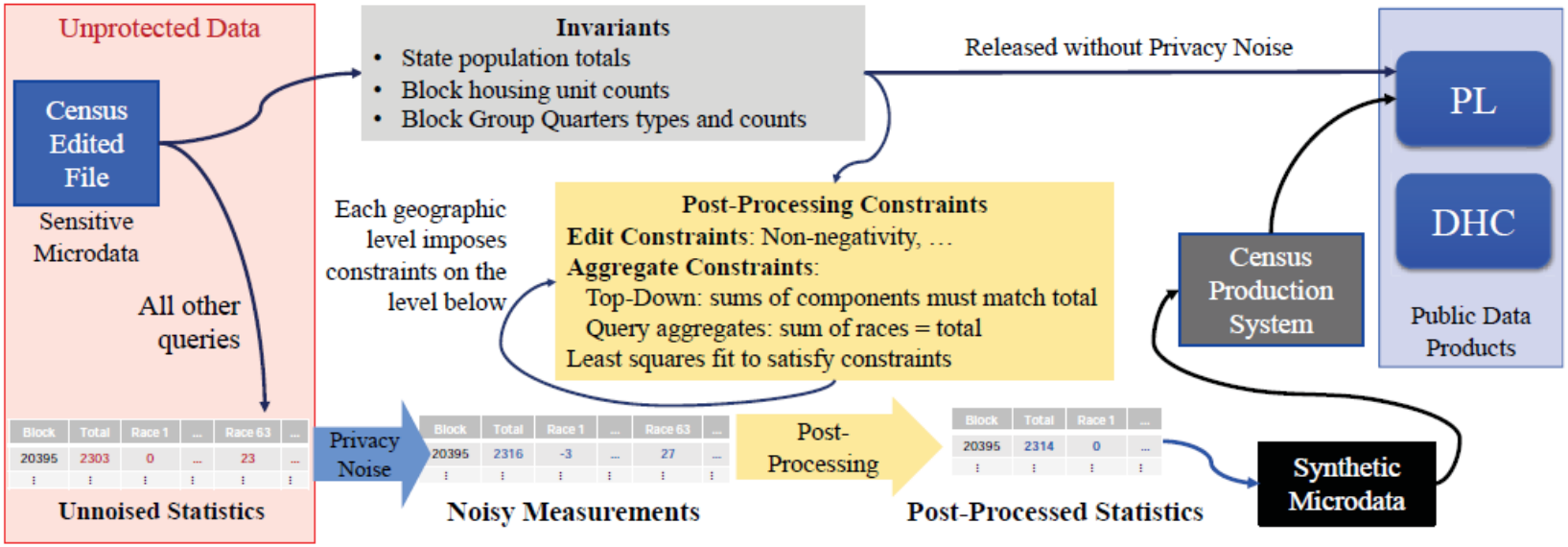



Figure 5-1: Process used to produce privacy-protected data products.

What happens to the **utility** of Census data when DP is applied?

# There are several uses of census data

- Reapportionment
- Redistricting
- Funding allocation (*billions* of $)
- Social science research
- Business decisions
- etc.

https://www.census.gov/programs-surveys/decennial-census/about/why.html

# There are several uses of census data

- Reapportionment
- Redistricting
- Funding allocation (*billions* of $)
- Social science research
- Business decisions
- etc.

**How are these uses impacted by DP?**
(i.e., What is the utility of census data protected under DP?)

https://www.census.gov/programs-surveys/decennial-census/about/why.html

# Utility vs. accuracy

- **Accuracy:** how close a published, differentially private estimate is to its non-differentially private counterpart

- **Utility:** “usefulness of a dataset or statistic for various societally beneficial purposes”
  - Sometimes studying utility may require qualitative approaches, but today we’ll stick to quantitative approaches.

https://www.cs.umd.edu/~kaptchuk/publications/nist24-dp-public-comment.pdf

# Utility vs. accuracy

- **Accuracy:** how close a published, differentially private estimate is to its non-differentially private counterpart

- **Utility:** "usefulness of a dataset or statistic for various societally beneficial purposes" **today**
  - Sometimes studying utility may require qualitative approaches, but today we'll stick to quantitative approaches.

https://www.cs.umd.edu/~kaptchuk/publications/nist24-dp-public-comment.pdf

# There are several uses of census data

- Reapportionment
- Redistricting
- Funding allocation (*billions* of $)
- Social science research
- Business decisions
- etc.

**today**

https://www.census.gov/programs-surveys/decennial-census/about/why.html

# Evaluating ("attacking") utility

1. Choose a dataset to be protected under DP
2. Consider a <u>specific</u> real-world use of that dataset
3. Formulate a metric (broadly defined) that would help understand the <u>utility</u> of the dataset for that specific task
4. Design & run simulations where DP noise is used to protect the data. Compare metrics on the DP-noised data vs. the dataset without protections*

*In practice, it may be hard to get access to the dataset without protections. In such a case, you may simulate the original data first.

# Evaluating ("attacking") utility

1. Choose a dataset to be protected under DP
2. Consider a <u>specific</u> real-world use of that dataset
3. Formulate a metric (broadly defined) that would help understand the <u>utility</u> of the dataset for that specific task
4. Design & run simulations where DP noise is used to protect the data. Compare metrics on the DP-noised data vs. the dataset without protections*

**When would we want to change the comparison in Step 4?**

*In practice, it may be hard to get access to the dataset without protections. In such a case, you may simulate the original data first.

# Part I: Redistricting

# Voting Rights Act (VRA) of 1965

- Landmark legislation passed by Congress & signed into law by President Lyndon B. Johnson during the civil rights movement
- Intended to uphold Black people's rights to vote and stop race discrimination in voting (e.g., literacy tests)
- **Section 2**: prohibits practices that "[result] in a denial or abridgement of the right of any citizen of the United States to vote on account of race or color"


Image: Library of Congress LC-U9-10344-16
Photographer: Marion S. Trikoso

https://www.archives.gov/milestone-documents/voting-rights-act
https://www.law.umich.edu/facultyhome/votingrights/Pages/SECTION-2-OF-THE-VOTING-RIGHTS-ACT.aspx

# Voting Rights Act (VRA) of 1965

- Landmark legislation passed by Congress & signed into law by President Lyndon B. Johnson during the civil rights movement
- Intended to uphold Black people's rights to vote and stop race discrimination in voting (e.g., literacy tests)
- **Section 2**: prohibits practices that "[result] in a denial or abridgement of the right of any citizen of the United States to vote on account of race or color"


Image: Library of Congress LC-U9-10344-16
Photographer: Marion S. Trikoso

**One way to abridge voting rights is to draw districts in a way that intentionally dilutes the voting power of minorities.**

https://www.archives.gov/milestone-documents/voting-rights-act
https://www.law.umich.edu/facultyhome/votingrights/Pages/SECTION-2-OF-THE-VOTING-RIGHTS-ACT.aspx

# Background on the VRA & gerrymandering

# How will DP applied to the U.S. Census impact the ability to enforce the VRA?

Today we'll see two approaches to answering this question, each with different conclusions.

SCIENCE ADVANCES | RESEARCH ARTICLE

SOCIAL SCIENCES

# The use of differential privacy for census data and its impact on redistricting: The case of the 2020 U.S. Census

Christopher T. Kenny[1], Shiro Kuriwaki[2], Cory McCartan[3], Evan T. R. Rosenman[4], Tyler Simko[1], Kosuke Imai[1,3]*

We will focus on their analyses around the detection of **packing** and **cracking** in plans drawn for the South Caroline State House.

# Simulation setup

**Goal**: Analyze how DP impacts conclusions drawn about racial biases in redistricting plans (specifically “packing” and “cracking”)

# Simulation setup

**Goal**: Analyze how DP impacts conclusions drawn about racial biases in redistricting plans (specifically “packing” and “cracking”)

1. Generate 100,000 realistic redistricting plans for the South Carolina State House, using each of three datasets:

# Simulation setup

**Goal**: Analyze how DP impacts conclusions drawn about racial biases in redistricting plans (specifically "packing" and "cracking")

1. Generate 100,000 realistic redistricting plans for the South Carolina State House, using each of three datasets:
    1. Published Census 2010 data (remember, these data were subject to swapping) ("**Census 2010**")
    2. Published Census 2010 data protected with TopDown, $\boldsymbol{\varepsilon}$ = 4.5 ("**DAS-4.5**")
    3. Published Census 2010 data protected with TopDown, $\boldsymbol{\varepsilon}$ = 12.2 ("**DAS-12.2**")

# Simulation setup

**Goal**: Analyze how DP impacts conclusions drawn about racial biases in redistricting plans (specifically "packing" and "cracking")

1. Generate 100,000 realistic redistricting plans for the South Carolina State House, using each of three datasets:
    1. Published Census 2010 data (remember, these data were subject to swapping) ("**Census 2010**")
    2. Published Census 2010 data protected with TopDown, $\boldsymbol{\varepsilon}$ = 4.5 ("**DAS-4.5**")
    3. Published Census 2010 data protected with TopDown, $\boldsymbol{\varepsilon}$ = 12.2 ("**DAS-12.2**")
2. Find signals of packing or cracking in 2010 districts (the enacted plan) by comparing between the estimated proportion of Black people in simulated districts using **Census 2010** compared to the enacted plan (also using **Census 2010**)

# Simulation setup

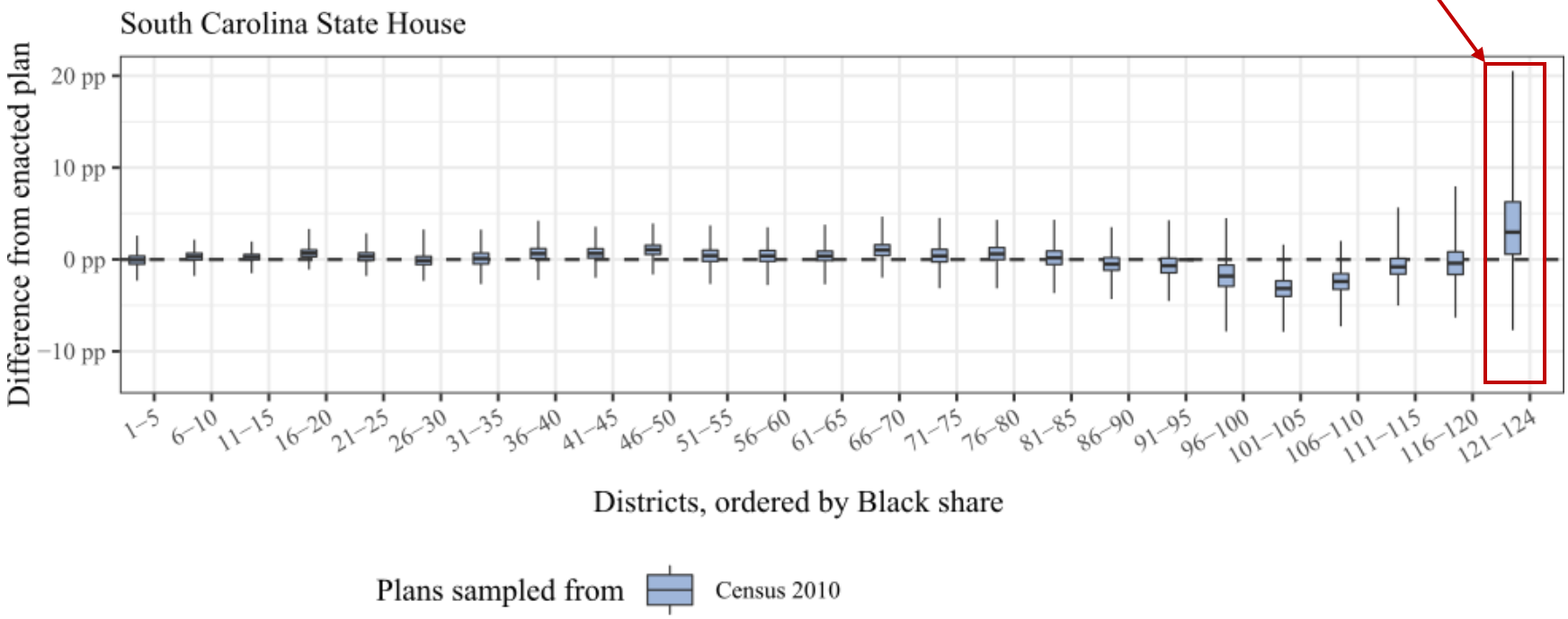
**Goal**: Analyze how DP impacts conclusions drawn about racial biases in redistricting plans (specifically "packing" and "cracking")

1. Generate 100,000 realistic redistricting plans for the South Carolina State House, using each of three datasets:
    1. Published Census 2010 data (remember, these data were subject to swapping) ("**Census 2010**")
    2. Published Census 2010 data protected with TopDown, $\boldsymbol{\varepsilon}$ = 4.5 ("**DAS-4.5**")
    3. Published Census 2010 data protected with TopDown, $\boldsymbol{\varepsilon}$ = 12.2 ("**DAS-12.2**")
2. Find signals of packing or cracking in 2010 districts (the enacted plan) by comparing between the estimated proportion of Black people in simulated districts using **Census 2010** compared to the enacted plan (also using **Census 2010**)
3. Observe whether signals of packing or cracking found above remain, disappear, or reverse with simulated plans based on **DAS-4.5** or **DAS-12.2**

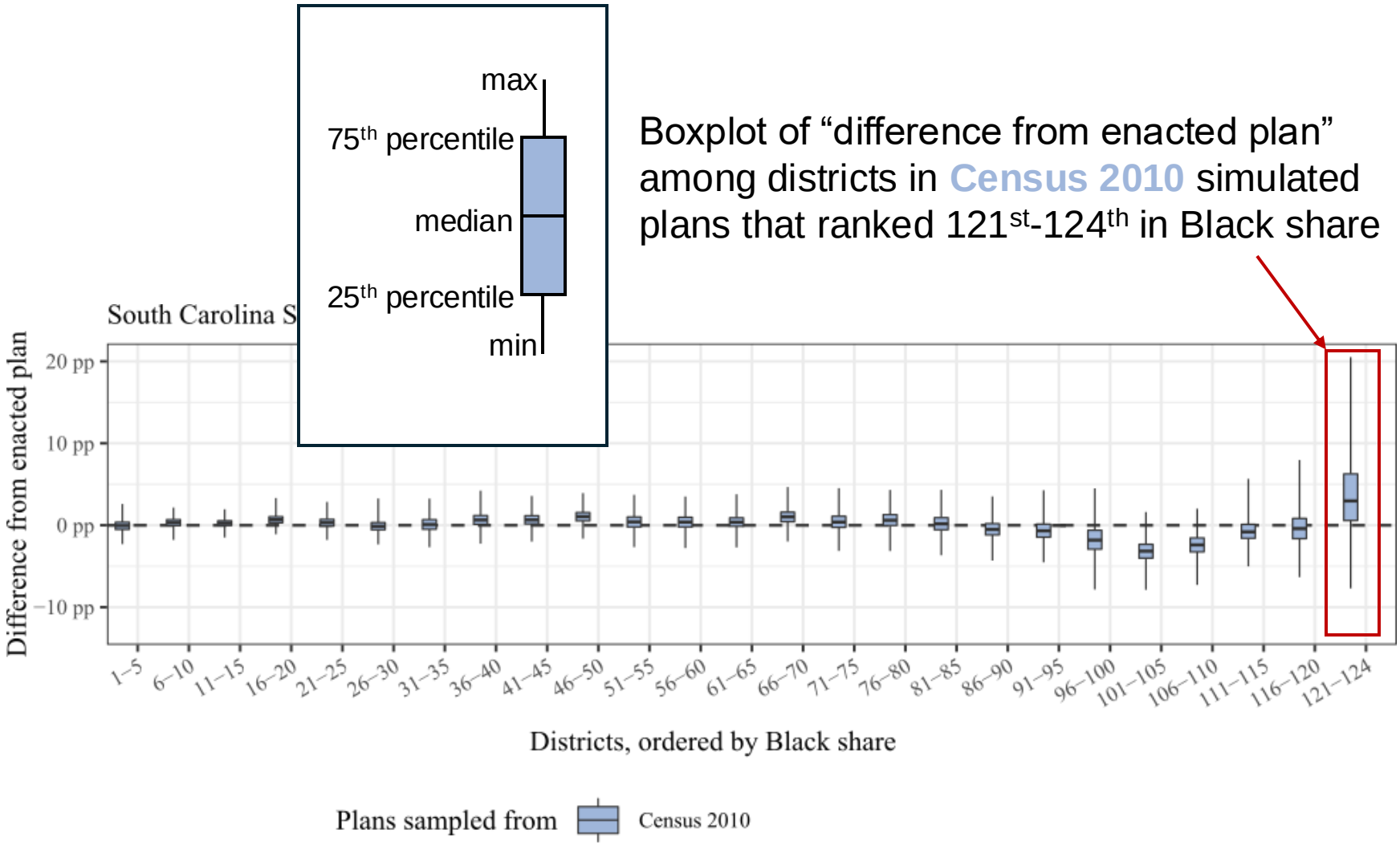
Difference from enacted plan =
Black share in simulated district – Black share in enacted district


Districts, in *ascending* order of Black share

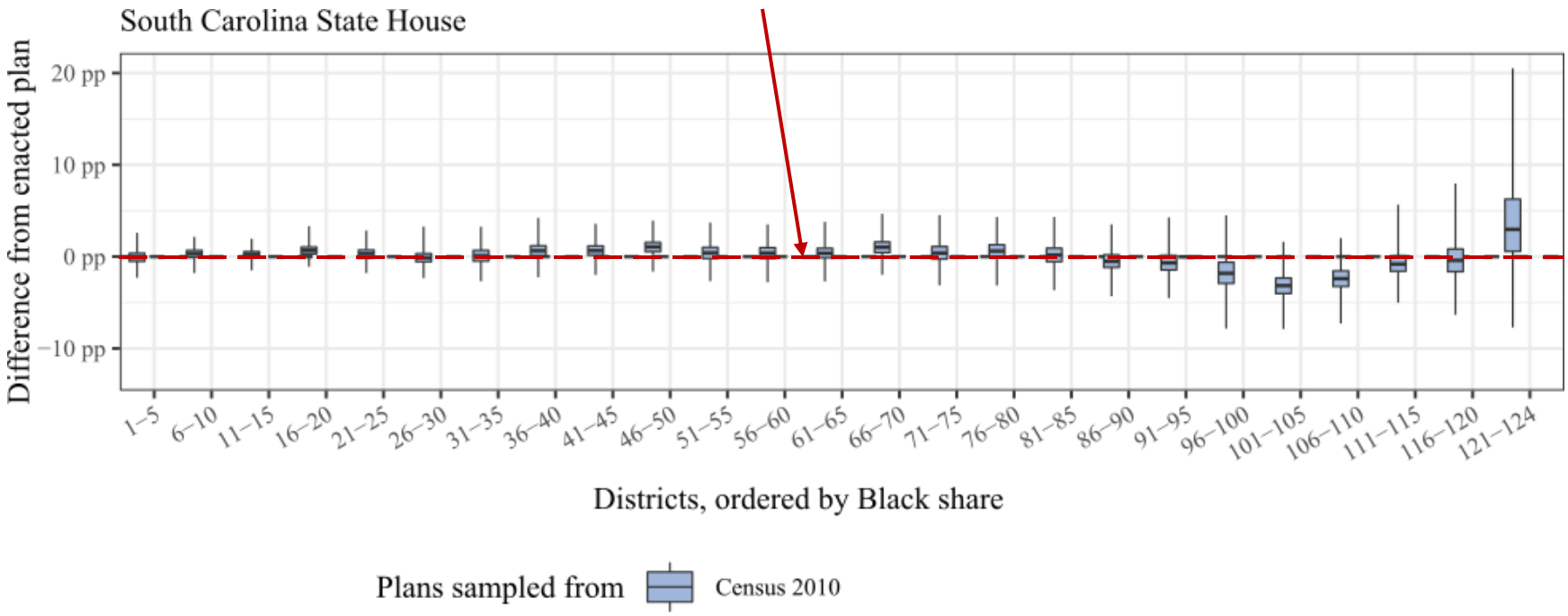
Boxplot of “difference from enacted plan” among districts in **Census 2010** simulated plans that ranked 121st-124th in Black share



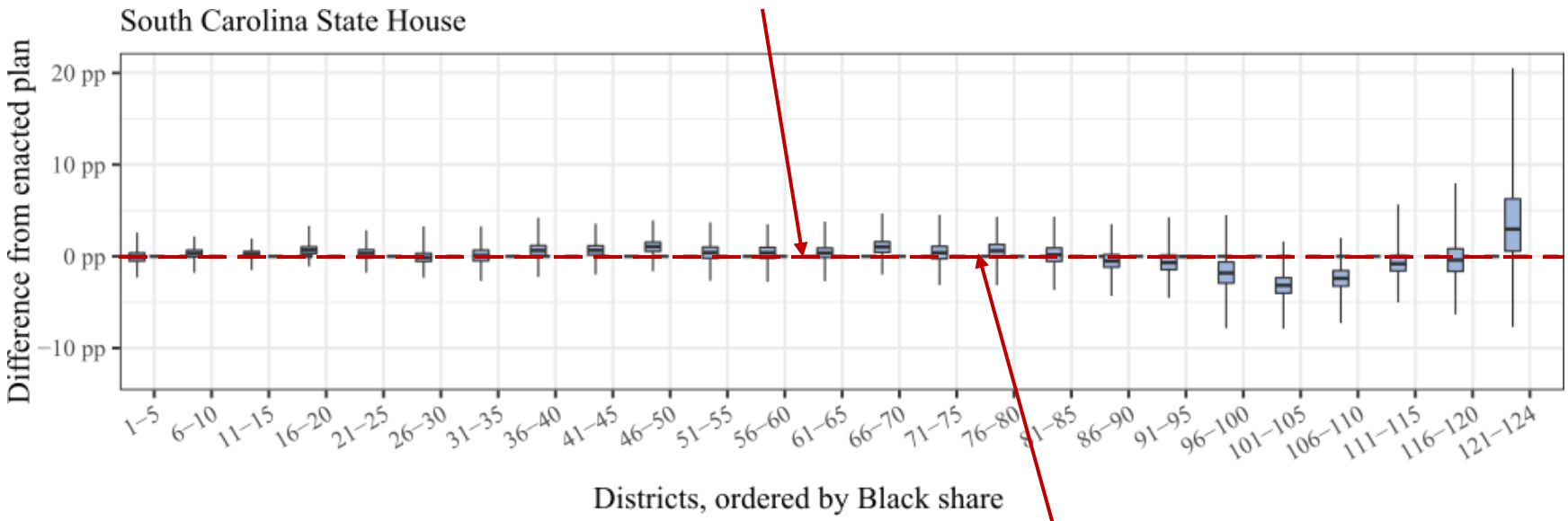
Boxplot of "difference from enacted plan" among districts in Census 2010 simulated plans that ranked 121st-124th in Black share



When “difference from enacted plan” > 0, there was higher Black share in simulated plans compared to the enacted plan (i.e., the enacted plan had *lower* Black share than what we would expect from a politically-neutral baseline).
This is evidence of “cracking” in the enacted plan.

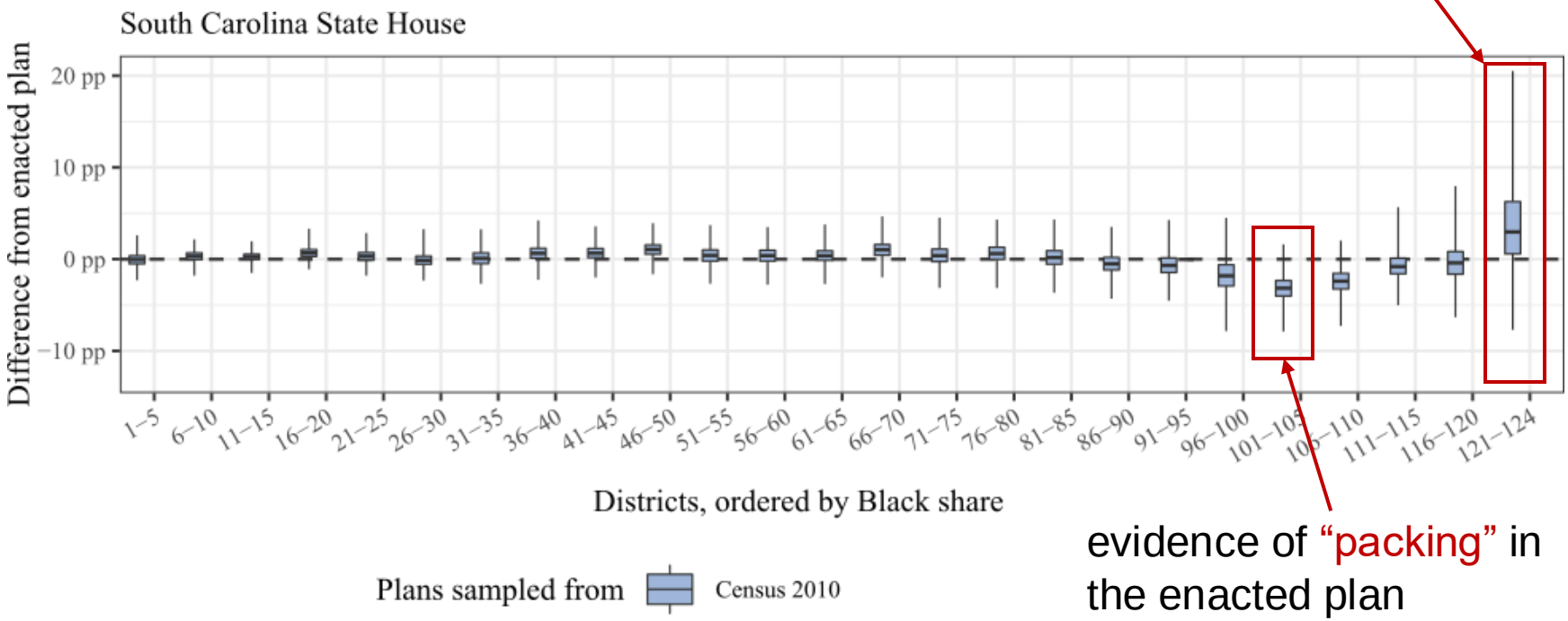


When “difference from enacted plan” > 0, there was higher Black share in simulated plans compared to the enacted plan (i.e., the enacted plan had *lower* Black share than what we would expect from a politically-neutral baseline).
This is evidence of “cracking” in the enacted plan.

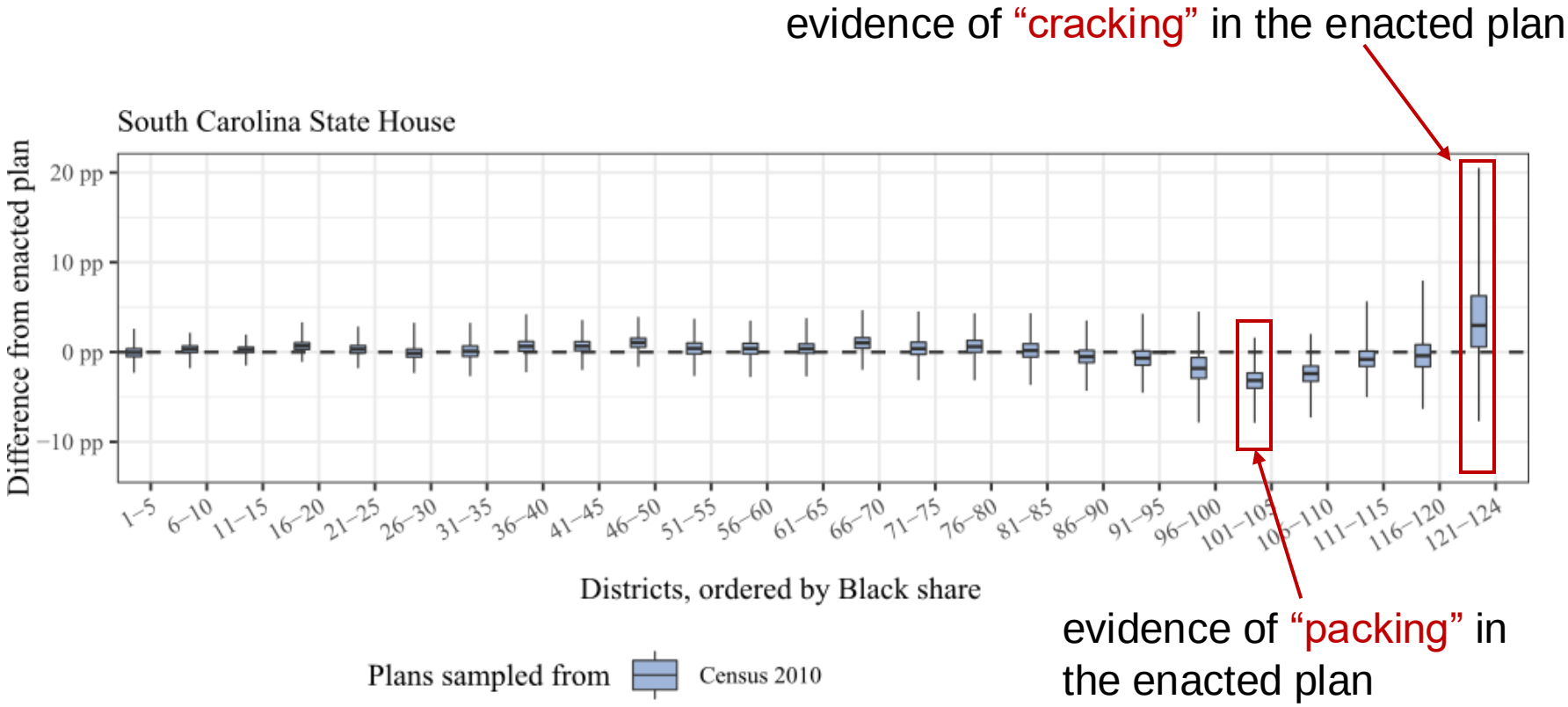



When “difference from enacted plan” < 0, there was lower Black share in simulated plans compared to the enacted plan (i.e., the enacted plan had *higher* Black share than what we would expect from a politically-neutral baseline). In other words, there is evidence of “packing” in the enacted plan.

evidence of “cracking” in the enacted plan

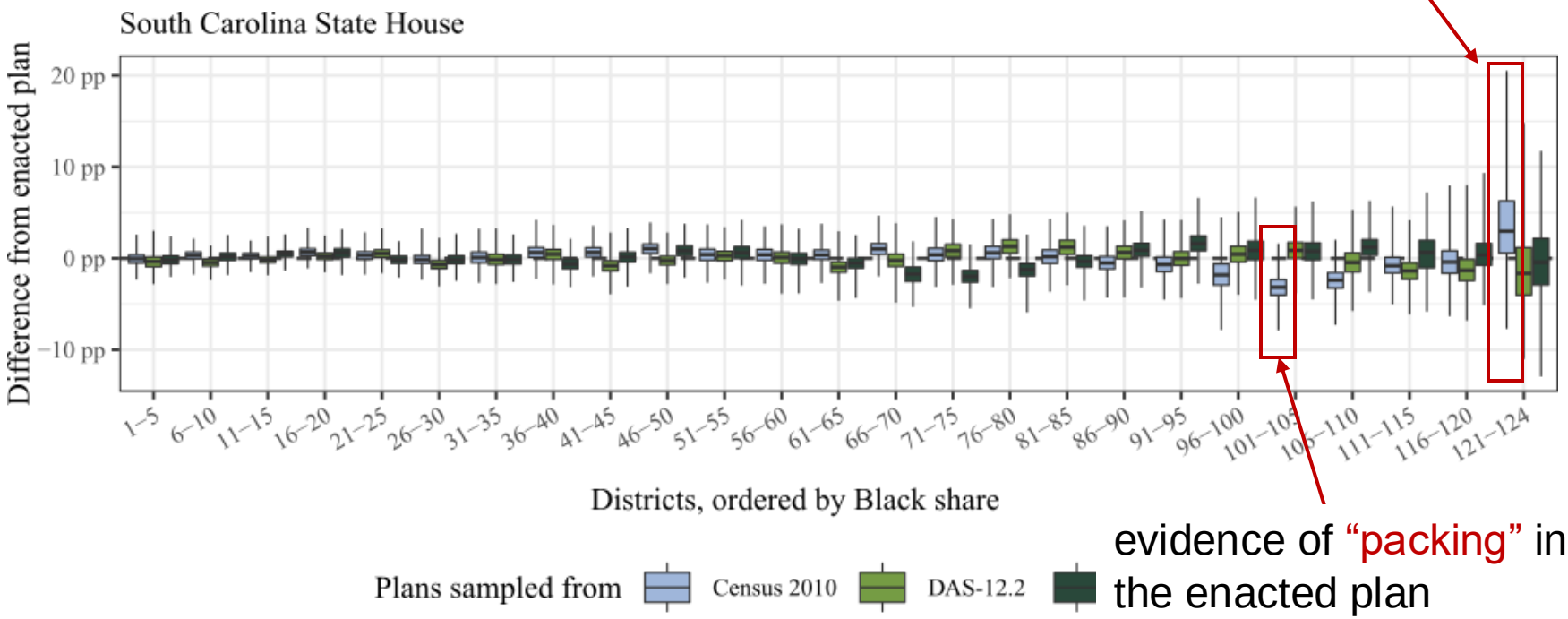



evidence of “packing” in the enacted plan



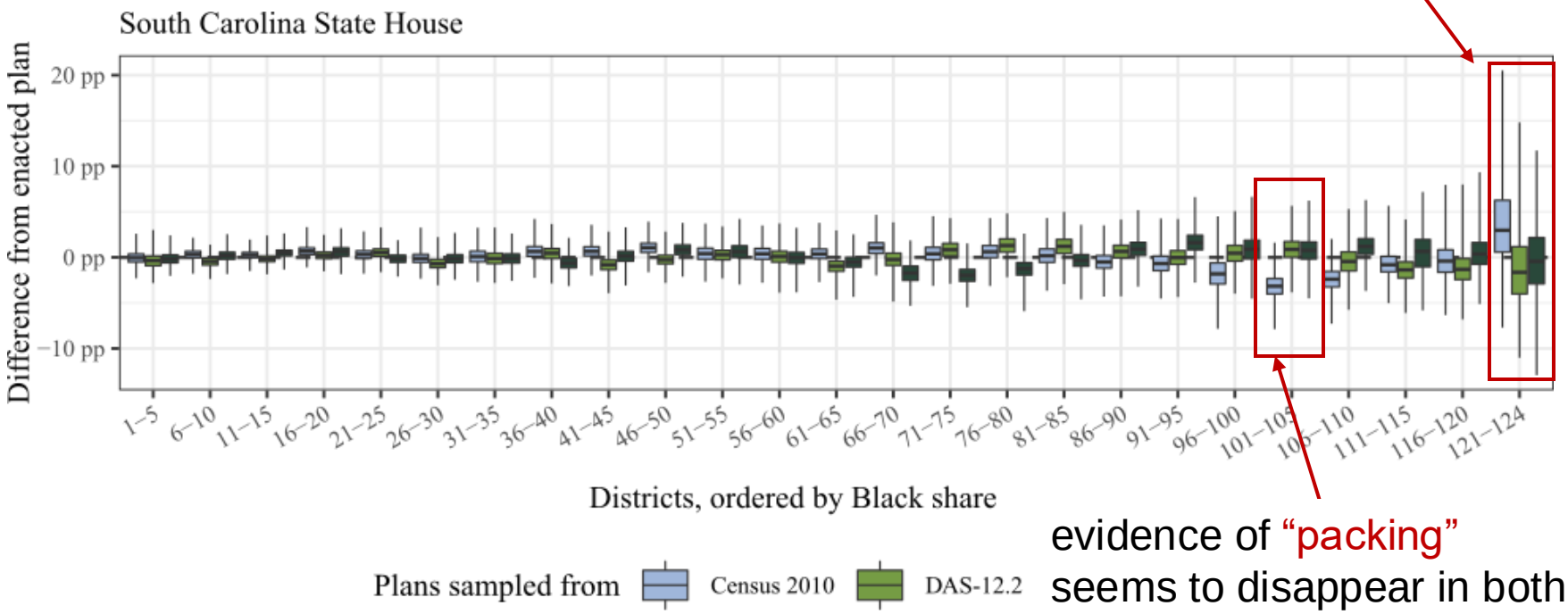

**Can we still detect evidence of cracking and packing using data protected under the new DAS?**

evidence of “cracking” in the enacted plan



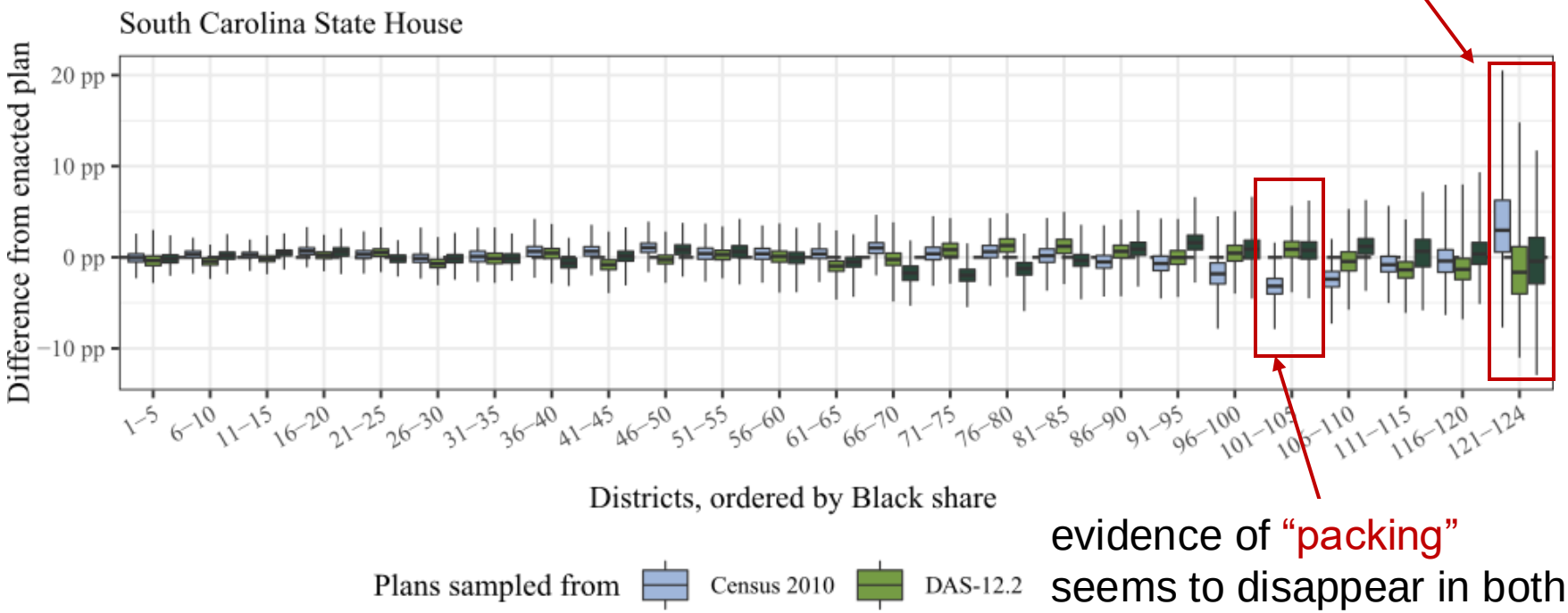

evidence of “packing” in the enacted plan

evidence of "cracking" seems to disappear in both DAS's simulated plans




evidence of "packing" seems to disappear in both DAS's simulated plans

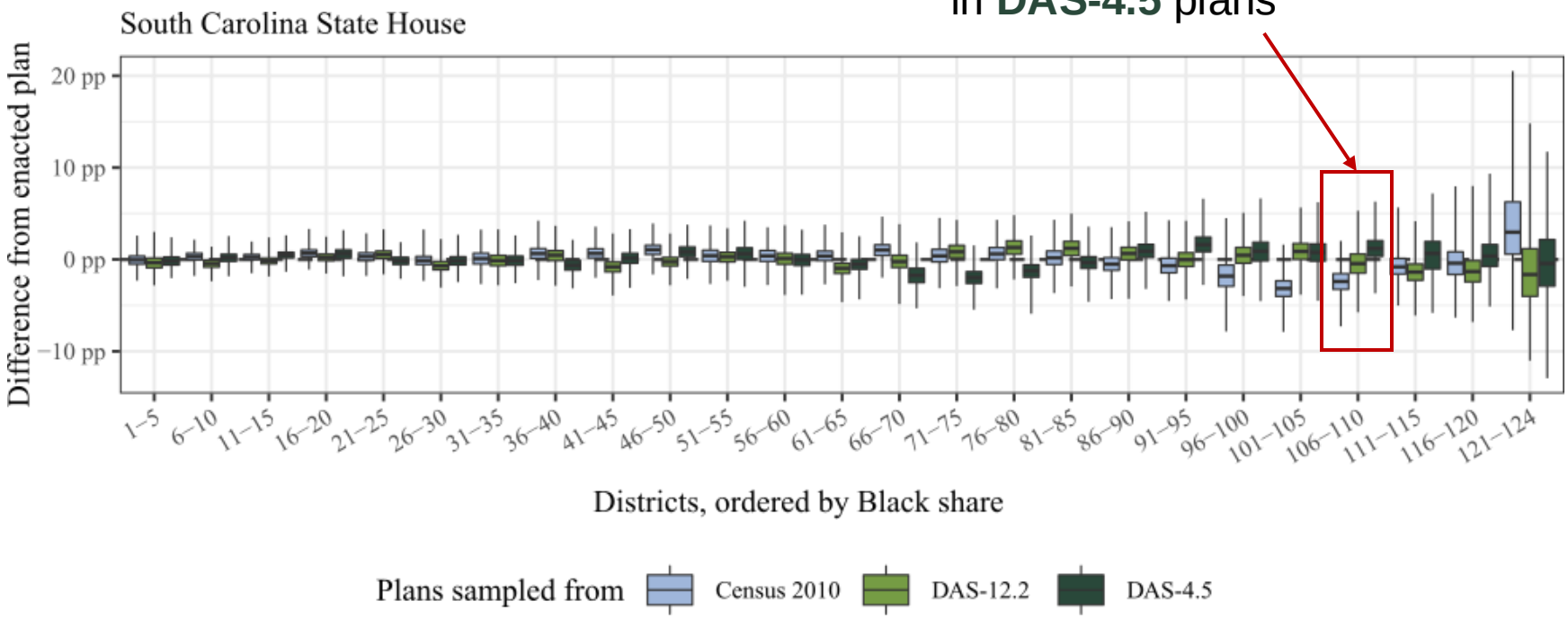
evidence of "cracking" seems to disappear in both DAS's simulated plans




evidence of "packing" seems to disappear in both DAS's simulated plans

**Can you find other disappearances or reversals of evidence?**

evidence of "packing" seems to reverse to evidence of "cracking" in **DAS-4.5** plans



# Census TopDown: The Impacts of Differential Privacy on Redistricting

**Aloni Cohen**
Hariri Institute for Computing and School of Law, Boston University, MA, USA

**Moon Duchin**
Department of Mathematics, Tufts University, Medford, MA, USA

**JN Matthews**
Tisch College of Civic Life, Tufts University, Medford, MA, USA

**Bhushan Suwal**
Tisch College of Civic Life, Tufts University, Medford, MA, USA

We will focus on their **ecological regression** analyses.

# Gingles factors

Demonstrating a violation of Section 2 of the VRA requires showing:

1. It's possible to create a district where the minority group is over 50% of the population
2. The minority group must be "politically cohesive"
3. The majority group also votes together as a bloc, such that it usually defeats the minority group's preferred candidate

# Gingles factors

Demonstrating a violation of Section 2 of the VRA requires showing:

1. It's possible to create a district where the minority group is over 50% of the population
2. The minority group must be "politically cohesive"
3. The majority group also votes together as a bloc, such that it usually defeats the minority group's preferred candidate

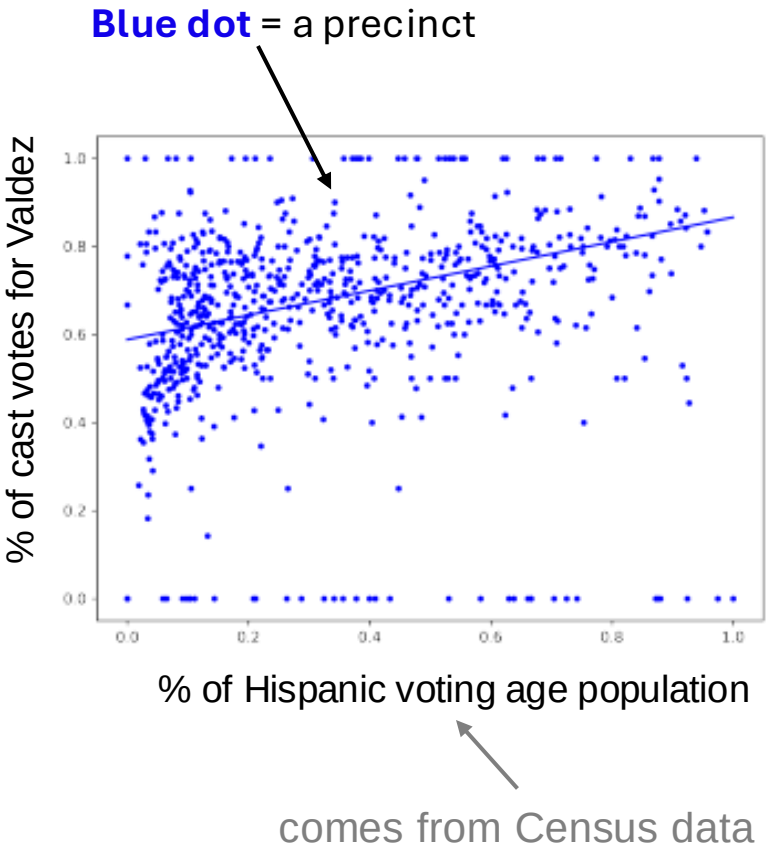
**= racially polarized voting**

# Simulation setup

**Goal**: Analyze whether it's possible to detect racially polarized voting with Census data protected under DP

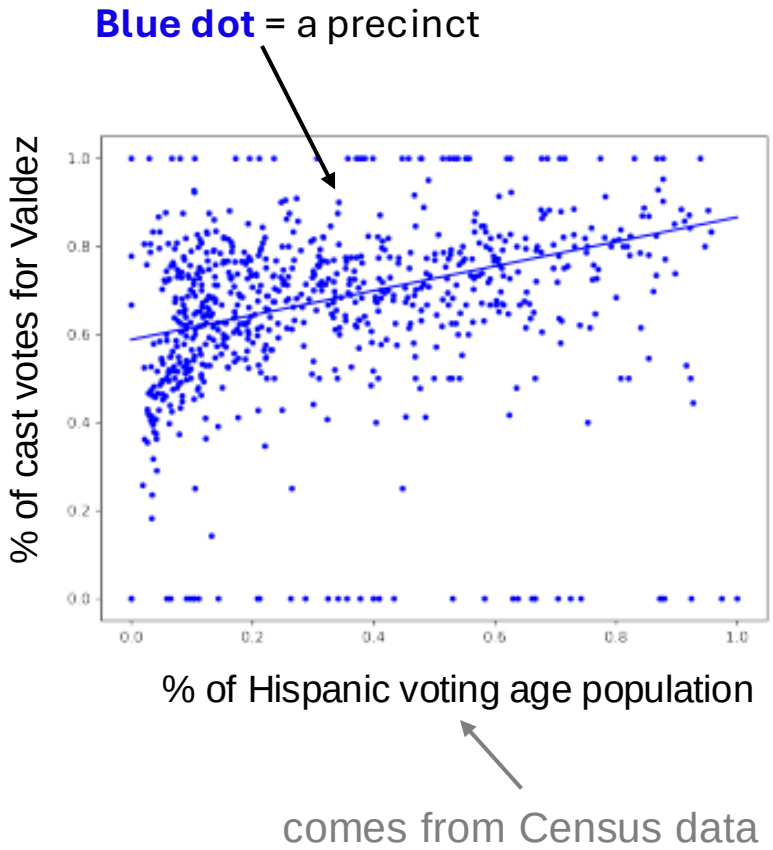
1. Reconstructed block-level 2010 microdata (i.e., person-level data) for Texas (but cannot quantify errors because they do not have access to the Census's microdata files)
2. Ran TopDown 16 times, $\boldsymbol{\varepsilon} = 1$ (equally split across geographical hierarchies)
3. Compared ability to detect racially polarized voting using reconstructed data ("un-noised data") vs. data protected using TopDown ("TopDown data")

# Ecological regression on un-noised data




**Ecological regression** on Dallas County precincts, showing support for Lupe Valdez for governor in the 2018 Democratic primary runoff

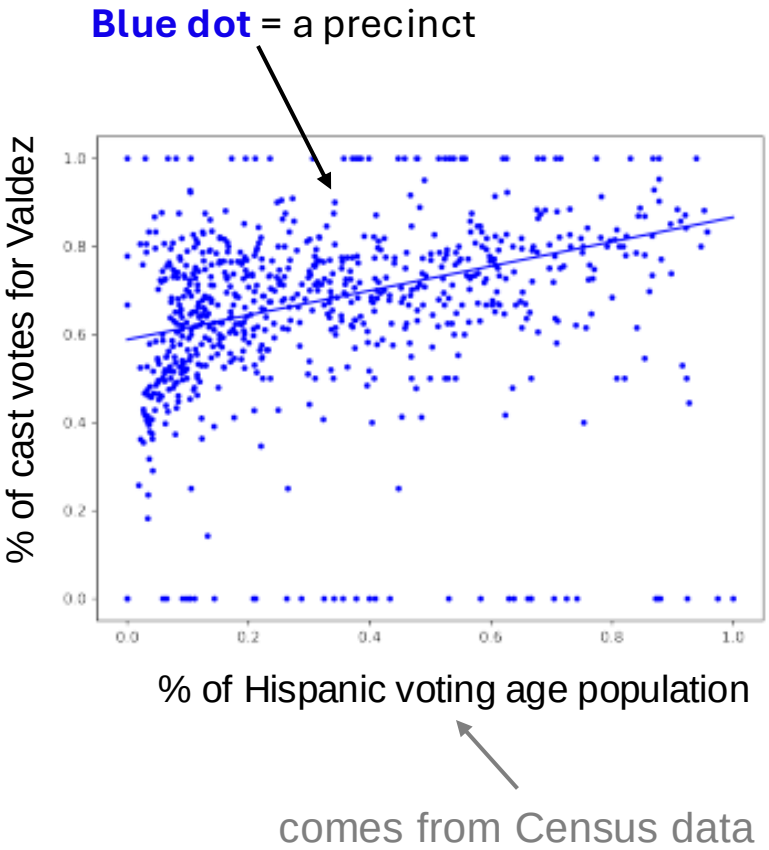
# Ecological regression on un-noised data




**Ecological regression** on Dallas County precincts, showing support for Lupe Valdez for governor in the 2018 Democratic primary runoff

Where the regression line intersects with $x = 1$ estimates support for Valdez in a 100% Hispanic population

# Ecological regression on un-noised data

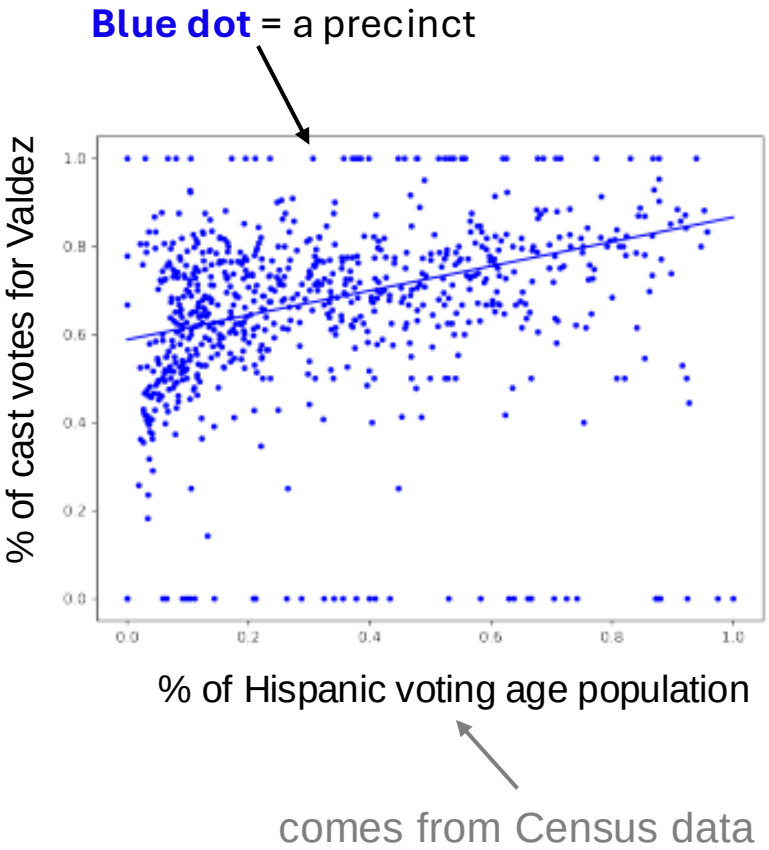



**Ecological regression** on Dallas County precincts, showing support for Lupe Valdez for governor in the 2018 Democratic primary runoff

Where the regression line intersects with $x = 1$ estimates support for Valdez in a 100% Hispanic population

Where the regression line intersects with $x = 0$ estimates support for Valdez in a 0% Hispanic (i.e., 100% non-Hispanic) population

# Ecological regression on un-noised data




**Ecological regression** on Dallas County precincts, showing support for Lupe Valdez for governor in the 2018 Democratic primary runoff

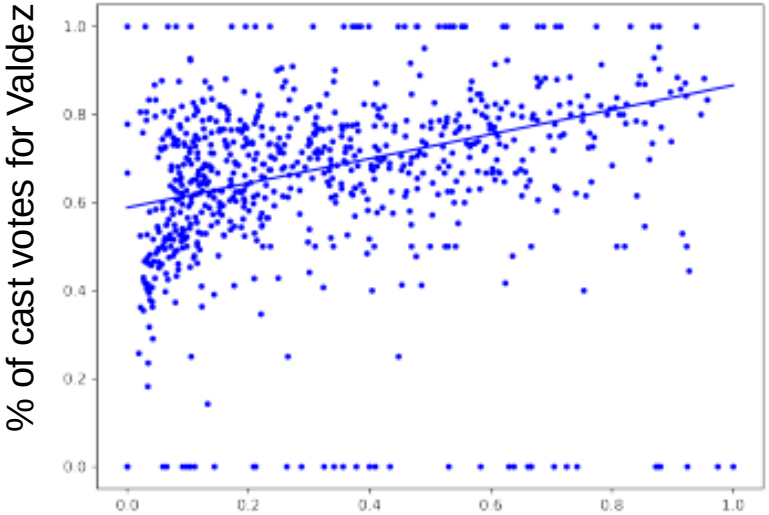
Where the regression line intersects with $x = 1$ estimates support for Valdez in a 100% Hispanic population

Where the regression line intersects with $x = 0$ estimates support for Valdez in a 0% Hispanic (i.e., 100% non-Hispanic) population

**What does a regression line with a large slope mean?**

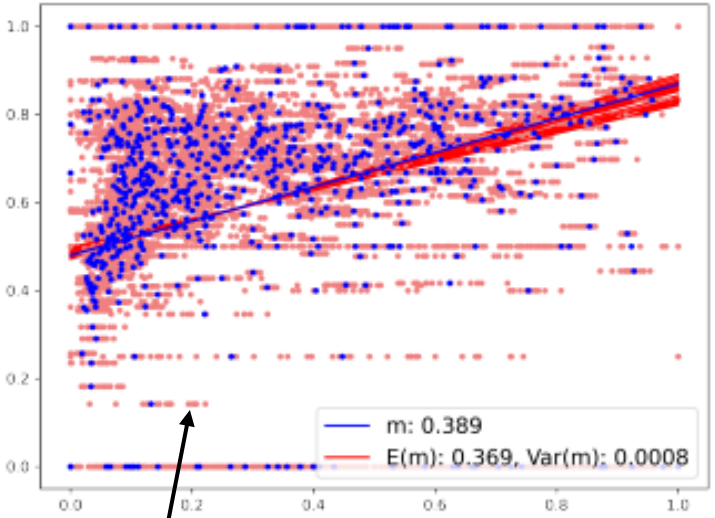
# Ecological regression on un-noised vs. TopDown data

Un-noised data



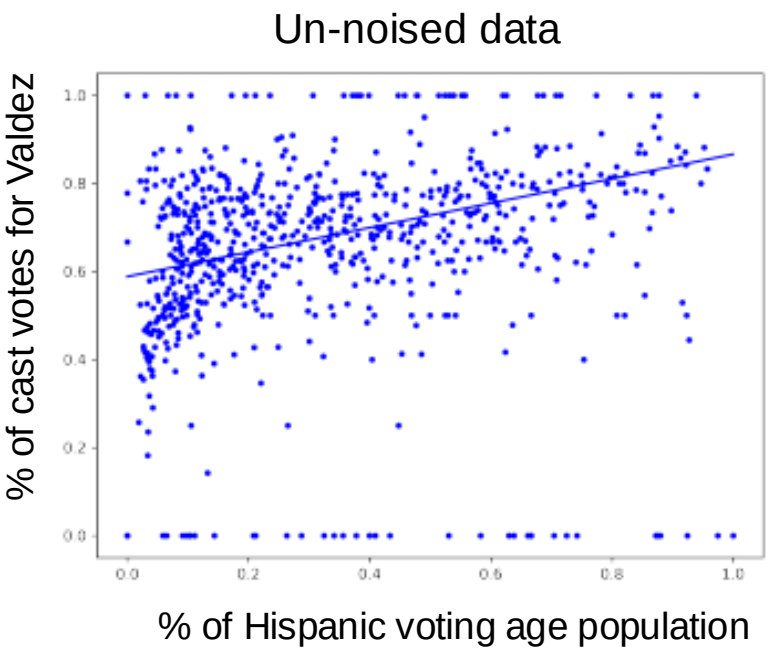

% of Hispanic voting age population

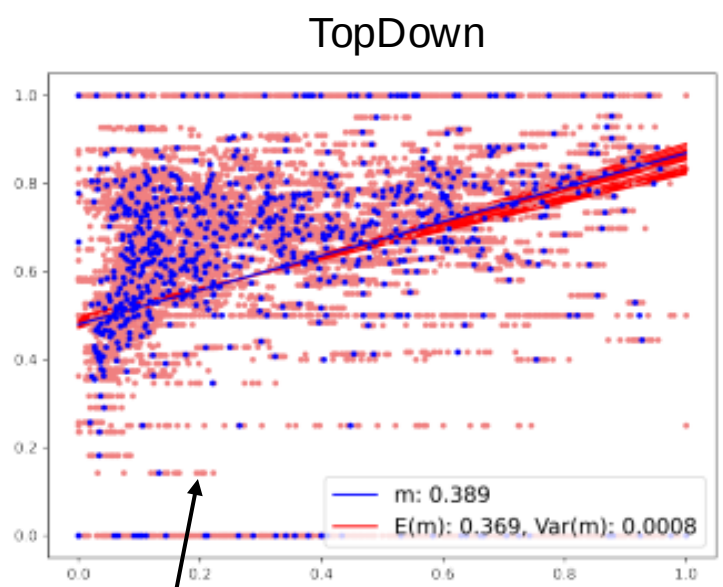
TopDown




**Pink dots** = 16 runs of TopDown
**Red lines** = regression lines corresponding to each of the 16 runs

# Ecological regression on un-noised vs. TopDown data

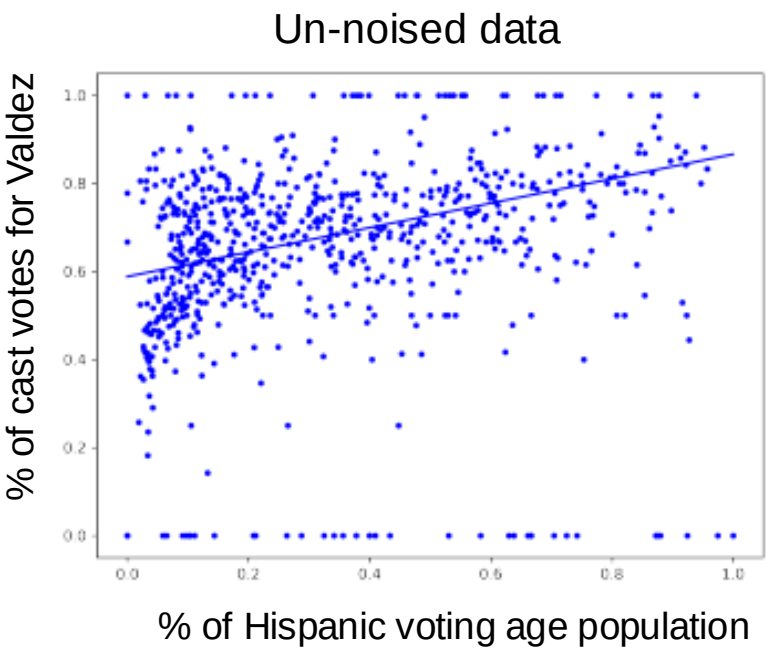


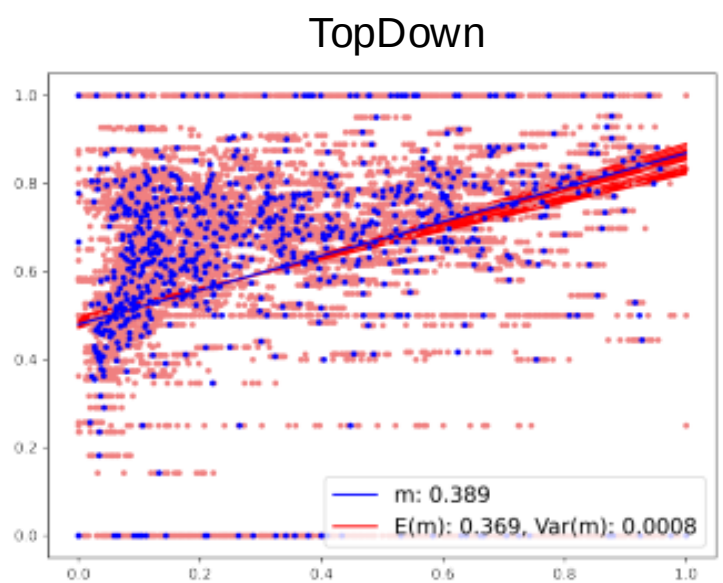




**Pink dots** = 16 runs of TopDown
**Red lines** = regression lines corresponding to each of the 16 runs

**Why do the pink dots jitter horizontally only?**

# Ecological regression on un-noised vs. TopDown data




**What do you think will happen if we remove precincts with few cast votes?**

# Ecological regression on un-noised vs. TopDown data

## Precincts with fewer than 10 cast votes filtered out


**The blue and red regression lines basically match.**

# Ecological regression on un-noised vs. TopDown data

## Precincts weighted by number of cast votes


**Again, the blue and red regression lines basically match.**

# Part II: Funding Allocation

POLICY FORUM | PRIVACY

# Policy impacts of statistical uncertainty and privacy

Funding formula reform may help address unequal impacts of uncertainty from data error and privacy protections

RYAN STEED , TERRANCE LIU, ZHIWEI STEVEN WU, AND ALESSANDRO ACQUISTI Authors Info & Affiliations

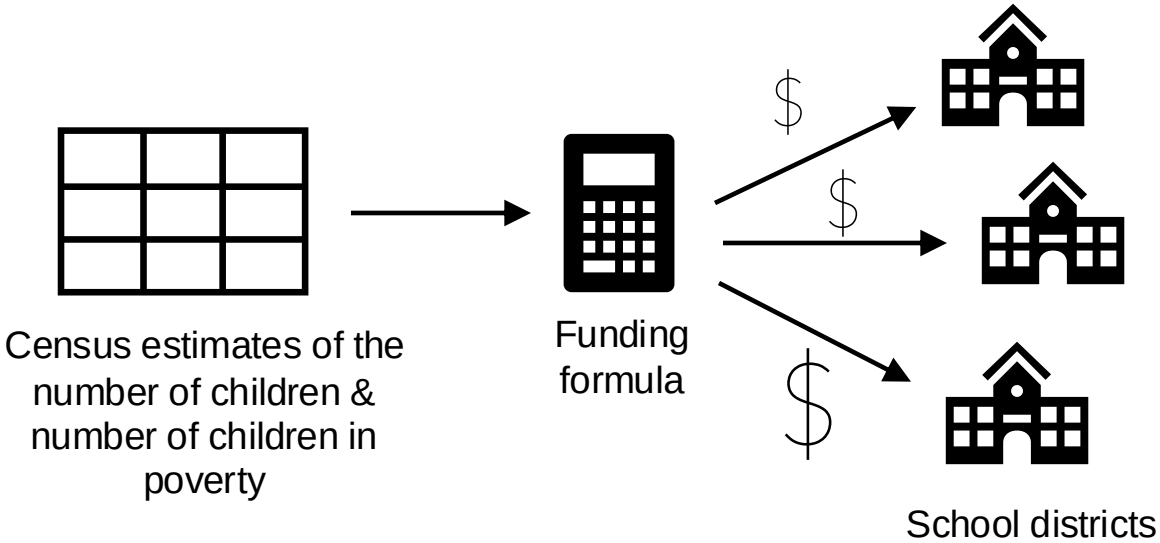
*SCIENCE* • 25 Aug 2022 • Vol 377, Issue 6609 • pp. 928-931 • DOI: 10.1126/science.abq4481

# Title I of the Elementary and Secondary Education Act

- Allocates funding to school districts with children in poverty
- "In 2021, the US federal government appropriated over $16.5 billion in Title I funds…to distribute to over 13,000 local education agencies (LEAs)”

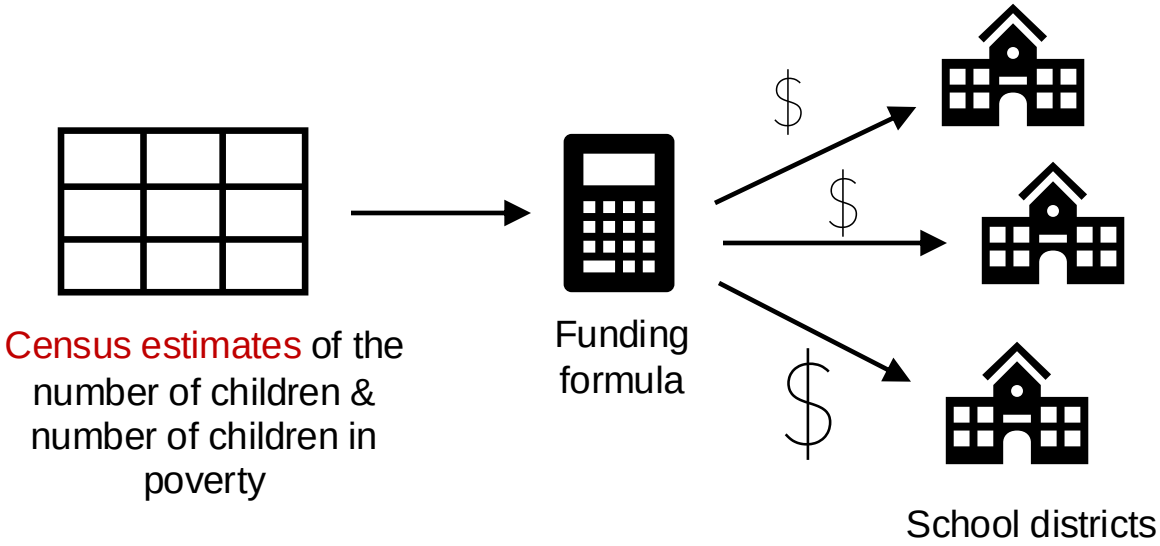
# Title I of the Elementary and Secondary Education Act

- Allocates funding to school districts with children in poverty
- "In 2021, the US federal government appropriated over $16.5 billion in Title I funds…to distribute to over 13,000 local education agencies (LEAs)”



# Title I of the Elementary and Secondary Education Act

- How would funding allocation change if DP were applied to census estimates?
- And how do downstream impacts of DP compare to impacts from other sources of data error?



# Simulate sources of error

Quantifiable error

Data deviations

Privacy deviations

# Simulate sources of error

## Quantifiable error

## Data deviations

e.g., error from other data sources used in the Small Area Income and Poverty Estimates; errors from converting county estimates → school district estimates

*Can model these deviations by sampling from a normal distribution centered at reported estimate*

## Privacy deviations

# Simulate sources of error

## Quantifiable error

## Data deviations

e.g., error from other data sources used in the Small Area Income and Poverty Estimates; errors from converting county estimates → school district estimates

*Can model these deviations by sampling from a normal distribution centered at reported estimate*

## Privacy deviations

e.g., error from DP mechanism

*Can model these by adding Laplace noise to American Community Survey estimates, which are used to generate estimates fed to the funding formula*

# Simulate sources of error

1. Simulate data deviations & privacy deviations
2. Compute formula-based allocations (i.e., entitlements)
3. Compare above allocations to the official allocations

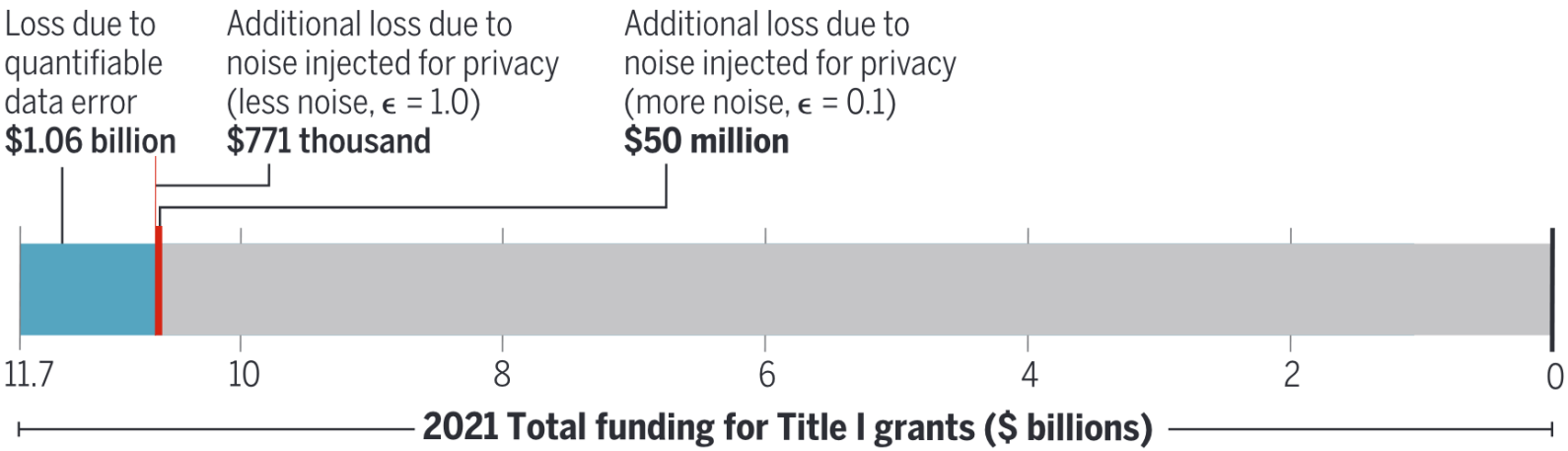
*Repeat 1,000 times*

# Simulate sources of error

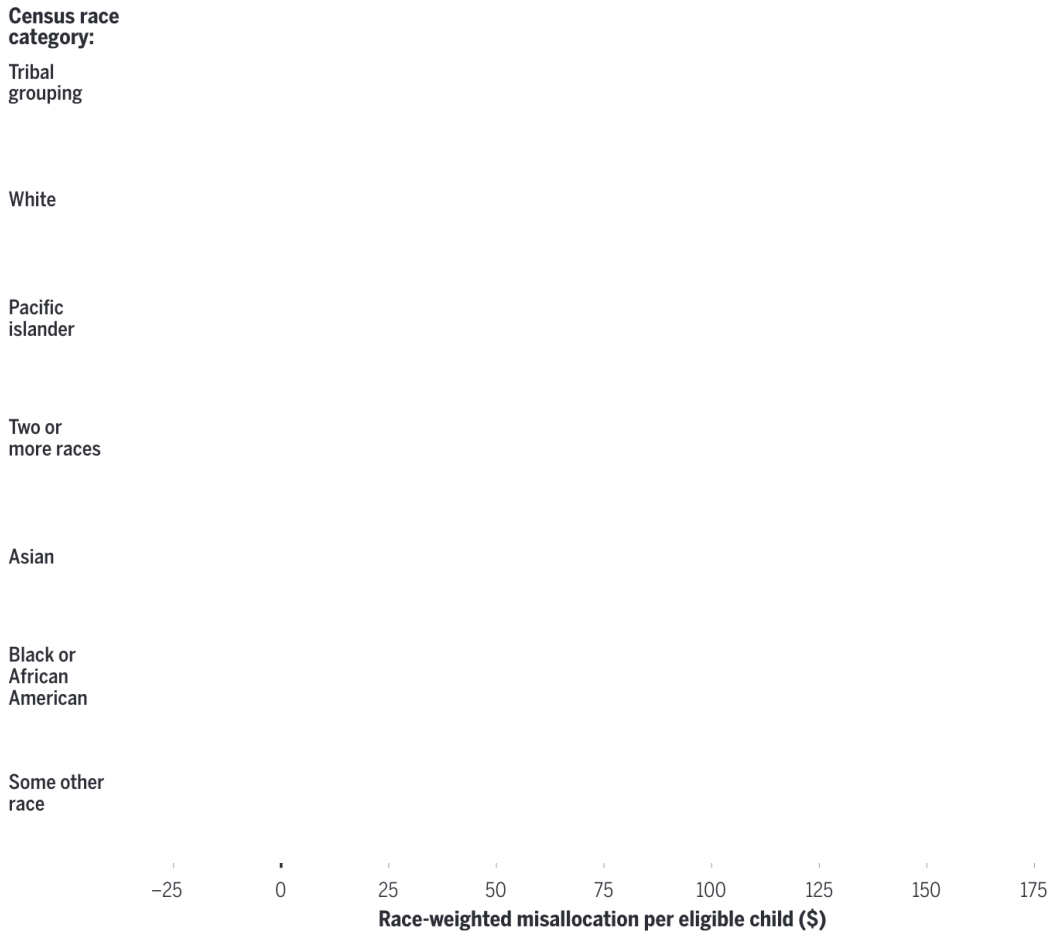
1. Simulate data deviations & privacy deviations (two privacy loss budgets)
2. Compute formula-based allocations (i.e., entitlements)
3. Compare above allocations to the official allocations

*Repeat 1,000 times*

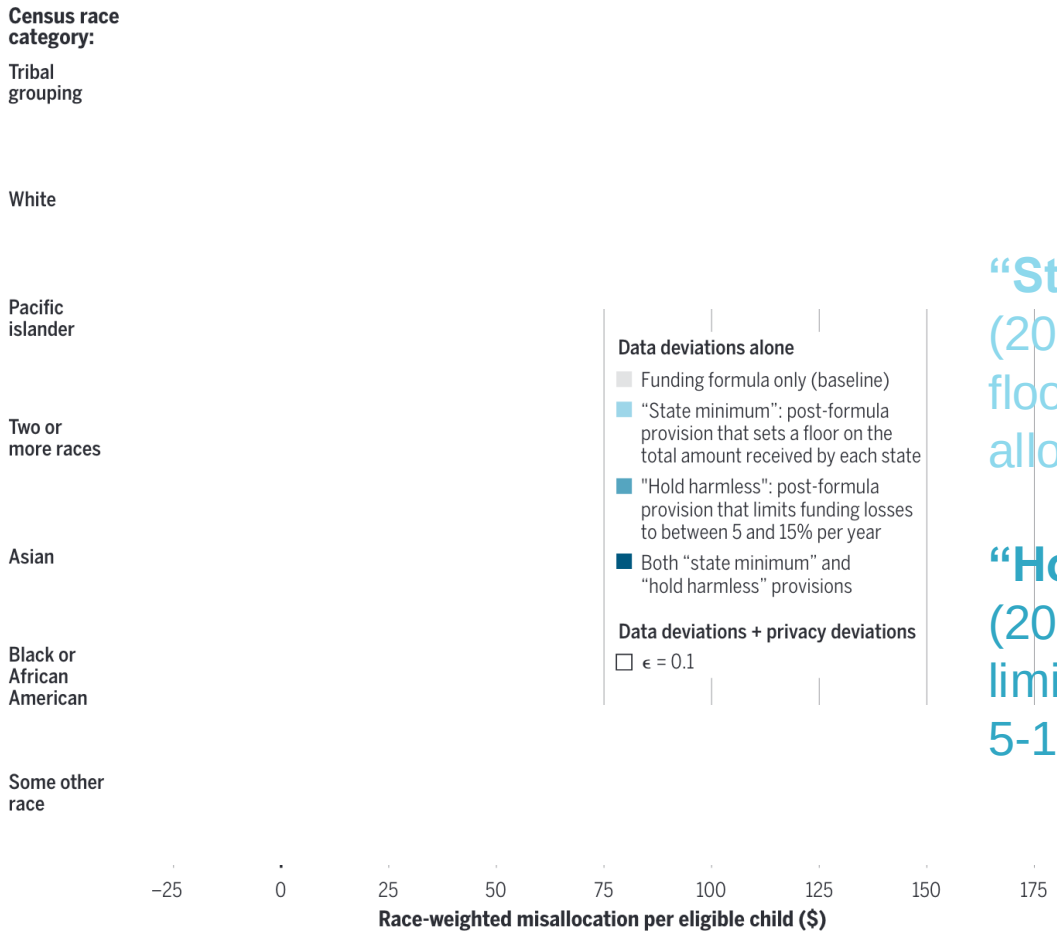
# Results


Loss due to quantifiable data error
$1.06 billion
Additional loss due to noise injected for privacy (less noise, ϵ = 1.0)
$771 thousand
Additional loss due to noise injected for privacy (more noise, ϵ = 0.1)
$50 million
11.7
10
8
6
4
2
0
2021 Total funding for Title I grants ($ billions)

# Results


Census race category:
Tribal grouping
White
Pacific islander
Two or more races
Asian
Black or African American
Some other race
−25
0
25
50
75
100
125
150
175
Race-weighted misallocation per eligible child ($)

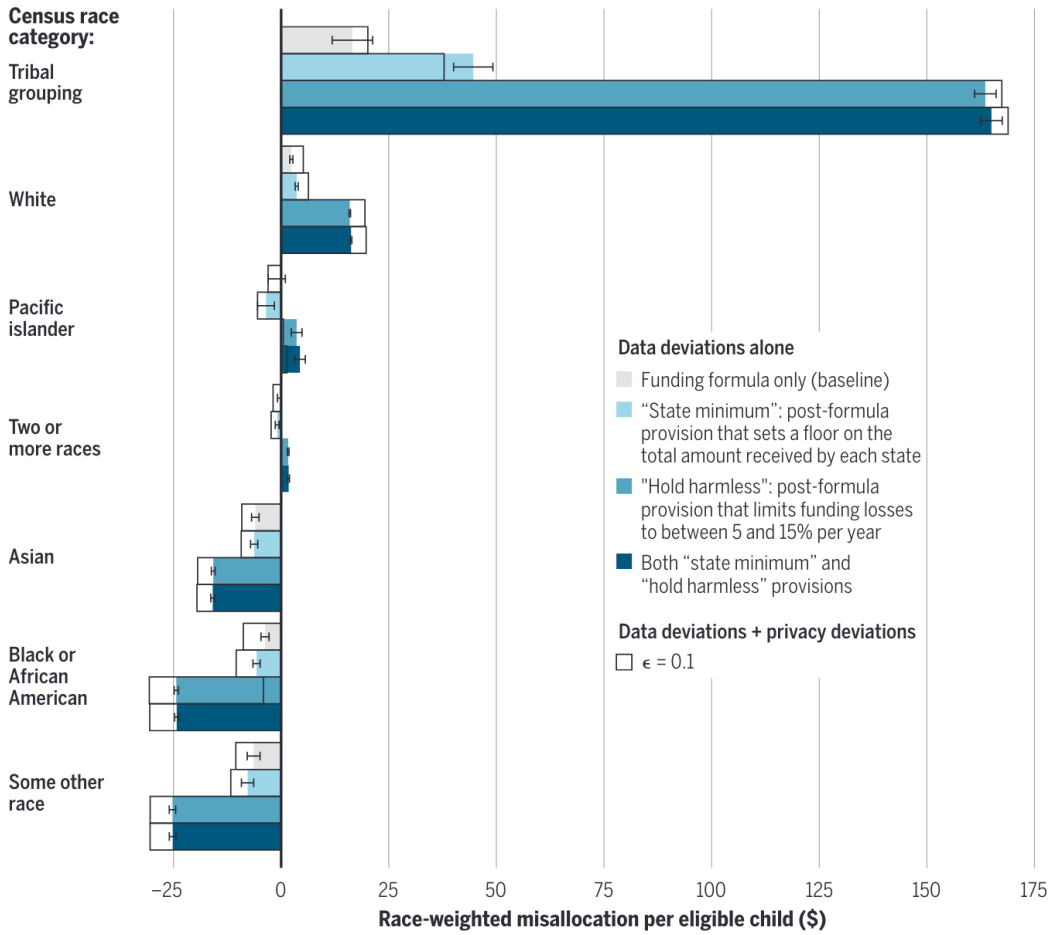
# Results




**"State minimum"** (20 U.S.C. §6333): floor on total state allocation

**"Hold harmless"** (20 U.S.C. §6332): limits funding loss to 5-15%

# Results


Census race category:
Tribal grouping
White
Pacific islander
Two or more races
Asian
Black or African American
Some other race
Data deviations alone
Funding formula only (baseline)
"State minimum": post-formula provision that sets a floor on the total amount received by each state
"Hold harmless": post-formula provision that limits funding losses to between 5 and 15% per year
Both "state minimum" and "hold harmless" provisions
Data deviations + privacy deviations
ϵ = 0.1
−25
0
25
50
75
100
125
150
175
Race-weighted misallocation per eligible child ($)

# Discussion

Choose one deployment from this list:

https://desfontain.es/blog/real-world-differential-privacy.html

and describe how you might attack its utility.

To evaluate utility quantitatively:

1. Choose a dataset to be protected under DP
2. Consider a specific real-world use of that dataset
3. Formulate a metric (broadly defined) that would help understand the utility of the dataset for that specific task
4. Design & run simulations where DP noise is used to protect the data. Compare metrics on the DP-noised data vs. the dataset without protections*

*In practice, it may be hard to get access to the dataset without protections. In such a case, you may simulate the original data first.

# Takeaways

- Utility expresses accuracy's impact on real-world outcomes (among other things, potentially)
- In debates about DP for the U.S. Census, investigations around utility (for redistricting, funding allocation) became cornerstones.
- To evaluate utility quantitatively:
  1. Choose a dataset to be protected under DP
  2. Consider a <u>specific</u> real-world use of that dataset
  3. Formulate a metric (broadly defined) that would help understand the <u>utility</u> of the dataset for that specific task
  4. Design & run simulations where DP noise is used to protect the data. Compare metrics on the DP-noised data vs. the dataset without protections*

*In practice, it may be hard to get access to the dataset without protections. In such a case, you may simulate the original data first.