

# Differential Privacy at the Census

Aloni Cohen

April 20

# Outline

- What is Census and how is it used?
- The 2020 Disclosure Avoidance System
  - TopDown
  - ToyDown
- Analyzing population errors in ToyDown
- Back to TopDown

We the People of  
insure domestic Tranquility, provide for the common defence  
and our Posterity, do ordain and establish this Constitution.

### Article I.

Section. 1. All legislative Powers herein granted shall be vested in a Congress of the United States, which shall consist of a Senate and House of Representatives.

Section. 2. The House of Representatives shall be composed of Members chosen every second Year by the People of the several States, and the Electors in each State shall have the Qualifications requisite for Electors of the most numerous Branch of the State Legislature.

No Person shall be a Representative who shall not have attained to the Age of twenty five Years, and been seven Years a Citizen of the United States, and who shall not, when elected, be an Inhabitant of that State in which he shall be chosen.

Representatives and direct Taxes shall be apportioned among the several States which may be included within this Union, according to their respective Numbers, which shall be determined by adding to the whole Number of free Persons, including those bound to Service for a Term of Years, and excluding Indians not taxed, three fifths of all other Persons. The actual Enumeration shall be made within three Years after the first Meeting of the Congress of the United States, and within every subsequent Term of ten Years, in such Manner as they shall by Law direct. The Number of Representatives shall not exceed one for every thirty Thousand, but each State shall have at least one Representative; and until such Enumeration shall be made, the State of New Hampshire shall be entitled to chuse three; Massachusetts eight; Rhode Island and Providence Plantations one; Connecticut five; New York six; New Jersey four; Pennsylvania eight; Delaware one; Maryland six; Virginia ten; North Carolina five; South Carolina five; and Georgia three.

When vacancies happen in the Representation from any State, the Executive Authority thereof shall issue Writs of Election to fill such Vacancies.

# Disambiguation

## Census Bureau

- a USA government agency

## A census

- an official count or survey of a population

## The decennial census

- a census carried out by the Census Bureau every 10 years
- “Counting everyone once, only once and in the right place”



# Decennial census

- Every 10 years
- Goal: count every person once and in the right place
- Few questions

# American Community Survey (and many other surveys)

- Every year
- Random sample of the population
- Many questions

**5. Please provide information for each person living here. If there is someone living here who pays the rent or owns this residence, start by listing him or her as Person 1. If the owner or the person who pays the rent does not live here, start by listing any adult living here as Person 1.**

**What is Person 1's name? Print name below.**

First Name MI

\_\_\_\_\_

Last Name(s)

\_\_\_\_\_

**6. What is Person 1's sex? Mark  ONE box.**

Male  Female

**7. What is Person 1's age and what is Person 1's date of birth? For babies less than 1 year old, do not write the age in months. Write 0 as the age.**

Age on April 1, 2020 Print numbers in boxes.

\_\_\_\_\_ years Month Day Year of birth

\_\_\_\_\_

**9. What is Person 1's race?**  
Mark  one or more boxes **AND** print origins.

White – Print, for example, German, Irish, English, Italian, Lebanese, Egyptian, etc.

\_\_\_\_\_

Black or African Am. – Print, for example, African American, Jamaican, Haitian, Nigerian, Ethiopian, Somali, etc.

\_\_\_\_\_

American Indian or Alaska Native – Print name of enrolled or principal tribe(s), for example, Navajo Nation, Blackfeet Tribe, Mayan, Aztec, Native Village of Barrow, Inupiat Traditional Government, Nome Eskimo Community, etc.

\_\_\_\_\_

<input type="checkbox"/> Chinese	<input checked="" type="checkbox"/> Vietnamese	<input type="checkbox"/> Native Hawaiian
<input checked="" type="checkbox"/> Filipino	<input type="checkbox"/> Korean	<input type="checkbox"/> Samoan
<input type="checkbox"/> Asian Indian	<input type="checkbox"/> Japanese	<input type="checkbox"/> Chamorro
<input checked="" type="checkbox"/> Other Asian – Print, for example, Pakistani, Cambodian, Hmong, etc. <input checked="" type="checkbox"/>	<input type="checkbox"/> Other Pacific Islander – Print, for example, Tongan, Fijian, Marshallese, etc. <input checked="" type="checkbox"/>	

\_\_\_\_\_

Some other race – Print race or origin.

\_\_\_\_\_

**8. Is Person 1 of Hispanic, Latino, or Spanish origin?**

**NOTE:** Please answer BOTH Question 8 about Hispanic origin and Question 9 about race. For this census, Hispanic origins are not races.

**19. About how much do you think this house and lot, apartment, or mobile home (and lot, if owned) would sell for if it were for sale?**

Amount – Dollars

\$ \_\_\_\_\_ .00

**20. What are the annual real estate taxes on THIS property?**

Annual amount – Dollars

\$ \_\_\_\_\_ .00

**OR**

None

**21. What is the annual payment for fire, hazard, and flood insurance on THIS property?**

Annual amount – Dollars

\$ \_\_\_\_\_ .00

**OR**

None

**22. a. Do you or any member of this household have a mortgage, deed of trust, contract to purchase, or similar debt on THIS property?**

Yes, mortgage, deed of trust, or similar debt

**23. a. Do you or any member of this household have a second mortgage or a home equity loan on THIS property?**

Yes, home equity loan

Yes, second mortgage

Yes, second mortgage and home equity loan

No → SKIP to **D**

**b. How much is the regular monthly payment on all second or junior mortgages and all home equity loans on THIS property?**

Monthly amount – Dollars

\$ \_\_\_\_\_ .00

**OR**

No regular payment required

**D** Answer question 24 if this is a MOBILE HOME. Otherwise, SKIP to **E**.

**HISPANIC OR LATINO ORIGIN**

Survey/Program: Decennial Census  
 Years: 2010  
 Table: P4

**AGE OF GRANDCHILDREN UNDER 18 YEARS LIVING WITH A GRANDPARENT HOUSEHOLDER**

Survey/Program: Decennial Census  
 Years: 2010  
 Table: P41

**GROUP QUARTERS POPULATION BY GROUP QUARTERS TYPE**

Survey/Program: Decennial Census  
 Years: 2010  
 Table: P42

**GROUP QUARTERS POPULATION BY SEX BY AGE BY GROUP QUARTERS TYPE**

Survey/Program: Decennial Census  
 Years: 2010  
 Table: P43

**POPULATION SUBSTITUTED**

Survey/Program: Decennial Census  
 Years: 2010  
 Table: P44

**ALLOCATION OF POPULATION ITEMS**

Survey/Program: Decennial Census  
 Years: 2010  
 Table: P45

**ALLOCATION OF RACE**

Survey/Program: Decennial Census  
 Years: 2010  
 Table: P46

Label	Congressional District 1 (113th Congress), Florida	Congressional District 2
▼ Total:	29,238	
▼ Male:	22,242	
▼ Under 18 years:	575	
▼ Institutionalized population (101-106, 201-203, 301, 401-405):	463	
Correctional facilities for adults (101-106)	97	
Juvenile facilities (201-203)	366	
Nursing facilities/Skilled-nursing facilities (301)	0	
Other institutional facilities (401-405)	0	
▼ Noninstitutionalized population (501, 601-602, 701-702, 704, 706, 801-802, 900...)	112	
College/University student housing (501)	26	
Military quarters (601-602)	5	
Other noninstitutional facilities (701-702, 704, 706, 801-802, 900-901, 903-9...)	81	
▼ 18 to 64 years:	20,883	
▼ Institutionalized population (101-106, 201-203, 301, 401-405):	10,816	
Correctional facilities for adults (101-106)	10,478	
Juvenile facilities (201-203)	62	
Nursing facilities/Skilled-nursing facilities (301)	226	
Other institutional facilities (401-405)	50	
▼ Noninstitutionalized population (501, 601-602, 701-702, 704, 706, 801-802, 900...)	10,067	
College/University student housing (501)	2,015	
Military quarters (601-602)	7,160	
Other noninstitutional facilities (701-702, 704, 706, 801-802, 900-901, 903-9...)	892	
▼ 65 years and over:	784	
▼ Institutionalized population (101-106, 201-203, 301, 401-405):	714	
Correctional facilities for adults (101-106)	112	
Juvenile facilities (201-203)	0	

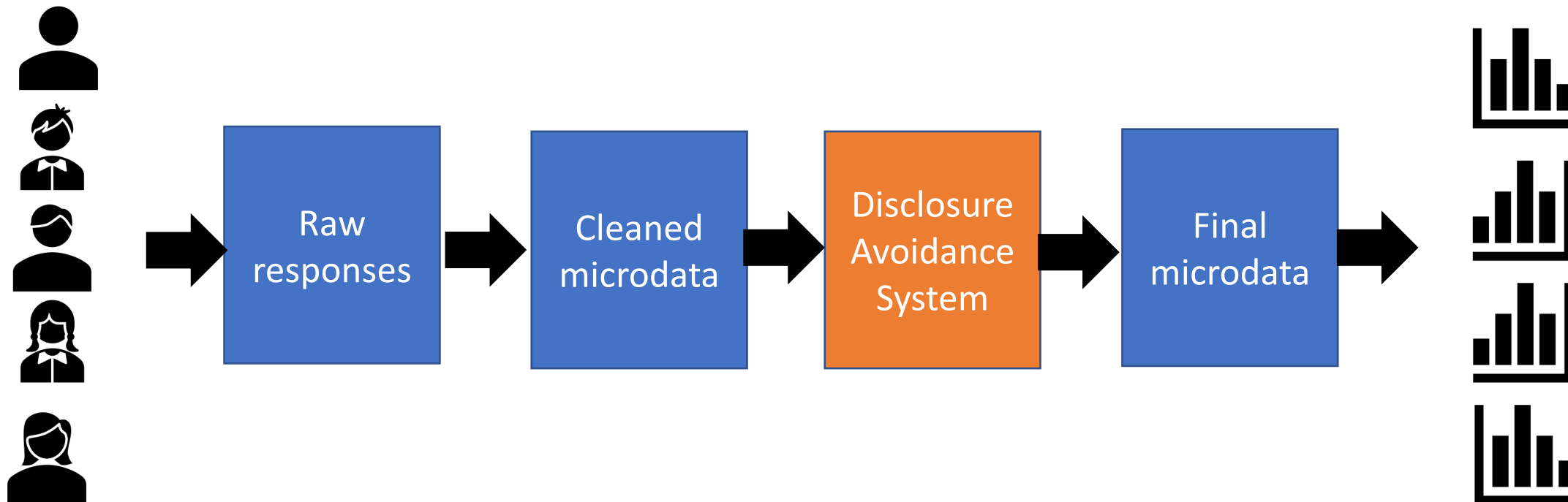
## The Census Confidentiality Protection Pledge:

ACCESS of WNY Advancement Project African American Ministers In Action Alliance for African Assistance American Civil Liberties Union American Constitution Society American Federation of Teachers American-Arab Anti-Discrimination Committee APIAVote Michigan Arab American Association of New York Page 5 of 9 Arab American Institute AREAA Arkansas United Asian & Pacific Islander American Health Forum Asian American Action Fund Asian American Federation Asian American Federation of Florida - South Region Asian American Legal Defense and Education Fund (AALDEF) Asian American Organizing Project Asian Americans Advancing Justice - AAJC Asian Americans Advancing Justice - Asian Law Caucus Asian Americans Advancing Justice - Atlanta Asian Americans Advancing Justice - Los Angeles Asian Americans Advancing Justice | Chicago Asian Americans United Asian and Pacific Islander American Vote Asian Community Resource Center Asian Counseling and Referral Service Asian Pacific American Institute for Congressional Studies (APAIACS) Asian Pacific American Labor Alliance Asian Pacific Development Center Blackstone Valley Tourism Council BPSOS, Inc. Brennan Center for Justice at NYU Law Brighton Park Neighborhood Council CAIR-Chicago California Calls Capeverdean American Community Development CASA Center for Civic Policy Center for Law and Social Justice at Medgar Evers College Center for Law and Social Policy (CLASP) Center for Southeast Asians Center for Urban Research, CUNY Graduate Center Chinese American Service League CIANA Citizens Committee for Children Coalición de Derechos Humanos Common Cause Page 6 of 9 Community Organizing and Family Issues (COFI) Community Resource Exchange (CRE) Crescent City Media Group/Center for Civic Action Demand Progress Education Fund Demos Dolores Huerta Foundation EMBARC Empowering Pacific Islander Communities (EPIC) Enlace Chicago Equality California Fair Count Fair Fight Action Faith and Community Empowerment (formerly KCCD) Fathers Who Care Florida Asian Services Florida Asian Women Alliance Florida Counts Census 2020 Forefront Fred T. Korematsu Institute Georgia

Association of Latino Elected Officials (GALEO) Girls Inc. Global Cleveland GreenLatinos Greenville NAACP #5522 Gum Moon Residence Hall Hispanic Federation Hispanics in Philanthropy Illinois Coalition for Immigrant and Refugee Rights Indian Horizon of Florida Inter University Program on Latino Research Japanese American Citizens League Japanese American Service Committee Justice for Chinedu Kentucky Youth Advocates Keystone Counts Korean Community Center of the East Bay Labor Council for Latin American Advancement Lambda Legal Page 7 of 9 Laotian American National Alliance Latino Community Fund (LCF Georgia) Latino Policy Forum LatinoJustice PRLDEF Lawyers' Committee for Civil Rights Under Law League of Conservation Voters League of United Latin American Citizens League of Women Voters of Connecticut League of Women Voters of the United States Lim Consulting Services Los Angeles United Methodist Urban Foundation Los Cien Sonoma Inc. Los Vecinos de Buford Highway LUNA Make the Road New Jersey MALDEF (Mexican American Legal Defense and Educational Fund) Market Insight Corporation Massachusetts Census Equity Fund MassCounts Mi Familia Vota Minnesota Council on Foundations Montana Women Vote Multicultural Media, Telecom and Internet Council (MMTC) NAACP NAACP Legal Defense and Educational Fund NANAY CEDC National Asian Pacific American Women's Forum (NAPAWF) National Association of Latino Elected and Appointed Officials (NALEO) Educational Fund National Association of Social Workers National Bar Association & President Alfreda Robinson National Center for Transgender Equality National Council of Asian Pacific Americans (NCAPA) National Council of Negro Women National Disability Rights Network National Education Association National Federation of Filipino American Associations National Hispanic Leadership Agenda National Hispanic Media Coalition Page 8 of 9 National Immigration Law Center National Korean American Service & Education Consortium (NAKASEC) National Latinx Psychological Association National LGBTQ Task Force National Network for Immigrant & Refugee Rights National Partnership for

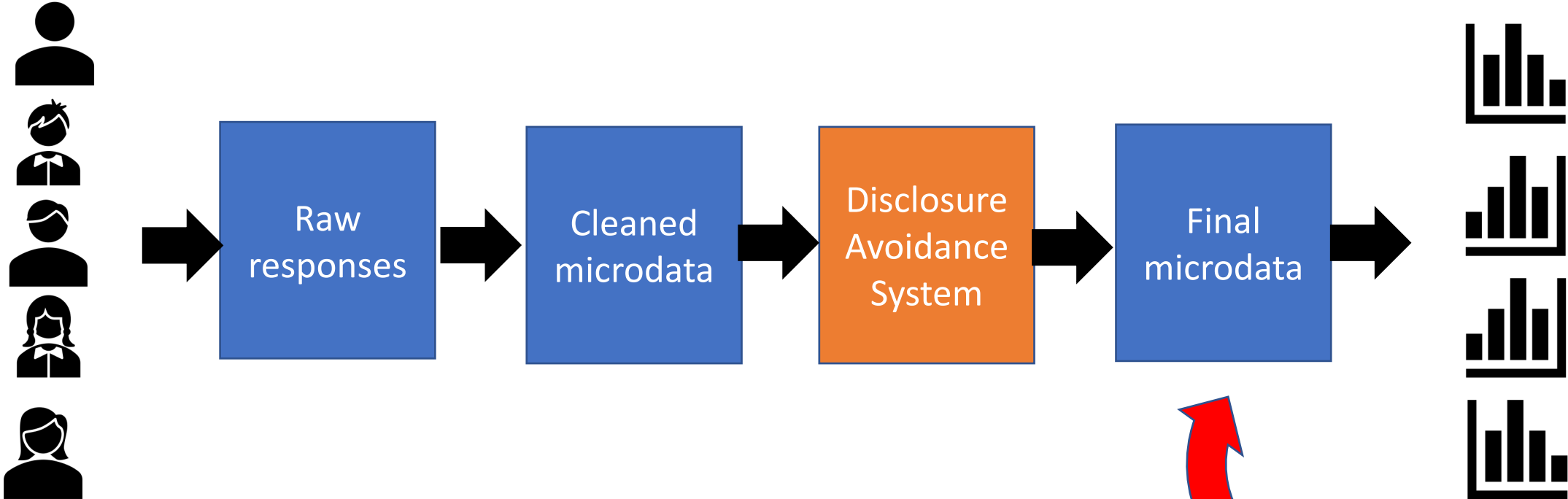
Women & Families National Redistricting Foundation National Tongan American Society National Urban Indian Family Coalition National Urban League Native American Rights Fund NETWORK Lobby for Catholic Social Justice New American Welcome Center at the University YMCA New Jersey Institute for Social Justice New York Immigration Coalition NextDayBetter NH Office of Strategic Initiatives NOBCO OCA South Florida Chapter OCA-Asian Pacific American Advocates Opportunity Council Orange County Asian and Pacific Islander Community Alliance (OCAPICA) Orange County Equality Coalition Pacific Asian Counseling Center Partnership for America's Children Partnership for the Advancement of New Americans People For the American Way PFLAG National PICO California Prison Policy Initiative Sanchez, Mowrer & Desiderio PC SBCC Thrive LA SER Jobs for Progress National Inc. Service Employees International Union (SEIU) Sierra Club Sikh American Legal Defense and Education Fund (SALDEF) Sikh Coalition South Asian Americans Leading Together Page 9 of 9 South Asian Bar Association of North America South Asian Network (SAN) Southeast Asia Resource Action Center (SEARAC) Southern Coalition for Social Justice Southern Poverty Law Center Action Fund Southwest Voter Registration Education Project Spanish Community Center Stanislaus County Office of Education Texas Progressive Action Network The California Endowment The Fund for New Jersey The Hyams Foundation The Leadership Conference Education Fund The Leadership Conference on Civil and Human Rights The National Queer Asian Pacific Islander Alliance (NQAPIA) The Resurrection Project Together We Count UAO Chapter of AKA UNAVSA UnidosUS Union for Reform Judaism United Cambodian Community United We Dream Univision Communications Inc. Vaughn Next Century Learning Center Virginia Civic Engagement Table Virginia Counts Coalition Voto Latino Foundation WAVE (Women for American Values and Ethics) William C. Velasquez Institute YWCA USA

# The data processing pipeline



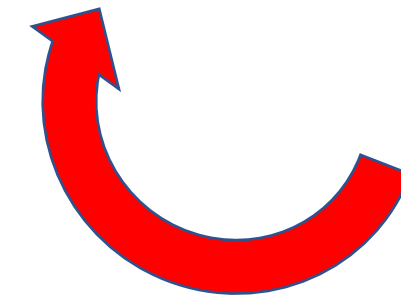


# 2010 reconstruction demonstration



## *To Reduce Privacy Risks, the Census Plans to Report Less Accurate Data*

Guaranteeing people's confidentiality has become more of a challenge, but some scholars worry that the new system will impede research.



Reconstruction:  
- Solving a constrained optimization problem



# How Census data is used

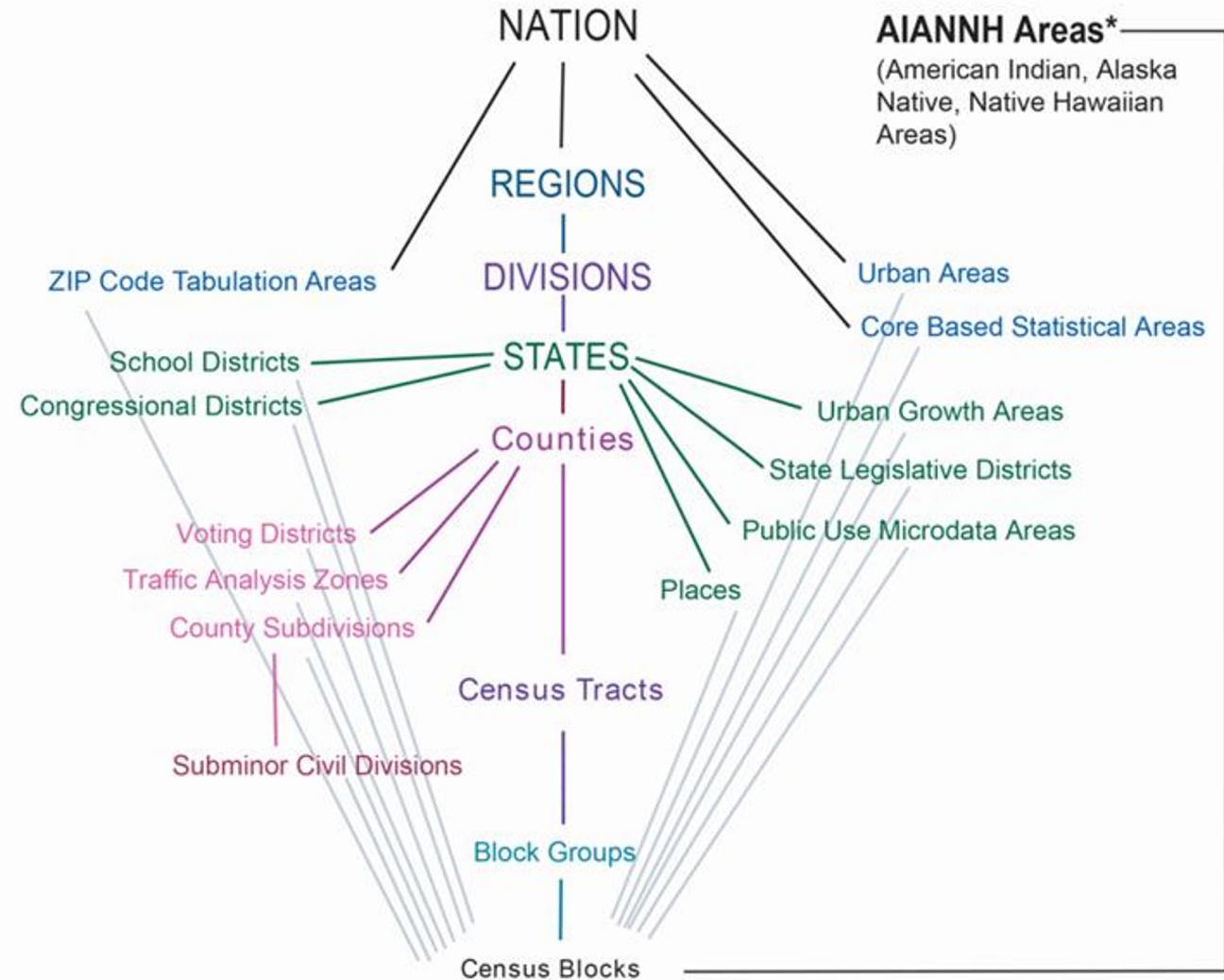
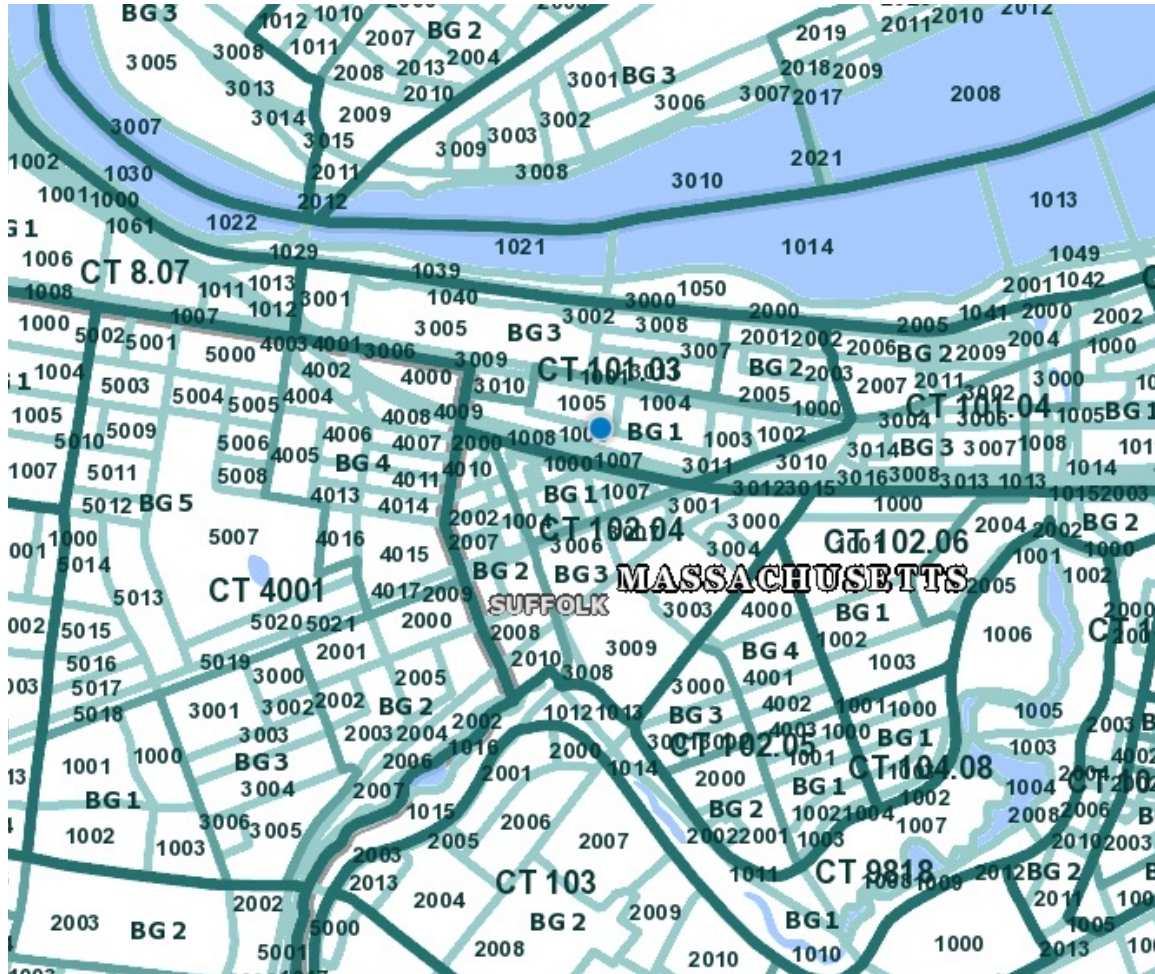
## Official Uses

- Congressional apportionment data
- Redistricting\*
- Designing representative samples for surveys
- Funding formulas

## Other uses

- Disaster planning
- Facility placement
- Budget
- Demographics / social science / economic research
  - How do food deserts affect economic mobility?
- ...

# The Census Hierarchy



# Redistricting

Every 10 years, across the country

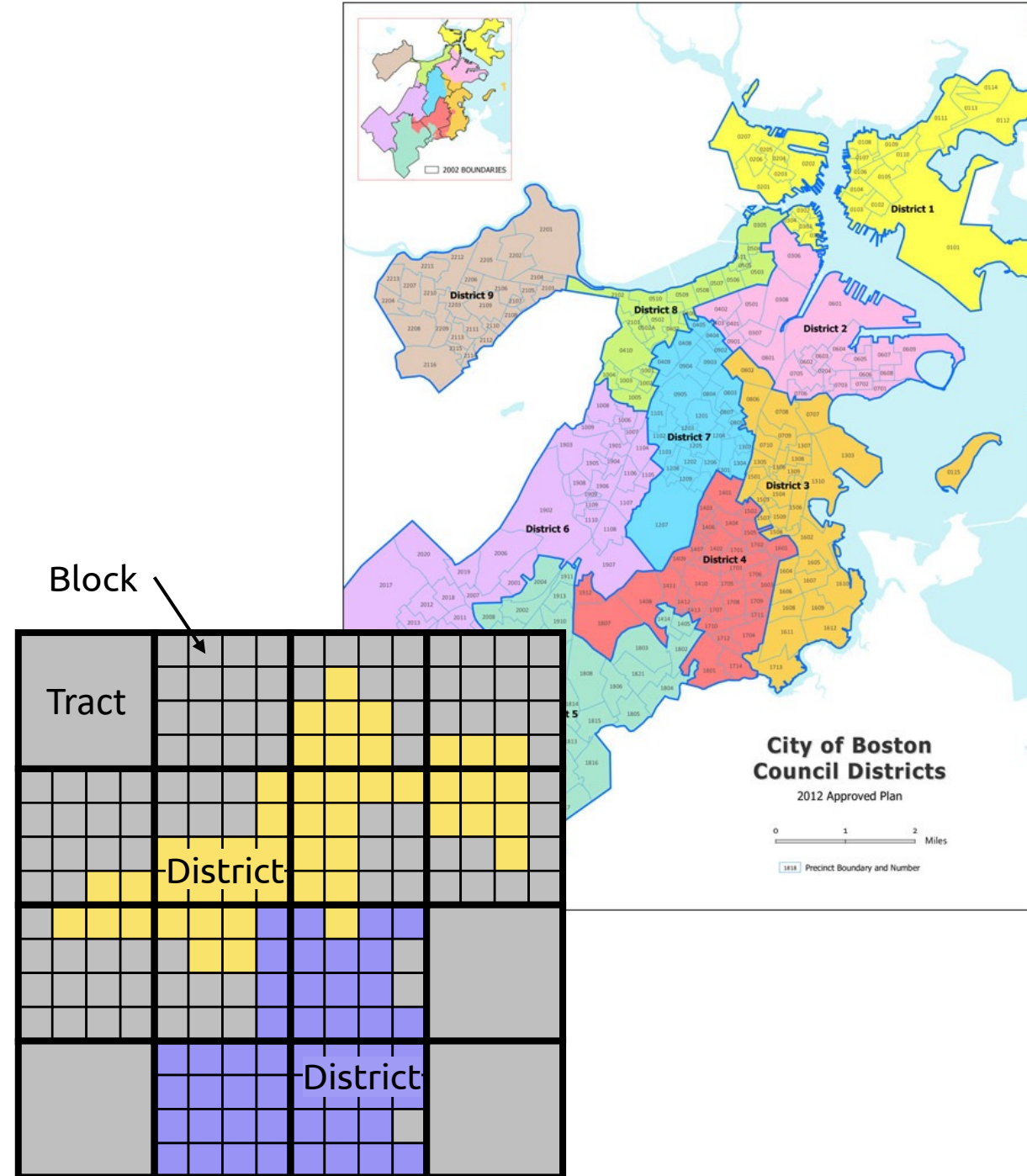
Built out of individual census blocks

Many different types / sizes

- Congress, state legislature, city council, school districts, ...

One person one vote

- Congressional districts are typically **balanced to 1 person** according to last decennial census
- Supreme Court has thrown out district plan for not justifying an 18 person deviation (Karcher)
- Other districts: 5—10% balance





# Racial gerrymandering

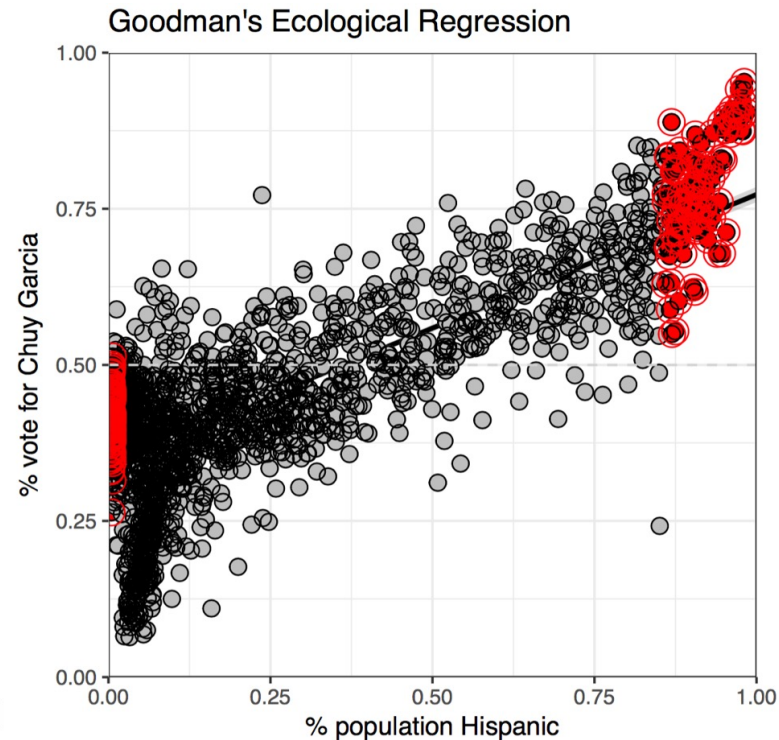
Strategically drawing districts to weaken the political power of a racial group

Illegal (unlike partisan gerrymandering)

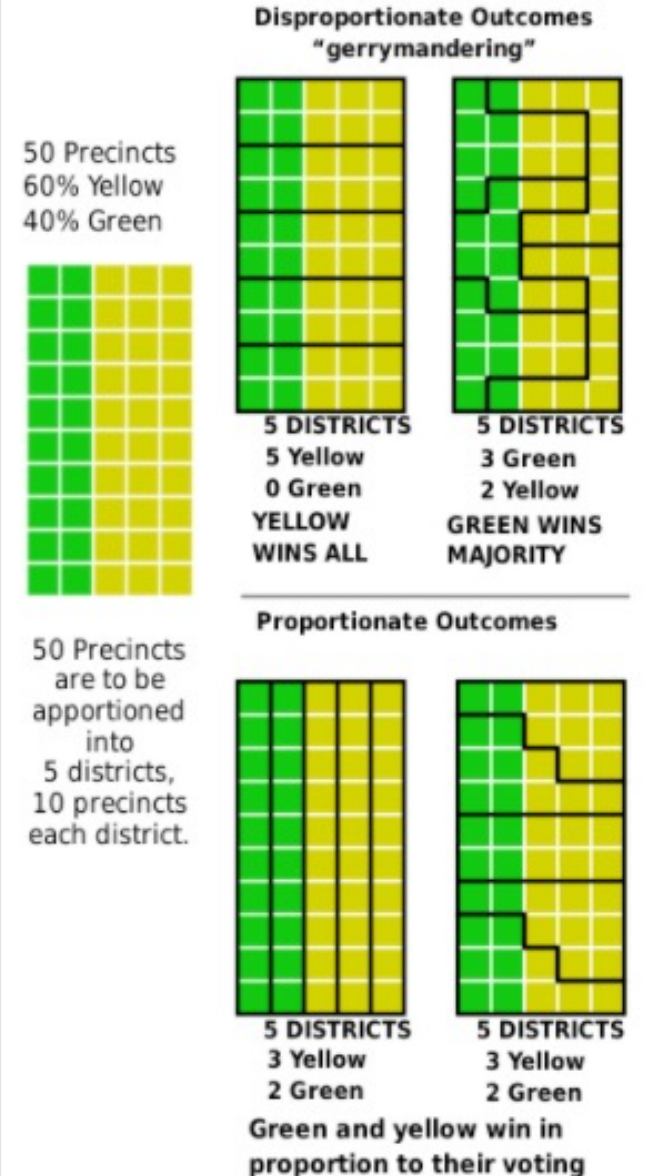
Gingles test

- Can you redraw the districts so that the racial group is a majority in more districts?
- Do voters of different races vote differently?
  - Regression using census data + election data

Chicago 2015 mayoral runoff (Rahm-Chuy)



Gerrymandering: drawing different maps for electoral districts produces different outcomes



Different ways to apportion electoral districts

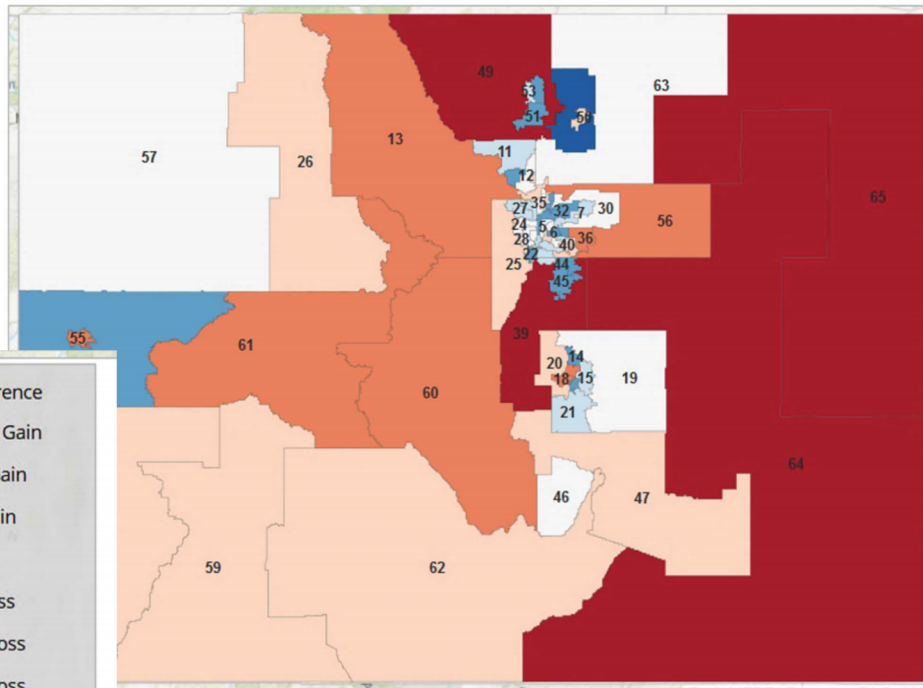


# The problem with DP (and every other source of noise)

## One Person One Vote

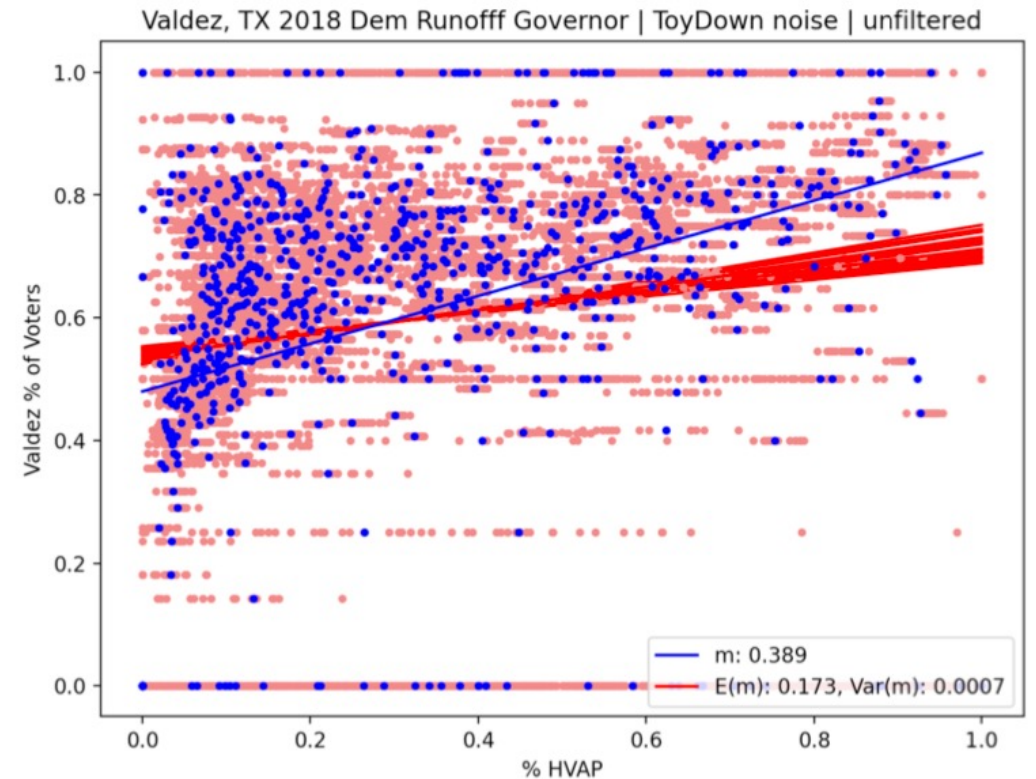
Can't draw equal population districts

Figure 4  
Population Differences across Colorado House Districts\*



## Racial gerrymandering

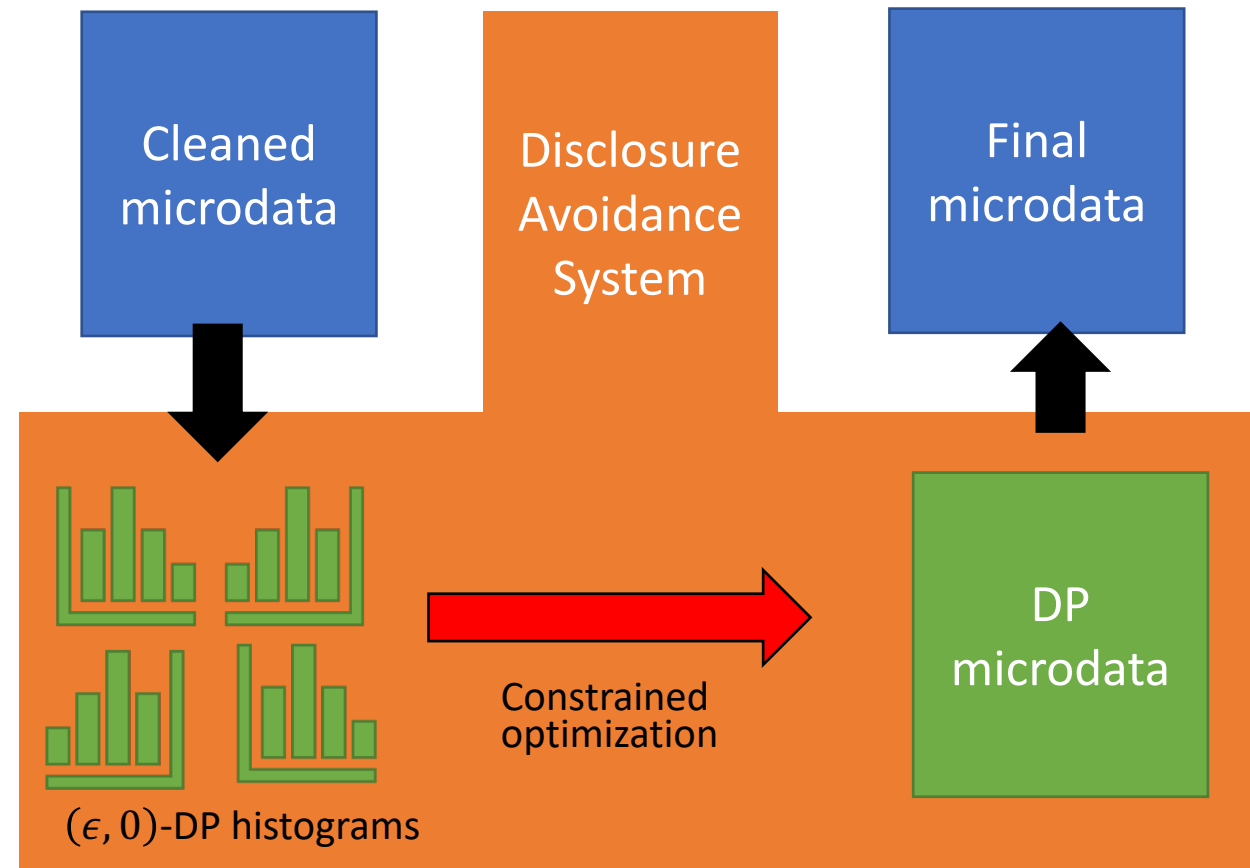
Noise attenuates linear regression signal





# Goals for DP disclosure avoidance system

- $(\epsilon, 0)$ -DP
- Tunable workload + budget
- DP microdata, not just raw statistics
  - $\Rightarrow$  all counts are non-negative
- Exact state-level populations
- Sanity constraints
  - Exact number of group quarters
  - $\#householders < \#households$
  - Children younger than parents
- Parallelizable



# The TopDown Algorithm

- Goals
- $(\epsilon, 0)$ -DP
- Tunable workload + budget
- DP microdata, not just raw statistics
- Exact state-level populations
- Sanity constraints
  - Exact number of group quarters
  - #householders < #households
  - Children younger than parents
- Parallelizable

## Inputs

- Census Hierarchy:  $H = (H_1, \dots, H_d)$ 
  - $h \in H_\ell$  is a unit of geography
  - E.g., tract, block, ...
- Workload of histograms  $Q_1, Q_2, \dots$ 
  - $q \in Q$  is a single histogram bin
  - E.g., number of 25yo male Black Hispanics in tract 123456
- Budget allocation:  $\epsilon$ 
  - Fraction  $\beta_\ell$  that level  $H_\ell$  gets
  - Fraction  $\beta_Q$  that histogram  $Q$  gets
- Dataset:  $\{a_h\}_{h \in H}$

## Stage 1: Noise infusion

- For every  $h \in H_\ell$  and every  $q \in Q$ :
  - $\tilde{a}_{h,q} \leftarrow \text{GeomMech}(q(\{a_h\}), \epsilon \cdot \beta_\ell \cdot \beta_Q)$

## Stage 2: Post-processing

- For each level  $\ell = 1, \dots, d$  from top:
  - Compute  $\{\tilde{a}_h\}_{h \in H_\ell}$
  - Minimizing  $\sum_{h \in H_\ell} \sum_q (q(\{a_h\}) - \tilde{a}_{h,q})^2$
  - Subject to
    - Consistent with level  $\ell - 1$
    - Non-negative
    - Lots of other constraints

## DP analysis

- Parallel composition
- Post-processing

# ToyDown: Simplifying TopDown

## Inputs

- Census Hierarchy:  $H = (H_1, \dots, H_d)$ 
  - $h \in H_\ell$  is a unit of geography
  - E.g., tract, block, ...
- Workload of histograms  $Q_1, Q_2, \dots$ 
  - $q \in Q$  is a single histogram bin
  - E.g., number of 25yo male Black Hispanics in tract 123456
- Budget allocation:  $\epsilon$ 
  - Fraction  $\beta_\ell$  that level  $H_\ell$  gets
  - Fraction  $\beta_Q$  that histogram  $Q$  gets
- Dataset:  $\{a_h\}_{h \in H}$

## Stage 1: Noise infusion

- For every  $h \in H_\ell$  and every  $q \in Q$ :
  - $\tilde{a}_{h,q} \leftarrow \text{Laplace}(q(\{a_h\}), \epsilon \cdot \beta_\ell \cdot \beta_Q)$

## Stage 2: Post-processing

- For each level  $\ell = 1, \dots, d$  from top:
  - Compute  $\{\tilde{a}_h\}_{h \in H_\ell}$
  - Minimizing  $\sum_{h \in H_\ell} \sum_q (q(\{\alpha_h\}) - \tilde{a}_{h,q})^2$
  - Subject to
    - Consistent with level  $\ell - 1$
    - Non-negative
    - Lots of other constraints

## DP analysis

- Parallel composition
- Post-processing

# ToyDown: Simplifying TopDown

## Inputs

- Census Hierarchy:  $H = (H_1, \dots, H_d)$ 
  - $h \in H_\ell$  is a unit of geography
  - E.g., tract, block, ...
- Budget allocation:  $(\epsilon_1, \dots, \epsilon_d)$
- Dataset:  $\{a_h\}_{h \in H}$

## Stage 1: Noise infusion

- For every  $h \in H_\ell$ 
  - $L_h \sim \text{Lap}\left(\frac{2}{\epsilon_\ell}\right)$
  - $\tilde{a}_h \leftarrow a_h + L_h$

## Stage 2: Post-processing

- For each level  $\ell = 1, \dots, d$  from top:
  - Compute  $\{\alpha_h\}_{h \in H_\ell}$
  - Minimizing  $\sum_{h \in H_\ell} (\alpha_h - \tilde{a}_h)^2$
  - Consistent with level  $\ell - 1$ 
    - Let  $\hat{h}$  be the parent node of  $h$
    - $\sum_{h' \in \text{Kids}(\hat{h})} \alpha_{h'} = \alpha_{\hat{h}}$

Output  $\{\alpha_h\}_{h \in H_L}$

# Plan: How does the error behave in ToyDown?

Total population error for a district

- District  $D$ , a collection of blocks
- $E_D = \sum_{h \in D} E_h = \sum_{h \in D} \alpha_h - a_h$

Low population error districts  $\rightarrow$   
Equal population districts

As a function of budget allocation  
across levels

- $\epsilon = \epsilon_1 + \epsilon_2 + \dots + \epsilon_d$
- Where to put the budget?

As a function of a district's shape

- How to draw low-error districts?



# Allocating budget across levels

Recall census spine:

- Nation, state, county, tract, block group, block

For the most accurate block-level counts, should we...

- Put more budget at the top (Vote: Thumbs up)
- Put more budget at the bottom (Vote: Clap)
- Something else (Vote: No)

# Notation

Hierarchy:  $H = (H_1, \dots, H_d)$

- Depth  $d$  tree
- Node:  $h \in H_\ell$ 
  - Children:  $h_1, h_2, \dots, h_{n_\ell}$
  - Parent:  $\hat{h}$

District:  $D \subset H_d$

Error:

- $E_h = \alpha_h - a_h$
- $E_D = \sum_{h \in D} E_h$

## Stage 1: Noise infusion

For every  $h \in H_\ell$

- $L_h \sim \text{Lap}\left(\frac{2}{\epsilon_\ell}\right)$
- $\tilde{a}_h \leftarrow a_h + L_h$

## Stage 2: Post-processing

For each level  $\ell = 1, \dots, d$  from top:

- Compute  $\{\alpha_h\}_{h \in H_\ell}$
- Minimizing  $\sum_{h \in H_\ell} (\alpha_h - \tilde{a}_h)^2$
- s.t.  $\sum_{i \in [n_\ell]} \alpha_{\hat{h}_i} = \alpha_{\hat{h}}$

(Simplifying: each  $h \in H_\ell$  has exactly  $n_\ell$  children)

# The root's error

Let  $h = 1$  be the root

$$L_1 \sim \text{Lap}\left(\frac{2}{\epsilon_1}\right)$$

No consistency constraint

$$\Rightarrow \alpha_1 = \tilde{a}_1 = a_1 + L_1$$

$$\text{Error: } E_1 = \alpha_1 - a_1 = L_1$$

$$\text{Expectation: } \mathbb{E}(E_1) = \mathbb{E}(L_1) = 0$$

$$\text{Variance: } \mathbb{V}(E_1) = 2 \left(\frac{2}{\epsilon_1}\right)^2 = \frac{8}{\epsilon_1^2}$$

## Stage 1: Noise infusion

For every  $h \in H_\ell$

- $L_h \sim \text{Lap}\left(\frac{2}{\epsilon_\ell}\right)$
- $\tilde{a}_h \leftarrow a_h + L_h$

## Stage 2: Post-processing

For each level  $\ell = 1, \dots, d$  from top:

- Compute  $\{\alpha_h\}_{h \in H_\ell}$
- Minimizing  $\sum_{h \in H_\ell} (\alpha_h - \tilde{a}_h)^2$
- s.t.  $\sum_{i \in [n_\ell]} \alpha_{\hat{h}_i} = \alpha_{\hat{h}}$

# Level $\ell$ error

## Thm

For non-root  $h_i$  on level  $\ell + 1$

$$E_{h_i} = L_{h_i} + \frac{1}{n_\ell} \left( E_h - \sum_j L_{h_j} \right)$$

**Observation:** error trickles down, but weakly

## Stage 1: Noise infusion

For every  $h \in H_\ell$

- $L_h \sim \text{Lap} \left( \frac{2}{\epsilon_\ell} \right)$
- $\tilde{a}_h \leftarrow a_h + L_h$

## Stage 2: Post-processing

For each level  $\ell = 1, \dots, d$  from top:

- Compute  $\{\alpha_h\}_{h \in H_\ell}$
- Minimizing  $\sum_{h \in H_\ell} (\alpha_h - \tilde{a}_h)^2$
- s.t.  $\sum_{i \in [n_\ell]} \alpha_{\hat{h}_i} = \alpha_{\hat{h}}$

## Proof

$$L_{h_i} \sim \text{Lap} \left( \frac{2}{\epsilon_{\ell+1}} \right)$$

## Consistency

- Want to find  $(\alpha_{h_1}, \dots, \alpha_{h_{n_\ell}})$  closest to  $(\tilde{a}_{h_1}, \dots, \tilde{a}_{h_{n_\ell}})$   
s.t.  $\sum_j \alpha_{h_j} = \alpha_h$
- Projecting a point to a plane
- Solution: divide the error equally among the terms

$$\alpha_{h_i} = \tilde{a}_{h_i} + \frac{1}{n_\ell} \left( \alpha_h - \sum_j \tilde{a}_{h_j} \right)$$

Substitute  $\tilde{a}_{h'} = a_{h'} + L_{h'}$ :

- $$\begin{aligned} \alpha_{h_i} &= (a_{h_i} + L_{h_i}) + \frac{1}{n_\ell} \left( \alpha_h - \sum_j (a_{h_j} + L_{h_j}) \right) \\ &= a_{h_i} + L_{h_i} - \frac{1}{n_\ell} \sum_j L_{h_j} + \frac{1}{n_\ell} (\alpha_h - a_h) \end{aligned}$$

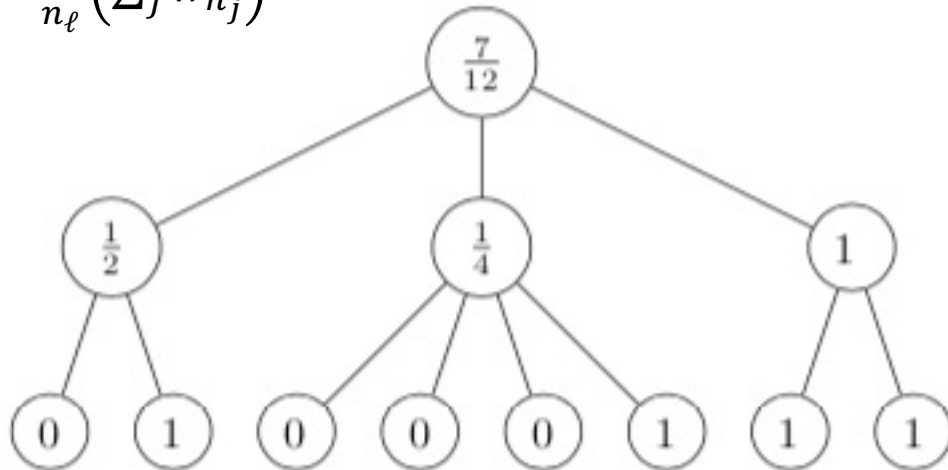
Rearrange, and substitute  $E_{h'} = \alpha_{h'} - a_{h'} = \alpha_h - a_h - \sum_j \tilde{a}_{h_j}$

# District error

## Weight of a node for district $D$

Fraction of subtree contained in the district

- For leaf  $h \in H_d$ :
  - $w_h = 1$  if  $h \in D$
  - $w_h = 0$  if  $h \notin D$
- For non-leaf  $h \in H_{<\ell}$ :
  - $w_h = \frac{1}{n_\ell} \left( \sum_j w_{h_j} \right)$



## Stage 1: Noise infusion

For every  $h \in H_\ell$

- $L_h \sim \text{Lap}\left(\frac{2}{\epsilon_\ell}\right)$
- $\tilde{a}_h \leftarrow a_h + L_h$

## Stage 2: Post-processing

For each level  $\ell = 1, \dots, d$  from top:

- Compute  $\{\alpha_h\}_{h \in H_\ell}$
- Minimizing  $\sum_{h \in H_\ell} (\alpha_h - \tilde{a}_h)^2$
- s.t.  $\sum_{i \in [n_\ell]} \alpha_{\hat{h}_i} = \alpha_{\hat{h}}$

## Thm

$$E_D = w_1 L_1 + \sum_{h \in H \setminus \{1\}} (w_h - w_{\hat{h}}) L_h$$

## Proof idea

- Prove general error for weighted sum of nodes on level  $\ell$  by induction + collecting terms

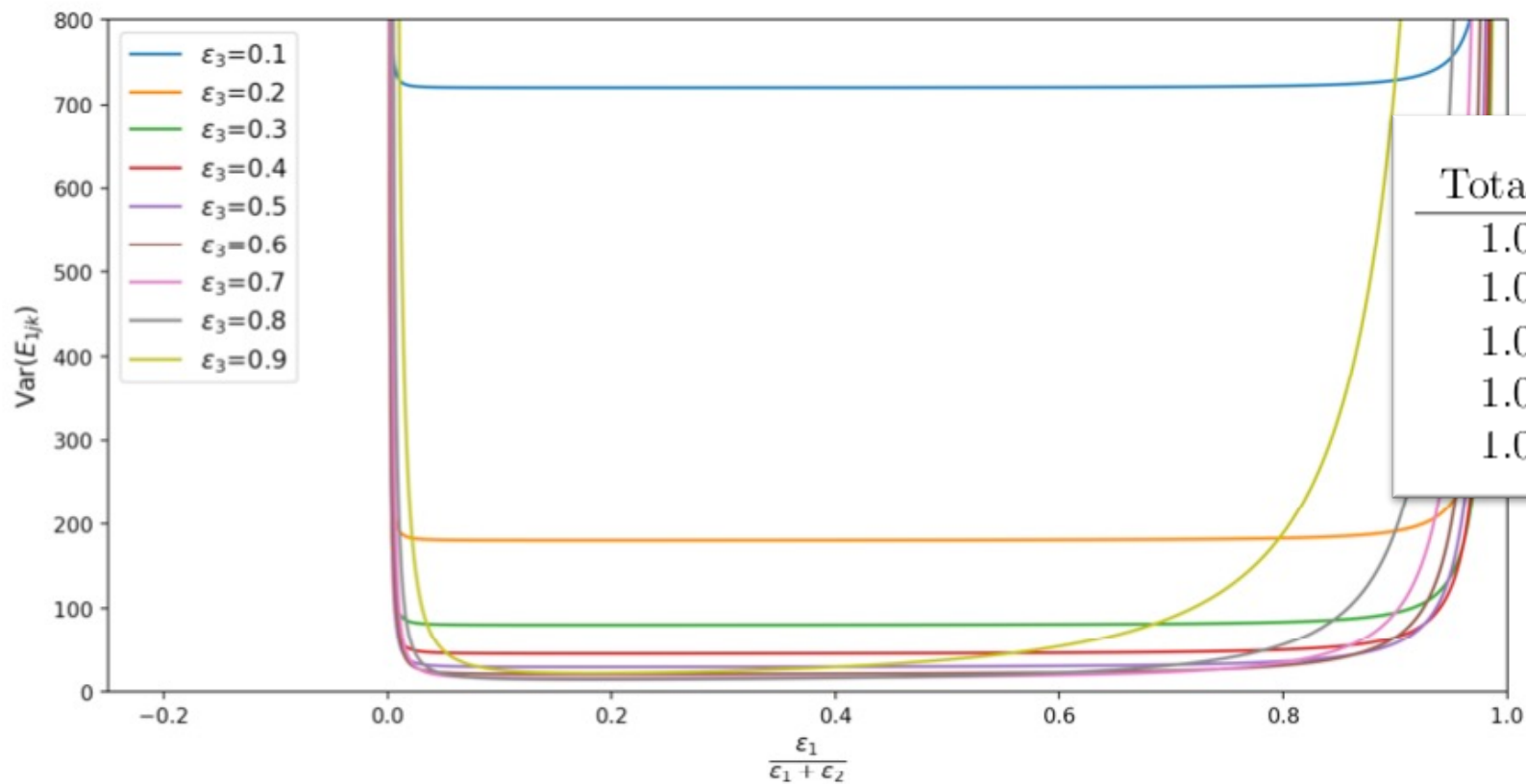


# Error for a single block

- $\mathbb{V}(E_{block}) = \frac{8w_1^2}{\epsilon_1^2} + \sum_{\ell=2,\dots,d} \left( \frac{8}{\epsilon_\ell^2} \sum_{h \in H_\ell} (w_h - w_{\hat{h}})^2 \right)$
- $w_h = \frac{1}{n_\ell \cdots n_{d-1}}$  if block is in subtree  $h$
- $w_h = 0$  if block is not in subtree

$$\mathbb{V}(E_{block}) \approx \sum_{\ell=1,\dots,d} \frac{8}{\epsilon_\ell^2 (n_\ell \cdots n_{d-1})^2}$$

Variance for single block:  $\sum_{\ell=1,\dots,d} \frac{8}{\epsilon_{\ell}^2 (n_{\ell} \cdots n_{d-1})^2}$



Total $\epsilon$	Allocation	$L^1$ error
1.0	(.16, .16, .16, .16, .16, .2)	0.03
1.0	(.2, .16, .16, .16, .16, .16)	0.03
1.0	(.1, .1, .1, .1, .1, .5)	0.02
1.0	(.02, .02, .02, .02, .02, .9)	0.03
1.0	(.66, .30, .01, .01, .01, .01)	0.09

**$L_1$ -error in Topdown**

A small selection of TopDown results on reconstructed Texas data from the 2010 Census. The allocation  $(\epsilon_1, \dots, \epsilon_6)$  goes from the nation to census blocks.

Figure 2: Error variance under ToyDown for a leaf node in the three-level hierarchy with  $n_1 = n_2 = 10$  and  $\epsilon = 1$ . The curves show varying  $\epsilon_3$  (colors) and the relative balance of  $\epsilon_1$  and  $\epsilon_2$  ( $x$ -axis). For these parameters, the minimum variance is 14.52 and occurs at roughly  $(\epsilon_1, \epsilon_2, \epsilon_3) = (0.038, 0.171, 0.791)$ .

# Allocating budget across levels

Recall census spine:

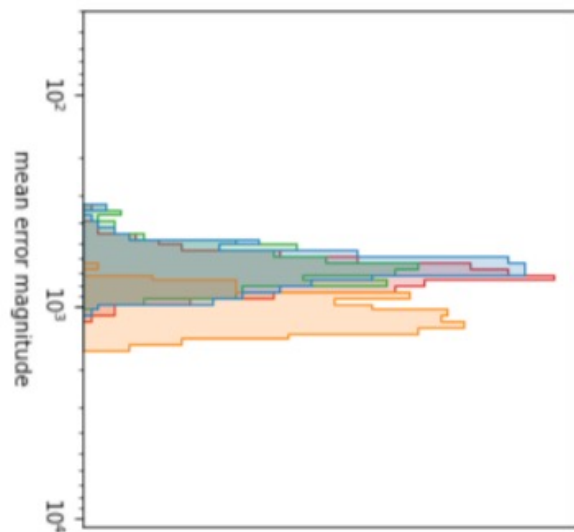
- Nation, state, county, tract, block group, block

For the most accurate **district-level** counts, should we...

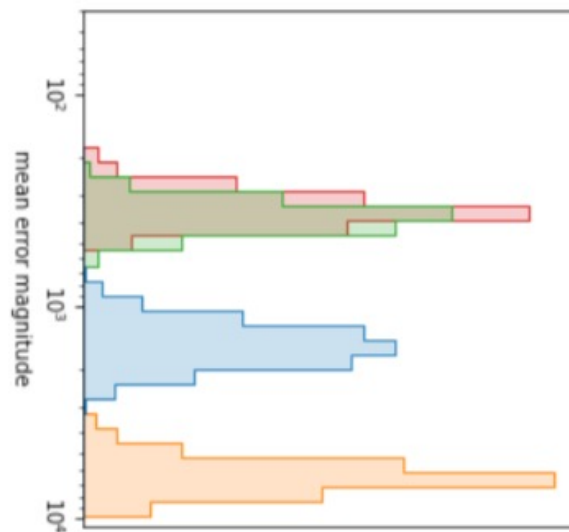
- Put more budget at the top (Vote: Thumbs up)
- Put more budget at the bottom (Vote: Clap)
- Something else (Vote: No)

# District level errors in ToyDown, experiments

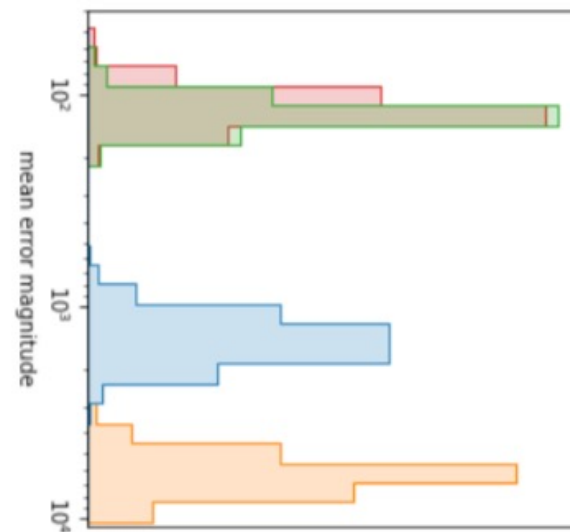
block-heavy



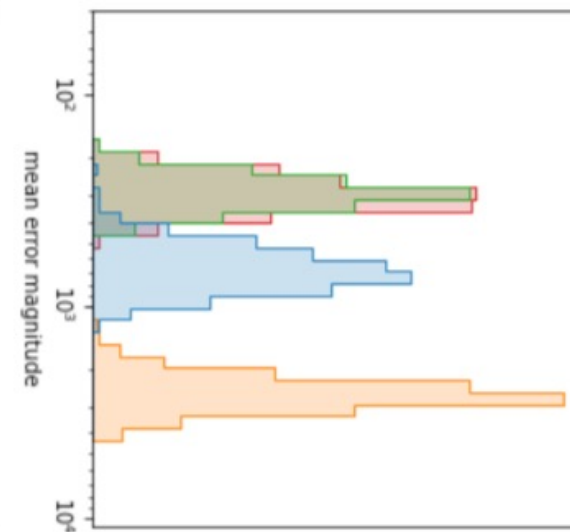
BG-heavy



tract-heavy



equal

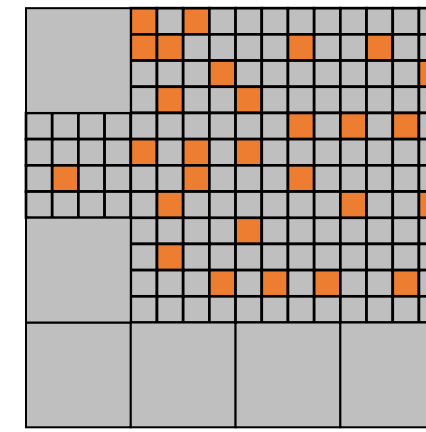
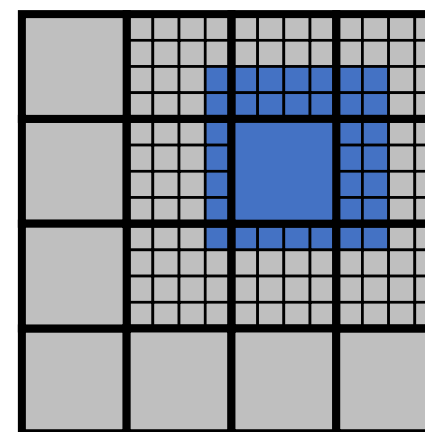
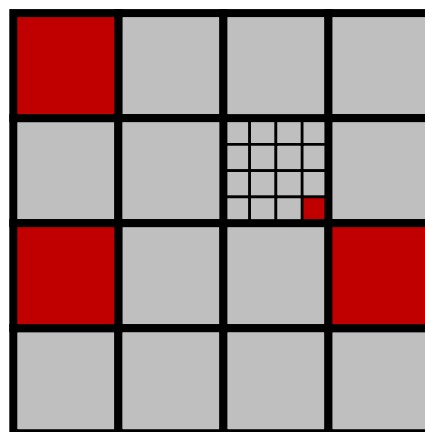


ToyDown

Red/Green – builds districts out of whole tracts

Blue – Squarish districts out of blocks

Orange – Random (discontiguous) blocks



# Fragmenting the hierarchy

Variance when  $\epsilon_1 = \dots = \epsilon_d = \frac{\epsilon}{d}$

- $\mathbb{V}(E_D) = \frac{8d^2}{\epsilon^2} (w_1^2 + \sum_{\ell=2,\dots,d} \sum_{h \in H_\ell} (w_h - w_{\hat{h}})^2)$
- $\mathbb{V}(E_D) = \frac{8d^2}{\epsilon^2} (w_1^2 + \sum_{h \in H} (w_h - w_{\hat{h}})^2)$

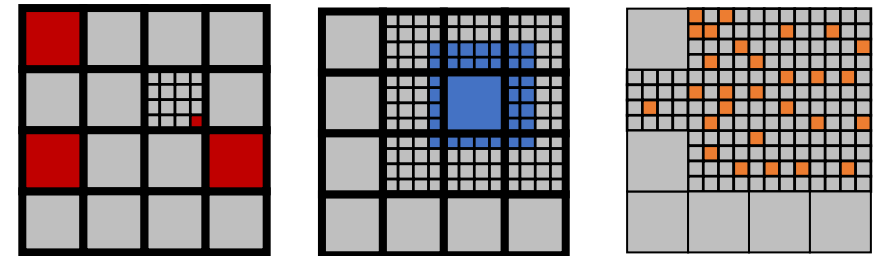
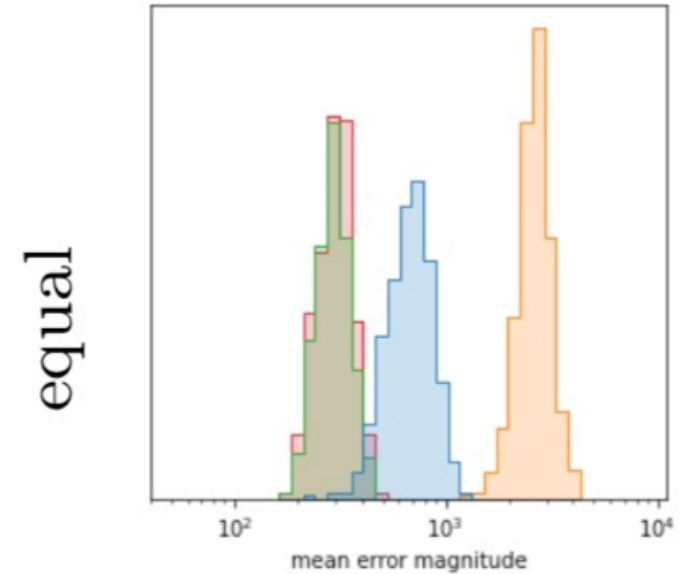
Fragmentation score of a district

- $\mathbb{F}(D) = \sum_{h \in H} (w_h - w_{\hat{h}})^2$

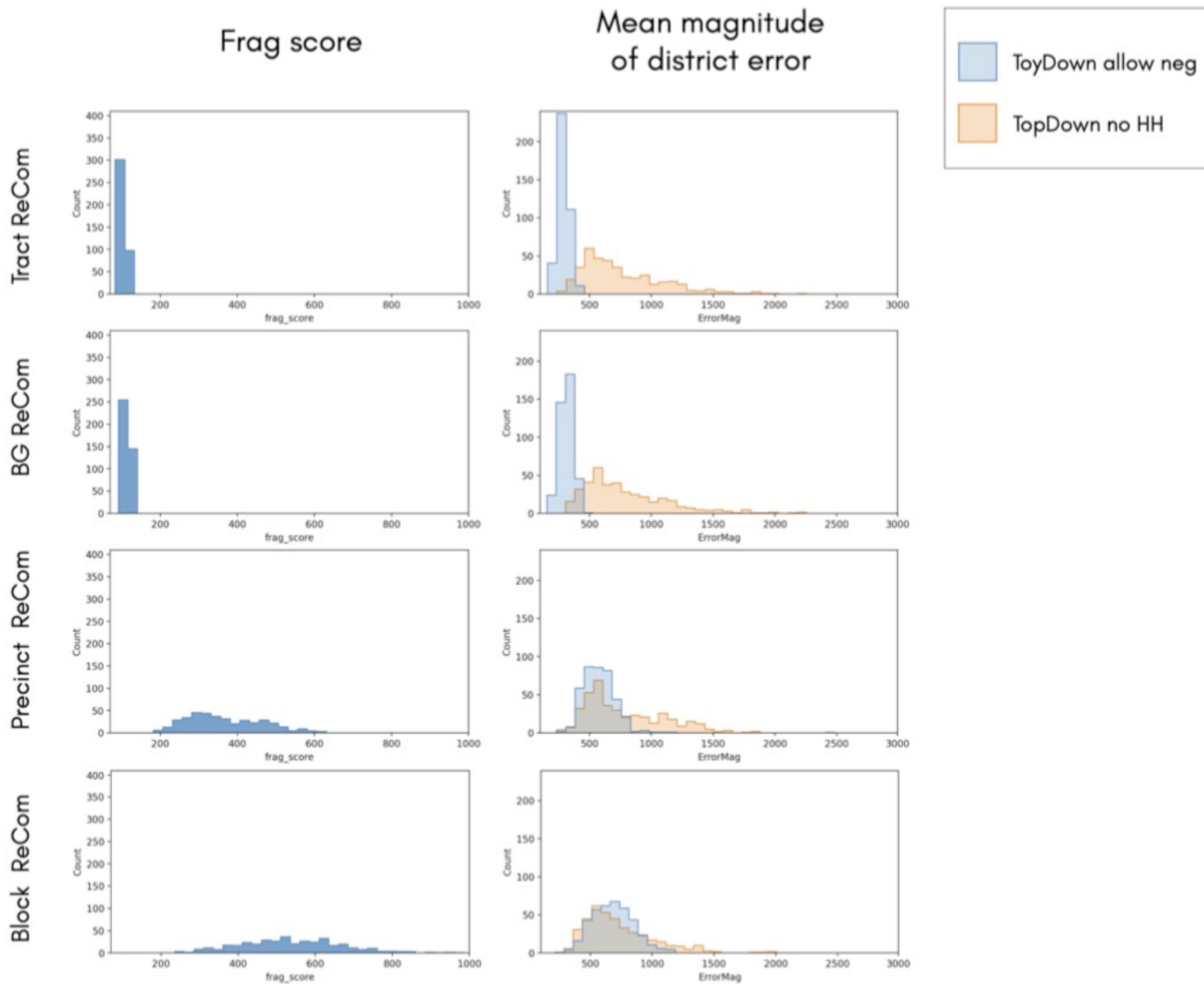
Example: Binary hierarchy, D contains  $\frac{1}{2}$  the blocks

- Left-Right  $\rightarrow \mathbb{F}(D_{left}) = 2 \cdot \frac{1}{4} = \frac{1}{2}$
- Even-Odd  $\rightarrow \mathbb{F}(D_{even}) = 2^d \cdot \frac{1}{4} = 2^{d-2}$

ToyDown



# ToyDown versus TopDown





# The effect of non-negativity

## Stage 1: Noise infusion

- For every  $h \in H_\ell$ 
  - $L_h \sim \text{Lap}\left(\frac{2}{\epsilon_\ell}\right)$
  - $\tilde{a}_h \leftarrow a_h + L_h$

## Stage 2: Post-processing

- For each level  $\ell = 1, \dots, d$  from top:
  - Compute  $\{\alpha_h\}_{h \in H_\ell}$
  - Minimizing  $\sum_{h \in H_\ell} (\alpha_h - \tilde{a}_h)^2$  s.t.:
  - Hierarchically consistent
  - **Non-negative**

Output  $\{\alpha_h\}_{h \in H_L}$

## Small counts bounce off 0

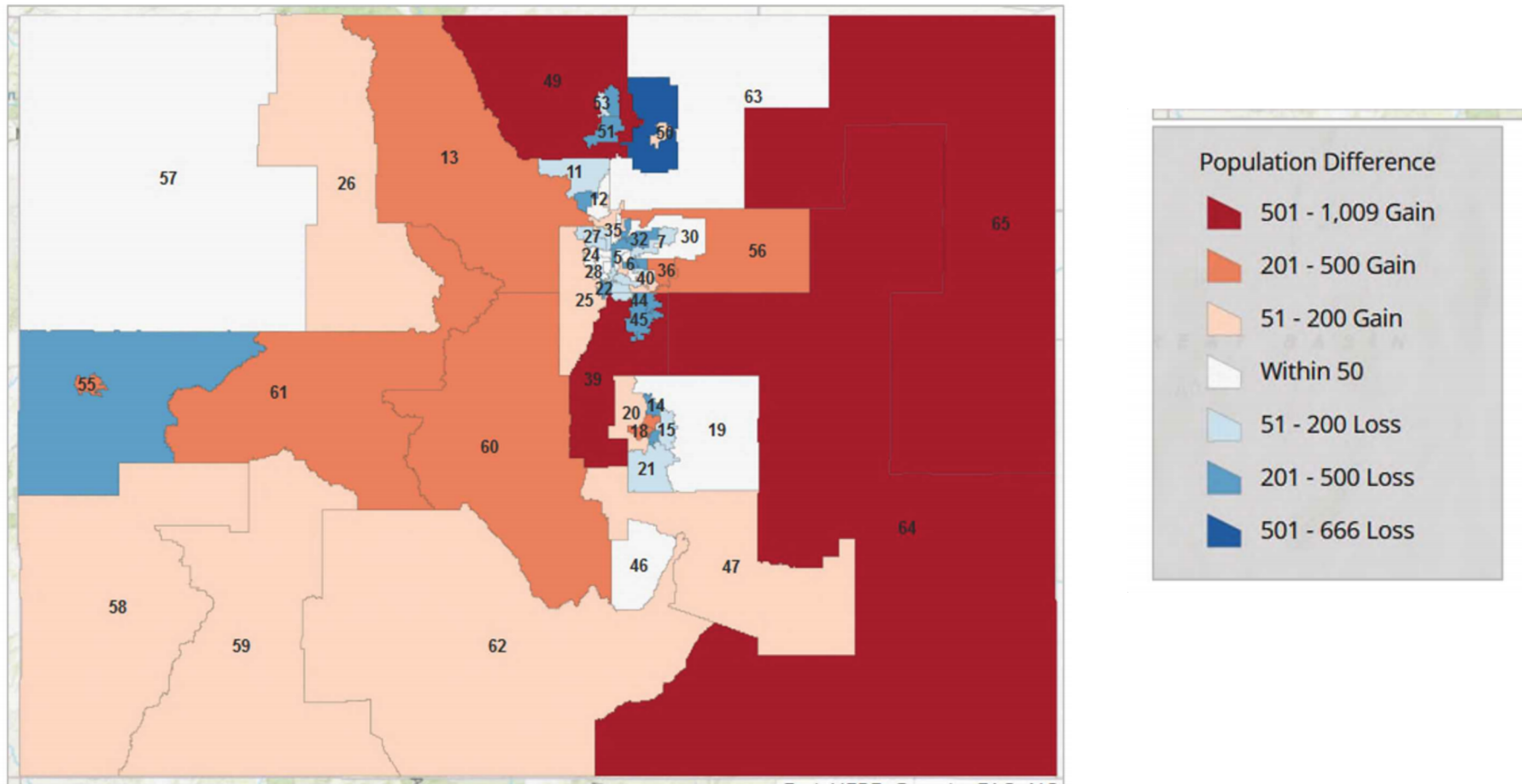
- Positively biased error
- $\mathbb{E}(E_h) > 0$  if  $a_h \approx 0$

## Large counts compensate

- Hierarchical consistency
- Negatively biased error
- $\mathbb{E}(E_h) < 0$  if  $a_h \gg 0$

# The effect of non-negativity

Figure 4  
Population Differences across Colorado House Districts\*



# ToyDown v TopDown

## Inputs

- Census Hierarchy:  $H = (H_1, \dots, H_d)$ 
  - $h \in H_\ell$  is a unit of geography
  - E.g., tract, block, ...
- Workload of histograms  $Q_1, Q_2, \dots$ 
  - $q \in Q$  is a single histogram bin
  - E.g., number of 25yo male Black Hispanics in tract 123456
- Budget allocation:  $\epsilon$ 
  - Fraction  $\beta_\ell$  that level  $H_\ell$  gets
  - Fraction  $\beta_Q$  that histogram  $Q$  gets
- Dataset:  $\{a_h\}_{h \in H}$

## Stage 1: Noise infusion

- For every  $h \in H_\ell$  and every  $q \in Q$ :
  - $\tilde{a}_{h,q} \leftarrow \text{Geometric}(q(\{a_h\}), \epsilon \cdot \beta_\ell \cdot \beta_Q)$

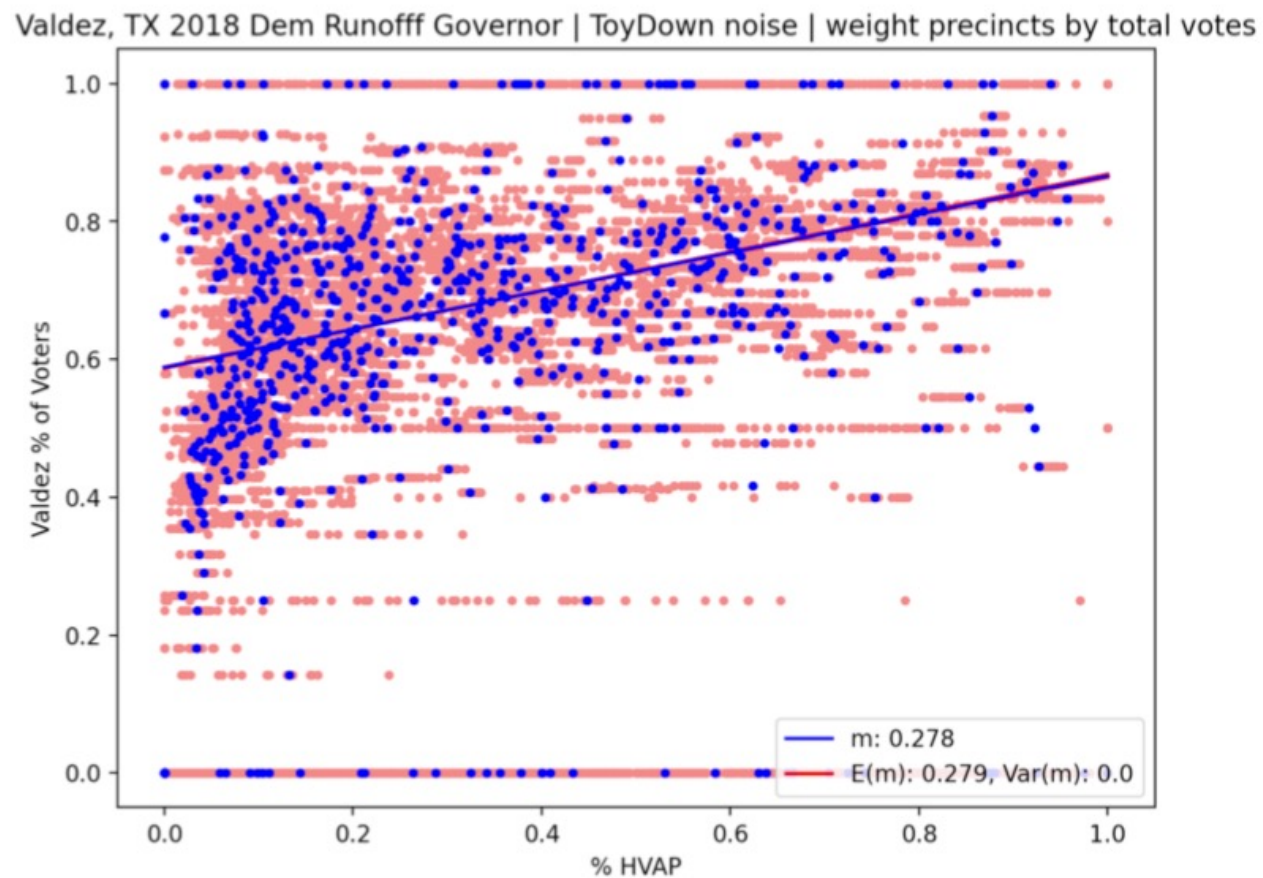
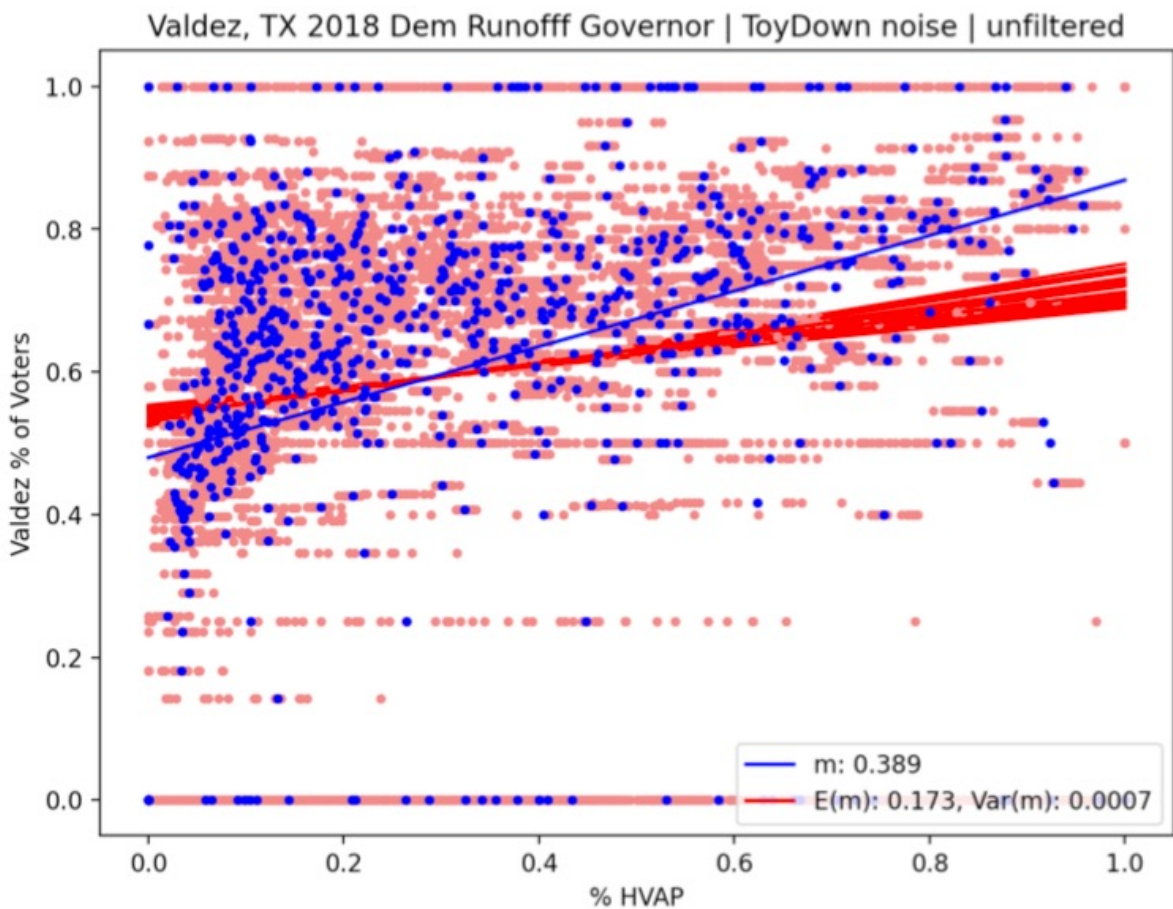
## Stage 2: Post-processing

- For each level  $\ell = 1, \dots, d$  from top:
  - Compute  $\{\tilde{a}_h\}_{h \in H_\ell}$
  - Minimizing  $\sum_{h \in H_\ell} \sum_q (q(\{a_h\}) - \tilde{a}_{h,q})^2$
  - Subject to
    - Consistent with level  $\ell - 1$
    - Non-negative
    - Lots of other constraints

## DP analysis

- Parallel composition
- Post-processing

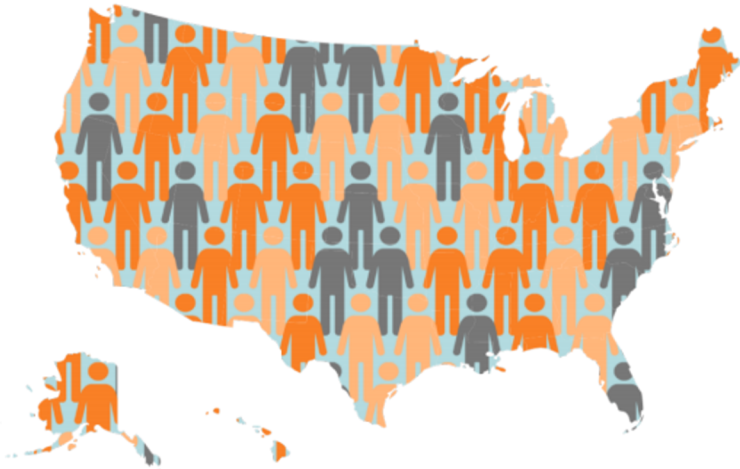
# Measuring Racial Polarization (gerrymandering)



**The Census Confidentiality Protection Pledge:**

For more than 200 years, the United States Constitution has required an enumeration of all persons

**PRELIMINARY REPORT:**  
*Impact of Differential Privacy & the 2020 Census  
on Latinos, Asian Americans and Redistricting*



THE STATE OF ALABAMA; ROBERT ADERHOLT, Representative for Alabama's 4th Congressional District, in his official and individual capacities; WILLIAM GREEN; and CAMARAN WILLIAMS,

Plaintiffs,

v.

UNITED STATES DEPARTMENT OF COMMERCE; GINA RAIMONDO, in her official capacity as Secretary of Commerce; UNITED STATES BUREAU OF THE CENSUS, an agency within the United States Department of Commerce; and RON JARMIN, in his official capacity as Acting Director of the U.S. Census Bureau,

Defendants.

CIVIL ACTION NO.  
3:21-cv-211-RAH-ECM-KCN

BRIEF OF *AMICI CURIAE* STATE OF UTAH AND 15 OTHER STATES IN SUPPORT OF PLAINTIFFS

**INTRODUCTION**

The States of Utah, Alaska, Arkansas, Florida, Kentucky, Louisiana, Maine, Mississippi, Montana, Nebraska, New Mexico, Ohio, Oklahoma, South Carolina, Texas, and West Virginia (*Amici* States) agree with Plaintiffs that the Secretary's intended use of differential privacy deprives states of accurate "[t]abulations of population" of state subparts to use in legislative apportionment and districting under 13 U.S.C. § 141(c). *Amici* States also agree that the Secretary can