Errors in 2020 Census Data Due to New Privacy Controls

A User's Guide

Jonathan Schroeder,
David Van Riper, Tracy Kugler
IPUMS & Minnesota Population Center

Virtual AAG Meeting – February 28, 2022

IPUMS.ORG

Disclosure Avoidance for the 2020 Census: An Introduction

Issued November 2021

United States Census Bureau

U.S. Department of Commerce
U.S. CENSUS BUREAU
census.gov

"The Handbook"

https://www.census.gov/library/publications/2021/decennial/2020-census-disclosure-avoidance-handbook.html

Consistency of Data Products and Formal Privacy Methods for the 2020 Census

Contact: Gordon Long — glong@mitre.org

JSR-21-02

January 2022

DISTRIBUTION A: Approved for public release; distribution unlimited.

JASON
The MITRE Corporation
7515 Colshire Drive
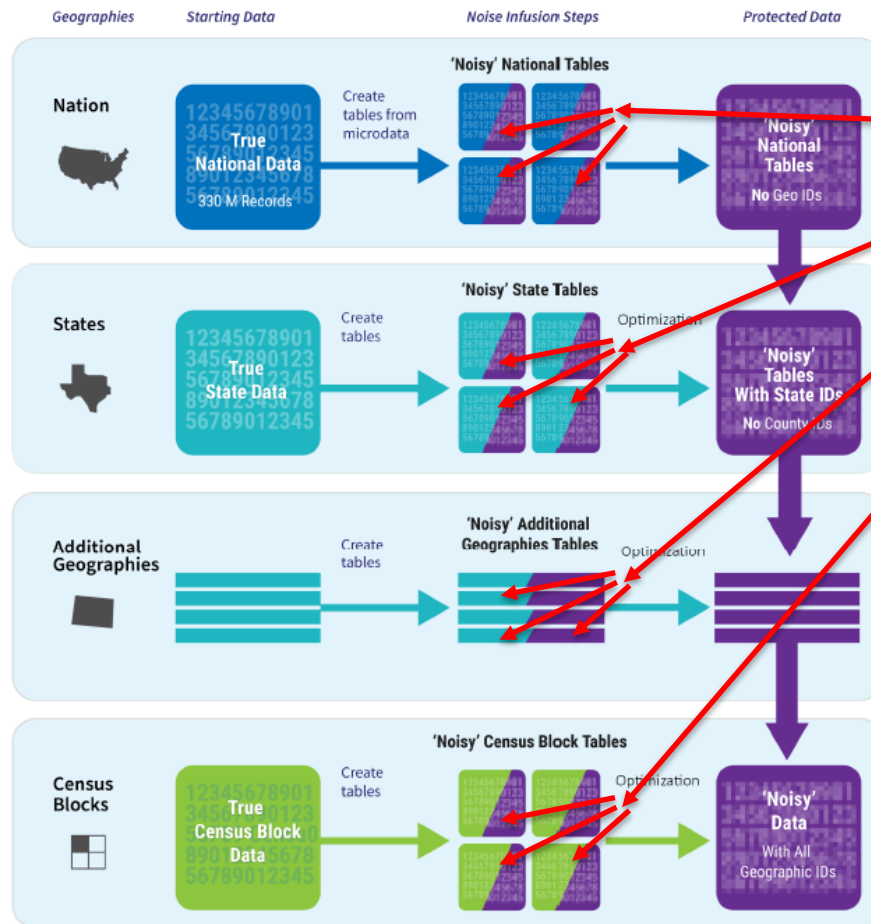McLean, Virginia 22102-7508
(703) 983-6997

JASON Report

"an independent group of scientists and engineers formed to work on problems of importance to the federal government"

https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/disclosure-avoidance/2020-das-updates/2022-01-31.html

"Top-Down Algorithm" (TDA)

"Privacy-Loss Budget" (PLB)

specified allocation

Figure 2.1. Creating Differentially Private Data for the 2020 Census Redistricting Files

Data Protection Process

Geographies	Starting Data	Noise Infusion Steps	Protected Data

Nation — True National Data, 330 M Records → Create tables from microdata → 'Noisy' National Tables → 'Noisy' National Tables, No Geo IDs

States — True State Data → Create tables → 'Noisy' State Tables → Optimization → 'Noisy' Tables With State IDs, No County IDs

Additional Geographies → Create tables → 'Noisy' Additional Geographies Tables → Optimization

Census Blocks — True Census Block Data → Create tables → 'Noisy' Census Block Tables → Optimization → 'Noisy' Data With All Geographic IDs

Handbook p. 13

IPUMS.ORG

Figure 2.2. **Standard Hierarchy of Select Geographic Areas**

"Geographic Spine"

"Off-Spine"

NATION

REGIONS

DIVISIONS

STATES

Counties

Census Tracts

Block Groups

Census Blocks

ZIP Code Tabulation Areas

School Districts

Congressional Districts

Voting Districts

County Subdivisions

Subminor Civil Divisions

AIANNH Areas
(American Indian, Alaska Native, Native Hawaiian Areas)

Urban Areas

Core Based Statistical Areas

Urban Growth Areas

State Legislative Districts

Places

Handbook p. 14

Figure 2.3. **Hierarchy for Disclosure Avoidance System Processing**

United States

↓ ↓

Non-AIANNH portions of 51 State/State Equivalent Areas | **AIANNH portions of 37 States**

↓ ↓

Non-AIANNH portions of Counties | **AIANNH portions of Counties**

↓ ↓

Non-AIANNH portions of Tracts and Block Groups | **AIANNH portions of Tracts and Block Groups**

↓ ↓

Non-AIANNH Census Blocks | **AIANNH Census Blocks**

Handbook p. 15
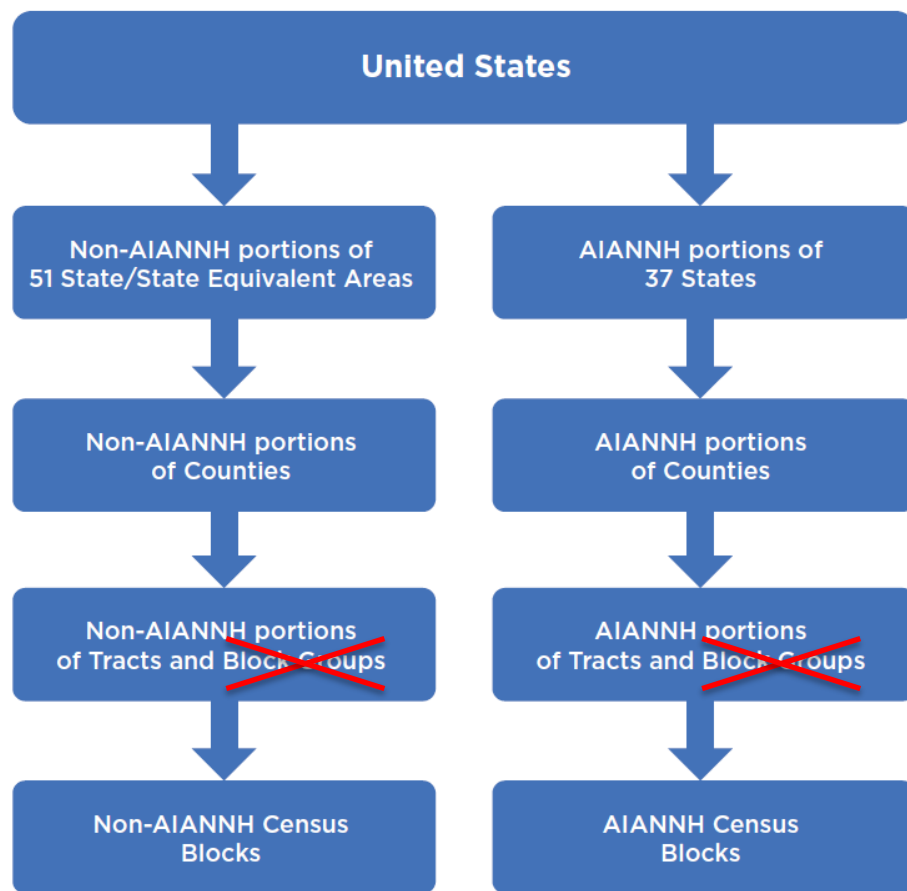
"Optimized spine"

Privacy-Loss Budget: People

Geographic level	Rho allocation
United States	104/4,099
State.	1,440/4,099
County	447/4,099
Tract	687/4,099
Optimized block group[1]	1,256/4,099
Block	165/4,099

[1] The Optimized Block Groups used within the TopDown Algorithm differ from tabulation block groups. These differences improve accuracy for "off-spine" geog... civil divisions. The use of optimized bl... and post-processing within the TopD... how the resulting data will be tabulate... be tabulated using the official tabulati... the Census Bureau's Geography Divisi...

Handbook p. 34

In most states, the District of Columbia, and Puerto Rico, these block aggregations (called "optimized block groups" in the technical documentation) were redefined to more closely approximate places (such as cities). In 12 states, blocks were aggregated to more closely approximate minor civil divisions (cities, boroughs, and towns/townships).

Handbook p. 16

Global Privacy-Loss Budget: People

Global rho	2.56
Global epsilon.	17.14
Delta	10^{-10}

Source: U.S. Census Bureau.

Privacy-Loss Budget: People

Geographic level	Rho allocation
United States	104/4,099
State	1,440/4,099
County	447/4,099
Tract	687/4,099
Optimized block group[1]	1,256/4,099
Block	165/4,099

[1] The Optimized Block Groups used within the TopDown Algorithm differ from tabulation block groups. These differences improve accuracy for "off-spine" geographies like places and minor civil divisions. The use of optimized block groups for measurement and post-processing within the TopDown Algorithm does not impact how the resulting data will be tabulated. All census data products will be tabulated using the official tabulation block groups as defined by the Census Bureau's Geography Division.
Source: U.S. Census Bureau.

Per Query Privacy-Loss Budget: People

Query	Geographic level and rho allocation					
	United States	State	County	Tract	Optimized block group[1]	Block
TOTAL (1 cell)[2]	N	3,773/4,097	3,126/4,097	1,567/4,102	1,705/4,099	5/4,097
CENRACE (63 cells)	52/4,097	6/4,097	10/4,097	4/2,051	3/4,099	9/4,097
HISPANIC (2 cells).	26/4,097	6/4,097	10/4,097	5/4,102	3/4,099	5/4,097
VOTINGAGE (2 cells)	26/4,097	6/4,097	10/4,097	5/4,102	3/4,099	5/4,097
HHINSTLEVELS (3 cells)	26/4,097	6/4,097	10/4,097	5/4,102	3/4,099	5/4,097
HHGQ (8 cells)	26/4,097	6/4,097	10/4,097	5/4,102	3/4,099	5/4,097
HISPANIC*CENRACE (126 cells).	130/4,097	12/4,097	28/4,097	1,933/4,102	1,055/4,099	21/4,097
VOTINGAGE*CENRACE (126 cells)	130/4,097	12/4,097	28/4,097	10/2,051	9/4,099	21/4,097
VOTINGAGE*HISPANIC (4 cells)	26/4,097	6/4,097	10/4,097	5/4,102	3/4,099	5/4,097
VOTINGAGE*HISPANIC*CENRACE (252 cells)	26/241	2/241	101/4,097	67/4,102	24/4,099	71/4,097
HHGQ*VOTINGAGE*HISPANIC*CENRACE (2,016 cells).	189/241	230/4,097	754/4,097	241/2,051	1,288/4,099	3,945/4,097

N Not applicable.

[1] The Optimized Block Groups used within the TopDown Algorithm differ from tabulation block groups. These differences improve accuracy for "off-spine" geographies like places and minor civil divisions. The use of optimized block groups for measurement and post processing within the TopDown Algorithm does not impact how the resulting data will be tabulated. All census data products will be tabulated using the official tabulation block groups as defined by the Census Bureau's Geography Division.

[2] The TOTAL query (total population) is held invariant at the state level. This note pertains to the interpretation of the entry in the State column of this row only. This rho allocation assigned to TOTAL at the state level is the amount assigned to the state-level queries for the total population of all American Indian and Alaska Native (AIAN) tribal areas within the state and for the total population of the remainder of the state, for the 36 states that include AIAN tribal areas.
Source: U.S. Census Bureau.

PLB allocation: Person queries

	US	State	County	Tract	OBG	Block	All
Total Population	0.000	32.352	8.321	6.403	12.746	0.005	59.826
Race	0.032	0.051	0.027	0.033	0.022	0.009	0.174
Hispanic	0.016	0.051	0.027	0.020	0.022	0.005	0.142
Voting Age	0.016	0.051	0.027	0.020	0.022	0.005	0.142
HHInstLevels	0.016	0.051	0.027	0.020	0.022	0.005	0.142
Household/GQ	0.016	0.051	0.027	0.020	0.022	0.005	0.142
Hispanic × Race	0.081	0.103	0.075	7.898	7.887	0.021	16.063
Voting Age × Race	0.081	0.103	0.075	0.082	0.067	0.021	0.428
Voting Age × Hispanic	0.016	0.051	0.027	0.020	0.022	0.005	0.142
Voting Age × Hispanic × Race	0.274	0.292	0.269	0.274	0.179	0.070	1.357
Detailed	1.990	1.972	2.007	1.969	9.628	3.876	21.443
Total	2.537	35.131	10.905	16.760	30.642	4.025	100.000

JASON Jan '22 p. 34

PLB allocation: Housing occupancy

Global Privacy-Loss Budget: Units

Global rho	0.07
Global epsilon.	2.47
Delta.	10^{-10}

Source: U.S. Census Bureau.

Privacy-Loss Budget: Units

Geographic level	Rho allocation
United States	1/205
State.	1/205
County	7/82
Tract	364/1,025
Optimized block group[1]	1,759/4,100
Block	99/820

Handbook p. 35

Puerto Rico allocation…

Privacy-Loss Budget: Units

Geographic level	Rho allocation
Puerto Rico.	5,047/876,580
Municipio.	18,746/219,145
Tract	94,451/262,974
Optimized block group[1]	281,911/657,435
Block	99/820

Handbook p. 37

Geographic level	Rho allocation
US	0.49
State	0.49
County	8.54
Tract	35.51
Optimized BG	42.90
Block	12.07

Expectations

1. Smaller counts have larger relative errors
   - Small areas (blocks)
   - Small population groups

2. "On-spine" levels have smaller errors
   - States, counties, tracts, "optimized block groups"

3. Total populations have smaller errors than other counts
   - Except for blocks and "off-spine" levels

■ These expectations mix in complex ways!

	US	State	County	Tract	OBG	Block	All
Total Population	0.000	32.352	8.321	6.403	12.746	0.005	59.826
Race	0.032	0.051	0.027	0.033	0.022	0.009	0.174
Hispanic	0.016	0.051	0.027	0.020	0.022	0.005	0.142
Voting Age	0.016	0.051	0.027	0.020	0.022	0.005	0.142
HHInstLevels	0.016	0.051	0.027	0.020	0.022	0.005	0.142
Household/GQ	0.016	0.051	0.027	0.020	0.022	0.005	0.142
Hispanic × Race	0.081	0.103	0.075	7.898	7.887	0.021	16.063
Voting Age × Race	0.081	0.103	0.075	0.082	0.067	0.021	0.428
Voting Age × Hispanic	0.016	0.051	0.027	0.020	0.022	0.005	0.142
Voting Age × Hispanic × Race	0.274	0.292	0.269	0.274	0.179	0.070	1.357
Detailed	1.990	1.972	2.007	1.969	9.628	3.876	21.443
Total	2.537	35.131	10.905	16.760	30.642	4.025	100.000

93.4%

JASON Jan '22 p. 34

Total Population by Summary Level, 2010
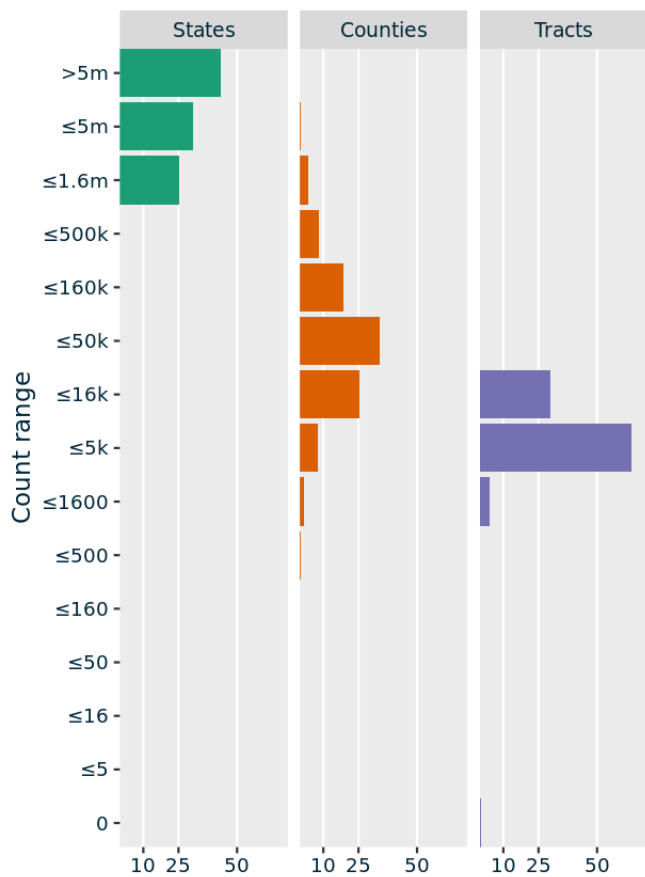
Total Population by Summary Level, 2010

IPUMS NHGIS

NHGIS GEOMARKER
NATIONAL HISTORICAL GIS

HOME | SELECT DATA | MY DATA | SUPPORT
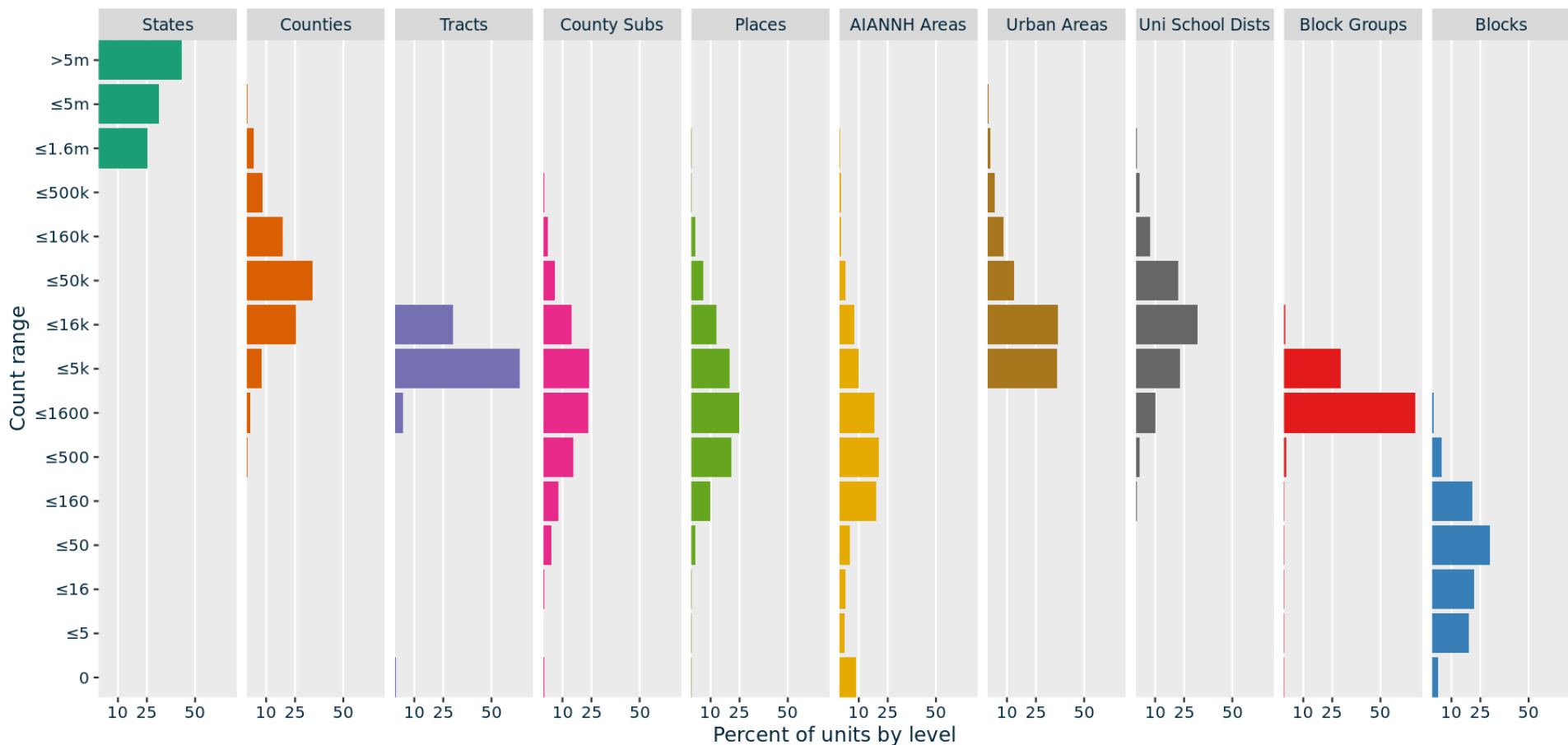
IPUMS NHGIS

ABOUT
REGISTER
DONATE TO NHGIS

DATA

BROWSE AND SELECT DATA
DOWNLOAD OR REVISE MY DATA
API

SUPPLEMENTAL DATA

GEOGRAPHIC CROSSWALKS
ENVIRONMENTAL SUMMARIES
PRIVACY-PROTECTED DEMO DATA
SABINS SCHOOL AREAS

DOCUMENTATION

DATA AVAILABILITY
OVERVIEW OF DATASETS
TABULAR DATA SOURCES
TIME SERIES TABLES
GIS FILES
REVISION HISTORY

SUPPORT

FAQ
USER GUIDE
USER FORUM

RESEARCH

CITING IPUMS NHGIS
IPUMS BIBLIOGRAPHY

PRIVACY-PROTECTED 2010 CENSUS DEMONSTRATI

The IPUMS NHGIS Privacy-Protected Demonstration Data link togethe
Census summary tables:

1. Original tables from the 2010 Census Summary Files
2. New tables based on experimental runs of the Census Bureau's
   System (DAS) applied to the original 2010 Census responses

The 2020 DAS represents a major change in the Census Bureau's app
privacy. The Privacy-Protected Demonstration Data facilitate assess
framework, enabling a broad range of users to investigate and provi
quality of data produced by different DAS versions, including both pr
the final production system.

- Purpose
- Feedback and Questions
- Technical Details
- Versions
  - Vintage 2019-10-29
  - Vintage 2020-05-27
    - Coverage
    - Parameters
    - Data Files
  - Vintage 2020-09-17
   - Coverage

Vintage 2021-06-08: Data Files

Nationwide Files

|		
State 📄	Place 📄	111th Congressional District 📄
County 📄	Alaska Native Regional Corporation 📄	State Legislative District (Upper Chamber) 📄
County Subdivision 📄	American Indian/Alaska Native/Hawaiian Home Land 📄	State Legislative District (Lower Chamber) 📄
Census Tract 📄	Core-Based Statistical Area (CBSA) 📄	School District (Unified) 📄
Block Group 📄	Urban Area 📄	American Indian Area/Alaska Native Area/Hawaiian Home Land (by State--County--Census Tract) 📄
County (by State--American Indian Area/Alaska Native Area/Hawaiian Home Land) 📄	County Subdivision (by State--American Indian Area/Alaska Native Area/Hawaiian Home Land--County) 📄	

Census Blocks by State

Alaska 📄	Louisiana 📄	Ohio 📄
Alabama 📄	Massachusetts 📄	Oklahoma 📄

Total Population: Errors > 2%

Solid fill: Discrepancies > 2%

Total Population: Errors > 25%
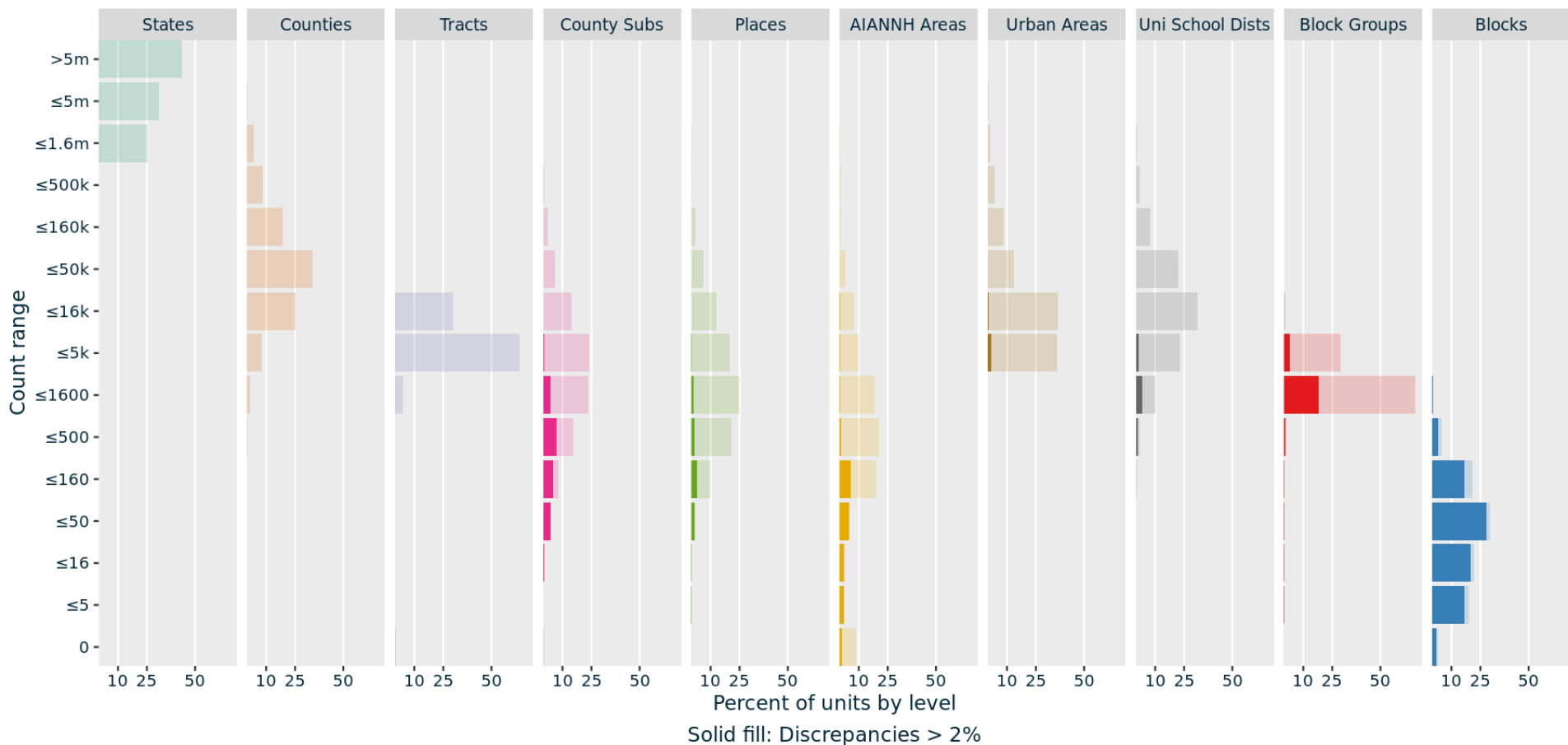
Solid fill: Discrepancies > 25%
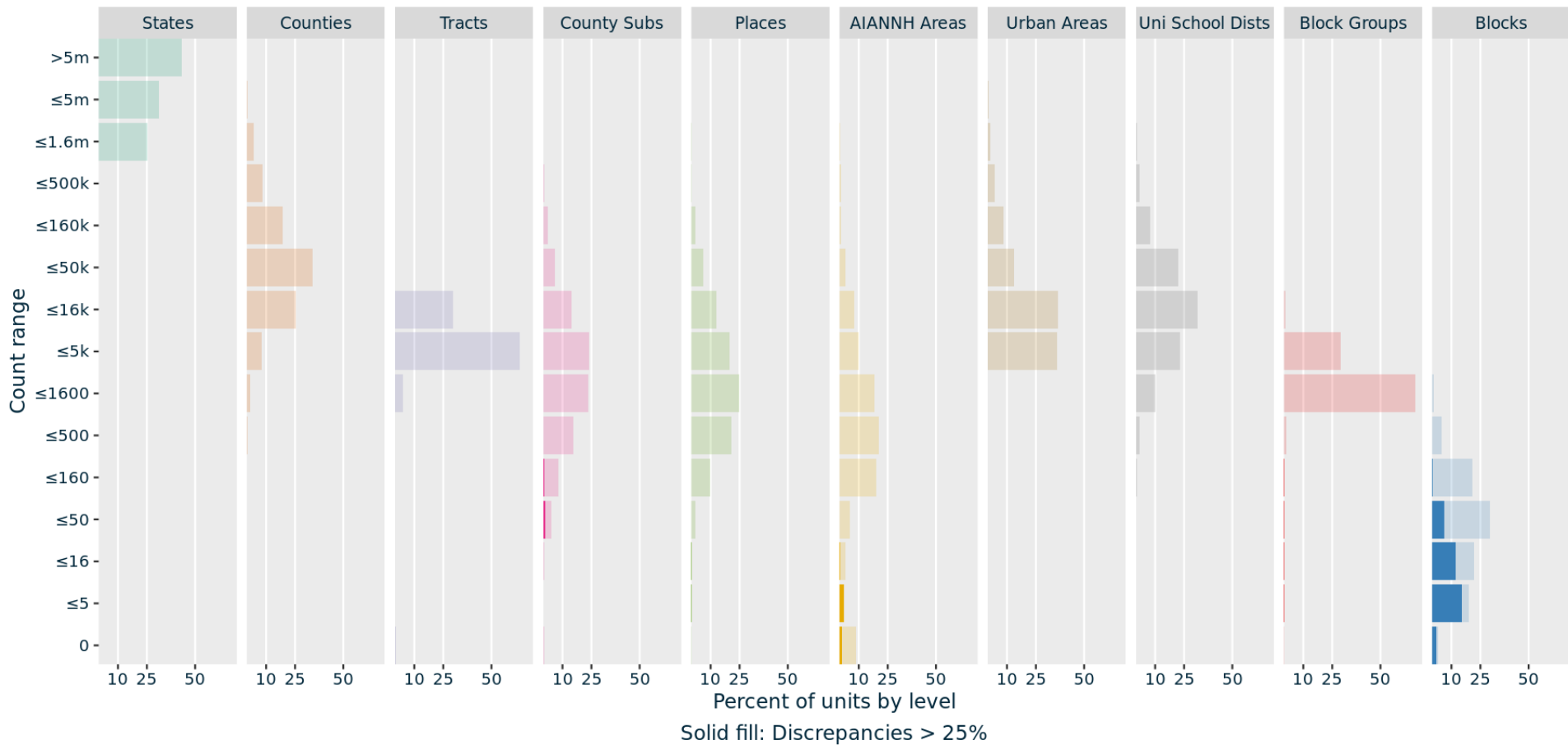
Persons under 18 Years: Errors > 25%

Solid fill: Discrepancies > 25%
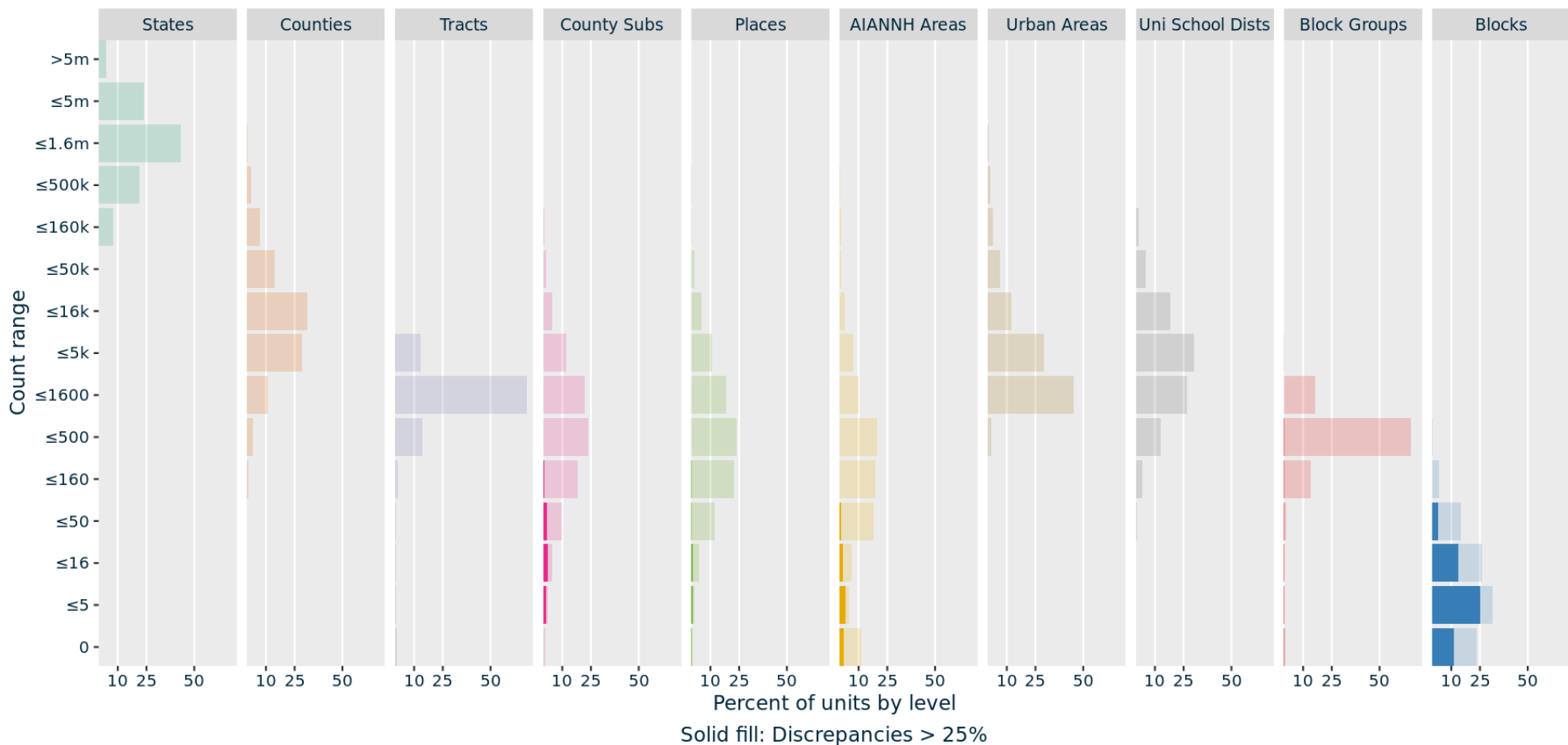
Hispanic or Latino: Errors > 25%

Solid fill: Discrepancies > 25%
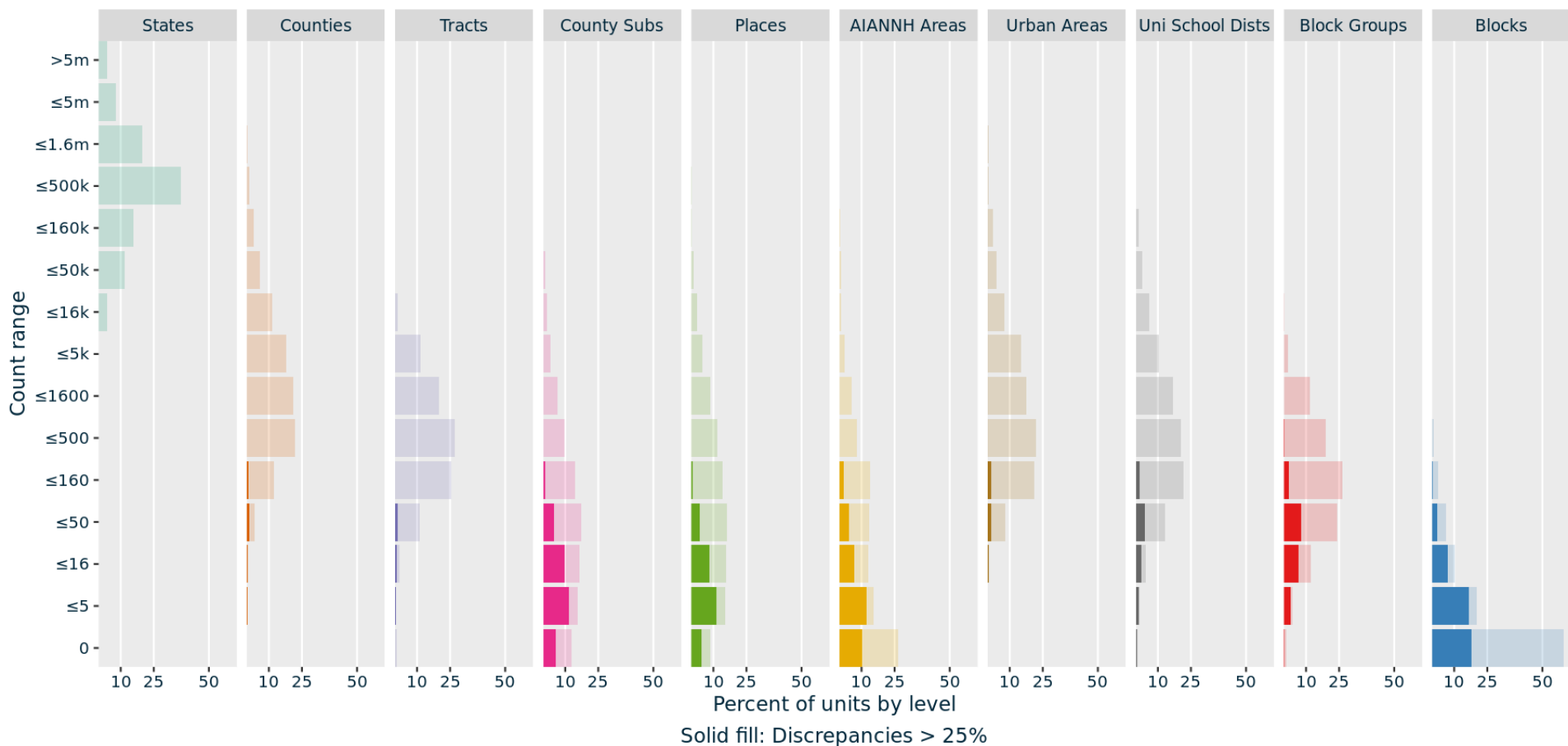
Native Hawaiian and Other Pacific Islander Alone, Not Hispanic: Errors > 25%

Solid fill: Discrepancies > 25%
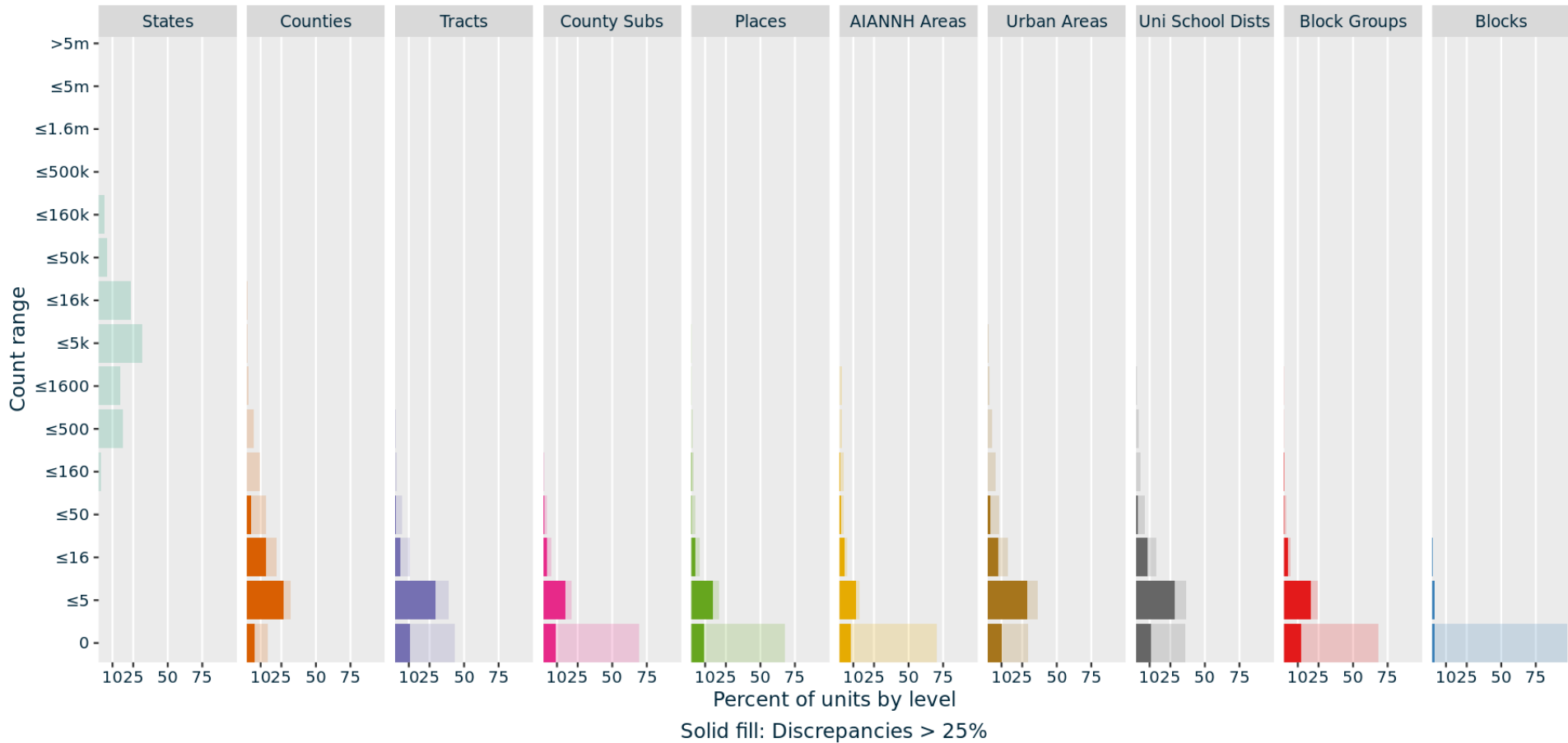
Vacant Housing Units: Errors > 25%

Solid fill: Discrepancies > 25%

Total Population: Errors > 10% | 2%

	County Subdivisions	Incorporated Places	Census-Designated Places

Not in MCD State

Quintile	County Subdivisions	Incorporated Places	Census-Designated Places
5	43,000	44,000	15,000
4	4,100	3,300	2,500
3	1,500	1,100	820
2	500	430	290
1	95	130	85

In MCD State*

Quintile	County Subdivisions	Incorporated Places	Census-Designated Places
5	20,000	34,000	11,000
4	2,900	4,500	2,700
3	1,100	1,600	1,300
2	510	630	590
1	150	200	190

In MI or NY

Quintile	County Subdivisions	Incorporated Places	Census-Designated Places
5	57,000	68,000	19,000
4	4,900	4,700	4,100
3	2,500	3,000	1,300
2	1,400	860	260
1	540	200	270

Percent with discrepancy > 10% | 2%
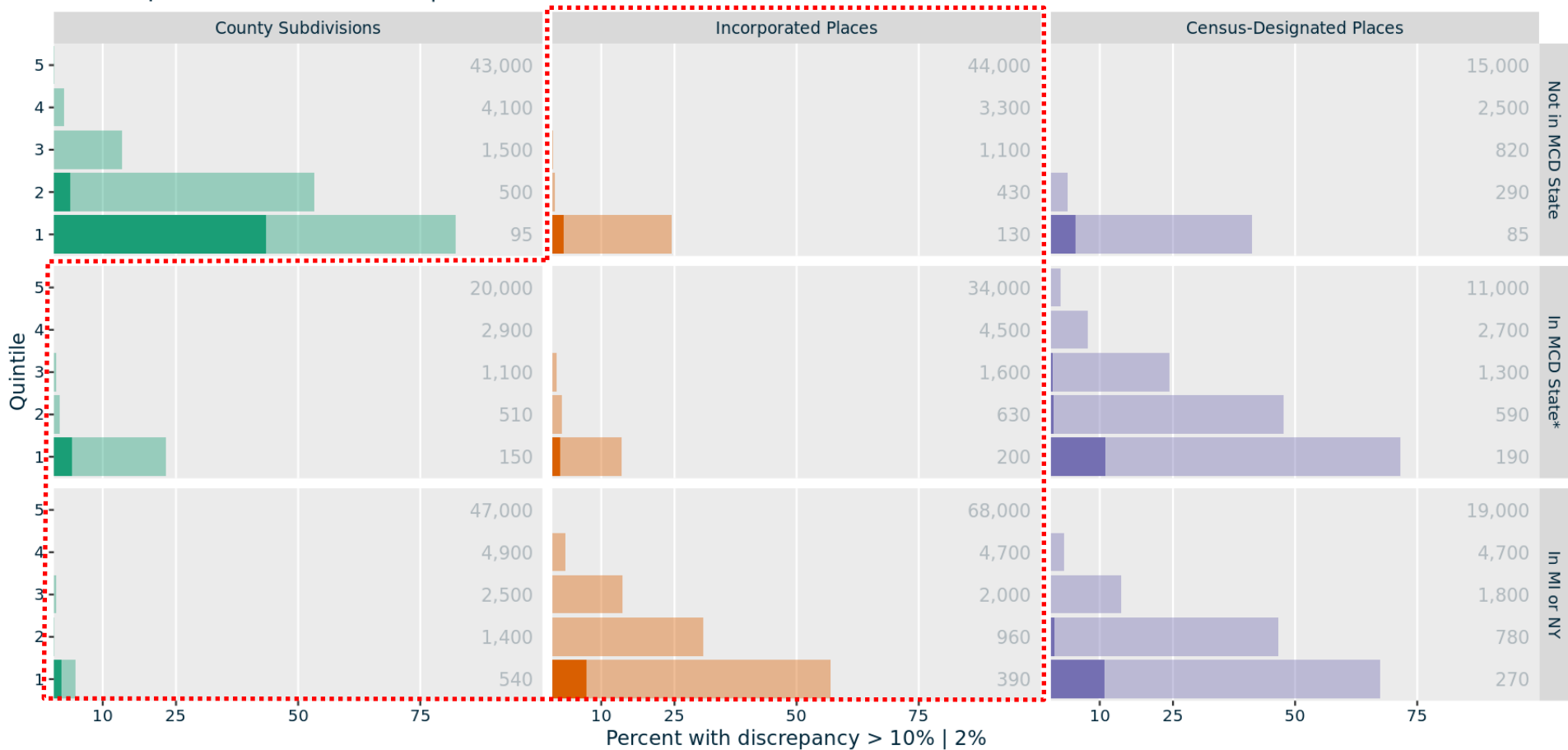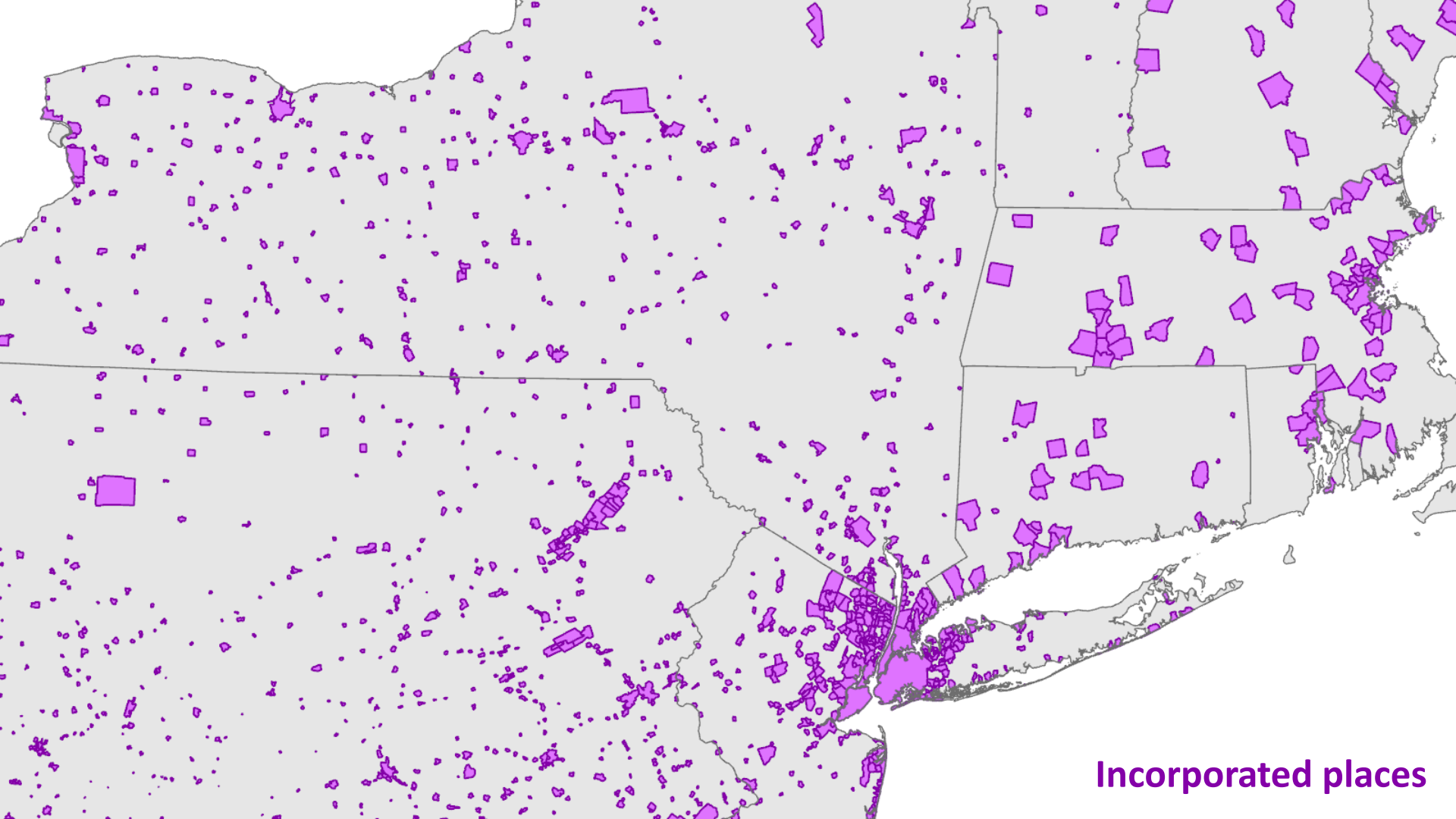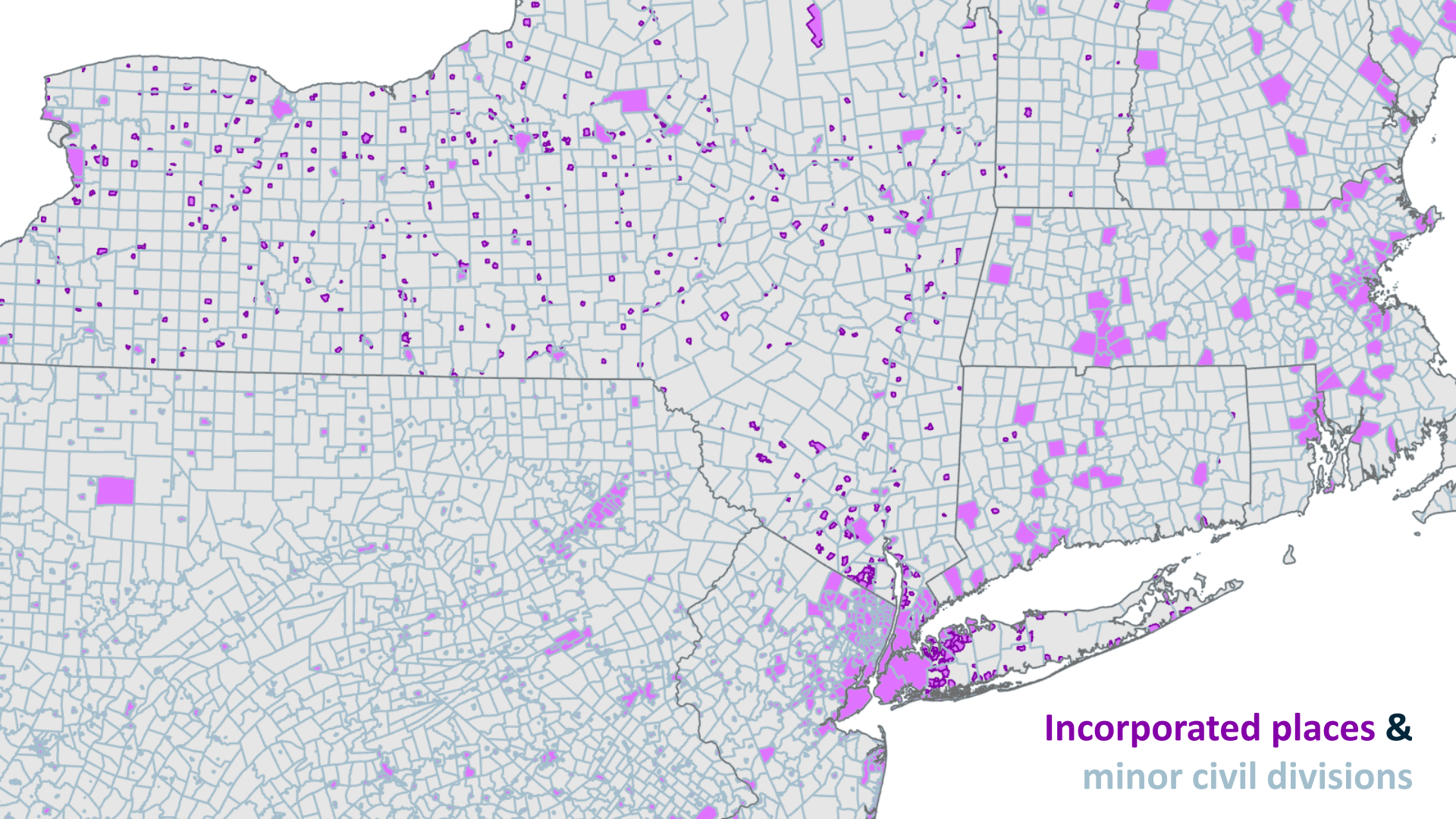
Total Population: Errors > 10% | 2%
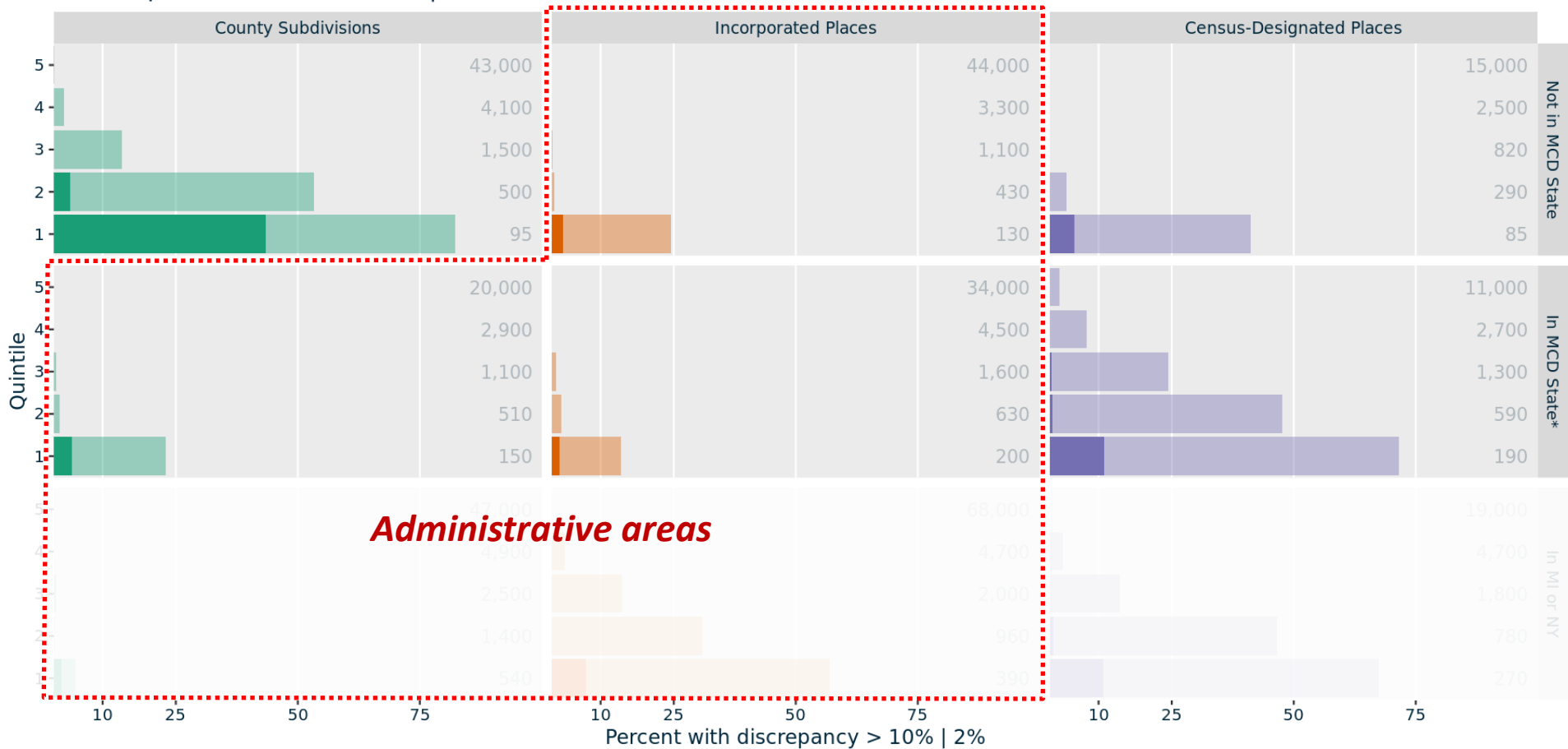
Total Population: Errors > 10% | 2%

Incorporated places

Incorporated places & minor civil divisions

Short straw for Michigan & NY villages

- 56 incorporated places have a population discrepancy of 50 or more
  - **52 of 56** are in Michigan or New York
  - **51 of 56** are population losses
- Worst case: Liberty, NY
  - 4392 → 4113 = -279, -6.4%

Some guidelines

- Be wary of small counts, especially < 100
  - Can you use larger areas or categories?
  - Could you aggregate up?
- Counts > 5,000 are broadly reliable
  - In states, counties, & tracts, counts > 500 are
- In MCD states, use county subdivisions. Elsewhere, use places.
- When possible, use total housing units…

Total Housing Units: Errors > 0%

	States	Counties	Tracts	County Subs	Places	AIANNH Areas	Urban Areas	Uni School Dists	Block Groups	Blocks
10	9,300,000	270,000	3,500	27,000	25,000	28,000	250,000	77,000	1,300	100
9	4,700,000	58,000	2,600	4,100	3,800	1,700	20,000	15,000	890	33
8	3,100,000	31,000	2,300	2,200	1,800	530	8,400	9,300	740	22
7	2,700,000	20,000	2,000	1,300	1,000	240	5,100	6,400	640	16
6	2,200,000	14,000	1,800	870	630	140	3,600	4,500	560	11
5	1,600,000	10,000	1,600	570	400	89	2,700	3,200	500	8
4	1,200,000	7,800	1,400	350	260	58	2,100	2,300	450	5
3	790,000	5,400	1,200	200	160	36	1,700	1,500	390	3
2	500,000	3,500	1,000	96	95	14	1,400	940	340	2
1	310,000	1,600	590	27	40	1	1,100	390	240	1

Decile

Percent with discrepancy > 0 %

Requests for Census Bureau

- Provide more info on ***optimized block groups***
  - The most accurate sub-tract units!
  - Huge potential value for boundaries & tables
- A solution to many problems: ***margins of error***
- If we're left to produce MOEs on our own, we could do that better with ***more demo data***
- Include all admin areas in optimized block groups

Acknowledgments

IPUMS.ORG