Understanding the 2020 Census Disclosure Avoidance System:

Production settings and DAS accuracy metrics for the P.L. 94-171 Redistricting Data Summary File

Michael Hawes and Matthew Spence
U.S. Census Bureau

July 1, 2021

Shape
your future
START HERE >

United States®
Census
2020

Acknowledgements

Any opinions and viewpoints expressed in this presentation are the author's own, and do not represent the opinions or viewpoints of the U.S. Census Bureau.

The statistics included in this presentation have been cleared for public dissemination by the Census Bureau's Disclosure Review Board (CBDRB-FY20-DSEP-001).

Shape
your future
START HERE >

United States®
Census
2020

Production Settings for 2020 Census P.L 94-171 Redistricting Data Summary File

FOR IMMEDIATE RELEASE: WEDNESDAY, JUNE 09, 2021

Census Bureau Sets Key Parameters to Protect Privacy in 2020 Census Results

JUNE 09, 2021
RELEASE NUMBER CB21-CN.42

SUBSCRIBE | RSS SMS EMAIL

JUNE 9, 2021 —The U.S. Census Bureau's Data Stewardship Executive Policy Committee (DSEP) announced it has selected the settings and parameters for the Disclosure Avoidance System (DAS) for the 2020 Census redistricting data (PL-94-171). The DAS uses a mathematical algorithm to ensure that the privacy of

Shape
your future
START HERE >

United States®
Census
2020

Stakeholder Feedback

We received a substantial amount of invaluable feedback on the April 2021 Demonstration Data.

Major themes included:

- Accuracy for AIAN tribal areas and other "off-spine" geographies
- Accuracy for places, Minor Civil Divisions, and tract-level data
- Bias (geographic and characteristic)
- Race and Ethnicity statistics
- Occupancy rates

Shape
your future
START HERE >

United States®
Census
2020

Key Parameters and Improvements

- **Privacy-loss Budget (PLB):**
  - **$\varepsilon = 17.14$ for persons**
  - **$\varepsilon = 2.47$ for units**

- **Improvements to the optimized geographic post-processing hierarchy**

- **Extra PLB allocated to Population counts**

- **Extra PLB allocated to Race and Ethnicity statistics**

- **Extra PLB allocated to occupancy rates at the block-group level and above**

Shape
your future
START HERE >

United States®
Census
2020

Privacy-loss Budget Allocations

https://www2.census.gov/programs-surveys/decennial/2020/program-management/data-product-planning/2010-demonstration-data-products/ProductionSettings20210608/2021-06-08-privacy-loss_budgetallocation.pdf

Shape
your future
START HERE >

United States®
Census
2020

Privacy-loss Budget Allocations
by geographic level

April 2021 Demonstration Data:

		rho Allocation by Geographic Level	
Global rho		1.05	
Global epsilon		10.3	
delta		10^{-10}	
		rho Allocation by Geographic Level	
US		51/1024	
State		153/1024	
County		78/1024	
Tract		51/1024	
Optimized Block Group*		172/1024	
Block		519/1024	

Production Settings:

		rho Allocation by Geographic Level	
Global rho		**2.56**	
Global epsilon		**17.14**	
delta		10^{-10}	
		rho Allocation by Geographic Level	
US		104/4099	
State		**1440/4099**	
County		**447/4099**	
Tract		**687/4099**	
Optimized Block Group*		**1256/4099**	
Block		165/4099	

Shape
your future
START HERE >

United States®
Census
2020

Privacy-loss Budget Allocations
by query

April 2021 Demonstration Data:

Query	Per Query rho Allocation by Geographic Level					
	US	State	County	Tract	Optimized Block Group*	Block
TOTAL (1 cell)		678/1024**	342/1024	1/1024	572/1024	1/1024
CENRACE (63 cells)	2/1024	1/1024	1/1024	2/1024	1/1024	2/1024
HISPANIC (2 cells)	1/1024	1/1024	1/1024	1/1024	1/1024	1/1024
VOTINGAGE (2 cells)	1/1024	1/1024	1/1024	1/1024	1/1024	1/1024
HHINSTLEVELS (3 cells)	1/1024	1/1024	1/1024	1/1024	1/1024	1/1024
HHGQ (8 cells)	1/1024	1/1024	1/1024	1/1024	1/1024	1/1024
HISPANIC*CENRACE (126 cells)	5/1024	2/1024	3/1024	5/1024	3/1024	5/1024
VOTINGAGE*CENRACE (126 cells)	5/1024	2/1024	3/1024	5/1024	3/1024	5/1024
VOTINGAGE*HISPANIC (4 cells)	1/1024	1/1024	1/1024	1/1024	1/1024	1/1024
VOTINGAGE*HISPANIC*CENRACE (252 cells)	17/1024	6/1024	11/1024	17/1024	8/1024	17/1024
HHGQ*VOTINGAGE* HISPANIC*CENRACE (2,016 cells)	990/1024	330/1024	659/1024	989/1024	432/1024	989/1024

Production Settings:

Query	Per Query rho Allocation by Geographic Level					
	US	State	County	Tract	Optimized Block Group*	Block
TOTAL (1 cell)		3773/4097**	3126/4097	1567/4102	1705/4099	5/4097
CENRACE (63 cells)	52/4097	6/4097	10/4097	4/2051	3/4099	9/4097
HISPANIC (2 cells)	26/4097	6/4097	10/4097	5/4102	3/4099	5/4097
VOTINGAGE (2 cells)	26/4097	6/4097	10/4097	5/4102	3/4099	5/4097
HHINSTLEVELS (3 cells)	26/4097	6/4097	10/4097	5/4102	3/4099	5/4097
HHGQ (8 cells)	26/4097	6/4097	10/4097	5/4102	3/4099	5/4097
HISPANIC*CENRACE (126 cells)	130/4097	12/4097	28/4097	1933/4102	1055/4099	21/4097
VOTINGAGE*CENRACE (126 cells)	130/4097	12/4097	28/4097	10/2051	9/4099	21/4097
VOTINGAGE*HISPANIC (4 cells)	26/4097	6/4097	10/4097	5/4102	3/4099	5/4097
VOTINGAGE*HISPANIC*CENRACE (252 cells)	26/241	2/241	101/4097	67/4102	24/4099	71/4097
HHGQ*VOTINGAGE* HISPANIC*CENRACE (2,016 cells)	189/241	230/4097	754/4097	241/2051	1288/4099	3945/4097

*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.

**The TOTAL query (total population) is held invariant at the state level. 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.

Shape
your future
START HERE >

United States®
Census
2020

Detailed Summary Metrics on Demographic Reasonableness

- Compare tabulated quantities at various geographic levels (e.g., Total Population at the county level, Asian Alone at the tract level) for a DAS run to published (i.e., swapped) 2010 tabulations.

- Metrics released for the Production Settings (June 2021) run include:

Accuracy	- Mean Absolute Error (MAE): What is the average absolute change (+/-)? - Mean Absolute Percent Error (MAPE): What is the average relative change (+/- %)
Bias	- Mean Error (ME): What is the average directional change? - Mean Algebraic Percent Error (MALPE): What is the average directional relative change?
Outliers	- How many geographies are above particular thresholds?

Shape
your future
START HERE >

United States®
Census
2020

Calculating Metrics: MAE

County	Published 2010 Population	PPMF 2010 Population	Error	Absolute Error
Autauga County, Alabama	54,571	54,581	10	10
Baldwin County, Alabama	182,265	182,263	-2	2
Barbour County, Alabama	27,457	27,455	-2	2
Bibb County, Alabama	22,915	22,922	7	7
Blount County, Alabama	57,322	57,321	-1	1
...	...	...	...	...
Loving County, Texas	82	77	-5	5
...	...	...	...	...

Mean Absolute Error: 4.91

Shape
your future
START HERE >

United States®
Census
2020

Calculating Metrics: MAPE

County	Published 2010 Population	PPMF 2010 Population	Percent Error	Absolute Percent Error
Autauga County, Alabama	54,571	54,581	0.0183%	0.0183%
Baldwin County, Alabama	182,265	182,263	-0.0011%	0.0011%
Barbour County, Alabama	27,457	27,455	-0.0073%	0.0073%
Bibb County, Alabama	22,915	22,922	0.0305%	0.0305%
Blount County, Alabama	57,322	57,321	-0.0017%	0.0017%
...	...	...	...	...
Loving County, Texas	82	77	-6.0976%	6.0976%
...	...	...	...	...

Mean Absolute Percent Error: 0.04%

Shape
your future
START HERE >

United States®
Census
2020

Total Population – County

County Total Population:
Mean Absolute Percent Error (MAPE) — All Counties

This translates to an average error of 82.2 persons

This translates to an average error of 1.75 persons

Mean Absolute Percent Error (MAPE)

Demonstration Product Vintage

Oct 2019 · May 2020 · Nov 2020 · Apr 2021(ϵ4.5) · Apr 2021(ϵ12.2) · Production Settings

Shape your future START HERE >

United States® Census 2020

Total Population – County

County Total Population:
Mean Absolute Percent Error (MAPE) — Least Populous Counties (Population Under 1,000)

This translates to an average error of 76.5 persons

This translates to an average error of 1.3 persons

Mean Absolute Percent Error (MAPE)

Oct 2019 · May 2020 · Nov 2020 · Apr 2021(ϵ4.5) · Apr 2021(ϵ12.2) · Production Settings

Demonstration Product Vintage

Shape your future
START HERE >

United States®
Census
2020

Total Population – Place

Place Total Population:
Mean Absolute Percent Error (MAPE) — All Incorporated Places

This translates to an average error of 85 persons

This translates to an average error of 3.6 persons

Mean Absolute Percent Error (MAPE)

10.0%
8.0%
6.0%
4.0%
2.0%
0.0%

Oct 2019   May 2020   Nov 2020   Apr 2021(ϵ4.5)   Apr 2021(ϵ12.2)   Production Settings

Demonstration Product Vintage

Shape your future START HERE >

United States® Census 2020

Total Population – Place

Place Total Population:
Mean Absolute Percent Error (MAPE) — Least Populous Incorporated Places (Population Under 500)

This translates to an average error of 43.7 persons

This translates to an average error of 1.9 persons

Mean Absolute Percent Error (MAPE)

Oct 2019 | May 2020 | Nov 2020 | Apr 2021(ϵ4.5) | Apr 2021(ϵ12.2) | Production Settings

Demonstration Product Vintage

Shape your future START HERE >

United States® Census 2020

Total Population – Tract

Tract Total Population:
Mean Absolute Error (MAE) — Tracts
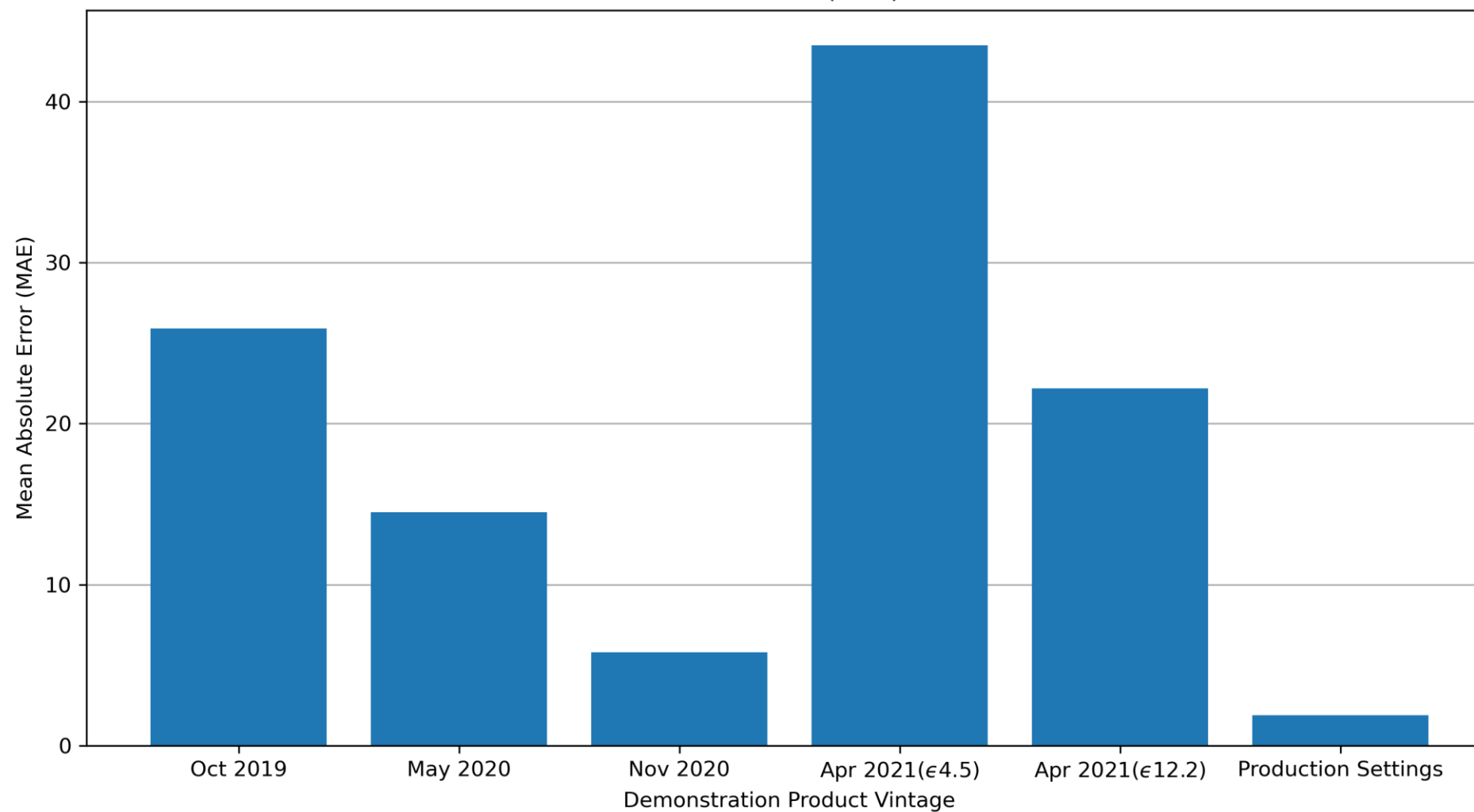
Total Population – Federal American Indian Reservation

Federal American Indian Reservation/Off-Reservation Trust Land
Total Population: Mean Absolute Error (MAE) — All Areas

2020CENSUS.GOV
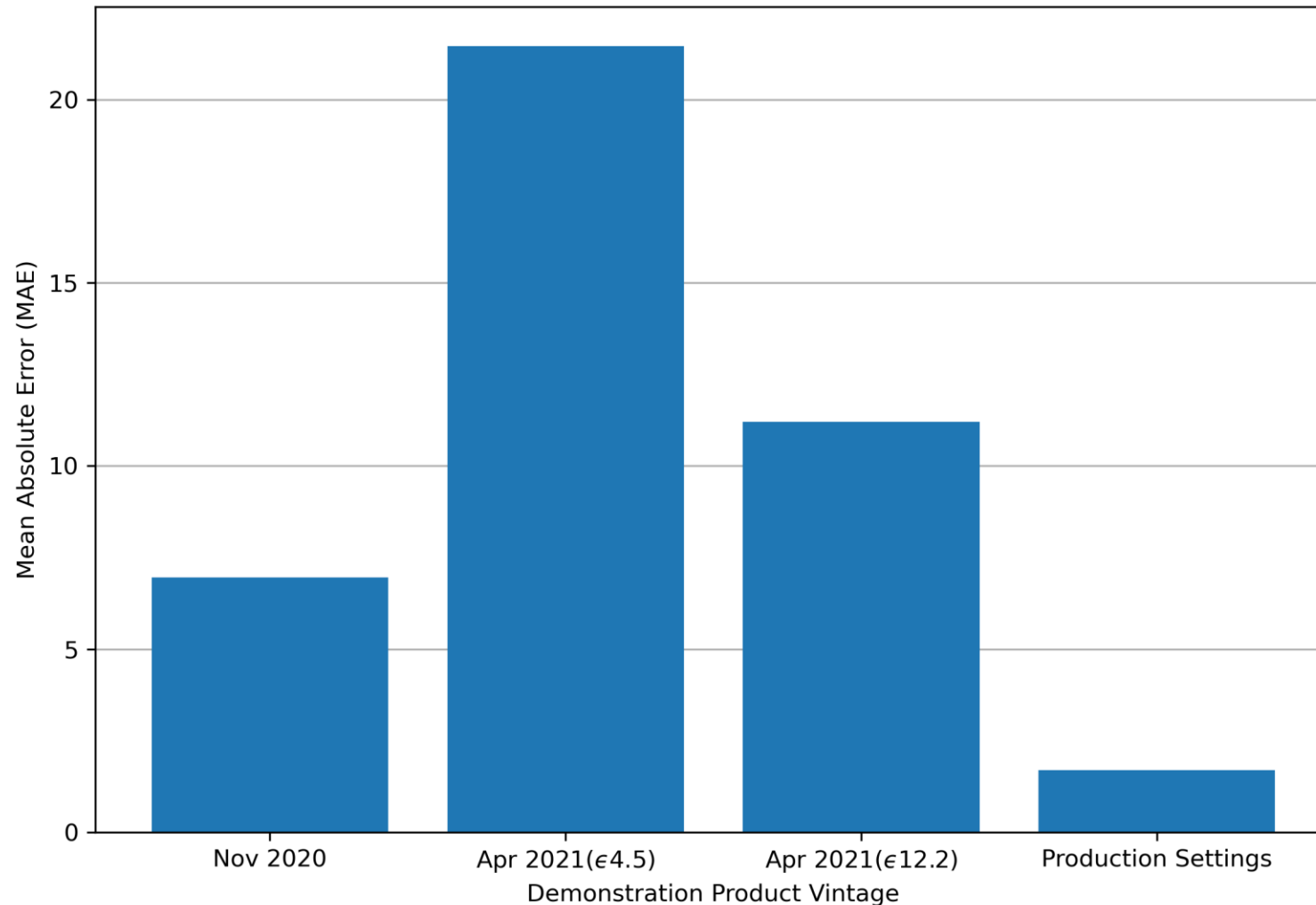
Total Population – Minor Civil Divisions

Minor Civil Division
Total Population: Mean Absolute Error (MAE) — All MCDs

Shape
your future
START HERE >

United States®
Census
2020
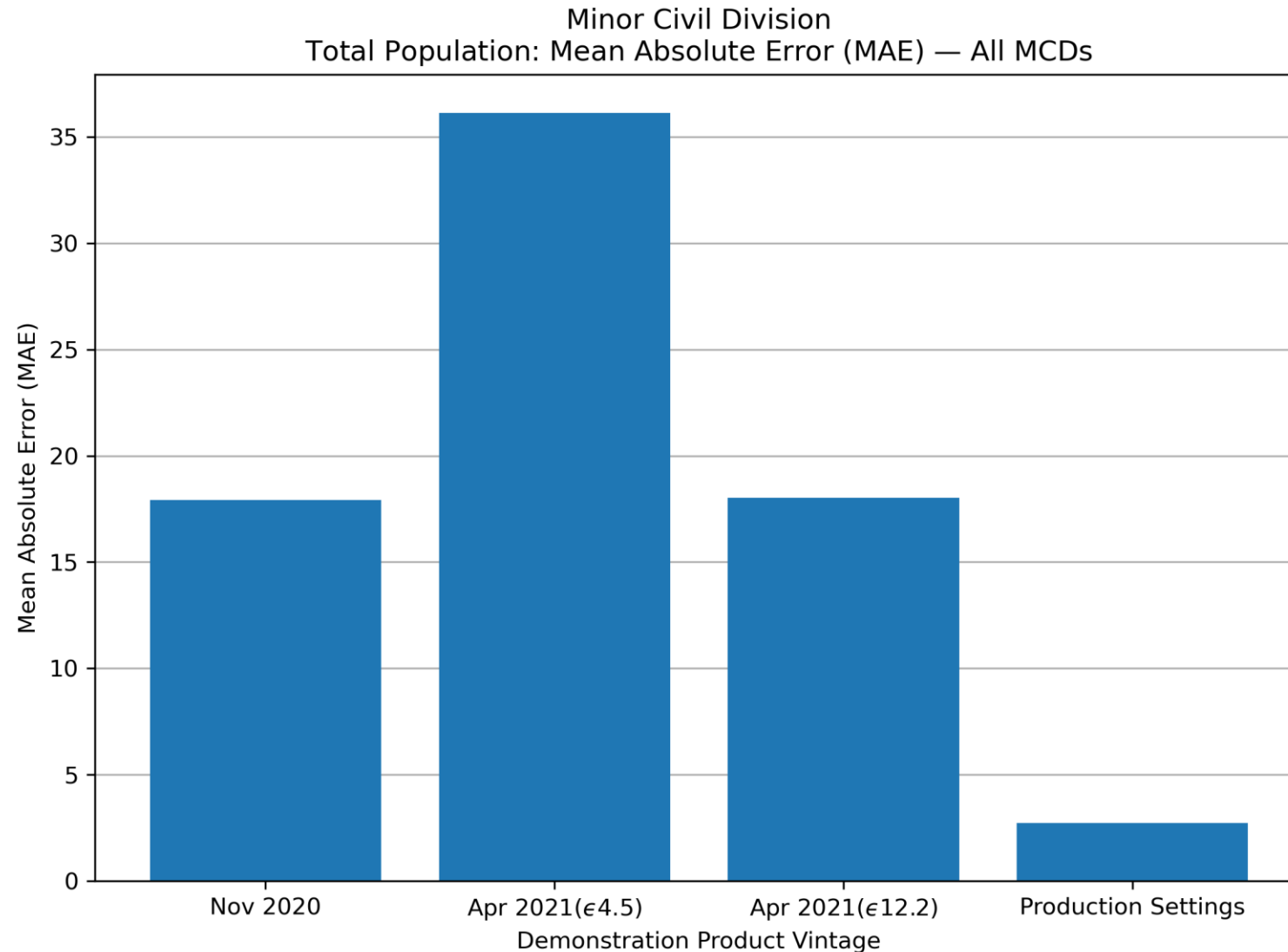
Total Population Aged 18+ – Various Geos.

Various Geographies — Population Aged 18 Years and Over: Mean Absolute Error (MAE)

2020CENSUS.GOV
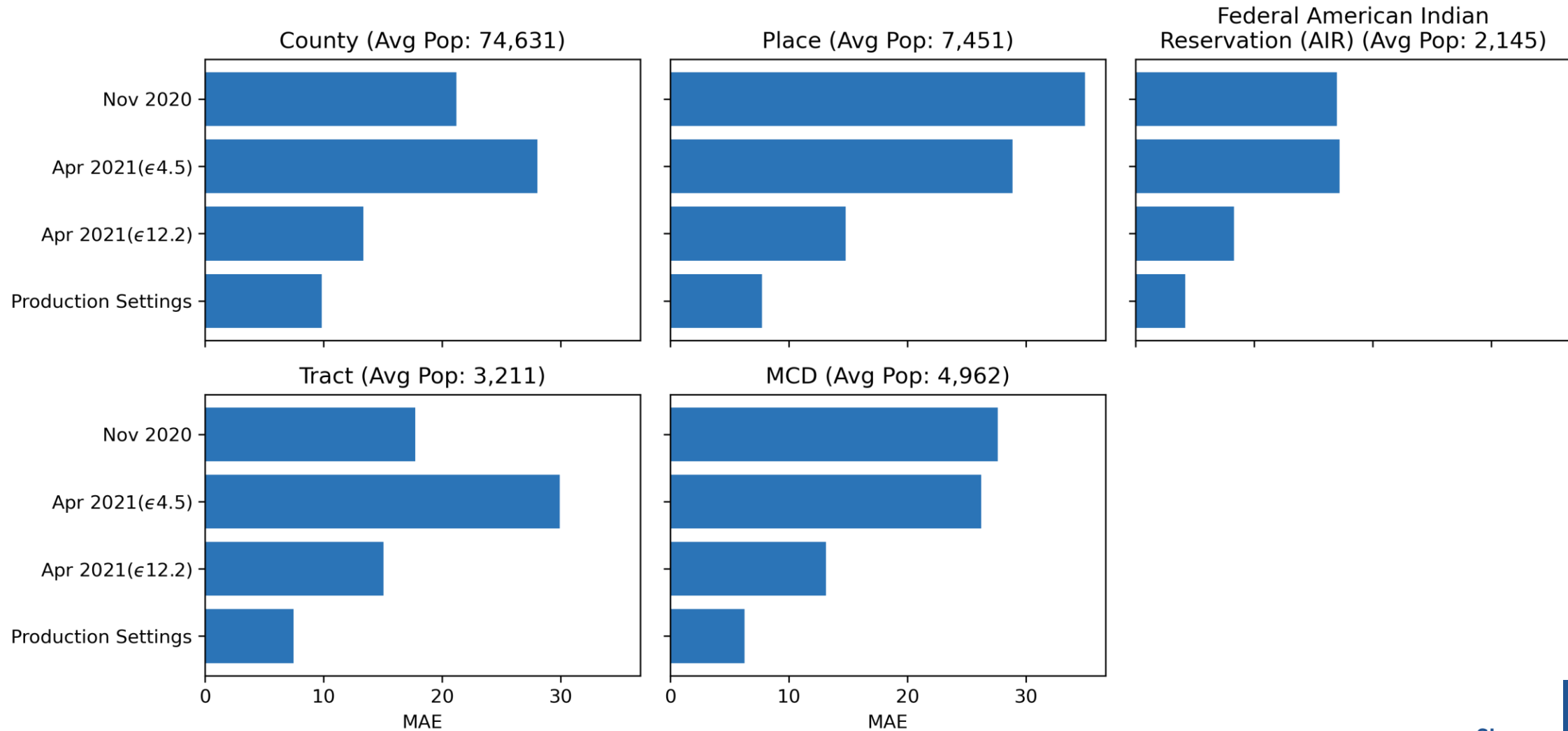
Race Alone – County

County — Race Alone Population: Mean Absolute Error (MAE)
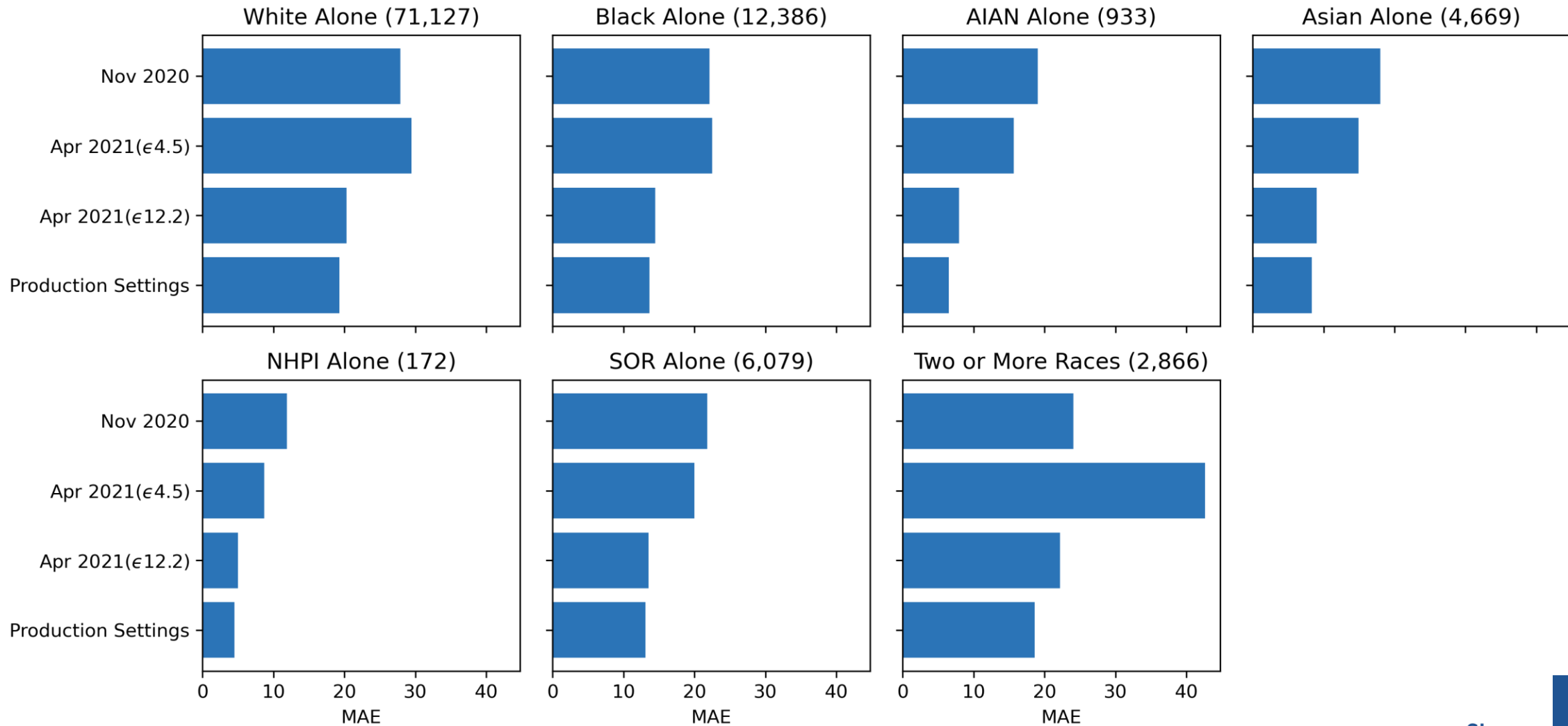
Race Alone – Tract
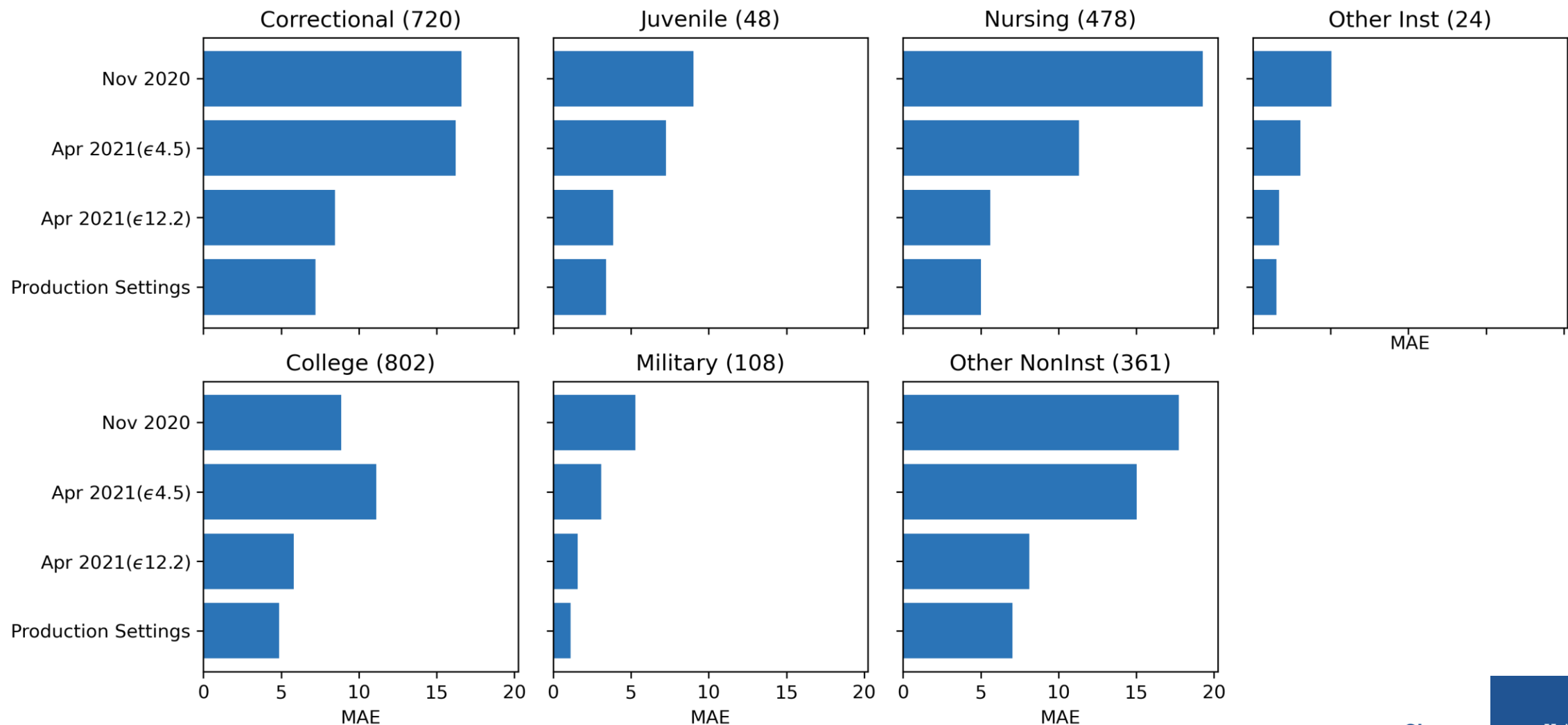

Tract — Race Alone Population: Mean Absolute Error (MAE)
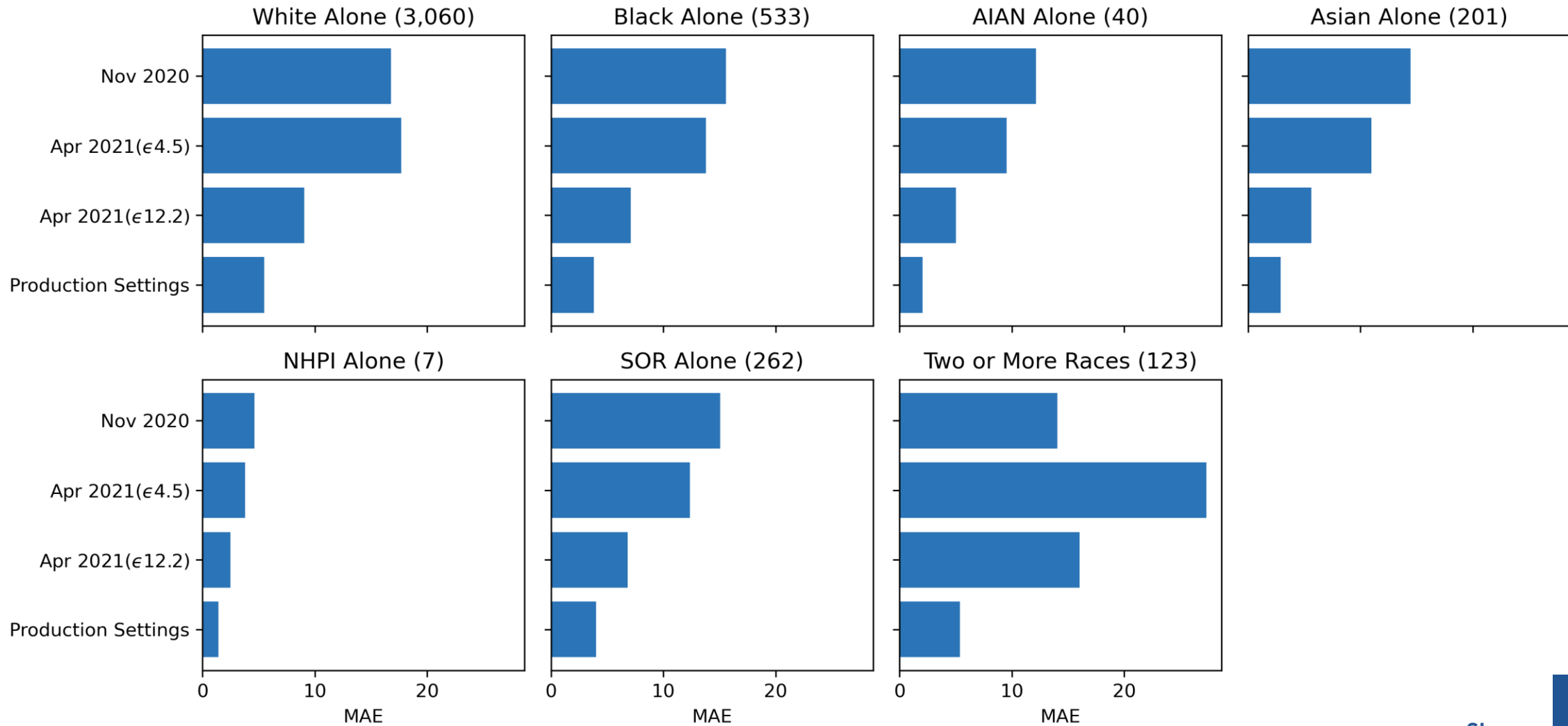
Race Alone – Federal AIR

Federal AIR — Race Alone Population: Mean Absolute Error (MAE)

2020CENSUS.GOV

Shape your future START HERE >

United States® Census 2020
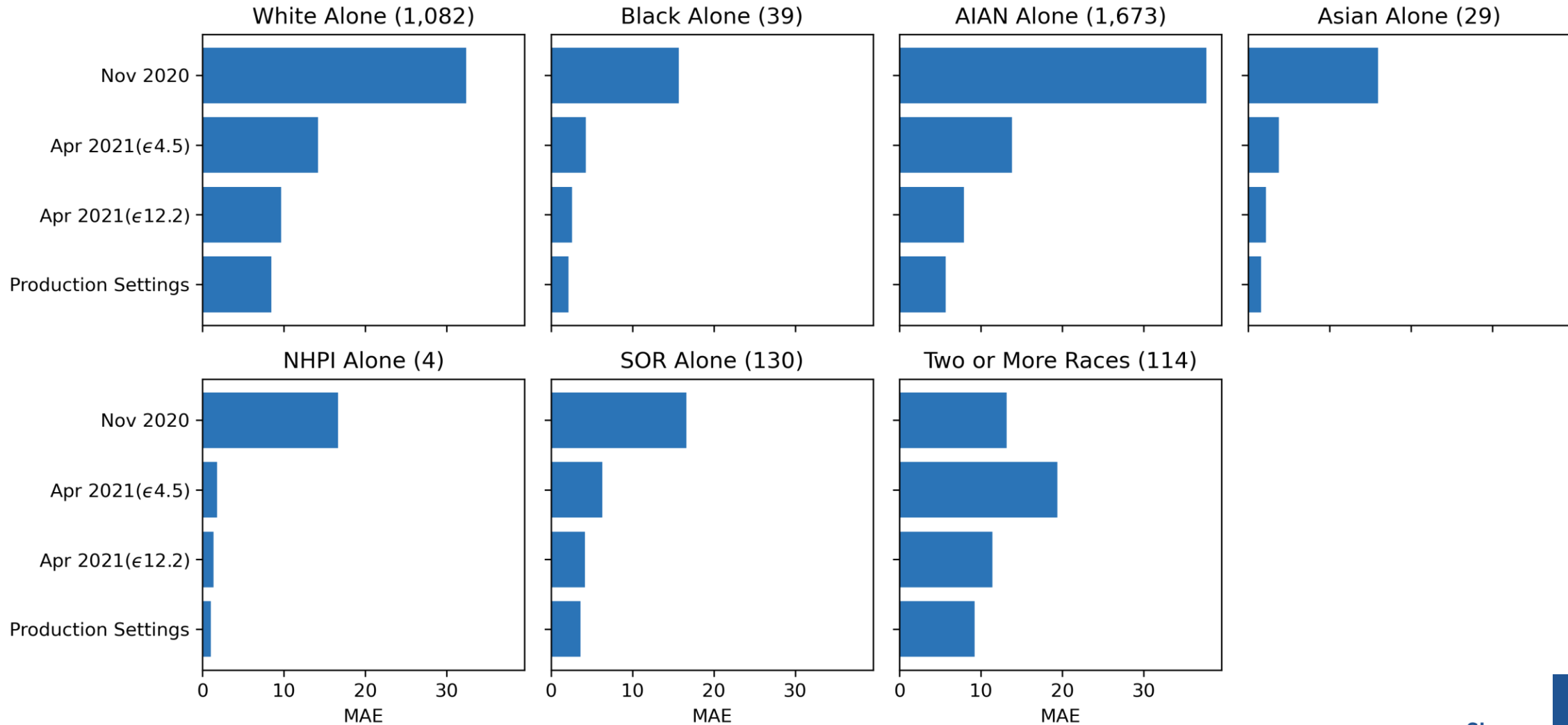
Race Alone or in Combination – County

County — Race Alone or In Combination (AOIC) Population: Mean Absolute Error (MAE)
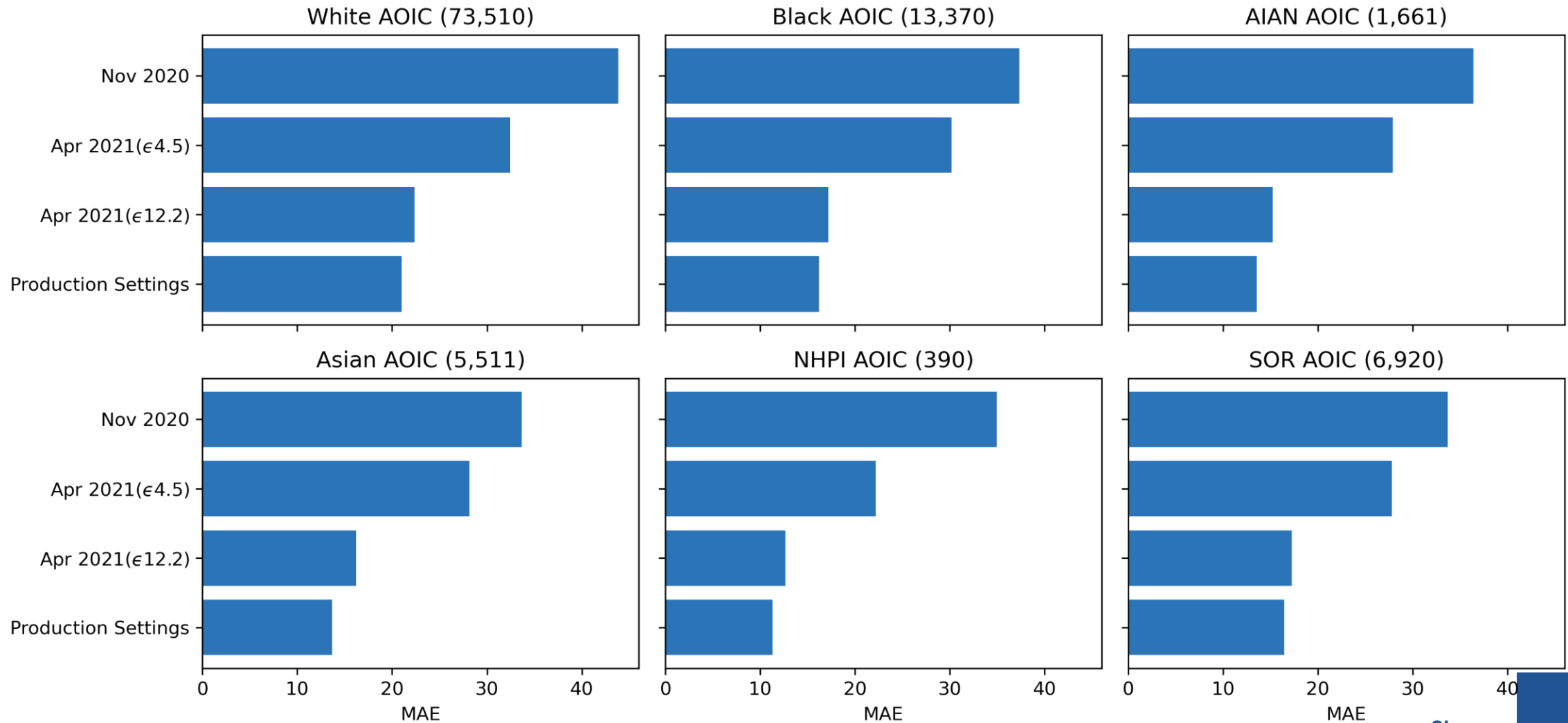
Shape your future
START HERE >

United States® Census 2020

Hispanic x Selected Race Alone – County

County — Hispanic x Race Alone Population: Mean Absolute Error (MAE)
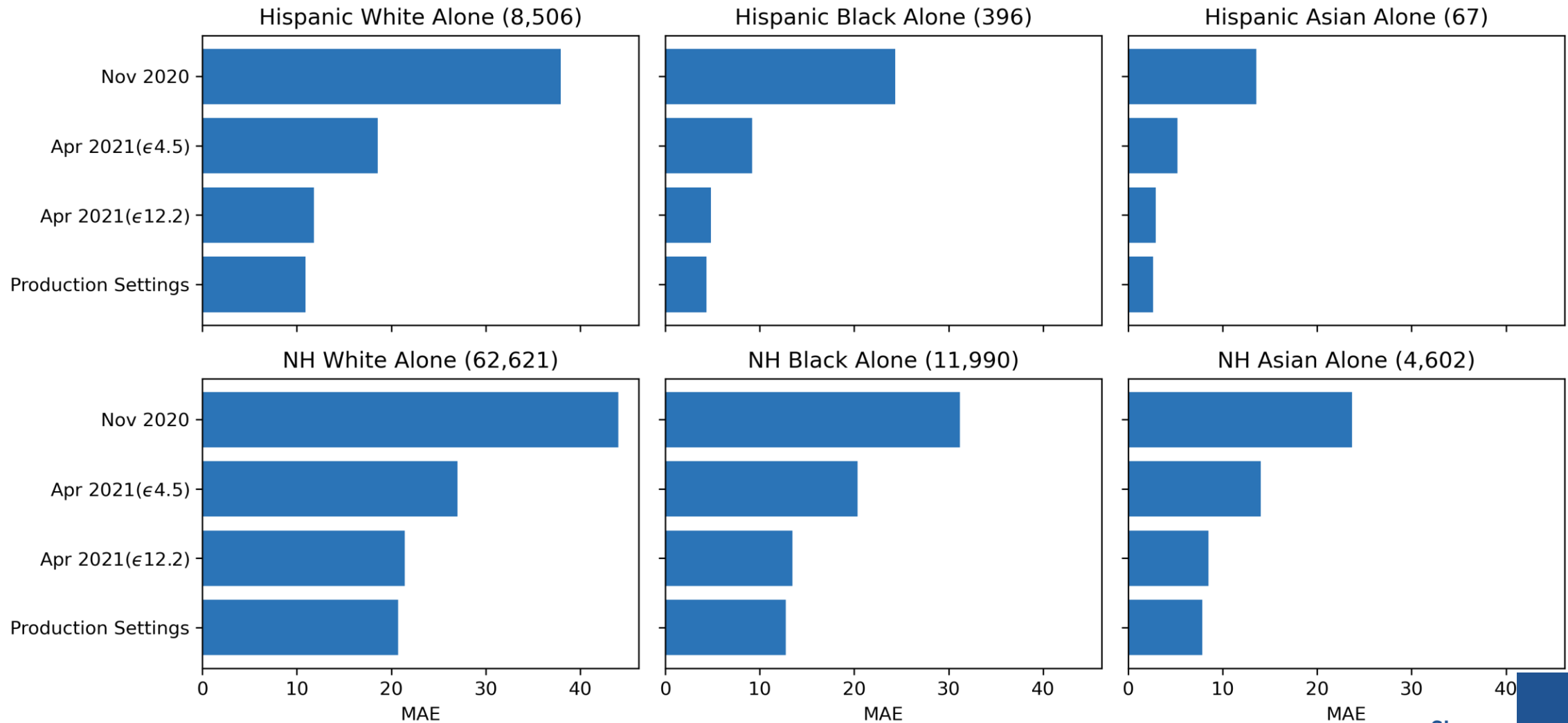
Hispanic x Selected Race Alone – County

County — Hispanic x Race Alone Population: Mean Absolute Error (MAE)
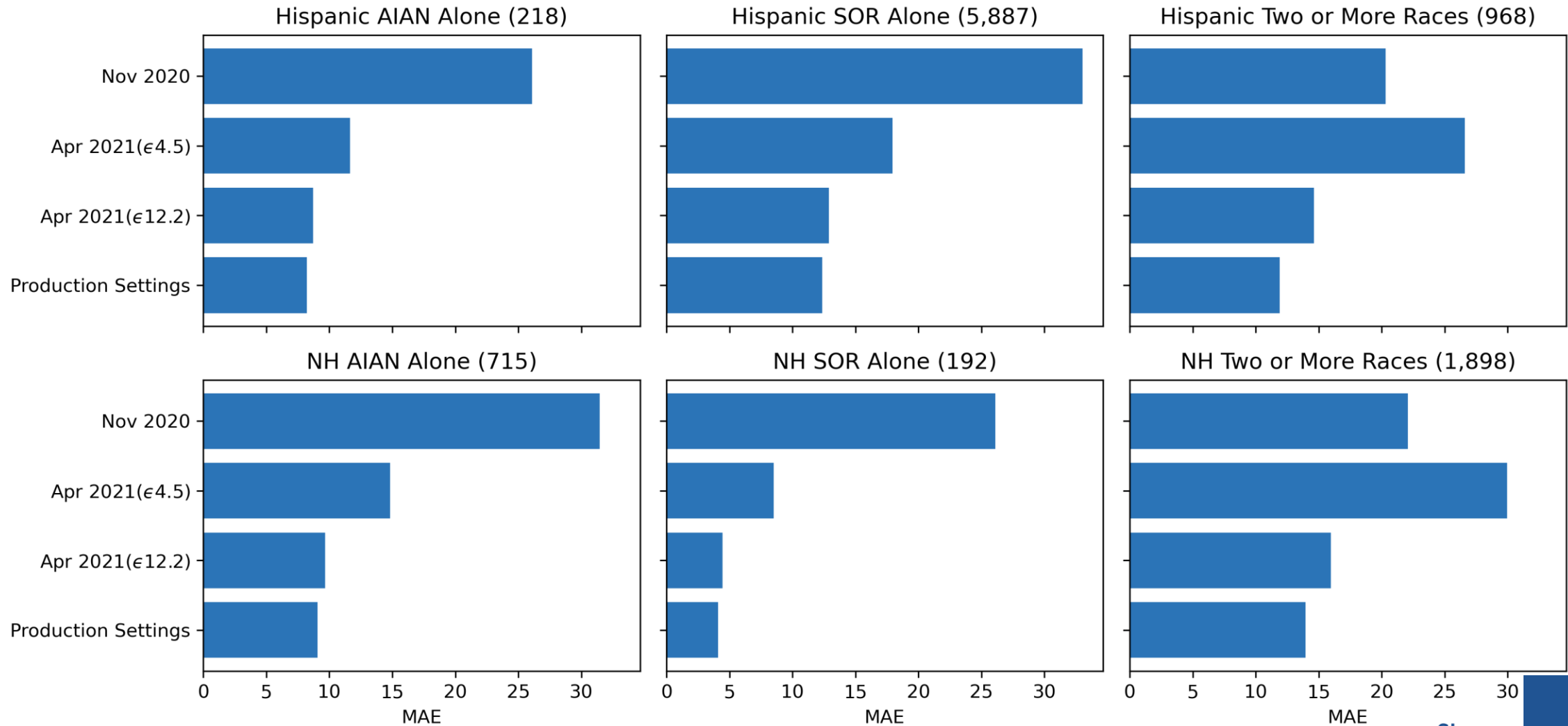
Hispanic x Selected Race Alone – Tract

Tract — Hispanic x Race Alone Population: Mean Absolute Error (MAE)
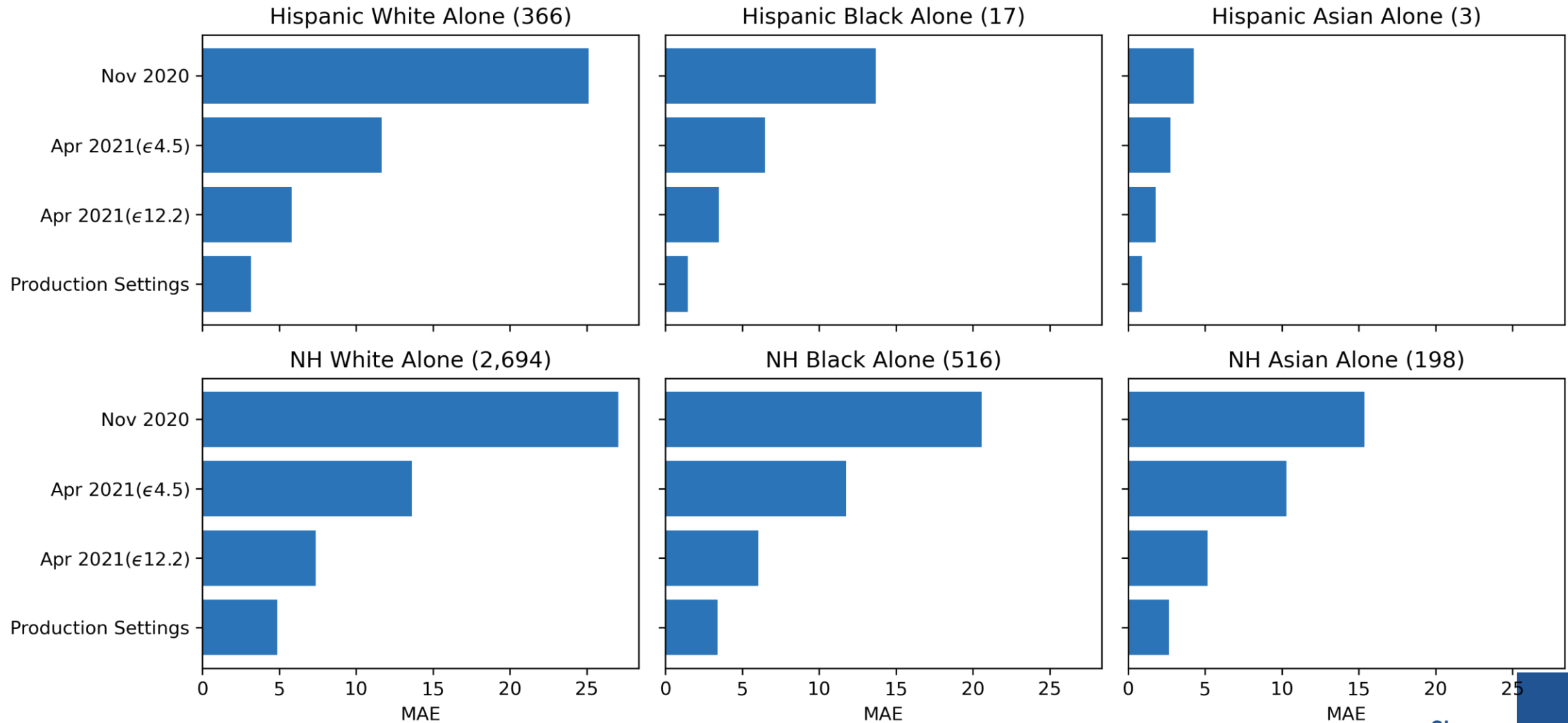
Hispanic x Selected Race Alone – Tract

Tract — Hispanic x Race Alone Population: Mean Absolute Error (MAE)
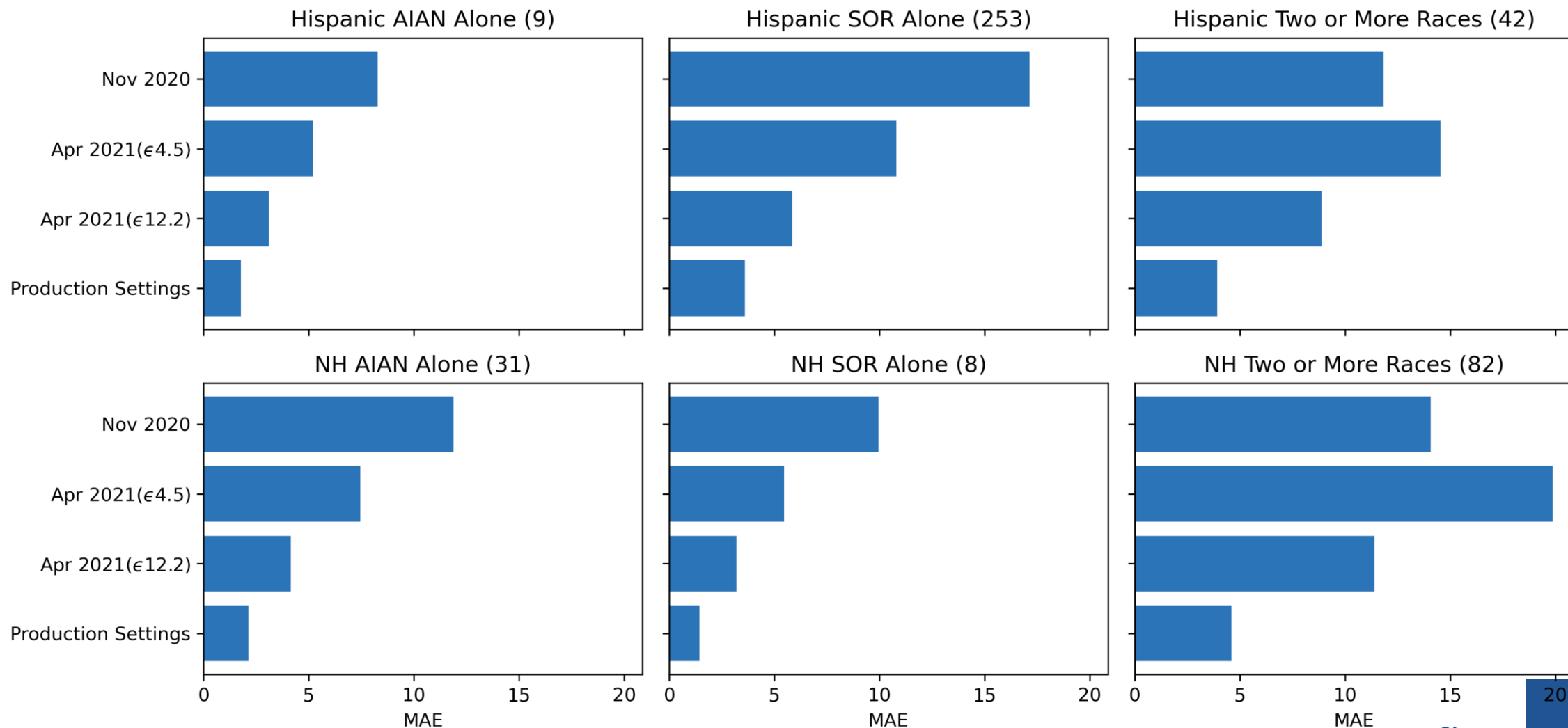
GQ Major Type – County

County — GQ Population by Major GQ Type: Mean Absolute Error (MAE)

GQ Major Type – Tract

Tract — GQ Population by Major GQ Type: Mean Absolute Error (MAE)

2020CENSUS.GOV

Shape
your future
START HERE >

United States®
Census
2020
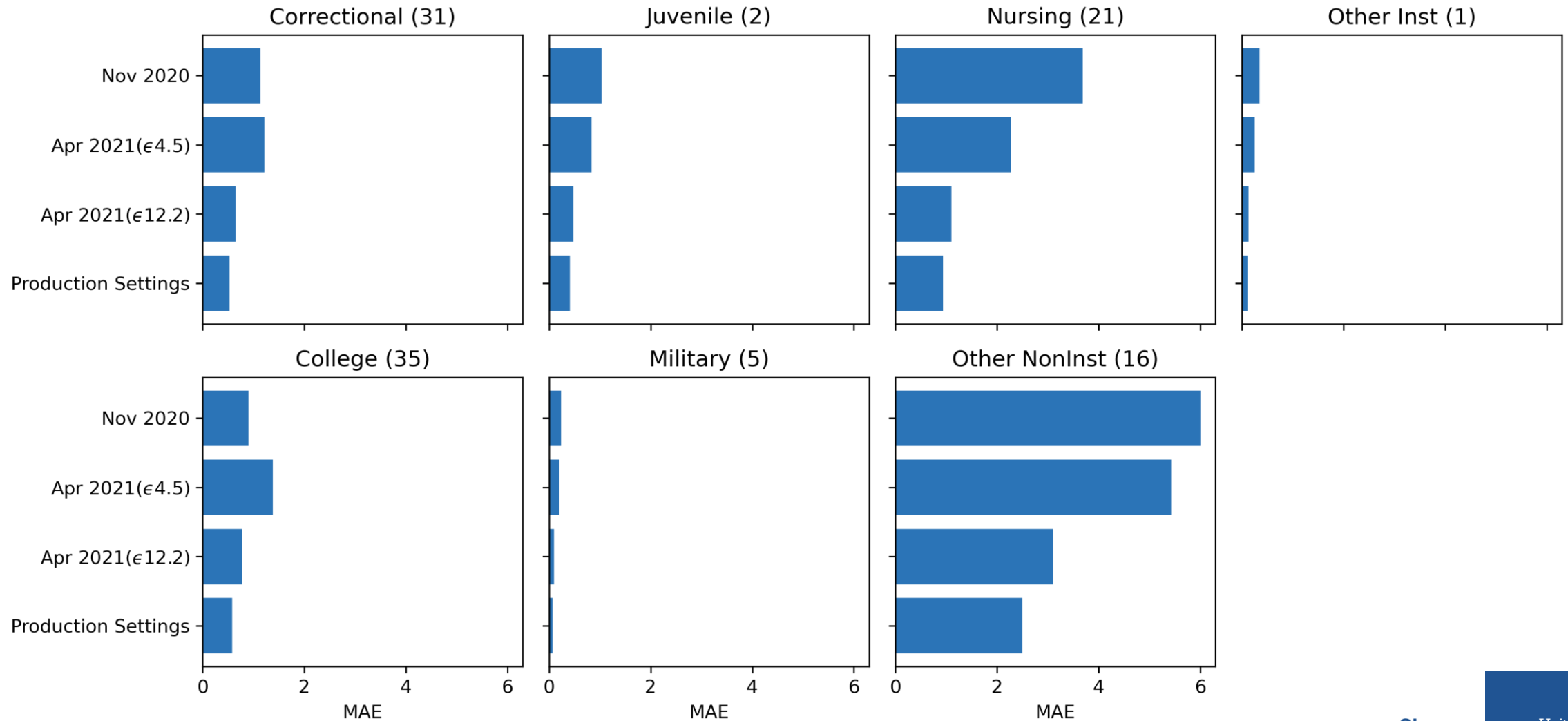
Addressing Bias

April 2021 PPMF

Block Groups

Diversity Quintile	Mean Difference In Total Population
0 – Least Diverse	5.04
1	4.24
2	0.99
3	-2.21
4 – Most Diverse	-8.07

Tracts

Diversity Quintile	Mean Difference In Total Population
0 – Least Diverse	15.95
1	11.15
2	3.01
3	-6.17
4 – Most Diverse	-23.94

Production Settings

Diversity Quintile	Mean Difference In Total Population
0 – Least Diverse	-0.375
1	1.009
2	0.997
3	-0.303
4 – Most Diverse	-1.352

Diversity Quintile	Mean Difference In Total Population
0 – Least Diverse	0.029
1	0.045
2	0.000
3	-0.020
4 – Most Diverse	-0.053

Shape your future START HERE >

United States® Census 2020

Total Population – Urban Block

Urban Block Total Population:
Mean Absolute Error (MAE) — Urban Blocks
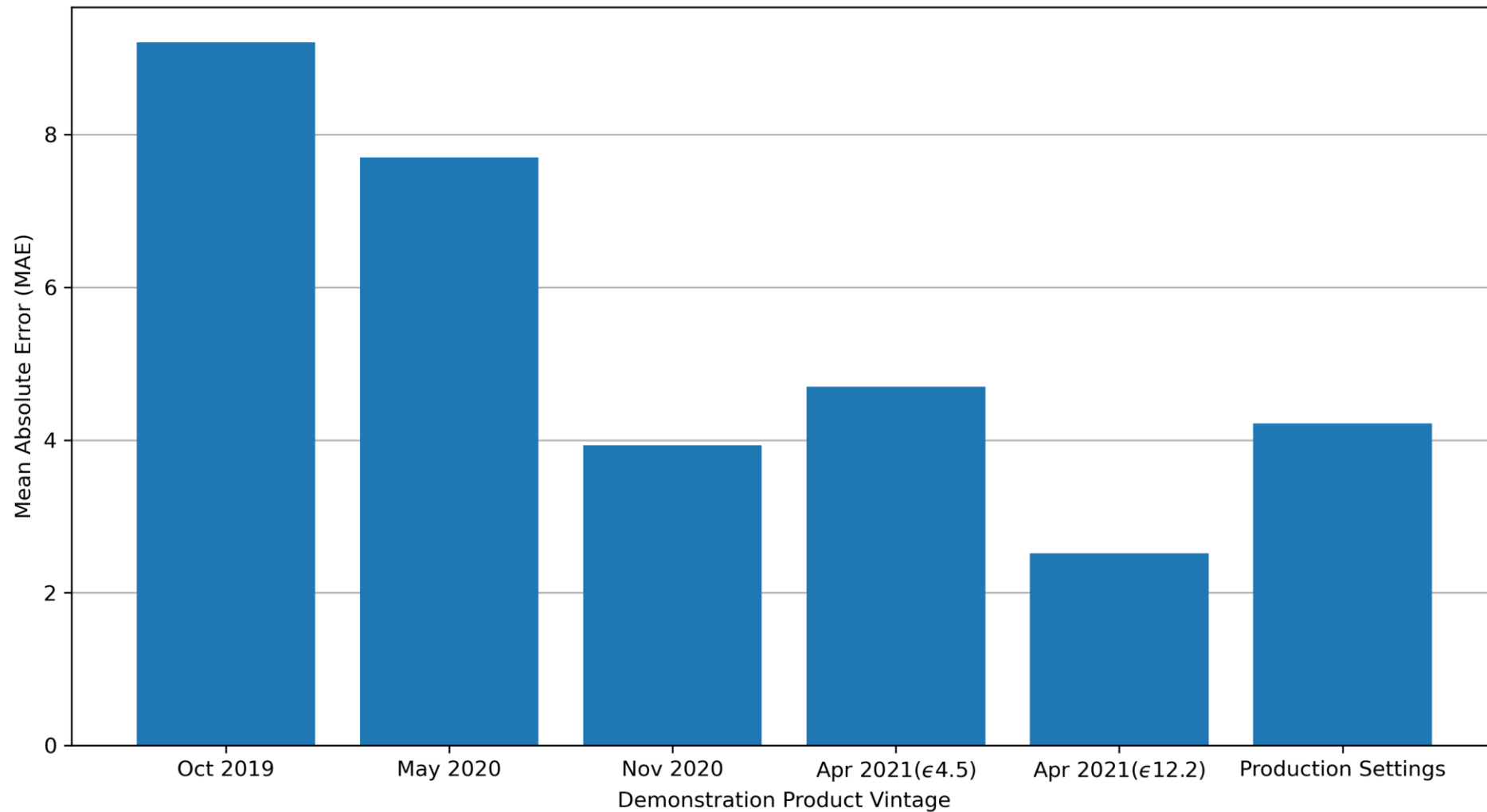
Total Population – Rural Block

Rural Block Total Population:
Mean Absolute Error (MAE) — Rural Blocks

your future
START HERE >

United States®
Census
2020
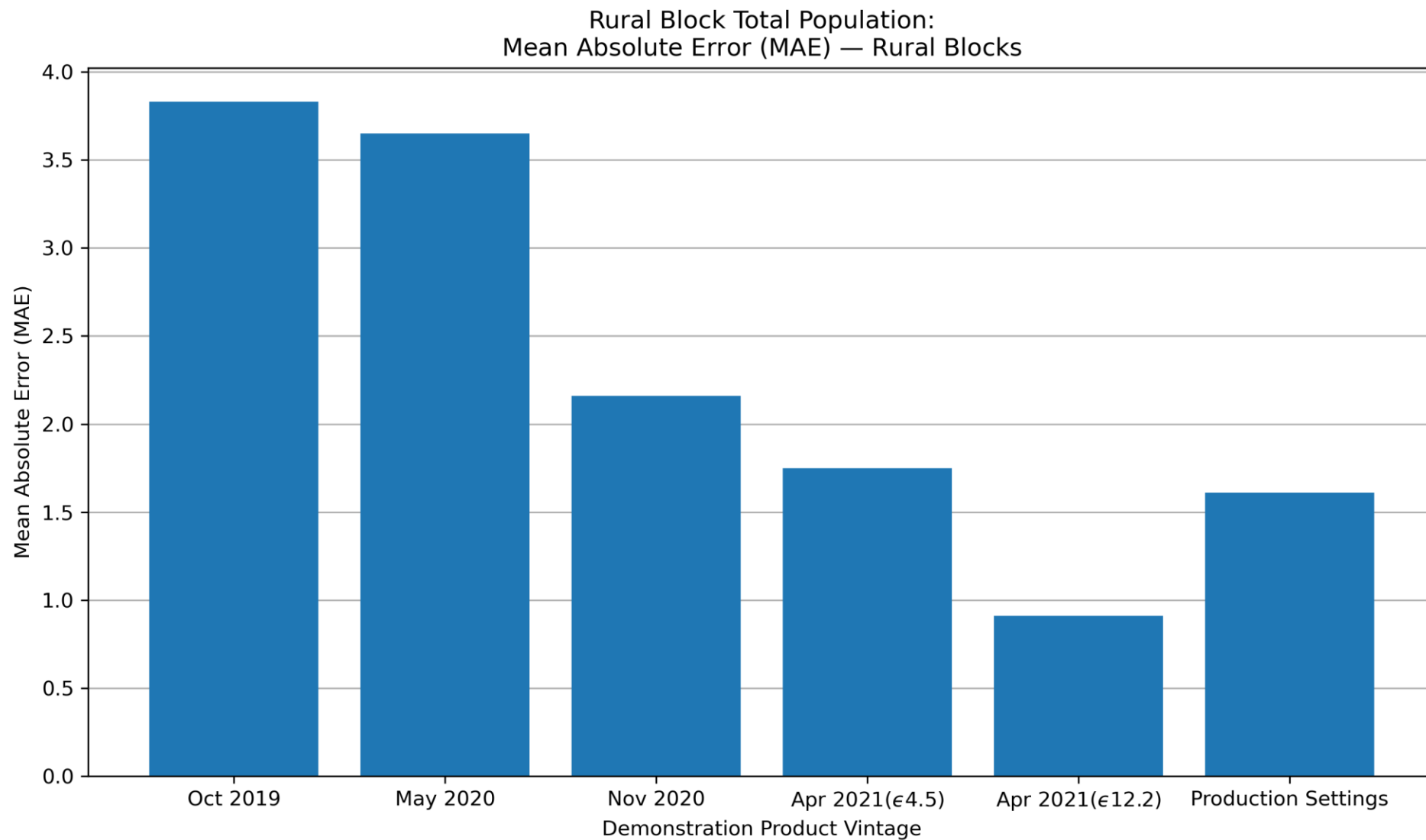
Block-Level Inconsistencies

Inconsistency	April 2021 ε12.2 Count of Blocks	Production Settings Count of Blocks
Occupied Housing Units > Household Population	203,519	303,984
Zero Occupied Housing Units; >0 Household Population	674,598	505,840
Zero Household Population; >0 Occupied Housing Units	77,947	148,836
Everyone in Block Under 18	90,534	163,884
> 10 Persons Per Household	87,342	121,376

Shape your future
START HERE >

United States®
Census
2020

Stay Informed:
Subscribe to the 2020 Census Data Products Newsletters

*Search "Disclosure Avoidance" at www.census.gov

Stay Informed:
Visit Our Website

*Search "Disclosure Avoidance" at www.census.gov
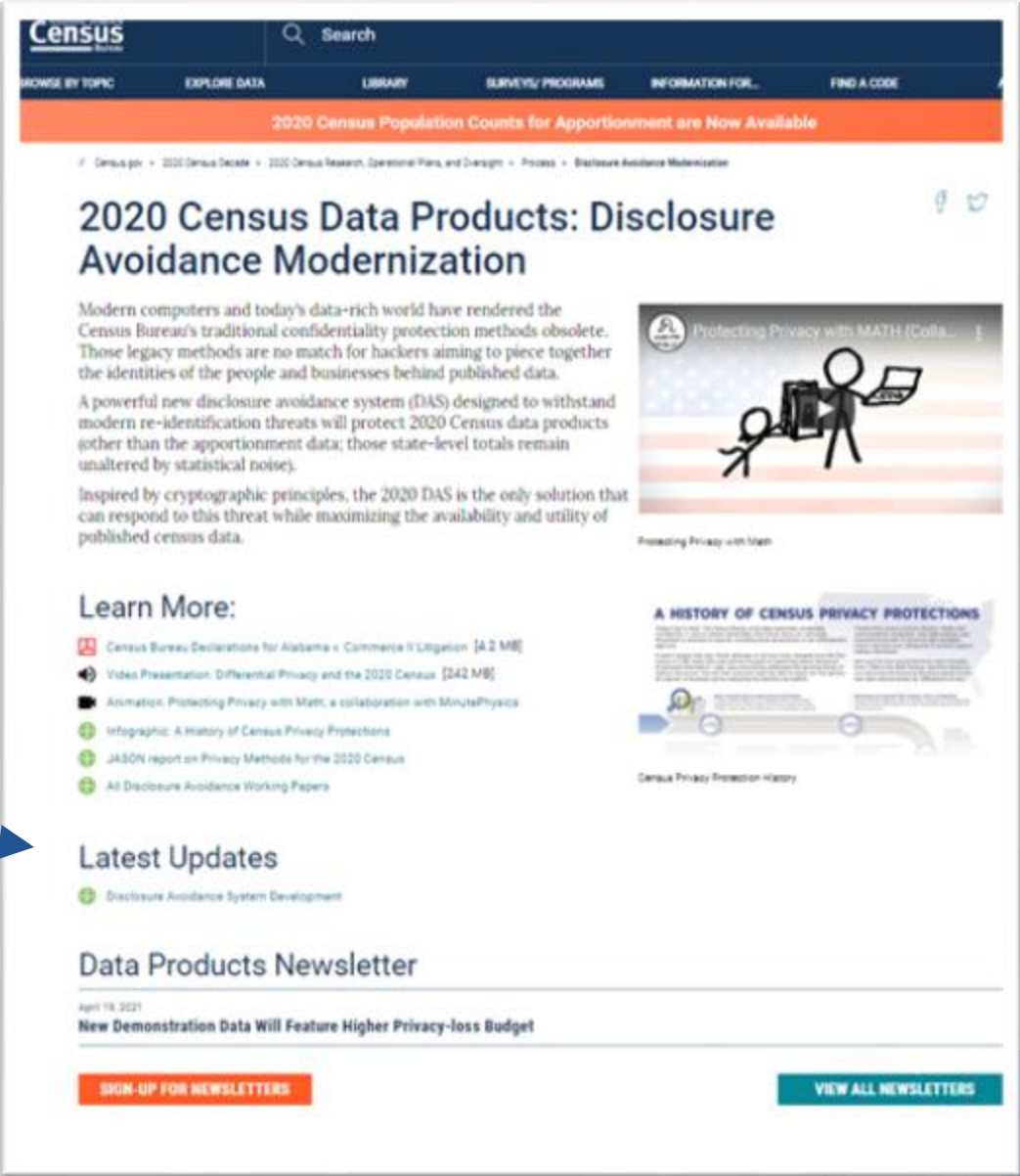
** New Video **

[Protecting Privacy in Census Bureau Statistics](#)

*Find it on our website and YouTube Page

Search "Disclosure Avoidance" at www.census.gov

Shape
your future
START HERE >

United States®
Census
2020

Questions?

Shape
your future
START HERE >

United States®
Census
2020