Understanding the 2020 Census Disclosure Avoidance System:

Analysis of Production Settings for Redistricting and Voting Rights Act Use Cases

Michael Hawes
Senior Advisor for Data Access and Privacy
August 10, 2021

Acknowledgements

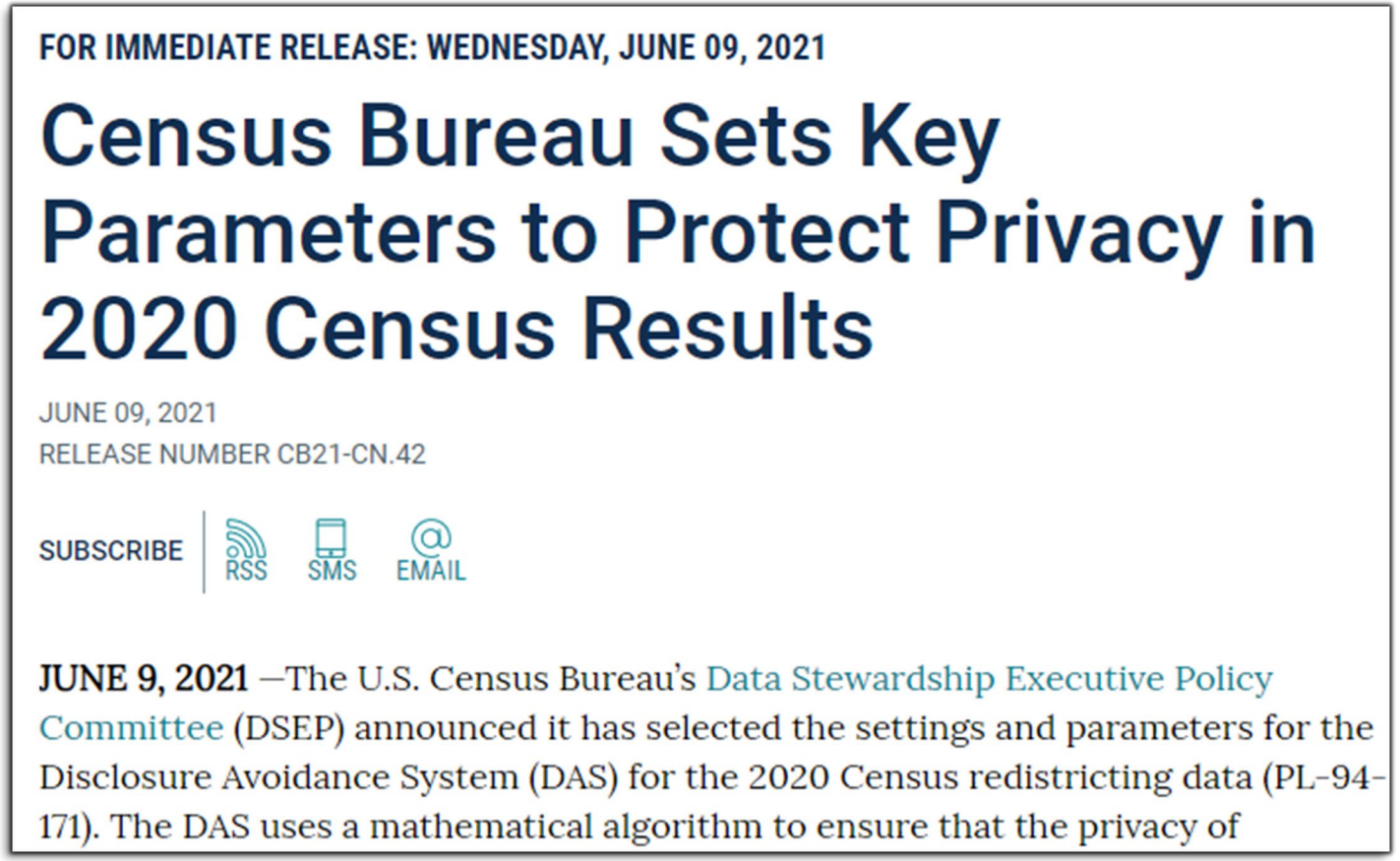
This presentation includes work by the Census Bureau's 2020 Disclosure Avoidance System development team, Census Bureau colleagues, and our collaborators, from the following Census Bureau divisions and outside organizations: ADCOM, ADDC, ADRM, CED, CEDDA, CEDSCI, CES, CSRM, DCMD, DITD, ESMD, GEO, POP, TAB, CDF, Econometrica Inc., Galois, Knexus Research Corp, MITRE, NLT, TI, and Tumult Labs.

We also acknowledge and greatly appreciate the ongoing feedback we have received from external stakeholder groups that has contributed to the design and improvement of the Disclosure Avoidance System.

For more information and technical details relating to the issues discussed in these slides, please contact the author at michael.b.hawes@census.gov.

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.

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

Stakeholder Feedback

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

Major themes included:

- **Accuracy for American Indian and Alaska Native 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**

Key Parameters and Improvements

- **Privacy-loss Budget (PLB):**
  - **ε= 17.14 for persons**
  - **ε= 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**

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

Privacy-loss Budget Allocations
by geographic level

April 2021 Demonstration Data:

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:

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

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.

Public Release of the 2020 Census P.L. 94-171 Redistricting Data Summary Files

The Census Bureau will release these data on its public FTP site on August 12, 2021.

The Census Bureau will release the same data in easier-to-use formats by September 30, 2021.

For more information, visit the Census Bureau's website at www.census.gov.

Production Settings 2010 Demonstration Data Release

In parallel with the release of the 2020 Census P.L. 94-171 Redistricting Data Summary Files, the Census Bureau will also be releasing the final set of 2010 Census Demonstration Data reflecting the final production settings used by the 2020 Census Disclosure Avoidance System to protect the 2020 Census redistricting data.

These are the same data used to produce the Production Settings Detailed Summary Metrics released on July 1, 2021.

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

Analyses of 2010 Census Demonstration Data for Redistricting and Voting Rights Act Use Cases

- **Empirical Study of Two Aspects of the Topdown Algorithm Output for Redistricting: Reliability & Variability**

  Updated version of Wright and Irimata (2021) study.

- **Analysis of the impact of DAS production settings on the identification of majority-minority districts**

Comparisons to Published 2010 Census Data

Both sets of analyses discussed today involve comparisons of the 2010 Census Demonstration Data (protected with the differentially private 2020 Census Disclosure Avoidance System) to published 2010 Census tabulations.

The 2010 Census used a form of noise infusion known as data swapping, wherein households' records were swapped across geographies to protect the confidentiality of individuals' census responses.

For the 2010 Census, the number of individuals and the number of voting-age individuals in each household were held invariant (no noise was added), but individuals' characteristics (e.g., race and Hispanic origin) were swapped across geographies.

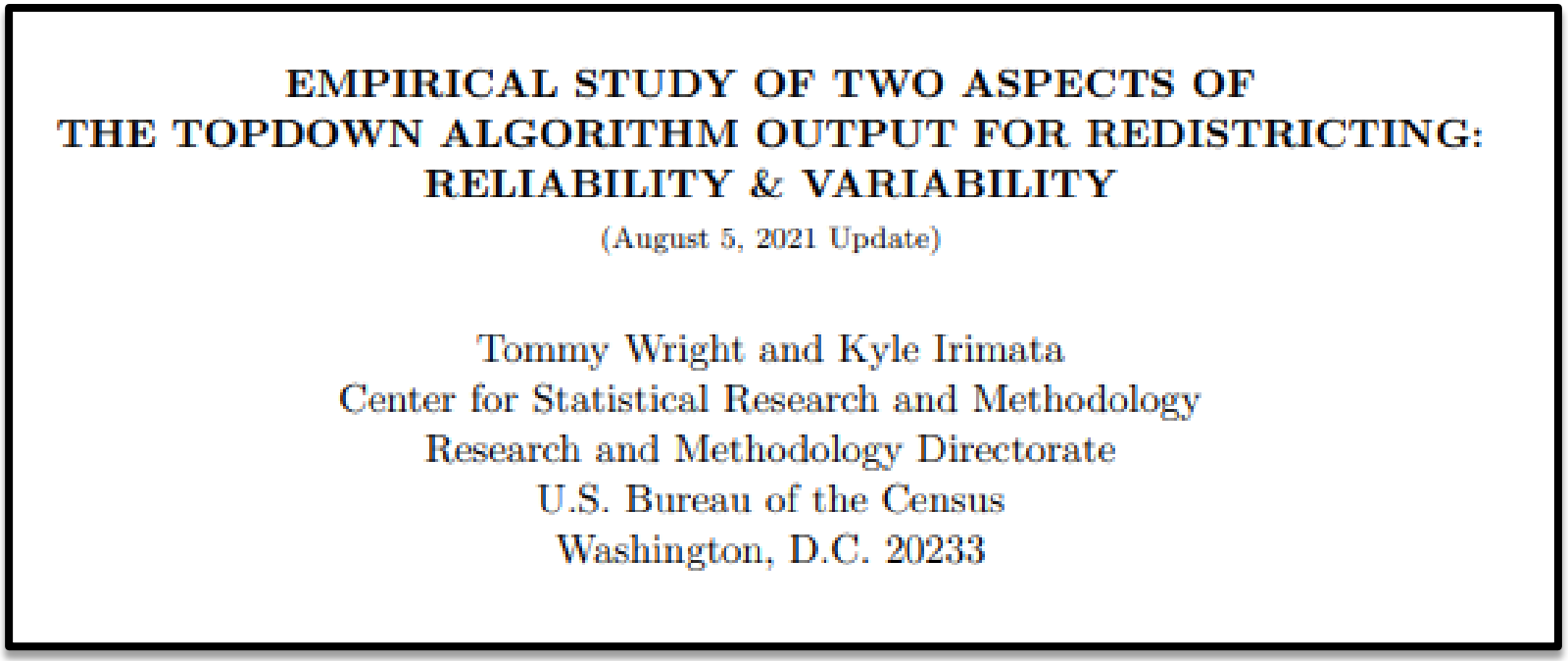
Differences between the Demonstration Data and the published 2010 Census tabulations presented in the following analyses reflect both the noise from the 2020 Census Disclosure Avoidance System and impact of the 2010 Census swapping methodology on characteristics data.

Average Error in Total Population Counts by Geographic Level

Geographic Level	Mean Absolute Error (MAE) (# of persons)
Counties	1.75
Minor Civil Divisions (MCD)	2.74
Incorporated Places	3.55
Tracts	1.93
Blocks (Urban blocks)	4.22
Blocks (Rural blocks)	1.61

Source: Production Settings Detailed Summary Metrics CBDRB-FY20-DSEP-001

Wright and Irimata (2021)



EMPIRICAL STUDY OF TWO ASPECTS OF THE TOPDOWN ALGORITHM OUTPUT FOR REDISTRICTING: RELIABILITY & VARIABILITY

(August 5, 2021 Update)

Tommy Wright and Kyle Irimata
Center for Statistical Research and Methodology
Research and Methodology Directorate
U.S. Bureau of the Census
Washington, D.C. 20233

Available at: https://www.census.gov/library/working-papers/2021/adrm/SSS2021-02.html

Wright and Irimata (2021)

Research Questions:

(1) What is the minimum TOTAL population of a district to have reliable characteristics of various demographic groups for redistricting purposes?

Because districts cannot be defined ex ante, this portion of their analysis uses two existing geographic levels as proxies for districts:

- Block Groups (“on spine”)
- Places and Minor Civil Divisions (“off spine”)

(2) How variable are data protected using the 2020 Disclosure Avoidance System for districts in Rhode Island and for three additional jurisdictions?

Wright and Irimata (2021)

Part 1: Reliability of demographic characteristics

Definition 1: Let $C_{SWA}(g)$ and $C_{TDA}(g)$ be two competing counts of the demographic group g associated with a block group (more generally, geographic district) whose total population counts are C_{SWA} and C_{TDA}, respectively. The **difference of ratios** is the absolute value of the difference between the SWA ratio $\frac{C_{SWA}(g)}{C_{SWA}}$ and the TDA ratio $\frac{C_{TDA}(g)}{C_{TDA}}$, given by:

$$DR_g = \left| \frac{C_{SWA}(g)}{C_{SWA}} - \frac{C_{TDA}(g)}{C_{TDA}} \right|. \quad (1)$$

Small values of the difference of ratios DR_g imply that the ratios for a group g due to SWA and TDA in the block group, MCD, or place are close.

Definition 2: When DR_g is sufficiently small while comparing a $C_{SWA}(g)$ count and corresponding $C_{TDA}(g)$ count for a demographic group g associated with a given block group, MCD or place, we say that the $C_{TDA}(g)$ count (or ratio) provides a **reliable characteristic** for the block group, MCD, or place.

Wright and Irimata (2021)

Part 1: Reliability of demographic characteristics

Table 1a: Block Group 240317044041 (564 HUs) Characteristics

($C_{TDA}(g)$ counts result from 2020 Census Redistricting Data Production Settings ($\epsilon = 17.14$ for persons) version of *TDA*.)

Demographic Group (g)[b]	$C_{SWA}(g)$	$C_{TDA}(g)$	$DR_g = \left\lvert \frac{C_{SWA}(g)}{C_{SWA}} - \frac{C_{TDA}(g)}{C_{TDA}} \right\rvert$
TOTAL	1,560	1,598	[c]
TOTAL18	1,198	1,229	[c]
TOTALHISP	133	141	$\left\lvert \frac{133}{1,560} - \frac{141}{1,598} \right\rvert = 0.0030$
TOTALNH	1,427	1,457	$\left\lvert \frac{1,427}{1,560} - \frac{1,457}{1,598} \right\rvert = 0.0030$
WHITENH	1,169	1,178	$\left\lvert \frac{1,169}{1,560} - \frac{1,178}{1,598} \right\rvert = 0.0122$
BLACKNH	36	54	$\left\lvert \frac{36}{1,560} - \frac{54}{1,598} \right\rvert = 0.0107$
AIANNH	10	8	$\left\lvert \frac{10}{1,560} - \frac{8}{1,598} \right\rvert = 0.0014$
ASIANNH	187	189	$\left\lvert \frac{187}{1,560} - \frac{189}{1,598} \right\rvert = 0.0016$
HPINH	5	2	$\left\lvert \frac{5}{1,560} - \frac{2}{1,598} \right\rvert = 0.0020$
OTHERNH	11	12	$\left\lvert \frac{11}{1,560} - \frac{12}{1,598} \right\rvert = 0.0005$
MLTMNNH	9	14	$\left\lvert \frac{9}{1,560} - \frac{14}{1,598} \right\rvert = 0.0030$

Wright and Irimata (2021)

Part 1: Reliability of demographic characteristics

Table 1a: Block Group 240317044041 (564 HUs) Characteristics

($C_{TDA}(g)$ counts result from 2020 Census Redistricting Data Production Settings ($\epsilon = 17.14$ for persons) version of *TDA*.)

Demographic Group (g)[b]	$C_{SWA}(g)$	$C_{TDA}(g)$	$DR_g = \left\lvert \frac{C_{SWA}(g)}{C_{SWA}} - \frac{C_{TDA}(g)}{C_{TDA}} \right\rvert$
TOTAL	1,560	1,598	$\left\lvert \frac{1,598 - 1,560}{1,560} \right\rvert = 0.024$
TOTAL18	1,198	1,229	$\left\lvert \frac{1,229 - 1,198}{1,198} \right\rvert = 0.026$
TOTALHISP	133	141	$\left\lvert \frac{133}{1,560} - \frac{141}{1,598} \right\rvert$
TOTALNH	1,427	1,457	$\left\lvert \frac{1,427}{1,560} - \frac{1,457}{1,598} \right\rvert = 0.0030$
WHITENH	1,169	1,178	$\left\lvert \frac{1,169}{1,560} - \frac{1,178}{1,598} \right\rvert = 0.0122$
BLACKNH	36	54	$\left\lvert \frac{36}{1,560} - \frac{54}{1,598} \right\rvert = 0.0107$
AIANNH	10	8	$\left\lvert \frac{10}{1,560} - \frac{8}{1,598} \right\rvert = 0.0014$
ASIANNH	187	189	$\left\lvert \frac{187}{1,560} - \frac{189}{1,598} \right\rvert = 0.0016$
HPINH	5	2	$\left\lvert \frac{5}{1,560} - \frac{2}{1,598} \right\rvert = 0.0020$
OTHERNH	11	12	$\left\lvert \frac{11}{1,560} - \frac{12}{1,598} \right\rvert = 0.0005$
MLTMNNH	9	14	$\left\lvert \frac{9}{1,560} - \frac{14}{1,598} \right\rvert = 0.0030$

Table 3: Proportion of Block Groups in Each Stratum for Three Criteria

(Proportion computations result from 2020 Census Redistricting Data Production Settings ($\epsilon = 17.14$ for persons) version of *TDA*.)

Population: United States (50 States & DC)

		Reliable Characteristics Criteria		
Stratum for Block Groups Using C_{SWA} for TOTAL	Number of Block Groups	Criterion I LDG $DR_g \leq 0.01$	Criterion II LDG $DR_g \leq 0.03$	Criterion III LDG $DR_g \leq 0.05$
$50 \leq C_{SWA} \leq 99$	128	0.1250	0.3594	0.5156
$100 \leq C_{SWA} \leq 149$	99	0.1818	0.5253	0.7071
$150 \leq C_{SWA} \leq 199$	124	0.1694	0.5565	0.7581
$200 \leq C_{SWA} \leq 249$	154	0.2662	0.6234	0.7922
$250 \leq C_{SWA} \leq 299$	209	0.2919	0.6459	0.8565
$300 \leq C_{SWA} \leq 349$	264	0.3636	0.7348	0.8902
$350 \leq C_{SWA} \leq 399$	407	0.3366	0.7346	0.8698
$400 \leq C_{SWA} \leq 449$	569	0.4077	0.7750	0.9315
$450 \leq C_{SWA} \leq 499$	915	0.4087	0.8284	0.9552
$500 \leq C_{SWA} \leq 549$	1,699	0.4197	0.8458	0.9588
$550 \leq C_{SWA} \leq 599$	3,238	0.4546	0.8684	0.9654
$600 \leq C_{SWA} \leq 649$	5,131	0.4578	0.8827	0.9751
$650 \leq C_{SWA} \leq 699$	6,683	0.4718	0.8927	0.9753
$700 \leq C_{SWA} \leq 749$	7,356	0.5007	0.9082	0.9826

Table 3
(continued)

$950 \leq C_{SWA} \leq 999$	8,723	0.5849	0.9512	0.9952
$1{,}000 \leq C_{SWA} \leq 1{,}049$	8,398	0.6044	0.9582	0.9952
$1{,}050 \leq C_{SWA} \leq 1{,}099$	8,345	0.6192	0.9646	0.9965
$1{,}100 \leq C_{SWA} \leq 1{,}149$	7,950	0.6244	0.9701	0.9972
$1{,}150 \leq C_{SWA} \leq 1{,}199$	7,860	0.6422	0.9763	0.9977
$1{,}200 \leq C_{SWA} \leq 1{,}249$	7,451	0.6515	0.9757	0.9988
$1{,}250 \leq C_{SWA} \leq 1{,}299$	7,124	0.6645	0.9749	0.9978
$1{,}300 \leq C_{SWA} \leq 1{,}349$	6,714	0.6822	0.9812	0.9988
$1{,}350 \leq C_{SWA} \leq 1{,}399$	6,507	0.6859	0.9866	0.9989
$1{,}400 \leq C_{SWA} \leq 1{,}449$	5,911	0.7090	0.9866	0.9992
$1{,}450 \leq C_{SWA} \leq 1{,}499$	5,617	0.7002	0.9858	0.9995
$1{,}500 \leq C_{SWA} \leq 1{,}549$	5,390	0.7330	0.9900	0.9994
$1{,}550 \leq C_{SWA} \leq 1{,}599$	4,856	0.7341	0.9866	0.9994
$1{,}600 \leq C_{SWA} \leq 1{,}649$	4,508	0.7420	0.9918	0.9998
$1{,}650 \leq C_{SWA} \leq 1{,}699$	4,325	0.7489	0.9908	0.9998
$1{,}700 \leq C_{SWA} \leq 1{,}749$	4,093	0.7669	0.9922	1.0000
$1{,}750 \leq C_{SWA} \leq 1{,}799$	3,689	0.7650	0.9938	0.9997
$1{,}800 \leq C_{SWA} \leq 1{,}849$	3,469	0.7530	0.9925	1.0000
$1{,}850 \leq C_{SWA} \leq 1{,}899$	3,252	0.7811	0.9945	0.9997
$1{,}900 \leq C_{SWA} \leq 1{,}949$	3,008	0.7793	0.9947	1.0000
$1{,}950 \leq C_{SWA} \leq 1{,}999$	2,832	0.7970	0.9965	1.0000
$2{,}000 \leq C_{SWA} \leq 2{,}049$	2,573	0.8022	0.9965	1.0000
$2{,}050 \leq C_{SWA} \leq 2{,}099$	2,356	0.7975	0.9966	1.0000
$2{,}100 \leq C_{SWA} \leq 2{,}149$	2,307	0.8331	0.9957	1.0000
$2{,}150 \leq C_{SWA} \leq 2{,}199$	2,033	0.8170	0.9975	1.0000
$2{,}200 \leq C_{SWA} \leq 2{,}249$	1,999	0.8354	0.9990	1.0000
$2{,}250 \leq C_{SWA} \leq 2{,}299$	1,892	0.8494	0.9984	1.0000

Table 3a: For Each Run, the Stratum and Stratum Proportion When 0.9500 First Exceeded

(Proportion computations result from 2020 Census Redistricting Data Production Settings ($\epsilon = 17.14$ for persons) version of *TDA*.)

Population: United States (50 States & DC)

		Criterion III LDG $DR_g \leq 0.05$
TDA Run	Stratum for Block Groups	Proportion When 0.9500 First Exceeded
1	$450 \leq C_{SWA} \leq 499$	0.9716
2	$400 \leq C_{SWA} \leq 449$	0.9596
3	$450 \leq C_{SWA} \leq 499$	0.9661
4	$400 \leq C_{SWA} \leq 449$	0.9543
5	$400 \leq C_{SWA} \leq 449$	0.9561
6	$400 \leq C_{SWA} \leq 449$	0.9508
7	$350 \leq C_{SWA} \leq 399$	0.9509
8	$450 \leq C_{SWA} \leq 499$	0.9541
9	$450 \leq C_{SWA} \leq 499$	0.9617
10	$450 \leq C_{SWA} \leq 499$	0.9661
11	$450 \leq C_{SWA} \leq 499$	0.9596
12	$450 \leq C_{SWA} \leq 499$	0.9683
13	$400 \leq C_{SWA} \leq 449$	0.9525
14	$400 \leq C_{SWA} \leq 449$	0.9543
15	$350 \leq C_{SWA} \leq 399$	0.9558
16	$450 \leq C_{SWA} \leq 499$	0.9650
17	$450 \leq C_{SWA} \leq 499$	0.9607
18	$400 \leq C_{SWA} \leq 449$	0.9596
19	$450 \leq C_{SWA} \leq 499$	0.9727
20	$350 \leq C_{SWA} \leq 399$	0.9582
21	$450 \leq C_{SWA} \leq 499$	0.9617
22	$450 \leq C_{SWA} \leq 499$	0.9683
23	$350 \leq C_{SWA} \leq 399$	0.9558
24	$450 \leq C_{SWA} \leq 499$	0.9628
25	$450 \leq C_{SWA} \leq 499$	0.9519

Table 4: Proportion of Places and MCDs in Each Stratum for Three Criteria

(Proportion computations result from 2020 Census Redistricting Data Production Settings ($\epsilon = 17.14$ for persons) version of *TDA*.)

Population: United States (50 States & DC)

Stratum for Places and MCDs Using C_{SWA} for TOTAL	Number of Places and MCDs	Reliable Characteristics Criteria		
		Criterion I	Criterion II	Criterion III
		LDG $DR_g \leq 0.01$	LDG $DR_g \leq 0.03$	LDG $DR_g \leq 0.05$
$50 \leq C_{SWA} \leq 99$	573	0.2182	0.5969	0.7923
$100 \leq C_{SWA} \leq 149$	622	0.4051	0.7540	0.9116
$150 \leq C_{SWA} \leq 199$	645	0.3442	0.8109	0.9473
$200 \leq C_{SWA} \leq 249$	598	0.4197	0.8361	0.9632
$250 \leq C_{SWA} \leq 299$	500	0.4860	0.9000	0.9760
$300 \leq C_{SWA} \leq 349$	448	0.5379	0.9152	0.9844
$350 \leq C_{SWA} \leq 399$	417	0.5731	0.9233	0.9808
$400 \leq C_{SWA} \leq 449$	399	0.6416	0.9449	0.9975
$450 \leq C_{SWA} \leq 499$	344	0.6424	0.9680	0.9913
$500 \leq C_{SWA} \leq 549$	341	0.6716	0.9765	0.9971
$550 \leq C_{SWA} \leq 599$	291	0.7113	0.9691	0.9966

Table 4
(continued)

$850 \leq C_{SWA} \leq 899$	222	0.7162	0.9955	1.0000
$900 \leq C_{SWA} \leq 949$	201	0.7562	0.9851	1.0000
$950 \leq C_{SWA} \leq 999$	210	0.7571	0.9952	1.0000
$1{,}000 \leq C_{SWA} \leq 1{,}049$	223	0.7982	0.9955	1.0000
$1{,}050 \leq C_{SWA} \leq 1{,}099$	157	0.8153	0.9873	1.0000
$1{,}100 \leq C_{SWA} \leq 1{,}149$	194	0.7423	0.9897	0.9948
$1{,}150 \leq C_{SWA} \leq 1{,}199$	178	0.8596	1.0000	1.0000
$1{,}200 \leq C_{SWA} \leq 1{,}249$	162	0.8395	1.0000	1.0000
$1{,}250 \leq C_{SWA} \leq 1{,}299$	174	0.8563	0.9885	1.0000
$1{,}300 \leq C_{SWA} \leq 1{,}349$	164	0.8659	0.9939	1.0000
$1{,}350 \leq C_{SWA} \leq 1{,}399$	166	0.8614	1.0000	1.0000
$1{,}400 \leq C_{SWA} \leq 1{,}449$	134	0.9030	0.9851	0.9925
$1{,}450 \leq C_{SWA} \leq 1{,}499$	153	0.8562	0.9935	1.0000
$1{,}500 \leq C_{SWA} \leq 1{,}549$	147	0.9320	1.0000	1.0000

Table 4a: For Each Run, the Stratum and Stratum Proportion When 0.9500 First Exceeded

(Proportion computations result from 2020 Census Redistricting Data Production Settings ($\epsilon = 17.14$ for persons) version of *TDA*.)

Population: United States (50 States & DC)

		Criterion III LDG $DR_9 \leq 0.05$
TDA Run	Stratum for Places & MCDs	Proportion When 0.9500 First Exceeded
1	$150 \leq C_{SWA} \leq 199$	0.9504
2	$200 \leq C_{SWA} \leq 249$	0.9548
3	$150 \leq C_{SWA} \leq 199$	0.9566
4	$150 \leq C_{SWA} \leq 199$	0.9504
5	$200 \leq C_{SWA} \leq 249$	0.9632
6	$200 \leq C_{SWA} \leq 249$	0.9632
7	$200 \leq C_{SWA} \leq 249$	0.9615
8	$200 \leq C_{SWA} \leq 249$	0.9615
9	$200 \leq C_{SWA} \leq 249$	0.9582
10	$150 \leq C_{SWA} \leq 199$	0.9643
11	$150 \leq C_{SWA} \leq 199$	0.9566
12	$150 \leq C_{SWA} \leq 199$	0.9504
13	$200 \leq C_{SWA} \leq 249$	0.9615
14	$150 \leq C_{SWA} \leq 199$	0.9550
15	$200 \leq C_{SWA} \leq 249$	0.9565
16	$150 \leq C_{SWA} \leq 199$	0.9550
17	$200 \leq C_{SWA} \leq 249$	0.9515
18	$150 \leq C_{SWA} \leq 199$	0.9519
19	$150 \leq C_{SWA} \leq 199$	0.9504
20	$200 \leq C_{SWA} \leq 249$	0.9532
21	$200 \leq C_{SWA} \leq 249$	0.9615
22	$200 \leq C_{SWA} \leq 249$	0.9548
23	$150 \leq C_{SWA} \leq 199$	0.9566
24	$150 \leq C_{SWA} \leq 199$	0.9550
25	$150 \leq C_{SWA} \leq 199$	0.9519

Analysis of Congressional and State Legislative Districts

Another type of defined geography that is not a part of this nesting includes Congressional districts and state legislative districts. As we will see with Rhode Island in Part II of this study report, each state has Congressional district(s) (CD), state legislative districts in an upper chamber (SLDU), and state legislative districts in a lower chamber (SLDL).

As with the summary display in Table 3a for block groups and the summary display in Table 4a for places and MCDs, we use results from the 25 runs for all "Congressional and state legislative districts". Altogether, we make use of all 7,167 (= 436 + 1,946 + 4,785) Congressional and state legislative districts in the United States. The Table below gives a few parameters for the national accounting of these districts.

	CD	SLDU	SLDL
Number of Districts	436	1,946	4,785
Min Population	526,283	13,629	3,173
Median Population	705,831	121,212	41,713
Mean Population	708,132	158,656	64,016
Max Population	989,415	940,612	470,325

Table 5: For Each Run, the Stratum and Stratum Proportion When 0.9500 First Exceeded

(Proportion computations result from 2020 Census Redistricting Data Production Settings ($\epsilon = 17.14$ for persons) version of *TDA*.)

Population: United States (50 States & DC)

		Criterion III LDG $DR_g \leq 0.05$
TDA Run	Stratum for Congressional & State Legislative Districts	Proportion When 0.9500 First Exceeded
1	$3,150 \leq C_{SWA} \leq 3,199$	1.0000
2	$3,150 \leq C_{SWA} \leq 3,199$	1.0000
3	$3,150 \leq C_{SWA} \leq 3,199$	1.0000
4	$3,150 \leq C_{SWA} \leq 3,199$	1.0000
5	$3,150 \leq C_{SWA} \leq 3,199$	1.0000
6	$3,150 \leq C_{SWA} \leq 3,199$	1.0000
7	$3,150 \leq C_{SWA} \leq 3,199$	1.0000
8	$3,150 \leq C_{SWA} \leq 3,199$	1.0000
9	$3,150 \leq C_{SWA} \leq 3,199$	1.0000
10	$3,150 \leq C_{SWA} \leq 3,199$	1.0000
11	$3,150 \leq C_{SWA} \leq 3,199$	1.0000
12	$3,150 \leq C_{SWA} \leq 3,199$	1.0000
13	$3,150 \leq C_{SWA} \leq 3,199$	1.0000
14	$3,150 \leq C_{SWA} \leq 3,199$	1.0000
15	$3,150 \leq C_{SWA} \leq 3,199$	1.0000
16	$3,150 \leq C_{SWA} \leq 3,199$	1.0000
17	$3,150 \leq C_{SWA} \leq 3,199$	1.0000
18	$3,150 \leq C_{SWA} \leq 3,199$	1.0000
19	$3,150 \leq C_{SWA} \leq 3,199$	1.0000
20	$3,150 \leq C_{SWA} \leq 3,199$	1.0000
21	$3,150 \leq C_{SWA} \leq 3,199$	1.0000
22	$3,150 \leq C_{SWA} \leq 3,199$	1.0000
23	$3,150 \leq C_{SWA} \leq 3,199$	1.0000
24	$3,150 \leq C_{SWA} \leq 3,199$	1.0000
25	$3,150 \leq C_{SWA} \leq 3,199$	1.0000

Wright and Irimata (2021)
Part 2: Variability across DAS runs

Part 2 of their analysis examines 25 independent runs of 2010 Census data through the 2020 Census DAS using the same production settings to assess variability across runs.

The resulting data were then assessed for their variability against a range of redistricting use cases:

- Rhode Island's Congressional Districts (2)
- Rhode Island's State Upper Legislative Districts (38)
- Rhode Island's State Lower Legislative Districts (75)
- DOJ-supplied jurisdictions
    - Panola County, Mississippi (2,180 blocks)
    - Tate County (School District), Mississippi (784 blocks)
    - Tylertown (Walthall County), Mississippi (136 blocks)

Wright and Irimata (2021)

Part 2: Variability across DAS runs

Definitions of Redistricting Measures of Variation. The measures defined here are all for a specific ϵ. Henceforth, and to simplify notation, we use S for SWA and T for TDA. Let

$$\begin{aligned} G &\equiv \text{the number of demographic groups;} \\ C_S(g) &\equiv \text{the population of group } g \text{ (2010 Census, SF1), for } g = 1, ..., G; \text{ and} \\ C_{Ti}(g) &\equiv \text{the population of group } g \text{ resulting from the } i^{th} \text{ } TDA \text{ run, for } i = 1, ..., 25. \end{aligned}$$

We have the following measures including two types of variation among the 25 TDA runs within group g: one relative to $\bar{C}_T(g)$ (see below) and another relative to $C_S(g)$.

(i) The **average population of group** g over the 25 TDA runs is

$$\bar{C}_T(g) \equiv \frac{C_{T1}(g) + C_{T2}(g) + \cdots + C_{T,25}(g)}{25}.$$

(ii) The **variation(1) among the population of group** g over the 25 TDA runs is

$$V(1)_g \equiv \frac{[C_{T1}(g) - \bar{C}_T(g)]^2 + [C_{T2}(g) - \bar{C}_T(g)]^2 + \cdots + [C_{T,25}(g) - \bar{C}_T(g)]^2}{25}.$$

(iii) The **relative variation(1) among the population of group** g over the 25 TDA runs is

$$RV(1)_g \equiv \frac{\sqrt{V(1)_g}}{\bar{C}_T(g)}.$$

(iv) The **average relative variation(1) among the population over the** G **groups** (essentially a coefficient of variation) is

$$AVERV(1) \equiv \frac{RV(1)_1 + RV(1)_2 + \cdots + RV(1)_G}{G}.$$

Wright and Irimata (2021)

Part 2: Variability across DAS runs

Figure 1

Jurisdiction	District	IDEAL POPULATION	*AVERV*(1)
Rhode Island	CD-01	526,283.50	0.006
Rhode Island	CD-02	526,283.50	0.008
Rhode Island	SLDU-01	27,699.10	0.036
Rhode Island	SLDU-02	27,699.10	0.051
Rhode Island	SLDU-03	27,699.10	0.045
Rhode Island	SLDU-04	27,699.10	0.048
Rhode Island	SLDL-01	14,034.20	0.067
Rhode Island	SLDL-02	14,034.20	0.069
Rhode Island	SLDL-03	14,034.20	0.057
Rhode Island	SLDL-04	14,034.20	0.104
Panola County, MS	D-01	6,941.40	0.362
Panola County, MS	D-02	6,941.40	0.353
Panola County, MS	D-03	6,941.40	0.344
Panola County, MS	D-04	6,941.40	0.369
Panola County, MS	D-05	6,941.40	0.280
Tate County Schools, MS	D-01	3,764.60	0.335
Tate County Schools, MS	D-02	3,764.60	0.355
Tate County Schools, MS	D-03	3,764.60	0.493
Tate County Schools, MS	D-04	3,764.60	0.449
Tate County Schools, MS	D-05	3,764.60	0.376
Tylertown, MS	D-01	402.25	0.748
Tylertown, MS	D-02	402.25	0.622
Tylertown, MS	D-03	402.25	0.596
Tylertown, MS	D-04	402.25	1.265

Wright and Irimata (2021)

Conclusions

The Key Empirical Message on Reliability:

"for any block group with a TOTAL count between 450 and 499 people or larger, and for MCDs and places between 200 and 249 or larger, the difference between the TDA ratio of the largest demographic group (LDG) and the corresponding SWA ratio for the LDG is less than or equal to 5 percentage points at least 95% of the time". No Congressional or state legislative district fails this test; that is, for these districts, the 5-percentage point criterion holds 100% of the time."

The Key Empirical Message on Variability:

"relative variability in the TDA [decreases] as we consider [larger] pieces of geography and population…At a high level, [the analysis] tends to show less relative variability using the 2020 Census redistricting data production settings version of the TDA than the 2021-04-28 version

Analysis of the impact of DAS production settings on the identification of majority-minority districts

Districts Examined:

Congressional Districts	436
State Upper Legislative Districts (SLDU)	1,946
State Lower Legislative Districts (SLDL)	4,785

Demographics Examined:

P1/P8 Tables (Total Population by Race):

- White Alone
- Black Alone
- AIAN Alone
- Asian Alone
- SOR Alone
- Black and Black+White

P2/P9 Tables (Total Hispanic/NH by Race):

- Hispanic
- NH White Alone
- NH Black Alone
- NH AIAN Alone
- NH Asian Alone
- NH SOR Alone
- NH Black and Black+White

P3/P10 Tables (VAP by Race):

- White Alone VAP
- Black Alone VAP
- AIAN Alone VAP
- Asian Alone VAP
- SOR Alone VAP
- Black and Black+White VAP

P4/P11 Tables (Hispanic/NH VAP by Race):

- Hispanic VAP
- NH White Alone VAP
- NH Black Alone VAP
- NH AIAN Alone VAP
- NH Asian Alone VAP
- NH SOR Alone VAP
- NH Black and Black+White VAP

NH = Not Hispanic

VAP = Voting Age Population

P1/P8 Tables (Total Population by Race)

White Alone

State	District	District Type	SF1 Total	DAS Total	SF1 White	DAS White	SF1 %White	DAS %White
27	062	Upper	79,225	79,289	39,665	39,639	50.01%	49.99%
6	46	Cong.	702,906	702,803	351,403	351,441	49.99%	50.01%

Black Alone

State	District	District Type	SF1 Total	DAS Total	SF1 Black	DAS Black	SF1 %Black	DAS %Black
1	085	Lower	45,229	45,206	22,651	22,581	50.08%	49.95%

P2/P9 Tables (Total Hispanic)

Total Hispanic:

State	District	District Type	SF1 Total	DAS Total	SF1 Hispanic	DAS Hispanic	SF1 %Hispanic	DAS %Hispanic
32	014	Lower	64,054	63,906	31,975	31,964	49.92%	50.02%

P2/P9 Tables (Total NH by Race)

NH White Alone:

State	District	District Type	SF1 Total	DAS Total	SF1 NH White	DAS NH White	SF1 %NH White	DAS %NH White
51	050	Lower	80,677	80,712	40,353	40,353	50.02%	50.00%*
32	009	Upper	128,882	128,641	64,380	64,343	49.95%	50.02%

NH Black Alone:

State	District	District Type	SF1 Total	DAS Total	SF1 NH Black	DAS NH Black	SF1 %NH Black	DAS %NH Black
37	043	Lower	79,233	79,025	39,545	39,551	49.91%	50.05%

NH AIAN Alone:

State	District	District Type	SF1 Total	DAS Total	SF1 NH AIAN	DAS NH AIAN	SF1 %NH AIAN	DAS %NH AIAN
30	015	Lower	9,595	9,610	4,807	4,744	50.10%	49.37%

P3/P10 Tables (Voting Age by Race)

VAP White Alone:

State	District	District Type	SF1 Total	DAS Total	SF1 VAP White	DAS VAP White	SF1 %VAP White	DAS %VAP White
27	40A	Lower	27,864	27,904	13,934	13,938	50.01%	49.95%

VAP Black Alone:

State	District	District Type	SF1 Total	DAS Total	SF1 VAP Black	DAS VAP Black	SF1 %VAP Black	DAS %VAP Black
26	005	Upper	196,028	196,209	98,101	98,071	50.04%	49.98%
37	058	Lower	61,968	61,983	30,954	31,003	49.95%	50.02%

VAP Black and Black+White:

State	District	District Type	SF1 Total	DAS Total	SF1 VAP Black	DAS VAP Black	SF1 %VAP Black	DAS %VAP Black
37	057	Lower	59,215	59,273	29,589	29,660	49.97%	50.04%
37	12	Cong.	544,436	544,179	272,110	272,256	49.98%	50.03%

P4/P11 Tables (Hispanic VAP)

Hispanic VAP:

State	District	District Type	SF1 Total	DAS Total	SF1 VAP Hisp.	DAS VAP Hisp.	SF1 %VAP Hisp.	DAS %VAP Hisp.
12	014	Upper	342,997	342,819	171,560	171,389	50.02%	49.99%
12	087	Lower	115,237	115,148	57,642	57,538	50.02%	49.97%

P4/P11 Tables (Not Hispanic VAP by Race)

VAP NH White:

State	District	District Type	SF1 Total	DAS Total	SF1 VAP NH White	DAS VAP NH White	SF1 %VAP NH White	DAS %VAP NH White
27	062	Upper	60,648	60,720	30,384	30,359	50.10%	50.00%*

VAP NH Black Alone:

State	District	District Type	SF1 Total	DAS Total	SF1 VAP NH Black	DAS VAP NH Black	SF1 %VAP NH Black	DAS %VAP NH Black
13	177	Lower	41,506	41,461	20,719	20,731	49.92%	50.00%
18	094	Lower	45,634	45,508	22,786	22,837	49.93%	50.18%
26	001	Upper	197,305	196,969	98,476	98,500	49.91%	50.01%
42	181	Lower	45,240	45,073	22,614	22,644	49.99%	50.24%

P4/P11 Tables (Not Hispanic VAP by Race)

VAP NH Black and Black+White:

State	District	District Type	SF1 Total	DAS Total	SF1 VAP NH Black	DAS VAP NH Black	SF1 %VAP NH Black	DAS %VAP NH Black
9	015	Lower	18,820	18,863	9,422	9,423	50.06%	49.95%
45	023	Lower	28,442	28,433	14,230	14,202	50.03%	49.95%
24	37A	Lower	30,391	30,391	15,118	15,199	49.74%	50.01%
36	05	Cong.	545,319	545,308	272,480	272,949	49.97%	50.05%
37	043	Lower	59,130	59,881	29,480	29,446	49.86%	50.01%

Conclusions

Comparison of the Production Settings Demonstration Data to the published 2010 Census tabulations identified 25 districts out of 7,167 (0.3%) where a demographic group *could* be considered to flip from majority to minority or vice-versa.

Flips occurred in both directions (11 groups went from majority to minority, 14 went from minority to majority).

No flips involved both a racial or ethnic group's total population and voting age population. (That is, districts drawn to have majorities of both total population and voting age population are more stable.)

All flips involved very small numbers of individuals in districts that were tightly drawn (usually within a few hundredths of a percent of the 50% mark) using the published 2010 Census tabulations (a level of precision that would be greatly impacted by the noise injected by the 2010 Census swapping algorithms).

Stay Informed:

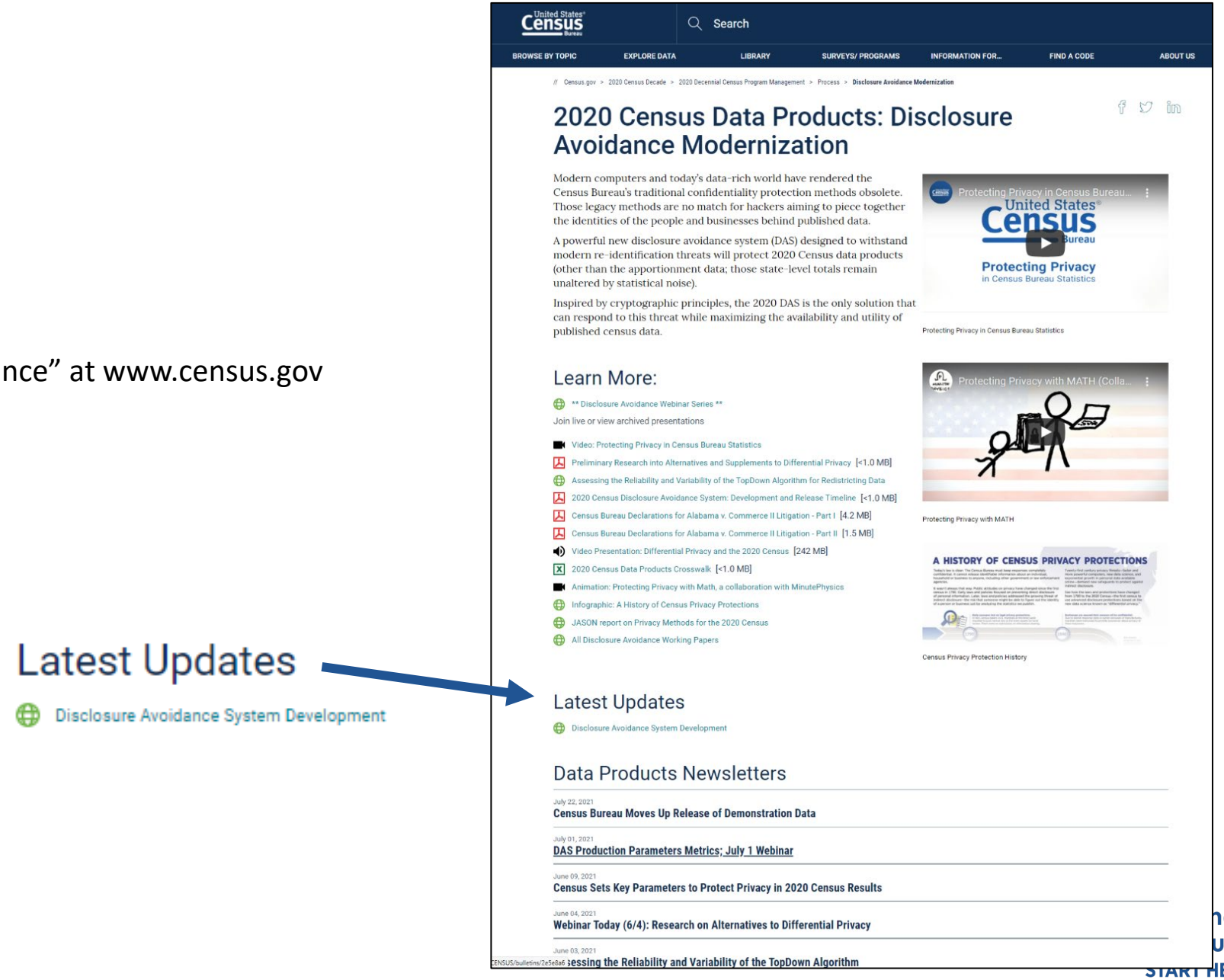
Subscribe to the 2020 Census Data Products Newsletters

*Search “Disclosure Avoidance” at www.census.gov

Stay Informed:

Visit Our Website

*Search “Disclosure Avoidance” at www.census.gov



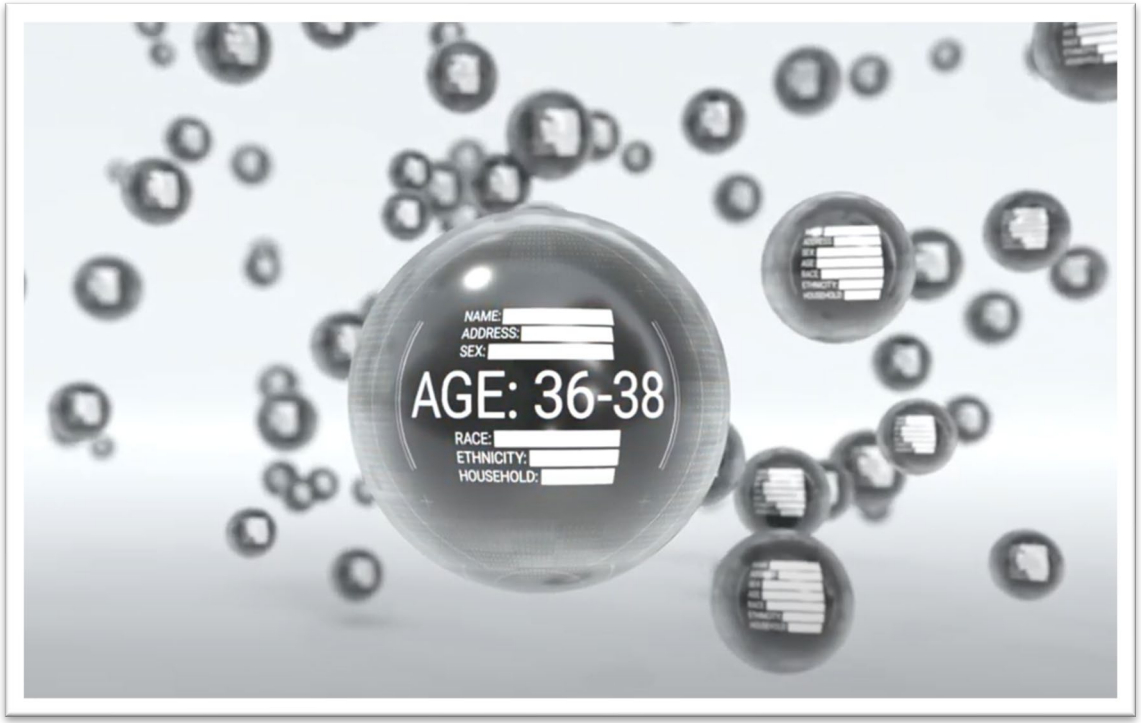
Latest Updates
Disclosure Avoidance System Development

** New Video **

Protecting Privacy in Census Bureau Statistics

*Find it on our website and YouTube Page

Search “Disclosure Avoidance” at www.census.gov



Questions?