



Option B: The Census (for experienced census users)

Presentation to the National Conference of State Legislatures Redistricting Seminar
10/26/2019

James Whitehorne
Chief – Census Redistricting & Voting Rights Data Office

Shape
your future
START HERE >

United States[®]
Census
2020

Agenda

- **2020 Residence Rules & Criteria**
- **Voting Rights Data**
 - Citizenship Voting Age Population Data
 - Section 203 Language Determinations
- **Disclosure Avoidance**
 - Differential Privacy
 - 2010 Demonstration Products
 - 2010 P.L. 94-171 Redistricting Data Demonstration Product
 - 2010 Demographic and Characteristic Demonstration Product

2020 Residence Criteria and Residence Situations

Juveniles in non-correctional treatment centers

- At their usual home address (previously counted at facility)
- At the facility if no usual home address

Religious Group Quarters

- At the facility (previously counted at home address)

College Students

- Where they live and sleep most of the time, as before

U.S. flagged maritime or merchant vessels between U.S. and foreign ports

- At usual home address (previously not counted)
- At port if no usual home address (previously not counted)

Federally Affiliated Count Overseas

- Military Deployed/Stationed

Prisoners at Correctional Facilities

- Counted at facility (as before)
- Addition of Group Quarters to P.L. 94-171 Data
- Geocoding Tool – webpage coming in November of 2019

Voting Rights Tabulations

Section 203 Language Determinations

- New determinations conducted every 5 years
- Identifies counties or townships where language assistance is needed for voting
- Calculated using the American Community Survey 5-year estimates and other Census data
- Last published in the Federal Register on 12/5/2016, next publication estimated 12/2021

Citizen Voting Age Population by Race and Ethnicity

- Initially created after Census 2000 (Census Long Form)
- Annual Tabulation (since 2011)
- Calculated using the American Community Survey 5-year estimates
- Added Congressional and State Legislative Districts for 2018 and future releases
- Typically released in the 1st week of February each year

Voting Rights Tabulations

Citizen Voting Age Population by Race and Ethnicity Data

- As stated by the Census Bureau's Chief Scientists at the Council of Professional Associations on Federal Statistics (COPAFS), "The Paperwork Reduction Act clearance package for the 2020 Census and the President's Executive Order 13880 commit the Census Bureau to releasing Citizen Voting-Age Population (CVAP) data by March 31, 2021."
- These data will be produced by combining administrative data from a number of federal, and possibly state, agencies into a separate micro-data file that will contain a "best citizenship" variable for every person in the 2020 Census.
- The citizenship file and the Census Edited File (CEF) will be simultaneously sent through the 2020 Disclosure Avoidance System, which will do the final record linkage and generate a confidentiality protected citizenship variable consistent with the redistricting data.
- CVAP data will be produced at the block-level and released to the public by March 31, 2021.

Voting Rights Tabulations

Citizen Voting Age Population by Race and Ethnicity Format

- No final decisions have been made regarding the methodology and format of the block-level CVAP data.
- No decisions have been made regarding the future of the American Community Survey-based CVAP data that have been produced annually since 2011.
- The Census Bureau's internal working group has set March 31, 2020 as the final date for determining the viability of each potential administrative data source on citizenship.

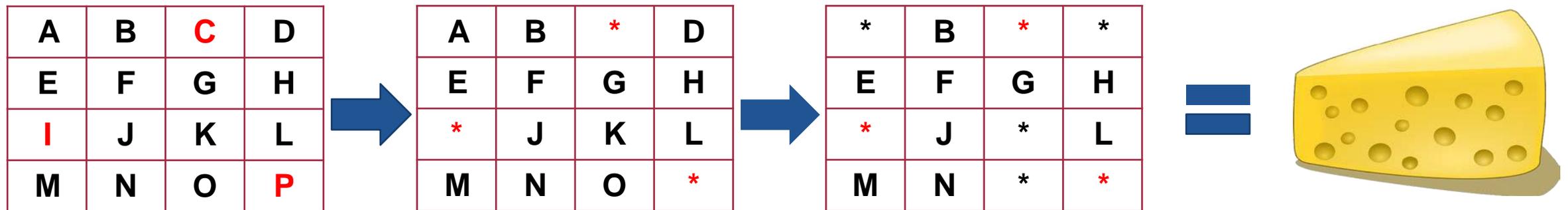
Disclosure Avoidance

- Disclosure avoidance is the process of making changes to data for the purpose of protecting a respondent's privacy and confidentiality in published statistics.
- An improper disclosure occurs when someone can use publicly available statistical information to correctly identify an individual or business who provided information. Using disclosure avoidance, the Census Bureau modifies or removes the characteristics that put confidential information at risk of disclosure.
- Although it may appear that a published table shows information about a specific individual, the Census Bureau has taken steps (such as “data swapping,” “noise injection,” and other statistical methods) to disguise the original data while making sure the results are useful.

Disclosure Avoidance

Suppression

Removing sensitive values from the data.



Disclosure Avoidance

Coarsening

Reducing the amount of detail in the data.

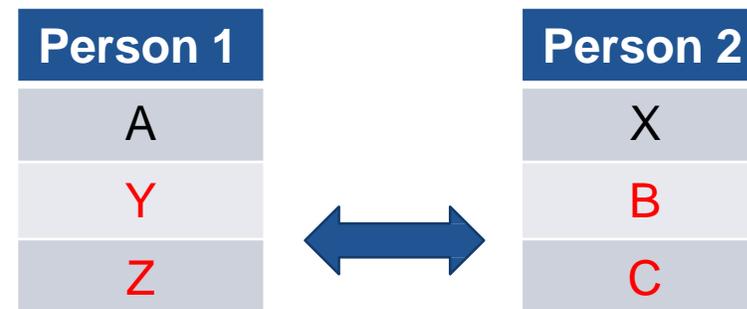
- Geographic aggregation
- Collapsing categories
- Rounding
- Reporting in ranges, etc.



Disclosure Avoidance

Data Swapping

- Data swapping is a disclosure avoidance method that “swaps” data between households in different locations that have similar characteristics on a set of variables.
- Which households were swapped is not public information. The selection process is highly targeted so it is applied to the data with the highest disclosure risk.
- Often, swapping occurs within a specific geographic area so there is no effect on the population or characteristics totals for that geographic area.
- Because of data swapping, users should expect that tables with cells having a value of one or two do not reveal information about specific individuals. As a consequence, these cells typically do not have a high degree of accuracy.



Disclosure Avoidance

https://www.census.gov/about/policies/privacy/statistical_safeguards/disclosure-avoidance-2020-census.html

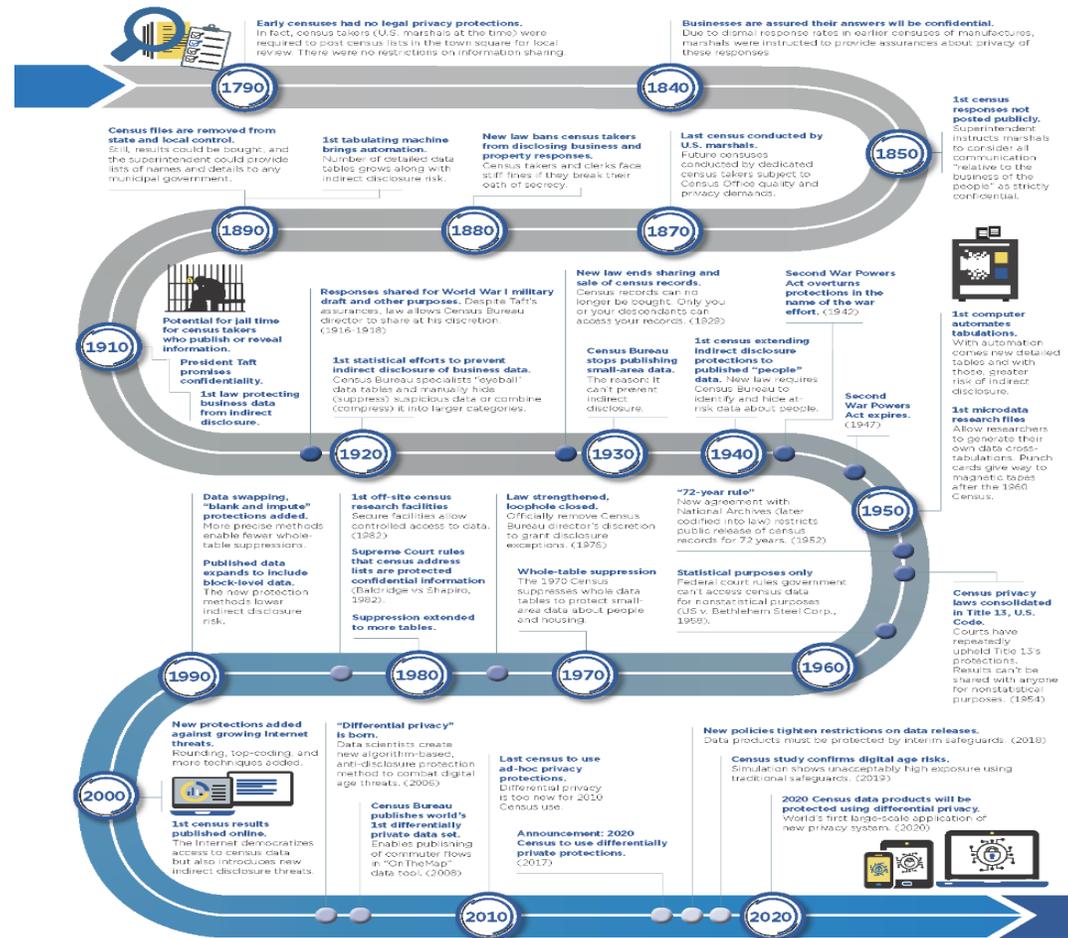
A HISTORY OF CENSUS PRIVACY PROTECTIONS

Today's law is clear: The Census Bureau must keep responses completely confidential. It cannot release identifiable information about an individual, household or business to anyone, including other government or law enforcement agencies.

It wasn't always that way. Public attitudes on privacy have changed since the first census in 1790. Early laws and policies focused on preventing direct disclosure of personal information. Later, laws and policies addressed the growing threat of indirect disclosure—the risk that someone might be able to figure out the identity of a person or business just by analyzing the statistics we publish.

Twenty-first century privacy threats—faster and more powerful computers, new data science, and exponential growth in personal data available online—demand new safeguards to protect against indirect disclosure.

See how the laws and protections have changed from 1790 to the 2020 Census—the first census to use advanced disclosure protections based on the new data science known as “differential privacy.”



Differential Privacy

- A new disclosure avoidance system (DAS) is needed to defend against new threats posed by today's technology: growing computing power, advances in mathematics, and easy access to large, public databases. Combined, these changes could allow highly sophisticated users to identify common data points between our published statistics, or between our statistics and outside databases. They could use these common threads to potentially identify the people or businesses behind the statistics.
- Our research shows that the risk of successful re-identifications is unacceptably large. We are committed to applying better and stronger protections with each advance in data science.
- To protect this data for now and into the future, the Census Bureau has committed to modernizing its approach to privacy protections by using Differential Privacy, also known as Formal Privacy.
- The specific method developed by the Census Bureau is being called the Top Down Algorithm since it starts at the national level. This was found to produce a better result than a bottom up approach.

Differential Privacy

- Differential privacy is the scientific term for a method that adds “statistical noise” to the tables we publish in a way that protects each respondent’s identity. Differential privacy uses statistical noise to slightly alter data so that the link between the data and a specific person or business can’t be certain.
- The idea of using statistical noise to protect privacy is not new: the Census Bureau has used similar techniques for decades.
- With differential privacy, the Census Bureau precisely controls the amount of statistical noise added using sophisticated mathematical formulas that allows us to assure that enough noise is added to protect privacy but not so much as to damage the statistical validity of our publications. The formulas allow us to balance between two opposing extremes: total accuracy and totally privacy.
- By documenting the properties of this noise, we can help data users determine if published estimates are suitable, or accurate enough, for their specific applications. We call this assuring “**fitness for use.**”

2010 Demonstration Products

What: The 2010 Census Demonstration Data Products apply the Top Down Algorithm of the new Census Disclosure Avoidance System (DAS) to the 2010 Census Confidential Data, that is the unprotected data from the 2010 Census that is not available publicly.

Why: The 2010 Census demonstration Data Products are designed to give data users a better understanding of the Census Bureau's 2020 DAS, which uses differential privacy as its core methodology.

The products will be the subject of a special workshop hosted by the Center for National Statistics (CNSTAT) on Dec. 11-12, 2019, where presenters will share their analyses of the products and the potential impact of the DAS on decennial data. Visit CNSTAT's website closer to the date for more information and viewing options.

Where: The data will be made available on the Census Bureau's website with documentation at: <https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/2020-census-data-products/2010-demonstration-data-products.html>

When: We anticipate publishing this data between October 28 and November 1, 2019

2010 Demonstration Products – P.L. 94-171

- **Source Data: 2010 Census Edited File** prior to application of 2010 disclosure avoidance processes
- **Disclosure Avoidance: planned 2020 Top Down Algorithm (Differential Privacy)**
- **Output: 2020 Census P.L. 94-171 Redistricting Data tables** - in the traditional Census Bureau summary file text file format - using the 2020 Census geoheader design
- **Scope: The 50 states, DC, and Puerto Rico**
- **Technical Documentation: 2018 Prototype P.L. 94-171 Redistricting Data Summary File**
- **Organization: Individual state (or state equivalent) .zip files + one National* .zip file.**
 - *the National file represents geographies that cross state boundaries. It is not a collection of merged state redistricting files.

2010 Demonstration Products – Demographic and Housing File

- **Source Data:** 2010 Census Edited File prior to application of 2010 disclosure avoidance processes
- **Disclosure Avoidance:** planned 2020 Top Down Algorithm (Differential Privacy)
- **Output:** a subset of the 2020 Census Demographic Housing and Characteristic Summary File - in the traditional Census Bureau summary file text file format - using the 2010 Census Summary File 1 geoheader design
- **Scope:** The 50 states, DC, and Puerto Rico
- **Technical Documentation:** specific 2010 Demonstration Data Product technical documentation
- **Organization:** Individual state (or state equivalent) .zip files + one National* .zip file.
 - *the National file represents geographies that cross state boundaries. It is not a collection of merged state redistricting files.

2010 Demonstration Products

Background

- The total population and voting age population counts are no longer invariant below the state level of geography. It is understood that there could be, and likely will be, differences between sub-state totals for any given piece of sub-state geography for these variables especially when reviewing the small area geography such as blocks. Our interest is in your use of these data in aggregate for your use cases.
- Since these datasets are produced from the 2010 Census Confidential Data, that is the unprotected data from the 2010 Census that is not available publicly, some differences you see in the data may be caused by the 2010 disclosure avoidance techniques applied to the previously released data.
- These data sets are only being published in the traditional summary file format, a series of text files containing a relational (link) field.

2010 Demonstration Products

Possible questions to ask:

- How are the deviations of my redistricting plan affected? Are they affected in a negative, neutral, or positive manner?
- Does this alter a remedy district in a manner that makes it no longer qualify as a remedy district?
- Does this affect my plan's ability to resist being challenged? If so, how different would my plan need to be to bring it back to a defensible state?
- Does this affect my <insert your analysis description>?
 - Is this easily remedied?
 - Does this require dramatic change to reproduce an outcome similar to what I previously received?

How to submit feedback or findings:

- Email : dcmd.2010.demonstration.data.products@census.gov

Other 2020 Census Updates

- August update to the 2020 Census webpage: <https://2020census.gov>
 - Next 2020 Census website update coming January 2020
- Mobile Questionnaire Assistance Operation
 - In Planning stage
 - Census employees deployed to hard-to-count areas to take responses and answer questions
 - Target resources dynamically towards areas experiencing lower response rates
- Completion of in-field address canvassing operation October 2019
- Final Voting District Verification January 3, 2020 to March 31, 2020
- Final pre-2020 Census Boundary and Annexation Survey (BAS) January 3, 2020 to May 31, 2020
 - Boundaries in effect as of January 1, 2020.

Thank You

James Whitehorne

Chief – Census Redistricting & Voting Rights Data Office

U.S. Department of Commerce
U.S. Census Bureau
4600 Silver Hill Rd.
Suitland, Maryland 20746

Office: 301-763-9051

Email: rdo@census.gov

Website: www.census.gov/rdo