

# Implications to Montana of Differential Privacy for Census Bureau 2020 Data Dissemination

Mary Craigle, RIS Bureau Chief  
MT Department of Commerce  
April 14, 2021

# Presentation Outline

What is Differential Privacy (DP) and the Census Bureau Disclosure Avoidance System (DAS)?

Why the Concerns?

Micro-files and Potential Impacts

Current Timing of Implementation and Actions



The Census Bureau's Disclosure Avoidance System (DAS) could have profound impacts on a decade of data quality, the equitable distribution of resources and the redistricting of/equal representation in state legislatures, county commissions, school districts, and city councils.





# HIGH LEVEL OVERVIEW OF DIFFERENTIAL PRIVACY (DP)



DP is a step  
beyond what  
traditionally has  
been  
Census Bureau  
Disclosure  
Avoidance

- The new disclosure rules were motivated by the threat of “**database reconstruction**” which is the ability to infer individual-level characteristics by relating various data tables. This is not the same as a person’s identity.
- Database reconstruction should not be confused with re-identification. To figure out a specific person’s identity, one would have to match the reconstructed Census microdata to an outside source (like *IBISWorld* or Nielsen) that provides the individual’s information.

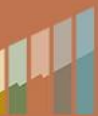


# Census Bureau

## Title 13

requirement to  
keep the  
information  
private – a  
further step

- Since 1962, the Census Bureau has interpreted “any particular establishment or individual” to mean an individual **whose identity can be determined**.
- Now the Census Bureau is saying it cannot release data about individuals, **even if the identity of those individuals is unknown**, because they could be identified using information from another source.
- The new interpretation asserts that it is **prohibited to reveal characteristics of an individual even if the identity of that individual is effectively concealed**.
- This is because the data records could be linked to an outside commercial database to determine personal identifiable information (PII) and then reveal the person’s identity.

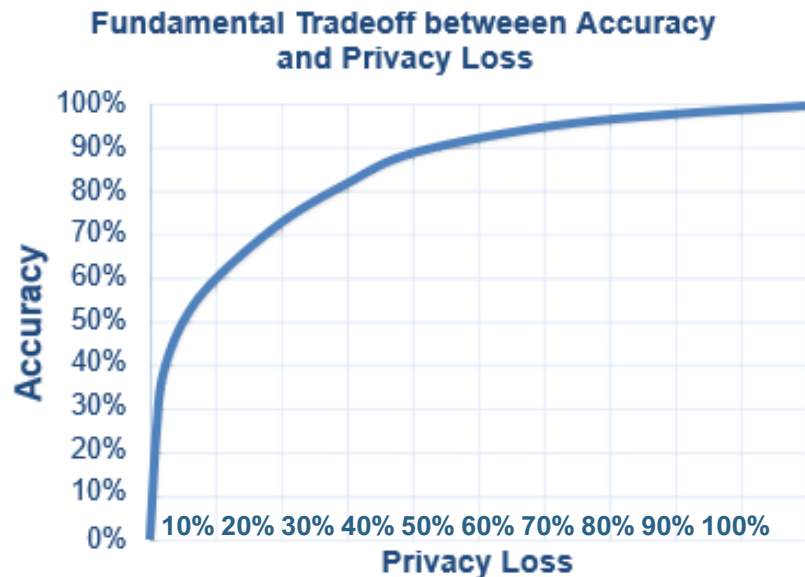


# What is Differential Privacy (DP)

It's called "differential privacy" because it **mathematically models** the privacy "differential" that each person experiences from having their data included in the Census Bureau's data products compared to having their record deleted or replaced with an arbitrary record.



$\epsilon$ -differential privacy is a mathematical definition for the privacy loss associated with any data release drawn from a statistical database.



The chosen blend of accuracy and privacy results in a measure called the “privacy loss budget” or “epsilon” ( $\epsilon$ ). An epsilon of zero results in perfect privacy but useless data, whereas an epsilon of infinity results in perfect accuracy but theoretically imperfect privacy.

The chosen trade-off point along the line must be allocated across the entire 2020 database for the release of any given data point – how much it adds or subtracts from the point.





# The Disclosure Avoidance System (DAS)

The Disclosure Avoidance System (DAS) is the name the Census Bureau has given to the IT system that uses the DP methodology to cloak record characteristics.

DAS is the formal privacy system the Census Bureau will be using on the 2020 Census Data and ALL other data products (ACS, Economic Census, etc.) including those prepared for other agencies (BLS, BEA, etc.)



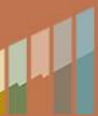
# The Disclosure Avoidance System (DAS) Relies on Injecting Statistical Noise into the Data with Formal Privacy Rules

**Advantages** of noise injection using differential privacy:

- Privacy guarantees are *future-proof*
- Privacy guarantees are ***provable***
- Privacy guarantees are *public and explainable*
- Protects against database reconstruction attacks (tunable)

**Disadvantages**

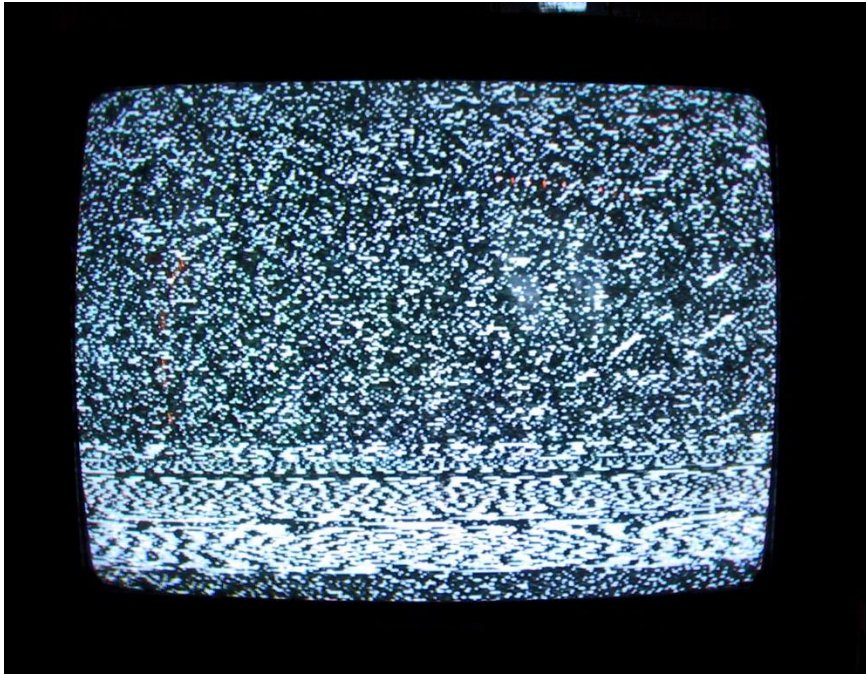
- Entire country must be processed at once for best accuracy
- Every use of the private data must be tallied in the ***privacy-loss budget***
- Loss of information due low accuracy or suppression



# HOW WILL DP IMPACT THE DATA?



# Key Takeaway



Unlike previous Censuses, only housing units are invariant - always exact number no matter the geography you chose.

Population, households, characteristics, etc. all have DAS “noise” added so the data no longer represents a true “census” especially at smaller geographies and populations.

This can create improbable outcomes and problematic data.



## Major Impacts of DP to Montana Data Are in Four Areas:



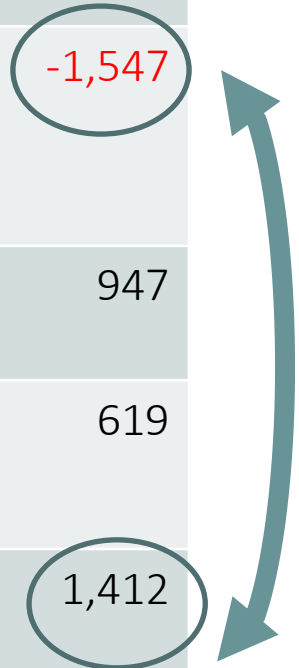
1. For data that is released, the smaller the geographic size, the more distorted the data. Only housing units are unadjusted.
2. Because of the loss of accuracy, much of the data and tables made available in past Censuses will be suppressed because of low accuracy.
3. Suppressed tables or tables not produced include detailed race and Hispanic origin tables, some family/household tables included in Summary File 1 in 2010 and **ALL** of Summary File 2 which had tribal affiliation.
4. The unreliability of the data could impact a variety of users including governments that rely on accurate counts for funding, planning and assessments. Researchers will lose their ability to reliably study the demographic and social characteristics of smaller locales. Businesses and non-profits may no longer have the data validity they require for planning, marketing, and investment purposes in some areas.



# Data Example #1 for Montana

## Distribution of Race Statewide As Reported in the 2010 Census and with DAS Applied

Race/Ethnicity	DAS Added 2010 Census	2010 Census	Difference
White Alone or in Combination	905,578	908,645	-3,067
American Indian/Alaskan Native Alone or in Combination	77,054	78,601	-1,547
Asian Alone or in Combination	11,429	10,482	947
Black/African American Alone or in Combination	8,536	7,917	619
Native Hawaiian /Pacific Islander Alone or in Combination	3,144	1,732	1,412



Analysis from Micro-file Data Set Produced November 2020



# DAS Impact Could Result in a Loss of Funding

According to the report “Counting for Dollars 2020,” Montana received \$5,061,000 for Indian Housing in FY 2016. Another \$974,103 was received for Native American Employment & Training.

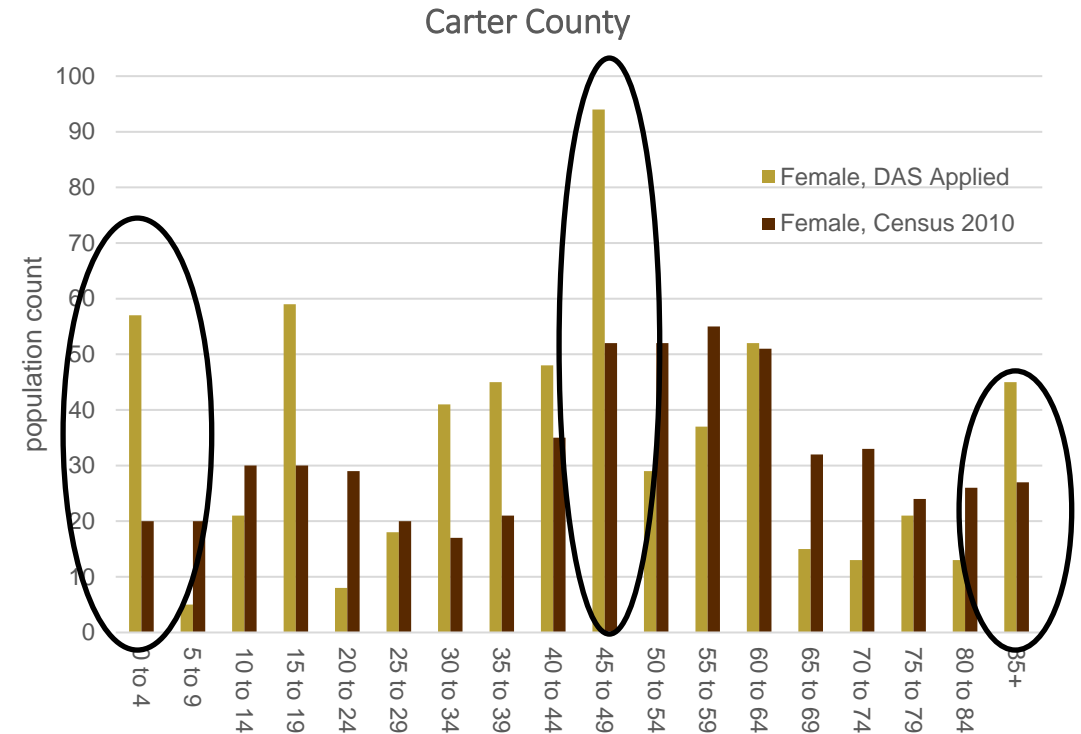
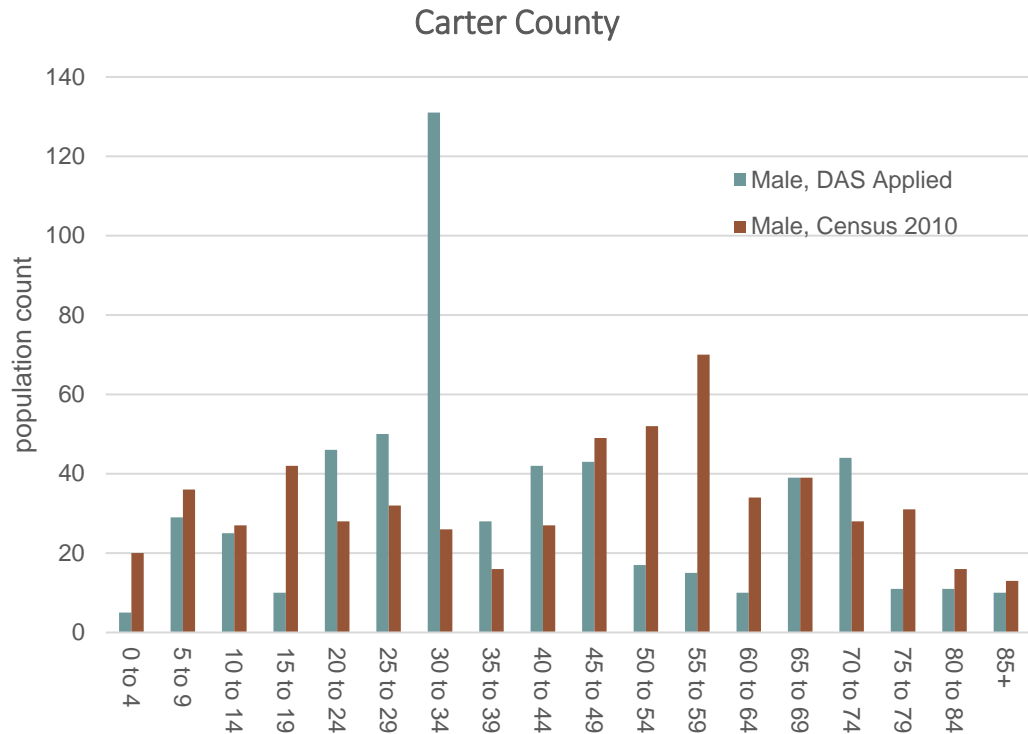
The amount of funding for these programs is based (in part) on the respondents who identified as single race or multi-race American Indian in the Census 2010 count.

[Counting for Dollars 2020: The Role of the Decennial Census in the Geographic Distribution of Federal Funds](#)  
[George Washington Institute of Public Policy](#)



# Data Example #2 for Montana

## Distribution of Age for Carter County by Gender As Reported in the 2010 Census and with DAS Applied



Analysis from Micro-file Data Set Produced July 2020 – Heather Zimmerman, DPHHS

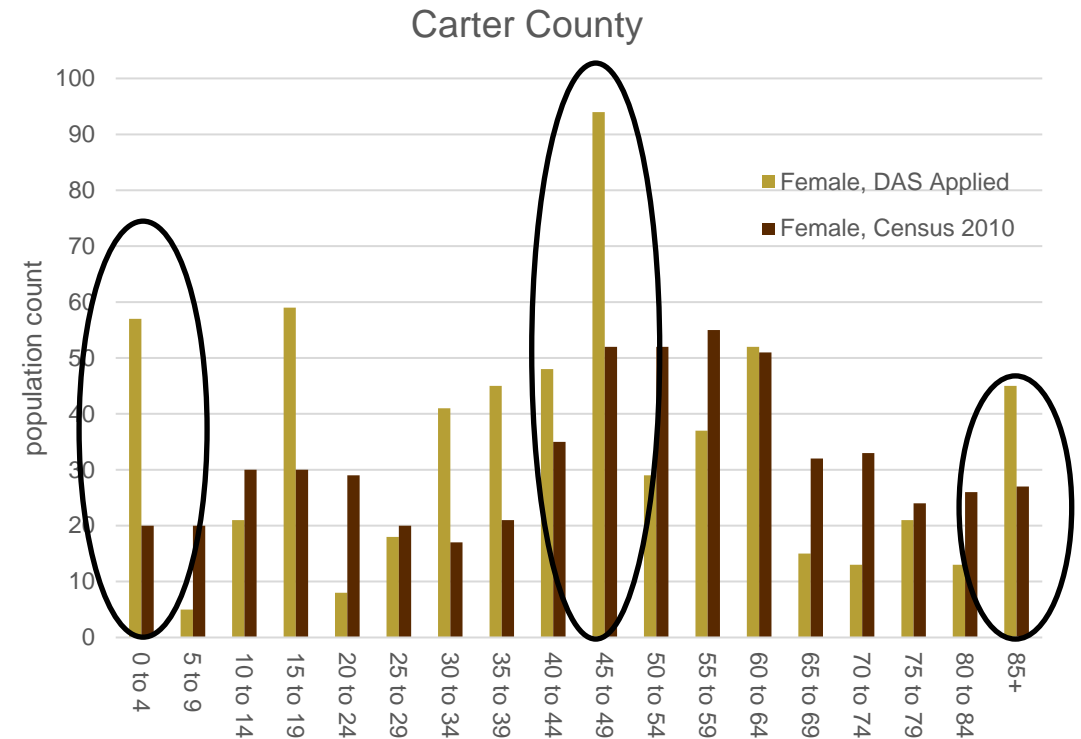




# DAS Impact Could Result in Poor Planning

The example of Carter County shows the potential impact to planning. Focusing on the distribution of females by age, you can see a large increase with DAS in three age groups: 0 to 4, 45 to 49, and 85+

The population overstatement could result in over-allocation of resources and programs specific to these three groups (pre-school, childcare, elder services) and undersupply to other cohorts.



# Impact of Size – Population Differences

## Larger Populations Have Less DP Impact

MT Town or City	DAS Added 2010 Census	2010 Census Total Pop	Difference	% Change
Ismay	40	19	21	111%
Opheim	127	85	42	49%
Billings	104,039	104,170	- 131	- 0.1%
Ennis	752	838	- 86	- 10%
Pony	78	118	- 131	- 34%
Inverness	33	55	- 22	- 40%

Analysis from Micro-file Data Set Produced November 2020

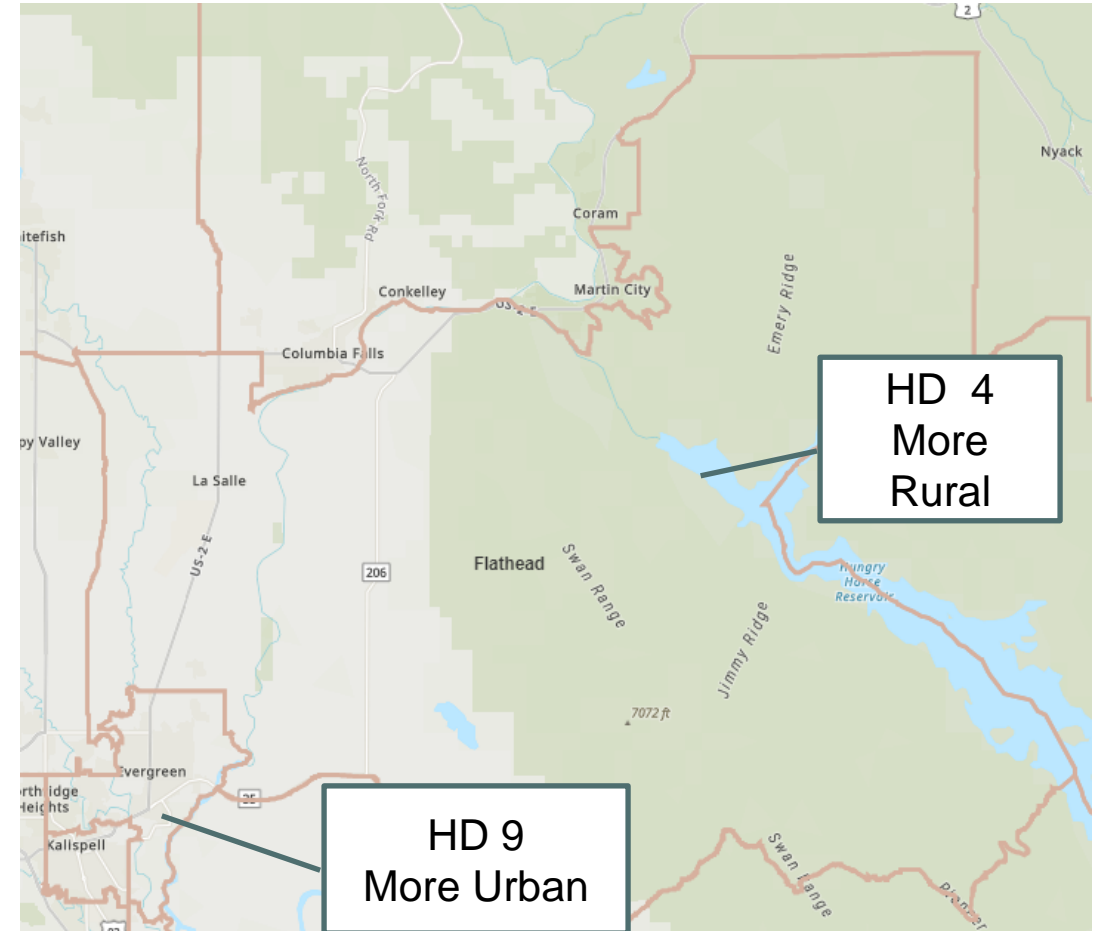


# Data Example #4 for Montana

Population Associated with Blocks Reported in the 2010 Census and with DAS Applied for Creating Boundaries - Shift from Urban to Rural (along with new definitions)

Legislative District	DAS Added 2010 Census	2010 Census	Difference
HD 4	10,018	9,915	+ 103
HD 9	9,945	9,999	- 54

 State House District Boundary

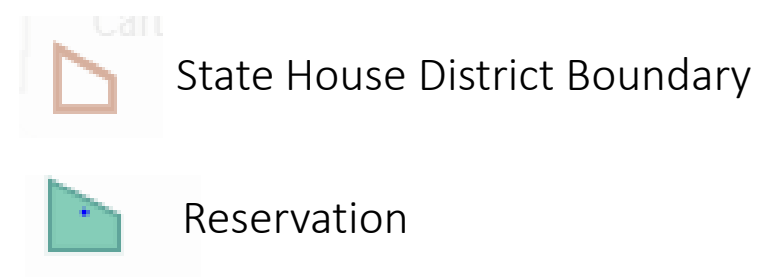
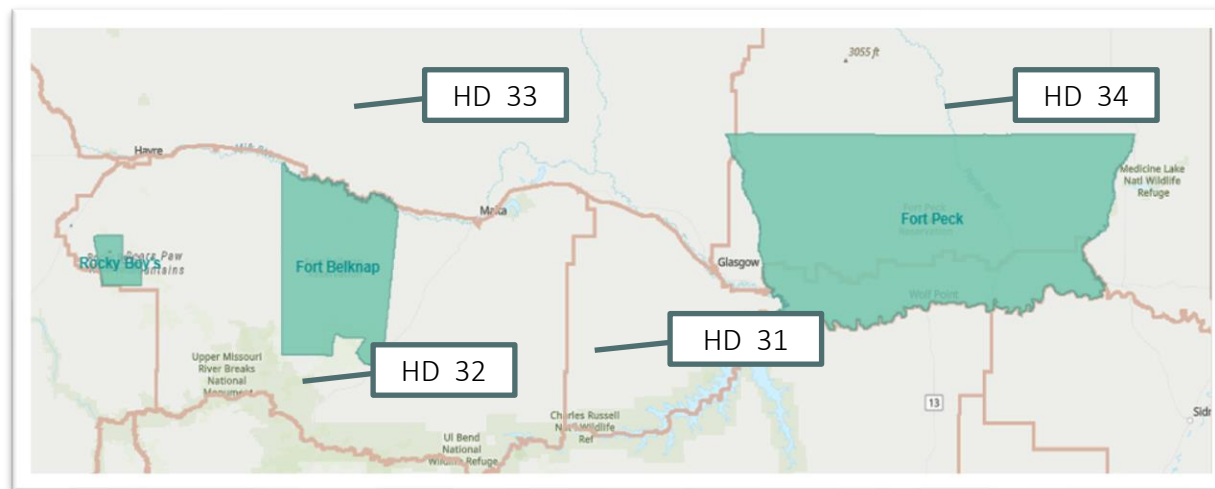


Analysis from Micro-file Data Set Produced November 2020



# Data Example #5 for Montana

## Population Associated with Blocks Reported in the 2010 Census and with DAS Applied for Creating Boundaries – Shifts in Race



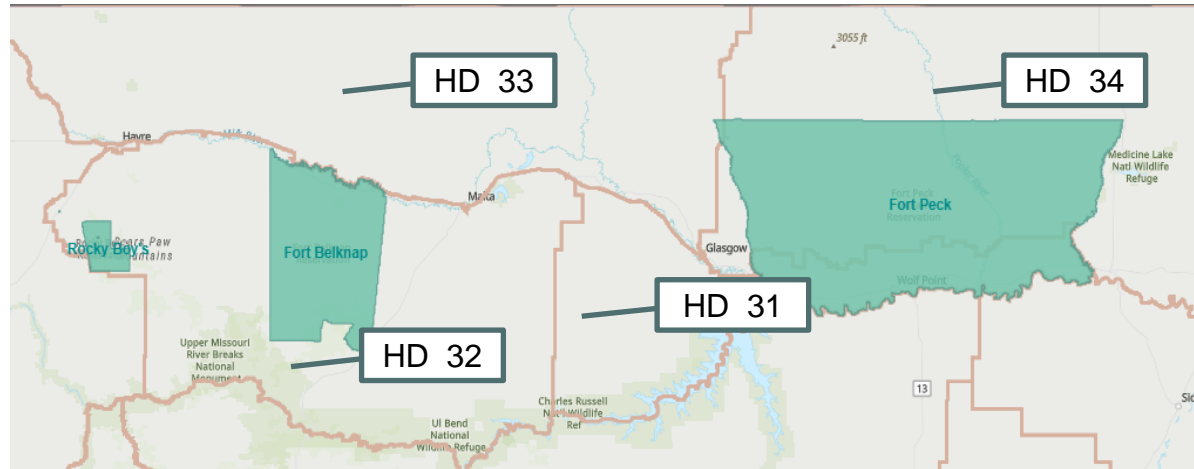
Legislative District	DAS Added 2010 Census - Total Pop	2010 Census Total Pop	Difference	% American Indian from 2010 Census (used in 2014 District Boundaries)
HD 31	9,698	9,837	- 139	71%
HD 32	9,956	9,800	+ 156	75%
HD 33	9,663	9,766	- 103	10%
HD 34	10,046	9,882	+ 164	7%

Analysis from Micro-file Data Set Produced November 2020



# Data Example #6 for Montana

Population Associated with Blocks Reported in the 2010 Census and with DAS Applied for Creating Boundaries – Shifts in Voting Age Population



Legislative District	DAS Added 2010 Census – Pop 18+	2010 Census Pop 18+	Difference
HD 31	6,626	6,627	- 1
HD 32	6,521	6,445	+76
HD 33	7,553	7,721	- 8
HD 34	7,775	7,721	+54

Analysis from Micro-file Data Set Produced November 2020



# Current Timing of Implementation

- April 30, 2021: Apportionment file released (no DAS applied)
- **April 30, 2021: New DAS Micro-files and Detailed summary metrics**
- **Early June 2021: Data Stewardship Executive Policy Committee (DSEP) finalizes policy decisions** on DAS design (e.g., geography, processing, invariants)
- **September 2021:** Census Bureau releases new micro-files and Detailed Summary Metrics from applying the production version of the DAS to the 2010 Census data; Census Bureau releases production code base for P.L. 94-171 redistricting summary data file and related technical papers.
- **September 30, 2021** Release of PL94-171 (redistricting files)



# Summary

- The Census Bureau continues to refine the algorithm to improve the accuracy of the data. The next data set is scheduled for release at the end of April 2021..
- Alabama has sued to prevent the use of DAS. Utah has filed an Amicus Brief and other states are expected to follow. Other states have sued about the Census count operations.
- The redistricting data will be coming out in August 30 for "interim" Redistricting in legacy (2010) delimited format and by September 30 for the standard release the Census data site.
- When the data is released, it will require lots of careful review of existing Montana Statutes, Administrative Rules, and policies to determine the impacts and if adjustments are needed. Also, the ability to challenge through CQR is still under discussion.



# What Montana Census Data Users Can Do

- ✓ Educate ourselves on DP and its impacts
- ✓ Follow the DAS Committee's work and understand how new decisions will affect Montana's data
- ✓ Work together to share knowledge and inform policy makers and others impacted
- ✓ Express concerns to the Census DAS Committee
- ✓ Start now on contingency planning of anticipated impacts





Questions



Census Website on Disclosure Avoidance / DP

[https://www.census.gov/about/policies/privacy/statistical\\_safeguards.html](https://www.census.gov/about/policies/privacy/statistical_safeguards.html)

Contact info: Mary Craigle, Commerce, 406.841.2742, [Mary.Craigle@mt.gov](mailto:Mary.Craigle@mt.gov)

THANK YOU!

