



Implications of new Census Bureau data privacy protocols for measuring the health status of Alaska Native peoples

Matthew Berman

Institute of Social and Economic Research
University of Alaska Anchorage
matthew.berman@alaska.edu

The 2020 Alaska Native Population Data Conference
Anchorage, Alaska
Thursday, January 30, 2020

At 90, Alaska Native woman will be 1st counted in US Census

By MARK THIESSEN 2 hours ago



TOKSOOK BAY, Alaska (AP) — Lizzie Chimiugak has lived for 90 years in the windswept western wilds of Alaska, born to a nomadic family who lived in mud homes and followed where the good hunting and fishing led.

Her home now is an outpost on the Bering Sea, [Toksook Bay](#), and she is about to become the first person counted in the U.S. Census, taken every 10 years to apportion representation in Congress and federal money.

It's the Law

*"To stimulate public cooperation necessary for an accurate census...Congress has provided assurances that information furnished by individuals is to be treated as confidential. Title 13 U.S.C. §§ 8(b) and 9(a) explicitly provide for nondisclosure of certain census data, and **no discretion is provided to the Census Bureau on whether or not to disclose such data**..." (U.S. Supreme Court, Baldrige v. Shapiro, 1982)*

Title 13, Section 9 of the United State Code prohibits the Census Bureau from releasing identifiable data "furnished by any particular establishment or individual."

Census Bureau employees are sworn for life to safeguard respondents' information.

Penalties for violating these protections can include fines of up to \$250,000, and/or imprisonment for up to five years!

Michael Hawes

Senior Advisor for Data Access and Privacy
Research and Methodology Directorate
U.S. Census Bureau

4

Shape
your future
START HERE >

United States
Census
2020



**UAA Institute of Social
and Economic Research**
UNIVERSITY of ALASKA ANCHORAGE

Census Bureau's dual (duelling) mandate: balance privacy concerns with public interest in understanding how the U.S. population is changing.

- U.S. Constitution requires count of the population every ten years
 - Census population counts used to determine seats in the U.S. House of Representatives
 - States use census counts to determine legislative district boundaries
- Census data required for numerous federal and state programs
 - Decennial census counts by race used to enforce the federal Voting Rights Act
 - Federal agencies use social and economic characteristics of communities, collected since 2005 in the American Community Survey (ACS), to allocate hundreds of billions in program funds annually
- Census privacy mandate
 - "[The Census Bureau] shall not make any publication whereby the data furnished by any particular establishment or individual ... can be identified." (Title 13 U.S.C. Chapter 1 § 9(a)(2).




Challenge of the Big Data Revolution

- Exponential increase in computing power
- Proliferation of massive databases of information from digital media
- Internet provides potentially unlimited data access to everyone
- Census staff have demonstrated the ability to reconstruct databases from published data products and identify individuals in Public Use Microdata Samples (PUMS) by combining public census bureau products with publicly available data from other sources
- Census Bureau (belatedly) joins the club.
 - Moves to adopt 21st Century privacy protocols long used in the private sector to protect anonymity of individual contributors of large databases
 - Differential privacy (Dwork, 2008): index of privacy loss from publication of summaries of a database

Census Bureau has historically used statistical methods to protect data privacy

- Suppression: omitting table cells with small counts

A	B	C	D
E	F	G	H
I	J	K	L
M	N	O	P

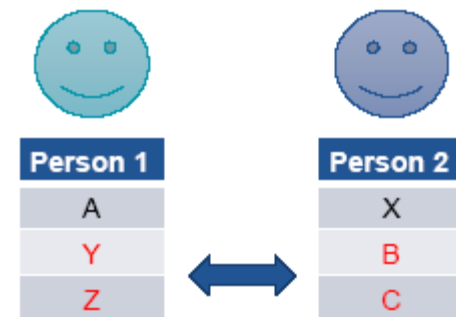


A	B	*	D
E	F	G	H
*	J	K	L
M	N	O	*

- Coarsening: combining categories, reducing geographic detail, reporting ranges, rounding, etc.



- Data swapping: switching some of one persons data with a neighbor



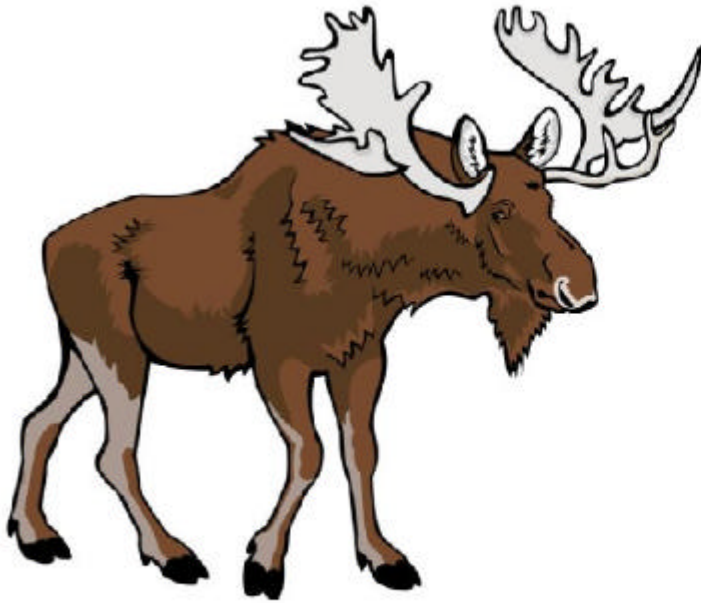
Differential privacy to be implemented in the 2020 Census

- What is differential privacy?
 - Assumes that the probability of identification of individuals in a database is proportional to the change in the table cell value caused by adding or removing an individual's data
 - To protect all persons, you have to protect the data from the person who would change the result the most
 - Fewer people included in the result means that more privacy is lost by publishing the number
- The database manager (Census Bureau) can use differential privacy to provide everyone a quantifiable degree of protection by adding noise to the tables. **That's right: deliberately falsifying the published numbers!**



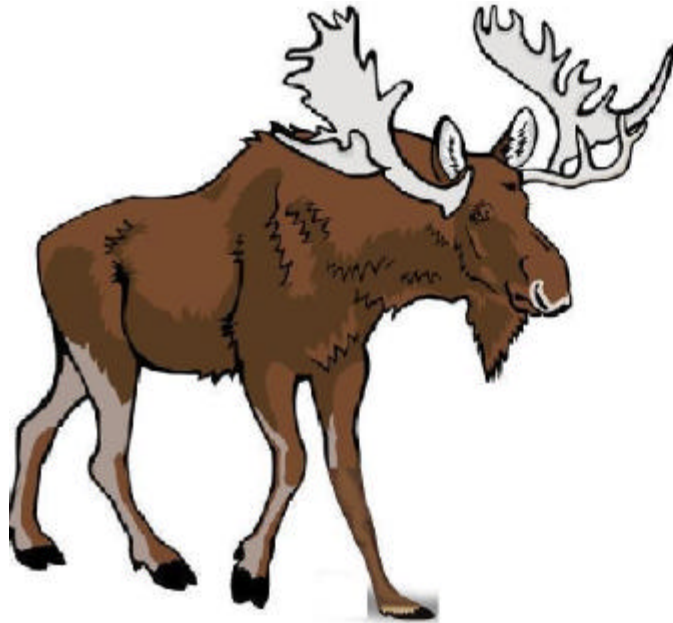
Two communities compared:

Actual counts



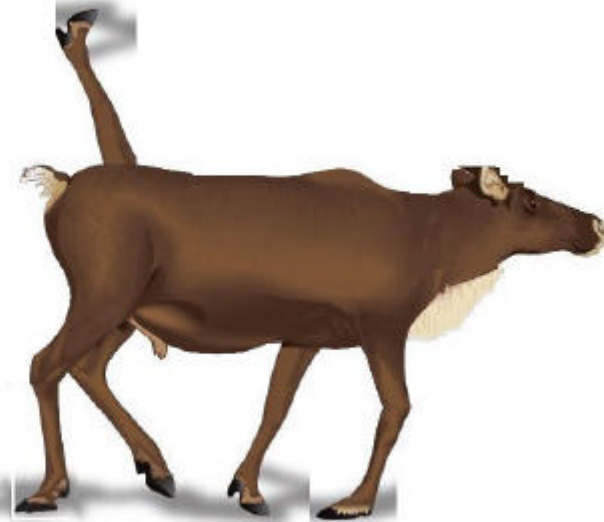
Two communities compared

Statistical disclosure avoidance techniques applied



Two communities compared:

Random noise added

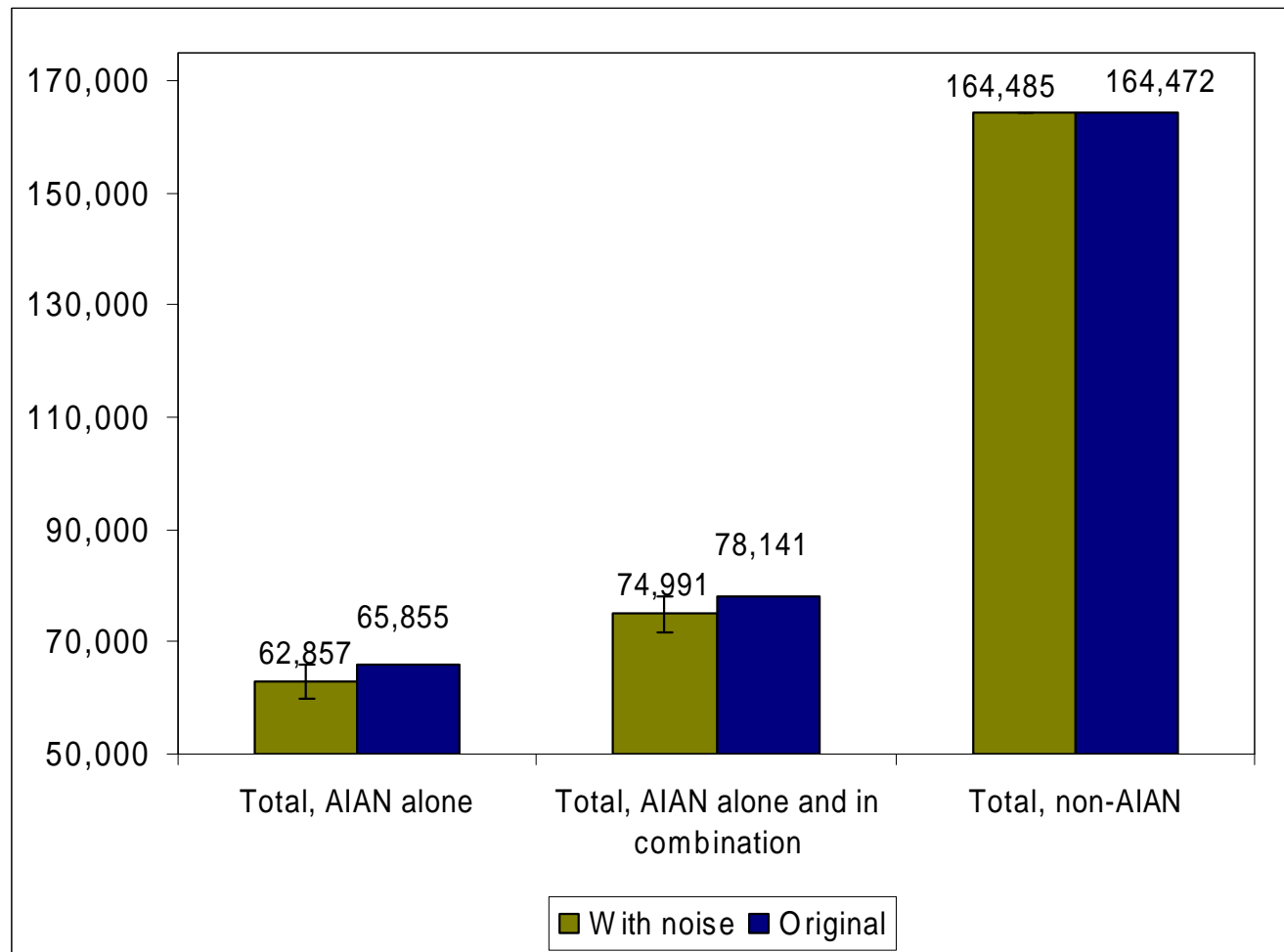


Differential privacy to be implemented in the 2020 Census

- To help the public understand how much noise they intend to add to the 2020 Census, they have released a set of tables from the 2010 Census to compare to the published 2010 results.

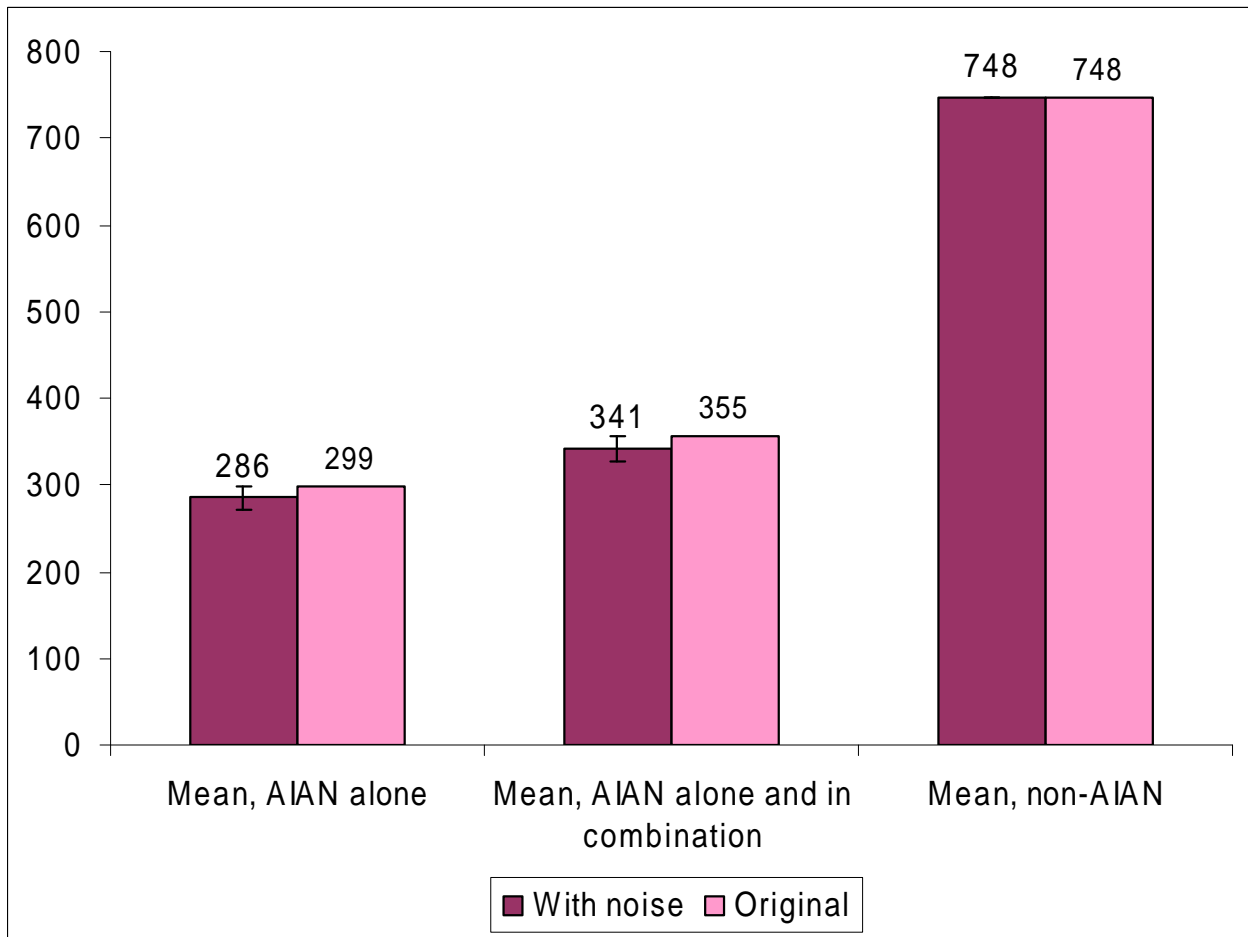
Effects of Noise on Accuracy of Published Data Census 2010 Demonstration Product

- Total population in ANVSAs



Effects of Noise on Accuracy of Published Data Census 2010 Demonstration Product

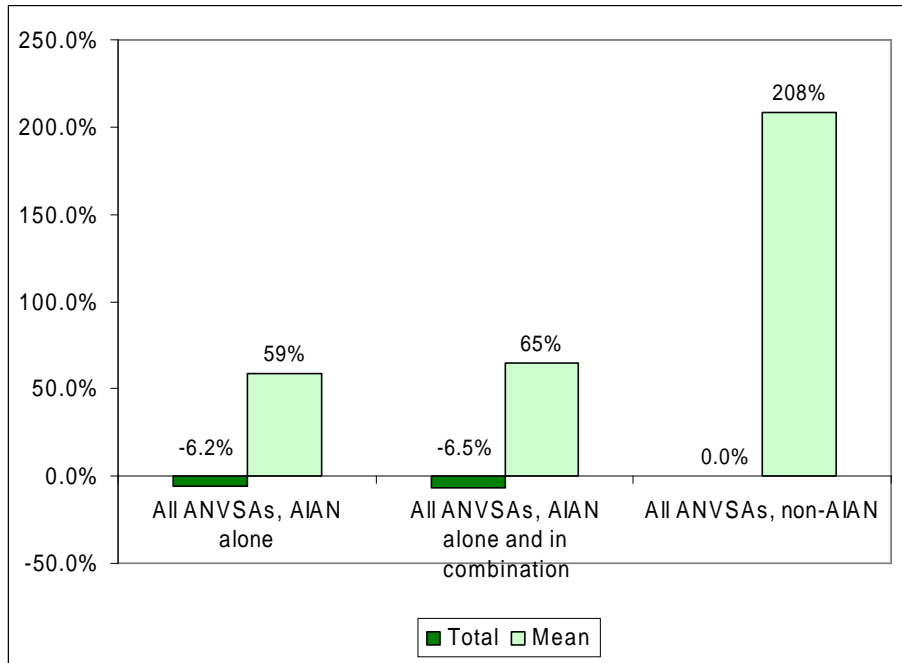
- Mean population of ANVSAs



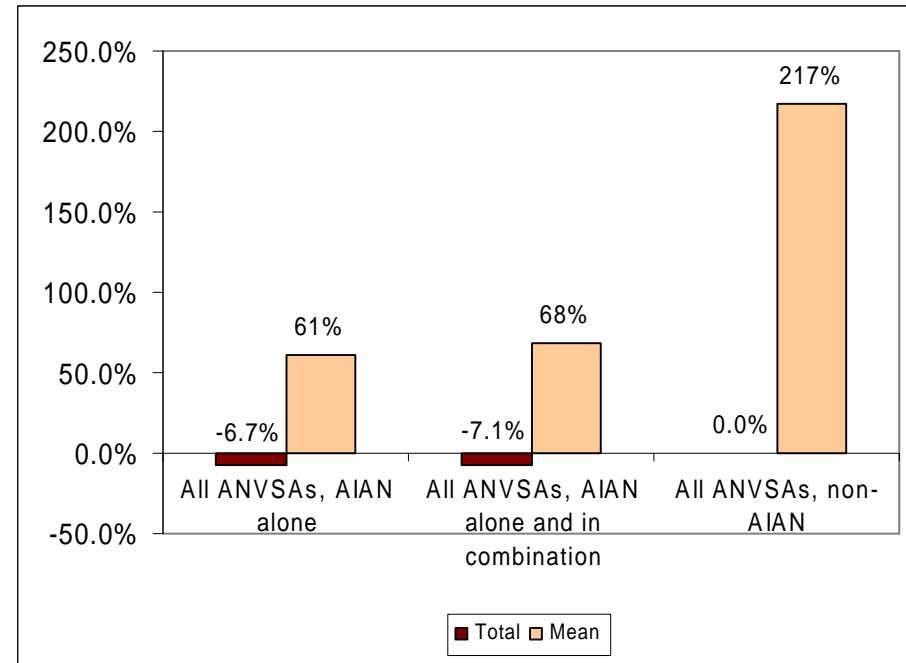
Effects of Noise on Accuracy of Published Data Census 2010 Demonstration Product

- Coefficient of variation from added noise, total population

All ANVSAs



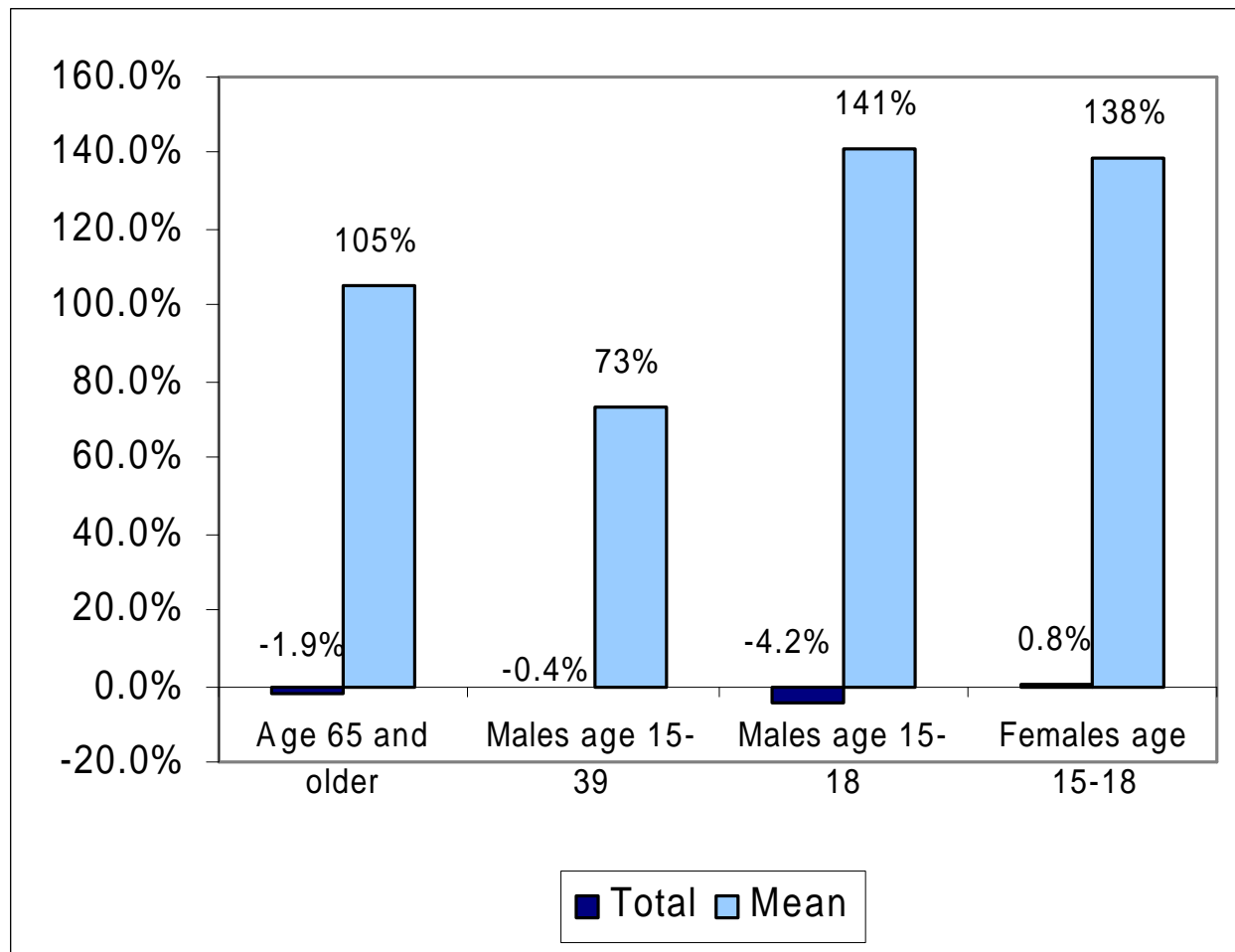
Occupied, non-urban ANVSAs



Effects of Noise on Accuracy of Published Data

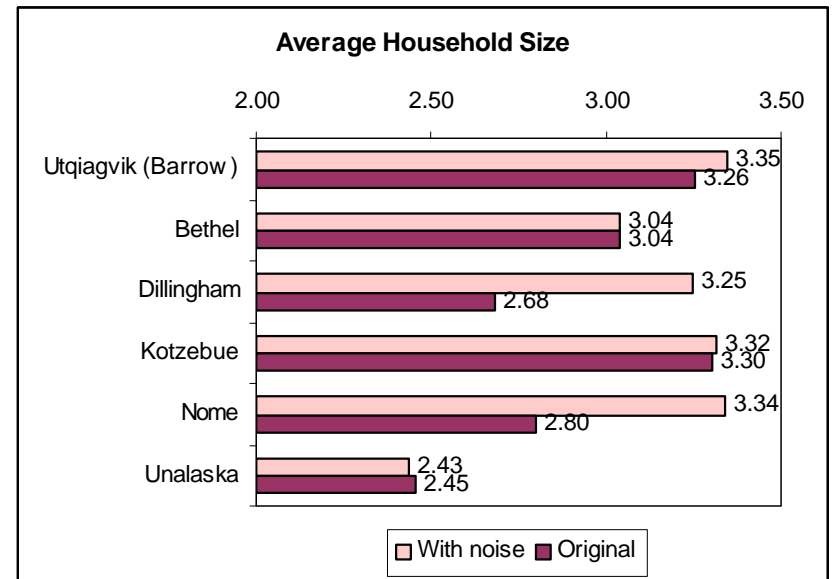
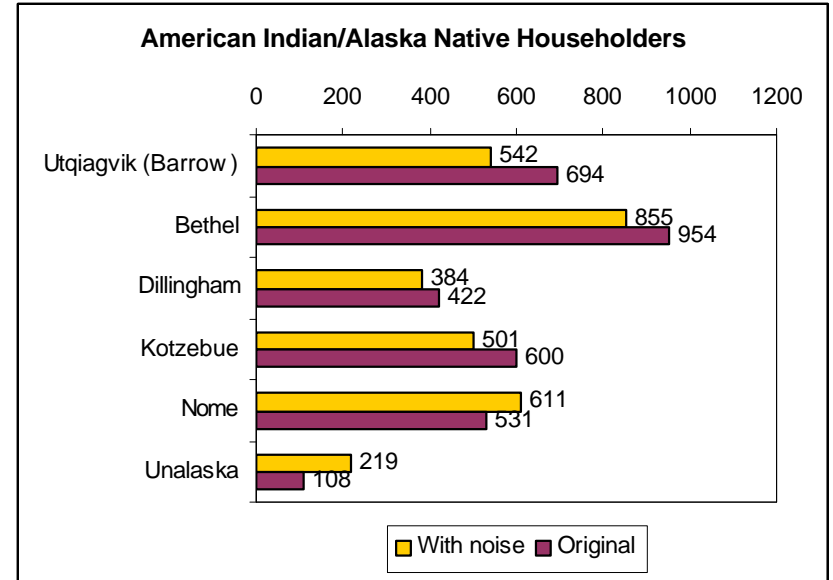
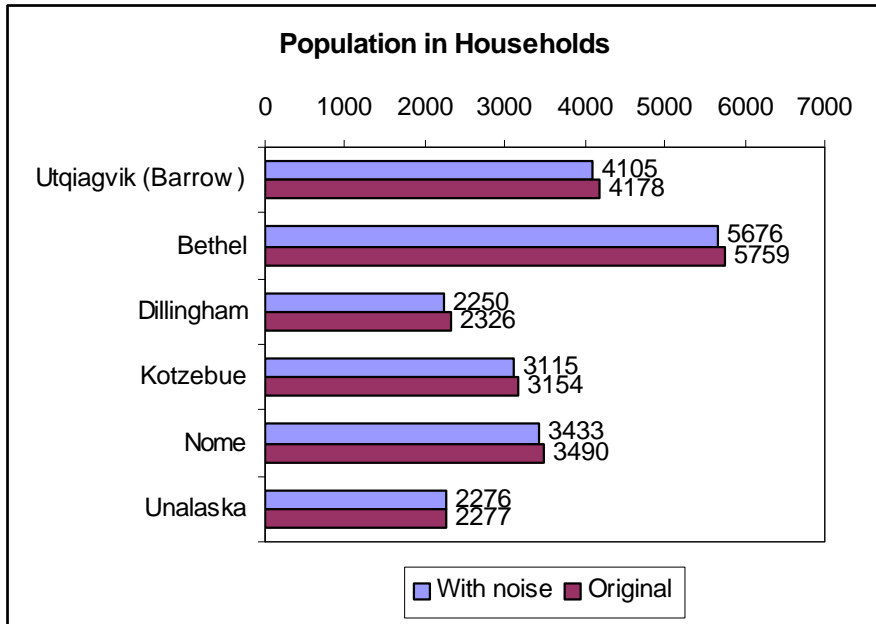
Census 2010 Demonstration Product

- Coefficient of variation from added noise, ANVSA population age-sex cohorts



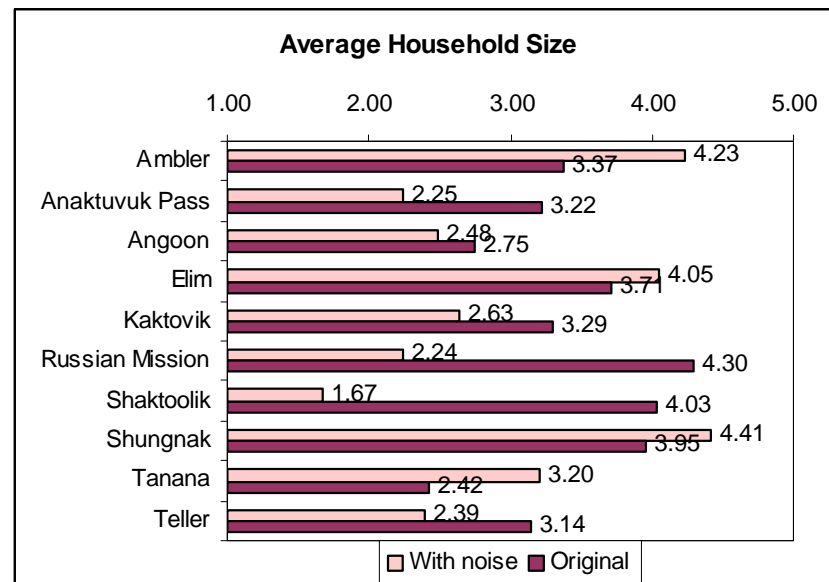
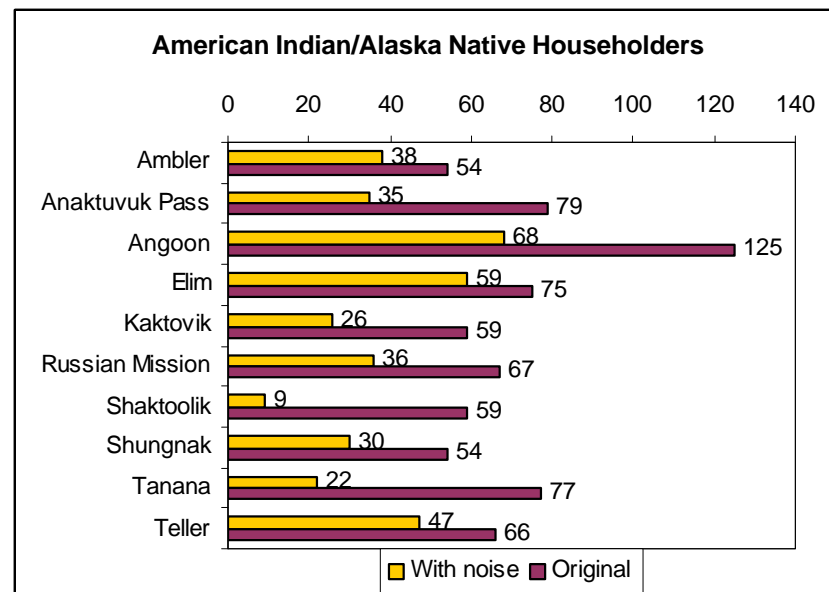
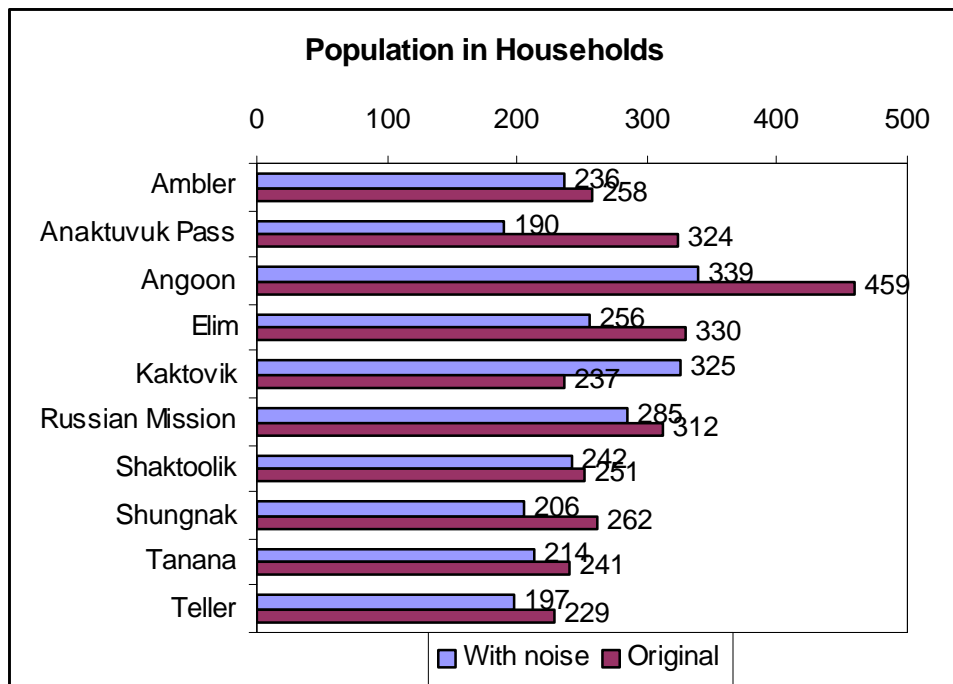
Census 2010 Demonstration Data Product

Comparing example privacy-enhanced tables to published tables: Rural Alaska regional hub communities



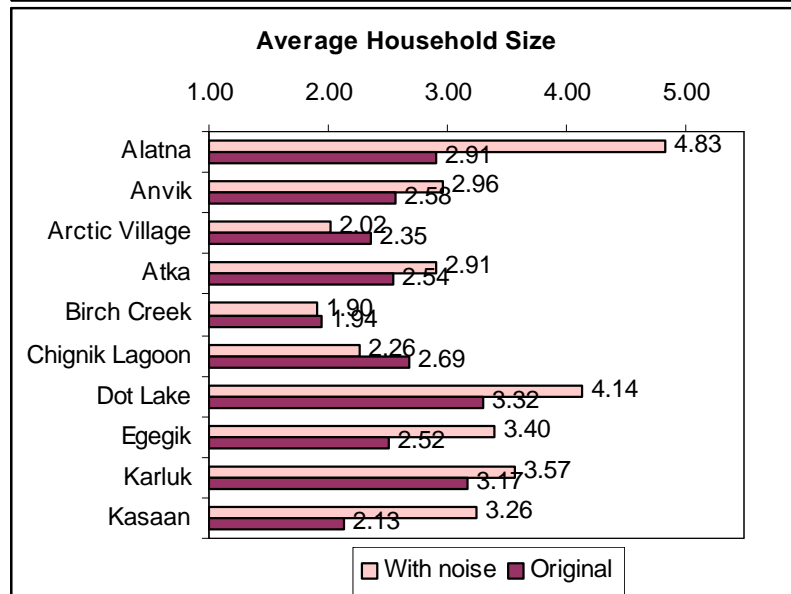
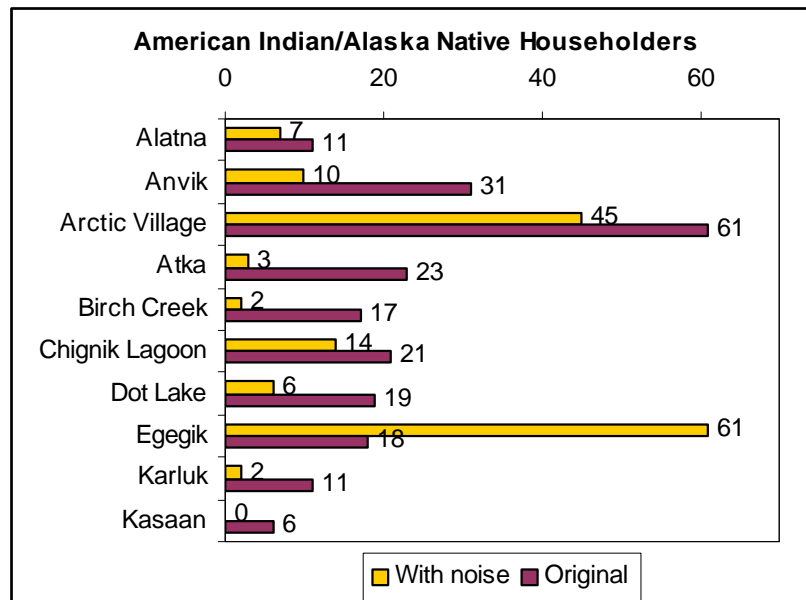
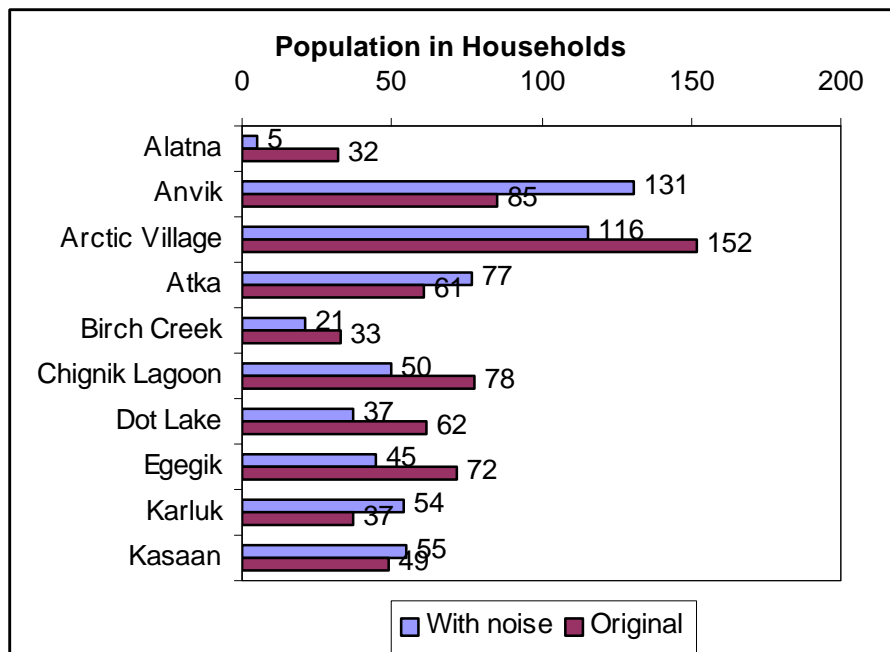
Census 2010 Demonstration Data Product

Comparing example privacy-enhanced tables to published tables: Rural Alaska mid-sized villages



Census 2010 Demonstration Data Product

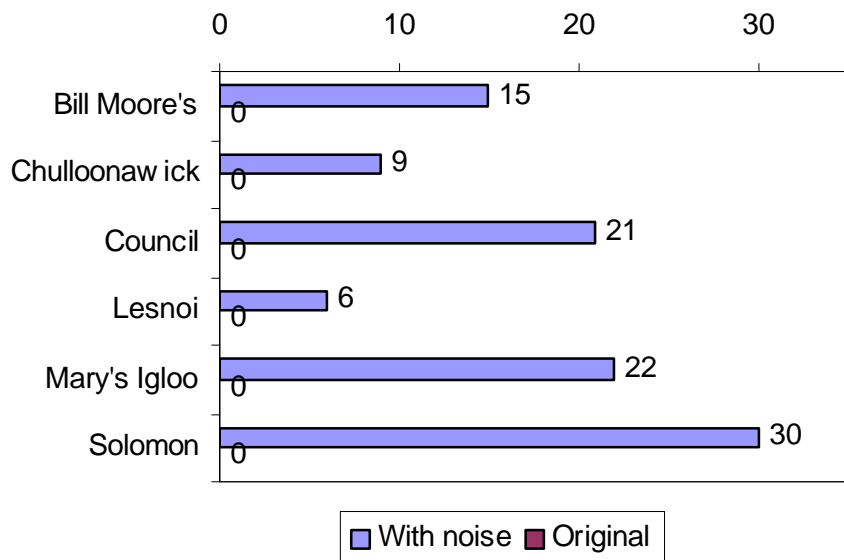
Comparing example privacy-enhanced tables to published tables: Rural Alaska small villages



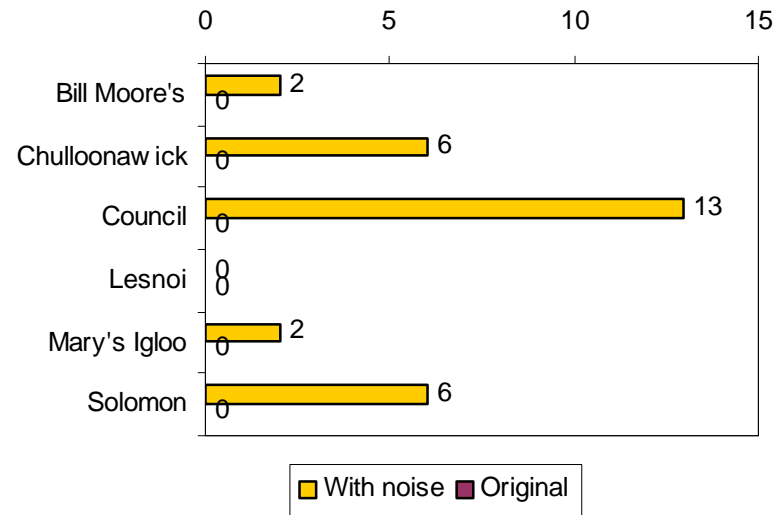
Census Bureau Demonstration Data Product

Comparing example privacy-enhanced tables to published tables:
Rural Alaska abandoned villages

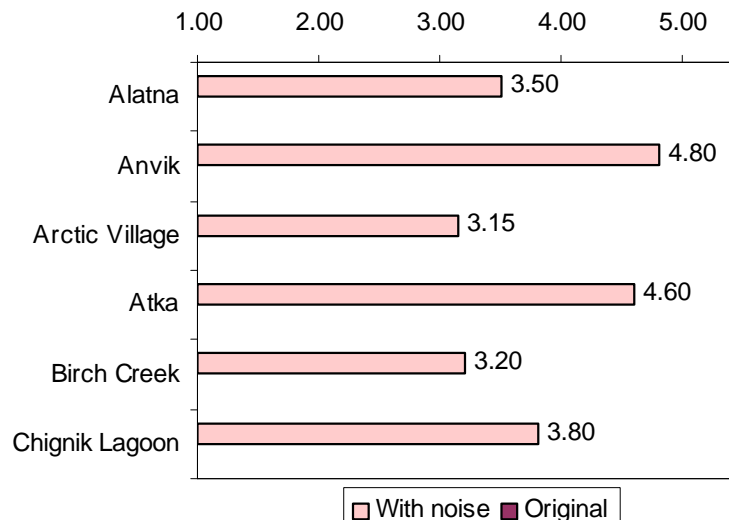
Population in Households



American Indian/Alaska Native Householders



Average Household Size



What can we infer from the Census 2010 Demonstration Product about prospects for the 2020 Census?

- Implementation of new privacy protocols will produce unusable demographic data for rural Alaska communities
- The Census Bureau cannot publish tables that provide meaningful demographic data on small areas and comply with Title 13 U.S.C. Chapter 1 § 9(a)(2)
- The Census Bureau cannot even publish tables that provide meaningful demographic data on small areas and guarantee census respondents a reasonable probability that their individual data cannot be identified.
- The Census Bureau needs to acknowledge that its mandate to provide information on the characteristics of the population of the United States conflicts with the mandate to protect confidentiality of respondents' data..
 - Enhanced privacy protection through implementation of advanced differential privacy techniques is a band-aid solution to a system that is broken
 - More fundamental change is required



If you think this is problematic, consider the American Community Survey (ACS)

- The ACS is the source of all the social and economic characteristics of the population besides age, sex, and race.
- Unlike decennial census data, which are counts of the whole population, ACS data comes from an annual sample.
 - Data are published for small areas as five-year moving averages
 - Non-response to individual questions imputed based on responses from “similar” individuals
 - Privacy protected by statistical disclosure methods such as data swapping, top and bottom coding, etc.
 - Margins of error for rural Alaska communities often as high as 30-50 percent. (This is in addition to non-sampling sources of error such as missing nearly half of household PFD income by not asking respondents about income of children under age 15.)
- Census Bureau says they will not try to add noise to ACS before 2025.
 - Will fewer tables be published?
 - What about the PUMS?



Where do we go from here?

Options to consider

- Stop publishing data for rural areas and tribes?
 - Even with the differential privacy approach, Census Bureau acknowledges that it cannot guarantee that every individual's data can be protected.
 - I.e., **CB acknowledges that they may be unable to comply with the law.**
- Change policy for what is considered private and confidential?
 - Is the number of people living in a particular housing unit really confidential?
 - Is a person's age and gender really confidential? (These are available for most people from public records)
 - Is a person's race confidential? (Like gender, race is self-identified).
- Change census enumeration protocol?
 - Ask respondents for informed consent (There is a long history of federal legislation related to human research subjects. Why should the census be different? The Constitution just mandates that individuals be counted, not described.)
 - Add “do not wish to disclose” as options for age, gender, and race
- Change the law: Congress decides what information is private?
 - Is some data about a person more private than other data?

