



Evaluating the Impact of Differential Privacy Using the Census Bureau's 2010 Demonstration Data Products and Implications for Data Users

In 2018, the Census Bureau announced a new set of methods for disclosure avoidance in public use data products, including aggregate-level tabular data and microdata derived from the decennial census and other Census Bureau surveys. The new approach, known as *differential privacy*, will result in a significant trade-off in data accuracy, potentially making the public data less useful for many scientific, program administration, and policy applications. Differential privacy techniques could make the release of useful microdata impossible and severely limit the utility of tabular small-area data. Therefore, programs that rely on accurate census counts and tabulations populations may need to identify alternative strategies.

The Texas Demographic Center has been reviewing and conducting analyses on the Census Bureau's 2010 Demonstration Data Products. In this information sheet, we communicate a few of concerns based on our review of the demonstration data for Texas counties and places.

2020 Census Data Products Planning Crosswalk

Part of the 2010 Demonstration Data Products includes a 2020 Census Data Products Planning Crosswalk. The crosswalk consists of table shells proposed to be included in the Redistricting Data Summary File and a comparison of the tables proposed for the 2020 Demographic and Housing Characteristics and the 2010 Summary File 1.

A brief review of the crosswalk reveals two variables will potentially not be made available for analysis in 2020 following differential privacy. These variables include average household size and household type, consisting of family, female/male headed households, etc. The average household size statistic is key to the TDC population estimates and projections program and we anticipate could also impact the work of other data users. The household type variable is frequently used by our office in describing populations most at risk of poverty, lack of health insurance, and other key socio-economic outcomes. We anticipate this variable is also pertinent to the work of many of our TDC affiliates and partners.

Analysis of 2010 Demonstration Data

We compared the demonstration data with the corresponding 2010 census data products at different geographic levels, using select characteristics. We compared the population aged 18 and over for Texas Senate and House districts. The differences in total population for the Senate districts were minimal, less than half a percent in general. For House districts, we found that four of the 150 districts had absolute differences greater than half a percent but still less than one percent. The race composition of the districts was also relatively consistent between the two datasets, with differences of generally less than one percent.

We also compared total population of Texas counties and places. Table 1 lists five counties with the biggest absolute percent

What is Differential Privacy?

In order to ensure respondent confidentiality in all data products as required by Title 13 and Title 26, the U.S. Census Bureau developed disclosure avoidance policies that have been implemented in different versions throughout the Bureau's history. More powerful computers, privacy threats, and the availability of personal online data require new safeguards be implemented to avoid disclosure of respondent data. The 2020 Census will use a new, powerful protection system known as differential privacy. Differential privacy uses advanced statistical methods to "swap data" and "insert noise" into the data in order to maintain the confidentiality of respondents. Differential privacy techniques will result in a trade-off in data accuracy and data usefulness.

Table 1: Differences in Population of Greater than Five Percent between 2010 Summary File 1 and 2010 Demonstration Data

County	SF1 Population	Demo Data Population	Percent Difference
McMullen	707	758	7.2%
Kenedy	416	451	8.4%
Kent	808	912	12.9%
Loving	82	94	14.6%
King	286	335	17.1%

Source: U.S. Census Bureau, 2010 Census Summary File 1 and 2010 Demonstration Data

differences, all of them having populations of less than 1,000 in 2010. Table 2 shows findings for places, including the Census Designated Places (CDPs), grouped by size of the population and the calculated mean absolute percent difference for each group. Similarly, the biggest differences were found among places with populations of less than 1,000. It is noteworthy that the 773 places with this population size comprise more than 40 percent of all Texas places.

Additionally, we looked at occupancy rates at the census tract level. Of the 5,265 census tracts in Texas, only 20 had full occupancy rates in the original 2010 summary file, whereas 1,190, or 22.6 percent, had full occupancy rates in the demonstration data.

These initial analyses reveal the implementation of differential privacy is most concerning in its potential impact on data accuracy at small geographies and population size.

Opportunities for Feedback

Census data users are encouraged to review the 2020 Census Data Products Planning Crosswalk to determine if variables pertinent to their analyses and research are impacted by differential privacy.

Table 2: Differences in Population by Place Population Size between 2010 Summary File 1 and 2010 Demonstration Data

Population	Number of Places	Mean Absolute Percent Difference
1-500	513	25.7%
501-1,000	260	10.1%
1,001-5,000	591	4.3%
5,001-10,000	143	1.7%
10,001 plus	241	0.7%

Source: U.S. Census Bureau, 2010 Census Summary File 1 and 2010 Demonstration Data

Additionally, organizations are encouraged to conduct their own analyses using the 2010 Demonstration Data and comparing it to the previously published 2010 Census data. Lastly, we encourage organizations to engage with the Census Bureau and provide input on what their analyses reveal and any other concerns they may have. Organizations can submit their feedback to the Census Bureau by emailing: dcmd.2010.demonstration.data.products@census.gov.

Useful Links:

U.S. Census Bureau 2010 Demonstration Data Products:

<https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/2020-census-data-products/2010-demonstration-data-products.html>

Disclosure Avoidance and the 2020 Census:

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

About the Texas Demographic Center

The Texas Demographic Center produces, interprets, and disseminates demographic information to facilitate data driven decision making for the benefit of Texans.



San Antonio Office

The University of Texas at San Antonio
501 West Cesar E. Chavez Blvd.
San Antonio, TX 78207-4415
Ph: 210-458-6543
Fax: 210-458-6541

Austin Office

P.O. Box 13455
Austin, TX 78711
Ph: 512-463-8390
Fax: 512-463-7632