The Effect of the Differential Privacy Disclosure Avoidance System Proposed by the Census Bureau on 2020 Census Products: Four Case Studies of Census Blocks in Alaska.

David A. Swanson, University of California Riverside, Riverside, CA
The Center for Studies in Demography and Ecology, University of Washington, Seattle, WA
Social Science Research Center, Mississippi State University, Starkville, MS
(email: dswanson@ucr.edu)

T. M. Bryan, Bryan Demographic Research, Richmond, VA
(email: tom@bryangeodemo.com

Richard Sewell, Alaska Department of Transportation and Public Facilities, Anchorage, AK
(email: resewell@gmail.com)

Abstract

The Census Bureau plans to introduce a new Disclosure Avoidance System known as Differential Privacy (DP) for its 2020 census data products. Using two DP demonstration product files provided by the Census Bureau, we assess the errors introduced by DP on census block data in Alaska in the form of four case studies and find them to be substantial by type and level. We use both the May 27th 2020 DP demonstration product and the most recent, the April 28th 2021 DP demonstration product relative to our four cases studies and compare the changes. This comparison is important because the Census Bureau reports that accuracy should improve because the privacy budget was increased in response to user complaints about poor accuracy. We find that the April 28th, 2021 release does produce more accurate data but that the level of accuracy remains unsuitable for use by those who work with small area data. Because it is likely that the results we found in Alaska will be found in other states, our examination leads us to conclude that it is likely that the errors introduced by DP of the type and at the level found in the most recent demonstration product files we examined will render the nation's block level data essentially unusable.

Introduction

The Census Bureau plans to introduce a new Disclosure Avoidance System known as Differential Privacy (DP) for its 2020 census data products (Abowd, 2020, Census Bureau 2020a, 2020b, 202c, 2020d, 2020e, 2020f, and 2020g). Our purpose in this paper is to assess the errors introduced by (DP) on census block data in Alaska in the form of four case studies.

Ruggles et al. (2019: 406) argue that DP goes far beyond what is necessary to keep data safe under census law and precedent and because it focuses on concealing individual characteristics instead of respondent identities, DP is a blunt and inefficient instrument for disclosure control. They go on to note that because the core metric of DP does not measure the risk of identity disclosure, it cannot assess disclosure risk as defined under census law, making it untenable for optimizing the privacy/usability trade-off.

Background

Covering 570,641 square miles of land, Alaska is the largest state but with the 2010 census showing only 710,231 people, it is the least densely-populated of the 50 states, at 1.24 people per square mile (Hunsinger et al. 2012: 8). The 2010 census (see below), organized the state into 45,292 census blocks, of which only 12,870 had one or more people, leaving 32,422 without any population. On average, there were 15.68 persons in each of these 45,292 census blocks. If we look at the 12,870 census blocks with at

least one person, there were 55.2 persons on average in each of these 12,870 blocks. These summary statistics make Alaska one of the states in which one would expect a high level of disclosure avoidance at the block level because there are so few people on average per block. This is a point to which we return in the final section.

Data

The application of DP is a brand new approach for the Census Bureau and is different from all prior Census Bureau initiatives in regard to disclosure avoidance. As a component of the DP initiative, the Census Bureau has released a series of "demonstration products" (Abowd, 2020, Census Bureau 2020a, 2020b, 202c, 2020d, 2020e, 2020f, and 2020g) that allow outside analysts and stakeholders to determine for their purposes the impact DP would have on Census data. These demonstration products generally contain:

- the most common, basic demographic and housing variables
- different levels of geography
- data as they were originally reported in the Summary Files (SF) in 2010, which reported actual census data with small privacy protection modifications
- trial data as they have been by adjusted (perturbed) DP

As the Census Bureau responded to User complaints about poor accuracy, the "privacy budgets" were changed in the demonstration products to provide higher levels of accuracy (Beveridge, 2021). Here, we examine the errors introduced by DP on 2010 census block data for Alaska in the form of four case studies. In our initial analysis, we employ the "demonstration product" for census blocks in Alaska released May 27th, 2020, fil (labeled as 2020527) with an epsilon level of 4.0, which was downloaded from the Minnesota Population Center's NHGIS site: https://nhgis.org/privacy-protected-demonstration-data. Against the results we find from the May 27th, 2020 file, we compare results from the most recent release, April 28th, 2021. (file labeled as 20210428) with an epsilon level of 10.3, which was downloaded from the same site.

In the analyses for case studies 1 through 3, we employed the cross-tabulation routine found in Release 12 of the NCSS Statistical System (https://www.ncss.com/software/ncss/). For case study 4, we sorted the blocks in descending order by the 2010 census total population, then used the logical "IF' function to examine differences between the 2010 census count and the DP count (match = zero; non-match =1), and summed the number of non-matches.

Results from the May 27th 2020 File

*Case 1: Children without Adults: How Did Differential Privacy turn three blocks into 765***?**

The 2010 census reported that there were three blocks in which 1 or more children (under age 18) were listed, but no adults (18 years and over). Of these three blocks, the first had one child, the second, five children, and the third had 15 children. It is likely that the last block has a facility where children reside in the presence of adults who themselves live elsewhere.

Out of 45,292 blocks, it is highly believable that there are three in which a total of 21 children reside without adults. However, DP produced 765 such blocks in which 3,381 children reside without adults - a highly unbelievable number

Case 2: Differential Privacy turned 1,252 blocks with one or more people of voting age into blocks with zero people of voting age

- In comparing the voting age populations reported by the 2010 census and the DP file, it was found that there are 1,252 blocks in which DP reported zero people of voting age while the 2010 census reported one or more persons of voting age in these same blocks.

Case 3: Differential Privacy turned 830 blocks with zero persons of voting age into blocks with one or more persons of voting age

- At the same time, DP turned 830 blocks in which the 2010 census reported zero persons of voting age into blocks with one or more persons of voting age.

Case 4: Of 12,870 blocks in which the 2010 census shows one or more persons, 12,366 of them (96%) show a different number of persons when DP is applied.

- Of these same 12,870 blocks, 12,009 of them (93%) show a different number of persons of voting age population (18 years and over) when DP is applied.

Results from the April 28th, 2021 File

*Case 1: Children without Adults: How Did Differential Privacy turn three blocks into 428***?**

The 2010 census reported that there were three blocks in which 1 or more children (under age 18) were listed, but no adults (18 years and over). Of these three blocks, the first had one child, the second, five children, and the third had 15 children. It is likely that the last block has a facility where children reside in the presence of adults who themselves live elsewhere.

Out of 45,292 blocks, it is highly believable that there are three in which a total of 21 children reside without adults. However, DP produced 428 such blocks in which 1,302 children reside without adults - a number that remains unbelievable.

Case 2: Differential Privacy turned 533 Blocks with one or more people of voting age into blocks with zero people of voting age

- In comparing the voting age populations reported by the 2010 census and the DP file, it was found that there are 533 blocks in which DP reported zero people of voting age while the 2010 census reported one or more persons of voting age in these same blocks.

Case 3: Differential Privacy turned 830 blocks with zero persons of voting age into blocks with one or more persons of voting age

- At the same time, DP turned 632 blocks in which the 2010 census reported zero persons of voting age into blocks with one or more persons of voting age.

Case 4: Of 12,866 blocks in which the 2010 census shows one or more persons, 11,801 of them (92%) show a different number of persons when DP is applied.

Discussion and Conclusion

Alaska was not subject to higher levels of DP Disclosure Avoidance than the other states in either of the two Demonstration Product" files (2020527 and 20210428) we have analyzed. Instead, the DP levels are reported as uniform across all states at an "epsilon" level of 4.0 and 10.3, respectively, for people (https://www.nhgis.org/privacy-protected-demonstration-data#v20210428_12-2). Given this and the low numbers of people found statewide in the 2010 census and its low number of 2010 census blocks, Alaska would appear to be a candidate for a higher level of DP disclosure avoidance than many other states.

Finding that in going from an epsilon of 4.0 in which DP produced 765 census blocks in which 3,381 children reside without adults to an epsilon of 10.3 in which DP produced 428 such blocks in which 1,302 children reside without adults remains very troubling, as are our other three comparisons

If DP is implemented at the avoidance level found in either of the two "Demonstration Product" files (2020527 and 20210428)) for census blocks in Alaska we examined in this study, it will affect almost all of the state's users of small area census data, from legislatures relying on the data to design Congressional Districts to comply with the law, to demographics vendors who supply clients with zip code level characteristics so businesses can make better decisions. Other end users such as health district administrators who need the data to tract health issues such as COVID-19, and businesses that use small area data such as zip codes, blocks and block groups to improve marketing stand to be dramatically impacted. Many government agencies also depend on accurate small area census data to make programs run efficiently and effectively and the biggest impact of DP will be in small areas. The data in small areas are typically used both directly where the small area is the unit of analysis and aggregated into higher levels of geography by these users. In the case of the latter, the errors introduced by DP tend to even out. However, in the case of the former, these users and their clients will be forced to deal with erroneous data if DP is implemented.

Because it is likely that the results we found in Alaska will be found in other states and perhaps at even higher levels of error, our examination leads us to conclude that it is likely the errors introduced by DP of the type and at the level found in the demonstration product files we examined will render the nation's block level data essentially unusable.

Acknowledgements

We are grateful to the Minnesota Population Center for assembling and making available the DP demonstration product file we use here. We also are grateful for advice and comments from Jan Vink and Bill O'Hare.

References

Abowd, J (2020). Modernizing Disclosure Avoidance: What We've Learned, Where We Are Now. https://www.census.gov/newsroom/blogs/research-matters/2020/03/modernizing_disclosu.html

Beveridge, A. (2021). Controversial Census Bureau Plan That Makes Data Less Accurate Goes to Court. (https://www.socialexplorer.com/blog/post/controversial-census-bureau-plan-that-makes-data-less-accurate-goes-to-court-11452).

Hunsinger, S., D. Howell, and E. Sandberg. (2012). Alaska Population Overview: 2010 Census and 2011 Estimates. Research & Analysis Section, Alaska Department of Labor, Juneau, AK (http://live.laborstats.alaska.gov/pop/estimates/pub/1011popover.pdf).

Ruggles, S., C. Fitch, D. Magnuson, and J. Schroeder. (2019). Differential Privacy: Implications for Social and Economic Research. American Economic Association Papers and Proceedings 109 (May): 403-408.

U.S. Census Bureau (2020a). 2020 Census Disclosure Avoidance Improvement Metrics, U.S Census Bureau, Washington DC., March 18, https://www2.census.gov/programs-surveys/decennial/2020/program-management/data-product-planning/disclosure-avoidance-system/2020-03-18-2020-census-da-improvement-metrics.pdf?#

U.S. Census Bureau (2020b). "2020 Census Data Products and the Disclosure Avoidance System, Hawes M. and Garfinkel. S. L., Planned presentation at the Census Scientific Advisory Committee meeting, March 26,

U.S. Census Bureau (2020c). DAS Updates, U.S Census Bureau, Hawes M. June 1 Washington DC., https://www2.census.gov/programs-surveys/decennial/2020/program-management/data-product-planning/disclosure-avoidance-system/2020-06-01-das-updates.pdf?#

U.S. Census Bureau (2020d). "Disclosure Avoidance and the Census," Select Topics in International Censuses, U.S. Census Bureau, October 2020. https://www.census.gov/library/working-papers/2020/demo/disclos-avoid-census.html

U.S. Census Bureau (2020e). "Disclosure Avoidance and the 2020 Census, U.S. Census Bureau," Washington DC., Accessed November 2nd. https://www.census.gov/about/policies/privacy/statistical_safeguards/disclosure-avoidance-2020-census.html

U.S. Census Bureau (2020f). Error Discovered in PPM, U.S. Census Bureau, Washington DC. https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/2020-census-data-products/2020-das-updates.html

U.S. Census Bureau (2020g). "2020 Disclosure Avoidance System Updates," U.S. Census Bureau, Washington DC., https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/2020-census-data-products/2020-das-updates.html