The Effect of the Differential Privacy Disclosure Avoidance System Proposed by the Census Bureau on 2020 Population Counts: Three Case Studies of Census Blocks in Washington.

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

Abstract

The Census Bureau plans to introduce a new Disclosure Avoidance System known as Differential Privacy (DP) for its 2020 census data products. Using the most recent (28 April, 2021) DP demonstration product file provided by the Census Bureau, I assess the errors introduced by DP on Washington's census block population data in the form of three case studies and find them to be substantial by type and level. Because it is likely that the results found in Washington will be found in other states, this examination leads me to conclude that it is likely that the errors introduced by DP of the type and at the level found in the demonstration product file I 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). My purpose in this paper is to assess the errors introduced by (DP) on census block data in Washington in the form of three 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 66,455.52 square miles of land with a 2019 population of 7,546,410, Washington has a population density of 113.56 persons per square mile (Office of Financial Management 2020). The 2010 census (see below) counted 6,729,540 persons and organized the state into 195,574 census blocks (Office of Financial Management, 2011). On average, there were 31.42 persons in each of these 34.41 195,574 census blocks in 2010 and an estimated 34.41 in them as of 2019.

In a letter dated February 6th, 2020, Mike Mohrman, the Washington State Demographer, wrote a letter to Steven Dillingham, then-Director of the U.S. Census Bureau outlining a host of problems with DP as applied at a lower level of Epsilon (which creates more error than a higher level of epsilon) to a range of Washington data (Mohrman, 2020). My examination of the more recent example of DP applied to Washington State can be viewed as a limited follow-up to the comprehensive examination outlined in Mohrman's letter, one that looks at the state from the perspective of less DP-introduced error.

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 as noted supra page
- trial data as they have been by adjusted (perturbed) DP

Here, I examine the errors introduced by DP on 2010 Census SF block data for Washington in the form of three case studies. I employ the "demonstration product" for census blocks in Washington labeled as 20210428, which I downloaded from the Minnesota Population Center's NHGIS site: https://nhgis.org/privacy-protected-demonstration-data. This file has an "epsilon level of 10.3 (U.S. Census Bureau, 2021). In the analysis, I utilize all of the 195,574 census blocks in Washington found in this file and the 2010 populations in them.

In the analyses for the three case studies, I employed the cross-tabulation routine found in Release 12 of the NCSS Statistical System (https://www.ncss.com/software/ncss/).

Results

Case 1: The 2010 census reported that there were 11 blocks in 96 children (under age 18) were listed, but no adults (18 years and over), a highly believable number, given the presence of juvenile facilities and other institutional settings for those under the age of 18. However, DP produced 2,217 such blocks in which 7,234 children reside without adults - a highly unbelievable number.
Case 2: Of 76,800 blocks in which the 2010 census reported zero population, DP turned 1,617 of them into blocks with people of voting age.
Case 3: Of 118,774 blocks in which the 2010 census reported one or more persons of voting age (18 years and over), DP turned 1,624 of them into blocks with zero people of voting age.

Discussion and Conclusion

If DP is implemented at the avoidance level found in the "Demonstration Product" file (20210428, with Epsilon = 10.3) for the population by census block in Washington I examined in this study, it will affect almost all of the state's users of small area census population 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 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 population data if DP is implemented at the level examined here (Epsilon =10.3).

This is troubling in that while the problems outlined in Mohrman's letter (Mohrman, 2020) have been reduced in this latest demonstration product, they likely remain substantial, given the results of my limited analysis.

Because the results I found in Washington are similar to the error levels found in the 2021048 population "demonstration products for both Alaska (Swanson, Bryan, and Sewell, 2021) and Mississippi (Swanson and Cossman, 2021), it is likely that similar levels of error will be found in other states and perhaps at even higher levels. As such, this examination leads me to conclude that it is likely the errors in population data introduced by DP of the type and at the level found in the 20210428 demonstration product file I examined will render the nation's block level population data essentially unusable.

Acknowledgements

I am grateful to the Minnesota Population Center for assembling and making available the DP demonstration product file used here.

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

Mohrman, M. (2020). WA_OFM_DAS_Respnse_letter (https://www.ncsl.org/Portals/1/Documents/Redistricting/WA_OFM_DAS_Response_Letter.pdf).

Office of Financial Management. (2020). *2019 Data Book*. Washington State Office of Financial Management. Olympia, WA (https://ofm.wa.gov/washington-data-research/statewide-data/washington-state-data-book)

Office of Financial Management (2011). *2010 State Demographic Profile*. Washington State Office of Financial Management (https://ofm.wa.gov/washington-data-research/population-demographics/decennial-census/census-2010/census-2010-data).

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.

Swanson, D., T. Bryan, and R. Sewell. (2021). "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." Presented at the (virtual) Symposium on Data Sciences and Statistics, June 4th, 2021

Swanson, D., and R. Cossman (2021). 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 Mississippi. (to be) presented at the Joint Statistical Meeting, American Statistical Association, Seattle, WA. August 8-12.

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

U.S. Census Bureau (2021). “4/28/21: Now Online: New Disclosure Avoidance System Update Meets or Exceeds Redistricting Accuracy Targets.” (https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/disclosure-avoidance/2020-das-updates.html).