

PRB

INFORM
EMPOWER
ADVANCE

2020 Census demonstration data: Privacy and accuracy issues

Beth Jarosz
@DataGeekB

December 2019

POPULATION REFERENCE BUREAU | www.prb.org

Overview

- Background
- Differential Privacy (DP)
- Fewer data products/tables
- 2010 demonstration data
- Call to action

USC Title 13 - Protection of Confidential Information

§9. Information as confidential; exception

(a) Neither the Secretary, nor any other officer or employee of the Department of Commerce or bureau or agency thereof, or local government census liaison, may, except as provided in section 8 or 16 or chapter 10 of this title or section 210 of the Departments of Commerce, Justice, and State, the Judiciary, and Related Agencies Appropriations Act, 1998 or section 2(f) of the Census of Agriculture Act of 1997—

(1) use the information furnished under the provisions of this title for any purpose other than the statistical purposes for which it is supplied; or

(2) **make any publication whereby the data furnished by any particular establishment or individual under this title can be identified**; or

(3) permit anyone other than the sworn officers and employees of the Department or bureau or agency thereof to examine the individual reports.

Existing Disclosure Avoidance Techniques

- Record-swapping
- Item imputation
- Whole household imputation
- Top- and bottom-coding

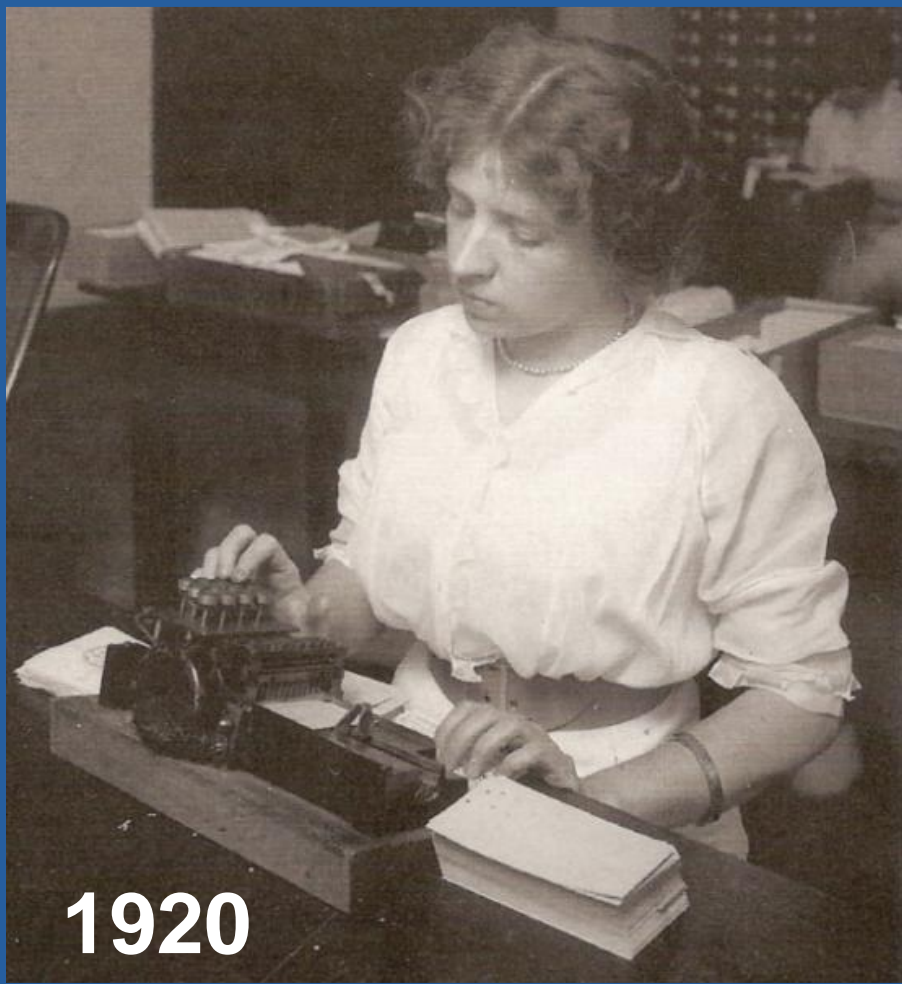
The New York Times

THE UPSHOT | To Reduce Privacy Risks, the Census Plans to Report Less Accurate Data

Luchsingers with another household living in another part of the state, who matched them on some key questions. This mechanism preserved their privacy, and kept summaries like the voting age population of the island correct, but also introduced some uncertainty into the data.



David Luchsinger and his wife, Debbie, out for a walk in their neighborhood, Liberty Island, in 2011, or one year after the 2010 census. Mustafah Abdulaziz for The New York Times



1920



2020

Images: https://www.census.gov/history/www/sights_sounds/photos/1920_photos.php and <https://flic.kr/p/YiyyBE>

Reconstruction & Re-identification

Simulated reconstruction/re-ID “attack” on 2010 Census

Publication	Released Counts (incl 0s)
PL94-171 Redistricting	2,771,998,263
Balance of Summary File 1	2,806,899,669
SF2	2,093,683,376
PUMS	30,874,554
Est. published statistics	7,703,455,862
Statistics/person	25

Source: <https://www.census.gov/content/dam/Census/newsroom/press-kits/2018/jsm/jsm-presentation-database-reconstruction.pdf>

Reconstruction & Re-identification

Simulated reconstruction/re-ID “attack” on 2010 Census

Reconstructed full database

- 46% complete

- 71%

Goal: New Disclosure Avoidance System (DAS)

Sources: https://www.census.gov/newsroom/blogs/random-samplings/2019/02/census_bureau_adopts.html and <http://blogs.cornell.edu/abowd/special-materials/245-2/>

Reconstruction & Re-identification

Simulated reconstruction/re-ID “attack” on 2010 Census

Reconstructed full database

- 46% complete re-ID
- 71% within +/- 1 year of age re-ID

But...

“more than half of these matches are incorrect, and an external attacker has no means of confirming them”

Sources: https://www.census.gov/newsroom/blogs/random-samplings/2019/02/census_bureau_adopts.html and <http://blogs.cornell.edu/abowd/special-materials/245-2/> and <http://users.hist.umn.edu/~ruggles/Articles/Privacy.pdf>



Image: <https://www.flickr.com/photos/domiriel/24832579995/>



Learn More About the Data

California data

- From Eddie Hunsinger at DOF

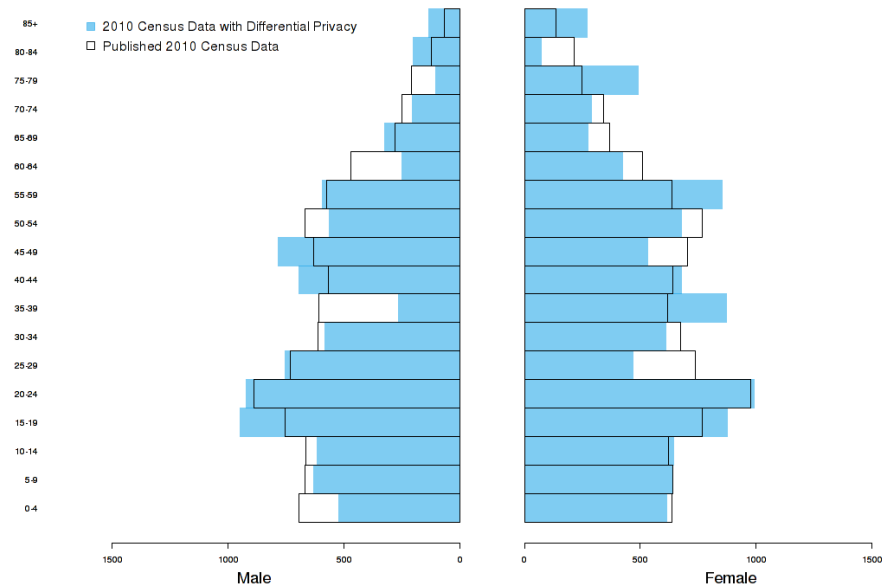
https://shiny.demog.berkeley.edu/eddieh/DP2010DemoData_CA/

Marina, CA

2010 population

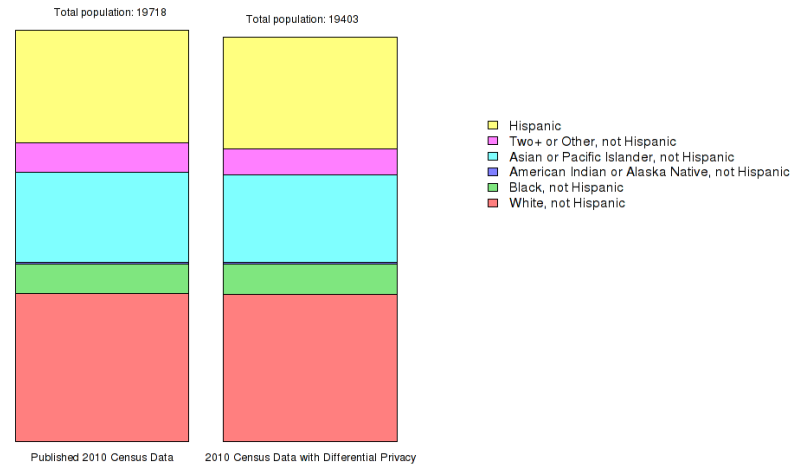
19,700

Population by Age and Sex, Marina city



Source: U.S. Census Bureau's 2010 Demonstration Data Products. Accessed via IPUMS NHGIS, University of Minnesota, November 2019.

Population by Race and Hispanic Origin Groupings, Marina city



Source: U.S. Census Bureau's 2010 Demonstration Data Products. Accessed via IPUMS NHGIS, University of Minnesota, November 2019.

Next Steps...

- National Academies workshop (webcast)
December 11th and 12th

http://sites.nationalacademies.org/DBASSE/CNSTAT/DBASSE_196518