CENTER FOR RENEWING AMERICA

Differential Privacy in the 2020 Census Explained

Topline: "Differential privacy" is a mathematical framework that introduces random noise into a data set to secure privacy by preventing the reverse engineering of characteristic data of individuals. Theoretically, this is done while maintaining accuracy; the more noise introduced, however, the less accurate the data becomes. For the first time in history, the U.S. Census Bureau used differential privacy on the data obtained during the 2020 Census. The downstream consequences of using differential privacy in the Census data set may have played a significant role in tilting the political scales favorably toward Democrats for apportionment and redistricting purposes, and other ramifications.

Key Problems: The usage of differential privacy in the census presents significant problems for policymakers and everyday citizens.

1. Structural Inaccuracy: Differential privacy's algorithmic processes preserve state-level totals while intentionally distorting characteristic data at each sub-level. No one can determine the extent of damage inflicted on both the geographic hierarchy (states down to block level) and the numerical accuracy within those hierarchies. Only a few Census Bureau staff have access to the actual data and algorithm, so no one can verify how falsified the data sets are, for example, a neighborhood, including both the characteristics of (e.g. age, race, sex, income, nationality) and actual number of people who live there.

2. Intentional Opacity: Local governments used to be able to analyze census data for errors, inform the Bureau of those errors, and obtain corrections. Because of differential privacy, local governments lack clarity on whether the errors are actual errors or the noise introduced by the differential privacy algorithm.

3. Citizenship Nullification: Even if the citizenship question is added to the Census, it will be impossible to ascertain the status of individuals so long as differential privacy is used. The algorithm will be able to mask characteristic data, including citizenship status, preventing the ability to determine total voter eligibility or how to draw maps to exclude those populations. Furthermore, the presence of illegal immigrants and those here legally increases the representational weight of large cities and suburbs, which draws political power away from rural, conservative areas and towards progressive populations in larger cities.

Conclusion: While protecting personal information is necessary and required, differential privacy intentionally imposes widespread instability in the numerical accuracy of Census data. Inaccurate data results in inaccurate redistricting processes by making it impossible to pinpoint only the population of citizens, and could result in bad apportionment data if citizenship characteristic data is scrambled as well. This results in unconstitutional outcomes.