one, very famous, court in East Texas seems to show that the problem is large, and with another court in West Texas recently sending its own "welcome to patent trolls," perhaps to compete with East Texas, the problems will only grow; see https://bit.ly/3H0yig3

George Neville-Neil,
Brooklyn, NY

Census Reconsiderations

Garfinkel, Abowd, and Martindale in "Understanding database reconstruction attacks on public data" (*Communications*, March 2019) highlight the dangers of database reconstruction attacks (DRAs). These U.S. Census Bureau methodologists illustrate DRAs using a small dataset of seven records, which they claim to be realistic because "the 2010 U.S. Census contained 1,539,183 census blocks in the 50 states and the District of Columbia with between one and seven residents." They protect the statistical output released for this dataset using cell suppression, more precisely "the rule of three," according to which cells sourced from fewer than three individuals are suppressed. The protected output is shown in Table 1 of their article, where (D) stands for a suppressed cell.

They show the original dataset can be uniquely reconstructed from their Table 1 by using a sophisticated SAT solver—in fact, reconstruction is feasible with much less apparatus: differencing entries in Table 1 and using simple logic and arithmetic reasoning success (see Muralidhar and Domingo-Ferrer, "Database reconstruction is very difficult in practice." UNECE/Eurostat Work Session on Stat. Data Confidentiality 2021.) Based on their analysis, they conclude protection of databases using suppression is inadequate. It turns out the reconstruction of Garfinkel et al. is based entirely on the incorrect application of (primary and complementary) suppression, so this conclusion is completely unwarranted:

▸ Under the rule of three, in addition to suppressing cells sourced from fewer than three individuals (primary suppressions), cells whose value would allow deriving the primary suppressions via subtraction must also be suppressed (complementary suppressions). See U.S. Census Bureau documents "American Community Survey: Data Suppression" and "Disclosure Avoidance and the Census." This requires cells in line 4A of Table 1 to be suppressed. Failing to suppress 4A allows differencing 2C and 4A to obtain 4B (the values for the only Black American male individual in the dataset).

▸ Releasing the median Age discloses the age of a single individual (when the group size is odd) or the average age of two individuals (when the group size is even). Hence, medians cannot be released under the rule of three.

▸ In contrast, there is no need to suppress the count and the mean in line 3A, since they affect three individuals (Single Adults) and do not need to be eliminated as complementary suppressions.

Without 4A, with 3A and without medians, there are thousands of datasets compatible with the remaining counts and means. No unique gender assignment to the seven individuals is possible. Also, uniquely reconstructing Age values without the medians is infeasible, as there are many Age assignments that fit the extant means and counts.

Hence, correct application of cell suppression can perfectly protect the example dataset. What the authors have proven, albeit unintentionally, is that simple disclosure prevention techniques, properly applied, are effective in preventing reconstruction of even very small (hypothetical) datasets created by senior U.S. Census Bureau methodologists to show the very opposite. Database reconstruction may not be so easy after all and this must be taken into account in the ongoing process of protecting the 2020 Census outputs.

Krishnamurty Muralidhar,
Norman, OK, and
Josep Domingo-Ferrer, Catalonia

Authors' response:

We thank Muralidhar and Domingo-Ferrer for their contribution to the example, and we note that the Census Bureau chose to use differential privacy rather than the alternative, massive cell suppression for the 2020 Census.

Simson Garfinkel, Arlington, VA,
John M. Abowd, Washington, D.C., and
Christian Martindale, Durham, NC

Communications welcomes your opinion. To submit a Letter to the Editor, please limit your comments to 500 words or less, and send to letters@cacm.acm.org

© 2022 ACM 0001-0782/22/06 $15.00

Coming Next Month in COMMUNICATIONS

Toward Verified Artificial Intelligence

***Communications of the ACM* Community A Letter from New Editor-in-Chief James Larus**

Expressive Querying for Accelerating Visual Analytics

Surveillance Too Cheap to Meter

The Keys to the Kingdom

Past, Present, and Future Language Models

Algorithms with Predictions

When Satisfiability Solving Meets Symbolic Computation

On Sampled Metrics for Item Recommendation

Plus, the latest news about building a practical quantum computer, how brain implants work, and blocking surveillance with makeup.