The 2020 Census Suggests That People Live Underwater. There's a Reason.

Technology advances forced the Census Bureau to use sweeping measures to ensure privacy for respondents. The ensuing debate goes to the heart of what a census is.

By Michael Wines

April 21, 2022

WASHINGTON — Census Block 1002 in downtown Chicago is wedged between Michigan and Wabash Avenues, a glitzy Trump-branded hotel and a promenade of cafes and bars. According to the 2020 census, 14 people live there — 13 adults and one child.

Also according to the 2020 census, they live underwater. Because the block consists entirely of a 700-foot bend in the Chicago River.

If that sounds impossible, well, it is. The Census Bureau itself says the numbers for Block 1002 and tens of thousands of others are unreliable and should be ignored. And it should know: The bureau's own computers moved those people there so they could not be traced to their real residences, all part of a sweeping new effort to preserve their privacy.

That paradox is the crux of a debate rocking the Census Bureau. On the one hand, federal law mandates that census records remain private for 72 years. That guarantee has been crucial to persuading many people, including noncitizens and those from racial and ethnic minority groups, to voluntarily turn over personal information.

On the other, thousands of entities — local governments, businesses, advocacy groups and more — have relied on the bureau's goal of counting "every person, only once and in the right place" to inform countless demographic decisions, from drawing political maps to planning disaster response to placing bus stops.

The 2020 census sunders that assumption. Now the bureau is saying that its legal mandate to shield census respondents' identities means that some data from the smallest geographic areas it measures — census blocks, not to be confused with city blocks — must be looked at askance, or even disregarded.

And consumers of that data are unhappy.

The area within Block 1012 on the southeast side of Chicago is said to have one home with 86 people living in it. Jamie Kelter Davis for The New York Times

"We understand that we need to protect individual privacy, and it's important for the bureau to do that," David Van Riper, an official of the University of Minnesota's Institute for Social Research and Data Innovation, wrote in an email. "But in my opinion, producing low quality data to achieve privacy protection defeats the purpose of the decennial census."

The Census Bureau says its privacy mechanisms are designed to move people only to census blocks with at least one residence. It suggested that vacant lots and rivers shown as hosting Chicagoans could at one point have had a residence, such as a houseboat or a since-demolished home, or that a coding error mistakenly marked such blocks as having one.

At issue is a mathematical concept called differential privacy that the bureau is using for the first time to mask data in the 2020 census. Many consumers of census data say it not only produces nonsensical results like those in Block 1002, but also could curtail the publication on privacy grounds of basic information they rely on.

They are also miffed by its implementation. Most major changes to the census are tested for up to a decade. Differential privacy has been put into use in a few years, and data releases already snarled by the pandemic have been delayed further by privacy tweaks.

Census officials call those concerns exaggerated. They have mounted an urgent effort to explain the change and to adjust their privacy machinery to address complaints.

But at the same time, they say the sweeping changes that differential privacy brings are not only justified but also unavoidable given the privacy threat, confusing or not.

"Yes, the block-level data have those impossible or improbable situations," Michael B. Hawes, the senior adviser for data access and privacy at the bureau, said in an interview. "That's by design. You could think of it as a feature, not a bug."

And that is the point. To the career data nerds who are the census's stewards, uncertainty is a statistical fact of life. To their customers, the images of census blocks with houses but no people, people but no houses, and even people living underwater have proved indelible, as if the curtain had been pulled back on a demographic Great Oz.

"They burst the illusion — an illusion that kept everybody thinking that these point estimates were always pretty good or the best possible," said danah boyd, a technology scholar who lowercases her name and has co-written a study of the privacy debate. "Census Bureau executives have known for decades that these small-area data had all sorts of problems."

The difference now, she said, is that everyone else knows it, too.

According to the 2020 census, there are three houses in Block 3002 with 13 people total, all 18 or younger. The area consists of: an empty lot, an auto repair shop and carwash, a law office and an empty storefront. Jamie Kelter Davis for The New York Times

Some history: Census blocks — there are 8,132,968 of them — began more than a century ago to help cities better measure their populations. Many are true city blocks, but others are larger and irregularly shaped, especially in suburban and rural areas.

For decades, the Census Bureau withheld most block data for privacy reasons, but relented as demand for hyperlocal data became insatiable. A turning point arrived in 1990: Census blocks expanded nationwide, and the census began asking detailed questions about race and ethnicity.

That added detail allowed outsiders to reverse-engineer census statistics to identify specific respondents — in, say, a census block with one Asian American single mother. The bureau covered those tracks by exchanging such easily identifiable respondents between census blocks, a practice called swapping.
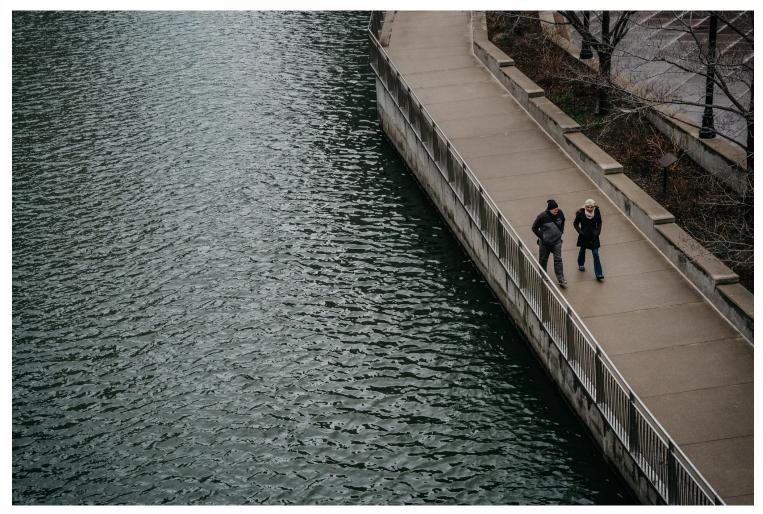
But by the 2010 census, the explosions of computing power and commercial data had barreled through that guardrail. In one analysis, the bureau found that 17 percent of the nation's population could be reconstructed in detail — revealing age, race, sex, household status and so on — by merging census data with even middling databases containing information like names and addresses.

Today, "any undergraduate computer science student could do a reconstruction like this," Mr. Hawes said.

The solution for the 2020 census, differential privacy, which is also used by companies like Apple and Google, applies computer algorithms to the entire body of census data rather than altering individual blocks. The resulting statistics have "noise" — computer-generated inaccuracies — in small areas like census blocks. But the inaccuracies fade when the blocks are melded together into one coherent whole.

The change brings the Census Bureau distinct advantages. While swapping is a crude way of masking data, differential privacy algorithms can be tuned to meet precise confidentiality needs. Moreover, the bureau can now tell data users roughly how much noise it has generated.

In data scientists' eyes, census block statistics have always been inaccurate; it's just that most users didn't know it. By that view, differential privacy makes census numbers more accurate and transparent — not less.

Because of privacy protections built into 2020 census data when it was processed, Block 1002 has a vacant home that houses 14 people. The area specified is actually a portion of the Chicago River where sightseeing boats dock and load tourists. Jamie Kelter Davis for The New York Times

Outsiders see things differently. A Cornell University analysis of the most recent data release in New York State concluded that one in eight census blocks was a statistical outlier, including one in 20 with houses but no people, one in 50 with people but no houses, and one in 100 with only people under 18.

Such anomalies will dwindle as algorithms are refined and new sets of data are released. Some experts say they still fear the numbers will be unusable.

Some civil rights advocates worry that noisy block data will complicate drawing political boundaries under the Voting Rights Act's provisions for minority representation, though others see no problem. Some experts who draw political maps say they have struggled with the new data.

Block anomalies posed no problem in larger districts, but they "caused real havoc in city council wards," said Kimball Brace, whose firm, Election Data Services, serves mostly Democratic clients.

Critics also fear that the bureau could limit publishing some important statistics only at the level of larger areas like counties, because census block numbers are unreliable.

Mr. Hawes, the bureau's privacy official, said that could happen. But because differential privacy restrictions are adjustable, "we're adding in some more of the lower-level geographic tables based on the feedback we've gotten," he said.

Such openness is a major shift in an agency where privacy is a mantra. The shift to differential privacy might be less rocky if the bureau better answered a basic question: "Since there's so much commercially available data out there, why do we care about protecting census data?" said Jae June Lee, a data scientist at Georgetown University who is advising civil rights groups on the change.

The answer, said Cynthia Dwork, a Harvard University computer scientist and one of four inventors of differential privacy, is that a new era of runaway technology and rising intolerance has made privacy constraints more important than ever.

Loosen them, she said, and census data could reveal subsidized housing tenants who take in unauthorized boarders to make ends meet. Or the data could be used by hate groups and the politicians who echo them to target people who do not conform to their preferences.

"Imagine a kind of weaponization, one where somebody decides to make a list of all the gay households across the country," she said. "I expect there will be people who would write the software to do that."

Michael Wines writes about voting and other election-related issues. Since joining The Times in 1988, he has covered the Justice Department, the White House, Congress, Russia, southern Africa, China and various other topics. More about Michael Wines