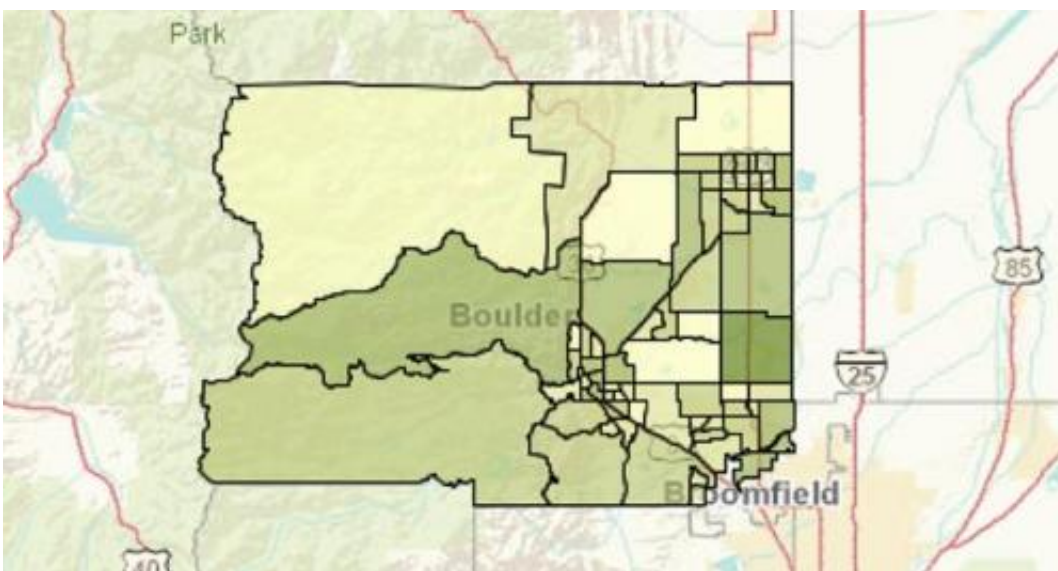


Professor works to shrink error margins in US census data

April 24 2013



A person searching through the massive expanse of data collected by the U.S. Census Bureau in search of details about a specific neighborhood may increasingly find statistics with colossal margins of error, such as an average income of \$50,000 plus or minus \$50,000.

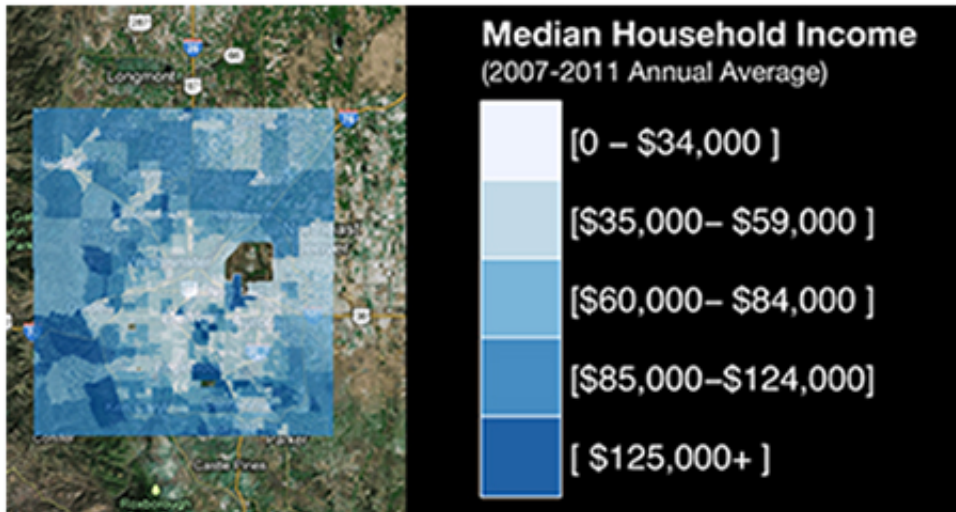
A geographer at the University of Colorado Boulder, one of eight nodes of the National Science Foundation's newly created [Census](#) Research Network, has been granted a five-year \$1.4 million grant to see if he can change that.

Assistant Professor Seth Spielman, director of the CU-Boulder Census Research Node, said the margin of error for neighborhood-level information collected by the U.S. [Census Bureau](#) hasn't always been so dismal. The quality of the data from the American Community Survey—the portion of the census that asks residents about their age, household makeup, education levels and income, among other facts—has been limited by the Census Bureau's budget and rigid census-reporting boundaries that have not changed in more than half a century, Spielman said.

The erosion of high-quality data affects a range of social services, since funding for those programs is often linked to [Census Bureau data](#) for metrics such as poverty.

Spielman thinks the key to reducing the margin of error lies in redefining how the boundaries are drawn around [neighborhoods](#) so that more similar people are grouped together. But to do that, he needs to dive into the highly secured data that actually bundles together information for individuals, including where they live, their race, how much they make and how many children they have.

"We want to understand what neighborhoods look like, and we think that by using individual-level data and [computer algorithms](#) we can redraw neighborhoods and get a more precise picture," Spielman said.



A map showing median household income based on data from the American Community Survey.

[Census data](#) on communities currently is available for small regions known as census tracts. When these groupings were originally made, in the 1960s, they were designed by local committees to delineate similar sections of cities so that individual neighborhoods could be studied. But as the decades have rolled by, the makeup of many of the census tracts has changed, and now some tracts encompass parts of multiple, widely varying neighborhoods. The disparity within the tracts, and the fact that fewer people are now being sampled in each tract, has inflated the margins of error.

Spielman is now using CU-Boulder's [Janus supercomputer](#) to test an algorithm that will allow for computer-assisted redrawing of neighborhood lines in the United States. Spielman doesn't propose that the old census tract lines be discarded, since it's important that tracts can continue to be compared over time. But the new neighborhood lines

might give people a more reliable way to understand what's going on inside a city.

The algorithm is still a work in progress. Spielman and David Folch, a postdoctoral researcher in CU-Boulder's geography department, are using the supercomputer to comb the ocean of government data for areas in which people are the most similar. Those similarities could include everything from race to family size to whether an individual commutes by bike or is a veteran.

"However we group things together, the best grouping is the grouping that results in a neighborhood that has the highest level of similarity," Spielman said. "For all the variables, we just want to maximize how similar the neighborhoods are."

Once the algorithm is finished, Spielman will apply it to individual-level data stored on secure servers in Washington, D.C. The resulting neighborhoods, however they may look, would not provide individual-level data to the public.

Provided by University of Colorado at Boulder

Citation: Professor works to shrink error margins in US census data (2013, April 24) retrieved 20 April 2024 from <https://phys.org/news/2013-04-professor-error-margins-census.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.