

THE WALL STREET JOURNAL.

OUR NAMES SHOW DIVERSITY GROWING—AND SHRINKING

By Paul Overberg
December 16, 2016

It takes fewer last names now than a decade ago to tally most of the U.S. population, a surprising drop for a country awash in diversity.

The top 25 names accounted for 8.7% of the population, up from 8.6% in 2000, according to newly released Census Bureau list of last names from Census 2010. The top 100 names covered 17.3% of the population, up from 17.1% in 2000.

Put another way, it took just 239 names to cover a quarter of the people in 2010, down from 244 names in 2000. And it took just 2,116 names to cover half of the population, down from 2,182 names in 2000.

At the same time, the list's rankings and racial breakdown for each last name shows the country continues to diversify. "Martinez" moved into the top 10 and "Perez" entered the top 25. In percentage terms, top gainers since Census 2000 were "Zhang" (111%), "Li" (93%), "Ali" (66%) and "Liu" (64%), and "Khan," "Vazquez" and "Wang" (63%).

The most common name, "Smith," was used by 2.44 million people, of whom 71% were white, down from 73% in 2000. About 23% of Smiths were black, 2% each were Hispanic or multiracial and 1% each were Asian or American Indian.

In the top 100 names, 33 had nonwhite majorities, up from 23 names in 2000. In the top 500 names, 138 had nonwhite majorities, up from 104 in 2000.

The crosscurrents occurred partly because of the growing ranks of Hispanic and Asian immigrants and their children. Their surnames lists are less diverse than the overall list, which reflects centuries of immigration from hundreds of nations and cultures. For the most common names where a majority of users were Hispanic, 86% of those names saw the Hispanic share of their users increase since 2000.

"There is some dispersion, but the consistency is still there," said William Frey, a Brookings Institution demographer. He also noted strong separation between names used by whites and blacks on one hand and Hispanics or Asians on the other. For 80% of the top 1,000 names, whites and blacks together make up 95% of the users. For another 10% of those names, Hispanics were 95% of the users. And for 2% of those names, Asians make up 95% of users.

In addition to guiding genealogists, the list has become something of a statistical Rosetta stone for businesses and government agencies. A key method, first described in 2009, uses the list, matched with a person's address and standard census data about a neighborhood, to produce the probability that the person belongs to a given race. Marc Elliott, a Rand Corp. statistician, developed the algorithm to help health-care firms and agencies improve care for minority patients. The Consumer Financial Protection Bureau has used this method to distribute funds from settlements with lenders accused of violating fair-lending laws.

And this year the federal government began using this method to calculate and publish race-specific data on the quality of care provided to seniors enrolled in Medicare Advantage plans.

In a separate approach, pollsters seeking small groups mixed in the population—such as minorities or people living in certain types of households—use the surname list to focus their sampling. This improves efficiency and lowers cost.

Elliott said the method works fairly well when race is unavailable and important to measure, but the address and last name of a person is available. He described one example where an agency wanted to study stresses faced by newly married low-income Hispanic couples in Los Angeles County. Researchers used census data and addresses from thousands of marriage licenses to focus on couples in likely neighborhoods. Then they used the surname list to target couples where both partners were very likely to be Hispanic. The result produced a sample where 92% of couples who were eventually contacted fit the criteria, cutting necessary screening almost in half.

The Census 2010 name list included 162,254 names covering 295 million people. The full list included 6.3 million names, including almost four million that were unique. The edited list released Thursday does not include any other information and omits names used by fewer than 100 people. The Census Bureau guarantees respondents that it will keep personally identifiable census responses confidential for 72 years.