PPIC | PUBLIC POLICY INSTITUTE OF CALIFORNIA

2020 Census Maps

California's Hard-to-Count Communities

Technical Appendix

CONTENTS

Tess Thorman and Sarah Bohn

Supported with funding from the California Community Foundation, the California Endowment, the California Health Care Foundation, and Silicon Valley Community Foundation

Appendix A. Variable Methodology and Notes

The interactive maps *2020 Census Maps: California's Hard-to-Count Communities* visualize vulnerable and hard-to-count communities at many levels of geography (counties, UC congressional districts, state assembly and senate districts, and census tracts). Here, we define the metrics used in the maps to visualize these communities and describe the data sources and methodologies for the creation of each metric.

Geographies and Suppression

Census tracts are small, relatively permanent geographies that the Census Bureau uses as part of its data collection and reporting. Tracts are the key building blocks of the maps; data for each variable begin at the tract level, and are aggregated to the county level or mapped onto legislative districts. When a tract falls into more than one legislative district, it is weighted according to the share of population of the overlapping tract-district area, using MABLE/Geocorr allocation factors. Exceptions to these methods are discussed below.

Tracts generally contain around 4,000 residents; the population in a tract commonly ranges between 1,200 and 8,000 people.[1] In California, the median population of a tract is about 4,500. For statistical accuracy, we suppress tract-level estimates when the estimated population falls below 250. This results in the omission of 28 out of 8,011 populated tracts, many of which are primarily composed of non-residential features such as parks and airports.

Mapped Variables

Likelihood of No Response

The Census Bureau identifies tracts whose residents are less likely to respond to initial census mailings by creating a measure called the Low Response Score. The measure uses a model of the relationship between current demographic and place characteristics (including, for example, age, poverty status, population density, and housing prices) and the mail nonresponse rate for the 2010 Census. It is best interpreted as the estimated share of households in a tract that would not self-respond to the census;[2] the score ranges in California from 2.0 percent, where almost no households would self-respond to the mail-in census, to 45.2 percent, where almost half would. In these maps, we show quintiles of the score across California tracts.

Demographic and place data for the Low Response Score come from the census tract–level 2012–16 American Community Survey (ACS) Summary Files (found in the 2018 Census Bureau Planning Database).[3] As a tract-level prediction, the Low Response Score cannot be aggregated to larger geographies, and we therefore show it at the tract level throughout these maps. The Census Bureau does not calculate a Low Response Score for tracts with zero housing units, tracts with less than the minimum number of qualifying addresses, and tracts whose geographic boundaries changed between 2010 and 2014. We further suppress scores for tracts with populations under 250 people, consistent with the other mapped metrics.

As an official metric of "hardness to count," the Low Response Score is a useful metric for visualizing the potential vulnerability of a community to omission from the census. The Census Bureau's Low Response Score,

---

[1] US Census Bureau. December 2012. *Geographic Terms and Concepts – Census Tract*.
[2] US Census Bureau. January 2018. *Response Outreach Area Mapper*.
For complete methodology, see Erdman, C. and N. Bates. 2017. The Low Response Score (LRS): "A Metric to Locate, Predict, and Manage Hard-to-Survey Populations." *Public Opinion Quarterly* 81(1): 144–156.
[3] US Census Bureau. June 20, 2018. *2018 Planning Database*.

however, explains just 55 percent of the variation in nonresponse rates to initial distribution of forms, and census tract populations may be significantly different in 2020 than they were from 2012 to 2016. Low Response Scores therefore cannot pinpoint all tracts that will be hard to count in the 2020 Census. These scores also do not indicate the underlying reasons a community might be hard to count.

African Americans, Latinos, or Native Americans

The percentage of a population that is African American, Latino/a, or Native American is calculated using 2012–16 ACS Summary File data. Given that these groups have historically been undercounted in the census, places where these subgroups make up larger shares of the population may be less likely to be counted accurately in 2020. We visualize them using a single metric for statistical purposes; pooling the three groups allows us to estimate populations in places where there are few individuals from any single undercounted race or ethnicity. The variable ranges from 0 to 100 percent across tracts.

These maps show quintiles of the shares of African Americans, Latinos, or Native Americans across places. For counties, we aggregate population counts from tracts to the county level; for legislative districts, we map population counts onto districts and recalculate percentages at those higher geographies. Tracts that cross district boundaries are split using population-weighted MABLE/Geocorr14 allocation factors. In addition to our practice of suppressing data from tracts with estimated populations of less than 250, we mark with an asterisk (*) tracts where the estimated population is less than 500, or the reported margin of error on the number of minority residents is greater than or equal to the estimated number of minority residents.

Noncitizens

Amidst concerns about privacy, fears of deportation, and the likely addition of a question about citizenship to the decennial census, immigrants and their families are probably less likely to respond to the 2020 Census. Places where noncitizens make up a larger share of the population may therefore be less likely to be counted accurately. The percentage of noncitizens measure uses 2012–16 ACS Summary File data, and ranges from 0 to 63.9 percent across tracts.

These maps show quintiles of the percentage of noncitizens across places. For counties, we aggregate population counts from tracts to the county level; for legislative districts, we map population counts onto districts and recalculate percentages at those higher geographies. Tracts that cross district boundaries are split using population-weighted MABLE/Geocorr14 allocation factors. In addition to our practice of suppressing data from tracts with estimated populations of less than 250, we mark with an asterisk (*) tracts where the estimated population is less than 500, or the reported margin of error on the number of noncitizens is greater than or equal to the estimated number of noncitizen residents.

Young Children

The percentage of a population that is under five years old is calculated using 2012–16 ACS Summary File data. Given that young children have historically been undercounted in the census, places where they make up a larger share of the population may be less likely to be counted accurately in 2020. Their share of the population ranges from 0 to 29.7 percent across tracts.

These maps show quintiles of the percentage of young children across places. For counties, we aggregate population counts from tracts to the county level; for legislative districts, we map population counts onto districts and recalculate percentages at those higher geographies. Tracts that cross district boundaries are split using population-weighted MABLE/Geocorr14 allocation factors. In addition to our practice of suppressing data from tracts with estimated populations of less than 250, we mark with an asterisk (*) tracts where the estimated

population is less than 500, or the reported margin of error on the number of young children is greater than the estimated number of young children.

Hard-to-Count Housing Score

The hard-to-count housing score identifies places that might have challenges counting residents based on the likelihood of their living in housing arrangements historically associated with being undercounted in the census.

Specifically, the score uses 2012–16 American Community Survey Summary File data at the tract level to replicate metrics in the 2016 Planning Database: share of occupied housing units that are rented, share of occupied rented housing units with more than one person per room (a standard measure for overcrowding used by the US Department of Housing and Urban Development), and share of occupied housing units that are mobile homes. We combine these into a single, composite metric. For each of these housing statistics, we divide census tracts into percentiles, assigning them a rank of 1 through 10. We then find the sum of the three ranks, and the quintile in which the sum falls, statewide, is the hard-to-count housing score that is used in these maps. Because the scores have no direct interpretation except as ranks, the component aspects—share of rentals, overcrowding, and mobile homes—are included in the tooltips of the maps.

Other indices, such as the Census Bureau's Low Response Score and Kissam's "Bad Master Address File" score, use similar housing characteristics in combination with other demographic and economic characteristics to identify hard-to-count tracts.[4] Our hard-to-count housing score draws on these examples, but unlike these models, our score is intended to help identify areas that might be particularly hard to count *due to housing circumstances alone*. Accordingly, our measure uses only three core, housing-related metrics and does not rely on any demographic data on population characteristics.

High-Speed Internet Connections

The ratio between the number of residential high-speed internet connections and the number of households in a tract is calculated by the Federal Communications Commission (FCC) as a measure of internet access. Given that the Census Bureau hopes to collect a majority of responses online in 2020, residents in areas with lower ratios of high-speed internet connections to households may have trouble being counted based on difficulties accessing the census online.

The FCC defines high-speed connections as at least 10 Mbps downstream and at least 1 Mbps upstream, and estimates number of residential connections using total connections and the percentage of residential connections that are reported by broadband providers. The number of households is a separate estimate, taken from 2011–15 ACS five-year estimates. This leads to some discrepancies: some tracts, for example, seem have more connections than households.[5] The FCC therefore produces ratios of connections to 1,000 households, rather than percentages of households, and allows for some ratios to exceed 1,000 per 1,000 households. It publicly releases tract-level ranges of the ratios (e.g., 600–800, or 800+ connections per 1,000 households); these are what are shown in the tract-level view of our interactive maps. At larger geographies, we show the population-weighted average range. In cases where tracts cross district boundaries, we split them using the MABLE/Geocorr14 allocation factors.

---

[4] Data gathered by Community Connect Labs in San Jose and Fresno show that some tracts have many housing units that will not receive census mail-in forms unless they are added to the Census Bureau's Master Address File (MAF). Using as an initial guide Kissam's "Bad MAF" score, a metric to identify tracts with many of these housing units, canvassers found 2–6 percent more units in tracts with high scores. The Bad MAF Score is a composite of weighted housing and population statistics, and therefore also reflects demographic and economic characteristics that may contribute to a tract's likelihood of having many hard-to-count homes. We map some of these characteristics in other categories in our interactive maps. Kissam, Ed. March 24, 2018. *An Effective Strategy To Reduce Census Undercount: Results from California Pilots of Community-Based Address Canvassing*. WKF Giving Fund.

[5] Industry Analysis and Technology Division, Wireline Competition Bureau. February 2018. *Internet Access Services: Status as of December 31, 2016*.

We use the most recent available data, current as of December 2016 (released April 2018).[6] The FCC does not calculate ratios for tracts that exclusively cover water, and assumes that tracts with nonzero populations but zero households have institutional, but not residential, connections. We suppress tract estimates with no occupied housing units in the 2012–16 ACS, in addition to those with fewer than 250 estimated residents.

Residential fixed high-speed internet connections are only one means of understanding geographic variation in internet access, and may not give a full picture of how internet access will impact the 2020 Census. It may be that people without high-speed connections at home have easy access to internet at work, through community institutions, or through smartphones and other handheld devices. Census canvassers following up with nonresponsive households may be less able to rely on internet access in some parts of the state than others, depending on service coverage. We believe that the ratio mapped here serves as an effective proxy for these conditions, until more detailed, recently added ACS questions on internet access have been included for enough years to provide tract-level estimates.

[6] Federal Communications Commission, Wireline Competition Bureau. April 2, 2018. *Residential Fixed Internet Access Service Connections per 1000 Households by Census Tract*.

PPIC | PUBLIC POLICY INSTITUTE OF CALIFORNIA

The Public Policy Institute of California is dedicated to informing and improving public policy in California through independent, objective, nonpartisan research.

Public Policy Institute of California
500 Washington Street, Suite 600
San Francisco, CA 94111
T: 415.291.4400
F: 415.291.4401
PPIC.ORG

PPIC Sacramento Center
Senator Office Building
1121 L Street, Suite 801
Sacramento, CA 95814
T: 916.440.1120
F: 916.440.1121