Staff Block Estimates - Methodology

Introduction

This document describes the methodology used by Federal Communications Commission staff to create intercensal block-level estimates of the number of housing units, households and population. The first section provides an overview of the methodology; the second section provides a description of the data sources; and the third section describes the methodology in more detail, providing illustrative examples.

The results of this approach are necessarily estimates. We attempt to minimize estimation errors by basing the estimates on an authoritative data source: U.S. Census Bureau county-level population estimates. And, as described below, we try to avoid over-stating the footprint of development – areas with housing units, households and/or population (e.g., to avoid creating an area unlikely to ever get broadband deployed but which we count as unserved). As with any such estimate, the larger the area of interest, the less these estimation errors matter.

Overview

The end product of this process is a table (in .csv format) with the number of housing units, households and population for each census block in the country.

The process begins with block-level counts of housing units, households and population from the decennial U.S. Census. Staff then take county-level estimates from the U.S. Census Bureau for HU and POP for mid-year and "distribute" any increases or decreases relative to the previous year's estimates along eligible roads as described below. This provides block-level estimates that, when summed to the county level, align with U.S. Census county-level estimates. Staff then calculate the HH count based on the HU count in each block.

There are five steps to determine block-level estimates

1. Determine areas ineligible for new development. These areas are meant to capture, for example, parkland where new development is much less likely
2. Determine road segments (TIGER edges) eligible for new development (i.e., find the roads along which new homes will be built and people will live)
3. Identify census blocks that contain with new roads indicative of new development; and add housing units and population along these new roads
4. Add or remove housing units and population to blocks to align with county-level estimates
5. Calculate the number of households in each block based on the number of housing units in the block

Note that the use of U.S. Census Bureau county-level estimates and the methods for adding and removing counts of housing units is consistent with the approach taken in developing cost models for high-cost Universal Service.[1]

Sources

- OpenStreetMap (OSM)
    - Downloads for the 50 states and DC
    - Downloads for the five territories: American Samoa, Guam, the Northern Mariana Islands, Puerto Rico and the Virgin Islands

[1] See, for example, Section Two of the cost model methodology documentation.

- U.S. Census Bureau TIGER shapefiles
    - Census block geometries: 2020 census blocks
    - Road segments: Census all lines (edges) from the current year.
- Additional U.S. Census data
    - 2020 block housing unit, household and population counts: Decennial Census P.L. 94-171 Redistricting Data
    - Annual population and housing unit estimates: County Population Totals
    - International Database (IDB)

Sources above have been updated to reflect the current benchmark datasets for 2020-2030. Commission staff used PostGIS (GIS-enabled Postgres) for the analysis; see, e.g., information about OpenGeo Suite. Commission staff read Census shapefiles into postgis using shp2pgsql; and read OSM data, in .osm.pbf format, into PostGIS using osm2pgsql.

All data sources are re-projected onto SRID 900913 to allow for geospatial analysis in a single SRID with a unit of meters.

Detailed description

Step One: Determine Areas Ineligible for New Development

While Commission staff are not aware of a data set that captures areas that are off-limits to new development, there are some areas where such development is less likely. National Parks are one clear example, but there are others (e.g., airports, wetlands, nature reserves, hospital campuses, cemeteries). Other areas, like university and college campuses or military installations, may see growth in population, but are less likely to see growth in housing units and households (with people living in group quarters like dormitories).

OpenStreetMap data include a rich set of descriptions to identify such areas. Staff used the following flags and values to identify areas that we assume will not have new development in this process:

a. leisure IN ('park','nature_reserve')
b. boundary = 'national park'
c. landuse IN ('forest','cemetery','golf course', 'landfill', 'military')
d. military IS NOT NULL
e. water IS NOT NULL or wetland IS NOT NULL or waterway='riverbank'
f. tourism = 'zoo'
g. aeroway = 'aerodrome' (i.e., airports)
h. amenity IN ('university','college','hospital')
i. natural IN ('water','wetland')

Staff create a table for each of the 56 states, territories and the District of Columbia with such areas that are treated as ineligible for development. Since OSM boundaries do not necessarily align well with U.S. Census shapes, staff buffer these polygons derived from OSM data to avoid situations such as placing a new home in a park by selecting a road that forms a park boundary.

The state-by-state data from OSM do not necessarily include all bodies of water, however. Therefore, we also exclude the area of any census block with no land area. This should exclude, for example, roads that cross bodies of water from having new housing units added to them.[2]

Staff append data for polygons for blocks without land area to the buffered "park" data to capture all the areas ineligible for new development.

There are examples of the impact of this approach in the next section.

Step Two: Determine Road Segments Eligible for New Development

Commission staff start with the Edges (all lines) file from the U.S. Census TIGER, and extract all roads (using roadflg='Y') for each county. Next, we determine whether each segment of road is eligible for placement of housing units and population along two dimensions: (a) road type, and (b) whether the road falls into an area determined to be ineligible above.

For road type, we consider two types as eligible based on the U.S. Census classification: (a) Secondary Road; and (b) Local Neighborhood Road, Rural Road, City Street.[3] This represents the majority of roads in most areas, and excludes, for example, Primary Roads, Ramps, Alleys and Private Roads.[4]

Next, our code identifies road segments from the Edges file that fall largely outside the ineligible area defined above. Excluding only roads that fall entirely in the excluded areas would mean that a road with only a tiny bit of its length outside a park could be selected for placement of a new housing unit or population which would likely fall inside the park. On the other hand, excluding any road that touched any part of a park would mean excluding too many roads. The method in place excludes any road with more than 25% of its length inside the area deemed ineligible.

See Figures 1 and 2 below for examples of the areas excluded using this approach.

[2] There are some census blocks in the middle of bodies of water that the US Census Bureau classifies as having land area, perhaps reflecting the area associated with roads; see, e.g., census blocks 121030254051000 and 121030268091033.

[3] Technically MTFCC of S1200 and S1400.

[4] MTFCCs of S1100, S1630, S1730 and S1740, respectively

Figure 1 Area north of Bethany Beach, Delaware showing roads excluded from eligibility for placement of new housing units and population in red. These ineligible roads are primarily in Delaware Seashore State Park and a National Guard facility. Base map © OpenStreetMap

Figure 2 San Francisco, California showing roads excluded from eligibility for placement of new housing units and population. Roads in red are excluded based on their location – for example, near t the embarcedaro, San Francisco VA hospital, major interstates, public alleys and San Francisco State University

Step Three: Place New Development Along New Roads

Identify blocks that contain new roads indicative of new development

It is difficult to identify newly constructed roads in the U.S. Census Edges file for a few reasons. First, while each road segment has a unique TLID ("Permanent edge ID"), Census assigns a new TLID when a road segment is redefined. For example in the 2010 Edge data the segment of K Street NW in Washington, DC with addresses from 52 (lfromadd) to 99 (rtoadd) has TLID=76235443; however in the 2015 Edge data that road segment is replaced by two others (TLID = 639337214 runs from 62 (lfromadd) to 81 (rtoadd) while TLID=639337215 extends from 76 (lfromadd) to 99 (rtoadd)).

Second, there may be new road segments in a later year that correct an omission in a prior year. See, for example, Figure 3 below, which shows a portion of K Street NW in Washington, DC overlaid with a satellite image of the area. Lines in orange represent roads in the 2010 TIGER Edge data, while blue lines represent roads in the 2015 TIGER Edge data (roads that appear in both will appear as blue). Between 2010 and 2015, the TIGER data corrected the data along K Street NW to reflect it as a divided road; but there was not new road added.

Figure 3 Area along K Street NW in Washington DC showing road lengths in 2015 (in blue) compared to 2010 (in orange). Imagery © 2016, Commonwealth of Virginia, DigitalGlobe, District of Columbia (DC GIS), Sanborn, U.S. Geological Survey, USDA Farm Service Agency

Furthermore, Census can re-classify a road that had been an ineligible road type in an earlier year as an eligible road type in a later year (see Step Two above for discussion of road types). In short, it is difficult to identify new roads from the Edge data alone.

Instead, staff calculate the amount of roadway in each census block and then determine the growth in eligible road types (secondary roads and local roads, as described in Step Two above) in each block over time.

The first step is to find the intersection between decennial block boundaries and current road segments for each county. Staff buffered decennial block boundaries by 0.1m to avoid situations where the road segments happen to cross the boundary. The next step is to find the intersection between those same blocks and the contemporaneous road segments for each county. Note that there can be changes to the definition of counties and their boundaries,[5] so it is important to align the areas over time.

Next, we calculate the total road length in each block in the end year and the start year. Finally, we test blocks to determine where the change in road length is more likely due to changes in the data set (e.g., corrections) than new road. The first test rejects blocks where changes in road length are small – 10m or less. The second test rejects blocks where total road length increases by less than half of the change in eligible road length – this excludes blocks where most of the increase in road length comes from reclassification of existing roads to secondary or neighborhood roads. The third test rejects blocks where the total length of new road is small compared to the perimeter of the block. This last test excludes, for example, where the second lane of a divided highway is added to a block.

This approach identifies blocks with new roads (see example below), while rejecting others where there do not appear to actually be new roads (including Figure 3 above). See, for example, Figures 4 and 5

[5] See, e.g., https://www.census.gov/programs-surveys/geography/technical-documentation/county-changes.html.

below showing census block 121199114001002 in Sumter County, Florida. Between 2010 and 2015, there were over 37.5 km (over 23 miles) of eligible roads added to this block alone.

Figure 4 Census block 121199114001002 in Sumter County, Florida, shaded blue. On the left, roads from the 2010 TIGER Edge data; on the right roads from the 2015 TIGER Edge data.

Figure 5 Close up of block 121199114001002 overlaid on satellite image showing residential development. Imagery © 2016 Digital Globe, Lake County, U.S. Geological Survey

There are examples where the approach appears to miss new development. See for example, Figure 6.

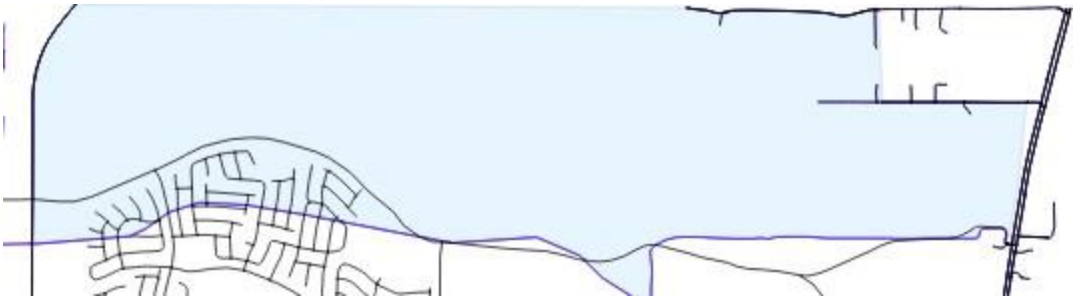


Figure 6 Census block 121199114001000 in Sumter County, Florida shaded blue, showing 2015 TIGER roads (North faces left in this figure); this block is just east of the block shown in Figures 4 and 5. This block is not identified as having new roads because of the amount of new road relative to the size of the block.

The drawbacks of assuming the presence of housing units and population where none actually exist is significant. The Commission could, for example, treat such areas as lacking broadband; but if those areas lack development, they are unlikely to ever have broadband deployed. Thus, the approach attempts to be conservative in identifying areas with new development. Unfortunately even with this approach, there appear to be areas identified as having new development that reflect corrections to the underlying data.

Figure 7 Census block 121199107001123 in Sumter County, Florida shaded in blue with 2015 TIGER road data overlaid. The two roads running north-to-south across the block are new in 2015 compared to 2010 and lead this block to be identified as having new roads and development. Imagery © 2016, DigitalGlobe, US Geological Survey

Add housing units and population along new roads

We next use the blocks identified as having new roads, and the length of new roads in those blocks, to determine the number of housing units and population to add. Complicating this is what appears may be a lag between the TIGER road data and the decennial census – it may take longer for roads to appear in the TIGER data than it takes for housing units, households and population to appear in the decennial census. For example, block 121199112001349 as shown below in Figure 8, was reported as having 440 housing units (and a population of only 373) in the 2010 Census figures despite having very few roads according to 2010 TIGER road data (left-hand side). A more complete road network appears in 2015 TIGER road data (right-hand side).

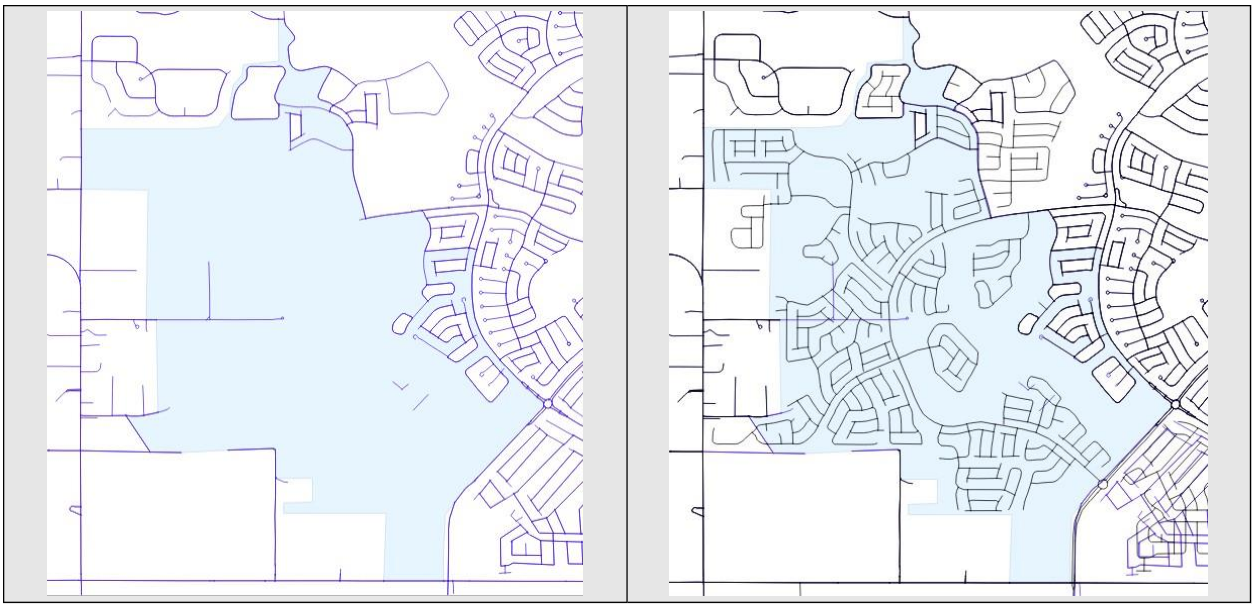


Figure 8 Census block 121199112001349 in Sumter County, Florida, shaded in blue. The left-hand side shows the roads according to 2010 TIGER data; the right hand side shows roads from 2015 TIGER data overlaid.

Therefore, the code treats blocks with a significant number of housing units or population differently than those with only a limited number of each. For blocks with fewer than five housing units or a population of ten in the decennial census, we assume one housing unit for every 30m (~100 feet) of new road length; and one person for every 15m. For blocks with more housing units or population, we assume one housing unit for each new road segment (new TLID).

In some cases, there is very large growth in road lengths in the current TIGER road data (perhaps due to corrections of prior road omissions rather than new road construction) in counties with relatively few housing units and population counts overall. Allocating one housing unit for every 30m of new road and one person for every 15m would mean putting close to – or on occasion more than – the total number of new housing units and all the new population of some counties in just one or two blocks. To avoid this, the code will place no more than 10% of the decennial housing units and population in any one census block due to road growth.

Step Four: Adjust Block-level Development to Align with County-level Estimates

Adding Development

Where needed, we add housing units and population along road segments. The first step in the process is to randomly pick an eligible road segment as identified in Step Two above. Note the use here of TIGER Edge data, where a long road is split into many segment by intersecting roads, rather than road data. This increases the likelihood of placing a location in a more densely settled area because there are more road segments in areas with more intersecting roads. For each road segment identified, the code finds a point a random length along the segment. Then, to ensure the selected point is in one block (not on the border between two blocks), we shift the point off the road by several meters; that offset is to a random

side of the street in a direction perpendicular to the end points of the street segment. Finally the code finds the census block associated with each point.

Note that it is possible for the offset point to be outside the county of interest (when the selected road segment makes up a border of the county and the offset moves the point outside rather than into the county). Due to such cases, there can be slightly fewer housing units or population added to counties than the total change in those quantities for that county; however, such discrepancies are typically small (compared, e.g., to the difference between the 2010 census count for a county and the 2010 county-level estimate for that same county).

About annual county-level housing unit and population estimates

Each year, the U.S. Census Bureau publishes population estimates for each county in the 50 states, the District of Columbia and Puerto Rico; and housing unit estimates for each county in the 50 states and the District of Columbia. We estimate the change in housing units for each county in Puerto Rico as proportional to the change in population in that county.

For American Samoa, Guam, the Northern Mariana Island and the Virgin Islands, we take population for the territory from the Census International Database (IDB); and allocate change in population in proportion to the decennial population. We then allocate change in housing units in proportion to population. This avoids allocating housing unit or population to counties (or county equivalents) that have no population in the most recent census.

As noted above, county definitions can change. Each year's county estimates reflect the then-current definition of county boundaries. For example, three counties in Alaska had their boundaries change between 2010 and 2015; U.S. Census Bureau estimates in 2015 reflect the 2015 definition of those counties in reporting the values for 2010. Thus there's a significant difference between the 2010 census figures and the 2010 estimates.

In order to account for changes to county definitions, this approach uses the difference between the current year estimates and the estimates for the most recent decennial census year, rather than the difference between current year estimates and decennial census counts. As noted above, however, this approach relies on block-level figures from the decennial census. *Therefore, because of difference between decennial census counts and estimates (even for counties that have not had a change in their definition), the county-level totals will typically not agree perfectly with U.S. Census county-level estimates.*

Removing housing units or population

Removing housing units or population randomly could lead to removing the only housing unit or person from a given census block, effectively making that block irrelevant for at least some policy-making purposes. Rather than risk dropping blocks in such a way, and recognizing that blocks with greater housing units and population are more likely to have reductions, staff remove housing units or population from blocks with higher starting counts.

The approach removes housing units or population from the block with the largest initial count. That process continues until the reduction target is met or until the block that had the highest number of housing units or greatest population has the same number as the second-highest. The process then removes housing units or population from the highest two until the target is met or until those two blocks have the same count as the third highest, and so on.

An illustrative example will highlight the method taken. Imagine a county consisting of five census blocks, each represented by a vertical line on the left-hand side of Figure 9 below. Each blue dot represents a housing unit in that block. If the count of housing units needs to be reduced by three, the three red dots shown in the right-hand side of Figure 9 would be removed (i.e., the block with the most housing units would have the count in that block reduced by three). To remove four to seven housing units, the green dots would be removed; and to remove eight to ten housing units the orange dots would be removed

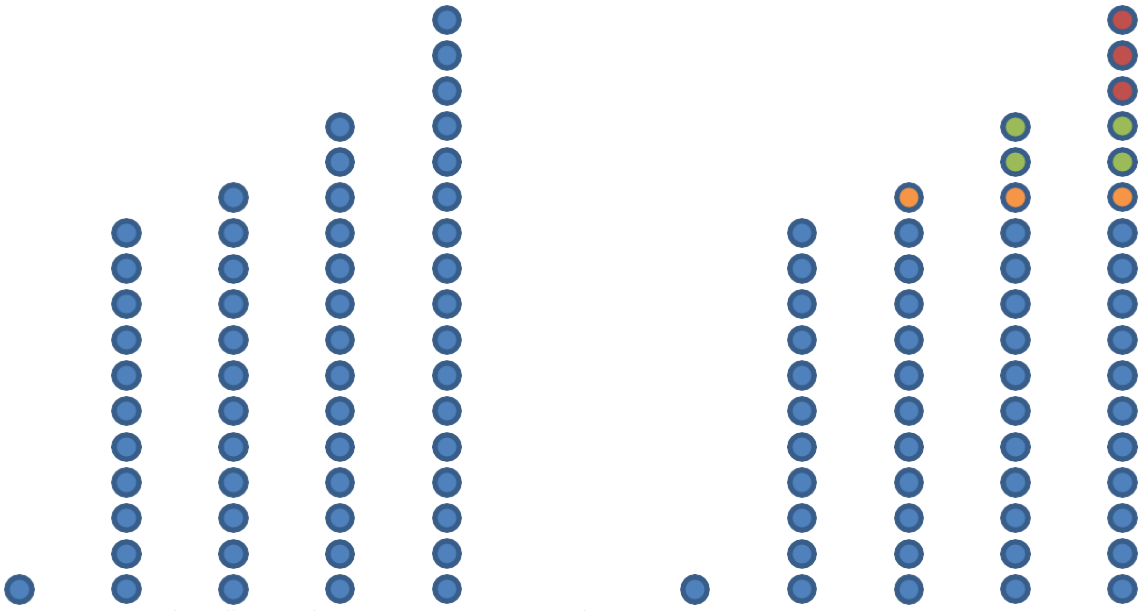


Figure 9 Example to illustrate how HU or POP are removed

The identification of new roads and placement of new housing units and population along those roads in steps two and three occurs independent of the

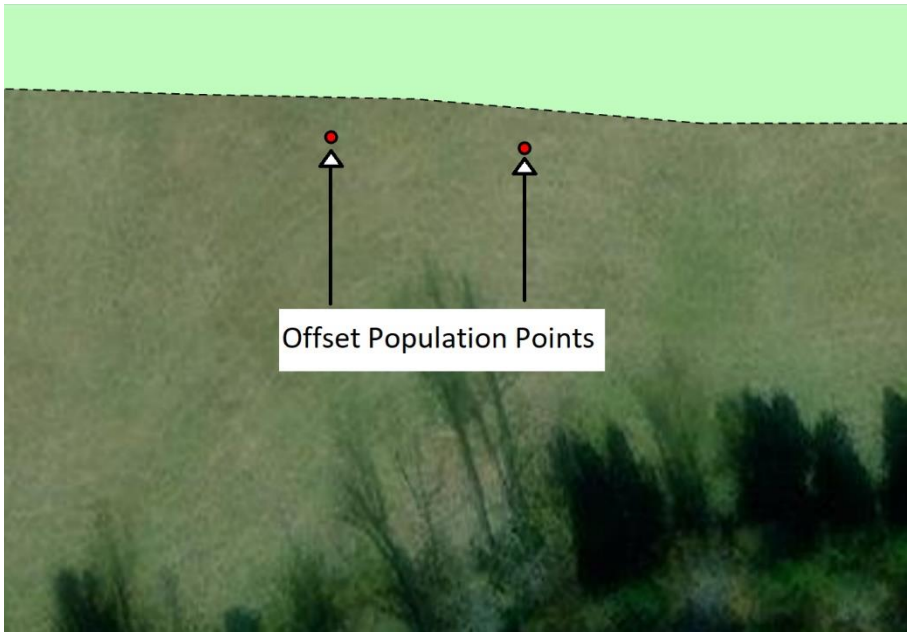
net change in housing units and population in each county. The number of housing units and population added or removed in step four reflects any such additions

Step Five: Calculate the Number of Households in Each Block

Since the U.S. Census Bureau does not release county-level estimates of households annually, Commission staff estimated the number of households in each block based on the number of housing units. In blocks where there are housing units in the decennial Census, the number of households in the current year is derived by taking the number of housing units estimated in the block for the current year and multiplying it by the ratio of households to housing units for that block. For blocks where there are no households in the decennial census, the code uses the ratio of households to housing units for the county in the decennial census instead.

This methodology produces household counts that may not be whole numbers in each block. Over larger geographic areas, like counties, however, staff believe this approach will provide a better estimate for households than rounding the household count to whole numbers in each block. Typically staff will round to whole numbers of housing units at the tract or county level.

Step Six: Sum and Assign Out-of-Block Offset Points Back to Blocks



It is possible for the offset point to be outside the county of interest. This occurs when a point is placed along a road that delineates a county's external border and is offset into an adjacent county.

Figure 10 Example of points offset out of county and into adjacent county

The county shown in light green in Figure 10 would contain two fewer people than the Census County Estimate for a given year. To re-assign such points back to a county, 500m buffers are constructed around each offset population point that is placed outside of a county. These buffers are intersected with a county and the area of that intersection is calculated. We then sort the area of the intersections

in a descending order and assign the out-of-block point to the county with largest intersection. This is done to accommodate points placed between two blocks in addition to being outside the county. Next, we compile a list of Block IDs and the sum of offset points to be re-assigned, and then add those values to each Block ID in the list.

Data Processing Notes: 2020 Census Block Population

Connecticut Population Estimates Beyond 2022

As of June 2022, the Census Bureau adopted the state of Connecticut's request to change the county-equivalent entities in the geographic hierarchy from eight counties to nine planning regions for the purposes of collecting, tabulating, and disseminating statistical data. Since Census Bureau county-level population estimates for 2020 and 2021 were reported for the eight legacy counties, it was difficult to use those data to determine county-level change in population estimates for 2022 (reported using the nine new county equivalents). To maintain continuity with our established FCC Staff Block Population Methodology, we ignored county estimates for Connecticut and used the change in state population estimates between 2021 and 2022 to inform the 2022 block population estimates (i.e., Connecticut was processed as if it were a single county).

Island Area Block Estimates

Within the Island Areas, population counts for some blocks and other geographies at lower levels on the central spine of the census geographic hierarchy were suppressed as part of the creation of the Demographic and Housing Characteristics file (see the technical documentation, Chapter 5, for additional details at https://www.census.gov/programs-surveys/decennial-census/technical-documentation/island-areas-censuses.html). In these situations, the population for the next-highest level of geography up the hierarchy that was left unassigned to a block was disaggregated down to those blocks with missing populations using the percentages of decennial housing as a weight to assign each person.

The Census Bureau does not produce annual county or state population estimates for the Island Areas. Instead, we used the population estimates for each Island Area found in the Census International Database and processed each as if it were a single county to obtain 2022 block population estimates.