CALIFORNIA
Water Boards

APPENDIX:
GENERAL GIS METHODOLOGY

LAST UPDATED: JUNE 2026

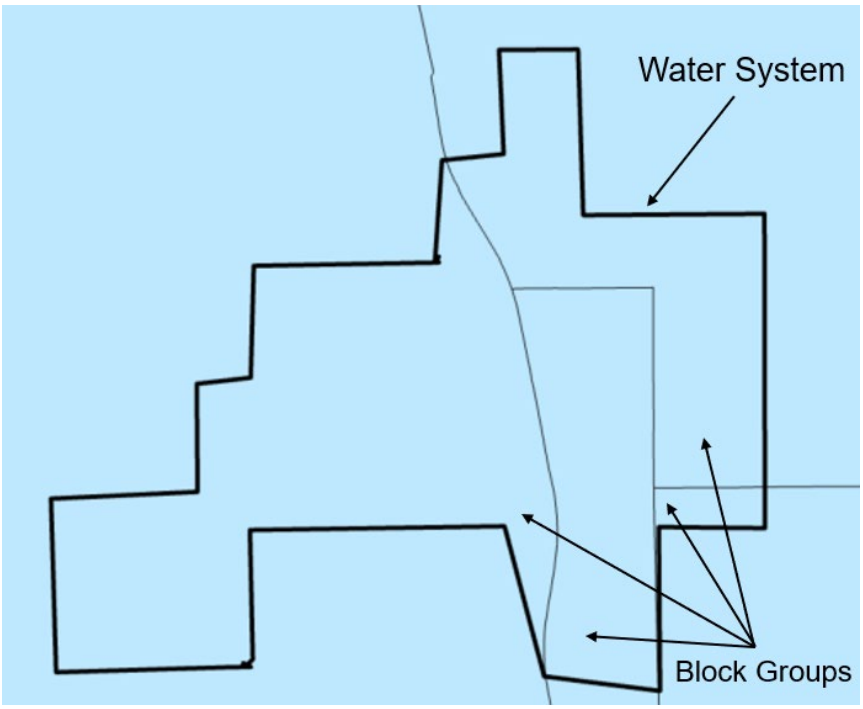
This appendix is related to the Drinking Water Needs Assessment. Learn more here: Drinking Water Needs Assessment.

CONTENTS

INTRODUCTION

Many demographic and socioeconomic datasets used in the Needs Assessment, including data derived from the American Community Survey (ACS) and CalEnviroScreen, are published for standardized geographic units such as census block groups, census tracts, counties, and other administrative boundaries. However, a key challenge arises when the geographic boundaries of interest – such as water system service area boundaries – do not align with census-defined geographies like block groups, census tracts, or counties. This mismatch can complicate efforts to accurately assess socioeconomic conditions within those areas.

Figure 1: Example of Census Area Mismatch Compared to Water System Service Area Boundary



Census data is typically organized into distinct geographic boundaries that describe the size of the surveyed area. The two most common census areas used by the Drinking Water Needs Assessment are block groups and census tracts. These are the two smallest geographic units for which ACS 5-Year Estimates are routinely published and provide the localized data necessary to calculate several of the indicators used in the Risk and Affordability Assessments. Other census areas – places and Zip Code Tabulation Areas (ZCTAs) – are

used in the calculation of median household income (MHI) and discussed in the Appendix: Median Household Income (MHI) and Economic Status Determination Methodology.[1]

Block Groups: A subdivision of a census tract (with each tract encompassing 1-9 block groups), the block group-level is the smallest geographic unit for which the U.S. Census Bureau publishes sample data.

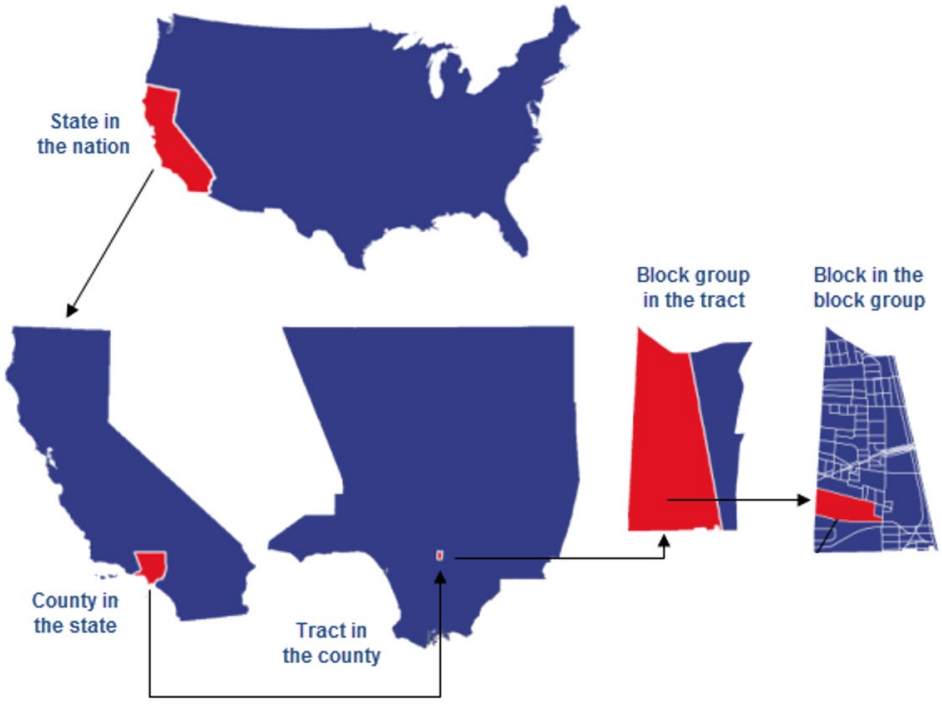
Census Tracts: Small, relatively permanent statistical subdivisions of a county that are uniquely numbered within each county with a numeric code. On average, there are approximately 4,000 inhabitants per census tract.

Places: Geographical subdivisions including legally incorporated cities or towns, as well as Census Designated Places (CDP), which are unincorporated areas defined for statistical purposes only. Importantly, places are not spatially continuous across California, meaning that some regions do not have an associated census place.

Zip Code Tabulation Areas: Statistical geographic areas created by the U.S. Census Bureau to approximate U.S. Postal Service ZIP Code service areas. ZCTAs are built by aggregating census blocks based on the most frequently occurring ZIP Code associated with addresses in those blocks. They are designed for tabulating census data and may not exactly match USPS ZIP Codes, which are defined for mail delivery and can change over time.

[1] Appendix: Median Household Income (MHI) and Economic Status Determination Methodology https://www.waterboards.ca.gov/drinking_water/certlic/drinkingwater/documents/needs/2026/2026mhi-calculation.pdf

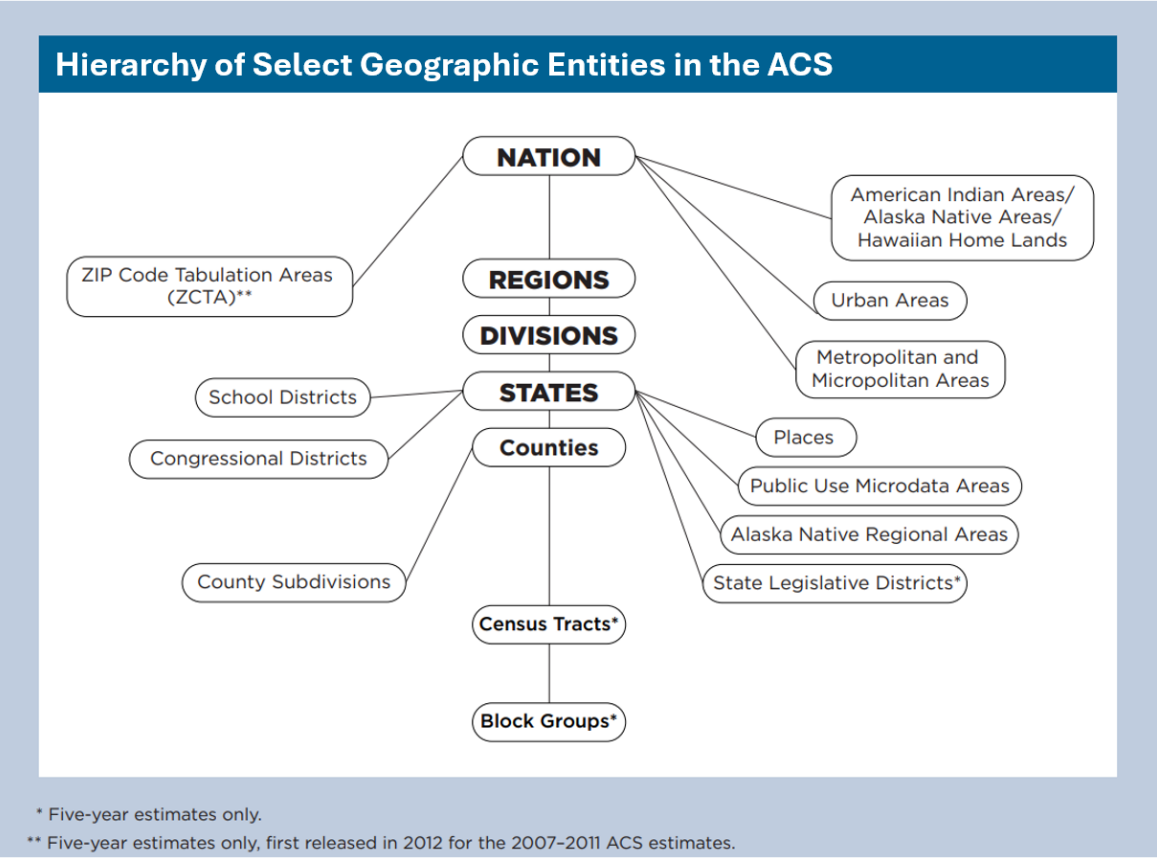
Figure 2: Geographical Census Boundaries[2]



Block groups are typically used as the primary geographic unit because they strike a balance between data availability and geographic precision, since they are the smallest geographic unit for which the U.S. Census Bureau publishes sample data (i.e., ACS 5-Year Estimates that survey a sample of the population each year). However, census tracts may be used in certain cases for the purposes of the Needs Assessment because the data are only published at that level – for example, the Housing Burden subcomponent of the Household Socioeconomic Burden indicator.

[2] United States Census Geography - ArcGIS Related Concepts
https://learn.arcgis.com/en/related-concepts/united-states-census-geography.htm

Figure 3: Hierarchy of Select Geographic Entities in the ACS[3]



The general methodology employs a spatial weighting approach that joins ACS data to spatial shapefiles for the selected census geography (block groups, census tracts, places, or ZCTAs), then intersects those with water system boundaries to calculate weighted averages for each system. Three weighting methods can be applied using this spatial join framework:

- **Area-Weighted:** Weights are calculated based on the proportional area of overlap between the census geography and the water system boundary.
- **Population-Weighted:** Weights are calculated based on the proportional census population within the overlapping areas.
- **Parcel Number-Weighted:** Weights are calculated based on the proportional number of residential parcels within the overlapping areas.

This approach allows socioeconomic estimates to be generated at the water system boundary level, despite misalignments between system boundaries and census geographies.

[3] ACS General Handbook Ch. 2. Geographic Areas Covered in the ACS https://www.census.gov/content/dam/Census/library/publications/2020/acs/acs_general_handbook_2020_ch02.pdf

Geospatial analyses performed for the Needs Assessment were conducted in R (version 4.6.0) using the sf (Simple Features) package. The sf package was used to perform spatial operations including intersections, area calculations, and geographic joins between water system service area boundaries and census geographies. The following guidance is tailored to workflows in R.

SPATIAL INDICATORS AND DATA SOURCES

The Socioeconomic and Demographic indicators are calculated using area-weighted GIS methodology for two census geography types, block groups and census tracts. Any indicator that requires GIS uses the following spatial data sources:

- California Drinking Water System Area Boundary Layer: SABL[4]
- Census Geography Boundaries: TIGER/Line Shapefiles[5]

Household Socioeconomic Burden

The Household Socioeconomic Burden indicator is used in both the Risk and Affordability Assessments and is intended to identify water systems that serve low-income communities experiencing both high poverty rates and high housing costs for households. These communities may already struggle to afford their current water bills with limited disposable income constrained by high housing costs and could face additional hardship if customer charges increase in the future. This indicator combines two metrics – Poverty Prevalence and Housing Burden – to capture the compounded financial strain on a water system's customers. Poverty Prevalence (from ACS data) and Housing Burden (from the U.S. Department of Housing and Urban Development's Comprehensive Housing and Affordability Strategy data) estimates are available at the block group and tract level, respectively, so spatial analysis is required to derive water system-level estimates of these indicator sub-components.

Required Data Sources:

- Ratio of Income to Poverty Level in the Past 12 Months (data table C17002): most recently available 5-Year Block Group-Level Estimates from ACS[6]
- Table 8 – Tenure by Household Income, Housing Cost Burden and Substandard Housing: most recently available 5-Year Census Tract-Level Estimates from Comprehensive Housing Affordability Strategy data, U.S. Department of Housing and Urban Development (HUD)[7]

[4] California Drinking Water System Boundaries (SABL) https://gispublic.waterboards.ca.gov/portal/home/item.html?id=fbba842bf134497c9d611ad506ec48cc
[5] TIGER/Line shapefiles (U.S. Census Bureau) https://www.census.gov/cgi-bin/geo/shapefiles/
[6] Census Bureau data table C17002 https://data.census.gov/table
[7] HUD Office of Policy Development and Research Comprehensive Housing Affordability Strategy (CHAS) data, https://www.huduser.gov/portal

Demographic Analyses

Demographic data is used alongside the results of the Needs Assessment to better understand the characteristics of populations most at risk. These data help provide important information about water systems that may be socioeconomically vulnerable and therefore especially affected by risk or affordability challenges. Demographic data points come from CalEnviroScreen (Draft 5.0 version) and ACS estimates. CalEnviroScreen is a screening tool developed by the Office of Environmental Health Hazard Assessment (OEHHA) to identify California communities facing socioeconomic, health, and environmental burdens. It includes a Population Characteristics Score (representing social and health vulnerabilities) and a Pollution Burden Score (representing exposure to pollutants), both ranging from 0 to 10, with 10 indicating highest vulnerability or burden. These scores are multiplied to generate a composite index at the census tract level that can range from 0 to 100. The ACS provides the following data at the block group level: poverty prevalence, linguistic isolation (the degree to which the population is limited English-speaking), household size, and the racial and ethnic breakdown of the population.

Required Data Sources:

- Draft CalEnviroScreen 5.0 Data (OEHHA): Census Tract-Level Estimates[8]
- Ratio of Income to Poverty Level in the Past 12 Months (ACS data table C17002): most recently available 5-Year Block Group-Level Estimates[9]
- Household Language by Household Limited English-Speaking Status (ACS data table C16002): most recently available 5-Year Block Group-Level Estimates
- Average Household Size of Occupied Housing Units in Tenure (ACS data table B25010): most recently available 5-Year Block Group-Level Estimates
- Hispanic or Latino Origin by Race (ACS data table B03002): most recently available 5-Year Block Group-Level Estimates

GENERAL AREA-WEIGHTED GIS METHODOLOGY

Calculation Steps:

Load California Drinking Water System Area Boundary Layer (SABL): The System Area Boundary Layer (SABL) maintained by the State Water Board is used to define public water system service area boundaries for spatial analyses. Previous versions of the Needs Assessment relied on a supplemental layer known as SABL+, which combined SABL boundaries with artificial service area boundaries created for water systems lacking mapped boundaries. These artificial boundaries were generated using a 0.5-mile buffer around the

[8] OEHHA CalEnviroScreen
https://oehha.ca.gov/calenviroscreen
[9] All Census Bureau data tables referenced in this section were retrieved from
https://data.census.gov/table

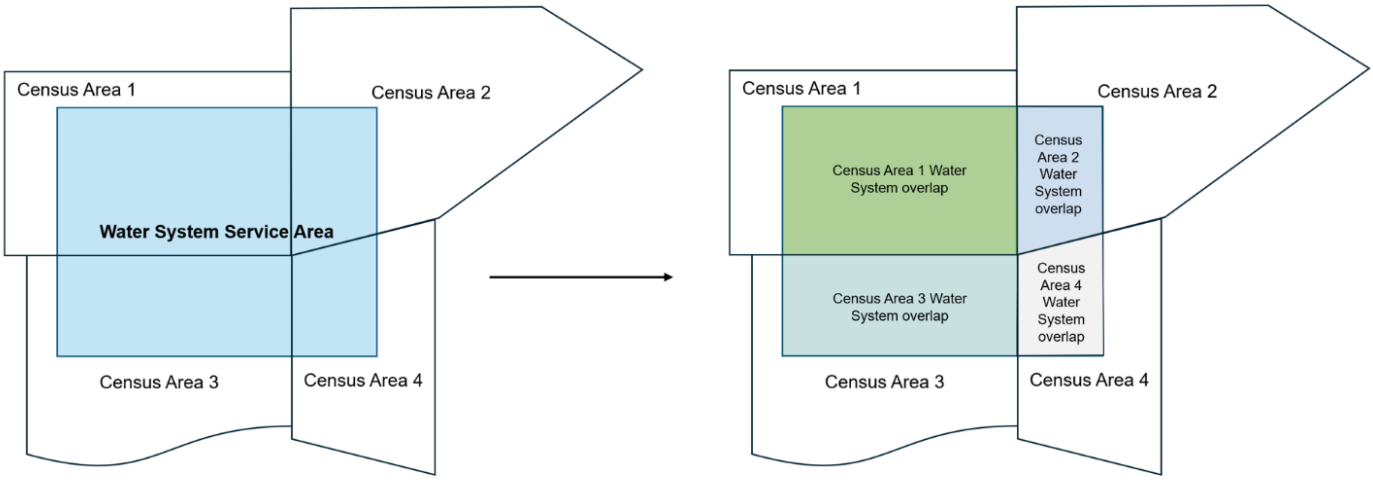
water system's distribution system facility and were used to approximate the potential service area of small systems. The State Water Board has since undertaken a comprehensive effort to add missing community water system boundaries and verify existing boundaries. As a result, all community water systems now have boundary information in SABL, eliminating the need for the SABL+ layer. Efforts to verify and refine system boundaries remain ongoing.

Clean Demographic Data: The demographic data are then prepared for analysis. This may include cleaning data for consistency, performing a margin of error adjustment, or combining multiple data points (for example, in the case of Poverty Prevalence and Housing Burden).

Join Demographic Data with Corresponding Census Geographies: To prepare spatial data for analysis, the cleaned demographic estimates were then joined to the corresponding TIGER/Line shapefile for block groups or census tracts using geographic identifier fields (GEOID). Spatially defined demographic data enables analysis of water system-level estimates through area-weighting, even when a system's service area boundaries overlap with more than one census area.

Intersect Water System Boundaries and Census Geographies: Service area boundaries from SABL were intersected with census boundaries containing demographic data using the sf (Simple Features) package in R. This analysis produced a new shapefile containing the geographic overlap between the service areas and census geographies.

Figure 4: Illustration of Census Area-Weighting Method



Determine Intersection "Weight": For each water system, intersection weights were calculated based on the portion of the system overlapping census areas containing non-missing demographic data. When a system intersected multiple census areas, and one or more of those areas were missing a particular demographic estimate, only the intersecting areas with non-missing values were included in the weighting. The total intersecting area with non-missing demographic data was used as the denominator to calculate each area's share of

the system-level weight. This ensured that missing data did not distort the final area-weighted calculation.

Perform Census Area-Weighted Calculation: After determining the value of a demographic variable, the water system-level estimate was calculated by weighting each intersecting census-area estimate by the percentage of the census geography area intersecting the water system.[10] This ensured that demographic characteristics were accurately represented based on the geographic overlap with the water system. The formula for this weighting is found in Equation 1 and an example calculation for a water system overlapping two census areas (A and B) is shown in Equation 2.

Equation 1: Water System Census Area-Weighted Demographic Estimate Calculation

$$\text{Demographic Estimate (Census Area−Weighted)} = \sum\left(\left(\frac{\text{Intersection Area}}{\text{Total Non−Missing Area}}\right) \times \text{Census Area Demographic Estimate}\right)$$

Equation 2: Example Area-Weighted Calculation

$$\text{Demographic Estimate (Census Area−Weighted)} = \left(\begin{matrix}\text{\% of WS intersecting}\\ \text{Census Area A}\end{matrix} \text{ x } \begin{matrix}\text{Demographic estimate}\\ \text{of Census Area A}\end{matrix}\right) + \left(\begin{matrix}\text{\% of WS intersecting}\\ \text{Census Area B}\end{matrix} \text{ x } \begin{matrix}\text{Demographic estimate}\\ \text{of Census Area B}\end{matrix}\right)$$

[10] This method implicitly assumes that the population of each water system is evenly distributed across its service area. In other words, if 50% of a system's area overlaps with a given block group, it is assumed that 50% of the system's population resides within that block group. While this area-weighting approach is commonly used due to data availability and straightforward calculation, it does not account for the actual distribution of people within the service area. This assumption may be particularly problematic in rural areas, where census geographies tend to cover large expanses with sparse or uneven population densities.