CALIFORNIA
Water Boards

APPENDIX:
UNIVERSAL GIS METHODOLOGY FOR THE APPLICATION OF AMERICAN COMMUNITY SURVEY DATA TO WATER SYSTEM BOUNDARIES

LAST UPDATED: JUNE 2025

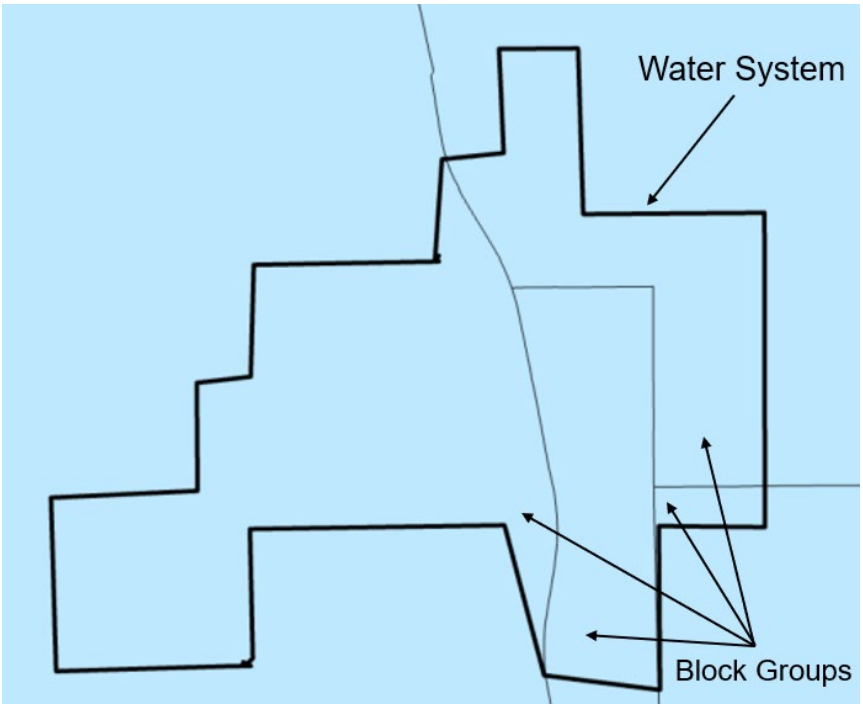
This appendix is related to the Drinking Water Needs Assessment. Learn more here: Drinking Water Needs Assessment

CONTENTS

INTRODUCTION

The American Community Survey (ACS) is an annual demographic survey program conducted by the United States Census Bureau and serves as a widely used source of socioeconomic data across the country and for California. Researchers rely on ACS data to analyze community-level conditions and trends for policy and academic purposes. 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 geographical boundaries that describe the size of the surveyed area. The smallest categories will be discussed for this analysis: Census Blocks, Block Groups, & Census Tracts.

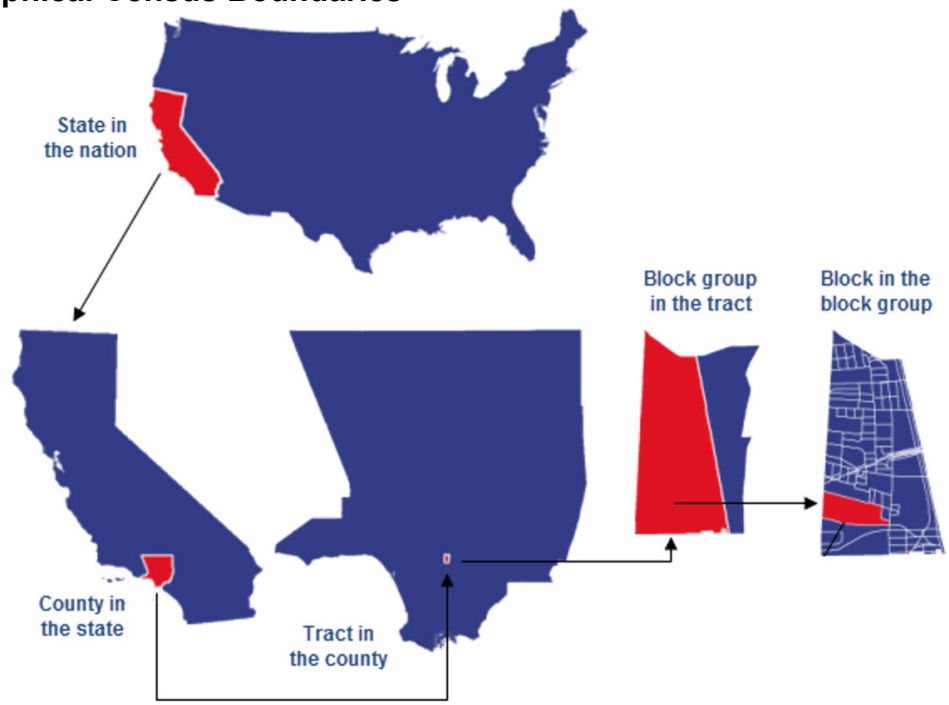
Census Blocks: The smallest geographic area for which the U.S. Census Bureau collects and tabulates decennial census data.

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.

Place: 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.

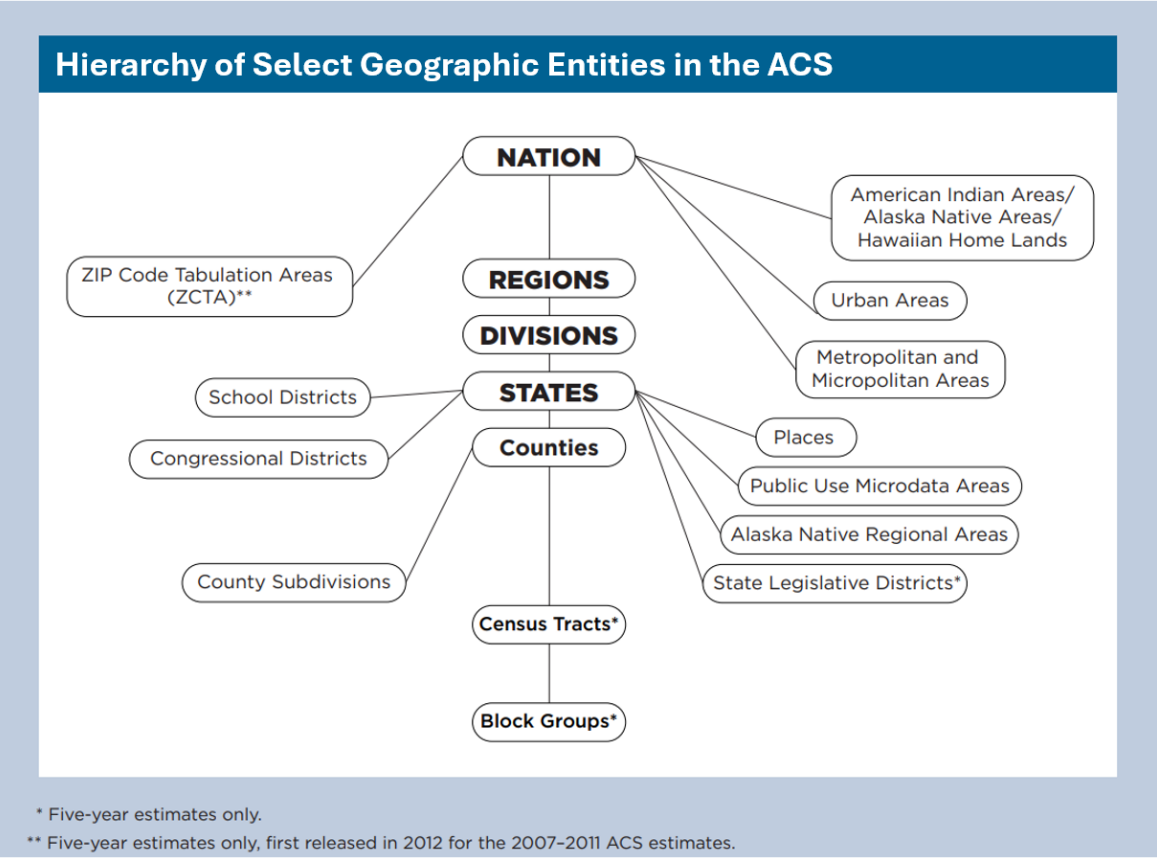
Figure 2: Geographical Census Boundaries[1]



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, other census geographies (such as tracts or places) may be used in certain cases for the purposes of the Needs Assessment, either because the data are only published at that level – for example, the housing burden subcomponent of the Household Socioeconomic Burden indicator – or to provide additional context, as in the case of the median household income calculation.

[1] United States census geography—Related Concepts | Documentation (arcgis.com)
https://learn.arcgis.com/en/related-concepts/united-states-census-geography.htm

Figure 3: Hierarchy of Select Geographic Entities in the ACS[2]



The general methodology is an area-weighted approach that involves joining ACS data to spatial shapefiles for the selected census geography (block groups, census tracts, or places), then intersecting those with water system boundaries to calculate area-weighted averages for each system. This process allows socioeconomic estimates to be generated at the water system boundary-level, despite misalignments between system boundaries and census geographies.

The Needs Assessment uses ArcGIS Pro (3.3.0) and the sf package in R (version 4.4.0) for geospatial analysis. The following guidance and code are tailored to workflows in these platforms.

SPATIAL INDICATORS AND DATA SOURCES

This section discusses the data points for the Needs Assessment that are calculated using the general area-weighted GIS methodology. Any indicator that requires GIS uses the following spatial data sources:

[2] 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

- Water system Service Area Boundary Layer: SABL[3]
- Census Geography Boundaries for Block Groups, Census Tracts, and Places: TIGER/Line Shapefiles[4]

Median Household Income (MHI)

MHI is a key data point used in the Risk and Affordability Assessments to identify water systems serving disadvantaged and severely disadvantaged communities (DAC/SDAC). DAC water systems are defined by Senate Bill 200[5], which established the Safe and Affordable Drinking Water Fund, as an area in which the MHI is less than 80% of the statewide average (SDAC is less than 60%). MHI and DAC status are also critical for assessing the affordability of water service provided to a system's customer base. The Affordability Assessment identifies DAC/SDAC community water systems and non-transient non-community water systems serving K-12 schools that have instituted customer drinking water charges exceeding the Affordability Threshold established by the State Water Board. This assessment is required to ensure compliance with state and federal drinking water standards and helps inform the State Water Board's annual Fund Expenditure Plan. Affordability is captured by the %MHI indicator (used in both the Risk and Affordability Assessments), which measures the annual system-wide average residential customer charges for 6 hundred cubic feet (HCF) of drinking water usage per month relative to the annual MHI within a water system's service area.

More details on the MHI calculation methodology can be found in the *Appendix: Median Household Income (MHI) and Economic Status Determination Methodology*.[6]

Required Data Source:

- Median Household Income in the past 12 months (data table B19103): 2019-2023 5-Year Estimates from U.S. Census Bureau's American Community Survey for block groups, census tracts, and places. Earlier ACS 5-Year Estimates (2016-2020, 2018-2022 and 2017-2021) were used selectively to address missing or suppressed data.[7]

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 communities experiencing both high poverty rates and high housing costs for low-income households. These communities may already struggle to afford their current water bills with limited disposable

[3] California Drinking Water System Boundaries
https://gispublic.waterboards.ca.gov/portal/home/item.html?id=fbba842bf134497c9d611ad506ec48cc
[4] TIGER/Line shapefiles (U.S. Census Bureau): https://www.census.gov/cgi-bin/geo/shapefiles/
[5] DWSRF Disadvantaged Community Definitions: A Reference for States
https://www.epa.gov/system/files/documents/2022-10/DWSRF%20DAC%20Definitions%20Report_October%202022%20Updates_FINAL_508.pdf
[6] Appendix: Median Household Income (MHI) and Economic Status Determination Methodology
https://www.waterboards.ca.gov/drinking_water/certlic/drinkingwater/documents/needs/2025/2025mhi-calculation.pdf
[7] American Community Survey Data Tables
https://data.census.gov/table

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 census 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): 2019-2023 5-Year Block Group-Level Estimates from ACS[8]
- Table 8 – Tenure by Household Income, Housing Cost Burden and Substandard Housing: 2017-2021 5-Year Census Tract-Level Estimates from Comprehensive Housing Affordability Strategy data, U.S. Department of Housing and Urban Development (HUD)[9]

Demographic Analyses

Demographic data is used alongside the results of the Needs Assessment to better understand the characteristics of populations most at risk. This 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 4.0 and ACS estimates. CalEnviroScreen is a screening tool developed by the Office of Environmental Health Hazard Assessment 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. The ACS provides data on 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:

- CalEnviroScreen 4.0 Data (2021): Census Tract-Level Estimates[10]
- Ratio of Income to Poverty Level in the Past 12 Months (data table C17002): 2019-2023 5-Year Block Group-Level Estimates from ACS[11]

[8] Census Bureau data table C17002 retrieved March 11, 2025 from https://data.census.gov/table/ACSDT5Y2023.C17002?g=040XX00US06$1500000&y=2023
[9] HUD Office of Policy Development and Research Comprehensive Housing Affordability Strategy (CHAS) data, retrieved January 27, 2025 from https://www.huduser.gov/portal/datasets/cp.html#data_2006-2021
[10] CalEnviroScreen 4.0 Data retrieved January 27, 2025 from https://oehha.ca.gov/calenviroscreen/download-data
[11] Census Bureau data table C17002 retrieved March 11, 2025 from https://data.census.gov/table/ACSDT5Y2023.C17002?g=040XX00US06$1500000&y=2023

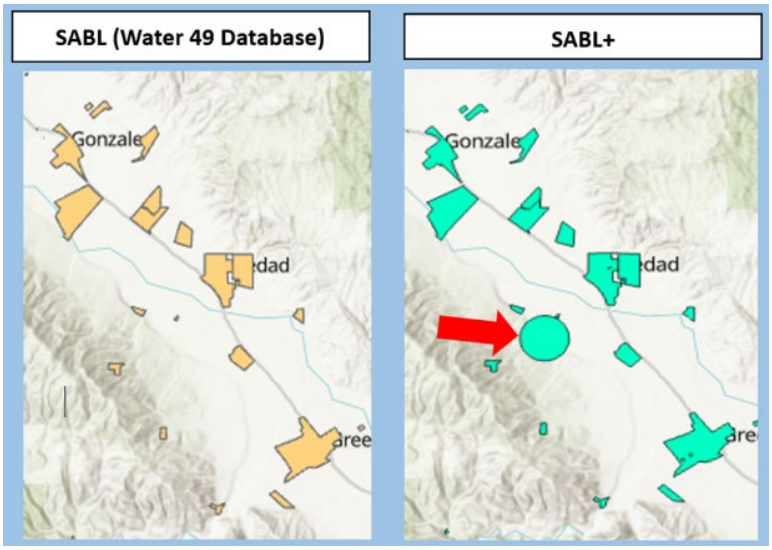
- Household Language by Household Limited English Speaking Status (data table C16002): 2019-2023 5-Year Block Group-Level Estimates ACS[12]
- Average Household Size of Occupied Housing Units in Tenure (data table B25010): 2019-2023 5-Year Block Group-Level Estimates ACS[13]
- Hispanic or Latino Origin by Race (data table B03002): 2019-2023 5-Year Block Group-Level Estimates ACS[14]

GENERAL AREA-WEIGHTED GIS METHODOLOGY

Calculation Steps:

Create System Area Boundary Layer+ (SABL+): First, water system boundaries from SABL were combined with artificial boundaries for water systems that were included in the Risk Assessment but did not have a known boundary in SABL. Artificial boundaries were generated for the purposes of the Needs Assessment by creating a 0.5-mile buffer around the location of the water system's distribution system facility. The 0.5-mile distance was chosen as an approximation to the potential service area since most of the missing systems are small water systems. The boundaries from SABL combined with the artificial boundaries are referred to as SABL+ and this layer is used in risk indicator calculations that require spatial analysis.[15]

Figure 4: Visual comparison of the Standard SABL layer to the SABL+ layer. Note the addition of the estimated water system boundary for the missing system (see red arrow)



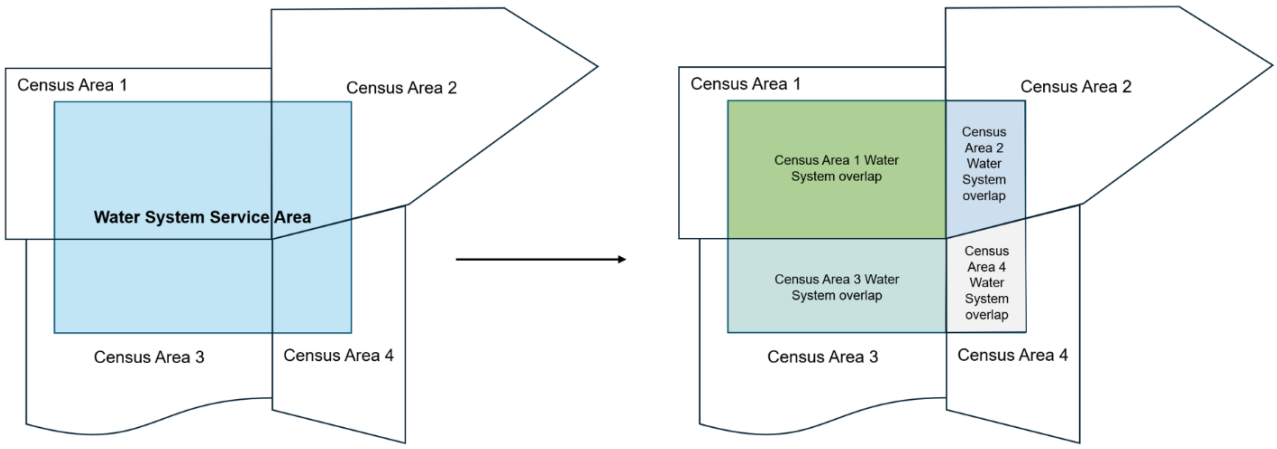
[12] Census Bureau data table C16002 retrieved April 3, 2025 from https://data.census.gov/table/ACSDT5Y2023.C16002?g=040XX00US06$1500000&y=2023
[13] Census Bureau data table B25010 retrieved April 4 2025 from https://data.census.gov/table/ACSDT5Y2023.B25010?g=040XX00US06$1500000&y=2023
[14] Census Bureau data table B03002 retrieved April 3 2025 from https://data.census.gov/table/ACSDT5Y2023.B03002?g=040XX00US06$1500000&y=2023
[15] SABL_Plus - Overview (ca.gov) https://gispublic.waterboards.ca.gov/portal/home/item.html?id=0e4c019a46454725b058edd90538732a

Clean Demographic Data: The demographic data (e.g., ACS block-group MHI estimates) are then prepared for analysis. This may include supplementing missing data (in the case of MHI calculation), 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, census tracts, or places 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 5: 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 valid (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 valid 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 poverty prevalence for each area of intersection, the system's overall poverty prevalence was calculated by weighting each intersecting poverty prevalence estimate by the percentage of the block group area

intersecting the water system.[16] This ensured that poverty prevalence values 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 Area-Weighted Demographic Estimate Calculation

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

Equation 2: Example Area-Weighted Calculation

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

[16] 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.