Population Data Sources Descriptions – Working Draft

CDPH/LHJ Populations Data Sources Task Force

10/2023

Data Sources:

* California Specific
  + California Department of Finance (DOF)
    - Main data used by many CDPH programs and LHJs, for many purposes
    - Recent issues have arisen with accuracy and errors
    - Not at census tract or zip code levels; no “disaggregated” race/ethnicity data
* US Census
  + Decennial Census
  + Census “Vintage” Estimates
  + American Community Survey (ACS) – direct
    - Used by many CDPH programs and LHJs for census tract and other sub-county work
    - Does not include all demographic detail needed for some purposes
  + American Community Survey (ACS) Public Use Microdata Sample (PUMS)
* Other Federal Data
  + SEER
* Commercial Data
  + ESRI
    - Used by CDPH and some LHJs for some purposes
    - Routine data does not include required information needed
    - May be mechanisms to get special California data
  + Claritas

| **Source** | **Description** | **Types of Data** | **Data and other Links** |
| --- | --- | --- | --- |
| California Department of Finance (DOF) | The Demographic Research Unit (DRU) of the California Department of Finance produces California population estimates and is designated as the single official source of demographic data for state planning and budgeting. Population data are used to establish appropriation limitations, distribute funds from a range of State and Federal programs, allocate capital outlay funds, and aid in the planning and evaluation of programs. DOF data have long been the main source of County (and LHJs, by extrapolation) and State population data, including by race/ethnicity, sex, and age, for many CDPH programs and many local health departments. While a core and useful source of data, there is no [legal] mandate that CDPH is required to use these data for analysis and planning purposes. There have been recent concerns about the reliability of DOF population estimates at the county level, by race/ethnicity and age. | P3 file has population projections at county level, by age (single years of age), sex, race/ethnicity (7 levels), for 2010 to 2060.  E-1 has total population estimates at the city level (and county and state) for the current year. E-4 has historical estimates at the same geography levels. | * [Complete P-3 Race/Ethnicity and Sex by Age for California and Counties](https://dof.ca.gov/forecasting/demographics/projections/) (July 1 projections) - county race/sex/age data * [E-1 Cities, Counties, and the State Population Estimates with Annual Percent Change](https://dof.ca.gov/forecasting/demographics/estimates/) (January 1 estimates) - city totals |
| Decennial Census | The Census is conducted every ten years to count the size of the US population and selected characteristics. It is the “gold standard” data backbone for many other data sources that estimate population sizes. Among many important purposes, census population counts are used to apportion the 435 seats of the U.S. House of Representatives among the 50 states for 10 years. The US Constitution mandates that the census asks questions of people in homes and group living situations, including how many people live or stay in each home, and the sex, age and race and ethnicity of each person | The core/basic data include population counts by sex and age, by race/ethnicity including multi-race combinations, at multiple geographic levels. | * [Race 2020 (P1 Redistricting Data)](https://data.census.gov/cedsci/table?q=p1) note that any California county can be selected, and the data easily downloaded) * [Ethnicity 2020 (P2 Redistricting Data)](https://data.census.gov/cedsci/table?q=p2) - Hispanic or Latino, and not Hispanic or Latino by race * Documentation (selected)   + [2020 Census National Redistricting Technical Documentation](https://www2.census.gov/programs-surveys/decennial/2020/technical-documentation/complete-tech-docs/summary-file/2020Census_PL94_171Redistricting_NationalTechDoc.pdf)   + [About 2020 Census Data Productshttps://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/release/about-2020-data-products.html](https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/release/about-2020-data-products.html) |
| Census “Vintage” Estimates | Each year the Census Bureau creates new “vintage” estimates of state and county populations, based on the latest data available, for a small set of core demographic characteristics. Each vintage includes revisions to previous years' estimates stretching back to the most recent census. | * Population counts at the state and county levels, by age (18 groups), sex, and race/ethnicity * Total population counts at city, metropolitan and micropolitan statistical areas * Total housing unit estimates at county level | * Current [Population and Housing Unit Estimates Tables](https://www.census.gov/programs-surveys/popest/data/tables.html) * [County Population by Characteristics: 2010-2019](https://www.census.gov/data/tables/time-series/demo/popest/2010s-counties-detail.html) |
| American Community Survey (ACS) – direct | The ACS is a nationwide demographic survey conducted by the US Census Bureau every year. It is the nation’s most current, reliable, and accessible data source for local statistics for critical planning topics including socioeconomic status, employment, housing, insurance and special populations including veterans. It collects data previously contained only in the long form of the decennial census. | Population sizes at multiple levels of geography (e.g. state, county, tract, ZCTA), by sex, age and race/ethnicity and characteristics including income/poverty, education, employment, ancestry, citizenship, language proficiency, migration, disability, housing characteristics, and others | * Data access (examples)   + [Table B01001 – Sex by Age](https://data.census.gov/cedsci/table?q=B01001%3A%20SEX%20BY%20AGE&g=0400000US06&tid=ACSDT5Y2020.B01001)   + [Table B03002 – Hispanic origin by race](https://data.census.gov/cedsci/table?q=B03002%3A%20HISPANIC%20OR%20LATINO%20ORIGIN%20BY%20RACE&g=0400000US06&tid=ACSDT1Y2021.B03002) * Documentation/information   + [Methodology](https://www.census.gov/programs-surveys/acs/methodology.html)   + [American Community Survey, General/Public, Information Guide](https://www.census.gov/content/dam/Census/programs-surveys/acs/about/ACS_Information_Guide.pdf) (excellent general overview and clear, general, methods) |
| American Community Survey (ACS) Public Use Microdata Sample (PUMS) | The ACS PUMS is not a separate data collection system from the “direct” ACS system described above. It is system for accessing and extracting approximately 250 person-level and approximately 200 housing-level variables available not available through regular published ACS tables. PUMS allows users to create custom variables and “cross tabulations” for local purposes. The only geography available with PUMS data is the State and Public Use Microdata Area, or “PUMA”. PUMAs are non-overlapping areas containing around 100,000 residents. For some counties, PUMAs can be aggregated/summed to obtain county-level estimates; for other, smaller counties, PUMAs cross county boundaries. |  | In general PUMS record-level data will be extracted via API or FTP, and processed by users using software such as R or SAS. But, some tabular data can be extracted using the Census tool here, <https://data.census.gov/mdat/#/>  Documentation   * https://www2.census.gov/about/training-workshops/2020/2020-03-11-acs-presentation.pdf * https://www.census.gov/programs-surveys/acs/microdata/documentation.html |
| SEER (Surveillance, Epidemiology, and End Results) | The National Cancer Institute (NCI) procures and maintains U. S. population data for the denominators of incidence and mortality rates. Annual mid-year population estimates stratified by age, gender, bridged race, and Hispanic ethnicity are available at county and census tract levels. County population estimates are produced under a collaborative arrangement between the U. S. Census Bureau and the National Center for Health Statistics with support from NCI through an interagency agreement. Census tract population estimates are produced by Woods & Poole Economics, Inc. (W&P) with support from NCI through a contract with methodology jointly developed by NCI, W&P, and the North American Association of Central Cancer Registries (NAACCR). | At tract or county level, race/ethnic (only 5 categories), age (19 group), sex | * Link for tract data: <https://seer.cancer.gov/censustract-pops/> * Link for county data: <https://seer.cancer.gov/popdata/> * Links to other data and information   + <https://seer.cancer.gov/data-software/datasets.html> |
| ESRI | Esri, a commercial source of population data, “is the global market leader in geographic information system (GIS) software, location intelligence, and mapping.” Along with an array of GIS/mapping products, they produce population estimates, including detailed estimates, for example by race/ethnicity, sex, and age at the census tract level. Some data are available publicly (?), some along with licenses for Esri GIS products, and some by direct purchase. | Age group (5 year) by sex by race/ethnicity (not OBS standard) available at multiple geographic level (city, zip code, county subdivisions, block groups, census tracts) | * [2022/2027 ESRI updated demographics methodology](https://storymaps.arcgis.com/stories/dd6c5388d48c448798613778644a1eaa) |
| Claritas | Commercial data source that might have needed population data. Used by some local health jurisdictions. More information is needed. | “Age, income, education, household composition, wealth, home ownership and more” | <https://claritas.com/data/> |