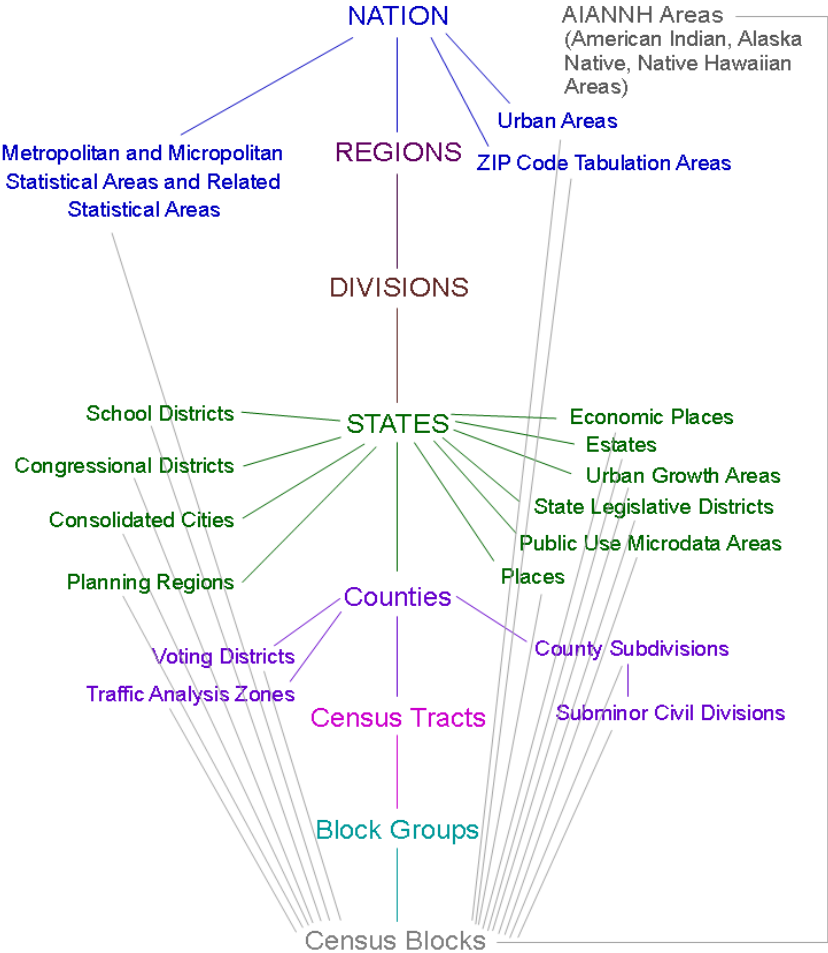


# Population Data Sources

## Decennial Census

All population estimates in the United States are ultimately based on the decennial census conducted at the beginning of each decade: 1990, 2000, 2010, 2020, etc. The US Census Bureau attempts to count each person living in the United States on April 1 of the census year and collects basic demographics: age, sex, race, and ethnicity for each person. This data is also the foundation of population estimates for non-census years with annual adjustments to the estimates based on data relating to population change: resident births, resident deaths, international immigration, and migration within the United States. Census data is available for the following geographic levels although demographic details are suppressed at smaller geographic levels to preserve privacy.

**Standard Hierarchy of Census Geographic Entities**



Decennial census data and intercensal/postcensal population estimates can be accessed through the [US census bureau](#) data portal. This access route includes data for all census geographies and by detailed race (each individual race category 'alone or in combination with another race') or single race (White, Black or African American, American Indian or Alaska Native, Asian, Native Hawaiian or other Pacific Islander, or more than one race). However, the data are often cumbersome to find and analyze because the interface just allows filtering of hundreds of pre-formatted tables and the demographic categories can't be adjusted to your needs. Census data is also made available in customizable queries through an Application Programming Interface (API). The API is more customizable but it requires more technical skill to use as well. The Census bureau does have an [API User Guide](#) to assist with getting started. Single Race population estimates at the county and state levels can also be accessed through [CDC Wonder](#). Estimates at the county-level are available in 5-year age groups and estimates at the state level are available in single-year, 5-year, and 10-year age groups.

Montana specific census data is also available through the Montana Department of Commerce, [Census and Economic Information Center](#) (CEIC). The CEIC not only makes census data available but also interfaces with the U.S. Census Bureau to help conduct the decennial census in Montana and is a great resource for any census related question. CEIC received (and shared with DPHHS) a data set with county level population by single year of age, gender, and detailed or single race for 2020-2022 from the census bureau. The data was available by request only because single year of age estimates by any other demographic factor are likely to be unstable. It is important to look at population estimates in the demographic sub-groups you are interested in before using the data for rate calculations to ensure there are at least 100 for each group. Estimates that are less than 100 are unstable and a larger geographic or demographic group should be used instead. The population estimates in this file should not be published or shared outside of DPHHS although it is acceptable to publish age-adjusted rates based on the data. The CEIC population data and the corresponding data dictionary can be accessed on the state network [here](#).

## Strengths

- Most accurate population estimates available, and the basis of all population estimates in the United States.
- Available in a variety of geographic levels.
- Includes detailed or single race estimates.
- CEIC data can produce estimates for custom age groups at the county level.

## Limitations

- Often difficult to find the data and to manipulate it for analysis.
- Geographies smaller than counties: census tracts, block groups, and census blocks can change with every census as they are defined by population targets and not legal boundaries. This might cause problems with analysis that spans before and after a census year.
- Counts are randomly adjusted to protect privacy, through a process called [Differential Privacy](#). This can cause significant problems for small populations because the "noise" introduced into the data has a larger effect when there are fewer observations. For

groups with less than 100 people population estimates are not reliable, and a larger geographic or demographic group should be used.

## National Center for Health Statistics

The National Center for Health Statistics (NCHS) produced annual population data at the county level with estimates by bridged race (White, Black or African American, American Indian or Alaska Native, Asian or Pacific Islander), Hispanic origin, sex, and single year age for each year from 1990 to 2020. Race bridging is a method used to make multiple-race and single-race data sufficiently comparable to permit estimation and analysis of race-specific statistics. The 1997 Office of Management and Budget (OMB) standards on race and ethnicity require Federal data collection programs to allow respondents to select more than one race category when reporting on their racial identity. This means that there are potentially 31 race groups (5 single-race and 26 multiple-race), depending on whether an individual selects one, two, three, four, or all five of the single-race categories. NCHS bridges the multiple-race group population counts to the four single-race categories specified above through use of a regression model with person-level and county-level covariates used to generate the probability of selecting each single-race category possible for a multiple-race group. One of the races selected will be assigned to each respondent with multiple races in proportion to these probabilities.

The most recent vintage of this data (2021) is saved on the Public Health and Safety Division shared drive: [here](#). NCHS bridged race population data are also available through [CDC Wonder](#).

These data have been the standard population data that were used to calculate all rates for surveillance data through 2020. However, NCHS will no longer produce these data so a different population data source must be used for surveillance data from 2021 forward.

### Strengths

- Ability to customize age groups and year spans.
- County level estimates are available.
- Easy to access and manipulate in SAS or other statistical software.

### Limitations

- Data for geographic areas smaller than the county are not available.
- Data are only available through 2020.
- Bridged race may cause under-counting of American Indian populations.

## American Community Survey

The American Community Survey (ACS) is conducted by the U.S Census Bureau on an ongoing basis to collect a wide variety of population and housing characteristics in addition to intercensal population estimates. The household survey collects the number of people living in the household as well as the following characteristics for each person:

- Relationship to person completing the survey
- Sex
- Age
- Ethnicity
- Race
- place of birth
- citizenship
- school attendance and educational attainment
- language other than English

- migration, where did they live a year ago?
- Health insurance coverage
- Hearing, vision, mobility, cognitive, self-care, or independent living impairments
- Marital status
- Veteran status
- Employment status
- Industry and occupation
- Transportation to work
- Income

The questionnaire also collects the following housing information:

- Type of housing (mobile home, apartment, single family, etc)
- Year the housing was built
- Lot size (for single family and mobile homes)
- Amount of agricultural product sales
- Number of rooms
- Number of bedrooms
- Presence of hot and cold running water, shower or bathtub, stove, and refrigerator
- Telephone service
- Computer and internet access
- Number of cars, vans, or trucks
- Main heating fuel
- Utility costs
- SNAP or food stamps benefits
- Rent/mortgage costs
- Insurance costs

ACS produces single year estimates for all geographies with a population of at least 65,000 people. For area with lower population, 5-year estimates are produced each year with the 5 most recent years of data. For example, the 2021 5-year estimates include data for 2017 to 2021 and the 2022 5-year estimates include data for 2018 to 2022. In Montana, only Cascade, Flathead, Gallatin, Lewis and Clark, Missoula, and Yellowstone counties currently have enough population for single year estimates. If any geography in your analysis require 5-year estimates you should use the 5-year estimates for all the geographies.

ACS data can be accessed through the same [US census bureau](#) data portal and API as decennial census data.

## Strengths

- Includes a wide variety of social determines of health and disability status that is not available in other population data sources.
- Includes estimates for even low population areas.
- Detailed race and single race estimates are available.

## Limitations

- Data is cumbersome to access and manipulate for analysis.
- Estimates for low population areas often have very large margins of error.

## National Cancer Institute

The National Cancer Institute (NCI) produces population data that is specifically intended for use in calculating cancer incidence and mortality rates. The data are based on the U.S. Census Bureau's Population estimates program but also include bridged race (with the same methodology as the NCHS bridged race data) and makes specific adjustments for significant limitations in census data related to an under-count of native Hawaiians and the huge migration of people out of Louisiana in 2005 due to Hurricane Katrina. For Montana, the

NCI county level population estimates should be identical to the NCHS data as neither of these limitations apply to Montana data. However, unlike NCHS, NCI does plan to continue producing bridged race population files past 2020. Unfortunately, each subsequent year will not be released until it is needed for analysis of cancer surveillance data, more than 2 years later. Each central cancer registry submits their final data for a calendar year in November 2 years after the close of that year; 2021 data will be submitted in November of 2023. The data from each registry is cleaned and combined into a national data set that is typically released in the following spring and the needed population data will be released at the same time. As such, the 2021 population data will not be available from NCI until the spring of 2024. NCI county population data can be downloaded [here](#).

NCI also provides census tract level data. The census tract level population data are produced by Woods & Poole Economic Inc. (W&P) using a hybrid regression, demographic, and proportional model jointly developed by NCI, W&P, and the North American Association of Central Cancer Registries. The data includes annual population estimates by age group (0, 1-4 years, 5-9, 10-14, ..., 80-84, and 85 and older), race/ethnicity (Non-Hispanic White, Non-Hispanic Black, Non-Hispanic American Indian and Alaska Native, Non-Hispanic Asian Pacific Islander, and Hispanic of any race), and gender for the years 2006 to 2020. All the data are based on Decennial Census 2010 census tract boundaries. Further information on the data and a request form to access the data are available [here](#).

## Strengths

- Only source for bridged race estimates past 2020.
- Both county and census tract level estimates are available by age, race, ethnicity, and sex.
- Data is available in easy to manipulate formats for analysis or through SEER\*Stat statistical software.

## Limitations

- Estimates are released more than a year after many public health surveillance data are finalized.

## Recommendations

- For analysis of data from 2020 or previous years → Use 2021 vintage NCHS population data
  - This will provide consistency with previous policy and allow the maximum comparability of estimates.
- For analysis of data from 2021 forward → Use CEIC population data
  - If the analysis doesn't include stratification based on race this data should be comparable to the NCHS population estimates and we will maintain the ability to produce custom age groups at the county level.
- If the data includes data from both prior to 2021 and after → Use a combination of the 2 population data files to correspond with the appropriate years.

- If the analysis includes stratification by race be sure to note the change in methodology of racial groupings and possibly break the trend line. See more detailed discussion of analysis by race [here](#).
- If these population data files do not meet your analysis needs, i.e. you need data at the census tract level, you should feel free to use the population data that best fits your needs.
- If there is a national standard for what population data should be used for your surveillance data (as there is for cancer surveillance data) you should follow the national standard so that MT estimates are comparable to the national estimates and other states.