

# Census Concepts

## Census Geography

Most census data is summarized geographically so that the statistics describe the population of different places as opposed to individuals, households, or businesses. The Census Bureau publishes data for a number of legal and statistical areas. Legal areas such as states, counties, and municipal civil divisions exist as a matter of fact, are defined by charters and boundaries codified by law, and have governments that provide services to residents within those areas.

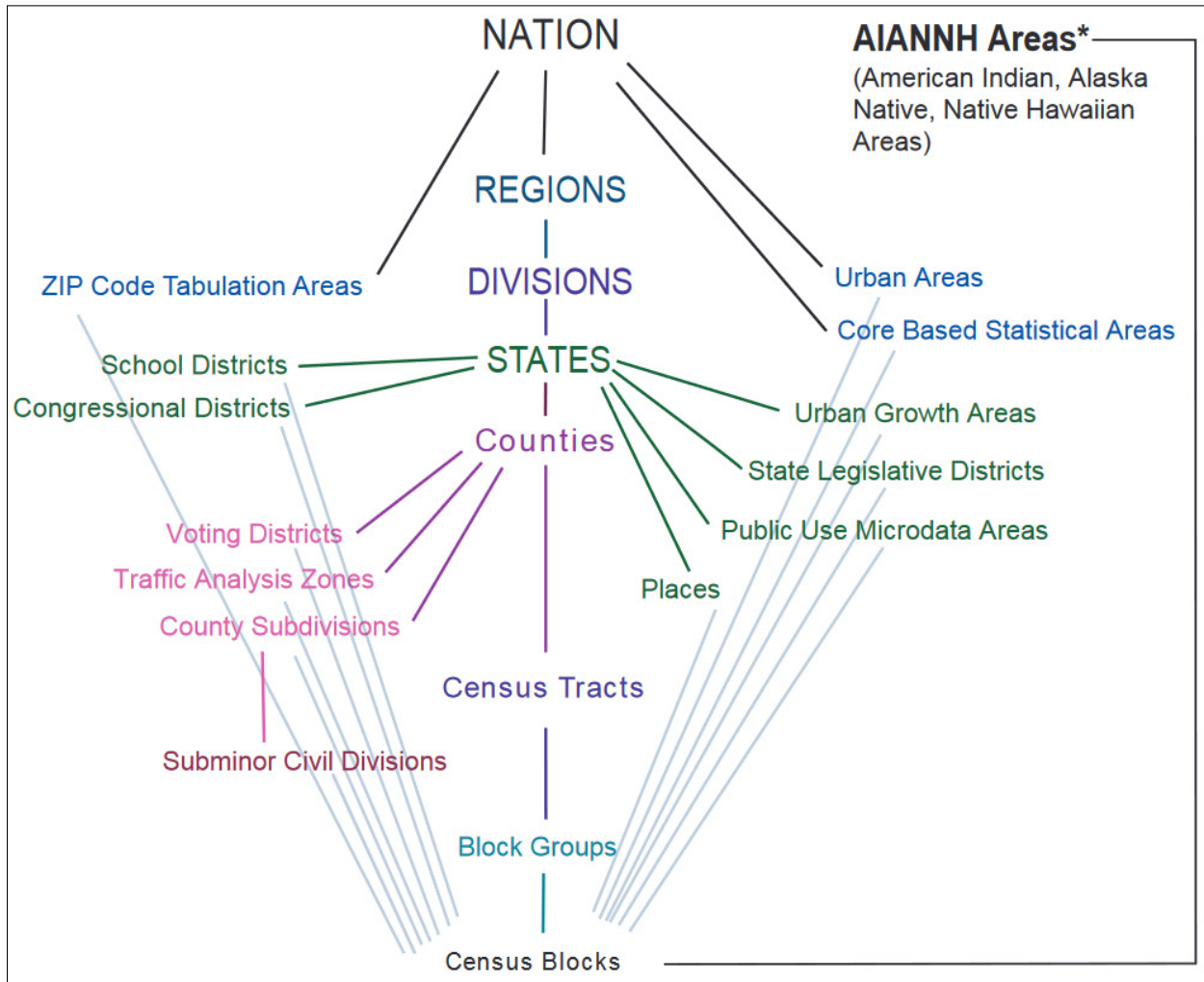
Statistical areas, such as census tracts, ZIP Code Tabulation Areas (ZCTAs), and Metropolitan Statistical Areas are defined by the Census Bureau or other government agencies for the sole purpose of presenting data for a specific purpose. For example, the Bureau delineates census tracts using legal boundaries and physical features such as roads and water bodies to create areas that are roughly equal in population size (a range of 1,200 to 8,000 people) for the purpose of making equal comparisons between places. The Bureau uses census blocks and its own methodology to approximate US Postal Service delivery areas as ZCTAs so that users will have access to data for this familiar geography. Metropolitan Areas are created by the Office of Management and Budget by aggregating counties to form functional socioeconomic areas based on population densities and shared commuting patterns. Some census geographies are hybrids of legal and statistical areas; the generic sounding Places geography consists of legally incorporated cities and towns and census designated places, which are concentrated population settlements that are identifiable by name but lack both governments and formal boundaries.

All of the census geographies fit within a hierarchy, where each is formed by smaller areas that nest within larger areas. The smallest areas are census blocks, whose boundaries are defined by physical features such as roads, railroads, and bodies of water.

Census blocks are aggregated to form larger statistical areas, such as census block groups, census tracts, and ZCTAs. Simultaneously, census block boundaries are constrained by legal areas such as counties and states and are designed not to cross these boundaries. Figure 3.1 illustrates how geographies are nested and constrained by each other. The vertical line that stretches from census blocks at the bottom to the nation at the top can be considered as the “primary trunk” of the census geographies, where the areas at the bottom nest within the areas directly above them without crossing their boundaries. For example, census tracts nest within counties and do not cross county boundaries, while counties nest within states. In contrast, there isn’t a line connecting tracts, places, and counties; census blocks nest within places, and places nest within states, which means that place boundaries may cross both tract and county boundaries.

This hierarchy poses practical considerations, the primary one being the selection of geographies a user makes when downloading data. Whether you use [data.census.gov](https://data.census.gov) or the Census Bureau’s API, when downloading data you can select areas one at a time or as a set if the areas nest. It’s possible to download data for all counties in a state or all states in the nation as they nest, and in some cases there are exceptions for skipping the middle geography if there is nesting above and below (e.g., all tracts in a state or all counties in the nation). Generally, it is not possible to download all places in a county or all tracts in a ZCTA as these areas don’t have a relationship with one another (places nest within states, and ZCTAs nest within nothing except the nation).

Each piece of census geography has a unique identifier called a GEOID. The long version of the GEOID contains codes that indicate the summary level that the geography is part of (its location in the census hierarchy), as well as an ANSI/FIPS code that indicates the specific geography. A short version of the GEOID consists of just the ANSI/FIPS code. For



**Figure 3.1**  
Hierarchy of census geographies (<https://www.census.gov/programs-surveys/geography/guidance/hierarchy.html>)

example, the full GEOID for Maricopa County, Arizona, is 0500000US04013. The first three digits, 050, indicate that the summary level is for counties. The additional zeros are reserved for a few special cases. The five digits after the US are the ANSI FIPS code, where 04 is the unique ID for the state of Arizona and 013 represents Maricopa County in Arizona. These codes allow users to relate data from different tables and to join data to boundary files in GIS (GIS is discussed in more detail in chapter 6).

The choice of geography for any analysis is going to be determined by its appropriateness of use, the availability of the data, estimate precision versus geographic detail, and limitations imposed by any external data used in conjunction with census data. For example, studying population by state would make sense if you are comparing trends between states or if the purpose of the study is to illustrate how distinct policies and laws in each state impact population

trends. Conversely, if the object is to study the distribution of the population across the country, states would be a poor choice as they vary significantly in size, shape, and population. Counties would be a better choice as they are smaller and more numerous. Likewise, using census tracts to study trends within a county or an urban area would be a better choice than using ZCTAs, as tracts have an equal population size and logical boundaries while ZCTAs do not. Despite their seeming familiarity, ZCTAs and ZIP codes are a challenging geography to work with and should be avoided when possible (Donnelly 2020). Data is not always available for all geographic areas; for example, block data is available only in the DEC and only for certain tables. Administrative data from other government agencies may be summarized only for states, counties, and metro areas, which could limit the choice of census data for your analysis. In datasets such as the ACS, there is a trade-off between geographic detail

and the precision of the estimate; the smaller the area, the less precise the estimate will be.

A final consideration is that census geography changes over time. The statistical areas created by the Census Bureau are redrawn every ten years prior to each DEC. Boundaries of blocks change as the physical and built environment changes, and blocks are often renumbered. Census tracts are designed to be as consistent as possible over time, but as they are defined by population size, they will be split as populations grow, aggregated as populations shrink, and redrawn to fit the changing landscape. Once they are redrawn, the census statistical areas remain relatively static and are modified only to correct boundary errors or to conform to changes in legal areas. Legal areas can change at any time as towns and cities acquire new land, incorporate, or unincorporate. The DEC captures what existed on the date it was taken, while annual ongoing programs such as the ACS and Population Estimates incorporate any changes that happened in a given year. We will discuss historical data and making comparisons over time in chapter 6. Researchers can explore the different geographies for their area of interest using TIGERweb, an interactive web map.

## Geographical Reference

*TIGERweb*

<https://tigerweb.geo.census.gov/tigerweb/>

*ANSI FIPS Codes*

<https://www.census.gov/library/reference/code-lists/ansi.htm>

## Subject Categories

Census data is also summarized by different categories of people and housing units. While the terms for many of these categories, such as household and family, may seem commonplace, in the census universe they have highly specific definitions. It's important to have an understanding of these terms in order to make informed choices, such as whether a measure of household income or family income would be appropriate for a particular analysis. This section summarizes the most salient categories: households and families, group quarters, age, sex, race and ethnicity, and housing units.

The entire population is subdivided into two large categories based on living arrangements. Most Americans live in households, which consist of one or more individuals who live together in a self-contained residential setting. Households are subdivided into family

households and nonfamily households; the term *families* is often used to simply refer to the first group. Families consist of at least two or more people who live in a residential setting and who are related to one another by blood, marriage, or adoption. In contrast, nonfamily households consist of people who live alone, nonmarried partners, roommates, and any situation where people living in a household setting are not related. In the 1950 census, 80 percent of all households were family households, whereas by the 2010 census, only 50 percent were. To answer our earlier question in light of this fact, using household income would be more appropriate when trying to generally measure income across the United States, whereas family income is pertinent only for studying that specific group of people who constitute families. One of the biggest changes in the 2020 census (and subsequent versions of the ACS) is that same-sex married couples are now counted as families. Prior to 2020, same-sex marriages were not explicitly tabulated, and all same-sex partnerships were counted as nonfamily households.

The population not living in households live in group quarters, a nonresidential setting where many unrelated people live together and share a common living space. Group quarters is subdivided into two groups. The institutionalized population lives in facilities where they have been committed to the institution and they are not permitted to come and go freely. This includes penitentiaries, psychiatric hospitals, and nursing homes. In contrast, the noninstitutionalized population lives in shared living quarters by choice for a common purpose. This includes military barracks, college dormitories, monasteries and convents, and homeless shelters. For certain types of analysis it's important to identify and exclude the institutionalized population, as they don't participate in the local economy. For both categories, the presence of this population can have an outsize influence when studying small communities or geographies, as these facilities tend to concentrate many people with shared characteristics in the same place. The opening or closing of a group-quarters facility can have a large impact on population change in small areas.

Most of the census datasets are cross-tabulated by age, sex, and race and ethnicity, and there are many data tables devoted specifically to these topics. Data on age is reported in various cohorts, such as one-, five-, and ten-year brackets, as well as by special age categories that have meaning based in the law or on the circumstances for which the data is reported, such as the population under eighteen or aged sixty-five and older. The Census Bureau does not summarize age data using generational categories such as millennial or generation X. Data for sex is defined as basic biological or anatomical sex as male and female. The census does not include any questions or data related

**Table 3.1.** Different approaches for reporting race, Clark County, Nevada, 2015–2019 ACS

Race/Ethnicity	Race Alone, as Reported	Hispanic as a Race
White	60.2%	42.8%
Black	11.7%	11.2%
American Indian/Alaska Native	0.9%	0.5%
Asian	9.7%	9.6%
Pacific Islander/Hawaiian Islander	0.8%	0.7%
Some Other Race	11.5%	0.4%
Multiracial	5.4%	3.8%
Hispanic/Latino	(of the total, 31.1%)	31.1%

Note: The margin of error for each group ranges from 0.1% to 0.3%.

to gender or sexual identity, although there has been ongoing debate on whether it should or will in the future (Wang 2018b).

Race was the first characteristic ever tabulated for the decennial census, and the categories have evolved over time as the nation has changed (Humes and Hogan 2009). These categories are defined in federal law by Directive 15, which was designed to ensure that all federal agencies use consistent racial and ethnic definitions when collecting and publishing statistics (Office of Management and Budget 1997). The original 1977 directive defined four racial categories: White or Caucasian, Black or African American, American Indian and Alaskan Native, and Asian, Pacific, or Hawaiian Islander. A special, separate ethnic category was established for Hispanic or Latino. The 1997 revision of the directive split Asians and Pacific or Hawaiian Islanders into two separate categories (making five racial groups), and an option for people to identify as multiracial was incorporated by allowing people to check multiple race categories on questionnaires.

The 1970 census marks the beginning of the racial categories as they are used today, even though it was conducted before Directive 15 was published. This was the first census where Americans were able to self-identify their race, as opposed to having to follow rules for how the question should be answered. It was also the first census where the majority of Americans received and returned the decennial census form by mail, as opposed to being visited and interviewed by a census enumerator. While the wording of the categories is different, they roughly align to the 1977 directive. The 1970 question about Hispanic or Latino ethnicity was hastily added to a small 5 percent sample form (Humes and Hogan 2009, 119), and the results are generally regarded as an undercount. The 1980 census was the first census that included all the racial and ethnic categories as part of the 100 percent count, and the 2000 census was the first to incorporate the 1997 guidelines that allowed for the tabulation of multiracial characteristics. Multiracial characteristics are tabulated in a number of different ways; in Race Alone tables, where the population is counted in

single-race categories along with a summary category for anyone who is of multiple races; and in total race tables, where the total number of people who identified as a particular race is counted.

One source of confusion for many data users is the treatment of the Hispanic and Latino population as an ethnicity as opposed to a racial group. There are two separate questions on both the DEC and ACS census forms that ask (1) What is your race? and (2) Are you Hispanic or Latino? This means that every person who is Hispanic or Latino is counted in one of the five racial groups or within a sixth optional Other race category. There are separate data tables that count people by race, that count by Hispanic and Latino ethnicity, and that cross-tabulate the two.

There are several ways that this data is commonly presented, in everything from research studies to news reports. The first approach presents the data as it was collected in the Race Alone tables. Data on race is presented for each of the five racial groups, the Other category, and a multiracial category for every person who selected more than one race, with a footnote indicating that X percentage of the total population is Hispanic or Latino. The second approach adapts the published data to the categories more commonly used in society. A race/ethnicity table where the two characteristics are cross-tabulated is used, where any person who identified as non-Hispanic is counted based on their race alone, while all Hispanic people are counted as Hispanic/Latino regardless of their identified race. In yet another iteration, the census includes Race Alone tables for many different variables for each individual race, labeled A through I. These tables include a Hispanic or Latino table, and a table for non-Hispanic Whites to separate Caucasian Whites from Hispanic Whites (as White is the most common racial category chosen by Hispanic or Latino people).

Table 3.1 illustrates the differences between the first two methods, using 2015–2019 ACS data for Clark County, Nevada, that shows the percentage of the population by race. Using the first method, data for race alone is presented as it is published, with the seven categories adding to 100 percent. A footnote

appears at the bottom indicating that 31.1 percent of this total population is Hispanic or Latino. Using the second method, Hispanic or Latino is treated as an eighth category, where people who identified as Hispanic are counted as such, regardless of their race, and the race categories represent non-Hispanic people of each race alone. Note the large decline in the White population and the near disappearance of the Other race population, as the majority of Hispanics and Latinos identify their race as one of these categories (Terry and Fond 2013).

There are arguments in favor of and against each of these approaches, and in favor of and against making Hispanic/Latino an actual race instead of a special ethnicity. After careful study of how this population responds to census forms, the US Census Bureau (2017) proposed recategorizing Hispanic and Latino as a race as opposed to an ethnicity. A number of other adjustments to racial categories were being actively considered prior to 2020 (Strmic-Pawl, Jackson, and Garner 2018). The Office of Management and Budget chose not to act on any of these proposals (Wang 2018a), and thus the racial categories used since the year 2000, and in some respects since 1970, will continue to be used for the 2020s. It remains to be seen how long these categories will remain relevant in the census and throughout society at large; one of the most significant findings in the 2020 census was the large increase in the multiracial population, which includes people who selected multiple categories (Frey 2021; Wang and Talbot 2021).

Lastly, the census also includes counts and characteristics of housing units. A housing unit is an individual, self-contained domicile so that an individual apartment or condo is counted as a single unit, equal to a single-family home. Buildings that are derelict, without a roof, door, or windows, are not counted as units. Units are counted as occupied if they are currently inhabited at the time of the count or survey, or vacant if they are not. Vacant units are classified based on their status: they are seasonal vacation units, are empty while for sale or rent, are empty as they were just constructed, or are unoccupied for some other reason. Occupied housing units are subdivided into owner-occupied and renter-occupied units under a concept called “tenure.” Occupancy status and tenure are counted in both the DEC and ACS. The ACS captures additional, detailed characteristics about housing.

## Tables and Universes

Census data is published in a series of tables of varying breadth and depth, where variables are grouped together based on characteristics they share. At the broadest level are the data profile tables, one for the

DEC and four for the ACS, categorized as social, economic, housing, and demographic data. These tables capture a selection of the most commonly sought variables in these datasets and are a good place to start for users who wish to familiarize themselves with the content of the census. Below the profiles are subject tables (referred to as quick tables in the 2010 census), which are greater in number. These tables include a narrower selection of related variables. For example, there is an income table with data on several measures of income (median, means, per capita, intervals) for households, families, and persons, cross-tabulated by age, sex, race, employment, and other variables. Both the data profile and subject tables include counts and percent totals.

Below the subject tables are the detailed tables, which are the narrowest and most specific. For example, a table on median household income includes nothing but that specific value, while separate tables capture the other variables that were published in the single subject table (household income with counts of households by income bracket, household income by race, etc.). The detailed tables contain only counts, and no percent totals.

The different groups of tables are named with prefixes that indicate the type of table, followed by a unique table number. The data profile tables begin with the letters DP (DP1 in the DEC and DP02 to DP05 for the ACS), while the subject (S) and quick tables (QT) have their own designations. Detailed tables in the DEC begin with a P or an H, indicating whether the table is for population or housing. The ACS detailed tables have a number of prefixes, the most common ones being B for Base table and C for Collapsed Table (a version of B with fewer categories). The Population Estimates and business tables have their own naming conventions.

Most of the graphic user interface (GUI)-based tools such as data.census.gov are built for accessing tables and don't allow you to select individual variables from multiple tables. When retrieving census data, users can select a broad number of variables in one or two tables and then subsequently narrow them down by extracting or deleting columns, before or after download. Alternatively, one can select several narrow, detailed tables with just the variables of interest and subsequently stitch the tables together into a whole after downloading. The API allows you to create a targeted selection of variables across many tables, but some knowledge of table structure is necessary for creating requests.

Each table is published for a specific subset of the population, referred to as a universe, that is relevant for that specific topic. For example, the universe for the population enrolled in school is the population aged three and above, as children younger than three wouldn't be attending school. The universe for housing

SCHOOL ENROLLMENT BY LEVEL OF SCHOOL FOR THE POPULATION 3 YEARS AND OVER		
Survey/Program: American Community Survey	Product: 2019: ACS 5-Year Estimates Detailed Tables ▾	
TableID: B14001	Universe: Population 3 years and over	
Wyoming		
Label	Estimate	Margin of Error
▾ Total:	559,774	±761
▾ Enrolled in school:	144,936	±1,606
Enrolled in nursery school, preschool	9,079	±619
Enrolled in kindergarten	7,213	±710
Enrolled in grade 1 to grade 4	31,568	±1,041
Enrolled in grade 5 to grade 8	31,236	±1,045
Enrolled in grade 9 to grade 12	28,319	±702
Enrolled in college, undergraduate years	29,971	±1,099
Graduate or professional school	7,550	±667
Not enrolled in school	414,838	±1,743

**Figure 3.2**  
Table universes illustrated with school enrollment in Wyoming, 2019 ACS

tenure (owner- or renter-occupied) is occupied housing units, as the tenure status of vacant housing units cannot be determined. It’s important to scrutinize the universe to avoid drawing false conclusions; the universe listed for any given table generally appears as “The Total,” but it should never be presumed that this total represents all people or all housing units. Figure 3.2 illustrates this point, with the school enrollment table from the ACS for Wyoming as depicted in data.census.gov. Note the universe is listed as the population aged three and above; the total at the top of the table refers to this population.

## References

Donnelly, Francis P. 2020. “The Trouble with ZIP Codes: Solutions for Data Analysis and Mapping.” *At These Coordinates*, May 11, 2021. <https://atcoordinates.info/2020/05/11/the-trouble-with-zip-codes-solutions-for-data-analysis-and-mapping/>.

Frey, William H. 2021. “New 2020 Census Results Show Increased Diversity Countering Decade-long Declines in America’s White and Youth Population.” *Brookings Metro*. August 13, 2021. Washington, DC:

The Brookings Institution. <https://www.brookings.edu/research/new-2020-census-results-show-increased-diversity-countering-decade-long-declines-in-americas-white-and-youth-populations/>.

Humes, Karen, and Howard Hogan. 2009. “Measurement of Race and Ethnicity in a Changing Multicultural America.” *Race and Social Problems* 1: 111–31.

Office of Management and Budget. 1997. “Revisions to the Standards of the Classification of Federal Data on Race and Ethnicity.” *Federal Register* 62, no. 210 (October 30): 58782.

Strmic-Pawl, Hephzibah V., Brandon A. Jackson, and Steve Garner. 2018. “Race Counts: Racial and Ethnic Data on the U.S. Census and Implications for Tracking Inequality.” *Sociology of Race and Ethnicity* 4, no. 1: 1–13.

Terry, Rodney L., and Marissa Fond. 2013. “Experimental U.S. Census Bureau Race and Hispanic Origin Survey Questions: Reactions from Spanish Speakers.” *Hispanic Journal of Behavioral Sciences* 35, no. 4: 524–41.

US Census Bureau. 2017. *2015 National Content Test: Race and Ethnicity Analysis Report*. February 28. Washington, DC: US Department of Commerce. <https://www.census.gov/programs-surveys/decen>

- nial-census/decade/2020/planning-management/plan/final-analysis/2015nct-race-ethnicity-analysis.html.
- Wang, Hansi Lo. 2018a. "Census to Keep Racial Categories Used in 2020." NPR. January 26. <https://www.npr.org/2018/01/26/580865378/census-request-suggests-no-race-ethnicity-data-changes-in-2020-experts-say>.
- . 2018b. "U.S. Census to Leave Sexual Orientation, Gender Identify Questions Off New Surveys." NPR. March 29. <https://www.npr.org/sections/the-two-way/2017/03/29/521921287/u-s-census-to-leave-sexual-orientation-gender-identity-questions-off-new-surveys>.
- Wang, Hansi Lo, and Ruth Talbot. 2021. "This Is How the White Population Is Actually Changing Based on New Census Data." NPR. August 22. <https://www.npr.org/2021/08/22/1029609786/2020-census-data-results-white-population-shrinking-decline-non-hispanic-race>.