[DC-06-028] United States Census Data

Abstract

The Census Bureau collects extensive numeric data on the residents of the United States as well ast the national economy. This is accomplished both through a decennial census as well as numerous other more frequent surveys. The decennial census is a fundamental basis of American democracy, mandated by the U.S. Constitution and essential for the equal representation in a democratic government. Numeric census data are maintained in vast collections of tables and organized at many different levels of geographies. From the Census website, the geographic and tabular data can be downloaded and then joined for display and analysis within a GIS. Because of the nature of individual data aggregated over areas and other matters, care must be taken to avoid statistical errors when undertaking spatial analyses.

Keywords: block groups, Census data, Census geography, population, tracts

Author & citation

Castagneri, J. (2019). United States Census Data. The Geographic Information Science & Technology Body of Knowledge (1st Quarter 2019 Edition), John P. Wilson (Ed.).
DOI: 10.22224/gistbok/2019.1.8.

Explanation

1. Definitions
2. Introduction
3. Census Geography
4. Census Data
5. Accessing and Using Census data within a GIS

1. Definitions

Participant Statistical Areas Program (PSAP): The 2020 Census Participant Statistical Areas Program (PSAP) allows local planning officials to review and update selected statistical area boundaries for decennial Census data tabulation following U.S. Census Bureau guidelines and criteria. The Census Bureau also will use the statistical areas defined for the Census to tabulate data for the annual American Community Survey (ACS) estimates and the Economic Census. There are two types of statistical geographies for review under the 2020 Census PSAP, standard and tribal. Tribal statistical geographies include additional tribal tracts and blocks groups plus other state level statistical areas for tribes without a federally recognized reservation. This topic covers standard statistical areas only.

Census Designated Places (CDP): CDPs are statistical geographic areas representing closely settled, unincorporated communities that are locally recognized and identified by name. CDPs are the statistical equivalents of incorporated places, with the primary differences being the lack of both a legally-defined boundary and an active, functioning governmental structure, chartered by the state and administered by elected officials. The

primary goal of the CDP program is to provide meaningful statistics for well-known, unincorporated localities.

Census Tracts (CT): The primary goal of the CT program is to provide a set of nationally consistent small, statistical geographic units, with stable boundaries, that facilitate analysis of data across time. Ideally, the boundaries of a CT remain the same between censuses making it possible to compare statistics for the same geographic area from decade to decade. Tracts have an optimum population of 4,000 persons and are subdivided once every ten years when growth within the tract exceeds 8,000 persons.

Block Groups **(BG):** BGs are statistical geographic divisions of a CT defined for the tabulation and presentation of data from the decennial census and selected other statistical programs. BGs are also used to tabulate and publish estimates from the ACS. BGs provide the geographic framework within which the Census Bureau defines and identifies census blocks. Each BG comprises a reasonably compact and contiguous cluster of census blocks; up to ten BGs can be contained within a single CT (up to nine standard BGs and one water BG). Block groups have a minimum population criterion of 600 persons and a maximum population of 3,000.

Blocks: As the smallest level of geography for which the census bureau collects and publishes data, census blocks are created by the Census Bureau every ten years after the re-definition of census tracts and block groups by local planning agencies as part of the PSAP. The nesting relationships of these geographies are crucial to spatial analysis. Census tracts must nest within counties, block groups nest within tracts, and census blocks nest within block groups. In this statistical hierarchy, there are no gaps and no overlaps in coverage.

2. Introduction

The decennial census is a fundamental basis of American democracy. The U.S. Constitution states in Article I, Section 2; “Representatives and direct Taxes shall be apportioned among the several States which may be included within this Union, according to their respective Numbers…” Since 1790, it has been the goal of every decennial census to count the entire resident population of the United States at the correct and current location where each person lives, and to count each person only once.

Census data helps explain the basis for equal representation in a democratic government. In addition to the decennial census, the Census Bureau conducts 130 surveys annually providing key indicators for our economy and health and welfare of our resident population. Businesses use census data and GIS to help locate stores and factories and to understand the demographics of the consumer. State and local governments use GIS and census data to plan communities and services based on age, income, and other demographic variables. Health officials use census data to establish health services and plan for effective programs providing aid to the underserved and uninsured.

The vast trove of geographic and demographic data released annually by the U.S. Census Bureau can be explored and analyzed readily through GIS. By linking decennial or American Community Survey (ACS) data with its corresponding geography within GIS, we begin to understand the landscape of our population and economy. Issues such as

redistricting, supply and demand, distribution of poverty, segregation, environmental risk, transportation planning, and many others become possible.

All census geographic and demographic data are free and available online from the Census Bureau's website at www.census.gov. In this topic, we will explore the basic information necessary to import and analyze census data within GIS. Before we can comfortably explore census data, we must also understand census geography.

3. Census Geography

The term "census geography" generally refers to that class of geographic information that is uniquely relevant to the Census Bureau. The fundamental building block of all census geography is the census block, appearing at the bottom the hierarchy (Figure 1). All other census geographies are built upon this basic polygon unit. A census block usually equates to a city block in urban areas. Census blocks are bounded by physical features such as roads, railroads and streams. They are also contained by city limits, county boundaries, congressional districts, census tracts, etc.

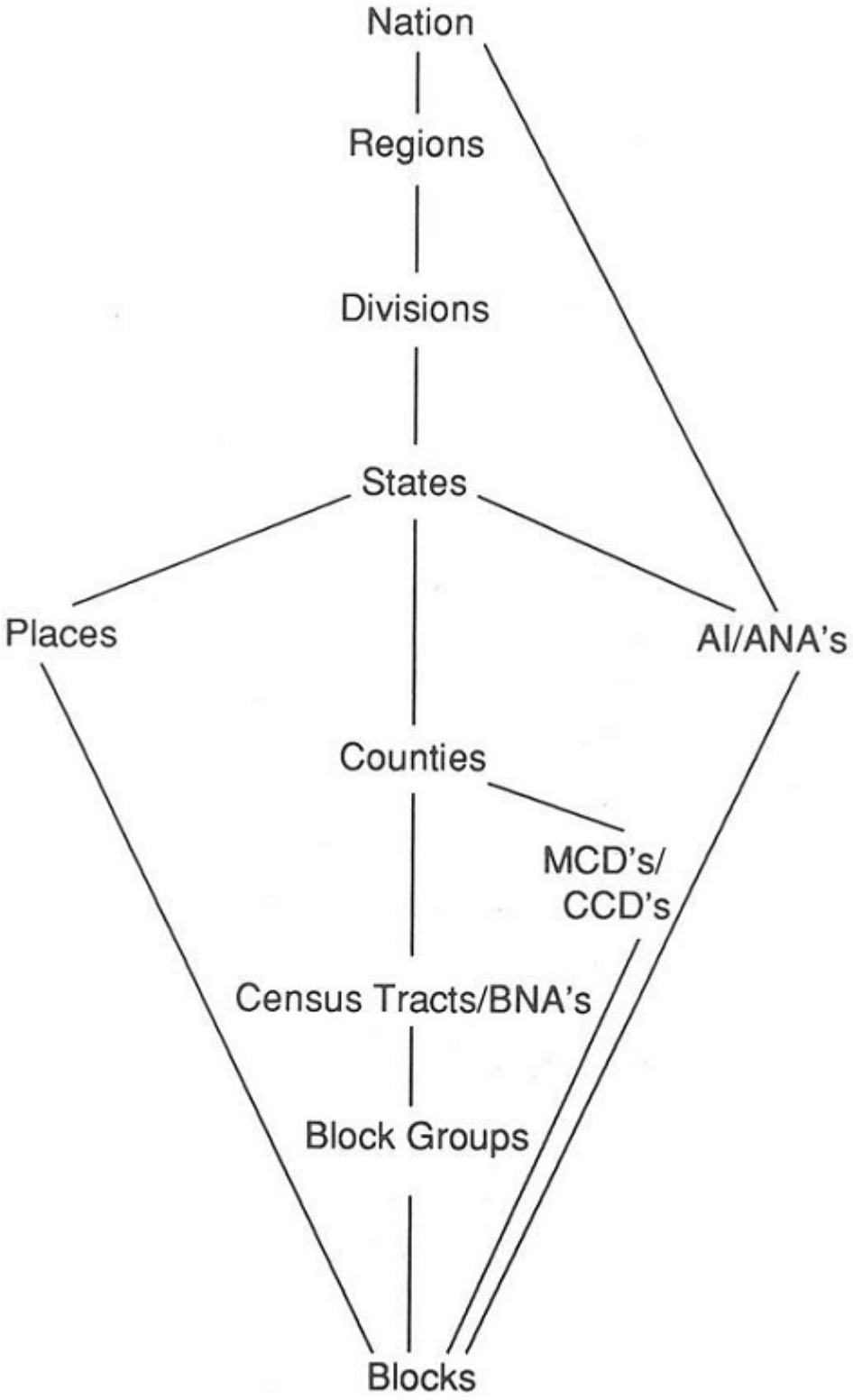


Figure 1. United States Census Geographies. Source: U.S. Census Bureau.

This leads us to another fundamental tenet of the Census, that all Census Geography can be broken down into two subclasses: 1) Political Geography and 2) Statistical Geography.

3.1 Political Geography

Political Geography includes entities that have functional governments with elected officials

and have the authority to enact local laws. These political entities may also have the ability to hold elections, incorporate or annex new land, and to tax their residents to provide services. Political geographies include:

- American Indian Reservations and tribal subdivisions
- States
- School Districts
- Counties, Boroughs, and Parishes
- Cities & Towns
- Minor Civil Divisions (MCDs) or Townships

Other political geographies include voting districts and school districts.

3.2 Statistical Geography

Statistical Geographies are those non-political geographic areas defined solely for the purposes of the demographic analysis of decennial census and American Community Survey (ACS) data. These geographies are defined every ten years as part of the decennial census through what is called the Participant Statistical Areas Program (PSAP). These areas include:

- Census Designated Places
- Census Tracts
- Block Groups
- Blocks

3.3 Topologically Integrated Geographic Encoding and Referencing (TIGER)

Along with a variety of linear features, all political and statistical geographies are stored in the Census Bureau's TIGER system. Developed in the 1980's in collaboration with the United States Geological Survey, and recently re-engineered as "MAF/TIGER," this massive system stores a variety of geospatial information including addresses, roads, highways, railroads, streams and rivers in addition to boundaries for political and statistical areas (Figure 2).

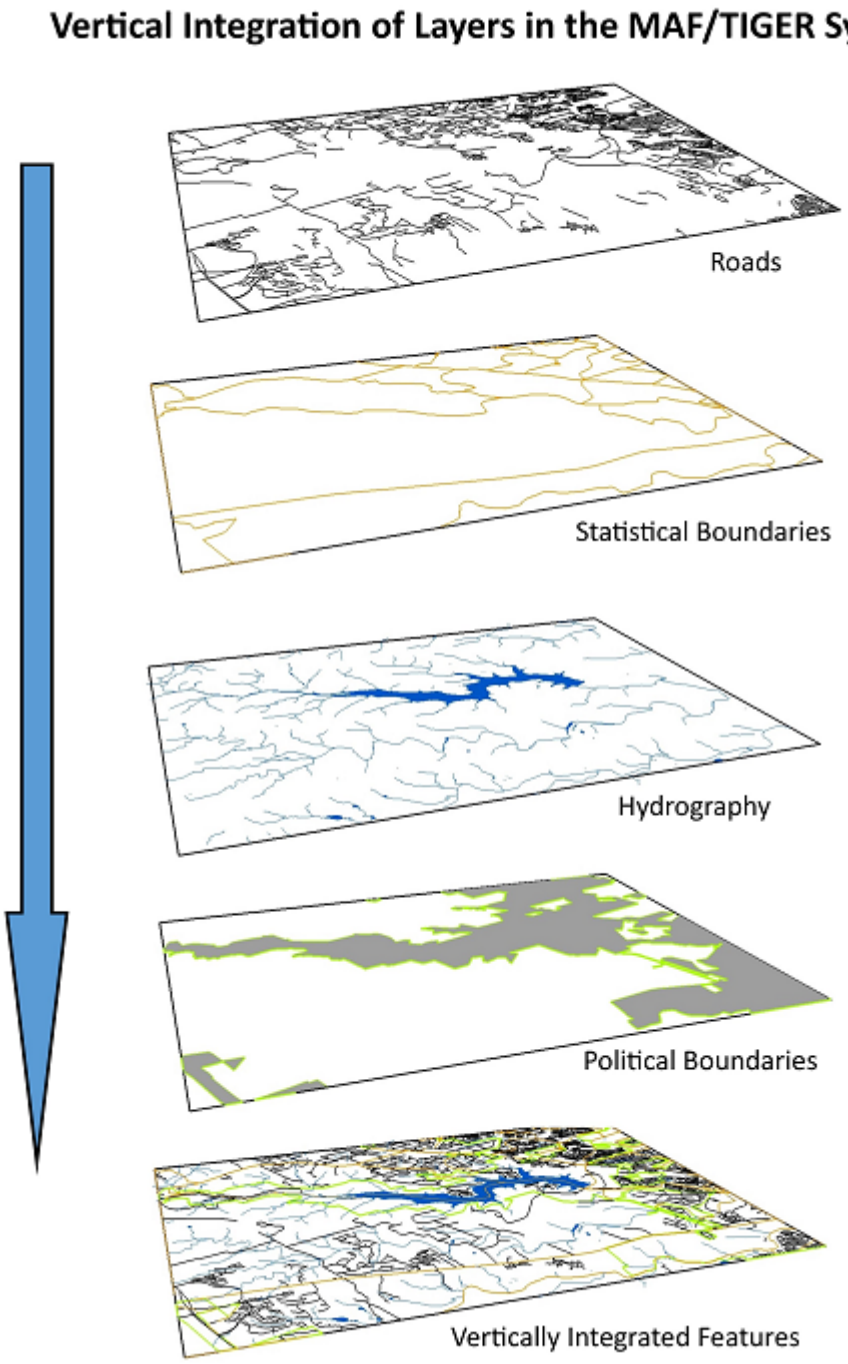


Figure 2. The MAF/TIGER system is designed to vertically integrate digital representations of spatial data.

Each of these features is interrelated through a set of topological rules that support and enforce a variety of spatial relationships critical to the Census Bureau and its data products. This hierarchical network of geography and the topological rules ensure that every square meter of the country is covered by CTs, BGs, and blocks. This wall-to-wall coverage is key to the utility of statistical geography as a spatial analysis framework (Figure 3).

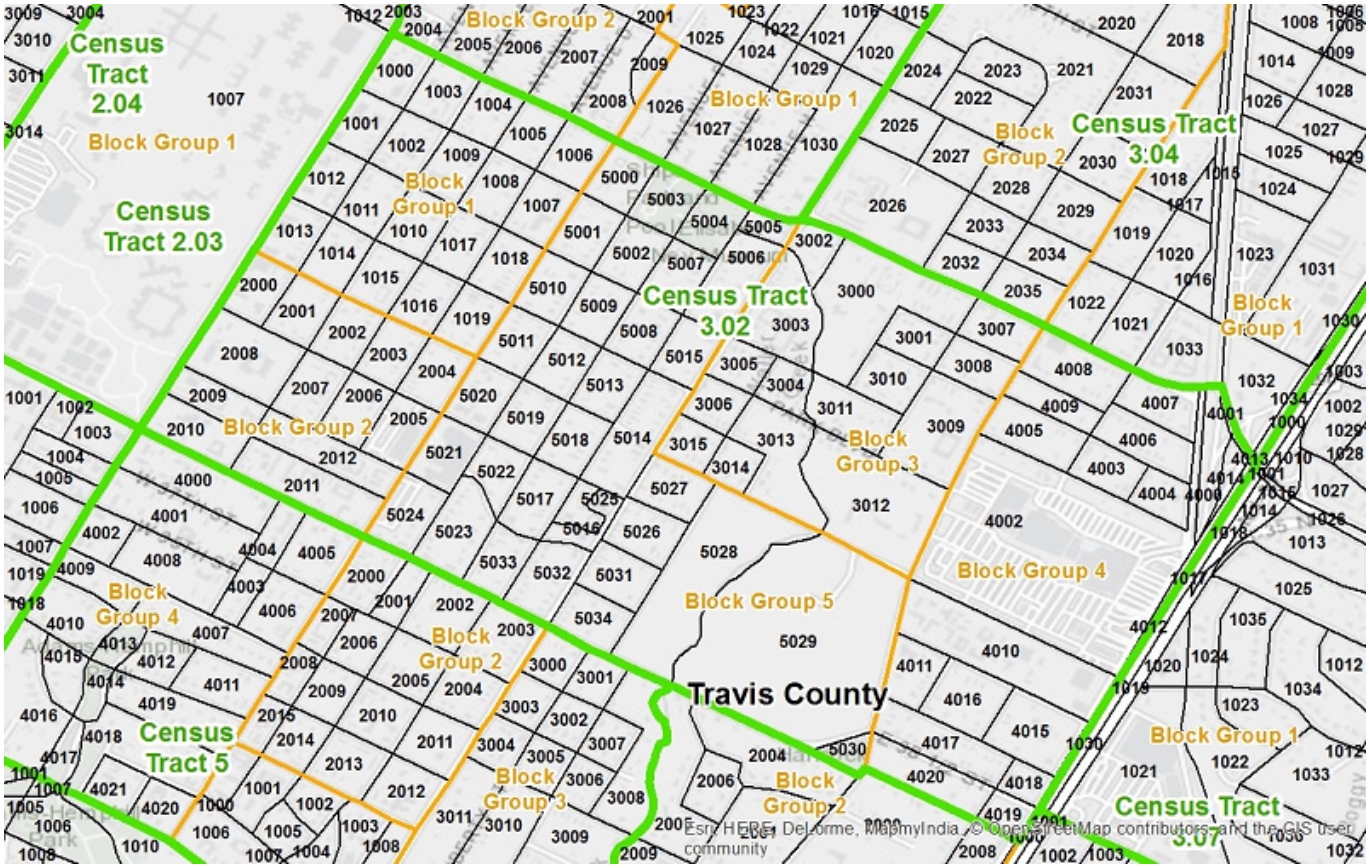


Figure 3. Nesting Relationships of Statistical Geographies. Source: author.

4.0 Census Data

When exploring the notion of pairing Census data with GIS, the volume of demographic data can be intimidating. The primary census data retrieval tool, currently American FactFinder, is context sensitive by geography. Thus it is helpful to choose your geography of interest first, then use FactFinder to determine what census data are available for that geography. Or, if your data variable is predetermined, FactFinder will display only the available geography for that particular variable. Once you determine the data of interest, you must decide on the level of geographic detail for your analysis. Generally speaking, statistical geography provides the best framework for spatial analysis of census data. While states and counties can provide useful distribution patterns, smaller geographies such as census tracts and blocks are better suited for sub-state or sub-county analysis.

To help understand it all, we can break Census data down into different classes of information. U.S. Census data fall into three general categories: People, Housing, and the Economy.

4.1 People and Housing

Data about people includes information on age, gender, educational enrollment and attainment, marital status, mobility, income, spending habits, jobs, disability, health care, commuting patterns, race & ethnicity and more. Housing is another general category of census data that includes number of occupants, structure type, structure age, plumbing &

kitchen facilities, vehicles available, housing value, insurance costs, heating & cooling, Internet connectivity, rent paid and more.

- **Decennial Census** data have been collected once every ten years since the first census of 1790, and this is the only full count of the population. These data are collected from every household in the United States and are used to meet the requirements of the Census Act (Title-13) within the constitution of the United States. Specifically, decennial census data are used to reapportion the 435 seats of the U.S. House of Representatives. Because it is a complete count of every household in the nation, data from the decennial census are available for all political and statistic geographies down to the census block level.
- **Sampled Census Surveys** are collected semi-regularly from a sample of the population for the purpose of supporting federal agency program requirements. Dozens of surveys are conducted by the Census Bureau each year, including the American Housing Survey (AHS), the Current Population Survey (CPS), and the American Community Survey (ACS). Data from these surveys are said to provide characteristics of the population since they are not derived from a complete enumeration, which only the decennial census provides.

The American Community Survey (ACS), first launched in 2005, is conducted annually from a nationwide random sample determined at the census tract level. It is the only annual survey that provides demographic characteristics below the county level. In fact, the ACS by itself is a treasure of over 4,000 population, housing, and demographic variables available down to the block group level. Add to this fact that the data are released annually, and the ACS is the most important data source for demographic analysis offered by the Census Bureau.

In 2017, the ACS sample was over 2.1 million housing units. The design of the ACS allows for annual releases of data from for any geography over 65,000 population. These data releases are referred to as **1-year data**. The **5-year data** are released annually based on a rolling 5-year average of each data variable and are available down to the block group level. For a discussion on the differences of ACS data, see When to use 1-year, 3-year or 5-year estimates on the Census ACS website.

4.2 Economy

Aside from population demographics and housing data, the Economic census provides data on the number of businesses by type and size, revenue, profits, ownership & history, output, imports and exports, and more. To measure the economic health of our country, the Economic Census is conducted once every 5 years by surveying 4 million businesses.

5. Accessing and Using Census Data in GIS

5.1 Locating and Downloading Census Data

All census data products are available from the Census Bureau's website (www.census.gov). To use Census data within a GIS, both the geographic data (the boundaries for the polygons of Census geographies, the geometric shapes) and the numeric data itself (in tables) must be available for **joining** within a GIS. The geographic data are

available to be downloaded as TIGER/Line files. For virtually all other non-spatial census data, the **American FactFinder** is the primary data retrieval tool. Information on how to use American FactFinder can be be found here: https://factfinder.census.gov/help/en/index.htm#

Once the correct table is identified and displayed using FactFinder, choose the "Download" action item near the top of the white tableview tab page. To keep track of what has been downloaded, it is recommended to have the annotations and data in a single file and to include descriptive data element names (Figure 4).

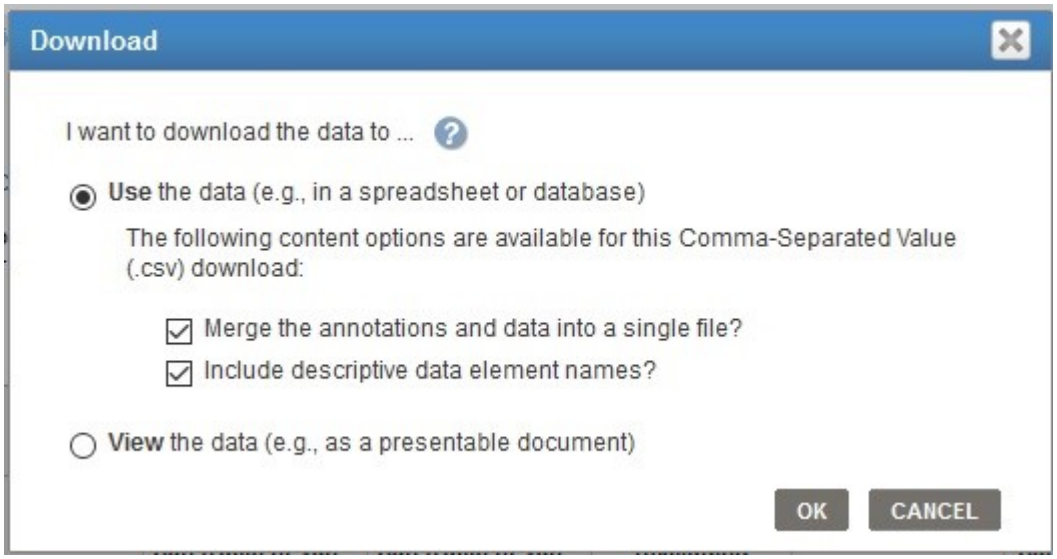


Figure 4. An example of the download dialog window from American FactFinder. Source: author.

Clicking OK, the system will ZIP four files together for download (Figure 5).

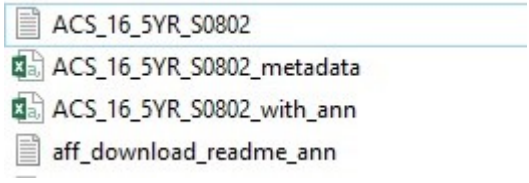


Figure 5. Sample of the collection of files that might be downloaded with Census data. Source: author.

Becoming an effective and efficient user of Census data requires confidence and competence in deciphering file names. In this particular example, the "ACS_16_5YR_S0802_with_ann" file contains the actual data, along with two header records. "ACS_16" means that these downloaded data are from the 2016 American Community Survey. The "5YR" indicates that the values are actually from a 5-year rolling average, so values averaged from the samples from 2012 - 2016. The "S0802" is the specific code for the data question of "what means of transportation people use to get to work." Another downloaded table, "ACS_16_5YR_S0802", will have metadata (field names

and descriptions) that explain the coded values of the data itself. It's a good idea to import the main data into a spreadsheet and the metadata into a separate tab for reference.

For those wanting programmatic access to a majority of census data tables, the Census API provides quick access to thousands of tables (including the ACS). By making customized requests via the API, users can retrieve customized and pre-formatted census data ready to use in GIS.

5.2 Joining Census Data to Census Geography

Once the desired tabular census data and the corresponding geography have been selected and downloaded, the matter of joining the data in GIS is predicated on a common data field in both datasets. This is often referred to as the "link" or "join" field. Both the GIS file and the attribute table containing census data to be joined must have identical fields with the same data type. The Census Bureau had this in mind and created a link field in all TIGER/line shapefiles and geodatabases called "**GEOID."** For joining census tracts to tract level ACS data, this field consists of a concatenated value that contains the 2-character state FIPS codes, 3-character County FIPs Code, and 6-character tract code. The resulting string will look similar to this: 08059002100, values that look like numbers but are stored and handled as text. These same link fields exist within the header of the demographic data when downloaded from the American FactFinder.

Cleanup and preparation will almost always be necessary before you perform the join execution within a GIS, particularly with data that have just only freshly been downloaded. Typically, from the data table you must delete the first header row and select the second row as the header record. You must then clean the header record of illegal characters, spaces, field names longer than 8 characters. Save the file as a new comma separated values (.csv) and/or a spreadsheet. The "join" process itself will vary based on the GIS program being used, and you should follow the steps indicated for this common process in your platform of choice.

5.3 Mapping and Analysis Considerations

Once the join is accomplished, normalizing the data may be necessary. Generally speaking, census tracts are designed to be directly comparable to one another without data normalization. However, some ACS data variables may require further consideration. For example, persons per household is not the same as persons per housing unit. These rates are using two different denominators: housing units and households. The data user should be careful to understand these denominators when creating a multivariate analysis of disparate variables.

In another example, population densities can be calculated using TIGER area measurements. However these area measurements are in square meters not square miles or square kilometers. Add to this the fact that many tracts contain large swaths of uninhabitable land and may skew the results significantly.

The Modifiable Areal Unit Problem becomes particularly evident when working with Census data, and matters of scale and zoning should be considered thoughtfully and carefully when undertaking GIS-based analyses. This is often troublesome when data aggregated within ZIP code areas are used, as they are frequently inconsistent across the agencies responsible for the geographic data (Krieger at al. 2002; Grubesic, 2008). Similarly, making

comparisons of Census variables over time can be problematic as both the variables and geographies can be inconsistent (Logan et al., 2014). Incorrect and inaccurate conclusions are a risk, though possible steps exist to mitigate the challenges (Gregory, 2002; Martin et al. 2002; Mennis, 2003). Numerous other studies have been conducted that consider these matters within specific disciplinary applications and should be consulted within each domain.

References

Gregory, I. N. (2002). The accuracy of aerial interpolation techniques: standardizing 19th and 20th century census data to allow long-term comparisons. Computers, Environment and Urban Systems, 26(4): 293-314.

Grubesic, T. H. (2008). Zip codes and spatial analysis: Problems and prospects. Socio-Economic Planning Sciences, 42(2): 129-149.

Krieger, N., Waterman, P., Chen, J. T., Soobader, M-J., Subramanian, S. V., and Carson, R. (2002). Zip Code Caveat: Bias Due to Spatiotemporal Mismatches Between Zip Codes and US Census–Defined Geographic Areas—The Public Health Disparities Geocoding Project. Journal of the American Public Health Association.

Logan, J. R., Xu, Z., and Stults, B. J. (2014). Interpolating U.S. Decennial Census Tract from as Early as 1970 to 2010: A Longitudinal Tract Database. The Professional Geographer, 66(3): 412-420.

Martin, D., Dorling, D., and Mitchell, R. (2002). Linking censuses through time: problems and solutions. AREA 34(1): 82-91.

Mennis, J. (2003). Generating Surface Models of Population Using Dasymetric Mapping. The Professional Geographer, 55(1): 31-42.