A Guide to Accessing Census Data for Municipalities

2023

Katelynn Conedera
University of Michigan
Taubman College of Architecture and Urban Planning

The US Census Bureau is the definitive source for geographically defined demographic data. However, among the many geographic categories available (from the national level all the way down to the census block), neither the Decennial Census nor the American Community Survey has a singular geographic category that contains every municipality in the United States. This presents a challenge for those researchers, practitioners, and other data users interested in the demographic characteristics of US municipalities.

Due to the complexity and variation in municipal government structures within and between states, the Census Bureau has developed two separate geographic categories for US localities (e.g., municipalities, unincorporated communities, etc.): 1) “Places”, which are population clusters that may or may not be consistent with a jurisdictional boundary (e.g., of a city) and are only provided for a subset of states, and 2) “County Subdivisions” (CSDs), which also may or may not be consistent with a jurisdictional boundary but are provided for the entirety of the US. Within the Places geographic category, municipalities are referred to as “Incorporated Places” and within the County Subdivisions category, municipalities are referred to as “Minor Civil Divisions” (MCDs). Together, these geographic categories are defined by the Census Bureau as municipalities[1].

The task for researchers interested in gathering information from the Census about a national set of municipalities, therefore, is to 1) collect information about both Places and County Subdivisions, and 2) ensure only those Places and County Subdivisions that represent a municipal jurisdiction are included, and 3) take care to address duplicates that are in both Place and County Subdivision data. Care must be taken to differentiate municipalities from other entities, as some non-governmental Places and County Subdivisions are named after nearby municipalities but may not have the same geographic boundaries.

Even within the Incorporated Places and Minor Civil Divisions geographic categories, separating out the municipalities can be difficult as each state has a different set of local government types with a wide range of functions. Because this can be such an arduous process, many researchers and practitioners just choose one of the two geographies and account for the limitations elsewhere. However, this can sometimes obscure true municipal boundaries and

[1] Code of Federal Regulations Title 15 section 70.2 (15 CFR § 70.2)

ascribe demographic data to the wrong populations. The purpose of this guide is to provide simple, easy to use information to help researchers and other Census data users be more precise and thorough when deciding which geographic entities best suit their purposes.

For those who need information on a specific set of municipalities but are unsure of where to start, this guide provides a summary-level list of the municipality types in each state and indicates their Census geographic category. Additionally, for those interested in collecting data across the United States, this guide instructs users on how to efficiently download the Place geography and supplement it with select data from the County Subdivision geography. This guidance was developed using the *2010 Guide to State and Local Census Geography*[2] and *2017 Census of Governments Individual State Descriptions*[3] released by the U.S. Census Bureau.

The accompanying spreadsheet contains a list of each municipal government type in each state, with one row per government. The spreadsheet includes two columns that indicate if that entity is present within either the Place or County Subdivision geographic categories, respectively. Finally, the last column is a dummy variable indicating which entities should be selected from each geographic category to achieve a full list of all municipal governments. The following readme guide instructs users on how best to use the information within the spreadsheet. For a real-world example of how this guide was used to conduct research using U.S. Census data, see *The Municipal Drinking Water Database*[4] linked below.

Making an Informed Choice

Census data users should be familiar with the uses and limitations of the various data products released by the US Census Bureau to ensure that the data is both selected and used appropriately. Because these data products support a wide variety of use cases, they often cannot be combined without modification. Additionally, though many Census data products appear to serve the same purpose, such as Decennial Census data modified to represent different geographic scales, each product has its own limitations and considerations that data users must be aware of to avoid issues such as overlapping geographic boundaries, wide margins of error, and mismatched time-series data.

Delineating between "Place" and "CSD" for geographic coverage

The Place and CSD geographies contain a combination of both municipalities (i.e., areas that coincide with a local government's jurisdiction) and non-municipal statistical entities (i.e., areas that share something in common besides falling under a local government's jurisdiction). Functionally, this means that neither the Place nor the CSD geography contains an exclusive list

[2] See https://www.census.gov/geographies/reference-files/2010/geo/state-local-geo-guides-2010.html

[3] See https://www.census.gov/library/publications/2019/econ/2017isd.html

[4] Hughes S, Kirchhoff CJ, Conedera K, Friedman M (2023) The Municipal Drinking Water Database. PLOS Water 2(4): e0000081. https://doi.org/10.1371/journal.pwat.0000081

of all municipalities in the United States. Due to the difference in criteria the Census uses to determine which entities appear in each geographic category, users must selectively subset both the Place and CSD geographic categories to capture every municipality. In short, the process of achieving a complete list of US municipalities is a two-step process of 1) collect information about both Places and County Subdivisions, and 2) ensure only those Places and County Subdivisions that represent a municipal jurisdiction are included, and 3) take care to address duplicates that are in both Place and County Subdivision data.

Places

In the **Places** geographic category, the statistical entities are called Census Designated Places (CDPs) and the municipalities are called Incorporated Places (IPs). CDPs are delineated by the Census Bureau for statistical purposes and have no relation to any existing municipal government boundaries. Incorporated Places correspond with municipal boundaries that meet incorporation requirements for each state, primarily being cities, towns, villages, and boroughs. The Places geography includes a large portion of municipalities in the United States, but the criteria for being an Incorporated Place necessarily excludes other types of active local governments such as townships and towns in certain states. Although townships are often active local governments with some or all of the same functions as IPs, the Census Bureau differentiates them from IPs due to the wide variation of governmental functions and structures both within and across states[5]. In states where towns have a more rural population distribution pattern similar to many townships,[6] towns are additionally not classified as IPs.

CSDs

Although IPs can be adequate for some use cases,[7] states that have many townships and rural towns have very limited coverage in the Places geographic category. Those towns and townships missing from the Places geography can instead be found in the CSD geography, though there are similar limitations as to which types of municipalities are included. The **CSD** geography was created to cover the entirety of the United States and includes municipal entities called Minor Civil Divisions (MCDs) and statistical entities called Census County Divisions (CCDs). Unlike the Places geography, the CSD geography does not allow for overlapping or semi-independent municipalities. For example, villages in Michigan are considered IPs, though they fall within and are only semi-independent from an encompassing township. Because these villages are governed by both the village government and the township government, the CSD geography only includes the broader township. Additionally, because the CSD geography by definition defines local governments in the context of how they subdivide counties, any municipality that falls across a county line is split into two entities.

---

[5] U.S. Census Bureau. *Geographic Areas Reference Manual Chapter 8: County Subdivisions* (1994). https://www2.census.gov/geo/pdfs/reference/GARM/Ch8GARM.pdf

[6] Wisconsin, New York, and all of New England

[7] The Places geography definition of CDPs by the Census Bureau can be helpful if you are interested in obtaining data on communities and population centers that may or may not have a legal boundary.

The CSD geography primarily contains all townships and select towns, in addition to any IPs that are not partially or completely governed by any other local government. It therefore contains many of the municipalities that are considered active local governments but do not meet incorporation criteria. Although the CSD geography is in many ways better suited for studies on municipal boundaries regardless of these limitations, the variation in local government structures has made it difficult to achieve full geographic coverage using MCDs alone. Because of this, only 29 states have MCDs available for data users, and only 12 of those are considered by the Census to be "strong-MCD" states[8]. Though all 29 states have usable MCD entities, only those 12 strong-MCD states have MCDs whose functions are directly comparable to Incorporated Places.

Ultimately a comprehensive "list" of municipalities that draws on Census data requires combining Incorporated Places and MCDs - with some caveats or tweaks that are described below.

Guidance for combining Place and CSD data

Large-scale Census data collection across multiple states will require a combination of Incorporated Places and MCDs to achieve full geographic coverage. For the 21 states that do not have MCDs, Incorporated Places are the best option for identifying all municipalities. This is because there is significant overlap across Place and CSD geographies where all cities, towns, and other municipal governments appear in both, but the CSD geography may contain non-incorporated entities.

The simplest route to achieving a list of all US municipalities is to download all Incorporated Places and supplement that list with non-Incorporated Places from the MCD geography. Not only does this provide a more streamlined and organized workflow to decrease the opportunity for user error, but this approach also ensures that municipalities that span across county lines do not appear as two distinct entities in the final list. The drawback of this approach, however, is that it introduces complexity and requires a more intentional approach to identifying which types of local government to include. There can be some compatibility issues when trying to merge a list of municipalities containing both Places and CSDs with data designed using only one of the two geographies. Many studies choose to accommodate for any limitations and include only Incorporated Places or MCDs because it is simpler. This guide aims to help streamline the process and encourage more precise and thorough examinations of US municipalities.

Avoiding Overlap when Combining Place and CSD Geographies

Although these entities may be identical between MCD and Incorporated Places datasets, the difference lies in the structure of the identifying FIPS codes. Even if there is 100% overlap between a Place and an MCD, the two entities will have different FIPS codes, as MCDs have an

[8] Strong-MCD states are those 12 states that have functioning governments at the sub-county level with full coverage across all counties.

For more information, see E. C. Castro Jr. & S. P. Hefter. U.S. Census Bureau Memorandum #ACS07-R-10 (11 June 2008). https://www.census.gov/content/dam/Census/library/working-papers/2008/acs/2008_Castro_01.pdf

additional 3 digits indicating the county they fall within. When joining data from a number of sources, one may run into both 7-digit and 10-digit FIPS codes[9]. If this is a concern, the TIGER/Line shapefiles for Places and MCDs can be overlapped in a geoprocessing software to identify those entities that have 100% boundary overlap. This will result in a key for any entity that has both a 7-digit and 10-digit FIPS code.

Because cities and towns for many states exist in both the MCD and Place geographies, care should be taken when combining the datasets. If it is *not* necessary to delineate between county representation, all Incorporated Places can be downloaded and supplemented with additional MCDs where appropriate. If, however, connection to the county government *is* important (or if the data must be compatible with a dataset that uses 10-digit FIPS codes), then the MCDs can be substituted for Places where appropriate.

Alternatives to this approach

Though there are other data products available that may better aid researchers in capturing demographic characteristics of US municipalities, the limitations of those approaches range from the lack of cross-compatibility with Census data to a need for more specialized knowledge of municipal functions. Ultimately, one must decide if the extra effort it takes to develop a thorough and customized list of municipalities best suits their needs given the other products provided by the Census Bureau.

For example, the Census of Governments would seem to be the best way to identify municipal governments, and it does provide a comprehensive listing of all local governments. However, the Census of Governments (CoG) is not cross-compatible with demographic data from the Census, as its ID structure is completely independent of any FIPS codes. A crosswalk file is necessary to bridge any CoG data with any Census data. Even with a crosswalk file, extra care is necessary when combining CoG data with another Census product, as it uses different methodologies and is collected only for years ending in 2 and 7.

One could also obtain the functional status codes (FUNCSTAT) from the TIGER/Line shapefiles and filter geographic entities by their governmental function rather than government type. However, FUNCSTAT codes provide a level of granularity beyond municipal type classifications (i.e., city, town) that is often unnecessary. Further, FUNCSTAT codes are based off of municipal type classifications and are simply a different way of displaying the same information. Though one could achieve the same list using either FUNCSTAT codes or municipal type classifications, this guide does not use FUNCSTAT codes because they require more knowledge of municipal government function than is necessary.

[9] Note that depending on the software used to view the data, some FIPS codes will be missing a leading 0, such as 123456789 instead of 0123456789. This can cause data compatibility issues if not addressed.

Accessing Census Data

There are currently two primary methods to downloading Census data directly from the Census Bureau: the interactive data.census.gov interface, and the more complicated FTP (file transfer protocol) site for advanced users familiar with accessing large data downloads using an FTP client or a coding language like R or Python. Though the FTP site is better for bulk data downloads and provides a broader selection of data than data.census.gov, it requires previous knowledge of how to download data from an FTP site. The following section aims to provide a brief, cursory explanation of both methods but only provides detailed instructions for accessing data.census.gov, which is sufficient for most users. For those who have access to third-party data repositories such as Social Explorer, these instructions can still be used to develop a list of all municipality FIPS codes that can then be joined to data from the source of your choosing.

The following geographic summary levels, as defined by the U.S. Census Bureau, will be accessed in this tutorial:

- 060 - State -> County -> County Subdivision
- 160 - State -> Place

Trade-offs: Ease of Access versus Data Availability

Data.census.gov is a relatively new website, having replaced the previous census data dissemination tool, American Factfinder, in 2020[10]. Though the user interface is more straightforward and accessible than American Factfinder, the Census Bureau is still working to migrate all of the data previously available into its new site. At this time, users can only access data starting in the year 2000 through this interface. Those who need data from previous years, including both ACS and Decennial Census data products, must access the FTP site[11]. For a more detailed chart showing what data products are available through data.census.gov, see "What data are available in data.census.gov" in the data.census.gov Frequently Asked Questions[12].

The Census Bureau's FTP site offers more data products and larger datasets at the expense of ease-of-use and download customizability. FTP users can, for example, access an entire Decennial Census summary file containing all available variables, whereas data.census.gov users must select from the provided tables or topics. However, those interested in downloading Place- and CSD-level data from the FTP site should consider using an FTP client or customized R/Python script, as data files are split by state and must be downloaded one-by-one. Another benefit of the FTP site is that it contains all of the TIGER/Line geographic shapefiles for all Census-defined geographies.

---

[10] U.S. Census Bureau Press Release. *American FactFinder Officially Retired: Data.census.gov Now Primary Data Tool* (31 Mar 2020). https://www.census.gov/newsroom/press-releases/2020/aff-data-census-gov.html

[11] Additional tables may be available at https://www.census.gov/data/tables.html

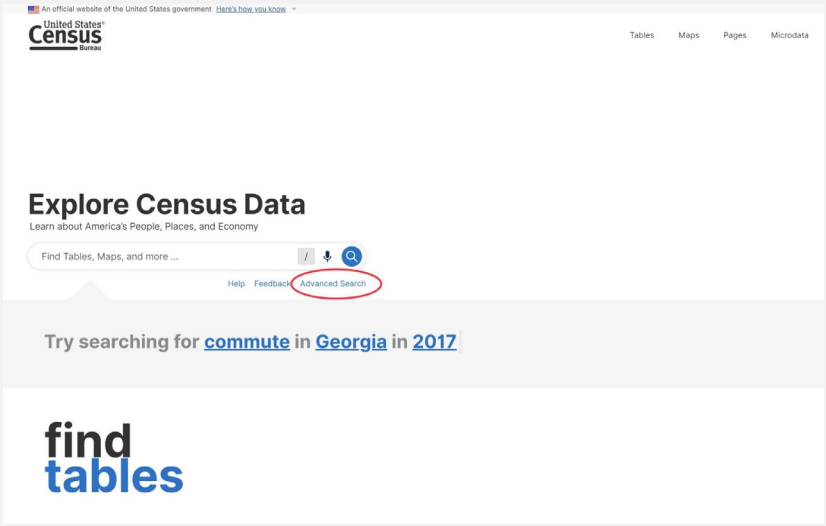
[12] https://www.census.gov/data/what-is-data-census-gov/guidance-for-data-users/frequently-asked-questions.html#data

Ultimately, it is up to the data user to determine if accessing the FTP site is necessary to access all of the data needed. Data.census.gov is easy to navigate and contains all of the data that the average user may need. The FTP site is best for experienced users and those who need large amounts of data or who need data not accessible through data.census.gov. It can be accessed at https://www2.census.gov/.

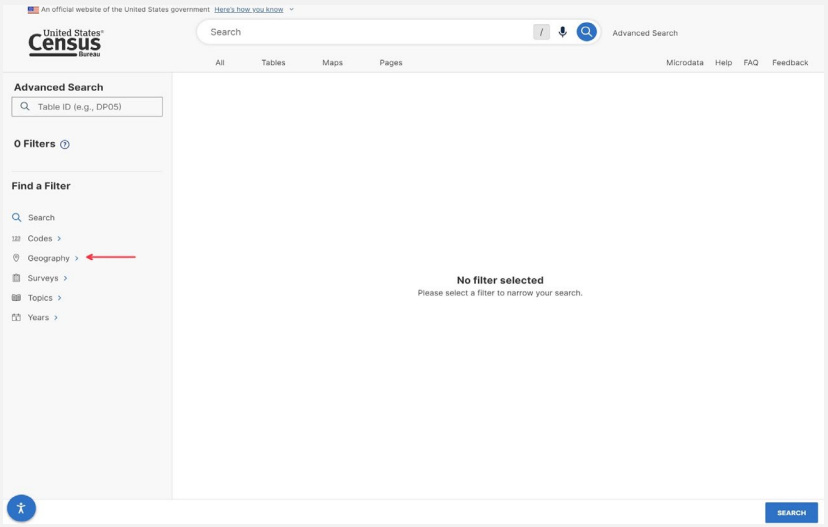
Using data.census.gov

Upon navigating to data.census.gov, users are instructed to enter a simple search term to begin their data exploration. The advanced search option is available for users with more specific needs and those already familiar with what types of data can be accessed using this tool. If you are accessing data.census.gov for the first time, or if you are not sure which table or data topic you are interested in accessing, take some time to explore the tool using the suggested search terms or browsing https://data.census.gov/table. Keep in mind that some tables are not available for all geographies. Once you have determined what data you wish to download, the following instructions will guide you in downloading that data for the Places and County Subdivision geographies.

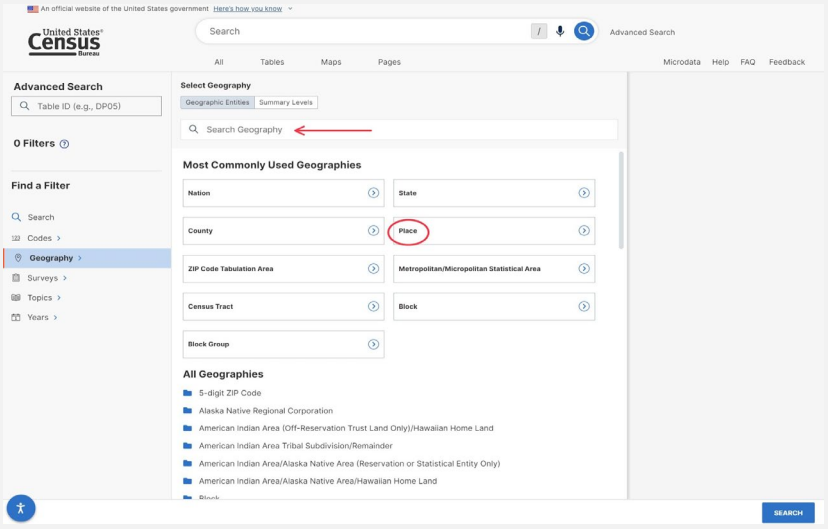
Step 1. Navigating back to https://data.census.gov, locate and click on the "Advanced Search" link under the search bar.



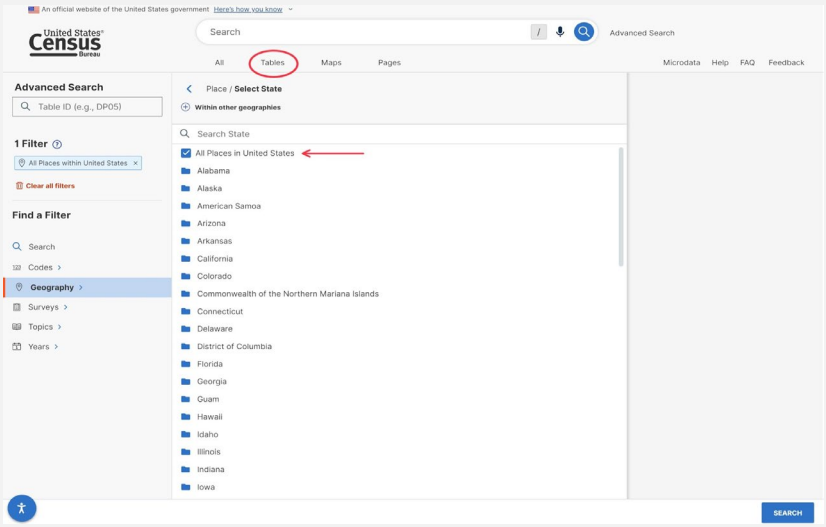
Step 2. On the advanced search page, there are a number of ways to filter data tables. Since some data is not available for Places or CSDs, we first want to filter by geography to exclude unavailable data tables. Select “Geography” in the “Find a Filter” section on the left to navigate to the geography selection page.



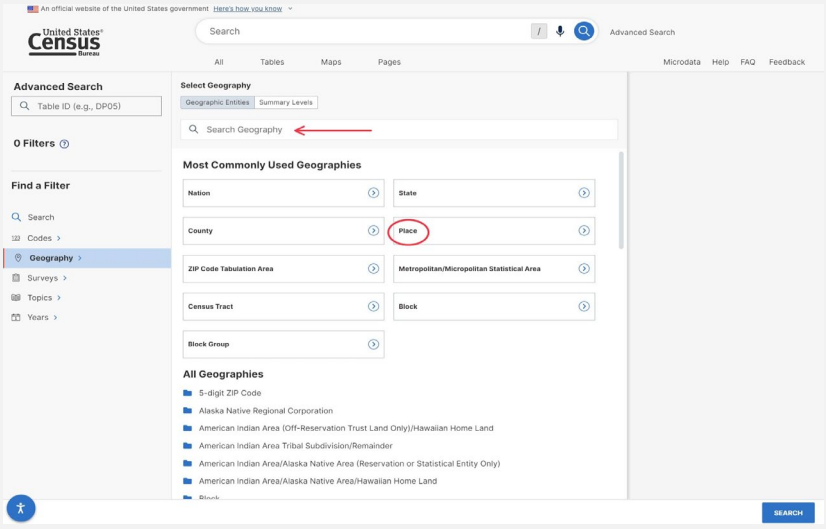
Step 3. Place is one of the most commonly used geographies, so it appears at the top of the page. Select “Place” under “Most Commonly Used Geographies,” or locate the search bar that says “Search Geography” and enter “Place.”



Step 4. Upon selecting “Place,” users will be prompted to select the state they want to limit the Place geography to. Unless you are only interested in a subset of states, select “All Places in United States” at the top of the list.

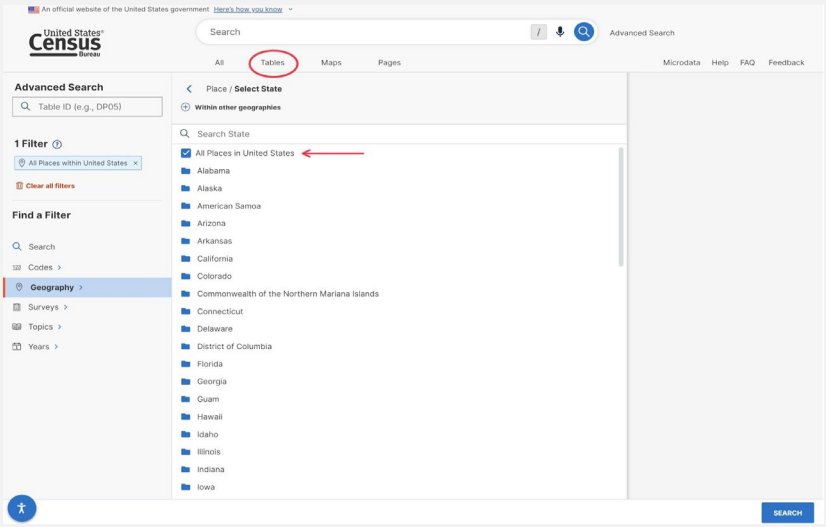


Step 5. Since County Subdivisions are not listed as a commonly used geography, we must locate them using the “Search Geography” search bar mentioned previously. To do this, navigate back to the “Geography” filter page and enter “County Subdivision” into the “Search Geography” search bar.

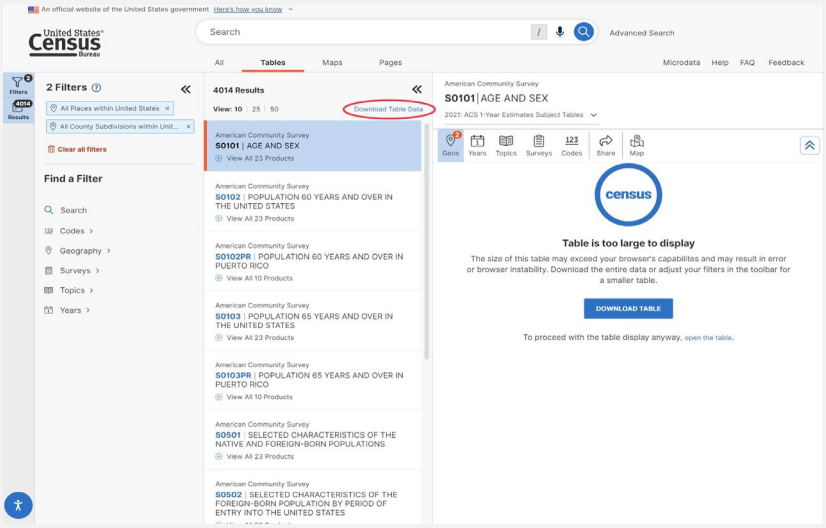


Step 6. Repeat the previous step by selecting “County Subdivision” in the search results and choosing “All County Subdivisions in United States.”

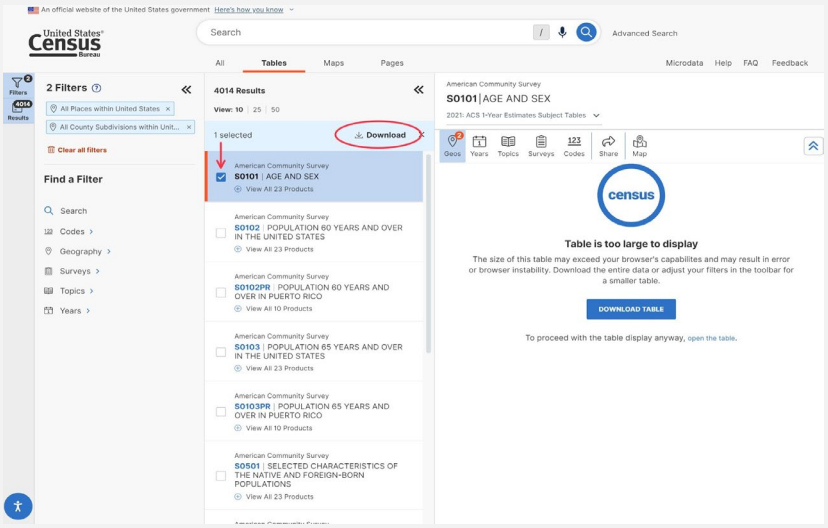
Once you have selected both all Places and all CSDs, you are ready to select your table. Enter any other search filters you require by using the “Find a Filter” options on the left, then navigate to the “Table” tab above to view all tables available based on your filters.



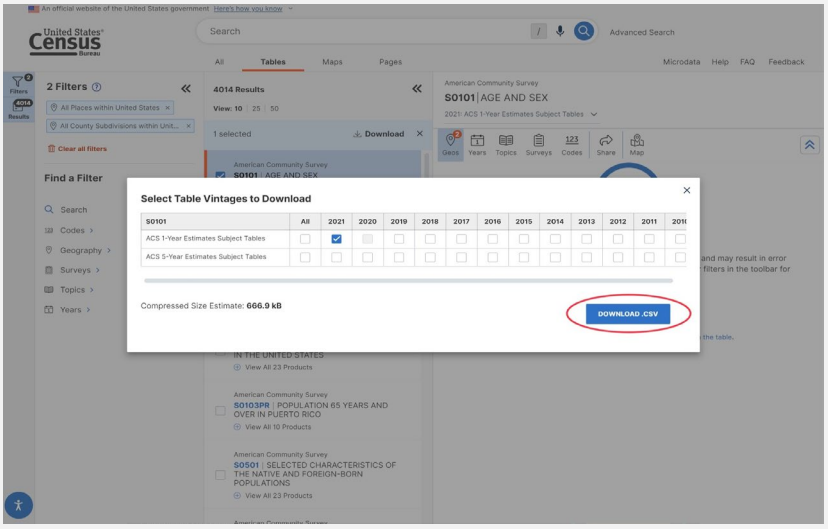
Step 7. The “Tables” tab lists all tables that fit your search terms and provides a preview of the selected table. Tables containing all Places and CSDs in the United States will likely be too large to preview. Once you’ve identified the table(s) you wish to download, click on “Download Table Data” at the top of the results list. This will allow you to select which tables you wish to download.



Step 8. Using the checkboxes next to each search result, select the table or tables that you wish to download. Then, click “Download” at the top of the results list. This will display a prompt allowing you to select which products/years you wish to download the tables for.



Step 9. This prompt will list all surveys and years that are available for download for your selected table. Select which tables you wish to download, and choose “Download .csv” to download your table(s). Your data download will consist of one zipped folder that, once unzipped, contains one .csv file each for each table you downloaded, in addition to a .csv file for column metadata and a text file with table notes.



FTP Access

More experienced users can use the Census FTP site for larger downloads. For those interested, the URL to the FTP directory is https://www2.census.gov/. To access the desired directory to download, replace “https://www2” with “ftp://ftp2” in the appropriate URL. Finally, use the user/password combination “anonymous:anonymous” to access the site.

A note on year and time-series analysis

If you plan on conducting a time series analysis with Census data, please consult Census documentation for comparing ACS data across time[13]. This guide is designed to be used with only one year of data at a time. However, users can create their own crosswalk files using Census relationship and comparability files[14] that show changes in geography from year to year. However, this should only be done with an understanding of the limitations of time-series analysis on ACS data[15].

Filtering for Municipal Governments

After downloading data containing all Places and County Subdivisions in your geographic scope, the next step is to filter all geographic entities for only Incorporated Places and Minor Civil Divisions, while accounting for duplicates that appear as both. This guide is meant to be software agnostic, meaning that you can use the statistical processing software you are most comfortable with. These are the basic steps to follow in order to use this Census guide correctly with any software. Since we used RStudio and select packages to process our own list of municipalities, we also provide some useful resources for RStudio users in the Appendix.

The following steps provide an overview of the final filtering and selection process. This guide is accompanied by an Excel spreadsheet ("Census Guide Final.xlsx") that contains several lists of local governments and exceptions that can be used to select all municipalities from the full list downloaded in the previous steps.

1. Load the Place and CSD geography datasets into your software
2. Conduct basic data cleaning tasks to prepare the separate datasets for merging, e.g.,
    a. Standardizing column names
    b. Removing undesired columns
3. Remove all Census-defined entities listed in sheet A. Census-Defined Entities
    a. These are all non-governmental entities such as CCDs and CDPs
    b. Note that at this point, all Places left will be Incorporated Places, and the following steps outline which CSDs to remove to only keep MCDs
4. Remove all non-municipal CSDs identified in sheet B. Nongov CSDs
    a. The actual entities you choose to remove may vary based on the definition of municipality or local government for different purposes
    b. Please see the "Notes" sheet for more detailed information about the governmental status of entities in each state

[13] U.S. Census Bureau. *Comparing ACS Data*. https://www.census.gov/programs-surveys/acs/guidance/comparing-acs-data.html

[14] For more information and to access relationship files, see the U.S. Census Bureau *Relationship Files* page at https://www.census.gov/geographies/reference-files/time-series/geo/relationship-files.html

[15] In the case of comparability issues or when using a crosswalk file, one should generate a TIGER/Line shapefile with all Places in the dataset and use it to select all entities in the "US MCDs TIGER/Line" shapefiles that have 100% overlapping boundaries. This can be used to generate a crosswalk file of alternative FIPS codes for a dataset.

5. Remove all other non-municipal CSD entities by state and county identified in sheet C. Nongov County CSDs
    a. This includes certain counties in Kentucky and Missouri whose County Subdivisions are nongovernmental
    b. Now all remaining entities are municipal; the dataset now contains Incorporated Places and Minor Civil Divisions, *with some duplicates*
6. Remove all CSD duplicates of Incorporated Places identified in sheet D. Duplicate Entities
    a. This will be done by municipality type; for example, you will remove all MCD cities from the identified states as they are already represented by the Incorporated Place cities
    b. Unless there is a reason to preserve MCD entities rather than Incorporated Place entities (for example, if the surrounding county is important and it is not an issue to split entities across county lines), we suggest removing all duplicates from the MCD dataset for simplicity

Note that consolidated cities and balances are not removed in this guide. This is a separate decision to be evaluated based on your own research needs and definition of municipal governments.

Appendix

This guide can be used with the statistical processing software of your choice, and in our own application we used RStudio to generate our list of municipalities. For your convenience, we have included some basic information about our process using RStudio in this appendix.

Basic instructional advice

For those unfamiliar with RStudio, or with the particular packages we reference, we have collected some resources provided by RStudio that may be particularly helpful.

- RStudio IDE
- Base R
- String Manipulation with stringr
- Data transformation with dplyr
- Working with Regular Expressions

Packages we used

We used the following packages to conduct our data cleaning process. These are not necessarily required to conduct data cleaning, but they helped streamline the process.

- stringr
- dplyr
- ngram
- sf