ALA American Library Association

US CENSUS DATA

CONCEPTS AND APPLICATIONS FOR SUPPORTING RESEARCH

Frank Donnelly

Library Technology Reports
Expert Guides to Library Systems and Services

MAY/JUNE 2022
Vol. 58 / No. 4
ISSN 0024-2586

Library Technology
REPORTS
Expert Guides to Library Systems and Services

US Census Data: Concepts and Applications for Supporting Research

Frank Donnelly

ALA TechSource
American Library Association

Library Technology
REPORTS

ALA TechSource purchases fund advocacy, awareness, and accreditation programs for library professionals worldwide.

Volume 58, Number 4
US Census Data: Concepts and Applications for Supporting Research
ISBN: 978-0-8389-3817-1
DOI: https://doi.org/10.5860/ltr.58n4

American Library Association
225 N. Michigan Ave., Suite 1300
Chicago, IL 60601-7616 USA
alatechsource.org
800-545-2433, ext. 4299
312-944-6780
312-280-5275 (fax)

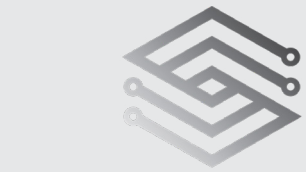
Editor
Samantha Kundert

Copy Editor
Judith Lauber

Production
ALA Production Services

Cover Design
Alejandra Diaz and ALA Production Services

Library Technology Reports (ISSN 0024-2586) is published eight times a year (January, March, April, June, July, September, October, and December) by American Library Association, 225 N. Michigan Ave., Suite 1300, Chicago, IL 60601-7616. It is managed by ALA TechSource, a unit of the publishing department of ALA. Periodical postage paid at Chicago, Illinois, and at additional mailing offices. POSTMASTER: Send address changes to Library Technology Reports, 225 N. Michigan Ave., Suite 1300, Chicago, IL 60601-7616.



About the Author

Frank Donnelly is the GIS and Data Librarian at Brown University Library in Providence, Rhode Island, where he supports geospatial and demographic research on campus and coordinates the library's data services program. He is an affiliated staff member of the Population Studies Training Center (PSTC) and the author of *Exploring the U.S. Census: Your Guide to America's Data* as well as several articles and reports on census data and the geography of US public libraries. Visit his website to learn more about his work: https://atcoordinates.info/.

Abstract

More than just a ten-year count, the US census is a collection of high-quality, geographically detailed, and free and open datasets that describe the demographic and socioeconomic characteristics of the nation on an ongoing basis. This issue of *Library Technology Reports* (vol. 58, no. 4), "US Census Data: Concepts and Applications for Supporting Research," provides readers with a crash course on the census: learn about the concepts on which the census is organized, the key datasets, accessing data online and through scripts via APIs, and considerations for using GIS, historical data, and microdata. Librarians will gain knowledge they can use for assisting members of their communities with census data and will see how the census can be used for library planning and research.

Contents

Chapter 1

Introduction

When most Americans think of the census, they think of the count of the population that happens every ten years. While the ten-year census is the Census Bureau's foundational program, it is actually part of an ecosystem of statistical programs that produce demographic, socioeconomic, housing, and business data on an ongoing basis. Most of this data is free and publicly available and is used to inform policy and research about all facets of American life. Libraries play key roles in assisting the Census Bureau with conducting these programs and help users access, understand, and apply census data to their work. Census data can be challenging to work with given its scope and complexity, and thus a variety of resources and tools have been created to make the data accessible to users with different needs and varying levels of technical skill.

This report provides a crash course on the census written for librarians who need a basic understanding of the data in order to help others. It will also be broadly useful to any student, researcher, or policy maker who is interested in using and understanding census data so they can apply it to their work. We will begin with an overview of the census: its role in American society in general and librarianship in particular, with an emphasis on its value as a free and open dataset. Then we will discuss three concepts that apply to all the census datasets, which are important to understand for both accessing data and analyzing it correctly: census geography, subject categories, and data tables. Given the large number of datasets, we will focus on the most frequently used series for general study of the population: the decennial census (DEC), the American Community Survey (ACS), Population Estimates, and two business datasets (Business Patterns and the Economic Census). With a basic understanding of concepts and datasets in place, we will explore some portals and tools for accessing data for users with varying degrees of technical skill, with an emphasis on data.census.gov and the Census Bureau's application programming interfaces (APIs).

Beyond the basics, we'll summarize the advanced topics that many researchers are likely to delve into first: working with census data and boundary files in geographic information systems (GIS), considerations and sources for working with historical census data and making comparisons over time, and working with microdata, which is anonymized samples of individual responses to census questionnaires. In covering this last topic we'll introduce a final dataset, the Current Population Survey (CPS). We will conclude with a summary of the applications of census data to library and information science (LIS) from a practitioner and academic research perspective, for studying library user populations, siting new facilities, and measuring accessibility and equity of services.

In covering these topics, we will highlight a number of useful portals and tools for accessing and working with data, as well as resources for learning more about particular concepts or topics. Throughout this report, we will highlight some of the changes associated with the recent 2020 census. Readers who want either a fuller or briefer treatment of the census can consult my book, *Exploring the U.S. Census: Your Guide to America's Data* (Donnelly 2020a) or my data brief *Exploring U.S. Census Datasets: A Summary of Surveys and Sources* (Donnelly 2020b).

Before we begin, some notes about terminology, as the word *census* applies to many different things. In this report I will refer to the agency as the "Census Bureau" or simply the "Bureau" and to the ten-year census as the "ten-year census" or "decennial census," abbreviated as DEC, or to a specific iteration, such as the 2020 census. I will use the term *the census* to generally reference datasets and products produced by the Census Bureau.

References

Donnelly, Frank. 2020a. *Exploring the U.S. Census: Your Guide to America's Data*. Los Angeles: SAGE.

———. 2020b. *Exploring U.S. Census Datasets: A Summary of Surveys and Sources*. WCIB Occasional Paper Series, no. 20. New York: Weissman Center for International Business, Baruch College/CUNY, February. https://zicklin.baruch.cuny.edu/wp-content/uploads/sites/10/2020/02/OPS-20-Exploring-Census-Datasets.pdf.

Chapter 2

Roles of the Census

The US decennial census is one of the oldest continuously conducted population counts in the modern era (Sweden began a regular census a few decades prior to the US; Whitby 2020, 87–88, 91) and the first conducted for the purpose of apportioning seats in a democratically elected legislature (Emigh, Riley, and Ahmed 2016a, 145–72). Mandated by Article 1, Section 2, of the Constitution, its original and principal role is to reapportion seats in the US House of Representatives among the states every ten years based on differential population growth between the states. Since the number of representatives each state has is based on its population, the seats must be redistributed as the population changes. Since the first census was conducted as a basic head count in 1790, the scope and role of the census has constantly expanded. The ten-year census grew to incorporate more questions, and additional data programs were added to satisfy other constitutional requirements and meet the needs of a growing federal statistical system to assess the state of the nation's population and economy (Anderson 2015, 2010). Federal laws and acts of Congress require the collection of different statistics through the census to assess policies and enforce laws. Beyond reapportionment, census data is used for redistricting. Congressional districts must be redrawn every ten years in states that gain or lose seats via the reapportionment process, and they must meet certain requirements in terms of population size and geographic area and shape.

Over the course of the twentieth century and into the twenty-first, census data was increasingly used by policy makers in models that are used to distribute federal funds to state and local governments to fund schools and roads, fight poverty and hunger, and supplement health care and unemployment programs (Anderson 2015, 156–223). In fiscal year 2017, $1.504 trillion were distributed to state and local governments, nonprofits, businesses, and households using census data (Reamer 2019). The role the census has in redistributing both political power and financial resources has established it as a vibrant institution in American society, one that is sharply debated and hotly contested. Some researchers have suggested that this makes the census more relevant to civic life in the United States relative to other countries, where census taking is viewed more as a distant, bureaucratic exercise with limited impact on people's lives (Emigh, Riley, and Ahmed 2016b, 147–76).

Ultimately, census data is collected to satisfy requirements of the Constitution, acts of Congress, and a large body of federal law. As a result, every question that's asked as part of a census survey has some basis in federal law (US Census Bureau 2017). For example, data on race is collected to measure and ensure that civil rights and voting rights are being upheld and is required by dozens of different federal statutes. The questions that are asked as part of each ten-year census and the ACS are reviewed and submitted to Congress for approval prior to each count (US Census Bureau 2018). The Census Bureau cannot ask questions simply because the results may be novel or interesting or because a particular stakeholder group wants the data to be collected.

The fact that census data is high-quality, transparent, free, and in the public domain ensures that it is used widely throughout American society. Anyone can access and use census data for any purpose, and as a result the census serves as a key piece of data infrastructure on which many products, policies, and studies are built. State and local governments use census data for everything from planning and policy work to emergency management and public health. In public health, census data serves as the denominator for measuring prevalence and risk of diseases, and its variables often serve as inputs in quantitative models (Wilson et al. 2017). Businesses use census data to understand and target markets and site new locations, while journalists use it to provide context for stories that describe communities or social issues. In

academic research, social scientists use the census to understand broad social and economic trends. These stakeholders both support and contest changes to the census as they impact their specific needs.

As high-quality open datasets, census data remains relevant in the age of big data as it is designed to answer specific research and policy questions while transparently accounting for bias and error. This stands in sharp contrast to big data, which is often not designed for the purpose it is applied to and is often proprietary and opaque (Kitchin 2014, 27–46). Census data can be used to ground-truth or test big data and derived datasets to gauge their accuracy or completeness (Donnelly 2020, 17–21). Given privacy and confidentiality regulations, there is some compromise in the level of detail and accuracy that the Census Bureau can provide. Individual responses to decennial census and ACS questionnaires are confidential and not published or distributed for seventy-two years from the date of release. All census data that's released is published either as summary data (summarized by geography or subject categories) or as anonymized microdata samples (samples of individuals' responses with personal information removed). Published data can be subjected to noise (artificially inflating or deflating values by a small amount) or nondisclosure of values below a certain limit to protect privacy.

Libraries play multiple roles when it comes to census data. On the data collection side, libraries help the Census Bureau with community engagement to promote census programs (the ten-year census in particular) and to encourage people to submit the questionnaires. Public libraries serve as venues for completing the census by providing forms and computers and hosting census volunteers who can answer questions (American Library Association 2020). On the data dissemination side, librarians help their patrons and communities navigate the census by helping users find and understand the data so they can use it for their research purposes. General reference librarians field questions about finding statistics (Bauder 2014, 1–15), while many academic libraries have data services or GIS librarians on staff to assist users with advanced research needs (Rice and Southall 2016, 1–14). Librarians also use census data as data consumers, as the census is a resource for understanding library user communities. By studying data on age, sex, race, income, language, educational levels, and computer access, libraries can make decisions about collections and services based on the characteristics of their communities.

References

American Library Association. 2020. *Libraries' Guide to the 2020 Census*, January 2020 Update. Chicago: American Library Association. https://www.ala.org/advocacy/govinfo/census.

Anderson, Margo J. 2010. "The Census and the Federal Statistical System: Historical Perspectives." *ANNALS of the American Academy of Political and Social Science* 631, no. 1 (September): 152–62, https://doi.org/10.1177/0002716210373721.

———. 2015. *The American Census: A Social History*. 2nd ed. New Haven, CT: Yale University Press.

Bauder, Julia. 2014. *The Reference Guide to Data Sources*. Chicago: ALA Editions.

Donnelly, Frank. 2020. *Exploring the U.S. Census: Your Guide to America's Data*. Los Angeles: SAGE.

Emigh, Rebecca Jean, Dylan Riley, and Patricia Ahmed. 2016a. *Antecedents of Census from Medieval to Nation States: How Societies and States Count*. New York: Palgrave Macmillan.

———. 2016b. *Changes in Censuses from Imperialist to Welfare States: How Societies and States Count*. New York: Palgrave Macmillan.

Kitchin, Rob. 2014. *The Data Revolution*. Los Angeles: SAGE.

Reamer, Andrew. 2019. "Brief 7: Comprehensive Accounting of Census-Guided Federal Spending (FY2017)." *Counting for Dollars 2020: The Role of the Decennial Census in the Geographic Distribution of Federal Funds*. Washington, DC: George Washington Institute of Public Policy, George Washington University. https://gwipp.gwu.edu/counting-dollars-2020-role-decennial-census-geographic-distribution-federal-funds.

Rice, Robin, and John Southall. 2016. *The Data Librarian's Handbook*. London: Facet.

US Census Bureau. 2017. *Subjects Planned for the 2020 Census and American Community Survey: Federal Legislative and Program Uses*. Washington, DC: US Department of Commerce, March 28. https://www.census.gov/library/publications/2017/dec/planned-subjects-2020-acs.html.

———. 2018. *Questions Planned for the 2020 Census and American Community Survey: Federal Legislative and Program Uses*. Washington, DC: US Department of Commerce, March 29. https://www.census.gov/library/publications/2018/dec/planned-questions-2020-acs.html.

Whitby, Andrew. 2020. *The Sum of the People: How the Census Has Shaped Nations, from the Ancient World to the Modern Age*. New York: Basic Books.

Wilson, R. T., S. H. Hasanali, M. Sheikh, S. Cramer, G. Weinberg, A. Firth, S. H. Weiss, and C. L. Soskolne. 2017. "Challenges to the Census: International Trends and a Need to Consider Public Health Benefits." *Public Health* 151 (October): 87–97. https://doi.org/10.1016/j.puhe.2017.05.015.

Chapter 3

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 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

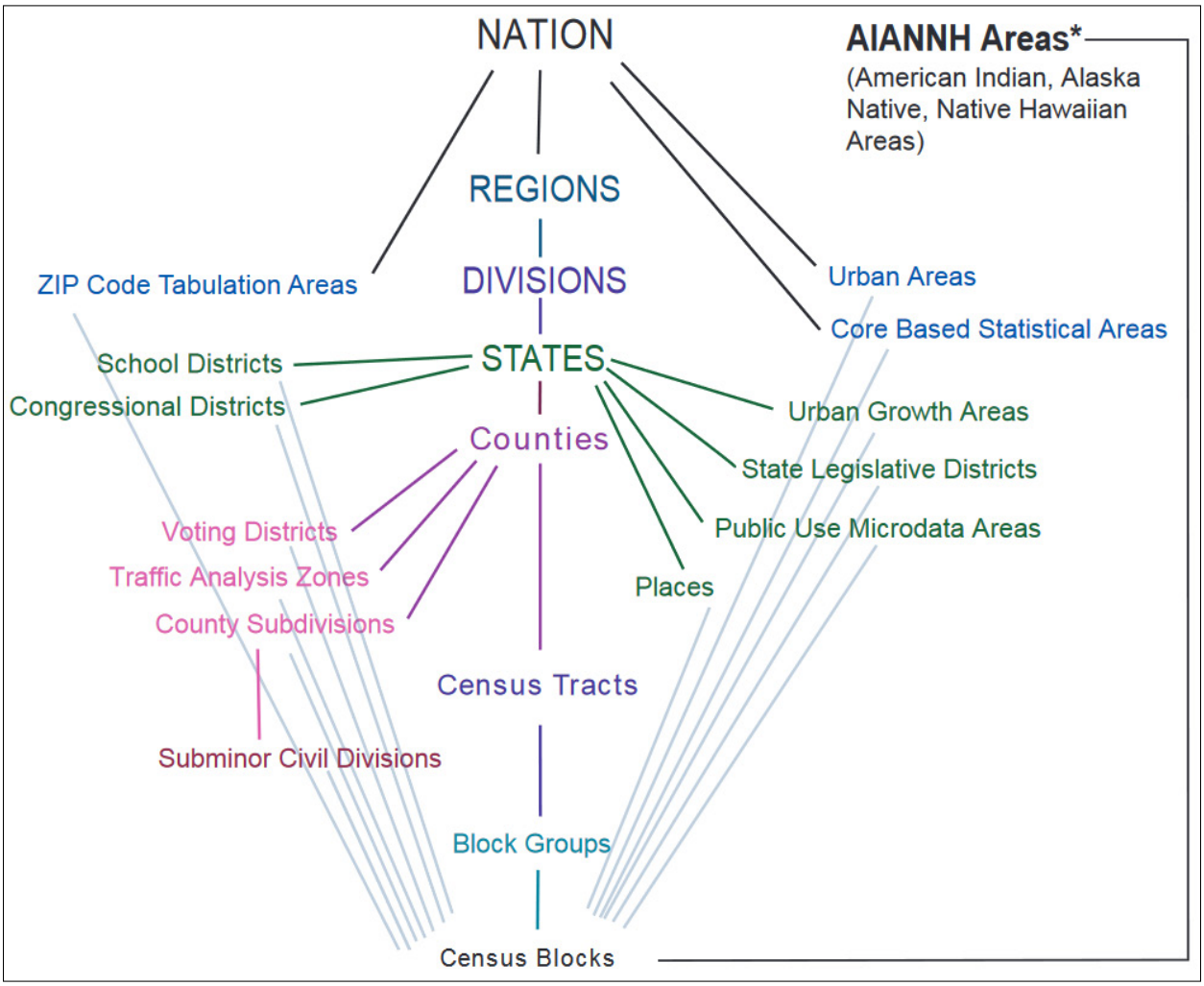


Figure 3.1
Hierarchy of census geographies (https://www.census.gov/programs-surveys/geography/guidance/hierarchy.html)

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/thetwo-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.

Datasets

Census Program Pages

The home pages for each census program. These contain news, release schedules, methodology, technical documentation, and tutorials. Direct downloads for smaller datasets are also available from these pages.

Decennial Census
https://www.census.gov/programs-surveys/decennial-census.html

American Community Survey
https://www.census.gov/programs-surveys/acs/

Population Estimates
https://www.census.gov/programs-surveys/popest.html

Business Patterns
https://www.census.gov/programs-surveys/cbp.html

Economic Census
https://www.census.gov/programs-surveys/economic-census.html

Decennial Census

The DEC is the cornerstone of the Census Bureau's programs. It is the original census dataset that has been collected since 1790, as stipulated by Article 1, Section 2 of the Constitution for the purpose of reapportioning seats in Congress between the states and redistricting Congressional districts within states. The DEC also serves as an input and a control for many other census data programs, such as the ACS and Population Estimates, as well as for datasets generated by other federal agencies.

The DEC is a 100 percent count of the entire population, conducted every decade in years ending in zero. It is conducted over a period of several months but uses a reference date of April 1 for determining residency. The concept of usual residence is employed, where each household is asked whether their current address is the place they usually live on April 1 (note that this is quite different from asking what a person's permanent or legal residence is). With the exception of foreign tourists and visitors on short-term visas, every person who resides in the United States is counted regardless of their status or living circumstances. The Census Bureau maintains a master address file for the entire nation, and all households are sent a census form that they are required by law to submit. For the 2020 census, people were able to submit their form online for the first time. The Census Bureau works with administrators of group-quarters facilities to count populations in these facilities and has a dedicated program for counting the homeless population.

The Bureau sends enumerators to certain hard-to-count areas to collect data in person and to households or quarters that don't respond to the initial form or follow-ups. The counting process for 2020 was noteworthy as it coincided with the COVID-19 pandemic, social unrest, and several severe weather events. The self-response rate for the 2020 census was 67 percent, which was comparable to the 2010 census, and after nonresponse follow-up, the Bureau estimates that over 99 percent of all addresses were reached (US Census Bureau 2020a). The count was noteworthy in that approximately 80 percent of all self-responses were submitted online and because item nonresponse (where a household left certain questions blank) was higher than in the past (US Census Bureau 2021d). While data for the vast majority of Americans is captured by self-response and nonresponse follow-up, the Bureau applies a number of imputation techniques to estimate missing data for those who are not directly reached.

Also new for 2020 is the adoption of a differential privacy mechanism, a mathematical approach with roots in cryptography that adds slight alterations or "noise" to published summary data and microdata to prevent reverse engineering of the census (in concert with third-party datasets) to reveal information about individuals. While the Census Bureau emphasizes that these new techniques are necessary given the current data privacy landscape and that the impact on most analyses is quantifiably minimal (US Census Bureau 2021b, 2021c), some demographic experts and policy makers have been critical of the changes (Wang 2021; IPUMS 2021). Another aspect of the new mechanism is a reduction in the number of tables that will be published and a decrease in their level of geographic detail, with fewer tables available at the block, block group, and tract levels.

Once the count is finished and the data is published, the Bureau uses two different techniques for estimating a possible undercount or overcount: a post-enumeration survey to a sample of households and demographic analysis using birth, death, and migration data from the Population Estimates program. Relatively speaking, the DEC is highly accurate; both the 2000 and 2010 census slightly overcounted the total population, while the 2020 census neither undercounted nor overcounted the total. However, in each case there was a differential undercount, where certain groups were missed while others were double-counted; young children, minorities, and people of low income tend to be undercounted.

The data is released in stages. By law the official population count of the US and each state must be submitted to the president by December 31 of the census year to launch the reapportionment process. The following March, the Public Redistricting files that contain basic population and housing counts down to the block level are released so that the redistricting process can begin. The following summer and fall, the most detailed census data is released on a rolling basis as part of the Summary File 1 series, with additional cross-tabulations by race and ethnicity published in Summary File 2 the following year (for the 2020 census, the term *Summary File* is slated to be replaced by the term *Demographic and Housing Characteristics File*). The timetable for the release of all the 2020 census products was delayed due to the COVID-19 pandemic and political upheaval. Apportionment data wasn't released until March 2021, and redistricting data was finally published in August 2021. The data shows that the population grew from 309 to 331 million between 2010 and 2020. This increase of 7.4 percent was the lowest rate of growth since the 1930s.

The contemporary DEC dataset consists of just basic demographic characteristics collected through the 100 percent count. These variables include age, sex, race, ethnicity, household and family relationships, group-quarters population, and housing unit occupancy and tenure. This focus on the basics was introduced in the 2010 census and continues for 2020. In contrast, from the 2000 census backward there were two questionnaires: a short form that collected the basics from the entire population and a long form that was sent to a large sample of the population (one in six households) that collected detailed socioeconomic and housing characteristics such as educational attainment, employment status, home value, and much more. The ACS was introduced in 2005 to capture all of these detailed characteristics on an annual basis so that in 2010 the decennial census could revert to being a simpler count.

Researchers should use the decennial census when they

- need precise counts of the population
- need block- and block-group-level data
- require just basic demographic variables
- do not require current data (particularly by decade's end)
- are studying small population groups
- are doing historical comparisons with prior decennial censuses

The DEC will not be your choice if you need detailed socioeconomic or housing characteristics, current data (particularly toward the end of a decade), or annual data.

American Community Survey

The ACS was launched in 2005 to provide detailed population and housing characteristics on an ongoing basis. It was designed to provide data in a way that was more timely and cost-effective than the DEC, while allowing the latter to focus more exclusively on counting the basic characteristics of the population. The ACS is a rolling sample survey of 292,000 addresses a month, for a total of 3.5 million addresses per year. The sample is stratified so that a certain number of addresses in every state are captured. The concept of current residence is used for determining residency, where a household is asked whether they are staying at the current address for at least two months. Sampled households are required by law to return the survey by mail or online, and the sample includes about 2.5 percent of all group-quarters addresses in the Bureau's master address file. Unlike the DEC, the Bureau follows up with only a sample of nonresponders and uses imputation techniques only for incomplete forms.

Each year, the sample data collected for the previous year is used to generate estimates for the total

Table 4.1. ACS Table Prefix Codes and Subjects

ID	Subject	ID	Subject
00	Unweighted Count (of the Sample)	15	Educational Attainment
01	Age and Sex	16	Language Spoken at Home
02	Race	17	Poverty Status
03	Hispanic or Latino Origin	18	Disability Status
04	Ancestry	19	Income
05	Citizenship Status, Year of Entry, Foreign Born	20	Earnings
06	Place of Birth	21	Veteran Status
07	Migration and Residence 1 Year Ago	22	Food Stamps/SNAP
08	Commuting and Place of Work	23	Employment and Work Status
09	Relationship to Householder	24	Industry, Occupation, Class of Worker
10	Grandparents and Grandchildren	25	Housing Characteristics
11	Household and Family Type	26	Group Quarters
12	Marital Status and History	27	Health Insurance Coverage
13	Fertility	28	Computer and Internet Use
14	School Enrollment	29	Citizen Voting-Age Population

population. A one-year period estimate is published for all geographic areas of the country that have at least 65,000 people. As the sample size is too small to publish reliable annual estimates for smaller areas, sample data for five years is aggregated to produce a five-year period estimate for all geographic areas down to the block group level. Each year, a new five-year period estimate is released by dropping out samples from the oldest year and adding samples from the latest year.

ACS statistics are published as estimates with margins of error at a 90 percent confidence level. For example, according to the 2019 ACS, the population of Providence, Rhode Island, was 179,875 +/− 40. This means that we are 90 percent confident that the population is somewhere between 179,835 and 179,915 people, and there is a 10 percent chance that the actual population falls outside this range. The smaller the population group or geographic area, the higher the margin of error will be. For instance, the population under age eighteen in Providence in 2019 was 36,020 +/− 3,401. In all published tables, ACS estimates are always provided in pairs with the estimate in one column and the margin of error in a second column. Estimates representing percent totals are also published with a margin of error.

Unlike the DEC statistics, ACS estimates are fuzzy intervals that can be used for generally characterizing an area and should never be interpreted as exact counts (Spielman, Folch, and Nagle 2014). This fuzziness is even greater for areas with less than 65,000 people, as we can characterize only a five-year time frame. For example, estimates for Newport, Rhode Island, are published for only a five-year period because the population of the city falls below the 65,000 threshold. We can say that the population of the city was approximately 24,663 +/− 34 between the years 2015 and 2019. If we wanted to make comparisons between Newport and Providence, we would need to use data from the five-year ACS for both places, since data is unavailable for Newport in the one-year series.

The great strengths of the ACS are that it provides annually updated data for large and small areas of the US and includes a broad range of detailed demographic and socioeconomic characteristics that are not captured in other census or government datasets. Table 4.1 summarizes the ACS subject categories and their prefix codes, which are used to identify and group the tables.

Its great weaknesses are its complexity, which makes it more challenging to work with, and reliability issues that are serious for small areas and population groups. It is not uncommon that ACS data is misinterpreted and used incorrectly. In particular, the margin of error is often ignored or misunderstood, and the data is inappropriately treated as a count (Jurjevich et al. 2018; Nesse and Rahe 2015). Ignoring the margin of error can lead to faulty conclusions regarding the significance of findings (Jung, Thill, and Issel 2019). Even though 3.5 million addresses are sampled, due to nonresponse and other issues, just over 2 million addresses are used to generate estimates in any given year. The level of uncertainty in the ACS has exceeded the Census Bureau's initial expectations, and some researchers have suggested that the block-group-level data is unusable for most applications, and even census-tract-level data does not provide reasonable certainty in estimating certain characteristics (Spielman, Folch, and Nagle 2014; Salvo et al. 2007). Due to a high nonresponse rate caused by disruptions from the COVID-19 pandemic, the Census Bureau is not publishing traditional one-year estimates for the 2020 ACS but is releasing a smaller number of experimental tables for large geographic areas instead (US Census Bureau 2021a).

Researchers can make compromises to counter the effect of low precision, essentially by using estimates generated from a larger sample. One approach is to use a five-year estimate even though a one-year estimate for a given place is available. For example,

the one-year 2019 estimate for the population under eighteen in Providence was 36,020 +/− 3,401, but if we used the five-year 2015–2019 estimate, the under-eighteen population was 40,156 +/− 1,006, a much smaller margin of error relative to the estimate. Another approach would be to use a smaller number of subcategories for a given variable, such as households by income classified into fewer income brackets. The Census Bureau provides some of these aggregates in the collapsed C tables (described in the previous chapter) for that purpose.

A third option would be to use a larger geography, sacrificing geographic detail for better estimate precision. Users can aggregate geographic areas or subject categories and use formulas to calculate margins of error for new estimates. For example, to add the population for two geographies together, you would sum the estimates and then take the square root of the sum of the squares of their margins of error to calculate a margin of error for the new estimate. For example, the Society Hill neighborhood in Philadelphia falls largely within census tracts 10.01 and 10.02. Using the 2015–2019 ACS, to create an aggregated neighborhood estimate of the number of people who take public transit to work, you would sum this population from Table B08006 for the two tracts:

```
273 + 436 = 709
```

And then take the square root of the sum of the squares of their margins of error:

```
√((80^2)+(151^2)) = 171
```

A statistic called the coefficient of variation (CV) is used to characterize the reliability of an estimate. To calculate a CV, you divide the margin of error by 1.645 (the constant *Z* value for a 90 percent confidence level), divide that result by the estimate, and then multiply by 100. Opinions vary on what an acceptable CV level is; as a general rule of thumb, CVs of 0 to 12 can be considered as highly reliable, 13 to 34 as medium reliability, and 35 and above as low reliability.

To calculate the CV for public transit commuters in the Society Hill neighborhood:

```
((171/1.645)/709)*100 = 15
```

This estimate would be interpreted as having medium reliability, as it has a value between 13 and 34. The CVs for the individual tracts, 10.01 and 10.02, are 18 and 21, respectively. While these are also of medium reliability, the aggregated neighborhood CV is lower and thus is considered to be more reliable.

A number of tools and resources exist to help researchers understand and work with ACS estimates (see the tools mentioned in the gray box under ACS Resources), and the Census Bureau provides recommendations with detailed examples in a series of guidebooks (US Census Bureau 2020b).

It's particularly important to recognize the fuzziness of estimates when comparing data over time. Given that many differences in values from one year to the next are more likely due to sampling variability rather than actual change, the one-year estimates do not lend themselves to being studied as a time series. There is a test for statistical difference that allows you to determine whether two different estimates are truly different. For the five-year estimates, only nonoverlapping periods should be compared; otherwise you would be comparing estimates that are generated from the same sample pool. For example, comparing estimates from 2010–2014 and 2015–2019 would be sound, as there is no sample overlap between the two series.

Researchers should use the ACS when they

- need detailed socioeconomic characteristics of the population
- are doing historical comparisons of these characteristics that are not captured in recent decennial censuses
- need current data
- can live with the fuzziness of the estimates

The ACS will not be your first choice if you need precise counts (particularly for small population groups and geographies), if you are studying basic demographic variables that are available in the DEC and timeliness is not an issue, or if you want an annual time series.

ACS Resources

ACS Calculator
Online calculators for significant difference and computing a new estimate from two existing estimates, from the Cornell Program on Applied Demographics
https://pad.human.cornell.edu/acscalc/index.cfm

ACS Tools
Excel spreadsheets with pre-built macros for significant difference and calculating new estimates, from the Economic, Demographic, and Statistical Research Unit of Fairfax County, VA (scroll to the bottom for the tools)
https://www.fairfaxcounty.gov/demographics/research-tools

Population Estimates

Whereas the DEC is a count and the ACS is a sample survey, the Population Estimates are generated from administrative records using a series of calculations. Using the DEC as a base, components (births, deaths, and migration) for different cohorts of the population (age, sex, race, Hispanic origin) are used to create new estimates for successive years. The cohort-component method is a basic approach used in demography; for a given place, data on births and deaths is used to calculate how many people were born in the first age group and how many people survived and thus moved from their current age group to the following one. Different sources for domestic and foreign migration data are used to determine how many people moved into and out of an area. The end result is a new estimate of the population for that given year. The process is repeated the following year using the previous year as a base.

The Census Bureau uses this approach to generate estimates for the nation, states, counties, and metropolitan areas each year. It is an iterative process carried out in stages to ensure that the sum of smaller geographies equals the whole. The data is produced in two sets: estimates of the components (births, deaths, migration) and of the characteristics (age, sex, race, Hispanic origin). A different methodology is used to generate basic population totals for places (cities and towns) and an estimate of total housing units for all geographies.

The Census Bureau releases the estimates in a series called a vintage. Each vintage contains the latest year of estimates, plus all previous years back to the previous DEC. The estimates for previous years may differ from previous vintages, as the Census Bureau modifies and improves its models and methodology based on new information. This means that an estimate for the year 2018 that appears in the 2019 vintage may be different from the estimate for 2018 that appeared in the 2018 vintage. For that reason, it's important for users of this data to wholly replace each vintage with the subsequent one. At the beginning of a new decade, a final vintage based on the DEC from ten years past is used to gauge the accuracy of the latest DEC. Following this release, two things happen. First, using the latest DEC the Bureau will go back and revise the annual estimates for the previous decade for a final time, based on what the latest count shows. This series is referred to as the Intercensal Estimates, and it becomes the definitive estimate series for that decade. Second, the subsequent vintage estimates for the next decade use the new DEC as the estimates' base.

Compared to the DEC and ACS, the estimates are a much smaller data series with far fewer tables and variables, and thus the dataset is easier to access and work with. In addition to data.census.gov and the API, users can easily search and browse through spreadsheets published on the Population Estimates Program website. The series is well suited for studying annual population change and basic demographic characteristics for large areas. The estimates are the only series that contains the components of population change—births, deaths, net domestic migration, and net foreign migration—and thus are the source for researchers interested in studying these components. Natality and mortality data in the US is compiled at the county level by local health departments, is reported up to the state level, and is collated at the federal level. Domestic migration data is sourced from the IRS and the Social Security Administration, while foreign migration data comes from multiple sources based on migration origin (i.e., US military overseas, US territories, foreign countries).

Researchers should use the Population Estimates when they

- need basic population counts and characteristics for large areas
- are studying annual population change
- need data on the components of population change

The population estimates will not be your first choice if you need more detailed characteristics or data for small geographic areas.

Business Establishments

The Census Bureau has been collecting and summarizing data on businesses almost as long as it has been collecting population data (Micarelli 1998). There are several different statistical programs that measure business activity, and in this section we will discuss two of the main programs: the Business Patterns and the Economic Census.

Both of these programs compile statistics on business establishments, which are defined as single physical locations where business is conducted or where services or industrial operations are performed. Establishments are assigned to industries, which are groups of businesses that produce similar products or provide similar services, using the North American Industrial Classification System (NAICS). Introduced in 1997, NAICS is used by government agencies in the US, Canada, and Mexico for classifying industrial activity, replacing the earlier Standard Industrial Classification (SIC) system. NAICS assigns businesses into broad groups and detailed divisions and subdivisions, with two-to-six-digit codes that indicate related groupings and the level of detail. Table 4.2 illustrates the different levels of NAICS for the Offices of Physicians Industry, with data from the 2019 Business Patterns for Rhode Island. This industry includes

Table 4.2. NAICS Codes illustrated for the Offices of Physicians Industry in Rhode Island, 2019 Business Patterns

NAICS Code, Title, and Level	Establishments	Employment
62 Healthcare and Social Assistance [sector]	3,113	87,067
- 621 Ambulatory Healthcare Services [subsector]	1,978	27,680
-- 6211 Offices of Physicians [industry group]	648	8,953
--- 62111 Offices of Physicians [industry]	648	8,953
---- 621111 Offices of Phys. (except mental health) [U.S. industry]	607	8,786
---- 621112 Offices of Phys., mental health [U.S. industry]	41	167

establishments where doctors are primarily engaged in the independent practice of general or specialized medicine or surgery, operating in private or group practices in their own offices or in the facilities of others, such as hospitals or medical centers.

The NAICS is revised every five years, in years that coincide with the Economic Census, to adapt the categories to fit a changing economy. Throughout NAICS there are miscellaneous categories that capture business activities that are not explicitly captured elsewhere. If an industry grows in importance over time, it may emerge as a distinct category in a revision, whereas an industry that declines in importance may be folded into a miscellaneous group. Each version of the NAICS is named for the year of its revision (e.g., NAICS 2017 and NAICS 2022). Concordances or crosswalks are published that allow researchers to relate the changing categories over time. The Economic Census and the Business Patterns use the NAICS version that coincides with their years (i.e., the 2017 Economic Census and the 2017 to 2021 Business Patterns use NAICS 2017).

The Business Patterns is an administrative dataset generated annually from the Business Register, a federal database of all businesses with paid employees in the United States. It is often referred to as the County Business Patterns and ZIP Code Business Patterns, the two most commonly used geographies in the series (which also includes states, metro areas, and Congressional districts). Unlike the population datasets, the Business Patterns and Economic Census report data by actual USPS ZIP codes instead of ZCTAs, as the ZIP information is simply scraped from the address record of the business. The Business Patterns series is relatively small and includes summary data on employees, establishments, and payroll by NAICS and geographic area. There are also tables that count establishments by employee size and as multiunit firms.

The Economic Census is a larger undertaking that generates data from counts, sample surveys, and administrative records. It is conducted every five years, in years ending in two and seven. It captures the same indicators included in the Business Patterns, as well as data on production and sales. Data is also published for the same geographies, plus cities and towns, and there are industrial as well as geographic summaries. The Economic Census is often used for studying trends as opposed to providing current statistics, as it takes several years from the time data is compiled to the time it is released. There are some slight differences between what industries the Business Patterns and Economic Census cover. Neither captures agriculture, government, or independent consultants and contractors, which are covered in different data programs (the US Department of Agriculture's Census of Agriculture and the US Census Bureau's Census of Governments and Nonemployer Statistics).

NAICS Reference and Other Business Datasets

NAICS
https://www.census.gov/naics/

Census of Agriculture
https://www.nass.usda.gov/AgCensus/

Census of Governments
https://www.census.gov/programs-surveys/cog.html

Nonemployer Statistics
https://www.census.gov/programs-surveys/nonemployer-statistics.html

Bureau of Labor Statistics Data Portal
https://www.bls.gov/data/

Like the population and housing datasets, the business datasets also fall under confidentiality regulations. Data on individual businesses is never reported; all data is summarized by geography and NAICS. There are a number of disclosure mechanisms with varying degrees of restriction designed to prevent users from reverse engineering the data to obtain establishment-level information. The Census Bureau may choose to inject noise into the employment or payroll values by inflating or deflating them by up to 10 percent, or may publish a range of values instead of a specific one, or may not publish the data at all if there are fewer than three establishments in a given industry or area or if one establishment comprises a large majority of employment or payroll. These policies present a range of challenges to data users. The disclosure mechanisms are indicated with footnotes that are embedded

in the data in place of actual values, which can create data processing challenges. Also, the sum of smaller parts will seldom equal the whole; if data is not disclosed at a six-digit NAICS level, it may be included in the four-digit level above it, or if data for an industry is not disclosed for a county, those statistics may be included in the state totals. Thus, aggregating data from smaller to larger parts leads to omission errors.

Both of these datasets count establishments and employees based on the location of the establishment. Beyond these series, there are other census datasets, like the ACS and CPS, that measure the labor force and typically count workers where they live (and in a few cases based on where they work). The Bureau of Labor Statistics, which collaborates with the Census Bureau in producing the CPS, also has a number of different data and survey programs for measuring business and labor activity.

Researchers should use the Business Patterns when they

- want annual data on establishments, employment, and wages by industry for large geographies
- want business data for ZIP codes and Congressional districts not published elsewhere

Researchers should use the Economic Census when they

- are doing long-term research, where timeliness is less important
- need data on production and sales for certain industries
- need data for ZIP codes, towns, and cities not collected elsewhere

References

IPUMS. 2021. "Changes to Census Bureau Data Products." IPUMS website. Minneapolis: Minnesota Population Center. https://www.ipums.org/changes-to-census-bureau-data-products.

Jung, Paul H., Jean-Claude Thill, and Michele Issel. 2019. "Spatial Autocorrelation and Data Uncertainty in the American Community Survey: A Critique." *International Journal of Geographical Information Science* 33, no. 6: 1155–75.

Jurjevich, Jason R., Amy L. Griffin, Seth E. Spielman, David C. Folch, Meg Merrick, and Nicholas N. Nagle. 2018. "Navigating Statistical Uncertainty: How Urban and Regional Planners Understand and Work with American Community Survey (ACS) Data for Guiding Policy." *Journal of the American Planning Association* 84, no. 2: 112–26.

Micarelli, William F. 1998. "Evolution of the United States Economic Censuses: The Nineteenth and Twentieth Centuries." *Government Information Quarterly* 15, no. 3: 335–77.

Nesse, Katherine, and Mallory L. Rahe. 2015. "Conflicts in the Use of the ACS by Federal Agencies." *Population Research and Policy Review* 34: 461–80.

Salvo, Joseph J., Arun Peter Lobo, Adam L. Willett, and Joel A. Alvarez. 2007. *An Evaluation of the Quality and Utility of ACS Five Year Estimates for Bronx Census Tracts and Neighborhoods*. New York: Population Division, New York City Department of City Planning. https://www1.nyc.gov/assets/planning/download/pdf/data-maps/nyc-population/eval_quality_utility_acs_five_year_estimates_bx.pdf.

Spielman, Seth E., David Foch, and Nicholas N. Nagle. 2014. "Patterns and Causes of Uncertainty in the American Community Survey." *Applied Geography* 46: 147–57.

US Census Bureau. 2020a. "2020 Census Response Rate Update: 99.98% Complete Nationwide." Press release no. CB20-CN.120, October 19. https://www.census.gov/newsroom/press-releases/2020/2020-census-all-states-top-99-percent.html.

———. 2020b. *Understanding and Using American Community Survey Data: What All Data Users Need to Know*. Washington, DC: US Department of Commerce, September. https://www.census.gov/programs-surveys/acs/guidance/handbooks/general.html.

———. 2021a. "Census Bureau Announces Changes for 2020 American Community Survey 1-Year Estimates." Press release no. CB21-123, July 29. https://www.census.gov/newsroom/press-releases/2021/changes-2020-acs-1-year.html.

———. 2021b. *Comparing Differential Privacy with Older Disclosure Avoidance Methods*. D-FS-GP_EN-0509. Washington, DC: US Department of Commerce, August. https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/disclosure-avoidance.html.

———. 2021c. *Protecting the Confidentiality of the 2020 Census Redistricting Data*. D-FS-GP-EN-049. Washington, DC: US Department of Commerce, August. https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/disclosure-avoidance.html.

———. 2021d. "2020 Census Operational Quality Metrics: Item Nonresponse Rates." *Random Samplings* (blog), August 25. https://www.census.gov/newsroom/blogs/random-samplings/2021/08/2020-census-operational-quality-metrics-item-nonresponse-rates.html.

Wang, Hansi Lo. 2021. "For the U.S. Census, Keeping Your Data Anonymous and Useful Is a Tricky Balance." NPR, August 2. https://www.npr.org/2021/05/19/993247101/for-the-u-s-census-keeping-your-data-anonymous-and-useful-is-a-tricky-balance.

Chapter 5

Accessing Data

There are many different websites and portals for accessing census data. As it's free and in the public domain, organizations take the data and package it in different ways to make it more accessible for certain groups of users and to add additional features, such as derived datasets, data visualizations, and map-based interfaces. Many portals make it easier to access data by filtering it down to certain subsets so that there is less to wade through. The Census Reporter was created by journalists to provide access to just the most recent iteration of the ACS, with convenient search mechanisms, infographics, and the ability to create basic shaded-area maps. The Missouri Census Data Center provides many different applications (for the entire US, not just Missouri). Its data profile tools provide just the five profile tables from the DEC and ACS in a few simple clicks, with the ability to compare up to four geographies. It flags ACS data for reliability based on its CV and provides graphs and charts for many indicators. Many state and local government agencies will repackage data for their jurisdictions on their websites, making it easier for local residents to access. There are also a number of proprietary library databases that focus on providing census and other government datasets, which we'll mention in the next chapter.

Librarians should consider the nature of each researcher's request and their technical capabilities when suggesting a specific tool or website and match the needs of the user to the resource. Some users would benefit from a simple source that entails just pointing and clicking, while others need a fuller range of capabilities and access to a greater number of datasets. It's always best to present users with a range of options so they can choose what works best for them.

In the following sections we will look at data.census.gov and the Census Bureau's API, which are two of the Bureau's primary portals for accessing a full range of datasets. Data.census.gov has both basic and advanced features that can satisfy the needs of novice and advanced users, while the API is suitable for users with a background in scripting or coding. We will briefly mention where you can find data briefs and research reports that summarize population trends.

Open Census Data Portals

Data.census.gov
https://data.census.gov/cedsci/

Census Data APIs
https://www.census.gov/data/developers/data-sets.html

Missouri Census Data Center
https://mcdc.missouri.edu/

Census Reporter
https://censusreporter.org/

Also check state and local government sources: regional planning, city planning, economic development, population divisions, labor departments.

Data.census.gov

Data.census.gov is the Census Bureau's primary portal for accessing many of its datasets. Launched a few years prior to the 2020 census, it was designed to provide a modern search-based interface for accessing data. The main page has a search box where users can type in a geography, topic, or table number to retrieve results that can be subsequently browsed and filtered. The search works well for casual browsing, and in particular for retrieving profiles on places (lots of data for one place) with helpful infographics.

The search concept has its limits as you are searching across data tables as opposed to documents

and text, which limits the effectiveness of keyword searching. Given the large volume of available data for dozens of datasets for dozens of years, a filter-based approach works best. By clicking on the advanced search option, users can limit the range of datasets to a small, usable selection that can be browsed through and downloaded more easily. Using the advanced search requires knowledge of the datasets and basic census concepts, all of which we have covered thus far. The five advanced search filters are as follows:

1. Dataset. Select the specific dataset to eliminate the others. For example, if you want data on educational attainment, that's something that's captured only in the ACS. If you want data for all the counties in a state, many counties have fewer than 65,000 people, so you would need to use the five-year ACS. Then you would choose between Profiles, Subject Tables, or Detailed Tables.
2. Year. A simple selection that filters out quite a bit. Choose the latest year if that's what you need. If you want decennial data, then your choice would be either 2020 or 2010. For the ACS, the year represents the one-year series and the last year in the five-year series, so the 2019 ACS refers to the 2019 one-year and 2015–2019 five-year estimates.
3. Geography. The nesting rules covered in chapter 3 largely apply here for selecting areas within other areas. Select all counties in the state or nation, or all census tracts in a county or state, or all places in a state, or select specific areas one by one.
4. Topic. This is the most open-ended of the options. Try selecting topics that you think are relevant to the specific variables you are looking for. If you are looking for DEC or ACS data profile tables, you can skip this step as there are only a few tables.
5. Industrial Code. Important for the business datasets if you want statistics for a particular industry.

Once the filters are selected and applied, an intermediary page appears that provides top results for tables, maps (as basic web mapping capabilities are part of the platform), and web pages on the Census Bureau website. Click right through to the tables to see all the options returned, and from there you can preview each table or start keyword searching if there are too many results. Download the data as a CSV for machine-readable files where each row is a geography and the variables are stored in columns. The CSV can be imported into a spreadsheet, scripting language, or any program. Download the data as an Excel spreadsheet for a human-readable version that resembles what's depicted on the screen. This format is fine for presentation and for looking up statistics, while the CSV is better for data processing, visualization, and analysis.

The ability to filter tables exists at multiple points in the process. You can filter at the beginning by doing the advanced search. Or, if you do a basic search and click your way through to a table, you can apply the filters on the table results screen. So if you did a basic search for educational attainment and got to a relevant table for the United States, in the table results you could apply a filter for a specific geography and year.

For users who want an exact listing of every single table to definitively evaluate available content, they can go to the individual program web pages for either the DEC or ACS on the Census Bureau's website. In the technical documentation section for each series there are PDF reports as well as Excel spreadsheets that list every table along with its ID number. The basic data.census.gov search works well for known-item searches; enter the table's ID number, move to the table results page, and from there apply filters for geography and year. For smaller datasets like the Population Estimates and Business Patterns, their program websites will have direct links to either spreadsheets or CSV files that you can browse through and download, bypassing data.census.gov altogether.

Census Bureau API

An application programming interface (API) allows users to tap into data repositories through a script or program, which offers a number of benefits. In using a GUI-based portal like data.census.gov, users must manually point and click to retrieve and download data. The download might include multiple tables that would need to be filtered or stitched together. A number of other manual tasks for processing and preparing the data for analysis may follow, likely in a spreadsheet package. Unless the user documented and shared the steps they took, their process would be opaque to subsequent viewers of the data. In contrast, by accessing an API through a script, coders can create highly specific queries to retrieve exactly what they want, and that data is pulled directly into their script where it can be processed, analyzed, or visualized and subsequently output to any number of data formats. The script allows for the automation of processes and serves as documentation that describes exactly how data was retrieved and manipulated. There is a growing need for librarians to become familiar with using APIs to assist researchers with accessing datasets, and this skill is a logical extension to the work many data librarians are increasingly doing (White and Powell 2019; Adams 2018).

There is a learning curve for writing scripts, but once you have a grasp on the basics, the technical process of retrieving data through an API is straightforward. The bigger challenge is having an understanding of census concepts, as these are used for structuring the data and must be understood for making requests. In a REST API, the user builds links

```
# Python Script for Census API
import requests,csv

#Set variables
year='2019'
dsource='pep'
dname='population'
cols='NAME,POP,DATE_DESC'
state='10'
county='*'
dcode='2,12'
keyfile='census_key.txt'
outfile='pop2019_de_counties.txt'

base_url = f'https://api.census.gov/data/{year}/{dsource}/{dname}'

#Read api key in from file
with open(keyfile) as key:
    api_key=key.read().strip()

#Retrieve data, print output to screen
data_url = f'{base_url}?get={cols}&DATE_CODE={dcode}&for=county:{county}\
    &in=state:{state}&key={api_key}'
response=requests.get(data_url)
popdata=response.json()
for record in popdata:
    print(record)

#Write data to CSV
with open(outfile, 'w', newline='') as writefile:
    writer = csv.writer(writefile, quoting=csv.QUOTE_ALL, delimiter=',')
    writer.writerows(popdata)
```

to data they wish to retrieve. Specific attributes that describe the data are passed into the link as variables. The program passes that URL to the API to request that specific data. If successful, the API returns the data to the user in a container that can be manipulated in the script. The methods and functions of the specific programming language are used to loop through the data that's returned to extract, process, visualize, or output it to a specific format.

Consider the sample script above, which uses the popular Python language to retrieve recent population estimates for the three counties of Delaware. At the top of the script, two modules are imported that provide additional functionality to the core Python language. The Requests module is popular for working with APIs, while the CSV module is used for manipulating data stored in a plain, delimited-text format. In the next portion of the script, a number of variables are set that will be passed into a URL. Every census program (DEC, ACS, estimates, etc.) has an API, and every iteration and series in each program has a specific web page that documents what its API provides. These pages describe how the data is structured, illustrate what variables and geographies are available for the given dataset, and provide examples. In this example, the year, dataset (PEP for Population Estimates Program), and a data name that varies with each census program are provided. In the Population Estimates, there are options for total population, components of change, and characteristics. Below the variables, a base URL that has these components is constructed; the variable names that appear in braces will be replaced by the actual variables stored at the top of the script.

More specific variables for the request are defined next. The specific columns to retrieve must appear as one string of text separated by a comma. For the geography, the ANSI/FIPS code that uniquely identifies Delaware is provided, and an asterisk for the county to return all counties in Delaware; alternatively specific counties could be passed as a string of text with the county ANSI/FIPS codes separated by a comma. Counties nest within states, so the request must be made in this manner. The `dcode` variable is a date

```
# Output from Python Script for Census API
['NAME', 'POP', 'DATE_DESC', 'DATE_CODE', 'state', 'county']
['Kent County, Delaware', '162350', '4/1/2010 population estimates base',
'2', '10', '001']
['Kent County, Delaware', '180786', '7/1/2019 population estimate', '12',
'10', '001']
['New Castle County, Delaware', '538484', '4/1/2010 population estimates
base', '2', '10', '003']
['New Castle County, Delaware', '558753', '7/1/2019 population estimate',
'12', '10', '003']
['Sussex County, Delaware', '197103', '4/1/2010 population estimates base',
'2', '10', '005']
['Sussex County, Delaware', '234225', '7/1/2019 population estimate', '12',
'10', '005']
```

code that is specific to this data file and indicates the years to retrieve; these codes are provided in the online documentation.

Most APIs require you to register with the organization in order to use them, and you must agree to its terms of service. Registering for a census API key is simple and free and can be done from the main census API website. The API key is an alphanumeric string that you must append to your requests to identify that it's coming from you. It's a best practice to not embed an API key in a script, but to store it in a file that you read in so that it's not exposed to others. When uploading scripts into a repository like GitHub, the key file should be ignored so that it is not included in the repository. In this example, the key filename is stored as a variable, and the file is stored in the same folder as the script.

Once the base URL is created, the file is opened and the key is read in. A fuller data URL is built by appending additional variables to it. This isn't strictly necessary; you could build one big link all at once, but this approach makes the code a bit more readable. If this script were longer, this approach would also allow for additional requests without having to re-create the base. Once the variables are passed into the URL, the actual link looks like this:

```
https://api.census.gov/data/2019
/pep/population?get=NAME,POP,DATE
_DESC&DATE_CODE=2,12&for=county:*&in=s
tate:10&key=APIKEYGOESHERE
```

Next, this URL is passed to the internet using the Requests module, and the outcome is saved in a variable called response. If all went well, the response can be read as a JSON object, which is a standard data format used by APIs to return data. The Census Bureau uses a simplified form of JSON, where the data is returned as a nested list, with each sublist representing a single record and each object in the list as a variable that we requested, separated by a comma. The script loops through the list and prints each record to the screen; the first record has the column headers, while the following records contain the data, in this case one record for each county and year as shown in the example above. Some variables were returned that weren't specifically requested, such as a date description that goes along with the date code, and the state and county ANSI/FIPS codes. At this point, you could loop through these records and manipulate them however you would like. In this example, we loop through the records and write each row out to a CSV file.

This is a minimally working example, and there are countless variations you can employ when writing your own programs. One common addition would be to incorporate protocols for handling errors. With the Requests module you can request status codes to indicate whether the API service is currently available and whether or not a request was correctly formed or allowed prior to processing data. For large requests it's a best practice to incorporate try and exception blocks, where you try to connect and retrieve data but do something else when an error occurs. Larger data requests will require you to loop through and do multiple iterations. To accommodate errors like an interruption in service, you would want to structure your code to hold on to data that has been retrieved and identify where the process has stopped so you can pick up where you left off when launching the script again. The default location for reading and writing files is the folder where the script is stored. Python's OS module can be used for navigating your file system and specifying different input and output locations.

Python is a popular, general purpose, open source scripting language and a good place to start if you are new to coding, but many programming languages can be used to make API requests. The R statistical language is a popular alternative, particularly for

researchers who plan to do statistical analysis. A recent issue of *Library Technology Reports* provides a crash course in R (Glowacka-Musial 2021), and Python and R have been reviewed for their suitability for librarians who wish to expand their programming skills (White and Powell 2019).

While using an API provides obvious benefits, there are certain scenarios where other options would be best. If a particular census table has all of the necessary data for a particular need, it's easier to simply download it from data.census.gov and then read the CSV file into a script for processing. Some datasets, such as the Population Estimates and Business Patterns, provide ready-to-use CSVs and spreadsheets that are straightforward to download and would save the time of writing out a script to do the job. On the other side of the spectrum, if you need a lot of data, like all the DEC or ACS tables for all geographies in a state, the Census Bureau's FTP site allows you to download this data in bulk. You can subsequently load it into a statistical package or relational database.

Reports and Data Summaries

While some users are looking for census data for doing their own analyses and supporting their writing or research, others may be seeking reports that summarize and provide context for the data. What are current population trends in the United States? What is the geographic concentration of different racial and ethnic groups? Are incomes growing or declining? Where are Americans moving? The Census Bureau publishes an extensive series of reports in its online library. Some are recurring series, such as changes in income year by year, while others represent special topics, such as the impact of the Great Recession on school enrollment. Most of the reports are tied to the study of a specific dataset, and you can filter reports by dataset, year, and topic. With each DEC, a series of reports is issued that studies population change, aging, changes in the composition and location of each racial and ethnic group, changing household and family relationships, and a summary of housing characteristics.

In addition to the Census Bureau, a number of nonprofit think tanks such as the Brookings Institute and the Pew Research Center regularly analyze census data and publish their latest findings in reports and extended blog posts. Demographic research centers at universities across the country will also publish data briefs and technical reports that study either national trends or local ones based on where the college is located.

Census Reports and Analysis

Census Bureau Reports
https://www.census.gov/library/publications.html

Brookings Institute: Demographics and Population
https://www.brookings.edu/topic/demographics-population/

Pew Research Center
https://www.pewresearch.org/

Population Reference Bureau
https://www.prb.org/what-we-do/focus-areas/u-s-census-american-community-survey-acs/

Carsey School of Public Policy, University of New Hampshire
https://carsey.unh.edu/publications

References

Adams, Richard Manly, Jr. 2018. "Overcoming Disintermediation: A Call for Librarians to Learn to Use Web Service APIs." *Library Hi Tech*, 36, no. 1: 180–90.

Glowacka-Musial, Monika. 2021. "Data Visualization with R for Digital Collections." *Library Technology Reports* 57, no 1 (January).

White, Philip, and Susan Powell. 2019. "Code-Literacy for GIS Librarians: A Discussion of Languages, Use Cases, and Competencies." *Journal of Map and Geography Libraries* 15: 45–67.

Chapter 6

GIS, Historical Research, and Microdata

GIS Data

Since most census data is geographic in nature, it lends itself well to geographic analysis. Geographic Information Systems (GIS) are collections of software and data for conducting geographic analyses and making maps. Special GIS data files called vector files store coordinates to form geometries that represent points, lines, and areas. Each file represents a particular type of feature that covers a specific extent: a file for counties for the US, a file of census tracts for a state, or a file of roads for a particular county. The files are georeferenced, which means they are drawn to scale using a specific spatial reference system that ties them to real locations on the earth. These reference systems allow GIS data files from multiple sources to be overlaid in a GIS project. In addition to the vector files, a different format called a raster represents continuous surfaces as a series of grid cells of equal size, where each cell has a value that represents something about the surface. Rasters are also georeferenced, and satellite imagery, air photos, and scanned paper maps such as topographic maps are stored in the raster format. For census mapping applications, rasters are useful as base maps to provide context for vectors.

Desktop GIS software provides a lot of power for cartography and geographic analysis, particularly in combination with GIS data from multiple sources. For example, with census tract boundaries, population data for those tracts, and a point layer of public libraries created from the Institute of Museum and Library Services data files, you can map census tracts by the population under eighteen and select all tracts that fall within a mile or two of a public library. This allows you to measure youth population near each library, as well as measuring the population that falls outside this zone. You can measure the distance from each tract to the nearest library to generate measures of accessibility for different areas. Vector layers can be overlaid on top of raster maps or web mapping services, such as the OpenStreetMap, to provide context for your layers. We will explore census mapping for LIS research in the final chapter.

The Census Bureau publishes vector GIS files that represent all of the legal and statistical areas that they publish data for, as well as files that represent features that are used for drawing these areas, including roads, railroads, water bodies, and other landmarks. These GIS files are part of the TIGER geographic database, which is the system the Census Bureau uses to update and maintain all of the geographies it uses for its operations. The features in this database follow the geographic summary level and nesting rules to maintain structural integrity of boundaries so that the boundaries of smaller features fit within the appropriate larger ones.

All of the TIGER files are published on the Census Bureau's website and can be freely downloaded for use in desktop GIS software and in web mapping applications. The vector format is published in several different file types, the most common being the ESRI shapefile, which is an open legacy format that's widely supported. Other options include Google KMLs, GeoJSON, and the native TIGER file format. Most vector formats can be used in any GIS package. ArcGIS Pro is the best-known proprietary GIS package, while QGIS is a popular free and open source alternative.

The Census Bureau publishes several different iterations of the TIGER files for different mapping purposes. The official TIGER files represent the precise legal and statistical boundaries delineated by the Bureau. Since the boundaries of features incorporate both land and water, the shapes may appear unusual at first glance and may not be ideal for thematic mapping. The Bureau publishes a derivative of TIGER called the Cartographic Boundary Files, where large coastal water bodies have been removed so that features better represent land areas. The linework of these files has been generalized to smooth out boundaries and remove small features like islands that

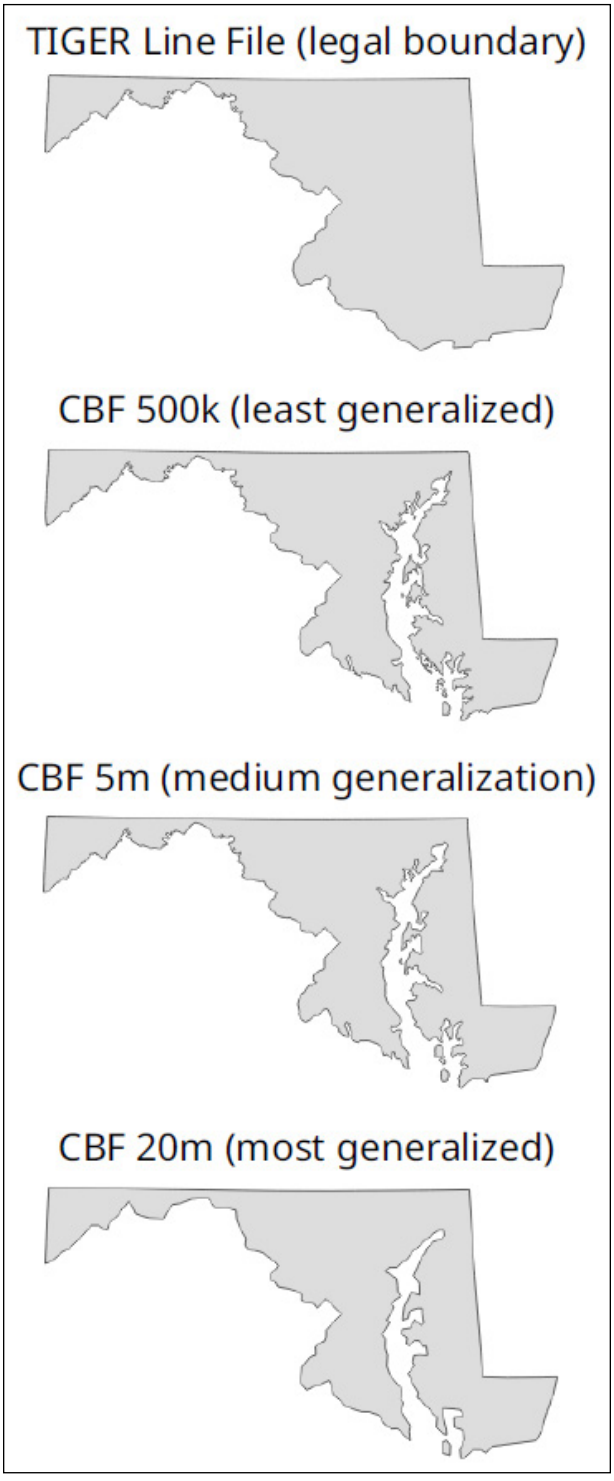


Figure 6.1
Boundary differences between the TIGER files and the Cartographic Boundary Files for Maryland

would not be visible at certain scales. There are different iterations of the boundary files that are appropriate for different scales; the most generalized would be appropriate for a map occupying a postcard, as opposed to the least generalized version that would be more appropriate for a poster. Figure 6.1 illustrates the differences.

Census GIS Resources

TIGER/Line Shapefiles
https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html

Cartographic Boundary Files
https://www.census.gov/geographies/mapping-files/time-series/geo/cartographic-boundary.html

Census Geocoder
https://geocoding.geo.census.gov/geocoder

In shapefiles, attributes about the features are stored in a table where each row represents a geographic feature and each column contains attributes that describe the feature. The features that are visually depicted in the software's map view and the records in the table are tied together so that you can use the attributes to label, filter, query, and thematically map the features. In GIS packages, attribute tables operate according to similar principles as tables in relational databases. Columns are designated to store text, integers, or decimals, and one column serves as a unique identifier or primary key. In most instances, vector files do not come with lots of attribute data preattached. Users can add nonspatial data tables, where each row represents a specific geography and columns contain population or socioeconomic data, to GIS software and then join that data table to a corresponding vector file that represents the same features, using a unique ID column that they share in common. This allows you to create thematic maps of the population data, such as shaded-area maps where ranges of population values are classified into categories assigned a specific color.

Mapping census data in GIS is a multistep process that involves downloading the TIGER file that represents the features, downloading population data that contains the attributes you wish to map for those features from data.census.gov, processing the features and data table to prepare them for analysis, adding both files to the GIS package, using the package's tools to join the GIS features to the table using the Census Bureau's GEOID column (which has the unique summary level and ANSI/FIPS codes), and using the GIS package tools to symbolize the features based on the attribute you wish to map. Most of the data processing steps are associated with taking large areas or large numbers of variables and creating subsets out of them or conversely stitching together multiple shapefiles or

data tables into a larger set. All the TIGER files and data tables from data.census.gov come with census identifiers in multiple forms, including the full GEOID field, so features and data tables can be readily joined. If you retrieve data from the API or another source, you will want to be sure that you also retrieve identifiers that will match the vector file.

While GIS gives you a lot of power and flexibility, the learning curve can be steep. The Census Bureau does package pre-joined vector files with a selection of basic population data to create special geodatabases. These can lower the curve by eliminating data processing work. For non-GIS users, the Bureau publishes a number of static maps depicting census boundaries, as well as selections of thematic maps. For users who want to do basic data exploration and mapping, data.census.gov and other free tools such as the Census Reporter provide basic web mapping capabilities, where you can select a type of geography and variable and make a map. There are also a number of proprietary library database products that public and academic libraries subscribe to. These tools add value by pulling a lot of disparate data together in one place and providing easy-to-use tools for creating thematic maps, doing basic querying and analysis, and downloading data.

Census Database and Mapping Products

PolicyMap
https://www.policymap.com/

SimplyAnalytics
https://simplyanalytics.com/

Social Explorer
https://www.socialexplorer.com/

GeoLytics
https://www.geolytics.com/

Historical Data

Historical analysis, whether it's studying recent or long-term trends, presents a number of challenges. In terms of access, most public data portals such as data.census.gov and the APIs do not archive data that's more than a few decades old. There are also few census products that compare change between two periods in a single table or file. The ACS comparative profiles compare two nonoverlapping five-year periods. There are no official DEC tables (i.e., part of the summary files with official table identifiers) that compare two successive censuses, but there are some special products published on the DEC program websites.

The primary public archive for all historical summary data is the National Historical Geographic Information System (NHGIS). This is one of many projects that's part of the IPUMS series of repositories created at the Minnesota Population Center at the University of Minnesota (Kugler and Fitch 2018). Users must register and create an account, but it's an academic, nonprofit project and registration is free. Users are asked to cite the NHGIS as the source and use the data for noncommercial purposes. Summary data from every DEC, all versions of the ACS, pre-twenty-first-century data from the Business Patterns, and several other special datasets are available. GIS vector boundary files for each decade are also available for depicting areas as they appeared for each decennial census. As in the advanced search in data.census.gov, users apply filters to choose a dataset, year, geography, and topic and can then browse through the returned tables and download them. Most of the data is presented nominally, the way it was originally published. The NHGIS has assembled a limited series of time series tables that allow you to compare the same or similar variables across several decades, from the 1970 census to the present.

One of the challenges in studying the census over time is that census geography changes; statistical areas are updated for each DEC, and legal areas change on an ongoing basis. One approach for grappling with this challenge is to use normalized instead of nominal data. Normalized data has been modified so that data from several points in the past has been altered to fit present-day boundaries, so comparisons within the same area can be made. Census tracts are used as the basis for creating normalized data, as they have been drawn and numbered in a way to insure some degree of consistency over time. Converting the past data to modern boundaries is achieved by aggregating tracts (if two old tracts were eventually combined into one), splitting tracts (if one old tract was subsequently split into two), or apportioning data in instances where a split was more complex.

The NHGIS publishes a series of normalized data tables (referred to as geographically standardized tables) from the 1990 to 2010 census using 2010 geography, for states down to block groups. The Longitudinal Tract Data Base produced at Brown University publishes a crosswalk for users to normalize their own data for census tracts from 1970 to 2010 using 2010 geography, and it publishes a limited number of normalized data tables for users to download (Logan, Xu, and Stults 2014). The GeoLytics company has been a long-time producer of normalized census data, and its Neighborhood Change product provides normalized tract data from 1970 to 2010. Once the data for the 2020 census is fully released, it is likely that each

organization will create new series with data normalized to 2020 boundaries.

Other resources allow users to adjust nontract geography over time. For each DEC, the Census Bureau publishes block relationship files that crosswalk blocks from the previous DEC to the current one, as blocks are completely renumbered and their boundaries can change. As blocks are the smallest geographies, they can be aggregated to form any geography in the hierarchy. The Bureau also publishes an ongoing list of changes to county boundaries from 1970 to the present that can be used to adjust data. While county-level changes do occur, they are less frequent in modern times and most adjustments are straightforward. The composition of metropolitan areas changes quite frequently, at least two or three times a decade. Since metro areas are composed of counties, county-level data can be aggregated to create consistent metro definitions for comparison over time.

Historical Census Data Resources

NHGIS
https://www.nhgis.org/

Longitudinal Tract Database
https://s4.ad.brown.edu/Projects/Diversity/Researcher/LTDB.htm

Census Block Relationship Files
https://www.census.gov/geographies/reference-files/time-series/geo/relationship-files.html

Changes to Counties, 1970 to Present
https://www.census.gov/programs-surveys/geography/technical-documentation/county-changes.html

Historical Census Reports
https://www.census.gov/prod/www/decennial.html

National Archives Census of 1940
https://1940census.archives.gov/

National Archives Census of 1950
https://1950census.archives.gov/

IPUMS Complete Count Data
https://usa.ipums.org/usa/complete_count.shtml

For contemporary historical analysis, the 1970 census is often the initial entry point of comparison for reasons described in chapter 3. It was the first census conducted under the basis of self-identification and the first conducted entirely through the mail. The race and ethnicity categories that we use today had their origins in 1970, although the 1980 census was the first where the standardized Directive 15 categories were used on the 100 percent count form. The digital predecessor to TIGER was developed for the 1970 census, so there were readily available map files to work from, and by this period a significant portion of the country was covered by census tracts (or some corollary), allowing for the creation of normalized files. Fewer geographies are available as you go back further in time, and fewer that cover the entire nation. Census tracts were introduced as a census geography in the 1940 census, and only for urban areas.

Questions on the forms and categories for summarizing data also change, and accommodating these changes is more difficult. As mentioned previously, changes for NAICS can be accommodated through using concordances from one version to the next, but while there is a bridge between the 1997 NAICS and the 1987 SIC system, there are quite a few differences between them. Adjusting dollar values for inflation is straightforward and an absolute must when comparing financial statistics. The Pew Research Center (2020) has created a chart that records how the nation's racial categories have changed over time. In some cases statistics for changing categories can be quantified or estimated, such as the shift to including a multiracial option for the first time in the 2000 census, but in many cases researchers can only note the differences and advise caution in making comparisons. Studies on how specific questions or categories have changed over time, such as native language or language spoken at home (Stevens 1999), can provide guidance for making comparisons.

After seventy-two years, the original responses to the census questionnaires are released by the National Archives, opening up new possibilities for historical research. Prior to 1960, census data was captured on ledgers as opposed to a form for a single household, and these ledgers were microfilmed and sent to libraries throughout the nation. For the 1940 census, scanned PDFs of the ledgers were published on the Archives website, which allowed for remote access. This data is not machine-readable (the responses were written in longhand) and specific persons can be located only if you know their address. With the address, you could identify the enumeration district where the street is located so you can find the file/ledger with data for that street. Private companies such as Ancestry.com digitize and make the records searchable through their products. For social science researchers, the IPUMS publishes a variety of machine-readable, full count data files as part of the IPUMS USA project. The public files have been de-identified (names and addresses removed), but the full individual records are available. Data from 1900 to 1940 is linked, so individuals can be followed over time.

Microdata and the Current Population Survey

Census *microdata* refers to person- or household-level records that represent a sample of responses to census questionnaires, with personal identifying information removed. Some researchers work with microdata because they are interested in studying population or socioeconomic trends from perspectives other than geographical ones, such as broad trends across an entire population and differences in trends between different age, sex, racial, and economic groups. Other researchers use microdata to create special cross-tabulations that are not present in the public summary data.

Microdata records include a number of special attributes called weighting variables, which are used to generate population-level estimates. The weight indicates the number of people or households in the general population that a particular respondent represents. For example, to generate an estimate of the American Indian or Alaska Native population sixty-five and over, you would sum the person weights for all records where the race and age meet these criteria. Public use microdata samples (PUMS) are generated for both the DEC and ACS and can be accessed from the DEC and ACS program web pages on the Bureau's website. IPUMS USA is an alternate choice for creating DEC and ACS microdata extracts, where users generate samples by choosing variables of interest.

To protect confidentiality, estimates for small areas such as census tracts cannot be generated. The Census Bureau has created a special geography called a Public Use Microdata Area (PUMA) for presenting reliable estimates for microdata. Built from blocks and nesting within states, PUMAs are designed to have a population size of approximately 100,000 people. They are identified by a five-digit number and a place name that indicates an area of coverage. Researchers need to be on guard when generating estimates for small population groups, as the sample size may not be large enough to generate reliable estimates. There is documentation for advising users how to estimate reliability.

Beyond the DEC and ACS, the Census Bureau conducts a number of smaller surveys for which it publishes both state and national statistics as well as the microdata for the survey. Of all these surveys, the Current Population Survey (CPS) is the largest and most widely used. Launched in the 1930s, it was one of the government's earliest efforts to employ sample surveying to produce population-level estimates. It is conducted jointly by the Census Bureau and the Bureau of Labor Statistics, and its original purpose was to generate monthly unemployment numbers, a primary role that it continues to serve. The CPS samples 60,000 households a month from a stratified sample that includes every state and the District of Columbia. Census field representatives conduct the surveys in person or over the phone, which makes the responses highly reliable and guarantees a response rate of over 90 percent. The CPS is unique among the census survey programs in that it is longitudinal. The same 60,000 households are sampled monthly for four months, then are rotated out of the survey for four months, and are rotated back in for a final four months. There are a core set of questions that are asked every month, plus a supplemental set of questions that are asked on a set but limited basis. For example, the Annual Social and Economic Survey (ASEC) is a set of detailed socioeconomic questions asked every March, while questions on voter registration and participation are asked every two years in November to coincide with federal elections. The CPS captures many of the variables that are included in the ACS, plus a substantive number that aren't captured elsewhere.

Given the sample size of the CPS, the Census Bureau and the Bureau of Labor Statistics generally publish national and state-level estimates only. Researchers can access all the microdata records from the CPS program page or can create specific extracts from the IPUMS CPS website. Since the survey is longitudinal, there are identifiers that allow researchers to track person and household responses across several months of the survey. There are a number of web-based tools for users who want to generate estimates without having to download and weight the records themselves. For example, the Census Bureau's Micro Data Access Tool (MDAT) and the IPUMS Analyzer allow you to generate estimates from samples in the ASEC.

Microdata Resources

IPUMS USA
https://usa.ipums.org/usa/

IPUMS CPS and Online Analyzer
https://cps.ipums.org/cps/

Current Population Survey
https://www.census.gov/programs-surveys/cps.html

Micro Data Access Tool (MDAT)
https://data.census.gov/cps/mdat/

Census Research Data Centers
https://www.census.gov/about/adrm/fsrdc/locations.html

In addition to the PUMS, there are also restricted microdata samples that researchers can access by submitting research proposals to the Census Bureau.

Strict procedures must be followed to ensure that the confidentiality of responses is not jeopardized. If accepted, the researcher can visit one of several Census Research Data Centers to work with the data onsite. The centers are located throughout the United States, usually at large research universities.

References

Kugler, Tracy A., and Catherine A. Fitch. 2018. "Interoperable and Accessible Census and Survey Data from IPUMS." *Scientific Data* 5: article 180007. https://www.nature.com/articles/sdata20187.

Logan, John R., Zengwang Xu, and Brian J. Stults. 2014. "Interpolating U.S. Decennial Census Tract Data from as Early as 1970 to 2010: A Longitudinal Tract Database." *Professional Geographer* 66, no. 3: 412–20.

Pew Research Center. 2020. *What Census Calls Us: A Historical Timeline*. Washington, DC: Pew Research Center, February 6. https://www.pewresearch.org/interactives/what-census-calls-us/.

Stevens, Gillian. 1999. "A Century of U.S. Censuses and the Language Characteristics of Immigrants." *Demography* 36, no. 3: 387–97.

Chapter 7

The Census in Library Applications

Census data is invaluable for practitioner-based library work as well as LIS research. Census data can be used to help public libraries understand their communities, so they can provide relevant services and meaningful collections. Coupled with GIS, census data can also be used to measure library access and availability and for the purpose of siting new libraries. In this chapter we will briefly review the literature and methods for using census data and GIS to study library issues.

Public libraries have a mission to serve communities within a defined legal area, such as a county, city, or town. Studying census data for the county or town can help the library understand the community it serves and can help inform decision-making. How many children are there relative to senior citizens? Is there are large population of residents who speak a language other than English at home? Is there a large percentage of people who don't have computers at home? What is the unemployment rate? The answers to all of these questions can inform library decisions in terms of what materials to buy and what programming to conduct in terms of workshops, lectures, and other events.

Beyond the legal service area, the LIS literature has recognized that there is a functional service area where the library attracts more people who live and work nearby as opposed to those who are farther away. In county and city library systems with multiple branches, studying the community that's in the neighborhood will be more meaningful than studying the entire city or county. Neighborhoods are informal areas that are not strictly defined, and the Census Bureau does not delineate or publish data for anything resembling a neighborhood. To study neighborhoods, one needs to use small census geography such as block groups or census tracts and aggregate the data for these geographies to form neighborhood-like areas based on formal and informal definitions used in that local area. Alternatively, one could select all census geographies at a given distance of a library and aggregate those areas to measure the population of the functional service area.

GIS is usually coupled with census data for doing these types of analyses, given its ability to collate data from multiple sources in the same geographic space, aggregate smaller geographies and data into larger units, and select or measure features based on geographical relationships or distance. The datasets produced by the Institute of Museum and Library Services (IMLS) as part of the annual Public Library Survey provide GIS and attribute data for library systems and outlets (individual library facilities) in the United States.

There are different methods for measuring library service areas, and the principal ones were outlined in an early study by Jue and colleagues (1999): container, fixed-radius or buffer, and a gravity-based or distance approach. For the container method, the census geography where the library falls serves as the de facto service area. This method is the least desirable, as the shape of the area may not meaningfully represent the true service area. The buffer approach is more widely used, where a circle is drawn around the library at a given distance and census areas within that circle are selected as part of the library's service area. An alternative to the buffer method is a Thiessen or Voronoi polygon approach, where a boundary is drawn between each library and its nearest neighbors to delineate nonoverlapping service areas for every library. With a distance-based approach, the distance from each census geography (an origin) is measured to the nearest library (a destination), and geographies are assigned to the library based on closest distance.

A number of factors must be considered in each of these analyses. For the buffer and Thiessen approaches, a method must be chosen for selecting the census areas that fall within the buffer or polygon, as the census areas may cross the boundaries of these features. One could assign all census areas that

partially fall inside a buffer to that library, or assign areas that have a certain percentage within the buffer, or assign areas whose geographic center is contained within the buffer (Schlossberg 2003). The population data of the areas could be apportioned based on the percentage of area that falls within the buffer. This approach was widely used prior to the introduction of the ACS but is more problematic today given the low precision of ACS estimates at the block group and tract levels and the difficulty with apportioning these estimates. The size of the census geographies is a key factor: while it is desirable to use the smallest areas possible to limit overlap between census features and the library buffer, ACS data becomes more unreliable the smaller the unit is, and block-level data is limited to the decennial census.

Another consideration for the buffer approach is how large the buffer should be. Most studies use a range of one to three miles based on past convention. This is based on how far people would be willing to walk or generally travel, but more simply it represents an area that would meaningfully constitute a neighborhood or community, which would typically be small. An alternative would be to use variable buffers, based on the size of the library (and its ability to attract people from larger distances) or whether the library is in an urban or rural community (which affects willingness to travel lesser or greater distances).

The distance-based approach eliminates this problem, as every census area would be associated with the nearest library based on distance. With distance-based approaches, the primary issue is how distance should be measured. Simple straight-line or Euclidean distance is the easiest to compute, but network distance that follows actual roadways and paths is more realistic. Again, the size of the census geography plays an important role, as the geography serves as an aggregated summary of every person's individual home. Distances would be measured from the central point or centroid of each geography to the closest library. A better alternative to geographic centroids are population centroids, which are summary points representing how population is distributed within an area. The Census Bureau publishes population centroids for each DEC for block groups, tracts, counties, and states, and this data can be plotted in GIS and associated with tabular census data.

Figure 7.1 illustrates a one-mile buffer approach with census tracts and an urban public library. Census tracts 214.01, 218, and 219.02 have been assigned to the central library's service area; as none of the tracts are perfectly contained within the one-mile buffer, they have been assigned based on whether the center of the tract falls within the buffer. The population data for these three tracts would then be summarized to describe this library's service area. The

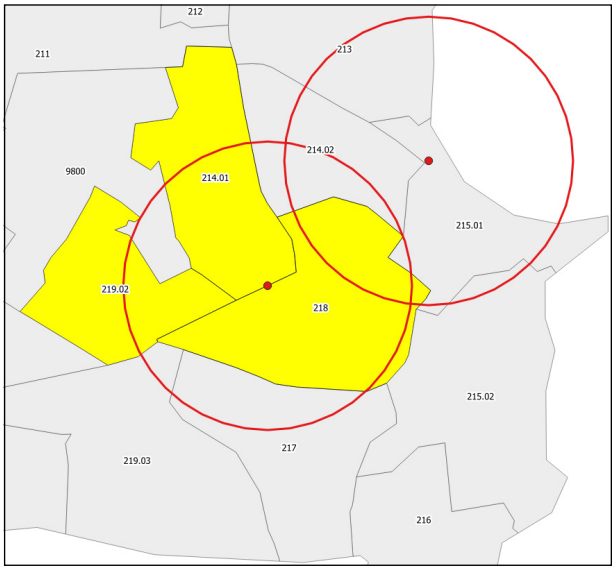


Figure 7.1
Using GIS to assign census tracts to libraries using one-mile buffers

second library to the northeast would be assigned the two tracts whose centers fall within its buffer. Smaller block groups could be used instead of tracts to minimize overlap, but if ACS data was being used, the precision of the estimates would be much lower. Alternatively, every tract could be assigned to a library based on distance to the closest branch.

Works by Palmer (1981) and Hayes and Palmer (1983) are among the earliest studies that reviewed the literature and measured library accessibility and service areas using a buffer-style approach in the pre-GIS era. Recent literature reviews by Bishop and Mandel (2010) and Mandel, Bishop, and Orehek (2020) summarize the applications of GIS to LIS research, which includes the use of GIS and census data for studying public library accessibility and distribution. Most of these studies focus on small areas such as counties or cities as case studies, while a few are national in scope. The earliest studies were practitioner-based and highlighted the utility of using GIS in conjunction with census data and library-generated data on circulation and patron registration to study library user communities, site new libraries, and understand library use and accessibility (Park 2012a; Bishop 2008; Hertel and Sprague 2007; Kinikin 2004; Ottensmann 1997). Other studies emphasized library policy in conjunction with social issues, such as the presence or absence of libraries in areas of high poverty (Jue et al. 1999), minority and low-income communities (Koontz, Jue, and Lance 2005), and the impact of library closures on different communities (Koontz, Jue, and Bishop 2009). Statistical analysis has been used to study the relationship between the locations of libraries and

the characteristics of users from both local (Japzon and Gong 2005) and national (Sin 2011) contexts. The application of buffer-based (Donnelly 2014) and distance-based (Donnelly 2015) approaches in conjunction with census data has been used to characterize the distribution of public libraries throughout the United States.

More recent work has focused on refining various aspects of the accessibility measurement process. The census counts people based on where they live, but many library users visit libraries based on where they work or don't make single trips to and from home and the library but combine them in a chain of trips (Park 2012b). Library users also use a variety of means to get to the library, including driving, public transit, biking, and walking, and this affects accessibility and distance that users are willing to travel (Allen 2019). It's important to consider these factors when studying library service areas and measuring accessibility and to remember that the approaches described here are all attempts to model reality. While a circular distance of one to three miles around a library may not perfectly measure the community or accessibility for all library users, it still serves as a general characterization of the area that likely constitutes the library's core population and generally measures the presence or absence of a library and whether a community will accrue its benefits or not.

The most recent studies demonstrate increasing internationalization, as researchers use census data in their countries and various GIS methods to study these issues in Wales in the UK (Page, Langford, and Higgs 2020; Higgs et al. 2018), Canada (Allen 2019), Hong Kong (Guo, Chan, and Yip 2017), South Korea (Lim and Park 2015), and Slovenia (Zakrajsek and Vodeb 2020), as well as in different cities in the United States (Cheng et al. 2021). These studies are also characterized by increasing complexity, where the focus has shifted to the refinement of models and methods and away from the practitioner and policy-driven applications of the earlier studies. There are exceptions, such as the studies in Wales, where the analyses are explicitly tied to the impact on society of cuts to library services and the need for informing public library policies. Recent practitioner-based works include studies on mapping rural library employment opportunities in Kansas (Lund 2017) and adult literacy outreach programs in Nashville (Rosichan 2020).

US census data is an invaluable public resource that supports policy and research about all facets of American society. For librarians, it's a resource that we rely on for meeting the diverse needs of our user communities, by connecting them to information they need to do their research, and for aiding our understanding of who our communities are so that we can better serve them.

References

Allen, Jeff. 2019. "Mapping Differences in Access to Public Libraries by Travel Mode and Time of Day." *Library and Information Science Research* 41: 11–18.

Bishop, Bradley Wade. 2008. "Use of Geographic Information Systems in Marketing and Facility Site Location." *Public Libraries* 47, no. 5: 65–69.

Bishop, Bradley Wade, and Lauren Mandel. 2010. "Utilizing Geographic Information Systems (GIS) in Library Research." *Library Hi Tech* 28, no. 4: 536–47.

Cheng, Wenting, Jiahui Wu, William Moen, and Lingzi Hong. 2021. "Assessing the Spatial Accessibility and Spatial Equity of Public Libraries' Physical Locations." *Library and Information Science Research* 43, no. 2: 101089. https://10.1016/j.lisr.2021.101089.

Donnelly, Francis P. 2014. "The Geographic Distribution of United States Public Libraries: An Analysis of Locations and Service Areas." *Journal of Librarianship and Information Science* 46, no. 2: 110–29.

———. 2015. "Regional Variations in Average Distance to Public Libraries in the United States." *Library and Information Science Research* 37: 280–89.

Guo, Yingqi, Chee Hon Chan, and Paul S. F. Yip. 2017. "Spatial Variation in Accessibility of Libraries in Hong Kong." *Library and Information Science Research* 39: 319–29.

Hayes, Robert M., and E. Susan Palmer. 1983. "The Effects of Distance upon Use of Libraries: Case Studies Based on a Survey of Users of the Los Angeles Public Library—Central Library and Branches." *Library and Information Science Research* 5: 67–100.

Hertel, Karen, and Nancy Sprague. 2007. "GIS and Census Data: Tools for Library Planning." *Library Hi Tech* 25, no. 2: 246–59.

Higgs, Gary, Samuel Jones, Mitchel Langford, and Jesse Heley. 2018. "Assessing the Impacts of Changing Public Service Provision on Geographical Accessibility: An Examination of Public Library Provision in Pembrokeshire, South Wales." *Environment and Planning C: Politics and Space* 36, no. 6: 548–68.

Japzon, Andrea C., and Hongmian Gong. 2005. "A Neighborhood Analysis of Public Library Use in New York City." *Library Quarterly* 75, no. 4: 446–63.

Jue, Dean K., Christie M. Koontz, J. Andrew Magpantay, Keith Curry Lance, and Ann M. Seidl. 1999. "Using Public Libraries to Provide Technology Access for Individuals in Poverty: A Nationwide Analysis of Library Market Areas Using a Geographic Information System." *Library and Information Science Research* 21, no. 3: 299–325.

Kinikin, JaNae. 2004. "Applying Geographic Information Systems to the Weber County Library System." *Information Technology and Libraries* 23, no. 3: 102–7.

Koontz, Christie M., Dean K. Jue, and Bradley Wade Bishop. 2009. "Public Library Facility Closure: An

Investigation of Reasons for Closure and Effects on Geographic Market Areas." *Library and Information Science Research* 31: 84–91.

Koontz, Christie M., Dean K. Jue, and Keith Curry Lance. 2005. "Neighborhood-Based In-Library Use Performance Measures for Public Libraries: A Nationwide Study of Majority-Minority and Majority White/Low Income Markets Using Personal Digital Data Collectors." *Library and Information Science Research* 27: 28–50.

Lim, Ho-kyun, and Sung Jae Park. 2015. "Designing a GIS-Based Planning and Support System for a Public Library Building Project." *Journal of Librarianship and Information Science* 47, no. 3: 254–64.

Lund, Brady. 2017. "Rural Library Opportunity Zones: Mapping Rural Library Employment Opportunities Using Quantum GIS." *Endnotes: The Journal of the New Members Round Table* 8, no. 1: 1–12.

Mandel, Lauren Heather, Bradley Wade Bishop, and Ashley Marie Orehek. 2020. "A New Decade of Uses for Geographic Information Systems (GIS) as a Tool to Research, Measure, and Analyze Library Services." *Library Hi Tech*, in press.

Ottensmann, John R. 1997. "Using Geographic Information Systems to Analyze Library Utilization." *Library Quarterly* 67, no. 1: 24–49.

Page, Nicholas, Mitchel Langford, and Gary Higgs. 2020. "Exploring Spatiotemporal Variations in Public Library Provision Following a Prolonged Period of Economic Austerity: A GIS Approach."*Area* 52, no. 2: 342–53. https://doi.org/10.1111/area.12575.

Palmer, E. Susan. 1981. "The Effect of Distance on Public Library Use: A Literature Survey." *Library Research* 3: 315–54.

Park, Sung Jae. 2012a. "Measuring Public Library Accessibility: A Case Study Using GIS." *Library and Information Science Research* 34: 13–21.

———. 2012b. "Measuring Travel Time and Distance in Library Use." *Library Hi Tech* 30, no. 1: 151–69.

Rosichan, Sharra. 2020. "Application of GIS in Mapping of Adult Literacy Outreach Programs and Potential Areas of Impact and Need." *Public Library Quarterly* 39, no. 4: 346–62.

Schlossberg, Marc. 2003. "GIS, the US Census and Neighbourhood Scale Analysis." *Planning, Practice, and Research* 18, no. 2–3: 213–17.

Sin, Se-Ching Joanna. 2011. "Neighborhood Disparities in Access to Information Resources: Measuring and Mapping U.S. Public Libraries' Funding and Service Landscapes." *Library and Information Science Research* 33: 41–53.

Zakrajsek, Franc J., and Vlasta Vodeb. 2020. "Agent-Based Geographical Modeling of Public Library Locations." *Library and Information Science Research* 42: 101013. https://doi.org/10.1016/j.lisr.2020.101013.

Notes

Notes

Library Technology REPORTS

Upcoming Issues	
July 58:5	**Designing Information Literacy Tutorials: Tips, Techniques, and Trends** by Yvonne Mery
Aug/Sept 58:6	**Improving Access and Delivery of Academic Content from Libraries** by Aaron Tay
October 58:7	**Mitigating Electronic Resource Access Issues** by Ashley Zmau and Holly Talbott

Purchase single copies in the ALA Store
alastore.ala.org

ALA TechSource
alatechsource.org

ALA TechSource, a unit of the publishing department of the American Library Association

A Note to Our Subscribers

As a subscription-based publication, ***Library Technology Reports*** will sunset with the December 2022 issue. After that time, it will be available for single-issue sales only. Subsequently, no new ***Library Technology Reports*** subscriptions or renewals will be accepted. If you have questions about your subscription, please contact us at subscriptions@ala.org or call toll free at (800) 545-2433, ext. 4299.