

Understanding and Using American Community Survey Data

What Researchers Need to Know

Issued March 2020



Acknowledgments

Linda A. Jacobsen, Vice President, U.S. Programs, Population Reference Bureau (PRB), and **Mark Mather**, Associate Vice President, U.S. Programs, PRB, drafted this handbook in partnership with the U.S. Census Bureau's American Community Survey Office. Other PRB staff who assisted in drafting and reviewing the handbook include **Beth Jarosz**, **Lillian Kilduff**, and **Paola Scommegna**. Some of the material in this handbook was adapted from the Census Bureau's 2009 publication, *A Compass for Understanding and Using American Community Survey Data: What Researchers Need to Know*, drafted by **Warren A. Brown**.

American Community Survey data users who provided feedback and case studies for this handbook include **Jean D'Amico**, **Brett Fried**, and **Sarah Kaufman**.

Nicole Scanniello, **Gretchen Gooding**, and **Charles Gamble**, Census Bureau, contributed to the planning and review of this handbook series.

The American Community Survey program is under the direction of **Albert E. Fontenot, Jr.**, Associate Director for Decennial Census Programs, **James B. Treat**, Assistant Director for Decennial Census Programs, and **Donna M. Daily**, Chief, American Community Survey Office.

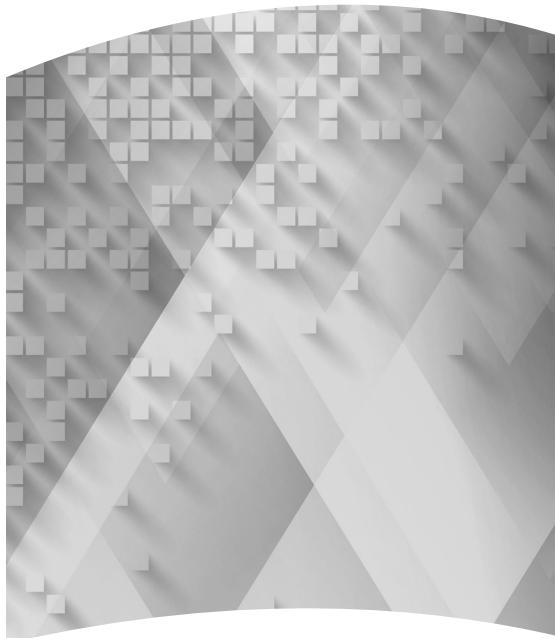
Other individuals from the Census Bureau who contributed to the review and release of these handbooks include **Grace Clemons**, **Barbara Downs**, **Justin Keller**, **Amanda Klimek**, **R. Chase Sawyer**, **Michael Starsinic**, and **Tyson Weister**.

Linda Chen, **Faye E. Brock**, and **Christine E. Geter** provided publication management, graphics design and composition, and editorial review for print and electronic media under the direction of **Janet Sweeney**, Chief of the Graphic and Editorial Services Branch, Public Information Office.

Understanding and Using American Community Survey Data

What Researchers Need to Know

Issued March 2020



U.S. Department of Commerce
Wilbur Ross,
Secretary

Karen Dunn Kelley,
Deputy Secretary

U.S. CENSUS BUREAU
Steven Dillingham,
Director

Suggested Citation

U.S. Census Bureau,
*Understanding and Using
American Community Survey Data:
What Researchers Need to Know*,
U.S. Government Printing Office,
Washington, DC, 2020.



U.S. CENSUS BUREAU

Steven Dillingham,

Director

Ron Jarmin,

Deputy Director and Chief Operating Officer

Albert E. Fontenot Jr.,

Associate Director for Decennial Census Programs

James B. Treat,

Assistant Director for Decennial Census Programs

Donna M. Daily,

Chief, American Community Survey Office

Contents

1. Topics Covered in the ACS	1
2. How Researchers Use ACS Data	3
3. Sampling Error in the ACS	12
4. Other Considerations in Working With ACS Data	15
5. Case Studies Using ACS Data.....	18
6. Additional Resources	32

This page is intentionally blank.

UNDERSTANDING AND USING AMERICAN COMMUNITY SURVEY DATA: WHAT RESEARCHERS NEED TO KNOW

The American Community Survey (ACS) is the nation's premier source of detailed social, economic, housing, and demographic characteristics for local communities.

This handbook describes how researchers can use ACS data to make comparisons, create custom tables, and combine ACS data with other data sources. It is aimed at researchers who are familiar with using data—summary tabulations and microdata records—from complex sample surveys.

What Is the ACS?

The ACS is a nationwide survey designed to provide communities with reliable and timely social, economic, housing, and demographic data every year. A separate annual survey, called the Puerto Rico Community Survey (PRCS), collects similar data about the population and housing units in Puerto Rico. The U.S. Census Bureau uses data collected in the ACS and the PRCS to provide estimates on a broad range of population, housing unit, and household characteristics for states, counties, cities, school districts, congressional districts, census tracts, block groups, and many other geographic areas.

The ACS has an annual sample size of about 3.5 million addresses, with survey information collected nearly every day of the year. Data are pooled across a calendar year to produce estimates for that year. As a result, ACS estimates reflect data that have been collected

over a period of time rather than for a single point in time as in the decennial census, which is conducted every 10 years and provides population counts as of April 1 of the census year.

ACS 1-year estimates are data that have been collected over a 12-month period and are available for geographic areas with at least 65,000 people. Starting with the 2014 ACS, the Census Bureau is also producing “1-year Supplemental Estimates”—simplified versions of popular ACS tables—for geographic areas with at least 20,000 people. The Census Bureau combines 5 consecutive years of ACS data to produce multiyear estimates for geographic areas with fewer than 65,000 residents. These 5-year estimates represent data collected over a period of 60 months.¹

For more detailed information about the ACS—how to judge the accuracy of ACS estimates, understanding multiyear estimates, knowing which geographic areas are covered in the ACS, and how to access ACS data on the Census Bureau’s Web site—see the Census Bureau’s handbook on *Understanding and Using American Community Survey Data: What All Data Users Need to Know*.²

¹ The Census Bureau previously released 3-year estimates based on 36 months of data collection. In 2015, the 3-year products were discontinued. The 2011–2013 ACS 3-year estimates, released in 2014, are the last release of this product.

² U.S. Census Bureau, *Understanding and Using American Community Survey Data: What All Data Users Need to Know*, <www.census.gov/programs-surveys/acs/guidance/handbooks/general.html>.

1. TOPICS COVERED IN THE ACS

The primary purpose of the American Community Survey (ACS) is to help Congress determine funding and policies for a wide variety of federal programs. Because of this, the topics covered by the ACS are diverse (see Table 1.1).

- Examples of social characteristics include disability, educational attainment, language spoken at home, and veteran status.

- Economic characteristics include employment status, health insurance, income, and earnings.
- Examples of housing characteristics include computer and Internet use, selected monthly owner costs, rent, and the year the structure was built.
- Demographic characteristics include age, sex, race, Hispanic origin, and relationship to householder.

Table 1.1. Population and Housing Data Included in the American Community Survey Data Products

Social Characteristics	Economic Characteristics	
Ancestry	Class of Worker	Plumbing Facilities ⁶
Citizenship Status	Commuting (Journey to Work)	Rent
Disability Status ¹	Employment Status	Rooms/Bedrooms
Educational Attainment	Food Stamps/Supplemental Nutrition Assistance Program (SNAP) ⁴	Selected Monthly Owner Costs
Fertility	Health Insurance Coverage ²	Telephone Service Available
Grandparents as Caregivers	Income and Earnings	Tenure (Owner/Renter)
Language Spoken at Home	Industry and Occupation	Units in Structure
Marital History ²	Place of Work	Value of Home
Marital Status	Poverty Status	Vehicles Available
Migration/Residence 1 Year Ago	Work Status Last Year	Year Householder Moved Into Unit
Period of Military Service		Year Structure Built
Place of Birth		
School Enrollment		
Undergraduate Field of Degree ³		
Veteran Status ²		
Year of Entry		
Housing Characteristics		Demographics Characteristics
	Computer and Internet Use ⁵	Age and Sex
	House Heating Fuel	Group Quarters Population
	Kitchen Facilities	Hispanic or Latino Origin
	Occupancy/Vacancy Status	Race
	Occupants Per Room	Relationship to Householder
		Total Population

¹ Questions on Disability Status were significantly revised in the 2008 survey to cause a break in series.

² Marital History, Veterans' Service-Connected Disability Status and Ratings, and Health Insurance Coverage were added in the 2008 survey.

³ Undergraduate Field of Degree was added in the 2009 survey.

⁴ Food Stamp Benefit amount was removed in 2008.

⁵ Computer and Internet Use was added to the 2013 survey.

⁶ One of the components of Plumbing Facilities, flush toilet, and Business or Medical Office on Property questions were removed in 2016.

Source: U.S. Census Bureau.

TIP: The ACS was designed to provide estimates of the **characteristics** of the population, not to provide counts of the population in different geographic areas or population subgroups. For basic counts of the U.S. population by age, sex, race, and Hispanic origin, visit the Census Bureau's Population and Housing Unit Estimates Web page.³

A good way to learn about all of the topics covered in the ACS is to explore the information available through the U.S. Census Bureau's data dissemination platform on data.census.gov.⁴ The Data Profiles in data.census.gov, which include the most frequently requested social, economic, housing, and demographic data, are

³ U.S. Census Bureau, Population and Housing Unit Estimates, <www.census.gov/popest/>.

⁴ U.S. Census Bureau, data.census.gov, <<https://data.census.gov>>.

useful for novice users who want to explore the range of topics that are available.⁵ Copies of ACS questionnaires for different years are also available on the Census Bureau's Web site.⁶

For more detailed information about the topics in the ACS, see the section on "Understanding the ACS: The Basics" in the Census Bureau's handbook on *Understanding and Using American Community Survey Data: What All Data Users Need to Know*.⁷

⁵ U.S. Census Bureau, data.census.gov, Data Profiles, <<https://data.census.gov/cedsci/all?q=dp>>.

⁶ U.S. Census Bureau, American Community Survey (ACS), Questionnaire Archive, <www.census.gov/programs-surveys/acs/methodology/questionnaire-archive.html>.

⁷ U.S. Census Bureau, *Understanding and Using American Community Survey Data: What All Data Users Need to Know*, <www.census.gov/programs-surveys/acs/guidance/handbooks/general.html>.

2. HOW RESEARCHERS USE ACS DATA

There are two main types of American Community Survey (ACS) data available for analysis: aggregate data and microdata. In aggregate (or summary) data, individual records are weighted and tabulated to create estimates for a range of geographic areas. In contrast, ACS microdata files include individual survey response records, with identifying information removed to protect the respondent's confidentiality. The type of ACS data researchers use depends on the specific variable categories and levels of geography needed for their analyses.

Using Aggregate ACS Data

Aggregate ACS data provide a good starting point for data users because they are relatively easy to access and are available for a broad range of geographic areas (for example, states, metropolitan statistical areas, cities, or counties). Many published ACS data tables are also disaggregated by age, sex, race/ethnicity, and other characteristics, enabling comparisons across different population subgroups. Every published table includes not only ACS estimates but also their associated margins of error (also known as levels of uncertainty). Researchers can use aggregate ACS data for a broad range of applications, including:

- Analyzing the relationship between economic status and health insurance coverage at the county level.
- Comparing patterns of marital status and family structure across different racial/ethnic groups.
- Investigating state-to-state migration flows and how they change over time.
- Analyzing economic data across neighborhoods to identify areas of concentrated poverty.

Researchers can also use aggregate ACS data to access estimates for small geographic areas such as census tracts—small subdivisions of counties that typically have between 2,500 and 8,000 residents. Census tracts are designed to follow the boundaries of neighborhoods; they encompass areas that are

homogeneous with respect to population characteristics, economic status, and living conditions. There are also more than 300 ACS data tables available for block groups—subdivisions of census tracts—that include between 600 and 3,000 people each. In the ACS, block groups are the smallest level of geography published. However, data users need to pay attention to sampling error associated with ACS estimates—especially when working with data for small geographic areas or population subgroups. See the section on “Sampling Error in the ACS” for more information.

For a list of published ACS data tables, users can download table shells that include information about table universes, category line numbers, and table IDs.⁸ ACS table shells are typically available 1 week before the data are released, allowing users to preview new table layouts in advance.

The U.S. Census Bureau provides access to published ACS tables through two main sources: data.census.gov and the ACS Summary File.⁹

Data.census.gov

Data.census.gov is the Census Bureau’s primary tool for accessing population, housing, and economic data from the ACS, the Puerto Rico Community Survey, the decennial census, and many other Census Bureau data sets.

Data.census.gov provides access to ACS data for a wide range of geographic areas including states, cities, counties, census tracts, and block groups.¹⁰ Researchers can access detailed ACS tables by using the “Advanced Search” feature, which allows users to conduct keyword searches or search by using pre-defined topics, geographies, years, surveys, or industry codes (see Figures 2.1 and 2.2).

⁸ U.S. Census Bureau, American Community Survey (ACS), Table Shells and Table List, <www.census.gov/programs-surveys/acs/technical-documentation/table-shells.html>.

⁹ U.S. Census Bureau, American Community Survey (ACS), Summary File Data, <www.census.gov/programs-surveys/acs/data/summary-file.html>.

¹⁰ U.S. Census Bureau, <<https://data.census.gov>>.

Figure 2.1. Advanced Search in Data.census.gov

The screenshot shows the homepage of Data.census.gov. At the top, the United States Census Bureau logo is visible. Below it, a large blue header reads "Explore Census Data". A sub-header states, "The Census Bureau is the leading source of quality data about the nation's people and economy." To the right of the text is a graphic featuring a computer monitor displaying a map and a laptop showing a chart. Below the main header is a search bar with the placeholder text "I'm looking for ...". A red box highlights the "Advanced Search" button below the search bar. At the bottom of the page, a grey footer bar contains the text "Source: U.S. Census Bureau, <<https://data.census.gov>>".

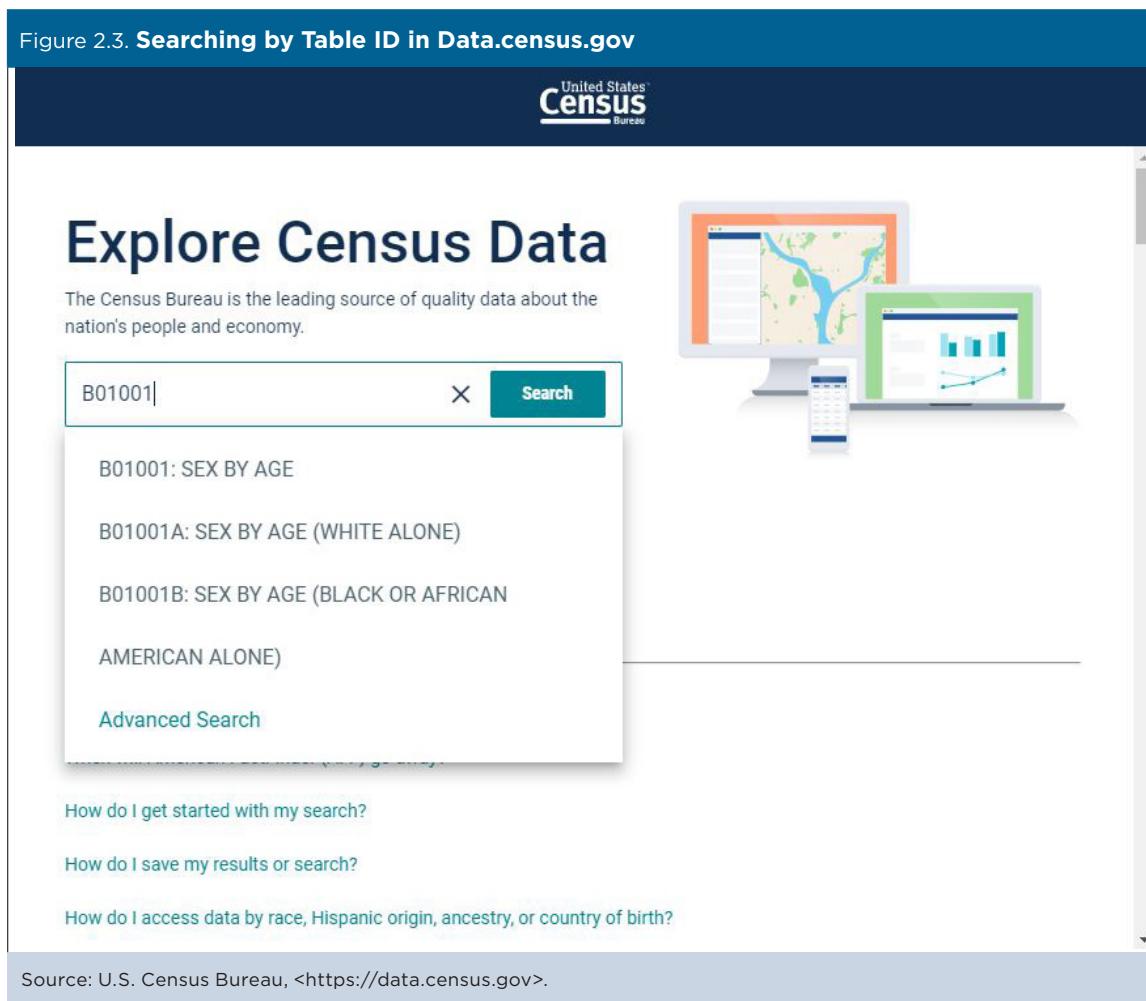
Figure 2.2. Advanced Search Filters in Data.census.gov

The screenshot shows the "Advanced Search" page. At the top, the United States Census Bureau logo is on the left, and a search bar with the placeholder "e.g. age, sex, population..." is on the right. Below the search bar, the title "Advanced Search" is displayed. A sub-section titled "Narrow search with filters" follows. Under this, a "FIND A FILTER" section includes a search input field with the placeholder "e.g. 336111 - Automobile Manufacturing" and a magnifying glass icon. To the right of this is a vertical sidebar with a red border containing a list of filter categories: "BROWSE FILTERS", "Topics", "Geography", "Years", "Surveys", and "Codes". At the bottom of the page, a grey footer bar contains the text "Source: U.S. Census Bureau, <<https://data.census.gov>>".

Researchers looking for a particular table can also use the search bar on the data.census.gov home page to search by Table ID. For example, typing “B01001” into

the search bar generates a list of relevant Sex by Age tables (see Figure 2.3). Click the “Search” button to view a list of relevant tables.

Figure 2.3. **Searching by Table ID in Data.census.gov**



Data users can also use data.census.gov to download multiple tables simultaneously (see Figure 2.4). These tables can be downloaded in either comma-separated values (CSV) or PDF format. After navigating to a list of relevant tables:

- Click “Download” on the left side of the screen.
- Use the checkboxes to select the table(s) you would like to download.
- Click “Download Selected” on the left side of the screen.

- Choose year(s) and type of estimates (1-Year or 5-Year).
- Choose File Type (CSV or PDF).
- Click “Download” at the bottom of the screen.

For more information about data.census.gov, view the Census Bureau’s release notes and answers to frequently asked questions about the site.¹¹

¹¹ U.S. Census Bureau, Data.census.gov: Census Bureau’s New Data Dissemination Platform Frequently Asked Questions and Release Notes, <<https://data.census.gov/assets/releasenotes/faqs-release-notes.pdf>>.

Figure 2.4. Downloading Multiple Tables Simultaneously in Data.census.gov

The screenshot shows the Data.census.gov website. In the sidebar, under 'ALL TABLES', several table thumbnails are listed, with 'SEX BY AGE' being the first one selected. A red box highlights the 'Download Selected (2)' button. In the main content area, a table titled 'SEX BY AGE' is displayed for the 'American Community Survey' from 2011 to 2017. A red box highlights the 'Download / Print / Share' section. Below it, a 'Select Table Vintages' grid allows users to choose specific years for each table. A red box highlights this grid. At the bottom right, a large red box highlights the 'DOWNLOAD' button.

Source: U.S. Census Bureau, <<https://data.census.gov>>.

6 Understanding and Using American Community Survey Data

ACS Summary File

Researchers with programming skills and access to statistical software can use the ACS Summary File to download and analyze ACS data.¹² The Summary File provides access to aggregate ACS data and includes information for geographic areas down to the block group level. It is useful for skilled programmers who want to access multiple ACS tables for large numbers of geographic areas. The ACS Summary File is designed for more advanced data users, so the Census Bureau recommends that users check to see if their tables of interest are easily available for download through data.census.gov before using this data product.

The ACS Summary File is a comma-delimited text file that contains all of the Detailed Tables for the ACS. The file is stored with only the data from the tables and without information such as the table title, description of the rows, or geographic identifiers. That information is located in other files that the user must merge with the data files to reproduce full tables. Users can merge these files through statistical packages such as R, Python, SAS, SPSS, or STATA.

The Summary File documentation provides users with all the information they need to access and process these data, including survey methods and links to sample SAS programs for processing the data files.¹³ The ACS Summary File can be downloaded as zipped files from the Census Bureau's FTP site.¹⁴ Developers can also access the Summary File through the Census

¹² U.S. Census Bureau, American Community Survey (ACS), Summary File Data, <www.census.gov/programs-surveys/acs/data/summary-file.html>.

¹³ U.S. Census Bureau, American Community Survey (ACS), Summary File Documentation, <www.census.gov/programs-surveys/acs/technical-documentation/summary-file-documentation.html>.

¹⁴ U.S. Census Bureau, American Community Survey (ACS), Data via FTP, <www.census.gov/programs-surveys/acs/data/data-via-ftp.html>.

Bureau's APIs.¹⁵ Separate ACS Summary Files are available for each 1-year and 5-year data release.

Using ACS Microdata

ACS aggregate data are available for a large number of topics, geographic areas, and population groups, but not every data need can be met through published tables. In these cases, researchers can use ACS microdata files to create custom estimates.

ACS microdata are individual records that include information about people and housing units in the survey with identifying information removed to protect each respondent's confidentiality. Microdata provide the flexibility to create custom tabulations or to investigate the relationship among characteristics captured by the survey questionnaire.

ACS microdata provide nearly unlimited possibilities for analysis, including:

- Estimating the population living below a specified income-to-poverty ratio (for example, family income below 185 percent of the poverty threshold).
- Studying the relationship between veteran status and income.
- Comparing poverty and unemployment estimates for women and men working in different occupational categories.
- Tracking trends in state-to-state migration among baby boomers since the Great Recession.

Most researchers access ACS microdata through the Census Bureau's Public Use Microdata Sample (PUMS) files. However, data can also be accessed through the Federal Statistical Research Data Centers (FSRDCs). Both sources of ACS microdata are described below.

¹⁵ U.S. Census Bureau, Developers, Available APIs, <www.census.gov/data/developers/data-sets.html>.

ACS Public Use Microdata Sample Files

Accessible through the Census Bureau's Web site, the ACS PUMS data allow data users to create their own tables with variables of their choosing.¹⁶

In general, the PUMS files are more difficult to work with than the premade tables on data.census.gov because data users need to use a statistical package to access the data. In addition, the responsibility for producing estimates from PUMS and judging their statistical reliability is up to the user. However, once a data user learns how to work with PUMS, the research possibilities are endless.

TIP: ACS PUMS data are not designed for statistical analysis of small geographic areas. The Census Bureau restricts the availability of information in microdata files that could be used to identify a specific housing unit or person, including detailed geographic information. Thus, the smallest geographic area available is the Public Use Microdata Area (PUMA), which has a minimum population of 100,000.

¹⁶ U.S. Census Bureau, American Community Survey (ACS), PUMS Data, <www.census.gov/programs-surveys/acs/data/pums.html>.

PUMAs are constructed based on county and neighborhood boundaries and do not cross state lines. Typically, counties with large populations are subdivided into multiple PUMAs, while PUMAs in more rural areas are made up of groups of adjacent counties. PUMAs are especially useful for rural areas because, unlike counties, they meet the 65,000-population threshold that is needed to provide ACS 1-year estimates. The value of using PUMA geography becomes apparent when looking at a state such as Kentucky (see Figures 2.5 and 2.6). The 2017 ACS 1-year estimates include data for only 13 of Kentucky's 120 counties, but they also include data for all 34 Kentucky PUMAs covering the entire state.

The ACS PUMS files include separate records for housing units and population. The housing unit records have unique identifiers that are repeated on each of the population records for people living in that housing unit. In this manner, housing unit characteristics can be merged with population records as needed for an analysis. For example, housing unit records contain variables on tenure (owner/renter status), so to analyze data on the demographic characteristics of homeowners, it is necessary to link the housing unit and population records.

Figure 2.5. Availability of ACS 1-Year Estimates for Kentucky: 2017

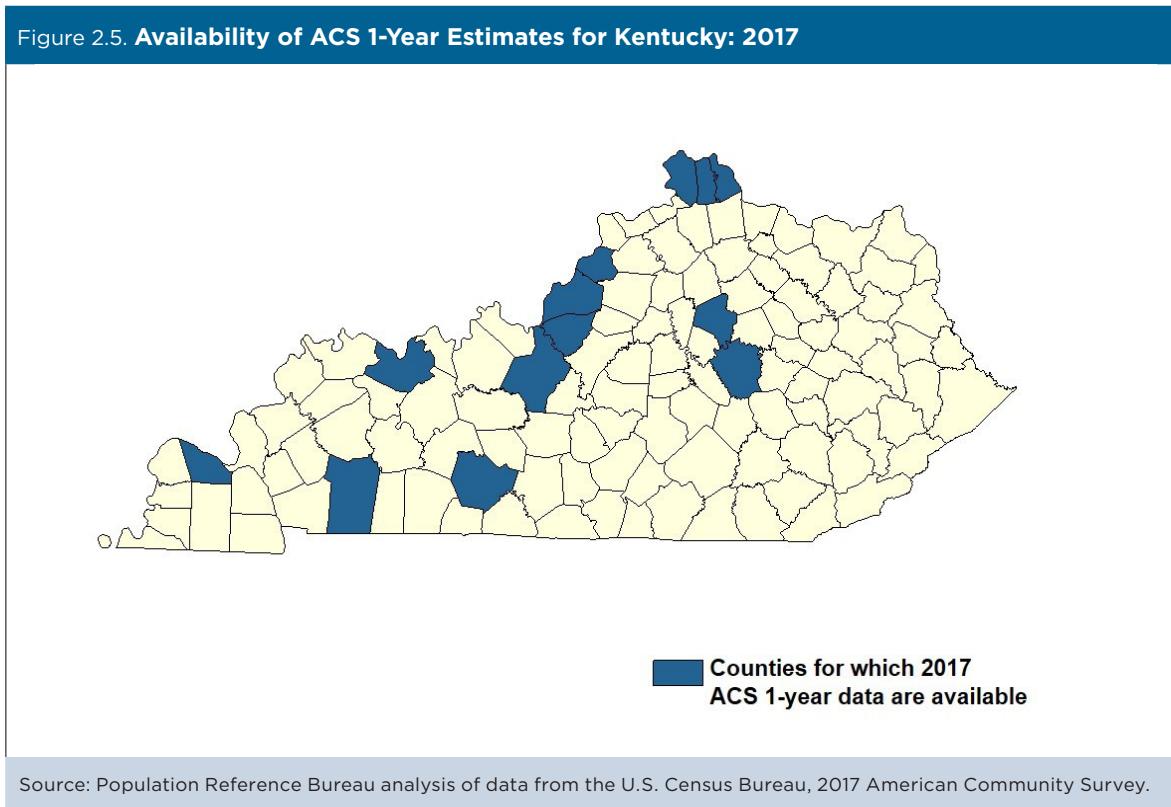
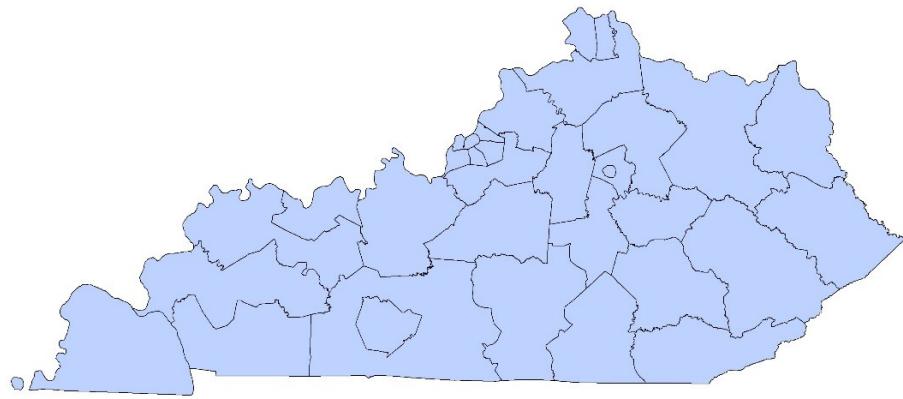


Figure 2.6. Public Use Microdata Areas in Kentucky



Source: U.S. Census Bureau, Cartographic Boundary Shapefiles—Public Use Microdata Areas (PUMAs) (2017 boundaries), <www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html>.

Each housing and person record is assigned a weight because the records in the PUMS files represent a sample of the population. The weight is a numeric variable expressing the number of housing units or people that an individual microdata record represents. The sum of the housing unit and person weights for a geographic area is equal to the estimate of the total number of housing units and people in that area. Since the ACS is not a simple random sample survey but rather a complex sample survey, the values of the weights vary. To generate estimates of the population based on the sample records, it is necessary to use the weights assigned to each of the records correctly. The Census Bureau provides basic tabulations

of weighted characteristics from the ACS PUMS that researchers can employ to verify the accuracy of their programming.

ACS microdata for recent years can be accessed through the PUMS Data Web page (see Figure 2.7).¹⁷ Separate files are available for ACS 1-year and 5-year estimates, from 2005 to the most recent data release. ACS data are also available for earlier years (2000–2004) through the Census Bureau's FTP site. Files are available in both CSV and SAS data set formats.

¹⁷ U.S. Census Bureau, American Community Survey (ACS), PUMS Data, <www.census.gov/programs-surveys/acs/data/pums.html>.

Federal Statistical Research Data Centers

The FSRDCs are partnerships between federal statistical agencies and leading research institutions.¹⁸ FSRDCs are secure facilities managed by the Census Bureau that provide secure access to a range of restricted-use microdata including ACS microdata. Unlike the ACS PUMS, which includes a representative subset of records from the ACS sample, the restricted data files contain all ACS records.

¹⁸ U.S. Census Bureau, Federal Statistical Research Data Centers, <www.census.gov/fsrdc>.

FSRDC researchers have access to computing capacity to handle large data sets and complex calculations. Standard statistical, econometric, and programming software, including R, Stata, SAS, MATLAB, and Gauss, are available in a Linux environment. FSRDC researchers can collaborate with other research data center researchers across the United States through the secure FSRDC computing environment.

Data access via an FSRDC requires a proposal and approval process, including background checks of researchers. The approval process, while straightforward, can take several months.

Figure 2.7. ACS PUMS Data Web Page

PUMS Documentation page.' The main content area contains two columns of data links. The left column includes 'Available on the FTP site' (PUMS files back to 2005 in CSV and SAS formats), '2018 ACS PUMS', '2017 ACS PUMS', '2016 ACS PUMS', and '2015 ACS PUMS'. The right column includes 'PUMS on data.census.gov' (Microdata access from 2005-current in beta), '2014-2018 ACS 5-year PUMS', '2018 ACS 1-year PUMS', '2013-2017 ACS 5-year PUMS', and '2017 ACS 1-year PUMS'. On the far right, there is a 'Related Information' section with links to 'Share Your ACS Data Story', 'Contact Us', 'Introduction to the Public Use Microdata Sample (PUMS) File', and a 'You May Be Interested In' section with links to 'RELATED TOPICS', 'Contact Us', 'AROUND THE BUREAU', 'Email Updates', 'MOST POPULAR', and 'Share Your ACS Data Story'. At the bottom left, a call-to-action button says 'Respond to the ACS Learn how'."/>

Source: U.S. Census Bureau, American Community Survey (ACS), PUMS Data, <www.census.gov/programs-surveys/acs/data/pums.html>.

The Census Bureau's FSRDC Program Management Office considers proposals from qualified researchers in social science disciplines consistent with the subject matter of the surveys and censuses collected by the Census Bureau.¹⁹ Proposals can be submitted at any time and must:

- Provide benefit to Census Bureau programs.
- Demonstrate scientific merit.
- Require nonpublic data.
- Be feasible given the data.
- Pose no risk of disclosure.

All FSRDC researchers must obtain Census Bureau Special Sworn Status—passing a moderate risk background check and swearing to protect respondent confidentiality for life, facing significant financial and legal penalties under Title 13 and Title 26 of the United States Code for failure to do so.²⁰

When researchers need to remove aggregated output, tables, or model coefficients from the secure environment, the output must be reviewed to ensure the confidentiality of survey respondents and that the output is consistent with the original proposal. Once the results pass disclosure review, the approved files are provided to the researcher or team outside of the secure computing environment, usually via e-mail. The researcher(s) can then produce reports, presentations, and other products outside of the secure environment.

Information about how to apply for FSRDC access is available on the Census Bureau's Web site.²¹

Blending ACS Data With Data From Other Sources

Researchers are increasingly blending ACS estimates with data from other sources to answer questions that the ACS alone cannot answer. There are two main methods analysts can use to combine data from different sources. The first method involves combining aggregate data based on a geographic identifier that is available in both data sets such as a county Federal Information Processing Standards (FIPS) code. For example, county-level social and economic estimates

¹⁹ U.S. Census Bureau, Center for Economic Studies (CES), <www.census.gov/programs-surveys/ces.html>.

²⁰ U.S. Census Bureau, History, Privacy & Confidentiality, <www.census.gov/history/www/reference/privacy_confidentiality/>.

²¹ U.S. Census Bureau, Center for Economic Studies (CES), Apply for Access, <www.census.gov/programs-surveys/ces/data/restricted-use-data/apply-for-access.html>.

from the ACS could be combined with county-level death rates to investigate relationships between county characteristics and mortality. State, county, and census tract variables available through the ACS are likely to be defined the same way in other data sets, enabling researchers to produce merged data files with expanded lists of variables for analysis.

The second method—available to Census Bureau staff and researchers with approved FSRDC projects— involves linking individual or housing unit records from the ACS with administrative records based on personal identifiers. For example, Census Bureau staff linked children in the ACS with records from the Internal Revenue Service, Department of Housing and Urban Development, Centers for Medicare and Medicaid Services, Department of Health and Human Services, and other sources to investigate the undercount of young children in the decennial census.²² ACS records were linked to administrative data using protected identification keys (PIKs)—anonymous identifiers that can be used to link records across different data sets.

The Census Bureau conducts a variety of research projects that combine administrative records and survey data to lower costs, increase efficiency, reduce respondent burden, and improve data quality. Some of these projects generate new social and economic statistics—such as the Small Area Income and Poverty Estimates Program.²³ Other projects investigate ways to use linked data to better measure family relationships, evaluate program participation, and improve coverage of hard-to-reach populations.²⁴

Researchers outside of the Census Bureau who are interested in working with linked ACS records can apply to do so through the FSRDCs. All FSRDC users must obtain Special Sworn Status and adhere to relevant ethics, confidentiality, and privacy protection procedures.

More information is available through the FSRDC Web site.²⁵

²² Leticia Fernandez, Rachel Shattuck, and James Noon, "The Use of Administrative Records and the American Community Survey to Study the Characteristics of Undercounted Young Children in the 2010 Census," CARRA Working Paper Series 2018 (no. 5), 2018.

²³ U.S. Census Bureau, Small Area Income and Poverty Estimates (SAIPE) Program, <www.census.gov/programs-surveys/saipe.html>.

²⁴ Amy O'Hara, Rachel M. Shattuck, and Robert M. Goerge, "Linking Federal Surveys with Administrative Data to Improve Research on Families," *The ANNALS of the American Academy of Political and Social Science*, 669 (no. 1): 63-74, 2016.

²⁵ U.S. Census Bureau, Federal Statistical Research Data Centers, <www.census.gov/fsrdc>.

3. SAMPLING ERROR IN THE ACS

Because the American Community Survey (ACS) is based on a sample, rather than all housing units and people, ACS estimates have a degree of uncertainty associated with them, known as sampling error. In general, the larger the sample, the smaller the level of sampling error. To help users understand the impact of sampling error on data reliability, the U.S. Census Bureau provides a “margin of error” (MOE) for each published ACS estimate. The MOE, combined with the ACS estimate, give users a range of values within which the actual, “real-world” value is likely to fall.

TIP: Sometimes ACS data users ignore the issue of sampling variability, which can be problematic when analyzing differences across small area estimates. Rather than considering that the estimates are derived from a complex sample survey, the estimates have sometimes been treated as values for the population. Data users should be careful in drawing conclusions about small differences between two ACS estimates because they may not be statistically different.

By presenting the MOE alongside the estimates, users can more easily determine whether differences they observe over time and space are statistically significant or within the bounds of random variation. The Census Bureau uses a 90 percent confidence level to determine the MOE in the published tabulations. Depending on the application, a user may wish to

increase the confidence level to 95 percent or 99 percent to conduct a more rigorous test of significant differences.

Tests of Statistical Significance for Aggregate ACS Estimates

The Census Bureau has produced a Statistical Testing Tool to make it easier for ACS data users to conduct tests of statistical significance when comparing ACS estimates (see Figure 3.1).²⁶

The Statistical Testing Tool consists of an Excel spreadsheet that will automatically calculate statistical significance when data users are comparing two or more ACS estimates. Data users simply need to insert the ACS estimate(s) and associated MOE(s) into the correct columns and cells in the spreadsheet. The results are calculated automatically. The result “Yes” indicates that estimates are statistically different and the result “No” indicates the estimates are not statistically different.²⁷

²⁶ U.S. Census Bureau, American Community Survey (ACS), Statistical Testing Tool, <www.census.gov/programs-surveys/acs/guidance/statistical-testing-tool.html>.

²⁷ This tool only conducts statistical testing on the estimates keyed in by the data user for comparison within the spreadsheet and it does not adjust the MOE when making multiple comparisons, nor incorporate a Bonferroni correction or any other method in the results of the statistical testing.

Figure 3.1. Statistical Testing Tool

The screenshot shows a Microsoft Excel spreadsheet titled "Statistical Testing Tool". The title bar includes the "Statistical Testing for Two Estimates" logo and the "United States Census Bureau". The spreadsheet has several sections:

- Purpose:** A text box explaining that the spreadsheet determines whether there is statistical evidence to conclude that two estimates are different from each other.
- Results:** A table with three rows: "Yes" (Estimates are statistically different), "No" (Estimates are NOT statistically different (or are statistically tied)), and "N/A" (Statistical testing is not applicable for one or both of the estimates).
- Navigation:** Buttons for "Overview", "Instructions", "Statistical Testing for Multiple Estimates", "Worked Example", and "Contact Us".
- Data Input:** A table with columns for "Label", "First Estimate", "First Margin of Error (MOE)", "Second Estimate", "Second Margin of Error (MOE)", and "Statistically Different?". The first row contains placeholder text: "1 Place your Geography of statistical variable keyword 2 in the gray column here. This text is here as an example. You may delete the text before using the spreadsheet.". The second row contains numerical data: First Estimate (48759, 6.2, 22147046, -), First Margin of Error (MOE) (+/-2,395, +/-0.1, +/-234,312, **), Second Margin of Error (MOE) (****, 6, 17997953, 27047), and Statistically Different? (Yes, No, Yes, N/A).
- Source:** A note at the bottom left states: "Source: U.S. Census Bureau, American Community Survey (ACS), Statistical Testing Tool, <www.census.gov/programs-surveys/acs/guidance/statistical-testing-tool.html>".

Calculating Margins of Error for Custom (User-Derived) Estimates

In some cases, researchers will need to construct custom ACS estimates by combining data across multiple geographic areas or population subgroups, or it may be necessary to derive a new percentage, proportion, or ratio from published ACS data. For example, one way to address the issue of unreliable estimates for individual census tracts or block groups is to aggregate geographic areas, yielding larger samples and estimates that are more reliable. In such cases, additional calculations are needed to produce MOEs and to conduct tests of statistical significance for the derived estimates. The section on “Calculating Measures of Error for Derived Estimates” in the Census Bureau’s handbook on *Understanding and Using American Community Survey Data: What All Data Users Need to Know* provides detailed instructions on how to make these calculations.²⁸ Each ACS data release is also accompanied by “Accuracy of the Data” documentation that includes formulas for calculating MOEs (see Figure 3.2).²⁹

²⁸ U.S. Census Bureau, *Understanding and Using American Community Survey Data: What All Data Users Need to Know*, <www.census.gov/programs-surveys/acs/guidance/handbooks/general.html>.

²⁹ U.S. Census Bureau, American Community Survey (ACS), Code Lists, Definitions, and Accuracy, <www.census.gov/programs-surveys/acs/technical-documentation/code-lists.html>.

Users should note that some of the general formulas for calculating MOEs for derived estimates produce approximations rather than exact MOEs. Advanced users may be interested in the Variance Replicate Tables, first released for the 2010–2014 ACS 5-year estimates in July 2016.³⁰ These augmented ACS Detailed Tables include sets of 80 replicate estimates, which allow users to calculate MOEs for derived estimates using the same methods that are used to produce the published MOEs in the premade tables from the Census Bureau. These methods incorporate the covariance between estimates that the approximation formulas in the “Accuracy of the Data” document do not include.

The Variance Replicate Tables are available for a subset of the 5-year Detailed Tables for 11 geographic summary levels including the nation, states, counties, census tracts, and block groups. These tables are released on an annual basis, shortly after the release of the standard 5-year data products.

Variance Replicate Tables documentation, including lists of tables and summary levels, is available on the Census Bureau’s Web site.³¹

³⁰ U.S. Census Bureau, American Community Survey (ACS), Variance Replicate Tables, <www.census.gov/programs-surveys/acs/data/variance-tables.html>.

³¹ U.S. Census Bureau, American Community Survey (ACS), Variance Replicate Tables Documentation, <www.census.gov/programs-surveys/acs/technical-documentation/variance-tables.html>.

Figure 3.2. Code Lists, Definitions, and Accuracy

The screenshot shows a web page with a dark blue header and a light gray footer. The header contains the title 'Code Lists, Definitions, and Accuracy'. Below the header is a sidebar with a dark blue background and white text, listing various links under categories like 'About the Survey', 'Respond to the Survey', 'News & Updates', 'Data', 'Guidance for Data Users', 'Geography & ACS', 'Technical Documentation', 'Methodology', 'Library', and 'Operations and Administration'. The main content area has a white background. It features a 'Code Lists, Definitions, and Accuracy' section with social media sharing icons (Twitter and Facebook). Below this is a timeline showing years from 2016 to 2012. Under each year, there are sections for 'Code Lists', 'Subject Definitions', 'Group Quarters Definitions', 'Comparison Guidance', 'Instructions for Applying Statistical Testing', and a 'Statistical Testing Tool'. A red box highlights the 'Accuracy of the Data' section in the bottom right, which contains links to '2012-2016 ACS Multiyear Accuracy (US)', '2012-2016 PRCS Multiyear Accuracy (Puerto Rico)', '2016 ACS 1-year Accuracy of the Data (US)', and '2016 PRCS 1-year Accuracy of the Data (Puerto Rico)'. The footer contains the source information: 'Source: U.S. Census Bureau, American Community Survey (ACS), Code Lists, Definitions, and Accuracy, <www.census.gov/programs-surveys/acs/technical-documentation/code-lists.html>'.

Calculating Standard Errors for ACS PUMS Estimates

Researchers using the microdata files need to calculate their own estimates of standard error due to sampling, using either a generalized variance function (generalized standard errors) or by using the replicate weights (direct standard errors). The Census Bureau presents both approaches in their Accuracy of the Public Use Microdata Sample documentation.³²

The Census Bureau notes that, “Direct standard errors will often be more accurate than generalized standard errors, although they may be more inconvenient for some users to calculate. The advantage of using replicate weights is that a single formula is

³² U.S. Census Bureau, American Community Survey (ACS), PUMS Technical Documentation, <www.census.gov/programs-surveys/acs/technical-documentation/pums/documentation.html>.

used to calculate the standard error of many types of estimates.”³³

With generalized standard errors, “design factors” are applied to “reflect the effects of the actual sample design and estimation procedures used for the ACS.”³⁴ There is a tendency for the formula for generalized standard errors to estimate higher standard errors than the replicate weights method. This means that by using the generalized formula instead of the replicate weights, one may be less likely to find statistically significant differences where they actually do exist.

³³ U.S. Census Bureau, Public Use Microdata Sample (PUMS), Accuracy of the Data, p. 16, 2016.

³⁴ U.S. Census Bureau, Public Use Microdata Sample (PUMS), Accuracy of the Data, p. 18, 2016.

4. OTHER CONSIDERATIONS IN WORKING WITH ACS DATA

Using ACS Data for Population and Housing Counts

Many researchers need data on the *number* of people and housing units in a given geographic area and how those numbers have changed over time. Such users need to understand that the American Community Survey (ACS) was designed to provide estimates of the *characteristics* of the population, not to provide counts of the population in different geographic areas or population subgroups. Therefore, data users are encouraged to rely more upon noncount statistics, such as percent distributions or averages, when using ACS estimates.

The U.S. Census Bureau's Population Estimates Program produces and disseminates the official estimates of the population for the nation, states, counties, cities, and towns, and estimates of housing units for states and counties.³⁵ For 2010 and other decennial census years, the decennial census provides the official counts of population and housing units.³⁶

The Census Bureau uses a weighting method to ensure that ACS estimates are consistent with official population estimates at the county level by age, sex, race, and Hispanic origin—as well as estimates of total housing units. ACS single-year estimates are controlled to population and total housing unit estimates as of July 1 of the survey year, while ACS 5-year estimates are controlled to the average of the July 1 population and housing unit estimates over the 5-year period.

Starting with the 2009 survey, ACS estimates of the total population of incorporated places (self-governing cities, towns, or villages) and minor civil divisions (county subdivisions, in 20 states where they serve as functioning governmental units) are also adjusted so they are consistent with official population estimates. However, ACS data for other statistical areas, such as Public Use Microdata Areas (PUMAs) or census tracts, have no control totals, which may lead to larger margins of error for population and housing unit estimates than in areas of similar size with control totals. In such cases, data users are again encouraged to rely more on noncount statistics such as percent distributions or averages.

³⁵ U.S. Census Bureau, Population and Housing Unit Estimates, <www.census.gov/popest/>.

³⁶ See, for example, the U.S. Census Bureau, "Census of Population and Housing, CPH-2," *Population and Housing Unit Counts report series*, <www.census.gov/prod/www/decennial.html>.

Comparing Geographic Areas

One of the main benefits of the ACS is the ability to make comparisons—over time, across different geographic areas, and across different population subgroups.

When making comparisons with ACS data, note that differences in survey design, questionnaire content and design, sample size, or geography may affect comparability of estimates. Researchers interested in making comparisons also need to pay attention to sampling error because differences between estimates may or may not be statistically significant. See the section on "Sampling Error in the ACS" for more information.

Data users also need to decide how to compare geographic areas with different population sizes. ACS estimates for areas with fewer than 20,000 people are provided only in the form of 5-year estimates. However, for larger areas with at least 65,000 people (or 20,000 people in the case of the 1-year Supplemental Estimates) both 1-year and 5-year data are available, so data users need to choose which estimates to use.³⁷

TIP: When comparing ACS estimates across different geographic areas or population subgroups, data users should avoid comparing ACS single-year estimates with ACS multiyear estimates. That is, 1-year estimates should only be compared with other 1-year estimates, and 5-year estimates should only be compared with other 5-year estimates.

Suppose a researcher wanted to compare veterans' characteristics in Athens, Texas—a small city southeast of Dallas—with veterans in Houston. Although the ACS publishes annual estimates on veterans for Houston, only 5-year estimates are available for Athens. Thus, data users should compare ACS 5-year estimates for Athens with ACS 5-year estimates for Houston, even though more recent, single-year estimates are available for Houston.

Another option for presenting ACS data for less populated areas is to show single-year estimates for large counties in Texas and then combine the remaining counties into a state "residual" by subtracting the available single-year data from the state total. Alternatively, data users could present ACS estimates for PUMAs, since they meet the 65,000-population

³⁷ One-year Supplemental Estimates are simplified versions of popular ACS tables available for geographic areas with at least 20,000 people.

threshold required for single-year estimates and are often used as a substitute for county-level data.³⁸

The Census Bureau provides additional guidance on Comparing ACS Data on their Web site.³⁹

Comparing ACS Data Over Time

TIP: When using 5-year estimates, data users are encouraged to compare ACS data over time based on nonoverlapping estimates. For example, it would be appropriate for a data user to compare the 2007–2011 ACS 5-year estimates to the 2012–2016 ACS 5-year estimates. However, it would not be appropriate for a data user to compare the 2011–2015 ACS 5-year estimates to the 2012–2016 ACS 5-year estimates.

Comparisons using ACS 1-year data are generally straightforward, but using multiyear estimates to look at trends for small populations can be challenging because they rely on pooled data for 5 years. For example, comparisons of 5-year estimates from 2011 to 2015 and 2012 to 2016 are unlikely to show much difference because 4 of the years overlap; both sets of estimates include the same data collected from 2012 through 2015.⁴⁰ The Census Bureau suggests comparing 5-year estimates that do not overlap—for example, comparing 2007–2011 ACS 5-year estimates with 2012–2016 ACS 5-year estimates.

There is a broader issue of how to use multiyear characterizations of an area to measure change over time. As the ACS program has moved forward, an entire series of multiyear estimates for various time intervals has become available. Data users now have access to nonoverlapping ACS 5-year estimates that have increased the value and utility of the data for monitoring trends in local communities. However, it is more challenging to capture rapid change in areas where only ACS 5-year estimates are available. For example, it was very difficult for local officials and planners to accurately assess changes in socioeconomic

³⁸ Although Public Use Microdata Areas typically follow county boundaries, this is not always the case, particularly in some New England states.

³⁹ U.S. Census Bureau, American Community Survey (ACS), Comparing ACS Data, <www.census.gov/programs-surveys/acs/guidance/comparing-acs-data.html>.

⁴⁰ While the interpretation of this difference is difficult, these comparisons can be made with caution. Users who are interested in comparing overlapping multiyear period estimates should refer to the section “Understanding Error and Determining Statistical Significance” in the Census Bureau’s handbook *Understanding and Using American Community Survey Data: What All Data Users Need to Know*, available at <www.census.gov/programs-surveys/acs/guidance/handbooks/general.html>.

characteristics accompanying expanded drilling in the Bakken oil fields in North Dakota—where there was a large influx of male workers—because the affected counties only received 5-year, rather than 1-year, ACS estimates.

TIP: Changes to ACS questions over time may also make it difficult to measure trends. For example, the Census Bureau made substantial changes to the 2008 ACS questions on labor force participation and the number of weeks worked. As a result, the Census Bureau recommends using caution when comparing 2008 and later labor force data with 2007 and earlier estimates.

The Census Bureau provides “New and Notable” information with each new ACS data release, including information about changes to tables that may affect users’ ability to measure trends over time.⁴¹ Data users should also consider changes in geographic boundaries, population controls, and inflation when analyzing trends with ACS data.

Geographic Boundaries

ACS data generally reflect the geographic boundaries as of the year the data are collected. While geographic boundary changes are somewhat infrequent, they do occur, and those changes can affect a data user’s ability to make comparisons over time. For example, congressional districts are redrawn every 10 years immediately following the decennial census. Congressional district data from the 2012 ACS reflect the new boundaries that were drawn after the 2010 Census, while ACS data for earlier years reflect the 2000 Census boundaries. Given the major changes to district boundaries after each census, a comparison of congressional district data between 2011 and 2012 is not feasible.

ACS data are also regularly updated to reflect local changes in geographic boundaries. For example, the city of Jurupa Valley, California, incorporated in July 2011. Data for this city was first published in 2012 and has been updated each subsequent year, but data are not available for Jurupa Valley for 2011 and earlier years. The Census Bureau does not revise ACS data for previous years to reflect changes in geographic boundaries. For more information, visit the Census Bureau’s Web page on Geography & ACS.⁴²

⁴¹ U.S. Census Bureau, American Community Survey (ACS), Data Releases, <www.census.gov/programs-surveys/acs/news/data-releases.html>.

⁴² U.S. Census Bureau, American Community Survey (ACS), Geography & ACS, <www.census.gov/programs-surveys/acs/geography-acs.html>.

Population Controls

The ACS uses a weighting methodology to ensure that ACS estimates are consistent with official Census Bureau population estimates by age, sex, race, and Hispanic origin. With each annual release of population estimates, the Population Estimates Program revises and updates the entire time series of estimates from the previous decennial census to the current year. However, ACS estimates for prior years are not revised or reweighted based on updated population estimates.

The change in the population estimates from 2009 to 2010 was particularly significant. The 2010 ACS 1-year data and 2006–2010 ACS 5-year data were controlled to population estimates that reflected the results of the 2010 Census. However, the 1-year and 5-year data for 2009 and earlier years used population estimates that were based on the 2000 Census.

TIP: Because the 2009 ACS and 2010 ACS 1-year estimates use controls that are based on different decennial census base years, data users need to use caution when making comparisons across these years. Specifically, estimates of the number of people in a given geographic area or population subgroup are not strictly comparable between these 2 years. However, rates and percentages—as well as monetary data, such as median income values—are generally comparable between the two periods.

Monetary Data

Data users also need to use caution in looking at trends involving income or other measures that are adjusted for inflation such as rental costs, home values, and energy costs.

For example, to compare published monetary data for the most recent year with data from the 2010 ACS, data users need to adjust the 2010 data for inflation based on a national-level consumer price index.

ACS multiyear estimates with dollar values are adjusted for inflation to the final year of the period. For example, the 2011–2015 ACS 5-year estimates are tabulated using dollars adjusted to 2015.

Note that inflation adjustment does not account for differences in costs of living across different geographic areas. For more information on the adjustment of ACS single-year and multiyear estimates for inflation, see the section on “Using Dollar-Denominated Data” in *Understanding and Using American Community Survey Data: What All Data Users Need to Know*.⁴³

⁴³ U.S. Census Bureau, *Understanding and Using American Community Survey Data: What All Data Users Need to Know*, <www.census.gov/programs-surveys/acs/guidance/handbooks/general.html>.

Comparisons With Data From the 2000 Census and the 2010 Census

The ACS was modeled after the long form of the decennial census, and data users interested in long-term trends can, in many cases, make valid comparisons between ACS and the 2000 Census (and earlier decennial census) estimates. Census Bureau subject-matter specialists have reviewed the factors that could affect differences between ACS and the 2000 Census estimates, and they have determined that ACS estimates are similar to those obtained from past decennial census sample data for most areas and characteristics.

However, differences in residence rules, universes (base reference totals against which all other characteristics are compared), and reference periods between the two surveys should be considered when making these comparisons. For example, the ACS data are collected throughout the calendar year, while the 2000 Census long form sampled the population as of April 1, 2000. Given the differences in the reference period, the two surveys may yield very different estimates for communities with large seasonal populations or those undergoing rapid change. The section on “Differences Between the ACS and the Decennial Census” in the handbook *Understanding and Using American Community Survey Data: What All Data Users Need to Know* provides more information about these differences.⁴⁴

The 2010 Census was a short-form only census, so it does not include all the detailed social, economic, and housing data available from previous censuses. However, data users can make valid comparisons between ACS estimates and basic characteristics from the 2010 Census including age, sex, race, Hispanic origin, household relationship, and housing tenure (homeowner or renter status). For basic counts of the U.S. population by age, sex, race, and Hispanic origin between censuses, data users are encouraged to use the Census Bureau’s official population estimates available on the Census Bureau’s Population and Housing Unit Estimates Web site.⁴⁵

For detailed guidance on comparing ACS and 2000 Census data, visit the Census Bureau’s Web page on Comparing ACS Data.⁴⁶

⁴⁴ U.S. Census Bureau, *Understanding and Using American Community Survey Data: What All Data Users Need to Know*, <www.census.gov/programs-surveys/acs/guidance/handbooks/general.html>.

⁴⁵ U.S. Census Bureau, Population and Housing Unit Estimates, <www.census.gov/programs-surveys/popest.html>.

⁴⁶ U.S. Census Bureau, American Community Survey (ACS), Comparing ACS Data, <www.census.gov/programs-surveys/acs/guidance/comparing-acs-data.html>.

5. CASE STUDIES USING ACS DATA

Case Study #1: Mobility and Economic Opportunity in New York City Neighborhoods

Skill Level: Intermediate/Advanced

Subject: Commuting/transportation challenges

Type of Analysis: Analysis of job opportunities, household income, and population size across New York City neighborhoods

Tools Used: Application Programming Interface (API), spreadsheet

Author: Sarah Kaufman, Assistant Director for Technology Programming at the New York University (NYU) Rudin Center for Transportation

The ability of a public transportation network to physically link residents to jobs has become a central point of concern for urban policy in an era of uneven unemployment and rapidly changing job markets. The economy of New York City is unique in North America due to the high proportion of residents using public transportation. In 2016, more than half of the population in New York City (56.6 percent) used some kind of public transportation to get to work, and an individual's ability to access a job is largely a function of how well their neighborhood is served by the public transportation system.

In a recent report, the Rudin Center for Transportation Policy and Management at NYU's Robert F. Wagner School of Public Service explored some of the key transportation challenges facing New York City residents, based on data from the American Community Survey (ACS) and other sources.⁴⁷ An accompanying interactive map enables users to explore the data for their neighborhoods.⁴⁸

Results showed disparities in transportation access. Furthermore, low levels of transit access were associated with lower income and employment among residents, while high levels of transit access were associated with higher income and employment.

Methods

Rudin Center staff analyzed and ranked 177 New York City neighborhoods based on access to job opportunities, household income, and population size. The rankings reflect the number of jobs that can be reached within 1 hour by public transportation. (A commute time of 1 hour or less was selected based on prior research showing that commuters prefer to travel less than 1 hour.)

Demographic data are from the U.S. Census Bureau's ACS 5-year estimates for ZIP Code Tabulation Areas (ZCTAs). ZCTAs are aggregations of census blocks that form "generalized areal representations of United States Postal Service (USPS) ZIP code service areas."⁴⁹

New York City fully contains 186 ZCTAs as defined in the 2010 Census. In this work, ZCTAs are only included as a unit of observation if they contain populations of at least 2,500 according to the 2008–2012 ACS. It should be noted that the estimates included in the report and interactive maps do not account for margins of error. However, the population threshold helps to ensure accurate demographic data exist within the ZIP code (unlike park areas), and to avoid small areas that would not be representative of a larger neighborhood. Of the 186 ZIP codes, 177 have a population of at least 2,500.⁵⁰

ACS estimates for ZCTAs were accessed through the Census Data API.⁵¹

To access 2008–2012 ACS 5-year employment and unemployment data from the Census Data API, enter the following query in your Web browser: <[https://api.census.gov/data/2012/acs/acs5?get=group\(B23025\)&for=ZIP%20code%20tabulation%20area:*](https://api.census.gov/data/2012/acs/acs5?get=group(B23025)&for=ZIP%20code%20tabulation%20area:*>)> as described in the steps below (see Figure 5.1).

1. Start your query with the host name: "<https://api.census.gov/data/>"

⁴⁷ The Rudin Center for Transportation Policy and Management, Mobility, Economic Opportunity and New York City Neighborhoods, November 2015, <<https://wagner.nyu.edu/impact/research/publications/mobility-economic-opportunity-and-new-york-city-neighborhoods>>.

⁴⁸ Datapolitan, NYC Neighborhoods: Mobility & Economic Opportunity, <www.datapolitan.com/job_access/>.

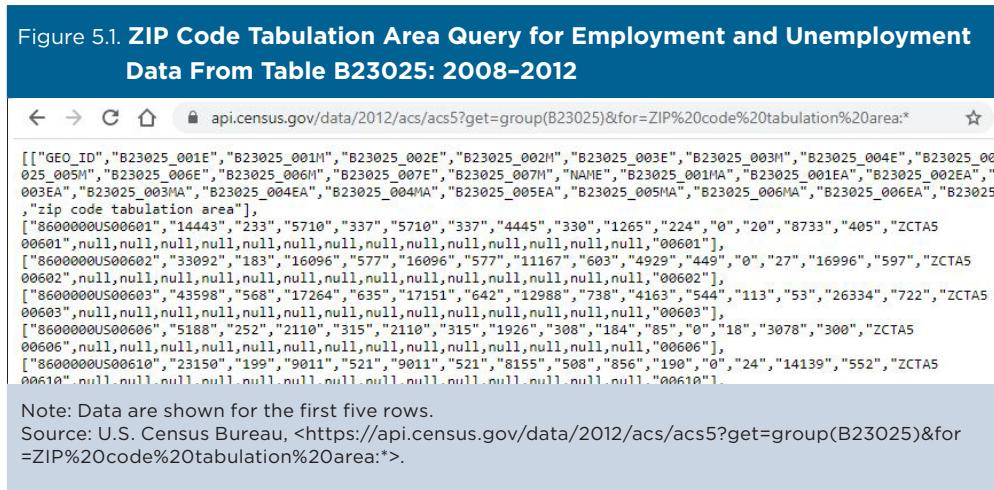
⁴⁹ U.S. Census Bureau, ZIP Code Tabulation Areas (ZCTAs), <www.census.gov/programs-surveys/geography/guidance/geo-areas/zctas.html>.

⁵⁰ In cases where ZCTA-level data were unavailable, census tract estimates were "cross-walked" to conform to ZCTA boundaries using an allocation algorithm provided by the Missouri Census Data Center.

⁵¹ U.S. Census Bureau, Census Data API User Guide, <www.census.gov/data/developers/guidance/api-user-guide.html>.

2. Add the data year (2012) to the URL: "<https://api.census.gov/data/2012>."
3. Add the data set name acronym for the ACS 5-Year Detailed Tables, and follow this base URL with a question mark: "<https://api.census.gov/data/2012/acs/acs5?>"
4. Add variables starting with a get clause, "get=:": "<https://api.census.gov/data/2012/acs/acs5?get=>"
5. Use the group feature to return all data items for Table B23025 (which contains labor force, employment, and unemployment details): "[https://api.census.gov/data/2012/acs/acs5?get=group\(B23025\)](https://api.census.gov/data/2012/acs/acs5?get=group(B23025))."
6. Add geography using a predicate clause starting with an ampersand (&) to separate it from your "get" clause and then a "for=" to identify geographic areas of interest:
[https://api.census.gov/data/2012/acs/acs5?get=group\(B23025\)&for=](https://api.census.gov/data/2012/acs/acs5?get=group(B23025)&for=)
7. Identify the geographic area(s) that you need (ZCTAs) by reviewing the list of geographies available for the 2008–2012 ACS 5-year Detailed Tables.⁵²
8. Because you need data for many ZIP codes, add a wildcard (*) to get all ZCTA values:
[https://api.census.gov/data/2012/acs/acs5?get=group\(B23025\)&for=ZIP%20code%20tabulation%20area:*](https://api.census.gov/data/2012/acs/acs5?get=group(B23025)&for=ZIP%20code%20tabulation%20area:*>)

After downloading the comma-separated file, we opened it in a spreadsheet to analyze the data.



To calculate the unemployment rate, we divided a ZCTA's unemployed population (B23025_005E) by its civilian labor force (B23025_003E). Using the example of Chelsea (North), ZCTA 10001, we calculated an unemployment rate of 9 percent (see Figure 5.2).

⁵² U.S. Census Bureau, <<https://api.census.gov/data/2012/acs/acs5/geography.html>>.

We repeated a similar process for all other ACS variables of interest. ACS data were then combined with information from the Google Maps Routing API and the Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) Origin-Destination Employment Statistics (LODES) data set.

Figure 5.2. Unemployment Rate for Chelsea (North) ZCTA 10001: 2008-2012

J5	A	B	C	D	E	F	G	H	I	J
1	geo_id	B23025_001E	B23025_001M	B23025_002E	B23025_002M	B23025_003E	B23025_003M	B23025_004E	B23025_004M	B23025_005E
2	8600000US10001	19026	918	13734	856	13734	856	12538	816	1196
	population ages 16 and older		labor force		civilian labor force		employed		unemployed	
3										
4										
5										unemployment rate = unemployed / civilian labor force 9%
6										

Source: Author's analysis of data from the U.S. Census Bureau, American Community Survey.

The Google Maps Routing API was used to estimate travel times between origins and destinations. The API can be queried with origin and destination pairs to output the estimated travel time according to Google's algorithm. This project used this service to generate a data set containing all ZIP code-level travel times in the region, which originated in New York City and terminated anywhere in the New York, New Jersey, or Connecticut region.

The LEHD data set provides employment counts by subcategories at the census block level. LODES provides a level of detail regarding employment that is not available from the ACS. LODES data were cross-walked from census blocks to ZIP codes using a Missouri Census Data Center tool.⁵³ Because census blocks are even smaller than the census tracts used for demographic data, there is essentially no loss of precision due to cross-walking to the much larger ZIP-code level. This report uses the LODES data for 2013, which were the most current at the time of publication.

Data points from the three aforementioned sources were merged together to create a single observation for each ZIP code in New York City. LODES data were downloaded for all of New York State, New Jersey, and Connecticut; this allowed job counts to be assigned to ZIP codes for the entire region. Google routing data were collected for journeys originating within a ZIP code in New York City, but ending in any ZIP code within the larger region. ACS data were collected for New York City only.

More detailed information about these methods is available in the report.

Results

The data show that mass transit access is associated with job opportunities and household income levels in most New York City neighborhoods.

The rankings, along with the summary chart below, show the swoosh-shaped relationship between transit and income in New York City: Neighborhoods with some, but insufficient transit access—those ranked in the middle third—faced higher rates of unemployment than those in the top or bottom third (see Figure 5.3).

Our partners at Datapolitan then turned the resulting data, for all ZIP codes, into an online, interactive application (see Figures 5.4 and 5.5).

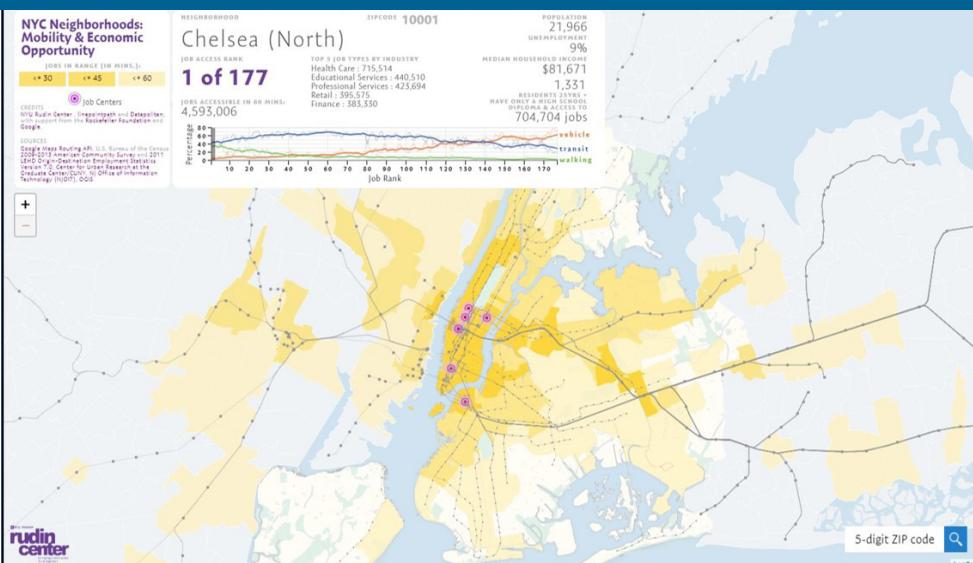
⁵³ Missouri Census Data Center, Geocorr 2014: Geographic Correspondence Engine, <<http://mc当地.missouri.edu/applications/geocorr2014.html>>.

Figure 5.3. New York Ranked Neighborhoods: Income, Unemployment, and Commuting: 2008–2012

Ranked Neighborhoods	Median Household Income	Unemployment Rate	Commute by Transit or Walking (avg)	Commute by Car (avg)
1-59	\$81,286	8.1%	79.1%	10.8%
60-119	\$46,937	12.6%	67.1%	27.6%
120-177	\$59,949	10.4%	44.2%	52.1%

Source: Author's analysis of data from the U.S. Census Bureau, American Community Survey.

Figure 5.4. Results for Chelsea (North) ZCTA 10001



Source: Datapolitan, NYC Neighborhoods: Mobility & Economic Opportunity, www.datapolitan.com/job_access/.

Figure 5.5. Results for Chelsea (North), ZCTA 10001



Source: Datapolitan, NYC Neighborhoods: Mobility & Economic Opportunity, www.datapolitan.com/job_access/.

Case Study #2: State Level Trends in Children's Health Insurance Coverage

Skill Level: Intermediate/Advanced

Subject: State-level trends in children's health insurance coverage

Type of Analysis: Analysis of changes in children's health insurance coverage over time

Tools Used: American Community Survey Public Use Microdata Sample (PUMS) files, statistical software, spreadsheet

Author: Brett Fried, Senior Research Fellow, State Health Access Data Assistance Center (SHADAC)

The State Health Access Data Assistance Center (SHADAC) is a multidisciplinary health policy research center affiliated with the University of Minnesota that focuses on state health policy. "State-Level Trends in Children's Health Insurance Coverage" (the "Kids' Report") is one of many reports that SHADAC produces at the state level to show trends over time in insurance coverage, access, cost, utilization, and outcomes, as well as in equity and economic measures.⁵⁴

Approach

To generate reports from the American Community Survey (ACS) PUMS files, SHADAC started by creating an analytical data set using SAS. The microdata allowed us to create custom variables such as a health insurance unit (HIU) in this data set. The HIU defines "family" based on who is likely considered part of a "family unit" in determining eligibility for either private or public coverage. HIU is a narrower definition of family, compared with the Census Bureau's general definition of family that groups all related members of a household into a family.⁵⁵ We also created Affordable Care Act (ACA)-relevant poverty-level categories—0 to 138 percent of the Federal Poverty Guideline (FPG); 139 to 400 percent FPG; and 401 percent FPG or more. To measure family poverty, income is totaled for all individuals in the health insurance unit. The income is divided by the FPG produced by the U.S. Department of Health and Human Services to calculate the income as a percentage of FPG. (In 2016, the federal poverty guideline for a family of four was \$24,300.) We used SAS to create the analytic data set that included the custom HIU and poverty-level categories and then transferred the data set using StatTransfer software into a STATA data set to produce relevant estimates.

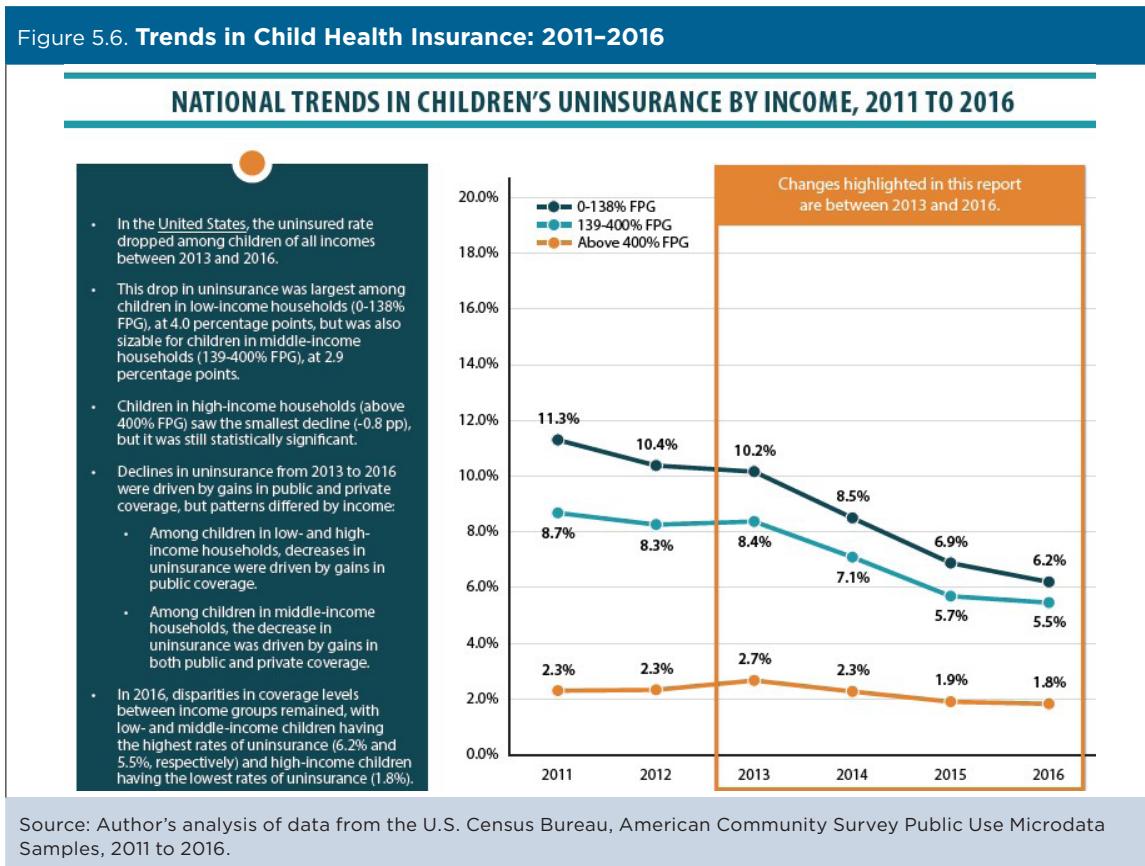
After transferring the data set into STATA, we created variables for other subjects of interest such as race/ethnicity and educational attainment. Then we produced estimates for all the custom HIU and income categories, broken down by coverage type, using STATA code. For example, we produced estimates of children by private coverage, public coverage, and uninsurance by three income categories from 2011 to 2016. If someone had more than one source of coverage, we considered private coverage as primary over public sources.

Next, we tested for statistically significant percentage-point differences in the estimates between 2013 (generally, pre-ACA implementation) and 2016 (post-ACA implementation). Percentage-point differences between years are reported in the tables. We produced three products from these estimates. The first product is a summary report where we use maps, tables, and figures to highlight the main findings. Estimates with coefficients of variation (standard error/estimate) greater than 30 percent are not included in the report (see Figure 5.6).

⁵⁴ SHADAC, "State-Level Trends in Children's Health Insurance Coverage," 2013–2016, <www.shadac.org/KidsReport2016>.

⁵⁵ SHADAC has a more detailed description of how we create the HIU in SHADAC I (Defining Family for Studies of Health Insurance Coverage), <www.shadac.org/publications/defining-family-studies-health-insurance-coverage>.

Figure 5.6. Trends in Child Health Insurance: 2011–2016



The second product is a set of 50-state tables. These detailed tables allow for cross-year comparisons between states from 2013 to 2016. Statistically significant differences between years at a 95 percent confidence level are indicated with an asterisk (see Figure 5.7).

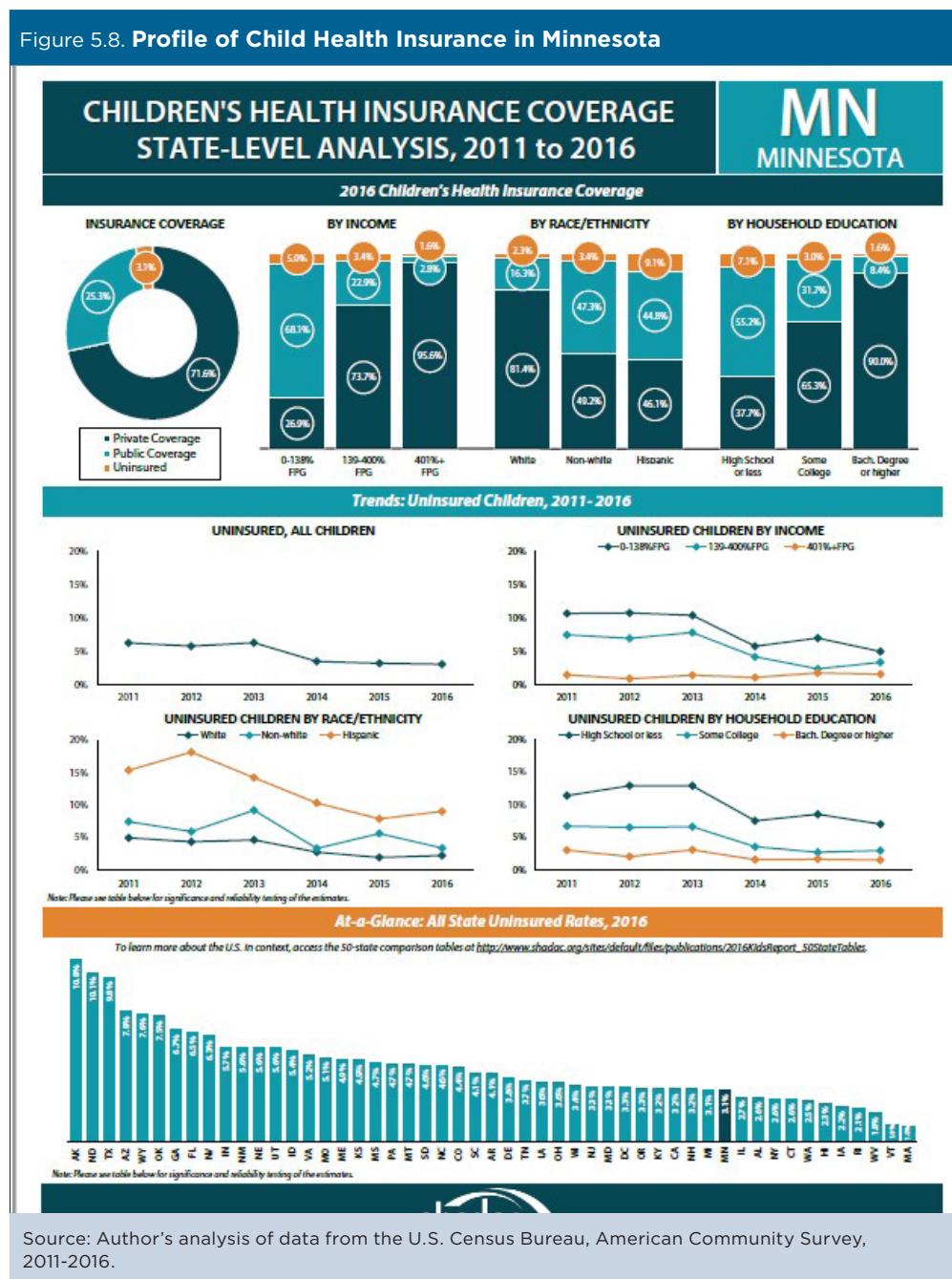
Figure 5.7. Trends in Health Insurance Coverage for Children by State: 2013–2016

TREND IN COVERAGE FOR CHILDREN

State	PRIVATE COVERAGE			PUBLIC COVERAGE			UNINSURED		
	2013 %	2016 %	Percent Point Change	2013 %	2016 %	Percent Point Change	2013 %	2016 %	Percent Point Change
Alabama	55.80%	55.4%	-0.4%	39.2%	41.9%	2.7% *	5.0%	2.6%	-2.3% *
Alaska	62.6%	58.6%	-4.0%	25.2%	30.6%	5.4%	12.2%	10.8%	-1.4%
Arizona	53.9%	55.6%	1.8%	33.4%	36.6%	3.2% *	12.7%	7.8%	-5.0% *
Arkansas	46.8%	48.1%	1.3%	46.7%	47.8%	1.1%	6.5%	4.1%	-2.4% *
California	54.7%	55.5%	0.8%	37.4%	41.3%	3.9% *	7.9%	3.2%	-4.6% *
Colorado	64.5%	63.6%	-0.9%	26.5%	31.9%	5.4% *	9.0%	4.4%	-4.5% *
Connecticut	66.5%	67.1%	0.6%	29.2%	30.4%	1.1%	4.3%	2.6%	-1.7% *

Source: Author's analysis of data from the U.S. Census Bureau, American Community Survey Public Use Microdata Samples, 2013 and 2016.

The third product is a set of individual state profiles. These two-page profiles provide “at-a-glance” graphic summaries of 5-year trends in children’s health insurance coverage for each state and the United States, including statistical comparisons (see Figure 5.8).



Findings

In the Kids' Report released in June 2018, we found that since the coverage provisions of the ACA took effect, children in the United States have seen significant declines in uninsurance, with the number of uninsured children dropping by 2.2 million, or 2.9 percentage points, between 2013 and 2016. These coverage gains were sustained despite an uncertain policy climate around the ACA. Drops in uninsurance were seen across demographic categories, and some of the largest coverage gains were made by groups of children who historically have had the highest rates of uninsurance: low-income, Hispanic, and non-White children, and children in households with low educational attainment. Despite coverage gains, coverage rates for these groups are still significantly below those of high-income children and White children, and coverage varies across states.

Lessons Learned

One of the lessons learned from this and other similar projects that include data for all states is that category definitions matter. For example, if the categories are too narrow, then estimates in many states will be suppressed (for example, if "American Indian and Alaska Native" is one of categories that is cross tabulated with children's coverage, then most state estimates will be suppressed due to small sample size and large margins of error around the estimates).

Impact

The SHADAC Kids' Report is updated annually as new data become available. The report is used as a resource by state and federal analysts, academic researchers, the media, nonprofits, advocacy groups, foundations, and the public, as well as by internal SHADAC coworkers. In the first 3 weeks after its release, the report was viewed nearly 250 times.

Case Study #3: Children Living in Areas of Concentrated Poverty

Skill Level: Advanced

Subject: Neighborhood poverty

Type of Analysis: Estimating the percentage of children who live in neighborhoods of concentrated poverty

Tools Used: American Community Survey (ACS) Summary File, statistical software (SAS), spreadsheet

Author: Jean D'Amico, Senior Research Associate, Population Reference Bureau (PRB)

Researchers largely agree that the residential clustering of poverty adversely affects the life chances of residents living in those high-poverty areas. There is also general consensus in the literature that the deleterious effects of residential concentrated poverty can occur once poverty rates reach a level of 20 to 40 percent. For this case study, we analyzed the percentage of children under the age of 18 living in areas of concentrated poverty—defined as census tracts with overall poverty rates of 30 percent or more.

Because we wanted to work with census-tract level data, we needed to use ACS 5-year data. We used the ACS Summary File because we needed data for a large number of geographic areas (every census tract in the nation).

Step 1. Extract the Data

There are several ways to access census tract data from the U.S. Census Bureau's Web site, including data.census.gov. However, for this example, we use a SAS macro program to extract data from the ACS Summary File. This program is intended for advanced users who need to extract data for many geographies at once.

Our first step is to download the SAS program that we need to merge the ACS estimate and margin of error files with the geography files.⁵⁶ The “5-Year Macros” program is designed to read in the ACS 5-year Summary File. The program includes detailed comments that guide users through each procedure and macro (see Figure 5.9).

Figure 5.9. Downloadable SAS Program to Extract ACS Summary File Data

The screenshot shows a web browser window with the URL <https://www.census.gov/programs-surveys/acs/technical-documentation/summary-file-documentation....> in the address bar. The main content area has a blue header "SAS Programs". Below it, text reads: "Contains SAS programs for each sequence per state, which can be used to convert each estimate and margin of error into SAS Datasets with table stubs". Underneath are two download links: "1-year SAS programs [41.0 MB]" and "5-year SAS programs [31.0 MB]". Below these is a link "Example of Creating a Table Using SAS [<1.0 MB]". Further down, there is a detailed description of the "5-Year Macros" program, followed by two more download links: "1-Year Macros [<1.0MB]" and "5-Year Macros [<1.0 MB]". At the bottom, a note states: "Source: U.S. Census Bureau, American Community Survey (ACS), Summary File Documentation, <www.census.gov/programs-surveys/acs/technical-documentation/summary-file-documentation...2016.html>."

⁵⁶ SAS programs for 2016 data can be found at <www.census.gov/programs-surveys/acs/technical-documentation/summary-file-documentation...2016.html> under the heading “SAS Programs.”

Step 2. Identify the Tables of Interest

Using the Sequence Number/Table Number Lookup file, we identify the tables needed to calculate our measure and note three key pieces of information for each: the table number, the sequence number, and the line numbers.⁵⁷

To determine the poverty rate of each tract, we need:

1. The table number (B17001).
2. The sequence number (48).
3. The line numbers needed for calculations (2 and 31) (see Figure 5.10).

Figure 5.10. Table, Sequence, and Line Numbers for Table B17001: Poverty Status in the Past 12 Months by Sex by Age

File ID	Table ID	Sequence Number	Line Number	Start Position	Total Cells in Table	Total Cells in Sequence	Table Title	Subject Area
ACSSF	B17001		48		7	59 CELLS	POVERTY STATUS IN THE PAST 12 MONTHS BY SEX BY AGE	Poverty
ACSSF	B17001		48				Universe: Population For Whom Poverty Status Is Determined	
ACSSF	B17001		48	1			Total:	
ACSSF	B17001		48	2			Income in the past 12 months below poverty level:	
ACSSF	B17001		48	3			Male:	
ACSSF	B17001		48	4			Under 5 years	
B17001 Continued								
ACSSF	B17001		48	29			65 to 74 years	
ACSSF	B17001		48	30			75 years and over	
ACSSF	B17001		48	31			Income in the past 12 months at or above poverty level:	
ACSSF	B17001		48	32			Male:	
ACSSF	B17001		48	33			Under 5 years	
ACSSF	B17001		48	34			5 years	

Source: U.S. Census Bureau, Sequence Number/Table Number Lookup file, 2016, <www.census.gov/programs-surveys/acs/technical-documentation/summary-file-documentation.2016.html>.

To determine the total population of children living in each tract, we need:

1. The table number (B09001).
2. The sequence number (34).
3. The line number needed for calculations (1) (see Figure 5.11).

Figure 5.11. Table, Sequence, and Line Numbers for Table B09001: Population Under 18 Years by Age

File ID	Table ID	Sequence Number	Line Number	Start Position	Total Cells in Table	Total Cells in Sequence	Table Title	Subject Area
ACSSF	B09001		34		7	10 CELLS	POPULATION UNDER 18 YEARS BY AGE	Children - Relationship
ACSSF	B09001		34				Universe: Population Under 18 years	
ACSSF	B09001		34	1			Total:	
ACSSF	B09001		34	2			In households:	
ACSSF	B09001		34	3			Under 3 years	

Source: U.S. Census Bureau, Sequence Number/Table Number Lookup, 2016, <www.census.gov/programs-surveys/acs/technical-documentation/summary-file-documentation.2016.html>.

⁵⁷ U.S. Census Bureau, American Community Survey (ACS), Summary File Documentation, <www.census.gov/programs-surveys/acs/technical-documentation/summary-file-documentation.html>.

Step 3. Download the Data

Now that we know which files we need, we can download them from the Census Bureau's File Transfer Protocol server.⁵⁸

Since we are interested in collecting tract-level data for the entire United States, and we are using SAS statistical software, we access the complete set of ACS 5-year Summary Files from the “5_year_entire_sf/” directory, which includes data for all census tracts in all states (see Figure 5.12).⁵⁹

Figure 5.12. Summary File Download: 2016				
	Name	Last modified	Size	Description
»	Parent Directory		-	
📁	1_year_by_state/	29-Aug-2017 17:01	-	
📁	1_year_entire_sf/	29-Aug-2017 17:30	-	
📁	1_year_seq_by_state/	29-Aug-2017 17:36	-	
📁	5_year_by_state/	22-Nov-2017 09:54	-	
📁	5_year_entire_sf/	12-Dec-2017 10:32	-	
📁	5_year_seq_by_state/	22-Nov-2017 15:47	-	
⬇️	2016_1yr_Summary_FileTemplates.zip	29-Aug-2017 16:55	2.3M	
⬇️	2016_5yr_Summary_FileTemplates.zip	22-Nov-2017 12:50	1.7M	

Source: U.S. Census Bureau, <https://www2.census.gov/programs-surveys/acs/summary_file/2016/data>.

Within the “5_year_entire_sf” directory, there are several files. We need to download files with 2016 tract-level ACS estimates and their associated margins of error. We also need the 2016 ACS geography files. We download and unzip the geography files (2016_ACS_Geography_Files.zip) and the estimate and margin of error files (Tracts_Block_Groups_Only.tar.gz) (see Figure 5.13).

Figure 5.13. Summary File Download: 2016				
	Name	Last modified	Size	Description
»	Parent Directory		-	
⬇️	2016_ACS_Geography_Files.zip	28-Dec-2017 11:45	34M	
⬇️	All_Geographies_Not_Tracts_Block_Groups.tar.gz	22-Nov-2017 10:46	6.3G	
⬇️	Tracts_Block_Groups_Only.tar.gz	22-Nov-2017 11:08	3.5G	

Source: U.S. Census Bureau, <https://www2.census.gov/programs-surveys/acs/summary_file/2016/data/5_year_entire_sf/>.

⁵⁸ U.S. Census Bureau, American Community Survey (ACS), Data via FTP, <www.census.gov/programs-surveys/acs/data/data-via-ftp.html>.

⁵⁹ Note: If we were interested in a specific state, we could save download time and disk space by downloading only that state. See the directory link for 5_year_by_state in Figure 5.12.

Step 4. Access and Analyze the Data

Now that we have the files we need, we can access the data using the 5-year macro program described in Step 1. The 5-Year Macro SAS program needs to be edited to reflect the file paths of our unzipped files. The macro program accesses the geography, estimate, and margin of error data and creates a single table for all geographies from the ACS Summary File. The final data set we create for our analysis includes all tracts (as separate rows) and the estimate and margin of error variables of interest for computing our measure (see Figure 5.14).

Figure 5.14. Selected Records and Variables From Summary File Table B17001 by Census Tract: 2016

STUSAB	LOGRECNO	STATE	TRACT	SEQUENCE	B17001e2	B17001m2	B17001e31	B17001m31	B09001e1	B09001m1
al	1762	1	20100	48	199	129	1811	248	456	126
al	1763	1	20200	48	483	210	1531	252	478	115
al	1764	1	20300	48	337	177	2799	364	701	198
al	1765	1	20400	48	125	97	4438	479	1075	230
al	1766	1	20500	48	998	622	9322	857	2852	349
al	1767	1	20600	48	237	151	3493	406	957	230
al	1768	1	20700	48	573	241	2450	337	758	175
al	1769	1	20801	48	446	244	2579	311	818	187
al	1770	1	20802	48	1639	495	9094	692	2900	382
al	1771	1	20900	48	594	341	5318	581	1487	244
al	1772	1	21000	48	436	193	2449	350	709	152
al	1773	1	21100	48	630	205	2617	322	662	149

Source: U.S. Census Bureau, ACS Summary File.

With the final data set complete, we begin constructing our measure by computing the poverty rate for each tract. Recall that Table B17001, line 2 (variable B17001e2) is the sum of those living below poverty. Table B17001, line 31 (variable B17001e31) is the sum of those living at or above poverty (see Figure 5.14, above). Therefore, the percentage of residents in a census tract who are living below poverty is calculated as:

$$\text{Percentage in Poverty} = \frac{\text{B17001e2}}{\text{B17001e2} + \text{B17001e31}} * 100;$$

SAS code:

PCTPOVERTY = B17001e2 / (B17001e2 + B17001e31) *100;

Figure 5.15. Calculating the Percentage in Poverty by Census Tract: 2016

STUSAB	LOGRECNO	STATE	TRACT	SEQUENCE	B17001e2	B17001e31	PCTPOVERTY
al	1834	1	953000	48	369	1108	25.0
al	1835	1	953100	48	1129	1769	39.0
al	1836	1	953200	48	850	3156	21.2
al	1837	1	953300	48	310	1481	17.3
al	1838	1	953400	48	715	1954	26.8
al	1839	1	953500	48	355	941	27.4
al	1840	1	200	48	1199	1835	39.5
al	1841	1	300	48	1101	1408	43.9
al	1842	1	400	48	1145	1566	42.2
al	1843	1	500	48	755	689	52.3
al	1844	1	600	48	845	663	56.0
al	1845	1	700	48	900	1711	34.5
al	1846	1	800	48	268	722	27.1

Source: Author's analysis of data from the U.S. Census Bureau, American Community Survey Summary File.

Next, we create a variable that will identify the number of children who live in tracts with poverty rates at or above 30 percent. We assign a value of zero to a variable when the poverty rate in the tract is below 30 percent. If the poverty rate in the tract is 30 percent or greater, the variable is equal to the child population of that tract. Recall that table B09001 line 1 (B09001e1) is the total population under 18 years (see Figure 5.16). The number of children who live in high-poverty tracts is computed as follows:

SAS code:

NUMCHILD = 0;

If PCTPOVERTY > = 30 then NUMCHILD = B09001e1;

**Figure 5.16. Calculating the Number of Children in High-Poverty Census Tracts:
2016**

STUSAB	LOGRECNO	STATE	TRACT	SEQUENCE	B17001e2	B17001e31	B09001e1	PCTPOVERTY	NUMCHILD
al	1834	1	953000	48	369	1108	337	25.0	0
al	1835	1	953100	48	1129	1769	872	39.0	872
al	1836	1	953200	48	850	3156	993	21.2	0
al	1837	1	953300	48	310	1481	402	17.3	0
al	1838	1	953400	48	715	1954	602	26.8	0
al	1839	1	953500	48	355	941	240	27.4	0
al	1840	1	200	48	1199	1835	518	39.5	518
al	1841	1	300	48	1101	1408	588	43.9	588
al	1842	1	400	48	1145	1566	660	42.2	660
al	1843	1	500	48	755	689	359	52.3	359
al	1844	1	600	48	845	663	383	56.0	383
al	1845	1	700	48	900	1711	629	34.5	629
al	1846	1	800	48	268	722	238	27.1	0
al	1847	1	900	48	768	2989	882	20.4	0
al	1848	1	1000	48	589	5519	1232	9.6	0
al	1849	1	1100	48	772	6240	1762	11.0	0
al	1850	1	1201	48	757	2270	495	25.0	0

Note: High-poverty is defined as a poverty rate at or above 30 percent.

Source: Author's analysis of data from the U.S. Census Bureau, American Community Survey Summary File.

The tract-level totals can be summed to larger levels of geography such as states or the entire United States. When we sum our variables by state, we create a new data set where the observations are the United States, the 50 states, the District of Columbia, and Puerto Rico. Our NUMCHILD variable reflects the number of children in the state (or the nation) who live in high-poverty tracts. The last step is to calculate the percentage of children living in high-poverty tracts for each of these areas.

To calculate the percentage of children in each state and the nation living in high-poverty tracts, we divide the number of children who live in high-poverty tracts (NUMCHILD) by the total population of children (B09001e1) (see Table 5.1).

SAS code:

PCTCHILD = NUMCHILD / B09001e1*100;

According to the 2012–2016 ACS 5-year estimates, a total of 9.4 million children under 18 years of age lived in a high-poverty neighborhood, representing 13 percent of all children in the United States.

Table 5.1. Number and Percentage of Children Living in High-Poverty Census Tracts by State: 2012–2016

State	B17001e2	B17001e31	B09001e1	PCTPOVERTY	NUMCHILD	PCTCHILD
United States	46,932,225	263,697,420	73,612,438	15.1	9,448,167	12.8
Alabama	868,666	3,851,926	1,105,189	18.4	178,052	16.1
Alaska	72,826	646,238	187,616	10.1	10,148	5.4
Arizona	1,165,636	5,407,887	1,619,618	17.7	372,624	23
Arkansas	542,431	2,338,973	707,234	18.8	112,970	16
California	6,004,257	31,908,887	9,140,283	15.8	1,385,724	15.2
Colorado	637,938	4,603,119	1,246,181	12.2	71,076	5.7
Connecticut	360,464	3,119,208	773,652	10.4	70,880	9.2
Delaware	109,448	799,525	204,192	12	7,684	3.8
District of Columbia	112,060	512,894	114,685	17.9	30,396	26.5
Florida	3,139,258	16,375,076	4,066,276	16.1	500,585	12.3
Georgia	1,746,894	8,082,162	2,495,175	17.8	399,602	16
Hawaii	148,577	1,227,351	308,216	10.8	16,719	5.4
Idaho	244,585	1,359,784	431,320	15.2	19,932	4.6
Illinois	1,753,731	10,794,807	2,990,629	14	336,198	11.2
Indiana	957,694	5,431,232	1,581,992	15	190,892	12.1
Iowa	369,828	2,635,980	727,514	12.3	23,598	3.2
Kansas	373,162	2,443,029	721,347	13.3	57,950	8
Kentucky	804,139	3,471,063	1,014,190	18.8	161,324	15.9
Louisiana	889,570	3,625,055	1,114,022	19.7	231,234	20.8
Maine	174,405	1,120,041	259,501	13.5	13,351	5.1
Maryland	576,835	5,242,728	1,347,810	9.9	64,433	4.8
Massachusetts	740,836	5,765,193	1,390,552	11.4	108,138	7.8
Michigan	1,575,066	8,108,799	2,227,763	16.3	376,431	16.9
Minnesota	577,196	4,749,823	1,282,098	10.8	69,041	5.4
Mississippi	645,553	2,247,538	732,235	22.3	189,066	25.8
Missouri	897,755	4,978,611	1,395,124	15.3	138,044	9.9
Montana	148,677	849,637	225,020	14.9	19,906	8.8
Nebraska	227,021	1,600,170	467,601	12.4	35,743	7.6
Nevada	417,257	2,381,286	664,632	14.9	76,599	11.5
New Hampshire	109,690	1,175,747	266,979	8.5	5,255	2
New Jersey	949,341	7,789,717	2,009,813	10.9	183,102	9.1
New Mexico	426,814	1,615,200	501,750	20.9	112,317	22.4
New York	2,967,564	16,218,498	4,226,409	15.5	765,921	18.1
North Carolina	1,631,704	8,053,807	2,287,826	16.8	287,088	12.5
North Dakota	79,314	631,041	167,717	11.2	8,665	5.2
Ohio	1,732,839	9,534,661	2,639,860	15.4	351,173	13.3
Oklahoma	621,155	3,138,895	952,325	16.5	115,631	12.1
Oregon	614,223	3,291,163	861,395	15.7	56,196	6.5
Pennsylvania	1,647,762	10,721,909	2,704,268	13.3	325,988	12.1
Rhode Island	140,161	873,755	212,406	13.8	31,531	14.8
South Carolina	806,422	3,886,844	1,085,779	17.2	137,943	12.7
South Dakota	115,300	707,034	209,615	14	23,176	11.1
Tennessee	1,100,169	5,286,582	1,494,925	17.2	218,833	14.6
Texas	4,397,307	21,936,698	7,132,476	16.7	1,197,938	16.8
Utah	338,808	2,562,657	905,196	11.7	28,761	3.2
Vermont	69,673	531,865	121,691	11.6	1,168	1
Virginia	921,664	7,139,228	1,865,556	11.4	91,426	4.9
Washington	883,256	6,056,366	1,604,595	12.7	88,038	5.5
West Virginia	318,060	1,474,748	379,848	17.7	35,707	9.4
Wisconsin	713,472	4,889,802	1,301,498	12.7	112,203	8.6
Wyoming	65,762	503,181	138,844	11.6	1,767	1.3
Puerto Rico	1,577,075	1,919,653	767,406	45.1	647,275	84.3

Note: High-poverty is defined as a poverty rate at or above 30 percent. The U.S. totals exclude data for Puerto Rico.

Source: Author's analysis of data from the U.S. Census Bureau, American Community Survey Summary File.

6. ADDITIONAL RESOURCES

U.S. Census Bureau, What is the ACS?

<www.census.gov/programs-surveys/acs/about.html>

**U.S. Census Bureau, Understanding and Using American Community Survey Data:
What All Data Users Need to Know**

<www.census.gov/programs-surveys/acs/guidance/handbooks/general.html>

U.S. Census Bureau, ACS Data Releases

<www.census.gov/programs-surveys/acs/news/data-releases.html>

U.S. Census Bureau, Geography and ACS

<www.census.gov/programs-surveys/acs/geography-accs.html>

U.S. Census Bureau, ACS Data Tables and Tools

<www.census.gov/acs/www/data/data-tables-and-tools/>

U.S. Census Bureau, Data.census.gov: Census Bureau's New Data Dissemination Platform Frequently Asked Questions and Release Notes

<<https://data.census.gov/assets/releasenotes/faqs-release-notes.pdf>>

U.S. Census Bureau, Public Use Microdata Sample (PUMS) Documentation

<www.census.gov/programs-surveys/acs/technical-documentation/pums.html>

U.S. Census Bureau, American Community Survey Design and Methodology Report

<www.census.gov/programs-surveys/acs/methodology/design-and-methodology.html>