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.

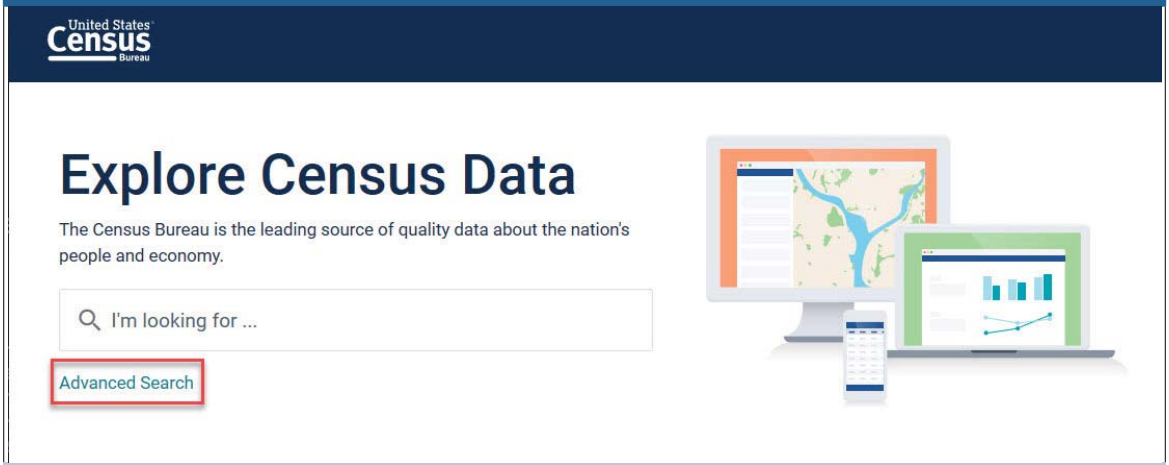
Data.census.gov provides access to ACS data for a wide range of geographic areas including states, cities, counties, census tracts, and block groups.[10] Researchers can access detailed ACS tables by using the "Advanced Search" feature, which allows users to conduct keyword searches or search by using predefined 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>.

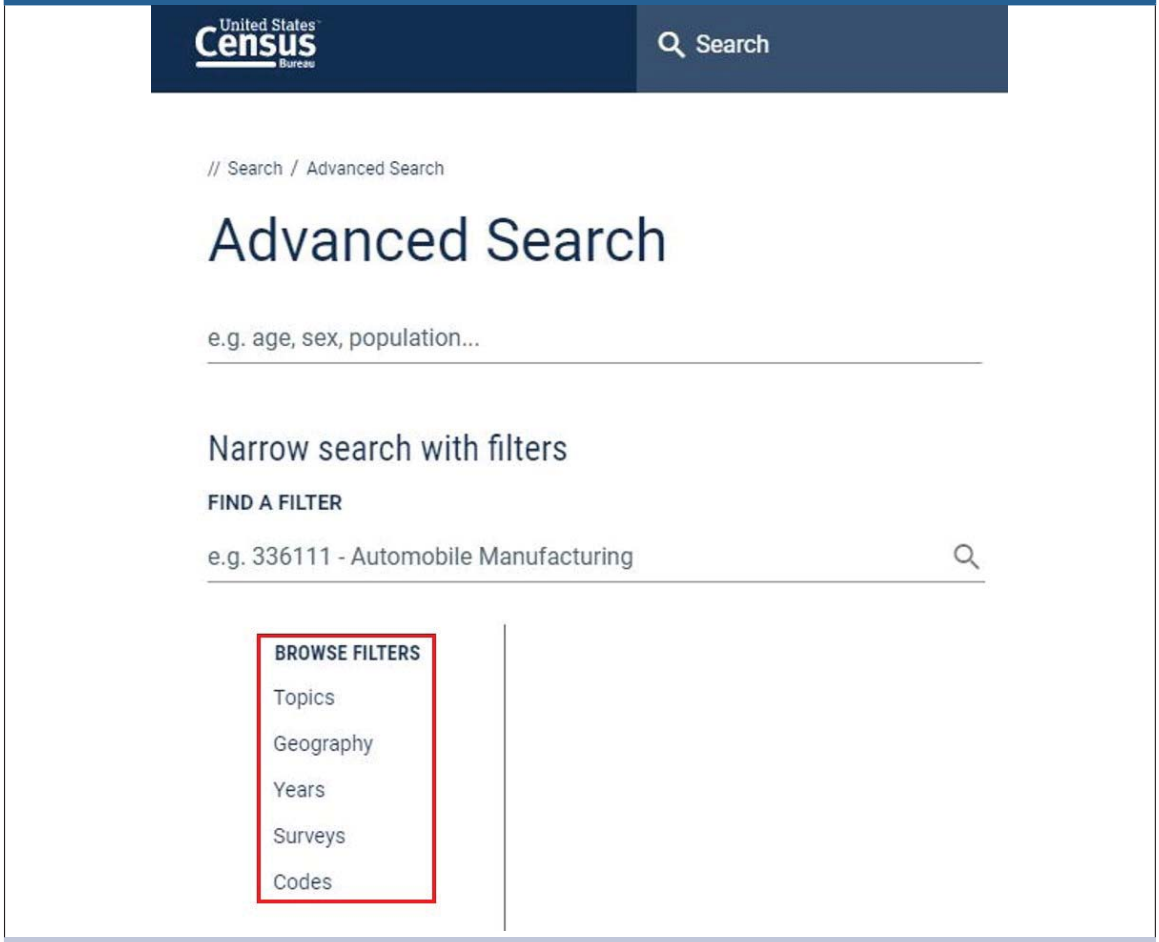
[10] U.S. Census Bureau, <https://data.census.gov>.

Figure 2.1. **Advanced Search in Data.census.gov**



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

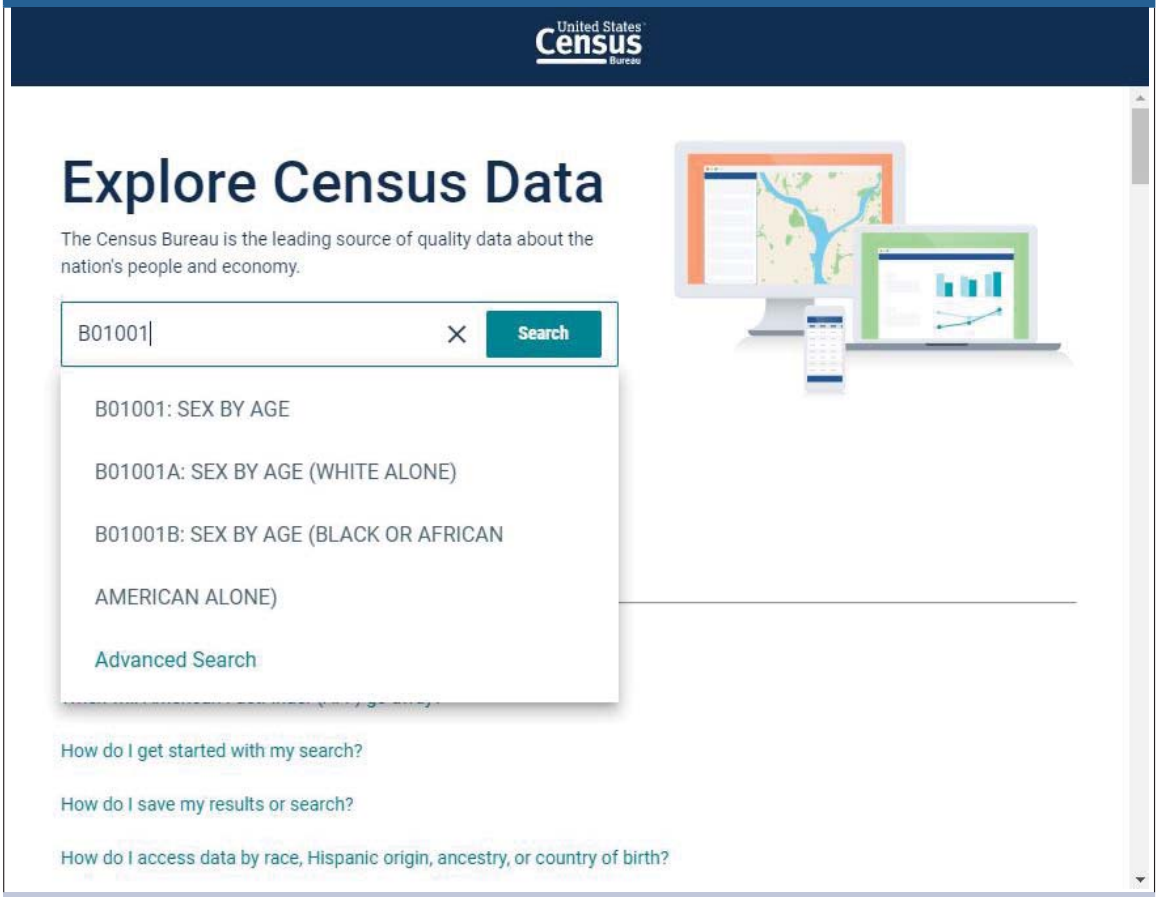
Figure 2.2. **Advanced Search Filters in Data.census.gov**



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**



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

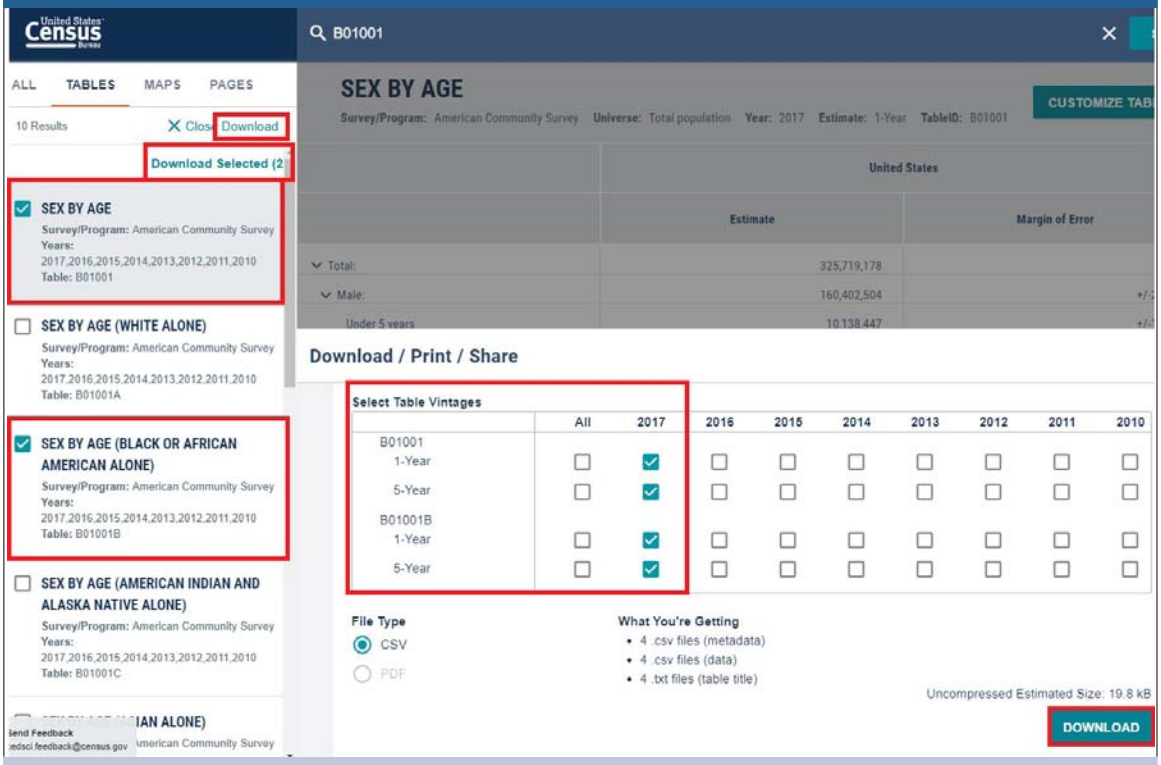
Data users can also use data.census.gov to download multiple tables simultaneously (see Figure 2.4). These tables can be downloaded in either comma-separated values (CSV) or PDF format. After navigating to a list of relevant tables:

- 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**



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

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 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.

[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>.

[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.

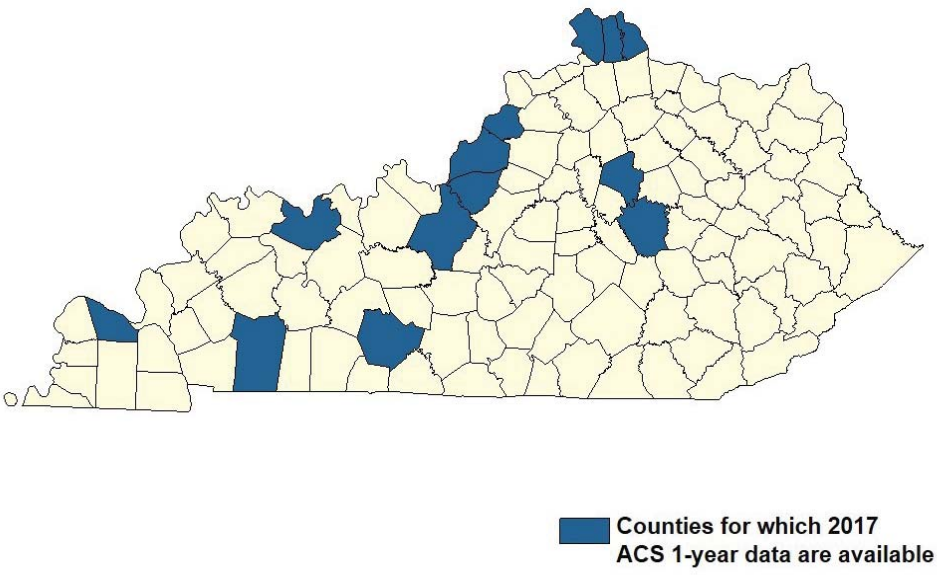
TIP: ACS PUMS data are not designed for statistical analysis of small geographic areas. The Census Bureau restricts the availability of information in microdata files that could be used to identify a specific housing unit or person, including detailed geographic information. Thus, the smallest geographic 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.

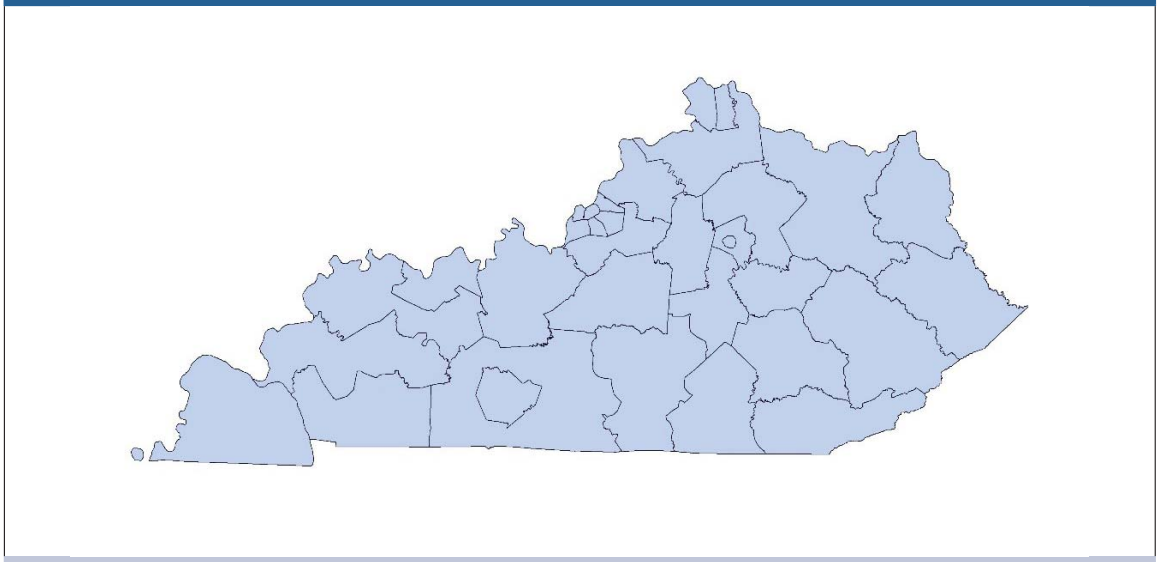
The ACS PUMS files include separate records for housing units and population. The housing unit records have unique identifiers that are repeated on each of the population records for people living in that housing unit. In this manner, housing unit characteristics can be merged with population records 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**



Source: Population Reference Bureau analysis of data from the U.S. Census Bureau, 2017 American Community Survey.

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

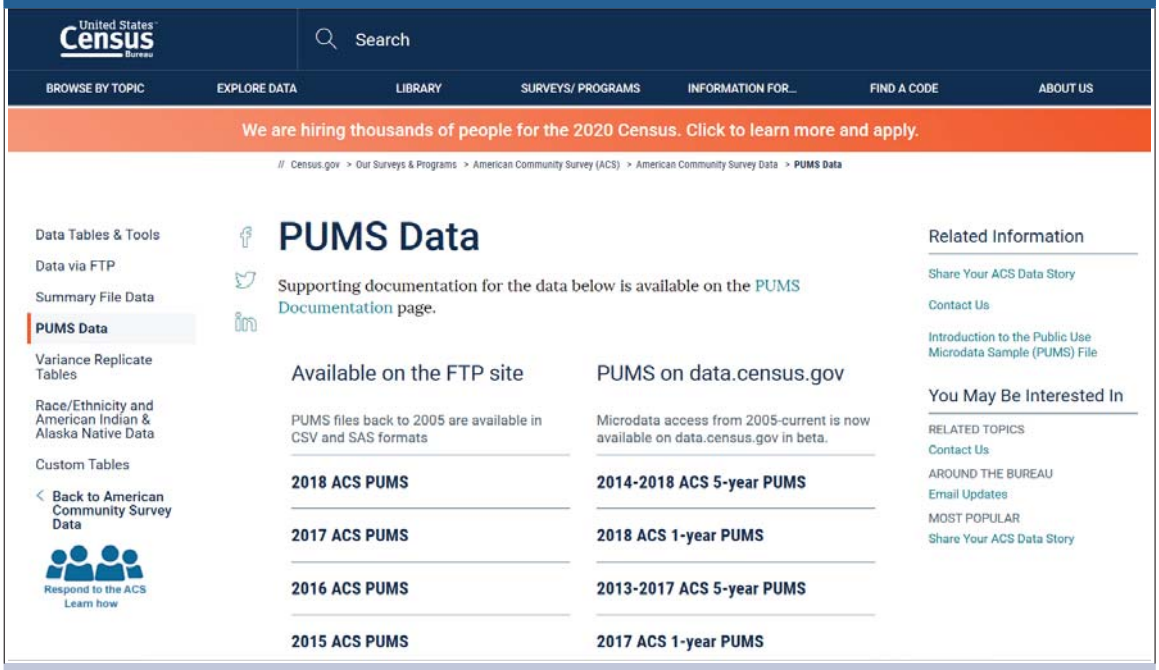
The FSRDCs are partnerships between federal statistical agencies and leading research institutions.[18] FSRDCs are secure facilities managed by the Census Bureau that provide secure access to a range of restricted-use microdata including ACS microdata. Unlike the ACS PUMS, which includes a representative 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**



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 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]

[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>.

[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>.