

1. ACS PUMS FILES: THE BASICS

The American Community Survey (ACS) microdata consist of individual records with information about the characteristics of each person and housing unit in the survey. The ACS Public Use Microdata Sample (PUMS) includes a subsample of the ACS microdata, devoid of personalized information. The PUMS represents about two-thirds of the responses collected in the ACS in a specific 1-year or 5-year period.

The U.S. Census Bureau produces ACS 1-year and 5-year PUMS files and typically releases these files 1 month after the release of the published ACS tables. The 5-year PUMS file is a combination of five 1-year PUMS files. The 1-year PUMS file includes records for about 1 percent of the total population and the 5-year file includes records for about 5 percent of the total population.

TIP: While PUMS data allow for more detailed and complex research techniques, the files are more difficult to work with than published tables. Data users need to use statistical software, such as SPSS, SAS, R, or Stata, to process PUMS data, and the responsibility for producing estimates from PUMS and judging their statistical significance is up to the data user.

There are two types of PUMS files, one for persons and one for housing units. The person-level file includes records for people, including those who live in group quarter facilities such as nursing homes or college dorms. The housing-level files include records pertaining to housing units, including vacant units.

The ACS PUMS is a weighted sample, and weighting variables must be used to generate estimates and standard errors that represent the population. The PUMS files include both population weights and household weights. Population weights should be used to generate statistics about individuals, and household weights should be used to generate statistics about housing units. (See the section on “Preparing ACS PUMS Data Files for Analysis” for more information.)

Protecting Confidentiality in the ACS PUMS

Title 13 legally requires the Census Bureau to keep all personal information strictly confidential.² Examples of measures taken to protect confidentiality in the PUMS include:

- Using only a subset of the full ACS sample to create the PUMS.
- Excluding names, addresses, and any information that could be used to identify a specific housing unit, group quarters unit, or person.
- “Swapping” (or exchanging) a small number of records with similar records from neighboring areas.
- Top-coding or bottom-coding answers to selected variables. Top-coding and bottom-coding involves truncating extreme values for certain variables. A list of top-coded and bottom-coded values is available on the Census Bureau’s PUMS Documentation Web page.³
- Limiting the geographic areas that can be identified in the PUMS. Data are available for the nation, regions, divisions, states, and Public Use Microdata Areas (PUMAs). The section on “Public Use Microdata Areas” provides more information.

TIP: Due to confidentiality protections and the fact that PUMS files are based on only about two-thirds of the ACS sample, estimates using the ACS PUMS may differ from estimates provided through the ACS Summary File or other published Census Bureau tables and profiles. You can verify that you have correctly accessed and tabulated data from the ACS PUMS file by replicating the values presented in “PUMS Estimates for User Verification” in the PUMS Technical Documentation.⁴

² U.S. Census Bureau, Data Protection and Privacy, Title 13 - Protection of Confidential Information, <www.census.gov/about/policies/privacy/data_stewardship/title_13_-_protection_of_confidential_information.html>.

³ U.S. Census Bureau, American Community Survey (ACS), PUMS Documentation, <www.census.gov/programs-surveys/acs/microdata/documentation.html>.

⁴ U.S. Census Bureau, American Community Survey (ACS), PUMS Documentation, <www.census.gov/programs-surveys/acs/microdata/documentation.html>.

Summary Data Versus Microdata

Summary data (also called aggregate or pretabulated data) are predefined weighted tabulations of person or housing characteristics. ACS summary tabulations are typically presented in tables, data profiles, and maps. Figure 1.1 provides a portion of a summary table on the means of transportation to work using the ACS 5-year estimates. In 2012–2016, 24,183 workers in Pittsburgh, Pennsylvania, got to work by bus or trolley.

The basic unit of analysis for summary data is a specific geographic entity such as a state, county, or place (including cities and towns). Estimates in summary tables can be added (or subtracted), but users are constrained by the categories that are available.

The benefit of using summary data is that estimates of population and housing characteristics are provided for geographic areas as small as census tracts (small

subdivisions of counties that typically have between 1,200 and 8,000 residents) and block groups (subdivisions of census tracts that typically have between 600 and 3,000 residents). Summary data in published tables are user-friendly because they are produced by Census Bureau analysts, informed by data user input, and usually include margins of error. A drawback to summary data is that the user is bound to a predetermined set of tabulations with fixed variable categories.

By contrast, the PUMS contains individual responses to the full range of topics on the ACS (without individually identifiable information). Table 1.1 provides an example of ACS 5-year PUMS microdata for a person in the state of Pennsylvania. The 2012–2016 ACS 5-year PUMS Data Dictionary provides information about each of the values in the table.⁵

⁵ U.S. Census Bureau, American Community Survey (ACS), PUMS Documentation, <www.census.gov/programs-surveys/acs/microdata/documentation.html>.

Figure 1.1. **Published Summary Data for Means of Transportation to Work in Pittsburgh, PA: 2012–2016**

Pittsburgh city, Pennsylvania		
	Estimate	Margin of Error
▼ Total:	148,176	+/-1,618
▼ Car, truck, or van:	96,009	+/-1,476
Drove alone	82,905	+/-1,571
▼ Carpooled:	13,104	+/-820
In 2-person carpool	11,234	+/-773
In 3-person carpool	1,270	+/-197
In 4-person carpool	401	+/-126
In 5- or 6-person carpool	146	+/-77
In 7-or-more-person carpool	53	+/-40
▼ Public transportation (excluding taxicab):	25,278	+/-983
Bus or trolley bus	24,183	+/-991
Streetcar or trolley car (carro publico ...	517	+/-147

Source: U.S. Census Bureau, <<https://data.census.gov>>, Table B08301: Means of Transportation to Work.

Table 1.1. Example of Microdata

RT	SERIALNO	SPORDER	ST	AGEP	JWTR	SEX
P	2016000009344	02	42	49	.	2
P	2016000009344	03	42	15	.	2
P	2016000009578	02	42	27	01	1
P	2016000009578	03	42	28	01	1
P	2016000009578	04	42	29	01	1
P	2016000009578	05	42	25	01	1
P	2016000021254	02	42	25	10	2
P	2016000024874	02	42	72	.	2
P	2016000025941	02	42	22	02	2
P	2016000025941	03	42	14	.	1
P	2016000025941	04	42	71	.	2

Source: U.S. Census Bureau, 2012–2016 American Community Survey Estimates, 5-Year Public Use Microdata Sample File.

In the highlighted row in Table 1.1:

- The variable for Record Type (RT) is “P,” indicating that this record comes from the PUMS person file.
- The SERIALNO (“2016000025941”) is a unique identifier for the housing unit.
- SPORDER (“02”) is a unique identifier of persons within a housing unit.
- ST is “42,” indicating the individual lives in Pennsylvania.
- AGEP is “22,” indicating the person was 22 years old at the time of the survey.
- JWTR, or Means of Transportation to Work, is “02” indicating the person commutes to work by bus or trolley.
- SEX is “2” indicating the person is female.

You can find values for all PUMS variables in the PUMS Data Dictionary.⁶ Here, missing values are shown as periods. A missing numeric value or a blank character value may be displayed differently depending on the software used. A missing value indicates that the person recode was not in the universe for the relevant variable. For example, as shown in the data dictionary, JWTR values are not available for respondents who are not workers.

⁶ U.S. Census Bureau, American Community Survey (ACS), PUMS Documentation, <www.census.gov/programs-surveys/acs/microdata/documentation.html>.

There are approximately 250 person-level variables and 200 housing-level variables available in the ACS PUMS. The PUMS files permit data users to analyze specific population groups and create custom variables that are not available through published tables. PUMS files allow more flexibility because data users can focus in on small population groups as well as investigate the relationship among survey questionnaire items by defining and creating unique combinations of person or household variables. Some examples include:

- Estimating the population living below a specified income-to-poverty ratio (for example, children living in families with income below 185 percent of the poverty threshold).
- Investigating the income and poverty status of Gulf War veterans.
- 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.
- Analyzing data for more detailed languages than those available through published tables in <<https://data.census.gov>>.