AMERICAN COMMUNITY SURVEY
2017 ACS 1-YEAR PUMS FILES
ReadMe

Prepared by
American Community Survey Office
U.S. Census Bureau
October 18, 2018

TABLE OF CONTENTS

I. OVERVIEW OF THE PUBLIC USE MICRODATA SAMPLE (PUMS)

The Public Use Microdata Sample (PUMS) contains a sample of actual responses to the American Community Survey (ACS). The PUMS dataset includes variables for nearly every question on the survey, as well as many new variables that were derived after the fact from multiple survey responses (such as poverty status). Each record in the file represents a single person, or, in the household-level dataset, a single housing unit. In the person-level file, individuals are organized into households, making possible the study of people within the contexts of their families and other household members. PUMS files for an individual year, such as 2017, contain data on approximately one percent of the United States population. PUMS files covering a five-year period, such as 2013-2017, contain data on approximately five percent of the United States population.

The PUMS files are much more flexible than the aggregate data available on American FactFinder, though the PUMS also tend to be more complicated to use. Working with PUMS data generally involves downloading large datasets onto a local computer and analyzing the data using statistical software such as R, SPSS, Stata, or SAS.

Since all ACS responses are strictly confidential, many variables in the PUMS files have been modified in order to protect the confidentiality of survey respondents. For instance, particularly high incomes are "top-coded," and uncommon birthplace or ancestry responses are grouped into broader categories. The PUMS files also provide a very limited set of geographic variables (explained more below).

II. PUBLIC USE MICRODATA AREAS (PUMAS)

While PUMS files contain cases from nearly every town and county in the country, towns and counties (and other low-level geography) are not identified by any variables in the PUMS

datasets. The most detailed unit of geography contained in the PUMS files is the Public Use Microdata Area (PUMA).

PUMAs are special non-overlapping areas that partition each state into contiguous geographic units containing no fewer than 100,000 people each. The 2017 ACS PUMS files rely on PUMA boundaries that were drawn by state governments after the 2010 Census. An interactive mapping application, TIGERweb, can be used to view PUMA boundaries from 2010. TIGERweb is available from the Census Bureau's web site at https://tigerweb.geo.census.gov/tigerwebmain/tigerweb_main.html.

To access the maps:

- Click on "TIGERweb Applications" on the upper left.
- Click "TIGERweb" on the left column.
- On the upper left, click on the circle to open/close menus.
- On the upper left, you should see choices: "Layers", "Legend", or "Task Results." Select "Layers."
- Under "Select Vintage:" choose "ACS 2017."
- Select "PUMAs, UGAs, and ZCTAs" on the left.
- Expand the "PUMAs, UGAs, and ZCTAs" box to see the choices: "2010 Census Public Use Microdata Areas" and "2010 Census ZIP Code Tabulation Areas.
- Click on the map to zoom, or move the zoom scale bar to zoom in closer to the map.
- When color fills the checked boxes next to "2010 Census ZIP Code Tabulation Areas" and "2010 Census Public Use Microdata Areas," deselect "2010 Census ZIP Code Tabulation Areas."

There are two additional resources that may help PUMS users understand and use PUMAs. They are the software MABLE, and the static maps published by the Census Bureau.

MABLE Software

The software MABLE was developed by the Missouri Census Data Center. It is found at: http://mcdc.missouri.edu/websas/geocorr12.html. The software allows data users to calculate the proportion of a PUMA's population that is within a county or other geographic area and then supplies you with the relevant PUMA codes.

Census Bureau Static Maps for PUMAs

Static maps for PUMAs may be found at: https://www.census.gov/geo/maps-data/maps/reference.html

To Access the static maps:

- Click on "Public User Microdata Areas (PUMAs)" in the middle of the screen.
- The section will expand to show three options. Choose the first option labelled "2010 Census Public Use Microdata Area (PUMA) Reference Maps".
- The section will expand again. Choose the state you are interested in from the drop down menu and click on "Go".
- A new webpage with the list of static maps for the relevant state will appear. Choose the geographic area of interest to see the relevant static PUMA map.

III. PUMS DOCUMENTATION

PUMS Introduction to the Public Use Microdata Sample (PUMS) File Webinar

Data users new to PUMS may find the Introduction to PUMS webinar to be a useful reference. The webinar may be found at: https://www.census.gov/programs-surveys/acs/guidance/training-presentations/acs-intro-pums.html.

PUMS Technical Documentation

The PUMS Technical Documentation is available at: https://www.census.gov/programs-surveys/acs/technical-documentation/pums/documentation.html. This page includes the following documents:

Subjects in the PUMS
This document provides a list of subjects available in the PUMS data. PUMS data users may also consult the ACS Subject Definitions located on the ACS Technical Documentation page, located at: https://www.census.gov/programs-surveys/acs/technical-documentation/code-lists.html.

PUMS Data Dictionary
The PUMS Data Dictionary provides PUMS variables, their values and the meaning of each of the values. Data users may access the Data Dictionary in three formats. There is a pdf version, a text version, and new for this year is a machine-readable comma-separated value (CSV) version.

The text version has been reorganized and formatted to make it easier for data users to read it into their program. The CSV version is also designed to be machine readable. The Census Bureau welcomes data users comments and suggestions on how to improve format and layout of the Data Dictionary.

PUMS Code Lists
The PUMS Code Lists provide categories for variables with a large number of coded responses such as ancestry or occupation. They provide the PUMS categories for the detailed codes. This allows data users to more clearly understand how each code maps to the broad categories used in PUMS.

PUMS Top Coded and Bottom Coded Values
This document contains the threshold value above which select variables are top-coded and the top-code value. Top-code threshold variables end in “TPCT” and the top-code value variables begin with “T_”. In addition, two variables are bottom-coded. The threshold value below which variables are bottom-coded end in “BPCT”, and the variables for the bottom-coded values begin with “B_”.

Accuracy of the PUMS
The Accuracy of the PUMS document provides descriptions of the sampling methodology, weighting methodology, confidentiality, and how to calculate SE using the design factor methodology.

PUMS Estimates for User Verification
PUMS estimates for selected housing unit and population characteristics are included so that data users may verify they are using PUMS data correctly. The estimates are weighted to represent the population. In addition, standard errors (SE) and margins of error (MOE) are also included for verification purposes. The standard errors provided in this document were computed using the successive difference replication (SDR) methodology, which uses PUMS replicate weights. See Section on PUMS Weights below for more information.

Unweighted Count of PUMS Records
Data users may confirm that they have all records for PUMS by examining the unweighted count of PUMS records. Previously, this information was provided at the national level in the introduction to the PUMS Accuracy of the Data document.

IV. OBTAINING PUMS DATA

PUMS data may be obtained in the following ways:

ACS Website
PUMS files can be accessed via the ACS website at https://www.census.gov/programs-surveys/acs/data/pums.html.

American FactFinder
PUMS Files are also accessible via American FactFinder at https://factfinder.census.gov.

DataFerrett
It is also possible to get PUMS data from the Census Bureau's DataFerrett, which has the additional feature of being able to make tables and perform basic analysis online. This tool is particularly useful for researchers who need a quick statistic or do not have access to statistical software. DataFerrett is available at https://www.census.gov/programs-surveys/acs/technical-documentation/pums/dataferrett.html.

V. PUMS FILE STRUCTURE

The ACS questionnaire contains "household" items that are the same for all members of the household (such as the number of rooms in the home) and "person" items that are unique for each household member (such as age, sex, and race). The ACS PUMS files are made available in this same structure. Researchers who are analyzing only household-level items can use the housing unit records files, whereas those using only person-level variables can use the population records files.

The population records files also contain records for persons in group quarter facilities (such as nursing homes or college dorms). The housing unit files contain place holder records for group quarters. The majority of the variables for housing unit records for group quarters are blank. The weights and replicate weights are zero. The group quarter place holder records exist so that data

users may obtain values for FS (Yearly food stamp/Supplemental Nutrition Assistance Program recipiency) for them.

Data users should note that PUMS files containing data for the entire United States (in contrast to individual state and state-equivalent files) are separated into multiple data files. These files contain unique records and must be concatenated in order to create a complete file for the nation. For example, users downloading the 2017 ACS 1-year PUMS files of United States Population Records will notice an "a" file and a "b" file. Each file contains about half the population records in the 2017 1-year PUMS dataset of the United States.

Basic Example of Combining PUMS Person and Housing Files
Below are instructions for concatenating the two PUMS person-level files, in the form of an italicized SAS program and pseudo-code.

Concatenate the two **person-level** files using the set statement:

```
data population;
   set psam_pusa psam_pusb;
run;
```

The 2017 ACS 1-year PUMS files of the United States Housing Records also contain an "a" file and a "b" file upon downloading. To create a complete housing-level file, the two files must be concatenated. Below are instructions for concatenating the two PUMS household-level files, in the form of an italicized SAS program and pseudo-code.

Concatenate the two **household-level** files using the set statement:

```
data housing;
   set psam_husa psam_husb;
run;
```

Some data users will need to use household and person items together. For instance, in order to analyze how the number of rooms in a home varies by a person's age, the merging of the household and person files will be required.

This merger must rely on the SERIALNO variable, which is the same in the household and person files. Below are instructions for merging the housing and population PUMS files, in the form of an italicized SAS program and pseudo-code.

1. First make sure the files are sorted by SERIALNO.

```
proc sort data=population;
   by serialno;
run;
```

```
proc sort data=housing;
   by serialno;
run;
```

2. Then merge the two files together using SERIALNO as a merge key. Note that in SAS, the 'in=' option will allow you to identify records from a specific file. The line 'if pop' retains only records from the population file.

```
data combined;
   merge population (in=pop) housing;
     by serialno;
   if pop;
run;
```

You should not merge the files unless the estimates you want require a merge. Note that there are many estimates that may be tabulated from the person file and from the household file without any merging. The suggested merge will create a person level file, so that the estimate of persons may be tallied within categories from the household file and the person weights should be used for such tallies.

Note also that the housing unit record files contain vacant housing units. There are no population records for these housing units. Please note that housing characteristics cannot be tallied from this merged file without extra steps to ensure that each housing weight is counted only once per household.

VI. WEIGHTS IN THE PUMS FILES

The ACS PUMS is a weighted sample, and weighting variables must be used to generate accurate estimates and standard errors. 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. The weighting variables are described briefly below.

PWGTP: Person's weight for generating statistics on individuals (such as age).

WGTP: Household weight for generating statistics on housing units and households (such as average household income).

WGTP1-WGTP80 and **PWGTP1-PWGTP80**: Replicate weighting variables, used for generating the most accurate standard errors for households or individuals.

PWGTP and WGTP may be used both to generate the point estimates and to generate standard errors when using a generalized formula. Replicate weights may only be used to calculate "direct standard errors." Direct standard errors are expected to be more accurate than generalized standard errors, although they may be more inconvenient for some users to calculate. Both

generalized and direct standard errors are explained in more detail in the Accuracy of the PUMS document. To find this document, go to https://www.census.gov/programs-surveys/acs/technical-documentation/pums/documentation.html.

Each housing unit and person record contains 80 replicate weights. To use the replicate weights to calculate an estimate of the direct standard error, first form the estimate using the full PUMS weight. Then form the estimate using each of the 80 replicate weights--providing both the full PUMS estimate and 80 replicate estimates. These estimates are then entered into the following formula, which is explained in more detail in the "Accuracy of the PUMS" document:

$$\mathrm{SE}(x) = \sqrt{\frac{4}{80}\sum_{r=1}^{80}(x_r - x_0)^2}$$

Where Xr is a replicate estimate from X1 to X80, and X is the full PUMS weighted estimate.

Data users who wish to learn more may consult the Variance Replicate Tables (VRT) Documentation, located at: https://www.census.gov/programs-surveys/acs/technical-documentation/variance-tables.html. Although the VRT documentation pertains to ACS data, the concepts provided in the documentation may be adopted for use with PUMS data.

The technical explanation of the ACS replicate weights is in Chapter 12 of the Design and Methodology document found at: https://www.census.gov/programs-surveys/acs/methodology/design-and-methodology.html.

For more information on the theoretical basis, please reference: Fay, R. and Train, G. (1995), "Aspects of Survey and Model-Based Postcensal Estimation of Income and Poverty Characteristics for States and Counties," Proceedings of the Section on Government Statistics, American Statistical Association, pp. 154-159, 1995.

Please note that many estimates generated with PUMS will be slightly different from estimates for the same characteristics published in American FactFinder. These differences are due to the fact that the PUMS files include only about two-thirds of the cases that were used to produce estimates on American FactFinder, as well as additional PUMS edits. More information on the PUMS sample design is available in the "Accuracy of the PUMS" document.

VII. VARIABLE CHANGES IN THE 2017 1-YEAR PUMS FILES

The 2017 1-year PUMS includes most of the variables that were included in the 2016 1-year PUMS.

There were also some variables with new codes, modified codes, or cosmetic changes to variable labels or value labels. See the 2017 1-year PUMS Data Dictionary for a complete listing of the variables and values contained in the 2017 1-year PUMS data file.

If PUMS users want to compare 2017 PUMS estimates to 2016 PUMS, or want to reuse a program written for 2016 PUMS data, the following table provides some information about variables with changes between 2016 and 2017. For more details on the changes, see both the 2016 and 2017 1-year PUMS Data Dictionaries and Code Lists from the PUMS Technical Documentation page found at https://www.census.gov/programs-surveys/acs/technical-documentation/pums/documentation.html.

Variables added since previous 1-year PUMS: None.

Variables deleted since previous 1-year PUMS: None.

Variables with new or modified codes since the previous 1-year PUMS: SERIALNO, ANC1P, ANC2P, CITWP, MIGSP, POBP, DECADE, MARHYP, YOEP, ADJINC, CITWP, YBL, FINCP, GRNTP, HINCP, PAP, PERNP, PINCP, SSIP, SSP.

Variables with cosmetic changes to variable labels or value labels: ADJHSG, AGS, CONP, ELEP, FFSP, FLAPTOPP, FRWATPRP, FS, FULP, GASP, HHT, INDP, INSP, INTP, MHP, MRGP, NP, NWAB, RETP, NWAV, NWLA, NWLK, NWRE, OIP, PUMA, RNTP, SEMP, SFR, SMOCP, SMP, TAXP, WAGP, WAOB, WATP, WGTP.

Detailed Changes by Variable (as compared to 2016 ACS 1-year PUMS Data Dictionary)

Variable	Description of Change
	New Variables
-	There were no new variables added for 2017.
	Variables with Code Changes
SERIALNO	Range updated for data year 2017. Length of SERIALNO changed to match PUMS 5-year SERIALNO length and format (Year + SERIALNO). It now has a length of 13.
ANC1P	Categories updated for data disclosure requirements.
ANC2P	Categories updated for data disclosure requirements
CITWP	Categories updated for data disclosure requirements.
HHLANP	Categories updated for data disclosure requirements.
LANP	Categories updated for data disclosure requirements.
MIGSP	Categories updated for data disclosure requirements.
POBP	Categories updated for data disclosure requirements.
	Variables with Annual Code Updates and Code Changes
DECADE	Most current year (2017) added to the last category (2010 or later). Category of “2000 or later” split into “2000-2009” and “2010 and later”.
MARHYP	Most current year (2017) added. Categories for 1937 and earlier collapsed into a single category for data disclosure requirements.
YOEP	Most current year (2017) added. Categories updated for data disclosure requirements.

Detailed Changes by Variable (Continued)

Variable	Description of Change
	Variables with Annual Code Updates
ADJINC	Adjustment factor updated.
CITWP	Most current year (2017) added.
YBL	Most current year (2017) added.
	Variables with Description Changes Only
ACCESS	Clarified definition for blank values.
ADJHSG	Year in title changed to 2017.
AGS	Title clarified to indicate no inflation factors is applied.
BROADBND	Clarified definition for blank values.
COMPOTHX	Clarified definition for blank values.
CONP	Title clarified to indicate that variable requires inflation factor adjustment.
DIALUP	Clarified definition for blank values
ELEP	Title clarified to indicate that variable requires inflation factor adjustment.
FFSP	Acronym SNAP added to title.
FLAPTOPP	Notebook Computer added to title.
FRWATPRP	Puerto Rico added to title.
FS	Acronym SNAP added to title.
FULP	Title clarified to indicate that variable requires inflation factor adjustment.
GASP	Title clarified to indicate that variable requires inflation factor adjustment.
HHT	Clarification of descriptions for values.
HISPEED	Clarified definition for blank values.
INDP	Clarification of descriptions for values.
INSP	Title clarified to indicate that variable requires inflation factor adjustment.
INTP	Title clarified to indicate that variable requires inflation factor adjustment.
LAPTOP	Clarified definition for blank values.
MHP	Title clarified to indicate that variable requires inflation factor adjustment.
MRGP	Title clarified to indicate that variable requires inflation factor adjustment.
NP	Clarification of descriptions for values.
NWAB	Clarification of Title.
RETP	Title clarified to indicate that variable requires inflation factor adjustment.
NWAV	Clarification of Title.
NWLA	Clarification of Title.
NWLK	Clarification of Title.
NWRE	Clarification of Title.
OIP	Title clarified to indicate that variable requires inflation factor adjustment.
OTHSVCEX	Clarified definition for blank values.
PUMA	Clarification of Title.
RNTP	Title clarified to indicate that variable requires inflation factor adjustment.
SATELLITE	Clarified definition for blank values.
SEMP	Title clarified to indicate that variable requires inflation factor adjustment.
SFR	Value description clarified.
SMARTPHONE	Clarified definition for blank values.

Detailed Changes by Variable (Continued)

Variable	Description of Change
Variables with Description Changes Only	
SMOCP	Title clarified to indicate that variable requires inflation factor adjustment.
SMP	Title clarified to indicate that variable requires inflation factor adjustment.
TABLET	Clarified definition for blank values.
TAXP	Title clarified to indicate no inflation factors is applied.
WAGP	Title clarified to indicate that variable requires inflation factor adjustment.
WAOB	Value description clarified.
WATP	Title clarified to indicate that variable requires inflation factor adjustment.
WGTP	Value description clarified.
Variables with Description Changes and Code Changes	
FINCP	Title clarified to indicate that variable requires inflation factor adjustment. Variable length changed to 7.
GRNTP	Title clarified to indicate that variable requires inflation factor adjustment. Variable length changed to 4.
HINCP	Title clarified to indicate that variable requires inflation factor adjustment. Variable length changed to 7.
PAP	Title clarified to indicate that variable requires inflation factor adjustment. Highest value changed from $99,999 to $30,000.
PERNP	Title clarified to indicate that variable requires inflation factor adjustment. Titles of values clarified.
PINCP	Title clarified to indicate that variable requires inflation factor adjustment. Titles of values clarified.
SSIP	Title clarified to indicate that variable requires inflation factor adjustment. Titles of values clarified.
SSP	Title clarified to indicate that variable requires inflation factor adjustment. Titles of values clarified.
Variable with Name Changes Only	
-	None.

In 2017, there was a need to collapse the categories of certain variables due to data disclosure requirements. For reference, those variables are:

- ANC1P
- ANC2P
- CITWP
- HHLANP
- LANP
- MARHYP
- MIGSP
- POBP
- RAC2P
- RAC3P
- YOEP

VIII. ADDITIONAL INFORMATION

Note on ADJINC Variable:

Divide ADJINC by 1,000,000 to obtain the inflation adjustment factor and multiply it to the PUMS variable value to adjust it to 2017 dollars. Variables requiring ADJINC on the Housing Unit file are FINCP and HINCP. Variables requiring ADJINC on the Person files are: INTP, OIP, PAP, PERNP, PINCP, RETP, SEMP, SSIP, SSP, and WAGP.

Note on ADJHSG Variable:

Divide ADJHSG by 1,000,000 to obtain the inflation adjustment factor and multiply it to the PUMS variable value to adjust it to 2017 dollars. Variables requiring ADJHSG on the Housing Unit files are: CONP, ELEP, FULP, GASP, GRNTP, INSP, MHP, MRGP, SMOCP, RNTP, SMP, and WATP.

Note that ADJHSG does not apply to AGS and TAXP because they are categorical variables that should not be inflation-adjusted. In addition, ADJHSG does not apply to VALP.

Note on PUMA Variable:

Public use microdata areas (PUMAs) are the smallest geography available on the PUMS. They are designed to have a population of 100,000 or more people. In order to uniquely identify a PUMA code it must be paired with the state FIPS code (ST). PUMA codes within a state are unique. However, multiple PUMAs in different states may share the state PUMA code.

Note on SOCP:

In cases where the SOC occupation code ends in X(s) or Y(s), two or more SOC occupation codes were aggregated to correspond to a specific Census occupation code. In these cases, the Census occupation description is used for the SOC occupation title.

These codes are pseudo codes developed by the Census Bureau and are not official or equivalent NAICS or SOC codes.

Codes to Identify NAICS Equivalents:
M = Multiple NAICS codes
P = Part of a NAICS code - NAICS code split between two or more Census codes
S = Not specified Industry in NAICS sector - Specific to Census codes only
Z = Exception to NAICS code - Part of NAICS industry but has a unique Census code

Additional information on NAICS and SOC groupings within major categories may be found at: https://www.census.gov/topics/employment/industry-occupation/guidance/indexes.html

Note on PUMS File Names for CSV Files

Data users may download PUMS data in either a CSV file or as a SAS file. Beginning with 2017 data, the CSV file will have the same name as the SAS file. For Person-level files, the name is "PSAM_P<ST>" and for Housing-level files, the name is "PSAM_H<ST>". Here, <ST> is the State FIPS code.

State names, abbreviations and FIPS codes may be found here: https://www.census.gov/geo/reference/ansi_statetables.html. Choose the "FIPS Codes for the States and the District of Columbia". For example, for Connecticut, <ST> is "09".

Note that for the National files, there are two files, an "A" and a "B" file. For Person-level data the names are "PSAM_PUSA" and "PSAM_PUSB". The Housing-level files are "PSAM_HUSA" and "PSAM_HUSB".

The "A" files contain data for Alabama (FIPS code "01") through Mississippi (FIPS Code "28"). The "B" files contain data for Missouri (FIPS code "29") through Wyoming (FIPS Code "56").

Puerto Rico data is not included in the national files. It is published as a state equivalent and has a State FIPS code of "72".

Note on the PUMS Data Dictionary Name

Please note that the name of the PUMS Data Dictionary has been changed to be more descriptive. It now has the form "PUMS_Data_Dictionary_<YYYY>". Where <YYYY> is the 4-digit year. For example, the pdf version is called: "PUMS_Data_Dictionary_2017.pdf".

Previous PUMS 1-year Data Dictionaries had the name "PUMSDataDict<YY>", where <YY> was the 2-digit. For example, for 2016, the name for the pdf version was PUMSDataDict16.pdf.

Additional Notes:

The Census Bureau occasionally provides corrections or updates to PUMS files. Data users may sign up for notifications and updates via the Census Bureau's E-mail Updates system at: https://service.govdelivery.com/accounts/USCENSUS/subscriber/new?category_id=USCENSUS_C12.

In addition, PUMS errata notes may be found here: https://www.census.gov/programs-surveys/acs/technical-documentation/errata.html

Data users may also email acso.users.support@census.gov with any PUMS-related questions.