AMERICAN COMMUNITY SURVEY
2013-2017 ACS 5-YEAR PUMS FILES
ReadMe

Prepared by
American Community Survey Office
U.S. Census Bureau
January 17, 2019

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 2013-2017 ACS PUMS 5-year 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 website 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.

A. 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.

B. 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 DOCUMENATION

A. 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.

B. 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:

1. **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.

2. **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.

3. **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.

4. **PUMS Top Coded and Bottom Coded Values**

To protect confidentiality, selected variables are top-coded. A process identifies records which exceed the specific top-code value. These records are replaced with the mean value of all records that are equal to or greater than the top-code value. This is done by state.

The PUMS Top Coded and Bottom Coded Values document contains the threshold value above which select variables are top-coded and the top-code value. Top-code threshold variables end in “TPCT”. Beginning in 2017, the top-code value variables begin with “T_”. Prior to 2017, the top-code value variables had the same name as the PUMS variable that was top-coded.

In addition, two variables are bottom-coded. Bottom-coding is similar to top-coding. It identifies records which are below the bottom-coded value and replaces them with the mean value of all records less than or equal to the bottom-coded threshold.

The threshold value below which variables are bottom-coded end in “BPCT”. Similarly to the top-codes, beginning in 2017 the variables for the bottom-coded values begin with “B_”. For 2016 and earlier, the bottom-coded variables are named “BINT” and “BSEM”.

5. **Accuracy of the PUMS**

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

6. **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 weights 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.

Note that the text version of the file (ending in "[LST]") is no longer published. The CSV and SAS versions are still available.

7. **PUMS Record Count**

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. CHANGES AND UPDATES TO METADATA FILES

A. Addition of PUMS Variable Type

The organization of the PUMS Data Dictionary has been updated. The first change is the addition of the variable type. That is, whether a variable is character or numeric. In the pdf, this is shown between the PUMS variable name and PUMS variable length. For example, the variable for Record Type (RT) is a character variable.

Example of PUMS Data Dictionary Variable Layout

```
RT          Character   1
      Record Type
            H .Housing Record or Group Quarters Unit
            P .Person Record
```

B. Organization of PUMS Variables in the Data Dictionary

Secondly, the PUMS variables are placed into groups. The Housing variables come first, followed by the Person variables. The variables are further divided into categories which are listed below.

Major Variable Organizational Categories

Data Dictionary Section	Description
HOUSING RECORD	
BASIC VARIABLES	Basic variables, such as geographic variables and inflation adjustment variables.
HOUSING UNIT VARIABLES	Housing variables pertaining to the Housing Unit
HOUSEHOLD VARIABLES	Housing variables pertaining to the Household
ALLOCATION FLAGS	Housing allocation flag variables
REPLICATE WEIGHTS	Housing replicate weight variables used for variance calculation
PERSON RECORD	
BASIC VARIABLES	Basic variables, such as geographic variables and inflation adjustment variables.
PERSON VARIABLES	Person Variables
RECODED PERSON VARIABLES	PUMS Person Variables created from other Variables.
ALLOCATION FLAGS	Person allocation flag variables
REPLICATE WEIGHTS	Person replicate weight variables used for variance calculation

C. Note on Data Dictionary CSV File

Beginning with 2017, the PUMS Data Dictionary is also available as a comma separated value (CSV) file. The intention is to provide a file that is machine readable. Therefore, only variables and their legal values are present in the file.

Example of Data Dictionary from CSV file

```
NAME,RT,C,1,"Record Type"
VAL,RT,C,1,"H","H","Housing Record or Group Quarters Unit"
VAL,RT,C,1,"P","P","Person Record"
NAME,SERIALNO,C,13,"Housing unit/GQ person serial number"
VAL,SERIALNO,C,13,"2013000000001","2017999999999","Unique
identifier"
```

D. Explanation of Variables in Data Dictionary CSV File

The table below describes the variables that appear in the CSV version of the PUMS Data Dictionary File. The position of the variables is given as the file contains no header. That is, the first line of the file is not the variable names.

PUMS Variables in Data Dictionary CSV File

Position	Variable	Description
1	Identifying Flag	"NAME" for information about the variable "VAL" for Legal values of the variable
2	PUMS Variable Name	PUMS variable name (e.g. RT, SERIALNO, AGEP, etc.)
3	Variable Type	"C" for Character variable "N" for Numeric variable Note: most variables are character variables
4	Length	Length of PUMS variable
5	Starting Legal Value	Starting value for legal value range.
6	Ending Legal Value	Ending value for legal value range.
7	Description	Descriptive name

For values that are not a range, the starting and ending legal values will be identical.

In addition, the first line of a variable (FLAG = "NAME") will have the variable's descriptive name in place of the Starting Legal Value. There will be no values for Ending Legal Value or Description.

E. Location of PUMS Design Factors

PUMS provides two methods for data users to calculate variances. One uses a generalized variance function (GVF) which involves design factors (DF). For 2016 and earlier, the design factors were published in the PUMS Accuracy document. The values were given in tables at the end of the pdf document.

Beginning in 2017, the design factors are now published in a comma separated value (CSV) file. A description of the variables is provided in the table below.

Note that in the 2012-2016 ACS 5-year PUMS Accuracy document a table of contents was provided in Attachment A. This allowed data users to quickly skip to a particular state's design factors. The ATTACHMENT variable is included in the CSV file to provide continuity.

PUMS Design Factor Variables in CSV File

Variable	Description
DF_ID	ID variable to identify the Characteristic with a numeric code.
YEAR	4-digit year
PERIOD	Time period (1-year or 5-year)
ATTACHMENT	A-1 through A-53. The attachments were referenced in the 2016 PUMS Accuracy document
STATE	State Name
ST	State FIPS Code
CHARTYP	Characteristic Type (either "POPULATION" or "HOUSING")
CHARACTERISTIC	Description of PUMS Design Factor Characteristic Group
DESIGN_FACTOR	Design Factor

V. OBTAINING PUMS DATA

PUMS data may be obtained in the following ways:

A. ACS Website

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

B. American FactFinder

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

C. DataFerrett:

PUMS data is also available via 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.

Note: DataFerrett only works with Microsoft Internet Explorer and the ESR version of Mozilla Firefox. DataFerrett does not work with Microsoft Edge or Google Chrome.

D. ACS FTP Site

The PUMS Files are also available through the file transfer protocol (FTP) site at: https://www2.census.gov/programs-surveys/acs/data/pums/.

VI. 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 the variable FS (Yearly food stamp/Supplemental Nutrition Assistance Program recipiency) for them.

PUMS files containing data for the entire United States are separated into four 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 2013-2017 ACS 5-year PUMS files of United States Population Records will notice an "a" file, a "b" file, a "c" file, and a "d" file. Each file contains about a quarter of the population records in the 2013-2017 5-year PUMS dataset of the United States.

A. 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 **person-level** files using the set statement:

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

The 2013-2017 ACS 5-year PUMS files of the United States Housing Records also contain an "a" file and, "b" file, "c" file, and "d" file upon downloading. To create a complete housing-level file, the files must also be concatenated. Below are instructions for concatenating the PUMS household-level files, in the form of an italicized SAS program and pseudo-code.

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

```
data housing;
   set psam_husa psam_husb psam_husc psam_husd;
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.

First make sure the files are sorted by SERIALNO.

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

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 do not need to merge the files unless the estimates you wish to create 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.

VII. 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. For the PUMS Person files, the full PUMS weight is called PWGTP. For the Housing files, it is called WGTP. 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:

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

Where X_r is a replicate estimate from X_1 to X_{80}, 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 may be 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.

VIII. VARIABLE CHANGES IN THE 2013-2017 5-YEAR PUMS FILES

The 2013-2017 5-year PUMS includes most of the variables that were included in the 2012-2016 5-year PUMS.

There were also some variables with new codes, modified codes, or cosmetic changes to variable labels or value labels. See the 2013-2017 5-year PUMS Data Dictionary for a complete listing of the variables and values contained in the 2013-2017 5-year PUMS data file.

If PUMS users want to compare 2013-2017 PUMS estimates to 2012-2016 PUMS, or want to reuse a program written for 2012-2016 PUMS 5-year data, the following table provides some information about variables with changes between 2012-2016 and 2013-2017. For more details on the changes, see both the 2012-2016 and 2013-2017 5-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 5-year PUMS:ACCESS, FACCESSP, BROADBND, FBROADBNDP, COMPOTHX, FCOMPOTHXP, DIALUP, FDIALUPP, HISPEED, FHISPEEDP, HOTWAT, FHOTWATP, LAPTOP, FLAPTOPP, OTHSVCEX, FOTHSVCEXP, PLMPRP, FPLMPRP, SATELLITE, FSATELLITEP, SMARTPHONE, FSMARTPHONP, SSMC, TABLET, FTABLETP, FHICOVP.

Variables deleted since previous 5-year PUMS: None.

Variables with new or modified codes since the previous 5-year PUMS: ADJHSG, ADJINC, AGS, ANC1P, ANC2P, CITWP, CONP, DECADE, ELEP, FER, FFSP, FINCP, FLANP, FS, FULP, GASP, GRNTP, HINCP, INDP, INSP, INTP, LANP, MARHYP, MHP, MIGSP, MRGP, NAICSP, OCCP, OIP, PAP, PERNP, PINCP, POBP, PUMA, RETP, RNTP, RWAT, RWATPR, SEMP, SERIALNO, SMOCP, SMP, SRNT, SSIP, SSP, SVAL, TAXP, TEL, WAGP, WATP, WKW, YBL, YOEP.

Variables with cosmetic changes to variable labels or value labels: AGEP, BDSP, COW, FES, HHT, JWMNP, OCCP, RMSP, RNTM, SOCP, VALP.

Detailed Changes by Variable (as compared to 2017 ACS 5-year PUMS Data Dictionary)

Variable	Description of Change
	New Variable
ACCESS	New Computer and Internet Variable for 2013-2017 PUMS 5-year
BROADBND	New Computer and Internet Variable for 2013-2017 PUMS 5-year
COMPOTHX	New Computer and Internet Variable for 2013-2017 PUMS 5-year
DIALUP	New Computer and Internet Variable for 2013-2017 PUMS 5-year
FACCESSP	New Computer and Internet Variable for 2013-2017 PUMS 5-year
FBROADBNDP	New Computer and Internet Variable for 2013-2017 PUMS 5-year
FCOMPOTHXP	New Computer and Internet Variable for 2013-2017 PUMS 5-year
FDIALUPP	New Computer and Internet Variable for 2013-2017 PUMS 5-year
FHISPEEDP	New Computer and Internet Variable for 2013-2017 PUMS 5-year
FHOTWATP	New Computer and Internet Variable for 2013-2017 PUMS 5-year
FLAPTOPP	New Computer and Internet Variable for 2013-2017 PUMS 5-year
FOTHSVCEXP	New Computer and Internet Variable for 2013-2017 PUMS 5-year
FPLMPRP	New Computer and Internet Variable for 2013-2017 PUMS 5-year
FSATELLITEP	New Computer and Internet Variable for 2013-2017 PUMS 5-year
FSMARTPHONP	New Computer and Internet Variable for 2013-2017 PUMS 5-year
FTABLETP	New Computer and Internet Variable for 2013-2017 PUMS 5-year

Variable	Description of Change
HISPEED	New Computer and Internet Variable for 2013-2017 PUMS 5-year
HOTWAT	New Computer and Internet Variable for 2013-2017 PUMS 5-year
LAPTOP	New Computer and Internet Variable for 2013-2017 PUMS 5-year
OTHSVCEX	New Computer and Internet Variable for 2013-2017 PUMS 5-year
PLMPRP	New Computer and Internet Variable for 2013-2017 PUMS 5-year
SATELLITE	New Computer and Internet Variable for 2013-2017 PUMS 5-year
SMARTPHONE	New Computer and Internet Variable for 2013-2017 PUMS 5-year
TABLET	New Computer and Internet Variable for 2013-2017 PUMS 5-year
FHICOVP	New Variable
SSMC	New Variable
	Variables with Annual Code Updates
ADJHSG	Annual Update of Inflation Factors
ADJINC	Annual Update of Inflation Factors
SERIALNO	Added 2017 SERIALNO codes and Removed 2012 SERIALNO codes
YBL	2017 Added as a category
	Variables with Annual Code Updates and Code Changes
YOEP	Updates due to new collapsing criteria and 2017 added
	Variables with Description Changes and Code Changes
ANC1P	Updates due to new collapsing criteria and changed spelling of "Luxembourger"
ANC2P	Updates due to new collapsing criteria and changed spelling of "Luxembourger"
MARHYP	Updates due to new collapsing criteria.
MIGSP	Updates due to new collapsing criteria and changed code for "Eastern Africa, Not Specified".
PAP	Added note to description indicating that variable is top-coded and range of legal values changed.
PERNP	Added note to description indicating that variable is top-coded and range of legal values changed.
PINCP	Added note to description indicating that variable is top-coded and range of legal values changed.
POBP	Update due to new collapsing criteria and changed code for "Eastern Africa, Not Specified"..
	Variables with Code Changes
CITWP	Updates due to new collapsing criteria and added 2017 as a category
DECADE	Changed "2000 or later" to 2 categories: "2000 - 2009" and "2010 and later".
FER	Category for "Suppressed for data year 2012 for selected PUMAs" removed as it is no longer necessary
	Variables with Description Changes
AGS	Update description to clarify no adjustment factor necessary
CONP	Note to use inflation factor added to description.

Variable	Description of Change
ELEP	Note to use inflation factor added to description.
FFSP	Added "SNAP" in description
FINCP	Note to use inflation factor added to description.
FLANP	Description changed to "Language other than English" for clarity.
FS	SNAP added to description.
FULP	Note to use inflation factor added to description.
GASP	Note to use inflation factor added to description.
GRNTP	Note to use inflation factor added to description.
HINCP	Note to use inflation factor added to description.
INDP	Clarification of descriptions of Industry Categories
INSP	Note to use inflation factor added to description.
INTP	Note to use inflation factor added to description.
LANP	Note added to some language categories to indicate when applicableand description changed to “Language other than English” for clarity.
MHP	Note added to description that variable is top-coded
MRGP	Note added to description that variable is top-coded
NAICSP	Clarification of category labels
OIP	Note added to description that variable is top-coded
PUMA	Clarified description
RETP	Note added to description that variable is top-coded
RNTP	Note added to description that variable is top-coded and minor description change
RWAT	Removed "2013 and later" in description
RWATPR	Removed "2013 and later" in description
SEMP	Note added to description that variable is top-coded
SMOCP	Note added to description that variable is top-coded
SMP	Note added to description that variable is top-coded
SRNT	Clarified description
SSIP	Note added to description that variable is top-coded and Changed range in description
SSP	Note added to description that variable is top-coded and Changed range in description
SVAL	Clarified description
TAXP	Note added to description that variable is top-coded and Changed range in description
TEL	Removed category for 2012 from variable and description
WAGP	Note added to description that variable is top-coded
WATP	Note added to description that variable is top-coded
WKW	Clarified description

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

IX. ADDITIONAL INFORMATION

A. 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.

B. 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.

C. 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.

D. 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 PUMS occupation code. In these cases, the PUMS occupation description is used for the SOC occupation title.

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

E. Note on Selected Values for INDP, NAICSP, and OCCP:

Some codes are pseudo codes developed by the Census Bureau and are not official or equivalent NAICS or occupation codes.

Pseudo Codes Values for Select Variables

Variable	Value	Description
INDP	9920	Unemployed And Last Worked 5 Years Ago Or Earlier Or Never Worked
NAICSP	9920	Unemployed And Last Worked 5 Years Ago Or Earlier Or Never Worked
OCCP	9830	MIL-Military, Rank Not Specified
OCCP	9920	Unemployed And Last Worked 5 Years Ago Or Earlier Or Never Worked

F. Codes to Identify NAICS Equivalents:

Data users may notice that some values of the PUMS variable NAICSP contain letters in addition to numbers. The table below provides an explanation of these letters.

Description of Special Letters in NAICSP Variable

Code	Description
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 may be found at:
https://www.census.gov/topics/employment/industry-occupation/guidance/indexes.html.

Note that NAICS stands for North American Industry Classification System and is pronounced as “nakes”.

G. Additional Information on PUMS Industry and Occupation Codes:

Data users may wish to consult the Code Lists on the PUMS Technical Documenation site (https://www.census.gov/programs-surveys/acs/technical-documentation/pums/documentation.html) for more information on how industry and occupation codes are mapped to PUMS industry and occupation codes.

For additional information on NAICS and SOC groupings within major categories visit:
https://www.census.gov/topics/employment/industry-occupation.html.

H. 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 four files, an "A", "B", "C", and "D" file. For Person-level data the names are "PSAM_PUSA", "PSAM_PUSB", "PSAM_PUSC", and "PSAM_PUSD". The Housing-level files are "PSAM_HUSA", "PSAM_HUSB", "PSAM_HUSC", and "PSAM_HUSD".

States in PUMS National Files

File	First State	First State FIPS Code	Last State	Last State FIPS Code
A	Alabama	01	Hawaii	15
B	Idaho	16	Mississippi	28
C	Missouri	29	Oregon	41
D	Pennsylvania	42	Wyoming	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".

I. Note on the PUMS Data Dictionary Name

The name of the PUMS Data Dictionary has the form "PUMS_Data_Dictionary_<YYYB>_<YYYE>". Where <YYYB> is the 4-digit year that begins the 5-year time period, and <YYYE> is the 4-digit year that ends the time period. For example, the pdf version is called: "PUMS_Data_Dictionary_2013_2017.pdf".

J. 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.