AMERICAN COMMUNITY SURVEY
2015-2019 ACS 5-YEAR PUMS FILES
ReadMe

Prepared by
American Community Survey Office
U.S. Census Bureau
January 14, 2021

Table of Contents

I. PURPOSE OF THIS DOCUMENT

This document is intended to provide resources and guidance for data users on how to use the Public Use Microdata Sample (PUMS) files.

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

The Public Use Microdata Sample (PUMS) files allow data users to create estimates for user-defined characteristics. The files contain a sample of the responses to the American Community Survey (ACS). The PUMS files includes variables for nearly every question on the ACS survey. Additional variables are also created from PUMS responses to provide data users with useful variables (such as poverty status) while protecting confidentiality and providing consistency within the PUMS files.

There are two types of PUMS files, one for Person records and one for Housing Unit records. Each record in the Person file represents a single person. Individuals are organized into households, making possible the study of people within the contexts of their families and other household members. In addition, the files contain people who reside in group quarters, such as nursing homes or college dormitories. The Housing Unit files contain records for single housing units, including vacant housing units.

PUMS files for an individual year contain data on approximately one percent of the United States population. PUMS files covering a five-year period contain data on approximately five percent of the United States population.

The PUMS files are much more flexible than the aggregate data provided in tables on data.census.gov, 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 replaced with a top-code value and uncommon birthplace or ancestry responses are grouped into broader categories. The PUMS files also limits the geographic detail below the state level. The only substate geography provided is the Public Use Microdata Area, or PUMA.

III. PUMS GEOGRAPHIES

The following sections provide an overview on geographies available in the PUMS files.

A. Geographies Available in PUMS

In order to protect confidentiality, a limited number of geographic summary levels are available on the PUMS files. They include region, regional division, state and Public Use Microdata Area (PUMA).

Regional division is a subdivision of the region summary level. An example of region is "Northeast", while a regional division is "New England". The PUMS variable for regional division is called "DIVISION".

In addition to the 50 states, there are also two state equivalents. They are the District of Columbia and Puerto Rico. Records for the District of Columbia are included in the PUMS files for the nation. Puerto Rico data is only available as a state-level file.

B. Overview of Public Use Microdata Areas (PUMA)

While PUMS files contain cases from nearly every town and county in the nation, 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 roughly 100,000 people at the time of their creation. They are created after each Decennial Census. The current PUMS files use the PUMA definitions created after Census 2010.

PUMAs are identified by a 5-digit code. Note that you must use the state variable (ST) along with the PUMA code to uniquely identify an individual PUMA. PUMA codes are unique within a state, but not between states. For example, the PUMA code "00100" is used in both Connecticut and North Dakota.

C. Note on Dual PUMAs

The current PUMA boundaries were first used beginning in 2012. For PUMS data for 2011 and earlier, the boundaries created after Census 2000 are used. Older multi-year records will contain both PUMA codes. For those files, PUMA codes from Census 2000 are called PUMA00, while the newer PUMA codes from Census 2010 are called

PUMA10. Data users will need to crosswalk their data to obtain a single PUMA geography. They may do so using allocation rates from Mable (see below).

D. Interactive Maps Using TIGERweb

The Census Bureau provides an interactive mapping application, called TIGERweb. Data users can view PUMA boundaries from 2010. TIGERweb is available 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. This will take you to a new page.
- On the upper left, you should see choices: "Layers", "Legend", or "Task Results". Select "Layers."
- Under "Select Vintage:" choose "ACS 2019".
- Select "PUMAs, UGAs, and ZCTAs" on the left.
- Click on the map to zoom, or move the zoom scale bar to zoom in closer to the map.
- Expand the "PUMAs, UGAs, and ZCTAs" box to see the choices: "2010 Census Public Use Microdata Areas" and "2010 Census ZIP Code Tabulation Areas".
- When the check boxes next to "2010 Census ZIP Code Tabulation Areas" and "2010 Census Public Use Microdata Areas" become active, deselect "2010 Census ZIP Code Tabulation Areas" to simplify the map.
- Alternately, you may check other boxes to add or remove geographic summary levels and other features, such as "Hydrography". Use the plus sign ("+") to see more detail for a particular selection.

E. Static Maps for PUMAs

Data users may be interested in static maps of PUMA boundaries. These may be found at: https://www.census.gov/programs-surveys/geography.html

To Access the static maps:

- Click on "Geographies" on the left hand side.
- Click on "Geography Reference Maps".
- Click on "More" to the right of the list of years. Change the year to "2010".
- Scroll down and click on "2010 Census Public Use Microdata Area (PUMA) Reference Maps".
- Choose the state you are interested in from the drop down menu.
- A list of state and PUMA codes with their names will appear. Click on the link at the top under "Access Public Use Microdata Area (PUMA) Maps". The link will have the state name followed by "Maps". For example, "Idaho Maps".
- You will be redirected to the relevant Census ftp site. Use the list of state and PUMA codes to select the relevant folder. For example, the state and PUMA code for "16 00300" has the description "Central Idaho--Lewiston City & Nez

Perce Reservation PUMA". Thus, you would click on the folder called "puma1600300/" to access these maps.

- Select the pdf that you are interested in.

F. Crosswalking Dual PUMA Codes Using GEOCORR and MABLE

The software MABLE was developed by the Missouri Census Data Center (MCDC), which is part of the Census Bureau's State Data Center program (https://www.census.gov/about/partners/sdc.html).

GEOCORR stands for "Geographic Correspondence Engine". According to the MCDC site it is an application that accesses the MABLE geographic database. More information may be found at: http://mcdc.missouri.edu/applications/geocorr.html.

Among other things, the software allows data users to calculate the proportion of a PUMA code from Census 2000 that lies within the new PUMA codes from Census 2010. It also provides the data user with an allocation factor so that they may crosswalk old PUMA codes to new PUMA codes.

IV. PUMS DOCUMENTATION

The list below provide links to documentation that are useful for PUMS users.

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.html.

B. PUMS Documentation

The PUMS documentation has been reorganized. Data users may now go to https://www.census.gov/programs-surveys/acs/microdata.html to find a central location for PUMS documentation. The page provides links to the PUMS technical documentation, such as the PUMS data dictionary as well as provide links to webinars on PUMS. In addition, there are links to access the data either through the FTP site or through the new online tool, called MDAT.

C. PUMS Handbook

A series of Handbooks are available that provide an overview of various aspects of the ACS. They are located at: https://www.census.gov/programs-surveys/acs/guidance/handbooks.html. The PUMS Handbook is called "Understanding and Using the American Community Survey Public Use Microdata Sample Files: What Data Users Need to Know".

V. PUMS DATA DICTIONARY

A. PUMS Data Dictionary Overview

The PUMS Data Dictionary provides the values for each PUMS variables, as well as labels for each value. For example, on the PUMS files, if the PUMS variable "REGION" has a value of "3", a data user may use the Data Dictionary to see that "3" means "South".

The PUMS Data Dictionary is published in three different formats. There is a text version, a pdf version, and a comma-separated values (CSV) version. The information in each version is equivalent to one another.

B. Explanation of Variables in Data Dictionary PDF File

Below is an example of the PUMS variable for Record Type (RT). The first line shows the PUMS variable name (RT), followed by "Character" to indicate it is a character variable and the number 1 to indicate that the length of the variable is one. The next line provides the descriptive title for the variable ("Record Type"). The remaining two lines provide the PUMS values for the variable and their appropriate labels. For example, when RT = "H", this stands for "Housing Record or Group Quarters Unit".

Example of PUMS Data Dictionary (PDF Version)

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

C. Explanation of Variables in Data Dictionary CSV File

Below is an example of how the Data Dictionary appears in the CSV version.

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"
```

The position of the variables (from left to right in the file) is provided in the table below. The file itself contains no variable names. That is, the first line of the file represents records for the data.

PUMS Variables in Data Dictionary CSV File

Position	Variable	Description
1	Identifying Flag	"NAME" for information about the variable "VAL" for 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 variable value range.
6	Ending Legal Value	Ending value for value variable range.
7	Description	Descriptive name

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

Note that the first line of a variable (Identifying Flag is "NAME") contains only five variables and will have the variable's descriptive name. Data users who read in the file expecting each record to contain seven variables will thus have the variable's descriptive name in place of the Starting Legal Value and the Ending Legal Value and Description variables will be blank.

The text and CSV versions may be read into statistical programs to create formats for the PUMS variables. Note that the CSV version was first published for 2017 PUMS data. For 2016 and earlier data, only the text version is available.

D. Organization of PUMS Variables in the Data Dictionary

The PUMS variables are placed into groups within the Data Dictionary. The Housing variables come first, followed by the Person variables. The variables are further divided into categories which are listed below. These sections and subsections are provided in the pdf and text versions of the Data Dictionary, but not the CSV version. The CSV version is intended to be machine-readable. Therefore, only variables and their values are present in that file.

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

VI. OBTAINING PUMS DATA

PUMS data may be obtained in multiple ways.

A. ACS Website

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

There are two columns shown. The column on the left links the FTP site. The column on the right links to the Microdata Analysis Tool (MDAT).

B. Microdata Analysis Tool (MDAT)

The Microdata Analysis Tool (MDAT) may be found at: https://data.census.gov/mdat/.

The tool may be used to create estimates online without the use of statistical software. Note that the tool may change in the future. It is still under development and in beta form.

C. 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/.

Choose the relevant year that you are interested in. Then choose the folder for the time period. For example, choose the folder marked “1-Year” to access 1-year PUMS files.

The document called "PUMS_file_naming_convention.pdf" explains the file naming convention system. The file names include state abbreviations. Data users may find a list of state names and abbreviations useful. The information may be obtained here: https://www.census.gov/library/reference/code-lists/ansi/ansi-codes-for-states.html. Click on "FIPS Codes for the States and District of Columbia" to obtain the state abbreviations.

VII. 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 may use the housing unit files, whereas those using only person-level variables may use the person files.

The person files also contain records for persons in group quarters 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 quarters place holder records exist so that data users may obtain values for the variable FS (Yearly food stamp/Supplemental Nutrition Assistance Program recipiency).

PUMS files containing data for the entire United States are separated into several files due to their size. For 1-year data, there are two files, and "a" and a "b" file. For 5-year PUMS data, there are four files ("a" through "d"). More information is provided at the end of this document.

A. Basic Example of Combining PUMS Person and Housing Files

Below are instructions for concatenating the two 1-year "a" and "b" PUMS files to create a single national file. The code is in italics and uses SAS programming code.

Concatenate the person-level files using the set statement:

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

Concatenate the household-level files using the set statement:

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

As mentioned above, if the data user is using 5-year PUMS files they will need to concatenate four files (psam_husa through psam_husd) together.

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, merge the household and person files together using the serial number variable (SERIALNO).

Note that vacant housing units are included on the housing files. There are no person-level records for these records.

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.

VIII. PUMS WEIGHTS AND NOTES ON CALCULATING VARIANCES

A. PUMS Weighting Variables

The ACS PUMS is a weighted sample. Weighting variables must be used in order to calculate estimates which represent the actual population. Weighting variables are also needed to generate accurate measures of uncertainty, such as the standard error or margin of error.

The PUMS files include both population weights (in the Person files) and household weights (located in the Housing files). Population weights should be used to generate statistics about individuals, and household weights should be used to generate statistics about housing units or households. 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: Replicate Household weighting variables, used for generating the standard error and margins of error for households.

PWGTP1-PWGTP80: Replicate Person weighting variables, used for generating the standard error and margins of error for individuals.

The PUMS Weighting variables (PWGTP and WGTP) may both be used to generate PUMS estimates. They are also used in the generalized variance formulas (GVF) method for calculating standard errors using the design factors. Replicate weights may only be used to calculate standard errors and margins of error using the successive difference replication (SDR) method. The SDR method may also be referred to as direct standard errors.

B. Successive Difference Replication Formula for Calculating Uncertainty

Successive Difference Replication (SDR) standard errors and margins of error are expected to be more accurate than generalized variance formulas (GVF) standard errors and margins of error, although they may be more inconvenient for some users to calculate. Both methods are explained in more detail in the Accuracy of the PUMS document, located on the PUMS Technical Documentation page: https://www.census.gov/programs-surveys/acs/microdata/documentation.html.

As previously mentioned, each housing unit and person record contains 80 replicate weights. To use the replicate weights to calculate an estimate of the SDR standard error, first calculate the PUMS estimate using the PUMS weight (either PWGTP or WGTP).

Then calculate 80 replicate estimates, using each of the 80 replicate weights. For example, for the first replicate estimate, use the first replicate weight (e.g. PWGTP1 instead of PWGTP, or WGTP1 instead of WGTP).

Calculate the variance by first taking the difference between each replicate estimate and the PUMS estimate. Square each of these differences, and then sum the 80 squared differences. Multiply this sum by the quantity 4/80. The 4 is required to remove bias, while 80 is present due to the 80 replicate estimates. The equation is below:

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

In the equation, x_r is a rth replicate estimate, and x is the full PUMS weighted estimate.

To obtain the standard error (SE), take the square root of the variance. To obtain a 90% confidence level margin of error, multiply the SE by 1.645.

C. ACS Variance Replicate Table Documentation

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

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

For more information on the theoretical basis for using the Successive Difference Replicate method, please see the paper entitled "Aspects of Survey and Model-Based Postcensal Estimation of Income and Poverty Characteristics for States and Counties" by Robert Fay and George Train (https://www.census.gov/content/dam/Census/library/working-papers/1995/demo/faytrain95.pdf).

Please note that many estimates generated with PUMS may be different from estimates for the same characteristics published on data.census.gov. These differences are due to the fact that the PUMS microdata is a sample of the full ACS microdata, and includes only about two-thirds of the records that were used to produce ACS estimates. Additional edits appropriate for PUMS were also made for confidentiality reasons.

IX. CHANGES TO PUMS VARIABLES FOR THE 2015-2019 PUMS 5-YEAR FILES

The 5-year PUMS files includes most of the variables that were included in the previous year. There were some new variables and five variables which were removed. There were also some variables with new codes, modified codes, or cosmetic changes to variable labels or value labels. See the 2015-2019 5-year PUMS Data Dictionary for a complete listing of the variables and values contained in the PUMS data files.

The PUMS Data Dictionaries may be found on the PUMS Technical Documentation page located at: https://www.census.gov/programs-surveys/acs/microdata/documentation.html.

Data users may wish to compare PUMS estimates to previous year estimates, or reuse a program written using previous year data. The PUMS Data Dictionary for the current year and previous year(s) may be used to find changes to the values for a specific variable.

Variables added since previous 5-year PUMS: CPLT, FJWTRNSP, FRELSHIPP, FWKWNP, HHT2, JWTRNS, RELSHIPP, WKWN.

Variables deleted since previous 5-year PUMS: FJWTRP, FRELP, JWTR, RELP, SSMC.

Variables with new or modified codes since the previous 5-year PUMS: ADJHSG, ADJINC, CITWP, MARHYP, SERIALNO, YBL, YOEP.

Variables that changed from numeric to character: HFL

Variables that changed from character to numeric: SPORDER

Variables with cosmetic changes to variable labels or value labels: ELEFP, FHINS6P, GRPIP, HINS6, HOTWAT, INDP, JWMNP, NAICSP, OCCP, OCPIP, PLMPRP, RMSP, SOCP, SVAL, TEL, WKEXREL, WKW.

Detailed Changes by Variable (compared to 2018 ACS 5-year PUMS Data Dictionary)

Variable	Description of Change
	New Variables
CPLT	New variable for Couple type
FJWTRNSP	New Allocation flag variable for JWTRNS
FRELSHIPP	New Allocation flag variable for RELSHIPP
FWKWNP	New Allocation flag variable for WKWN
HHT2	New variable for Household/family type (includes cohabiting) Note: HHT remains on the PUMS files
JWTRNS	Replaces JWTR
RELSHIPP	Replaces the variable RELP
WKWN	Numeric variable for Weeks Worked Note: WKW remains on the PUMS files
	Deleted Variables
FJWTRP	Allocation flag for JWTR removed and replaced with FJWTRNSP
FRELP	Allocation flag for RELP removed and replaced with FRELSHIPP
JWTR	Removed and replaced with JWTRNS
RELP	Removed and replaced with RELSHIPP
SSMC	Removed variable for Same-sex married couple households
	Variables with Metadata Changes
HFL	Data dictionary entry corrected from numeric to character variable
SPORDER	Data dictionary entry corrected from character to numeric variable
	Variables with Annual Code Updates
ADJINC	Annual Update of Inflation Factors
ADJHSG	Annual Update of Inflation Factors
CITWP	New bottom-coded value (1944 and earlier) and Current year (2019) added
MARHYP	New bottom-coded value (1940 and earlier) and Current year (2019) added
YBL	Current year (2019) added
YOEP	Current year (2019) added
	Variables with Code Changes
ADJHSG	Updated Housing Inflation Adjustment factors for 2019
ADJINC	Updated Income Inflation Adjustment factors for 2019
SERIALNO	Removed 2014 Serial Numbers and added Serial Numbers for 2019
	Variables with Cosmetic Changes

Variable	Description of Change
ELEFP	Values for blank and 1 (Included in rent or in condo fee) now on separate lines in Data Dictionary
FHINS6P	Changed variable description
GRPIP	Percent sign (%) changed to 'percent'
HINS6	Changed variable description
HOTWAT	Removed wording about records from 2013
INDP	Minor changes to select label descriptions
JWMNP	Changed "at" to "from" in category label for bbb (values that are blank)
NAICSP	Minor changes to select label descriptions
OCCP	Minor changes to select label descriptions
OCPIP	Percent sign (%) changed to 'percent'
PLMPRP	Added note about Flush Toilets included for 2016 and earlier records
RMSP	Changed lowest value from 0 to 1
SOCP	Minor changes to select label descriptions
SVAL	Clarified SVAL = "1" definition for records prior to 2016
TEL	Added 2019 TEL = "8" (Suppressed values)
WKEXREL	Changed "<" to "less than" in labels for values
WKW	Corrected label for WKW = "6"

In 2017, there was a need to collapse the categories of certain variables due to data disclosure requirements. In 2018, the variables for Industry and Occupation were also updated. The list of collapsed variables are:

- ANC1P
- ANC2P
- CITWP
- HHLANP
- INDP
- LANP
- MARHYP
- MIGSP
- NAICSP
- OCCP
- POBP
- RAC2P
- RAC3P
- SOCP
- YOEP

X. SUPPRESSION OF VARIABLE VALUES FOR SELECT GEOGRAPHIES

A. Suppression for Ancestry Variables (ANC, ANC1P, ANC2P) and Divorced (MARHD)

Due to issues found with the ACS data, Ancestry variables (ANC, ANC1P, and ANC2P) and the variable for Divorced in the Past 12 Months (MARHD) for selected geographies are suppressed. Records which were suppressed for ANC and MARHD were assigned a value of '8'. Records which were suppressed for ANC1P and ANC2P were assigned a value of "000".

More information may be found on the ACS Errata Notes page for 2018, located at: https://www.census.gov/programs-surveys/acs/technical-documentation/errata.html. The relevant notes are called "Data Collection Error in Philadelphia County, PA" and "Data Collection Error in Delaware".

B. Suppression for Telephone (TEL)

Problems in the collection of data on the availability of telephone service (TEL) led to suppressing this variable in various PUMAS in 2015 and 2016. Records which were suppressed were assigned a value of '8' for the Telephone variable (TEL).

For the affected geographies, please see the errata note located at : https://www.census.gov/programs-surveys/acs/technical-documentation/errata/118.html

XI. EXPLANATION OF VARIABLE CHANGES

This section is for additional information on changes that occurred to the PUMS Variables.

A. Note on New and Deleted Variables (CPLT, RELSHIPP, and JWTRNS)

Due to changes from the ACS content test, some variables were removed and replaced with the new version. Prior year data was crosswalked using PUMS variables.

Couple Type (CPLT) replaced the Same Sex Married Couple Household variable (SSMC). The Relationship variable (RELP) was replaced with the new Relationship to Reference Person variable (RELSHIPP). Finally, the previous Means of Transportation to Work variable (JWTR) was replaced with the new Means of Transportation to Work variable (JWTRNS).

The old allocation flag variables were also removed and replaced with the new allocation flags. Note that SSMC and CPLT do not have allocation flag variables associated with them.

B. Note on New Variables (HHT2 and WKWN)

The new variable for Household/Family Type (includes cohabitating) called HHT2 was added to the PUMS files. The older Household/Family Type variable (HHT) was not removed.

In addition, a numeric version of the Weeks Worked During Past 12 Months variable (WKWN) was added to the PUMS files. The categorical version, called WKW, remains on the file. Because WKW is categorical, it could not be crosswalked over to WKWN. The numeric values of WKWN were placed into categories and added to create WKW for user's convenience.

C. New Bottom-codes (CITWP and MARHYP)

The variables for Year of naturalization write-in (CITWP) and Year Last Married (MARHYP) had the bottom-coded years changed to match the 2019 PUMS 1-year values.

D. Note on Metadata Changes (HFL and SPORDER)

The type (character or numeric) in the data dictionary for the variables Home Heating Fuel (HFL) and Person Number (SPORDER) were corrected. The variable type in the PUMS files did not change.

XII. ADDITIONAL NOTES AND USEFUL INFORMATION

A. Rounding Rules for Income Variables

PUMS Income variables are subject to rounding rules, displayed in the table below.

Range	Rounding Rule	Unrounded Example	Rounded Example
0	0	No Rounding	0
$0 < X \leq 7$	4	6	4
$7 < X \leq 999$	Nearest ten	12	10
$999 < X \leq 49{,}999$	Nearest hundred	5,234	5,200
$X > 49{,}999$	Nearest thousand	54,123	54,000
$-7 \leq X < 0$	-4	-6	-4
$-999 \leq X < -7$	Nearest negative ten	-12	-10
$-49{,}999 \leq X < -999$	Nearest negative one hundred	-5,234	-5,200
$X \leq -49{,}999$	Nearest negative thousand	-54,123	-54,000

B. Note on the PUMS Design Factors

The PUMS design factors are periodically updated as the need arises. For example, if new variables are included on the PUMS file, additional design factors may be added. The design factors are not updated every year. In 2017 they were published as a csv file. Prior to that they were located at the end of the PUMS Accuracy document in a set of tables. The design factors for the 2015-2019 PUMS 5-year data are the same as the ones published for the 2014-2018. There are no changes.

C. Note on Income and Earnings Inflation Factor (ADJINC)

Divide ADJINC by 1,000,000 to obtain the inflation adjustment factor and multiply it to the PUMS variable value to adjust it to 2018 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.

D. Note on Housing Dollar Inflation Factor (ADJHSG)

Divide ADJHSG by 1,000,000 to obtain the inflation adjustment factor and multiply it to the PUMS variable value to adjust it to 2018 dollars. Variables requiring ADJHSG on the Housing Unit files are: CONP, ELEP, FULP, GASP, GRNTP, INSP, MHP, MRGP, SMOCP, RNTP, SMP, TAXAMT, and WATP.

For PUMS 1-year data, ADJHSG has a value of 1000000. That is, a housing dollar inflation factor of 1. Consult the PUMS Data Dictionary for the values for the 5-year files.

E. Inflation Factor Adjustment for AGS, TAXAMT, TAXP, and VALP

Note that TAXAMT is inflation adjusted. In the past TAXP was not due to it being a categorical variable. In addition, the Housing Dollar Inflation factor is not applied to

Property Value (VALP). Data users may apply the inflation adjustment to VALP if they wish to compare VALP and TAXAMT using the same criteria.

Note that ADJHSG does not apply to AGS because it is a categorical variable. If data users convert the categories in AGS to a numeric value (for example, using the midpoint of the range of each category), then they may apply the inflation factor.

F. Note on Public Use Microdata Area (PUMA) Variable

Public use microdata areas (PUMAs) are the smallest geography available on the PUMS files. They are designed to have a population of roughly 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.

G. Note on Standard Occupational Classification codes (SOCP)

In cases where the Standard Occupational Classification (SOCP) codes ends in X(s) or Y(s), two or more SOC occupation codes were aggregated to correspond to a specific PUMS SOCP 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/code-lists.html.

H. Note on Selected Values for Industry and Occupation (INDP, NAICSP, OCCP, and SOCP)

Some codes are pseudo-codes developed by the Census Bureau and are not official or equivalent NAICS, industry, or occupation codes. NAICS stands for "North American Industry Classification System" and is pronounced "nakes".

Pseudo-Codes Values for Select Variables

Variable	Value	Description
INDP	9920	Unemployed, With No Work Experience In The Last 5 Years Or Earlier Or Never Worked
NAICSP	999920	Unemployed, With No Work Experience In The Last 5 Years Or Earlier Or Never Worked
OCCP	9830	MIL-Military, Rank Not Specified
OCCP	9920	Unemployed, With No Work Experience In The Last 5 Years Or Earlier Or Never Worked
SOCP	999920	Unemployed, With No Work Experience In The Last 5 Years Or Earlier Or Never Worked

I. Codes to Identify NAICS Equivalents

Data users may notice that some values of the PUMS variable NAICSP contain letters in addition to numbers. NAICS stands for "North American Industry Classification

System" and is pronounced "nakes". 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/code-lists.html.

J. 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/microdata/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 see the Industry and Occupation page, located at: https://www.census.gov/topics/employment/industry-occupation/guidance/code-lists.html.

K. Note on PUMS Data Dictionary and Blank Values

Records in PUMS that are not within the universe for a variable are given blank values. For example, for the PUMS variable Educational Attainment (SCHL), the universe is for people age 3 or older. Person records with an age less than 3 have a blank value for SCHL.

The PUMS Data Dictionary represents blank values as a series of b's. For example, Educational Attainment has a length of 2. In the PUMS Data Dictionary, blank values for SCHL are displayed as "bb". The PUMS files do not use b's to denote blanks. Instead, they are either a numeric blank (for numeric variables) or a character blank value (for character variables).

L. Note on Top-coded and Bottom-coded Variables

To protect confidentiality, selected PUMS variables are top-coded or bottom-coded. That is, a threshold value is calculated. For top-coded variables, all values at or above the threshold value are replaced with the top-code value. For bottom-coded variables, all values at or below the bottom-code threshold are replaced with the bottom-code value.

The threshold value is calculated by identifying the top half percent or top three percent value using the distribution of the full microdata.

The following variables use the half percent threshold: AGEP, BDSP, ELEP, GASP, INSP, JWMNP, MRGP, RMSP, RNTP, SSP, TAXP, VALP, WAGP, WATP.

The following variables use the three percent threshold: CONP, FULP, INTP, MHP, OIP, PAP, RETP, SEMP, SMP, SSIP.

M. 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/library/reference/code-lists/ansi.html. Choose the "State and State Equivalents" link. FIPS Codes are 2-digit codes. For example, for Connecticut, <ST> is "09".

Note that for the National files, there are multiple files. For PUMS 1-year data there is an "A" and "B" file. For Person-level data the names are "PSAM_PUSA" and "PSAM_PUSB". The Housing-level files are "PSAM_HUSA" and "PSAM_HUSB".

States Contained in PUMS 1-year National Files

File	First State	First State FIPS Code	Last State	Last State FIPS Code
A	Alabama	01	Mississippi	28
B	Missouri	29	Wyoming	56

For the 5-year data, 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 Contained in PUMS 5-year 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".

N. 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 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.

PUMS Design Factor Variables in CSV File

Variable	Description
YEAR	4-digit year
PERIOD	Time period (1-year or 5-year)
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

O. Note on HOTWAT, PLM, PLMPRP, and RWAT

Some variables pertaining to plumbing are used only for Puerto Rico or only for the United States.

Variable	Description
HOTWAT	Water heater (Puerto Rico only)
RWAT	Hot and cold running water
PLM	Complete plumbing facilities
PLMPRP	Complete plumbing facilities for Puerto Rico
FHOTWATP	Water heater allocation flag (Puerto Rico only)
FPLMPRP	Complete plumbing facilities allocation flag for Puerto Rico

P. BUS, FBUSP, TOIL, and FTOILP Discontinued in 2016

Variables BUS, FBUSP, TOIL and, FTOILP were discontinued in 2016 due to questionnaire changes. However, these variables remain on the 2015-2019 file. These variables retain their values for data for 2015. For 2016 and later year, the variables BUS and TOIL have a value of '9'. The variables FBUS and FTOIL were assigned a value of '0'.

Data users should use caution when dealing with the variables BUS, TOIL, FBUSP and FTOILP. Five year estimates for these variables cannot be derived from the current 5-year PUMS file as there is no data for 2016 or later.

These variables were retained on the file as they were previously included as components of the variables SVAL and PLM, respectively. The removal of the variables caused a change in the components of these variables between data years 2015 and 2016. For 2015, BUS was included as a component in SVAL; for 2016 and later, SVAL does not consider the presence of a business on the property. Likewise, PLM required a flush toilet

in 2015; in 2016 and later, complete plumbing is not defined by TOIL. Keeping the 2015 data values on the 2015-2019 ACS 5-year PUMS files allows users to create recodes of PLM and SVAL that will be comparable across all five data years.

For more information, see the User Note at: https://www.census.gov/programs-surveys/acs/technical-documentation/user-notes/2018-01.html

Q. Issue with Own Child (OC) and Related Child (RC)

For records from 2016 and earlier, group quarters records assigned the variables Own Child (OC) and Related Child (RC) a value of zero ('0'). For 2017 and later years, GQ records have a blank value for OC and RC. These variables are housing variables and therefore group quarters records should not be included in the universe of OC or RC. There is currently no plan to reissue data files.

R. Errata and User Notes for ACS and PUMS

Data users may wish to read the original errata or user notes released to the public. In addition, this document does not include any notes related to ACS errata or user notes. While some ACS errata and user notes pertain to specific tables or data products, others refer to broader issues that data users may find relevant to using the PUMS data files.

Data users may find errata notes here: https://www.census.gov/programs-surveys/acs/technical-documentation/errata.html.

In addition, user notes are located here: https://www.census.gov/programs-surveys/acs/technical-documentation/user-notes.html.

S. 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.

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