American Housing Survey

Getting Started with the Public Use File: 1997 - 2013

v1.0

Updated May 18, 2021

U.S. Census Bureau, Department of Commerce
Department of Housing and Urban Development

Contents

Overview

The purpose of this document is to introduce American Housing Survey (AHS) users to the revised 1997-2013 AHS public use file (PUF) microdata. AHS users who have used the originally published 1997-2013 PUF microdata will notice changes in the revised files, such as:

- Fewer tables in the relational PUF structure
- Fewer weights
- Removal of noninterviews from the PUF
- Standardized supporting files and documentation

The remainder of this guide is organized into sections, with each section addressing an important PUF microdata topic. If applicable, comparisons will be made with the original 1997-2013 PUFs or PUFs from other years.

Using the 1997-2013 PUF microdata requires a statistical program such as SAS, STATA, or R. Although it is technically possible to use PUF microdata in Microsoft Excel or Access, users will find doing so difficult. The appendix in this document contains examples of merging datasets and creating tables with commonly used AHS variables in SAS.

PUF users who do not have the resources to purchase a statistical program such as SAS or STATA can obtain a free program such as R or Python, which has downloadable packages for data management and analysis.

Revisions to the 1997-2013 PUFs

In May 2021, HUD and Census released revisions to all national and metro PUFs from 1997 through 2013 to increase usability for data users. The list below summarizes the revisions made to the files, which are also covered in more detail throughout this guide.

1. Variables without usable information were removed from the files. This is covered in the Variable Formatting section of this guide.
2. Standardized formatting rules were applied to all variables. This is also discussed in more detail in the Variable Formatting section.
3. Non-interview sample along with variables used for non-interview data were removed from the files. This is covered in more detail in the Respondent Types section.
4. Datasets were combined and condensed where possible to reduce the number of files required to use the PUFs. More detail on this process is included in the File Structure section.
5. A standardized flat and CSV file format was created for each survey year, which is also described in the File Structure section.
6. Electronic codebook entries for all historic files were reviewed for accuracy, and corrections and updates were applied. This is described in more detail in the Interactive Codebook section.
7. A standard value label package (VLP) was created for each set of PUFs. This is explained more in the section on Value Label Packages.

Sample Design

Throughout its history, the AHS has always consisted of a longitudinal national sample (hereinafter referred to as AHS-N) and independent longitudinal samples of selected metropolitan areas (hereinafter referred to as AHS-MS). Over the length of a longitudinal sample in the AHS, the same housing units are

interviewed across multiple survey years, with new housing units added to the sample each year to represent new construction. While this guide focuses on the 1997-2013 period, it's important to note that the AHS-N sample is longitudinal from 1985-2013, and many of the AHS-MS samples are longitudinal with years earlier than 1997. From 1985-2013, the AHS-N sample was surveyed in every odd-numbered year. The AHS-MS samples were surveyed periodically throughout the same period, but unlike national, could also exist in even-numbered years. For a full listing of the metro areas sampled in each year, see https://www.census.gov/content/dam/Census/programs-surveys/ahs/tech-documentation/Quick%20Guide%20to%20Metro%20Area%20Histories%201974%20to%20Current.pdf.

The AHS-N Sample and Sample Changes

The AHS-N sample is designed to be representative of the nation as well as the four large Census regions and nine Census divisions within the U.S. The national sample was not designed to produce estimates of individual metropolitan area, states, or other areas smaller than the Census division.

The AHS-N sample underwent several changes during the 1985-2013 period. The table below summarizes the changes and their impact on the ability to produce national estimates.

Description of Change	Impact
Big 6 Oversample These cases were added to the AHS-N sample for 1995, 1999, 2003, 2009, 2011, and 2013 files. These cases are in the Chicago, Detroit, Los Angeles, New York, Northern New Jersey, and Philadelphia Metropolitan Statistical Areas. All six metropolitan areas were included in the AHS-N in 1995, 1999, 2003. For 2009 and 2013, Chicago, Detroit, New York, Northern New Jersey, and Philadelphia were included in the AHS-N. For 2011, Los Angeles was included in the AHS-N.	• This change meant that PUF users wishing to produce estimates for the six metropolitan areas would need to use the AHS-N PUF. • This change had no impact on the ability to produce national estimates from the AHS-N PUF.
HUD-Assisted Housing Units Oversample HUD added approximately 5,200 HUD-assisted renter cases to the AHS-N sample for 2011 and 2013.	• This change had no impact on the ability to produce national estimates from the AHS-N PUF. • This change gave users the ability to produce national estimates of HUD-assisted households.
1997 Sample Drop Some sample from earlier years were temporarily removed in 1997 and reinstated in 1999.	• This change had no impact on the ability to produce national estimates from the AHS-N PUF.
2007 Sample Drop These cases were reduced from the sample in 2007 for budget reasons. They were all reinstated in 2011 as part of the larger sample.	• This change had no impact on the ability to produce national estimates from the AHS-N PUF.

The AHS-MS Sample and Sample Changes

The data from the independent metropolitan surveys are contained in a single metropolitan dataset for their year. Within that dataset, the cases from each metropolitan area are identified by the SMSA variable. This is important because different metro areas could use the same CONTROL value.

The original release of the 2011 AHS survey data was an exception to this arrangement. In that survey, the metropolitan and national cases were originally released in a single combined dataset. The 2011 files were revised in April 2019, splitting the metropolitan and national cases into separate datasets like the other years in this time period. A more detailed description of these changes can be found in the 2011 version control document found at https://www2.census.gov/programs-surveys/ahs/2011/AHS%202011%20National%20PUF%20Version%20Control.pdf.

The cases from the six metropolitan areas conducted as oversamples in the national survey are always part of the national dataset and have no metro-specific weights that are publicly available. These are explained in further detail in the AHS-N sample section.

Like with the AHS-N sample, the independent AHS-MS samples are also longitudinal, meaning that they return to the same housing units in each survey year. Data users can use CONTROL and SMSA to link AHS-MS sample longitudinally. For the 2011 and 2013 metro files, the CONTROLM and SMSA codes will need to be linked to CONTROL and SMSA in earlier metro files. Full documentation on the longitudinal history of each AHS-MS sample can be found in the Metropolitan Oversample Histories: 1973-2013 guide at https://www.census.gov/programs-surveys/ahs/tech-documentation/help-guides/metro_oversamp_hist_2013.html.

Survey Design/Mode

From 1973 through 1996, the AHS was conducted using a paper survey instrument administered by Census Bureau Field Representatives (FRs) who visited respondents in their homes to conduct interviews in person. Beginning in 1997, the survey switched to computer-assisted personal interviewing (CAPI), in which the instrument is a program running on the FR's laptop computer. Most interviews were conducted in person, although some were conducted by telephone.

Evolution of Dependent Interviewing

In the paper questionnaire, values for a limited number of variables were carried forward using a paper control card that was kept for each survey year. After development of the electronic questionnaire in 1997, the AHS had the capability to expand the number of values carried forward electronically. A more detailed overview of this dependent interviewing can be found at https://www.census.gov/content/dam/Census/programs-surveys/ahs/tech-documentation/1985-2013/Dependent_Interviewing_1985_to_2013.pdf.

File Structure

Separate PUFs are published for the integrated national longitudinal sample and the metropolitan area samples during the 1997-2013 period. These PUFs are referred to as the national PUF and the metropolitan area PUF.

As in recent years, HUD and the Census Bureau are providing the PUF in two structures—relational and flat. The structure of the PUFs is discussed in the next two sections. From 1985 through 1996, HUD and the Census Bureau provided the PUF in only one format structure called the flat file. Beginning in 1997, HUD and the Census Bureau provided the PUF in both a relational structure and a flat file.

All PUFs are published in SAS and CSV formats. Besides the difference in file type, two other important differences exist between the SAS and CSV formats. First, the SAS files will include descriptive column labels. Second, the representation of "not applicable" and "missing/refused" is different.

PUF Flat File Structure

In a flat file, there is one file that contains all the data. Each record represents a housing unit. One-to-many relationships (such as demographic data for persons) are represented by numbered sets of variables. For example, there will be a RACE1, SEX1, GRAD1, etc. for the first person in the household, and another set for the second person (RACE2, SEX2, GRAD2, etc.). If a housing unit has fewer than the maximum number of such entities, the extra sets of variables are filled with missing values.

In the May 2021 revisions to the 1997-2013 files, a standard file flattening format was created and applied to all years. This corrects some issues in the original flattened files which sometimes contained formatting inconsistencies between survey years, additional empty variables, inconsistent or lack of labeling, and other minor formatting problems. The revisions also include updates to the flattened files to make them consistent with the revised relational PUFs.

The following standards are now applied to all years:

- For AHS-N PUFs, there will be always be exactly one row in the flattened file for each unique CONTROL number in that year's HOUSEHOLD file. For AHS-MS PUFs, there will always be exactly one row for each unique combination of CONTROL and SMSA.
- Flattened variables are numbered consecutively (e.g., AGE1, AGE2, AGE3, AGE4, etc.).
- Flattened variables will only be created if there is at least one observation with non-missing data (e.g., if the maximum household size in a survey year is 18, the flattened AGE variables will go up to AGE18 for that year).
- Variables without data will be set to a missing value (“.” for numeric or “ ” for character). For example, AGE4 will be a “.” for a household with only three members.
- Flattened file names will use the following convention: ahs[4-digit year][n for national or m for metro]. For example, “ahs1998m” or “ahs2011n”.

PUF Relational File Structure

In a relational database, the data for a given survey are divided among more than one file. Some of the files have one record per housing unit, but in others there may be many or no records for a given housing unit. For example, the file for person data has one record per person. A vacant unit would have no records in this file, while a unit with a large household may have many.

The PUF relational structure in 1997-2013 is similar to other year PUFs. The May 2021 revisions combined files that were previously separated to reduce the overall number of files needed. If a record is out of universe and doesn't exist on one of the files being combined (such as renter-occupied units not being on the original MORTG file), all variables on the smaller file for that record are set to a not applicable value. The following chart describes each table within the relational structure.

1997-2013 Relational Table Name	Description	Changes from Original PUFs
HOUSEHOLD	Contains one record for each housing unit. Also includes certain householder information, such as race and sex. The householder information is also on the PERSON file, but is duplicated in the HOUSEHOLD table for ease of use.	Household level tables from prior versions (HOUSHLD, WEIGHT, NEWHOUSE, REPWGT, TOPPUF, TOPICAL, RATIOV, OWNER, and MORTG) were combined into one HOUSEHOLD table. Note that unlike 2015, the mortgage file MORTG was at the household level.
PERSON	Includes one record for each person in the household. Each person in the person table received a unique person number PLINE.	Person level tables from the original JTW files were merged into the PERSON file.
PROJECT	Includes one record for each home improvement project undertaken by owner-occupied households within two years of the interview.	The original HOMIMP file was renamed to PROJECT to be more consistent with naming conventions with other survey years. Otherwise, this file functions the same.
RMOV	Includes one record for each recent mover group in a household (up to three), and contains variables from the recent mover section.	This file was not impacted by revisions. It also does not exist in 2015 or later, since all recent mover variables were flattened into the HOUSEHOLD files for those years.
OMOV	Includes one record for each out-mover group	This file was not impacted by revisions. It is also unique to topical data collected only in the 2013 AHS.

The PERSON, PROJECT, RMOV, and OMOV tables each have a many-to-one relationship with the HOUSEHOLD table. They can be linked to the HOUSEHOLD table using the variable CONTROL. CONTROL can also be used to link records from year to year. Metropolitan PUFs should be matched on CONTROL and SMSA.

Similarly, to match housing units longitudinally across survey years, data users can link by CONTROL for national and CONTROL and SMSA for metro sample. Because the CONTROL variable is formatted differently for 2011 and 2013 metro sample, linking those metro areas to prior years requires using CONTROLM and SMSA, which can then be linked to CONTROL and SMSA in prior years.

For data users familiar with the original 1997-2013 PUF relational files, below is a table with a description of all the files originally included in the 1997-2013 PUFs and how they were altered when the files were revised.

Relational Table	Years Used	Description
HOUSHLD	1997-1999	Includes one record for the housing unit, which includes the household occupying the unit. Holds most of the general information about the unit and the household. This was merged into the HOUSEHOLD table when the files were revised.
NEWHOUSE	2001-2013	Includes one record for the housing unit, which includes the household occupying the unit. Holds most of the general information about the unit and the household. This table replaces HOUSEHLD, TOPPUF, and WEIGHT. Also holds all the data from the householder's PERSON table record, so that these can be accessed without linking tables. (The name comes from "new househld file."). This was merged into the HOUSEHOLD file when the files were revised.
WEIGHT	1997-1999	Includes one record for each housing unit. Holds the survey weighting variables. This was merged into the HOUSEHOLD file when the files were revised.
TOPPUF	1997-1999	Includes one record for each housing unit. Holds variables that were recoded in the editing process. This was merged into the HOUSEHOLD file when the files were revised.
RMOV	1997-2013	Includes one record for each recent mover group in a household (up to three). Holds variables from the recent mover module. This file name and structure was not altered for the revisions.
OWNER	1997-2013	Includes one record for each rental occupied or vacant unit. Holds a variable on the presence of an owner or manager on site. This was merged into the HOUSEHOLD file when the files were revised. Variables that were originally exclusive to the OWNER file were set to Not Applicable values for owner occupied units.
MORTG	1997-2013	Includes one record for each owner-occupied unit with a mortgage or home equity loan. Holds sets of variables on up to three mortgages and three home equity loans. This was merged into the HOUSEHOLD file when the files were revised. Variables that were originally exclusive to the MORTG file were set to Not Applicable values for units that were not eligible to be on the original MORTG file.
HOMIMP	1997-2013 (except 1998 metro)	Includes one record for each home improvement project undertaken by an owner-occupied household in the past two years. This file was renamed to PROJECT to be consistent with 2015 and later survey years, but otherwise was not structurally changed during the revisions.

Relational Table	Years Used	Description
JTW	1997-2009 (except 1998 metro)	Includes one record for each worker in the household. Holds data about commuting (journey to work). This was merged into the PERSON file during the revisions. Variables that were originally exclusive to the JTW file were set to Not Applicable value for household members who were excluded from JTW due to not having a commute.
PERSON	1997-2013	Includes one record for each person in the household. Holds demographic and other data about the person. This file name and structure was not changed during the revisions.
RATIOV	1998-2013 (except 2007)	Includes one record per renter household meeting a threshold of high cost-to-income ratio. Holds data about coping methods. This was merged into the HOUSEHOLD file when the files were revised. Variables that were originally exclusive to the RATIOV file were set to Not Applicable values for units that were not eligible to be on the original RATIOV file.
REPWGT	2007-2009 national only	Includes one record per housing unit. Holds replicate weights. These are included in NEWHOUSE in 2011-2013. This was merged into the HOUSEHOLD file when the files were revised.
OMOV	2013	Includes one record per out-mover group. Holds some of the data from the topical module on doubled-up households. This file name and structure was not changed during the revisions.
TOPICAL	2013	Includes one record per household. Holds the household-level data from the 2013 topical modules. This was merged into the HOUSEHOLD file when the files were revised.
TYPEC	2011	Holds the data for "Type C" non-interview cases. The records in this file do *not* link to any other file in 2011, only longitudinally. (In other survey years, these cases are included in NEWHOUSE or HOUSHLD, as applicable). This file was deleted during the revisions because its information is included in the National Sample Case History File.

CSV Structure

All PUFs in 1997-2013 are available in both SAS and Comma Separated Value (CSV) formats. The CSV format allows data users to import the data when using statistical packages that aren't compatible with SAS file formats.

The May 2021 revisions to the 1997-2013 files introduced a standardized CSV format for each PUF. This fixes inconsistencies in CSV formatting with the original files and creates CSVs for years where they did not originally exist. In addition, the CSV files are consistent with the revised PUFs. The following standards are applied to the CSV files:

- Variable names on the CSV datasets will match their names on the SAS version exactly.
- Variable values on the CSV datasets will match the SAS version exactly, except for missing and non-applicable values.
- The length of variables on the SAS and CSV datasets will match.
- Character variables in SAS will have single quotes (‘ ’) around their values on the CSV file. Numeric values will not.
- SAS values of ‘ ’, ‘D’, ‘R’, and ‘N’ for character variables and ., .D, .R for numeric variables will be assigned a value of -9 (Not Reported) on the CSV files. Note that this differs from the 2015 and later values for N because, in a few instances, an ‘N’ exists on files prior to 2015 (such as for the HDSB variable) where it is considered “Not Reported” rather than “Not Applicable”.
- SAS values of ‘B’ for character variables and .B for numeric variables will be assigned a value of -6 (Not Applicable) on the CSV files.

1997-2013 Interactive Codebook

A codebook is a necessary tool for working with the PUFs. Prior to 2015, HUD and the Census Bureau maintained the AHS Codebook as a living document that spanned surveys conducted from 1997 through 2013. HUD and the Census Bureau made updates to the AHS Codebook each time a new survey was published. Other, static AHS codebooks covered surveys prior to 1997.

Starting with the 2015 PUF, HUD and the Census Bureau implemented an entirely new approach to maintaining the AHS Codebook called the AHS Codebook Interactive Tool. This tool is a web-based utility that includes all the information previously contained in the AHS Codebook but with enhanced functionality. This approach to presenting codebook-style information will enable HUD and the Census Bureau to easily update the AHS Codebook if corrections or additions are required.

The AHS Codebook Interactive Tool now includes both AHS-N and AHS-MS PUFs from 1973 to present, which covers all publicly available AHS data. The data includes variables on the AHS PUFs from 1973-2013 and includes both IUF and PUF variables for 2015 to present. The Codebook was also updated to be consistent with the May 2021 revisions to the 1997-2013 files. The current Interactive Codebook is located at https://www.census.gov/programs-surveys/ahs/tech-documentation/codebooks/ahs-codebook-tool.html.

Several features are built into the AHS Codebook tool to meet the needs of AHS PUF users. The following table describes various features.

Feature	Description
Filter by survey year	This feature enables users to filter the AHS Codebook by survey year. Users wishing to download an entire codebook for a specific year could use this feature.
Filter by topic or subtopic	This feature enables users to filter the AHS Codebook by a particular topic and, if desired, a subtopic. This feature is helpful for viewing all the variables in a particular topic.
Download button	This feature enables users to download a selected portion of the AHS Codebook. Format options include PDF and CSV.
Mini-codebook download	This feature enables users to download a PDF of all the variables in any specified survey year. The file includes the following fields: topic, subtopic, variable name, and description. The purpose of the file is to enable users to quickly review all the variables in the selected PUF and IUF.

Feature	Description
Historic view	When a variable's metadata has no changes, multiple years are grouped together into one entry. When a variable's metadata has changes for different years, the tool displays a separate entry. This feature allows users to easily identify when a variable's metadata has changed over time.

AHS = American Housing Survey. IUF = internal use file. PUF = public use file.

1997-2013 Internal Use Files

HUD and the Census Bureau make every attempt to include every possible AHS variable on the PUFs. However, our duty to protect respondents' confidentiality prevents public release of some variables. We maintain another file, the internal use file (IUF), that includes such variables.

If a 1997-2013 AHS PUF user cannot complete an analysis because one or more AHS variables are not available on the PUF, HUD and the Census Bureau may be able to help in two ways. First, HUD and Census Bureau staff can run special tabulations, with disclosure methods applied, using IUF Only variables. Second, HUD and Census Bureau staff can assist AHS users with the process of obtaining access to the AHS IUF. Data users can contact an AHS staff member with special table requests at ahsn@census.gov or at 301-763-3235.

AHS users should note this process requires the following steps. More details on the process are available at https://www.census.gov/programs-surveys/ces/data/restricted-use-data/apply-for-access.html.

- Submitting a detailed proposal.
- Submitting to a background check and obtaining a non-sensitive federal security clearance.
- Agreeing to become subject to legally binding disclosure restrictions.
- Willingness to travel to a Census Bureau Federal Statistical Research Data Center.[1]
- Willingness to pay a fee to access the Research Data Center.

Unlike the 2015 and later IUFs, variables unique to the 1997-2013 IUFs are not all documented in a publicly available data dictionary. Some limited information on historic IUF files is available at https://www.census.gov/content/dam/Census/programs-surveys/ahs/tech-documentation/American_Housing_Survey_AHS_Using_the_Internal_User_File_IUF.pdf. Data users can reach out to the AHS staff contact listed above to verify whether an IUF has the variables needed.

Respondent Types

Though the AHS visits the same housing units each cycle, units may be removed or added to the sample for various reasons throughout the survey process. HUD and the Census Bureau created the National Sample Case History File to track all sampled housing units in the integrated National sample. AHS users who want to know why various sample cases were not surveyed or were nonresponses for 1997-2013 can review this file.[2]

The 1997-2013 AHS PUFs include occupied interviews, usual residence elsewhere interviews, and vacant interviews. The variable STATUS indicates which of these three types of interviews was conducted.

[1] For more information, see https://www.census.gov/fsrdc.
[2] For more information, see https://www.census.gov/programs-surveys/ahs/tech-documentation/help-guides/1985-2013/national-sample-case-history_2013.html.

While many of the original PUFs from this time period included non-interview sample, the May 2021 revisions removed noninterviews to be consistent with formatting in 2015 and later files.[3]

As in other years, noninterviews are classified as one of three types indicated in the following table.

Noninterview Type	Reason for Noninterview	Eligibility for Future Surveys
Type A	Respondent refused the interview or could not be contacted.	Housing unit remains in the sample for future surveys.
Type B	Housing unit was currently not habitable but could be habitable in a future year.	Housing unit remains in the sample for future surveys.
Type C	Housing unit was demolished or destroyed.	Housing unit is permanently removed from the sample.

Variable Formatting

The 1997-2013 PUFs include more than 1,400 unique variables across all relational tables, plus edit flags and replicate weights. These variables come in two types—numeric and categorical.

The May 2021 file revisions applied the following standardization rules to the variable formats:

- All variable names were fully capitalized.
- Labels matching the e-codebook descriptions were applied to all files.
- Variables without usable data were removed from the PUFs. This includes variables that had only missing values, only reported one unique value, or if they were determined to be artifacts of data processing that were not useful for publication. A full list of variables removed during revisions can be found in the Appendix.
- Some ineligible values (such as underscores or other special character) were set to a standard missing value.
- Variable types were updated to be consistent across survey years and with the use of the variable. A variable was set to numeric if it has a natural representation as a number, such as a count, dollar, or percent value. All other variables were stored as character values. While variable type will not change for most variables covering this time period, the standards for the original files stored larger categorical variables as numeric. A full list of variables with type changes during revisions can be found in the Appendix.
- Leading spaces were removed from character variables with a length greater than one, and these variables were padded with leading 0s where appropriate.
- A small number of duplicated person level records were resolved (this only impacts the 2002 PERSON file).
- Some sample had only 11-digits in the CONTROLM variable on the 2011 Metro PUF. These were updated to add a trailing 0, which allows them to match to the 12-digit CONTROL variables used in previous years.

Many of the 1997-2013 PUF variables represent the same or similar concepts as other years. However, many variable names are different. Because the 2015 AHS included an entirely new sample, HUD and the Census Bureau took the opportunity to change variable names to make them more user friendly and to

[3] The following variables were also deleted from the files when non-interviews were removed: BBLDG, BOARDU, EXPOSE, DISAS, DFIRE, and NOINT. These variables were removed because they were only applicable for non-interviewed sample.

bring them in line with other surveys. More details on this process are available in the 2015 Getting Started Guide located at https://www.census.gov/programs-surveys/ahs/tech-documentation/help-guides/2015-later/puf_start.html.

Due to the numerous changes in variable names, HUD and the Census Bureau have produced a crosswalk file comparing 2015 PUF variable names with PUF variable names from the 2011 AHS and 2013 AHS, which is available at https://www.census.gov/programs-surveys/ahs/data/2015/ahs-2015-public-use-file--puf-/ahs-2015-national-public-use-file--puf-.html.

Replicating PUF Estimates in Tables

Data users often wish to replicate estimates published in the AHS Summary Tables. HUD and the Census Bureau publish documentation called "Table Specifications," which explains how to replicate most summary tables. Table specification files are located in the Data section of the Census AHS website (https://www.census.gov/programs-surveys/ahs/data.html). To locate the table specifications, choose the year of survey you want, click on the link to the PUF, and then choose the National PUF or the Metropolitan PUF.

From 1997 through 2009, the table specifications included pseudocode (generally in SAS format). In 2011 and 2013, the specifications were reformatted to use excel files designed to correspond to each individual AHS summary table. The specifications for each table now show the variables that go into each row and column in a more transparent and intuitive manner.

The estimates in the publication tables are created with IUF files, and thus, may not exactly match estimates created from the PUF due to disclosure methodology that are applied to anonymize the public use data. The Topcoding and Disclosure section of this document covers these methods in more detail. Unlike the 2015 and later specifications, some of the internal variables used in the 1997-2013 files do not exist in public codebook documentation. In these cases, HUD and Census make every effort to show how these variables are created in the documentation available. The weighting section of this document discusses the correct weights to use to replicate published estimates in different years of the survey.

Not Applicable vs. Not Reported

The 1997-2013 PUFs differentiate between respondents who didn't know, refused, or generally gave no response to the question. This differs from more recent PUFs which combine these into one "not reported" category. Below is a summary of how missing or not applicable data are identified in the 1997-2013 files.

SAS files[4]

SAS allows for more than one missing value code, represented by letters, even in numeric variables. The 1997-2013 AHS uses the following codes to represent missing values. For numeric variables, these codes will be preceded by a ".", such as ".B".

- B: Not applicable
- D: Don't know
- R: Refused
- N: Not reported (exists for very few variables, such as HDSB)

[4] SAS programming tip: It's a good practice to place the statement MISSING B D R; at the beginning of your program to help SAS recognize missing values.

In addition, some numeric variables will have the (.) system missing code, and some character variables will have a blank (‘ ’) missing value code. These are most often due to respondents missing the question at the time of the interview because they weren’t eligible at the time or quit the interview early.

However, for some variables it could represent the absence of a positive response, such as a missing parent line number meaning that the person’s parent doesn’t live in the household. In another example, a missing value in a mark all that apply variable (where an ‘X’ means they have the characteristic) means the respondent did not pick that response from the answer categories. In this case, it can usually be treated as a “No” response, and is documented as such in the codebook.

CSV files

Because other statistical software has trouble importing the SAS missing value codes, especially for numeric variables, the CSV files are recoded. Note that the CSV program used is standardized to be consistent with the 2015 and later PUFs, so not reported values are consolidated into -9 codes.

- -6: Not applicable
- -9: Not reported

Geography

As a national survey with a relatively small sample size, the AHS is limited in the geographic detail it can reveal on the PUF. There are two reasons for this. First, the combination of the survey’s rich data and a localized geographic variable could compromise the confidentiality of the survey’s respondents. Second, the small number of cases in a local geography do not comprise a sample size sufficient for statistically valid estimation.

The table below summarizes the geographic variables available on the 1997-2013 PUFs. For a more comprehensive discussion of AHS geography, including disclosure avoidance methodology and changes in definitions across survey years, see the 1985-2013 AHS geography guide at https://www.census.gov/programs-surveys/ahs/tech-documentation/help-guides/1985-2013/geography--public-use-file--1985-2013.html.

Geography Variable	Years	Description
REGION	1997-2013 National	The four Census regions.[5]
DIVISION	2011-2013 National	The nine Census divisions. For some cases on the PUF, divisions 5 (South Atlantic) and 6 (East South Central) and divisions 8 (Mountain) and 9 (Pacific) are combined to prevent disclosure.
STATE	1998-2011 Metro	This variable identifies state codes for sample in metropolitan areas. It cannot be used to make state level estimates. Data users are advised to use caution with this variable, because it is masked for some sample to prevent disclosure.

[5] A list of states that make up each Census Region and Division can be found at https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf.

Geography Variable	Years	Description
COUNTY	1998-2011 Metro	This variable identifies county codes for sample in metropolitan areas. It cannot be used to make county level estimates. In some cases, the variable may refer to a combined set of counties. For more information on the county code, see the codebook. Data users are advised to use caution with this variable, because it is masked for some sample to prevent disclosure.
DEGREE	1997-2013 National	This variable categorizes units into one of six climate categories based on climate data from its county.
CMSA	1997-2013 National	The Consolidated Metropolitan Statistical Area (CMSA) is a grouping of two or more metropolitan statistical areas. Users are advised that these are not representative in AHS.
SMSA	1997-2013 National and Metro	This code is the primary identifier of a metropolitan area in both the national and metro files. It can be used to make estimates of the indicated area on metro PUFs, but users are advised caution when using SMSA on the national PUFs. For national, SMSA is only representative for "big 6" sample in years where the oversample is on the file, and in those cases, it sometimes needs to use a collection of SMSA codes.
METRO	1998-2013 Metro	Identifies the primary and up to five secondary central cities.
METRO3	1997-2013 National	Recode identifying combinations of whether the unit is in a metro area, in a central city, and whether it is urban or rural.
PMSA	1998 Metro	Based on the 1990 Primary Metropolitan Statistical Area code for the metropolitan areas surveyed in 1998.
PSUDOTCT	2002-2009 Metro	Contains a number representing a pseudo Census tract for the year in which the sample was originally drawn. This is not present in all years. Generally speaking, this variable does not correspond with actual Census tracts, so it should only be used to identify sample that may be clustered together geographically, not to make tract level estimates.

Geography Variable	Years	Description
ZONE	1998-2011 Metro	Area of greater than 100,000 population constructed by HUD and usually based on central cities of the OMB MSA or PMSA at the time the sample was drawn. However, when possible, the ZONE codes were updated to reflect current central cities and their respective boundaries. Zone definition files are available for 2007-2013 on the AHS Metropolitan PUF web pages[6].

Core and Rotating Topics

The 1997-2013 AHS questionnaires are organized into topics, where a topic is a set of related questions about a particular subject matter. HUD and the Census Bureau group all the topics into two types—core and rotating. Core topics are administered in every survey. In some cases, questions in the core topics may be edited or deleted, or HUD and the Census Bureau may add a question to a core topic. When a question is edited, removed, or added to a core topic, the change is intended to be permanent.

Rotating topics may appear in one survey, not appear in the next survey, but reappear in a future survey. AHS users interested in reviewing the specific questions in the rotating topics can review the "items booklet" for each survey year located in the Questionnaires section at https://www.census.gov/programs-surveys/ahs/tech-documentation/def-errors-changes.html.

To maximize the number of rotating topics that can be included in a survey year, beginning with the 2013 AHS, HUD and the Census Bureau adopted a strategy of splitting the AHS sample in two groups, then administering some topics to Group 1 and some topics to Group 2. To produce estimates with the 2013 rotating topics, PUF users must use a special split sample weight pertaining the rotating topic group. These weights are discussed in the weighting section of this document.

More information on rotating topics can be found in the Topical Module History Report 2009-2019 located at https://www.census.gov/programs-surveys/ahs/tech-documentation/help-guides/topmod_hist_report_0919.html. This document also includes information on the reason behind HUD's decisions to include particular topical modules.

The following table lists the rotating topics, their groups for split weighting purposes, and the AHS Interactive Codebook topic (or category) in which they can be found.

Survey Year	Topic	Topical Module Group	Topic in the Codebook Database
2009	Hurricane Katrina	Not Split	New Orleans Variables
2011	Healthy Homes	Not Split	Healthy Homes
2011	Home Accessibility	Not Split	Accessibility
2013	Public Transportation	Group 1	Public Transportation
2013	Community Involvement	Group 2	Neighborhood, Neighborhood Social Capital Subtopic

[6] Shapefiles can be found at https://www.census.gov/programs-surveys/ahs/data.html by selecting the year of interest from the tabs at the top (2007-2013), clicking the link to the AHS Public Use File (PUF), selecting Metro PUF, then scrolling to the Mapping Files section.

Survey Year	Topic	Topical Module Group	Topic in the Codebook Database
2013	Doubled-Up Households	Not Split	Doubling Up
2013	Supplemental Neighborhood Quality	Group 2	Neighborhood Features (Includes some core)
2013	Disaster Preparedness	Group 1	Disaster Planning
2013	Delinquent Payments and Notices	Not Split	Delinquency

Weights

Historically, the PUFs have included multiple weighting options, however this has often caused confusion for data users as only one weight is generally used for nearly all estimates

To resolve this, Census deleted some weighting variables from the 1997-2013 PUFs during the May 2021 revisions so that only the recommended weights are on the files. The table below lists all weighting variables originally published as part of 1997-2013 along with a description summarizing how they should be used and whether they were altered in the PUF revisions.

Variable Name	PUF Availability	Description
WEIGHT[7]	1997-1999 National	This is the primary weight used to replicate published estimates.
WGT90GEO	2001-2013 National	This is the primary weight used to replicate published estimates.
WEIGHT	2001-2013 National	This was a continuation of the weighting methodology used for the 1997-1999 National files. WEIGHT was removed from the 2001-2013 PUFs during the revisions and replaced by WGT90GEO to generate official estimates.
WEIGHT	1998-2013 Metro	This is the primary weight used to replicate published estimates for the independent metropolitan sample (excludes "Big 6" metro areas that are supplemental to the national sample).

[7] Weights used to generate national estimates can also be used to generate applicable "big 6" supplemental metro estimates in years they are represented in the national file. However, users are advised that differences caused by disclosure avoidance methods that mask the SMSA value for some cases could contribute to larger differences than what they may see in a national estimate.

Variable Name	PUF Availability	Description
PWT	1997-2013 National and Metro	This is a basic or "pure" weight that reflects the inverse probability of selection without additional adjustments. Advanced users can use this information to create custom weights, including longitudinal weighting. To reduce confusion, this variable was removed from the PUFs during revisions, but remains on the Sample Case History files.
SPLTWGT1	2013 National	National level weight for estimates using 2013 topical public transportation and disaster planning variables.
SPLTWGT2	2013 National	National level weight for estimates using 2013 topical neighborhood and social capital variables.
REPWGT1-160	2011-2013 National and Metro	Replicate weights used for variance estimation for 2011 and 2013. For more information, see the variance estimation guide at https://www.census.gov/programs-surveys/ahs/tech-documentation/help-guides/quickguide_repweight.html.
SUBWGT, SUBWGTREP1-160	2011-2013 National	National subsidized renter weight and its replicate weights
SP1SUBWGT, SP1SUBWGTREP1-160	2013 National	National subsidized renter weight and its replicate weights for estimates using topical public transportation and disaster planning variables.
SP2SUBWGT, SP2SUBWGTREP1-160	2013 National	National subsidized renter weight and its replicate weights for estimates using topical neighborhood and community involvement variables.
WGTMETRO	Combined 2011 National and Metro File	This weight was originally included in the combined 2011 national and metro file because it already contained the national WEIGHT variable, but was renamed back to WEIGHT when the 2011 files were revised and split into separate national and metro PUFs.

Regarding weights, 1997-2013 AHS PUF users should note three other pieces of information—

- Weights are assigned only at the household level; there are no person-level weights present in the AHS. For additional information about how the weights were constructed, users should consult the "Sample Design, Weighting, and Error Estimation" documents published with each year's PUFs[8].
- In 2013, for some AHS respondents in topical modules that are asked of only one group in the split sample, the value of SPLTWGT1 or SPLTWGT2 will be 0, but the value of WGT90GEO will be positive. This happens when the respondent answers enough survey questions to be considered a completed interview, but does not fully respond to questions in the topical modules to be considered sufficient for publication purposes.
- In general, nearly all STATUS 1-3 housing units have a positive weight on the PUF. However, a small number of housing units are given a positive weight only when regularly occupied, but a weight of 0 if vacant. This occurs with certain types of housing units such as a house boats or RVs that may not be considered a housing unit when vacant. Beginning in 2015, these types of units were set to a Type B non-interview when not occupied.
- For additional information about how replicate weights were constructed, users should consult "Guide to Estimating Variances for the American Housing Survey" located at https://www.census.gov/programs-surveys/ahs/tech-documentation/help-guides/quickguide_repweight.html.

Edits and Imputation

In some instances, AHS respondents provide answers that HUD and the Census Bureau believe are in error due to conflicts with other responses. Additionally, AHS respondents may not provide a response to certain questions if they do not know or refuse to answer.

In the case of an inconsistent response, for example if a respondent reports the year in which they moved into the unit being before their year of birth, Census may edit their year moved to match their birth year. This is referred to as a consistency edit.

When a respondent does not give a response, the value for the AHS variable may be imputed to fill in missing values. Variables that are edited or imputed are released with an edit flag, which is listed for each variable in the AHS Interactive Codebook. Edit flags in the 1997-2013 PUFs always start with a "J", but unlike the 2015 and later files, it will not always be followed by the exact name of the base variable and will not exist on the file unless the variable is edited or if it is created from multiple other variables with their own edit flags.

Edit flag variables in the 1997-2013 files will generally follow the response codes below.

Blank or -9: No edit or allocation
1: Variable was edited (consistency edits)
2: Variable was imputed (hot deck)
3: Variable was imputed (regression)
4: Variable was both edited and imputed

[8] Sample Design, Weighting and Error Estimation documents can be found in the Accuracy of the Data section at: https://www.census.gov/programs-surveys/ahs/tech-documentation/def-errors-changes.html.

Unlike 2015 and later, many recoded variables will not have an edit flag. For this reason, researchers using the 1997-2013 PUFs are advised to consider using the edit flags for variables that are part of the recode (such as individual income type variables in the case of the household income recode).

Additional details on editing and imputation in this time period, including an overview of the different methods, can be found in the Item Nonresponse and Imputation: 1997-2013 document located at https://www.census.gov/programs-surveys/ahs/tech-documentation/help-guides/1985-2013/nonresponse_imputation_2013.html.

Topcoding and Disclosure

To ensure respondent confidentiality, the Census Bureau requires that disclosure avoidance procedures be applied to certain PUF variables. For some variables, this means they cannot be published on the PUFs in any form. This is mentioned in the Internal Use File and Geography sections of this document.

A handful of other disclosure avoidance procedures are applied to 1997-2013 PUF variables. These procedures include collapsing into larger categories, adding noise to some variables, and topcoding. To maintain confidentiality, unusually high (and, in a few cases, low) values in the AHS dataset are replaced with maximum (or minimum) values, called top (or bottom) codes. Between the years of 2003 and 2013, the AHS PUF web pages contain a Topcodes file that documents the specific values for affected variables[9]. For some variables in the 1985-2013 PUFs, the original variable is retained in the IUF.

For additional information about disclosure avoidance methods, users should consult the 1985-2013 disclosure avoidance documentation available at https://www.census.gov/content/dam/Census/programs-surveys/ahs/tech-documentation/1985-2013/Disclosure_Avoidance_1985_to_2013.pdf.

Value Label Packages

The 1985-2013 PUFs include a significant number of categorical variables. When using the PUFs, adding descriptions to a categorical variable's values can be a labor-intensive task. For example, a data user might want to add the descriptive names ("mobile homes", "single-family detached", "single-family attached") that correspond to a categorical variable's values ("1", "2", "3") in order to display each value's actual meanings in tabulated output.

For this reason, HUD and Census publish a value label package with each PUF. The package is a zip file that includes instructions for applying value labels to categorical variables and a spreadsheet that maps categorical variable values to descriptive names. Although the instructions are generally for SAS users, non-SAS users should be able to use the contents of the Value Labels Package.

Though value label packages existed for the original 1997-2013 PUFs, they were not always consistently formatted and were often manually maintained, which led to variables missing entries in some years. The 1997-2013 PUF revisions applied a standard Value Label Package format, which pulls directly from the Interactive Codebook response descriptions.

[9] Topcodes files can be found at https://www.census.gov/programs-surveys/ahs/data.html by selecting the year of interest from the tabs at the top (2003-2013), clicking the link to the AHS Public Use File (PUF), selecting National PUF or Metro PUF, then scrolling to the Topcodes section.

Appendix

Variables with Type Changes During Revisions

Variable Name	Variable Description	Type Change	Years
AIR	Room air conditioner	Numeric to Character	2013N, 2013M, 2011N, 2011M, 2009N, 2009M, 2007N, 2007M, 2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
APTFL	Number of floors in the apartment itself	Character to Numeric	1998M, 1997N
CELLAR	Unit has a basement	Numeric to Character	2013N, 2013M, 2011N, 2011M, 2009N, 2009M, 2007N, 2007M, 2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
DBIN1REAS	Reason first in-mover group left previous home	Numeric to Character	2013N, 2013M
DBINREAS	Most important reason person left previous home	Numeric to Character	2013N, 2013M
DBOUT1REAS	Reason first out-mover group left	Numeric to Character	2013N, 2013M
DBOUTREAS	Reason former co-occupant left home	Numeric to Character	2013N, 2013M
FCOKST	Fuel used by cookstove	Numeric to Character	2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
FFRPL	Fuel used by fireplace without inserts	Numeric to Character	2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
FFRPLI	Fuel used by fireplace with inserts	Numeric to Character	2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
FPLWK	Unit has useable fireplace	Numeric to Character	2013N, 2013M, 2011N, 2011M, 2009N, 2009M, 2007N, 2007M, 2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
FPORTH	Fuel used by portable electric heaters	Numeric to Character	2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
FRSIT	First site mobile home ever placed on	Numeric to Character	2013N, 2013M, 2011N, 2011M, 2009N, 2009M, 2007N, 2007M, 2005N, 2004M, 2003N, 2002M, 2001N, 1999N

Variable Name	Variable Description	Type Change	Years
GRAD	Educational level of person	Numeric to Character	2013N, 2013M, 2011N, 2011M, 2009N, 2009M, 2007N, 2007M, 2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
HEINF1	Current interest rate on first home equity line of credit - fraction	Numeric to Character	1997N
HEINF2	Current interest rate on second home equity line of credit - fraction	Numeric to Character	1997N
HEINF3	Current interest rate on third home equity line of credit - fraction	Numeric to Character	1997N
HEQUIP	Main heating equipment	Numeric to Character	2013N, 2013M, 2011N, 2011M, 2009N, 2009M, 2007N, 2007M, 2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
HHGRAD	Educational level of householder	Numeric to Character	2013N, 2013M, 2011N, 2011M, 2009N, 2009M, 2007N, 2007M, 2005N, 2004M, 2003N, 2002M, 2001N
HHMVG	Recent mover group of householder	Character to Numeric	2013N, 2013M, 2011N, 2011M, 2009N, 2009M, 2007N, 2007M, 2005N, 2004M, 2003N, 2002M, 2001N
HHNATVTY	Country of birth of householder	Numeric to Character	2001N
HHREL	Relationship of householder to household	Numeric to Character	2013N, 2013M, 2011N, 2011M, 2009N, 2009M, 2007N, 2007M, 2005N, 2004M, 2003N, 2002M, 2001N
HJOB	Reports to work at same place each day	Numeric to Character	1997N
HKDAMCSM	Estimated construction start month	Numeric to Character	2009M
HKDAMGNM	Estimated repair completion month	Numeric to Character	2009M
HKDAMGYM	Estimated construction completion month	Numeric to Character	2009M
HKDAMRSM	Estimated repair start month	Numeric to Character	2009M
HKMOVR	Forced to move because of Hurricane Katrina	Numeric to Character	2009M

Variable Name	Variable Description	Type Change	Years
HKNORES	Living in New Orleans during Hurricane Katrina	Numeric to Character	2009M
HKPERM	Permanently settled	Numeric to Character	2009M
HKSHAR1	Lived in shared residence since Hurricane Katrina	Numeric to Character	2009M
LOONCL	Percent of first mortgage to be paid in the final balloon payment - calculated	Numeric to Character	2013N, 2013M, 2011N, 2011M
MGRESA	Shopped around before choosing lender for first mortgage	Numeric to Character	2011N, 2011M
MGRESA2	Shopped around before choosing lender for second mortgage	Numeric to Character	2011N, 2011M
MGRESA3	Shopped around before choosing lender for third mortgage	Numeric to Character	2011N, 2011M
MOVGRP	Recent mover group	Character to Numeric	2009N, 2009M, 2007N, 2007M, 2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
MVG	Recent mover group	Character to Numeric	2013N, 2013M, 2011N, 2011M, 2009N, 2009M, 2007N, 2007M, 2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
NATVTY	Country of birth of person	Numeric to Character	2001N
NUMBLOW	Number of times fuses blown or breakers tripped, past 90 days	Character to Numeric	2013N, 2013M, 2011N, 2011M, 2009N, 2009M, 2007N, 2007M, 2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
NUMCOLD	Number of times heating equipment broke down	Character to Numeric	2013N, 2013M, 2011N, 2011M, 2009N, 2009M, 2007N, 2007M, 2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N

Variable Name	Variable Description	Type Change	Years
NUMTLT	Number of toilet breakdowns, 6 hours or more	Character to Numeric	2013N, 2013M, 2011N, 2011M, 2009N, 2009M, 2007N, 2007M, 2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
OAFUEL	Fuel used by other central air conditioning unit	Numeric to Character	2013N, 2013M, 2011N, 2011M, 2009N, 2009M, 2007N, 2007M, 2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
PERSINT	Personal / telephone interview	Numeric to Character	1997N
REL	Relationship of person to reference person	Numeric to Character	2013N, 2013M, 2011N, 2011M, 2009N, 2009M, 2007N, 2007M, 2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
REUAD	Reason unit was added to sample	Numeric to Character	2013N, 2013M, 2011N, 2011M, 2009N, 2009M, 2007N, 2007M, 2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
SAFEU5KD	Children under five live in or regularly visit the household	Numeric to Character	2011N, 2011M
SMSA	1980 design PMSA code	Numeric to Character	2009N, 2007N, 2005N, 2003N, 2001N, 1999N, 1997N
TPARK	Number of mobile homes in group	Numeric to Character	2013N, 2013M, 2011N, 2011M, 2009N, 2009M, 2007N, 2007M, 2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
TRAN	Method of transportation to work	Numeric to Character	2009N, 2009M, 2007N, 2007M, 2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1997N
TYPE	Housing unit type	Numeric to Character	2013N, 2013M, 2011N, 2011M, 2009N, 2009M, 2007N, 2007M, 2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
VACANCY	Vacancy status	Numeric to Character	2013N, 2013M, 2011N, 2011M, 2009N, 2009M, 2007N, 2007M, 2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
WHYMOVE	Main reason moved	Numeric to Character	2013N, 2013M, 2011N, 2011M, 2009N, 2009M, 2007N, 2007M, 2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
WLINEQ	Worked at all last week	Numeric to Character	1997N

Variable Name	Variable Description	Type Change	Years
ZMVGRP	Total number of mover groups in household	Character to Numeric	2013N, 2013M, 2011N, 2011M, 2009N, 2009M, 2007N, 2007M, 2005N, 2004M

Variables Dropped During Revisions

Variable Name	Reason Dropped	Years
ACCESS	Non-interview variable	2013N, 2013M, 2011N, 2011M, 2009N, 2009M, 2007N, 2007M, 2005N, 2003N, 2002M, 2001N, 1999N, 1998M
ADDTN3	No valid data	2013N, 2013M, 2011M
ADJDEP3	No valid data	2013N, 2013M
ADJFIX3	No valid data	2013M
ADJRTF3	No valid data	2013M
AMMRT4	No valid data	2013N, 2013M, 2009M, 1998M
AMTM2	No valid data	2013M, 2009M
AMTM3	No valid data	2013N, 2013M
APTAD	No valid data	2002M, 1998M
APTCH	No valid data	2002M, 1999N, 1998M, 1997N
APTCM	No valid data	2002M, 1998M
APTSP	No valid data	2002M, 1998M
ARM3	No valid data	2013M
BALAMT3	No valid data	2013N, 2013M, 2011N, 2011M
BBLDG	Non-interview variable	2013N, 2013M, 2011N, 2011M, 2009N, 2009M, 2007N, 2007M, 2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
BEDX	No valid data	1998M
BILLEF	No valid data	2011N, 2011M, 2009N, 2009M, 2007N
BILLEG	No valid data	2011N, 2011M, 2009N, 2009M, 2007N
BILLEO	No valid data	2011N, 2011M, 2009N, 2009M, 2007N
BILLET	No valid data	2011N, 2011M, 2009N, 2009M, 2007N
BILLEW	No valid data	2011N, 2011M, 2009N, 2009M, 2007N
BILLFE	No valid data	2009M, 2007M
BILLFG	No valid data	2009M, 2007M, 2003N, 2002M, 1999N, 1998M
BILLFO	No valid data	2009N, 2009M, 2007M, 2005N, 2004M
BILLFT	No valid data	2011N, 2011M, 2009N, 2009M, 2007N, 2007M, 2005N, 2002M, 2001N
BILLFW	No valid data	2011N, 2011M, 2009N, 2009M, 2007N, 2007M, 2001N
BILLGE	No valid data	2007M
BILLGF	No valid data	2011N, 2011M, 2009N, 2009M, 2007N, 2007M, 2003N, 2002M, 1998M, 1997N
BILLGO	No valid data	2011N, 2011M, 2009N, 2009M, 2007N, 2007M

Variable Name	Reason Dropped	Years
BILLGT	No valid data	2011N, 2011M, 2009N, 2009M
BILLGW	No valid data	2011N, 2011M, 2009N, 2009M, 2007N, 2007M, 2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
BILLOE	No valid data	2009M, 2007M
BILLOF	No valid data	2011N, 2011M, 2009N, 2009M, 2007N, 2007M, 2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
BILLOG	No valid data	2009M, 2002M
BILLOT	No valid data	2011N, 2011M, 2009N, 2009M, 2007N, 2001N
BILLOW	No valid data	2011N, 2011M, 2009N, 2009M, 2007N, 2007M, 2002M, 1998M
BILLTF	No valid data	2011N, 2009N, 2009M, 2007M, 2005N, 2004M, 2002M, 1999N, 1997N
BILLTO	No valid data	2009M, 2007M, 2005N, 2003N, 2002M, 2001N, 1999N, 1998M
BILLTW	No valid data	2011N, 2011M, 2009N, 2009M, 2007N, 2007M, 2005N, 2004M, 2003N, 2002M, 2001N, 1998M, 1997N
BILLWF	No valid data	2009N, 2009M, 2005N, 2004M, 2003N, 2002M
BILLWO	No valid data	2009N, 2007M, 2002M, 2001N, 1999N, 1998M, 1997N
BLOON2	No valid data	2009M
BLOON3	No valid data	2013M
BOARDU	Non-interview variable	2013N, 2013M, 2011N, 2011M, 2009N, 2009M, 2007N, 2007M, 2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
CASH3	No valid data	2013N, 2013M, 2011N, 2011M
CHANGEDRECORD	Processing variable	2007N, 2005N
CROPSL	No valid data	2009M, 2004M, 2002M
DAFUR	No valid data	2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
DCOKST	No valid data	2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
DELECT	No valid data	2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
DFIRE	Non-interview variable	1998M, 1997N
DFLOT	No valid data	2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
DFRPL	No valid data	2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
DFRPLI	No valid data	2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
DHOTH	No valid data	2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
DHPMP	No valid data	2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N

Variable Name	Reason Dropped	Years
DISAS	Non-interview variable	2013N, 2013M, 2011N, 2011M, 2009N, 2009M, 2007N, 2007M, 2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
DPLF	No valid data	2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
DSTEAM	No valid data	2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
DSTOVE	No valid data	2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
EXPOSE	Non-interview variable	2013N, 2013M, 2011N, 2011M, 2009N, 2009M, 2007N, 2007M, 2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
EXTLN2	No valid data	2009N, 2009M, 2005N
FIXED3	No valid data	2013M
FRSTRM3	No valid data	2013M
GPMWP3	No valid data	2013M
HALB	No valid data	1998M, 1997N
HEBAL3	No valid data	2009M, 2007N, 2007M, 2005N, 2004M, 2003N, 2002M, 2001N, 1999N
HEBAM2	No valid data	2009M
HEBAM3	No valid data	2013N, 2009N, 2009M, 2007N, 2007M, 2005N, 2004M, 2003N, 2002M, 2001N
HECR2	No valid data	2009M
HECR3	No valid data	2013N, 2009N, 2009M, 2007N, 2007M, 2005N, 2004M, 2003N, 2002M, 2001N, 1999N
HELMP3	No valid data	1997N
HEPMT2	No valid data	2009M
HEPMT3	No valid data	2013N, 2009N, 2009M, 2007N, 2007M, 2005N, 2004M, 2003N, 2002M, 2001N
HISBUILT	Processing variable	2004M
HISTRY	No valid data	1998M
HPMP	No valid data	2009M
HYBARM3	No valid data	2013M
HYBMYR3	No valid data	2013N, 2013M, 2011M
IMPRV3	No valid data	2013N
INCPR2	No valid data	2009M
IO3	No valid data	2013M
LISCH	No valid data	2002M, 1998M
LOON2	No valid data	2009M
LOONCL3	No valid data	2013N, 2013M, 2011N, 2011M
MAXADJ3	No valid data	2013M
MXDJTM3	No valid data	2013N, 2013M, 2011N, 2011M

Variable Name	Reason Dropped	Years
MXINTF3	No valid data	2013N, 2013M, 2011N
MXINTR3	No valid data	2013N, 2013M, 2011N
MXINTW3	No valid data	2013N, 2013M, 2011N
NONINT	Non-interview variable	2013N, 2013M, 2011N, 2011M, 2009N, 2009M, 2007N, 2007M, 2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
OLDMSFLG	No valid data	2011N
ORINTF3	No valid data	2013M
ORINTR3	No valid data	2013M
ORINTW3	No valid data	2013M
OTBUP	No valid data	2007N, 2007M
PMAMT2	No valid data	2009M
PMAMT3	No valid data	2013N, 2013M, 2011N, 2011M
PMT4	No valid data	2013N, 2013M, 2009M, 1998M
PTCHAM3	No valid data	2013M
PWT	Redundant weighting variable	2013N, 2013M, 2011N, 2011M, 2009N, 2009M, 2007N, 2007M, 2005N, 2004M, 2003N, 2002M, 2001N, 1999N, 1998M, 1997N
R_SAMEHH	Processing variable	1997N
REDPA2	No valid data	2007M
SAMEDU	No valid data	1998M
SHOCK3	No valid data	2013N
STATUS	No valid data	1998M
TADJ	No valid data	2011M, 2009M, 2007M, 2004M, 2002M
TALWIR	No valid data	2009M, 2007M, 2004M, 2002M
TASB	No valid data	2009M, 2007M, 2004M, 2002M
TIMBOM3	No valid data	2013M
TPBPAI	No valid data	2011M, 2009M, 2007M, 2004M, 2002M
TPBPIP	No valid data	2011M, 2009M, 2007M, 2004M, 2002M
TPBSOL	No valid data	2011M, 2009M, 2007M, 2004M, 2002M
TRADON	No valid data	2009N, 2007M, 2004M, 2002M
TREP	No valid data	2011M, 2009M, 2007M, 2004M, 2002M
TWATER	No valid data	2009M, 2007M, 2004M, 2002M
UNPBAL4	No valid data	2013N, 2013M
USFCAM	No valid data	1999N, 1998M
USFCHG	No valid data	1998M
VARM3	No valid data	2013M
WINUS	No valid data	2007N, 2007M

Variable Name	Reason Dropped	Years
WEIGHT	Redundant weighting variable	2013N, 2011N, 2009N, 2007N, 2005N, 2003N, 2001N

Example SAS Code

This section provides some examples of using AHS data so that you can see what you need to do. Each example begins with a description of the general procedure, followed by the SAS code for users of that software

Example 1: Generating a Table

To build a simple crosstabulation of AHS data, you need to specify the variables to tabulate and the weighting variable. STATUS is the interview status of a sample case. NUNIT2 is the type of structure. They take these values (as you can find in the Interactive Electronic Codebook):

STATUS
Interview status

1: Occupied interview
2: URE (Usual Residence Elsewhere) interview
3: Vacant interview
4: Noninterview

NUNIT2
Are these living quarters in a…

1: One-unit building, detached from any other building
2: One-unit building, attached to one or more buildings
3: Building with two or more apartments
4: Manufactured (mobile) home
B or -6: Not applicable

An example of SAS code that can be used to generate the table is below:

```
PROC FREQ DATA=AHS2013N.HOUSEHOLD; /* This line specifies which dataset to use as input. */
TABLE STATUS*NUNIT2; /* This section specifies which variables to use in the tabulation. */
WEIGHT WGT90GEO; /* This section specifies the weighting variable. */
RUN;
```

Running the code above gives the following output.

The FREQ Procedure

Table of STATUS by NUNIT2

STATUS(Interview status) NUNIT2(Structure type)

Frequency Percent Row Pct Col Pct	1	2	3	4	Total
1	7.432E7 55.95 64.15 89.12	6619456 4.98 5.71 87.32	2.799E7 21.07 24.16 84.17	6917166 5.21 5.97 80.40	1.159E8 87.22
2	1150865 0.87 62.84 1.38	99266.7 0.07 5.42 1.31	437491 0.33 23.89 1.32	143917 0.11 7.86 1.67	1831539 1.38
3	7918751 5.96 52.27 9.50	862098 0.65 5.69 11.37	4825916 3.63 31.86 14.51	1542157 1.16 10.18 17.93	1.515E7 11.40
Total	8.339E7 62.78	7580820 5.71	3.326E7 25.04	8603240 6.48	1.328E8 100.00

Example 2: Linking Files

Another common procedure when using the AHS PUFs is to link variables referring to the same housing unit that exist on two different files. In this example, we create a file called MATCH that has both the 2011 and 2013 interview status variables for any housing unit that exists on either the 2011 or 2013 national HOUSEHOLD dataset.

The code below uses an SQL statement to merge the files, then creates a list tabulation of the 2011 and 2013 STATUS variables.

```
PROC SQL;
   CREATE TABLE MATCH AS /* Specify the name of the output dataset */
SELECT AHS11.STATUS AS STATUS2011,
AHS13.STATUS AS STATUS2013 /* Rename the STATUS variable on the 2011 and 2013 files. */
FROM AHS2011N.HOUSEHOLD AHS11
FULL JOIN AHS2013N.HOUSEHOLD AHS13 /* Specify files to link. */
        ON AHS11.CONTROL = AHS13.CONTROL; /* Files are linked by CONTROL. */
QUIT;

PROC FREQ DATA=MATCH;
        TABLES STATUS2011*STATUS2013 / LIST MISSING; /* Create list table of the variables. */
RUN;
```

After running the code above to link the 2011 and 2013 HOUSEHOLD files by CONTROL, the FREQ procedure generates the following output.

The FREQ Procedure

STATUS2011	STATUS2013	Frequency	Percent	Cumulative Frequency	Cumulative Percent
	1	15355	19.50	15355	19.50
	2	199	0.25	15554	19.75
	3	2193	2.78	17747	22.54
1		7427	9.43	25174	31.97
1	1	41355	52.51	66529	84.48
1	2	224	0.28	66753	84.77
1	3	3021	3.84	69774	88.60
2		103	0.13	69877	88.73
2	1	275	0.35	70152	89.08
2	2	182	0.23	70334	89.31
2	3	383	0.49	70717	89.80
3		1176	1.49	71893	91.29
3	1	3112	3.95	75005	95.24
3	2	276	0.35	75281	95.59
3	3	3469	4.41	78750	100.00

Note that the above output is unweighted and includes missing values. STATUS does not have any missing values in any survey year, so we know that the missing values in the above frequency mean that the sample unit does not exist in that year. For example, the first line can be interpreted as there being 15,355 sample units that exist on the national 2013 HOUSEHOLD file with a STATUS of 1 which don't have a record on the 2011 national HOUSEHOLD file.

Files can be linked across years when they are longitudinal, or between datasets in the same year.