

American Housing Survey

Getting Started With the Public Use File: 2015 and Beyond

Updated June 2019

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

Contents

| | |
|---|----|
| Release Notes for 2017 (V3.0)..... | 3 |
| Overview..... | 4 |
| A Brief Review of the 2015 Sample Design..... | 4 |
| PUF Availability..... | 5 |
| PUF Relational File Structure..... | 5 |
| PUF Flat File Structure..... | 6 |
| PUF Codebook Interactive Tool..... | 6 |
| PUF and Internal Use File..... | 7 |
| Type of Respondents Included in the PUF..... | 8 |
| Variable Types and Names in the PUF..... | 8 |
| Using the PUF to Replicate Estimates in the AHS Summary Tables..... | 9 |
| Coding of “Not Applicable” and “Not Reported” in the PUF..... | 9 |
| Geographic Indicators in the PUF..... | 10 |
| Core and Rotating Topical Modules..... | 10 |
| Weights in the PUF..... | 11 |
| Edits and Imputations..... | 12 |
| Topcoding and Other Disclosure Avoidance Techniques..... | 12 |
| Value Labels Package..... | 15 |

Release Notes for 2017 (V3.0)

The 2017 AHS PUF does not include the Eviction topical module. HUD and Census intend to release these data at a later date.

Overview

The purpose of this document is to introduce American Housing Survey (AHS) users to the 2015 and later AHS public use file (PUF) microdata. AHS users who have used prior-year PUFs (that is, 2011, 2013, and so on) may notice similarities between the more recent PUF and prior-year PUFs. However, there are important differences, including—

- Integration of the 15 largest metropolitan areas into the National sample PUF.
- Fewer tables in the relational PUF structure.
- Fewer weights.
- Name changes to many variables.
- Removal of noninterviews from the PUF.
- New interactive codebook (AHS Codebook Interactive Tool) and reorganization of codebook variables into topics and subtopics.

The remainder of this guide is organized into sections, with each section addressing an important PUF topic. If applicable, comparisons will be made with prior-year PUFs.

Using the PUF requires a statistical program such as SAS, STATA, or R. Although it is technically possible to use the PUF in Microsoft Excel or Access, users will find doing so difficult. 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 Data Analysis Laboratory (Pandas).

A Brief Review of the 2015 Sample Design

For 2015, the U.S. Department of Housing and Urban Development (HUD) and the U.S. Census Bureau drew an entirely new sample for the AHS. The 2015 AHS sample is composed of an integrated National sample and independent metropolitan area samples. The National sample is described as integrated, because it incorporates a few different types of samples. The integrated National sample includes—

- A representative sample of the nation.
- Representative oversamples of each of the 15 largest metropolitan areas.
- A representative oversample HUD-assisted housing units.

HUD and the Census Bureau’s intent is to survey the entire integrated National sample once every 2 years. As such, it is a longitudinal panel with a 2-year survey cycle.

The independent metropolitan area samples include representative samples of selected metropolitan areas. For 2015, the 10 selected metropolitan areas represent one-half of what HUD and the Census Bureau refer to as the “Next 20” group of metropolitan areas (the second half was included in the 2017 AHS). The Next 20 group of metropolitan areas is a subset of

metropolitan areas ranging from 16th to 51st largest by population.¹ HUD and the Census Bureau intend to survey each member of the Next 20 group of metropolitan areas once every 4 years. As such, the Next 20 group of independent metropolitan area samples is a longitudinal panel with a 4-year survey cycle.

PUF Availability

Similar to prior-year PUFs, separate PUFs will be published for the integrated National sample and the independent metropolitan area samples. These PUFs are referred to as “National PUF” and “Metropolitan Area PUF.”

As in recent years, HUD and the Census Bureau are providing the PUF in two formats—relational and flat. The structure of the PUFs is discussed in the next two sections.

Also as in recent years, all PUFs will be published in SAS and CSV formats. Beside 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.²

PUF Relational File Structure

The PUF relational format is similar to prior years’ format. However, the PUF includes fewer tables than in prior years. The following table describes each table within the relational structure.

| 2015 and later Relational Table Name | Description | Comparison With Prior-Year PUFs |
|--------------------------------------|---|--|
| HOUSEHOLD | Includes one record for each housing unit. Also includes certain householder information, such as race and sex. The householder information is also present on the PERSON table but is duplicated in the HOUSEHOLD table for ease of use. | <ul style="list-style-type: none"> • Prior-year tables NEWHOUSE, REPWGT, RATIOV, TOPICAL, and OWNER are now included in HOUSEHOLD. • Prior-year table RMOV has been flattened and integrated into HOUSEHOLD. Up to three mover groups are presented in the HOUSEHOLD table. Variable prefixes denote |

¹ For more information about how the Next 20 group of metropolitan areas was selected, see “Metropolitan Area Selection Strategy: 2015 and Beyond” on the Census Bureau’s AHS website (“2015 Redesign” portion of the “Operations and Administration” page).

² See the subsequent section, Coding of “Not Applicable” and “Not Reported” in the PUFs.

| | | |
|----------|---|---|
| | | the mover group number (that is MVG1XXXX). |
| PERSON | Includes one record for each person in the household, up to 16 persons. For 2015 and forward, each person in the PERSON table receives a unique person number (PLINE), and this number will be unique for the life of the longitudinal sample. In other words, if a person moves out of the housing unit, that PLINE number will not be reused. | Prior-year table PERSON is functionally equivalent to the new PERSON table. |
| MORTGAGE | Includes one record for each mortgage or similar debt owned by the household. The PUF includes the first three mortgages (values 1–3); the IUF includes all mortgages reported by the respondent (up to 5). | Prior-year table MORTG was organized as a flat file, with one record per housing units and replicate variable sets to accommodate more than one loan. |
| PROJECT | Includes one record for each home improvement project undertaken by the household within the past 2 years. At most, there are 26 records per household, reflecting up to 26 projects. | Prior-year table HOMIMP is functionally equivalent to the new PROJECT table. |

IUF = internal use file. PUF = public use file.

The PERSON, MORTGAGE, and PROJECT 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.

PUF Flat File Structure

The flat file structure includes the HOUSEHOLD table combined with flattened versions of the PERSON, MORTGAGE, and PROJECT tables. The flattened format is similar to the format in prior years. For instance, suppose a household has three persons. The variable RACE from the PERSON table in the relational structure will be flattened into RACE1, RACE2, and RACE3 in the flat file structure. The number appended to RACE reflects the number of the person. The same convention is used for the MORTGAGE and PROJECT tables.

PUF Codebook Interactive Tool

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 have 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. The 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.

Currently, the AHS Codebook Interactive Tool includes surveys conducted from 1997 through 2015.³ As such, the tool includes all variables that have appeared on any AHS PUF from 1997 through 2013, plus all new variables created for 2015 and later

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 particular topic and, if desired, 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 either the 2015 or later AHS or 2013–1985 AHSs. 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. |

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

PUF and Internal Use File

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. The AHS Codebook Database contains an “Availability” section that takes two possible values— PUF or IUF Only. When a variable is listed as IUF Only, it means the variable exists on the IUF but, for various reasons, is not on the PUFs.

If a AHS PUF user cannot complete an analysis because one or more AHS variables is 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 using IUF Only variables. Second, HUD and Census Bureau staff can assist AHS users with the process of obtaining access to the AHS IUF. However, AHS users should note this process requires—

- Submitting a detailed proposal.
- Submitting to a background check and obtaining a non-sensitive federal security clearance.

³ Work is in progress to expand the interactive codebook to include every year of the AHS, which began in 1973.

- Agreeing to become subject to legally binding disclosure restrictions.
- Willingness to travel to a Census Bureau Federal Statistical Research Data Center.⁴
- Willingness to pay a fee to access the Research Data Center.

Type of Respondents Included in the PUF

The AHS PUF includes occupied interviews, usual residence elsewhere interviews, and vacant interviews. The variable INTSTATUS indicates which of these three types of interviews was conducted. The 2015 and later AHS PUF does not include noninterviews, unlike some prior-year PUFs that did include noninterviews. HUD and the Census Bureau created the National Sample Case History File to track all sampled housing units in the integrated National sample and the Next 20 independent metropolitan area samples. AHS users who want to know why various sample cases were not surveyed or were nonresponses for 2015 (or any future year) can review this file.⁵

As in past years, noninterviews are classified as one of three types as 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 Types and Names in the PUF

The PUF includes more than 420 variables across the four relational tables, plus edit flags and replicate weights. The PUF variables come in two types—numeric and categorical. Numeric variables are those that have a natural representation as a number. Examples include counts, dollars, and percent values. All other variables that do not meet the test of having a natural representation as a number are classified as categorical variables.

Many of the PUF variables represent the same or similar concepts as in past years. However, most variable names have changed since 2013. Because the 2015 AHS includes an entirely new sample, HUD and the Census Bureau took the opportunity to change variable names in order to make them more user friendly and to bring them in line with other surveys, most notably the

⁴ For more information, see <https://www.census.gov/fsrdc>.

⁵ For more information, see <https://www.census.gov/programs-surveys/ahs/tech-documentation/help-guides/national-sample-case-history.html>.

American Community Survey (ACS). For example, several prior-year AHS variables relating to the type and size of the housing unit have been replaced with a new variable, BLD, which corresponds to the ACS variable BLD.

Another notable change beginning with 2015 is several instances in which multiple variables were collapsed into one variable. For instance, the 2015 PUF variable FINROOMS represents a combination of five variables found in the 2013 PUF and prior-year PUFs. A similar strategy was used for heating and cooling and other variables.

A few variables, such as REL, have the same name but use different codes.

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. The crosswalk will help users map new variable names to old variable names. This variable name crosswalk is available on the Codebook Interactive Tool website.

AHS PUF users are strongly encouraged to carefully review the variables in their analysis, especially categorical variables, to ensure proper understanding of what the values of the variables represent.

Using the PUF to Replicate Estimates in the AHS Summary Tables

A frequent question of AHS PUFs users is, “how can I replicate the results in the AHS summary tables?” In past years, HUD and the Census Bureau have published SAS code that provides instructions on how to replicate most summary table estimates. This code was referred to as “table specifications.”

Starting with the 2013 PUF, HUD and the Census Bureau created a new product called “[AHS summary table specifications](#).” This product represents a substantial improvement from the table specification published prior to 2013. It is organized to correspond exactly to each AHS summary table that is currently available in the AHS Table Creator. Moreover, extensive SAS code was replaced with more generalized code, so non-SAS users can better understand how to replicate various summary table estimates.

For 2015 and future years, HUD and the Census Bureau will continue to produce AHS summary table specifications. Estimates calculated using the PUFs might not be exactly the same as those published in the AHS summary tables due to disclosure avoidance adjustments applied to the PUFs.

Coding of “Not Applicable” and “Not Reported” in the PUF

Prior-year PUFs implemented a coding scheme when a variable did not have a response from a respondent. The coding scheme included four categories: “not applicable,” “not reported/missing,” “don’t know,” and “refused.” Users of prior-year AHS PUFs have routinely commented that they found the “not reported/missing” category to be confusing, especially when the values were blank or null values. Moreover, they did not understand the difference between “not reported” and “refused.”

Starting with 2015 the PUF includes a new coding scheme for nonresponse. There are two categories—“not reported” and “not applicable.” The “not reported” category means that the respondent did not provide an answer to the question, either because they refused or because they did not know. The “not reported” category is coded .M in the SAS files and -9 in the CSV files.

The “not applicable” category means the question was not asked of the respondent, because the question was not *in scope* for the respondent. For instance, if a respondent reported living in a single-family detached home, they would not receive questions related to mobile home features. The “not applicable” category is coded .N in the SAS files and -6 in the CSV files.

If AHS PUF users want to better understand why a response is recorded as “not applicable,” they should consult the Conditions in Universe section of the AHS Interactive Codebook. AHS PUFs users should also note that the Codebook now lists conditions *in* universe, a change from the previous codebook’s conditions *not in* universe.

Geographic Indicators in the PUF

The 2015 and later PUF will not include many of the geographic indicators included in prior-year PUFs due to disclosure avoidance procedures required by the Census Bureau.

The 2015 and later National PUF will include two geographic variables—Census Division and OMB13CBSA. The variable OMB13CBSA is the 2013 Office of Management and Budget Core Based Statistical Area (CBSA) code. However, this variable is populated with values for only the 15 largest metropolitan areas. The metropolitan area PUF will include OMB13CBSA but not Census Division.

HUD and the Census Bureau may add additional geographic variables at a later date, including adding additional CBSA codes to the variable OMB13CBSA.

As a final note, PUF users may notice that a handful of variables in the subtopic “Nearby Features” are not included in the PUF, although they are present in the IUF. Because these variables are inherently geographic, they pose a disclosure risk. As such, Census Bureau disclosure avoidance procedures require they not be included in the PUF.

Core and Rotating Topical Modules

The 2015 and later AHS questionnaires are organized into question modules, where a module is a set of related questions about a topic. HUD and the Census Bureau group all the modules into two types—core and rotating topical. Core modules are administered in every survey. In some cases, questions in the core modules may be edited or deleted, or HUD and the Census Bureau may add a question to a core module. When a question is edited, removed, or added to a core module, the change is intended to be permanent.

Rotating topical modules 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 topical modules can review the document “AHS Items Booklet.”

To maximize the number of rotating topical modules 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 topical modules to Group 1 and some modules to Group 2.

There is one important piece of information that users must consider when using variables derived from rotating topical module questions. To produce estimates with these variables, PUF users must use a special split sample weight pertaining the topical module group. These weights are discussed in the section on Weights in the PUF.

The following table lists the rotating topical modules, their topical module groups, and where the variables are located in the AHS Interactive Codebook.

| Survey Year | Topical Module | Topical Module Group | Topic in the Codebook Database |
|-------------|--------------------|----------------------|--------------------------------|
| 2015 | Arts and Culture | Group 1 | Arts |
| 2015 | Food Security | Group 1 | Food Security |
| 2015 | Housing Counseling | Group 1 | Housing Counseling |
| 2015 | Healthy Homes | Group 2 | Healthy Homes |
| 2017 | Commuting | Group 1 | Commuting |
| 2017 | Disaster Planning | Group 2 | Disaster Planning |

Weights in the PUF

The 2015 and later PUF will have three weighting variables, as well as corresponding replicate weights. The following table describes each weight and its corresponding replicate weights.

| Weighting Variable Name | Description | Replicate Weights |
|-------------------------|---|-----------------------------|
| WEIGHT | Use WEIGHT to produce National and metropolitan area estimates for all variables other than those found in topical modules. | REPWEIGHT1– REPWEIGHT160 |
| SP1WEIGHT | Use SP1WEIGHT to produce National and metropolitan area estimates for variables in the topical module Group 1. | SP1REPWGT1– SP1REPWGT160 |
| SP2WEIGHT | Use SP2WEIGHT to produce National and metropolitan area estimates for variables in the topical module Group 2. | SP2REPWGT1– SP2REPWGT160 |

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

- For some AHS respondents in topical module Group 1 (Group 2), the value of SP1WEIGHT (SP2WEIGHT) will be 0, but the value of WEIGHT 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 Group 1 (Groups 2) topical modules to be considered sufficient for publication purposes.

- 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 integrated National and independent metropolitan area “Sample Design, Weighting, and Error Estimation” documents.
- For additional information about how replicate weights were constructed, users should consult “Guide to Estimating Variances for the American Housing Survey.”

Edits and Imputations

In some instances, AHS respondents provide answers that HUD and the Census Bureau deem likely to be incorrect. When this happens, the AHS variable containing the responses is edited.

In some instances, AHS respondents do not provide answers to certain questions. In such cases, the value for the AHS variable may be imputed. It is important to note that imputations are not made for every variable with a missing value.

There are two resources AHS PUF users can use to further investigate edits and imputations. First, nearly every variable in the 2015 AHS PUF includes a corresponding variable representing an edit or imputation flag. In keeping with the convention used in prior AHS PUFs, the edit-imputation variable will be named the same as the variable to which it refers but will begin with the letter J. For instance, the edit-imputation variable for the PUF variable TOTROOMS is JTOTROOMS. The name of the corresponding J variable can be found in the AHS Interactive Codebook in the Edit Flag Variable field. The Codebook also includes an Imputation Strategy field (to be populated at a later time).

The second resource AHS PUF users can use to further investigate edits and imputations is the forthcoming document “Imputations and Allocations: 2015 and Later.” This document provides the details concerning the procedures used to impute missing values.

Topcoding and Other Disclosure Avoidance Techniques

To ensure respondent confidentiality, the Census Bureau requires that disclosure avoidance procedures be applied to certain PUF variables. Some variables are not included in the PUF because of disclosure avoidance. The section on Geographic Indicators mentions that geographic indicators in the PUF are limited due to disclosure avoidance.

A handful of other disclosure avoidance procedures are applied to 2015 and later PUF variables. These procedures include topcoding, noise, perturbation, and collapsing. For instance, the PUF variable LOTSIZE represents collapsing the answers to the lot size questions into various lot size groups.

Initially, all mortgage and financial characteristics in the 2015 and 2017 PUF were withheld due to the potential risk to disclosure. In April 2019, a subset of these withheld variables were released after further analysis, and in some instances, noise injection.

Mortgage and financial categorical variables released in April 2019: HOWBUY, DWNPAYPCT, FIRSHOME, LEADINSP, MORTLINE, LOANTYPE, MORTNEW, REFI, INTPMT, TAXPMT, and MISCPMT.

Mortgage and financial numeric variables released in April 2019 following noise injection: PMTAMT, PROTAXAMT, INSURAMT, INTRATE, MARKETVAL, HOAAMT, TOTBALAMT, LOTAMT, MORTPURCH, MAINTAMT. Note the TOTHCAMT and MORTAMT are calculated using the noise-injected values of their components.

Mortgage and financial variables withheld from the PUF: MORTSTAT, MORTTYPE, MORTDOC, LENMOD, REFILWINT, YEARBUY, MORTYEAR, UNPBALAMT, PMTFREQ, FXDPM, REFILWPAY, PRICE, MORTTERM, PRIPMT, MINPM, REFILWPER, DWNPAYSRC, MORTARM, INSPMT, INTPM, REFIINCPER, FORSALE, BALLOONAMT, PMIPMT, RATEPM, REFICSH, LOTVAL, MORTADDTN, PMIAMT, ADJPM, REFIEXTLN, MORTSUB, OTHAMT, OTRPM, REFIOTH, MORTSRC, PTCHYR, REFICSHAMT, HELOCBAL, MORTGOV, PTCHAM, REFILWINT, HELOCAD, HELOCL

Noise Injection Process and Impact

Noise factors were created using random, independent pulls from a Laplace distribution with a mean (location) of 1 and a scale value of $k*(1/\sqrt{N})$. N being the number of responses in geographic area and k being a proportional multiplicative factor determined in conjunction with the Census Disclosure Review Board. Original values are then multiplied by these noise factors to create the noisy data. Noise factors are independent across variables and years.

Noise is applied at the variable level in a single year for all observations. Thus, the mean noise factor for a variable in a given year is 1. This limits the impact of noise injection at the aggregate level. However, the more a variable is sliced, the higher the risk that the mean noise factor will deviate from 1. See the impact of this below.

Original values of 0 were not impacted by the multiplicative noise injection. Thus, the more 0's a noise-inject variable has, the more mean noise will deviate from 1. We recommend calculating means using non-zero values. For example, the mean of INSURAMT for those that paid at least \$1 in insurance.

How will noise impact my cross-sectional analysis?

For the National Sample excluding the Top 15 metro areas, the impact of noise injection in cross-sectional analysis is minimal. Cross-sectional analysis using noise-injected data from the AHS metro area should not be used for analysis smaller than mean by three-groups. If your analysis requires this level of granularity, use of the internal use file will be required.

Are certain variables impacted more by noise injection?

LOTAMT and MORTPURCH are impacted by noise-injection more than other noise-injected variables. LOTAMT is uncommon and are more at-risk for disclosure, thus, more noise had to be injected. MORTPURCH is impacted by editing. There are many cases of MORTPURCH grouped around and at 100 and noise-injection would often increase these values above 100. Noise-injected values above 100 were edited and rounded down to 100. Thus, mean noise is no longer 1 for MORTPURCH.

For the National Sample excluding the Top 15 metro areas, the mean absolute noise factor is around 7% for LOTAMT and 2% for MORTPURCH. For the metro areas c, the mean absolute noise factor is around 34% for LOTAMT and 6% for MORTPURCH.

When using metro area data, we recommend not using LOTAMT in analysis that requires filtering LOTAMT by more than one group. For other uses, we recommend binning LOTAMT into large bins to absorb the impact of noise injection.

When using metro area data, we recommend binning noise-injected MORTPURCH or to truncate/bin the upper end of the data to absorb the impact of post-noise editing.

How will noise impact my longitudinal analysis?

For the National Sample that is not part of the Top 15 Metro areas, the impact of noise injection for longitudinal analysis will be minimal. Changes in noise-injected variables across years 1% or less will be hidden at this aggregate level. For example, a 1% increase in MARKETVAL from 2015 to 2017 will not be seen in the noise-injected data.

For the Top 15 Metro areas, which are interviewed every survey cycle, the impact of noise injection for longitudinal analysis for all noise-injected variables excluding LOTAMT and MORTPURCH is as follows:

- True fluctuations in means of non-0 values across years less than 2-3% will be hidden in the noise-injected data (e.g. the change in the mean of MARKETVAL from 2015 to 2017).
- True fluctuations in means of non-0 values by group across years less than 3% will be hidden in the noise-injected data (e.g. change in the mean of MARKETVAL by FIRSHOME from 2015 to 2017).
- True fluctuations in means of non-0 values by two groups across years less than 4% will be hidden in the noise-injected data (e.g. change in the mean of MARKETVAL by BEDROOMS and FIRSHOME from 2015 to 2017).

Users should note a few pieces of information regarding topcoding, noise, collapsing, and other disclosure avoidance techniques in PUF.

- If disclosure avoidance techniques have been applied to a variable, it will be denoted in the AHS Interactive Codebook in the field Disclosure.
- For some variables in the PUF, the original variable is retained in the IUF. For instance, the PUFs include a variable called LOTSIZE that presents lot size in categories. The IUF includes LOTSIZE_IUF, which is the raw lot size value, as reported by the respondent.

- For additional information about disclosure avoidance methods, users should consult the forthcoming document “Topcoding and Confidentiality Measures: 2015 and beyond.”

Value Labels Package

The PUF includes a significant number of categorical variables. For AHS users of numerous PUF categorical variables, altering a categorical variable’s values (for example, 1, 2, 3) such that they reflect descriptive names (for example, mobile homes, single-family detached, single-family attached) can be a labor-intensive task.

In keeping with prior-year practice, HUD and the Census Bureau created an [AHS Value Labels Package](#). 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.