



AMERICAN HOUSING
SURVEY

Using the Internal Use File (IUF)

LAST UPDATED: OCTOBER 2022

U.S. DEPARTMENT OF HOUSING AND URBAN DEVELOPMENT
U.S. CENSUS BUREAU
U.S. DEPARTMENT OF COMMERCE





Contents

1. Overview	1
2. Why Use the Internal Use Files?	1
3. Internal Use Files Content and Structure	2
4. Internal Use File Geography Variables	5
5. Internal Use File Latitude/Longitude Coordinates	6
6. Internal Use File Race Variables	6
6.1. Race and Nativity	6
7. Internal Use File Recoded Variables	6
8. Sampling Variables – National Longitudinal Sample Files	7
8.1. Key Sampling Variables	7
9. Gaining Access to the IUF	8
10. Contact Information	8
Appendix A.	9

List of Exhibits

Exhibit 2.1. Differences between the PUF and IUF	1
Exhibit 4.1. Key Geography Variables on the IUF <i>MASTER</i> File.....	5
Exhibit 6.1. Coding of RACE1-RACE5 on the IUF	6
Exhibit 7.1. Variables Based on Especially Complex Calculation	6



1. Overview

The purpose of this help guide is to provide information about the American Housing Survey (AHS) Internal Use Files (IUFs) which contain full information collected from AHS respondents. The IUFs are available only to U.S. Census Bureau (Census Bureau) staff or authorized researchers who have obtained Special Sworn Status.

2. Why Use the Internal Use Files?

Each survey year, HUD and the Census Bureau release public use file (PUF) microdata to the general public. The AHS PUFs includes *most* of the data collected on the AHS. However, to maintain the privacy of the AHS respondents, many AHS variables are not released on the PUFs or are altered in some fashion (i.e., recoded, topcoded, or bottomcoded). This is especially true for AHS variables that reflect political or economic geography (i.e., census tract, county), or represent concepts that are inherently spatial.

Some AHS users may have data needs that cannot be fully supported by the PUF. To accommodate those users, HUD and the Census Bureau have created AHS IUF for each survey year. The AHS IUF contains geographic identifiers (e.g., census tract) that do not appear on the PUF and also includes data that has not been topcoded or aggregated for confidentiality reasons. In addition, the IUFs contain variables that can be used to calculate the correct standard errors for a complex random sample. IUF records, however, have been edited and imputed as needed. Exhibit 2.1 shows the differences between the PUF and the IUF.

Exhibit 2.1. Differences between the PUF and IUF

PUF	IUF
Topcoded/Bottomcoded and perturbed (Age) data to protect the privacy and confidentiality of respondents	Non-topcoded/bottomcoded data and age is not perturbed
Some variables, such as amenities available in the community, may be recoded to a single Yes/No variable	All individual variables collected in the survey are available
Other variables, such as some variables tracking details on why an unit was not interviewed, are not available	All variables in the interview are available
Merged/Masked geography	Detailed geography (including 1980 census tract)
Recoded/Collapsed information on home improvement	Detailed information on home improvement at the job-level
No information on sampling frames	Detailed information on sampling frames
In some metropolitan areas some cases may be dropped for disclosure purposes, as the population count in certain areas may no longer meet the threshold for disclosure by the Census Bureau (post-Hurricane Katrina New Orleans, for example)	All cases are available
No recoded/computed variables that are used to create the publication tables	A file including the recoded/computed variables used to create tables are included
No information from <i>MASTER</i> file	<i>MASTER</i> file which includes detailed information on sampling frame, geography, etc. is included



3. Internal Use Files Content and Structure

The main documentation for the AHS data is the PUF codebook, available on the Census Bureau website (<https://www.census.gov/programs-surveys/ahs.html>). The dataset consolidation applied to the 1997-2013 PUFs in the May 18th, 2021 revision and the 1973-1996 PUFs in the October 19th, 2022 revision has not yet been applied to the IUFs, so the IUFs are still set up using the file structure detailed below. In general, the AHS IUF internal variable names and file structure corresponds to that of the PUFs prior to the May 18th, 2021 and October 19th, 2022 revisions. There are some exceptions, some of which are described in this document.

The IUF consists of ALL variables on the PUF, including computed variables that are released on the PUF (i.e., variables that are created by combining or recoding other variables). The only difference between variables on the PUF and IUF are that all IUF variables are the actual data, whereas some variables on the PUF have been topcoded or bottomcoded and have been otherwise adjusted to protect the privacy and maintain the confidentiality of the respondent.

All data files described below are in SAS format. For each survey year, the IUF dataset consists of multiple SAS files.:

HOUSHLD/NEWHOUSE

This file used to be called *HOUSHLD* (until 2005) and is now called *NEWHOUSE*. The file has the same name in the PUF and IUF and contains most of the survey data. It has one record per household and the IUF includes more variables than the PUF.

PERSON

This file includes a single record for each member of the household. This means there can be multiple records for each occupied household and there are no records for vacant, usual-residence-elsewhere and noninterview housing units. This file includes all person-level data including race/ethnicity, income, age, disability status, etc. of each person in the household. Like on the PUF, the variable *MVG* on the person file can be used with *MOVGRP* (mover group), which is available in the *RMOV* (recent mover) data file to determine to which mover group each person belongs. Beginning in 2009, the *MOVGRP* variable on the *RMOV* file was renamed to *MVG* to make the merge easier.

HOMIMP

This file includes information on the home improvement projects conducted on each housing unit. In the PUF, the data are one record for each home improvement job reported, whereas on the IUF the data are one record per household with each job shown in their own variables.

RMOV

This file includes information on Recent Movers and is at the mover group level and is the same for both the IUF and PUF.

JTW



This file has information on an individual's journey to work and is at the person level. It is the same for both the IUF and PUF. This file no longer exists from 2011 onwards, as the *JTW* questions were removed.



MORTG

The information on mortgages is available on this file. It is at the household level and includes information on all owner-occupied homes with a mortgage. The only major differences between the PUF and the IUF for this file are topcodes.

RATIOV

This is a household level file that includes the ratio verification variables. There is no difference between the IUF and the PUF.

OWNER

This is a household level file with information on the on-site owner or manager and there is no difference between the PUF and the IUF.

TBLRCD

Many recodes are computed to create the tables released by the Census Bureau on the AHS data. These recoded variables are available in this file. It is a household level file. There is no PUF equivalent for it, as this file is only released with the IUF.

MASTER

This is a household level file, which is also unique to the IUF (no PUF equivalent). This file has sampling frame information on each housing unit and includes fields identifying detailed geography (down to the 1980 census tract level) for each unit.

NONTOPPUF

In years prior to 2007, this file included the recodes that are included on the PUF. Unlike on the PUF, these variables have not been topcoded and do not have values suppressed or masked. It should be noted that many of these variables or near equivalents appear on other IUF files. These variables have been incorporated in the *NEWHOUSE* file starting with the 2007 national longitudinal sample IUF.

TOPICAL

Information from survey modules dealing with specific topics, for example, use of public transportation or household planning for disasters. These modules are not part of every AHS survey (in 2011 and 2013 surveys only).

OMOV

Information on persons who have left the unit since the last survey; this information is organized by groups of out-movers (in 2013 survey only).

REPWGT

Information needed to calculate variances. Contains multiple weights that can be used to generate alternative estimates of a variable for the purpose of calculating variances (in 2011 and 2013 surveys only).



TYPEC

Cases that were shown on previous year metropolitan area longitudinal oversample or national longitudinal sample public use files (PUFs) but do not appear on the current year PUF (in 2011 and 2013 surveys only).

4. Internal Use File Geography Variables

Exhibit 4.1. Key Geography Variables on the IUF MASTER File

Variable Name	Census Geography
CBNCOD90	1990 central city/balance code
CBUR80	1980 central city/balance/urban/rural code
CENSTATE	1960 Census state code
CMSA80	1980 consolidated MSA code
CMSA90	1990 consolidated MSA code
COOLDAY	Cooling degree days
COUNTY80	1980 FIPS county code
COUNTY90	1990 FIPS county code
FIPSTATE	FIPS state code
HEATDAY	Heating degree days
MDCOD80	1980 design MCD/CCD code
MDCOD90	1990 design MCD/CCD code
MSASTA80	MSA status – 80 definition
MSASTA90	MSA status – 90 definition
PLCODE80	1980 design Census place code
PLCODE90	1990 design Census place code
PMSA80	1980 design MSA/PMSA code
PMSA90	1990 design MSA/PMSA code
REGION	Census region
TRACT80	1980 tract code
TRCTSF80	1980 tract suffix
UACODE80	1980 design urbanized area code
UACODE90	1990 design urbanized area code
UASIZE90	1990 design urbanized area size
URBRUR80	1980 design urban/rural code
URBRUR90	1990 design urban/rural code
ZONE	Zone code (metropolitan only)

Federal Information Processing Standards (FIPS) codes can be found at:
<https://www.census.gov/library/reference/code-lists/ansi.html>

Metropolitan Statistical Area (MSA) codes can be found at: <https://www.census.gov/programs-surveys/metro-micro/geographies/geographic-reference-files.html>



5. Internal Use File Latitude/Longitude Coordinates

An internal use SAS dataset (*geoskinny.sas*) is available for use with the IUF. It is designed to match to the internal “*newhouse.sas*” file using the *CONTROL* variable. This file provides *WGS84* coordinates in addition to census 2000 vintage tract, block group, and block codes. For 2011 and 2013, it also includes 2010 vintage tract, block group, block, and PUMA codes, as well as the census 2000 vintage PUMA codes.

6. Internal Use File Race Variables

For the national longitudinal sample and metropolitan area longitudinal oversample surveys between 1997 and 2002, *RACE* is coded identically on both the IUF and the PUF. Beginning in 2003, respondents were allowed to choose “one or more” races. The IUF stores these answers in *RACE1* (first race mentioned) through *RACE5* (last race mentioned). The PUF includes only the variable *RACE*, which recodes these answers into 21 categories such as “White only.”

Exhibit 6.1. Coding of RACE1-RACE5 on the IUF

Code	Race
1	White
2	Black or African American
3	American Indian or Alaska Native
4	Asian
5	Native Hawaiian or Other Pacific Islander
6	Other - DO NOT READ

6.1. Race and Nativity

Nativity (country of birth) information has been collected since the 2001 national longitudinal sample file, but appears on the PUF in aggregated form (e.g., Portugal and the Azores are grouped into a single category). Nativity codes for the IUF are contained in Appendix 1.

7. Internal Use File Recoded Variables

The *TBLRCD* file contains the recodes used internally to produce the AHS publication tables. Most of these variables are fairly simple recodes, but may be of use to users seeking to reproduce the AHS publication tables. These variables may allow users to save some effort and avoid “reinventing the wheel.” Five variables that are based on especially complex calculations, and hence probably of the most interest to users, are:

Exhibit 7.1. Variables Based on Especially Complex Calculation

Variable Name	Description
OTPINR	Outstanding principal and interest (rnd)
POORR	Household income as % of poverty level (rnd)
POVLVL	Poverty value that corr. to lookup table
ZSMHCM	Monthly housing costs w/ maintenance
ZSMHCN	Monthly housing costs w/o maintenance



8. Sampling Variables – National Longitudinal Sample Files

The AHS is a complex random sample. Standard errors calculated from the data will be larger than those calculated using formulas that assume a simple random sample. This section discusses the key features of the AHS sample design, and discusses the IUF variables that correspond to those features.

The AHS sample is stratified in two ways: by Primary Sampling Unit (PSU) and by sample frame. A PSU is a county or a group of counties. The AHS sample was drawn by dividing the country into PSUs and then randomly sampling PSUs.

More specifically, large PSUs (called “self-representing” PSUs) were drawn with certainty. Smaller PSUs (“non-self-representing”) were divided into strata based on geography and characteristics from the 1980 census.¹ Finally, one PSU was randomly chosen from each strata.

A difficulty with this scheme is that there is only one PSU per stratum, while at least two PSUs per stratum are required to calculate the standard errors. Hence, the Census Bureau has combined pairs (sometimes triplets) of PSUs into “pseudo-strata” of similar PSUs for the purpose of calculating SEs.

Within each PSU, housing units were randomly chosen from four sample frames (lists of housing units). The 1980 census frame contains housing units constructed before 1980. The permit frame contains housing units built since 1980 in areas where a building permit is required to authorize construction. The special areas frame contains housing units built since 1980 outside of permit-issuing areas, a small number of rural areas. Finally, the group quarters frame contains units in group quarters. These group quarters units are not considered housing units and are not in the interviewed sample. A sample is drawn from these units nonetheless, because some of these units may later be converted to housing units.

8.1. Key Sampling Variables

PSU80

1980 design stratification PSU

SEGMTYPE

Can be used to identify which sample frame unit comes from

Frames defined using SEGMTYPE

Permit Frame—SEGMTYPE = 4,

13 GQ Frame—SEGMTYPE = 12

Special Area Frame—SEGMTYPE = 6, 10

Unit Frame—SEGMTYPE =1, 2, 3, 7, 8, 9, 11

SEGMNT80

The last three digits can also be used to identify which sample frame unit comes from Frames defined using SEGMNT80 are:

Permit Frame—001-249

GQ Frame—250-274



Special Area Frame—275-299

Unit Frame—300-999

STRPSU80

Pseudo-strata

PSUTYP80

1980 design PSU type (self-representing/non-self-representing). Values of 1, 4, or 5 mean self-representing, 2 or 3 mean non-self-representing.

In a statistical procedure such as SAS proc means, a user would specify strata as self-representing*frame*pseudo-strata, and specify cluster (within strata) as PSU80.

Note that the AHS national longitudinal sample files from 1985-2013 are based on a 1980 sample design. 1990 and 2000 sample design.¹

9. Gaining Access to the IUF

Access to the IUF is available through the secure Research Data Center (RDC) network. Approved researchers must obtain a Special Sworn Status (SSS) from the Census Bureau. SSS researchers are sworn for life to protect the confidentiality of the data they access. There are a number of partnering universities and research institutions throughout the United States staffed with a Census Bureau employee. For information on how to apply and the types of proposals that are acceptable see <https://www.census.gov/about/adm/ced/apply-for-access.html>.

10. Contact Information

AHS Information Owner: Tamara Cole, Survey Director, Housing Surveys 1-301-763-4665, tamara.a.cole@census.gov

SEHSD AHS staff: 1-301-763-3235 or 1-888-518-7365 (toll free) or by email at ahsn@census.gov.

¹ For a more detailed description of the AHS sample design and formation of the strata, see “American Housing Survey, A Quality Profile,” Current Housing Reports H121/95-1, available at <http://www.census.gov/programs-surveys/ahs/research/publications/h12195-1.html>. Variables are also included on the file, for internal use (i.e., for field staff), but played no role in drawing the sample.



Appendix A.

Exhibit A.1. Coding of NATVTY (Country of Birth) on the IUF

Code	Country	Code	Country	Code	Country
57	United States	312	El Salvador	184	Lithuania
72	Puerto Rico	555	Elsewhere	224	Malaysia
96	Outlying Area of the U.S. (American Samoa, Guam, U.S. Virgin Islands, Northern Marianas, Other U.S. Territory)	139	England	315	Mexico
200	Afghanistan	417	Ethiopia	252	Middle East
375	Argentina	148	Europe	436	Morocco
185	Armenia	507	Fiji	126	Netherlands
245	Asia	108	Finland	514	New Zealand
501	Australia	109	France	316	Nicaragua
102	Austria	110	Germany	440	Nigeria
130	Azores	421	Ghana	468	North Africa
333	Bahamas	138	Great Britain	304	North America
202	Bangladesh	116	Greece	142	Northern Ireland
334	Barbados	340	Grenada	127	Norway
103	Belgium	313	Guatemala	462	Other Africa
310	Belize	383	Guyana	527	Pacific Islands
300	Bermuda	342	Haiti	229	Pakistan
376	Bolivia	126	Holland	253	Palestine
377	Brazil	314	Honduras	317	Panama
205	Burma	209	Hong Kong	385	Peru
206	Cambodia	117	Hungary	231	Philippines
301	Canada	210	India	128	Poland
353	Caribbean	211	Indonesia	129	Portugal
318	Central America	212	Iran	132	Romania
378	Chile	213	Iraq	192	Russia
207	China	119	Ireland/Eire	233	Saudi Arabia
379	Colombia	214	Israel	140	Scotland
311	Costa Rica	120	Italy	234	Singapore
337	Cuba	343	Jamaica	156	Slovakia/Slovak Rep
155	Czech Republic	215	Japan	449	South Africa
105	Czechoslovakia	216	Jordan	389	South America
106	Denmark	427	Kenya	134	Spain
338	Dominica	218	Korea/S. Korea	136	Sweden
339	Dominican Republic	221	Laos	137	Switzerland
380	Ecuador	183	Latvia	237	Syria
415	Egypt	222	Lebanon		

U.S. Department of Housing and Urban Development
Office of Policy Development and Research
Washington, DC 20410-6000



October 2022