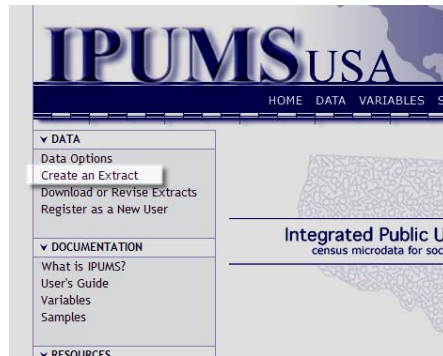


Instructions for Using the IPUMS Data Extraction System

The IPUMS-USA data extraction system allows researchers to make extracts of the data oriented to their specific research needs and available computing resources. In practice, researchers never require all variables and all cases from a census or survey year. Researchers can design subsamples incorporating the subset of variables and years pertaining to the specific population(s) of interest to them.

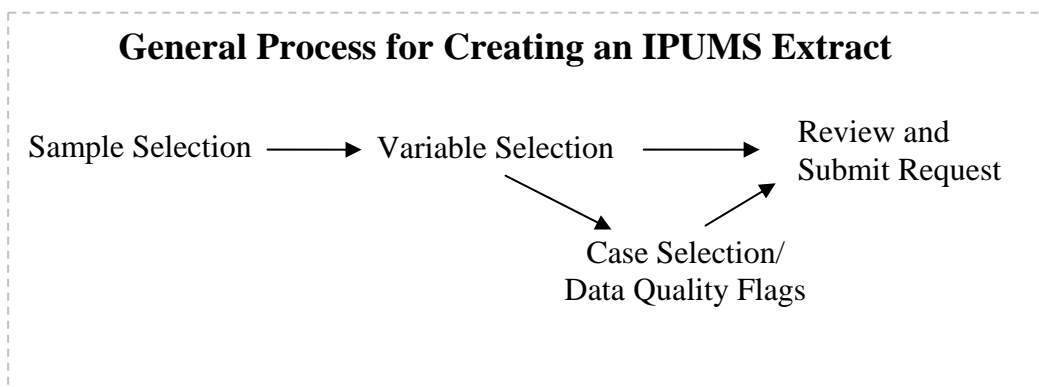
From the home page, click on 'Create an Extract'.



Login- On the login page, you are prompted for your e-mail address. The e-mail address acts as your password and provides us with a means of contacting you and constructing a unique file name for your data extract. (To use the extract system, you must first [register](#).)

The Extract Interface

The extract procedure involves a series of web pages, with the contents of each depending on selections made on the previous page. On each page, you will make the desired (and required) selections and click the 'Continue' button to advance to the next screen.



Step 1 -- Sample Selection

The IPUMS extract system makes available many samples. On the first page, you must choose the density of the sample(s) you would like to extract.

Sample Density - The "small" and "tiny" densities are intended for testing and quick turnaround; they include only a fraction of the cases available from any given sample. The "regular" samples are typically 1 percent samples of each year. The "large" samples are 5 percent samples available only for recent years.

Create a New Extract

Sample Density [help](#)

Tiny Tiny samples contain approximately 5,000 adults and their households.

Small Small samples contain approximately 50,000 adults and their households.

Regular Most of the regular samples are 1% but the density of these samples varies. Refer to the [online documentation](#) for more information.

Large The large samples are 5% samples and are available for 1980-2000.

Regular & Large You can mix regular and large samples in the same extract

Sample Selection List [help](#)

Only the most commonly requested samples

All of the available samples

[Continue to Sample Selection](#)

Sample Selection List - The final selection on the first page affects what choices you will have on the next screen. If you would like to be allowed to choose from among all available samples (in years that have more than one), you must choose "All of the available samples." Otherwise, the system defaults to presenting one sample per year, the one determined to have the broadest application. This option only pertains to the "regular" sample densities.

On the second page you will define the general characteristics of your desired extract.

Extract File Type - Choose the preferred file structure for your extract: hierarchical (household record followed by person records), rectangular ("flat", with all household information attached to respective household members), or household records only. The system defaults to rectangular format, which is the overwhelming choice of researchers.

Variable Listing - The remaining selections affect what you see on the next screen. You can choose to be presented with only the most commonly-requested variables or with all the variables available in any of the samples selected. This listing of variables can be presented in alphabetic order or grouped thematically.

Sample Selection

[Continue to Variable Selection](#)

Extract File Type [help](#)

Rectangular

Hierarchical

Household records only

Variable Selection List [help](#)

All of the available variables

Only the most commonly requested variables

Samples [help](#)

[1980 5% sample](#)

[1990 5% sample](#)

[2000 5% sample](#)

[Continue to Variable Selection](#)

Samples - Finally, you select the particular census sample or combination of samples you want. If on the previous page, you selected "regular" density and "all available samples," you will have more than one sample to choose from in some years.

Step 2 -- Variable Selection

On the variable selection page, you will be able to select variables and request case selections and data quality flags for those variables.

Requesting Data Quality Flags

If you would like to include data quality flags in your extract, check the box near the top of the page (as shown below). All flags corresponding to selected variables will be presented on the next screen.

Jump to variable groups

Household Record Technical Variables Geographic Variables Group Quarters Variables Economic Characteristic Variables Dwelling Characteristic Variables Appliances, Mechanical, Other Variables Constructed Household Variables	Person Record Technical Variables Family Interrelationship Variables Demographic Variables Race, Ethnicity, and Nativity Variables Education Variables Work Variables Income Variables Migration Variables Disability Variables Veteran Status Variables Place of Work and Travel Time Variables Other Variables
---	--

Variable Availability Key

	All Years	X	Available in all samples in this year
		.	Not available in this year
	1990	S	State sample (5% state), 1% Flat sample
		M	Metro sample (1% metro)
		E	Elderly sample
		L	Labor Market Areas
	2000	1	1% census sample
		5	5% census sample, 1% flat sample
		S	2000 American Community Survey

Data Quality Flags [help](#)

Select data quality flags

Technical Variables (Household) [top](#)

Detailed Version	General Version	Variable	Label	Case Selection	1990	2000
<input checked="" type="checkbox"/>		YEAR	Census year		x	x
<input type="checkbox"/>		DATANUM	Data set number		x	x
<input type="checkbox"/>		SPH1	Household serial number		v	v

Selecting Variables

Only those variables available for the particular samples selected in Step 1 are displayed as options. If you have selected multiple samples, all variables occurring in any of the specified samples are displayed.

On the right side of the variable selection page, there is a column for each year selected on the previous (sample selection) page. Only columns for selected years are displayed, with the symbols showing the availability of each variable across years. An "X" indicates that the variable is available for all individuals or households in a given year, an "SL" means that the variable is only available for sample-line individuals in 1940 and 1950, and so on. All symbols are explained in the 'Variable Availability Key' near the top of the page. The variable availability key is generated specifically for the samples you selected in Step 1.

To select a variable to be included in your extract, click on the check box to the left of the variable. Descriptions of the variables are available by clicking on the variable name.

Demographic Variables (Person)		
Detailed Version	General Version	Variable
<input checked="" type="checkbox"/>	<input type="checkbox"/>	RELATE
<input type="checkbox"/>		AGE
<input type="checkbox"/>		SEX
<input type="checkbox"/>		MARST
<input type="checkbox"/>		CHBORN
<input type="checkbox"/>		All Demographic Variables

Demographic Variables (Person)		
Detailed Version	General Version	Variable
<input type="checkbox"/>	<input type="checkbox"/>	RELATE
<input checked="" type="checkbox"/>		AGE
<input type="checkbox"/>		SEX
<input type="checkbox"/>		MARST
<input type="checkbox"/>		CHBORN
<input type="checkbox"/>		All Demographic Variables

Selecting General Versions

Many of the variables are available in both a general and a detailed version. The difference between the two versions is outlined on the codes page which can be linked to from the variable description. For variables that also include a general version, you may select the general version by clicking on the box in the general version column.

Demographic Variables (Person)			
Detailed Version	General Version	Variable	Label
<input checked="" type="checkbox"/>	<input type="checkbox"/>	RELATE	Relationship to household head
<input type="checkbox"/>		AGE	Age
<input checked="" type="checkbox"/>		SEX	Sex
<input type="checkbox"/>		MARST	Marital status
<input type="checkbox"/>		CHBORN	Children ever born
<input type="checkbox"/>		All Demographic Variables	

Requesting Case Selection

Some variables are categorical and users are able to select persons or households who fit certain criteria based on values for that variable. For instance, users are able to extract data only for persons over the age of 15 by using the case selection feature in conjunction with the AGE variable. To do so, check the box in the case selection column.

Detailed Version	General Version	Variable	Label	Case Selection	199
<input checked="" type="checkbox"/>	<input type="checkbox"/>	RELATE	Relationship to household head	<input checked="" type="checkbox"/>	x
<input type="checkbox"/>		AGE	Age	<input type="checkbox"/>	x
<input checked="" type="checkbox"/>		SEX	Sex	<input type="checkbox"/>	x
<input type="checkbox"/>		MARST	Marital status	<input type="checkbox"/>	x
<input type="checkbox"/>		CHBORN	Children ever born	<input type="checkbox"/>	x
<input type="checkbox"/>		All Demographic Variables			

Step 3 -- Data Quality Flag and Case Selection

Case Selection

Step 3 provides for case selection. Only those variables chosen for case selection in Step 2 will appear on this page. Case selection limits the extract to include only cases that contain the values you select for the listed variables.

You have the option of selecting only those individuals with the selected characteristics or entire households containing any individual with the selected characteristics.

Case Selection

Individual/Household Case Selection

Include only those persons meeting case selection criteria
 Include all persons in the household of person meeting case selection criteria

Case Selection Across Variables

"AND" across variables
 "OR" across variables

Age (Person)

000 Less than 1 year old

001 1

002 2

003 3

004 4

005 5

006 6

007 7

008 8

009 9

Race (Person)

1 White
 2 Black/Negro
 3 American Indian or Alaska Native
 4 Chinese
 5 Japanese
 6 Other Asian or Pacific Islander
 7 Other race, nec
 8 Two major races
 9 Three or more major races

There is also a choice of a logical "or" or "and" in applying the case selection criteria. The default choice is "and."

Finally, for the variables for which you have requested case selection, you will select specific values.

Age (Person)

009 9

010 10

011 11

012 12

013 13

014 14

015 15

016 16

017 17

018 18

Race (Person)

1 White
 2 Black/Negro
 3 American Indian or Alaska Native
 4 Chinese
 5 Japanese
 6 Other Asian or Pacific Islander
 7 Other race, nec
 8 Two major races
 9 Three or more major races

Data Quality Flag

If, in Step 2, you chose to include data quality flags, the flags corresponding to variables selected in Step 2 will appear on this page. Select the ones you would like to include in your extract. For more information on data quality flags, read our [“Data Quality Flags Introduction.”](#)

Data Quality Flag Selection

Technical Flags (Household)

Flag Name	Flag Description	1980	1990	2000
<input type="checkbox"/> QAUGMENH	For whole records	.	.	5, 1
<input type="checkbox"/> QSUBSTIH	For whole records	.	.	5, 1
<input type="checkbox"/> All Technical Flags				

Group Quarters Flags (Household)

Flag Name	Flag Description	1980	1990	2000
<input checked="" type="checkbox"/> QGQ	For GQ	X	.	.
<input type="checkbox"/> All Group Quarters Flags				

Technical Flags (Person)

Flag Name	Flag Description	1980	1990	2000
<input type="checkbox"/> QAUGMENT	For whole records	.	X	5, 1
<input type="checkbox"/> All Technical Flags				

Step 4 -- Extract Summary

In the final step, you can review your selections on a summary page. A series of gray bars near the bottom of the screen allow you to jump to specific pages in the extract process to edit selections on those screens.

When you are satisfied with your extract request, submit it for processing. The system will inform you via e-mail when the extract is ready and will provide instructions for downloading the files. At any time, you can access your completed data files by clicking on the “Download or Revise Extracts” link (under “Data” on the left side) from most IPUMS web pages.

Command File Format - All data extracts are compressed automatically. The system produces ASCII column-format data, but it will generate SAS, SPSS, or Stata command files to facilitate reading the data into any of these statistical packages. The command files contain the column locations of variables, variable labels, and value labels for categorical variables.

For specific instructions on how to open and use an extract in PC-based statistical software, please see our [Instructions for Opening an IPUMS Extract](#).

Modifying Previous Extracts

The IPUMS system allows you to modify previous extracts and resubmit them. Click “Download or Revise Extracts” from most IPUMS web pages. You can load an old extract request file with all of your previous selections already made and edit only those parts that you wish to alter. You can then submit these revisions and a new extract will be requested, leaving your original extract intact and available for download or for other revisions.

