# Instructions for Using the IPUMS Data Extraction System

The IPUM-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 populations of interested to them.

From the home page, click on **'Select Data'**.
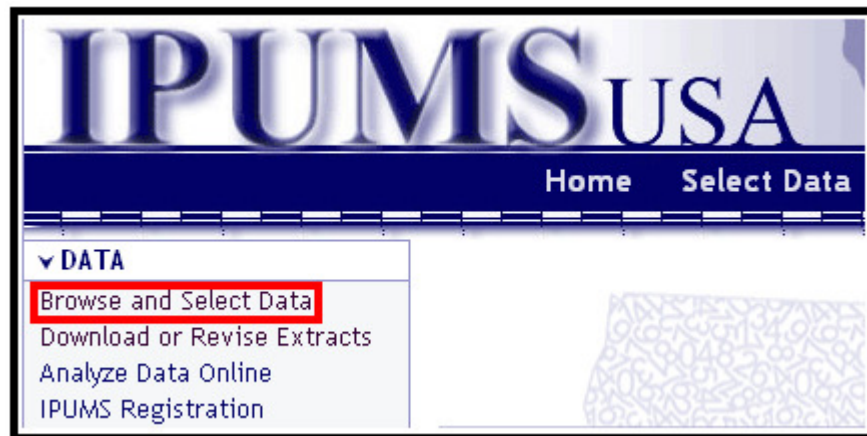
The 'Browse and Select Variables' screen that follows allows you to browse and select the variables that you are interested in, select relevant samples, and set extract options.

To view or select household-level variables, hover over **'Household'** and then click on the desired variable category. To view or select person-level variables, hover over **'Person'** and click the category of your choice. Alternatively, you can view and select variables alphabetically by hovering over **'A-Z'** and selecting the relevant letter. You can also use the **search functionality** to locate the variable that you're looking for.
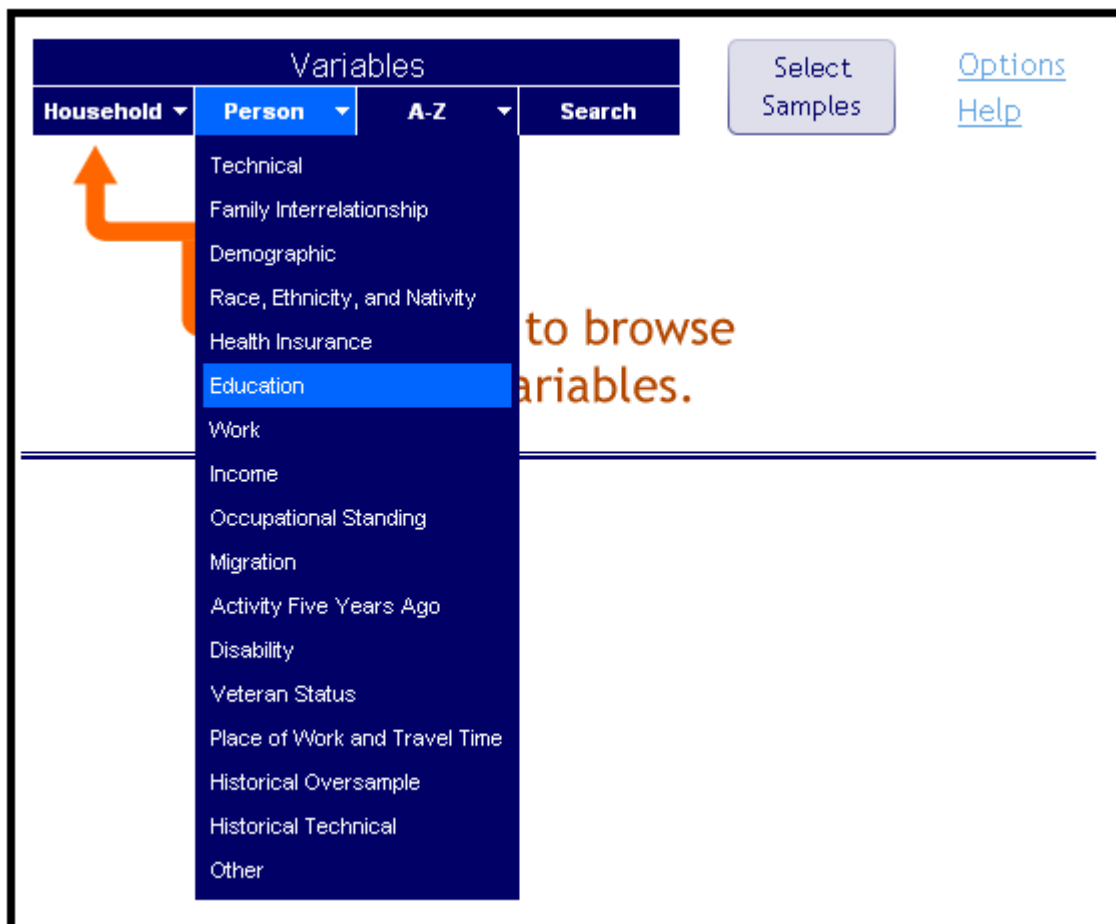
The screen that follows contains all **variables matching the criteria you selected**, and indicates the **samples for which these variables are available**.

In the example above, the 'X' for 'SCHOOL' under '2009 acs' indicates that the SCHOOL variable is available in the 2009 ACS sample. The '.' For 'HIGRADE' under '2009 acs' indicates that the HIGRADE variable is not available in the 2009 ACS sample.

To view the variable description for a particular variable, simply click on the **variable name**. You can also click on the **"Codes"** link to view the possible codes and labels for the variables, along with the years in which each code was available as a response.

Once you've clicked through to the variable description, you can view the availability, universe, codes, or enumeration text by simply clicking the relevant link at the top of the variable description page. If you're satisfied that the variable suits your needs, click 'Add to Cart' in the upper-right-hand corner. You can click "Back" in your browser window to return to the variable browse and select screen.

If you know which variables you want to use without needing to read the description, you can always click the "+" button next to the desired variable. Once a variable is in your data cart, this button will instead display as a checked box. Clicking the checked box will remove the variable from your cart.

Clicking the "Select Samples" button will allow you to pick the samples for which you'd like to obtain data.

After checking the boxes next to the samples you desire, click 'Submit sample selections' to save your choices.
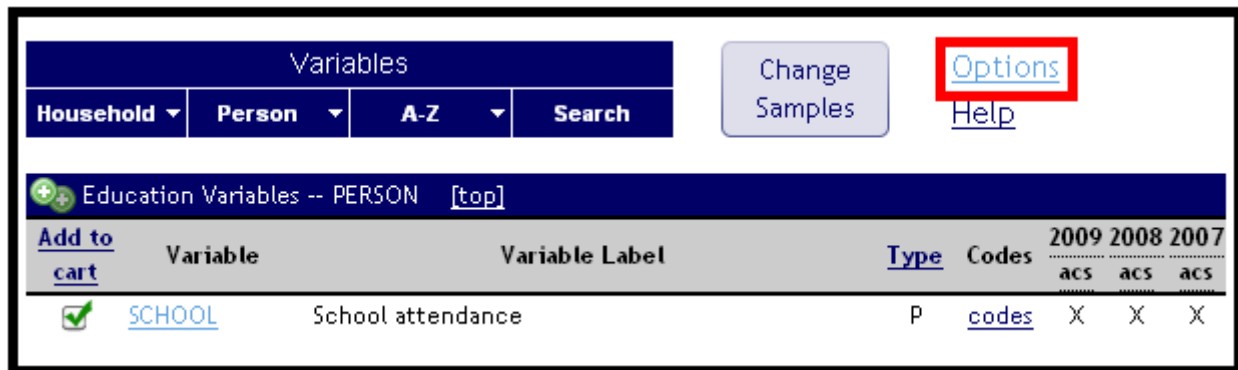
You can revise the selected samples at any time by clicking the 'Change Samples' button.

You can edit display options by clicking the "Options" link in the upper-right corner.

The View Options page allows you to tell the extract system how you want the variables to be displayed. You can opt to have them display one group at a time or all together. You can elect to be shown detailed or summary availability information. The system allows you to select whether you want to show variables that are not available for the selected samples or not. Finally, you can choose chronological or the reverse chronological ordering. Click 'Apply Selections' to save your choices.

Once you've selected the samples and variables of your choice, click **'View Cart'**.

The page that follows lists the **variables pre-selected by the extract system** as well as any **variables you selected**. You can remove variables from your extract in this step by clicking the **relevant checkbox in the 'in cart' column**. If the samples and variables are as you wish, click **'Check out: Create Data Extract'**.

At this point, you'll be asked to log in or register. Registered users can enter their log in details and click "Submit". If you're not yet registered, you'll be taken back to your data cart to continue the process once you've submitted the registration form.

The 'set extract options' screen that follows allows you to generate **either a rectangular or hierarchical file**. By default, IPUMS data are rectangularized – that is, the extract system puts the household information on the person records and drops the separate household record. If you'd prefer a hierarchical file you can choose that here. Then, click **'Continue to next step'**.

The 'set variable options' page lets you select cases, attach characteristics, and include flags.

You can include data quality flags for certain variables by checking the boxes next to the variable name. Checking the box at the top of the **'Include flags' column** will select all data quality flags available for the variables and samples you have chosen.

After you've checked the desired boxes, click **'Continue to next step'**.

The 'select cases' feature allows users to limit their dataset to contain only records with specific values for selected variables, such **as persons age 65 and older**. Multiple variables can be used in combination during case selection. Selections for multiple variables are additive, each being implicitly connected by a logical "AND" for processing purposes. When you've made your selections, click **'Continue to next step'**.

The data extract system can attach a characteristic of a person's mother, father, or spouse as a new variable on the person's record. It can also attach the characteristics of the household head. For example, using the variable 'Age', it can make a new variable for **'Age of Spouse'**. Again, click **'Continue to next step'**.

Finally, the 'Review extract request' screen summarizes your request…

Review extract request (help)

Extract File Type and Estimated Size

File Type: Rectangular
Estimated size: 387.4 MB

Samples

Sample	Density	Custom Density	Note
2007 ACS	1.0%	--	2007 ACS sample
2008 ACS	1.0%	--	2008 ACS sample
2009 ACS	1.0%	--	2009 ACS sample

Variables and Case Selection

Type	Variable	Label	Case Selection
H	YEAR	Census year	--
H	DATANUM	Data set number	--
H	SERIAL	Household serial number	--
H	HHWT	Household weight	--
H	GQ	Group quarters status	--
P	PERWT	Person weight	--
P	RELATE (general)	Relationship to household head [general version]	--
P	RELATED (detailed)	Relationship to household head [detailed version]	--
P	AGE	Age	--
P	SEX	Sex	--
P	MARST	Marital status	--
P	SCHOOL	School attendance	--
P	AGE_SP	Age [of spouse]	--

You can also make desired changes desired changes at this point using the appropriate buttons on the screen.

Use these buttons to make changes to your extract.

Samples and variables

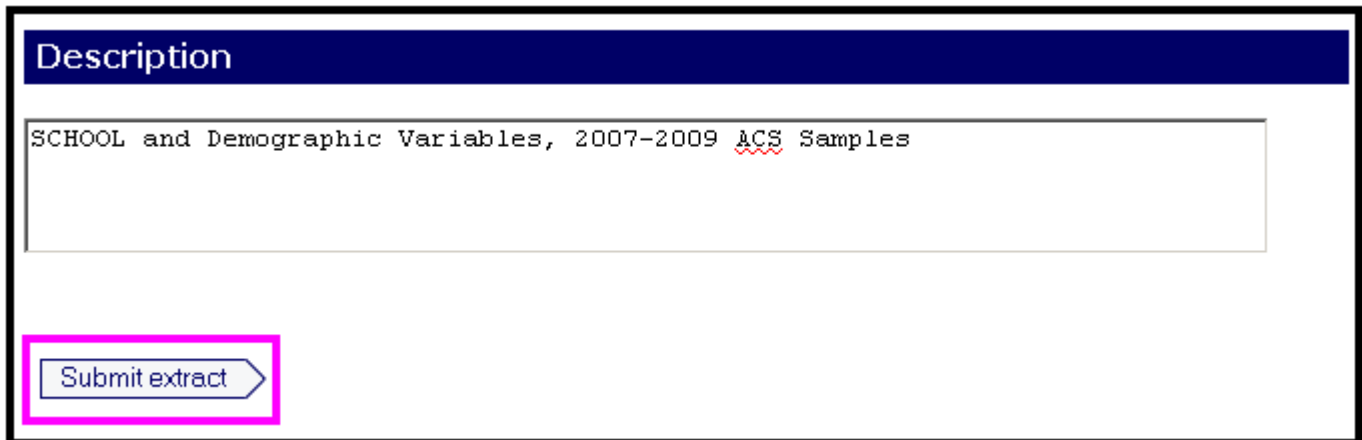Extract options

Variable options

Attached characteristics

Sample sizes

If you're satisfied with your request, enter a description of your extract for your own reference and click **'Submit extract'**.

Extracts typically take only a few minutes, but can take a few hours or even more than a day if they are very large. You'll receive an email once your extract is complete, at which time you'll be able to download the data using the link provided in the email or by clicking "Download Your Extract" in the upper-left corner.