User Guide

Select and Download Data in the Data Finder

The IHGIS Data Finder guides you through the process of creating a customized data extract consisting of selected data tables and accompanying metadata. You will browse available data tables by dataset, select the tables you wish to download, review and submit your extract request, and download your extract.

Apply filters

Click the green "DATASETS" button to choose one or more datasets to browse. This will open the datasets filter pane.

FILTER » OPTIONS » REVIEW

APPLY FILTERS ❓

DATASETS

No filters selected.
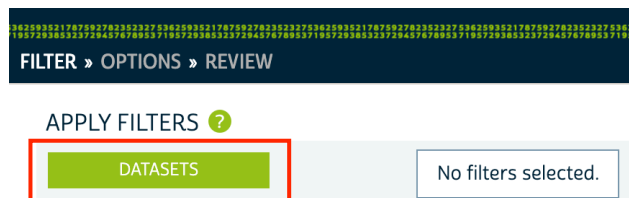
You can filter datasets by world Region or by Decade by using the tabs on the left side of the filter pane.

Click the green '+' next to each dataset you wish to browse, then click "SUBMIT"

Note: The dataset filters only affect which tables are shown in the data selection grid. They do not affect your data table selections or data cart. So you can filter on one dataset, select tables from that dataset, change the filter to a different dataset (removing the first dataset from the filters), and retain the tables you selected from the first dataset in your cart.

Future versions of IHGIS will include additional filters, such as country, year, topic, and/or type of dataset.

Select tables

After applying dataset filters, data tables available for those datasets will be listed in the "SELECT DATA" grid. Add data tables to your cart by clicking the green '+' on the left side.

You can change the number of tables listed in each page of the grid with the dropdown at the top and bottom center.

You can remove tables from your cart by clicking the green check mark.

Data cart

The tables you select will appear in your Data Cart. If you wish to review your selections, click "SHOW SELECTIONS," which will open a view similar to the Select Data grid, showing only the tables you have selected. You may remove tables from your cart by clicking the green checkmark in this view.

To move to the Options step, click "CONTINUE" in the Data Cart.

Data options

The Data Options step serves as a review of the tables you have selected. The number of selected tables and number of files requested is shown for each dataset. You will receive one file for each geographic level available for each table you have selected. For example, if a table is available at the Nation and District levels, you will receive two files, one with the national level data, and one with the district-level data. This structure facilitates joining to GIS files of the unit boundaries for mapping and spatial analysis. Future versions of IHGIS will enable you to select which geographic level(s) you receive for each table.

You can view a summary of your selected tables and their available geographic levels for each dataset by clicking the "*X tables*" link under Tables Selected.

DATA OPTIONS ❓

SOURCE TABLES

DATASET	SELECTIONS	
	TABLES SELECTED	FILES REQUESTED
Agricultural Census: AL2012ag	1 table	2
Population Census: DZ1998pop	1 table	2

Move to the final Review and Submit step by clicking "CONTINUE" in the Data Cart.

Review and submit

You may enter a description of your selections as a reminder of what is included in your extract. This description will appear on the downloads page.

FILTER » OPTIONS » REVIEW

REVIEW AND SUBMIT ❓

SOURCE TABLES
  2 source tables

DESCRIPTION ❓

Albania agricultural land and Algeria marital status

SUBMIT

Click "SUBMIT" to start processing your extract. You must be logged in to submit an extract. If you are not already logged in, you will be prompted to do so.

You will receive an email notification when your extract is ready to download. (Most IHGIS extracts should complete within a few minutes.)

Download

When your extract is complete, download your data by clicking the "tables" link under "DOWNLOAD TABLE DATA."

You may also revise or resubmit your extract using the respective links.

To access your Extracts History page from elsewhere in the system, click "MY DATA" from the top menu.

What's in an Extract

You will receive your extract as a .zip file, which you can open with your favorite unzipping utility. Inside the .zip file, you will find one or more data files and four metadata files.

data files

metadata files

The metadata file names begin with "ihgis" and your extract number.

Data file names consist of:
• Two-letter country code
• Year
• Dataset type code (e.g., "ag" for agricultural census, "pop" for population census)
• Three-character table code
• Geographic level code (g0 is national, g1 is the largest subnational units, etc.)

Codebook

The ihgisXXXX_codebook.txt file is a human-readable summary of the contents of your extract. It includes basic information about the datasets, tables, and variables, as well as the recommended citation for IHGIS.

The other metadata files are provided as comma-separated values files to facilitate importing them into statistical packages or other software tools.

Geographic level metadata

The ihgisXXXX_geog.csv file provides the name of each geographic level that is included in one or more tables in your extract.

	A	B	C
1	dataset	geog_level	geog_level_label
2	AL2012ag	g0	Nation
3	AL2012ag	g1	Prefectures
4	DZ1998pop	g0	Nation
5	DZ1998pop	g1	Wilayas

Table metadata

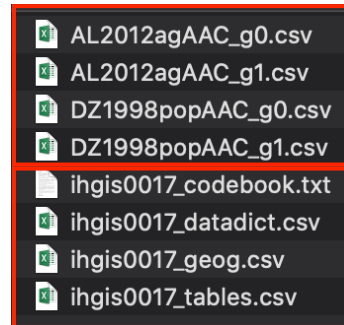The ihgisXXXX_tables.csv file provides detailed metadata for each table in your extract. *dataset* and *table* codes are provided separately as well as in a concatenated *dataset_table* field.

table_num is the designation of the table in the source document

source_pub_eng is the title of the document in which the table was originally published. It may be a translation into English of the original native-language title.

If any footnotes are present for the tables, they will be included in this file.

	A	B	C	D	E	F	G	H	I	J
1	dataset	table	dataset_table	title	table_num	table_universe	geog_levels	geog_level_labels	source_pub_eng	country
2	AL2012ag	AAC	AL2012ag.AAC	Agricultural holdings with	3	Agricultural holdings	g0; g1	Nation; Prefectures	Preliminary Results of Agricu	Albania
3	DZ1998pop	AAC	DZ1998pop.AAC	Population by Marital Sta	3	Resident Population of Or	g0; g1	Nation; Wilayas	Population and Housing Cens	Algeria

Data dictionary (variable metadata)

The ihgisXXXX_datadict.csv file provides detailed metadata for the variables (columns) in the tables in your extract. This information is key to interpreting the data in the data files.

dataset and *table* match the file name of the data file(s) that contain the listed variables.

table_var codes provide the link to the column headers in the data files.

label describes the variable represented in the corresponding column in the data files.

data_year is the year represented by the data in a given column, which may be different from the year of the dataset. For example, a table may describe population growth over time, with population counts from several years prior to the census.

universe describes the scope of who or what is covered by the variable. For example, data on marital status or economic activity may only cover persons over a certain age.

agg_method describes the general method used to aggregate information from individual census responses to calculate the summary values in the table. The most common aggregation methods are count and percent.

agg_detail provides additional aggregation details necessary to fully describe how the variable was calculated. For example, aggregation details may include units of measurement, numerators and denominators of ratios, or scaling factors.

	A	B	C	D	E	F	G	H
1	dataset	table	table_var	label	data_year	universe	agg_method	agg_detail
2	AL2012ag	AAC	AAC001	Total agricultural holdings	2012	Agricultural holdings	count	
3	AL2012ag	AAC	AAC002	Agricultural holdings with agricultural land	2012	Agricultural holdings	count	
4	AL2012ag	AAC	AAC003	% Agricultural holdings with agricultural land	2012	Agricultural holdings	percent	
5	DZ1998pop	AAC	AAC001	Single	1998	Resident Population of Ordinary and Collective Households	count	
6	DZ1998pop	AAC	AAC002	Married	1998	Resident Population of Ordinary and Collective Households	count	
7	DZ1998pop	AAC	AAC003	Divorced	1998	Resident Population of Ordinary and Collective Households	count	
8	DZ1998pop	AAC	AAC004	Widowed	1998	Resident Population of Ordinary and Collective Households	count	
9	DZ1998pop	AAC	AAC005	Not stated	1998	Resident Population of Ordinary and Collective Households	count	
10	DZ1998pop	AAC	AAC006	Total	1998	Resident Population of Ordinary and Collective Households	count	

Data files

Each data file contains the data from a table in the source document for one geographic level. (In cases where the source document included separate tables for sub-national geographic units, those tables have been combined into nation-wide data files.)

GISJOIN codes provide the link between rows of data and polygons in the GIS boundary files. You may join data files to shapefiles in a GIS package using the GISJOIN field in both files.

The next set of columns (*g0, g1, g2*…) provides the names of the geographic units and their parent units.

The remaining columns provide the actual data. The codes in the header row (e.g., AAA001) correspond to variable descriptions in the codebook and data dictionary.

	A	B	C	D	E	F	G	H	I
1	GISJOIN	g0	g1	AAC001	AAC002	AAC003	AAC004	AAC005	AAC006
2	DZ01	Algeria	ADRAR	208624	89837	3785	9345	23	311615
3	DZ14	Algeria	CHLEF	578439	255259	4757	19982	258	858695
4	DZ25	Algeria	LAGHOUAT	207603	99603	2533	7171	216	317125
5	DZ34	Algeria	OUM EL bouaghi	341441	161760	2917	13021	31	519170
6	DZ06	Algeria	BATNA	639857	290862	7166	24674	65	962623
7	DZ08	Algeria	BEJAIA	553121	272943	3392	27241	143	856840
8	DZ09	Algeria	BISKRA	380561	175456	4585	15225	32	575858
9	DZ07	Algeria	BECHAR	149210	67195	2961	6173	6	225546
10	DZ10	Algeria	BLIDA	506328	252212	4909	20630	203	784283

Mapping with IHGIS Boundary Files

IHGIS provides GIS shapefiles delineating the boundaries of the geographic units described in the data tables. The boundary files enable you to map data in the tables by linking them to the boundaries.
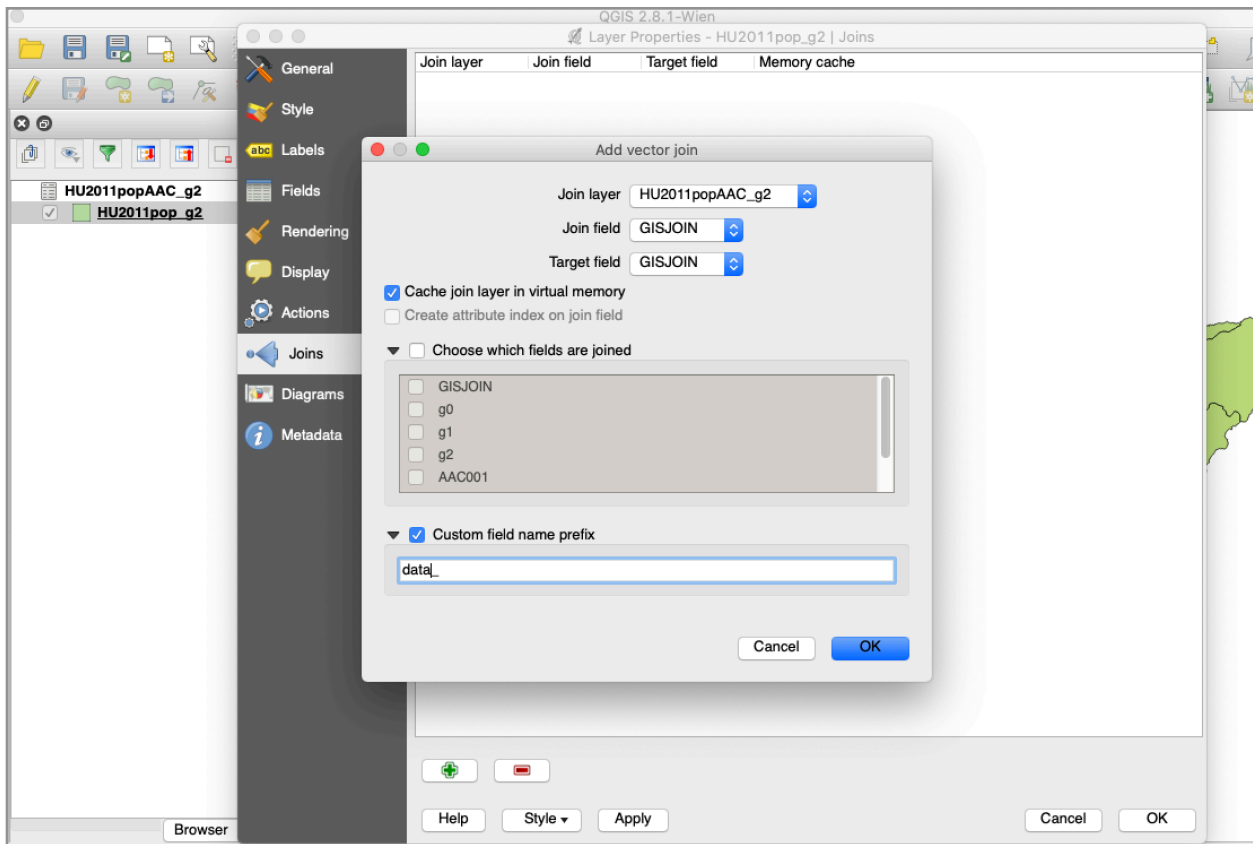
Download GIS boundary files

IHGIS boundary shapefiles are available from the Geography & GIS page. This page also provides more detailed information about IHGIS boundary shapefiles. If a dataset has more than one geographic level, for example, states and counties, a separate shapefile is provided for each level. The links on the Geography & GIS page will download .zip files containing the component files that make up each shapefile. In future versions of IHGIS, boundary files will be available through the Data Finder.

Join data tables to boundary files

After downloading and unzipping a shapefile, you can open it in a GIS package. You can then join data from a data file to create a map. This video tutorial illustrates the process in ArcGIS using NHGIS data and shapefiles. The process is very similar for IHGIS data and shapefiles. The steps below use QGIS (a free, open source GIS package). The process and concepts will be similar in other GIS packages.

1) Add both the shapefile and data table to a map. For this example, we will map county population density in Hungary using the Hungary 2011 county (g2) shapefile from the Geography & GIS page and the "Change in population density" table (HU2011popAAC_g2) from the Data Finder. (In QGIS, you should add the data table using the Add Delimited Text Layer dialog to ensure that data values are treated as numeric rather than strings.)

2) Open the properties for the shapefile and go to the Joins panel.

3) Add a new join, with the data file as the join layer and GISJOIN as both the Join field and Target field.

4) The attribute table for the shapefile should now include columns from the data table.

5) To map a variable, open the properties for the shapefile and go to the Style panel. For a numeric variable, select Graduated, then select the variable column you want to display. You will need to refer to the codebook.txt or datadict.csv file that came with your extract to interpret the content of each column. Select the number of classes, color ramp, classification mode, and any other options.
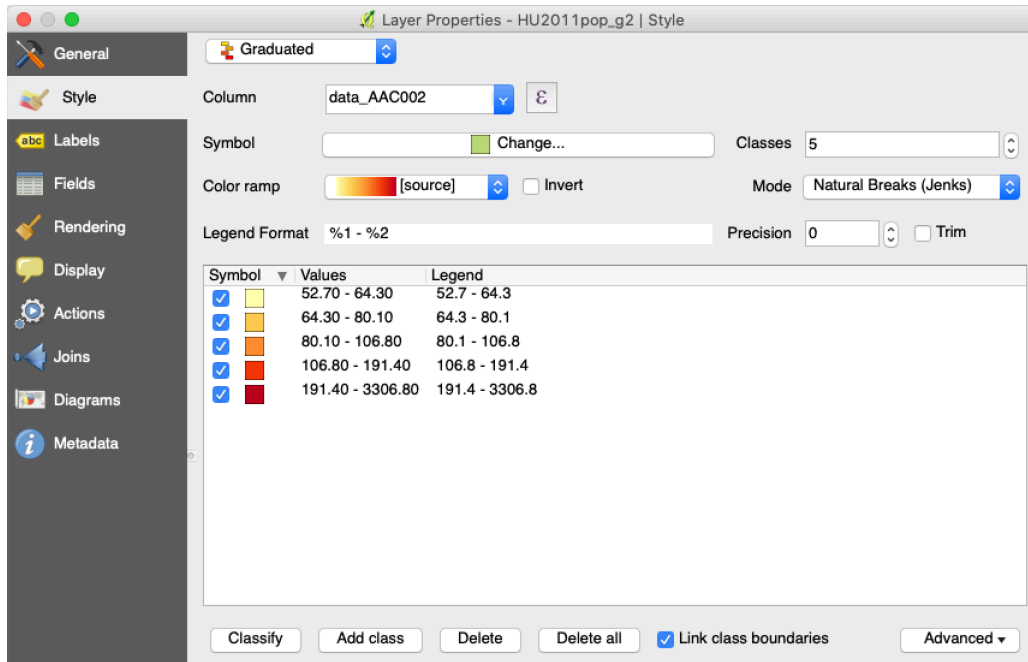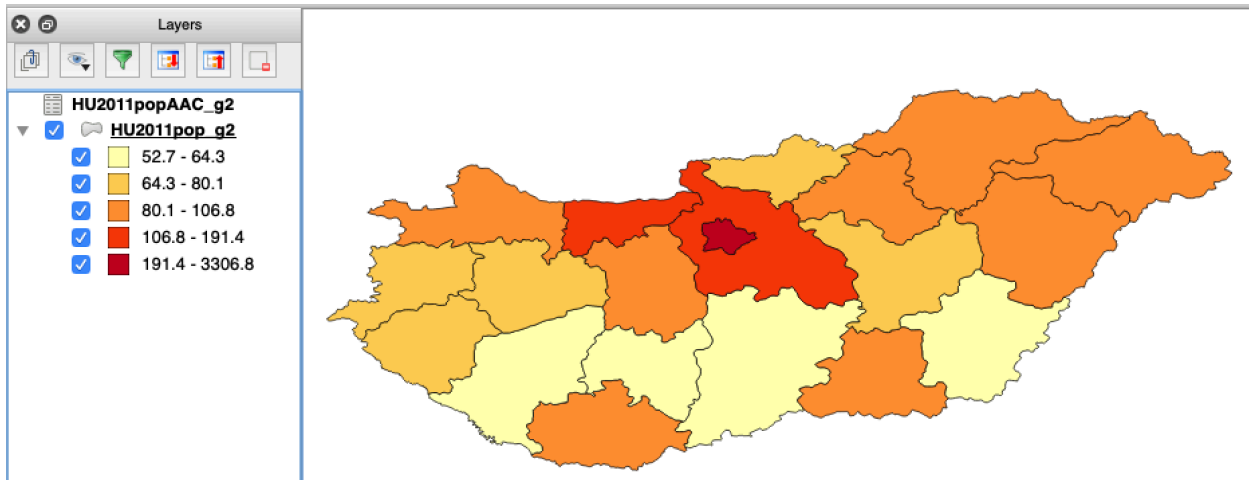
Joining a data table to a shapefile

Shapefile attribute table with joined variables from a data table

Setting a graduated style to map a data table variable

Map of a data table variable

Joining microdata sample tabulations to IPUMS International shapefiles

IHGIS data tabulated from IPUMS International microdata samples can be joined to boundary files available from the IPUMS International GIS Boundary Files page. IHGIS data files with "g1" or "g2" in the file name correspond to IPUMS International year-specific boundaries. IHGIS data files with "ga" or "gb" in the file name correspond to IPUMS International harmonized boundaries.

Due to differences in naming and coding conventions, IPUMS International shapefiles do not include GISJOIN fields that can be used directly to join to IHGIS data tables. You can, however, construct fields that can be used to join data tables to boundaries based on available information.

Joining g1 tables to year-specific boundaries
Create a new field in the shapefile attribute table that concatenates the 2-character country code with the IPUMyyyy (where yyyy is the year) field in the shapefile. For example, for the Austria 1991 dataset, the formula for the new field in QGIS is, `concat('AT',"IPUM1991")`. The data table can then be joined based on GISJOIN in the data table and the new field in the shapefile.

Joining g2 tables to year-specific boundaries
There is some inconsistency among datasets for g2 coding conventions. One of the following methods should work.

Method 1: Create a new field in the IHGIS data table that consists of the last 6 digits in the GISJOIN code. Be sure the new field is created as a text field to retain any leading zeroes. For example, for the Austria 1991 dataset, the formula for the new field in QGIS is, `right("GISJOIN", 6)`. This new field can then be joined to the IPUMyyyy field in the shapefile.

Method 2: Create a new field in the shapefile attribute table that concatenates the 2-character country code, the PARENT field, and the field containing level-specific unit codes. (This final field will have a name based on the name of the geographic level, e.g., for Benin, the g2 level is communes and the name of the field is like COMN1992.) For example, for the Benin 1992 dataset, the formula for the new field in QGIS is, `concat('BJ',"PARENT","COMN1992")`. This new field can then be joined to the GISJOIN field in the IHGIS data file.

Joining ga or gb tables to harmonized boundaries
Create a new field in the shapefile attribute table that concatenates the 2-character country code with the GEOLEVEL<1/2> field in the shapefile. For example, for the Benin 1992 harmonized 2nd level, the formula in QGIS is, `concat('BJ', "GEOLEVEL2")`. The new field can then be joined to the GISJOIN field in the data table.

Note that the tabulated data in IHGIS are based on IPUMS International data as as of September 2018. In some cases, the IHGIS data tables may not align with the current shapefiles. If you encounter such a case, please let us know by emailing ipums@umn.edu.