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.

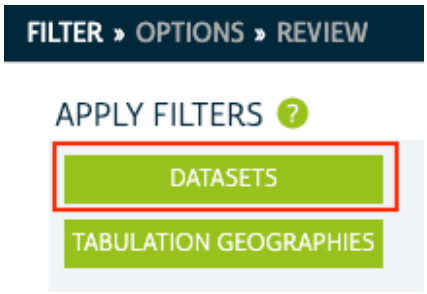
Apply filters

IHGIS offers two types of filters, Datasets and Tabulation Geographies. You can use either or both of these filters to narrow down the list of tables shown in the data selection grid.

Note: 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.

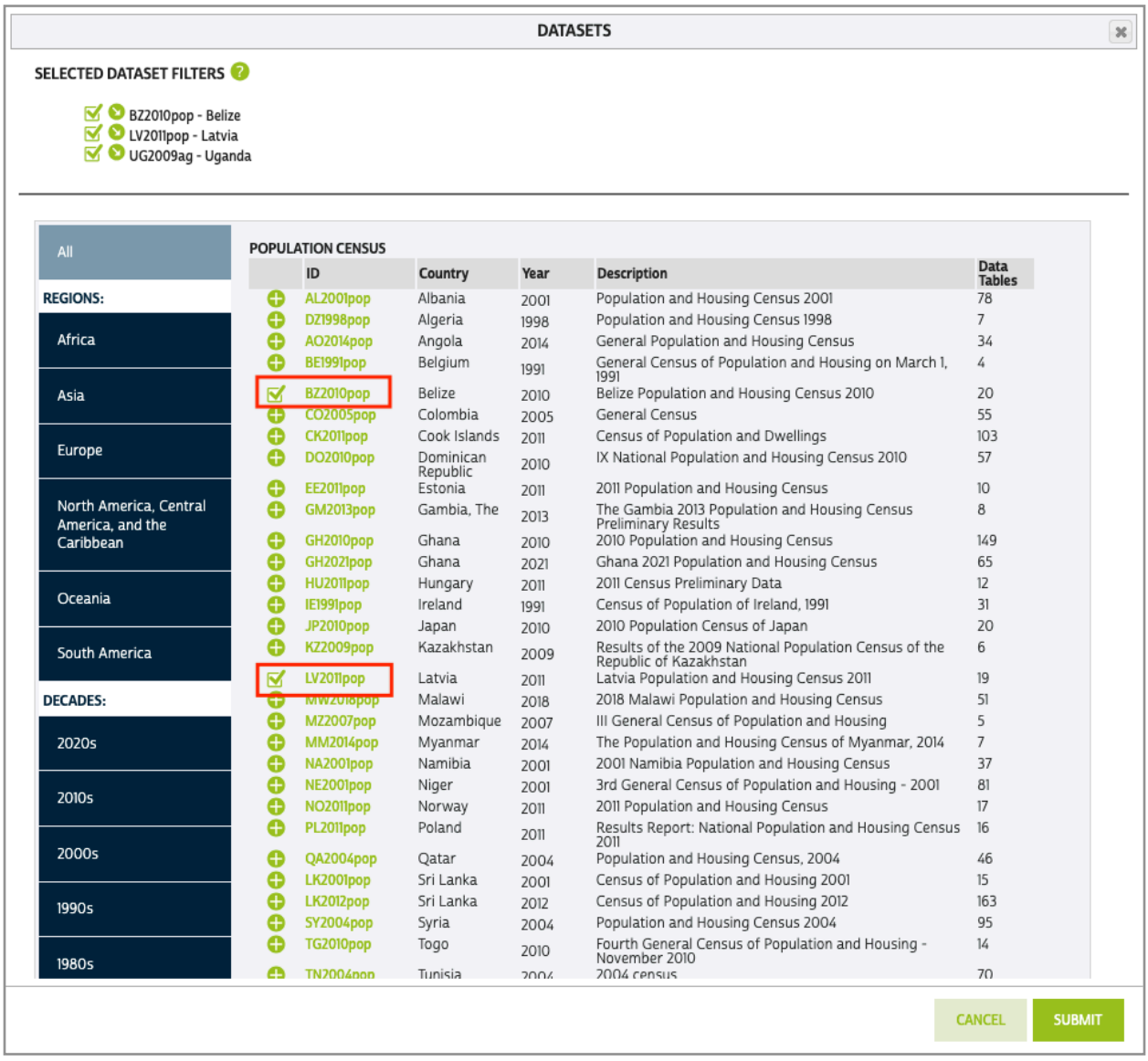
Filter by Dataset

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



Datasets are organized by type, including Population Census, Agricultural Census, and Microdata Sample Tabulations. You can filter datasets by world Region or by Decade by using the tabs on the left side of the filter pane.

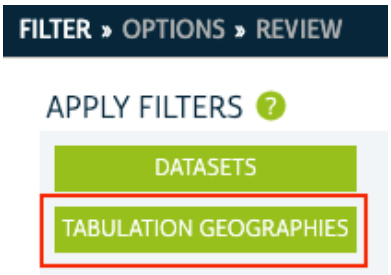
Click the green ‘+’ next to each dataset you wish to browse, then click “SUBMIT”



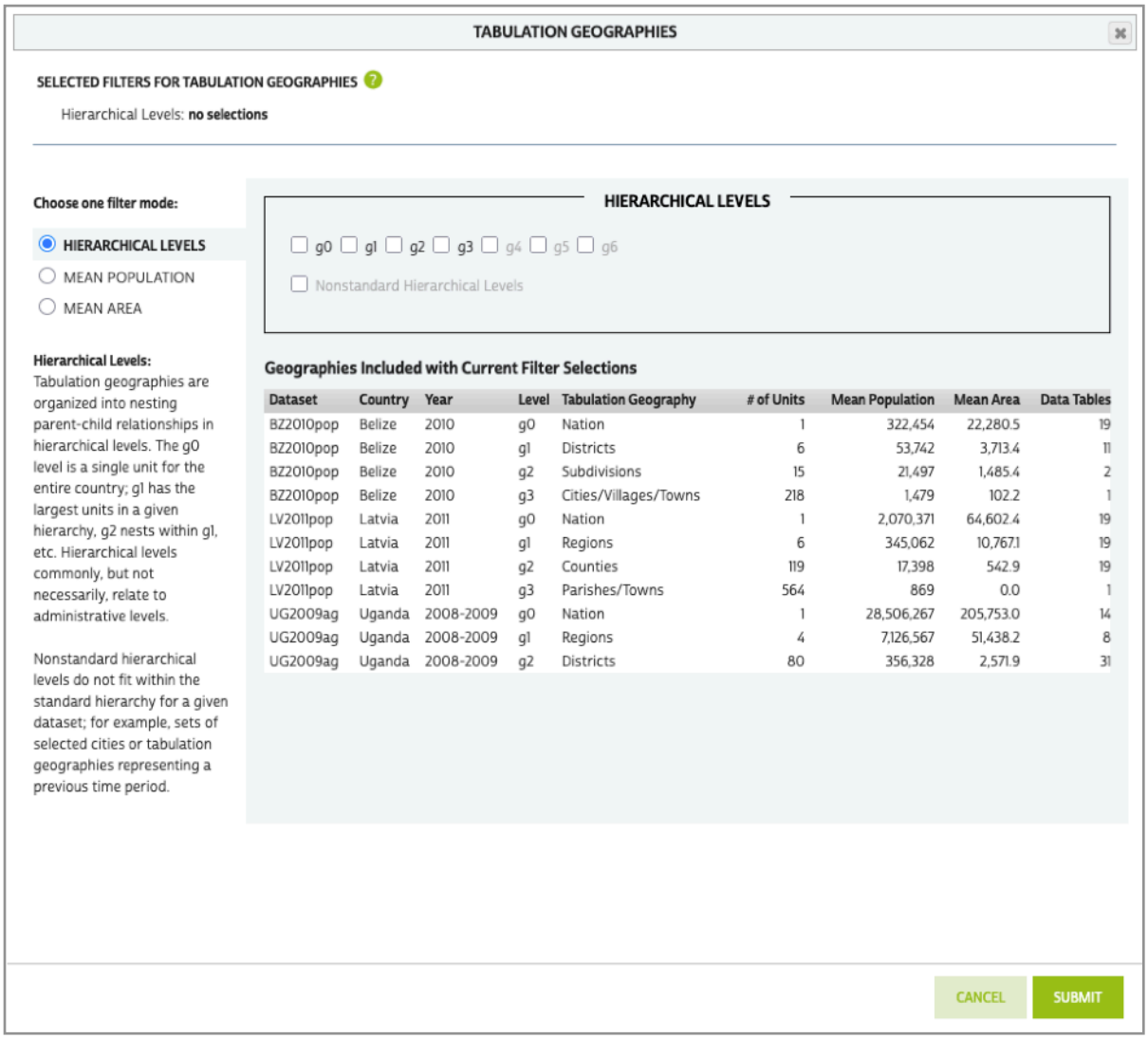
NOTE: The question mark icons throughout the site link to additional information about features of the interface.

Filter by Tabulation Geography

Tabulation geographies are sets of geographic units for which census data are tabulated. Tabulation geographies are organized in hierarchies, with smaller units (e.g., counties, districts) nested within larger units (e.g., states, provinces). In IHGIS, levels of the hierarchy are numbered with the coarser units having smaller numbers and finer units having larger numbers. For example, g2 units nest within g1 units. Some datasets also include tabulation geographies that do not fit within the standard hierarchy, such as urban centers or historical configurations of units. These tabulation geographies are called nonstandard hierarchical levels and are designated with letters, e.g., 'ga.'

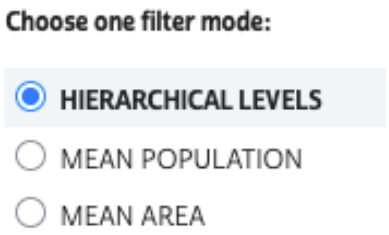


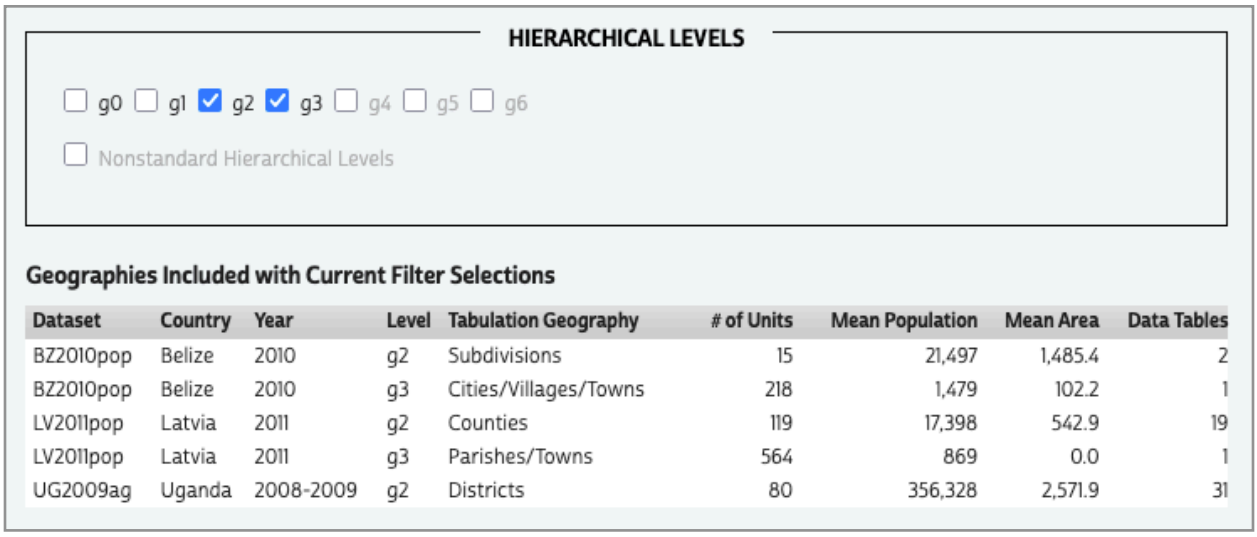
To display only tables that are available for particular tabulation geographies, click the "TABULATION GEOGRAPHIES" filter button to open the tabulation geographies filter pane. If you have applied a dataset filter, the tabulation geographies associated with those datasets will be listed.



Tabulation geographies can be filtered based on hierarchical level, mean population, or mean area.

To filter on hierarchical levels, click the check boxes for the level(s) of interest. This will narrow the list of tabulation geographies in the filter pane to just the selected hierarchical levels.

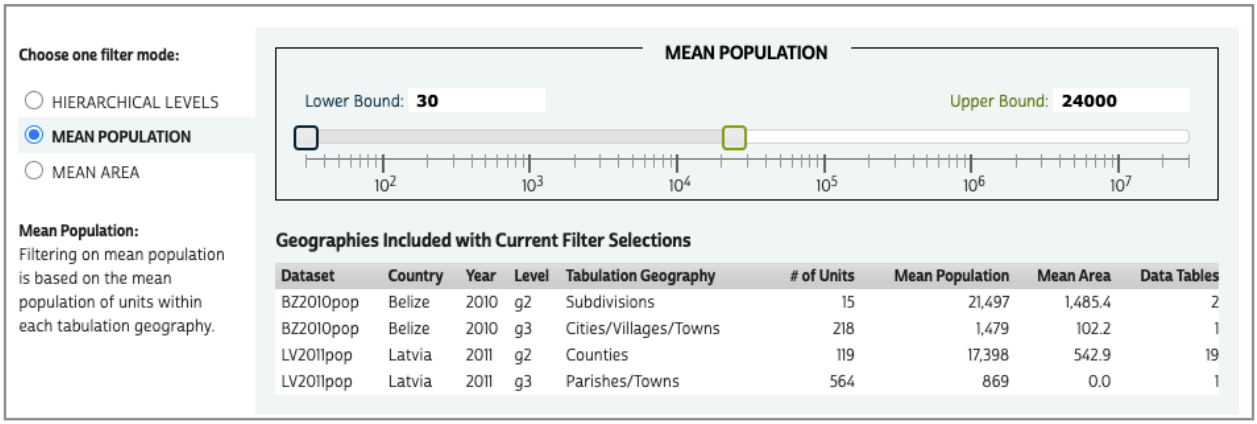




Dataset	Country	Year	Level	Tabulation Geography	# of Units	Mean Population	Mean Area	Data Tables
BZ2010pop	Belize	2010	g2	Subdivisions	15	21,497	1,485.4	2
BZ2010pop	Belize	2010	g3	Cities/Villages/Towns	218	1,479	102.2	1
LV2011pop	Latvia	2011	g2	Counties	119	17,398	542.9	19
LV2011pop	Latvia	2011	g3	Parishes/Towns	564	869	0.0	1
UG2009ag	Uganda	2008-2009	g2	Districts	80	356,328	2,571.9	31

To filter based on mean population or mean area, click the corresponding radio button. This will switch the filter mode from hierarchical level check boxes to a slider bar allowing you to select a lower and upper bound. Note that the slider uses a logarithmic scale.

Means are calculated across the units in each tabulation geography. For standard hierarchical levels, in which the units exhaustively cover the country, the mean is effectively the total population or area of the country divided by the number of units in the tabulation geography.



Dataset	Country	Year	Level	Tabulation Geography	# of Units	Mean Population	Mean Area	Data Tables
BZ2010pop	Belize	2010	g2	Subdivisions	15	21,497	1,485.4	2
BZ2010pop	Belize	2010	g3	Cities/Villages/Towns	218	1,479	102.2	1
LV2011pop	Latvia	2011	g2	Counties	119	17,398	542.9	19
LV2011pop	Latvia	2011	g3	Parishes/Towns	564	869	0.0	1

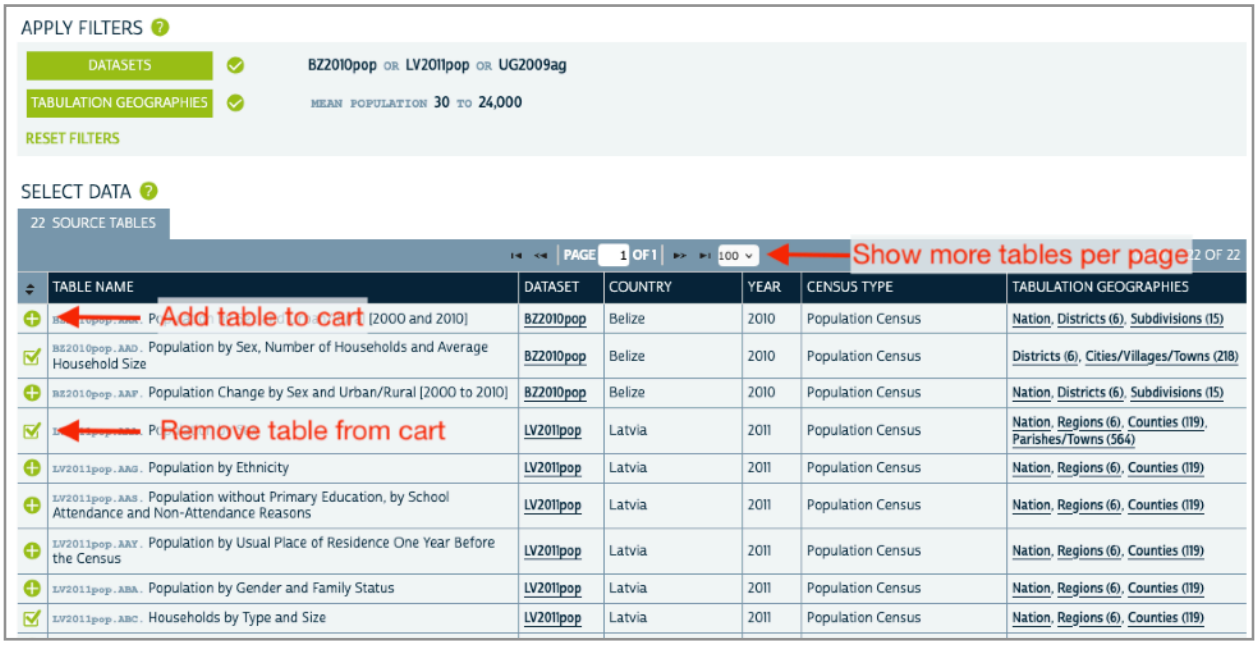
Note that only one filter mode—hierarchical level, mean population, or mean area—can be applied at a time. Click “SUBMIT” to view tables that are available for the tabulation geographies meeting the selected criteria.

Select tables

After applying dataset and/or tabulation geography filters, data tables available for those datasets/geographies 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.



You can also view metadata about the datasets and tabulation geographies by clicking the links in those columns. See the View Metadata section, below, for more detail.

Data cart

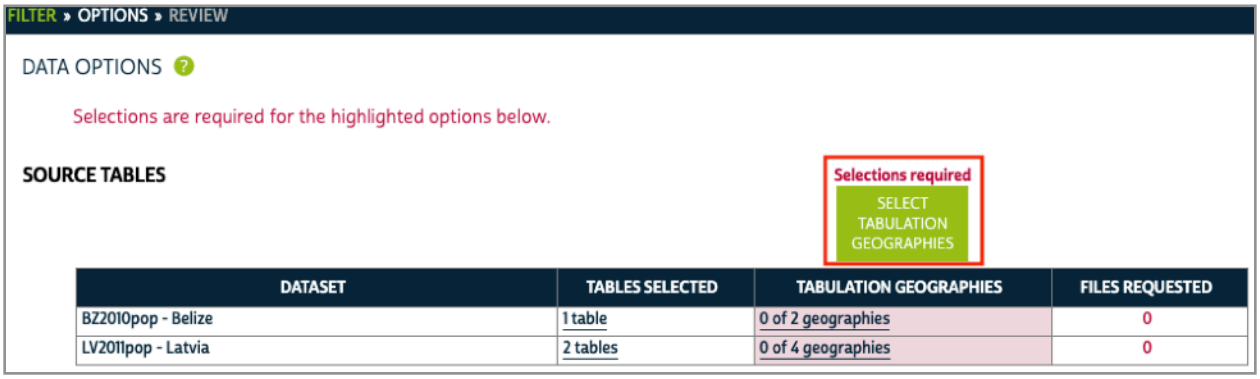
The tables you select will appear in your Data Cart. If you wish to review your selections, click “SHOW SELECTIONS,” which will open a view similar to the Select Data grid, showing only the tables you have selected. You may remove tables from your cart by clicking the green checkmark in this view.

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

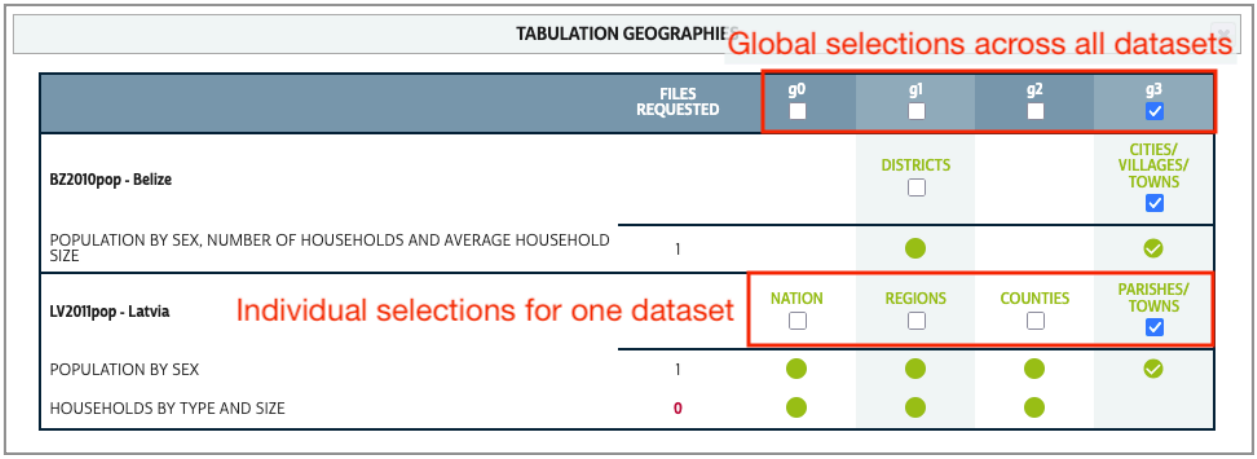
Data options

In the Data Options step you will select the tabulation geographies you want to include in your extract. You will receive one file for each selected tabulation geography/table combination (provided the tabulation geography is available for that table). This structure facilitates joining to GIS files of the unit boundaries for mapping and spatial analysis.

Click the “SELECT TABULATION GEOGRAPHIES” button to to open the Tabulation Geography Selection pane.

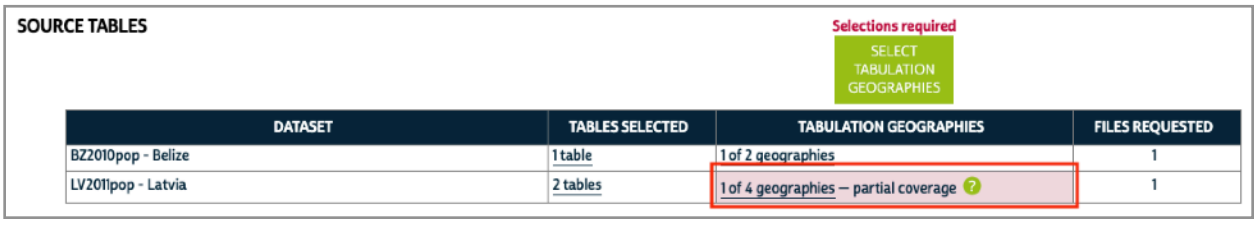


You can select hierarchical levels to apply across all datasets or individual tabulation geographies by dataset.

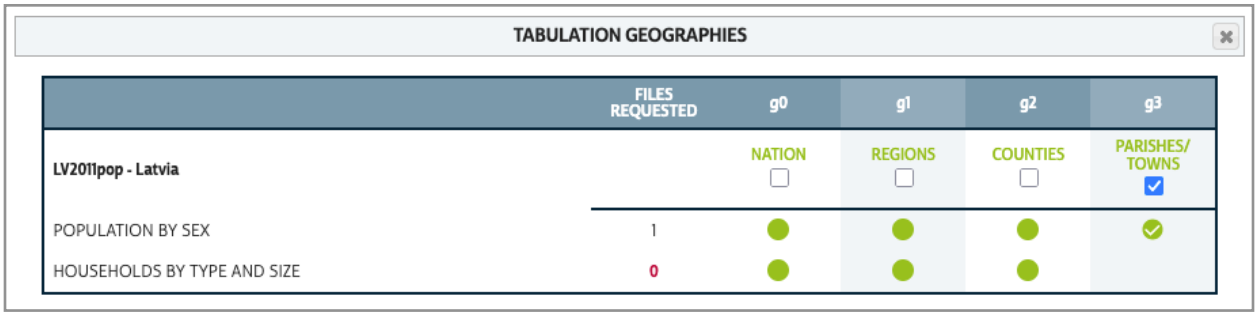


Each green check mark represents a tabulation geography/table combination for which you will receive a data file. You must select at least one available tabulation geography for each table in your cart. Tables for which no available tabulation geography has been selected will have a red “0” under “Files Requested” in the Tabulation Geography Selection pane.

On the Data Options page, you will be alerted about tables without available tabulation geographies selected with red shading and a “partial coverage” indicator.

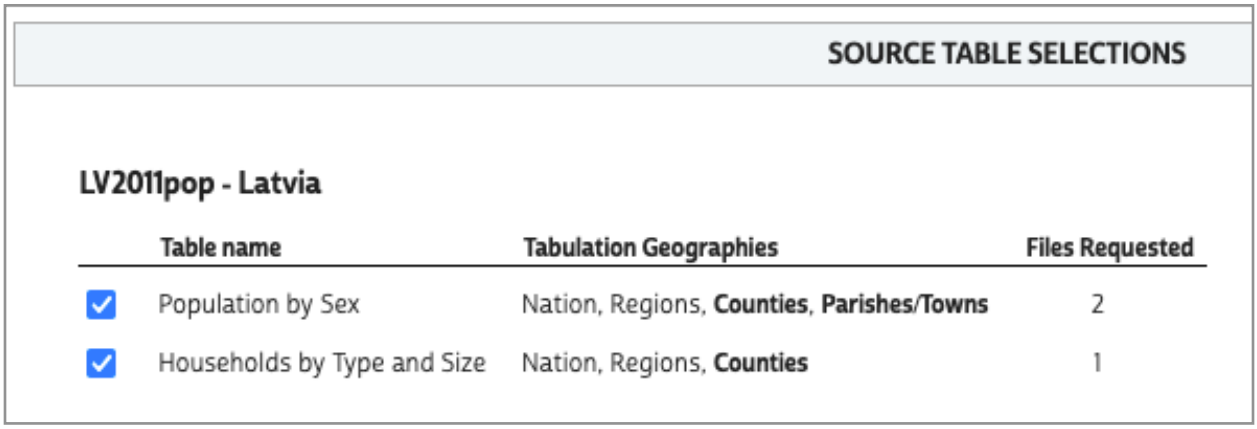


You can open a Tabulation Geography Selection pane for an individual dataset by clicking the "X of Y geographies" link. This pane will show only the tabulation geographies for that dataset, and will not include the global selection options.



If you had a tabulation geography filter applied when you added a table to the Data Cart, tabulation geographies meeting the filter criteria will be selected by default. Because tabulation geography selections are applied at the dataset level, the geographies will also be selected for any other tables in the dataset.

The Data Options page also allows you to view a summary of your selected tables and their available geographic levels for each dataset by clicking the "*X* tables" link under Tables Selected.



Tabulation geographies in bold have been selected. You can remove a table from your Data Cart by clicking the blue check mark.

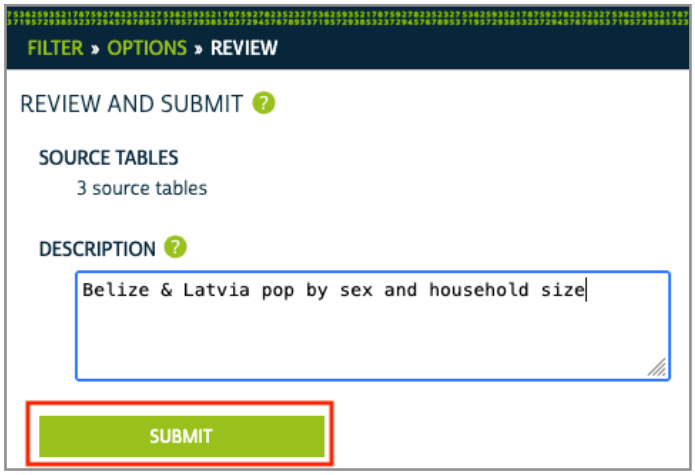
Move to the final Review and Submit step by clicking "CONTINUE" in the Data Cart.

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.

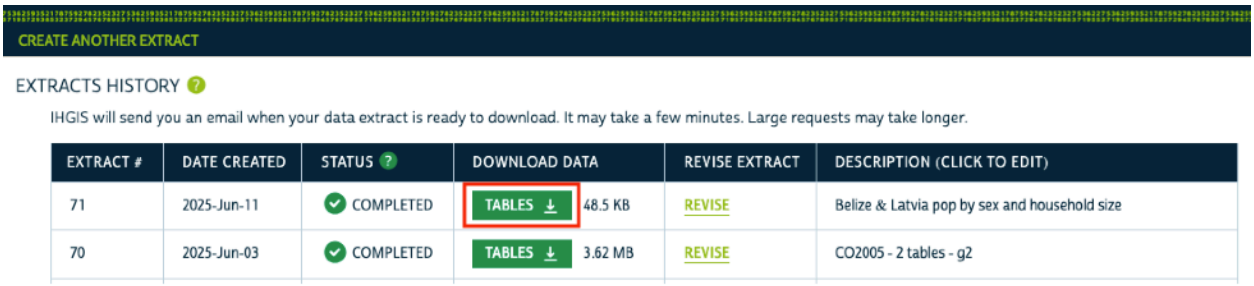
Click “SUBMIT” to start processing your extract. You must be logged in to submit an extract. If you are not already logged in, you will be prompted to do so.

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 DATA.”

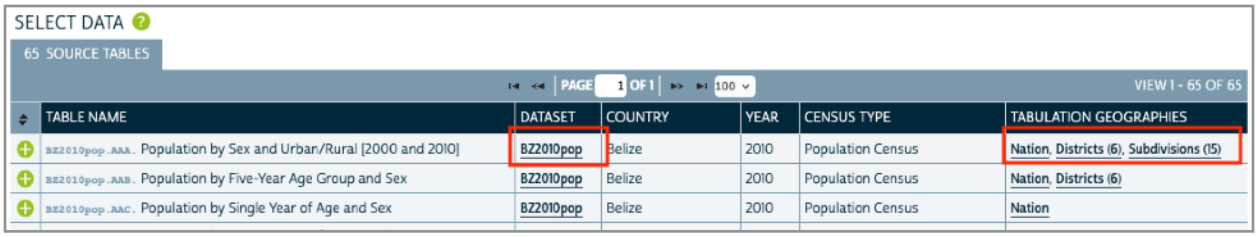


Clicking the “REVISE” link will return you to the extract builder with the selections for that extract in place, allowing you to make changes.

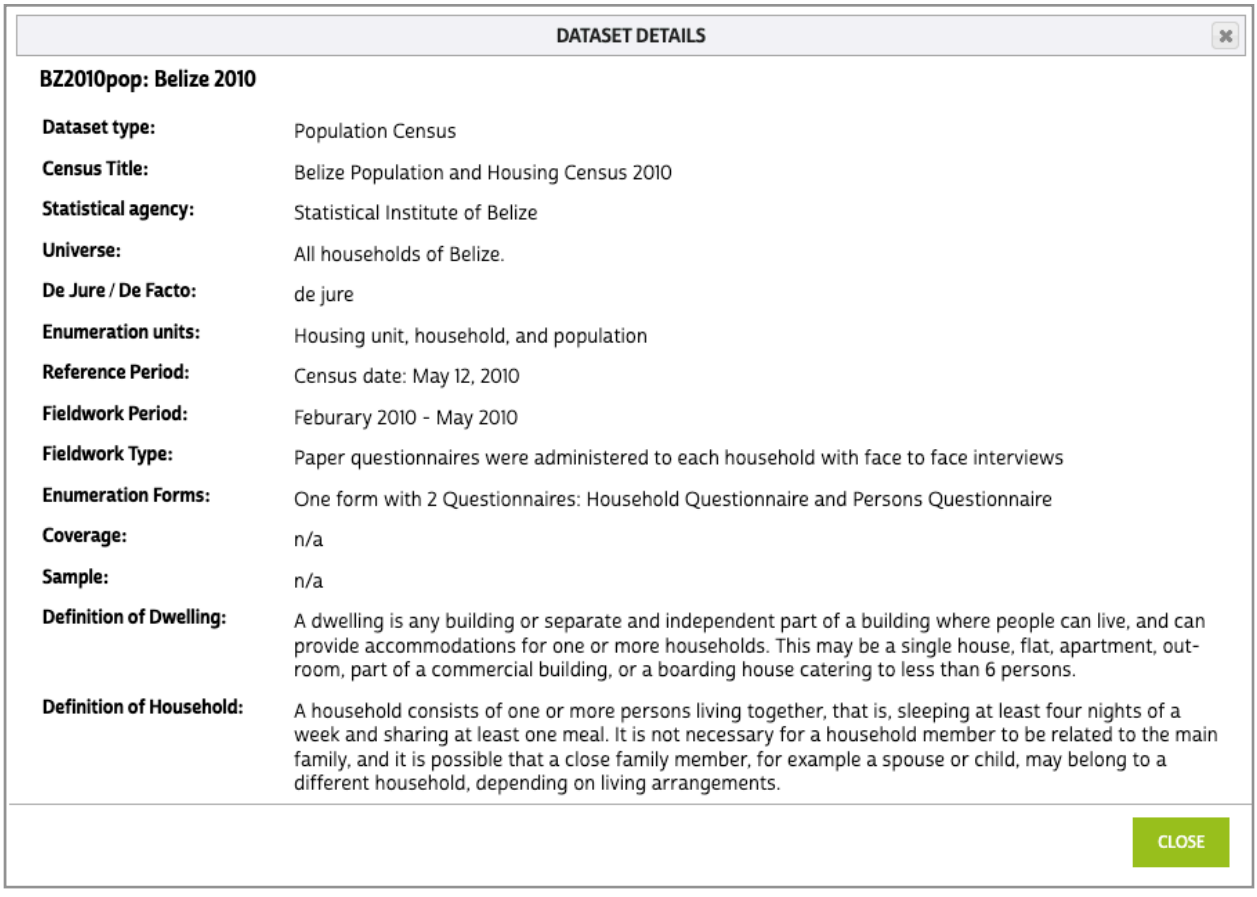
To access your Extracts History page from elsewhere in the system, click “MY DATA” from the top menu.

View Metadata

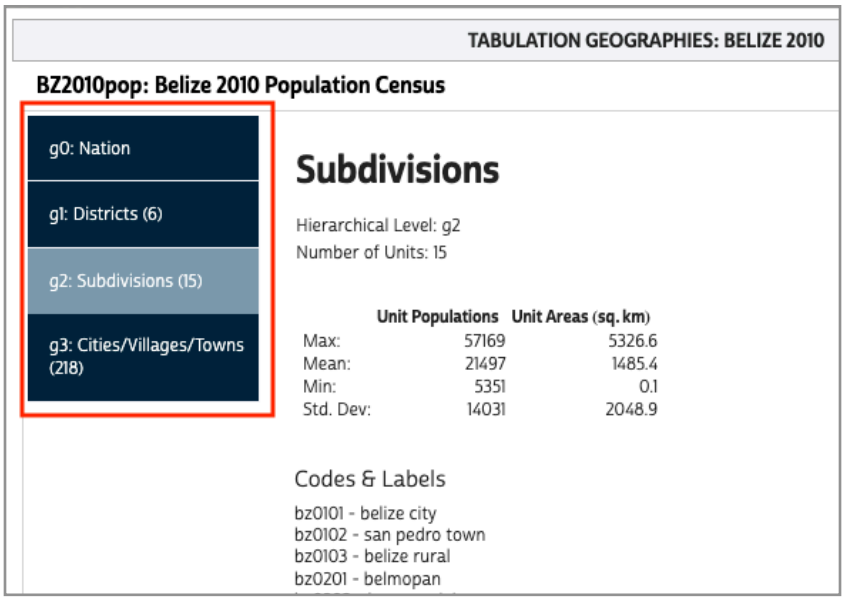
You can view metadata for datasets and tabulation geographies within the IHGIS Data Finder. You can access both types of metadata from links in their respective columns in the data selection grid, as well as elsewhere in the system.



Dataset metadata provides basic information about the census and how it was conducted, as well as definitions of dwellings, households, and group quarters.

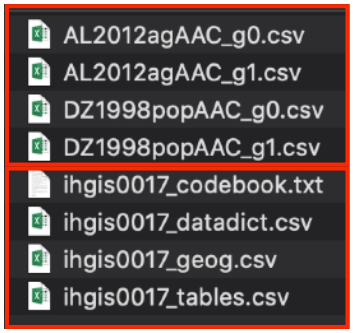


Tabulation geography metadata includes the number of units, summary statistics for unit population and area, and a sampling of codes and labels. The codes and labels list is not exhaustive, but is intended to provide an indication of the type of units that make up the geography. This may be useful if you are familiar with the country and know the names of several of the units at the level you are looking for. You can switch between tabulation geographies within a dataset using the tabs on the side of the panel.



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
- Hierarchical 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.

Tabulation geography metadata

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

	A	B	C
1	dataset	tabulation_geog	tabulation_geog_label
2	BZ2010pop	g3	Cities/Villages/Towns
3	LV2011pop	g2	Counties
4	LV2011pop	g3	Parishes/Towns

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	tabulation_geogs	tabulation_geog_labels	source_pub_eng	country
2	BZ2010pop	AAD	BZ2010pop.AAD	Population by Sex, Number of H	n/a	Total population	g3	Cities/Villages/Towns	Belize Population and Hou	Belize
3	LV2011pop	AAA	LV2011pop.AAA	Population by Sex	SG11-01	Resident population	g2; g3	Counties; Parishes/Towns	Latvia Population and Hou	Latvia
4	LV2011pop	ABC	LV2011pop.ABC	Households by Type and Size	TSG11-28	Private households	g2	Counties	Latvia Population and Hou	Latvia

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	BZ2010pop	AAD	GISJOIN	GIS join match code	2010			
3	BZ2010pop	AAD	AAD001	Total	2010	Total population	count	
4	BZ2010pop	AAD	AAD002	Male	2010	Total population	count	
5	BZ2010pop	AAD	AAD003	Female	2010	Total population	count	
6	BZ2010pop	AAD	AAD004	Households	2010	Households	count	
8	LV2011pop	AAA	GISJOIN	GIS join match code	2011			
9	LV2011pop	AAA	AAA001	Total population	2011	Resident population	count	
10	LV2011pop	AAA	AAA002	Males	2011	Resident population	count	
11	LV2011pop	AAA	AAA003	Females	2011	Resident population	count	
12	LV2011pop	ABC	GISJOIN	GIS join match code	2011			
13	LV2011pop	ABC	ABC001	Private households	2011	Private households	count	

Data files

Each data file contains the data from a table in the source document for one tabulation geography. (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.

GISJOIN	g0	g1	g2	g3	AAD001	AAD002	AAD003	AAD004	AAD005
BZ010327	Belize	belize	belize rural	st. paul's bank	153	79	74	37	4.14
BZ010328	Belize	belize	belize rural	western paradis	1258	599	659	348	3.61
BZ010329	Belize	belize	belize rural	willows bank	185	97	88	46	4.02
BZ010330	Belize	belize	belize rural	other - belize	818	511	307	272	3.01
BZ020101	Belize	cayo	belmopan	belmopan	13931	6775	7156	3463	4.02
BZ020201	Belize	cayo	benque viejo	benque viejo	6148	3057	3091	1416	4.34
BZ020301	Belize	cayo	san ignacio/santa elena	san ignacio	10489	5129	5360	2598	4.04
BZ020302	Belize	cayo	san ignacio/santa elena	santa elena	7389	3622	3767	1753	4.22
BZ020401	Belize	cayo	cayo rural	arenal	612	320	292	118	5.19
BZ020402	Belize	cayo	cayo rural	armenia	1395	723	672	278	5.02
BZ020403	Belize	cayo	cayo rural	billy white	586	286	300	113	5.19
BZ020404	Belize	cayo	cayo rural	blackman eddy	534	287	247	110	4.85

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.

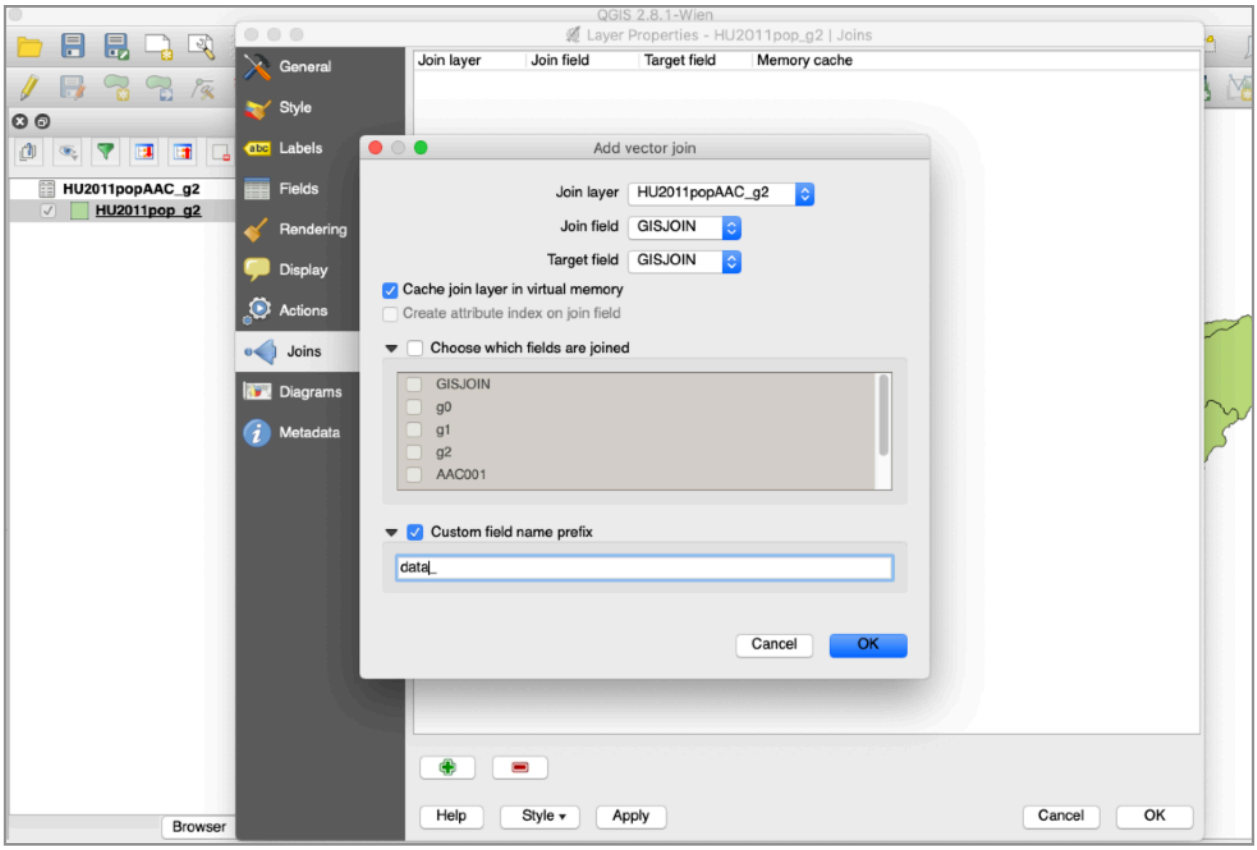
Download GIS boundary files

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 tabulation geography, for example, states and counties, a separate shapefile is provided for each geography. 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 (2001 and 2011)” table (HU2011popAAD_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.

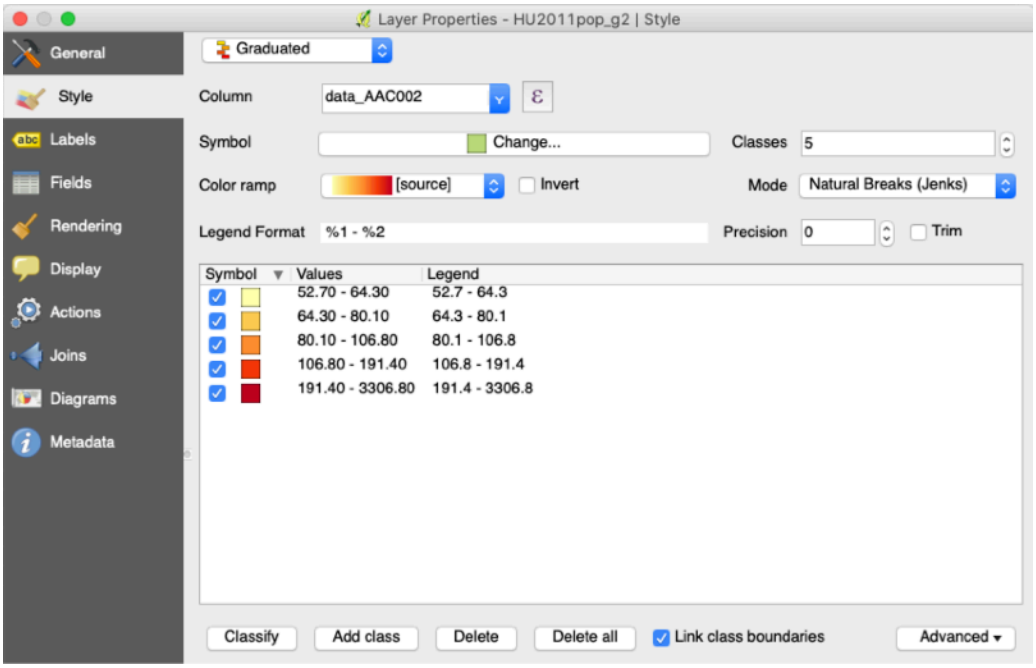


Joining a data table to a shapefile

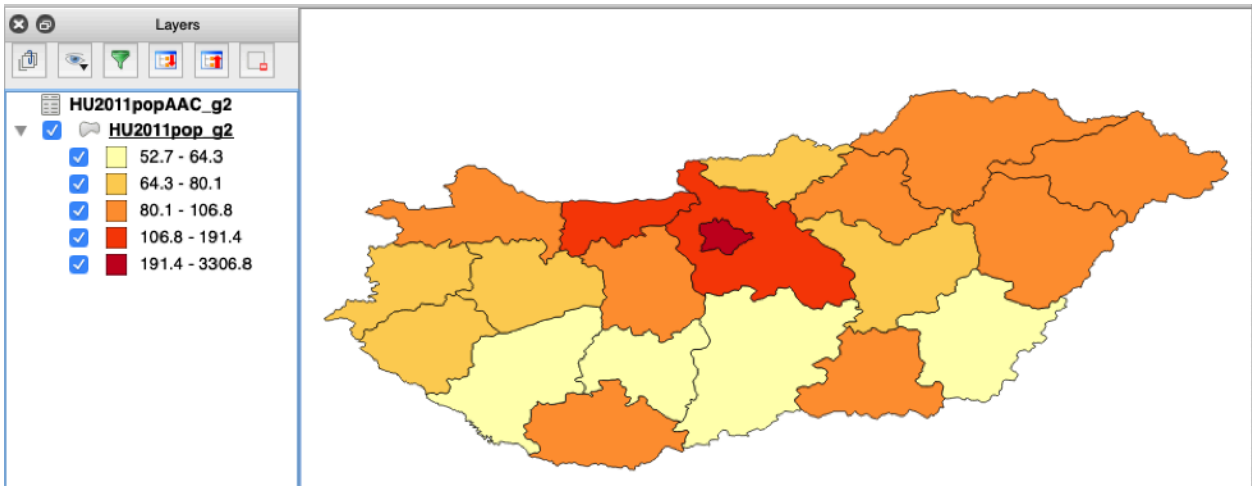
Attribute table - HU2011pop_g2 :: Features total: 20, filtered: 20, selected: 0

	label	parent	GISJOIN	data_g0	data_g1	data_g2	data_AAC001	data_AAC002	data_AAC003
0	Bács-Kiskun	HU33	HU331	Hungary	southern gre…	bács-kiskun	64.7	61.7	-3
1	Baranya	HU23	HU231	Hungary	southern tra…	baranya	92	87.6	-4
2	Békés	HU33	HU332	Hungary	southern gre…	békés	70.7	64.3	-6
3	Borsod-Abaú…	HU31	HU311	Hungary	northern hun…	borsod-abaúj…	102.7	95.3	-7
4	Budapest	HU10	HU101	Hungary	central hung…	budapest	3385.7	3306.8	-79
5	Csongrád	HU33	HU333	Hungary	southern gre…	csongrád	101.7	98.1	-4
6	Fejér	HU21	HU211	Hungary	central trans…	fejér	99.6	98.1	-2
7	Gyor-Moson-…	HU22	HU221	Hungary	western tran…	győr-moson-…	104.3	106.8	3
8	Hajdú-Bihar	HU32	HU321	Hungary	northern gre…	hajdú-bihar	89	88.4	-1
9	Heves	HU31	HU312	Hungary	northern hun…	heves	89.6	85.4	-4
10	Jász-Nagyku…	HU32	HU322	Hungary	northern gre…	jász-nagyku…	74.5	69.7	-5
11	Komárom-Es…	HU21	HU212	Hungary	central trans…	komárom-es…	139.8	135.2	-5
12	Nógrád	HU31	HU313	Hungary	northern hun…	nógrád	86.5	80.1	-6
13	Pest	HU10	HU102	Hungary	central hung…	pest	169.6	191.4	22
14	Somogy	HU23	HU232	Hungary	southern tra…	somogy	55.5	52.7	-3
15	Szabolcs-Sz…	HU32	HU323	Hungary	northern gre…	szabolcs-sza…	98.1	94.7	-3
16	Tolna	HU23	HU233	Hungary	southern tra…	tolna	67.4	62.5	-5
17	Vas	HU22	HU222	Hungary	western tran…	vas	80.4	77.3	-3
18	Veszprém	HU21	HU213	Hungary	central trans…	veszprém	82.3	78.8	-4
19	Zala	HU22	HU223	Hungary	western tran…	zala	78.6	74.9	-4

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.