Using the SABINS Data Finder

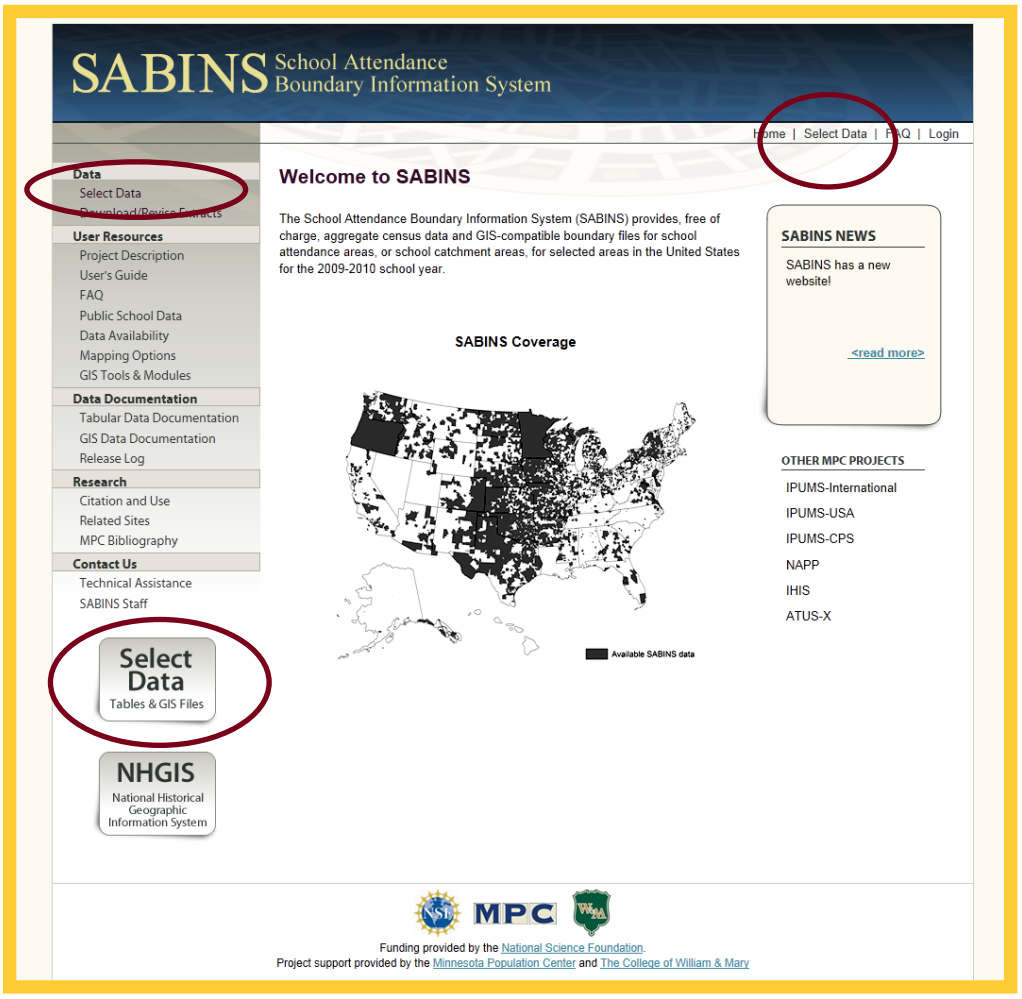
August 2011

This guide shows users how to use the School Attendance Boundary Information System (SABINS) to access GIS boundary files and data tables for school attendance boundaries in the 2009-2010 school year. The SABINS data finder provides aggregate data in a variety of formats, including those compatible with popular spreadsheet and statistical packages (i.e. Excel, SAS, SPSS, and Stata) and GIS software like Esri ArcGIS.

SABINS consists of 13 GIS boundary files, one for each K-12 grade level in school districts located throughout the nation. In addition, tabular data from the US Census 2010 Summary File 1 and PL 94-171 are also available for each grade level geography.

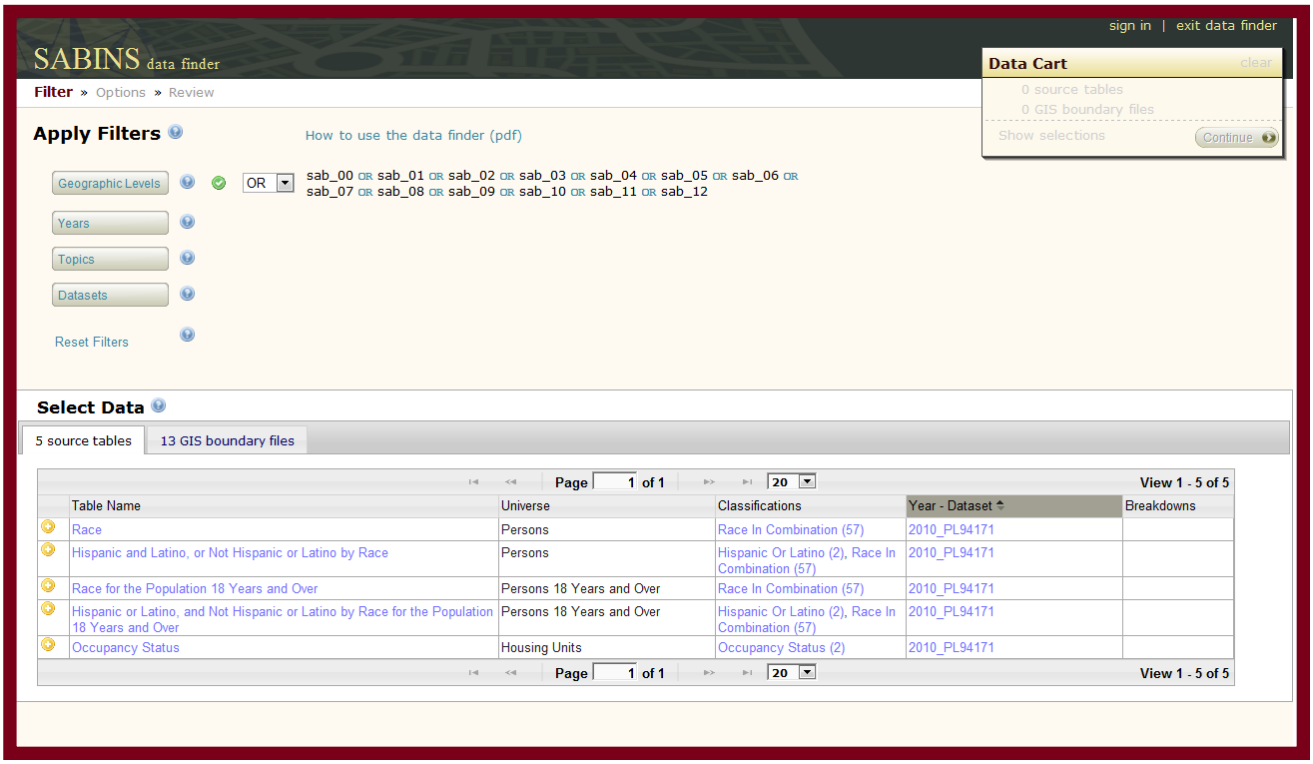
Funding provided by the National Science Foundation.
Project support provided by the Minnesota Population Center and The College of William and Mary.

Open the Data Finder



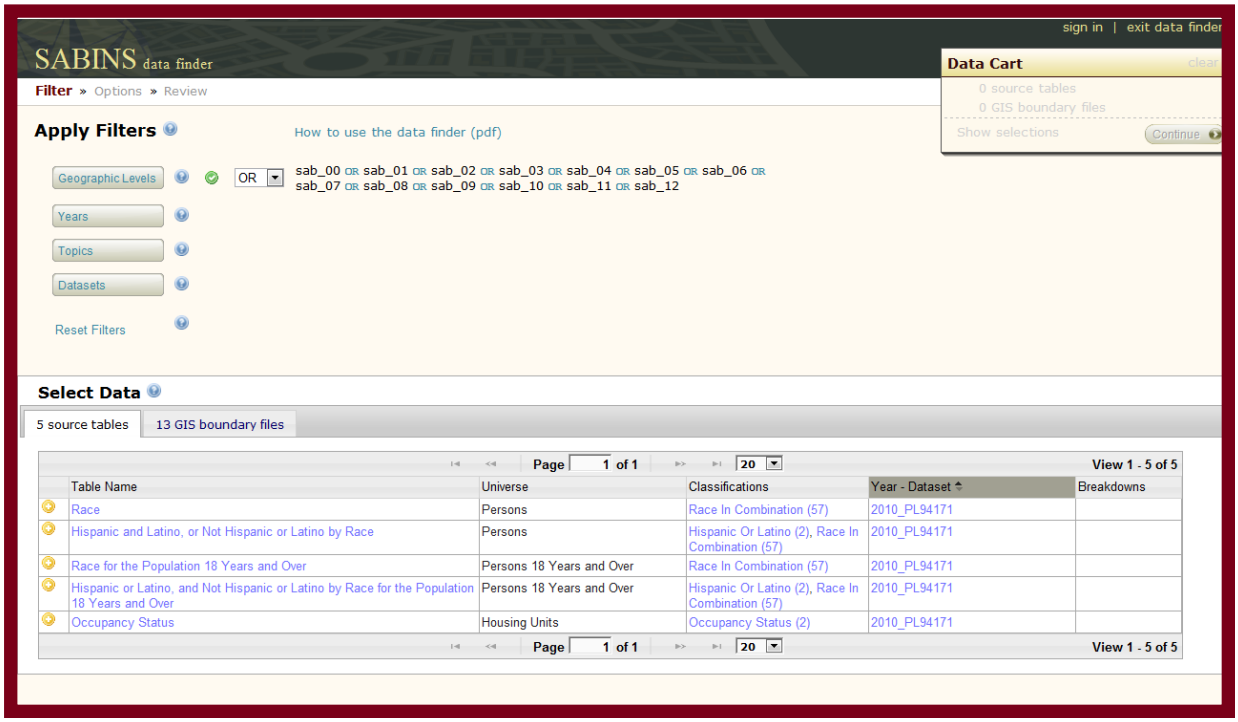
From the home page, click a "**Select Data**" link to proceed immediately to the SABINS data finder. You do not need to log in to browse and select data. Before submitting your extract request, however, you must log in or create an account at that time.

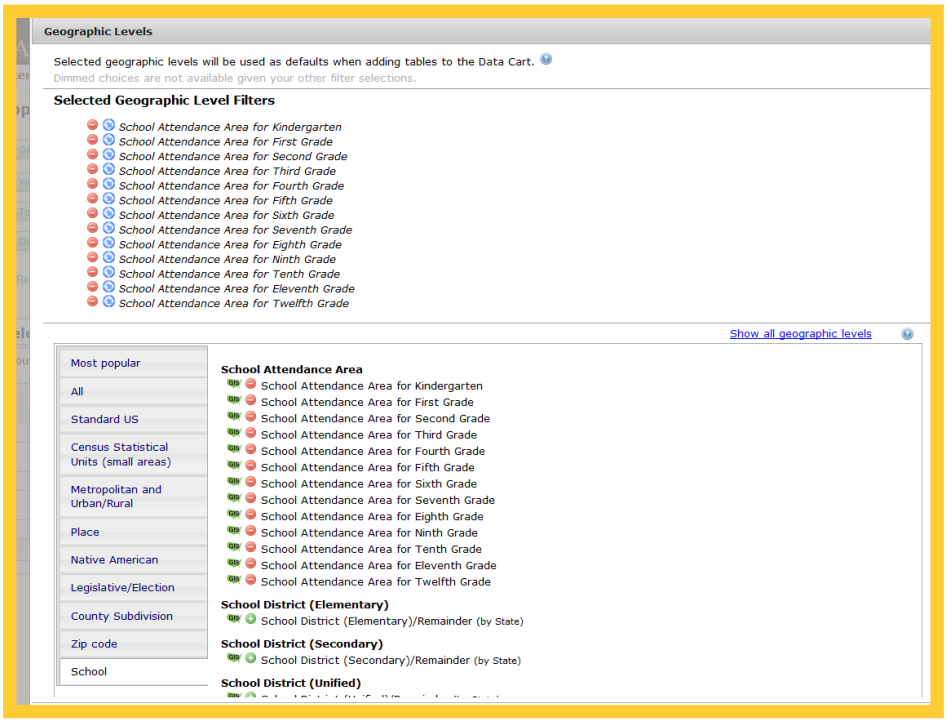
If you already have an account through the Minnesota Population Center that you use with other data projects like NHGIS, IPUMS, or NAPP, that account is the same as for SABINS.



Step 1 - Filter

The SABINS data finder has made it easy to find the school attendance boundary data by preselecting the appropriate filter. Filters are the tool users use to only show tables and GIS boundary files that they are interested in. With SABINS, each K-12 grade level is considered a separate geographic level much like census tracts or counties are separate geographic levels.



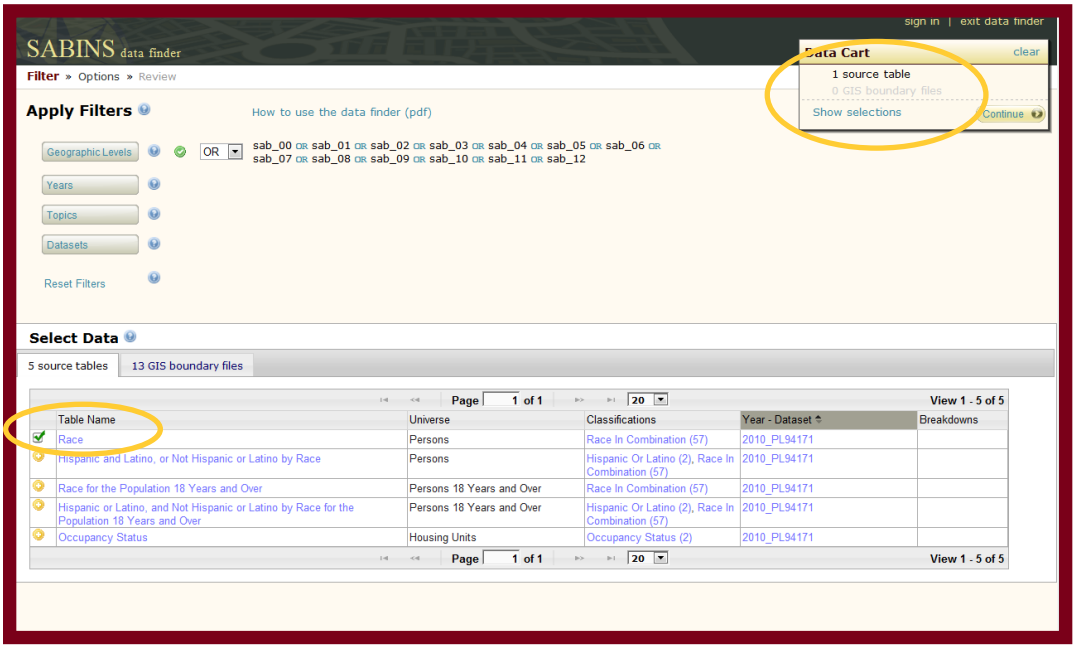


Question Mark icons contain additional, helpful information.

Step 1 - Filter

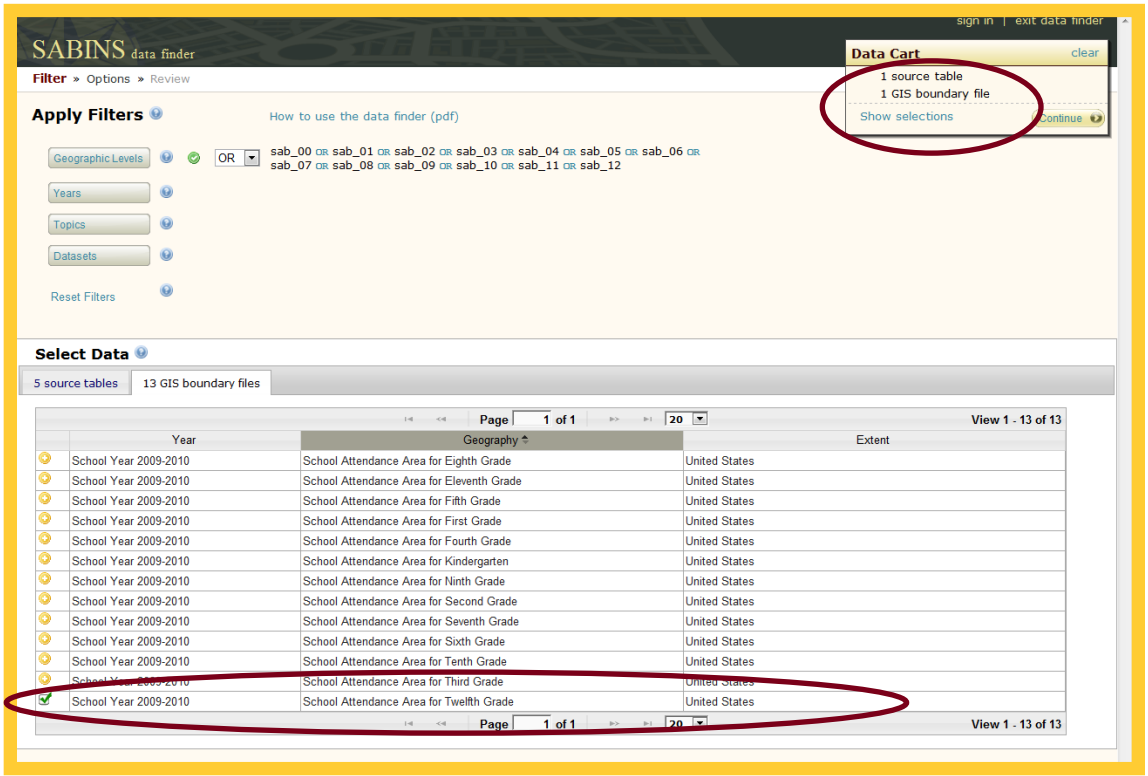
Clicking the plus sign next to a table, such as 'Race', adds that table to the Data Cart. Clicking the check mark would remove the table from the Data Cart. Switching over to the GIS boundary files tab, clicking the plus sign next to any of the 13 files, like the 'School Attendance Area for Twelfth Grade' adds it to the Data Cart.

Once you have finished filtering and making selections, you can click on *Show Selections* to review the tables and GIS files selected, or simply click **Continue** to move on to the next step. Conversely, you can start over if you wish by clicking the *clear* on the Data Cart to remove everything selected.



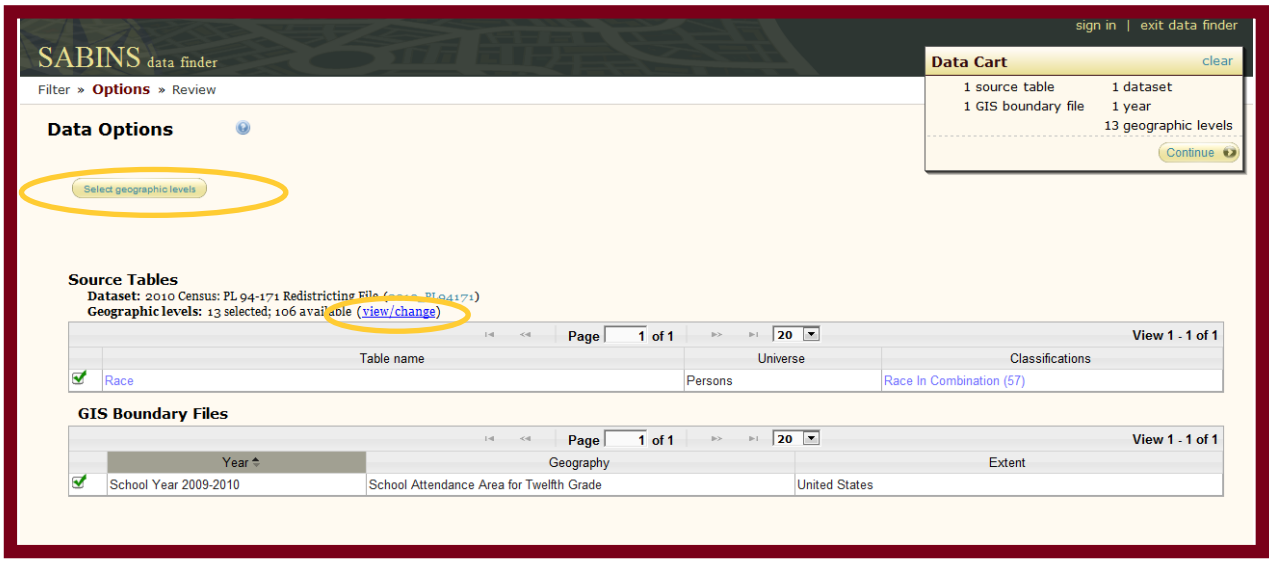
Hint!

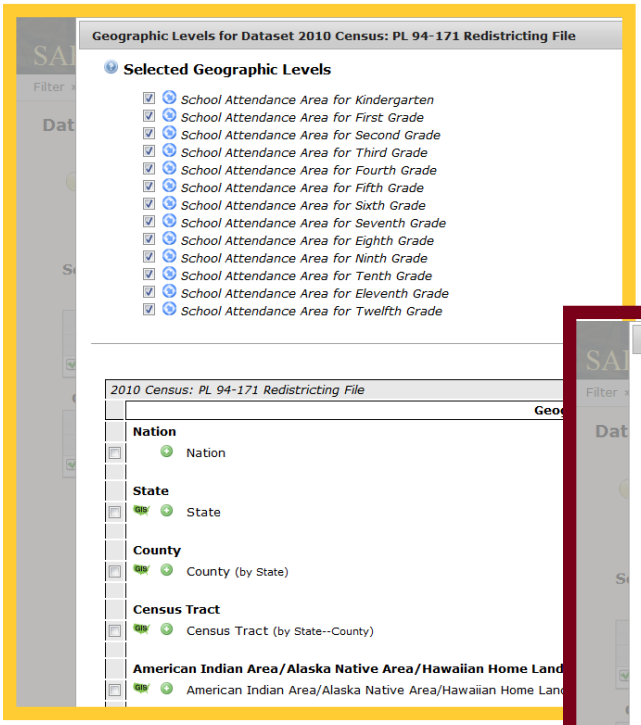
Click on blue text to reveal more detailed information.



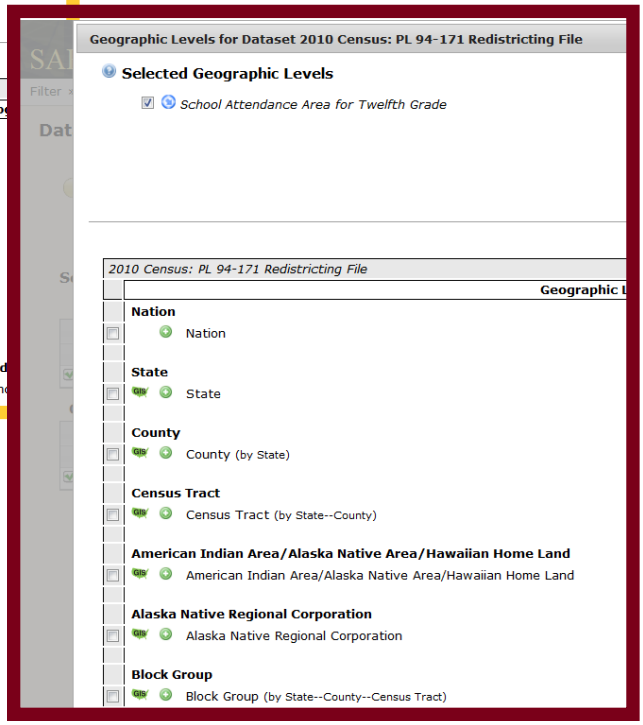
Step 2 - Options

Once the data have been chosen, the next step is to define for which Geographic Level you want data. Note that currently 13 geographic levels are selected. Clicking either on the **Select geographic levels** button, or the blue *view/change* text will open a window where the geographic selection level can be made. Let's unselect all of the geographic levels except for 'School Attendance Area for Twelfth Grade', as it will match with the GIS file we selected.



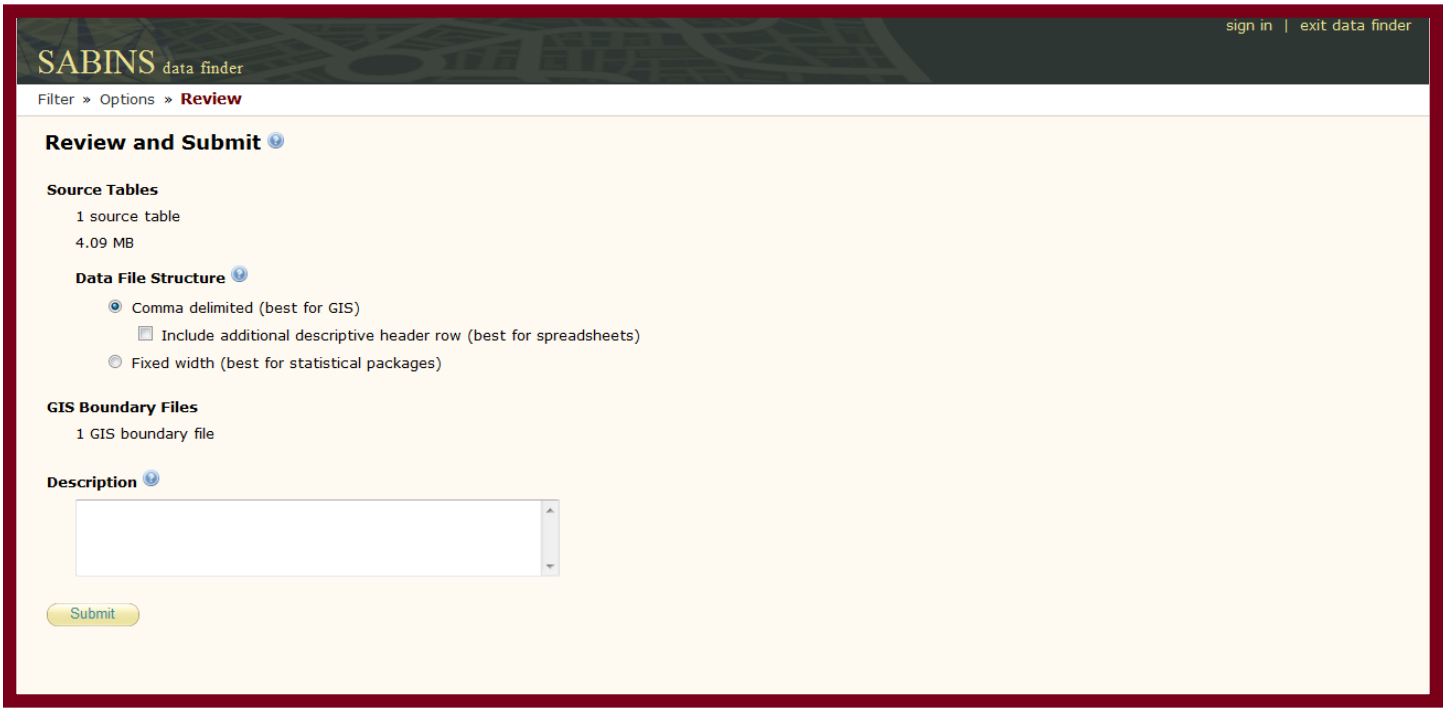


Click on **Submit** to confirm your selection. Notice that in the Data Cart, it now says that one geographic level is selected, rather than 13 as before. Select **Continue** in the Data Cart to proceed to the next step.



Step 3 - Review

Once you have selected the table and GIS file we want along with the desired geographic level, the final step before submitting your request is to choose the format in which you want the data tables. GIS files only come in the shapefile format, so no options are available. The data tables, however, can be prepared in a **Comma delimited** format, either with or without **additional descriptive header rows**; or the data can be downloaded in a **Fixed width** format.



Comma delimited is the ideal choice for using the data in a GIS. The field names consist of a unique ten character code created by SABINS/NHGIS that does not contain spaces or other special characters. This allows for the .csv file (comma delimited) to be added to the GIS without errors. This format can also be used with a spreadsheet software like Excel. To decipher the nondescript field names, you will need to use the codebook file that is automatically included in your data extract. This is a .txt file that includes important information about your extract. In addition to the field name key, the codebook also lists the table and dataset names you downloaded, along with citation information.

If your intentions are only to use Excel or another spreadsheet software, you may be better served by clicking the **Include additional descriptive header row** box. This option will return the same .csv file, but it will include an extra row of field names that can be read and understood without using the codebook. The format of these descriptive field names, however, prevents the data from being used in a GIS without error.

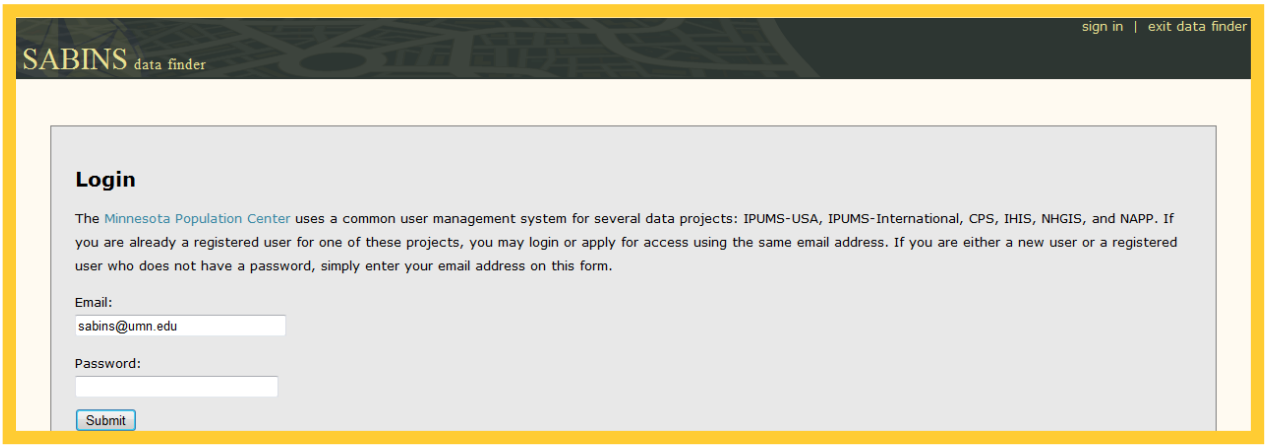
Fixed width is the best format to select when you will be using the data in a statistical package like SAS, SPSS, or Stata. This download option will provide the .dat fixed width file, in addition to three separate command files, one each for the aforementioned statistical packages. The command files are a .sas for SAS, .sps for SPSS, and .do for Stata. The .txt codebook file is also included in the download.

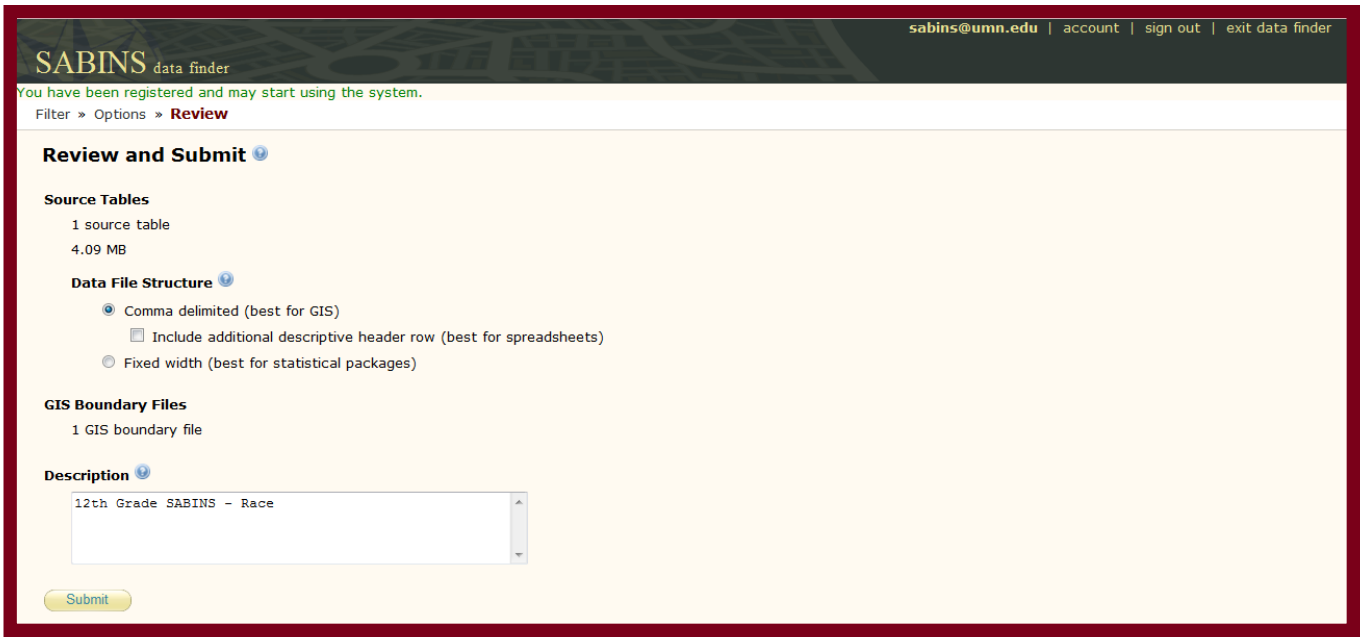
Step 3 - Review

For this example, because we are selecting a GIS boundary file, we will choose 'Comma delimited (best for GIS)' option. While only one data file structure can be chosen for each extract request, this tutorial will show you what the output would also look like if we chose the 'Fixed width (best for statistical packages)' option instead.

After making your selection on the data file structure, it is recommended that a brief description of your extract request be typed into the Description text box. This description will appear in the Extracts History page and the extract completion email sent to the user, and is the most efficient way of keeping track of each extract. Ideally, the description should be no more than a few key words. In our example, lets call our extract "12th Grade SABINS - Race." This way, you can easily remember what is included in the extract.

Once you click **Submit**, you will be taken to your Extracts History page if you are already logged in to SABINS. Otherwise, you will be prompted to login or set up a new account. Creating an account is easy with SABINS. Just enter your email address, leave the Password field blank, and click **Submit**. You will then be prompted to complete a simple account registration. Once completed, you will return to the previous page where you can submit your extract request.





Final Steps

Once on the Extracts History page, you can watch the progress of your extract request by refreshing your browser page. You'll notice that the status changes as it transitions from queued, in progress, and complete. Otherwise, you may exit SABINS and simply wait to receive the extract completion email with download link that will be sent to your email address on file.

The data tables and GIS are downloaded separately by clicking on *data* and *gis*. Both are zipped folders that contain the data you requested.

In addition to simply downloading your extract, from the Extracts History page you can revise your extract, resubmit your extract, or change its description.

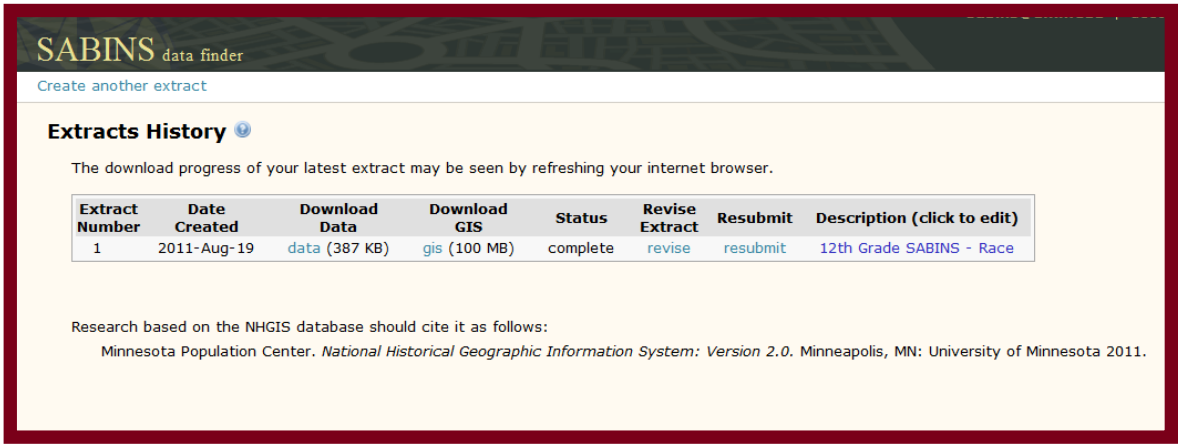
Revise Extract is very useful as it allows you to create a new extract with all the parameters preset from the original extract. If you want both the comma delimited and fixed width files, for example, revising the extract would be the fastest way of getting the two formats.

Resubmit allows you to request the same exact extract again. The downloadable data is cleared off the Extracts History page every two weeks to preserve storage space, while the parameters of the extract are maintained indefinitely. This method benefits SABINS immensely while keeping users' inconvenience to a minimum.

Description can be altered by clicking on the existing text or blank space if no text was entered.

In Process

Complete

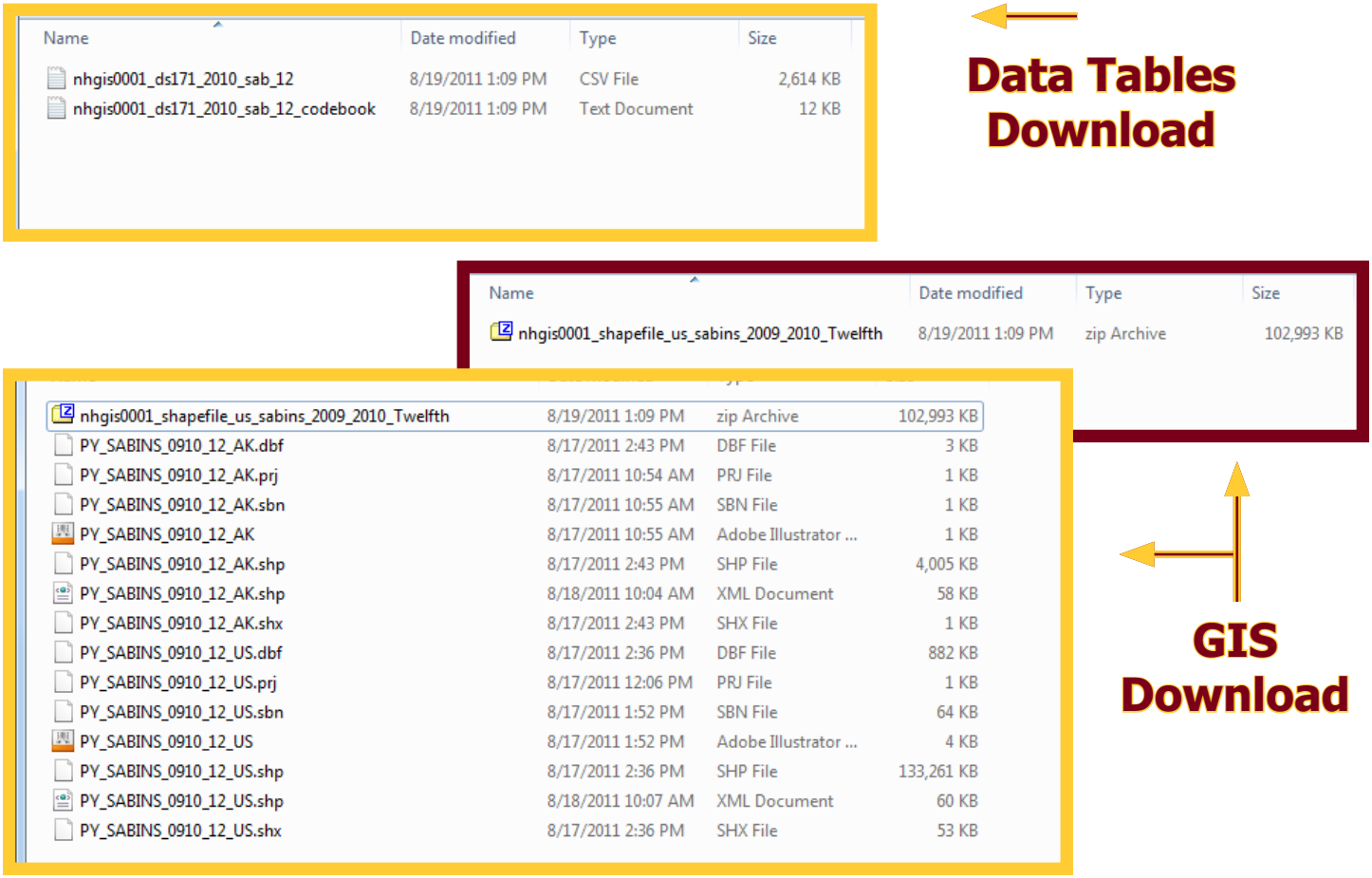


Final Steps

Download Your Data

After you have clicked to download your extracts, you'll be prompted to open or save the zipped folder(s). Inside each zipped folder, again, will be multiple files. In our example, the zipped data folder called *nhgis0001_csv.zip* will contain two files. The first is named *nhgis0001_ds171_2010_sab_12.csv* and is the comma delimited file that can be brought into a GIS or opened in a spreadsheet software. The second file is named *nhgis0001_ds171_2010_sab_12_codebook.txt* and is the codebook that includes the key to understanding the field names along with other information about the downloaded data. In both instances, your naming structure may vary slightly as the *0001* that follows *nhgis* is the sequential extract number. In our case, it is the very first extract we have ever done.

The zipped folder with the GIS data is typically much larger and contains zipped folders for each shapefile group, with each of those folders containing several files. In our example, the internal zipped folder is called *nhgis0001_shapefile_us_sabins_1920_2010_Twelfth.zip* and in it is 14 separate files. These 14 files, however, are just two shapefiles that can be opened in a GIS. One is for the state of Alaska and the other is the contiguous United States. Take care when moving or copying these files to keep the seven files for each shapefile together. If the seven files for each shapefile are separated from each other, the shapefile will not open in a GIS. Ideally, file management of shapefiles should be done within Esri ArcCatalog or a similar open source software.

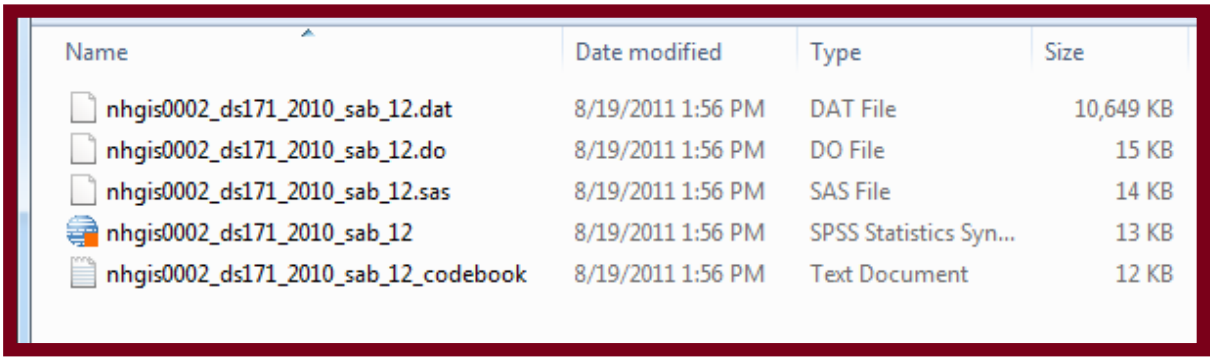


Final Steps

Remember that each geographic (grade) level will have this same number of files. Thus, selecting all 13 grade levels will return 13 .csv files, 13 .txt codebook files, and 13 zipped shapefiles folders that each contain the 14 files that make up the contiguous US and Alaska shapefiles.

Fixed width output example

As mentioned prior, if Fixed width instead of Comma delimited was selected as the data file structure, the output would contain a number of different files instead of the one .csv and .txt file for each grade level. By revising the extract request, we can go back and select the fixed width option and resubmit the extract. Since this is now our second extract, our zipped folder with the data is called *nhgis0002_fixed.zip* and contains five files. Three command files (.sas, .do, .sps), the .txt codebook, and the .dat fixed width data file are included. Again, these files will be repeated for each grade level, creating 65 total files if every grade level is selected.



Name	Date modified	Type	Size
nhgis0002_ds171_2010_sab_12.dat	8/19/2011 1:56 PM	DAT File	10,649 KB
nhgis0002_ds171_2010_sab_12.do	8/19/2011 1:56 PM	DO File	15 KB
nhgis0002_ds171_2010_sab_12.sas	8/19/2011 1:56 PM	SAS File	14 KB
nhgis0002_ds171_2010_sab_12	8/19/2011 1:56 PM	SPSS Statistics Syn...	13 KB
nhgis0002_ds171_2010_sab_12_codebook	8/19/2011 1:56 PM	Text Document	12 KB

Final Notes

Too many columns for GIS - In addition, it is important to understand that the number of fields in the .csv can be very large when multiple tables are selected. Older versions of Excel cannot handle files with more than 255 columns in a worksheet and will simply cut off the remaining columns. Even more problematic is that ArcGIS (pre 10.1) also limits tables to 255 columns (fields), except in geodatabase tables, and will also simply exclude all other columns beyond that range. Another unfortunate quirk when importing .csv files directly into ArcGIS is that when the .csv contains exceeds 2048 rows (shown as *2000 in the attribute table), it will not include the extra rows UNLESS in the attribute table you immediately click the arrow to go to the last record. Then it will process and include all of the records. Exceeding 2048 rows is commonplace with SABINS data due to its nationwide extent and thousands of records.

Join data in GIS - It is very important to realize that even when downloading data tables and GIS files in the same extract that the data tables are not attached automatically to the GIS files. This 'join' must be done by you, inside of the GIS. NHGIS has made it very simple, however, to successfully complete a 'join' operation. In both the data tables and the attribute table within each GIS shapefile, you will find a field called *GISJOIN* that is the key to linking the separate files together. For detailed information on completing the 'join' operation, please review the *Using SABINS with GIS* user's guide.