

How to Import More Than 256 Columns of Data into a GIS

September 2011

This tutorial shows users how to add comma delimited .csv files into Esri ArcGIS that contain more than 256 columns of data. Under normal operations, ArcGIS is unable to import more than 256 columns of a single .csv file due to a Microsoft software limitation. Additional fields are stripped away and the 256 column .csv file is opened without notification that fields were removed.

While most users who typically work with smaller tables and datasets will never encounter this issue, it is common to perform an NHGIS data extract of one or more tables that results in a .csv file exceeding 256 columns. For example, one particular table from the 2000 Census has more than 1400 columns of data. If a user attempts to bring the .csv file with this one table into ArcGIS, the vast majority of the data are stripped away.

Users do have options, however, to make sure all of their data can be utilized by ArcGIS. Two common workarounds are detailed in this tutorial document. First, the Quick Import tool available through the Data Interoperability extension is used. Second, the large .csv file is broken apart into smaller, less than 256 column .csv files using Microsoft Excel and are then rejoined and exported in an alternative format within ArcGIS.

These instructions are based on Esri's ArcGIS 10 software package. While the steps are the same, those using a version of ArcGIS 9 may notice their screen looks different than the images shown here. In addition, other GIS packages (MapInfo, Intergraph's GeoMedia, GRASS, etc.) may yield different results when adding large .csv files to map documents.

This tutorial assumes that you have already downloaded tabular data with more than 256 columns and in the .csv format from NHGIS onto your computer and have unzipped all files.

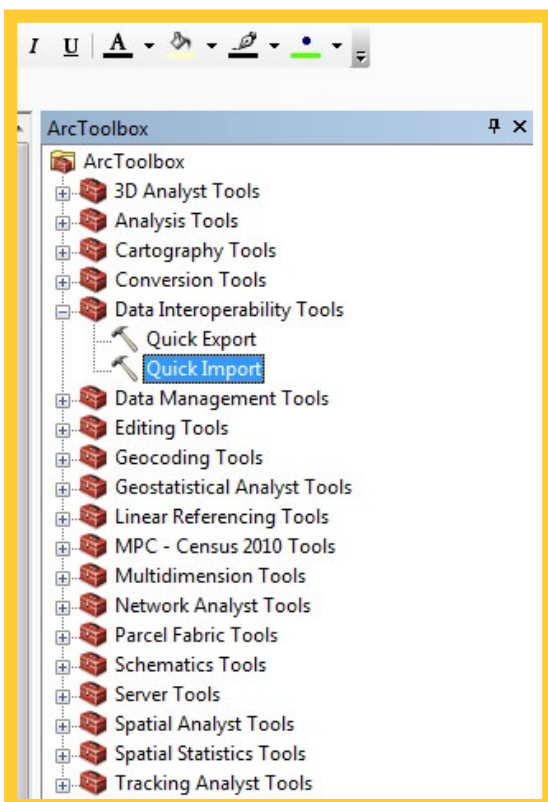
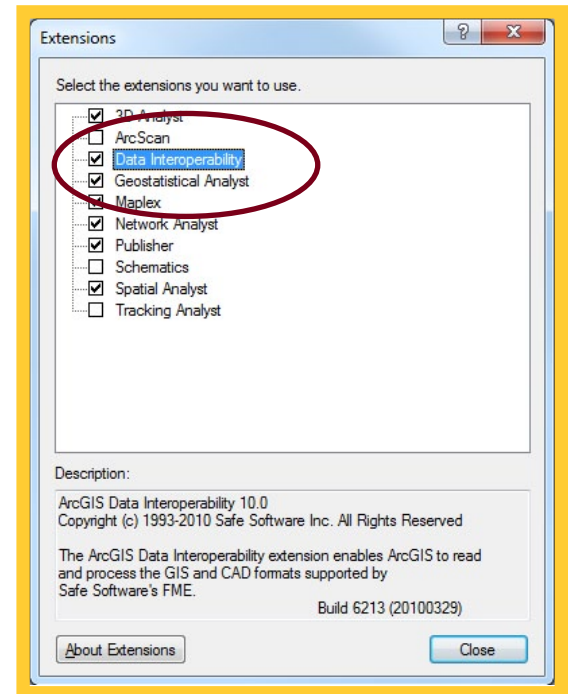
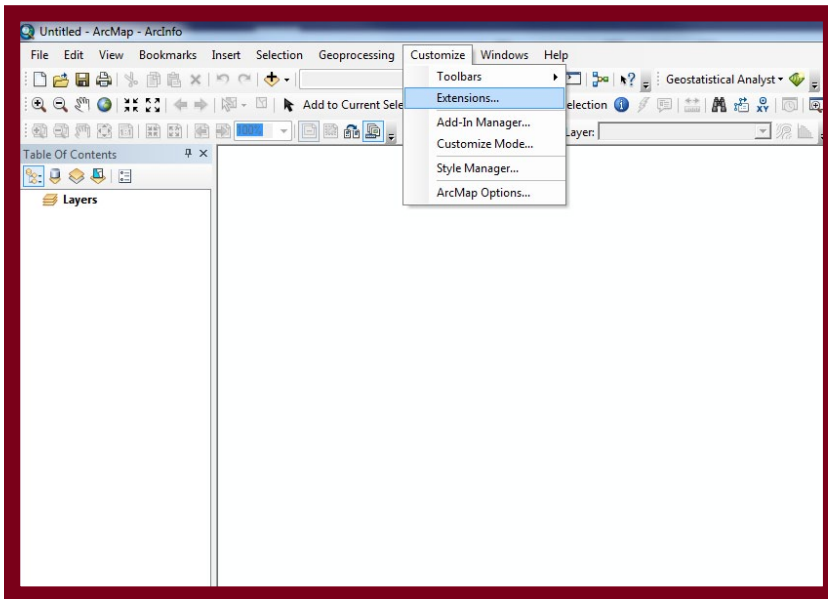


Funding provided by the National Science Foundation and National Institutes of Health.
Project support provided by the Minnesota Population Center.

Method 1: Quick Import using Data Interoperability

The premise behind Data Interoperability is that it allows data to be more easily transferred between ESRI and non-ESRI native formats. The comma delimited .csv file downloaded from NHGIS will be converted to a file geodatabase table that is not confined to the 256 column limit.

That Data Interoperability extension does not come standard with ArcGIS, but many NHGIS users have access to it through their company, university, or student trial license. If available, turn on the extension by selecting Customize>Extensions and then clicking the check box next to Data Interoperability on.

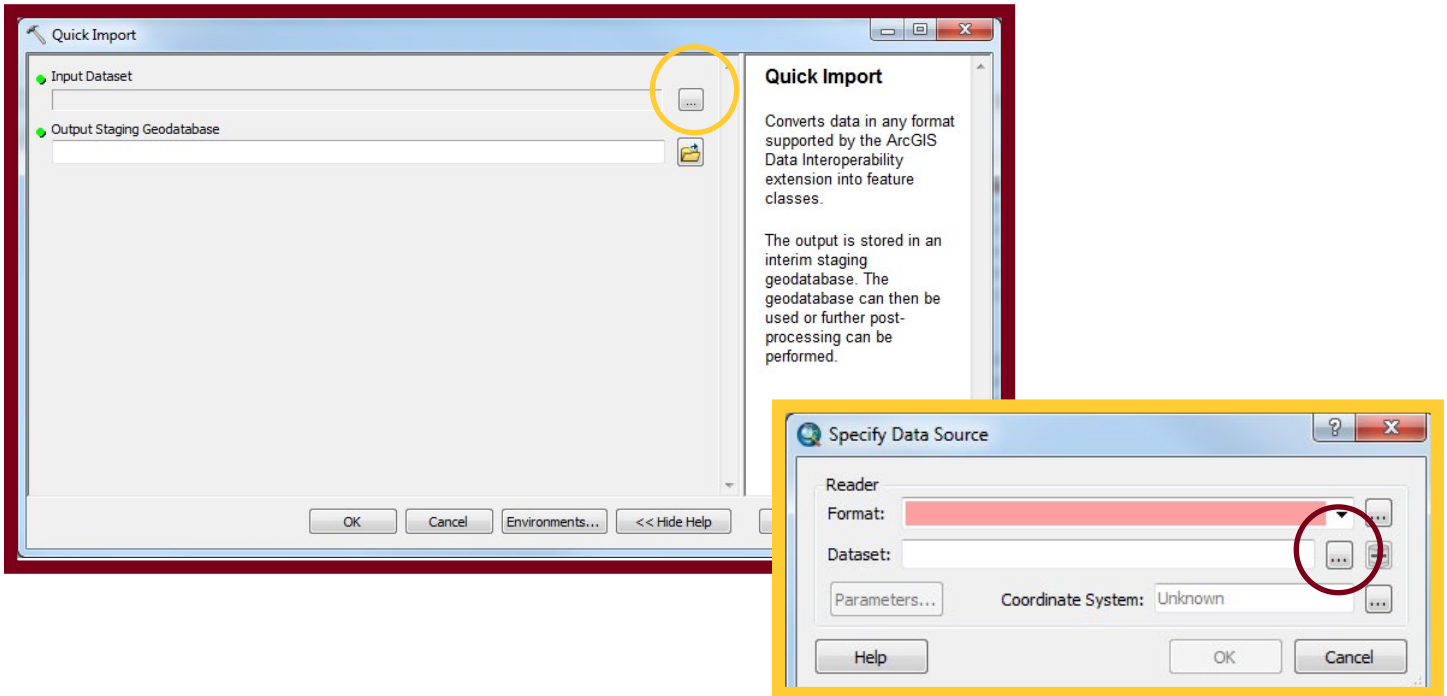


Once the Data Interoperability extension is turned on, you will be able to utilize the Quick Import tool within ArcToolbox. It is found within the Data Interoperability Tools.

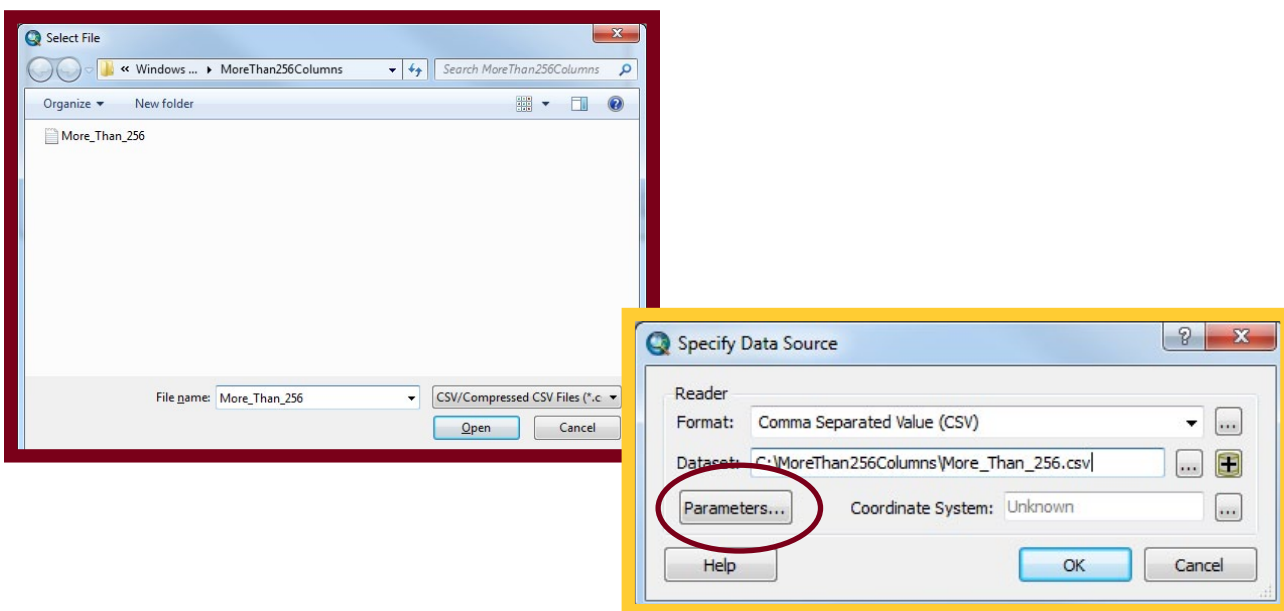
In your blank map document, click to open the Quick Import tool.

Method 1: Quick Import using Data Interoperability

With the Quick Import window open, click the Input Dataset browse button. This opens the Specify Data Source dialogue box, where you will select the .csv file you wish to import.

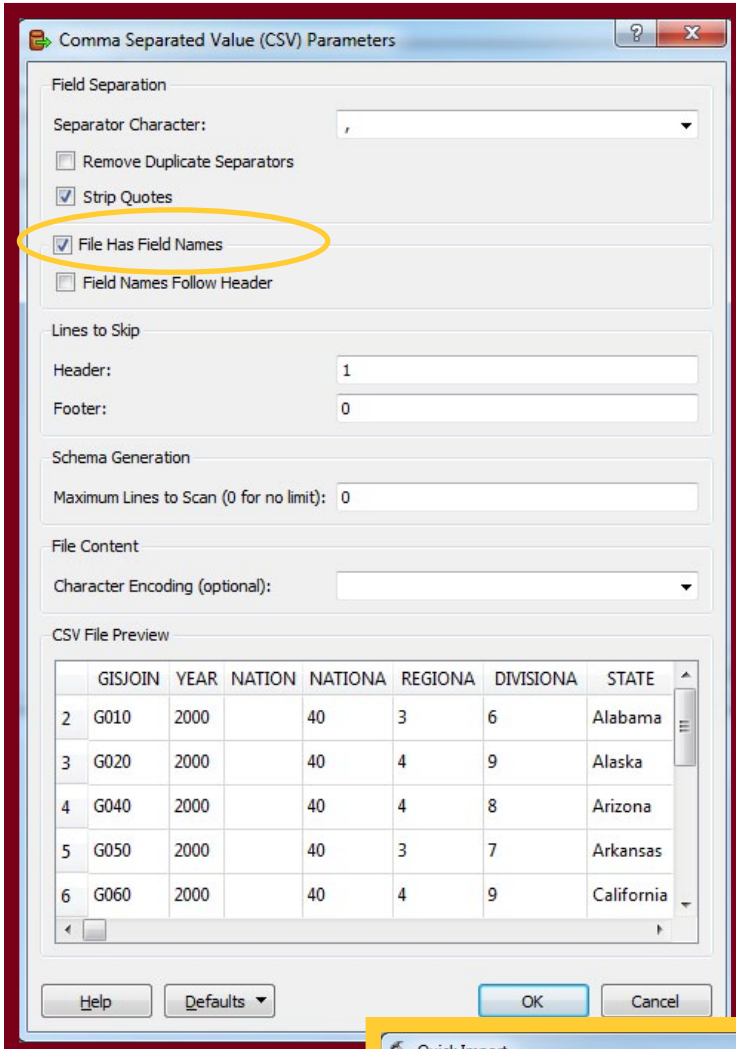


Using the Dataset browse button, navigate to the .csv file you wish to import. In this case, it is a file called "More_Than_256.csv." Once you select the appropriate file, click the Parameters button.



Method 1: Quick Import using Data Interoperability

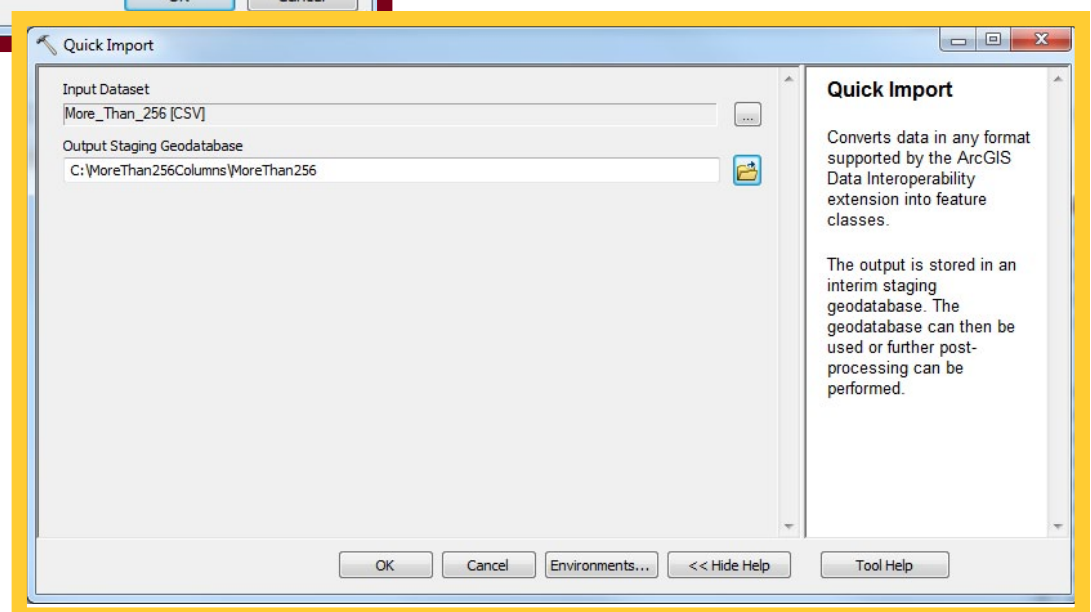
Inside the Parameters dialogue box, a number of options are available to the user. Most important for our purposes is the "File Has Field Names" box. Check this box on, and click OK, and OK again on the Specify Data Source dialogue box.



Finally, select a location and name for the Output Staging Geodatabase. In this example, I have simply called the geodatabase "MoreThan256." and placed it in the same folder as the existing .csv file. The file extension will automatically be added.

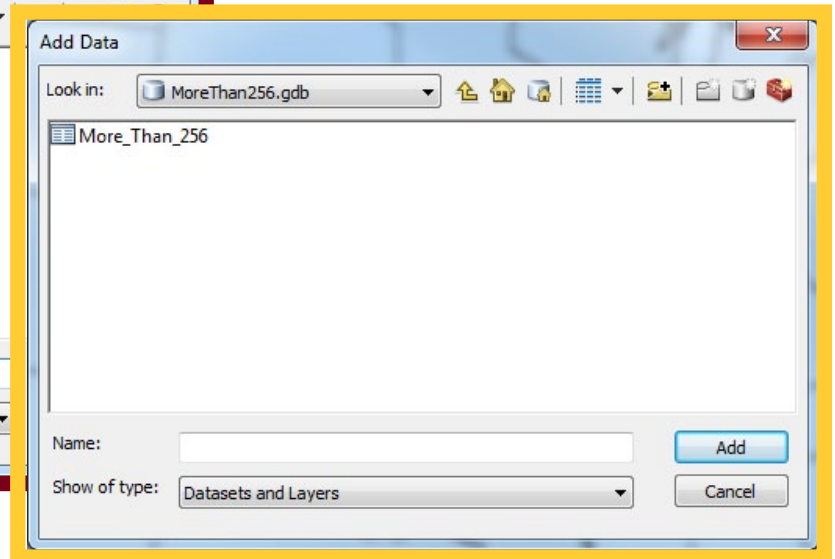
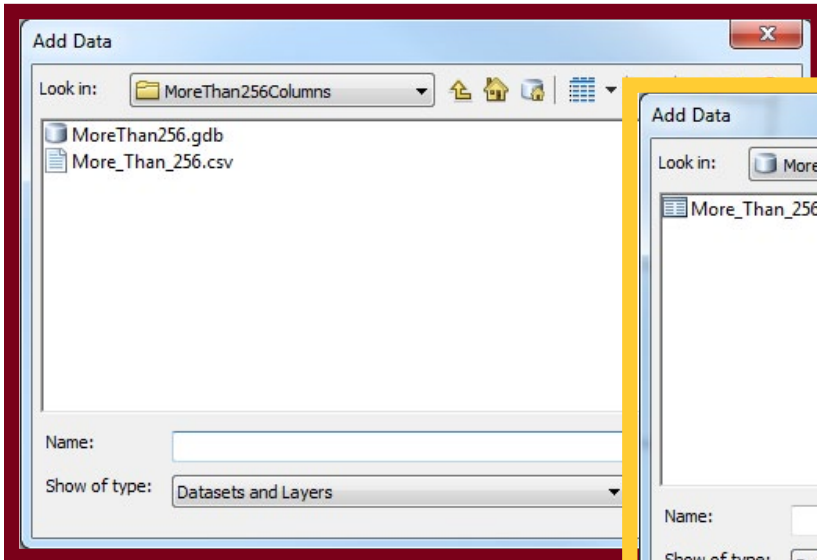
This geodatabase is not meant to be the table's permanent home necessarily, but rather, as a temporary location where you can decide what to do with it.

In fact, if you select an existing geodatabase as the Output Staging Geodatabase, the tool may either fail to run successfully or the existing geodatabase content will be removed and replaced with this imported table.



Method 1: Quick Import using Data Interoperability

Once the Quick Import tool successfully runs, you can add your table to ArcGIS with all columns intact! This example shows the last column as being "FUM379", whereas the .csv file directly added to ArcGIS maxed out at the 256th column call "FUM218", indicating that 161 additional columns of data are now included in the table that previously were not.



The Original .CSV File

FUM213	FUM214	FUM215	FUM216	FUM217	FUM218
10268	10684	10637	10960	11188	10351
210	206	217	219	230	226
1519	1567	1546	1545	1576	1496
3890	4011	4057	4233	4147	4103
20450	20927	21577	22387	22626	21116
1464	1474	1577	157	1581	1484
2883	2982	3074	3173	3181	2971
1334	1398	1513	1514	1457	1394
2706	2800	2855	2813	2753	2369
21949	23214	23371	23832	24090	23119
21568	22209	22370	22830	23597	21828
192	209	190	182	184	148
54	50	65	56	62	48
18661	19275	18611	18141	18931	17657
4794	5004	5068	4126	5027	4678
664	676	707	717	663	624
1466	1522	1480	1541	1599	1512
2548	2719	2710	2749	2734	2645
13921	14371	14377	14517	14121	13878
73	77	77	79	90	101
12751	13410	13674	13827	13987	12988
3203	3300	3382	3531	3430	3153
13876	14607	15064	15661	15725	13154
1911	2075	2016	2075	2024	1860
9832	10369	10094	10348	10381	9820
6247	6527	6447	6508	6527	6114
38	20	31	37	24	25
661	710	728	681	772	717
1246	1356	1326	1380	1333	1303

The Geodatabase Table After Quick Import

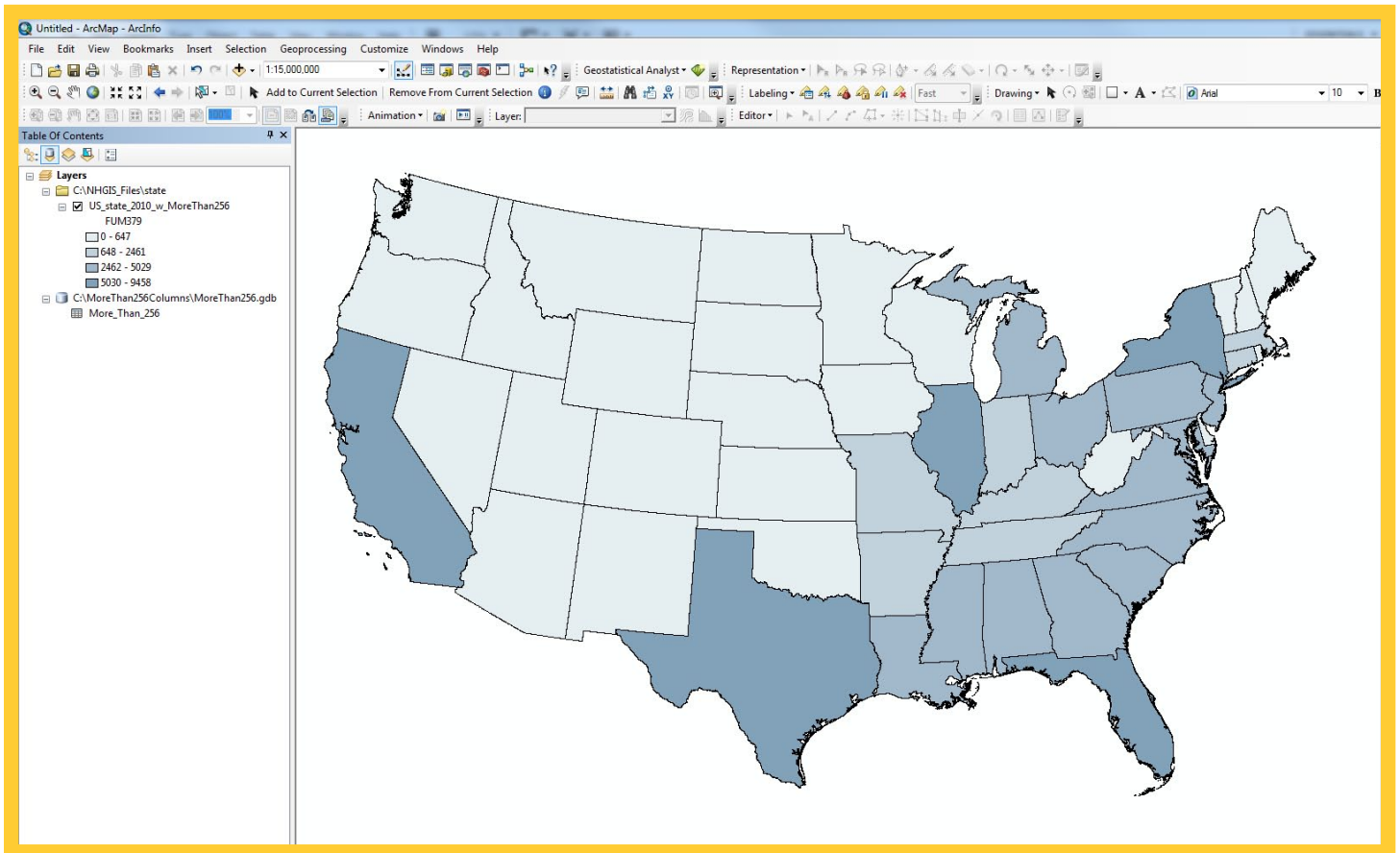
FUM374	FUM375	FUM376	FUM377	FUM378	FUM379
4010	3883	3652	3778	3527	3510
37	33	32	34	19	17
420	369	378	369	304	325
1233	1245	1062	1252	1065	1119
7326	7398	6622	6498	5663	5831
450	419	397	409	372	347
1037	967	875	897	795	756
473	489	409	388	405	411
1525	1640	1517	1623	1435	1441
7232	6993	6517	6612	5954	5821
6364	6292	5753	5827	5154	5029
10	26	14	17	10	18
4	4	5	2	4	4
6793	6619	6064	6067	5319	5513
1640	1684	1560	1619	1472	1358
121	164	134	112	119	118
442	453	423	430	385	406
916	909	854	859	846	824
4595	4508	4129	4289	3861	3784
10	5	9	8	11	4
4783	4704	4259	4291	3879	3751
1049	968	888	925	793	781
4482	4506	4212	4326	3971	4020
244	244	219	213	219	197
3264	3043	2790	3018	2605	2717
2088	2186	1972	1982	1783	1820
6	6	2	7	4	0
205	193	149	170	156	127
439	444	362	362	28	349

Notice how the number of columns has changed between the .csv file and the geodatabase table in ArcGIS.

Method 1: Quick Import using Data Interoperability

Given that the geodatabase created by Quick Import is intended as a staging geodatabase, you may benefit from either copying the table to your project's geodatabase (if you have one) or join the table to a corresponding shapefile downloaded from NHGIS using the "GISJOIN" field.

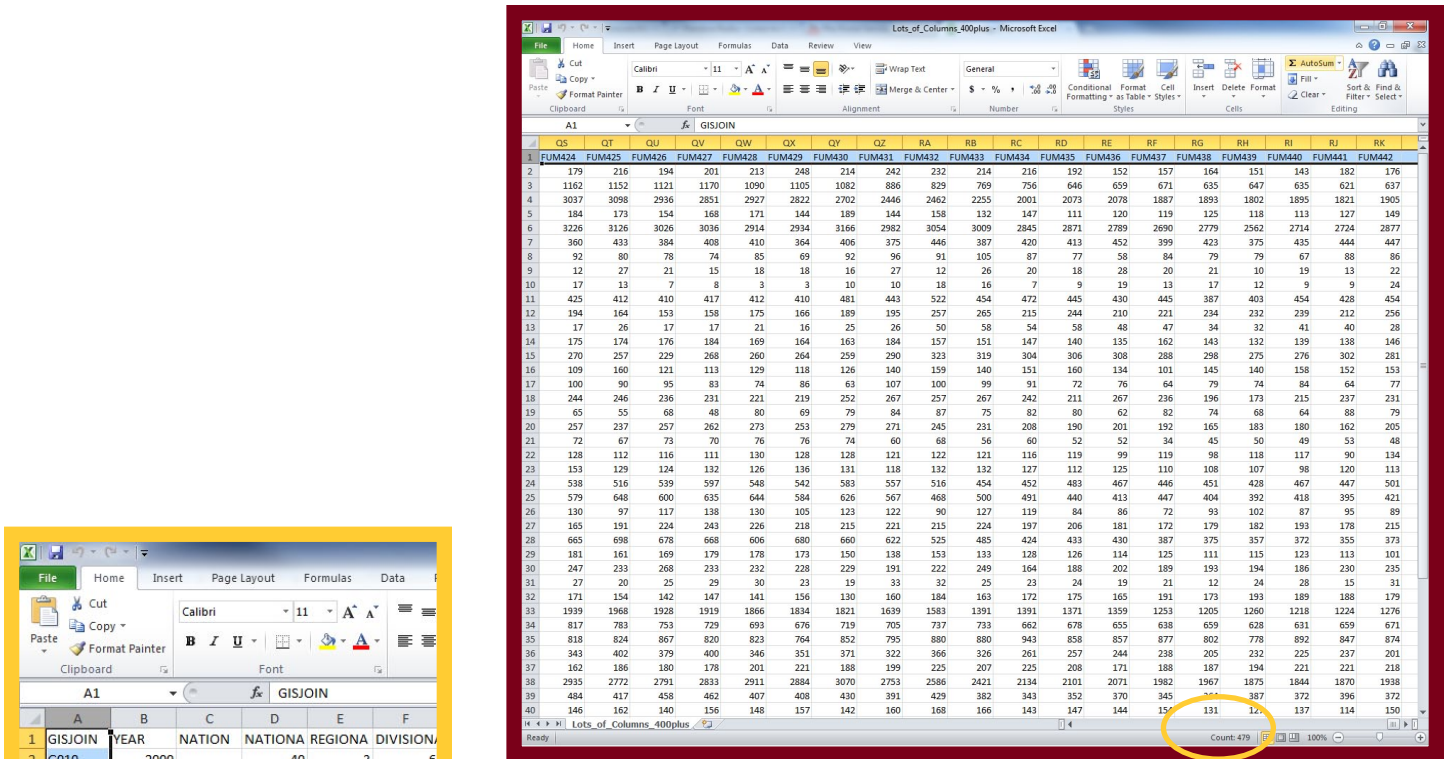
Additional information on creating joins can be found in the *How to Join NHGIS Data and Boundary Files* User's Guide found on the NHGIS website.



The new geodatabase table permanently joined and exported to a US States shapefile.

Method 2: Break Large .CSV Into Several Small .CSV

Another option for getting more than 256 columns of data from a .csv file into ArcGIS is to manually break the .csv apart using Excel or other software, and then join the small .csv files, one at a time, to a shapefile or feature class inside ArcMap or ArcCatalog. Note that this method is far from ideal, and if users have the Data Interoperability extension available, the Quick Import method is preferred.



Notice that the first column in this 479 column .csv file is the "GISJOIN" field. This field is the key used to link the NHGIS shapefiles to the .csv data. As you cut and paste columns of the large .csv into each smaller file, you must include the "GISJOIN" field with each smaller .csv file.

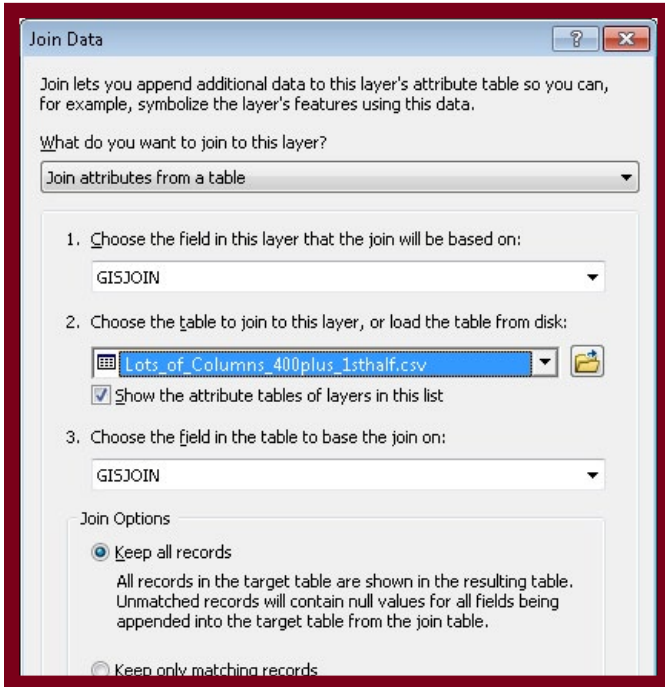
In this example, the 479 column .csv file has been broken into 2 smaller .csv each approximately having 240 columns of data. You may choose to break up your .csv by as few or as many columns (less than 256, of course!) as you wish. This .csv could have been split into five files each having less than 100 columns, if one desired.

While the end goal is to recombine these small .csv files in a different file format in ArcGIS, you should be cognizant of how the file is broken up. For example, if your .csv contains data from two different tables, you may wish to split the file at the break in tables. This way, if you chose not to recombine the .csv files in ArcGIS, your smaller files would each contain an entire table, rather than having a table split between two separate files.

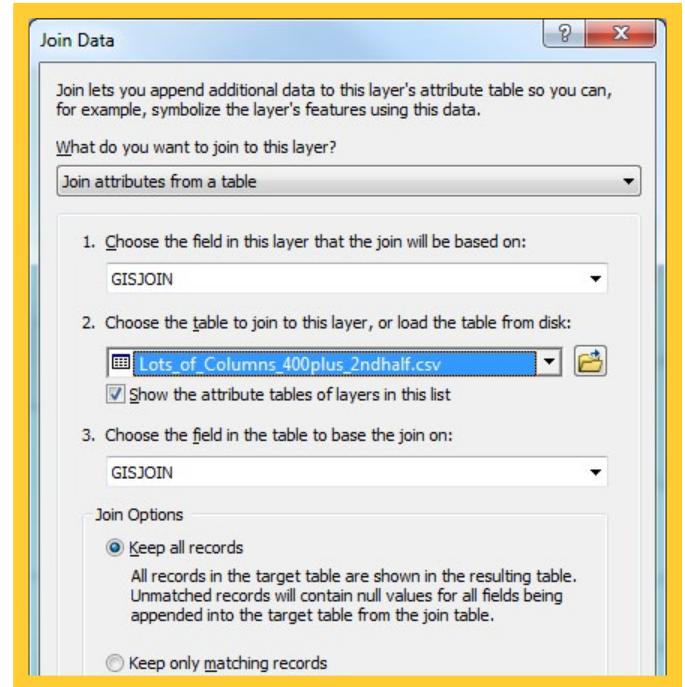
Method 2: Break Large .CSV Into Several Small .CSV

In this example, our 479 column .csv file has been broken into two separate .csv files. One is called "Lots_of_Columns_400plus_1sthalf.csv" and the other is "..._2ndhalf.csv." Each file contains the GISJOIN field that was found in the first column of the original .csv file. In ArcMap, these two files are added, alongside an NHGIS shapefile for the states in the year 2000.

Using a join operation, both .csv files are attached to the shapefile one at a time.



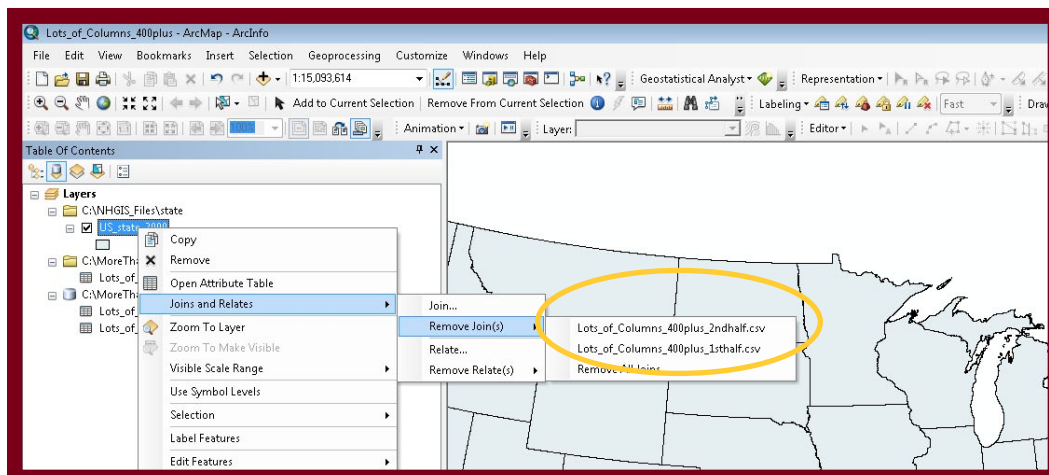
1st Join



2nd Join

Alternatively, one .csv can be joined to the other (again using GISJOIN as the joining field) if you do not want the data connected to a shapefile or if no shapefile exists for the data.

Additional information on creating joins can be found in the *How to Join NHGIS Data and Boundary Files* User's Guide found on the NHGIS website.



Two separate joins on the one shapefile

Method 2: Break Large .CSV Into Several Small .CSV

The smaller .csv files joined to the shapefile, the final step is to export the data. In this example, the shapefile with the US states is now joined with the two separate .csv files that together contain over 400 columns of data and is exported with the name "Shapefile_w_400plus.shp."

While this step is not required to utilize the data in analysis and mapping, it is highly recommended. The complexity of multiple joins with files containing hundreds of columns may cause notable computer slowdowns with ArcGIS if the data is not exported to a new shapefile or geodatabase feature class.

