

How to Join NHGIS Data Tables and GIS Boundary Files

August 2011

This tutorial shows users how to join shapefiles and aggregate data downloaded from the National Historical Geographic Information System data access system. 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 have different methods for joining data, so check your software's help documentation for more information.

This tutorial assumes that you have already downloaded a zipped shapefile and a corresponding zipped aggregate data file from NHGIS to your computer and have unzipped both files. Be aware that the content of the now unzipped shapefile folder is additional files that must also be unzipped before being used.

Please note, this tutorial also assumes that you selected the download option of "Comma delimited (best for GIS)" and did NOT select the option to "Include descriptive header row (best for spreadsheets)".

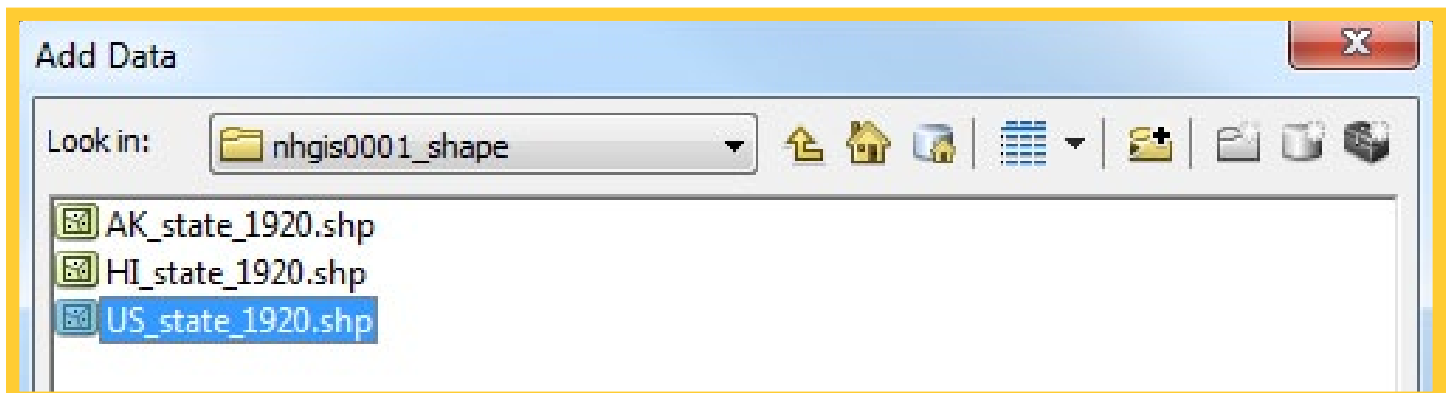
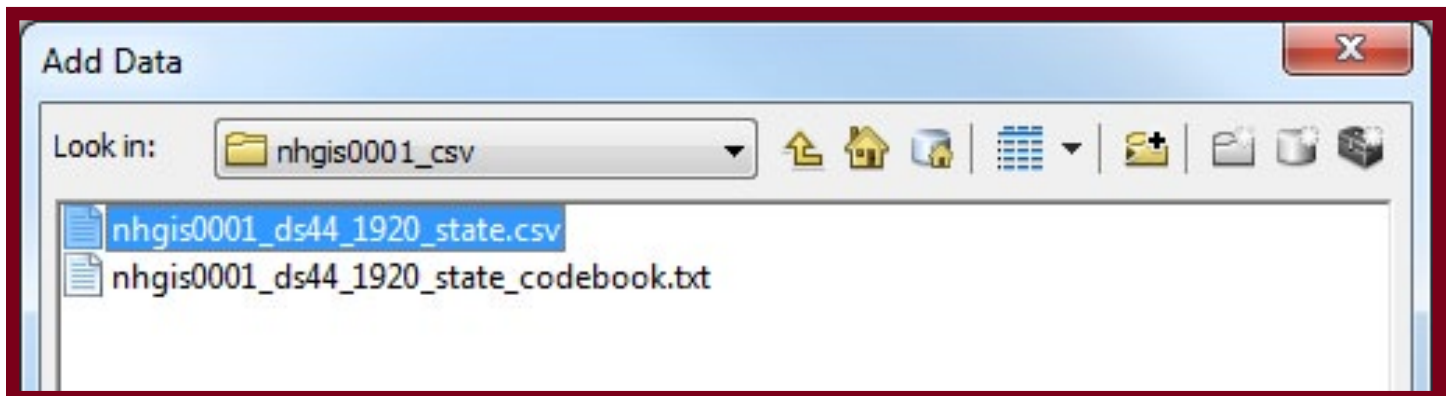
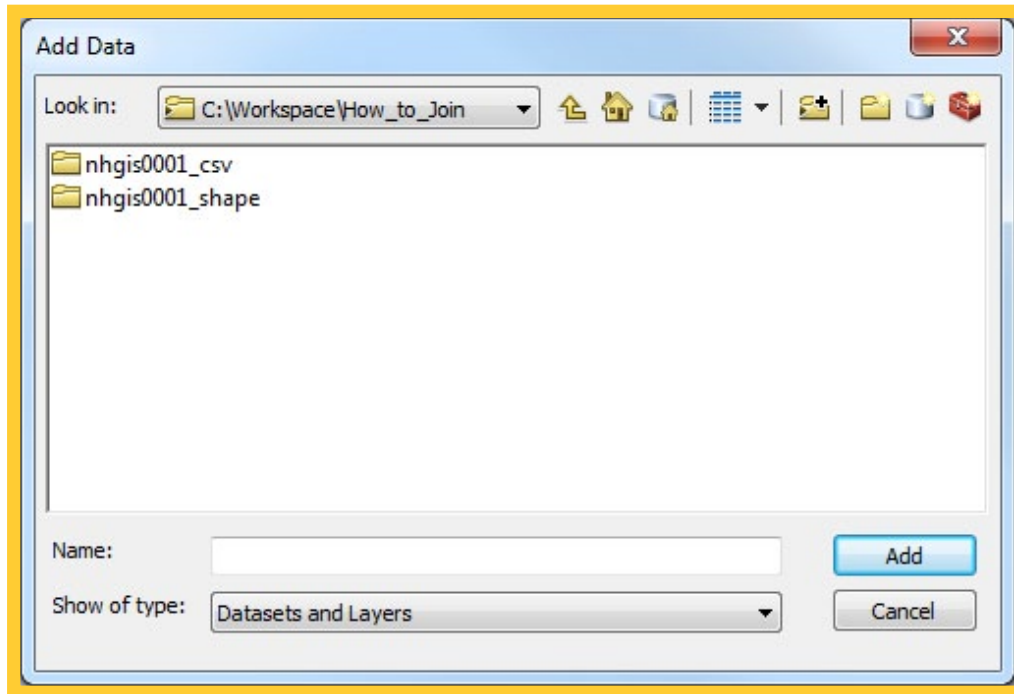
This tutorial will use the data extract completed as Scenario 1 of the User's Guide. 1920 Census data on the number of farms per state needs to be ***joined*** to a shapefile that contains every state in the United States for the year 1920.



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

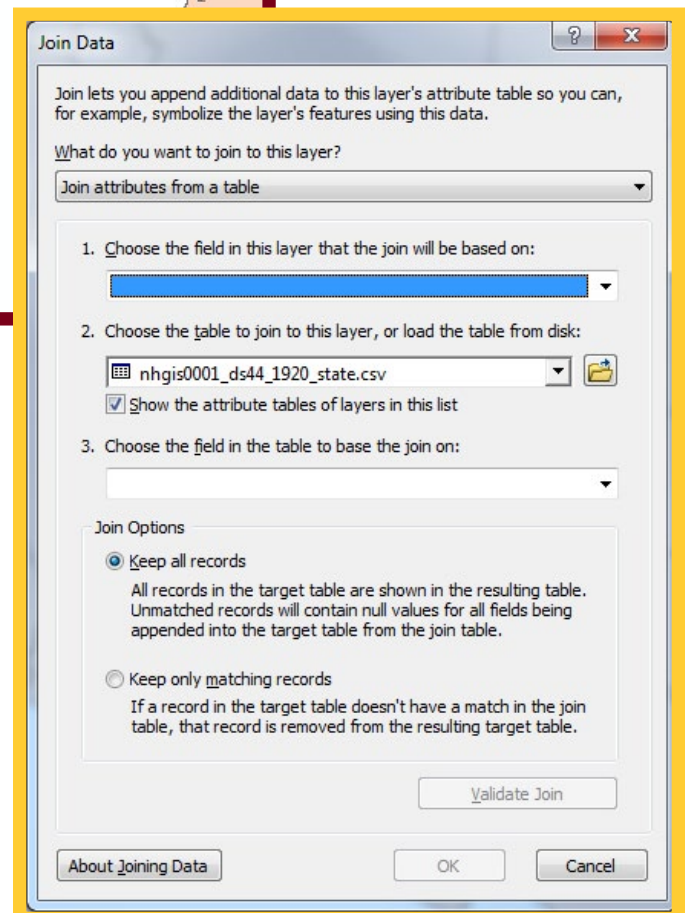
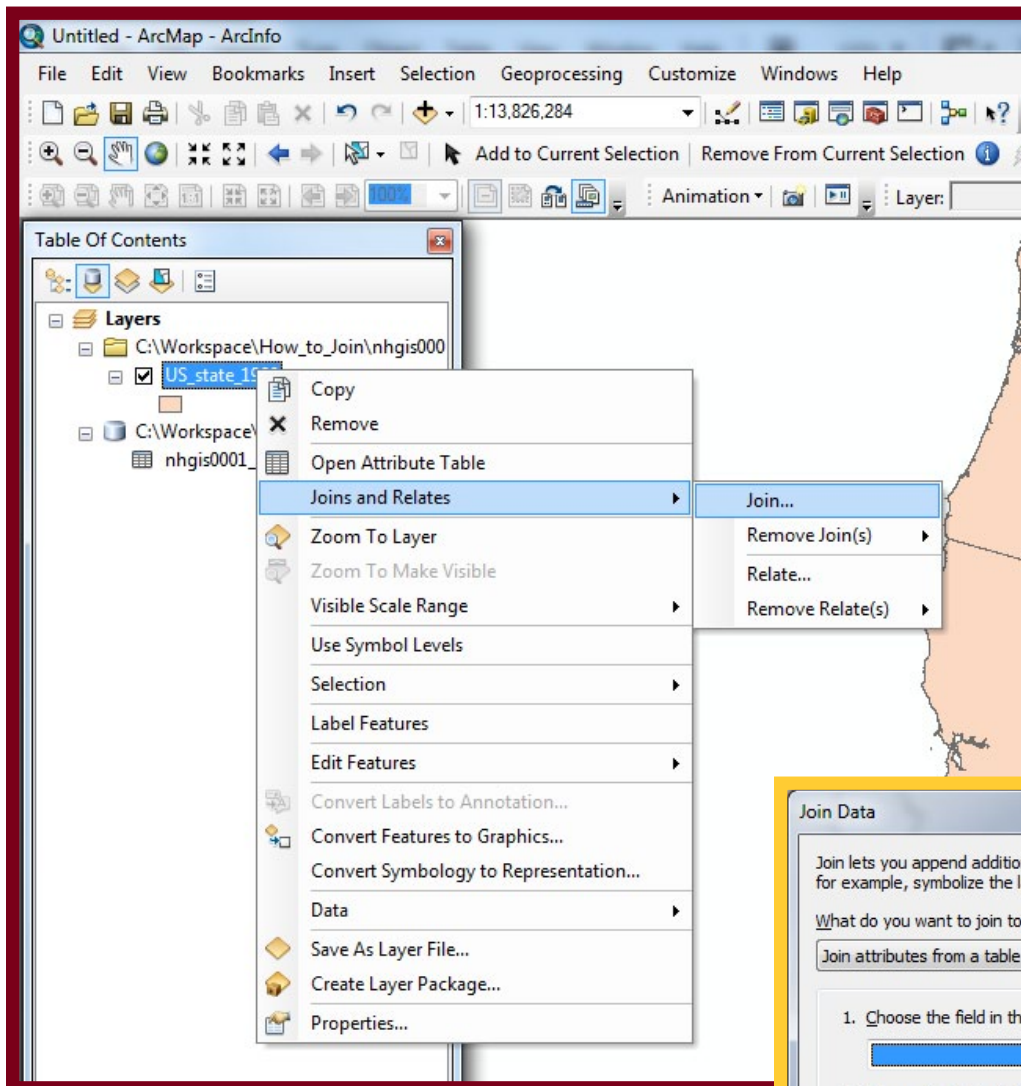
Step 1: Add data to your map document

In your map document, click the Add Data button and navigate to the folder where you saved your NHGIS extract. Click on the ..._csv folder and add the .csv file (not the .txt codebook file) to your map. Next, click the Add Data button and navigate to the ..._shape folder and add the desired shapefile. In our example, we want to use the US_state_1920.shp which is the lower 48 states. Note that the states (or territories as they were in 1920) of Alaska and Hawaii are included as separate shapefiles.



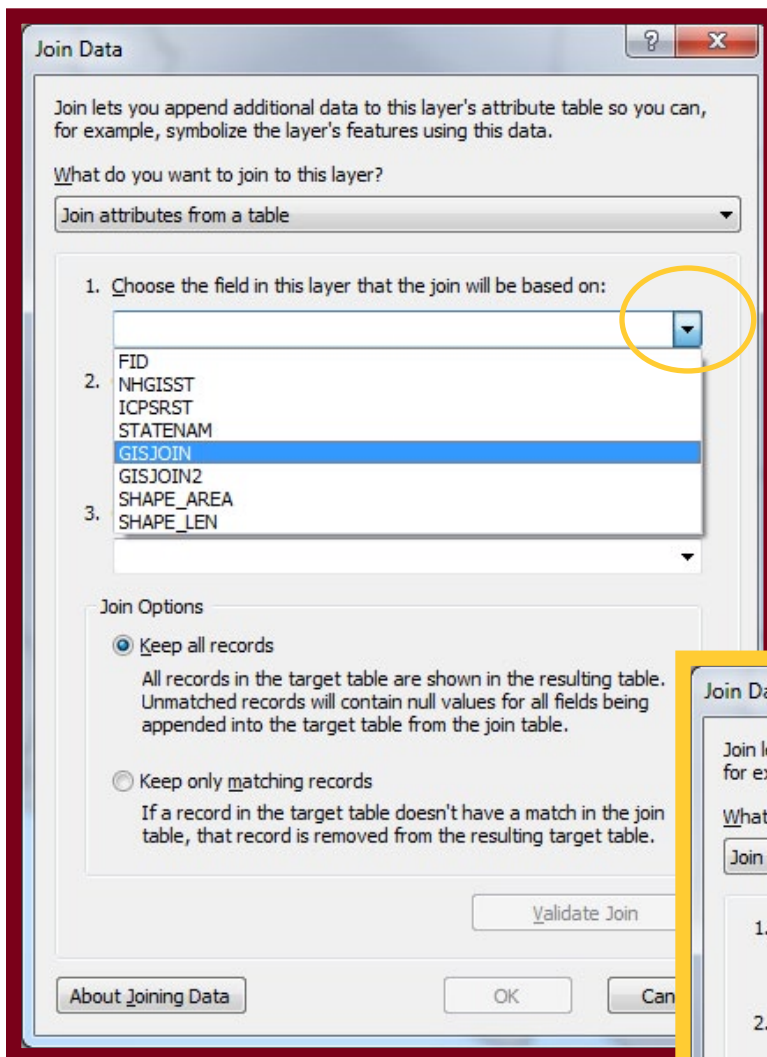
Step 2: Open the Join dialog box

Right-click on the shapefile in the Table of Contents and then select **Joins and Relates** from the context menu. Finally, click on "Join..." to open the Join dialog box.



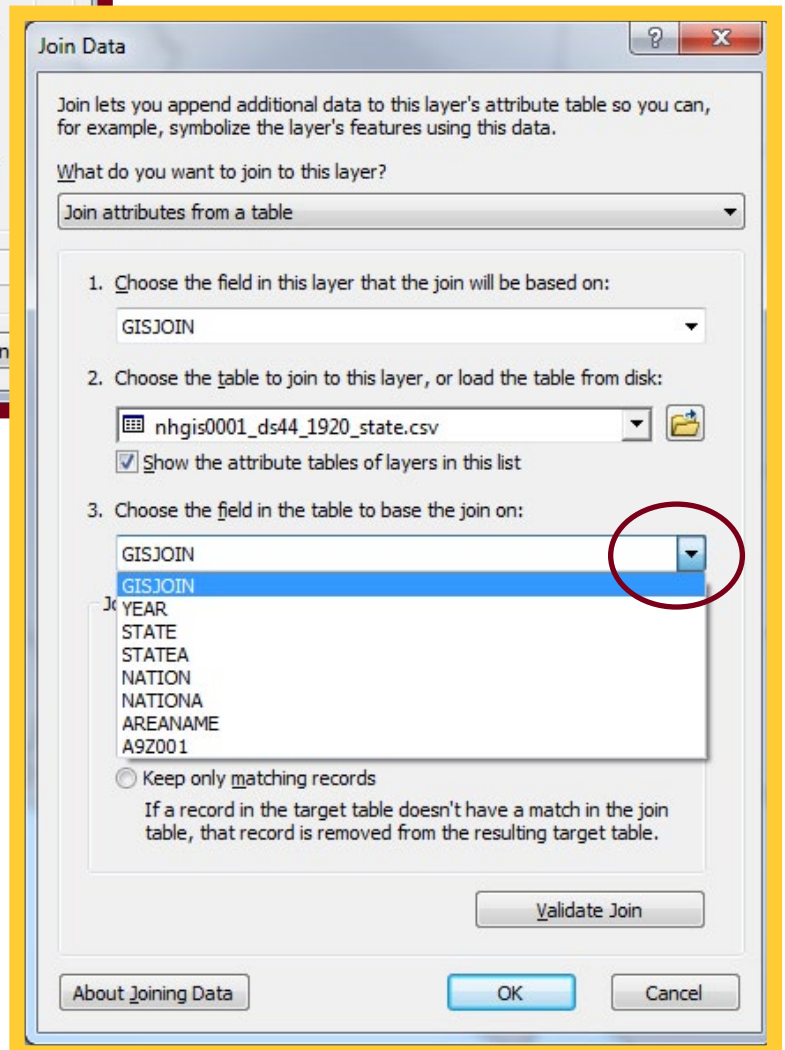
Step 3: Choose your Join options

The Join dialog box prompts you to choose the fields in the shapefile and aggregate data file to base the join on.



First, choose the attribute in the shapefile. All NHGIS shapefiles contain an attribute called GISJOIN, which is the unique identifier for every geographic feature in the file. Click on the down arrow (circled in gold) and select GISJOIN from the list of attributes.

Second, choose the table you want joined to your shapefile. In our example, it defaults to the correct .csv file because it is the only other file in the map. With other tables present, it may not default to your desired table, so always make sure the correct table is chosen.

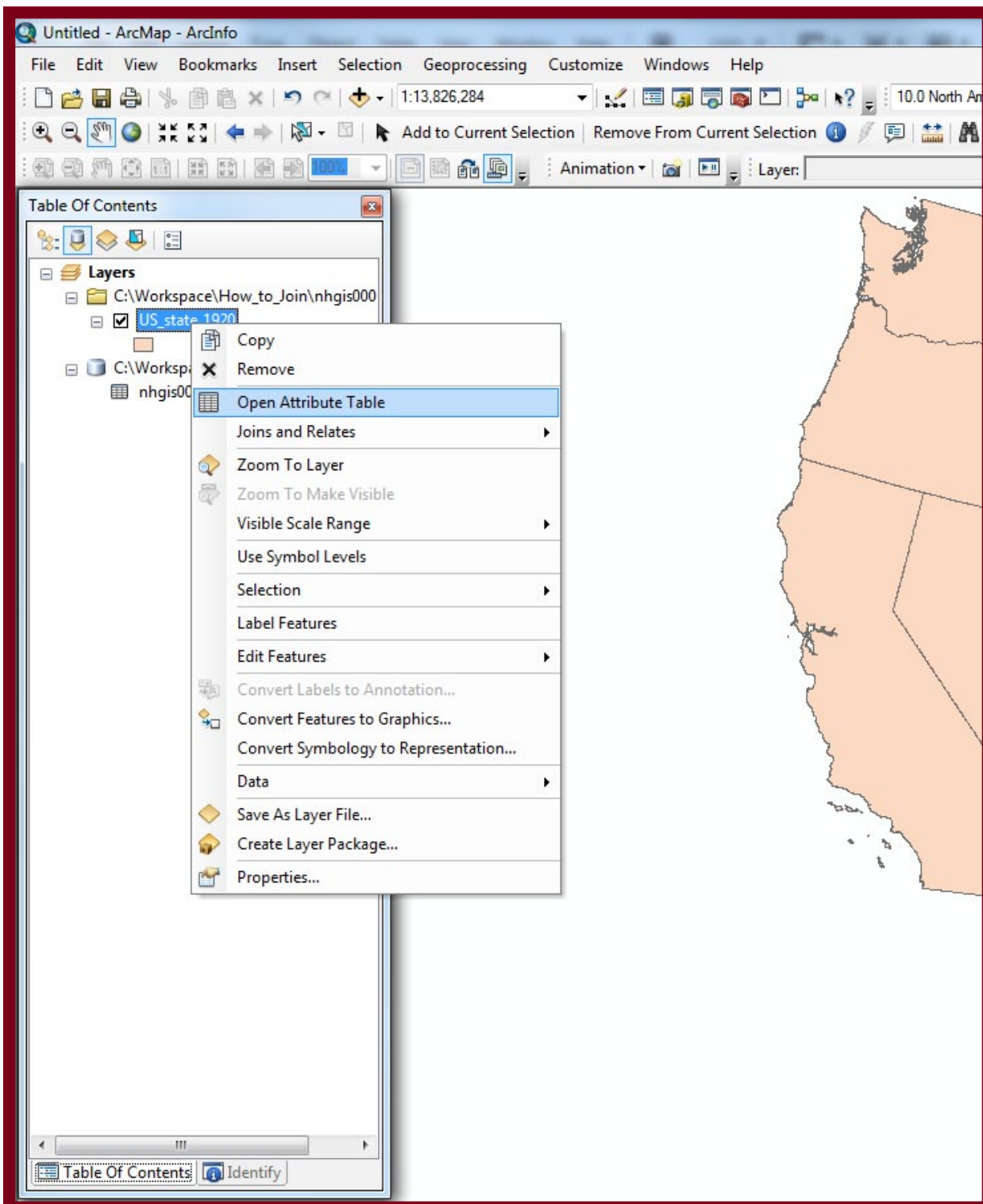


Third, choose the attribute in the aggregate data file. All NHGIS aggregate data files contain an attribute called GISJOIN, which is a unique identifier for every geographic feature in the file. Click on the down arrow (circled in maroon) and select GISJOIN from the list of attributes.

Step 4: Verify the Join

To verify the join, open the attribute table of the shapefile. If the join worked correctly, you will see all the attributes from the shapefile and the aggregate data file.

To open the attribute table of the shapefile, right-click on the shapefile and select **Open Attribute Table** from the context menu.



Step 4: Verify the Join

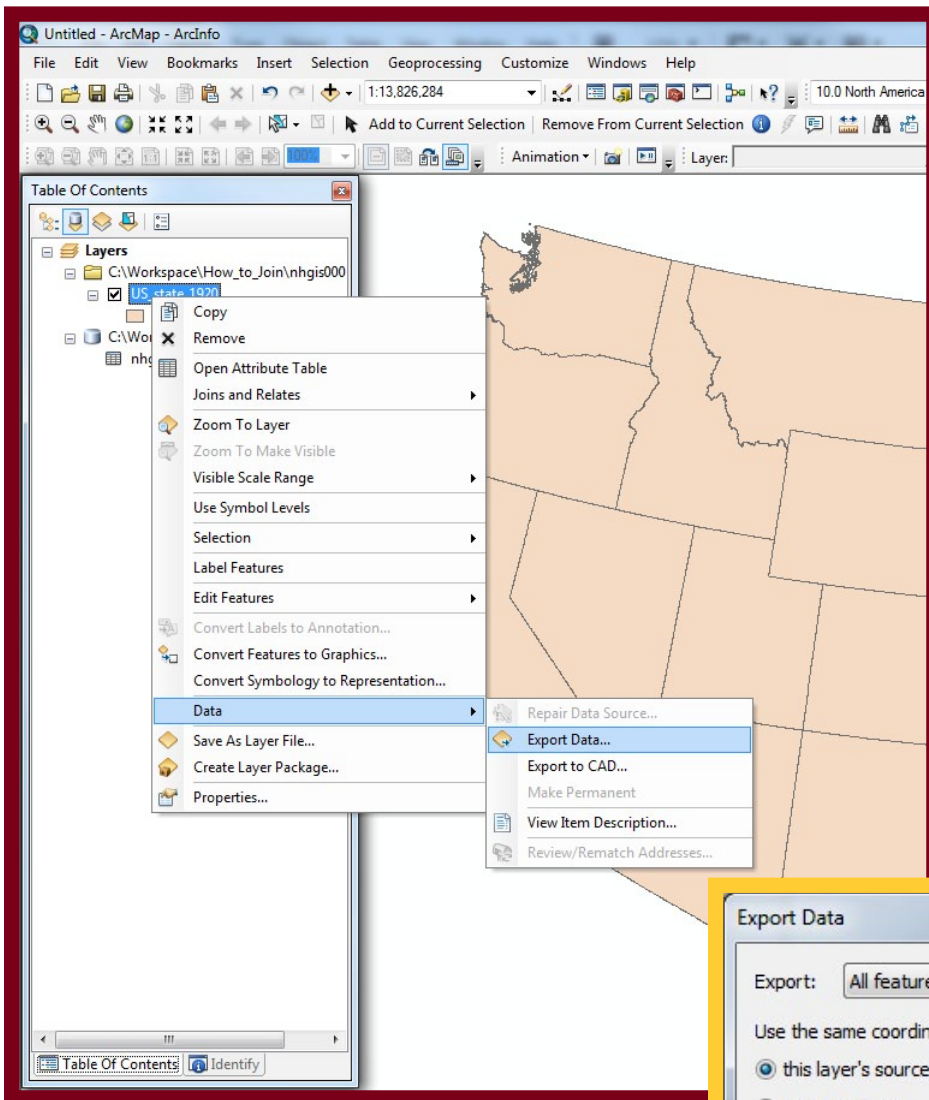
This will open the shapefile's attribute table. You will notice that the attribute table now contains the data from the aggregate data file along with the data from the shapefile. Notice that there are now two fields called 'GISJOIN'. The first GISJOIN attribute (circled in maroon) is from the shapefile and the GISJOIN attribute (circled in gold) is from the aggregate data file. Please note that the joined aggregate data traditionally follows all of the attributes originally available on the shapefile.

Notice that the last field (A9Z001) is the data from our data table, and it can now be mapped!

FID	Shape *	NHGISST	ICPSRST	STATENAM	GISJOIN	GISJOIN2	SHAPE_AREA	SHAPE_LEN	GISJOIN	YEAR	STATE	STATEA	NATION	NATIONA	AREANAME	A9Z001
0	Polygon	120	43	Florida	G120	120	146829778555	165701.06295	G120	1920	Florida	120	<Null>	<Null>	Florida	54005
1	Polygon	510	40	Virginia	G510	510	103661132151	7983527.9611	G510	1920	Virginia	510	<Null>	<Null>	Virginia	186242
2	Polygon	230	2	Maine	G230	230	84140536266.3	7529480.8460	G230	1920	Maine	230	<Null>	<Null>	Maine	48227
3	Polygon	400	53	Oklahoma	G400	400	181036866732	2678384.63438	G400	1920	Oklahoma	400	<Null>	<Null>	Oklahoma	191988
4	Polygon	200	32	Kansas	G200	200	213096312615	2003684.97128	G200	1920	Kansas	200	<Null>	<Null>	Kansas	165286
5	Polygon	050	42	Arkansas	G050	050	137733197260	2171272.49317	G050	1920	Arkansas	050	<Null>	<Null>	Arkansas	232604
6	Polygon	280	46	Mississippi	G280	280	123453258391	2908321.3185	G280	1920	Mississippi	280	<Null>	<Null>	Mississippi	272101
7	Polygon	190	31	Iowa	G190	190	145741540602	1862375.02113	G190	1920	Iowa	190	<Null>	<Null>	Iowa	213439
8	Polygon	300	64	Montana	G300	300	380340758698	3161857.41468	G300	1920	Montana	300	<Null>	<Null>	Montana	57677
9	Polygon	550	25	Wisconsin	G550	550	145301942368	2886959.73343	G550	1920	Wisconsin	550	<Null>	<Null>	Wisconsin	189295
10	Polygon	480	49	Texas	G480	480	685670896867	12470402.8007	G480	1920	Texas	480	<Null>	<Null>	Texas	436033
11	Polygon	530	73	Washington	G530	530	174889493247	6146884.73276	G530	1920	Washington	530	<Null>	<Null>	Washington	66288
12	Polygon	250	3	Massachusetts	G250	250	21013355034.3	2705616.03803	G250	1920	Massachusetts	250	<Null>	<Null>	Massachusetts	32001
13	Polygon	210	51	Kentucky	G210	210	104659205429	2171569.46932	G210	1920	Kentucky	210	<Null>	<Null>	Kentucky	270626
14	Polygon	010	41	Alabama	G010	010	133741975435	2455229.82422	G010	1920	Alabama	010	<Null>	<Null>	Alabama	256099
15	Polygon	340	12	New Jersey	G340	340	19594133178.2	2580834.35743	G340	1920	New Jersey	340	<Null>	<Null>	New Jersey	29702
16	Polygon	290	34	Missouri	G290	290	180533553620	2390682.378	G290	1920	Missouri	290	<Null>	<Null>	Missouri	263004
17	Polygon	110	98	District of Columbia	G110	110	159396851.326	113301.002223	G110	1920	District Of Columbia	110	<Null>	<Null>	District of Columbia	204
18	Polygon	350	66	New Mexico	G350	350	314917369534	2392339.47805	G350	1920	New Mexico	350	<Null>	<Null>	New Mexico	29844
19	Polygon	040	81	Arizona	G040	040	295256634850	2396297.46642	G040	1920	Arizona	040	<Null>	<Null>	Arizona	9975
20	Polygon	130	44	Georgia	G130	130	1522212842801	3821114.51332	G130	1920	Georgia	130	<Null>	<Null>	Georgia	310732
21	Polygon	260	23	Michigan	G260	260	150593768991	6672516.42275	G260	1920	Michigan	260	<Null>	<Null>	Michigan	196447
22	Polygon	330	4	New Hampshire	G330	330	24000085437.4	1086363.16689	G330	1920	New Hampshire	330	<Null>	<Null>	New Hampshire	20523
23	Polygon	380	36	North Dakota	G380	380	18310743509	2117094.24214	G380	1920	North Dakota	380	<Null>	<Null>	North Dakota	77690
24	Polygon	390	24	Ohio	G390	390	106869524366	1865383.89331	G390	1920	Ohio	390	<Null>	<Null>	Ohio	256695
25	Polygon	410	72	Oregon	G410	410	251101405295	3232270.68497	G410	1920	Oregon	410	<Null>	<Null>	Oregon	50206
26	Polygon	160	63	Idaho	G160	160	216293338847	2950547.64031	G160	1920	Idaho	160	<Null>	<Null>	Idaho	42106
27	Polygon	060	71	California	G060	060	409391746395	5906720.22595	G060	1920	California	060	<Null>	<Null>	California	117670
28	Polygon	560	68	Wyoming	G560	560	253972132576	2041178.71177	G560	1920	Wyoming	560	<Null>	<Null>	Wyoming	15748
29	Polygon	310	35	Nebraska	G310	310	200343832822	2185374.35786	G310	1920	Nebraska	310	<Null>	<Null>	Nebraska	124417
30	Polygon	270	33	Minnesota	G270	270	218532331586	3223066.05584	G270	1920	Minnesota	270	<Null>	<Null>	Minnesota	178478
31	Polygon	500	6	Vermont	G500	500	24900407113	915268.764597	G500	1920	Vermont	500	<Null>	<Null>	Vermont	29075
32	Polygon	470	54	Tennessee	G470	470	109151322774	2111488.59883	G470	1920	Tennessee	470	<Null>	<Null>	Tennessee	252774
33	Polygon	170	21	Illinois	G170	170	145917639248	2139234.707	G170	1920	Illinois	170	<Null>	<Null>	Illinois	237181
34	Polygon	460	37	South Dakota	G460	460	199728001492	2114721.40987	G460	1920	South Dakota	460	<Null>	<Null>	South Dakota	74637
35	Polygon	370	47	North Carolina	G370	370	127827660541	8727205.40644	G370	1920	North Carolina	370	<Null>	<Null>	North Carolina	269763
36	Polygon	090	1	Connecticut	G090	090	12872097915.1	1011849.6755	G090	1920	Connecticut	090	<Null>	<Null>	Connecticut	22655
37	Polygon	450	48	South Carolina	G450	450	79925512231	5077259.95363	G450	1920	South Carolina	450	<Null>	<Null>	South Carolina	192693
38	Polygon	360	13	New York	G360	360	125747667291	5322050.49978	G360	1920	New York	360	<Null>	<Null>	New York	193195
39	Polygon	440	5	Rhode Island	G440	440	2826795940.73	747925.263919	G440	1920	Rhode Island	440	<Null>	<Null>	Rhode Island	4083
40	Polygon	180	22	Indiana	G180	180	93711487232.6	1722971.29358	G180	1920	Indiana	180	<Null>	<Null>	Indiana	205126
41	Polygon	490	67	Utah	G490	490	219884448585	1975274.48767	G490	1920	Utah	490	<Null>	<Null>	Utah	25662
42	Polygon	080	62	Colorado	G080	080	269601383513	2102136.87398	G080	1920	Colorado	080	<Null>	<Null>	Colorado	59934
43	Polygon	240	52	Maryland	G240	240	25709868254.3	6413762.32206	G240	1920	Maryland	240	<Null>	<Null>	Maryland	47908
44	Polygon	320	65	Nevada	G320	320	286351578735	2372543.24352	G320	1920	Nevada	320	<Null>	<Null>	Nevada	3163
45	Polygon	540	56	West Virginia	G540	540	62754843839	2026386.70051	G540	1920	West Virginia	540	<Null>	<Null>	West Virginia	87289
46	Polygon	100	11	Delaware	G100	100	5131438942.8	663761.939092	G100	1920	Delaware	100	<Null>	<Null>	Delaware	10140
47	Polygon	420	14	Pennsylvania	G420	420	117287756886	1725806.33026	G420	1920	Pennsylvania	420	<Null>	<Null>	Pennsylvania	202250
48	Polygon	220	45	Louisiana	G220	220	1193854440501	12188602.2681	G220	1920	Louisiana	220	<Null>	<Null>	Louisiana	135463

Recommended Step 5: Make the Join Permanent

A join is not a permanent creation that will stay on the shapefile if opened in another map document. Instead, it is considered temporary because it only exists within the confines of the map document in which it was created. New users to GIS are encouraged to make the join permanent by exporting the shapefile, while the join is in place, as a new shapefile. A permanent join allows you to easily open the tabular data in different map documents with ease, preserves your original shapefile as it was, prevents repeating steps later on, and reduces processing time when doing using tools from the ArcToolbox.



You may find it not necessary to make the join permanent, however, for a number of reasons as well. First and foremost, it is an extra step that takes extra time that you may not consider worthwhile. You may also find it simpler to organize your data if you are not creating new shapefiles that have the same features but different data. In addition, with a small dataset you may not be as affected by reduced processing time when running analyses with the join intact.

If you decide that you do want to make the join permanent, you must *'export'* the shapefile while it is joined to the .csv file. To do this, right-click on the shapefile in the Table of Contents and select **Data** from the context menu and select "Export Data..." Save the new shapefile where ever you wish, click OK, and a new shapefile will be created. You can then add it to your map.

