

## How to Join NHGIS Shapefiles with NHGIS Aggregate Data

David Van Riper

3/23/2007

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 software package. Other GIS packages (MapInfo, Intergraph's GeoMedia, etc.) may have different methods for joining data, so check the help documentation for more information.

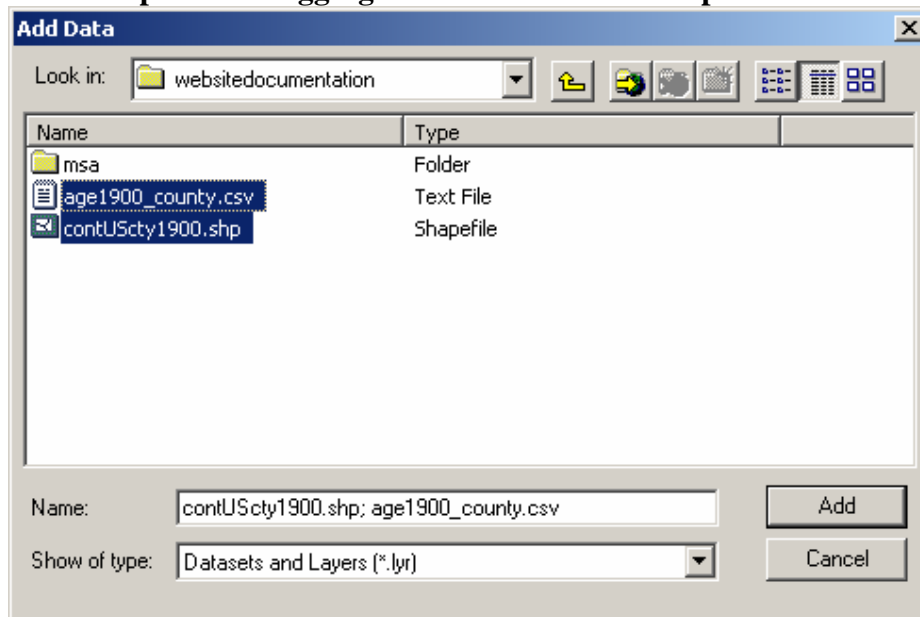
This tutorial assumes that you have downloaded a zipped shapefile and corresponding aggregate data file to your computer and unzipped the shapefile. (NOTE: When selecting your aggregate data format, make sure to download the *GIS Comma-Separated Values* file. The *Comma-Separated Values* will not work in a GIS.)

Before you may use the aggregate data file in ArcGIS, you must modify the file name so ArcGIS can read it properly. NHGIS aggregate data filenames contain characters that ArcGIS considers illegal; therefore, those illegal characters must be removed before using the data file in ArcGIS. A data file from NHGIS may have the following name:

vanriper@geog.umn.edu-102-gis.csv

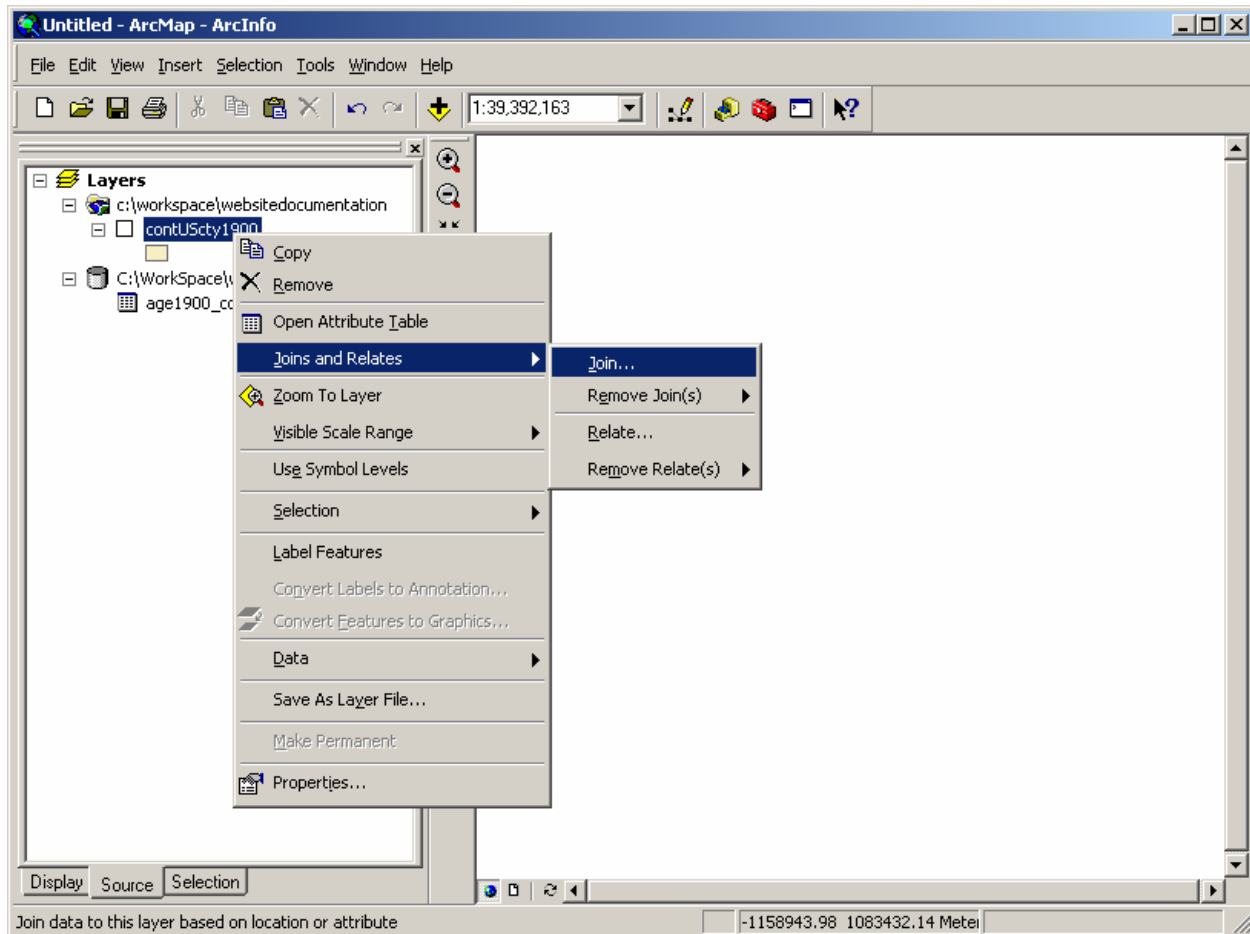
To use this file in ArcGIS, you must remove the “@”, “.”, and “-“ from the name. For example, you may rename the file **vanriper@geog.umn.edu-102-gis.csv** to **age1900\_county.csv**. After this initial work and selecting to create “a new emptying map” option in ArcMap, you can follow steps 1-3.

### Step 1 – Load the shapefile and aggregate data file into ArcMap.



## Step 2 – Open the Join dialog box.

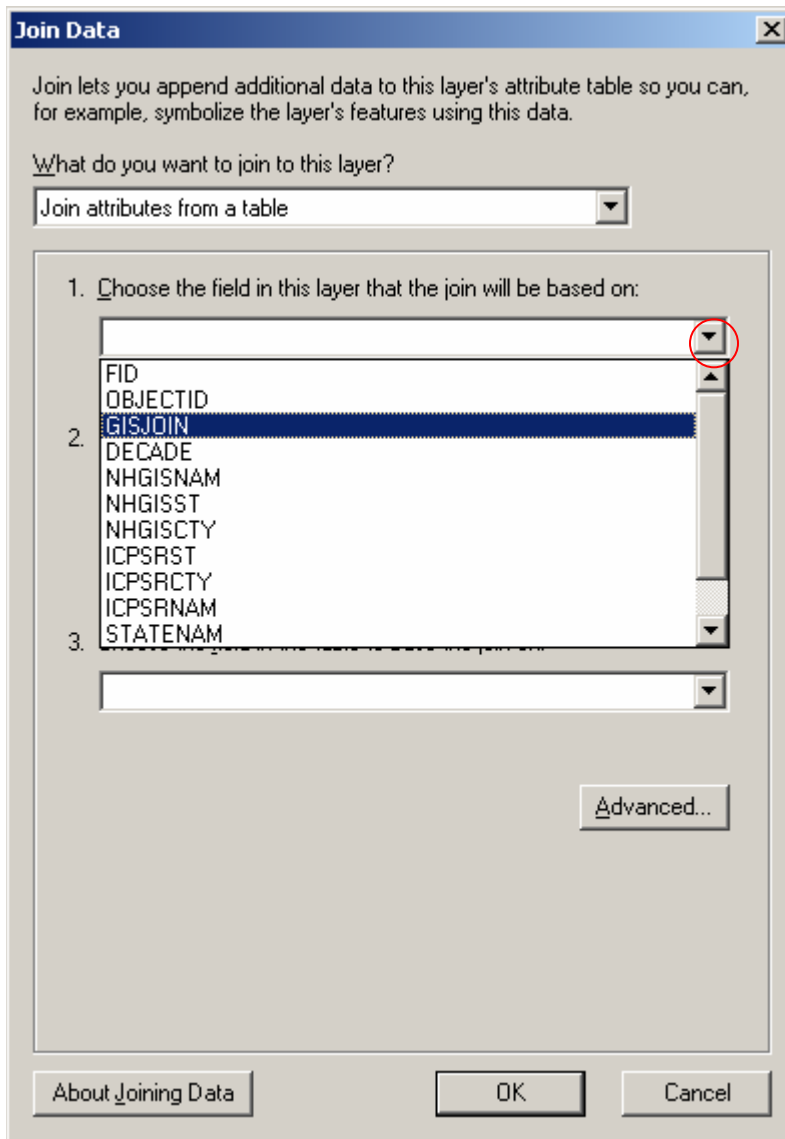
Right-click on the shapefile in the Table of Contents and then select *Joins and Relates* from the context menu. Finally, click on Join... to open the Join dialog box.



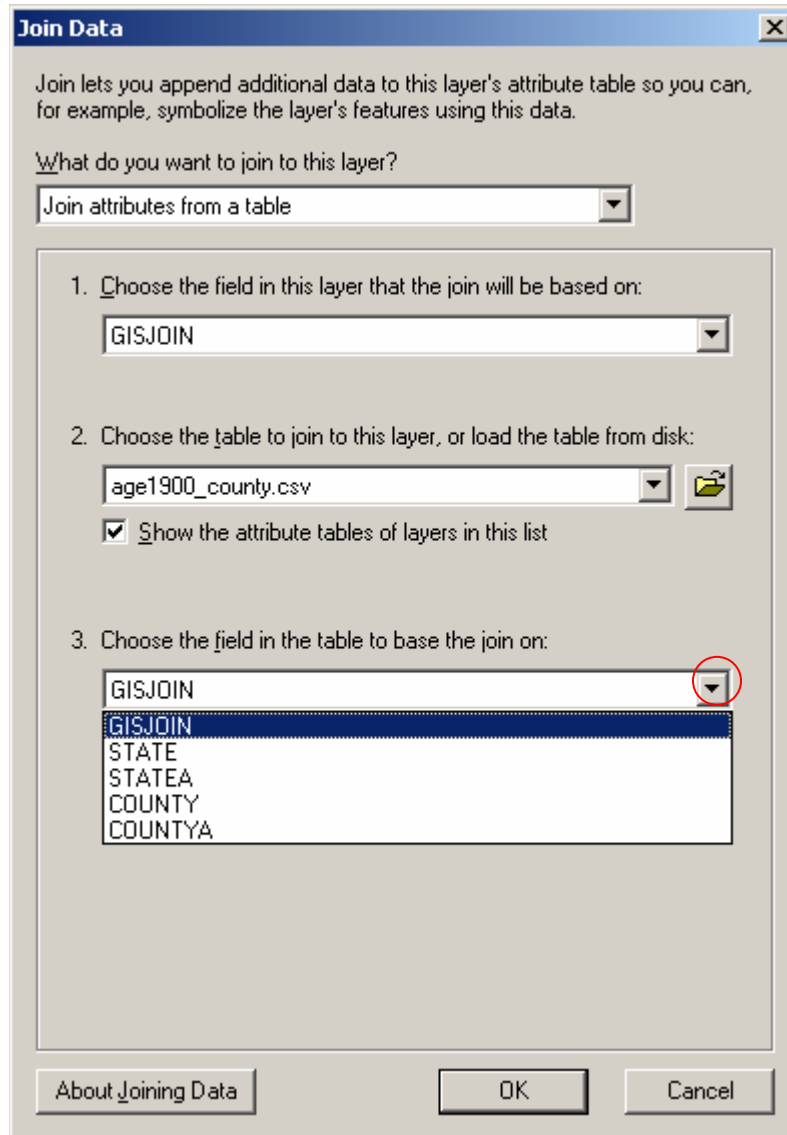
### Step 3 – Choose the common fields in the shapefile and aggregate data file

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 red) and select ***GISJOIN*** from the list of attributes.



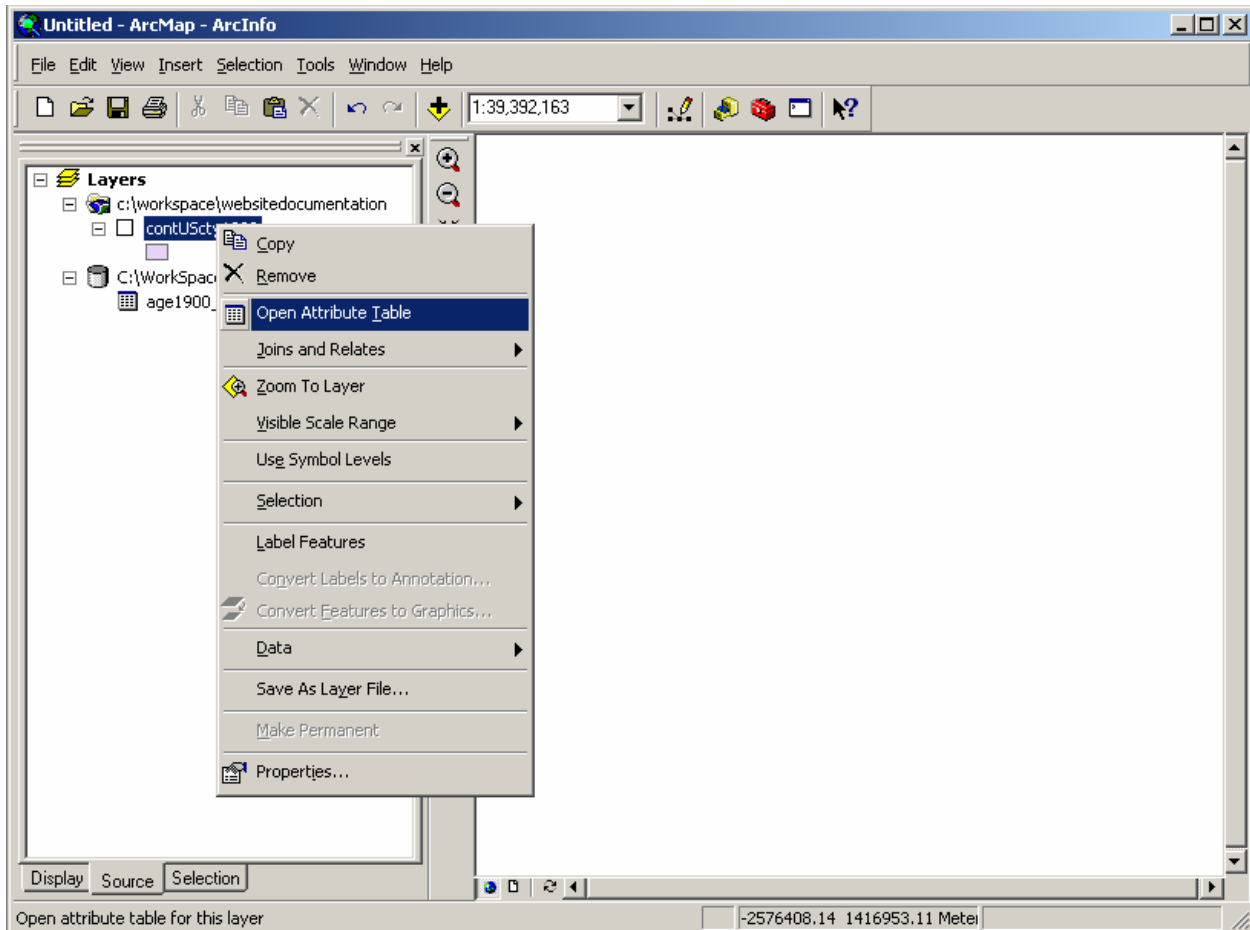
Second, choose the attribute in the aggregate data file. All NHGIS aggregate data files contain an attribute called *GISJOIN*, which is a unique identifier for every geographic feature in the file. Click on the down arrow (circled in red) and select *GISJOIN* from the list of attributes. Then, click **OK** to execute the join.



**Step 3 – Examine the result of the join.**

Open the attribute table of the shapefile and 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.



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. The *contUScty1900.GISJOIN* attribute (circled in red) is from the shapefile and the *GISJOIN* attribute (circled in blue) is from the aggregate data file. Please note that the joined aggregate data traditionally follows all of the attributes originally available on the shapefile. For illustrative purposes we moved the two joined fields closer together.

| contUScty1900.GISJOIN | GISJOIN | contUScty1900.DECADE | contUScty1900.IHGISIAM |
|-----------------------|---------|----------------------|------------------------|
| 2000350               | 2000350 | 1900                 | Cowley                 |
| 2000370               | 2000370 | 1900                 | Crawford               |
| 2000390               | 2000390 | 1900                 | Decatur                |
| 2000410               | 2000410 | 1900                 | Dickinson              |
| 2000430               | 2000430 | 1900                 | Doniphan               |
| 2000470               | 2000470 | 1900                 | Edwards                |
| 2000490               | 2000490 | 1900                 | Elk                    |
| 2000530               | 2000530 | 1900                 | Ellsworth              |
| 2000550               | 2000550 | 1900                 | Finney                 |
| 2000570               | 2000570 | 1900                 | Ford                   |
| 2000590               | 2000590 | 1900                 | Franklin               |
| 2000610               | 2000610 | 1900                 | Geary                  |
| 2000630               | 2000630 | 1900                 | Gove                   |
| 2000650               | 2000650 | 1900                 | Graham                 |
| 2000670               | 2000670 | 1900                 | Grant                  |