

## Introduction

A common need in GIS is to map characteristics, or attributes, of different geographic areas. These maps are called thematic maps. Examples of thematic maps might include maps of the population of cities, counties, or states; election results by state, county or election district; or incidence of particular species within mapped areas or units (such as wetlands). You may collect this attribute data yourself or obtain it from another source.

To create a thematic map, you need an appropriate basemap (spatial data), and the attribute data you wish to display.

**Spatial data:** A digital basemap displaying boundaries, roads, or some other geographic feature(s). Spatial data is georeferenced in a real world coordinate system and may contain some attribute information associated with its features. In a GIS, information about the features in a spatial dataset is stored in the dataset's attribute table.

**Attribute data:** Descriptive information that describes the geographic features represented in your spatial dataset. Attribute data must be related via a unique identifier to a geographic feature. For example, to map election results by county, the table of election results (attribute data) must contain a column that uses the same county identifier as the attribute table of the spatial dataset. In this case, that identifier could be the county name, an abbreviation, the county FIPS code, or any other unique identifier common to both datasets.

ArcMap provides two methods to associate data stored in tables with geographic features: **join** and **relate**. When you join two tables, you append the attributes from one onto the other based on a field common to both tables. In contrast, relating tables defines a relationship between two tables - also based on a common field - but doesn't append the attributes of one to the other. **Join** tables when the table on which you're performing the join has only one value for each record in the second table (for example, joining the countries listed in one table with their capitols, listed in a second table). **Relate** tables when the table on which you're performing the join may have multiple values in the second table associated with each record (for example, joining the countries listed in one table with all of their major cities, listed in a second table).


When the attribute data you want to combine with spatial data is properly formatted and the datasets have a common unique identifier for the appropriate geographic area, joining the two is straightforward. When there is no common identifier, you'll have to edit one or both datasets to create one.

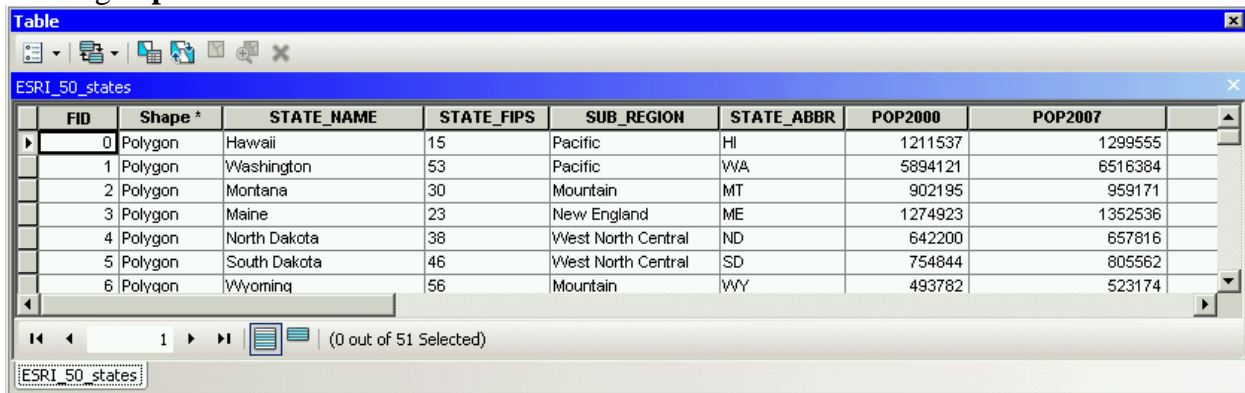
## Part 1: Data Collection and Editing

In this exercise, we'll start with a shapefile of state boundaries for the United States. We'll also look at an Excel spreadsheet containing data on the home state for incoming freshmen at Cornell, and edit it so that it can be joined to the attribute table for a spatial dataset containing state boundaries.

The data files you'll need for this exercise are located in the directory `\GIS\ArcGIS_2\Attribute\` in the **Classroom Files** folder (use the desktop shortcut to get there). Check to see if there is a folder named "WorkSpace" at the top level of the `C:\` drive on your computer. If no such directory exists, create one. It's important that there be no spaces or special characters anywhere in the path directories of your data files, so be sure to create this directory at the top level of the `C:\` drive. Copy the entire **Attribute** folder to `C:\WorkSpace` on your computer. The spatial data files are from ESRI, the makers of ArcGIS software. These data files and other data from ESRI are available on DVD-ROM stored at the Stone Computing bookshelf in Mann Library. The attribute data are actual data on incoming freshmen at Cornell for fall 2004, courtesy of the Office of Institutional Research and Planning.

### Part 1.1: Add Spatial Data to ArcMap

1. Start ArcMap (go to **Start>All Programs>ArcGIS>ArcMap 10**). Click **OK** when asked you if you'd like to begin with a **new empty map**. Maximize the window and dock (drag) any toolbars so they're out of the way.
2. To add your data to the map view, click the **Add Data** button . Navigate to your `C:\WorkSpace\Attribute` folder and select **ESRI\_50\_states.shp**. Click **Add**.
3. Open the attribute table for **ESRI\_50\_states** by right-clicking on the layer name in the table of contents and selecting **Open Attribute Table**.



FID	Shape *	STATE_NAME	STATE_FIPS	SUB_REGION	STATE_ABBR	POP2000	POP2007
0	Polygon	Hawaii	15	Pacific	HI	1211537	1299555
1	Polygon	Washington	53	Pacific	WA	5894121	6516384
2	Polygon	Montana	30	Mountain	MT	902195	959171
3	Polygon	Maine	23	New England	ME	1274923	1352536
4	Polygon	North Dakota	38	West North Central	ND	642200	657816
5	Polygon	South Dakota	46	West North Central	SD	754844	805562
6	Polygon	Wyoming	56	Mountain	WY	493782	523174

Notice the fields (column headings) across the top of the table. We'll need to select one of these to use to join our attribute data to this table. Recall that we want to display the number of incoming freshmen from each state, so we're displaying data by state. The fields that would be good candidates to use in the join are `STATE_NAME`, `STATE_FIPS`, and `STATE_ABBR`. If one of these fields already exists in our attribute dataset, we're half way there. Close the attribute table.

*Note:* Do not be tempted to use the `FID` field to join tables; it is a number that is automatically generated by ArcGIS and has no geographic meaning.

4. **Save** your map to your `C:\WorkSpace\Attribute` directory (choose **Save** under the **File** menu).
5. Minimize the ArcMap window, but don't exit the program.

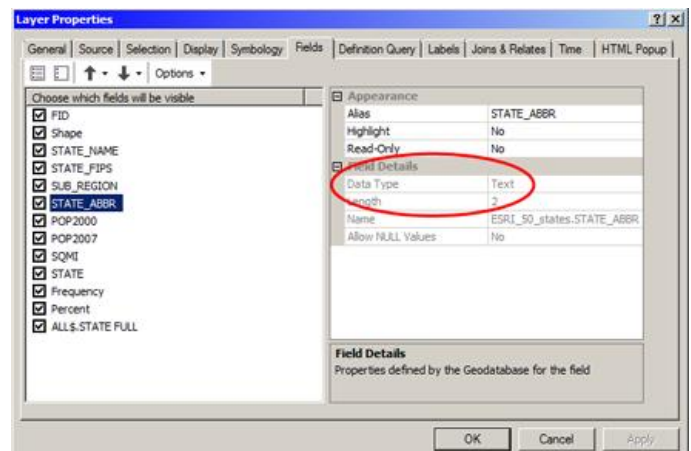
## Part 1.2: Edit Attribute Data

File and data formatting are the most common sources of problems when trying to join data. Recent versions of ArcGIS allow direct linking and joining of individual Excel (.xls and .xlsx) spreadsheets, which helps to make sure that values. This allows the use of OLE (Object Linking and Embedding), where data can be maintained and updated outside of ArcGIS. (When the Excel file is updated, the new data automatically appears within ArcGIS.) Note that Excel tables are read-only in ArcGIS, and when they are open in ArcGIS they will be read-only in Excel, so you'll have to close your ArcGIS project before editing the file in Excel. If you want to be able to edit your tables in ArcGIS, then you'll need to export from Excel as a .dbf or .csv file.

1. From Windows Explorer, open (double click) the Excel file (“**2004 home state for GIS.xls**”) from **C:\Workspace\Attribute** folder. Notice that this file consists of three separate sheets --see the tabs in the lower left corner, and make sure the “ALL” sheet is selected. The first two columns headings, “STATE” and “STATE FULL”, correspond to the “STATE\_ABBR” and “STATE\_NAME” fields in the attribute table of our shapefile, and either one would be a good choice to use to join the data. (We'll use the state abbreviations for this exercise.) Also note that the “Frequency” column contains the data on the number of incoming students.

To avoid potential problems with some of the older tools in ArcGIS, we recommend taking some precautions when preparing your tabular data:


2. The first row of the table should contain column headings, rather than data. Column headings should begin with a letter, and only contain letters, numbers, or underscores. Note that the column names in the table can be different than the corresponding columns in the shapefile attribute table.
  - Change the “STATE FULL” heading to “STATE\_FULL”
3. The filename should also begin with a letter, and only contain letters, numbers, or underscores. (Of course the file extension needs to be preceded by a period.)
  - From the **File** menu, choose **Save As** and name the new file “**StudentData2004.xls**”, saving it within the **C:\Workspace\Attribute** folder.
4. It's often a good idea to check the data type (number, text string, etc.) for the fields on which you are basing the join, especially if you are using zipcodes, FIPS codes, or other identifiers that look like numbers but are actually text strings that may start with a zero. The data types for these columns must be the same in both tables. The sample data requires no changes, but here's how to force a column to use a text datatype.
  - In Excel, select the entire “STATE” column by clicking the column A header, then right-click it and select “Format Cells”. The data type will be highlighted. The “General” category in Excel means that no explicit datatype has been defined for the cells. Select the “Text” category and click “OK”. Save the file.
  - In the ArcMap table of contents, right-click the “ESRI\_50\_states” layer and select “Properties”, then click the “Fields” tab, and then choose the “STATE\_ABBR” field. Look on the right to confirm that this field has a “Text” data type. Look to see the datatypes of some of the other fields, then click “Cancel” to close the Properties window.

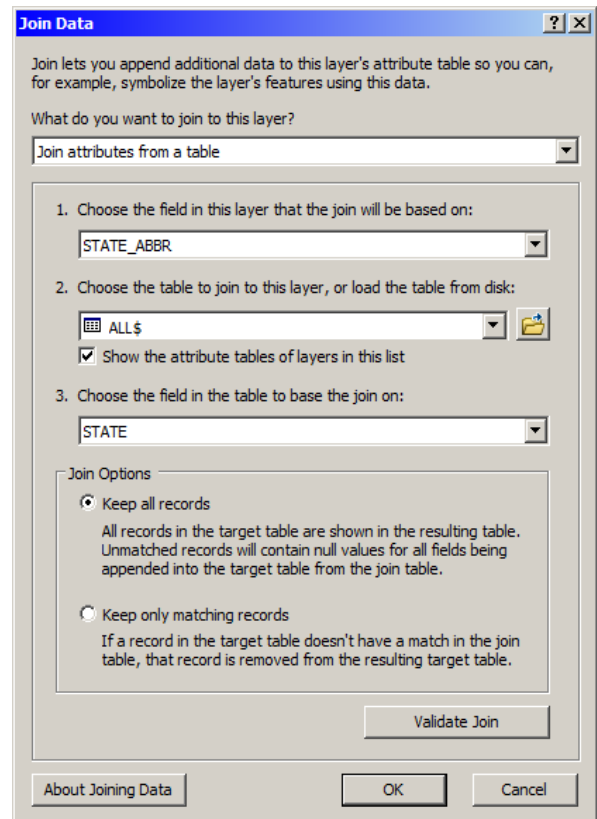


## Part 2: Joining Tables and Creating a Thematic Map

In this exercise, you'll join the file containing your attribute data to the attribute table for the spatial dataset, and display the data on a map.

### Part 2.1: Joining Tables

1. In ArcMap, add your new **StudentData2004.xls** file. Click the Add Data button  as done previously, select the file and click **Add**. You can now select the individual worksheet you wish to use. In this case, select **ALL\$**, and click **Add**.
2. Right click and open the **ALL\$** table to make sure ArcMap is able to read your file. If the table values do not appear as expected, you may need to go back and re-edit your Excel spreadsheet.
3. Right click the **ESRI\_50\_states** layer and select **Joins and Relates**, then select **Join**.
4. The **Join Data** dialog box will open. Set the parameters as pictured here, and then click **OK**.



5. Right click on the **ESRI\_50\_states** layer and open the attribute table. Scroll all the way to the right to see that the last four columns are those that have been joined from the Excel spreadsheet. Notice that there are some **<Null>** values in the row for South Dakota. This is because the Excel spreadsheet did not include a row for that state, since there were no new Cornell students from South Dakota in 2004.
6. Close the attribute table.

Even though the tables appear to be combined at this point, they are actually only linked via OLE (Object Linking and Embedding). If you anticipate that you will be updating the spreadsheet in the future, you may want to leave the joined tables as they are. But sometimes what we want is a new shapefile that contains all the data in a single layer. For this exercise, we will permanently join the tables in a new spatial file.

7. Right-click on the **ESRI\_50\_states** layer and select **Data > Export Data**. Select **All features** from the **Export** pull-down menu and make sure **Use the same coordinate system as: this layer's source data** is checked. Save your file in the **C:\Workspace\Attribute** directory and name it **HomeStates\_2004.shp**, and **Save as type: Shapefile**. Click **Save**.
8. Click **OK**. When you are asked if you want to add the data to the map as a layer, click **YES**.
9. After the new shapefile is added to the Table of Contents, turn off (uncheck) the **ESRI\_50\_states** layer and make sure the **HomeStates\_2004** layer is turned on (checked).
10. Right click and open the attribute table for **HomeStates\_2004**. Notice that all the fields have been combined. (If there were any columns in the Excel file with the same name as a column in the shapefile, the joined name may have been slightly altered to avoid duplicate names.) Close the attribute table.

*Note: You might have noticed that there were records in the original Excel ALL spreadsheet that didn't correspond to states (e.g. "INTERNATIONAL" and "ARMED FORCES"). Because there was no corresponding record in the spatial attribute table, these data do not appear in the joined table. Also, as mentioned earlier, there was no record in the Excel file for the state of South Dakota. Consequently, South Dakota does appear in the resulting joined table, and in the first version of the joined table, you would have seen <Null> in the cells with missing data. In the exported table, ArcMap automatically inserted a zero in the **Frequency** (number of students) column. It's worth considering what might happen to data when you join tables that have missing data, or records that are unique to one of those tables. If you want to restrict a join to only records that match, use the **Keep Only Matching** option in the **Join Data** dialog box. (The default option keeps all records from the join table, regardless of a match, in the output table.)*

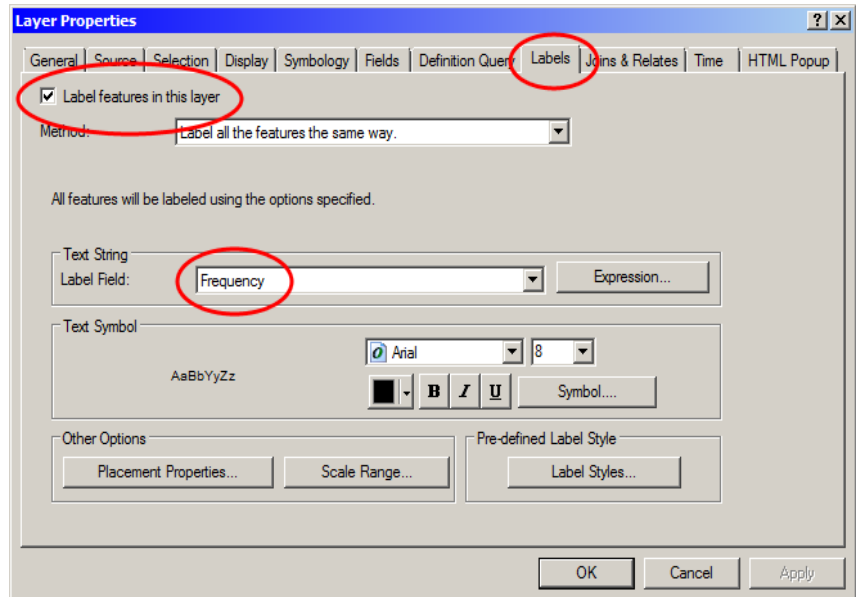
11. Since we now have all the data we need in the HomeStates\_2004 layer, we can remove the other extra layers and table (**ESRI\_50\_states** and StudentData2004 **ALL\$** spreadsheet). Simply right-click each one and select **Remove**.
12. **Save** your map.

## Part 2.2: Display Your New Data on a Map

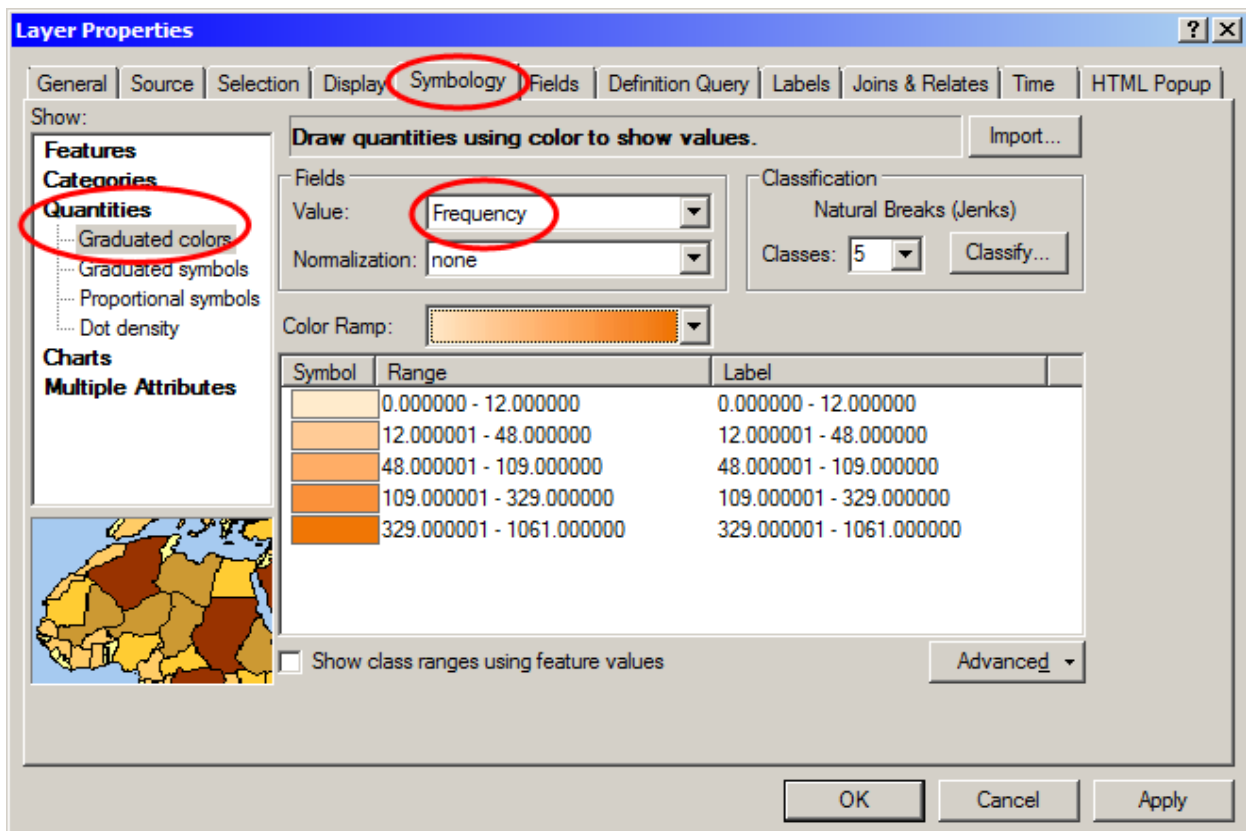
In this section, we'll use labels to display the number of incoming students per state on the map, and create a map layout for printing.

1. Label the map with the number of students from each state:
  - Right-click on the **HomeStates\_2004** layer and select **Properties**.
  - Click the **Labels** tab in the **Properties** window.
  - Check the box that says "Label features in this layer".
  - Select "**Frequency**" from the **Label Field** pull-down menu.

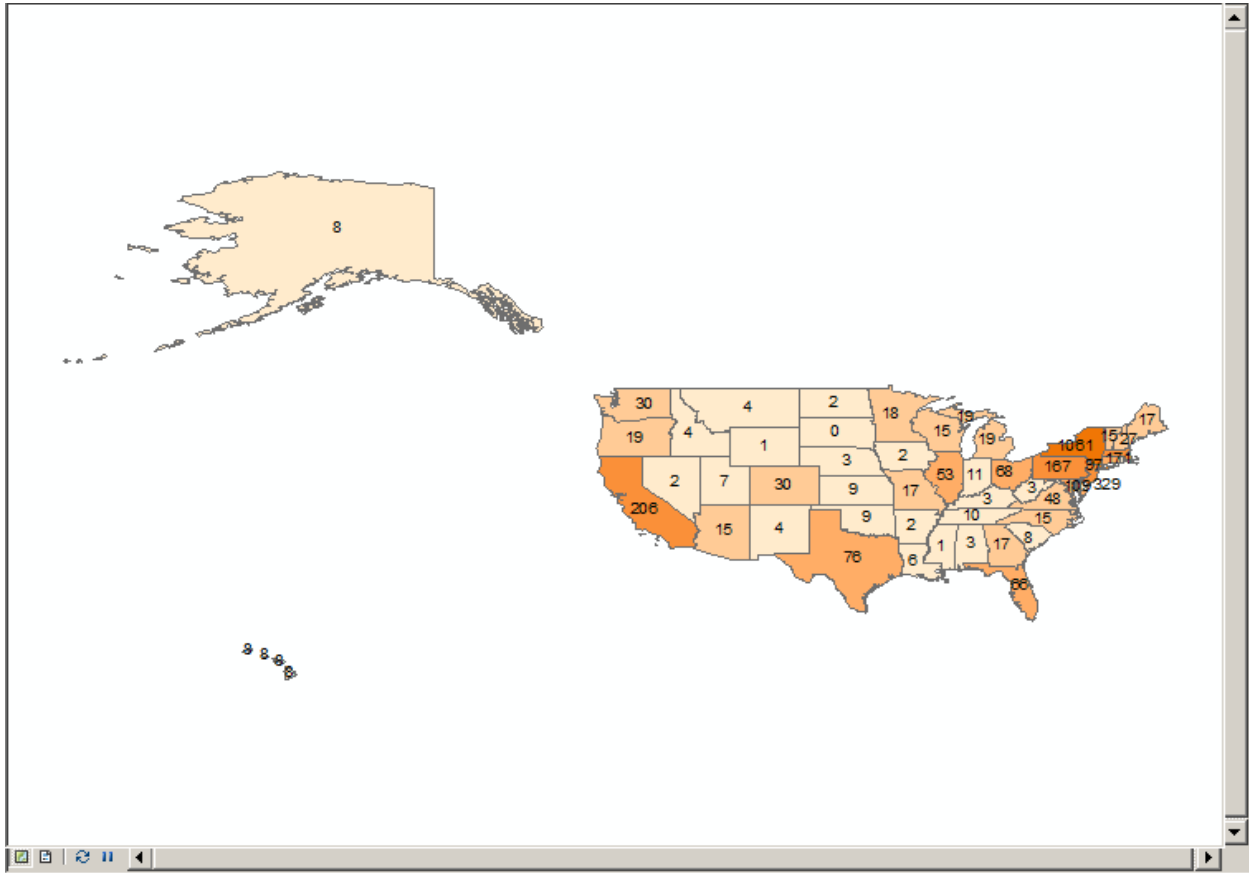
Notice that you can also change the font style, size, and color in this box. Make changes if you want to, then click **OK**.



2. Color each state according to the number of students:
  - In the table of contents, right-click "**HomeStates\_2004**", select **Properties...**, then the "**Symbology**" tab.
  - Choose **Quantities** > **Graduated colors**, and set the **Value** dropdown to "Frequency".
  - Choose a color ramp that goes from light to dark (but not too dark, so that the labels will still be legible).

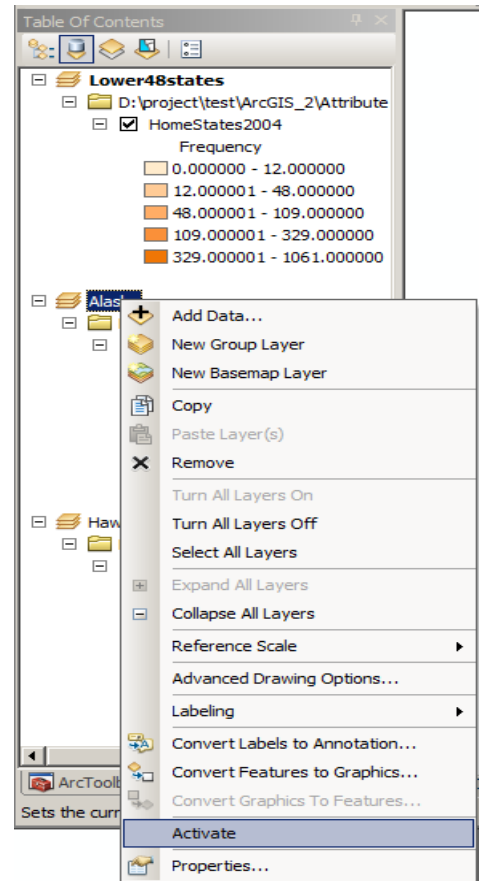




You've probably also noticed that the projection distorts the state of Alaska quite a bit, that there are multiple data labels for the state of Hawaii, and that there is a lot of extra white space in the display. We'll solve these problems by creating different data frames for the lower 48 states, Alaska, and Hawaii, selecting more appropriate projections for each frame, and adjusting the labeling.



3. Create a new data frame by selecting **Data Frame** from **Insert** menu.
4. Name this new frame by **clicking twice** (slowly) on the **New Data Frame** text and entering "Alaska" in the Name box. If you accidentally double-click instead, the Data Frame Properties box opens, and you can add the name under the general tab then click OK.
5. Add the **HomeStates\_2004** layer to the **Alaska** data frame by right-clicking on the **HomeStates\_2004** layer, selecting **Copy**, right-clicking on the **Alaska** data frame, and selecting **Paste layer(s)**.
6. Repeat steps 3-5 to create a third data frame, but name this one **Hawaii**.
7. Rename the original data frame **Lower48states**.

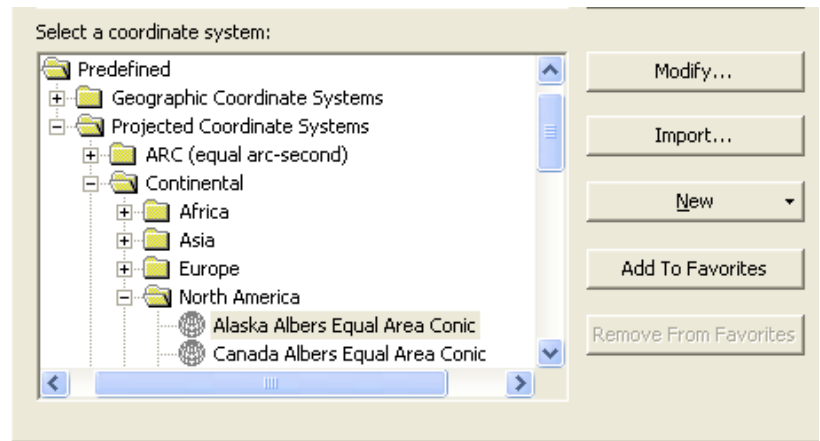
8. You should now have three data frames listed in the table of contents, each with a HomeStates\_2004 data layer. The active frame's name is in **bold** text in the table of contents. Activate the **Alaska** data frame by right-clicking its name and selecting **Activate**.



9. Use the mouse scroll wheel or the **zoom**  and **pan**  buttons to zoom in and center the state of Alaska in the data view display. (Alternatively, open the attribute table and select Alaska by clicking the box to the left of the row for Alaska, close the table, right click the **Alaska>HomeStates2004** layer and choose **Selection>Zoom to selected features**. You can clear a selection from the same **Selection** menu.

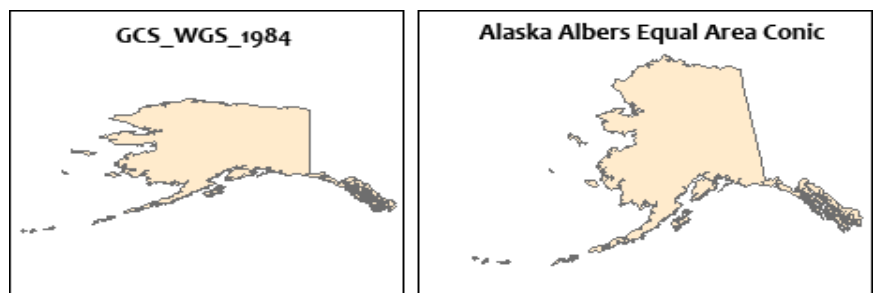
10. Change the projection for this frame by double clicking on the data frame name to open the **Data Frame Properties** box, and select the **Coordinate System** tab. Notice that the current coordinate system is “GCS\_WGS\_1984”, which is a common flavor of latitude/longitude coordinates.

11. Using a coordinate system that produces less distortion for a given area will improve the appearance of the map. For Alaska, one such system is the **Alaska Albers Equal Area Conic**, which you will find under **Predefined\Projected Coordinate Systems\Continental\North America folder**. Select this system in the dialog box, and click **OK**.



Notice how the shape of Alaska (and any other visible states) has changed.

Feel free to experiment with other projections if you have time. When you're finished, you may have to zoom and pan again to get just the state of Alaska in the frame.



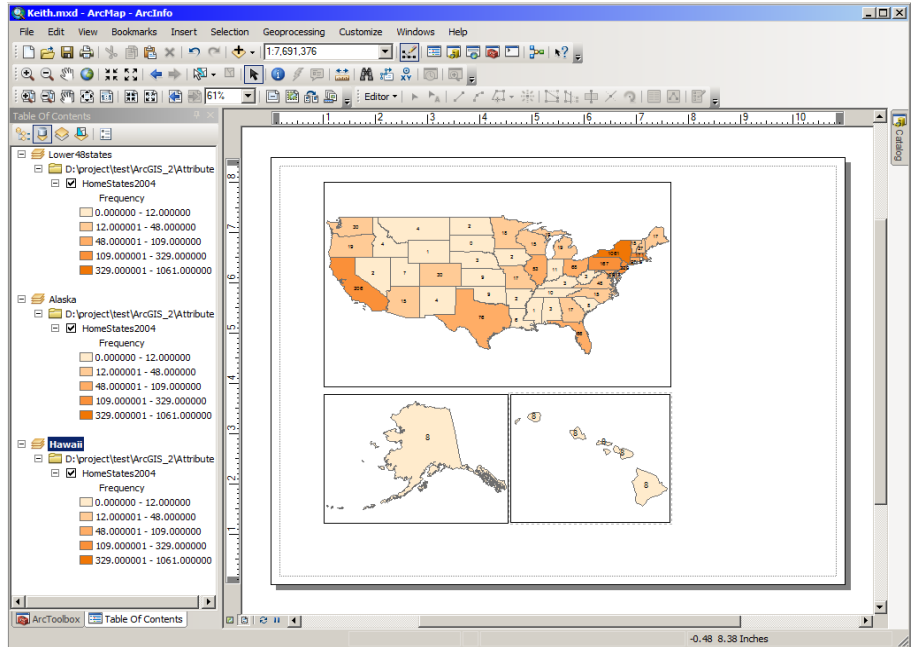
12. Change the **Hawaii** data frame coordinate system to **Hawaii Albers Equal Area Conic** (found in the the same coordinate system folder used in step 11).

13. Change the **Lower48states** data frame coordinate system to **USA Contiguous Albers Equal Area Conic**.

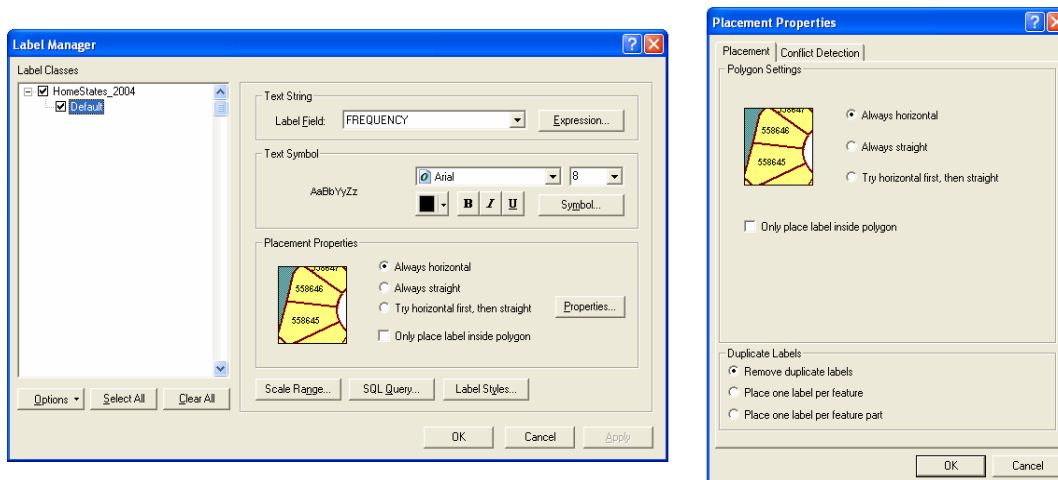
14. **Save** your map.

### Part 2.3: Customize the Page Layout

15. Change from Data view to Layout view in ArcMap using the **View** menu > **Layout View**. (There is also a very small layout view button in the lower left corner of the map display.) You should see all three data frames appearing in the layout, in random sizes and positions.
16. Change the page orientation from portrait to landscape by choosing **Page and Print Setup** from the **File** menu. Change the orientation to **Landscape** in the **Printer Setup** section. Click **OK**.
17. Select, resize, and position the frames so they do not overlap (click the **Refresh** button, the two arrows next to the **Layout** button, if you need to refresh the display).



18. The **Hawaii** data frame is displaying a data label for each island. To fix this, right click on the **Hawaii** data frame in the table of contents, and select **Labeling** > **Label Manager**. Click on **Properties** in the **Placement Properties** area. Select the **Placement** tab if needed. Check **Remove duplicate labels** in the **Duplicate Labels** area of the dialog box. Click **OK**, then **OK** again.



19. Look at the options available to you from the **Insert** menu at the top of the screen. You might want to insert a **Title**, some explanatory **Text**, or other pictures or objects. The **North arrow**, **Scale bar**, **Neatline**, and **Legend** choices will bring up a wizard or dialog box to guide you through the process of adding them to the layout view. Use these tools to complete your map to your satisfaction.