

Exercise 15. Mapping Census 2000 Data

Purpose: In this exercise you will utilize *ArcMap* and existing digital data sources to produce a choropleth map of Los Angeles County. While other mapping packages are available, all CSU campuses have access to this software through a site license and so it will be used as an example.

GIS packages like *ArcGIS* are proving useful for general mapping purposes and they do offer the added advantage of being able to link an entire symbol set to a data attribute. Thus groups of symbols can be designed and modified quickly. *ArcGIS* also can convert all GIS layers into *Adobe Illustrator* layers. That graphic software is very useful for finishing the map and for creating posters from other graphic elements.

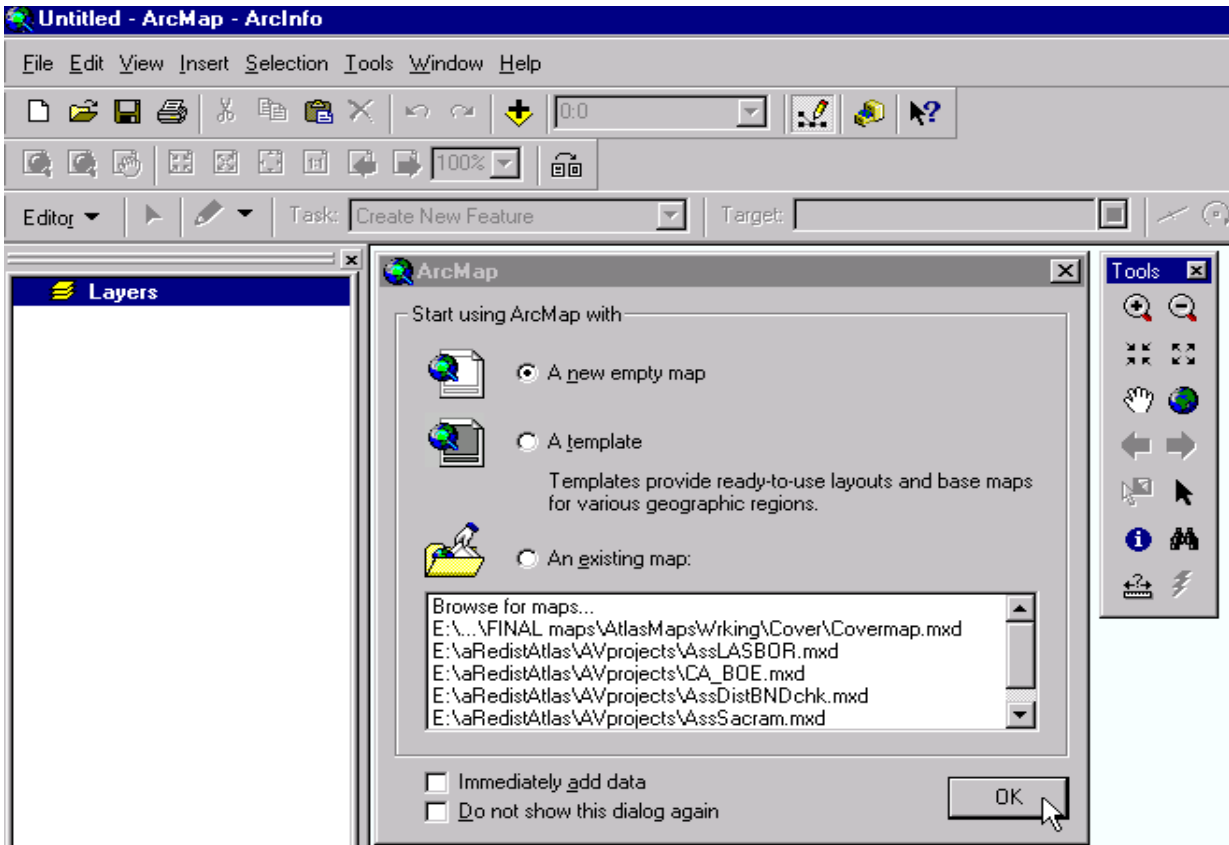
To produce the map, two files will be needed. One is the boundary file that contains the positions of all points that describe the outlines of counties or other geographic areas. The second is the data file that must be linked to the boundary file in order to map some variable of interest. Both files must contain a variable that uniquely identifies each geographic unit and these will serve to join the records of both files into one long record. Then, values from the second file can be displayed graphically within the boundaries.

Loading Files for Mapping

1. To begin, locate *ArcGIS* and *ArcMap* software on your machine. Then copy the *Mapping* folder to your machine.
2. Make sure the California county shapefile (*Caco*) is present.
3. From the *Start* button in *Windows* locate *ArcGIS*, the *ArcMap* option, and open it.

The window shown below will open.





The left part of the window in the illustration is the *Table of Contents*. A default *Layers* icon is shown below which any added data sets or layers will be listed.

The larger window in the center is the *Display* window (now mostly covered by the ArcMap window and it shows your map once the layers have been loaded.

On the far right is a *Tools* menu that contains mostly browsing features. Move the cursor over each to see what they do - which is pretty self explanatory.

In the middle of the screen is a gray window giving you the option of opening a blank template, using an existing template, or opening an existing project. Templates are basically predefined map layouts that are invaluable for doing a series of similar maps. They may be simply layout windows or may contain partially completed maps.

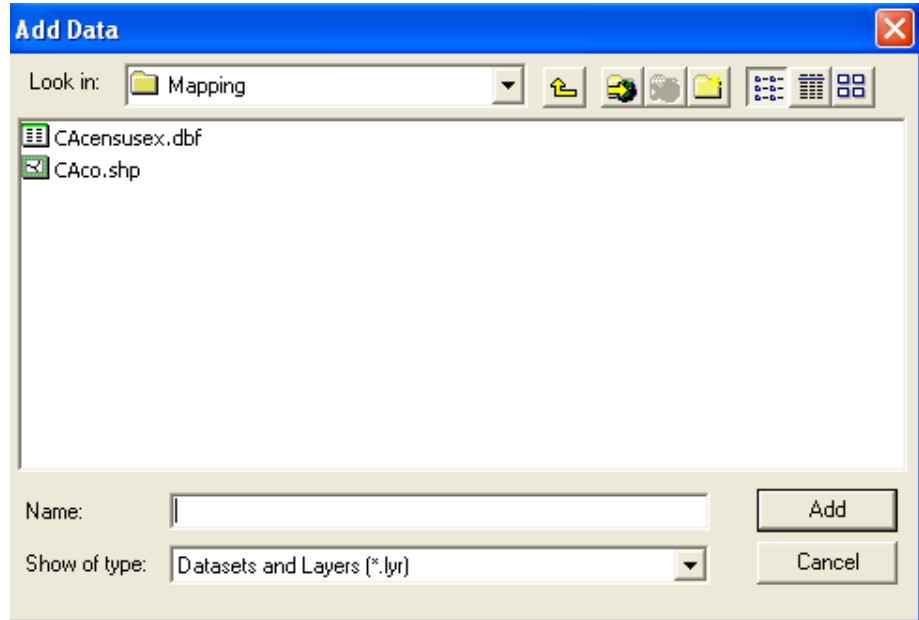
The most important icon for now is the diamond and cross at the top of the main menu. This allows you to add data to your project.



4. Click the *OK* button on the center window and then the *Add Data* icon at the top of the screen.

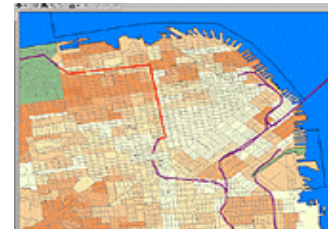
5. When the *Add Data* window opens (see below) select the California county outline boundary file named *Caco*. Then click the *Add* button.

Note that *ArcGIS* represents different data types with different icons. Point data are represented with three dots, line data with a line, and polygons with three joined polygons. *Dbf* files like *Ancestries* are shown with some column-like dashes.



Most mapping projects are composed of multiple layers though, in this exercise, we will keep things to an absolute minimum. The type of map you are going to create cartographers refer to as a *choropleth* map. In *Arcmap* it is called a *Graduated Color* map.

Basically statistical numbers are displayed within their sampled areas. Here that would be county units. Also, the data values are assigned to several classes so that the map takes on a sort of “quilt” effect from the colors assigned to the several classes. Cartographers have spent considerable effort on finding appropriate methods to determine the appropriate number of classes and where break points should occur within a distribution. The default method used in *Arcmap*, inappropriately called *natural breaks*, is an excellent starting point for setting break points in a distribution. As for the number of classes, the rule of thumb is to choose between three and eight. Fewer are needed when there are fewer areas such as with this map of California counties.



The choropleth map is a very common type of statistical or thematic map type, but it does have some important caveats.

1. Because of difference in area sizes, displaying numerical totals is usually not appropriate. The obvious result of this is that large areas will always appear in the highest classes and small areas will usually appear in the lower classes. For example, if you mapped number of children, retirees, singles, and executives the resulting patterns would all be virtually identical - reflecting the distribution of total population. Usually you want to map a percentage, percent change, density, or some other statistic adjusted for population or area.

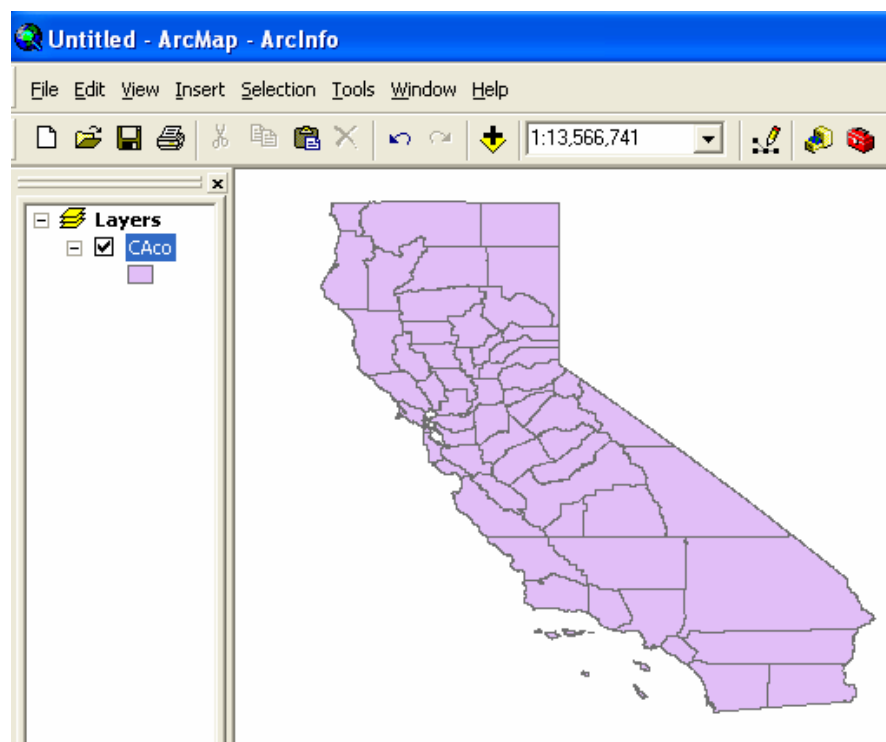
2. It is possible that important areas are missed because they are small and an inset map of an enlarged subarea might be necessary. A good example are the boroughs of New York that contain significant numbers of people but are virtually invisible on a page-size map of the entire United States. Because of the larger sizes of areas in the Western United States, one often gets a greater sense of importance for these areas than for those in the Eastern United States. In your map, San Francisco County is very small.

3. The sequence of colors selected for the map categories should increase in impact with the values in the classes. Thus the highest category should visually stand out most. The best way to accomplish this is to let the value of a color change with increasing magnitude. In ArcMap you can pick a strong, dark color for the highest category and a pale color for the lowest category and create a ramp. The software will calculate a series of transition colors between the extremes. All too often maps are created showing categories with different hues such as red, green, yellow, and blue and this makes estimating the values of intermediate categories difficult.

4. Be careful about areas with very small total populations since they can generate very high percentages. For example, in 1980 one census tract in Los Angeles County had over 12% American Indian. There were only 8 people in the tract and one was an American Indian. In some cases you might set a minimum population threshold to reduce this effect.

5. Always place a few locational references on these maps to help readers identify locations. These might include major cities, roads, rivers, or other significant regions. Other than countries or states, people are probably unfamiliar with exactly where polygons are located. In the case of our map, displaying San Francisco, Sacramento, Los Angeles, and San Diego would be helpful.

When the county layer is added, you will see a list of layers and a map as shown right. Any symbols shown are default and these can be changed. The map also has no projection. The lack of projection causes California to be 20-30% wider than it should be.



Another common problem in GIS is that data sets may have different coordinate systems or different datums. Usually if the coordinates are different, the second set will not appear on the map. If the datums are different, the layers will be offset at larger scales. ArcGIS will usually warn you when problems are encountered.

B. Setting up a Projection

Currently all the coordinates are in latitude and longitude and you can see the location values of your cursor in the bottom portion of the window.

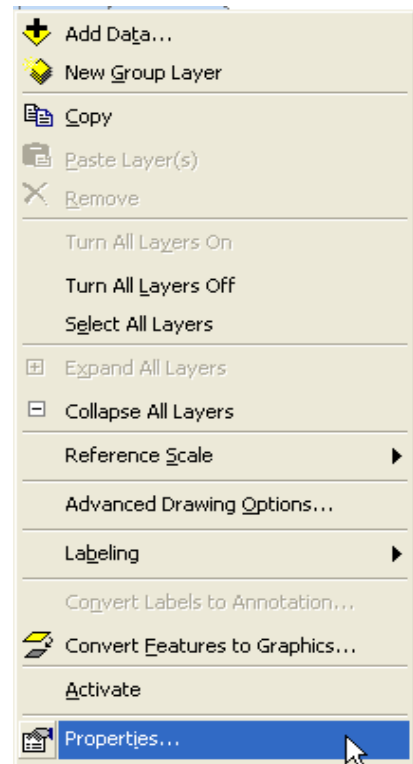
118°7'12.91"W 33°36'55.42"N

The default map is represented as if the coordinates were cartesian (x,y). This creates what is called a platte carree projection that has a great deal of distortion in higher latitudes.

Unprojected maps are increasingly appearing in the literature which indicates that many people are unaware or don't care about the distortion of area and directions being presented. The horizontal exaggeration exists even if a very small area is mapped at a very large scale. One could potentially make some serious measurement errors from such maps, and so it is worth a few moments to learn to assign a projection to a map. Fortunately we can create a projection "on-the-fly" in ArcGIS that does not change the actual coordinates. Let's pick a common projection for mid latitude areas, *Albers Equal Area Conic*.

The conic projection is typically used for mid-latitude places and especially for those with a prominent east-west dimension like the U.S., Canada, Europe, Russia, China, and Australia. For portraying geographic distributions a projection that preserves area also is desirable so that the sizes of places across the map are in proportion to their size on the earth. Your concern is to locate a central meridian in the middle of your desired area and two standard parallels that divide the vertical extent of your area of interest more or less into thirds.

For larger areas the Universal Transverse Mercator projection is often used. Each UTM zone covers six degrees of longitude and for the continental U.S. the first zone is Zone 10 on the west coast (includes northern California) followed by Zone 11 for Southern California. The zones proceed eastward in six degree steps. Furthermore, each zone is divided into a north and south half and all coordinates are laid out on a regular grid in meters. The origin for the grid is on the equator 500,000 meters to the west of the central meridian of the zone. All values in the zone are positive. One of the UTM projections is often useful for mapping one or more counties or even smaller areas.



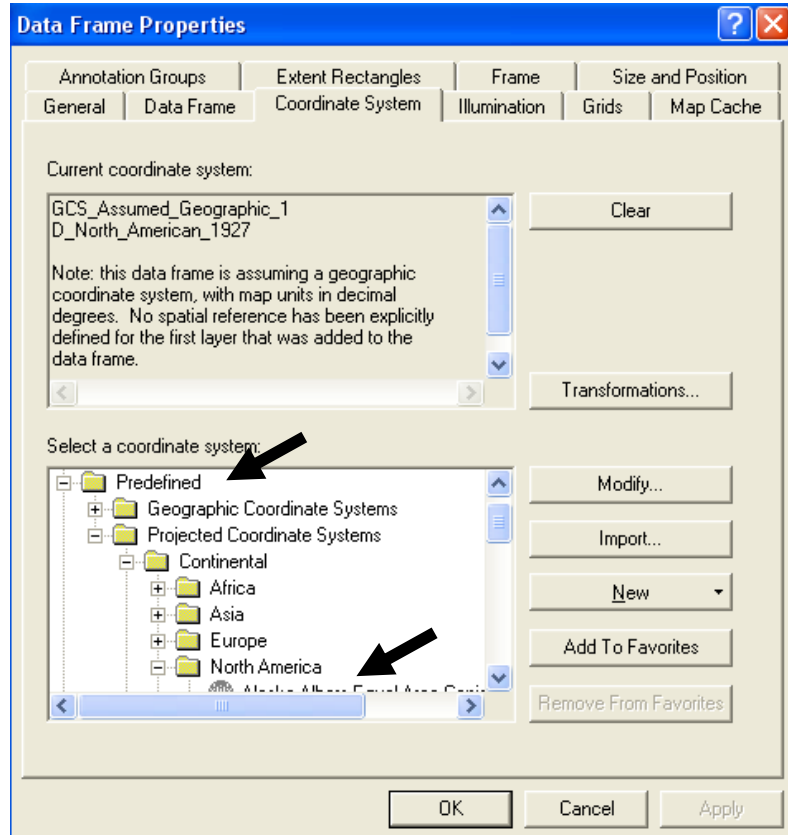
6. Right-click on the map or on the yellow *Data Frame* icon in the *Table of Contents*.

7. From the popup menu at right scroll down and select the *Properties* option.

Note well, that you will do a lot of "right-clicking" in *ArcGIS*. In particular, you will be accessing the *Properties* option of this menu.

When the *Properties* window opens make sure the *Coordinate System* tab is selected in the new window. This map has a global coordinate system (GCS) based on the *North American datum of 1927*.

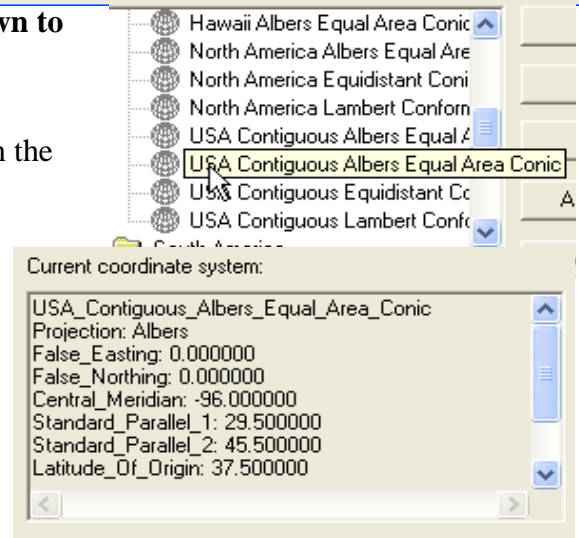
8. To select a projection from the *Select a coordinate system* window (see lower window), open the *Predefined* directory and click on the *Projected Coordinate Systems* directory, then the *Continental* directory, and finally the *North America* directory.



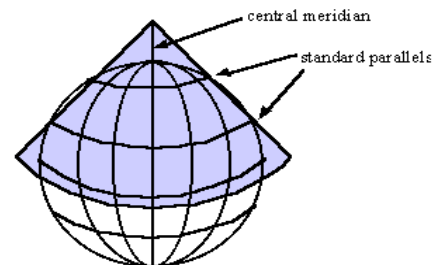
A list of projections will appear.

9. Under the *North America* directory scroll down to the *USA Contiguous Albers Equal-Area Conic*.

The projection properties will now be listed in the *Current Coordinate System* window.

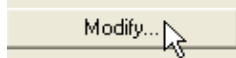


The cone is usually centered over a pole and its surface contacts the earth along a parallel. The best representation of the earth occurs along that parallel (called the *Standard Parallel*) and as you can see, conic projections are ideal for mid-latitude regions. Usually the cone is made to cut into the earth at one parallel and re-emerge at another to further improve the area of coverage. Thus, two standard parallels are called for along with a central meridian to center the map on.



You will need to make a couple changes to this projection since it was set for a map centered on North America not California. You will have to move the central meridian westward and pick two lines of latitude that lie well within California. Note that longitude values in the Western Hemisphere are given negative values.

10. From the set of buttons in the right of the *Projection Coordinate System* window select the *Modify* button.

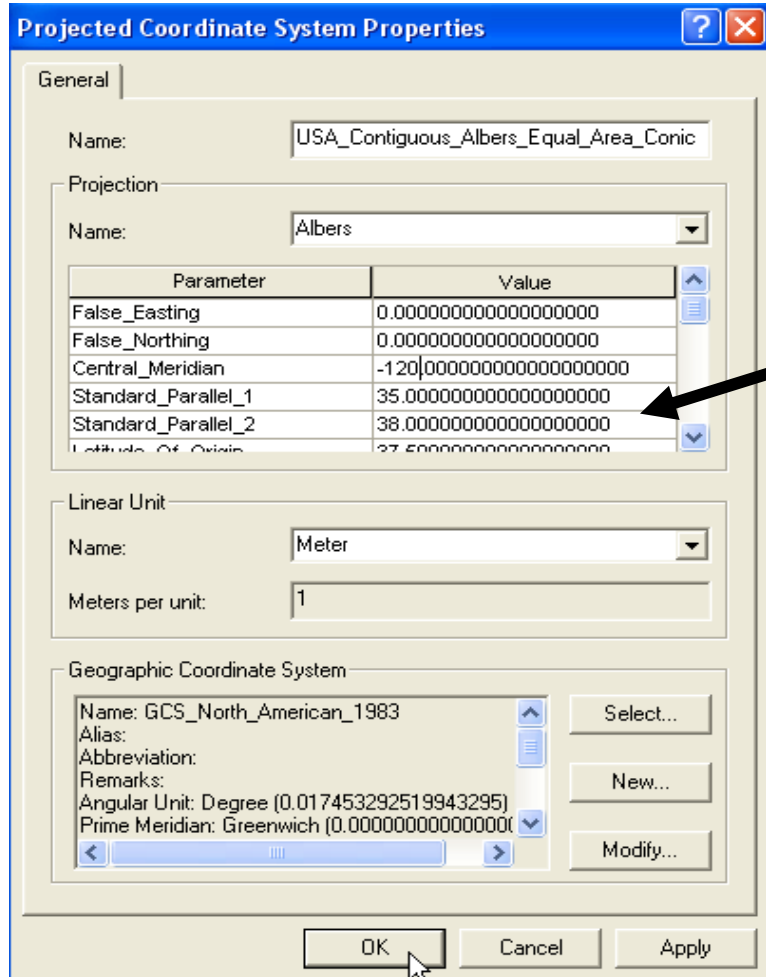


When the *Projection Properties* window opens set the *Central Meridian* to *-120*. (note the negative sign), the *First Standard Parallel* to *35*, and the *Second Standard Parallel* to *38*. Click *OK*.

Note the change in the shape of California.

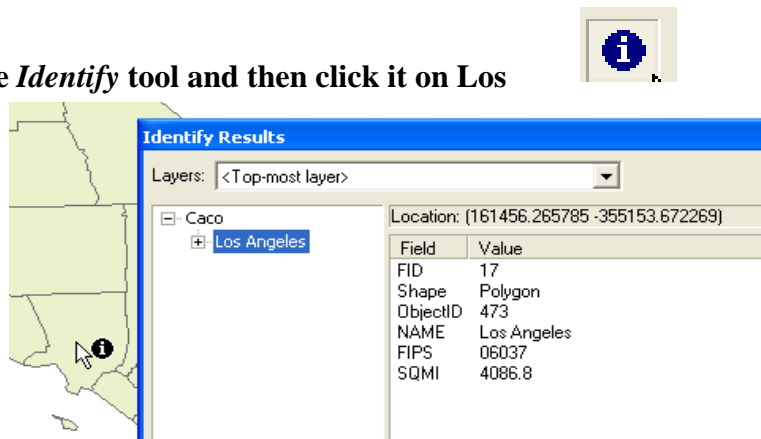
FRUSTRATION ALERT!!

This would be a really good time to save your *ArcMap* project. Select the *File* menu and save the *mxd* file in your working directory. Should *ArcMap* fail, you can open the program at this stage of completion. Note that once you have created this map file you should not move it or the map layers.



11. From the *Tools* menu select the *Identify* tool and then click it on Los Angeles County.

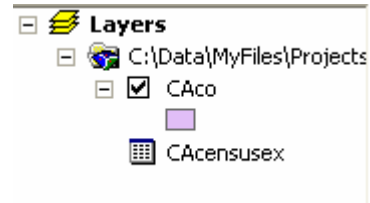
The *Identify Results* window right will open listing the attributes of the area.



Note what variables come with the California County boundary file. Check the form of the geographic ID of the county and its name (*GEOID2*). Fortunately, the Census Bureau in its files does create unique FIPS code IDs for the counties as shown here.

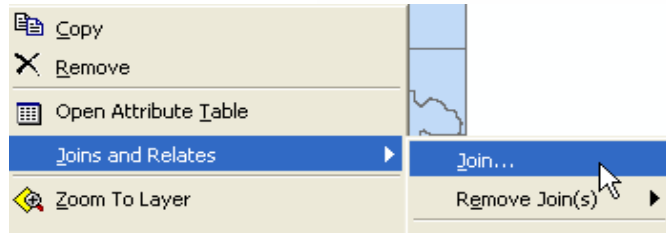
C. Importing Data into the Map

1. Click on the *Add Data* icon and add the *CAcensusEx* file. Note the icon in the *Table of Contents* for a data table.

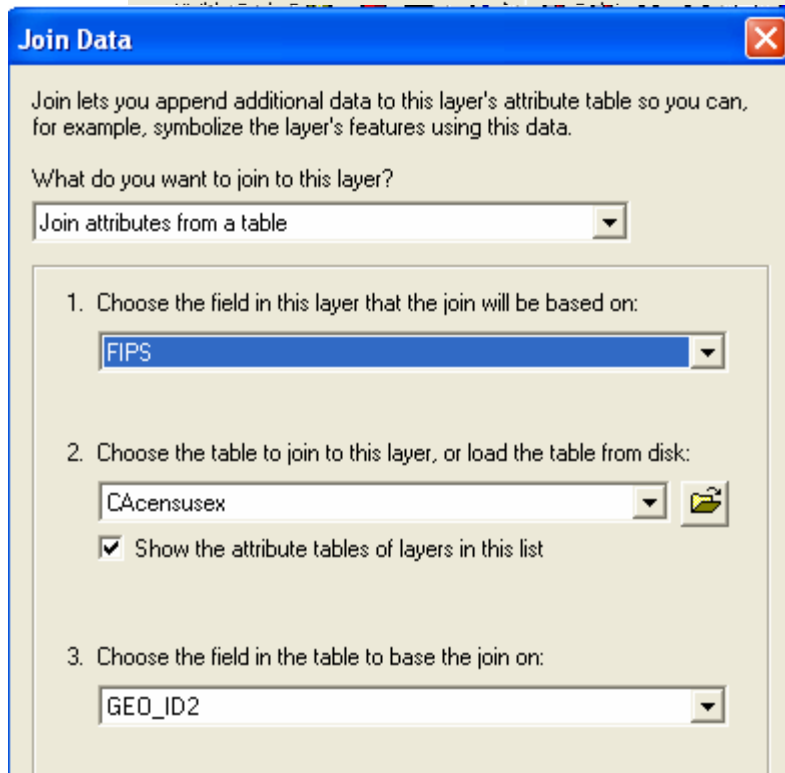


You will need to join this data table to the attributes of the boundary file in order to map the information.

2. Right-click on the *Caco* layer in the *Table of Contents* and from the popup menu select *Joins and Relates > Join*.



3. From the *Join Data* window make sure that *FIPS* (your ID field in the boundary file) is selected for the first join field at 1.



Next at 2, select the name of your data table, *CAcensusex*.

Then at 3 select the *GEO_ID2* variable from the data table. Then click *OK*. Say *Yes* to creating an index.

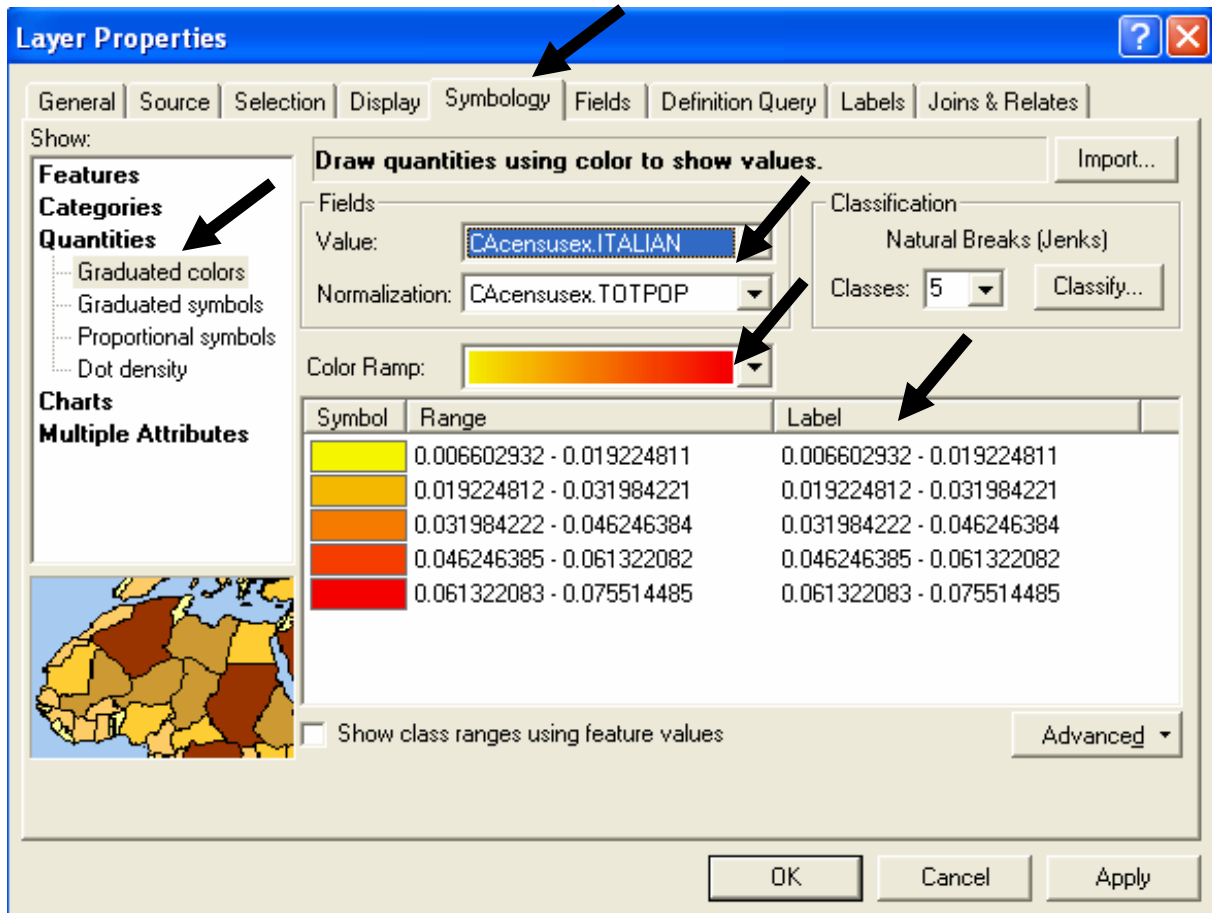
4. Again right-click on the *Caco* layer and select the *Open Attribute Table* option. Scroll to the right to see the new appended variables. Note that if you see the word *<null>* that there was a problem with the join.

Attributes of CAco							
CAco.STATE_FIPS	CAco.CITY_FIPS	CAco.FIPS	CAcensusex.OID	CAcensusex.GEO_ID2	CAcensusex.GEO_NAME	CAcensusex.TOTPOP	
06	093	06093	46	06093	Siskiyou County, California	44301	
06	015	06015	7	06015	Del Norte County, California	27507	
06	105	06105	52	06105	Trinity County, California	13022	
06	089	06089	44	06089	Shasta County, California	163256	

Close the *Attribute Table* when satisfied.

D. Making the Map

1. Right-click on the *Caco* layer and choose the *Properties* option. When the window below opens, select the *Symbology* tab.
2. In the left-side window click on the *Quantities* label and note the four map types shown. Make sure *Graduated colors* is chosen.

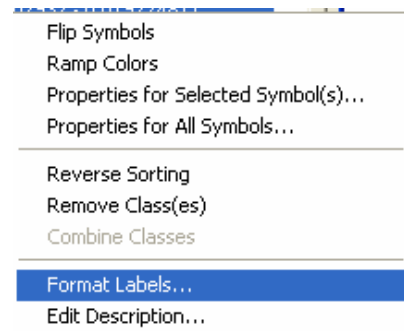


3. In the *Value:* window select the *Italian* variable and in the *Normalization* window select the *Totpop* variable. This will create a proportional value that can be converted to a percent.

A default set of 5 classes with a default color ramp will appear.

4. Click on the triangle to the right of the *Color Ramp* window and then select the yellow to red color ramp from the popup list.

5. Note that all the class labels are small decimals. To convert these to percents with fewer decimal values right-click on one of the label values. Select the *Format Labels...* option.



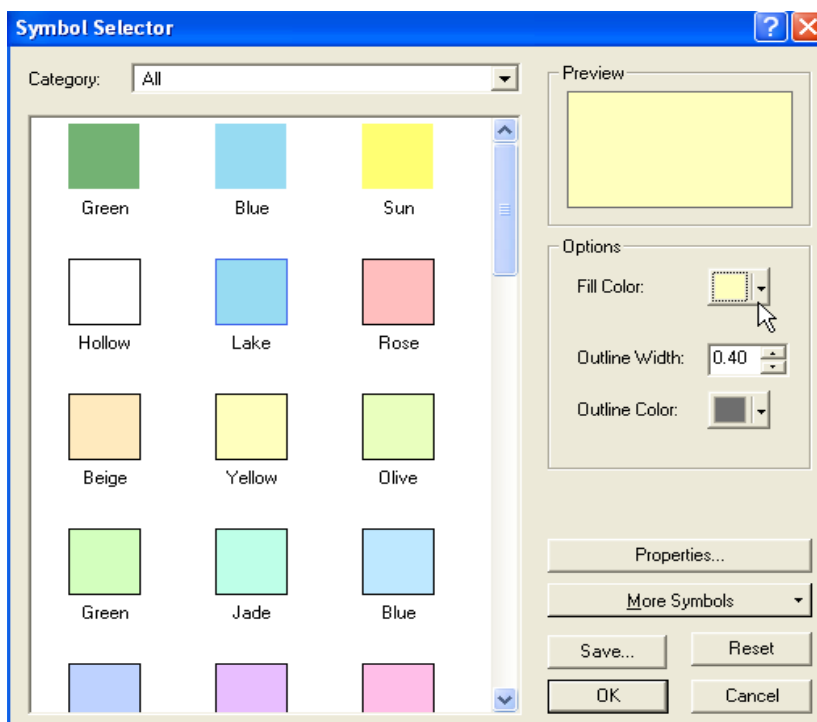
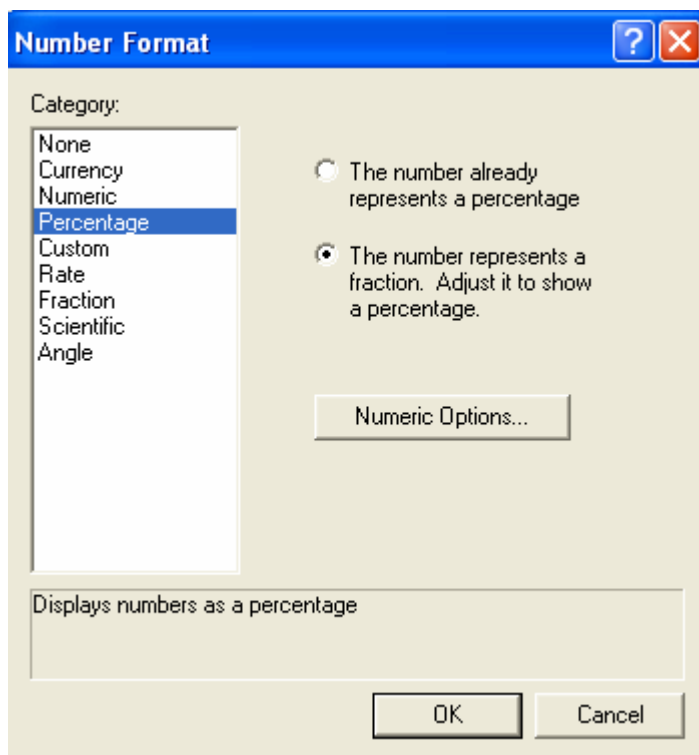
6. In the *Number Format* window select the *Percentage* category. Then click the button for *The number represents a fraction*.

7. Click the *Numeric Options* button. Set the number of decimals to 1 and then click *OK*.

Click *OK* a second time to see the map.

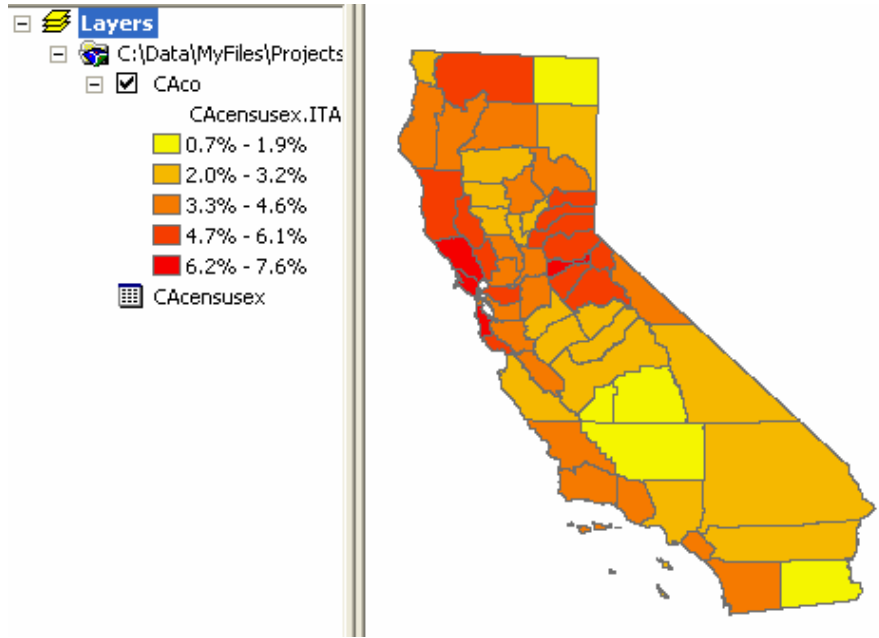
Also in the *Properties* window above you can change the number of classes, the classing method, the color ramp, and you can manually re-label the classes under the *Label* column.

8. Double-click on the yellow color symbol in the *Table of Contents* to bring up the *Symbol Selector* window. Here you can change its fill shade to a different shade or the stroke color and width. Change the yellow to a lighter shade and then repeat the step for a few of the other lower classes. When done, click *OK* twice.



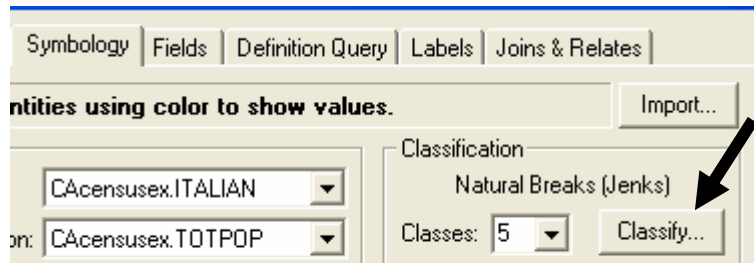
9. When the map at right of Percent Italians appears, look at it and the class values to see if it seems reasonable. In other words, check your work.

Any problems?



10. Again right-click on the *Caco* layer and select the *Properties* option. From the *Symbols* tab window select the *Classify* button.

In the *Classification* window you can modify the classes in various ways.

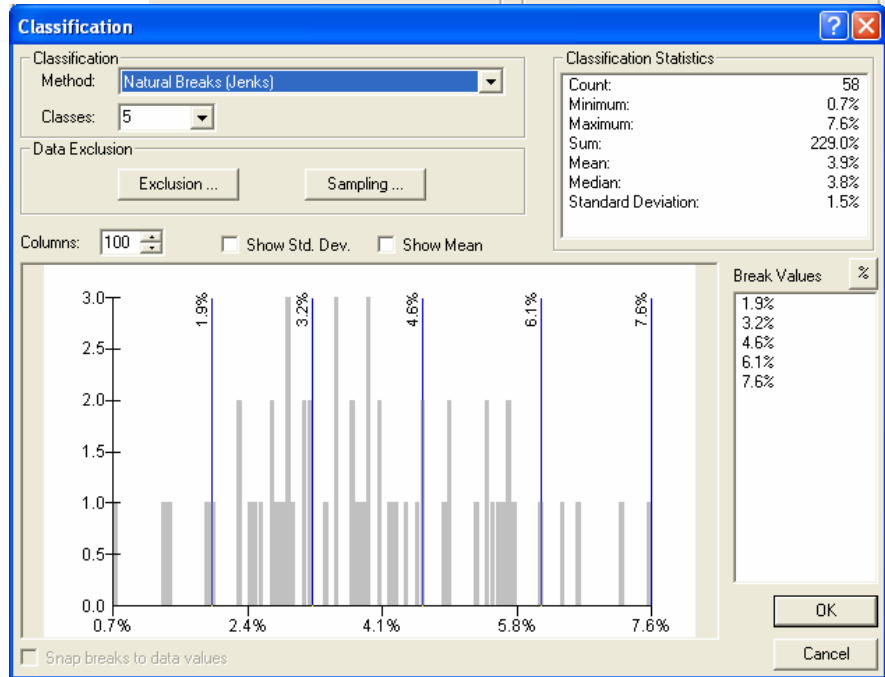


11. From the *Classification Statistics* window (right) what is the count of counties? _____

What is the minimum percentage? _____

What is the maximum percentage? _____

What are the number of classes? _____



Look at the current class break values and the shape of the distribution in the frequency diagram. You also may drag the blue lines if you want to manually shift the class breaks.

Using the *Exclusion* feature you could exclude values that fail to meet a chosen criteria. For example you could eliminate any counties with percentages less than 0.1 or with fewer than 1000 persons. On the map these could be assigned a unique color.

12. Close the *Classification* window.

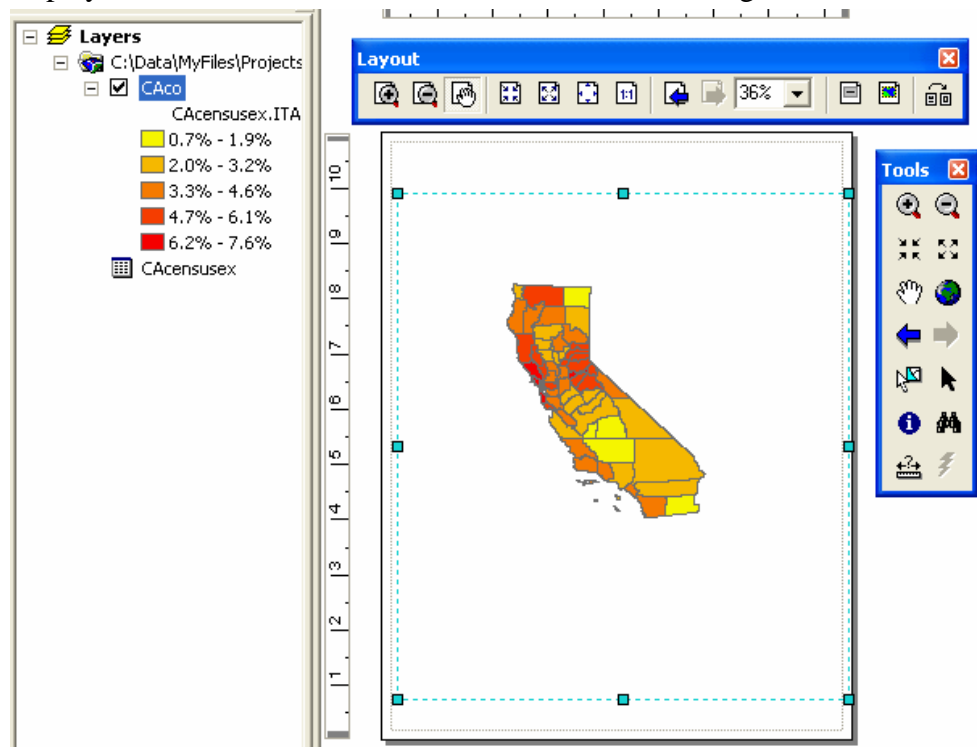
E. The Layout Window

ArcGIS provides two views of your map, a *Data View* and a *Layout View*. The former is used to compose a particular map and only one *Data View* can be displayed at a time. The latter is used to prepare a map for publication and might display several data views, plus a legend, scale, and other graphic elements.

1. From the *View* menu select *Layout View*. From here you can compose your final map for printing.

When the *Layout* window opens you will note the map in the *Data View* is displayed inside a dashed line. This sets the map display size and it can be changed by dragging its handles. Outside the dashed line are two additional lines. The first indicates the printing area of the output device and the second indicates the size of the paper the map will be printed on. On the far right of the display is the old *Tools* menu. These can be used to enlarge, reduce, or move the map within its data frame. You should NOT use these tools to zoom or move the layout page since they act on the map itself.

To move around the *Layout* window you will notice another set of tools designed for it at the top of the page above. Remember that because menus can be dragged that they may be in different locations on your machine.



2. Use the original *Tools* menu and *Zoom in* tool to size the map to fit nicely within the page.

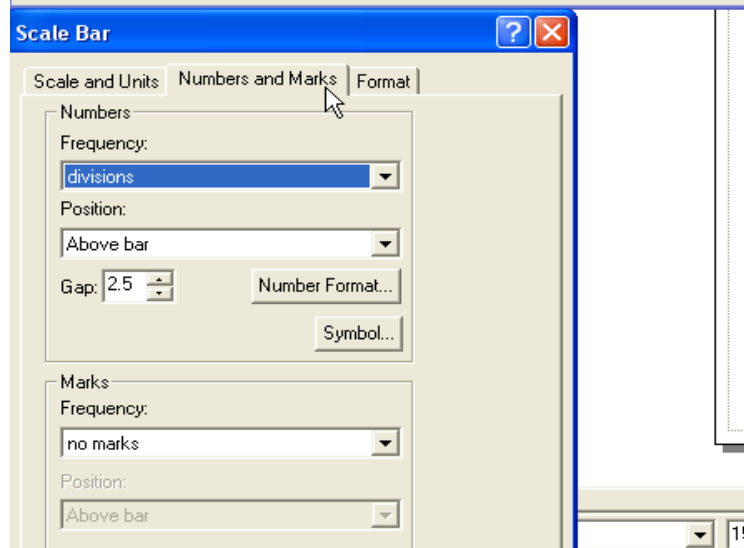
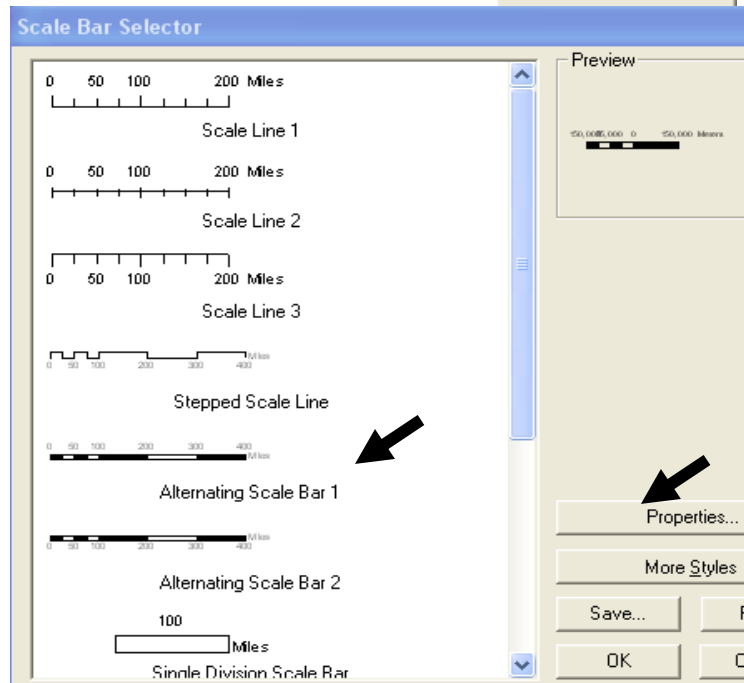
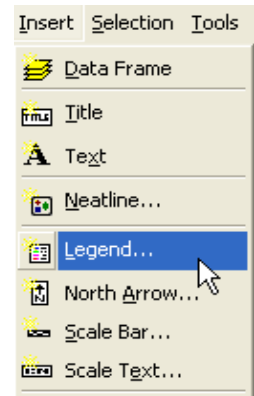
3. From the top menu select the *Insert* menu and and the *Legend* option. A *Legend Wizard* will present you with a number of design options, but for now accept all the defaults.

You can make some changes later by double-clicking on the legend itself. Basically you can add a box or background, change the nature of the symbol boxes, and determine what items are included. Not every map symbol needs to be in the legend and some would argue that the subtitle *Legend* is redundant.

4. Now drag a box over the area on the map page where you would like to place a legend.

From the *Insert* menu you may also add a north arrow and a bar scale. However, remember that on intermediate and small scale maps a north arrow is not appropriate since the direction varies over the map.

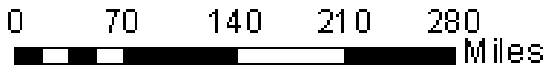
5. Select the *Insert* menu and the *Scale bar* option. Choose *Alternating Scale 1* and then select the *Properties* button. Click the *Numbers and Marks* tab and set the value to *Divisions*.



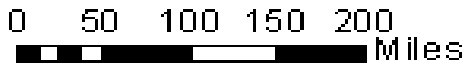
6. Now select the *Divisions and Units* tab. Set the *Division Units* to *Miles* and de-select the button to *Show one division below zero*.

Click *OK* to return to the main *Scale Bar Selector* window and then *OK* again to see the bar scale on the map.

It is quite possible your bar scale will come out similar to that shown. Unfortunately these odd break points are not as helpful as they should be. A bar scale should have evenly divisible units from which interpolation is not an effort. You will notice that the first break would be at 17.5 miles.

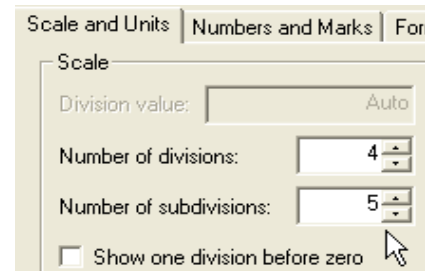
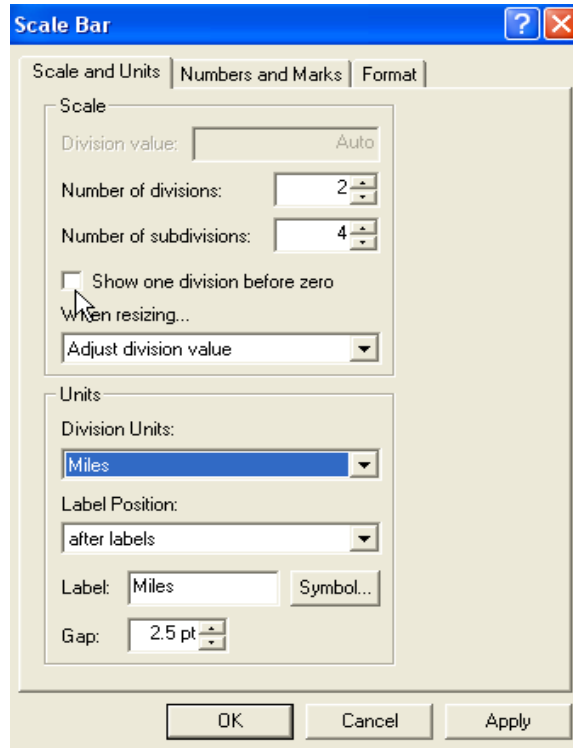
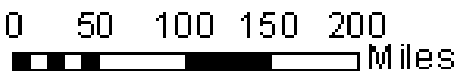


To fix this problem you can double-click on the bar to adjust it or, more simply, drag one of its handles to the left or right until the units are evenly divisible.



However, the problem with this bar scale is that there are four subdivisions at the beginning and so the breaks will occur at odd numbers.

By double-clicking on the bar scale and selecting the *Scale and Units* tab, the number of subdivisions can be set to 5 so that the subdivisions are even between 0 and 50.



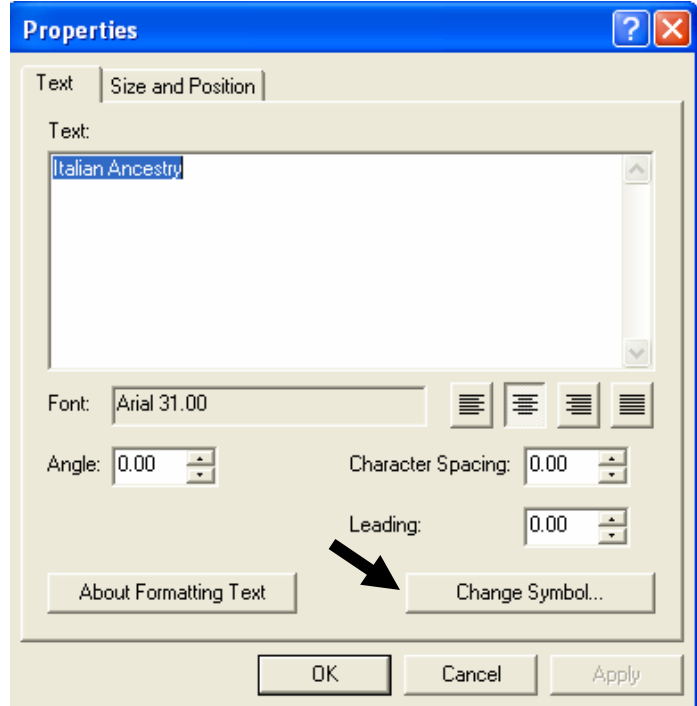
7. Select the *Insert* menu and the *Title* option. *ArcGIS* will open a box at the top of the map in which you can type your title. For this map it is *Italian Ancestry*.



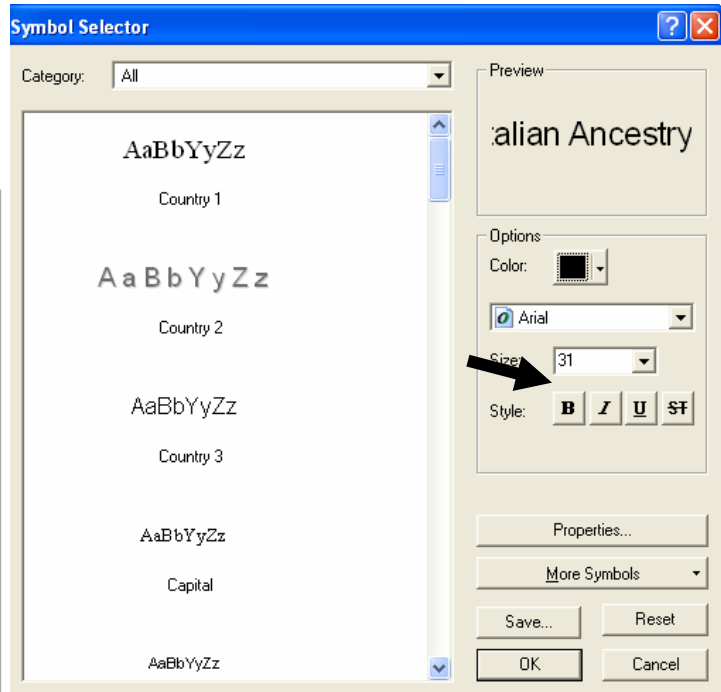
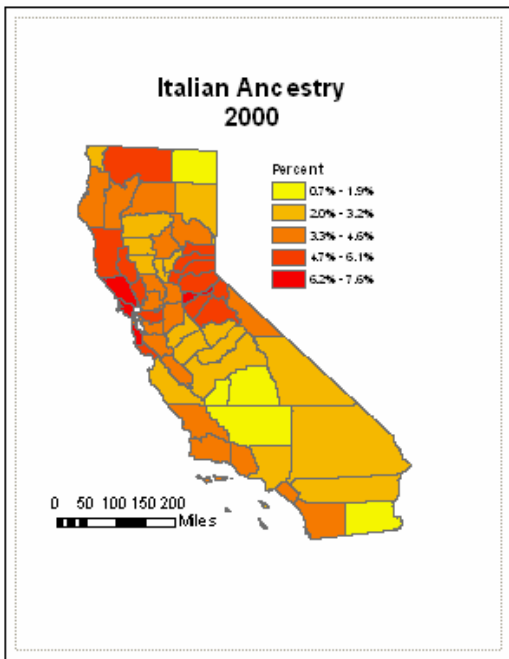
Italian Ancestry

To change the character of the title double-click on it to open the *Properties* window. Then select the *Change Symbol* button at the bottom. (See both right.)

8. Change your text (Font) to bold by clicking on the *B* button. Then click *OK* twice to see the result. Note you can make some other cosmetic changes to the legend labels by clicking once on their names in the *Table of Contents* and retyping a new value there.



9. Your map is now done. Print a copy for yourself and look at other census variables if you wish.



F. Querying the Map

One of the significant applications of software like ArcGIS is the ability to perform a variety of queries and analyses with one or more sets of data. This is generally beyond the scope of this module, but a few simple queries and tabulations of the data can be useful.

1. Select *View > Data View*. From the *Tools* menu select the *Identify* tool and click on one of the counties with a high percentage of Italians. A window of attributes will pop open. Look over the different variable values for the county.

Unfortunately, the list does not include the percent Italian which was computed in the *Properties* window. To work with this information we would need to create a new field in the *Attribute Table* and then populate it with the calculated percent Italian.

2. Close the *Identify Results* window. Then select *Selection > Select by attributes*..

From this window we can write a query to select a subset of counties from the map.

Double-click on “*CAcensusex.ITALIAN*” variable.

Click on the \geq button.

Enter *10000* at the end of the statement as shown right.

Click *Verify* to check your syntax and then click *OK*. All counties with 10,000 or more Italians will be highlighted. In this way we can see not only where the percentage of Italians is high, but where large numbers are involved. Note we could select counties based on other variables as well.

3. Close the *Select by Attributes* window. Note that in *ArcGIS* you can export selected geography into a new layer or the data into a new spreadsheet.

The image shows two overlapping windows from ArcGIS. The top window is titled "Identify Results" and shows a table of attributes for a selected feature named "Napa". The bottom window is titled "Select By Attributes" and shows a list of variables with "CAcensusex.ITALIAN" selected. The query text in the bottom window is "SELECT * FROM CAco_CAacensusex WHERE: 'CAcensusex.ITALIAN' >= 10000".

Field	Value
CAco.FID	12
CAco.Shape	Polygon
CAco.ObjectID	27
CAco.NAME	Napa
CAco.STATE_NAME	California
CAco.STATE_FIPS	06
CAco.CNTY_FIPS	055

Method: Create a new selection

Variables list:
"CAcensusex.IRISH"
"CAcensusex.ISRAELI"
"CAcensusex.ITALIAN"
"CAcensusex.LITHU"
"CAcensusex.NORWEG"
"CAcensusex.POLISH"

Query: "CAcensusex.ITALIAN" >= 10000

4. Right-click on the *Caco* layer in the *Table of Contents*. Then select *Open Attribute Table* from the popup menu.

Note that in the *Attribute Table* that the selected counties are highlighted. At the bottom of the table you can see that 24 of the 58 counties have been selected.

5. On the bottom of the table click on the *Selected* button. Only the selected counties will be displayed. Then click the *All* button.

Attributes of

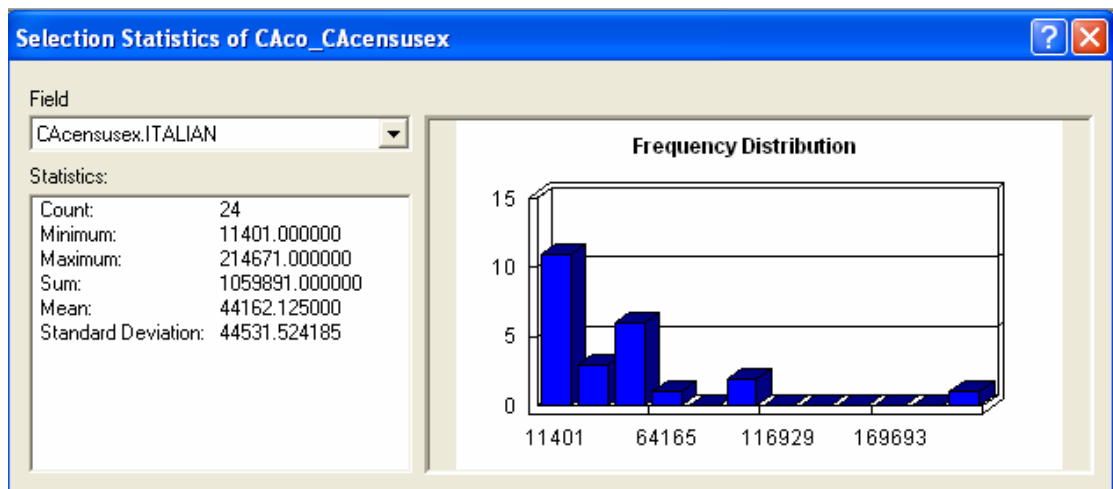
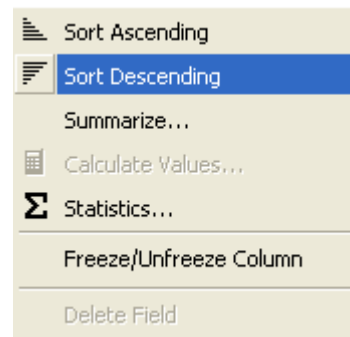
	CAcensusex.ISRAELI	CAcensusex.ITALIAN	CAcensusex.LITHU	CAcensus
	0	2525	56	
	0	619	16	
	0	461	8	
	0	6446	131	
	24	5851	110	
	0	1565	38	
	4	4247	50	
	0	766	5	
	253	17654	584	
	0	2879	74	
	0	449	0	
	28	5925	177	
	0	6861	136	
	124	34632	700	
	27	14624	404	
	511	53527	1107	
	643	49120	1659	
	553	32433	1162	
	348	48846	787	

Record: 1 Show: All Selected Records (24 out of 58 Selected.)

6. Right-click on the top of the *CAcensusex.ITALIAN* column. From the popup menu select the *Sort Descending* option. Look over the column. What counties have the most Italians?

7. Again right-click on the top of the *CAcensusex.ITALIAN* column. Then select the *Statistics...* option.

Note the resulting statistics are relevant to your selected subset of counties. When done, close the *Selection Statistics* window.



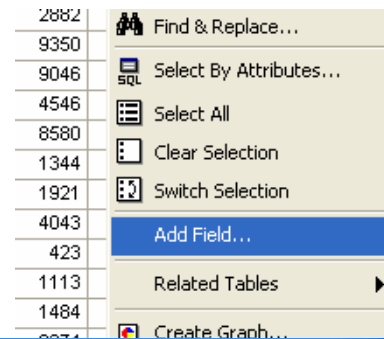
8. Select *Selection > Clear Selected Features*.

9. Another way to select features is to use the *Select Features* tool from the *Tools* menu. You may click and drag a rectangle over the map or hold the shift key and click on a series of counties to select them. Use *Selection > Clear Selected Features* to release the selected counties.

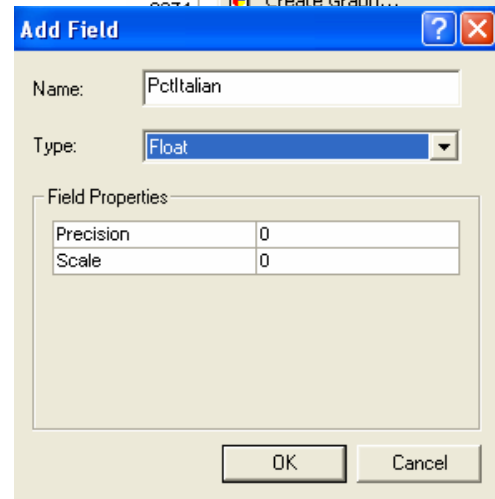


10. Right-click on the *Caco* layer and select the *Open Attribute Table* option to make the attribute table visible.

11. At the bottom of the *Attribute Table* click the *Options* button. From the popup menu select the *Add Field...* option.



12. In the *Add Field* window enter the Name of *PctItalian* and set the *Type* to *Float*. Leave the *Precision* and *Scale* set to 0. The former is the number of characters and the latter is the number of decimals. *ArcGIS* can determine these values.



Click *OK* and note the new field will be added at the end of the list of boundary file attributes. (i.e. ahead of the joined variables)

13. Look for the *Editor Toolbar* symbol at the top of the screen. If it is not visible click *Tools > Editor Toolbar*.



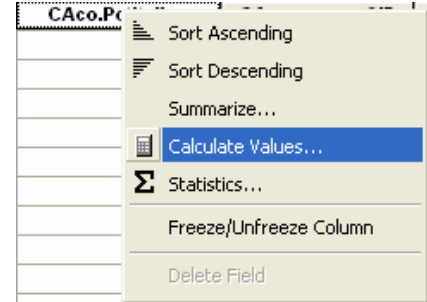
14. Locate the *Editor* menu and select the *Start Editing* option.



In the *Attribute Table* the fill behind the column labels will turn white.

Attributes of				
	CAco.FIPS	CAco.PctItalian	CAcensussex.OID	CAcensussex.GEO_ID2
	06047	0	23	06047
	06069	0	34	06069
	06051	0	25	06051
	06109	0	54	06109
	06007	0	3	06007
	06091	0	45	06091

15. Right-click on the top of the new column (which is currently filled with 0s) and select *Calculate Values* from the popup menu.



16. In the *Field Calculator* window double-click on the item *CAcensussex:ITALIAN*. Note spaces are required between items in the equation.

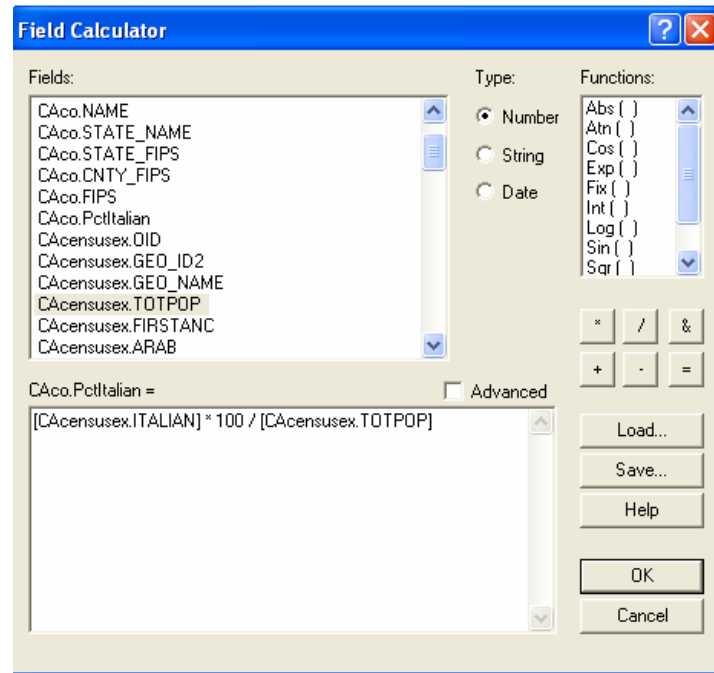
Select the * symbol.

Enter a space and type 100.

Enter a space and select the / symbol.

Enter a space and double-click on *CAcensussex:TOTPOP*

Click *OK* to run the calculation. The new percent Italian field will be populated.



17. Again locate the *Editor* button and select the *Stop Editing* option. Save your edits.

You now have a permanent value for percent Italian that can be sorted and processed just as the count of Italians was.

G. Exercises

1. Select one of the other ancestries in the *Attribute Table* and map it by percent of total population. Create a new field with the actual percentages. In what counties is the group especially concentrated? In what counties are there large numbers? Can you offer any explanation about why the group is located where it is? Consider time of initial settlement, and particular occupations.
2. Make choropleth maps for an ethnic group of both the number and the percent of total population. Compare the patterns on the two maps. Are they similar? Why do you think they might be different?
3. Compute the correlation of two census variables in a statistics program. Export both the correlations and the residuals making sure that you have useable geographic identifiers as well. Map the two variables, the correlation values, and the residuals. Describe the patterns.
4. Calculate the percent ethnic and make choropleth maps for counties and census county divisions. How do the two patterns differ? Is either one more helpful or descriptive?