ENVR/GEOG 260 – Introduction to Geographic Information Systems
LAB 3 – Working with Shapefiles and Databases

Due date: Thursday, October 31, 2002

NOTE: This lab is fairly extensive, but there are many instances where I ask you to address specific items or questions in your lab write-up. **These will be denoted by bold face and underlined text**. Be sure that you address these items in your write-up, as this is where much of your grade will come from!

As in previous labs, you will be using the Getting to Know ArcGIS Desktop as a primary reference book for many of the functions in this exercise.

Introduction

Database management is an essential tool that must be mastered in order to effectively work with spatial data. Many of the database operations can be performed directly within the GIS, and ArcGIS has many the basic and advanced database SQL operations built in to the query system. The purpose of this lab is to allow you to become familiar with how to manipulate and query database information associated with shapefiles. You will be using the data that can be obtained off the Blackboard site for this class in the "Course Documents" section (which will also contain the PDF file of this document).

Data

The data are in "zipped" format, which means you need to "unzip" the data to your Zipdisk or network drive prior to starting the lab. I suggest you create a folder specifically for this Lab 3 assignment, and place the data files into that folder. (Proper organization in advance of starting any project will make things go much more smoothly later in the course!) The data are included in a file called "lab3data.zip".

The basic data unit in ArcGIS is called the "shapefile," which is actually a term used to describe a combination of files that include the spatial and attribute information you are seeking to analyze. The basic shapefile structure contains a minimum of three files:

- the shapefile itself, containing the spatial data (<filename>.shp)
- an attribute database file (<filename>.dbf)
- an index file that helps to "glue" the whole package together (<filename>.shx)

These files must be in the same place for ArcGIS to be able to use the data. Also, it is very unwise to attempt to make changes to these files outside ArcGIS (for example, with a text editor or a database program), since ArcGIS is expecting a particular structure to the files.

You should have the following files in your folder after uncompressing the data file:

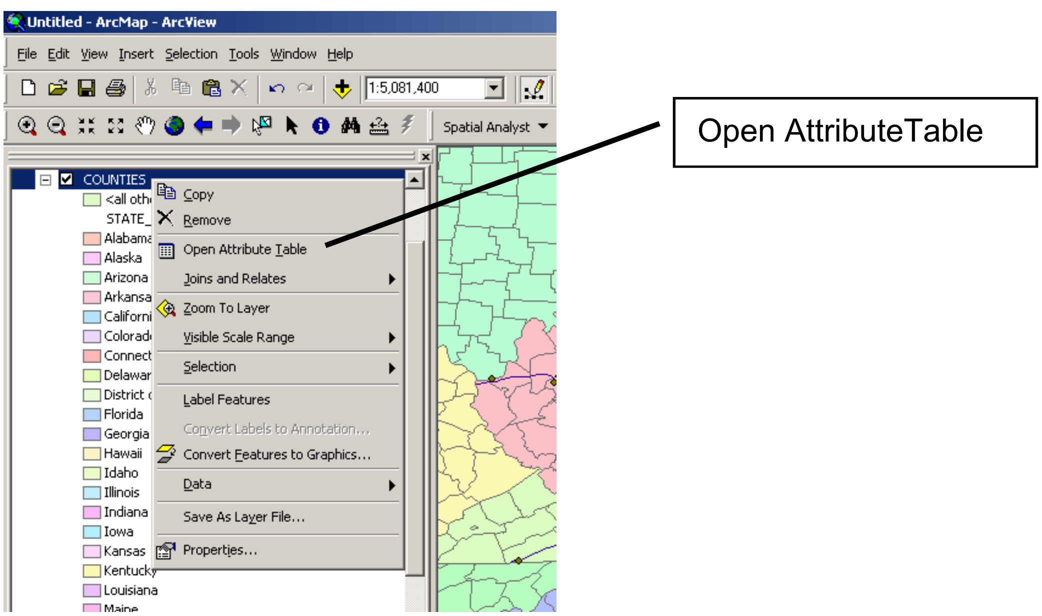
- counties.shp (and associated files) (United States county-level data)
- cities_atl.shp (and associated files) (mid-Atlantic states major cities)
- roads_atl.shp (and associated files) (mid-Atlantic states major highways)
- ce1980t.dbf (Census data from 1980) and ce1980t.txt (metadata)
- lab3.pdf (this file)

LAB PROCEDURES

Part 1. Selecting and subsetting data

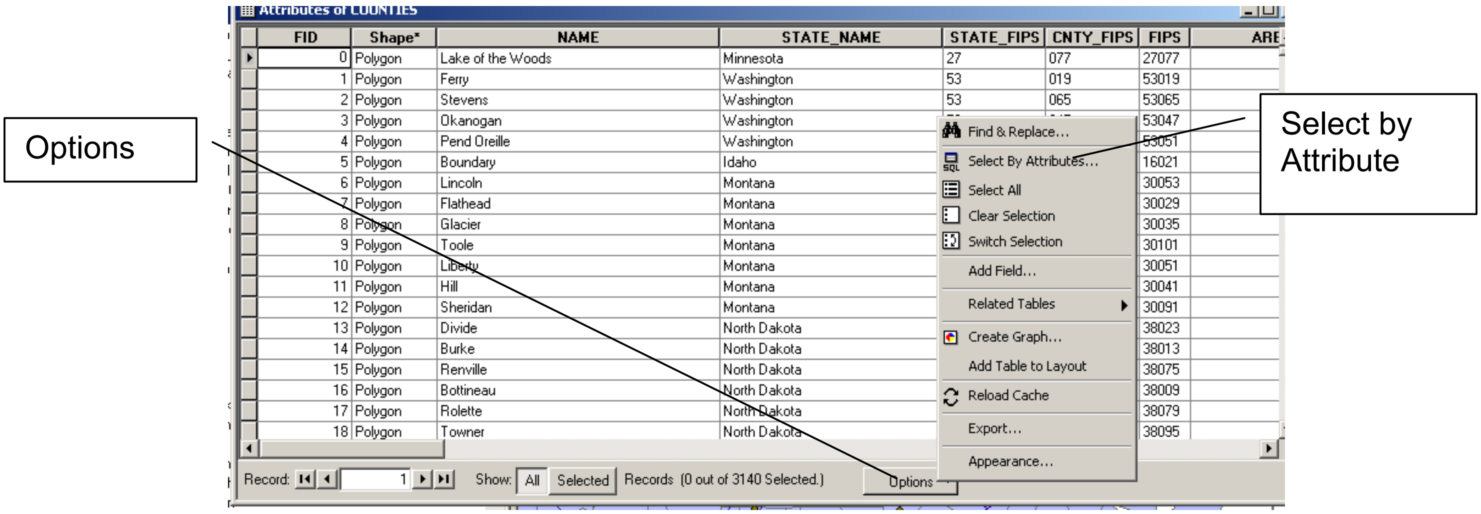
Start a new project, and create a view that contains the county-level U.S. data ("counties.shp"). You will notice that the data encompass the entire United States, including Alaska and Hawaii, and for this reason Virginia is difficult to visualize. Use the zoom tool to zoom into the region around Virginia. Double-click on the shapefile name in the table of contents (TOC) of the view, and change the classification scheme to show each state as a unique value, using "STATE_NAME" as the classification field. (**Include a screenshot of the view in your report**.)

Look at the attribute table associate with this shapefile by right-clicking on the Counties layer and selecting "Open Attribute Table" table tool:

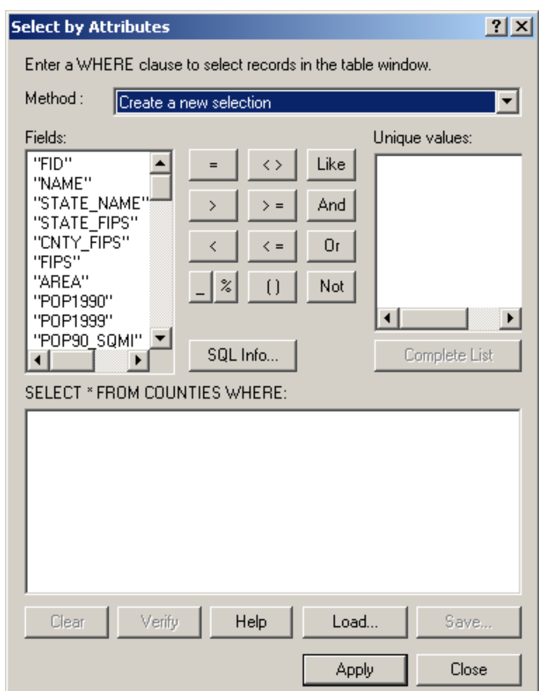


After doing this, you should see a spreadsheet-like table appear with the view, which contains all of the available attribute information associated with the shapefile. (The table is quite extensive, and contains a great deal of demographic information gleaned from previous censuses.)

Because the shapefile contains all the county-level information for the entire United States, it becomes rather cumbersome to use if you only wish to look at a few selected states. Let's select out the Middle Atlantic states by performing a "query", using the "select by attribute" tool located under "Options" in the attribute table:



You will get the following dialog box on your screen:



Notice that the fields contained in the attribute table are listed, along with a variety of mathematical and Boolean operators.

Select the following states from the database (use the instructions in GTKAD for reference):

Virginia
Maryland
District of Columbia
Delaware
North Carolina
West Virginia

(**<u>NOTE</u>**: As mentioned in class, it is very important that you use the proper syntax to select the records, otherwise you might not get the proper results! When in doubt, write it out by hand before entering it into the query box!)

When you have entered in the proper query, make sure you have selected “Create a new selection” under “Method”, and click “Apply”. **<u>What happens to the map when you do this?</u>** Examine the attribute table, scrolling through the entire table. **<u>What changes have occurred in the appearance of the table?</u>**

Notice that the selected records are scattered throughout the attribute table. In order to make it easier to examine the selected records, you can view only the selected records by clicking “Selected” next to “Show” at the bottom of the table. **<u>What do you notice about the table’s appearance at this point? Has this had any impact on the appearance of the map in the view?</u>**

With the mid-Atlantic states selected, you can add these to your TOC as a new layer by right-clicking on the “Counties” layer, going to “Selection>Create Layer from Selected Features”. This will place the feature subset at the top of the TOC. (NOTE: This does not create a new shapefile; if you want to create a shapefile from the selected set, right-click on the layer, go to “Data>Export date” and give it a new name; this will permanently save the selected file for future use.) Give it a new title, and open its attribute table. **<u>What differences do you see between the new table and the table from the original “Counties” table, in terms of entries?</u>** Turn off the original Counties layer, and apply a unique classification Symbology to the new layer. Zoom to the new layer by right-clicking on the layer, and clicking on “Zoom to Layer”.

<u>Part 2. Adding and manipulating fields to the database</u>

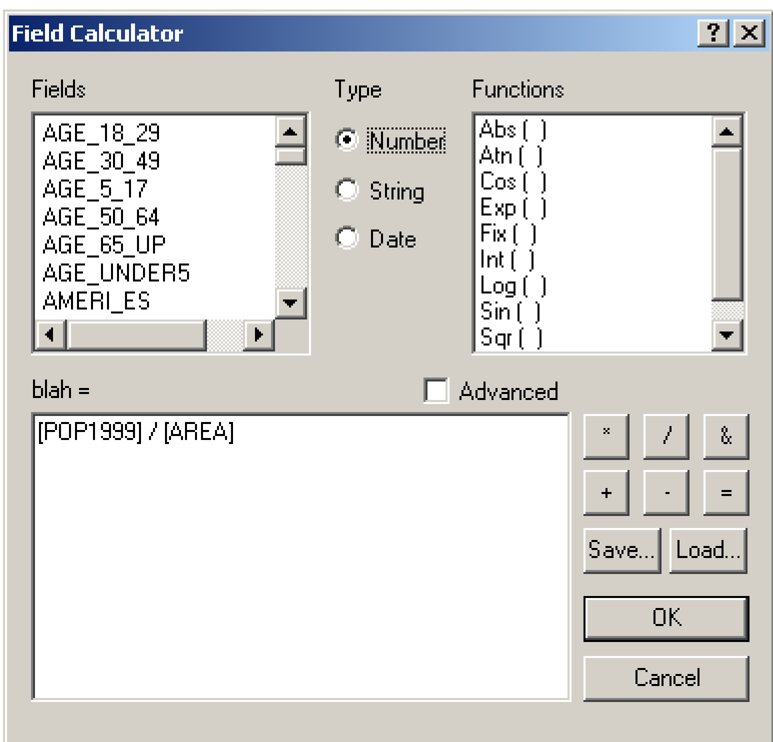
One of the more useful features of the ArcGIS system is the ability to directly manipulate the attribute table, in order to create derived measures of spatial variability. This is done by activating the table to be modified, and adding additional fields to the database.

Activate the subset layer and activate the attribute table for that layer. Enlarge the table so that you can better see the field titles. (Note that by using the horizontal scroll bar you can easily see the fields for the entire table.) For this next exercise, we wish to see how population density for each county in our study region has changed from 1990 to 1999.

You will note that the table already has the county population density (measured in people per square mile) calculated for 1990 ("Pop90_sqmi"), but the population density for 1999 is missing. Therefore, we must calculate this figure ourselves. Also note that the population density for 1990 is accurate to only the nearest whole number. (This is important, since you want to be careful about doing arithmetic, especially division, on numbers that equal zero. If you find yourself with the potential for "divide-by-zero" situation, you will need to recalculate the field using a higher level of precision. For this exercise, you may assume that this is not an issue.)

Open the attribute table for the layer, and add a new field called "Pop99_sqmi" (Options>Add Field…), providing the name of the field and other pertinent information (for this example, set the precision to three (3) decimal places). Note that the new field is added to the end of the table horizontally.

With the field selected (by clicking on the field title), right-click on the field heading and choose Calculate Values. This brings up a dialog box very similar in looks and operation to the Select by Attribute tool:



Using the same procedure as with queries, enter the appropriate formula:

Pop1999 / Area

This will calculate a value for the entire series of records in the database. Once you have completed this, it is a good idea to save your work.

You can now recast the layer in terms of the population densities. (For convenience, turn off all the layers from the map <u>except</u> the layer containing your 6-state map.) Double-click on the layer TOC entry, and change the symbology to "graduated color," and choose an appropriate color scheme. Choose "pop90_sqmi" as the classification field. Choose 10 classes, and choose "quantile" as the type of classification method. After examining the class breakpoint, change the scale to a better "rounded" scale with more intuitive numbers. Also, take this chance to give the layer a new name, something like "1990 Population Density"). **<u>What happens in the view, both in the map and in the TOC? (Be sure to include a screenshot in your write-up.)</u>**

For ease of comparison, make two copies of this layer in the map, and change one of the layers to display the 1999 population density. To copy a layer, right-click on the layer title and choose "copy". Next, right-click on the map title (usually "Layers" unless you have renamed it) and choose "paste layer". A new copy of the same layer will be placed on the map, which can be manipulated independently of the other layers. Rename this layer to something like "1999 Population Density", and display it using the "pop99_sqmi" field, and using the same color scale and divisions as you used in the 1990 population density map. **<u>Be sure to include a screenshot in your write-up.</u>**

<u>Compare the two population density layers. Do you notice any spatial patterns evolving in the map? What sorts of places (urban, rural, etc.) seem to be experiencing the most and least growth? Are the regions that are going down in population, relative to other areas? Why might this be the case?) How easy or difficult is it to see the changes using this direct comparison method?</u>

One of the most effective ways of gauging the amount of change in several related variables (such as those that are used to create the population density figures) is to create some sort of <u>*index*</u> that reflects temporal variability in the variables. In this case, an effective index to address the degree of change would be to find the difference between the population density figures, and then to plot the results. This can be done very easily by creating a new field in the attribute table and calculating the difference between the 1999 figures and the 1990 figures (i.e., pop99_sqmi minus pop90_sqmi). Perform this operation, using the same procedures as before (remembering to save your work as you go!) When this has been done, copy a new layer (as before), and change the plotted variable to the newly created field. Rename this layer to something appropriate (your choice), and create a new legend using 10 classes and a "blue to red" color ramp scheme. <u>In this case, rework the individual classes such that increasingly deeper shades of blue represent increasingly negative values of population density loss, and increasing shades of red show increases in density.</u> **<u>How does this method of showing population density change compare with the previous method? Are spatial patterns emerging? What sorts of explanations might account for the patterns?</u>** Add the "cities_atl.shp" and "roads_atl.shp" shapefiles as layers, and arrange the layers so that you can see the cities and roads overlaid upon the map. **<u>What sorts of physical features (cities, roads, etc.) might account for the observed changes? (Be sure to include this map in your write-up.)</u>**

Repeat this procedure, except this time, calculate the percentage change in population density going from 1990 to 1999. This can be calculated using the following formula:

(pop99_sqmi – pop90_sqmi)/pop90_sqmi * 100

This will yield numbers (in percent) that relate to the percentage gain or drop in population density from 1990 to 1999. **<u>Perform the same analysis as in the previous step, and readdress the same questions. In addition, discuss which method (actual numerical change versus percentage change) is more appropriate for the data being used here.</u>**

<u>Part 3. Joining external databases to a layer</u>

A very powerful capability of most GIS's is the ability to attach external databases to your spatial data, thereby increasing your query capabilities. This is typically done by "joining" the external database (often referred to as the "source" table) to the attribute table of your shapefile (called the "destination" table). **(Refer to Chapter 9 of GTKAD for details on joining tables, and be careful which file you use as the "source" and "destination" tables!)**

Included in your data set is a census file containing the 1980 census data. Add this table to your TOC, and join this data to the database file for your 1980 layer, using the county-level FIPS code as the common field between the databases. ("FIPS" stands for "Federal Information Processing Standard", the standard way of identifying political boundaries in Census data.) **<u>What type of database relationship (one-to-one, one-to-many, etc.) has been formed when these tables are joined in this fashion?</u>** Once this is accomplished, you will be able to access the information in that database file as if it were completely contained in the shapefile database file. **<u>Using the newly-added information, repeat the operations in the previous section to create maps showing the increase and/or decrease in population density from 1980 to 1990, again creating an appropriate color scheme for your legend. *Be sure to make your legends consistent between the two time periods, so that direct comparisons can be made between the decades.* What sorts of changes have occurred in the spatial variability of population density distribution between the 1980's and 1990's? Can you think of some reasons why these changes might be taking place? What might this say about the overall shifts in population during these timeframes, in terms of why people change locations? (As always, include a copy of the resulting map in your lab write-up.)</u>**