Data Exploration with ArcGIS

Working with Tables

Tutorial goals:

This tutorial will help you become familiar with table functions in ArcGIS, and learn to manipulate fields and incorporate data from other sources. You should also learn to perform structured query functions and carry out basic statistical calculations.

Tutorial tasks:

1. Attribute table functions:
   A. compute a new variable based on existing variables
   B. match data from other source

2. Save data
   A. save as a layer
   B. save as a shape

3. Query and map aggregation
   A. map aggregation- the dissolve function
   B. selection by attributes

Data and shapefile:

In this tutorial, we will use the state1990 data from the "extension" file that was used in tutorial 2, which contains the state shapefile and an Excel file of regional codes. If you did not complete tutorial 2 the file can be downloaded from Tutorials webpage, it is located under "Getting Started with ArcGIS2" in the data section. Use Winzip to uncompress the file. In a later section (adding data from a spreadsheet) you will use an excel spreadsheet file from the extension.zip file.

The shapefile State1990.shp is the state map of the United States with demographic data from Census 1990. There are fifty-one records in the file, which contains the following variables:

Variable name	Variable label
STATE_NAME	state name
STATE_FIPS	state FIPS code
SUB_REGION	division code
STATE_ABBR	state abbreviation
POP1990	population 1990
HOUSEHOLDS	number of households 1990
MALES	number of males
FEMALES	number of females
WHITE	non-Hispanic white population
BLACK	non-Hispanic black population
ASIAN_PI	Asian population
HISPANIC	Hispanic population
AGE_UNDER5	age below 5
AGE_5_17	age 5-17
AGE_18_29	age 18-29
AGE_30_49	age 30-49
AGE_50_64	age 50-64
AGE_65_UP	age 65 and above
NEVERMARRY	never married
MARRIED	married
SEPARATED	separated
WIDOWED	widowed
DIVORCED	divorced
HSEHLD_1_M	number of Male-headed households
HSEHLD_1_F	number of Female-headed households
HSE_UNITS	total housing units
VACANT	total vacant housing units
OWNER_OCC	number of owner occupied housing units
RENTER_OCC	number of renter occupied housing units
MEDIAN_VAL	median housing value

Starting ArcMap:

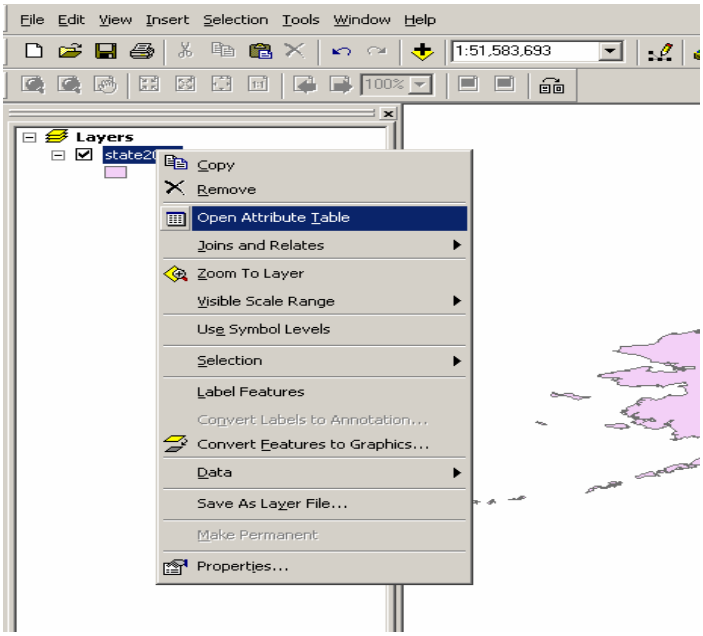
- Click the **Start** button on the Windows Taskbar.
- Point mouse to **All Programs**
- Point mouse to **Instructional**, then to **ArcGIS**, and click **ArcMap**.

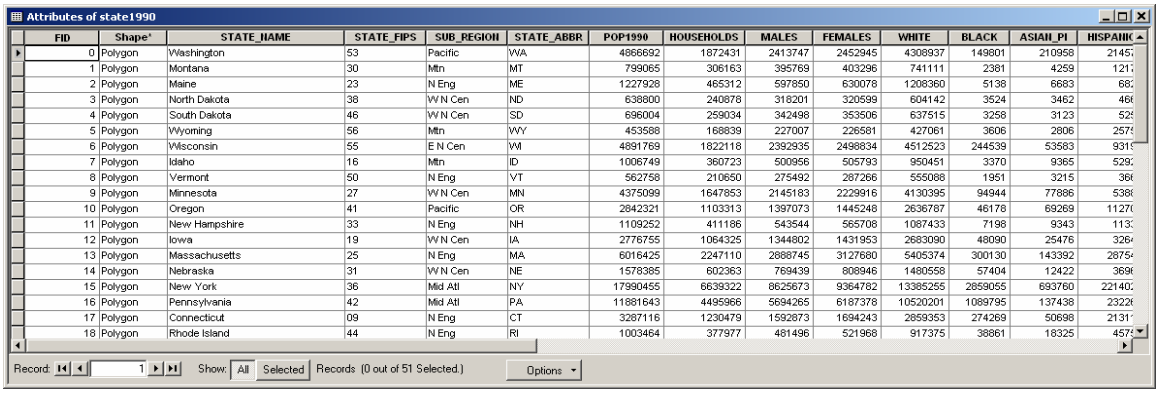
Click **Add Data** icon, then locate your file folder, **Select** the shape file and add to the map. (Check the projection and coordinate system for layer and for data frame.)

1. Attribute table functions:

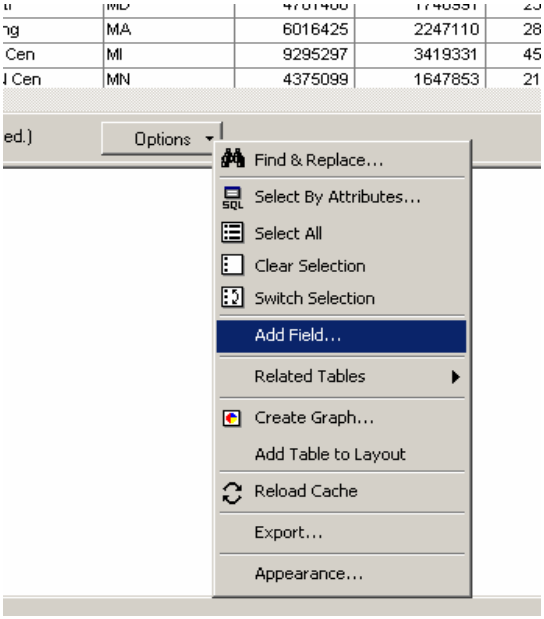
A. Add a new variable to the table

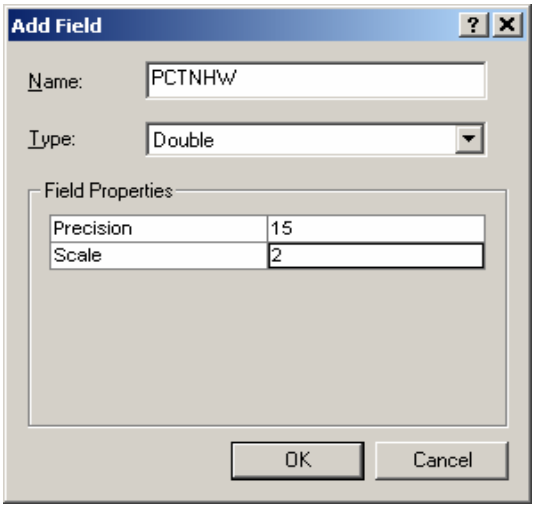
a. **Right click** the layer **state1990** in the **TOC** panel, select **Open Attribute Table**:





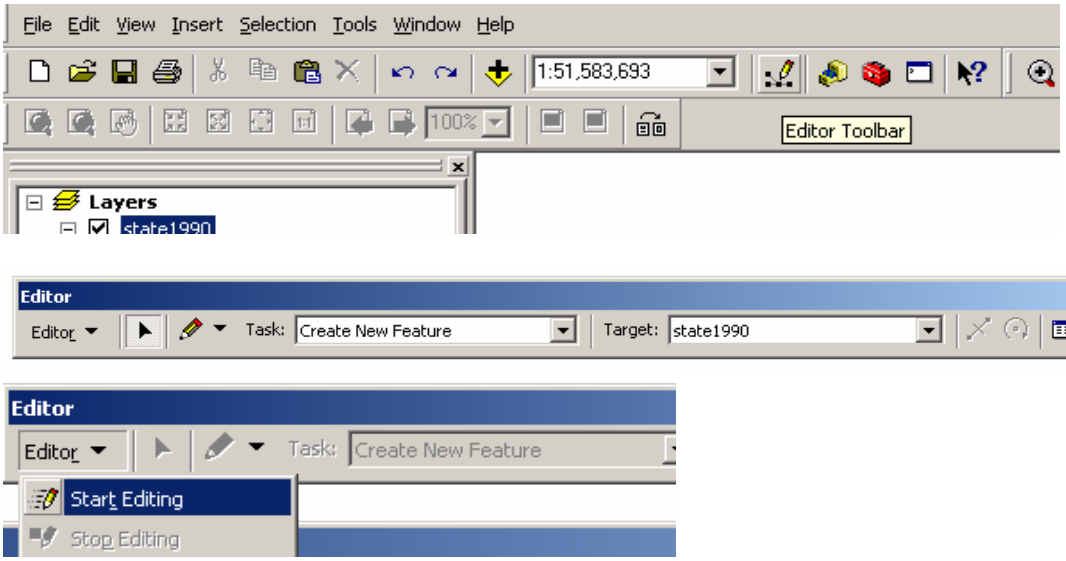
b. A window of the **Attributes of state1990** appears, **Click the Options** button at the bottom of the window, select **Add Field...** from the choices:



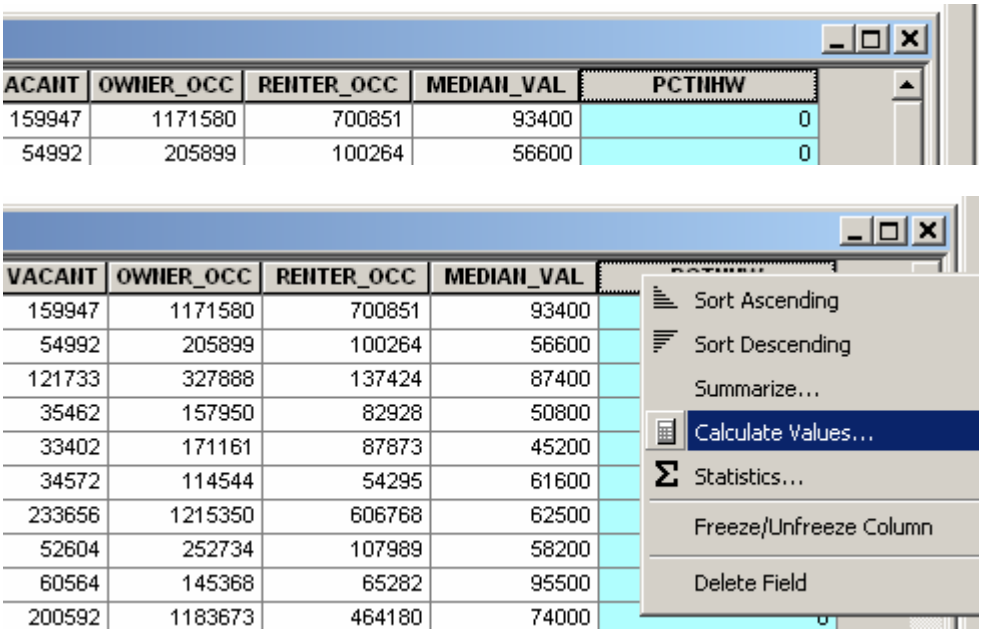


c. In the **Add Field** dialogue box, give name **PCTNHW** to the new field; specify **Type** as **Double**, **Precision** 8, and **Scale** 2. **Click OK**. A column should be added to the end of the table.

Click the **Editor Toolbar** button on the **main toolbar**. The **Editor Toolbar** appears, **Click** the **Editor** button, then select **Start Editing**:



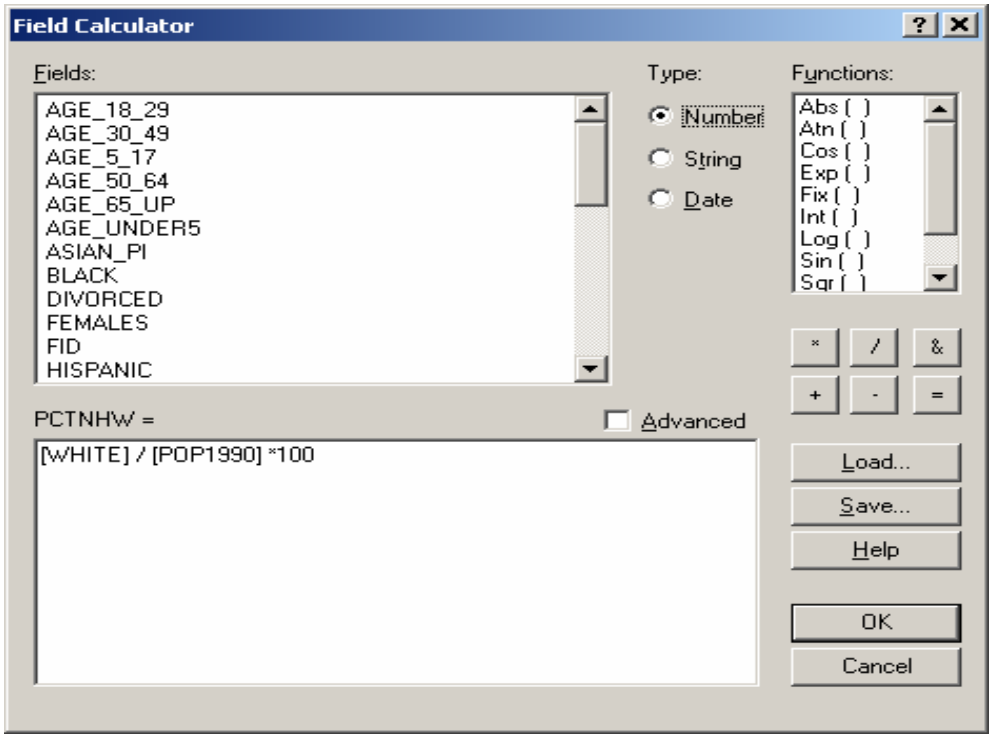
d. Go back to the **Attribute Table**, **Right Click** the newly created field **PCTNHW** at the end of the table > **Calculate Values...** (Notice the last option in this panel is **Delete Field**.)



ACANT	OWNER_OCC	RENTER_OCC	MEDIAN_VAL	PCTNHW
159947	1171580	700851	93400	0
54992	205899	100264	56600	0

VACANT	OWNER_OCC	RENTER_OCC	MEDIAN_VAL
159947	1171580	700851	93400
54992	205899	100264	56600
121733	327888	137424	87400
35462	157950	82928	50800
33402	171161	87873	45200
34572	114544	54295	61600
233656	1215350	606768	62500
52604	252734	107989	58200
60564	145368	65282	95500
200592	1183673	464180	74000

e. Compute percent of non-Hispanic white with the following formula in the **Field Calculator. Then Click OK.**



The new field should be populated with percentage values. **Click the Editor** button, then select **Save Edits, then Stop Editing**. Then **Close** the **Editor Toolbar.**

B. Add data from a spreadsheet

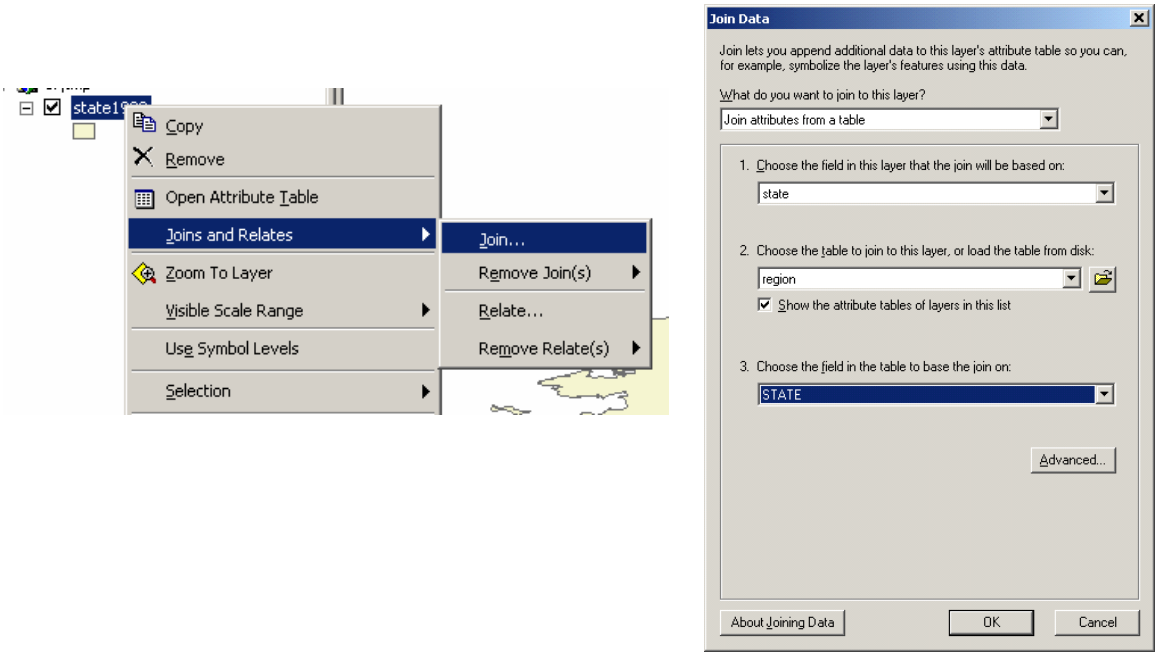
a. **Export stateregion.xls into a DBase file.**

Start Excel, open stateregion.xls. (It contains two variables: state and region.) Go to File > Save As..., select file type DBF4 (DBase IV) (.dbf), then Click Save.

b. Create a matching variable in the shape file:
Go to **ArcMap, Open Attribute Table > Options > Add Field**... type state as name, choose **Double** as type, **Precision** 9, and **Scale** 0, then **Click OK**.

c. Calculate values of **State** equals to **STATE_FIPS**.

d. **Right Click** the **state1990** layer in the **TOC** panel > **Joins and Relates > Join...**



In the **Join Data** window, choose **state** as base field, choose **region.dbf a**s source file, choose **state** as the base field**, Click OK. Hit Enter** at the window of **Create Index.**

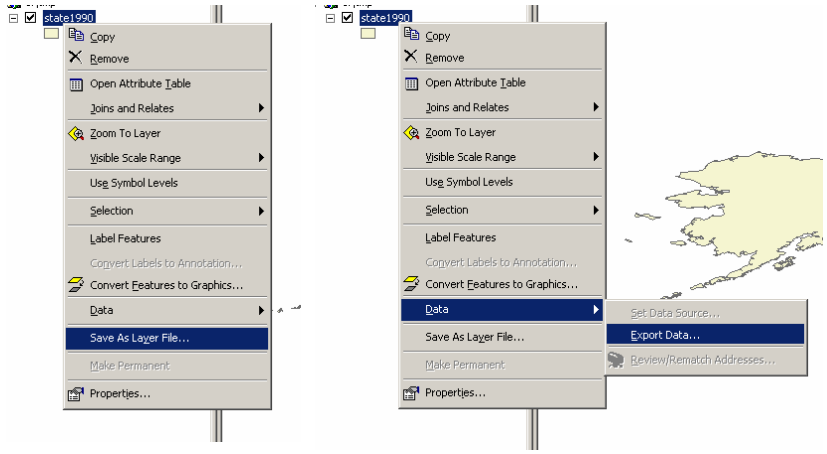
Open Attribute Table again, three new columns appear in the end of the table.

2. Save data
A. Save as a layer file

Right Click the state1990 layer in the **TOC** panel > **Save As Layer File**, locate your data folder, type file name **state1990add**. (Notice the default file extension.)

B. Save as another shapefile

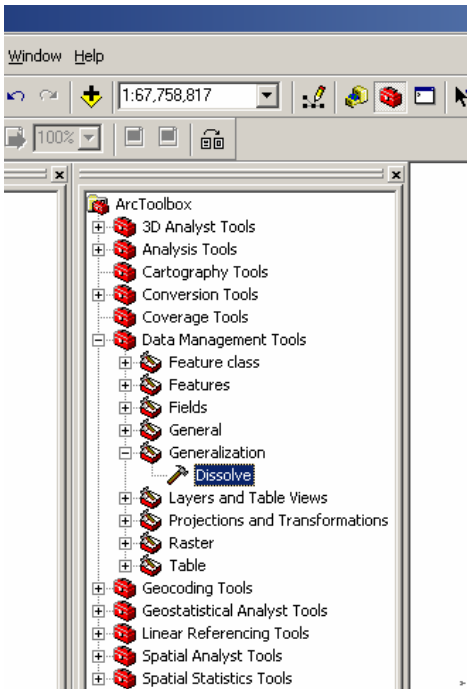
Right Click the **state1990** layer in the **TOC** panel > **Data > Export Data...** In the **Export Data** window type **state1990add** as new shape name, then **Click Ok**.

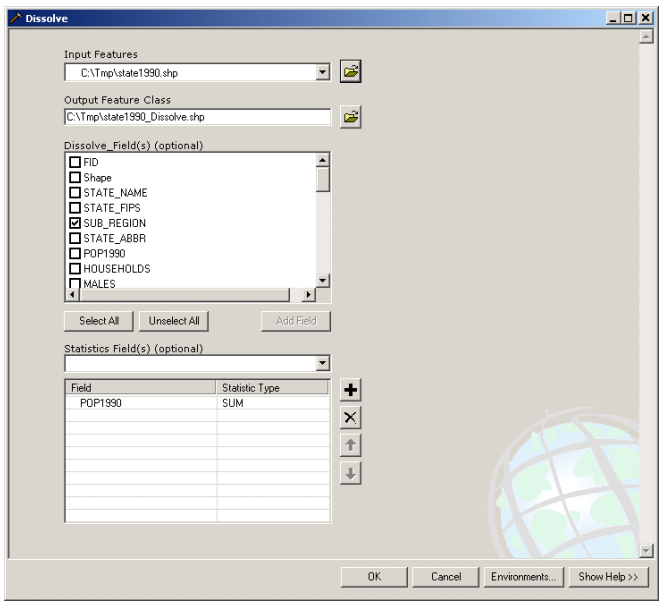


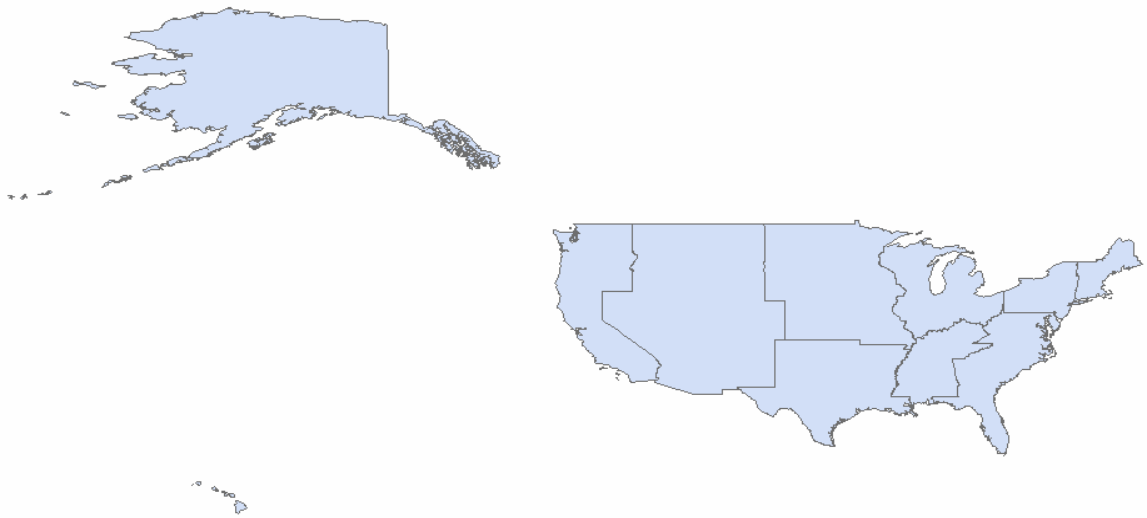
3. Query and map aggregation
A. Map aggregation- the dissolve function

a. Open **ArcToolBox, > Data Management Tools > Generalization > Dissolve...**

b. In the **Dissolve** window, select **state1990.shp** as input file, accept the default output file name, select **SUB_REGION** as dissolve field, select **pop1990** in the statistics field, click **Statistic Type** and choose **Sum**, then **Click OK**. In the new layer, the states are aggregated into Census divisions with division population total.

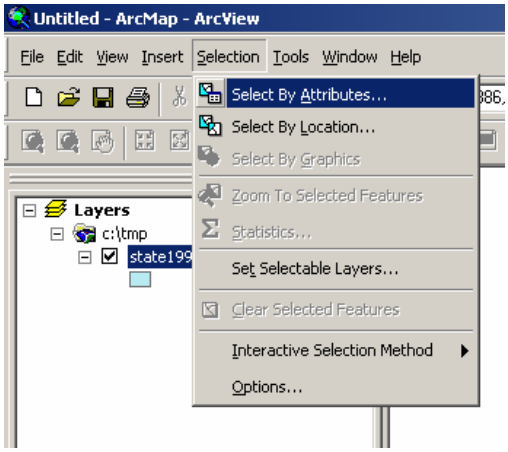


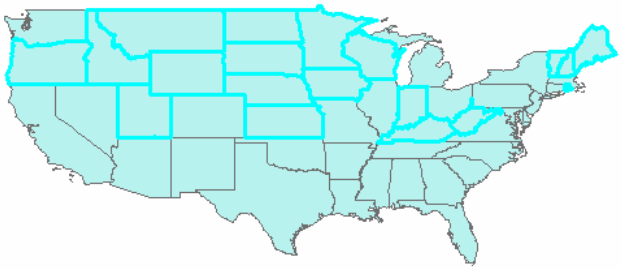


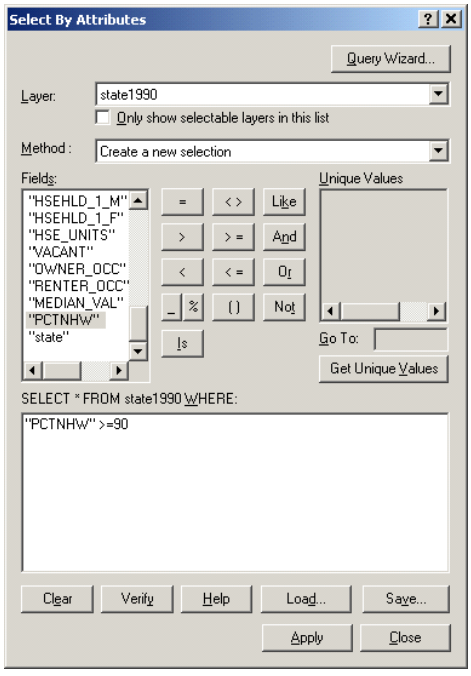


B. Query

a. Go to **main menu, Selection > Select By Attributes... select states with at least 90% non-Hispanic white population, then Click Apply.**



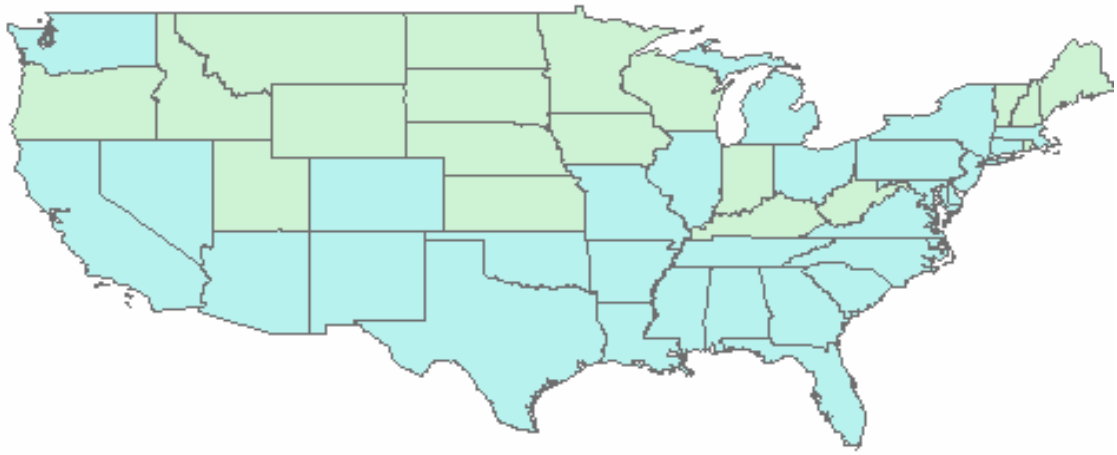




Reflective Research Questions:
How many states fit that criterion?
Do you see clusters?

b. Save the selected features in a separate shapefile: **query1.shp**. (**Right Click** the state1990 layer in the **TOC** panel > **Data > Export Data...)**

c. Add that layer on top of the states map.



Exercises to do on your own based on the data you have explored:

1. Create new variables of percent male, percent minority, and percent population between age 18 and 49. Report the standard statistics for each.

2. Aggregate the states into regions using state1990add.shp. Print a map of regions in the US. Report regional measures of the following variables: percent of non-Hispanic black population, percent of divorced people, and percent of female-headed households.

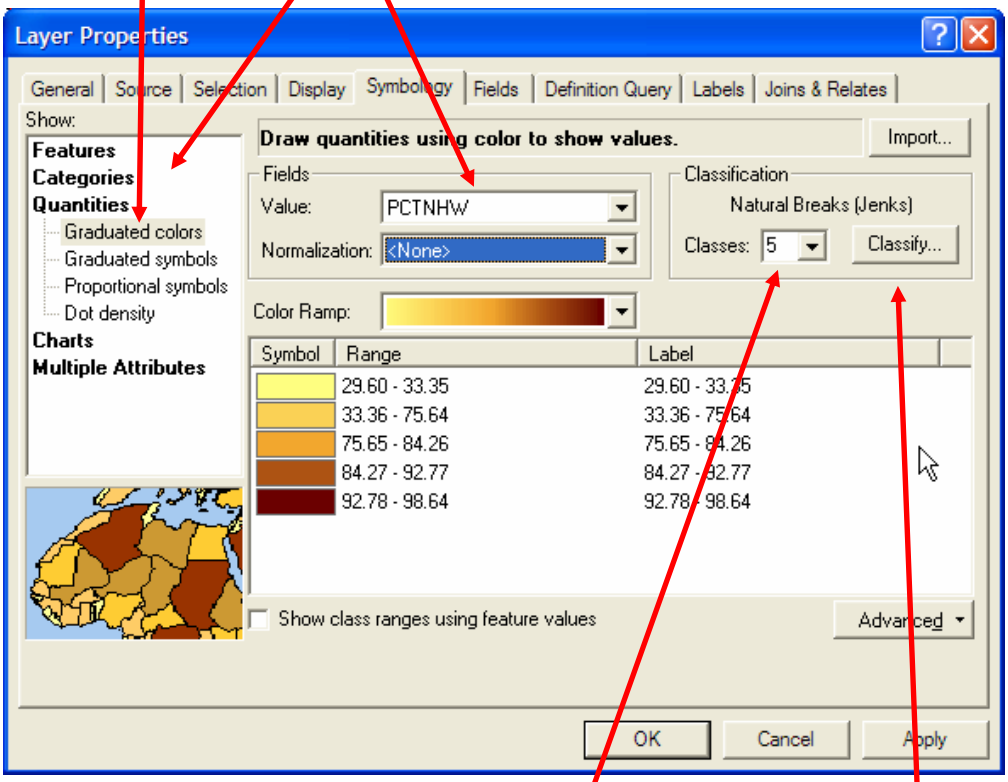
3. Identify the states with at least a quarter minority population and describe their cluster patterns. Print a map of only those states with labels.

Map Classification Tools

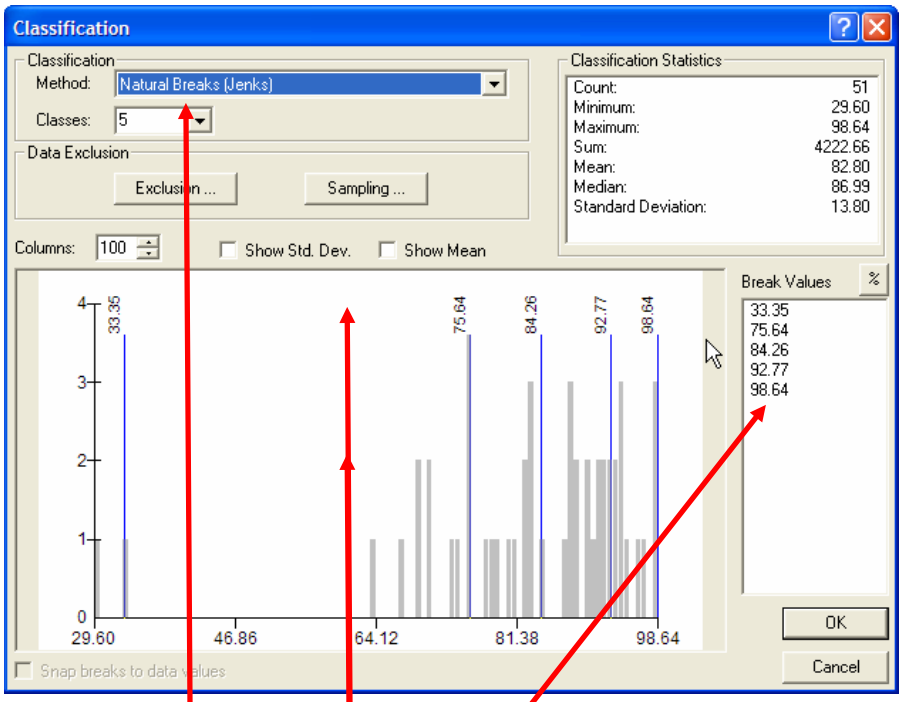
So far you have followed instructions to create national, regional, and state maps of different variables. Furthermore you have learned how to create new data fields and join data from outside sources. For the layers you have been working with, particularly the States1990 layer to which you added PCTNHW, open up the

symbology window (**right-click the layer in the TOC then select properties and then choose the symbology tab**).

In an earlier tutorial, and perhaps during your work on this one, you mapped distributions of variables. In our Getting Started with ArcGIS 2 you mapped a nominal/categorical variable in which all observations (countries) fell into one of 4 categories (of ecomonic development). With census data, such as the data that is part of this tutorial you have variables that are continuous/ratio, presenting you with numerous mapping options. When mapping continuous data you want to **open the quantities option on the left hand side** of the symbology window. Once you have selected the type of map you want to make (colors, symbols, or density) you need to **choose a variable** that is consistent with this type of map (it must by a continuous numeric variable) such as the variable you created earlier (PCTNHW). For the exercise you are going to **create a choropleth map that assigns a color to an area** (enumeration zone) based on its value for a certain variable (graduated colors).



The default classification is a natural breaks classification which is an Idiographic method of classifying data that relies on the specific, non-statistical, details of the data. While this might be a good starting point it is important to be comfortable with the numerous classification schemes that are available to you and understand the implications of choosing one over another. In the main symbology window the **number of classes can be changed**, but you can also work in **more depth with classification by clicking the "classify..."** button. Clicking on "classify" will bring up the following window.



You have several options you can explore in this window. You can change create/use a new classification scheme based on several rule systems; **use the "Method" pull down** menu to see your options. You can select a scheme that is based on the percentile of observations in each class (often called percentile or quantile, although the latter is often confused with quartile and quintile), the statistical distribution of the data, equal sized intervals, defined intervals, or you can create your own set of classes based on some exogenous information (something outside the distribution of data, probably related to phenomena being mapped). In the manual method (as with all methods, in fact) you can manipulate the break points for each class visually and manually in this window. Visually you can select (grab) the blue vertical lines that indicate class breaks and move (or delete) them in the central data distribution window. You can also **manually select the numeric break points** and delete them or edit them in the smaller "Break Values" window. This might be useful if you are mostly happy with an automatic classification scheme (standard deviation, quantile, etc.) but want to increase or decrease the size of a single class. Some aspects of the classification window will be unclickable depending on the type of classification you chose.

In order to better understand the classification process and some of the decisions that go into the selection of break points and numbers of classes you should read the "classifying data" article linked from the first module at http://www.s4.brown.edu/s4/about.htm, it provides a thorough description of

different classification schemes and provides some useful guidelines for their application.