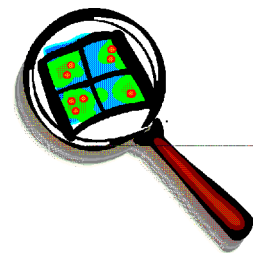
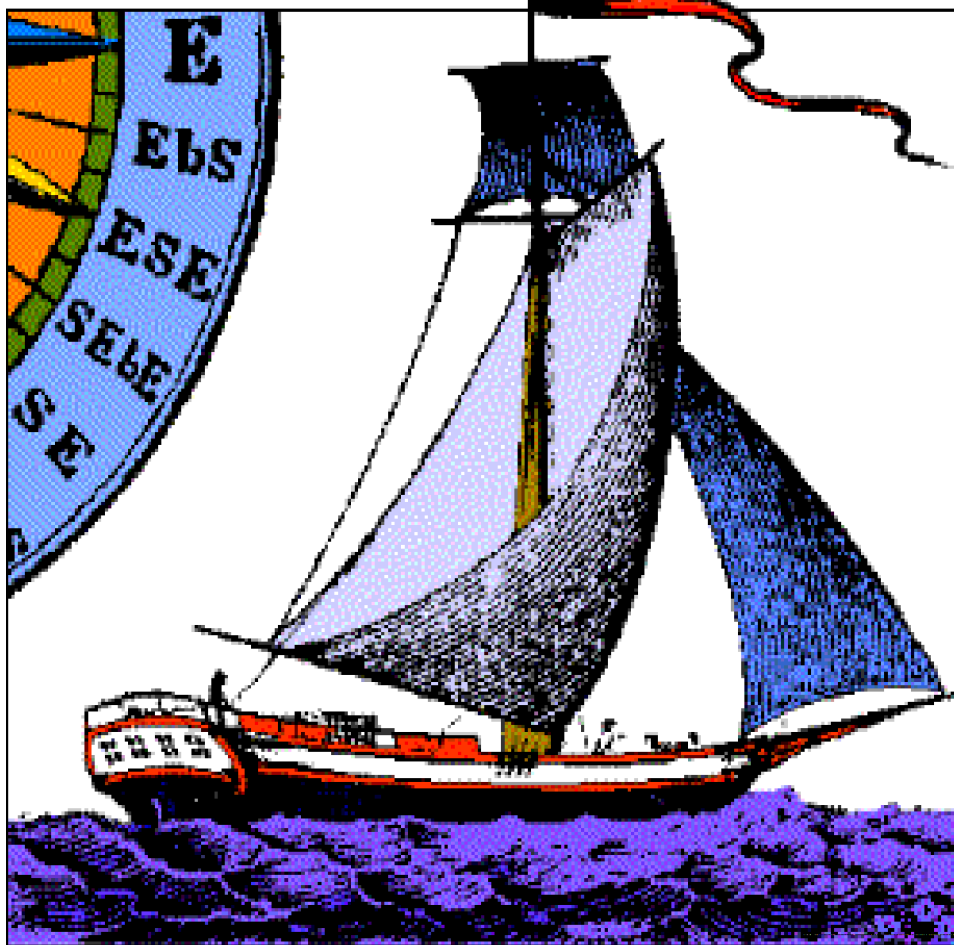


Working with ArcVoyager

Understanding Data



ArcView
for Schools and Libraries

ENVIRONMENTAL SYSTEMS RESEARCH INSTITUTE, INC.

Working with ArcVoyager

Understanding Data

One way of defining a GIS or geographic information system is that it is...

Information about the real world represented by points, lines, and areas that span a range of geography and that can be explored with a computer.

This snapshot statement emphasizes one of the critical components of a GIS—*data*. Without data, there are no maps, charts, tables, or graphs in a GIS. It would be like having a sports car but no gas. It looks good but it won't take you anywhere. In this case, without the *data* gas you would have a powerful software tool with nothing to display or analyze.

So when we talk about data, what are we talking about? The following discussions will give you answers and a starting point as you begin using ArcView and ArcVoyager to explore places and topics on the planet that are of interest to you. They will also point out some areas of caution.

GIS Data: Real World

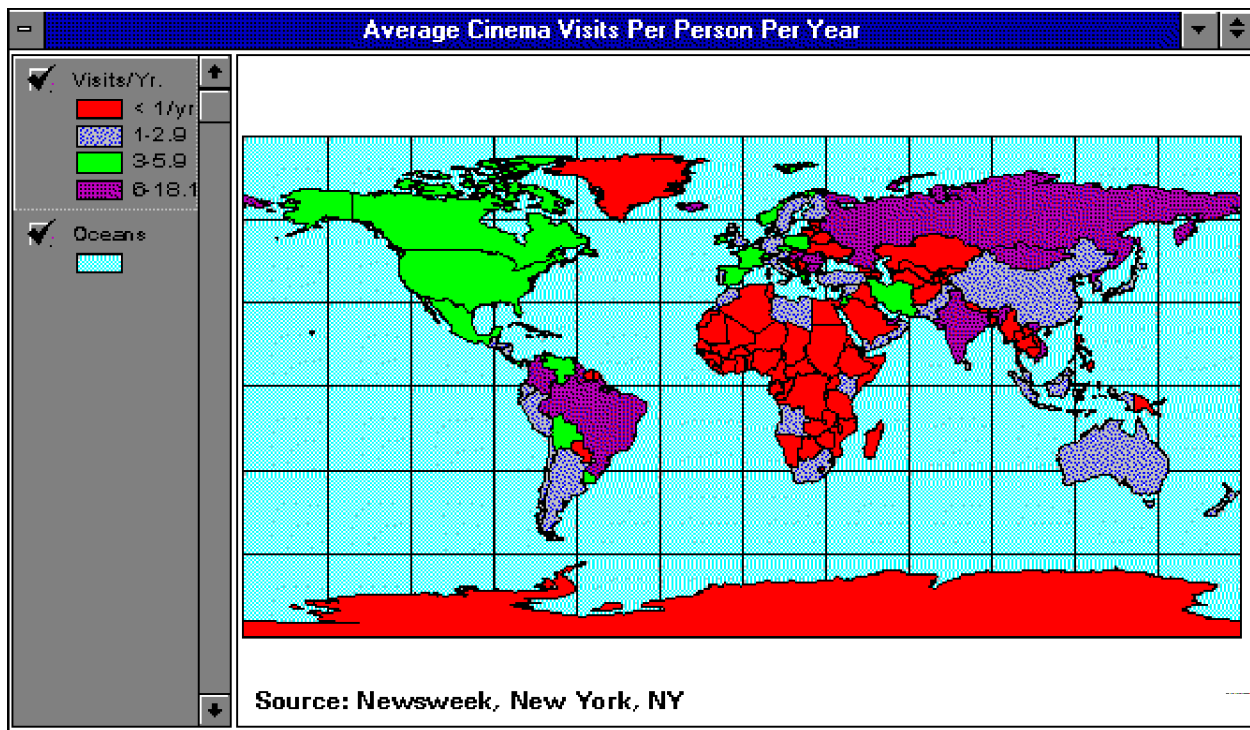
In another area of ArcVoyager, *Understanding Geographic Information Systems*, we explored the idea that GIS is about everyday life. The news story snippets there refer to some of the kinds of things that we are all regularly exposed to—the weather, parks, the environment, emergencies, the economy, people, traffic, neighborhood issues. Since these things surround us and sort of saturate our lives, we may not necessarily think about representing them as data, as information, but we can. If we can portray these things as data then we have to ask ourselves if we can picture them in a map. For instance...

Think about the weather. You may conjure up a picture of clouds on a stormy day but you can also bring up a picture of today's national weather map with all of its various patterns.

Think about a park. You may see a forest-covered hill in a local preserve but you could see a map that pinpoints the boundaries of that park and the location of its nature center, trails, lakes, and the hill you see.

Think about traffic. You may imagine a bumper-to-bumper scene at rush hour but you can also envision a map of city highways and roads highlighting areas of congestion.

So, many of the things that make up our day-to-day lives and those of people across the world can be represented as information that can be mapped. Since they can, they can be mapped using a GIS. Also, notice that the word is represent. Data and maps are not the people, the parks, the environment, the emergencies, and so on. They are depictions and tools that help us consider everyday problems and opportunities in ways that stretch beyond ourselves.



For example, if you were the owner of an international theater chain looking for new global markets, you might want to know something about people's movie habits around the world using mapped data that shows annual cinema visits by country on a per capita basis. In *Designing Global Adventures: Point Me*, you can explore this yourself. As you can see from the map above, the information will not be the people but a measure that when mapped helps you visualize and make decisions.

GIS Data: Shapes, Layers, and Types

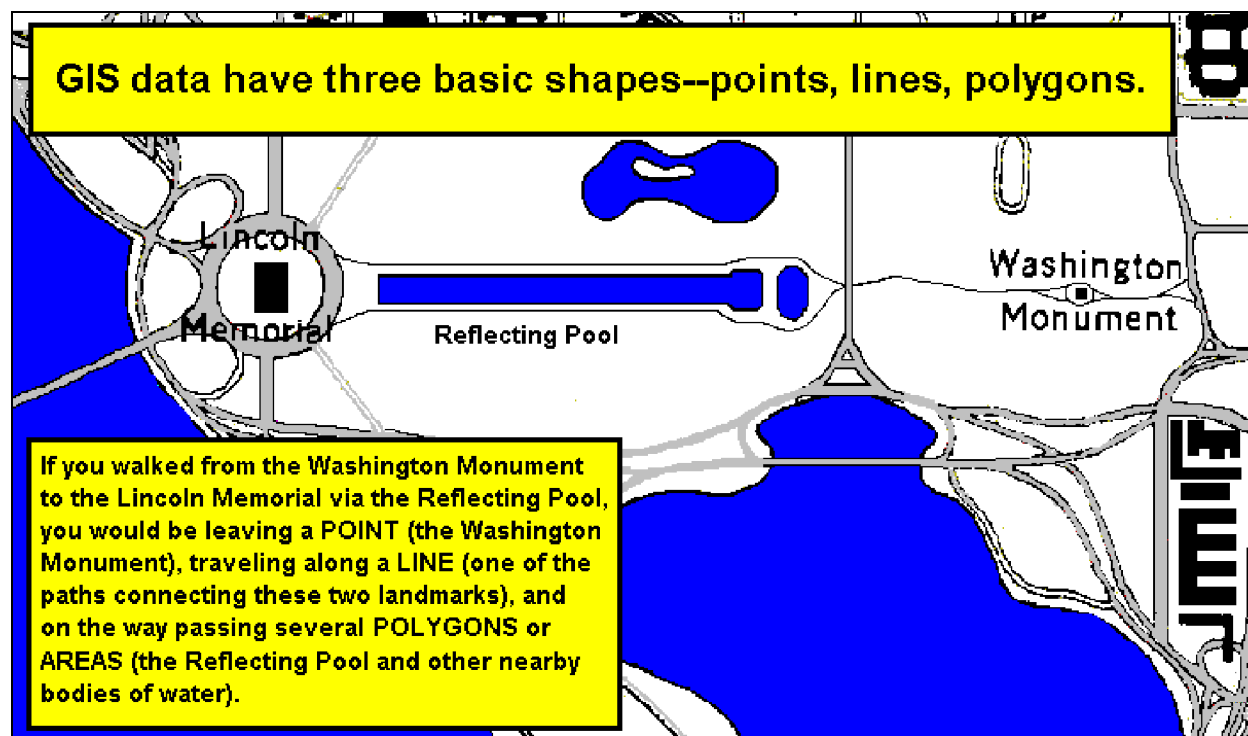
Since maps produced using a GIS are representations of the real world, how do we represent the data that we end up mapping? There are many ways in which we can represent and classify geographic data. We can look at them as...

- *Points, lines, and areas*
- *Layers*
- *Images, locational features, and statistical themes*

Points, lines, and areas

If you look at a map you might use to find your way across town or hike through a park, you will see various kinds of representations for things and spaces that are on the ground. Basically, these are displayed using three general shapes *points*, *lines*, and *areas* (which are also referred to as *polygons*).

For example, let's say that you are visiting Washington, DC and you are strolling from the Washington Monument to the Lincoln Memorial via the Reflecting Pool. If you mapped your walk, you would be leaving a *point*—the Washington Monument, traveling along a *line*—one of the paths connecting the two landmarks, and passing several areas (or polygons)—the Reflecting Pool and the other nearby bodies of water.



These three classes describe the basic shapes of geographic data found in a GIS.

- A point has no length or spatial area. A point is defined as a single X,Y coordinate such as something found at a particular latitude and longitude. (In this case, it's not just X marks the spot, but X and Y.)
- Lines represent linear features (even if they have bends). Lines have length but no spatial area. A line can also be thought of as a string of X,Y coordinates that have been connected from a beginning point to an ending point, like a straight line.
- Areas or polygons have length, spatial area, and a perimeter. A polygon is defined as a string of X,Y coordinates with the same beginning and ending point. For instance, a circle, a rectangle, or a many-sided shape works here.

To help clarify the concepts, here are several examples for each class. Note that some can be in more than one category depending upon scale. Also, if you were studying a single topic (such as the oil industry) you could have data with all three shapes (see the italicized text).

- Points—fire hydrants, mountain peaks, cities, telephone poles, *oil wells*
- Lines—rivers, highways, railroads, contour lines, utility lines, *oil pipelines*
- Areas (or polygons)—cities, countries, climate zones, biomes, agricultural belts, *oil-producing regions within a country*.

Layers

An important aspect of a GIS is its ability to overlay or stack various sets of geographic data one on top of another. It is much like making a sandwich—an Earth sandwich.

A single location on the planet has many characteristics. For instance, we might look at a single area in terms of *people, climate, roads, rivers, history, agriculture, archeology, geological structure, political alignments*, and so on. Because they are all about a particular spot, they weave together to compose a geographic mosaic of that location. This ability to stack or layer data gives us the opportunity to study relationships, see patterns, and the like.

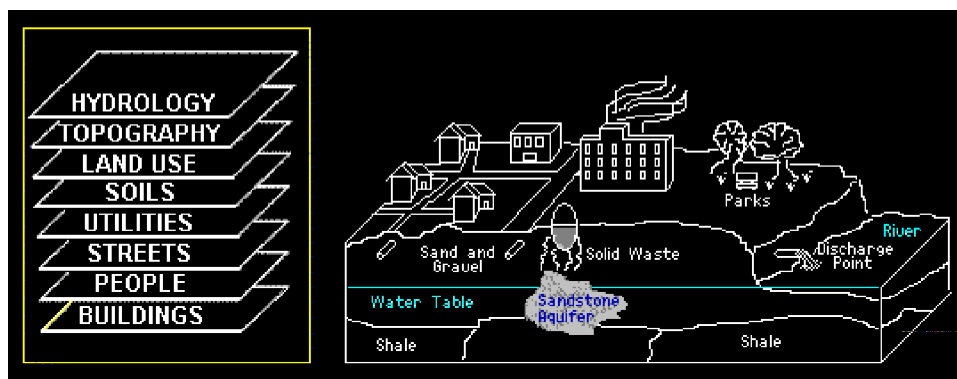
One way to begin looking at GIS data layers is by breaking them into two major groups—physical and human geographic layers.

- *Physical geographic layers* include information about the natural features of an area.
- *Human geographic layers* relate to information about people, their structures, or their interactions with the environment.

Some of the kinds of physical and human geographic layers that are found in a GIS include

- *Ecology*—such as types of plants and animals, biome class, ecological hazards and threats
- *Climate*—such as precipitation, cloud and wind patterns, temperature
- *Hydrology*—such as rivers, streams, lakes, reservoirs, aquifers
- *Topography*—the physical features of a place or region such as elevation
- *Geology*—such as the shape of the land in an area (mountains, plains, etc.), types of rocks, and geological risk factors
- *Land use*—essentially any use to which humans put the earth's surface such as agricultural, commercial, industrial, recreational, and residential
- *Soil classes*—the quality or characteristics of the soil in various locations such as sand, loam, or clay or a soil's suitability for a specific land use

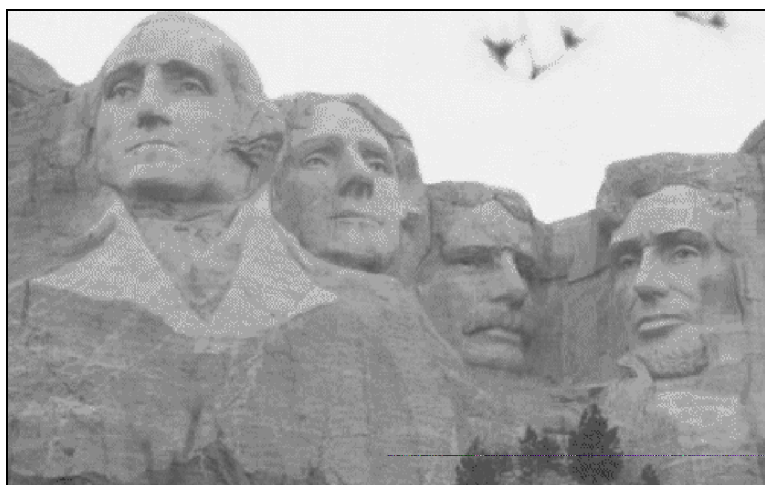
- **Land parcels**—such as the piece of property associated with a house, apartment building, or business
- **Utilities**—the above and below ground network of wires, cables, pipes, and tanks carrying electricity, water, sewage, and communications
- **Streets**—the network of roads, highways, bridges, tunnels, and other thoroughfares
- **Other human characteristics**—such as demographic, socioeconomic, historical, political, or environmental characteristics of a place or region



It is also useful to note that GIS data about a place include information depicting conditions above, on, and below the ground.

Images, locational features, and statistical themes

Images. GIS applications including ArcView are capable of displaying images as well as other information about locations. While you can bring computer-ready versions of photographs into ArcView such as a shot of Mount Rushmore, the images that are best suited for use in a GIS are those that are *georeferenced*. This means that these images have geographic coordinates associated with them and they can be correctly placed on the Earth. This, in turn, means that other layers of information (such as roads or country boundaries) can be placed on top of them.



Here are two examples from images that have been geographically tagged. One is taken from a *satellite image* that displays a small segment of a Landsat photograph just east of Garden City, Kansas. (Note the circular areas. They are irrigated fields.)



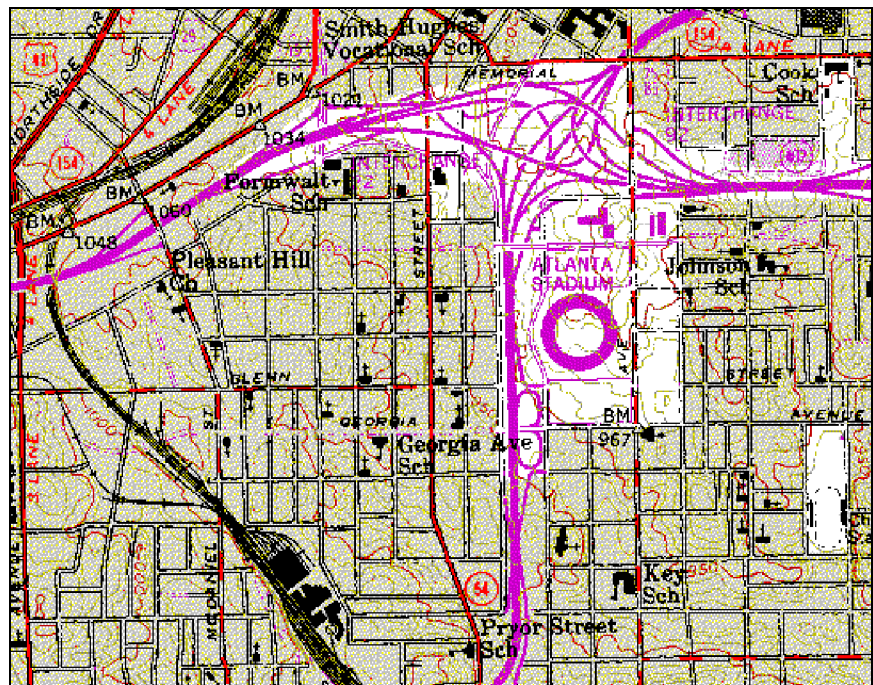
The other is from a *scanned topographical map* produced by the U.S. Geological Survey. It shows an area near the Atlanta stadium. (Note that this depiction does not show any of the construction for the 1996 Olympics.) The images from which these shots were clipped are geographically registered.

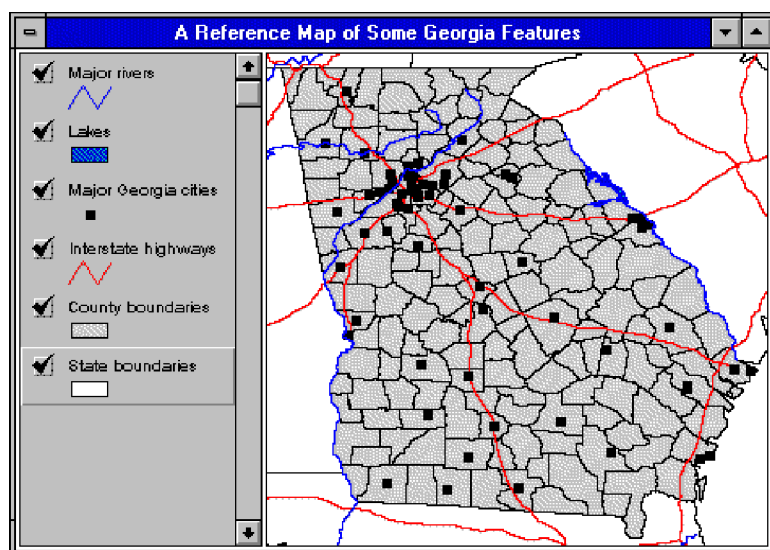
Locational features. One way to define feature data is that they make up the layers of a GIS that ***provide a general locational context for data in other layers.***

What does that mean? A good mental snapshot to use is the kinds of information that you generally see on a standard state highway map. Some of the characteristics, landmarks, and other things you usually find are political boundaries (like the state outline and boundaries for counties or similar areas), cities and towns, rivers and streams, lakes and reservoirs, and some kind of locator grid system (such as lines of latitude and longitude). On the next page you will see an example of a reference map of some Georgia features.

Statistical themes (or attributes). With the items mentioned above, you can create a good reference map. But imagine that you want to know something about those features. The data that help you to do this are statistical theme (or attribute) layers. They ***represent characteristics about features*** whether they are points, lines or areas.

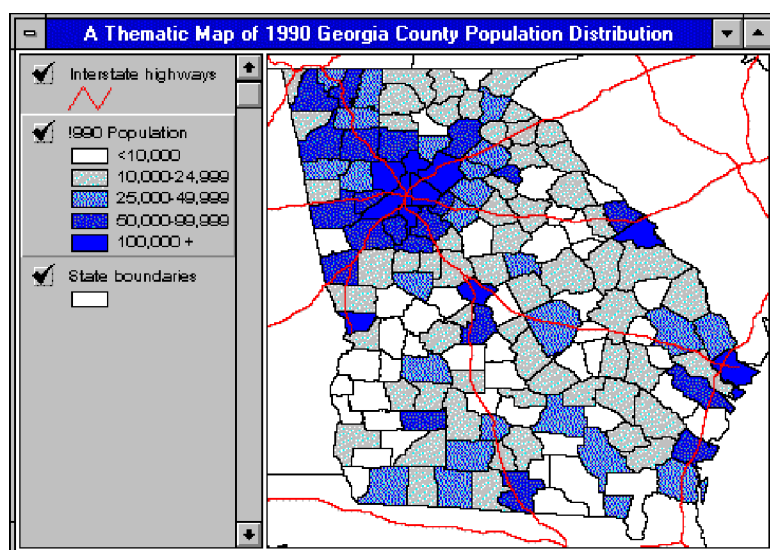
For instance, picture that you want to investigate county population distribution across Georgia. In this case, you'll need population statistics for each individual county that you can combine with individual county geography. By combining





the statistics (population) with the needed geographic features (counties as areas), coding the statistical data into ranges, and adding colors or patterns for each range, you can create a thematic map of Georgia population distribution like the one shown here.

It's like having a cake mold (or a geographic container) into which you pour various kinds of cake batters (or statistics). You can keep using the same mold but continue to change its contents. In the ArcView for Schools and Libraries data bundle some of the richest statistical theme layers are for countries of the world, U.S. states, and U.S. counties.



As you get more involved with GIS, you will hear the word *coverages*. It's nothing complicated. It's just part of the GIS language.

A coverage is in essence the combination of locational feature data and statistical themes (or

attributes) about a single type of geographic data, such roads or counties. Together, they represent the data that can answer a GIS software application's questions of

"How do I draw this particular kind of thing (say mountain peaks) in correct geographic location?"

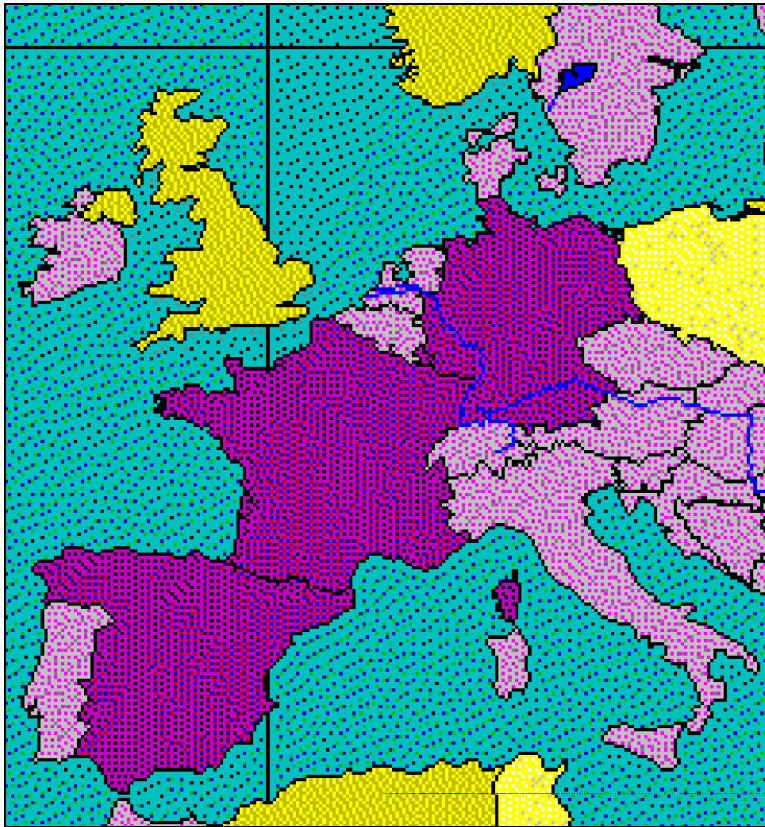
"Now that I know how and where to map mountain peaks, what can I map about them (like elevation ranges)?"

GIS Data: Geographic

While GIS data are geographic, not all levels of geography are contained in any one GIS.

What does that mean? Basically, a GIS can allow you to display and analyze data for any number of geographic locations around the globe but to do so requires that you have the data—the locational feature and statistical theme data about those places.

For instance, a GIS used by the World Health Organization may look different from one used by the U.S. Centers for Disease Control. And these will be very different from one used by the Fulton County Health Department. While there will be crossover among these three, the things that will be most different will be the geographic focus and the exact kinds of geographic data they each have at their disposal.



So, just because you may have a map displayed on your GIS showing the boundaries of European countries doesn't mean that you can continue to zoom in to see the road network for Paris. In order to see that road network, you must have the needed data available to you.

Geography found in the ArcView bundle and ArcVoyager

The ArcView for Schools and Libraries data bundle and the specialized data sets associated with ArcVoyager provide you with a good range of geography and thematic data but it is limited. On the following page are two example schematic views of geographic areas focusing

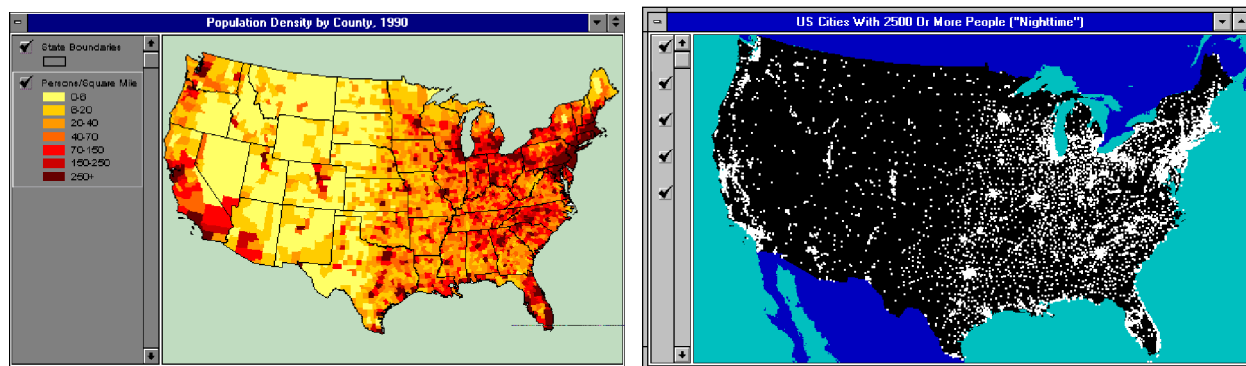
mainly on politically defined areas. The first looks at global geography, the second emphasizes the United States. Both highlight the geographic areas you will find as you work with the ArcView and ArcVoyager data sets. In terms of thematic mapping with the ArcView and ArcVoyager data sets, the richest geographic areas to use are countries of the world, U.S. states, and U.S. counties.

Some Global Geographic Areas		
Geographic Hierarchy	Geography in ArcView and ArcVoyager	Data in ArcView and ArcVoyager
The world	yes	feature data
Continents	yes	feature data
Countries	yes	feature and statistical data
Subnational boundaries	some	some statistics for Canada, Mexico, more U.S.
Cities	yes	point features, minimal statistics
Subcity, neighborhoods	no	no

Some U.S. Geographic Areas		
Geographic Hierarchy	Geography in ArcView and ArcVoyager	Data in ArcView and ArcVoyager
Nation	yes	feature data, some 48 states only
States	yes	feature and statistical data
Federal lands	yes	feature data
Counties and equivalent areas	yes	feature and statistical data
Cities	yes	point features, minimal statistics
Census tracts and other subcounty census geography	Atlanta only	feature and limited statistical data
ZIP Codes	yes	feature data, no statistics
Street networks, etc.	Atlanta only	feature data

Although there are geographic and data holes in the above listings, the holes can be filled. Another important aspect about a GIS is that it can be updated. New geographic data can be added once it is in a recognizable computer format. This also means that you can add new statistical data sets for geographic areas you already have in your GIS. You can also create new geographic areas such as point locations by knowing the latitude and longitude of those points.

Therefore, a critical step to take before you create a map is to consider the geographic focus that your question or research requires. After that, you must determine if you have the needed feature and statistical information. If you don't, is it possible to get or create it? If not, can you work with what you have? For instance, simply saying that you want to map *the*



population of the contiguous 48 states can net very different displays depending upon interpretation. It can result in a map of population density across counties or it might end up looking at urban areas as centers of population concentration.

GIS Data: Computerized

Working with a GIS means working with computerized data. All of the locational feature, statistical theme, and image data sets must be recognizable to the GIS software, such as ArcView. For you, this means several things. You need to know how to access the data files. *Where are files located and what's in a file once you find it?* It also means knowing something about data file formats. *What files can and cannot be used in this software?*

Data file locations and contents

Since we have supplied a number of pre-created data sets for use in specific areas of ArcVoyager, the initial need to know about the location of data sets is minimal. However, the first time you wish to add an additional data set, you will need to know where you can find them. You will need an understanding of and skills in using directories and files. If these concepts are new to you, we suggest that you begin by investigating *Understanding the PC Computer* section of *Exploring Key Concepts: Teach Me*. This area will point out specific Windows and general personal computer skills needed in navigating directories and files.

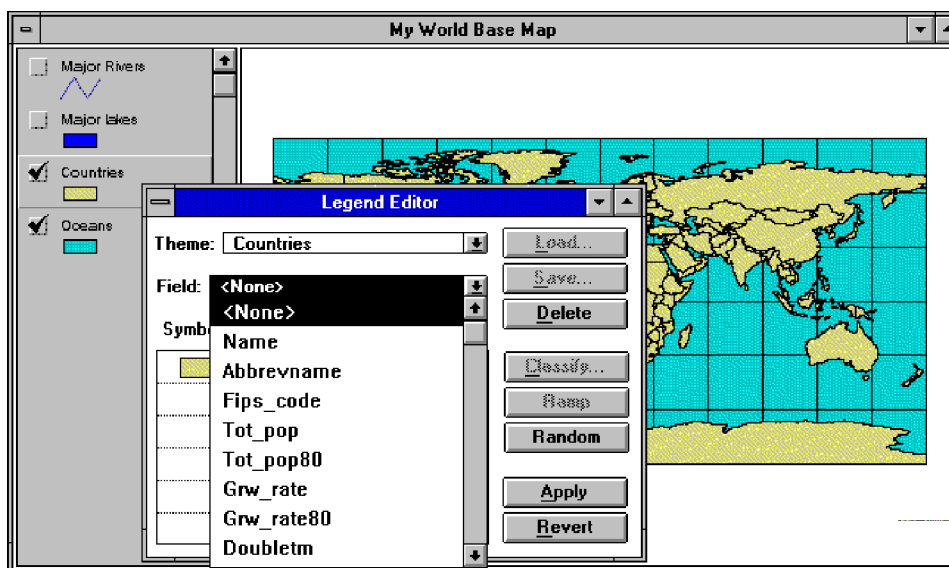
Files and fields

As you work with the data sets in *Designing Global Adventures: Point Me*, you will quickly be exposed to the notion of fields within specific files. A field is much like a single column of information you might see in a table of numbers. A commonplace example is the weather information for cities around the world you find in the daily newspaper. Columns in the table

might show yesterday's weather and today's forecast. If we turned such a table into a computerized database, the two columns would represent fields.

For instance, one of the data sets we have added for your use in *Designing Global Adventures: Point Me*, is a single file—DEMOG.DBF. Within this file, are 28 fields with country-level mappable data such as population, agricultural land, and GNP. The DEMOG.DBF data file is also comprised of records or countries. Below is a snapshot from the beginning of that file showing a handful of fields and records. (**Note:** We

demog.dbf					
Fips_code	Abbrevname	Tot_pop	Tot_pop80	Grw_rate	Grw_rate80
AF	Afghanistan	19903010	15950000	2.34	2.57
AG	Algeria	24453010	18669010	2.90	3.07
AJ	Azerbaijan	7450787	-99	1.60	-99.00
AL	Albania	3204000	2671000	1.92	0.04
AM	Armenia	3415566	-99	0.80	-99.00
AO	Angola	9694001	7723001	0.00	0.00
AR	Argentina	31883010	28237010	1.19	1.55
AS	Australia	16765000	14692000	1.40	1.22
AU	Austria	7598001	7554001	0.04	0.04
AY	Antarctica	-99	-99	-99.00	-99.00
BC	Botswana	1217000	902000	3.09	3.50
BE	Belgium	9886001	9847001	0.08	0.10
BG	Bangladesh	111590000	87510000	2.48	2.76
BH	Belize	184208	143925	2.57	2.24
BK	Bosnia/Herz	4364000	-99	0.50	-99.00



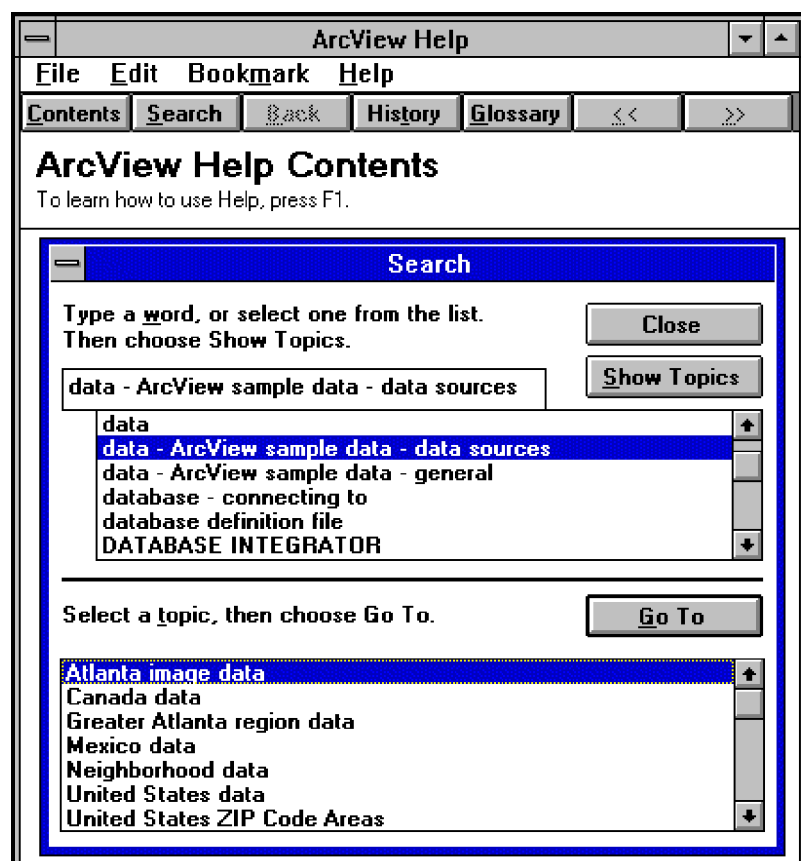
have highlighted a few countries in the above image. In some of the fields shown, you will see special codes **-99** and **-99.00**. These values indicate fields where data are missing or are not available.)

Another way you will encounter fields is in using the *Legend Editor*. The

Legend Editor, among other things allows you to change the data field you are mapping for the particular *theme* you have selected (such as Countries) in the map display's *Table of Contents* (the left side of the ArcView map display). If you wanted to map total population (TOT_POP) for the countries in the DEMOG.DBF database. You would select that field name. If you wanted to see something different, you would click on one of the other statistical fields. The same is true for any data file you use.

Field names and data sources and dictionaries

By working with GIS data files and their fields, you will be faced with field names such as TOT_POP and GRW_RATE in the DEMOG.DBF file. These are fairly clear as to what they are about. However, you cannot tell the year of the data from the names. Since field names



are limited to ten characters, other names may not be as obvious—PCT_F_WORK. This is where data dictionaries or data source explanations are invaluable. Currently, in ArcView and ArcVoyager, you have three data source areas to explore for explanations of some of the data you will be viewing and using. Inside *Viewing World Snapshots: Show Me*, you will find a hypertext section called *Learning About Atlas Data Sources*. This will give you a brief snapshot of where the information is from and anything else immediately important.

Another area of data source help is inside ArcView Help. By searching on the keyword

data (and subsequently clicking to *data - ArcView sample data - data sources*) you can explore background information for some of basic data sets that came with ArcView and which have been included inside ArcVoyager such as DEMOG.DBF.

You will also find information about the other data sets which are part of the standard ArcView bundle available in the *Introduction to Other ArcView for Schools and Libraries Data Sets* section of *Creating New Worlds: Turn Me Loose*.

In *Creating New Worlds: Turn Me Loose* you will notice an additional data heading, *Access to Other Cool Data Sets*. Presently, this area is essentially empty. However, we will be adding database descriptions to this area in the next edition of ArcVoyager. In the meantime, we have created a document that explains other ArcVoyager data sets currently used in *Designing Global Adventures: Point Me*. The guide is *Working with ArcVoyager, Designing Global Adventures: Data Dictionaries*.

ArcView and ArcVoyager data formats

The data sets you will most often use in ArcView and ArcVoyager are ready to go. Like file locations and contents, there is not much you need to know initially about data formats. However, as you get more involved with ArcView and ArcVoyager, you will want to work

with some of the other data sets available for your use as well as design some of your own. To do that, you need to know about the kind of file formats that these software applications crave. Below are some file formats associated with particular ArcView data types.

ArcView Data Types and File Formats		
Type of data	What's it for	File formats to know
Locational feature data	"How do I draw it" data	Special files created using ESRI software. Files to know are shape (.SHP) files in ArcView or coverages in ARC/INFO (ESRI's top-end GIS)
Statistical theme data	"What's this place like" data	Table files like those used in a spreadsheet or database application. Files to know are .DBF and .TXT
Image data	"Picture is worth a thousand words" data	Graphics and picture files like those used in a graphics application. The format to immediately know is .TIF

Creating your own data

You can create your own data for use in ArcView and ArcVoyager. However, we recommend that you learn your way around the software and data you have before launching into this effort. Here are some basic guidelines on types of data you may want to create, what software you will need to create them, and the specific file types you can produce for ArcView.

Creating Your Own Data for ArcView		
Data type to create	What you'll need	What files you want for use in ArcView
Locational feature data	ArcView	.SHP files for new geographic areas, lines, and points
Statistical theme data	ArcView, a spreadsheet or database package	.DBF or .TXT files for new characteristics about places
Image files	A good graphics package and image files (such as a personal photo CD)	.TIF files for geo-registered and non-geographically registered pictures about locations