

## **OBJECTIVES**

*Define GIS*

*Define spatial data for graphic and image map layers*

*Review the national infrastructure for spatial data*

*Review the unique capabilities of GIS*

*Demonstrate how GIS can be used for health applications*

*Introduce ArcGIS and its user interface*

## ***GIS Tutorial 1***

# **Introducing GIS and health applications**

Geographic information systems (GIS) is a technology with unique and valuable applications for policy makers, planners, and managers in many fields, including public health and health care. GIS software and applications enable you to visualize and process data in ways never before possible. The purpose of this book is to provide you with hands-on experience using the premier GIS software package ArcGIS Desktop in the context of health applications. You need not have any previous experience using GIS.

The next section describes GIS, its inputs, and special capabilities, followed by a discussion of health issues and GIS applications. We also preview the upcoming tutorials in this book, then introduce you to ArcGIS software in a short tutorial.

# What is GIS?

GIS is a multidisciplinary software system that engages geographers, computer scientists, social scientists, planners, engineers, and others. Consequently, it has been defined from several different perspectives (e.g., see Clarke 2003). We prefer a definition that emphasizes GIS as an information system: *GIS is a system for input, storage, processing, and retrieval of spatial data.* Except for the additional word “spatial,” this is a standard definition for an information system. Spatial components include a digital map infrastructure, GIS software with unique functionality based on location, and new mapping applications for organizations of all kinds. Next, we define and discuss these distinctive aspects of GIS.

## Spatial data

Spatial data is information about the locations and shapes of geographic features, in the form of either vector or raster data. Graphic maps, also known as vector maps, are created with layers that have features drawn using points, lines, and polygons. Geographic objects have a variety of shapes, but all of them can be represented as one of three geometric forms—a polygon, a line, or a point. A polygon is a closed area with a boundary consisting of connected straight lines. For example, figure 1.1 is a map with two polygon map layers (state and county boundaries), a line layer of rivers, and a point layer of cities with populations of 250,000 or more. The state and county boundaries are coterminous, that is, they share boundaries and do not overlap each other. We used fill color within the county polygons to show the mortality of lung cancer for white males, which has some striking geographic patterns to be discussed later in this tutorial.

Associated with individual point, line, or polygon features are data records that provide identifying and descriptive data attributes. For example, in figure 1.1 the labels for the names of states and cities come from tables of attribute records associated with each map layer. You will revisit this map in the tutorial where you will use ArcGIS to explore map layers and spatial patterns of cancer mortality.



**Figure 1.1** Lung cancer mortality per 100,000 white males, 1970–1994.

Sources: (a) ESRI Data & Maps 2005 and (b) The Cancer Mortality Maps and Graph Web site: The National Cancer Institute

Raster maps are aerial photographs, satellite images, or images created with software that are stored in standard digital image formats, such as tagged image file format (TIFF) or Joint Photographic Experts Group (JPEG). An image file is a rectangular array, or raster, of tiny square pixels. Each pixel has a single value and solid color, and corresponds to a small, square area on the ground, from 6 inches to 3 feet on a side for high-resolution images. (Pixels are also often referred to as cells.) Accompanying the image files are world files that provide georeferencing data, including the upper left pixel's location coordinates and the width of each pixel in ground units. The world file provides the data needed, so the GIS software needs to assemble individual raster datasets into larger areas and overlay them with aligned vector datasets.

Viewed on a computer screen or on a paper map, a raster map can provide a detailed backdrop of physical features. In figure 1.2, an aerial photograph overlaid with vector map layers, shows locations where serious injuries of child pedestrians occurred in relation to public playgrounds. The two boundaries surrounding the parks are 1,250- and 2,500-foot buffers used to study injury rates near parks. You will explore the GIS data behind this map in the exercises in this tutorial and then in depth in tutorial 7, where you will create and use buffers similar to those in figure 1.2.

Map layers have geographic coordinates, projections, and scale. Geographic coordinates for the nearly spherical world are measured in polar coordinates, angles of rotation in degrees, minutes, and seconds, or decimal degrees. The (0,0) origin is generally taken as the intersection of the equator with the prime meridian (great circle) passing through the poles and Greenwich, England. Longitude is measured to the east and west of the origin for up to 180° in each direction. Latitude is measured north and south for up to 90° in each direction.

The world is not quite a sphere because the poles are slightly flattened and the equator is slightly bulged out. The world's surface is better modeled by a spheroid, which has elliptical cross sections with two radii, instead of the one radius of a sphere. The mathematical representation of the world as a spheroid is called a datum; for example, two datums



**Figure 1.2** Locations of serious injuries to child pedestrians in Eastern Pittsburgh, Pennsylvania.

Sources: (a) Children's Hospital of Pittsburgh; (b) Pennsylvania Spatial Data Access (PASDA) and USGS; (c) City of Pittsburgh Department of City Planning.

commonly used for North America are NAD 1927 and NAD 1983. If you use the same projection, but with two different datums, then each corresponding map will have small but noticeable differences in coordinates.

A point, line, or polygon feature on the surface of the world is on a three-dimensional spheroid, whereas, features on a paper map or computer screen are on flat surfaces. The mathematical transformation of a world feature to a flat map is called a projection. There are many projections, some of which you will use in tutorial 4. Each projection has its own rectangular coordinate system with a (0,0) origin conveniently located so that coordinates generally are positive and have distance units, usually feet or meters.

Necessarily, all projections cause distortions of direction, shape, area, and length in some combination. So-called conformal projections preserve shape at the expense of distorting area. Some examples are the Mercator and Lambert conic projections. Equal area projections are the opposite of conformal projections: they preserve area while distorting shape. Examples are the cylindrical and Albers equal area projections (Clarke 2003, 42–44).

Map scale is often stated as a unitless, representative fraction; for example, 1:24,000 is a map scale where 1 inch on the map represents 24,000 inches on the ground, and any distance units can be substituted for inches. Small-scale maps have a vantage point far above the earth and large-scale maps are zoomed in to relatively small areas. Distortions are considerable for small-scale maps but negligible for large-scale maps relative to policy, planning, and research applications.

GIS maps are composites of overlaying map layers. For large-scale maps such as in figure 1.2, the bottom layer can be a raster map with one or more vector layers on top, placed in order so that smaller or more important features are on top and are not covered up by larger contextual features. Small-scale maps, such as figure 1.1, often consist of all vector map layers. Each vector layer consists of a homogeneous type of feature: points, lines, or polygons.

### **Digital map infrastructure**

GIS is perhaps the only information technology that requires a major digital infrastructure: We also refer to the map layers of the infrastructure as basemaps, namely, a collection of standards, codes, and data designed, built, and maintained by government. Vendors provide valuable enhancements to the digital map infrastructure, but for the most part, it is a public good financed by tax dollars. Without this infrastructure, GIS would not be a viable technology.

The National Spatial Data Infrastructure (NSDI), developed by the Federal Geographic Data Committee ([www.fgdc.gov](http://www.fgdc.gov)), “. . . encompasses policies, standards, and procedures for organizations to cooperatively produce and share geographic data.” Associated with the NSDI is the Geospatial One-Stop Web site ([gos2.geodata.gov](http://gos2.geodata.gov)) for access to spatial data. Next we will briefly review several components of the NSDI.

Perhaps most useful for health applications are the TIGER/Line maps provided by the U.S. Census Bureau. These maps are available by states and counties for many classes of layers. These classes, and examples of each, follow:

- *Political layers*—states, counties, county subdivisions (towns and cities), and voting districts
- *Statistical layers*—census tracts, block groups, and blocks
- *Administrative layers*—ZIP Codes and school districts
- *Physical layers*—highways, streets, rivers, streams, lakes, and railroads

You can download TIGER/Line map layers in GIS-ready formats at no cost from a variety of Web sites, including the U.S. Census Bureau site ([www.census.gov](http://www.census.gov)) and the ESRI site ([www.esri.com](http://www.esri.com)). Steps for doing so are found in tutorial 5.

Corresponding to TIGER/Line maps of statistical boundaries is census data tabulated by census and federal information processing standards (FIPS) codes ([www.census.gov/geo/www/fips/fips.html](http://www.census.gov/geo/www/fips/fips.html)) and census-area codes. FIPS codes are “. . . a standardized set of numeric or alphabetic codes issued by the National Institute of Standards and Technology (NIST) to ensure uniform identification of geographic entities through all federal government agencies.” Data from the decennial census is available at no cost from [www.census.gov/geo/www](http://www.census.gov/geo/www). Steps for downloading census data and preparing it for GIS use with TIGER/Line maps are also found in tutorial 5.

The U.S. Geological Survey (see [www.usgs.gov/aboutusgs](http://www.usgs.gov/aboutusgs)) is the “largest water, earth, and biological science and civilian mapping agency . . . [and it] collects, monitors, analyzes, and provides scientific understanding about natural resource conditions, issues, and problems.” Among its products useful for health applications are its 1:24,000-, 1:25,000-, 1:63,360-scale topographic maps, known as digital raster graphic (DRG) maps in scanned image format, and its digital orthophoto quarter quadrangle (DOQQ) aerial photographs (such as are used in figure 1.2). Full national coverage of the most recent DOQQs and 90 percent of DRG maps are available at nominal cost from the Seamless Data Distribution System at [edcns17.cr.usgs.gov/EarthExplorer](http://edcns17.cr.usgs.gov/EarthExplorer).

Local governments provide many of the large-scale map layers in the United States, and in this data most features are smaller than a city block. Included are deeded land parcels and corresponding real property data files on land parcels, structures, owners, building roof footprints, and pavement digitized from aerial photographs. Often you can obtain such map layers and data for nominal prices from local governments.

### ***Unique capabilities of GIS***

Historically, maps were made for reference purposes. Examples of reference maps are street maps, atlases, and the USGS topographic maps. It wasn't until GIS, however, that analytic mapping became widely possible. For analytic mapping, an analyst collects and compiles map layers for the problem at hand, builds a database, and then uses GIS functionality to provide information for understanding or solving a problem. Before GIS, analytic mapping was limited to a few kinds of organizations, such as city planning departments. Analysts did not have digital map layers, so they made hard-copy drawings on acetate sheets that could be overlaid and switched in and out, for example, to show before-and-after maps for a new facility such as a baseball stadium. With GIS, however, anyone can easily add, subtract, turn on and off, and modify map layers in an analytic map composition. This capacity has led to a revolution in geography and an entirely new tool for organizations of all kinds.

As figures 1.1 and 1.2 show, maps use symbols that are defined in map legends. Graphical elements of symbols include fill color, pattern, and boundaries for polygons; width, color, and type (solid, dashed, etc.) for lines; and shape, color, and outline for points. A GIS analyst does not apply symbols individually to features, but applies and renders a layer at a time based on attribute values associated with geographic features.

For example, given a code attribute for schools with values public, private, and parochial, a GIS analyst can choose a green, circular, 10-point marker for public schools; a blue, square, 8-point marker for private schools; and an orange, triangular, 8-point marker for parochial schools. Those three steps render all schools in a map layer with desired point markers.

Similarly, we created the color-shaded county map layer in figure 1.1 based on an attribute that provides the lung cancer mortality rate of white males by county. A map that uses fill color in polygons for coding is called a choropleth map. In this case it shows an equal-interval numeric scale, rendered using a gray monochromatic color scale. The darker the shade of gray, the higher the interval of the numeric scale. By making selections and setting parameters, the GIS analyst accomplishes all of this coding and rendering with a simple graphical user interface.

Most organizations generate or collect data that includes street addresses, ZIP Codes, or other georeferences. GIS is able to spatially enable such data, that is, add geographic coordinates or make data records joinable to boundary maps. Geocoding, also known as address matching, uses street addresses as input and assigns point coordinates to address records on or adjacent to street centerlines, such as in the TIGER/Line street maps. Geocoding uses a sophisticated program with built-in intelligence—similar to a postal delivery person's when delivering your mail—that can interpret misspellings, variations in abbreviations, and so on.

Policy, planning, and research activities often require data aggregated over space and time, rather than individual points. For example, in a study of demand patterns for locating a satellite medical clinic, it may be desirable to aggregate patient residence data to counts per census tract or ZIP Code boundaries for a recent year. GIS has the unique capacity to determine the areas in which points lie, using a spatial join or overlay function, and this enables the analyst to count points or summarize their attributes (e.g., using sums or averages) by area.

This brings us to the last GIS capability to be described in this section: proximity analysis. As an example, we conducted a study to determine whether the decline in the use of senior centers was affected by the distance from the seniors' residences (Johnson, et al. 2005). We geocoded the senior centers that provided human services and where the senior citizens lived, placing them as points on a map, and included the total target population using census data by city block. Then we used ArcGIS to create buffers around the facilities with a radius of 0.5 miles, 1 mile, 1.5 miles, 2.0 miles, and so on. We used spatial joins to assign buffer identifiers to residence points and blocks, count clients by buffer area, and sum population by block group.

Next, through careful subtraction of counts and sums, we were able to get the total number of clients and the target population in each ring around facilities (e.g., 0.5 miles, 1 mile, 1.5 miles), calculate use rates for each ring by dividing the count of clients by the sum of target population, and plot the relationship of use rate versus mean distance of each ring from facilities. We found that the use rate by the clients, who were elderly users of the senior center facilities in our county, declined rapidly with distance, given their mobility limitations. The policy implication of this result is that senior centers need to be located in areas with high

densities of elderly populations. You will conduct similar buffer analyses in tutorials 7, 9, and 10 in a variety of health contexts.

### ***Tutorial 1: Introducing GIS and health applications***

Health care is a large, growing, and complex sector of economies around the world. The United States, like other countries, faces many challenges in providing the best health care for its citizens and at the lowest possible cost. Currently many health informatics systems are manual and/or nonintegrated (*The Economist* 2005). By nonintegrated, we mean that interacting organizations have information systems in-house that are not connected with each other in ways that allow sharing of data that could be beneficial. This has led to top-heavy administrations, high costs of transactions (Hagland 2004), medical errors, and duplication of efforts, such as unnecessary medical tests (Protti 2005). Additional health-policy issues arise, for example, from too much emphasis on treatment of sickness and not enough on prevention of illness (Kennedy 2004); enormous numbers of uninsured persons; the overuse of emergency rooms; nursing shortages; and lack of preparedness for bioterrorism (Featherly 2004).

One clear trend in health policy, research, planning, and management is the increasingly important role of health informatics. More systems will become automated and integrated, with large costs but even larger benefits. What does this mean for GIS applications? One consequence is that there will be even more data available for possible input to GIS—additional data on patients, facilities, programs, and events that include disease incidence, medical diagnostics, and treatments. Much of the additional data will have street addresses, ZIP Codes, or other location elements that will make them applicable to GIS processing.

This book has a sampling of health GIS applications that cut a wide path across the landscape we have just described. Next is a brief summary of the tutorials ahead.

### ***Tutorial 2: Visualizing health data***

First is a simple public health application for visualizing breast cancer mortality rates at the state and county levels across the United States. In the tutorial, you will work with breast cancer data and lung cancer data. While valuable for providing a snapshot of cancer mortality, the primary purpose of the application is to get you comfortable with map navigation in ArcGIS. ArcGIS provides many ways to change scales and views of a map in search of information, as well as ways to work with the data records behind the map features. You will get experience with some of them in this tutorial.

### ***Tutorial 3: Designing maps for a health study***

In tutorial 3 we will learn about uninsured populations in a state and their health-care financial needs. In the tutorial, you will analyze where to locate programs that provide health-related financial support for uninsured populations in Texas counties. You will use an advanced map with a bivariate analysis to contrast the magnitude of uninsured populations with measures of poverty by county.

GIS can provide many kinds of outputs. In this tutorial you will build stand-alone map layouts, with components such as multiple map frames, legends, and scale bars to use in presentations and reports. In the exercises you will use U.S. Census data to explore neighborhoods for areas in which to promote walking for reducing obesity.

#### ***Tutorial 4: Projecting and using spatial data***

What are the most appropriate ways to map disease incidence (numbers of cases) versus prevalence (numbers of cases per population of 10,000). In the first health application you will visualize a communicable disease, HIV infection, and AIDS. You will build two kinds of map layers, one for disease incidence based on polygon centroids (points) using size-graduated point markers and the other for disease prevalence represented by fill color for polygons to encode data.

A second application further pursues the problem of child obesity. Here your objective will be to identify green spaces in the vicinity of schools for possible use in physical education programs.

GIS issues pursued for these applications involve importing spatial data into ArcGIS and projecting map layers for the application and geographic scale at hand. The exercises in this tutorial address the same GIS issues, but move into an entirely new application area—preparation of emergency evacuation routes to medical facilities and shelters.

#### ***Tutorial 5: Downloading and preparing spatial data***

Where are the concentrations of a city's older houses that are likely to have lead-based paint? Do children who have elevated levels of lead in their blood live in those areas? The purpose of this study is to identify clusters of children with elevated levels for targeting screening programs. You will work with blood samples and census data on housing built before 1970, when lead-based paints were still used.

In this tutorial you will download and prepare U.S. Census data. You must clean up the data by renaming variables and deleting rows that do not conform to data table formats, modifying census tract identifiers in the table so that they match comparable identifiers in the census tract map layer, and joining them to a downloaded census tract boundary map. Ultimately, you will produce a very nice bivariate map using choropleth and dot density displays. In that map, you will place randomly located points within polygons in proportion to an attribute of interest. The exercises in this tutorial seek additional explanation of observed clusters by using additional census variables.

#### ***Tutorial 6: Geocoding tabular data***

General spatial information is available from basemaps, but how do you get your organization's data for points on a map? Data of interest often includes point locations of patients' or clients' residences, health care or other service delivery locations, such as the scenes of traffic accidents. If data includes street addresses, ZIP Codes, or other spatial identifiers, then GIS has the tools to plot points of interest to use for analysis.



For this case study, you need to spatially enable some data currently in tabular form. You will geocode existing facilities within a county so they can be placed on a map as points. You also will place patients on a ZIP Code map over a wide area. With this data mapped, you can readily see potential service gaps. Then, in an exercise, you will map some suitability measures for a health clinic location, to aid in identifying locations in gap areas that would be an attractive site for a new facility.

### ***Tutorial 7: Preparing and analyzing spatial data***

What are some neighborhood factors leading to child pedestrian injuries in a city? Is poverty a factor? What about the lack of safe public areas for play? In this tutorial you will explore the determinants of serious child pedestrian injuries for the purpose of designing prevention programs. The basis of your study is a sample of serious injury data that has been geocoded and can be compared to census data on poverty and map layers for streets, neighborhoods, and parks that have playgrounds and playing fields.

The GIS work includes preparatory steps for extracting study-region maps from county maps and then focuses on detailed proximity analyses using park buffers, like those seen in figure 1.2.

### ***Tutorial 8: Transforming data using approximate methods***

How can health-care analysts combine data from different, incompatible polygon boundary sets? Often the spatial unit of analysis for a health study will be a custom set of polygon boundaries, designed for the phenomenon at hand. An example is the hospital service areas and hospital referral regions of the Dartmouth Atlas of Health Care Project of the Center for the Evaluative Clinical Sciences at Dartmouth Medical School ([www.dartmouthatlas.org](http://www.dartmouthatlas.org)) in Hanover, New Hampshire. While appropriate for studying patterns in the quality of health care across the country, these custom areas have the limitation of not sharing boundaries with census statistical areas (i.e., they are noncoterminous sets of boundaries). Thus, it is not possible to directly use census data for supportive analysis of the custom areas.

Another common case is regional analysis in which individual records have only ZIP Codes as spatial identifiers, so ZIP Codes become the *de facto* unit of spatial analysis. Detailed census variables, from U.S. Census Summary File 3 on income, poverty, educational employment, and so on, are not easily available at the ZIP Code level. Fortunately, advanced GIS functionality, using spatial joins, can produce some very accurate approximations (i.e., apportionments) for transforming data from one set of polygons to another incompatible set.

### ***Tutorial 9: Using ArcGIS Spatial Analyst for demand estimation***

This tutorial is an introduction to the Spatial Analyst extension for ArcGIS Desktop. Spatial Analyst uses or creates raster datasets comprised of grid cells to display data that is distributed continuously over space as a surface. In this tutorial you will prepare and analyze a demand surface map for the location of heart defibrillators in Pittsburgh, with demand based on the number of out-of-hospital cardiac arrests with potential bystander help. You will also learn how to use Spatial Analyst's raster calculator to create a poverty index surface combined with several census data measures from block and block group polygon layers.

### ***Tutorial 10: Case study: Studying food-borne disease outbreaks***

Finally, this tutorial and tutorial 11 provide a change of pace—opportunities to apply and extend the GIS skills and health applications you have learned in the previous nine tutorials to new case studies. We provide source data and guidelines for analysis, with a broad outline of steps; however, it is up to you to carry out the GIS work somewhat independently. In this tutorial, you will prepare map layers, which includes geocoding, as the basis for analyzing outbreaks of food-borne illnesses. Then you will use data to simulate the impact of an outbreak. You will do a proximity analysis based on patterns in reported disease cases.

### ***Tutorial 11: Case study: Forming a national ACHE chapter***

The book concludes with a second case study, following a similar setup to that in tutorial 10. Staff members of the American College of Healthcare Executives (ACHE) want you to use GIS to help them set up ACHE chapters across the country that provide educational and other services to health-care professionals.

You will perform a buffer analysis of existing affiliates that propose to become ACHE chapters. The buffers will help determine territories served and gaps that suggest where new chapters should be established. You will do some work interactively using ArcGIS, but then for steps that must be done repeatedly over time, you will build an ArcGIS model that generates a macro to automate steps.

### ***Introduction to ArcGIS and map documents***

ArcGIS Desktop is available in three license levels: ArcView, ArcEditor, and ArcInfo. All three of the licensed ArcGIS Desktop products look and work the same, though they differ in how much they can do.

This book is designed for use with ArcGIS Desktop 9.3 software with an ArcView license. ArcView, the most popular member of this collection and most widely used GIS package in the world, is a full-featured GIS software application for visualizing, managing, creating, and analyzing geographic data. ArcEditor adds more GIS editing tools to ArcView. ArcInfo is the most comprehensive GIS package, adding advanced data conversion and geoprocessing capabilities to ArcEditor. While this book was written for ArcView, you can use it with ArcEditor and ArcInfo as well.

Also available as part of ArcGIS is the free ArcReader mapping application, which allows users to view, explore, and print maps. Finally, ArcGIS has numerous extensions that can be added to ArcView, ArcEditor, and ArcInfo. Some major extensions include ArcGIS 3D Analyst for three-dimensional rendering of surfaces, ArcGIS Network Analyst for routing and other street network applications, and ArcGIS Spatial Analyst for generating and working with raster maps.

ArcView consists of two application programs, ArcCatalog and ArcMap. ArcCatalog is a utility program that has file browsing, data importing and converting, and file maintenance functions (such as create, copy, and delete)—all with special features for GIS source data. For managing GIS source data, you will use ArcCatalog instead of the Microsoft Windows utilities My Computer or Windows Explorer.

GIS analysts use ArcMap to compose a map from basemap layers, and then they can carry out many kinds of analyses and produce several GIS outputs. A map composition is saved in a map document file with a name chosen by the user and the .mxd file extension. For example, you will soon open Tutorial1A.mxd.

A map document stores pointers (paths) to map layers, data tables, and other data sources for use in a map composition, but does not store a copy of any data source. Consequently, map layers can be stored anywhere on your computer, local area network, or even on an Internet server, and be part of your map document. In this book, you will use data sources available from the data CD accompanying the book that you will install on your desktop computer's hard drive.

### ***Installing ArcView and the health tutorial data CD***

This book includes a DVD with the 180-day trial version of ArcGIS Desktop 9.3. See appendix C for instructions on installing ArcView and the data CD, also accompanying the book. You must successfully install ArcView and the data to complete the tutorials in this book.

# Introducing the ArcGIS user interface

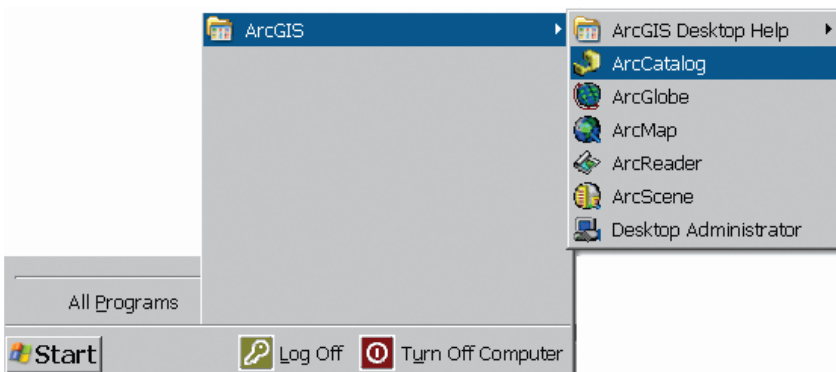
The following steps will acquaint you with the functionality and user interfaces of ArcMap and ArcCatalog. You will start by using ArcCatalog to browse the data sources used in figure 1.1 (see p. 2), and then examine the completed project itself. You will learn how to build, modify, and query data in the remaining tutorials.

In the exercises that follow, you need to be at your computer to carry out the numbered steps. Screen captures accompanying the steps show you important dialog boxes and output. Occasionally we have added “Your Turn” exercises after a series of steps. It’s critical that you do these brief exercises to start internalizing the processes covered.

## Launch ArcCatalog

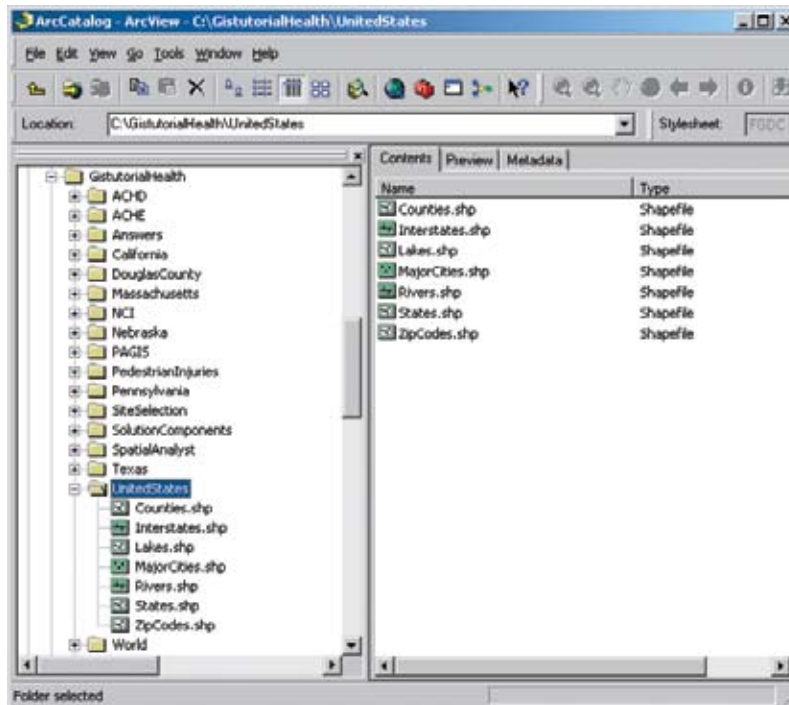
- 1 From the Windows taskbar, click Start, All Programs, ArcGIS, ArcCatalog.

Depending on how ArcGIS has been installed, you may have a different navigation menu or a name other than ArcGIS.



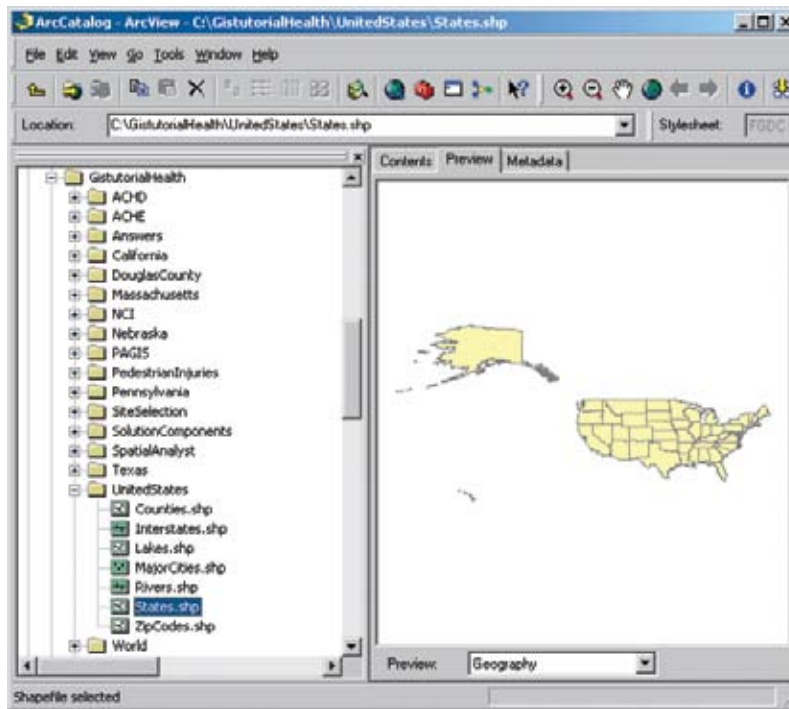
- 2 The left panel of ArcCatalog is called the Catalog tree. It is used to navigate to the data on your computer or network server, much like Windows Explorer. In the Catalog tree, navigate to your **GistutorialHealth** folder by clicking the small plus signs to expand the folders in which it’s installed. (The default location for this folder is **C:\GistutorialHealth**, but it may be in a different location depending on where you or your instructor installed it. If you cannot find this folder, make sure it was installed as described in appendix B.) After navigating to the **GistutorialHealth** folder, expand the United States folder located inside it.

The right panel, called the Catalog display, contains three tabs: Contents, Preview, and Metadata. When you choose the Contents tab, the datasets in the current folder are listed. The datasets currently listed represent spatial data, and the icon next to each file name indicates what type of geometry the data is built with: point, line, or polygon.



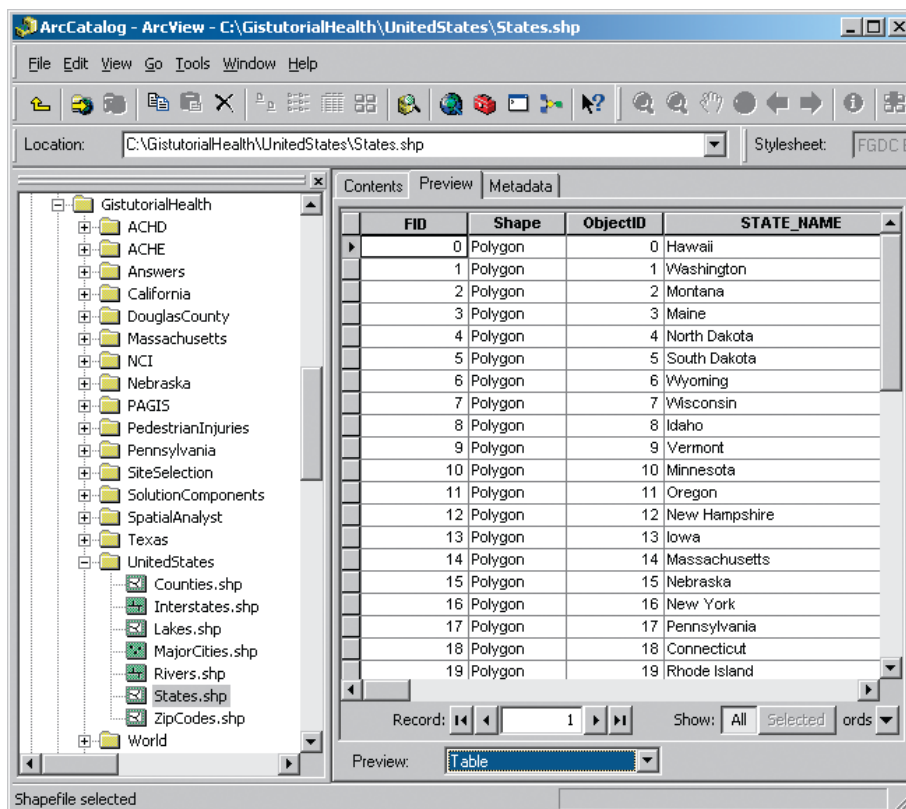
- 3 In the Catalog display, click States.shp, then click the Preview tab at the top of the right panel.

Previewing data this way allows you to get a quick glimpse of the data without actually loading it into a map. You can also use this tab to preview the contents of a table.



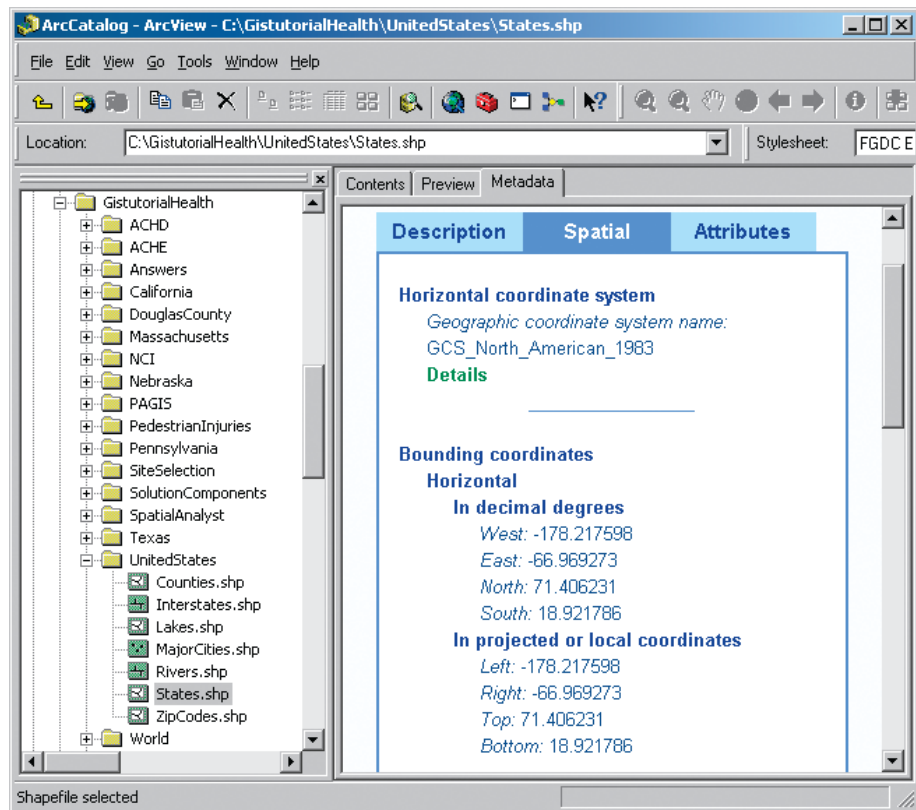
- 4 At the bottom of the Catalog display, click the Preview drop-down arrow and click Table, then use the horizontal scroll bar to view the attribute fields in the table.

Each record in the table corresponds to one of the state polygons you previewed in the previous step, and as you can see, there are quite a few attributes stored for each state, most of which are demographic. For example, by reading across the table, you could identify that the state of Washington is in the Pacific subregion and had a population of 5,894,121 in 2000. We used the STATE\_NAME attribute to label states in figure 1.1.



- 5 Click the Metadata tab at the top of the Catalog display and click the Spatial tab in the resulting display. (If the Spatial tab does not display, make sure the Stylesheet drop-down list is set to FGDC ESRI.)

Metadata describes data; it is data about data. For example, you can see that States has a geographic coordinate system (latitude and longitude) with certain bounding coordinates for the rectangle framing the map layer. Also available are descriptions of the data and a list of attributes with their data types, under the Description and Attributes buttons.



6 Click the Contents tab in the Catalog display.

## YOUR TURN

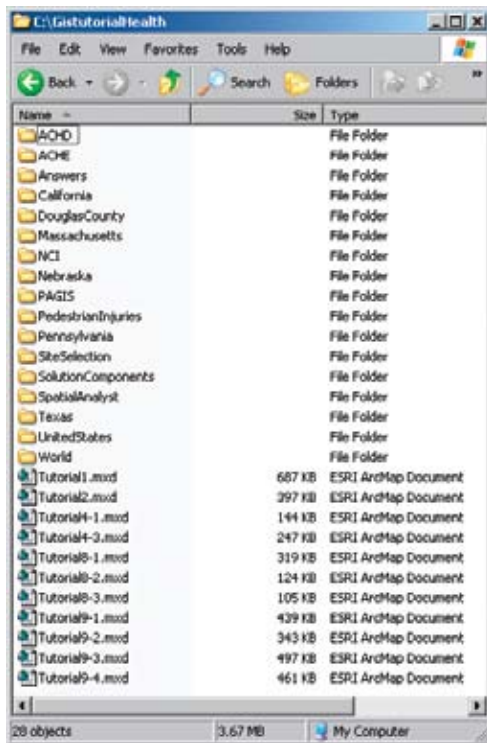
Explore additional layers in the United States folder. When finished, click the contractor button with the minus sign to the left of the United States folder icon in the Catalog tree. Then close ArcCatalog.

### Review data source types

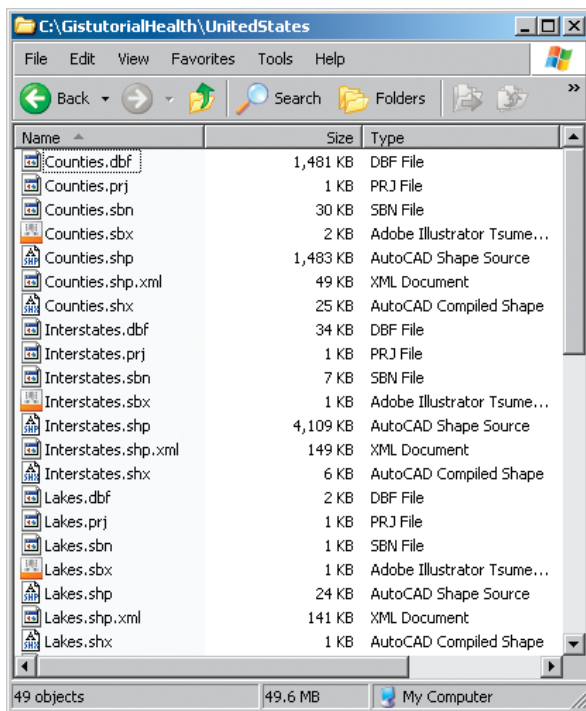
You will do most GIS file maintenance work in ArcCatalog, though it is instructive to view GIS files in a conventional browser. Next you will examine two common ESRI file formats used in GIS: a shapefile and a personal geodatabase. A shapefile map layer has three or more files with the same name but different file extensions, all stored in the same folder. A personal geodatabase is a Microsoft Access database file (.mdb extension) that has one or more map layers stored as relational database tables. Both of these data types are very common in the GIS industry, with the geodatabase being a more modern form.

1 From the Windows taskbar, click Start, My Computer. (The path to My Computer may differ depending on which operating system you are using.)

- 2 Browse to where your **GistutorialHealth** folder is installed (e.g., C:\GistutorialHealth).



- 3 Double-click the **United States** folder to view its shapefiles.





Let's review the MajorCities shapefile, which actually consists of seven files, all with the name MajorCities.

- The MajorCities file with the .shp extension has the features' geometry and coordinates. In this case, each record has a point and an x,y location. For line and polygon layers, each shape record has coordinates of a line segment or a polygon.
- The MajorCities file with the .dbf extension has the feature attribute table in dBASE format. This file can be opened and edited as a Microsoft Excel spreadsheet and an Access database, but such work must be done carefully and without deleting or adding records or changing the order of rows. This could result in corrupted data. The relationship between the .shp and .dbf files of a shapefile depends on one-to-one physical arrangement of records in both files.
- The .sbx and .shx files contain indexes for speeding up searches and queries. The .prj file is a simple text file that has the map projection parameters of the layer.
- Finally, the .shp.xml file has the layer's metadata and can be opened in a Web browser for reading.

**4 Click View, Go To, Up One Level in My Computer.**

**5 Click the Solution Components folder icon, then click the Chapter 5 folder.**

This folder contains a personal geodatabase, LeadStudy.mdb. A single personal geodatabase can have one or more vector map layers and data tables. Never open a personal geodatabase in Microsoft Access! Always use ArcCatalog to maintain map layers and data tables stored in a personal geodatabase; otherwise, you will likely corrupt the layers.

**6 Close My Computer.**

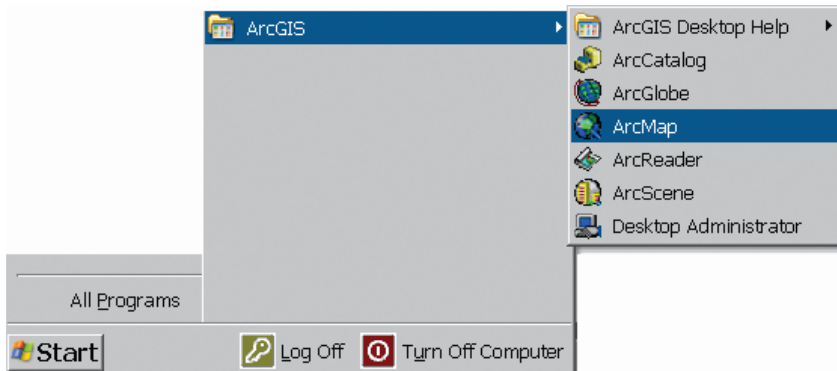
## YOUR TURN

Launch ArcCatalog and explore the personal geodatabase in the **\GistutorialHealth\SolutionsComponents\Chapter5** folder. Preview its tables and map layer. Then close ArcCatalog.

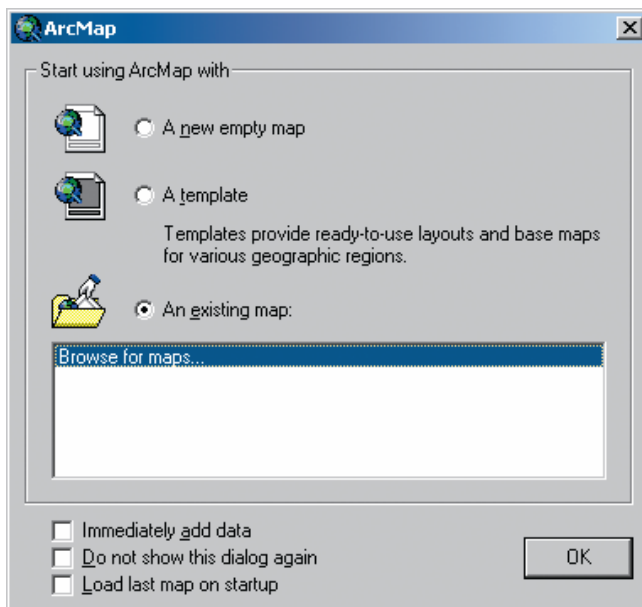
## Launch ArcMap

- 1 From the Windows taskbar, click Start, All Programs, ArcGIS, ArcMap.

Depending on how ArcGIS and ArcMap have been installed, you may have a different navigation menu or a name other than ArcGIS.



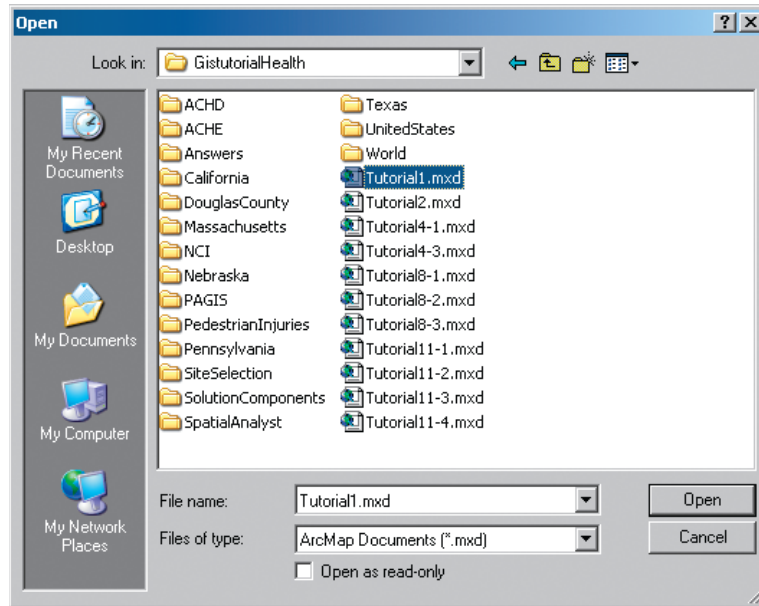
- 2 In the resulting ArcMap window, click the An existing map radio button, then click OK.



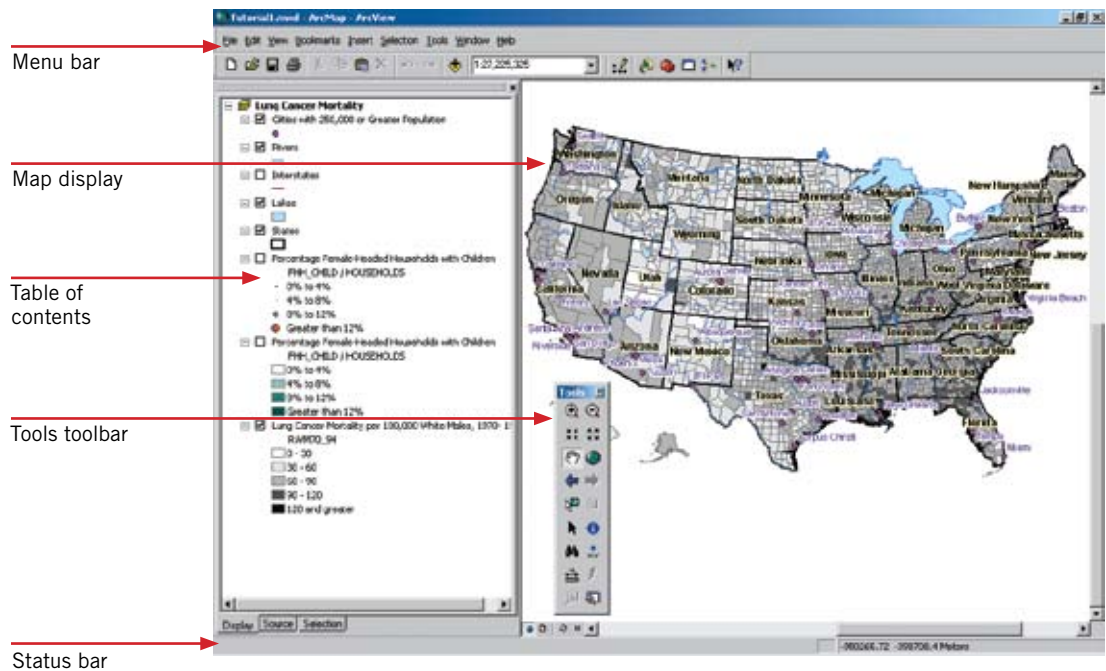
- 3 Browse to the drive and folder in which the **GistutorialHealth** folder is installed (e.g., **C:\GistutorialHealth**).

4 Click the **Tutorial1.mxd** map document icon, then click **Open**.

Note: The .mxd extensions may not be visible, depending on how your operating system is configured.



Tutorial1.mxd opens in the ArcMap application window. If you do not see the table of contents, click **View, Table of Contents** to add this panel. The map currently contains the contiguous United States showing demographics related to female-headed households and lung cancer mortality rates. The major components of the ArcMap interface are also identified.



Let's review the major components of the ArcMap application window.

- The map display is where the datasets loaded into the map are drawn.
- The table of contents lists all the data in the map document and allows you to toggle their visibility and access their properties. Currently, the Display tab is selected at the bottom of the table of contents showing the order of map layers and their legends. By clicking the Source tab at the bottom of the table of contents, you can also see the folder and file path to the source data of map layers and tables. ArcMap draws maps from the bottom up in the table of contents, so big feature layers, such as Lung Cancer Mortality, that would cover smaller feature layers, such as Rivers, must go on the bottom. You can see that only some of the layers are turned on, those with checked boxes.
- The Tools toolbar has frequently used tools and can be docked, if desired, by dragging it by its blue top to any boundary in the ArcMap interface. Similarly, you can undock it by dragging and dropping it to the desired location.
- The menu bar has some items common to most Windows application packages, plus some unique to GIS.
- The status bar displays the map coordinates of the cursor location in the map display. In the graphic on the previous page, the cursor is currently over Omaha, Nebraska, and the coordinates for Omaha are being displayed on the status bar.

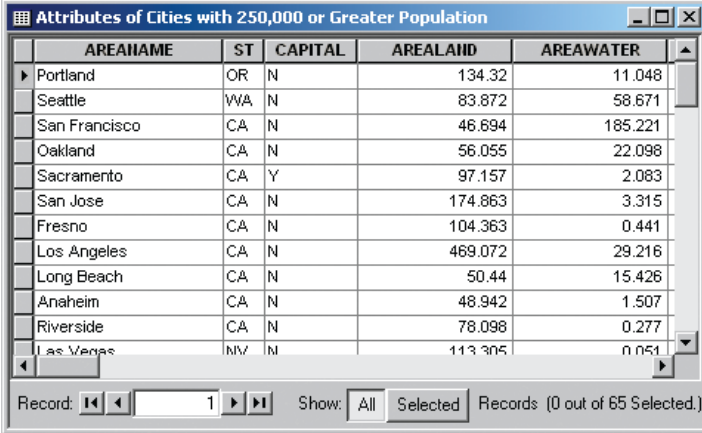
## YOUR TURN

Experiment with turning map layers on and off by checking and unchecking the boxes to the left of each map layer in the table of contents. Start with layers currently turned off, from the top down, and turn each one on and then off. Then try various combinations. Keep in mind that the Female-Headed Households with Children layer is a strong indicator of poverty. Of course, not every such household is in poverty, but the tendency is strong. When finished, return the map document to its original condition, with layers turned on and off as seen on the previous page.

## View map layer attribute tables

- 1 In the table of contents, right-click Cities with 250,000 or Greater Population. Select Open Attribute Table.

Attributes of Cities with 250,000 or Greater Population opens. This is the same sort of table that you previewed in ArcCatalog.



AREA NAME	ST	CAPITAL	AREALAND	AREAWATER
Portland	OR	N	134.32	11.048
Seattle	WA	N	83.872	58.671
San Francisco	CA	N	46.694	185.221
Oakland	CA	N	56.055	22.098
Sacramento	CA	Y	97.157	2.083
San Jose	CA	N	174.863	3.315
Fresno	CA	N	104.363	0.441
Los Angeles	CA	N	469.072	29.216
Long Beach	CA	N	50.44	15.426
Anaheim	CA	N	48.942	1.507
Riverside	CA	N	78.098	0.277
Las Vegas	NV	N	113.305	0.051

- 2 Visually scan the map for a few cities listed in the table, verifying that there is a point plotted for every record in the table.
- 3 Close the table.

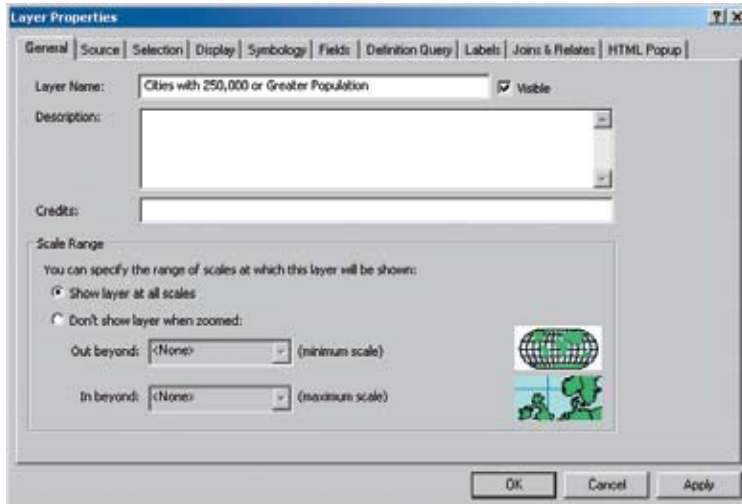
## YOUR TURN

Open a few other tables and scan the contents.

## View map layer properties

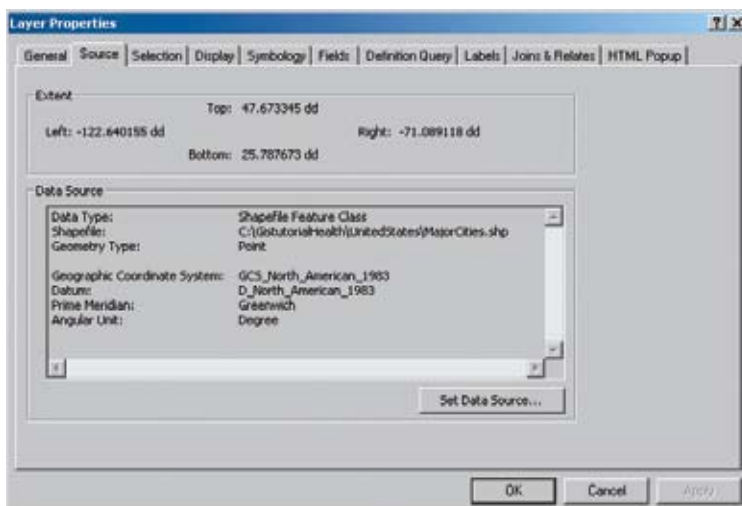
- 1 In the table of contents, right-click **Cities with 250,000 or Greater Population** and in the resulting context menu, click **Properties**. In the resulting **Layer Properties** dialog box, click the **General** tab.

The **Layer Properties** is an important dialog box for managing the way layers behave and appear in ArcMap. We will not attempt to describe all the properties that can be set, but will just review useful ones to know about for now. For example, on this panel of the **Layer Properties** you can change the layer name by typing the name of your choice into the **Layer Name** text box.



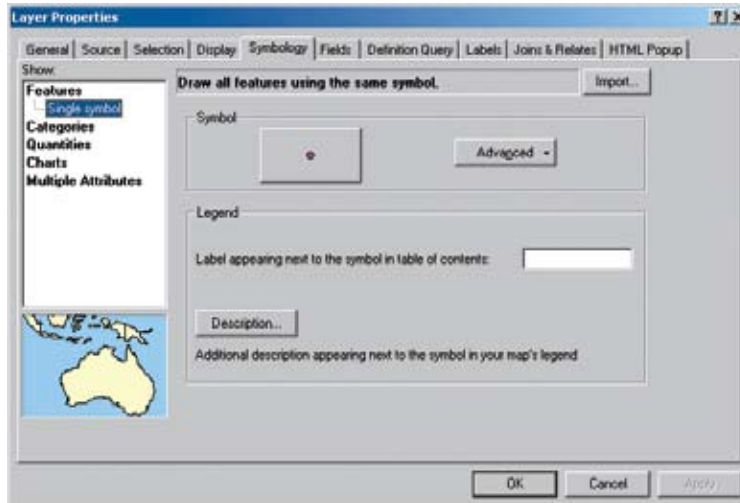
- 2 In the **Layer Properties** dialog box, click the **Source** tab.

This window provides bounding coordinates of the map layer (i.e., its extent), the data type, location on the server or computer, geometry type, and projection.



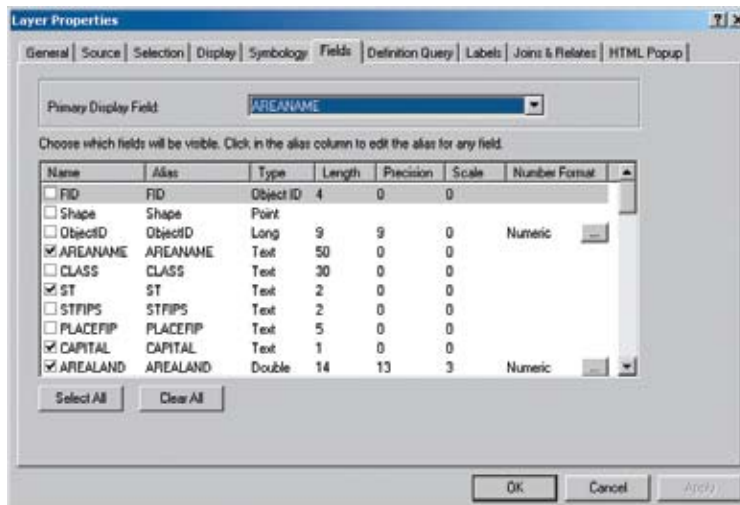
### 3 Click the Symbology tab.

This tab provides many options for symbolizing vector map layers. The cities layer uses a single symbol, a purple circular point marker with a black outline, chosen by the authors.



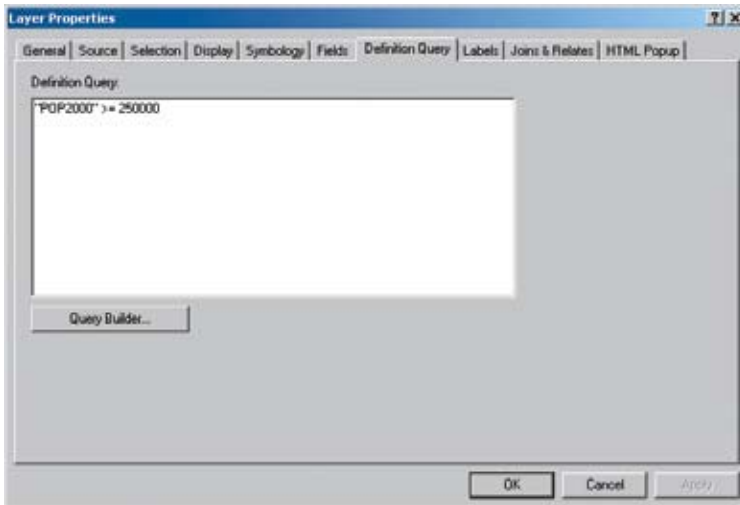
### 4 Click the Fields tab.

This tab lists the layer's attribute fields and their properties.



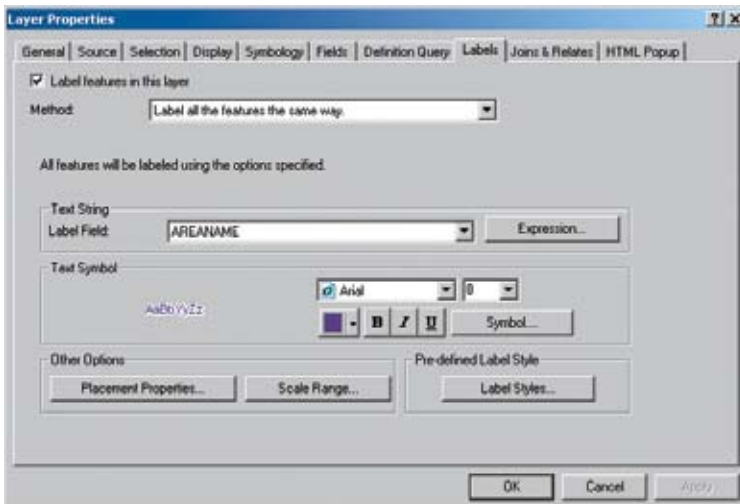
### 5 Click the Definition Query tab.

Use this tab to restrict what is displayed within a map layer. In this case, we used the POP2000 attribute to limit cities displayed to those with a population of 250,000 or more in the year 2000. This only changes the display and does not remove the records from the source data.



## 6 Click the Labels tab.

This tab allows you to label features, if you wish, using an attribute from the feature attribute table. We used AREANAME, which is the name of cities, with a purple font and white halo mask. We also turned labeling on with the check box at the top of the form.



## 7 Close Layer Properties.

# YOUR TURN

Examine the properties of a few more layers, but do not make any changes. When finished, close ArcMap.



## ***Use the map***

Fundamentally, when you use GIS for analysis, you observe spatial patterns related to a problem, phenomenon, or issue of interest. In this case, the phenomenon is mortality of white males from lung cancer.

Clearly we can see strong spatial patterns with concentrations of mortality along the southeastern coast of the United States and other locations. We can recognize some correlates on the map, based on spatial arrangement alone. For example, the band of high mortality along the eastern border of Kentucky is in that state's coal mining belt. Perhaps the white males succumbing to lung cancer in that area tend to be coal miners.

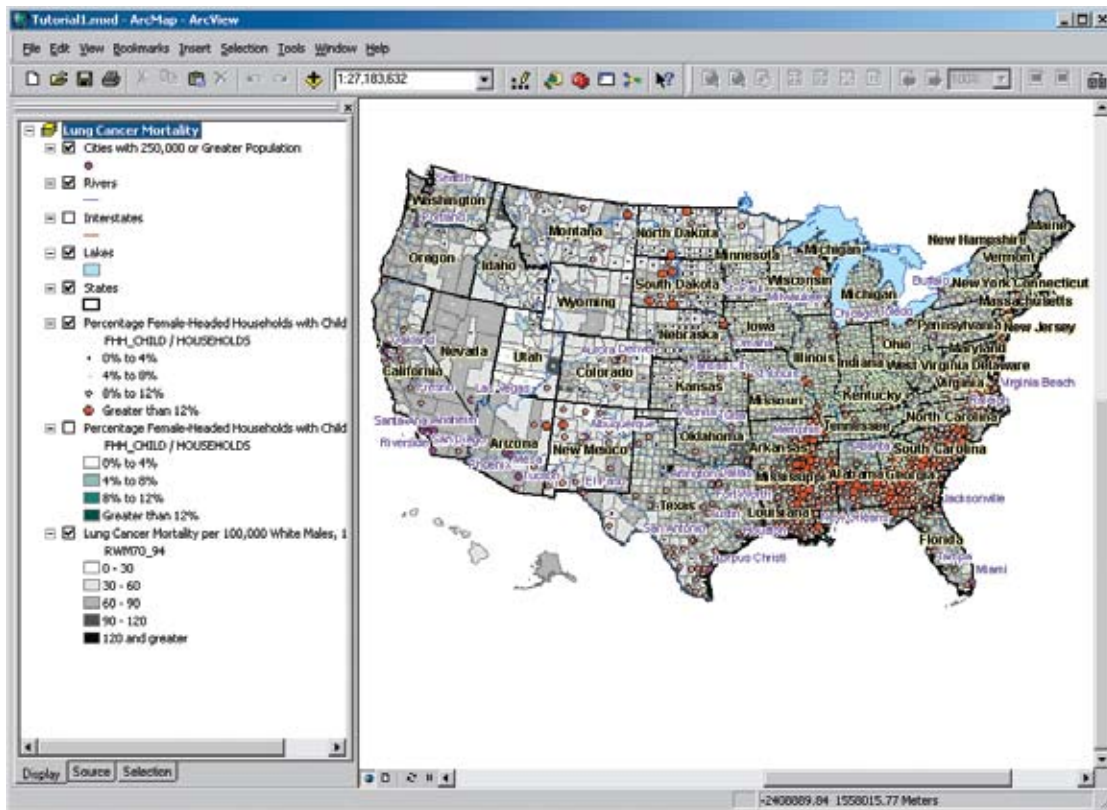
Some of the southeastern peak areas for mortality correspond to tobacco-growing areas of Virginia, the Carolinas, and Georgia, but high mortality areas run beyond the tobacco fields. Perhaps being near to tobacco-growing areas increases the likelihood of smoking and, therefore, lung cancer. A question then becomes: how far beyond the immediate tobacco growing areas could such an influence exist? Regardless, it cannot account for the high mortality observed along the Mississippi River.

An additional correlation is poverty, which contributes to many health problems and perhaps increases exposure to factors leading to lung cancer. A strong indicator of poverty, and the only poverty indicator available in our U.S. counties data, is the percentage of female-headed households with children. As an example of this indicator's effectiveness, in Pittsburgh at the tract level the simple correlation between percentage of female-headed households with children and percentage of total population below the poverty line is 0.67. Let's see if there is any visual evidence of a positive correlation of this poverty indicator with white male lung cancer incidence.

As you will see, we have used a size-graduated point marker and dichromatic color scale for this poverty indicator. We used an equal-interval numeric scale of 4 percent width, with the darker shade of green signifying the lowest value interval and the darker shade of orange signifying the highest value interval. Also, the larger the diameter of circular point marker used, the higher the interval of the indicator variable.

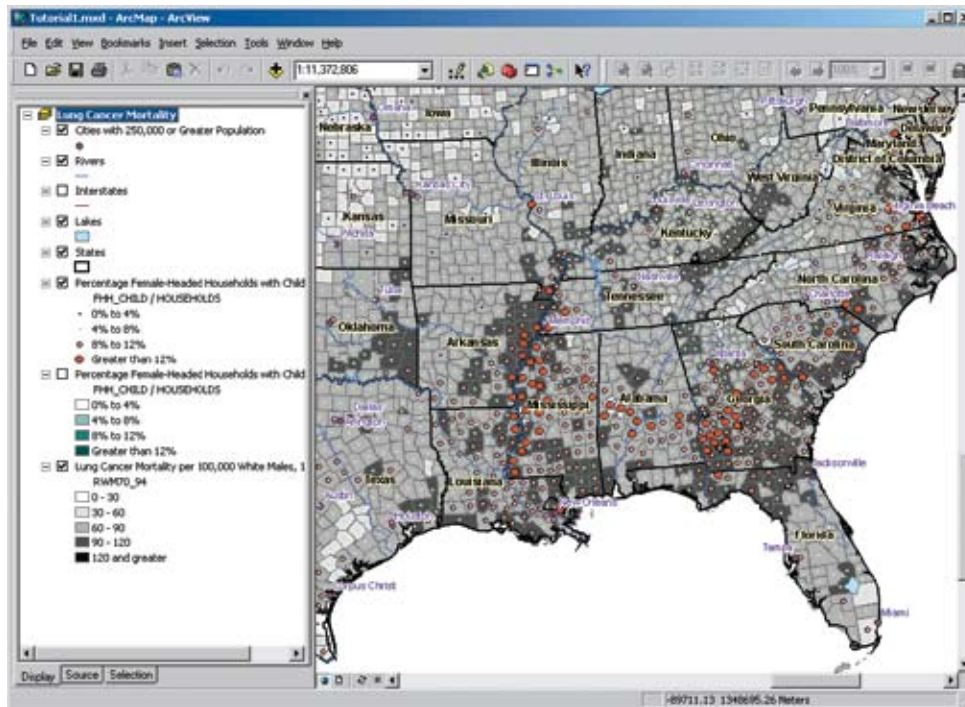
### **1 Open ArcMap and open Tutorial1.mxd. Check the box to the left of the Percentage Female-Headed Households with Children map layer in the table of contents to turn that layer on.**

Immediately, you can see that this variable is promising as a correlation to the male lung cancer deaths because it is high in many of the same areas where lung cancer mortality is high. The current map scale, however, is zoomed out too far. You need a closer look at the southeastern United States. We have built a spatial bookmark for this purpose, which you will use next, to get an uncluttered, zoomed-in view of the map. Notice that the map scale in the graphic on the following page is 1:27,183,632 for the entire United States. It will be different on your screen, because your screen is likely a different physical size than ours. Make a note of your scale.



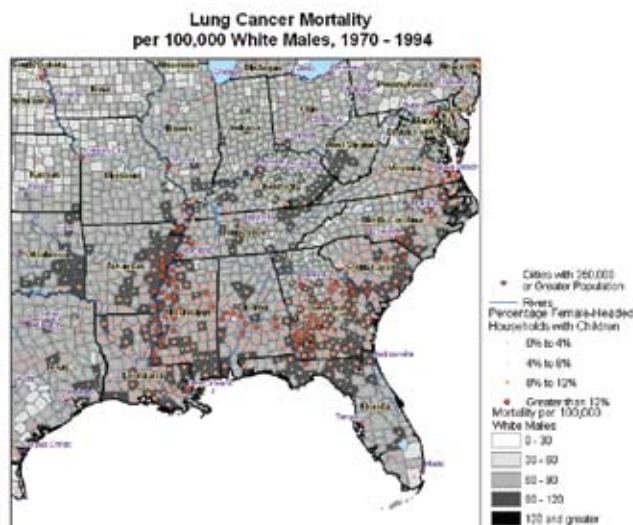
## 2 From the Bookmarks menu, click SouthEast.

That helped considerably. You can see that there is a visual correlation, with both variables tending to be high in the same areas. Clearly, further examination would require a multivariate model. We have identified some promising variables: coal production, tobacco crops, adjacency to tobacco-growing areas, and poverty. For the present, however, you need a professional-quality map that could be used in a Microsoft Office Word document or PowerPoint presentation. We have built a map layout for this purpose, which you will view next. Note that the map scale is much larger when zoomed in to 1:11,372,806, as shown in the following graphic.



3 From the View menu, click Layout View.

Below is the resultant map, which you will later learn how to export for use in a document. ArcMap provides two general views: data view and layout view. Typically you use data view to interact with your map data through browsing, symbolizing, and editing. You can also interact with your data in layout view, but its primary purpose is to finalize map compositions. In layout view you can add north arrows, scale bars, legends, and other map elements.



4 From the Bookmarks menu, click U.S.

5 Close ArcMap. Do not save your changes.

## Summary

GIS is a fascinating and valuable information technology, enabling spatial processing and visualization of data in ways never before possible. At the most basic level, GIS makes it easy to quickly compose and render maps from base layers. You can easily turn map layers on and off to study spatial patterns and correlations. You saw how GIS connects visual maps and underlying attribute data when you labeled map features and rendered numeric scales for variables using choropleth and graduated point marker maps. In later tutorials, you will discover additional, powerful uses for attribute data connected to map features.

Two components make GIS analysis possible for applications such as in health care. The first is the national map infrastructure—a collection of basemaps, census and other data, and geocodes for uniquely identifying areas. In the United States, this resource is a public good, paid for by tax dollars, and available for your use at no or nominal cost. Commercial map layers are also available at reasonable cost, and in many cases are essential for applications. This tutorial introduced important components of map layers, including geographic and rectangular coordinate systems, projections between the two types of coordinate systems, and map scale.

The second component necessary for applications is user-friendly and powerful GIS software. ArcGIS Desktop is the world's leading GIS software and the software used in this book. This tutorial covers the primary uses and interfaces of the two major ArcGIS application packages: the ArcCatalog utility for browsing and maintaining spatial data files, and ArcMap for creating maps and analyzing data. This tutorial asks you to examine the input map layers and GIS data for studying lung cancer mortality in white males. The spatial patterns you observed suggest some possible correlations of the mortality data, including poverty, tobacco growing, and coal mining.

We should probably add a third necessary component for successful GIS applications: GIS analysts. There is a lot to learn, and knowing how to use GIS is not a common skill. Nevertheless, this book will help you use GIS productively and efficiently, making the best use of your limited and valuable time.

## References

- Clarke, K. C. 2003. *Getting started with geographic information systems*. 4th ed. Upper Saddle River: Prentice Hall.
- Featherly, K. 2004. Battling bioterror. *Healthcare informatics online*. [www.healthcare-informatics.com/issues/2004/02\\_04/cover.htm#disease](http://www.healthcare-informatics.com/issues/2004/02_04/cover.htm#disease) (no longer available).
- Hagland, M. 2004. Harnessing efficiency. *Healthcare informatics online*. [www.healthcare-informatics.com/issues/2004/02\\_04/cover.htm#workflow](http://www.healthcare-informatics.com/issues/2004/02_04/cover.htm#workflow) (no longer available).
- Johnson, M., W. L. Gorr, and S. Roehrig. 2005. Location of service facilities for the elderly. *Annals of operations research* 136:1.
- Kennedy, E. May 13, 2004. Statement on the introduction of the health care modernization, cost reduction, and quality improvement act. [kennedy.senate.gov/~kennedy/statements/04/05/2004513C28.html](http://kennedy.senate.gov/~kennedy/statements/04/05/2004513C28.html) (no longer available).
- Protti, D. J. 2005. The use of computers in health care can reduce errors, improve patient safety, and enhance the quality of service-there is evidence. [www.connectingforhealth.nhs.uk/worldview/protti2](http://www.connectingforhealth.nhs.uk/worldview/protti2).
- The Economist*. 2005. Special report: IT in the health-care industry. April 30, 2005, 65.

## Exercise assignment 1-1

# Benchmark health GIS Web sites

### **Problem:**

Many health organizations make maps available online. Examples include the World Health Organization at [www.who.int/csr/mapping/en](http://www.who.int/csr/mapping/en), the Centers for Disease Control at [www.cdc.gov/nchs/gis.htm](http://www.cdc.gov/nchs/gis.htm), and the National Cancer Institute at [www3.cancer.gov/atlasplus](http://www3.cancer.gov/atlasplus). Note: Some URLs may change. If you cannot access the sites for the World Health Organization, Centers for Disease Control, or National Cancer Institute, try to find similar sites. The objective of this assignment is to investigate the unique capabilities of GIS and its applications to health by exploring these sites. Start by studying the above Web sites. Then browse the Internet to find an interesting health-related Web site that uses GIS (excluding the sites already listed above).

Provide an analysis of the Web site in a Microsoft PowerPoint presentation. Include the following in your presentation:

- Title page with title, Web site URL, and your name
- Purpose of the Web site and its use of GIS
- Screen captures of the Web site (about half a dozen) with maps
- A brief analysis of the GIS that includes:
  - GIS content provided
  - GIS functionality provided
  - Effectiveness in providing information
  - Ease of use
- Summary of what makes the Web site valuable in terms of functionality and content



### **What to turn in:**

If you are working in a classroom setting with an instructor, you may be required to submit the PowerPoint slides that you created in the exercise.

#### **PowerPoint slide**

C:\GistutorialHealth\Answers\Assignment1\Assignment1.ppt