

# Chapter 21

## Geographic Information Systems

Carol L. Hanchette

**Abstract** Geographic information systems provide powerful tools that can enable public health practitioners to analyze and visualize data and to make informed decisions in a timely and relevant manner. Since the publication of the first edition of *Public Health Informatics and Information Systems*, GIS has become increasingly more accessible and widely used. It has also become more powerful as new applications are developed and more spatial statistics are incorporated into GIS software programs. Many public health professionals—in epidemiology and disease surveillance, environmental health, and community assessment—are using GIS as a tool for analysis and decision-making. While the educational background of such professionals often does not include GIS, it is important for these GIS users to understand basic geographic and GIS concepts and to be able to interpret and critically analyze GIS maps created by others. Eventually, as such part-time GIS users become more familiar with the technology and its wide range of applications, they will go beyond mapping and begin to use GIS for more sophisticated forms of spatial analysis. However, GIS users must recognize that GIS is not a panacea; they must be aware of its limitations. Some of these limitations are tied to issues of map scale and the accuracy and completeness of available data; others concern the proper use of visualization and spatial analysis tools.

**Keywords** Geographic information systems • Topologically Integrated Geographic Encoding and Referencing • GIScience • Spatial data • Topology • Attribute data • Geocode • Geocoding • X-coordinate • Y-coordinate • Geographic coordinate system • Map projection • Scale • Vector • Raster • Point • Line • Polygon • Choropleth mapping • Address matching • Distance • Spatial query • Buffer functions • Overlay

---

C.L. Hanchette, PhD  
Department of Geography, University of Louisville,  
Louisville, KY 40292, USA  
e-mail: carol.hanchette@louisville.edu

analysis • Point-in-polygon overlay • Polygon overlay • Spatial join • Areal interpolation • Spatial statistics • Color ramps • Desktop GIS • Shapefile • Metadata • Equal interval • Quantile • Natural breaks • Mean • Standard deviation

### Learning Objectives

1. Describe how geographic information systems (GIS) can be used to analyze public health information.
2. Identify specific GIS functions that can be applied to health data analysis.
3. Explain the limitations of GIS software and spatial data.
4. Discuss the emerging technologies that have implications for GIS use in public health.

## Overview

Geographic information systems provide powerful tools that can enable public health practitioners to analyze and visualize data and to make informed decisions in a timely and relevant manner. Since the publication of the first edition of *Public Health Informatics and Information Systems*, GIS has become increasingly more accessible and widely used. It has also become more powerful as new applications are developed and more spatial statistics are incorporated into GIS software programs. Many public health professionals—in epidemiology and disease surveillance, environmental health, and community assessment—are using GIS as a tool for analysis and decision-making. While the educational background of such professionals often does not include GIS, it is important for these GIS users to understand basic geographic and GIS concepts and to be able to interpret and critically analyze GIS maps created by others. Eventually, as such part-time GIS users become more familiar with the technology and its wide range of applications, they will go beyond mapping and begin to use GIS for more sophisticated forms of spatial analysis. However, GIS users must recognize that GIS is not a panacea; they must be aware of its limitations. Some of these limitations are tied to issues of map scale and the accuracy and completeness of available data; others concern the proper use of visualization and spatial analysis tools.

This chapter describes what a geographic information system (GIS) is, how it works, and the contributions it can make to analysis and decision making in public health. Commonly-used functions and limitations are also discussed.

## Introduction

During the past few years, the contribution of information technology to the practice of public health has become increasingly apparent and has led to the emergence of the discipline of *public health informatics*, defined as “...the systematic application

of information and computer science and technology to public health practice, research and learning [1]”. Savel and Foldy [2] highlighted three functions of public health informatics: (1) the study and description of complex systems, such as disease transmission models; (2) innovative use of data collection and information to improve the efficiency and efficacy of public health systems; and (3) the implementation and maintenance of systems that achieve the first two functions. Geographic information systems hold the potential to make significant contributions to all three.

A *geographic information system* (GIS) is a computer mapping and analysis technology, consisting of hardware, software, and data, all of which allow large quantities of information to be viewed and analyzed in a geographic context. It has nearly all of the features of a database management system, with a major enhancement: Every item of information in a GIS is tied to a geographic location. Lasker et al. [3] identified three basic types of information needs essential to public health services: (1) data collection and analysis, (2) communication, and (3) support in decision-making. Geographic information systems have enormous potential to contribute to the analysis of population-based public health with their ability to support all three types of information needs. With geographic information systems, public health professionals can manage large quantities of information; map and model the distribution of diseases and health care resources; analyze the relationships among environmental factors, socioeconomic environments, and disease outcomes; determine where to locate a new hospital or clinic; and even make decisions about the development or implementation of health policy.

## The Importance of GIS and Its Contribution to Public Health

Many introductory texts on medical geography and epidemiology begin with a reference to John Snow, the London physician who mapped cholera cases in the Soho District of London during the cholera epidemic of 1854 [4]. Snow was able to show that these cases clustered around the Broad Street pump. The closure of the pump, through the removal of the pump handle, and subsequent reduction in cases supported Snow’s contention that cholera was a water-borne disease.

Perhaps more interesting than Snow’s map, however, was his “medical detective” work preceding the 1854 epidemic and following the epidemic of 1849, which helped him to recognize the association between contaminated water and cholera. The cholera epidemic of 1849 killed over 52,000 people in Great Britain and over 13,000 in London alone [5]. While Snow published a brief account of this epidemic in 1849, he continued to carry out research over the next few years, leading to an 1854 edition of *On the Mode of Communication of Cholera* that was a more substantial work [6].

In that later account, Snow noted the association between cholera, poverty, elevation, and the water supply of the various London districts. A fascinating reconstruction, mapping, and geographic analysis of these associations was provided by Cliff and Haggett [5]. As the authors noted, “these associations result in some striking geographical distributions...” such as the higher mortality rates in areas adjacent to the River Thames and the relationship between cholera and water supplies of London Districts [5]. At that time, a number of metropolitan water companies

were supplying water to the city from a myriad of sources—some directly from the Thames, others from reservoirs. Cholera mortality was linked to contaminated water supplies provided by companies drawing their water directly from the Thames.

Today's technology makes it possible to carry out an analysis such as Snow's in a very small amount of time, at the desktop. Imagine Dr. John Snow at his desk with a powerful computer mapping and information system. On his computer screen, he has maps of London districts, their water supplies, and the locations of cholera cases. In addition, his water supply map database contains information about characteristics of the water, such as pH factor and water source. He also has a map of soils, with information about their characteristics, and a digital elevation model. With the tools available in a geographic information system, Dr. Snow could do point mapping of cholera cases, calculate distances to water sources, and examine the relationship of cholera incidence to water source, water type, soils, and elevation.

Snow's work provides an indication of how a GIS can benefit public health practice. Medical geographers, epidemiologists, statisticians and health practitioners have been carrying out mapping and spatial analysis for centuries, but have been doing it "longhand," so to speak. Some of the classic geographic research on probability mapping [7], disease diffusion and modeling [8], the spatial organization of cancer mortality [9], cardiovascular disease [10], and the allocation of health services [11] would have benefited from the use of GIS, or, more specifically, from the combination of GIS and statistical analysis software—all used some combination of mapping, spatial analysis, and statistical analysis.

Obviously, GIS is needed for more efficient processing and analysis of geographic data. It is also needed to integrate public health data from a wide range of sources, to perform population-based public health analyses, and to provide sound information on which to base decisions. Geography is a great integrator: Nearly every entity of public health information is located somewhere in space, whether it be a county, a ZIP code, a dot on a map, a hospital room, or even a point within the human body. GIS provides a means of integrating all this information through a spatial referencing system.

GIS technology, then, has much to offer public health practitioners. Perhaps most importantly, the analysis and display of geographic data is an efficient and effective means of providing data for decision-making; for example, how to implement lead screening guidelines [12]. GIS also permits the development of new types of data, the establishment of data partnerships and data sharing, and the development of new methods and tools for use by public health professionals [13, 14]. An additional benefit of GIS is that it can be used for quality control of health datasets. Geographically-based logical consistency checks can be carried out to verify the accuracy of geographic identifiers in health datasets.

The use of GIS among public health professionals has been on the increase, but it is still not a mainstream activity. In 1999, two editions of the *Journal for Public Health Management and Practice* were devoted entirely to GIS applications [15, 16]. In 2002, the first volume of *International Journal of Health Geographics* was published. Cromley and McLafferty have now published two editions of their *GIS and Public Health* textbook [17, 18]. For nearly a decade, Environmental Systems Research Institute (ESRI), the developer of ArcGIS software, has sponsored

a “Health GIS” conference; in 2013, the Urban and Regional Information Systems Association (URISA) will host its fourth “GIS and Public Health” conference.

The activities outlined above illustrate both the current importance of GIS and its potential to contribute to the ongoing assessment of health status in a community within the context of the second essential service of public health, namely the capacity to detect, “. . .diagnose and investigate health problems and health hazards in the community” [19]. It is therefore essential that public health managers and front-line practitioners develop a deeper understanding of GIS. This includes an understanding of some of the limitations of GIS, as well as an appreciation for the tool’s extraordinary capacity to support both the analysis of data and the presentation of that data in a way that is often more intuitively comprehensible to policy-makers, practitioners and community groups than the presentation of bare statistics and facts.

## What Is GIS?

What is a geographic information system? Many definitions exist. Essentially, it is a system of computer hardware and software that allows users to input, analyze, and display geographic data. Clarke refers to GIS as (1) a toolbox, (2) an information system, and (3) an approach to science [20]. As a toolbox, GIS is a software package that contains a variety of tools for processing, analyzing, and visualizing spatial data. Public health professionals might use these tools to map infant mortality rates across a state, identify areas with underserved populations, maintain an infectious disease surveillance system, or model environmental exposures to toxic substances.

As an information system, a GIS consists of a series of databases that contain observations about features or events that can be located in space and, hence, mapped and analyzed. This component of GIS includes a focus on data structures. GIS also functions as a means of spatial data storage [21]. Information that previously was on physical maps now can be stored in digital format in a GIS.

In some circles, the meaning of GIS has shifted from “geographic information system” to “geographic information science,” sometimes referred to as *GIScience* [22]. *GIScience* refers to the science behind the technology – the disciplines and technologies that have contributed to the development of today’s GIS software. These disciplines include geography, cartography, geodesy, photogrammetry, computer science, spatial statistics, and a wide range of physical and social sciences.

## Theoretical Foundations and the Development of GIS

As a science, the theoretical roots of GIS lie in geography, cartography, and spatial analysis. Certain paradigms in the discipline of geography have had a strong impact on the development of GIS technology. In the mid-1950s, geography experienced a shift from integrated, regional science approaches to a paradigm that embraced logical positivism and the quantitative revolution. Logical positivism incorporated a

theory of knowledge that was based on empiricism (sensory experience) and required deductive instead of inductive reasoning and laws of probability. In geography, this involved a heavy use of mathematics and statistics. Emerging computer technology contributed to this shift by providing faster computations and a means of storing and retrieving vast quantities of information [23]. During this time, methods of spatial analysis that had been developed earlier in the century were automated, and many new spatial/statistical methods were developed. For example, Glick used the concept of spatial autocorrelation to examine cancer mortality in Pennsylvania [24]. Other schools of thought in geography, such as the landscape and human ecology schools, which focused on the relationships between humans and their environment, had an impact on the development of automated mapping techniques to store and map environmental information.

Many US federal government agencies were important to the evolution of GIS technology and the development of digital cartographic data, perhaps most notably the US Bureau of the Census. In 1967, the agency piloted the use of digital geographic files (streets and census blocks) for a pilot project in New Haven, Connecticut. These files, the Geographic Base File Dual Independent Map Encoding files (otherwise known as GBF/DIME files), were used in urban areas for the 1970 and 1980 censuses. Today, the most commonly used spatial data in the country are probably the US Bureau of the Census TIGER/Line files, usually referred to as simply TIGER (Topologically Integrated Geographic Encoding and Referencing system) files. These were first used in 1990.

In the late 1980s, the move away from mainframe computers toward workstation and PC technologies resulted in dramatic changes to GIS software and functionality. Most notably, software became increasingly easy to use with the development of graphical user interfaces and menu-driven systems, and large collections of digital datasets were developed for use with the software. Today, computer users with a day's training or less can easily begin using GIS. This facility of use has obvious advantages, but there are drawbacks as well. After all, geographic data are complex. Without a sound knowledge of basic geographic principles, data issues, and map design, it is easy for an uninformed user to make errors, to mislead, and to be misled.

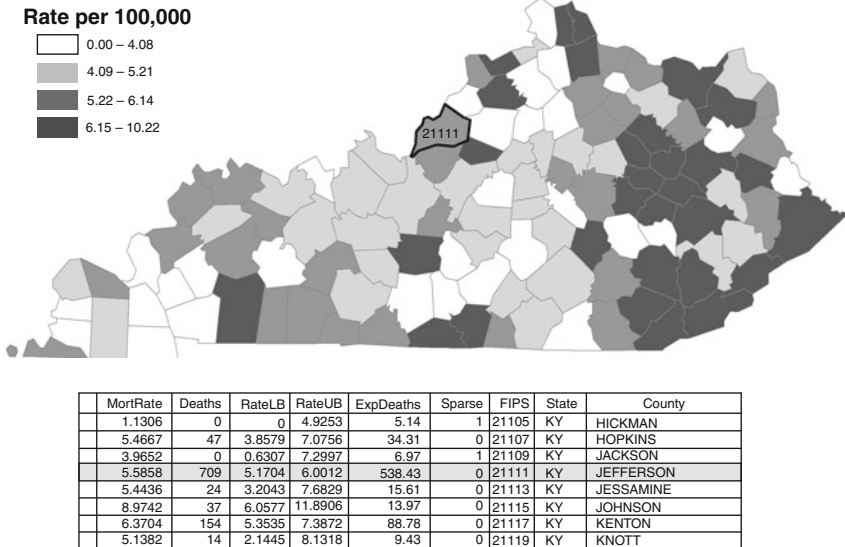
## How Do Geographic Information Systems Work?

Geographic Information Systems have several important concepts in common: the relationship between spatial and attribute data, map projections and spatial referencing systems, map scale, and spatial data representation.

### *Spatial and Attribute Data*

Although recent developments in hardware and database management software have led to the development of many new data structures, we can think of GIS data as having two components. The first is *spatial data*, consisting of geographic

**Cervical Cancer Mortality Rates for Kentucky Counties  
All Women, 1970–2004**



**Fig. 21.1** Spatial and attribute data for Kentucky counties. The record for Jefferson County is highlighted in the table, along with its corresponding map location. The FIPS code for Kentucky is 21, and the FIPS code for Jefferson County is 111, providing a combined FIPS code and unique identifier of 21111. This value is contained in the table’s FIPS field. The Kentucky county boundary file has a FIPS code associated with each county, and the attribute data are linked to the appropriate boundary through this geocode (Data Source: National Cancer Institute)

coordinates that provide information about the location and dimensions of features on earth and the relationships among these features. These spatial data are stored in a *topologic* data structure—a data structure that maintains information about the spatial relationships among features, such as adjacency, connectivity, and containment.

The second component is *attribute data*. Most people who use standard spreadsheet or database software think of these data as ‘columns’ or ‘fields.’ In other words, these are variables that describe the non-spatial aspects of the database, such as the total population of a given county, or its lung cancer mortality rate. Attribute and geographic data are linked through a *geocode*, a geographic identifier that is contained in both data components. This geocode can be a county name or a state name, a ZIP code, a street address, or some other numeric code. Standard numeric codes or geocodes for states and counties were developed by the National Institute of Standards and Technology (NIST) as part of the Federal Information Processing Standard (FIPS). Figure 21.1 displays a map of Kentucky showing cervical cancer mortality rates. The spatial data on the map are the Kentucky county boundaries. Attribute data are contained in the table below the map and are represented on the map by a series of shading patterns. Each record contains information for a single county; in this case, the information includes a county name, its FIPS code, and the cervical cancer mortality rate.

Most federal geographic data, such as census data, use a set of FIPS codes. However, the federal codes are not always used by state agencies or other organizations. Geographic files, such as the county boundary file in Fig. 21.1, often contain more than one set of geocodes. If health agencies in the state of Kentucky coded health data by county name, for example, these data could be mapped using county name as a geocode, so long as that information was also contained in a field in the spatial database and no county names were misspelled.

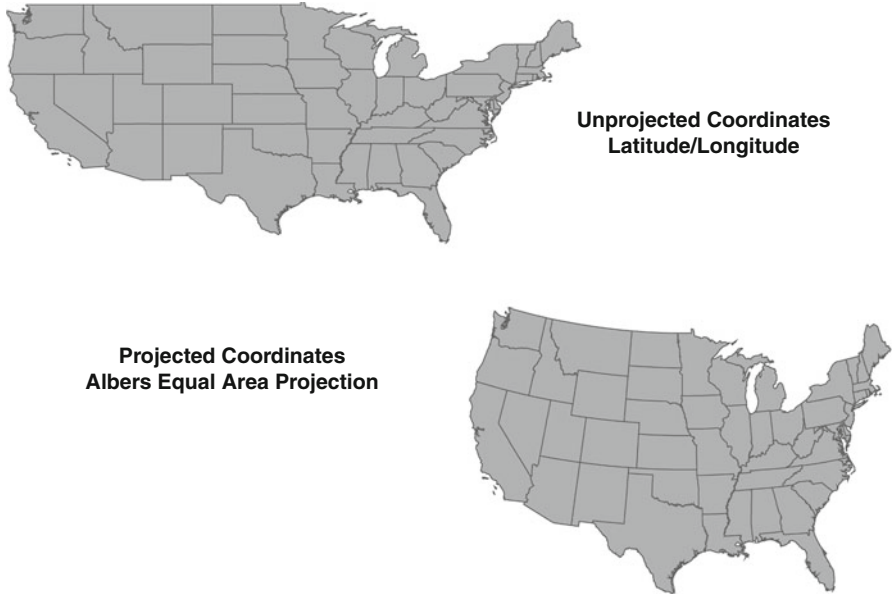
Attribute data come in a wide range of formats from a variety of sources. One of the challenges of using health and demographic data in a GIS is working with different data formats and structures. Attribute data are typically stored in tables, where columns represent fields or variables and rows represent cases or observations. These tables or files are often stored in a database. The original data may be stored in proprietary software such as Oracle (Redwood City, CA), SAS (Cary, NC), IBM SPSS (Armonk, NY), Microsoft (Redmond, WA) Access databases or Excel spreadsheets, or other formats such as delimited text files. Linking these data to spatial data usually requires importing them into the GIS software, so users must be knowledgeable about the native format of the GIS software they're using and which database formats can be imported. Spreadsheets and databases are not the same, and importing spreadsheets into GIS software can be problematic. For instance, spreadsheets can contain random text cells or column names that don't conform to database standards. Many GIS users view dBase (.dbf) as a preferred file transfer format because it is readable by many GIS software applications and requires little or no formatting. Recent developments in both GIS and database management software allow direct linkage between some GIS applications and database management systems.

For years, the main database management system utilized by GIS applications has been the relational model, where two or more tables can be linked easily via a common identifier, or key. This is the method by which attribute data are linked to spatial data using a common geocode. Recent trends in the larger GIS software applications are toward the use of object-oriented databases, which are capable of modeling complex spatial objects. These spatial objects contain not only attributes, but also the methods and procedures that operate on them.

## ***Map Projections and Coordinate Systems***

In a GIS, all geographic features, such as hospital location, county boundaries, and street networks, must be defined in terms of a common frame of reference, or coordinate system. Coordinates are defined by their distance from a fixed set of axes. In general, an *x-coordinate* refers to an east/west location; a *y-coordinate* defines a north/south location. Features on the earth can be located with the *geographic coordinate system*, which uses latitude for a north/south position and longitude for an east/west position. However, this system pinpoints location on a spherical earth. Maps and computer monitors, on the other hand, are flat. Therefore, the transformation of features from a three-dimensional sphere to a two-dimensional surface,





**Fig. 21.2** Unprojected and projected coordinates (Map Source: ESRI, Redlands, CA)

known as a *map projection*, must take place in order for the system to produce accurate mapping and analysis. Because degrees of longitude vary in actual distance across the globe (i.e., they converge at the poles), projections are used to establish a grid system with uniform units of measurement and to reduce the distortion in unprojected map coordinates.

Map projection is a science in and of itself. Projections are mathematical transformations of endless variety and, while they reduce the distortion inherent in geographic coordinates, they all involve some sort of distortion of shape, area, direction, or distance. Imagine drawing a map on the entire outside of an orange, then trying to remove and flatten the peel while maintaining the integrity of the map features. While it takes time and experience to learn which projections are best suited for a particular application, it is important for the new GIS user to understand that all map layers used in an application must use the same projection and coordinate system. Indeed, this is one of the strengths of GIS: Multiple map layers can be overlaid and relationships among them can be analyzed and displayed when they are tied to a common coordinate system.

Many geographic databases are stored as unprojected data—i.e., as latitude/longitude coordinates. These coordinates are a sort of *lingua franca*, a standard data exchange format, and must be projected using GIS software for more accurate analysis and visualization. Projections and/or coordinate systems commonly used in the US include (1) state plane coordinate systems, (2) Albers Equal Area projection, (3) Lambert Conformal Conic projection, and (4) Universal Transverse Mercator (UTM) projection. Figure 21.2 displays a map of the continental United

States in latitude/longitude coordinates (unprojected) and in Albers Equal Area coordinates (projected). More information about map projections can be found in Harvey [25].

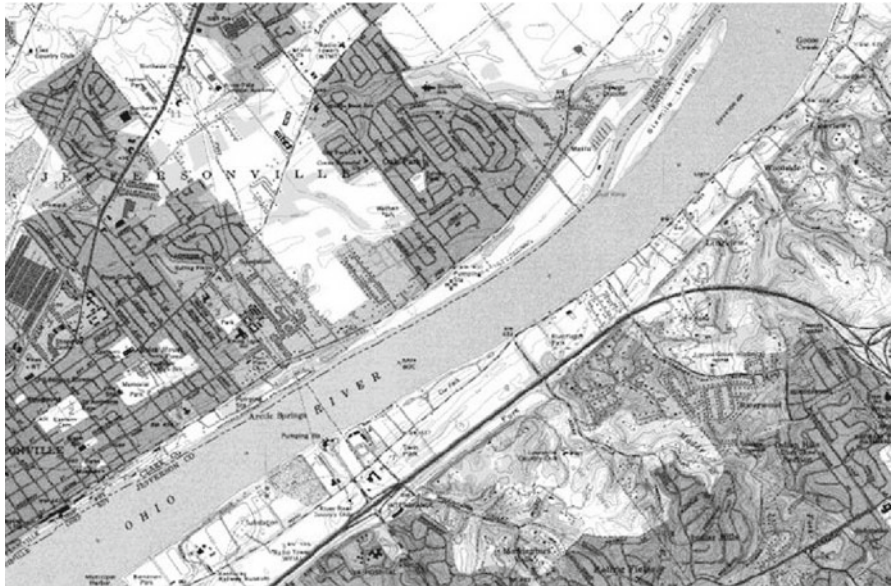
## ***Scale***

Scale refers to the ratio of a distance on a map to the corresponding distance on the ground. A scale of 1/100,000 (usually represented as 1:100,000) means that 1 inch on the map is equal to 100,000 inch on the real earth. The ratio is true for any unit of measurement (1 cm on the map is equal to 100,000 cm on the ground). Large-scale maps show more detail than small-scale maps. The concept of scale can be confusing because the larger the denominator in the fraction is, the smaller the scale is. In other words, a map at a scale of 1:12,000 is a larger-scale map than one at 1:2,000,000. Smaller-scale maps are generally used to show a larger area (such as the world or the US), whereas larger-scale maps can be used to “zoom in” to a smaller area (such as a city or a neighborhood). Because many map details are lost in smaller-scale maps, scale has an important effect on the precision of location. Figure 21.3 shows an area of Louisville, Kentucky, represented at two different scales. It is important to remember that although GIS software allows users to zoom in and out to different scales, the amount of detail in a map depends entirely on the scale of the source map!

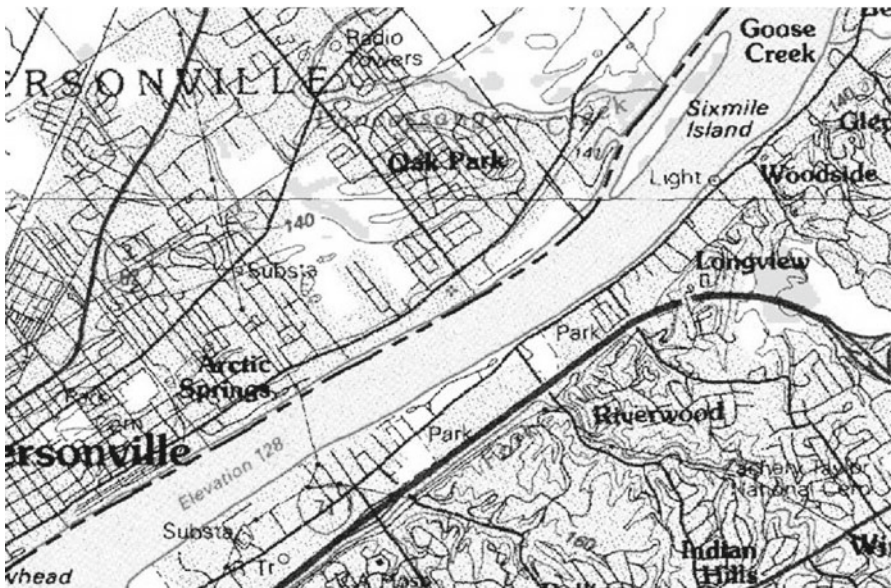
## ***Representations of Spatial Data***

Most spatial data in a GIS are either feature-based or image-based, often referred to as *vector* or *raster*, respectively. Vector data are represented by feature types that resemble the way we visualize and draw maps by hand—by use of (1) *point*, a single  $x, y$  location (example: a residence); (2) *line*, a string of coordinates (example: a road); and (3) *polygon*, a chain of coordinates that define an area (example: a county boundary).

Satellite images, digital aerial photography, and other forms of remotely-sensed data are the most commonly used raster data. These data are stored, not as features, but as a series of pixels or grid cells. Both types of data can (and should) be registered to a real-world coordinate system for display and analysis. Figure 21.4 displays examples of feature (vector) and image (raster) data and the ability of the GIS software to overlay these by use of a common coordinate system. Most computer and smart phone users are now familiar with the satellite imagery used in mapping applications such as Google Maps (Mountain View, CA). Satellite images and other remote sensing data have become increasingly important for monitoring and modeling human health [26].

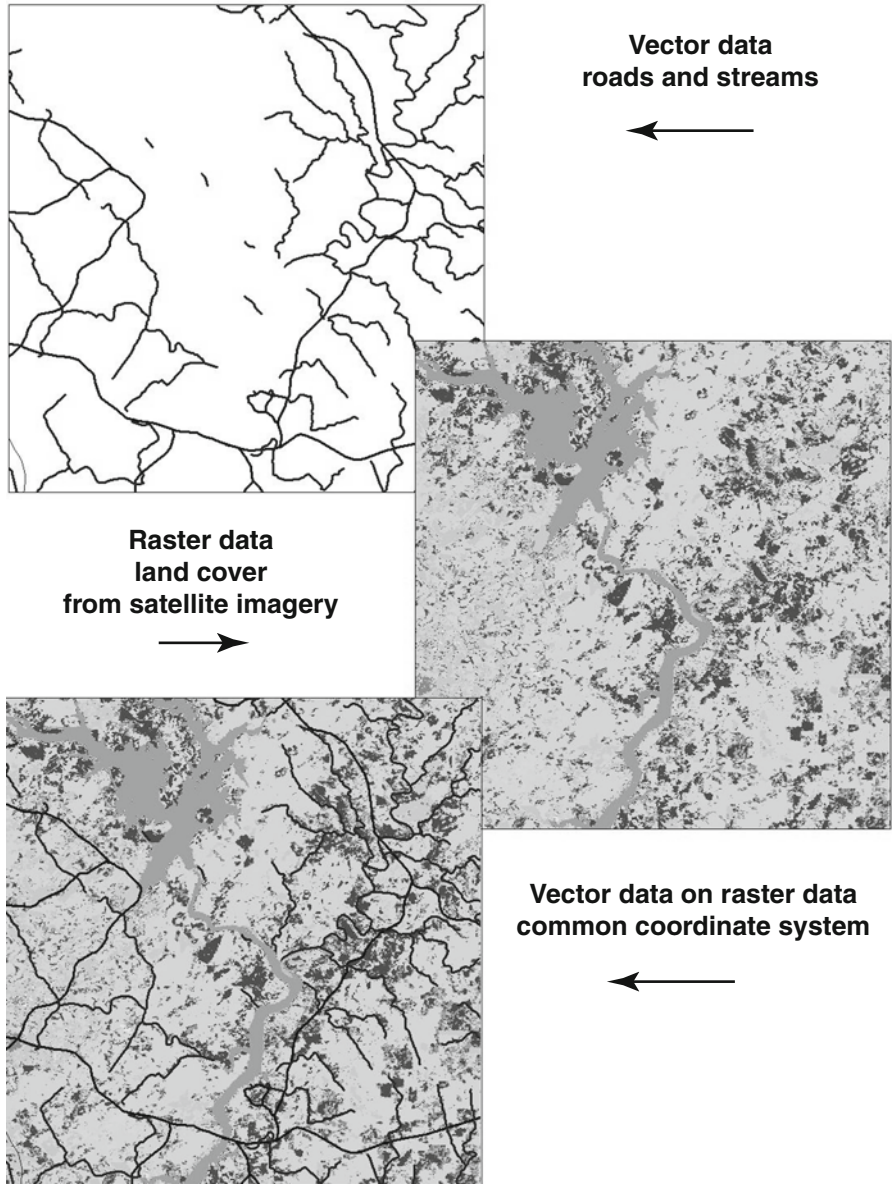


1:24,000



1:100,000

Fig. 21.3 A portion of Louisville, Kentucky, shown at two map scales (Source: US Geological Survey)



**Fig. 21.4** Vector and raster data (Map Source: ESRI, Redlands, CA; USGS National Land Cover Data)

***Functionality: Mapping and Spatial Analysis for Health Applications***

A discussion of GIS functions used for public health applications can be found in Vine et al. [27] and Cromley and McLafferty [18]. Some of the more generic functions are described in the next few paragraphs. For the beginning GIS user, the most

heavily utilized application of GIS probably will be the display of map layers and the production of thematic maps, most likely shaded (choropleth) maps. Thematic maps show the distribution of a variable, or theme, such as disease mortality, across space and are very important for understanding patterns of health outcomes.

*Choropleth mapping* assigns different shades or colors to geographic areas, according to their values; it was the technique used to produce the map in Fig. 21.1. In health applications, it may be used with counties, ZIP codes, health service areas, census tracts, or other geographic units to show the distribution of health outcomes, socio-demographic characteristics, health services, or other relevant variables. Because correct interpretation of the message or pattern displayed on a choropleth map is so critical to analysis and decision-making, a more detailed discussion of choropleth map production is provided in a later section in this chapter concerning visual display of spatial data.

*Automated address matching* can be used to map clinics, patient residences, and other locations that contain street addresses. *Address matching* is a term often used synonymously with *geocoding*, but it is actually only one of many methods of geocoding. Essentially, an address, such as 525 Fuller Street, is a geocode—it refers to a specific location along Fuller Street. Address matching works by comparing a specific street address in a database to a map layer of streets. If the map layer contains relevant information about the street name and the range of addresses along that street, the software can interpolate the location of the address and place it along the street. Most computer and smart phone users are familiar with address-matching functions: they use software such as Google Maps (Mountain View, CA), Yahoo! Maps (Sunnyvale, CA), MapQuest (Denver, CO), or other global positioning system (GPS) software to provide travel directions. Most GIS software allows the user either to enter addresses interactively (one at a time), or to process an entire database of addresses in batch mode.

*Distances* among geographic features can be determined with nearly all GIS or mapping software. In health applications, distances often are needed to analyze access to health care or to model exposure to an environmental contaminant, among other things. Most GIS software allows users to determine distances either interactively or in batch mode through the use of a distance function. In the case of batch mode determination, the distance calculation is stored in a variable that may be used for later analysis, such as regression or some sort of exposure modeling [28].

*Spatial query* allows a GIS user to query the attribute database and display the results geographically. For instance, a user could make a query to display the location of all rabies cases that have occurred in a county during the past year, or to show all census tracts in which more than 50 % of households have a household income below the poverty rate. Queries also can be based on distance: A GIS can be used to display all ZIP codes within a 25-mile radius of a particular health clinic or to show all patients within 15 miles of a field phlebotomist.

*Buffer functions* can define and display a region or “ring” of specified radius around a point, a line, or an area. GIS software allows the user to define the width of the buffer—i.e., the distance of the outside edge of the buffer from the feature boundary. A 150-m buffer might be created to determine the number of residences close to a toxic release event. A 25-m buffer zone around major roads could identify

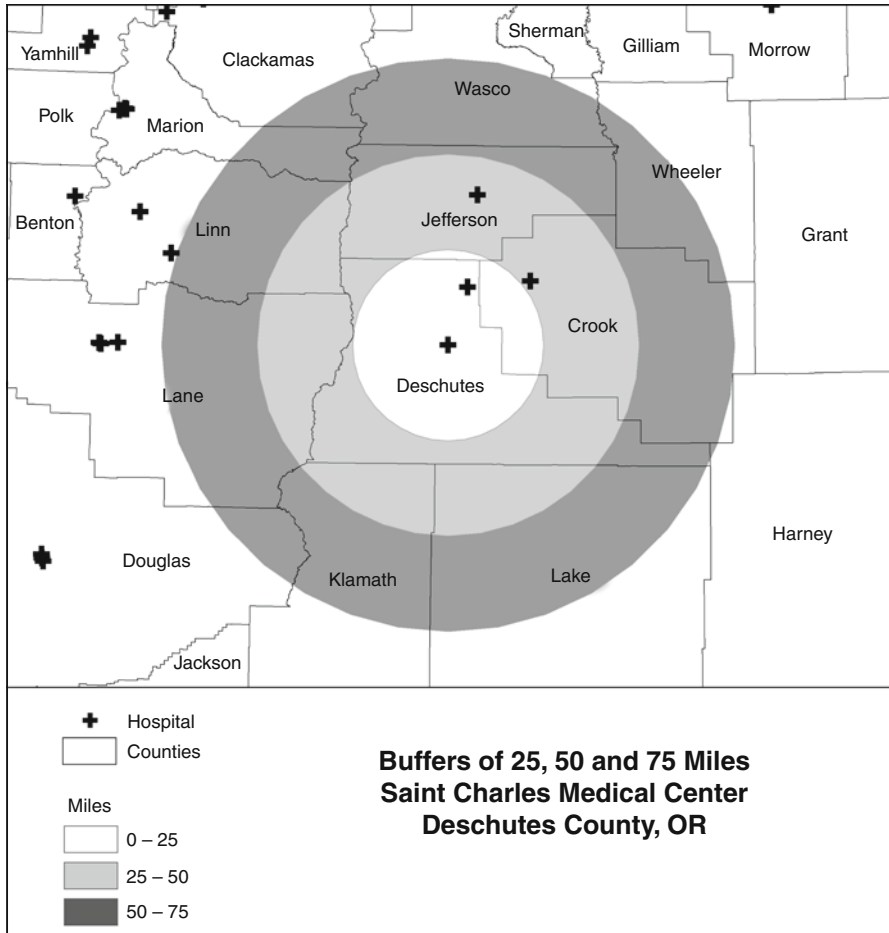


Fig. 21.5 Buffer function (Map Source: ESRI, Redlands, CA)

areas with potential lead hazards in soil from past use of leaded gasoline. Figure 21.5 shows buffers of 25, 50, and 75 miles from Saint Charles Medical Center in Bend, Oregon. Another hospital is located within 25 miles of Saint Charles Medical Center and there are three hospitals within 50 miles of the center.

*Overlay analysis* allows GIS users to integrate feature types and data from different sources. It is not to be confused with visual overlay, which occurs when several map layers are registered to a common coordinate system and displayed together, as in Fig. 21.4. Overlay analysis involves some spatial data processing and results in the creation of new data or modification of existing data. Two commonly used types of overlay analysis are *point-in-polygon overlay* and *polygon overlay*.

*Point-in-polygon overlay* is used to determine which area, or polygon, a point or set of points lies in, or whether a point lies inside or outside a particular geographic

area. For example, a point map of patient residences might be overlaid on a map layer of census tracts to determine which census tract each patient resides in. This application is important when a user is examining the association of census variables, particularly socioeconomic ones, with health outcomes. Some GIS software refers to this process as a *spatial join*.

*Polygon overlay* can be used to create a new map layer from two existing polygon map layers, when their boundaries are not coincident. For example, a ZIP code map layer can be overlaid on a layer of primary sampling units to obtain a map layer showing all complete and partial ZIP codes within a sampling area. Polygon overlay is sometimes used to estimate populations within a geographic area where boundaries differ from census boundaries, using an *areal interpolation* method. This method operates in a “cookie cutter” fashion to create new polygons; population is then prorated by comparison of the area of the new polygon to that of the original. An example of *areal interpolation* is shown in Fig. 21.6.

While these are only a few examples of GIS functions, they are all commonly used in health applications and are easy to learn. Many other functions exist, ranging from relatively simple techniques to complex methods of spatial modeling. Lai et al. [29] have written an excellent text on spatial approaches to disease analysis which discusses more advanced methods.

## Spatial Statistics

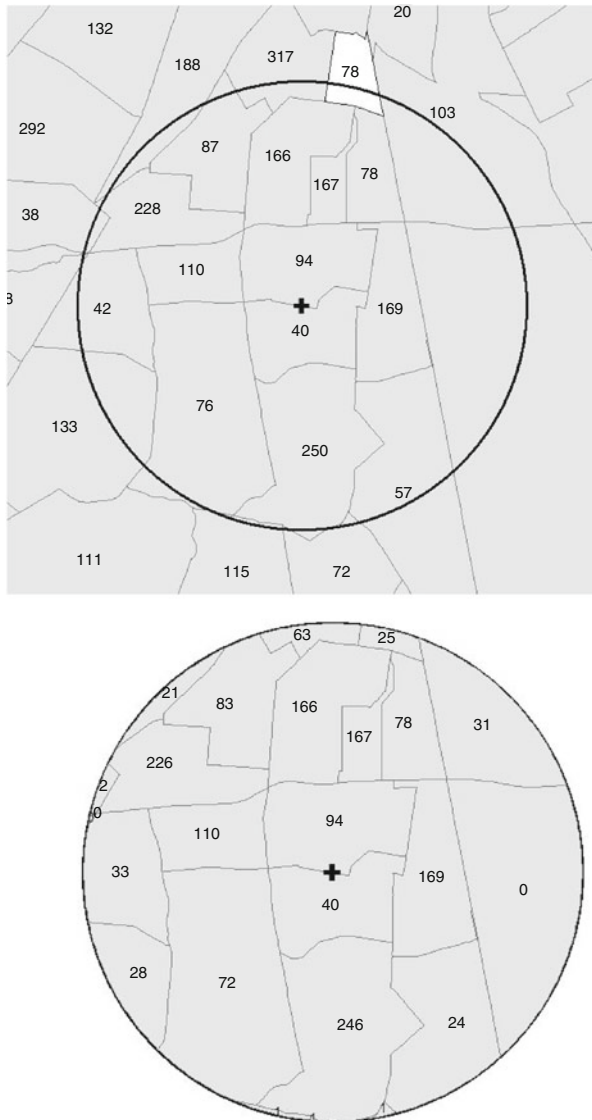
There are many time-honored spatial analysis techniques that recently have been incorporated into some of the more widely used GIS software products. Still, using statistical or more advanced spatial analysis techniques often requires additional programming, or reformatting GIS data for use with statistical software such as SAS or SPSS. Currently, two free software programs are extremely useful for the geostatistical analysis of health data: (1) SaTScan, developed for the analysis of disease clusters [30] and (2) GeoDa, which performs spatial data analysis, visualization, spatial autocorrelation and modeling procedures [31]. Figure 21.7 shows how GeoDa’s Local Index of Spatial Autocorrelation (LISA) method can be used to identify regions with statistically high cervical cancer mortality rates.

## Visual Display of Spatial Data

The proper display of spatial data requires an understanding of cartographic design, levels of measurement, and the wide range of symbols and color schemes that can be used to represent feature and image data. A thorough treatment of this subject is beyond the scope of this chapter, but it can be found in any number of cartography references and primers [32–36]. Unfortunately, the proliferation of GIS and the development of user-friendly interfaces to GIS software has made it easy for the

**Fig. 21.6** Areal interpolation. The top diagram shows a hospital in the center, with a three-mile buffer, overlaid on census tracts. Each tract is labeled by the number of women 45 years or older in poverty. The areal interpolation method estimates counts by determining the percentage of area inside the buffer and applying it to the count. The number of women in the tract just north of the hospital is 94. Since this tract is entirely contained in the buffer, all will be counted. The number of women in the highlighted tract on the buffer's northern perimeters is 78. However, only 32 % of this tract is inside the buffer, so only 25 women will be counted ( $78 \times 32\% = 25$ ). This process is accomplished by clipping the census tract map with the buffer in a cookie-cutter fashion, then comparing the tract area with the old area. Each tract in the bottom diagram is labeled by the number obtained from areal interpolation. The total population of women within the 3-mile buffer is 1680

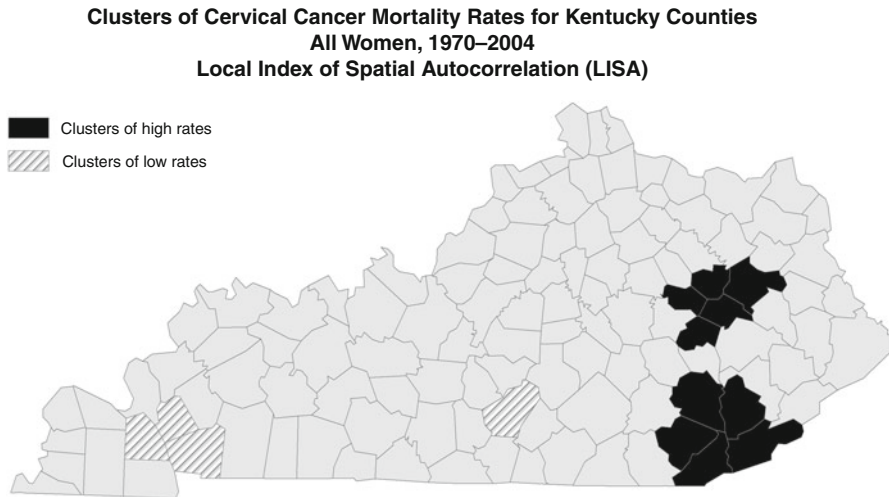
**Mammography Screening for Women in Poverty  
3-Mile Radius of Hospital**



“cartographically illiterate” to produce bad maps. Bad maps can result from the improper use of map projections, unfamiliarity with basic principles of map design, lack of understanding of data type and distribution, and poor symbol choice.

Because choropleth maps are so frequently produced and because they convey such a powerful image of the distribution and quantity of phenomena, two critical





**Fig. 21.7** LISA cervical cancer (Data Source: National Cancer Institute)

aspects of their production are discussed briefly in this chapter: (1) grouping data into classes for mapping and (2) appropriate use of symbols for choropleth mapping.

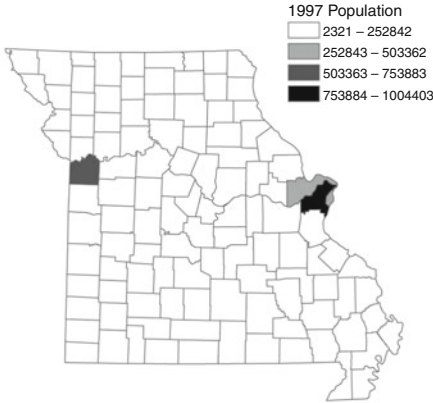
## Grouping Data into Classes for Mapping

The way in which data are grouped or classified has a strong effect on the appearance of the map and can result in maps that look very dissimilar but use the same set of data. The mapmaker must determine how many categories or classes to use and the intervals, or cut-off points, for each class. Most shaded maps use from three to six classes, which are represented in the legend. Most GIS or mapping software provides users with a number of options for classifying numeric data. Four commonly used methods are (1) equal interval, (2) quantile, (3) natural breaks, and (4) mean and standard deviation. Figure 21.8 provides examples of these methods. The viewer can discern immediately how different each of these maps looks, but they all use the exact same data!

Generally, there is no consistent “right” or “wrong” classification method to use for classifying data, but some methods are more appropriate for certain data distributions. The mean and standard deviation method is probably used least, because the general public may not understand the concept of standard deviation. A disadvantage of using the equal interval method is that, because classes are determined by dividing the range of data, and not by data distribution, it is possible to have data classes with no observations. In this case, a class (and associated shade) would be represented in the legend, but not on the map. Probably the best rule of thumb for those who are uncertain is to use either the natural breaks or the quantile method. In fact, the quantile method has been supported for epidemiological rate mapping [37].

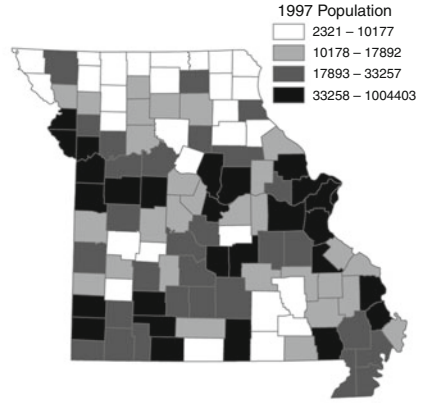
### Equal interval

The range of the data is determined by subtracting the lowest value from the highest value. The range is then divided by the desired number of classes, usually four or five, to determine the beginning and end values for each class.



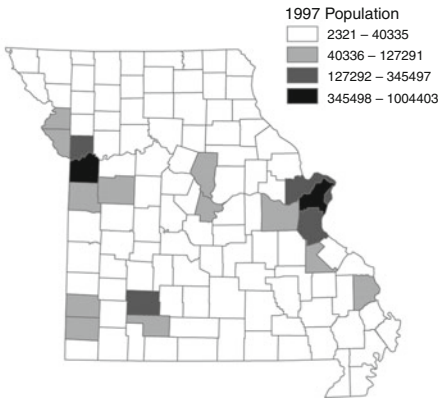
### Quantiles

The data are arranged in sequence from low to high values. The observations are then separated into the desired number of classes so that each class contains the same number of observations, or geographic units.



### Natural breaks

Natural breaks are points where there are gaps in the distribution of the data, i.e. fewer or no observations. These break points are often used as dividing points for the classes.



### Mean and standard deviation

The mean is computed and established as the center of the data distribution. Class intervals are determined by the standard deviation, a measure that determines the spread of the data around the mean.

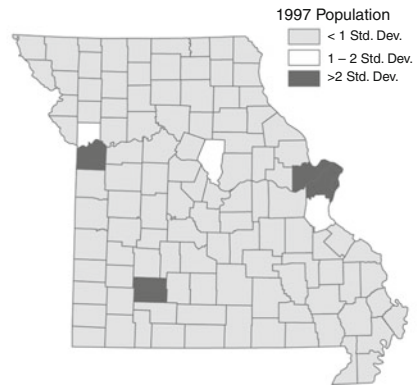


Fig. 21.8 Data grouping methods for choropleth mapping (Map Source: ESRI, Redlands, CA)

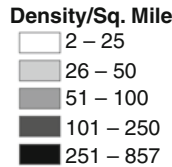
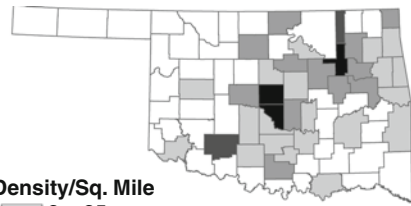
## Appropriate Use of Symbols for Choropleth Mapping

With the availability of color in computer hardware and software, it is tempting to use a wide range of colors in map production. However, a user working with numeric data should choose colors and shading patterns that communicate the map's

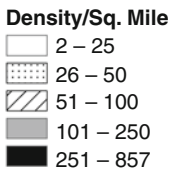
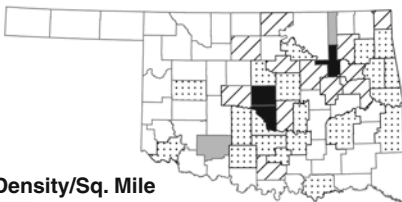
### Use of Map Symbols for Choropleth Mapping of Numeric Data

Different values of the same hue, a progression of hues, or black and white shading can be used to show patterns that are intuitive to the viewer. Dot and hatch patterns also can be used effectively. The map in the lower right corner shows a pattern resulting from a poor choice of shades.

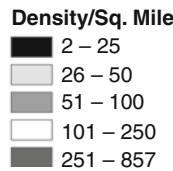
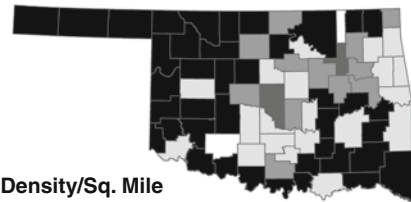
#### Black, white and grey shading



#### Dot and hatch patterns



#### Inappropriate use of symbols



**Fig. 21.9** Use of map symbols for choropleth mapping of numeric data (Map Source: ESRI, Redlands, CA)

message as clearly as possible and reflect the value of the data so that the patterns on the map are intuitive to the viewer.

In color terminology, *hue* refers to the name of the color (e.g., red, blue, green) and *value* is the lightness or darkness of a hue [38]. In general, it is best to use light colors for low data values and intense or dark colors for high data values. A gradation of values for one hue works well with numeric data, as does a range of hues from light to dark. These configurations of colors are often available in GIS or mapping software as *color ramps*, a range of hues or colors set up in the software that the user can quickly apply to numeric data. When producing a map series, color and shading patterns should be standardized for consistent interpretation across the series, such as the patterns used in the *Atlas of Cancer Mortality in the United States, 1950–94* [39]. Figure 21.9 provides examples of both appropriate and inappropriate use of symbols.

Maps are often produced for publications or reports. When color maps are too expensive to produce, the map’s message can be conveyed effectively in black and white. Gray shades can be used in place of a range of colors. However, gray shades

do not always print or copy well, and solid black can obscure boundaries, text, and other features. Dot and hatch patterns can be a more effective way to present the information.

### ***GIS Implementation: Software and Hardware***

In previous decades, GIS implementation strategies have focused on the acquisition of hardware and software, the collection of data, and aspects of managing the system, including organization and staffing. While all of the considerations addressed by these formal implementation strategies remain important today, the rapid evolution of computing technologies and increasing availability of geospatial data have resulted in a wide range of products and ‘apps’ that offer varying levels of GIS functionality. This flexibility provides the technological basis for a continuum of organizational models and implementation strategies. At one end of the continuum, an individual uses GIS on a computer, tablet, or smart phone; at the other extreme is enterprise GIS, where an entire organization uses GIS.

Much of the GIS software today falls under the general category of *desktop GIS*, which runs on a personal computer instead of being executed from a more powerful server. The GIS software and data reside on the personal computer. Over the past two decades, GIS software has become increasingly user-friendly with easy-to-use graphical user interfaces consisting of menus and tool bars. Many inexpensive, user-friendly GIS or mapping software products are now available. *Desktop GIS* does not include the broad category of web-based GIS, a technology that provides access to mapping capabilities through the use of a Web browser such as Internet Explorer, Firefox, or Safari.

GIS software falls into two primary categories: commercial (proprietary) and open source. The former costs money; the latter refers to free software whose source code can be modified by various programming languages. Some open source software is referred to as FOSS (free and open-source software). Up-to-date information about GIS platforms and their functionality can be found in Dempsey [40] and Steiniger and Hunter [41]. The suite of GIS software products developed by ESRI (Redlands, CA), such as ArcGIS, currently has the highest market share of the proprietary GIS software products. Estimates of ESRI market share range from 70 % [40] to 30 or 40 % [42]. Other products with high market share include Pitney Bowes MapInfo (Troy, NY) and GE Smallworld (Atlanta, GA). One of the better-known open source products is GRASS (Geographic Resources Analysis Support System) initially developed in 1982 by the US Army Corps of Engineers [43].

In order to evaluate hardware and software needs, GIS users in public health must determine which GIS organizational model meets their needs, the availability and format of digital geographic data, and how their GIS activities will be integrated with other research or operational units. In many cases, a powerful PC with desktop software will be sufficient. With more sophisticated systems, such as those used in

a departmental or an enterprise GIS, larger investments in data servers and software will be necessary. Potential GIS users should check with other units in their organization to determine whether any existing GIS software license agreements are in place. For example, some states have a statewide agreement that allows employees of government agencies or universities to access the licensed software. No matter which GIS system is utilized, spatial data are always disk-space-intensive. Geographic data files are large; a user should have plenty of hard drive space available.

### ***Spatial Data Access and Development***

In the 1980s and early 1990s, the primary bottleneck in GIS implementation was the need to develop and/or acquire high quality geographic data, a factor that was (and still is) often underestimated. Fortunately, during the past several years, there has been a proliferation of digital spatial data as a result of improvements in technology, the ever-increasing use of GIS, and coordination efforts by federal, state, and local government agencies, such as the Federal Geographic Data Committee (FGDC). Many of these spatial data layers are part of the FGDC's National Spatial Data Infrastructure (NSDI). They are free or can be purchased at a minimal cost from federal or state agencies. Others are sold by private vendors who have either created spatial data themselves or else added value to spatial data obtained from government and other sources.

Probably the most commonly-used spatial data in the country are the US Bureau of the Census TIGER/Line files. These files, usually referred to as simply TIGER files, were first produced for the 1990 census and contain map layers for census geography, physical landmarks, rivers and streams, transportation networks, and other features. These spatial data files can be linked with the census attribute data files for mapping and analysis of census socio-demographic variables. The street network data can be used for address matching. TIGER files can be obtained at no cost from the US Census Bureau web site [44]. Commonly-used census units are blocks, block groups and tracts. A relatively new statistical unit, the ZIP Code Tabulation Area (ZCTA), consists of an aggregation of census blocks that closely approximates a post office ZIP code area. ZCTA is beneficial for many health professionals because it allows them to link the ZIP code information in many health datasets with census socio-demographic data with greater accuracy than has been possible in the past. Of course, ZIP codes are relatively small geographic units, so users need to be cautious about HIPAA regulations and statistical small numbers issues.

As a result of the FGDC's work on the National Spatial Data Infrastructure, many states now have spatial data clearinghouses, which are often web-based 'go to' locations for free and trusted geospatial data downloads. Many of the available vector databases are in *shapefile* format, a vector spatial data format developed by ESRI, but recognized by a number of other GIS software products.

A web search on ‘GIS data’ yields many pages of results, but the links may not lead users to trusted data sources; therefore, information about the creation and lineage of spatial data is critical. The FGDC spent several years developing a standard for *metadata* that describes the content and quality of a spatial database, or, in FGDC’s words, “data about data.” Metadata provides important information about who developed the database, the scale of the original data, the time period of the content, and attribute and positional accuracy. While metadata does not guarantee the quality of the data, it does provide important information with which a user can determine appropriateness of the data’s use. Metadata is usually in XML format. Developing metadata is time consuming; therefore it might not accompany all spatial data. The metadata standard has been adopted by federal agencies as well as many state and local agencies.

GIS data for public health applications are often created by linking health attribute data from state and local government agencies to geographic boundary files by geocode. For instance, county-level mortality data can be linked to a state’s county boundary file by county code. Health datasets that contain ZIP code fields can be linked to a ZIP code boundary file. Many public health datasets are created through the address matching process, described previously.

### ***Web-Based GIS***

Internet map server technology allows nearly anyone with access to a web browser to produce maps and perform rudimentary spatial analysis. Most people are probably familiar with Google Maps or Bing Bing (Microsoft, Redmond, WA), and the ability to view map data and aerial imagery, turn layers on and off, and obtain driving directions. Google Earth, a free software download, provides more layers and functions. The FGDC manages the Geospatial Platform [45], a portfolio of geospatial data from trusted sources that includes a mapping application. The United States Geological Survey (USGS) has developed The National Map Viewer with a wide variety of map layers available for viewing and download [46]. With web-based GIS, geographic information is provided via a client–server model where an application server accesses data from a data server or data warehouse and provides the data to a client using a map server application.

In the past few years, the number of health-related map servers has proliferated. A few examples include the National Cancer Institute’s *Cancer Mortality Maps*, where the user can define anatomical site, time period, spatial unit, number of class intervals, and color scheme [47]. The Centers for Disease Control and Prevention (CDC) hosts a number of interactive atlases, such as the one for heart disease and stroke, which provides county-level mapping [48]. Not all map servers work well with all browsers. One of the emerging technologies in GIS is cloud computing, where powerful servers store data and provide applications over the internet. In this environment, spatial data, GIS software, and applications are part of the cloud infrastructure and accessible via a number of hardware options, including mobile devices.

## ***GIS Training***

All organizational models of GIS require personnel with high levels of technical competence to develop the databases and applications that provide effective, high quality analysis and results for decision support. Somers [49] made a distinction between (1) full-time GIS users, (2) part-time GIS users, and (3) support staff. Full-time GIS users are often technicians, analysts, or managers, who have educational backgrounds in geography or GIS; part-time users might have backgrounds in a field of expertise, such as environmental health, with training in the use of GIS.

For the most part, learning how to use GIS or desktop mapping software is not difficult or time-consuming, a fact that can be deceptive because it obscures the complexity of GIS. GIS software vendors often offer their own training courses and many universities now offer online postbaccalaureate certificates in GIS, such as the one offered through Pennsylvania State University's World Campus [50].

GIS users in the public health fields have additional concepts that they must master, many of which can be gleaned from a course in epidemiology or biostatistics. These concepts include the use of rates, statistical variation involving the use of small numbers in either the numerator or denominator, the concept of rate adjustment, and the impact of different standard populations on rates. In addition, state and local public health GIS users need to have a sound understanding of the ecological fallacy in the analysis of cause-and-effect relationships, i.e. that one cannot make assumptions about individuals based on group-level data, and of issues involved in modeling exposure to environmental factors.

## ***Social and Institutional Issues***

Individual and organizational users of GIS typically need to address a number of social and institutional issues. These issues include confidentiality, security and data access, coordination with other agencies, and organizational politics.

## ***Protected Health Information and HIPAA***

Many health datasets contain sensitive information. Patient addresses and other geocodes can serve as individual identifiers. Consequently, public law mandates that agencies and researchers maintain the confidentiality of patient records and health statistics. The Health Insurance Portability and Accountability Act (HIPAA) sets out detailed regulations on the dissemination of personal health information (PHI), including geography [51]. HIPAA regulations mandate that all geographic subdivisions smaller than a state must be removed before the data is considered de-identified enough for publication. The exception to this is the 3-digit ZIP code, an area generic enough to protect privacy and produce meaningless results. GIS

users must be very cautious about which maps are produced for internal use vs. those that are distributed to the public or shown in presentations. Some researchers have suggested that HIPAA restrictions have had a negative impact on public health research in a GIS context [52]. One of the best approaches is to discuss the project with an Institutional Review Board (IRB) member; it may be possible to obtain a HIPAA waiver.

### ***Security and Data Access***

Many of the security and data access concerns are closely related to data privacy and confidentiality issues discussed earlier in Chap. 9. All of the major computer operating systems have security features that can restrict access to files and data through the use of logins, passwords, and encryption software. In addition, firewalls are often set up to limit access from outside an organization. It is critical to have competent system administration and information technology staff to handle data security issues. GIS users need to think carefully about the data on their personal computers and USB devices to prevent security breaches.

### ***Coordination with Other Agencies***

In addition to federal coordination agencies, such as the FGDC, many states and regions are involved in data sharing and coordination activities. For instance, the Louisville Metro (KY) health department has access to a wealth of spatial data developed by the Louisville/Jefferson County Information Consortium (LOJIC) [53]. Coordination activities provide GIS users with opportunities for: sharing data and applications; keeping abreast of developments in the technology; training; and access to important information for decision-making, such as software purchases.

### ***Organizational Politics***

The impact of organizational politics on GIS operations should not be overlooked. For example, upper level managers might veto GIS applications that address politically sensitive or controversial issues. In addition, reorganization in government agencies, common and usually political, can have either positive or adverse impacts on GIS operations. Moreover, GIS is a technology that nearly everyone wants. Consequently, the location of a GIS unit in the organizational structure in an agency can affect which projects receive priority and/or funding.



## **GIS Limitations**

While GIS is a powerful tool that is increasingly easy to use, GIS users must recognize the limitations of the software and the spatial data and make attempts to work around those limitations. Some common limitations are discussed below.

### ***Accuracy and Completeness of Spatial Data***

Mapping and spatial analysis can be severely impacted by the quality of the geographic data. In addition, errors can be propagated during data processing or modeling activities. Coordinate precision, i.e., the number of significant digits that are stored for each coordinate, plays a role in some of these errors, as does the use of different map projections. Three good rules to follow are to (1) never assume that a geographic database is free of error, (2) acquire the metadata and read it to obtain information about the creation of the data, and (3) whenever possible, develop methods of assessing data quality.

### ***Accuracy and Completeness of Attribute Data***

Inaccuracies also exist in non-spatial databases. Character fields may have misspellings, and numeric fields may have data entry errors. As with spatial databases, quality control procedures should be developed to the extent possible, as illustrated by the following example. In 1998, the author conducted extensive mapping and geographic analysis using one of the public health screening databases maintained by the State of North Carolina [54]. During this study, it became apparent that many of the county geocodes in the database were incorrect. The author compared data from 1994–96, consisting of 265,492 records, to a master lookup table containing City, County and ZIP Code fields to check for correspondence in the screening database, and discovered that only 158,552 records (59.72 %) contained accurate and/or complete information. Some incorrect geocodes resulted from laboratory manual data entry errors (i.e., typos, which are easy to make since most geocodes are numeric), while others resulted from confusion over city and county names: many North Carolina towns and counties have the same names but very different locations. These types of errors are common, and in this case went unnoticed until these data were used in a GIS.

### ***Currency and Time Period of Data Content***

One data characteristic that is often neglected is that of time. When were the data collected? When were they last updated? It is easier to obtain funds to create GIS databases than to maintain and update them. Currency has been a serious issue

with census data, which are commonly used in health analyses. Prior to the implementation of the American Community Survey, census data were only collected every 10 years. Thus, a study of 1998 mortality had to utilize 1990 census data, or intercensal estimates. Now, census data are collected continually via the American Community Survey, but such timely data are not always available for smaller geographic areas such as block groups.

### ***Address Matching Issues***

Address matching is commonly used with health datasets to create a map layer of points showing facility locations or patient residences. The proliferation of street network data by private companies over the past few years has resulted in much greater accuracy in both urban and rural areas. However, not all addresses will match a street database – for example, there may be typographic errors, or multiple units in a large apartment complex– and the user will need to make decisions about how to process the ‘rejects.’ Many health surveys obtain information about mailing address, which sometimes differs from address of residence. For epidemiologic studies, it is important to remember that address of residence does not always infer location of exposure. Also, an address provides no indication of residential mobility: information about previous addresses or length of residence at current address is rarely contained in health datasets.

### ***Use of ZIP Codes***

Many health datasets do not contain an address field, and attempts to conduct sub-county analyses may therefore be limited to the use of ZIP codes. However, ZIP codes were developed by the US Postal Service for the delivery of mail, not for geographic analysis and mapping. Unlike census units (e.g., tracts, block groups) ZIP codes were not intended to be homogeneous with respect to socio-demographic variables. Although census data are now tabulated by ZCTA, the heterogeneity of populations within a specific ZCTA can still lead to averaging of values. ZIP codes can also cross county lines. One additional problem with ZIP code boundaries is that they change over time. Therefore, health data from 2006, for example, should not be mapped using a 2010 ZIP code file. Sometimes there is no choice but to use available data. In such a case, a user should always document the source of the data and its time period.

### ***Scale and Precision of Location***

Metadata should include information about the processes used to create the database. For example, the scale of the source map has a great impact on the

coordinate precision of a feature's location. The location of features digitized from a large-scale map will be more precise than those obtained from a small-scale map. The precision of point data is dependent on the method used to locate the points. Points that have been address-matched to a street network will generally be more precise than points matched to a ZIP code centroid, but less precise than those matched to the centroid of a tax parcel (i.e., property).

### ***Proximity vs. Exposure***

In epidemiologic studies, it is important to remember that proximity to a feature, such as a hazardous waste site, does not always imply exposure. Beware of associations gleaned from map overlay or geographic analysis. GIS is a wonderful tool for understanding relationships among features and for generating hypotheses about etiology, but GIS must be supplemented with standard epidemiological methods when analyzing spatial correlates of health outcomes.

### **Summary**

In summary, GIS is an information system, an approach to science, and a powerful set of analysis and visualization tools that can be used by public health professionals to enhance their analysis and understanding of public health issues and to provide a basis for sound decision-making. GIS is deceptively easy to use; however, geographic data, spatial or epidemiologic analysis, and GIS information systems are more complex than they appear to the casual user. The effective use of GIS requires a combination of good training and experience. In the years ahead, that training and experience will grow in importance as GIS becomes an increasingly powerful and common tool in the practice of public health.

#### **Review Questions**

1. Explain three ways in which GIS can be useful to public health practitioners.
2. Describe the difference between *spatial* and *attribute* data in a GIS.
3. Define *raster* and *vector* data.
4. Why is it important to understand cartographic principles such as map projections and data classification?
5. What steps must be taken to protect sensitive information in health datasets?
6. GIS is a powerful tool, but what are some of its limitations?
7. What is *metadata* and why is it important?
8. Explain the principles underlying (1) the use of colors in maps that display data and (2) the principles for appropriate use of black and white.

## References

1. O'Carroll PW. Introduction to public health informatics. In: O'Carroll PW, Yasnoff WA, Ward ME, Ripp LH, Martin EL, editors. *Public health informatics and information systems*. New York: Springer; 2002. p. 3–15. p.5.
2. Savel TG, Foldy S. The role of public health informatics in enhancing public health surveillance. *Morb Mortal Wkly Rep*. 2012;61(03):20–4.
3. Lasker RD, Humphreys BL, Braithwaite WR. Making a powerful connection: the health of the public and the National Information Infrastructure. Report of the U.S. Public Health Service Public Health Data Policy Coordinating Committee; Washington, DC. 1995.
4. Meade MS, Emch M. *Medical geography*. 3rd ed. New York: The Guilford Press; 2010.
5. Cliff AD, Haggett P. *Atlas of disease distributions: analytic approaches to epidemiological data*. Oxford: Blackwell Publishers; 1988. p. 5.
6. Snow J. On the mode of communication of cholera. 2nd ed. London: Churchill Livingstone; 1854.
7. Choynowski M. Maps based on probabilities. *J Am Stat Assoc*. 1959;54:385–8.
8. Cliff AD, Haggett P, Ord JK. Forecasting epidemic pathways for measles in Iceland: the use of simultaneous equation and logic models. *Ecol Dis*. 1983;2:377–96.
9. Glick BJ. The spatial organization of cancer mortality. *Ann Assoc Am Geogr*. 1982;72:471–81.
10. Meade MS. Cardiovascular disease in savannah, Georgia. In: McGlashan ND, Blunden JR, editors. *Geographical aspects of health: essays in honor of Andrew Learmonth*. London: Academic; 1983. p. 175–96.
11. Gould PR, Leinbach TR. An approach to the geographical assignment of hospital services. *Tijdschrift voor Econ En Soc Geogr*. 1966;57:203–6.
12. Hanchette CL. GIS and decision making for public health agencies: childhood lead poisoning and welfare reform. *J Public Health Manag Pract*. 1999;5:41–7.
13. Richards TB, Croner CM, Rushton G, Brown CK, Fowler L. Geographic information systems and public health: mapping the future. *Public Health Rep*. 1999;114:359–73.
14. Winn DM. The long island breast cancer study project. *Nat Rev Cancer*. 2005;5:986–94.
15. Richards TB, Croner CM. Issue focus: geographic information systems in public health, part 1. *J Public Health Manag Pract*. 1999;5(2):1–106.
16. Richards TB, Croner CM. Issue focus: geographic information systems in public health, part 2. *J Public Health Manag Pract*. 1999;5(4):1–82.
17. Cromley EK, McLafferty SL. *GIS and public health*. New York: The Guilford Press; 2002.
18. Cromley EK, McLafferty SL. *GIS and public health*. 2nd ed. New York: The Guilford Press; 2012.
19. Turnock BJ. *Public health—what it is and how it works*. 3rd ed. Sudbury: Jones and Bartlett; 2004. p. 184.
20. Clarke KC. *Getting started with geographic information systems*. Upper Saddle River: Prentice-Hall; 1999.
21. Jones C. *Geographical information systems and computer cartography*. Essex: Longman; 1997.
22. Goodchild MF. What is geographic information science? NCGIA Core Curriculum in GIScience. 2007. Available at: <http://www.ncgia.ucsb.edu/giscc/units/u002/u002.html>. Cited 17 Mar 2013.
23. Haggett P. *Geography: a global synthesis*. London: Pearson-Prentice Hall; 2001.
24. Glick B. The spatial autocorrelation of cancer mortality. *Soc Sci Med*. 1979;13(2):123–30.
25. Harvey F. *A primer of GIS: fundamental geographic and cartographic concepts*. New York: The Guilford Press; 2008.
26. Beck LR, Lobitz BM, Wood BL. Remote sensing and human health: new sensors and new opportunities. *Emerg Infect Dis*. 2000;6(3):217–27.
27. Vine MF, Degnan D, Hanchette C. Geographic information systems: their use in environmental epidemiologic research. *Environ Health Perspect*. 1997;105:598–605.

28. Vine MF, Stein L, Weigle K, Schroader J, Degnan D, Tse C, Hanchette C, Backer L. Effects of the immune system associated with living near a pesticide dump site. *Environ Health Perspect.* 2000;108(12):1113–24.
29. Lai PC, So FM, Chan KW. *Spatial epidemiological approaches in disease mapping and analysis.* Boca Raton: CRC Press; 2008.
30. Kulldorff M. A spatial scan statistic. *Communications in Statistics: Theory and Methods.* 1997;26:1481–96.
31. Anselin L, Syabri I, Kho Y. *GeoDa: an introduction to spatial data analysis.* *Geogr Anal.* 2006;38(1):5–22.
32. MacEachren A. *Some truth with maps: a primer on symbolization and design.* Washington, DC: Association of American Geographers; 1994.
33. Monmonier M. *How to lie with maps.* Chicago: The University of Chicago Press; 1991.
34. Monmonier M. *Mapping it out: expository cartography for the humanities and social sciences.* Chicago: The University of Chicago Press; 1993.
35. Krygier J, Wood D. *Making maps: a visual guide to map design.* New York: The Guilford Press; 2005.
36. Brewer CA. *Designing better maps: a guide for GIS users.* Redlands: ESRI Press; 2005.
37. Brewer CA, Pickle L. Evaluation of methods for classifying epidemiologic data on choropleth maps in series. *Ann Assoc Am Geogr.* 2002;92(4):662–81.
38. Robinson AH, Morrison JL, Muehrcke PC, Kimerling AJ, Guptill SC. *Elements of cartography.* 6th ed. New York: John Wiley and Sons; 1995.
39. National Institutes of Health (NIH), National Cancer Institute (NCI). *Atlas of cancer mortality in the United States, 1950–94.* NIH Publication No. 99–4564. Bethesda: NIH; 1999.
40. Dempsey C. *Implementing GIS.* 2008. Available at: <http://www.gislounge.com/implementation-of-a-gis/>. Cited 17 Mar 2013.
41. Steiniger S, Hunter AJS. The 2012 free and open source GIS software map – a guide to facilitate research, development and adoption. *Comput Environ Urban Syst.* 2013;39:136–50.
42. Schutzberg A. Update: Esri has 40+ % of GIS Marketshare. *Directions Magazine.* 21 Nov 2011. Available at: <http://apb.directionsmag.com/entry/esri-has-40-of-gis-marketshare/215188>. Cited 17 Mar 2013.
43. Wikipedia. 2011. Available at: [http://en.wikipedia.org/wiki/List\\_of\\_geographic\\_information\\_systems\\_software](http://en.wikipedia.org/wiki/List_of_geographic_information_systems_software). Cited 17 Mar 2013.
44. US Census Bureau. TIGER Products. Available at: <http://www.census.gov/geo/maps-data/data/tiger.html>. Cited 6 Sep 2013.
45. Federal Geographic Data Committee. Geoplatform. Available at: <http://www.geoplatform.gov>. Cited 17 Mar 2013.
46. US Geological Survey. National Map. Available at: <http://www.nationalmap.gov>. Cited 31 Mar 2013.
47. National Cancer Institute. Cancer mortality maps. Available at: <http://ratecalc.cancer.gov/>. Cited 17 Mar 2013.
48. Centers for Disease Control and Prevention. Interactive atlas of heart disease and stroke. Available at: <http://apps.nccd.cdc.gov/DHDSPAtlas/>. Cited 17 Mar 2013.
49. Somers R. Organizing and staffing a successful GIS: organization strategies. *URISA Journal.* 1995;7:49–52.
50. Pennsylvania State University. Postbaccalaureate certificate in geographic information systems. Available at: <http://www.worldcampus.psu.edu/degrees-and-certificates/geographic-information-systems-gis-certificate/overview>. Cited 17 Mar 2013.
51. US Department of Health and Human Services. Health information privacy. Available at: <http://www.hhs.gov/ocr/privacy/>. Cited 17 Mar 2013.
52. Tellman N, Litt ER, Knapp C, Cheng J, Radonovich LJ. The effects of the health insurance portability and accountability act privacy rule on influenza research using geographical information systems. *Geospat Health.* 2010;5(1):3–9.
53. Louisville/Jefferson County Information Consortium (LOJIC). 2013. Available at: <http://www.lojic.org/>. Cited 17 Mar 2013.