1 Introduction to Geographic Information Systems

Chapter Description

This chapter introduces, defines, and explains three primary components and associated concepts common to any geographic information system (GIS), including appropriate examples relevant to social science research methods. The chapter serves as a foundation for the following sections of the text. We illustrate the versatility and wide-ranging applicability of GIS.

Chapter Objectives

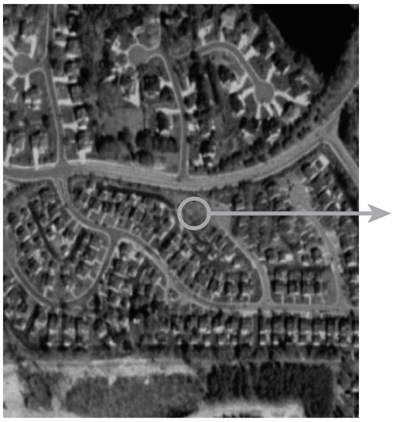
- Present a working definition of a geographic information system.
- Demonstrate the use of GIS as a social science research method.
- Present examples showing how GIS can be used in social science research situations.

After reading this chapter, you should be able to perform the following tasks:

- Identify spatial perspectives in your own area of social science research.
- Present your own definition of a GIS.
- Understand how a GIS is different from other analysis environments you are familiar with.

What Is a Geographic Information System?

In its simplest form, a geographic information system (GIS) is a system designed to store, manipulate, analyze, and output map-based, or spatial, information. In practice, the functions of a GIS can be carried out by hand, using only paper,



Data Sheet

Survey Number: 294
Street Address: 325 Burnaby Court
City, State: Fraser, TX
Age of respondent: 37
Gender of respondent: Male
Highest level of education: B.A.
Number of adults living in household: 2
Profession of respondent: Stock Broker
Profession of other adults in household: Teacher
Annual Household Income: $86,000
Number of children under 18 in household: 3

Figure 1.1 Linking real-world locations to data via a GIS-compatible spatial reference. On the left is an example of a United States Geological Survey aerial photograph of a suburban location. This photo shows the world in much the same manner as you might see from the window of an airplane. When collecting data in this area—for example, via a mailed survey—you would have the option of retaining spatial information by recording something about the specific place the data were collected (e.g., street address, census block, neighborhood). On the right are tabular data that are associated with one such surveyed household, as recorded on a survey form. Used together, this spatial and tabular information would be useful in doing a GIS-based analysis.

pencil, and a ruler (a surprising number of people still do it this way!). Of course, this is not practical or efficient for many research applications.

When we refer to spatial information or data, we simply mean that the information is linked to a specific location, for example, a street address. Figure 1.1 provides an example of a real-life view of the world, as represented in an aerial photograph tied to the associated data about the world similar to what you might choose to collect or analyze. These tabular data are related to the world via its location, or spatial information.

Although there is no single universally agree-on definition of a GIS, GIS professionals do agree on several general principles. First, a GIS requires a combination of computer hardware and software tools. Second, a GIS requires data, and these data must posses a spatial or location component. Third, a GIS requires knowledgeable individuals to develop the database and carry out the data processing. Most of these tasks can be accomplished by anyone with a little basic computer knowledge, which we discuss in this book. Although GIS software has become much easier to use since the introduction of graphical user interfaces, GIS programs and much of the underlying geographic theory require people to have a basic understanding of maps and map analysis. Last, and perhaps most important, a GIS is a system for analysis. That is, a GIS is useful for the examination, display, and output of information gleaned from the data that are stored and maintained in the system. The focus of this book is to provide you with the necessary understanding of mapping concepts and spatial analysis as well as the analytical approach needed to perform GIS-based research.

Understanding GIS

GIS are best understood by breaking down the terminology and developing an appreciation for the application of GIS to various analysis situations. In particular, how can a researcher's area of interest and the associated data be placed into a GIS context and how can GIS technology enhance the analysis and understanding of data? In the following sections, we review GIS in detail to establish a basis for successful use of GIS in applied social science research.

The "G" in GIS

The geographic component of GIS is simultaneously obvious, confusing, and difficult to get a handle on. From an early age, we all develop an understanding that the location of people and places can be marked on a map and, furthermore, that connections can be made between these locations. What we may not have is a good understanding of the scientific basis for mapping—that is, the numerous issues of scale, coordinates, control datums, and so forth. Even to those readers familiar with some of these concepts, very few, other than mapping professionals, have a deep conceptual understanding of the mathematical algorithms behind these concepts and the potential errors that result from various combinations and interactions of such data.

Fortunately, most of these underlying issues are addressed for us through the GIS software, so it is not essential to have a deep understanding of these concepts. What is important for you to understand is that there are a few concepts that will be important to pay attention to, even if you don't understand the intimate details of how they work. You can think of this as analogous to knowing the difference between VCRs and DVD players. You know these are different tools with different strengths and weaknesses, but selecting the right one does not require you to understand the intricate details of their function. What is important is that you know which format to ask for when you go to rent a movie at the local video shop so that the media selected fits the player you own.

In social science research, the value of the geographic context may not always provide an obvious research benefit. Many social science studies focus on social, economic, cultural, and survey data that have limited if any spatial question associated with them. For example, do pregnant women who are better educated or wealthier receive higher quality prenatal care? Perhaps the more telling question would be, where are the prenatal clinics located relative to available public transportation, child care, and so on? One might use census data to conduct a statistical analysis of census block groups and levels of prenatal care, but this analysis may miss an important location component. Often, when explored in conjunction with other map-based location information (e.g., where the blocks are located relative to other important components), a more complete understanding of the causative relationships can be obtained. Furthermore, from an applied standpoint, the geographic component can help in determining where to best locate and spend limited resources to help improve the situation.

In reality, almost all information researchers collect about people, their communities, and their environments can be tied to a geographic location. For example, we may survey people at their home address or by a geographic unit such as a census block or city of residence. All of these locations can be easily mapped to a location. Furthermore, if privacy is a concern, data may be aggregated to a larger geographic unit to mask specific, personal information. In short, if you can answer the question, "Where was the data collected?" then GIS is an appropriate means for storing and analyzing the data.

Difficulties With the "G"

The geographic context may be difficult to collect because determining the exact location of a piece of data on the ground is not always easy to accomplish or, for reasons of privacy, may not be permissible. When mapping people, we face an additional challenge: People move around, may be without a home, or otherwise may be difficult to tie to a particular location. However, because geographic data are the heart of the GIS, knowing a location of some kind is an essential part of the GIS process (even if it must be spatially degraded or detached from the exact, true location).

For example, if you were doing a study of homeless individuals, you might be better able to define their location at the level of a particular neighborhood they call home than you could a particular street address. Furthermore, even in studies where mappable locations are available, there may be privacy issues that necessitate degrading that information. In other words, even if you have specific addresses of your respondents, you might choose to degrade the data to census blocks, neighborhood, or even the city level to maintain privacy dictated in ethical research. Choosing the level of spatial detail (which is actually a scale question, as discussed in further detail in Chapter 2) is an important part of the GIS process.

Perhaps more difficult to map, though equally important in social research, are conceptually mapped features. Such features might include data about perceptions, ideas, or interactions. For example, social networks or interactions between individuals may be mapped in such a way that people who are emotionally close are located conceptually close together, whereas individuals who are casual acquaintances are mapped at a greater distance. Lines connecting people on the map could represent a social distance rather than true geographic distance. On such maps, referred to as cartograms, the distance between mapped data is scaled to a variable or index value other than distance. In the case of social ties, this might be an index representing the strength of a particular relationship.

A second difficulty in GIS mapping relates to the variability that occurs in time and space. Most data at present are collected as a snapshot in time. We have a more difficult task obtaining data over a series of short time steps to reliably map changes or trends in the data. Furthermore, because many of the things that we may map—especially individual people—move over time, there is an extra dimension of analysis to consider. Do we locate survey respondents based on their home address or place of employment or on where the respondents are most likely to be at a particular time of the day or week? This decision would be most significantly influenced by the question under study; there are no set answers.

Using computer animation, map data can be transitioned from static to dynamic. However, this type of mapping is still limited by the difficulty and expense of collecting data at a high frequency (temporal scale) as well as by software limitations in incorporating data instantly as it is collected. Fortunately, there are few social science applications that necessitate true real-time analysis. The primary goal for researchers considering GIS as a tool in their analysis is to make such decisions before collecting the data.

Additional considerations relate to the spatial representation of data. Often, in social science research, privacy is of the utmost concern. Data are typically lumped to mask individual data points representing individual respondents. A serious trade-off resulting from the lumping, or degrading, of data in this fashion is that the true, raw data may be permanently lost and no longer available for future research. One result of this process is that an enormous amount of redundant data collection occurs in situations where the simple recategorization of existing data in different but equally valuable combinations could provide for the exploration of vastly different questions.

A simple example of data degradation is the grouping of income levels into categories, a common practice in survey research. Categorical information such as < \$15,000 and \$15,001–\$35,000 provides no means for a later study to distinguish individuals with incomes between \$20,001 and \$30,000. In a mapping context, it could be useful to link people or ideas to specific locations, but, more commonly, data are collected by larger geographic regions, such as census blocks or other political boundaries. The same problem as demonstrated with income occurs with census blocks. A census block doesn't show the internal distribution of data in the census block. Are the households equally distributed across the area, or is clustering of the households hidden in the simplified data? Data that are degraded can no longer be recategorized to explore new or different questions. Of course, these are not simple issues to address because anonymity is an essential component of many social science questions; however, to the extent possible, when data are maintained in near-original, detailed form, the possibilities for analyses both within and outside the GIS become much greater.

Expanding the "G"

Mapping attitudes, ideas, social networks, and countless other human constructs should be viewed as equally valid as mapping the latitude and longitude of a data point on the ground. Numerous opportunities, limited only by the creativity of the researcher, allow GIS to extend into realms not envisioned by the traditional geographies originally programmed into the software. The question that remains to be addressed is how one can develop an appropriate mapping context to represent concepts such as social interaction, desirability of a community, or social ties. Developing an index value or relationship between data points that can be used in place of physical distance as traditionally mapped is one means for visualizing data in a mapping context.

The "I" in GIS

The information component of GIS relates to the database aspect of the software. Databases are specialized software programs designed for the storage, organization, and retrieval of information. GIS software packages can read or directly interact with data from almost any data management and analysis software. There may be some data translations necessary to facilitate the movement of an existing research database into a GIS software package. It is useful to note that many of the fundamental baseline data sets one might need in answering a question are already available in GIS-ready formats.

In particular, data from the United States Census, as well as data from many state- and local-level data sets, are available through online sources or via the appropriate government office. Numerous university sources also provide GIS-ready data, as do a variety of private firms. Much of the data are freely accessible, whereas a good number of commercial databases are available for purchase. Deciding to use free versus commercial data most often comes down to your needs and experience. Commercial data are often reviewed for quality control purposes and are presented in ready-to-use formats directly compatible with commercial GIS software packages. Many free sources are similarly prepared, although some free sources may require more effort and manipulation to make them compatible with a particular GIS package. Your choice of data source may depend only on your budget. If you do not have the resources to purchase data or to collect your own new data, you may need to explore free options. Of course, if you use free data, there could be a trade-off in the time it takes to prepare the data for use in your analysis, or the necessary data simply may not be available.

Of course, you can also use any of your own data collected via surveys, interviews, observations, or almost any other means, which becomes vastly useful input to the GIS. In fact, as long as a researcher intends the data to end up in a computer in digital form, regardless of the particular software involved, it will be accessible to a GIS. Chapter 6 discusses data preparation in detail, including the import and export of data between software packages. In fact, with a little foresight, perhaps through one or two additional questions or notations on the data sheet, data can be collected to facilitate its easy incorporation into GIS. These additional data are the collection of some form of either real or conceptual location information tied to a base map chosen by the researcher. Thus, the information aspect of GIS is the easy part because almost everyone working in social sciences is already familiar with the process of collecting and coding data. Most are also familiar with entering data into a computer for analysis of one kind or another.

Extending the "I"

Most of the issues related to making information more accessible to GIS relate to the upfront organization and structure of data storage, coding, and format. These issues are not unique to GIS but are essential considerations for all data collection and analysis. Additional benefits of the computer and GIS are the additional storage opportunities available. Multimedia capabilities of the computer allow the linkage of photographs, sound or movie files, and scanned information. This provides a significant opportunity for raw data preservation, thus maintaining complete, detailed information.

For example, a key-informant interview or oral history could be recorded on tape in its entirety or a traditional dance recorded on video. These records can be converted into sound and video computer data files and linked to the GIS map as discussed in Chapter 6. When the location associated with that video or sound file is clicked on the GIS map, the complete recording becomes available to the researcher. Data that are coded or summarized from open-ended surveys or interviews for purposes of analysis are simultaneously available in their entirety to a researcher who may opt for a different coding scheme or analysis at a later time.

Thus, beyond the many analytical and data presentation benefits we explore in this text, the GIS can serve as a place to store a wide variety of digital data sets collected in the course of a particular study. And so far as these data can be maintained in raw form, this information can be used in new and different analyses at a later time. As one builds a library of GIS data sets, opportunities for extending the life and use of any individual data set are greatly enhanced.

The "S" in GIS

The system necessary to carry out the development and analysis in GIS includes a variety of hardware and software components, in addition to people who can make them work. Most academic institutions and government agencies and many private consultants may have these capabilities. Building GIS from the ground up varies greatly in cost and required training. The important considerations for a researcher are the trade-offs between doing the GIS work on your own and turning some aspects over to a more highly trained GIS analyst. Regardless of the decision made, it is most important to ensure that there is an upfront understanding of the data structure and format necessary to achieve compatibility with a minimum of difficulty, this being a component that you, the researcher, needs to take charge of. We discuss the conceptualization of a GIS-based analysis in detail in Chapter 8.

Difficulties With the "S"

GIS were largely developed with traditional geography in mind, and these original systems have areas of improvement. Since the initial development of GIS, the use and thus the capabilities of GIS have expanded across a wide variety of disciplines and applications in both the public and private sectors. These systems are further limited by the map data model, consisting of points, lines, and polygons (discussed in Chapter 2). This model assumes that all data can be linked to a very specific, discreet location and that lines can be drawn to explicitly delineate the boundary between data categories. Of course, many data sets are not so clearly defined, especially when dealing with social science data, which tend to be less geographically specific, either because they must be degraded to protect privacy or because you are analyzing conceptual maps as opposed to geographic maps.

Although there have been efforts to develop fuzzy GIS systems that allow locations and boundaries to be less definite, these systems are not yet available in the mainstream. Therefore, the existing GIS data models do not necessarily fit the analysis being conducted, and a researcher is sometimes forced to come up with a creative solution to make nondiscreet data fit into a discreet data model (data models are discussed in detail in Chapter 2). Therefore, one caution is that it is all too easy to allow the capabilities of the software to dictate the analysis carried out, or conversely, the analysis never attempted. This is one reason why GIS technology has taken longer to make its way into social science research than it did the natural sciences (as much as 40 years). Nonetheless, GIS did begin to find their niche in social science applications soon after their early development, most notably their use as a tool for the collection, storage, and analysis of U.S. census data beginning with the 1970 U.S. Census.

Those early years of GIS focused on natural resources for two simple reasons. Natural resource managers in Canada, with vast amounts of information to map, originated GIS technology in the 1960s. It was relatively easy to map the location of a forest, which doesn't move and doesn't change quickly, and the computer provided an excellent platform to accomplish that task in a large government agency. Working out methods to map nondiscreet data, such as mobile human populations, concepts, and attitudes more typical of social science research, would come later, thus explaining why GIS have taken substantially longer to make their way into daily use by social science researchers.

Furthermore, early GIS were traditionally expensive and complicated to use. This made GIS a difficult technology to bring to small community groups or individuals and often required a researcher to act as an intermediary between the data provider and the GIS analysis system. To provide the additional capabilities of a spatial analysis requires the researcher to develop an understanding of spatial data characteristics, a multitude of data types, spatial and topological analysis, and spatial modeling. It is crucial that the social scientist be equipped to discuss the application of GIS technology to their data, even if a GIS expert eventually is employed to carry out the database development, analysis, or output of results from the GIS. Fortunately, powerful personal computers are now relatively inexpensive (especially compared to computers of the 1960s and 1970s, when GIS technology was in early development). In combination with easier to use GIS software, the technology is much more accessible to the end user. Many spatial analysis questions that may previously have required the expertise of a GIS specialist can now be performed directly by the researcher. Learning the ins and outs of any GIS program, as well as the essential underlying concepts, is still a relatively time-consuming process. However, with some time and effort put into learning the fundamental concepts and organization of GIS and spatial analysis, you can go a long way in answering interesting and important questions in the social sciences.

As GIS has grown in popularity, the number of trained GIS professionals has also expanded tremendously. Thus, when your questions or analyses go beyond your own experience, there are a number of people you can go to for additional assistance. In particular, check for experts in GIS at local universities, community or technical colleges, and government agencies. In many areas, there are even GIS user groups that meet regularly to share ideas, assist with questions, or show off new projects and capabilities of their favorite GIS software. Visiting the Web site of your GIS software can often provide information on these local groups as well as access to online communities of users who can be of assistance.

Summary

As a research tool, GIS provide a wide range of opportunities for examining relationships in space, often by incorporating additional information not traditionally considered in social science research. As GIS continues to become more accessible to nonexpert users, there are a wide range of new and creative opportunities in social science research. As with any new tool, GIS can be used incorrectly and can provide misleading results. Any researcher familiar with standard statistical analysis software can relate to this concept. Use of the tool is not simply a matter of collecting a data set and pushing buttons in the software. Good research requires collecting appropriate data in an appropriate fashion and then conducting an intelligently thought out analysis that makes sense in light of the hypothesis or question being addressed.

In the following chapters, we discuss several of the essential concepts all GIS users should keep in mind when developing a data set for analysis in a GIS. We also address the major GIS data formats along with their strengths, weaknesses, and applications. With these few, but important pieces of information, you will be better prepared to use GIS as a tool in your own research applications, while avoiding many of the common pitfalls encountered by new users of these tools.

Relevant Web Sites

The first two Web sites listed in this section are portals to all things GIS. In addition to basic information about GIS, they link to information about software, training, data, and a variety of other useful resources.

GIS.com: This is a general GIS portal managed by the Environmental Systems Research Institute (ESRI), one of the major GIS software providers. The site offers a variety of general information and resources relevant to getting started with GIS technology. http://www.gis.com/

The GIS Lounge: This is a general GIS portal offering a variety of general information and links related to GIS technology, software, data, and other related resources. http://gislounge.com/

Cartographic Communication: This link regarding cartographic communication is part of a larger geography education Web site developed by Kenneth E. Foote and Shannon Crum, The Geographer's Craft Project, Department of Geography, The University of Colorado at Boulder. http://www.colorado.edu/geography/gcraft/notes/cartocom/cartocom_f.html

Center for Spatially Integrated Social Sciences (CSISS): CSISS recognizes the growing significance of space, spatiality, location, and place in social science research. It seeks to develop unrestricted access to tools and perspectives that will advance the spatial analytic capabilities of researchers throughout the social sciences. CSISS is funded by the National Science Foundation under its program of support for infrastructure in the social and behavioral sciences. http://www.csiss.org/