Cambridge University Press
978-0-521-67170-5 - Historical GIS: Technologies, Methodologies and Scholarship
Ian N. Gregory and Paul S. Ell
Excerpt
More information

1

GIS and its role in historical research: an introduction

1.1 INTRODUCTION

Until the mid-1990s, most historians or historical geographers would not have heard of a Geographical Information System (GIS), despite its widespread use in other disciplines. Since then there has been a rapid increase in awareness of the potential of GIS in historical research, such that a new field, historical GIS, has emerged. This book will demonstrate that historical GIS has the potential to reinvigorate almost all aspects of historical geography, and indeed bring many historians who would not previously have regarded themselves as geographers into the fold. This is because GIS allows geographical data to be used in ways that are far more powerful than any other approach permits. It provides a toolkit that enables the historian to structure, integrate, manipulate, analyse and display data in ways that are either completely new, or are made significantly easier. Using these tools and techniques allows historians to re-examine radically the way that space is used in the discipline. As with any new approach, enthusiasm for GIS is not without risks. GIS originated in disciplines that use quantitative and scientific approaches in a data-rich environment. Historical geography is rarely data-rich; in fact, data are frequently incomplete and error-prone. As a discipline, historical geography employs both quantitative and qualitative approaches and is rightly suspicious of overly scientific or positivist methodologies. Further, many researchers are introduced to GIS as a software tool rather than an approach to scholarship. Although at its core GIS is software-based, to be used effectively it has to be seen as an approach to representing and handling geographical information that provides scholars with information on both *what* features are and *where* they are located. The researcher using GIS should be asking 'what are the geographical aspects of my research question?' rather than 'what can I do with my dataset using this software?'

This book describes how GIS can be used in historical research. Historical GIS is a highly inter-disciplinary subject combining historical scholarship with expertise in using GIS. GIS stresses the geographical aspects of research questions and datasets. This book focuses on this approach in historical research. As it is written for an audience of historians, historical geographers and others with an interest in the past, it will not say much about traditional historical approaches, which will already be familiar to the reader. The philosophy of the book is that GIS has much to offer to our understanding of history and that it must be used appropriately, critically and innovatively so that the limitations of spatially referenced data are understood, and that GIS is used in ways that are appropriate for historical research.

The first section of this chapter describes the advantages of GIS's ability to handle Geographical Information. The chapter then turns to describing how GIS can be productively used in historical research. The evolution of the use of GIS in geography and the controversies that ensued are then described so that lessons can be learnt by historians. Finally, we look at how GIS is spreading into historical scholarship and how it should be best used.

These are themes that will be further developed as the book progresses. Chapters 2, 3 and 4 introduce the fundamentals of GIS data and tools. Chapter 2 examines how GIS models the world and how this enables and limits what can be done with GIS. Chapter 3 looks at one of the more onerous and arguably under-valued aspects of historical GIS research: creating databases. Chapter 4 then explores the basic tools that GIS offers to the historian. Chapters 5 to 8 build on these basic concepts to show the different approaches to applying GIS to historical research. Chapter 5 evaluates how the mapping and visualisation abilities offered by GIS can allow exploration of the spatial patterns of the past. In Chapter 6 we explore how time can be added to basic GIS functionality to enable the study of temporal change. Chapter 7 goes beyond desktop GIS to examine the potential for data retrieval and integration over the internet. Chapter 8 then looks at how GIS can be used to perform the quantitative analysis of spatial patterns. Finally, in Chapter 9 we provide a critique of the various ways in which GIS is being used by historians following both quantitative and qualitative approaches to advance their understanding of the geographies of the past. In this way we offer an in-depth critique of technologies and methodologies of historical GIS, and explore how these are changing scholarship in historical geography.

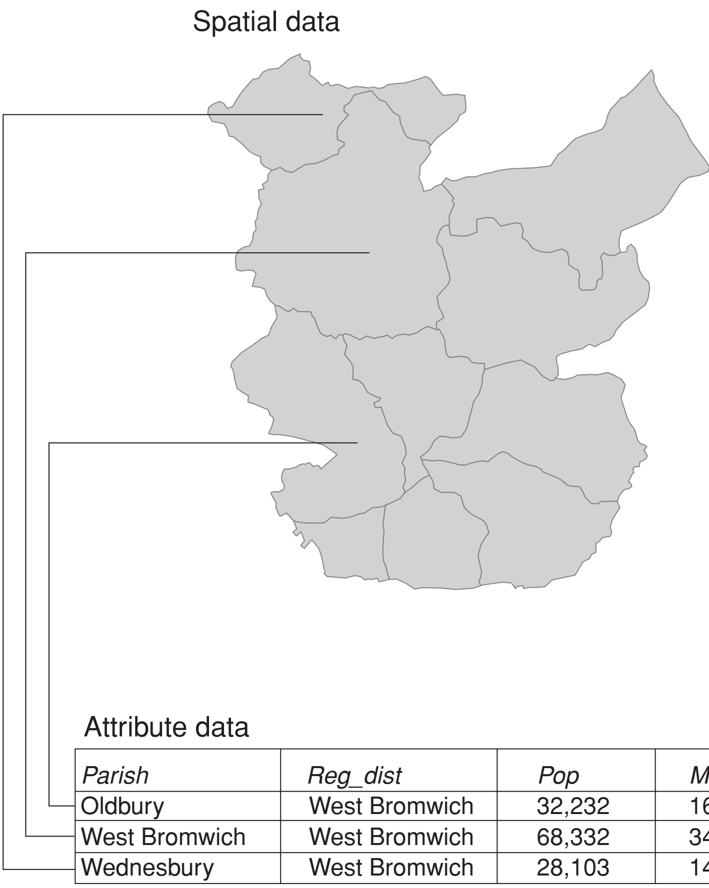
1.2 DEFINITIONS: GEOGRAPHICAL INFORMATION AND GEOGRAPHICAL INFORMATION SYSTEMS

Geographical Information Systems (GIS) and its related fields have a maze of definitions that vary between authors and depend heavily on context (Chrisman, 1999).

There are, however, some basic terms related to the subject that are important to understand. The best place to start is by defining Geographical Information (GI). At its broadest, any information that refers to a location on the Earth's surface can be considered as Geographical Information. In practice, this covers almost all information so it is perhaps more helpful to regard information as GI when it includes some information about location which is relevant to the problem at hand. The location of the railways that make up a transport network or the location of the boundaries that define an administrative system are obvious examples of Geographical Information. Census data are also Geographical Information as they consist of statistics on clearly bounded locations on the Earth's surface such as enumeration districts (EDs), census tracts, registration districts or counties. Information on historical monuments or data on hospital mortality rates can also be GI if they include information on where the features are located. Qualitative sources can also be GI. Examples include texts referring to various places, drawings or photographs of buildings around a town, and sound recordings of place names. The lack of a clearly defined location need not mean that information is not GI, although it may limit the uses to which it can be put. For example, information about 'the area around London' or a Chinese province for which no accurate boundaries exist (or may ever have existed) can still be GI.

What then are Geographical Information Systems? The narrowest answer to this is to regard GIS as a type of software. In simple terms, a GIS is a computer package that is designed to represent Geographical Information effectively. It is, therefore, a system that allows us to handle information about the location of features or phenomena on the Earth's surface. This is usually done by combining a database management system (DBMS) with a computer mapping system. Thematic information that says *what* a feature is is stored as a row of data in the DBMS. Technically, this is referred to as *attribute data*. Each row of attribute data is linked to information on *where* the feature is located. This is termed *spatial data* and is stored using co-ordinates but is usually represented graphically using the 'mapping system' (Cowen, 1990). 'Mapping system' is perhaps an over-simplification, as in addition to providing the ability to draw maps, this deals with all the functionality that is explicitly spatial, including responding to queries such as 'what is at this location?' and calculating distances, areas and whether features are connected to each other. A GIS software package is thus a geographical database management system as is shown in Fig. 1.1.

Moving beyond this limited definition involves defining GIS not as a type of software but as the tools that the software offers. A widely quoted definition was produced by a government enquiry into the use of Geographical Information which defined GIS as '[a] system for capturing, storing, checking, integrating, manipulating, analysing



Attribute data

Parish	*Reg_dist*	*Pop*	*Males*	*Females*
Oldbury	West Bromwich	32,232	16,120	16,112
West Bromwich	West Bromwich	68,332	34,371	33,961
Wednesbury	West Bromwich	28,103	14,211	13,892

Fig. 1.1 An example of a GIS as a database showing both spatial and attribute data. In this example the spatial data represent some English parishes while the attached attribute data are statistical data taken from the census.

and displaying data which are spatially referenced to the Earth' (Department of the Environment, 1987: p. 132). Other authors list slightly different characteristics, but it is broadly agreed that the key abilities of GIS are that it allows a geographical database to be created and the data in it to be manipulated, integrated, analysed and displayed.

Both of these approaches are bottom-up definitions in that they define GIS by looking at the technology or the functionality that it offers. It is also useful to take a top-down approach that asks what using GIS offers to the scholar. Researchers using GIS often talk about 'following a GIS approach'. This involves making use of both components of the data, the spatial and the attribute, to look explicitly at how spatial patterns vary across the Earth's surface. Although it can be argued that this is what geographers have always done, a GIS approach is distinguished by both the power and the limitations of how data are represented in GIS software, something that will be described in detail in Chapter 2. It is important to realise, however, that it is not necessary to use GIS software to follow a GIS approach, although it is usually advantageous to do so.

'Following a GIS approach' effectively must involve conducting high-quality geographical research in ways that are enabled by GIS. It is therefore worth considering briefly what the discipline of geography is concerned with. Johnston (1983) argues that geography is the study of places and the inter-relationships between them. He says that there are two main aspects to the geographer's approach: the vertical, which attempts to study individual places and the components that make them up in detail, and the horizontal, which attempts to study the relationships that link separate locations. It is the ability to do both, he argues, that gives geography its integrity as a discipline. Geography is thus fundamentally concerned with space. Horizontally it is concerned with the flows of people and ideas. Vertically most places can be broken down into increasingly small units to the level of the individuals and households that have distinct locations within a place. Space, therefore, frequently determines the arrangement of how people interact with each other, and with the natural and man-made environment. It is important also to note that a place is rarely, if ever, an un-sub-dividable whole. Instead, it is made up of smaller components that will also have locations. These include, for example, households within a village, or tracts within a city. As GIS is fundamentally concerned with locating data in space, clearly it has much to offer to the researcher who is interested in studying these vertical and horizontal geographical relationships.

As well as applying GIS to academic study, there is a strong research agenda that looks at how best to handle Geographical Information in appropriate and effective ways. This has become known as Geographical Information Science (GISc) (Goodchild, 1992a). The term 'science' is perhaps an unfortunate one from the historian's point of view as it implies a strongly quantitative approach. In fact, there is a significant amount of work in the GISc arena that is relevant to qualitative historical research and where historians have much to offer GI scientists. This is particularly true of how to handle qualitative data in a GIS environment: an area where significant advances have been made by historians, as will be discussed in Chapter 9.

As well as the GIS and GISc arenas, GIS is closely linked to three other research areas: spatial analysis, geocomputation and scientific visualisation. Spatial analysis (variously also referred to as spatial data analysis, spatial statistics, spatial modelling and geographical data analysis[1]) refers to the use of Geographical Information in a statistical environment. It seeks to explore spatial patterns; therefore, unlike with conventional statistical approaches, results will vary depending on where the observations are located (Fotheringham, 1997). One example of spatial analysis is testing the locations of patients suffering from a disease to see if they cluster in certain parts of the study area. Another would be testing to see if high values of one dataset tend to occur near high values of another. Spatial analysis can be done without using GIS – indeed the sub-discipline emerged before GIS – but using the two together has clear advantages as GIS software allows the spatial component of the data to be handled effectively – for example, to calculate distances or to display results through maps. The use of spatial analysis with GIS is returned to in Chapter 8.

Geocomputation is a newer term. Whereas spatial analysis uses the power of statistics to obtain knowledge from geographical data, geocomputation uses computing power and computing science. It has been defined as '[t]he eclectic application of computational methods and techniques "to portray spatial properties, to explain geographical phenomena, and to solve geographical problems"' (Couclelis, 1998: p. 17). Geocomputation emphasises using highly computationally intensive techniques such as artificial intelligence in a high-performance computing environment (Openshaw, 2000). Unlike spatial statistics, geocomputation is a very young sub-discipline whose value in academic research has yet to be fully established.

Scientific visualisation, otherwise known as visualisation in scientific computing (ViSC) (Wood and Brodlie, 1994), involves representing data graphically so that they can be explored and understood. Mapping is an obvious form of scientific visualisation that can be used effectively to explore Geographical Information. Other less explicitly spatial forms of scientific visualisation, such as graphs, are also important ways of visualising Geographical Information. Digital technology allows us to go beyond these traditional forms of visualisation into areas such as interactive maps or diagrams that can be queried and changed, animated maps that show change over time and virtual worlds that allow exploration of a detailed representation of a re-created landscape. All of these techniques are well suited to exploring Geographical Information and complement a GIS approach well (Visvalingham, 1994). Visualisation is returned to in Chapter 5.

GIS software provides a database management system that is capable of representing Geographical Information in ways that make use of both its attribute, which says what the data is, and its location, saying where it is. This 'dual-representational

[1] All of these terms can be defined in slightly different ways, but the differences are too specific to be relevant here.

framework' (Peuquet, 1994) is what distinguishes GIS, and it opens a range of functionality to allow researchers to handle Geographical Information effectively. Using a GIS approach involves making use of both of these components of data simultaneously. As will be discussed later in this chapter, this is not radically new; however, modern computing power means that many of the problems and complexities that have hampered this approach in the past can now be handled. We are thus in a position to incorporate effective use of both components of Geographical Information into historical geography.

1.3 GIS AND THE THREE COMPONENTS OF DATA

The previous section introduced the concept that Geographical Information has two components: the attribute that says what the object is, and the spatial that describes where it is located. In reality, most information has a third component: time.

Geographers and historical geographers as diverse as J. Langton and D. Massey have long argued that to understand a phenomenon fully requires a detailed understanding of all three of its components. Langton (1972) argued this from the perspective of systems theory. He argued for what he termed *diachronic* analysis in which the parameters in the system could be measured as frequently as possible so that the effects of a process could be measured. In this way, inappropriate or unrealistic assumptions such as a system being in equilibrium could be avoided. Implementing this effectively would require large amounts of attribute, spatial and temporal information to measure the impact of change over time and space and thus allow an understanding of the process to be developed. Without detail in all three components, the understanding that could be gained will inevitably be simplistic, heavily reducing the effectiveness of this form of analysis.

More recently, Massey (1999 and 2005) also stressed the importance of an integrated understanding of space and time. She argued that to understand an individual place at a set point in time we need to understand the chronology of how the place developed. This provides a single story or example of how a place develops and thus how a process can operate. To gain a fuller understanding requires multiple stories giving a variety of examples of how different places develop differently. These authors, writing nearly thirty years apart, and from very different perspectives, have a common thread: that gaining understanding requires an understanding of attribute, space and time together.

In an ideal world, information on all three components of the data would be available. Unfortunately, this is not often the case, especially when working with

Cambridge University Press
978-0-521-67170-5 - Historical GIS: Technologies, Methodologies and Scholarship
Ian N. Gregory and Paul S. Ell
Excerpt
More information

Table 1.1 *The representation of geographic data in various formats*

	Fixed	Controlled	Measured
Census data	Time	Location	Attribute
Soils data	Time	Attribute	Location
Topographic map	Time	Attribute	Location
Weather reports	Location	Time	Attribute
Flood tables	Location	Time	Attribute
Tide tables	Attribute	Location	Time
Airline schedules	Location	Attribute	Time

Source: adapted from Langran and Chrisman, 1988: p. 2.

historical sources. Langran and Chrisman (1988) argue that in many instances, providing a detailed measurement of one of the three components of data requires the second to be controlled or limited and the third to be completely fixed. One example of this is the census: to measure information about people and households as an attribute accurately, the census arbitrarily subdivides the country into discrete spatial units, to control location, and is taken on a specific night so time is fixed. Soils data provide a different example. Soils are first subdivided into arbitrary classes, controlling attribute, and the survey is again done at a fixed time. The boundaries between the different soil types are recorded, so location is accurately measured. A variety of examples of this type are shown in Table 1.1.

Representing a source provides a second layer of complexity. Again, choices are frequently made that further erode the detail inherent in at least one of the three components. To return to the example of the census, many researchers have analysed the census using a database management system. These have no concept of location beyond place name and administrative hierarchy. As these are extremely crude forms of spatial information, location is no longer simply controlled; it is effectively removed. Less extreme examples include simplifying the spatial component by aggregation or using approximate locations.

The previous section described how GIS enables attribute and location to be handled in an integrated manner. It is not, however, explicitly designed to handle time, although strategies can be developed to work around this, as will be discussed in Chapter 6. This means that GIS has the ability to handle the detail of all three components of the data contained in a source. To illustrate the importance of this, suppose that there is a database of information on historic artefacts such as pieces of pottery. This contains attribute information that describes the pottery, locational information that gives the co-ordinates where the artefact was found and perhaps where it was made, and temporal information on when the pottery was made and perhaps when it was found. Any database could store this information but only GIS is able to handle it effectively. For example, in response to the query 'where has pottery of a certain kind dating from the 1850s been found?', the GIS would map these results and may show, for example, that these were only found in one quite limited area. Repeating the query for later dates may reveal if and how the pottery spread out over time. Even in this simple way, using GIS allows the researcher to make improved use of all three of the components of the data up to the limitations of the source.

Census data provide a more complex example. As described above, a census measures the number of people by controlling location and fixing time. Using GIS allows the attribute and locational components to be handled together, as the GIS will store the attribute information combined with a graphical representation of the administrative units used to publish the data. This allows the exploration of two components of the data rather than just one. Exploring change over time from the census requires multiple censuses to be represented. GIS enables this by integrating data through location and by using maps and spatial analysis as tools for comparison. It does not automatically solve the problems inherent in the sources, such as changes in the questions asked and the administrative areas used, but it does, as will be returned to in later chapters, provide a way into these problems.

A GIS approach, therefore, helps the researcher to make full use of all three components of a source. In theory, it makes no requirements on the approach that the researcher adopts. GIS is applicable to any form of historical scholarship, quantitative or qualitative, as long as the research is concerned with information on theme, location and, if appropriate, time. In practice, however, GIS requires the data to be structured in certain ways that are often not sympathetic to many forms of historical sources, particularly those containing uncertainty, ambiguity and missing data. It can also be argued that GIS is better at handling quantitative, clearly defined sources such as the census, rather than qualitative or vaguer sources such as textual descriptions. This will be returned to in later chapters but, as the book will demonstrate, these 'softer' sources can be used effectively in GIS.

1.4 THE BENEFITS AND LIMITATIONS OF GIS

From the above discussion it should be clear that a GIS is a form of database management system that allows the researcher to handle all three components of information in an integrated manner with a particular emphasis on location. Gregory *et al.* (2001a) group the benefits of using GIS into three categories: the organisation of historical sources, the ability to visualise these sources and the results gained from analysing them, and the ability to perform spatial analysis on the data. None of these are impossible without GIS, but all may be made significantly easier by using it.

Each item of data in a GIS database contains a co-ordinate-based reference to its location on the Earth's surface. This provides a framework for organising a database that has many benefits. Most obviously, it allows the researcher to query the database to ask where objects are and how their location relates to the location of other objects. Co-ordinates are a particularly useful tool for integrating data from diverse sources. To return to the examples above, it may be that a researcher wants to investigate the relationship between pottery and certain population characteristics recorded in the census – for example, to do with occupations. As both the pottery database and the census database have co-ordinate-based spatial data we are able to integrate the two and find which census zone each piece of pottery was found in. It may be that the datasets use different co-ordinate systems or different map projections, but GIS software should provide the tools necessary to standardise these. Once the two have been integrated the researcher can examine the relationship between the two datasets using whatever methodology he or she sees fit. Chapter 4 describes this in detail.

A further use of co-ordinates in data integration is associated with resource discovery. With the growth of the internet there is a wealth of data available online. However, finding appropriate data can be difficult. The use of metadata, data that describes a dataset, has become a standard way of providing information to online catalogues (Green and Bossomaier, 2002). Co-ordinates are a particularly useful form of metadata. For example, a researcher may be interested in a particular part of the country, say the hundred[2] of Eyhorne in Kent. As this is not a common way of referring to this area it is likely that searching for the name 'Eyhorne' will result in many datasets of relevance to this place being missed. If co-ordinates describing the extent of Eyhorne are used instead, the fact that, for example, our pottery database or census data contains co-ordinates lying within this extent can be discovered. This is discussed in Chapter 7.

The second advantage offered by GIS is the ability to visualise data, particularly through mapping. In GIS the map is no longer an end product; it is now a research tool. As soon as the GIS database is created it can be mapped. This means that the spatial patterns within the data can be repeatedly re-explored throughout the research process, greatly enhancing our ability to explore and understand spatial

[2] Hundreds were a type of English administrative unit that was in use until the nineteenth century but have since fallen out of use.