Volume 67 Number 1 Spring 1991

Guest Commentary

Geographic Information Systems

MICHAEL F. GOODCHILD
University of California
Santa Barbara, California

INTRODUCTION

It has been difficult to work with maps or geographic data in the past few years without encountering the term GIS, or geographic information system. Indeed, a significant industry valued in the hundreds of millions of dollars (Bylinsky 1989) has grown up to meet the expanding demand for this technology. Although the field was first defined in the 1960s, it has benefited enormously from 1980s interest in high technology, and from the vast improvements in hardware and software that have occurred in the past three decades. The current explosion of interest in GIS actually dates only from the early 1980s, when hardware costs fell below a widely acceptable threshold and certain essential software components became sufficiently widely available.

Many attempts have been made to define GIS and to distinguish it from related technologies such as CAD (computer assisted design). But the most striking distinguishing characteristic of a GIS is geography—the fact that every record or digital object in a GIS has an identified geographical location. Given the importance of location in marketing, it is perhaps surprising that there has been so little discussion of GIS in the market research literature. Annitto and Cromley (1988) find that compared to resource management and land use planning, there have been few attempts to apply GIS to market analysis. Two recent guest commentaries in the *Journal of Retailing* on information technology and its role in marketing

---

The National Center for Geographic Information and Analysis is supported by the National Science Foundation, grant SES 88-10917.

make no mention of GIS at all (Achabal and McIntyre 1987; Beaumont 1988a).

At its simplest, a GIS merely provides geographic access to records. Even at this level, however, substantial technology is required to handle the geographic component. GIS needs specialized hardware for input (digitizers, map scanners) and output (plotters and graphic displays), and specialized techniques for storing the coordinates of complex geographic objects, including networks of lines (roads) and reporting zones (e.g., ZIPS, census tracts, or counties). But with these enhancements, a conventional database can answer simple geographical queries. Besides the usual information system functions, a GIS can give the user the capability to point at a map display and ask for the attributes of a particular place, or to ask the system to display the location of an object given its attributes: ''tell me the demographic characteristics of this area'' or ''show me the locations of all of my regular customers.''

This interpretation of GIS as a geographical key to conventional records accounts for perhaps half of all applications, and well over half of the industry's current value. Local governments use such capabilities to keep track of land ownership, assessment, and taxation. Utility companies use GIS for managing their distributed facilities, such as underground pipes, valves, and transformers. Systems are used to answer routine queries, and to maintain and update maps. The acronym AM/FM (Automated Mapping/Facilities Management) is commonly used to refer to this class of GIS applications.

CAD technology is also designed to deal with spatial objects, and is similar to GIS in many ways. But GIS differs at this level in linking spatial objects with their distinguishing attributes, providing the ability to access records through their geographical locations. Not surprisingly, many CAD vendors also offer GIS products (e.g., Intergraph Corp. and Prime Computervision), and many successful GIS architectures exploit tight integrations between CAD packages such as AutoCAD and database management systems.

Geographic Relationships

This linkage between attributes and geography is only one distinguishing characteristic of GIS. Although it is useful to be able to retrieve records geographically, the real power of a GIS derives from its ability to link records based on their relative geographic locations. Geographic relationships can take an enormous variety of forms, and Table 1 gives common examples of relationships between the three types of geographic object:

TABLE 1

Examples of Relationships Between the Three Types of Geographic Object

	Point	Line	Area
Point	Is nearest to		
Line	Is located on	Crosses	
	Is nearest to	Intersects	
Area	Lies within	Crosses	Overlaps with
	Is outside	Touches	Is adjacent to

point, line, and area. For example, a point might represent a customer location, a line might be a street, and an area might be a ZIP code. Thus a customer and a ZIP code might be related by the *lies inside* relationship; a customer and a street might be related by the *located on* relationship. Relationships between objects can be binary (relationship is either true or false), as in these examples, but can also carry attributes based on continuous scales of measurement; for example, the distance between a customer and a retail store.

The ability to handle relationships between objects gives a GIS enormous power as an analytic and modeling tool. Besides answering simple queries about objects such as pipes, utility companies also exploit the modeling power of a GIS to analyze flows, linking pipes together by relationships such as *is connected to* or *flows into*. The needs of GIS for handling complex sets of relationships between different classes of objects turn out to be very well satisfied by the relational model (Codd 1990), and the majority of current GIS products exploit relational database technology in one form or another.

It is useful to distinguish between three types of relationships. The first are those that allow complex objects to be constructed from simple ones, such as the relationship between a line and the points that define it. These relationships are essential if the system is to handle a full range of geographic objects, and they are computed and stored as the database is built.

The second are those that can be computed geometrically from the locations and shapes of the objects, such as *contains* or *is nearest to* or *crosses*. Most GISs have the ability to compute and store this second type of relationship and use the results in analysis. For example, it is possible to assign points to areas (e.g., customers to census tracts) in most GISs using the "point-in-polygon" operation. The more sophisticated systems allow a certain amount of uncertainty in many of these operations, as geographic locations are never known with absolute accuracy (Blakemore 1984).

The third category of relationships are those that cannot be computed from the geometry, and must therefore be input directly. These include the *intersects* relationship between streets, which is not the same as *crosses* because of the possibility of overpasses, and is therefore important in any traffic routing application. Two-dimensional geometry alone is not sufficient to determine whether two crossing streets actually intersect at grade.

GIS Functionality

As with any information system, it is essential that a GIS provide routine housekeeping functions, including data input and storage. Since the data is geographically referenced, a GIS should be able to generate a map output. However it is able to offer much greater power than a simple automated mapping system for two reasons: because of its analytic functions; and because of the importance of relationships between objects, which are not significant in simple mapping.

The basis of GIS power lies in its ability to carry out several unique classes of functions:

- computing spatial relationships between objects;
- using spatial relationships to assign new attributes to objects; e.g., the distances of customers from a shopping center location;
- computing new objects from existing ones; e.g., a trade area (area object) from customer locations (point objects);
- analyzing the attributes of a class of geographically referenced objects; e.g., selecting all customers with income >$30,000 and displaying their locations;
- analyzing multiple classes of objects; e.g., forecasting trips to shopping centers based on characteristics of neighborhoods (area objects), shopping centers (point objects), and trips (point-area relationships).

CURRENT GIS PRODUCTS

Current GIS products run on almost every imaginable hardware platform, from the PC to the mainframe, and under almost every currently available operating system. The *GIS Sourcebook* is an excellent survey of vendor products published annually by GIS World Inc., Fort Collins, Colorado, and general information on the field is widely available through the US periodicals *GIS World* and *Geo Info Systems* and the UK periodical *Mapping Awareness*.

A comprehensive GIS is an immensely complex software product, combining the database capabilities of a relational database management system, the graphics capabilities of a CAD system, and specific functionality for handling such unique issues as map projections and geographic relationships. In this environment most vendors inevitably seek specialized market niches, although some, notably ESRI (Environmental Systems Research Institute, Redlands, California), market strongly into all application areas with a consistent product running under Unix, VMS, VM/CMS, Pios and other proprietary operating systems, as well as a PC version under MS-DOS.

Existing systems also differ in their underlying data model. Geographic variation is inherently complex, and the expression "the closer you look the more you see" is almost universally true. In marketing, demographic data become less and less accurate as they are aggregated, so here too there is a direct relationship between accuracy and geographic scale. A GIS's data model is the set of rules used to represent and approximate geographical reality in digital form. For example, smoothly curving streets are often represented by sequences of straight lines connecting points. Data models vary from one system to another, and two GISs applied to an identical application might well use entirely different methods of digital representation.

In the current industry, data models differ in two important ways. One concerns the basic representation of a single object. A road, for example, can be represented as a set of lines or vectors connecting discrete points, or by "turning on" a selection of cells in a regular array or raster. In general, current systems fall into one camp or the other, although there is increasing integration.

The second distinction, and one that is particularly relevant for marketing applications, concerns the representation of a partition of an area into reporting zones, for example ZIPs. One approach sees a single zone as the basic object, and uses this as the unit of representation. The other recognizes that zones are not independent of each other, as they must share common boundaries, exhaust the space, and not overlap. The basic object in this representation is the arc, or stretch of common boundary between two zones. In the complex and often confusing informal terminology of GIS, the latter approach is often called "topological."

Of the alternatives, vector seems much more appropriate for marketing applications than raster, and has been much more successful in the marketplace, because of its greater ability to represent discrete objects such as streets, customer locations, and reporting zones. The topological approach is preferred because it leads to greater efficiency in processing, and to less redundancy in databases.

This view of the world as populated by discrete objects is probably appropriate for almost all marketing applications, but in other areas it may be more useful to look at geographical variation as if it were continuous. This field view is common in analyzing topography, aerial photographs and images from space, and climatic variables. Although fields and objects have sometimes been seen as competing approaches to GIS, roughly corresponding with raster and vector systems respectively, the current consensus is that both are necessary, and that any comprehensive product must accommodate both.

GIS APPLICATIONS IN MARKETING

Applications of GIS technology in marketing range in sophistication from simple data exploration to numerical modeling. In this section I look at four distinct types, that together cover the majority of uses.

Spatial Organization of Information

The ability to examine data in its spatial context is enormously important. A simple map gives the eye immediate access to geographical proximity and therefore to the role of distance as an explanatory factor. A customer name, "John Doe," conveys very little, but if John Doe's street address is used to assign geographic coordinates, then he can be related to the known characteristics of his neighborhood, and his proximity to various shopping alternatives. The geocoding process takes a simple street address and matches it to a database containing address ranges for every link in the street network. In this way customer lists can be converted into simple dot maps, or coupled with census data and displayed as maps of market penetration.

Market analysts must cope with a variety of data problems, including the propensity of reporting zones to change through time. Although some sets of zones are relatively stable, the lower levels of census geography such as enumeration districts which are essential for market analysis are revised at every census, making longitudinal analysis difficult. Postal zones such as ZIP codes have become increasingly popular recently because of the ease with which they can be linked to address records, but unfortunately the quality control of postal zone geography is poor. Hence market analysts frequently face the need to transfer data from one set of zones to another, when the second set fails to respect the boundaries of the first. In general, basis change describes any transfer of attributes from one set of spatial objects to another; the primary application is the transfer of socioeconomic and demographic statistics from one set of reporting zones to another. The influence of variable reporting zones on the results of analysis is commonly called the "modifiable areal unit problem."

Beaumont (1988b) sees basis change as the major area of application of GIS technology in marketing, based on the GIS's ability to handle multiple sets of spatial objects, to compute the relationships between objects, and to base estimates on those relationships. Goodchild and Lam (1980) showed how the GIS operation of polygon overlay could be used to compute the areas of overlap between reporting zones, and how these could be used as coefficients in estimation. Flowerdew and Green (1989) have extended the technique further, and Arbia (1988) has looked at the overall problem of basis change within a GIS environment.

Exploratory Spatial Analysis

A map is a static presentation of data that is very efficient at allowing the eye to explore geographical distributions and certain types of spatial relationships. But with a complex dataset, the single perspective offered by a map may be too limiting to be of much value. Recently there have been several significant efforts to use GIS technology to provide market analysts with tools for exploring geographical markets. In a sense this exploratory spatial analysis or ESA is the equivalent for spatial data of Tukey's exploratory data analysis (Tukey 1977), which is a set of simple methods for exploring patterns and structures in numerical data.

One obvious advantage of ESA is that it can avoid the high cost of transporting a real estate specialist to a site. With access to appropriate data and the necessary tools to explore and visualize, one could carry out detailed site evaluations from a remote digital environment. Emergency management services are already using GIS systems that allow attributes to consist of images (e.g., NOAA's CAMEO), so that a photographic view of a burning building can be retrieved before the fire truck arrives at the scene, allowing the crew to develop a detailed strategy in advance.

The data needed to assess a market is multidimensional, and varies geographically as well as temporally. A map is limited to a single geographical slice of this multidimensional data matrix, for one variable and one point in time. In the digital environment it should be possible to explore the data matrix in a much more versatile manner, and the Voyager system provides a recent example of this type of ESA. The user is able to open several windows showing different views of the dataset; one might contain a spatial view, and one a temporal view. When the user points a cursor at a location within the spatial window, the temporal window shows how a particular variable has changed through time at that location. A third window might be used to select the variable of interest.

Another simple type of ESA might be applied to site selection. Suppose the key variable is the total expenditure within one mile of the site, and we wish to rate sites on this basis from maps of average-per-capita expenditure and population density. This would be difficult, as the eye is not good at combining information from two different maps and integrating the result within one-mile circles. But in a digital environment it is easy to imagine a system that allows the user to move an integrating circle around the screen, continuously updating the desired variable. The value might even be discounted to allow for the locations of existing competitors. This kind of capability has been available for some time in the form of dial-up reports on single sites, but technology has now advanced to the point where it is feasible on a real-time, continuous basis.

Monmonier has described a different kind of ESA called ''geographical brushing'' (Monmonier 1989). To understand the relationship between two variables, the user moves a brush within a spatial window. The locations that fall within the current position of the brush are analyzed, and a bivariate plot is displayed within a second window.

Site Modeling

The analytic aspects of early GIS technology were directed toward simple functions like measuring area, selecting and counting objects, and computing relationships. However the range of possible analyses can be extended very quickly by linking the GIS to a standard statistical package, or by incorporating such a package into the GIS itself. This allows the full range of statistical methods to be combined with the GIS's power in working with spatial relationships, and in displaying results in their geographical context.

One common form of site modeling is the use of regression analysis to relate a store's sales to its physical and marketing attributes, such as signage, parking spaces, floor area, traffic conditions, and demographics. The coefficients of each variable allow us to predict sales for a new site, or the effects of changing one or more of an existing site's attributes, such as new signage or reduced parking.

There are several benefits to be gained by performing regression analyses within a GIS environment. The GIS provides an efficient way of generating some of the necessary variables, particularly demographic counts within each store's surrounding area. A map of residuals from the analysis may provide useful insights into the factors that have not been included in the regression model, or into geographical variation in the model's performance.

Spatial interaction models (Fotheringham and O'Kelly 1989) can be used to analyze and predict consumer spatial behavior, but are notoriously difficult to calibrate within standard statistical packages. In part this is because all spatial interaction models work with several types of spatial objects at once: attributes of neighborhoods, destinations, and trips. In part it is because the functions for working with spatial objects that exist in a GIS are not present in standard statistical packages. For example, the user has functions available within a GIS for measuring trip lengths through street networks, or for locating a representative centroid within a neighborhood, or for aggregating customers from points to census tracts.

Spatial-Decision-Support Systems

Location-allocation is the term given to a set of techniques for searching an area for the optimum sites for a number of central facilities, and many applications of location allocation to retail site modeling have been described (Ghosh and Rushton 1987). Simple location-allocation packages are available, but they suffer from several disadvantages. Although they produce the optimum sites for a particular objective, inevitably the final decision must pay attention to factors that have not been included in the optimization. Therefore it helps to know how much optimality must be sacrificed when the solution must be modified by new factors. It is also useful to have efficient ways of building the database for a location-allocation analysis, and of displaying the results geographically.

Recently there has been significant interest in the concept of a spatial-decision-support system (SDSS), based on a GIS database. Densham and Goodchild (1989, 1990) have described the current state of development of SDSS, and proposed a research agenda. Like a DSS, an SDSS must address the needs of a user faced with a poorly structured problem, and provide the interface and knowledge to help the user work with a set of well-structured models.

OUTLOOK

If a GIS is a system for handling spatially referenced information, and if spatial referencing is very significant in marketing because of the importance of location, then why is this area of GIS application lagging behind others at this time? In this section I look at current impediments to the use of GIS technology in market analysis, and the prospects for the future.

Data

The technology available to market researchers has changed enormously over the past few decades, as computers, statistical packages, and now GIS software have become available. However, although much data is now available from data collection agencies in digital form, there has been remarkably little change in the process by which data are collected and distributed. The census, which continues to be the major source of archival data, is still carried out every ten years, then tabulated and distributed in tabular form for reporting zones.

We have seen how geocoding allows records to be placed in their geographical context, by matching addresses to street files. In the US the primary data tool for geocoding is the TIGER database, which is in the process of replacing the earlier DIME. Both databases were developed for internal census purposes: as geocoding tools they suffer from incomplete coverage and currency that is geared to the ten-year census cycle. As a result, a 70 percent "hit" rate is considered good for a typical customer mailing list—30 percent of addresses will not be successfully geocoded.

With the technology currently available it would be a simple matter to build and maintain a complete, current national database of all streets, which could successfully geocode close to 100 percent of addresses. The problems that have thus far prevented this are almost entirely administrative: lack of a consistent, uniform method of addressing in rural areas (Gribb, Czerniak, and Harrington 1990), problems of coordinating input, and lack of a federal agency with the appropriate mandate. In general, data collection agencies have been hard pressed to maintain traditional operations, far less undertake new activites in response to the opportunities offered by a new technology.

Software

The GIS software products currently available for marketing applications fall into two broad categories: general toolboxes that offer generic capabilities across a wide range of applications (e.g., ESRI's ARC/INFO); and packages that have been developed specifically for marketing and similar applications (e.g., Strategic Mapping's ATLAS*GIS or MapInfo). The first group have been slow to exploit marketing applications for several reasons: more traditional markets, such as resource management, have been strong; marketing has not been perceived as a large market in relation, say, to county government; and a package sufficiently versatile to handle the wide range of marketing applications requires extensive training.

The second group have been more successful. The simple mapping capabilities offered by packages such as SAS and Strategic Mapping's ATLAS*GRAPHICS have sold well to market analysts, and more recent packages have offered significant GIS functionality.

In the longer term a convergence seems likely with the development of more specialized marketing capabilities within the generic GIS packages, particularly better access to statistical analysis, and with wider capabilities within the specialized marketing packages. However any package is ultimately constrained by its own data model and conceptual framework, and it is often difficult to add sophistication to a simple conceptual design.

Hardware

The current explosion of interest in GIS began around 1980 when the new range of super-minicomputers typified by the VAX became available: for the first time the computing power necessary to support GIS became accessible for a price that was affordable to many resource-management agencies. Since 1980 cost-effectiveness has improved by at least an order of magnitude, so that the $250,000 minimal GIS hardware configuration of 1980 is now available based on workstations for under $25,000.

On the other hand, the desktop mapping market is a more recent development of the PC world, particularly the AT and 386-based machines. GIS packages developed specifically for market research applications tend to be PC-based, and there is a significant hardware barrier between them and the workstation-based generic GISs.

The State of GIS

At the highest level of abstraction, GIS is an emerging science of spatial information. It deals with how to collect, compile, store, analyze, and display spatial data within a digital environment, raising explicit questions that have previously always been implicit within spatial analysis, such as the measurement of the accuracy of spatial data (NCGIA 1989). It deals with the theory of spatial relationships and the representation of geographical variation within formal models.

At a more immediate level, however, GIS suffers from all of the problems of a rapidly expanding industry. It lacks a clear focus, a set of institutional structures around which it might be organized. It has no icons and few heroes. There is confusion about what it is, how to find out about it, what to buy. These are very early days in GIS, particularly in application fields like marketing. Although progress has been made, the lack of a clear consensus on these issues remains a significant impediment to the use of GIS in market analysis.

In the long run, much of our intuitive knowledge about the world is organized spatially. The residential location of a customer is used extensively in market research to impute other characteristics, particularly from the aggregate attributes of the neighborhood. Location is particularly important in retailing, which serves the needs of a dispersed population from a few central locations. These points suggest that in the long term ESA may be of more significance than EDA, and that the potential market for GIS is at least comparable to that for the major statistical packages. We may be at the early stages of a sustained interest in the geographical context of activities. Geography can be used to organize intuitive knowledge, though it will never substitute for economics or psychology as a basis for understanding human spatial behavior. Nevertheless, the spatial context is essential if market research is to understand the implications of economic or psychological processes.

REFERENCES

Achabal, D. D., and S. H. McIntyre (1987), "Guest Editorial: How Information Technology Is Reshaping Retailing," *Journal of Retailing,* **63,** 321–325.

Annitto, R. N., and R. G. Cromley (1988), "MARKMAP: A GIS for Small Scale Market Area Analysis," in R. T. Aangeenbrug and Y. M. Schiffman, (eds.), *Proceedings, International Geographic Information Systems (IGIS) Symposium: The Research Agenda,* Washington, D.C.: NASA, **III,** 37–44.

Arbia, G. (1988), *Spatial Data Configuration in the Statistical Analysis of Regional Economics and Related Problems,* Dordrecht: Kluwer.

Beaumont, J. R. (1988a), "Guest Commentary: An Overview of Decision Support Systems for Retail Management," *Journal of Retailing,* **64,** 361–374.

——— (1988b), "Store Location Analysis: Problems and Progress," in N. Wrigley, (ed.), *Store Choice, Store Location and Market Analysis,* London: Routledge.

Blakemore, M. (1984), "Generalization and Error in Spatial Databases," *Cartographica,* **21** (2,3):, 131–139.

Bylinsky, G. (1989), "Managing with Electronic Maps," *Fortune* (April).

Codd, E. F. (1990), *The Relational Model for Database Management: Version 2,* Reading, MA: Addison-Wesley.

Densham, P. J., and M. F. Goodchild (1989), "Spatial Decision Support Systems: A Research Agenda," *Proceedings, GIS/LIS '89,* Bethesda, MD: ASPRS/ACSM, **2,** 707–716.

——— (1990), *Research Initiative 6: Scientific Report for the Specialist Meeting,* Technical Report 90–5, Santa Barbara, CA: National Center for Geographic Information and Analysis.

Flowerdew, R., and M. Green (1989), Statistical Methods for Inference Between Incompatible Zoning Systems," in M. F. Goodchild and S. Gopal, (eds.), *Accuracy of Spatial Databases,* Basingstoke: Taylor and Francis.

Fotheringham, A. S., and M. E. O'Kelly (1989), *Spatial Interaction Models: Formulations and Applications,* Dordrecht: Kluwer.

Ghosh, A., and G. Rushton (1987), *Spatial Analysis and Location-Allocation Models,* New York, NY: Van Nostrand Reinhold.

Goodchild, M. F., and N. S. Lam (1980), "Areal Interpolation: A Variant of the Traditional Spatial Problem," *Geoprocessing,* **1,** 297–312.

Gribb, W. J., R. J. Czerniak, and J. A. Harrington, Jr. (1990), "Rural Addressing and Computer Mapping in New Mexico," *Professional Geographer,* **42** (4), 471–480.

Monmonier, M. (1989), "Geographical Brushing: Enhancing Exploratory Analysis of the Scatterplot Matrix," *Geographical Analysis,* **21,** 81–84.

NCGIA (1989), "The Research Plan of the National Center for Geographic Information and Analysis," *International Journal of Geographical Information Systems,* **3,** 117–136.

Tukey, J. W. (1977), *Exploratory Data Analysis,* Reading, MA: Addison-Wesley.