December 1998

Presenting Map-based Information When Using Geographic Information Systems

Cheri Speier
Michigan State University

Morgan Swink
Michigan State University

Presenting Map-based Information When Using Geographical Information Systems

Cheri Speier
Morgan Swink
Michigan State University

Abstract

Geographic information systems (GIS) have taken on an increasingly important role supporting organizational decision making. The effectiveness of GIS as a decision support tool comes primarily from the visual display of data in the form of maps. When presenting information as a geographic map, the level of data aggregation used potentially affects aspects of task complexity such as information load and the potential for pattern recognition by the user. Other task attributes related to data aggregation effects include problem size, data dispersion, and users' spatial orientation skills. Results from an experiment indicate that all of these map information characteristics have significant influence on decision performance.

Introduction

Business entities are increasingly using geographic information systems (GIS) to support organizational decision making (Payne, 1993; Wagoner and Masser, 1996). GIS is a tool for storing and displaying spatially or geographically related data (Keenan, 1997). Most GIS research to date has compared information presented in tables to that presented in maps on decision making performance (Crossland, Wynne, and Perkins, 1995; Smelcer and Carmel, 1997). However, there has been little research investigating "how" to present map-based information (Keenan, 1997; Mennecke, 1997; McMaster and Shea, 1992). Burrough (1992) suggests over 2000 combinations of factors exist influencing a map-based information display. Therefore, both GIS designers and users need to better understand how to present map-based information to improve GIS-based decision performance (Burrough, 1992).

The most challenging aspect of presenting map-based information is to manage the overall complexity inherent in such data-laden problems (McMaster and Shea, 1992). The complexity of map-based information is generated from a number of sources: number of observations, level of aggregation, spatial variability, (Burrough, 1992; McMaster and Shea, 1992). Complexity increases as these various factors interact and the features of the map become less easy to visualize (McMaster and Shea, 1992).

A key decision when presenting map-based information pertains to the level of data aggregation displayed by the map (i.e., the degree to which the data is summarized). Ideally, the modeling and presentation of data in the GIS should accurately reflect all significant operating levels relevant to the decision. However, due to the enormous size and complexity of the data supporting these decisions, decision modelers are forced to aggregate the data in ways that make decision modeling tractable given the magnitude of the data set. Prior research on data aggregation using tabular information indicates data aggregation directly influences decision quality (Barefield, 1972; Chervany and Dickson, 1974). However, the influences of data aggregation on decision processing and performance using GIS and map-based information are not well understood.

Theory Development

Complexity theory provides a useful framework for discussing decision making processes and decision effectiveness. One objective complexity measure suggests that complexity is a function of the number of distinct acts/information cues that need to be processed (Wood, 1986). Other researchers have conceptualized complexity as a function of the number of "knowledge states" possible in the task (Newell and Simon, 1972; Card, Moran, and Newell, 1983). Recent research suggests that the number of knowledge states present in geographic problems explains task difficulty to a great degree (Smelcer and Carmel, 1997).

Problem Size

One way to increase the objective complexity of map-based information is to increase the number of data points that need to be examined in order to make a decision. Several studies have shown that decision performance degrades for geographic decision making tasks as the number of decisions and knowledge states increase (Taylor and Iwanek, 1980; Crossland, Wynne, and Perkins, 1995; Swink and Robinson, 1997). The theory underpinning these findings states that increased problem size degrades decision performance due to increases in information load. Accordingly, we forward the following hypotheses.

H1a: Larger problems result in lower decision quality than smaller problems.
H1b: Larger problems result in greater decision time than smaller problems

Data Aggregation

For most problems solved with map-based information (with or without a GIS), it is common practice to aggregate individual data points into geographic clusters (representing customer accounts, franchise location, etc.). Aggregation reduces the scope of the problem to be formulated and the amount of data required, thereby reducing complexity.

Maps enable decision makers to use visual heuristics, reducing the number of outcome paths available for examination (i.e., reducing problem complexity) (Smelcer and Carmel, 1997). Additionally, decision makers' abilities in recognizing or establishing patterns in visual data influence performance outcomes for various types of geographic problems (Krolak, Felts, and Nelson, 1972; Scriabin and Vergin, 1975; Taylor and Iwanek, 1980). Taylor and Iwanek (1980) demonstrated that decision makers produced better solutions for facility location and transportation problems when more customer zones were included in the problems. They suggested that increasing the number of customer elements facilitated resolution by enabling decision makers to identify structures or patterns inherent within the map. This rationale led us to the following hypotheses.

> H2a: Problems containing disaggregated data result in higher quality decisions than those with highly aggregated data.
>
> H2b: Problems containing disaggregated data result in greater decision time than those with highly aggregated data.

Data Dispersion

Previous research points to the dispersion of data as a key component of problem complexity for map-based visualization problems. Taylor and Iwanek (1980) found that decision makers had the most difficulty solving facility location problems when customer demands were uniformly dispersed, containing no apparent groupings. When demand dispersion is uniform or has a very low variance, no patterns can be discerned from the data inhibiting effective decision making. Conversely, highly dispersed data should simplify problem solving by producing stronger, more recognizable patterns. Therefore, we state the following hypotheses.

> H3a: Highly dispersed data results in higher decision quality than less dispersed data.
>
> H3b: Highly dispersed data results in less decision time than less dispersed data.

Research Method

A laboratory experiment implementing a three-factor within subjects design was used to test the hypotheses. Between subjects factors included two levels of problem size and four levels of data aggregation. The within subjects factor was represented by three levels of demand dispersion. Each subject solved two facility location problems. The order in which subjects solved the problems was counterbalanced to reduce learning effects and subjects were randomly assigned to experimental treatments.

Subjects participating in the experiment were undergraduate students at a large university. Subjects volunteered and received participation points equivalent to 1% towards their course grade. Subjects were trained and had actively used the GIS to complete coursework prior to their participation in the experimental session.

Problem Solving Task and GIS

The experimental task environment consisted of a facility network design problem (Hax and Candea, 1984) requiring decision makers to determine the number and locations of facilities to open which minimized the total cost of fulfilling customer product demand. Subjects solved these problems using an interactive, menu-driven GIS representative of those used in industry and has been validated in prior research (Robinson and Swink, 1994; Swink and Robinson, 1997). The GIS unobtrusively recorded all of the user's candidate network configurations, decision times, final solution configuration, and final solution cost.

Subjects were also given documents detailing customer product demand data. The documents identified potential facility sites and provided transportation distances among customer zones and facility sites. Also provided were fixed annual facility operating costs, production cost rates, transportation rates, and demand figures for each customer zone. Geographic data was presented in the form of a map of the continental United State with potential facility sites represented on the map. In addition, each demand zone was identified and numbered; zones were shaded a to indicate the overall demand for product associated with that zone. Five different levels of shading were used to indicate the levels of demand in the various zones.

Independent Variables

The experiment included small problems operationalized as 20 potential facility sites and large problems operationalized as 40 potential facility sites. These problem sizes have been validated in prior use with DSS (Swink, 1995; Swink and Robinson, 1997).

Customer demand was represented by the demand for product required by a specific geographic zone. Clustering demand into 100 to 200 zones is common practice for warehouse location in commercial GIS (Ballou, 1994). Therefore, four levels of demand aggregation (50, 100, 200, and 400) were created

Three levels of demand dispersion were used in the experiment (low variance, medium variance, high variance). The demand dispersion based on census population figures served as the medium demand variance treatment. Initially, product demands using the population-based dispersion were determined for each of the 400 zones. The square root values of the 400 zone demands were computed and used to create the low demand variance problems. Similarly, the 400 zone population values were squared in order to produce the high variance demand problems. The demand values were summed and zones aggregated to create the 200, 100, and 50 zone problems for each of the low, medium, and high variance demand dispersions respectively.

Dependent Variables

Two measures of decision performance were evaluated in the experiment: decision quality and decision time. Projected total system operating cost is a commonly used decision performance measure when evaluating facility network design decisions. We assessed the quality of each subject's best solution by computing the percentage deviation of the user-generated solution's operating cost from the optimal solution's operating cost (DEV).

We measured decision time as the total time in minutes required to generate and evaluate candidate solutions and to select the best solution from the candidates.

Results

A three-way ANCOVA was executed for each of the dependent variables using spatial orientation, problem order, and effort as covariates. A substantial number of subjects generated optimal solutions on small problems producing a skewed, non-normal distribution of the DEV variable. To reduce the non-normality of the treatment distributions, we used the square-root values of the DEV variable in conducting statistical tests (Scheffe, 1959).

Decision Quality

For decision quality, main effects for data dispersion and problem size were both significant ($p < .001$ in both cases), supporting H1a and H3a. Decision quality for smaller problems was superior to that produced for larger problems. For dispersion, decision quality for highly varied demands was superior to decision quality for the other problems. There was no significant main effect related to the level of data aggregation, providing no support for H2a.

Decision Time

For decision time, the problem size and data aggregation main effects were significant ($p < .001$ for both effects). There was no significant effect for demand dispersion. These results support H1b and H2b, but not H3b.

Discussion

A summary of the findings as they relate to each of our hypotheses is presented in Table 1. The results support previous research in suggesting that increased problem size detrimentally affects decision making. Our data show degradations of 40 to 50 percent in decision quality and decision time when the problem size is doubled. The size effect was the most consistent of any we studied. Moreover, problem size appears to work in combination with data aggregation and data dispersion to increase the time required for decision makers to arrive at a satisfactory decision.

Data disaggregation appears to help decision makers who are solving geographic problems containing patterns of highly varied geographic data. Increased detail has little effect on decision quality for problems with less varied demand dispersions. This finding is consistent with the notion that patterns in geographic data are sharpened as additional detail is provided. We posit that decision makers are able to more effectively perceive and make use of these stronger, more easily identified patterns.

Additionally, decision makers are evidently able to handle increased information load from increasing levels of detail without significant impact on decision quality. Our problems did not appear to reach a threshold of information overload, assuming such a threshold exists. However, additional complexity due to information load is indicated by the additional time required by decision makers to generate good decisions, especially when solving large problems.

As described previously, the problems we included in the experiment span normal levels of aggregation for many geographic-based decisions. Our findings suggest that this common clustering of 100 to 200 zones is an appropriate level of aggregation. This level provides benefits to decision quality (for high demand dispersions) without significant detriment to decision time.

Previous research suggests that recognizable patterns or dominant flows in geographic data may enhance the quality of decision making (Trybus and Hopkins, 1980; Taylor and Iwanek, 1980). To the extent that increased variation in data dispersion increases the strength of geographic patterns, our study supports such a theory. In addition, it appears that sufficient levels of disaggregation are necessary for dominant patterns to emerge resulting in improved decision quality.

Research of visually-based solution methods for facility layout decisions suggests that moderate data variance is the most difficult for decision makers to process. Our results imply that decision makers need more time to process moderate demand variance in comparison to low demand variance. We take special interest in these findings given that our operationalization of moderate demand variance was based upon actual population values in the U.S.--a level of variance frequently found in practice.

Conclusions

This research builds upon and extends previous studies of map-based information presentation and task complexity. However, it is unique in its examination of factors related to the design and use of GIS. Therefore, examining factors that influence the more effective design and use of GIS is valuable to practitioners as well as to the academic community. The results demonstrate that the effects on experienced complexity are significant and should be considered in GIS design.

Table 1. Summary of Support for Hypotheses

Independent Variable	Hypothesis	Support	p-value
Increased Problem size	H1a: lower decision quality	Yes	.001
(n=480)	H1b: greater decision time	Yes	.001
Increased Data Disaggregation	H2a: greater decision quality	No	.424
(n=480)	H2b: greater time	Yes	.001
Increased Data Variation	H3a: superior decision quality	Yes	.001
(n=480)	H3b: less required decision time	No	.431

References

References are available upon request (cspeier@pilot.msu.edu; mswink@pilot.msu.edu).