Abstract

With the production release of SAS/GIS™ Software, organizations will be acquiring spatial and attribute data that is best suited for their particular industry segment and application needs.

Spatial data is available in a variety of coverages including streets, zip codes, address ranges, census tracts, and block groups. Demographic data is one popular form of attribute data and is readily available for various industry segments such as health care, insurance, retail, and banking.

This paper examines the spatial and attribute data that is available for use with SAS/GIS™ Software from Geographic Data Technologies and Claritas Inc. It describes how both types of data can be used to create a Business Geographic system with industry specific examples.

Introduction

As with any emerging technology in the computer industry, the recent movements in the Business Visualization and Multimedia market are now surfacing within the SAS System. Businesses are working with larger amounts of data than traditional graphics procedures and output can handle. Often, business professionals have difficulty understanding the meaning behind large amounts of numbers. Translating that data into visual information makes it more understandable.

That’s precisely the idea behind Business Visualization and Multimedia: to enhance the way data is presented to make it more meaningful. Enhancements can include video clips, images, maps, volume visualizations, and even virtual reality environments. These new tools are soon to be fully available in 1995 from the Institute.

The emerging Business Visualization and Multimedia technology makes pictures worth a thousand numbers. In other words, pictures are the numbers of tomorrow. Hardware advances have also helped to bring the technology to the forefront of applications in recent times. The increase in workstation processing speed and power and the parallel decrease in hardware costs will make Business Visualization and Multimedia tools increasingly more accessible and justifiable. In addition, PC platforms are now being shipped with CD drives as standard equipment which will enable even larger data distribution to the masses. Moreover, the new operating systems soon to be available on PC’s, such as WIN95, will add support for 3-D graphics libraries previously found only on more expensive workstation platforms.

As the need for information has expanded, the ways in which that information can be delivered have multiplied. But as Business Visualization and Multimedia technology begin to take off, the fundamental principles have remained the same: to provide users from many industries with efficient, easy-to-use tools for viewing and analyzing their data from different perspectives. One of the key perspectives will most certainly be Business Geographic Information Systems.

What is a Geographic Information System?

A Geographic Information System (GIS) is a tool for organizing and analyzing data that can be referenced spatially. Many types of data have a spatial aspect—they can be tied to a physical location. A GIS helps you analyze your data in its spatial context.

GIS systems are based on spatial data management, a horizontal technology with broad applications in many different market segments. Each has its own unique set of requirements for data, software, and distribution. GIS has enjoyed a number of years of relatively stable growth within well-defined market segments like government and environmental applications. The GIS software market is now in the midst of a dramatic transition: it is moving away from being a technology that existed almost exclusively in government and utility applications towards a mainstream business application.

What is Business Geographics?

GIS have traditionally been the domain of technical users. But the focus is now shifting toward the use of a GIS as a core of a company’s overall information delivery system. The massive amounts of data that a GIS can handle allows a user to layer different types of information, graphics, and text.

‘Business Geographics’ is an implementation of GIS technology that allows users to be smarter in their use of geographics. Better access to maps and locational data is beginning to affect companies. Early adopters have found that business geographics improves existing business processes and makes new ones possible. Currently there is tremendous growth in private companies’ acceptance of business geographics. A successful business GIS will perform two critical functions. First, it allows synthesis and analysis of geographically related data. Second, it allows the information developed using GIS analysis techniques to be displayed in a geographic context.

Types of Data

There are two broad types of data associated with a SAS/GIS Software application: spatial data and attribute data. Spatial data contain the coordinates for the features in a map. In SAS/GIS Software, there are three types of features: points, lines, and polygons. Attribute data are linked to the spatial data and contain additional information related to features in the layers. For example, attribute data for a map could be the price and address for a house or the population data for a census tract. Linking attribute data and spatial data in a visual manner is a key concept of Business geographic information systems and can provide a unique look at your data.

In the not too distant past, the cost of the spatial data kept GIS technology from being widely implemented in mainstream business applications. It is estimated that just a few years ago, over 80% of the cost of a GIS was the spatial data itself. Now, some of the spatial data is available for minimal cost and is distributed on a variety of media formats including CD ROM. One of the key differences between traditional GIS applications and Business Geographics is related to data. The geography that fuels business geographics applications has already been digitized by others. Street maps, highways, ZIP+4 codes, census tracts, cities, counties, states, countries, etc., are available in digital form. Spatial coordinates of stores and shopping centers as well as a wealth of demographic data are readily available. Business geographics users primarily buy data, not digitize it.
Where Does Spatial Data Come From?

SAS/GIS Software can directly import seven different kinds of spatial data. TIGER (Topologically Integrated Geographic Encoding and Reference) files are composed of digital map data for 1990 census geographic areas with basic map features (streets, rivers, railroads, etc). Also, for 345 core metropolitan areas, the names, address ranges and ZIP codes are included. The Census Bureau has made this public domain data available for a nominal fee of $250 per CD ROM. DLG (Digital Line Graphs) files are available from the U.S. Geological Survey. DXF (Drawing Interchange) files are from most CAD programs, like AutoCAD® software (AutoCAD is a trademark of AutoDesk, Inc.). Polygons can be imported from SAS/GRAPH® map data files which (a required product for SAS/GIS Software), includes maps covering most of the world. ARC/INFO® files can be imported that have been exported to an uncompressed ARC/INFO transport file (ARC/INFO is a trademark of ESRI Corp.).

The first step in a SAS/GIS Software application is to determine what spatial data to use and to make decisions about what level of geography is appropriate. For example, is the county level enough, or do you need to display street level maps? This concept of “rightsizing” the level of geography is a second major difference between traditional GIS applications and Business Geographics. Many traditional GIS users have been involved in some form of land management. A city, for example, must manage information about roads, parks, and sometimes utilities. Business geographics users do not tend to be land managers. The land is a backdrop for the data that interests them. Roads, for example, are a backdrop on which to place customers. ZIP codes are a backdrop for sales data. Census tracts are a backdrop for demographics. The bottom line is that business geographic applications are tightly focused on specific business needs.

Since SAS/GIS Software is integrated with the SAS System, you can take advantage of the client/server tools to store and distribute large amounts of spatial data. One possible scenario could be to use a server to store a spatial coverage of a nationwide zip code file and distribute state zip code coverages to individual workstations.

Where Does Attribute Data Come From?

The second step in building a SAS/GIS Software application is to identify the attribute data sets. An attribute data set contains information about features on the map that is linked to the map. The linkage is achieved by specifying variables in the attribute data set and composite associations in the spatial definition that have the same values. The three main uses of associated attribute data sets are layer themes, actions, and where selections. For example a Spatial Definition might have a composite association named TRACT that defines the census tract boundaries. A data set with tract level demographics could then be linked to the map. This would allow you to create a theme by which to color the map. A theme could show the number of households earning greater than $75,000 annually, or number of government workers in each tract.

There are two broad categories of attribute data: in-house and out-of-house. In-house attribute data can be any SAS data set or data view. By taking advantage of middleware technology within the SAS System, attribute data associated with the map is available to the user in virtually any format on any platform such as DB2, Oracle, Ingress, Sybase, DBF, DIF, and WK1 files. Out-of-house data can be purchased from independent data vendors such as Claritas.

Claritas is a leading vendor of demographic data. Third party demographic information can be used to augment internal corporate information to precisely define, analyze and target consumer markets. Claritas offers special databases for segmentation/lifestyle, business and retail, consumer expenditure, healthcare, media, bank/financial, and data for special areas like crime, the environment, and high schools. Claritas is best known for the PRIZM lifestyle segmentation system which was introduced in the early 1970s. PRIZM defines every micro-neighborhood in the U.S. in terms of demographically and behaviorally distinct types, or “clusters.” PRIZM offers an easy way to identify, understand and target consumers. The third generation of PRIZM contains 62 clusters, an increase of more than 50% over the previous system.
The Key Link: How attribute data is tied to spatial data

SAS/GIS Software offers two methods for linking your attribute data to the map: Action Definitions and GEOCODING.

**Action Definitions**

As one of the most powerful capabilities offered in SAS/GIS Software, Action Definitions provide a third dimension to the attribute data linked to the map. Immediate or deferred, an action is a process performed on the attribute data corresponding to the currently selected features. For example, by selecting a point on a map that represents a house, a user may also see an image of the house as well as information on square footage, price, age, etc. The user saw the relationship between the house and school zones, major thoroughfares, shopping areas, and recreational facilities (visually on the map). With SAS/GIS actions, however, choosing the point gave the user unique access to the attribute data.

**Display 2** Example of an Image and Browse Action

There are seven action types available in SAS/GIS Software.

**SAS/GIS Action Types:**

1. **Browse Action** - displays linked attribute data in a FSBROWSE window
2. **View Action** - displays linked attribute data in a FSVIEW window
3. **Image Action** - displays an image associated with a feature
4. **Data Action** - creates a subset of the linked attribute data
5. **Program Action** - executes a user-written SAS program on a subset of the linked attribute data
6. **Drill Down Action** - loads a new map based on the selection of map features
7. **Spatial Info Action** - displays the spatial data for the currently selected map feature

Actions are defined in the Action Definition Window. This window shows the currently defined actions and lets you edit any action's parameters or create new actions.

**Display 3** Defining a Browse Action

The Data Link field, common to all actions, is used to define how the attribute data is associated with the map. In edit mode, choosing the right arrow displays the GIS Attribute Data Sets Window. This window shows all currently defined links and allows you to create new links.

**Display 4** Edit mode for link definition

The first step in linking attribute data to a map is to name the link and provide the attribute data set name. The Spatial Vars listbox is already initialized with the spatial data base variables. Once the attribute data set name is provided, the Dataset Vars listbox is also filled in. The link is created by choosing one variable from each listbox that contains an identical mapping attribute. In this example, the variable feaname, found in both the attribute data and spatial data base, contains the address of the house. When the point from the map (100 Oak Street) is selected, the corresponding value from the attribute data set is found and the related data becomes available. Other common features, such as census tracts or blocks, zip codes, or a variety of municipal districts, can be used to link spatial and attribute data.

**GEOCODING**

Geocoding is a process that adds map coordinates to address data sets. By matching variables found in both the spatial and address data bases, lat/long or (x,y) coordinates are found and written to the corresponding address data set observation.

The Geocoding Window sets the spatial and address data sets and the address variable names used in the matching process. The Values to Add Listbox lists additional spatial variables that can be added to the address data set. When all the fields are filled in, choosing the GEOCODE push button starts the process.

822
The house point layer, described in the previous example, was added to the map using this technique. A real estate data base contained the addresses for listings in the map area. Geocoding was used to find the mapping coordinates for each address and add them to the real estate data base. These mapping coordinates were then used to add the house locations as a point layer.

Some address information may not match the spatial data exactly. Your spatial data base from 1990, for example, may not have a new street name or a matching address range. In these cases, SAS/GIS searches for as close a match as possible and ranks the results in an automatic variable called _score_. The higher the score, the more exact the match.

If an address cannot be found, but a correct zip code is provided, SAS/GIS Geocoding can assign a map coordinate based on the zip code centroid. This centroid is taken from a data base provided in the SASHELP library. This coordinate assignment is helpful if your goal is to see the number of houses (or customers) by zip code, rather than the exact location.

Once the spatial and attribute data is linked together with techniques like geocoding, a business geographics environment has been created. The next and final step is to determine what actions to define and to build vertical applications that reflect your business need. One key advantage of SAS/GIS Software is that you have the flexibility to use the entire SAS System to process the attribute data. In other words, you can use all of the report writing, statistical analysis, visualization and multimedia tools to customize the output associated with your map.

Vertical industry areas in which SAS/GIS Software is particularly effective as a business geographics solution include banking, insurance, health care, pharmaceutical, and retail. In addition, any business site with a Master Customer Information File (MCIF) will find the SAS/GIS Software solution to be helpful in visualizing their customer base.

Business geographics examples include competitive and demographic analysis, distribution analysis, prospect analysis, response tracking, site analysis, target marketing, and territory allocation. An insurance company could use the SAS/GIS Software solution for risk analysis, underwriting, claims processing, catastrophe management, and customer service. Healthcare examples include disease clustering, patient analysis, and outcomes analysis.

Site selection is a common example of how Business geographics can be used. Imagine that you are the owner of a grocery store in Mecklenburg county, NC and you want to expand your business. You have collected demographic data from a survey to current customers and you have acquired additional demographic data from Claritas to reflect the demographics of other parts of the area. Based on statistical analysis, you have determined that household income of greater than $50K is the best predictor of grocery sales in other areas. SAS/GIS Software allows you to easily visualize this information and quickly pinpoint other geographic areas that fit the desired demographic make-up.

One of the most useful and visual functions of SAS/GIS Software is the ability to create thematic layers. The value of a response variable or theme variable in a data set is used to determine the coloring for the layer. For our example, the concentration of families that earn more than $50K would serve as the theme variable for the map layer.
Wake) that surface as counties with large numbers of high income families. A browse action could also be set up to show the specific information for a county when the county is selected by the user. The Browse action shown below details the demographic data provided by Claritas for Wake County with an accompanying legend.

Another action that can be employed when creating vertical business geographics applications is a drill action that allows the user to load a new map. In our example, after determining that Wake County is a good demographic match, we could then drill down to Wake county from the map of North Carolina. As shown in the figure below, we have also added point layers to the map. The two point layers represent existing grocery stores (our competition) and real estate for sale. As with the map of NC, we create a thematic map at the county level to reflect the value of household income > $50K.

Based on the thematic colors, it’s easy to determine that there is one census tract that has a high concentration of households earning > $50K with little existing competition and available real estate. Using the zoom tool from the tool palette, you could then rubberband that area of the map which turns on the street layer as shown below.

As a final step, you could apply the same statistical model that predicted that household income was the best predictor of grocery sales to the attribute data for the selected tracts. By doing so, you get an answer to the question of how much revenue would be generated in a particular tract if a new store was built there.

Summary

In this paper, several of the issues surrounding the building of SAS/GIS Software applications are presented. A site selection example is used to illustrate how a Business Geographics solution offers an intuitive environment for visualization.

In this example, the steps necessary to build a SAS/GIS Software application from scratch have not been presented. The application presented does not include an example of the entire toolset available within SAS/GIS Software for importing data, customizing maps, creating actions, selecting features and performing tasks. When building applications with SAS/GIS Software there could potentially be several different levels of interaction with the software.

One level of interaction would be the person who is charged with the system administrator role. This person would perform the importing of the spatial data. They would also determine where the desired attribute data is located and surface that data to the SAS/GIS environment. This role might also include tasks for map customization and setting up the actions via the action definition window.

A second level of interaction would be the person or group of people who decide questions about the application itself. For example, decisions about what level of detail the maps need to be (state, county, tract, block, or zip?), as well as determine what
are the actions to be defined, i.e., should there be a tabular report or a bar chart? This person or group of people needs to also determine what attribute data the actions will use.

A third level of interaction with SAS/GIS Software is the end user. This person or group of people only need to know two things. The first is how to use the selection tools in the map window to choose the map features of interest. The second is to know which actions do what, i.e., know that the Report action produces a tabular report or that the Graph action produces a bar chart (these are the actions that were produced at the second level).

Of course these levels of interaction could all be the same individual or group of individuals. These three levels serve to point out some of the internal issues involved in constructing and supporting a SAS/GIS Software application environment.

**Why Choose SAS/GIS Software?**

There are two key differences between SAS/GIS Software and other GIS products available today: 1) the ability to access attribute data in virtually any format on any platform and 2) the ability to bring the power of the entire SAS System to bear on the management, analysis, and presentation of attribute data.

First, with SAS/GIS Software you can surface attribute data with the toolset for accessing data and distributing data found within the SAS System. Not only are existing SAS data sets and SAS Data views available to you, the ability to tie into SAS/ACCESS®, SAS/SHARE®, and SAS/CONNECT® mean that any data on virtually any platform is potential attribute data for a SAS/GIS Software application.

Secondly, it should be noted that one of the most powerful aspects of SAS/GIS Software software is the Program action type. One of the features that makes SAS/GIS Software a unique and powerful solution for Business Geographics solutions is the tremendous flexibility a user has to manage, analyze, and present the attribute data with the SAS System toolset. A site can choose from a wide variety of reporting tools like Proc Print, Proc Tabulate, and Proc Report. The SAS System analysis toolset is very robust and includes state of the art statistical and decision support procedures. SAS/GIS Software can also take advantage of the other tools for Business Visualization and Multimedia delivery within the SAS System to present attribute data graphically such as images, volume visualizations, 3-D bar and pie charts, and video. Existing SAS applications could be easily integrated in the SAS/GIS environment with the Program action type. In addition, a program generator like SAS/ASSIST® software could be used to produce source code that can be used with the Program action type.