C H A P T E R

1

Introducing SAS/GIS Software

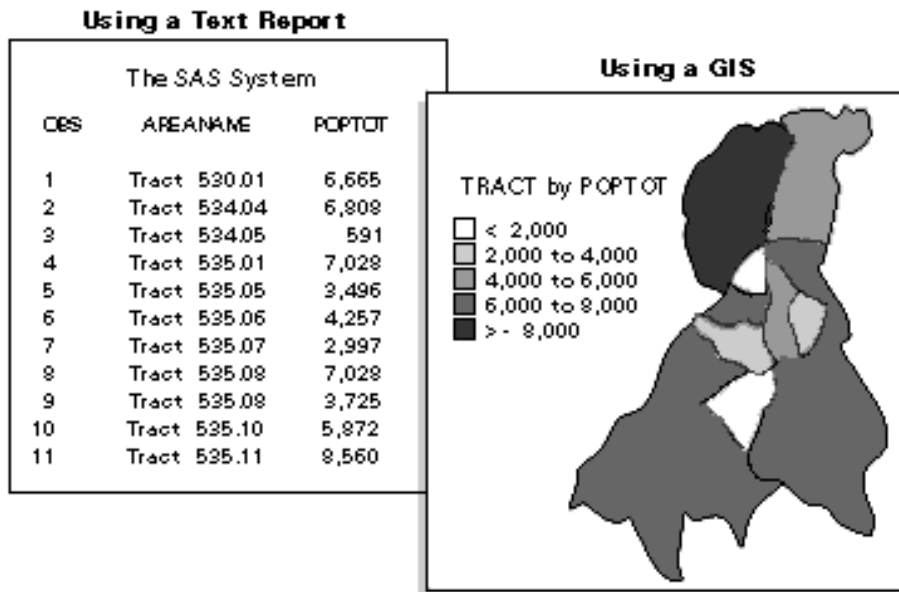## Introduction to Geographic Information Systems

SAS/GIS software provides an interactive geographic information system within the SAS System. A *geographic information system* (GIS) is a tool for organizing and analyzing data that can be referenced spatially; that is, it is data that can be tied to physical locations. Many types of data have a spatial aspect, including demographics, marketing surveys, customer addresses, and epidemiological studies. A GIS helps you analyze your data in the context of their location.

For example, if you need to evaluate population data for census tracts, you could view the information in tabular format. However, consider how much easier and more effective it would be to view the demographic information in the context of the geography of the tracts as in Figure 1.1 on page 4. When viewing information that has a spatial component, you may find it easier to recognize relationships and trends in your data if you view the information in a spatial context.

Figure 1.1   Evaluating Spatially Referenced Data

Using a Text Report

OBS	AREANAME	POPTOT
	The SAS System	
1	Tract 530.01	6,665
2	Tract 534.04	6,808
3	Tract 534.05	591
4	Tract 535.01	7,028
5	Tract 535.05	3,496
6	Tract 535.06	4,257
7	Tract 535.07	2,997
8	Tract 535.08	7,028
9	Tract 535.08	3,725
10	Tract 535.10	5,872
11	Tract 535.11	8,560

Using a GIS

TRACT by POPTOT

☐ < 2,000
☐ 2,000 to 4,000
☐ 4,000 to 6,000
☐ 6,000 to 8,000
■ > - 8,000

SAS/GIS software enables you to do more than simply view your data in their spatial context. It also enables you to interact with the data by selecting features and performing actions that are based on those selections. SAS/GIS software draws on capabilities of the SAS System and enables you to access, manage, analyze, and present your data easily.

Features of the SAS System

The SAS System provides a powerful programming language and components called procedures that allow you to perform many different kinds of analysis and data management functions, as well as produce many different types of text-based and graphical presentation output. Combined with other features of the SAS System, the SAS language and its procedures make possible an unlimited variety of applications from general-purpose data processing to highly specialized analyses in diverse application areas, including the following:

☐ Access raw data files and data in external databases such as DB2, SQL/DS, Rdb/VMS, OS/2 Database Manager, Lotus 1–2–3, dBASE, and SYSTEM 2000.

☐ Manage your data using tools for data entry, editing, retrieval, formatting, and conversion.

☐ Analyze your data using descriptive statistics, multivariate techniques, forecasting and modeling, and linear programming.

☐ Present your data using reports and business and statistical graphics.

The SAS System is portable across computing environments. Your SAS applications function the same and produce the same results regardless of the operating environment on which you are running SAS to process your data. However, some features, such as the interactive GIS window, are not supported on all platforms.

Data in SAS/GIS Applications

SAS/GIS software uses two basic types of data:

Spatial Data
> Contain the coordinates and identifying information describing the map features such as streets, rivers, and railroads.

Attribute Data
> Are the information that you want to use for analysis or presentation. This information must be spatial in nature. Sales figures for each of your store locations, population data for each county, and total income for each household in a region are examples of information that are spatial in nature because the information applies to a specific geographic feature.

For example, the U.S. Census Bureau distributes both types of data:

TIGER Line files
> Contain spatial information that you can use to build maps.

Summary Tape files
> Contain population and other demographic information that you can link to the map features.

Attribute data provide the information that you want to analyze, and spatial data provide the context in which you want to analyze it. For example, consider the SAS/GIS map shown in Display 1.1 on page 5. Spatial data provide the boundaries for the map areas, and attribute data provide the population information that is used to color the map areas.

Display 1.1 Spatial and Attribute Data in SAS/GIS Maps
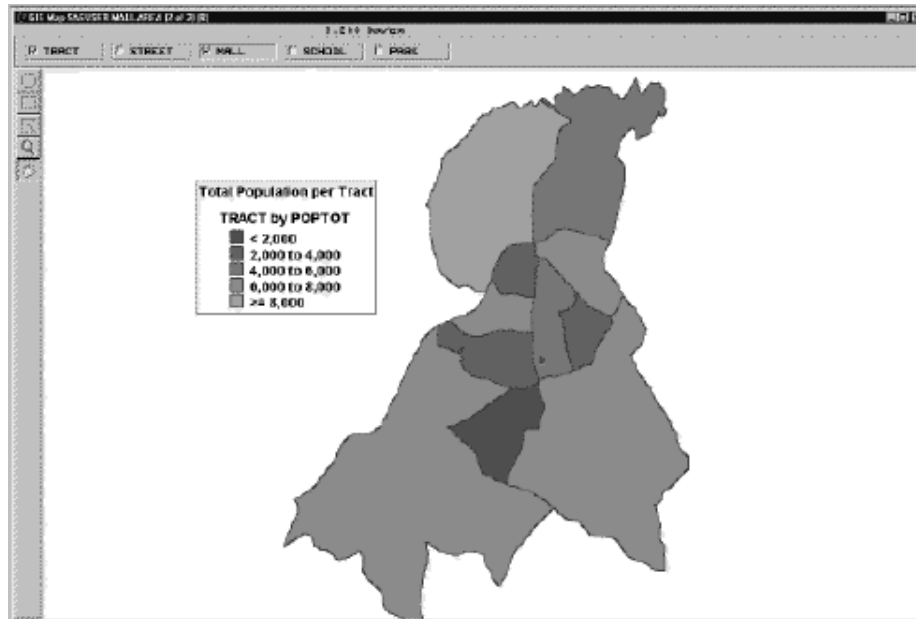

Spatial Data

Spatial data contain the coordinates and identifying information that is necessary to draw maps. For SAS/GIS software, spatial data are stored in SAS/GIS spatial

databases, which consist of collections of SAS data sets and SAS catalog entries. The primary method for creating a SAS/GIS spatial database is through the SAS/GIS Import facility, either in batch or in interactive mode. You can also use the GIS procedure to create, modify, and manage the catalog entries in a spatial database.

Features in the spatial data are organized into layers. A *layer* is a collection of all the features in the map that share some common characteristic. The various physical aspects of the map—political boundaries, roads, railroads, waterways, and so forth—are assigned to layers according to their common spatial data values. Some features can appear in multiple layers. For example, a street can also be a ZIP code boundary and a city boundary line. The street could appear in three layers: one containing the streets, one containing the ZIP code boundaries, and one containing the city boundaries.

Three types of layers can be represented in SAS/GIS maps: points, lines, and areas. For example, the collection of all the points in the map that represent park locations can be organized into a point layer for parks, the collection of all the lines in the map that represent streets can be organized into a line layer for streets, and the collection of all the areas that represents census tracts can be organized into an area layer for tracts. When the various layers are overlaid, they form a map, as shown in Figure 1.2 on page 7.

Figure 1.2 Layers Forming a SAS/GIS Map
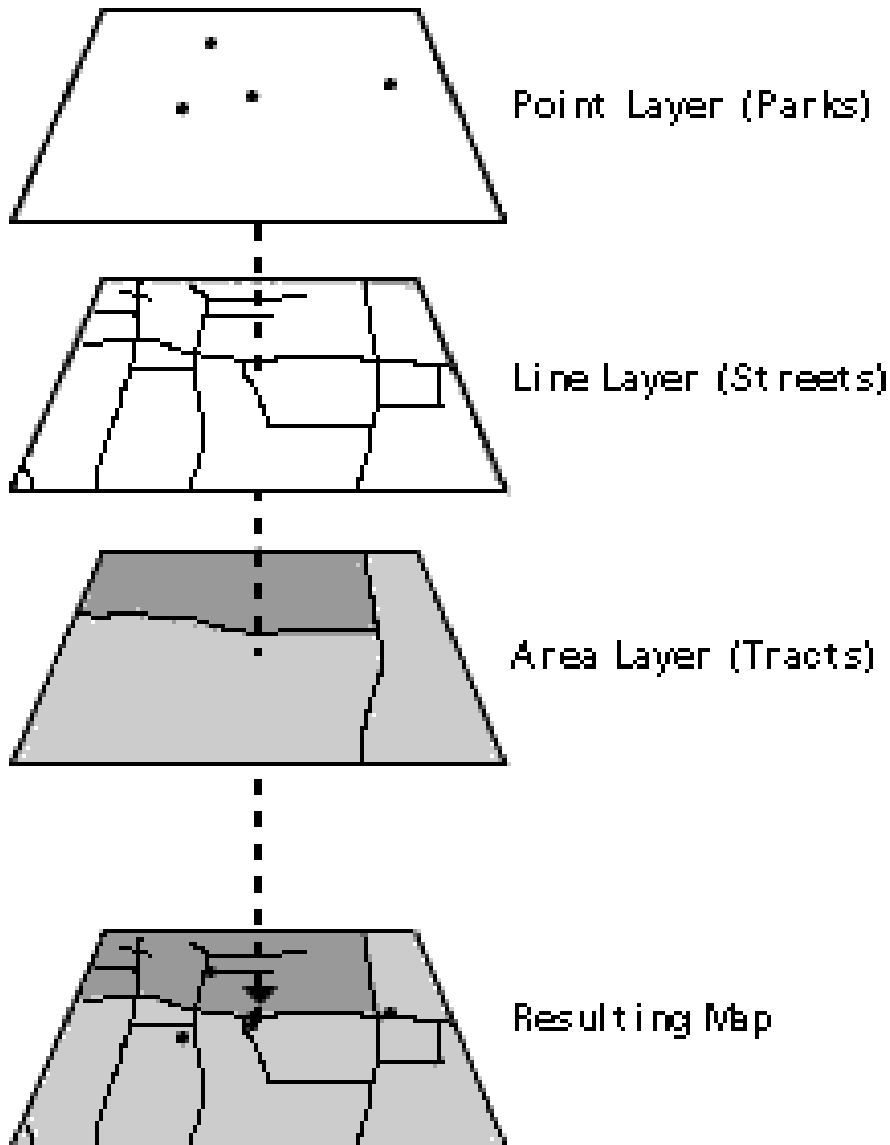
A layer can be displayed as either static or thematic. When a layer is displayed as static, it uses the same graphical characteristics (color, line, width, and so forth) for all features in that layer. For example, a street layer could use the same color and line style to display all the streets. When a layer is displayed as thematic, it uses different graphical characteristics to classify the features in that layer. For example, a theme representing sales regions could use different colors to show the quarterly sales performance of each region. A theme in a layer representing highways could use different line widths to show the classes of roads. A layer can have multiple themes stored in it, and you can easily change which theme is currently displayed.

In SAS/GIS software, maps display only the portion of the spatial data that falls within a given coverage. A *coverage* defines a subset of the spatial data that is available to a map. The coverage can include all the spatial data in the database, or only selected portions. For example, a spatial database may contain geographic data for an entire country, but a coverage may restrict the portion that is available for a given map to only

one region. You can define more than one coverage for each spatial database, although a map uses only one coverage at a time.

Most operations in SAS/GIS software use composites of spatial data variables rather than the actual spatial data variables themselves. *Composites* identify the relationships and purpose of the variables in the spatial data.

For example, if the spatial data have variables STATEL and STATER that contain the state ID codes for the left and right sides of each feature, then the spatial database could define a composite named STATE that identifies the relationship between these variables and specifies that they delineate state areas in the map. You would use the STATE composite, rather than the actual STATEL and STATER variables, to link state areas in the map to attribute data for the corresponding state.

See Appendix 1, "Details of SAS/GIS Spatial Databases," on page 135 for more information about the structure of SAS/GIS spatial databases.

Attribute Data

The second type of data that are used in a GIS is attribute data. In SAS/GIS software, your attribute data must be stored in either a SAS data set or a SAS view. SAS views allow you to transparently access data in other formats. For example, you can create a SAS/ACCESS view to access data in a database such as DB2 using the SAS/ACCESS to DB2 software. A DATA step view or an SQL view also allows you to access an external file, or any other type of data for which you can create a SAS view. Once your data are accessible either as a SAS data set or through a SAS view, they can be linked to your spatial data for use in labeling, analysis through an action, or theming. For instance, your spatial data might represent a county and contain information for city boundaries, census tract boundaries, streets, and so forth. An attribute data set with population information for each census tract can be linked to a map using the corresponding tract composite in the spatial data.

Some of the ways in which you can use attribute data in SAS/GIS software include the following:

□ Using values in your attribute data as labels. For example, you could use attribute data containing population data to provide the text of labels for each of the census tracts.

□ Using the values in your attribute data as themes for layers. For example, you could use attribute data containing average household income data as a theme for the census tract layer.

   See Chapter 5, "Customizing Maps," in *SAS/GIS Software: Usage and Reference, Version 6* for more information about assigning themes to map layers.

□ Defining actions that display or manipulate the attribute data when features are selected in the map. This way, you can explore your attribute data interactively rather than simply view static results. The actions can range from simple, such as displaying observations from an attribute data set that relate to features in the map, to complex, such as submitting procedures from SAS/STAT software to perform statistical analyses.

Actions can be defined to do the following:

□ Display observations from the attribute data sets that relate to the selected map features.

□ Open additional maps that relate to selected map features.

□ Display images that relate to the selected map features.

□ Interactively subset the attribute data sets according to the subset of selected map features.

☐ Submit SAS programs.

☐ Issue SAS commands.

☐ Issue host commands.

☐ Display and edit information for the selected map features.

☐ Organize area features into groups that are based on the attribute data.

See Chapter 4, "Performing Actions for Selected Map Features" in *SAS/GIS Software: Usage and Reference, Version 6* for more information on defining and performing actions.

Designing a SAS/GIS Spatial Database

One of the first steps in a SAS/GIS project is determining the design of your SAS/GIS spatial database. The database should include all of the spatial data that the user wants to see and all of the associated attribute data that the user needs to use for analysis or presentation purposes.

Although your first tendency with a new software product may be to begin using it immediately after you install it, take some time to draw up an overview of the system goals and data requirements that you will need for your database before you begin creating it. Do not rush into building your database. The time you spend designing your database initially will save you time and expenses later in the project. Remember, a well-designed database is easier to maintain and document, and you can extend it for future GIS projects.

Use the following guidelines when determining the information you want to include in a database:

☐ Identify the initial objective of the project and its ultimate goal. Consider any requirements that may have been imposed on it. Determine their feasibility for initial implementation and, as best as possible, the impact of any future demands upon them.

☐ Identify the attribute data that are necessary to illustrate the project objectives. Determine if you have these data or can obtain them.

☐ Identify the spatial features that you want on your map, for example, states, cities, rivers, roads, railroads, airports, and so forth.

Once you have determined a preliminary list of the data that you will need, use these additional factors to help evaluate and refine your list:

☐ To utilize attribute data for map actions, themes, or labeling, the attribute data set must contain the same identification information as the spatial feature that it describes so that you can link between them. For example, if one of the items on your attribute data list is the Sales Revenue for each store, along with the Store ID Number, you probably want to include the actual location in longitude and latitude for each Store ID Number on your spatial data list. You can then place a marker at the store location and also visualize and analyze the corresponding attribute data for each store.

☐ Do not use more detail than you need. If your store locations request the customer ZIP code at the cash register, don't assume that you need ZIP code boundaries on your map. ZIP code boundaries may be far too small for your purposes if you have stores nationwide. You may decide instead that the 3-digit ZIP code boundaries provide fewer, yet more appropriately sized, areas for your analysis. You can summarize your attribute data to the 3-digit ZIP code level and use it for your analysis, reducing both the amount of spatial data and attribute data that you need. As long as it is appropriate for your analysis, decreasing the amount of required spatial and attribute data reduces storage space and improves

performance. Reducing the level of detail in the spatial data also saves money if you have to purchase the data.

□ If you plan to summarize your attribute data to a matching level of your spatial data, make sure the two types of data have a common level that you can use. For example, ZIP code boundaries can cross not only county boundaries, but also state boundaries, so there is usually not a one-to-one correspondence between ZIP codes and states or counties. If the only information that ties your attribute data to your spatial data is ZIP codes, you will have difficulties using your ZIP code level attribute data if you include only state or county boundaries in your spatial data.

For specific, smaller areas of the country, a one-to-one correspondence may exist that will allow you to summarize your attribute data to a higher level. However, ZIP codes can change frequently, and this correspondence may be lost. Also, because ZIP codes change, you must be able to account for these changes when performing historical analyses. For example, if you are comparing sales in a specific ZIP code area over a ten-year period, make sure that the area remained constant during that period.

Using the SAS/GIS Interface

If you have no previous experience with SAS System window applications, some of the user interface features of SAS/GIS software may be unfamiliar to you.

Starting SAS/GIS Software

Use the following steps to start a SAS/GIS software session:

1 Open a SAS session.

2 From the SAS menu bar, select

| Solutions | ► | Analysis | ► | Geographic Information System |

Or Type **GIS** in the SAS Command Box or on any SAS command line.

Using the Mouse

SAS/GIS software is designed to use two mouse buttons.

□ The *mouse selection button* selects map features and activates various tools. To *drag* with the mouse, you hold down the mouse selection button while moving the mouse.

□ The *mouse menu button* activates pop-up menus. The pop-up menu that appears depends on the pointer's position when you press the mouse menu button.

By default, the left mouse button is the selection button and the right mouse button is the menu button, but most hosts permit you to change this (for example, to accommodate left-handed users).

Using Field Value Controls

In most places where you must supply a value in a SAS/GIS window, you will see either a pull-out control or a pull-down control, or both, presented in conjunction with the value entry fields, as shown in Display 1.2 on page 11.

Display 1.2 Typical Field Value Controls

Selecting a pull-down control displays a list of valid choices for the option. Display 1.3 on page 11 shows the list that is displayed by selecting the pull-down control for the `Style` field in Display 1.2 on page 11.

Display 1.3 List Displayed by the Style Pull-down Control

Selecting a pull-out control opens a new window in which you can interactively select appropriate values. Display 1.4 on page 11 shows the window opened by selecting the pull-out control for the `Color` field in Display 1.2 on page 11.

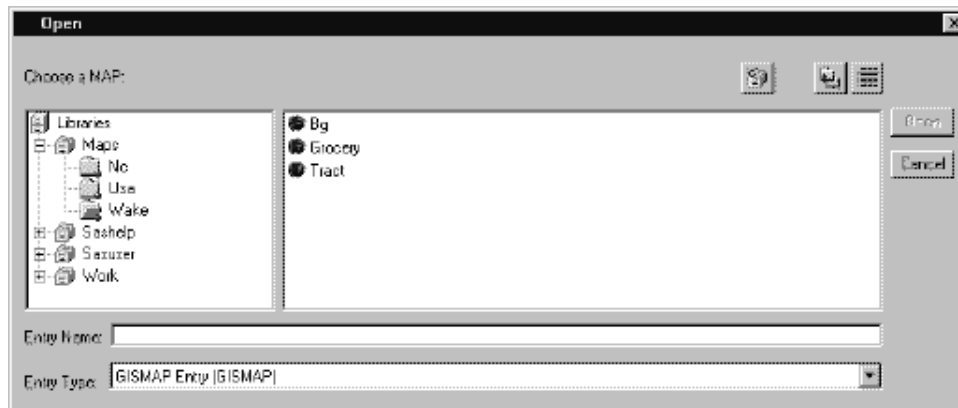**Display 1.4** Window Opened by the Color Pull-out Control

Selecting Maps and SAS Data Sets

Whenever you need to specify the name of a SAS data set or SAS catalog entry, SAS/GIS software opens a File Open window like the one shown in Display 1.5 on page 12.

Display 1.5   Typical File Open Window

The window provides an intuitive way to find the SAS data set or catalog entry that you need. It also makes it impossible to enter an invalid name, since only those choices that are appropriate for the operation you are performing are presented for selection.

To select each level of the SAS name from the tree view in the file requestor window, double-click the mouse selection button on your choice. Once you make a selection, the list of available choices for the next level of the name is displayed.

Getting Started with SAS/GIS Software

This book does not attempt to cover all of the fundamentals of using SAS/GIS software. For an introduction to the basic tasks that you can perform, see the online tutorial that is included in SAS/GIS software. To start the tutorial, make the following selections from the GIS Map window's menu bar:

Help   ▶   Getting Started with SAS/GIS Software   ▶   Begin Tutorial...

The tutorial creates sample maps and attribute data and leads you step by step through the following tasks:

□ Displaying maps.

□ Selecting the types of feedback that are provided about the displayed map.

□ Using the zoom tool to zoom in on selected areas of the map.

□ Using the pan tool to move the map within the window.

□ Modifying, adding, and removing layers in the map.

□ Using attribute data as a theme for a layer.

□ Adding legends that explain how features are represented.

□ Selecting features and using actions to explore the attribute data.

□ Saving changes to the map.

After you have used the online tutorial to become familiar with the basics of using SAS/GIS software, you can refer to *SAS/GIS Software: Usage and Reference, Version 6, First Edition* for additional information about using SAS/GIS software and for detailed reference information about the features of SAS/GIS software.