

Introduction to SAS/GIS™ Software

Jack Bulkley
SAS Institute Inc

Abstract

SAS/GIS™ Software is a new part of the SAS® System to be available as a Version 6.09 add-on. As more and more data becomes available to enterprises, new tools are needed to make sense of this massive amount of information. SAS/GIS Software is a tool for handling spatially related data. The data importing, querying, analysis, and display will be covered as well as the strong connection to the rest of the SAS System.

Introduction

SAS/GIS Software adds the ability to analyze, display and interact with geographic data to the SAS System. This type of powerful tool is needed to make sense of the massive amount of data becoming available to enterprises. Much data has a spatial aspect that tells where it is located. GIS is a tool to manage this spatial aspect to visualize potentially important relationships in the data. Since SAS/GIS Software is still under development, some features may change but this paper describes the basic functionality that SAS/GIS Software will provide. The most powerful aspect is its strong ties to the rest of the SAS system.

While SAS/GIS Software is new, GIS has been used for many years. Up till now GIS has been primarily used in areas where the data and geography were very closely tied. Enterprises charged with managing areas (whether they be forests, cities, utility networks, or the whole country) have used GIS as a tool. These organizations have been able to gain enough efficiency to justify the often high costs of creating the digital spatial data bases they need. However, some of this spatial data is now available for a minimal cost and this makes GIS a cost-effective solution for more types of enterprises than in the past.

There is not enough room here to completely describe all of the spatial data that is available, but here is brief description of some of the data available. The U.S. Census Bureau has developed TIGER data as part of the 1990 census. TIGER is digital map data for 1990 census geographic areas, basic map features (streets, rivers, railroads, etc). Also for 345 core metropolitan areas, the names, address, ranges and ZIP code are included. The Census Bureau has made this public domain data available for a nominal fee of \$250 per CD-ROM, where, for example, all of North Carolina is on one CD-ROM.

The Digital Chart of the World (DCW) was produced for the U.S. Defense Mapping Agency. The DCW is a comprehensive 1:1,000,000-scale vector basemap of the world. It consists of cartographic, attribute, and textual data stored on CD-ROMs. The primary source for the database is the U.S. Defense Mapping Agency Operational Navigation Chart series. This is the largest-scale unclassified map series in existence that provides consistent, continuous global coverage of essential basemap features.

The U.S. Geological Survey has various digital data in their DLG format. This planimetric data represents the cartographic line information usually portrayed on a map. DLG's are available for 1:20,000, 1:24,000, 1:48,000, 1:62,500, 1:63,360, 1:100,000, and 1:2,000,000 scale maps. A CD-ROM of the 1:2,000,000 scale data is available. DLG's are sold in five categories:

1. U.S. Public Land Survey System including township, range and section.
2. Boundaries including state, county, city, and other civil divisions.
3. Transportation including roads and trails, railroads, pipelines, and transmission lines.
4. Hydrography including streams and water bodies.
5. Hypsography including contours.

Geographic Data Technology, Inc. (GDT) provides Dynamap/2000™ data, a nationwide, up-to-date, topologically correct street center-line database. GDT also has other geographic data including ZIP code boundaries, census boundaries, and highways.

Basic Concepts

GIS is a tool for making sense of spatially related data. Thus understanding the data leads to understanding GIS. There are two basic types of data used in the GIS, the spatial data base and data related to the spatial data base. The first type, spatial data, contains the coordinates and identifying information for the various map features. In SAS/GIS Software, a feature is any point, line or polygon that can be visually represented on a map. For example, using TIGER data, SAS/GIS Software can create a spatial data base that contains information for city boundaries, census tract and block boundaries, streets, etc.

The second type of data is other data sets containing information about features on the map. This type of data is associated with the map by ID variables. For example, a data set with population for each census tract is connected to the spatial data base by the ID variable, TRACT. This type of data is already widely used in SAS for analysis, summary, reporting, and graphing. SAS/GIS Software moves beyond just drawing maps by allowing interaction with the data to explore relationships in the data.

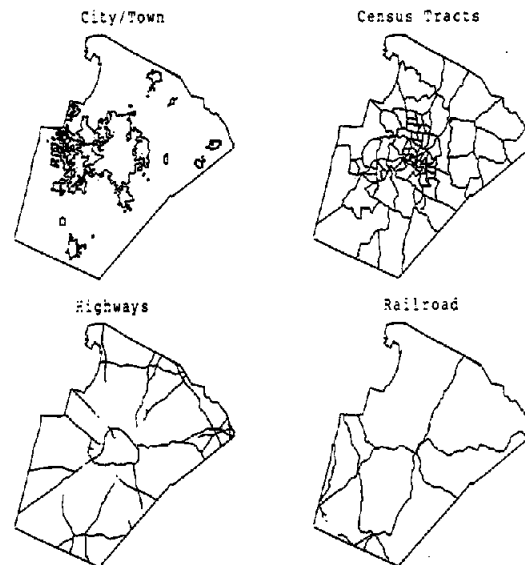
The basic operation of SAS/GIS Software is to select features of interest on the map and then do things with the data associated with those features. The things that you do are called actions and there are six types of actions, each described below. As well as the type, each action definition tells when the action is to be performed. An immediate action is performed as soon as the selected features on the map change. A deferred action will not be performed until requested by the user from the map window menus. Here is a brief description of each action type.

1. A **browse action** shows the observations in a data set that correspond to the selected features in the map in a FSBROWSE window task. A screen catalog entry can be specified to customize the appearance and performance of the FSBROWSE window.
2. A **view action** shows the observations in a data set that correspond to the selected features in the map in a FSVIEW window task. A formula catalog entry can be specified to customize the FSVIEW window.
3. A **spatial information action** shows the values from the spatial data base for the current feature.
4. A **data action** creates a data set that is a subset of a data set, choosing only the observations that correspond to the selected map features.
5. A **report action**, like a data action, creates a data set that is a subset of a data set, choosing only the observations that correspond to the selected map features, and then also submits SAS statements to run. This is the most powerful action because it allows direct access to the powerful analysis and report writing components of the SAS System.
6. An **image action** shows an image that corresponds to the current selected feature in the map. The name of the image file is stored in a variable in the associated data set.

The map definition tells everything about the map you are going to see and use, including the coverage, layers, actions and other map options. The coverage tells which spatial data base and which part of it to use. For example, a cover-

age might indicate to use a spatial data base of the Southeast U.S., but only read and show the data for Florida on the map.

A layer is a logical subset of features from the spatial data base. It might be all the roads, the county polygons, or points representing bank locations. The various layer definitions of a map define all of the features you can see on the map. If a feature is not part of one of the map's layers, it will not be on the map. The following pictures show some of the layers that can be defined from TIGER data for Wake County, North Carolina.



Implementing an Application Using SAS/GIS Software

SAS/GIS Software can be used for many different applications. The following is a brief description of a fairly simple application, using data available from the U.S. Census Bureau.

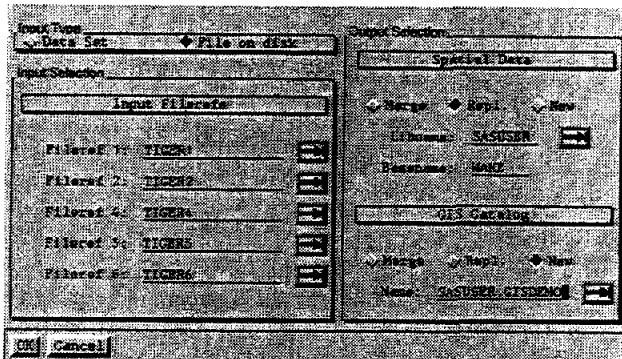
Define Application and Determine Data Needs

While describing an interactive application on paper makes it difficult to capture the true feel of the application, the following description should give a taste of what SAS/GIS Software can do. It is very straight forward to setup an application to examine Summary Tape File (STF) data with a spatial data base created from TIGER data. STF data comes in several varieties. The STF1A data contains 1050 variables from the 1990 census describing demographics down to the census block level. TIGER data contains the matching spatial information.

Import Data

Upon receiving the desired data, the interactive importing

screens are used to read in the TIGER spatial data and STF1A demographic data. The importing window takes the specified input files and creates data sets and catalog entries for the spatial data base, a coverage, basic layers, and a map. Very little information needs to be provided, only the location of the files to be imported and a place to store the spatial data base and the SAS/GIS catalog. The importing process will create enough layers, coverages, and maps to allow viewing the data without any further definitions.



Setting Up the Map

The basic map created during the importing allows a quick view of the spatial data base. With TIGER data this would include the following layers: COUNTY, HIGHWAY, CITY, and TRACT. A new map with just these layers can be created simply in the map window from the map generated by importing by:

1. Removing the unwanted layers,
2. Reordering the remaining layers,
3. Saving the map into a new catalog entry.

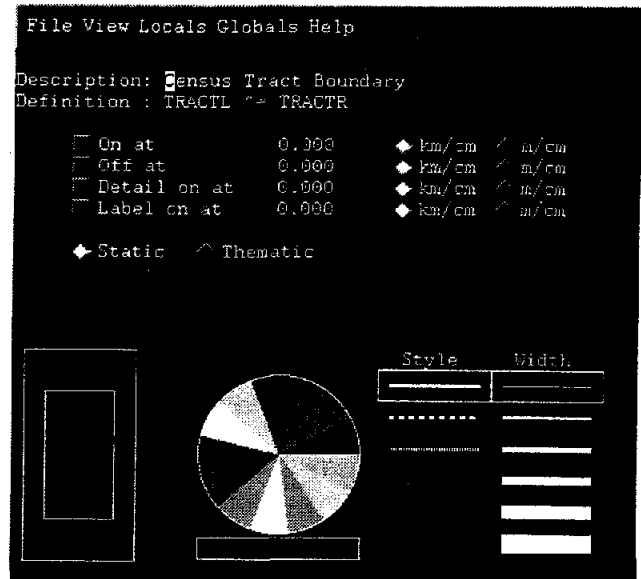
That would leave a map looking like the following.



The map window has a tool palette along the left side, and layer boxes across the top. The tools can be used for various selection methods as well as panning and zooming the map display. The layer buttons can be used to turn on and off the layers. A pop-up menu for each layer gives further control including making a layer selectable, adding labels, and allowing further editing of the layer definition in a separate window.

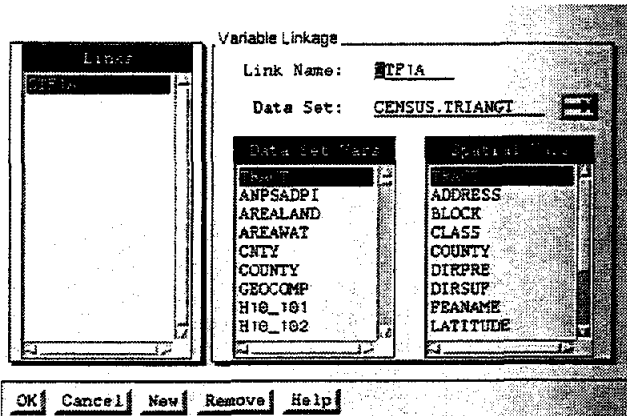
Changing a Layer

In the layer edit window there are controls to cause a layer to be automatically turned on or off as the scale of the map display changes. This allows a layer with much detail, such as all streets, to be off when viewing a whole county, but turned on when zoomed in to a neighborhood.

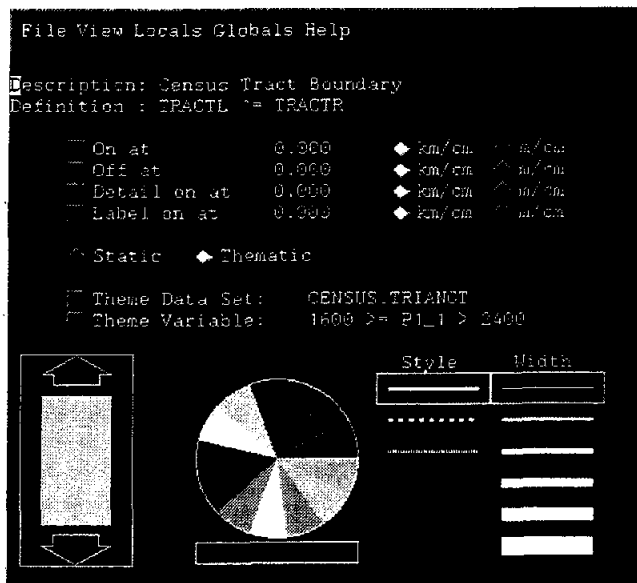


Layers can be static or thematic. In a static layer, all of the features are drawn in the same color and style. In a thematic layer, the value of a data set variable determines the color and style. To display a variable from the STF1A data, change the layer definition of TRACT to include a theme data set and variable.

Choosing *Thematic* in the layer window radio causes the data set window to come up. A data set can be associated with the map by entering the data set name and then selecting the corresponding variables from the data set and the spatial data base. In this case, the STF1A data set and the spatial data base both have a variable named TRACT which contains the census tract number. This association is stored as part of the map definition so this process does not have to be repeated.

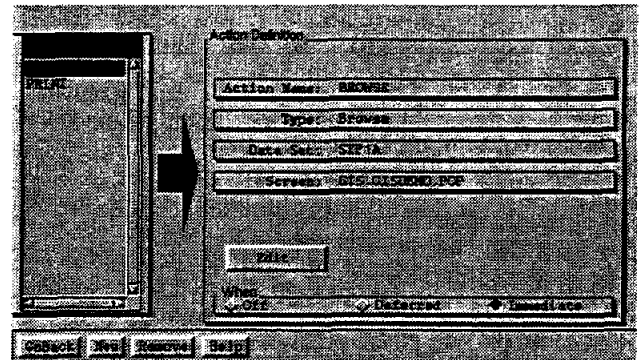


Now after choosing the POP100 variable, which shows the total population for the census tract, the area at the bottom of the layer window can be used to set the color for the various population ranges. The range values are shown next to the variable and the color in the area in the bottom left corner of the window. The arrows in this area are used to move thru the ranges. The range in the figure is from 1600 to 2400.



Defining Actions

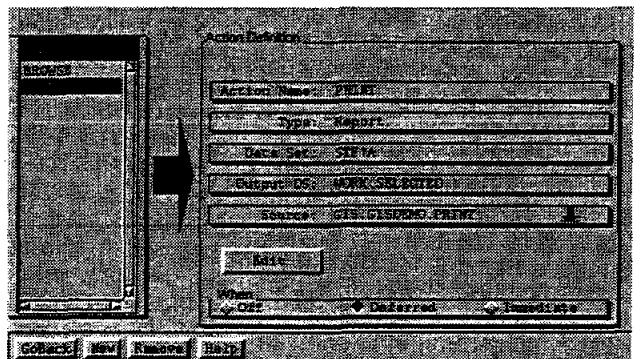
To define an action to display the STF1A variables for a selected census tract, create a browse action with a custom FSBROWSE screen. This will be an immediate action so each time a different tract is selected, different data will appear in the FSBROWSE window.



Define a report action

Another useful action to define would be a report action to print some of the the STF1A variables associated with the selected tracts. A report type action using the STF1A associated data set can be defined to create a data set, WORK.SELECTED, and submit the code stored in a source entry, SASUSER.GISDEMO.PRINT.SOURCE. This particular entry contains the following code:

```
proc print data=SELECTED;
    var TRACT
        POP100
        HU100
        P2_1 /* families */
        P5_1 /* male */
        P5_2 /* female */;
run;
```



Querying the Map

There are several ways that features can be selected from the map. Using the mouse, features can be selected by clicking, dragging a rectangle or circle, or drawing a polygon. Features can also be selected by their attribute values using an SQL where clause builder. The various selection mechanisms can be combined. For example, you could drag out

a 10 mile circle around a retail business and select all of the census tracts in the circle. Then use a WHERE clause to select only the tracts with 400 or more people over age 65.

Other Examples

While it would be impossible to list all of the possible applications for SAS/GIS Software, I will describe a few.

Real Estate

Another application that most people can relate to is real estate. Using a street database from TIGER or Dynamap/2000 you can add point data to represent available houses for sale. Then you could drag out a 10 mile circle around a business and select all of the available houses in the circle. Then use a WHERE clause to select only the houses with 4 or more bedrooms and a garage. Actions could be defined to allow browsing data about the house (like number of bedrooms and baths, size of garage, square feet, price, etc.) and showing a picture of the house in a separate window.

Market Research

A market research application to examine and build a demographic profile of customers can allow better target marketing and data to coordinate consumer preference studies. A simple start would be to view customer locations on a map. Then exam the demographics of the census blocks or tracts where they live. Next select other blocks or tracts with the same demographics. These are also potential customers. Are they on the mailing list? Draw symbols at the points from the addresses in the mailing list. If so, why are they not buying? Draw the competition on the map. Is the competition closer? Is there a barrier, like a river? The asking and answering of these kinds of questions lead to a better understanding of a market and more efficient marketing expenditures.

Conclusion

SAS/GIS Software will add a powerful tool to the SAS System that allows enterprises to explore and display the spatial relationships in their data. The software is highly interactive, easy to use and does not require a sophisticated knowledge of GIS to setup and use. These are only the highlights of features for a Version 6.09 release of SAS/GIS Software. Other features will be available for both Version 6.09 and beyond.

Further Information

For more information on U.S. Census Data Products including Summary Tape Files and TIGER contact: Customer Services, Bureau of the Census, Washington, DC 20233, Voice: (301) 763-4100 Fax: (301) 763-4794.

For more information on the U.S. Defense Mapping Agen-

cy Digital Chart of the World (DCW) contact: USGS Open File Section, Box 25286, Denver, CO 80225, Voice: (303) 236-7476.

For more information on the U.S.G.S. DLG Files contact: Reston-ESIC, U.S. Geological Survey, Room 1C402, Reston, VA 22092, Voice: (703) 648-6045.

For more information on the Dynamap/2000 contact: Geographic Data Technology, Inc., 13 Dartmouth College Highway, Lyme, NH 03768-9713, Voice: (800) 331-7881.

SAS is a registered trademark of SAS Institute Inc. SAS/GIS is a trademark of SAS Institute Inc. Dynamap/2000 is a registered trademark of Geographic Data Technology, Inc., Lyme, NH.