1. Abstract

The U.S. and world mapping capabilities of SAS/GRAPH® are used by analysts at Eastman Kodak Company to display sales territory data and the results of consumer surveys. However, these users often do not have sufficient SAS background to write the SAS code required to produce these maps.

This paper discusses our approaches to enable novice SAS users to:

- Create a response data set, if necessary, using SAS/FSP®
- Subset a response data set
- Generate maps
- Output the maps on a shared IBM3287 printer in a VM/CMS environment.

2. Introduction

The mapping applications at Eastman Kodak Company are usually one of the following four types:

- U.S. state and county mapping
- World and regional mapping
- Choropleth mapping with symbols
- User-defined territory mapping

The AUTOMAP system fully supports the first two types of mapping because they are the most popular. A brief overview of each of the above mapping applications follows.

U.S. state and county mapping: Figure 2.1a shows the states map. The county map with empty counties within state outlines is presented by Figure 2.1b. Finally, Figure 2.1c presents the mapping of subset counties in a multi-state area, while retaining the outer boundaries of the states.

World and regional mapping: These maps may depict the whole world as in Figure 2.2a or regions, such as the AAA (Africa, Asia and Australia), Latin America or European regions. Figure 2.2b shows the Latin America regional map.

Symbol choropleth mapping: This is a mapping technique in which symbols are overlaid on a geographical area. For instance, a communication network can be overlaid on the European region (see Figure 2.3a).

User-defined territory mapping: Often, a user's response variable is mapped by sales territory or ADI (Area of Dominant Influence). These geographic coordinates are not supplied as map data sets with SAS/GRAPH®. With the GREDUCE procedure, these boundaries can be created. Figure 2.4a and Figure 2.4b present maps using user-defined boundaries.

3. Annotating a Map

Figure 3.1 shows a map generated with AUTOMAP default annotations. The SAS code used to generate this map is shown in Figure 3.1a. As you can see, there is no title, the variable name appears in the legend and each value in the data set has an entry in the legend. Figure 3.2 has a title. This is added to the map through the use of the SAS TITLE statement. (See Figure 3.2a.) Figure 3.3 has a footnote below the map. Notice that the size of the map has been adjusted to accommodate the title and footnote. The footnote was added through the use of a FOOTNOTE statement. (See Figure 3.3a.) No additional annotation is added in Figure 3.4 but the legend title has been changed from "POP" to "Population." This is caused by the addition of the LABEL statement. (See Figure 3.4a.) To set up ranges for the legend instead of the default single data values (Figure 3.5), the FORMAT procedure was used to reformat the data values. A FORMAT statement instructing SAS® to use these reformatted values is also added to the GMAP procedure code. (See Figure 3.5a.) Figures 3.6 and Figures 3.7 show the same map with all of the annotation as a PRISM and BLOCK map, respectively. The change in map type is caused by changing the map type keyword in the GMAP Procedure. (See Figure 3.6a and Figure 3.7a.)

4. Using the AUTOMAP System

PROC GMAP requires two SAS data sets in order to produce a map - a response data set, containing the values of the variable to be mapped, and a map data set, containing the geographic coordinates for the boundaries of each unit area. AUTOMAP has provisions to help the user create both of these data sets, if necessary.

Very often the response data set already exists. The data, collected as a result of a consumer survey, may have been analyzed using SAS® or territory information may exist in a SAS data set for on-going retrieval and analysis. If the data to be mapped already resides in a SAS data set with the appropriate geographic identification (ex., state FIPS code, country ID) then AUTOMAP will use the existing SAS data set.

However, sometimes the data to be mapped is not resident in a SAS data set. In this case, the user can use the SAS/FSP® facility of AUTOMAP to generate a response data set. The user is prompted for the geographic area of the map - U.S., Latin America, Africa, etc. and the variable name. A screen is then available for each geographic unit. All the user needs to do is enter the response variable value. AUTOMAP supplies the
geographic identification necessary. (See Figure 4.1)

In order to create the proper map data set, AUTOMAP asks the user to supply the name of the response data set and the name of the variable to be mapped. At this point, the user is given the opportunity to subset the response data set. If subsetting is desired, all the user has to do is enter a logical statement. AUTOMAP then asks the user if all geographic units should be shown on the map or just those with geographic id's in the response data set. AUTOMAP generates the appropriate map data set based on this information. Therefore, the user only has to be concerned about the response data set because AUTOMAP will generate the appropriate map data set.

AUTOMAP also prompts the user for the type of map desired - choropleth, block or prism. The user can specify the type of range labels and the break points for the ranges. The user can annotate the map with a legend label and up to ten titles and ten footnotes.

The user is asked for the output device type for the map. AUTOMAP determines the shading patterns based on the device type. If IBM 3287 output is requested, a GOUT data set is created which can be spooled to a shared IBM 3287 printer. This capability allows a user to produce maps from a non-graphics device.

See Figures 4.2 through 4.4 for an example AUTOMAP session.

5. Internal System

The AUTOMAP system is actually a collection of SAS programs and SAS map data set libraries which reside on a CMS mini disk. Since this mini disk is an extension of the user's A disk, a user can invoke AUTOMAP with the use of the GETSAS command during a SAS session or in a SAS command before the SAS session.

There are two customized SAS map data libraries, MAPS and MAPW, in the system. The first one, MAPS, contains US state and county boundaries. The second one, MAPW, contains boundaries of other countries. The US county map data set was projected, reduced and then broken down by state. For example, the county map data set of New York is MAPS.NY. For county maps, this structure of the MAPS library increases the speed of processing map data sets and plotting maps. Unfortunately, the same method cannot be applied on the MAPW library because the SAS projection method cannot handle data for the whole world at once. Consequently the whole world map data set was broken down by region and then projected and reduced. However, this means that different regions cannot be overlaid.

The density of the MAPW library is 2. This means that only approximately 8% of the points in the original world map data set are used but most countries retain their appearance. The exceptions, Hong Kong, Singapore and Puerto Rico, are too small to appear on a map. Those small islands are important for our users so circles are placed on a map to indicate them.

The core parts of the system are three SAS program generators: a user response data set generator, a SAS mapping program generator and a map data set generator. Based on individual need, those program generators create SAS programs by the use of "PUT" statements. This method was chosen because of dynamic user needs.

The mapping system begins by prompting the user for information needed to build a SAS mapping program, AUTOMAP SAS A, on the user's A disk. Based on user needs, the AUTOMAP program may have all or part of the following SAS code:

1. GOPTIONS 6 TITLE
2. PATERN 7 FOOTNOTE
3. CMS FILEDEF 8 PROC GMAP
4. PROC SORT 9 LABEL
5. PROC FORMAT 10 FORMAT
6. CMS MASLIE DEL;

Once this is done, the map data set generators are intelligent enough to know how to reshape the map data set. This is particularly important in the mapping of subset countries in a multi-state area, while retaining the outer boundaries of the states. Finally, the system executes AUTOMAP and produces maps.

6. Development Considerations

In developing this system, the primary considerations were (1) programming ease; (2) expansion flexibility; (3) map production on various types of graphic devices; (4) operation by less experienced users.

(1) Programming ease: SAS® was chosen as the programming language for AUTOMAP because of the richness of its programming facilities. Since AUTOMAP is a SAS application system, it runs without leaving the SAS session.

(2) Expansion flexibility: The system is written in a modular form so that expansion can be done easily. In the past two years, several major enhancements have been made by adding new modules to the system.

(3) Map production on various types of devices: Since SAS/GRAPH® is device independent, this was easy to implement. But in the past two years, the availability of plotting devices for our large user community has been a problem. This is largely due to the fact that graphic devices cannot be shared by CMS users.

Recently, a graphic virtual machine which drives an IBM 3287 printer has been designed. If a user chooses an IBM 3287 as the output
device, it is not necessary to attach the printer to the user's virtual machine. The AUTOMAP system produces a GOUT file, spools this file from a user virtual machine to a graphic machine, and produces maps on the IBM 3287 device. This approach allows an IBM 3287 printer to be shared by many CMS users. The feedback from users has been very positive and the number of IBM 3287 printers under this configuration is growing in our installation.

(4) Operation by less experienced users: The system is menu driven but not full screen. All the user has to do in order to produce a map is to follow AUTOMAP's instructions.

Although the AUTOMAP facility was originally designed with the novice SAS user in mind, it is also used by experienced SAS programmers to generate "skeleton" SAS code for mapping applications. This reduces the programming time required for new applications. It also reduces the maintenance required for these applications.

7. Future Enhancements

We have had great success with the use of the current system; however, we plan to add two enhancements to the system: the ability to generate class intervals for a response data set and several new map data sets containing user-defined boundaries.

The class interval is used to group the response data values so that they can be represented by a few categories. As we have seen, each category (e.g. range) and the corresponding pattern are shown in the legend of a map. Thus, the class interval has a great impact on a map's interpretation. An inappropriate class interval can misrepresent a response data set. In the AUTOMAP system, a user must provide the upper and lower data limits of categories (e.g. ranges) to construct a legend. Our plan is to introduce some methods for classifying a response data set. Based on the selected method, the system will perform some analysis on the user's response data set and determine the class intervals. Before plotting a map, a user will have an opportunity to preview and change the distribution of the response data set within the class intervals.

The GREMOVE procedure enables the user to create user-defined boundaries, such as sales territory or ADI (Area of Dominant Influence) from a map data set of standard geographic boundaries, for example, states and counties. We plan to introduce these user-defined boundaries into the AUTOMAP system so that a user of sales territory or ADI maps can take full advantage of these facilities in the AUTOMAP system.
**SELECT THE REGION TO BE MAPPED**

1. **AMERICA**
2. **EUROPE**
3. **AFRICA**
4. **THE WHOLE WORLD**

**VALID SELECTIONS ARE 1 2 3 4 5 6**

**ENTER RESPONSE VARIABLE NAME**

---

**Select data set: WORK.WORK**

<table>
<thead>
<tr>
<th>Command</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACTION</td>
<td>1</td>
</tr>
<tr>
<td>INDEX</td>
<td>1</td>
</tr>
<tr>
<td>IDN</td>
<td>90</td>
</tr>
<tr>
<td>NAME</td>
<td>BELGIUM</td>
</tr>
<tr>
<td>FORDER</td>
<td></td>
</tr>
</tbody>
</table>

**[[Figure 4.1]]**

---

**Select data set: WORK.WORK**

<table>
<thead>
<tr>
<th>Command</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACTION</td>
<td>1</td>
</tr>
<tr>
<td>INDEX</td>
<td>1</td>
</tr>
<tr>
<td>IDN</td>
<td>90</td>
</tr>
<tr>
<td>NAME</td>
<td>FRANCE</td>
</tr>
<tr>
<td>FORDER</td>
<td></td>
</tr>
</tbody>
</table>

**[[Figure 4.1]]**

---

**What Kind of Map**

1. **Choropleth**
2. **Political**
3. **Block**
4. **Topographical**

**Enter Name of Response Variable**

**[[Figure 4.3]]**

---

**Do you need title?**

1. Yes
2. No

**Enter Title**

**Do you need ranges for response variable?**

1. Yes
2. No

**Do you need range labels?**

1. Yes
2. No

---

**What Type of Map**

1. All states in US
2. Selected states in US
3. All counties in states
4. Selected counties in states

---

**[[Figure 4.4]]**

---

**Northeastern Region**

---

**[[Figure 4.5]]**