ZOOMMAP AND OTHER SPECIAL TECHNIQUES
FOR SAS MAPPING APPLICATIONS

Warren Repole Jr., Info Tech Inc.

INTRODUCTION

Maps can be invaluable when data of a geographic nature needs to be displayed and understood. PROC GMAP is adequate for the majority of SAS/GRAPH* mapping applications. However, there are some types of maps which a user with only a limited knowledge of SAS* programming and PROC GMAP is unable to produce because of the complex programming techniques required.

Among the most requested special applications are "subset" maps and maps with labels. Maps with labels are almost trivial to produce with the addition of the ANNOTATE facility to SAS/GRAPH. Subset maps, which can be much more difficult to generate, fall into two major categories:

1. Close-up maps in which we "zoom in" to show a small area of interest within a larger region.
2. Regional maps in which a large area is "split up" into smaller individual maps.

To permit more extensive use of the mapping capabilities of the SAS System, a set of algorithms has been implemented using the %MACRO language to assist with the creation of subset maps and maps with labels. These macros have been developed to allow users who are familiar with PROC GMAP, but who do not have sophisticated programming skills, to produce useful maps. To permit maximum flexibility, each technique has been implemented with numerous options which can be selected by the user. While these macros are helpful under current versions of the SAS System, the addition of the SAS/GRAPH ANNOTATE feature will greatly enhance their usefulness.

CLOSE-UP MAPS: %ZOOMMAP

Subsetting a large map area to focus on a small area of interest can be a difficult and bothersome task. The default process used by PROC GMAP automatically "subsets" the map coordinate data set so that only the ID areas which have non-missing response values are shown. However, when these areas are not contiguous, the resulting map suffers from an "island" effect (Figure 1). Even when the areas are contiguous, the map may have an irregular or unrecognizable shape.

Specifying the ALL option on the PROC GMAP statement will retain all intervening and surrounding boundaries, but the close-up effect needed to emphasize the area of interest is lost (Figure 2).

The %ZOOMMAP macro creates a "close-up" map data set in three steps.

1. The area of interest is determined by enclosing all non-missing response areas in a rectangular "window" as shown in Figure 2. When the ANNOTATE facility is being used to label points on the map, the window is defined to enclose all X-Y ANNOTATE coordinates.
2. The response window is adjusted to conform with the dimensions of the graphics device chosen.
3. After the device-specific window is established, the appropriate subset of the original map coordinate data set is obtained.

The %ZOOMMAP macro requires projected map coordinates. Coordinates for ANNOTATE labels must be projected in the same scale as the map coordinates. (The %BONNE macro described later makes equivalent projection of the map and label coordinates simple.) The user specifies the map ID variable(s) along with a single response variable or an ANNOTATE data set.

Additional options available with %ZOOMMAP include:

1. A border drawn around the close-up area.
2. An expansion factor used to add a surrounding "cushion" to the close-up area.
3. Removal of internal boundary lines of non-response areas (as with PROC GREMOVE).
The user also chooses the graphics device and whether to use the ROTATE graphics option. The default graphics configuration is that of the IBM 3279. If another device is desired, a SAS data set must be defined which contains some of the graphics device characteristics and parameters available through PROC CTESTIT (e.g., ROWS, COLUMNS, GNRATIO). The %ZOOMMAP macro will search this "GPARMS" data set to obtain the proper device configuration.

Figure 3A shows a sample %ZOOMMAP application without the BORDER option, produced on a ZETA plotter. Figure 3B is the output when the BORDER option is selected. Figure 4 highlights several counties which defined the close-up area. The GREMOVE option has eliminated all non-response county borders. Figure 5 resulted from using %ZOOMMAP with an ANNOTATE data set containing the locations of the cities indicated on the map.
Another difficulty with mapping applications is the projection of map coordinates. Using PROC GPROJECT on different sets of coordinates can yield incompatible results. In order to obtain compatibility between several sets of coordinates, they must be projected together in a single PROC GPROJECT step and later separated. This is the current case with an ANNOTATE data set; the ANNOTATE points must be combined with the map coordinates before projection and separated after projection.

The `%MACRO BONNE` uses an equal-area projection algorithm which can calculate projected coordinates point by point, just as calculations are performed within a DATA step (one observation at a time). If the central coordinates specified for the algorithm are held constant, many coordinates can be projected separately with absolute compatibility.

One of the features that makes `%BONNE` such a powerful tool is its ability to automatically calculate central coordinates and save them in a SAS data set for later use. This option was utilized to produce Figure 5. The ANNOTATE data points were projected using `%BONNE` and the central coordinates were saved. These central coordinates were used as the basis for the projection of the map coordinates. This approach is much simpler, and more efficient, than the combine-project-separate process required with PROC GPROJECT.

Like other conic projection algorithms, only coordinates located on one side of the Equator can be processed. Coordinates in the Northern and Western Hemispheres are assumed to have positive latitude and longitude, although these assumptions can be waived by the user. The algorithm performs well for all locations within the continental United States (Figure 6).

Other options are available for `%BONNE` including the conversion of degrees to radians before projection.

```sas
%BOSS(INMAP=POLETS,
OUTMAP=POLETS,
DEGREES=YES,
OUTBASE=CENTR);
DATA MAP; SET MAPS.STATES;
%BONNE(INBASE=CENTR);
RUN;
```
When a large area is divided into smaller regions, PROC GMAP can produce a separate map for each region through the use of a BY statement. For example, if REGION is defined in the RESPONSE data set, regional maps of the United States, one region per graphics "panel", could be created by

```plaintext
PROC GMAP DATA=RESPONSE MAP=MAPS.US, BY REGION;
```

Unfortunately, when some ID areas lack response values, the "island" effect can result (Figures 7A and 7B), as with close-up maps. The ALL option does not solve the problem (Figures 8A and 8B). The addition of other non-geographic variables to the BY statement, such as YEAR, makes the situation considerably more complex.

By filling the "gaps" in the response data, the problem of regional maps can be solved. The general solution implemented by the %SPLTMAP macro determines all combinations of the ID variable(s), the region-defining variable(s), and the additional BY variable(s). Combinations which do not have corresponding response values are added with a user-specified pseudo-response value.

In order to generate combinations of region and ID variables, a reference data set must be created to associate each ID value with a region value. This "link" data set is required for the %SPLTMAP macro.

Figures 9A and 9B illustrate a case in which the U.S. is divided into regions and the individual regions have incomplete response data. The %SPLTMAP macro was used with one additional BY variable (MODEL) specified.

Figures 10A and 10B show another problem associated with using a single map coordinate data set to produce multiple maps. If the regions are sufficiently far apart, the projection of their coordinates can produce a slight "tilt" in the individual maps.

To solve the "tilt" problem, an option is available for %SPLTMAP which uses a version of the %BONNE projection algorithm (%SPBONNE) to project the coordinates of each region separately. The independently projected regions (in this case, single states) are shown in Figures 10C and 10D.

**NOTE**

The combinations of region, ID and other BY variables are generated using the cartesian product function of a procedure developed at Info Tech, PROC GENJOIN.
SUMMARY

The SAS System, including SAS/GRAPH, is a powerful tool for the preparation and display of information. When information is of a geographic nature, PROC GMAP is relatively easy to use to generate maps.

However, some custom map applications require special techniques which may be beyond the programming level of the user. The %ZOOMMAP, %BONNE, and %SPLTMAP macros help perform sophisticated tasks with limited coding. With the assistance of these macros, and the addition of the ANNOTATE facility to SAS/GRAPH, many types of useful maps are literally at the user's fingertips.

FOR FURTHER INFORMATION

Warren Repole Jr.
Info Tech, Inc.
P. O. Box 14545
Gainesville, FL 32604
(904) 375-7624

A TYPICAL USE OF %SPLTMAP

%SPLTMAP(INDATA=MAPDATA,
OUTDATA=NEWDATA,
RESPVAR=SALES,
MISSING=0,
IDVARS=STATE,
PANLVAR=REGION,
BYVARS=YEAR MODEL,
MAPLINK=REGIONS);

PROC GMAP DATA=NEWDATA MAP=MAP;
BY YEAR MODEL REGION;
ID STATE; CHORO SALES;
REFERENCES


TRADEMARK NOTICE

SAS and SAS/GRAPH are registered trademarks of SAS Institute Inc., Cary, N.C. USA.