PROBLEMS AND CONSIDERATIONS IN SAS/GRAPH MAP GENERATION

Mitchell J. Tuller, The Rand Corporation

INTRODUCTION

The mapping capabilities now available within SAS/GRAPH make it a very powerful and flexible tool. Unfortunately, unless considerable care is taken in the preparation of maps, much of the impact and utility is lost. This paper describes some of the important techniques and procedures that one can use to ensure that each map portrays precisely the information intended.

BACKGROUND

The release of SAS 7.9.5 included among its many changes a set of procedures solely for mapping. The procedures are PROC GMAP, PROC GREMOVE, PROC GPROJECT, and PROC GREDUCE. In addition, the PATTERN global statement was enhanced to allow one to prescribe the characteristics of the desired map. A number of SAS format boundary coordinate ("map datasets") were also contained in the release. The map datasets have plotting points for the states and counties of the United States and for the Canadian provinces. Available for an extra fee are boundary coordinate data for the entire world. The world map dataset consists of data for the world (both projected and nonprojected) as a whole, and broken down by continent. Each continent is further divided into either countries or political subdivisions.

The associated manuals do a fair job of describing the data and the attributes of the data, but do little to orient a user toward the production of effective maps. In fact, there has been little literature on the subject of computer-generated color maps. The subject is still relatively new, and procedures and methods have only recently begun to be disseminated. The intent here is to cull the more prevalent and useful map construction techniques and apply them to SAS/GRAPH.

DIGITIZATION

Digitization can best be defined as the translation of data into numerical digits. Put more explicitly, it is the transformation of data from a coordinate system (i.e. latitudes and longitudes) into numerical values. The device that is used to digitize maps is, appropriately enough, called a digitizer. It is simply a flat-bed tablet on the surface of which the manuscript map is placed. When a cursor-like device is traced over the desired areas, an electronic signal is transmitted indicating the location of the cursor in an x-y coordinate system.

There are two methods of digitizing data. They differ in the method of sampling the data and in the density of information recorded. In the "point mode" technique, selected data points are recorded at the digitizer operator's discretion. This technique is used when a low density of points is required for a particular application. In the "stream mode" technique of digitization, points are sampled at precise time or distance intervals, resulting in a very high density of points.

Each technique has both advantages and drawbacks. Point mode data capture produces sparse data samples. The need for post-processing is usually reduced or eliminated. But intricate nuances in the line being traced will probably not be reflected in the final product. Stream mode data recording will probably capture intricate nuances but will also densely sample relatively simple areas. Further processing of the recorded data is usually required to eliminate redundant entries.

Digitization is the most accurate method of translating data from maps into a machine readable form, but it is still far from perfect. The output data can only be as good as the source material. There are limitations to the accuracy and sensitivity of the digitizer itself. Intricate nuances in the source material may be outside the tolerance level of the digitizer. In practice, the most common source of error is human, ranging from the operator being too lazy or careless to properly trace the line to barely perceptible twitches and spasms in the operator's hand as he moves the cursor.

Creating SAS Datasets from Coordinate Data

At this point, it should be obvious that users are not restricted to SAS coordinate datasets. A user with access to a digitizer is fully capable of creating his own coordinate datasets. But there are several important considerations.

The most important thing to remember is that PROC GMAP can only draw closed polygons. This means that whatever data points are used, they must describe a closed figure; the data must be arranged so that the line segments that represent the borders are in the proper order. Because most commercially available map databases do not exhibit this organization, it is up to the user to rearrange the line segments so that the geographic units form closed boundaries. Also, PROC GMAP can not handle polygons inscribed within other polygons. It may be necessary to modify the data to avoid that situation. One solution is to join the inscribed polygon to the outer one by a line.

The choice of measurement units for the stored data is also important. When map coordinates are provided on a computer tape, the latitudes and longitudes are usually expressed in degrees rather than radians. These data will eventually have to be converted to radians because the trigonometric functions required to project the map work only in radians. The decision when to convert is dependent largely on the particular
use of the data. If a great deal of coordinate
dataset manipulation is anticipated, it is probably
best to leave the data in degrees because most
people are used to working in degrees. However,
if the data are static and human intervention is
minimal, it would probably be more economical to
store the data in radian format. Datasets
provided by SAS are in radian format.

A third consideration regarding the storage of
the coordinate data is whether the data should be
saved in a projected form. This has a bearing on
the units in which the data are saved; if the
coordinate data is whether the data should be
stored data are in projected form, they are
saved in a projected form. This has a bearing on
the projection being used. In particular,
neither in radians nor degrees, but in arbitrary
units that depend on the type of projection. The
general rule regarding whether to save the data
in projected form is as follows: if the coordinate
data will not be subsetted to such a degree that
the projection is inappropriate and the user has
no need to keep the data in raw form, then it is
most economical to save the coordinate data in a
projected form. Otherwise, it is best to save it
unprojected. Occasionally one needs to transform
projected coordinate data back into latitudes and
longitudes. This may be possible but it depends
on the projection being used. In particular,
reversing a cylindrical or azimuthal projection
will usually involve straightforward mathematics,
but reversing a conic projection may be
impossible unless the standard parallels are
known.

A final point to be made regarding the creation of
a map database concerns its density. A dataset
rich in data points is obviously more expensive to
process than a sparse one. High density is
wasteful if the scale of the map is not sufficient
to exhibit the richness. Hardware constraints
are another consideration. If the projection
equipment is incapable of displaying intricate
detail present in a database, there is no point in
having the richness. Once again, the decision
whether or not to save a rich database lies in its
potential application. If one were generating the
same map often, it would be advisable to have a
database consisting solely of the points required
to make the map. If however, storage space is a
consideration, there might be little choice but to
extract from a master database. In this case, it is
obvious that the master database must contain
sufficient points to satisfy all possible subsetting
needs.

MAP PROJECTIONS

The map projection used is one of the most
important decisions in the construction of a map.
All maps need to be projected because the earth
is round and the piece of paper on which the map
is drawn is flat. The term map projection can be
defined as "an orderly or systematic arrangement
of the earth grid on a plane surface." There are
over a hundred map projections defined and each
one possesses a feature not found in the others.
The choice of map projection depends upon what
areas of the earth are being displayed and what
characteristics in that area the user wants to
emphasize.

Distortion is inherent in all maps. It is simply
impossible to project a round surface onto a flat
surface and have all areas of the map stay in
true proportion, shape, distance, and direction
with one another. The purpose of map
projections is to minimize distortion in one or two
selected areas. For example, one can accurately
represent azimuthity (true direction) but only
at the expense of scale, shape, and distance.

PROC GPROJECT currently provides three
different projections: Alber's equal-area,
Lambert's conformal, and the Gnomonic
projection. But it is not clear to most users
when one projection should be used over
another. Alber's equal-area projection is used
when one wants to preserve areal relationships
on areas of continental size or less. It displays
the United States quite nicely, and any distortion
increases northward and southward from the
standard parallels. Lambert's conformal also
displays the United States quite well, and
preserves true angular relationships across the
map. It is limited to regions of continental size or less. The Gnomonic projection
displays all great circles (the path representing
the shortest distance between two points on the
earth) as straight lines. Distortion on this
projection increases quite rapidly from the center
of the map outward. Like the Alber and Lambert
projections, the Gnomonic projection is best when
limited to a continental maximum area.

There are several other projections that one
might find useful with PROC GMAP. The fact that
PROC GPROJECT can not produce them is no
limitation. They just need to be implemented in a
DATA step. Some of the more popular
projections, their properties and uses are:

1. Mercator
A conformal (true bearing) projection
suitable for plotting the entire world. Distortion
increases in a poleward fashion. Most often used for
navigational purposes.

2. Cylindrical equal-area
An equivalent (equal area) projection
also suitable for plotting the entire
world. Distortion increases north and
south from the equator or standard
parallels. Often used for mapping
distributions when areal size is of
major importance.

3. Stereographic
A conformal projection suitable for
hemispheric areas. Distortion
increases from the center outward.
Used when true direction is
important.

4. Rectangular
A compromise (a projection in which
total distortion is minimized by
allowing minor amounts of distortion
in all areas) projection well suited for small areas. Distortion increases in northward and southward directions. Easy to construct, it gives a reasonable diagrammatic representation of the data.

Mathematical formulas on the implementation of these and many other projections can be found in geography or cartographic textbooks. Another excellent source is the Pearson reference.

It is important to remember that the projection formulas of PROC GPROJECT expect the latitudes and longitudes (y and x) to be in units of radians and the sign of the longitude to be the opposite of conventional notation. Under standard notation, north latitudes are positive, south latitudes negative, east longitudes positive, and west longitudes negative. PROC GPROJECT expects east longitudes to be negative and west longitudes positive. This is not true of the projection formulas found in most reference sources.

COLORING AND LEGENDS

The ability of computers to generate color maps quickly and cheaply is a relatively recent phenomenon, and consequently there has not been a great deal of material available regarding sensible and effective map shading. It is imperative that the map user be able to quickly understand and memorize the legend. Ideally, the user would internalize the legend, that is, grasp its meaning with little to no thought. There are several examples of this: a large symbol intuitively means more of a particular quantity than does a smaller one and a gradual shading within a particular color implies a monotonic change in the quantity being portrayed.

These concepts are fine and should be used when possible, but they are not always possible with SAS/GRAPH. The options available to a SAS/GRAPH user are limited to color (hardware permitting) and shading patterns. For maps displaying a single-response variable, there are two options. One may represent the different response levels by different colors or by different shading patterns. The use of different shading patterns is best when the data are continuous. The increasing thickness of the shading lines represents the natural ordering of the data values. It is not good practice to take data that are naturally ordered (e.g., continuous distributions) and display them with a scheme that contains an arbitrary ordering (e.g., random colors). Colors are more useful for data that are categorical and for which there is no preconceived notion about what colors are associated with what values. Descriptive variables such as county type and metropolitan adjacency and most other variables that contain discrete responses are natural choices for a color-display scheme.

The user finds a new set of problems when dealing with a map that contains two variables. With bivariate maps, color retinal relationships are not clearcut. Black is the darkest color and white the lightest, but what the middle colors are and in what order they fall is not intuitively obvious. For maps with one discrete variable and one continuous variable, the univariate rules apply. Even if both variables are discrete, one can choose shading patterns judiciously and provide enough variation in the shading patterns so as not to give the appearance of continuity. The big problem occurs when both variables are continuous. SAS/GRAPH lacks the sophistication to properly handle this situation and there is no clearcut general solution.

Map Dataset Modification Techniques

When referring to SAS map datasets, dataset modification usually means one of three things: extraction, projection, or pirating. Extraction may be a combination of PROC GREDUCE, PROC GREMOVE, and a DATA step. Projection involves either PROC GPROJECT or a DATA step. The term pirating used here refers to the technique of extracting data from the map dataset to create a response dataset.

The "filtering" procedures PROC GREDUCE and PROC GREMOVE are fully explained in their respective documentation, but there is another commonly used filtering method that is not implemented in any procedure. The method of filtering by distance is similar to the E1, E2, ..., E5 options of PROC GREDUCE, but instead of assuming a flat earth and using the Euclidean Distance formula, it compensates for the curvature of the earth by using the Great Circle Distance formula. Use of this formula guarantees accuracy over the whole globe and is most appropriate where large distances are involved or extreme northern or southern points are being referenced.

Implementing one's own projection is similarly done in a DATA step. Care must be taken to properly choose a standard parallel or parallels in those projections where required. A good rule of thumb in choosing a pair of standard parallels is to make each one approximately one-sixth of the way from the edge of the page. This allows the middle two-thirds to enjoy minimal distortion.

When working with a new map dataset, one may wish to simply display the map but does not have a response dataset readily available. One technique for getting around this problem is to extract variables from the map dataset itself and use them as the response dataset. Alternatively, the user may subset the map dataset by keying on the ID variable. If the map dataset is sorted by the ID variable, it can easily be subsetted by using the FIRST.Id variable feature in a DATA step, creating a response dataset containing one observation for every level of the ID variable.

SUMMARY

The ease with which SAS allows a user to generate maps introduces a whole new set of
considerations to the user. By utilizing a mixture of common cartographic techniques and data processing fundamentals, the user can create very complex maps with little effort. But unless the user possesses a real understanding of the tools available to him, they can never be utilized to their fullest. There are many more topics that could be discussed in a paper of this type but the intent here is simply to scratch the surface. It is hoped that the reader will be encouraged to explore more deeply some of the subjects discussed and gain greater benefit from SAS/GRAPH.

REFERENCES


Pearson, Frederick II, Map Projection 1977 Equations (March), Naval Surface Weapons Center, Dahlgren Laboratory. Report TR-3624.
