Abstract

In our job, we produce several books and studies a year about the nonprofit sector and the charitable behavior of Americans. We use different components of the SAS® system to do the "leg" work (data manipulation, statistics, and graphics) necessary for producing a publishable product. This paper concentrates on the graphical aspect of production, specifically using some of the features of SASGRAPH in designing and constructing various styles of maps.

We try to utilize the SAS system as much as possible in production even though other commercial packages, such as Corel or Mapinfo, may be better suited to the production of graphics and maps. We stay with SAS because we use it in the initial step of data management, and we find working with one system as much as possible is more efficient than continually transferring data and other information to different software packages.

Introduction

Maps are an efficient and effective way of displaying a lot of information in one picture. At times, we need to show different types of information about both the country and regions and would like to do so concisely and in a manner that is easily understood by our audience. Since we are dealing with information based on geography, maps more effectively display the information than designs like pie charts or bar graphs do. The map provides a visual perspective within which to view the information in a way that charts and graphs cannot.

There are all sorts of maps and map options, so deciding on the type of visual demonstration is the first step. SAS allows for elaborate and complicated maps, however we have found that the more basic and simple maps can be just as effective at information display. All work was done on a Sun Sparc Workstation running version 6.07 of SAS under UNIX.

The Map

Our latest study required that we show the distribution of nonprofit organizations in the country by census region and in the context of the distribution of religious congregations and the U.S. population. We narrowed our options for presenting this information via maps down to two. Our first option, and the one we chose, joined the states into the four main census regions, yet separated ("exploded") these regions from each other, while maintaining the general shape of the country. This feature of mapping allowed us to emphasize the point that the information presented referred to the regions and easily conveyed to the audience what states are included in the various regions without us having to note this. The critical information was communicated via text boxes displayed over the regions. (See figure 1.)

We found this presentation to be visually more pleasing than our second option. That option outlined the regions over a map of the country and did not explode the regions apart from each other. The information was communicated by text boxes in the same manner as before. (See figure 2.) While this presentation was just as effective as the first map, it was slightly more complicated as it required the use of two colors to distinguish the regions from each other: one color was used for the outline of the continent and the states, and the second color outlined the regions. The first map only used one color to draw both the continent, regions, and states. Given our desire to keep our displays as simple as possible, in both presentation and production, the first map appealed more to us.

The different types of information we needed to communicate to our audience could have been conveyed via bar graph. However, the graph did not emphasize as well as the map did distinctive aspects of the information we were conveying. With the map we were able to show the distribution of nonprofit organizations by region in the context of geographic size, an important distinction for us. The graph pointedly noted how the regions differed in the distribution of organizations, but could not convey that this was distinctive because of the varying geographic size of each region.

The graph was also not as well suited to our presentation needs as the map was because it required more effort on the part of the audience to comprehend the meaning. The graph required use of a legend if we were to include all of the information we wanted to.

The Code - Map 1

We use SASGRAPH to create our maps because we do not have access to SASGIF yet. SASGRAPH is not always user friendly. For example, SASGRAPH operates with a grid system, which requires trial and error when placing items on the map. Another problem is that while the graphics editor is extremely useful in perfecting the placement of items and the general appearance of the map, once the map has been saved and the graphics editor has been shut down, it is not possible to reopen the graphics editor and edit selected items of the map. This is because when the map is saved, the items and segments of information that have been overlaid on top of the map outline become one piece and are no longer dealt with separately by the graphics editor. Therefore, using the graphics editor to enhance the basic map is not without drawbacks. However, while there are some drawbacks to using SASGRAPH, it is not complicated, just time consuming.

The first step in the mapping process is to create a regional dataset that takes the coordinates of the individual states found in the U.S. map dataset which is located in the map library and to then place these states into the four main census regions. Notably, since SAS does not provide an option in PROC GMAP to explode sections of maps, you have to modify the coordinates of each region in order to move states and regions about the page. This first program will create figure 1.

/* set up printer */
fileout out1 pipe 'lpr -PQMSColor':
/* set graphics options */
OPTIONS
reset=all
device=XColor
target=QMSColor
gsfmode=replace
Hactor=4
gsfile=133
gsname=out1
gsplot=250,90x
gsplot=landscape
cs=white
cs=black

/* get the state coordinates from US map dataset */
DATA regions; set us1;
length division $2 ;

/* divide country into regions based on state flp codes, and then modify the x,y coordinates to form four "exploded" regions */
if state in (23,33,50,25,8,44,36,42,34) then do;
x = x+.06; y = y-.05;
division='NE';
end;
if state in (54,10,04,11,51,37,45,13,12,21,7,28,1,40,5,48,22) then do;
x=x+.05; y=y-.05; division='S';
end;
if state in (55,26,17,18,39,38,27,46,19,31,20,29) then do;
x=x-.070; division='W';
end;
if state in (90,16,56,32,49,8,55,4,62,15) then do;
x=x-.05; division='W';
end;
run;

The second step is to create an annotate dataset which takes advantage of the polygon and label functions to produce the text boxes. At this point placement of the text boxes becomes a matter of the trial and error.

/* Create Annotate Dataset for the four regional text boxes and national totals */
DATA mapannot;
length text $45 color function style $8;
retain years '2' position '6' maxys '5';

/* draw the West poly */
when = 'a';
color='cyan'; size=1.3;
function = 'poly'; style='msolid'; line=1;
x = .290; y = .455; output;
x = .200; y = .05; output;
x = -.360; y = .05; output;
x = -.390; y = .145; output;

/* label it */
color='white'; style='centx'; size=1.51;
position = '5'; function = 'label'; x = .290; y = .14; text = 'WEST'; output;
position = '6';
x = -.370; y = .12; text = 'Congregations'; output;
position = '4';
x = -.250; y = .12; text = 'Charities'; output;
position = '5';
x = .210; y = .12; text = '(21.4%)'; output;
position = '6';
function = 'label'; x = -.370; y = .10; text = 'Congregations'; output;
position = '4';
x = -.250; y = .10; text = 'Charities'; output;
position = '5';
x = .210; y = .10; text = '(16.1%)'; output;
position = '6';

/* draw the South poly */
when = 'a';
color='cyan'; size=1.3;
function = 'poly'; style='msolid'; line=1;
x = .050; y = .055; output;
function = 'polycont';
x = .230; y = .055; output;
x = .230; y = -.15; output;
x = -.050; y = .15; output;
x = -.050; y = .055; output;

/* label it */
color='white'; style='centx'; size=1.51;
position = '5'; function = 'label'; x = .140; y = .06; text = 'SOUTH'; output;
position = '6';
x = .060; y = .08; text = 'Charities'; output;
position = '4';
x = .160; y = .06; text = 'Population'; output;
position = '5';
x = .220; y = .06; text = '(28.0%)'; output;
position = '6';
function = 'label'; x = .060; y = .10; text = 'Congregations'; output;
position = '4';
x = .160; y = .10; text = 'Population'; output;
position = '5';
x = .220; y = .10; text = 'Charities'; output;
position = '6';

/* draw the Northeast poly */
when = 'a';
color='cyan'; size=1.3;
function = 'poly'; style='msolid'; line=1;
x = .200; y = .255; output;
function = 'polycont';
x = .380; y = .16; output;

position = '4';
x = -.250; y = .06; text = 'Population'; output;
position = '6';

/* draw the Midwest poly */
when = 'a';
color='cyan'; size=1.3;
function = 'poly'; style='msolid'; line=1;
x = .040; y = .135; output;
function = 'polycont';
x = .120; y = .135; output;
x = .120; y = .04; output;
x = -.060; y = .04; output;
x = -.060; y = .135; output;

*/

970
The Code - Map 2

This second program will create figure 2. The first step of this program is similar to that of the previous program, but with modifications to account for the contiguous regions. The first step takes a variable titled state, which is a two character variable, and creates the four main census regions from it and also a fifth dummy region for Alaska and Hawaii. The necessity of doing this becomes apparent later on. We then create a temporary variable called state_tmp which converts the variable state into a numeric variable, with the values corresponding to the fit code.

/* create dataset that divides the states into 4 census regions */
data state1; set lib2.state90x (where=state);
length division $;
division = 'None';
if state in ('ME' 'NH' 'VT' 'MA' 'CT' 'RI' 'NY' 'PA' 'NJ') then division = 'Northeast';
if state in ('WV' 'DE' 'MD' 'DC' 'VA' 'NC' 'SC' 'GA' 'FL' 'KY' 'TN' 'MS' 'AL' 'OK' 'AR' 'TX' 'LA') then division = 'South';
if state in ('WI' 'MI' 'IL' 'IN' 'OH' 'MI' 'MN' 'SD' 'IA' 'NE' 'KS' 'MO') then division = 'Midwest';
if state in ('MT' 'ID' 'WY' 'NV' 'UT' 'CO' 'NM' 'AZ' 'WA' 'OR' 'CA') then division = 'West';
if state in ('AK' 'HI') then division = 'ZZZZ';
/* you need this, otherwise AK & HI will be displayed in black */
if division ne 'None';
/* exclude US territories & foreign organs */
est_tmp = estlps(state);
/* convert `character` state into numeric fit code */
drop state;
run;

The next step is to recreate the variable titled state as a numeric variable since the newly created dataset must be merged by a numeric "state" with the library dataset that creates the map of the United States.

/* create `state` since `state` was a character previously and now needs to be a numeric */
data state1; set state1;
state = est_tmp;
PROC sort; by state;
drop est_tmp;
PROC sort; by division;
run;

/* merge x,y, coordinates found in map library onto state1 ds */
data state1; merge state1 lib1.us; by state; run;
PROC sort; by division;
run;

This step removes the individuals state boundary lines from the map, creating the regional map. This is a necessary step because eventually this map will overlay a map of the individual states.

/* remove individual state boundary lines */
data state1; set state1 OUT=restmap;
PROC gremove; by division;
run;

Next, you must create an annotate dataset that will create both the text boxes and label the individual states. To get the coordinates for the state labels, use the map "uscenter" found in the map library. Note that for some selected states, mainly those on the northeast coast, there are two sets of coordinates. We selected the first set for four states, and the last set for all remaining states.
Since we are labelling each individual state, we wanted to make sure that the state label was not visible from behind the text boxes. We took care of that by changing the value of 'when' for any state names that appeared within the boundaries of the text boxes.

/* Create Annotate Dataset to label individual states, and make the text boxes */
DATA maparena:
length text 20; color function style $;
retain state yxs &c;

/* has the x,y coordinates of the center of each state */
function = 'label'; style = 'centx'; when = 'a';
/* for eastern states lib1.uscenter gives you the option for the center within the state or outside of it in the atlantic, take the first set of x,y for these four states */
if state in (10 11 44 24) then do;
  if first.state then do;
    size = 1.35; color = 'black';
    text = ftext(state);
  /* if the state name is within a text box, hide it */
  if (x >= -.38 & x <= -.20 & y >= .05 & y <= .145) then when = 'b';
  output;
end;
else delete;

/* the text boxes. */
JMP:
end;

/* change x,y and the size of names of three states */
if state = '15' then do;
  x = -.190; y = -.17; end;
/* Hawai'i coord */
if state = '34' then do;
  size = 1.20; x = .28000; y = .077; end;
/ New Jersey coords */
if state = '9' then do;
  size = 1.20; y = .11500; end;
/ CT coords */
if state in (25,33,50) then size = 1.20;
/* shrink MA, NH, & VT */
color = 'black';
text = ftext(state);
/* if the state name is within a text box, hide it */
if (x >= -.32 & x <= -.140 & y >= .05 & y <= .145)
  if (x >= -.06 & x <= .12 & y >= -.04 & y <= .135)
    if (x >= -.06 & x <= -.23 & y >= -.15 & y <= .055)
      if (x <= .2 & x <= .38 & y <= -.16 & y <= .255) then when = 'b';
  output;
JMP:
end;

/* draw the West poly */
when = 'a';
color = 'cyan'; size = 1.3;
function = 'poly'; style = 'msolid'; line=1; x = -.320; y = .145; output;
function = 'polycount';
x = -.140; y = .145; output;
x = -.140; y = .05; output;
x = -.320; y = .05; output;
x = -.320; y = .145; output;
/* label it */
color = 'white'; style = 'centx'; size = 1.51;
position = '5'; function = 'label'; x = -.310; y = .12; text = 'WEST'; output;
position = '6';
x = -.230; y = .14; text = 'WEST'; output;
position = '4';
x = -.190; y = .12; text = 'taxi'; output;
position = '6';
x = -.190; y = .12; text = 'Population'; output;
position = '6';
x = -.150; y = .12; text = '21.4%'; output;
position = '6';

/ draw the Midwest poly */
when = 'a';
color = 'cyan'; size = 1.3;
function = 'poly'; style = 'msolid'; line=1; x = -.060; y = .135; output;
function = 'polycount';
x = .120; y = .135; output;
x = .120; y = .04; output;
x = -.060; y = .04; output;
x = -.060; y = .135; output;
/* label it */
color = 'white'; style = 'centx'; size = 1.51;
position = '5'; function = 'label'; x = .050; y = .06; text = 'MIDWEST'; output;
position = '6';
x = .050; y = .11; text = 'Charlies'; output;
position = '4';
x = .070; y = .11; text = '32,368'; output;
x = .110; y = .11; text = '24.3%'; output;
position = '6';
function = 'label'; x = -.050; y = .09; text = 'Midwest'; output;
position = '4';
x = .070; y = .09; text = '74,394'; output;
x = .110; y = .09; text = '26.4%'; output;
position = '6';
function = 'label'; x = -.050; y = .07; text = 'All'; output;
position = '4';
x = .070; y = .07; text = '108,760'; output;
x = .110; y = .07; text = '27.3%'; output;
position = '6';
function = 'label'; x = -.050; y = .05; text = 'Population'; output;
position = '4';
x = .070; y = .05; text = '59,669,000'; output;
x = .110; y = .05; text = '24.0%'; output;
position = '6';

/* draw the South poly */
when = 'a';
color = 'cyan'; size = 1.3;
function = 'poly'; style = 'msolid'; line=1; x = .050; y = .05; output;
function = 'polycount';
x = .230; y = .05; output;
x = .230; y = .15; output;
x = .050; y = .15; output;
x = .050; y = .05; output;
/* label it */
color = 'white'; style = 'centx'; size = 1.51;
position = '5'; function = 'label'; x = .140; y = .06; text = 'SOUTH'; output;
position = '6';
x = .060; y = .08; text = 'Charlies'; output;
position = '4';
x = .182; y = .08; text = '37,365'; output;
x = .220; y = .08; text = '28.0%'; output;
position = '6';
function = 'label'; x = .060; y = .10; text = 'Congregations'; output;
position = '4';
x = .182; y = .10; text = '105,479'; output;
x = .220; y = .10; text = '40.5%'; output;
position = '6';
function = 'label'; x = .060; y = .12; text = 'All'; output;
position = '4';
x = .180; y = .12; text = '142,844'; output;
x = .220; y = .12; text = '36.5%'; output;
position = '6';
function = 'label'; x = .060; y = .14; text = 'Population'; output;
position = '4';
x = .180; y = .14; text = '85,446,000'; output;
x = .220; y = .14; text = '34.4%'; output;

972
Finally we produce three maps. The first one is of the United States with the states labeled in black and with cyan text boxes. The state boundaries are in red. The next map produces an outline of the continental United States with the four regions outlined in black. Note that if you don't use the where clause and exclude Alaska and Hawaii, they will be displayed. The last map overlays the regional map over the state map to produce a map with four distinct census regions.

```
* draw the us map with text boxes and state names */
PROC GMAP DATA=data1 MAP=map1 GOUT=mapcat9;
ID state;
chboro division=OUTLINE=red notitle discrete annotate=map;
quit;
GOPTIONS factor=7;
* draw the Regional Map without the states, in a different color */
PROC GMAP DATA=states MAP=map3 GOUT=mapcat9 all;
ID division; where division ne 'zzzz';
so AK and HI do not show up */
chboro division=OUTLINE=black notitle discrete empty=white;
TITLE ;
note h=2 in c=black j=1 center 'Figure 2' ;
=c by Census Region ;
move=(1,1) h=.12 in 'Note: AK and HI are in the WEST Region.;';
quit;
* Reply the Maps (Regional over State) */
PROC gplay label=sas=mapcat9 toc=sasgif temp=template=whole;
trophy=1:1 1:2 :
run;
quit;
```

**Conclusion**

As evidenced above, none of the programming was difficult or complicated in itself. Rather, the placement of the map and its items was time consuming. The SAS technical support has indicated that SASGIFS and version 6.09 may rectify many of the problems witnessed with SAS/GRAFX and may be more "user friendly."

Making SAS/GRAFX and PROC GMAP do what you want them to do is just a matter of patience, will power, and ingenuity.

Author contact:
Stephen Noga, Director of Technical Services and Programs
Heather Gorski, Research Associate
INDEPENDENT SECTOR
1828 L Street, NW
Washington, DC 20036
(202) 223-8100
INDSEC@gwuvm.gwu.edu