Creating Winning Graphs

Deb Cassidy, Cardinal Health, Inc., Dublin, OH
Lori Griffin, Hoechst Marion Roussel, Inc., Kansas City, MO

Abstract

This workshop will walk users through two graphs that have won in the SUGI graphics competition. One graph uses PROC GPLOT with several 'tricks' to place all the information on the graph. In addition to the use of an annotate dataset, this graph uses options such as offset and blank title statements. The graph won the competition at SUGI 20.

The other graph focuses on the use of PROC GMAP and won first place in Best Presentation of Data-Color at SUGI 18. It shows data for potential sales by county for an individual state plus the 'nearby' section of the surrounding states. The map is used to truly show data for areas such as the Cincinnati metropolitan area which is in three states (Ohio, Kentucky and Indiana). Individual state maps would be misleading for this type of data. The graph also uses annotate to label selected cities on the map.

General Information

There are several books on good visual design of graphics as well as numerous SUGI and regional conference papers that focus on SAS/GRAPH®. We suggest you do read some of this material. You will find that not all the authors agree. Many times the author is writing general guidelines so there can be exceptions. Someone out there will probably have a 'better' way of doing the graphs presented here. This workshop focuses on writing the code but we do want to emphasis two points.

First, you've probably heard the saying "a picture is worth a thousand words". Before you write one line of code, you should know those "words". A graph should stand on its own without explanatory text. However, "ten thousand words" are not always better than a "thousand words". Don't put too much in a graph. Once people learn annotate, they often want to use it all the time and the message of the graph is often is lost in excess information or frills.

For example, the emphasis for the map is the state of Ohio. If I only included Ohio, I would miss the areas in other states that a sales representative would consider "nearby" and very important. On the other hand, if I included all the surrounding states, the map would be viewed as a regional map and Ohio would be just a section of that map. You should also have others review the graph to see if they "see" the same words. Many times it is better to have someone who doesn't anything about the data to review the graph.

SAS/GRAPH® is a registered trademark of SAS Institute Inc. in the USA and other countries. © indicates USA registration.

Authors:
Deb Cassidy
Senior SAS Programmer/Analyst
Cardinal Health, Inc.
5555 Glendon Court
Dublin, OH 43016
(614) 717-7136
Dcassidy@Cardhealth.com

Lori D. Griffin
Senior Applications Analyst
Hoechst Marion Roussel, Inc.
10400 Hickman Mills Dr.
Kansas City, Missouri 64137
(816) 966-7044
USMARNZ7@lbmmail.com
Graph Code - Lori's drug graph

***** SITE SPECIFIC SECTION - This section will depend on your operating system and output device(s). This code was done under CMS to a post-script printer; this section was completely omitted under SAS 6.11 for Windows to a post-script printer.

%LET PGM=SUGI21; /* NAME OF THE SAS PROGRAM */

options noxwait;
%let color_black;
%let fttext=zapf;
%let btext=zapfb;

FILENAME PSFILE "&PGM..PS";
GOPTIONS DEVICE=PS GSFNAME=PSFILE GSFLEN=132 gsfmode=replace
   GACCESS=SASGEDIT GPLOG="2S210D"X;
GOPTIONS ROTATE=LANDSCAPE ROTATE;

***** END OF SITE SPECIFIC SECTION;

%let color_black;
%let fttext=zapf;
%let btext=zapfb;

goptions hpos=80 vpos=32 ftext=&fttext ctext=&color;

proc format;
   value doselog
      1='25'
      2='50'
      3='100'
      4='150'
      5='200'
      6='300'
      7='Comparator';

data hilo;
   input sort doselog value n;
   cards;
1 7 39 368
1 7 45 .
1 7 51 .
2 1 18 74
2 1 32 .
2 1 45 .
2 2 23 186
2 2 37 .
2 2 45 .
2 3 38 163
2 3 45 .
2 3 54 .
2 4 45 594
2 4 49 .
2 4 53 .
2 5 38 456
2 5 43 .
2 5 48 .
2 6 34 79
Hands-on Workshops

```sas
proc sort data=hilo;
  by dose log;
data anno; set hilo;
  by dose log;
xsys='2';
ysys='2';
length text $16. line 3. color $8. ;
color='BLUE';
function='label'; x=6.5; y=14.5; text='Historical'; output;
function='label'; x=6.5; y=9.5; text='Control Rate'; output;
xsys='3';
function='label'; x=97; y=12; text='11%'; output;
function='label'; x=92.8; y=49; size=.6; text='Comparator'; output;
function='label'; x=7; y=7; position='0'; size=.7; style='special'; text='0'; output;
style='ftext'; position='5';
function='label'; x=93.1; y=46; size=.6; text='95% Confidence'; output;
function='label'; x=93.1; y=43; size=.6; text='Interval'; output;
/* automatically appends the sample size */
  if first.dose log then do;
    if dose log = 7 then do;
      function='label'; x=52.0; y=40; position='5'; size=.6; text='n='; output;
      function='label'; x=7; y=39; position='0'; size=.6; text='left(n)'; output;
    end;
  else do;
    xsys='2';
    function='label'; x=dose log; y=value-3; position='5'; size=.6; text='n='; output;
    function='label'; x=7; y=39; position='0'; size=.6; text='left(n)'; output;
  end;
end;
/* this is the highlighted area, which could also be automated but 
has not been in this example */
xsys='2'; position='5';
function='move'; x=7; y=39; output; ** \ *
function='draw'; x=7.2; y=45; line=1; output; ** > *
function='draw'; x=7; y=51; line=1; output; ** / *
function='move'; x=1; y=39; output;
function='draw'; x=7; y=39; line=20; output;
function='move'; x=1; y=45; output;
function='draw'; x=7; y=45; line=20; output;
function='move'; x=1; y=51; output;
function='draw'; x=7; y=51; line=20; output;
function='move'; x=1; y=39; output;
function='bar'; x=7; y=51; style='solid'; l=3; color='YELLOW'; output;
title1 h=1.1 f=ftext J=L "Figure 1" J=C
  "DrugA vs Comparator" j=C a=-90 ' ";
title2 h=1.1 f=ftext "% Complete Response by Dose";
footnote1 h=.6 j=1 f=special 'O' f=ftext
  "Comparator includes patients from DrugB and DrugC";
footnote2 h=.5 J=L F=ftext "Note: This is fictitious data. ";
symbol1 i=hiloct COLOR=&color;
```
symbol2 i=hiloctj COLOR=&color;

proc gplot data=hilo anno=anno;
format doselog doselog.;

plot value*doselog=sort / haxis=axis1
vaxis=axis2
name='graph1'
vref=11
lv=2
nolegend;

axis2 order=(0 to 80 by 10)
minor=none
label = (f=&btext a=90 h=1 "% Complete Response");

axis1 offset=(3pct, 10pct)
minor=none
value=(h=.9)
order=(1 to 7 by 1)
label = (f=&abtext h=1 'Dose (mg)');
run;
quit;

Graph Code - Deb's map

*** Fake data is used for the SUGI presentation. To use this example in your situation, you will need the create a dataset with a count by state & county. SAS provides a function to convert state abbreviations to the state FIPS code. You can manipulate the map datasets to convert a county name to a county code. If your original data is only available by zip code rather than county, you can purchase data from the post office and other sources to do the conversion. SAS Institute used to include this data as part of the maps datasets but it was difficult to keep the dataset up-to-date. ;

data custct;
  infile cards;
  input state county count;
  cards;
  39   5   1
  39  13   2
  39  17   2
  39  25   1
  39  29   1
  39  35  12
  39  39   2
  39  41   6
  39  49  17
  39  55   2
  39  57   1
  39  61  14
  39  63   2
  39  81   3
  39  87   1
  39  93   3
  39  95   4
  39  97   4
  39  99   3
  39 103   2
  39 113   7

631
Hands-on Workshops

**NOTE:** This code could be written in a macro to process several states.

1) Create an annotate dataset from the USCITY dataset. The USCITY dataset has X and Y values for the US map so new X and Y values must be created based on the longitude and latitude values.

Only include cities in the states and longitude/latitude to be graphed.

Only include cities with a population over 100,000.

This dataset will need to be projected along with the map. Create dummy state and county variables for that step.

2) Create a dataset from COUNTY for the selected states.

3) Combine the cities and counties datasets and project them. Restrict latitude and longitude.

4) Adjust the X and Y values to separate states.

5) Output the cities and counties to separate datasets.

---

632
Hands-on Workshops

6) Create choro map.
   * Use ALL option to print the empty counties.
   * Use DISCRETE option with format to group counts into proper patterns.

*******************************************************************************
data cities;
   set maps.uscity(drop=x y);
   if state in (39,42,18,54,26,21);
   if pop>100;
   text=city;
   function='label';
   position='5';
   xsys='2';
   ysys='2';
   size=.5;
   when='A';
   x=long*arcos(-1)/180;
   y=lat*arcos(-1)/180;
   if 77.5<=long<=86.5 and 37.5<=lat<=42.5;
   state=101;
   county=999;
   keep text function position xsys ysys size when x y state county;
run;

data region; set maps.county /* was counties - density??*/
   if state in (39 s 42,18,54,26,21);
   *if density!=
run;

data regcity; set cities region;
run;

proc gproject data=regcity out=projmap
   longmin=77.5
   longmax=86.5
   latmin=37.5
   latmax=42.5;
   id state county;
run;

proc format;
   value count
      1-5='1-5'
      5-1.0''='5-10'
      11-high=-'11+';

data projmap;
   set projmap;
   if state=18 then x=x-.005; /* indiana*/
   else if state=26 then y=y+.005; /* michigan*/
   else if state=42 then x=x+.005; /* pa*/
   else if state=21 then do;
      x=x-.0025;
      y=y+.005; /* ky - separate from wv*/
   end;

633
else if state=54 then do;
  x=x+.0025;
  y=y-.005; /* wv - separate from ky*/
end;
run;

data citiesp regp;
  set projmap;
  if state>100 then output citiesp;
  else output regp;
run;

pattern1 value=solid c=yellow;
pattern2 value=solid c=cyan;
pattern3 value=solid c=red;

options ftext=duplex;
proc gmap map=regp data=custct anno=citiesp all;
  choro count / discrete
c  empty=black
c  coutline=black
c  legend=legend1;
c  id state county;
c  format count count.;
c  legend1 label=('# Potential Customers');
c  title1 h=2 'ACME Analysts, Inc.';
c  title2 h=2 'Areas for Potential Customers';
c  title3 h=1.5 'OHIO AND VICINITY';
c  footnote j=1 h=.5 'NOTE: THIS IS NOT REAL DATA';
run;
Figure 1

DrugA vs Comparator

% Complete Response by Dose

 Comparator includes patients from DrugB and DrugC

Note: This is fictitious data.
ACME ANALYSTS, INC.
AREAS FOR POTENTIAL CUSTOMERS
OHIO AND VICINITY

NOTE: THIS IS NOT REAL DATA