Simplifying SAS/GRAPH® Production Using Utility Macros

Donald W. Peterson III, Progressive Software Computing Inc, Wilmington, DE
Lothar T. Tremmel, Amgen, Thousand Oaks, CA

ABSTRACT

In an effort to make graph production more efficient, we developed several 'utility macros' to take care of some frequent problems associated with creating a graph: The macro %GROPTION frees the user from the burden of figuring out the appropriate GOPTION statements to set up SAS/GRAPH®. Other utility macros are useful to preprocess data so as to facilitate their graphical representation: %GRJITTR resolves the problem of overlapping symbols by doing appropriate data manipulations before the GPLOT call; %GRCONFID summarizes data into means and confidence intervals in a convenient format so they can be represented nicely with GPLOT in conjunction with a particular SYMBOL statement.

INTRODUCTION

For occasional SAS/GRAPH users, the production of one presentation quality graph can easily take a whole day. Working at a CRO, we felt the urgent need to shorten this time. Efforts to develop fully automated macros to deliver graphs with a push of a button were not successful, mainly by the following reason: Macros deliver standard graphs which still need to be adjusted to specific needs. Different from data sets delivered by number-crunching macros, graphs can neither be edited nor post-processed.

However, we still found it possible to make graph production more efficient by simplifying certain time consuming routine tasks associated with graph production. Such tasks comprise settling up the appropriate GOPTIONS statements to obtain certain desired graph properties and output formats. This task can be extremely burdensome for unsophisticated SAS users; most of us know how tricky it can be to find the correct GOPTIONS for certain layout properties such as desired margin widths, or to figure out a good Word Perfect compatible driver. Once these things are figured out for a particular platform, the solution should be preserved and made widely available. One way is to write a macro such as %GROPTION which sets up SAS/GRAPH® based on a user-friendly, installation-unspecific parametrization.

Other tasks associated with graph production comprise the transformation of data into a format more appropriate for plotting. Operations like "Jittering" of data points to resolve overlapping plotting points are cumbersome enough so they are rarely done even when their use is indicated; the macro %GRJITTR reduces this task to a simple macro call. Whenever summary statistics are to be plotted rather than the original data, the summarization step can also be automated, as we will demonstrate with the macro %GRCONFID.

THE %GROPTION MACRO

The %GROPTION macro is designed to facilitate the generation of headings to graphical output, as well as the generation of the GOPTION statement. Graphical headings usually contain the same format, but the actual text differs. The %GROPTION macro allows for these changes by having the user provide them as parameters to the macro. These may include company name, the project, the system date and time, and the type of report (Clinical/Statistical Report). The GOPTION statement is used prior to graphics generation to set margins, filename, device driver and paper orientation (landscape or portrait). The macro eliminates the need to know the syntax required to generate the headings and margins, etc. for all graphical output. This feature also maintains consistency among displays.
THE %GRJITTR MACRO

Overlapping plotting symbols are a common problem in x-y plots. As a solution, Cleveland (1985) proposed to randomly jitter data points. The %GRJITTR macro implements this idea, as well as other methods of non-random shifting of points.

Input:

<table>
<thead>
<tr>
<th>Parent</th>
<th>Description and Valid Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>data</td>
<td>Name of input data set</td>
</tr>
<tr>
<td>X</td>
<td>Name of x variable coordinate</td>
</tr>
<tr>
<td>Y</td>
<td>Name of y variable coordinate</td>
</tr>
<tr>
<td>Z</td>
<td>Name of group variable</td>
</tr>
<tr>
<td>xjitter</td>
<td>The jittered x and y coordinates, resp.</td>
</tr>
<tr>
<td>yjitter</td>
<td>Defaults: x, y, z.</td>
</tr>
<tr>
<td>shiftx</td>
<td>Shift size in x and y direction, as an inverse fraction of the axis range. Defaults: 200 for shiftx (meaning 1/200 of x axis range), 150 for shifty</td>
</tr>
<tr>
<td>shifty</td>
<td>Options</td>
</tr>
<tr>
<td>debug</td>
<td>YES for diagnostic output</td>
</tr>
<tr>
<td>out</td>
<td>Name of output dataset. Default is the name of the input dataset.</td>
</tr>
</tbody>
</table>

Output: The macro sets up SASGRAPH by generating and executing appropriate TITLE and GOPTION statements. The project, pgm, titles, company, drug, and protocol parameters are used to display consistent titles on the output graph. The target option is to select printer, screen or WordPerfect as output device. The %GROPTION macro is used in conjunction with other macros and the actual plot code to generate output.

If TARGET is not equal to S (Screen), the macro sets up SAS graph to generate a GSF file, using &PGM to construct a file name:

```
TARGET PATH
IBM or HP path/&PGM.gsf
WP    path/&PGM.gwp
```

Importing graphs into WordPerfect: If TARGET=WP is selected, the HP7475A driver will be used as the default driver because in the past this one has proved to be a 'robust' driver leading to acceptable results in almost all cases when importing graphs into WordPerfect. However, based on experimentations by Griffin and ourselves, the following drivers may lead to better results in special cases: HPGL, CGMWPA, CGMWP, CGMWPL, CGM. Therefore, the user can use the driver= parameter to overwrite the default.

```plaintext
Output: If no output data set is specified, the macro appends to the input dataset two new variables &XJITTER and &YJITTER containing the jittered x and y coordinates. If an output dataset is specified, the macro generates a dataset &OUT which contains all the variables of the input dataset.
```
Information Visualization

dataset, plus the two new variables. If DEBUG=YES is specified, a comprising list output is produced as well, including a full account of original and changed x and y values for all overlapping data points.

Algorithm for random jittering: The extreme x values XMIN and XMAX and the extreme y values YMIN and YMAX are determined, based only on data points with both coordinates nonmissing. The maximum shift amounts SX and SY are determined as follows:

\[ SX = (X_{\text{MAX}} - X_{\text{MIN}}) / \text{shiftx} \]
\[ SY = (Y_{\text{MAX}} - Y_{\text{MIN}}) / \text{shifty} \]

A shift angle Z_ANGLE is determined randomly, based on the uniform distribution (every angle is equally likely). The jittered coordinates are calculated as follows:

\[ X_{\text{jitter}} = X + \cos(Z_{\text{angle}}) \times SX \]
\[ Y_{\text{jitter}} = Y + \sin(Z_{\text{angle}}) \times SY \]

By this method, a shift in each direction is equally likely, and both the minimum shift distance and the maximum shift amount in x and y direction are defined.

Within each group of points with the same x and y coordinates, the first point is not jittered. Hence, the coordinates of points with unique coordinates remain untouched.

Cautions: Points are only considered overlapping if their x and y values are exactly the same. Hence, it may happen that a visually exact overlap is not resolved by jittering because numerically the values are slightly different. Therefore, appropriate rounding of x and y is recommended before the macro is called. (A future version of GRJITTR could consider points as overlapping if they are visually overlapping, based on a fraction of the axis ranges.)
The %GRCONFID macro was designed to generate a data set with three Y variables for each X: Mean, lower and upper confidence interval plot.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>data</td>
<td>Input data set</td>
</tr>
<tr>
<td>out</td>
<td>Output data set</td>
</tr>
<tr>
<td>group</td>
<td>Group variable (e.g., treatment)</td>
</tr>
<tr>
<td>x</td>
<td>X variable</td>
</tr>
<tr>
<td>y</td>
<td>Y variable</td>
</tr>
<tr>
<td>shift</td>
<td>Horizontal shift to avoid overlap of curves for different treatments (default=0.2)</td>
</tr>
<tr>
<td>percent</td>
<td>100 times coverage probability of confidence interval (default=95)</td>
</tr>
</tbody>
</table>

Output: Global macro variables for plot statement:

&Z_XX = Name of x variable  
&Z_YY = Name of y variable  
&Z_GROUP = Name of classification variable

If no output data set is specified, the macro appends to the input data set three new variables &Z_XX, &Z_YY and &Z_GROUP. &Z_YY contains the three Y values for each X values: The mean, and the two confidence boundaries.

Usage Notes: The output data can be plotted with:

    PLOT &Z_YY*&Z_XX=&Z_GROUP....

If symbol statements with i=HILOBJ are used, this produces graphs with connected means and with boxes to indicate the confidence intervals. HILO specifies a solid vertical line connecting the minimum and maximum Y values for each X value. BJ indicates to connect maximum and minimum values with a bar and joins the means or close values.
CONCLUSIONS

Even when a complex task such as graph production cannot be fully automated, certain subtasks can still be simplified. Apart from speed gains, quality gains are also obtained by user-friendly implementations of graphing methods like jittering which otherwise would scarcely be used in practice because they are too cumbersome. Supplying an automated setup procedure for the graphics system has additional advantages: Different graphs follow the same layout (same title fonts, sizes, justifications etc.; same header structure), and the actual graphic generating programs become portable because the code becomes platform independent.

REFERENCES


Griffin, Lori: "Importing SAS Graphic files into WordPerfect" Marion Merrel Dow Inc., Kansas City, MO (Unpublished manuscript)

ACKNOWLEDGEMENTS

The authors are indebted to Daphne Ewing for reviewing earlier versions of the manuscript, and to Cesar Kuroki for his ideas which initiated our efforts.

SAS® and SAS/GRAPH® are registered trademarks of SAS Institute, INC. in the USA and other countries.

Authors' addresses:

Donald W. Peterson III, 110 Yeakel Ave, Erdenheim, PA 19038. Tel: (610) 832-3020 via PA relay center (1-800-654-5988). E-Mail: donald_peterson_1@sbphrd.com

Lothar T. Tremmel, Amgen, 1840 DeHaviland Drive, Mail stop 172-2C, Thousand Oaks, CA 91320-1799. Tel: (805) 447-9385. E-Mail: ltremmel@amgen.com