AN INTRODUCTION TO SAS/GRAPH SOFTWARE FOR NEW OR NERVOUS USERS

NEIL HOWARD, MARriott CORPORATION

ABSTRACT

How would you like to make graphics an integral part of your everyday reporting? A picture is still worth a thousand words and tables, and SAS/GRAPH allows user-friendly creation of presentation and information graphics. This tutorial illustrates simple, logical processes to move from number crunching tabular reports to emphatic, concise graphical displays with no pain! It will address the basic details of SAS/GRAPH that will enable the user to make a quick transition from base product reporting facilities to graphics procedures. Material from the Color Graphics I course taught by the SAS Institute will be presented throughout the body of this paper.

INTRODUCTION

Flipping through the SAS/GRAPH Users Guide gives you a vivid indication of the exotic, brilliant and magical things you can produce with this product. Closer scrutiny of the code hints at a degree of difficulty that clouds the new user's ability to visualize his problem and data in those forms.

The missing link or bridge into the graphics world comes with understanding that the SAS/GRAPH procedures are just an extension of the procedure set you already know. Additional procedure options and keyword statements exist for greater picture enhancement and control. Data preparation techniques remain in the familiar DATA step environment. The key is starting simple and working up.

Recommended supplemental documentation and instruction are available. The SAS/GRAPH Users Guide, the SAS/GRAPH Guide to Hardware Interfaces, and Technical Report P-164: Additional SAS/GRAPH Hardware Interfaces are excellent resources. Especially recommended are the Color Graphics I and Color Graphics II courses offered by the SAS Institute, Inc. The first course covers the fundamentals of SAS/GRAPH, illustrating the capabilities of each procedure and how to use the options to achieve the desired results. The second course focuses on customized graphics, both customizing procedure output and producing stand-alone custom graphics, using the Annotate facility. The texts for these courses are the best learning and teaching tools. These SAS Views (course notes and text) are an excellent entry into the details of SAS/GRAPH and the approach new users should take.

In reviewing the Proceedings from past SUGI (SAS Users Group International) conferences, I was amazed to find few papers addressing the new or nervous SAS/GRAPH user. This tutorial will open up the possibilities for using SAS/GRAPH for basic reporting and beyond, and provides a synopsis of basic material from the Color Graphics I course.

SETUP

The Color Graphics I course bases its examples on information about an office products company with two divisions: a manufacturing division with 2 production facilities, and a distribution division consisting of 6 distribution centers. Data is stored in 5 SAS data sets representing manufacturing production, manufacturing inventory, distribution inventory, and orders shipped and orders received at each distribution center. This data is also incorporated into the examples in this tutorial.

TEXT

Titles and Footnotes are available with all the Graphics procedures to place text on the picture. The number of options available for these statements makes them extremely flexible in terms of placement, readability, and emphasis. Titles and Footnotes are numbered; you can use up to ten. You can also control the height of the letters, fonts, and colors with the H=, F=, and C= options. The colors are limited to those available on your particular device type. The fonts are listed in the SAS/GRAPH Users Guide. Letter height may be specified in screen cells, inches, centimeters or percent of the screen. Justification can be controlled with the J= option values of L or LEFT,
R or RIGHT, and C or CENTER. You will see various titles and footnotes appearing in the figures, although the statements have not been included in the code.

**PLOTTING**

The office products company needs to look at the inventory of desk phones in New York for each month in 1983. We must first create a subset of the distribution inventory data set for SITE=1, product number=1142111 and year=1983. Figure 1 illustrates the same information displayed with PROC PRINT, then with GGPLLOT, the SAS/GRAPH procedure for graphing one variable against another, using the following code:

```
PROC PRINT DATA=SUBSET;
  VAR MONTH QUANTITY;
RUN;
PROC GPLOT DATA=SUBSET;
  PLOT QUANTITY * MONTH;
RUN;
```

The two procedures are equally simple to use. The GPLOT offers more opportunity for enhancements.

The basic syntax for GPLOT is:

```
PROC GPLOT options;
  PLOT yvariable*xvariable/options;
RUN;
```

where:

- `yvariable` = vertical axis
- `xvariable` = horizontal axis.

The PROC statement options include DATA= for specifying the input data set. PLOT statement options allow for control of the color, size, appearance of the physical picture. Figure 2 indicates the use of the keyword SYMBOL statement that is used with PROC GPLOT, providing further enhancement, in this case, by connecting the points with an interpolation technique of JOIN and by using an "I" from the SPECIAL font .3" high to represent the points on the plot:

```
PROC GPLOT DATA=SUBSET;
  PLOT QUANTITY * MONTH;
  SYMBOL V=I F=SPECIAL H=.3 IN I=JOIN;
RUN;
```

**CHARTING**

Another requirement is to produce a report showing the number of orders received at each distribution site. This requires some simple statistics off the "orders received" data set, which has 3645 observations. You can count by site with:

```
PROC FREQ DATA=SASDATA.ORDREC;
  TABLES SITENAME;
RUN;
```

and get a precise report with the frequency count for each site, the cumulative frequency, the percent of total and the cumulative percent. Alternatively, you can try a vertical bar chart. The basic syntax is:

```
PROC GCHART options;
  VBAR variable / options;
RUN;
```

where the VBAR indicates the physical form the chart will take, in this case vertical bars. The "chart" variable (character or numeric) indicates how many vertical bars will appear: one per value of the chart variable. Options of the VBAR statement allow control of such things as the axes, the statistic used for determining what the height of the bar represents, and how colors and fill patterns will be used in the chart.

The GCHART invocation (with no options) shown in Figure 3 (with the PROC FREQ output) presents the same basic information as the frequency distribution using the same number of lines of code, giving the user an excellent alternative to table presentations. PROC GCHART allows for vertical and horizontal charts, block charts, pie charts and star charts. There is a keyword statement for each type with appropriate options.

The charting procedure handles both character and numeric input, but user beware. Suppose we need to look at the number of orders shipped from each distribution site. A frequency distribution can easily tally for the numeric variable SITE in the "orders shipped" data set:

```
PROC FREQ DATA=ORDSHIP;
  TABLES SITE;
RUN;
```
But it should be noted that GCART treats all numeric chart variables as continuous, and will attempt to determine reasonable midpoints at which to construct bars. It is necessary to use the DISCRETE option to tell GCART to construct a bar for each unique value of a chart variable that is numeric but categorical.

The meaning of the height or length of a bar can be changed by using the SU= and TYPE= options which allow the user to select a statistic to be used in calculating the height and indicate which variable is to be used in the calculation. Figure 5 illustrates the PROC FREQ output that would produce the percent of orders received from each site; and the GCART with the change to the default statistic from frequency to percent to achieve the same results. The following code was used:

```latex
PROC FREQ DATA=ORDREC;
   TABLES SITENAME;
RUN;

PROC GCART DATA=ORDREC;
   VAR SITENAME / TYPE=PERCENT;
RUN;
```

Further, the resulting horizontal bar chart illustrates that we're summing over AMOUNT for each SITENAME (each bar). Note the change in the keyword statement to produce a horizontal bar chart:

```latex
PROC GCART DATA=ORDREC;
   HBAR SITENAME / SUNVAR=AMOUNT;
RUN;
```

The mean amount for orders received at each distribution center can easily be calculated with PROC MEANS:

```latex
PROC MEANS DATA=ORDREC MEAN;
   BY SITENAME;
   VAR AMOUNT;
RUN;
```

Changing the TYPE= variable to MEAN in the options of the HBAR alters the representation of the bars in Figure 6. Looking at manufacturing production data set, we want to display the average production levels for two different products, and then compare the average production levels of the same two products for 1983, 1984, and 1985. Run on a subset of the data, the following code gives us numbers:

```latex
PROC MEANS DATA=NPROD MEAN;
   BY YEAR;
   VAR PRODNO;
RUN;
```

GCHART can effectively convey this information in the form of block charts. We can see the data graphically with:

```latex
PROC GCHART DATA=NPROD;
   BLOCK PRODNO / TYPE=MEAN SUNVAR=QUANTITY;
RUN;

PROC GCHART DATA=NPROD;
   BLOCK YEAR / TYPE=MEAN SUNVAR=QUANTITY
   GROUP=PRODNO DISCRETE;
RUN;
```

Notice with a chart variable of year (a numeric variable), the discrete option is used. The "side by side" effect with product number is generated using the GROUP= option; in this case, it is illustrating the equivalent of the BY statement in the PROC MEANS.

Circular charts are a very popular form of graphic presentation. In comparing the orders received at each site in 1985, a PROC FREQ run on a subset of the data would produce predictable tabular output. Using the PIE keyword statement in following code:

```latex
PROC GCHART DATA=SUBSET;
   PIE SITENAME;
RUN;
```

show pie slices proportional to the frequency distribution for each sitename. That Seattle received the most orders is instantly visible. Figure 8 also depicts the average inventory levels for desks in a pie chart where patterns have been used:

```latex
PROC GCHART DATA=SUBSET2;
   PIE SITENAME / SUNVAR=QUANTITY TYPE=MEAN;
   PATTER1 V=PS C=CYAN;
   PATTER2 V=PE C=RED;
   PATTER3 V=P2X C=BLACK;
   PATTER4 V=P2 C=GREEN;
   PATTER5 V=P260 C=RED;
   PATTER6 V=Z9135 C=BLACK;
RUN;
```
The syntax for the pattern stateleDt has not been covered, and it should be noted that pattern requests do vary between procedures. Be sure to look them up.

Although the colors cannot be seen in this publication, begin to think about how colors and patterns can be used purposefully to highlight a region, emphasize a high or low quota, or just draw the eye to a specific point.

Figure 9 illustrates the use of the ACROSS= option, again a technique producing a BY statement effect. Here, the average production rates for four products are shown for sites 90 and 91:

```
PROC GCHART DATA=MPPROD;
PIE MPPROD / SUBVAR=QUANTITY TYPE=MEAN
   GROUP=SITE ACROSS=2;
RUN;
```

**MAPPING**

Although mapping could be its own tutorial, Figure 10 geographically displays the orders received from each state, as calculated by PROC MEANS with the results stored in a SAS data set -- the response data set. Any given map requires two data sets -- a map data set with the coordinates of the areas in which to show response, and a response data set which contains the response value for each area. PROC GMAP links these two data sets, and using default or user-defined patterns and colors, compares the level of response by state (or whatever the "map" variable is):

```
PROC GMAP DATA=STATS MAP=US;
   ID STATE;
   CHORO TUTORIAL/DISCRETE;
RUN;
```

More statements and options were used to generate the map in Figure 10, but this gives an idea of the structure of the procedure call. This provides a more obvious visual draw to states with the highest number of orders than perhaps a PROC PRINT listing the state codes and corresponding count of orders. The SAS/GRAPH Users Guide provides extensive information on mapping.

**CONCLUSIONS**

This tutorial has presented a brief comparison of selected base product reporting procedures with some corresponding graphics. What is presented here is by no means conclusive or comprehensive. But it is hoped that SAS/GRAPH has been presented in a friendly way, providing a comfort level for integrating its use into your SAS routine.

**PARTING ADVICE**

1. Keep things simple.
2. Graphics procedures follow the same basic structure as base SAS procedures.
3. Investigate your graphics environment. Check out the OPTIONS statement. Try PROC GTESTIT.
4. Evaluate your data, the intended audience, and your message when choosing procedures.
5. Choose the appropriate procedure.
6. Pre-process your data if necessary.
7. Run your data through a basic, vanilla invocation of the graphics procedure with no fancy options, evaluating the results.
8. Enhance your pictures. You have a taste of the kinds of options that exist.
9. Use the Users Guide index and examples for details of the options and statements touched on in this tutorial.
10. Have fun!!

**ACKNOWLEDGMENTS**

The author would like to thank the staff of SAS Consulting Services, Inc., for their support with the resources for this paper.

**BIBLIOGRAPHY**

PROC PRINT

MONTH QUANTITY
1  794
2  757
3  778
4  746
5  731
6  781
7  797
8  771
9  772
10  765
11  703

Figure 1

PROC GPLOT

Desk Phones in Stock, New York, 1983

Figure 2

PROC GPLOT with options

Desk Phones in Stock, New York, 1983

Figure 3

PROC FRQ

SITENAME FREQUENCY
Atlanta  178
Chicago  860
Dallas  367
Denver  692
New York  511
Seattle  1037

PROC GCHART

Number of Orders Received at Each Site
**PROC FREQ**

<table>
<thead>
<tr>
<th>SITE</th>
<th>FREQUENCY</th>
<th>PERCENT</th>
<th>CUMULATIVE FREQUENCY</th>
<th>CUMULATIVE PERCENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>512</td>
<td>14.3</td>
<td>512</td>
<td>14.3</td>
</tr>
<tr>
<td>2</td>
<td>191</td>
<td>5.3</td>
<td>703</td>
<td>19.7</td>
</tr>
<tr>
<td>3</td>
<td>1028</td>
<td>28.8</td>
<td>1731</td>
<td>48.4</td>
</tr>
<tr>
<td>4</td>
<td>319</td>
<td>8.9</td>
<td>2050</td>
<td>57.4</td>
</tr>
<tr>
<td>5</td>
<td>829</td>
<td>23.2</td>
<td>2879</td>
<td>80.6</td>
</tr>
<tr>
<td>6</td>
<td>684</td>
<td>19.4</td>
<td>3573</td>
<td>100.0</td>
</tr>
</tbody>
</table>

**PROC GCHART -- numeric variable**

**PROC GCHART -- DISCRETE option**

**Figure 4**

**PROC MEANS**

<table>
<thead>
<tr>
<th>SITE NAME</th>
<th>FREQUENCY</th>
<th>PERCENT</th>
<th>CUMULATIVE FREQUENCY</th>
<th>CUMULATIVE PERCENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atlanta</td>
<td>178</td>
<td>4.9</td>
<td>178</td>
<td>4.9</td>
</tr>
<tr>
<td>Chicago</td>
<td>860</td>
<td>23.6</td>
<td>1038</td>
<td>28.5</td>
</tr>
<tr>
<td>Dallas</td>
<td>367</td>
<td>10.1</td>
<td>1405</td>
<td>38.5</td>
</tr>
<tr>
<td>Denver</td>
<td>692</td>
<td>19.0</td>
<td>2097</td>
<td>57.5</td>
</tr>
<tr>
<td>New York</td>
<td>511</td>
<td>14.0</td>
<td>2608</td>
<td>71.6</td>
</tr>
<tr>
<td>Seattle</td>
<td>1037</td>
<td>28.4</td>
<td>3645</td>
<td>100.0</td>
</tr>
</tbody>
</table>

**Vertical Bar -- TYPE=PERCENT**

**Horizontal Bar -- TYPE=SUM**

**Figure 5**

152
PROC MEANS by SITENAME

VARIABLE MEAN
--- SITENAME=Atlanta ----
AMOUNT 1930.73
--- SITENAME=Chicago ---
AMOUNT 9983.98
--- SITENAME=Dallas ---
AMOUNT 3996.62
--- SITENAME=Denver ---
AMOUNT 8013.73
--- SITENAME=New York ----
AMOUNT 6046.30
--- SITENAME=Seattle ----
AMOUNT 11976.41

PROC MEANS
VARIABLE MEAN
----- PRODNO=763A ----- QUANTITY 301.12
----- PRODNO=8492 ----- QUANTITY 363.67

PROC MEANS by YEAR

YEAR PRODNO MEAN
1983 Receivers 287.683
1983 Bases 383.538
1984 Receivers 301.123
1984 Bases 363.670
1985 Receivers 265.130
1985 Bases 388.570

BLOCK CHART

Mean Production Levels for Bases vs. Receivers

Figure 7

153