This graphics tutorial does not lend itself well to a paper, so this will be a summary of some of the slides used.

I will be demonstrating some of the features of the PROC GPLOT and PROC GCHART procedures in SAS/GRAPH. Before I get started, I would like to repeat some of the suggestions I presented last year on graphics in general.

Editorial Section

Good Graphics Should:
- Show the data clearly
- Display your conclusions
- Avoid distorting the truth
- Avoid ornamentation
- Be interesting and aesthetic

As you use these PROCs and any other graphics keep these rules in mind to obtain graphics that are clear and understandable.

Some Don'ts:
- Don't let software defaults determine the final output.
- Don't get carried away with fonts, patterns, colors or 3d effects.
- Don't forget what you're trying to show.
- Don't omit or distort information.
- Don't use the wrong type of chart.

And some ...

Steps for Creating a plot:
1) Know your audience
2) Decide what you want to show
3) Make a sketch
4) Start simple
5) Add title and symbols
6) Check colors and shading for impact
7) Does the plot show what you want?
8) Check for chart junk
9) Refine and simplify if possible
10) Try the plot out on someone

These steps can improve your graphics abilities and make it easier to create great graphics.

Introduction

The GPLOT and GCHART procedures are quite powerful and flexible, so much so that it is sometimes hard to get them to do what you want. I will be discussing the commonly used options and providing step by step examples of the effect of the options on the output. I encourage you to try the examples yourself and recommend that you set up your own SAS data set and try all the options one at a time to familiarize yourself with their function.

Some features of the GPLOT and GCHART procedures:

**PROC GPLOT**

Features:
- Automatically scales the values on axes or axes and tick marks can be explicitly defined.
- Provides horizontal and vertical reference lines.
- Plots can be superimposed.
- A variety of symbols can be used to represent points.
- Colors, symbols, line styles and interpolation method can be selected.

**PROC GCHART**

Features:
- Charts numeric and character variables.
- Automatically selects intervals for continuous numeric variables or midpoints can be explicitly defined.
- Summarizes descriptive statistics, including:
  - Frequency - # of observations at each midpoint
  - Cumulative Frequency Counts
  - Percentages
  - Cumulative Percentages
  - Sums
  - Means

Enough proselytizing.
PROC GPLOT provides the standard SAS/Graph options:

**PROC GPLOT options**
- DATA=SASdataset
- GOUT=graphicscatalog
- ANNOTATE=SASdataset
- ANNO=SASdataset

The general form of a PROC GPLOT statement is:

PROC GPLOT options

This statement by itself will create a plot, but it is all the other options that create a meaningful plot. For example you have complete control over the color of reference lines and their line type (dotted/dashed):

PROC GPLOT
  PLOT yvariable*xvariable.../options
  • HREF=values
  • CHREF=color
  • LHREF=linetype
  • VREF=values
  • CVREF=color
  • LVREF=linetype
  • VAXIS Examples

You can include frames around the plot and specify the color of text, axes or the frame itself.

PROC GPLOT
  PLOT yvariable*xvariable.../options
  • NOAXES
  • FRAME
  • FR
  • CFRAME=color
  • C=color
  • GAXIS=color
  • G=color
  • GTEXT=color
  • CT=color

PROC GPLOT allows plots to be superimposed using the overlay option and provides options for specifying legends or even deleting legends.

**PROC GPLOT**

The VAXIS or HAXIS commands provide flexibility in specifying axis endpoints, tick marks, direction and labelling.

PROC GPLOT
  PLOT yvariable*xvariable.../options
  • VAXIS=values
  • HAXIS=AXISn
  • VMINOR=n
  • HMINOR=n
  • VZERO
  • HZERO
  • VREVERSE

Some examples are shown:

**VAXIS Examples**

<table>
<thead>
<tr>
<th>Specification</th>
<th>The code</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAXIS=10 to 100 by 10</td>
<td>VAX 10,20,30,40</td>
</tr>
<tr>
<td>VAXIS=10,20,30,40</td>
<td>VAX 10,20,30,40</td>
</tr>
<tr>
<td>VAXIS=10,20 to 40 by 10</td>
<td>VAX 10,20,30,40</td>
</tr>
<tr>
<td>VAXIS=10,20 to 50 by 10</td>
<td>VAX 10,20,30,40</td>
</tr>
<tr>
<td>VAXIS=10,20,30,40</td>
<td>VAX 10,20,30,40</td>
</tr>
<tr>
<td>VAXIS=10,20,30,40</td>
<td>VAX 10,20,30,40</td>
</tr>
</tbody>
</table>

1324
Other standard SAS/GRAF statements allow specification of symbols, patterns, titles, and footnotes.

PROC GLOT options
- SYMBOL options
- PATTERN options
- TITLE options
- FOOTNOTE options
- NOTE options
- BY variables

Symbols include letters and numbers and the following special symbols:

Patterns include the following or may be designed yourself.

Let's TalkWidgets
Let's look at a standard widget start-up company (ACME Widget Co.) reviewing its first nine months of sales.

Our initial data set looks like this:
Our first plot using PROC GPLOT looks like this:

Although not bad for two lines of code, there are some needed refinements. The first would be to join the points using the interpolation option of the symbol statement and add square markers for the points.

The interpolation options are:

**Miscellaneous Interpolation Option**
- I=None
- I=JOIN
- I=NEEDLE
- I=STEPxx
- I=Sxxxxx
- I=STDxxxx
- I=SPLINE
- I=SPLINEP
- I=Lx
- I=HLOxxxx
  
  xxx=T B C or J or combinations

Using I=JOIN we obtain:

It might be useful to try a linear regression through the sales figures to determine the growth in sales. The R-series interpolation provides this capability.

**R-Series Interpolation Option**
- I=Rxxxxxx

We will use I=RLCIM95 which stands for Regression Line & Confidence Limits on the Mean at 95% confidence.

This creates this plot:
Other regression options allow curve fitting using spline and cubic regression techniques. Try changing the I= parameters to get a feel for how these options work.

In order to determine if ACME Widgets will make their year end goal of 100 sales we can extend the axes using the VAXIS and HAXIS statement:

Finally we can add TITLE and FOOTNOTE statements to complete our plot.

To compare sales of two groups we can add the other data set and a classification variable (in this case G) and change the PLOT statement to the form PLOT Y * X = G. The program and data set look like this:
PROC GCHART provides a variety of graphical outputs.

**PROC GCHART**

- Horizontal bar
- Vertical bar
- Block
- Pie
- Star

The statements are basically as follows:

**PROC GCHART**

- HBAR variables... / options
- VBAR variables... / options
- BLOCK variables... / options
- PIE variables... / options
- STAR variables... / options

Many of the options are similar to the PROC GPLOT options with extensions to include options for statistical summaries and subgrouping.

These standard SAS/GRAPH statements can be used.

**PROC GCHART**

- PATTERN options
- BY variables
- AXISn options
- LEGENDn options
- TITLE options ‘text’
- FOOTNOTE options ‘text’
- NOTE options ‘text’

I will be describing primarily the HBAR and VBAR options. The common AXIS options are shown. Note the ability to sort by ascending or descending size of bars. The AXES options are the same as those for the GPLOT procedure. MIDPOINTS = option can be used to explicitly define the midpoints as a list of values.

**PROC GCHART**

- HBAR variables... / options
- or VBAR variables... / options

Options:
- ANNOTATE=SASdataset
- ASCENDING or DESCENDING
- AXIS=values
- AXISn
- RAXIS=values
- RAXISnn
- GAXIS=axisnumber
- MIDPOINTS=values

The grouping options shown provide the ability to make stacked bar charts by specifying the SUBGROUP = variable option. DISCRETE tells the system that the numeric variable is not continuous.

**PROC GCHART**

- HBAR variables... / options
- or VBAR variables... / options

Options:
- G100
- GROUP=variable
- GSPACE=number
- SUBGROUP
- SPACE=number
- LEGEND=legend number
- LEVELS=number
- DISCRETE
- MINOR=n

Other options provide for color, frame, and text selections.

**PROC GCHART**

- HBAR variables... / options
- or VBAR variables... / options

Options:
- CAXIS=color
- CFigure=color
- CFRAME=color
- CFigure COLOR=color
- CTEXT=color
- FRAME FR
- PATTERNID=method
The main power of the Proc GCHART procedure lies in its ability to automatically summarize data. The procedure can automatically determine intervals for midpoints of continuous numeric data and can summarize data by frequency, cumulative frequency, percent, cumulative percent, and by sum and mean.

The following options specify what the variables in the data set represent.

PROC GCHART
HBAR variables... / options
or VBAR variables... / options
Options:
* FREQ
* CFREQ
* PERCENT
* CPERCENT
* SUM
* MEAN
* SUMVAR
Example: FREQ=variable
(SUMVAR = must be specified for SUM + MEAN)

These options specify what types of statistical summary the bars represent.

PROC GCHART
HBAR variables... / options
or VBAR variables... / options
Options:
* TYPE=FREQ
* TYPE=CFREQ
* TYPE=PERCENT
* TYPE=CPERCENT
* TYPE=SUM
* TYPE=MEAN

In addition, these are options for getting rid of axes, statistics, symbols, legends.

PROC GCHART
HBAR variables... / options
or VBAR variables... / options
Options:
* NOAXIS
* NOAXES
* NOSTATS
* NOSYMBOL
* NOLEGEND
* NOZEROS

First attempts at using Proc GCHART can be frustrating as the procedure often provides default summarization of data without being requested. Our first attempt using our ACME Widget Co. data provides the following unusual graph.

The data was summarized as frequency of the variable Amt By Month and Month was assumed to be a continuous numeric variable so the midpoints were automatically picked for us.

A solution is to add a SUMVAR = Amt option and a DISCRETE option to tell the system that month is not a continuous variable. In addition, the midpoints = values option is used to explicitly specify the intervals to be used for month. Add a VAXIS statement and some titles and footnotes and the result looks like this:
By splitting up the data set into the actual sales figures by widget type, we obtain the following:

DATA SALES:
INPUT: WIDGET 5 ANY WIDGET 6:
CARDS:
JAN 13 A
JAN 6 B
JAN 14 C
JAN 7 B
FEB 16 A
FEB 10 B
FEB 16 C
FEB 10 D
MAR 22 A
MAR 14 C
MAR 19 D
MAR 12 B
APR 24 A
APR 14 B
APR 12 C
APR 9 D
MAY 30 A
MAY 12 B
MAY 12 C
MAY 11 D
JUN 26 A
JUN 16 B
JUN 32 C
JUN 12 D
JUL 25 A
JUL 17 B
JUL 23 C
JUL 12 D
AUG 22 A
AUG 22 B
AUG 22 C
AUG 16 D
SEP 22 A
SEP 22 B
SEP 22 C
SEP 22 D

A chart of this data using the SUBGROUP = variable option looks like this:

This chart shows trends in the data by showing fewer subgroups and using PATTERN statements to choose the colors and patterns (it looks better in color).

Notice the use of the FRAME option, the HREF= 100, and the use of MIDPOINT= 'JAN' 'FEB' etc. to show the months.
This chart is an example of the use of PROC GCHART to summarize data. The data set was created using the NORMAL function and the data was summarized as shown.

I hope these examples are useful in demonstrating the PROC GLOT and PROC GCHART procedures. The best way to learn these procedures is to try them out and I encourage you to experiment with the options until you get that great graph you were looking for.

No graphics topic would be complete without a 3-d chart. The block option shows the widget data in "skyscraper" chart format. I don't find this representation very clear, but it does seem to be popular. The block charts often require a change to the HPOS or VPOS options in the OPTIONS statement to fit all the information. This in turn often makes lettering very small.

Try to only use 3-d charts for 3-d data (ie. 3 variables). Two variable 3-d charts are invariably less clear than line or bar charts of the same data.