Beyond Default Colors: An Introduction to Customizing Graphic Output

Janet Lasher, University of California, San Francisco

Abstract
This paper serves as an introduction to customizing SAS/GRAPH® output. It covers the basic form of the most frequently used and flexible SAS/GRAPH statements which are used to manipulate the default graphic output with simple techniques readily achievable by the beginning SAS® programmer. It is assumed that you already know how to create default graphic output, and you are now in the position of trying to alter these graphs.

Introduction
Not all of the options available with the statements described below have been covered. These options can be found in the SAS/GRAPH manual. It is my intention to explain the general form and methods of these customizing statements, not to cover every option available. I hope that with this introduction, you will have the confidence and knowledge to open the SAS/GRAPH manual and try graphics with a minimum of stress. As with trying to learn the SAS system for the first time, in learning SAS/GRAPH syntax you may feel overwhelmed with all of the statements, procedures, device names and fonts in the manual. Try to learn some of the basic statements covered here and you will start to build a vocabulary and feel of how to go about creating graphs with your SAS system data sets.

The form of a SAS/GRAPH program is basically the same as any other SAS system program. You start with a SAS data set and run procedures. The difference is that the procedures you run create graphs. These procedures, like GLOT, have special statements to run the procedure. If you have access to a graphic output device during the SAS session, or route your output to one, your graph will be output without any additional specialized programming. In general, most procedures and statements that relate to the creation of graphics begin with the letter G, like GCHART (a procedure that creates graphic device charts similar to the CHART procedure in base SAS). In addition, your program may include setting system options, similar to the OPTIONS statement. In the case of graphics, these options are called GOPTIONS, for graphic-specific options. These procedures and statements are explained in detail in the SAS/GRAPH manual. So with the exception of a few new statements, the general form of the SAS/GRAPH program to create graphic output is the same as running a regression analysis. The following is a brief introduction to how to make your graphic output look like your own.

The first problem that you will have with creating graphs is that, although SAS/GRAPH is fairly easy to use, it still takes time. You should leave adequate time to create your graphs. As you become familiar with the possibilities of SAS/GRAPH, you will try more complicated things and will become more critical of the final output. Therefore, you will be spending time "tweaking" your graph, such as changing the size of the third title, or the color of the tick marks on the horizontal axis, and this takes time. Make sure that when you are going to work on a graphic project, you bring your manual with you. You will need the SAS/GRAPH manual as a reference for all of these great options you want to use.

One of the best ways to start a graphic project is to find an example in the manual that is similar to the graph you want to produce. This enables you to start with a program that you can modify for your needs. If you can do this you can start making presentation-quality graphic output without a thorough understanding of all of the options of GRAPH. If you can't find an example that meets your needs, you should read through the examples in the manual similar to the graphs you want to produce, to see how the statements are put together to create the graph.

GOPTIONS Statement
The GOPTIONS statement defines the system options to be set for the SAS/GRAPH session. These options are set at the invocation of a SAS session, and remain in effect until the end of the session, unless you issue another GOPTIONS statement later in the program. The default GOPTIONS set for your installation can be found by running a PROC GOPTIONS. The form of the GOPTIONS statement is the keyword GOPTIONS followed by all the options to be set for this session (e.g. GOPTIONS NOTEXT82 DEVICE=IBM3179). These options are described fully in Chapter 6 of the SAS/GRAPH manual. Some of these options include setting the device type, the default title height, and the size of the entire graph.

Since SAS Institute has changed the way that titles and other forms of text are defined, the GOPTION, NOTEXT82, should be set on for the syntax to function as described in the Version 5 SAS/GRAPH manual, and as described in this paper. If you are not running Version 5 of SAS Institute products, be aware that there are substantial differences in syntax between the 1982 version and the new Version 5 products.

The DEVICE= option will describe the type of output device you will be using, such as a graphic terminal or pen plotter. A list of valid devices can be found in Appendix 3 of the SAS/GRAPH manual. This listing of the device names includes the default color order for each device. There are additional devices that have been added to SAS/GRAPH since the printing of this manual. If necessary, these devices can be explained by your on-site consultant.

Two other useful GOPTIONS, HSIZE= and VSIZE=, provide one way for you to control the ultimate size of your graphs. These options are used to describe the size in, say, inches, to scale your graph when it is output. This can be very useful if you need a very small graph. Please note you should try to retain the same relative dimensions of the normal sheet of paper so that the graph does not become too distorted.

Text Statements
There are basically three ways to get text onto your graphic output, using the TITLE, FOOTNOTE, or, NOTE statements. Each of these statements can be controlled by using the same options, so by learning a few of the most frequently used text control options you will be starting to create a vocabulary that is used throughout the SAS/GRAPH language. These options allow you to control the color, font (what the letters and symbols look like), height, and justification of the text on your graph.

The TITLE statement allows up to ten lines of titles using the TITLEn where n is the number of the line on which the text will appear. The FOOTNOTEn statement is like the TITLE statement, with the exception that the FOOTNOTEn are pushed to the bottom of the page. There are also ten lines available for FOOTNOTE. The NOTE statement will use the space between the TITLEn and FOOTNOTEn. Please note, there
When you first start to create graphs using SAS/GRAPH, you find that mysterious colors and heights of titles appear on your graph. And you didn’t say anything about using that black pen your getting. SAS/GRAPH has a default set of colors, heights, fonts, and justification for all graphs. By automatically creating this set of defaults for graphs, SAS/GRAPH allows you to quickly create graphs. The default color list for your device is provided in Appendix 3 of the SAS/GRAPH manual. SAS/GRAPH uses the first default color for all titles on your graph; this color is generally black. The default font for all text is the COMPLEX font, a plain font with block-type letters. The height of the text is a height of 1 character unit except for the first TITLE, which is printed using a HEIGHT=2.

If you don’t want any titles, please include a TITLE statement in your program anyway. If there is no TITLE statement in your program, you will accidentally get a title that says “SAS” with a height of 2. To avoid this, include the statement TITLE; in your program. If you want to customize or change these defaults you will have to learn about some more options.

The following statements are the most often used to manipulate text. Each text statement, TITLE, FOOTNOTE, and NOTE, appears before the text that they affect. There can be multiple occurrences of each of these options which will cause the text to change as it appears across the line. Remember all text must be enclosed in quotes. If you intend to change any of the options within the same text statement, you should include an extra space enclosed within the text between each word. Without this space the text will change when the options change, but the words will not be separated properly.

If you would like to change the font (the type of letters to be printed), you can specify the font type on the text definition statement using the font option, FONT = fontname. The fontnames are listed on page 32 of the GRAPH manual. These fonts are described using the FONT= (or F=) option after the text keyword (such as TITLE), and before the text to be printed in that font. If necessary, multiple FONT= options can be used in a single text statement (e.g., TITL3 FONT= triplex "This is a title").

As a rule, fonts are proportionally spaced, meaning that each letter does not require the same amount of space. In addition, many fonts have a Uniform version that will use the same amount of space for all letters and spaces. This is quite useful if you are trying to line up text on multiple lines. To specify these fonts, a U is placed after the regular font name, as in ITALICU. In addition, some of the ‘bolder’ fonts have an Empty option that will print only the outline of the letters (the center being EMPTY). This option is invoked by adding an E suffix onto the fontname (e.g. XSWISSE).

The FONT procedure is quite a useful tool to use to preview a particular font. This procedure is described in the GRAPH manual, and if printed or plotted, can become a permanent record of what the font actually looks like. Each letter is output in the font requested with the actual character printed below in block letters.

The HEIGHT of the text can also be changed. As mentioned earlier, the first TITLE is printed in character cells of height 2, while the remainder of the text statements are printed using ‘1’ cell height. To control the size of the text the option HEIGHT= n or H= n is placed before the text. This number can be any value of cell heights, like 1.5 or 3. If you prefer to describe the height of the text you may add H= n IN to describe the text in inches, or H= n CM to specify heights in centimeters.

To control the color of the text the option COLOR = colorname is used. These colors are listed in Appendix 3 of the SAS/GRAPH manual in default order and by name (notice that black is first on most devices). If you are creating your output on a pen plotter and you are in control of the pen mounts, you may want to use the colorname in the option statement as if it were a variable, descriptive of the pen you want to mount (e.g., BLK7 for a 7mm black pen). The actual color names can be used wherever you like. In addition, you may include a COLOR=color list statement in your GOPTIONS section. This GOPTION will specify the order for the colors to be used by the SAS system as your default color order. These colors usually include RED, GREEN, BLUE, and BLACK, or whatever. If you are lucky enough to have a really fancy output device, you can use up to 256 named colors as described in Chapter 3 of the SAS/GRAPH manual.

As stated before, any of these options can be used more than one time on the same text statement. If you are creating a graph and would like to have, for example, a footnote with one of the words in the italic font and the rest of the footnote in the triplex font, you must include multiple FONT= options in your text statement. For example:

```
FOOTNOTE  height=0.5 COLOR = orange
          FONT=triplex 'Thank you'  F =italic ' -Janet';
```

This will result in the text Thank you in the triplex, and -Janet in the italic font. In addition, notice the spaces inside the quotation marks with the text -Janet. This is necessary so that the two portions of text are separated by a space in the resulting FOOTNOTE. Note, since there is only one height and color option in this statement, all text will be produced using the same color and height. This can really be useful if you need to use a special symbol from one of the symbol fonts (e.g. GREEK, MATH, or MUSIC) in your text. There are samples of these and other fonts in Appendix 1 on the SAS/GRAPH manual.

Ok, you say all of these options are really great, but I want all of my TITLE1 statements to be displayed in EMPTY, GREEN, XSWISS letters. Another way of controlling text in a more general way is to include a few options in your GOPTION statement. The options FITITLE= and CITITLE=, if placed on the GOPTIONS statement, will control the font for all TITLE1 statements and the color for all TITLE1s in this program. The GOPTIONS statement is included in your program before any G-PROC statements and looks something like:

```
GOPTIONS TITLE=xswiss CITITLE=green ;
```

This can be very helpful if you are trying to put together a graph for a journal that requires a specific style, height and color for all titles or figures submitted. By default all text lines will be centered on the printed page. This, too, can be altered with yet another option to justify the text, the J= option. There are three methods of justification, =C to center (the default), =R to move the text to the right-most edge of the graphic, and =L to move the text to the left side of the output. The right and left justification will move the text to the edge of the printable page, so if you are going to right or left justify your text you may want to include leading blanks or trailing blanks to have a small margin of space before or after the text.

Another useful option is the UNDERLIN = n option. This option will underline any text after the option is invoked.
valid values for n are 0, 1, 2, or 3, which are used to increase the thickness of the line under the text (one being the thinnest line and three being the thickest line available). To turn off the underline, reset the UNDERLIN option to its default, UNDERLIN=0.

Getting even fancier, a box can be drawn around the text of a TITLE, FOOTNOTE, or NOTE statement. Similar to the UNDERLIN option, the BOX=n option also has four possible values for n, 0, 1, 2, or 3. The BOX option is processed after the text of the statement has been output, so this option should be used only once in a single text statement. The color of this box will be the same color as the last COLOR= option. If you would like the box a specific color, or a color other than the color of the text on the statement, you should include a COLOR= statement just before the BOX= option, but after the color defining the text on the statement. For example, FOOTNOTES C=blue F=xswiss "Make a box here" C=red BOX=3; will create a red box using a thick line around the blue, empty, xswiss text.

There are other options that can help customize this BOX and are described in Chapter 4 of the GRAPH manual.

Often we include a BY statement when processing data in a basic SAS data set. For many of the SAS/GRAPH procedures the BY statement is also used. When you include a BY statement your output will include a line under your titles stating what the by-variable is. This can really cause stress, since no amount of playing with the title statements will get rid of or alter the appearance of this line of text. However, the by-variable line of text can also be customized on the OPTIONS statement. The option CBY= controls the color of the by-statement, HBY= the height of the by-statement, and FBY= the fonts to be used for the by-text line. The valid values for these options are the same as for the COLOR=, HEIGHT=, and FONT= text options. Additionally, you can force the by-text line right off your graph by stating OPTIONS HBY=0. This forces the height of the by-text to zero, so nothing will be printed under your titles.

The options described above are not the only options available that allow you to customize the text on your SAS/GRAPH output. However, they will help you to understand the basic structure of these SAS/GRAPH statements, and will get you started on the way to creating easy and customized text on your graphic output. Other options allow you to rotate lines of text, to print text at an angle, to move text to a specific location on the graph, and many others. Each of these additional options have similar uses in all of the TEXT statements.

**PATTERN Statement**

The PATTERN= statement can be used to change the patterns output in the following SAS/GRAPH procedures: GCHART, GCONTOUR, GMAP, and GPLOT. Each of these SAS/GRAPH procedures have their own idiosyncrasies associated with the use of PATTERN statements. They are described in each of the chapters describing these procedures in the manual and should be referenced as needed. Generally, the pattern statement will allow you to change the type of fill used in empty spaces created by graphs, such as the fill pattern used to fill in a bar or pie chart, or a region or state in a map. Each pattern statement can have a number associated with it that can be used to specify which portion of the graph should be filled in. The pattern statement should be placed before the description of the graph in the program, e.g., before the PLOT statement in GPLOT, or the VBAR statement in a GCHART procedure.

The PATTERN statement begins with the keyword PATTERNn. The n is optional, but is useful to help you keep track of multiple pattern statements within one procedure. The PATTERN statement, like the text statements, has an array of options which allow easy control. The COLOR= option allows you to control the color of the fill. The VALUE= option specifies the actual pattern, or hashing. Not all values for the VALUE= option are available for all procedures. For example, GCHARTs (particularly PIE charts) have their own value options as described in GCHART. Most generally the options used for these fill patterns range from Empty, to Solid with a variety of hashes that angle to the right, left, or cross.

There is a graph called the Pattern Selection Guide on page 58 of the SAS/GRAPH manual. This guide is read in the following way. Read down the far right side of this pattern graph and across the striped bars. Which letter is associated with the hashing you want (R for right upward, X for crossed, or L for left upward marks)? Then choose the density of the lines by using the numbers across the bottom of the picture. Note for the E (empty) and S (solid) options, the slant and density are not used at all. To create the VALUE= option on a pattern statement, first start with the option clause VALUE= and follow with the desired hash and the density values as read off the Guide. So if you need a fairly dense cross hatched fill in red, your PATTERN statement would look something like: PATTERN COLOR= red VALUE= x; obviously this can become more complicated, since we can define these patterns ourselves using the VALUE option as described below the graph in the manual. For most applications the patterns provided in this Pattern Selection Guide are sufficient.

The PATTERN statements can have a number associated with it like a title can, as in PATTERN1. These numbers force the use of patterns in the order of the descriptor number. When using GCHART the PATTERNs are used in conjunction with the the values of the SUBGROUP option, in GPLOT with the AREA option, and with each value to be used in creating maps. PATTERNn statements are additive. If you specify an option, say COLOR = on a PATTERN1 statement, you can reuse the PATTERN1 statement within the same procedure by omitting the option for color, and adding, say a VALUE= option. The color will remain the same as on the first PATTERN1 statement, and the new value option will now be output on your graph. To restart your numbering scheme you may either initialize all options to default by issuing a PATTERN statement and a null value, e.g., color =, or by reusing the number and issuing new values for all defined values on the previous PATTERN statement.

**SYMBOL Statement**

The SYMBOL= statement can be used anywhere in a SAS program and is used by GPLOT to define the characteristics of lines and points plotted. The SYMBOL= statement is placed before the GPLOT procedure is begun. Each symbol retains the same definition until another SYMBOL statement ending with the same number appears in the program. If the number is left off the SYMBOL statement, SYMBOL1 is assumed.

The SYMBOL statement can be used in different ways in the GPLOT procedure. The first way, using numbered SYMBOL= statements allows you to specify the symbol options for each plot, and then refer to this number in the PLOT statement. If you specify more than one plot on a PLOT statement and do not include the n-value for the SYMBOL, SAS/GRAPH will use the symbol statements starting with SYMBOL1 for the first plot mentioned, SYMBOL2 for the second plot, etc. For example:

- SYMBOL1 COLOR=blue V=diamond;
- SYMBOL2 COLOR=green V=star;
- PROC GPLOT;
- PLOT height * weight = 1 ;
This GPLOT will create one graph using the SYMBOL1 statement, plotting blue diamonds for each point on the graph. The SYMBOL2 statement may be used later in the program.

The second way the SYMBOLn statement is used is with the BY statement. As you are plotting two variables by another, the GPLOT procedure will plot the observations corresponding to the value of the BY-variable with the ordered SYMBOL statements. In other words, if you have the BY-variable, CLASS, that can have the values 3 or 7, you would define two SYMBOL statements (SYMBOL1 & SYMBOL2). The data points would be plotted using the SYMBOL1 definitions for all points with the CLASS value equal to 3, and using the SYMBOL2 definitions for the points with the CLASS value of 7. So, in this case it is important to have SYMBOL statements that correspond to each value of the BY-variable.

As with the PATTERNn statement, the SYMBOLn statements are also additive. To restart your numbering scheme and symbol options you may either initialize all options to default by issuing a SYMBOLn statement, or reuse the number issuing new values for all previously defined values as in the PATTERNn statement.

The SYMBOL statement has many other options that allow the specification of plot characters, line types, colors, and interpolation techniques. These options work in the same general way as the options described before, starting with color. The COLOR= colorname option specifies the color for the corresponding plotted points. The V= symbol option defines the symbol to be used for the point to be plotted. The valid values for this option are the letters A through W, the special symbols, and the numbers 0 through 9, as shown on page 62 of the SAS/GRAPH manual. For example, if you wanted a red plus-sign plotted, you could use this statement:

```
SYMBOL1 COLOR=RED V=PLUS;
```

As with all other statements that refer to custom text, the SYMBOL statement also has a F= fontname and H=height option to define the font and height to use when plotting the symbols. Additionally, a W=width option can be used to specify the thickness of any plotted lines. The L=number, option is used to specify the type of line to be drawn on the plot (a solid line is the default value). Examples of the available line types are shown on page 63 of the SAS/GRAPH manual.

The I=(interpolation type) option is used to specify how the data points are to be connected. The types of interpolation include: =NONE (leave the points unconnected), =JOIN (join the points with a straight line), and =NEEDLE (draw a vertical line from the x-axis up to each data point). Another possible value for the interpolation option is to specify that a regression line be calculated and drawn through the data points. This option, =Rxxxxx has values to allow you to describe which type of regression is used. The first x defines the type of regression line to be fit, either L (linear), Q (quadratic), or C (cubic). The next x is optional and can have only one value, 0. This states that the intercept is to be forced through the origin of the plot. If you do not want this option, leave the 0 from the definition. The last three x’s can take the values CLM or CLI, either to show the confidence limits for the Mean or the Individual predicted values. Finally, you may specify the confidence level by including 90, 95, or 99 after the CLM or CLI. The line style used for the confidence limits will be one greater than the L= option states, so if your regression line is to be plotted using L=10, the confidence limit lines will be plotted using L=11. Therefore a valid SYMBOL statement using the R= option could look like:

```
SYMBOL2 COLOR=blue I=rcdm95;
```