SAS/GRAPH® Software: Doing More with Less ANNOTATE
Irene Mendelson, Hoffmann-La Roche, Inc., Nutley, NJ

ABSTRACT

Annotation, an advanced technique to embellish basic graphs in SAS/GRAPH®, is at times an arcane process. It involves first of all recognizing what features in the graphic desired can be produced with the basic tools and then deciding how to translate the additional lines, text, or shapes desired into SAS instructions in the annotation dataset. In the last few years, improvements in SAS/GRAPH have provided alternatives to annotation in certain cases. This paper will touch on these improvements and outline some rules of thumb that the author has found useful in distinguishing when to use annotation and when a simpler technique will do.

INTRODUCTION

The best place to start creating a graphic is to break it down into its basic elements and compare them to the basic patterns in SAS/GRAPH. The basic graphs that can be produced include text only slides, bar charts, pie charts, line graphs, maps, three-dimensional plots, and contour plots. To create these basic graphs, you need to identify what the graph variables will be and focus on other backdrops features, such as headings and footnotes, symbol types or shading patterns, axis labels, axis text, and axis orientation.

With the basic pattern of the graphic fixed, you can move on to the extras. These extras include additional text, lines, or shapes. The exact position of the extras must be determined as well as their color and size, and, in specific situations, line type (dotted, etc.) or type of shape (bar, etc.). Some extras can be added with the addition of simple options to basic SAS code; others will require the detailed specification and activation of an annotation dataset.

Before we dive into a few graphs that require annotation, we will review the following examples which can be produced using some of the newer options or statements:

- Numbers at end of bars on bar chart
- Regression equation on line graph
- Half-bar or bar chart (lines on bar chart)
- Legend on line graph

TEXT: NUMBERS ABOVE BARS

With vertical or horizontal bar charts, people may wish to make it easier for the exact value represented by the bar length to be quickly read off the graph. The option to use here with either the VBAR or the HBAR statement is the SUM option. In the example SAS program below, the data is drawn from experiments measuring the enzyme levels for different doses of a compound called ABT (1).

Simplified SAS program code:

```sas
DATA MEANMICE;
  INPUT DRUGABT MEANP450;
  CARDS;
  0 1.297
  25 0.714
  100 0.451
RUN;

TITLE1 HEIGHT=1.6
  'LIVER P-450 CONCENTRATION VERSUS DOSE OF ABT';
FOOTNOTE1 HEIGHT=1 JUSTIFY=LEFT
  'WITH SUM OPTION ON VBAR STATEMENT';
PROC OCHART DATA=MEANMICE;
  VBAR DRUGABT / SUMVAR=MEANP450
    SUM
      DISCRETE
      RAXIS=AXIS1
      MAXAXIS=AXIS2;
    AXIS1 LABEL = (A=90 R=0 HEIGHT=1
      'MEAN P-450 CONCENTRATION')
      ORDER = 0 TO 1.7 BY .1;
    AXIS2 LABEL = (HEIGHT=1 'DOSE OF DRUG ABT');
RUN;
```

The output is shown below in figure 1.

LIVER P-450 CONCENTRATION VERSUS DOSE OF ABT

[Graph showing liver P-450 concentration versus dose of ABT]

TEXT: REGRESSION EQUATION

Another common item that people may wish to add to a line graph illustrating the results of a linear regression is the equation itself. This is done with the REGEQN option on the PLOT statement, in conjunction with turning on the linear regression via the INTERPOL= option on the SYMBOL statement (2). In the example SAS program below, the data is modified from the calibration process of a laboratory instrument. Simplified SAS program code:

```sas
DATA REGWGT;
  INPUT X Y;
  YINV = 1/Y;
  CARDS;
  20 0.1186
  20 0.1165
  40 0.2313
  40 0.2313
  70 0.4063
RUN;

title 'LIVER P--450 CONCENTRATION VERSUS DOSE OF ABT';
proc gplot;
  plot1 y*x / regeqn
  regplot y=x;
  format yinv y inv.
run;
```

Figure 1
DATA MICEABT;
INPUT DRUGABT PROTEIN ENZYME;
P450 = ENZYME/PROTEIN;
CARDS;
0 11.0 12.9
0 10.7 15.4
0 10.7 13.7
25 10.7 7.3
25 9.1 6.6
25 13.6 10.0
100 8.5 3.9
100 7.9 3.0
100 15.0 7.7
400 14.9 6.5
400 10.6 4.7
400 10.1 4.5;
RUN;

TITLE1 HEIGHT=1.5 JUSTIFY=CENTER 'HALF ERROR BARS';
PROC GCHART DATA=MICEABT;
VBAR DRUGABT / TYPE=MEAN
SUMVAR=P450 DISCRETE ERRORBAR=TOP
RAXIS=AXIS1 MAXIS=AXIS2;
AXIS1 LABEL=(A=90 R=0 HEIGHT=1 'P-450 CONCENTRATION')
ORDER=0 TO 1.7 BY .1;
AXIS2 LABEL=(HEIGHT=1 'DOSE OF DRUG ABT');
RUN;

The output is shown below in figure 3.

The output is shown below in figure 3.

The output is shown below in figure 3.

Figure 2

LINES: HALF-ERROR BAR CHART

A more specialized chart used frequently by scientists is the error bar chart, which gives a visual measure of the variability of the data. The bar heights indicate the mean experimental values and are bracketed by lines whose length in each direction represents one standard error of the mean. Since the full error bar chart would be cluttered below the edge of each bar, we concentrate on the half-error bar chart, where only the top of the error bar is shown. The option to add to the VBAR statement is ERRORBAR=TOP (9).

Simplified SAS program code:
HALF ERROR BARS

Figure 3

TEXT, LINES, SHAPES: LEGENDS

Another area of recent SAS/GGRAPH improvements is in drawing legends automatically as opposed to painstakingly sketching out each element of the legend, i.e. put text here, put graph symbol or bar pattern here, draw box around legend. The LEGEND statement is the key and works well for contour plots and maps. Its use is more restricted with bar charts and line graphs (4). In the example below, the data for the line graph had to be transformed to fit into a PLOT statement requirement of the form Y * X = Z. This data compares the blood concentration over time of administering a drug orally (P.O.) versus intravenously (I.V.).

The legend placement desired is to have the symbol and text explanation for the Z dimension one to a line inside the axis area on the graph. The option ACROSS=1 restricts the symbol/text explanation combination to one per line while the POSITION=INSIDE option indicates put the legend inside the axis area. MODE=SHARE allows the legend to share the axis area.

The output is shown below in figure 4.
BLOOD LEVELS OF TOTAL RADIOACTIVITY AFTER DELAPRIL HCl

Figure 4

ANNOTATION ALTERNATIVE

While it is nice to keep the options and statements discussed above in mind when preparing more complex graphics, there are still situations where annotation must be used. Annotation involves three steps:

- Determining what extras to put on the graphic
- Specifying how to put these extras on
- Activating the annotation instructions

Determining what to put on involves using powers of pattern recognition while specifying how to put the extras on involves constructing an annotation dataset. Activating the annotation instructions is the simplest step of all and requires the addition of the ANNOTATE= option on the SAS/GRAPH procedure line (5).

First we will outline a five-step strategy for specifying how to put annotation extras on the graphic and then illustrate how to translate this strategy into SAS code through two examples:

- Labelling the peaks in a chromatogram
- Adding a legend to a grouped bar chart

ANNOTATION STRATEGY

A good annotation strategy begins with answering five questions. The first four apply to adding text, lines, or shapes: (1) what part of the graphics area will be used? Inside the axis area or the entire screen area? (2) where will the annotation element be put? or what x- and y- coordinates will be used to place it on the graphic? (3) what should be done? add text, draw a line, move to another position, draw a bar? (4) how should this be done? what color, font style or typeface, size, or line type to use? The final question applies to text extras only: (5) what orientation to the x- and y-positional coordinates should be used and what text to put there?

Moving closer to actual SAS code, each question can be mapped to appropriate special variables to be put into the annotation dataset. The first question involves specifying what is called the frame of reference and picking values of the variables XSYS, YSYS, and HSYS. The second question involves the X and Y variables for the vertical and horizontal directions on the graphic. The third question translates into picking values of the FUNCTION variable: 'LABEL' for text; 'DRAW' for draw a line; 'MOVE' for move to the x- and y-coordinates specified in the X and Y variables; 'BAR' for drawing a bar. The fourth question translates into picking which of the variables COLOR, STYLE, SIZE, LINE, etc., to include in the annotation dataset. The fifth question involves picking a value for the POSITION variable and putting the text desired into the TEXT variable.

TEXT: LINE GRAPH PEAK LABELS

One of the simplest annotation techniques is to place text on a graph. A good example is labelling selected peaks in a special line graph used by scientists called a chromatogram. We have eyeballed the original data (not included in this paper due to the length of the dataset) and picked out the x- and y-coordinates of the three peaks of interest. Text identifying the specific substance whose presence the peak represents will be positioned directly above each of the corresponding points in the line graph.

Following the five-step annotation strategy outlined above, we begin by identifying the portion of the graphic area or the frame of reference for the additional text. The area between the axis lines is the data area and, since we have the x-y coordinates of each peak, we will use the data value frame of reference. This translates into the annotation dataset value of 2 for both XSYS and YSYS (HSYS is left to its default as is often the case). To answer the where question, the x-y coordinates for the peaks are (13.45,32), (16.30,125), and (23.29,9), which will become three pairs of values for the X and Y variables. The answer of the what to do question for text is the value of 'LABEL' for the FUNCTION variable. The answer to how to do this is to specify the font to be used as the value for the STYLE variable, in this case 'DUPLEX'. Finally, to position this text directly above each x-y coordinate, we use the value of '2' for the POSITION variable and use the values of 'HEF', 'DAF', and 'F' respectively for the TEXT variable.

Simplified SAS code (minus the data file):

FILNAM CHROMDAT 'R177.DAT';
DATA CHROM177;
INFILE CHROMDAT;
LENGTH LINE $80 TIME DPM 4;
INPUT LINE 1-80;
TIME = LEFT_SCAN(LINE,1,' ') Y 60;
DPM = LEFT_SCAN(LINE,3,' ');
DROP LINE;
RUN;

DATA ANNO177;
LENGTH FUNCTION $8 POSITION $1;
RETAIN XSYS YSYS '2' FUNCTION 'LABEL' POSITION '2';
STYLE = 'DUPLEX';
INPUT X Y TEXT $;
CARDS;
13.45 32 HEF
16.30 125 DAF
23.20 9 F
;
TITLE1 H=1.5 'CHROMATOGRAMS OF BRAIN LEVELS OF TOTAL RADIOACTIVITY';
PROC GPLOT DATA=CHROM177
ANNOTATE=ANNO177;
PLOT DPM * TIME = 1/
  VAXIS = AXIS1
  HAXIS = AXIS2;
AXIS1 LABEL = (A=90 R=0 H=1 'TOTAL DPM')
ORDER = 0 TO 400 BY 50
WIDTH = 2
OFFSET = (2);
AXIS2 LABEL = (H=1 'TIME (MINUTES)')
WIDTH = 2
OFFSET = (2);
SYMBOL V=NONE I=J C=BLACK W=2;
RUN;

The output is shown below in figure 5.

![Chromatograms of Brain Levels of Total Radioactivity](image)

**Figure 5**

**LEGENDS: GROUPED BAR CHART**

As noted above, adding legends to certain types of bar charts must still be done by using annotation (2). With data on the effect of norepinephrine dosing on vasculature, we wish to highlight at each dose the relative effect on the aorta vs. the pulmonary artery. For this reason, the grouped bar chart is our choice, not the subgroup bar chart that easily lends itself to adding a legend through the LEGEND statement.

Following the five-step annotation strategy outlined above, we begin by identifying the portion of the graphic area or the frame of reference for the legend. As with the chromatogram example above, the area between the axis lines, or the data area, is the correct choice. But, since the legend text and bar elements are not tied to the heights of the bars but are more easily positioned using x- and y-coordinates as a percentage of the respective total axis length, we will use the data percentage frame of reference. This translates into the annotation dataset value of 1 for both the XSYS and YSYS variables.

Then, the remaining three or four questions are answered separately for the text and bar portions of each line of the legend. We will focus on the aorta only line. To answer the where question for text, the (x,y) coordinates of (78,92) are selected with a little trial and error for how the legend placement looks on the graph. The answer of the what to do question for text is the value of ‘LABEL’ for the FUNCTION variable. The answer to how to do this is to specify the font to be used as the value for the STYLE variable, in this case ‘DUPLEX’. Finally, to position this text centered on and to the right of (78,92), we use the value of 6 for the POSITION variable and use the value ‘AORTA’ for the TEXT variable.

For the bar portion of the legend, we must first tell SAS to move to the x- and y-coordinates of the bottom left hand corner of the bar and then draw the bar through the top right-hand corner x- and y-coordinates. This translates into two separate annotation instructions or observations in the annotation dataset, one for the move and the other for the drawing of the bar. For the move instruction, the only questions that need answers are the where and the how questions, which are answered by specifying X as 70 and Y as 90 with the value of ‘MOVE’ for the FUNCTION variable. For the draw the bar instruction, the where question is answered by determining the top right-hand x- and y-coordinates to be (76,93). The answer of the what to do question for a bar is the value of ‘BAR’ for the FUNCTION variable. The answer of how to draw the bar is to specify ‘E’ for the STYLE variable, 0 for a solid line type for the LINE variable, and ‘BLACK’ for the COLOR variable.

**Simplified SAS code:**

```sas
DATA NOREPH;
  INPUT ORGAN $ 1-12 DOSE DIVISION ;
  CARDS ;
  AORTA  1 37
  AORTA  2 46
  AORTA  3 44
  PULM. ARTERY 1 24
  PULM. ARTERY 2 32
  PULM. ARTERY 3 31 :
  PROC FORMAT;
    VALUE DOSEFMT 1='First'
                   2='Second'
                   3='Third';
  RUN;

DATA ANNON;
  LENGTH FUNCTION COLOR $ 8 POSITION $ 1
                   TEXT $ 12 STYLE $ 6;
  RETAIN XSYS YSYS 1 : ;
  INPUT FUNCTION $ 1-6 X 7-8 Y 10-11 STYLE $ 13-18
                   POSITION $ 20 LINE 22 TEXT $ 24-35 COLOR $ 37-43 ;
  CARDS ;
  MOVE 70 90 . . . . BLACK
  LABEL 78 92 DUPLEX 6. AORTA BLACK
  MOVE 70 80 . . . . .
  BAR 76 83 X2 . . . . BLACK
  LABEL 78 82 DUPLEX 6 . PULM. ARTERY BLACK
  RUN ;
  TITLE 'EFFECT OF NOREPINEPHRINE PRIMING ON VASCULATURE ;
  FOOTNOTE H=1.5 DOSE OF NOREPINEPHRINE ;
  PATTERN1 V=E C=BLACK ;
  PATTERN2 V=X2 C=BLACK ;
  PROC GCHART DATA= NOREPH ANNOTATE= ANNON ;
  AXIS1 LABEL=(A=90 R=0 H=1.5 'DIVISIONS ')
           ORDER = 15 TO 60 BY 5
           MINOR = NONE ;
  AXIS2 LABEL = NONE
           VALUE = NONE ;
  AXIS3 LABEL = NONE
           FORMAT DOSE DOSEFMT . ;
  VBAR ORGAN /
  SUMVAR=DIVISION GROUP=DOSE
  PATTERN1= MIDPOINT
  MAXIS=AXIS2
  RAXIS=AXIS1
  GAXIS=AXIS3
  ;
  RUN ;
```

The output is shown below in figure 6.
CONCLUSION

With recent enhancements to SAS/GRAFP, more options and statements are available to add text, lines, and shapes to the basic types of graphs. This paper has reviewed specific examples for putting numbers at the end of bars on a bar chart, a regression equation at the bottom of a line graph, half-error bars on a bar chart, and a legend on a line graph. Beyond these examples are more situations that still require the use of annotation. This paper has outlined a five-step strategy to help translate the annotation extras into specific types of SAS variables in the annotation dataset. In the final analysis, the best rule of thumb to determine if annotation is needed is to use your powers of pattern recognition to see if your graphic matches up with the patterns of the simple situations described here and in the SAS manuals before travelling down the path of annotation.

REFERENCES


