How Not to Lose Your Head When Enhancing SAS/GRAPH Output

Arthur L. Carpenter
California Occidental Consultants

KEY WORDS
ANNOTATE, GRAPHICS, FRANCE, GSLIDE, GANNIO, FUNCTION

INTRODUCTION
The ANNOTATE Facility within SAS/GRAPH has a reputation for causing new graphics users head aches and frustration. Much of the confusion is caused by the approach that ANNOTATE uses to gather the information it needs to place text, lines, and symbols on the plot or graph. ANNOTATE is not like the rest of SAS. Directions are not passed to the ANNOTATE Facility using a PROCedural call. Specific statements are not used to define parameters. Information is passed to a PROCedure through a SAS data set. To make matters a bit more arcane, the user of the ANNOTATE data set is not free to select variable names or their attributes. This is very contrary to both the traditional DATA step and the PROC step. This paper will introduce the concept of a data set which is used to transmit informational values to a PROCedure. Selected ANNOTATE variables will be introduced and discussed, and the resulting ANNOTATE data set will be used to enhance graphics output.

ANNOTATE DATA SET (WHERE, WHAT, and HOW)
The purpose of the ANNOTATE data set is to pass information to the appropriate PROCedure. Individual PROCedures are/were not designed to accommodate the flexibility required for the types of graphics enhancements possible through the ANNOTATE Facility. In the data step, the user creates an ANNOTATE data set which contains the functional information directly applicable to the graphics output.

Although an ANNOTATE data set can contain over twenty variables that have specific meanings, the new user need not master all of them prior to creating an ANNOTATED plot or graph. Basically the two questions of WHERE and WHAT need to be answered by the information contained in the data set. Often this information is conveyed by the three primary ANNOTATE variables; FUNCTION, X, and Y. FUNCTION tells ANNOTATE what to do and X and Y tell it where to do it. Most of the other ANNOTATE variables are used to enhance or supplement these three.

Figure 1

WHERE: Positioning
WHAT: Functionality
HOW: Using Annotate Data

Figure 2

WHERE:

Define available area

XSYS & YSYS

Figure 3

Each of these system values indicates if the coordinate system is to be measured in units (inches or cm as defined in the GOPTIONS) or as a percentage of the available area. Using the simple plot shown in Figure 3, the available area can...
be coded using XSYS and YSYS. These areas can be the 'Graphics Area' (Figure 4), 'Window Area' (Figure 5), or 'Screen Area' (Figure 6).

The 'Graphics Area' will be defined by the extent of any axis that may be plotted. 'Window Areas' include all of the graph except the border areas utilized by titles and footnotes. And finally the 'Screen Area' includes the entire graphics area.

The choices can be a bit daunting and while the power, like cake and other desserts, is nice, what we really need is a staple, bread. Two of the twelve choices of XSYS and YSYS will satisfy most of our ANNOTATE needs. Data driven applications will most often use 'absolute data' ('2') and text placement applications the 'absolute window percent' ('5').

The WHERE is further defined using the numeric coordinate variables X and Y, Figure 7, (and occasionally Z, but we won't talk about that, this being a family book often used put people to sleep). The values of X and Y can be defined explicitly as in figures 8 & 9 or may be data driven. In either case X is used to define horizontal coordinates and Y, of course, the vertical.

```
* define the label;
xsyz = '2'; ysys = '2';
x = 49; y = 7;
text = 'S'; style = 'marker';
position = '5'; output;
text = '49.7'; style = 'text';
position = '9'; output;
```
WHAT:

The character variable FUNCTION provides the information of WHAT is to be done. Virtually all ANNOTATE data sets will have this variable defined for all observations. This variable provides the user the ability to express what is to be done. Consequently it is one of the best places for a new user to start when creating an ANNOTATE data set. Values of FUNCTION include MOVE and DRAW with a pen, add a LABEL, and to add SYMBOLs, BARs and POLYgons to a graph (Figure 10). Usually when a FUNCTION is defined, one or more supplemental variables will also be defined. The SAS/GRAPH User's Guide (Release 6.03) TABLE 7.3 notes which of these variables is typically used with the basic FUNCTIONS.

WHAT

Define the task

FUNCTION

MOVE

DRAW

LABEL

Figure 10

The FUNCTION='MOVE' picks up and moves the pen to the specified location. When followed by a FUNCTION='DRAW' a line is drawn to the second location. Using a series of MOVEs and DRAWs one is able to sketch a simple to complex diagram (Figure 11). The window percentage coordinate system was used for both X and Y. X=50 indicates a position 50% of the way across the graphics window as measured from the left side (Figure 12).

Specify a Label

FUNCTION='LABEL';
X=50; Y=50;
TEXT='JULY 14, 1787';

Figure 13

Other variables are available to enhance the text's color, size, and font. The variables which may be used with FUNCTION='LABEL' are designated in TABLE 7.3 of the SAS/GRAPH User's Guide. And further detailed in the section entitled "Outline of Use: Basic Functions" in the same manual.

TEXT attributes are controlled in much the same way as they are in a SAS/GRAPH TITLE or FOOTNOTE. The size of the text is controlled by the SIZE variable in much the same way as H= is used in a
TITLE statement. Font selection is through the STYLE variable which corresponds to the F=, and COLOR specifies the color as does C= in titles.

DATA MARIE;
LENGTH FUNCTION COLOR STYLE $8;
FUNCTION='LABEL';
COLOR='BLUE'; STYLE='SCRIPT';
SIZE=4;
TEXT='JULY 14, 1789';
X=30; Y=15;
OUTPUT;
PROC GANNO ANNO=MARIE;
TITLE H=2 F=SIMPLEX 'Bastille Day';
RUN;

HOW:
Once created, the ANNOTATE data with its functions, coordinates and associated variables must be passed to a PROCedure capable of using it. The display of ANNOTATE data commands can be accomplished in one of two basic ways, either through PROC GANNO or through the use of the ANNO= option (Figure 15).

CREATING THE ANNOTATE DATA SET
The ANNOTATE data set can be created in any of the ways that a SAS data set is created. In the previous section the data set MARIE is created using assignment statements, however, this becomes cumbersome if the data set is large. Often, if there are many commands, such as when drawing, a raw or flat file is created and read as data using the INPUT statement. When the graphics enhancement depends on an established SAS data set, that data set can sometimes be used to build the ANNOTATE data set as well.

* Using a flat file;
DATA TOWER;
LENGTH FUNCTION $8;
RETAIL XSYS YSYS '5' LINE 2;
INPUT FUNCTION X Y;
CARDS;
MOVE 30 30
DRAW 30 32
DRAW 45 45
DRAW 30 30
 PROC GANNO ANNO=TOWER;
 TITLE H=2 F=SIMPLEX 'Eiffel Tower';
RUN;

Existing SAS data sets can be used as the input to create an ANNOTATE data set. The two do not necessarily need to be
distinct.

PROC PRINT DATA=ACID;

OBS TEMP TIME ACID
1 16 5 25
2 20 12 15
3 15 23 8
4 12 41 3

DATA ANNPLT; SET ACID;
LENGTH FUNCTION $8 TEXT $2;
RETAIN XSYS YSYS '2' STYLE 'SIMPLEX'

FUNCTION 'LABEL' POSITION '6';
DROP TEMP TIME ACID;
TEXT=PUT(ACID,2.);
PROC G PLOT DATA=ACID ANNO=ANNPLT;
PLOT TEMP*ACID;
TITLE H=2 F=SIMPLEX 'Temp. & Cone.';
'RETAIN XSYS YSYS '2' STYLE 'SIMPLEX'
'Text=PUT(ACID,2.);
PROC G PLOT DATA=ACID ANNO=ANNPLT;
PLOT TEMP*ACID;
TITLE H=2 F=SIMPLEX 'Temp. & Conc.';
'Concentration';
RUN;

ANNPLT is essentially similar to MARIE, and only really differs in its method of creation. The variable POSITION determines where the text string is to be placed relative to the (X, Y) location. POSITION='6' indicates that the text will start immediately to the right of the plotted symbol.

CHOOSING THE ANNOTATE VARIABLES

The process of selecting the variables to include in the ANNOTATE data set should always start with FUNCTION. The value of the FUNCTION variable often will determine what other variables are needed. The User's Guide is organized to describe those variables that are associated with each value that FUNCTION can take on. The Release 6.03 SAS/GRAPH User's Guide (P. 130) details other variables that could have been selected. These include font selection (STYLE), size of the text (SIZE), color of text (COLOR), and the position of the text relative to the designated coordinate (POSITION).

Next, in all but two of the functions, the location on the graph must be selected. The coordinates are usually placed in the variables X and Y. The coordinate system may also need to be selected.

This process can be summarized as:

I Select a FUNCTION (Table 7.3)

II Select support variables (Table 7.3 and function text)

III Select coordinate system and coordinate variables

IV Assign values to the ANNOTATE data set

SUMMARY

The construction of ANNOTATE data sets is often viewed as a daunting task for users who are unfamiliar with the approach used by SAS/GRAPH to assimilate the ANNOTATE commands. The ANNOTATE data set can be seen as a way to transfer these commands into a SAS/GRAPH PROCedure.

The construction of the data set is fairly straightforward as long as the user starts with the definition of the FUNCTION of each observation in the ANNOTATE data set. The FUNCTION determines for the most part what other variables will be needed. After selection of the value of FUNCTION and appropriate support variables, the coordinates (X and Y) need to be supplied. These coordinates may be in the units of the plotted data or in units which define the graphics window itself.

ABOUT THE AUTHOR

Arthur L. Carpenter has over fourteen years of experience as a statistician and data analyst and has served as a senior consultant with California Occidental Consultants, CALOXY, since 1983. His publications list includes a number of papers and posters presented at SUGI and he has developed and presented several courses and seminars on statistics and SAS programming.

CALOXY offers SAS contract programming and in-house SAS training nationwide.

AUTHOR

Arthur L. Carpenter
California Occidental Consultants
4239 Serena Avenue
Oceanside, CA 92056-5018
(619) 724-8579

REFERENCES


TRADEMARK INFORMATION

SAS and SAS/GRAPH are registered trademarks of the SAS Institute, Inc., Cary, NC, USA.