Little Orphan Annotate:
How To Dress Up SAS/GRAPH® Output

Arthur L. Carpenter
California Occidental Consultants

KEY WORDS
ANNOTATE, FUNCTION, XSYS, YSYS, SAS/GRAPH, GPLOT

ABSTRACT
The ANNOTATE Facility within SAS/GRAPH has a reputation for causing new graphics users headaches and frustration. Many abandon hope of making use of this important graphics tool. Much of this 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 much of the rest of SAS®. Directions are not passed to the ANNOTATE Facility using a PROCedural call. Specific statements are not used to define parameters. Instead, information is passed to a PROCedure through a specialized 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. Although this is very contrary to both the traditional DATA step and the PROC step, it does not mean that using ANNOTATE needs to be difficult.

This workshop is designed for users new to ANNOTATE and 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. Specific course topics include: GSLIDE (drawing a picture), and GPLOT (adding text and labels to the plot and to specific data points).

THE ANNOTATE FACILITY
The ANNOTATE facility is included within SAS/GRAPH. It acts as a bridge between the procedure selected by the user and the user's desire to customize the graphics output. The power of the ANNOTATE facility is accessed through the use of a specialized data set. ANNOTATE looks for variables with specific names and attributes, and the values taken on by these variables in turn instruct ANNOTATE as to the user's intentions.

The data set itself is fairly rigidly defined in terms of the variables that it is to contain and the attributes that these variables must have.

THE ANNOTATE DATA SET
The purpose of the ANNOTATE data set is to pass information to the appropriate PROCedure. Individual PROCedures 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 at first it seems clumsy to pass specific information to a PROCedure through the use of dedicated data sets, PROCedures are actually designed to accept, interpret, and respond to SAS data sets. Therefore, an ANNOTATE data set can contain the functional information and PROCedure statements that could not be included in the PROC step itself. The result is a stronger and more flexible approach.

An ANNOTATE data set can contain over twenty variables that have specific meanings. Fortunately, however, the new user need not master all of them prior to creating an ANNOTATE plot or graph. Basically the two questions of: WHAT TO DO?, and WHERE TO DO IT?, 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.

The ANNOTATE data set is processed one observation at a time. The value of the variable FUNCTION is evaluated for each observation. The value that FUNCTION takes on determines which of the other variables in the observation may have information applicable to the FUNCTION and, therefore, to the observation being processed. The action specified by the FUNCTION along with whatever modifiers are appropriate takes place before the next observation is read.
Table 1 shows those ANNOTATE variables discussed in this paper and whether or not they are used with various values of the FUNCTION variable. Shown are three of the values that can be taken on by the variable FUNCTION. On an observation that has FUNCTION='DRAW', for instance, the variables SIZE and LINE will be used when defined. However, when FUNCTION='MOVE' the variable SIZE will be ignored.

WHAT TO DO

Since ANNOTATE is used primarily to enhance a graph, the first step for the programmer faced with using the ANNOTATE facility is determine what needs to be done. The answer will usually take the form of something like: 'add a label,' 'include a legend in the upper right hand corner,' or 'draw a triangle.' This information is passed to ANNOTATE through specific variables in the annotate data set.

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 with 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. Usually when a FUNCTION is defined, one or more supplemental variables will also be defined (Table 1). The SAS/GRAPH User's Guide (Release 6.03 - TABLE 7.3) notes which of these variables is typically used with the basic FUNCTIONs.

FUNCTION='LABEL'

Labels can easily be added to a plot by specifying FUNCTION='LABEL'. This alerts ANNOTATE that a text string (contained in the variable TEXT) is to be placed on the graph.

Other variables are available to enhance the text's color (COLOR), size (SIZE), and font (STYLE). The variables which may be used with FUNCTION='LABEL' are designated in the SAS/GRAPH User's Guide (Release 6.03 - TABLE 7.3). And further detailed in the section entitled "Outline of Use: Basic Functions" in the same manual.

TEXT attributes are controlled in a similar fashion 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.

FUNCTION='MOVE' and FUNCTION='DRAW'

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. When drawing the variable LINE may be used to specify the style of line to be used (solid, dashed, etc.). The line thickness may also be controlled through the use of SIZE.

WHERE TO DO IT

Almost all uses of ANNOTATE require information on location, i.e. WHERE on the graph should the annotation be placed. The variables X and Y provide the coordinates that the specified FUNCTION is to take place. The physical location on the graph depends on the coordinate system which can be selected by using the variables XSYS and YSYS. Although these character variables can take on one of twelve 'system' values as shown in the SAS/GRAPH User's Guide (Version 6.03 - Table 7.3 and Figure 7.1, Version 6.06 - Table 18.3 and Figure 18.2), 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'). 'Absolute data' (XSYS & YSYS='2') places the point according to the values of the horizontal and vertical axes that are plotted on the
'Absolute window percent' (XSYS & YSYS='5'), however, uses percentages of the graph page based from the lower left corner.

The WHERE is further defined using the numeric coordinate variables X and Y. These variables may be defined explicitly in a data step or may be data driven. In either case, X is used to define horizontal coordinates and Y, of course, the vertical. Where a particular value of X will be located depends on the value assigned to XSYS. When XSYS='5' (window percentage) a value of X=50 will be plotted in the middle of the page. However, when XSYS='2' (data value) the placement depends on the horizontal axis. If the axis ranges from 0 to 55, X=50 will be located on the far right.

HOW TO GET IT DONE

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 in one of the other SAS/GRAPH procedures, such as, GPLOT or GSLIDE.

ANNOTATE data sets can be utilized by most of the PROCedures in SAS/GRAPH. Usually ANNOTATE is used to enhance graphics output through the use of the information in an ANNOTATE data set, however, PROC GANNO only produces graphics as directed by ANNOTATE. In each of these PROCedures a PROC option is used to designate the annotate data set. This option takes the form of ANNOTate=datasetname.

Consider the ANNOTATE data set ORPHAN. It could be made available to the following PROCs:

```
PROC GANNO ANNOTate=ORPHAN;
PROC GPLOT DATA=PLOTDATA ANNO=ORPHAN;
PROC GPLOT DATA=PLOTDATA;
PLOT VVAR=HVAR / ANNOTate=ORPHAN;
```

CREATING THE ANNOTATE DATA SET

The ANNOTATE data set can be created in any of the ways that a SAS data set is created. Small control files that add only a label or two are often created using assignment statements.

In the following example, the 'Absolute window percent' coordinate system (XSYS & YSYS = '5') 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.

```
DATA ANNIE;
RETAIN XSYS YSYS '5';
LENGTH FUNCTION COLOR STYLE $8;
FUNCTION='LABEL';
COLOR='BLUE'; STYLE='SCRIPT';
SIZE=4;
TEXT='HOME WANTED FOR GIRL WITHOUT EYES';
X=50; Y=75;
OUTPUT;
Y=30; TEXT='HAS DOG, WILL TRAVEL';
OUTPUT;

PROC GANNO ANNO=ANNIE;
TITLE H=5 F=SIMPLEX 'Orphanage Classifieds';
RUN;
```

In the previous section the data set ANNIE 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. The input file should contain all necessary information required by ANNOTATE. Variables not used by ANNOTATE will be ignored as will variables not required by a particular function.

```
* USE THE CARDS STATEMENT TO PRESENT THE DATA;
DATA DIAMOND;
LENGTH FUNCTION $8;
RETAIN XSYS YSYS '5' LINE 1;
INPUT FUNCTION X Y;
CARDS;
MOVE 0 3
DRAW 3 0
DRAW 5 4
DRAW 6 3
DRAW 0 3
DRAW 6 3
DRAW 3 0
DRAW 3 4
DRAW 0 3
DRAW 4 3
DRAW 2 3
DRAW 3 4
DRAW 4 3
DRAW 5 4
DRAW 2 3
DRAW 3 0
DRAW 4 3

PROC GSLIDE ANNO=DIAMOND;
TITLE1 'EXAMPLE 4';
TITLE2 H=3 F=SIMPLEX "Warbuck's Diamond";
RUN;
```

When the graphics display depends on an established SAS data set, that data set can sometimes be used as the basis from
which to build the ANNOTATE data set as well. The data set to be plotted and the annotate data set do not necessarily need to be distinct, variables that are not used by ANNOTATE will be ignored during the annotation process.

*LABEL ALL POINTS WITH A OZONE LEVEL > 2;
DATA ANNPLOT; SET SASUSER.SF88AIR;
LENGTH FUNCTION $8 TEXT $5;
RETAIN XSYS YSYS '2' STYLE 'SIMPLEX'
FUNCTION 'LABEL' POSITION '6';
IF 03 GE 2;
X=MONTHi;
Y=03i;
TEXT=PUT(03,5.2) ,
PROC GPLOT DATA=SASUSER.SF88AIR
ANNO=ANNPLOT;
BY STATION;
PLOT 03 * MONTH,
SYMBOLI L=l V=NONE I=JOIN,
TITLE1 'EXAMPLE 5';
TITLE2 H=2 F=SIMPLEX 'OZONE LEVELS
IN 1988';
RUN;

ANNPLOT is essentially similar to ANNIE, 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 SAS/GRAPH User's Guide (Release 6.03 - starting on page 127) 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, for nearly all 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
II Select support variables
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 straight-forward 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 fifteen 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 statistics consultations, 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.

APPENDIX

Additional workshop examples.

* EXAMPLE6.SAS

* READ DATA FROM A FILE TO CREATE AN ANNOTATE DATA SET.
* USE PROC GLIIDE AND ANNOTATE TO DRESS UP A AIR QUALITY PLOT;
* USING DATA VALUES, ADD A STATION LABEL AT THE BOTTOM OF THE PLOT;
* OPTIONS NOCHARACTERS NOFILL NOCELL;
* OPTIONS CBACK=GREY GACCESS='SASGASTD>DEV/TTY' DEVICE=VGA;

PROC PRINT DATA=SASUSER.SF88AIR;
DATA ANNPLT; SET SASUSER.SF88AIR;
LENGTH FUNCTION $8 TEXT $15;
RETAIN XSYS YSYS '2' STYLE 'SIMPLEX' FUNCTION 'LABEL' POSITION '6';
IF N=1 THEN DO;
- )(-=6;
Y=1;
TEXT = STATION;
OUTPUT;
END;
XSYS='2'; YSYS='2';
IF 03GE2;
X=MONTH; Y=03;
TEXT=PUT(03.S.2);
OUTPUT;
RUN;
PROC GPLOT DATA=SASUSER.SF88AIR ANNO=ANNPLT;
PLOT 03 * MONTH;
SYMBOL1 L=1 V=NONE I=JOIN;
TITLE1 'EXAMPLE 6';
TITLE2 H=2 F=SIMPLEX 'OZONE LEVELS IN 1988';
RUN;

* EXAMPLE7.SAS

* READ DATA FROM A FILE TO CREATE AN ANNOTATE DATA SET.
* USE PROC GLIIDE AND ANNOTATE TO DRESS UP A AIR QUALITY PLOT;
* USING PERCENTAGE VALUES, ADD A STATION LABEL AT THE BOTTOM OF THE PLOT;
* OPTIONS NOCHARACTERS NOFILL NOCELL;
* OPTIONS CBACK=GREY GACCESS='SASGASTD>DEV/TTY' DEVICE=VGA;

PROC PRINT DATA=SASUSER.SF88AIR;
DATA ANNPLT; SET SASUSER.SF88AIR;
LENGTH FUNCTION $8 TEXT $15;
RETAIN XSYS YSYS '2' STYLE 'SIMPLEX' FUNCTION 'LABEL' POSITION '6';

PROC GPLOT DATA=SASUSER.SF88AIR ANNO=ANNPLT;
PLOT 03 * MONTH;
SYMBOL1 L=1 V=NONE I=JOIN;
TITLE1 'EXAMPLE 7';
TITLE2 H=2 F=SIMPLEX 'OZONE LEVELS IN 1988';
RUN;

* EXAMPLE8.SAS

* DEMONSTRATE THE POSITION VARIABLE WITH FUNCTION=LABEL;
* OPTIONS NOCHARACTERS NOFILL NOCELL;
* OPTIONS CBACK=GREY GACCESS='SASGASTD>DEV/TTY' DEVICE=VGA;

DATA PLTDAT;
INPUT POS $ X Y;
CARDS;
1 1
2 2
3 3
4 4
5 5
6 6
7 7
8 8
9 9
A 10 10
B 11 11
C 12 12
D 13 13
E 14 14
F 15 15
RUN;
DATA ANNDAT;
SET PLTDAT;
LENGTH FUNCTION $8 TEXT $15;
RETAIN XSYS YSYS '6' STYLE 'SIMPLEX' FUNCTION 'LABEL' ;
TEXT = 'POS' , !! POS;
POSITION , POS;
RUN;
PROC GPLOT DATA=PLTDAT ANNO=ANNDAT;
PLOT Y X;
SYMBOL1;
TITLE1 'EXAMPLE 8';
TITLE2 H=2 F=SIMPLEX 'LABEL POSITIONS';
RUN;

* EXAMPLE9.SAS

* DEMONSTRATE THE XSYS AND YSYS VARIABLES USING FUNCTION=FRAME;
* OPTIONS NOCHARACTERS NOFILL NOCELL;
* OPTIONS CBACK=GREY GACCESS='SASGASTD>DEV/TTY' DEVICE=VGA;

DATA PLTDAT;
INPUT X Y;
CARDS;
1 1
15 15
RUN;
DATA ANNDAT;
LENGTH FUNCTION $8;
RETAIN FUNCTION 'FRAME' X Y LINE 1 SIZE 3;
COLOR='BLUE XSYS=': '2' YSYS='2'; OUTPUT;
COLOR='RED XSYS=' '3' YSYS=' '3'; OUTPUT;
COLOR='GREEN XSYS=' '5' YSYS=' '5'; OUTPUT;
RUN;

553
PROC GPLOT DATA=PLTDAT ;
    PLOT VERT * HORZ I=JOIN V=NONE;
    TITLE 'EXAMPLE 9';
    TITLE2 H=2 F=SIMPLEX 'USING FRAME TO DEMONSTRATE XSYS & TITS';
    FOOTNOTE H=1.5 F=SIMPLEX C=WHITE 'XSYS & TITS ='
        C=BLUE  '2'
        C=RED   '3'
        C=GREEN '5';
    LABEL VERT = 'VERTICAL LABEL';
RUN;