THE ANNOTATE FACILITY:
A QUICK START TO AN EASY OVERVIEW

Arthur L. Carpenter

ABSTRACT
The ANNOTATE facility is included within SAS/GRAPH™ and acts as a bridge between the procedure selected by the user and the user's desire to customize the graphics output. The ANNOTATE facility can be used to; position text or symbols anywhere on the graph, control text characteristics, draw line segments and polygons, and otherwise enhance the user's graph.

Using ANNOTATE need not be difficult and can be easily introduced by presenting the different fundamentals of the specialized ANNOTATE 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.

Unlike most SAS data sets, the ANNOTATE data set is fairly rigidly defined in terms of the variables that it is to contain and the attributes that these variables must have. Because individual PROCedures were not designed to accommodate the flexibility required for the types of graphics enhancements possible through the ANNOTATE facility, the ANNOTATE data set is used to pass information to the appropriate graphics PROCedures.

This tutorial will discuss the important attributes of the ANNOTATE data set and how the user can select the correct variables and values.

THE ANNOTATE FACILITY
The ANNOTATE Facility can be used to:
• position text or symbols anywhere on the graph
• control text color, font, and size
• draw line segments of any length or thickness
• draw polygons of any style, size or shape

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 ANNOTATEd 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

<table>
<thead>
<tr>
<th>WHAT</th>
<th>WHERE</th>
<th>HOW</th>
</tr>
</thead>
<tbody>
<tr>
<td>FUNCTION</td>
<td>X Y XSYS YSYS</td>
<td>POSITION SIZE STYLE LINE TEXT</td>
</tr>
<tr>
<td>LABEL</td>
<td>* * *</td>
<td>*</td>
</tr>
<tr>
<td>MOVE</td>
<td>* *</td>
<td>*</td>
</tr>
<tr>
<td>DRAW</td>
<td>* * *</td>
<td>* *</td>
</tr>
</tbody>
</table>

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.

Many of the supporting variables convey different information to ANNOTATE depending on the value of the variable FUNCTION. When FUNCTION = 'label' the variable SIZE refers to the height of the text to be printed, but when FUNCTION = 'DRAW' SIZE relates to the width of the line to be drawn.

CHOOSING THE ANNOTATE VARIABLES

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.

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 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. These might 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

WHAT TO DO

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 Reference Manual (Release 6.06) TABLE 18.1, pp. 472-473 lists the ANNOTATE variables and TABLE 18.2, p.472 lists the values that the variable FUNCTION can take on.

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).

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 (this may be an imaginary or virtual pen on many plotters and displays) 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 Reference Manual (Version 6.06 - Figure 18.2 p. 476), 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 graphics output area percent' ('3'). 'Absolute data' (XSYS & YSYS = '2') places the point according to the values of the horizontal and vertical axes that are plotted on the graph. 'Absolute output area percent' (XSYS & YSYS = '3'), 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 = '3' (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 56, X = 50 will be located on the far right.

1424
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, G PLOT 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=datanamename.

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

```
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.

PROC GANNO

The GANNO PROCedure is designed to display information contained in ANNOTATE data sets. No other data or graphics information is shown.

In the following example, the 'Absolute graphics output area percent' coordinate system (XSYS & YSYS = '3') was used for both X and Y. X = 60 indicates a position 50% of the way across the graphics window as measured from the left side.

```
DATA SANDY;
LENGTH FUNCTION $8;
RETAIN FUNCTION 'LABEL' XSYS YSYS '3' X Y 50;
TEXT='A Dog Called Sandy';
PROC GANNO ANNO=SANDY;
QUIT;
```

The variables SIZE and STYLE can be used to adjust the size of the letters and the font used to make the label.

```
DATA SANDY;
LENGTH FUNCTION $8;
RETAIN FUNCTION 'LABEL' XSYS YSYS '5' X 50;
STYLE='brush'; ... SIZE=4;
TEXT='A Dog Called Sandy';
PROC GANNO ANNO=SANDY;
RUN;
```

PROC GSLIDE With ANNOTATE

TITLES and FOOTNOTES can be used to advantage when the ANNOTATE data set is used in one of the other SAS/GRAPH PROCedures, such as, PROC GSLIDE.

```
* USE PROC GSLIDE AND ANNOTATE TO CREATE A CLASSIFIED AD FOR ANNIE:
DATA ANNIE;
LENGTH FUNCTION COLOR STYLE $8;
RETAIN FUNCTION 'LABEL' XSYS YSYS '5' X 50;
COLOR='BLUE';
STYLE='SCRIPT';
SIZE=4;
TEXT='Home Wanted Y=75;
OUTPUT;
SIZE=2;
Y=50;
STYLE='DUPLEX';
TEXT='GIRL - WITHOUT EYES';
OUTPUT;
SIZE=3;
Y=30;
STYLE='TRIPLEX';
COLOR='GREEN';
TEXT='Has Dog Will Travel';
```

1425
In the previous section the data set ANNIE is created using assignment statements, however, this becomes cumbersome if the data set is large. This is especially true when using ANNOTATE to draw figures. An alternate way to build the ANNOTATE data set is to create a raw or flat file which can be 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 6 3
  DRAW 5 4
  DRAW 1 4
  DRAW 0 3
  DRAW 6 3
  MOVE 1 4
  DRAW 2 3
  DRAW 3 4
  DRAW 4 3
  DRAW 5 4
  MOVE 2 3
  DRAW 3 0
  DRAW 4 3
;
PROC GSLIDE ANNO=DIAMOND;
  TITLE1 'EXAMPLE 4';
  TITLE2 H=3 F=SIMPLEX "Warbuck's Diamond";
RUN;

The size and position of the diamond can be changed by adjusting the values taken on by X and Y.

* USE THE CARDS STATEMENT TO PRESENT THE DATA;
DATA DIAMOND;
  LENGTH FUNCTION $8;
  RETAIN XSYS YSYS '5' LINE 1;
  INPUT FUNCTION X Y;
  X = X*10 + 20;
PROC GSLIDE ANNO=DIAMOND;
  TITLE1 'EXAMPLE 4';
  TITLE2 H=3 F=SIMPLEX "Warbuck's Diamond";
RUN;

Using Established SAS Datasets
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.
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.

A label can be added to the bottom of the plot by including one more observation into the ANNOTATE data set.

```
*LABEL ALL POINTS WITH A OZONE LEVEL > 2;
DATA ANNPLT; SET HERE.SF88AIR;
LENGTH FUNCTION $8 TEXT $15;
RETAIN XSYS YSYS '2' STYLE 'SIMPLEX' FUNCTION 'LABEL' POSITION '6';
IF 03 GE 2;
  X=MONTH; Y=03;
  TEXT=PUT(03,5.2);
OUTPUT;
END;
PROC GPLOT DATA=HERE.SF88AIR ANNO=ANNPLT;
PLOT 03 * MONTH;
SYMBOL1 L=1 V=NONE I=JOIN;
TITLE1 'EXAMPLE 8';
TITLE2 H=Z F=SIMPLEX 'OZONE LEVELS IN 1988';
RUN;
QUIT;
```

The label for SAN FRANCISCO could also be placed by using the screen percentages. This is often easier as it does not require knowledge of the data prior to generating the graphic. Notice that the variables XSYS and YSYS are used to control the coordinate system. The station label is placed using screen percentages ('3') while the labels for the data points rely on the data values ('2').

```
*LABEL ALL POINTS WITH A OZONE LEVEL > 2;
DATA ANNPLT; SET HERE.SF88AIR;
LENGTH FUNCTION $8 TEXT $15;
RETAIN XSYS YSYS '2' STYLE 'SIMPLEX' FUNCTION 'LABEL' POSITION '5';
IF _N_=1 THEN DO;
  XSYS='3'; YSYS='3';
  STYLE='DUPLEX'; SIZE=1.5;
  X=50;
  Y=65;
  TEXT=STATION;
  OUTPUT;
END;
IF 03 GE 2;
  XSYS='2'; YSYS='2';
  STYLE='SIMPLEX';
  SIZE=1;
  X=MONTH; Y=03;
  TEXT=PUT(03,5.2);
  OUTPUT;
RUN;
PROC GPLOT DATA=HERE.SF88AIR ANNO=ANNPLT;
PLOT 03 * MONTH;
SYMBOL1 L=1 V=NONE I=JOIN;
TITLE1 'EXAMPLE 8';
TITLE2 H=Z F=SIMPLEX 'OZONE LEVELS IN 1988';
RUN;
QUIT;
```
The value of POSITION was changed to '5' in Figure 8. Notice that the labels are now partially covered by the line.

Relationship Between POSITION and FUNCTION = 'LABEL'
The POSITION variable is used to place the text relative to the point established by X and Y. The variable can take on 15 values.

* DEMONSTRATE THE POSITION VARIABLE WITH
* FUNCTION=LABEL;
DATA PLTDAT;
LENGTH POS $1;
INPUT POS $ X Y;
CARDS;
1 1 1
2 2 2
3 3 3
4 4 4
5 5 5
6 6 6
7 7 7
8 8 8
9 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 STYLE 'SIMPLEX' XSYS YSYS '2' FUNCTION 'LABEL';
TEXT 'POS':: pos;
POSITION = POS;
RUN;
PROC GPLOT DATA=PLTDAT ANNO=ANNDAT;
PLOT Y * X;
SYMBOL 1;
TITLE1 'EXAMPLE 9';
TITLE2 H=2 F=SIMPLEX 'LABEL POSITIONS';
RUN;

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.

TRADEMARK INFORMATION
SAS/GRAPH is a registered trademark of SAS Institute, Inc. in the USA and other countries. ™ indicates USA registration.

REFERENCES


ABOUT THE AUTHOR
Arthur L. Carpenter has over seventeen 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. Art has served as a steering committee member and president of the Southern California SAS User’s Group, a Section Chair at the Western Users of SAS Software regional conference, WUSS, and in various positions at SUGI. He has developed and presented several courses and seminars on statistics and SAS programming and has taught for Colorado School of Mines, University of Redlands, and University of California at San Diego. His latest book, co-authored with Charlie Shipp, Visualization of Scientific Data: Using SAS/GRAPH Without Annotate, is being published by SAS Institute’s Books by Users.

CALOXY offers SAS contract programming and in-house SAS training nationwide, including a three day course on SAS/GRAPH. This presentation is based on a Chapter in the SAS/GRAPH course.

AUTHOR
Arthur L. Carpenter
California Occidental Consultants
4239 Serena Avenue
Oceanside, CA 92056-6018
(619) 724-8579

1428