Annotate facilitates data-driven graphs

Meera G. Kumar and Jeffrey Cortez
Bridgewater Crossings, NJ

ABSTRACT

This paper gives an example of how we used the annotate dataset in SAS/GRAPH to make the plot symbol vary in size in proportion to the sample size of our clinical data.

INTRODUCTION

A wise person once said that “a picture is worth a thousand words”. This expression, commonly used in the context of art, is also true in the world of SAS graphics. A graph has the power to condense information from thousands and thousands of data points into a space often smaller than an eight-by-eleven-and-a-half piece of paper. It can also communicate a story in a matter of seconds that could otherwise be understood in minutes or hours.

There are many parts of a graph that can be used to communicate to the reader. All should be well-understood since they are powerful tools at your disposal to tell the story of your clinical data. We wish to focus primarily on one of the less obvious and most commonly overlooked part of the graph - the symbol size. What can you communicate with the symbol size, you may ask? The answer is quantitative information. Since symbol size is defined on a numeric scale you can have it represent some quantity. In a way, modifying the symbol size is like adding another axis or dimension to your plot. As will be demonstrated the vehicle that allows us to control the symbol size as well as other features of the graph is the SAS annotate facility.

EXAMPLE

Suppose you are plotting a set of statistics. They could be means, medians, relative risks, p-values, etc. Then suppose that these statistics come from samples of different sizes, some that are perhaps much larger than others. You may wish to have your graph highlight or emphasize the statistics that come from the larger samples. One way to do this would be by increasing the size of the symbol proportionate to the sample size. For example, a statistic from a sample of size 20 could have a symbol that is twice as big as the symbol for a statistic from a sample of size 10.

Consider the following relative risk plot (see figure 1) for various subgroups.
The dots represent the relative risks. They vary in size based on the sample size of each subgroup, so that the bigger subgroups get the bigger dots. The relative risks from the smaller subgroups are so small that they are almost undetectable which is exactly our intention since we wish to de-emphasize these less reliable statistics. A 95% confidence interval is shown for each relative risk using bars to the left and right of each symbol.

Now let’s examine the code (see figure 2) that generated the annotate dataset for this graph and explain what each piece is doing.

CODE

```plaintext
1   /* Create annotate dataset **/
2
3   data anno;
4     set relriskdata;
5
6     length text $30 color function $8;
7
8   /* display statistics */
9
10    function='label'; xsys='1'; ysys='1';
11      x=58; y=99;
12    color='black'; when='a'; position='6'; size=1.3;
```
EXPLANATION

The annotate facility in SAS/GRAPH is highly powerful and enables the user to customize the output to the needs of the study. The customization is done with the use of the annotate dataset which describes the function along with its attributes. In other words, it is a set of commands that gets applied to the graphics output.

The annotate dataset is very unique because each observation in the dataset contains complete instructions for drawing a plot or any other graphic. The three main components of the instruction are: what to draw (the action variable), where to draw (the positioning variable) and how to draw (the attribute variable). The annotate facility only looks for these predefined variables and ignores the presence of any other variables. The dataset is built carefully with the use of ‘OUTPUT’ statements after providing instructions for each functionality.

Typically, when an annotate dataset is defined there is no set statement. However, since the annotate dataset is no different than any other SAS dataset, we could use a SET statement to enable the ANNOTATE facility as is done in the above code.

The first section where function='label' in the above code (lines 10-14) is instructing SAS to output the text 'Trt A (n/N), Trt B (n/N)' in the upper right-hand corner of the graph. More specifically, the x and y values are controlling the coordinates of the text along the x and y axis, and the xsys and ysys values are controlling the units in which x and y are specified. The variables xsys and ysys are set equal to ‘1’ which instructs SAS that x and y are specified as a percentage of the distance from the origin to end of the axis. In other words, our text is plotted exactly 58% of the distance from the origin to the end...
of the x-axis, and 99% of the distance from the origin to the end of the y-axis. This is just one of several ways that x and y coordinates can be specified using the annotate facility. Another popular method is to specify exact coordinates along an axis, rather than expressing coordinates as a percentage distance along an axis.

The next two sections where function='label' in the above code (lines 15-25), is instructing SAS to output text that is constructed using values from the variables from the input dataset “relriskdata”. The x and xsys variables are again instructing SAS to output the text at an x-axis coordinate that is exactly 58% of the distance from the origin to the end of the x-axis. However, the y-coordinate is now defined using the value of the variable nparm from the input dataset “relriskdata”. The variable ysys is set equal to ‘2’ which instructs SAS that the value of y is an exact coordinate on the y-axis, as opposed to a percentage distance on the axis.

The section where function='symbol’ in the above code (lines 30-33), is the place where we specify that the size for the symbol is to determined based on the sample size ‘N’ which comes from the input dataset “relriskdata”. The max() function fine tunes the actual size of the symbol. The user can choose any number that best fits their need. In our case, we chose max(N/5000, 1500/5000) to accommodate all the data we were plotting. In the event that the sample size was less than 1500, the max() function ensures that the symbol size is at least 1500/5000 or 0.3. If the symbol size were any smaller, it would basically be too small to the extreme that we might not see it at all on the graph.

The final step in creating the plot is to reference the annotate dataset inside of a GPLOT procedure as shown below (see figure 3). The annotate dataset “anno” is referenced in line 9 as one of the options for the plot statement. The display characteristics of the x and y axis such as the label and the scale are controlled in lines 2 – 5.

```
1  symbol1 v=none i=join color=black r=100 width=12;
2  axis1 offset=(5,5) pct value=(a=0 j=r h=16pt) order=(1 to 11 by 1)
3  label=(h=16pt 'Subgroup');
4  axis2 offset=(10,10) pct value=(h=16pt) order=(0.5 to 1.75 by 0.25)
5  label=(h=16pt '<-------------- Relative Risk (RR) ------------->');
6
7  proc gplot data=ds1;
8  plot nparm*xvar=nparm / vreverse nolegend noframe vaxis=axis1
9                 haxis=axis2 annotate=anno
10             href=1 chref=black lhref=3 ;
11  format nparm ord.;
12  run;
13  quit;
```

**Figure 3**

**CONCLUSION**

The annotate facility in SAS/GRAPH is an extremely powerful tool to know and learn. A good understanding of the basics gives the user the ability to enhance the graph and take control of the expected output. The example given here is just one of the many instances demonstrating the power and use of the annotate facility.
SAS is a registered trademark of SAS Institute, Inc., Cary, NC.

Contact Information
Your comments and questions are valued and encouraged. Contact the authors at:

Meera G. Kumar          Jeffrey Cortez
Sanofi-Aventis          Sanofi-Aventis
200 Crossing Boulevard, PO Box 6890      200 Crossing Boulevard, PO Box 6890
Mail Box: BX2-406A           Mail Box: BX2-406A
Bridgewater, NJ 08807        Bridgewater, NJ 08807
Work Phone: (908)304-6478    Work Phone: (908)231-5051
Email: meera.kumar@sanofi-aventis.com  Email: jeffrey.cortez@sanofi-aventis.com