Paper 408-2011

Using SAS/GRAPH® Annotate to Create Customized Control Charts

Sailesh Vezzu, Educational Testing Service, Princeton, NJ

ABSTRACT

The Annotate facility in SAS/GRAPH is a very powerful tool to generate customized plots. This paper is a direct result of one of my projects at work, where I had more than one observation at the same point in the plot. In this paper, I introduce a creative way to represent multiple numbers of observations at various points in the plot and show how to implement this using SAS® Annotate. I also show how to use SAS Annotate to add custom notes for some points in the plot and to make some outlying observations stand out in the plot by using a different color or symbol.

INTRODUCTION

Several of the projects I am involved with in the Research and Development division at Educational Testing Service (ETS) include processing data that is received from various institutions and releasing statistical reports. The data analysis and nature of these reports is beyond the scope of this paper, but I will demonstrate the process used to keep track of the timelines for processing these reports. These projects require the data from the institution to be processed and the reports to be released within 35 business days of receiving the data.

This paper will present control charts that provide a way to present the timelines for processing these reports. The data in this paper is for demonstration purposes only and is not actual data.

Figure 1 below shows a sample input data set to the program.

Figure 1. Sample Input Data Set

A typical input data set for this program includes the type of study, the study release date, and the number of business days to process a study. As you can see from the data above, several studies get released on the same day. As a result, several observations could fall on the same point in the plot when plotting the data. There are a few ways to represent several observations at the same point in the plot. These include jittering and bubble plots. In bubble plots, the size of the bubble represents the number of observations at a given point. In this paper, I will show how to use the GPLOT procedure with SAS/GRAPH Annotate to draw a control chart with bubbles and highlight points in the plot that didn't meet the deadline.

PROC GPLOT WITH ANNOTATE

To draw the control chart, use PROC GPLOT to draw a line plot and use the ANNOTATE option to draw bubbles for points in the plot with more than one observation. You can also use the ANNOTATE option to draw lines for mean, lower, and upper control limit.

Once you know what you want to accomplish with the ANNOTATE option in SAS, the next step is to build an appropriate annotate data set. The annotate data set is like any other SAS data set, but with specific key variables and values. Each observation in the annotate data set gives instruction to SAS to draw graphics or perform a specific action. Examples of the instructions include draw a blue line from the point with co-ordinates (a, b) to the point with co-ordinates (c, d) and draw a circle with a specific radius at point (x, y) on the plot.

CONSTRUCTING THE ANNOTATE DATA SET

In this section, I will describe how to build the annotate data set in several DATA steps and put all of these data sets together using one DATA step by using a SET statement.

To draw bubbles of appropriate size when you have more than one observation at the same point in the plot, you first have to get the number of observations at a given point in the plot.

```
/* &x_var. and &y_var. are global macro variables assigned using %let statements */
/* %let x_var = StudyReleasedDate; %let y_var = nbd_processing; */
proc means data = pro n;
      var &y_var.;
      by &x_var. &y_var.;
      ods output summary = ns_for_each_point(rename=&y_var._N=n);
run;
```

The data set "ns_for_each_point" would have the number of observations at a given point in the plot. The next step is to create bubbles with the size of the bubbles proportional to the number of observations.

```
/* &deadline. is a global macro variable assigned using %let statement */
/* %let deadline = 35; */
data anno_bigger_dots;
      length function $15. color $15. text $100. style $10.;
      set ns_for_each_point;
      if n > 1;
      function = 'symbol';
      text='circle';
      style='none';
      x=&x_var.;    y =&y_var.;
      xsys='2';     ysys='2';
      size=n;       color='blue';
      position='5';
      if nbd_processing > &deadline. then color = 'red';
run;
```

Use the ODS output data set "ns_for_each_point" to construct this part of the annotate data set to draw bubbles with the size proportional to the number of the observations at a given point. Then limit the data set to the points on the plot with more than observation. The value of the 'function' variable in the annotate data set specifies the drawing action, which in this case tells SAS to draw a symbol, and the value of the variable 'text' tells SAS the symbol to draw, which in this case is a circle. Set the 'size' variable to the frequency of the number of observations for each point in the plot. The x and y variables of the data set indicate where the circle should be drawn, while the xsys and ysys variables tell SAS what co-ordinate system to use. To highlight the points in the plot that didn't meet the deadline, use an "IF" statement and change the color to red if the number of business days it took to process a study was more than 35 days.

The code below sets up the part of the annotate data set to label points in the plot with the number of observations that fall at the point.

```
/* &display_ns. is a global macro variable assigned using %let statement */
data anno_ns;
      set  ns_for_each_point;
      length function $15. color $15. text $100. style $10.;
      %if (%upcase(&display_ns.) = "Y") %then %do;
            function= 'label'; text=strip(n); color='black'; style='';  x=&x_var.; y
            =&y_var.;    xsys='2'; ysys='2';size=1; position='3';
            if nbd_processing > &deadline. then do; color = 'red'; size=1.5; end;
      %end;
run;
```

The lower and upper control limits of the control chart are computed as mean +/- 2*(sd). Use the MEANS procedure to compute the mean and standard deviation.

```
proc means data = pro n mean std;
      var &y_var.;
      ods output summary = ss1;
run;
data ss2(keep=n mean stddev lcl ucl);
      set ss1;
      n=&y_var._n;
      Mean = &y_var._Mean;
      Stddev = &y_var._stddev;
      lcl = max(0, (&y_var._Mean - (2 * &y_var._StdDev)));
      Ucl = &y_var._Mean + (2 * &y_var._StdDev);
run;
```

The code below shows how to construct the part of the annotate data set to draw a line for Mean. Use the value "move" for the "function" variable to move the cursor to (0, *Mean*), and then use the value "draw" for the "function" variable in the next observation to draw the line in the plot from (0, Mean) to (100, Mean).

```
data anno_other_lines(drop=n mean stddev lcl ucl);
      set ss2;
      length function $15. color $15. text $100. style $10.;
      function='move'; text='';color='';style='';x=0; y=Mean; xsys='1';
      ysys='2';size=.;position='5'; output;
      …
      function='draw'; text='';color='darkgrey';style='none';x=100; y=Mean; xsys='1';
      ysys='2';size=3;position='5'; output;
      …
run;
```

Drawing lines for lower control limit, upper control limit and the deadline is similar. Use the value "move" for the "function" variable to move the cursor to (0, value), and then use the value "draw" for the "function" variable to draw a line from (0, value) to (100, value).

To add legends and labels to the control chart, use the code below.

```
data anno_labels;
      …
      function='label'; text='Number of Studies Released'||strip(put(n, 8.0);
      …
run;
```

Use a DATA step to combine all the parts of the annotate data sets into one data set.

```
data anno;
      set anno_bigger_dots anno_ns anno_other_lines anno_labels anno_legend;
run;
```

To draw the control chart along with all the customized options, use ANNO= option in the PLOT statement of the PROC GPLOT.

```
options orientation=landscape;
symbol1 value = dot interpol=join c=blue h=0.5 l=1;
/* &path. (output file location), &fn. (name), &dt1.(date1), and &dt2.(date2) are global
macro variables assigned using %let statement */
axis1  order = (0 to 40 by 5)
            label = (angle=90 j=c height = 2 'Number of Business days');
axis2  order = ("&dt1."d to "&dt2."d by month)
            label = (height = 2 'Study Release Date');

ods pdf file = "&path.\&ofn._&dt1._&dt2..pdf" bookmarklist=none;
       proc gplot data =pro;
               plot &y_var. * &x_var./ vaxis=axis1 haxis=axis2 anno=anno;
       run;
ods pdf close;
```

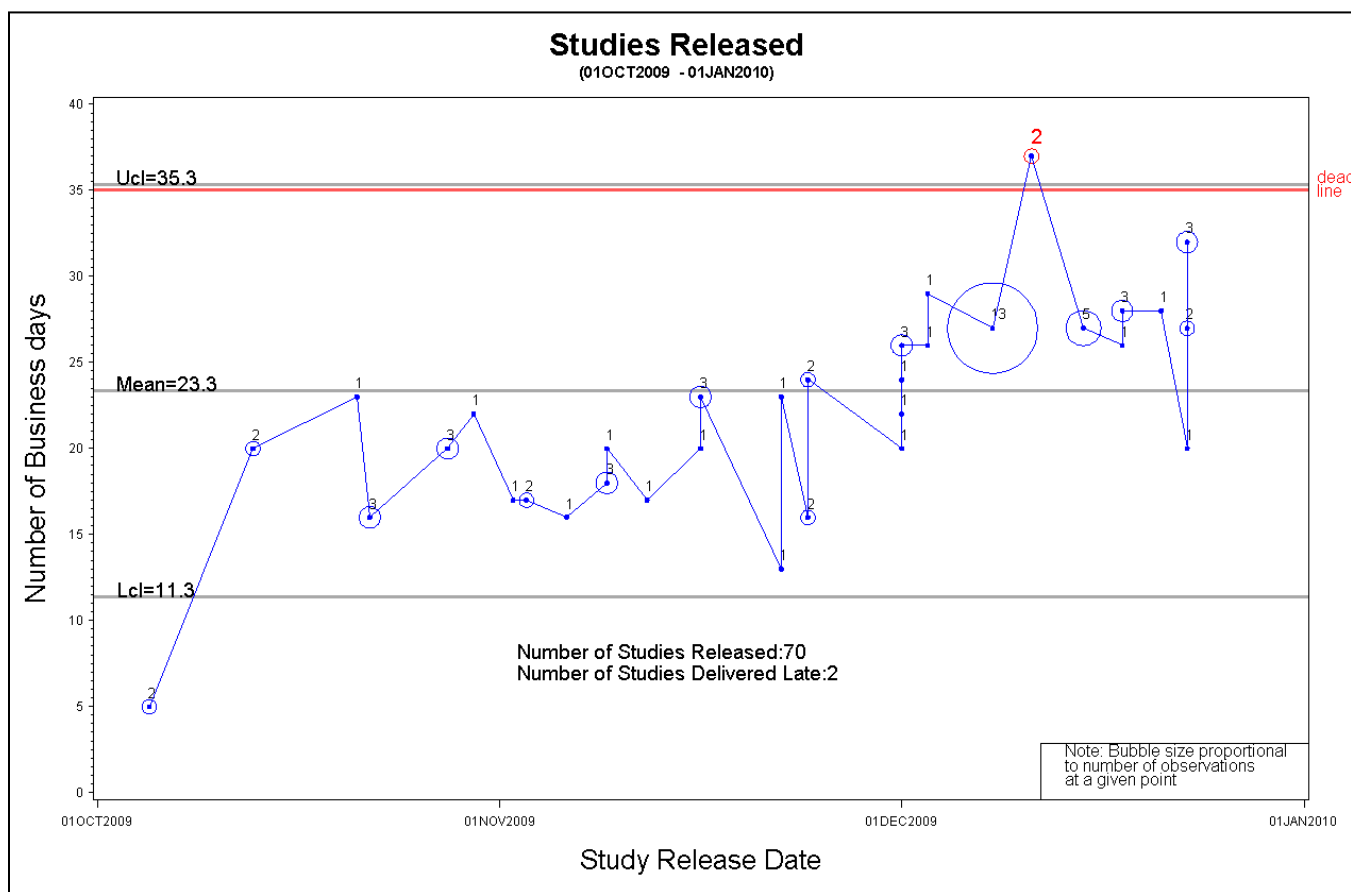Figure 2 below shows the control chart created using the PROC GPLOT along with the ANNO option.

Figure 2. Control Chart

CONCLUSION

This paper only shows a simple example of what can be done with SAS/GRAPH Annotate. Applying the same principle used in this paper, the Annotate facility can be used to customize and enhance output from any SAS/GRAPH procedure.

REFERENCES

SAS Institute Inc. 2010. *SAS/GRAPH® 9.2 Reference, Second Edition*. Cary, NC: SAS Institute Inc.

ACKNOWLEDGEMENTS

I would like to thank my colleagues and friends at ETS for supporting this paper and my professional development and my beautiful wife Meg for her endless support in all I do and her help with editing this paper.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:
 Sailesh Vezzu
 Educational Testing Service
 Rosedale Road, MS 20T
 Princeton, NJ 08541
 (609) 734 -5786
 svezzu@ets.org

TRADEMARKS