Creating Bar Charts and Scatter Plots on the Same Page, Deciphering SAS® ANNOTATE in PROC G PLOT
Rita Tsang, Averion Inc., Framingham, Massachusetts

ABSTRACT
The SAS GPLOT and GCHART procedures are powerful tools for generating graphical displays of clinical data, as they are usually employed to create scatter plots and bar charts respectively. However, we often need to display both a scatter plot and a bar chart within the same graph. The GCHART procedure cannot generate scatter plots and it does not have the ability to superimpose graphical displays. How do we get around this limitation?

Previously, my approach to the problem was to use a combination of the SAS DATA step and GPLOT procedure (Tsang, 2005). This paper graduates from the DATA step by demonstrating how the powerful SAS ANNOTATE can be used to easily accomplish the same task.

As a bonus, this paper will also explore the functionality of the new SAS version 9 PROC GBARLINE, which, was created to tackle the limitation of GCHART (SAS Institute, 2005).

INTRODUCTION
This paper will delineate two main approaches to create bar charts and scatter plots on the same graphical display, namely using DATA MANIPULATION and SAS ANNOTATE. The SAS ANNOTATE function will be deciphered to show how easy the task can be accomplished. The new GBARLINE procedure available in SAS version 9 will also be explored in this paper.

THE DATA
Our task at hand is to display the urine output data (Urine__) in a vertical bar chart and plasma creatinine data (Labval) in a scatter plot by assessment date (Asdt) for each patient. Here is a sample display of the Urine__ data.

<table>
<thead>
<tr>
<th>Site</th>
<th>Randno__</th>
<th>Asdt</th>
<th>Labval</th>
<th>Urine__</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1101</td>
<td>14MAR2004:20:00:00</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1101</td>
<td>15MAR2004:04:00:00</td>
<td>595</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1101</td>
<td>15MAR2004:04:20:00</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1101</td>
<td>15MAR2004:12:00:00</td>
<td>329</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1101</td>
<td>15MAR2004:14:45:00</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1101</td>
<td>15MAR2004:16:45:00</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1101</td>
<td>15MAR2004:18:45:00</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1101</td>
<td>15MAR2004:20:45:00</td>
<td>1.1</td>
<td>48</td>
</tr>
<tr>
<td>3</td>
<td>1101</td>
<td>15MAR2004:22:45:00</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1101</td>
<td>16MAR2004:00:45:00</td>
<td>53</td>
<td></td>
</tr>
</tbody>
</table>

1. DATA MANIPULATION
The following DATA step is a simple manipulation of the clinical data to create additional records of zero value for urine output (Urine1) and missing value for plasma creatinine (Labval) at each assessment time point. By creating the extra records, we can create vertical bars for the urine output data at each assessment time point by linking the extreme values, zero and the original values, using the INTERPOL=BOX00F option in the GPLOT SYMBOL statement. With the plasma creatinine values set to missing, SAS GPLOT simply will not display those observations.
data urine;
set urine__;
by site randno__ asdt;
urine1=urine__; output;
if urine__>. then do;
  urine1=0;
  labval=.; output;
end;
run;

In PROC GPLOT, we can produce box and whisker plots by specifying INTERPOL=BOX in the SYMBOL statement (SAS Institute, 2002). By varying the options in the INTERPOL=BOX <options>, the bottom and top edges of the box can be redrawn at a desired location. For example, INTERPOL=BOX00 will display a box ranges from the high and low extremes without displaying the whiskers.

Borrowing the idea of box plots, we can employ the following SYMBOL statement for the urine output, to link the original urine output value (high extreme) with the newly created records of zero value (low extreme). This SYMBOL statement will connect the high and low extreme values on the x-axis (I=BOX00F) with no plot symbol (V=NONE). The I=BOX00F option will fill the box with the color black (CV=BLUE). The BIWIDTH option specifies the width of the box. The width is about 0.4 cm in the current graphical display.

Symbol1 v=none i=box00f cv=blue bwidth=2;

CREATING A SCATTER- PLOT SUPERIMPOBED ON THE BAR-CHART: A SECOND PLOT STATEMENT
The SYMBOL statement for plasma creatinine is defined with the letter P as the plot symbol (V='P'). The symbols will be joined by a line in a scatter plot.

symbol2 h=0.17 in v='P' color=green cv=green i=join l=1;

The following axis statements are defined for the assessment date (horizontal axis: AXIS1), urine output (left vertical axis: AXIS2), and plasma creatinine (right vertical axis: AXIS3) respectively.

axis1 label=(h=0.17 in f=triplex "Assessment date and time") value=(h=0.135 in f=triplex) minor=none order=("14MAR2004:12:00:00"dt to "19MAR2004:12:00:00"dt by dtday) offset=(1 pct) color=black;
axis2 label=(h=0.17 in a=90 f=triplex "Urine Output (CC)") minor=none order=(0 to 2100 by 100) offset=(5 pct) value=(h=0.135 in f=triplex) color=black;
axis3 label=(h=0.17 in a=270 f=triplex "Plasma Creatinine (mg/dl)") minor=none order=(0 to 2 by 1) offset=(5 pct) value=(h=0.135 in f=triplex) color=black;

By employing two PLOT statements in GPLOT, we can superimpose a scatter plot of plasma creatinine on the bar chart of urine output with different scales on the same graph (SAS Institute, 2002). The first PLOT statement (PLOT URINE1*ASDT) displays the bar chart of urine output by assessment date. The second PLOT statement (PLOT2 LABVAL*ASDT) displays a scatter plot of plasma creatinine by assessment date. The OVERLAY option allows the second PLOT statement to be superimposed on the bar chart. Figure 1 shows the graphical display using DATA MANIPULATION

proc gplot data=urine;
plot urine1*asdt/ overlay noframe nolegend haxis=axis1 vaxis=axis2;
plot2 labval*asdt / noframe nolegend haxis=axis1 vaxis=axis3;
run;
quit;
2. SAS ANNOTATE

The alternative approach in creating bar charts is by employing the ANNOTATE data set to customize the graphical display in SAS GPLOT. In an ANNOTATE data set, each observation represents a command to draw a graphics element or to perform an action. The SAS ANNOTATE macros can be used within a DATA step to simplify the process of creating ANNOTATE observations (SAS Institute, 2002).

The following DATA step shows how an ANNOTATE data set can be easily created by using the BAR macro, %BAR (x1, y1, x2, y2, color, line, style), in the ANNOTATE facility.

```sas
%annomac;

data annot;
  set urine__;
  by site randno__ asdt;
  if urine__>.;
  retain xsys ysys '2';
  %bar(asdt-3600,0,asdt+3600,urine__,green,0,solid);
  format x datetime16.;
run;
```

The ANNOMAC macro needs to be first compiled (%ANNOMAC) before any ANNOTATE macros can be accessed in SAS. The ANNOTATE data set is named ANNOT. XSYS and YSYS are ANNOTATE variables which define the system coordinates to be used. The value 2 references the minimum and maximum of the data values as the major drawing area. The BAR macro has the following parameters X1, Y1, X2, Y2, COLOR, LINE, STYLE. This macro will draw a bar from Y1 (0 value) to Y2 (the actual Urine output value), with a width from X1 (assessment date/time – 3600 seconds) to X2 (assessment date/time + 3600 seconds). The width of the bars can be adjusted by changing the values of X1 and X2. The bar is filled by the color green with a solid outline, and the parameter LINE=0 draws an outline all around the bar.

By using the BAR macro, the values for the ANNOTATE variables are automatically created. Here is a sample display of the ANNOTATE data set:

<table>
<thead>
<tr>
<th>Xsys</th>
<th>Ysys</th>
<th>Color</th>
<th>X</th>
<th>Y</th>
<th>Function</th>
<th>Line</th>
<th>Style</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>green</td>
<td>14MAR04:19:00:00</td>
<td>0</td>
<td>MOVE</td>
<td>0</td>
<td>solid</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>green</td>
<td>14MAR04:21:00:00</td>
<td>27</td>
<td>BAR</td>
<td>0</td>
<td>solid</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>green</td>
<td>15MAR04:03:00:00</td>
<td>0</td>
<td>MOVE</td>
<td>0</td>
<td>solid</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>green</td>
<td>15MAR04:05:00:00</td>
<td>595</td>
<td>BAR</td>
<td>0</td>
<td>solid</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>green</td>
<td>15MAR04:11:00:00</td>
<td>0</td>
<td>MOVE</td>
<td>0</td>
<td>solid</td>
</tr>
</tbody>
</table>

In the ANNOTATE data set, the values for X and Y are created by the pairs of (X1, Y1) and (X2, Y2) values in the BAR macro. The MOVE function in the ANNOTATE facility will move the starting point to the specified (X, Y) values. The BAR function in the next observation will draw a green bar from the current (X, Y) position to the starting point defined in the previous observation.

In SAS GPLOT, we will not be plotting any line or symbol for the urine output since the ANNOTATE data set will be used to accomplish the task. The SYMBOL statement for the urine output will be defined as:

```sas
Symbol1 v=none i=none cv=green;
```

By employing the ANNOTATE= option in the first PLOT statement, the ANNOTATE data set ANNOT will be used to create bar chart for the urine output data. The second PLOT statement (PLOT2) will display the scatter plot of plasma creatinine. Figure 2 shows the graphical display using SAS ANNOTATE in PROC GPLOT:

```sas
proc gplot data=urine__;
  plot urine__*asdt/ annotate=annot overlay noframe nolegend haxis=axis1 vaxis=axis2;
  plot2 labval*asdt / noframe nolegend haxis=axis1 vaxis=axis3;
run;
quit;
```
3. PROC GBARLINE IN SAS VERSION 9

The new SAS GRAPH procedure GBARLINE may provide an effective solution to the limitation of superimposing bar charts and scatter plots on one graph for data with limited distinct groupings on the X-axis. The following procedure shows how a bar chart and a scatterplot can be displayed on the same output:

```sas
axis1 label=(h=0.17 in f=triplex "Assessment date and time")
value=(h=0.135 in f=triplex) minor=none value=(h=0.07 in f=triplex a=-45)
color=black;

proc gbarline data=urine__;
format asdt datetime16.;
bar asdt/ sumvar=urine__ width=2 discrete noframe space=1 axis=axis2
maxis=axis1;
plot /sumvar=labval axis=axis3;
run;
quit;
```

The BAR statement generates a bar chart of the urine output on the left axis using urine__ as the summary variable (SUMVAR=urine__) and the PLOT statement displays a scatter plot of the creatinine lab values on the right axis using labval as the summary variable (SUMVAR=labval). The DISCRETE option in the BAR statement generates a midpoint for each unique value of the assessment date/time. These midpoints are the values of the bar variable that identify categories of data on the X-axis.

In the GCHART procedure, the ORDER= option in the AXIS statement cannot be used for calculating midpoint values (SAS Institute, 2002). The same rule applies to the GBARLINE procedure. Therefore the ORDER= option in the AXIS statement for the assessment date/time (AXIS1) is omitted.

The MIDPOINTS= option in the BAR statement is not used here to specify midpoint values, e.g. MIDPOINTS= "14MAR2004:12:00:00"dt to "19MAR2004:12:00:00"dt by dday, because this option will only consolidate all data points based on the values specified in the value list.

Figure 3 shows the graphical display using PROC GBARLINE. Notice that although some data are collected on an irregular basis, the GBARLINE procedure displays each date/time by the same interval. As a result, we cannot see the trend of the safety data as in figures 1 and 2.

CONCLUSION

This paper has shown how a bar chart and a scatter plot can be created on the same graphical output using DATA MANIPULATION and SAS ANNOTATE in PROC GPLOT. The SAS version 9 PROC GBARLINE provides a viable alternative for accomplishing the same task for small distinct groupings of data or data that are evenly-spaced on the X-axis. For data that are clustered or irregularly spaced, the GPLOT procedure produces a more meaningful and interpretable graphical display.

REFERENCE


ACKNOWLEDGEMENT
The author would like to express her appreciation to the following individual for her invaluable comments and suggestions in this paper:

Pat Wesolowski, Averion Inc.

CONTACT INFORMATION
The author may be contacted at:
Rita Tsang
Averion Inc.
4 California Avenue
Framingham, MA 01701
(508)416-2694
Fax: (508)416-2797
E-mail: rtsang@averioninc.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.
Figure 1 Creating Bar Charts and Scatter Plots using Data Manipulation

Urine Output, and Creatinine Levels for Patient 1101

\[ P = \text{Plasma Creatinine} \]
Figure 2 Creating Bar Charts and Scatter Plots using SAS ANNOTATE

Urine Output, and Creatinine Levels for Patient 1101

P = Plasma Creatinine
Figure 3 Creating Bar Charts and Scatter Plots using PROC GBARLINE
Urine Output, and Creatinine Levels for Patient 1101

P = Plasma Creatinine