Graph Your SAS® Off
Karena Kong, InterMune Inc. Brisbane, CA.

ABSTRACT
This paper demonstrates three different SAS® procedures for creating graphs. For illustration purposes, the bubble plot, which is created by these procedures, shows the ratio of broadband users (DSL, Cable, Other) ranked by population ("List of countries"). The three procedures are SAS/Graph - GPLOT, Statistical Graphics (SG) – SGPLOT and Graphics Template Language (GTL) – PROC TEMPLATE with SGRENDER. This paper will discuss the advantages and disadvantages between each one. Based on the comparisons, it recommends which procedure should be used to create a similar graph.

INTRODUCTION
SAS® version 9.3 offers a wealthy collection of procedures to create graphs. A similar graph can be created by using the traditional SAS/GRAPH procedures, or the relatively new and easy to use Statistical Graphics (SG) procedures, or the more complex and flexible Graphics Template Language (GTL).

SAS makes graph creation easy as long as you know which procedure to use. SAS offers several different ways to complete this task. Well, why do I want to use SAS, you might ask? Why don’t I just plot the data in Excel or another program? That’s because, by using SAS to create graphs, you automate the programming process, reduce errors from mistyping; create consistent and easy to understand graphics that you can insert into any presentation. By using SAS, the graphs can be re-generated when input data changes. Furthermore using SAS to create graphs makes presenting data fun.

STATISTICAL GRAPHICS PROCEDURES

Figure 1. Graph created using SG procedures

Figure 1 was created using Statistical Graphics (SG) procedures. SG procedures begin with SG prefix, e.g. SGPLOT, SGSCATTER and SGPANEL. This procedure is available in SAS Version 9.2 and later. When creating Figure 1, three different plots were layered on top of one another.
1. A bubble plot was used to display the online population that is using 'DSL', 'Cable' or 'Other'. The size of the bubble is proportional to the ratio of technology used for that country.

2. To make it visually more appealing, a needle plot was used to plot up to the highest value thus taking on the appearance of reference lines in the graph.

3. A scatter plot is used for the data values along the bottom of the graph. Instead of using symbol characters, the MARKERCHAR option takes the data values from the dataset and places them on the graph.

In the previous example, it is advantageous that the data values are automatically plotted using the MARKERCHAR option. Therefore if the data ever changes, you only need to rerun the program, and the integrity of the graph remains. If a change in the appearance of the graph is desired, you can use SG to layer any other types of graphs, including histograms, bar charts, series plots and step plots.

Tip: I drew opaque bubbles after the needle plot to cover the line inside the bubble

```sas
PROC SGPLOT DATA=broaddata NOAUTOLEGEND;
NEEDLE X=country Y=maxnum /LINEATTRS=(COLOR=grey) TRANSPARENCY=.8;
BUBBLE X=country Y=num SIZE=size/group=type fill FILLATTRS=(color=white);
BUBBLE X=country Y=num SIZE=size/group=type fill TRANSPARENCY=.4 name='a';
XAXIS DISPLAY=all LABEL=' ' FITPOLICY=STAGGER;
YAXIS label='Percent Penetration' offsetmin=.17;
SCATTER X=country Y=online/ MARKERCHAR=total Y2AXIS;
Y2AXIS offsetmax=.87 LABEL=' ' DISPLAY=NONE;
REFLINE label /label LABELPOS=MIN AXIS=Y2 LINEATTRS=(THICKNESS=0);
KEYLEGEND 'a'/ location=inside position=topright noborder title='';
FOOTNOTE1 HEIGHT=1 JUSTIFY=LEFT COLOR=grey "Countries ranked by population. Bubble size is the ratio of broadband technology / Country's total online population";
RUN;
```

SAS/GRAPH GPLOT PROCEDURES

Figure 2. Graph created using SAS Graph procedures

SAS/Graph procedures are the traditional methods for creating graphs. It has been available in SAS since Version 6. To create a similar graph using SAS/Graph procedures requires use of the Annotate Facility. Figure 2 has two bubble statements to plot 'DSL', 'Cable'. The annotation dataset is created by using the %ANNOMAC macro to draw the 'Other', data labels, the line plot and legend. The disadvantage of using this method
is that every time data changes, you will have to manually update the annotation dataset which could lead to errors. The annotation dataset below contained seventy-six observations.

![Graph](image)

```sas
AXIS1 LABEL=NONE VALUE=(ANGLE=0) OFFSET=(5,5) ORDER=('China' 'USA' 'Japan' 'Germany' 'France' 'UK' 'South Korea' 'Russia' 'Brazil' 'India' 'Italy' 'Mexico' 'Spain' 'Canada' 'Turkey') width=1;
AXIS2 ORDER=(0 to 50 by 10) LABEL=NONE OFFSET=(3,3)
MAJOR=(HEIGHT=1.5) MINOR=(HEIGHT=1) WIDTH=1; SYMBOL1 INTERPOL=Needle
VALUE=NONE LINE=1 COLOR=black; LEGEND VALUE=(COLOR=blue HEIGHT=1);

PROC GPLOT DATA=broaddata ANNO=anno;
BUBBLE dsl*country=dsize /HAXIS=axis1 VAXIS=axis2
BFILL=SOLID BCOLOR=blue DATAORDER;
BUBBLE2 cable*country=csize/HAXIS=axis1 VAXIS=axis1
BFILL=SOLID BCOLOR=pink DATAORDER NOAXIS;
RUN;

%ANNOMAC
DATA anno; LENGTH function color style $8;
%system((4,2)) /*System Coordinates*/
%line(5,0,109.5,0, VOC900BB,1,1); /*Horizontal Reference Line*/
%line(10,0,10,5,0,VOC900BB,1,1); /*China Vertical Reference Line*/
%label(10,-1,'116.7',purple, 0,0,1,ariel);/*China Total Subscribers*/
%label(10,-3,'7.7',purple, 0,0,1,ariel);/*China Population Online*/
%line(16.8,0,16.8,16.4,V07900BB,1,1); /*USA Vertical Reference Line*/
%label(16.8,-1,'83.3',purple, 0,0,1,ariel);/*USA Total Subscribers*/
%label(16.8,-3,'27.2',purple, 0,0,1,ariel);/*USA Population Online*/
%slice(97.49, pie,360,.5,pink,solid,1); /*DSL Legend Label*/
%label(102.49,'DSL',purple, 0,0,1,ariel);
%slice(97.47, pie,360,.5,blue,solid,1); /*Cable Legend Label*/
%label(102.47,'Cable',purple, 0,0,1,ariel);
%slice(97.45, pie,360,.5,green,solid,1); /*Other Legend Label*/
%label(102.45,'Other',purple, 0,0,1,ariel);
RUN;
```

**Figure 3. The annotation dataset.**

**Tip:** X-axis tick values look much nicer horizontal

### GRAPHICS TEMPLATE LANGUAGE PROCEDURES

The graph below was created with the Graphics Template Language (GTL), which is available in SAS Version 9.2 or later. Figure 4 uses the LAYOUT OVERLAY option for the bubble and needle plot. I used the INNERMARGIN option to plot the data values in a BLOCKPLOT.

GTL is a very powerful tool that can create robust and complex graphs. There are options to overlay different graphs including those with different axes. You can control every attribute, color and label in the output. The only disadvantage being that it is not as simple as using SG procedures. There is a slight learning curve but once you understand GTL, you will have complete control to display your information with graphs.
Graph Your SAS Off, continued

**PROC TEMPLATE;**
DEFINE STATGRAPH bubbly;
BEGINGRAPH;
ENTRYTITLE 'Ratio of Broadband Users by Country and Population Penetration';
LAYOUT overlay / XAXISOPTS=( label=' ')
YAXISOPTS=(GRIDDISPLAY=on GRIDATTRS=(COLOR=lightgray PATTERN=dot)
OFFSETMIN=0.05 LABEL='Percentage Penetration');
BUBBLEPLOT X=country Y=num SIZE=size/NAME ='bubble' DATASKIN=sheen GROUP=type DATATRASPARENCY=0.3;
NEEDLEPLOT X=country Y=maxnum/LINEATTRS=(COLOR=grey);
INNERMARGIN;
BLOCKPLOT X=country BLOCK=total / DISPLAY=( values label)
CLASS=online REPEATEDVALUES=true VALUEVALIGN=center VALUEHALIGN=start;
ENDINNERMARGIN;
DISCRETELENGENED 'bubble' /ACROSS=3 BORDER=false LOCATION=inside HALIGN=right VALIGN=top;
ENDLAYOUT;
ENDGRAPH;
END;
RUN;

**PROC SGRENDER DATA=broaddata**
**TEMPLATE="bubbly";**
**RUN;**

Figure 4. Graph created using GTL procedures
Tip: The create the labels for on the Y2 axis in the SG Figure and the GTL figure, I created another variable 'Online'
see Appendix for dataset.

CONCLUSION
In general, a user should use SG procedures as much as possible to achieve the desired graphic presentation, since
SG procedures are quick and easy to learn in order to produce professional looking graphs. There are limitations
within SG procedures. A user will be able to create more complex graphs by using GTL with the SGRENDER
procedure. Although there is steep learning curve of GTL, it is well worth the time investment. There are so many
possibilities in GTL, which is outside the scope of this paper. Another advantage of GTL is that a user defined
template can be used to generate graphs from different source data by using SGRENDER procedure. SAS/GRAPH
procedures are still good alternatives, if the user already has SAS macros set up to generate common graphs.

So which method is the right one for you? Well, that depends on which version of SAS you are working with. If you
have SAS 9.1 or earlier, you have no choice but to use SAS Graph Procedures. If you have SAS 9.2 or later you
have two additional options. You can choose between SG Procedures or GTL. If you really want to get a grasp of
creating graphs and really enjoy a good challenge every now and then, try GTL. If you are on a deadline and just
need to get it done, use SG Procedures. Either way you will have a beautiful graph in the end.
REFERENCES

ACKNOWLEDGMENTS
I want to thank Elizabeth Li for being my mentor and my friend. Sanjay Matange for his valuable input.

CONTACT INFORMATION
Your comments and questions are valued and encouraged. Contact the author at:

- Name: Karena Kong
- Enterprise: InterMune
- Address: 3280 Bayshore Blvd.
- City, State ZIP: Brisbane, CA 940050
- Work Phone: 415-466-2442
- E-mail: kkong@intermune.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.
**APPENDIX**

The following dataset below was used in the SG and GTL figure

<table>
<thead>
<tr>
<th>country</th>
<th>type</th>
<th>num</th>
<th>size</th>
<th>maximum</th>
<th>pop</th>
<th>online</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>China</td>
<td>DSL</td>
<td>6.3</td>
<td>81.815181818</td>
<td>6.3</td>
<td>7.0</td>
<td>736895000</td>
<td>7.7</td>
</tr>
<tr>
<td>USA</td>
<td>DSL</td>
<td>10.9</td>
<td>40.073529412</td>
<td>14.6</td>
<td>3.0</td>
<td>90432043</td>
<td>27.2</td>
</tr>
<tr>
<td>Japan</td>
<td>DSL</td>
<td>7.3</td>
<td>27.756653992</td>
<td>14.8</td>
<td>1.9</td>
<td>2498291030</td>
<td>26.3</td>
</tr>
<tr>
<td>Germany</td>
<td>DSL</td>
<td>27.5</td>
<td>89137380192</td>
<td>27.9</td>
<td>7.1</td>
<td>714222444</td>
<td>31.5</td>
</tr>
<tr>
<td>France</td>
<td>DSL</td>
<td>23.7</td>
<td>94.599992621</td>
<td>29.7</td>
<td>6.0</td>
<td>601632000</td>
<td>31.4</td>
</tr>
<tr>
<td>UK</td>
<td>DSL</td>
<td>24.1</td>
<td>79.016399443</td>
<td>24.1</td>
<td>4.3</td>
<td>433747700</td>
<td>30.5</td>
</tr>
<tr>
<td>S. Korea</td>
<td>DSL</td>
<td>5.5</td>
<td>17.151162791</td>
<td>17.9</td>
<td>0.3</td>
<td>351032000</td>
<td>34.4</td>
</tr>
<tr>
<td>Russia</td>
<td>DSL</td>
<td>5.4</td>
<td>46.95621739</td>
<td>5.4</td>
<td>8.0</td>
<td>801632000</td>
<td>11.5</td>
</tr>
<tr>
<td>Brazil</td>
<td>DSL</td>
<td>4.7</td>
<td>56.348913589</td>
<td>4.7</td>
<td>7.7</td>
<td>77553000</td>
<td>8.1</td>
</tr>
<tr>
<td>India</td>
<td>DSL</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.9</td>
</tr>
<tr>
<td>Italy</td>
<td>DSL</td>
<td>20.9</td>
<td>98.120065728</td>
<td>20.9</td>
<td>2.9</td>
<td>2895436487</td>
<td>21.3</td>
</tr>
<tr>
<td>Mexico</td>
<td>DSL</td>
<td>7.6</td>
<td>77.227722727</td>
<td>7.6</td>
<td>0.9</td>
<td>945971323</td>
<td>10.1</td>
</tr>
<tr>
<td>Spain</td>
<td>DSL</td>
<td>18</td>
<td>81.09081081</td>
<td>18</td>
<td>1.9</td>
<td>154714794</td>
<td>22.2</td>
</tr>
<tr>
<td>Canada</td>
<td>DSL</td>
<td>13.1</td>
<td>43.521594684</td>
<td>16.5</td>
<td>5.9</td>
<td>123915701</td>
<td>30.1</td>
</tr>
<tr>
<td>Turkey</td>
<td>DSL</td>
<td>5</td>
<td>73.70721707</td>
<td>5</td>
<td>8.1</td>
<td>801632000</td>
<td>12.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>country</th>
<th>type</th>
<th>num</th>
<th>size</th>
<th>maximum</th>
<th>pop</th>
<th>online</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>China</td>
<td>Cable</td>
<td>14.6</td>
<td>53.67167598</td>
<td>1.2</td>
<td>12.6</td>
<td>1216935932</td>
<td>116.7</td>
</tr>
<tr>
<td>USA</td>
<td>Cable</td>
<td>4.2</td>
<td>15.99531749</td>
<td>1.4</td>
<td>3.2</td>
<td>140897432</td>
<td>33.5</td>
</tr>
<tr>
<td>Japan</td>
<td>Cable</td>
<td>3.2</td>
<td>10.203642173</td>
<td>1.2</td>
<td>0.3</td>
<td>81917552</td>
<td>25.5</td>
</tr>
<tr>
<td>Germany</td>
<td>Cable</td>
<td>1.6</td>
<td>5.089414013</td>
<td>1.6</td>
<td>0.3</td>
<td>324112200</td>
<td>20.3</td>
</tr>
<tr>
<td>France</td>
<td>Cable</td>
<td>6.4</td>
<td>20.983600575</td>
<td>1.8</td>
<td>4.4</td>
<td>120479280</td>
<td>18.8</td>
</tr>
<tr>
<td>UK</td>
<td>Cable</td>
<td>10.6</td>
<td>30.810953408</td>
<td>1.7</td>
<td>6.0</td>
<td>177956200</td>
<td>15.0</td>
</tr>
<tr>
<td>S. Korea</td>
<td>Cable</td>
<td>1.9</td>
<td>23.45793123</td>
<td>1.9</td>
<td>0.5</td>
<td>31390000</td>
<td>15.5</td>
</tr>
<tr>
<td>Russia</td>
<td>Cable</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.9</td>
</tr>
<tr>
<td>Brazil</td>
<td>Cable</td>
<td>2.2</td>
<td>21.782736219</td>
<td>2.2</td>
<td>0.3</td>
<td>238930056</td>
<td>10.8</td>
</tr>
<tr>
<td>Mexico</td>
<td>Cable</td>
<td>4.1</td>
<td>18.4584698</td>
<td>2.4</td>
<td>3.3</td>
<td>42073925</td>
<td>10.3</td>
</tr>
<tr>
<td>Spain</td>
<td>Cable</td>
<td>16.8</td>
<td>56.146173402</td>
<td>1.7</td>
<td>5.0</td>
<td>171344722</td>
<td>10.1</td>
</tr>
<tr>
<td>Canada</td>
<td>Cable</td>
<td>0.3</td>
<td>2.4390243902</td>
<td>0.3</td>
<td>0.3</td>
<td>27366285</td>
<td>5.1</td>
</tr>
</tbody>
</table>

The following dataset below was used in the GPLOT figure

<table>
<thead>
<tr>
<th>DSL</th>
<th>Cable</th>
<th>dsize</th>
<th>csize</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.3</td>
<td>0</td>
<td>81.818181818</td>
<td>0</td>
</tr>
<tr>
<td>10.9</td>
<td>14.6</td>
<td>40.073529412</td>
<td>53.676470588</td>
</tr>
<tr>
<td>7.3</td>
<td>4.2</td>
<td>27.756653992</td>
<td>15.969581749</td>
</tr>
<tr>
<td>27.9</td>
<td>3.2</td>
<td>89.137360192</td>
<td>10.223642173</td>
</tr>
<tr>
<td>29.7</td>
<td>1.6</td>
<td>94.585697261</td>
<td>5.095514013</td>
</tr>
<tr>
<td>24.1</td>
<td>6.4</td>
<td>79.016399443</td>
<td>20.983600575</td>
</tr>
<tr>
<td>5.9</td>
<td>10.6</td>
<td>17.151162791</td>
<td>30.813953488</td>
</tr>
<tr>
<td>5.4</td>
<td>0</td>
<td>46.95621739</td>
<td>0</td>
</tr>
<tr>
<td>4.7</td>
<td>1.9</td>
<td>58.024691356</td>
<td>23.456790123</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>20.9</td>
<td>0</td>
<td>38.1220657280</td>
<td>0</td>
</tr>
<tr>
<td>7.8</td>
<td>2.2</td>
<td>77.227722772</td>
<td>21.78278218</td>
</tr>
<tr>
<td>18</td>
<td>4.1</td>
<td>81.091081081</td>
<td>18.468484686</td>
</tr>
<tr>
<td>13.1</td>
<td>16.3</td>
<td>43.521594684</td>
<td>56.146179402</td>
</tr>
<tr>
<td>9</td>
<td>0.3</td>
<td>73.170731707</td>
<td>2.4390243902</td>
</tr>
</tbody>
</table>