Paper 309-4013

Graph Your SAS® Off

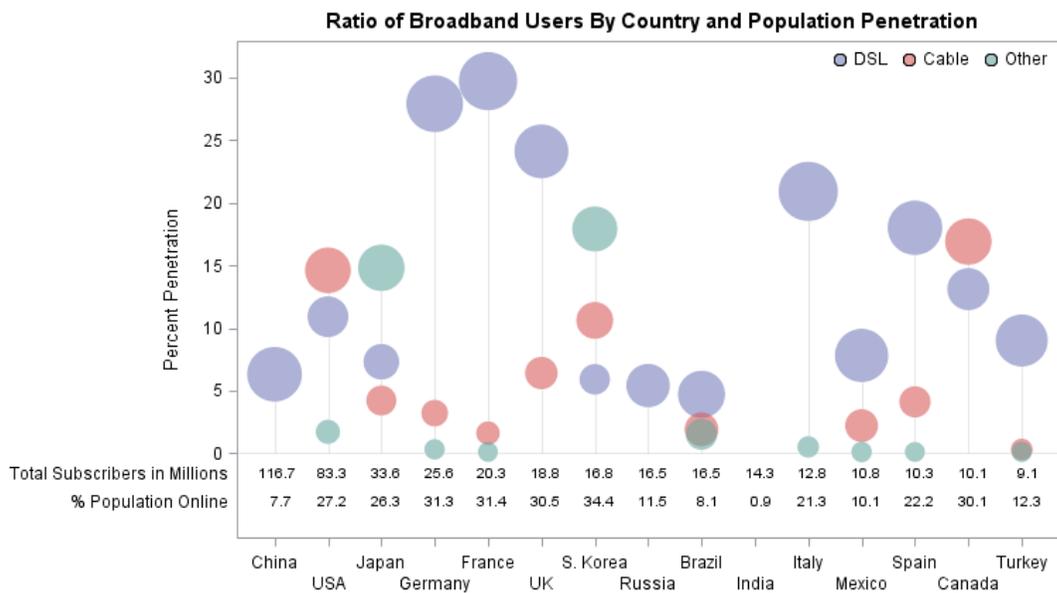Karena Kong, InterMune Inc. Brisbane, CA.

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

```
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";
```
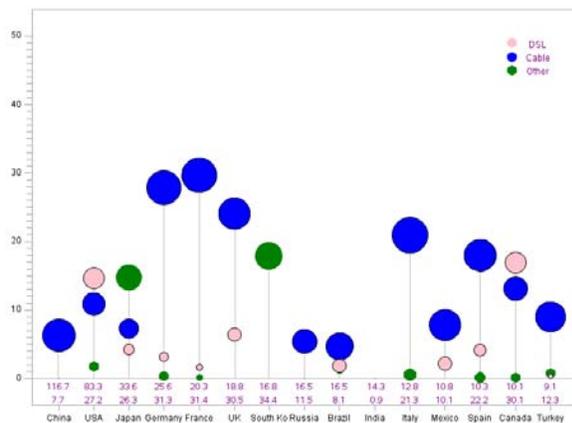
SAS/GRAPH GPLOT PROCEDURES

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.

```
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=axis2
  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,V07900BB,1,1);/*Horizontal
                          Reference Line*/
%line(10,0,10,6.3,V07900BB,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(16.8,0,16.8,14.6,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(16.8,1.7,pie,360,.5,green,solid,1);/*USA
                              other*/
%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;
```

function	color	style	XSYS	YSYS	HSYS	X	Y	LINE	SIZE	ANGLE	ROTATE	TEXT
MOVE	V07900BB		4	2		5	0	.	.	.	.	
DRAW	V07900BB		4	2		109.5	0	1	1	.	.	
MOVE	V07900BB		4	2		10	0	1	1	.	.	
DRAW	V07900BB		4	2		10	6.3	1	1	.	.	
LABEL	purple	ariel	4	2		10	-1	1	1	0	0	116.7
LABEL	purple	ariel	4	2		10	-3	1	1	0	0	7.7
MOVE	V07900BB	ariel	4	2		16.8	0	1	1	0	0	7.7
DRAW	V07900BB	ariel	4	2		16.8	14.6	1	1	0	0	7.7
LABEL	purple	ariel	4	2		16.8	-1	1	1	0	0	83.3
LABEL	purple	ariel	4	2		16.8	-3	1	1	0	0	27.2
MOVE	V07900BB	ariel	4	2		23.5	0	1	1	0	0	27.2
DRAW	V07900BB	ariel	4	2		23.5	14.8	1	1	0	0	27.2

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.

```
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
                       datatransparency=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;
                       discretelegend 'bubble' /across=3 border=false location=inside halign=right
                                                                                  valign=top;
            endlayout;
    endgraph;
end; run;
proc sgrender data=broaddata
template="bubbly";
run;
```
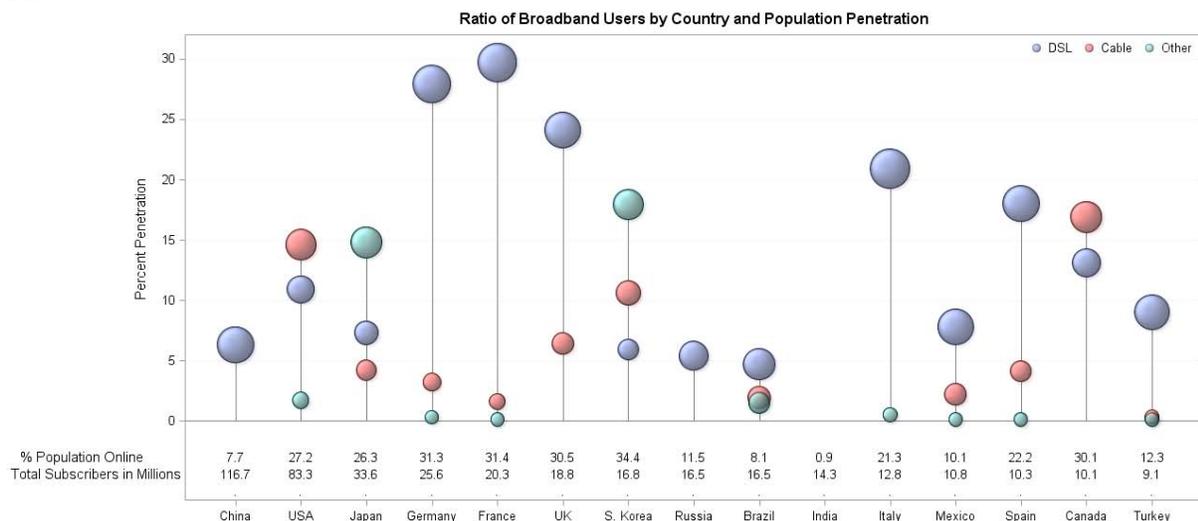
Figure 4. Graph created using GTL procedures

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.

ACKNOWLEDGMENTS

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

REFERENCES

"List of Countries by Number of Broadband Internet Subscriptions." *Wikipedia*. Wikimedia Foundation, 11 Jan. 2012. Web. 01 Nov. 2012.
<http://en.wikipedia.org/wiki/List_of_countries_by_number_of_broadband_Internet_subscriptions>.

CONTACT INFORMATION

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

> Karena Kong
> InterMune Inc.
> 3280 Bayshore Blvd.
> Brisbane, CA 94005
> 415-466-2442
> 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

country	type	num	size	maxnum	pop	online	total
China	DSL	6.3	81.818181818	6.3	734895000	% Population Online	7.7
USA	DSL	10.9	40.073529412	14.6	908459704.3	% Population Online	27.2
Japan	DSL	7.3	27.756653992	14.8	244825910.8	% Population Online	26.3
Germany	DSL	27.9	89.137380192	27.9	714222144	% Population Online	31.3
France	DSL	29.7	94.585987261	29.7	601632900	% Population Online	31.4
UK	DSL	24.1	79.016393443	24.1	453747570	% Population Online	30.5
S. Korea	DSL	5.9	17.151162791	17.9	99056103	% Population Online	34.4
Russia	DSL	5.4	46.956521739	5.4	89100000	% Population Online	11.5
Brazil	DSL	4.7	58.024691358	4.7	77550000	% Population Online	8.1
India	DSL	.	0	0	0	% Population Online	0.9
Italy	DSL	20.9	98.122065728	20.9	268545646.6	% Population Online	21.3
Mexico	DSL	7.8	77.227722772	7.8	84581733.6	% Population Online	10.1
Spain	DSL	18	81.081081081	18	184714794	% Population Online	22.2
Canada	DSL	13.1	43.521594684	16.9	132817507.1	% Population Online	30.1
Turkey	DSL	9	73.170731707	9	81888795	% Population Online	12.3
China	Cable	.	0	.	0	Total Subscribers in Millions	116.7
USA	Cable	14.6	53.676470588	.	1216835934.2	Total Subscribers in Millions	83.3
Japan	Cable	4.2	15.969581749	.	140858743.2	Total Subscribers in Millions	33.6
Germany	Cable	3.2	10.223642173	.	81917952	Total Subscribers in Millions	25.6
France	Cable	1.6	5.0955414013	.	32411200	Total Subscribers in Millions	20.3
UK	Cable	6.4	20.983606557	.	120497280	Total Subscribers in Millions	18.8
S. Korea	Cable	10.6	30.813953488	.	177965202	Total Subscribers in Millions	16.8
Russia	Cable	.	0	.	0	Total Subscribers in Millions	16.5
Brazil	Cable	1.9	23.456790123	.	31350000	Total Subscribers in Millions	16.5
India	Cable	.	0	.	0	Total Subscribers in Millions	14.3
Italy	Cable	.	0	.	0	Total Subscribers in Millions	12.8
Mexico	Cable	2.2	21.782178218	.	23856386.4	Total Subscribers in Millions	10.8
Spain	Cable	4.1	18.468468468	.	42073925.3	Total Subscribers in Millions	10.3
Canada	Cable	16.9	56.146179402	.	171344722.9	Total Subscribers in Millions	10.1
Turkey	Cable	0.3	2.4390243902	.	2729626.5	Total Subscribers in Millions	9.1
China	Other		0		0		

The following dataset below was used in the GPLOT figure

DSL	Cable	dsize	csize
6.3	0	81.818181818	0
10.9	14.6	40.073529412	53.676470588
7.3	4.2	27.756653992	15.969581749
27.9	3.2	89.137380192	10.223642173
29.7	1.6	94.585987261	5.0955414013
24.1	6.4	79.016393443	20.983606557
5.9	10.6	17.151162791	30.813953488
5.4	0	46.956521739	0
4.7	1.9	58.024691358	23.456790123
0	0	0	0
20.9	0	98.122065728	0
7.8	2.2	77.227722772	21.782178218
18	4.1	81.081081081	18.468468468
13.1	16.9	43.521594684	56.146179402
9	0.3	73.170731707	2.4390243902

5