ABSTRACT

Meaningful and interpretable visualization is critical to turn data into knowledge. Almost every software systems provide settings to create default graphs. Although creating default graphs provided by these software is straightforward, creating customized graphs for specialized analysis and specific interpretation are challenging. Furthermore, in most analysis, data are collected and summarized by groups. To create customized graph for group data for specific layouts and desired appearances are even more complicated. This paper provides a step by step, start-from-scratch approach to create a customized graph for group data using SAS® Graph Template Language (GTL).

INTRODUCTION

GTL is a powerful language to visualize data. An extension to the Output Delivery System (ODS), GTL is designed to create complex analytical graphics that are not available from traditional SAS procedures. GTL is powerful and flexible, but its power and flexibility come with some complexity. Syntax and style of GTL are perceived deviant from conventional SAS syntax, which makes most SAS users feeling not knowing enough to use GTL. When there are needs to create customized graphs, most people feel not sure how to start.

In fact, any complex graph starts from simple graph designs. Understanding the basic graphics structures and GTL language elements can enable us to establish a thinking process which can help us to first build points of references to start with, then moving from simple patterns to a highly complex graph. Following the thinking process and GTL language elements, to create a highly complex, customized graph from basic patterns are approachable.

STEP 1: ANALIZE INPUT DATASET AND TARGET GRAPH

Before starting writing codes to create plot, the first step is to examine the desired graph, figuring out the basic patterns then relating them to the data. For example, the target plot is to display means with confidence intervals at different visits for four individual groups from the dataset in Table 1. The dataset contains groups, visit weeks, means, lower and upper confidence intervals. In the plot, x-axis is the visit week, y-axis illustrates the means and confidence interval bands. To make the graph interpretable, each group for each visit week needs to form a cluster.
Table 1. Sample Data

<table>
<thead>
<tr>
<th>groupcd</th>
<th>visit</th>
<th>Mean</th>
<th>Lower CI</th>
<th>Upper CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Eastern</td>
<td>Week 1</td>
<td>45</td>
<td>25.955413024</td>
</tr>
<tr>
<td>2</td>
<td>Eastern</td>
<td>Week 2</td>
<td>40.875</td>
<td>25.137655134</td>
</tr>
<tr>
<td>3</td>
<td>Eastern</td>
<td>Week 3</td>
<td>40.5</td>
<td>23.940127575</td>
</tr>
<tr>
<td>4</td>
<td>Eastern</td>
<td>Week 4</td>
<td>38.25</td>
<td>20.551986459</td>
</tr>
<tr>
<td>5</td>
<td>Eastern</td>
<td>Week 5</td>
<td>27.5</td>
<td>12.776709223</td>
</tr>
<tr>
<td>6</td>
<td>Eastern</td>
<td>Week 6</td>
<td>22.5</td>
<td>10.510934271</td>
</tr>
<tr>
<td>7</td>
<td>Eastern</td>
<td>Week 7</td>
<td>20.5</td>
<td>7.8210450527</td>
</tr>
<tr>
<td>8</td>
<td>Eastern</td>
<td>Week 8</td>
<td>16.25</td>
<td>3.8755315022</td>
</tr>
<tr>
<td>9</td>
<td>Central</td>
<td>Week 1</td>
<td>93.75</td>
<td>74.367615403</td>
</tr>
<tr>
<td>10</td>
<td>Central</td>
<td>Week 2</td>
<td>84.125</td>
<td>76.789358769</td>
</tr>
<tr>
<td>11</td>
<td>Central</td>
<td>Week 3</td>
<td>91.375</td>
<td>73.62604454</td>
</tr>
<tr>
<td>12</td>
<td>Central</td>
<td>Week 4</td>
<td>80.75</td>
<td>71.377659488</td>
</tr>
<tr>
<td>13</td>
<td>Central</td>
<td>Week 5</td>
<td>53.75</td>
<td>41.947464206</td>
</tr>
<tr>
<td>14</td>
<td>Central</td>
<td>Week 6</td>
<td>51.125</td>
<td>29.620903841</td>
</tr>
<tr>
<td>15</td>
<td>Central</td>
<td>Week 7</td>
<td>39.75</td>
<td>21.27764903</td>
</tr>
<tr>
<td>16</td>
<td>Central</td>
<td>Week 8</td>
<td>29.5</td>
<td>13.64671895</td>
</tr>
<tr>
<td>17</td>
<td>Mountain</td>
<td>Week 1</td>
<td>95.75</td>
<td>89.974555968</td>
</tr>
<tr>
<td>18</td>
<td>Mountain</td>
<td>Week 2</td>
<td>92.375</td>
<td>85.861679287</td>
</tr>
<tr>
<td>19</td>
<td>Mountain</td>
<td>Week 3</td>
<td>92.25</td>
<td>88.526032625</td>
</tr>
<tr>
<td>20</td>
<td>Mountain</td>
<td>Week 4</td>
<td>93.875</td>
<td>86.545451344</td>
</tr>
<tr>
<td>21</td>
<td>Mountain</td>
<td>Week 5</td>
<td>79.5</td>
<td>71.686306891</td>
</tr>
<tr>
<td>22</td>
<td>Mountain</td>
<td>Week 6</td>
<td>73.25</td>
<td>53.900144804</td>
</tr>
<tr>
<td>23</td>
<td>Mountain</td>
<td>Week 7</td>
<td>55.5</td>
<td>25.898672214</td>
</tr>
<tr>
<td>24</td>
<td>Mountain</td>
<td>Week 8</td>
<td>23.875</td>
<td>4.1655193511</td>
</tr>
<tr>
<td>25</td>
<td>Pacific</td>
<td>Week 1</td>
<td>85</td>
<td>75.80371875</td>
</tr>
<tr>
<td>26</td>
<td>Pacific</td>
<td>Week 2</td>
<td>86</td>
<td>78.579319856</td>
</tr>
<tr>
<td>27</td>
<td>Pacific</td>
<td>Week 3</td>
<td>81.686866667</td>
<td>73.056086706</td>
</tr>
<tr>
<td>28</td>
<td>Pacific</td>
<td>Week 4</td>
<td>82.666666667</td>
<td>76.253709135</td>
</tr>
<tr>
<td>29</td>
<td>Pacific</td>
<td>Week 5</td>
<td>51.166666667</td>
<td>19.53324033</td>
</tr>
<tr>
<td>30</td>
<td>Pacific</td>
<td>Week 6</td>
<td>45.333333333</td>
<td>16.818898195</td>
</tr>
<tr>
<td>31</td>
<td>Pacific</td>
<td>Week 7</td>
<td>17.833333333</td>
<td>7.4443614988</td>
</tr>
<tr>
<td>32</td>
<td>Pacific</td>
<td>Week 8</td>
<td>12.166666667</td>
<td>5.8231162688</td>
</tr>
</tbody>
</table>

**STEP 2: UNDERSTAND THE BUILDING BLOCKS OF GTL**

Two procedures are essential building blocks in GTL: PROC TEMPLATE and PROC SGRENDER. PROC TEMPLATE defines basic plot types and their appearances; PROC SGRENDER associates the data to the graph template.

The structures of PROC TEMPLATE and SGRENDER are outlined as below:

```plaintext
proc template;
  Define statgraph template-name;
  BeginGraph / <options>;
    layout overlay;
      type-of-plot
```
DEFINE statement of PROC TEMPLATE creates specialized SAS templates, that are used for controlling the appearance of ODS output. STATGRAPH is a reserved graph template from PROC TEMPLATE. Running PROC TEMPLATE compiles and saves the template; running PROC SGRENDER creates actual graph.

**STEP 3: START FROM SCRATCH**

Analyzing the plot types, data we have and desired graph, we need a scatter plot to display means and their confidence interval bands, a series plot superimposes on the scatter plot and connects the means for each visit, and each group in same visit week need to form a cluster.

Following is the example to build basic plot types in GTL:

```sas
proc template;
    define statgraph meanplot;
        begingraph;
            layout overlay / xaxisopts=(griddisplay=on)
                             yaxisopts=(griddisplay=on);
                scatterPlot X=visit y=mean
                                  /YErrorUpper=eval(uppercl)
                                  YErrorLower=eval(lowercl);
            endlayout;
        endgraph;
    end;
run;

proc sgrender data=final template=meanplot;
run;
```

**STEP 4: ADD GROUP DATA**

Following is the syntax showing adding option 'group' after scatterplot statement:

```sas
proc template;
    define statgraph meanplot;
        begingraph;
            layout overlay / xaxisopts=(griddisplay=on)
                             yaxisopts=(griddisplay=on);
                scatterPlot X=visit y=mean
                                  /YErrorUpper=eval(uppercl)
                                  YErrorLower=eval(lowercl);
                group=group;
            endlayout;
        endgraph;
    end;
run;
```
STEP 5: PLACE DIFFERENT GROUPS IN ONE CLUSTER

In the following example, type=discrete and groupdisplay=cluster options are to create four separate lines in one cluster:

```sas
proc template;
    define statgraph meanplot;
    begingraph;
        layout overlay / xaxisopts=(griddisplay=on type=discrete) yaxisopts=(griddisplay=on);
        scatterPlot X=visit y=mean /group=groupcd
            YErrorUpper=eval(uppercl)
            YErrorLower=eval(lowercl);
        endlayout;
    endgraph;
end;
run;

proc sgrender data=final template=meanplot;
run;
```

STEP 6: CREATE A GAP BETWEEN EACH CLUSTER

To create gaps between groups, we use dummy datasets. Dummy datasets and variables in the dummy datasets are like virtual variables that create space in between lines thus create gaps but will not display images and values.

Figure 1 displays the plot after dummy datasets are set with the original dataset. Gaps between clusters for each visit are created:
**STEP 7: SUPERIMPOSE A SERIES PLOT TO CONNECT DOTS FOR EACH GROUP**

We want to have lines to connect the means of individual group for each week. One of GTL's most powerful feature is the flexibility to combine different types of plots together. To superimpose a series plot to the scatter plot, we just need to add a series plot statement:

```sas
proc template;
    define statgraph meanplot;
    begingraph;
        entrytitle '';
        layout overlay / xaxisopts=(griddisplay=on type=discrete) yaxisopts=(griddisplay=on);
            seriesplot x=visit y=mean/group=groupcd;
            scatterPlot X=visit y=mean/group=groupcd groupdisplay=cluster
            YErrorLower=eval(lowercl); YErrorUpper=eval(uppercl);
        endlayout;
    endgraph;
endproc;
```
Figure 2 presents a series plot superimposing on a scatter plots:

Figure 2.

**STEP 8: DISPLAY ATTRIBUTES**

Followings are examples for adding attributes such as ‘label’, ‘labelattrs’, ‘tickvalues’, and ‘tickdisplaylist’:

```plaintext
proc template;
  define statgraph meanplot;
  begingraph;
    layout overlay /xaxisopts=(griddisplay=on type=discrete
                          label="Visit"
                          labelattrs = (size=9pt))
     yaxisopts=(griddisplay=on
                label="% Maximum"
                labelattrs = (size=9pt)
    tickvalueattrs  = (size=9pt)
```
Figure 3 is the presentation after using the default attribute options:

```plaintext
linearopts=(viewmin=0 viewmax=120
tickvalueclist=(0 10 20 30 40 50 60 70 80
                90 100 110 120)
tickdisplaylist=('0' '10' '20' '30' '40'
                '50' '60' '70' '80' '90'
                '100' '110' '120'))
seriesplot x=visit y=mean/group=groupcd
  lineattrs=(thickness=2) name="SP";
ScatterPlot X=visit y=mean/group=groupcd
  markerattrs=(size=8)
  YErrorUpper=eval(uppercl)
  YErrorLower=eval(lowercl)
  groupdisplay=cluster;
  discretelegend "SP" /title="Area " exclude=('Group5' 'Group6');
endlayout;
endgraph;
end;
run;
```

Figure 3.
STEP 9: BRING IN PROFESSIONAL QUALITY WITH STYLE TEMPLATE

Default attributes are handy and easy to use, but when it comes to customize graph, default attributes usually fall short. GTL further provides ‘style template’ for the need of customizing appearance of graphs, especially for the case of group data in which further differentiation among groups are often requested. Generally, default attributes are not enough for individual group to display its own character. GTL ‘style template’ is a great tool to let each individual group having its own color and style.

Following is the structure of PROC TEMPLATE. Using GraphData1 – GraphDataN. Attributes such as colors, line patterns, and marker symbols for each group can customize and differentiate nicely:

```plaintext
proc template;
  define style styles.name;
    parent = styles.default;
    style GraphData1 from GraphData1 / markersymbol =
      linestyle =
      contrastcolor =
    ;
  run;
```

Although only four groups in the graph, two additional styles are needed for the dummy groups. To apply the style to a graph, STYLE= option is used in the ODS HTML statement to specify the style name.

Following are style template used to create Figure. 4:

```plaintext
proc template;
  define style styles.blue;
    parent = styles.default;
    style color_list from color_list
      "Abstract colors used in graph styles" / 'bgA'   = cxfcffff;
      style GraphData1 from GraphData6 / markersymbol = "circle"
        linestyle = 2
        contrastcolor = darkyellow
      ;
      style GraphData2 from GraphData5 / markersymbol = "Circle"
        linestyle = 3
        contrastcolor = black
      ;
      style GraphData3 from GraphData3 / markersymbol = "star"
        linestyle = 5
        contrastcolor = darkorange
      ;
      style GraphData4 from GraphData4 / markersymbol = "asterisk"
        linestyle = 43
        contrastcolor = darkred
```
; style GraphData5 from GraphData2 /
  markersymbol = "square"
  linestyle = 27
  contrastcolor = blue
; style GraphData6 from GraphData1 /
  markersymbol = "DiamondFilled"
  linestyle = 42
  contrastcolor = green
; end;
run;

STEP 10: FINAL ADJUSTMENT, ADD TITLES AND FOOTNOTES

Final size of the output graph can also customize when bringing ODS. Titles, footnotes, size of font, type of fonts can always defined easily. In discretelegend statement, the appearance of legend can customize; order=rowmajor and across=2 are to display legends in two rows.
Figure 4.

The codes below create final plot Figure 4:

```plaintext
ods graphics on / height=6.6in width=10in outputfmt=png;

title1 h=11pt j=c "Results" j=r "First 8 Week" font="Courier";
title2 h=8pt j=c "Product A" j=r "Department" font="Courier";

footnote h=8pt j=l "From Sept 1 to Oct 29, 2016" j=r "Page 1 of 1"
font="Courier";

proc template;
    define statgraph meanplot;
        begingraph;
            beginlayout overlay /
                xaxisopts=(griddisplay=off type=discrete
                           label="Visit"
                           labelattrs = (size=9pt))
                yaxisopts=(griddisplay=off
                           label="% Maximum"
                           labelattrs = (size=9pt)
                           tickvalueattrs = (size=9pt)
                           linearopts=(viewmin=0 viewmax=120
                                      tickvaluelist=(0 10 20 30 40 50 60 70 80
                                                    90 100 110 120)
                                      tickdisplaylist=('0' '10' '20' '30' '40'
                                                      '50' '60' '70' '80'
                                                      '90' '100' '110' '120'))
            );

            seriesplot x=visit y=mean/group=groupcd
                groupdisplay=cluster
                lineattrs=(thickness=2) name="SP";

            scatterplot x=visit y=mean/group=groupcd
                groupdisplay=cluster
                markerattrs=(size=8)
                YErrorUpper=eval(uppercl)
                YErrorLower=eval(lowercl)
                errorbarattrs=(thickness=2);

            discretelegend "SP" /title="Area" exclude=('Group4' 'Group5')
                order=rowmajor across=2;
        endlayout;
        endgraph;
    end;
run;

proc sgrender data=all template=meanplot;
run;
```
CONCLUSIONS

The above steps demonstrate how simple plain patterns of plots can evolve to a highly complicated, customized graph. Understanding GTL language structures, knowing patterns of plots available in GTL are the tools to build points of references to start with; to examine data and analyze target graph are the steps to shape our thinking process. With the step by step approach, we can use GTL to create customized graph when challenge occurs.

REFERENCES


Cartier, Jeff, SAS Institute INC., Cary, NC. A Programmer’s Introduction to the Graphics Template Language.

CONTACT INFORMATION

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

Elva Chen
Pharmacyclics, An AbbVie Company
995 E. Arques Avenue
Sunnyvale, CA 94085-4521
408-215-3100
echen@pcyc.com