PharmaSUG 2024 - Paper 10019-

Dressing Up a Spaghetti Plot with SG Attribute Maps

Yao Fang, Sanofi;

ABSTRACT

SG Attribute Maps are commonly used in Bar plots to show categorical outcomes. However, using SG Attribute Maps to control visual attributes in Spaghetti plot (ex: assign different colors attributes according to different x-axis values on the same line) poses unique challenges of interpretation and implementation. We want to prepare a Spaghetti plot over time with mean AT levels of each participant. Within the same participant, the plot has different colors for different time periods. We can specify the display attributes for the data of each dose by group.

In this scenario, existing problems include: 1) multiple independent lines or accumulative lines, 2) stacked lines. We propose a method in data processing step to solve these problems: creating dummy records to connect the points belong to different time periods within the same participant. Then using macro to draw the individual line for each participant on the same page. This plot is intuitive to interpretate. It provides visualization of AT levels which are within the therapeutic range on different time periods.

This paper introduces how to use SG Attribute Maps to control visual attributes in Spaghetti plot. It also provides how to prepare a pre-defined graph data for the SGPLOT procedures.

INTRODUCTION

Sometimes we need to provide Spaghetti plot that indicates the values changed along with time within one participant. Normally the plot will be displayed as several lines based on participants, and each line for each participant. whichis easy to be producted. We just need to use 'GROUP=SUBJID' option in SG procedures, and it will help to group by participants.

If we want to assign different colors attributes to different dose periods on the same line,what should we do? The methods I chose was SG Attribute Maps. Attribute map data sets are used in the SG procedures to associate data values with visual attributes. The ID and VALUE variables are required for every SG attribute map data set. VALUE specifies the group data value that is assigned to the attributes. The variable which could be used to map different dose periods will be the group data. We need to create a group data variable with a DATA step before using SG Attribute Maps in SG procedures. To connect different dose period lines, dummy records of group data variable for dummy nodes are required. For example, group data value for first line is 'dose1' and for second line is 'dose2'. The first line will not be connected to the second line due to different group data values. Therefore, dummy nodes need to be created to solve this issue. For example, the end of the first line will be repeated as dummy nodes but group data value changed to 'dose2', then the second line will extend to the dummy nodes, which is also the end of the first line. Creating dummy nodes is the key to connect several different dose period lines.

Another question, if SG Attribute Maps used in SG procedures, the ID and VALUE variables are required. If they are not found, a warning is written in the SAS log, and the data set is ignored. VALUE specifies the group data value that is assigned to the attributes. The value must be valid for the data group that is assigned in the plot statement with the GROUP= option. That means we cannot use 'GROUP=SUBJID' option to draw lines group by participants, only GROUP=VALUE is accepted in SG Attribute Maps. Here the difficulty is how to put all lines for all participants into one page. The method I used was to draw each line for each participant independently, then repeat for all participants by macro.

EXAMPLE DATA SET

The data set that is used to draw Spaghetti plot follows a very specific format. It should be set up in such a way that the SG procedures determines both the group data value and the dummy nodes, which to connect two individual dose period lines. The required data structure was chosen based on ease of programming and construction. It requires the following:

•One row per participant per time per group data, each participant contains several rows with different time and group data.

•Dummy records for dummy nodes should be created.

The code below will create a data set for a mock study that will be used in the examples in this paper. The codes in red are used to generate dummy records:

```
Data test1;
Subjid="Subject 1"; exdose="dose 1"; avisit="Month 1"; aval=15; n=1; output;
Subjid="Subject 1"; exdose="dose 1"; avisit="Month 2"; aval=1; n=1; output;
Subjid="Subject 1"; exdose="dose 1"; avisit="Month 3"; aval=13; n=1; output;
Subjid="Subject 1"; exdose="dose 2"; avisit="Month 3"; aval=13; n=1; output;
Subjid="Subject 1"; exdose="dose 2"; avisit="Month 4"; aval=1; n=1; output;
Subjid="Subject 1"; exdose="dose 2"; avisit="Month 5"; aval=15; n=1; output;
Subjid="Subject 1"; exdose="dose 2"; avisit="Month 6"; aval=19; n=1; output;
Subjid="Subject 1"; exdose="dose 3"; avisit="Month 6"; aval=19; n=1; output;
Subjid="Subject 1"; exdose="dose 3"; avisit="Month 7"; aval=12; n=1; output;
Subjid="Subject 1"; exdose="dose 3"; avisit="Month 8"; aval=5; n=1; output;
Subjid="Subject 1"; exdose="dose 3"; avisit="Month 9"; aval=16; n=1; output;

Subjid="Subject 2"; exdose="dose 1"; avisit="Month 1"; aval=5; n=2; output;
Subjid="Subject 2"; exdose="dose 1"; avisit="Month 2"; aval=11; n=2; output;
Subjid="Subject 2"; exdose="dose 1"; avisit="Month 3"; aval=3; n=2; output;
Subjid="Subject 2"; exdose="dose 2"; avisit="Month 3"; aval=3; n=2; output;
Subjid="Subject 2"; exdose="dose 2"; avisit="Month 4"; aval=10; n=2; output;
Subjid="Subject 2"; exdose="dose 2"; avisit="Month 5"; aval=5; n=2; output;
Subjid="Subject 2"; exdose="dose 2"; avisit="Month 6"; aval=9; n=2; output;
Subjid="Subject 2"; exdose="dose 3"; avisit="Month 6"; aval=9; n=2; output;
Subjid="Subject 2"; exdose="dose 3"; avisit="Month 7"; aval=2; n=2; output;
Subjid="Subject 2"; exdose="dose 3"; avisit="Month 8"; aval=15; n=2; output;
Subjid="Subject 2"; exdose="dose 3"; avisit="Month 9"; aval=6; n=2; output;
run;
```

Figure 1 is a Snapshot of the data set test1 that is created by the previous code. The records in red frame are dummy records for dummy nodes:

	Subjid	exdose	avisit	aval	n
1	Subject 1	dose 1	Month 1	15	1
2	Subject 1	dose 1	Month 2	1	1
3	Subject 1	dose 1	Month 3	13	1
4	Subject 1	dose 2	Month 3	13	1
5	Subject 1	dose 2	Month 4	1	1
6	Subject 1	dose 2	Month 5	15	1
7	Subject 1	dose 2	Month 6	19	1
8	Subject 1	dose 3	Month 6	19	1
9	Subject 1	dose 3	Month 7	12	1
10	Subject 1	dose 3	Month 8	5	1
11	Subject 1	dose 3	Month 9	16	1
12	Subject 2	dose 1	Month 1	5	2
13	Subject 2	dose 1	Month 2	11	2
14	Subject 2	dose 1	Month 3	3	2
15	Subject 2	dose 2	Month 3	3	2
16	Subject 2	dose 2	Month 4	10	2
17	Subject 2	dose 2	Month 5	5	2
18	Subject 2	dose 2	Month 6	9	2
19	Subject 2	dose 3	Month 6	9	2
20	Subject 2	dose 3	Month 7	2	2
21	Subject 2	dose 3	Month 8	15	2
22	Subject 2	dose 3	Month 9	6	2

Step2: In order to draw lines group by participants in PROC SGPLOT step, different time variables (avisit_1 for Subject 1, avisit_2 for Subject 2 and so on) are created to be x-axis value:

```
 data test2;
     set test1;
          array tab[*]$ avisit_1-avisit_2;
               do i=1 to dim(tab) ;
                    if n=i then tab[i]=avisit;
               end ;
               output;
run;
```

	Subjid	exdose	avisit	aval	n	avisit_1	avisit_2
1	Subject 1	dose 1	Month 1	15	1	Month 1	
2	Subject 1	dose 1	Month 2	1	1	Month 2	
3	Subject 1	dose 1	Month 3	13	1	Month 3	
4	Subject 1	dose 2	Month 3	13	1	Month 3	
5	Subject 1	dose 2	Month 4	1	1	Month 4	
6	Subject 1	dose 2	Month 5	15	1	Month 5	
7	Subject 1	dose 2	Month 6	19	1	Month 6	
8	Subject 1	dose 3	Month 6	19	1	Month 6	
9	Subject 1	dose 3	Month 7	12	1	Month 7	
10	Subject 1	dose 3	Month 8	5	1	Month 8	
11	Subject 1	dose 3	Month 9	16	1	Month 9	
12	Subject 2	dose 1	Month 1	5	2		Month 1
13	Subject 2	dose 1	Month 2	11	2		Month 2
14	Subject 2	dose 1	Month 3	3	2		Month 3
15	Subject 2	dose 2	Month 3	3	2		Month 3
16	Subject 2	dose 2	Month 4	10	2		Month 4
17	Subject 2	dose 2	Month 5	5	2		Month 5
18	Subject 2	dose 2	Month 6	9	2		Month 6
19	Subject 2	dose 3	Month 6	9	2		Month 6
20	Subject 2	dose 3	Month 7	2	2		Month 7
21	Subject 2	dose 3	Month 8	15	2		Month 8
22	Subject 2	dose 3	Month 9	6	2		Month 9

ATTRIBUTE MAP DATA SETS

Attribute map data sets are used in the SG procedures to associate data values with visual attributes. When we have determined which attributes we want to apply to the group data in a graph, we can create a SG attribute map data set. Each observation defines the attributes for a particular data group. An observation uses reserved variable names for the attribute map identifier (ID), the group value (VALUE) and the attributes.

We create the data set using the same methods that we use to create any SAS data set. The main distinctions are that the SG attribute map data set uses reserved keywords for its variable names, and each observation represents the attributes for a particular data group. The most common used method for creating data sets is with a DATA step.

Here is an example of an SG attribute map data set called discreteMap. The observations in this data set contain the attribute map identifier (ID), the group value (VALUE) and the attributes (LINECOLOR, LINETHICKNESS).

This DATA step creates the SG attribute map data set discreteMap. The ID values for the attribute map are ID.

```
Data discreteMap;
    length value LineColor $8. ;
    Id='X'; Value="dose 1"; LineColor='red'; LineThickness=3; output;
        Id='X'; Value="dose 2"; LineColor='yellow'; LineThickness=4; output;
```

```
Id='X'; Value="dose 3"; LineColor='green'; LineThickness=5; output;
run;
```

Figure 2 is a table of the SG Attribute Map Data Set discreteMap.

	value	LineColor	Id	LineThickness
1	dose 1	red	X	3
2	dose 2	yellow	X	4
3	dose 3	green	X	5

The data set has these characteristics:

1. The value of the ID variable is referenced in one or more plot statements within the procedures.
2. The values of the VALUE variable are valid data group values. These values are case sensitive. The data group is assigned in the plot statement with the GROUP= option.
3. The values for LINECOLOR= is valid SAS colors. We can specify colors using the same color schemes that are supported by SAS/GRAPH software.

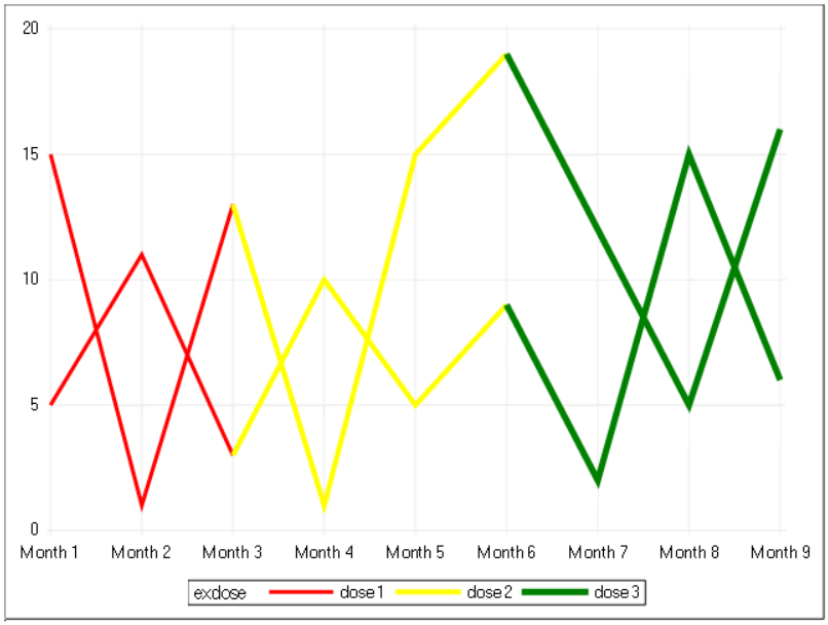
When a SG attribute map data set is processed, the SG procedures looks for the values of specific variables in the SG attribute map data set. The procedures uses these values to associate visual attributes with group data. Variables in the SG attribute map data set have predefined names. In each observation, the procedures looks only for variables with those names. Other variables can be present, but they are ignored.

GRAPHICAL COMPONENTS

This procedures generates the graph. The SGPLOT statement references the previous test2 data set and the discreteMap attribute map data set. The plot statement references the attribute map ID (ID). The options that reference the attribute map (data set and ID) are highlighted.

```
ods graphics / reset attrpriority=color;
proc sgplot data=test2 noborder dattrmap=discreteMap subpixel;
 %macro sub(i=);
   %if avisit_&i. ne . %then %do;
     series x=avisit_&i. y=aval / group=exdose attrid=X;
          %end;
 %mend;
 %sub(i=1);%sub(i=2);
  xaxis display=(noline noticks nolabel) grid;
  yaxis display=(noline noticks nolabel) grid;
run;
```

Here I use macro(%sub) to draw individual line for each participant respectively, the advantage is we can check each line for each participant by controlling macro to run.



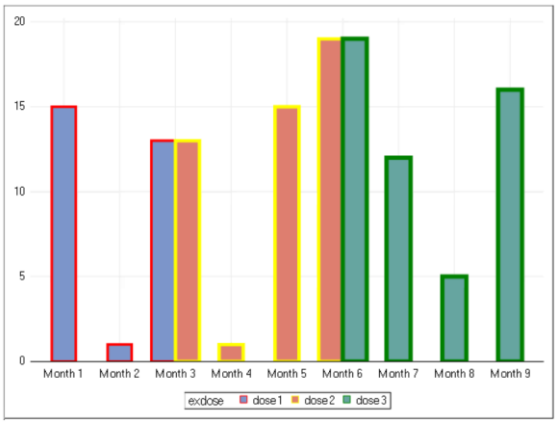
ALL PLOT STATEMENTS SUPPORT ATTRIBUTE MAPPING

In the SGPLOT and SGPANEL procedures, any plot statement that supports GROUP variables supports attribute mapping. In these procedures, all plot statements support attribute mapping except the following: DENSITY, HISTOGRAM, KEYLEGEND and REFLINE in SGPANEL procedures; DENSITY, ELLIPSE, HISTOGRAM, INSET, KEYLEGEND and REFLINE in SGPLOT procedures.

The example data used for previous Spaghetti plot was also used in below chart and panel plots.

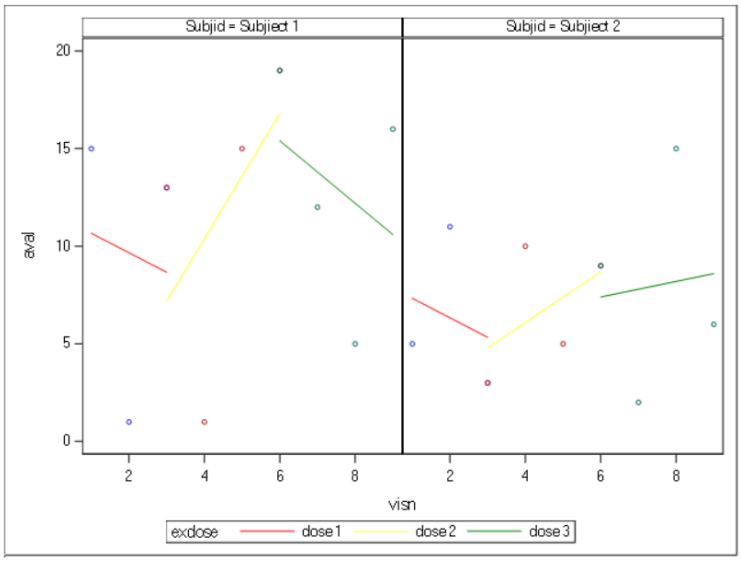
Chart:

```
ods graphics / reset attrpriority=color;
proc sgplot data=test2 noborder dattrmap=discreteMap subpixel;
    vbar avisit_1 / response=aval group=exdose groupdisplay=cluster attrid=X;
  xaxis display=(noline noticks nolabel) grid;
  yaxis display=(noline noticks nolabel) grid;
run;
```



Panel:

```
ods graphics / reset attrpriority=color;
proc sgpanel data=test2 dattrmap=discreteMap;
panelby subjid;
reg x=visn y=aval / group=exdose attrid=X;
run;
```



CONCLUSION

This paper introduced how to use SG Attribute Maps to control visual attributes in Spaghetti plot. The paper provides a simple and easy approach, and an overall idea of the data sets that are required for graph. The methods described in this paper can be applied to a large population to see the whole trend for all subjects under the influence of Investigational Medicinal Product by different dose regimen .

ACKNOWLEDGMENTS

Many thanks to Fan Zhang and Alice Chen, for their unfailing patience, inspiration, and technical assistance!

CONTACT INFORMATION

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

Yao Fang
Sanofi
Yao.fang@sanofi.com