

## PAPER DG04

# Annotate your SG Procedure Graphs

Mekhala Acharya, EMMES Corp., Rockville, MD

## ABSTRACT

In SAS®9.2, the SAS/Graph SG procedures offered a plethora of new options. One notable exception was the Annotate facility. Often graphs are needed with additional informative text and enhancements which are data driven. The use of the SAS/Graph ODS Graphics Editor is not recommended for annotation in clinical research, although it can be used effectively to edit and annotate graphs that are created by a wide variety of SAS procedures. This paper shows the effectiveness of using alternate SAS procedures to annotate. Sample code has been provided to show how a lot of the annotate code of traditional SAS graphs can be replaced using very little code. Code reduction to nearly one-third has been seen when using alternate methods to Annotate.

## INTRODUCTION

Annotate has been traditionally used with SAS graphics when additional enhancements to graph output was needed. You need to generate a special data set of graphics commands from which you can produce graphics output. This data set is referred to as an 'Annotate dataset'.

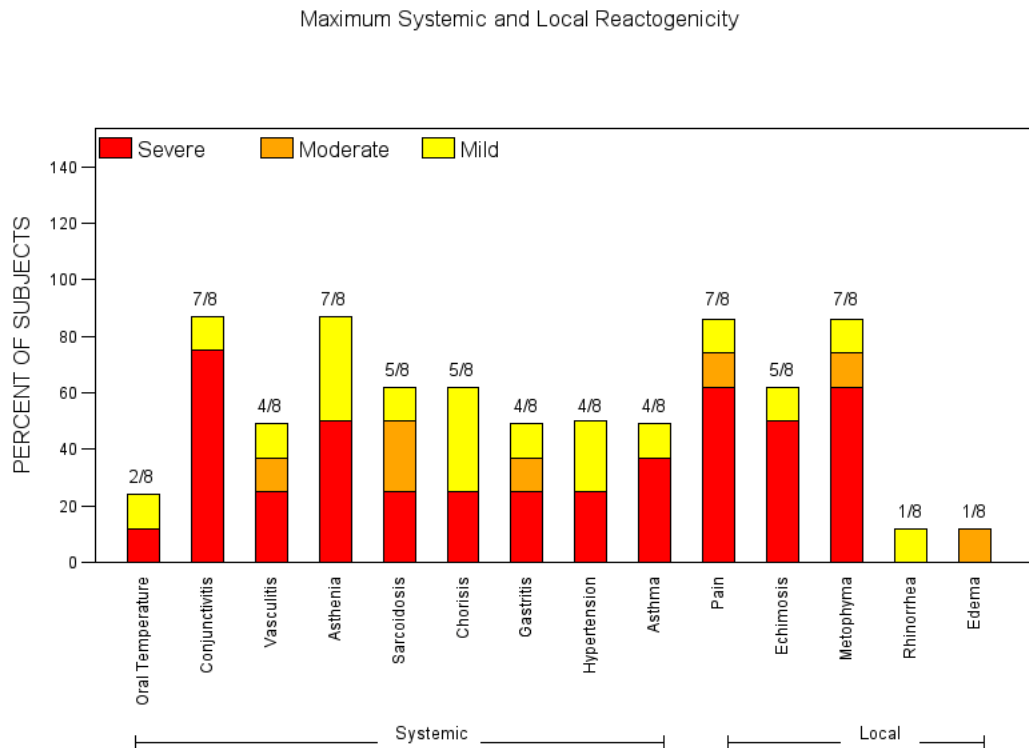
SAS 9.2, introduced statistical procedures to support ODS graphics to create graphs in the context of a specific analysis. GTL (Graphics template language) provided a powerful syntax for creating custom graphs. The graph templates provided by SAS System could be modified for use with SAS/STAT procedures. Both SG procedures and GTL play a vital role in the exploration of data and in the construction of specialized displays for analysis. SG procedures, including SGPLOT, SGPANEL, SGSCATTER and SGRENDER, enable users to easily create many types of graphs.

1. PROC SGPLOT creates plots overlaid with statistical features on a single set of axes
2. PROC SGPANEL creates paneled displays for the values of one or more classification variables
3. PROC SGSCATTER creates scatter plot matrices depending on the plot statement that you use

Now with SAS 9.3, we can add custom labeling or other graphical features to the SG procedures by using annotation support in SG procedures.

## ANNOTATE BASICS

In traditional SAS graphs annotate works well giving us the flexibility to add items to make our output perfect. Consider the following output.



**Figure 1: Using PROC GCHART and Annotate**

Two important features are needed:

1. The frequency counts on top of each bar.
2. The division between systemic and local events.

The annotate code for the above graph requires that the events be divided by type on the x-axis with labels 'Systemic' and 'Local'. The calculation of frequency of the most severe reactions is fairly easy, but the placement on the bar needs annotate features and so does the classification between systemic and local reactions. The code for placing the labels on the x-axis is shown below.

If you wanted to label each point in the plot with the person's name, you could construct an ANNOTATE data set as shown below. In this data step, FUNCTION = 'LABEL' writes the TEXT at the position given by X and Y in the plot.

<pre>DATA _anno;   set class;   xsys = '2'; ysys= '2';   x = height; y = weight;   function='label';   text=name;   position='2'; PROC PRINT;</pre>	<p>The annotate dataset is then called in PROC GCHART</p>	<pre>PROC GCHART data=;   vbar / discrete   anno=_anno   subgroup=severity   type=freq   freq=pct   space=4   raxis=axis1   maxis=axis2   legend=legend1;</pre>
---	---	---

Similarly other functions such as 'MOVE' and 'DRAW' are used to draw the graphic element. Additional goptions are needed for color coding the severity and labeling the axis.

```
goptions reset=all;
goptions reset=goptions ftext="arial" rotate=portrait
  htext=3 hsize=10 in vsize=6 in
  gsfmode=replace prompt gunit=pct gsfname=figs device=png;
filename figs "figname.png";

legend1 label=none value=(justify=1 h=3.0 'Severe' 'Moderate' 'Mild')
  position=(top left inside) across=3 shape=bar(2.5,2.5) ;

pattern1 v=solid c=red;
pattern2 v=solid c=orange;
pattern3 v=solid c=yellow;

axis1 order=(0 to 150 by 20) major=(h=1) minor=none
  label=(h=3.0 angle=90 'PERCENT OF SUBJECTS') v=(h=2.0) offset=(, 5 pct);
axis2 label=none order=(1 to 14) offset=(, 5 pct)
  value=(angle=90 rotate=0 h=2.0
'      Oral Temperature' 'Conjunctivitis' 'Vasculitis' 'Asthenia' 'Sarcoidosis' 'Chorisis'
'Gastritis' 'Hypertension' 'Asthma' 'Pain' 'Echimosiis' 'Metophyma' 'Rhinorrhea' 'Edema');
Title1 "Maximum Systemic and Local Reactogenicity";
Title4 ' ';
```

There is no doubt that the annotate facility provides nearly endless capability for customizing graphic output. However, a complete understanding of coordinate systems and units for coordinate values is essential for successfully creating an annotate dataset.

## PROPOSED METHODS

We will look at two different methods. Both methods use PROC SG PANEL to generate the graphs.

### METHOD 1

The analysis data set is used to create the static format values for the frequency of the severity of each reactogenicity event. The analysis data set looks like this:

SYMPTOM	TOTAL	SPECIFIC	SEVERITY	PCT	CLASS
Sarcoidosis	8	5	SEVERE	25	SYSTEMIC
Rhinorrhea	8	1	MILD	12.5	LOCAL

where 'total' is the total number of symptoms, 'specific' refers to count of number severity specific events.

<p>Using static values of the analysis data set, the values of pct can be formatted to obtain the fractions to be placed above each of bars of the bar chart.</p>	<pre>PROC FORMAT value st      low-0='1/8'               1-15 ='2/8'               20-30='4/8'               50-60='5/8'               60-70='7/8'               ; run;</pre>
---	---

This works on the assumption that the data is either frozen or will not undergo any further changes. To overcome this we look at method 2.

### METHOD 2

Consider the same analysis data set.

SYMPTOM	TOTAL	SPECIFIC	SEVERITY	PCT	CLASS	TEXT
Sarcoidosis	8	5	SEVERE	25	SYSTEMIC	5/8
Rhinorrhea	8	1	MILD	12.5	LOCAL	1/8

1. The text value is computed by taking the ratio of the specific event and the total number of reactogenicity events.
2. We create a second SAS data set that holds one observation for each unique value of the text.
3. Next we use the data from the new SAS data set to create format values. There are three variables that are required in the formatting data set: FMTNAME, START, and LABEL.
4. Instead of specifying a starting value of 'LOW' and an ending value of 'HIGH' we need to use a different variable called 'HLO'.

NOTE: Here we use HLO='o' which is Other (both Start and End values are ignored)  
For the data set above we create a format \$ST is SAS. This dynamically changes the format as the data in the analysis data set changes for every run of the SAS program.

```
DATA work.sevfmt;
  length label $ 8;
  set work.tmp (keep=pct text rename=(pct=start text=label)) end=__last;
  retain fmtname "st" type "n";

  end=start;

  output;

  if __last = 1 then
    do;
      hlo = "o";
      label = "other";
      output;
    end;
run;

PROC FORMAT library=work cntlin=work.sevfmt;
run;
```

## SGPANEL

The SGPANEL procedure creates a panels of graph cell for the values of one or more classification variables, in our case the two kinds of reactogenicity events. Using a simple PROC FORMAT the various events and their severity frequency can be easily plotted. This eliminated a lot of the annotate code and reduces the length of the program to nearly one-third of its size.

```
PROC SGPANEL data=;
  panelby reacto /novarname uniscale=row;
  vbar symptom/stat=sum group=severity response=pct datalabel;
  rowaxis tickvalueformat=best12. values=(0 to 100 by 20) label='PERCENT OF SUBJECTS';
  label severity='Severity' symptom='Symptom';
  format symptom $symp. severity $sev. pct st.;
run;
```

In the code above:

1. reacto is the type of symptom (local or systemic)
2. severity refers to the severity of the symptom itself (mild, moderate or severe)

- pct refers to the percentage of the subjects who have a severe symptom which are displayed on top of each bar as a ratio of the numbers of subjects with a severe symptom to the total number of subjects having the symptom.

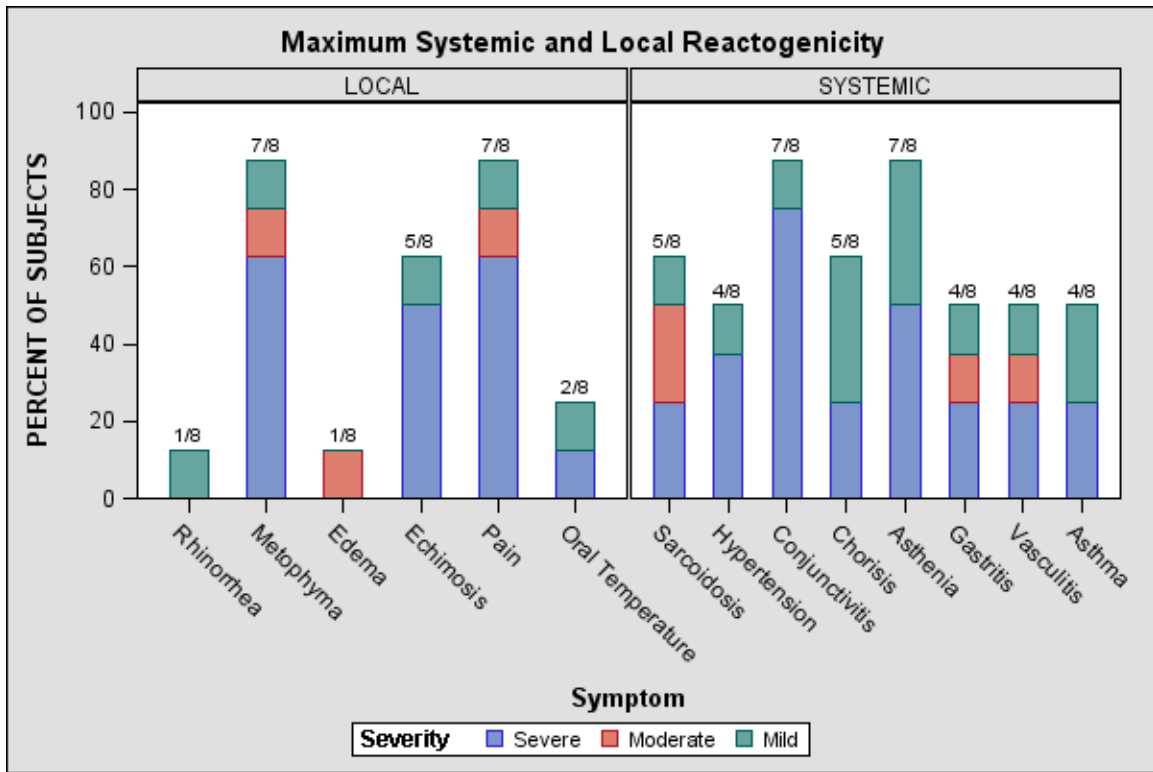


Figure 2: Using PROC SGPNEL

The TEMPLATE procedure enables you to customize the appearance of your SAS output graph. The following code uses the Styles.Default template that SAS provides, but includes changes made to the style in order to customize the output's appearance in our case: color. When using the GraphData# style element, you can use the COLOR= style attribute.

```

PROC TEMPLATE;
define style styles.mystyle;
parent=styles.default;
style graphdata1 from graphdata1 / contrastcolor=gray color=red;
style graphdata2 from graphdata2 / contrastcolor=gray color=orange;
style graphdata3 from graphdata3/contrastcolor=gray color=yellow;
end;
run;

```

The style is invoked using the ods statement “ods listing style=mystyle” before the proc sgpanel statement.

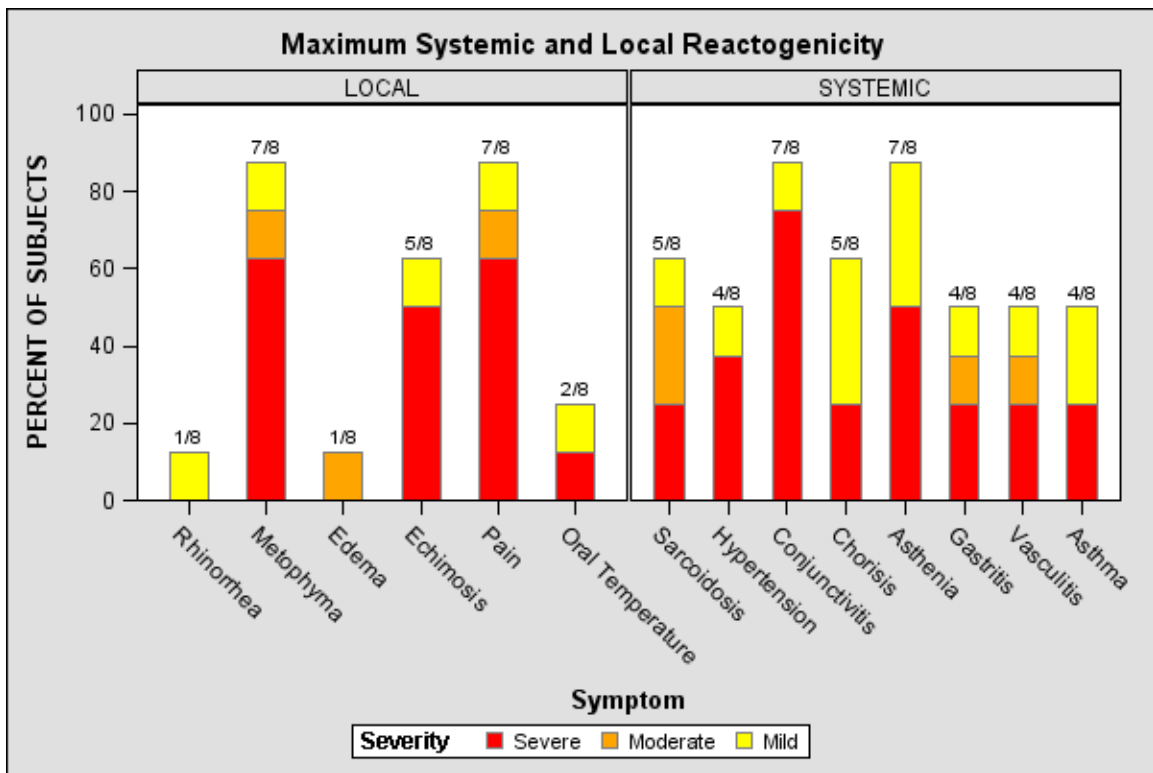


Figure 3: PROC SGPanel with Styles Template

## CONCLUSION

This has only been an introduction of the many and varied ways to use a formatting data set and SGPanel in place of traditional annotate methods. SAS online documentation provides more information on the use of GTL to accomplish the same task, the analysis of which is beyond the scope of this paper.

## REFERENCES

- Philip R Holland. Why Should You Be Using the New SG (Statistical Graphics) Procedures in SAS® 9.2?  
SAS Global Forum 2011, 427-2011
- Chuck Kincaid. SGPANEL: Telling the Story Better.  
SAS Global Forum 2010, 234-2010.
- Wendi L. Wright. Creating a Format from Raw Data or a SAS Dataset.  
SAS Global Forum 2007, 068-2007.

## CONTACT INFORMATION

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

Name: Mekhala Acharya

Enterprise: EMMES Corp.

Address: 301 N. Washington Street

City, State ZIP: Rockville, MD 20850

Phone: (301)251-1161

E-mail: macharya@emmes.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.