Three Ways to Add Texts to Graphics in PROC SGPLOT

Yaqi Jia, Johns Hopkins University, Baltimore, MD

ABSTRACT

The procedure PROC SGPLOT in Base SAS® is a powerful procedure to create various graphics. We can customize graphics adding features such as color, pattern, and size, within each plot statement. However, sometimes graphics are not enough. We may wish to add texts to graphics. For example, adding correlation coefficient 'R = 'in a scatter plot helps demonstrate how the y axis variable is correlated with the x axis variable. Or, we want to add true odds ratio (95% CL) results beside their forest plots. Base SAS® version 9.4 provides multiple ways to accomplish this goal. This paper will discuss three powerful and easy-to-learn ways to add texts to the graphics using special statements and options in the procedure PROC SGPLOT: INSET statement, TEXT statement, and the optional SGANNO=anno dataset in PROC SGPLOT statement.

INTRODUCTION

ODS Graphics introduces a whole new way to generate high-quality graphs using Base SAS® version 9.2 and later version. There are more than 60 statistical procedures to produce graphs using ODS graphics. They all start with the letters 'SG'. Among them, the procedure PROC SGPLOT is a common used and powerful way to produce a wide range of plot types including box, dot, bar, needle and others (SAS Institute Inc., 2018, p. 744). The SGPLOT procedure can also allow us to specify colors, marker symbols, and other attributes of the plot features. The INSET and TEXT statements are two ways to add simple text to the graph. A third powerful way to add text to graphs is to use SGANNO=ANNOTATE DATASET option in the PROC SGPLOT statement. The 'annotate dataset' is a SAS dataset that contains the commands for creating the annotation elements. This dataset can be created using the same methods that are used to create any SAS data set. The main distinctions are that the SG annotation data set uses reserved keywords for its variable names, and each observation represents a command to draw an annotation element (SAS Institute Inc., 2018, p. 1611). When SGANNO option is used in the PROC SGPLOT procedure, the annotate dataset will be linked and text will be added to the graph. This paper was written using SAS 9.4 M14.3, and will be a helpful resource for all industries and all levels of proficiency with SAS.

REVIEW OF PROC SGPLOT

The SGPLOT procedure creates one or more plots and overlays them on a single set of axes. We can use the PROC SGPLOT to create statistical graphics such as scatter plots, box plots, and regression plots. By customizing statements or adding options, we can control the appearance of our graph such as line patterns, colors and thickness and add additional features such as legends. Prior to demonstrating how adding text improves the readability of output from the SGPLOT procedure, we will illustrate how customization can enhance the appearance of statistical graphs in general. We will create a scatter and ellipse plot that show how the markerattrs and lineattrs statements can improve the understanding of graphs:

```
ODS GRAPHICS/RESET;
PROC SGPLOT DATA=sashelp.class;
   SCATTER Y=height X=weight / MARKERATTRS = (SIZE=10 COLOR=black);
   ELLIPSE Y=height X=weight / LINEATTRS = (COLOR=red THICKNESS=4);

RUN;
```

The results are shown in Figure 1

Figure 1. Scatter and ellipse plot using PROC SGPLOT
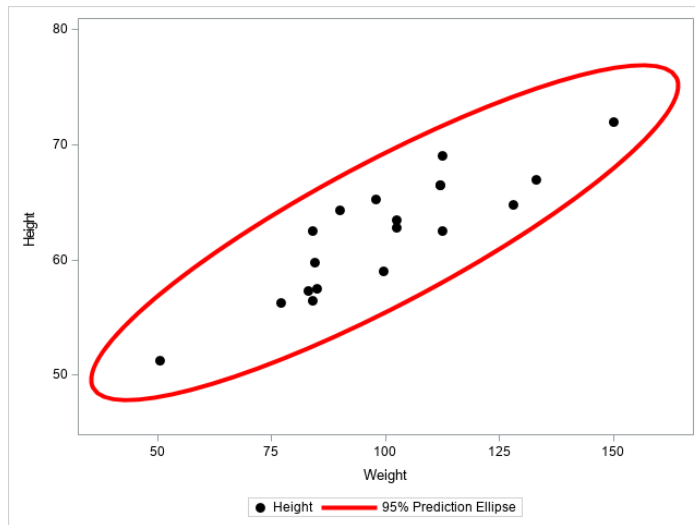
THREE WAYS TO ADD TEXT TO GRAPHICS

SAS provides various ways to add text to graphics by using different options, statements and procedures. Not every method is easy to learn and use. Here are three easy and powerful ways to enhance statistical graphics. We usually wish to add the correlation coefficient R, e.g. 'R = 0.8', to a scatter plot to show how the y-axis variable and x-axis variable are correlated. Or, we want to add labels besides some specific lines or markers. Thirdly, we may want to have a more flexible tool to add text to our graphics. We will show three examples below that accomplish these goals, using the INSET statement, the TEXT statement, and the SGANNO option in PROC SGPLOT.

EXAMPLE 1: INSET STATEMENT

The first simple way to add descriptive text to graphics is INSET statement. INSET statement can be added to any type of graph (SAS Institute Inc., 2018, p. 987).

Syntax:

INSET "text-string" <... "text-string-n"> <*/option(s)*>;

*"text-string": specifies one or more quoted text strings. Each string is placed on a separate line in the text box (for example, "My line 1" "My line 2").

Or:

INSET (label-list) <*/option(s)*>;

*(label-list): Specify your label-value pairs as follows: ("label-1" ="value-1" ... "label-n" = "value-n")

Table 1: Options of INSET statement

Options	Summary
BACKCOLOR=*color*	Specifies the background color of the inset.
BORDER | NOBORDER	Specifies whether to display a border around the text box.
LABELALIGN=LEFT | CENTER | RIGHT	Specifies how the labels are aligned when you specify label-value pairs.
OPAQUE	Forces the inset background to be opaque rather than transparent.
POSITION=*position-value*	Specifies the position of the text box within the plot. Eight positions: BOTTOM BOTTOMLEFT BOTTOMRIGHT TOP TOPLEFT TOPRIGHT LEFT RIGHT
TEXTATTRS=*style-element <(options)> | (options)*	Specifies the appearance of the text in the text box.
TITLE="*text-string*"	Specifies a title for the text box.
TITLEATTRS=*style-element <(options)> | (options)*	Specifies the appearance of the title.
VALUEALIGN=LEFT | CENTER | RIGHT	Specifies how text values are aligned.

An example:

```
PROC SGPLOT DATA=sashelp.class;
   SCATTER Y=height X=weight / MARKERATTRS = (SIZE=10 COLOR=black SYMBOL=circlefilled);
   INSET   ("(*ESC*){UNICODE alpha}" = "0.05"
           "R(*ESC*){sup '2'}" = "0.78"     ) / BORDER TEXTATTRS = (SIZE=12 COLOR=red);
RUN;
```

Figure 2: Scatter plot with text using INSET statement
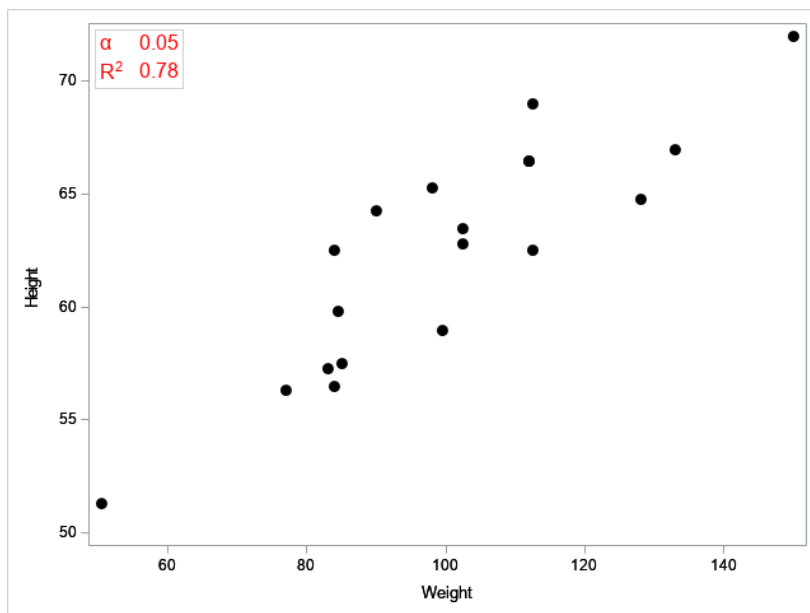
This example is using INSET (label-list) to add text to scatter plots. The SCATTER statement draws the scatter plot with y-axis=height and x-axis=weight and also controls the dots features using "markerattrs" option. The INSET statement adds "alpha 0.05" and "R^2 0.78" to the top left of the Figure 2 (SAS itself will choose one of 8 positions to insert the text if you don't specify its position.) If you want to choose a specific position, you can use POSITION option to place text within the plot). The BORDER option will add the border around the text box. The TEXTATTRS option can specify the appearance of the text, such as color, size or style. In this INSET statement, we use the second syntax: INSET (label list), which should have this syntax: INSET ("label-1" = "value-1" "label-2" = "value-2). Here (*ESC*) {Unicode alpha} is the label-1 and 0.05 is the value-1. (*ESC*) is the default ODS escape character.

EXAMPLE 2: TEXT STATEMENT

The TEXT statement displays the associated text values at (X, Y) locations in the graph. The text can be numbers or characters (SAS Institute Inc., 2018, p. 1159).

Syntax:

TEXT X=variable Y=variable TEXT=variable </option(s)>;

Required Arguments:

X=variable: specifies the variable for the X axis.

Y=variable: specifies the variable for the Y axis.

TEXT=variable: specifies the variable for the text values that are used for the markers.
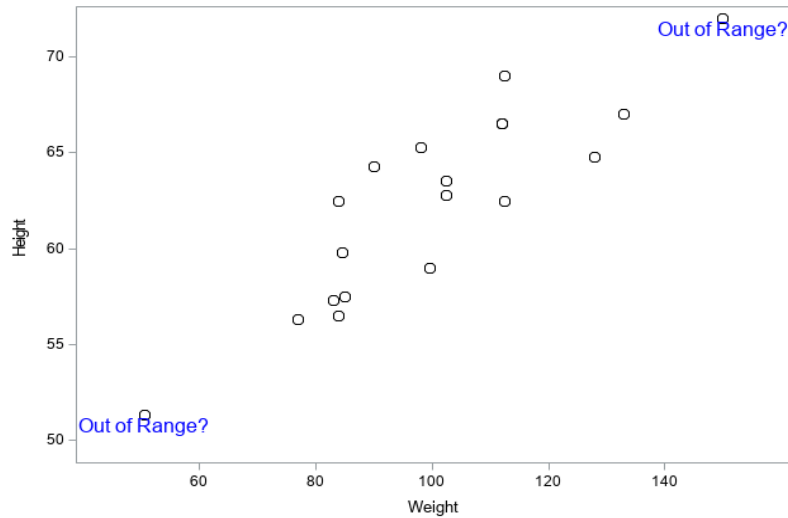
An example:

```
DATA class;
  SET sashelp.class;
  IF height=72   and weight=150  THEN text='Out of Range?';
  IF height=51.3 and weight=50.5 THEN text='Out of Range?';
RUN;

PROC SGPLOT DATA=class;
  SCATTER Y=height X=weight / MARKERATTRS = (SIZE=10 COLOR='black');
  TEXT    Y=height X=weight TEXT=text
          /POSITION=bottom TEXTATTRS = (SIZE=12 COLOR='blue');
RUN;
```

The data step creates a new variable 'text' and assigns a value 'Out of Range' to two subjects, one with height=72 & weight=150, the other with height= 51.3 and weight=50.5.

The TEXT statement in PROC SGPLOT adds the text 'Out of Range' to the scatter plot. Y=height and X=weight provides the text location and TEXT=text provides the text contents. Since only two subjects were assigned the text 'Out of Range' in the DATA step, there are only two circles with text 'Out of Range' next to them in the Figure 3. The option 'POSITION' specifies the position of the text with respect to the location of the data point. The option 'TEXTATTRS' specifies the color and font properties of the marker text.

Figure 3: Using a TEXT statement to add text to a scatter plot

EXAMPLE 3: PROC SGPLOT STATEMENT WITH THE SGANNO OPTION

The option SGANNO in the PROC SGPLOT statement specifies the SG annotation data set that you want to use. SG annotation data set is a SAS data set that contains the commands for creating one or more annotation elements to a graph. In other words, the data set contains all the information about what and how you want to add to a graph using this data set (SAS Institute Inc., 2018, p. 1611). Table 2 shows what can be added to a graph. We are focusing on the TEXT function. Table 3 is an example of how the TEXT function of SG annotation data set looks like. The variable 'FUNCTION' in the data set tells SAS what function you want to add to the graph. Here we want to add text, so all the values are 'text'. All other variables in the data set tell SAS what features you want to assign to the text. The variable 'LABEL' is the content of the text. X1 and XC1 specify the text position on X-axis (X1 is for numeric data, XC1 is for character data). Similarly, Y1 and YC1 specify the text position on Y-axis. XSPACE and YSPACE specifies the drawing space of the annotation's X coordinate and Y coordinate, respectively. They are like the unit of x1 or y1. The value for Xspace and Yspace can be one of the followings (see Table 4: Draw space values and Description and Figure 4: Explanation of draw space of a SAS graph). The meaning of other features are similar to those in common graphs: for example, textsize is the size of text, etc. A forest plot is used as an example to explore how SGANNO works.

Table 2: Summary of SG Annotation Functions

Function	Description
ARROW	Draws an arrow annotation.
IMAGE	Specifies a graphic file to use for an image annotation.
LINE	Draws a line annotation.
OVAL	Draws an oval or circle annotation.
POLYCONT	Continues drawing a polygon that was begun with the POLYGON function, or a line that was begun with the POLYLINE function.
POLYGON	Specifies the beginning point of a polygon.

|	
POLYLINE	Specifies the beginning point of a polyline, which is a connected series of line segments.
RECTANGLE	Draws a rectangle or square annotation.
TEXT	Places text in the graph output.
TEXTCONT	Continues a text string.

Table 3: The dataset (Anno_AKI) as an example for TEXT function of SG annotation

function	label	x1	x1space	yc1	y1space	anchor	textweight	width	textsize	fillcolor	display
text	Anti-angiogenic biomarker	1.5	graphpercent	Anti-angiogenic biomarker	datavalue	left	bold	200	10	white	all
text	VEGFR1	3	graphpercent	VEGFR1	datavalue	left		200	8	white	all
text	1.56 (1.31, 1.87)	-2	wallpercent	VEGFR1	datavalue	right		200	8	white	all
text	Pro-angiogenic biomarkers	1.5	graphpercent	Pro-angiogenic biomarkers	datavalue	left	bold	200	10	white	all
text	VEGF	3	graphpercent	VEGF	datavalue	left		200	8	white	all
text	0.89 (0.82, 0.98)	-2	wallpercent	VEGF	datavalue	right		200	8	white	all
text	PGF	3	graphpercent	PGF	datavalue	left		200	8	white	all
text	0.69 (0.55, 0.87)	-2	wallpercent	PGF	datavalue	right		200	8	white	all

Table 4: Draw space values and Description in a SG annotation dataset

Draw Space Values	Description
DATAPERCENT	Positioned and scaled as a percentage with respect to the data area.
DATAPIXEL	Positioned and scaled as pixels with respect to the data area.
DATAVALUE	Positioned and scaled with respect to the data values.
GRAPHPERCENT	Positioned and scaled as a percentage of the graph area.
GRAPHPIXEL	Positioned and scaled as pixels with respect to the graph area.
LAYOUTPERCENT	Positioned and scaled as a percentage of the layout area.
LAYOUTPIXEL	Positioned and scaled as pixels with respect to the layout area.
WALLPERCENT	Positioned and scaled as a percentage of the wall area.
WALLPIXEL	Positioned and scaled as pixels with respect to the wall area.

Figure 4: Explanation of draw space of a SAS graph
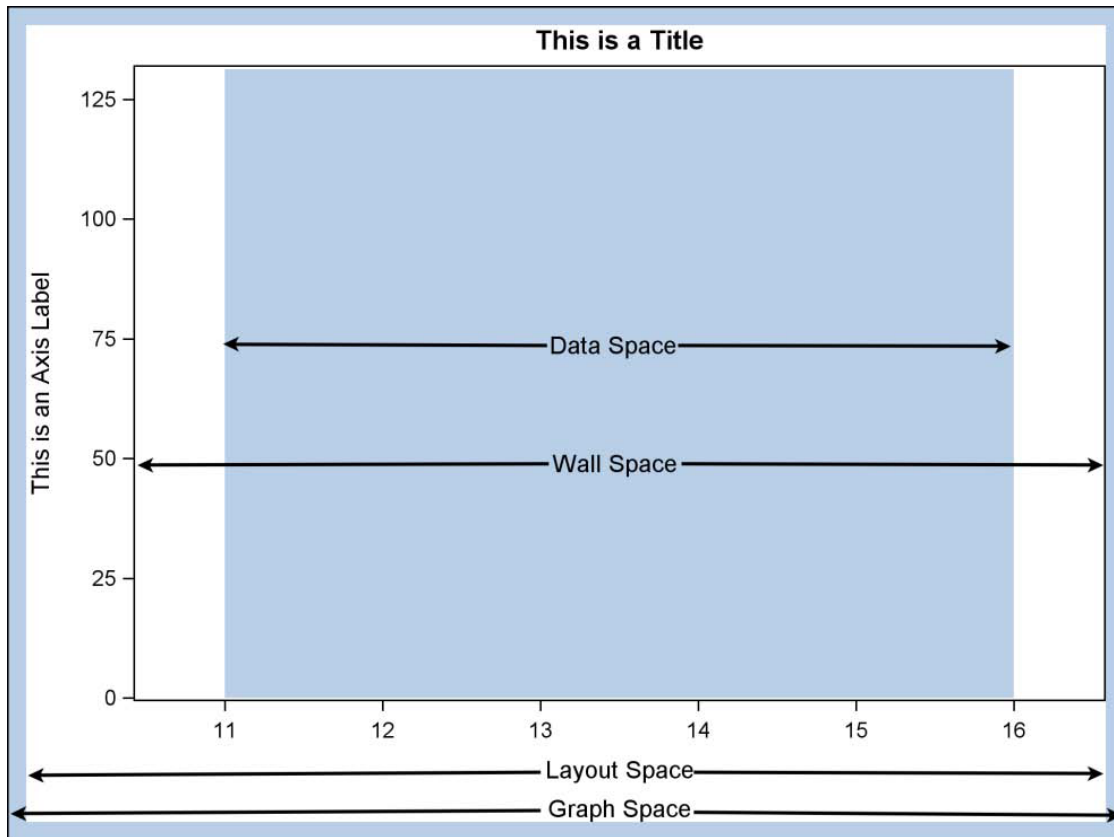
The dataset AKI contains all the information that will be used to draw the forest plot (See Table 5). The SGPLOT procedure code shown below draws a forest plot by using SCATTER statement and HIGHLOW statement. The right part of

Figure **5** is generated by these two statements. The REFLINE statement draws the dash line x=1 for the xaxis. The YAXIS statement customizes the yaxis features. The 'REVERSE' option makes the yaxis tick values display in reverse order and 'DISPLAY=NONE' specifies that none of the features of the yaxis are displayed.

The 'SGANNO=anno_aki' option in PROC SGPLOT statement links it to the dataset 'anno_aki', which contains all the information that how the text can be added to the left of graph (See Table 3 for the contents of anno_aki). The first row of the anno_aki tells SAS to add the text 'Anti-angiogenic biomarker' at the position of xaxis = 1.5 graphpercent and yaxis=Anti-angiogenic biomarker with TEXTSIZE=10 and TEXTWEIGHT=bold. The third row of the anno_aki tells SAS to add '1.56 (1.31, 1.87)' to the graph at the position of xaxis= -2 wallpercent and yaxis = 1.56 (1.31, 1.87). '-2 wallpercent' means at the 2% percent left far away from the wall space left margin. Since the 'anchor' for this row is 'right', the furthest right of the '1.56 (1.31, 1.87)' text is aligned with -2wallpercent, which is a little further left from 0.5 of xaxis in Figure 5. In general, each row in the dataset anno_aki adds one text to the graph.

The SAS codes for Figure 5:

```
PROC SGPLOT DATA=AKI NOCYCLEATTRS NOAUTOLEGEND NOWALL NOBORDER
            SGANNO=anno_aki DATTRMAP=attrmap_aki PAD=(LEFT=70% BOTTOM=3% TOP=10%);
   REFLINE  1/ AXIS=x LINEATTRS=(PATTERN=2);
   SCATTER  Y=varlabel X=AKI_OR / GROUP=varlabel ATTRID=my_id;
   HIGHLOW  Y=varlabel HIGH=AKI_UCI LOW=AKI_LCI / HIGHCAP=serif LOWCAP=serif;
   YAXIS    REVERSE DISPLAY=none;
   XAXIS    TYPE=log LOGBASE=2 VALUES=(0.5,0.75,1,1.5,2) LABEL='Adjusted Odds Ratio';
RUN;
*The dataset 'anno_aki' is shown in Table 3 (Table 3: The dataset (Anno_AKI) as an example for TEXT
function of SG annotation);
```
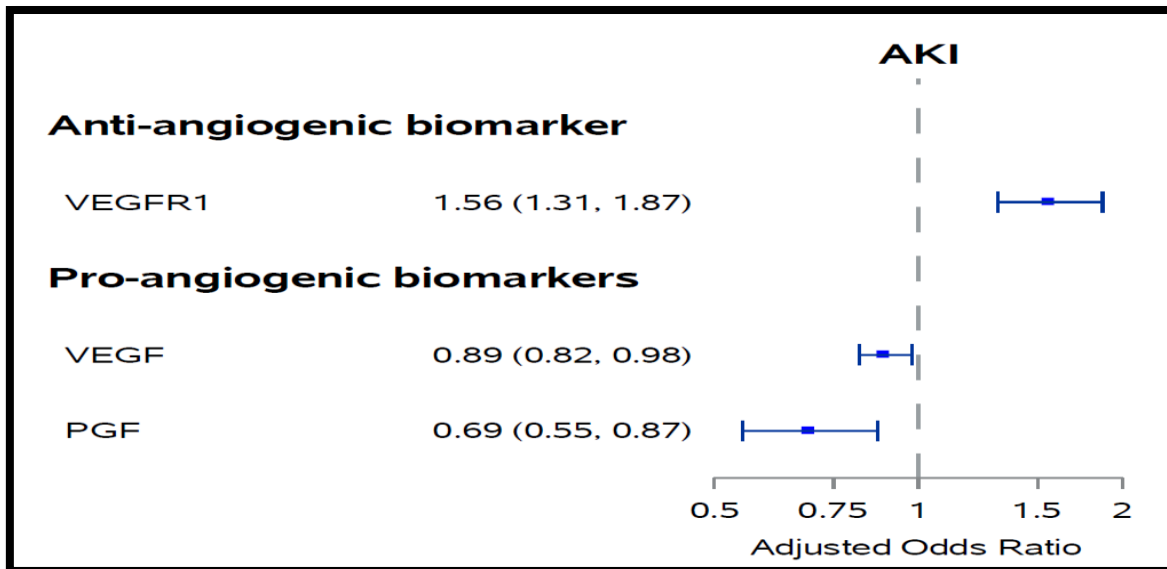
Table 5: The dataset 'AKI' used for forest plot

varname	varlabel	AKI_or	AKI_lci	AKI_uci	AKI_orci
Header	Anti-angiogenic biomarker				
VEGFr1	VEGFR1	1.56	1.31	1.87	1.56 (1.31, 1.87)
Header	Pro-angiogenic biomarkers				
VEGF	VEGF	0.89	0.82	0.98	0.89 (0.82, 0.98)
PGF	PGF	0.69	0.55	0.87	0.69 (0.55, 0.87)

Figure 5: Forest plots with TEXT added

CONCLUSION

These three ways of enhancing SGPLOT output are very useful and straightforward to pick up. The choice of which way is best for depends on your needs. If you only need to add simple text, such as the coefficient of a parameter and/or p-value somewhere inside the figure, try the INSET statement. If you want to add data values-related text inside the figure, try TEXT statement. If you want to add more flexible text inside or outside the figure, use the option SGANNO in the PROC SGPLOT statement. This option will link it to a

dataset which contains all the information on what content you want to add to your statistical graphic and how that content should be displayed.

REFERENCES

Slaughter, S. J., & Delwiche, L. D. (2010). Using PROC SGPLOT for Quick High-Quality Graphs. *SAS Global Forum 2010*, 156.

SAS Institute Inc. (2018). INSET Statement. In *SAS@ 9.4 ODS Graphics: Procedures Guide, Sixth Edition* (p. 987). Cary, NC: SAS Institute Inc.

SAS Institute Inc. (2018). TEXT statement. In *SAS@ 9.4 ODS Graphics: Procedures Guide, Sixth Edition* (p. 1159). Cary, NC: SAS Institute Inc.

SAS Institute Inc. (2018). Annotating ODS Graphics. In *SAS@ 9.4 ODS Graphics: Procedures Guide, Sixth Edition* (p. 1611). Cary, NC: SAS Institute Inc.

Recommended Reading

- *Base SAS® Procedures Guide*
- SAS® 9.4 ODS Graphics: Procedures Guide, Sixth Edition

ACKNOWLEDGMENTS

CONTACT INFORMATION

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

Yaqi Jia
Johns Hopkins University
+1 443-287-4712
yaqi.jia@jhu.edu