

Creating a Publication Quality Graphic with SAS®

Charlotte Baker, Florida Agricultural and Mechanical University

ABSTRACT

Graphics are an excellent way to display results from multiple statistical analyses and get a visual message across to the correct audience. The combination of SG procedures, such as PROC SGPLOT, and ODS statements in SAS allow for the creation of custom graphics that meet the expectations of scientific journals and are excellent for research presentations or handouts. While these are excellent tools, first time users may experience difficulties when attempting to utilize them. This paper will describe two methods for creating a publication quality graphic in SAS 9.4 and, more specifically, solutions for some issues encountered when doing so.

INTRODUCTION

The graphical presentation of data is both an art and a science. Often, researchers are required to present results of studies in a manner that is clear not only to other researchers but also to the general public. Graphics such as line graphs and bar charts can be very useful as tools to accomplish this. One caveat of graphical display is that the graph or chart needs to be readable if expanded for display in a presentation or minimized if used in a print journal. It also needs to look good on websites and in PDF documents. While there are many programs that can be used to create graphics, SAS contains tools that can be combined to do a more than sufficient job, namely the statistical graphics (SG) procedures and the Output Delivery System (ODS).

There are many reasons why one would choose to use a single program to complete both statistical analysis and the graphing of that data. However, it is arguably more important that SAS's SG procedures can create many different types of graphs based on many different statistics and allows the complete customization of the output. This means that no matter where the graphic needs to be used, it will reflect exactly what the researcher wanted someone to see. What program is used to make the graphic matters quite a bit when journals typically provide resolution, color, and/or size criteria for graphics submitted along with manuscripts. The evolution of the SG procedures along with ODS means that SAS can create graphics that can be used for such purposes. This paper will demonstrate two methods of creating a publication quality graphic using PROC SGPLOT.

SCENARIO

You are a researcher in a team that has completed a study of stroke hospitalizations due to sickle cell disease over a period of six data collection periods. We are interested in graphing information to show whether the proportion of stroke hospitalizations for children with sickle cell disease has changed over the data collection periods for four age groups. Our four age groups are 1-4 years, 5-9 years, 10-14 years, and 15-18 years. To find out whether there was a trend in the proportion for each age group, we used a linear regression. We want to create a single line graph (see Figure 1) with a regression trend line for each age group and a marker for the proportion of hospitalizations for each age group by year.

METHOD 1

You need a graphic that can be used quickly for a slide presentation. The screen for the presentation is large but the factor of most concern is the file size so that the presentation loads quickly. You choose to create a TIFF image.

The code to create a basic graph for this scenario requires information for the x-axis and the y-axis. Because we want to have a linear regression line to show the trend, we need to use a REG statement. The REG statement fits a regression line to your data.

```
ODS GRAPHICS ON;
PROC SGPLOT DATA = trenddata;
  REG X=year Y=ageper1;
  REG X=year Y=ageper2;
  REG X=year Y=ageper3;
  REG X=year Y=ageper4;
  XAXIS DISCRETEORDER=DATA VALUES=(1997 2000 2003 2006 2009 2012);
  YAXIS VALUES =(00 05 10 15 20 25 30 35 40 45 50 55 60);
RUN;
```

The SG procedures use ODS GRAPHICS to provide the best image possible so these need to be turned on. Because we are trying to plot four separate lines (one for each group), this code shows four separate REG statements. In each REG statement, "year" is the name of the variable that contains the information that should be on

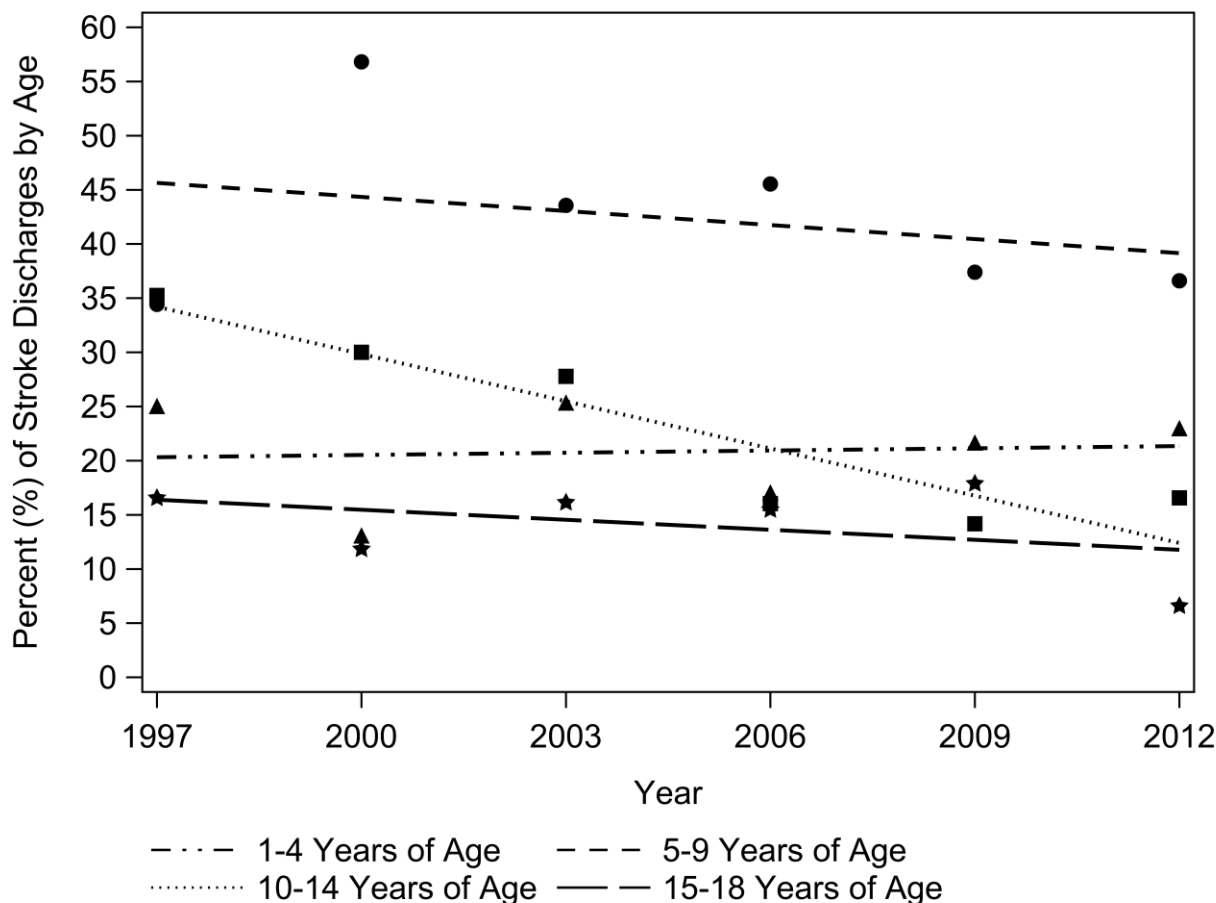


Figure 1. Presentation Quality Graphic – TIFF at 1200 DPI (Baker et al)

the x-axis. In this case, we are plotting the information on the y-axis by year and the variable for year is “year”. The first line we want to plot is for the first age group (1-4 years of age). We are plotting the proportion of stroke hospitalizations due to sickle cell disease for that age group on the y-axis and the variable that represents this proportion is “ageper1”. The XAXIS statement allows us to specify the options for what the x-axis looks like. The discreteorder option specifies the order of the tick marks on the x-axis. In this case we want them in the order of the data. The values we want to appear (instead of the default values) are the actual years of data collection that we studied – 1997, 2000, 2003, 2006, 2009, and 2012. The YAXIS statement allows us to specify the options for how the y-axis looks. Because we wanted to start our axis at 0 and because we were familiar with the age proportions that we were plotting on the y-axis, we made the values for the y-axis range from 0 to 60 in increments of 5. Each of these values will have a tick-mark on the y-axis.

While the code above will provide a graph that can certainly be used for presentation, it is often a goal to be sure that whoever is viewing the graph can decipher the different items being plotted. Because of this, we want to be sure to be more specific about how each of the linear regression lines look and how the markers for each proportion for each age look. We do this with additional options added to the REG statements.

```
PROC SGPLOT DATA = trenddata;
  REG X=year Y=ageper1 / LINEATTRS = (COLOR = BLACK PATTERN = DASHDOTDOT)
  MARKERATTRS = (SYMBOL = TRIANGLEFILLED COLOR = BLACK) LEGENDLABEL = "1-4 Years of
  Age";
  REG X=year Y=ageper2 / LINEATTRS = (COLOR = BLACK PATTERN = DASH) MARKERATTRS =
  (SYMBOL = CIRCLEFILLED COLOR = BLACK) LEGENDLABEL = "5-9 Years of Age";
  REG X=year Y=ageper3 / LINEATTRS = (COLOR = BLACK PATTERN = DOT) MARKERATTRS =
  (SYMBOL = SQUAREFILLED COLOR = BLACK) LEGENDLABEL = "10-14 Years of Age";
  REG X=year Y=ageper4 / LINEATTRS = (COLOR = BLACK PATTERN = LONGDASH) MARKERATTRS
  = (SYMBOL = STARFILLED COLOR = BLACK) LEGENDLABEL = "15-18 Years of Age";
  XAXIS DISCRETEORDER=DATA VALUES=(1997 2000 2003 2006 2009 2012);
  YAXIS VALUES=(00 05 10 15 20 25 30 35 40 45 50 55 60);
  RUN;
```

To add options to the REG statement (or any specific graph statement in PROC SGPLOT), insert a forward slash after identifying the X and Y variables and follow it with the options you want. In this study, we wanted to change the color and pattern of each line, the shape and color of each marker that identified the proportions, and add a specific label for the graph's legend. LINEATTRS is the option that changes the attributes of the line you see in the graph. Because we are changing more than one attribute (color and pattern), we need to put the attributes we are changing inside parentheses. To specify colors in SAS/GRAPH®, one has several options including RGB, CMYK, and SAS Registry Color options. In this example, we used a recognized SAS Registry Color – black. Just as there are many options to change the color of a line, there are many options to change the pattern of the lines that appear for our REG line. We selected to use a longdash pattern for the fourth regression line.

The markers in this example are the items that appear on the graph to indicate where on the graph the proportion for a specific age group in a specific year. We change what the markers look like by using the MARKERATTRS option. In this example, we wanted to change the shape of the marker (done with the SYMBOL option) and the color of that marker. Just like the color for the LINEATTRS option, we change the color for the MARKERATTRS option using a SAS registry color. Because the fourth REG statement is for the 15-18 year old age group, each marker on the graph for that group will be a filled in star shape (starfilled). There are approximately two dozen marker symbols to choose from. See the reference section for links to further examples and details for specifying colors, patterns, and markers.

If you were to print this graph using the code above, you would have a graph that has everything you want in the graph but it is missing labels for the axes and the fonts for the labels and axes will be smaller than is typically warranted when making graphics for professional purposes. To correct this, we add more options to the XAXIS and YAXIS statements. The graph also does not have a legend to explain to us what the lines and markers mean. To add one, we add the KEYLEGEND statement.

```
PROC SGPLOT DATA = trenddata;
  REG X=year Y=ageper1 / LINEATTRS = (COLOR = BLACK PATTERN = DASHDOTDOT)
  MARKERATTRS = (SYMBOL = TRIANGLEFILLED COLOR = BLACK) LEGENDLABEL = "1-4 Years of
  Age";
  REG X=year Y=ageper2 / LINEATTRS = (COLOR = BLACK PATTERN = DASH) MARKERATTRS =
  (SYMBOL = CIRCLEFILLED COLOR = BLACK) LEGENDLABEL = "5-9 Years of Age";
  REG X=year Y=ageper3 / LINEATTRS = (COLOR = BLACK PATTERN = DOT) MARKERATTRS =
  (SYMBOL = SQUAREFILLED COLOR = BLACK) LEGENDLABEL = "10-14 Years of Age";
  REG X=year Y=ageper4 / LINEATTRS = (COLOR = BLACK PATTERN = LONGDASH) MARKERATTRS
  = (SYMBOL = STARFILLED COLOR = BLACK) LEGENDLABEL = "15-18 Years of Age";
  XAXIS VALUEATTRS = (FAMILY = ARIAL SIZE = 18pt) DISCRETEORDER=DATA VALUES=(1997
  2000 2003 2006 2009 2012) LABEL = "Year" LABELATTRS = (SIZE =18pt);
  YAXIS VALUEATTRS = (FAMILY = ARIAL SIZE = 18pt) LABEL = "Percent (%) of Stroke
  Discharges by Age" LABELATTRS = (SIZE =18pt) VALUES =(00 05 10 15 20 25 30 35 40 45
  50 55 60);
  KEYLEGEND / NOBORDER LOCATION = OUTSIDE POSITION = BOTTOMLEFT VALUEATTRS = (FAMILY
  = arial SIZE = 18pt);
RUN;
```

The VALUEATTRS option allows us to change the attributes of the tick mark value labels for each axis. For the YAXIS, we specified that any text for the y-axis should be in 18-point Arial font. In this example, we chose 18-point font because that is the smallest recommended font size for anything presented in a slideshow. SAS accepts font sizes to be specified in point size, pixels, inches, centimeters, millimeters, or percentages. In order for the font selected in the FAMILY option to actually appear in the graph, it must be a TrueType font and must be present on the computer being used to make the graph. The LABEL option is where you can specify what text should appear on the axis. For the x-axis, we decided to simply write "Year". By using the LABELATTRS option, we specified that "Year" should also appear in 18-point font. KEYLEGEND specifies that we need a legend for our graph. In this case, we did not want a border for the legend and wanted it outside the main graph area. We wanted it to appear instead to the bottom left outside the graph. We wanted the information in the key to appear in Arial 18-point font.

We are now ready to output this graph as a TIFF image as we decided in the beginning. To do this, the HTML output needs to be turned off and the LISTING output needs to be turned on. This is very important as the default in current SAS 9.4 installations is for HTML to be turned on. In addition, one must specify where the graph should be saved. The GPATH option for ODS LISTING allows us to do this.

```
ODS HTML OFF;
ODS LISTING GPATH = "C:\tiff";
```

Without further specifications, an image will be saved but it will be in PNG format. To change the image format, we need to add the IMAGEFMT option to the ODS GRAPHICS statement.

```

ODS HTML OFF;
ODS LISTING GPATH = "C:\tiff";
ODS GRAPHICS ON / IMAGEFMT = TIFF;
PROC SGPLOT DATA = trenddata;
REG X=year Y=ageper1 / LINEATTRS = (COLOR = BLACK PATTERN = DASHDOTDOT)
MARKERATTRS = (SYMBOL = TRIANGLEFILLED COLOR = BLACK) LEGENDLABEL = "1-4 Years of
Age";
REG X=year Y=ageper2 / LINEATTRS = (COLOR = BLACK PATTERN = DASH) MARKERATTRS =
(SYMBOL = CIRCLEFILLED COLOR = BLACK) LEGENDLABEL = "5-9 Years of Age";
REG X=year Y=ageper3 / LINEATTRS = (COLOR = BLACK PATTERN = DOT) MARKERATTRS =
(SYMBOL = SQUAREFILLED COLOR = BLACK) LEGENDLABEL = "10-14 Years of Age";
REG X=year Y=ageper4 / LINEATTRS = (COLOR = BLACK PATTERN = LONGDASH) MARKERATTRS
= (SYMBOL = STARFILLED COLOR = BLACK) LEGENDLABEL = "15-18 Years of Age";
XAXIS VALUEATTRS = (FAMILY = ARIAL SIZE = 18pt) DISCRETEORDER=DATA VALUES=(1997
2000 2003 2006 2009 2012) LABEL = "Year" LABELATTRS = (SIZE =18pt);
YAXIS VALUEATTRS = (FAMILY = ARIAL SIZE = 18pt) LABEL = "Percent (%) of Stroke
Discharges by Age" LABELATTRS = (SIZE =18pt) VALUES =(00 05 10 15 20 25 30 35 40 45
50 55 60);
KEYLEGEND / NOBORDER LOCATION = OUTSIDE POSITION = BOTTOMLEFT VALUEATTRS = (FAMILY
= arial SIZE = 18pt);
RUN;

```

We now have a graphic that we can use in our presentation!

METHOD 2

After your presentation, you decide to submit your study results for publication in a manuscript that requires the image to be a TIFF image that is 1200DPI. Your current method of creating a TIFF image will not create an image with that type of resolution (the default is 100 DPI in HTML). The alternatives are to use the IMAGE_DPI option for ODS or the DPI option for ODS. IMAGE_DPI will work in ODS LISTING or ODS HTML to create images but neither allows for the creation of a TIFF image. Because I could use a free program outside of SAS to convert the PNG image that DPI and IMAGE_DPI could provide, I elected to create a PNG image. In addition to creating the image in HTML, I elected to create it in a PDF and an RTF file as well.

```

ODS HTML DPI = 1200;
ODS PDF DPI = 1200 FILE = "C:\tiff\graph.pdf";
ODS RTF DPI = 1200 FILE = "C:\tiff\graph.doc";
PROC SGPLOT DATA = trenddata;
REG X=year Y=ageper1 / LINEATTRS = (COLOR = BLACK PATTERN = DASHDOTDOT)
MARKERATTRS = (SYMBOL = TRIANGLEFILLED COLOR = BLACK) LEGENDLABEL = "1-4 Years of
Age";
REG X=year Y=ageper2 / LINEATTRS = (COLOR = BLACK PATTERN = DASH) MARKERATTRS =
(SYMBOL = CIRCLEFILLED COLOR = BLACK) LEGENDLABEL = "5-9 Years of Age";
REG X=year Y=ageper3 / LINEATTRS = (COLOR = BLACK PATTERN = DOT) MARKERATTRS =
(SYMBOL = SQUAREFILLED COLOR = BLACK) LEGENDLABEL = "10-14 Years of Age";
REG X=year Y=ageper4 / LINEATTRS = (COLOR = BLACK PATTERN = LONGDASH) MARKERATTRS
= (SYMBOL = STARFILLED COLOR = BLACK) LEGENDLABEL = "15-18 Years of Age";
XAXIS VALUEATTRS = (FAMILY = ARIAL SIZE = 12pt) DISCRETEORDER=DATA VALUES=(1997
2000 2003 2006 2009 2012) LABEL = "Year" LABELATTRS = (SIZE =12pt);
YAXIS VALUEATTRS = (FAMILY = ARIAL SIZE = 12pt) LABEL = "Percent (%) of Stroke
Discharges by Age" LABELATTRS = (SIZE =12pt) VALUES =(00 05 10 15 20 25 30 35 40 45
50 55 60);
RUN;
ODS PDF CLOSE;
ODS RTF CLOSE;

```

To change the DPI, one can use the DPI option and add a value that it should be. In this example, DPI = 1200 was the way to change the DPI of the image to 1200 DPI. Be sure to close the PDF and RTF ODS locations.

A serious problem generally occurs when trying to select a large DPI. SAS will quickly stop processing and provide the following error in the log:

```

Java virtual machine exception. java.lang.OutOfMemoryError: Java heap space. NOTE:
The SAS System stopped processing this step because of errors.

```

This message means that the java virtual machine in SAS does not have enough memory available to use to create the graphic. You can give it more memory to use by making changes in the root folder of your computer where SAS is installed. Directions to solve this problem are available from the SAS Knowledge Base (see references). By following these directions, you can give SAS access to more memory to use from your computer and it will be able to create the graphic you want to see. Any missteps in implementing the directions from that page can result in your computer or SAS becoming unusable so be sure that you're comfortable performing the procedure before attempting.

Once your SAS installation has enough Java Heap Space, try the same syntax again and enjoy the result! (Figure 1)

CONCLUSION

There are several tools and methods to create graphics for professional use. The PROC SGPLOT and ODS statements are excellent tools for using SAS to create these graphics for any type of presentation from a slideshow to a PDF to a print journal. It is important to know what type of graphics each ODS option can create and display. It is also important to know how to change the details of each aspect of the graph to ensure that it can be read and viewed as intended. Despite the sometimes tedious nature of creating graphics in SAS for professional uses, it is a welcome alternative to other software as it can create very detailed and sharp looking images.

REFERENCES

1. Slaughter, Susan J and Delwiche, Lora D. 2010. "Using PROC SGPLOT for Quick High-Quality Graphs." *Proceedings of the SAS Global Forum 2010*. Available at <http://support.sas.com/resources/papers/proceedings10/154-2010.pdf>.
2. SAS Institute Inc. "Sample 48138: Remove the border around the axis area and make the background white in graphics produced by PROC SGPLOT". October 26, 2012. Available at <http://support.sas.com/kb/48/138.html>.
3. SAS Institute Inc. "Setting the Resolution of Your Graphic". Available at <http://support.sas.com/documentation/cdl/en/graphref/65389/HTML/default/viewer.htm#n19kyqdftrtqxqwc10x79xxtv18xh.htm>.
4. SAS Institute Inc. "Using Fonts with Universal Printers and SAS/GRAPH Devices". Available at <http://support.sas.com/documentation/cdl/en/lrcon/65287/HTML/default/viewer.htm#n04qo8tw2712qln1mzugjoorkj7j.htm#p0is74urmmze08n1ou7f0qof6xcf>.
5. SAS Institute Inc. "PROC SGPLOT Concepts". Available at <http://support.sas.com/documentation/cdl/en/grstatproc/62603/HTML/default/viewer.htm#a003199399.htm>.
6. SAS Institute Inc. "Graph Size and Resolution". Available at http://support.sas.com/documentation/cdl/en/statug/68162/HTML/default/viewer.htm#statug_odsgraph_sect046.htm.
7. SAS Institute Inc. "Usage Note 31184: "Java virtual machine exception" message issued with ODS Graphics and large data sets." *SAS Knowledge Base*. August 31, 2010. Available at <http://support.sas.com/kb/31/184.html>.
8. Baker, C., Grant, A. M., George, M. G., Grosse, S. D., & Adamkiewicz, T. V. (2015). Contribution of Sickle Cell Disease to the Pediatric Stroke Burden Among Hospital Discharges of African-Americans—United States, 1997–2012. *Pediatric Blood & Cancer*, 62(12), 2076–2081. doi:10.1002/pbc.25655

Download code for the graphic: <https://github.com/dr baker11/SESUG2016.git>

CONTACT INFORMATION

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

Charlotte Baker
Florida Agricultural and Mechanical University Institute of Public Health
1515 S. Martin Luther King Jr Blvd
Tallahassee, FL 32307
charlotte.m.h.baker@gmail.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.