Making Your SAS® Results More Meaningful with Color
Kirk Paul Lafler; @sasNerd

Abstract
Color can help make your SAS® results more meaningful. Instead of producing boring and ineffective results, users are able to enhance the appearance of their output, documents, reports, tables, charts, statistics, and spreadsheets to highlight and draw attention to important data elements, details, and issues, including using color in headings, subheadings, footers, minimum and maximum values, ranges, outliers, special conditions, and other elements. Color can be added to text, foreground, background, rows, columns, cells, summaries, totals, and traffic lighting scenarios. Topics include using color to results, documents, reports, tables, charts and spreadsheets can be enhanced with color, effectively add color to PDF, RTF, HTML, and Excel spreadsheet results using PROC PRINT, PROC FREQ, PROC REPORT, PROC TABULATE, and PROC SGPLOT and Output Delivery System (ODS) with style.

Introduction
The English language-idiom, “A picture is worth a thousand words” has received a lot of traction in the digital age. Over the years numerous studies have been conducted to determine how color can be used to help humans assimilate information from pictures, images and visuals along with the best use of color in reports, tables and spreadsheets. Shugars (2018) describes that 4.5% of the world’s population possesses some level of color vision deficiency with the largest segment of this population being affected by red and green colors. Consequently, many experts suggest that the best way to gain the largest audience participation is to replace red and green colors with yellow and darker blue colors. In another very different study conducted by social media scientist, Dan Zarrella (2013), found that tweets with images are 94% more likely to be retweeted than tweets without images.

It’s commonly understood that data patterns and differences are not always obvious from tables and reports. This is where graphical output and its ability to display data in a meaningful way have a clear advantage over tables, reports and other traditional communication mediums. Through the use of effective visual techniques the ability to better understand data is often achieved – a need that appeals to data analysts, statisticians and others. With the availability of ODS Statistical Graphics the SAS user community has a powerful set of tools to produce high-quality graphical output for data exploration, data analysis, and statistical analysis.

This paper introduces SAS users to the world of ODS Statistical Graphics, its features, capabilities, and syntax associated with the SGPLOT, SGPanel and SGSCATTER procedures. Numerous tips and guidelines are presented to effectively use visuals and color to assimilate information contained in reports, tables and spreadsheets.

Data Set Used in Examples
Each example illustrated in this paper uses a Movies data set (table) consisting of six columns: title, length, category, year, studio, and rating. Title, category, studio, and rating are defined as character columns with length and year being defined as numeric columns, shown below.

<table>
<thead>
<tr>
<th>Title</th>
<th>Length</th>
<th>Category</th>
<th>Year</th>
<th>Studio</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brave Heart</td>
<td>117</td>
<td>Action</td>
<td>1995</td>
<td>Paramount Pictures</td>
<td>R</td>
</tr>
<tr>
<td>Casablanca</td>
<td>103</td>
<td>Drama</td>
<td>1942</td>
<td>MGM / UA</td>
<td>PG</td>
</tr>
<tr>
<td>Christmas Vacation</td>
<td>97</td>
<td>Comedy</td>
<td>1939</td>
<td>Warner Brothers</td>
<td>PG-13</td>
</tr>
<tr>
<td>Coming to America</td>
<td>116</td>
<td>Comedy</td>
<td>1950</td>
<td>Paramount Pictures</td>
<td>R</td>
</tr>
<tr>
<td>Dances with Devils</td>
<td>130</td>
<td>Romance</td>
<td>1936</td>
<td>Columbia Tijuana</td>
<td>R</td>
</tr>
<tr>
<td>Dexter's Fa Cup</td>
<td>105</td>
<td>Drama</td>
<td>1935</td>
<td>Warner Brothers</td>
<td>PG-13</td>
</tr>
<tr>
<td>Forrest Gump</td>
<td>140</td>
<td>Drama</td>
<td>1994</td>
<td>Paramount Pictures</td>
<td>PG-13</td>
</tr>
<tr>
<td>Fight Club</td>
<td>127</td>
<td>Drama</td>
<td>1995</td>
<td>Paramount Pictures</td>
<td>PG-13</td>
</tr>
<tr>
<td>Vieda</td>
<td>125</td>
<td>Action</td>
<td>1975</td>
<td>Universal Studios</td>
<td>R</td>
</tr>
<tr>
<td>Jurassic Park</td>
<td>127</td>
<td>Action</td>
<td>1993</td>
<td>Universal Pictures</td>
<td>PG-13</td>
</tr>
<tr>
<td>Lethal Weapon</td>
<td>110</td>
<td>Action/Comedy</td>
<td>1987</td>
<td>Warner Brothers</td>
<td>R</td>
</tr>
<tr>
<td>Michael</td>
<td>106</td>
<td>Drama</td>
<td>1990</td>
<td>Warner Brothers</td>
<td>PG-13</td>
</tr>
<tr>
<td>National Lampoon's Vacation</td>
<td>98</td>
<td>Comedy</td>
<td>1980</td>
<td>Warner Brothers</td>
<td>PG-13</td>
</tr>
<tr>
<td>Poltergeist</td>
<td>115</td>
<td>Horror</td>
<td>1982</td>
<td>MGM / UA</td>
<td>R</td>
</tr>
<tr>
<td>Reds</td>
<td>120</td>
<td>Action</td>
<td>1976</td>
<td>MGM / UA</td>
<td>R</td>
</tr>
<tr>
<td>Scarecrow</td>
<td>170</td>
<td>Action/Comedy</td>
<td>1987</td>
<td>Universal Studios</td>
<td>R</td>
</tr>
<tr>
<td>Silence of the Lambs</td>
<td>118</td>
<td>Drama</td>
<td>1991</td>
<td>Orion</td>
<td>R</td>
</tr>
<tr>
<td>Star Wars</td>
<td>124</td>
<td>Action/Sci-Fi</td>
<td>1977</td>
<td>Lucas Films Ltd</td>
<td>R</td>
</tr>
<tr>
<td>The Hunt for Red October</td>
<td>135</td>
<td>Action</td>
<td>1990</td>
<td>Paramount Pictures</td>
<td>PG-13</td>
</tr>
<tr>
<td>The Wizard of Oz</td>
<td>104</td>
<td>Adventure</td>
<td>1939</td>
<td>MGM / UA</td>
<td>R</td>
</tr>
<tr>
<td>Titanic</td>
<td>194</td>
<td>Drama</td>
<td>1997</td>
<td>Paramount Pictures</td>
<td>PG-13</td>
</tr>
</tbody>
</table>
The Brain and Human Perception

Human perception is the process of understanding how the brain recognizes, organizes and interprets the sensory stimulations in the world around us. We should care that our visuals are perceived in the way we want them to be when presenting data. Before developing and using any visual, we should have firm answers to the following questions. How do we want our visuals to be perceived by others? Are our visuals being understood by everyone viewing them? Finally, are our visuals being perceived the same way by different viewers?

To answer these questions about perception, Few (2008) suggests that context helps to shape our perception. He further qualifies this by stating, “Perception is influenced by what surrounds the visual.” To better understand this point, the image displayed in Figure 1, below, shows three colored rectangles (i.e., gray, blue and red) with a horizontal bar inside each rectangle. The intent of this image is to show that visuals can cause our brain to perceive things that are not true. Case in point – when observing the image are you able to discern which end of the horizontal bar in the center of each rectangle is darker? If you answered the left end of each horizontal bar is darker, then you’d be mistaken. The correct answer is the horizontal bar within each colored rectangle is the same shade at both ends. The inability to see this represents an optical illusion produced by the gradient differences in the colored background rectangles. Still don’t believe that each horizontal bar has the same color shading from the left edge to the right edge? Cover each of the colored background rectangles to unveil the same shading at both ends of each horizontal bar.

The optical illusion produced by the background color gradients produce two important tips that are worth following.

**Tip #1:** Make sure the background color being used is consistent and avoid gradients altogether.

**Tip #2:** Use background and foreground colors that contrast sufficiently with the visual being used.

Graphical Design Principles

Good graphical design begins with displaying data clearly and accurately. Data (and information) should be conveyed effectively and without ambiguity. Unnecessary information often distracts from the message, therefore it should be excluded. In their highly acclaimed book, Statistical Graphics Procedures by Example (2011), Mantange and Heath share this message about graphical output, “A graph is considered effective if it conveys the intended information in a way that can be understood quickly and without ambiguity by most consumers.”

Gestalt Principles of Visual Perception

Gestalt, borrowed from the world of psychology, means “unified whole.” The theory of visual perception was developed by German psychologists in the 1920s. For those interested, more information can be found at, [http://www.scholarpedia.org/article/Gestalt_principles](http://www.scholarpedia.org/article/Gestalt_principles).
Making Your SAS® Results More Meaningful with Color, continued

The SGPLOT Procedure Plot Type, Visual and Description
Whether the data you work with is large or small, or some size in between, the SGPLOT procedure is a powerful tool for handling many of your graphical needs. You’ll be able to let your visuals do the talking by helping your audience see hidden, or hard to see, things in your data, while avoiding the obvious by surprising and engaging your audience. PROC SGPLOT creates single-cell bar charts, box plots, bubble plots, dot plots, histograms, line plots, scatter plots, and an assortment of other plot types quickly and easily. The SGPLOT procedure supports the following plot types and output results.

<table>
<thead>
<tr>
<th>Plot Type</th>
<th>Visual</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bar</td>
<td><img src="image" alt="Bar Chart" /></td>
<td>Bar charts show the relationships among variables and the distribution of a categorical variable as horizontal (HBAR) or vertical (VBAR) output.</td>
</tr>
<tr>
<td>Box</td>
<td><img src="image" alt="Box Plot" /></td>
<td>Box charts are often used during data analysis for the purpose of showing the shape of a distribution, its central value, and its variability.</td>
</tr>
<tr>
<td>Bubble</td>
<td><img src="image" alt="Bubble Plot" /></td>
<td>Bubble plots produce three-dimensional output to help visualize data and show the relationship between categorized circles (bubbles).</td>
</tr>
<tr>
<td>Dot</td>
<td><img src="image" alt="Dot Plot" /></td>
<td>Dot plots provide an effective way to compare frequency counts within categories or groups of values.</td>
</tr>
<tr>
<td>Histogram</td>
<td><img src="image" alt="Histogram" /></td>
<td>Histograms provide an effective way to view continuous data where the categories represent ranges of data.</td>
</tr>
<tr>
<td>Line</td>
<td><img src="image" alt="Line Plot" /></td>
<td>Line plots provide a quick and easy way to show and organize the frequency of data as horizontal (HLINE) or vertical (VLINE) output.</td>
</tr>
</tbody>
</table>
Pie charts illustrate data proportionality by displaying the quantities of one or more (maximum of six) categories in the form of pie slices.

 Scatter plots are most effectively used to show how much one variable is affected by another.

 Series plots provide a convenient way to visualize numeric response data over a continuous axis.

 Vector charts provide a visual of the x and y values along with the length and direction of data values.

**Best Practice Graphical Techniques**

Best practice graphical techniques emphasize essential visual elements in a graph or chart for maximum effectiveness. The following best practice graphical techniques are valuable considerations as you design and develop visual output.

- Group similar elements together to differentiate them from other elements
- Use distinct colors to set elements apart
- Use different shapes to distinguish points on a line graph
- Varying 2-dimensional elements in a line graph offers an effective way to indicate differences, as well as how much difference exists
- Length serves as an excellent way to communicate differences between elements
- Size can show that differences exist but doesn’t show the amount of difference between elements
- Width is good for showing that differences exist but fails to show the actual difference between elements
- Color saturation is useful for showing distinctions and differences between elements.

**Tips for Designing Effective Bar Graphs**

- Avoid misleading scales by starting quantities at zero
- Verify that the vertical axis is about 25% shorter than the horizontal axis
- Verify that bars are all the same width
- Verify that bars are arranged in a logical sequence
- Use tick marks to display amounts
- Verify that the space between bars is about half as wide as the bar itself.
Making Your SAS® Results More Meaningful with Color, continued

**Tips for Designing Effective Line Graphs**

- Avoid misleading scales by starting quantities at zero
- Verify that the vertical axis is about 25% shorter than the horizontal axis
- Use grid lines to display quantities and amounts.

**Tips for Designing Effective Pie Charts**

- Limit the number of pie slices to no more than six
- Specify the largest pie slice at the top and work clockwise in decreasing order
- Combine “small” slices into a “Miscellaneous” or “Other” slice
- Label each pie slice inside the pie slice
- Emphasize a particular pie slice by color or by extracting it from the pie itself.

**Creating Visuals and Graphics with Statistical Graphic Procedures**

With the availability of ODS Statistical Graphics the SAS user community has a powerful set of tools to produce high-quality graphical output for data exploration, data analysis, and statistical analysis. All that is required to leverage the power available in ODS Statistical Graphics is a Base SAS® license. With that, ODS Statistical Graphics and its many capabilities are available to SAS software users in the form three procedures: SGPLOT, SGSCATTER and SGPANEL.

**The SGPLOT Procedure by Example**

The syntax for the SGPLOT procedure is displayed below.

```sas
PROC SGPLOT DATA=data-set-name > < options > ;
  Plot-statement(s) plot-request-parameters < / options > ;
RUN ;
```

**Creating a Bar Chart**

Bar charts are a commonly used graph type for displaying and comparing the quantities, frequencies or other measurements of discrete categories or groups. The selected variable is displayed along the horizontal (x-axis) and the frequency along the vertical (y-axis). The next SGPLOT displays the distinct values for the RATING variable using a VBAR statement.

```sas
TITLE 'Vertical Bar Chart' ;
PROC SGPLOT DATA=MOVIES ;
  VBAR RATING / GROUP=RATING ;
RUN ;
```
Creating a Grouped Vertical Bar Chart

Like the bar chart in the previous example, a grouped bar chart is a commonly used visual for displaying color and comparing the quantity, frequency or other measurements for discrete categories or groups. A grouped bar chart displays the selected variable along the horizontal (x-axis) and uses a grouping variable to display the various sub-category values for each discrete value of the VBAR variable. The next SGPLOT displays the distinct value(s) for the RATING variable and the distinct value(s) for the grouped variable, CATEGORY, along the horizontal axis, and the frequency value along the vertical axis in a VBAR statement. Since the mean value is computed and displayed at the top of each grouped bar for the CATEGORY variable, a FORMAT statement is specified to display values in the desired formatting.

```
TITLE 'Grouped Bar Chart' ;
PROC SGPLOT DATA=MOVIES ;
   FORMAT Length 5.2 ;
   VBAR RATING / RESPONSE=Length
               STAT=Mean
               GROUP=CATEGORY
               GROUPDISPLAY=Cluster
               DATALABEL ;
RUN ;
```

Creating a Stacked Vertical Bar Chart

Like the bar charts in the previous two examples, a stacked bar chart displays and compares the quantity, frequency or some other measurement for discrete categories or groups. A stacked bar chart displays the distinct values for the selected variable of the VBAR statement and stacks the grouping variable, designated with the GROUP= option) along the horizontal (x-axis). The next SGPLOT displays the distinct value(s) for the CATEGORY variable and stacks the distinct value(s) for the grouped variable, RATING, along the horizontal axis, and the frequency value along the vertical axis in a VBAR statement.

```
TITLE 'Stacked Bar Chart' ;
PROC SGPLOT DATA=MOVIES ;
   VBAR CATEGORY / RESPONSE=Length
               GROUP=RATING ;
RUN ;
```
Creating a Horizontal Box Plot

Box plots are useful for displaying data outliers and for comparing distributions for continuous variables. Box plots split the data into quartiles where the range of values traverses the first quartile (Q1) through the third quartile (Q3). The median of the data is represented by a vertical line drawn in the box at the Q2 quartile. Box plots also display the range (or spread) of the data indicating the distance between the smallest and largest value. The next SGPLOT shows the LENGTH variable in an HBOX statement with the RATING variable specified in the CATEGORY= option. The LENGTH variable is displayed on the horizontal (x-axis) and the categorical variable, RATING, is displayed on the vertical (y-axis) of the horizontal box plot output.

```
TITLE 'Horizontal Box Plot';
PROC SGPLOT DATA=MOVIES;
  HBOX Length /
    CATEGORY=Rating;
RUN;
```

Creating a Bubble Plot

Bubble plots provide a three-dimensional way to visualize data. Considered as a variation of the scatter plot, a bubble plot replaces the dots with bubbles providing an interesting way to view a set of values in three-dimensions. Visuals are created using a horizontal and vertical axis for the purpose of determining whether a possible relationship exists between one variable and another, along with a third variable to determine the size of each bubble. The relationship between two variables is often referred to as their correlation. Bubble plot output is produced with the SGPLOT procedure and a BUBBLE statement. In the following example, the resulting bubble plot contains the data points for the average LENGTH variable on the horizontal (x-axis) and the average YEAR variable on the vertical (y-axis) along with the number of Movies in each Rating group from the MOVIES data set.

```
PROC SQL;
  CREATE TABLE Movies_Summary as
    SELECT Title, Rating,
    COUNT(Rating) AS Count_Rating,
    AVG(Length) AS Avg_Length,
    AVG(Year) AS Avg_Year
     FROM Movies
     GROUP BY Rating;
QUIT;

PROC SGPLOT DATA=Movies_Summary;
  BUBBLE X=Avg_Length Y=Avg_Year
    SIZE=Count_Rating /
    GROUP=Rating
    DATALABEL=Rating;
RUN;
```
Creating a Histogram

Histograms are vertical bar charts that display the distribution of a set of numeric data. They are typically used to help organize and display data to gain a better understanding about how much variation the data has. Much can be learned from viewing the shape of a histogram. The various histogram shapes include: Bell-shape – displays most of the data clustered around the center of the x-axis, Bimodal – displays two data values occurring more frequently than any other, Skewed Right or Skewed Left – displays values that tend to be occurring around the high or low points of the x-axis, Uniform – displays data equally (no peaks) across a range of values. The next SGPLOT shows the LENGTH variable specified in a HISTOGRAM statement with the MOVIES data set as input. The LENGTH variable is displayed on the horizontal (x-axis) of the histogram output.

Combining a Histogram with a Density Plot

Histograms can also have a density curve applied to display the distribution of values for numeric data. The next SGPLOT shows the LENGTH variable specified in a HISTOGRAM statement combined with the LENGTH variable displayed in a density plot with the DENSITY statement. As before, the LENGTH variable is displayed on the horizontal (x-axis) of the histogram output.
Creating a Pie Chart

Pie charts display data proportionality by using one or more pie slices (maximum of six slices) to illustrate quantity. They are typically used to show the quantity of data categories by displaying pie slices to gain a better understanding of the quantity each category represents. Visualization experts caution users when selecting pie charts because they are difficult to compare data across different pie charts. The following Pie chart template is produced using the Base SAS’ TEMPLATE procedure, or statistical graphic language (SGL) along with the SGRENDER procedure. The pie chart shows the quantities of each RATING variable (i.e., “G”, “PG”, “PG-13”, and “R” rated movies) as pie slices.

```
proc template;
  define statgraph PieChart;
    begingraph;
      entrytitle "Pie Chart by Movie Rating";
      layout gridded / columns=1;
        layout region;
          piechart category=Rating;
        endlayout;
      endlayout;
    endgraph;
  end;
run;
quit;
proc sgrender data=mydata.Movies template=PieChart;
run;
quit;
```

Pie charts can also be created using much less code with PROC SGPIE. In the next example, a PROC SGPIE and the PIE statement are specified to show the quantities of each RATING variable (i.e., “G”, “PG”, “PG-13”, and “R” rated movies) as pie slices.

```
proc sgpie data=mydata.movies;
  title 'Pie Chart by Movie Rating';
  pie rating / datalabeldisplay=all;
run;
```
Creating a Scatter Plot

Scatter plots are typically produced to display data points on a horizontal and vertical axis for the purpose of determining whether a possible relationship exists between one variable and another. The relationship between two variables is often referred to as their correlation. The SCATTER plot output is produced with the SGPLOT procedure and a SCATTER statement. The resulting plot contains the data points for the LENGTH variable on the horizontal (x-axis) and the YEAR variable on the vertical (y-axis) from the MOVIES data set.

```
TITLE 'Basic Scatter Plot' ;
PROC SGPLOT DATA=MOVIES ;
  SCATTER X=Length Y=Year ;
RUN ;
```

Creating a Series Plot

Series (or Time Series) plots are useful for displaying trends on paired data at different points in time. Series plots display the time variable along the horizontal (x-axis) and data values along the vertical (y-axis). The next SGPLOT selects the YEAR (or time) variable along the horizontal axis and LENGTH along the vertical axis in a SERIES statement.

```
PROC SORT DATA=Movies
  OUT=Sorted_Movies ;
  BY Year ;
RUN ;

TITLE 'Series Plot' ;
PROC SGPLOT DATA=Sorted_Movies ;
  SERIES X=Year Y=Length /
    GROUP=Rating ;
RUN ;
```
The SGSCATTER Procedure by Example

Whether the data you work with is large or small, or some size in between, the SGSCATTER procedure is a powerful tool for handling many of your graphical needs. PROC SGSCATTER provides single-statement control to a number of scatter plot panels and matrices. The SGSCATTER procedure supports the following plot statements.

<table>
<thead>
<tr>
<th>Plot Statement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plot</td>
</tr>
<tr>
<td>Compare</td>
</tr>
<tr>
<td>Matrix</td>
</tr>
</tbody>
</table>

The syntax for the SGSCATTER procedure is displayed below.

```
PROC SGSCATTER < DATA=data-set-name > < options > ;
  Plot variable1 * variable2 / < options > ;
  Compare X=(variable(s)) Y=(variable(s)) < / options > ;
  Matrix variable(s) < / option(s) > ;
RUN ;
```

Creating a Scatter Plot with SGSCATTER

SGScatter plots display data points on a horizontal and vertical axis for the purpose of determining whether a possible relationship exists between one variable and another. The relationship between two variables is often referred to as their correlation. The SCATTER plot output is produced with the SGSCATTER procedure and a PLOT statement. The resulting scatter plot contains the data points for the LENGTH variable on the horizontal (x-axis) and YEAR variable on the vertical (y-axis) from the MOVIES data set.

```
TITLE 'SGScatter Plot';
PROC SGSCATTER DATA=MOVIES;
  PLOT Year * Length / GROUP=Rating 
                  DATALABEL;
RUN;
```
Creating a Scatter Matrix Plot with SGSCATTER

Scatter matrix plots display data points on a horizontal and vertical axis for the purpose of determining whether a possible relationship exists between one variable and another. The SCATTER plot output is produced with the SGSCATTER procedure and a MATRIX statement. The resulting scatter plot contains a two-pair scatter matrix plot from the MOVIES data set.

```
TITLE 'SGScatter Plot Matrix' ;
PROC SGSCATTER DATA=MOVIES ;
  MATRIX Year Length ;
RUN ;
```

The SGPANEL Procedure by Example

The SGPANEL procedure provides the ability to create a single-panel of plots or charts using one or more classification variables. Unlike a single-page plot or chart as is produced with the SGPLOT procedure, the SGPANEL procedure is able to produce a panel of plots or charts in a single image, or multiple plots or charts displayed in multiple panels. The result is a panel of plots or charts that display relationships among variables. The SGPANEL supports a number of plot and chart types including histograms, horizontal bar charts, vertical bar charts, horizontal box plots, and vertical box plots. The SGPANEL procedure supports the following statements.

SGPanel Statement

```
SGPanel Statement
  Panelby
  Other Plot / Chart Statements
```

The syntax for the SGPANEL procedure is displayed below.

```
PROC SGPANEL < DATA=data-set > < options > ;
  PANELBY classvar1 < classvar2 . . . < classvarn > / options ;
  < plot / chart statement > ;
RUN ;
```

Creating a Panel of Histograms with SGPANEL

A panel of plots and charts created with the SGPANEL procedure displays values of one or more classification variables. The purpose of a panel of plots is to be able to compare one or more variables with one another. The next example shows a panel of histograms created with the SGPANEL procedure and a HISTOGRAM statement. The resulting panel of histograms contains the number of movies (SCALE=Count) by category corresponding to the year the movie was produced from the MOVIES data set.
PROC SGPANEL DATA=MOVIES;
PANELBY CATEGORY / ROWS=3 COLUMNS=4;
HISTOGRAM YEAR / SCALE=COUNT;
RUN;

Creating a Panel of Bar Charts with SGPANEL

A panel of bar charts can be created with the SGPANEL procedure using one or more classification variables. As before, the purpose of a panel of bar charts could be for comparing one or more variables values against another. The next example shows a panel of vertical bar charts created with the SGPANEL procedure and a VBAR statement. The resulting panel of vertical bar charts contains the average movie lengths (STAT=Mean) for each category of movie corresponding to the movie rating (i.e., G, PG, PG-13, and R) from the MOVIES data set.

TITLE 'VBAR Chart with SGPANEL';
PROC SGPANEL DATA=MOVIES;
PANELBY CATEGORY / ROWS=3 COLUMNS=4;
VBAR RATING / GROUP=Rating RESPONSE=Length STAT=Mean DATALABEL;
RUN;

Using Color in Results, Reports, Tables, Charts and Spreadsheets

The use of color in results, reports, tables, charts and spreadsheets can help to organize, engage, promote, and encourage greater comprehension with an audience. Color can help make SAS® results, reports, tables, charts and spreadsheets more professional and meaningful. Instead of producing boring and ineffective results, users can enhance the appearance of their end product to highlight and draw attention to important data elements and issues, including headings, subheadings, footers, minimum and maximum values, ranges, outliers, special conditions, and other elements. Color can be added to text, foreground, background, row, column, cell, summary, and total with amazing color and traffic lighting scenarios. The following examples illustrate the use of the ODS Statistical Graphic procedures to produce high-quality graphical output for data exploration, data analysis, and statistical analysis.
Using Color in Bar Charts with SGPLOT

Colors and the various color schemes can be applied to chart elements in many ways. Pre-set or fixed colors are available “right-out-of-the-box” with the SGPLOT procedures. Options are available to change the “default” colors and color schemes to address specific needs and requirements. In the next example, the SGPLOT procedure produces a vertical bar (VBAR) chart that uses pre-set “default” colors.

```
TITLE 'VBAR Chart with SGPLOT' ;
PROC SGPLOT DATA=MOVIES ;
  VBAR Rating / GROUP=Rating 
      DATALABEL 
      FILL 
      GROUPDISPLAY=Cluster ;
RUN ;
```

In the next example, the SGPLOT procedure is specified to produce a vertical bar (VBAR) chart with “custom” colors. By defining an attribute map (ATTRMAP) SAS data set with information about the values and colors to use during the creation of the bar chart, users have the ability to override and control which colors, and color schemes, they desire. In producing the vertical bar chart with the SGPLOT procedure, the DATTRMAP= procedure option references the attribute map data set, the GROUPDISPLAY=Cluster and ATTRID=Rating VBAR statement options are specified.

```
DATA ATTRMAP ;
  INPUT @1 ID $6. 
      @8 Value $5. 
      @14 FillColor $6. ;
  DATALINES ;
  Rating G     Green 
  Rating PG    Blue 
  Rating PG-13 Yellow 
  Rating R     Red 
; 
RUN ;
TITLE 'VBAR Chart with SGPLOT' ;
PROC SGPLOT DATA=MOVIES 
   DATTRMAP=ATTRMAP ;
  VBAR Rating / GROUP=Rating 
      DATALABEL 
      FILL 
      GROUPDISPLAY=Cluster 
      ATTRID=Rating 
; 
RUN ;
```
Using Color in Series Plots with SGPLOT

As was shown in the previous section, colors and various color schemes can be applied to plot elements. Pre-set or fixed colors are available “right-out-of-the-box” with the SGPLOT procedures. Options are available to change the “default” colors and color schemes to address specific needs and requirements. In the next example, the SGPLOT procedure produces a series (SERIES) and scatter (SCATTER) plot that uses the pre-set or “default” colors.

PROC SORT DATA=mydata.Movies
   OUT=work.Movies_Sorted ;
   BY Year ;
RUN ;
TITLE 'Response over Year by Rating' ;
PROC SGPLOT DATA=work.Movies_Sorted ;
   SERIES X=Year Y=Length /
      GROUP=Rating
      LINEATTRS=(THICKNESS=3) ;
   SCATTER X=Year Y=Length /
      GROUP=Rating
      MARKERATTRS=(SYMBOL=CIRCLEFILLED SIZE=11) ;
   XAXIS DISPLAY=(NOLABEL) ;
   YAXIS LABEL='Response' ;
RUN ;

In the next example, the SGPLOT procedure is specified along with a user-defined attribute map (ATTRMAP) to produce a series (SERIES) and scatter (SCATTER) plot with “custom” colors. By defining an attribute map (ATTRMAP) SAS data set with information about the values and colors to use during the creation of the series plot, users have the ability to override and control which colors, and color schemes, they desire. In producing the series plot with the SGPLOT procedure, the DATTRMAP= procedure option references the attribute map data set, the GROUPDISPLAY=Cluster, the ATTRID=Rating, and DATALABEL=Rating SERIES statement options are specified.

DATA ATTRMAP ;
   INPUT @1 ID $6.
      @8 Value $5.
      @14 LineColor $6.
      @21 MarkerColor $6. ;
   DATALINES ;
   Rating G Green
   Rating PG Blue
   Rating PG-13 Yellow
   Rating R Red
   ;
RUN ;
PROC SORT DATA=Movies
   OUT=work.Movies_Sorted ;
   BY Year ;
RUN ;
TITLE 'Response over Year by Rating' ;
PROC SGPLOT DATA=work.Movies_Sorted
   DATTRMAP=ATTRMAP ;
   SERIES X=Year Y=Length /
GROUP=Rating
LINEATTRS=(THICKNESS=3)
GROUPDISPLAY=Cluster
ATTRID=Rating
DATALABEL=Rating ;
SCATTER X=Year
    Y=Length / GROUP=Rating
    MARKERATTRS=
        (SYMBOL=CIRCLEFILLED SIZE=11)
    ATTRID=Rating
    XAXIS DISPLAY=(NOLABEL) ;
    YAXIS LABEL='Response' ;
RUN ;

Using Color in Excel Spreadsheets

The use of background and foreground colors in Excel spreadsheets should be chosen carefully and contrast sufficiently with the visual being used. This becomes ever so obvious as we examine the content of an Excel spreadsheet, illustrated in Figure 2, below. As shown, the blue background and black foreground (text) colors make the spreadsheet content difficult to read, particularly because of the insufficient contrast between the black text and blue background colors. Also, the colors used in the traffic lighting for the “Vehicle MSRP” column are difficult to read and distinguish. Color contrast issues should, and must, be avoided to have the results be readable to an audience.

Figure 2. Color contrast issues identified in an Excel spreadsheet

By specifying the following PROC FORMAT, PROC REPORT and ODS Excel code to address the color contrast issues identified and shown in the aforementioned Excel spreadsheet, we’re able to make the spreadsheet’s content more readable and, as a result, more appealing. The code begins by defining the traffic lighting scenario (i.e., “G-rated” movies are shaded in ‘Green’, “PG-rated” movies are shaded in ‘Light Blue’, “PG-13-rated” movies are shaded in ‘Orange’, and “R-rated” movies are shaded in ‘Red’) we want using a user-defined format. Then, we specify an ODS Excel statement to create our Excel spreadsheet. Finally, a PROC REPORT is specified to produce the desired layout for the content of our Excel spreadsheet along with the traffic lighting scenario we want displayed for each of the movie ratings. The resulting Excel spreadsheet, shown in Figure 3, illustrates the “default” style, along with the background and foreground color customizations that were specified to resolve the color contrast issues described earlier.
**SAS Code to Customize Excel Spreadsheet:**

```sas
PROC FORMAT;
   Value $RatingFmt
      'G'   = 'Green'
      'PG'  = 'Light Blue'
      'PG-13' = 'Orange'
      'R'   = 'Red';
RUN;

ODS Excel file='D:/Movie Rating with Traffic Lighting.xlsx'
   style=styles.HTMLBlue;

PROC REPORT DATA=WORK.Movies_Sorted NOWINDOWS
   STYLE(Header)={BackGround=White ForeGround=Black Font=(Calibri,10pt,Bold)};
   COLUMNS Title Length Category Year Studio Rating;
   DEFINE Title    /
      DISPLAY    'Movie Title'        WIDTH=25 ;
   DEFINE Length   /
      DISPLAY    'Movie Length'       WIDTH=12  CENTER ;
   DEFINE Category /
      DISPLAY    'Movie Category'     WIDTH=20 ;
   DEFINE Year     /
      DISPLAY    'Year of Movie'      WIDTH=13  CENTER ;
   DEFINE Studio   /
      DISPLAY    'Studio'             WIDTH=25 ;
   DEFINE Rating   /
      DISPLAY    'Movie Rating'       WIDTH=12  CENTER
      STYLE(Column)=[FontWeight=bold BackGround=$RatingFmt.];
RUN;

ODS Excel close;
```

**Customized Excel Spreadsheet:**

![Figure 3. Color contrast issues addressed in an Excel spreadsheet](image)

**Conclusion**

Over the years numerous studies have been conducted to determine how color can be used to help humans assimilate information from pictures, images and visuals along with the best use of color in reports, tables and spreadsheets. Shugars (2018) describes that 4.5% of the world’s population possesses some level of color vision deficiency with the largest segment of this population being affected by red and green colors. Consequently, many experts suggest that the best way to gain the largest audience participation is to replace red and green colors with yellow and darker blue colors. In another very different study conducted by social media scientist, Dan Zarrella (2013), found that tweets with images are 94% more likely to be retweeted than tweets without images.
SAS users are encouraged to use ODS Statistical Graphics procedures: SGPLOT, SGPANEL and SGSCATTER. Numerous tips and guidelines were introduced to use visuals and color to assimilate information contained in reports, tables and spreadsheets. Through the use of effective visual techniques the ability to better understand data can be achieved. Good graphical design begins with displaying data clearly, accurately, and without ambiguity, and when unnecessary information distracts from the message, it should be excluded.

**References**


**Acknowledgments**

The author thanks the 2021 PharmaSUG Conference Committee for accepting my abstract and paper; the 2021 PharmaSUG Executive Committee for organizing and supporting this virtual conference event; SAS Institute Inc. for providing SAS users with wonderful software; and SAS users everywhere for being the nicest people anywhere!
Making Your SAS® Results More Meaningful with Color, continued

Trademark Citations
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.

About the Author
Kirk Paul Lafler is a lecturer and adjunct professor at San Diego State University; an advisor and adjunct professor at the University of California San Diego Extension; and teaches SAS, SQL, Python, R and Excel courses, seminars, workshops, and webinars to students, professionals and users around the world. Kirk has been a SAS user since 1979 and is the author of several books including, PROC SQL: Beyond the Basics Using SAS, Third Edition (SAS Press. 2019) along with papers and articles on a variety of SAS topics. Kirk has also been selected as an Invited speaker, educator, keynote and section leader at SAS conferences; and is the recipient of 25 “Best” contributed paper, hands-on workshop (HOW), and poster awards.

Comments and suggestions can be sent to:

Kirk Paul Lafler
SAS® Consultant, Application Developer, Programmer, Data Analyst, Educator and Author
E-mail: KirkLafler@cs.com
LinkedIn: https://www.linkedin.com/in/KirkPaulLafler/
LinkedIn: https://www.linkedin.com/in/Order-of-Magnitude-Analytics/
Twitter: @sasNerd