Making Your SAS® Results More Meaningful with Color

Kirk Paul Lafler, Software Intelligence Corporation, Spring Valley, California

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.

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 were presented to effectively use visuals and color to assimilate information contained in reports, tables and spreadsheets.

It's commonly understood that data patterns and differences are not always obvious from tables and reports. This is where graphical output and their 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.

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.

	Title	Length	Category	Year	Studio	Rating
1	Brave Heart	177	Action Adventure	1995	Paramount Pictures	R
2	Casablanca	103	Drama	1942	MGM / UA	PG
3	Christmas Vacation	97	Comedy	1989	Warner Brothers	PG-13
4	Coming to America	116	Comedy	1988	Paramount Pictures	R
5	Dracula	130	Horror	1993	Columbia TriStar	R
6	Dressed to Kill	105	Drama Mysteries	1980	Filmways Pictures	R
7	Forrest Gump	142	Drama	1994	Paramount Pictures	PG-13
8	Ghost	127	Drama Romance	1990	Paramount Pictures	PG-13
9	Jaws	125	Action Adventure	1975	Universal Studios	PG
10	Jurassic Park	127	Action	1993	Universal Pictures	PG-13
11	Lethal Weapon	110	Action Cops & Robber	1987	Warner Brothers	R
12	Michael	106	Drama	1997	Warner Brothers	PG-13
13	National Lampoon's Vacation	98	Comedy	1983	Warner Brothers	PG-13
14	Poltergeist	115	Horror	1982	MGM / UA	PG
15	Rocky	120	Action Adventure	1976	MGM / UA	PG
16	Scarface	170	Action Cops & Robber	1983	Universal Studios	R
17	Silence of the Lambs	118	Drama Suspense	1991	Orion	R
18	Star Wars	124	Action Sci-Fi	1977	Lucas Film Ltd	PG
19	The Hunt for Red October	135	Action Adventure	1989	Paramount Pictures	PG
20	The Terminator	108	Action Sci-Fi	1984	Live Entertainment	R
21	The Wizard of Oz	101	Adventure	1939	MGM / UA	G
22	Titanic	194	Drama Romance	1997	Paramount Pictures	PG-13

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.
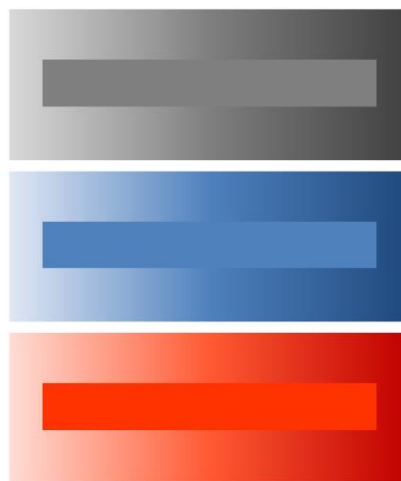

Figure 1. Optical illusion produced by background color gradients

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.
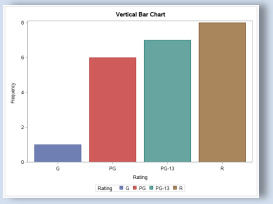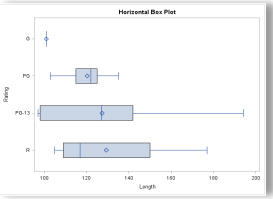
Graphical Design Principles

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."*
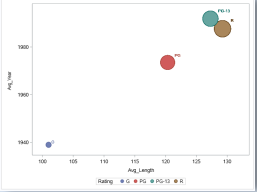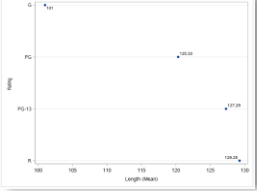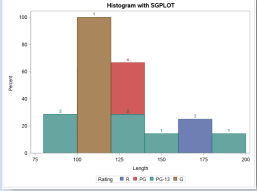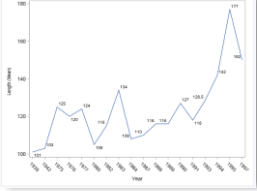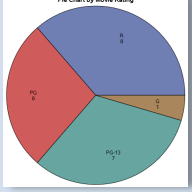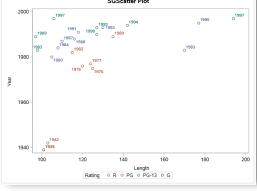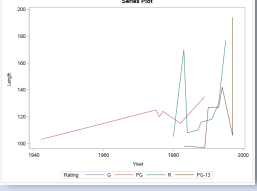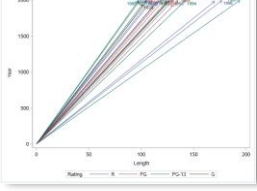
Gestalt Principles of Visual Perception

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.

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.

Plot Type	Visual	Description
Bar		Bar charts show the relationships among variables and the distribution of a categorical variable as horizontal (HBAR) or vertical (VBAR) output.
Box		Box charts are often used during data analysis for the purpose of showing the shape of a distribution, its central value, and its variability.

Bubble		**Bubble plots produce three-dimensional output to help visualize data and show the relationship between categorized circles (bubbles).**
Dot		**Dot plots provide an effective way to compare frequency counts within categories or groups of values.**
Histogram		**Histograms provide an effective way to view continuous data where the categories represent ranges of data.**
Line		**Line plots provide a quick and easy way to show and organize the frequency of data as horizontal (HLINE) or vertical (VLINE) output.**
Pie		**Pie charts illustrate data proportionality by displaying the quantities of one or more (maximum of six) categories in the form of pie slices.**
Scatter		**Scatter plots are most effectively used to show how much one variable is affected by another.**
Series		**Series plots provide a convenient way to visualize numeric response data over a continuous axis.**
Vector		**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

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.

```
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.

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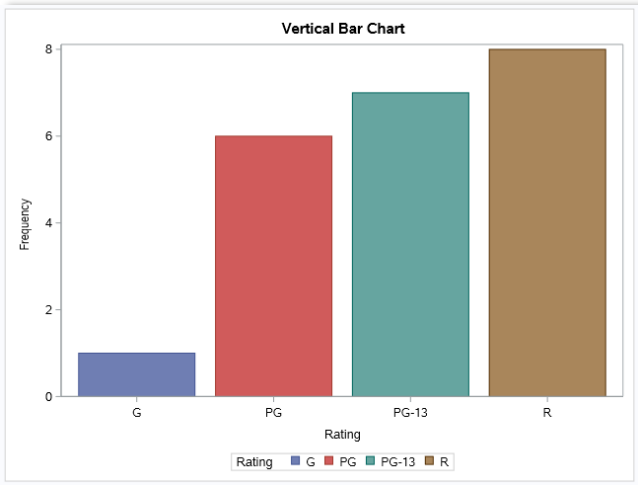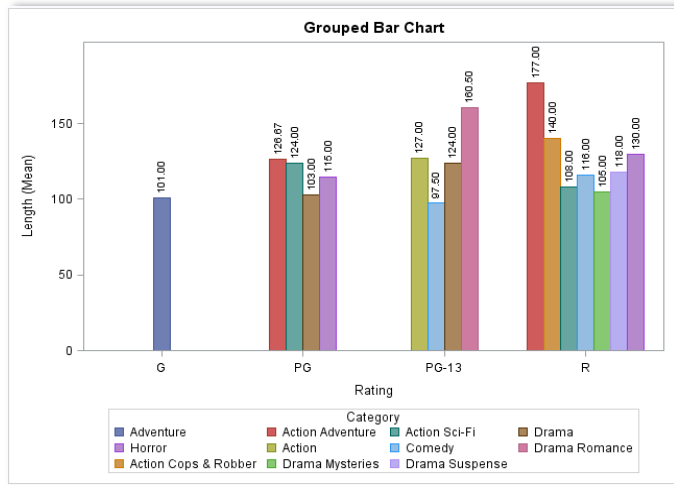

Creating a Grouped Vertical Bar Chart

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 ;
```
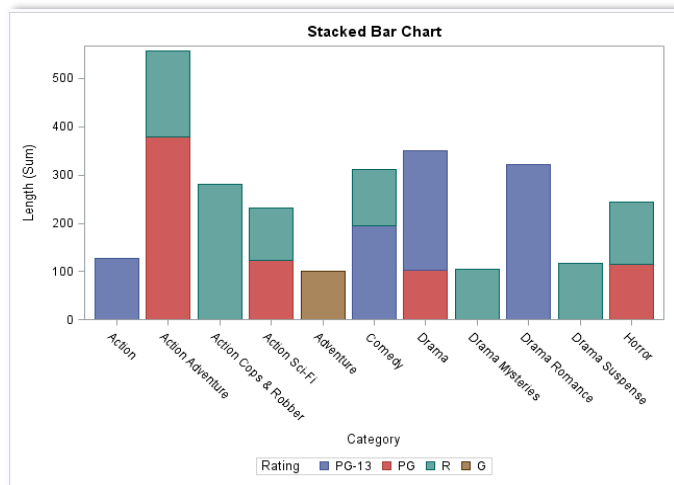
Creating a Horizontal Box Plot

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 ;
```
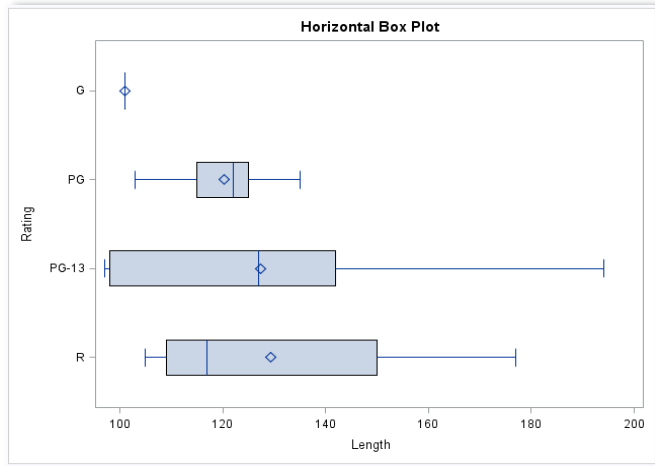

Creating a Bubble Plot

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 ;
```
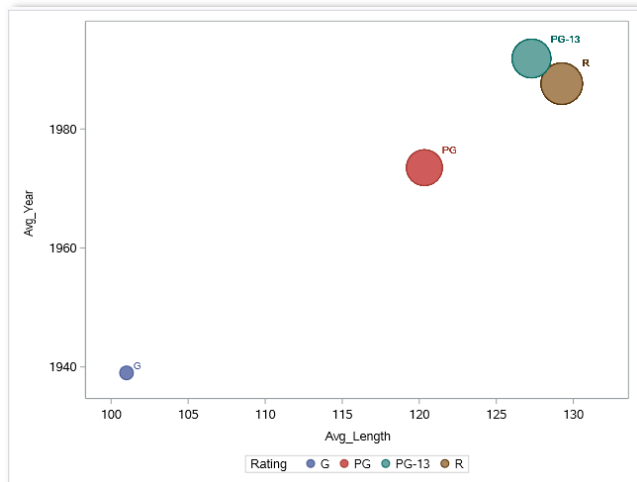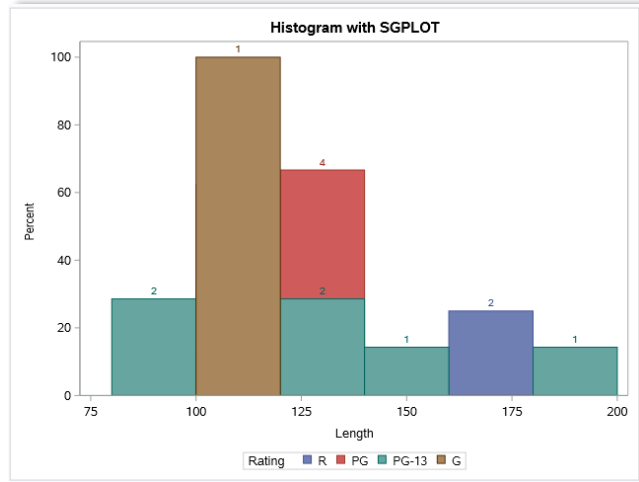
Creating a Histogram

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.

```
TITLE 'Histogram with SGPLOT' ;
PROC SGPLOT DATA=MOVIES ;
  HISTOGRAM Length / GROUP=Rating
               DATALABEL=Count ;
RUN ;
```


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.

```
TITLE 'Histogram with Density' ;
PROC SGPLOT DATA=MOVIES ;
  HISTOGRAM Length /
              DATALABEL=Count ;
  DENSITY Length ;
RUN ;
```
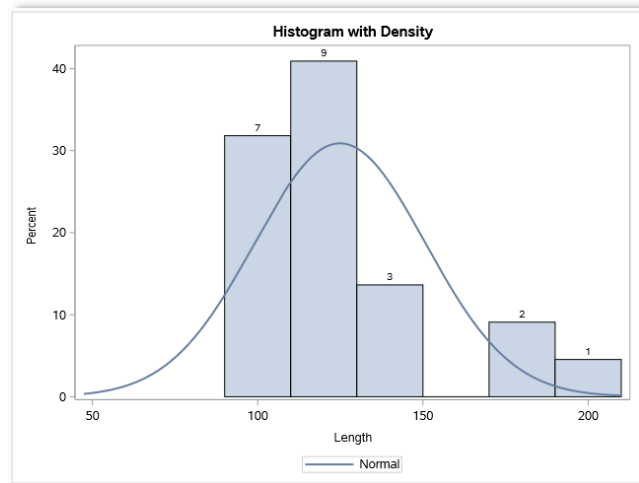

Creating a Pie Chart

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 ;
```
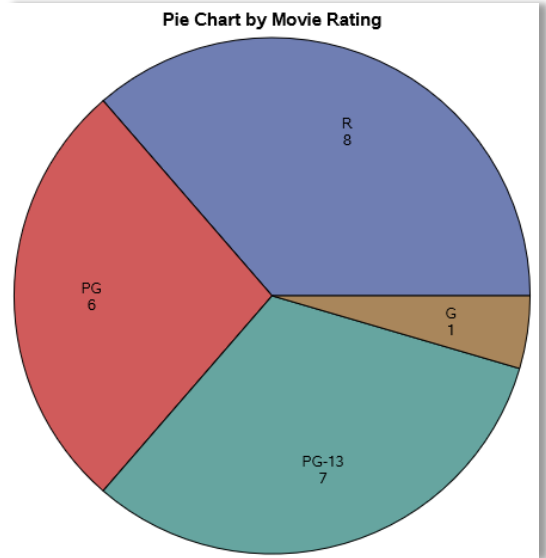
Creating a Scatter Plot

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 ;
```
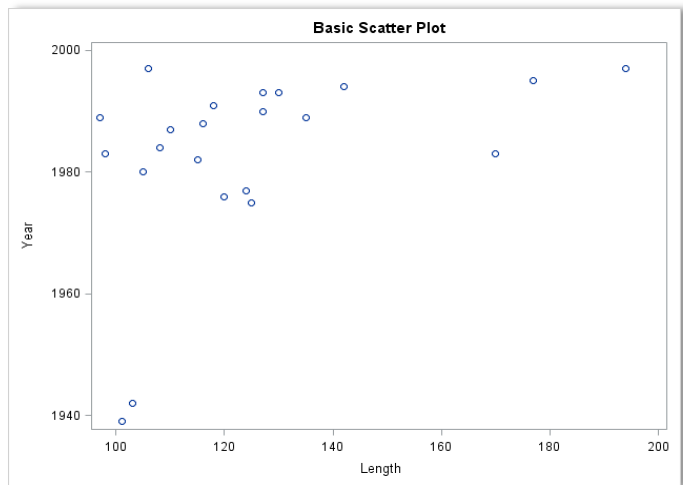
Creating a Series Plot

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 ;
```
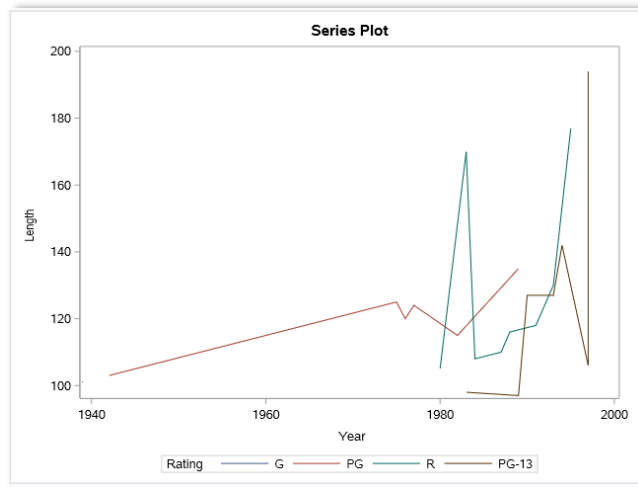

The SGSCATTER Procedure by Example

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.

Plot Statement
Plot
Compare
Matrix

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 ;
```
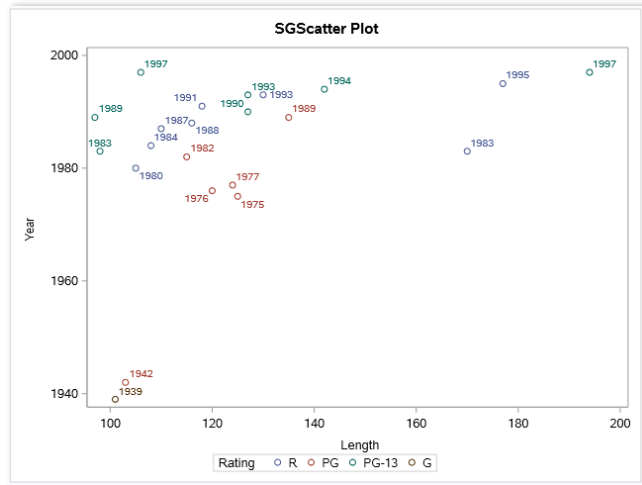
Creating a Scatter Matrix Plot with SGSCATTER

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 ;
```
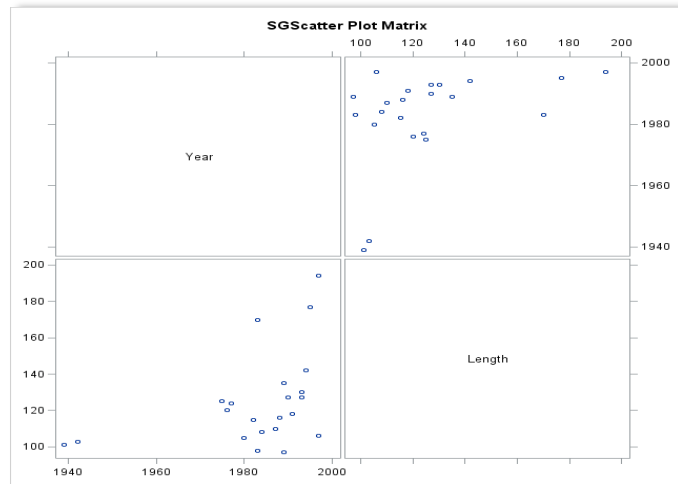
The SGPANEL Procedure by Example

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 hart 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
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 ;
```
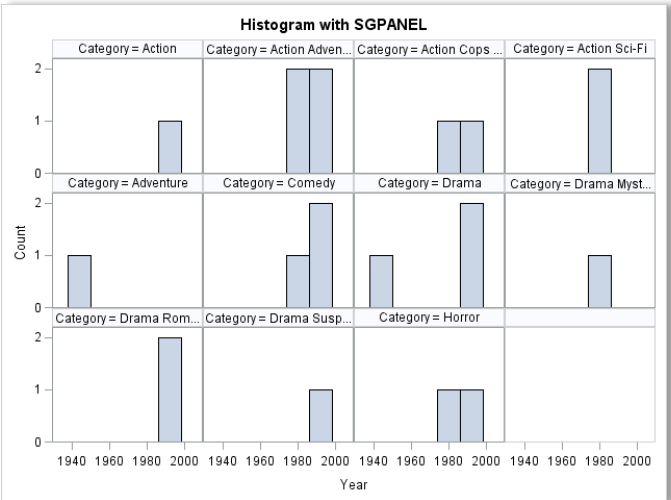
Creating a Panel of Bar Charts with SGPANEL

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 ;
```
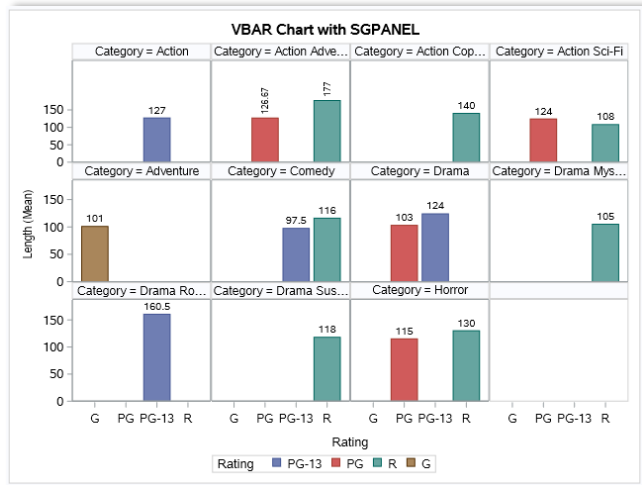
## Using Color in Results, Reports, Tables, Charts and Spreadsheets

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 ;
```
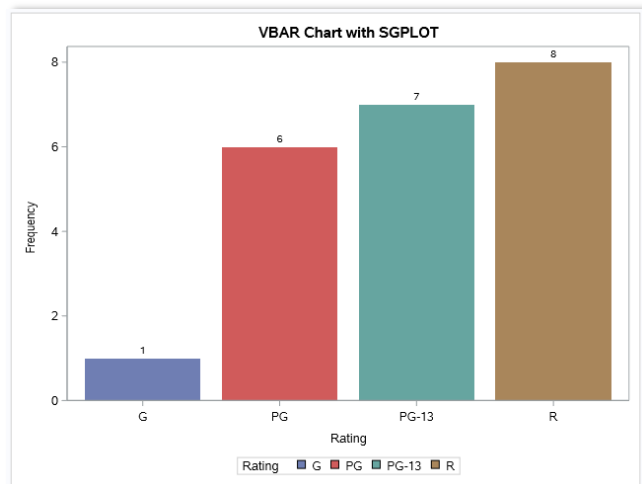
In the next example, the SGPLOT procedure is specified to produce a vertical bar (VBAR) chart with "custom" colors. By defining an attribute map (ATTRMAP) SAS data set with information about the values and colors to use during the creation of the bar chart, users have the ability to override and control 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 ;
```
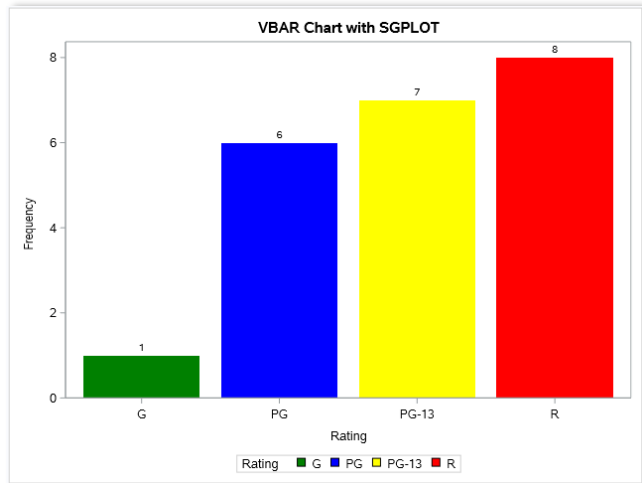

Using Color in Series Plots with SGPLOT

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 ;
```
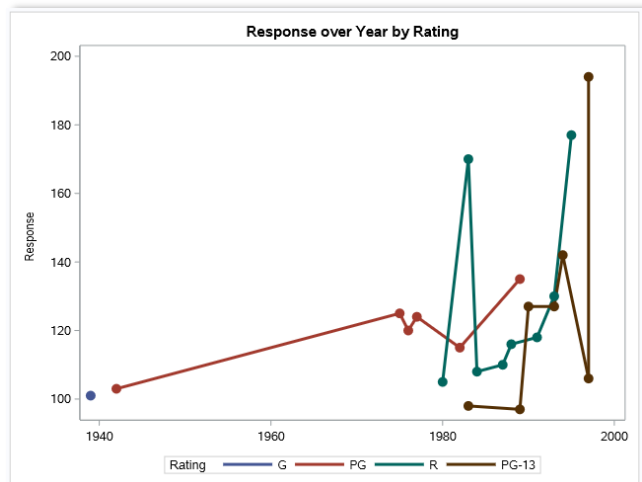

In the next example, the SGPLOT procedure is specified along with a user-defined attribute map (ATTRMAP) to produce a series (SERIES) and scatter (SCATTER) plot with "custom" colors. By defining an attribute map (ATTRMAP) SAS data set with information about the values and colors to use during the creation 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 ;
```
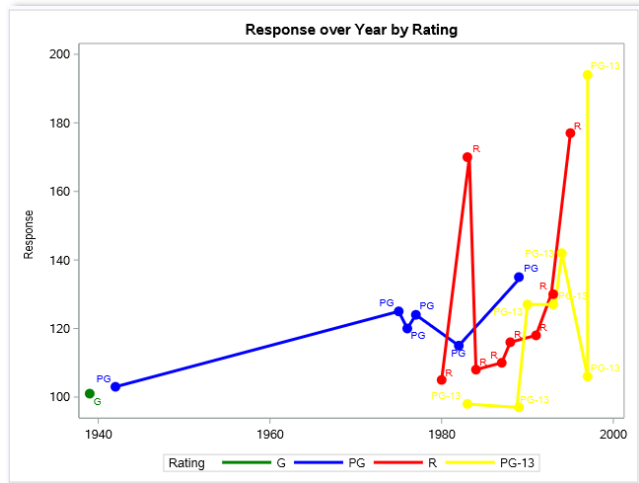
Using Color in Excel Spreadsheets

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


ODS Excel file='/folders/myfolders/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:

	A	B	C	D	E	F
1	Movie Title	Movie Length	Movie Category	Year of Movie	Studio	Movie Rating
2	Brave Heart	177	Action Adventure	1995	Paramount Pictures	R
3	Casablanca	103	Drama	1942	MGM / UA	PG
4	Christmas Vacation	97	Comedy	1989	Warner Brothers	PG-13
5	Coming to America	116	Comedy	1988	Paramount Pictures	R
6	Dracula	130	Horror	1993	Columbia TriStar	R
7	Dressed to Kill	105	Drama Mysteries	1980	Filmways Pictures	R
8	Forrest Gump	142	Drama	1994	Paramount Pictures	PG-13
9	Ghost	127	Drama Romance	1990	Paramount Pictures	PG-13
10	Jaws	125	Action Adventure	1975	Universal Studios	PG
11	Jurassic Park	127	Action	1993	Universal Pictures	PG-13
12	Lethal Weapon	110	Action Cops & Robber	1987	Warner Brothers	R
13	Michael	106	Drama	1997	Warner Brothers	PG-13
14	National Lampoon's Vacation	98	Comedy	1983	Warner Brothers	PG-13
15	Poltergeist	115	Horror	1982	MGM / UA	PG
16	Rocky	120	Action Adventure	1976	MGM / UA	PG
17	Scarface	170	Action Cops & Robber	1983	Universal Studios	R
18	Silence of the Lambs	118	Drama Suspense	1991	Orion	R
19	Star Wars	124	Action Sci-Fi	1977	Lucas Film Ltd	PG
20	The Hunt for Red October	135	Action Adventure	1989	Paramount Pictures	PG
21	The Terminator	108	Action Sci-Fi	1984	Live Entertainment	R
22	The Wizard of Oz	101	Adventure	1939	MGM / UA	G
23	Titanic	194	Drama Romance	1997	Paramount Pictures	PG-13

Figure 3. Color contrast issues addressed in an Excel spreadsheet

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.

This paper introduced 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 were presented to effectively 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. It's commonly understood that data patterns and differences are not always obvious from reports, tables, and spreadsheets that omit visuals and color. 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

Allison, Robert (2014), *"How to create a bubble plot in SAS University Edition,"* Copyright 2014 by Robert Allison, blogs.sas.com, SAS Institute Inc., Cary, NC, USA.

Bessler, LeRoy. 2014. "Communication-Effective Data Visualization: Design Principles, Widely Usable Graphic Examples, and Code for Visual Data Insights." Proceedings of the SAS Global 2014 Conference, Cary, NC, USA: SAS Institute Inc. Available at http://support.sas.com/resources/papers/proceedings14/1464-2014.pdf.

Few, Stephen (2008), *"Practical Rules for Using Color in Charts,"* Copyright 2008 by Stephen Few, Perceptual Edge.

Kincaid, Chuck (2010), *"SGPANEL: Telling the Story Better,"* Proceedings of the 2010 SAS Global Forum (SGF) Conference, COMSYS, Portage, MI, USA.

Lafler, Kirk Paul; Joshua M. Horstman and Roger D. Muller (2017), *"Building a Better Dashboard Using SAS® Base Software,"* Proceedings of the 2017 Pharmaceutical SAS Users Group (PharmaSUG) Conference.

Lafler, Kirk Paul; Joshua M. Horstman and Roger D. Muller (2016), *"Building a Better Dashboard Using SAS® Base Software,"* Proceedings of the 2016 Pharmaceutical SAS Users Group (PharmaSUG) Conference.

Lafler, Kirk Paul (2016), *"Dynamic Dashboards Using SAS®,"* Proceedings of the 2016 SAS Global Forum (SGF) Conference, Software Intelligence Corporation, Spring Valley, CA, USA.

Lafler, Kirk Paul (2015), *"Dynamic Dashboards Using Base SAS® Software,"* Proceedings of the 2015 South Central SAS Users Group (SCSUG) Conference, Software Intelligence Corporation, Spring Valley, CA, USA.

Lafler, Kirk Paul (2015), *"Dynamic Dashboards Using SAS®,"* Proceedings of the 2015 SAS Global Forum (SGF) Conference, Software Intelligence Corporation, Spring Valley, CA, USA.

Matange, Sanjay and Dan Heath (2011), Statistical Graphics Procedures by Example, SAS Institute Inc., Cary, NC, USA. Click to view the book at the SAS Book store.

Sams, Scott (2013), *"SAS® BI Dashboard: Interactive, Data-Driven Dashboard Applications Made Easy,"* Proceedings of the 2013 SAS Global Forum (SGF) Conference, SAS Institute Inc, Cary, NC, USA.

Slaughter, Susan J. and Lora D. Delwiche (2010), *"Using PROC SGPLOT for Quick High-Quality Graphs,"* Proceedings of the 2010 SAS Global Forum (SGF) Conference, SAS Institute Inc, Cary, NC, USA.

Shugars, Michelle (2018), *"Using color to increase the Visual Effectiveness of Reports,"* Copyright 2018 by Michelle Shugars, https://www.megaputer.com/report-colors-increase-visual-effectiveness/.

Umar, Hafiz (2015), *"Data Visualization: Choose The Right Chart Type,"* Copyright 2015 by Hafiz Umar, https://www.linkedin.com/pulse/data-visualization-choose-right-chart-type-hafiz/.

Zarrilla, Dan (2013), *"Use Images on Twitter to Get More ReTweets,"* Copyright 2013 by Dan Zarrilla, http://danzarrella.com/use-images-on-twitter-to-get-more-retweets/.

Zdeb, Mike (2004), *"Pop-Ups, Drill-Downs, and Animation,"* Proceedings of the 2004 SAS Users Group International (SUGI) Conference, University at Albany School of Public Health, Rensselaer, NY, USA.

Trademark Citations

About the Author

Kirk Paul Lafler is an entrepreneur, founder and consultant at Software Intelligence Corporation, and has been using SAS software since 1979. As a SAS application developer, programmer, certified professional, provider of SAS consulting services, mentor, lecturer and adjunct professor at San Diego State University, advisor and adjunct professor at University of California San Diego Extension, and emeritus sasCommunity.org Advisory Board member, Kirk has taught SAS courses, seminars, workshops and webinars to thousands of users around the world.. As the author of seven books including PROC SQL: Beyond the Basics Using SAS, Third Edition (SAS Press. 2019) and Google® Search Complete (Odyssey Press. 2014); and hundreds of papers and articles; Kirk has been selected as an Invited speaker, trainer, keynote and section leader at SAS International, regional, special-interest, local, and in-house user group conferences and meetings; 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

Software Intelligence Corporation

E-mail: KirkLafler@cs.com

LinkedIn: https://www.linkedin.com/in/KirkPaulLafler/

LinkedIn: https://www.linkedin.com/in/Order-of-Magnitude-Analytics/

Twitter: @sasNerd