Graphical Representation of Clinical Data using Heat Map Graphs

Pavani Potuganti, and Sree Nalluri, Eliassen Group Biometrics and Data Solutions

ABSTRACT

A heat map graph is a visualization technique of representing data in a two-dimensional form using variations in colors to show different data values like extreme colors represents extreme values. The goal of heat map is to provide a colored visual summary of information and giving a better insight. Heat Maps are commonly used to monitor region wise weather, thermal values, geographical heights, and migrations of people etc. In clinical trial domains heat map graphs can be used to represent adverse events by treatment, metabolic abnormalities, infection rate in children vs young adults or male vs female etc. This paper demonstrates the relative efficiency of treatment groups versus placebo with a practical example using SAS® procedure and techniques enhancing the graph.

INTRODUCTION

Graphs can be effective to visualize the data which might be difficult to see in tabular form. If you want to display as much information as possible in as little space, ways like heat maps can be helpful. Heat maps are versatile and self-explanatory and are used across various industries. Data values are presented as rectangular areas or cells with varying colors and patterns here are usually clear. The darker the color, the higher the value. Using these graphs, we can plot either numeric or non-numeric data using a solid color or gradient of colors to indicate different ranges within data values.

HEAT MAPS AND RELATED SAS® PROCEDURE

SAS tools are of great help in creating heat maps. While there are other ways like Graph Template Language etc. to create heat map graphs, as of SAS® 9.4, statements like HEATMAPPARM and HEATMAP in SGPLOT and SGPANEL procedures can be used to create heat maps. The syntax of HEATMAPPARM and HEATMAP statements is mostly same for both SGPLOT and SGPANEL procedures. This paper primarily focuses on SGPLOT.

HEATMAPPARM statement creates a plot, which represents the values of a response variable with different colors. This statement requires X, Y and either one of COLORRESPONSE or COLORGROUP arguments. The COLORRESPONSE argument only supports numeric variables and colors the cells continuously using a gradient of colors.

In order to demonstrate heat maps in this paper, a sample dataset with 5270 observations and consisting of 4 variables like subject id, system organ class of adverse events, and treatment group variables have been used. The frequency count of adverse events for each SOC by treatment group has been calculated as shown below and this dataset SOCFREQ, consisting of the treatment group variable TRT, system organ class variable AESOC and frequency count variable COUNT has been used in Figure 1 - Figure 4.

```
proc freq data=sample;
 table trt*aesoc/ out=socfreq;
run;
```

PROCEDURE SGPLOT WITH HEATMAPPARM STATEMENT AND RELATED OPTIONS:

A few examples of heat maps created using SGPLOT procedure with HEATMAPPARM statement are presented in Figure 1 - Figure 4. These heat maps show the frequency of adverse events in each SOC by treatment group. Based on the colors we can see which system organ class has more adverse events or least adverse events and in which treatment group.

The graph below plots the independent variables AESOC on Y-axis against TRT (Treatment variable) on X-axis and colors the cells to represent the values of the response variable COUNT (Frequency count) using HEATMAPPARM statement. Since we are using COLORRESPONSE argument here, the cells are colored continuously using a color gradient. The auto legend on the right side of the plot depicts the frequencies using different shades of default colors, where the darkest blue corresponds to the least frequency and the darkest red corresponds to the highest frequency. The OUTLINE option is selected so that all the cells created will have a black outline.

Figure 1 shows a heat map using all the options mentioned above and default assignment of colors.
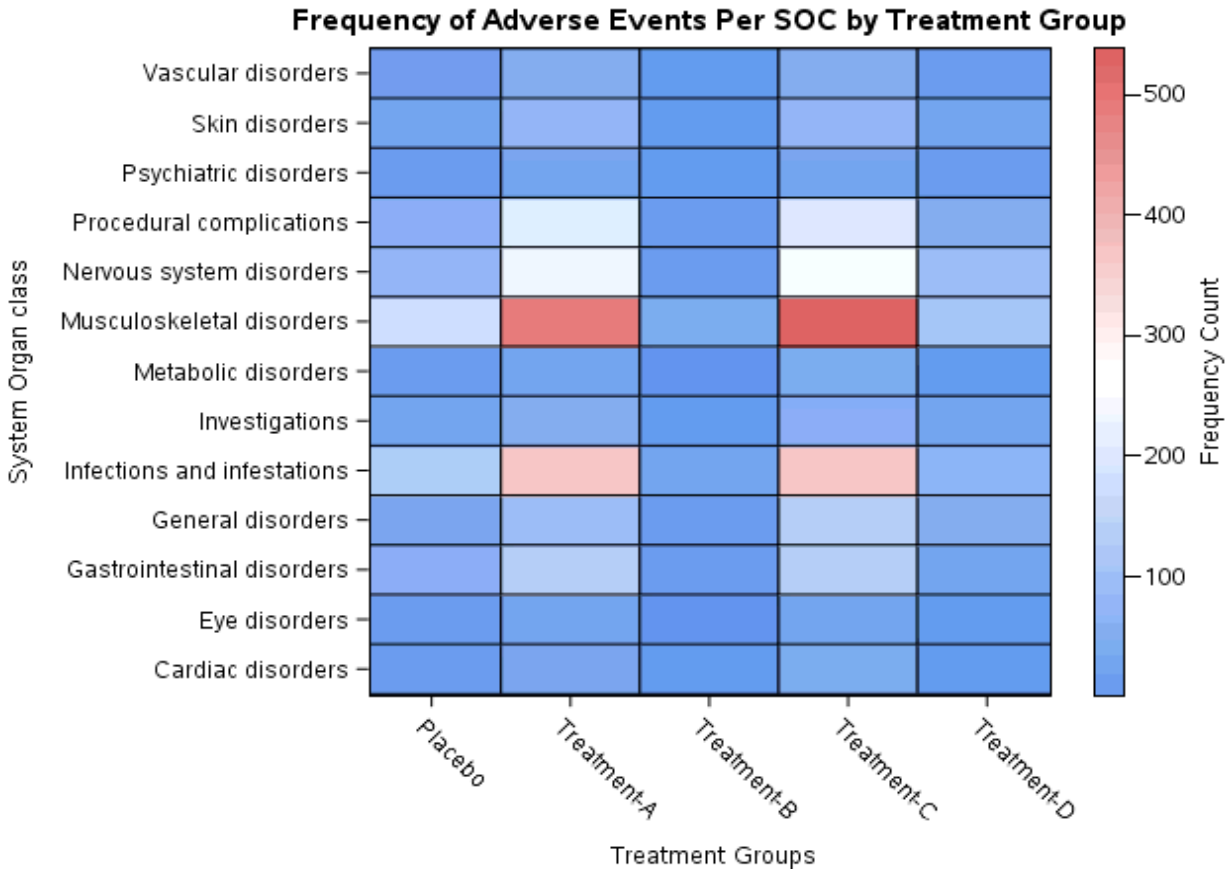

Figure 1. Heat Map with default color assignments and auto legend

The code for the graph is shown below:

```
proc sgplot data=socfreq;
   title "Frequency of Adverse Events Per SOC by Treatment Group";
   heatmapparm x=trt y=aesoc
   colorresponse=count / outline;
   yaxis label="System Organ class";
   xaxis label="Treatment Groups";
run;
```

Assigning Gradient of Colors:

Figure 2: If the user wants to use choice of colors instead, then we can specify the colors using COLORMODEL option as shown below. And instead of auto legend we use GRADLEGEND here, which is the gradient legend that can be used only in conjunction with COLORRESPONSE option. Here the GRADLEGEND displays shades of green, yellow, and red as mentioned in the COLORMODEL= option. This example also shows the assignment of display values for the X-axis tick values using VALUESDISPLAY= option and display of the counts in the cells using the TEXT option.

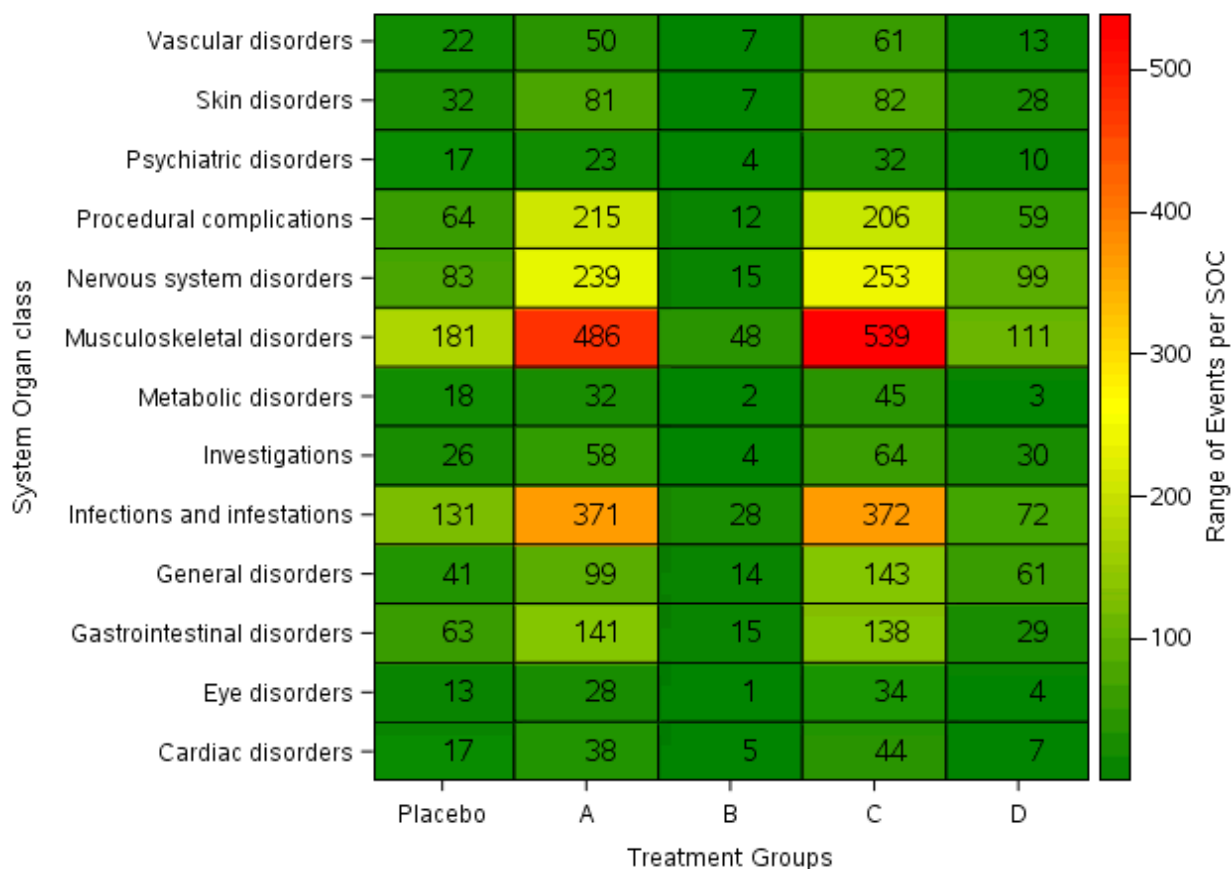The output with gradient of colors is shown in Figure 2.

Figure 2. Heat Map using COLORMODEL and GRADLEGEND options

The program to generate figure 2 is shown below.

```
proc sgplot data=socfreq noautolegend;
  heatmapparm x=trt y=aesoc
   colorresponse=count / colormodel=(green yellow red) outline ;
   text x=trt y=aesoc text=count / textattrs=(size=11pt);
   gradlegend  /title='Range of Events per SOC';
   yaxis label="System Organ class";
   xaxis label="Treatment Groups" values=
     ('Placebo' 'Treatment-A' 'Treatment-B' 'Treatment-C' 'Treatment-D')
     valuesdisplay=('Placebo' 'A' 'B' 'C' 'D');
 run;
```

DATACOLORS and KEYLEGEND Options:

Figure 3 illustrates how COLORGROUP argument varies from COLORRESPONSE. When COLORGROUP option is used, the cells are colored discretely and COLORGROUP can be used for numeric and character variables unlike COLORRESPONSE which can be used only for numeric variables. Another way to assign specific colors using STYLEATTRS DATACOLORS option is shown below. The variable COLR is derived to set the range, and this is used in COLORGROUP= option. Since the GRADLEGEND is not supported by COLORGROUP option, KEYLEGEND statement can be used to display the range values instead.

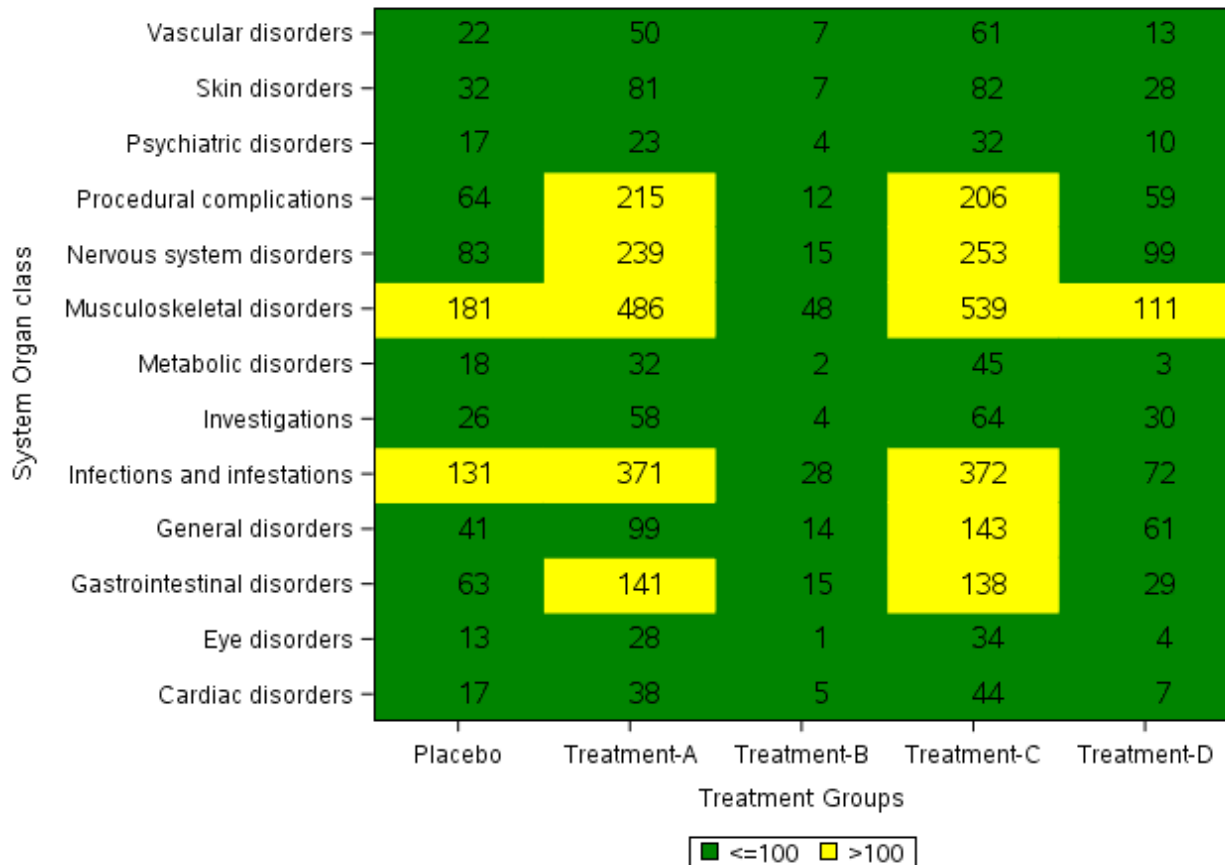 Let us see the effect of COLORGROUP related arguments in Figure 3.

Figure 3. Heat Map using DATACOLORS and KEYLEGEND options

Figure 3 does not have the cells outlined since the OUTLINE option is not used. The code to generate the figure is shown below.

```
data socfreq;
 set socfreq;
   if count <= 100 then colr='<=100';
   else if count > 100 then colr='>100';
 run;

 proc sgplot data=socfreq noautolegend;
    styleattrs datacolors =(green yellow );
    heatmapparm x=trt y=aesoc
    colorgroup=colr / name='range';
    text x=trt y=aesoc text=count / textattrs=(size=11pt);
    yaxis label="System Organ class";
    xaxis label="Treatment Groups";
    keylegend 'range';
 run;
```

RATTRID and FITPOLICY Options:

Figure 4: Assigning specific colors for a specific range of values to control the visual attributes of the graph can be achieved using RATTRID option. This option specifies the value of the ID variable in a range attribute map dataset. The range of values and the color for each range can be defined in the attribute map dataset and this dataset is used as RATTRMAP value. This example also shows the use of FITPOLICY option, which splits the tick values of the axis when there is not much room to fit in one line.

The program to generate figure 4 is shown below.

```
/* Range Attribute Map Dataset */
data attrmap;
  retain id "myid";
  length min 4. max $5.;
   input min max $ color $;
   datalines;
   0  25 green
   26 50 yellow
   51 150 orange
   151 _max_ red
   ;
run;


 proc sgplot data=socfreq rattrmap=attrmap noautolegend;
    heatmapparm x=trt y=aesoc
    colorresponse=count/ rattrid=myid outline name='range';
    text x=trt y=aesoc text=count / textattrs=(size=8pt);
    gradlegend 'range'/title= 'Range of Events per SOC';
    yaxis fitpolicy=split label="System Organ class";
    xaxis label="Treatment Groups";
 run;
```

The output for this code is shown in Figure 4 in which Y-axis values of System Organ Class have been split into two lines.
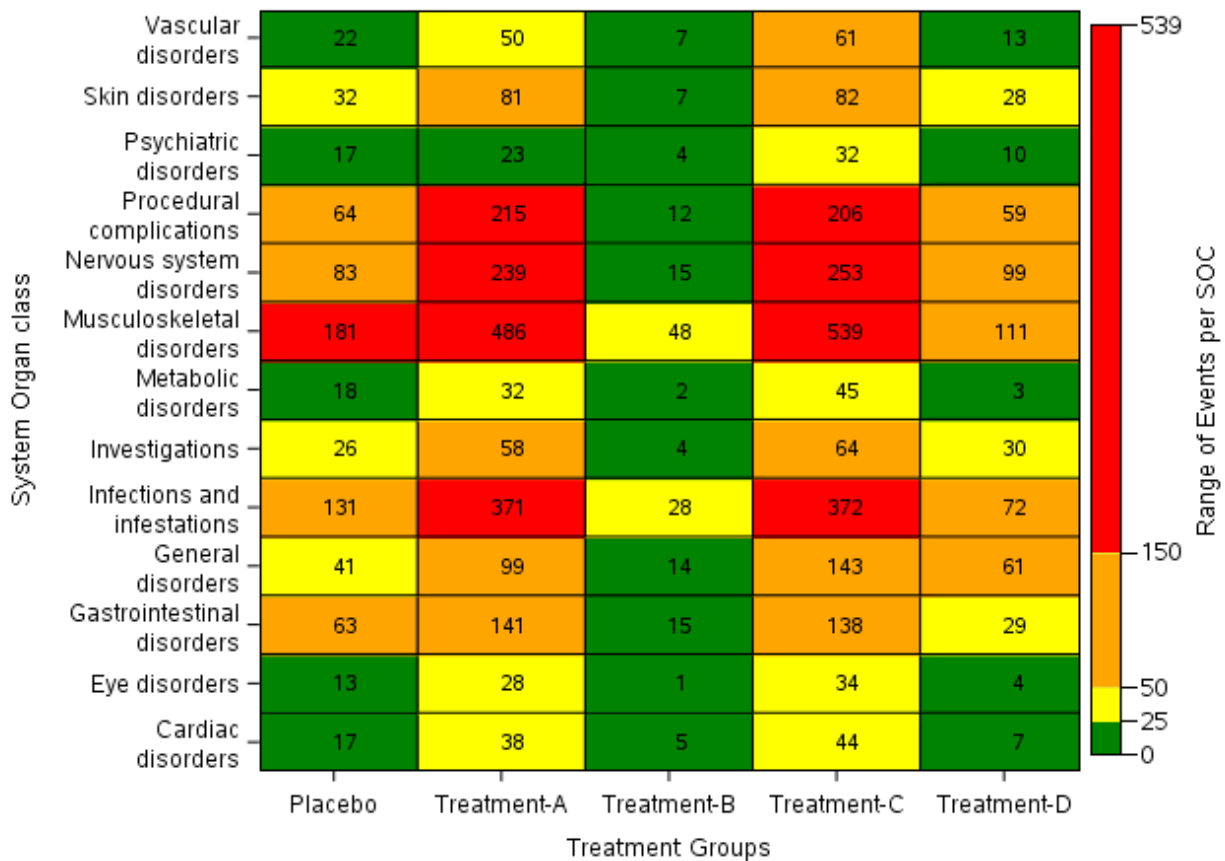
Figure 4. Heat Map using RATTRID and FITPOLICY options

In Figure 1 – Figure 4 we can see how different options help to enhance the figures to better visualize the data as you can see in most of the treatment groups Musculoskeletal disorders are more when compared to others.

PROCEDURE SGPLOT WITH HEATMAP STATEMENT:

Heat map plots can also be generated using Procedure SGPLOT with HEATMAP statement. The main advantage to HEATMAP statement over HEATMAPPARM statement is that we can control continuous numeric response data into desired number of bins and bin sizes. We can also show desired statistic like mean, sum, pct etc. which is calculated over all observations that fall into the bin. We can also overlay other plots like REG, LOESS, ELLIPSE etc. in the same graph. HEATMAP statement requires only X and Y variables unlike HEATMAPPARM where we need either COLORRESPONSE or COLORGROUP as well.

Figure 5 - Figure 7 show HEATMAP statement with options like COLORSTAT, NXBINS, NYBINS, SHOWYBIN, XBINSTART etc. The sample data used for these figures consist of subject id, height, weight, baseline pain score of subjects and treatment group variables.

Figure 5 is a default heat map, plotting the mean of response variable PAIN using X variable WEIGHT and Y variable HEIGHT without any binning options. COLORSTAT option is used to specify the desired statistic, which in here is MEAN of the COLORRESPONSE variable PAIN.

```
proc sgplot data=sample1;
   heatmap x=weight y=height/
   colorresponse=pain colorstat=mean colormodel=(green yellow red);
run;
```
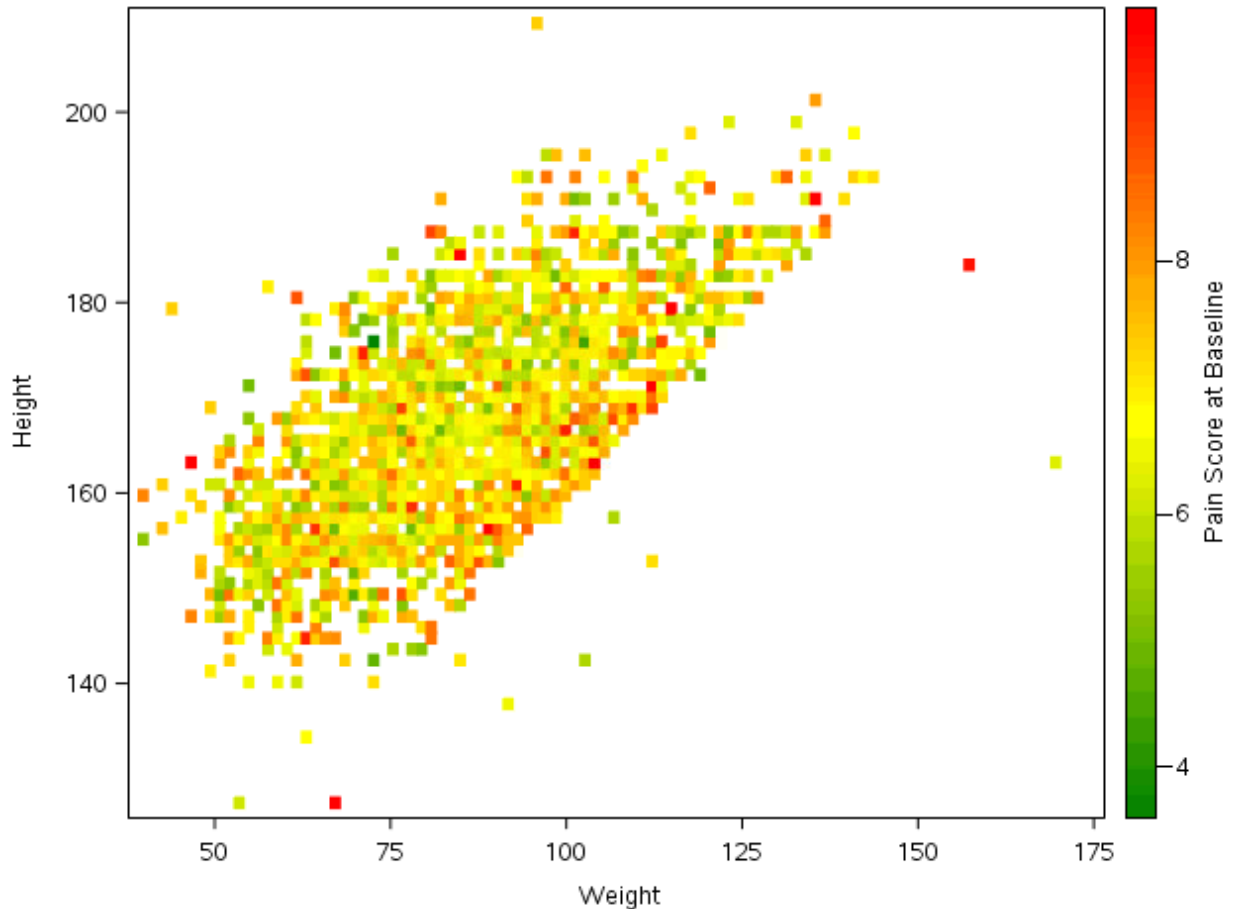
The output for the code is shown below in Figure 5.

Figure 5. Using HEATMAP Statement

HEATMAP Statement with Binning Options:

Figure 6 shows binning options for X and Y axis variables. NXBINS and NYBINS specify the number of bins for X and Y variables respectively. Since SHOWYBINS option is used here, bins will be used as the basis for Y-axis tick marks. YBINSIZE option increases the size of the bin along the Y-axis. In this example you can also see the ELLIPSE statement used to overlay the graph.

SAS sometimes ignores XBINSIZE and YBINSIZE options, to accommodate data when both BINSIZE and NXBINS or NYBINS are specified.

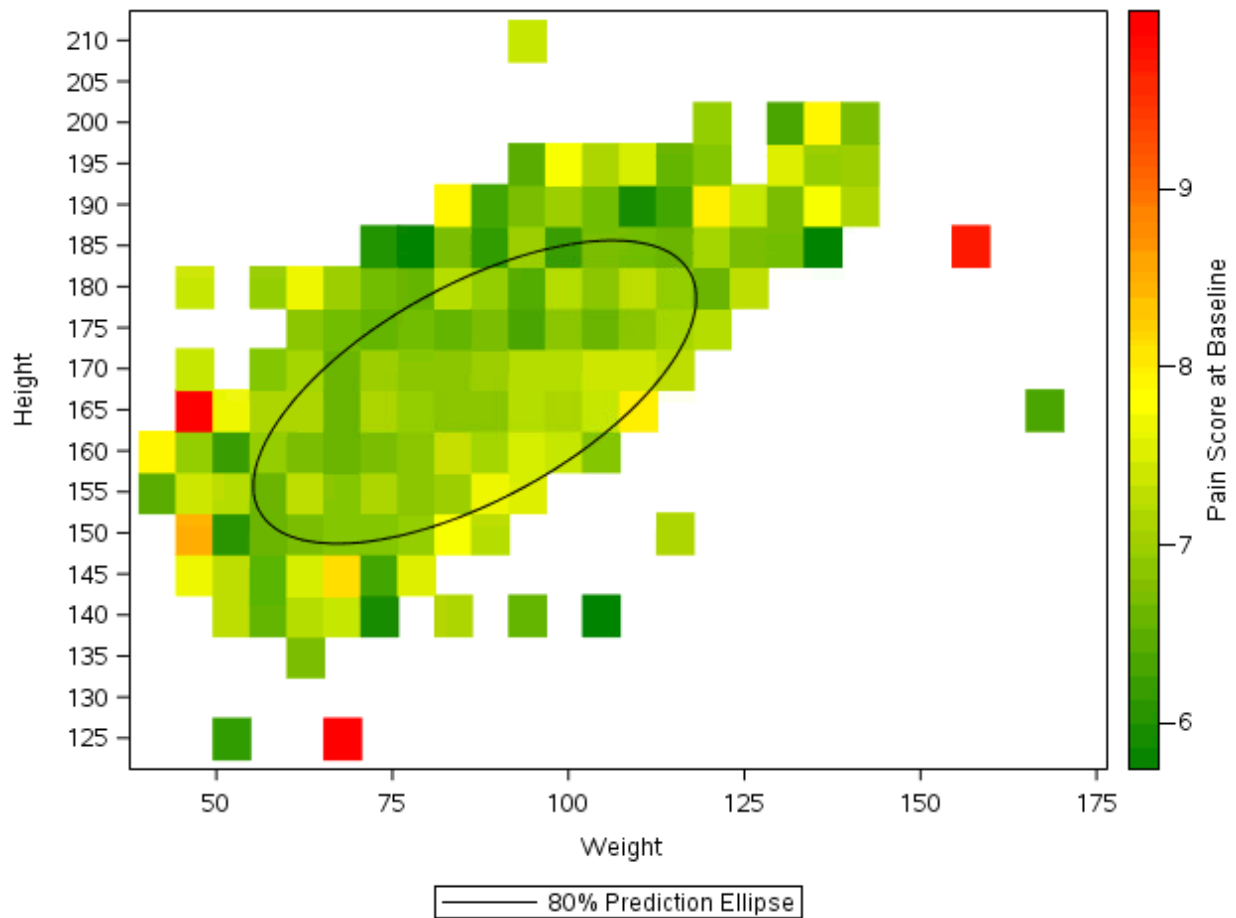The output generated using the binning options is shown in Figure 6.

Figure 6. Using HEATMAP Statement binning options

The code to generate Figure 6 is shown below:

```
proc sgplot data=sample1;
  heatmap x=weight y=height/ nxbins=25 nybins=25 showybins
  colorresponse=pain colorstat=mean colormodel=(green yellow red);
  ellipse X=weight Y=height/alpha=.2;
run;
```

If we use the YBINSIZE= option in the above code, the plot will be drawn as shown below in Figure 7.
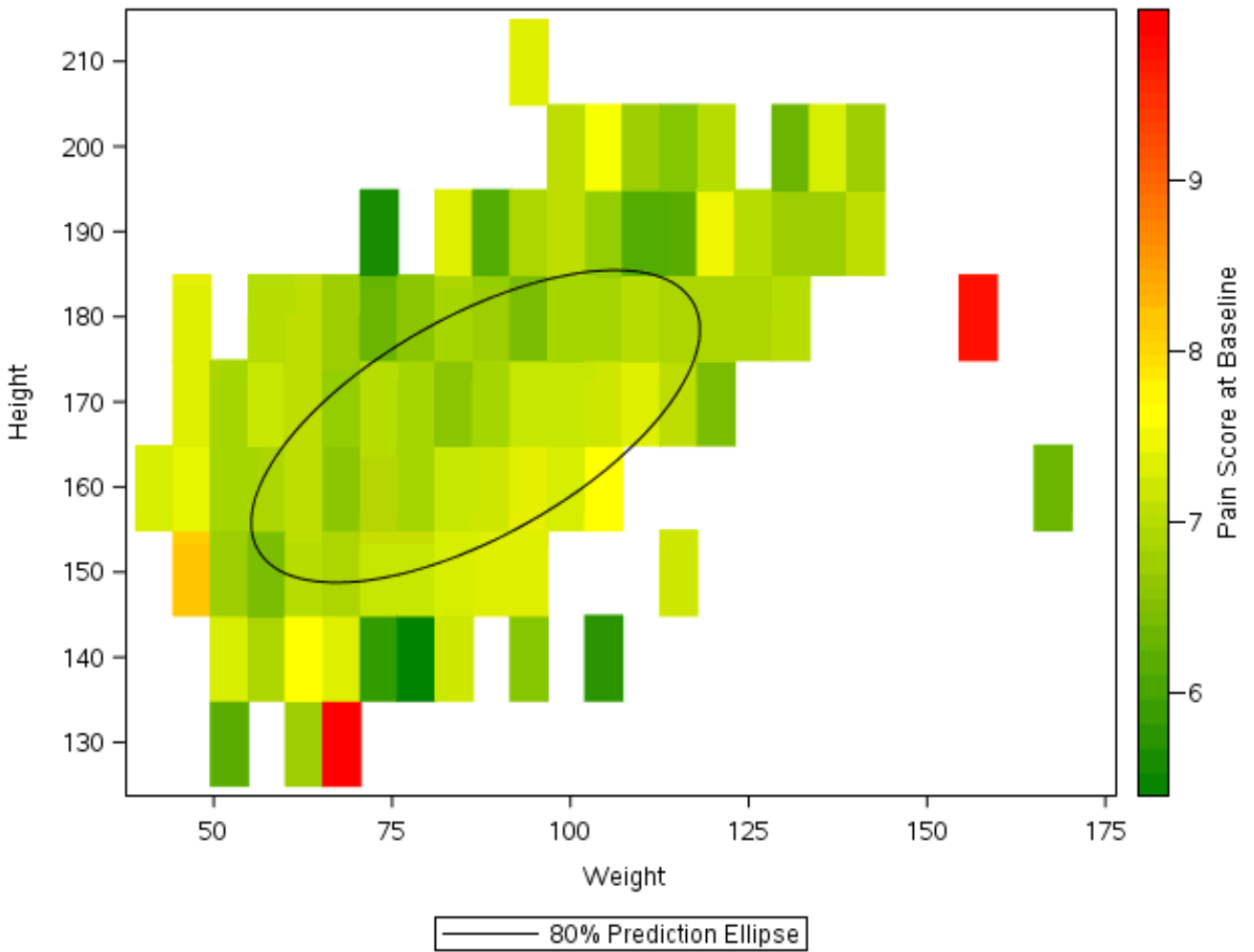
Figure 7. Using YBINSIZE option

CONCLUSION

Heat map plot generated using SAS can be an effective method to represent complex and huge amount of data graphically. SAS® 9.4 procedures like SGPLOT and its related statements like HEATMAPPARM and HEATMAP can be used to create such graphs as illustrated in this paper. This paper demonstrates heat map graph generation with most used options on a sample dataset.

REFERENCES

Prashant Hebbar and Sanjay Matange. 2016. "Graph A Million with the SGPLOT Procedure." *Proceedings of the SAS Global 2016 Conference.* Available at https://support.sas.com/resources/papers/proceedings16/SAS4341-2016.pdf

Rick Wicklin's blog "Create a discrete heat map with PROC SGPLOT". Available at

https://blogs.sas.com/content/iml/2019/07/15/create-discrete-heat-map-sgplot.html

ACKNOWLEDGMENTS

CONTACT INFORMATION

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

Pavani Potuganti
Eliassen Group Biometrics and Data Solutions
PPotuganti@eliassen.com

Sree Nalluri
Eliassen Group Biometrics and Data Solutions
SNalluri@eliassen.com