Making Graphs Using SAS

Sanjay Matange, SAS Institute Inc.

Topics for Today

- Brief history of ODS Graphics.
- Review of some key concepts.
- Sneak preview of new features in SAS 9.40M5.
- Blog articles and Communities conversations.

Prior to SAS 9.2 – Using the SAS/GRAPH Procedures

- The SAS/GRAPH product provided multiple procedures to create graphs.
- Each procedure specialized in the type of graphs it made.
  - GPLOT procedure for Scatter, Series, Needles and more.
  - GCHART procedure for Categorical graphs like Bar and Pie.
  - GMAP for mapping.
  - GBarLine for combination Bar and Line.
  - GCONTOUR, G3D for other specialized cases.
- SAS/GRAPH procedures are popular and continue to be supported.

Prior to SAS 9.2 – Using the SAS/GRAPH Procedures

- Separate procedures meant plots could not be combined.
  - To create a Bar-Line graph, we needed a new procedure.
- Output was created in SAS internal GRSEG format.
- Combining graphs together in a layout needs use of GREPLAY.
  - This sometimes had issues with aspect ratios of individual graphs.
- Annotation was used extensively to customize the graph.
  - Annotation is an external process, so such programs do not scale well to different data.

Prior to SAS 9.2 – Using the SAS/GRAPH Procedures

- Most analytical procedures did not produce graphs automatically.
- For such procedures, to visualize results one had to run the procedure, save the data, and then use a SAS/GRAPH procedure to create the graph.
- This means that every user had to become proficient with graph syntax.
- This also meant that every user created their own graphs, so there was no consistency.
- Transient data generated while the analytical proc is running was lost.

Starting with SAS 9.2 – Using ODS Graphics

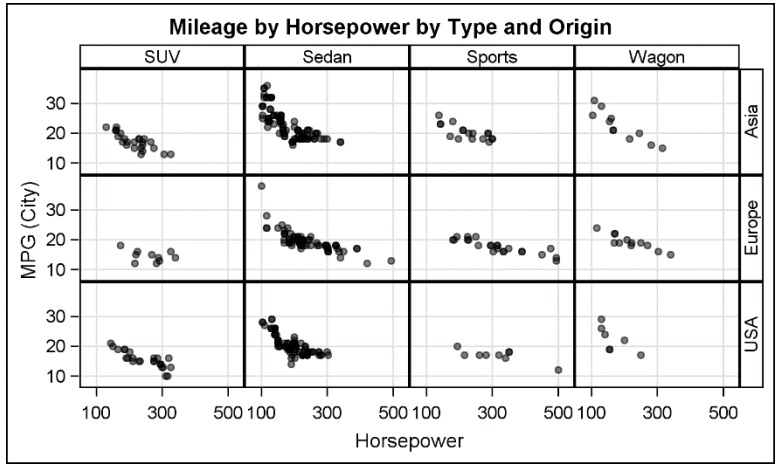
- SAS analytical procedures create graphs automatically.
  - This ensures all users get consistent graphs, including transient information.
  - No need for analysts to learn graph syntax to create the graphs.
  - Procedures use graph templates and GTL to create the graphs.
- GTL is the underlying engine for creating all graphs.
- SAS users can also use GTL to create their custom graphs.
- Simplified “80-20” wrappers are provided on GTL feature set.
  - SG Procedures provide a syntax wrapper on GTL to create graphs.
  - SG Designer provides a GUI wrapper on GTL to create graphs.

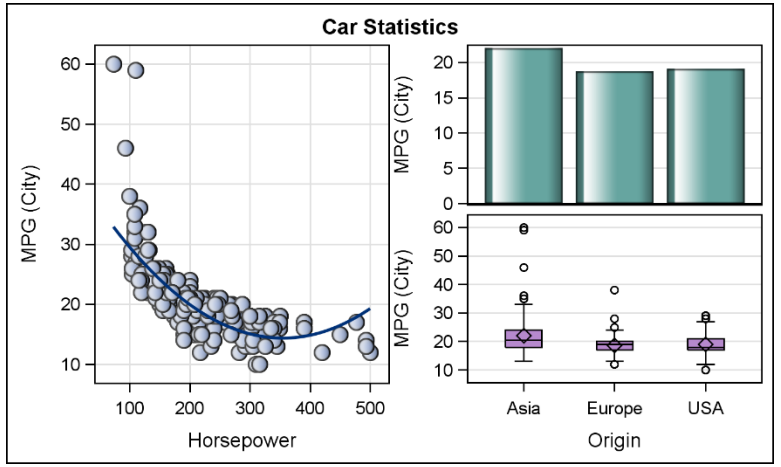
ODS Graphics v/s SAS GRAPH

- ODS Graphics includes automatic graphs from analytical procedures, GTL, SG Procedures and Designer.
- ODS Graphics works differently from SAS/GRAPH procedures.
- ODS Graphics output goes to open ODS destinations like HTML and PDF.
- ODS Graphics produces output in industry standard formats like PNG, PDF, SVG and many more.
- ODS Graphics produces high resolution graphs. Setting high DPI scales all aspects of the graph, including marker size and line thickness and pattern.
- ODS Graphics does NOT honor any GOPTIONS.

Key Features of ODS Graphics

- GTL is the underlying engine for rendering all graphs.
  - This ensures consistency between output from all ODS Graphics sources.
- GTL uses the concept of LAYOUTS and PLOTS to define graphs.
  - LAYOUTS decide where a plot is displayed.
  - PLOTS decide how the data is to be displayed



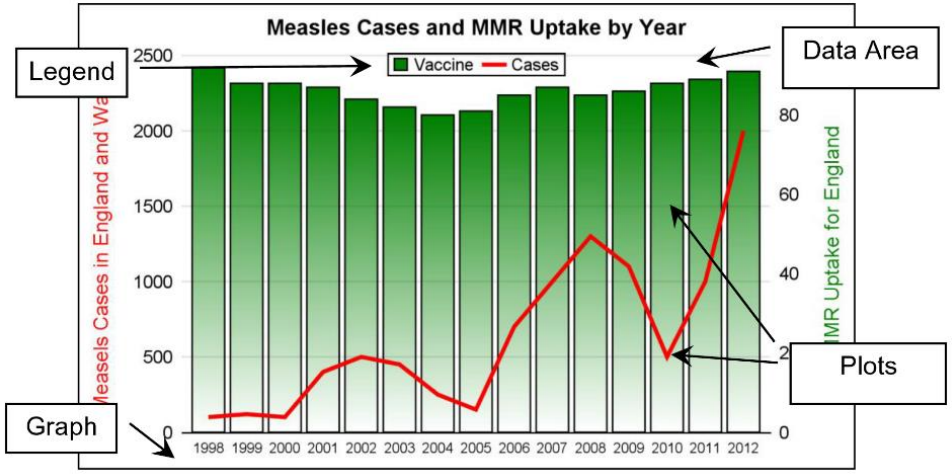


SG Procedures

- SG Procedures include
  - SGPLOT Procedure for single-cell graphs.
  - SGPANEL Procedure for Classification Panels
  - SGSCATTER Procedure for comparative scatter plots and Matrix.
- Two other SG Procedures are:
  - SGRENDER Procedure to render a graph using GTL template.
  - SGDESIGN Procedure to render a Designer built (SGD) graph to ODS destination.

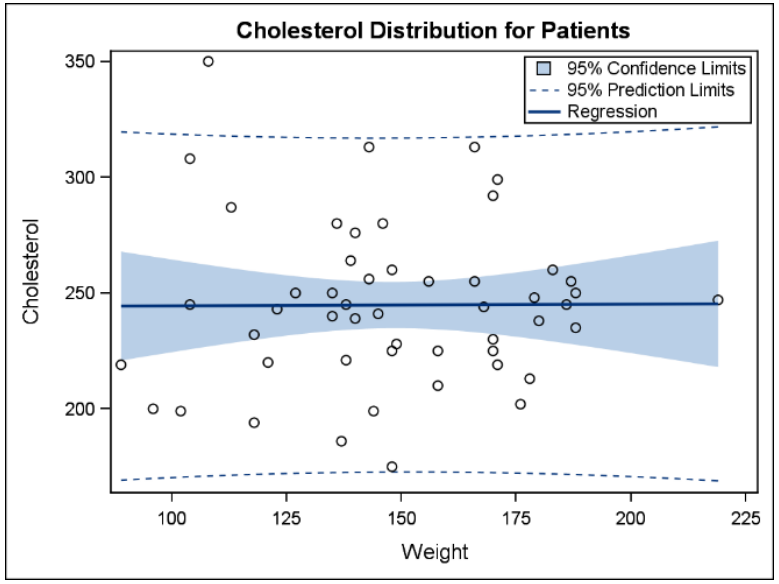
SPLOT Procedure Concepts

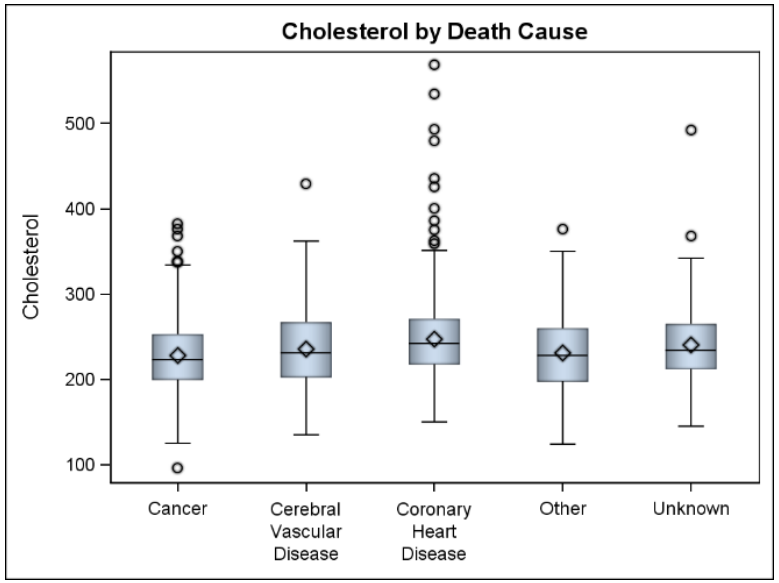
- SGPLOT procedure creates “Single-Cell” graphs.
- These graphs have one main data area, with multiple overlaid plots.
- The data area can have upto 4 axes, X (bottom), X2(top), Y(left), Y2(right).
- Any plot can be associated with one of the X and one of the Y axes.
- You can add multiple legends and insets.



Single Plot Statement Graphs

- Many graphs can be built using just one plot statement.
- A Regression plot uses one REG statement.
  - The display includes the observations, the fit line and the confidence bands.
- A Box plot uses one VBOX or HBOX statement.
  - The display includes the Q1-Q3 box, the Mean, the Median values, the Whiskers and the outliers.

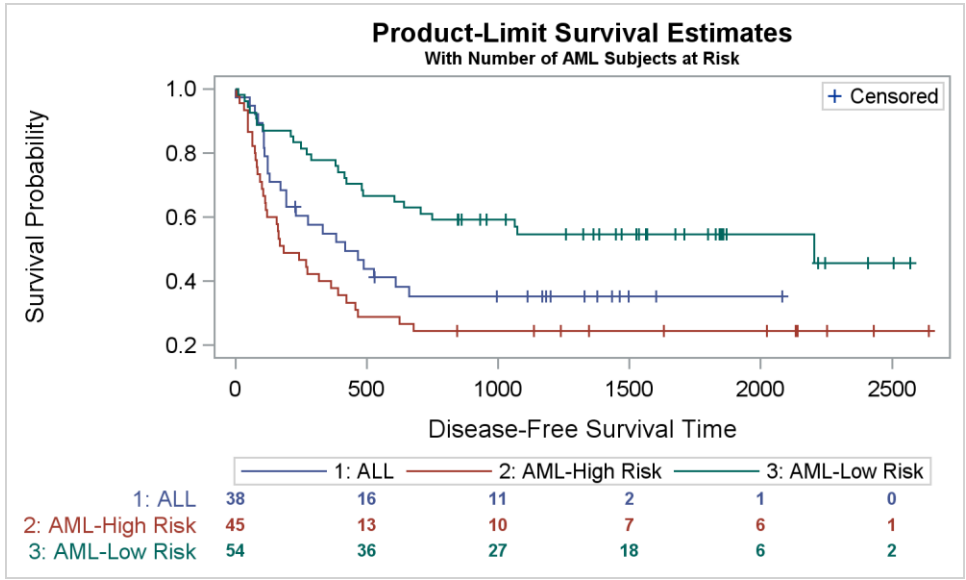




Building Graphs in Layers

- Complex graphs can be built by combining plots as components.

```
proc sgplot data=SurvivalPlotData;
  step x=time y=survival / group=stratum name='s';
  scatter x=time y=censored /
          markerattrs=(symbol=plus) name='c';
  scatter x=time y=censored / group=stratum
          markerattrs=(symbol=plus);
  xaxistable atrisk / x=tatrisk
          class=stratum
          colorgroup=stratum;
  keylegend 'c' / location=inside
          position=topright;
  keylegend 's';
run;
```



Colors and Other Attributes for Classification

- Discrete colors, symbols or line patterns for classification are assigned from the GraphData1-12 Style Elements.
- These are assigned in sequence based on the order in which the values are encountered by the software.
- Prior to SAS 9.4, changing classification attribute required change of the values defined in the active Style.
- Starting with SAS 9.4, these can be changed in procedure or GTL code.

Attribute Priority

- Normally, all attributes are rotated together through the (12) colors, (7) Symbols and (11) line patterns.
- This is called ATTRPRIORITY=NONE. Most Styles follow this process and is set in the Style.
- One can also have ATTRPRIORITY=COLOR. In this case, only colors are rotated using 1st Symbols and 1st Line Pattern constant till all 12 colors are used. Then, we shift to 2nd Symbol and 2nd Line Pattern, and rotate through all the 12 colors. And so on.
- HTMLBlue Style (HTML destination) uses ATTRPRIORITY=COLOR.
- You can change this behavior in ODS GRAPHICS statement or in GTL code.

Discrete Attributes Map

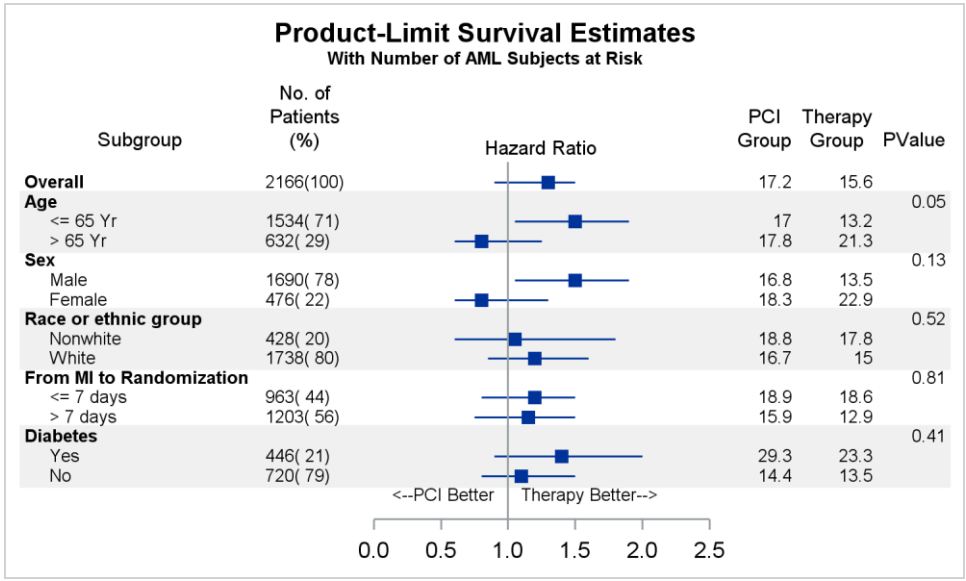
- Classification attributes are assigned based on the order the values are encountered in the data.
- First value gets the first GraphData Element, and so on.
- If the order of the values change, the assignment changes.
  - So, Drug A could be Blue today and Red tomorrow.
- The best way to ensure that specific classification VALUES get specific attributes, use Discrete Attributes Map.
- This is like a format, and defines by value the attributes to be used.
  - So, we can say “IBM” should be blue, and “John Deere” should be green.
- You can also ask all the values from map to be displayed in legend.

Range Attributes Map

- We can use ColorResponse role with many plots like Scatter, or HeatMap.
- Normally, the default 3-Color ramp will be used.
  - The lowest value in data will get 1st color.
  - The highest value in the data will get the 3rd color.
  - Other values will get a color that is interpolated between the 3 colors.
- This means sometimes 80 degrees will be red and sometimes 100 will be red. Two graphs side by side could have different color mappings.
- To ensure consistent color mapping, use Range Attribute Map.
  - In this, you can define by VALUE the color to be assigned.
  - So, you can always get consistent color mapping.

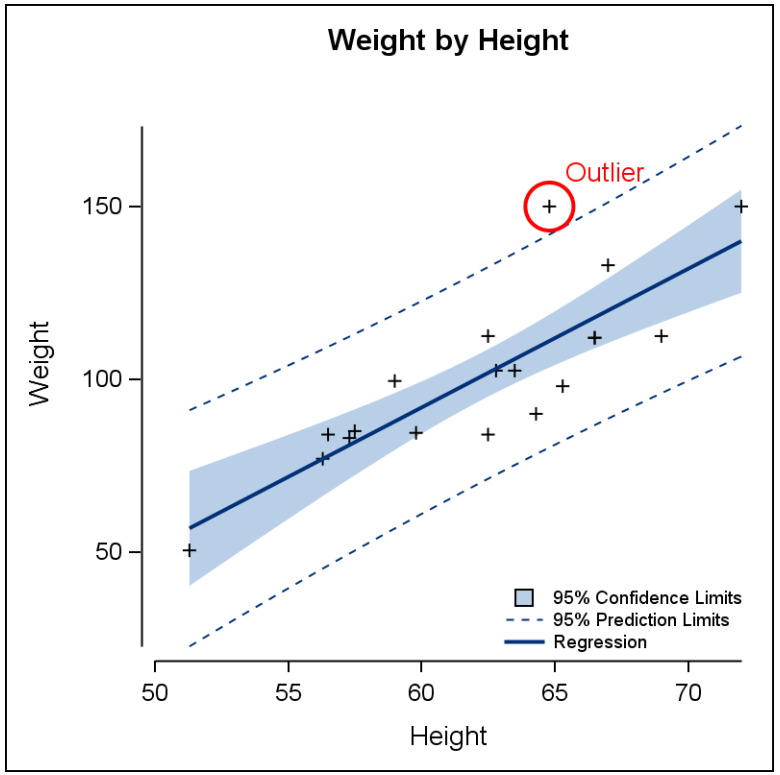
Using Plot Layers v/s Annotation

- Complex graphs can be built by combining plots as components.
- Most customizations that are within the data area can be made using plot layers.
- Plot layers work with axes to reserve the space needed automatically.
- AxisTable components can be used to add textual data inside or outside data area. Appropriate space is reserved automatically.
- The same program can scale well for different data.



Using Annotation

- Annotation is best for markup of a graph after it is created as shown in the graph below.
- Annotation is most useful when the markup has to span from inside data space to outside.
- One drawback of Annotation is the axes are not aware of it.
- Annotation does not scale well to different data. If values become much larger in a table displayed using annotation, it has to be adjusted by hand.



New Features for SAS 9.40M5

- SAS 9.40M5 will be the 6th release of SAS 9.4.
- We have continued to add significant new features in these releases.
- As some of you have commented on the communities page, it is important to know the release number of SAS as ODS Graphics keeps improving.
- You will see many new features based on user requests including:
  - Fill Patterns for all plot types including box, highlow, ellipse, band, polygon.
  - Reverse order of legend entries. Useful for stacked bar charts.
  - Data label options for Hbar.
  - Using Reflines for custom banding.
  - Text plot position by column.

Blog Topics

Most blog article topics are motivated from the following:

- New features released that address some long standing issues.
- Discussion on the Communities page that leads to a solution for a specific user that could also be of use to other users.
- Discussion with co-workers that lead to some interesting graphs like the recent “Stem and Leaf” plot.
- Redoing some graph done using R with the R-like look and feel.
- “Getting Started” topics for users who are new to this subject.
- Which ones do you like?

User Interaction and Feedback is Gratifying

- In early days of ODS Graphics, we got a lot of questions from users on “Can you do this”.
- Such questions led us to find ways to create the requested graph, often with much coding. It also led us to include new features to make the same easier.
- Now, we are getting fewer such questions. This is good and bad.
  - Good if it means the software has reached some maturity.
  - Bad as we don’t get more ideas for enhancing the software.
  - So, keep the questions coming.
- Comments, good or bad on the blog articles also help.

Participating on the Communities page.

- Both Dan Heath and I monitor the Communities page on SAS/GRAPH and ODS Graphics.
- It is really nice to see many SAS users (you guys) answering the questions on the communities page.
- It is preferable if the answer came from another SAS user as it gives the person asking more confidence in the solution.
- If no answer is forthcoming, or, if there is another way to do the same, we will chime in.
- It is always challenging to encourage users to provide some code they have tried, along with sample data and SAS release version.

Thank you for attending