

SAS Macros for S-PLUS Clinical Graphics within Production Reporting Environment

Sven J. Knudsen, Insightful Limited, Basingstoke, UK
Charlie Roosen, Insightful Limited, Basingstoke, UK

ABSTRACT

SAS® has long been recognized as a leading application for data processing, and *S-PLUS™* as a leading application for statistical graphics. A common approach is to use SAS for data processing, tables and listings; and S-PLUS for graphs from the transformed SAS data sets. The new *S+Connector for SAS®* gives a powerful environment for integrating S-PLUS graph creation into a SAS analysis with no need for S-PLUS programming. This paper demonstrates how the SAS programmer can use the S+Connector to easily design graph templates in the new *Insightful Clinical Graphics* point-and-click application, then generate graphs from the templates using SAS macros. We show how graph templates are created from a built-in pallet that comprises a broad set of validated clinical displays, how these pre-validated templates can be used in SAS, and that a simple set of SAS macros governs good practice and standards. We discuss CDISC data support, the validation process of graph templates and macros, and inclusion of graph validation metadata and the S-PLUS verbose execution log in the SAS log.

INTRODUCTION

S-PLUS® has long been renowned for its powerful environment for creating graphs that maximizes our ability for visual decoding. S-PLUS is designed for statistical visualization and supports a wide range of modern graph types with features that adhere to key graphics principles such as those originally espoused by thought leaders Tukey (1977), Tufte (1983) and Cleveland (1993). This has been achieved by using Trellis™ graphics and more recently the S-PLUS Graphic Object Model (GOM). In contrast to many graphics engines, S-PLUS offers scripting, cross platform support, and batch mode execution with verbose log file. This makes S-PLUS widely used for graphics production in pharmaceutical companies.

The Insightful Clinical Graphics (iCG) solution enables point-and-click creation of advanced S-PLUS graphics without programming while generating the code behind. Prior to the introduction of iCG with its central graph library, graphs have been re-created, requiring more time and expense in the quality assurance steps. With iCG, statisticians and programmers can create graphs as templates for reuse on bound or new data. The template may be saved for downstream use in the existing (validated) production reporting systems; as such, these jobs and resulting graphs need little, if any, further validation.

To facilitate such reporting in companies who have a validated SAS production environment, iCG can be integrated using the *S+Connector for SAS*. In production, SAS will call S-PLUS Server with requests to create graphs based on iCG graph templates. There are three sets of macros:

1. The S+Connector for SAS set of macros
2. A set of helper macros for graph and reference elements settings
3. A set of graph templates macros

The iCG SAS macros are aimed for programmers who need to create SAS programs for simultaneous data listing, tabular and graphical reporting. The general set of S+Connector for SAS macros (1.) gives the programmer full control of configuration, graph, verbose log, and script integration. General graph attributes such as titles, axis settings and reference lines are handled via helper macros (2.). A template macro is then used to set graph type specific parameters (3), repurposing the graph template with bound or unbound defaults on new data.

CREATING GRAPH TEMPLATES

Graphs are added to iCG by creating a graph template. This is pretty straightforward using the point-and-click iCG client (Figure 1). Templates can be based on the existing iCG graph library and saved off to the same place. The iCG templates include metadata that assist downstream searching and validation status and therefore governs standard operating procedures in place in pharmaceutical companies.

PhUSE 2008

Once the template has been saved, the SAS programmer has to integrate and perform a series of quality assurance tests.

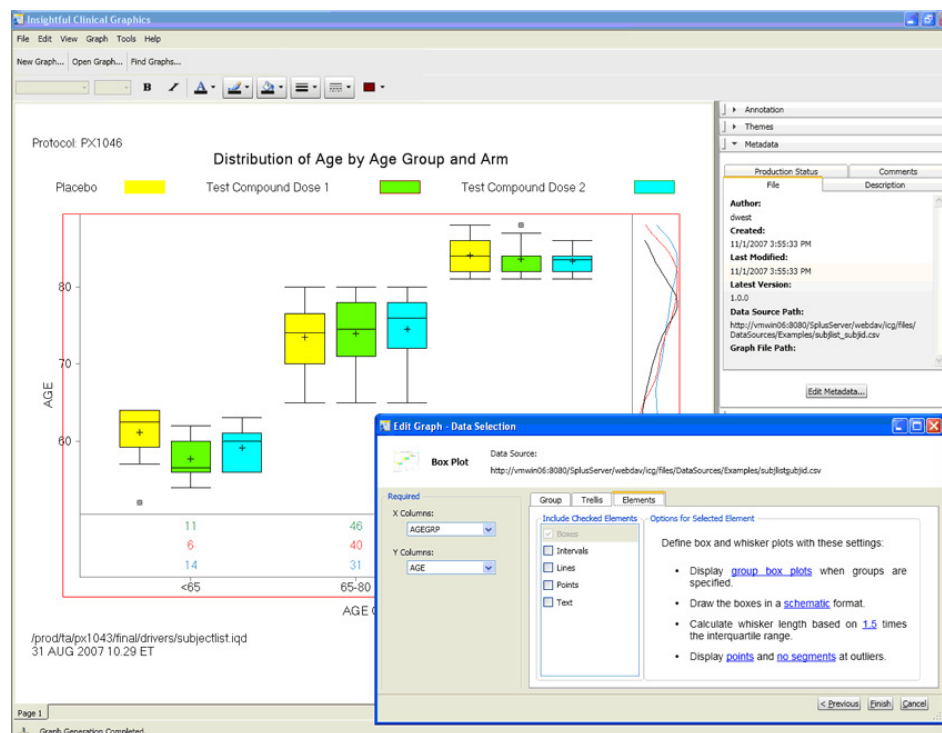


Figure 1: A screenshot of the iCG point-and-click client.

INTEGRATION

The iCG SAS macro centers on a few important macros:

- **icg_Configuration**: defines global SAS macro variables required by the iCG macro library, usually located in the AUTOEXEC
- **icg_CreateGraph**: executes a graph document, controls output location, and log and script integration
- **icg_InsertParameter**: a general parameters handler for the iCG template; for example, the data set name, variable names or details like the percentiles used in a box-whiskers plot say.

The **icg_InsertParameter** is a generic macro that can control template parameters, and works on a name-value basis. The set of available template parameters is available as a SAS data set, making it easy to investigate and identify the parameter we need to set. However, some parameters are used more often than others across all graphs. The set of helper macros provide the ability to specify such general properties:

- **icg_Page**, **icg_Theme**: page format, margin settings, and theme to be used (color or black-and-white)
- **icg_Title**: headers, footers, titles and sub-titles
- **icg_GraphTable**: plot aspect ratio, number of columns and rows for Trellis layout
- **icg_Axes**: axes settings such as labels, range and tick marks, scale type such as symmetric, linear or log
- **icg_Legend**: legend properties such as location, title, and visibility
- **icg_ReferenceXLines**, **icg_ReferenceYLines**: horizontal or vertical reference lines with text labels
- **icg_ReferenceCurve**: reference curves such as slope lines or fitted curves
- **icg_ReferenceArrow**, **icg_ReferenceText**: arrows and text annotation

The helper macros are all based on **icg_InsertParameter** and can be extended further.

The final set of macros, the template macros, is also using **icg_InsertParameter** as the central handler. In the majority of the cases it is sufficient to specify data set and variable names required creating the graph, and the template macro is no longer than a few lines. The template will include the specification of the graph type in use. However, if required, **icg_InsertParameter** can be used to change the specification; for example, change the symbol type used in a bubble plot or turn confidence intervals in a survival plot on or off. This flexibility is handy for ad-hoc reporting and exploratory usage, and can, obviously, also be used in the production environment.

VALIDATION AND TESTING

With iCG templates the graph is defined once, enabling others to re-use the template across their functional areas. Further, iCG templates created on existing pre validated templates require no extra validation and testing. As such, production template macros and resulting graphs need no further validation and testing.

When creating a SAS template macro we need to test that the parameters are passed on properly. To aid this process, the runtime S-PLUS script and verbose log are both included in the SAS log file for inspection. The log file will also hold template meta-data and its validation status.

On a more general system development level, for example when creating helper macros, we need to perform unit and regression tests. These can be automated using an image processing tool. A good (and free) suite can be found in ImageMagick® (<http://www.imagemagick.org/>). This tool can compare two images and provide a metric that summarizes the binary difference, and also produces the difference as an image. Any change can quickly be detected by putting tests on the produced metric and a difference can be captured for visual decoding of potential bugs. ImageMagick can read over 100 formats, including WMF, PDF, PNG, JPG and PS.

Once installed, ImageMagick® is quite easy to use. Assume we have two graph files gr.png and grBaseline.png. The following SAS X command compare the two images and store the metric in a output text file calle out.txt:

```
x 'command/c compare -metric RMSE gr.png grBaseline.png null: 2>out.txt';
```

The out file out.txt a single line where the first number is zero if there if there is no difference. The following SAS X command produce the difference between the two images called diff.png; see Figure 2 below:

```
x 'command/c convert gr.png grBaseline.png -compose difference -composite diff.png';
```

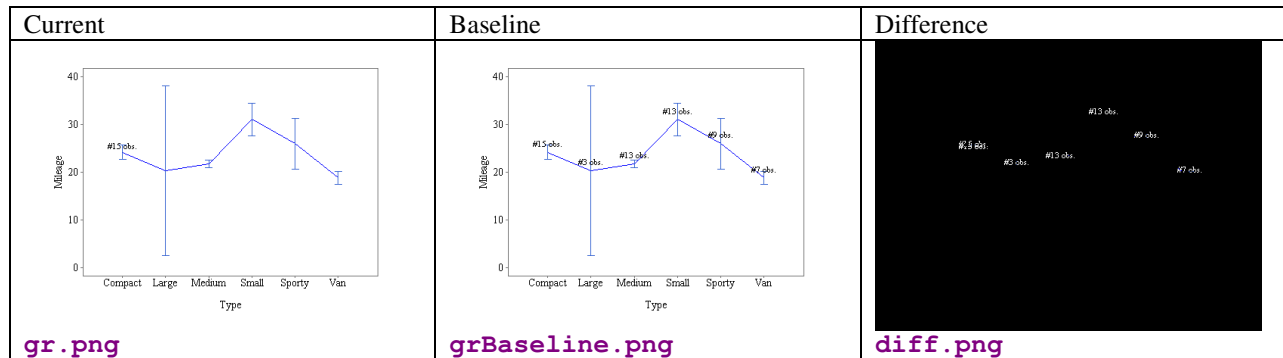


Figure 2 Example of a current graph being compared to a baseline or gold graph. The last column shows the difference where we can clearly see white text labels. This quickly draws our attention to a potential bug (in this case it was)

SUMMARY

The benefits of iCG are easily identified through its client server architecture, providing a well defined workflow for creation, re-purposing and enforcement of best-practice statistical graphics for clinical data analysis. The workflow reduces, if not eliminates, the manual QA process through pre-validated graphs.

Production graphics jobs are supported either directly through the server side S-PLUS engine or through a server-side SAS engine. Programmers can use a simple set of configuration, helper and template macros to create SAS programs for simultaneous data listing, tabular and graphical reporting.

Over the last 20 years of working with pharmaceutical companies Insightful has distilled a graphics solution to address the needs of clinical groups including:

- Creation of best-practice, statistical graphics for clinical data analysis (statistics, programming)
 - Create informative and custom statistical graphics (point-click) from a comprehensive palette of graph types: dot plot, box plot, line plot, etc.
 - Add new statistical graphics for use throughout end-user community
- Repurpose graphs across studies and functional areas (clinical, programming, statistics)
 - Create and save a customized graph as a template for reuse on bound or new data
 - Repurpose graphs in instream clinical data review and study reporting

PhUSE 2008

- Repurpose graphs in analysis of different trial/endpoint by different person
 - Data standards (CDISC) are layered into the graph file
- Reuse graphs in exploratory and validated areas (statistics, programming)
 - Automated creation of script to run in validated (typically UNIX) environment
 - Oracle / SAS and submission reporting
- Point-and-Click graph styling and customization (with bound data) for presentation and publication (publishing)
 - Style a graph into corporate power point templates or scientific journal style
 - Save new styles as themes for reuse
 - No change to scientific graph content

REFERENCES

Cleveland, W. (1993). *Visualizing Data*. Hobart Press, Summit NJ.

O'Connell, M. (2007). Graphical analysis of clinical safety and efficacy data. DIA Presentation. http://www.insightful.com/news_events/2007dia.

Tufte, E. R. (1983). *The Visual Display of Information*, Graphics Press, Cheshire CT.

Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley, Reading

CONTACT INFORMATION

Your comments and questions are valued and encouraged. The contact author can be reached as follows:

Sven J. Knudsen
Insightful UK
5th Floor Network House
Basing View
Basingstoke, Hampshire
RG21 4HG
Tel: +44 (0) 1256 339800
Fax: +44 (0) 1256 339839
E-mail: sknudsen@insightful.com
Web: www.insightful.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries.

S-PLUS and all other Insightful Corp. product or service names are registered trademarks or trademarks of Insightful Corporation in the USA and other countries.

® Indicates USA registration.

Other brand and product names are trademarks of their respective companies.