INTRODUCTION

This graphics scaling algorithm was developed to solve a problem in the graphics display module of a large, automated pharmaceutical data analysis system (TOXSAS). Users of TOXSAS were not pleased with the default graphics axes being produced by the GPLOT and PLOT procedures. The users complained that the axes were cluttered, the end points not "nice", and the graphs unpredictable. In some cases, the users disliked the graphs enough to re-enter the data on a PC and use a PC based graphics package to produce the output, an obviously inefficient remedy for a supposedly automated system.

When enhancements to TOXSAS were planned, a request was made to improve the graphical presentation of the data and to use something other than the default axis values produced by the GPLOT and PLOT procedures. Since a very large amount of data was processed and this was to be an automated system, it was not practical to process the data, have the user examine the data, determine the appropriate graphing parameters, and then execute the graphics module. Although the users might not know what specific data was being extracted, they knew how they wanted to see it: well displayed data, usually in the middle 70-80% of the graph with "nice" end points. The concept of "nice" end points is a subjective one, intuitively obvious to the user but more difficult to quantify for a computer program. It generally meant, for example, end points at 0 and 20 when the actual data values ranged from 3 to 18 or 700 and 1100 when actual data values ranged from 750 to 1000.

Several examples of the default presentation compared to the graphs produced using the algorithm are displayed at the end of this paper.

ENVIRONMENT

The data analysis and output modules - data listings, tables, and graphics - use base SAS® and SAS/GRAPH® software (Version 5.18), and run under the VAX/VMS® operating system. A user interface allows the desired laboratory measurements, statistical analyses and output formats to be specified; the data is then extracted and processed in batch mode. Although developed under Version 5.18 for the VAX operating system, the algorithm described in this presentation has been enhanced slightly to take advantage of more sophisticated macro features unavailable in that version. The code presented here was developed under Version 6.03 and should be transportable to almost any system.

CONSTRAINTS

The users specified the following criteria for the new graphics system:
- Axes should not be cluttered
- End points should be "nice" values
- Data should fill 70% to 80% of the graph's display area
- A graph should display 4 or 5 major tick marks
- A graph should also display appropriate number of minor ticks

Achieving these objectives presented several challenges:
- The wide variety of data being processed meant that no assumptions could be made on the types of data to be graphed.
- The differing ranges of data made determining the 70-80% spread difficult.

- Rounding the end-point values had to be based on the magnitude of the range, the value of the numbers being rounded, and whether the minimum or maximum end point was being determined.
- Rounding had to maintain the 70-80% spread as much as possible.
- The final end points had to allow for evenly spaced major tick marks.

APPROACH

The approach was to develop an algorithm that would calculate the appropriate end points, find the value of the step (which then determined the major tick marks), identify the appropriate number of minor tick marks, and provide this information automatically to the graphics procedure. Although the macro facility under VMS was limited, it did allow the creation of macro variables which were then passed for use in the graphics module of TOXSAS.

The algorithm was structured as follows:

- `PROC SUMMARY`
  to find observed minimum and maximum values
- `DATA STEP`
  calculate range of observed minimum and maximum
  - add 5% to the maximum value
  - subtract 5% from the minimum value
    This established rough end points at 95%.
  - to gauge and adjust for the magnitude of the range: calculate LOGFACT and POWER variables
    This step was critical because it reduced the impact of the magnitude of the range which caused problems with the CEIL and FLOOR functions.
  - calculate end points
    Using the FLOOR and CEIL functions, and the LOGFACT and POWER variables, "nice" end points were calculated.
  - adjust minimum or maximum if necessary
    If all values were positive but the process produced a minimum that was negative or alternatively, if the reverse was true, then that end point was set to 0.
  - determine appropriate step amount and number of major tick marks
    4, 5, 3 were the desired number of major tick marks. The best fit was determined using the MOD function and the calculated value of the step. If none fit evenly, then the appropriate end point would be pushed out enough to produce an even fit.
- determine appropriate number of minor tick marks
  A similar process occurred here except if no fit was possible then a single minor
  tick mark was specified.

- assign values to macro variables
  The values needed for the VAXIS and HAXIS statements, i.e., vmin, vmax,
  vstep, vtick - were placed into macro variables using CALL SYMPUT.

PROC GPLOT
run the graphs requested using the values
of the macro variables

---

CODE DESCRIPTION
As described here, this application uses two macros to invoke
and execute the algorithm. The results of this application
produces the macro variables needed to specify the parameters
needed in the GPLOT or PLOT procedure. Macro SETUP
obtains the needed information - data set name, x-axis variable, y-axis variable, as well as the adjustment factor, and invokes macro
AXIS, which executes the algorithm. An illustration of the major
actions occurring in this application is displayed in Figure 1. Macro setup
is invoked with the parameters specifying the
information needed in order to run the algorithm and calls macro
axis which then determines the minimum and maximum values,
the step and number of minor tick marks. These are passed via
macro variables to the graphics procedure to produce the
desired graphics output.

---

MACRO SETUP

MACRO AXIS

PROC SUMMARY
Find observed minimums and maximums

DATA STEP
Determine range between minimum and maximum
Create new end points by adding or subtracting
5% to the minimum and maximum
Calculate LOGFACTOR and POWER variables to reduce
impact of magnitude and create 'nice' numbers
Calculate nice end points
Adjust minimum and maximum if needed
Determine appropriate number of tick marks
Assign values to macro variables

---

Figure 1
SPECIFIC APPLICATIONS

The 3 examples presented here illustrate some of the problems that occurred with the default axis scaling, and a view of the same graphs after the algorithm produced the necessary parameters.

Example 1: In this case, using the PROC PLOT procedure, the default results in the axis having a series of values clustered in the center of the vertical axis. The graph produced by TOXSAS has the points evenly distributed over the entire axis.

Example 2: the single observation case. Note that the default places the value of that 1 point in the middle of the axis, with no end points. Although the placement of the observation doesn't change after using the new algorithm, the users were more pleased with a specified minimum and maximum.

Example 3: the users were not pleased with the "non-nice" end points or the cluttered tick marks presented on this graph. The alternative graph presented "nice" major tick marks as opposed to printing every value from 7.98 to 8.18.
Example 2

TOXICOLOGY ANALYSIS SYSTEM

BODY WEIGHT AT TWO DAY INTERVALS

Example 3

TOXICOLOGY ANALYSIS SYSTEM

Body Weight at Two Day Intervals
CONCLUSION

The real strength of this scaling algorithm lies in its predictability. Using the default options with the PROC PLOT or PROC GPLOT procedures may produce perfectly acceptable output, but in an automated system such as TOXSAS, the uncertainty is not welcome. Furthermore, adding or deleting footnotes or titles from a graph may impact the default output presentation, adding another element of uncertainty. The macro environment allows for the results of the scaling algorithm to be easily used with a graphics procedure, especially necessary in an automated system. This scaling tool, used in a macro environment, was able to provide a consistent way to output a wide variety of data to our users satisfaction.

SAS and SAS/GRAPH are registered trademarks of SAS Institute Inc., Cary, NC.

VAX/VMS is a registered trademark of the Digital Corporation.

ACKNOWLEDGEMENTS

I want to thank Roger Staum, Medical Research Division-American Cyanamid, for his invaluable ideas and suggestions in the development of this application as well as in the preparation of this presentation.

AUTHOR CONTACT

The author may be contacted for questions or comments:

Marcia S. Murto
Lederle Laboratories
190/208C
N. Middletown Road
Pearl River, NY 10965
(914) 732-2852