

## Efficient Graphing of Basic Data Structure – Let Data Decide Scale of Y-Axis for Themselves

KuenHung Lin, PAREXEL International Co., Ltd, Taipei, Taiwan

### ABSTRACT

When creating information-rich graphics, one of difficult challenges is how to set appropriate scale of y-axis. As creating figures based on Basic Data Structure (BDS) with several parameters, user must try each parameter one-by-one to find suitable scale. In this paper, we would like to introduce an idea that let data decide optimized minimum, maximum and increment of y-axis for each parameter automatically with fixed number of major ticks. The advantage of this method is that user does not have to adjust the scale of y-axis per parameter manually. User can do efficiently while creating BDS graphic, such as line chart for various parameters in laboratory data.

### ACRONYM

AVAL: Analysis Value

BDS: Basic Data Structure

GCD: Greatest Common Divisor

### NOTATION

$y_{max}$ : Maximum value in y-axis

$y_{min}$ : Minimum value in y-axis

$dy_{max}$ : Decimal number of  $y_{max}$

$dy_{min}$ : Decimal number of  $y_{min}$

$dy$ : Minimum between  $dy_{max}$  and  $dy_{min}$

$gy_{max}$ : Maximum tick value of y-axis for graphic

$gy_{min}$ : Minimum tick value of y-axis for graphic

$\Delta$ : Increment as  $(gy_{max} - gy_{min})/\text{number of major ticks}$

### INTRODUCTION

For graphing of BDS, x-axis is usually corresponding to time frame. In other words, the x-axis is defined by study materials, such as protocol and statistical analysis plan (SAP).

Because the ranges of AVAL are widely different for parameters in BDS, we relatively struggle to deal with y-axis. If we try each parameter one-by-one to find suitable scale of graphic, it is lack of efficient. Hence, we would like to introduce an idea let data decide optimized minimum, maximum and increment of y-axis for each parameter automatically with fixed number of major ticks.

The rest of this paper is organized as follows. First, we use two data sets to create three graphics to describe the motivation of this study. Second, we introduce the proposed algorithm. And then, we use the data sets of motivation to illustrate the proposed algorithm. Finally, some concluding remarks are given at the end of this paper.

## FRAMEWORK AND MOTIVATION

In this section, we use two examples with different scale of y-axis to illustrate the motivation of this study. For the first data set, we would like to raise two practical issues as below.

- Figure 1-1 shows the major ticks of y-axis are created tick by SAS® automatically. There are not enough major ticks of y-axis as reference of AVAL because clinical trial reports need more information in y-axis for reference. In our experience, 10 major ticks of y-axis are more appropriate.
- Figure 1-2 shows the major ticks of y-axis are assigned to improper number of major ticks of y-axis by user manually. The displayed output does not provide a clear profile of the data.

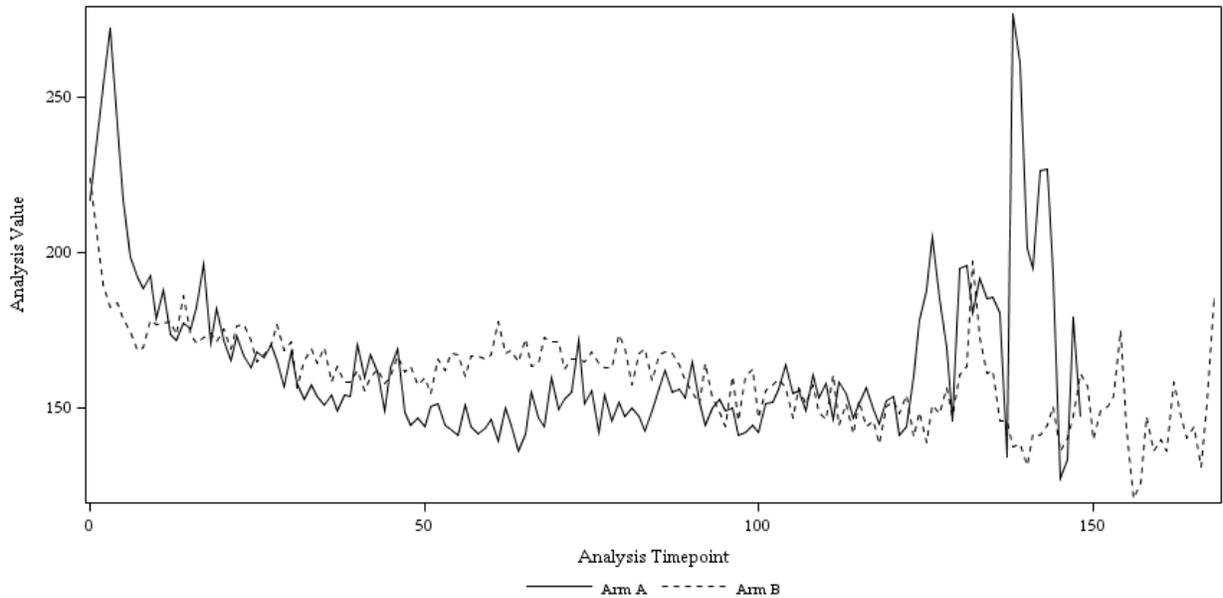


Figure 1-1. Line chart with default ticks

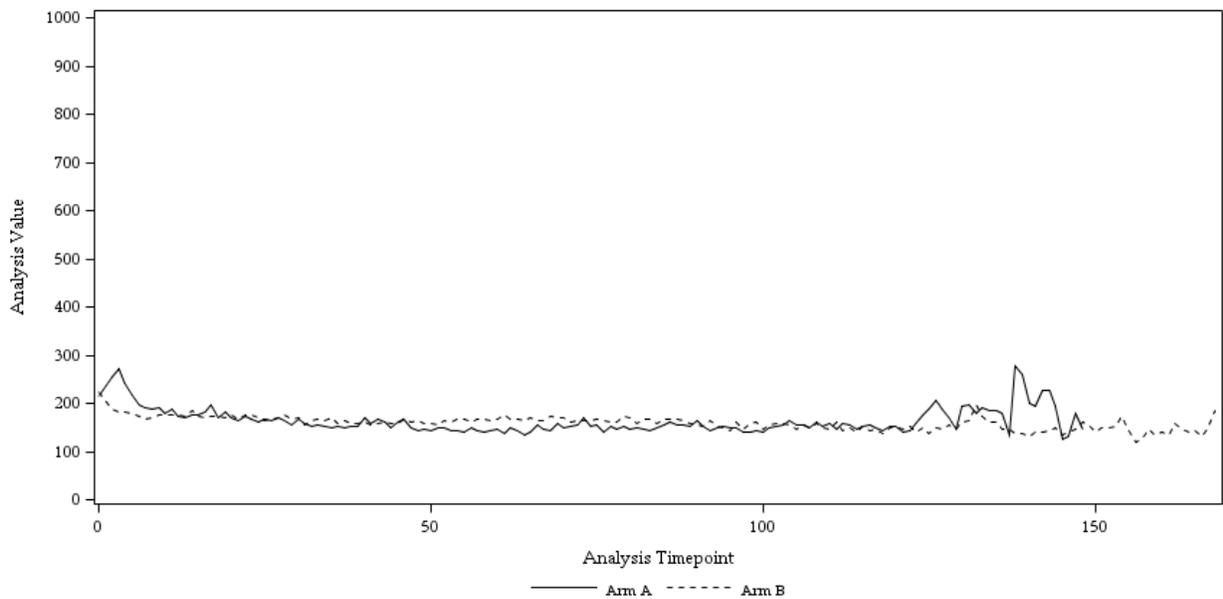
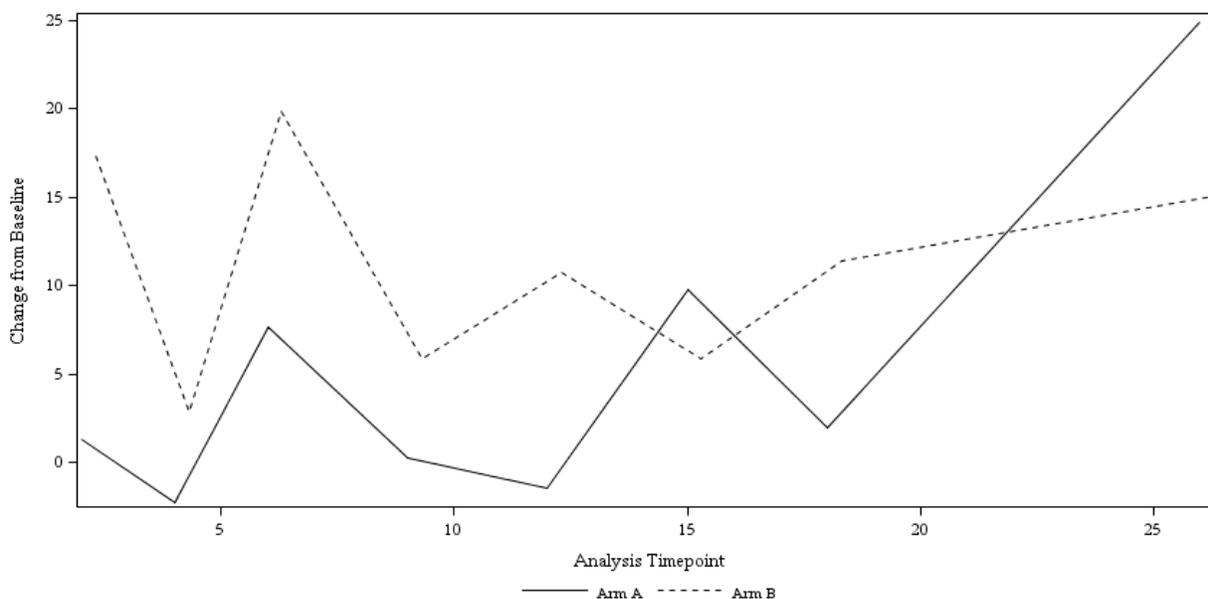


Figure 1-2. Line chart with improper ticks

For the second data set, I would like to discuss another practical issue as below

- Figure 2 shows the major ticks of y-axis is lack of negative tick. There will be a challenge, if client would like to add negative tick and keep major tick of 0 as reference.

For those three issues, we would like to illustrate the proposed algorithm in the next section.



**Figure 2. Line chart with lack of negative ticks**

## LET DATA DECIDE SCALE OF Y-AXIS FOR THEMSELVES

At first, we need data to introduce themselves. Data will provide the bound of analysis value, such as  $y_{max}$  and  $y_{min}$ . If AVAL is derived,  $y_{max}$  and  $y_{min}$  will be rounded to specify decimal number, instead of automatic rounding by SAS. And then, we will apply  $y_{max}$  and  $y_{min}$  to below algorithm of the macro %AutoTicks to obtain optimized minimum, maximum and increment of y-axis.

- Step1: Selecting the minimum from  $dy_{max}$  and  $dy_{min}$  as decimal number of y-axis, for example one decimal number is 1 and another is integer, the values of tick will be integer on y-axis.
  - Corollary: The decimal number of  $dy_{max}$  and  $dy_{min}$  which is not equal to  $dy$  is greater or equal to one.
- Step2: Rounding  $y_{max}$  and  $y_{min}$  based on  $dy$ .
- Step3: For y-axis more intuitively, we will assign the greatest value from those values which are smaller than  $y_{min}$ , and the last digit is 0 or 5 to  $gy_{min}$ . Similarly, we will assign the smallest value from those values which are greater than  $y_{max}$ , and the last digit is 0 or 5 to  $gy_{max}$ . We assign number of major ticks as 10 consequently the increment of last digit will also be 0 or 5.
- Optional: To solve the issue has been mentioned in Figure 3 when  $gy_{max} > 0$  and  $gy_{min} < 0$ , we follow below steps to re-assign scale of y-axis.
  - Step1: If absolute of  $gy_{max}$  and  $gy_{min}$  are equal, we can directly use current  $\Delta$  because the number of major ticks for negative is the same with positive values.
  - Step2: If absolute of  $gy_{max}$  and  $gy_{min}$  are not equal, we use GCD of  $gy_{max}$  and  $gy_{min}$  to obtain number of major ticks for negative and positive values. Because GCD is only applied to integer value, if  $dy > 0$ ,  $gy_{max}$  and  $gy_{min}$  must to re-scale as integer. And then, if summation of number of major ticks for negative and positive values smaller than number of major ticks, we get back GCD of  $gy_{max}$  and  $gy_{min}$  to original scale to obtain  $\Delta$ .
  - Step3: If absolute of  $gy_{max}$  and  $gy_{min}$  are not equal and summation of number of major ticks for negative and positive values greater than number of major ticks, we will check difference between absolute of  $gy_{max}$  and  $gy_{min}$ . If the difference is few, we will use the maximum value between  $gy_{max}$  and  $gy_{min}$  to replace the minimum value between  $gy_{max}$  and  $gy_{min}$ . And then, we can refer to optional step1 to assign scale of y-axis.

- Step4: If absolute of  $gy_{max}$  and  $gy_{min}$  are not equal, summation of number of major ticks for negative and positive values greater than number of major ticks and difference between absolute of  $gy_{max}$  and  $gy_{min}$  are huge, it means there is not enough space to display positive, 0, and negative major ticks at the same time. For the scenario, we will not re-assign scale of y-axis.
- Finally, the algorithm will provide the optimized major ticks information of y-axis by macro variable which is composed of “ $gy_{max}$  to  $gy_{min}$  by  $\Delta$ ” depend on which graphic procedure is used.

## CASE APPLICATION

The macro %AutoTick will be called using the following structure:

```
%AutoTick(
  _lib = WORK,
  _ds = ,
  _var = AVAL,
  _par = PARAM,
  _deci = ,
  _NumMajTicks = 10,
  _gcd = ,
  _proc = SGPLOT,
  _tickvar = _ytick
);
```

**Table 1. Macro Parameters**

Name	Description	Allowed Values	Default Value
_lib	Library of Data set to graph		WORK
_ds	Data set to graph		REQUIRED
_var	Variable of value in y-axis		AVAL
_par	Variable of the graphic parameter		PARAM
_deci	Decimal number of $y_{max}$ and $y_{min}$ . Default is $dy$ .	BLANK   Integer	BLANK
_NumMajTicks	Number of major ticks in y-axis	Positive integer	10
_gcd	Flag to specify whether to add major ticks of negative value and keep major tick of 0 in y-axis when $gy_{max} > 0$ and $gy_{min} < 0$	BLANK   Y	BLANK
_proc	Specify which procedure is used to create graphic	GPLOT   SGPLOT   TEMPLATE	SGPLOT
_tickvar	Name of macro variable which is composed of “ $gy_{max}$ to $gy_{min}$ by $\Delta$ ”		_ytick

Now, we use two data sets in frame work and motivation section to illustrate the proposed algorithm.

## HOW TO ADD ENOUGH AND APPROPRIATE MAJOR TICKS

For the first data set with structure as Table 2,

**Table 2. Data Structure of the first data set (motivation1)**

PARAM	AVISITN	AVAL	TRTAN
Motivation1	0	216.44897959	1
Motivation1	0	223.98039216	2
Motivation1	2	253.53333333	1
Motivation1	2	189.78	2
Motivation1	3	272.54761905	1
Motivation1	3	182.25714286	2

there are some notes as below,

- Because there is some AVAL with cyclic decimal, the decimal number of y-axis should be given. According to Figure 1.1, we will assign 0 to decimal number of y-axis.
- $y_{max}$  and  $y_{min}$  are maximum and minimum value of ALL AVAL without considering strata in actual treatment group, such as TRTAN.

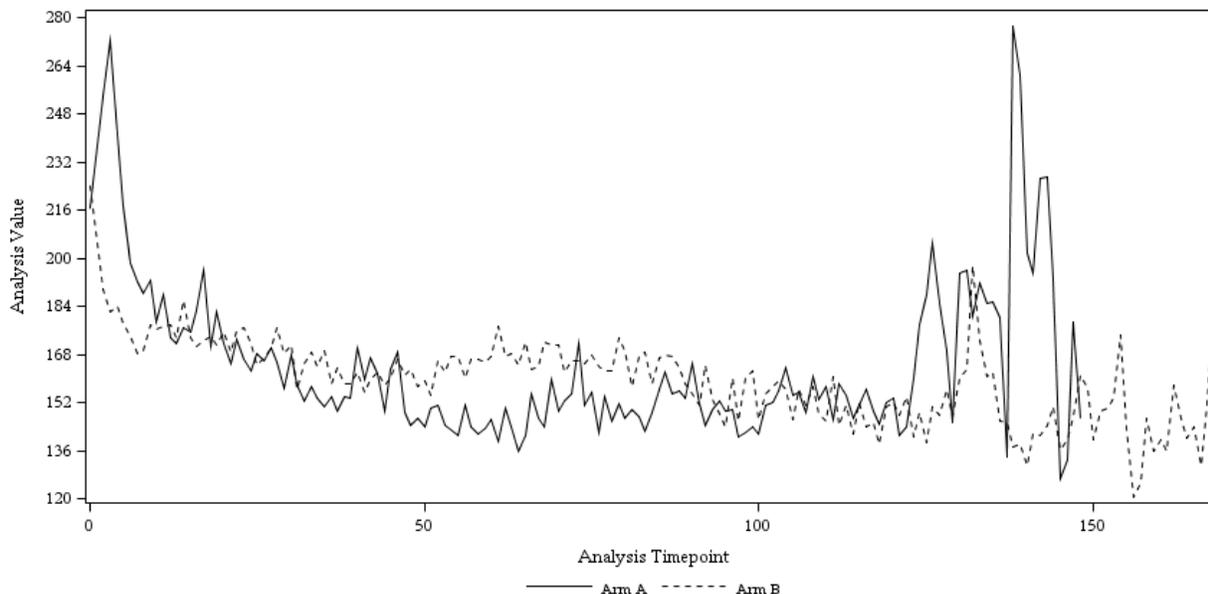
The following call to the macro was implemented:

```
%AutoTick(
  _lib = WORK,
  _ds = motivation1,
  _var = AVAL,
  _par = PARAM,
  _deci = 0,
  _NumMajTicks = 10,
  _gcd = ,
  _proc = SGPLOT,
  _tickvar = _ytick
);
```

And then, we applied return macro variable (&\_ytick) to YAXIS option of graphic procedure (SGPLOT) to create Figure 3, such as

```
yaxis label = "Analysis Value" values = (&_ytick.);
```

Compared with Figure 1.1 and Figure 1.2, Figure 3 has more enough and appropriate major ticks to provide more clear profile of the data.



**Figure 3. Line chart of analysis value with appropriate ticks**

### HOW TO ADD NEGATIVE MAJOR TICKS AND KEEP MAJOR TICKS OF ZERO

For the second data set which has the same structure with Table 2, there are some notes as below,

- When we assign 1 to decimal number of y-axis,  $y_{max}$  and  $y_{min}$  will be 25.0 and -2.5. According to our method which were mentioned above, we will use GCD function to obtain integer number of major ticks for negative and positive value in y-axis, both are 10 and 1. Hence, there are at least 12 major ticks in y-axis to display positive, negative, and zero value in y-axis.
- If we assign 0 to decimal number of y-axis,  $y_{max}$  and  $y_{min}$  will be 25 and -5. And then, the number of major ticks is 7. However, we still assign 12 to number of major ticks at that time. It means the option of number of major ticks is upper bound. And, we can obtain different number of major ticks based on different decimal number.

- If the number of major ticks by GCD method is larger than our macro option setting, and absolute  $y_{max}$  and  $y_{min}$  are obviously difference. For the scenario, the GCD method will not be used to re-assign scale of y-axis to add negative major ticks and keep major ticks of zero.

The following call to the macro was implemented:

```
%AutoTick(
  _lib = WORK,
  _ds = motivation2,
  _var = AVAL,
  _par = PARAM,
  _deci = 1,
  _NumMajTicks = 12,
  _gcd = Y,
  _proc = TEMPLATE,
  _tickvar = _ytick
);
```

And then, we applied return macro variable (&\_ytick) to YAXIS option of graphic procedure (TEMPLATE) to create Figure 4, such as

```
yaxisopts = (label = "Change from Baseline" linearopts = (&_ytick.) display = ALL);
```

Compared with Figure 2, Figure 4 has more major ticks, such as major ticks of negative value and zero.

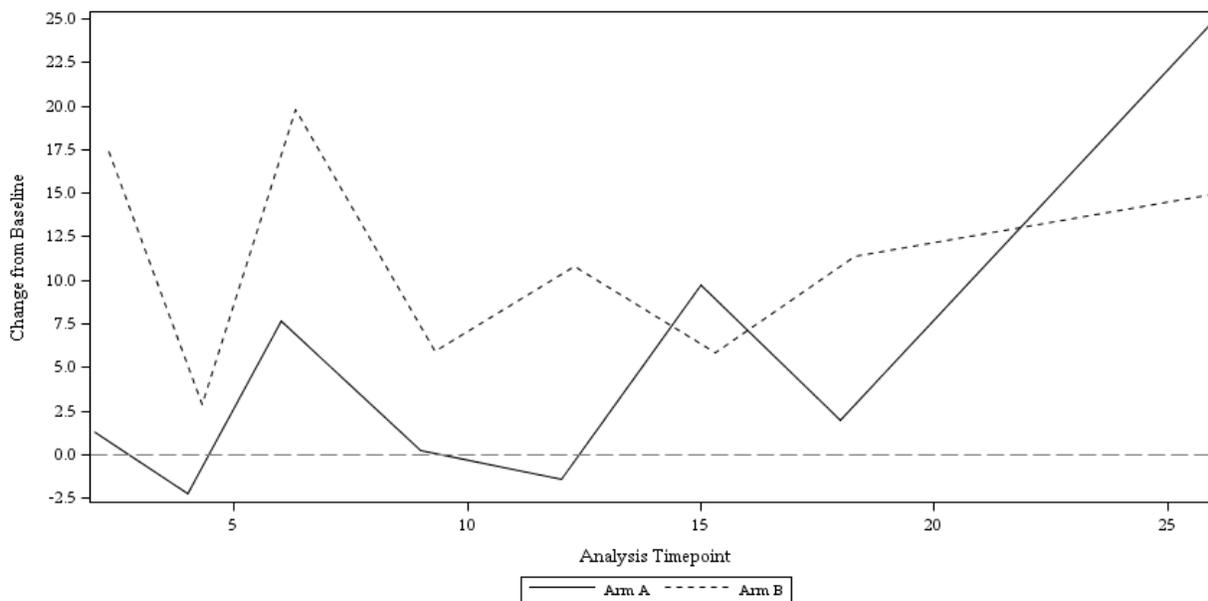


Figure 4. Line chart of change from baseline with appropriate ticks

## CONCLUSION

The macro introduced in this article was designed to let data decide scale of y-axis for themselves. The most advantage of our method is user can do efficiently while creating BDS graphic without trying each parameter one-by-one to find suitable scale. However, if there is any unusual value, such as outlier, data will also directly let us know. In this scenario, our method can't provide customized scale of y-axis to exclude outlier. Hence, further development of the macro is anticipated to cater to some extreme cases. To let data tell us their profile without limit.

## REFERENCES

[1] Watts, Perry (2009), "Generate a Customized Axis Scale with Uneven Intervals in SAS® Automatically" Proceedings of the SAS Global Forum 2009, paper 192-2009. Available at <http://support.sas.com/resources/papers/proceedings09/192-2009.pdf>

[2] Groulx, Adrienne (2006), "*Graphing Laboratory Data and an Introduction to the Custom Macro %GRPAHME*" Proceedings of the 31th Annual SAS Users Group International 2006, paper 088-31. Available at <http://www2.sas.com/proceedings/sugi31/088-31.pdf>

[3] Peszek, Iza (2000), "*How to Customize Axes in PROC GLOT*" Proceedings of the 25th Annual SAS Users Group International 2000, paper 110-25. Available at <http://www2.sas.com/proceedings/sugi25/25/cc/25p110.pdf>

## ACKNOWLEDGMENTS

The author would like to thank Robert Tai for valuable comments that help improve the manuscript.

## CONTACT INFORMATION

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

Name: KuenHung Lin

Enterprise: PAREXEL International Co., Ltd

Address: 22F, Far Glory International Center, No 200, Sec. 1 Keelung Road, Taipei, Taiwan 11071, ROC

E-mail: [KuenHung.Lin@PAREXEL.com](mailto:KuenHung.Lin@PAREXEL.com)

Web: <http://www.PAREXEL.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. ® indicates USA registration.

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