

How to Prevent Producing Misleading Graphical Displays

Julie A Oates

ABSTRACT

Table displays often summarize the data clearly, but graphical displays have the pictorial element that most people prefer, especially when giving a presentation. Our ability to understand graphical information varies between individuals based upon their knowledge and purpose for reviewing the graphical displays. This paper highlights the importance of creating clear and concise graphics so as not to mislead the reviewer.

INTRODUCTION

During my programming career I have had the opportunity to produce many graphical displays. The choice of graphic and format layout is critical in the prevention of producing those misleading and deceptive graphics we have all seen in the past, whether it be during presentations or within the media.

This paper highlights the considerations you should take to avoid a graphical nightmare. I am looking at graphics in their simplest form and cover the following areas and where possible show plots to highlight the various issues:

- Titles and Footnotes
- Axes
- Symbols
- Use of colour
- Multiple plots on one page
- Type of plot

TITLES & FOOTNOTES

When considering the titles and footnotes to add to a plot, you need only consider adding information which will enhance the plot.

It is a fine line between too little information and crowding the plot with oversized or too many titles and footnotes and hence reducing the plot size. If you reduce the plot size you could be losing vital information on the plot.

AXES

Axes define the scale of the plot and badly selected axes can completely ruin the point of creating the graphical display. If the axes are too large you will suppress any obvious differences: too small and you could magnify a minute insignificant difference. If the axes are not logically labelled (i.e. ticks; the dash marks on the axes, label etc), the result could mislead and confuse the reviewer.

Where possible plots should start at zero unless there is a compelling reason not to.

EXAMPLES– TITLES/FOOTNOTES AND AXES

Figure 1A (bad example) shows:

- Too many titles and footnotes which shrinks the size of the actual plot
- Badly defined axes
 - The axes are not labeled clearly
 - The scale of the axes are inappropriate for data

Figure 1B (improved example) shows

- More concise titles and footnotes
- The algorithm is referenced in the footnote, rather than written out in full
- Axes are clearly labeled and have appropriate scale.

Figure 1A: Titles/Footnotes & Axes - How not to do it!

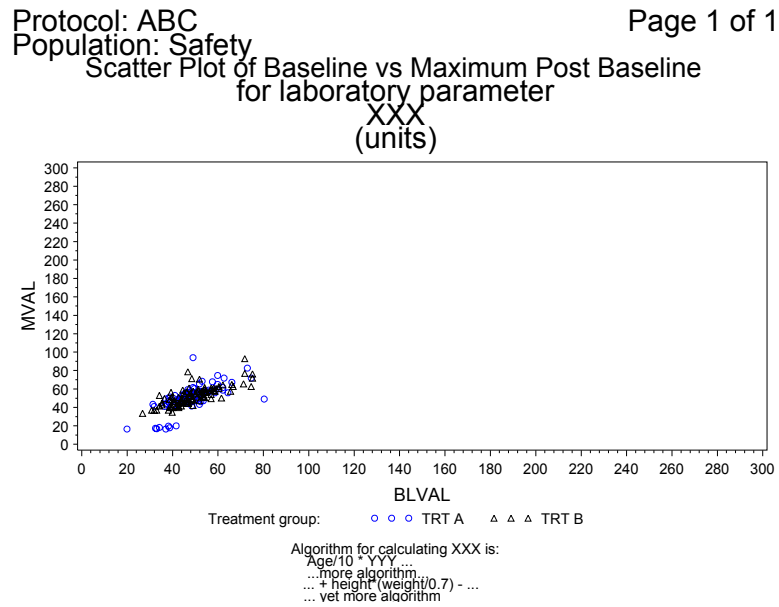
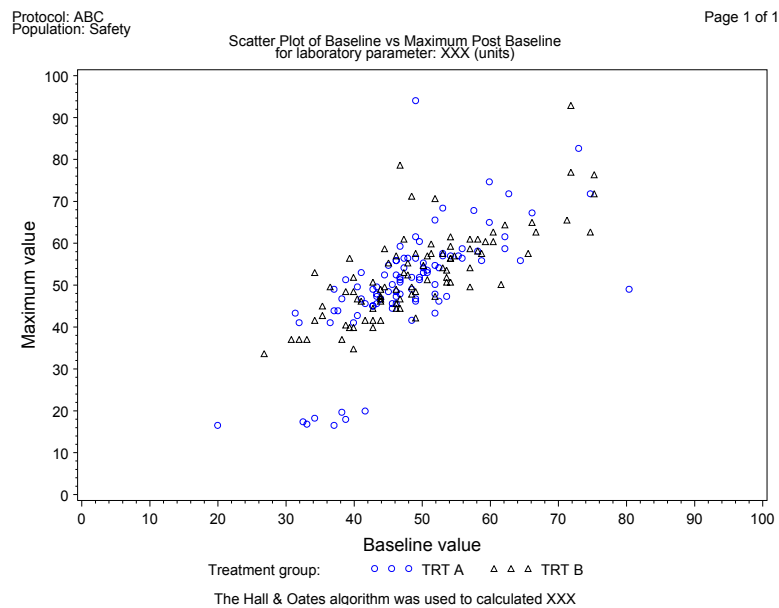


Figure 1B: Titles/Footnotes & Axes - Better



SYMBOLS

I used to consider selecting symbols for a plot to be a personal choice and to some degree it still is, however careful selection ensures you are not misrepresenting the data, for example biasing a particular group.

I recently had to produce a plot of lab values with a particular event vs those with no event to show that having the event had no effect on the distribution of the “Lab” value. My first thoughts were to use a filled-in circle for those with an event and an empty circle for those with no event. Logically this seemed to make sense, but as you can see in Figure 2A, the “event” group appears to have more values plotted, though I have plotted exactly the same values for both groups.

USE OF COLOUR

Care needs to be taken when selecting colours for your plot. If the graphical display is only intended to be reviewed on paper (i.e. in black & white), the colours you choose would only appear in various shades of grey. The contrast is often insufficient in the definition between the groups. For those displays which will be viewed in colour, choose contrasting colours and do not choose colours which highlight one particular group more than the other. Keep away from the colour yellow as it is very difficult to view on a white background, see the second main title in Figure 2A.

Note: At last year’s PhUSE conference LeRoy Bessler produced an excellent poster regarding colour “Effective Communication with Colour”

EXAMPLES– SYMBOLS & COLOURS

Figure 2A (bad example) shows:

- Inappropriate selection of symbols in this case places more emphasise on one group over the other
- Choice of colour is crucial; I have exaggerated the use of colour in this plot. Can you see the title in yellow? I have even increased the font size, but it is still difficult to read.

Figure 2B (improved example) shows:

- The symbol choice has improved, this time both groups look the same even when colour has been introduced, preventing bias.
- Ideally titles/footnotes should always be black, but for the purposes of the previous example I presented the colour yellow on a white background. This is now much simpler to read when all titles are displayed using black text

Figure 2A: Symbols & Colours – A bad choice!

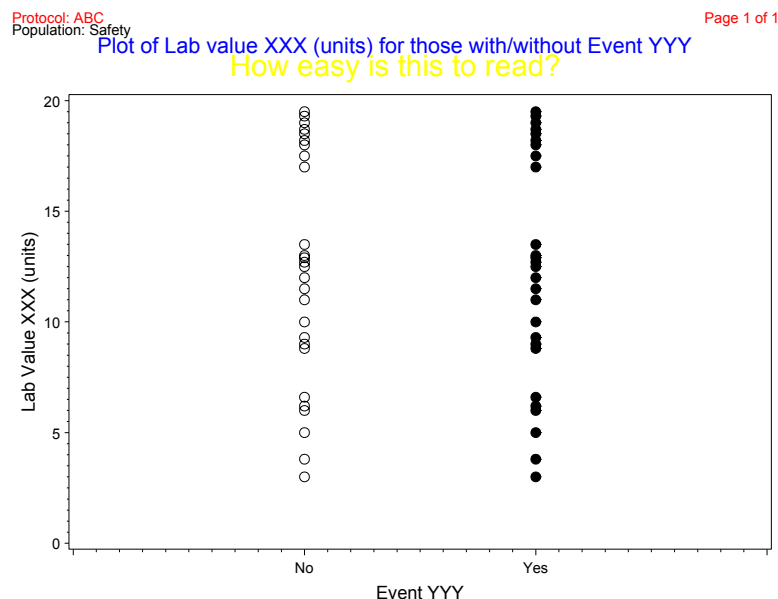
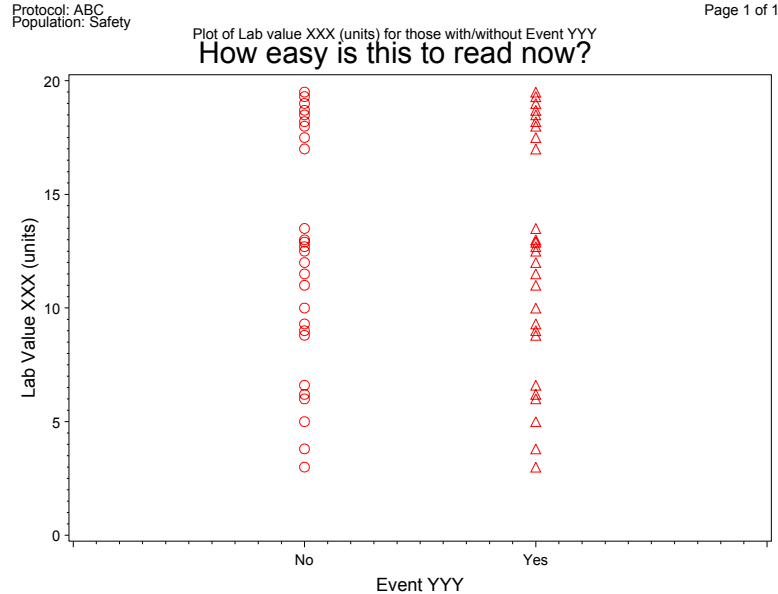


Figure 2B: Symbols & Colours – Better**MULTIPLE PLOTS ON ONE PAGE**

This can be very useful especially when producing related plots or plots which are repeated many times. i.e. patient profiles.

I have been reliably informed the new S-Plus tools are ideal for multiple panel plots, but I am more familiar with the SAS graphical functionality.

SAS already provides standard templates which can be called using PROC GREPLAY, but if you'd prefer to create your own take a look at the standard templates to understand how to set these up. The templates are based upon X and Y coordinates of each of the four corners of each plot.

The main considerations when creating multiple plots on one page are:

- Subtle differences could be overlooked, as each plot will be much smaller to view
- The down scaling of the individual plots needs to be carefully thought through, as resizing a plot could change the overall message compared to presenting one per page.
- An obvious one, ensure the plot sizes are all the same
- Label the plots clearly
- For comparative plots ensure the axes are the same to allow for fair comparisons between each plot (see an example of how this principle has been overlooked)

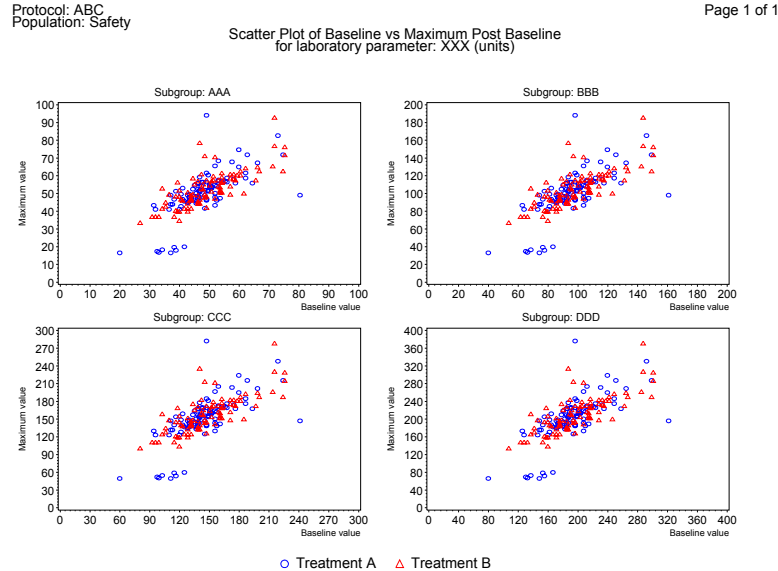
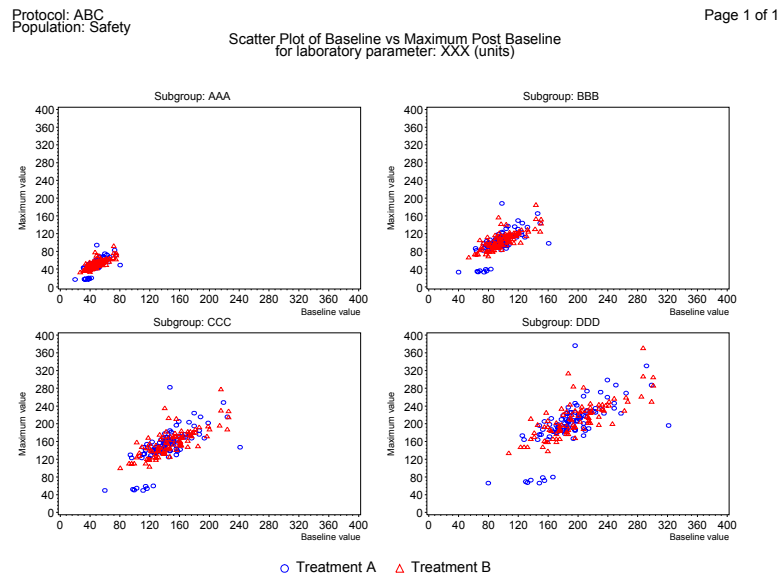
EXAMPLES – MULTIPLE PLOTS ON ONE PAGE

Figure 3A shows:

- Four plots of the same laboratory parameter for different subgroups. These are comparative plots, which look identical on a first glance, but with a closer look these all have different axis ranges which is very deceptive

Figure 3B shows:

- The same four plots as in the previous plot but this time on the same axes scales. It is now clear to see that there is a vast amount of difference between the subgroups which was not apparent from the previous plot.

Figure 3A: Multiple plots – Are these the same? Look closely at the axes**Figure 3A: Multiple plots – Better****TYPE OF PLOT**

Histograms, Scatter plots and Box & Whisker plots are just a few of the many type of plots available when considering how best to represent the data.

Take into consideration the type of data you have, discrete or continuous and select the plot most appropriate to the data.

I personally steer away from 3D plots as these can be easily misunderstood by the reviewer. Two dimensional plots with additional options can be less misleading.

The additional options which provide greater functionality are:

- Annotate – this has greatly been improved and simplified over the years with the functions available. It allows you to add almost anything you desire to your basic plot. i.e. Symbols, text, arrows, Ns etc
- Overlay – you can overlay more than one plot on top of each other. Using overlay allows for 2 axes to be created, one on the left hand side and one on the right.

PhUSE 2006

- Note – I have used “note” more recently for adding the number of subjects at each visit into a “table” just below the box and whisker plot.

CONCLUSION

This paper has only allowed me to skim the surface of the world of graphical displays and the considerations to ensure accurate representation of your data without bias or misinterpretation.

Hopefully I have given an insight to the main basic areas to consider allowing you to start producing clear and concise graphical displays.

Here is a recap:

- Ensure titles and footnotes explain the graphical display clearly, but be careful not to crowd the plot
- Define the axes; clearly label each axis and ensure the axes ranges compliment the data
- Select the symbols which provide enough contrast between the groups of interest but do not place bias towards one group in particular
- Choose the colours carefully, selecting contrasting colours without bias
- Be consistent between your choice of symbol and colour across multiple graphical displays
- Avoid yellow as this is often difficult to view against a white background
- Multiple plots on one page can be effective as long as the information from the individual plots are not lost (due to size reduction) and each plot is clearly labelled
- Ensure comparative plots have the same axis scales
- Selecting the type of plot is crucial to ensure the graphical display represents the data in the best way possible. Consider the options and don't be afraid to question the graphical choice requested if you do not think it compliments the data
- Use 3D plots with caution, as these can often complicate and mislead the eye of the reviewer
- 2D plots with additional options such as using the annotate facility, overlay and notes provide extra information on the plot without clouding the overall message

Remember the reviewer often only has a moment to “get the message”, so in short, keep it simple, keep it clear and the graphical display should provide the reviewer with the message intended.

REFERENCES

PhUSE 2005 Posters

Title: Effective Communication with Colour [PS18]

Author: LeRoy Bessler

SAS online help

ACKNOWLEDGMENTS

I would like to thank my colleagues who gave up their time to provide honest feedback regarding the contents of this poster and paper.

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.