IS A PICTURE REALLY WORTH A THOUSAND WORDS?

James A. Sannerud, Compumetrics, Inc.

ABSTRACT

The abundance of computer hardware and the power of statistical software have simplified the time consuming process of creating presentation quality graphics. SAS/GRAPH provides the SAS software user the ability to produce high quality graphical output. Unfortunately, hardware and software cannot compensate for a lack of skill in graphical design. Unless the analyst has an appropriate understanding of the principles of effective graphical design, the presentation of the data may not convey the intended message. A poorly designed or misleading graph is equivalent to a grammatical error in speech or text. This paper explores the subject of graphical design and discusses both effective and ineffective examples of each.

INTRODUCTION

Graphics are a means of communicating information and ideas. As with any other form of communication, you must first understand what you want to communicate. Does the data require a graphical presentation at all? In many cases the answer will be no. Rather than force data into an inappropriate medium, consider the use of tables or data embedded in text. Text is useful for explaining or interpreting data, or when comparing only two numbers. For example:

Net earnings increased from $1.34 last year to $1.52 this year.

More complex data can be presented in tabular form. Ordering the table by data value facilitates comparisons. For example:

<table>
<thead>
<tr>
<th>Division</th>
<th>Contribution to Revenue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eastern</td>
<td>47%</td>
</tr>
<tr>
<td>Western</td>
<td>34%</td>
</tr>
<tr>
<td>Central</td>
<td>19%</td>
</tr>
</tbody>
</table>

A well designed graph will facilitate the analysis of large volumes of data. The graph should capture and hold the viewers attention, inviting further analysis. In this way, the viewer is more likely to remember the graph than he or she is likely to remember a page of text. Graphics can permit detection of relationships and trends in data by making patterns obvious.

How do you judge the graphic format most appropriate for your data? First understand the data and the message you want to communicate. Answer some basic questions to help you focus on the best format, such as are the data discrete or continuous. Of course, you should have at least a basic knowledge of the procedures available in SAS/GRAPH and some of the more useful options. Visualize the graph and sketch it with pencil and paper. Experiment with different formats until you are satisfied. Once you are able to roughly draw the graph, you should be able to produce it with SAS/GRAPH.

EFFECTIVE GRAPHICAL DESIGN

The first step in designing an effective graph is to become familiar with the data. It is critical that you understand its content and scale. The designer must also have familiarity with statistics as well as a basic understanding of available graphic formats. The next step is to decide on a presentation format. The choice of presentation format is determined by several other factors including the objective of the graph, the nature of the data, and its intended audience.

The objective of a graphic presentation is the accurate communication of information in a clear and concise manner. Good graphics make hidden patterns in data obvious. The graph should be clean and attractive, free of confusing clutter and extraneous information. The viewer should be drawn into the graph and captivated by its content. Its message should be self-evident and not require excessive footnotes, tedious legends, or pages of accompanying explanations. Sets of related graphics should use a consistent format for ease of interpretation.

The nature of the data may itself determine the basic
presentation format. For example, historical data such as closing stock prices are best presented as a time series whereas data related to geographic area such as population or weather information are best presented on a map. In some cases the nature of the data may dictate the use of text or a table. Graphics are more visually effective in a landscape rather than portrait format. Landscape format, with the horizontal axis longer than the vertical, allows the eye to move naturally from left to right and is easier to label.

It is important to gauge the sophistication of the audience before you design a graph. Because graphics are a form of communication, the graph must be comprehensible by those you are communicating with, be they senior management, your peers, readers of a technical journal, or the general public. In any group, the population is diverse, hence you must appeal to all levels. A semi-log plot might be readily understandable to some readers of a technical journal, but baffling to many others. Misuse or overuse of graphics can insult the intelligence of an audience. Graphics need not dominate a presentation, but rather should highlight the salient points. Remember: not all information has to be communicated with pictures.

Most important, the graph must tell the truth, the whole truth, and nothing but the truth. The data must be accurate and must not intentionally or unintentionally mislead the viewer. Data, like speech, must not be taken out of context. People are naturally skeptical about statistics and their graphical representation. Even accurate data are subject to interpretation problems if not presented clearly. The most common causes of ineffective graphics I'll refer to as: muddled message, sloppy scaling, poor perspective, ignored index, and contempt of context. Each will be explored in more detail.

Muddled Message

Leave art to the artists. While it can be argued that a fine graphic is a work of art, we are more interested in communicating information to help the viewer make a decision rather than trying to use as many features of SAS/GRAPH as possible. It is unfortunate that many business graphics are discarded or ignored because the designer has forgotten this basic premise. Think of how you would feel after spending hours of your time creating a "work of art" only to have people not use it because they did not understand it. Also, graphics are not intended to impress your management with your artistic ability, but rather with your ability to communicate. Evaluate your work critically, because others certainly will.

Avoid the use of pie charts. Pie charts, one of the most popular graphics in use today, are also some of the least effective. By definition, pie charts represent 100% of some quantity and how that quantity is subdivided. With pie charts, it is cumbersome to both label and order the slices. Comparisons between pie slices are more difficult for the viewer than comparisons between horizontal or vertical bars. If used at all, pie charts should be limited to four or five slices. It is almost always possible to substitute a simple or stacked bar chart for a pie chart. Please do not compare two pie charts. The only design worse than a pie chart is multiple pie charts.

Many problems of graphical clutter are caused by abuse of grid lines, cross hatching, and color. Grid lines, if used at all, should be used sparingly and should be thin. An occasional reference line is preferable. When creating a graph it is easy to err on the side of using more ink than is necessary. Omit all superfluous lines, axes, and titles: all clutter that detracts from the message of the graph. For example, eliminate the redundant labelling of a value in a bar chart where the value is easily derived by looking at the axis or at the height of the bar. Excessive ink may also be found within the confines of the graph itself. When dealing with discrete versus continuous data, avoid the tendency to connect discrete data points (the "connect the dots" syndrome), implying a continuous function, when a dot or needle plot would be more truthful.
Titles and labels are as important to the total effect of the graph as the choice of graph type. Nevertheless, one could easily detract from the value of a graph by misusing titles, legends, or axes. It is best to stick to one or two font styles, and use upper and lower case for maximum readability rather than using all capital letters. Use fonts like SWISS or Helvetica for titles and other labels where clarity is critical. Excessive style changes distract from the effect of the graph, as does excessive labelling.

We have all seen the optical illusion know as the Moire effect, where a graph appears to vibrate or shimmer. This is caused by cross hatching, stripe patterns, and equally spaced vertical or horizontal bars. Moire effect is visually distracting and can be avoided by using gray shading or by annotating the chart with text.

Color is a frequently abused graphic technique in the misguided quest for a pretty picture. Save color for emphasis; use strong colors sparingly. Despite attempts to assign a hierarchy to colors similar to the familiar visual spectrum, most viewers must refer to the legend to interpret colors used this way. It is thus more effective to use shades of gray to indicate density patterns.

SLOPPY SCALING

Mistakes in scaling of either the horizontal or vertical axis commonly contribute to the incorrect interpretation of a graph. Unexpected origins, breaks in continuity, and changes in granularity can deceive the viewer regardless of whether or not the graph is correctly labeled. When comparing relative differences in values, the origin of the Y axis should always be zero. If the origin is not zero, the percentage difference could be perceived as significant, when in reality it is minimal. For example, plotting summer temperatures on the Y axis with an origin at 70 degrees will overemphasize normal temperature fluctuations compared with a graph whose origin is at zero degrees.

Breaks in the scale often depicted by a jagged line on an axis cannot be tolerated. Distortions similar to those produced by a non-zero axis result when large sections of the X or Y axis are omitted. When the data range is extreme, such as Y values of 1, 10, 1000, and 10000, a logarithmic scale may be more appropriate.

Space on a scale must allow for missing values. Missing data could cause inadvertent breaks in an axis, as, for example, when plotting daily temperatures where a few days of data are missing. In this case, a zero value would clearly be incorrect; use a special missing value character instead. Avoid using two sets of X or Y axes and do not change the granularity of the axis in midstream. For example, if you are plotting data values by month, do not shift to weekly values for the most current month.

Labels cannot compensate for a poorly designed graph; an effective graphic must stand on its own. Labels, no matter how clear and well intentioned, may not be immediately noticed. The viewer may think he or she understands the graph at a glance and turn the page, never realizing the deception. Graphics for group presentation do not allow the viewer the luxury of time to study labels and legends. Viewers in the back of the room may be unable to read the labels and legends and may draw erroneous conclusions from the picture.
POOR PERSPECTIVE

Data should be presented in the least number of dimensions and in the least complex form required, lest it result in a misuse of proportion. Visual representation of data must be directly proportional to its numerical value. This error is most frequently manifested by the presentation of one dimensional data in the form of a two dimensional picture graph, even when the data are better suited for a table or simple bar graph. We have all seen picture graphs in newspapers; recently the number of allied and enemy armed forces deployed have been depicted as little airplanes, tanks, or soldiers, where a small soldier may represent one thousand troops and a larger soldier, ten thousand troops. Many of these picture graphs are misleading because the viewer may interpret the value as the area of the symbol rather than the height of the symbol. Typically created by an analyst who is trying to be an artist, there is nothing pretty about this kind of picture. Choose the simple elegance and unmistakable clarity of a bar chart over a visually distorted two dimensional picture.

Worse yet are three dimensional graphs. If you believe you need to use a three dimensional graph, think again. Are you just trying to be an artist? There are valid uses for three dimensional graphics, but they are rare. In fact, shadow effects on three dimensional bar charts may distract the viewer, or worse yet, may mislead the viewer into mistaking the shadow for meaningful data. When using PROC GMAP, choropleth maps may not be flashy, but they are usually easier to interpret than surface, block, or prism maps.

IGNORED INDEX

Ignoring the effects of the changing value of a number over time leads to misleading graphics. The most common example of this occurs with money. Ignoring the effects of inflation will tend to overstate current amounts relative to past amounts unless values are plotted in constant, inflation-adjusted dollars.

Another error is caused by changing a reference point or index without adjusting past data. For example, when a stock splits two for one, all shareholders receive two shares for every one they hold, and the price is halved, so there is no net change in value. Plotting stock price without adjusting for the split would incorrectly indicate a large drop in value. To circumvent this problem, all past data must either be adjusted or disregarded.

CONTEMPT OF CONTEXT

As stated in a popular axiom, statistics can be produced which will support any conclusion. Contempt of context, or taking data out of context, is one way it is done. Data, like speech, must never be taken out of context. Data taken out of context is not statistically significant, but the viewer of the graph will not see the larger context. The graph may tell the truth, but not the whole truth.
Two points define a line but do not make a trend. Normal data variation must be considered, and a sample population large enough to support the hypothesis must be chosen. In addition, there may be other influences on the data from outside the study which can affect its content. For example, economic data such as housing starts is best presented “seasonally adjusted” to account for the normal slowdown in building during the winter months.

Zooming in or exploding one small section of a cyclical graph without showing the surrounding data is really a case of misrepresentation by omission, and can lead to erroneous conclusions. Politicians are infamous for this when “proving” the benefits of partisan legislation. For example, several years ago proponents and opponents of the national 65 mile per hour speed limit law both used the same set of highway accident statistics to produce graphs supporting their respective positions. Each graph was correct for the small subset of data it represented; it was not correct, however, to imply that this data, taken out of context, represented the entire effect of the legislation. Other, unrelated factors, both cyclical and extraordinary, had skewed data and prejudiced statistics. Cyclical factors include the typically higher accident rates during holiday periods; extraordinary factors include unusually bad weather and the expanded presence of police patrols.

CONCLUSION

When developing graphics it is best to experiment with several formats before deciding on an optimal layout for your final graph. Become familiar with SAS/GRAPH procedures and their options, and do not forget the effectiveness of text and tables for presenting small amounts of data. Graphics are best deployed when conveying large quantities of data and complex relationships.

Clarity and accuracy are key: informative graphics need not be works of art. Rather than diluting the effect of a presentation with spurious graphs, keep graphics to a minimum and make every graph count. Effective graphics are like spices: used sparingly and in the right combination, they can liven up any presentation; used incorrectly or in excess, they can spoil even the best presentation.