Paper 049-2011

Finding the Best Display Type for Your Data

Varsha Chawla, SAS, Cary, NC
Lisa Whitman, SAS, Cary, NC

ABSTRACT

With the new release of SAS BI Dashboard 4.3, many new display types are available for dashboard designers to use and take advantage of. Common questions that many dashboard users have will be addressed, including, which display types should be used for various types of data, and how to leverage the different display types so that information is communicated clearly and effectively. Explore the various new charts and displays that can help communicate information, including speedometers, bubble plots, vector plots, waterfall charts, and many other new display types in SAS BI Dashboard 4.3. Learn when to use these displays and how they display your data to create dashboards that are usable, attractive, and efficiently communicate the point you are trying to get across.

INTRODUCTION

Dashboards consist of many different display types, designed to give users what they want out of dashboard: a quick overview of the current status, trends, or forecasts of certain types of information. With a myriad of display type options, the dashboard designer has to keep several goals in mind. These goals vary by organization, audience, and purpose of the dashboard. Most dashboard designers aim to provide information in a way that makes the dashboard easy to read, interpret, and spot important changes in data points, as well as be interesting and pleasing to look at. Variety might not come to mind as your primary goal when designing a dashboard, but it can help set data apart and reveal patterns or information that would not be as apparent when all data is represented using the same display type. Dashboards should be designed so that all of the information is displayed on screen, so it is important to put considerable thought into which display types will convey the most about of information most clearly in the limited amount of space. This paper will introduce new data visualizations offered in SAS BI Dashboard 4.3 and reveal which display types are best for different types of data. Design tips and common pitfalls will also be highlighted so that you can design dashboards that are as usable and effective as they can be.

POPULAR DISPLAY TYPES

The most commonly used display types in dashboards include line charts, and bar charts, and pie charts. Many dashboard users choose these display types because most people are familiar with them. These are the displays that are taught to us in school and we can almost guarantee that anyone from a statistician to a sales person is familiar with a pie chart, line chart, and bar chart. We also predict that these displays will continue to be used heavily in dashboards. However, even the most popular display types have their downsides. The following are cases where it is beneficial to use these display types:

Line charts

Line charts are used for a wide variety of data visualization applications. When stacking lines, they can be used to compare the trend or individual values for several variables. Use line charts when the change in a variable or variables clearly needs to be displayed and/or when trending or rate of change information is of value. Don't just pick a line chart because you have data points. For example, if you simply have ten data points that need to be displayed; the best option is to list these in some order. To categorize them or show variables for multiple values, simply list them in a table. Line charts use a lot of valuable space on a dashboard, so think about whether or not the relationship between data points really needs to be conveyed.

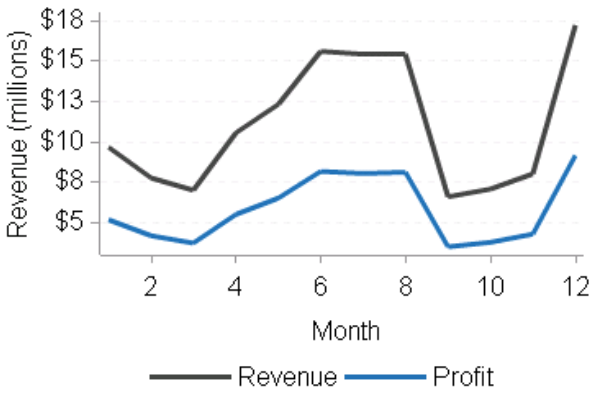

Figure 1. An example of a dual line chart

Bar charts

Bar charts are one of the most commonly used charts that allow the user to quickly compare the value for different variables, values, or categories of data. Use a simple bar chart when values are distinct enough so differences in the bars can be detected by the human eye. When values (bars) are very close together or a large number of values (bars) are displayed, it becomes more difficult to easily compare bars against each other. Bars can have different colors to indicate a particular status or range. Coloring the bars works best when most bars are in a different range or status, as shown in the example below. When all bars are in the same range or status, the color becomes insignificant and needs to be removed or kept consistent.

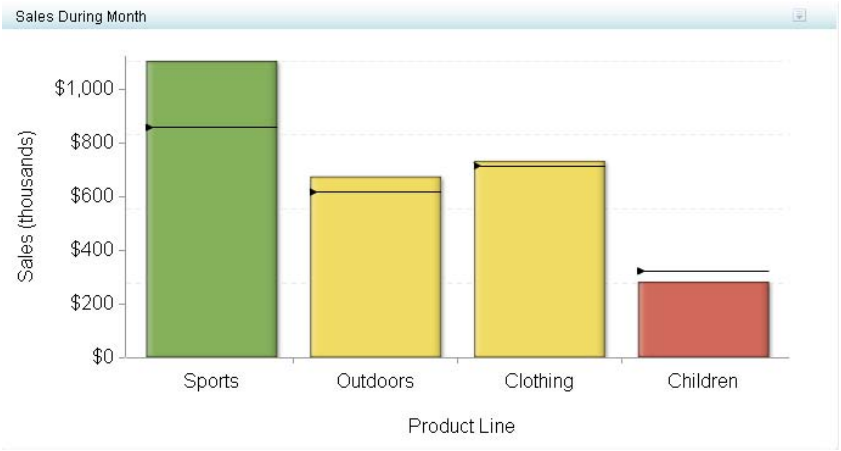

Figure 2. An example of a bar chart with reference lines

Pie charts

Although pie charts are very commonly used in dashboards, they can be difficult to interpret because of the fact that the human eye has a difficult time estimating area and comparing angles. It is also difficult to compare pie slices that are not right next to each other or have similar values. Avoid pie charts with many slices due to these difficulties. Pie charts work best when there are limited components (five or less) and when including text and percentages in either a legend or on the pie itself so the dashboard user does not have to guess the meaning and value of each slice. Note that pie charts also take up quite a bit of real estate since they are round. Their round shape does not make them an ideal display type for effective use of space on dashboards.

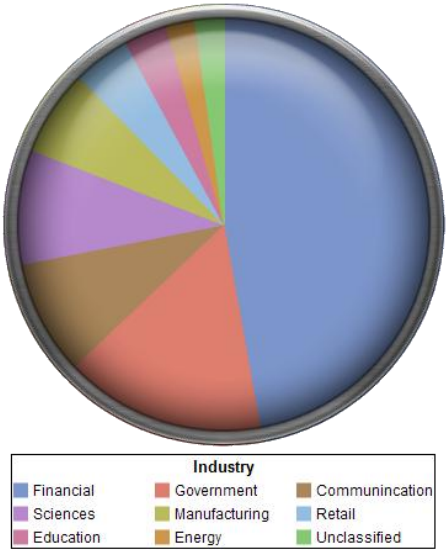

Figure 3. An example of a pie chart

Labels can be displayed on or near the slices, or in a legend as shown. Values should also be displayed with the labels. Notice how difficult it is to guess the values of the categories without the values shown with the labels.

COMMONLY USED DISPLAY TYPES

In addition to line charts, bar charts, and pie charts, bubble plots and scatter plots are also frequently used in dashboards. They work best when more than just a handful of data points are present, but at the same time, there is a fine balance between too few and too many data points that can be accurately represented by these display types. These plots are useful for identifying patterns in the distribution of data points. The following are cases when you should be using bubble plots and scatter plots:

Bubble plots

Bubble plots are most commonly used when the user has a need to define and compare data points across two dimensions and want to use the diameter of the bubbles or circles as a third dimension to represent the size of each data point compared against the two dimensions. An example of this is when the diameters of the bubbles are used to represent company size, the x-axis represents company growth, and the y-axis represents current company revenue. Use transparent bubbles when some bubbles overlap and all bubbles need to be visible. Use colored bubbles to group similar bubbles by categories. Avoid bubble plots when the bubbles start overlapping to the point where it is difficult to see individual bubbles or when bubbles are almost identical in size and position.

Scatter plots

Scatter plots should be used when many data points are present and there is interest in figuring out whether or not a correlation exists between the data points. For example, to figure out whether or not there is a correlation between the outside temperature and the number of forest fires in a given region, each forest fire could be plotted at the outside temperature at the time of the forest fire. When all data points are plotted, the user is able to either visually determine whether or not data points are related, or, can determine this by using an equation for the correlation between data points by using best-fit procedures. Note that this is not automatically done by simply creating the scatter plot; best-fit procedures have to be done after the plot has been created. Without best-fit estimates, scatter plots are still useful for getting a rough sense of how spread out or closely related data points are, and quickly identifying if patterns are present in the data point distribution. Avoid using scatter plots when you have less than five data points, as a bar chart or simple table might be more clear and effective in displaying the information.

Figure 4. An example of a scatter plot

LESS COMMONLY USED DISPLAY TYPES

Although the following display types are not extensively used in dashboards, they can definitely provide value to the dashboard user, if the right opportunity to use them exists, and if the dashboard audience is able to easily interpret them. Here is when these display types should and should not be used:

Scatter histograms

A scatter histogram is similar to a scatter plot with the addition of a horizontal and vertical histogram along the top and right side of the graph. The purpose of the horizontal and vertical bars is to show a count of the data points taken as columns and rows. Use a scatter histogram when it is necessary to display both individual data points *and* summaries of the points. An example of this might be to test scores for a class over the last six years. The individual test scores for each year can be displayed as data points on the graph itself, while the histogram displays the count of the scores in bar format.

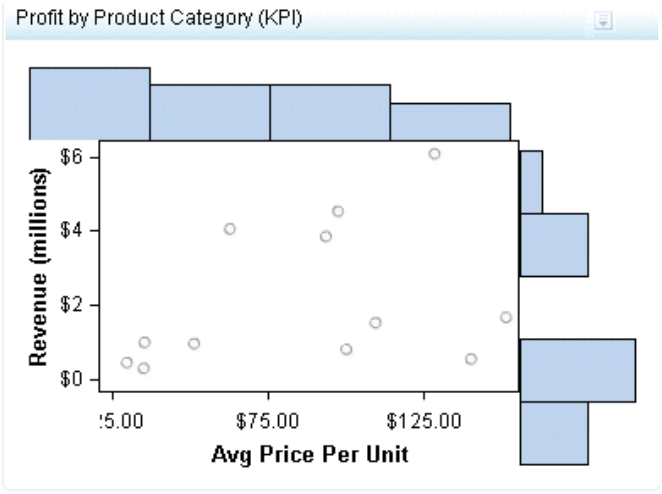

Figure 5. An example of a scatter histogram

Range maps

Range maps should be used when comparing data points in terms of position and range. The ranges on a range map are defined by range colors and border each other so that every part of the map represents a certain range. Use range maps when there are many data points in different ranges. For example, range maps can be used when

representing the body weight of males across different ranges: underweight, healthy, overweight, and obese. Each data point should fall into one of these ranges on the map. Range maps are preferable to bar charts with ranges when it is necessary to see how wide spread the data points are, in addition to how they compare to each other and what range they are in. Range maps are similar to scatter plots, but include colored ranges covering the chart, so that it is instantly apparent what category or range a data point is in.

Tree maps

Tree maps use shapes (usually rectangles) to represent data. Rectangular tiles vary according to size to represent a value for different categories. Because the size of each tile is proportional to the value, not to the category, it provides a clear overview of the status or contribution of each category. An example of this is when each tile represents rice production in different countries. The tile for Vietnam would be quite large compared to the tile for the US, even though the US is a larger country. Because of this difference, the dashboard user can clearly see who the top contributors are in a single map. Tree maps are therefore most effective when there is a need to display a lot of information in a small amount of space. The individual tiles can also be colored to represent ranges. Don't expect much detail to come out of a tree map, however. Tool tips and legends can reveal details of each tile, but the main point is for the dashboard user to get a rough sense of top contributors (good or bad), or the largest categories for a given map.

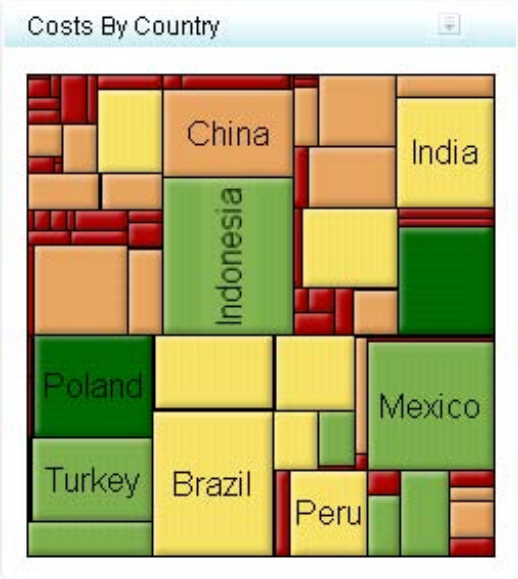

Figure 6. An example of a tree map

Vector plots

Vector plots should only be used when direction and magnitude (value) are both present in the data and have to be displayed. Unlike a line chart, a vector plot can show how two variables interplay with each other over a third variable which identifies each individual vector. Note that vector plots are less commonly used in business and more commonly used when dealing with physics and earth sciences. However, vector plots can be used to represent certain aspects of business. For example, a vector plot can be used to determine the direction of sales, given the resources and effort that went into making a sale. Vector plots will be much less familiar to dashboard users, so should be used when it is absolutely imperative that both direction and value are represented. They also take quite some time to analyze, so provide as much information about what the plot displays in the title as possible. Some additional labels or text will also help guide the user.

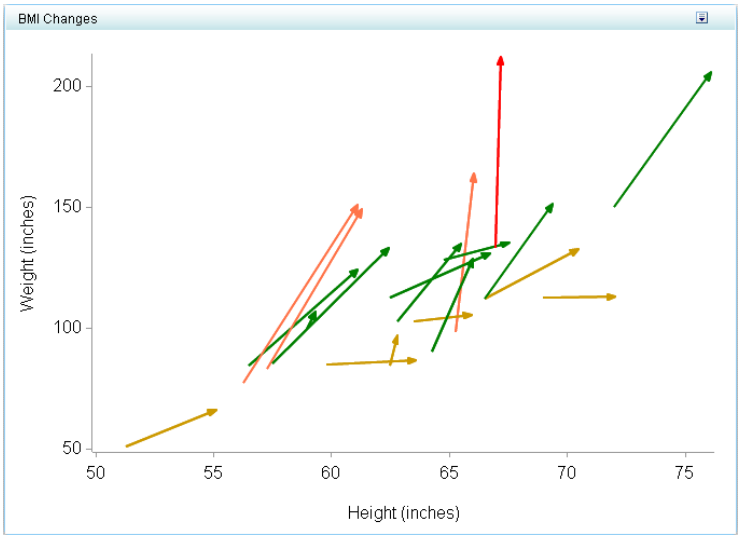

Figure 7. An example of a vector plot

Needle plots

Needle plots should be used when working with many data points that all extend from the same baseline. A common example of needle plot is the magnitude of earthquakes, which are represented by the needles in a needle plot. The horizontal axis usually represents time. Use a needle plot whenever there are too many data points to represent the data as a bar graph or scatter plot, and when it needs to be clear that each data point extends from the same baseline. Because many data points are usually present in needle plots, the individual needles sometimes become irrelevant, and the overall trend of the top of the needles becomes more apparent.

Waterfall charts

Waterfall charts are most often used when representing account or revenue data. They display how an initial value is increased or decreased by intermediate values. Use waterfall charts when details about how a value was increased or decreased need to be displayed in addition to the beginning and ending balances or values. Waterfall charts are most often related to finances and beginning and ending balances. They should not be used when the end results are the only items that are of importance. Use waterfall charts when working with a financialy savvy audience, or executives who need to be informed of how balances change over time.

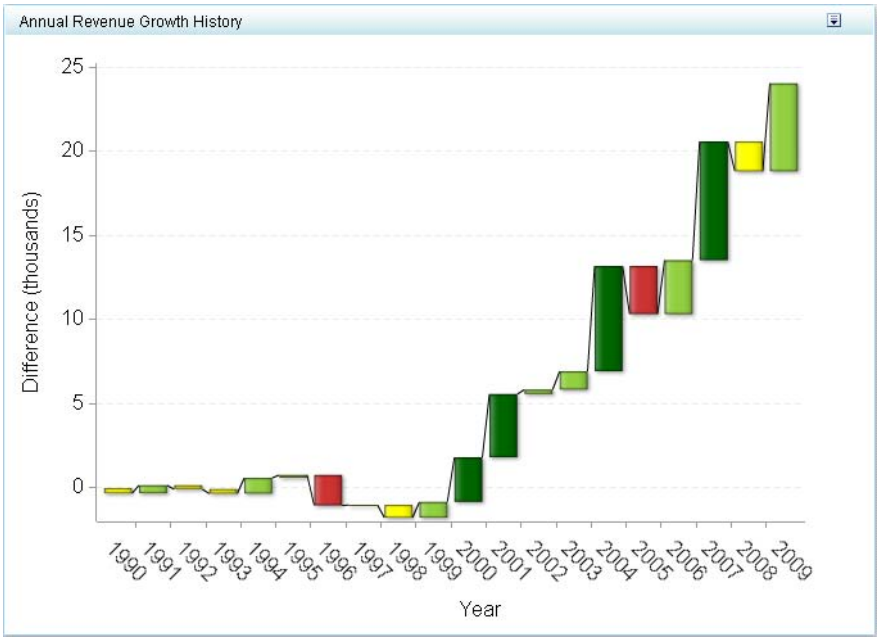

Figure 8. An example of a waterfall chart

GENERAL GUIDELINES FOR USING DISPLAY TYPES

In summary, display types only provide value if they are clear, accurate, and usable to the end user. To take advantage of the many different displays that can be used in SAS BI Dashboard 4.3, take the time to think about what the display type should reveal before selecting it for use in a dashboard. The following are some questions to ask yourself and to consider when choosing a display type:

1. What is the purpose of the display type?
2. What needs to stand out in the display type?
3. What data am I looking to display? Is it all or part of the data that I have?
4. How many data points am I working with?
5. How familiar is the dashboard user with the data that I am displaying?
6. How familiar is the dashboard user with the display type that I will be using? Will they be able to easily interpret it?

The answers to these questions, in conjunction with the information about each display provided above, should help you determine which display type is best for the message you are trying to convey.