perceptual edge

Introducing Bandlines

Sparklines Enriched with Information about Magnitude and Distribution

Stephen Few, Perceptual Edge
Visual Business Intelligence Newsletter
January/February/March 2013

When Edward Tufte originally introduced sparklines, he did not intend them specifically for dashboards. Tufte embedded no information in sparklines themselves about the magnitude of values, but illustrated two ways to incorporate this: 1) by labeling the first and last or the lowest and highest values in the data set, or 2) by using a horizontal band of light gray to mark the range within which values usually fell. The need for information about magnitude in sparklines themselves was somewhat mitigated by the fact that Tufte often displayed them in collaboration with other means of expression, such as by embedding them in text, which provided information about magnitude. When displayed in context of this sort, it is acceptable for magnitude information to be read (i.e., processed verbally) rather than scanned visually. When displayed in dashboards, however, if information about magnitude is needed, it must be incorporated visually into the sparklines themselves to be discerned efficiently.

The most common method that is used to add magnitude information to sparklines marks the highest and lowest value in the series with a simple dot and uses text to label those values, either in place or off to the side, as illustrated below:

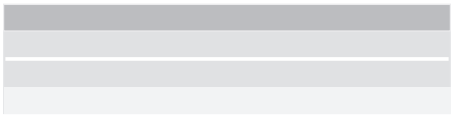
This approach is not ideal on a dashboard, however, because text requires reading, which involves verbal processing in the brain, which is slower than visual processing. This slows comparisons of magnitude among a series of sparklines to a crawl. To support rapid monitoring in a dashboard, we must encode the magnitude information using visual elements. For this purpose, I've created a version of sparklines that I call *bandlines*.

Bandlines use horizontal bands of color in the background of the plot area to display information about the distribution of values. This information is similar to that found in a box plot. To do this you must gather information about how values related to the measure that will be featured in the sparkline are distributed during a period of history that usually extends further into the past than the values that will appear in the sparkline itself. For example, for a sparkline that will display 30 time-series values—one for each of the last 30 days—you might gather information about how its values were distributed during the last six months or year. You need five primary facts about a set of time-series values to construct a bandline: 1) the lowest value, 2) the 25th percentile (i.e., the point at and below which the lowest 25% of the values reside), 3) the median (a.k.a., the 50th percentile, the point at and below which 50% of the values reside), 4) the 75th percentile (i.e., the point at and below which 75% of the values reside), and 5) the highest value. With these five values to mark the boundaries, you've divided the full distribution of values into four ranges, called quartiles (quartile 1 at the bottom through 4 at the top), which each contain one quarter of the values.

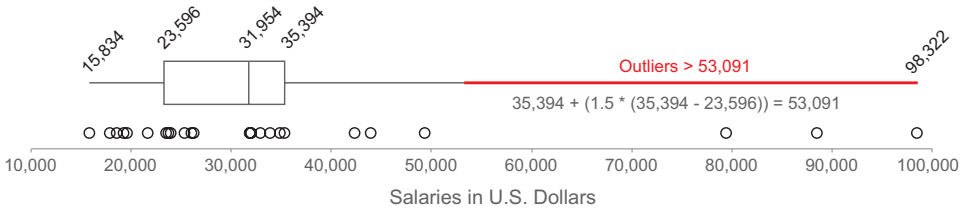
The quartiles are represented visually in bandlines using horizontal bands of color in the background. Three shades of a single hue, usually gray, in sequence from light to dark are used to divide the full spread of the

distribution into three sections: the lightest for the first quartile, a medium shade for the midrange (a.k.a., the interquartile range, consisting of the second and third quartiles), and the darkest for the fourth quartile.

Within the midspread band of color, a horizontal white line is used to mark the median's position, which divides the midspread into the second and third quartiles.



And finally, within the first and fourth quartiles, if any values qualify as outliers, they are individually highlighted. Outliers are determined just as they are in box plots. A distance of 1.5 times the midspread above the 75th percentile and below the 25th percentile marks the points beyond which values qualify as outliers.



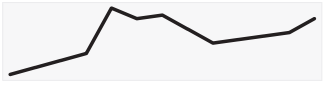
Low outliers are identified by solid black data points (dots) along the line and high outliers are identified by white data points with a thin black border.

The following features of the time series are revealed in bandlines:

Feature	Visual Representation
Trend	Overall direction of change (mostly up, relatively flat, mostly down)
Pattern of Change	Varying slopes along the line
Magnitude of Values	A combination of vertical position of values along the line and background fill color
Magnitude of Change	A combination of the slope of a line segment and in some cases the line's migration from one band of background color to another
Variability	Variation of vertical positions along the line and in some cases the line's migration from one band of background color to another

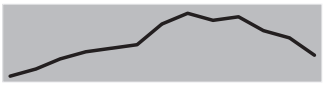
The following examples (below and on the next page) illustrate how bandlines appear and are interpreted:

Values are spread across all quartiles. This time series trends upwards and spans an extensive range of values.



Values all fall within the bottom quartile. This time series trends upwards within a narrow range of low values.

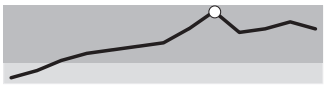
Values all fall within the mid-spread: some above and some below the median. This time series trends downwards and spans the middle two quartiles.



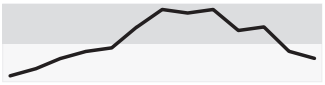
Values all fall within the top quartile. This time series trends upwards slightly within a narrow range of high values.

Values all fall within the bottom quartile and one is a low outlier. This time series trends upwards within a narrow range of low values.

Values all fall within the top quartile and the last two are high outliers. This time series trends upwards within a narrow range of high values.



Values fall within the midspread (slightly) and top quartile and one is a high outlier. This time series trends upwards across a larger than usual range of mostly high values, one of which is extreme.



Values fall within the midspread and bottom quartiles in roughly equal proportions. This time series trends up slightly but remains among the lower half of values.

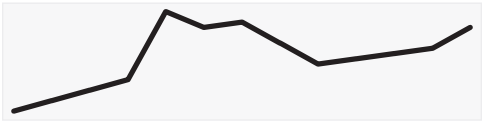
Notice that the bandlines in these examples were all scaled to fill the vertical space: in each the lowest value is positioned at the bottom and the highest at the top of the plot area. This was done to provide the clearest possible picture of the trend and pattern of change in each bandline. When more precise magnitude comparisons are required across a set of related time series, a set of bandlines may all share the same quantitative scale, just as we often do with sparklines. However, when a common scale is used and values vary significantly in magnitude, with some items having much greater values than others, lines that represent relatively low values often look flat, which hides information about patterns and trends of change. One of the advantages of bandlines over plain sparklines is the fact that they may be independently scaled to fill the vertical space of the plot area to optimally feature patterns and trends of change but provide information about magnitude that will allow the person viewing them to differentiate variability within a small range from greater variability.

Basing the bands on quartiles causes the ranges that they represent to vary in the quantitative distances that they span. The full spread of values is not divided into four equal quantitative ranges, but rather into ordinal groups, each consisting of 25% of the values. If the line extends equal distances into two of the quartiles, this does not necessarily mean that its values ventured in equal proportions into those two ranges, because the ranges can differ in quantitative distance. For example, the bandline below is fully contained within the midspread and appears to extend roughly as far into the third quartile (above the white median line) as it does into the second quartile (below the median). The lower boundary of the second quartile at the bottom and the upper boundary of the third quartile at the top are not visible, however, so we can't tell if one covers a greater quantitative distance than the other. If the second quartile extends across a greater distance than the third, then the proportion into which the sparkline ventures into the third quartile is greater.

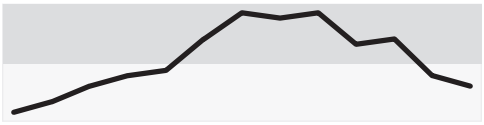
Although this might strike you as a problem, it provides useful information. A line that appears to extend equally into two ranges does in fact extend equal magnitudes into them, which users will understand intuitively, but by

delineating quartiles rather than equal quantitative distances, we gain information about the values contained in a bandline relative to the norm, which is especially useful when monitoring performance.

Despite the useful information about magnitude that is provided by bandlines, scaling to fill the vertical space sometimes comes with a cost. Restricting information about magnitude to a maximum of six distinct distribution ranges—low outlier, first quartile, midspread (divided into second and third quartiles by the median), fourth quartile, and high outliers—will fail to provide adequate precision for some uses, especially when a time series resides entirely within a single quartile. For example, by viewing the following bandline, we know that all values are in the bottom quartile and that none qualify as outliers, but we can't discern if the values vary only slightly or greatly within the quartile.

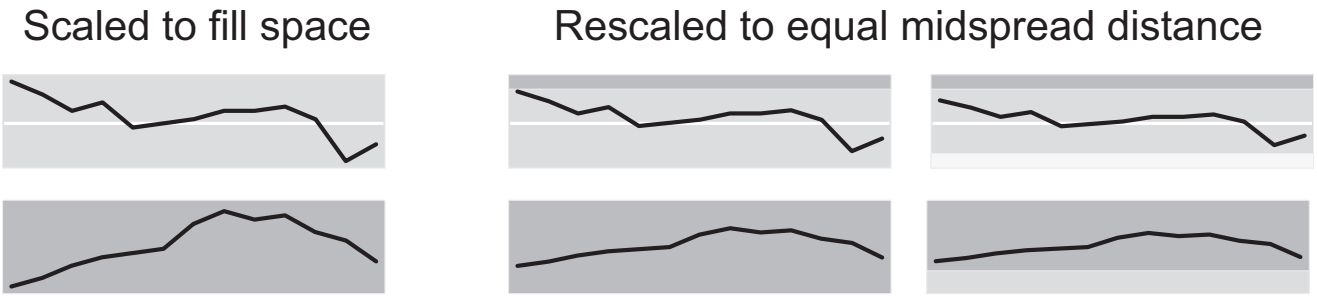


Even when the line spans two quartiles, as it does below, we know that values ventured roughly equal distances into the first and second quartiles, but nothing about the degree. It's possible that values varied only to a slight degree near the 25th percentile (i.e., the boundary between the first and second quartile).



For most performance-monitoring purposes, knowing the quartiles in which values reside is adequate information about magnitude, but we must sometimes provide greater precision. The trick is to do this in a way that doesn't complicate interpretation or slow down perception. This can be achieved in two ways.

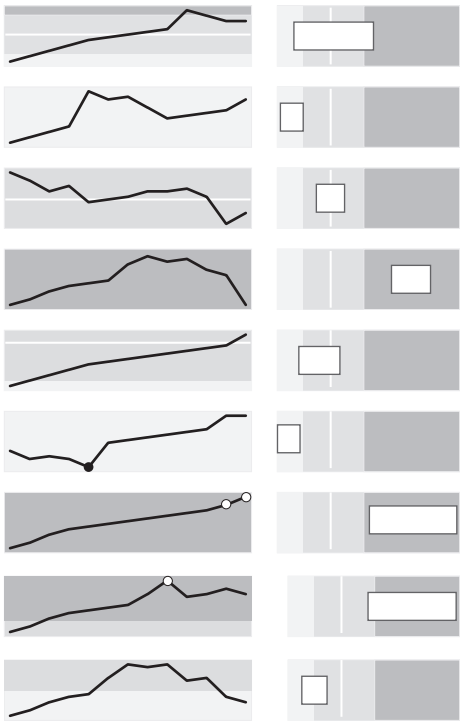
The first method modifies the scale of the bandline by filling the vertical space only when values span a quantitative distance that equals or exceeds the distance of the midspread. When they span a smaller distance, the scale is increased to equal the distance of the midspread. This does not mean that the entire midspread will always appear in the plot area of the bandline, but that the scale from bottom to top will equal the same distance that lies between the bottom and top of the midspread. The two bandlines below on the left, if scaled in this manner, could look like either of the two alternatives on the right:



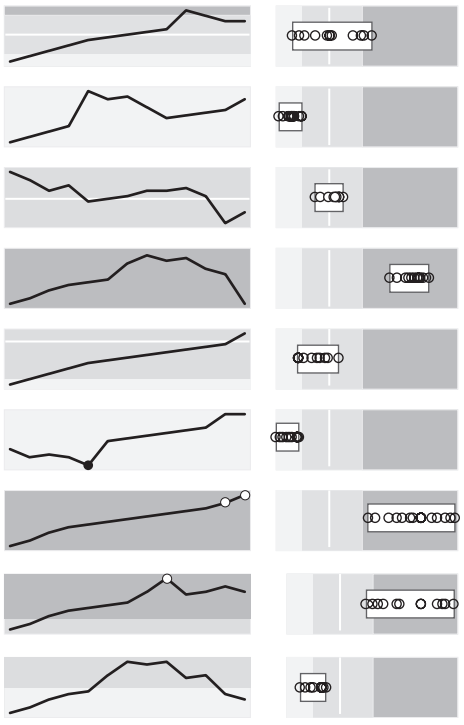
By using this scaling method, time series that vary little in value will be prevented from having the exaggerated appearance of significant variation. Lines that represent little variation will appear more flat and will not fill the vertical space.

The other method is more informative, but it requires additional space. The information contained in bandlines may be supplemented by displaying range bars alone or a combination of range bars and strip plots next to

them to provide distribution information in a way that is easier and faster to perceive and compare. I call these *sparkstrips*. Here's a version next to the set of bandlines that we've seen before, which uses range bars alone.



When this is done, the bands in bandlines need not span a midspread of space because the magnitude information that this method strives to reveal is provided more effectively in the sparkstrips. They are always consistently scaled to display the full range across which the values are historically spread, not just the range represented by a particular bandline. This provides an efficient way to compare the distributions among bandlines, plus information about magnitude and variability as well. By displaying strips plots (a set of dots to mark each value in the series) on top of the range bars, richer information about the distribution's shape is provided without overcomplicating the sparkstrips, illustrated below.



Bandlines may be modified to represent other meaningful ranges besides quartile-based distributions. In the Student Performance Dashboard below, I used the colored bands to represent grade-based ranges of performance (A, B, C, D, and F) because these ranges are more meaningful to most teachers who would use this dashboard.

(Click here to view a full-sized version of the dashboard.)

When I first introduced bullet graphs in 2005, no products could support them at that time, but this changed quickly as software vendors recognized their value. Although some products might already provide the programming flexibility needed to produce bandlines and sparkstrips, none support them as ready-made charts. All vendors are welcome to add this functionality, however, without restraint. Because I have copyrighted the content of this publication to establish bandlines and sparkstrips as my extension of sparklines, but have not restricted their use through patent protection, anyone may implement them but no one may prevent others from copying the design.

Discuss this Article

Share your thoughts about this article by visiting the Introducing Bandlines thread in our discussion forum.

About the Author

Stephen Few has worked for nearly 30 years as an IT innovator, consultant, and teacher. Today, as Principal of the consultancy Perceptual Edge, Stephen focuses on data visualization for analyzing and communicating quantitative business information. He provides training and consulting services, writes the quarterly *Visual Business Intelligence Newsletter*, and speaks frequently at conferences. He is the author of three books: *Show Me the Numbers: Designing Tables and Graphs to Enlighten*, Second Edition, *Information Dashboard Design: The Effective Visual Communication of Data*, and *Now You See It: Simple Visualization Techniques for Quantitative Analysis*. You can learn more about Stephen's work and access an entire library of articles at www.perceptualedge.com. Between articles, you can read Stephen's thoughts on the industry in his blog.