Wrapping Graphs to Extend Their Limits

Stephen Few, Perceptual Edge
Visual Business Intelligence Newsletter
July/August/September 2013

Some of the graphs that we rely on to examine and compare quantitative data are unnecessarily constrained in the number of values that we can simultaneously view. For example, it would ordinarily be difficult to display more than 100 bars or so in a horizontal bar graph on a typical computer screen. Occasions when it would be useful to exceed this limit are increasingly frequent, however. This limitation was in part what motivated Ben Shneiderman to invent the treemap many years ago, but treemaps are only appropriate for part-to-whole data sets and they surpass the limits of a bar graph by encoding values in ways that are less effective perceptually—rectangle sizes and colors—than the lengths of bars that share a common baseline. So, we could benefit from a way to extend the number of values that can be displayed in a bar graph well beyond the limits of their current design. In this article, I'm proposing an enhanced version of a bar graph that allows the number of values that can be simultaneously viewed to increase significantly. I call it a *wrapped bar graph*.

The capacity of a horizontal bar graph can be dramatically increased by splitting it into multiple columns and wrapping a series of bars across them, first filling the leftmost column from top to bottom, and then the next, and so on. This wrapping technique is common in textual displays, such as in the wrapped columns that appear in newspapers and magazines.

In the following example of a wrapped bar graph, a graph that would ordinarily be limited to 63 visible bars (without scrolling) has been expanded in capacity to 315 bars by splitting the graph into five columns of 63 bars each.
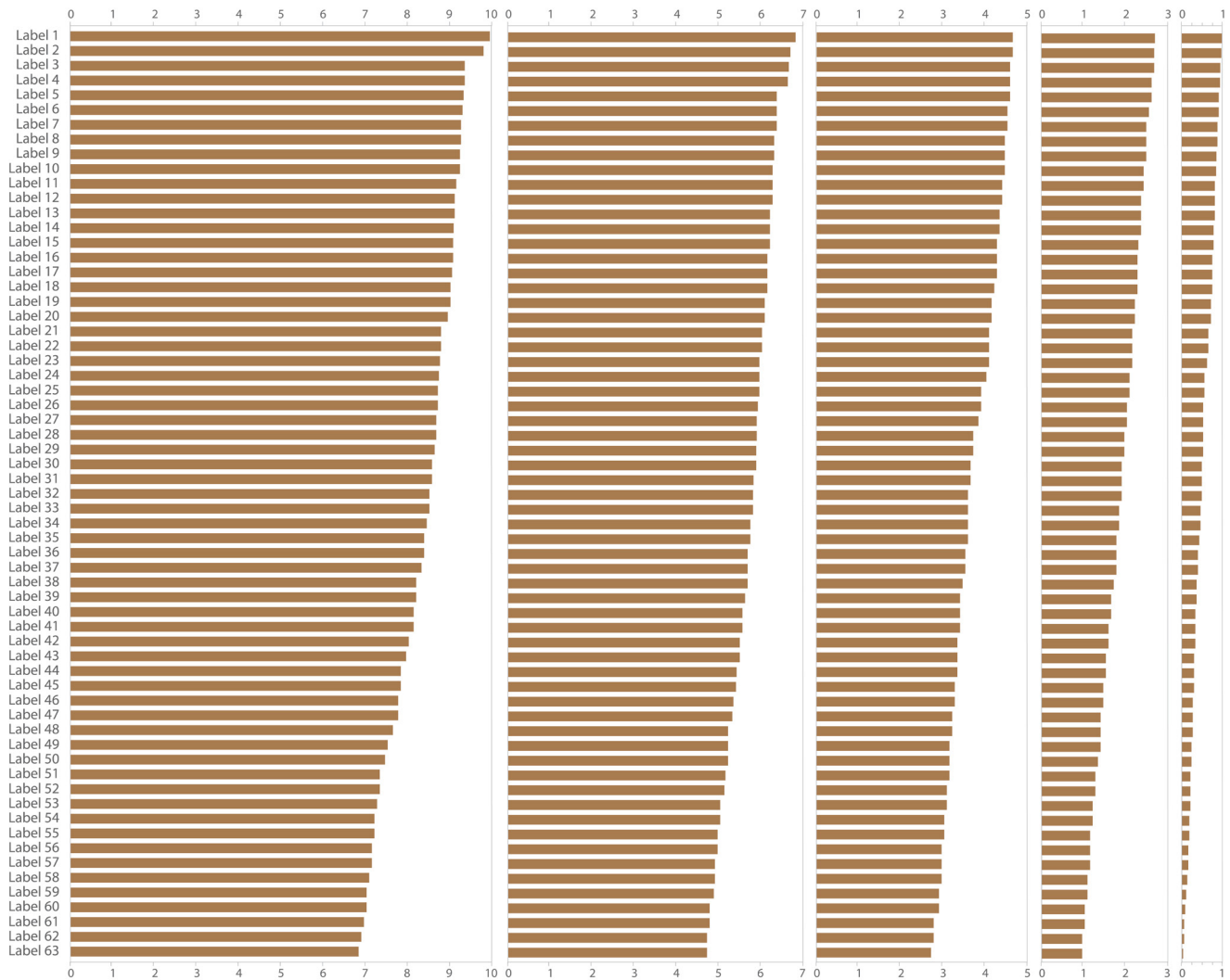
Horizontal screen space has been optimized by sorting the bars in order from largest to smallest and limiting the quantitative scale associated with each column to the range that's needed for the values in that column. Notice that the scale extends to a value of ten in the leftmost column but only to a value of seven in the second column, and so on. Because the distance between intervals along each scale remains consistent, accurate comparisons can be made between bars in different columns, even though the columns are different widths.

With a wrapped bar graph, to further optimize horizontal screen space, only one column of bars would ordinarily be labeled at any one time, defaulting to the leftmost column. To see the labels in other columns, one of two methods can be used: 1) hovering over a bar with the mouse to access its label and precise value in a tooltip, or 2) selecting a particular column of bars to cause its labels to be displayed and the column that had been labeled to collapse into a non-labeled state, as illustrated below.
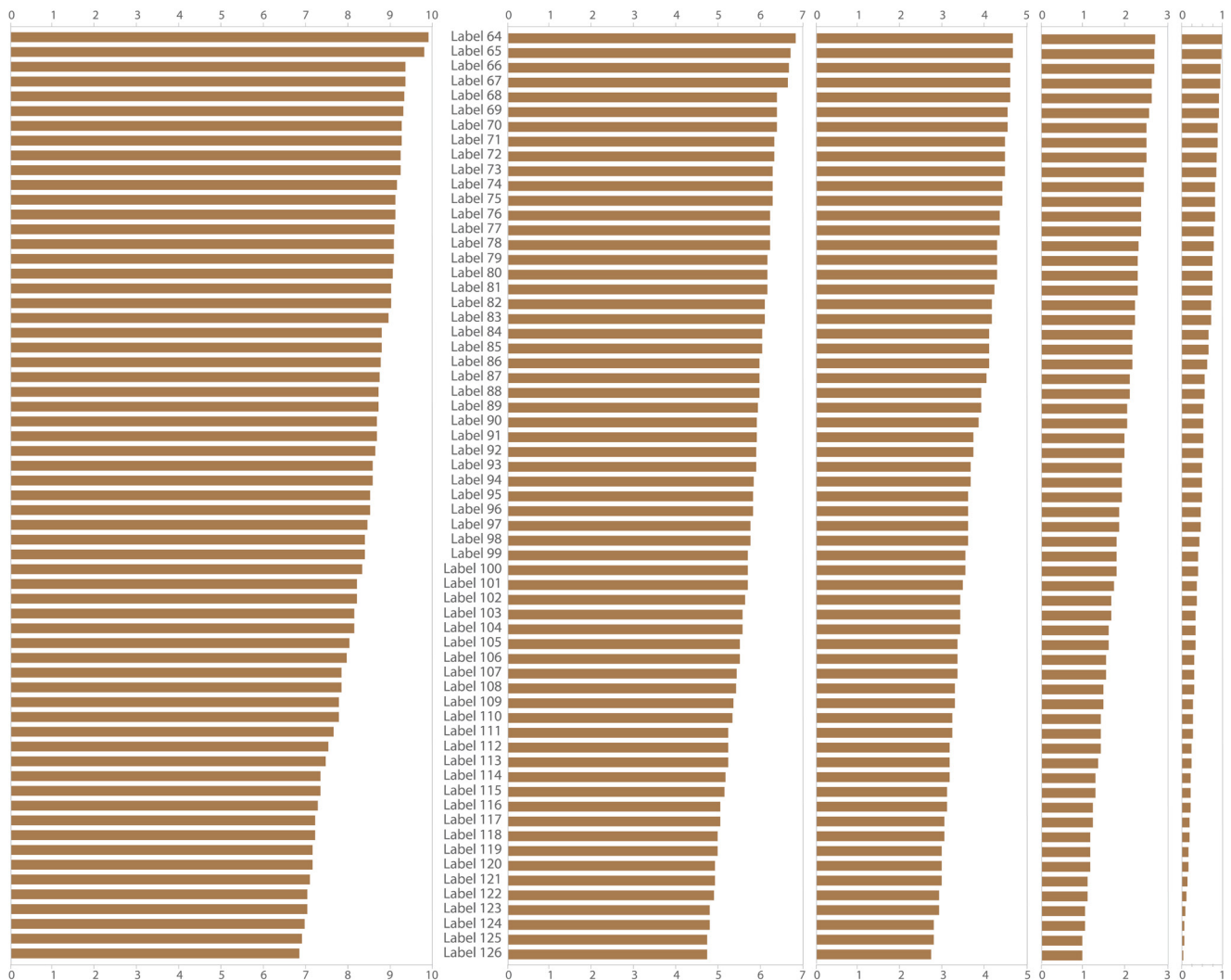
Because there are times when we would want to simultaneously identify and compare bars that are not located in the same column, additional functionality could be provided to search for particular items (e.g., bars representing the countries "France" and "Italy"), so they could either be highlighted in place or all other values could be filtered out of the display. Other means of filtering could be enabled as well. Filtering data could be handled by control-clicking or shift-clicking to select the bars that we wish to see and then clicking a filter control to cause all unselected bars to disappear, or vice versa. Remaining bars would then be rearranged to optimally fill the space. To accomplish this, the number of columns could decrease and the width of the bars could increase dynamically, based on an algorithm that maintains optimal perceptibility based on the screen's dimensions and resolution.

An implementation of wrapped bars should incorporate an algorithm for automatically determining the number of columns and bar widths that will optimally support comparisons. In the following example, forcing a set of bars across five columns that could have been displayed in a single column would undermine our ability to easily compare the values.
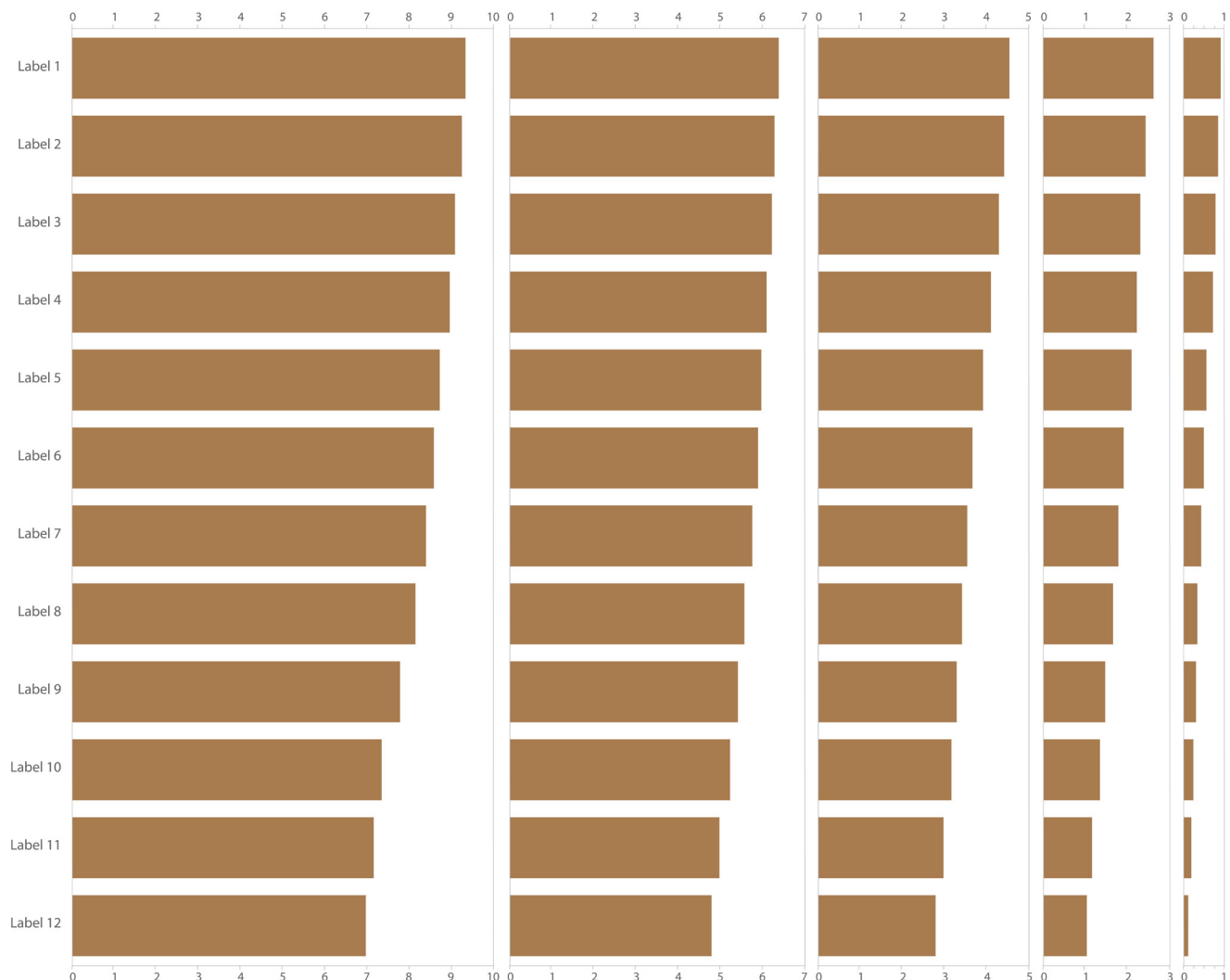
The option to turn off column wrapping should always be no more than a click away, which would automatically turn on scrolling if necessary to fit all of the bars into a single column. Ideally, wrapped bar graphs should not be provided as a separate type of graph from a normal horizontal bar graph, but as an option that could be easily turned on or off as needed.

Additional quantitative context could be displayed in the background of the wrapped bar graph to support easier comparisons between bars that are located in different columns. Something that I have done with normal bar graphs and also with bullet graphs in the past—including information about how the full set of values are distributed—lends itself to this situation as a useful solution.

In the following example, I've added information about the full distribution of values in the graph using three vertical references lines. Starting from the left, the first reference line marks the data set's 25th percentile (i.e., that point at and below which the lowest 25% of the values reside), the second slightly thicker line marks the median value (a.k.a., 50th percentile), and the third line marks the 75th percentile (i.e., the point at and above which the highest 25% of the values reside).
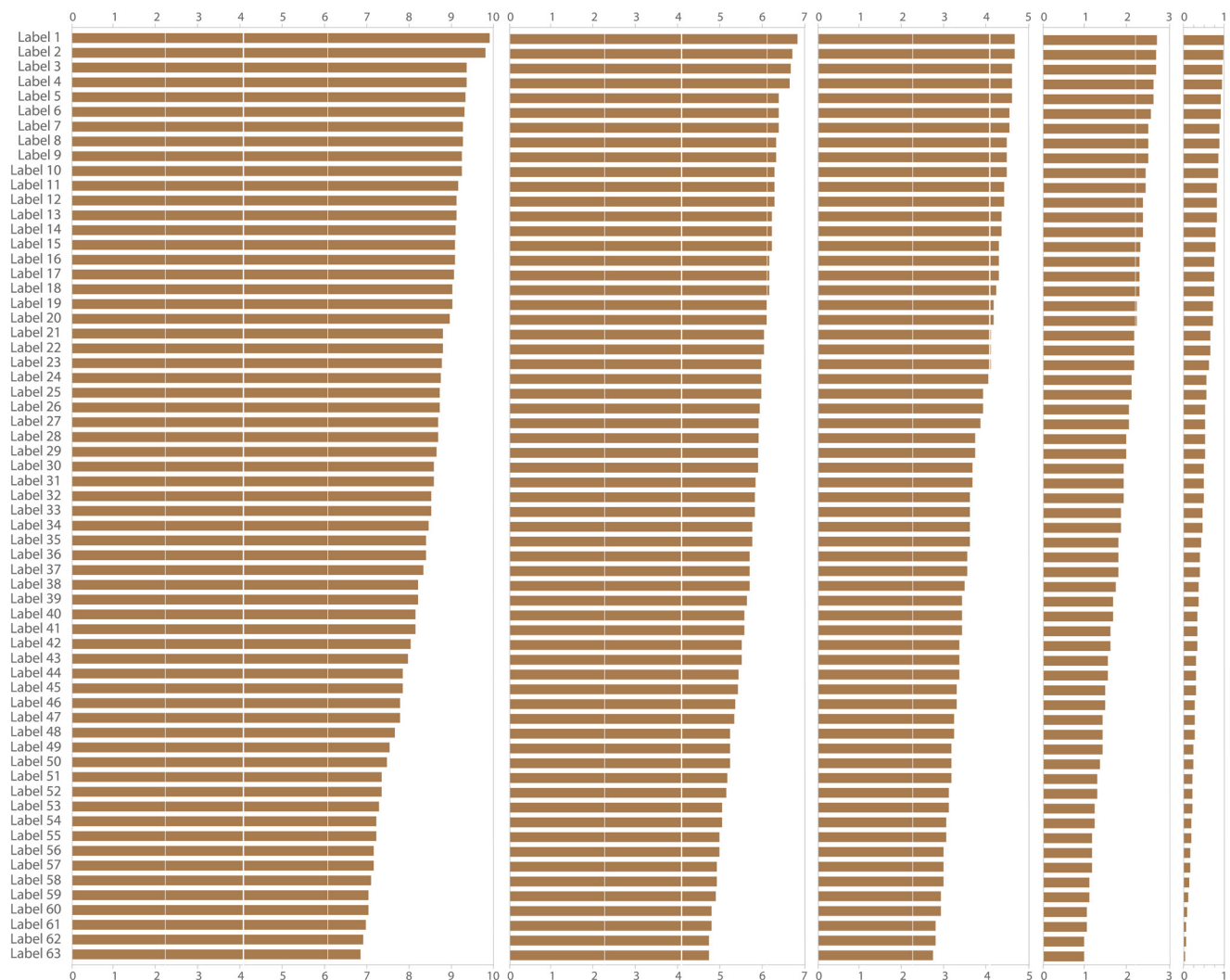
Notice that columns of bars other than the leftmost display less and less of the full range across which values are distributed, such that the values in the rightmost column are all contained in the first quartile (i.e., the range where the lowest 25% of the values reside). Without overcomplicating the graph, this contextual information about the data set's distribution makes it easier to compare bars to one another, especially when they aren't located in the same column. It also makes it easy to see how each value relates to the entire set.

Vertical reference lines that appear on top of the bars are not the only way that distribution information can be displayed. I experimented with several potential methods. I decided to make the reference lines only visible in the foreground of the bars rather than as continuous lines that appear behind the bars because the latter approach did not work as well. In the following example, you can see that the reference lines don't stand out clearly enough and they create a distracting visual effect where the lines peek out between the bars.
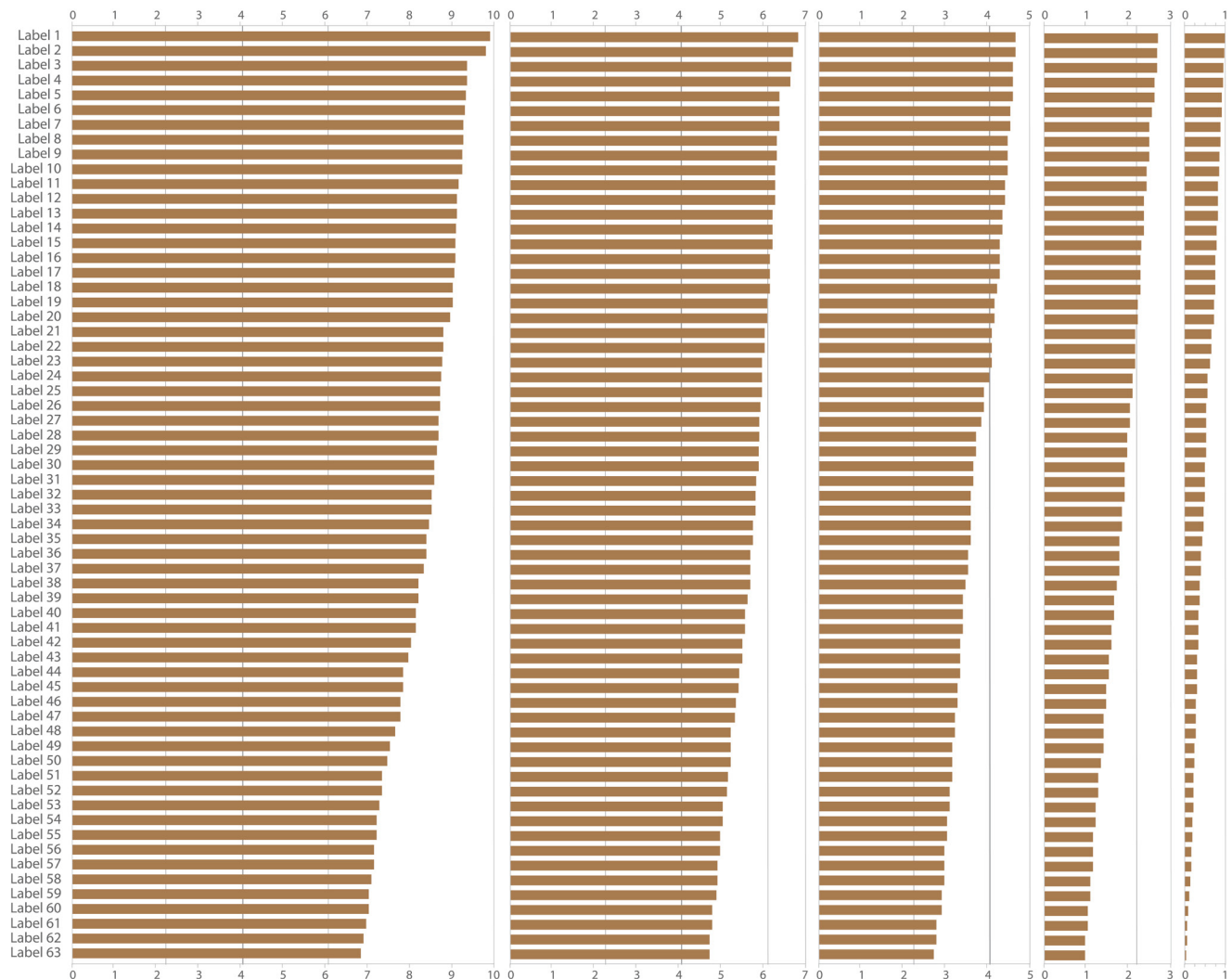
These problems could perhaps be eliminated by increasing the space between the bars, but this would limit the number of bars that could be displayed—not an acceptable compromise. Displaying continuous gray lines in the foreground didn't work any better. When reference lines are used to mark the 25th percentile, median, and 75th percentile, light lines that appear in the foreground seem to produce the best result.

I also experimented with the combination of a shaded area to mark the range between the 25th and 75th percentiles (a.k.a. the midspread) and a white line to mark the median, both positioned behind the bars in the background, but this produced a visual effect that looked a bit like venetian blinds, which I and others found annoying (see below).
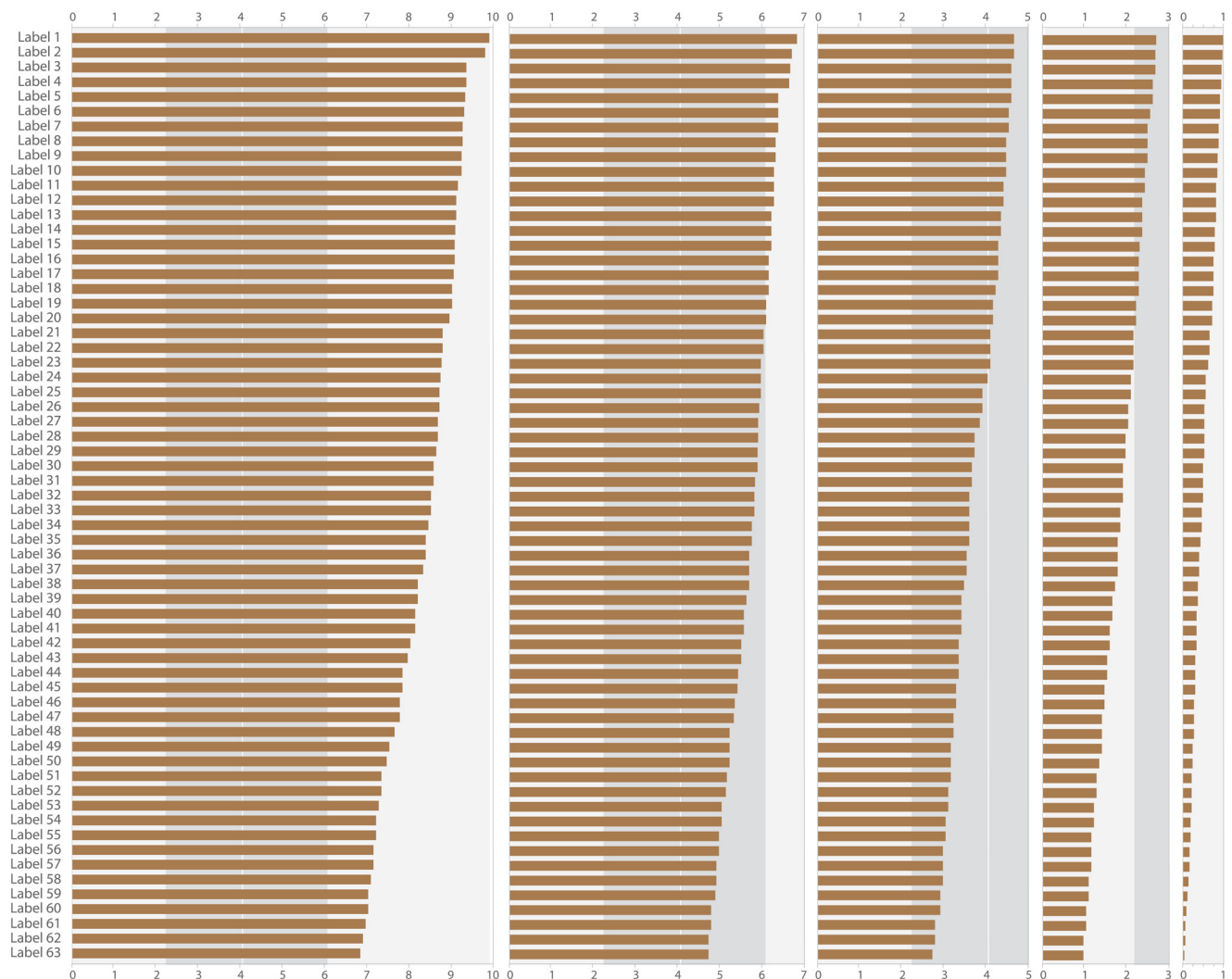
For this reason, I gave up on this approach, but then Maureen Stone, who now works in Tableau Software's research lab, pointed out that this annoying visual effect could be eliminated by overlaying the shaded region and line on top of the bars, as illustrated on the next page.
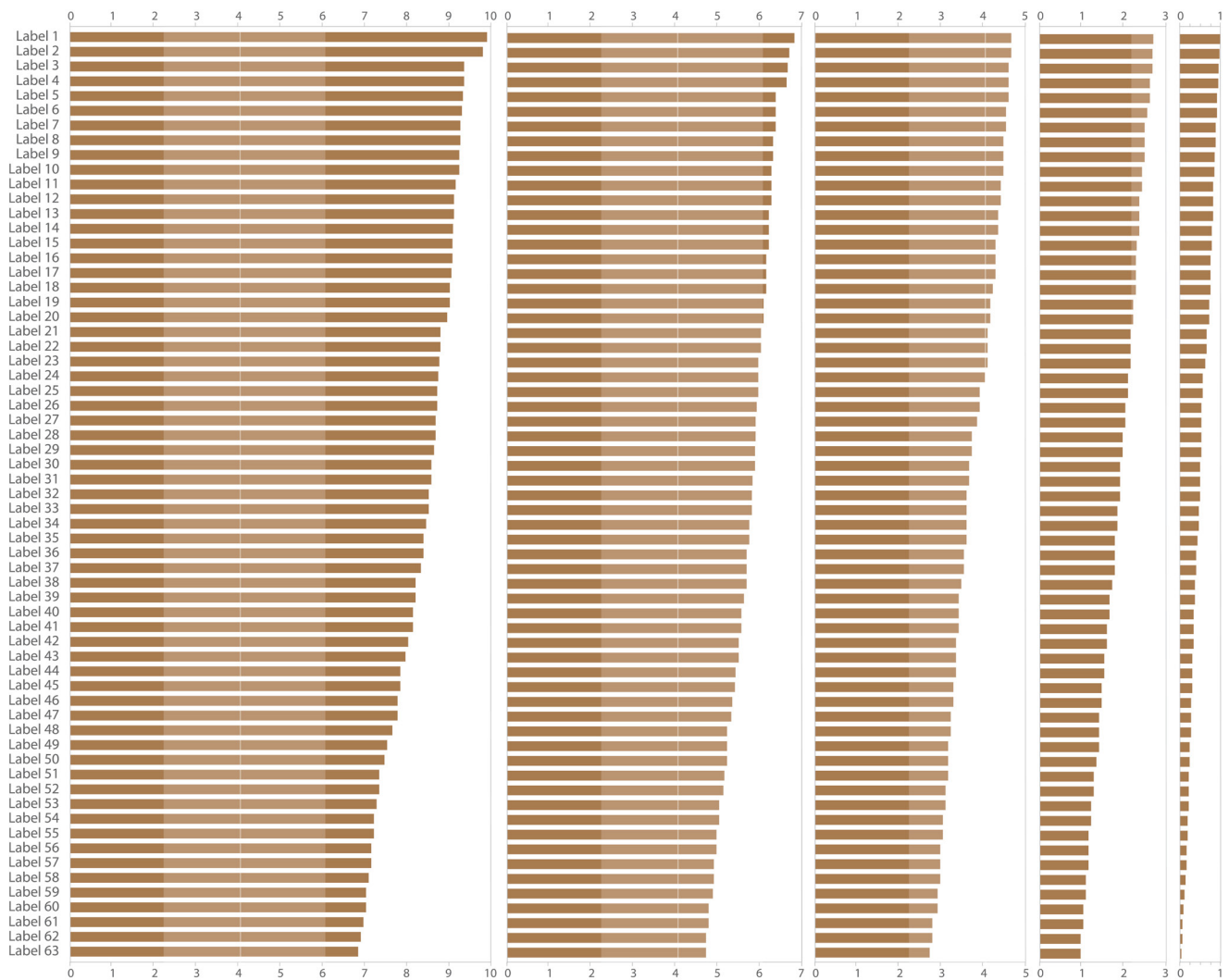
In some respects this works better than three reference lines, for it makes the midspread easy to see and displays it more intuitively as a range of values and not just two independent points for the 25th and 75th percentiles. You'll see in a moment, however, that this method won't work for all occasions.

Just as treemaps can display two variables simultaneously—one quantitative variable using rectangle sizes and second variable, either quantitative or categorical, using colors—bars in a wrapped graph can display a second variable using color variation as well. The second variable can be categorical, in which case distinct hues would be used (e.g., distinct colors for geographical regions) or variation in color intensity can be used to encode a quantitative variable. The following example illustrates the use of color intensity from light to dark to encode a second quantitative variable. (Only the upper portion of the graph is shown here and in a few examples that follow to save space.)
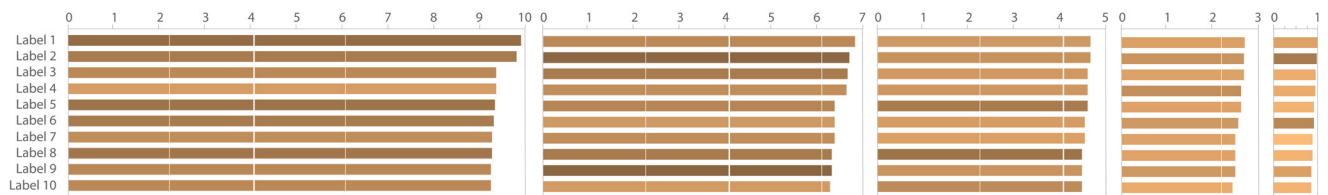
I used vertical reference lines to display distribution information in this example, because when variation in color intensity is used to display information, using a lighter color for the midspread causes the quantitative information encoded as color intensity to be altered in this region, which complicates comparisons and introduces unnecessary visual complexity, as shown below.
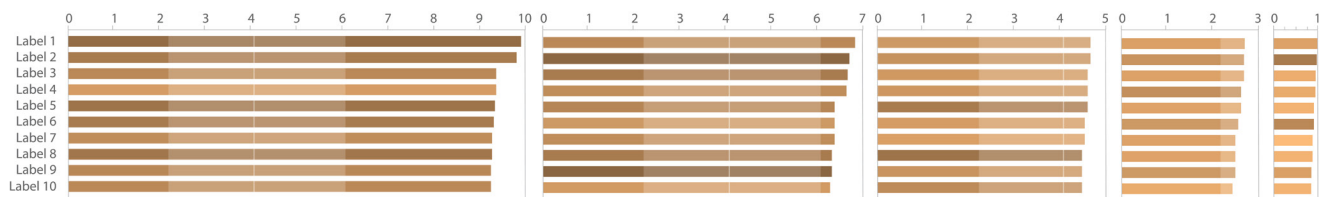
When distribution information is not included in a wrapped bar graph to support easier magnitude comparisons, grid lines may be included to lend extra precision when interpreting a bar's magnitude or when comparing bars. Just as reference lines worked best when displayed in the foreground of the bars rather than as continuous lines, the same is true of grid lines. In the example below, vertical grid lines mark the position of each value along the quantitative scale and have smaller intervals for greater precision in rightmost column.
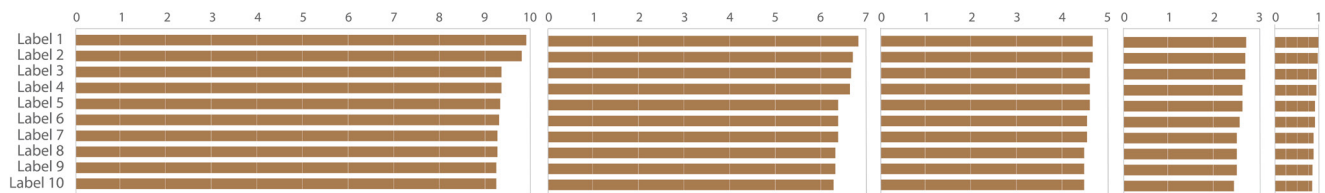

Wrapped graphs need not be limited to bars. Dot plots are often useful for displaying large sets of discrete values, especially when we don't want to start the quantitative scale at zero, which is required whenever bars are used. In the example below, all values fall within the 90 to 100 range, so by narrowing the scale to begin at 90 and end at 100, we can more easily see differences than we could if bars that extended from a base value of zero were used instead.
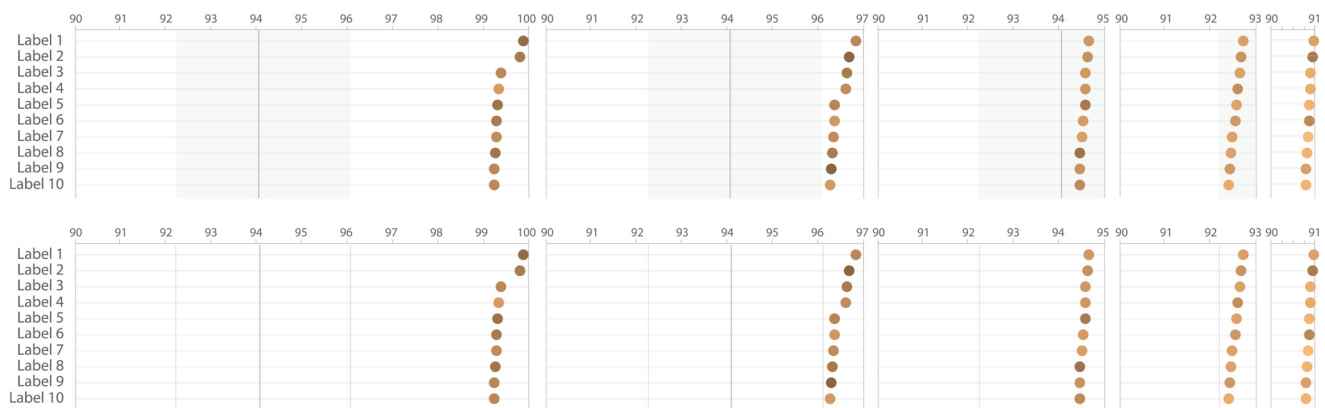
When dots are used, either shaded ranges or reference lines can be shown in the background to provide distribution information. Notice that wrapped dot plots appear less cluttered than wrapped bar graphs. Because bars are much larger objects that fill most of the screen, they have more visual weight than dots, which creates a busier, more crowded appearance. For this reason, wrapped dot plots will often be preferable to wrapped bar graphs even when the quantitative scale begins at zero. Both have their uses, however, so ideally, it should be possible to easily and quickly shift between a wrapped bar graph and a wrapped dot plot at any time. When shifting to a wrapped dot plot, if the lowest value in the data set is far from zero, the quantitative scale should be automatically be narrowed to begin slightly below the lowest value.

Wrapped bar graphs and dot plots can accommodate both positive and negative values simultaneously, just as normal versions of these graphs can do, while still optimizing horizontal space. In the following example, notice that where negative values appear in the rightmost column, the quantitative scale only goes as far as necessary in both the positive and the negative direction to accommodate the values.
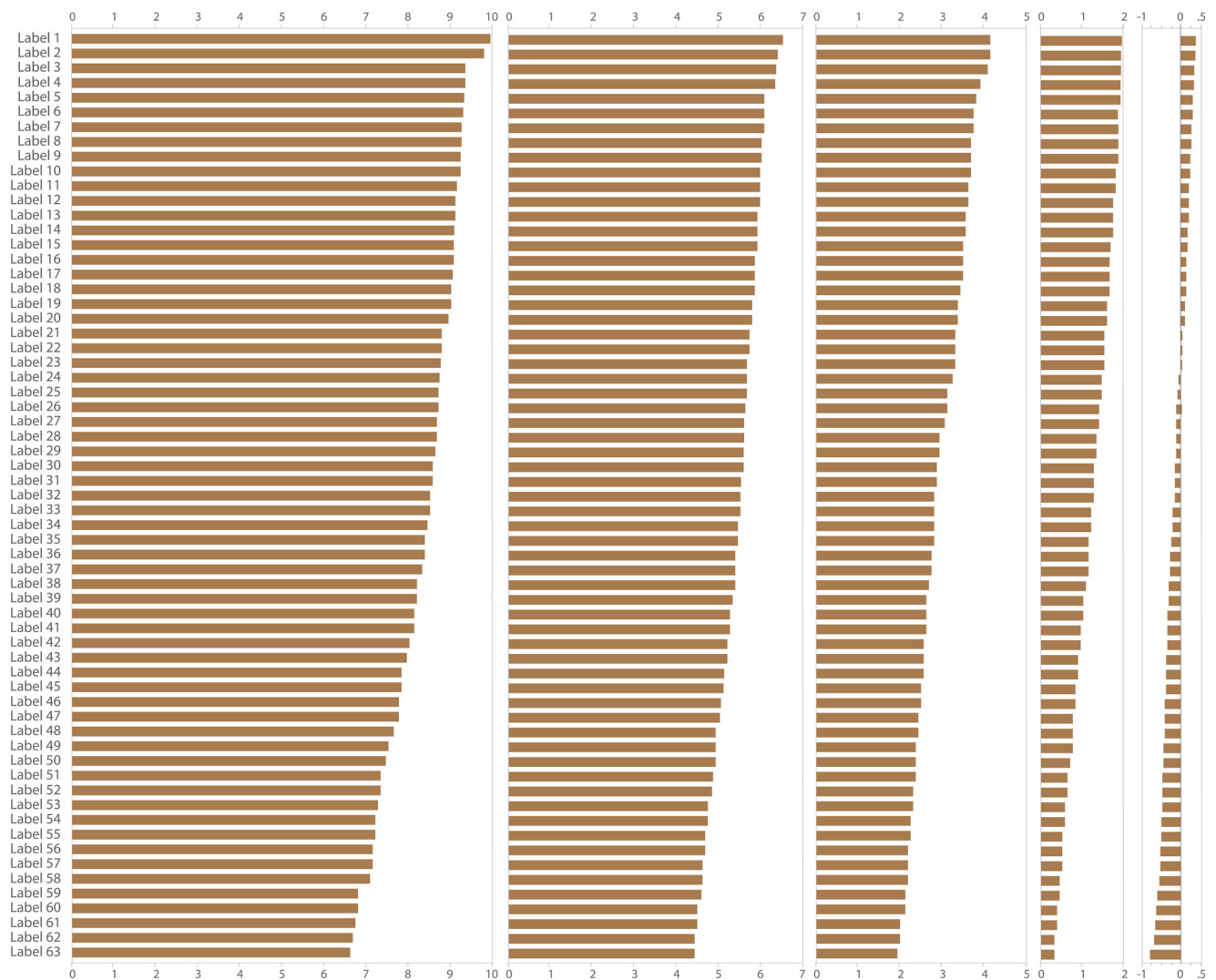
A wrapped graph could be arranged with the quantitative scale running vertically on the Y-axis rather than horizontally on the X-axis (e.g., using vertical rather than horizontal bars) with the values wrapping across a series of rows rather than columns, but the labels would need to oriented vertically, which would make them difficult to read (see below).
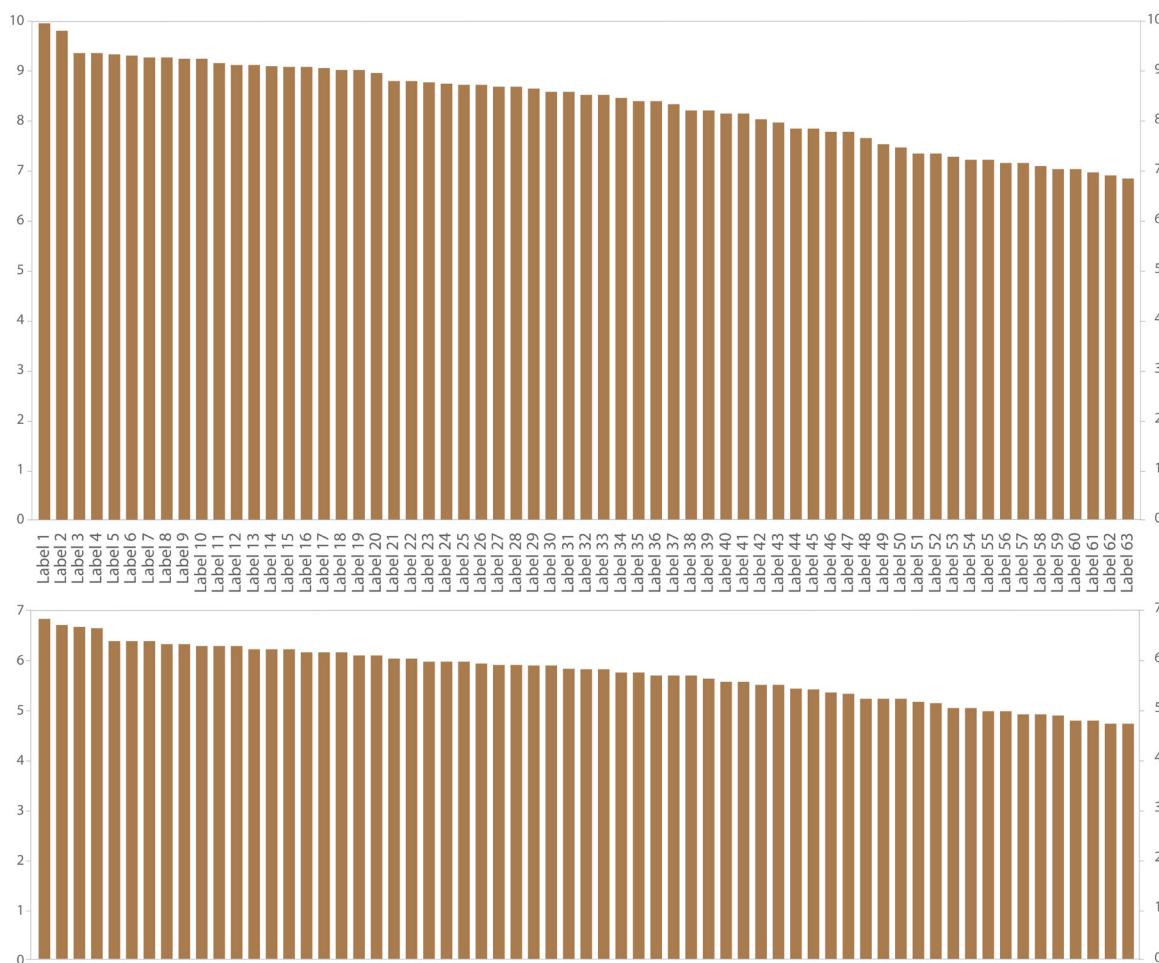
This arrangement offers no advantage over the other, except when displaying time series values, which work best when arranged horizontally from left to right. It wouldn't make sense, however, to sort time-series values from highest to lowest. With this in mind, I'm inclined to skip this arrangement altogether.

Wrapped bar graphs and dot plots can accommodate a large number of values; not as many as a treemap, of course, but a significant increase over regular bar graphs and dot plots. As such, they can accommodate cases that fall between normal bar graphs or dot plots and a treemap. And, unlike a treemap, wrapped bars and dots can be used when the values don't represent parts of a whole. This simple extension to the design of bar graphs and dot plots could expand their usefulness and thereby increase the insights that they potentially present to our eyes. If you agree, encourage your software vendor to support them.

Discuss this Article

Share your thoughts about this article by visiting the <u>Wrapping Graphs to Extend Their Limits</u> 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 <u>library</u> of articles at <u>www.perceptualedge.com</u>. Between articles, you can read Stephen's thoughts on the industry in his <u>blog</u>.