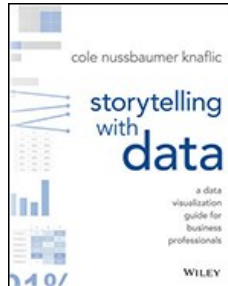


Chapters *To Go*



Storytelling With Data: A Data Visualization Guide For Business Professionals

by Cole Nussbaumer Knaflic
John Wiley & Sons (US). (c) 2015. Copying Prohibited.

Reprinted for YI LIN, CVS Caremark

yi.lin@cvscaremark.com

Reprinted with permission as a subscription benefit of **Books24x7**,
<http://www.books24x7.com/>

All rights reserved. Reproduction and/or distribution in whole or in part in electronic, paper or other forms without written permission is prohibited.



Chapter Six: Dissecting Model Visuals

Up to this point, we've covered a number of lessons you can employ to improve your ability to communicate with data. Now that you understand the basics of what makes a visual effective, let's consider some additional examples of what "good" data visualization looks like. Before covering our final lesson, in this chapter we will look at several model visuals and discuss the thought process and design choices that led to their creation, utilizing the lessons we've covered.

You'll notice some similar considerations being made across the various examples. When creating each example, I thought about how I want the audience to process the information and made corresponding choices regarding what to emphasize and draw the audience's attention to as well as what to de-emphasize. Because of this, you will see common points raised around color and size. The choice of visual, relative ordering of data, alignment and positioning of elements, and use of words are also discussed in a number of cases.

This repetition is useful to reinforce the concepts I'm thinking about and resulting design choices across the various examples.

Each visual highlighted was created to meet the need of a specific situation. I'll discuss the relevant scenarios briefly, but don't worry too much about the details. Rather, spend time looking at and thinking about each model visual. Consider what data visualization challenges you face where the given approach (or aspects of the given approach) could be leveraged.

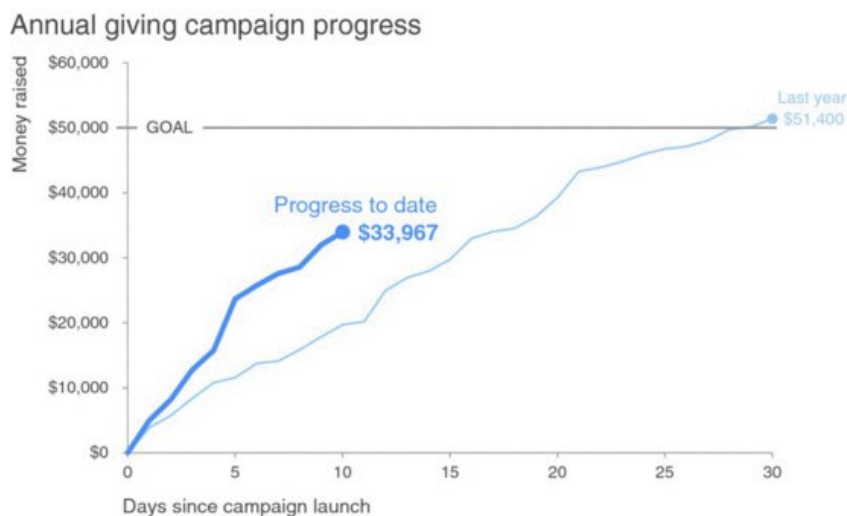


FIGURE 6.1: Line graph

Company X runs an annual month-long "giving campaign" to raise money for charitable causes. Figure 6.1 shows this year's progress to date. Let's consider what makes this example good and the deliberate choices made in the course of its creation.

Words are used appropriately. Everything is titled and labeled, so there's no question about what we are looking at. Graph title, vertical axis title, and horizontal axis title are present. The various lines in the graph are labeled directly, so there's no work going back and forth between a legend and the data to decipher what is being graphed. Good use of text makes this visual accessible.

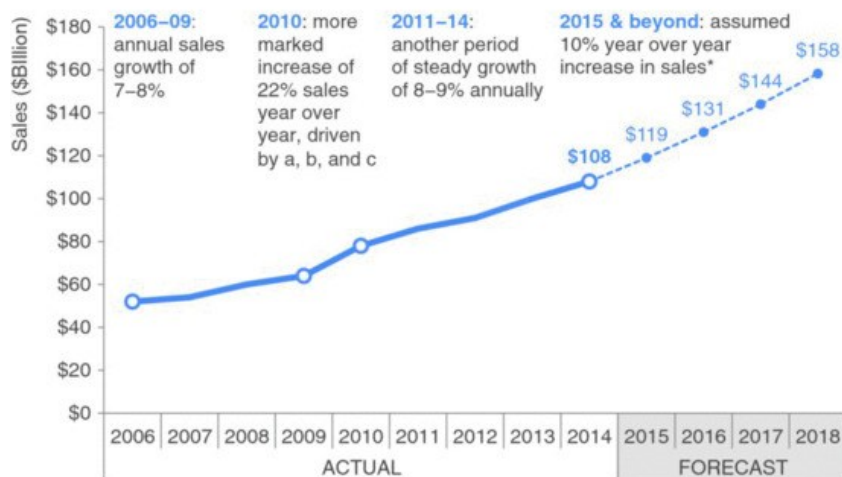
If we apply the "where are your eyes drawn?" test described in Chapter 4, I briefly scan the graph title, then I'm drawn to the "Progress to date" trend (where we want the audience to focus). I almost always use dark grey for the graph title. This ensures that it stands out, but without the sharp contrast you get from pure black on white (rather, I preserve the use of black for a standout color when I'm not using any other colors). A number of preattentive attributes are employed to draw attention to the "Progress to date" trend: color, thickness of line, presence of data marker and label on the final point, and the size of the corresponding text.

When it comes to the broader context, a couple of points for comparison are included but de-emphasized so the graph doesn't become visually overwhelming. The goal of \$50,000 is drawn on the graph for reference, but is pushed to the background by use of a thin line; both the line and text are the same grey as the rest of the graph details. Last year's giving over time is included but also de-emphasized through the use of a thinner line and lighter blue (to tie it visually with this year's progress, but without competing for attention).

A couple of deliberate decisions were made regarding axis labels. On the vertical y-axis, you could consider rounding the numbers to thousands—so the axis would range from \$0 to \$60 and the axis title would be changed to "Money raised (thousands of dollars)." If the numbers were on the scale of millions, I probably would have done this. For me, however, thinking about numbers in the thousands isn't as intuitive, so rather than mess with the scale here, I preserved the zeros in the y-axis labels.

On the horizontal x-axis, we don't need every single day labeled since we're more interested in the overall trend, not what happened on a specific day. Because we have data through the 10th day of a 30-day month, I chose to label every 5th day on the x-axis (given that this is days we're talking about, another potential solution would be to label every 7th day and/or add super-categories of week 1, week 2, etc.). This is one of those cases where there isn't a single right answer: you should think about the context, the data, and how you want your audience to use the visual and make a deliberate decision in light of those things.

Sales over time



Data source: Sales Dashboard; annual figures are as of 12/31 of the given year.

*Use this footnote to explain what is driving the 10% annual growth forecast assumption.

FIGURE 6.2: Annotated line graph with forecast

Figure 6.2 shows an annotated line graph of actual and forecast annual sales.

I often see forecast and actual data plotted together as a single line, without any distinguishing aspects to set the forecast numbers apart from the rest. This is a mistake. We can leverage visual cues to draw a distinction between the actual and forecast data, easing the interpretation of the information. In Figure 6.2, the solid line represents actual data and a thinner dotted line (which carries some connotation of less certainty than a solid, bold line) represents the forecast data. Clear labeling of Actual and Forecast under the x-axis helps reinforce this (written in all caps for easy scanning), with the forecast portion set apart visually ever so slightly via light background shading.

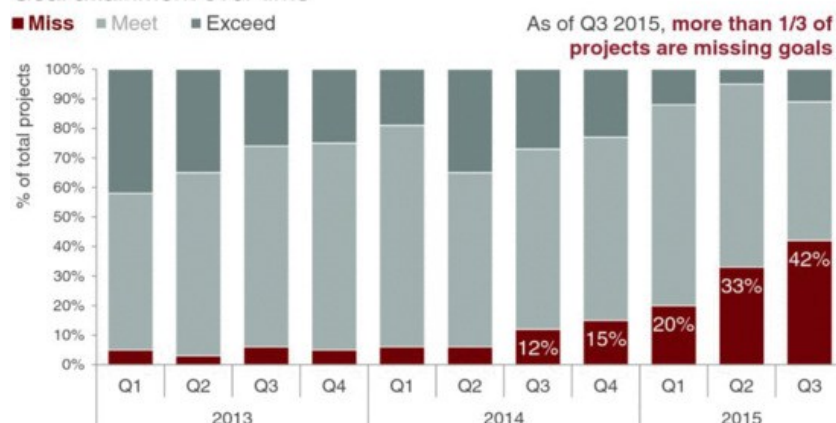
In this visual, everything has been pushed to the background through the use of grey font and elements *except* the graph title, dates within the text boxes, data (line), select data markers, and numeric data labels from 2014 forward. When we consider the visual hierarchy of elements, my eye goes first to the graph title at the top left (due to both position and the preattentive larger dark grey text discussed in the prior example), then to the blue dates in the text boxes, at which point I can pause and read for a little context before moving my eye downward to see the corresponding point or trend in the data. Data markers are included only for those points referenced in the annotation, making it a quick process to see what part of the data is relevant to which annotation. (Originally, the data markers were solid blue, but I changed to white with blue outline, which made them stand out a little more in a way that I liked; the forecast data markers are smaller and solid blue, because white with blue outline there looked overly cluttered against the dotted lines.)

The \$108 numeric label is bold. This is emphasized intentionally, since it is the last point of actual data and the anchor for the forecast. Historical data points are not labeled. Instead, the y-axis is preserved to give a general sense of magnitude, since we want the audience to focus on relative trends rather than precise values. Numeric data labels *are* included for the forecast data points to give the audience a clear understanding of forward-looking expectations.

All text in the visual is the same size except where intentional decisions were made to change it. The graph title is larger. The footnote is deemphasized via smaller font and a low-priority placement at the bottom of the visual so that it is there to aid interpretation as needed, but doesn't draw attention.

Model visual #3: 100% stacked bars

Goal attainment over time



Data source: XYZ Dashboard; the total number of projects has increased over time from 230 in early 2013 to nearly 270 in Q3 2015.

FIGURE 6.3: 100% stacked bars

The stacked bar chart in Figure 6.3 is an example visual from the consulting world. Each consulting project has specific goals associated with it. Progress against those goals is assessed quarterly and designated as “Miss,” “Meet,” or “Exceed.” The stacked bar chart shows the percentage of total projects in each of these categories over time. As with prior examples, don’t worry too much about the details here; instead, reflect on what can be learned from the design considerations that went into creating this data visualization.

Let’s first consider the alignment of objects within this visual. The graph title, legend, and vertical y-axis title are all aligned in the upper-left-most position. This means our audience encounters how to read the graph before they get to the data. On the left-hand side, the graph title, legend, y-axis title, and footnote are all aligned, creating a clean line on the left side of the visual. On the right-hand side, the text at the top is right-justified and aligned with the final bar of data that contains the data point being described (leveraging the Gestalt principle of proximity). This same text box is aligned vertically with the graph legend.

When it comes to focusing the audience’s attention, red is used as the single attention-grabbing color (primary red tends to be too loud for me, so I often opt instead for a burnt-red shade as I did here). Everything else is grey. Numeric data labels were used—an additional visual cue signaling importance given the stark contrast of white on red and large text—on the points we want the audience to focus: the increasing percentage of projects missing goals. The rest of the data is preserved for context, but pushed to the background so it doesn’t compete for attention. Slightly different shades of grey were used so you can still focus on one or the other series of data at a time, but it doesn’t distract from the clear emphasis on the red series.

The categories fall along a scale from “Miss” to “Exceed,” and this ordering is leveraged from bottom to top within the stacked bars. The “Miss” category is closest to the x-axis, making change over time easy to see because of the alignment of the bars at the same starting point (the x-axis). Change over time in the “Exceed” category is also easy because of the consistent alignment along the top of the graph. The change over time in the percentage of projects that meet their goals is harder to see because there is no consistent baseline at either the top or bottom of the graph, but given that this is a lower-priority comparison, this is OK.

Words make the visual accessible. The graph has a title, the y-axis has a title, and the x-axis leverages super-categories (years) to reduce redundant labeling and make the data more easily scannable. The words at the top right reinforce what we should be paying attention to (we will talk about words much more in the context of storytelling in Chapter 7). The footnote contains a note about the total number of projects over time, which is helpful context that we don’t get from the visual directly due to the use of 100% stacked bars.

Model visual #4: leveraging positive and negative stacked bars

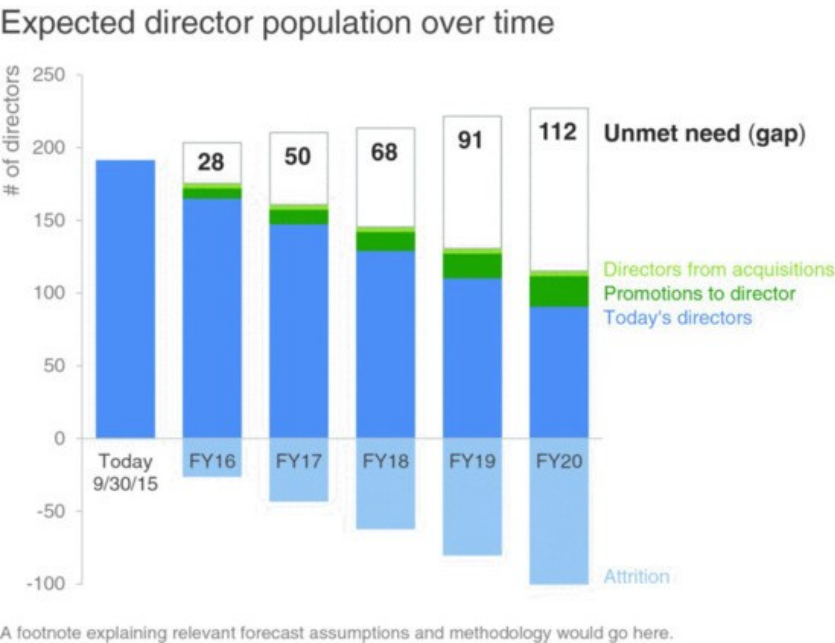


FIGURE 6.4: Leveraging positive and negative stacked bars

Figure 6.4 shows an example from the people analytics space. It can be useful to look forward to understand expected needs for senior talent and identify any gaps so they can be proactively addressed. In this example, there will be increasing unmet need for directors given assumptions for expected additions to the director pool over time through acquisitions and promotions and the decrease to the pool over time due to attrition (directors leaving the company).

If we consider the path our eyes take with Figure 6.4, mine scan the title, then go directly to the big, bold, black numbers and follow them to the right to the text that tells me this represents “Unmet need (gap).” My eye then goes downward, reading the text and glancing back leftward to the data each describes, until I hit the final series, “Attrition,” at the bottom. At this point, my eyes sort of bounce back and forth between “Attrition” and “Unmet need (gap)” portions of the bars, noting that there is some increase in the total number of directors over time as we look left to right (likely as the overall company grows and the need for senior leaders increases as a result), but that the majority of the unmet need is due to attrition of the current director pool.

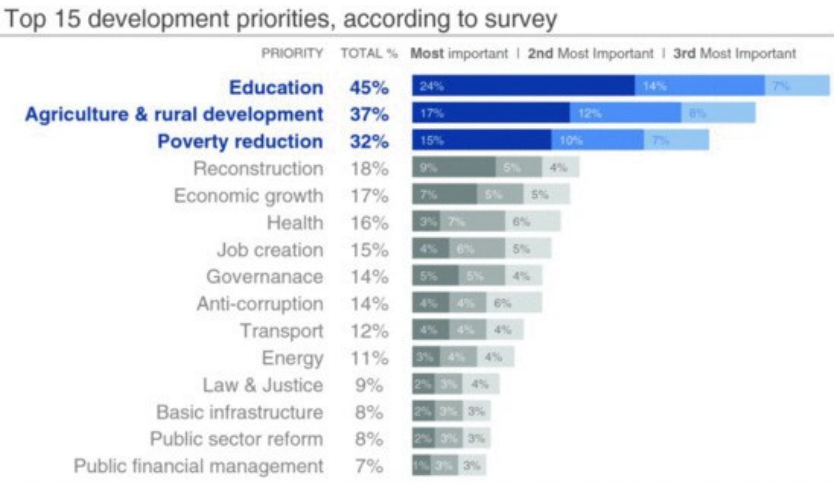
Intentional choices were made when it comes to the use of color throughout this visual. “Today’s directors” are shown in my standard medium blue. The exiting directors (“Attrition”) are shown in a less saturated version of the same color to tie these together visually. Over time, you see less of the blue falling above the axis and an increasing proportion falling below the axis as more and more directors attrite. The negative direction of the “Attrition” series reinforces that this volume represents a decrease to the director pool. Directors added through acquisitions and promotions are shown in green (which carries positive connotation). The unmet need is depicted by an outline only, to visually show empty space, reinforcing that this represents a gap. The text labels on the right are each written in the same color as the given data series they describe, except “Unmet need (gap),” which is written in the same big, bold, black text as the data labels for this series.

The ordering of the various data series within the stacked bars is deliberate. “Today’s directors” is the base, and as such is shown beginning at the horizontal axis. As I mentioned previously, the negative “Attrition” series falls below that in a negative direction. Above “Today’s directors” are the additions: promotions and acquisitions. Finally, at the top (where our eye hits sooner than the subsequent data), we encounter the “Unmet need (gap).”

The y-axis is preserved so the reader has a sense of total magnitude (both in the positive and negative direction), but it is pushed to the background via grey text. Only those specific points we should pay attention to—the “Unmet need (gap)”—are labeled directly with numerical values.

All text in the visual is the same size *except* where decisions were made to further emphasize or de-emphasize components. The graph title is larger. The axis title “# of directors” is slightly larger to ease the reading of the rotated text. The “Unmet need (gap)” text and numbers are bigger and bolder than anything else in the visual, as this is where we want the reader to pay attention. The footnote is written in smaller text, so it is there as needed but does not draw attention. By making it grey and in the lowest-priority position at the bottom of the visual, we further de-emphasize the footnote.

Model visual #5: horizontal stacked bars



N = 4,392. Based on responses to item, When considering development priorities, which one development priority is the most important? Which one is the second most important priority? Which one is the third most important priority? Respondents chose from a list. Top 15 shown.

FIGURE 6.5: Horizontal stacked bars

Figure 6.5 shows the results of survey questions on relative priorities in a developing nation. This is a great deal of information, but due to strategic emphasizing and de-emphasizing of components, it does not become visually overwhelming.

Stacked bars make sense here given the nature of what is being graphed: top priority (in first position in the darkest shade), 2nd priority (in second position and a slightly lighter shade of the same color), and 3rd priority (in third position and an even lighter shade of the same color). Orienting the chart horizontally means the category names along the left are easy to read in horizontal text.

The categories are organized vertically in descending order of “Total %,” giving the audience a clear construct to use as they interpret the data. The biggest categories are at the top, so we see them first. The top three priorities are emphasized specifically through the use of color (the narrative that accompanied the original version of this visual focused on these). This color is leveraged for the category name, total % and stacked bars of data. This consistent color ties the components together visually.

One decision point when graphing data is whether to preserve the axis, label the data points (or some data points) directly, or both. In this case, the numeric data labels within the bars have been preserved, but de-emphasized with smaller text (oriented to the left, which creates a clean line as you scan down the data labels for the “Most important,” making it feel slightly less cluttered than right- or center-oriented text that would vary in position across each of the bars). The data labels were further de-emphasized through the color they are written in: a light shade of blue or grey that doesn’t create as stark a contrast as white labels on a colored bar. The x-axis was eliminated altogether. Here, we implicitly assume that the specific values are important enough to label. Another scenario may call for a different approach.

As we noted with a number of the previous examples, words are used well in this visual. Everything is titled and labeled. The titles “Priority” and “Total %” are written all in caps for easy scanning. The legend for the interpretation of the bars appears immediately above the first bar of data with the keywords “Most,” “2nd,” and “3rd” bolded for emphasis. Additional detail is described in the footnote.

In closing

We can learn by examining effective visual displays and considering the design choices that were made to create them. Through the examples in this chapter, we've reinforced a number of the lessons covered up to this point. We touched on the choice of graph type and ordering of data. We considered where our eyes are drawn and in what order due to strategies employed to emphasize and de-emphasize components through the use of color, thickness, and size. We discussed the alignment and positioning of elements. We considered the appropriate use of text that makes the visuals accessible through clear titling, labeling, and annotation.

There is something to be learned from every example of data visualization you encounter—both good and bad. When you see something you like, pause to consider *why*. Those who follow my blog (www.storytellingwithdata.com) might be aware that I'm also an avid cook, and I often parlay the following food metaphor into data analytics: in data visualization, there is rarely (if ever) a single “right” answer; rather, there are flavors of good. The examples we've looked at in this chapter are the haute cuisine of charts.

That said, different people will make different decisions when faced with the same data visualization challenge. Because of this, inevitably I've made some design choices in these visuals that you might have handled differently. That's OK. I hope by articulating my thought process that you can understand why I made the design choices I did. These are considerations to keep in mind in your own design process. Of primary importance is that your design choices be just that: intentional.

Now you're ready for the final storytelling with data lesson: **tell a story**.