

POINTS OF VIEW

Points of review (part 2)

I will continue to demonstrate how judicious choice of graphical representations can improve visual communication. Here I will focus on data figures.

The power and primary purpose of graphs is to reveal connections in data. As opposed to tables, in which there is little visual association between individual values, graphs and charts depend on readers to form patterns. In reading graphs, we observe individual data points, keep each of them in memory and construct an image from the constituents. The entire process can be exceedingly fast and attest to the power of visual perception. Graphical encoding needs to support the detection and assembly process of reading graphs.

We are more accurate at certain types of visual estimation than others (September 2010 column)¹. For example, to understand relative differences between categories, a standard bar chart might be easier to read than a pie chart, particularly to appreciate the direction and magnitude of change (Fig. 1). Small differences are more readily apparent when we compare length of bars (Fig. 1c) than sizes of pie slices (Fig. 1a)².

Pie charts can be useful. Although they are not intended to show complex relationships, pie charts do well to depict parts of a whole. *The Wall Street Journal Guide to Information Graphics*³ suggests an ordering of slices to aid reading: place the largest wedge to the right of 12 o'clock, the second largest to the left of 12 o'clock and the remainder counter-clockwise descending in size (Fig. 1d). In this way, the largest (and presumably most important) wedges end up at the top. With the two largest slices sharing a vertical edge, we can rely on reading angles to estimate proportion.

When we need to show several dimensions of data at once, the multivariate scatter plot is one solution. With these displays of data, the challenge is in choosing representations that allow us to distinguish the qualities within and between parameters. In an example published figure that relies on position, color, color value and size to represent different aspects of the data (Fig. 1b)², it is difficult to pick out the eight sizes of data points, 11 shades of yellow and 13 shades of blue. One way to reduce the busyness is to limit the color value

Figure 1 | Certain visual encodings are easier to read. (a,b) Analysis of genetic interactions. Adapted and reprinted from *Nature Methods*². (c) A bar chart showing data from the pie chart in a. (d) A method for ordering slices of a pie chart. (e) Multiple views to show overlapping data from b. Former 'yellow' and 'blue' categories are shown in purple and green, respectively.

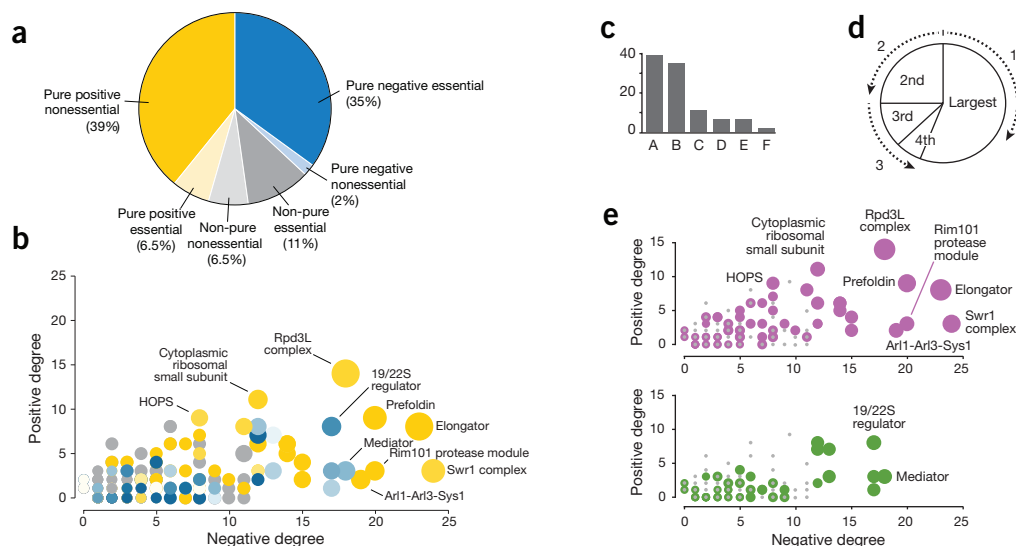
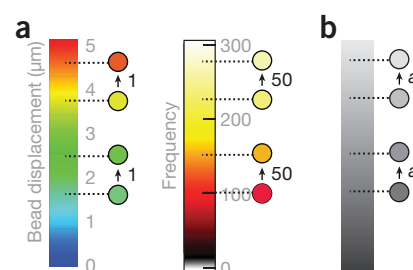


Figure 2 | Color is not ideal for presenting quantitative data. (a) Shifts in color scales (circles) are not visually commensurate with change in value. Reprinted from *Nature Methods*^{2,5}. (b) A gradation from 10–90% black produces even transitions.



and size scales to several ranges (for example, 0–3, 4–7 and others). Additionally, only plotting the parameters that matter most to convey the intended message will also reduce visual complexity. In the graph in **Figure 1c**, color value actually has a very limited role; it is not explicitly keyed in the original figure legend. But because of the severe data occlusion problem, it might be most helpful to separately plot the former yellow and blue categories each in gray (Fig. 1e).

Color is not ideal for representing quantitative information. In the above example, yellow is particularly problematic. It has an extremely restricted value range so there is not much difference between the lightest and deepest yellow. With color scales such as the rainbow spectrum, uneven transitions in color can break the correspondence between color and numerical value (August 2010 column)⁴. In **Figure 2a**, two color scales from recent journal articles are shown^{1,3}. In each instance, I sampled colors equal distance apart at two locations. The same incremental change in value does not equate to the qualitative difference between the pairs of color spots (Fig. 2a). Color can introduce considerable biases in data presentation. When we must represent values with color, a gradient of 10–90% black produces a consistent visual scale (Fig. 2b).

Next month I will cover another fundamental of design: typography.

Bang Wong

1. Wong, B. *Nat. Methods* **7**, 665 (2010).
2. Baryshnikova, A. *Nat. Methods* **7**, 1017–1024 (2010).
3. Wong, D. *The Wall Street Journal Guide to Information Graphics* (W.W. Norton and Company, New York, New York, USA, 2010).
4. Wong, B. *Nat. Methods* **7**, 573 (2010).
5. Legant, W. *Nat. Methods* **7**, 969–971 (2010).

Bang Wong is the creative director of the Broad Institute of the Massachusetts Institute of Technology and Harvard and an adjunct assistant professor in the Department of Art as Applied to Medicine at The Johns Hopkins University School of Medicine.