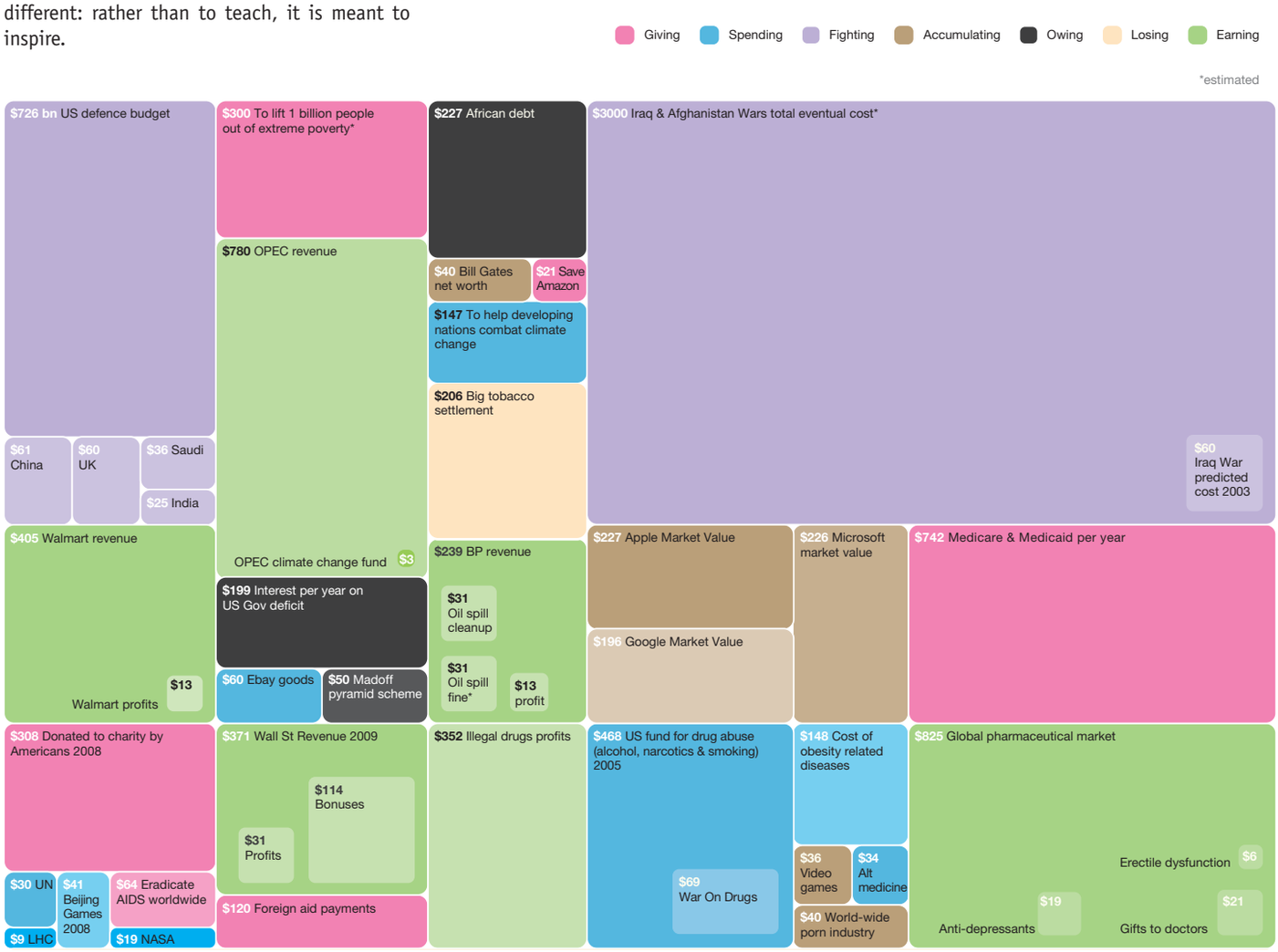
Making information beautiful – and clear

Information is presented too often in ways that bore the reader, or put him off, or simply baffle him. **Julian Champkin** looks at four examples that do better things with data

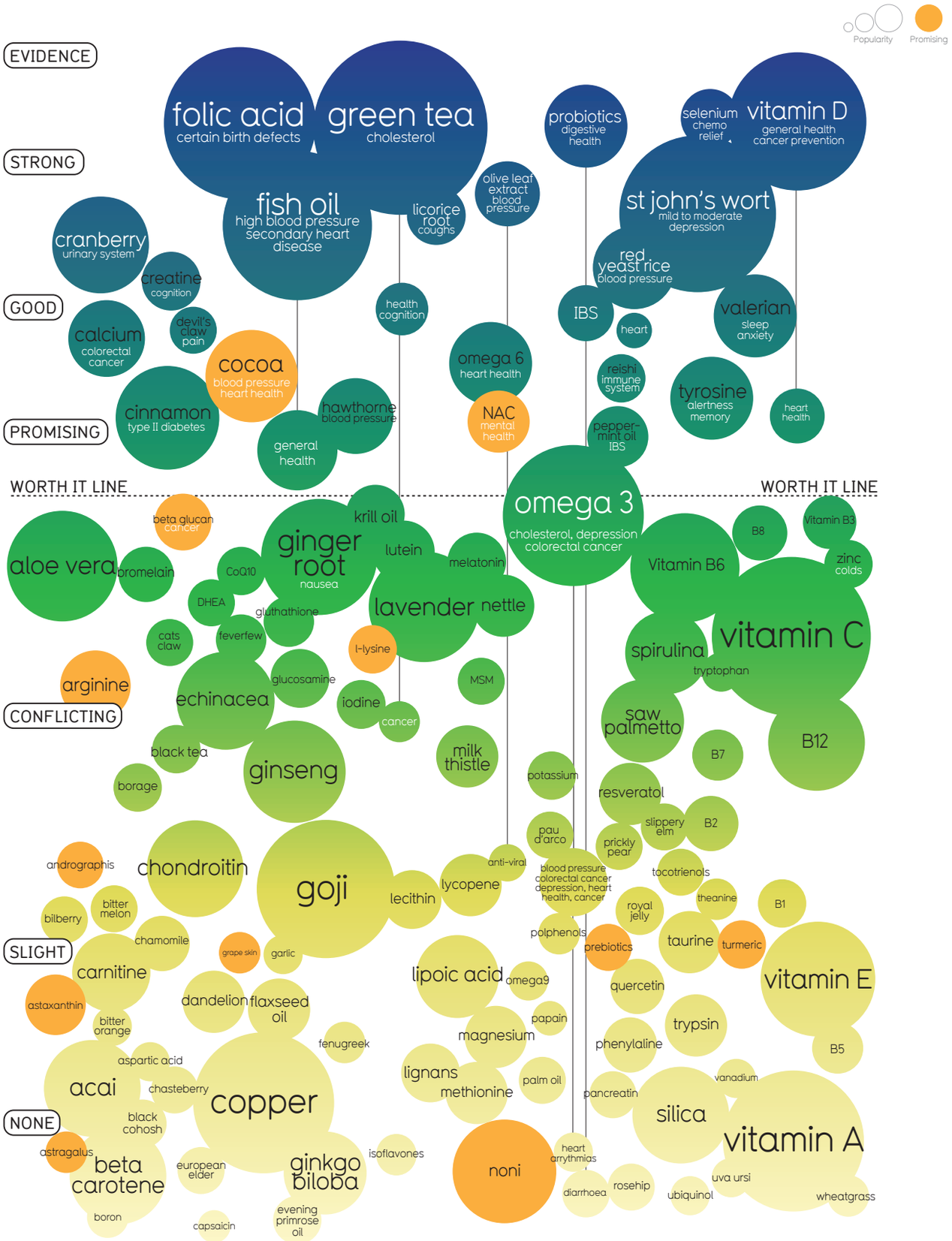
Usually this section tries to help the aspiring student to use one tool or another in the great armoury of techniques that are available to statisticians. This month's Toolkit is slightly different: rather than to teach, it is meant to inspire.

When all the statistics of a project have been done, the results have to be communicated; increasingly, they have to be communicated to non-specialists or to the general public. Data can be presented as tables; but to a non-statistician few things are as off-putting as a table of figures.

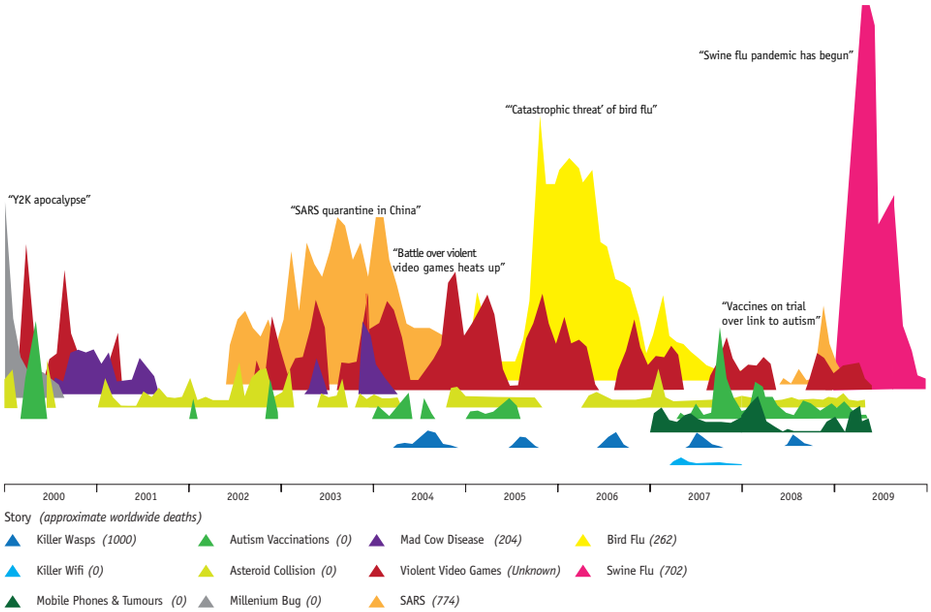


The Billion-Dollar-o-Gram, 2009 figures. Sources: *NYTimes*, *The Guardian*, BBC, CNN and other media reports. Note: some slight visual cheating to make things fit.

© 2011 The Royal Statistical Society



Snake oil? Scientific evidence for popular dietary supplements showing tangible health benefits when taken orally by an adult with a healthy diet. Source: English language placebo-controlled double-blind human trials on PubMed.org, US Office of Dietary Supplements, Herbmed.org, European Medicines Agency



Mountains out of molehills: a timeline of global media scare stories. Source Google News (worldwide deaths at time of print)

Detailed technical reports may be necessary; on the other hand they may not be, and one picture is worth a thousand words.

Whole books can, have, and should be written about visual presentation of data. One of them is *Information is Beautiful*, by David McCandless.

McCandless is not a statistician; he is a journalist. His skill is in communicating things. His diagrams have appeared in the pages of the *Guardian* newspaper and elsewhere; here we present four of them. They are not perfect. But they are different, and they are clear, and they convey the data in ways that every reader can understand – and that may even stick in the mind.

In a good graphic, the user reads the graphic, not the caption

How much did the Gulf oil spill affect BP's profits? If Wall Street bankers' bonuses were otherwise applied could it eliminate AIDS worldwide, and pay for a moderate-sized Olympic Games on top? And just how much is the war in Afghanistan costing, compared with what was predicted? Is that a lot or a little compared to, say, the nation's Medicare? The "Billion-Dollar-o-Gram" shows such things, by area. (A fuller version of it, incidentally, shows just how huge the world financial crisis was – as great as the rest of the chart combined.) Area, rather than length or height, is what the eye judges size by. All such charts of course are only as good as their data, and though sources are given the many caveats, uncertainties and problems of definition are of necessity skated over.

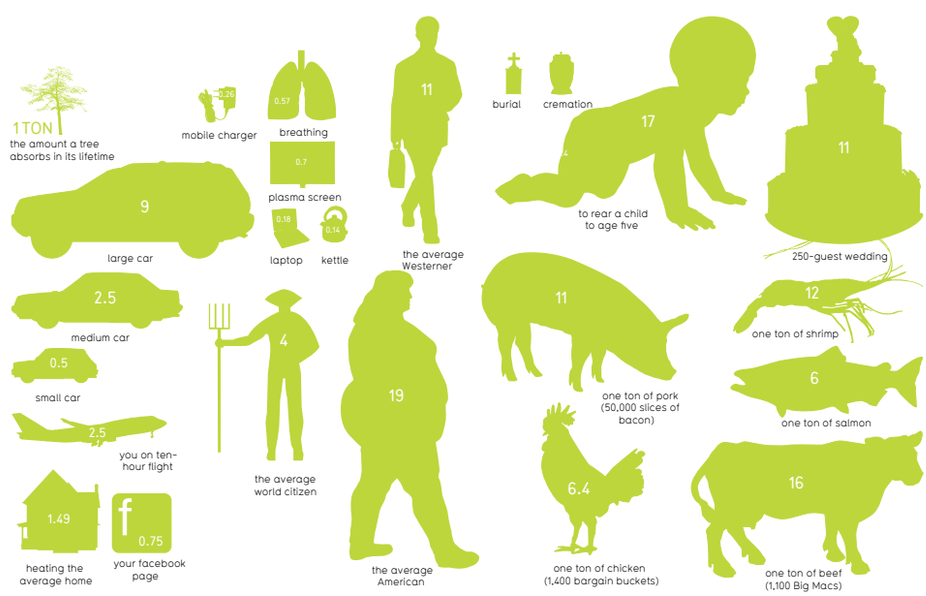
Trying to show two or more variables on one diagram is notoriously difficult, but it can be done: the "Snake Oil" bubble-chart shows, simultaneously and clearly, four – the popularity of the treatment, its promise, the evidence for its effectiveness, and what it is supposed to treat. Its message is clear at a glance – though the meaning of the vertical lines could do with clarification. (For a classic graphic that shows no fewer than six variables, from army casualties and geography to direction of advance and what the weather was like, and using only two colours at that, see Minard's famous diagram of Napoleon's march on Moscow (*Significance*, September 2008).)

Only a purist would complain about the lack of a vertical axis scale in "Mountains out of Molehills"; the story it tells is clear. The importance of a story in no way reflects the media space that is given to it. "Tons of Carbon" puts some surprising data before you: does your Facebook page really emit half as much carbon as heating your house? That the numbers are again represented by the area, not the height, of the outlines could be made more plain. And if the ton of shrimp really represents 12 tons of carbon, shouldn't it be the same size as the pig and the wedding cake?

Which in itself is a plus: errors can be spotted. The degree of redundancy – size with the number inside – itself forms a cross-check. A similar mistake in a plain table of the data just would not be noticed.

These are sunspots. The mark of a good graphic is that the user reads the diagram, not the caption. Information should be in the picture, not in the words underneath. These pass that test with flying colours.

They are not intended to be copied; rather to hint at what is possible. For many other examples, ingenious, informative and beautiful, see *Information is Beautiful*, by David McCandless, published by Collins at £20.00. The US title is *The Visual Miscellaneum: A Colorful Guide to the World's Most Consequential Trivia*, published by Collins Design at $26.99.



Tons of carbon emissions per year (unless otherwise stated). Source: *New York Times*, Environmental Protection Agency, IPCC, Energy Information Administration, UNESCO