Statistical graphics

(Hadley Wickham, Rice University)

1 Introduction

The field of statistical graphics is extremely broad. This article is my personal take on it, focussing on the connections to statistics and the results most important for creating insightful graphics. I have tried to select references that are accessible and informative, if not always the first published on the topic.

Graphics are useful because they display data in a way that is easy for the brain to parse and interpret. All displays of data are not equally accessible: we need to learn a little cognitive psychology to guide our choices. This is the topic of Section 2, which discusses three topics important to visualisation: the perception of continuous values, colour and differences. Section 3 gives a broad overview of static graphics, including the origins the most common statistical graphics, plots of statistical summaries, graphics for high-dimensional data, display spatial data, and the grammar of graphics, a theoretical framework that connects together many data graphics.

2 Perception

A small amount of knowledge about perception goes a long way in helping to understand and design effective visualisations. Here I touch on the aspects of perception that I think are most important: reading continuous values, Section 2.1, seeing colour, Section 2.2, and spotting differences, Section 2.3. Many of the most common graphical problems are caused by a lack of knowledge of these topics.

2.1 Perception of continuous values

Cleveland and McGill [10, 11] asked how we can display continuous data so that it can be most accurately decoded. Building on psychological theory, the authors performed an experiment asking participants to judge the relative proportion of two numbers displayed in different ways: as the height of bars, the length of lines, the areas of circles, and so on. The error rates associated with each graphical encoding gives an ordering from easiest to perceive to hardest. Heer and Bostock [28] recently reproduced this work on a large scale and with several new perceptual tasks, and found very similar results, as shown in Figure 1.

In brief, perceptual difficulty falls roughly into five groups from easiest to hardest: position along a common scale (T1, T2, T3), length and angle (T4, T5, T6), area (T7, T8, T9), volume and curvature, shading and colour saturation [10]. This ranking is extremely important because it provides a basic guide for the construction of new graphics: map the most important variables to the graphic properties that are the easiest to perceive.

However, these results don't tell the full story about perception. Area and angle are ranked lowly on this scale, which suggests that pie charts do not support accurate comparison. This is true, but comparing two numbers is only one question that can be asked of a plot. For other challenges, like comparing compound proportions, the pie chart has been found to be superior to the bar chart [38, 40]. Whenever declaring one chart "better" than another, we must be very careful about our definition of better.

2.2 Colour

Colour is important. It is not useful for the accurate decoding of continuous variables, but it is a powerful indicator of trend, and very useful displaying discrete variables. The perception of colour is a complex topic, but knowing a little bit about colour theory can help make the most of colour in your graphics and help you to avoid some common pitfalls. The focus of this section is on colour spaces, the numerical description of colour.

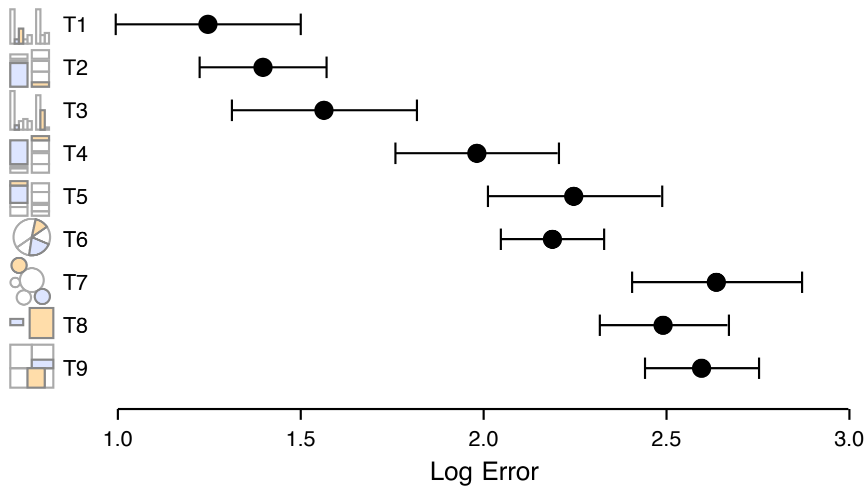

Figure 1: Error rates associated with comparison of two numbers displayed in various ways. Log 2 means displayed with 95% confidence intervals. Used with permission from XXX YYY.

You are probably most familiar with the RGB colour space which represents colours as a mixture of red, green and blue light. While this is an accurate model for how the eye perceives light, it does not match the brain's internal colour model. A better model is the hcl colour space, which breaks colour down into hue, chroma and luminance components [51]. Figure 2 shows five slices out of the space: luminance, or brightness, varies between the slices; hue varies with angle; chroma, or intensity, varies with distance from the origin. Compared to the RGB space, which is a simple cube, the hcl space has a much more complex shape because of the irregularity of colour perception. A more thorough introduction to colours is provided by Stone [41].

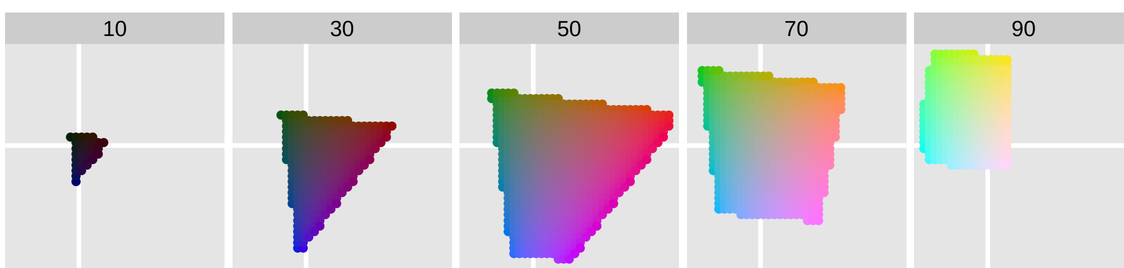

Figure 2: Five luminance slices (at 10, 30, 50, 70 and 90) from the hcl colour space. Not shown are slices at luminance 0 and 100 which are black and white dots respectively. Hue is given by the angle and chroma by the distance from the origin (marked with white lines). The range of possible chroma varies with both hue and luminance.

The hcl colour space is important for a few reasons. It is a perceptual space, which means that a unit of distance has an (approximately) equal perceptual difference everywhere in the space; this is not the case for RGB. Chroma and luminance are ordered, and so are useful for continuous variables, while hue is not ordered, so is appropriate for discrete variables. Colours with moderate chroma and luminance tend to be more aesthetically appealing and restful on the eyes. Zeileis et al. [51] discuss a number of colour schemes that take advantage of these properties. Brewer [5, 6] combine the principles with user testing to develop a set of schemes tailored for maps, and made available at `http://colorbrewer2.org/`.

Figure 3 illustrates three colour scales. The top two use the hcl colour space, and illustrate one scale with equal chroma and luminance at each end and one scale with varying luminance: there is a clear ordering when luminance is varied, but not when both chroma and luminance are constant. The bottom scale illustrates why the hcl space is so important. The rainbow colour scheme is path through the RGB colour space, but we do not see a smooth transition from colour to colour: instead we see discrete regions of roughly constant colour punctuated by abrupt changes. For this reason, rainbow schemes are best avoided.

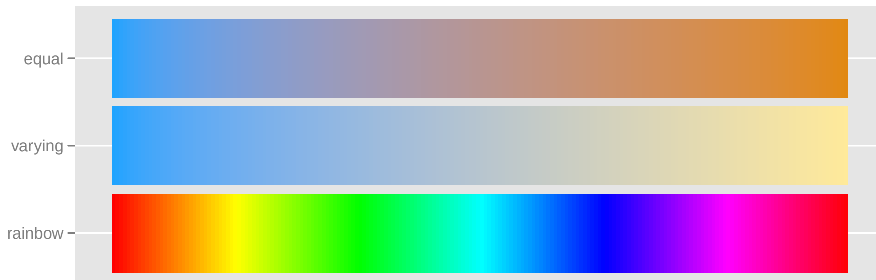

Figure 3: Three colour scales. The top two use hcl, and the bottom one uses RGB. The rainbow colour scheme has severe problems: even though it is a smooth path through rgb space, we not perceive it as such.

Another important fact about colour is that 5-10% of men have some sort of red-green colour impairment. For this reason, red-green colour contrasts should be avoided (heatmaps are a frequent offender) and redundant encodings should be used where possible (i.e. use both shape and colour in a scatterplot). Vischeck (`http://www.vischeck.com/`) and Color Oracle (`http://colororacle.cartography.ch/`) are useful tools for simulating colour blindness and testing for potential problems.

2.3 Spotting differences

Determining the distance between two curves is a surprisingly challenging task because the brain computes the shortest distance between lines, rather than the horizontal difference, is typically most important statistically. Figure 4 illustrates this problem with two curves offset by a constant. It looks like the curves are closer together on the left and further apart on the right. Even taking the relatively strong step of drawing vertical lines is of limited help. Be very cautious when displaying multiple curves and expecting viewers to compare them: it is far better to calculate the differences and draw them directly.

We are also surprisingly bad at spotting differences over time, and often fail to spot both dramatic and gradual changes. I recommend watching the following youtube videos which illustrate our inability to spot changes over time: `http://youtu.be/mAnKvo-fPs0`, `http://youtu.be/vJG698U2Mvo`, `http://youtu.be/1nL5ulsWMYc`. They are short and surprising, and you will be amazed by how easy it is to miss obvious patterns.

3 Static graphics

Understanding perception helps create new graphics, but if you don't have a solid understanding of what graphical forms have been explored, you'll spend a lot of time recreating the wheel. This section discusses some of the most important existing graphics, starting with a look back at where the most common statistical

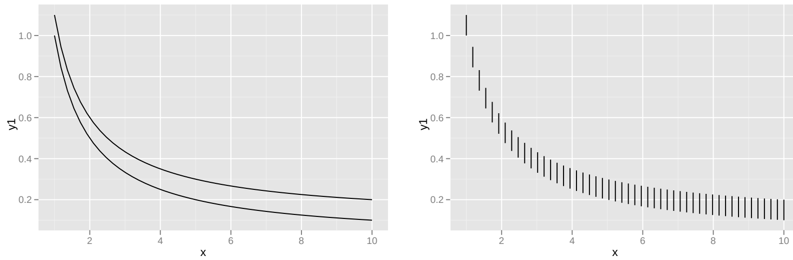

Figure 4: Top curve: $y = 1/x$, bottom: $y = 1/x + 0.1$. (Left) Each function drawn with its own line: lines look much closer on the left than the right. (Right) Vertical difference between the functions accentuate by drawing evenly spaced vertical lines. between the Even emphasising the vertical distance by drawing vertical lines doesn't help, a strong optical illusion remains.

graphics have come from (Section 3.1), continuing on to discuss plots that display statistical summaries (Section 3.2) and plots specially designed for high-d data (Section 3.3), and finishing with a look the grammar of graphics (Section 3.5), a theoretical framework that ties together many existing graphics.

3.1 Classic graphics

Many of the most commonly used graphics of today have been around for a long time. Playfair invented the bar and line charts in 1786, and the pie chart in 1801 (both works now available as Playfair et al. [35]). The scatterplot was invented in the 1830s [19, 39], although it wasn't until 1975 that the scatterplot matrix was seen in print [25]. Michael Friendly's graphical milestones, `http://www.datavis.ca/milestones/`, is an excellent resource to explore the historical development of these and many other plots.

3.2 Displaying summary statistics

Displaying raw data is only practical when the amount of data is small. This is a particular problem for 1d data: it is very difficult to see the distribution of points plotted along a single axis, because many points will be overlaid on one another. This is the problem of *overplotting*. This leads to the development of the histogram in the 1840's [18] and the boxplot in the 1970's [43]. This section focusses on two other important techniques for dealing with large datasets: improvements to the histogram for 1d and scatterplot smoothers for 2d.

Histograms are problematic because they can be extremely sensitive to the position of the origin: shifting the origin by a small amount (or even changing from right-open left-closed to left-open right-closed intervals) can dramatically change the resulting shape. The average shifted histogram (ASH) [37] is a simple (and computationally efficient) way of overcoming this problem: compute histograms with multiple starting points and then average over them. A more sophisticated alternative is the kernel density estimate (KDE) [vad014] [36], which computes the density at each location from a weighted estimate of neighbouring points and yields a smooth estimate with desirable statistical properties. A plot of a KDE is often called a density plot. Figure 5 illustrates the histogram, ASH and density plot for a mixture of two normals.

Scatterplots of large data suffer a similar problems to dotplots: overplotting obscures the bulk of the data. They also suffer from a new problem: extreme points may distract the eye and mislead the reader about the true relationship. We can alleviate both problem by fitting a smooth curve through middle of the data, helping draw the eye towards the signal, not the noise. A simple approach parallels the histogram: divide the data into bins along the x axis, then display the mean of the y values in each bin [16]. Cleveland [9] built on this approach in a similar way to the KDE, smoothly varying the weights used to estimate the

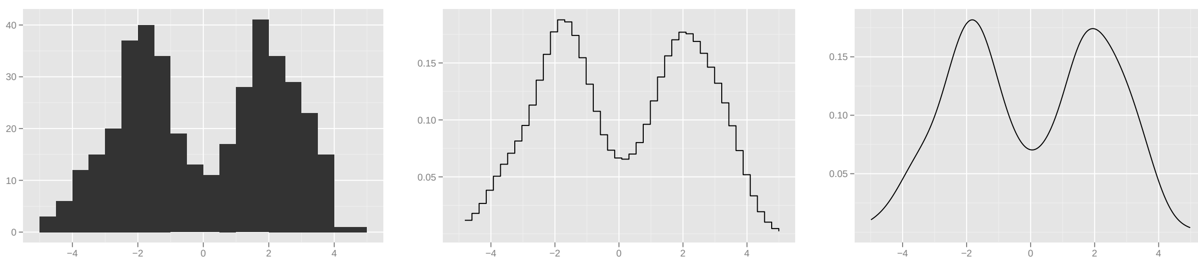

Figure 5: Three displays of a mixture of two normals. From left to right: histogram, average shifted histogram and density plot. Parameters selected to give approximately the same degree of smoothness.

mean at each location. This technique is called lowess (locally weighted scatterplot smoothing) or loess. Figure 6 illustrates these two methods.

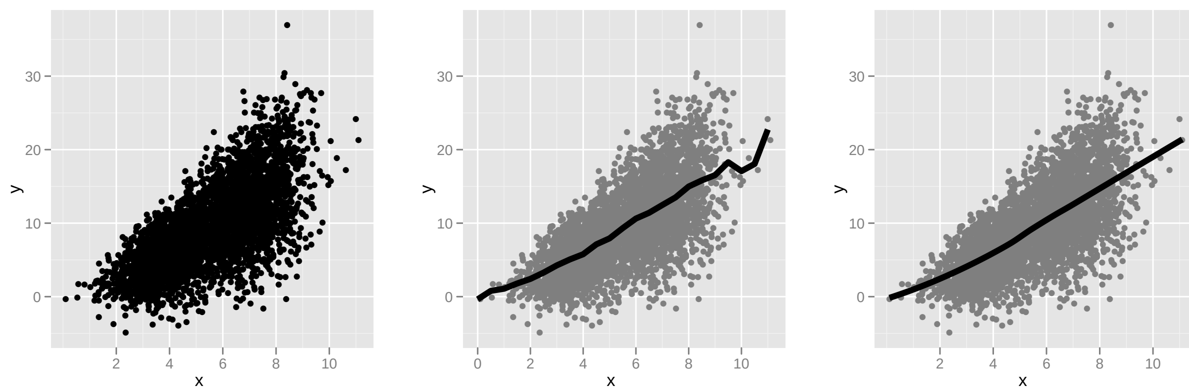

Figure 6: Three displays of a scatterplot of 4000 points. Raw data (left), raw data overlaid with binned means (centre), and raw data overlaid with a loess curve (right). It is easier to see the subtle non-linearity when a summary curve is added.

A related technique is "smoothing by weighted average of rounded points" [24], which applies the computationally efficient ASH approach to generate scatterplot smoothers for 2d (and higher). Other useful techniques that come from a more explicit modelling background are generalised additive models [vag007] and smooth non-parametric quantile regression [vnn091].

Density techniques also readily extend to 2d (the 2d histogram and 2d KDE) and can also be useful for dealing with overplotting. A complication is the choice of bin shape: there is only one way to slice up the real line, but many ways to partition the real plane. The simplest approach is to use rectangles, but these can produce distracting visual artefacts because of the gridding. An alternative is to use hexagons Carr et al. [7], which, surprisingly, are no more computationally complex than squares. Figure 7 illustrates three 2d density estimates. Compared to smoothing-based techniques which focus on the distribution of y conditional on x, density based techniques tend to focus on the joint distribution of y and x.

One complication of summary methods is that we now have a parameter to select: how wiggly do we want the curve? (The binwidth for the histogram, the bandwidth for KDEs and the span for loess). There have been many automated criteria proposed, typically focussed on minimising the integrated mean square error, but some have argued that this is not the right criteria for data analytic graphics [13]. It is often better for displays to be slightly undersmoothed, so that you can spot suspicious deviations, but still do some smoothing by "eye". Additionally, there may not be one optimal value, but a small set of values each of which

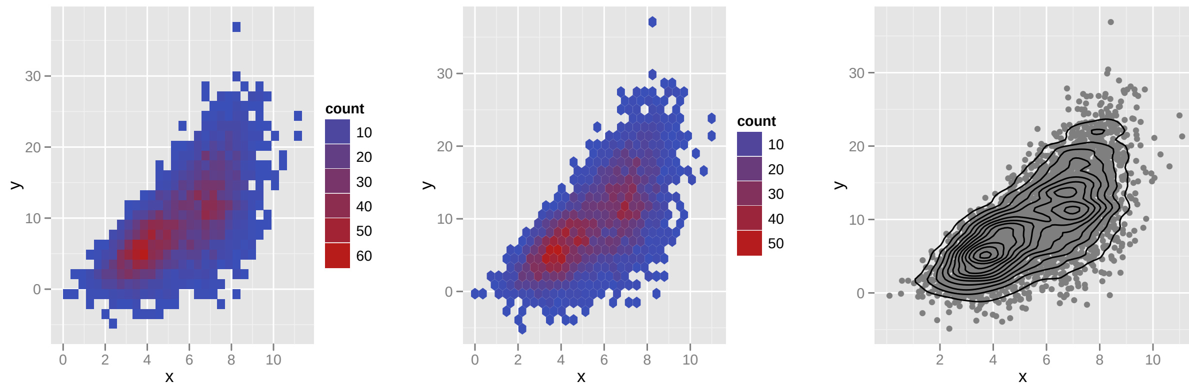

Figure 7: Four displays of a mixture of two normals, as in Figure 6. From left to right: histogram, average shifted histogram and density plot. Parameters selected to give approximately the same degree of smoothness.

illustrates different features of interest. This is analogous to the problem of aspect ratio selection [27].

3.3 Multidimensional data

Other plots have been invented specifically to better display multidimensional data:

- Mosaic plots [17, 26, 29] are a tool for high-dimensional categorical data that recursively partition space according to counts in contingency table. A recent extension that includes bar charts and treemaps is the product plots framework [47].

- Parallel coordinate plots, first used by Maurice d'Ocagne in 1885, were rediscovered and formalised by Inselberg [32] and popularised in statistics by Wegman [44]. They work by displaying data in a projective geometry. Axes are displayed in parallel, and observations are drawn as lines connecting their values on each axis. There has been much work done on the problem of selecting optimal orderings. See Hurley and Oldford [31] for a recent approach.

- Glyph plots convert display each observation as a glyph or icon, where data values are mapped to various properties of the glyph, such as length of lines [1, 33], or (infamously) as elements of a human face [8].

These plots tend to work with best with a moderate numbers of dimensions (10-20). There are fewer off-the-shelf methods available for very high-dimensional data sets, and analysis demands a tight integration between modelling and visualisation tools. One popular tool is the heatmap: a matrix display that colours tiles according to value. The ordering of rows and columns ("seriation") is critical for making informative displays. Hahsler et al. [23] provides a review of the relevant literature and describes their implementation in an eponymous R package. Heatmaps exploded in popularity in the biological literature in the 1990s, but ancestors have been around since the late 19th century [50]. A common problem with heatmaps is the use of the red-green colour scale: 5-10% of men have a red-green colour weakness and have difficulty distinguishing positive and negative values.

3.4 Spatial data

Spatial data is data associated with locations; points, lines and polygons. This section focusses on the most challenging visualisation problem: visualising data associated with polygons. This is a challenge because the two most easily compared attributes (x and y position) are already taken. There are two popular basic approaches: fill each polygon with texture or colour, or distort the polygon so that its area is proportional to its value.

Figure 8 shows two *choropleth maps*, where colour is mapped to value, here state population. The left plot is *unclassed*, with population is mapped to a continuous colour palette. This can make patterns difficult to discern for highly skewed distributions (like population). An alternative, shown on the right, is the *classed* choropleth, where value is binned prior to plotting. As with the histogram, there are many ways to bin the data, but a good rule of thumb is to bin the data into 5-9 categories each containing approximately the same number of observations. [4] performs a thorough literature review and user evaluation to justify this choice.

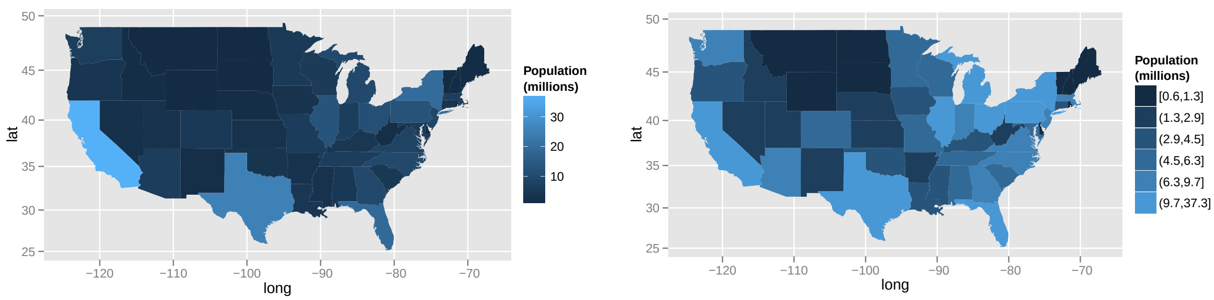

Figure 8: (Left) An unclassed choropleth map. (Right) A classed choropleth map with six classes, each containing approximately the same number of observations.

A problem with choropleth maps is that polygons with small areas are hard to see, regardless of how extreme their value is [22]. This is particularly problematic because small geopolitical areas often have high population densities. This problem is well illustrated in the US: the states in Pacific Northeast are small with large populations, while states adjacent to the Rockies have larger areas and small populations. Breaking into smaller divisions often doesn't help either: zip codes, for example, were designed to contain approximately equal numbers of people. Gastner et al. [21] provides a good discussion of how these issues relate to understanding the US presidential elections.

One variation of the choropleth is the *dot density map*, shown in Figure 9, where each polygon is filled with of randomly placed dots, with the number of dots proportional to the value. This requires careful choice of number and size of dots but can be an effective display. Note that the reader must be cautioned that an individual dot doesn't mean anything: it's the density of dots within a region that conveys information.

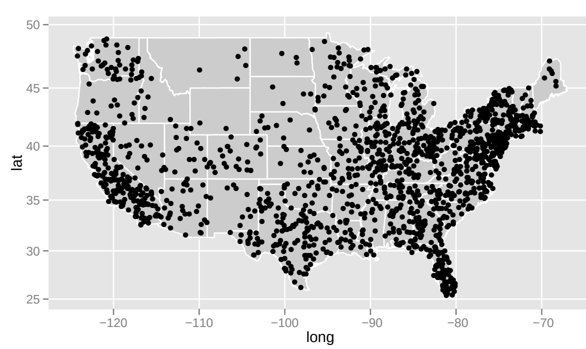

Figure 9: A dot density plot. Each dot represents 200,000 people.

An alternative approach altogether is the *cartogram* where each polygon is distorted so that its area is proportional to the desired value. Figure 10 shows two simple examples. The left figure is a non-contiguous cartogram [34] where each state has been resized so its area is proportional to its population. The right figure replaces the irregular state polygons with circles, and is a step towards a Dorling cartogram [14], which also requires repositioning of locations to avoid overlaps and recreate the original connectedness as much as possible. These two graphics shows the central tradeoff of cartograms: making areas easier to compare tends to make the spatial context less recognisable.

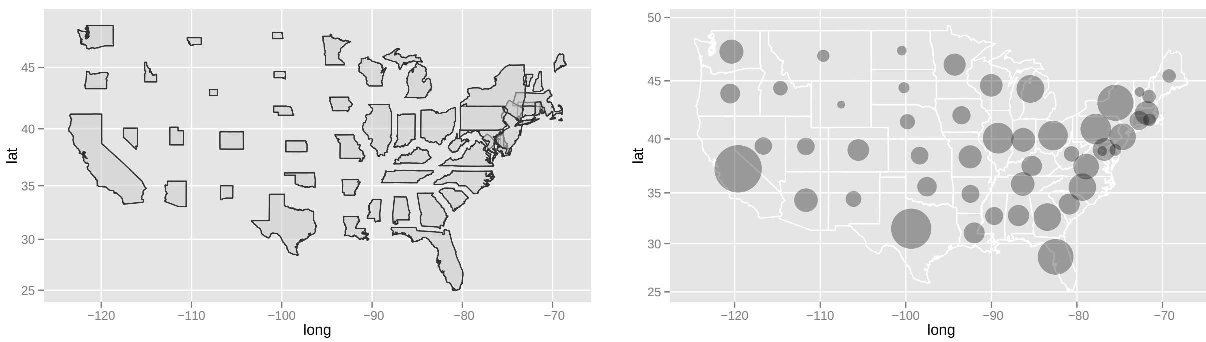

Figure 10: (Left) A non-contiguous cartogram, with states sized according to their population. (Right) Arbitrary areas are difficult to compare, so each state is replaced by a circle located at the state's center.

There are many other types of cartograms; Dorling [15] and Tobler [42] provide good summaries from a cartographic perspective. Other interesting work has been coming out of computer science visualisation communities, such as the constraint based approach of House and Kocmoud [30], or the diffusion based approach of Gastner and Newman [20].

3.5 The grammar of graphics

To do more than understand graphics as a collection of special forms, it's necessary to divine some underlying order, and start to understand graphics as combinations of interdependent components. This process was started in Bertin's semiology of graphics in 1967 [2] (but was not translated to English until 1983 [3]), and made extensive, rigorous and backed up by software in the "grammar of graphics" [48, 49]. It was enhanced in [45, 46] and coupled with an open-source implementation (used to make the majority of graphics in this article).

Bertin's semilogy and Wilkinson's grammar are useful because they transition from a typology of named graphics to a richer theory that helps to "bring together in a coherent way things that previously appeared unrelated and which also will provide a basis for dealing systematically with new situations" [12]. They provide a framework that makes it easier to describe and reason about graphics, and with computational support, make it easier to go from plots in your head to plots on the page. You think about how your data will be represented visually, then describe that representation using a declarative language. The declarative language, or grammar, provides a set of independent building blocks, similar to nouns and verbs, that allow you to build up a plot piece by piece. You focus on describing what you want, leaving it up to the software to draw the plot.

Figure 11 demonstrates one small insight gained from the grammar of graphics: polar coordinates turns stacked bar plots (or spine plots) into pie charts and bullseye charts.

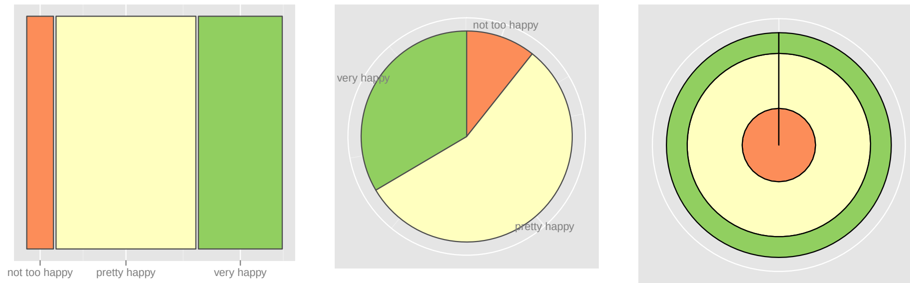

Figure 11: (Left) A stacked barchart showing the distribution of happiness. (Middle) A pie chart: the bar chart in polar coordinates with θ mapped to x. (Right) A "bullseye" chart: the bar chart in polar coordinates with θ mapped to y. Key: very happy, pretty happy, not too happy

References

[1] E. Anderson. A semigraphical method for the analysis of complex problems. *Proceedings of the National Academy of Sciences of the United States of America*, 43(10):923, 1957.

[2] J. Bertin. *Semilogie Graphique*. Mouthon-Gauthiers-Villars, Paris, 1967.

[3] J. Bertin. *Semiology of Graphics*. University of Wisconsin Press, Madison, WI, 1983.

[4] C. Brewer and L. Pickle. Evaluation of methods for classifying epidemiological data on choropleth maps in series. *Annals of the Association of American Geographers*, 92(4):662–681, 2002. URL `http://scripts.cac.psu.edu/users/c/a/cab38/Brewer_Annals.pdf`.

[5] C. A. Brewer. Color use guidelines for mapping and visualization. In A. MacEachren and D. Taylor, editors, *Visualization in Modern Cartography*, pages 123–147. Elsevier Science, 1994.

[6] C. A. Brewer. Guidelines for use of the perceptual dimensions of color for mapping and visualization. In *Color Hard Copy and Graphic Arts III, Proceedings of the International Society for Optical Engineering (SPIE), San Jose*, volume 2171, pages 54–63, 1994.

[7] D. B. Carr, R. J. Littlefield, W. L. Nicholson, and J. S. Littlefield. Scatterplot matrix techniques for large N. *Journal of the American Statistical Association*, 82(398):424–436, 1987.

[8] H. Chernoff. The use of faces to represent points in k-dimensional space graphically. *Journal of the American Statistical Association*, 68(342):361–368, 1973. URL `http://www.jstor.org/stable/2284077`.

[9] W. Cleveland. Robust locally weighted regression and smoothing scatterplots. *Journal of the American statistical association*, pages 829–836, 1979.

[10] W. S. Cleveland and R. McGill. Graphical perception: Theory, experimentation and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387):531–554, 1984.

[11] W. S. Cleveland and R. McGill. An experiment in graphical perception. *International Journal of Man-Machine Studies*, 25(5):491 – 500, 1986. doi: DOI:10.1016/S0020-7373(86)80019-0.

[12] D. R. Cox. Some remarks on the role in statistics of graphical methods. *Applied Statistics*, 27(1):4–9, 1978.

[13] L. Denby and C. Mallows. Variations on the histogram. *Journal of Computational and Graphical Statistics*, 18(1):21–31, 2009. doi: 10.1198/jcgs.2009.0002. URL `http://pubs.amstat.org/doi/abs/10.1198/jcgs.2009.0002`.

[14] D. Dorling. Map design for census mapping. *Cartographic Journal, The*, 30(2):167–183, 1993.

[15] D. Dorling. Area cartograms: Their use and creation. Concepts and Techniques in Modern Geography 59, Quantitative Methods Research Group of the Royal Geographical Society, 1996. URL `http://qmrg.org.uk/files/2008/11/59-area-cartograms.pdf`.

[16] M. Ezekiel. *Methods of Correlation Analysis.* John Wiley and Sons., New York, 2nd edition, 1941.

[17] M. Friendly. A brief history of the mosaic display. *Journal of Computational and Graphical Statistics*, 11 (1):89–107, 2002.

[18] M. Friendly. Milestones in the history of data visualization: A case study in statistical historiography. *Classification—the Ubiquitous Challenge*, pages 34–52, 2005.

[19] M. Friendly and D. Denis. The early origins and development of the scatterplot. *Journal of the History of the Behavioral Sciences*, 41(2):103–130, 2005. doi: 10.1002/jhbs.20078.

[20] M. Gastner and M. Newman. Diffusion-based method for producing density-equalizing maps. *Proceedings of the National Academy of Sciences of the United States of America*, 101(20):7499, 2004.

[21] M. Gastner, C. Shalizi, and M. Newman. Maps and cartograms of the 2004 us presidential election results. *Advances in Complex Systems*, 8(1):117–124, 2005.

[22] K. W. Haemer. Area bias in map presentation. *The American Statistician*, 3(2):19–19, 1949.

[23] M. Hahsler, K. Hornik, and C. Buchta. Getting things in order: An introduction to the r package seriation. *Journal of Statistical Software*, 25(3), 2008. URL `http://www.jstatsoft.org/v25/i03`.

[24] W. Härdle and D. Scott. Smoothing in low and high dimensions by weighted averaging using rounded points. *Computational Statistics*, 7:97–128, 1992.

[25] J. A. Hartigan. *Clustering Algorithms*. Wiley, 1975.

[26] J. A. Hartigan and B. Kleiner. A mosaic of television ratings. *The American Statistician*, 38(1):32–35, 1984.

[27] J. Heer and M. Agrawala. Multi-scale banking to 45 degrees. *IEEE Transactions on Visualization and Computer Graphics*, 12(5), 2006. URL `http://vis.berkeley.edu/papers/banking/`.

[28] J. Heer and M. Bostock. Crowdsourcing graphical perception: Using mechanical turk to assess visualization design. In *CHI 2010*, 2010.

[29] H. Hofmann. Constructing and reading mosaicplots. *Computational Statistics and Data Analysis*, 43(4): 565–580, 2003.

[30] D. House and C. Kocmoud. Continuous cartogram construction. In *Proceedings of Visualization '98*, pages 197–204. IEEE, 1998.

[31] C. B. Hurley and R. W. Oldford. Pairwise display of high-dimensional information via eulerian tours and hamiltonian decompositions. *Journal of Computational and Graphical Statistics*, 19(4):861–886, 2010. doi: 10.1198/jcgs.2010.09136. URL `http://pubs.amstat.org/doi/abs/10.1198/jcgs.2010.09136`.

[32] A. Inselberg. The plane with parallel coordinates. *The Visual Computer*, 1:69–91, 1985.

[33] B. Kleiner and J. A. Hartigan. Representing points in many dimensions by trees and castles. *Journal of the American Statistical Association*, 76(374):260–269, 1981. ISSN 01621459. URL `http://www.jstor.org/stable/2287820`.

[34] J. Olson. Noncontiguous area cartograms. *The Professional Geographer*, 28(4):371–380, 1976.

[35] W. Playfair, H. Wainer, and I. Spence. *The commercial and political atlas and statistical breviary*. Cambridge Univ Pr, 2005. URL `http://books.google.com/books?id=MMUcXPLqnI0C&printsec=frontcover`.

[36] D. Scott. *Multivariate density estimation*. Wiley Online Library, 1992.

[37] D. W. Scott. Averaged shifted histograms: Effective nonparametric density estimators in several dimensions. *The Annals of Statistics*, 13:1024–1040, 1985.

[38] I. Spence. Visual psychophysics of simple graphical elements. *Journal of Experimental Psychology: Human Perception and Performance*, 16(4):683, 1990.

[39] I. Spence and R. F. Garrison. A remarkable scatterplot. *The American Statistician*, 47(1):12–19, 1993.

[40] I. Spence and S. Lewandowsky. Displaying proportions and percentages. *Applied Cognitive Psychology*, 5(1):61–77, 1991.

[41] M. Stone. Representing colors as three numbers. *IEEE Computer Graphics and Applications*, 25(4): 78–85, 2005. URL `http://www.stonesc.com/pubs/Stone%20CGA%2007-2005.pdf`.

[42] W. Tobler. Thirty five years of computer cartograms. *Annals of the Association of American Geographers*, 94(1):58–73, 2004.

[43] J. W. Tukey. *Exploratory Data Analysis*. Addison–Wesley, preliminary edition, 1970.

[44] E. J. Wegman. Hyperdimensional data analysis using parallel coordinates. *Journal of the American Statistical Association*, 85(411):664–675, 1990.

[45] H. Wickham. *ggplot2: Elegant graphics for data analysis*. useR. Springer, July 2009.

[46] H. Wickham. A layered grammar of graphics. *Journal of Computational and Graphical Statistics*, 19(1): 3–28, 2010.

[47] H. Wickham and H. Hofmann. Product plots. *IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis '11)*, 2011. [26% acceptance rate.].

[48] L. Wilkinson. *The Grammar of Graphics*. Statistics and Computing. Springer, 1999.

[49] L. Wilkinson. *The Grammar of Graphics*. Statistics and Computing. Springer, 2nd edition, 2005.

[50] L. Wilkinson and M. Friendly. The history of the cluster heat map. *The American Statistician*, 63(2): 179–184, 2009. doi: 10.1198/tas.2009.0033.

[51] A. Zeileis, K. Hornik, and P. Murrell. Escaping RGBland: Selecting colors for statistical graphics. Report 61, Department of Statistics and Mathematics, Wirtschaftsuniversität Wien, Research Report Series, November 2007. URL `http://epub.wu.ac.at/1692/`.