35. Data Visualization for Human Perception

by Stephen Few

Data visualization is the graphical display of abstract information for two purposes: sense-making (also called data analysis) and communication. Important stories live in our data and data visualization is a powerful means to discover and understand these stories, and then to present them to others. The information is abstract in that it describes things that are not physical. Statistical information is abstract. Whether it concerns sales, incidences of disease, athletic performance, or anything else, even though it doesn't pertain to the physical world, we can still display it visually, but to do this we must find a way to give form to that which has none. This translation of the abstract into physical attributes of vision (length, position, size, shape, and color, to name a few) can only succeed if we understand a bit about visual perception and cognition. In other words, to visualize data effectively, we must follow design principles that are derived from an understanding of human perception.

As the saying goes, "a picture is worth a thousand words" - often more - but only when the story is best told graphically rather than verbally and the picture is well designed. You could stare at a table of numbers all day and never see what would be immediately obvious when looking at a good picture of those same numbers. Allow me to illustrate. Here's a simple table of sales data - a year's worth - divided into two regions:

2009 Sales (thousands of U.S. $)

Region	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
Domestic	1,983	2,343	2,593	2,283	2,574	2,838	2,382	2,634	2,938	2,739	2,983	3,493	31,783
International	574	636	673	593	644	679	593	139	599	583	602	690	7,005
Total	2,557	2,979	3,266	2,876	3,218	3,517	2,975	2,773	3,537	3,322	3,585	4,183	38,788

Figure 35.1

This table does two things extremely well: it expresses these sales values precisely and it provides an efficient means to look up values for a particular region and month. But if we're looking for patterns, trends, or exceptions among these values, if we want a quick sense of the story contained in these numbers, or we need to compare whole sets of numbers rather than just two at a time, this table fails.

Now look at the following picture of the same information in the form of a line graph:
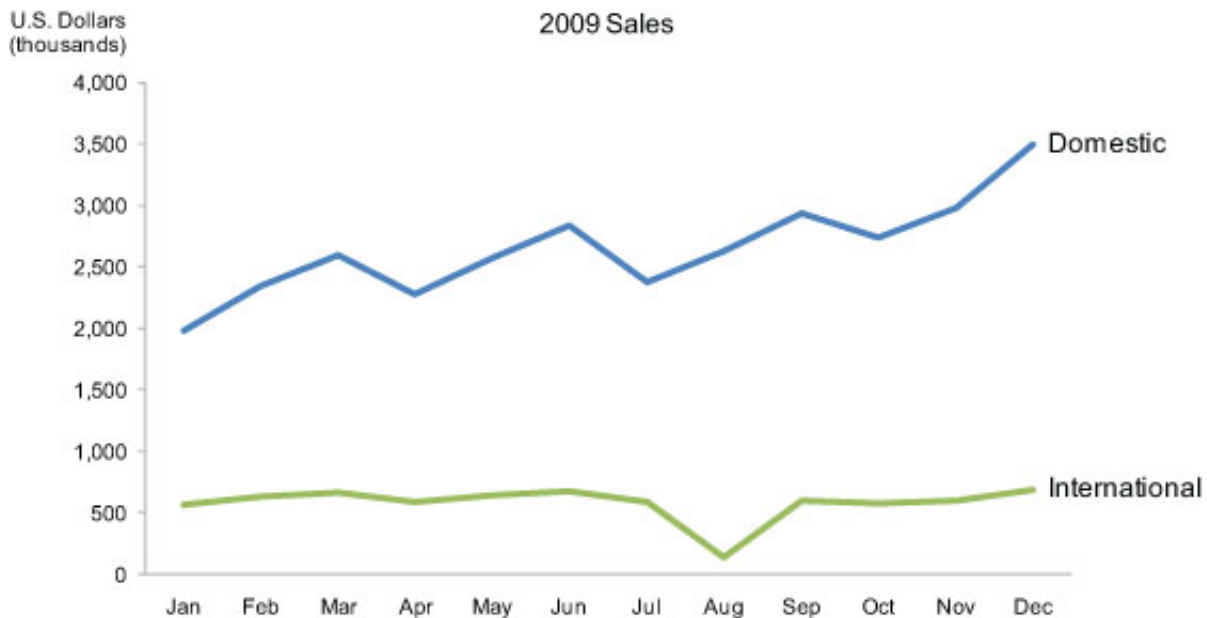
Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms below.

Figure 35.2

Several facts now leap into view:

- Domestic sales were considerably and consistently higher than international.
- Domestic sales trended upward over the year as a whole.
- International sales, in contrast, remained relatively flat, with one glaring exception: they decreased sharply in August.
- Domestic sales exhibited a cyclical pattern - up, up, down - that repeated itself on a quarterly basis, always reaching the peak in the last month of the quarter and then declining dramatically in the first month of the next.

What these numbers could not communicate when presented as text in a table, which our brains interpret through the use of verbal processing, becomes visible and understandable when communicated visually. This is the power of "data visualization."

Although data visualization usually features relationships between quantitative values, it can also display relationships that are not quantitative in nature. For instance, the connections between people on a social networking site such as Facebook or between suspected terrorists can be displayed using a node and link visualization. In the following example, people are the nodes, represented as circles, and their relationships are the links, represented as lines that connect them.
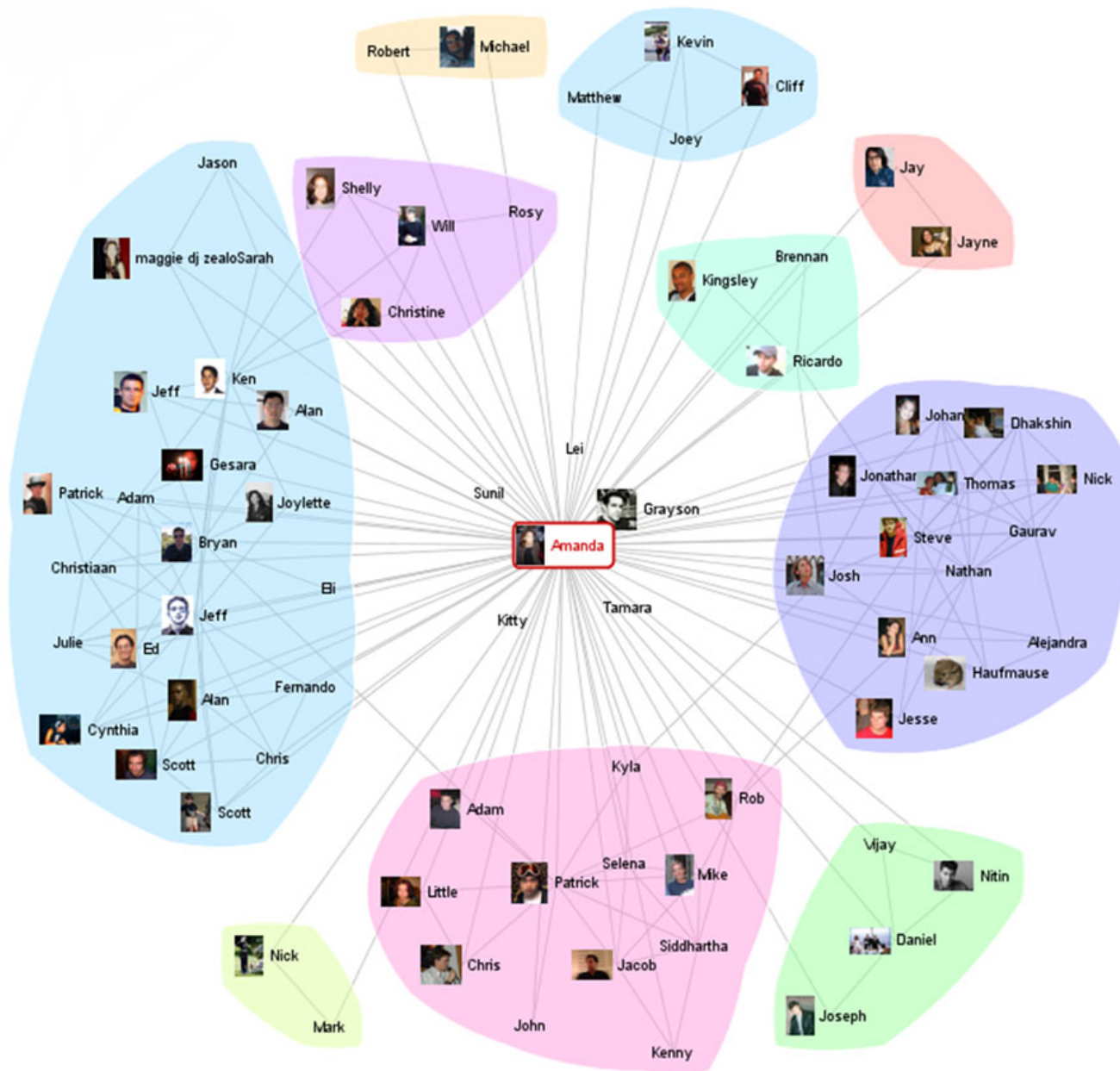
Courtesy of Jeffrey Heer and Danah Boyd using Vizster. Copyright: CC-Att-ND (Creative Commons Attribution-NoDerivs 3.0 Unported).
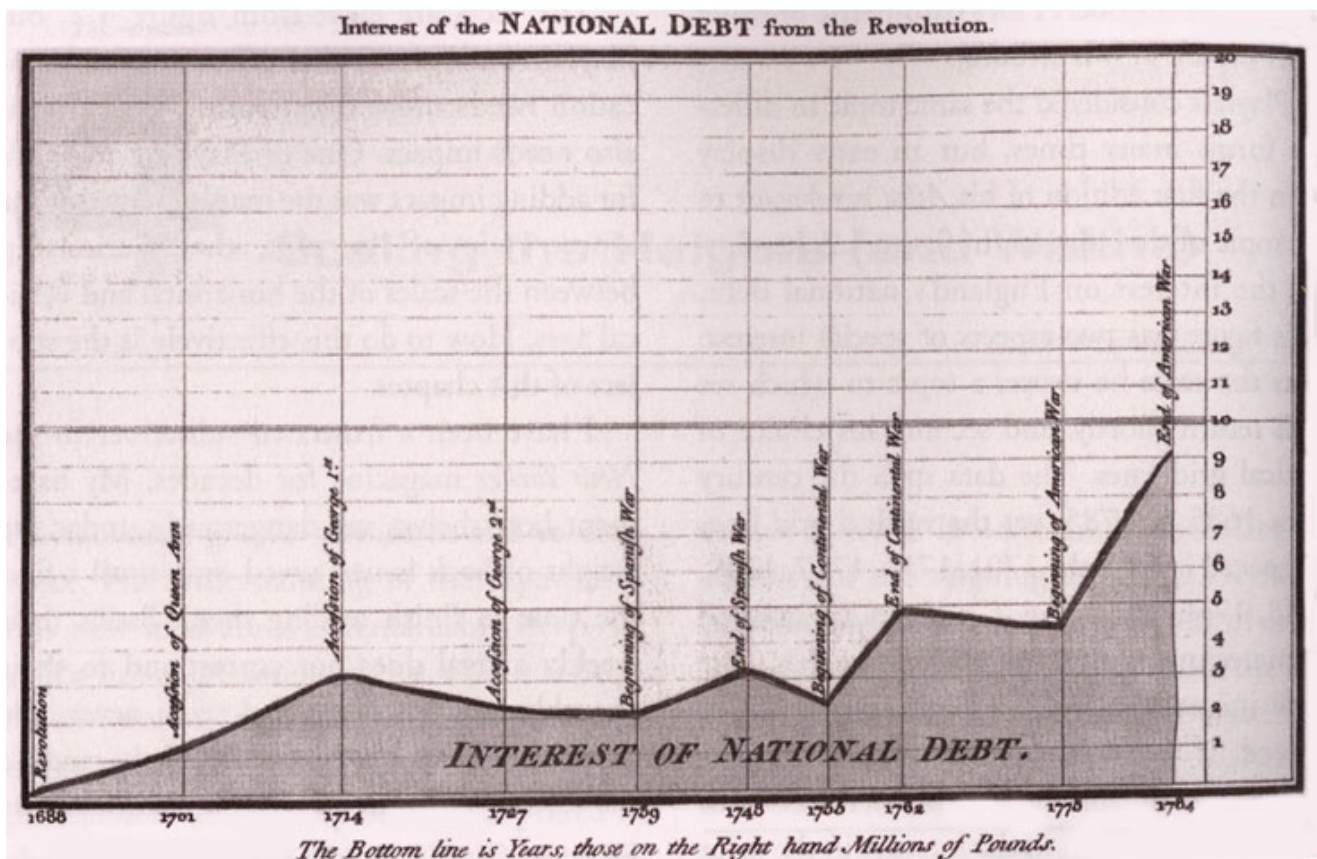
Figure 35.3

Visualizations that feature relationships between entities, such as the people in the example above, can be enriched with the addition of quantitative information as well. For example, the number of times that any two people have interacted could be represented by the thickness of the line that connects them.

35.1 Data Visualization in Historical Context

People have been arranging data into tables (columns and rows) at least since the 2nd century C.E., but the idea of representing quantitative information graphically didn't arise until the 17th century. For this innovation we have the French philosopher and

mathematician Rene Descartes to thank. He developed a two-dimensional coordinate system for displaying values, consisting of a horizontal axis for one variable and a vertical axis for another, primarily as a graphical means of performing mathematical operations. It wasn't until the late 18th century that we began to exploit the potential of graphics for the communication of quantitative data, for which we have the Scotsman William Playfair to thank. Playfair pioneered many of the graphs that are commonly used today. He was the first person to use a line moving up and down as it progressed from left to right to show how values changed through time, as in the example below. He also invented the bar graph, and on one of his off days he invented the pie chart, which we have since found relatively ineffective, because it encodes values as visual attributes (primarily the area of each slice as well as the angle that it forms in the center of the pie) that we cannot easily perceive and compare.

Courtesy of William Playfair (1759-1823). Copyright: pd (Public Domain (information that is common property and contains no original authorship)).

Figure 35.4: Playfair included this graph in his *The Commercial and Political Atlas* (1786) to argue against England's policy of financing colonial wars through national debt.

The use of quantitative graphs gradually increased over the years, but their methods and effectiveness evolved little until the second half of the 20th century. Jacques

Bertin laid the foundation for much of the progress that's been made during the last half a century with the publication in 1967 of the book *Semiologie graphique* (*The Semiology of Graphics*, Bertin 1967). His work was pivotal because he discovered that visual perception operated according to rules that could be followed to express information visually in ways that represented it intuitively, clearly, accurately, and efficiently.

The person who really introduced us to the power of data visualization as a means for exploring and making sense of quantitative data was the Princeton statistics professor John Tukey, who in 1977 gave form to a whole new statistical approach called *exploratory data analysis*.

In 1983, the person working in the field today whose name is recognized above all others, Edward Tufte, published his groundbreaking book *The Visual Display of Quantitative Information*. In it he pointed out that there were effective ways of displaying data visually and then there were the ways that most people were doing it, which didn't work very well. Also working to improve data visualization practices around this time was William Cleveland, who extended and refined data visualization techniques for statisticians.

Soon thereafter, a new research specialty emerged in the academic world, which was coined "information visualization." In their 1999 book *Readings in Information Visualization: Using Vision to Think*, Stuart Card, Jock Mackinlay, and Ben Shneiderman collected the best academic work that had been done by that time into a single volume and made its discoveries accessible beyond the walls of academia (Card et al 1999).

Since the turn of the 21st century, data visualization has been popularized, too often in tragically ineffective ways as it has reached the masses through commercial software products. Gratefully, amongst the bevy of products that promote data visualization in ways that feature superficially appealing aesthetics above useful and effective data exploration, sense-making, and communication, there are a few serious contenders for our attention who are helping us fulfill its potential in practical and powerful ways.
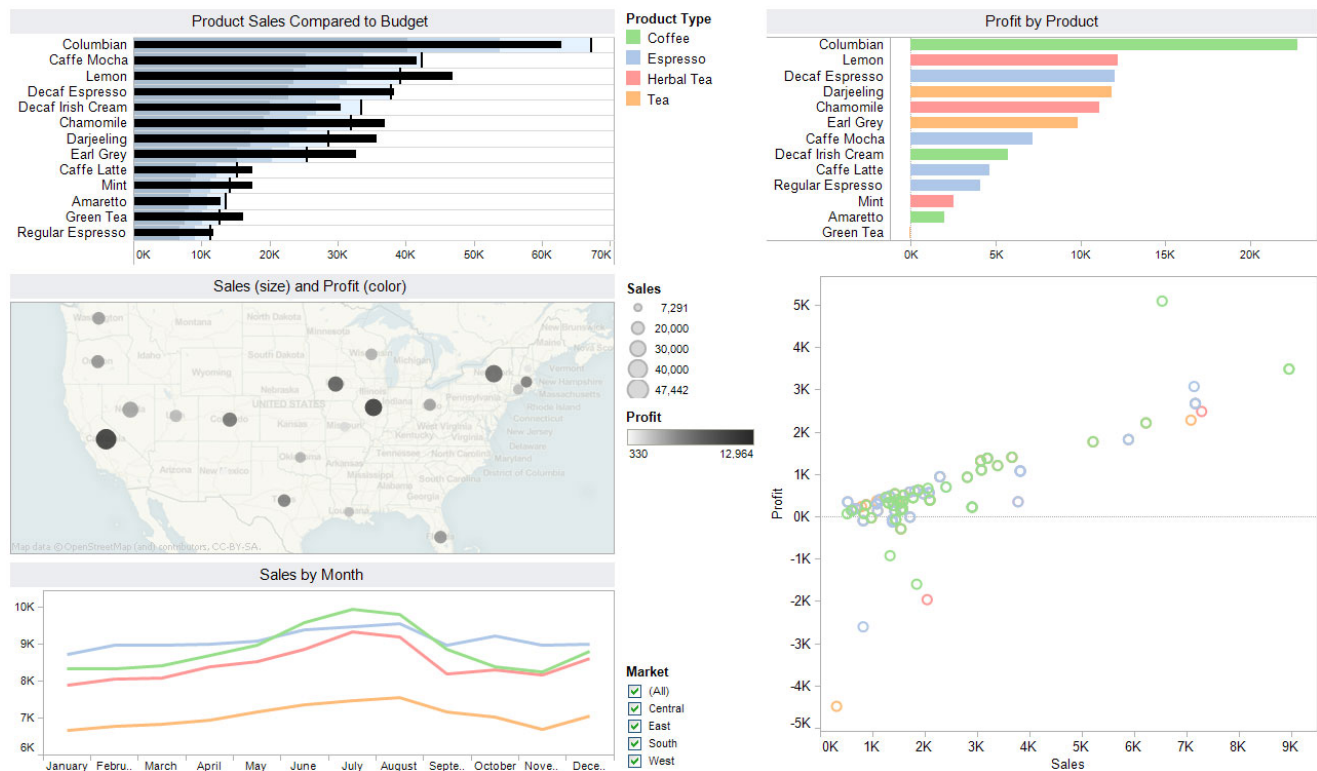
Figure 35.5: This display, consisting of multiple views of the same data set, was created using Tableau Software, one of the few software vendors that currently understand data visualization.

Among those who have contributed to our understanding of data visualization, Colin Ware has done the most to base its practice on an understanding of human perception. Ware's two excellent books - *Information Visualization: Perception for Design* (Ware, 2004) and *Visual Thinking for Design* (Ware 2008) - compile, organize, and explain what we have learned from several scientific disciplines about visual thinking and cognition and apply that knowledge to data visualization.

35.2 Pictures for the Eyes and Mind

Data visualization is only successful to the degree that it encodes information in a manner that our eyes can discern and our brains can understand. Getting this right is much more a science than an art, which we can only achieve by studying human perception. The goal is to translate abstract information into visual representations that can be easily, efficiently, accurately, and meaningfully decoded. Consider a case when you need to help people understand the primary causes of death in America contained in the following table:

Causes of Death	Deaths per Year
Heart disease	616,067
Cancer	562,875
Stroke (cerebrovascular diseases)	135,952
Chronic lower respiratory diseases	127,924
Accidents (unintentional injuries)	123,706
Alzheimer's disease	74,632
Diabetes	71,382
Influenza and Pneumonia	52,717
Nephritis, nephrotic syndrome, and nephrosis	46,448
Septicemia	34,828
All other causes	577,181
Total	2,423,712

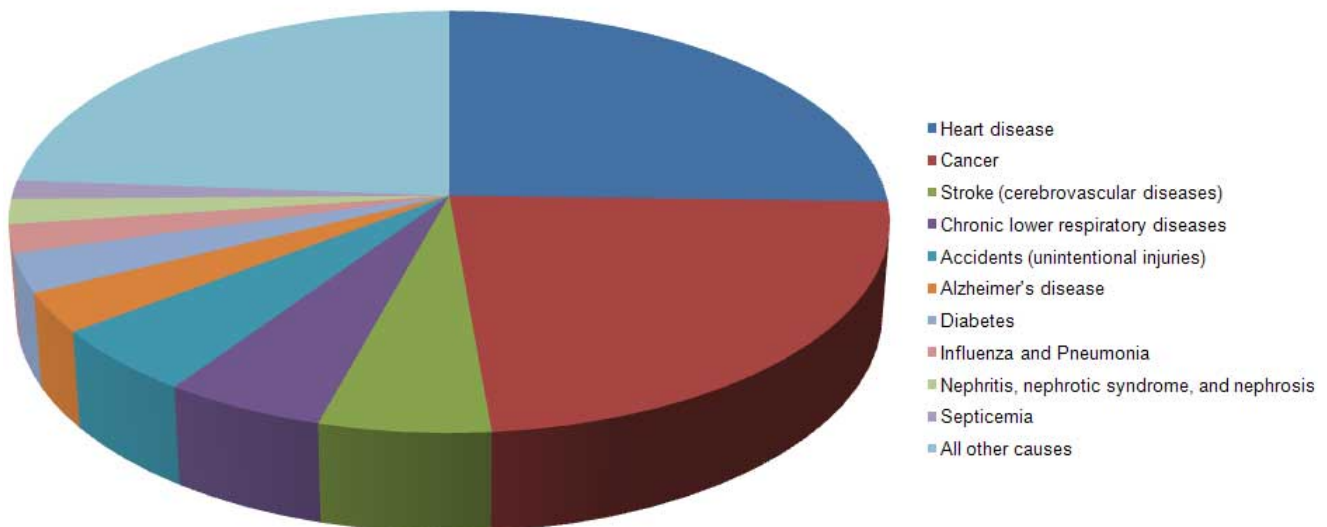Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms below.

Figure 35.6

To achieve this goal, the display should achieve the following:

- Clearly indicates how the values relate to one another, which in this case is a part-to-whole relationship - the number of deaths per cause, when summed, equal all deaths during the year.
- Represents the quantities accurately.
- Makes it easy to compare the quantities.
- Makes it easy to see the ranked order of values, such as from the leading cause of death to the least.
- Makes obvious how people should use the information - what they should use it to accomplish - and encourages them to do this.

The traditional way to display this information graphically involves a pie chart, illustrated below.

Total Deaths in American by Cause in 2007

Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms below.

Figure 35.7

How well does this pie chart satisfy our criteria for effectiveness? Let's consider each of the requirements.
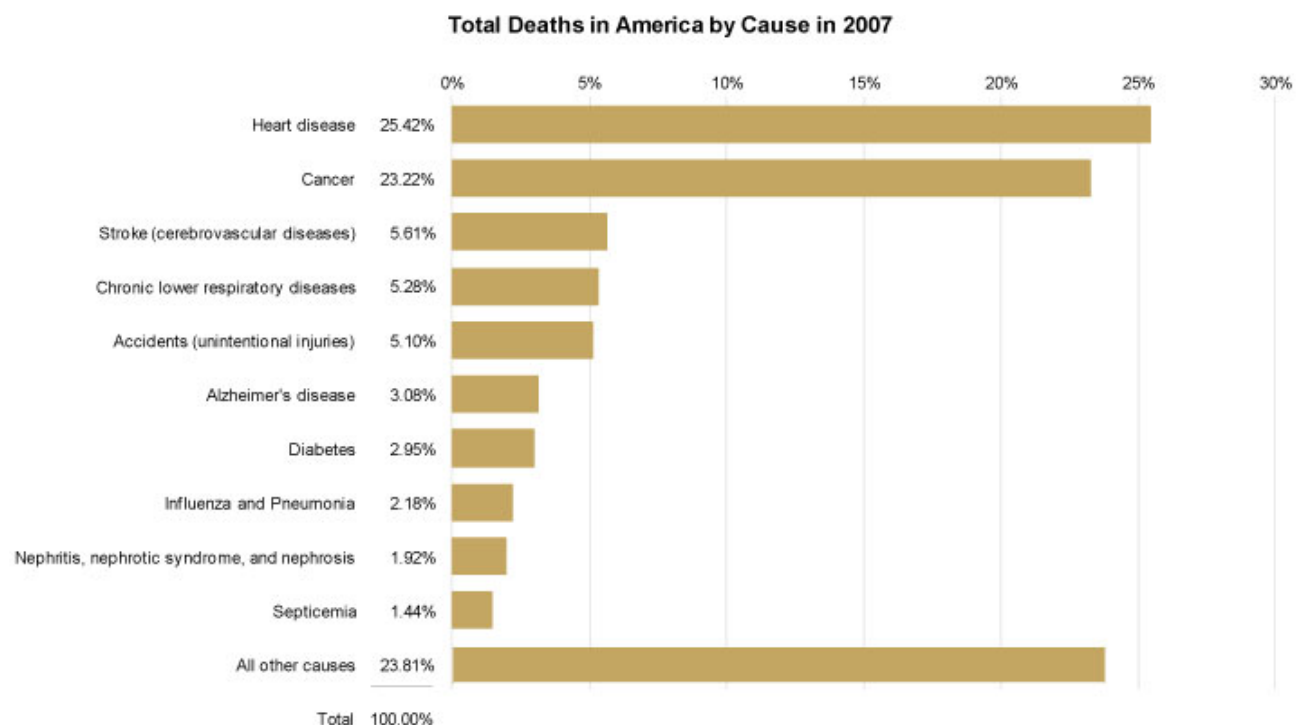
- *Clearly indicates the nature of the relationship? Yes.* The primary strength of a pie chart is the fact that it clearly indicates a part-to-whole relationship between the values.

- *Represents the quantities accurately? No.* Pie charts encode values redundantly through the use of three visual attributes: the area of each slice, the angle formed by each slice at the center of the pie, and the length of the each slice along the pie's perimeter. Even when the area, angle, and perimeter of each slice is calculated properly, it fails in that we cannot perceive any one of these attributes accurately. Visual perception in humans has not evolved to support accurate decoding of areas, angles, or distance along a curve.

- *Makes it easy to compare the quantities? No.* Because we cannot perceive the values accurately, we also cannot compare them easily or accurately. Furthermore, in this particular pie chart, because a legend has been used to label the slices, we are forced over and over to look up the meaning of the slices we wish to compare by finding the right color, which is often difficult to discriminate. The fact that this pie chart has been rendered in 3-D also complicates the simple act of comparison because the perspective skews the relative size and shape of the slices, making slices on the bottom appear larger and more salient than similarly sized slices on the top.

- *Makes it easy to see the ranked order of values? No.* Even though the slices are

displayed in ranked order from the highest value (heart disease) at the top and continuing clockwise to the smallest, excluding the final "All other causes" slice, this ranking isn't obvious, because it's difficult to compare the slices. For example, the red cancer slice appears to be larger than the blue heart disease slice due to the 3-D effect, which has given it more visual weight. Effects such as the 3-D rendering of this pie chart are sometimes used to intentionally mislead.

- *Makes obvious how people should use the information? Partially.* Although the pie chart succeeds in encouraging people to compare the slices to understand the relative contributions of each part to the whole, it fails to support this operation effectively.

Given the ways in which this pie chart has failed to match human perception, let's consider an alternative form of display. The following bar graph displays the same set of values, but in a way that can be more readily perceived.

Total Deaths in America by Cause in 2007

Heart disease	25.42%
Cancer	23.22%
Stroke (cerebrovascular diseases)	5.61%
Chronic lower respiratory diseases	5.28%
Accidents (unintentional injuries)	5.10%
Alzheimer's disease	3.08%
Diabetes	2.95%
Influenza and Pneumonia	2.18%
Nephritis, nephrotic syndrome, and nephrosis	1.92%
Septicemia	1.44%
All other causes	23.81%
Total	100.00%

Figure 35.8

Let's review the effectiveness of this bar graph using the same criteria as before.

- *Clearly indicates the nature of the relationship? Yes.* In and of itself, a bar graph does not declare the part-to-whole nature of the relationship between these values, because, unlike pie charts, bar graphs can be used to display other relationships as well. This particular bar graph, however, includes components that make the nature

of the relationship clear, including the title ("Total Deaths...") and especially the column of values that add up to 100%.

- *Represents the quantities accurately? Yes.* The horizontal position at which each bar ends and the length in relation to the quantitative scale along the x-axis both encode these values in a way that can be accurately perceived. Unlike areas, angles, and the lengths of curved lines that don't share a common baseline, 2-D position and the length of straight linear objects such as these bars, which share a common baseline and run parallel to one another are visual attributes that we can perceive with a high degree of accuracy.

- *Makes it easy to compare the quantities? Yes.* Because we can perceive these values accurately when encoded as bars, it is also quite easy to compare them. Notice how easy it is to see differences in the lengths of these bars that could not be easily seen when comparing the slices of the pie. Also notice that when each bar shares the same color, unlike the pie's slices, which varied in color, our eyes are encouraged to compare the bars because of that likeness. And because the bars are labeled directly with the names of the causes of death, we must no longer do the work that a legend requires when comparing the values.

- *Makes it easy to see the ranked order of values? Yes.* Because differences in the bar's lengths are easy to perceive, the fact that they are ranked from highest to lowest, except for the final "All other causes" bar, is obvious. By arranging the bars in ranked order, we've also made comparisons much easier by placing those causes of death that are closest in value near one another in the graph.

- *Makes obvious how people should use the information? Yes.* The fact that these bars should be compared to understand the varying degree to which these causes of death contribute to total deaths is intuitively obvious.

The point of comparing the perceptual effectiveness of the pie chart and bar graph has not been to make a case against pie charts (although this case deserves to be made), but to illustrate how **we should always judge a visualization's merits by the degree to which we can easily, efficiently, accurately, and meaningfully perceive the story that the information has to tell**. To do this, we must understand the perceptual strengths and weakness of various graphical means for displaying particular stories. To do this, we must understand perception.

35.3 Data Visualization and Human Perception

Data visualization is effective because it shifts the balance between perception and

cognition to take fuller advantage of the brain's abilities. Seeing (i.e visual perception) which is handled by the visual cortex located in the rear of the brain, is extremely fast and efficient. We see immediately, with little effort. Thinking (i.e. cognition), which is handled primarily by the cerebral cortex in the front of the brain, is much slower and less efficient. Traditional data sensemaking and presentation methods require conscious thinking for almost all of the work. Data visualization shifts the balance toward greater use of visual perception, taking advantage of our powerful eyes whenever possible.
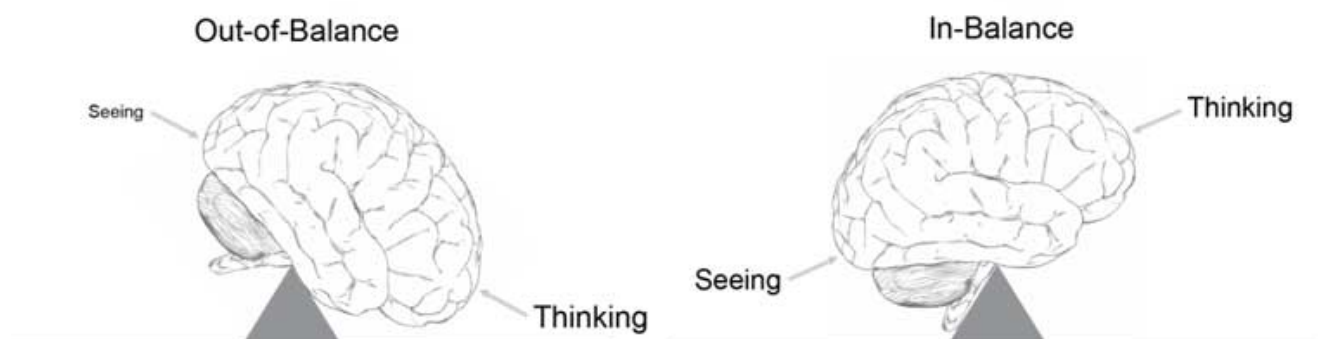
Figure 35.9

One of the earliest contributions to the science of perception was made by the Gestalt School of Psychology. The original intent of this effort when it began in 1912 was to uncover how we perceive pattern, form, and organization in what we see. The founders observed that we organize what we see in particular ways in an effort to make sense of it. The result of the effort was a series of Gestalt principles of perception, which are still respected today as accurate descriptions of visual behavior. Here are a few of the principles that can inform our data visualization efforts:
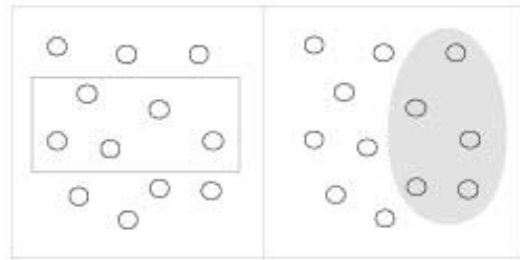
Proximity	Objects that are close together are perceived as a group.	
Similarity	Objects that share similar attributes (e.g., color or shape) are perceived as a group.	
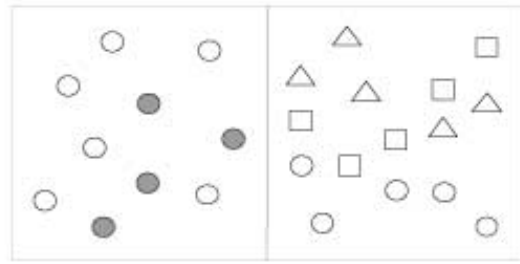
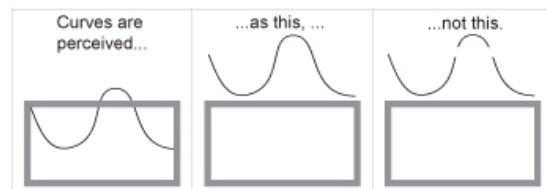| **Enclosure** | Objects that appear to have a boundary around them (e.g., formed by a line or area of common color) are perceived as a group. | |
Closure	Open structures are perceived as closed, complete, and regular whenever there is a way that they can be reasonably interpreted as such.	
Continuity	Objects that are aligned together or appear to be a continuation of one another are perceived as a group.	
Connection	Objects that are connected (e.g., by a line) are perceived as a group.	
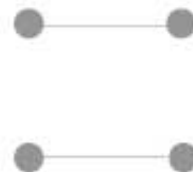
New insights into visual perception and cognition are arising from work in various disciplines besides information visualization, such as human factors and human-computer interaction, but none are more ground-breaking than those arising from the cognitive sciences, especially cognitive psychology. Today, with new and improved

technologies and methodologies for brain exploration, opportunities to improve the perceptual effectiveness of data visualization abound. Two areas of study in particular are especially useful:

- preattentive visual processing
- mechanisms and limitations of attention and memory

One of the great strengths of data visualization is our ability to process visual information much more rapidly than verbal information. Preattentive visual processing is that part that automatically occurs in the brain prior to conscious awareness. It consists of several stages, each handled by specialized neurons that are tuned to detect particular attributes of the visual information contained in light that reflects off the surfaces of objects in the world, which is then stitched together into a picture in our mind's eye of that object. We can use these basic attributes, such as differences in length, size, hue, color intensity, angle, texture, shape, and so on, as the building blocks of data visualization. When we do so in an informed manner, we have the ability to transfer much of the work that is needed to decode the contents of a visual display, such as a graph, from the slower conscious, energy intensive parts of the brain to the faster parts of the brain that require less energy, which results in more efficient cognition.

Studies in attention and memory are revealing our surprisingly limited ability to hold multiple items simultaneously in awareness. This recognition leads us to augment attention and memory by relying on external forms of information storage. One of the most powerful ways to do this is to encode information visually, which allows more information to be chunked together into the limited slots available in working memory. Another method is to place several views of information in front of our eyes at one time, thus extending our ability to explore data multidimensional and from multiple perspectives to make comparisons and see connections to a degree that would be impossible if we had to consume these views one at a time, due to the limits of working memory. Good data visualization techniques and technologies, properly used, can extend our thinking into new realms of analytical sensemaking, and we are still only beginning to tap into this potential.

35.4 Future Directions

What's most needed in the field of data visualization, as in other fields, is not always

what's most exciting or not even what's particularly innovative. Sometimes we simply need to make it easier to do those things that work. One example of this is the effort of a few software vendors to build data visualization best practices right into the tools, such as in the form of defaults, thereby making it easier and less time-consuming to do what works and harder and more costly to do what doesn't. Besides these simple, straightforward but often overlooked improvements, a few other areas offer the potential for enrichment, such as the following:

- The integration of geo-spatial and network displays (such as node and link diagrams) with other forms of display for seamless interaction and simultaneous use.
- Technological support for collaborative data sensemaking to bring the complementary advantage of multiple brains together.
- The application of data visualization beyond descriptive statistics to the realm of predictive analytics, such as through the use of interactive predictive visual models,
- Tighter integration of data mining algorithms to find meaningful patterns with data visualization to provide a better way to review and explore those patterns.
- Improved human-computer interface devices for interacting with data visualization in a more rapid and seamless manner.

All of these are being pursued to some degree, but could be exploited more quickly if more researchers focused on solving real problems that we face in the world today.

35.5 Where to Learn More

Several universities have developed graduate programs that are dedicated to the study and advancement of data visualization. The University of Maryland, Stanford, the University of North Carolina, the University of California, Berkeley, and Georgia Tech are a few of the finest. Although several periodicals in the broader fields of computer graphics and human-computer interaction include articles about data visualization, only one academic journal features the field exclusively: *Information Visualization Journal*, published quarterly by Palgrave Macmillan. A few smaller publications focus on making data visualization practical and accessible to a broader audience, such as the *Visual Business Intelligence Newsletter*. Conferences dedicated to the field are also few. The oldest, IEEE's *VisWeek*, which includes the InfoVis and VAST (Visual Analytics Science and Technology) sub-conferences that are dedicated entirely to data visualization, remains the largest and perhaps best of the conferences, but significant

work in the field also appears in other conferences of broader perspective, such as CHI
(Computer-Human Interaction) and SIGGRAPH.

CHI - Human Factors in Computing Systems

2011 2010 2009 2008 2007 2006 2005
2004 2003 2002 2001 2000 1999 1998
1997 1996 1995 1994 1993 1992 1991
1990 1989 1988 1987 1986 1985 1983
1982

SIGGRAPH - International Conference on Computer Graphics and Interactive Techniques

2002 2001 2000 1999 1998 1997 1996
1995 1994 1993 1992 1991 1990 1989
1988 1987 1986 1985 1984 1983 1982
1981 1980 1979 1978 1977 1976 1975
1974

InfoVis - IEEE Symposium on Information Visualization

2005 2004 2003 2002 2001 2000 1999
1998 1997 1995

Refreshing exceptions, including Tableau Software and TIBCO Spotfire, both spin-offs
of academic work, SAS JMP, which arose from a deep understanding of statistics, and
a few other relatively small vendors, are gradually stealing the attention they deserve
from the big software companies - especially business intelligence vendors - that
dominate the market. Apart from product vendors, a few research laboratories and
consultancies are also contributing to the development and application of the field,
including Microsoft Research, Pacific Northwest National Laboratory, Flowing Media,
Oculus Info, and Perceptual Edge.

Several good books have been written about data visualization. The following, in
chronological order, are especially useful for surveying the field and as a source of
basic instruction:

Tufte, Edward R. (1983): *The Visual Display of Quantitative Information.* Cheshire, CT, Graphics Press	All four of Tufte's books are exceptional, but his first is the best. It makes an inspiring case for graphical excellence.
Cleveland, William S. (1994): *The Elements of Graphing Data.* Hobart Press	Data visualization practices focused on the needs of statisticians.
Harris, Robert L. (2000): *Information Graphics: A Comprehensive Illustrated Reference.* Oxford University Press, USA	An encyclopedic reference for information graphics.
Card, Stuart K., Mackinlay, Jock D. and Shneiderman, Ben (eds.) (1999): *Readings in Information Visualization: Using Vision to Think.* Academic Press	An overview of the best academic research in the field as of the publication date.
Few, Stephen (2004): *Show Me the Numbers: Designing Tables and Graphs to Enlighten.* Analytics Press	An accessible, practical, and comprehensive guide to the design of tables and graphs for communication.
Ware, Colin (2008): *Visual Thinking: for Design.* Morgan Kaufmann	An eloquent introduction to visual perception and cognition as it relates to data visualization.
Few, Stephen (2009): *Now You See It: Simple Visualization Techniques for Quantitative Analysis.* Analytics Press	An accessible and practical guide to data visualization for analysis.

Many blogs and online discussion forums feature data visualization - some thoughtfully, based on expertise, and some with the shallowness that is often found on the Web. Here are a few of the best:

- Tufte.com (Edward Tufte)
- Perceptual Edge (Stephen Few)
- Eager Eyes (Robert Kosara)
- Visual Complexity (Manuel Lima)
- Flowing Data (Nathan Yau)
- Pictures of Numbers (Mike Dickison)

- Instant Cognition (Clint Ivy)

35.6 Commentary by Ronald A. Rensink

35.6.1 Four Futures and a History

Stephen Few provides a nice overview of the reasons why we should design data visualizations to be effective, and why it's important to understand human perception when doing so. In fact, he's done this so well that I can't add much to his arguments. But I can, however, push the basic message a bit further, out into the times before and after those he discusses. Out into areas that are not as well known, or not really developed, where new opportunities and new dangers may lie...

Perhaps the best place to begin is the beginning. Discussing the beginning of visualization is not without its problems, if only for the fact that there exist several different kinds of visualization - for example, data visualization, information visualization, and scientific visualization. But whatever adjective used, we generally find a history more extensive than commonly imagined. For example, although Descartes did contribute to the graphic display of quantitative data in the 17th century, graphs had already been used to represent things such as temperature and light intensity three centuries earlier. Indeed, as Manfredo Massironi discusses in his book (Massironi, 2002; p. 131), quantities such as displacement were graphed as a function of time as far back as the 11th century. But while these facts may be of interest in their own right, the more important point is that techniques in graphic representation have been developed over many centuries, and many of these techniques have been subsequently forgotten - perhaps fallen out of vogue, or never found wide use to begin with. But the reasons for their dismissal may not necessary apply in this day and age. Indeed, several techniques might lend themselves quite well to modern technology, and so might be worth resurrecting in one form or other. Books such as Massironi's are helpful in discovering such possibilities.

On to the future. Or more precisely, on to ways of further developing useful connections between visualization and psychology. To begin with, there is potential for considerably more integration between vision and visualization than currently exists; much more processing could be offloaded to the viewer's visual cortex. As Stephen

Few mentions, one way of doing so is by making use of simple preattentive properties such as length, orientation, and hue. But recent work in vision science has shown that the preattentive level of vision contains far more visual intelligence than that. Among other things, preattentive processes can determine shadows, extract three-dimensional orientation, and link scattered elements of the image into unified groups. These abilities could be exploited in higher-powered visualizations. Another area of recent progress is our understanding of visual attention and scene perception. Our visual perception of the world seems to be based on a just-in-time architecture in which attention is directed to the right object at the right time. If the co-ordination mechanisms involved can be handled correctly, it would open up the prospect of "seeing" abstract datasets in a way that is as natural and effortless as seeing the physical world. (A brief overview of these developments and their implications can be found in Rensink, 2002.)

A related opportunity is the greater use of visual analogy (or metaphor). Here, the emphasis is no longer on bypassing conscious thought, but on using modes of thought best suited for reasoning about visuospatial objects and processes. For example, when reasoning about physical force, a highly useful metaphor is the directed line, or arrow. A more modern example is the desktop, which allows a user to reason about possible actions on their computer. As in the case of visual perception, many - if not most - developments to date have been based on a relatively shallow understanding of the mechanisms involved. But given that cognitive scientists have learned much more about metaphor, it may be time to consider its use in a more sophisticated fashion. Ultimately, visualizations might be able to create mental images that correspond in a natural way to the structure of any process or task. (For an interesting discussion of this, see Paley, 2009.)

A third direction of potential importance is the creation of more powerful evaluation methods based on the methodologies developed in experimental psychology. Psychologists have spent centuries learning what to do (and not to do) to obtain precise measurements of various aspects of human behaviour. It would be good to learn from this. Of course, some of these techniques have already been adapted to evaluation. But as in the case of cognitive and perceptual mechanisms, the transfer of knowledge here is far from complete, and there is much that could still be done. For example, consider evaluating how well a given scatterplot design conveys the correlation in a dataset. In the past, this was done by presenting the viewer with the scatterplot and asking for a numerical estimate of the (perceived) correlation. But a

more powerful approach is to borrow the experimental methodology of measuring just noticeable differences (jnds): the viewer is presented with two side-by-side scatterplots, and asked to choose the more correlated one. Results based on this approach show both precision and accuracy to be specified over all correlations by two functions governed by only two parameters. As a consequence, a given scatterplot design can be completely evaluated based on just two simple measurements. (For details, see Rensink and Baldridge, 2010.)

A final direction to consider - perhaps the most challenging of all - is to develop a systematic way of ensuring that visualization designs make optimal (or at least, good) use of human perception and cognition. In theory, this could result in a "science of design". In practice, this might not be possible, if only because the number of possible designs is so immense and our understanding of human cognition so incomplete. But it may be possible to follow the example of several other areas of design, and aim for a set of principles that would at least constrain the space of possibilities to consider. For example, constraints based on physical forces or material properties can be applied to any architectural design, determining whether or not it is viable. There is no *a priori* reason why a similar approach would not also work for visualization. The efforts of Bertin are perhaps a start in this direction, providing suggestions about the kinds of graphic representation that might be applied to various kinds of problems. Work by Tufte, Mackinlay, Ware, and others have extended this further. But however useful these suggestions are, we are still a long way from a solid foundation for thinking about effective visualizations. Many foundational issues are still poorly understood. What is really going on in a visualization? Is there a way to describe this process precisely and objectively? Is it even possible in principle to determine if a given visualization draws upon the perceptual and cognitive resources of the viewer in an optimal way? The answers to these questions and others like them will be difficult to find. But they will determine the extent to which we can enable humans and machines to best combine their respective strengths.

35.6.2 References

- Massironi, M. (2002). *The Psychology of Graphic Images: Seeing Drawing, Communicating*. Matwah NJ: Erlbaum.
- Paley, W.B. (2009). Interface and mind. *it – Information Technology*, **51**: 131-141.
- Rensink, R.A. (2002). Internal vs. external information in visual perception. *Proceedings of the Second International Symposium on Smart Graphics*: 63-70.

[Smart Graphics 2; Hawthorne, NY, USA.]

- Rensink R.A., and Baldridge G. (2010). The perception of correlation in scatterplots. *Computer Graphics Forum*, 29: 1203-1210.

35.7 Commentary by Naomi B. Robbins

Stephen Few wrote an excellent description of data visualization and the necessity for designing graphics to take advantage of our knowledge of human perception and cognition. In this commentary I question who is responsible for the myriad of visualizations that ignore this knowledge: the software vendors, the software users or others? In addition, I point out important work that deserves greater exposure on the integration of geo-spatial and other forms of data display, a topic on Few's most-needed list. I end with additional sources for learning more.

35.7.1 Responsibility for perceptual problems with many data visualizations

Few's article states:

> 66 Since the turn of the 21st century, data visualization has been popularized, too often in tragically ineffective ways as it has reached the masses through commercial software products. 99

Certainly, software vendors are responsible for offering many graph forms that hinder rather than help the reader to understand the data. The vendors offer graphs to wow the audience rather than to communicate clearly and they create demand for ineffective graphs. But they are not solely responsible for the myriads of graphs with perceptual problems.

People learn from what they see and they see many ineffective graphs. The software users then demand software that allows them to imitate these ineffective designs. This gets us in a chicken and egg situation: Do vendors produce these awful visualizations because their customers demand them, or do the customers become attracted to them when they see what vendors market?

An example of the ineffective ways includes pseudo-third dimensions in bar charts. Figure 1 shows a pseudo-three-dimensional bar chart in Excel. Almost no one reads it

correctly. I describe other problems with this graph in *Creating More Effective Graphs* [1].
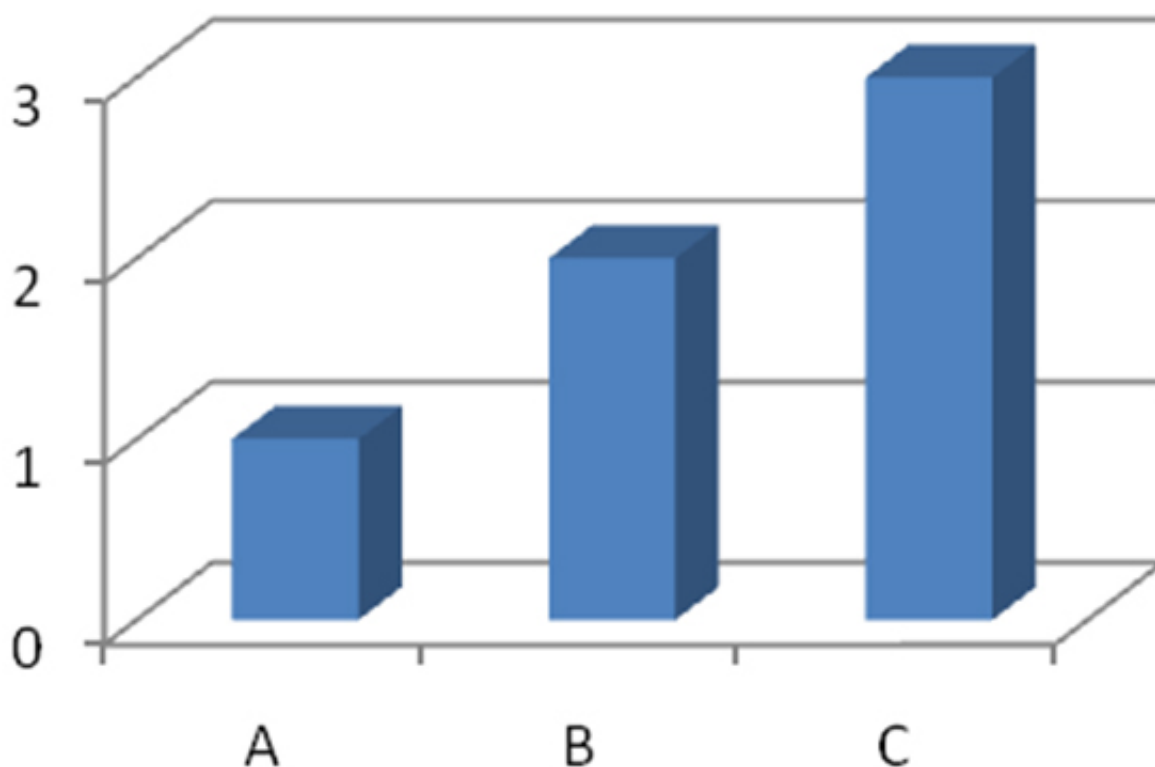
Figure 35.1: Almost no one reads this simple chart correctly. The numbers plotted are 1, 2, and 3. Plot it yourself in Excel if you don't believe me

A number of graphic artists have made major contributions to the field of data visualization. However, there are some graphic artists who have no appreciation of numbers and don't realize that the representation of numbers in graphs should be proportional to the numbers they represent. As a result, it is common to see graphs that are not drawn to scale.

Some graph designers want to give the impression of better performance than is actually the case and intentionally design graphs that mislead to achieve this impression. Other graph designers may be more concerned with demonstrating their technological abilities or artistic abilities than in communicating clearly and accurately. Until recently, our educational system did not provide training in communicating numbers. Today, there are some excellent courses at the college level but the majority of people receive little, if any, training in presenting numerical information. Therefore,

many graph designers are unaware of the principles of effective graphs. Some of the problems occur from a lack of proofreading and careless errors.

As an analogy, a current style in fashion is high-heeled shoes. A quick search on "dangers of high heels" revealed that there has been an increase in the number of bunion operations on wearers of high heels as well as foot pain, back pain and neck pain. In some cases the Achilles tendon grows shorter. Balance is affected so that the risk of falls is greater. The list of problems goes on and on. Is the shoe designer, the shoe manufacturer, the retail outlet that sells the shoes or the customer who buys them responsible for this increase in medical problems? Is this situation analogous to the data visualization one? Both cause serious problems: poor business decisions in one case and pain and suffering as well as unnecessary medical expenses in the other. I hope that these questions stimulate interesting discussion.

35.7.2 Integration of geo-spatial displays with other forms of display

In his section on future directions, Few mentions areas that offer the potential for enrichment including the integration of geo-spatial displays with other forms of display for seamless interaction and simultaneous use. Several researchers have made advances in this area. For example, the micromap designs of Dan Carr [1] and [2] add a geographic context to statistical information, allowing for the joint exploration of statistical and geographic patterns in data. As illustrated in Figure 2, statistical graphics, here dots, are linked to small maps by color. In the first row, we can see that Maryland is represented by red dots and so Maryland is shaded red on the right-hand map. Sorting by poverty level, we see that not only are poverty and education inversely related, but that there is a geographic clustering of southern U.S. states by these variables.
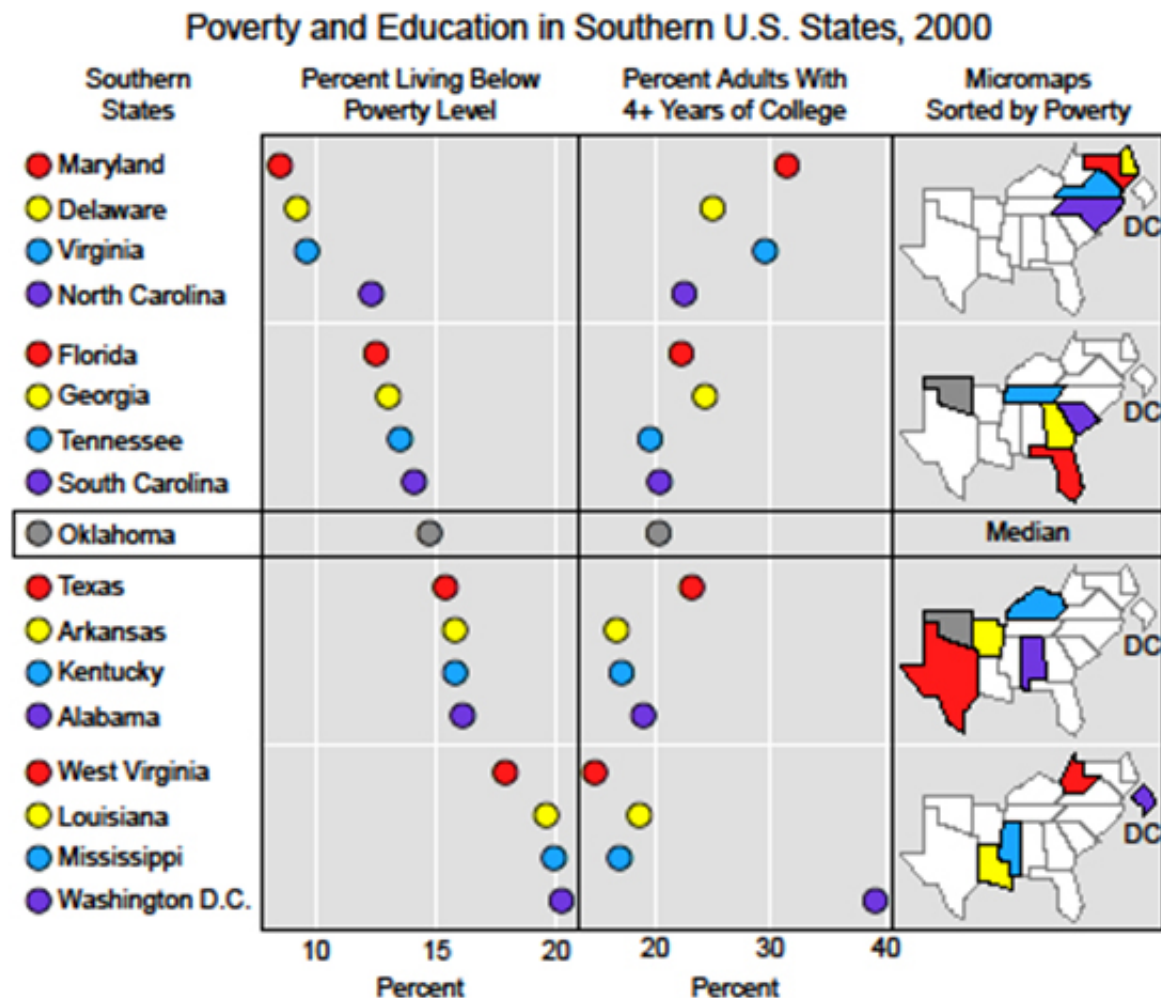
Figure 35.2: An example of a micromap design from Carr and Pickle [1]

35.7.3 Where to Learn More

Data visualization does not belong to a single academic discipline. Statisticians, computer scientists, psychologists, graphic designers and others practice and contribute to data visualization. The university programs and resources that Few mentions lean heavily towards computer science. A few excellent programs joining statistical graphics with computer science are available at George Mason University, Iowa State, and the University of Augsburg. There are many others. I will leave it to other commentators to add excellent programs in cognitive psychology and graphic design. The Journal of Computational and Graphical Statistics, a joint publication of the American Statistical Association, the Institute of Mathematical Statistics and the Interface Foundation of North America is another academic journal on the topic. The Statistical Computing Statistical Graphics Newsletter (SCGN) is another informal publication. Although the Joint Statistical Meetings are not exclusively devoted to

statistical graphics and data visualization, there are as many sessions sponsored by the Statistical Graphics Section as many a smaller conference contains.

One addition I would make to the "what's needed" list is better communication between the computer scientists, graphic designers, psychologists and statisticians. More joint conferences and attending each other's conferences would help each discipline benefit from the research of the others.

35.7.4 References

1. Carr, Daniel B. and Linda Williams Pickle. 2010. *Visualizing Data Patterns with Micromaps.* Chapman and Hall/CRC, Boca Raton, FL.
2. Carr DB, Wallin JF and Carr DA. Two new templates for epidemiology applications: Linked micromap plots and conditioned choropleth maps. *Statistics in Medicine* 19:2521-38, 2000.
3. Robbins, Naomi B. 2005. *Creating More Effective Graphs*. John Wiley and Sons, Hoboken, NJ.

35.8 Commentary by Robert Kosara

35.8.1 Metaphors and Interaction

One important topic Stephen Few only mentions briefly in his very well-written and comprehensive piece is interaction. While static charts and visualizations are undoubtedly useful, they make little use of the immense computing power that is readily available to us today. Interaction in visualization enables the fast exploration and discovery of data patterns that the user may not even have expected. It is also possible to reduce the amount of data shown at the same time, providing clearer visualizations, while still giving the user the option to get that information on demand at any time.

Ben Shneiderman captured the role of interaction in his famous visual information seeking mantra (Shneiderman, 1996): *overview first, zoom and filter, then details on demand*. Abstract information spaces require an overview so the user has an idea where to even find data, but then it is necessary to zoom in to see details. Filtering data is important when dealing with larger datasets. Finally, details on what is shown (and also what is not shown) can be retrieved by the user as needed. All of these steps require interaction, where the user tells the visualization what he or she wants to see.

35.8.2 Simple Interactions

Among the simplest interactions are tooltips or other data displays that appear when the user points at a part of a visualization. Take the causes of death bar chart in Few's article above: the numbers could be shown purely on demand, perhaps including not just percentage but total number. Also, a vertical line could be drawn from the end of the active bar to the scale at the top, to make it easier to see the bars in context.

This type of interaction is effortless and easy to discover: just move your mouse over the display and see if anything happens. Displaying numbers in charts is also rather common. But the real power comes from the more advanced interactions.

35.8.3 Linking and Brushing

Brushing lets the user selects data points that get highlighted in one or more views of the same data. When several views are involved, the fact that all of them highlight the same data points is commonly referred to as *linking* (and the views are called *coordinated multiple views*). Consider this example of linked bar charts of data about passengers on the Titanic. Each bar chart represents one data dimension (class, gender, age, and survived), and shows a histogram of how many people were in each of the categories.
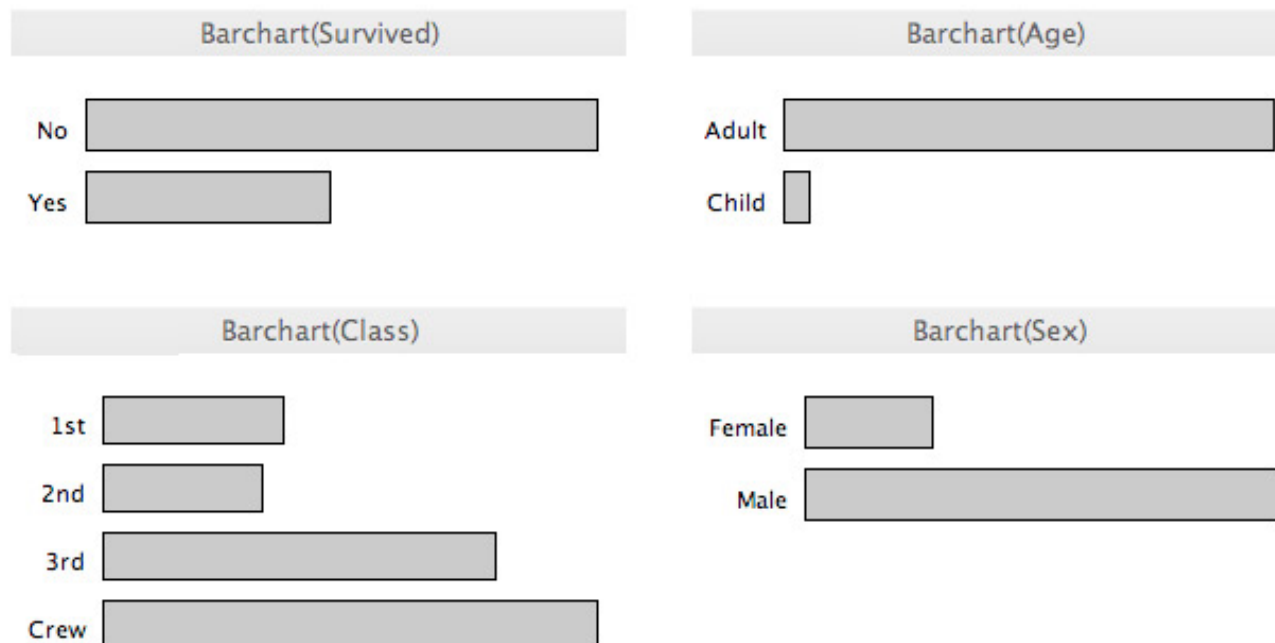
Courtesy of Robert Kosara. Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms below.

Figure 35.1

To find out how many people survived in each category, we will select the relevant bar, which will brush those data points in all the views. We can now compare survival rates for different sexes, classes, etc. by looking at how much of their respective bars is highlighted.
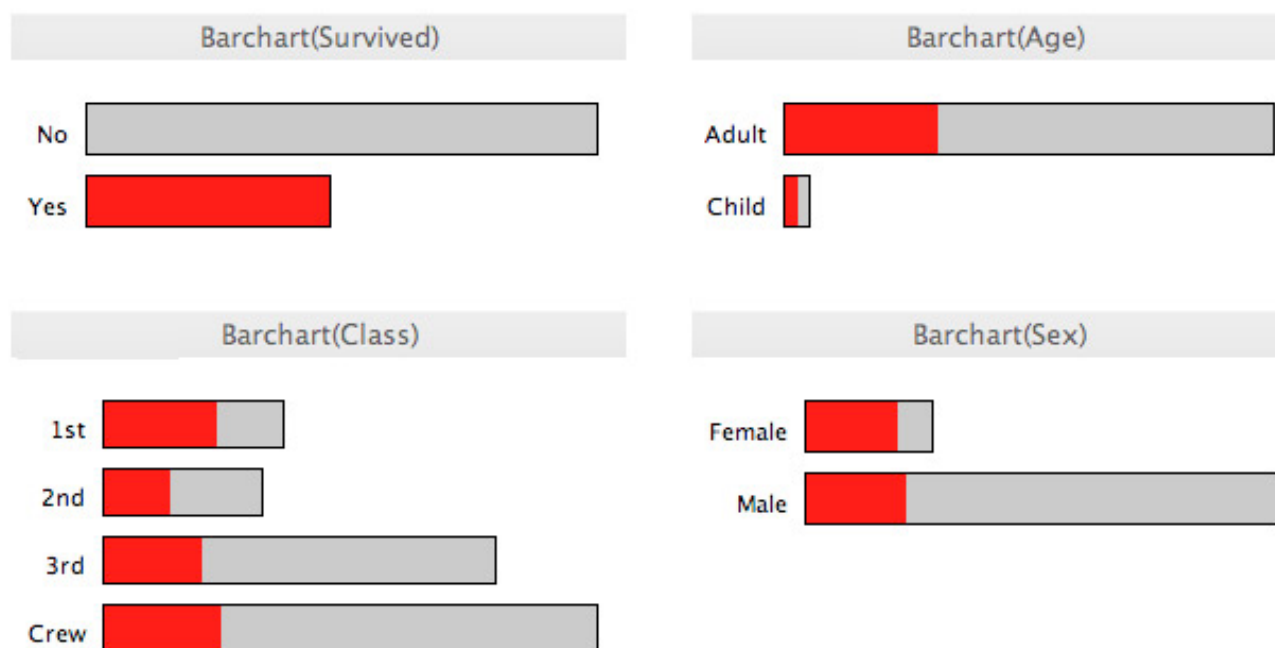
Courtesy of Robert Kosara. Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms below.

Figure 35.2

The mechanism is very similar for individual data points rather than summary data like in this example. Brushing and linking make it possible to find out high-dimensional relationships in the data by trying out different possibilities.

35.8.4 Metaphors and Structure

Metaphors have a somewhat complicated history in visualization. There is not even a clear understanding what a metaphor even is: many people talk about visual metaphors when they mean different ways of depicting data, but others use them specifically for somewhat embellished visualizations (flowers growing to represent traffic in chat rooms, etc.).

What I want to add here is a combination of both, perhaps best summarized as structure: how do the relationships between elements in the visualization influence how people read the data? Caroline Ziemkiewicz and I have done work on this topic, and have found that the big-picture structure plays a bigger role than most people would assume.

When comparing different types of tree visualizations, we found that different studies had come to different conclusions as to which method works better based on which metaphor was used in the question: A being contained in B, or A being below B in the hierarchy. We did a study and found that there was, indeed, a compatibility effect between the linguistic metaphor used in the question and the visual metaphor of the visualization (Ziemkiewicz and Kosara, 2008).

We recently showed that there is an apparent effect of gravity between objects in a visualization that can distort the perception of distance (Ziemkiewicz and Kosara, 2010).

35.8.5 The Future

While we know a lot about how to create reasonable visualizations, there is still a lot we do not know or are not yet aware of. Even seemingly basic knowledge like how the layout of a visualization influences our reading of the data still needs more work to be understood and turned into useful recommendations and best practices.

Interaction is not exactly a new topic in visualization research, but is still rather rudimentary in many visualization and charting programs. To really unlock the power of

visualization, these programs will need more advanced capabilities as well as ways to educate their users about their interactive features. Visualization has a lot more to offer than what most people are aware of today.

35.8.6 References

- Shneiderman, Ben (1996): The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In: VL 1996 1996. pp. 336-343.
- Caroline Ziemkiewicz, Robert Kosara, The Shaping of Information by Visual Metaphors, Transactions on Visualization and Computer Graphics (Proceedings InfoVis), vol. 14, no. 6, pp. 1269-1276, 2008.
- Caroline Ziemkiewicz, Robert Kosara, Laws of Attraction: From Perceived Forces to Conceptual Similarity, Transactions on Visualization and Computer Graphics (Proceedings InfoVis), vol. 16, no. 6, pp. 1009-1016, 2010.

<   Experimental Methods in Human-Computer Interaction

Human-Robot Interaction   >

Learnt something new? Share with your friends:

in   y   f   G+   ✉

35.10 References

The Encyclopedia
of Human-Computer
Interaction

The Encyclopedia of Human-Computer Interaction, 2nd Ed.

BY MADS SOEGAARD AND RIKKE FRIIS DAM (EDS.)

READ OFFLINE

Read this book offline by downloading it or buying a hardcopy version.

PDF version   >

CHAPTER CONTENTS

AUTHOR(S)

Stephen Few

Stephen Few has over 20 years of experience as an innovator, consultant, and educator in the fields of business intelligence (a.k.a. data warehousing and decision support) and information design. Through his company, Perceptual Edge, he focuses on the effective analysis and presentation quantitative business information. Stephen is recognized as a world leader in the field of data visualization. He teaches regularly at conferences such as those presented by The Data Warehousing Institute (TDWI) and DCI, and also in the MBA program at the Haas School of Business at U. C. Berkeley. He is also the author of the book "Show Me the Numbers: Designing Tables and Graphs to Enlighten" (Analytics Press).

View profile

COMMENTARIES BY

Ronald A. Rensink

I am interested in vision-the various ways that humans, animals, and computers use light to see. I believe that vision involves constraints that apply to any system, and that the most successful visual systems are based on very general

information-processing strategies. As such, my approach is to examine biological systems (including humans) to see how they operate, and then to look at these mechanisms from a computational point of view to see if they embody more general principles. Among other things, these more general principles can provide a scientific basis for

the design of visual interfaces that can interact with human visual systems in an optimal way.My research interests
include: 1. Human vision what is attention, and how does it operate? what is space, and how do we represent it? what
are objects, and how do we represent them? how are scenes represented? 2. Computational vision how do "quick an…

View profile

Naomi B. Robbins

Naomi B. Robbins is the author of Creating More Effective Graphs, published by John Wiley (2005). She is a
consultant, keynote speaker, and seminar leader who specializes in the graphical display of data. She trains
employees of corporations and organizations on the effective presentation of data. She also reviews documents and
presentations for clients, suggesting improvements or alternative presentations as appropriate. Naomi received her
Ph.D. in mathematical statistics from Columbia University, M.A. from Cornell University, and A.B. from Bryn Mawr
College. Dr. Robbins was an officer of the Statistical Graphics Section of the American Statistical Association (ASA).
She has served the New Jersey Chapter of the ASA as President, Vice-President, Secretary, Treasurer, and Chair of the
Advisory Committee, and was the first chapter member to be awarded the Chapter Service Award. She had a long…
career at Bell Laboratories before forming NBR, her consulting practice.

View profile
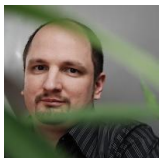
Robert Kosara

I am an Assistant Professor at the Department of Computer Science, College of Information Technology, at the
University of North Carolina at Charlotte (UNCC), where I am also a member of the Charlotte Visualization Center. I
received both my Ph.D. (2001) and M.S. degrees from Vienna University of Technology (Vienna, Austria). Before
coming to Charlotte, I worked at the VRVis Research Center and the in-silico pharmaceutical research company
Inte:Ligand. My research is in Information Visualization (InfoVis) and Visual Analytics. The goal of these fields is to
translate data into images that we can interact with and read to understand the underlying data.

View profile

UPCOMING COURSE

Conducting Usability Testing

STARTS IN

1 | **23** | **44** | **56**
days | hrs | mins | secs

93 %
BOOKED

VIEW COURSE

Web Design for Usability

STARTS IN

2 | **23** | **44** | **56**
days | hrs | mins | secs

90 %
BOOKED

VIEW COURSE