

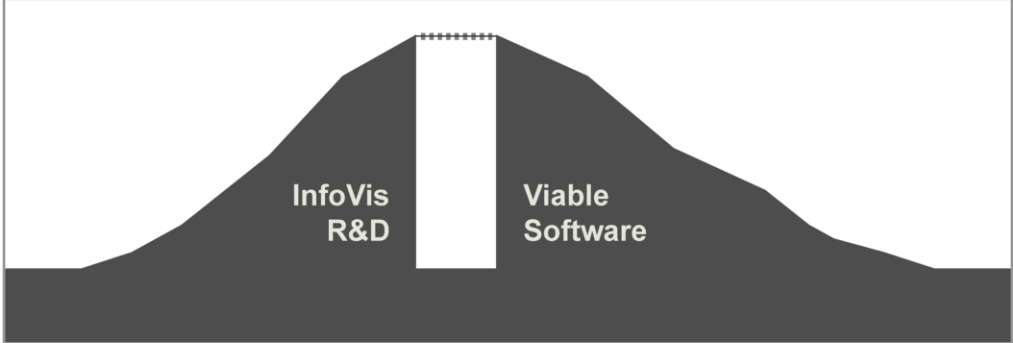
Bridging the Chasm

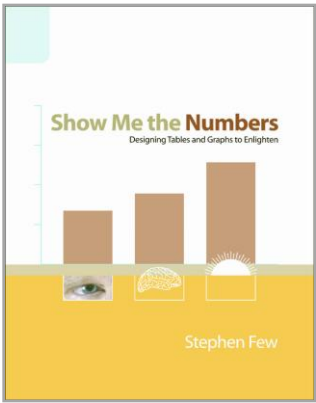
Between Infovis and the World Out There

From the theoretical to the practical,
the innovative to what really works,
and the cool to what's really needed

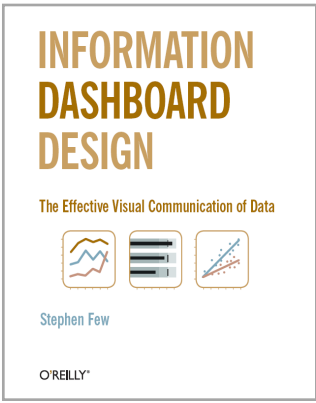
Stephen Few

Perceptual Edge and the University of California, Berkeley

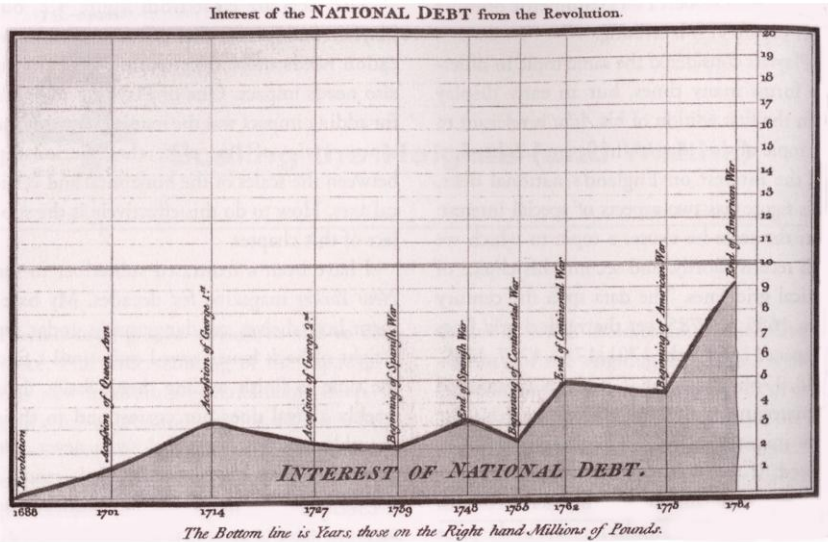




Stephen Few
Principal, Perceptual Edge
Author of
Show Me the Numbers
and
Information Dashboard Design



Data visualization is not new.



William Playfair – 1786

In 1786, a roguish Scot – William Playfair – published a small atlas that introduced or greatly improved most of the quantitative graphs that we use today. Prior to this, graphs of quantitative data were little known.

Today, 220 years later, graphs are common.

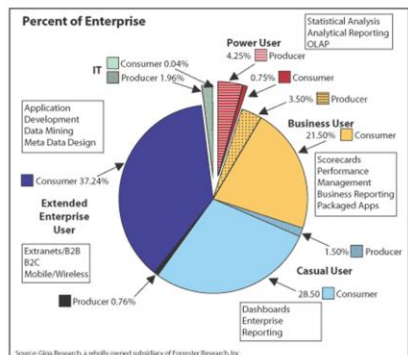
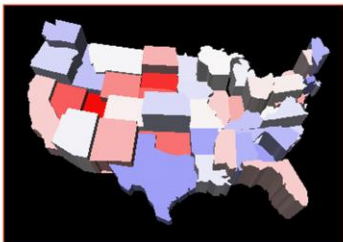
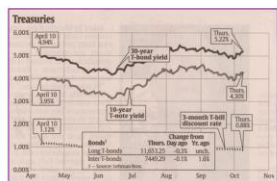
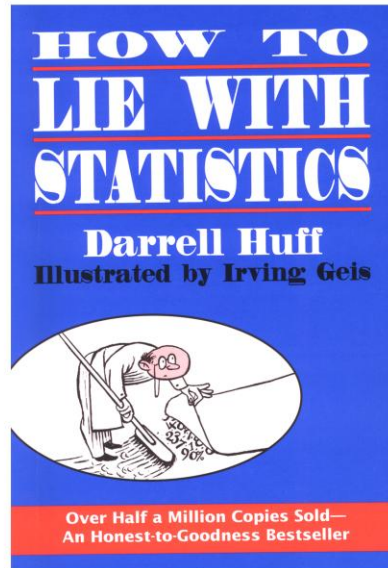


Figure 2: Analytic End-User Spectrum of Usage

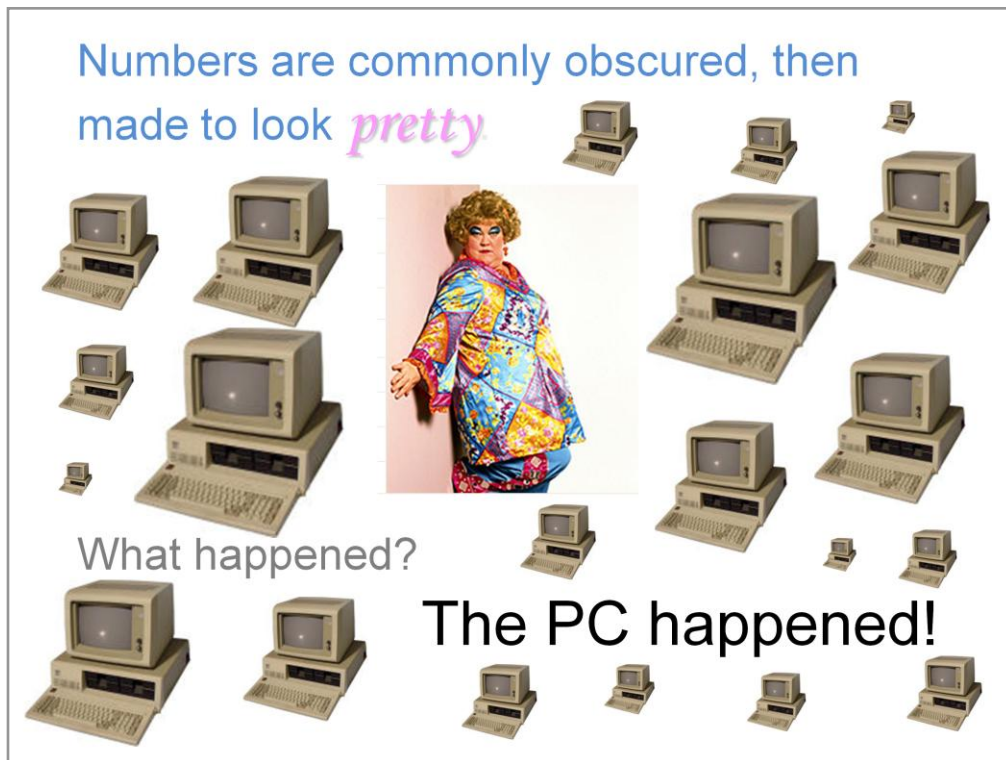
Today, 220 years later, partly due to the arrival of the PC, graphs are commonplace, fully integrated into the fabric of modern communication. Surprisingly, however, Playfair's innovative efforts – sprung from meager precedent – are still superior to most of the graphs produced today.

A powerful language,
with such promise,
is largely being wasted!

Intentional deceit is no longer our biggest problem.



In 1954, Darrell Huff wrote his best-selling book about how people often intentionally use graphs to spread misinformation, especially in favor of their own products or causes. Today, vastly more misinformation is disseminated unintentionally because people don't know how to use charts to communicate what they intend.



When the PC was introduced, software soon made the arduous task of table and graph creation as easy as 1-2-3 (literally "Lotus 1-2-3", the software that was the first to legitimize the PC as a viable tool for business). Unfortunately, this improvement in ease and efficiency was not accompanied by instruction in visual design for communication. People today think that if they know how to click with the mouse to create a table or graph, they know how to present data effectively.

"In the two centuries since [the invention of the first graphs], ...charts have become commonplace. With the advent of modern computer tools, creating graphs from data involves trivial effort. In fact, it has probably become too easy. Graphs are often produced without thought for their main purpose: to enlighten and inform the reader."

Jonathan G. Koomey, *Turning Numbers into Knowledge*, Analytics Press, 2001.

Why do I care?

- People desperately need what only we can provide
- We have an opportunity to contribute something worthwhile to the world

What is the goal?

To provide InfoVis products and techniques that
really work and really matter

What keeps us from this goal?

- Many infovis researchers lack the necessary aptitudes, expertise, and skills
- Much infovis research is poorly done
- Much infovis research doesn't matter
- Much infovis development doesn't work

"I have been increasingly bothered by the lack of reality in academic research. University-based research can be clever, profound, and deep, but surprisingly often it has little or no impact either upon scientific knowledge or upon society at large. University-based science is meant to impress one's colleagues: What matters is precision, rigor, and reproducibility, even if the result bears little relevance to the phenomena under study. Whether the work has any relevance to broader issues is seldom addressed...Most academic study is designed to answer questions raised by previous academic studies. This natural propensity tends to make the studies ever more specialized, ever more esoteric, thereby removed even further from concerns and issues of the world."

Donald A. Norman, *Things That Make Us Smart: Defending Human Attributes in the Age of the Machine*, 1993, Basic Books, New York, pp. xii and xiii.

People have many perceptions of InfoVis;
many are intimidating and inaccurate.

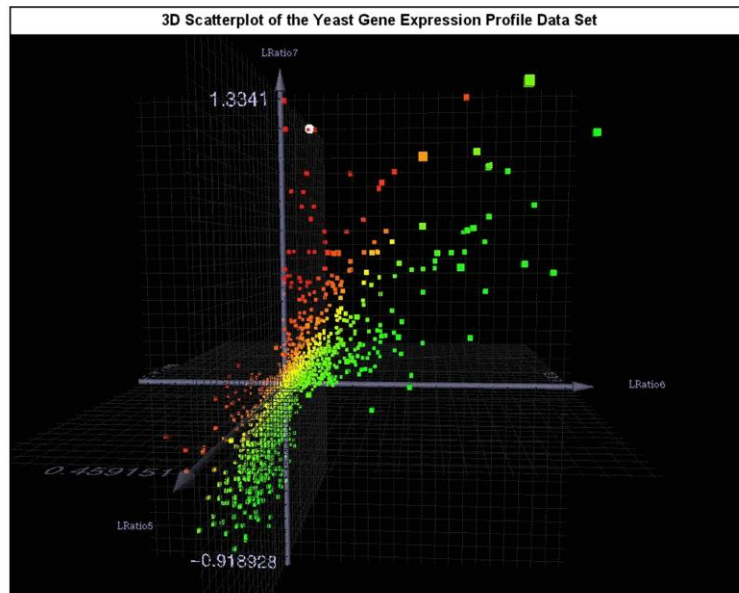
Infovis?

Mysterious scientific displays



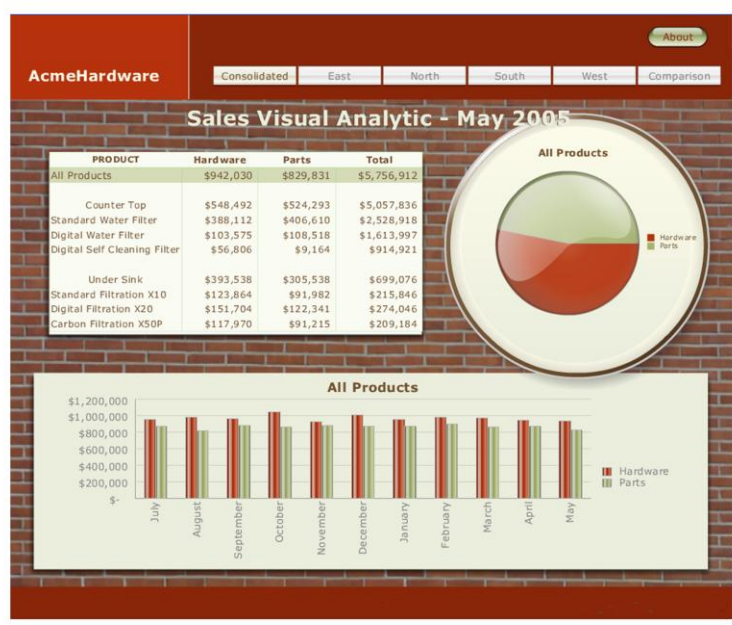
Some people associate data visualization with complex scientific visualizations, such as X-rays, which can seem deeply mysterious even to the brightest of us, unless we've been training in their interpretation.

Complex statistical displays



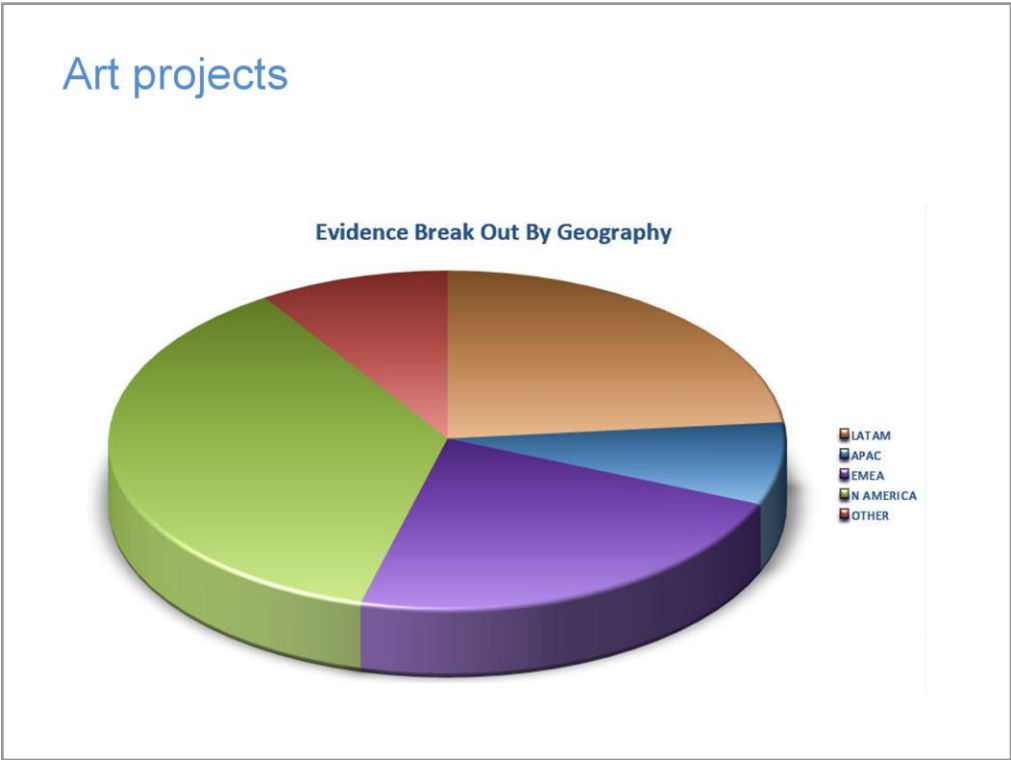
Some people associate data visualization with sophisticated statistical visualizations, lined with mathematical formulas and unfamiliar terms, which can intimidate even the bravest of us, unless we've studied advanced statistics.

Flashy, video-game-like displays



Many people associate data visualization with popular business software such as vibrant dashboards with their video-game-like displays of gauges and meters, and recognize them for the superficial displays that they often are. This is disheartening for those who demand more substance and less dazzle. Far too many buyers of business intelligence software fall for dazzling claims and demonstrations—pure fluff when they need real substance. Visualizations that jump off the screen with bright colors and dancing widgets might be entertaining at first, but quickly become just plain annoying when they fail to deliver real value.

There are products out there that give data visualization a bad name, but beyond these there exists tremendous potential that is seldom tapped. Some let a bad meal from a fast food restaurant doom them to a lifelong fast.



No software is more often used to make sense of data than Excel. Like it or not, Excel sets expectations for what software ought to do. With the release of Excel 2007, Microsoft introduced an entirely new charting engine, which gave them the chance to dramatically improve the products data visualization capabilities. Unfortunately, other than providing the ability to make prettier charts, the new release of Excel failed to improve the product. Its emphasis on eye-candy encourages people to dress their presentations up as flamboyant displays, rather than focusing on presenting information clearly and accurately.

People who need us don't always know this is true.

Definition of Business Intelligence (BI):

An interactive process for exploring and analyzing structured and domain-specific information to discern trends or patterns, thereby deriving insights and drawing conclusions.

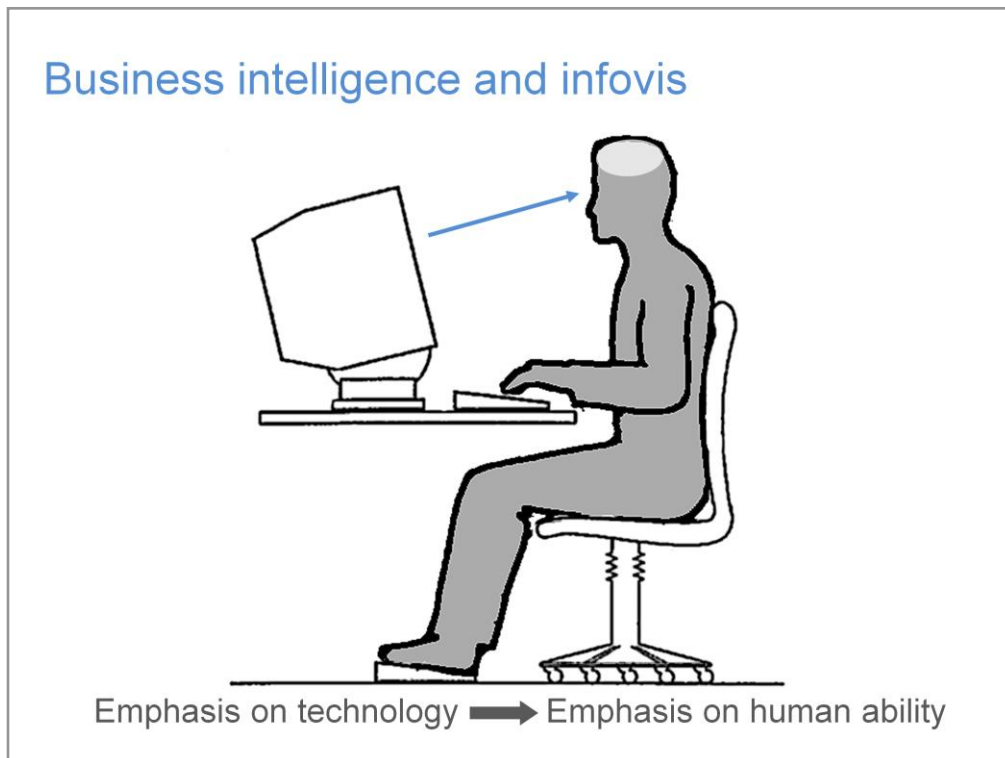
(Source: Glossary at www.gartner.com)

The part of the software industry that provide products for making sense of and presenting business information is call business intelligence (BI).

Gartner , the organization that introduced the term, defines BI as:

An interactive process for exploring and analyzing structured and domain-specific information to discern trends or patterns, thereby deriving insights and drawing conclusions.

This is precisely what infovis does, yet visualization is only now becoming respected and sought after aspect of business intelligence.



To date, business intelligence has focused on technology and project methodology, resulting in great advances. Today, we have huge and fast warehouses of information. Now it's time to focus on the true essence of business intelligence – important, meaningful, and actionable information – and the most powerful resources for tapping into its value are those that engage the tremendous capacities of *human perception and intelligence* to make sense of and communicate information.

Many organizations aren't effectively analyzing the data they do have to improve their business.

What's more troubling, perhaps, is many companies that purchase powerful analysis and business-intelligence tools don't use them effectively. Users of these products often generate the most basic and obvious reports and never get their hands dirty with the deep-analysis tools.

By ignoring these products' deep-analysis capabilities, organizations could be missing important trends—information that might show, for example, where a company is losing business.

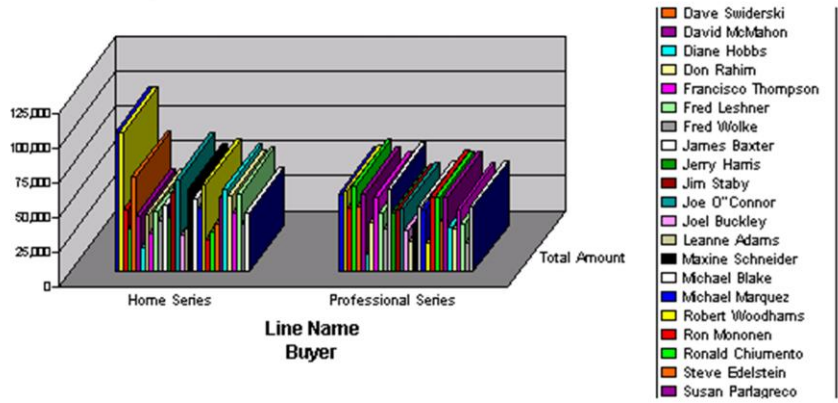
Companies might also see that business decisions that were made based on basic information were wrong and ended up costing money in the long run.

eWeek, "With Data Analysis, Less Isn't More, Jim Rapoza, Ziff Davis Media Incorporated, June 7, 2004.

Most Americans are poor quantitative thinkers...When people are innumerate – when they do not know how to make good use of available quantitative information – they make uninformed decisions.

What the Numbers Say: A Field Guide to Mastering Our Numerical World, Derrick Niederman and David Boyum, Broadway Books, New York, 2003, page 229.

Most business intelligence vendors know almost nothing about infovis.



Their attempts are usually primitive and ineffective

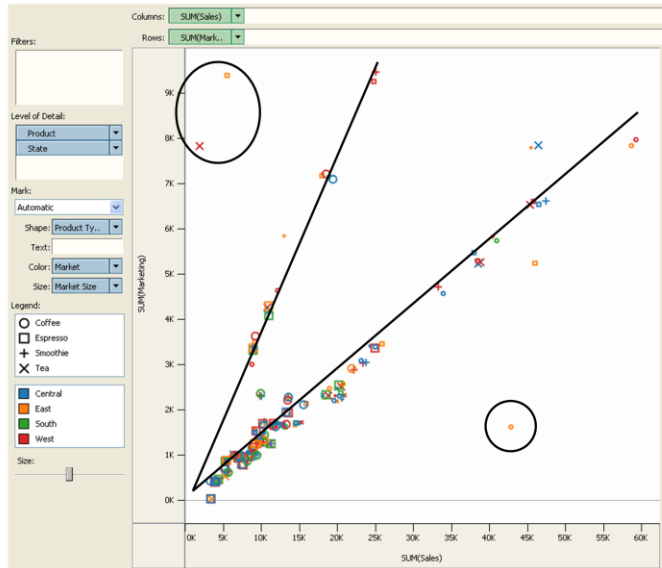
Traditional BI relies heavily on text displays.

| | | | | | |
|-------------------|-----------------|---------|---------|--------|---------|
| Columns: Market | | | | | |
| Rows: Product T.. | | Product | | | |
| | | Central | East | South | West |
| Coffee | Amaretto | 44,144 | 9,415 | | 29,427 |
| | Columbian | 91,938 | 152,639 | 69,694 | 97,431 |
| | Decaf Irish C.. | 82,595 | 20,684 | 36,875 | 57,865 |
| Espresso | Caffe Latte | | | 49,501 | 65,605 |
| | Caffe Mocha | 111,979 | 52,483 | 45,039 | 60,271 |
| | Decaf Espres.. | 77,861 | 25,013 | 48,772 | 96,534 |
| | Regular Espr.. | | 77,142 | | |
| Smoothie | Blueberry | 30,547 | 38,615 | | 45,110 |
| | Chocolate | 117,096 | 6,947 | 35,646 | 82,569 |
| | Lemon | 70,092 | 84,456 | 46,454 | 101,953 |
| Tea | Darjeeling | 96,567 | 44,755 | | 90,752 |
| | Earl Grey | 105,290 | 20,854 | | 89,036 |
| | Green Tea | 16,737 | 36,436 | | 51,415 |

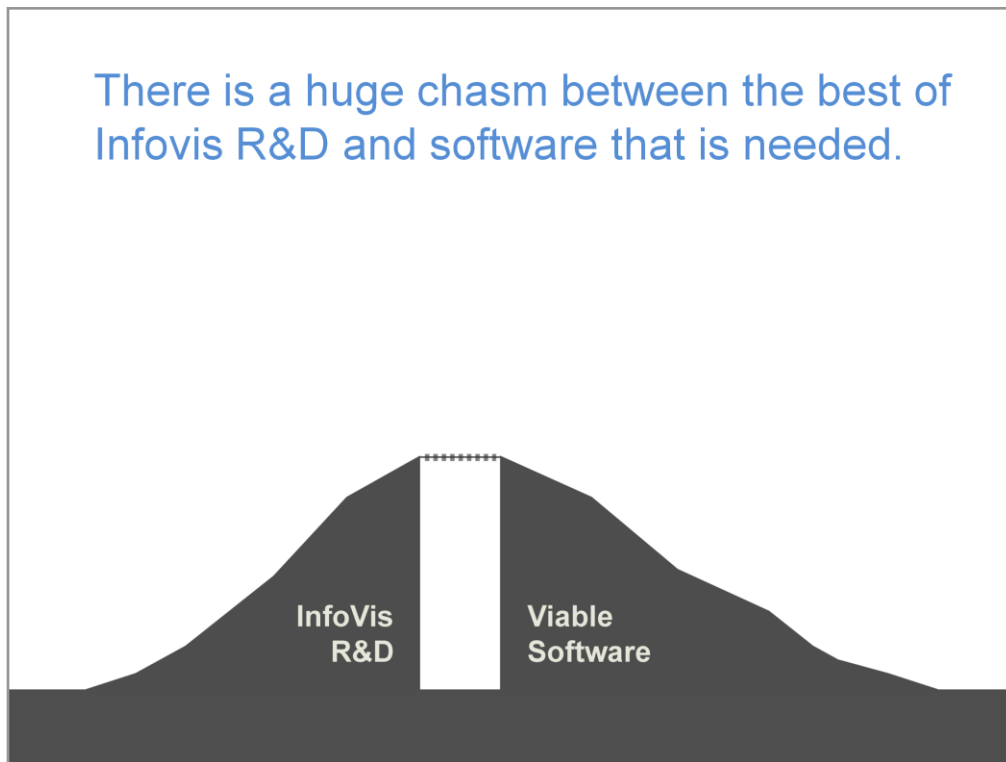
Traditional BI relies too heavily on text-based data displays, primarily on tables, such as pivot tables. Tables of text are powerful tools for some tasks, but they don't do a good job of revealing trends, patterns, and exceptions that are best shown by giving shape to the data, thus making them easy to see, which is precisely what visualizations do.

In this example of beverage sales by region and product, the tabular display must be read in its entirety to answer even simple questions such as, "for what products and in which regions are sales the greatest?" This is a slow process, even with such a small table, and you must rely heavily on memory and be very careful to get it right.

Infovis brings otherwise invisible patterns, trends, and exceptions to light.



Properly designed visualizations, however, can answer more complicated questions in an instant.



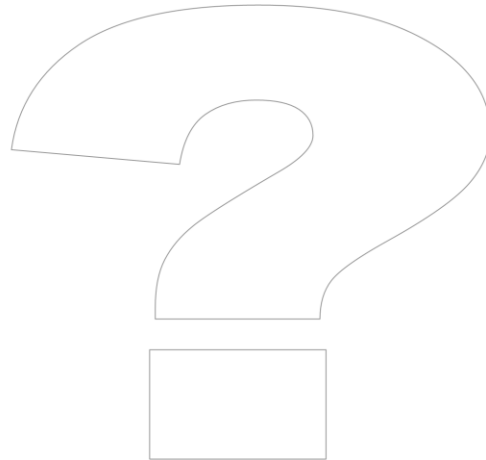
Too little infovis research and development finds its way into software products that are broadly used. Researchers who want their work to make a difference in the world can take practical steps to bridge the chasm that separates them from the new or existing software products that could incorporate their work.

A wide chasm continues to separate infovis researchers from the software vendors that could benefit from their work. Software companies (including most business intelligence vendors) are responsible for half of the chasm. With few exceptions, they know little about information visualization and are doing almost nothing to change this. Researchers, however, are responsible for the other half. This is largely due to three failures:

- Much infovis research has no practical application.
- Much infovis research produces incomprehensible visualizations and ineffective functionality.
- Much infovis research is not presented in an understandable and compelling manner.

Researchers who want their work to matter must respond to the actual needs of people. Pursuing what interests you is great if it happens to coincide with what's needed; otherwise, it might end up mattering to no one but you. Producing work that looks cool but is incomprehensible is a similar waste. Perhaps the greatest travesty, however, is great research that matters and works, but remains on the shelf because you've failed to present it in way that informs and inspires.

What aptitudes, expertise, and skills do
infovis researchers need to bridge this gap?



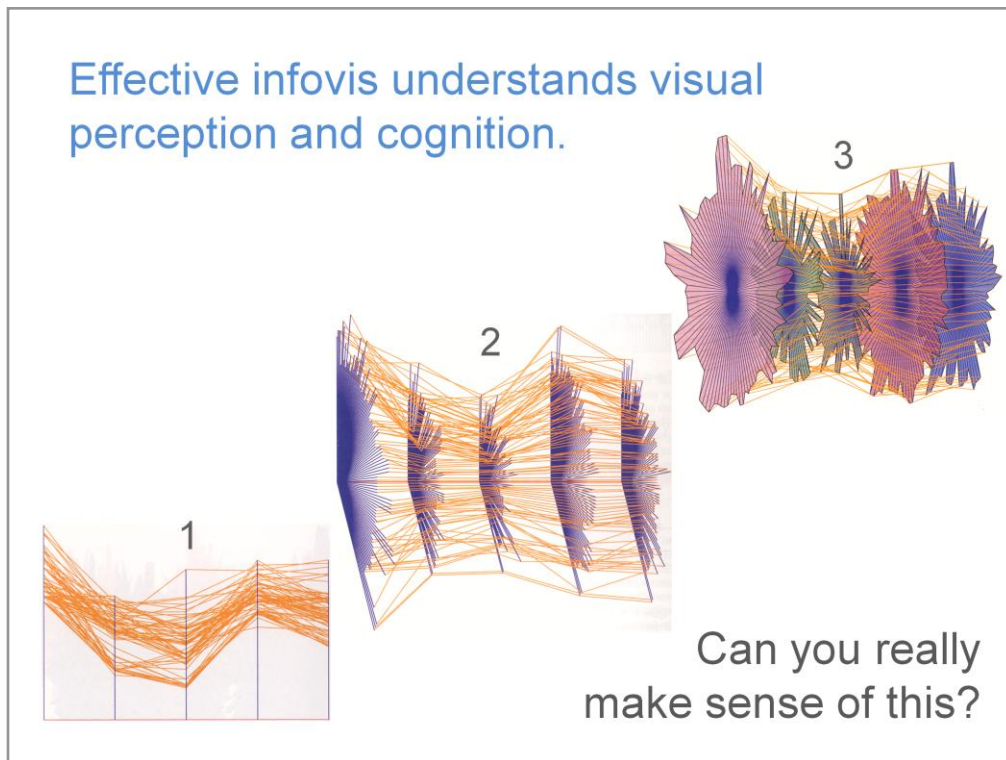
Here are a few that I think are important, but are too often missing:

- Knowledge of visual perception, as it applies to data presentation
- Knowledge of data analysis, including actual experience doing it
- Knowledge of what people need (both sophisticated, statistically trained users and those who are not)
- Design orientation
- Knowledge of the scientific method and how to design a good research study
- Ability to communicate in plain language

Useful References:

Things That Make Us Smart: Defending Human Attributes in the Age of the Machine, Donald Norman

The Design of Everyday Things, Donald Norman



“A general shortcoming [of information visualization] is that, up to now, mainly technical issues have been the focus of discussion. The prerequisites of the user for dealing with information visualizations and making sense of visualizations have not gained much attention in the past. It is important to develop new technologies in alignment with the changing demands of the user, because the user is the one who has to interact with the information visualizations. Therefore, it is necessary to include the experience and know-how of more user-oriented sciences, like Psychology. According to Marshall (2001), information visualizations often lack comprehensibility.”

Sigmar-Olap Tergan and Tanja Keller, Editors, *Knowledge and Information Visualization*, 2005, Springer-Verlag, Berlin Heidelberg, page 11.

Source of the images above: Elena Fanea, Sheelagh Carpendale, and Tobias Isenberg, “An Interactive 3D Integration of Parallel Coordinates and Star Glyphs”, IEEE Symposium on Information Visualization 2005.

Effective infovis can be decoded without overwhelming effort.

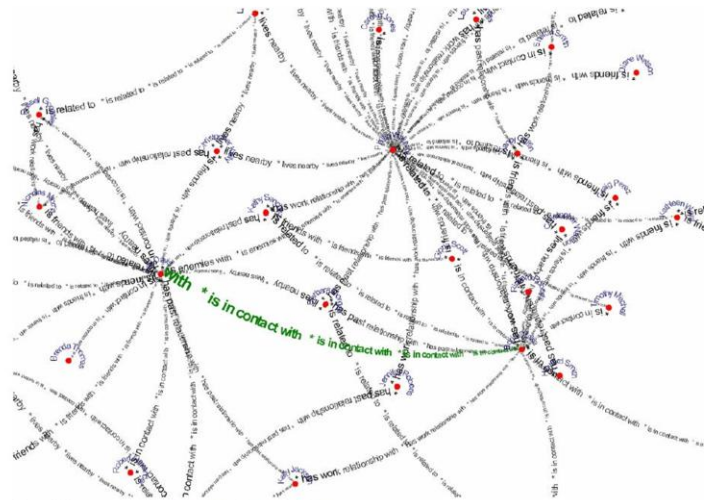


Figure 6: A screen snapshot of the GreenArrow visualization.

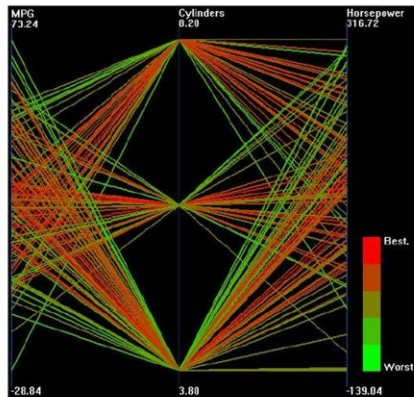
Be honest—can you read this?

“Indeed, the power of software to implement increasingly complex, colorful, and rapidly accessible knowledge visualizations may encourage researchers and information designers to develop representational schemes that will overwhelm, or at least unduly tax, the cognitive powers of the very users they are supposed to be serving.”

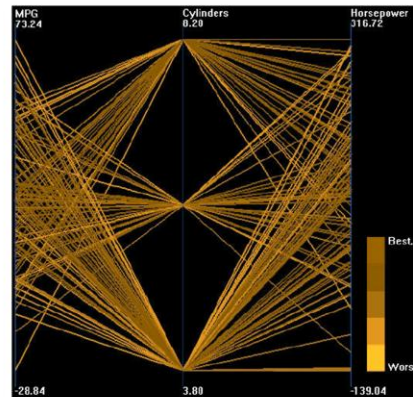
Sigmar-Olap Tergan and Tanja Keller, Editors, *Knowledge and Information Visualization*, “Representational Correspondence as a Basic Principle of Diagram Design”, Christopher F. Chabris and Stephen M. Kosslyn, Springer-Verlag, Berlin Heidelberg, 2005, p. 37.

Source of the image above: “Have Green – A Visual Analytics Framework for Large Semantic Graphs”, Wong, et. al., IEEE Symposium on Visual Analytics Science and Technology 2006, October 31-November 2.

Effective infovis uses color effectively.



Avoid red and green together



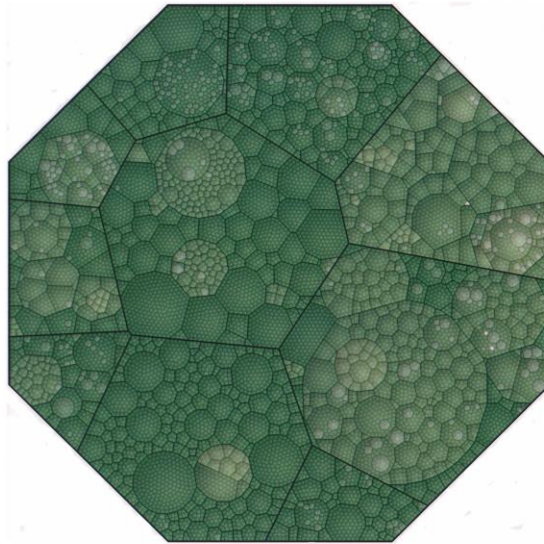
Here's how it looks if you're colorblind

I cringe every time I see a presentation at an InfoVis conference that exhibits an ignorance of one of the simplest facts about visual perception: approximately 10% of males and 1% of females are color blind, and most of them cannot distinguish red and green.

Source of the image on the left: "Exploratory Visualization of Multivariate Data with Variable Quality", Xie, et. al., IEEE Symposium on Visual Analytics Science and Technology 2006, October 31-November 2.

The image on the right was created by using Visccheck color deficit simulation software.

Effective infovis addresses the real-world needs of people.



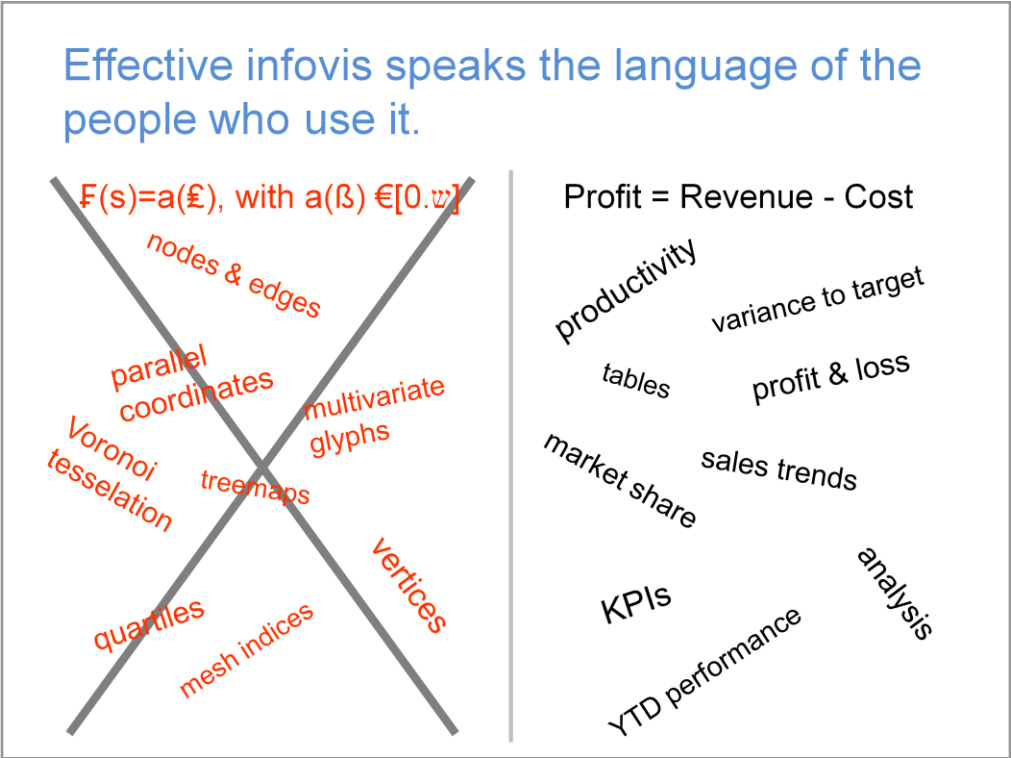
Cool!!!!

But what
can I do
with it?

“Professor Fred Brooks long ago encouraged researchers to focus on a ‘driving problem’. His advice remains potent, especially for those who are entranced with colorful animated displays and elaborate statistical manipulations. Explorers of the vast multidimensional spaces are more likely to make important discoveries if they keep their mind’s eye focused on solving their driving problem. They are also more likely to experience those wonderful Aha! moments of insight that are the thrill of discovery.”

Ben Shneiderman in the Foreword to Chaomei Chen’s book *Information Visualization: Beyond the Horizon, Second Edition*, 2004, Springer-Verlag, London, page ix.

Source of the image above: Michael Balzer and Oliver Deussen, “Voronoi Treemaps”, IEEE Symposium on Information Visualization 2005.



Most of the people (though certainly not all) who attend a conference like InfoVis understand the terms that appear above on the left, but most of the people who actually rely on InfoVis software to do their jobs don't. It isn't up to them to learn our language.

Grice's communication maxims

4 categories:

- Quantity
- Quality
- Relevance
- Manner

Paul Grice was a 20th century philosopher whose work ventured into the realm of linguistics. He is well known for his communication maxims, which attempt to describe the characteristics of courteous and effective communication.

Every one of these maxims apply to all forms of communication, including the words that you write in a research paper and the presentations that people make using InfoVis software.

Grice's communication maxims: Quantity

1. Make your contribution to the conversation as informative as necessary.
2. Do not make your contribution to the conversation more informative than necessary.

Grice's communication maxims: Quality

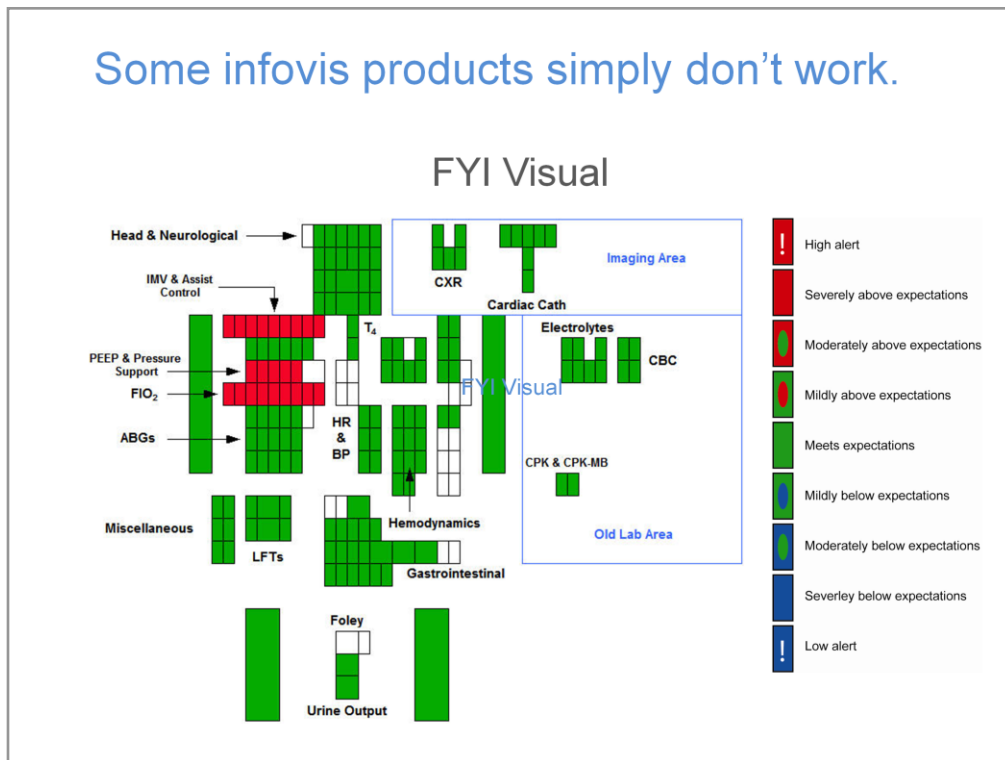
1. Do not say what you believe to be false.
2. Do not say that for which you lack adequate evidence.

Grice's communication maxims: Relevance

Be relevant (that is, say things related to the current topic of conversation).

Grice's communication maxims: Manner

1. Avoid obscurity of expression.
2. Avoid ambiguity.
3. Be brief (avoid unnecessary wordiness).
4. Be orderly.



I recently wrote an article to tell the story of FYI Visual, a product that was developed by Dr. Michael Lesser to address a real and important need, but simply doesn't work, because the person who did the initial research and developed the product did not have the expertise nor research method that was required to do so effectively. You can find the article at <http://www.perceptualedge.com/articles/06-26-07.pdf>.

When used for patient diagnosis (its original intent), FYI Visual presents a full set of medical data on a single screen or page. Each measure of the patient's health is represented by a single KEGS (rectangle, which serves as a "knowledge enhanced graphical symbol"). The KEGS' are arranged into related collections of measures, called a Kegset. An attempt is made to arrange this Kegset to roughly emulate the shape of the human body, with measures regarding the head positioned at the top, and so on. The position of each KEGS determines what it measures. Despite the good intentions of this structure, to make sense of the display, you must memorize the meaning of each KEGS, based on its location. I assume that most of the eight hours of training that is required for a physician to learn how to use an FYI Visual display involves memorizing, first, the meaning of each of the symbols (rectangles that contain various colors and patterns), and second, the meaning of each location.

Based on the key on the right, I assume that an actual display used by a physician would exhibit KEGS (rectangles) with a greater variety of colors. Some would appear as a solid color, others would include an oval of a different color within, and extreme values would include an exclamation point. Looking at this partial example below, you can get a sense of what's involved.

This is far too complex for rapid decoding.

| Type: Multistate Symbols | |
|--|-------|
| Unique name (alias) | set 1 |
| Unique name (alias) | set 2 |
| Unique name (alias) | set 3 |
| Unique name (alias) | set 4 |
| In addition to these symbols, custom symbols can be assigned to multistate KEGS. | |

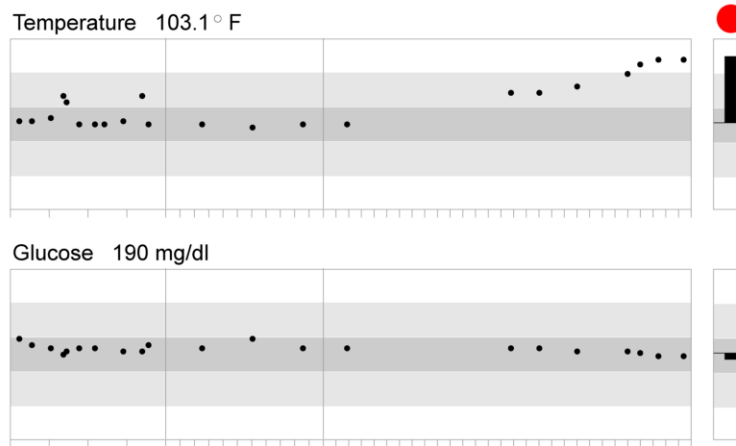
| Type: Binary Symbols | |
|--|-------------|
| State | Symbol Sets |
| Unique name (alias) | set 1 |
| Unique name (alias) | set 2 |
| Unique name (alias) | set 3 |
| Unique name (alias) | set 4 |
| In addition to these symbols, custom symbols can be assigned to binary KEGS. | |

| Type: KEGS Modifiers | Symbol |
|--------------------------------------|--------|
| Out of Paradigm | |
| Missing Data or No Confidence | |
| No Data Expected | |
| Agged Data or Less Confidence (for) | |
| Missing Symbol | |
| Flyout Indicator (for) | |

Here is a full library of KEGS, with a separate color and pattern for each possible meaning in the vocabulary. A great deal of research has been done over many years to determine how visual perception works and the visual attributes that can be used most effectively to encode meanings for clear and efficient input and understanding. Do the colors and shapes that FYI Visual uses to encode values make use of this research? No, they do not. It is not clear why Lesser chose these particular colors and patterns, contained in rectangles and why he considered an oval of one color inside a rectangle of another color better than the alternatives, such as variations of color intensity from light to dark. I suspect that he used his own imagination and judgment to make these choices, rather than taking advantage of the available body of empirical research. This is unfortunate.

Is it sufficient to reduce quantitative values that measure health, to a discrete set of values that express low, normal, and high, thereby losing the precision of the original values? Can a physician make an informed diagnosis based only on a set of recent measures, without historical information? There is not enough information or context to make informed judgments. The intention to consolidate patient information into a single display and express it in low-normal-high terms is a good one, but FYI's solution provides only a fraction of the effectiveness that better designs can provide, which are built on a solid understanding of visual perception and cognition.

Here's an alternate solution.



A promising way to display patient information, which uses variations of conventional graphs (dot plots), was proposed by Seth M. Powsner and Edward R. Tufte in the mid 1990s ("Graphical Display of Patient Status," *The Lancet*, August 6, 1994; also described in Tufte's 1997 book *Visual Explanations*). I've put together a partial example of a similar display above, which could be developed using something as simple as Excel .

To complement the time-series dot plots that Powsner and Tufte proposed, I've added a bullet graph on the right (my own invention) for each item to prominently feature the most recent measure. Rather than trying to force the different measures of health into an anatomical arrangement as Lesser has done, they could be more simply arranged in logical groups of related measures, with headings to identify the groups. Groups of related measured could be quickly compared using the bullet graphs, with the historical information readily available whenever it is needed to put what's happening into context. Red icons, or something similar that appears only next to measures that are critical, whether high or low, could be used to highlight them for rapid scanning.

By displaying the information on a computer screen (perhaps one that is handheld), physicians could quickly select the measures that they wish to focus on, causing all others to temporarily disappear, leaving a much simpler set of data. Similarly, they could rapidly grab and toss out any one or more measures that aren't needed. Measures could be arranged in any order to support comparisons. Details about each historical measure could be accessed by hovering with the mouse pointer over any single data point, and by zooming into the graph, allowing it to temporarily fill the screen and add more information.

This is but one of perhaps several designs for displaying patient information that would provide the richer set of data that are needed to perform diagnoses, and would do so in a way that would require almost no training, because the information has been mapped in a more natural and effective manner to visual representations.

Some vendors begin with a bad idea and then try to use it for everything.



Despite the ineffectiveness of FYI Visual, the company has had some success selling to the U.S. military and has been trying desperately for years to conquer the business intelligence market.

Above is a copy of a submission that FYI made to a data visualization competition, which I judged for *DM Review* magazine in 2005. The purpose was to design a sales dashboard, based on data that I provided. Of all the entries, this was the least effective. Frankly, it was absurd. It was based on this sort of evidence that I formed the opinion that FYI was trying to sell itself as a general BI solution when it in fact had no capability, nor understanding of data visualization, to merit its use for BI.

What should infovis help people do?

Think
and
Communicate

“Above all else show the data.”

Edward Tufte

It is helpful to remind ourselves from time to time what InfoVis is meant to help people do: to think about data to make sense of it and then to communicate what was found to others.

This well known Edward R. Tufte quote above is from *The Visual Display of Quantitative Information* (1983).

Infovis defined

Information visualization is the use of computer-supported interactive visual representations of abstract data to amplify cognition.

Card, Mackinlay, & Shneiderman (1999)

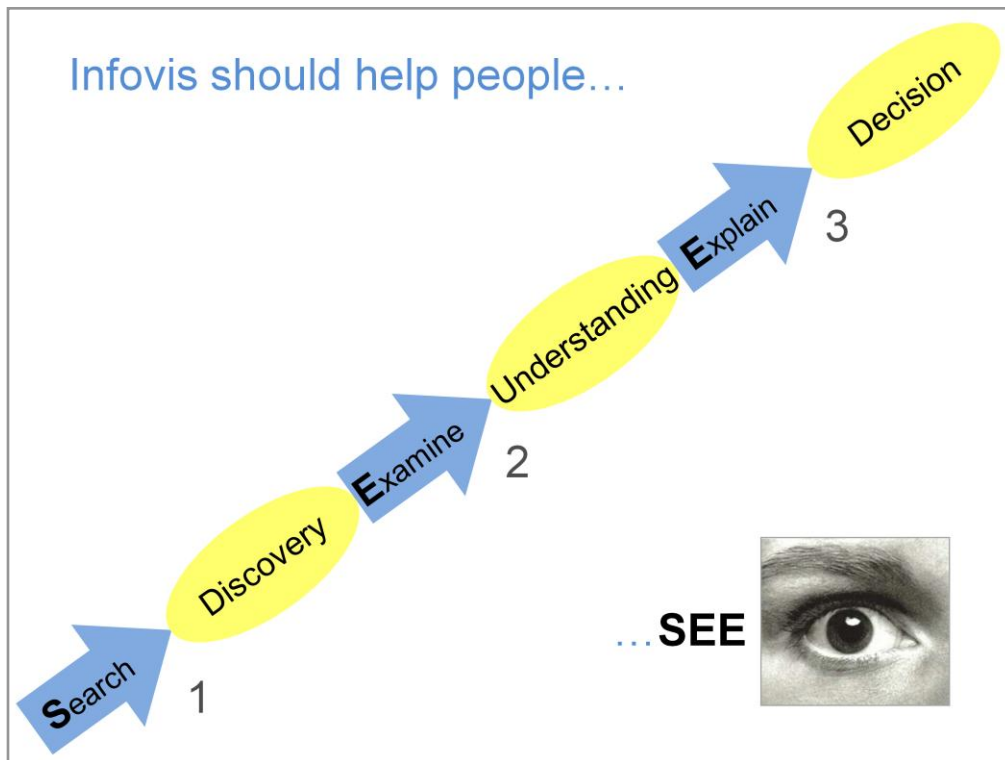
The three authors of *Readings in Information Visualization: Using Vision to Think* (1999), Stuart Card, Jock Mackinlay, and Ben Shneiderman, provided an excellent working definition of information visualization for us.

Their definition features the following:

- *computer-supported* – The visualization is displayed by a computer, usually on a computer screen.
- *interactive* – The visualization can be manipulated directly and simply by the user in a free-flowing manner, including such actions as filtering the data and drilling down into details.
- *visual representations* – The information is displayed in visual form using attributes like location, length, shape, color, and size to make a picture of the data and thereby reveal patterns, trends, and exceptions that might not be seen otherwise.
- *abstract data* – Information such as quantitative data, processes, and relationships, as opposed to visual representations of physical objects, such as geography or the human body.
- *amplify cognition* – Interacting with these visualizations extend our ability to think by assisting memory and representing the information in ways that our brains can easily comprehend.

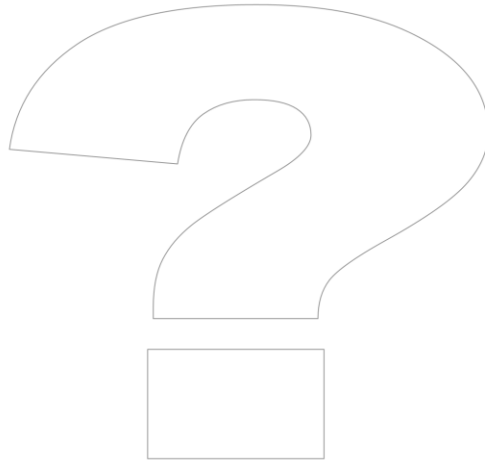
Information visualization can be used for a variety of purposes, especially:

- Searching for specific information in a large collection of data
- Exploring data in an effort to discover anything that is meaningful and useful, whether anticipated or not
- Sense-making (i.e., examining it in an effort to understand it, to recognize particular characteristics of the data and understand what they mean)



All data analysis begins with (1) searching through the data to discover potentially meaningful facts, then involves (2) examining that data more closely to understand it, including what caused it to occur, so that you can then (3) explain what you've learned to those who can use that knowledge to make good decisions. Most of what we need to recognize and understand in our business data is not all that complicated.

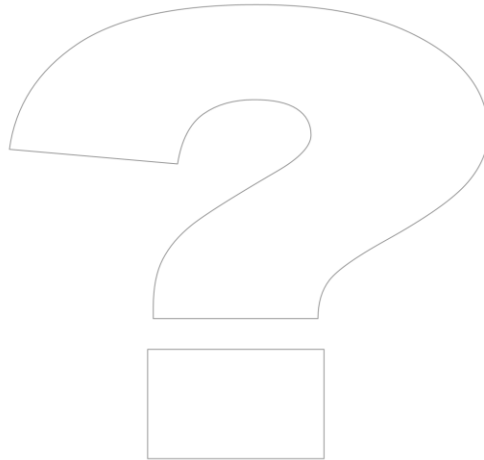
How can we determine what the world needs from infovis?



Here are a few suggestions:

- Spend time working in the real world doing things that can benefit from Infovis.
- Assist people in the real world in their efforts to make sense of information. Observe what causes them to struggle.
- Become familiar with the business intelligence industry.
- Examine existing products to determine what they lack.

What is most needed from infovis today?



Here are a few of the needs that I'm aware of:

- Users must learn basic skills of data analysis and communication
- Products that do the basics simply and effectively
- Products that make it easy to shift seamlessly between different visualizations and different slices of data
- Products that support the transition from data analysis to presentation seamlessly
- Software that makes it easy to tell statistical stories
- Environments that support analytical collaboration
- Better human-computer interfaces
- Custom solutions to a host of specific real-world needs
- Reduction of problems related to over plotting

I read an email a few days ago in which a business intelligence consultancy listed the questions it was hoping the infovis community and this conference would address. Here's the list:

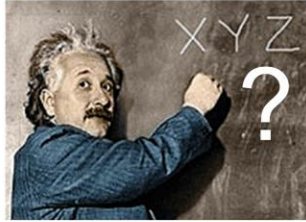
- What's the most distracting, but commonly seen visualization method? In other words, what methods are counter productive?
- Data is growing exponentially. What is the industry doing to identify the correct data to visualize? Can this be automated, or will it always take an expert/artist/scientist to figure it out?
- What initiatives/advances are being made to enable visualization for the masses? Applications? Techniques? Mediums?
- If a picture says a 1000 words, what does an animation tell? Is it better? Are there principles that work for static images, but that do not for motion?
- How can I do visualization if I'm not a programmer?
- What are some of the best ways to visualize flows, progressions or timelines?
- How can this stuff become more mainstream and not just academic?
- Why is this conference so dominated by academics? Does that give us a clue as to why good visualization isn't pervasive in business tools as opposed to scientific tools?
- Shouldn't "design" and "information visualization" be married? Are they even dating yet?

What does it take to make good decisions?

Good
Data

Understanding

Good
Decisions



Software, no matter how sophisticated, is useless if the people who use it don't possess the fundamental skills of data analysis. Data analysis is for one purpose: to enable good decisions. Does one need to be an Einstein to make sense of data?

Most data analysis is not complex.

Only 10% requires

- Sophisticated techniques
- Complex statistics

90% requires only

- Simple visual techniques
- Minor statistics

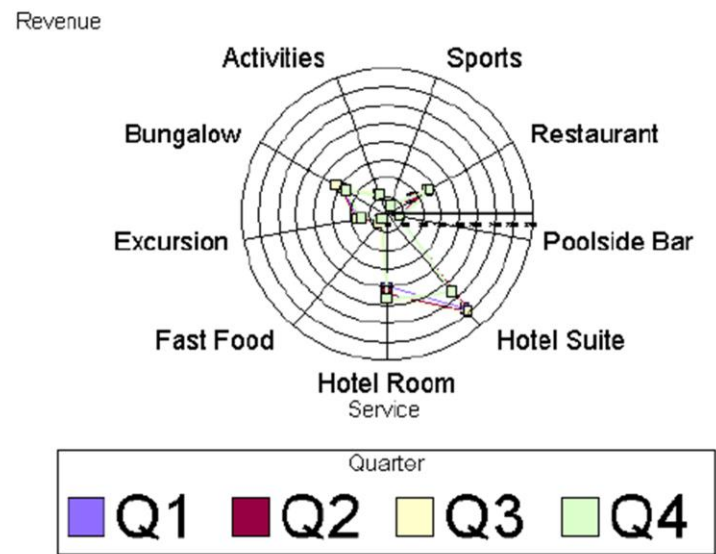
Though simple, people must learn these skills and software must support them .

Most of the data analysis that is needed in the normal course of business requires relatively simple data visualization techniques, leaving little that requires the sophisticated techniques of statistical and financial analysis. If you search for resources that teach data analysis skills, you'll find many books and courses that present the sophisticated techniques needed by the few, but few resources if any that teach the simple techniques that most of us need to make sense of business data. The skills that most of us need to infuse our businesses with needed insights can be learned without a background in statistics, but these skills don't come naturally – they must be learned. You must develop expertise, but it is expertise that can be easily learned with the proper direction and practice. You must learn to see particular patterns in data that are meaningful.

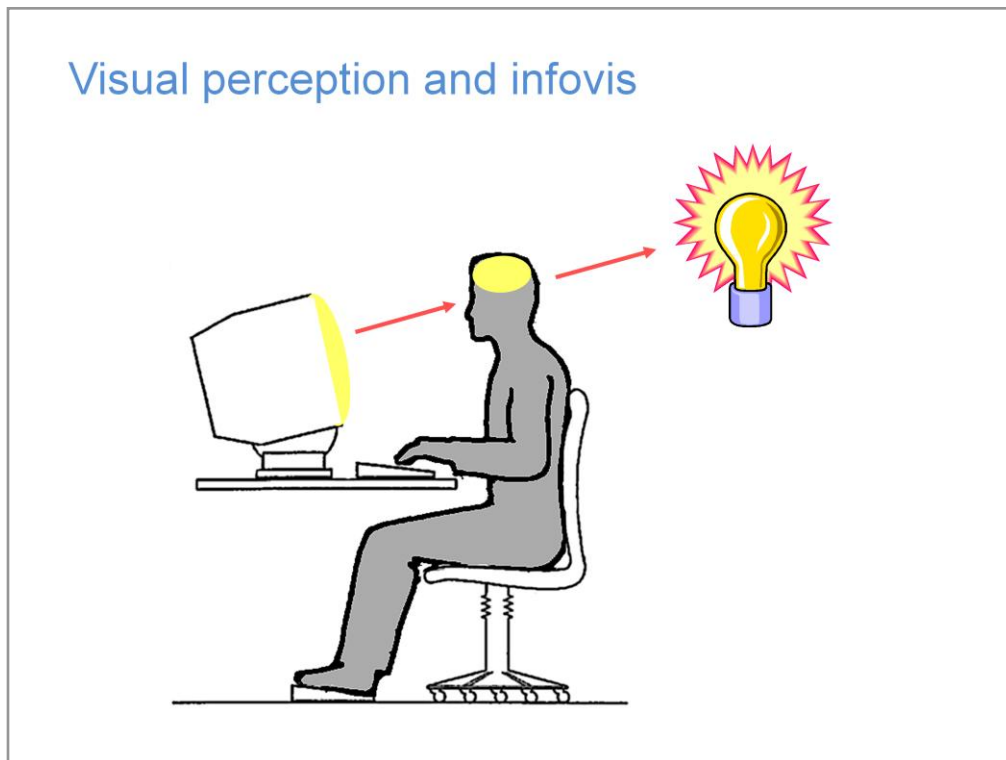
People can learn pattern-detection skills, although the ease of gaining these skills will depend on the specific nature of the patterns involved. Experts do indeed have special expertise. The radiologist interpreting an X-ray, the meteorologist interpreting radar, and the statistician interpreting a scatter plot will each bring a differently tuned visual system to bear on his or her particular problem. People who work with visualizations must learn the skill of seeing patterns in data.

Information Visualization: Perception for Design, Second Edition, Colin Ware, Morgan Kaufmann Publishers, 2004, page 209.

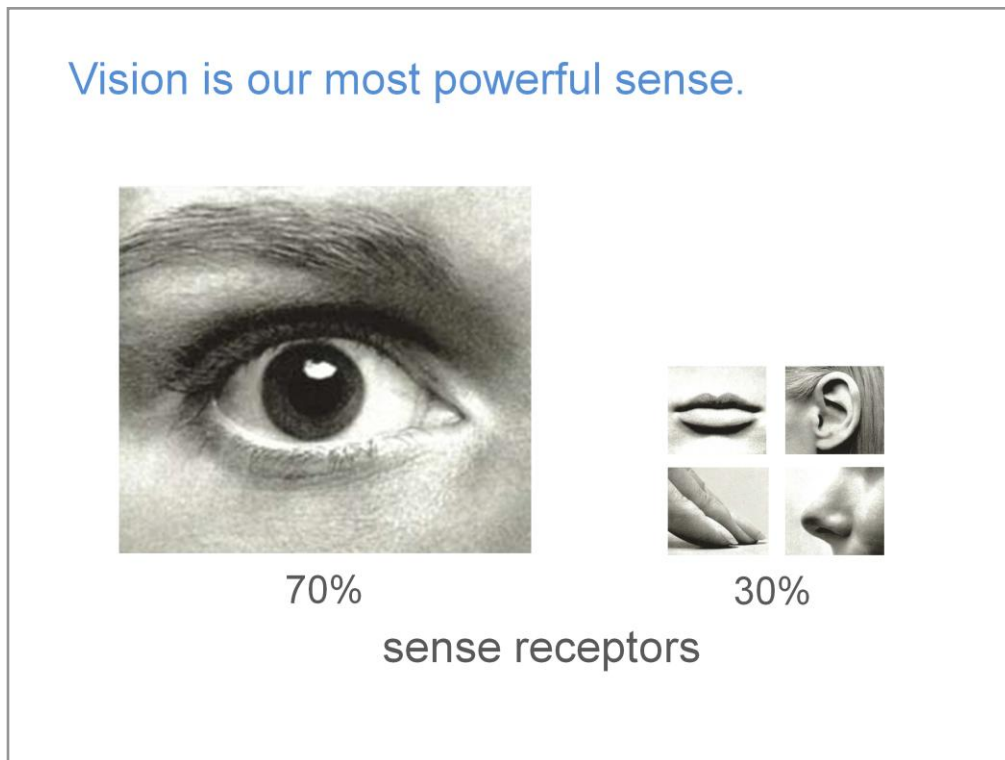
Much of what the world needs is the simple stuff done better.



While we're trying to create the next generation of information visualization, most of the commercial software available today handles even the most basic visualizations without even an elementary understanding of what works. There is a great need in the software industry for the expertise that properly trained infovis professionals have to offer.



Even perfect data can become difficult to understand if it is not presented to our eyes in a clear manner. Much of the software available for analyzing data does a poor job of presenting data. This is especially true of typical business software that presents data visually in the form of graphs. It is the responsibility of InfoVis software to present information to our eyes in ways that match how we see and how we think.



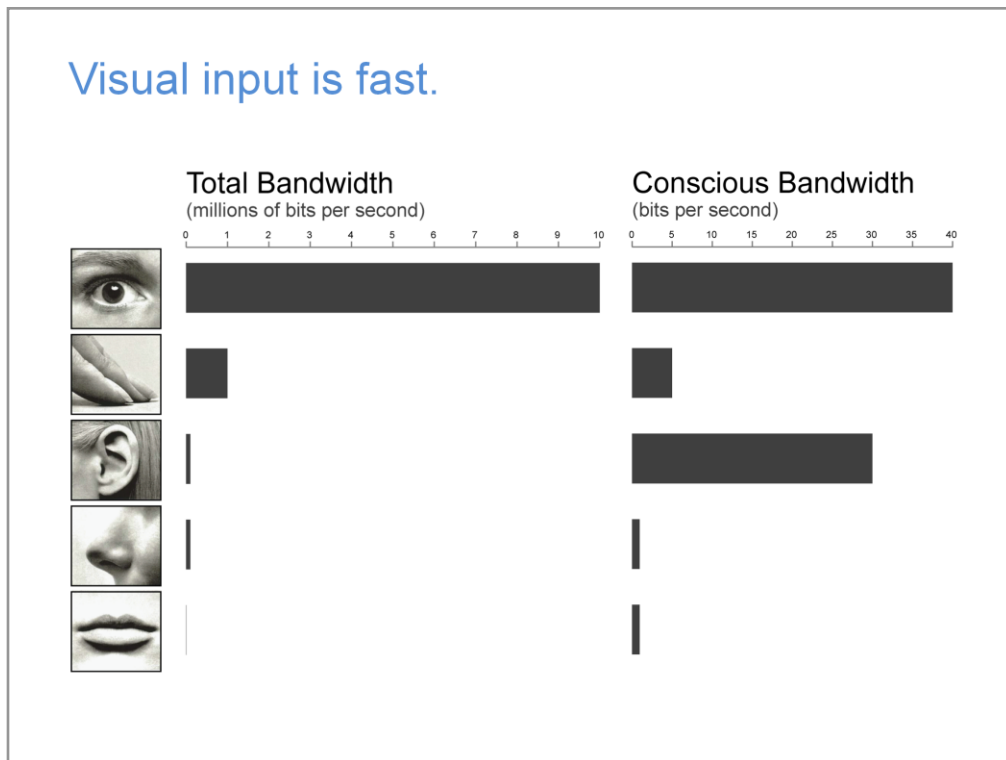
Human perception is amazing. I cherish all five of the senses that connect us to the world, that allow us to experience beauty and an inexhaustible and diverse wealth of sensation. But of all the senses, one stands out dramatically as our primary and most powerful channel of input from the world around us, and that is vision. Approximately 70% of the body's sense receptors reside in the eye.

Perhaps the world's top expert in visual perception and how its power can be harnessed for the effective display of information is Colin Ware, who has convincingly described the importance of data visualization. He asks:

Why should we be interested in visualization? Because the human visual system is a pattern seeker of enormous power and subtlety. The eye and the visual cortex of the brain form a massively parallel processor that provides the highest-bandwidth channel into human cognitive centers. At higher levels of processing, perception and cognition are closely interrelated, which is the reason why the words 'understanding' and 'seeing' are synonymous. However, the visual system has its own rules. We can easily see patterns presented in certain ways, but if they are presented in other ways, they become invisible... The more general point is that when data is presented in certain ways, the patterns can be readily perceived. If we can understand how perception works, our knowledge can be translated into rules for displaying information. Following perception-based rules, we can present our data in such a way that the important and informative patterns stand out. If we disobey the rules, our data will be incomprehensible or misleading.

Information Visualization: Perception for Design, Second Edition, Colin Ware, Morgan Kaufmann Publishers, 2004, page xxi.

Perhaps the best known expert in data visualization, Edward Tufte, says: "Clear and precise seeing becomes as one with clear and precise thinking." (*Visual Explanations*, Edward R. Tufte, Graphics Press: Cheshire, CT. 1997 page 53.)



Stating that 70% of the bodies sense receptors reside in the eye perhaps tells an overly simplified story of what is going on. A slightly different picture unfolds when we compare the five senses in terms of total bandwidth (millions of bits of information per second), but vision still clearly dominates with 10 million bits per second versus 1 million for touch (via receptors in the skin), all the way down to only 1,000 bits per second of taste perception. To complete the picture of how each channel of sensory perception contributes to our overall awareness, however, we should consider only that part of perception of which we are conscious. From this perspective, vision still dominates with 40 bits per second of information, but hearing isn't far behind with 30, followed by touch with 5, and smell and taste with 1 each. Despite the reduced dominance of vision in the realm of consciousness, the fact that it dominates so dramatically in total sense receptors and bandwidths of perception means that it offers a vastly richer spectrum of what it can communicate tell us.

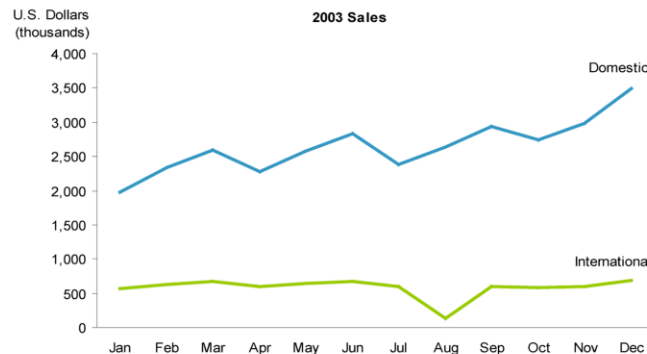
(Source: Tor Norretranders, *The User Illusion: Cutting Consciousness Down to Size*, Viking Press, New York, 1998)

Visualizations give shape to the data.

2003 Sales (U.S. dollars in thousands)

| | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|---------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 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 |
| International | 574 | 636 | 673 | 593 | 644 | 679 | 593 | 139 | 599 | 583 | 602 | 690 |
| | \$2,557 | \$2,979 | \$3,266 | \$2,876 | \$3,218 | \$3,517 | \$2,975 | \$2,773 | \$3,537 | \$3,322 | \$3,585 | \$4,183 |

versus



The presentation of data as text, such as you see in this table, is perfect when you need precise values or when the purpose is to look up or compare individual values, but not when you wish to see patterns, trends, and exceptions, or to make comparisons. When this is your goal, visualizations work best.

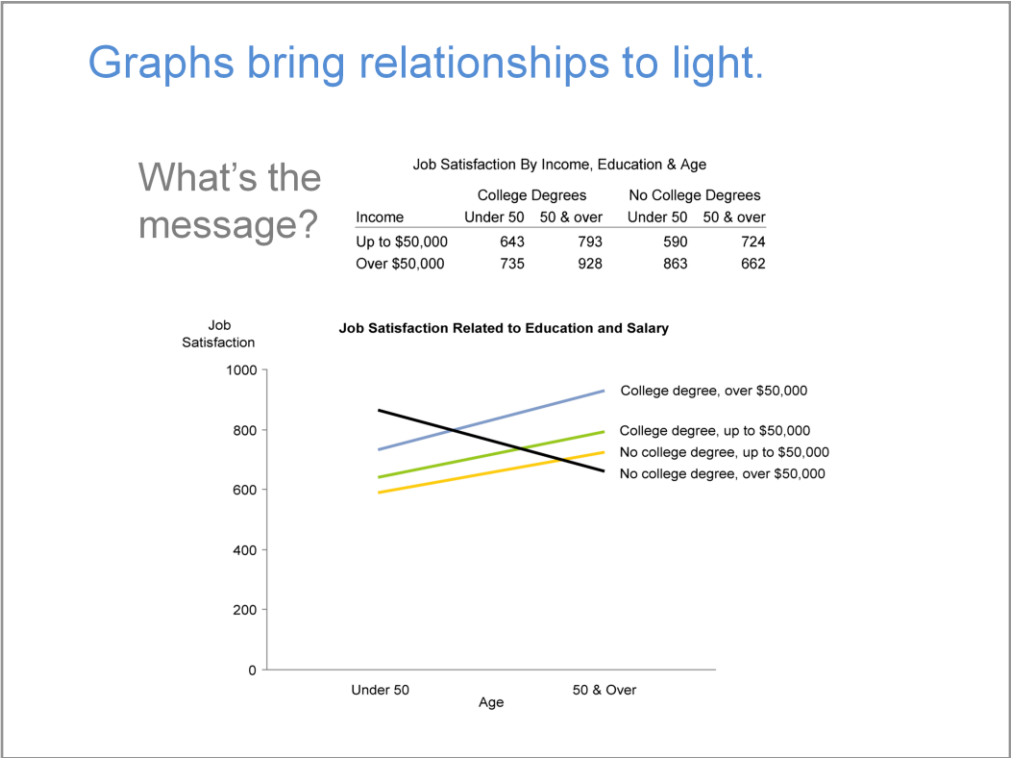
When data is presented visually, it is given shape, and from its shape we can easily glean insights that would take a long time to piece together from the same data presented textually, if ever. This graph of the same data that appears in the table makes brings to light several of the stories contained in the data that weren't obvious before, and it did so instantly.

When] we visualize the data effectively and suddenly, there is what Joseph Berkson called 'interocular traumatic impact': a conclusion that hits us between the eyes.

Visualizing Data, William S. Cleveland, Hobart Press, 1993, page 12.

Modern data graphics can do much more than simply substitute for small statistical tables. At their best, graphics are instruments for reasoning about quantitative information. Often the most effective way to describe, explore, and summarize a set of numbers – even a very large set – is to look at pictures of those numbers. Furthermore, of all methods for analyzing and communicating statistical information, well-designed data graphics are usually the simplest and at the same time the most powerful.

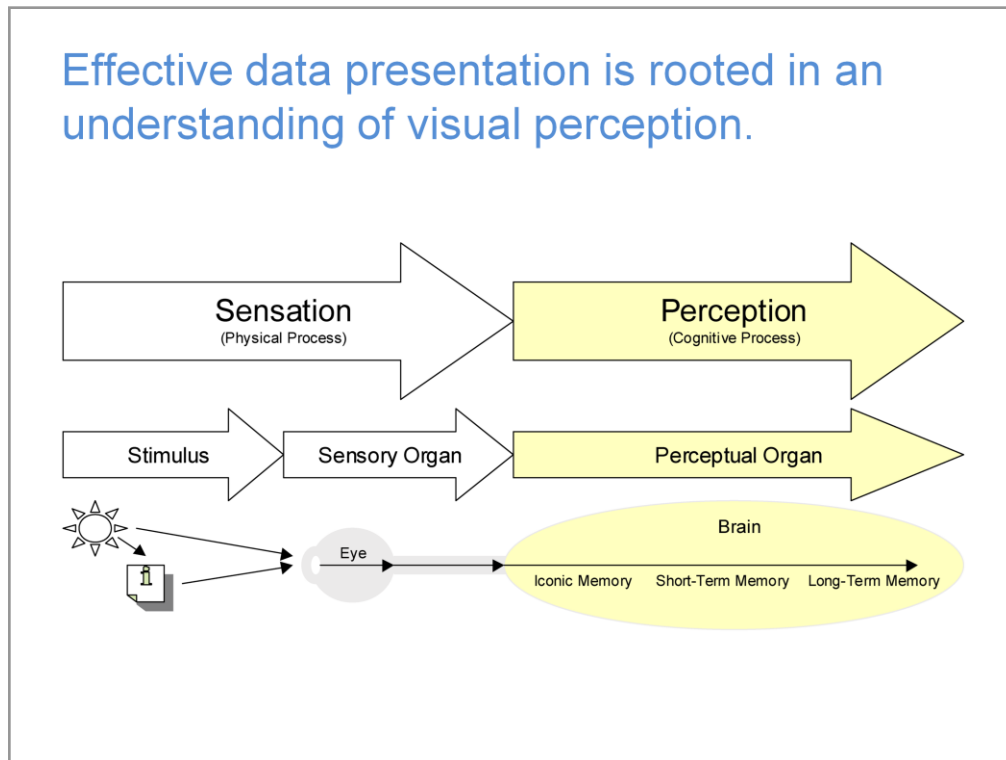
The Visual Display of Quantitative Information, Edward R. Tufte, Graphics Press: Cheshire, CT 1983, Introduction.



The fact that job satisfaction for employees without a college degree decreases significantly in their later years doesn't jump out at you when you examine the table, but it is immediately obvious when you examine the graph.



The type of graph that is selected and the way it's designed also have great impact on the message that is communicated. By simply switching from a line graph to a bar graph, the decrease in job satisfaction among those without college degrees in their later years is no longer as obvious.

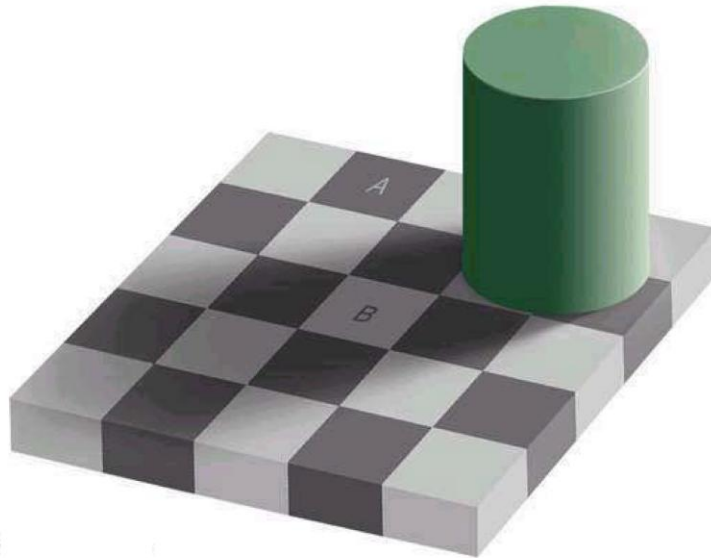


To know how to present information visually in an effective way, you must understand a little about visual perception – what works, what doesn't, and why. We won't delve into this deeply, but it is worth learning a bit about two aspects of visual perception that apply directly and powerfully to visual data analysis: the *pre-attentive attributes* of visual perception and the *limits of working memory*.

Useful References:

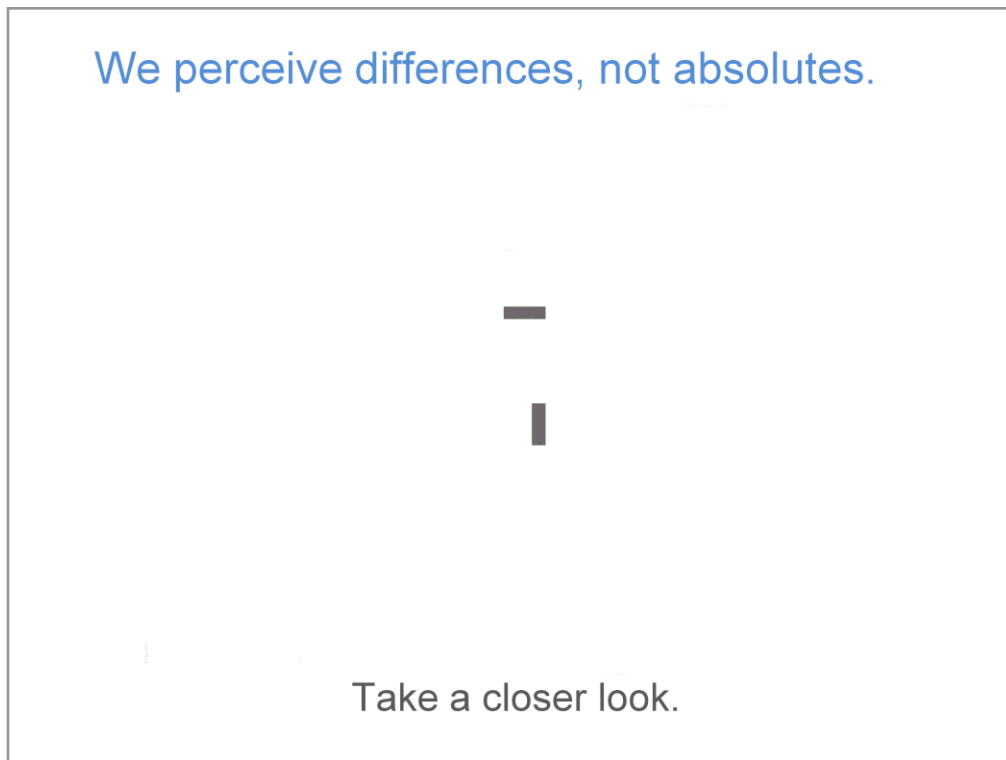
Information Visualization: Perception for Design, Second Edition, Colin Ware, Morgan Kaufmann Publishers, 2004.

Visual perception is not just camera work.



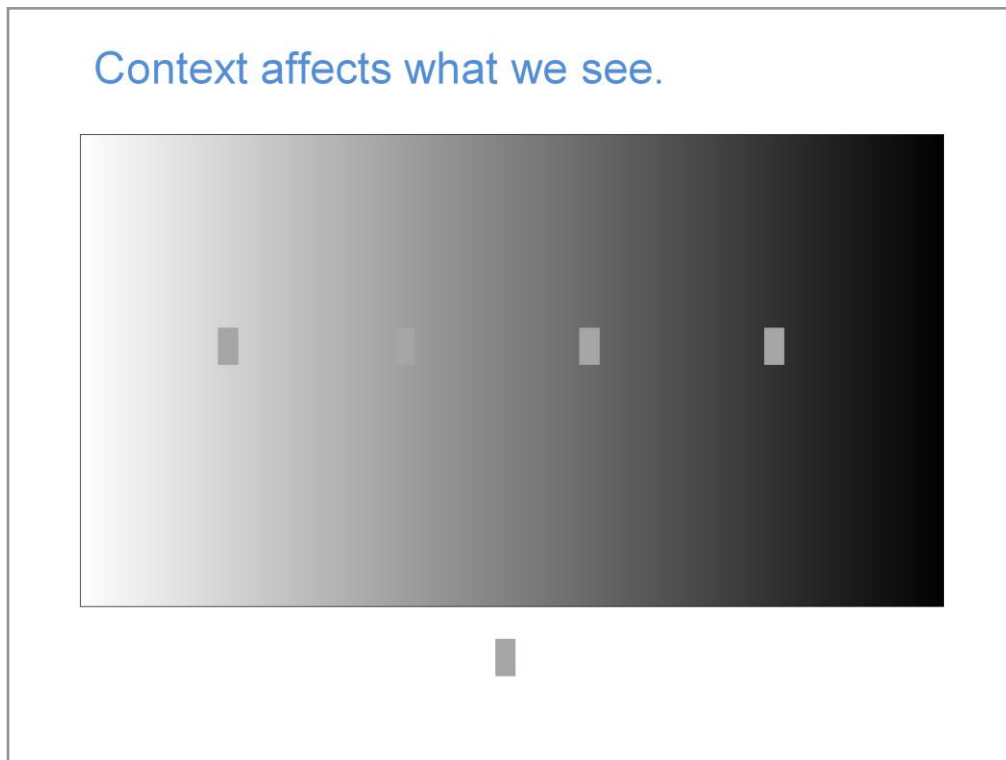
Square A is darker than B, right?

Actually, squares A and B are exactly the same color.



What we see is not a simple recording of what is actually out there. Seeing is an active process that involves interpretations by our brains of data that is sensed by our eyes in an effort to make sense of it in context. The presence of the cylinder and its shadow in the image of the checkerboard triggers an adjustment in our minds to perceive the square labeled B as lighter than it actually is. The illusion is also created by the fact that the sensors in our eyes do not register actual color but rather the difference in color between something and what's nearby. The contrast between square A and the light squares that surround it and square B and the dark squares that surround it cause us to perceive squares A and B quite differently, even though they are actually the same color, as you can clearly see above after all of the surrounding context has been removed.

The ability to use graphs effectively requires a basic understanding of how we unconsciously interpret what we see.



This image illustrates the surprising effect that a simple change in the lightness of the background alone has on our perception of color. The large rectangle displays a simple color gradient of a gray-scale from fully light to fully dark. The small rectangle is the same exact color everywhere it appears, but it doesn't look that way because our brains perceive visual differences rather than absolute values, in this case between the color of the small rectangle and the color that immediately surrounds it.

Among other things, understanding this should tell us that using a color gradient as the background of a graph should be avoided.

Perception is sometimes serial and slow
and sometimes parallel and immediate.

987349702756479021947286240924060370804702890727
803208029007305901270238008374082078720272008083
247802602703793715709701379706674620970941027806
927979709123097230919592750927309272197873497260

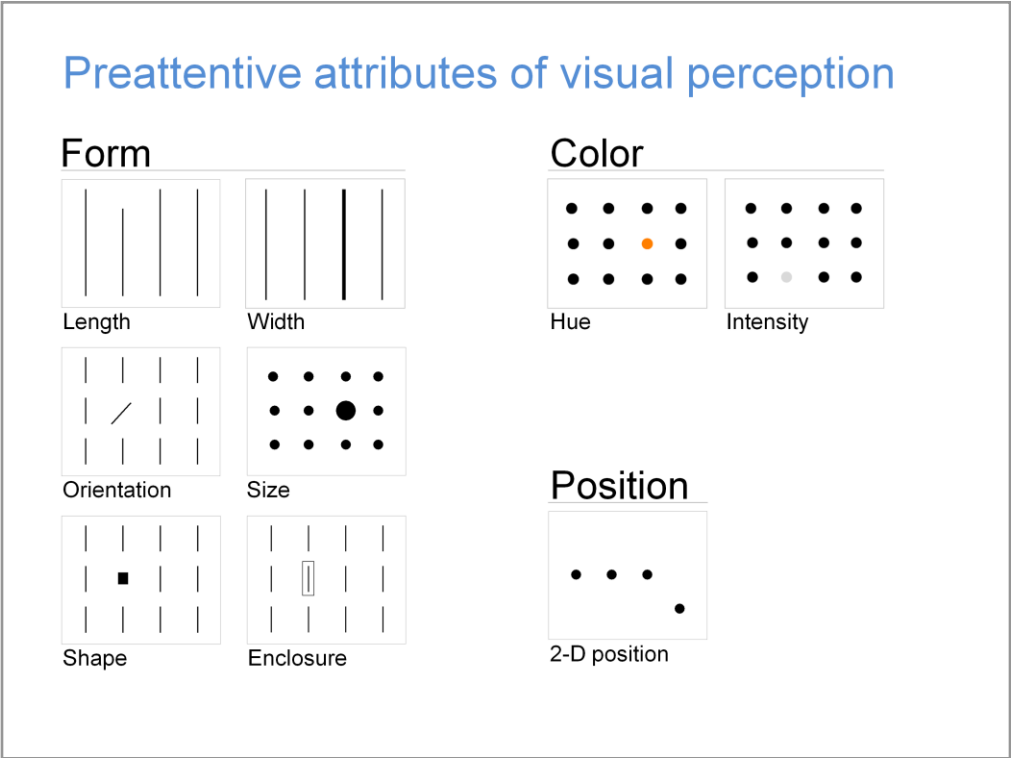
9873497027**5**6479021947286240924060370804702890727
80320802900730**5**901270238008374082078720272008083
24780260270379371**5**709701379706674620970941027806
927979709123097230919**5**927**5**0927309272197873497260

How many fives are in this list of numbers?

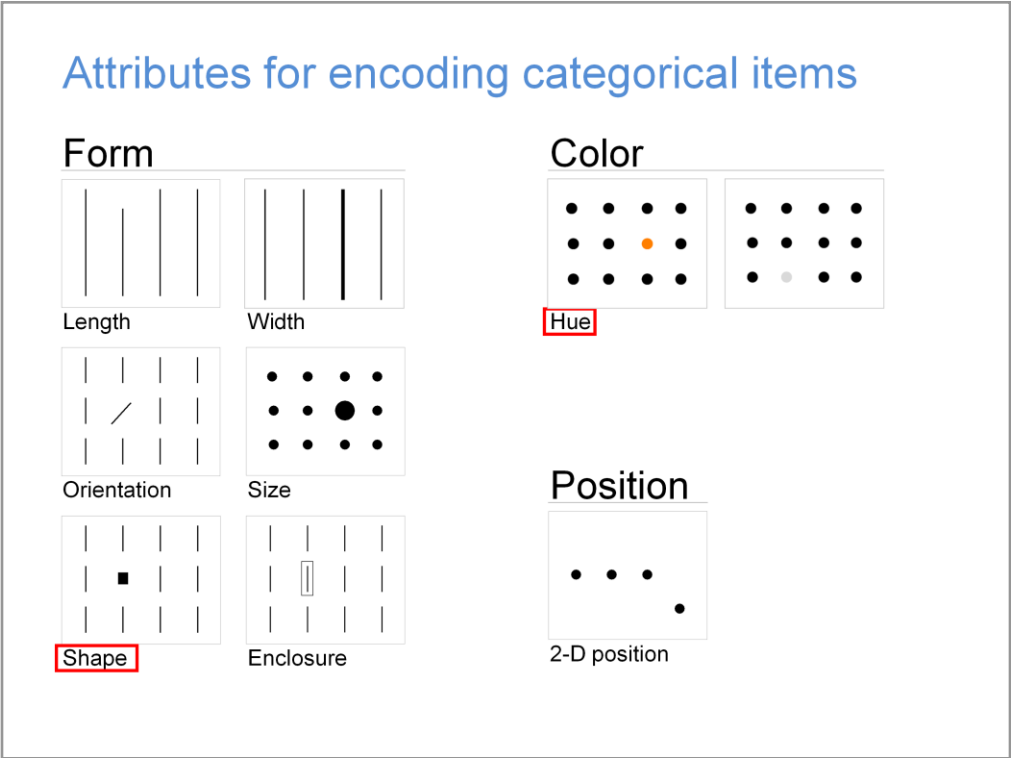
Text, the written form of verbal language, must be processed serially. Because the top list above consists of digits without spaces to group them into separate multi-digit numbers, you must read them one digit at a time.

In the bottom list, however, the fives pop out immediately. The bottom list is exactly the same as the top, except for one simple visual difference: they are a darker color. This single distinction made them immediately perceptible.

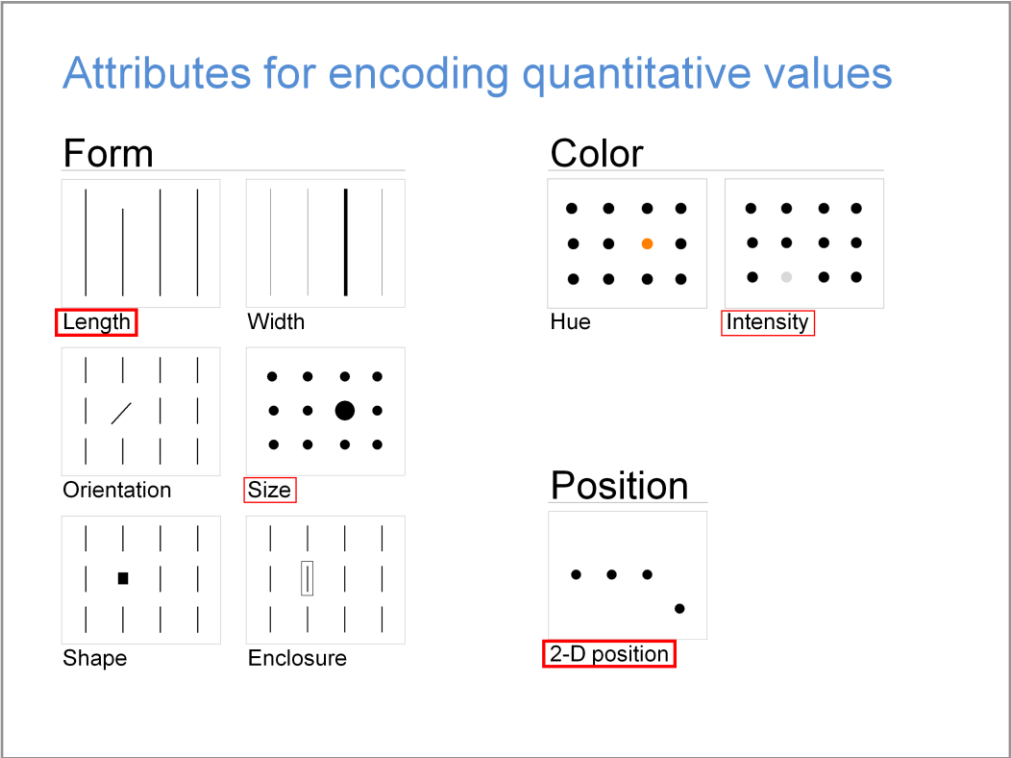
Some visual attributes are easier to see and distinguish than others. The most powerful of these are called *pre-attentive attributes* because we perceive them immediately, without conscious thought.



We perceive some visual attributes immediately, without conscious thought. We call these *preattentive attributes* of visual perception. The full list of visual attributes that we perceive pre-attentively is larger than the list above. These are the ones that are most useful to us when examining data visually in a static display.

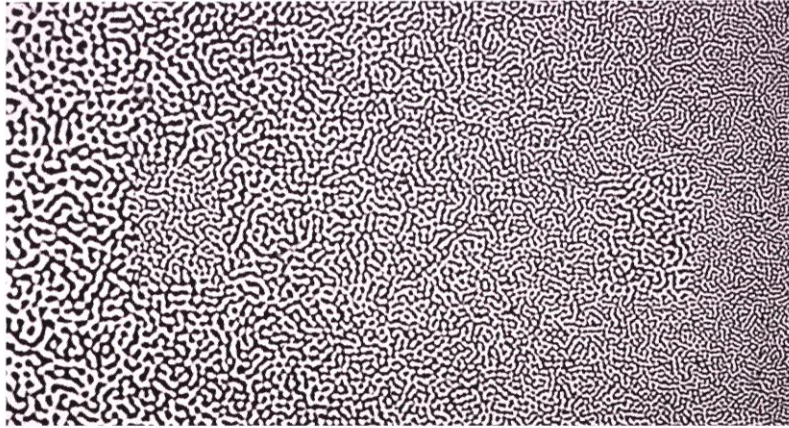


The preattentive attribute that provides the best means to visually distinguish data sets in a graph is hue. Simple shapes like squares, circles, triangles, diamonds, X's, and +'s (plus signs) can also be used, but they don't work as well as distinct colors. Distinct hues work best of all as long as you're not color blind.



Some of these preattentive visual attributes are perceived quantitatively (i.e., some values are greater than others), which are marked with red boxes. The visual attributes that are marked with pale red boxes are perceived quantitatively but not as powerfully as line length and 2-D position.

We don't notice everything that we see.



We notice contrasts to the norm.

We do not notice everything in our field of vision. Visual perception is selective and must be, for an awareness of everything out there would overwhelm us. We tend to notice two types of visual phenomena:

1. Contrasts to the norm
2. Distinct patterns that we are searching for specifically.

For this reason, to successfully see meaning in the data, we must visually encode interesting exceptions to pop out, and we must know the patterns that we wish to see and search for them one at a time.

In the image above, the two sections of texture that stand out, one left of center and one right of center are exactly the same. They differ only from the surround field of texture because the lines that form the texture pattern are smaller than those that surround them on the left and larger than those that surround them on the right.

(Note: The texture image appears in Colin Ware, *Information Visualization: Perception for Design*, Second Edition, Morgan Kaufmann Publishers: San Francisco, CA, 2004.)

We don't notice everything that we see.



We notice patterns that we know to look for.

There is an image in the picture of the rose that is only noticeable if we know to look for it. Once primed with the image of the dolphin, however, we can easily spot it in the rose.

(Note: The image of the rose was found at www.coolbubble.com.)

We don't remember everything that we see.



We only clearly remember that to which we attend.

In addition to understanding visual perception, visual analysis tools must also be rooted in an understanding of how people think. Only then can they recognize and support the cognitive operations that are necessary to make sense of information.

Memory plays an important role in human cognition. Because memory suffers from certain limitations, visual analysis tools must be able to augment memory.

The example above illustrates one of the limitations of short-term memory. We only remember that to which we attend. Any part of this image that never gets our attention will not be missed when we shift to another version of the image that lacks that particular part. If we don't attend to it, we might notice the change from one version of the image to the next, but only if the transition shift immediately from one to another, without even a split second of blank space between them.

In addition to not remembering, we also don't clearly see that on which we don't focus. To see something, we must focus on it, for only a small area of receptors on the retinas of our eyes are designed for high-resolution vision.

(Source: This demonstration of change blindness was prepared by Ronald A. Rensink of the University of British Columbia. Several other examples of this visual phenomenon can be found at <http://www.psych.ubc.ca/%7erensink/flicker/download/index.html>.)

Effective visualizations therefore must...

- Make meaningful information stand out in contrast to what's not worth our attention
- Encode meaningful information as patterns that we can recognize and interpret
- Direct us to focus on what matters

Visualizations can augment working memory.

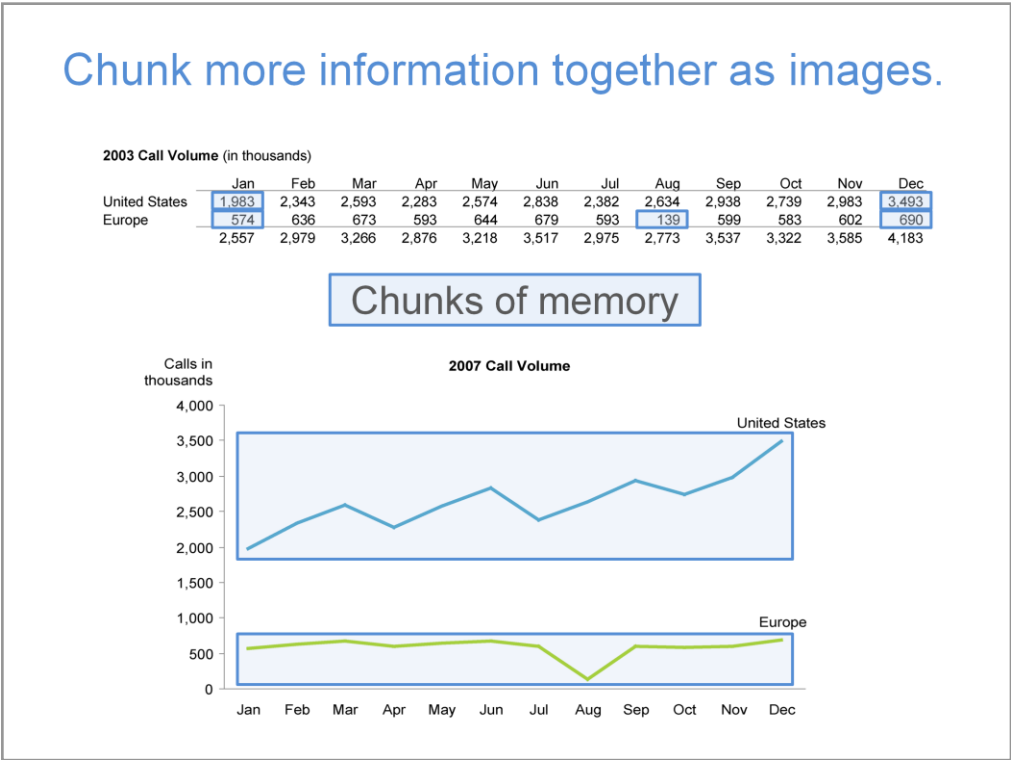
We only retain about four chunks of visual information at a time

therefore visualizations should

place all information that should be compared within eye span

and

encode groups of related values as visual patterns that can be stored as chunks in memory.



Memories are stored as chunks of information. A chunk can be as small as a single tiny fact (for example, revenue equals \$56,384 for the quarter) or a larger set of facts that you've learned to think about as a single complex unit (for example, a trend line on a time-series graph that shows revenue increasing from month to month throughout the year.) The better you get at seeing and understanding meaningful patterns and relationships in data, the better able you are to store more data as a single chunk. Working memory – also called short-term memory – is where information is stored while we are thinking about something. It is like the working memory, or RAM, in a computer. Our brains are constantly swapping chunks of information in and out of working memory from either what we perceive in the outside world or from the more permanent storage of long-term memory. There is a limit to the amount of information that can be held in working memory at any one time, which is estimated by researchers to be about four chunks.

Because working memory is limited in this way, when examining and analyzing data, it is important to place as much information within eye span as possible. If you see patterns in a graph and then try to compare them to patterns in another graph on a different screen, you won't remember everything that you were looking at previously. You'll end up bouncing back and forth between displays, wasting time and getting very frustrated in the process.

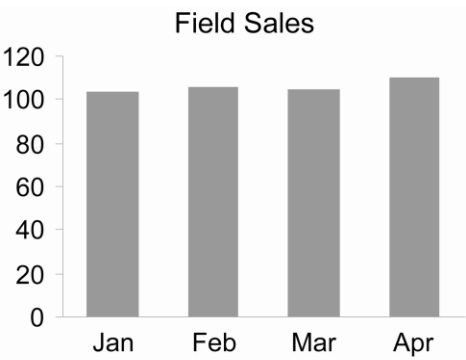
Avoid fragmented displays.

All data that you wish to compare ought to reside within eye span.

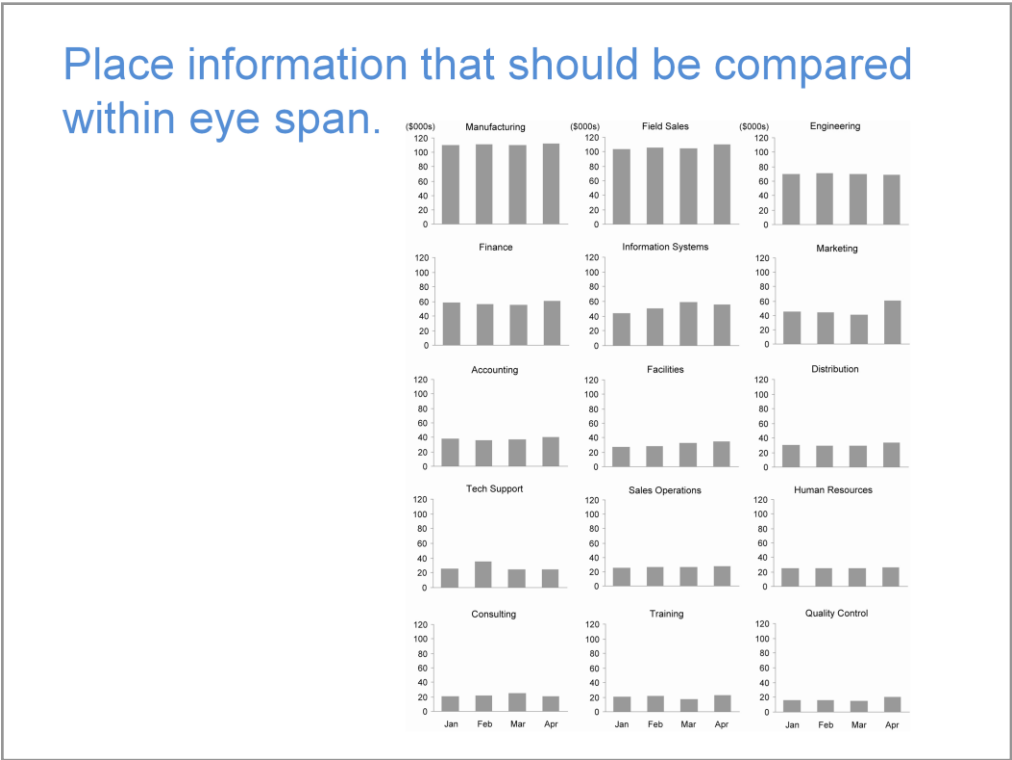
If you must scroll or switch from screen to screen to see it, good comparisons are impossible.

It is very difficult with most software to combine all of the information that you want to see together on a single screen without needing to scroll. You often end up bouncing from screen to screen to see separately what you would ideally like to see together in order to make comparisons and get a sense of the big picture.

Separate displays cannot be compared.



This slide displays six graphs in quick succession, but it is almost impossible to remember what appeared in any of the graphs after they are no longer visible. For instance, do you remember which department had the lowest expenses in the month of March?



Now, however, with expenses for 15 separate departments visible at the same time, this display serves as an external aid to short-term memory, making it easy to make comparisons.

Many of the best visualizations are also the simplest.

Elegance through simplicity

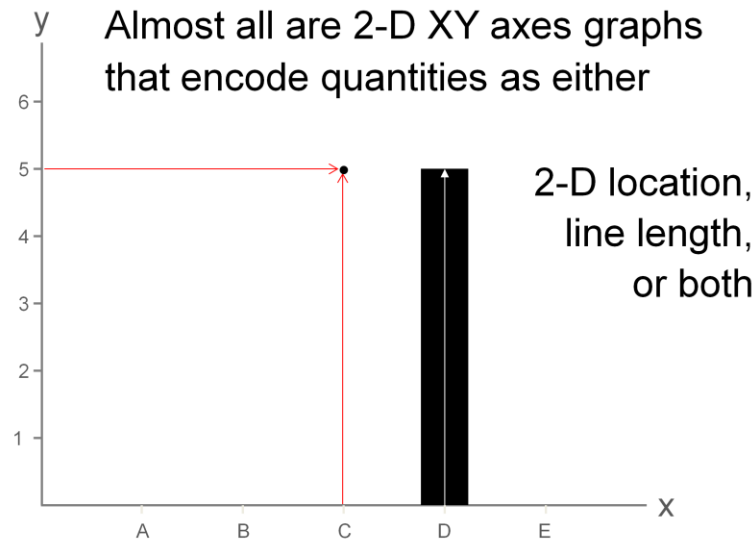
Simplify, simplify, simplify.

Henry David Thoreau

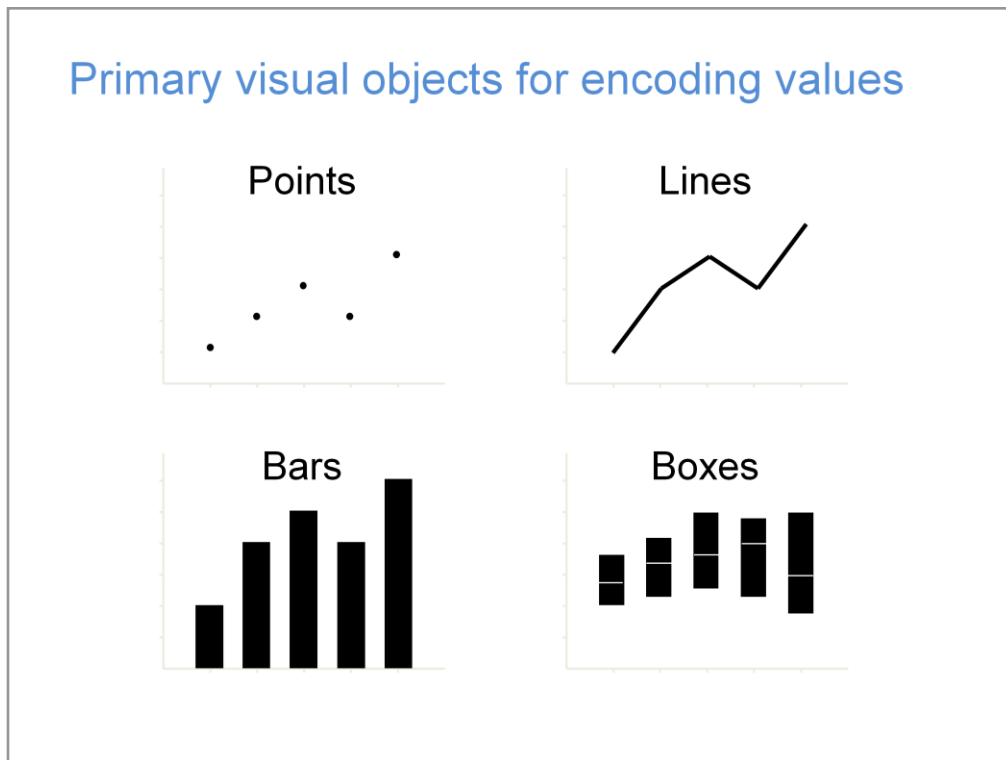
When is one expression better than another for analysis? Basically, when the data are more simply described, since this implies easier and more familiar manipulations during analysis and, even more to the point, easier and more thorough understanding of the results.

The Collected Works of John W. Tukey, John W. Tukey, Wadsworth, Inc.: Belmont, CA, 1988, page 12.

Effective graphs for quantitative analysis...

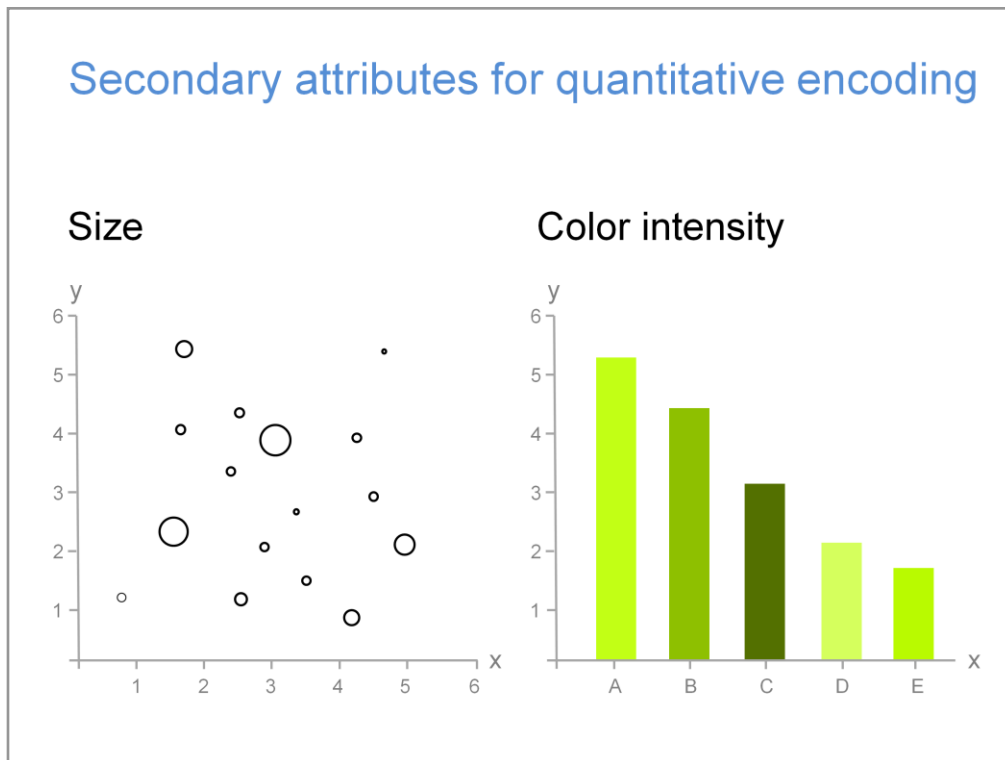


Despite the fact that software that is used for visual data analysis usually includes a broad assortment of graph types, only a few work well for the analysis of typical quantitative business data. All the useful graphs are of the 2-D XY type and use either the 2-D position of a data object (data points, data points along a line, and the endpoints of bars) or line length (length of bars) in relation to the quantitative axis to encode quantitative values.



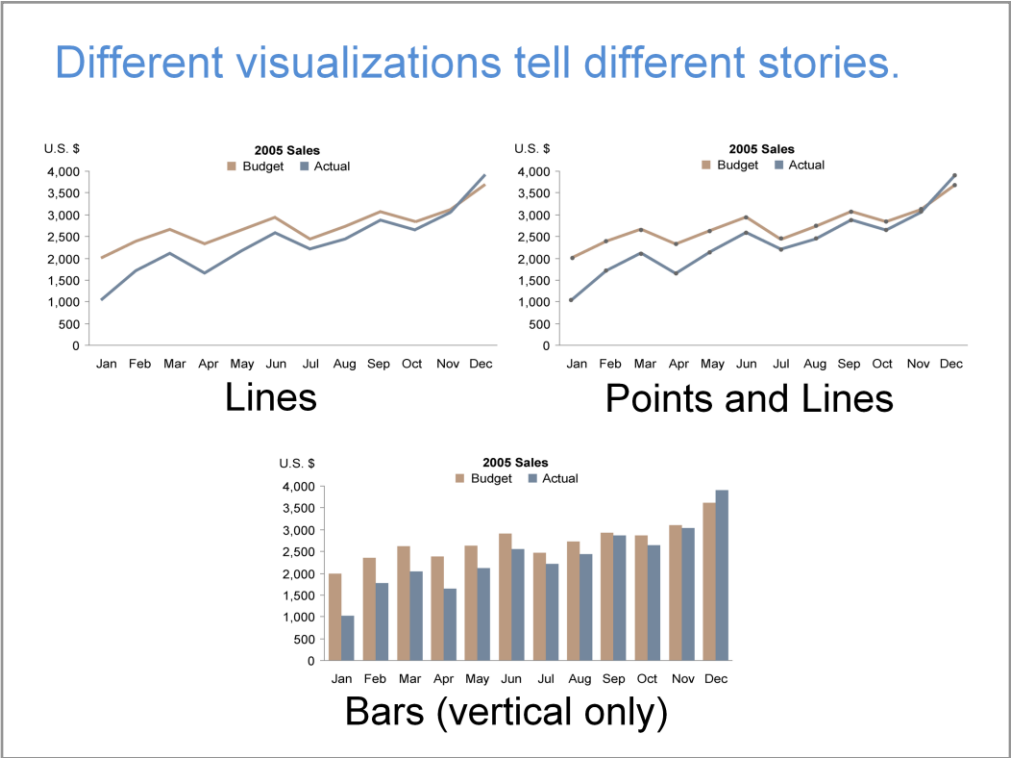
For typical quantitative data analysis, only four objects are needed to encode data in graphs:

- Points (small circles, squares, triangles, etc.)
- Lines (with or without data points on them to mark the values)
- Bars
- Boxes (similar to bars, but used to display distributions of values from the lowest to the highest and usually points of interest in between)



In the scatter plot on the left, the size of the data point is being used to encode a third quantitative variable, because 2-D position along the X and Y axes has already been used. As you can see, precise magnitude comparisons based on the sizes of these circles are not possible, but you can roughly perceive these differences in value.

In the bar graph on the right, two quantitative variables have been encoded in the bars: one as the heights of the bars in relation to the Y axis and a second as the color intensity of the bar ranging from light green for low values and dark green for high values. If you imagine that the heights of the bars encode revenues for various sales regions and that color intensity encodes profits, a graph such as this could be used to determine that the region with the greatest revenue falls somewhere near the center in terms of profits, whereas the third ranking region in revenues earns the highest profits. Precisely how much greater the profits are for region C compared to region A, however, you cannot determine, but if your objective doesn't require this precision, a graph of this type can be quite useful.

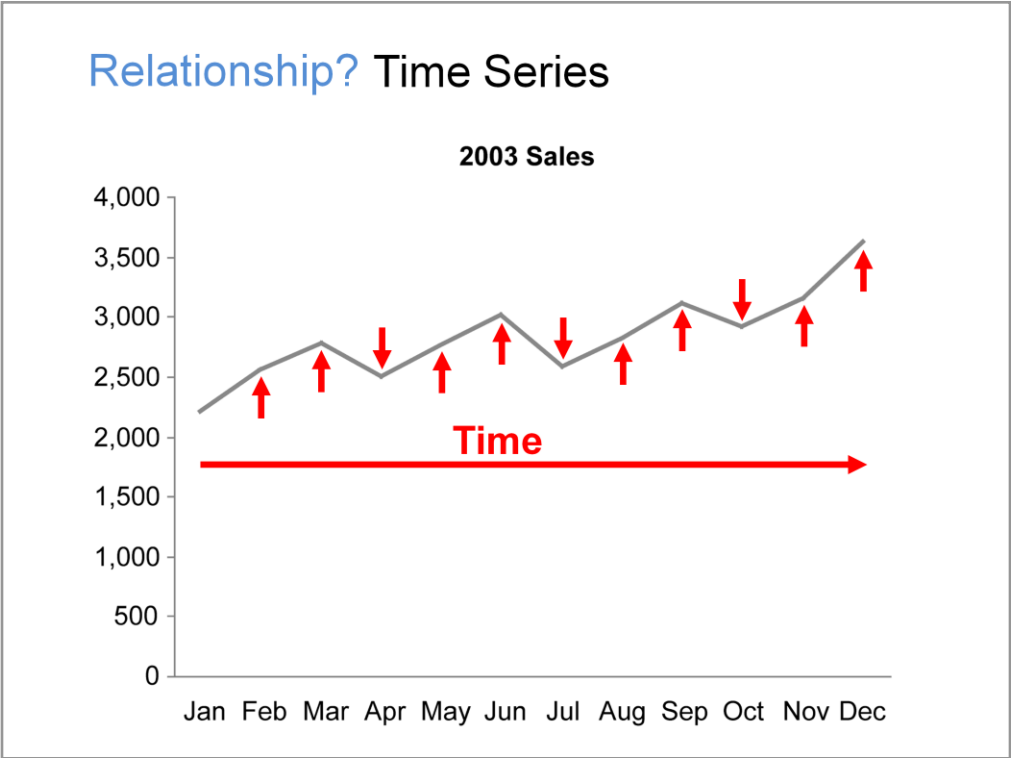


Lines do a great job of showing the flow of values across time, such as consecutive months of a year. The movement from one value to the next in this case represents change, giving meaning to the slope of the line: the steeper the slope, the more dramatic the change. If you want your message to emphasize individual values, such as the value for each month, however, bars do the job nicely. This is especially true when you graph multiple data sets, such as revenue and expenses, and you want to make it easy to compare these values for individual units of time, such as the month of September.

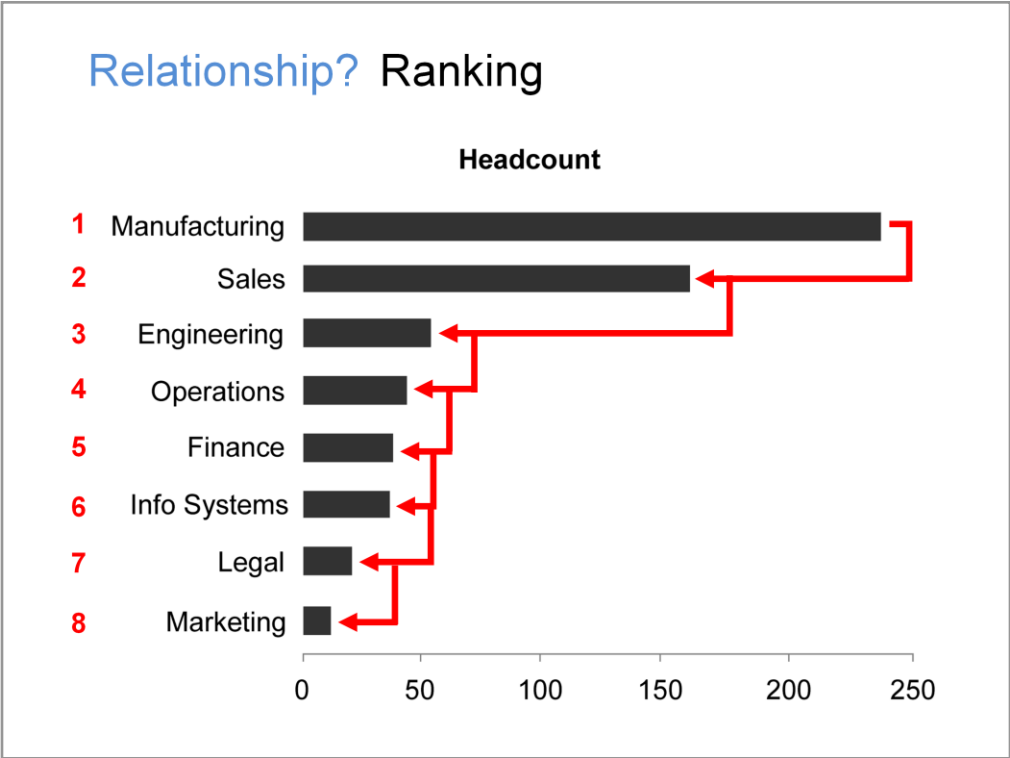
Quantitative visualizations feature only a few relationships.



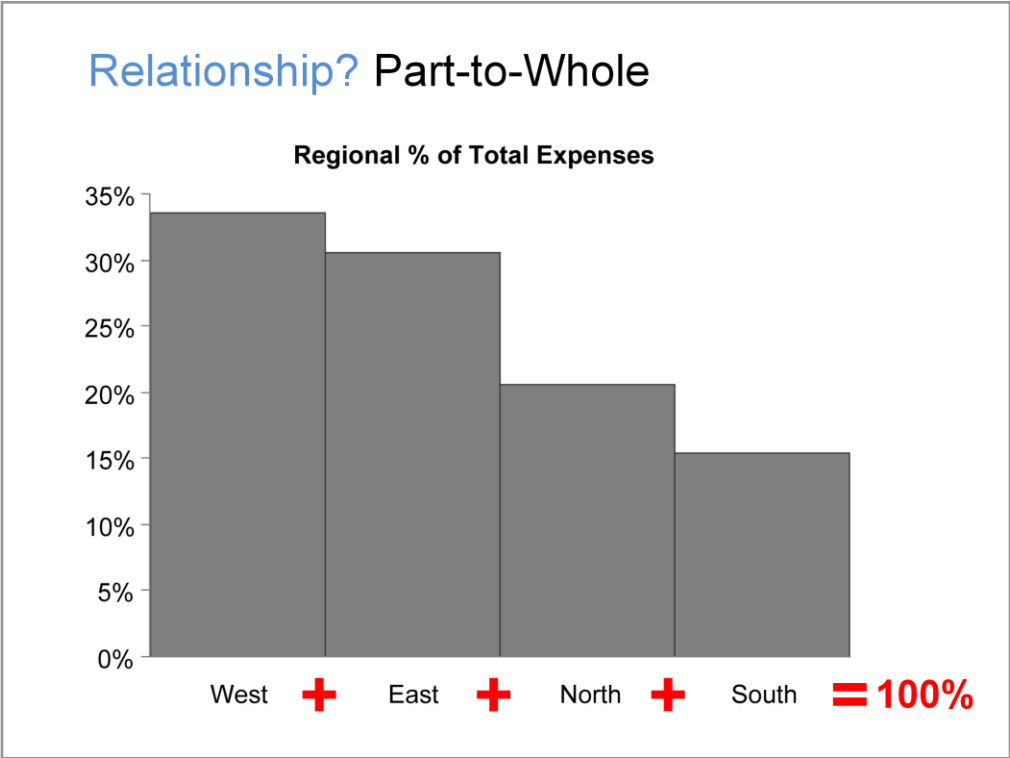
In my work with business people, I help them focus on the limited number of relationships that they typically need to feature when presenting quantitative data and the best ways to do so.



A time-series graph has a categorical scale that represents time, subdivided into a particular unit of time, such as years, quarters, months, days, or even hours. These graphs provide a powerful means to see patterns in the values as they march through time.

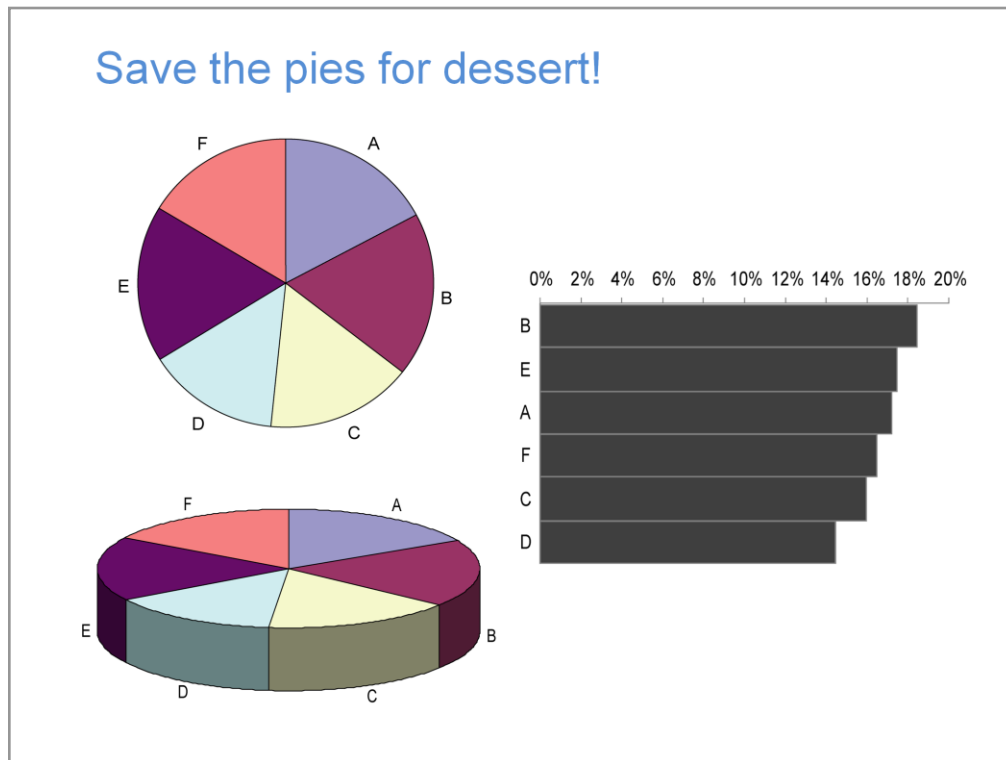


Ranking graphs show the sequence of a series of categorical subdivisions, based on the measures associated with them.



A part-to-whole graph shows how the measures associated with the individual categorical subdivisions of a full set relate to the whole and to one another.

Notice that I didn't use a pie chart to display this relationship. For comparing the magnitudes of the parts and seeing how they are ranked, bar graphs are superior.

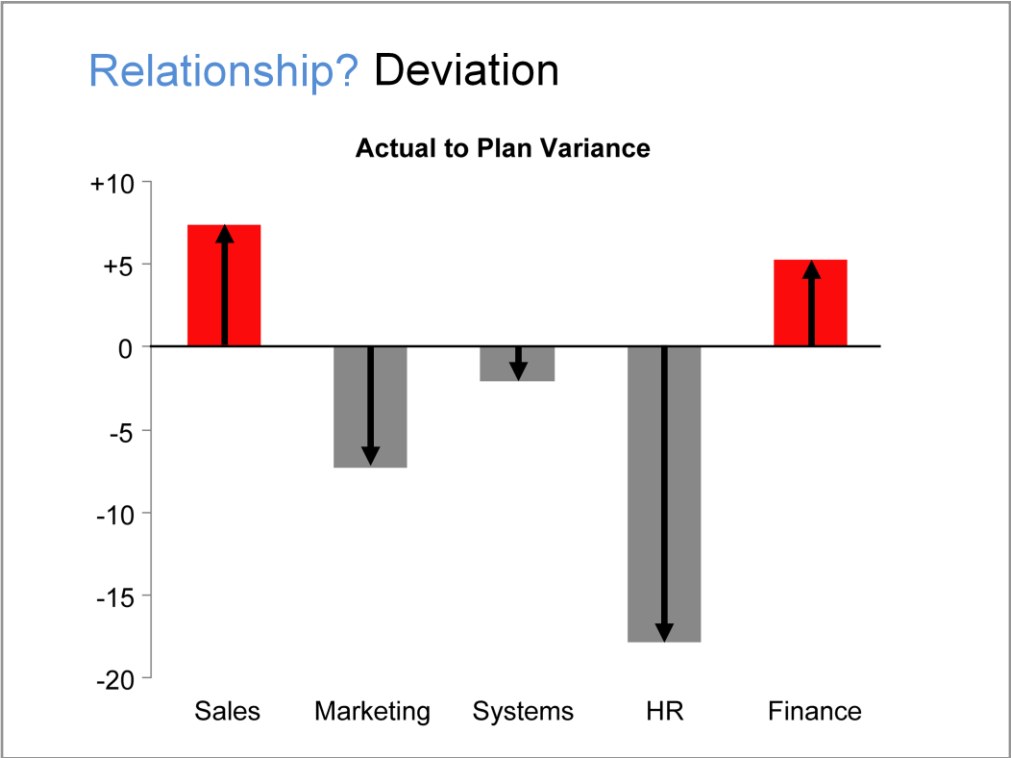


Pie charts use 2-D areas and the angles formed by the slices to encode quantitative values. Unfortunately, our perception of 2-D areas and angles as measures of quantity is poor.

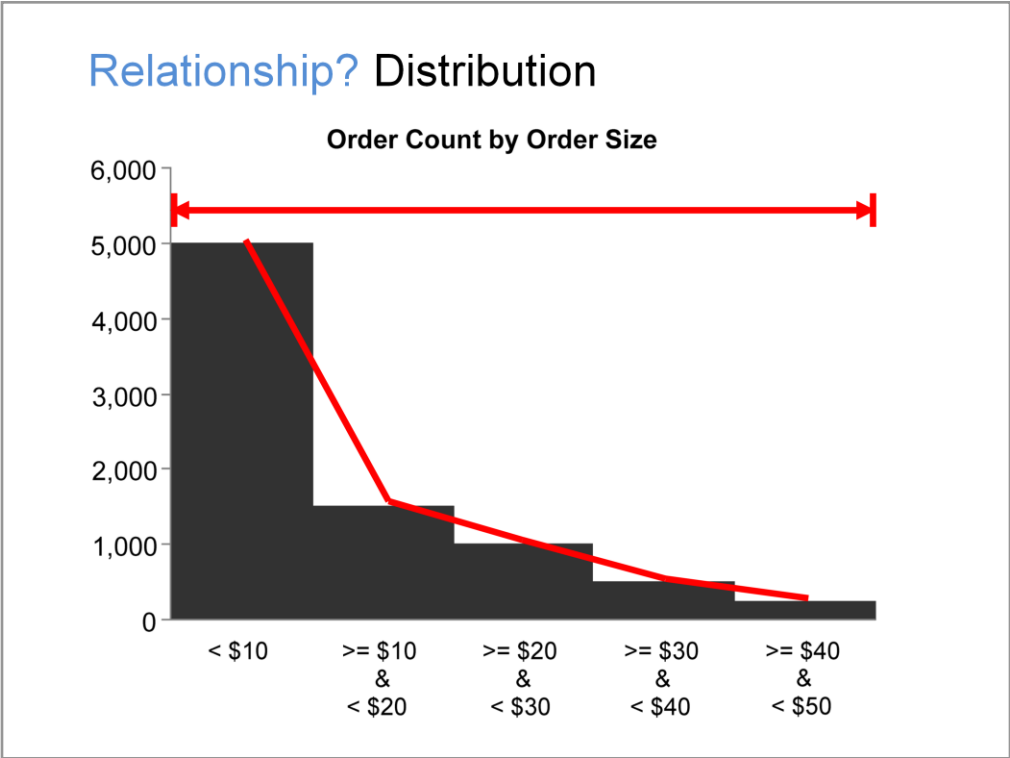
Try using either one of the pie graphs to put the slices in order by size. Can't do it, can you? Now see how easy this is to do when the same data is encoded in a bar graph.

Coda Hale once expressed his opinion of pie charts quite colorfully:

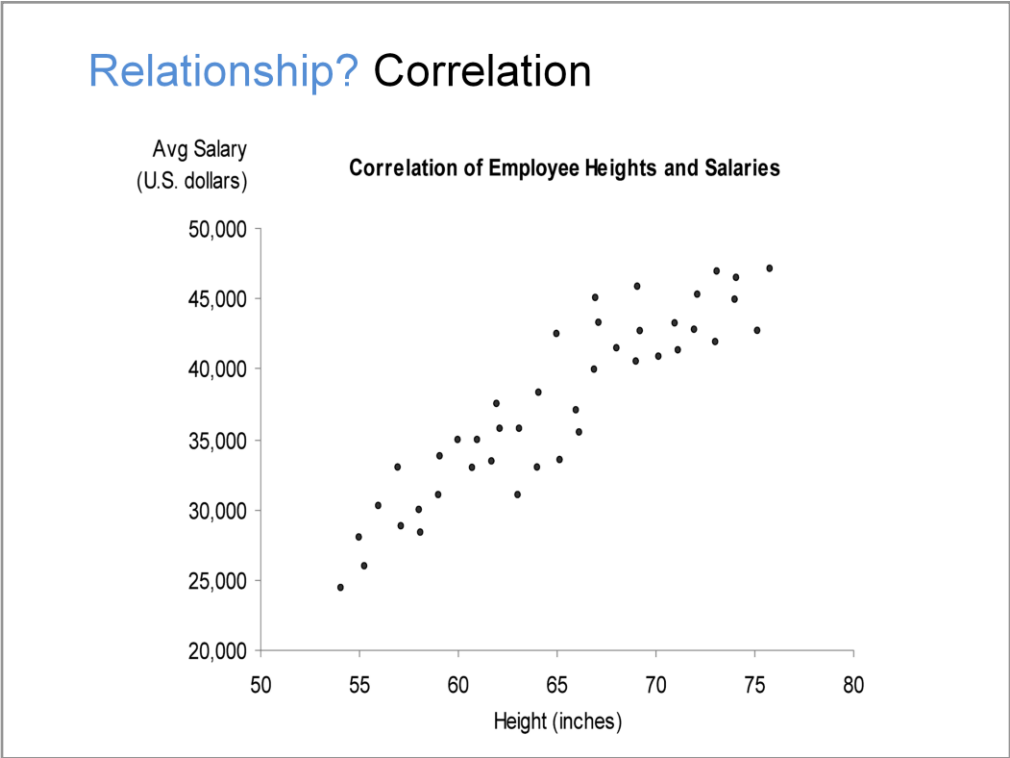
Pie charts are the information visualization equivalent of a roofing hammer to the frontal lobe...[Piecharts] have no place in the world of grownups, and occupy the same semiotic space as short pants, a runny nose, and chocolate smeared on one's face. They are as professional as a pair of assless chaps. Anyone who suggests their use should be instinctively slapped.



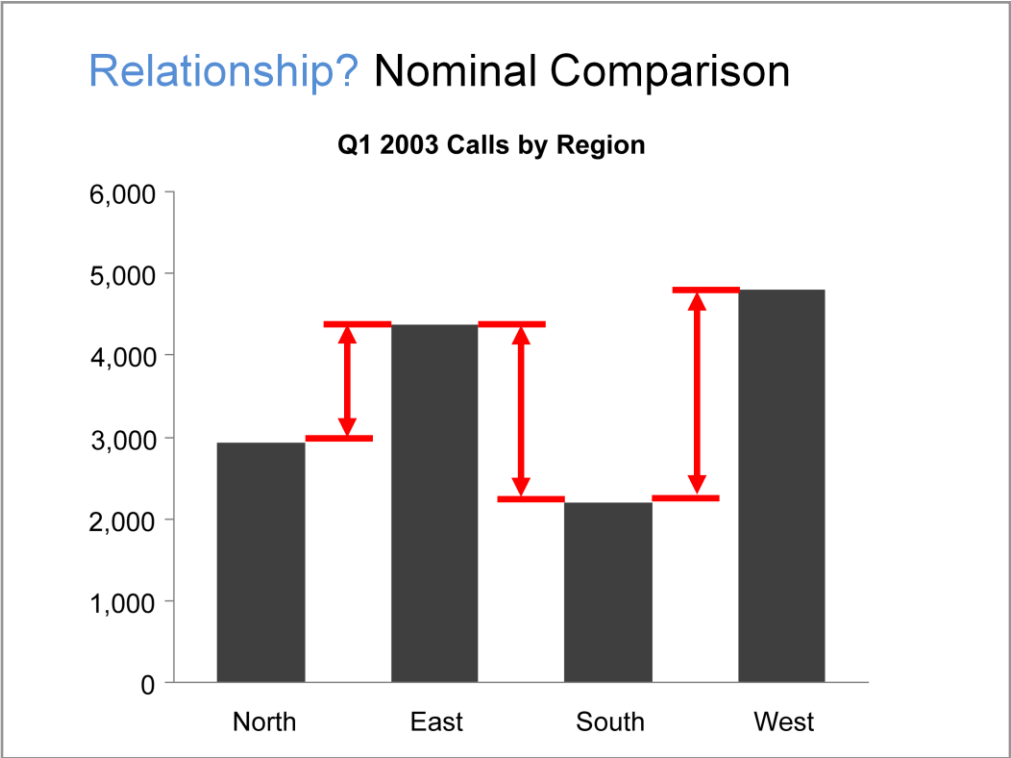
A deviation graph shows how one or more sets of values differ from a reference set of values.



This type of distribution graph, called a frequency distribution, shows the number of times something occurs across consecutive intervals of a larger quantitative range. In a frequency distribution, a quantitative scale (in this case the range of dollar values of orders) is converted to a categorical scale by subdividing the range and giving each of the subdivisions a categorical label (“< \$10”, and so on).



A correlation graph shows whether two paired sets of measures vary in relation to one another, and if so, in which direction (positive or negative) and to what degree (strong or weak). If the trend line moves upwards, the correlation is positive; if it moves downwards, it is negative. A positive correlation indicates that as the values in one data set increase, so do the values in the other data set. A negative correlation indicates that as the values in one data set increase, the values in the other data set decrease. In a scatter plot like this, the more tightly the data points are grouped around the trend line, the stronger the correlation.

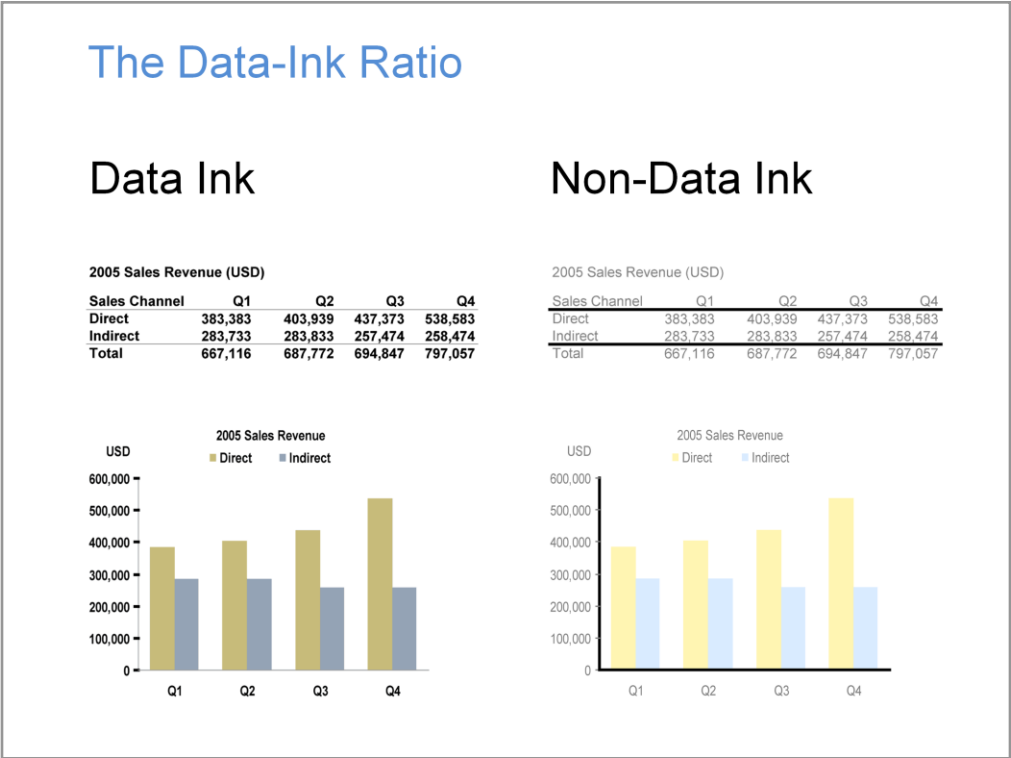


When items relate to one another nominally, they have no particular order. Whenever you find yourself creating a graph with only a nominal relationship, ask yourself if you could improve it by showing another relationship as well, such as a ranking or a part-to-whole.

The seven common relationships in graphs

- Time-series
- Ranking
- Part-to-whole
- Deviation
- Distribution
- Correlation
- Nominal comparison

Without reviewing the last few slides, unless you must as a reminder, try to describe a real-world example of each type of relationship.



According to Edward Tufte, tables and graphs consist of two types of ink: data ink and non-data ink. He introduced the concept of the “data-ink ratio” in his 1983 classic *The Visual Display of Quantitative Data*. He argued that the ratio of ink used to display data to the total ink should be high. In other words, ink that is used to display anything that isn’t data should be reduced to a minimum.

Steps in the design process

1. Reduce the non-data ink.

Remove unnecessary non-data ink.
De-emphasize and regularize the
remaining non-data ink.

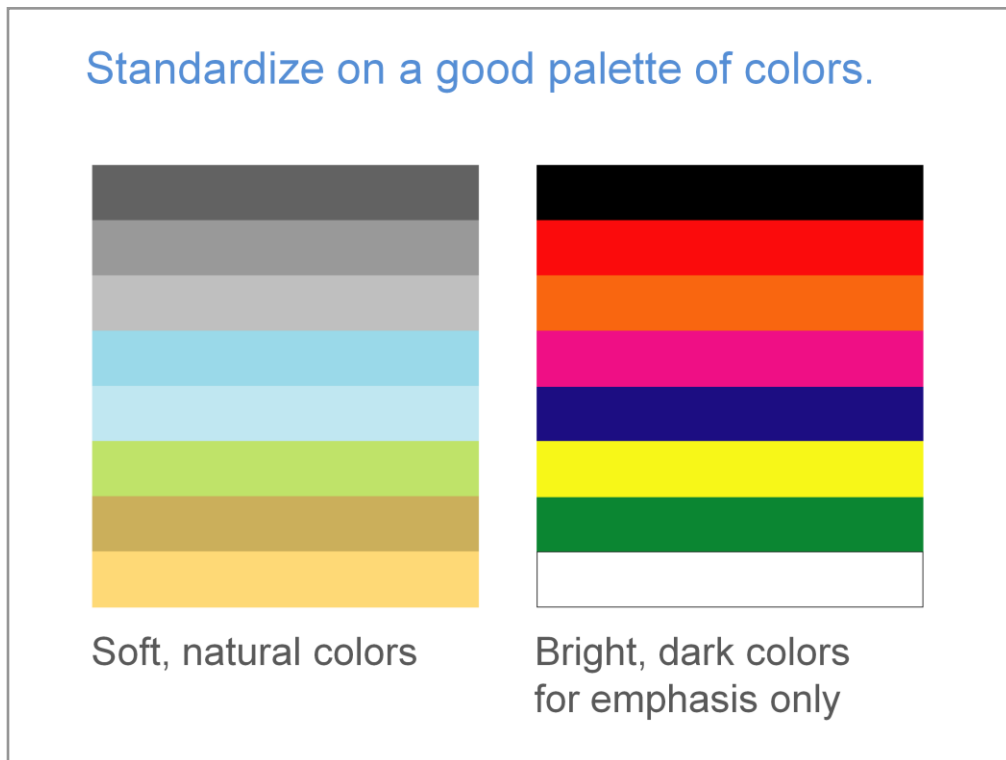
2. Enhance the data ink.

Remove unnecessary data ink.
Emphasize the most important
data ink.

“In anything at all, perfection is finally attained not when there is no longer anything to add, but when there is no longer anything to take away.” Antoine de St. Exupery

John Maeda, in *The Laws of Simplicity*, offers a maxim about design simplicity, which I have massaged into the following statement:

Simplicity is about eliminating the obvious (and everything else that doesn't support your purpose), and enhancing the meaningful.



Color is commonly misused in data displays, which is a shame, because the principles for its effective use are not difficult to master.

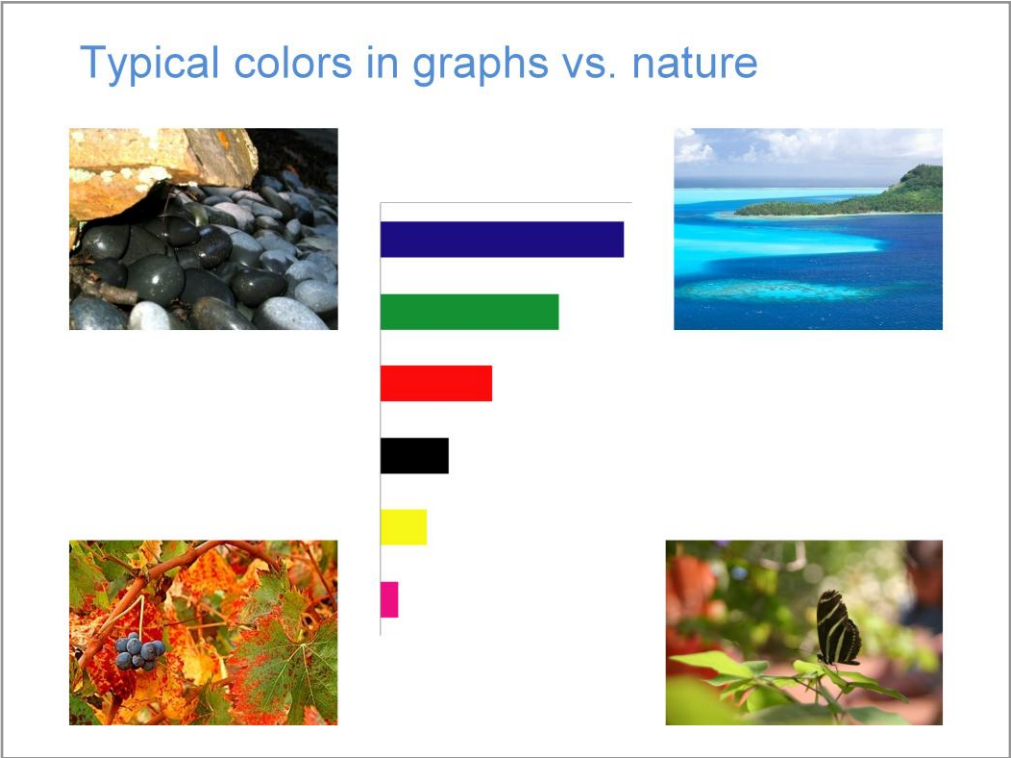
Soft, natural earth tones work best for everything except data that needs to stand out above the rest. Use bright, dark colors only for highlighting data. If your software allows you to customize your color palette, it will definitely save you time to do this once, then rely on those colors for all of your displays.

Useful References

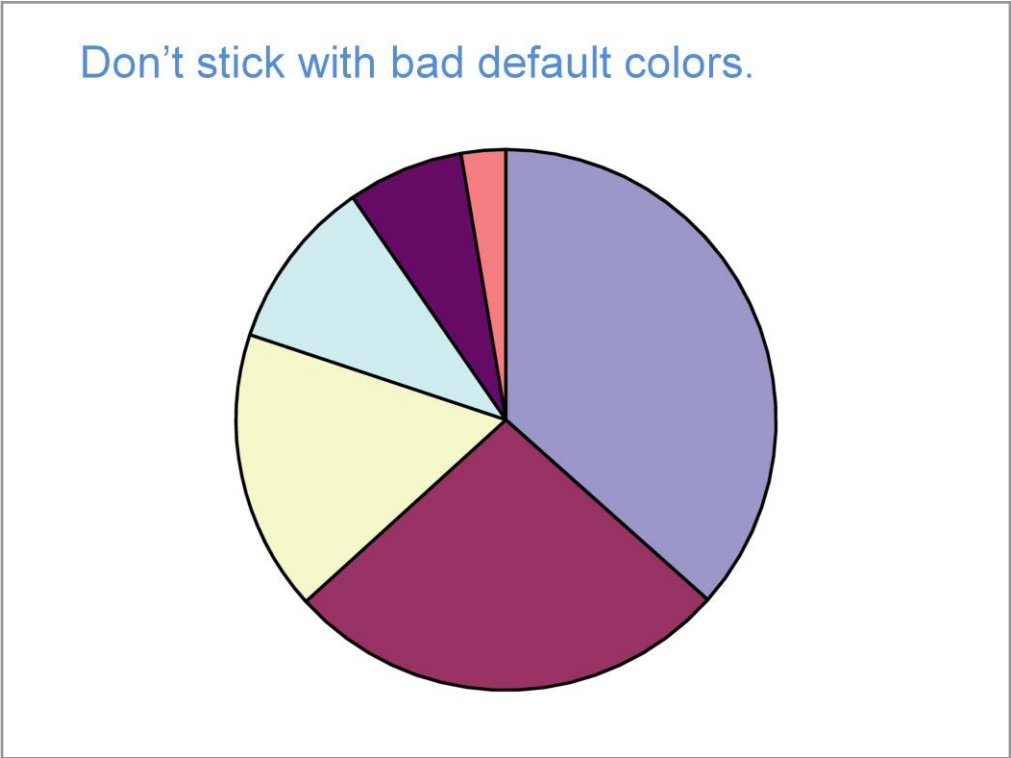
Designing Better Maps, Cynthia Brewer

www.colorbrewer.org, Cynthia Brewer

A Field Guide to Digital Color, Maureen Stone

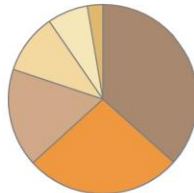
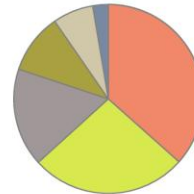
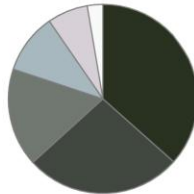
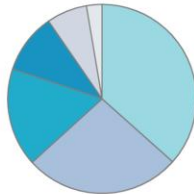


Fully saturated colors are found less often in nature than you might assume. The bright, saturated colors that appear in the bars of this graph cannot be found anywhere in these four photographs of the natural world.



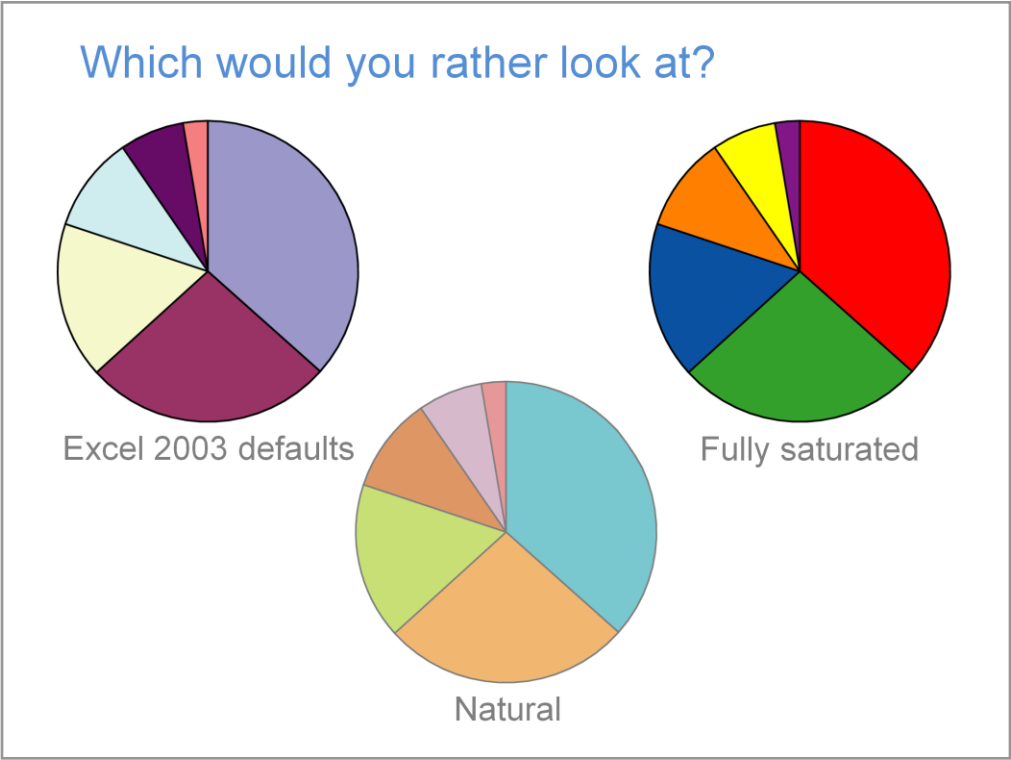
Software products often set the default colors that are used in graphs very poorly. These are the defaults that come with Microsoft Excel 2003. They are not all overly saturated, but they are poorly balanced. Notice how some of the colors, such as the purples, stand out much more than the others.

Nature's colors are so much more polite.

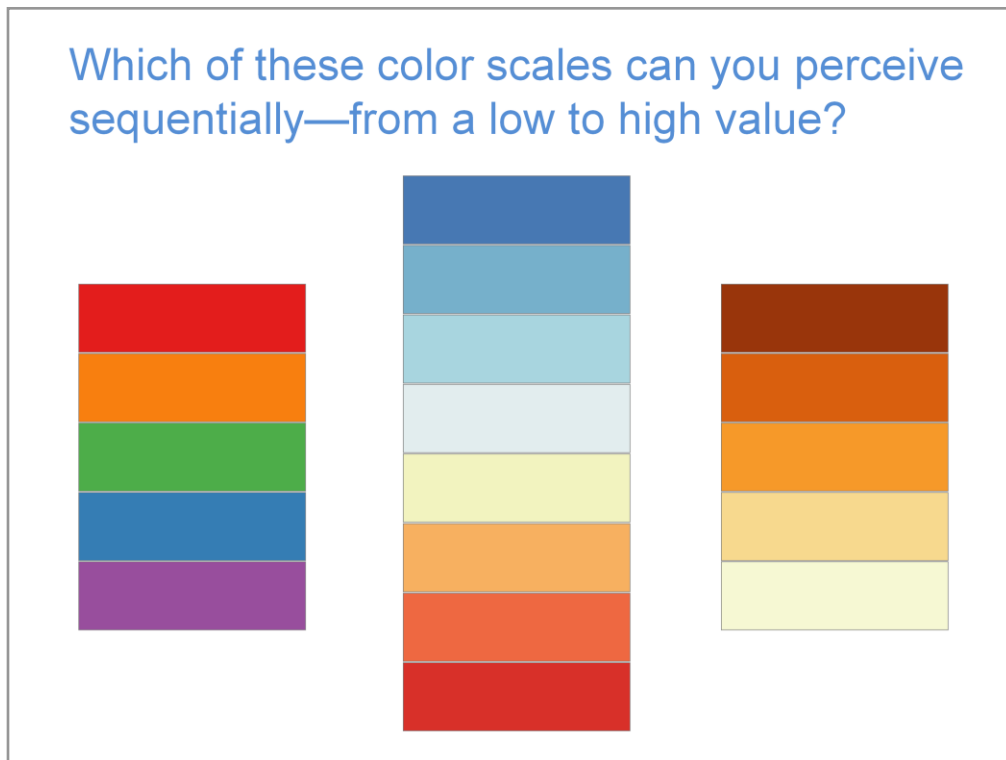


These charts illustrate colors that can be found in nature. It is not my intention to suggest that any of these are ideal color palettes for data presentation, but merely to show how the colors that are commonly found in nature are more pleasant to look at than those that are often provided as defaults in software.

By the way, by displaying these palettes of colors in the form of pie charts, I am not suggesting that you should use this type of graph. A pie chart just happens to provide a nice means to show several color together, but it doesn't provide an effective means of displaying data.



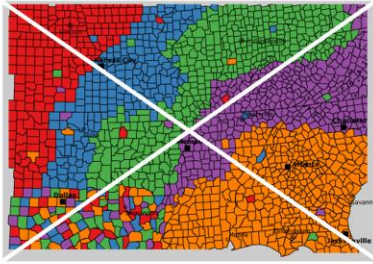
Notice that, despite the softness of the colors in the example of natural colors, they still do the job of separating the sections of these pies just as well as the other examples, but do so in a manner that is much more pleasant to look at.



As you can see, we do not perceive hues in a sequential manner, even when they are arranged according to their spectral values as the example on the left is. The best way to encode sequential values using color is to vary color intensity, especially through variations of lightness, from light to dark. This can be done by varying the intensity of a single hue or by varying hue to some degree as well, as shown in the example to the right, as long as you differentiate the colors primarily by intensity as well.

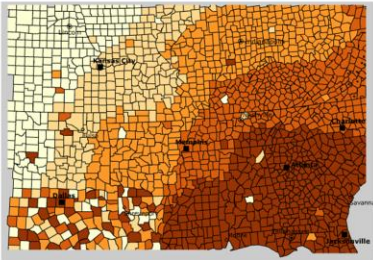
The color scale could be used to encode a single sequential series of values, from the lowest value at the bottom to the highest at the top. Notice that the color scale in the middle consists of two distinct hues (blue in the upper half and red in the lower half). The different expressions of each of these hues vary sequentially. The blue increase in intensity as they proceed upwards, and the reds increase in intensity as they proceed downwards. This is an example of a diverging scale, which is one that has a breakpoint somewhere in the middle, with one set of values the proceeds upwards and one that proceeds downwards from that breakpoint. A typical example of values that a diverging color scale can encode is a set that consists of both positive and negative values, with zero as the breakpoint in the middle. Another example might be values that are above the norm (for example, above average) and those that are below the norm.

Use appropriate colors for quantitative scales.

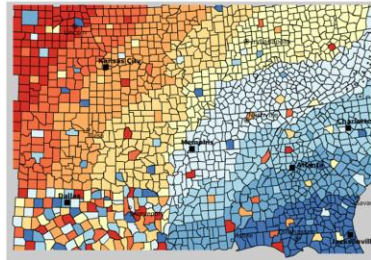


We do not perceive hues sequentially.

But we do perceive color intensities sequentially.



Single sequence scale

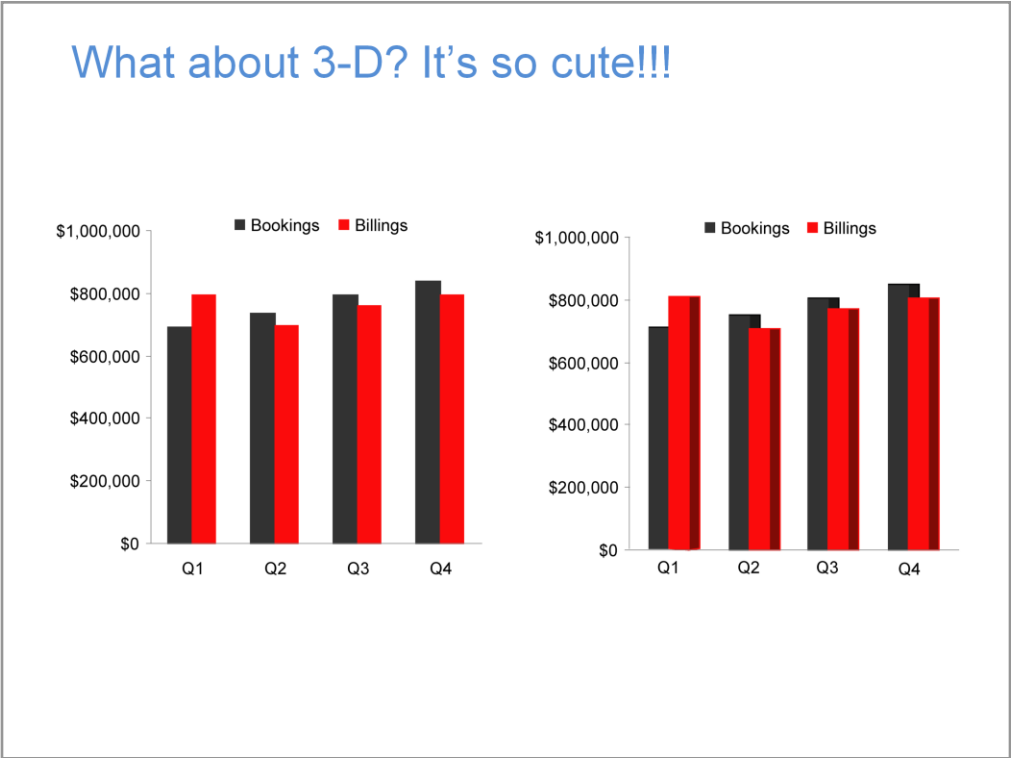


Diverging sequential scale

Here we see the same color scales that appear on the previous slide, but this time shown on a map to illustrate how they might be used for geospatial displays.

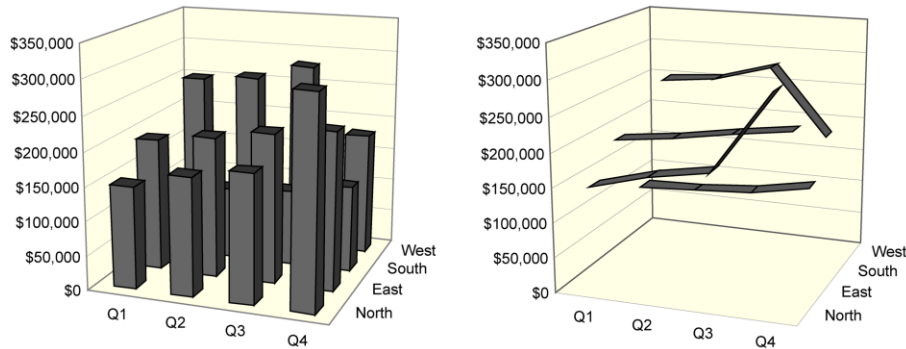
One of the experts in the use of color to encode data on maps, and about geospatial information design in general, is Cynthia Brewer of Penn State University. Dr. Brewer is a cartographer in the geography department. Take advantage of her free *Color Brewer* application at www.ColorBrewer.org. It is worthwhile as well to get a copy of her book *Designing Better Maps: A Guide for GIS Users*, published by ESRI Press in 2005.

As you can see in the top map, different hues work very effectively for assigning discrete categorical distinctions to data (see the top example), but not for displaying a range of quantitative values.



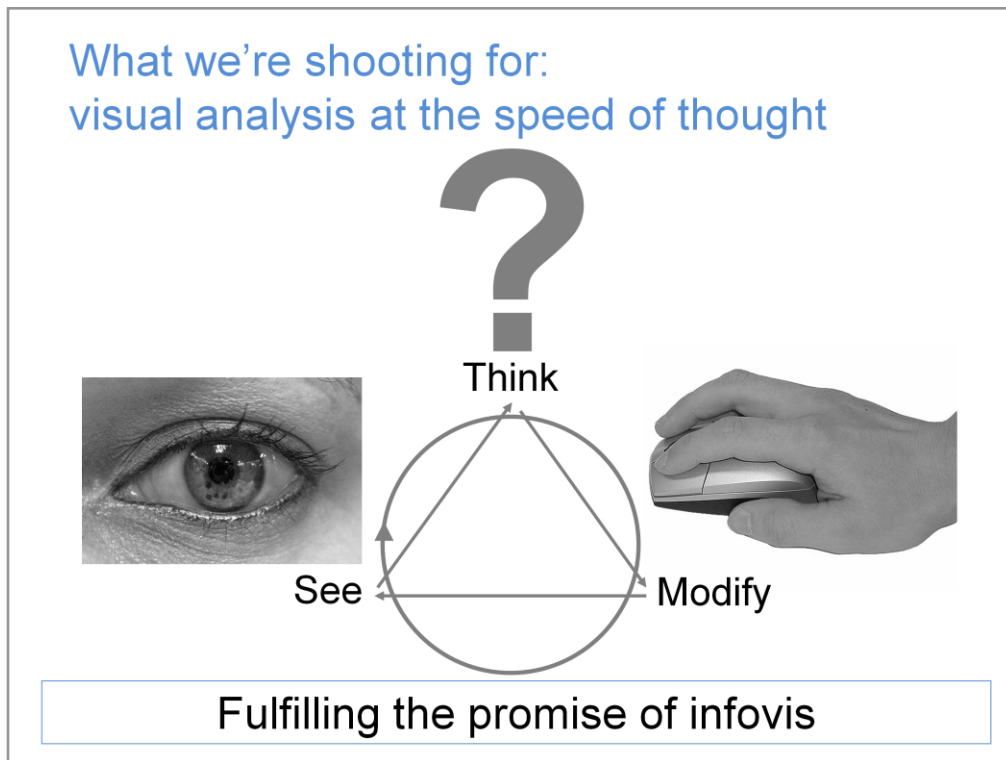
A 3rd dimension without a corresponding variable is not only meaningless, it makes it more difficult to decode the data.

But what if there's a 3rd variable?



Can you determine which of the lines in the graph on the right represents the South region? Are you sure?

A 3rd dimension with a corresponding variable is too hard to read.



Direct dynamic interaction with the visualized data allows you to see something in the data visualization and interact with it directly to filter out what you don't need, drill into details, combine multiple variables for comparison, etc., in a way that promotes a smooth flow between seeing something, thinking about it, and manipulating it, with no distracting lags in between.

Shneiderman's mantra for effective navigation

“Overview first,
zoom and filter,
then details-on-demand.”

When new recruits by intelligence organizations are trained in spy craft, they are taught a method of observation that begins by getting an overview of the scene around them while being sensitive to things that appear abnormal, not quite right, which they should then focus in on for close observation and analysis.

A visual information-seeking mantra for designers: ‘Overview first, zoom and filter, then details-on-demand.’

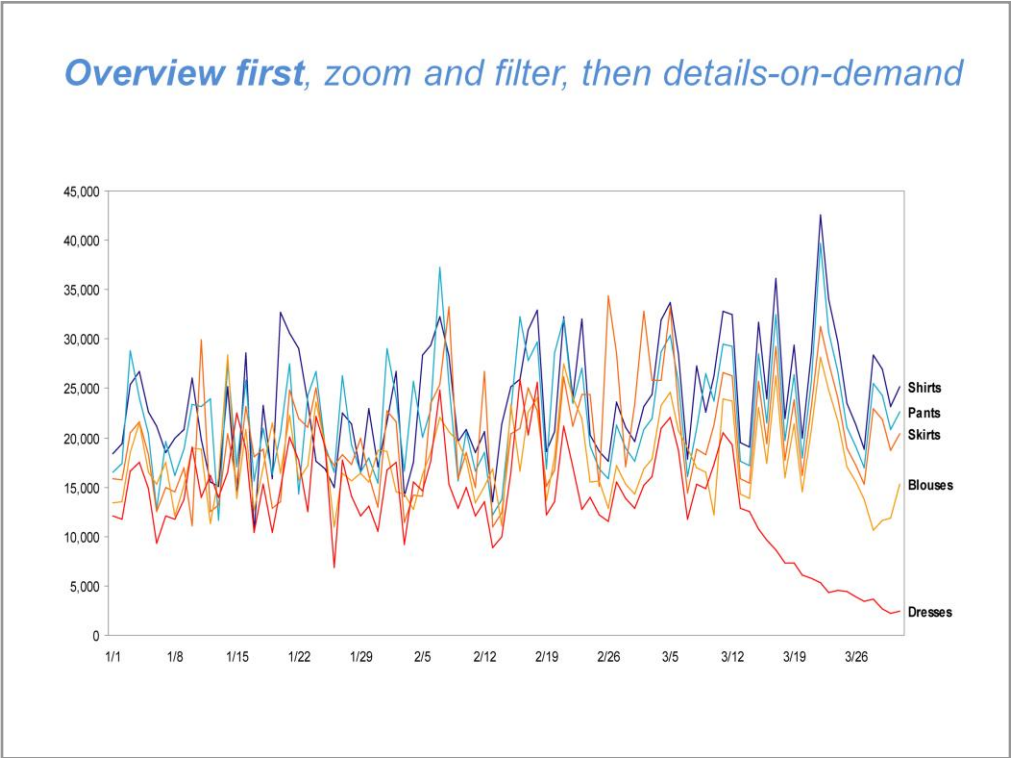
Readings in Information Visualization: Using Vision to Think, Stuart K. Card, Jock D. Mackinlay, and Ben Shneiderman, Academic Press, San Diego, California, 1999, page 625.

Having an overview is very important. It reduces search, allows the detection of overall patterns, and aids the user in choosing the next move. A general heuristic of visualization design, therefore, is to start with an overview. But it is also necessary for the user to access details rapidly. One solution is overview + detail: to provide multiple views, an overview for orientation, and a detailed view for further work.

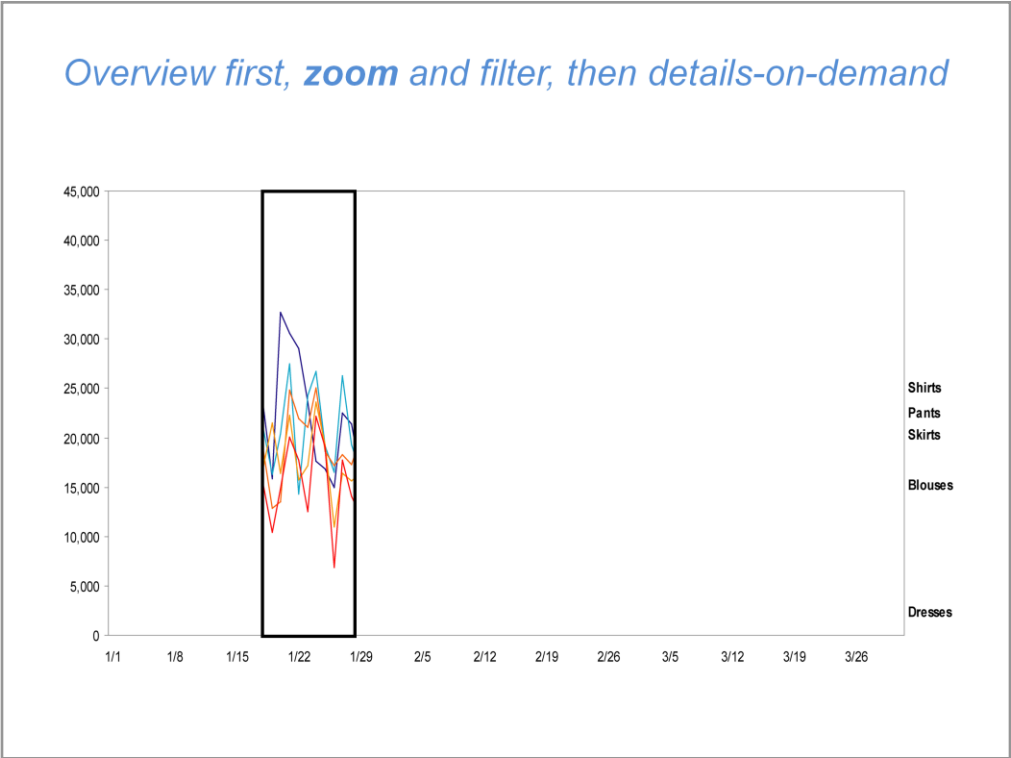
Ibid., page 285.

Users often try to make a ‘good’ choice by deciding first what they do not want, i.e. they first try to reduce the data set to a smaller, more manageable size. After some iterations, it is easier to make the final selection(s) from the reduced data set. This iterative refinement or progressive querying of data sets is sometimes known as hierarchical decision-making.

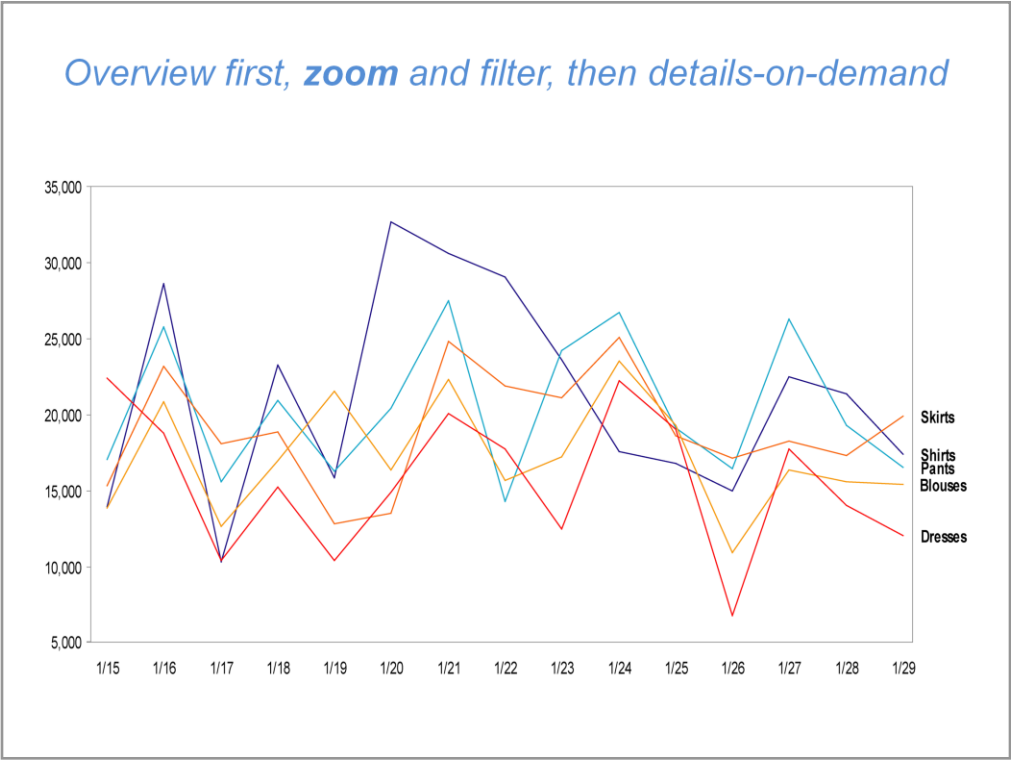
Ibid., page 295.



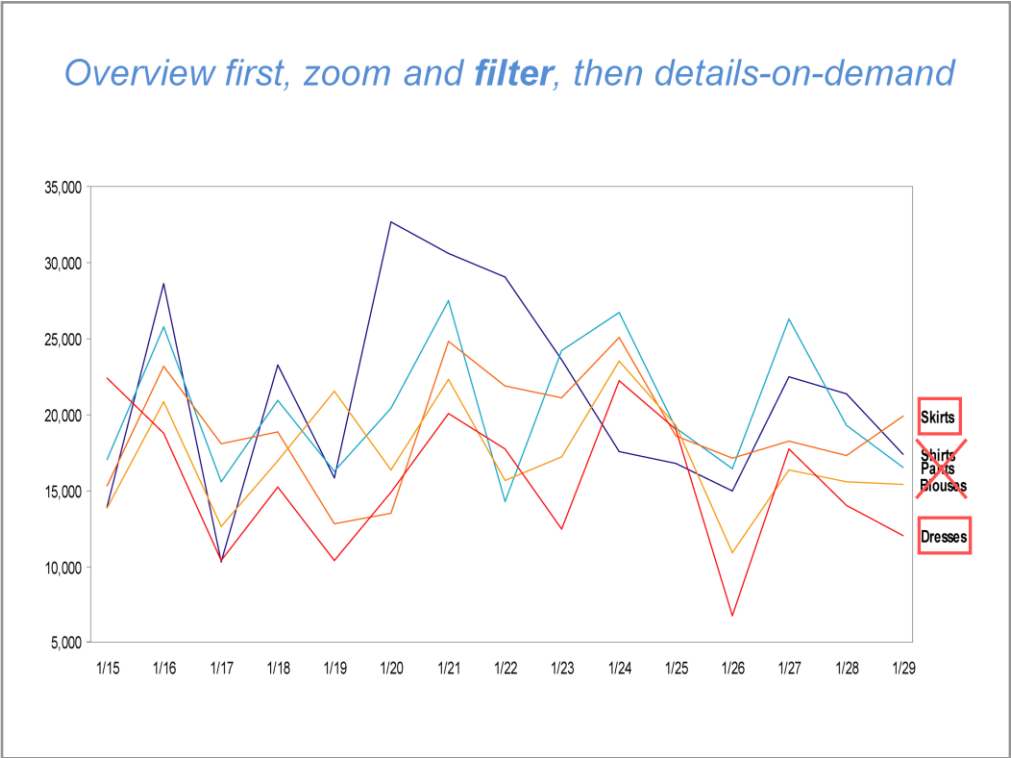
Ben Shneiderman’s technique begins with an overview of the data – the big picture. Let your eyes search for particular points of interest in the whole.



When you see a particular point of interest, then zoom in on it.



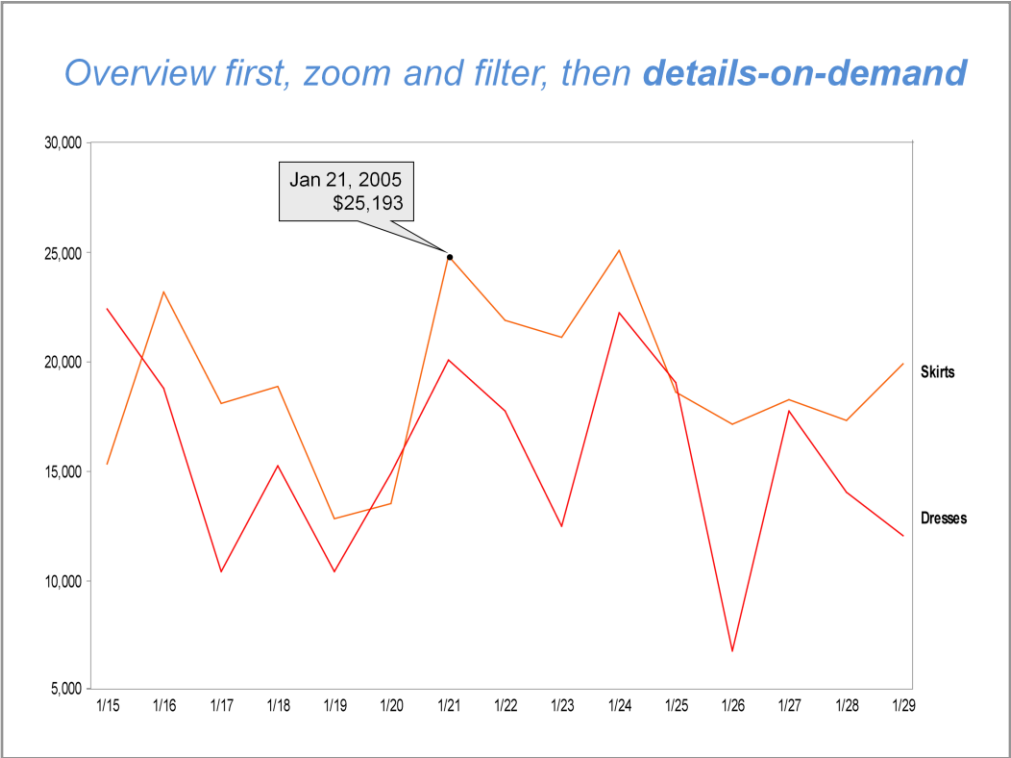
Once you've zoomed in on it, you can examine it in greater detail.



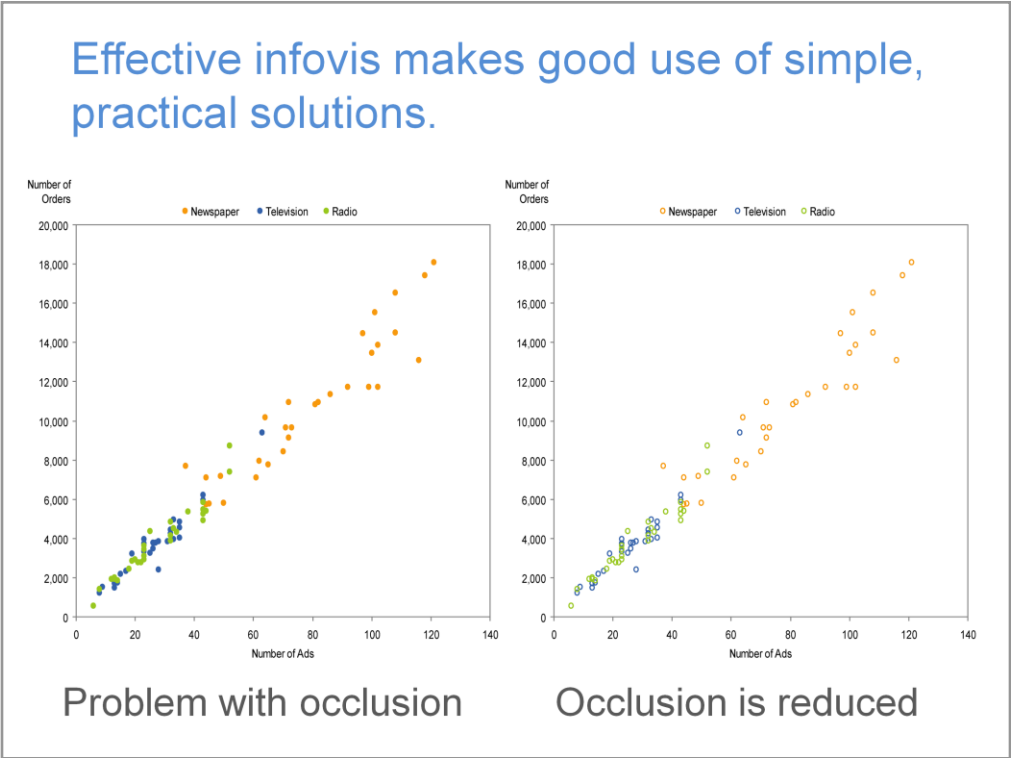
Often you must remove data that is extraneous to your investigation in order to better focus on the relevant data.



Filtering out extraneous data removes distractions from the data under investigation.



Visual data analysis relies mostly on the shape of the data to provide needed insights, but there are still times when you need to see the details behind the shape of the data. Having a means to easily see the details when you need them, without having them in the way when you don't works best.



If many data points are close together and overlap one another in the scatter plot, this problem can be easily reduced by removing the fill color from the points, leaving only their outlines. This makes it easier to see when data points overlap.

10 questions customers should always ask about infovis software

1. Can you make sense of the data with little training?
2. Does it make meaningful patterns, trends, and exceptions easy to see and interpret?
3. Does it encode quantitative data accurately?
4. Does it avoid visual distractions?
5. Is it pleasant to look at?
6. Can you use it to easily answer real business questions?
7. Can you interact with the data without losing your train of thought?
8. Does it encourage you to do things right and discourage you from doing things wrong?
9. Can you easily find what you need?
10. Does it provide the flexibility required to do everything you need to make sense of the data?

Let me suggest 10 questions that you should ask of visual analysis software when you're out there doing your shopping.

- When you look at it, can you make sense of the data with relatively little training?
- Does it make the most meaningful patterns, trends, and exceptions easy to see and interpret?
- Does it encode quantitative data accurately (that is, based on the visual display itself, can you accurately interpret the quantities it encodes and accurately compare them to one another)?
- Does it avoid features that distract you from the data?
- Is it pleasant to look at for long periods of time without undo visual fatigue?
- Can you use it to answer real business questions? If so, how long does it take? Anyone who is responsible for making sense of data in the real business world knows that questions that are hard to answer with the tools you have end up not getting asked.
- When you use it to interact with the data (for example, to filter the data), can you do so in a manner that support your flow of thought about the data without interruption?
- Does it encourage you to do things right and discourage you from doing things wrong (for example, to use chart types that make no sense given the nature of the data that you are analyzing)?
- Can you easily find and use the types of visualizations, data interactions, and calculations that are needed most often without wading through intimidating menus and lists to get to them?
- Does it provide the flexibility required to easily display and interact with the data in the full range of meaningful ways that come to mind as you work to make sense of the data?

Effective commercial products emerge where
infovis research and the real world intersect.

