Table and Graph Design
for Enlightening Communication

Stephen Few, Perceptual Edge
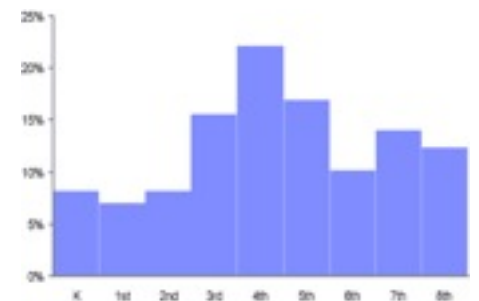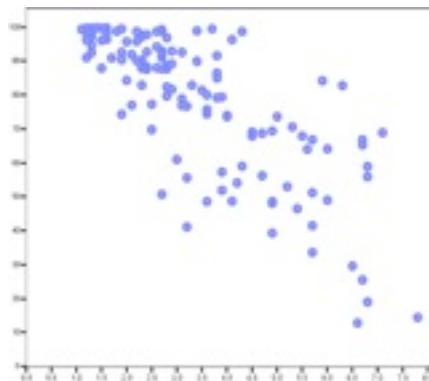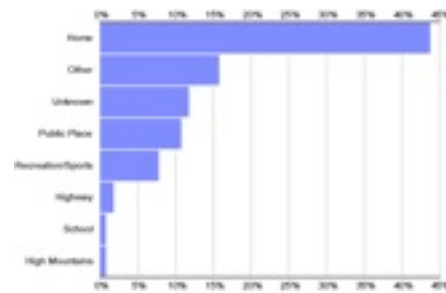
Stephen Few, Principal, Perceptual Edge
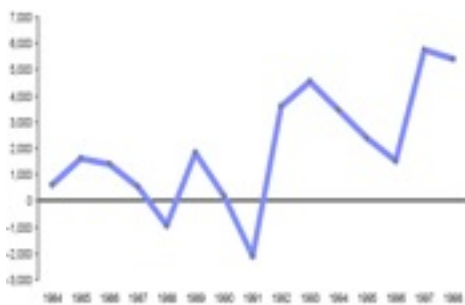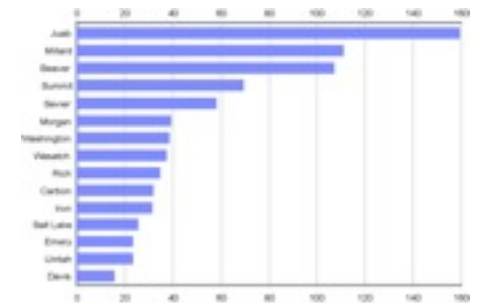sfew@perceptualedge.com
(510) 558-7400

We all have important stories to tell.

Some stories involve numbers.

I work with people and organizations of all types. If they've requested my services, they have one thing in common—they all have important quantitative information to present, and they recognize they could be doing it much better.

We've been telling them with graphs for quite awhile.

Interest of the NATIONAL DEBT from the Revolution.

INTEREST OF NATIONAL DEBT.

The Bottom line is Years, those on the Right hand Millions of Pounds.
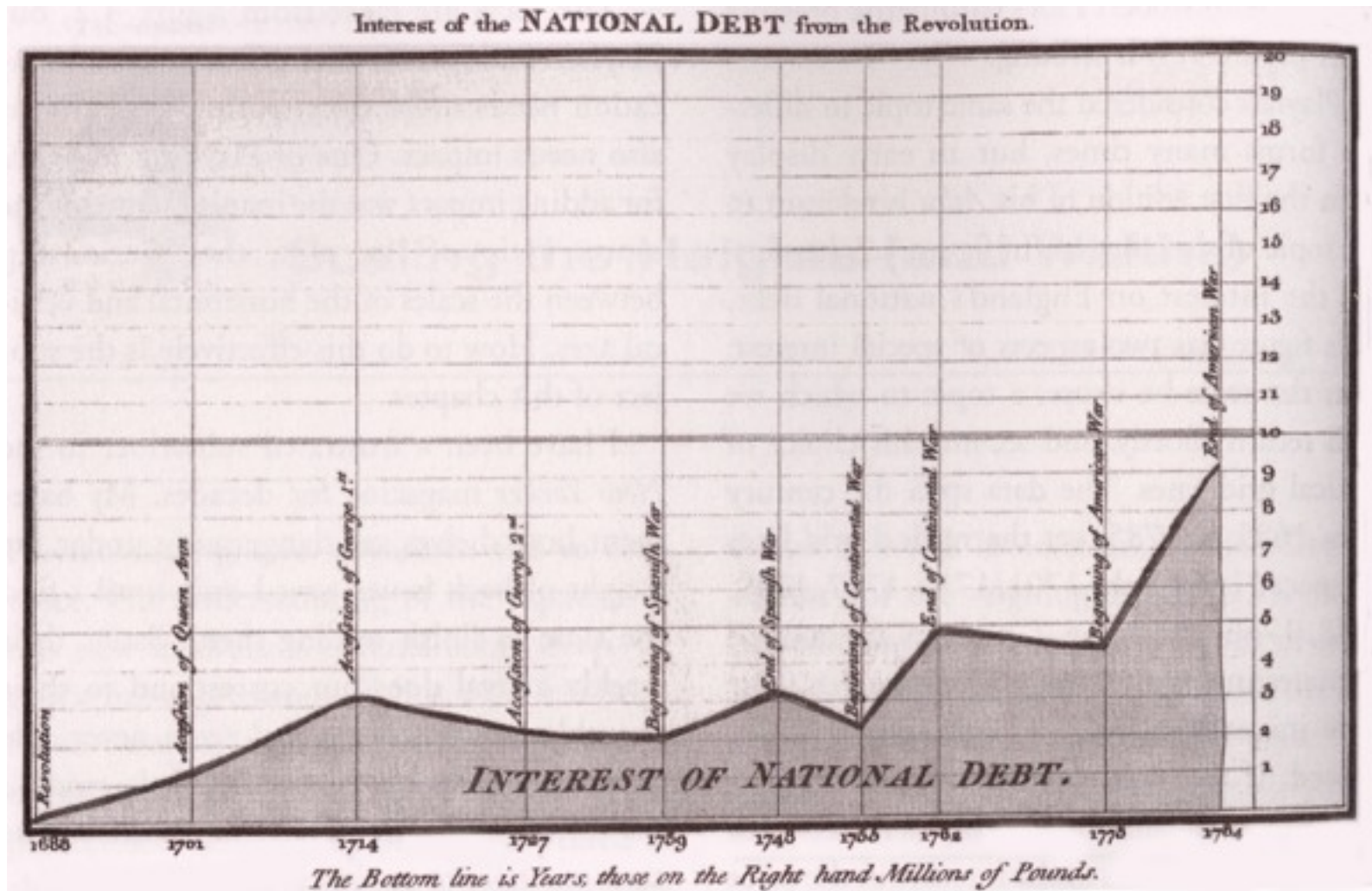
William Playfair – 1786

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.

(Source: This graph was included in Playfair's *The Commercial and Political Atlas* in 1786 to make a case against England's policy of financing colonial wars through national debt.)
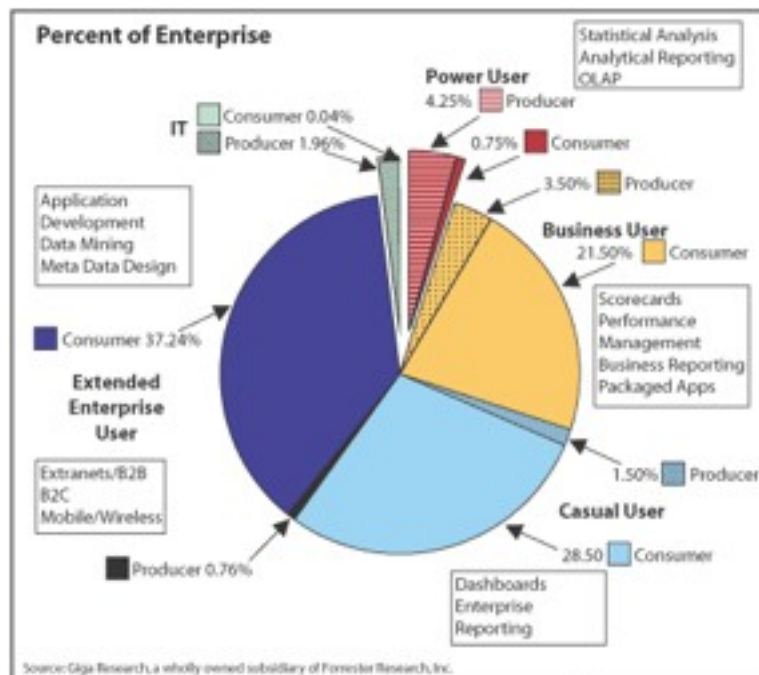
Graphs are now commonplace.
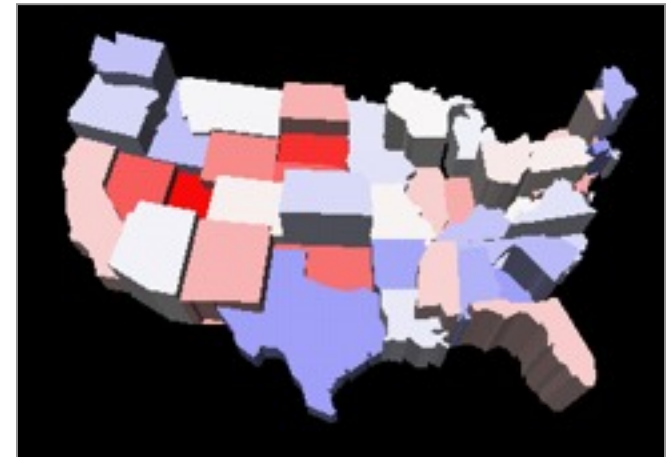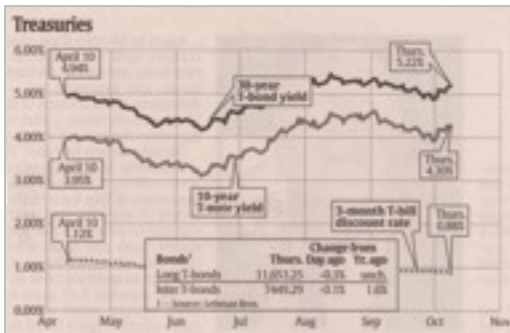

Figure 2: Analytic End-User Spectrum of Usage

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.

But most graphs today communicate poorly.

Floods of data collect in ever-expanding, stagnant pools.
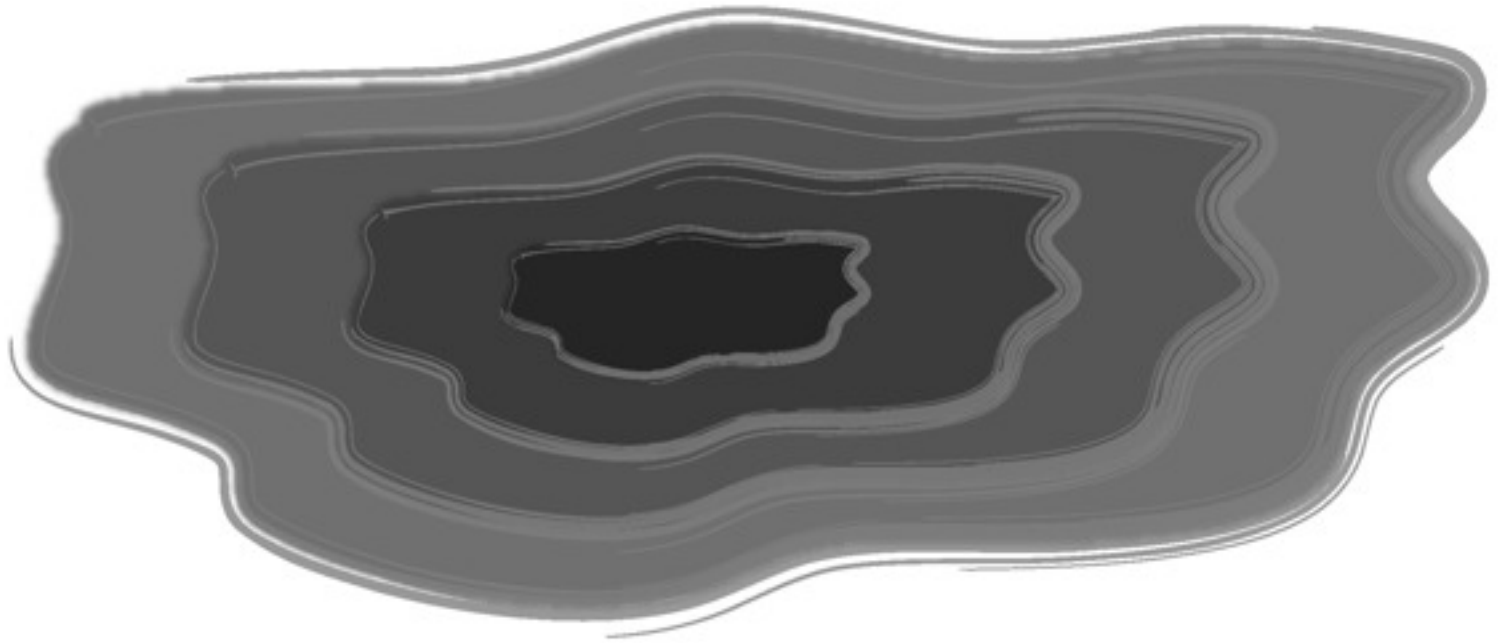
Information is useless until we understand what it means and can clearly communicate that meaning to those who need it, those whose decisions affect our world.

Finally...

Effective network monitoring has arrived!

- Near real-time

- Phenomenally user-friendly

- Instant insight→effective response

Imagine that you've been invited to another of those many meetings that you're required to attend. You're one of several managers in the IT department. Like most meetings, this one begins with the light of a projector suddenly illuminating a screen. Bursting with excitement, a young fellow at the front of the room announces that you will now receive a daily report that will inform you how the network is being utilized, and then the graph on the next slide appears.

Network Volume by Application
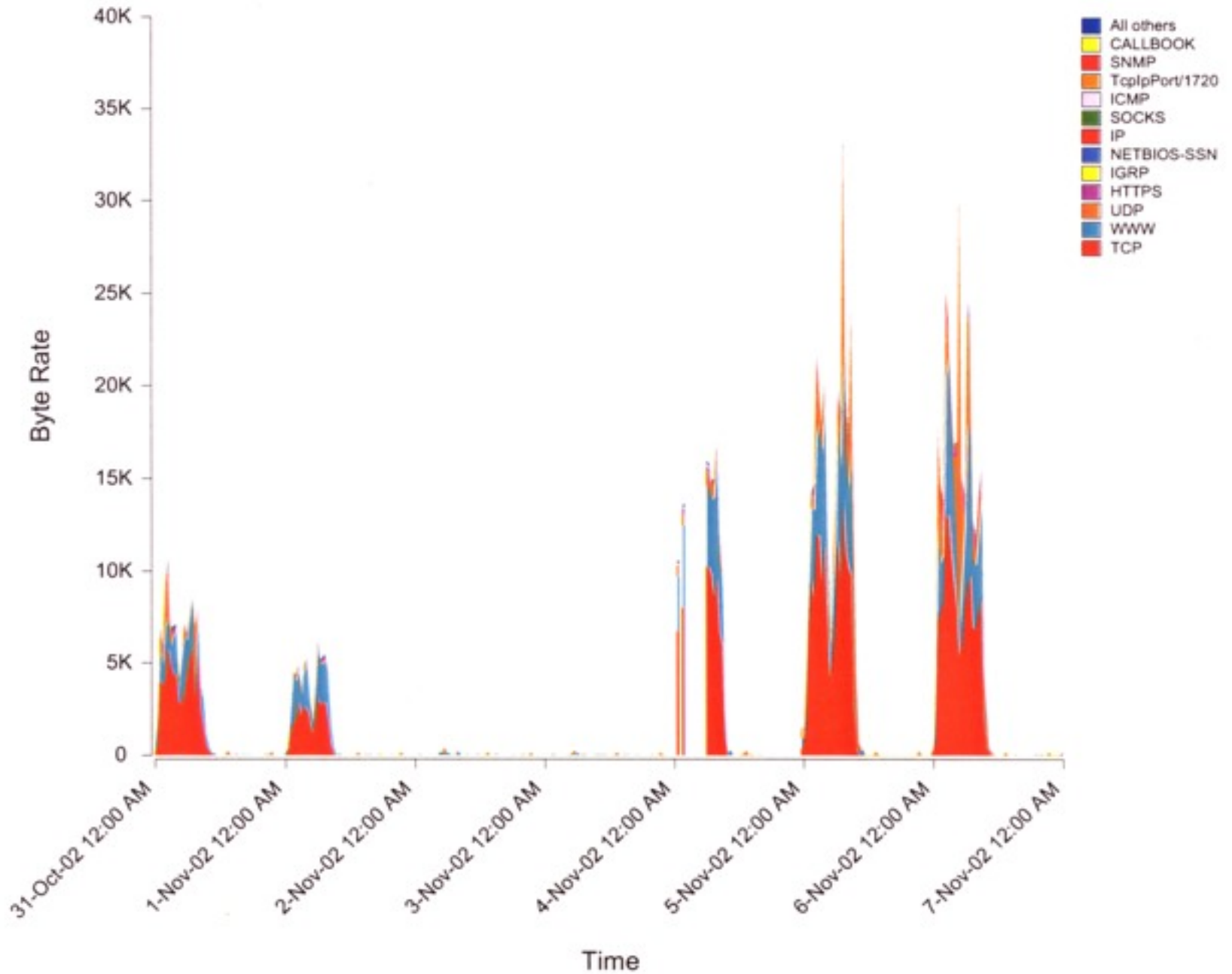
You stare at this graph intently, trying your best to keep any hint of confusion from showing on your face. From your peripheral vision you can see that the CIO (Chief Information Officer) is smiling broadly and nodding with obvious understanding. You and everyone else in the room begin to nod enthusiastically as well. You feel dumb, because you have no idea what this graph is trying to say. What you don't realize is that you are not alone.
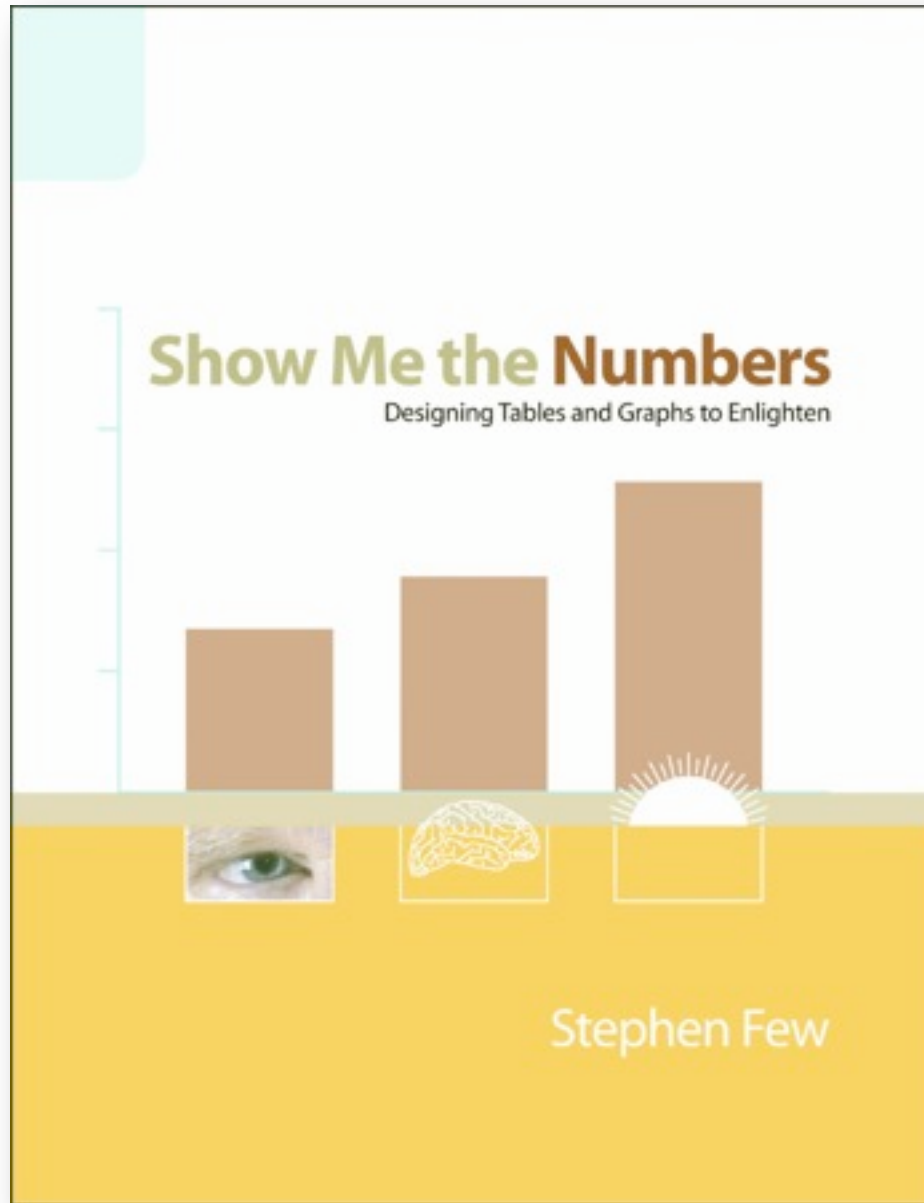
In 2004, I wrote the book *Show Me the Numbers: Designing Tables and Graphs to Enlighten* to help people like you respond in practical ways to the challenges that you face every day when presenting quantitative information.

We are awash in data.

"Just show me the numbers!"
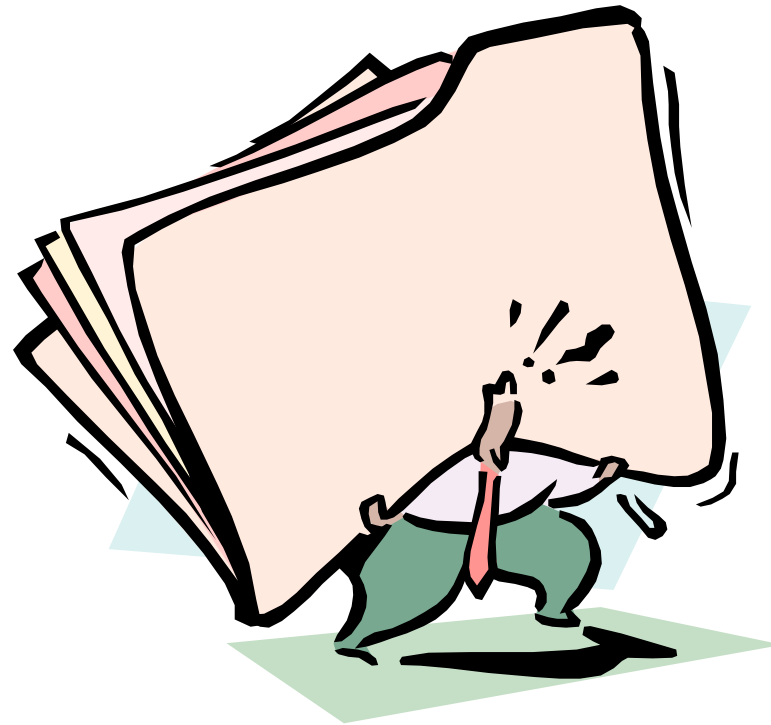
The phrase, "Just show me the numbers," is one that I heard from time to time on the lips of business people, especially popular those responsible for sales organizations, frantic to know how sales are going. They couldn't afford to wade through lengthy reports and unnecessary detail; they just want to see the important numbers right now!
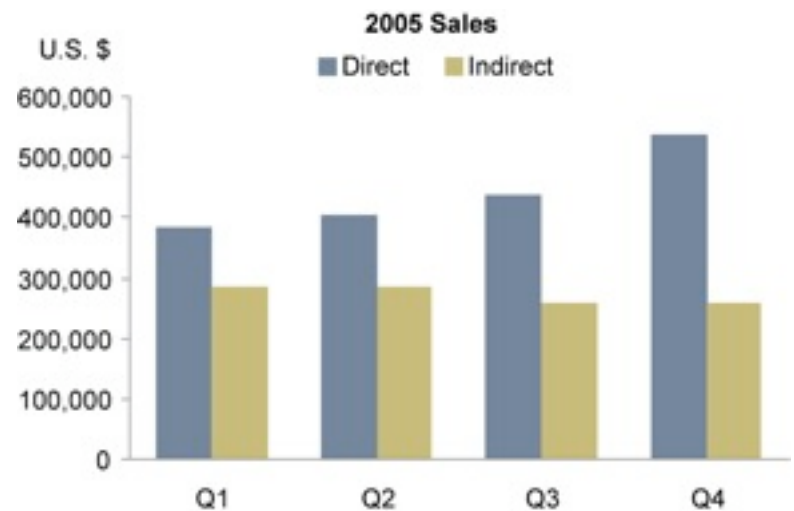
Everyone is scrambling for metrics, key performance indicators (KPIs), scorecards, and digital dashboards. Quantitative data is what we rely on most to measure the health of our organizations, to identify opportunities, and to anticipate the future.

Despite great progress in our ability to gather and warehouse data, we're still missing the boat if we don't communicate the numbers effectively. Contrary to popular wisdom, information cannot always speak for itself.

You can spend millions of dollars to build the most robust and pristine data warehouse in the world, running on the most powerful hardware, and accessed by state-of-the-art Business Intelligence software, but if the people who work with the data don't know how to make sense of it or how to present it clearly to decision makers, your investment is wasted.

Quantitative information is primarily communicated through **tables and graphs**.

Expenses				
Department	Jan	Feb	Variance	Change %
Sales	9,933	9,293	-640	-6%
Marketing	5,385	5,832	+447	+8%
Operations	8,375	7,937	-438	-5%
Total	$23,693	$23,062	-$1,327	-3%

2005 Sales

- Direct
- Indirect

(bar chart, U.S. $ from 0 to 600,000; quarters Q1, Q2, Q3, Q4)

But few communicate effectively. **Why?**

Why? Few people are trained.

Why? Few people recognize the need.

Why? Few examples of good design exist to expose the problem.

"Poor documents are so commonplace that deciphering bad writing and bad visual design have become part of the coping skills needed to navigate in the so-called information age." Karen A. Schriver, *Dynamics in Document Design*, John Wiley & Sons, Inc., 1997.

"The public is more familiar with bad design than good design. It is, in effect, conditioned to prefer bad design, because that is what it lives with. The new becomes threatening, the old reassuring." (Kevin Mullet and Darrel Sano, *Designing Visual Interfaces*, Sun Microsystems, Inc., 1995 – quoting Paul Rand, *Design, Form, and Chaos*)

Effective communication is not always intuitive – it must be learned.

# Intentional deceit is no longer our biggest problem.
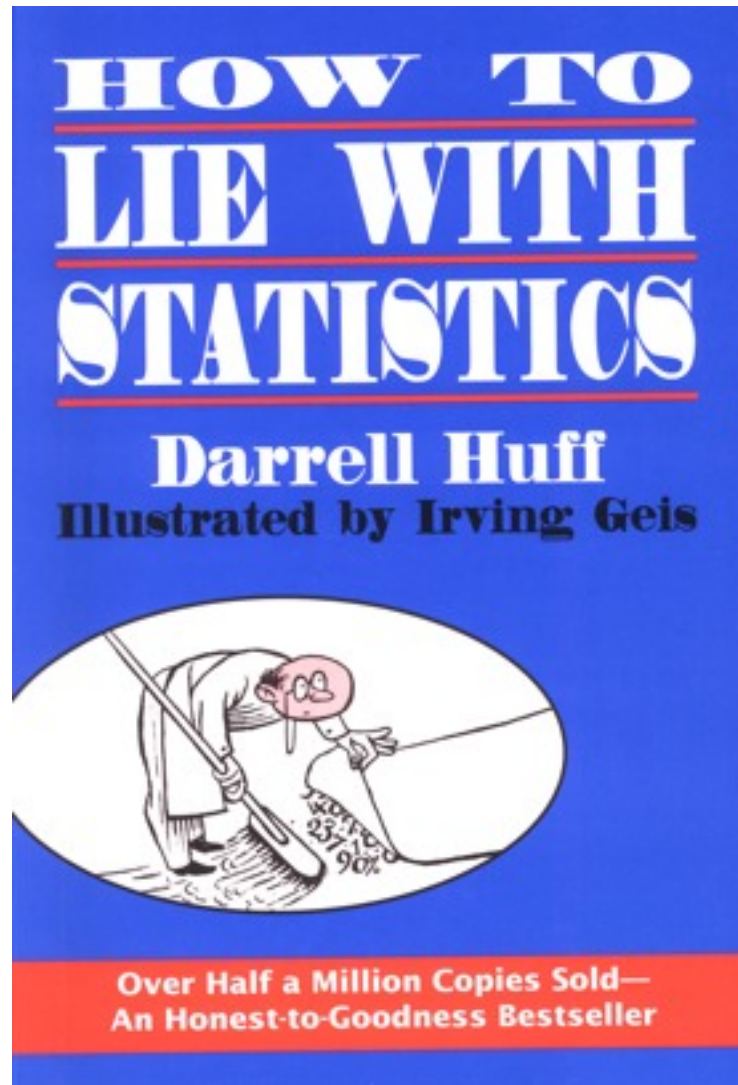
In 1954, Darrell Huff wrote his best-selling book about how people were often intentionally using statistics, including graphs, to spread misinformation, especially in favor of their own products or causes. Today, vastly more misinformation is disseminated unintentionally because people don't know how to use charts to communicate what they intend.

Numbers are commonly obscured, then dressed up to look *sexy*.
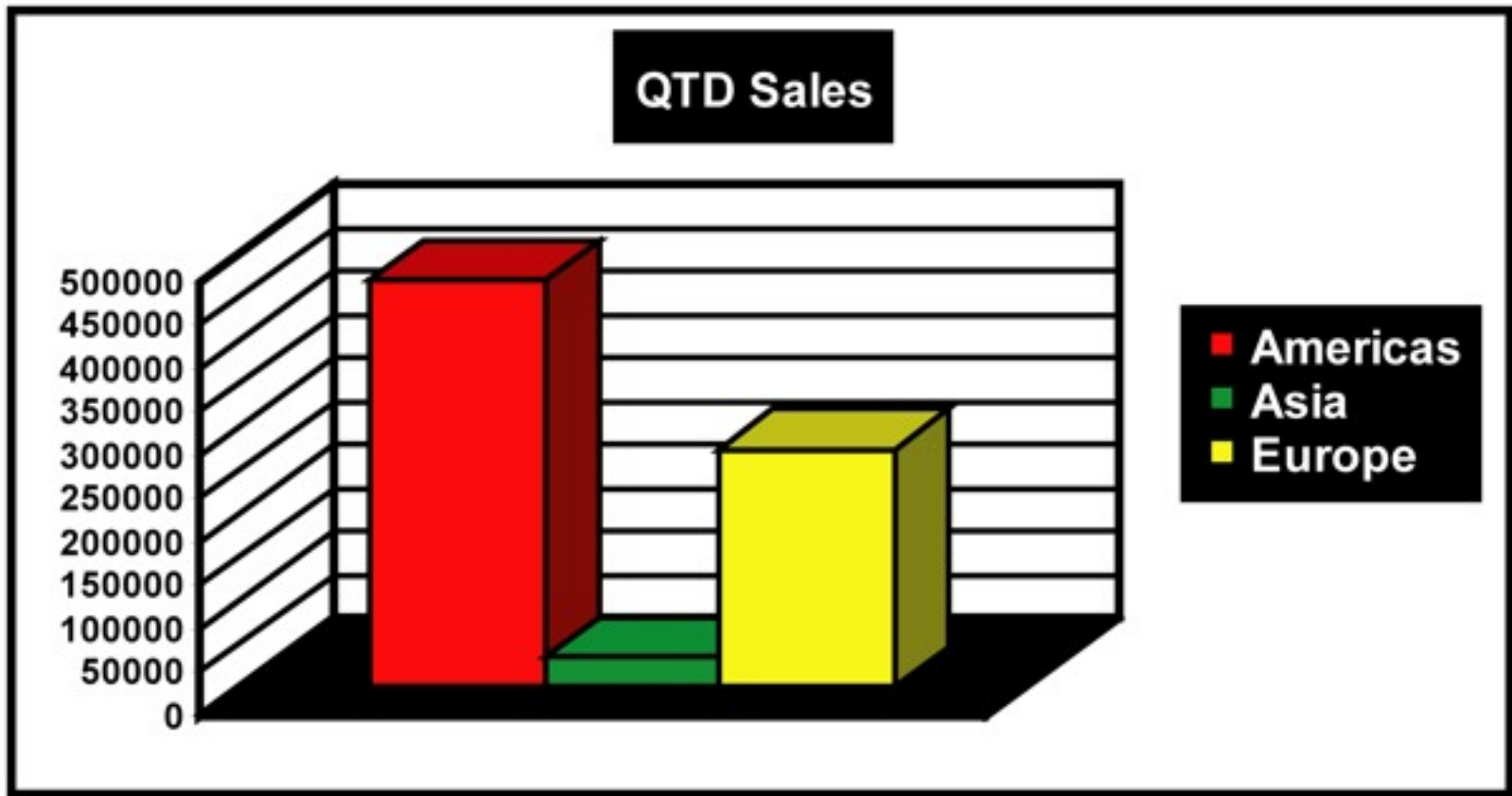
Why?

2007 SALES chart
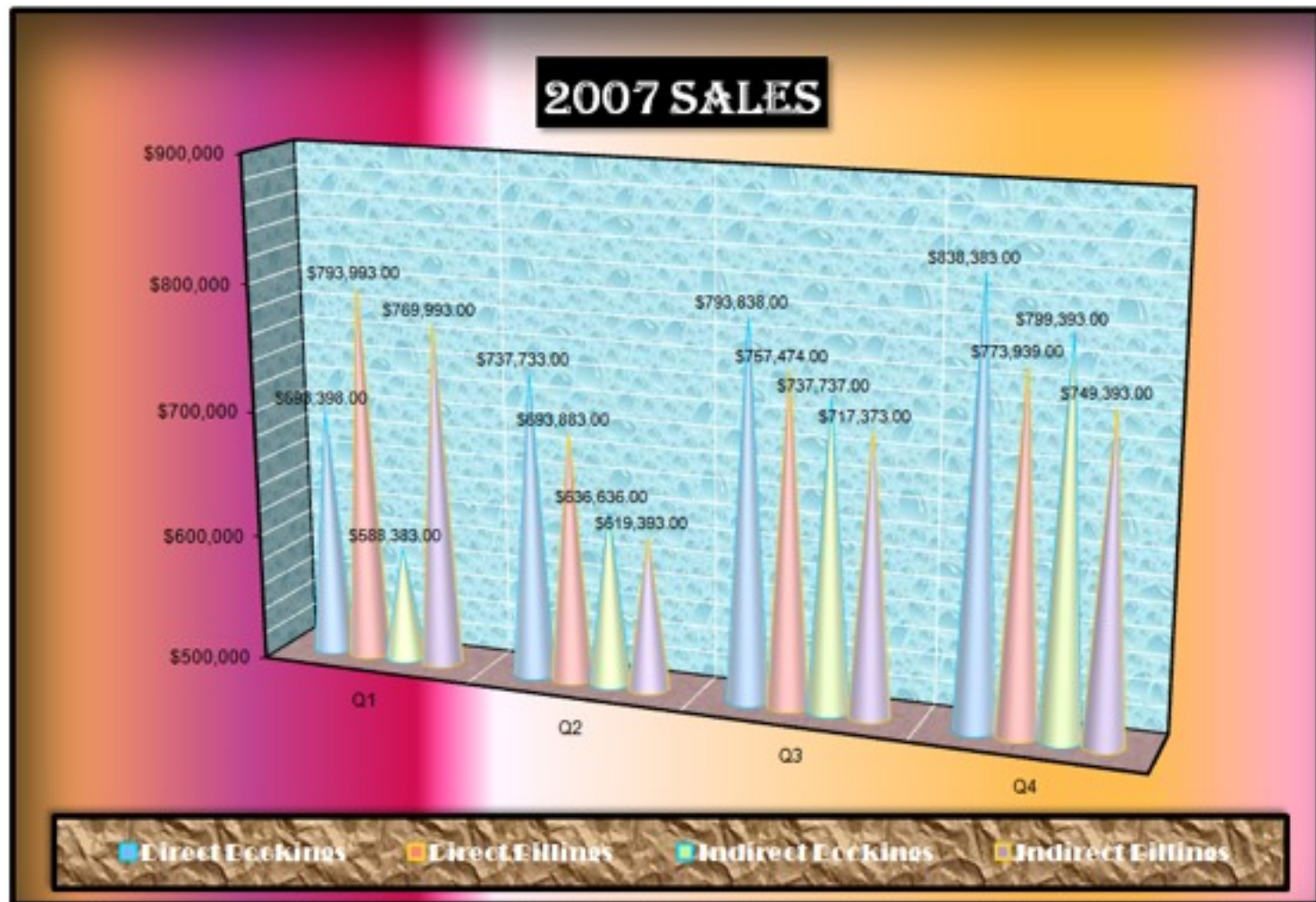
When the PC was introduced, software soon made the arduous task of table and graph creation as easy as 1-2-3 (literally "Lotus 1-2-3", the software that was the first to legitimize the PC as a viable tool for business). Unfortunately, this improvement in ease and efficiency was not accompanied by instruction 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

I can talk about this all day, but the best way to make my point convincingly is to show you.

Example #1 - Poor

QTD Sales

Value
500000
450000
400000
350000
300000
250000
200000
150000
100000
50000
0

- Americas
- Asia
- Europe

What does this graph tell you? Is the resulting information worth the effort?

Example #1 - Good

2003 Q1-to-Date Regional Sales
March 15, 2003

	Sales (U.S. $)	Percent of Total Sales	Current Percent of Qtr Plan	Qtr End Projected Sales (U.S. $)	Qtr End Projected Percent of Qtr Plan
Americas	469,384	60%	85%	586,730	107%
Europe	273,854	35%	91%	353,272	118%
Asia	34,847	5%	50%	43,210	62%
	$778,085	100%	85%	$983,212	107%

Note: To date, 83% of the quarter has elapsed.

This table presents the same information that appears in the graph and more, but it does so clearly and simply. One common problem in the display of quantitative information is that people often choose the wrong medium of display – a graph when a table would work better and vice versa. Too seldom do report developers consider their message and carefully design its presentation to communicate that message effectively.
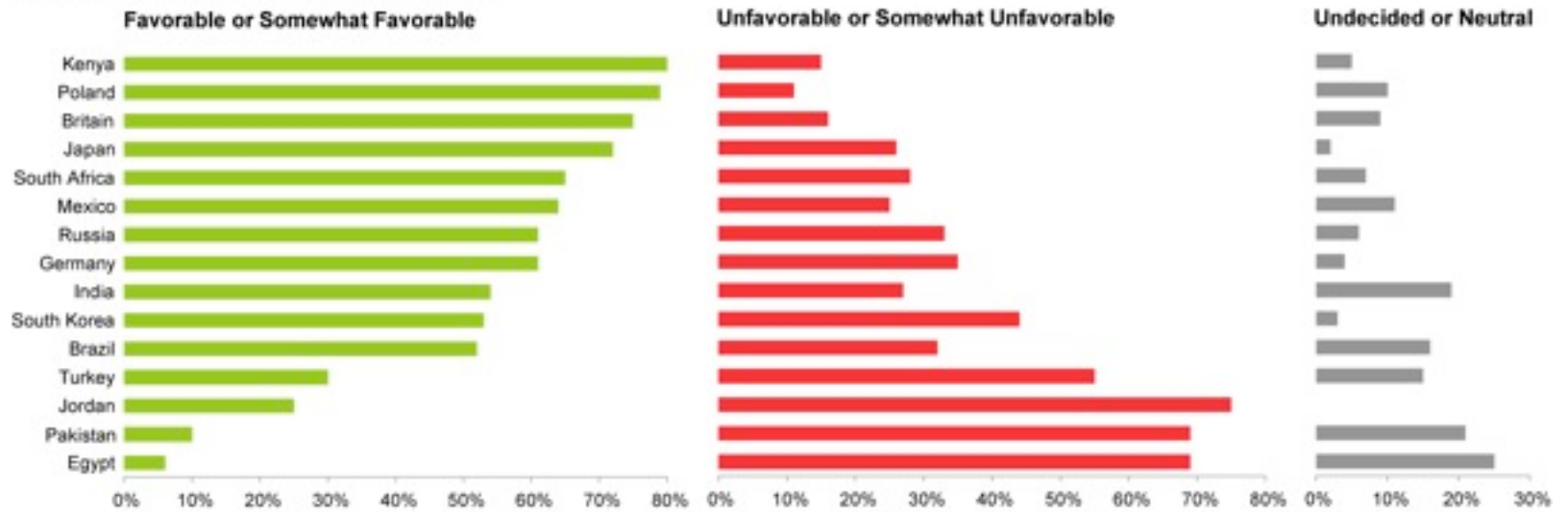
Example #2 - Poor

Favorable or Unfavorable View of the U.S.	
Brazil: % with somewhat or very favorable opinion of the U.S.:	52%
Brazil: % with somewhat or very unfavorable opinion of the U.S.:	32%
Mexico: % with somewhat or very favorable opinion of the U.S.:	64%
Mexico: % with somewhat or very unfavorable opinion of the U.S.:	25%
Britain: % with somewhat or very favorable opinion of the U.S.:	75%
Britain: % with somewhat or very unfavorable opinion of the U.S.:	16%
Germany: % with somewhat or very favorable opinion of the U.S.:	61%
Germany: % with somewhat or very unfavorable opinion of the U.S.:	35%
Russia: % with somewhat or very favorable opinion of the U.S.:	61%
Russia: % with somewhat or very unfavorable opinion of the U.S.:	33%
Poland: % with somewhat or very favorable opinion of the U.S.:	79 %
Poland: % with somewhat or very unfavorable opinion of the U.S.:	11%
South Africa: % with somewhat or very favorable opinion of the U.S.:	65%
South Africa: % with somewhat or very unfavorable opinion of the U.S.:	28%
Kenya: % with somewhat or very favorable opinion of the U.S.:	80%
Kenya: % with somewhat or very unfavorable opinion of the U.S.:	15%
India: % with somewhat or very favorable opinion of the U.S.:	54%
India: % with somewhat or very unfavorable opinion of the U.S.:	27%
Japan: % with somewhat or very favorable opinion of the U.S.:	72%
Japan: % with somewhat or very unfavorable opinion of the U.S.:	26%
South Korea: % with somewhat or very favorable opinion of the U.S.:	53%
South Korea: % with somewhat or very unfavorable opinion of the U.S.:	44%
Egypt: % with somewhat or very favorable opinion of the U.S.:	6%
Egypt: % with somewhat or very unfavorable opinion of the U.S.:	69%
Pakistan: % with somewhat or very favorable opinion of the U.S.:	10%
Pakistan: % with somewhat or very unfavorable opinion of the U.S.:	69%
Turkey: % with somewhat or very favorable opinion of the U.S.:	30%
Turkey: % with somewhat or very unfavorable opinion of the U.S.:	55%
Jordan: % with somewhat or very favorable opinion of the U.S.:	25%
Jordan: % with somewhat or very unfavorable opinion of the U.S.:	75%

I found this table on the Web site for Bill Moyers' public television show "Now". I felt that it provided important information that deserved a better form of presentation. In this case the story could be told much better in visual form.

Example #2 - Good

Current World Opinions About the U.S.A

Favorable or Somewhat Favorable | Unfavorable or Somewhat Unfavorable | Undecided or Neutral

Countries listed (top to bottom): Kenya, Poland, Britain, Japan, South Africa, Mexico, Russia, Germany, India, South Korea, Brazil, Turkey, Jordan, Pakistan, Egypt

Source: 2004 study conducted by the Pew Center.

This series of related graphs tells the story in vivid terms and brings facts to light that might not ever be noticed in the table.

Example #1

Departmental Expenses
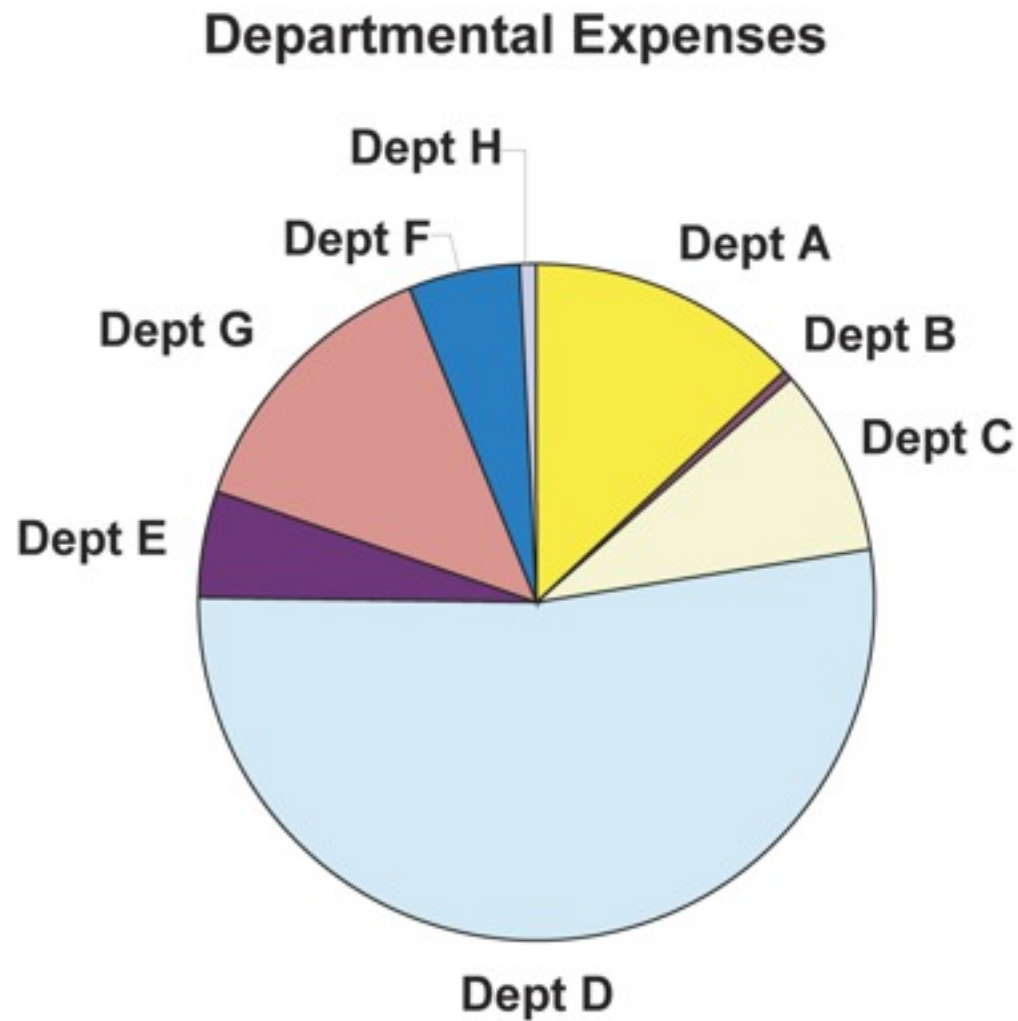
The purpose of this graph is to display how *Department G* is doing regarding expenses compared to the other departments. Is the message clear?

Often, when someone creates a graph that appears inadequate somehow, they try to fix it with sizzle, as in the next slide.
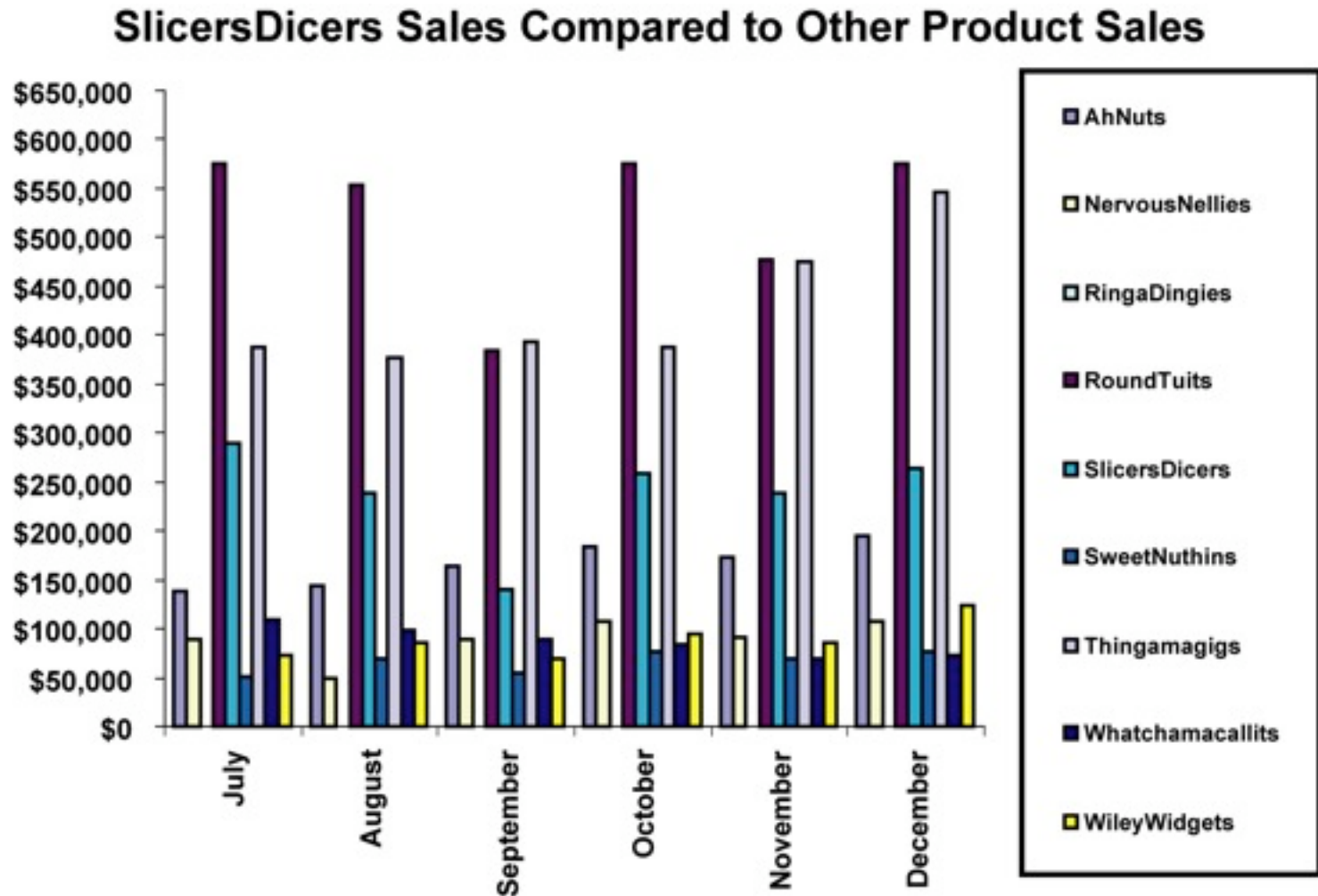
Example #1

Departmental Expenses
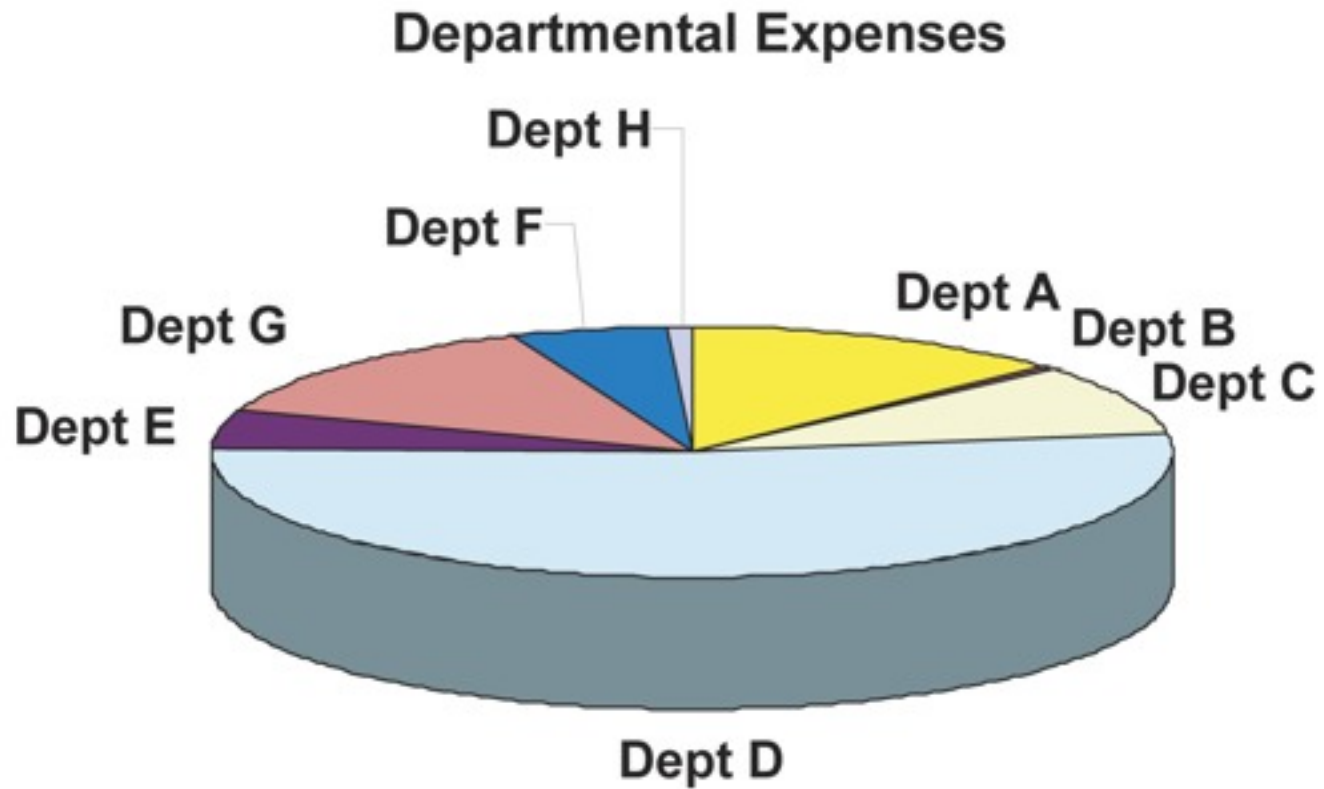
Does the addition of 3D improve this pie chart? Definitely not. In fact, it actually makes it harder to read.

Example #1 - Improved

Departmental Expenses

Department	Percentage	
Dept D	52.44%	████████████████████
Our Dept	**13.46%**	█████
Dept A	13.03%	█████
Dept C	8.85%	███
Dept F	5.37%	██
Dept E	4.95%	██
All Others	1.90%	█
	100.00%	
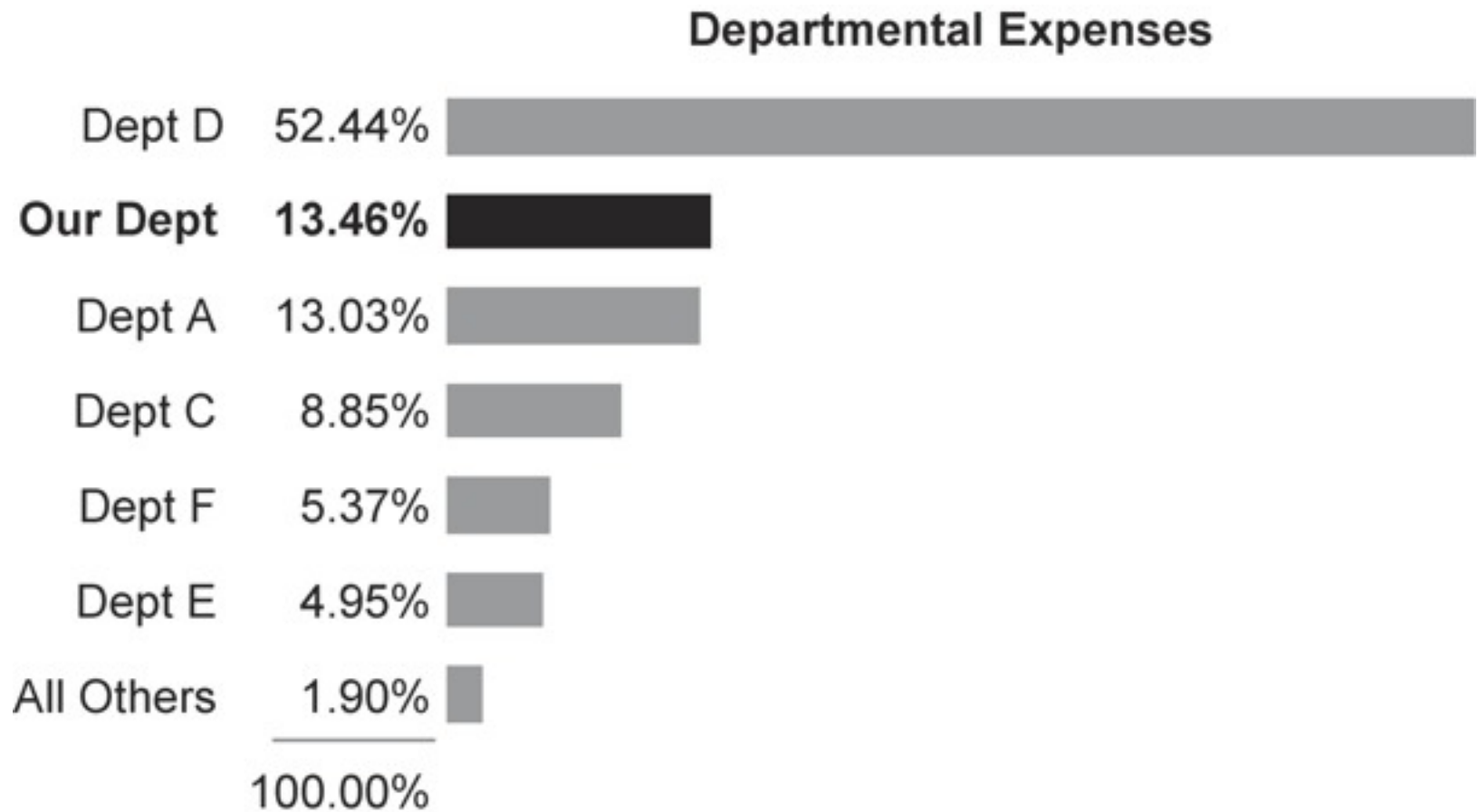
Though it lacks flash and dazzle, this simple bar graph tells the story elegantly.

Example #4 - Poor

SlicersDicers Sales Compared to Other Product Sales

Legend:
- AhNuts
- NervousNellies
- RingaDingies
- RoundTuits
- SlicersDicers
- SweetNuthins
- Thingamagigs
- Whatchamacallits
- WileyWidgets

Months: July, August, September, October, November, December

Y-axis: $0 to $650,000 in $50,000 increments

Without the title, could you determine the purpose of this graph? The design of a graph should clearly suggest its purpose.

In the general field of design, we speak of things having "affordances." These are characteristics of something's design that declare its use; a teapot has a handle and a door has a push-plate. Graphs should also be designed in a manner that clearly suggests their use.

Besides the lack of affordances, what else about this graph undermines it ability to communicate?

Example #4 - Good

Sales of SlicersDicers Compared to Sales of Other Products
July - December, 2003

(SlicersDicers' sales are displayed as black reference lines of 100%; the orange lines represent the average monthly sales for July through December.)
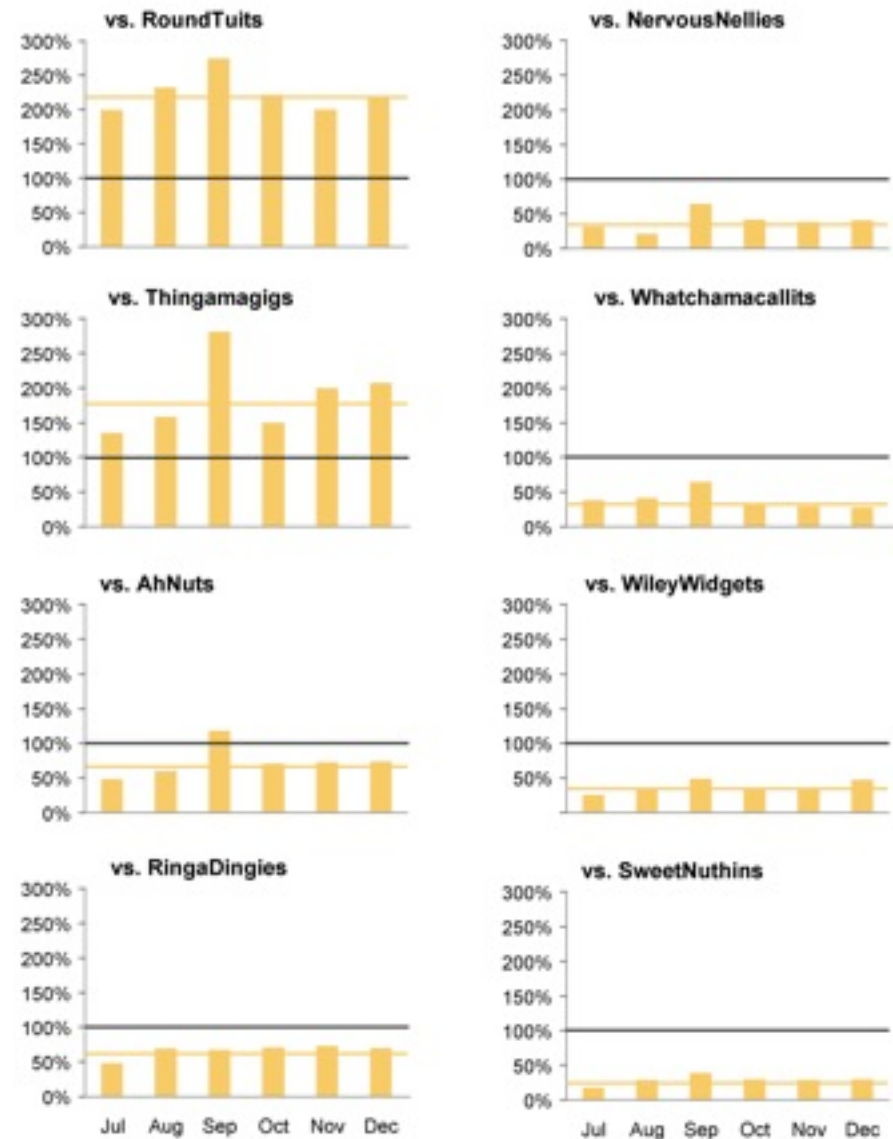
The design of this quantitative message ties clearly to its purpose. It is obvious to the reader that its intention is to compare the performance of SlicersDicers to that of the other eight products.

This solution uses a technique called "small multiples" – a series of related graphs that differ only along a single variable, in this case the various products. This technique has been known for over 20 years, but I bet you've never used software that makes this easy to do.

Warning!

Most software vendors make good design difficult.
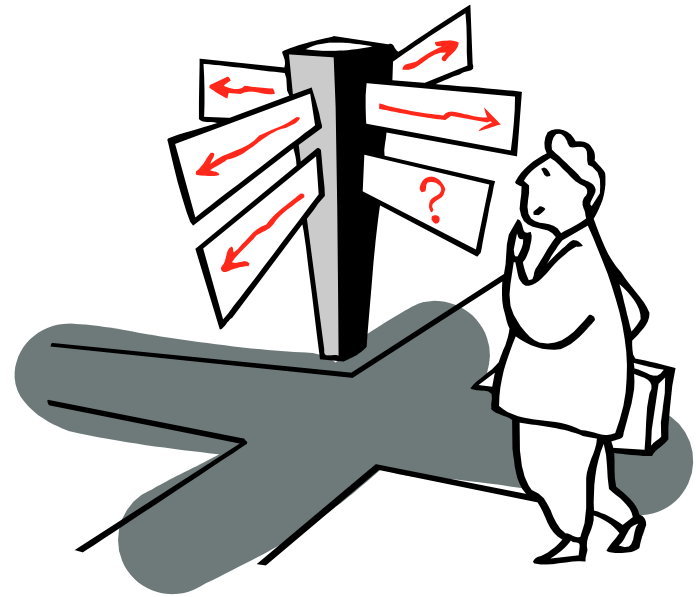
They encourage poor design by:

- providing useless features and gizmos
- providing formatting defaults that undermine a clear display of the data
- producing documentation that demonstrates poor design
- marketing flash and dazzle, rather than good design

As an independent analyst and educator who focuses on uses of data visualization for business, it is my role to be the voice of reason, constantly asking the question: "But does it work?" No matter how cool it looks or how much it dazzles the eyes, does it help you discover, make sense of, or communicate what's meaningful in the data? A great deal of what is being marketed by software vendors and even what appeals to buyers simply doesn't work. This is especially true of the big software companies, with few exceptions. The best data visualization products by and large are being offered today by relatively small software companies that have close ties to the academic research community.

Let's take a quick tour of several graph examples from the user documentation and Web sites of several software vendors to illustrate my point.

Example: Sales by company

Sales Detail by Account — Canada

Legend:
- ACME Corporation
- Eastside Chemical
- BTR Gas
- United Power
- High Tech Manufacturing
- Optical Light
- Great White Plumbing
- Art and Sound Corporation
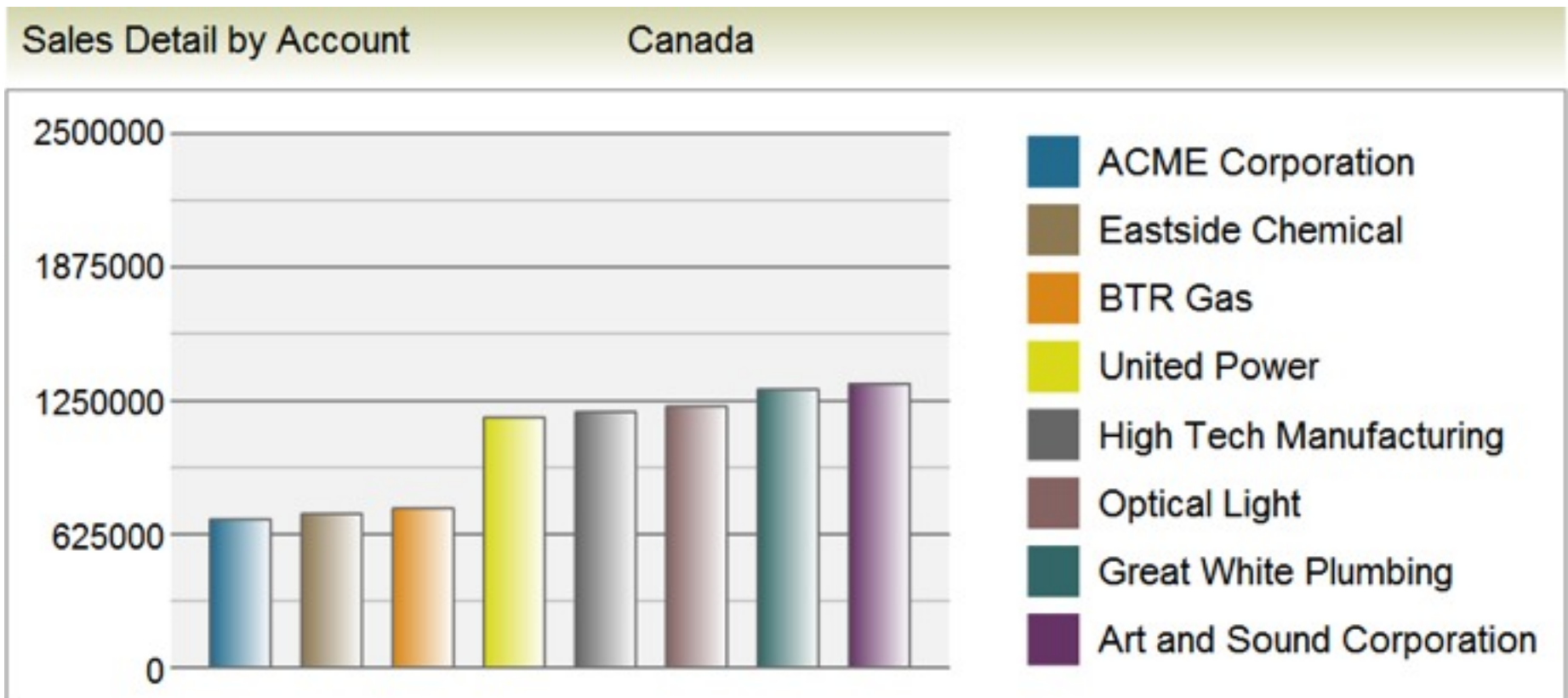
Aren't the colors pretty? Unfortunately, the lighting effects on the bars make it harder than necessary to match the colors in the legend to the corresponding bars. In fact, why would you ever need a legend to label bars when you can always label them along the axis?

(Source: Website of KearanDashboard.)

Example: Totals of something or other by year
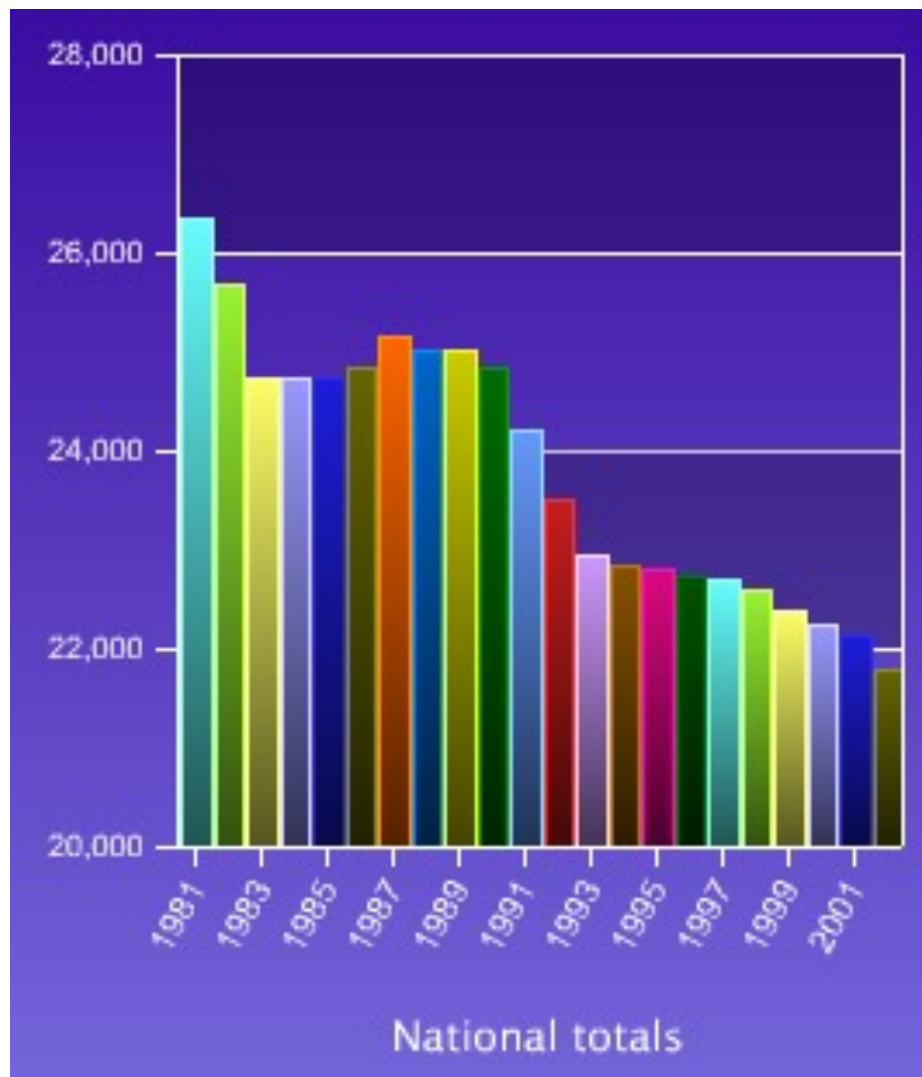
This graph gets extra points for the creative use of color – a bit too creative, don't you think? What do the different colors mean?

(Source: Website of Corda Technologies, Incorporated.)

Example: Total medals by country and type
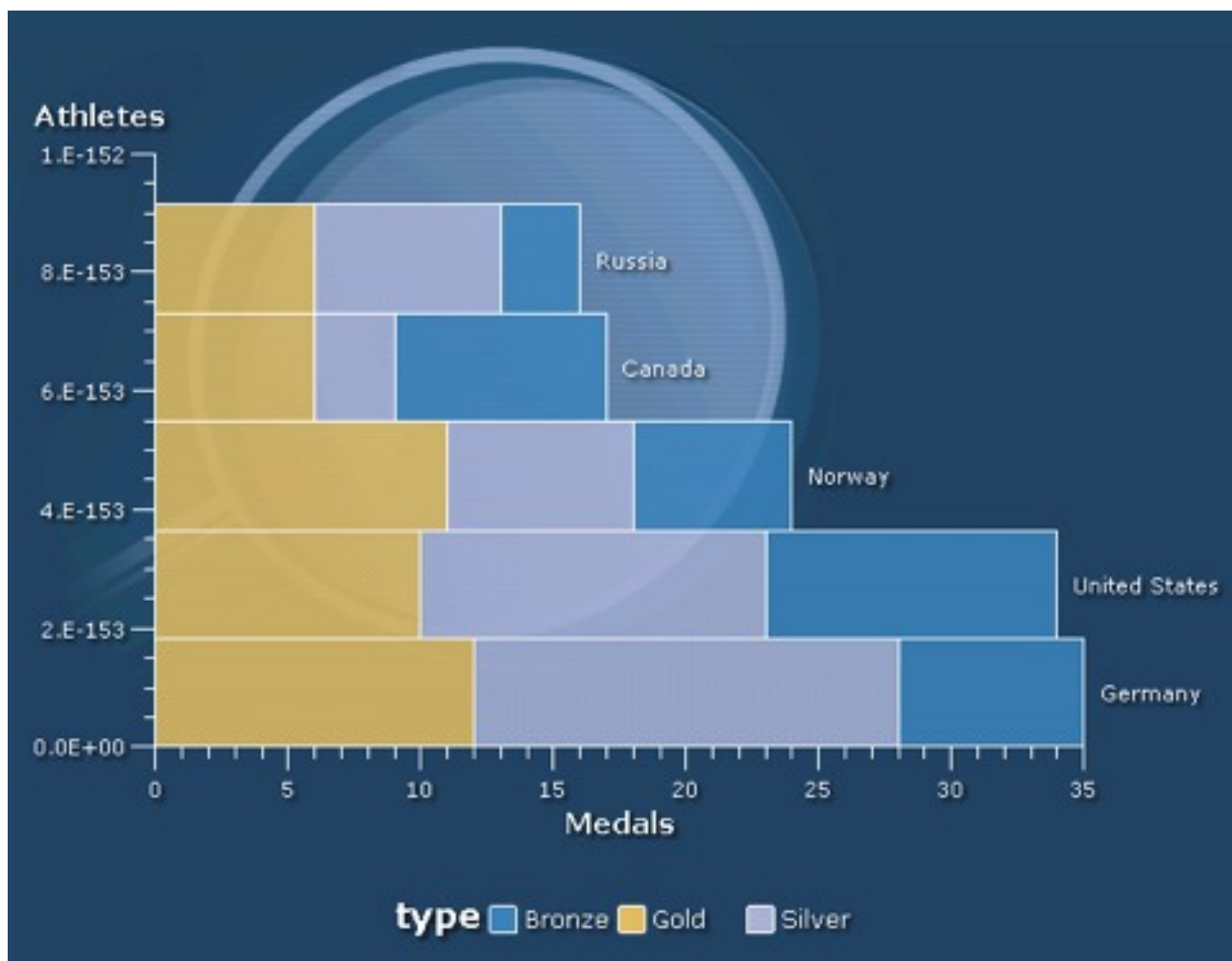
I guess the round object in the background is a medal. Even if it looked more like a medal, it would still do nothing but distract from the data itself. Can you make sense of the quantitative scale along the vertical axis?

(Source: Website of SAS Institute Inc.)

Example: New sales opportunities by lead quality

Legend:
- 1-Excellent
- 2-Very High
- 3-High
- 4-Fair
- 5-Poor
- Field Sales: 2-Very High
- Unspecified

Pie chart values: 13.4%, 15.6%, 0.1%, 17.3%, 14.7%, 17.5%, 21.5%
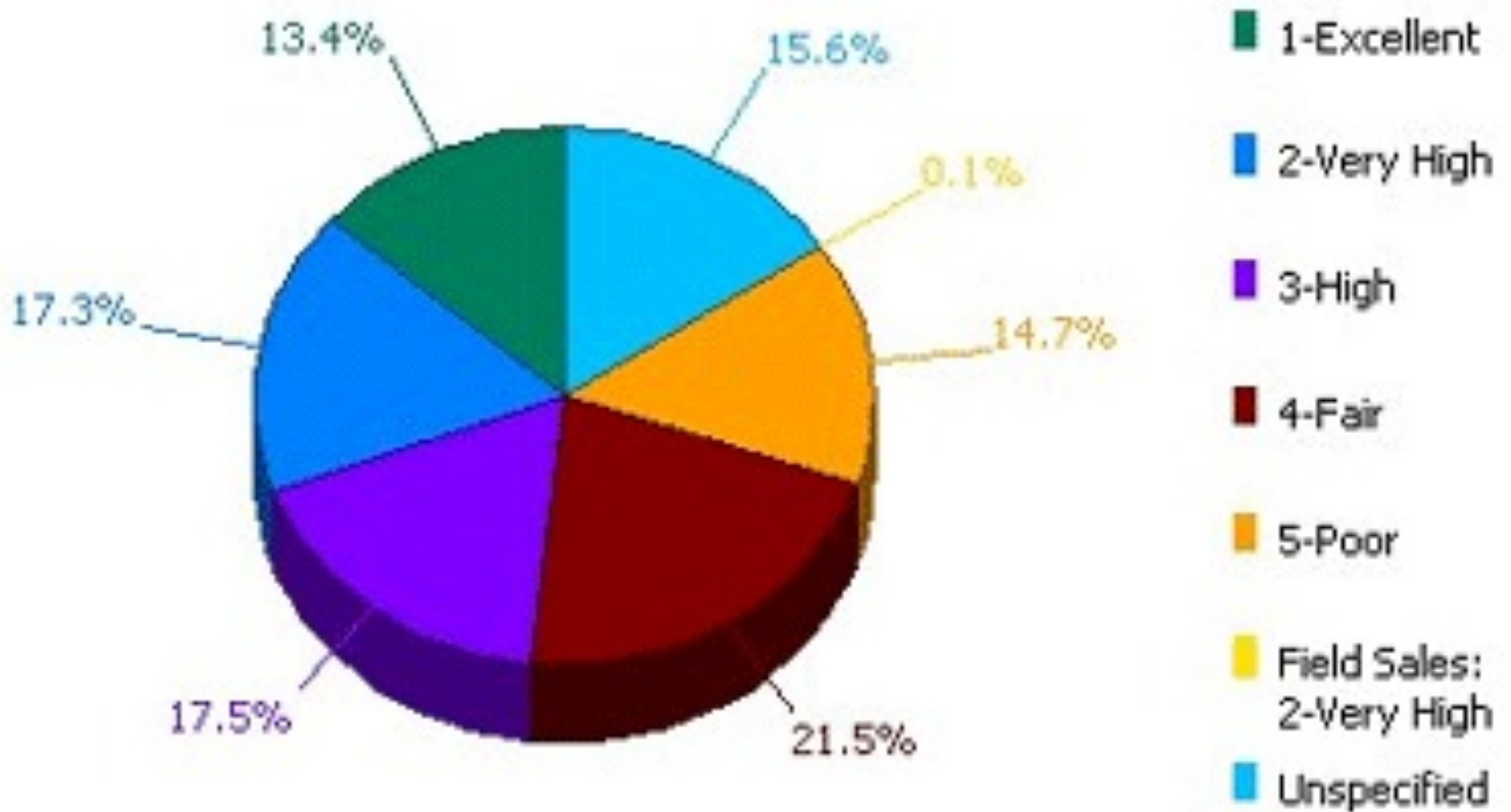
Notice the effort that is involved in shifting your focus back and forth between the pie chart and the legend to determine what each slice represents, especially given the fact that the order of the items in the legend does not match the order in the pie. Also notice how slices that are different in value often appear to be the same size.

(Source: Website of Siebel Systems, which has since been acquired by Oracle Corporation.)

Example: Revenue by service line

$2,218,124.00

$513,570.00

$554,830.00

Revenue

Service Line

Accommodation

Food & Drinks

Recreation
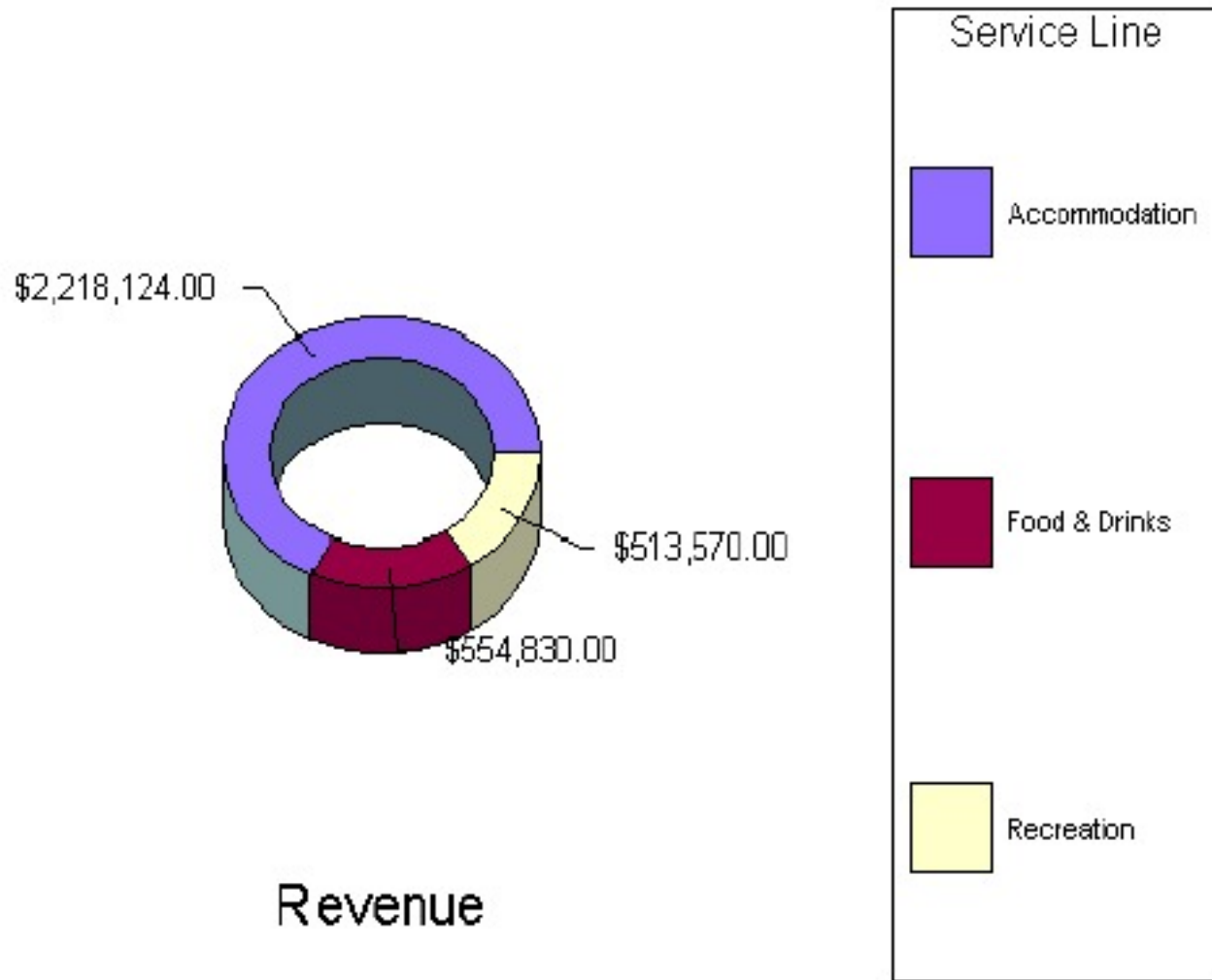
Even turning a pie into a donut doesn't make it any more palatable. A donut chart is just a pie with a hole in it.

(Source: User documentation of Business Objects.)

Example: Revenue by resort and month
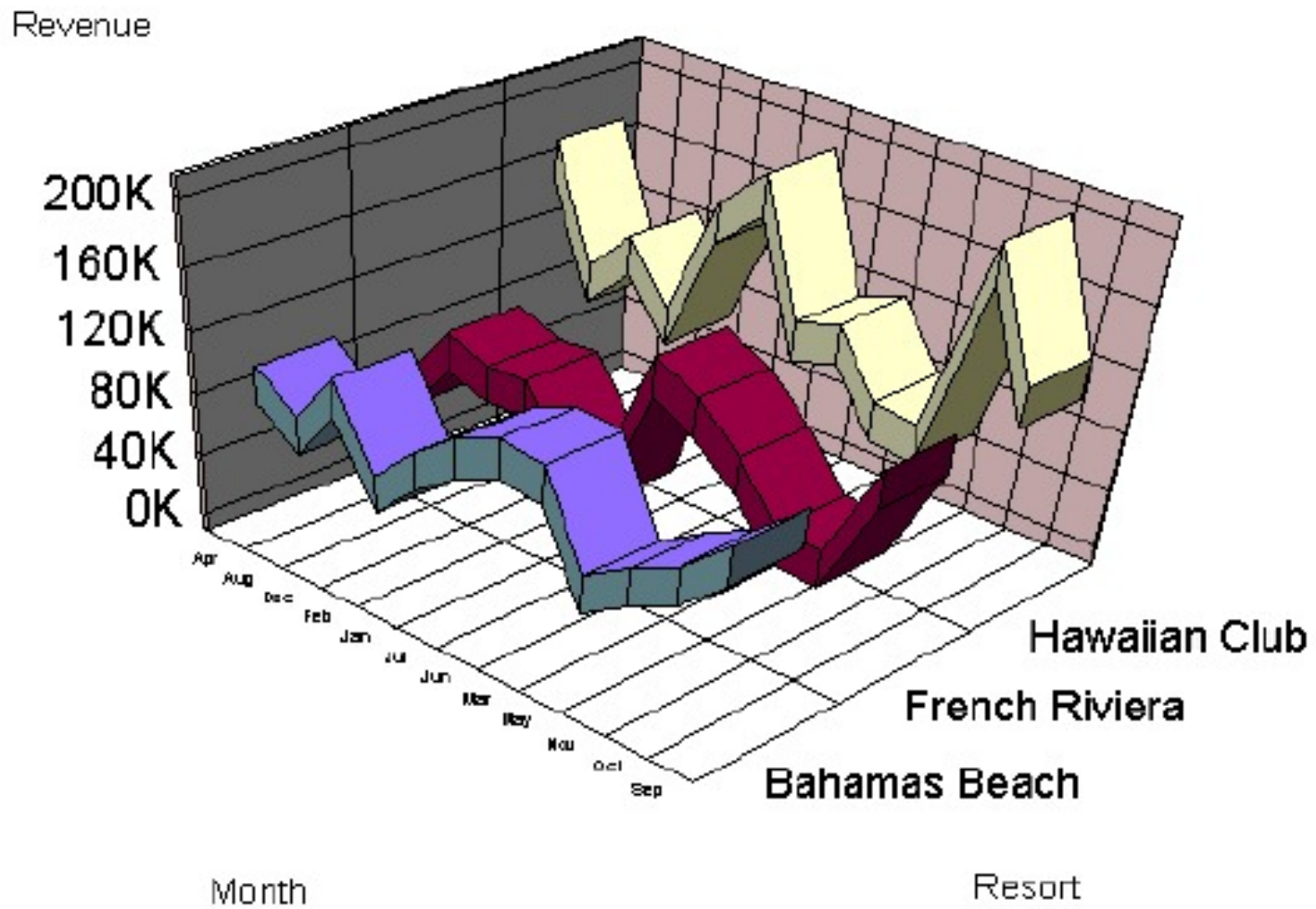
But 3-D lines are the height of fashion. And time trends with the months sorted in alphabetical rather than chronological order are so much more creative.

(Source: User documentation of Business Objects.)

Example: Revenue by sales channel and quarter
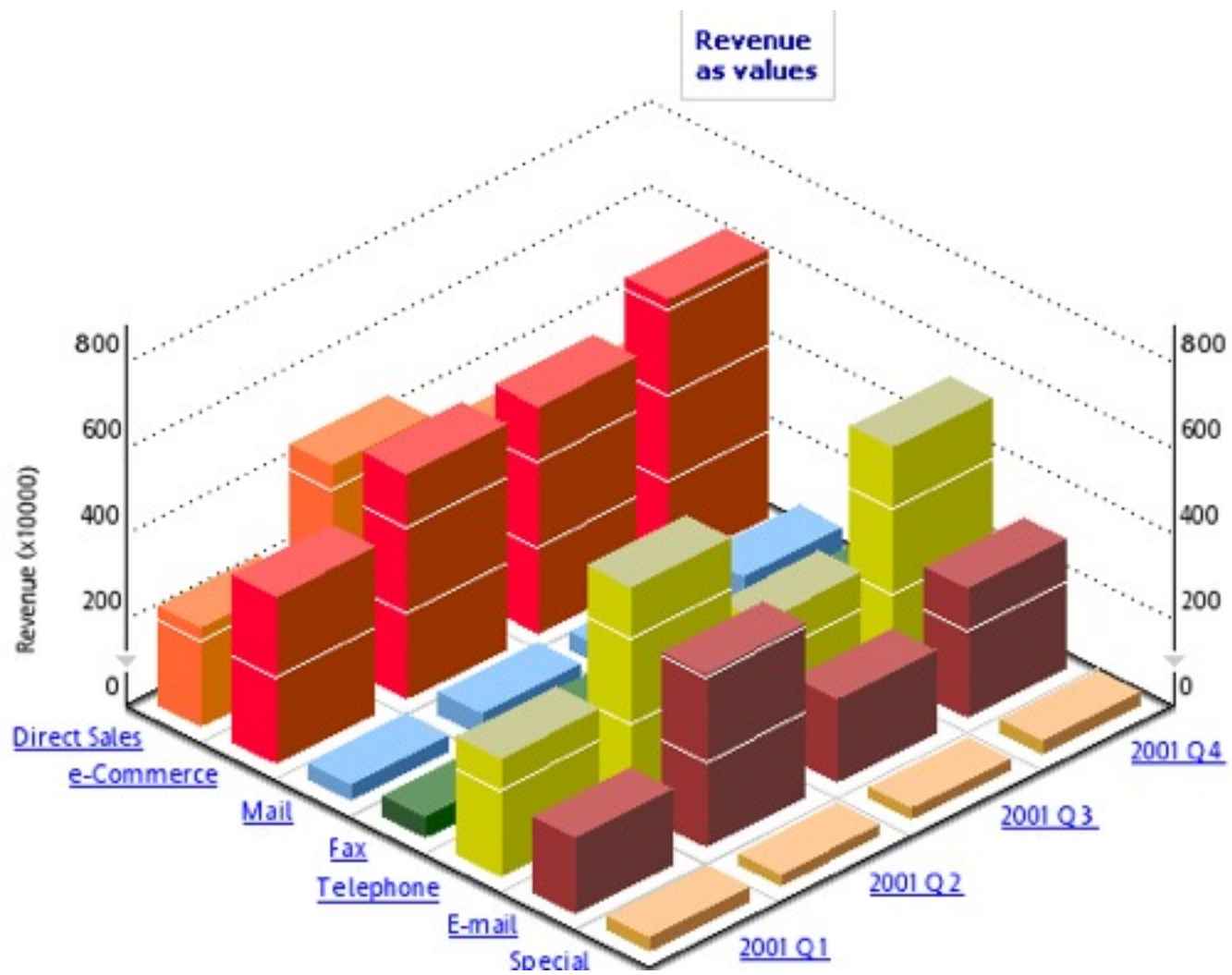
No matter how bright the bars, you can't see them if they're hidden behind others. Can you determine fax revenue for Q3 or direct sales revenue for Q4? This problem, when something is hidden behind something else, is called *occlusion*.

(Source: Website of Cognos Incorporated.)

Example: Revenue by product and quarter
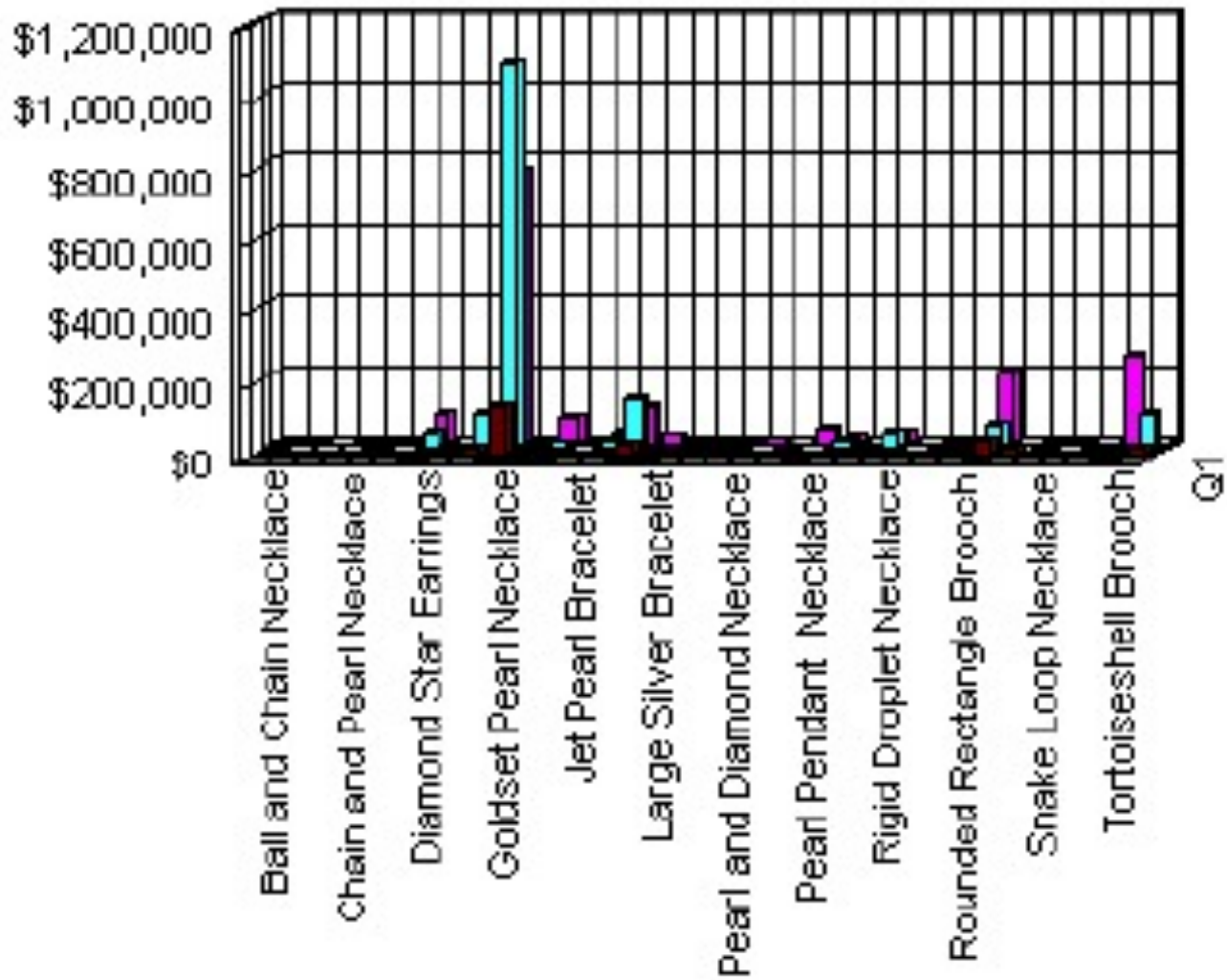
Most of the bars in this graph are so short, they're barely visible, and impossible to interpret. Notice that this graph contains four quarters worth of data, but the sole label of "Q1" suggests otherwise. And what do you think of the dark grid lines? They make this graph look a little like a prison cell from which the numbers will never escape!

(Source: User documentation of Business Objects.)

Example: Expenses by GL account category and account
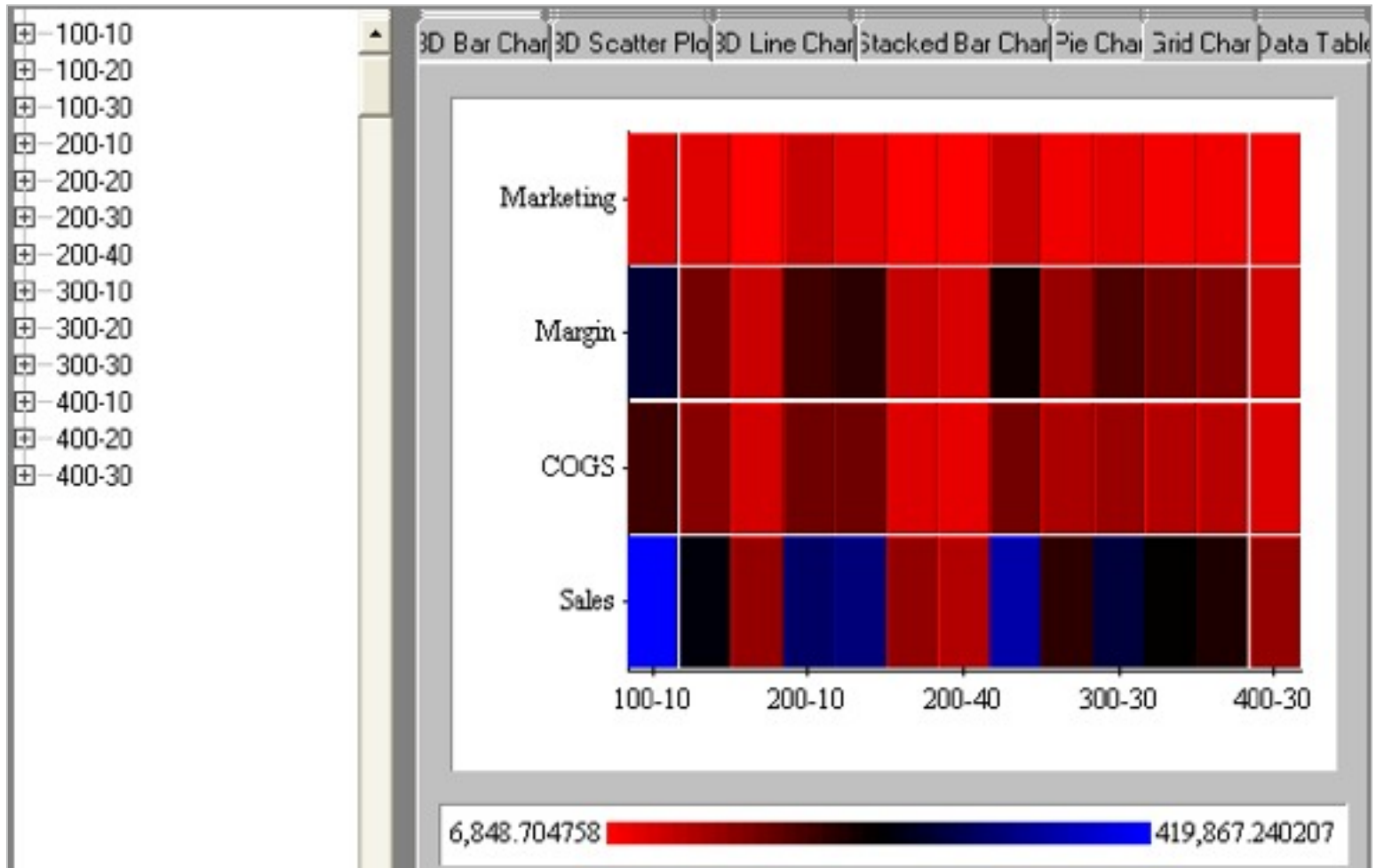
I call this a "Scottish Graph," because it looks a lot like the tartan print of a kilt. This is just plain absurd. Does this vendor really believe that people can interpret numbers as a continuum of color ranging from red through black to blue? The quantitative key at the bottom is a hoot. It includes six decimal places of precision for what must be dollars, but you'd be lucky to interpret any of the numbers in the graph within $10,000 of the actual value.

(Source: Website of Visualize, Inc.)

Example: Hotel revenue by service and quarter
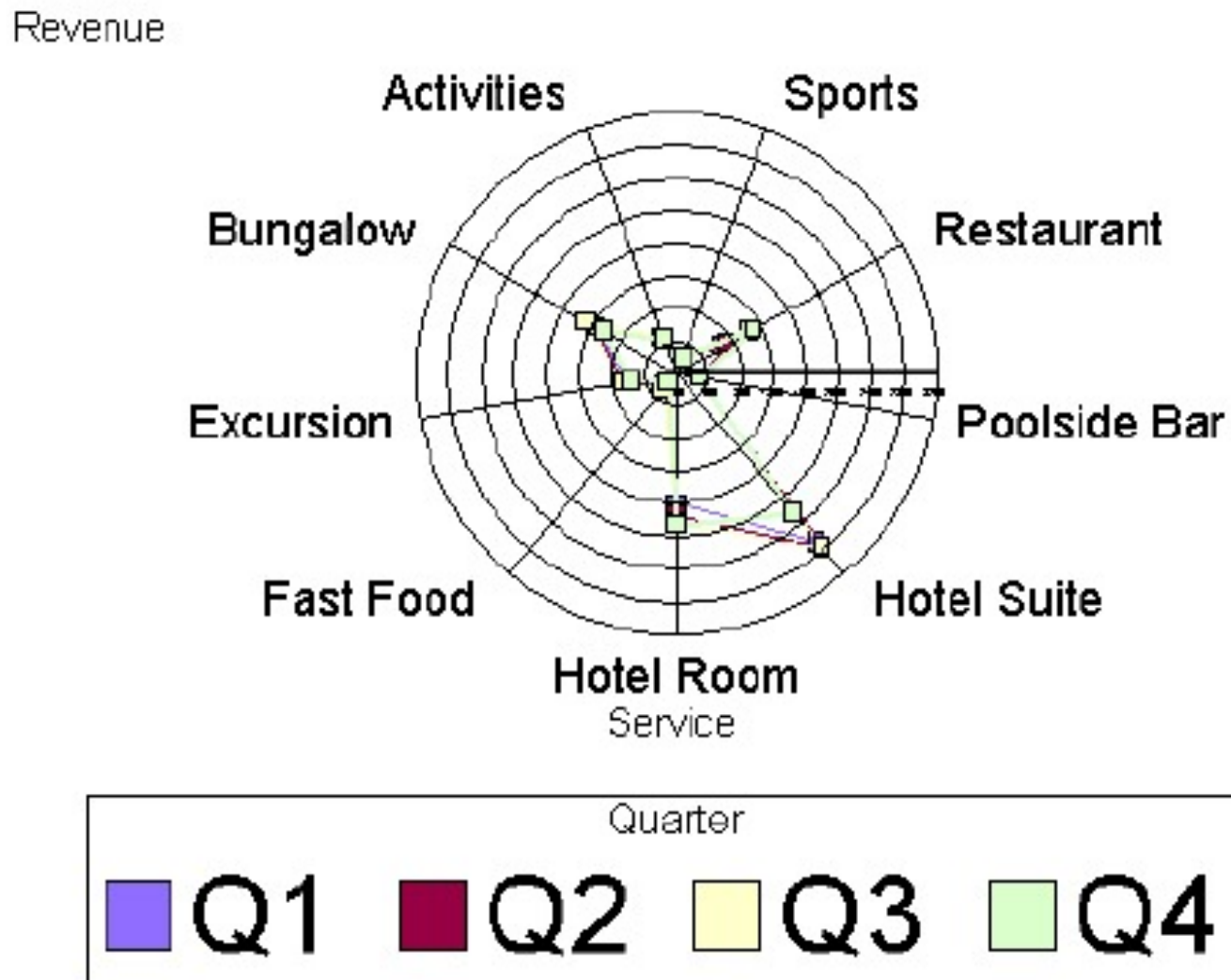
Good thing the quarters are labeled so largely in the legend, otherwise I'd never be able to make sense of this radar graph.

(Source: User documentation of Business Objects.)

Example: Revenue vs. costs per month

Costs vs. Revenue

January February March April May

☐ Revenues ☐ Operating Costs
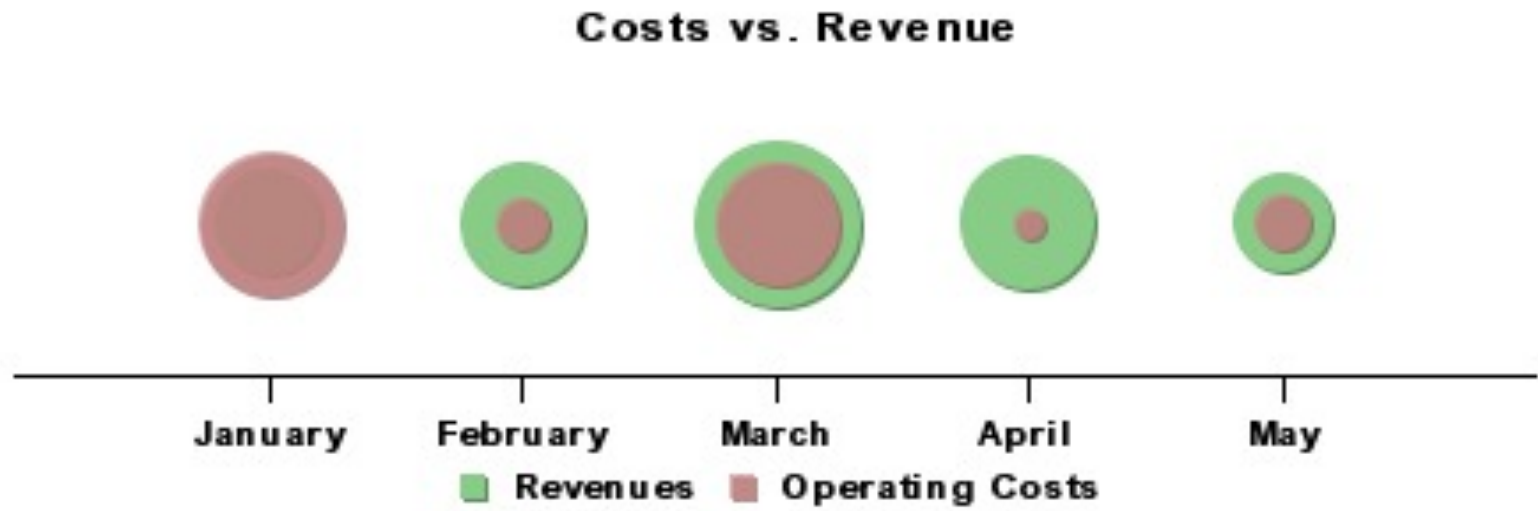
Circles within and behind circles. Pretty! Pretty silly that is.

(Source: Website of Visual Mining, Inc.)

Example: Your guess is as good as mine
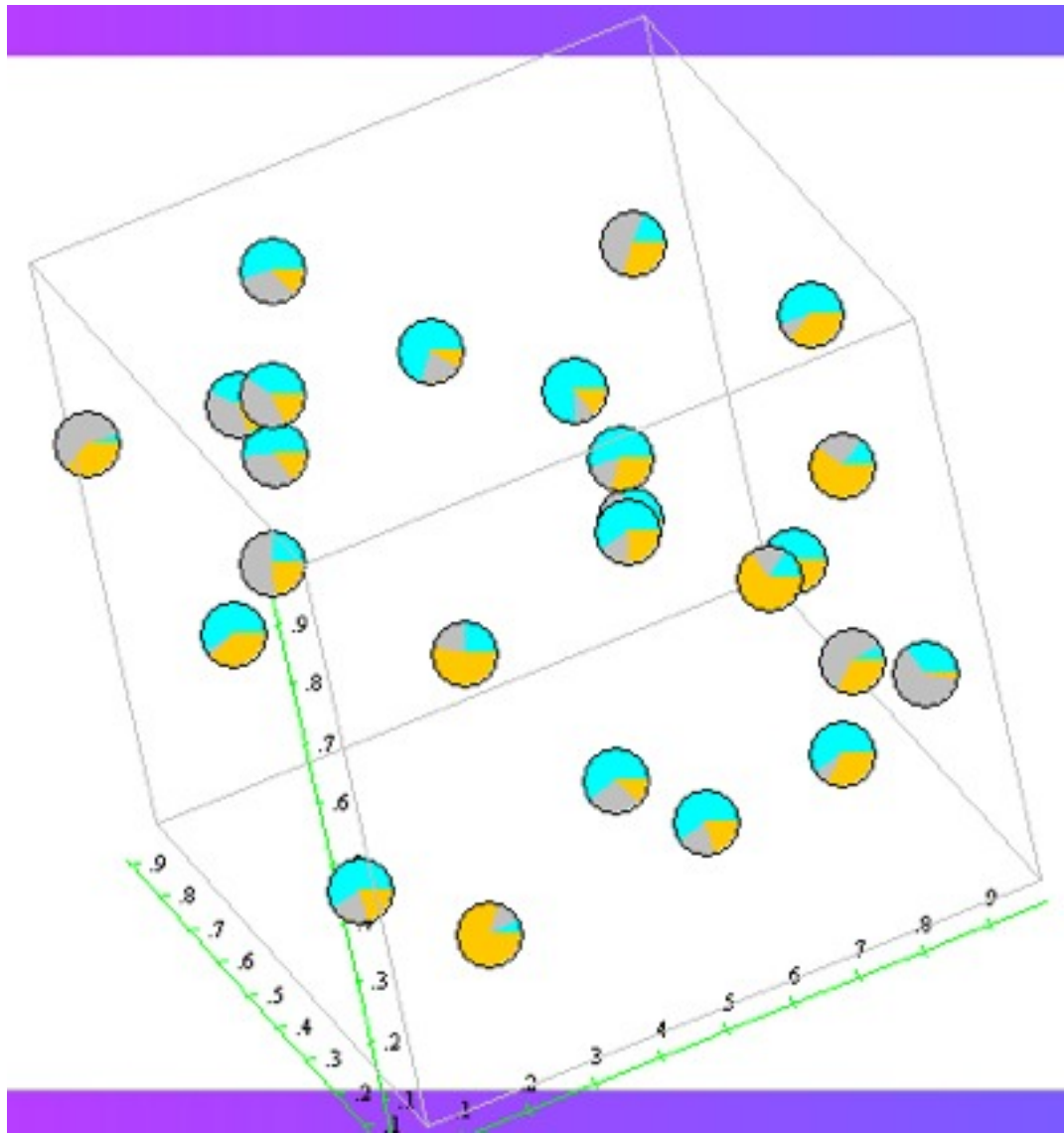
Pies in 3-D space. Awesome!

(Source: Website of Visualize, Inc.)

Tables and graphs help us…

Think

and

Communicate

"Above all else show the data."

Edward Tufte

This Edward R. Tufte quote is from his milestone work, *The Visual Display of Quantitative Information*, published by Graphics Press in 1983.

In tables and graphs:

- The message is in the data.
- The medium of communication, especially for graphs, is visual.
- To communicate the data effectively, you must understand visual perception – what works, what doesn't, and why.

Grice's conversational 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 conversational maxims, which attempt the describe the characteristics of polite conversation.

Every one of these maxims of conversation apply equally well to the communication of quantitative information in the workplace. We'll strive in this workshop to translate these maxims into effective and polite communication via tables and graphs.

Grice's conversational 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 conversational 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 conversational maxims: Relevance

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

Grice's conversational maxims: Manner

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

Fundamental challenges of data presentation

1. Determining the medium that tells the story best

2. Designing the visual components to tell the story clearly

Either

Product	Units Sold	Revenue
Food	34,837	746,383
Beverage	42,374	845,984
Total	77,211	1,592,367

or
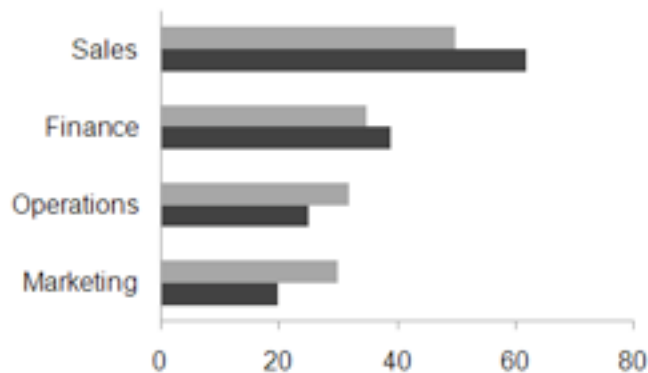
and
which kind?

From

to

Revenue
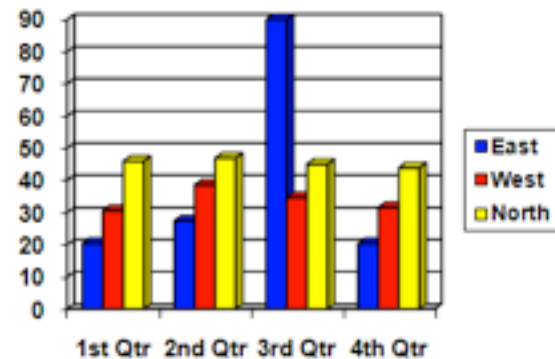
1. You begin by determining the best medium for your data and the message you wish to emphasize. Does it require a table or a graph? Which kind of table or graph?

2. Once you've decided, you must then design the individual components of that display to present the data and your message as clearly and efficiently as possible.

The solutions to both of these challenges are rooted in an understanding of visual perception.

Match the display to your message.

Indirect Portion of Domestic Sales is Declining

Indirect Portion of International Sales is Holding Steady

Legend: ■ Direct ■ Indirect

Indirect Sales Percentage of Total Sales

International holding steady

Domestic declining
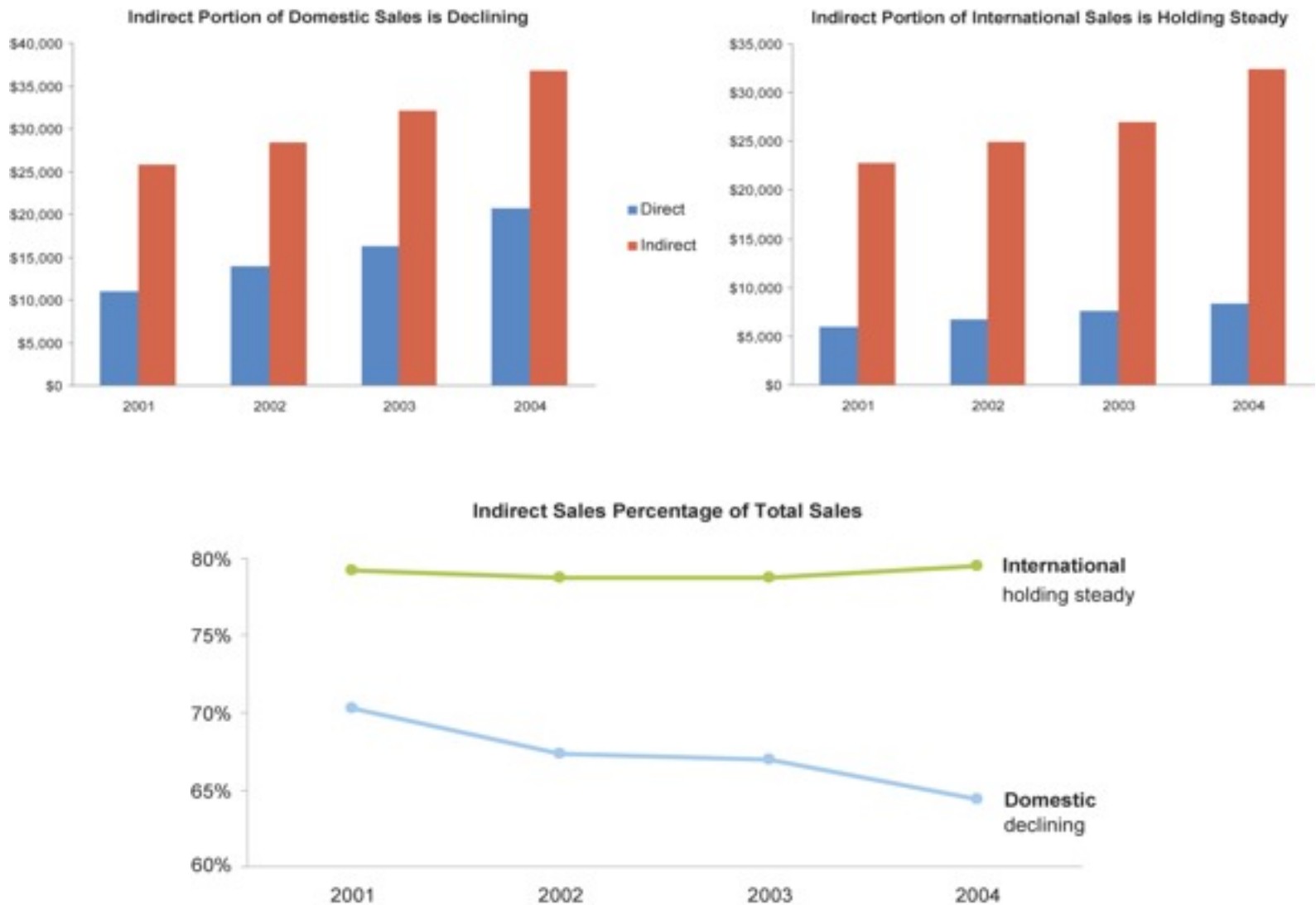
In the top example, the message contained in the titles is not clearly displayed in the graphs. The message deals with the ratio of indirect to total sales – how it is declining domestically, while holding steady internationally. You'd have to work hard to get this message the display as it is currently designed.

The bottom example, however, is designed very specifically to display the intended message. Because this graph is skillfully designed to communicate, its message is crystal clear. A key feature that makes this so is the choice of percentage for the quantitative scale, rather than dollars.

Eeney, meeney, miney, moe – what type of display should I use this time?

Employment Candidate Review

or
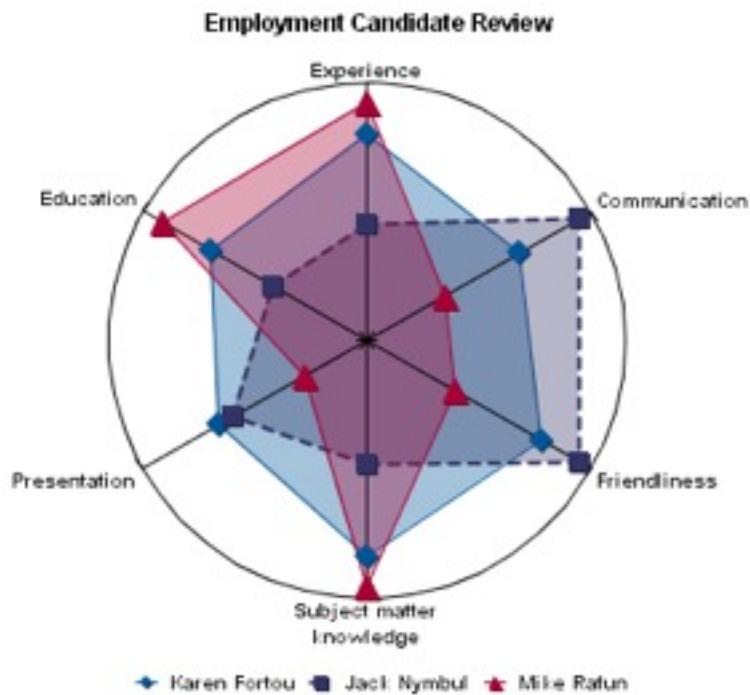
Employment Candidate Review

	Candidates		
Rating Areas	Karen Fortou	Mike Rafun	Jack Nymbul
Experience	4.00	4.50	2.50
Communication	3.50	2.00	5.00
Friendliness	4.00	2.00	4.50
Subject matter knowledge	4.00	5.00	2.50
Presentation	3.00	1.50	2.75
Education	3.50	4.50	2.00
Average Rating	3.67	3.25	3.21

If you were the person responsible for creating a means to display the relative merits of candidates for employment, you would have to choose the best visual means to communicate this information. You shouldn't just choose any old method, and especially not a particular one because it looks the snazziest. A graph isn't always the best way to get your message across. In this case a simple table would do the job better.

(Source: the radar graph on the left was found on the Website of Visual Mining, Inc.)

Tables work best when…

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Annual
1990	127.4	128.0	128.7	128.9	129.2	129.9	130.4	131.6	132.7	133.5	133.8	133.8	130.7
1991	134.6	134.8	135.0	135.2	135.6	136.0	136.2	136.6	137.2	137.4	137.8	137.9	136.2
1992	138.1	138.6	139.3	139.5	139.7	140.2	140.5	140.9	141.3	141.8	142.0	141.9	140.3
1993	142.6	143.1	143.6	144.0	144.2	144.4	144.4	144.8	145.1	145.7	145.8	145.8	144.5
1994	146.2	146.7	147.2	147.4	147.5	148.0	148.4	149.0	149.4	149.5	149.7	149.7	148.2
1995	150.3	150.9	151.4	151.9	152.2	152.5	152.5	152.9	153.2	153.7	153.6	153.5	152.4
1996	154.4	154.9	155.7	156.3	156.6	156.7	157.0	157.3	157.8	158.3	158.6	158.6	156.9
1997	159.1	159.6	160.0	160.2	160.1	160.3	160.5	160.8	161.2	161.6	161.5	161.3	160.5
1998	161.6	161.9	162.2	162.5	162.8	163.0	163.2	163.4	163.6	164.0	164.0	163.9	163.0
1999	164.3	164.5	165.0	166.2	166.2	166.2	166.7	167.1	167.9	168.2	168.3	168.3	166.6
2000	168.8	169.8	171.2	171.3	171.5	172.4	172.8	172.8	173.7	174.0	174.1	174.0	172.2
2001	175.1	175.8	176.2	176.9	177.7	178.0	177.5	177.5	178.3	177.7	177.4	176.7	177.1
2002	177.1	177.8	178.8	179.8	179.8	179.9	180.1	180.7	181.0	181.3	181.3	180.9	179.9

- Used to look up individual values
- Used to compare individual values
- Data must be precise
- You must include multiple units of measure
- You wish to show both details and their sums

What do graphs do well?

Feature patterns, trends, and exceptions.

The saying, "A picture is worth a thousand words," applies quite literally to quantitative graphs. By displaying quantitative information in visual form, graphs efficiently reveal information that would otherwise require a thousand words or more to adequately describe.

In this example of purely manufactured data to illustrate my point…

Joseph Berkson once stated what happens quite powerfully: When we visualize the data effectively and suddenly, we experience "interocular traumatic impact"—a conclusion that hits us between the eyes.

Graphs bring patterns to light.

What's the message?

Job Satisfaction By Income, Education & Age

Income	College Degrees		No College Degrees	
	Under 50	50 & over	Under 50	50 & over
Up to $50,000	643	793	590	724
Over $50,000	735	928	863	662

Job Satisfaction Related to Education and Salary

Job Satisfaction

College degree, over $50,000
College degree, up to $50,000
No college degree, up to $50,000
No college degree, over $50,000

Age — Under 50, 50 & Over
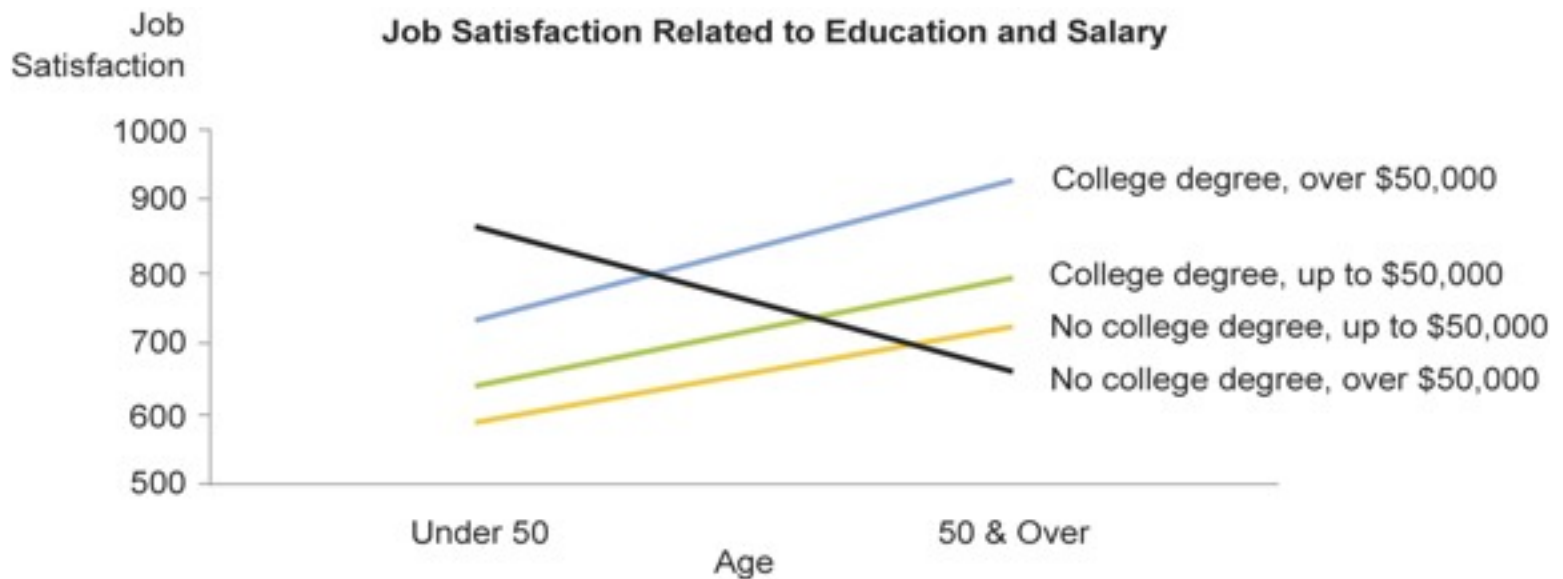
The fact that job satisfaction for employees without a college degree decreases significantly in their later years doesn't jump out at you when you examine the table, but it is immediately obvious when you examine the graph.

…but only if you design them correctly.

Job Satisfaction Related to Education and Salary (line graph)
- Job Satisfaction axis: 500–1000
- Age axis: Under 50, 50 & Over
- College degree, over $50,000
- College degree, up to $50,000
- No college degree, up to $50,000
- No college degree, over $50,000

Job Satisfaction Related to Education and Salary (bar graph)
- College degree, up to $50,000
- College degree, over $50,000
- No college degree, up to $50,000
- No college degree, over $50,000
- Job Satisfaction axis: 0–1000
- Age: Under 50, 50 & Over

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.

Poor graphs can be transformed.
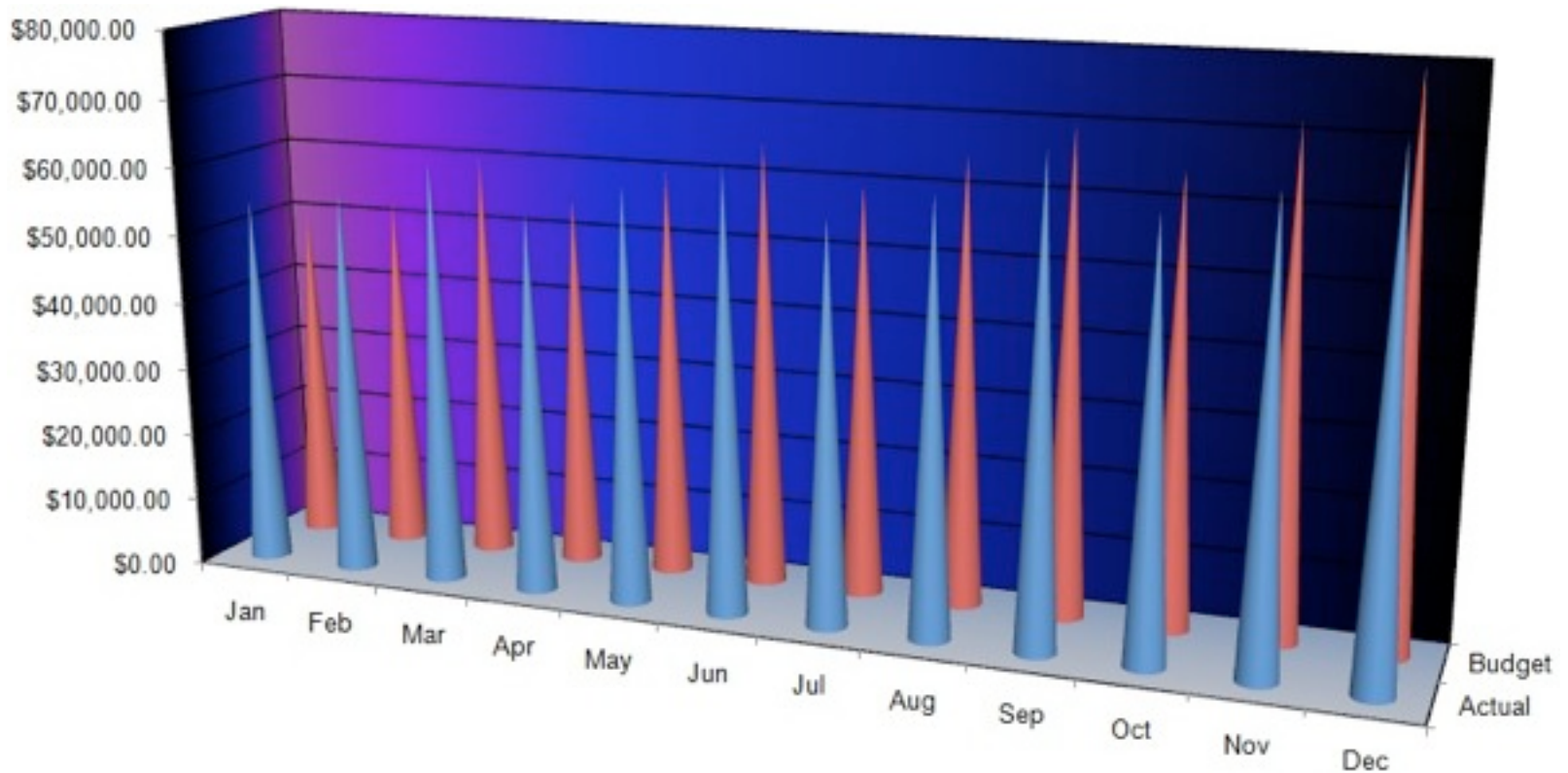
This is the kind of graph that software products, including Excel, encourage us to create. They give us an infinite selection of poorly-designed graphs from which to choose. What we really need, however, is a small selection of graphs that really work.

Using this graph, try to see the pattern of change across the months in actual expenses. Try to determine one of the actual values. Try to compare actual expenses to the budget across time.

Let's transform this graph into one that communicates.

Poor graphs can be transformed.
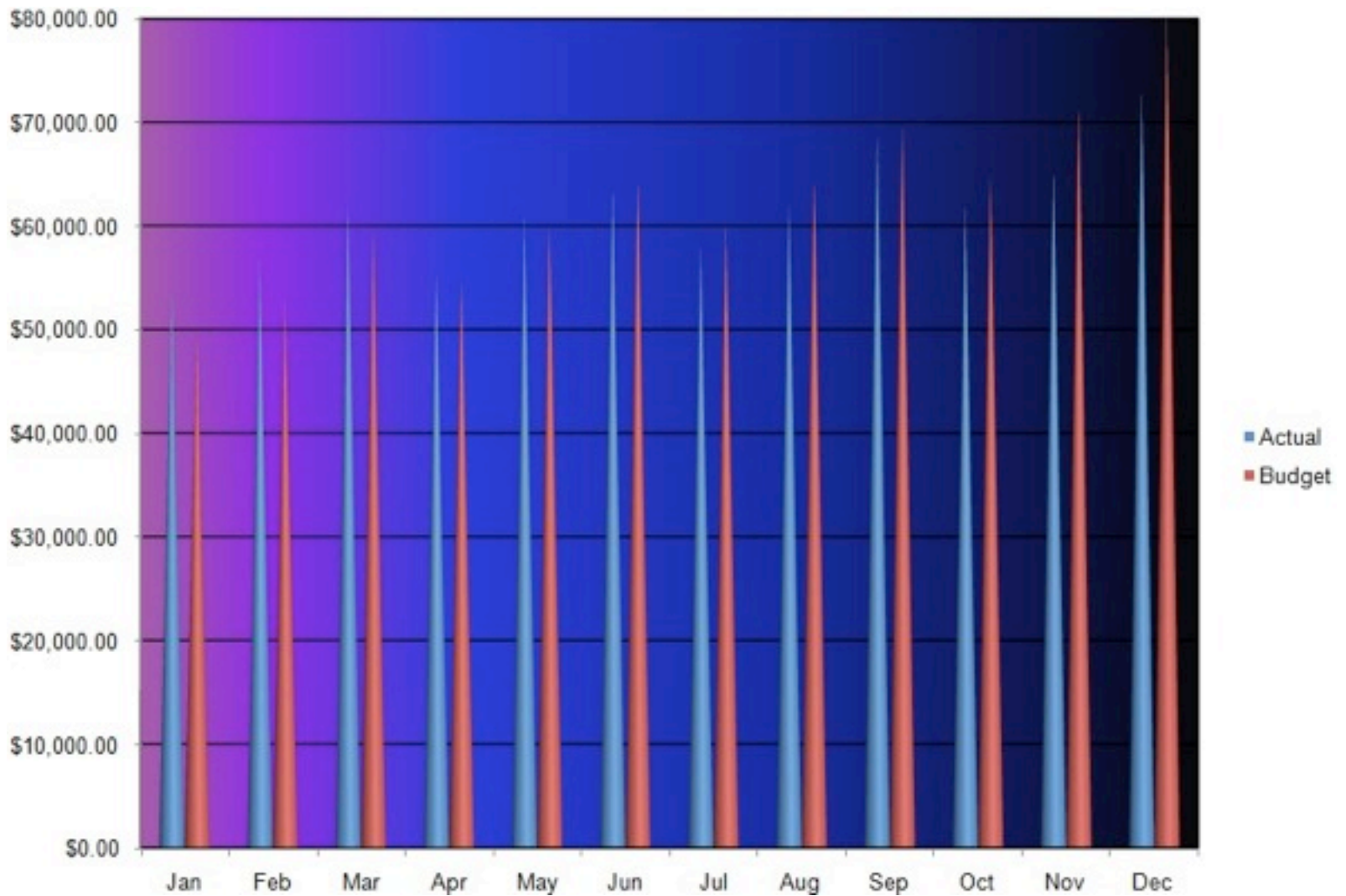
We have now removed the useless 3-D effects and angle, which makes the data easier to read.

Poor graphs can be transformed.
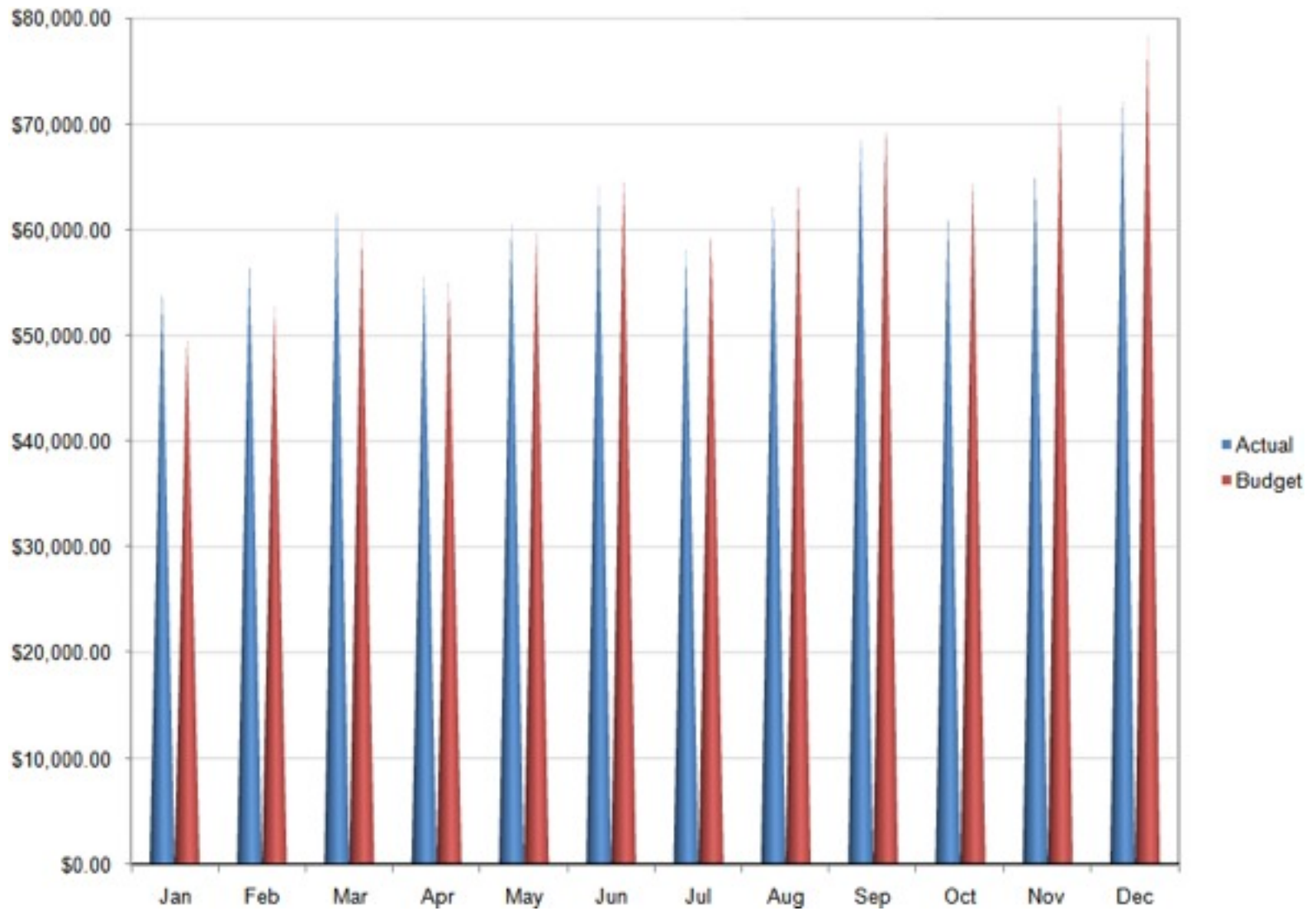
We have now removed the background fill color.

Poor graphs can be transformed.
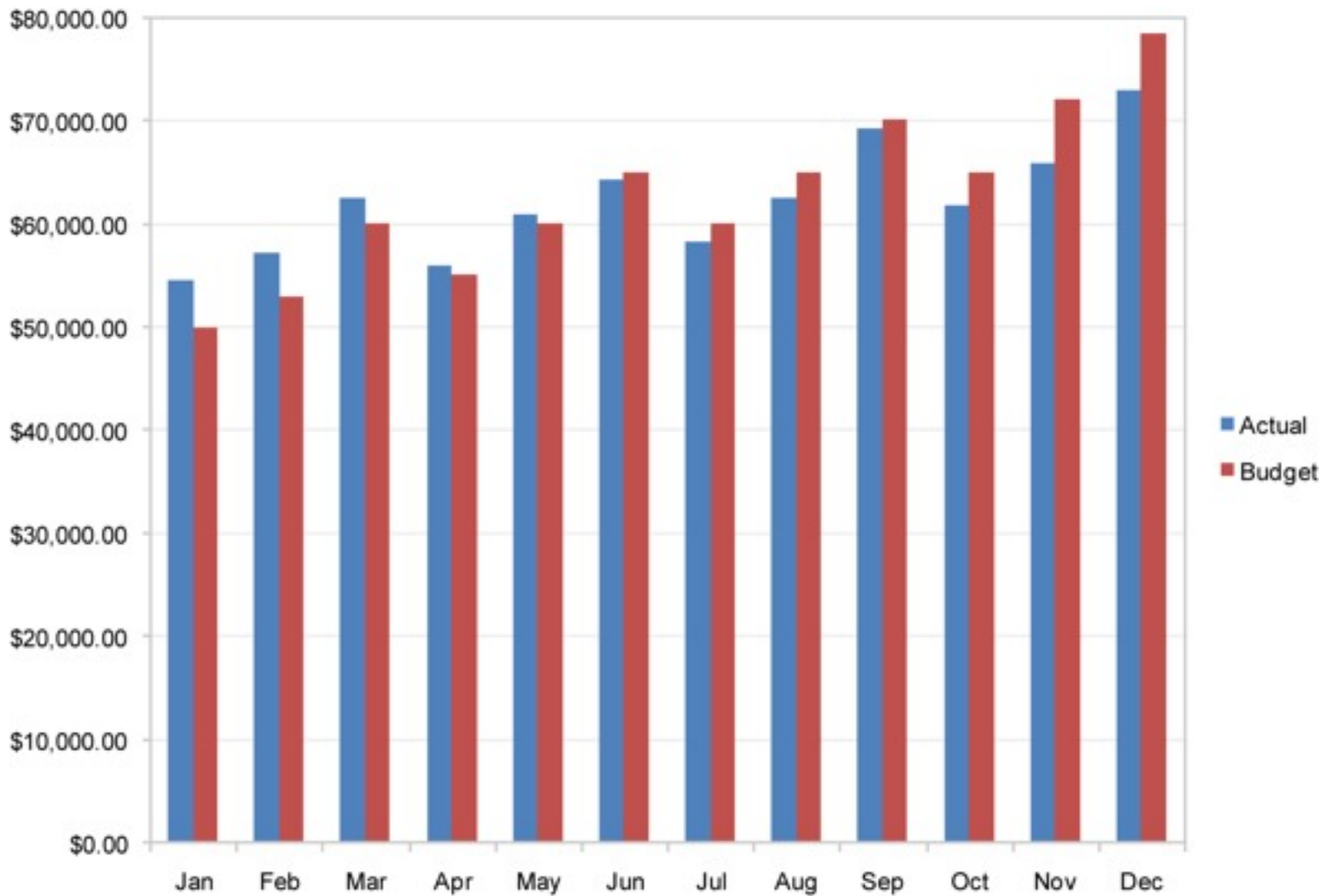
We have now replaced the silly cones with regular bars.

Poor graphs can be transformed.
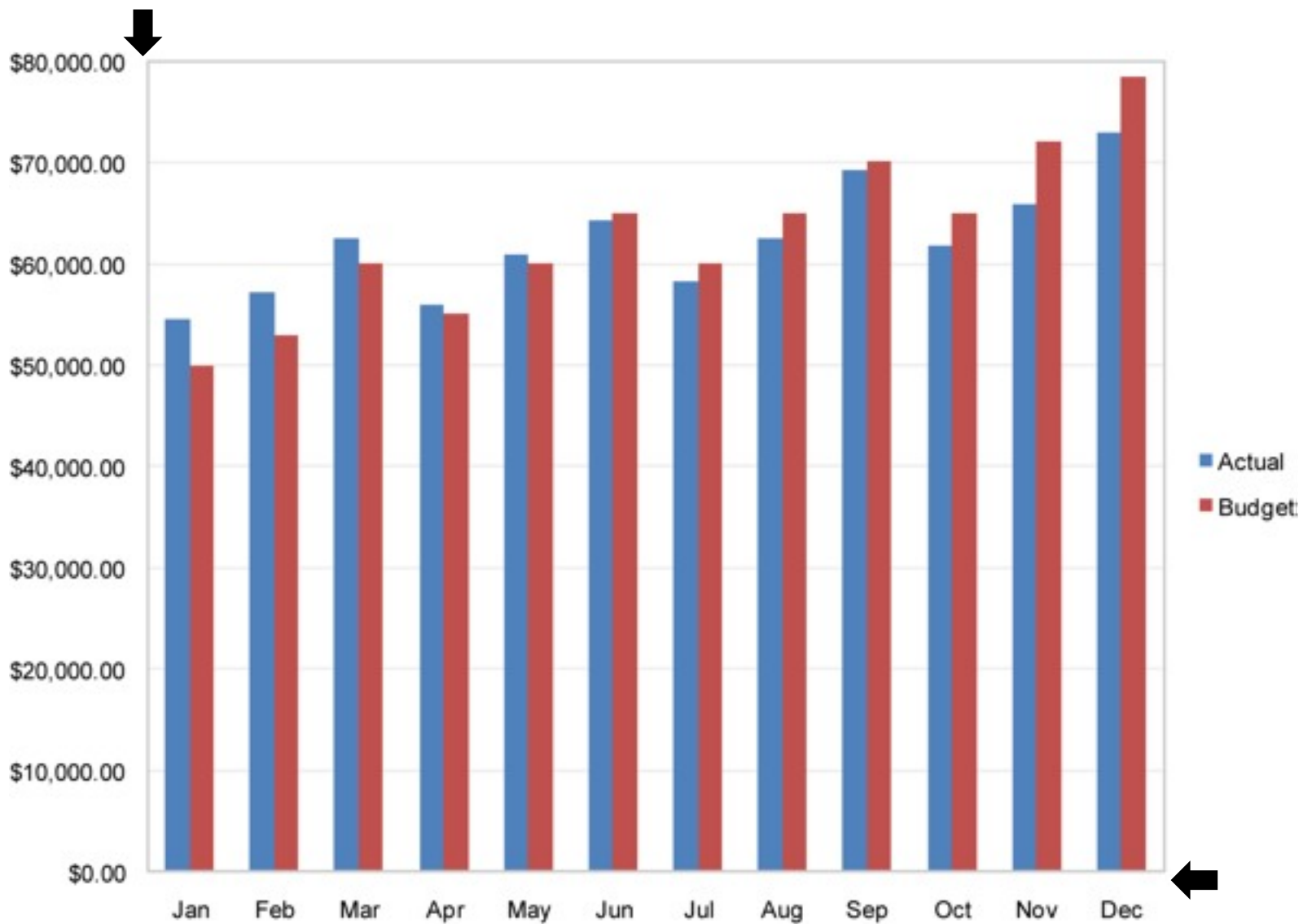
We have now removed the tick marks, which aren't necessary. Tick marks are not needed to separate the months along the X-axis and because horizontal grid lines are being displayed, there is no need for tick marks on the Y-axis either.

Poor graphs can be transformed.
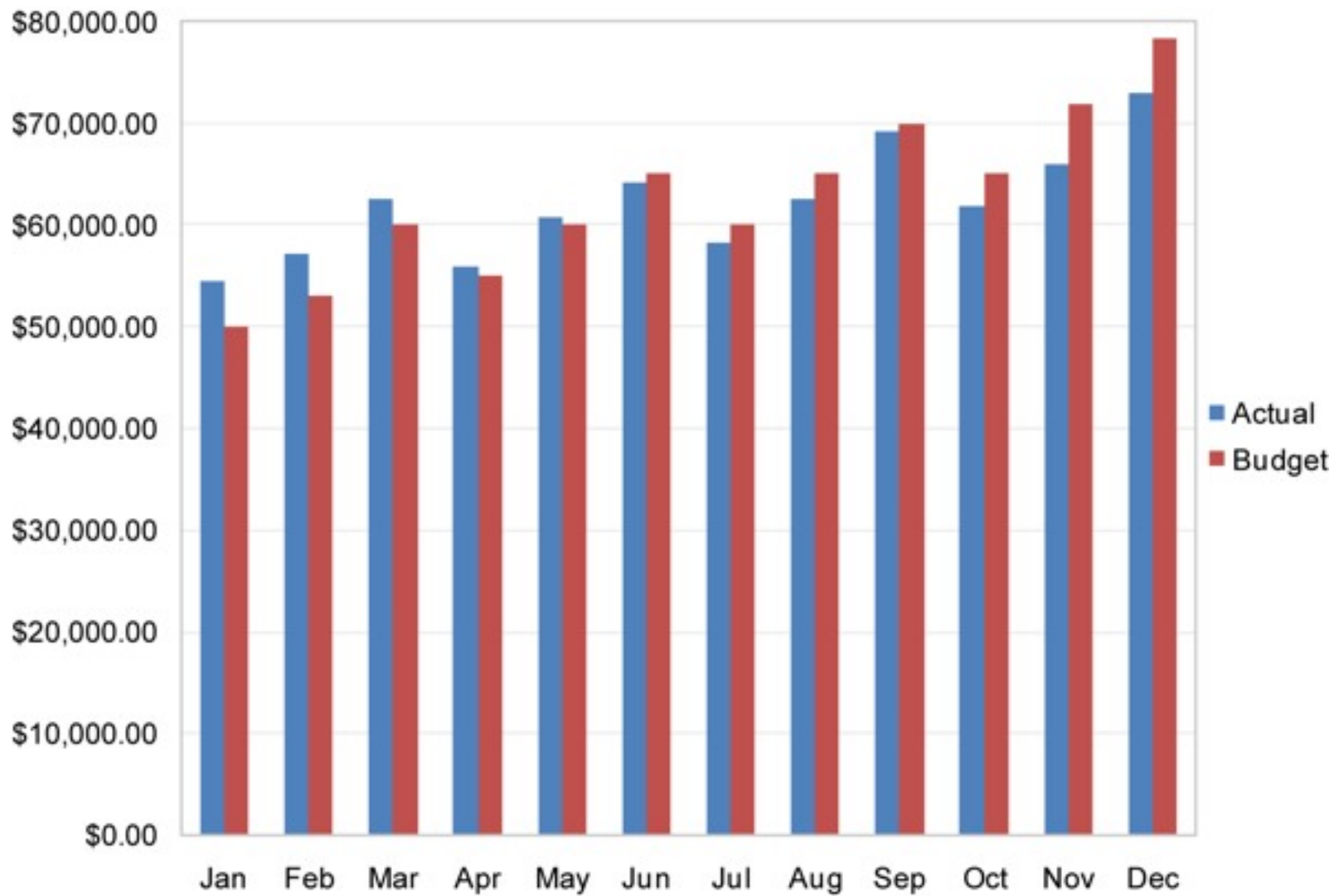
We have now enlarged the text, making it easier to read.

Poor graphs can be transformed.
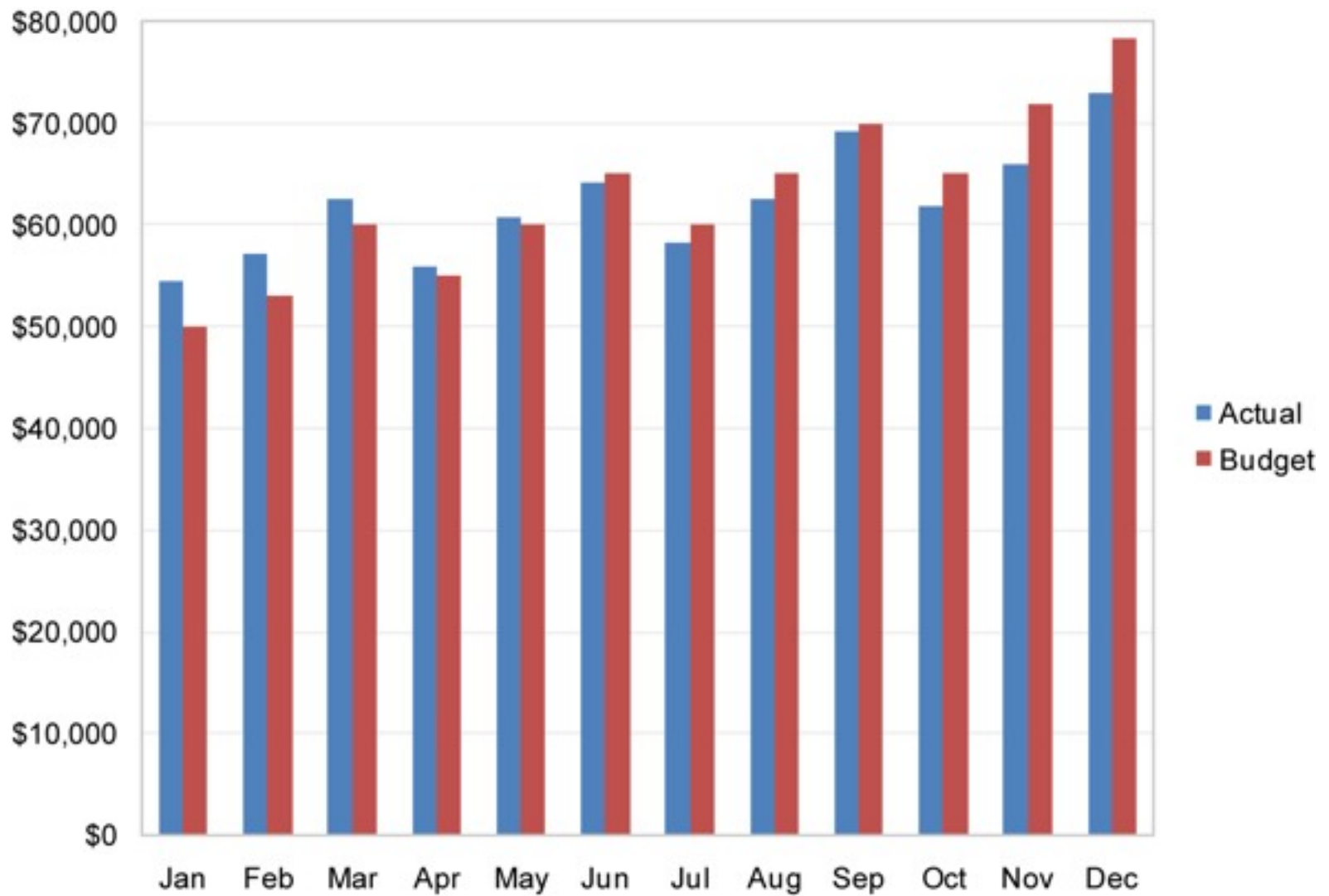
We have now removed the unnecessary decimal places in the dollar amounts along the Y-axis.

Poor graphs can be transformed.
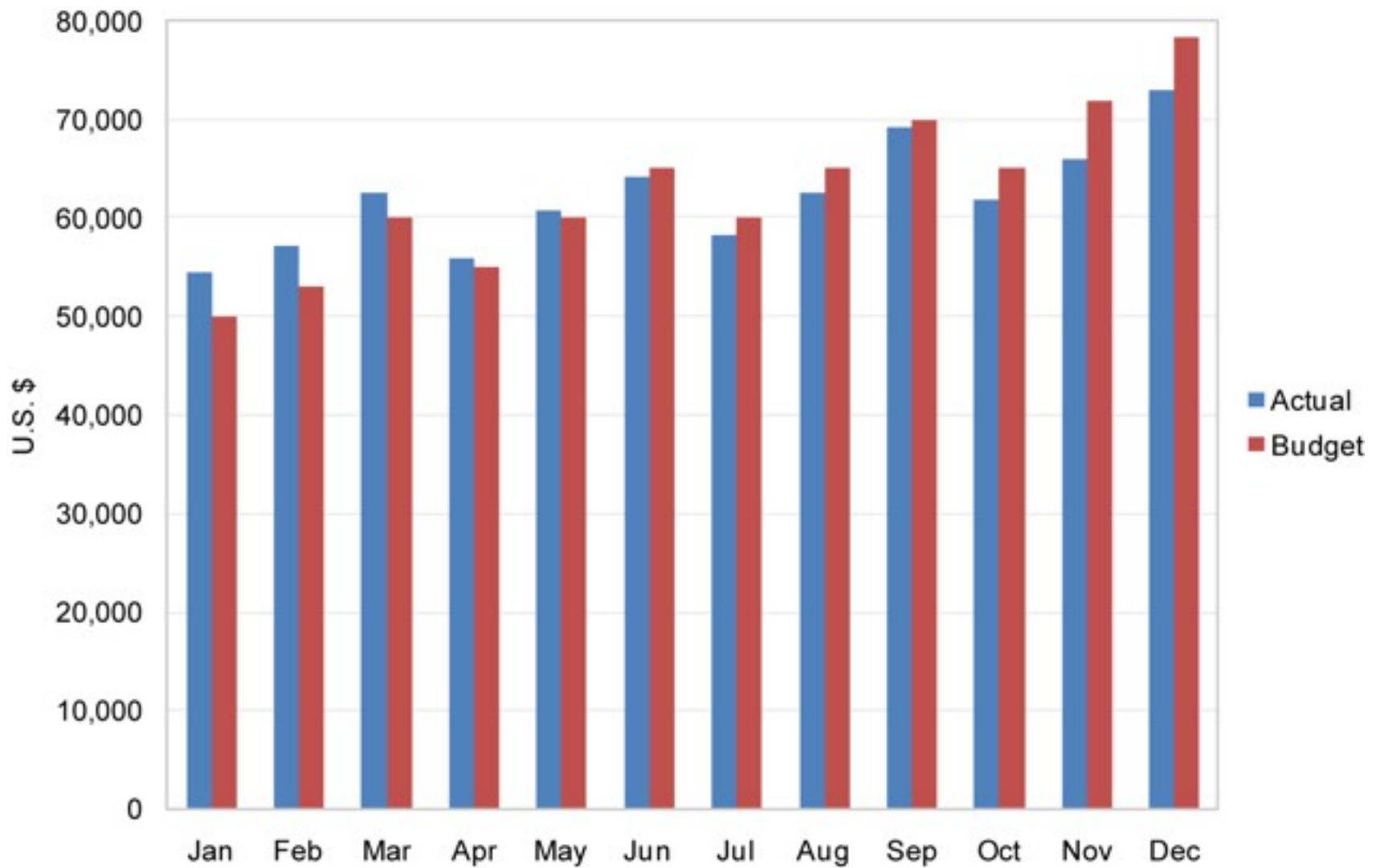
We have now removed the redundant dollar signs and labeled the unit of measure (U.S. $) clearly.

Poor graphs can be transformed.
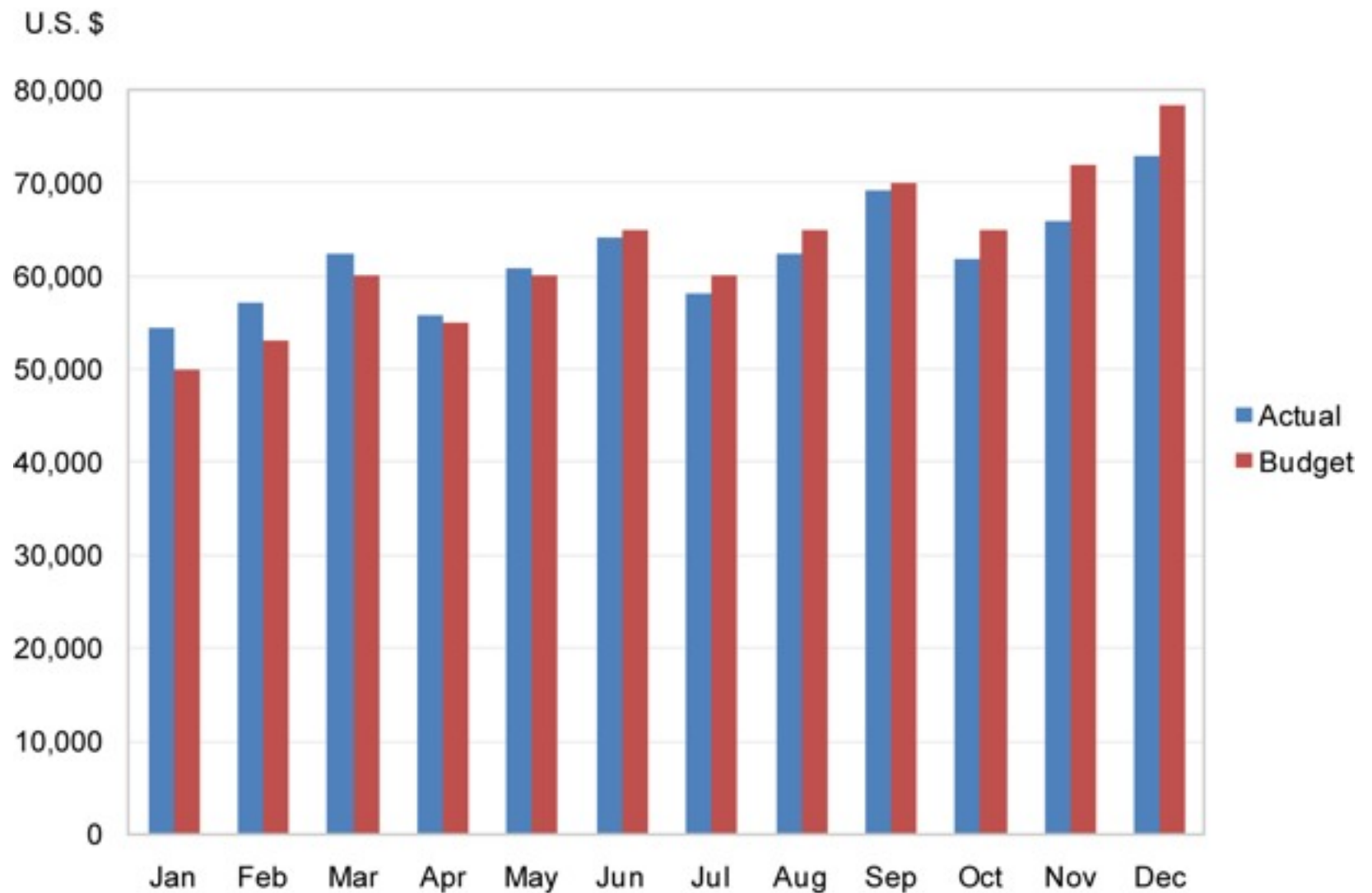
We have now reoriented the Y-axis label to the horizontal and placed it above the axis to make it easier to read.

Poor graphs can be transformed.
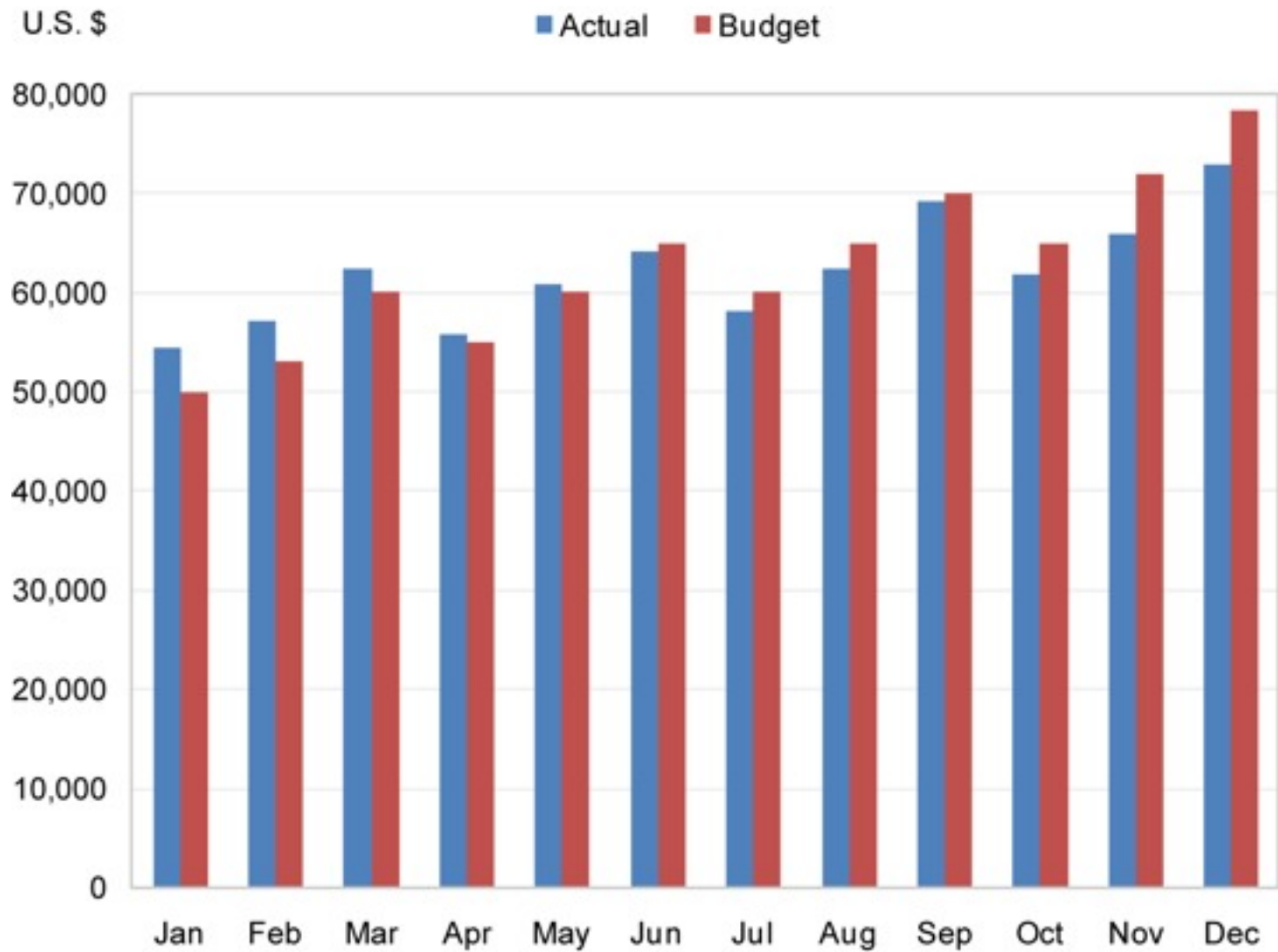
We have now reoriented and repositioned the legend to make it easier to associate it with the data bars.

Poor graphs can be transformed.
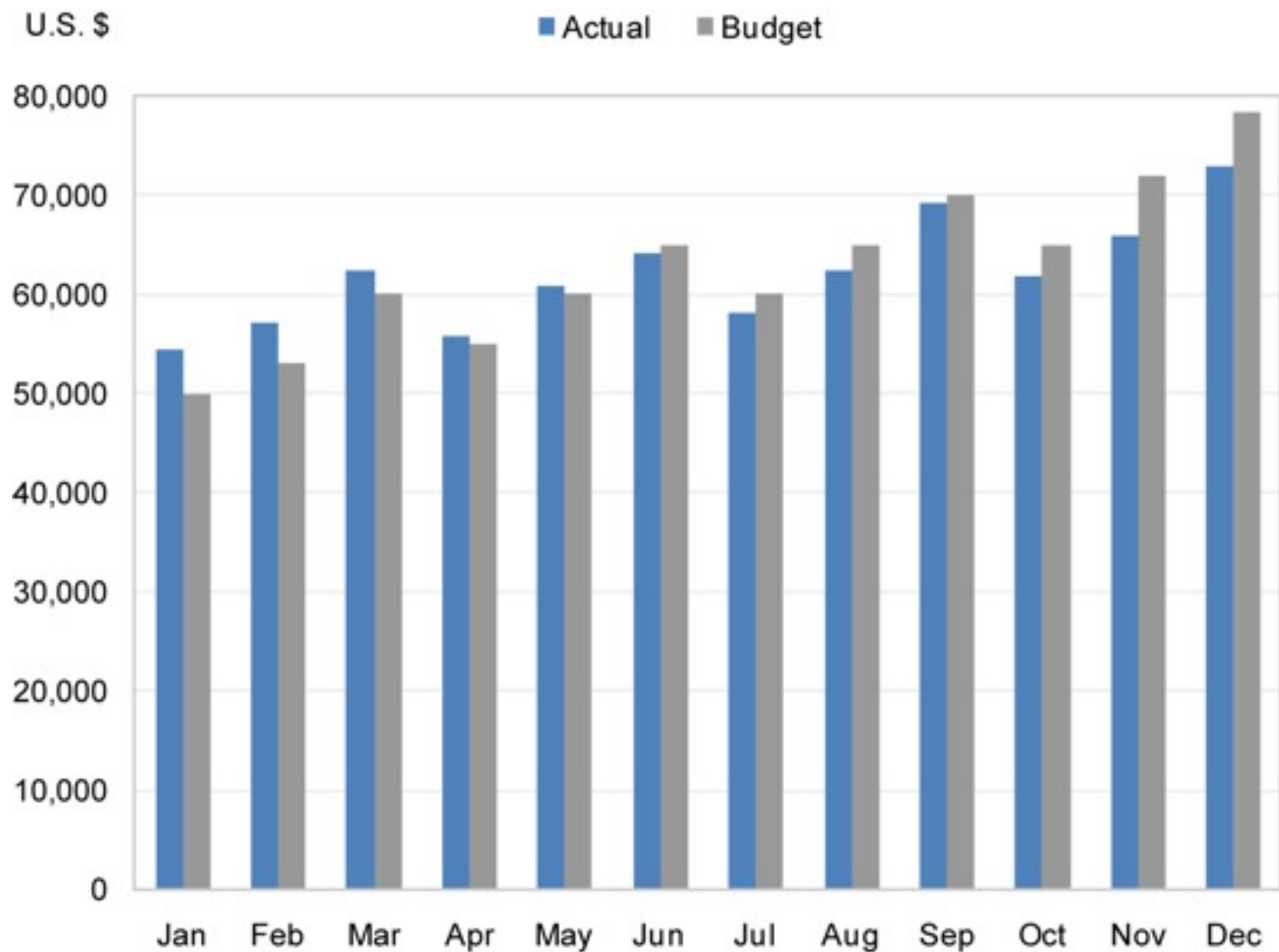
We have now changed the color of the Budget bar to be more visually pleasing in relation to the blue Actual bars. Changing from the color red also removed the possibility people interpreting the data as something bad or a warning, which red is often used to represent.

Poor graphs can be transformed.
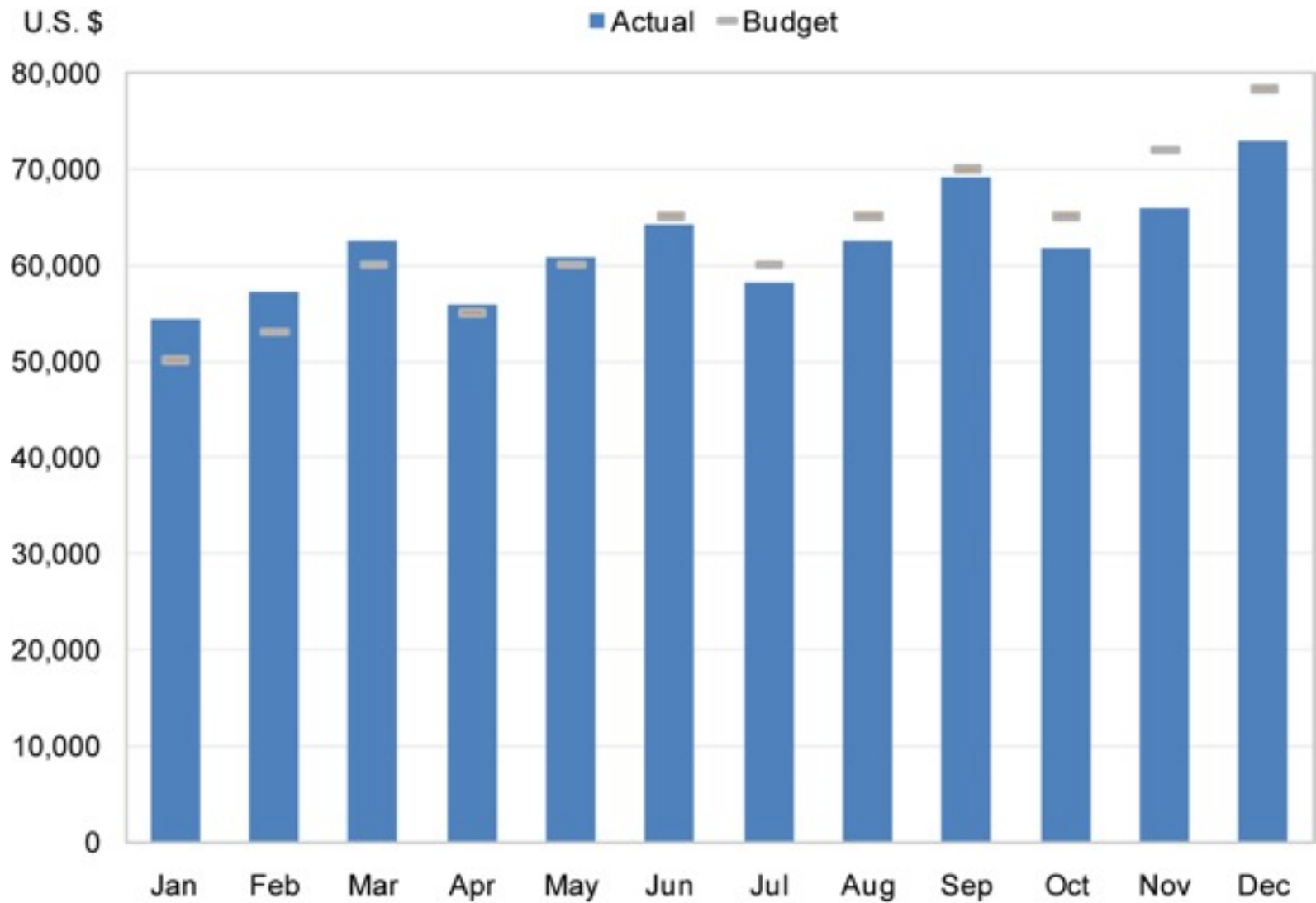
We have now reduced the visual salience of the Budget values, because they are less important that the Actual values, and have done so in a way that reduced clutter.

Poor graphs can be transformed.
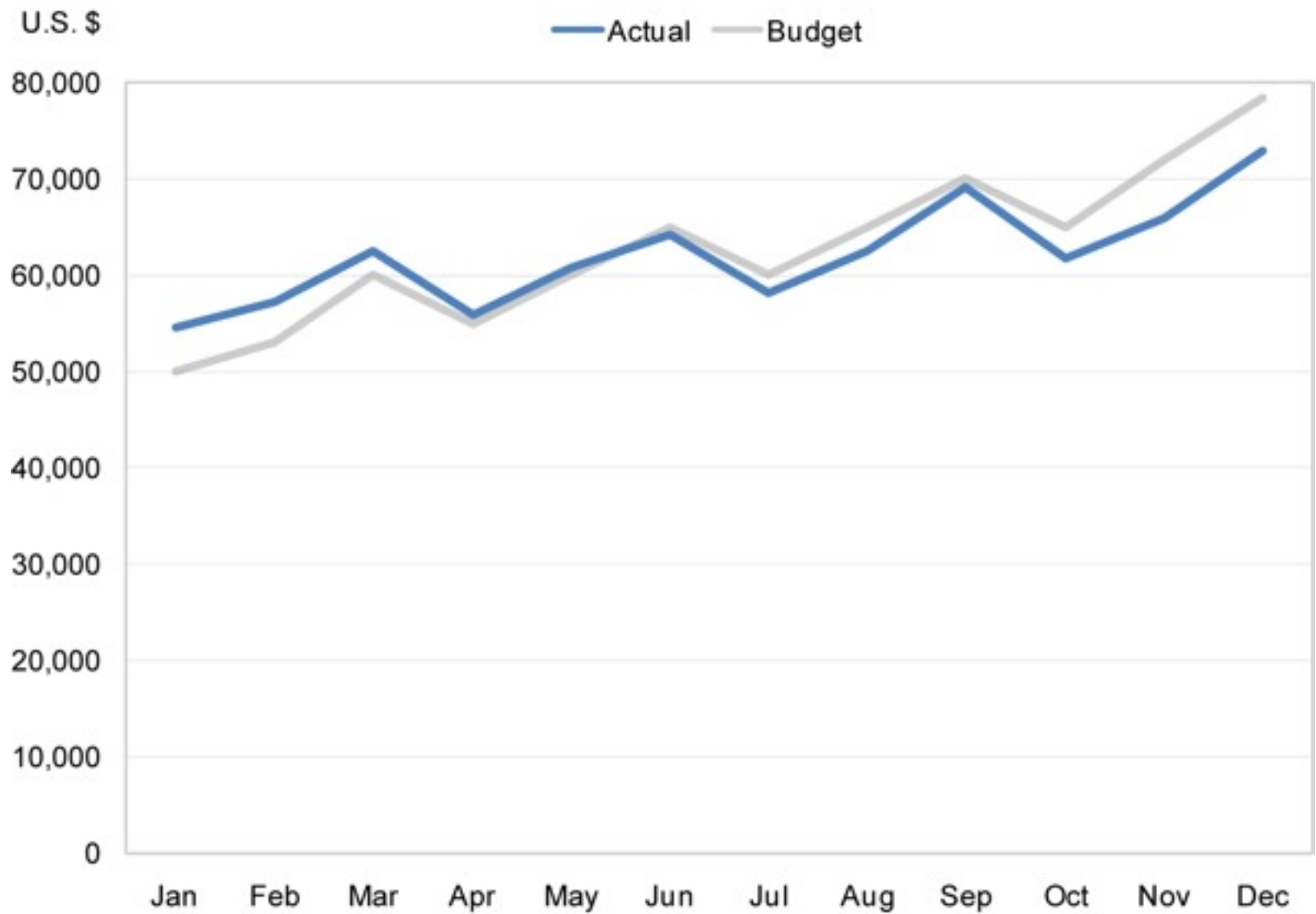
We have now made it much easier to see the pattern of change through time by using lines rather than bars to represent the data.

Poor graphs can be transformed.
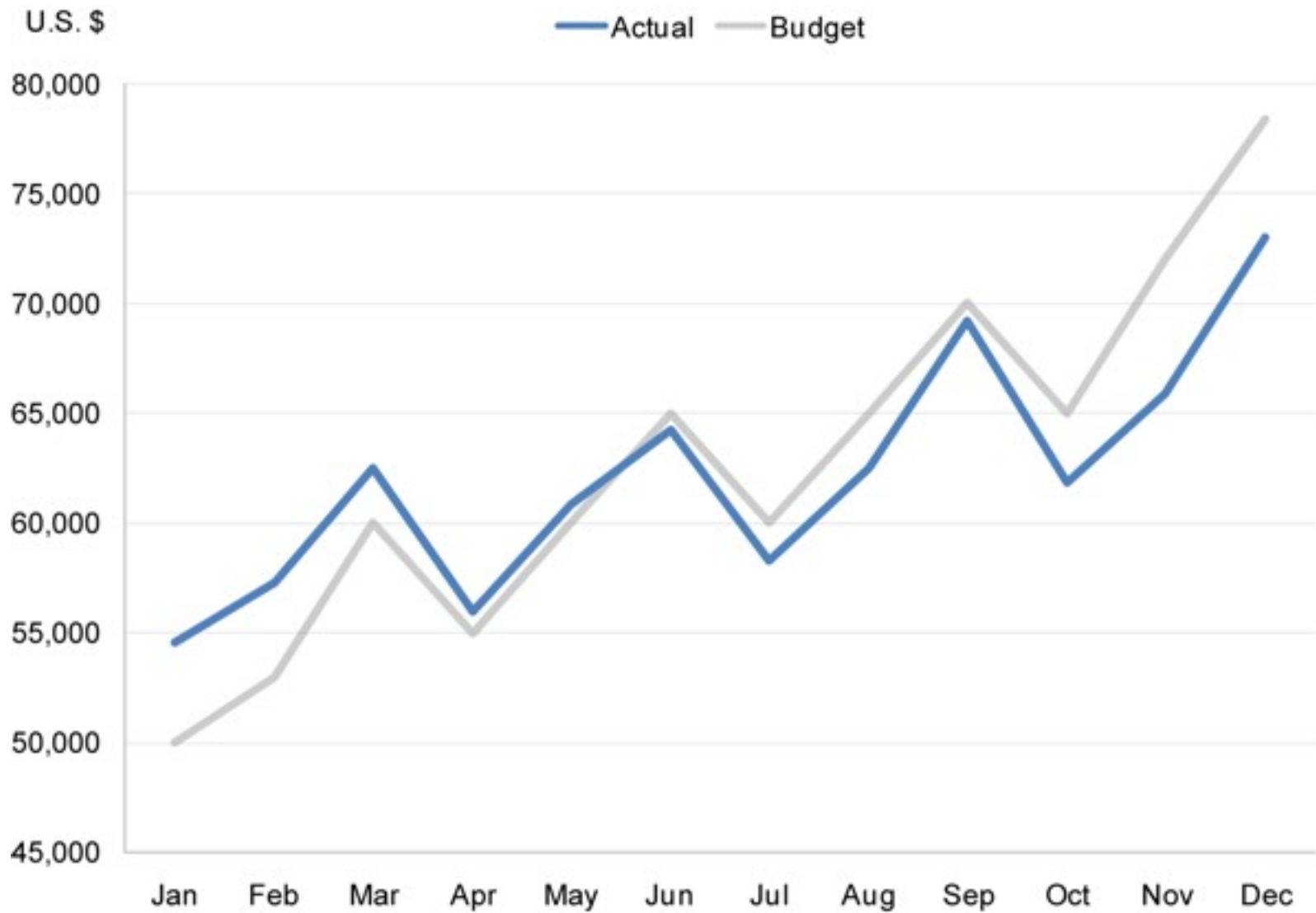
We have now made it much easier to examine the differences between actual expenses and the budget by spreading them across more space.

Poor graphs can be transformed.
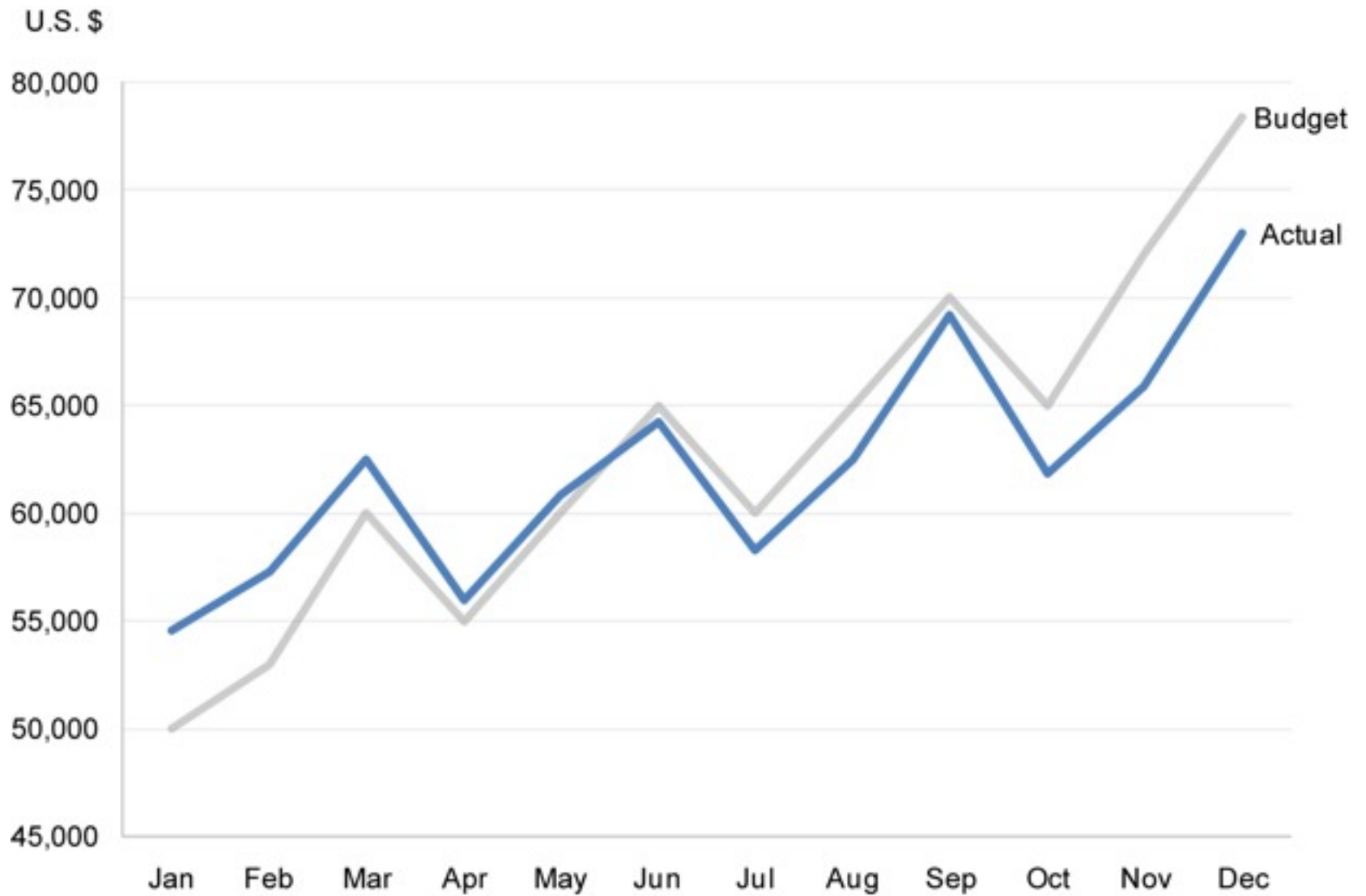
We have now labeled the lines directly, removing the need for a legend.

Poor graphs can be transformed.
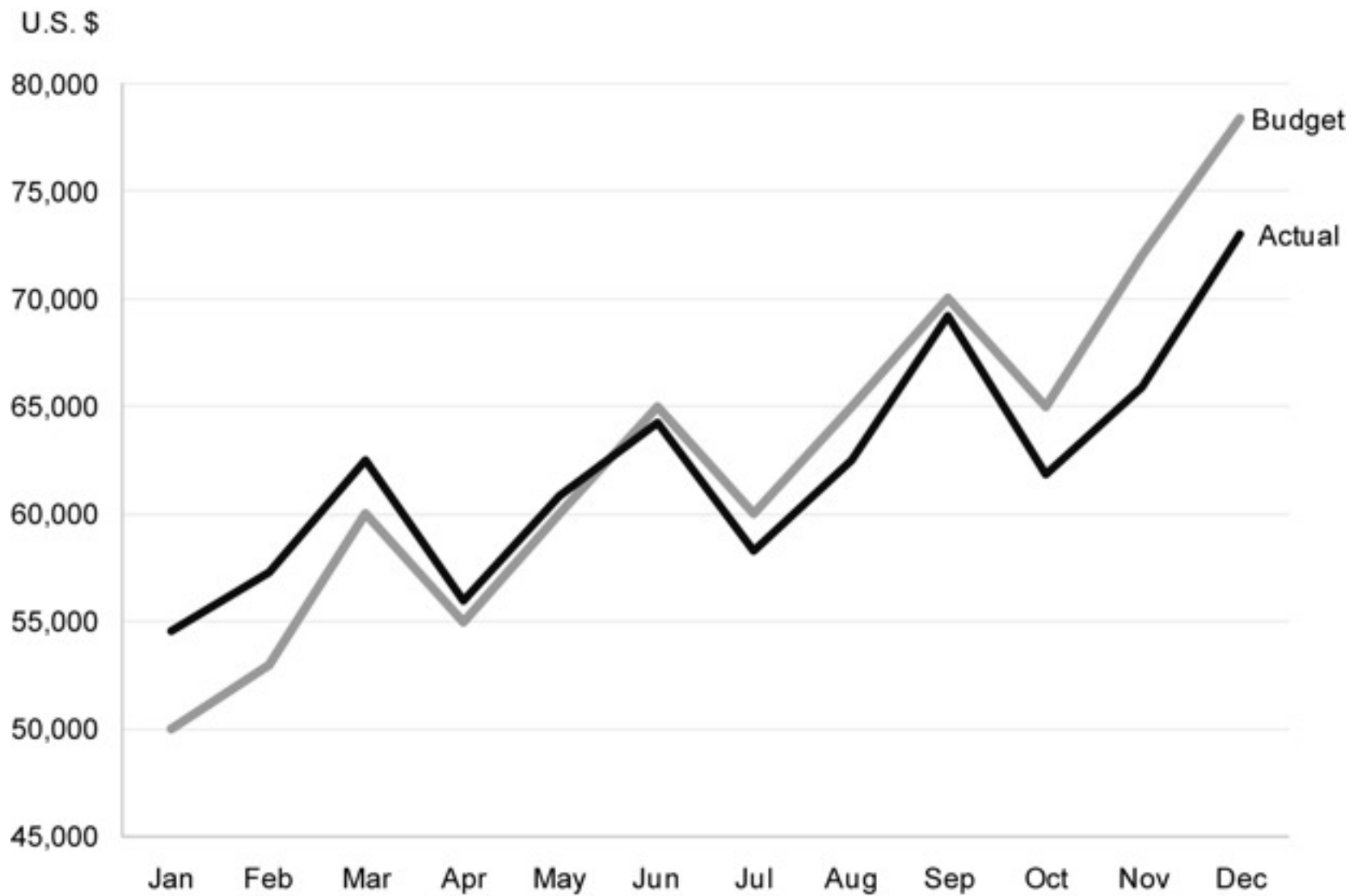
We have now changed the lines to two shades of gray to guaranty that even if the graph is printed on a black-and-white printer or photocopier, they will still look distinctly different from one another.

Poor graphs can be transformed.
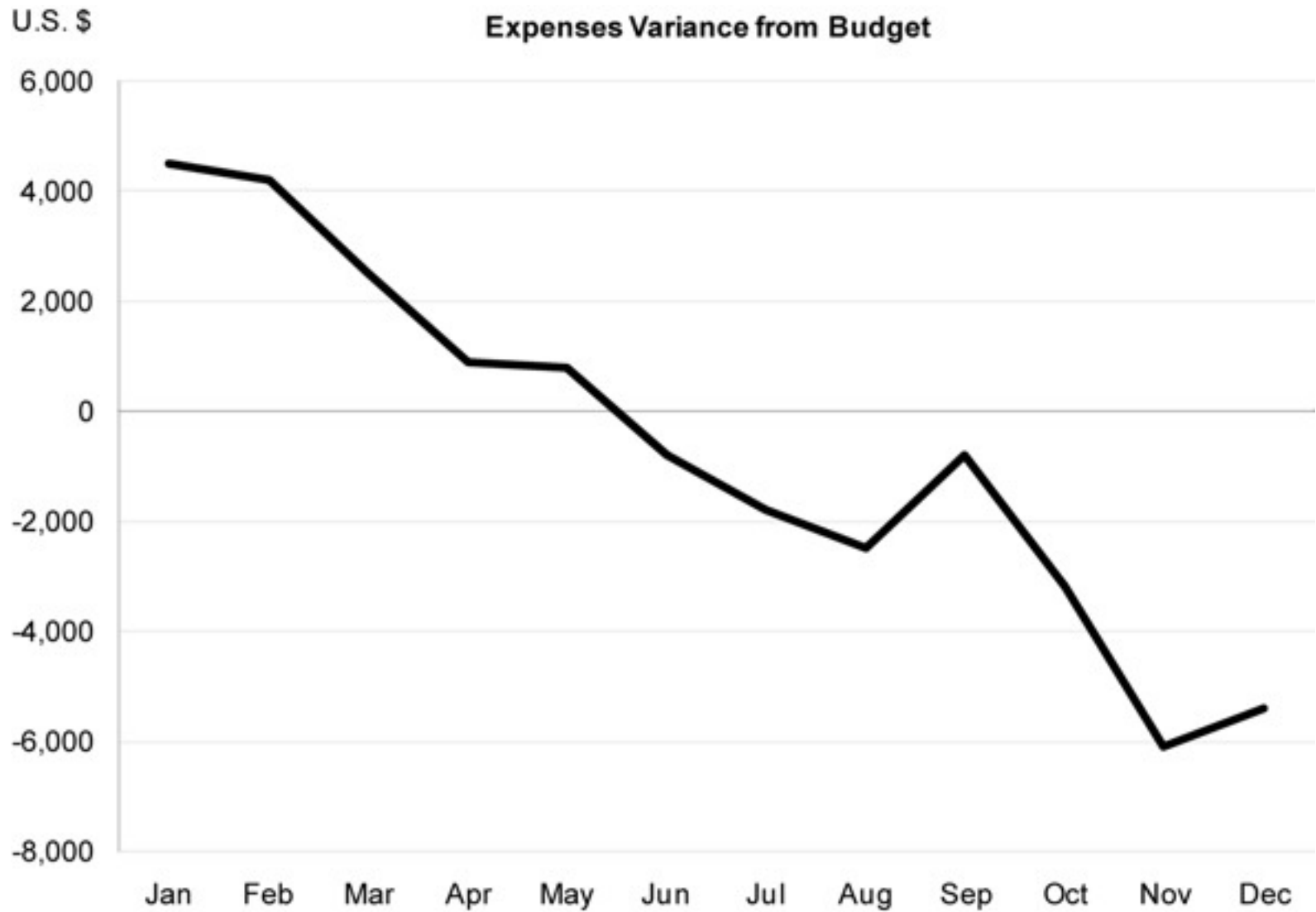
We have now represented the variance of actual expenses from the budget directly, as a single line.

Poor graphs can be transformed.

Expenses Percentage Variance from Budget

As our final step, we have expressed variance as a percentage, to provide a better measure of performance.

From incomprehensible to clear and simple
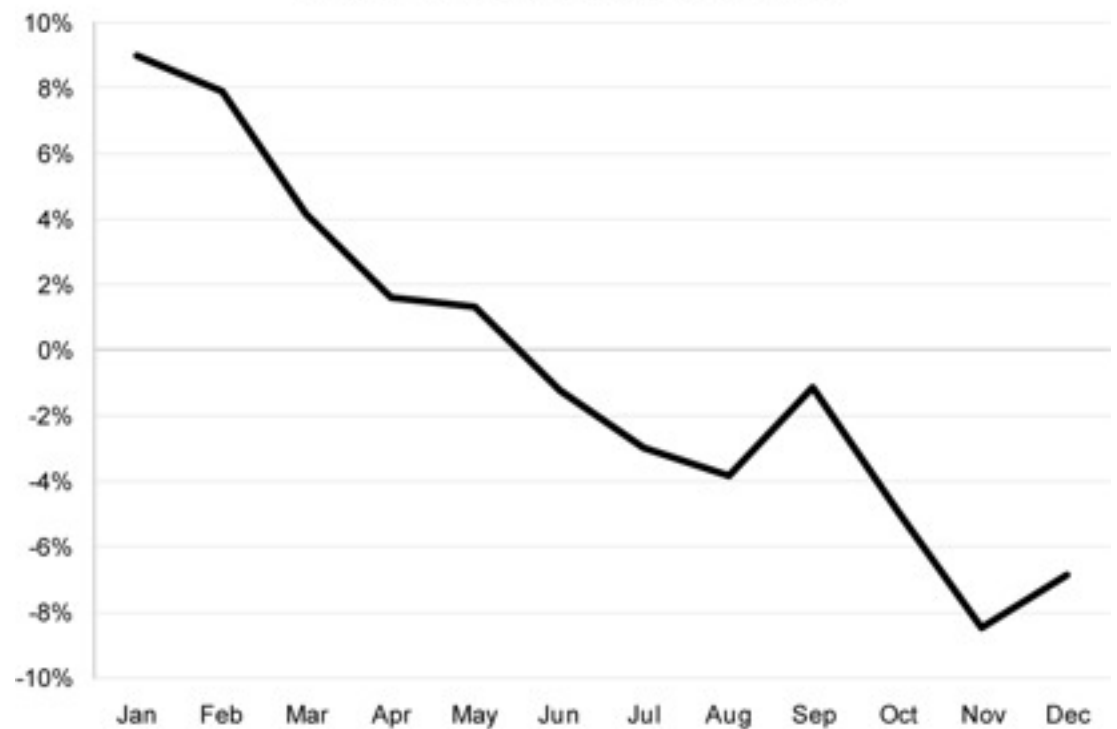
Our final solution, which we produced in sixteen steps, could have easily been our original solution. It usually takes no longer to design effective graphs than those that communicate poorly, if at all.

Quantitative displays consist of two parts.

1. Quantitative values (a.k.a. measures)

2. Categorical labels (a.k.a. dimensions)

2005 Sales (USD)

Sales Channel	Q1	Q2	Q3	Q4
Direct	383,383	403,939	437,373	538,583
Indirect	283,733	283,833	257,474	258,474
Total	667,116	687,772	694,847	797,057

2005 Sales Revenue (USD)

Sales Channel	Q1	Q2	Q3	Q4
Direct	383,383	403,939	437,373	538,583
Indirect	283,733	283,833	257,474	258,474
Total	667,116	687,772	694,847	797,057
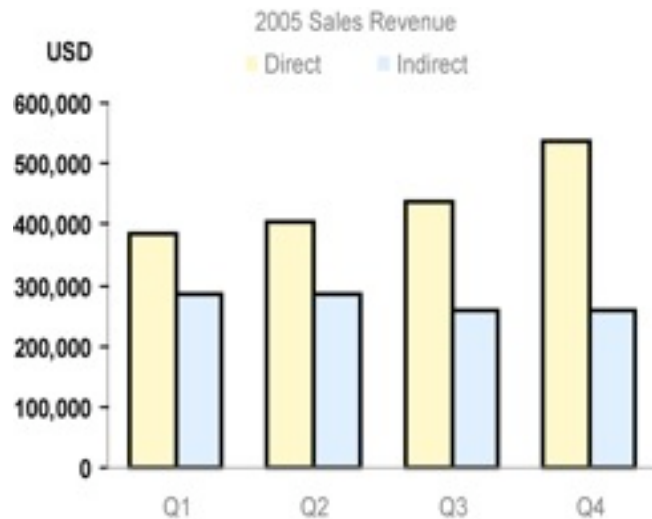
Quantitative values are numbers – measures of something related to the business (number of orders, amount of profit, rating of customer satisfaction, etc.). Categorical subdivisions break the measures down into meaningful groups and give them meaningful labels.

Scales are either quantitative or categorical.

Quantitative
How much?
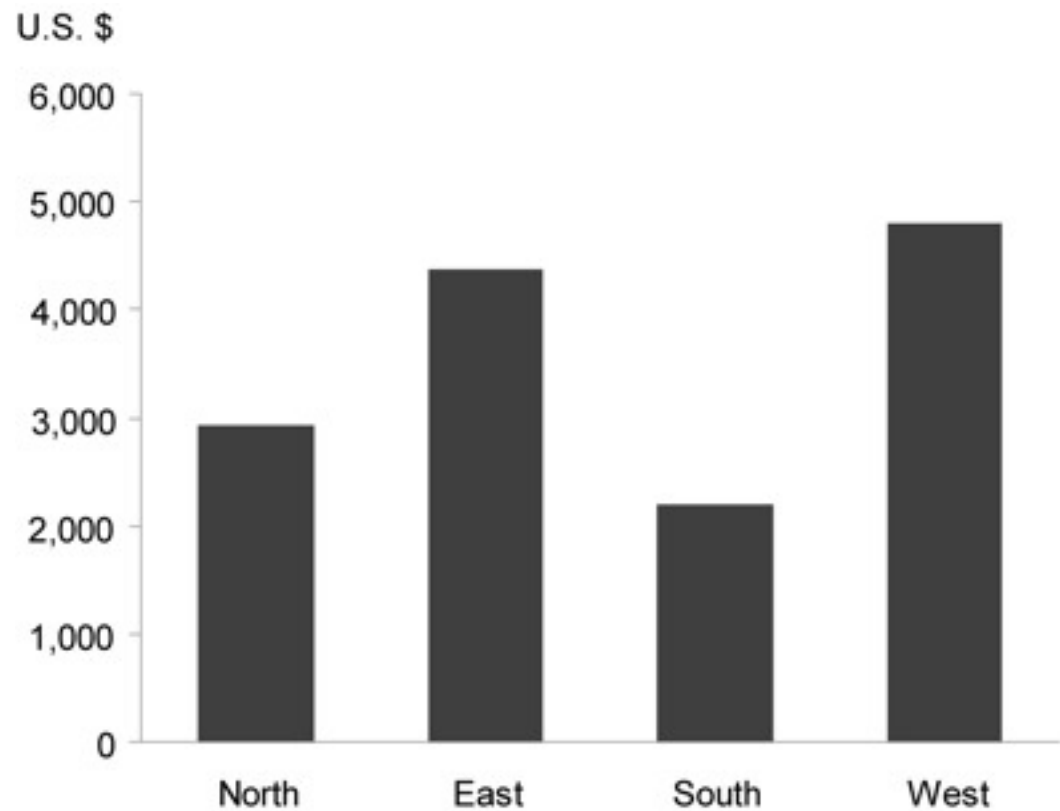
Categorical
What?

A quantitative scales consists of a range of values that is used to measure something. A categorical scale identifies what is being measured, listing the separate instances of a variable, the items in a category.

Types of categorical scales

Nominal

Sales	Operations	Engineering	HR	Marketing	Accounting

Ordinal

1st	2nd	3rd	4th	5th	6th

Interval

|					
0-99	100-199	200-299	300-399	400-499	500-599

?

January	February	March	April	May	June

Categorical scales come in several types, three of which are common in graphs:

- Nominal: The individual items along the scale differ in name only. They have no particular order and represent no quantitative values.
- Ordinal: The individual items along the scale have an intrinsic order of rank, but also do not represent quantitative values.
- Interval: The individual items along the scale have an intrinsic order, which in this case does correspond to quantitative values. Interval scales are a marriage of quantitative and categorical scales. They begin as a range of quantitative values, but are converted into a categorical scale by subdividing the range into small ranges of equal size, each of which is given a label (e.g., "0-99" or ">= 0 and <100").

The last scale above, consisting of January through June, is an interval scale. It is ordinal, in that the months have an intrinsic order, but it is an interval scale because months are equal subdivisions of time, and time is quantitative in that units of time can be added, subtracted, and so on.

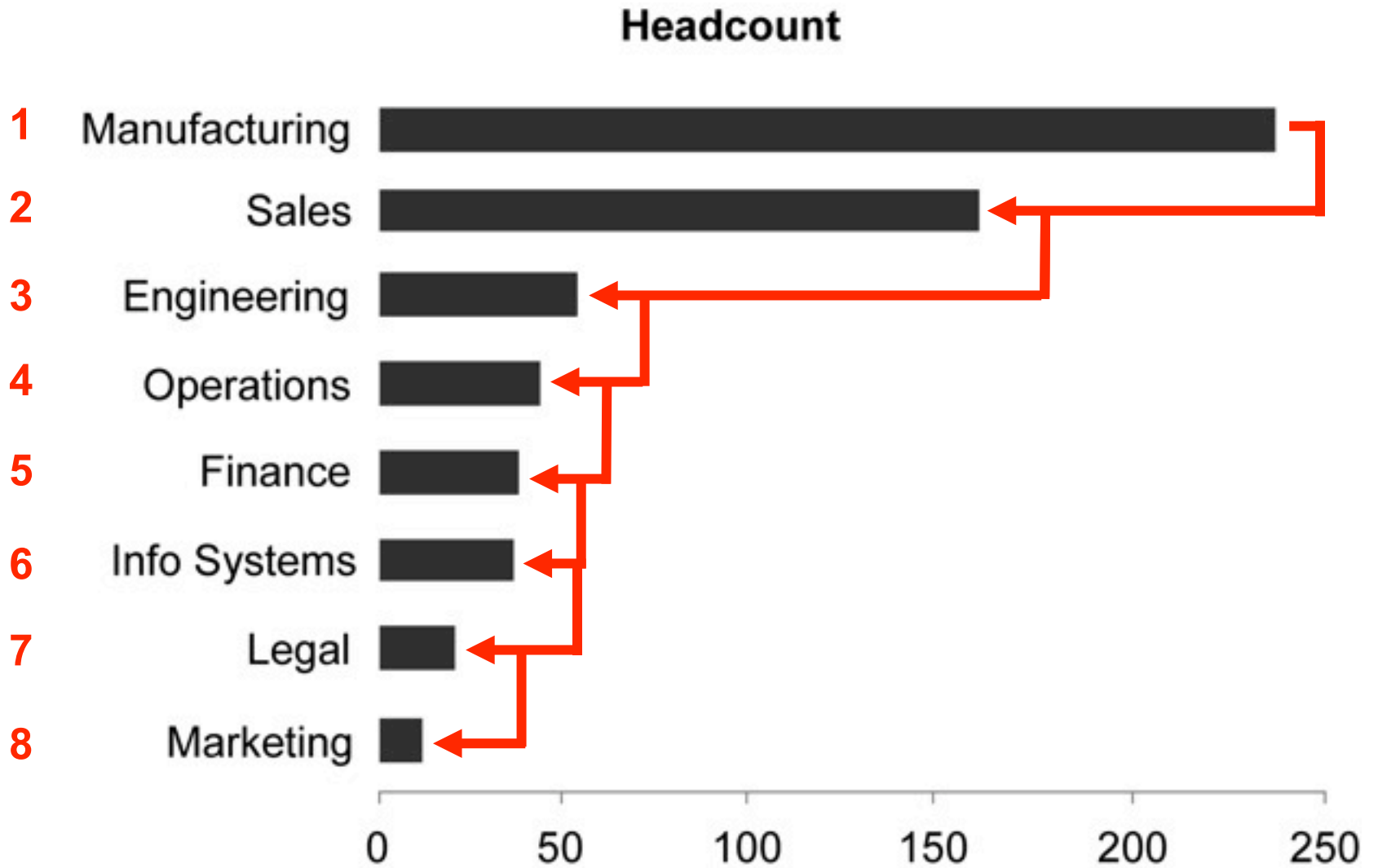# Quantitative messages always feature relationships.
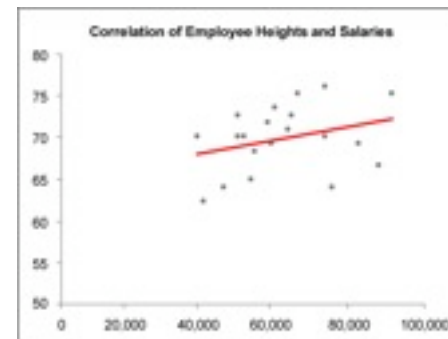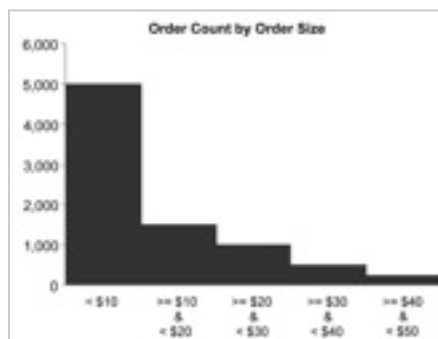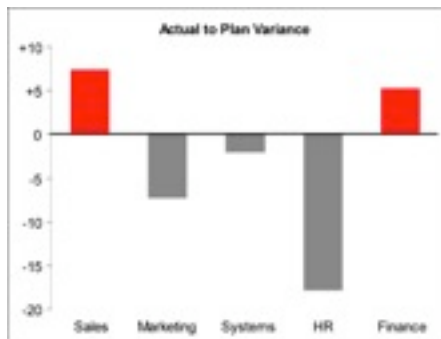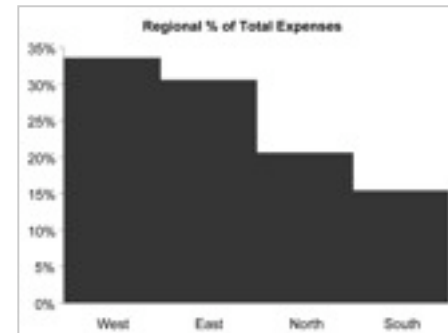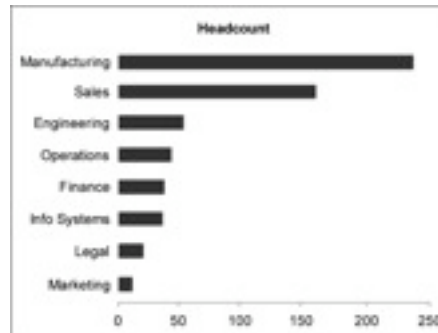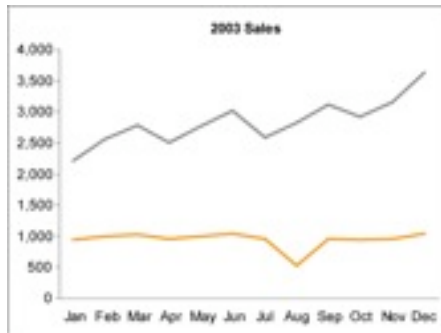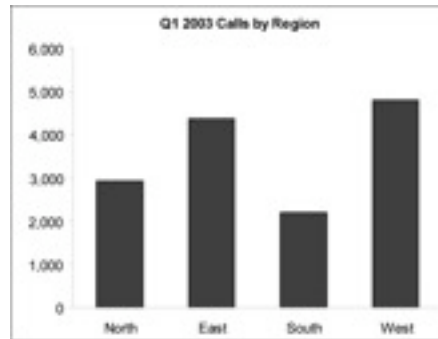
Each of these graphs illustrates a different type of quantitative relationship. Just as in life in general, the interesting and important content of a graph always involves relationships.

Relationship? Time Series

2003 Sales
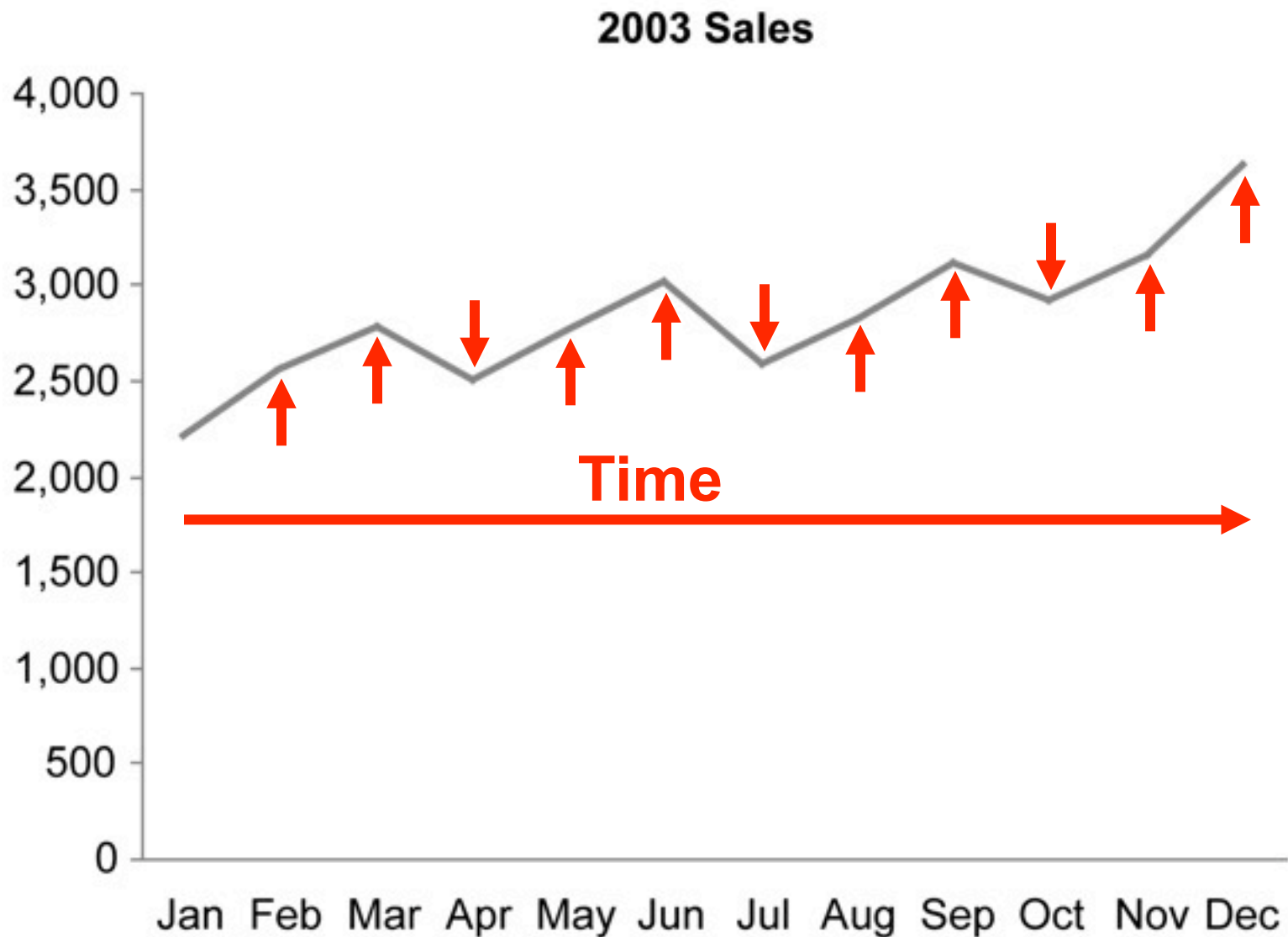
A time-series graph has a categorical scale that represents time, subdivided into a particular unit of time, such as years, quarters, months, days, or even hours. These graphs provide a powerful means to see patterns in the values as they march through time.

Relationship? Ranking

Headcount

Rank	Department
1	Manufacturing
2	Sales
3	Engineering
4	Operations
5	Finance
6	Info Systems
7	Legal
8	Marketing

0   50   100   150   200   250

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

Relationship? Part-to-Whole

Regional % of Total Expenses


West **+** East **+** North **+** South **= 100%**
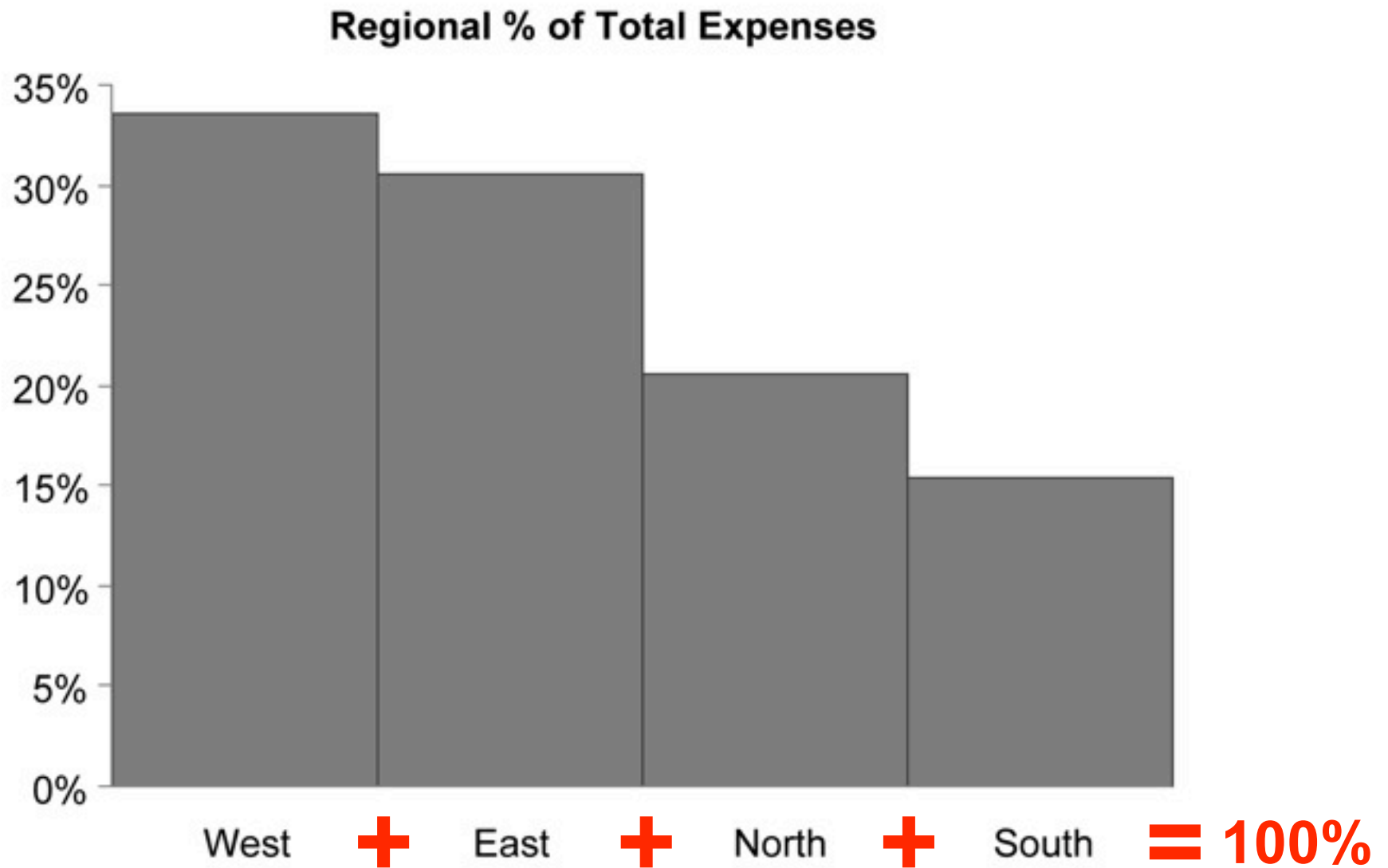
A part-to-whole graph shows how the measures associated with the individual categorical subdivisions of a full set relate to the whole and to one another.

Relationship? Deviation

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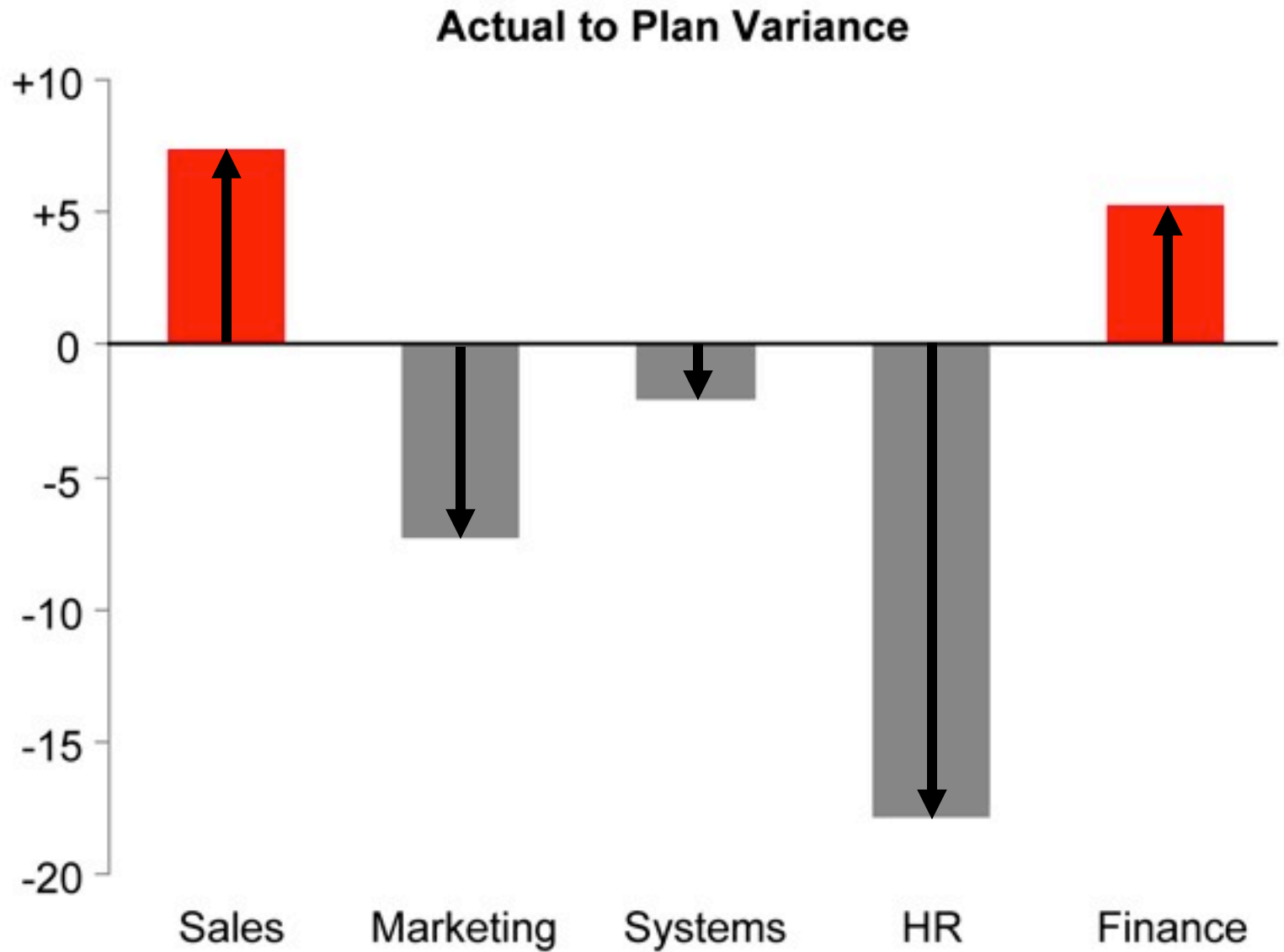
A deviation graph shows how one or more sets of values differ from a reference set of values.

Relationship? Distribution

Order Count by Order Size
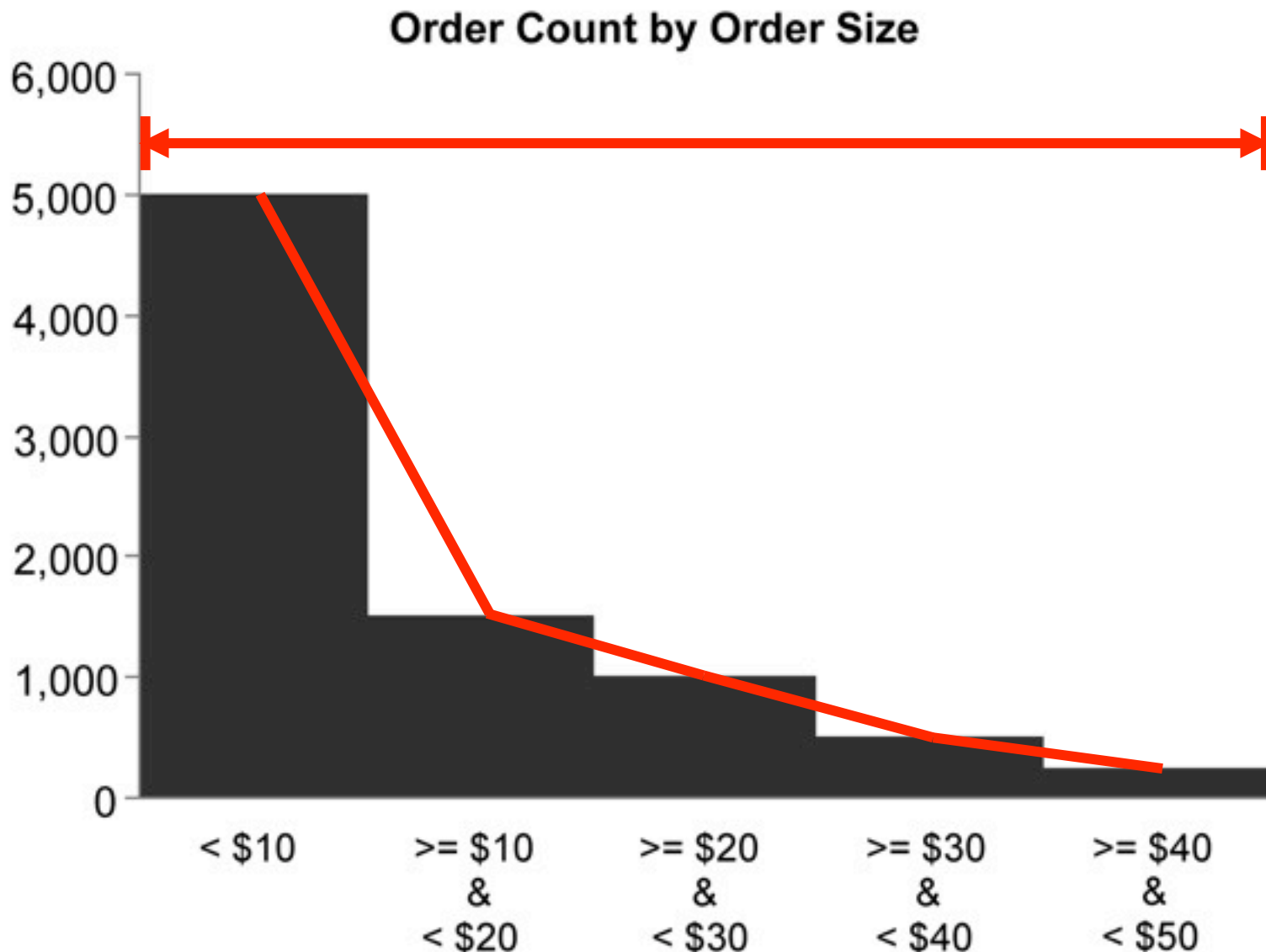
This type of distribution graph, called a frequency distribution, shows the number of times something occurs across consecutive intervals of a larger quantitative range. In a frequency distribution, a quantitative scale (in this case the range of dollar values of orders) is converted to a categorical scale by subdividing the range and giving each of the subdivisions a categorical label ("< $10", and so on).
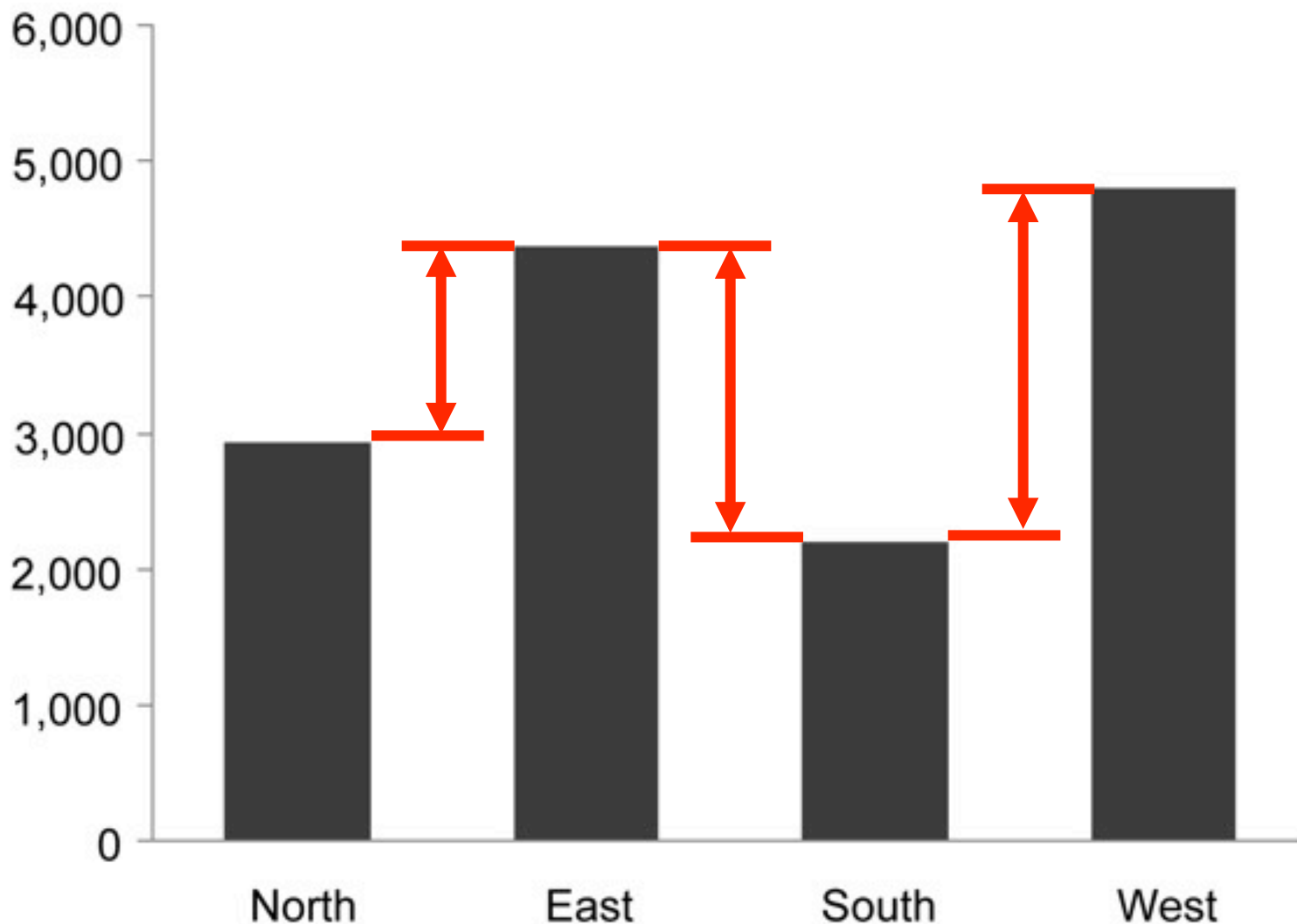
Relationship? Correlation
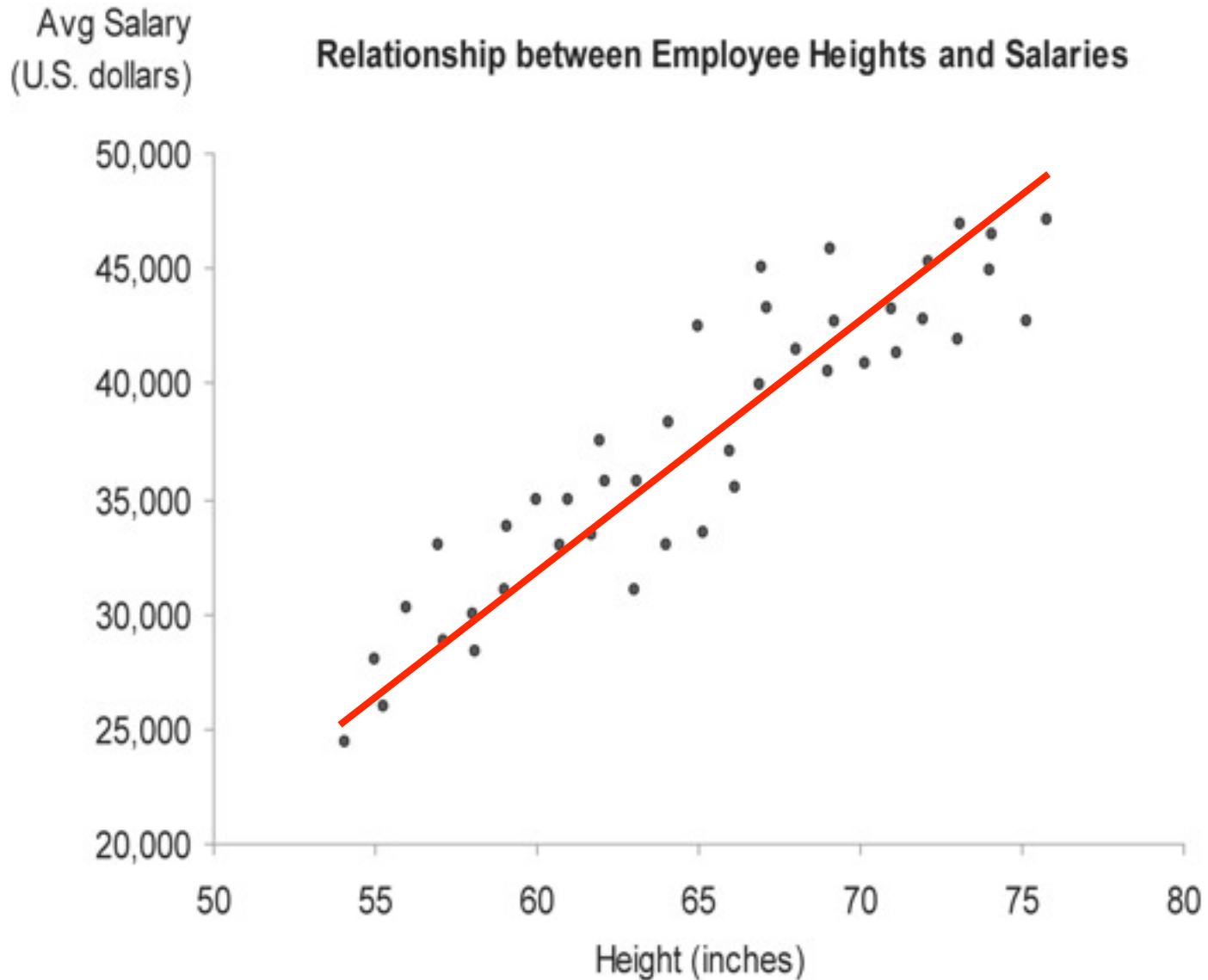
A correlation graph shows whether two paired sets of measures vary in relation to one another, and if so, in which direction (positive or negative) and to what degree (strong or weak). If the trend line moves upwards, the correlation is positive; if it moves downwards, it is negative. A positive correlation 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.

Relationship? Nominal Comparison

Q1 2003 Calls by Region

The term nominal means "in name only." 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.

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.

Selecting the right medium is half the battle.

Once you've selected the best means to display your data and message, the next big challenge is to design the visual components of the display to state your message clearly, without distraction, and to highlight what's most important.

Which would you rather look at?

What if you are Jessica's dermatologist?

Dressing things up is appropriate for advertising, because the illusion pleases and sells. When you're responsible for discovering the truth and understanding it, however, makeup only gets in the way.

Data-Ink Ratio

Data Ink

Non-Data Ink

According to Edward Tufte, tables and graphs are made up 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.

Table design practice

YTD International Revenue

Product	Jan	Feb	Mar	Apr	May	Jun	Total
Disk Drives	$93,993.00	$84,773.00	$88,833.00	$95,838.00	$93,874.00	$83,994.00	$541,305.00
Monitors	$87,413.00	$78,838.00	$82,614.00	$89,129.00	$873,020.00	$78,114.00	$1,289,128.00
Printers	$90,035.00	$2,120,400.00	$85,093.00	$91,803.00	$899,210.00	$80,457.00	$3,366,998.00
Computers	$92,736.00	$83,640.00	$87,645.00	$94,557.00	$92,619.00	$82,871.00	$534,068.00
Memory Sticks	$3,624,500.00	$77,785.00	$81,510.00	$87,938.00	$86,136.00	$77,070.00	$4,034,939.00
Sound Cards	$88,832.00	$80,118.00	$83,956.00	$90,576.00	$88,720.00	$79,382.00	$511,584.00
Video Cards	$82,614.00	$74,510.00	$78,079.00	$84,236.00	$82,509.00	$73,825.00	$475,773.00
RAM	$85,092.00	$76,745.00	$80,421.00	$86,763.00	$84,985.00	$76,040.00	$490,046.00
Scanners	$87,645.00	$79,048.00	$82,834.00	$89,366.00	$87,534.00	$78,321.00	$504,748.00
Input Devices	$90,275.00	$81,419.00	$85,319.00	$920,470.00	$90,160.00	$80,671.00	$1,348,314.00
Total	$4,423,135.00	$2,837,276.00	$836,304.00	$1,730,676.00	$2,478,767.00	$790,745.00	$13,096,903.00

The primary purpose of this table is primarily to communicate the total 2003 year-to-date revenue for each product to the nearest U.S. dollar.

Improve the design of this table, keeping the following steps in mind:

1. Reduce the non-data ink.
   - Subtract what's unnecessary.
   - De-emphasize and regularize what remains.
2. Enhance the data ink.
   - Subtract what's unnecessary.
   - Emphasize what's most important.

Simple design changes make a difference

YTD International Revenue

Product	Jan	Feb	Mar	Apr	May	Jun	Total
Disk Drives	$93,993.00	$84,773.00	$88,833.00	$95,838.00	$93,874.00	$83,994.00	$541,305.00
Monitors	$87,413.00	$78,838.00	$82,614.00	$89,129.00	$873,020.00	$78,114.00	$1,289,128.00
Printers	$90,035.00	$2,120,400.00	$85,093.00	$91,803.00	$899,210.00	$80,457.00	$3,366,998.00
Computers	$92,736.00	$83,640.00	$87,645.00	$94,557.00	$92,619.00	$82,871.00	$534,068.00
Memory Sticks	$3,624,500.00	$77,785.00	$81,510.00	$87,938.00	$86,136.00	$77,070.00	$4,034,939.00
Sound Cards	$88,832.00	$80,118.00	$83,956.00	$90,576.00	$88,720.00	$79,382.00	$511,584.00
Video Cards	$82,614.00	$74,510.00	$78,079.00	$84,236.00	$82,509.00	$73,825.00	$475,773.00
RAM	$85,092.00	$76,745.00	$80,421.00	$86,763.00	$84,985.00	$76,040.00	$490,046.00
Scanners	$87,645.00	$79,048.00	$82,834.00	$89,366.00	$87,534.00	$78,321.00	$504,748.00
Input Devices	$90,275.00	$81,419.00	$85,319.00	$920,470.00	$90,160.00	$80,671.00	$1,348,314.00
Total	$4,423,135.00	$2,837,276.00	$836,304.00	$1,730,676.00	$2,478,767.00	$790,745.00	$13,096,903.00

YTD International Revenue (USD)

Product	Jan	Feb	Mar	Apr	May	Jun	Total
Memory Sticks	3,624,500	77,785	81,510	87,938	86,136	77,070	4,034,939
Printers	90,035	2,120,400	85,093	91,803	899,210	80,457	3,366,998
Input Devices	90,275	81,419	85,319	920,470	90,160	80,671	1,348,314
Monitors	87,413	78,838	82,614	89,129	873,020	78,114	1,289,128
Disk Drives	93,993	84,773	88,833	95,838	93,874	83,994	541,305
Computers	92,736	83,640	87,645	94,557	92,619	82,871	534,068
Sound Cards	88,832	80,118	83,956	90,576	88,720	79,382	511,584
Scanners	87,645	79,048	82,834	89,366	87,534	78,321	504,748
RAM	85,092	76,745	80,421	86,763	84,985	76,040	490,046
Video Cards	82,614	74,510	78,079	84,236	82,509	73,825	475,773
Total	4,423,135	2,837,276	836,304	1,730,676	2,478,767	790,745	13,096,903

# We must understand how people see.
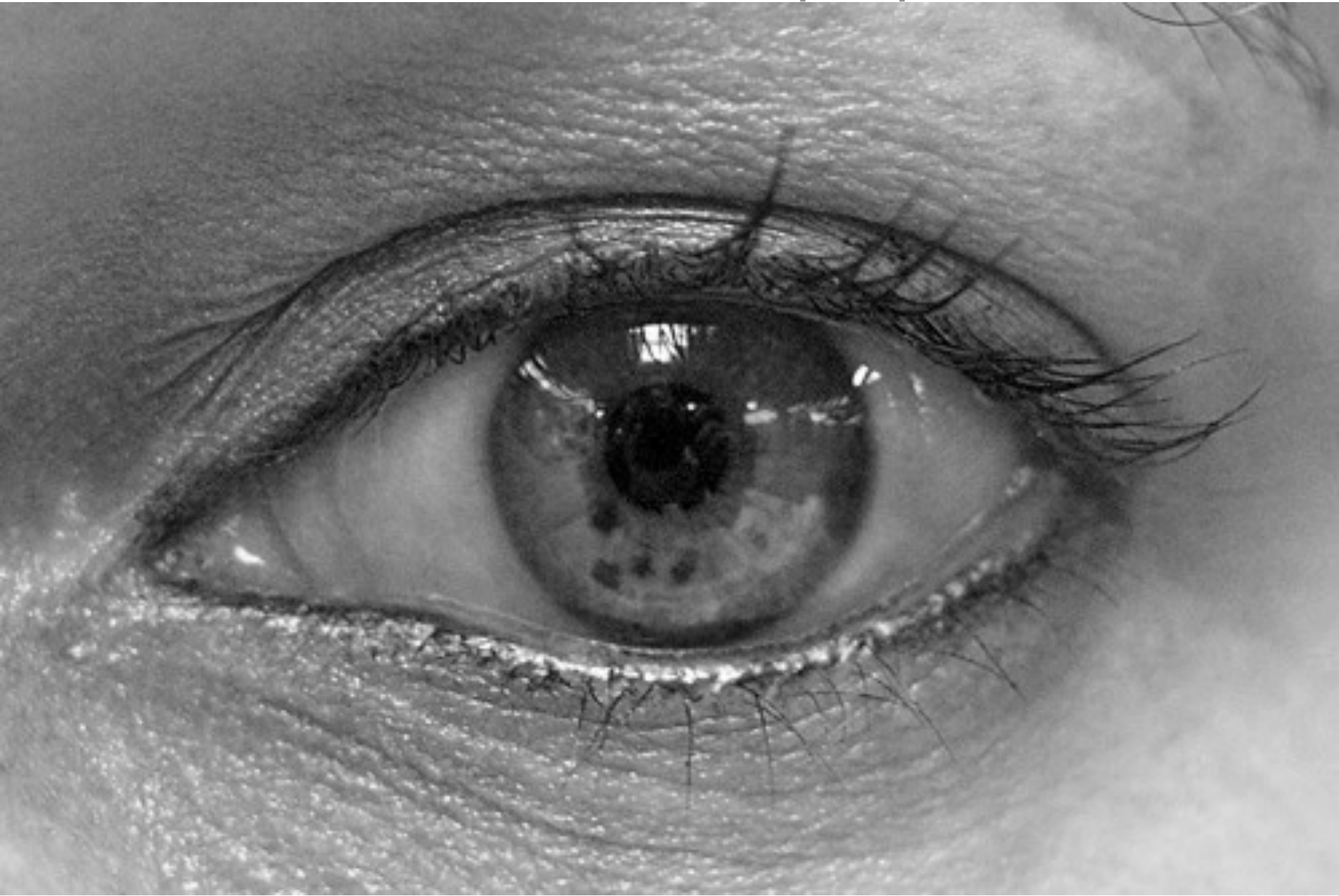
To know how to present information visually in an effective way, you must understand a little about visual perception – what works, what doesn't, and why.

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… 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…If we can understand how perception works, our knowledge can be translated into rules for displaying information… If we disobey the rules, our data will be incomprehensible or misleading.

Colin Ware, *Information Visualization: Perception for Design*, 2nd Edition, Morgan Kaufmann Publishers, San Francisco, 2004.

Visual perception is not just camera work.
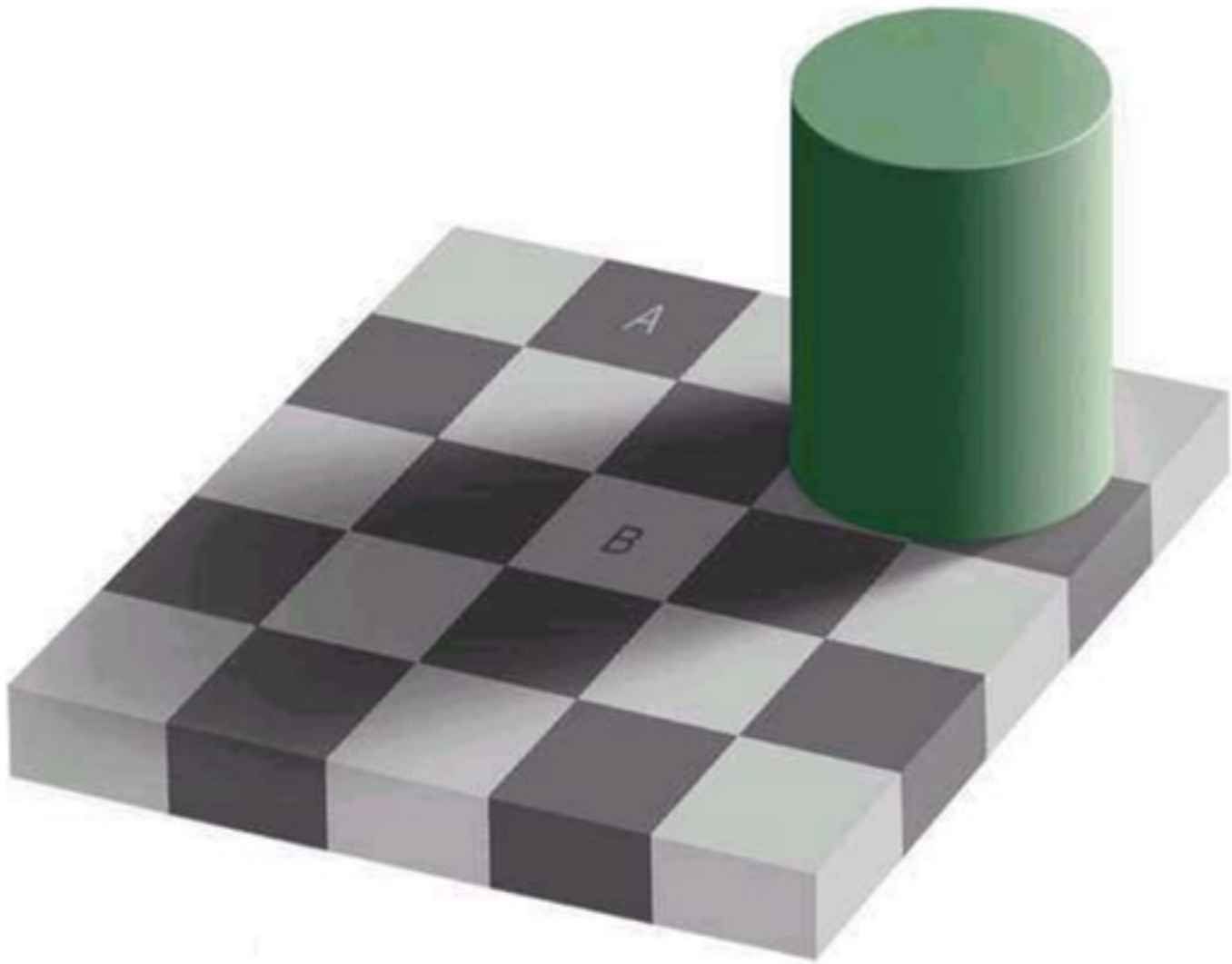
Square A is darker than B, right?

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

We perceive differences, not absolutes.

Take a closer look.

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.

Context affects what we see.
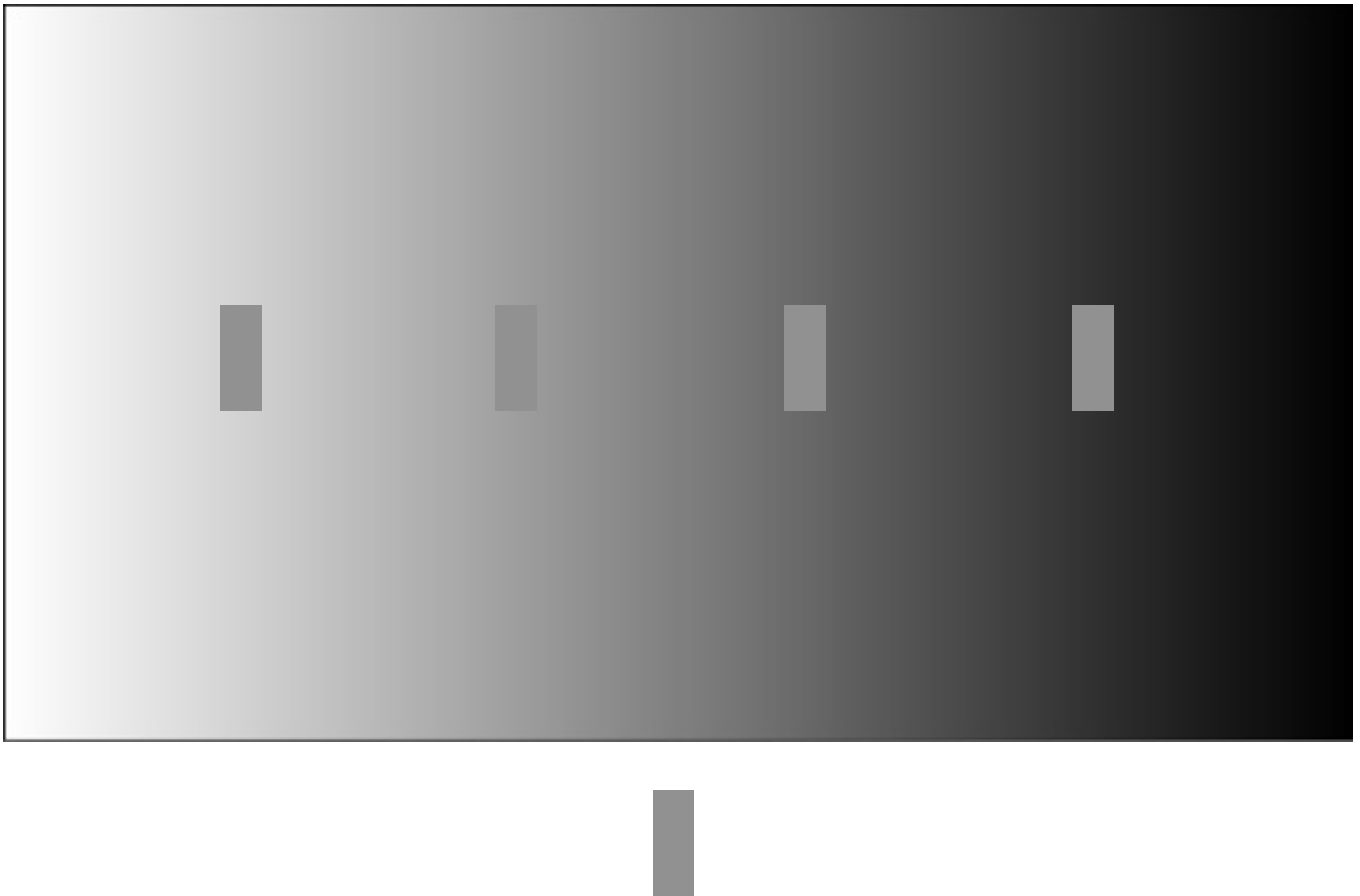
This image illustrates the surprising effect that a simple change in the lightness of the background alone has on our perception of color. The large rectangle displays a simple color gradient of a gray-scale from fully light to fully dark. The small rectangle is the same exact color everywhere it appears, 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.

From light to perception
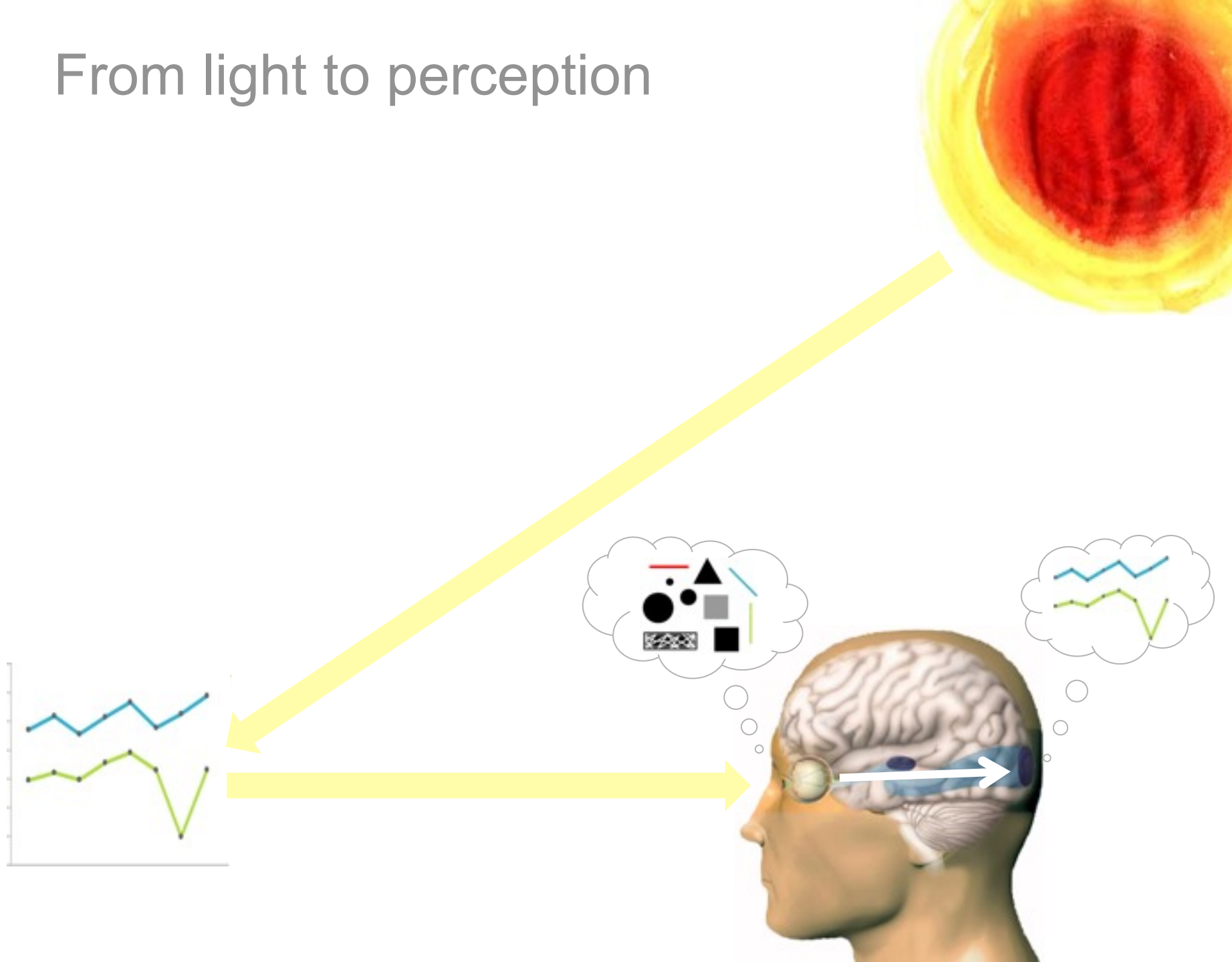
- Our eyes sense light that bounces off the surfaces of objects in the world.

- What we perceive as an object is built up as a composite of several visual properties—building blocks of vision.

- Although we perceive this composite of properties as a whole object, we can still distinguish the properties that compose it.

- These properties are individual attributes of light that our eyes are tuned to sense.

- These attributes include 2-D location, length, width, area, shape, color, orientation, and so on.

- We perceive these attributes preattentively, which means that this perception takes place without and in advance of conscious thought.

- Because it works in this way, with millions of receptors in our eyes working together at once, preattentive perception is an extremely fast parallel process.

- When data is represented in the form of preattentive visual attributes, we are able to perceive a great deal of information rapidly, and capable of storing an entire object made up of many properties as a single chunk of memory.

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

98734970275647902194728624092406037080470289072780320802900730590127023800837408207872027200808324780260270379371570970137970667462097094102780692797970912309723091959275092730927219787 3497260

98734970275647902194728624092406037080470289072780320802900730590127023800837408207872027200808324780260270379371570970137970667462097094102780692797970912309723091959275092730927219787 3497260

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.

Pre-attentive attributes of visual perception

Form
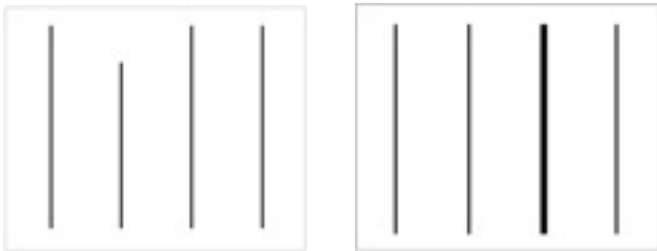
➡️Length   ⇨Width
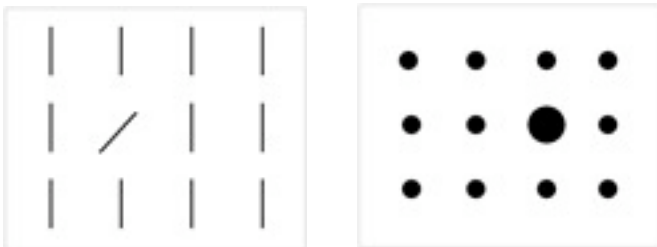
Orientation   ⇨Size
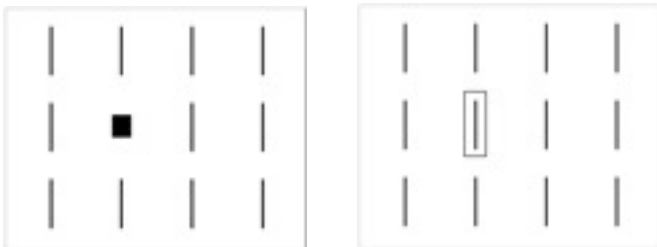
Shape   Enclosure

Color
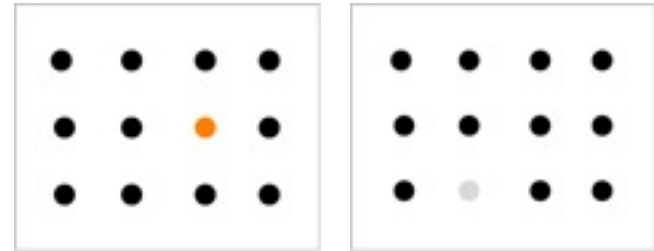
Hue   ⇨Intensity

Position

➡️2-D position

The full list of visual attributes that we perceive pre-attentively is larger than the list above. These pre-attentive attributes, however, are the ones that are most useful to us when presenting data visually.

Some of these visual attributes are perceived quantitatively (i.e., some values are greater than others), which are marked with red arrows. The visual attributes that are marked with a pale red arrow are perceived quantitatively but not as powerfully as those with the bright red arrow.

Table Design

What about grids and rules in tables?

Product	Jan	Feb	Mar	Apr	May	Jun
Product 01	93,993	84,773	88,833	95,838	93,874	83,994
Product 02	87,413	78,839	82,615	89,129	87,303	78,114
Product 03	90,036	81,204	85,093	91,803	89,922	80,458
Product 04	92,737	83,640	87,646	94,557	92,620	82,872
Product 05	83,733	75,520	79,137	85,377	83,627	74,826
Total	447,913	403,976	423,323	456,705	447,346	400,264

Product	Jan	Feb	Mar	Apr	May	Jun
Product 01	93,993	84,773	88,833	95,838	93,874	83,994
Product 02	87,413	78,839	82,615	89,129	87,303	78,114
Product 03	90,036	81,204	85,093	91,803	89,922	80,458
Product 04	92,737	83,640	87,646	94,557	92,620	82,872
Product 05	83,733	75,520	79,137	85,377	83,627	74,826
Total	447,913	403,976	423,323	456,705	447,346	400,264

Product	Jan	Feb	Mar	Apr	May	Jun
Product 01	93,993	84,773	88,833	95,838	93,874	83,994
Product 02	87,413	78,839	82,615	89,129	87,303	78,114
Product 03	90,036	81,204	85,093	91,803	89,922	80,458
Product 04	92,737	83,640	87,646	94,557	92,620	82,872
Product 05	83,733	75,520	79,137	85,377	83,627	74,826
Total	447,913	403,976	423,323	456,705	447,346	400,264

Product	Jan	Feb	Mar	Apr	May	Jun
Product 01	93,993	84,773	88,833	95,838	93,874	83,994
Product 02	87,413	78,839	82,615	89,129	87,303	78,114
Product 03	90,036	81,204	85,093	91,803	89,922	80,458
Product 04	92,737	83,640	87,646	94,557	92,620	82,872
Product 05	83,733	75,520	79,137	85,377	83,627	74,826
Total	447,913	403,976	423,323	456,705	447,346	400,264

Product	Jan	Feb	Mar	Apr	May	Jun
Product 01	93,993	84,773	88,833	95,838	93,874	83,994
Product 02	87,413	78,839	82,615	89,129	87,303	78,114
Product 03	90,036	81,204	85,093	91,803	89,922	80,458
Product 04	92,737	83,640	87,646	94,557	92,620	82,872
Product 05	83,733	75,520	79,137	85,377	83,627	74,826
Total	447,913	403,976	423,323	456,705	447,346	400,264

Product	Jan	Feb	Mar	Apr	May	Jun
Product 01	93,993	84,773	88,833	95,838	93,874	83,994
Product 02	87,413	78,839	82,615	89,129	87,303	78,114
Product 03	90,036	81,204	85,093	91,803	89,922	80,458
Product 04	92,737	83,640	87,646	94,557	92,620	82,872
Product 05	83,733	75,520	79,137	85,377	83,627	74,826
Total	447,913	403,976	423,323	456,705	447,346	400,264

Tip: Use the least visible means to support the function of the non-data ink.

Using white space

Product	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Product 01	93,993	84,773	88,833	95,838	93,874	83,994	84,759	92,738	93,728	93,972	93,772	99,837
Product 02	87,413	78,839	82,615	89,129	87,303	78,114	78,826	86,246	87,167	87,394	87,208	92,848
Product 03	90,036	81,204	85,093	91,803	89,922	80,458	81,191	88,834	89,782	90,016	89,824	95,634
Product 04	92,737	83,640	87,646	94,557	92,620	82,872	83,626	91,499	92,476	92,716	92,519	98,503
Product 05	86,245	77,785	81,511	87,938	86,136	77,071	77,773	85,094	86,002	86,226	86,043	91,608
Product 06	88,833	80,119	83,956	90,576	88,720	79,383	80,106	87,647	88,582	88,813	88,624	94,356
Product 07	82,614	74,511	78,079	84,236	82,510	73,826	74,498	81,511	82,382	82,596	82,420	87,751
Product 08	85,093	76,746	80,421	86,763	84,985	76,041	76,733	83,957	84,853	85,074	84,893	90,384
Product 09	87,646	79,048	82,834	89,366	87,535	78,322	79,035	86,475	87,399	87,626	87,440	93,095
Product 10	90,275	81,420	85,319	92,047	90,161	80,672	81,406	89,070	90,021	90,255	90,063	95,888

Product	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Product 01	93,993	84,773	88,833	95,838	93,874	83,994	84,759	92,738	93,728	93,972	93,772	99,837
Product 02	87,413	78,839	82,615	89,129	87,303	78,114	78,826	86,246	87,167	87,394	87,208	92,848
Product 03	90,036	81,204	85,093	91,803	89,922	80,458	81,191	88,834	89,782	90,016	89,824	95,634
Product 04	92,737	83,640	87,646	94,557	92,620	82,872	83,626	91,499	92,476	92,716	92,519	98,503
Product 05	86,245	77,785	81,511	87,938	86,136	77,071	77,773	85,094	86,002	86,226	86,043	91,608
Product 06	88,833	80,119	83,956	90,576	88,720	79,383	80,106	87,647	88,582	88,813	88,624	94,356
Product 07	82,614	74,511	78,079	84,236	82,510	73,826	74,498	81,511	82,382	82,596	82,420	87,751
Product 08	85,093	76,746	80,421	86,763	84,985	76,041	76,733	83,957	84,853	85,074	84,893	90,384
Product 09	87,646	79,048	82,834	89,366	87,535	78,322	79,035	86,475	87,399	87,626	87,440	93,095
Product 10	90,275	81,420	85,319	92,047	90,161	80,672	81,406	89,070	90,021	90,255	90,063	95,888

What is the least visible means that you can use to delineate columns and rows or to group data? White space. In most cases, white space is all that is needed. When rows of data are so wide that it is difficult to track across a single row, you can increase the amount of white space, but should generally refrain from making it taller than the height than the row of text.

Using fill colors

To delineate rows or columns

Product	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Product 01	93,993	84,773	88,833	95,838	93,874	83,994	84,759	92,738	93,728	93,972	93,772	99,837
Product 02	87,413	78,839	82,615	89,129	87,303	78,114	78,826	86,246	87,167	87,394	87,208	92,848
Product 03	90,036	81,204	85,093	91,803	89,922	80,458	81,191	88,834	89,782	90,016	89,824	95,634
Product 04	92,737	83,640	87,646	94,557	92,620	82,872	83,626	91,499	92,476	92,716	92,519	98,503
Product 05	86,245	77,785	81,511	87,938	86,136	77,071	77,773	85,094	86,002	86,226	86,043	91,608
Product 06	88,833	80,119	83,956	90,576	88,720	79,383	80,106	87,647	88,582	88,813	88,624	94,356
Product 07	82,614	74,511	78,079	84,236	82,510	73,826	74,498	81,511	82,382	82,596	82,420	87,751
Product 08	85,093	76,746	80,421	86,763	84,985	76,041	76,733	83,957	84,853	85,074	84,893	90,384
Product 09	87,646	79,048	82,834	89,366	87,535	78,322	79,035	86,475	87,399	87,626	87,440	93,095
Product 10	90,275	81,420	85,319	92,047	90,161	80,672	81,406	89,070	90,021	90,255	90,063	95,888

To group sections

Product	Jan	Feb	Mar	Apr	May	Jun	Total
Product 01	93,993	84,773	88,833	95,838	93,874	83,994	541,305
Product 02	87,413	78,839	82,615	89,129	87,303	78,114	503,414
Product 03	90,036	81,204	85,093	91,803	89,922	80,458	518,516
Product 04	92,737	83,640	87,646	94,557	92,620	82,872	534,072
Product 05	83,733	75,520	79,137	85,377	83,627	74,826	482,220
Total	447,913	403,976	423,323	456,705	447,346	400,264	2,579,526

When you don't have enough room to adequately delineate rows through the use of white space alone, the subtle use of color works quite effectively.

Subtle shades of color can also be used to group sections of data separate from the rest.

Highlighting data

Product	Jan	Feb	Mar	Apr	May	Jun	Total
Product 01	93,993	84,773	88,833	95,838	93,874	83,994	541,305
Product 02	87,413	78,839	82,615	89,129	87,303	78,114	503,414
Product 03	90,036	81,204	85,093	91,803	89,922	80,458	518,516
Product 04	92,737	83,640	87,646	94,557	92,620	82,872	534,072
Product 05	83,733	75,520	79,137	85,377	83,627	74,826	482,220
Total	447,913	403,976	423,323	456,705	447,346	400,264	2,579,526

Border (enclosure)

Product	Jan	Feb	Mar	Apr	May	Jun	Total
Product 01	93,993	84,773	88,833	95,838	93,874	83,994	541,305
Product 02	87,413	78,839	82,615	89,129	87,303	78,114	503,414
Product 03	90,036	81,204	85,093	91,803	89,922	80,458	518,516
Product 04	92,737	83,640	87,646	94,557	92,620	82,872	534,072
Product 05	83,733	75,520	79,137	85,377	83,627	74,826	482,220
Total	447,913	403,976	423,323	456,705	447,346	400,264	2,579,526

Fill color (contrasting hue)

Product	Jan	Feb	Mar	Apr	May	Jun	Total
Product 01	93,993	84,773	88,833	95,838	93,874	83,994	541,305
Product 02	87,413	78,839	82,615	89,129	87,303	78,114	503,414
Product 03	90,036	81,204	85,093	91,803	89,922	80,458	518,516
Product 04	92,737	83,640	87,646	94,557	92,620	82,872	534,072
Product 05	83,733	75,520	79,137	85,377	83,627	74,826	482,220
Total	447,913	403,976	423,323	456,705	447,346	400,264	2,579,526

Boldfaced text (greater line width)

Product	Jan	Feb	Mar	Apr	May	Jun	Total
Product 01	93,993	84,773	**88,833**	95,838	93,874	83,994	541,305
Product 02	87,413	78,839	**82,615**	89,129	87,303	78,114	503,414
Product 03	90,036	81,204	**85,093**	91,803	89,922	80,458	518,516
Product 04	92,737	83,640	**87,646**	94,557	92,620	82,872	534,072
Product 05	83,733	75,520	**79,137**	85,377	83,627	74,826	482,220
Total	447,913	403,976	**423,323**	456,705	447,346	400,264	2,579,526

Colored text (contrasting hue)

Product	Jan	Feb	Mar	Apr	May	Jun	Total
Product 01	93,993	84,773	88,833	95,838	93,874	83,994	541,305
Product 02	87,413	78,839	82,615	89,129	87,303	78,114	503,414
Product 03	90,036	81,204	85,093	91,803	89,922	80,458	518,516
Product 04	92,737	83,640	87,646	94,557	92,620	82,872	534,072
Product 05	83,733	75,520	79,137	85,377	83,627	74,826	482,220
Total	447,913	403,976	423,323	456,705	447,346	400,264	2,579,526

The best means to highlight sections of data in tables involves using one of these four visual attributes. The power of a border to group and highlight data was first recognized by the Gestalt school of psychology, well known for its many breakthroughs in the understanding of visual perception. The Gestalt principle of enclosure asserts that any means of enclosing something visually, such as a surround box or circle or a background fill color, serves as a powerful means to set it apart from the rest.

Aligning numbers

Sales	Sales	Sales
93,883.39	93,883.39	93,883.39
5,693,762.32	5,693,762.32	5,693,762.32
483.84	483.84	483.84
674,663.39	674,663.39	674,663.39
548.93	548.93	548.93
3,847.33	3,847.33	3,847.33
$6,467,189.20	$6,467,189.20	$6,467,189.20

Numbers should always be right aligned to the decimal point or to the right-most digit when there is no decimal point. When parentheses are used to identify negative numbers, the numbers should still be right aligned to the decimal point or to the right-most digit when there is no decimal point. When displaying percentages, it is best to always include the percentage sign (%) and to make the number of decimal places consistent for all the numbers in a given column so that the decimal points are right aligned.

Aligning text

Product Code	Product Name	Product Code	Product Name	Product Code	Product Name
A1838	2-Door Sport	A1838	2-Door Sport	A1838	2-Door Sport
A89	4-Door Sport	A89	4-Door Sport	A89	4-Door Sport
J98488	2-Door Luxury	J98488	2-Door Luxury	J98488	2-Door Luxury
J3883	4-Door Luxury	J3883	4-Door Luxury	J3883	4-Door Luxury
K9288	2-Door Truck	K9288	2-Door Truck	K9288	2-Door Truck
K38733	4-Door Truck	K38733	4-Door Truck	K38733	4-Door Truck

Cust Code	Preferred?	Cust Code	Preferred?
193847394	Y	193847394	Y
109388484	N	109388484	N
187466463	N	187466463	N
174563553	N	174563553	N
175357736	Y	175357736	Y
167374565	Y	167374565	Y

Text should be left aligned. Text includes numbers that do not represent quantitative data, such as the product and customer codes in these examples. The only exception to the left-alignment rule is when the text values in a given column all contain the same number of characters and that number of characters is considerably less than the number of characters in that column's label. In this case it works best to center the text, which causes it to be separated from the column to its left by more white space.

Displaying dates

Date	Date	Date	Date
12/17/02	12/17/02	12/17/02	12/17/02
1/2/03	01/02/03	1/2/03	01/02/03
1/17/03	01/17/03	1/17/03	01/17/03
2/9/03	02/09/03	2/9/03	02/09/03
10/29/03	10/29/03	10/29/03	10/29/03
12/1/03	12/01/03	12/1/03	12/01/03
1/1/03	01/01/03	1/1/03	01/01/03

Whether dates are aligned to the left, right, or center is not important. What is important is that all date that appear in the same column contain a consistent number of digits to express the month (that is, two), a consistent number of digits to express the day (that is, two), and a consistent number of digits to express the year (that is, either two or four). This will cause the same parts of each date (that is, the month, day, or year parts) to be aligned with one another from one row to the next.

Using descriptive text

2003 Bookings Revenue by Product
as of July 1, 2003
(U.S. Dollars)

Product	Jan	Feb	Mar	Q1 Mo Avg	Apr	May	Jun	Q2 Mo Avg
Product 01	93,993	84,773	88,833	89,200	95,838	93,874	83,994	91,235
Product 02	87,413	78,839	82,615	82,956	89,129	87,303	78,114	84,849
Product 03	90,036	81,204	85,093	85,444	91,803	89,922	80,458	87,394
Product 04	92,737	83,640	87,646	88,008	94,557	92,620	82,872	90,016
Product 05	86,245	77,785	81,511	81,847	87,938	86,136	77,071	83,715
Product 06	88,833	80,119	83,956	84,303	90,576	88,720	79,383	86,226
Product 07	82,614	74,511	78,079	78,401	84,236	82,510	73,826	80,191
Product 08	85,093	76,746	80,421	80,753	86,763	84,985	76,041	82,596
Product 09	87,646	79,048	82,834	83,176	89,366	87,535	78,322	85,074
Product 10	90,275	81,420	85,319	85,671	92,047	90,161	80,672	87,626
Total	$884,886	$798,085	$836,307	$839,759	$902,255	$883,765	$790,751	$858,924

Please note that the Q1 and Q2 columns display average monthly bookings revenue, not total revenue for the quarter.

(Source: This data was extracted from the Enterprise Data Warehouse and has not been altered, except for rounding the summary figures that appear in this table to whole dollars. Questions may be directed to the Enterprise Reporting Group by emailing erg@company.com or calling (888) 555-3923.)

A seemingly small point which deserves special mention is that one should get in the habit of making all charts, tables, reports, and problems complete in themselves. It should be possible to learn from the sheet, six months after its completion, what the problem was all about, who gathered the data and calculated it, who drew the graph, what were the important conditions at hand, what were the sample size and the units, the date, etc. Only in this way can we make sure that our records will not grow "cold."

I. W. Burr, Professor (1941 – 1974), Purdue Univers

Tables and graphs should always include the following descriptive text:
- Titles that describe the content
- Unit of measure
- Date range and the date the data was collected
- Notes when explanations are needed
- Contact information

2007-2008 State Summary of Education Progress Report

Louisiana Department of EDUCATION

Districts	Student Characteristics											
	Population [2]		Students by Gender and Ethnicity							English Proficiency		Poverty
	Number of Schools	Number of Students	Percent Female	Percent Male	Percent American Indian	Percent Asian	Percent Black	Percent Hispanic	Percent White	Percent of LEP [3] Students	Percent of Non-LEP Students	Percent of Students Eligible for Free & Reduced Price Lunch [4]
	2007-2008	2007-2008	2007-2008	2007-2008	2007-2008	2007-2008	2007-2008	2007-2008	2007-2008	2007-2008	2007-2008	2007-2008
Louisiana [1]	1,471	681,038	48.9	51.1	0.6	1.4	46.0	2.7	49.2	1.7	98.3	63.2

All

		ELA Performance Level				ELA Proficiency Status		
		1	2	3	4	Below Standard	At/Above Standard	All
Grade								
3	Number	121	400	1093	130	521	1223	1744
	Percent	6.9	22.9	62.7	7.5	29.9	70.1	100.0
4	Number	97	462	1092	71	559	1163	1722
	Percent	5.6	26.8	63.4	4.1	32.5	67.5	100.0
5	Number	15	438	1086	128	453	1214	1667
	Percent	0.9	26.3	65.1	7.7	27.2	72.8	100.0
6	Number	3	571	1019	64	574	1083	1657
	Percent	0.2	34.5	61.5	3.9	34.6	65.4	100.0
7	Number	9	590	1043	39	599	1082	1681
	Percent	0.5	35.1	62.0	2.3	35.6	64.4	100.0
8	Number	52	774	820	21	826	841	1667
	Percent	3.1	46.4	49.2	1.3	49.6	50.4	100.0
All	Number	297	3235	6153	453	3532	6606	10138
	Percent	2.9	31.9	60.7	4.5	34.8	65.2	100.0

Graph
Design

How 2-D X and Y axes graphs work
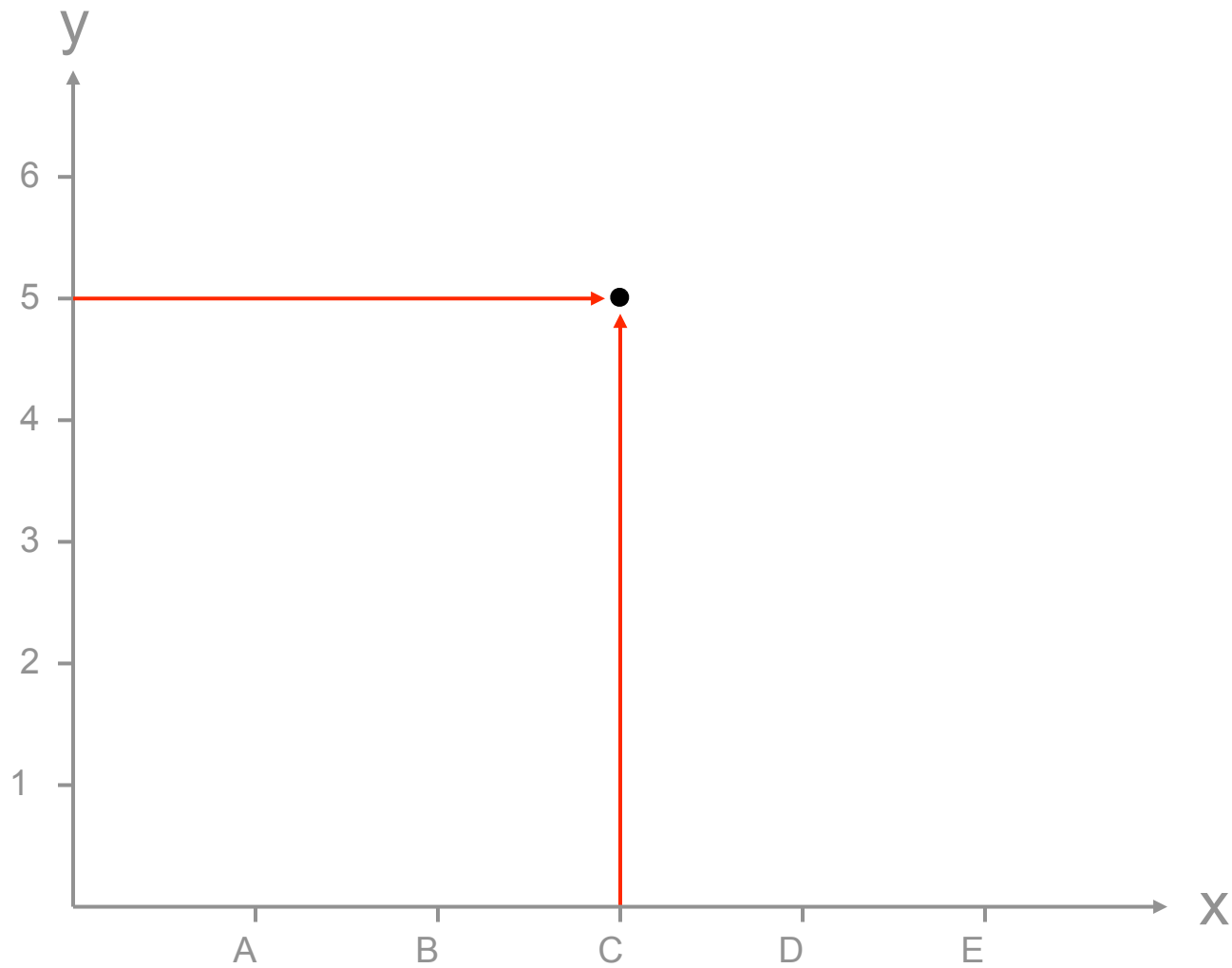
Most 2-D graphs work by means of a system of coordinates, encoding each value as a coordinate along two perpendicular axes, X (horizontal) and Y (vertical). Rene Descartes, the 17th century mathematician and philosopher ("I think, therefore I am.") invented this XY coordinate system for use in mathematics, not for communicating quantitative data.

Values can be encoded using four objects.

Points
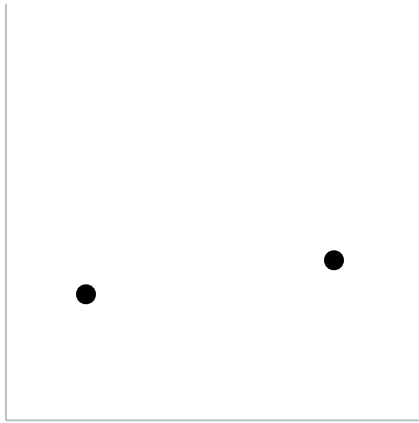
Bars
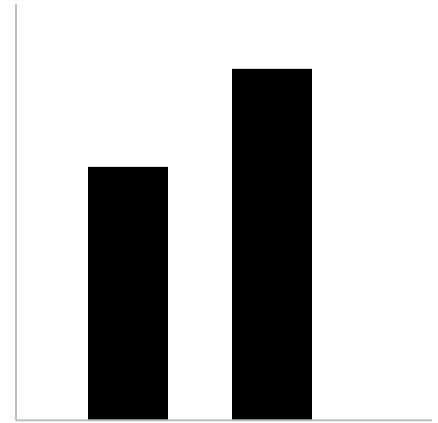
Lines
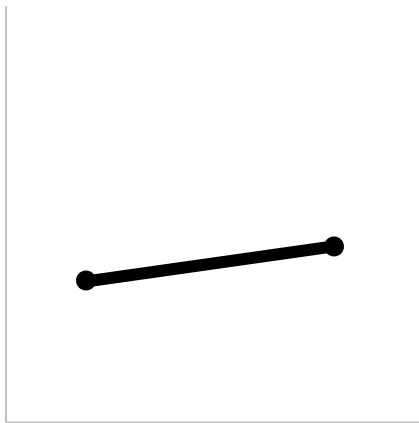
Boxes
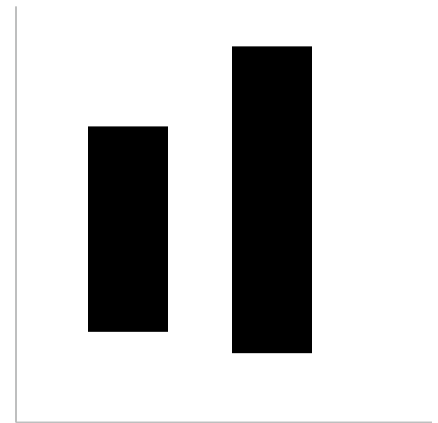
Four different objects can be used to visually encode data in graphs:

- The simplest is a point, which for all practical purposes has zero dimensions, for it marks a point in space, but has neither height nor width.
- The second is a line, which can be thought of as a point that has been extended in a particular direction to become a one-dimensional object.
- The third is a rectangle, or what we call a bar when it is used in graphs, which you can think of as a line to which the additional dimension of width has been added to turn it into a two-dimensional object.
- The last and probably least familiar is a box, which is just like a bar, except that both ends are used to mark a value in the graph.

Data encoding objects: Points
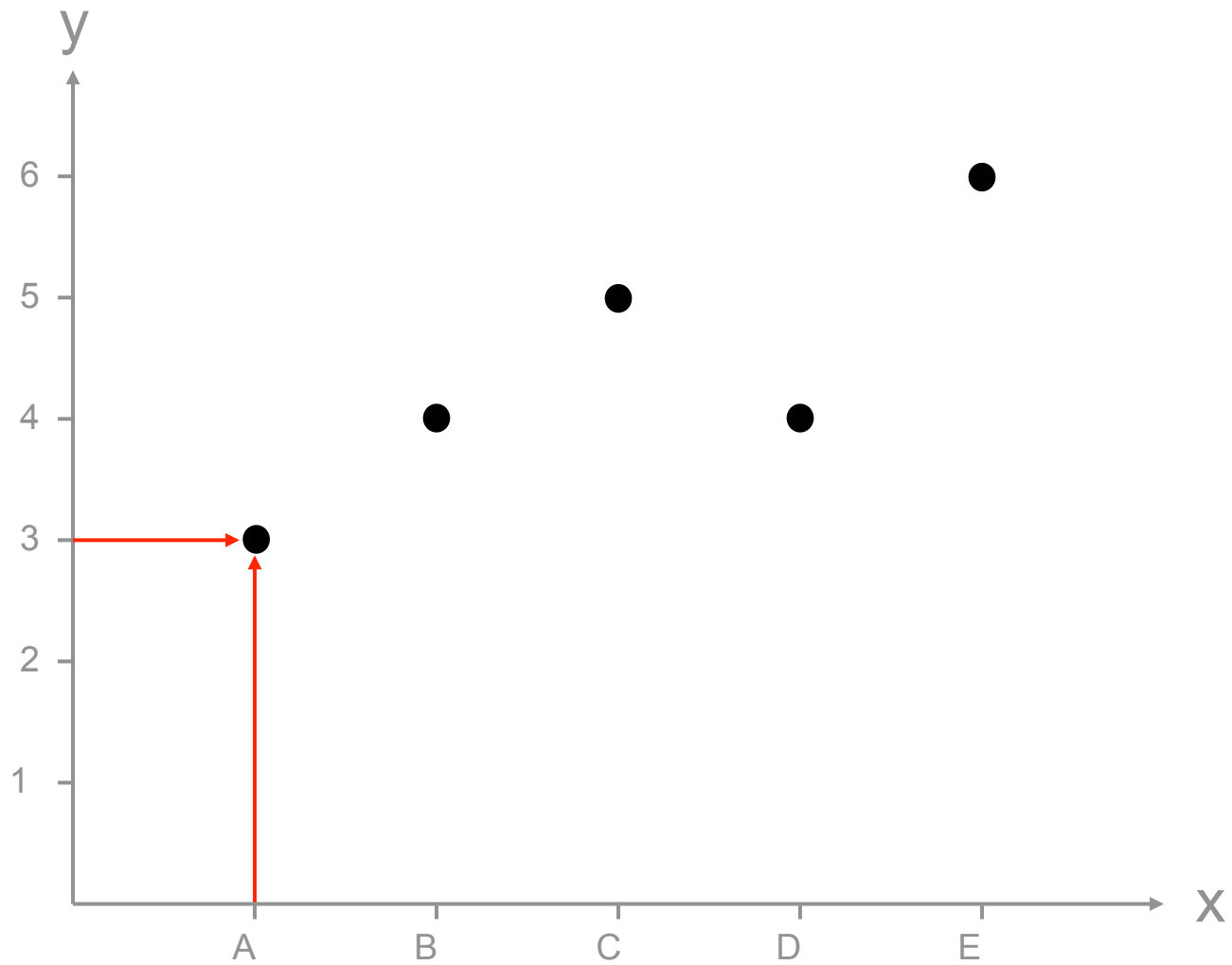
Points encode individual values as 2-D position.

Data encoding objects: Lines
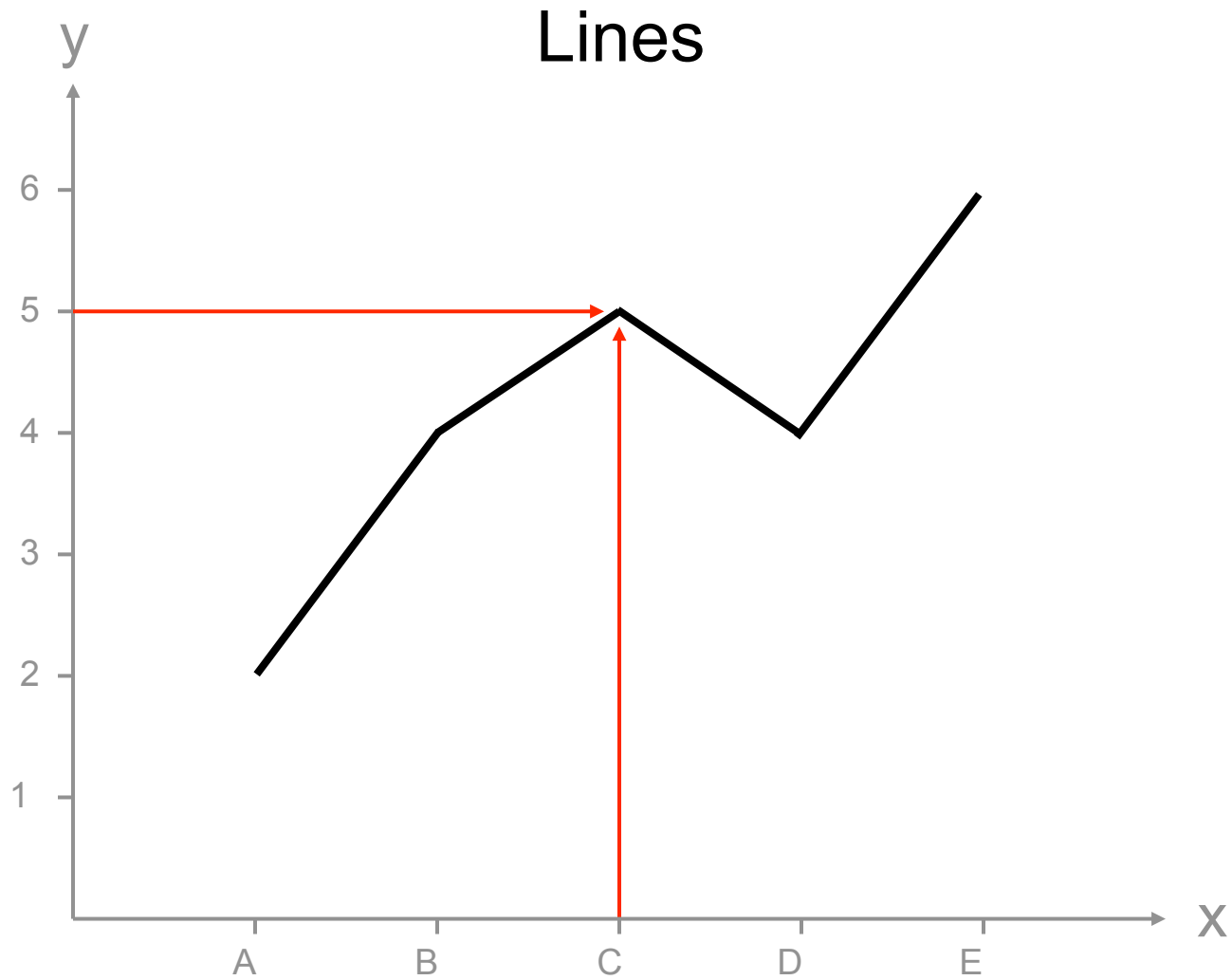
Lines

Lines encode individual values as 2-D position (at each data point connected by the line), but by connecting the data values the added characteristics of slope and direction also carry information.

Data encoding objects: Bars
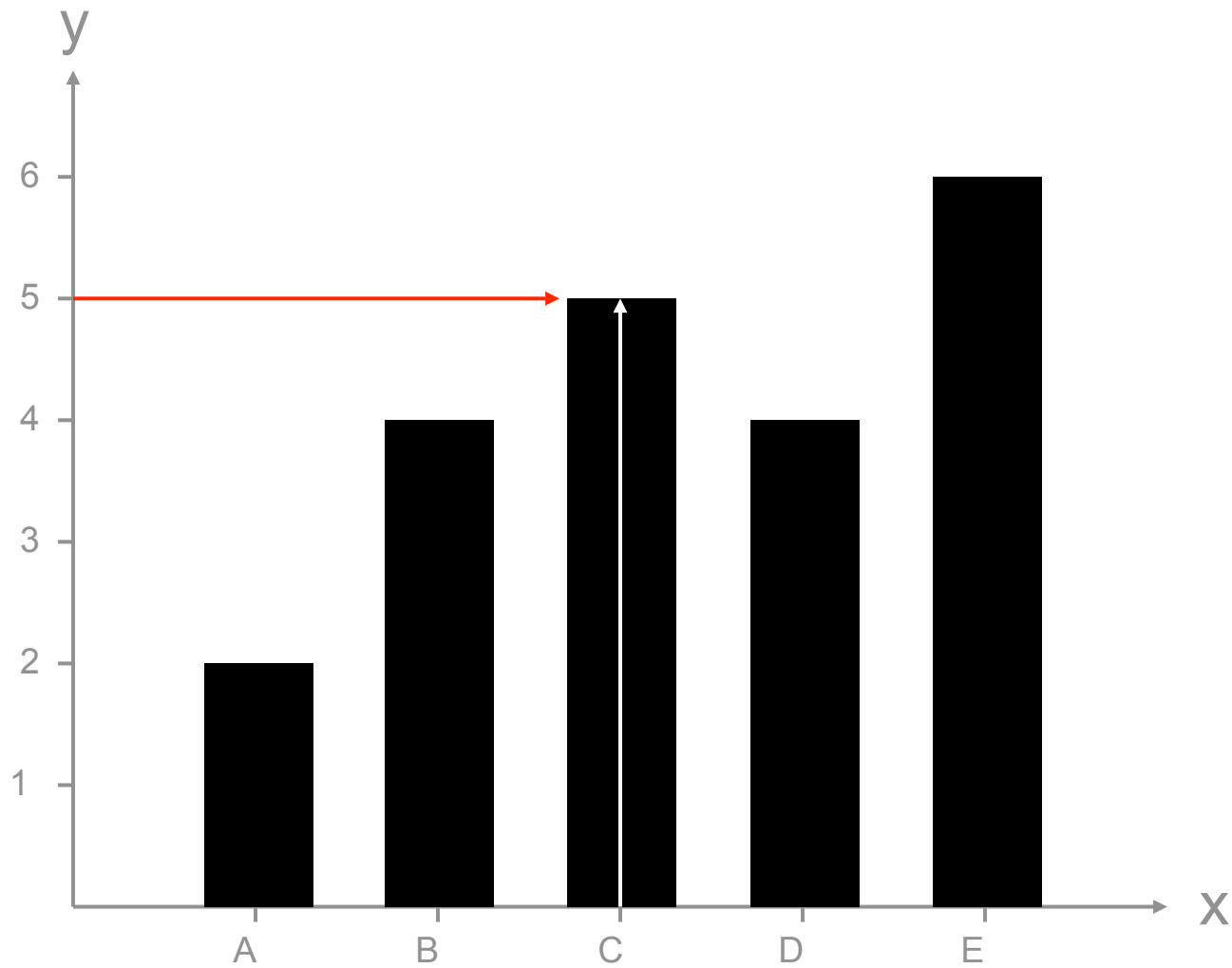
Bars encode individual values as 2-D position at the endpoint of the bar, and also as height (vertical bars) or length (horizontal bars).

Data encoding objects: Boxes
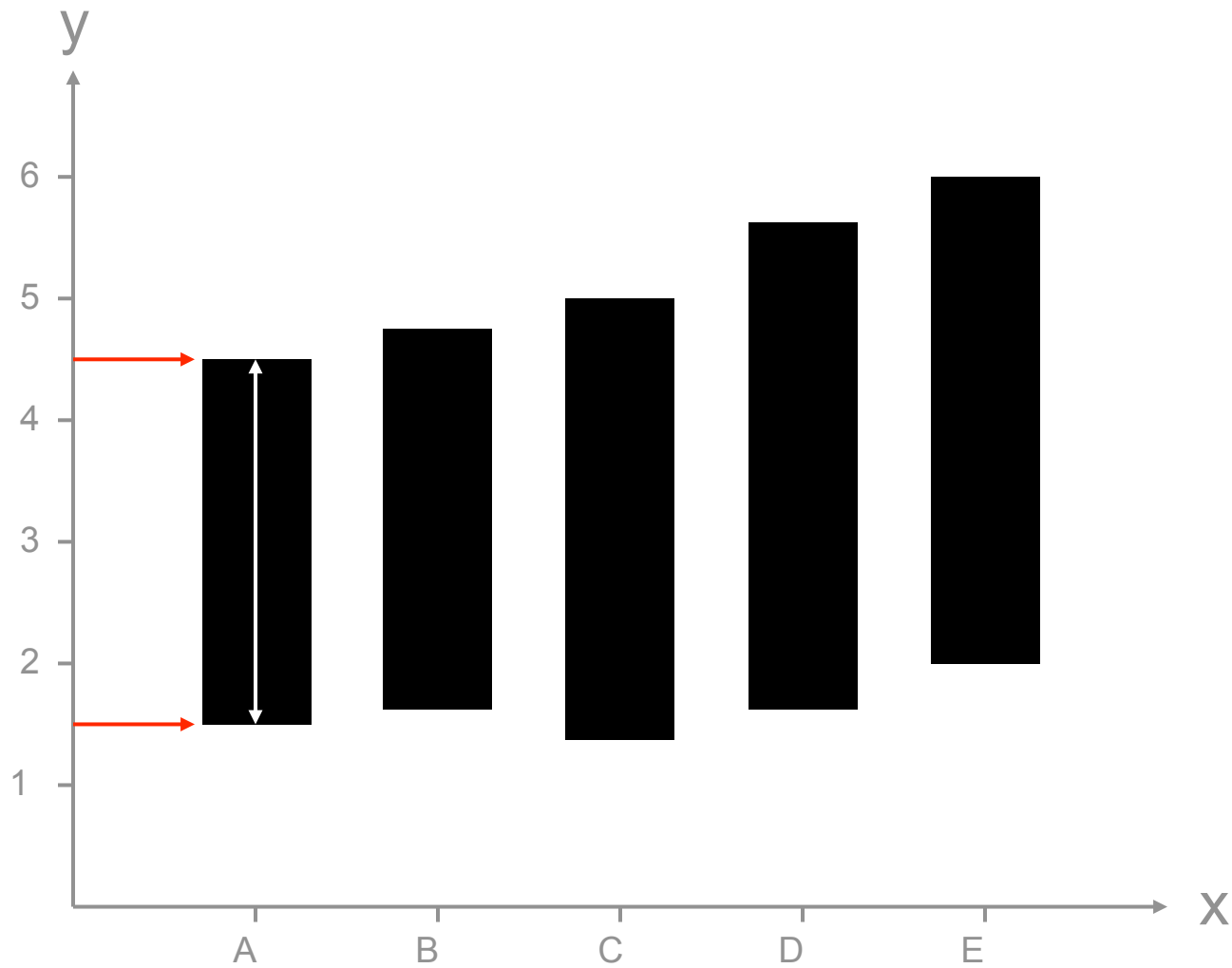
Boxes look and mostly function like bars, except that they encode two values, rather than one. They represent a range of values by using the one end of the box to encode the lowest value in the range and the other end to encode the highest value. Like bars, boxes encode values both as 2-D location based on the position of each end, and as line length based on the length of the box.

Useful variations of data encoding objects
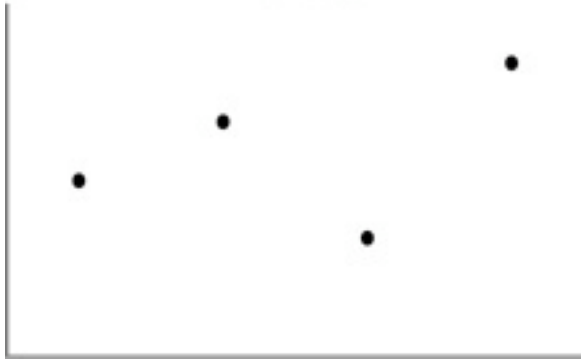
Points

Lines
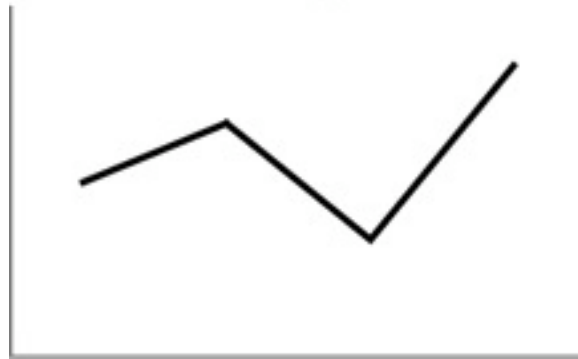
Points and Lines

Vertical Bars
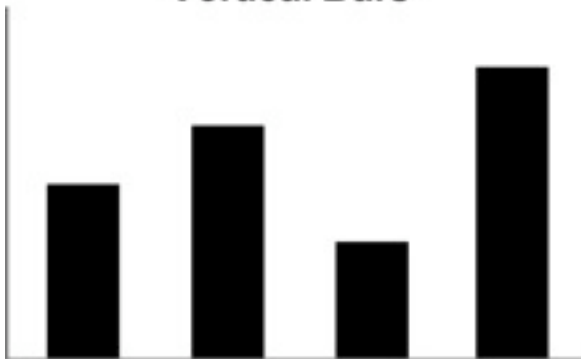
Horizontal Bars

Boxes (Points and Bars)

Six variations of these four data encoding objects work well in XY graphs. Each has its own strengths and weaknesses.

What meanings do these bars suggest?
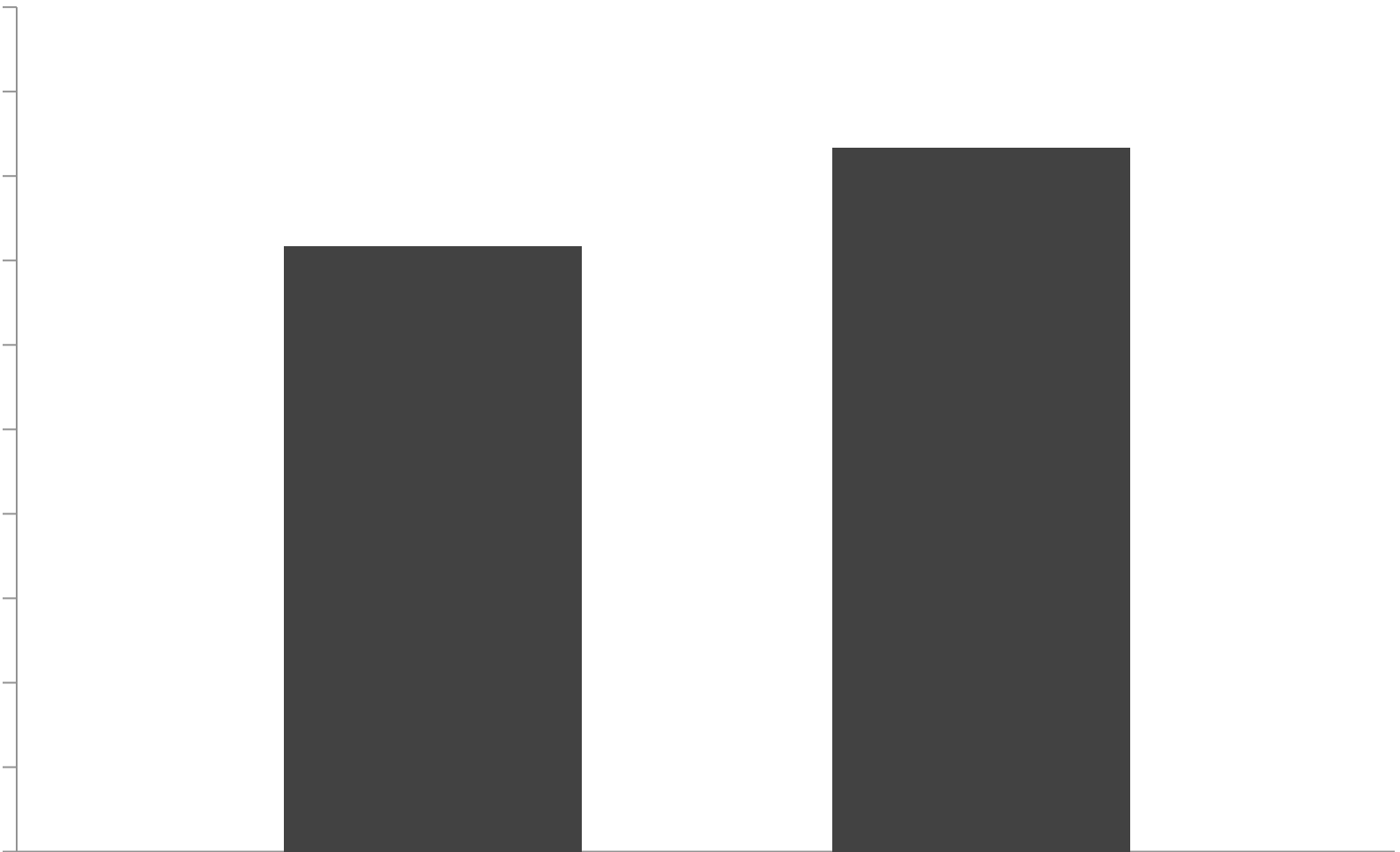
When we see two objects like these bars, arranged in this manner, what stands out is the difference in their heights, which encourages us to compare their heights and notice that one is taller than the other. This is what bars are good at: displaying differences in magnitudes and making it easy for us to compare these differences.

The strengths of bars

Q1 2003 Planned vs. Actual Sales by Region
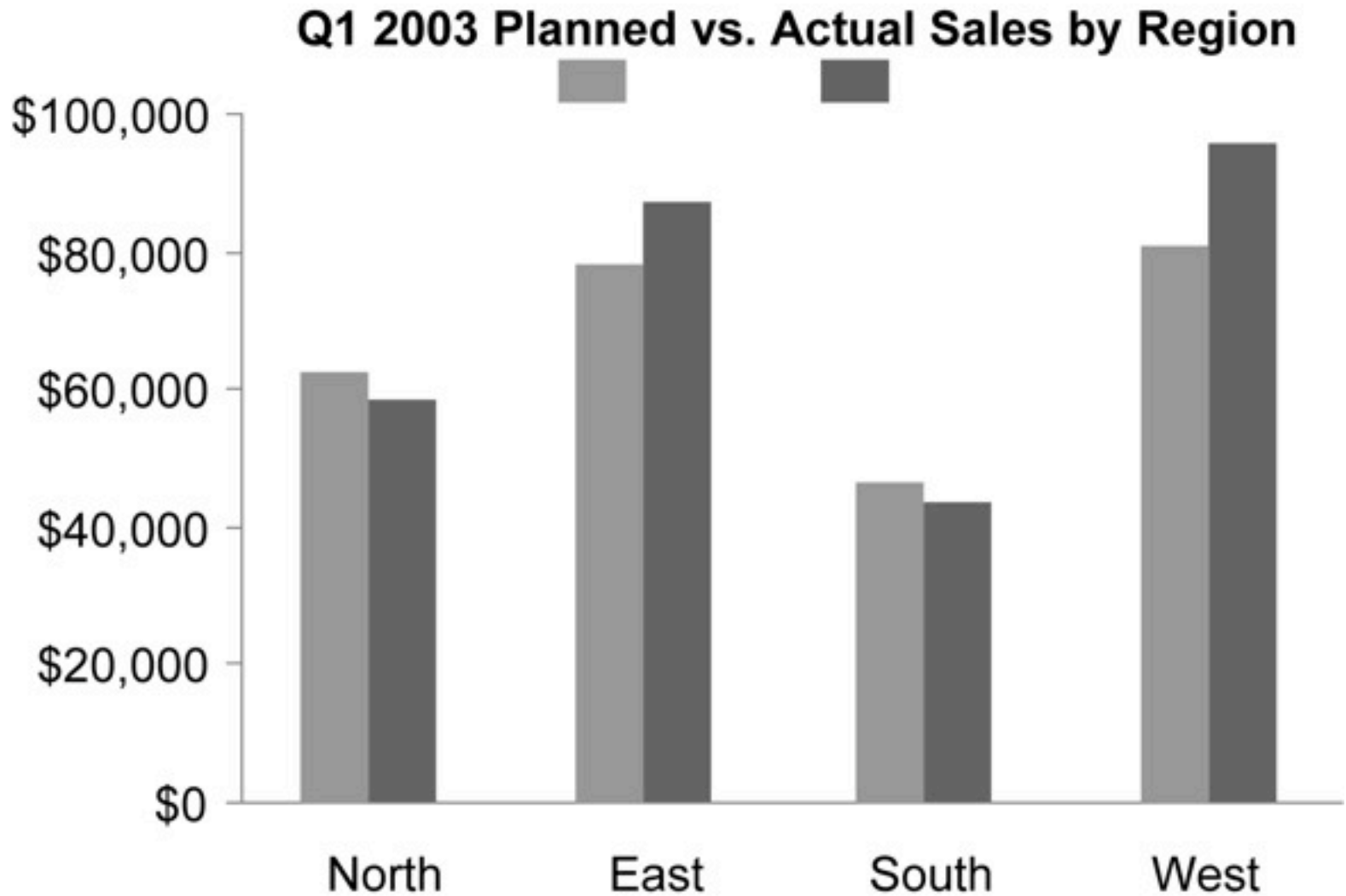
The strength of bars, because they have such great visual weight, like great columns rising into the sky, is their ability to emphasize individual discrete values. Because bars encode quantitative values in part as line length, they must always begin at a value of zero, otherwise their length does not correspond accurately to their quantitative value.

What's wrong with this bar graph?

In a recent industry survey that measured customer loyalty, MicroStrategy outscored all of the competition.

		60%	70%	80%	90%
1	MicroStrategy				
2	Applix TM1				
3	SAP BW				
4	Microsoft AS				
5	MIS Alea				
6	Oracle OLAP Servers				
7	Business Objects				
8	Hyperion Essbase				
9	Oracle Discoverer				
10	Cognos PowerPlay				

Something is terribly wrong with this graph. Can you see the problem?

Bars should always begin at zero.
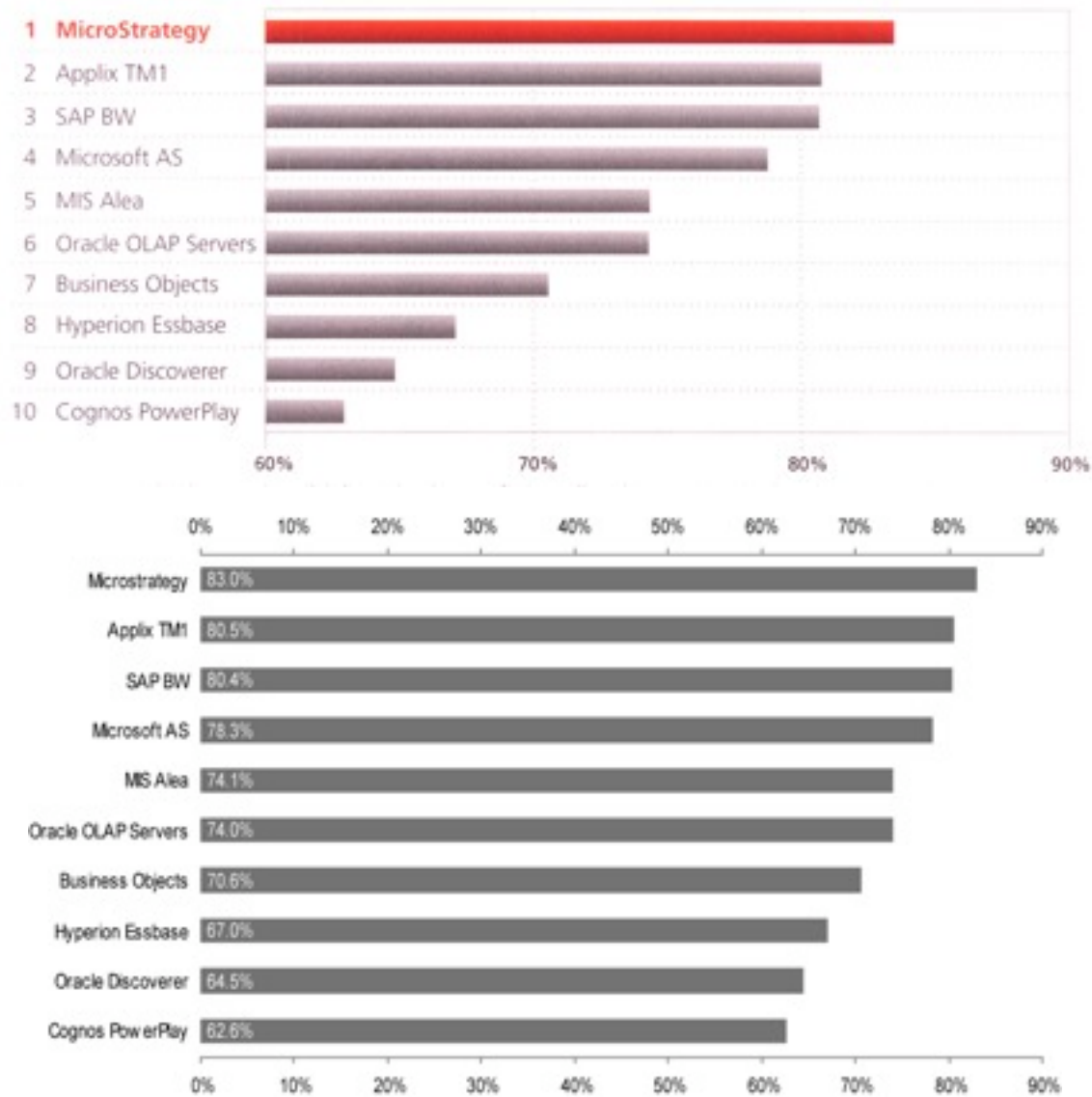
Because bars encode quantitative values in part as line length, they must always begin at a value of zero, otherwise their length does not correspond accurately to their quantitative value.

MicroStrategy's customer service score is only 33% greater than the lowest scoring product, Cognos PowerPlay, but the difference in the lengths of these two bars graph is 650%. That's a lie factor of 617% (650% / 33% = 617%). Notice how different the data looks when encoded accurately in the bottom graph.

What meanings does this line suggest?
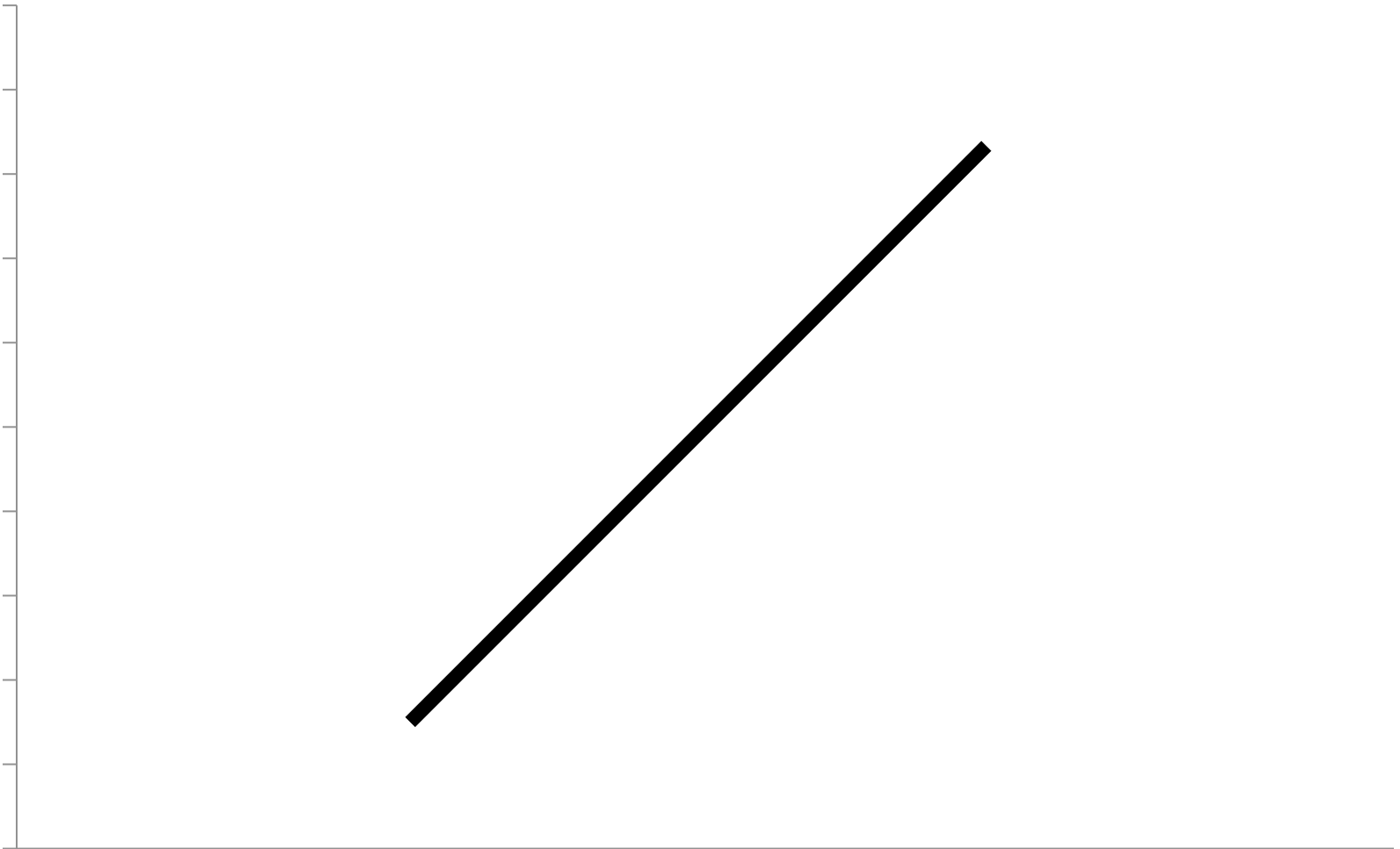
When we see a line – in this case one that is angled upwards from left to right – we are inclined to interpret it as increase, something moving upwards. This is precisely what lines are good at. They do a great job of showing the shape of change from one value to the next, such as change through time.

The strengths of lines

U.S. Dollars — 2002 Sales. Line chart showing monthly sales values from January through December, ranging from about 2,000 in Jan rising with fluctuations to about 3,500 in Dec.
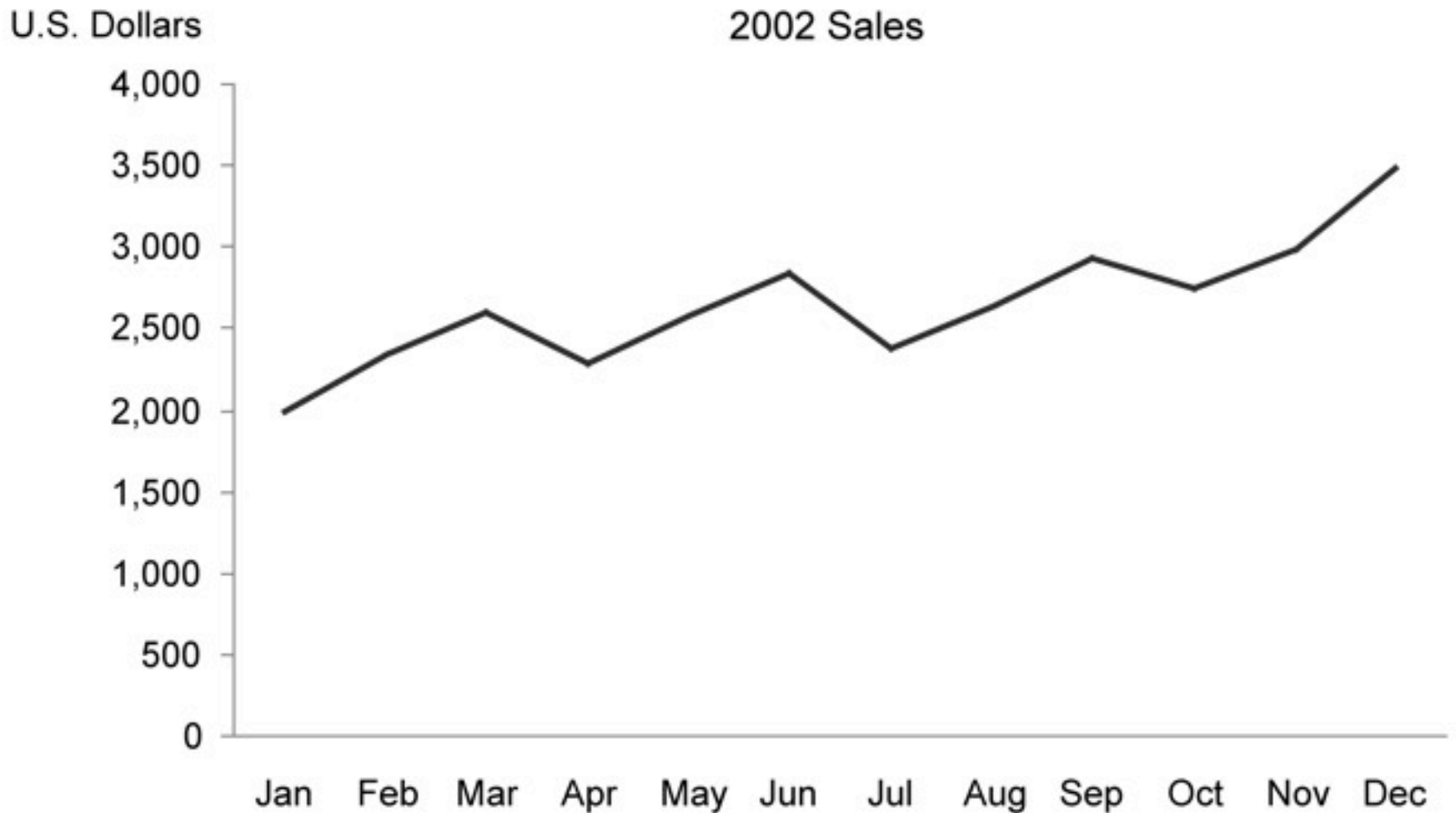
The strength of lines is their ability to emphasize the overall trend of the values and the nature of change from one value to the next. They should only be used to encode continuous variables along an interval scale, never discrete variables. The most common examples of continuous variables are those corresponding to a time series (continuous units of time) or a frequency distribution (contiguous ranges of quantity, such as 0-5, 6-10, 11-15, and so on).

Misuse of lines

Q1 2003 <u>Planned</u> vs. <u>Actual</u> Sales by Region
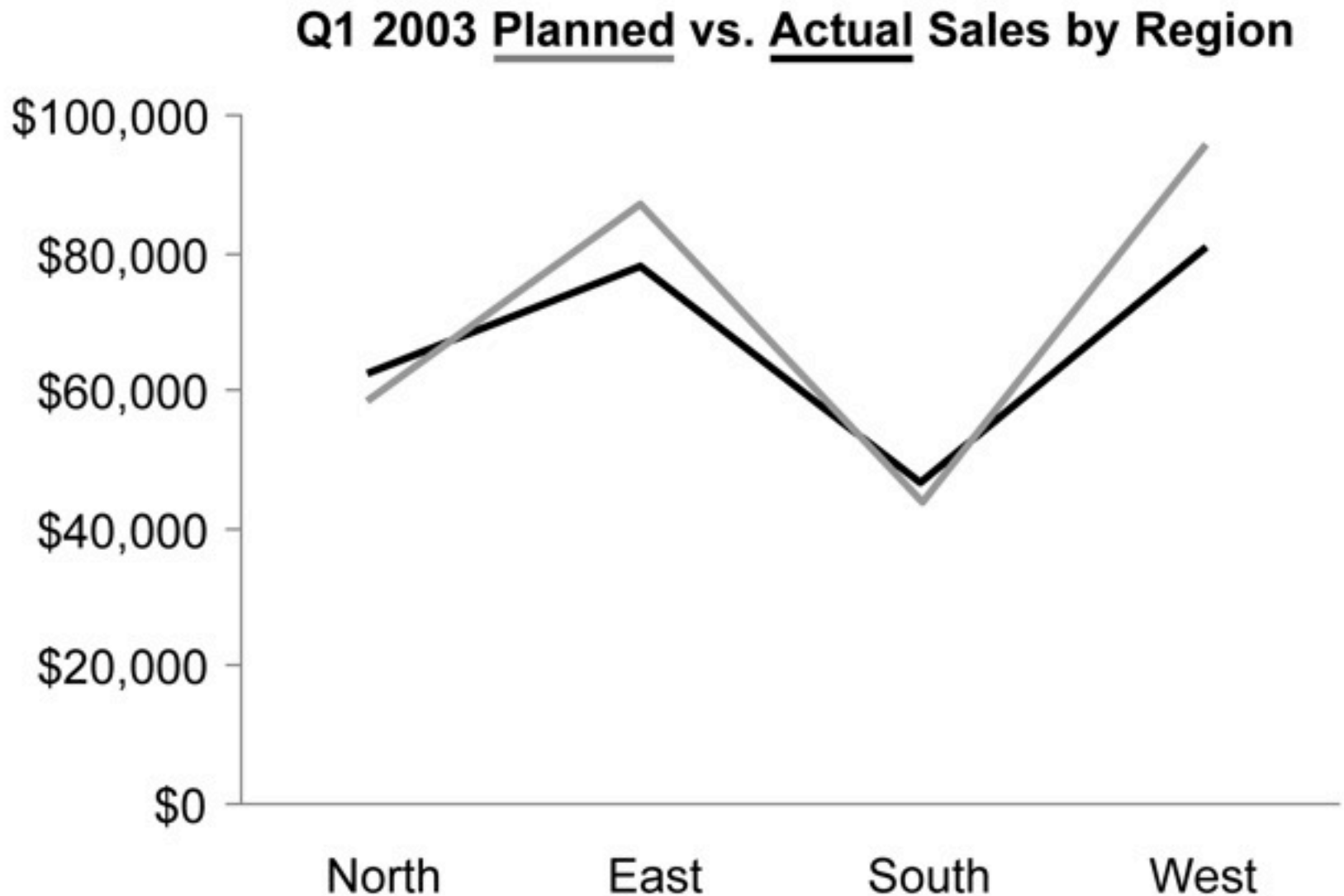
It is never appropriate to use lines to connect values along a nominal scale, as shown above. Doing so suggests a connectedness in the data that doesn't exist, and presents slopes and patterns that have no meaning. It is also not appropriate to use lines to connect values along an ordinal scale, except to encode a value that is cumulative from one item to the next along the ordinal scale. This works, because cumulative values along an ordinal scale are intimately connected from one to the next, not only in their order, which in itself would not warrant the use of lines, but also in that each new value incorporates the sum of all those that came before it.

What's wrong with this graph?

Map key:
Cities visited

- 1979-1980
- 1981-1985
- 1986-1990
- 1991-1995
- 1996-2000
- 2001-2004

Total number of cities visited

Horizontal axis: 1979-80, 1981-85, 1986-90, 1991-95, 1996-00, 2000-04
Vertical axis: 0, 50, 100, 150, 200
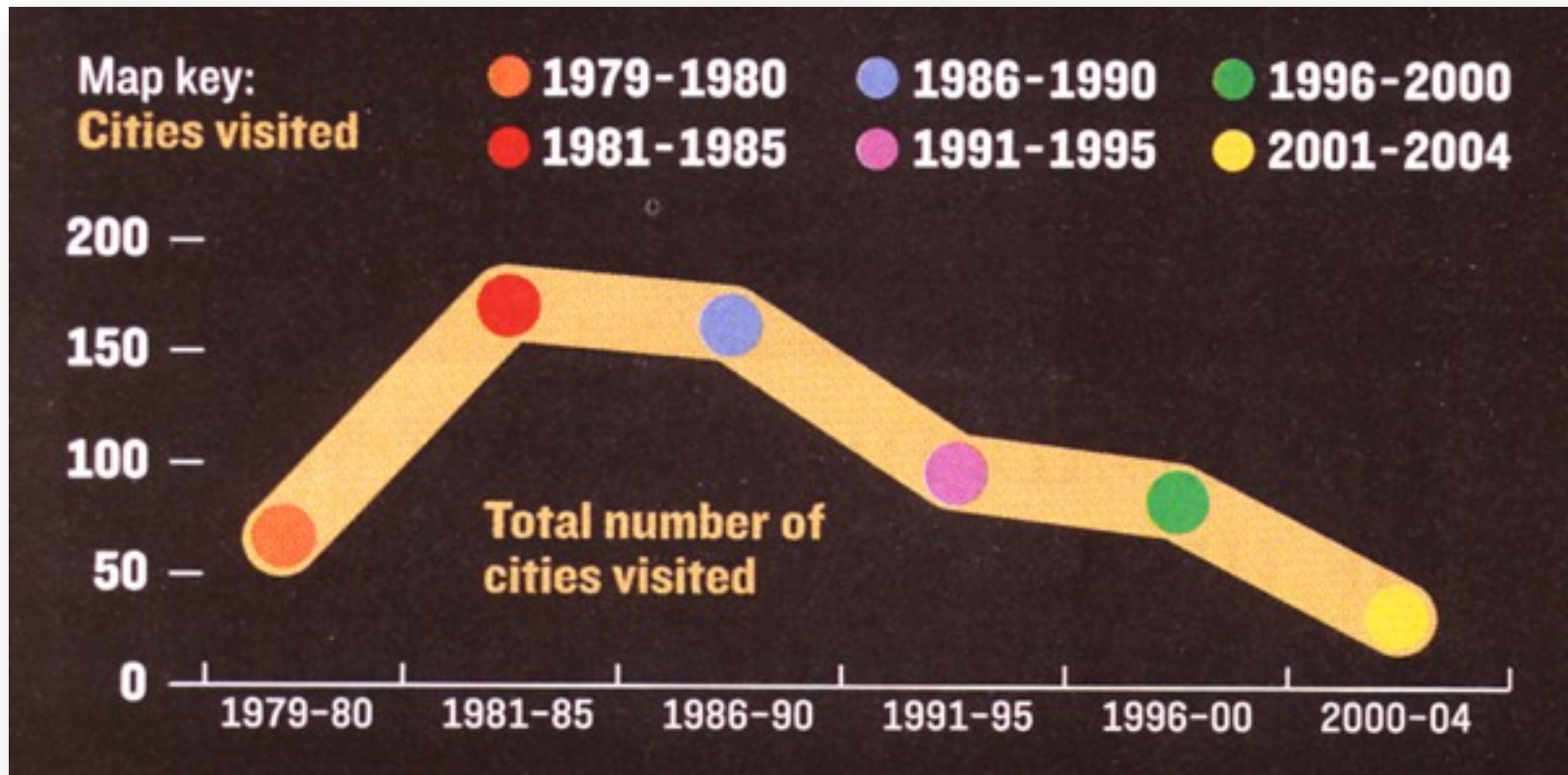
This graph appeared in the November 11, 2005 issue of Newsweek magazine. It displays the frequency of travel done by Pope John Paul II during his many years of service. The scale along the horizontal axis is an interval scale, but it has a problem: the intervals aren't equal. Notice that the first interval only covers two years, while each of the others covers five years. This causes the distribution to look as if the pope traveled relatively little during the first few years and then increase his travels dramatically in the next few. Another problem in this graph is the unnecessary dual labeling of the intervals, both along the horizontal axis and in the legend using color coding. This is not only unnecessary, it is also distracting.

What meanings do these points suggest?
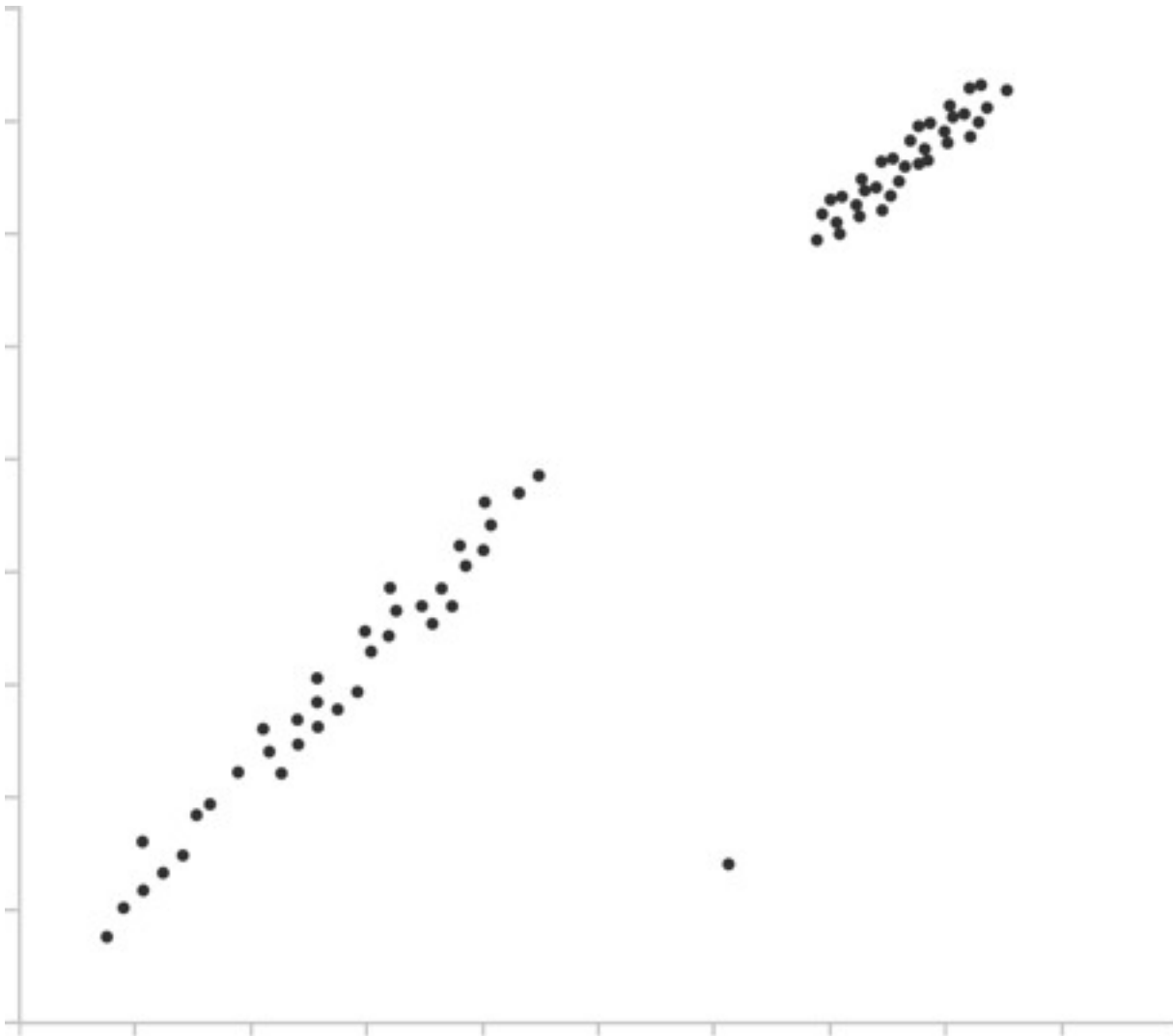
Data points like these, especially in a scatterplot, encourage us to notice such things as clusters of values, linear arrangements of values, gaps in values, and values that are much different from the norm (that is, exceptions or outliers). These are precisely the kinds of patterns that are meaningful in correlation relationships.

The strengths of points
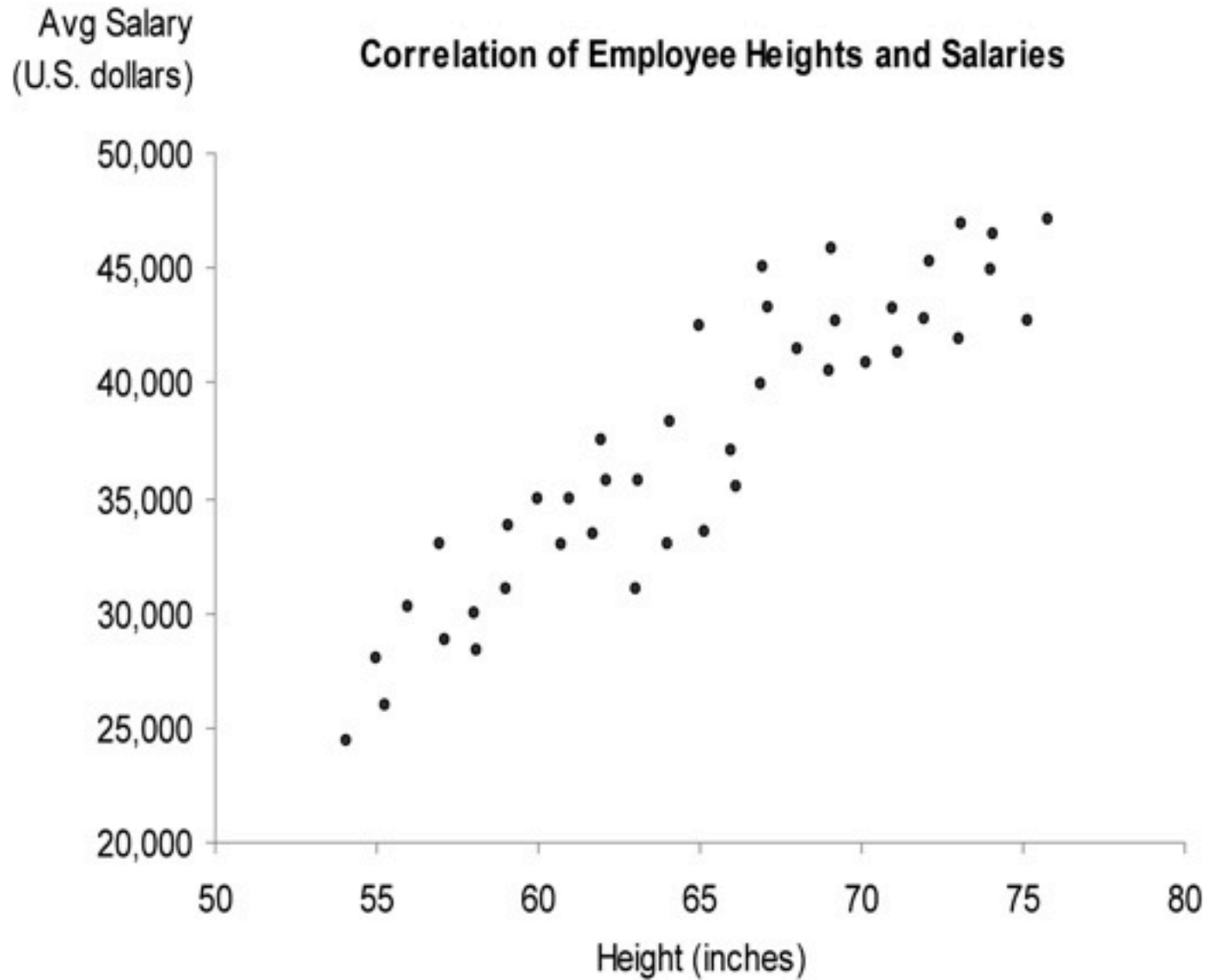
The unique strength of points is their ability to encode values along two quantitative scales aligned with two axes simultaneously.

Another strength of points
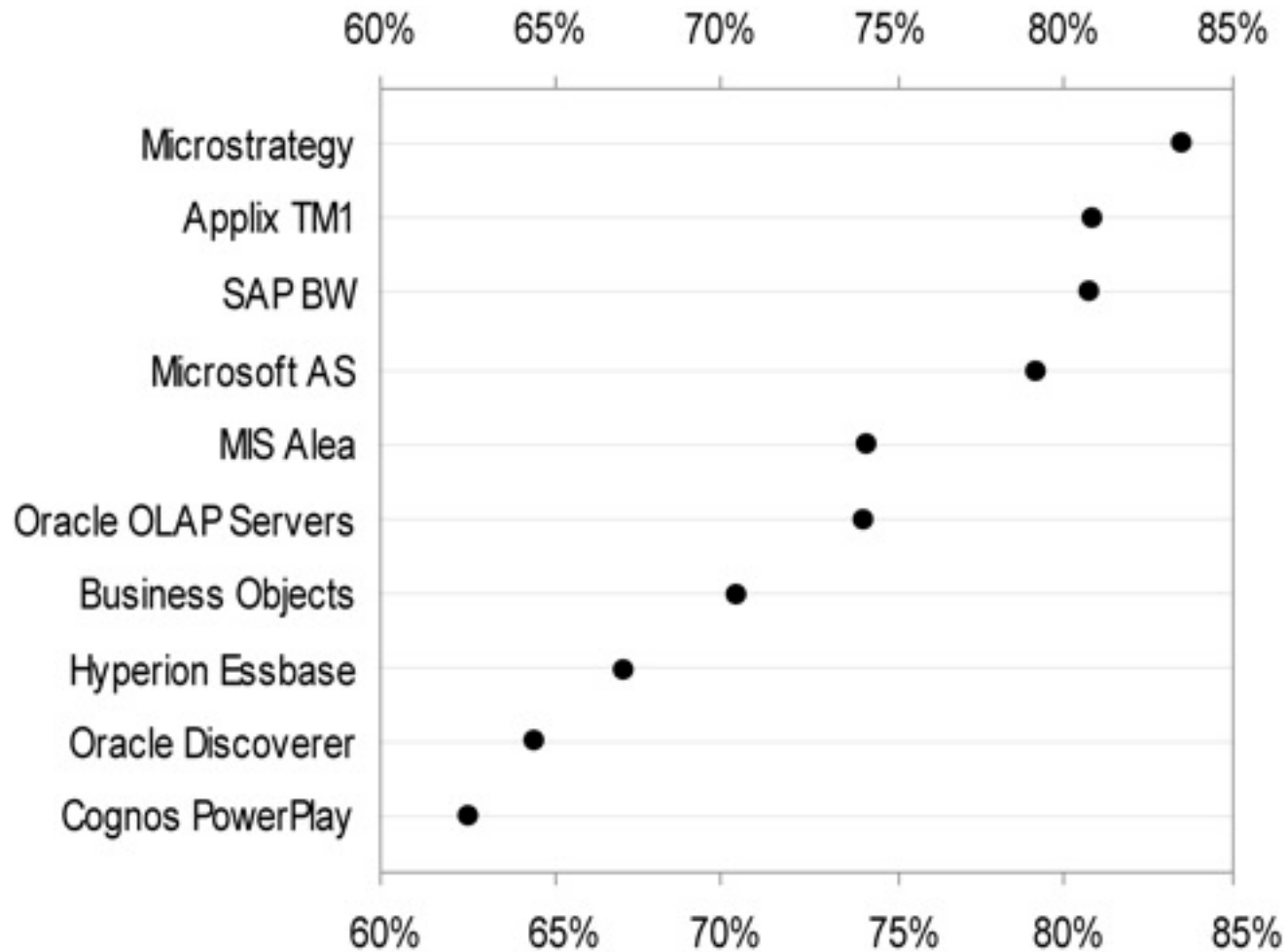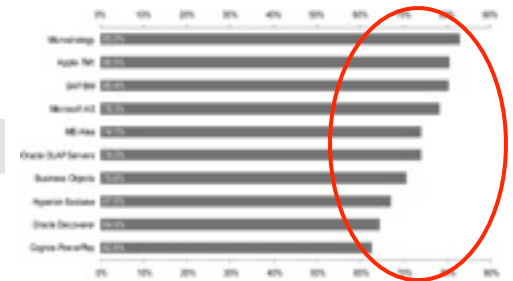
Points can also be used when you would normally use bars if there is a significant advantage to narrowing the quantitative scale such that zero is not included. When bars are used, the quantitative scale must include zero as the base for the bars, because otherwise the lengths of the bars would not accurately encode their values.

The strength of points and lines together
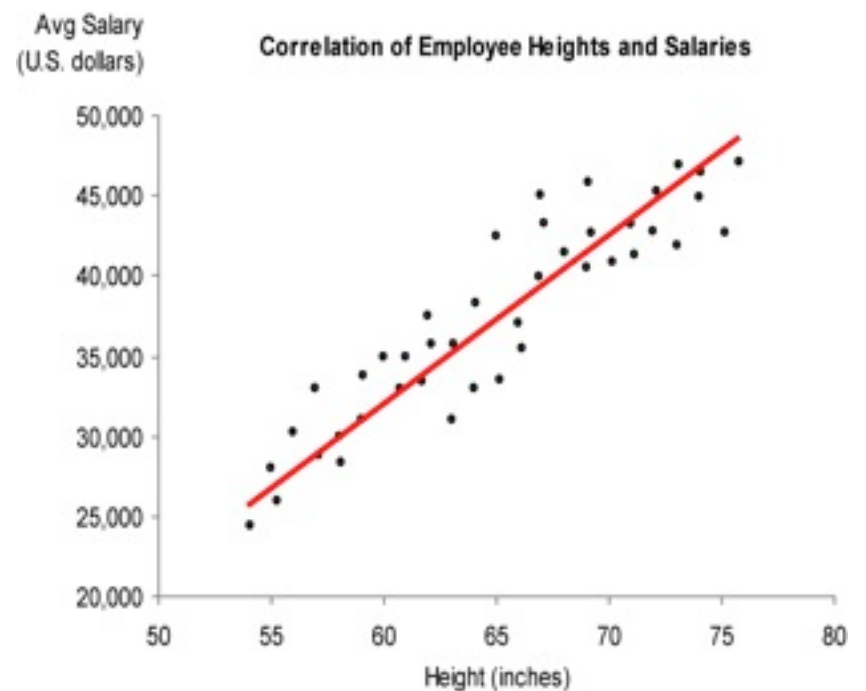
Point and lines can be combined in a single graph in two ways:

- Lines connect the individual points.
- Lines encode the overall shape of the points in the form of a trend line.

The strength of using them together in these ways is the ability to simultaneously emphasize individual values and their overall shape.

What meanings do these boxes suggest?
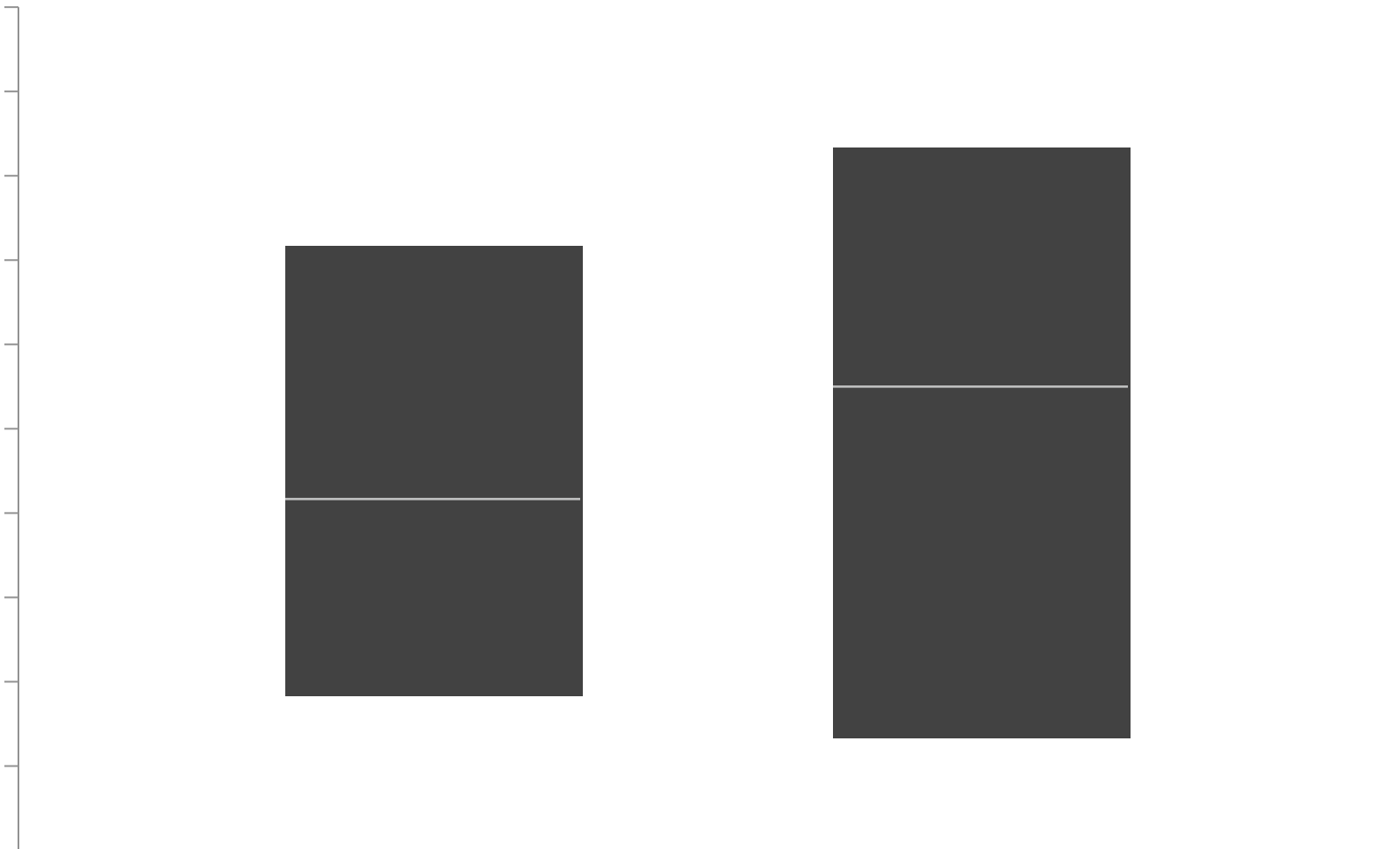
These rectangles are similar to bars, but because they don't share a common baseline, we tend to notice the difference between the positions of their tops, the difference between the positions of their bottoms, and the difference between their lengths. This is precisely what these rectangles are designed to help us do. Rather than bars, these are called boxes, which are used in a kind of graph called a box plot. Each box represents the distribution of an entire set of values: the bottom represents the lowest value, the top the highest value, and the length the full spread of the values from lowest to highest. The mark that divides the box into two sections—in this case a light line—indicates the center of the distribution, usually the median or mean. A measure of center or average gives us a single number that we can use to summarize an entire set of values. Notice how your eyes are encouraged compare the different positions of these box's centers—the fact that on average, the values represented by the box on the right are higher than those on the left.

The strength of boxes
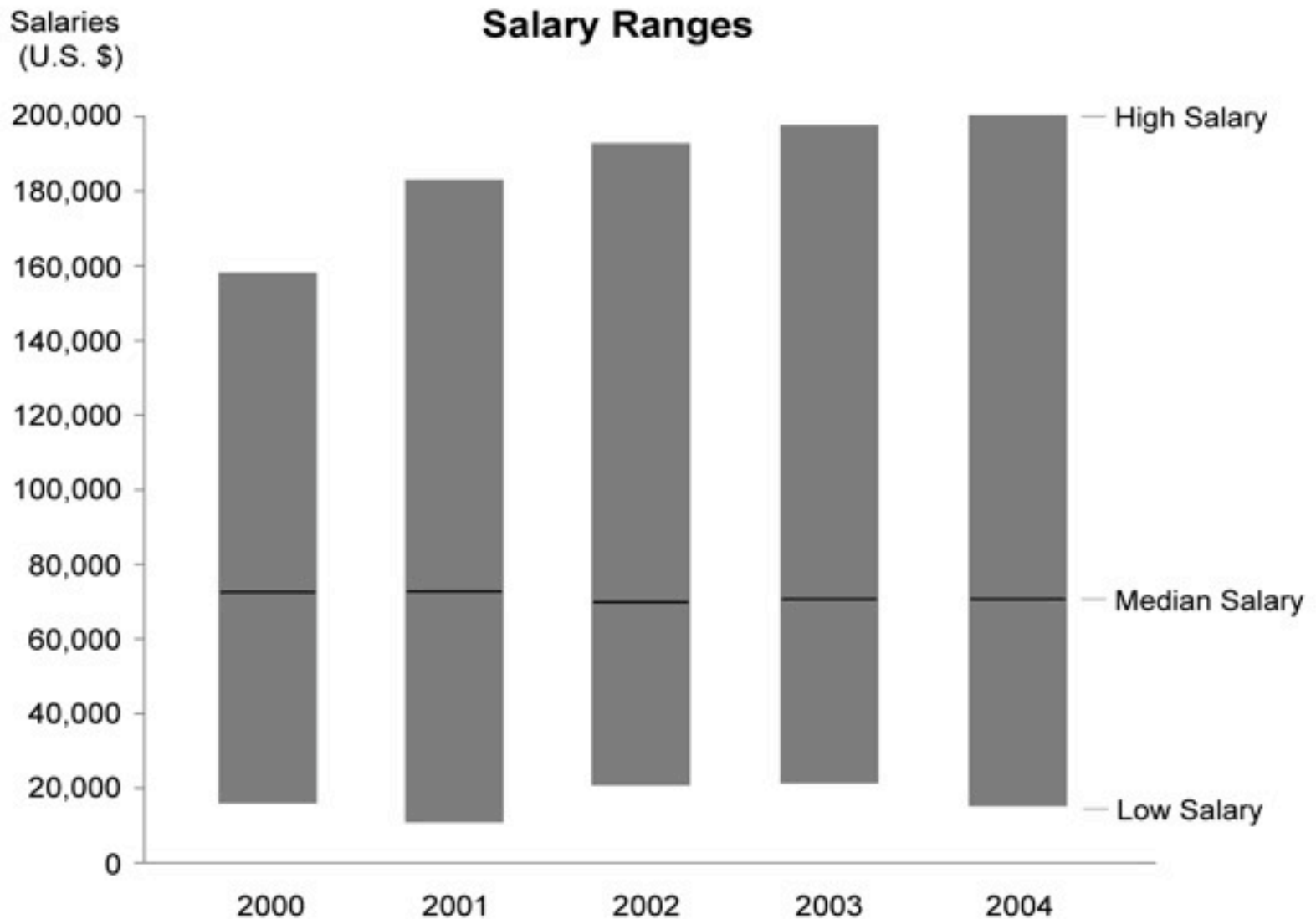
Because boxes have a two ends that can both be used to mark a position along a quantitative scale, they can be used to encode the range from lowest to highest of a full set of values. A data point in some form, such as the short line on the examples above, can be used to mark the center of the range such as the median of the full set of values. This combination of points and bars provides a powerful way of summarizing the distributions of multiple sets of values.

When boxes and points are combined in this or similar ways to encode a distribution of values, it is called a *box plot*.

Why not encode data as 2-D areas?

1

?
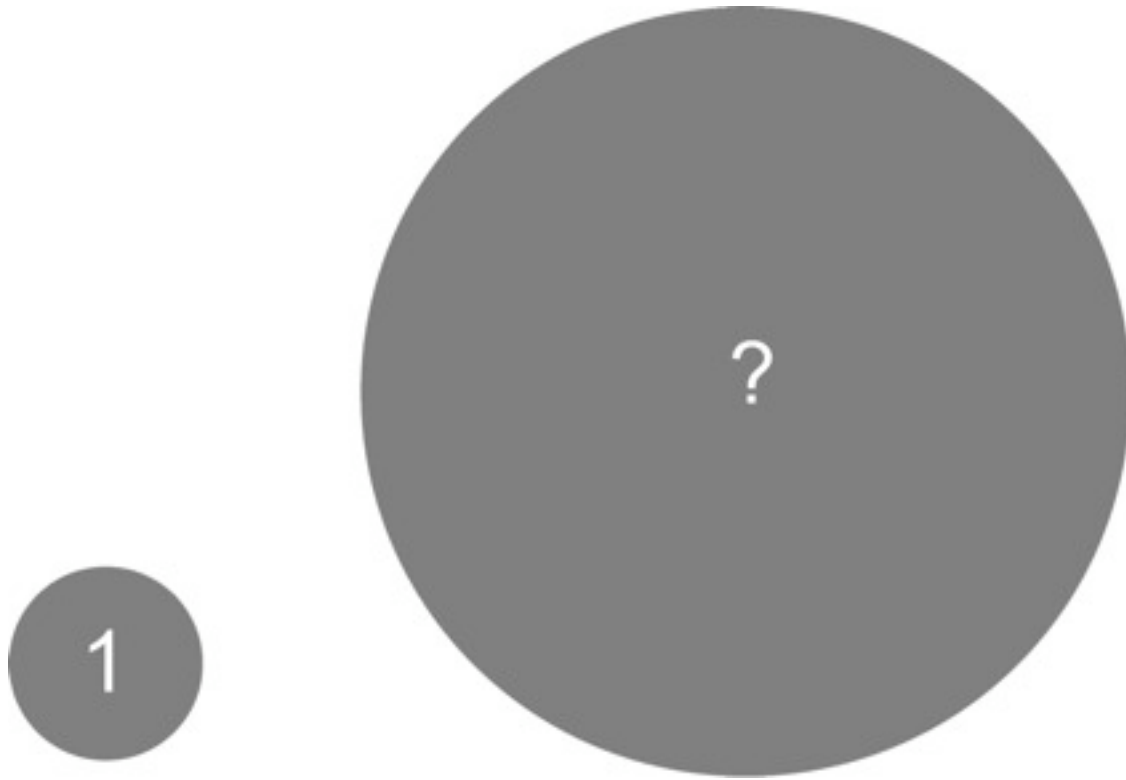
How big is the large circle? 16

Our visual perception of 2-D area is poor. It is difficult for us to accurately compare the sizes of 2-D areas.

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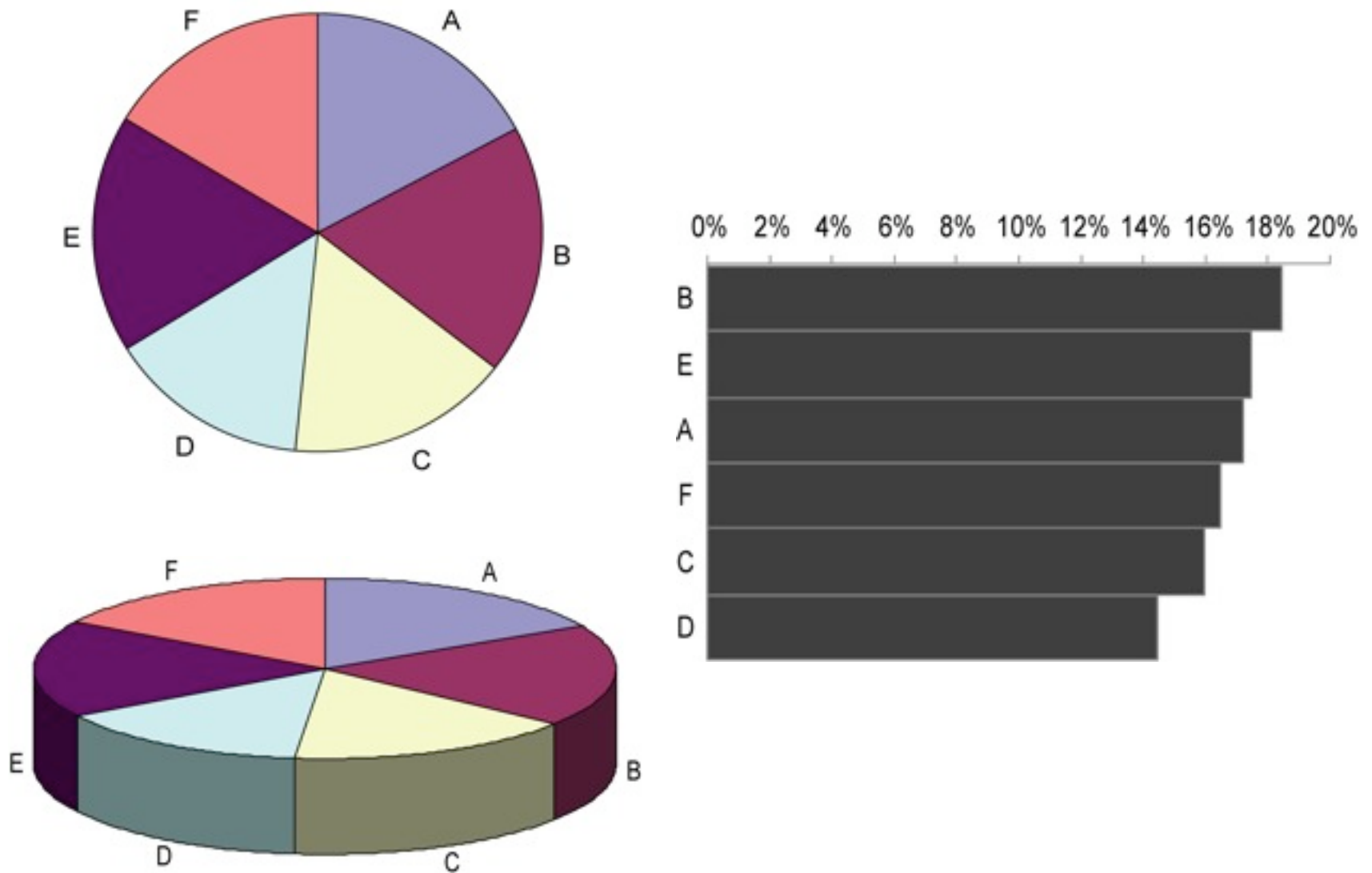
Pie charts use 2-D areas and the angles formed by slices to encode quantitative values. Unfortunately, our perception of 2-D areas and angles as measures of quantity is poor.

Since all graphs have one or more axes with scales, there must be one on a pie chart, but where is it? The circumference of the circle where its quantitative scale would appear, but it is rarely shown.

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.*

Exhibit 12: 2007-2008 Ethnic Composition of the 681,038 Public School Student Population

Hispanic 2.7%

American Indian 0.8%

White 49.2%

Black 46.0%

Asian 1.4%
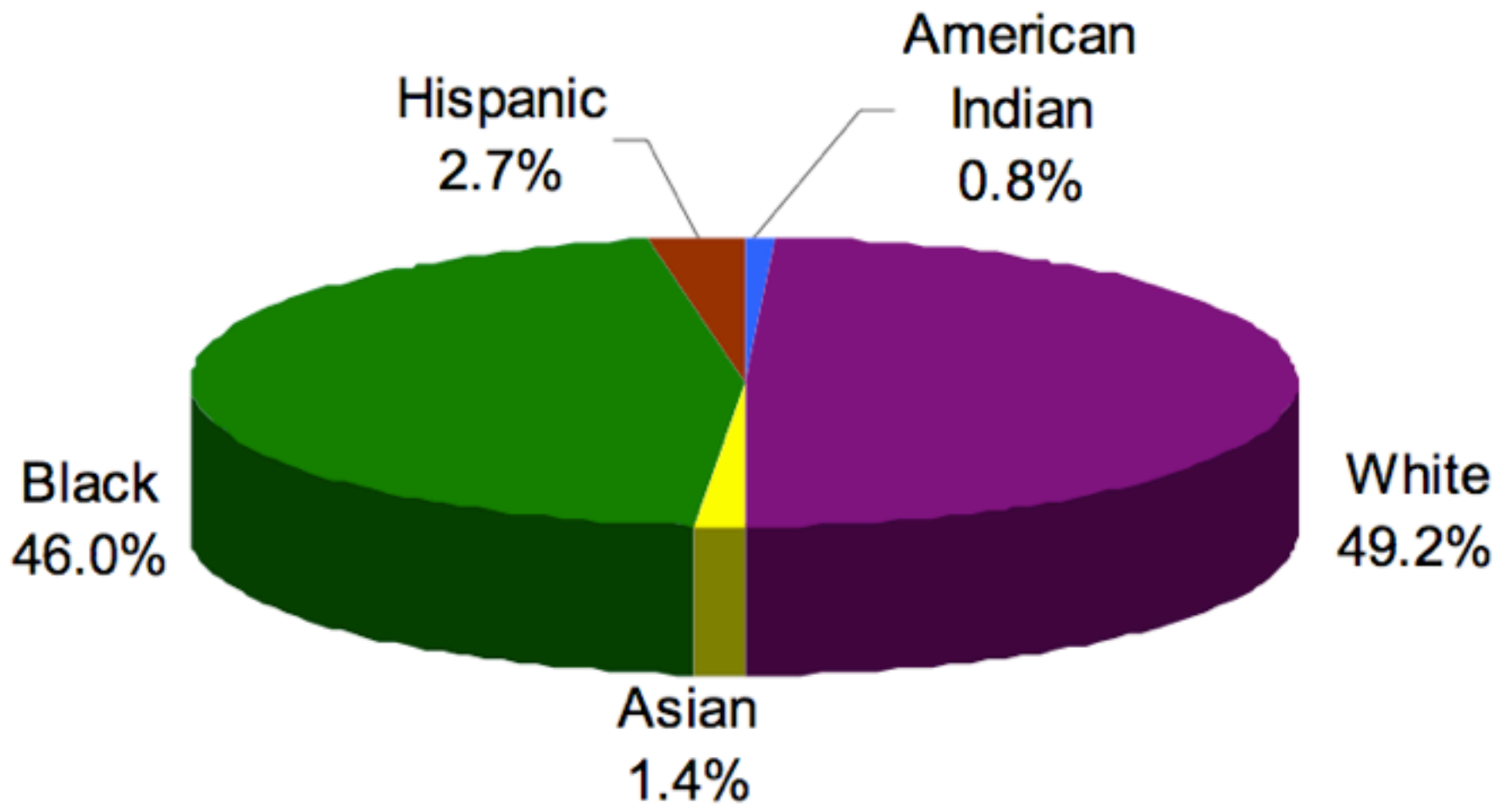
Save the pies for dessert!

Legend:
- Pie I have eaten
- Pie I have not yet eaten

People are too often impressed by displays that don't work.
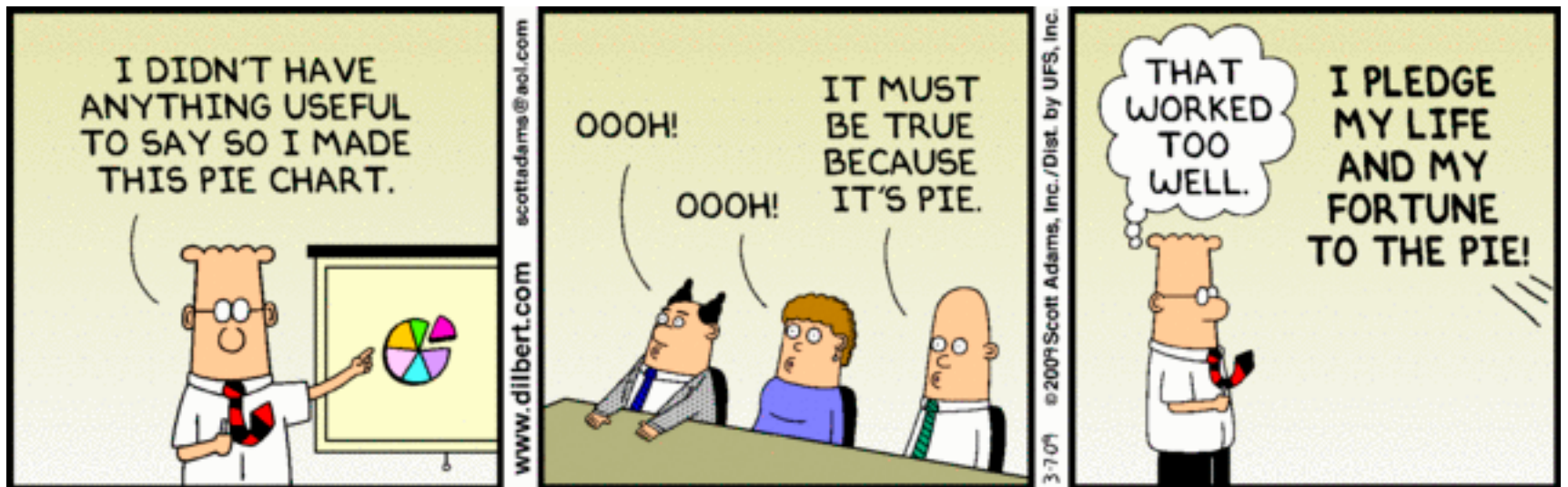
But only until they must actually use them for something real.

Which objects display each relationship best?

	Points	Lines	Bars	Boxes
Nominal comparison				
Time series				
Ranking				
Part-to-whole				
Deviation				
Distribution				
Correlation				

During the course of the next few slides we are going to identify which visual objects (points, lines, points and lines, bars, and boxes) do the best job of encoding each type of quantitative relationship.

Nominal Comparison

2003 Sales

Nominal Comparison

2003 Sales

Bars
(vertical or horizontal)
or
replace bars with points

2003 Sales
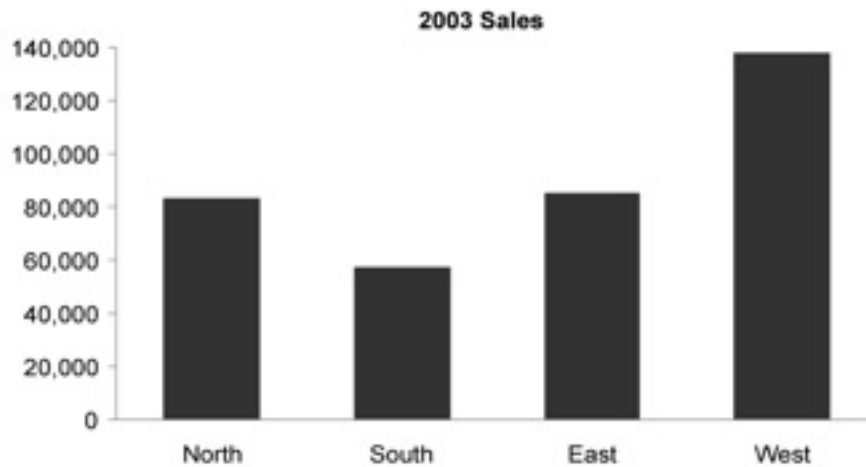
(U.S. $ in thousands)
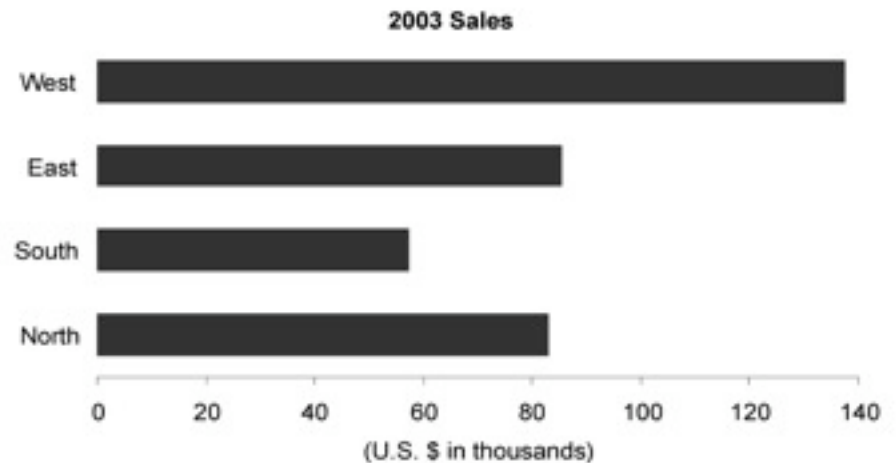
Why isn't it appropriate to use a line to encode a nominal comparison? Because the slope of the line as it moves from data point to data point would suggest change between different instances of the same measure. The difference between each region's sales is meaningful, but the movement from one region's sales to the next does not represent a change. Bars work best because they emphasize the independent nature of each region's sales.

Time Series
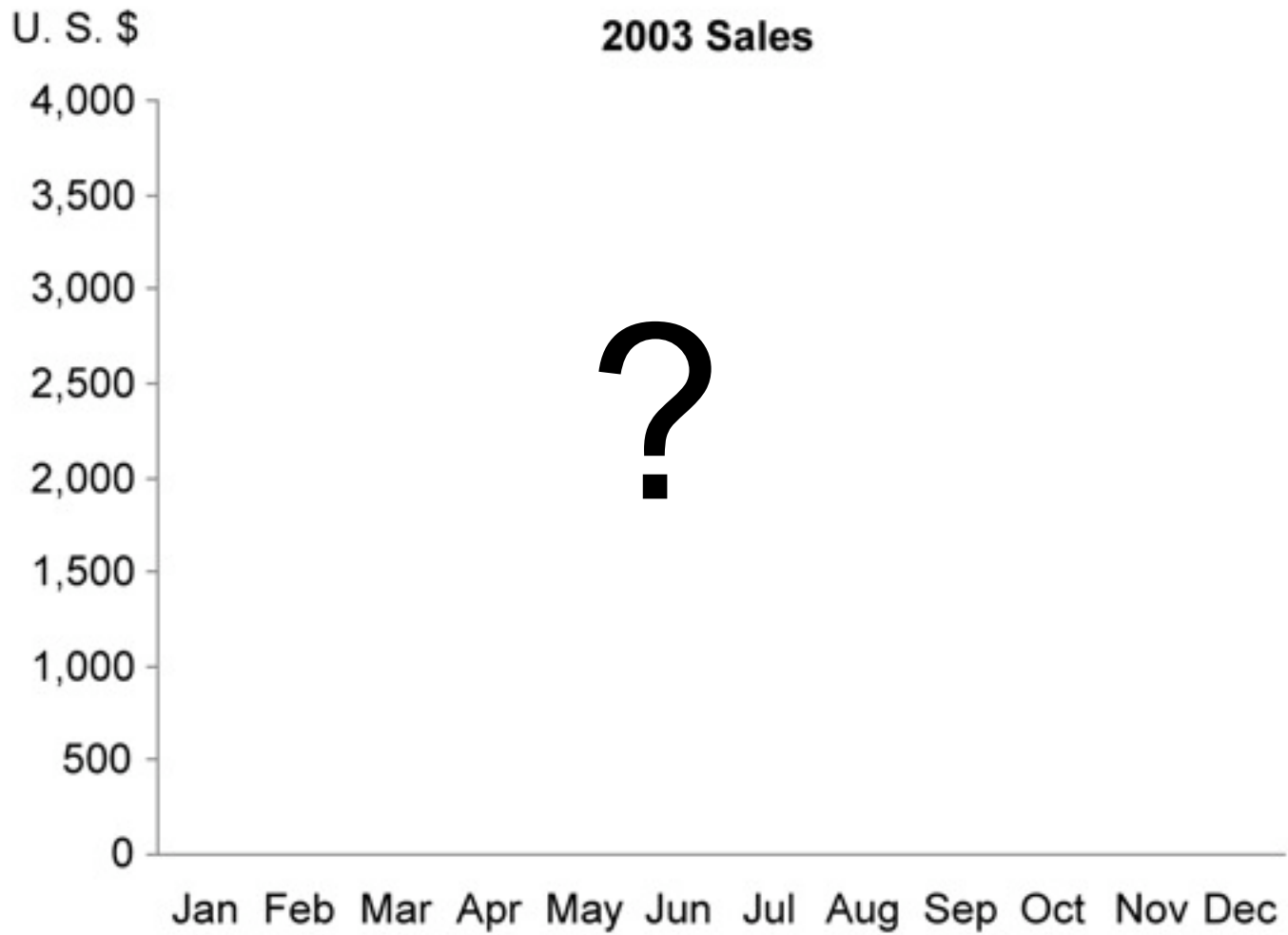
Time Series
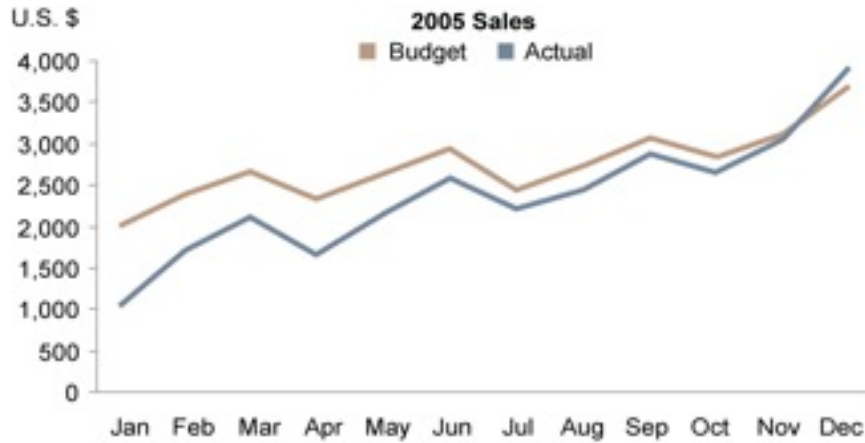
Lines
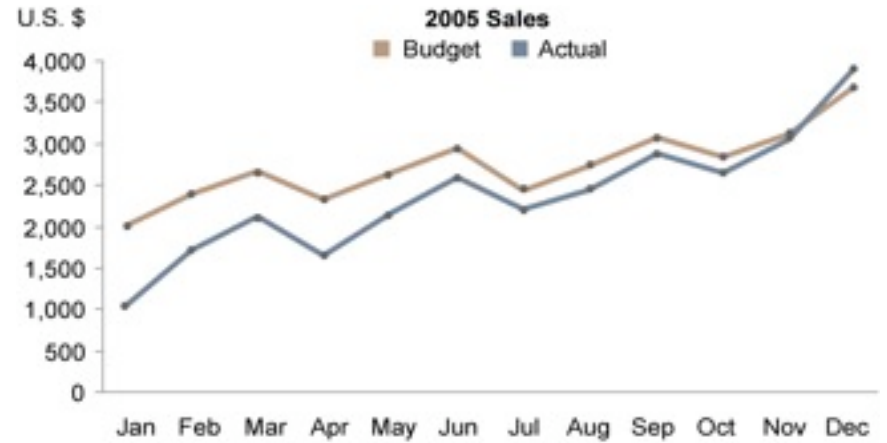
Points and Lines
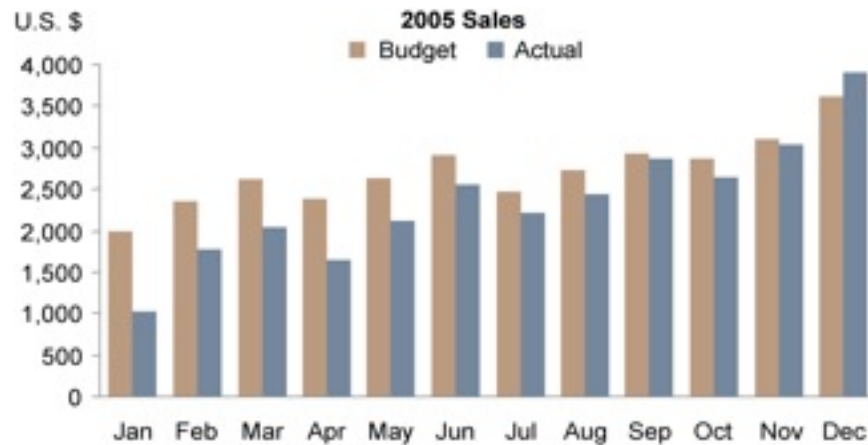
Bars (vertical only)

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.

Ranking

Units Sold

400,000	
300,000	
200,000	?
100,000	
0	

Russia Czech Republic Slovakia United Arab Emirates Saudi Arabia Egypt

Ranking

Units Sold

Bars
(vertical or horizontal)
or
replace bars with points

Units Sold
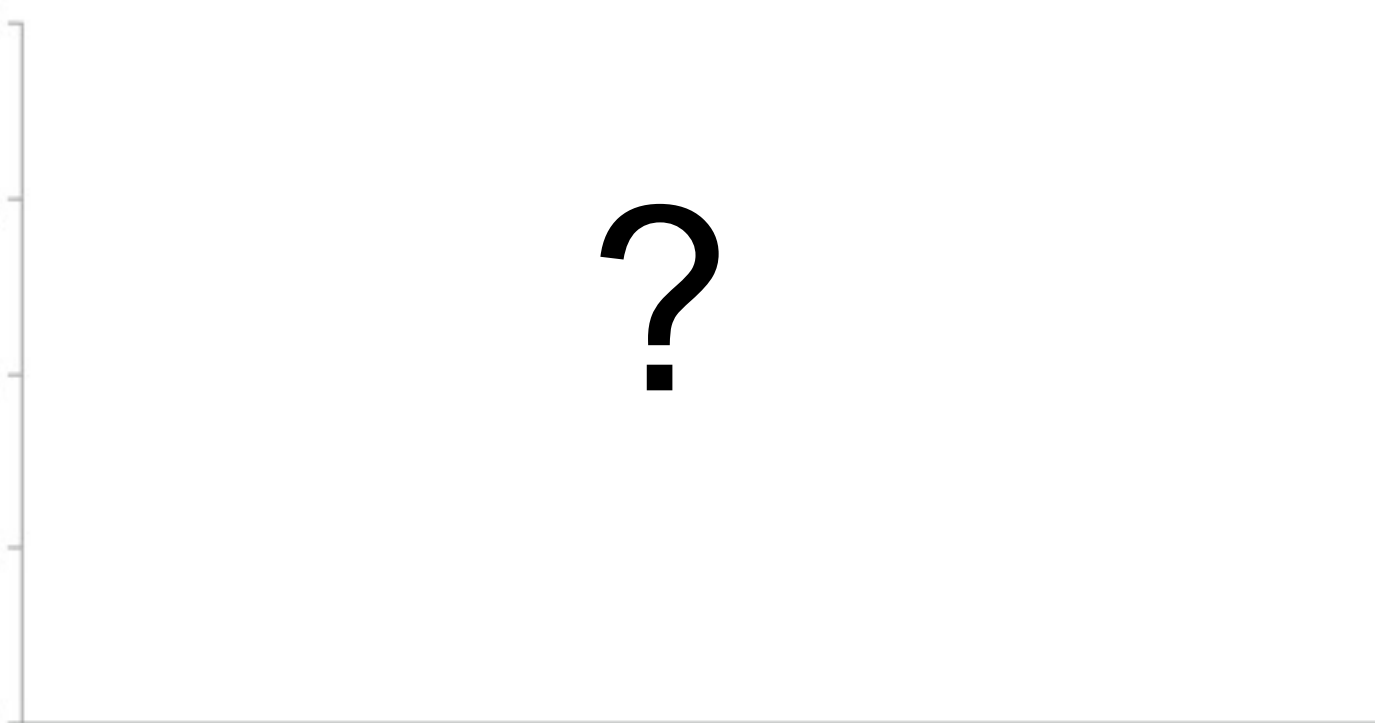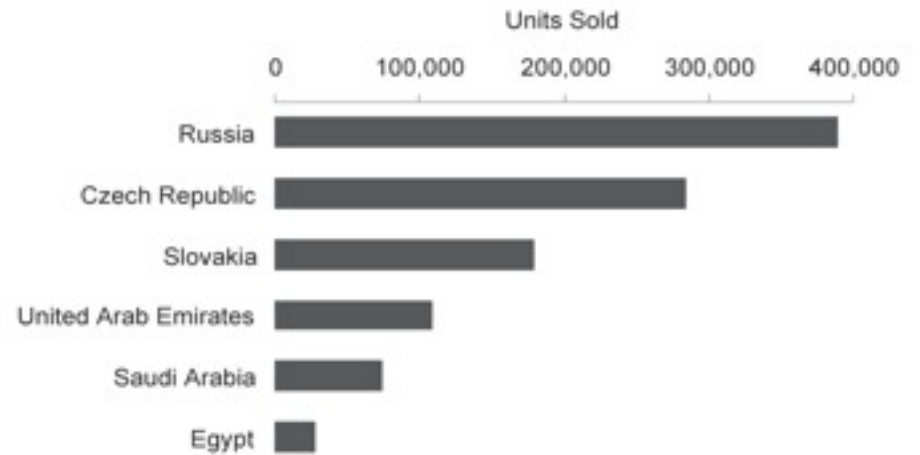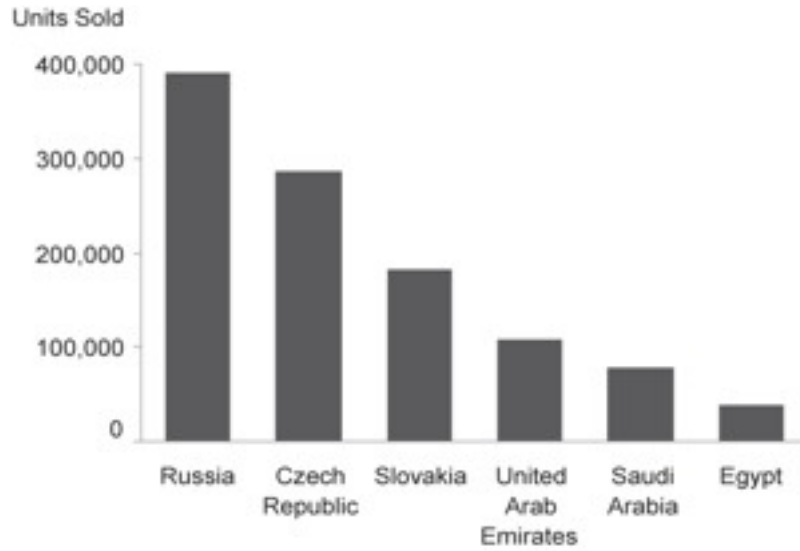
To emphasize the high values, sort the values in descending order from left to right or top to bottom; to emphasize the low values, sort the values in ascending order from left to right or top to bottom.

Part-to-Whole

Percentage of Total Sales

40%

30%

20%

10%

0%

?

West   East   North   South

Part-to-Whole

Percentage of Total Sales

West	East	North	South	Total
38.1%	24.4%	24.0%	14.5%	100%

Bars
(vertical or horizontal)

Percentage of Total Sales

West	38.1%
East	24.4%
North	24.0%
South	14.5%
Total	100.0%
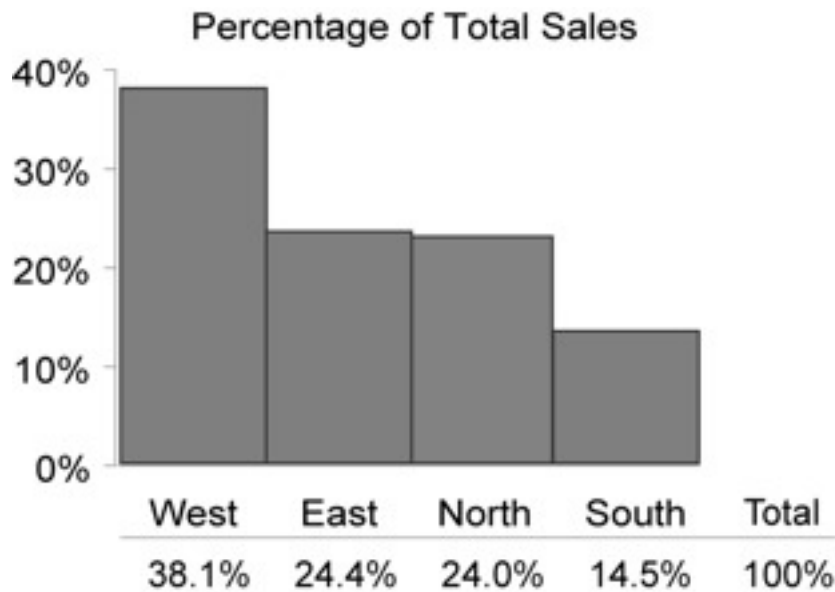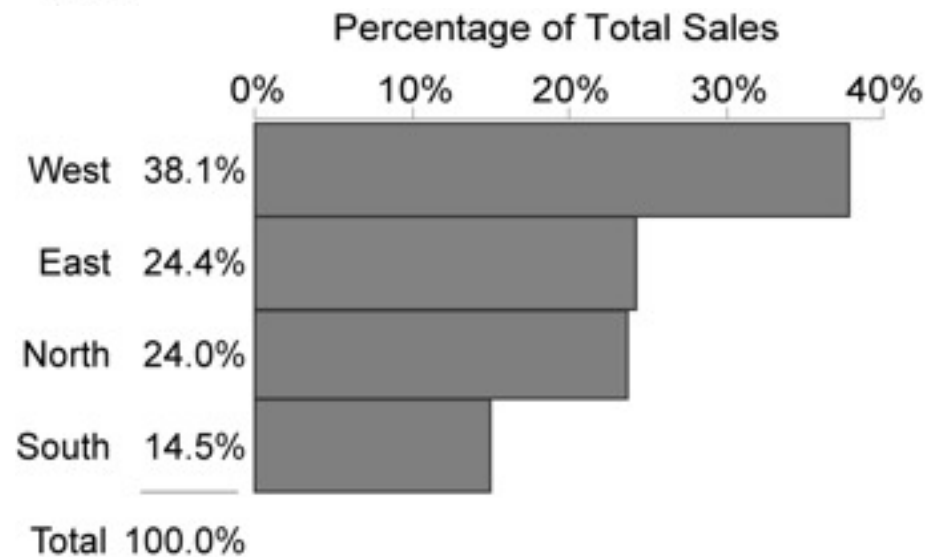
You will no doubt notice that the graph that is generally used to display part-to-whole relationships, the ubiquitous pie chart, has been left out. Despite the problems inherent in all forms of area graphs, the one useful characteristic of a pie chart is the fact that everyone immediately knows that the individual slices combine to make up a whole pie.

When bars are used to encode part-to-whole relationships, the fact that the bars add up to 100% – a whole – is not as obvious, even though percentage is the unit of measure. To alleviate this problem, be sure to make the part-to-whole nature of the data obvious in the graph's title. In the above examples, the fact that the individual bars, which represent regional sales, add up to total sales, is stated explicitly in the title. Another way that you can subtly indicate the fact that the bars represent parts of a whole is to position them so that they are touching one another, thus visually reinforcing their intimate connection of parts of a common whole.

What about stacked bars?

Percentage of Total Sales (100% stacked bar — West, East, South, North)

2003 Sales (stacked bars by quarter — West, South, East, North), U.S. $ (thousands)

Percentage of Total Sales (bar chart — West, East, North, South)

2003 Sales (clustered bars by quarter — North, East, South, West), U.S. $ (thousands)
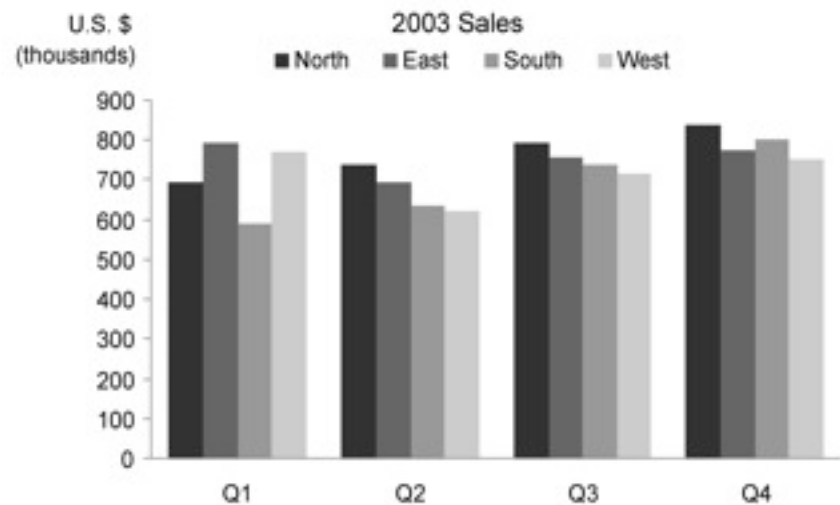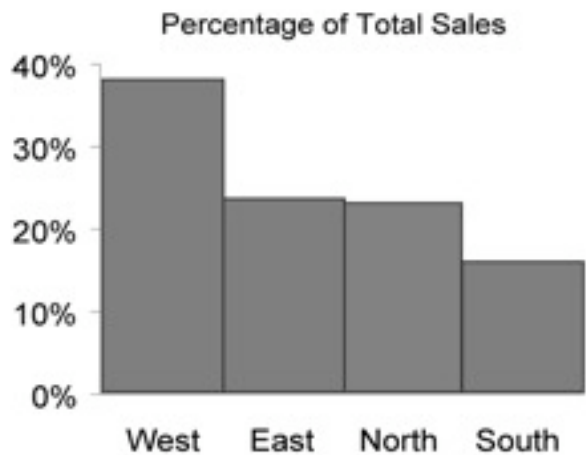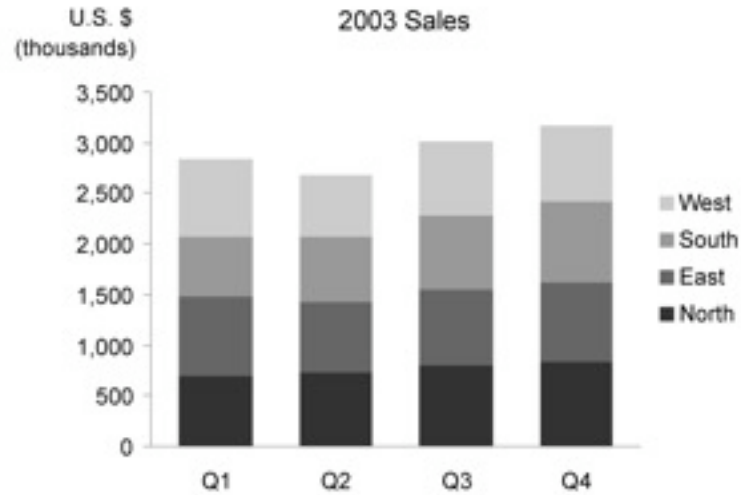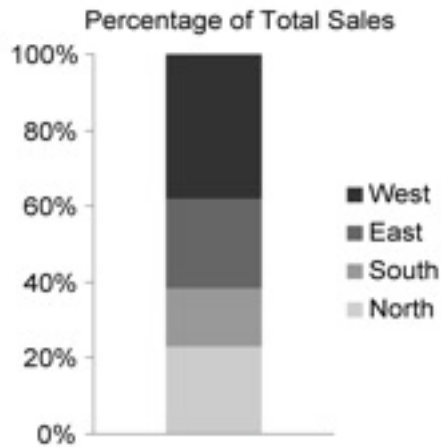
It is harder to compare and assign values to bars that are stacked, rather than side by side. Use stacked bars only when you need to display a measure of the whole as well as its parts.

Deviation

Headcount
Actual to Plan

?
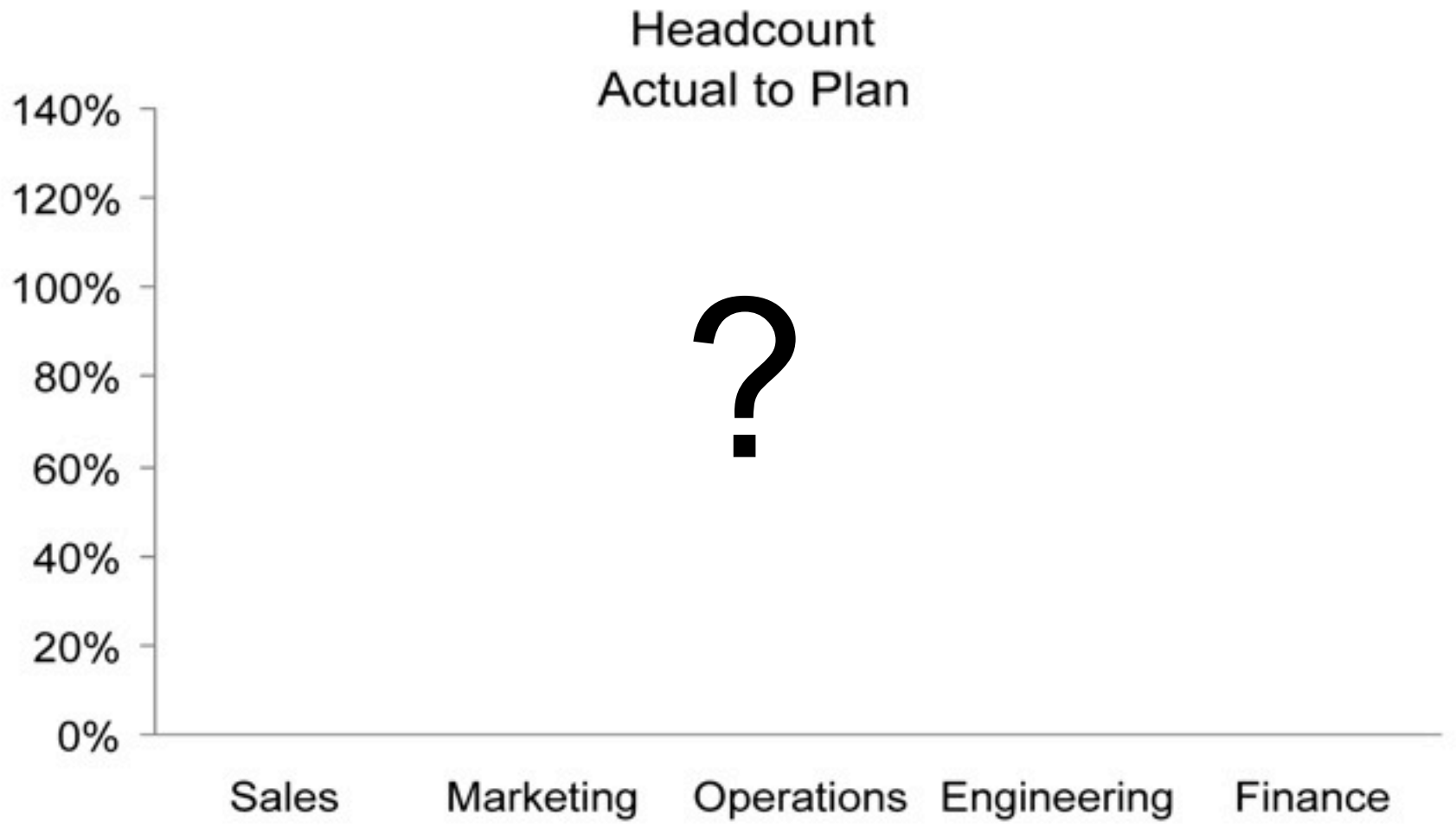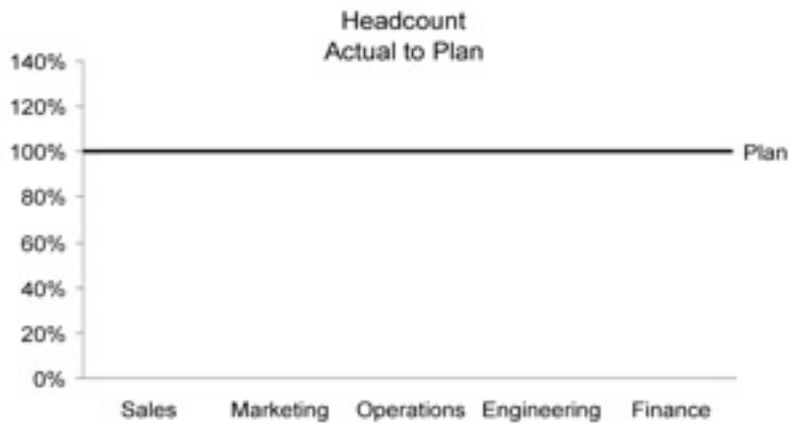
Deviation

Reference Lines (always)

Bars

Lines

The use of a reference line makes it clear that the main point of graphs like those pictured above is to display how one or more measures deviate from some point of reference. Although deviations need not be expressed as percentages, when they are, especially as plus or minus percentages, the intention of the graph to focus on deviation becomes crystal clear. When bars are used to encode deviations, the reference line should always be set to a value of zero with the bars extending up (positive) or down (negative) from there.

Distribution

Shipping Performance

Number of Orders (y-axis): 0, 20, 40, 60, 80, 100

Number of Days (x-axis): 1, 2, 3, 4, 5, 6, 7, 8

?

Single distribution

Histogram

Number of Orders

Shipping Performance

100 — 80 — 60 — 40 — 20 — 0

Number of Days: 1 2 3 4 5 6 7 8
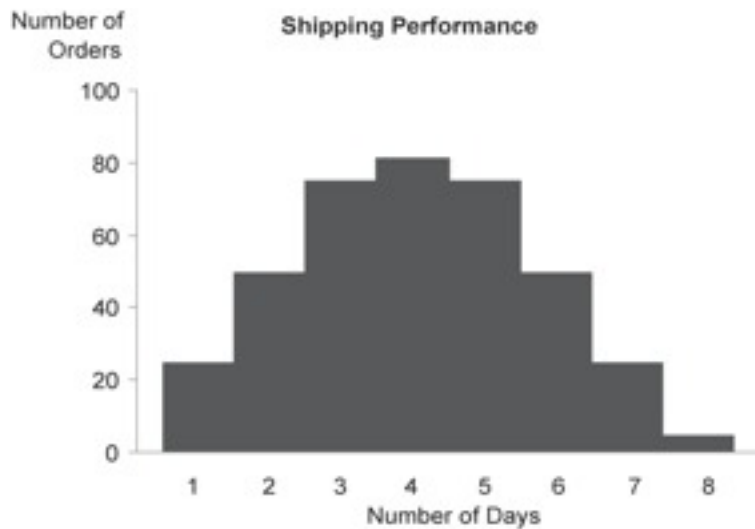
Bars (vertical only)

Frequency Polygon

Number of Orders

Shipping Performance

100 — 80 — 60 — 40 — 20 — 0

Number of Days: 1 2 3 4 5 6 7 8

Lines

Strip Plot

10   20   30   40   50   60   70
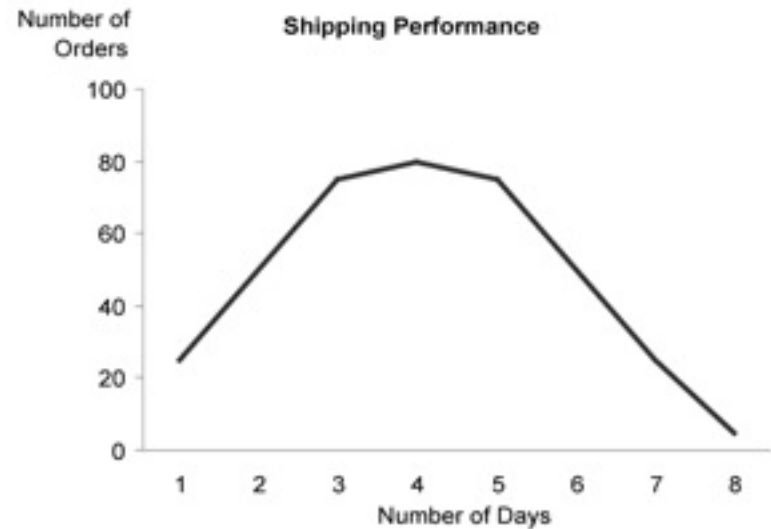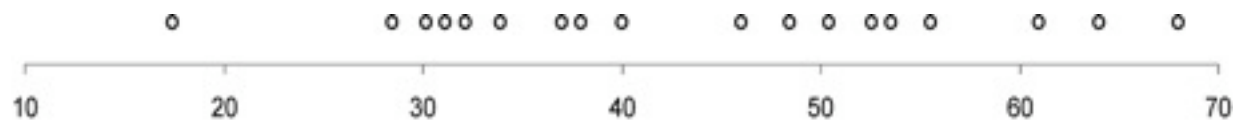
Points

Histograms and frequency polygons both do a wonderful job of displaying frequency distributions; the only difference is whether you intend to emphasize the values of individual intervals or to emphasize the overall shape of the distribution. When you have only a few data points, strip plots work quite well, providing the detail that would be lost in the other two graphs.

Multiple distributions
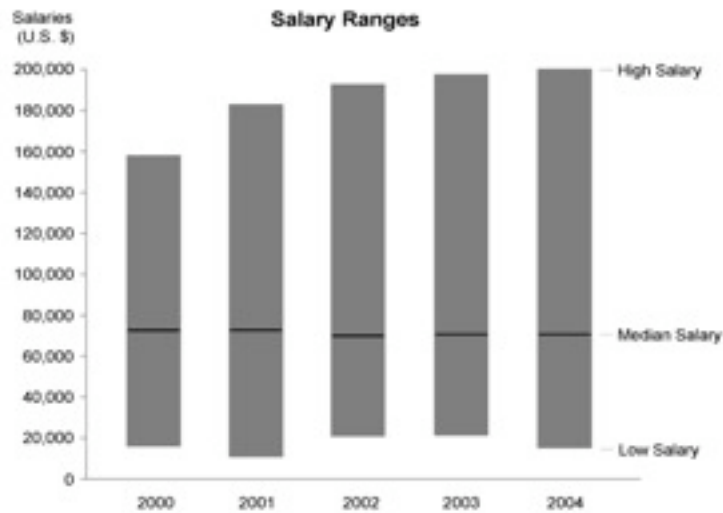
Boxes
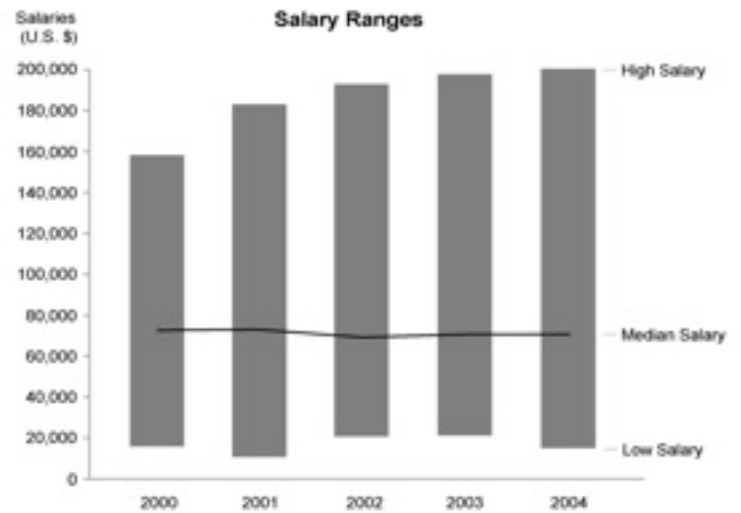
Boxes with Lines

Box Plots

Points

Strip Plot

Box plots, which use range bars to encode the spreads of distributions and some form of point to mark measures of center – averages, such as means and medians – are usually the best way to display multiple distributions so they can be compared. Box plots can become very sophisticated, with boxes (or bars) that encode measures of distribution other than the full range, such as quartiles, and individual data points to display outliers. If you want to emphasize the change from one distribution to another across an interval scale, lines can be used to mark the center in place of the points. When you have only a few values per distribution, points in the form of a strip plot sometimes work best for showing the details of the distributions.

Multiple five-point distributions

Salary Ranges, Medians, and Percentiles
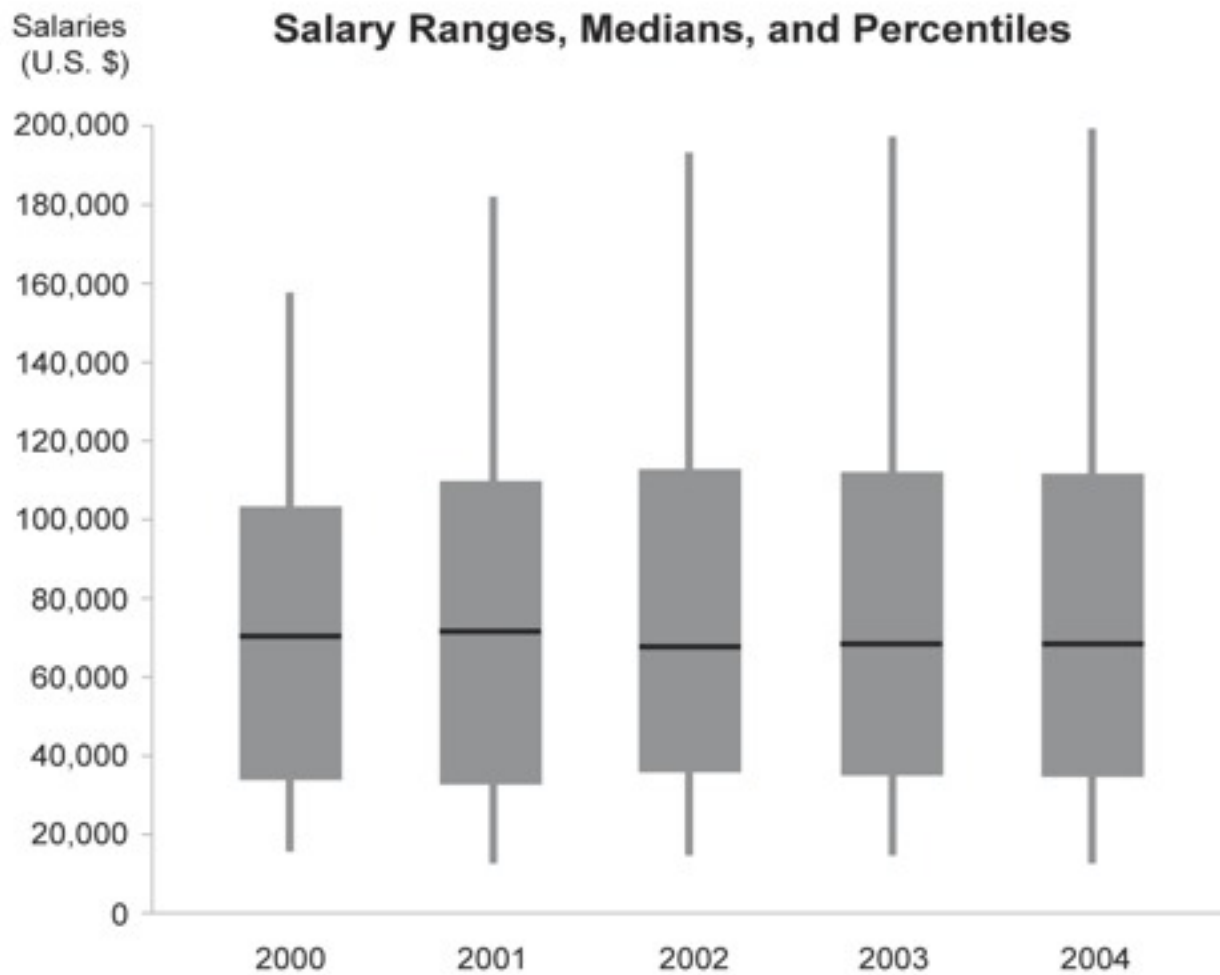
There are many ways to visualize multiple distributions. Some provide more insight than others. The best approach depends on how much detail you want to communicate.

In a 5-point distribution, the full range of values is subdivided into quartiles to display the distribution of the top 25%, upper mid 25%, lower mid 25%, and bottom 25% provide the most insight of these three approaches.

Correlation

Correlation of Employee Heights and Salaries
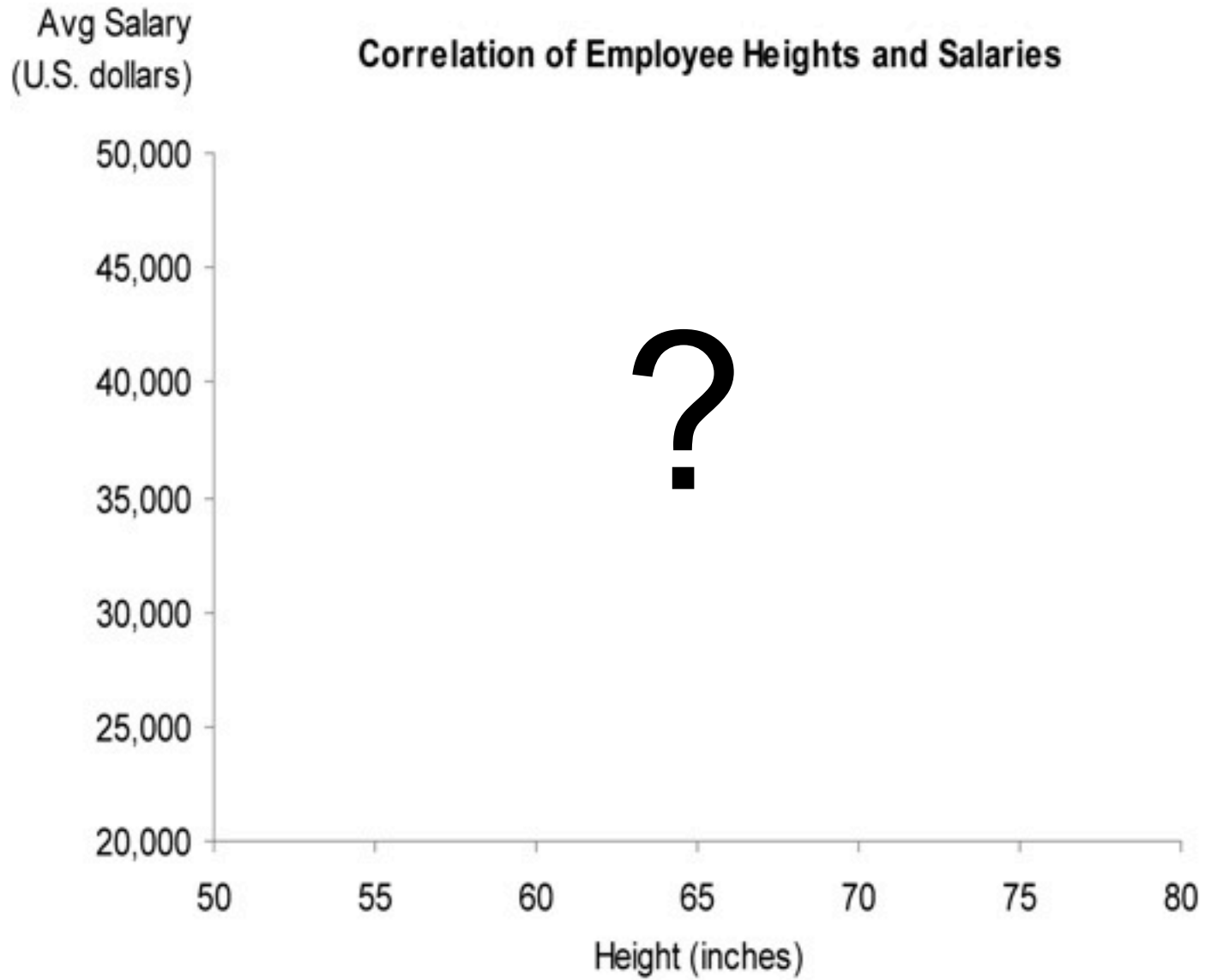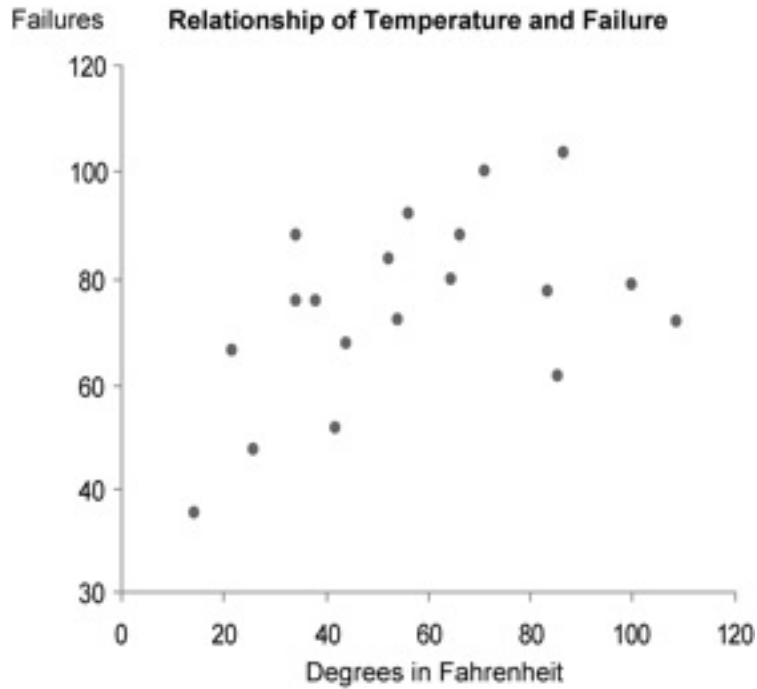
Correlation

Scatterplots
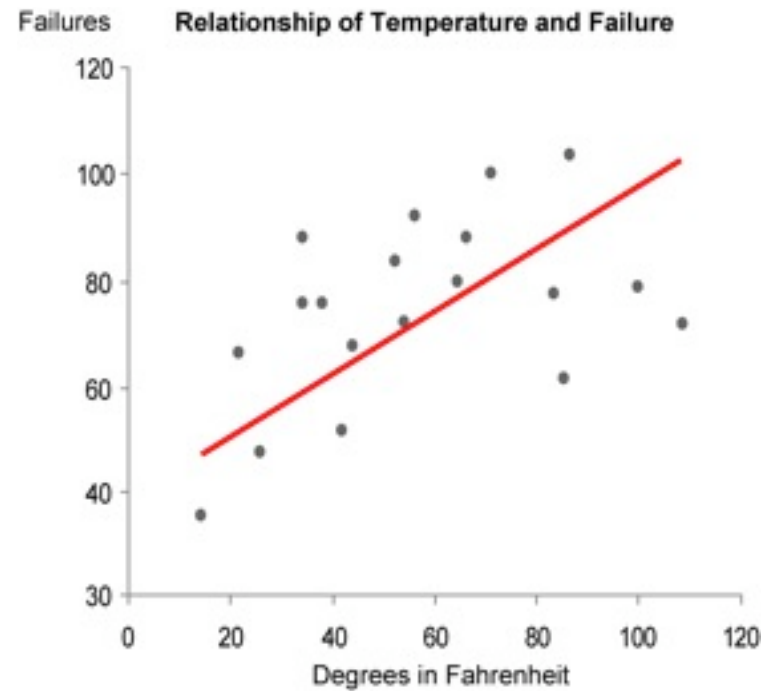
Points

Points and Trend Lines

When you must display correlations to folks who aren't familiar with scatter plots, and you don't have the time nor opportunity to provide the instruction that they need, there are effective ways that you can use bars to encode the correlation of two data sets, especially if you have a relatively small set of values. For an introduction to paired bar charts and correlation bar charts, please refer to *Show Me the Numbers: Designing Tables and Graphs to Enlighten*.

YONKERS PUBLIC SCHOOLS

Math 3−8 Results, 2009

DistrictWide

Grade	Math Proficiency Status	FREQ.	CUM. FREQ.	PCT.	CUM. PCT.
03	Below Standard	204	204	11.49	11.49
	At/Above Standard	1571	1775	88.51	100.00
04	Below Standard	346	346	19.85	19.85
	At/Above Standard	1397	1743	80.15	100.00
05	Below Standard	352	352	20.63	20.63
	At/Above Standard	1354	1706	79.37	100.00
06	Below Standard	566	566	33.49	33.49
	At/Above Standard	1124	1690	66.51	100.00
07	Below Standard	451	451	26.69	26.69
	At/Above Standard	1239	1690	73.31	100.00
08	Below Standard	771	771	46.11	46.11
	At/Above Standard	901	1672	53.89	100.00

PERCENT

Math Performance Level   1   2   3   4

M09 Summary.sas   DistrictWide Summary

Exhibit 15: Dropout Rates for Grades 7-12 and Overall (9-12) 2003-2004 through 2006-2007 Percent of Students That Dropped Out

Grade	2003-2004	2004-2005	2005-2006	2006-2007
Grade 7	1.4%	1.5%	2.6%	1.6%
Grade 8	3.4%	3.3%	4.2%	2.8%
Grade 9	7.7%	7.6%	7.8%	7.7%
Grade 10	6.6%	6.2%	6.2%	6.0%
Grade 11	6.3%	6.4%	6.4%	6.3%
Grade 12	9.0%	7.6%	7.1%	7.1%
Overall (9-12)	7.4%	7.0%	6.9%	6.9%

■ 2003-2004 ■ 2004-2005 ■ 2005-2006 ■ 2006-2007
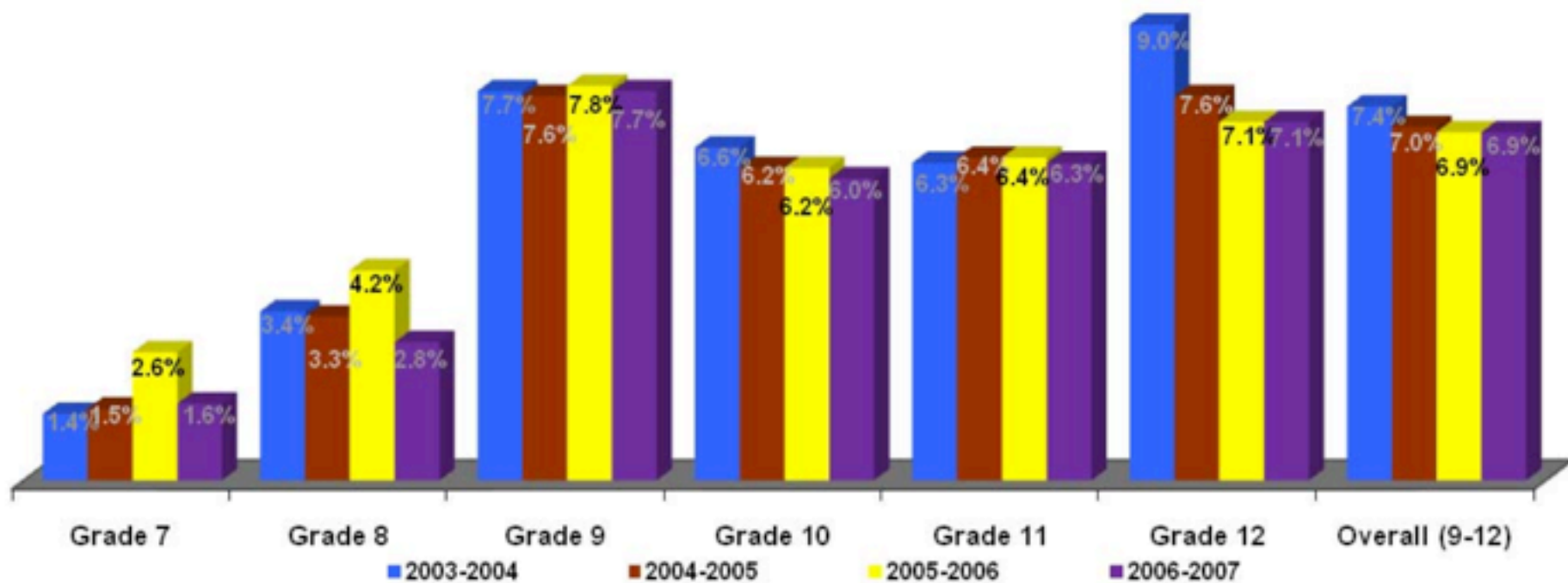
Remember the steps in the design process?

?

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.

1. Reduce the non-data ink.
   a) Subtract unnecessary non-data ink.
   b) De-emphasize and regularize what remains.
2. Enhance the data ink.
   a) Subtract unnecessary data ink.
   b) Emphasize the most important data ink.

Now let's apply these steps to the design of graphs.

What about grid lines in graphs?
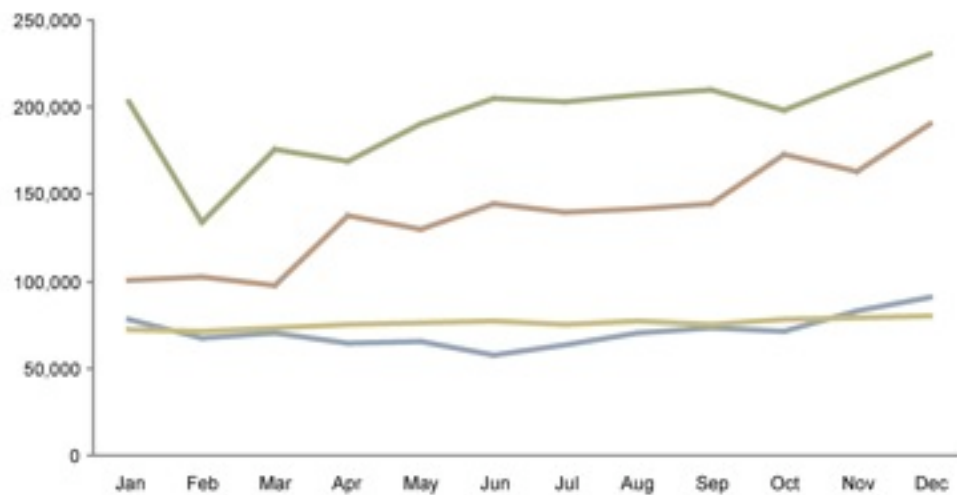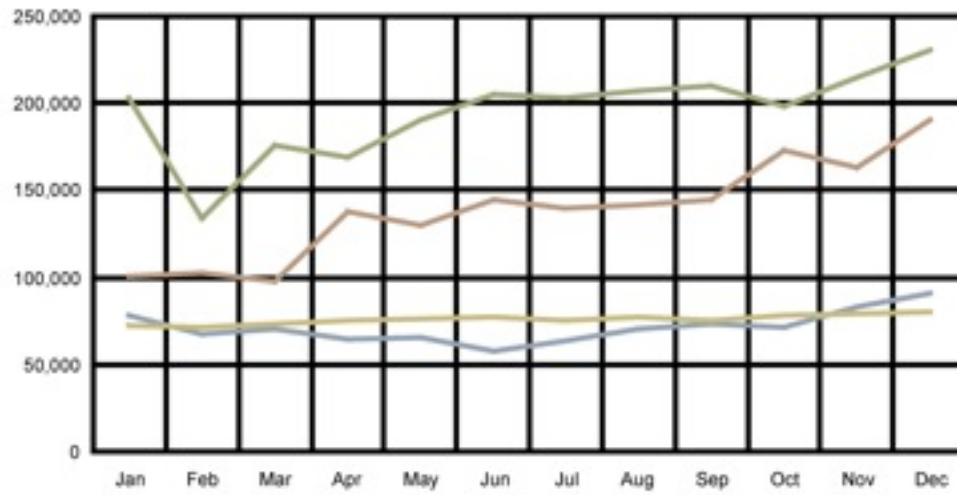
Grid lines are rarely useful and dark grid lines are never useful. They make it very difficult to pick out the shape of the data imprisoned behind the grid lines.

Grid lines clarify subtle differences.
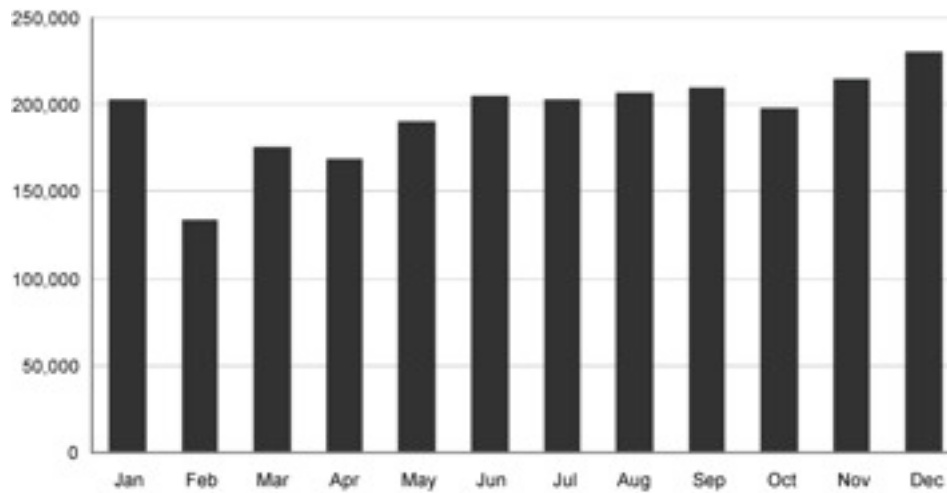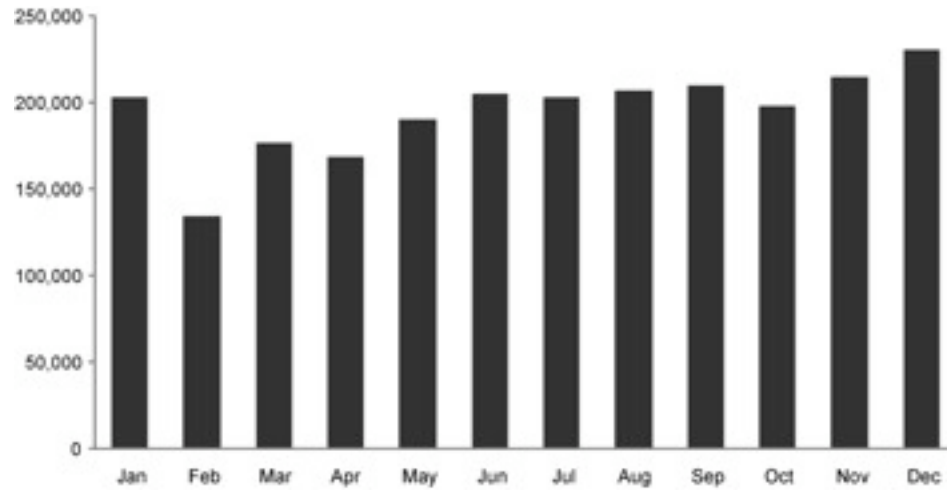
Light lines along the quantitative scale assist in making subtle distinctions.

Grid lines delineate sections.
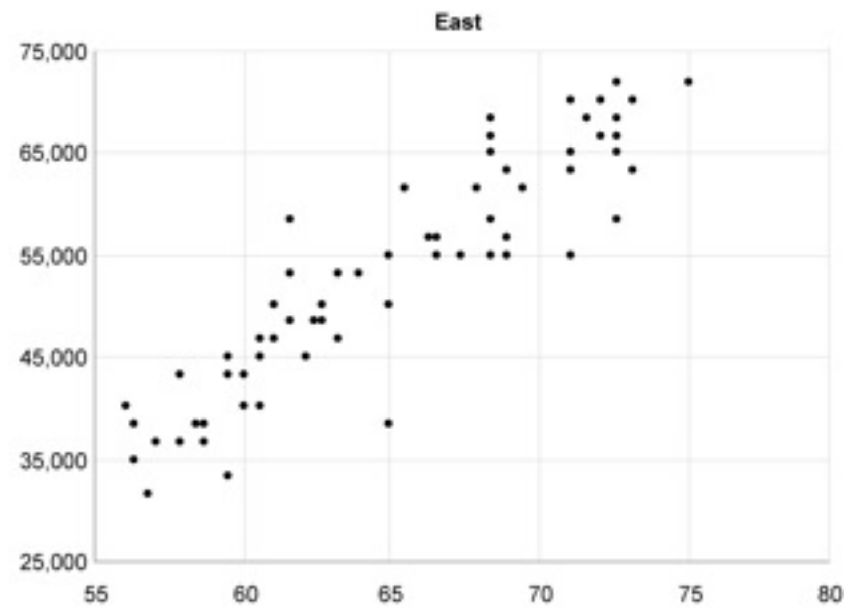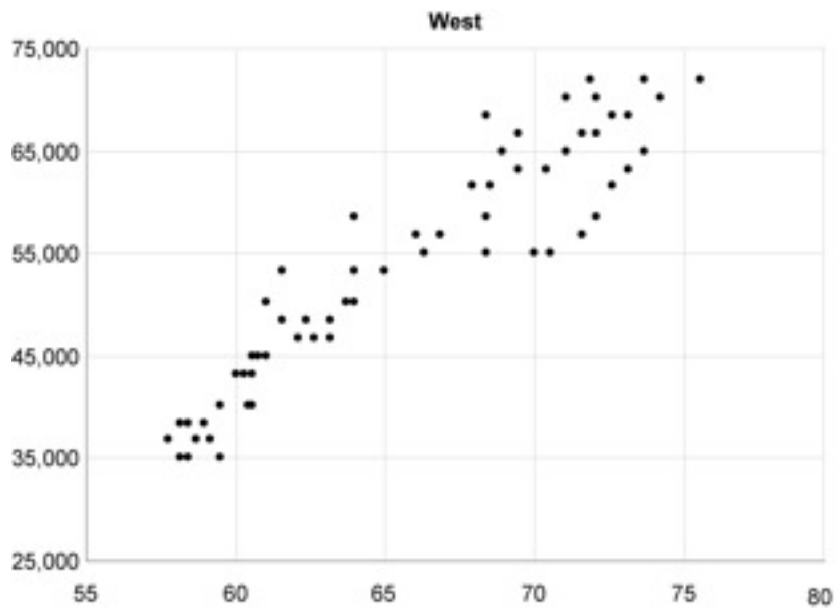
Light grids assist in reading and comparing subsections of graphs.

What about 3-D? It's so cute!!!
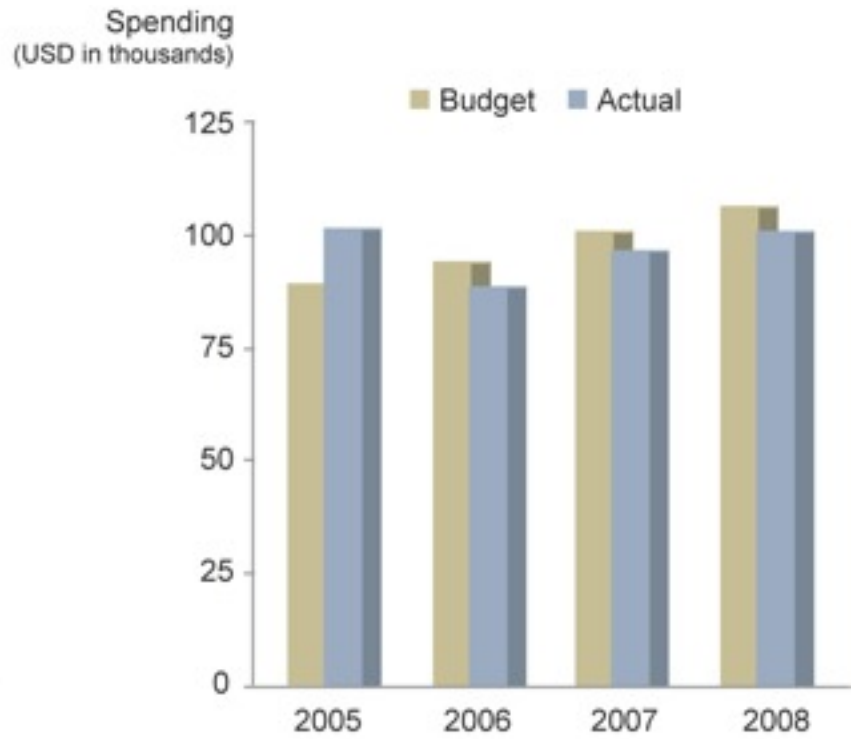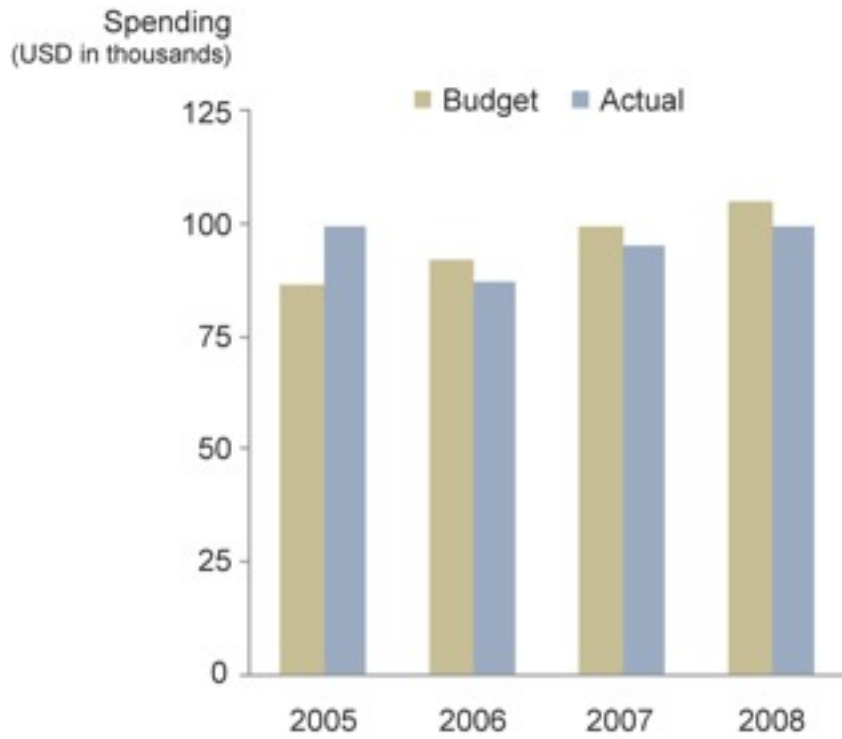
Adding a third dimension of depth to the bars on the right without adding a corresponding third variable is not only meaningless, it makes it more difficult to decode the data.

But what if there's a 3rd variable?
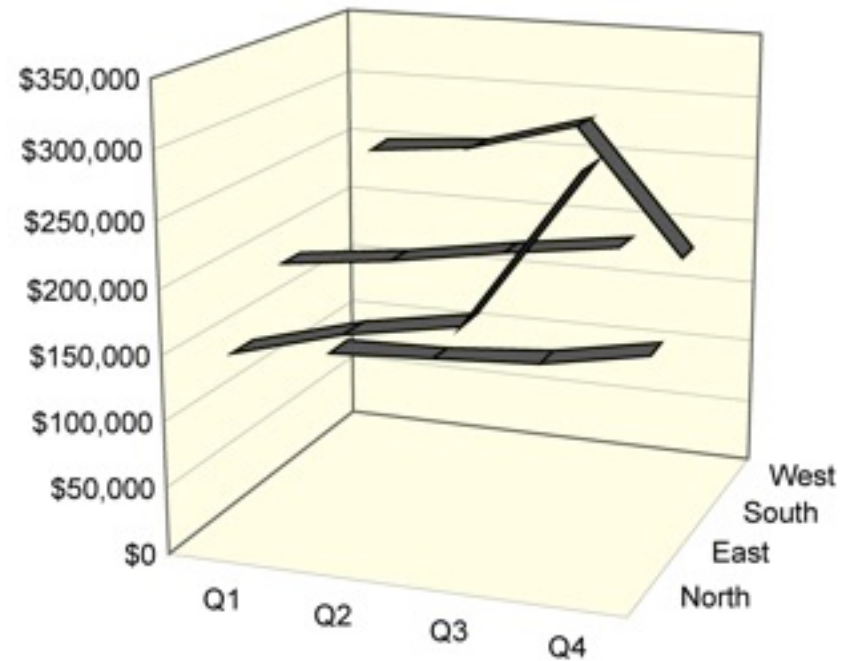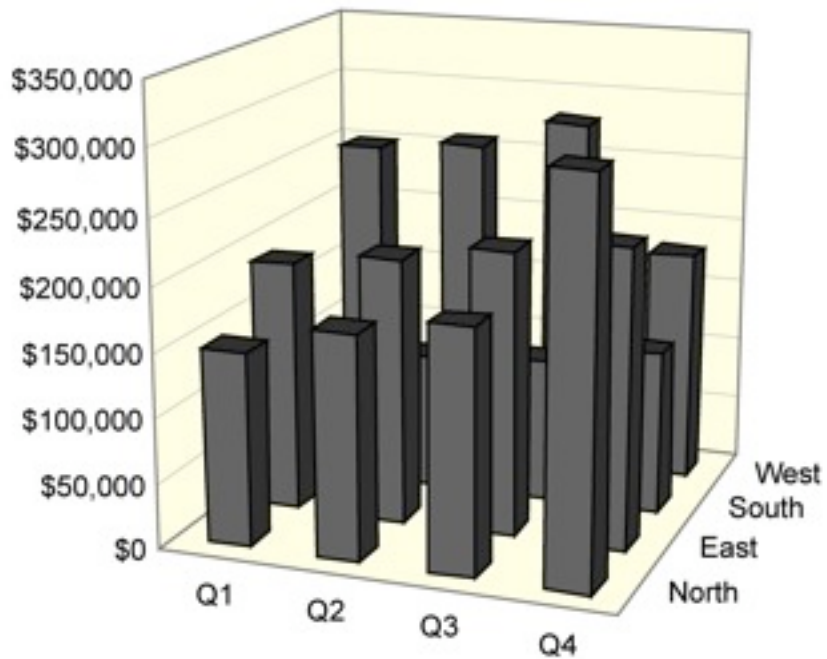
Can you determine which of the lines in the graph on the right represents the East region? Are you sure?

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

YONKERS PUBLIC SCHOOLS

ELA Grade 4, 1999 through 2005

Percent By Performance Level

All Students,   Districtwide,   by Year
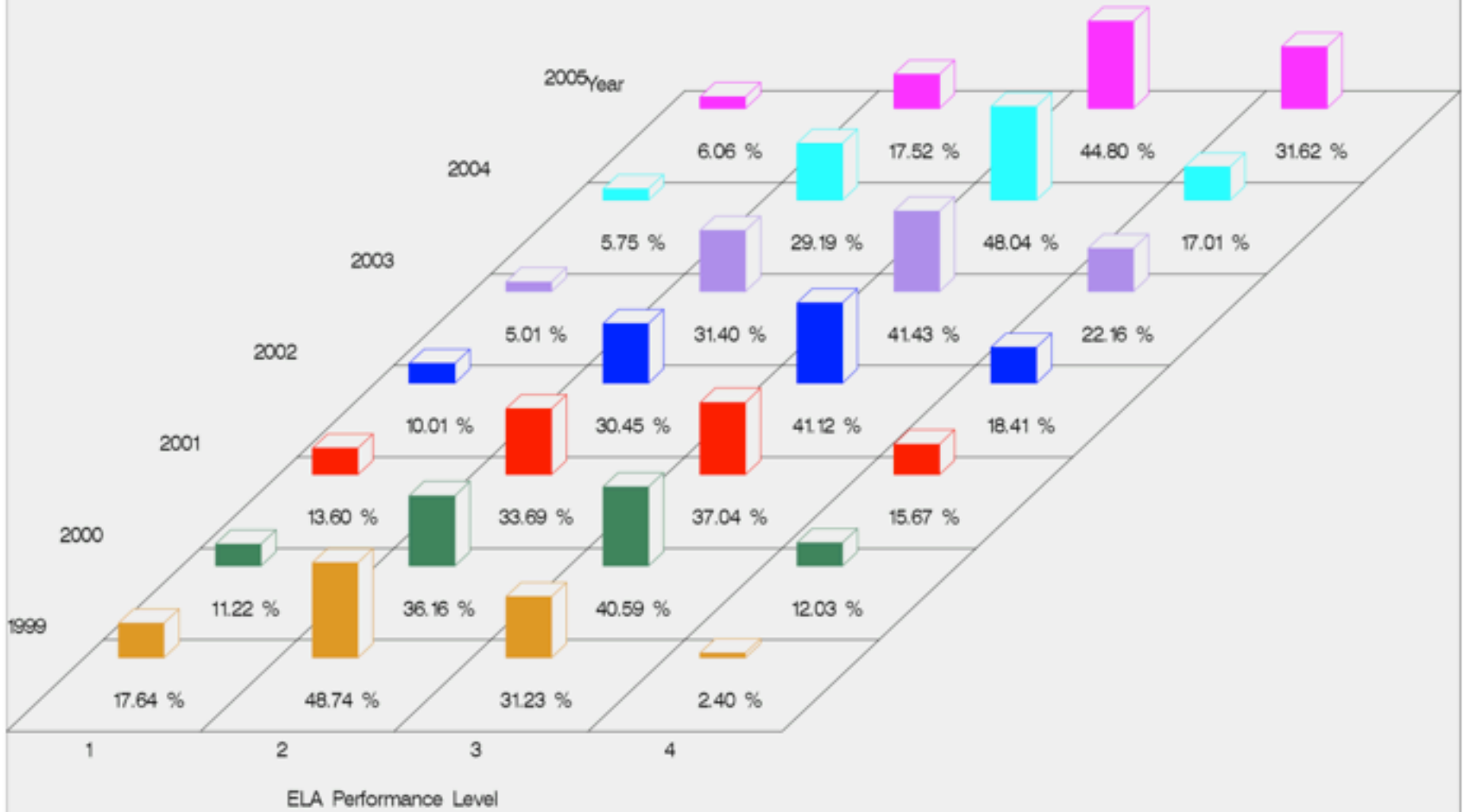
Year

2005

2004 — 6.06 % | 17.52 % | 44.80 % | 31.62 %

2003 — 5.75 % | 29.19 % | 48.04 % | 17.01 %

2002 — 5.01 % | 31.40 % | 41.43 % | 22.16 %

2001 — 10.01 % | 30.45 % | 41.12 % | 18.41 %

2000 — 13.60 % | 33.69 % | 37.04 % | 15.67 %

1999 — 11.22 % | 36.16 % | 40.59 % | 12.03 %

17.64 % | 48.74 % | 31.23 % | 2.40 %

| 1 | 2 | 3 | 4 |

ELA Performance Level

Don't make people solve a 3-D puzzle.
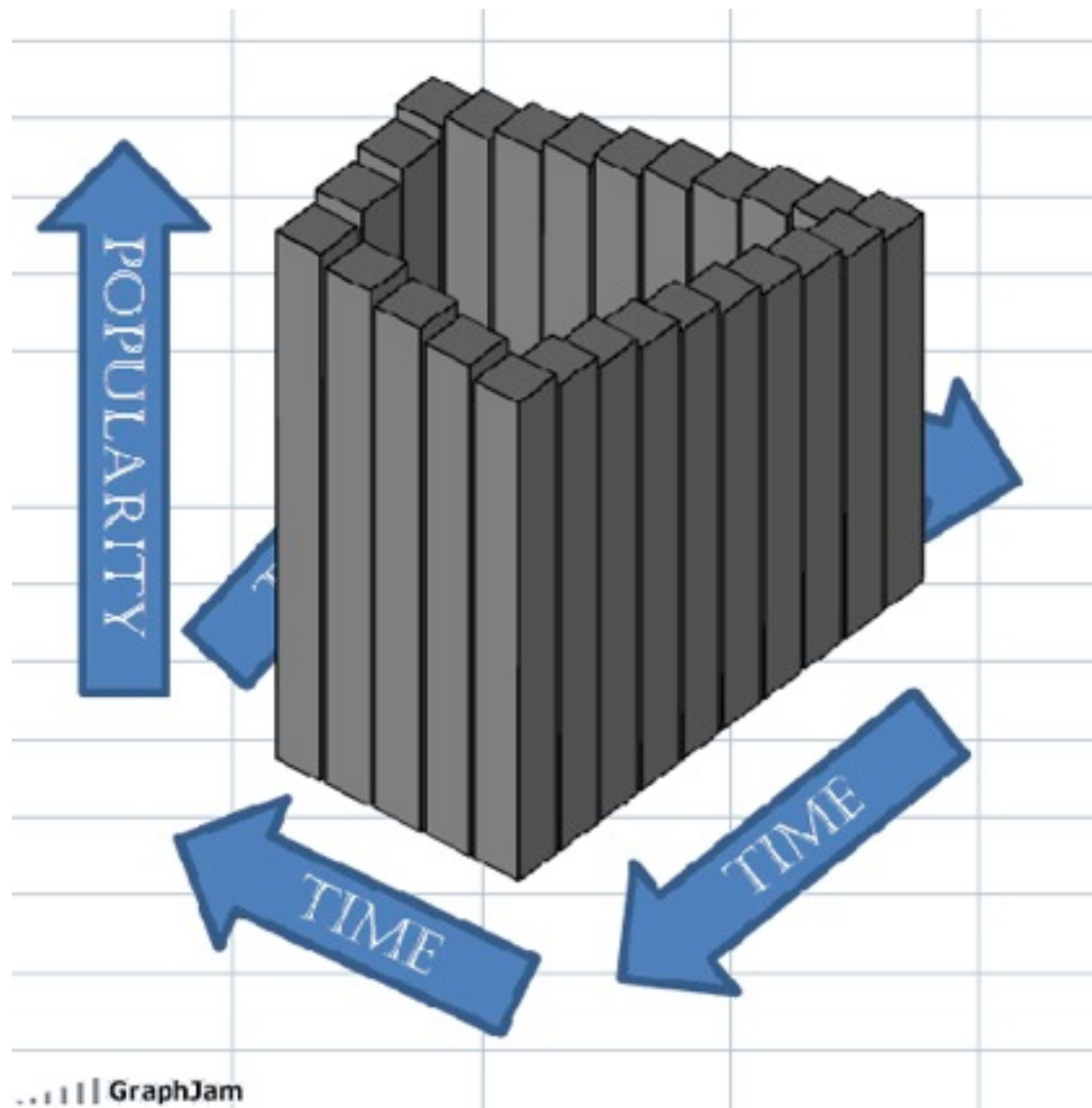
This chart of Escher's changing popularity through time was created by B. Brucker. I found it at www.GraphJam.com.

Can you ever show too much?
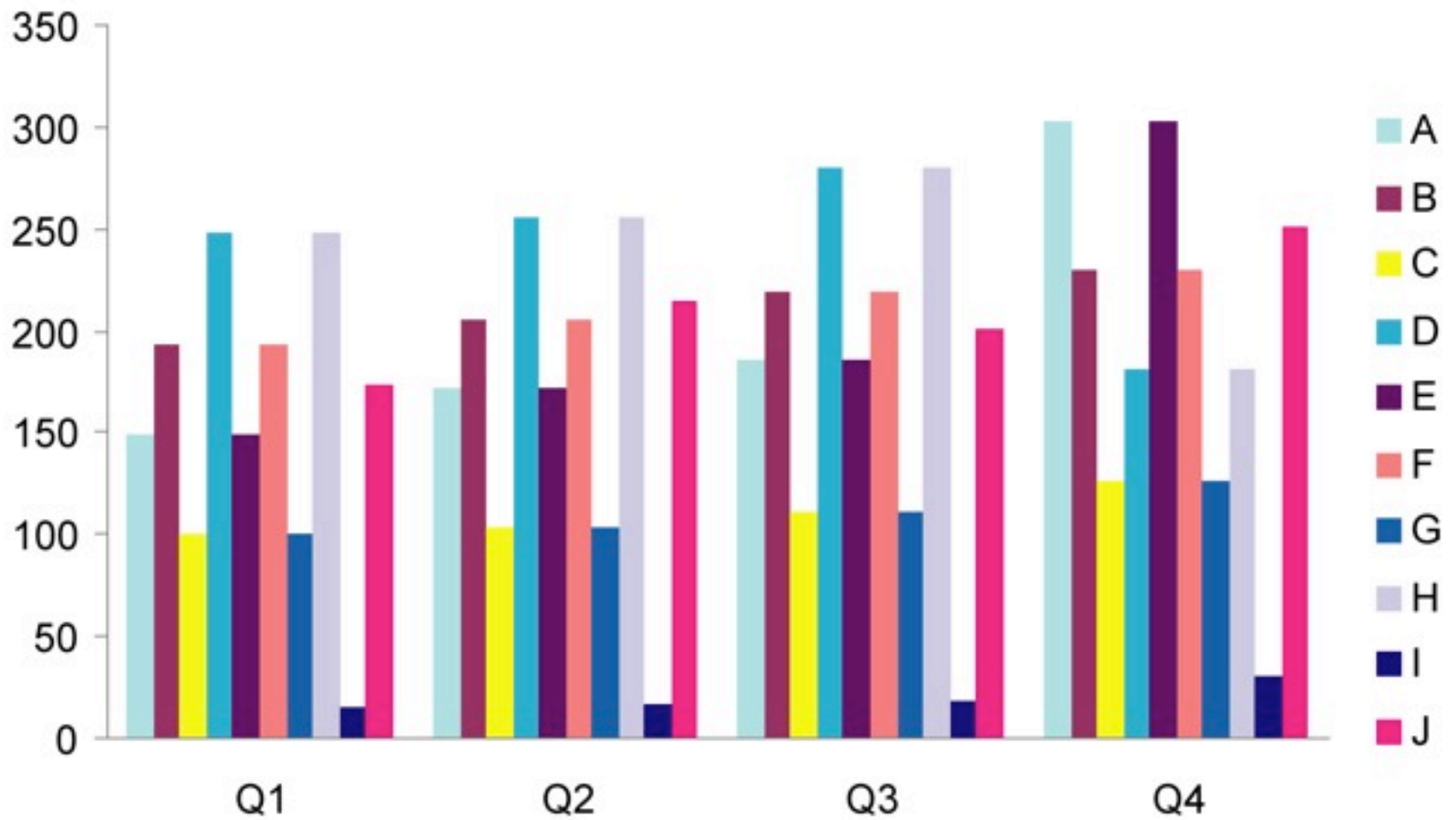
Limit the number of categorical subdivisions to around five, except when encoded by position along an axis. Short-term memory can only maintain about four chunks of information at a time.

Is more color always better?

Lots of bright color is great is you're a preschooler. For adults, frequent use of bright colors accosts visual perception. Pastels and earth tones are much easier to look at. Use bright colors only to make particular data stand out above the rest.

Color choices make a difference.
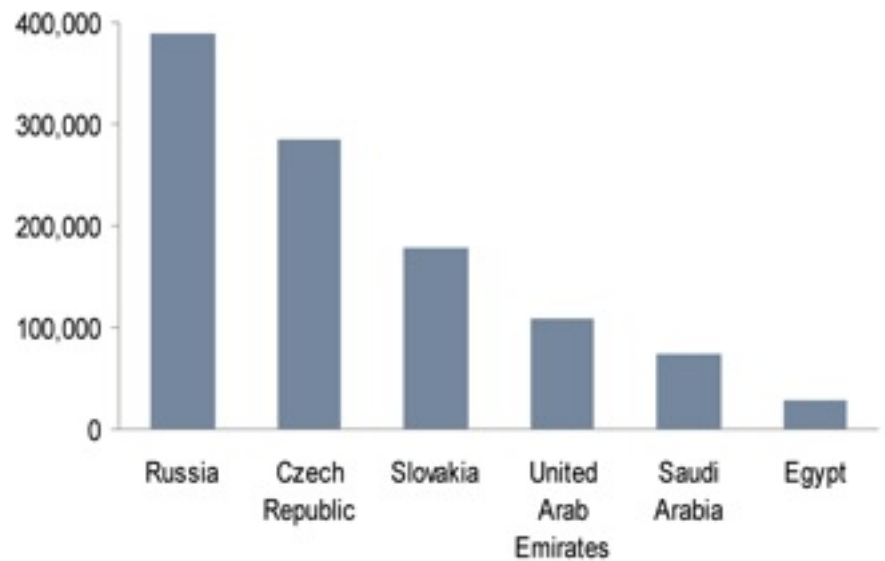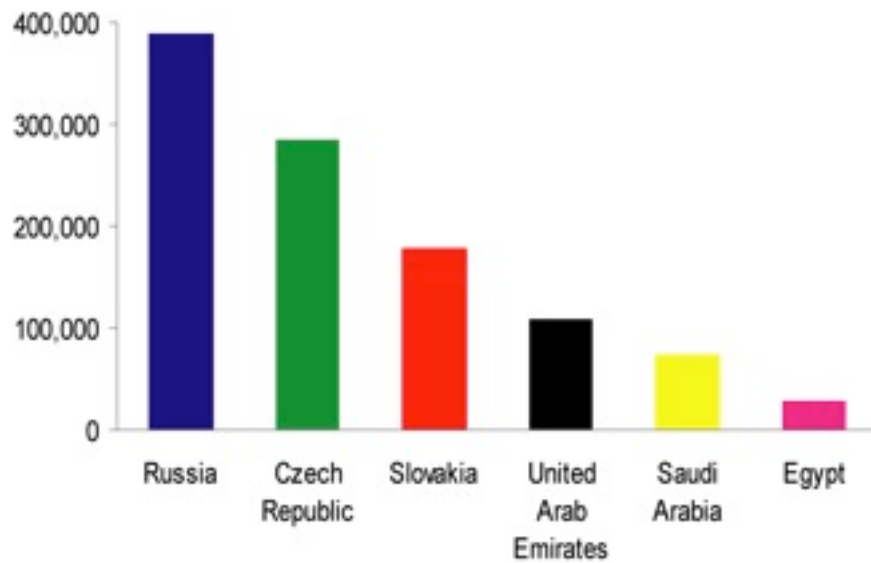
The top graph varies the colors of the bars unnecessarily. We already know that the individual bars represent different countries. Varying the colors visually separates the bars by making them look different from one another, but we want them to look alike to encourage people to compare them and to see the ranking pattern that they form as a whole.

Standardize on a good palette of colors.
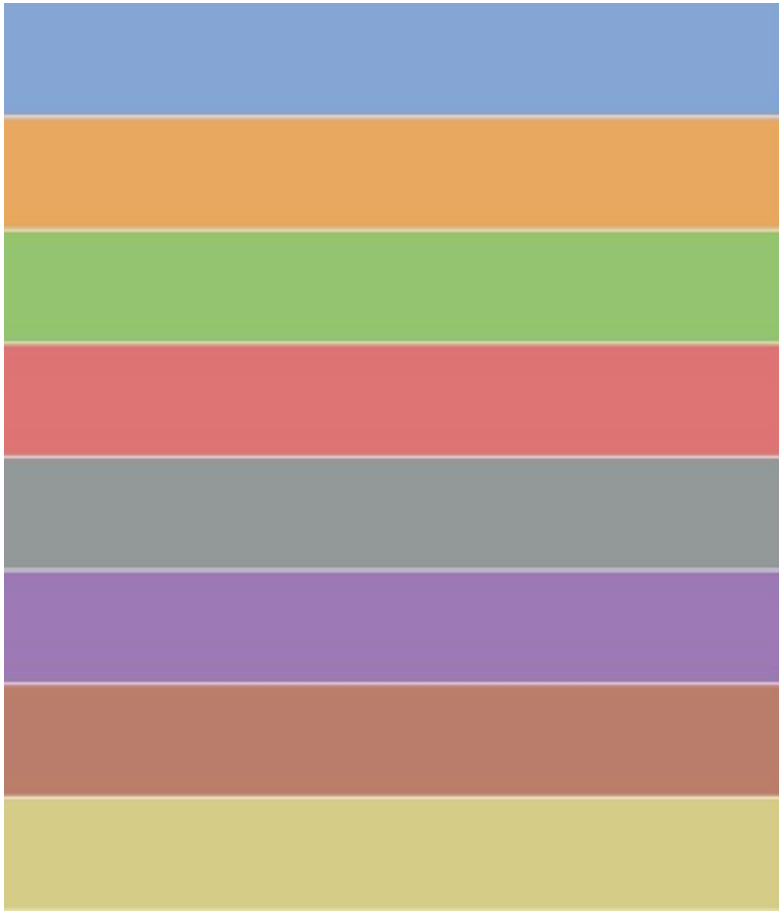
Soft, natural colors for normal use
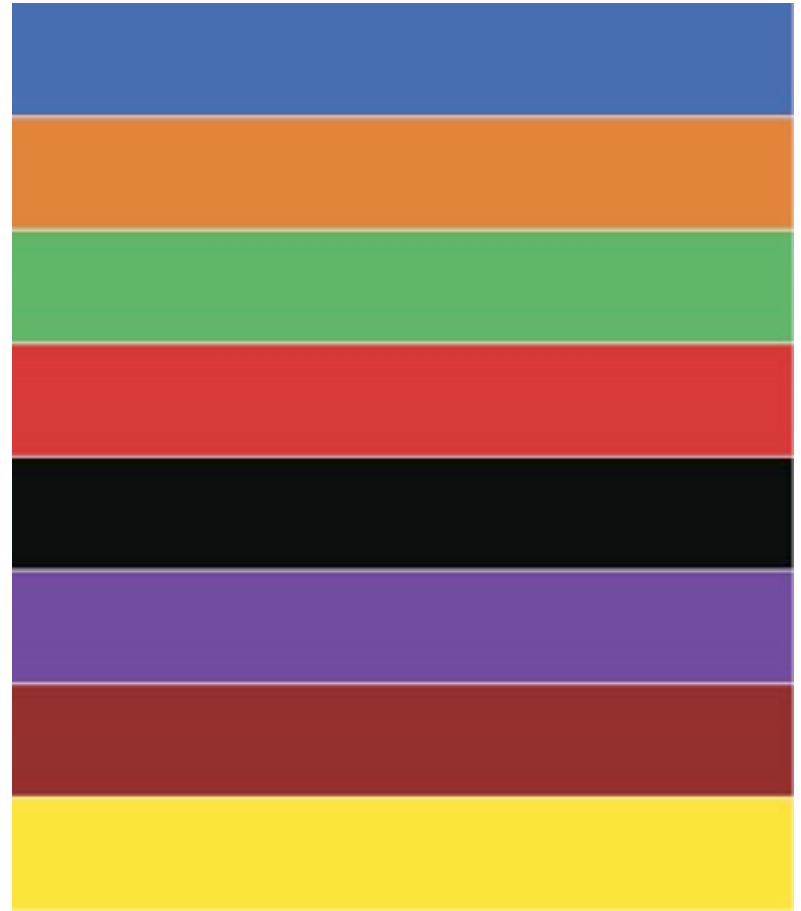
Fully saturated, dark colors for emphasis only

Soft, natural earth tones work best for everything except data that needs to stand out above the rest. Use colors that are fully saturated or dark 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.

One of the best resources for selecting effective colors for data visualization is the free Color Brewer application that was developed by Cynthia Brewer for use on maps, which can be found at www.colorbrewer.org.

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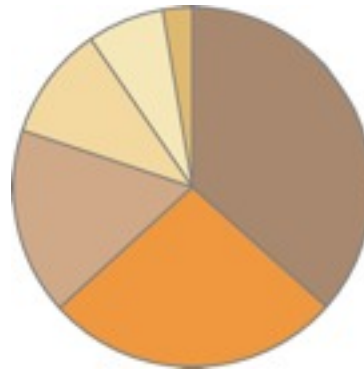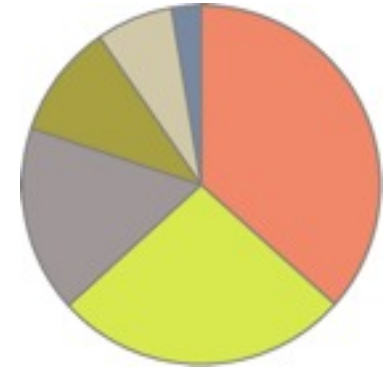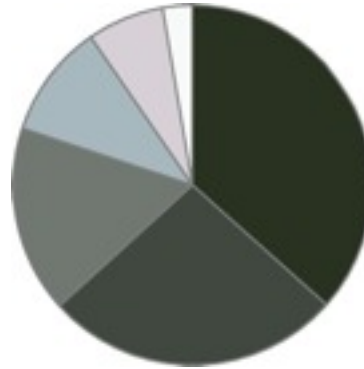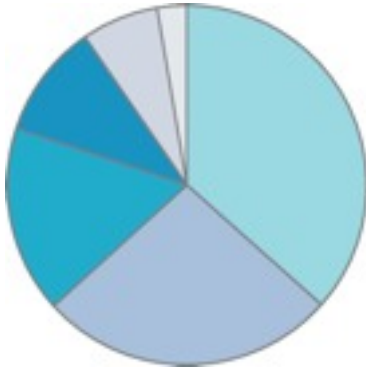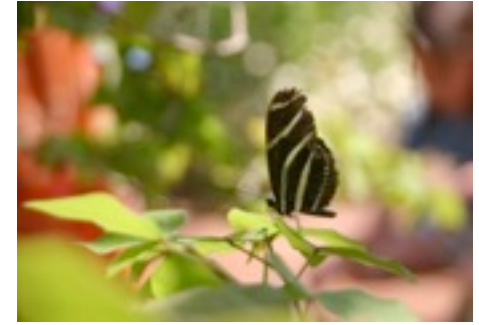
Fully saturated colors are found less often in nature than you might assume. The bright, saturated colors that appear in the bars of this graph cannot be found anywhere in these four photographs of the natural world.

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.

Which would you rather look at?
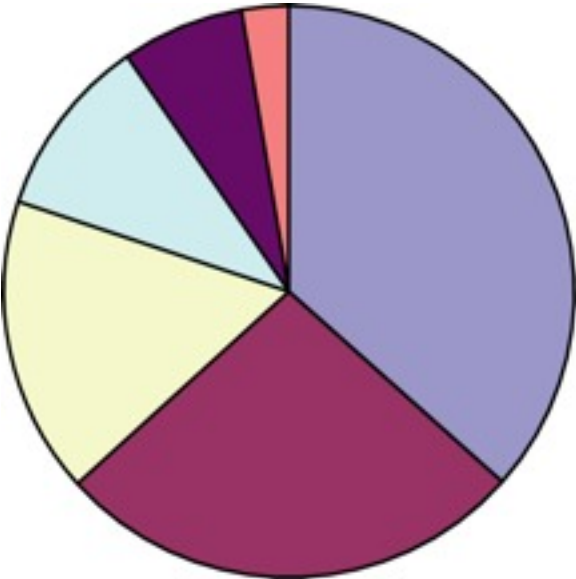
Excel 2003 defaults

Fully saturated
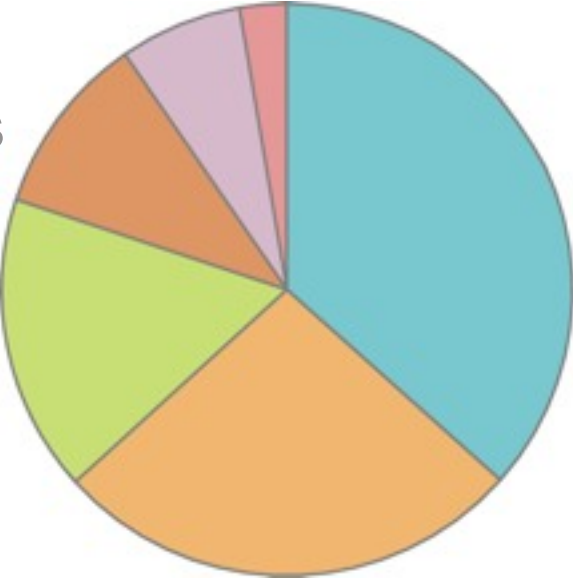
Natural

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.

Design for truth.
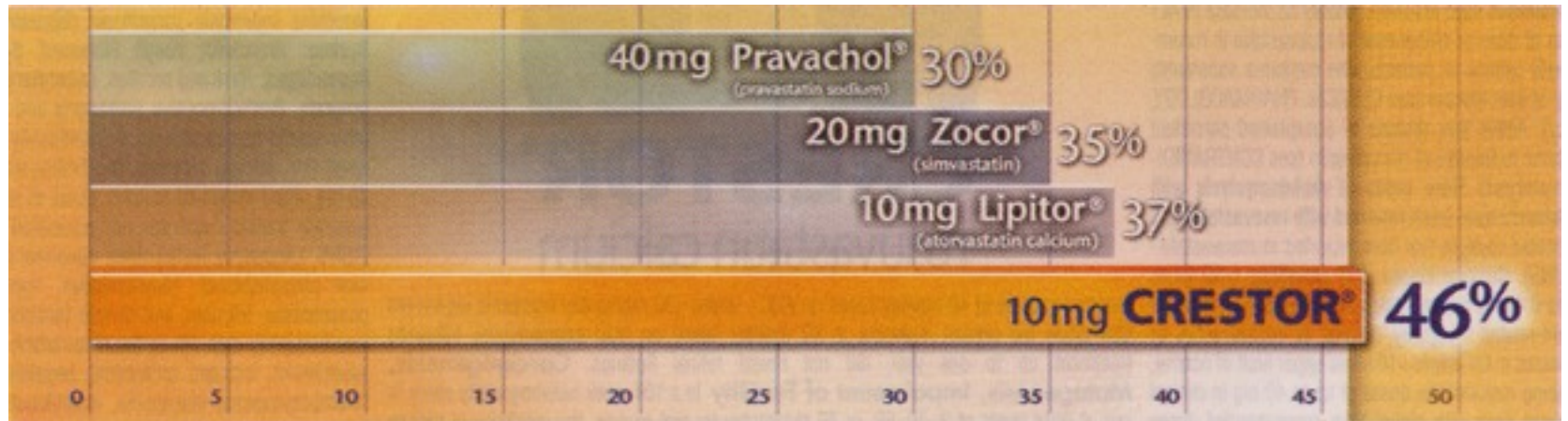
How has the effectiveness of Crestor been exaggerated?

It is appropriate to highlight a particular measure compared to others. This graph, however, does more than simply highlight Crestor; it has been visually manipulated to exaggerate its effectiveness in reducing cholesterol relative to the competition.

Highlighting and boosting the perception of Creator's effectiveness were accomplished by:

- Enlarging and boldfacing the Crestor label and 46% measure.
- Using a bright color for the Crestor bar compared to dim, transparent colors of the other bars (notice that you can see the grid lines through them).
- Using a color gradient in the background, ranging from dark orange to light orange, to increase the visual weight of the bar that extends farthest into the lighter region to the right.
- Rendering the Crestor bar alone using 3D, which add the extra dimension of depth to pump up its size.

Best location for the quantitative scale
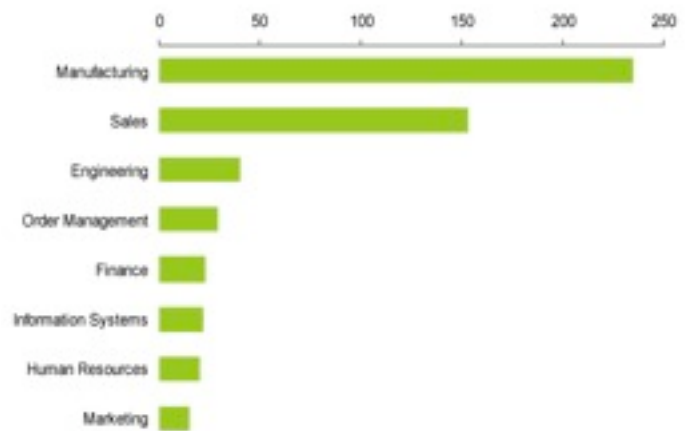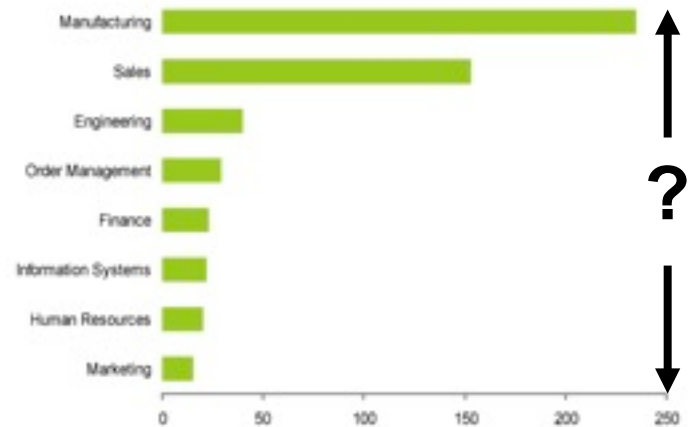
Quantitative scales are usually placed on the left or bottom axis, but there is no reason that they can't be placed on the right or top axis if that offers an advantage. As a general rule, place the quantitative scale on the axis that is closest to the most important data in the graph. For instance, in the time-series graph above, if you want to make it easier to read the December value because where the year ended is more important than where it began, then the scale belongs on the right axis. If both ends of the graph are equally important and the graph it is particularly wide or tall such that some data would be far from the quantitative scale if you had to pick a side, you can place the quantitative scale on both sides (left and right or top and bottom).

Best range for the quantitative scale

Whenever bars are used to encode values, the range of the quantitative scale must include zero. This is because the length of the bar encodes its quantity, which won't work without zero. Bars that encode positive values extend up (vertical bars) or to the right (horizontal bars) and those that encode negative values extend either down or to the left.

When lines or points are used to encode the values, however, the quantitative scale can be narrowed, for zero isn't required. Often, there is an advantage to setting the range of the quantitative scale to start just below the lowest value and end just above the highest value, thereby filling the data region of the graph with values without wasted space where no values exist. When this is done, however, you must make sure that people know it.

Why bother with horizontal bars?

Top Sales Representatives
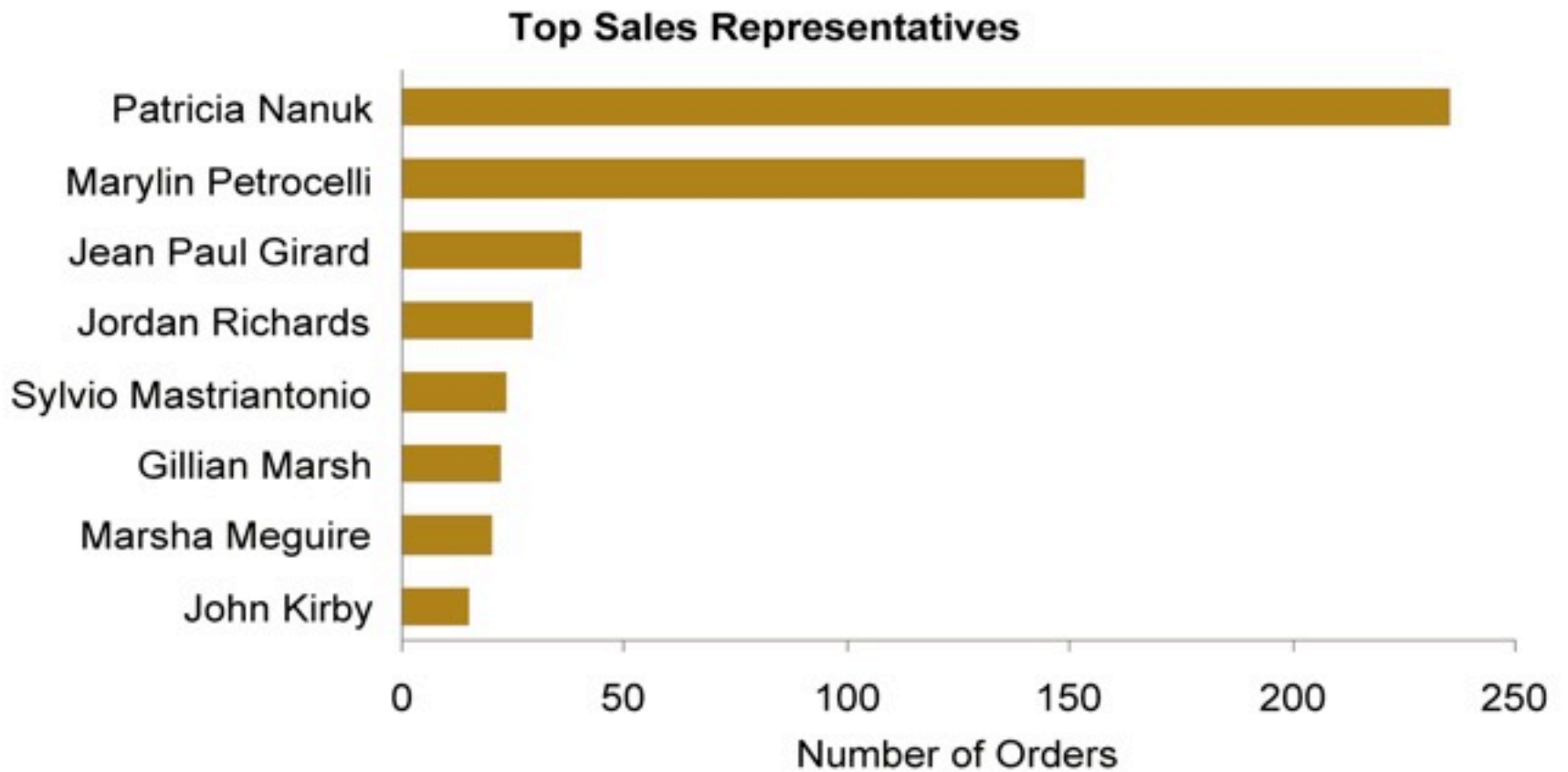
Long and/or large numbers of categorical labels fit better along the vertical axis.

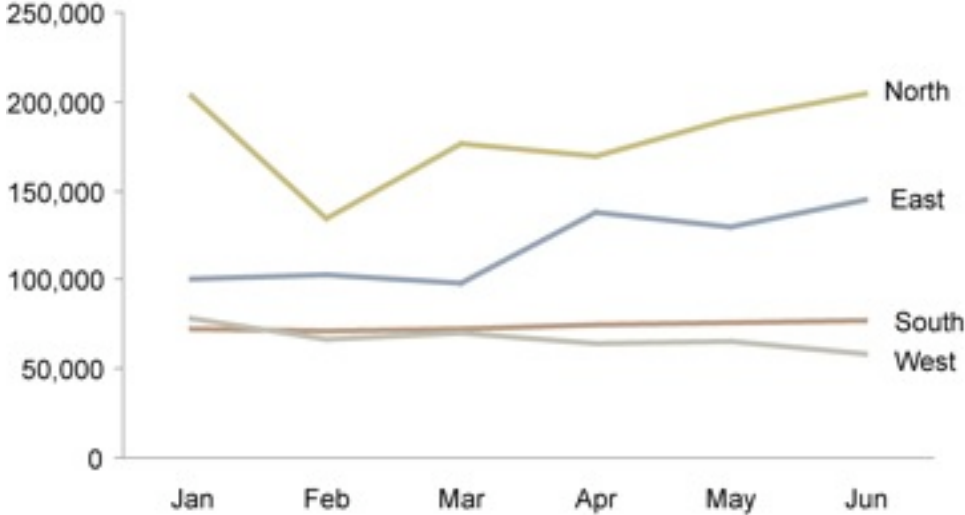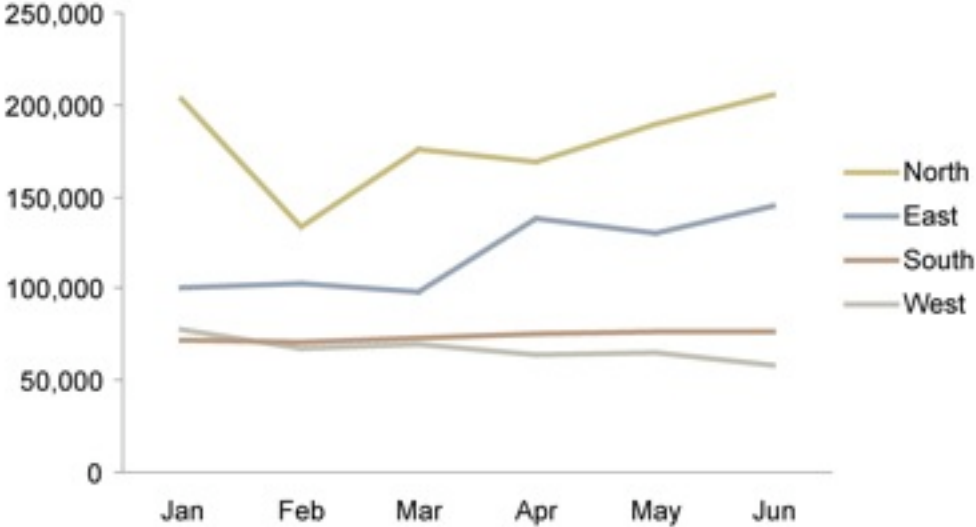# How can you minimize problems associated with legends?
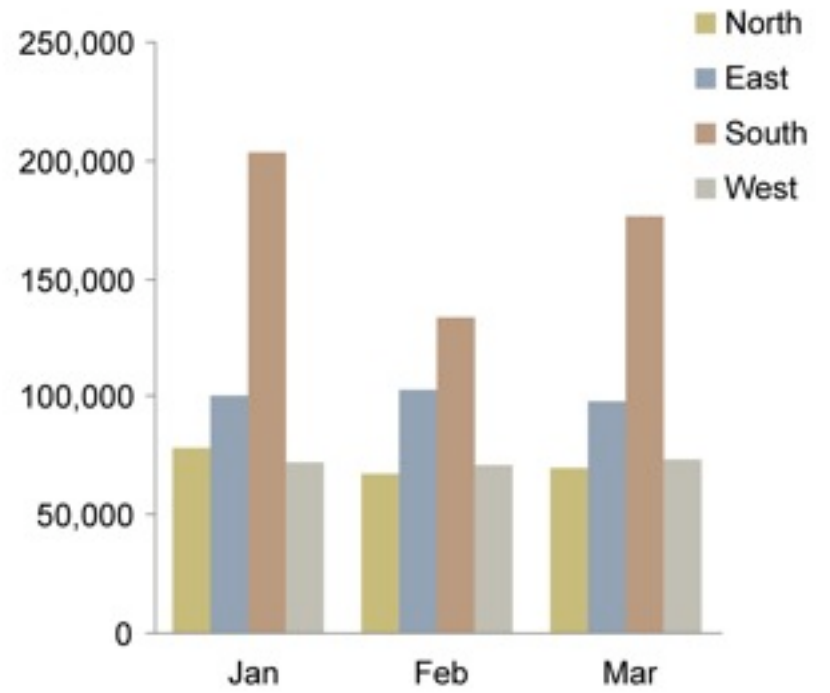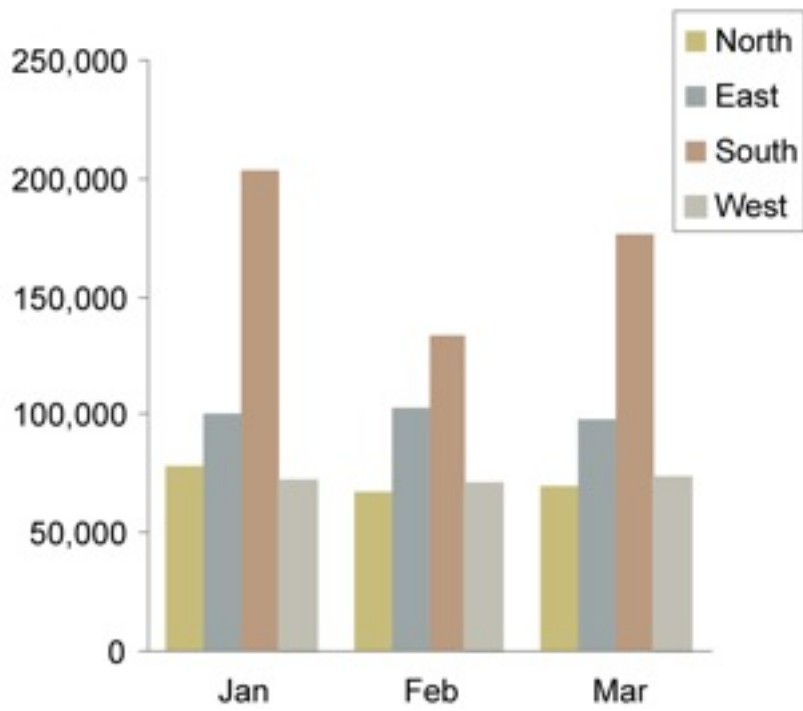
Borders around legends rarely add value.

How else can you improve legends in bar graphs?

2002 Sales

U.S. $

Legend: North, East, South, West
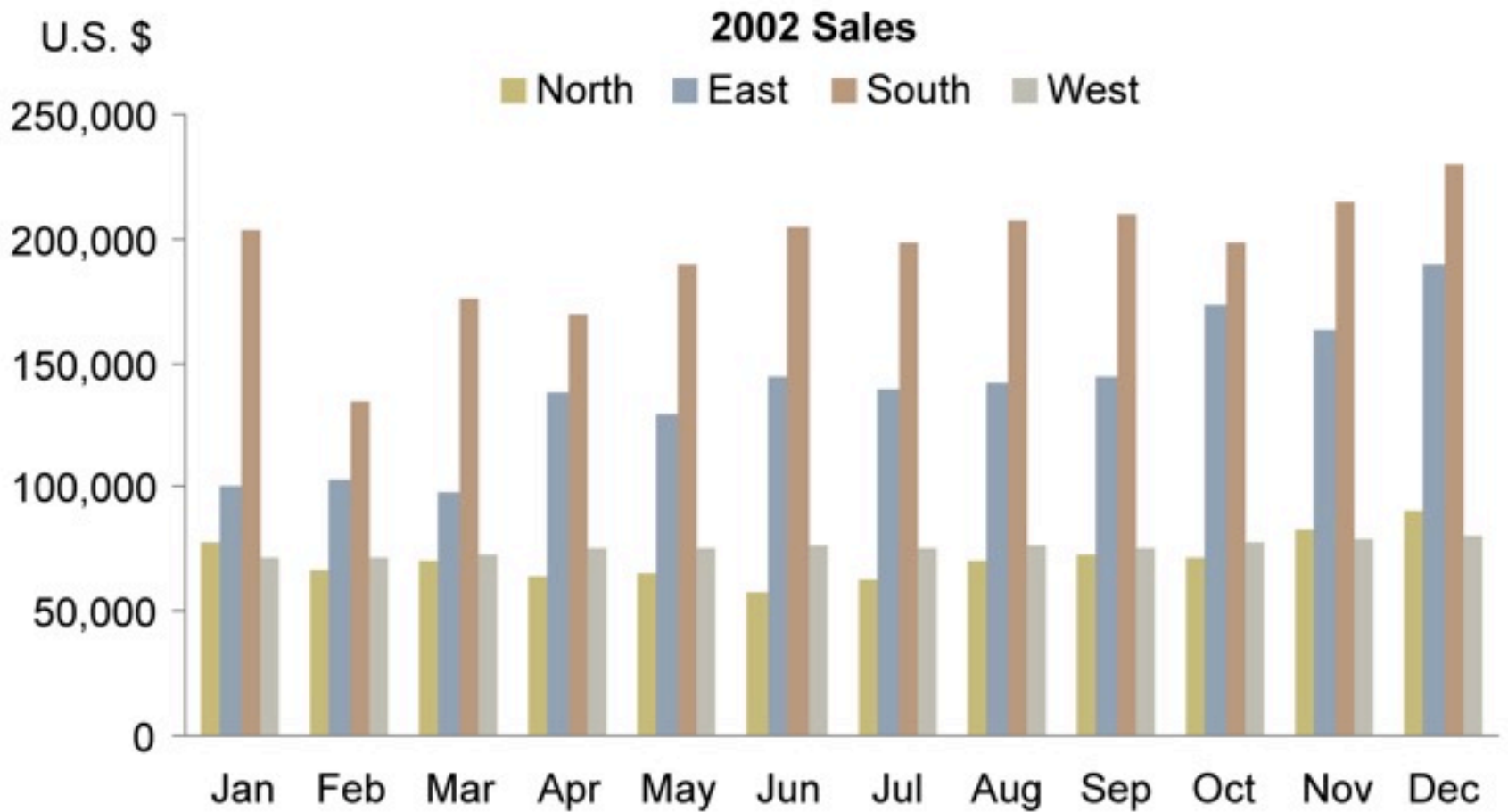
You can arrange the labels horizontally in the same order as the bars. This links the legend more tightly with the bars.

What about legends for line graphs?

Place the labels next to the corresponding lines whenever possible. This saves the reader considerable time.

How can you display additional variables?
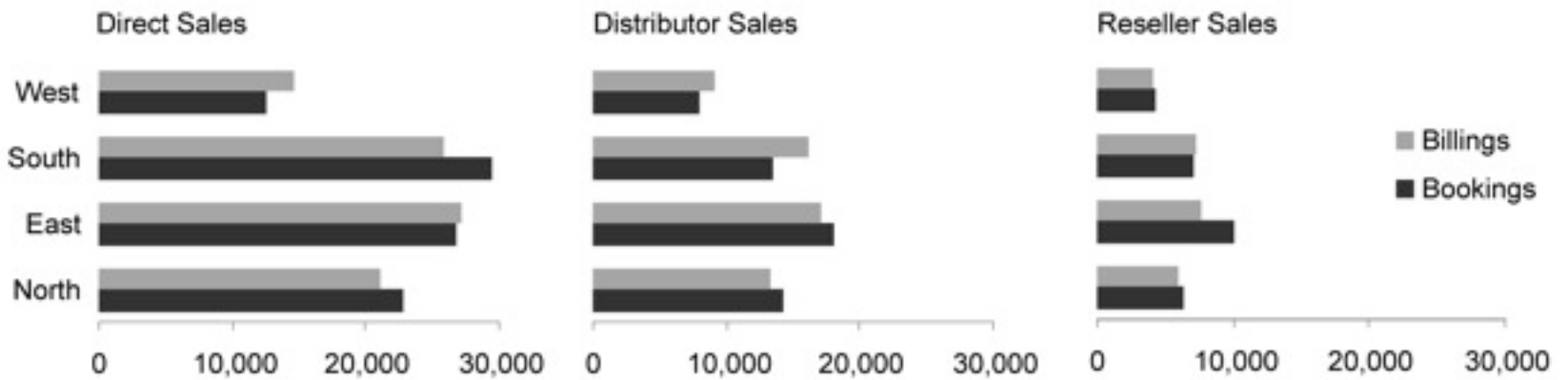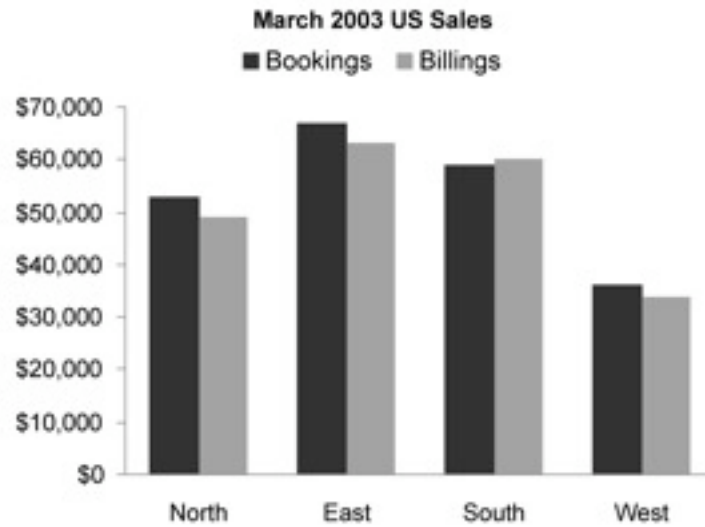
For instance, let's say you want to break the data in the top graph into three sales channels per region. You can do this by displaying a series of related graphs, each representing a different instance of the variable.
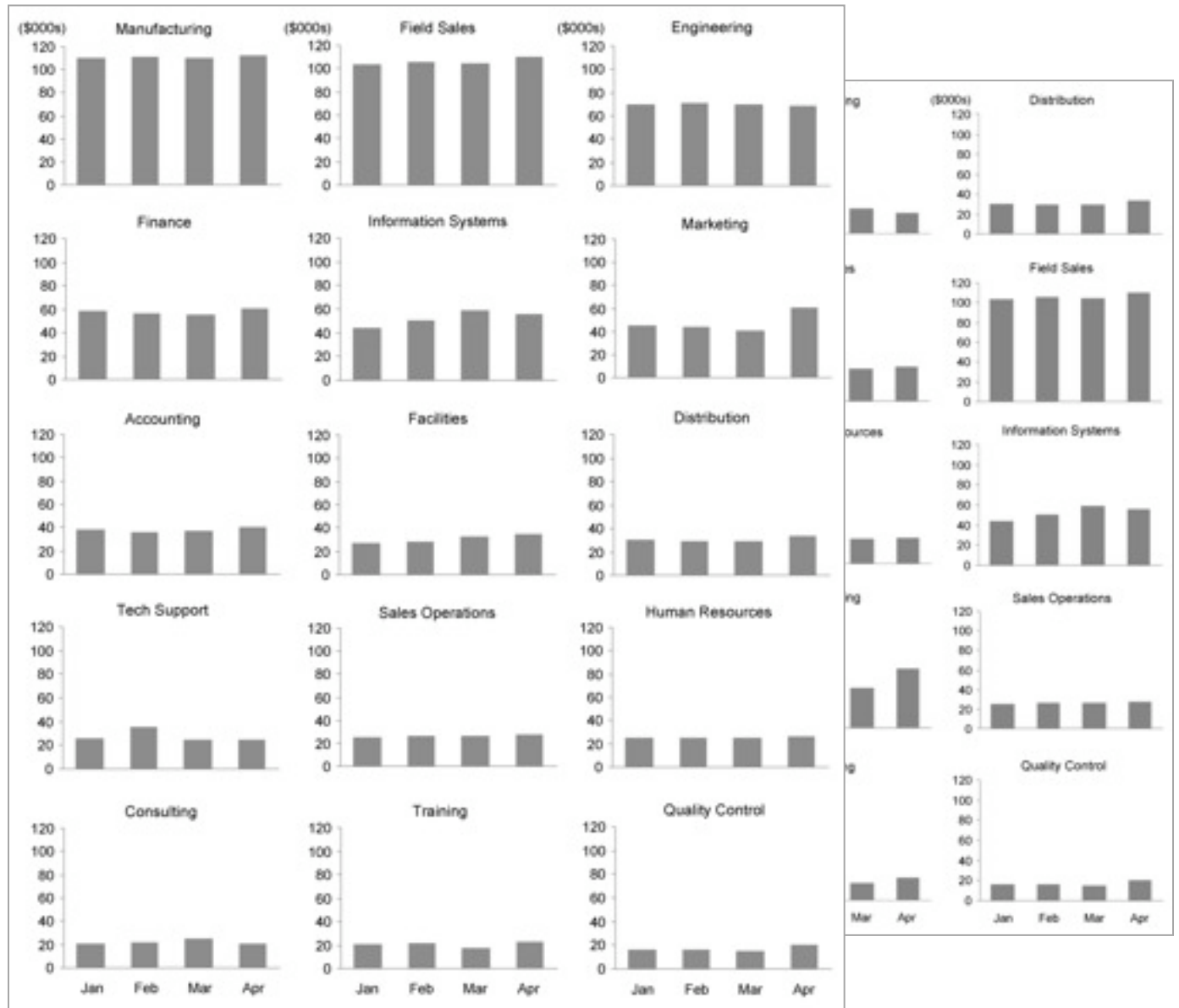
You can display many graphs within a single eyespan.

- Vary nothing but a single variable
- Arrange graphs in a logical sequence
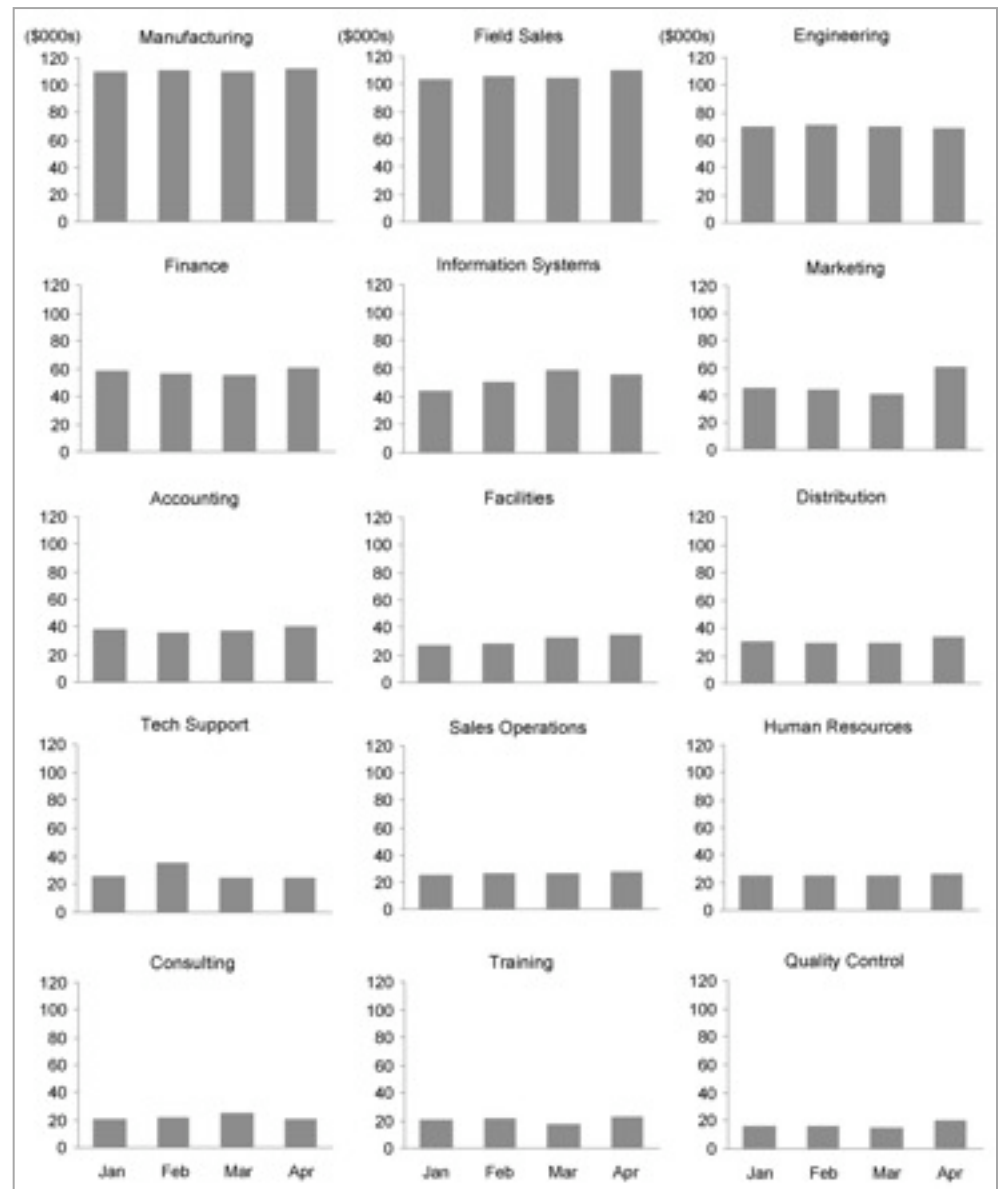
A series of related graphs arranged in a row, column, or matrix is what Edward Tufte calls "small multiples."

The order really does make a difference!

It's amazing what a significant difference sorting the graphs makes. When sorted by value, they are much easier to compare.

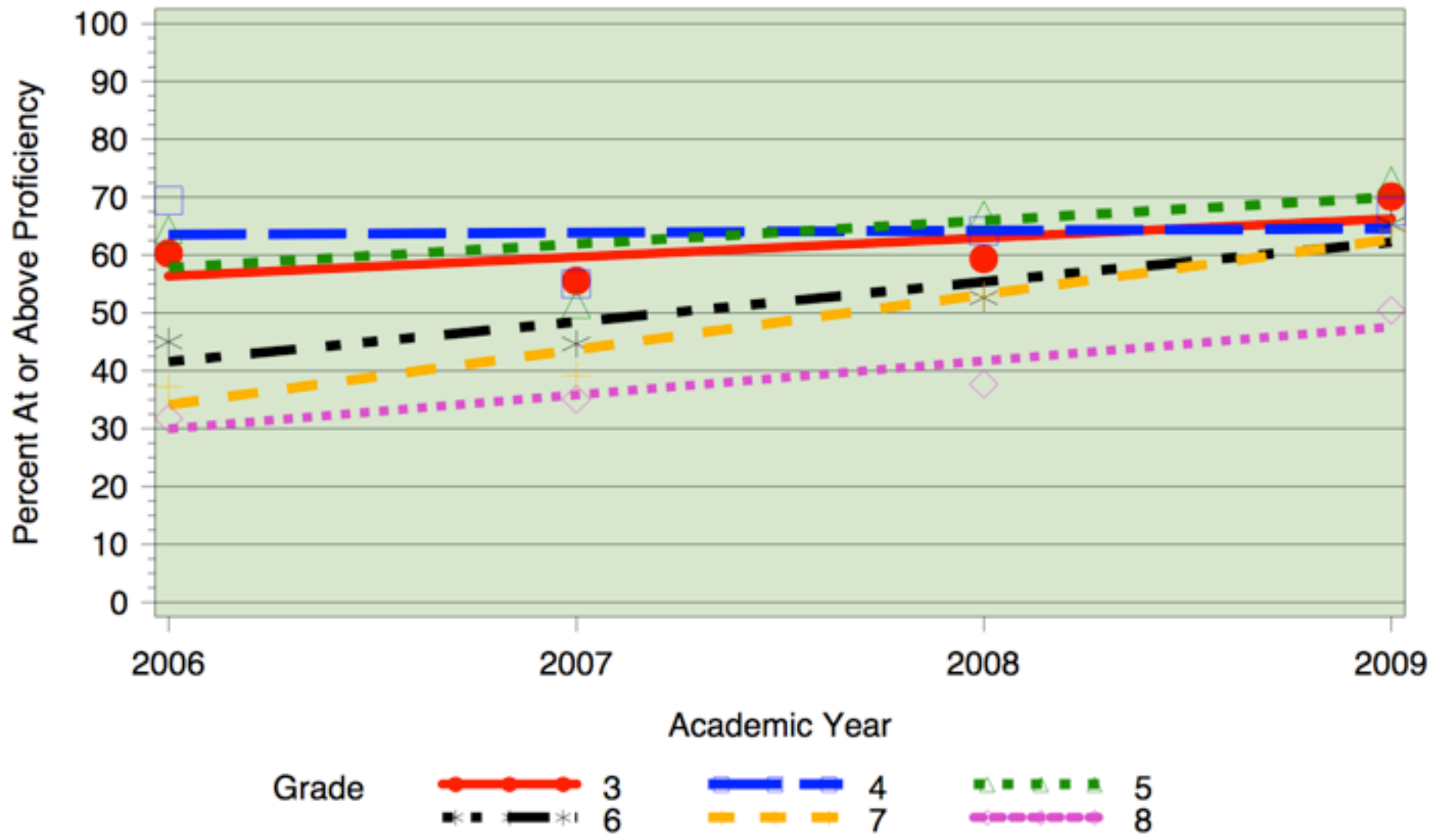# Enrollment by Ethnicity, Texas Public Schools, 1997-98 Through 2007-08
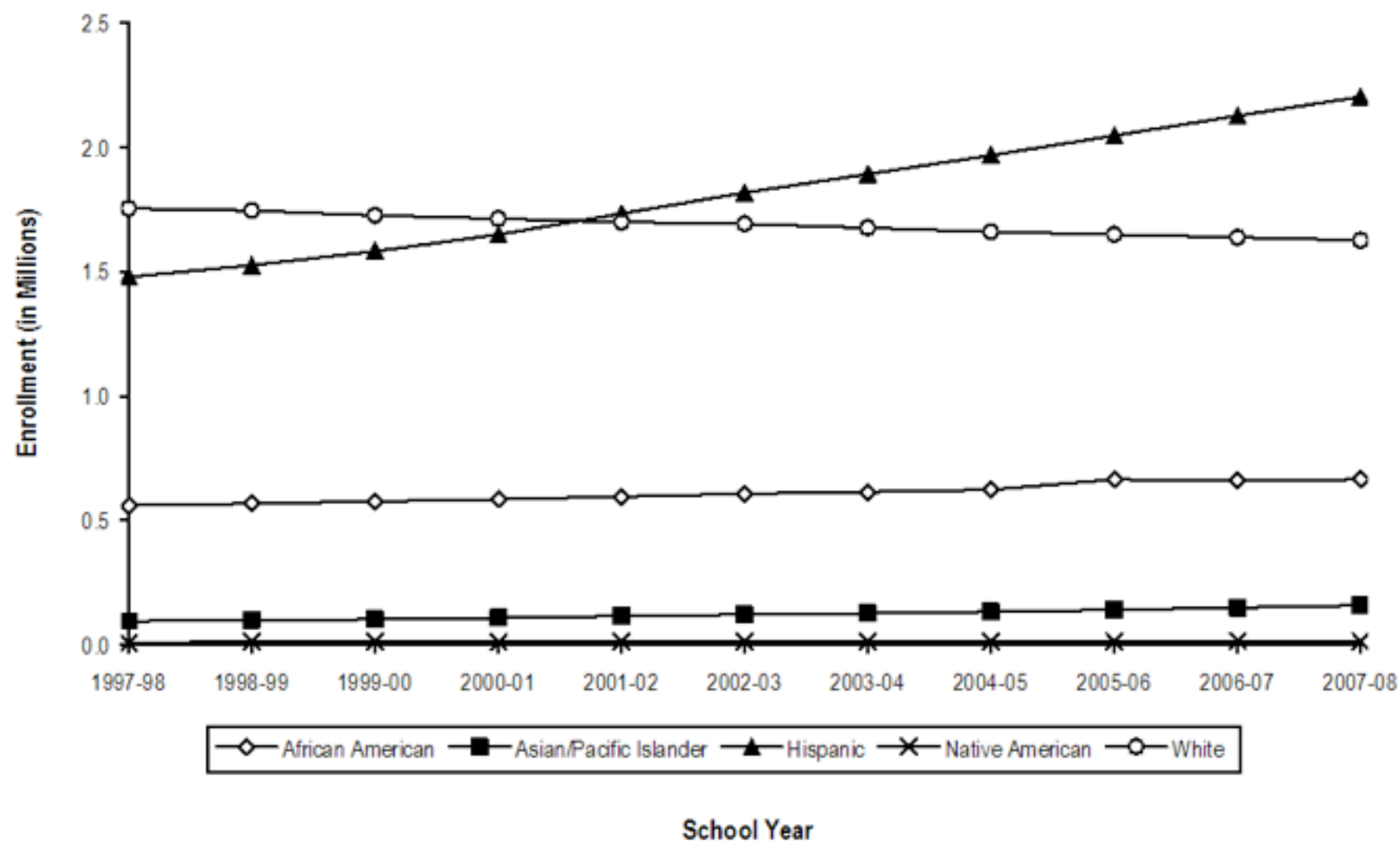
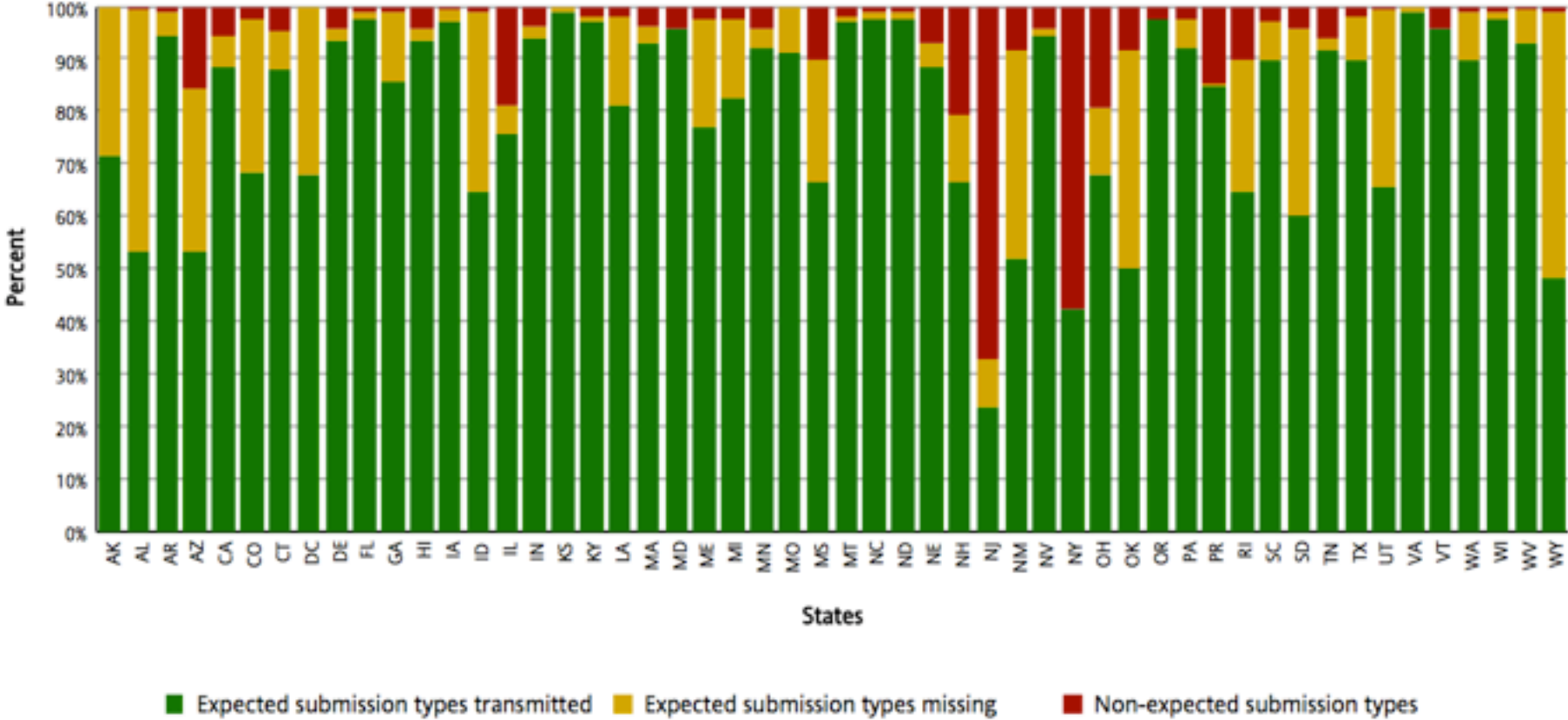Source. Enrollment in Texas Public Schools, 2007-08 (Texas Education Agency, 2009).

YONKERS PUBLIC SCHOOLS

ELA 3-8 Results, 2006 2007 2008 and 2009

DistrictWide

Percentage of expected ESS files transmitted, non-expected or missing by state

You have a choice.

The good news is, although the skills required to present data effectively are not all intuitive, they are easy to learn. The resources are available, but it won't happen unless you recognize the seriousness of the problem and commit yourself to solving it. It is up to you.