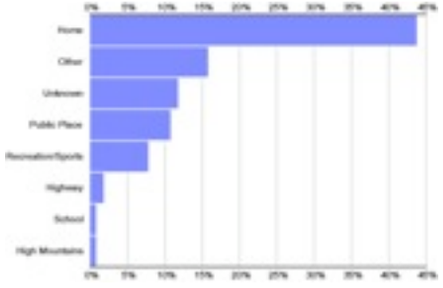
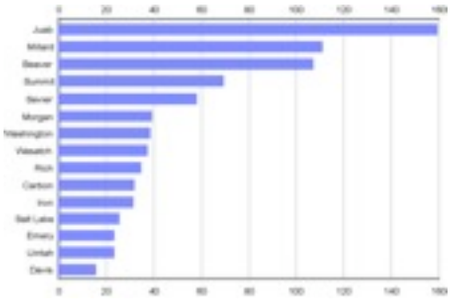
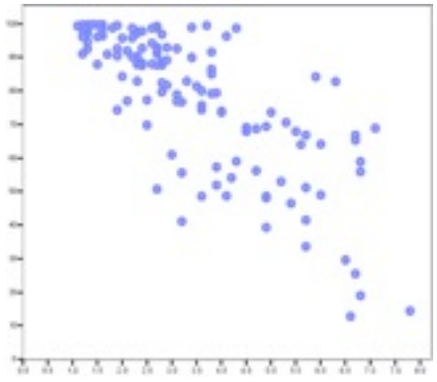
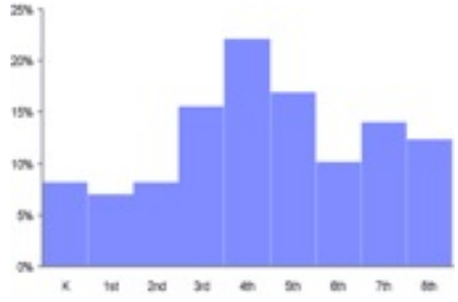
Table and Graph Design for Enlightening Communication

Stephen Few, Perceptual Edge



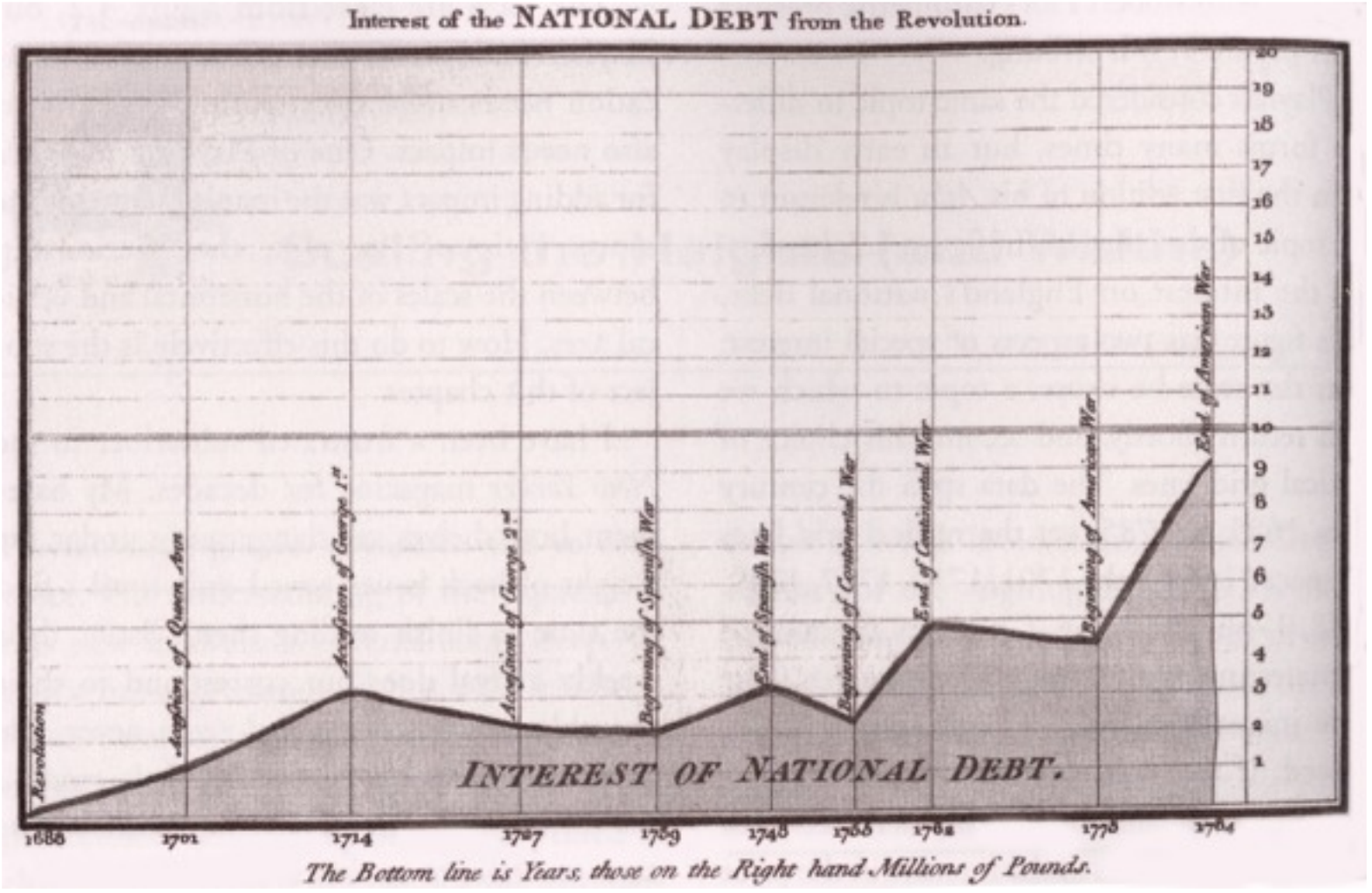






Stephen Few, Principal, Perceptual Edge
sfew@perceptualedge.com
(510) 558-7400

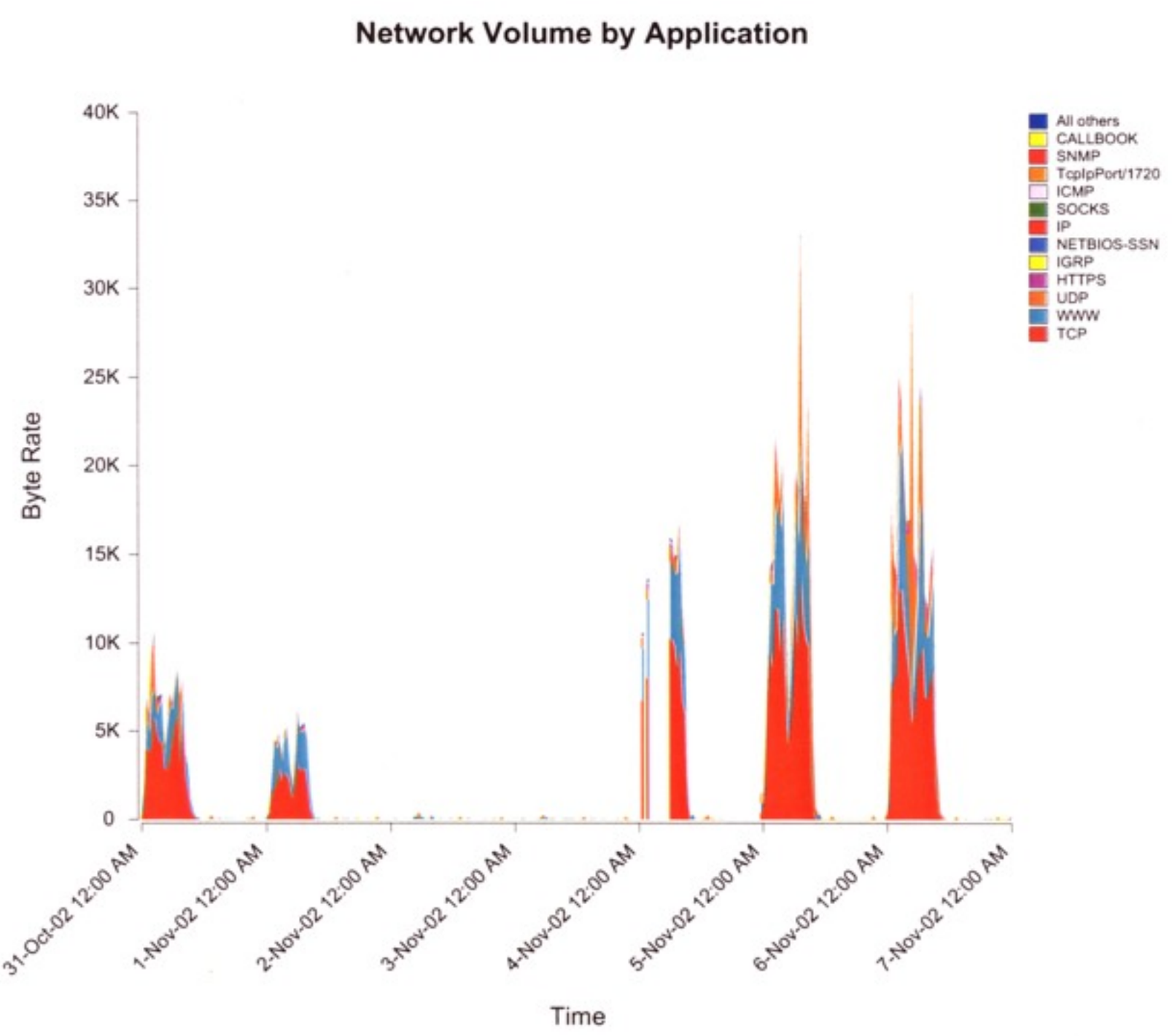
We've been telling them with graphs for quite awhile.



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.

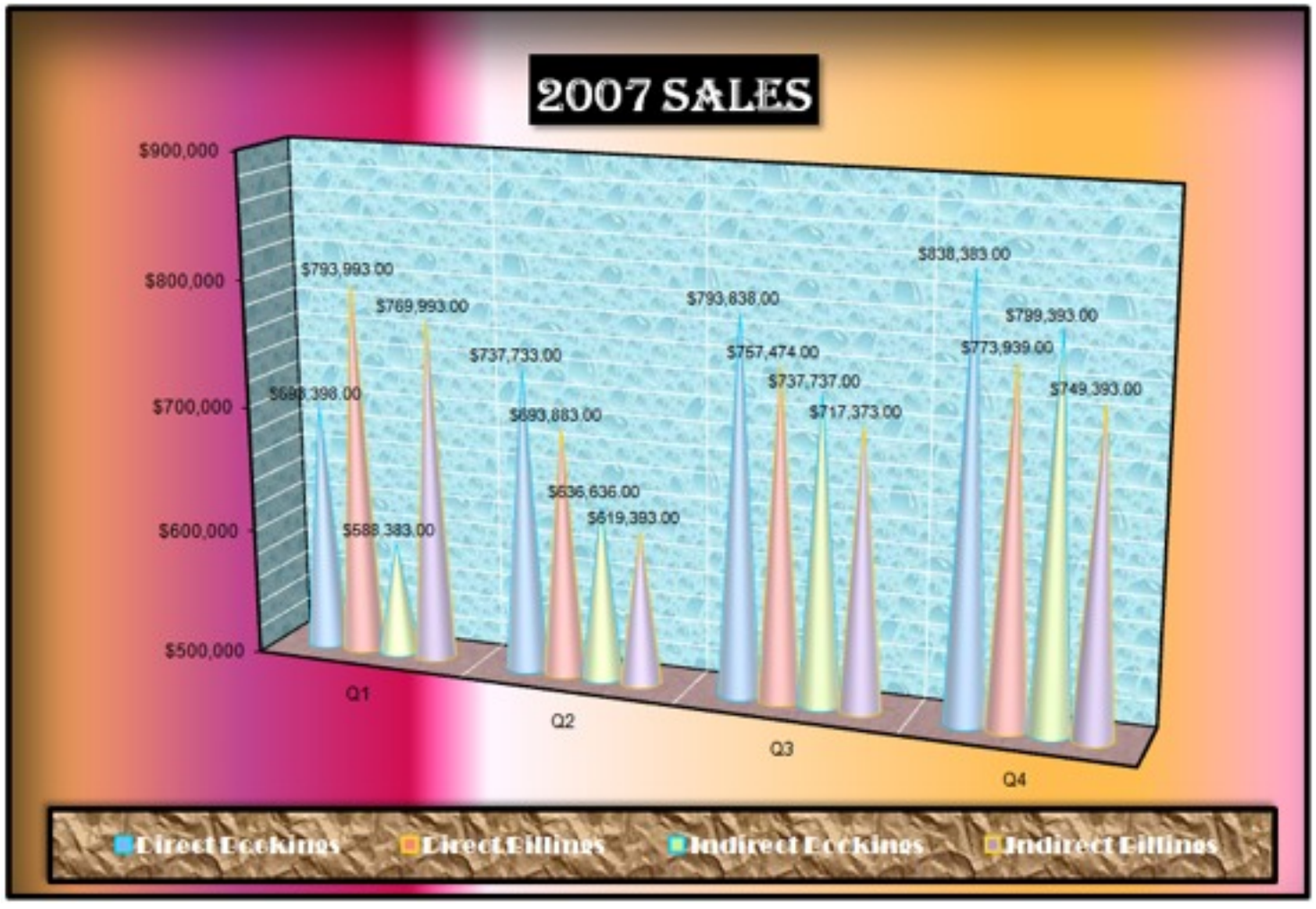
(Source: This graph was included in Playfair's *The Commercial and Political Atlas* in 1786 to make a case against England's policy of financing colonial wars through national debt.)



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.

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

Why?

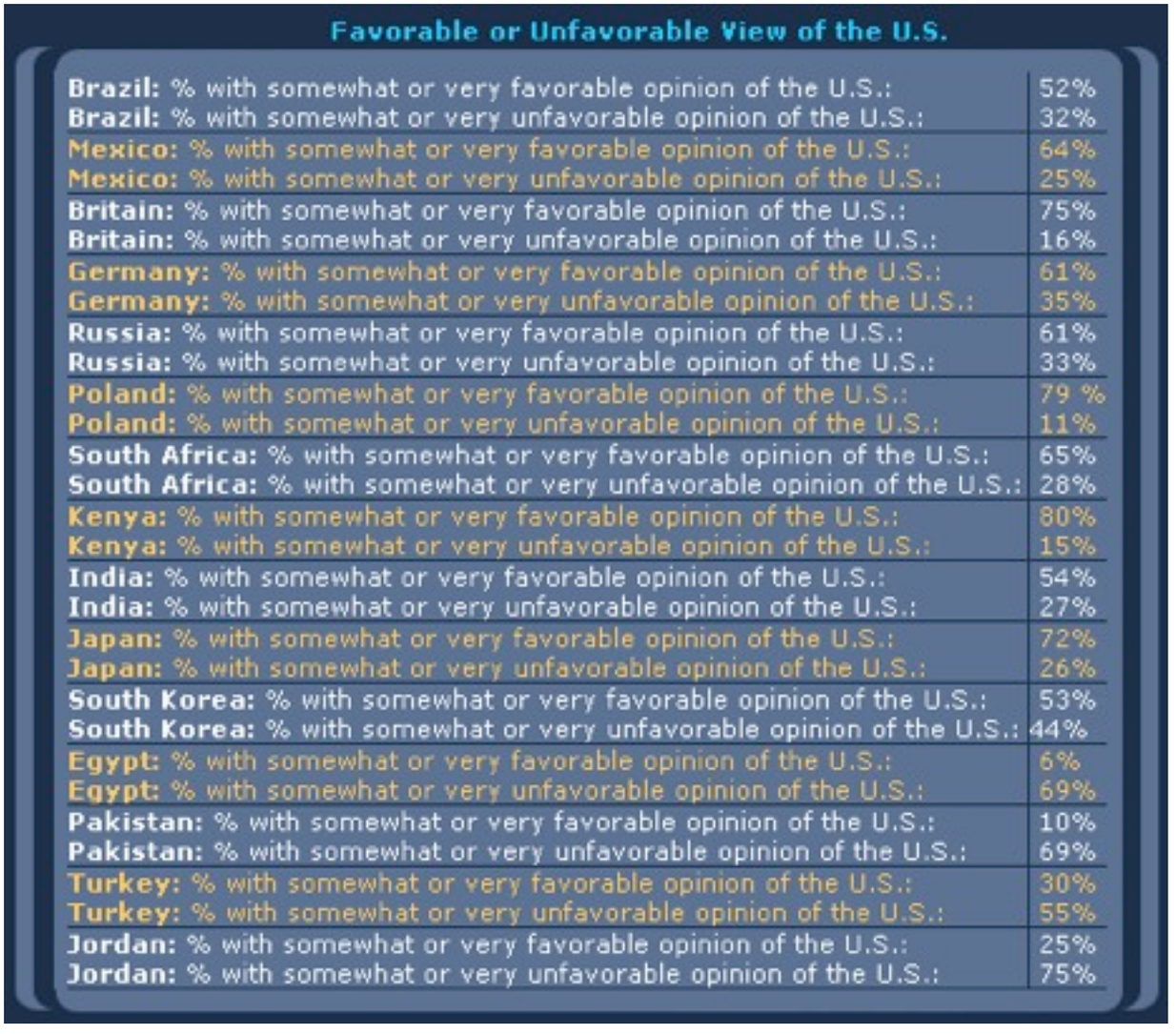


When the PC was introduced, software soon made the arduous task of table and graph creation as easy as 1-2-3 (literally “Lotus 1-2-3”, the software that was the first to legitimize the PC as a viable tool for business). Unfortunately, this improvement in ease and efficiency was not accompanied by instruction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

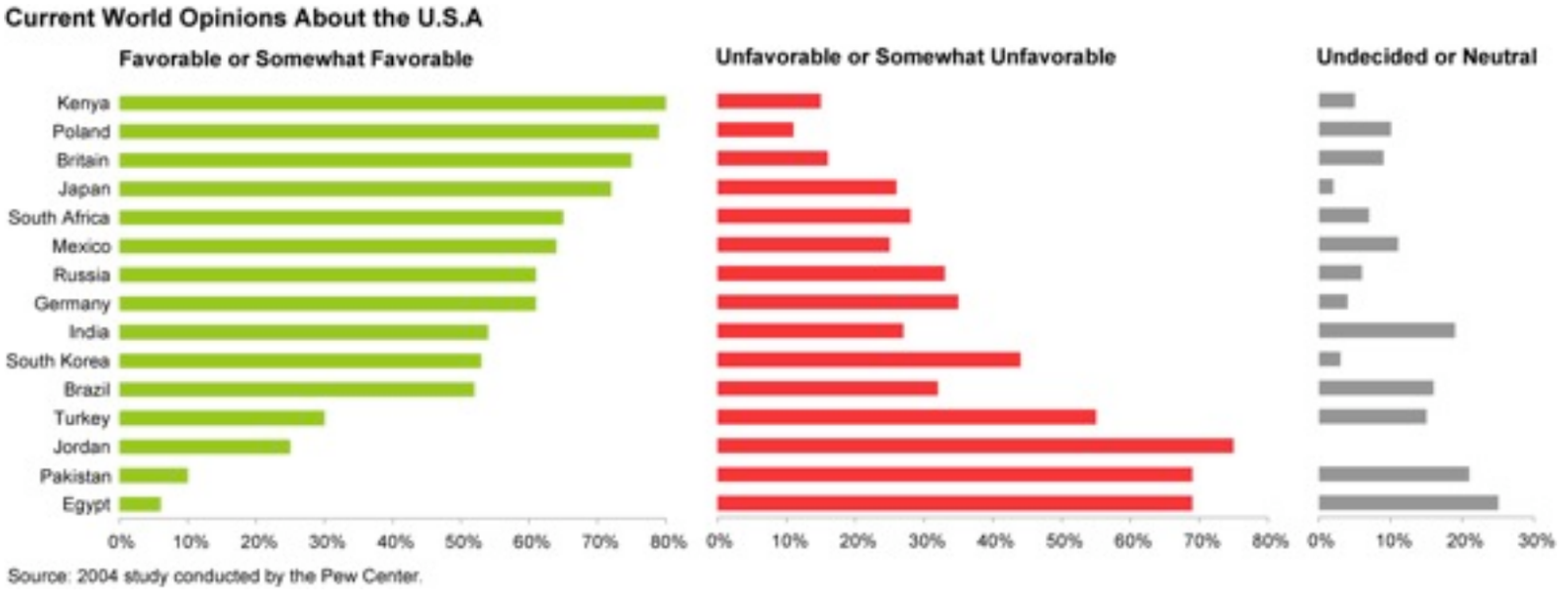


Favorable or Unfavorable View of the U.S.

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%

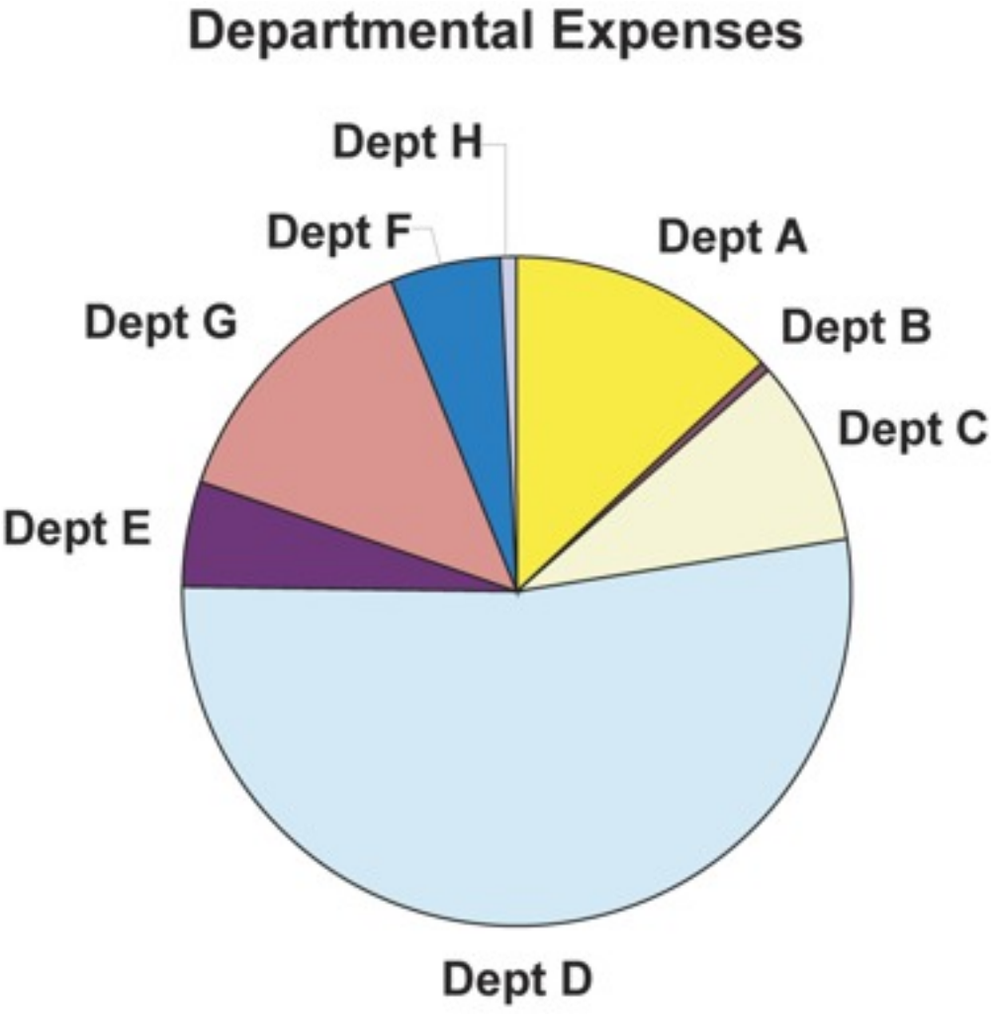
I found this table on the Web site for Bill Moyers' public television show "Now". I felt that it provided important information that deserved a better form of presentation. In this case the story could be told much better in visual form.

Example #2 - Good



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.

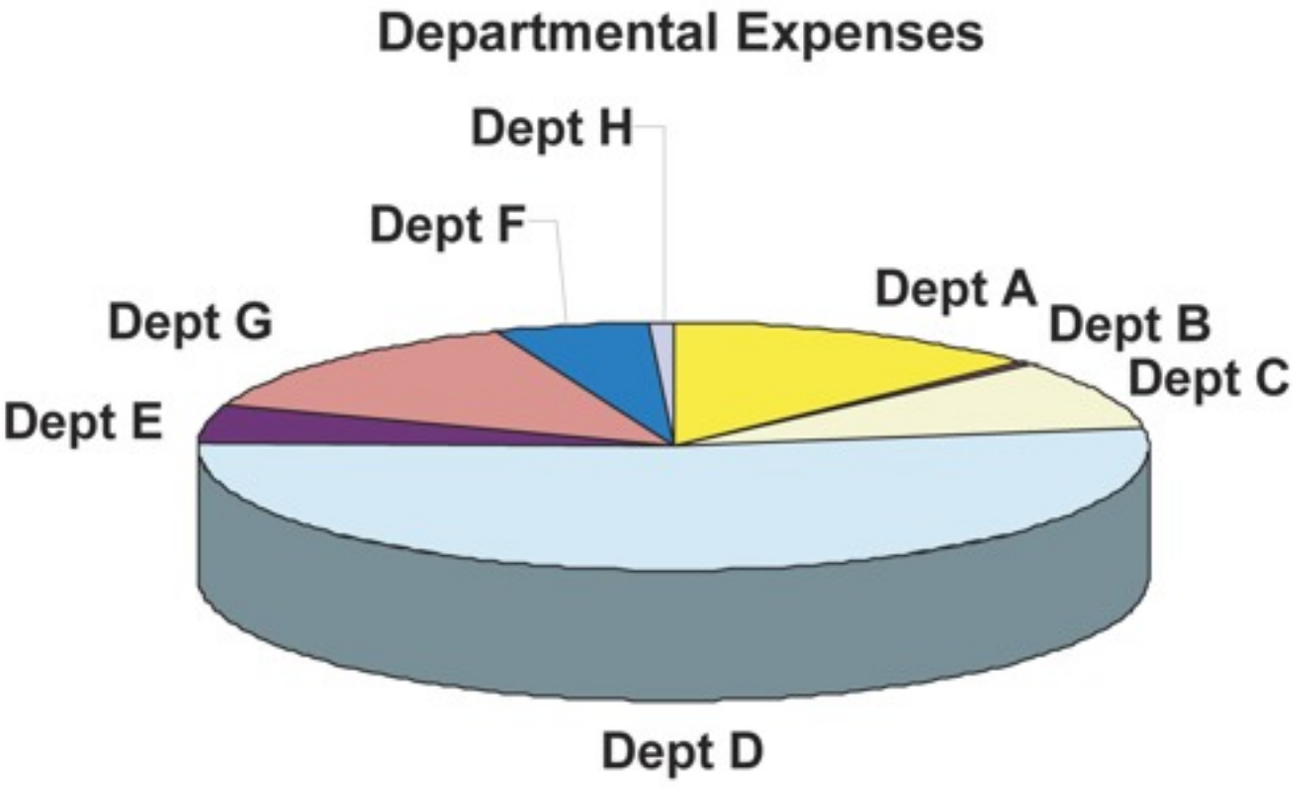
Example #1



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

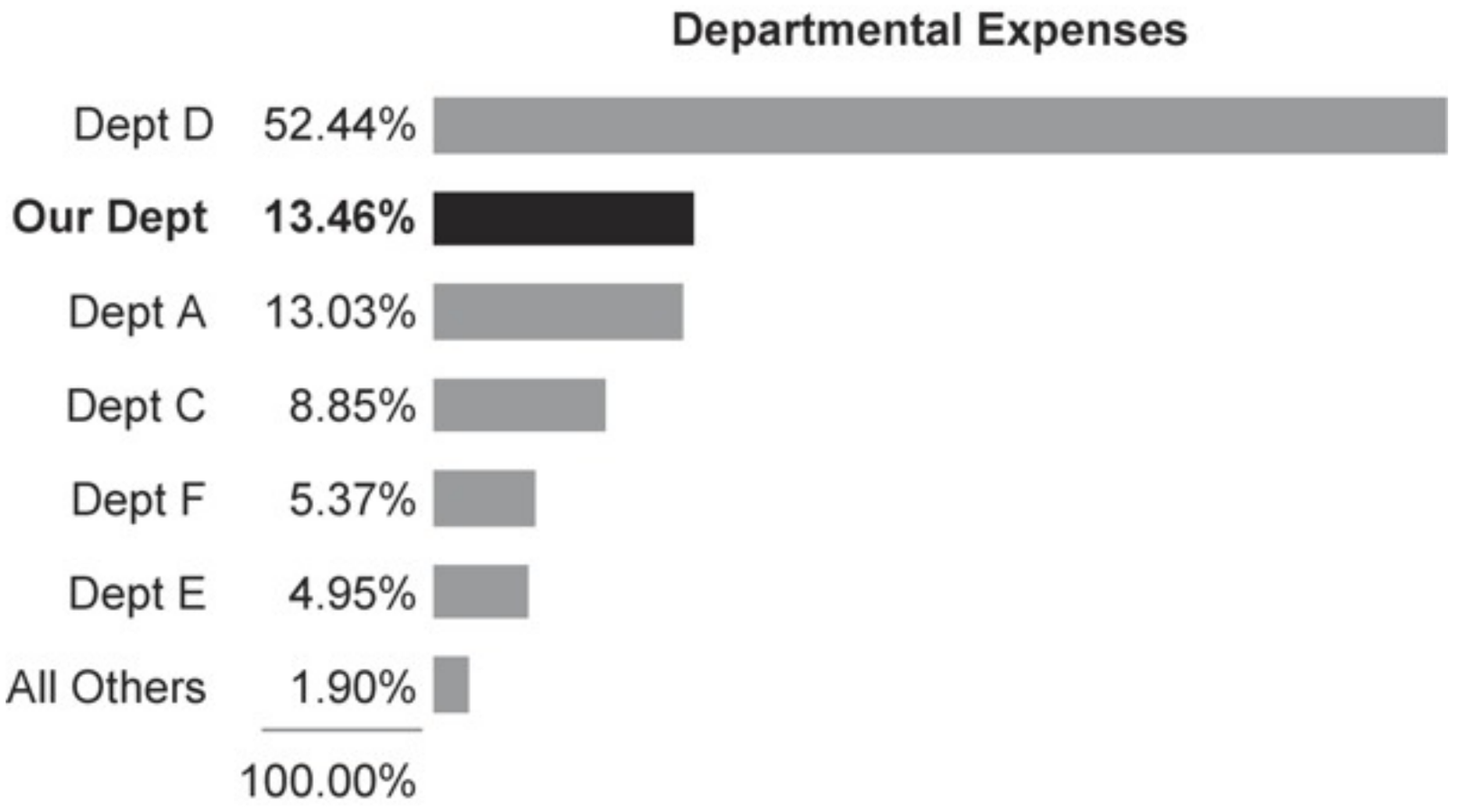
Often, when someone creates a graph that appears inadequate somehow, they try to fix it with sizzle, as in the next slide.

Example #1



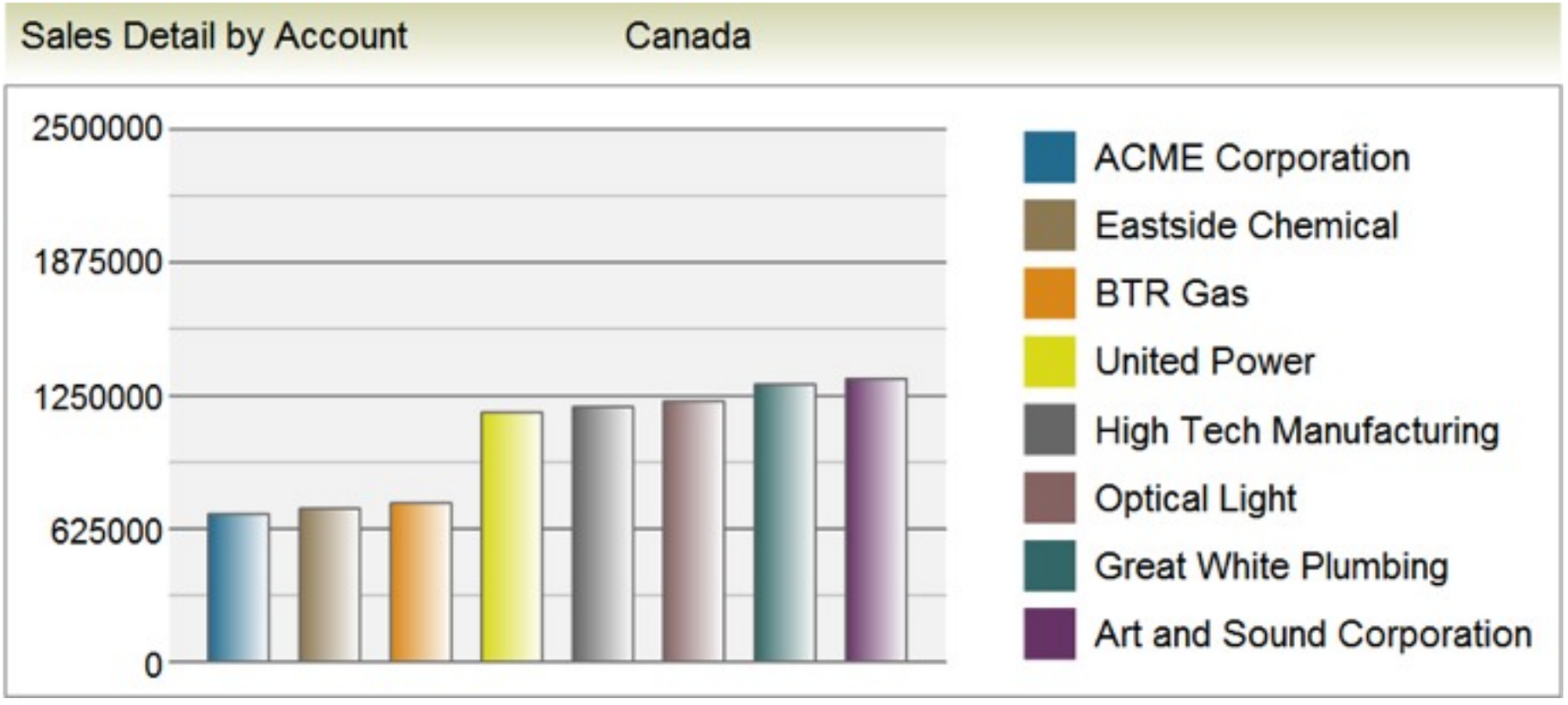
Does the addition of 3D improve this pie chart? Definitely not. In fact, it actually makes it harder to read.

Example #1 - Improved



Though it lacks flash and dazzle, this simple bar graph tells the story elegantly.

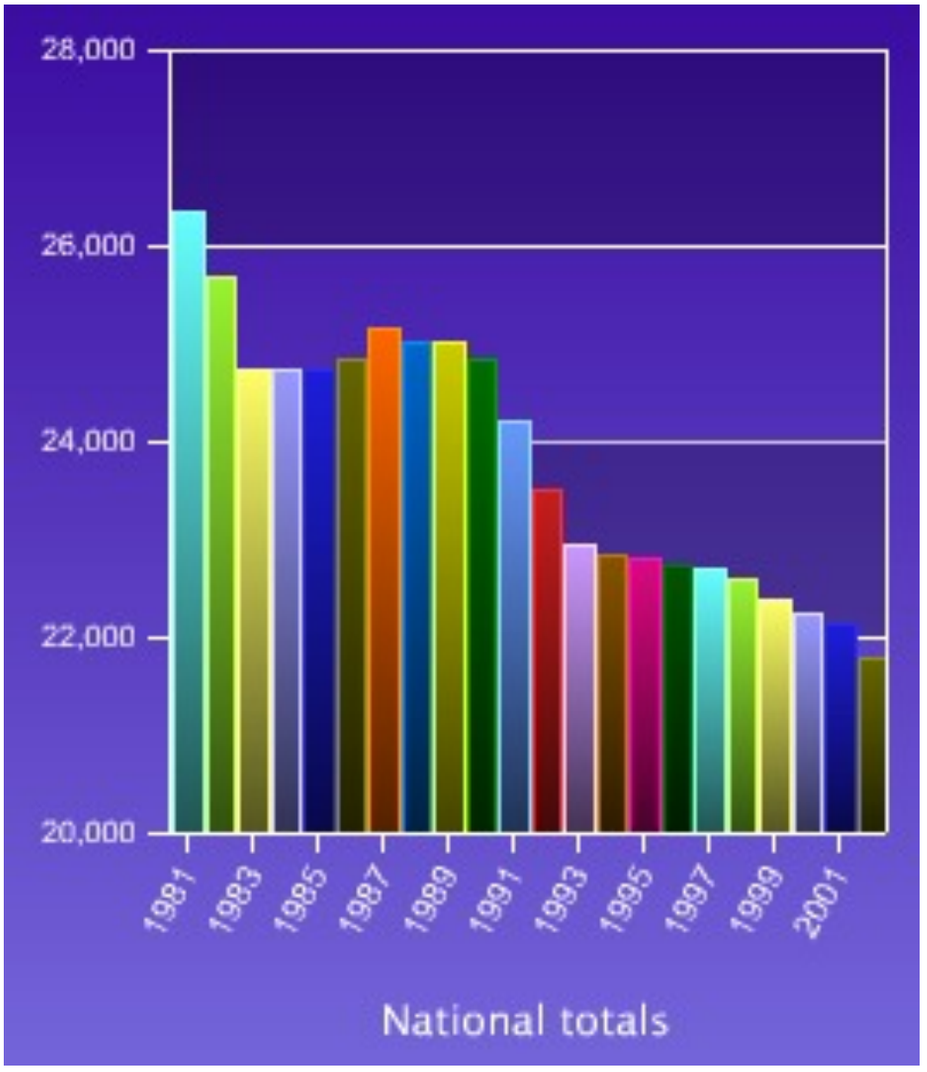
Example: Sales by company



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.)

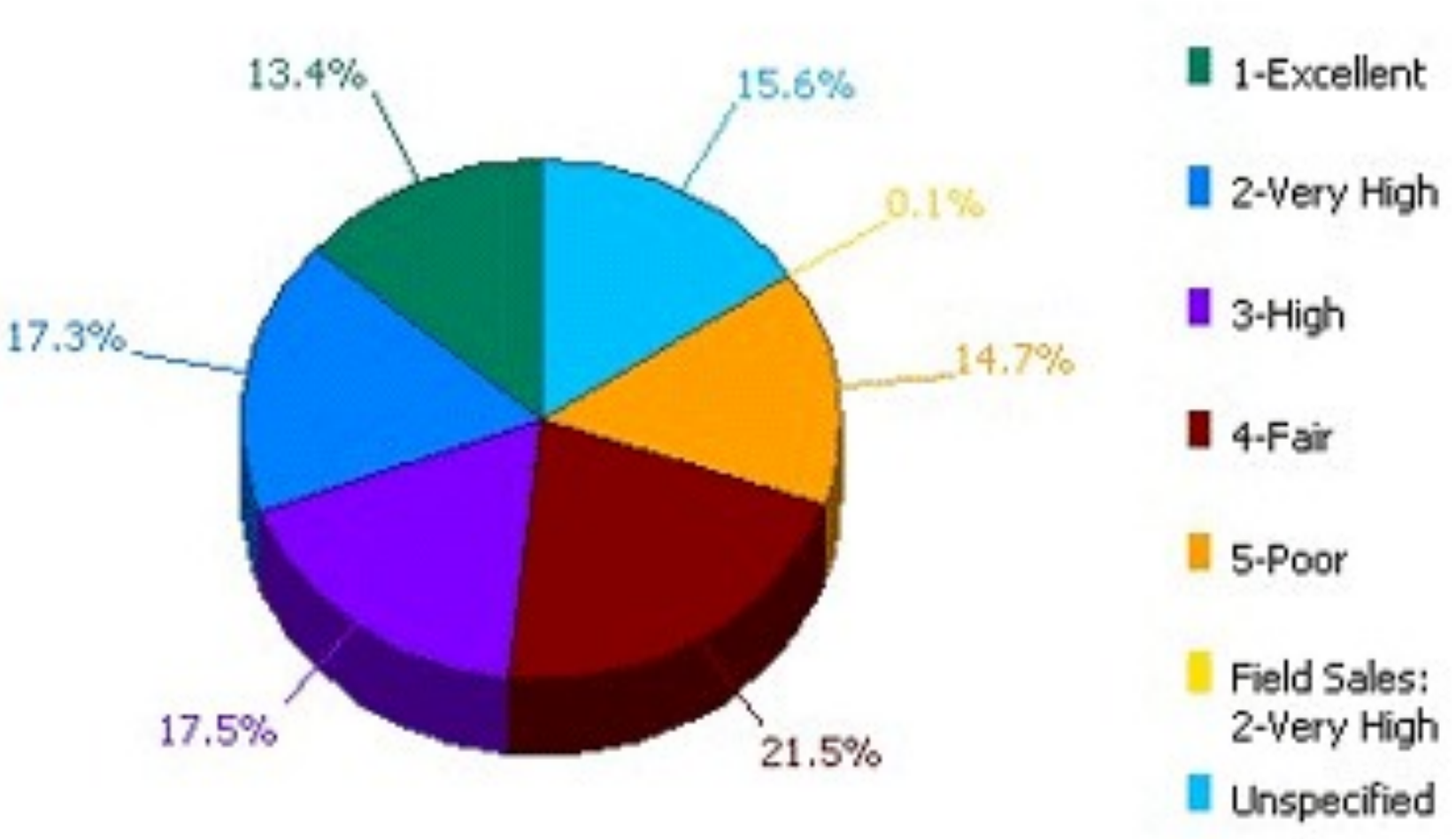
Example: Totals of something or other by year



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.)

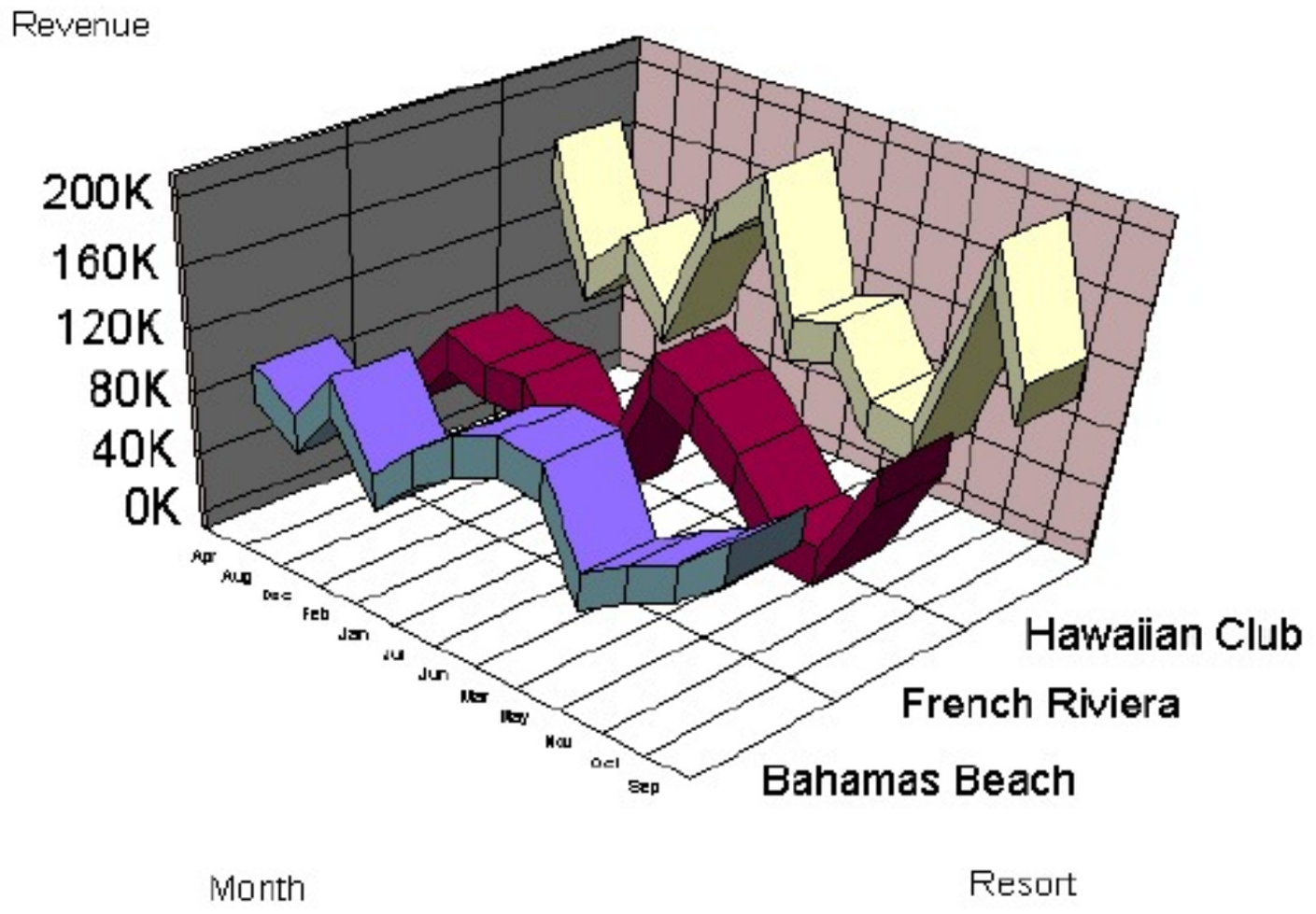
Example: New sales opportunities by lead quality



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.)

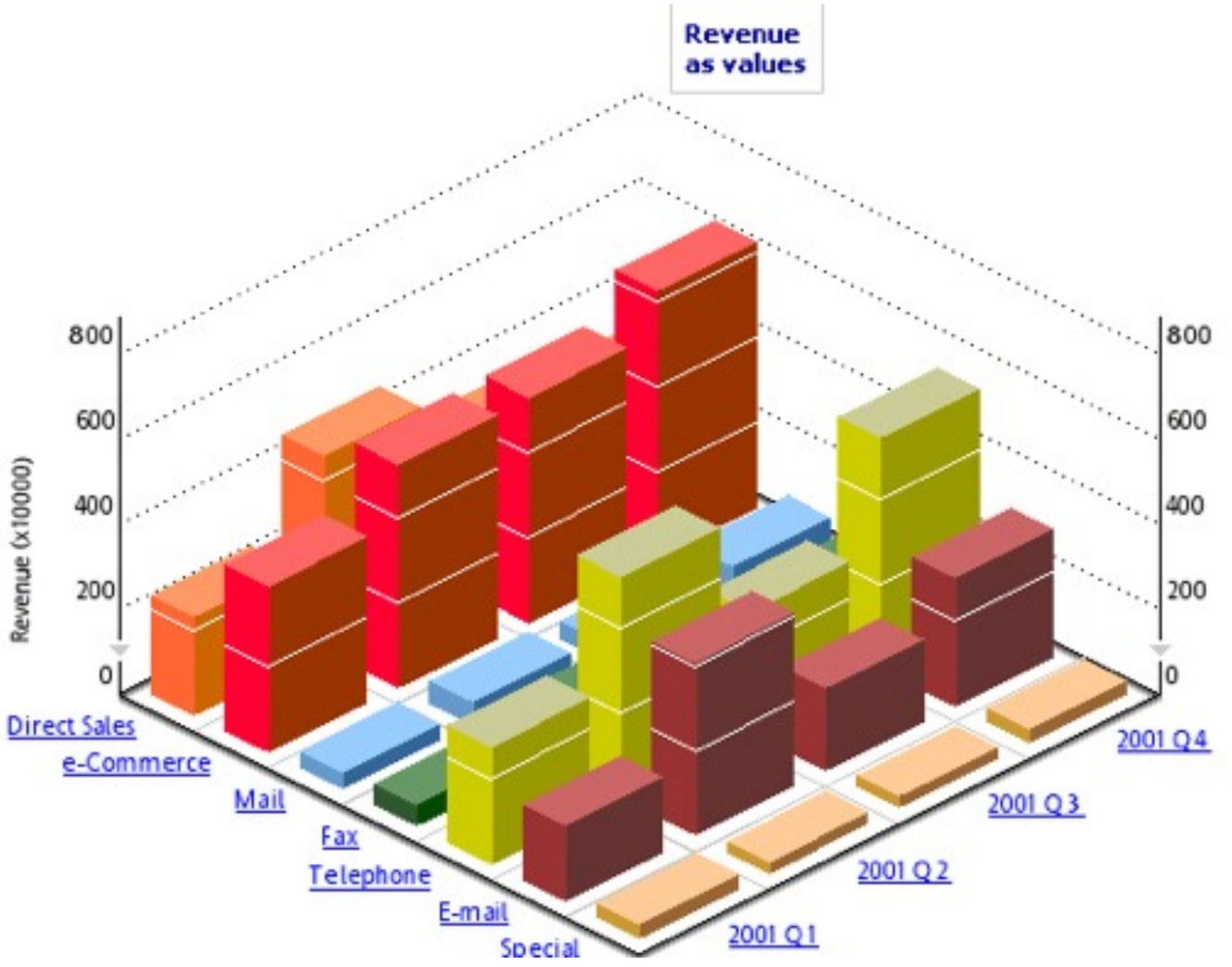
Example: Revenue by resort and month



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.)

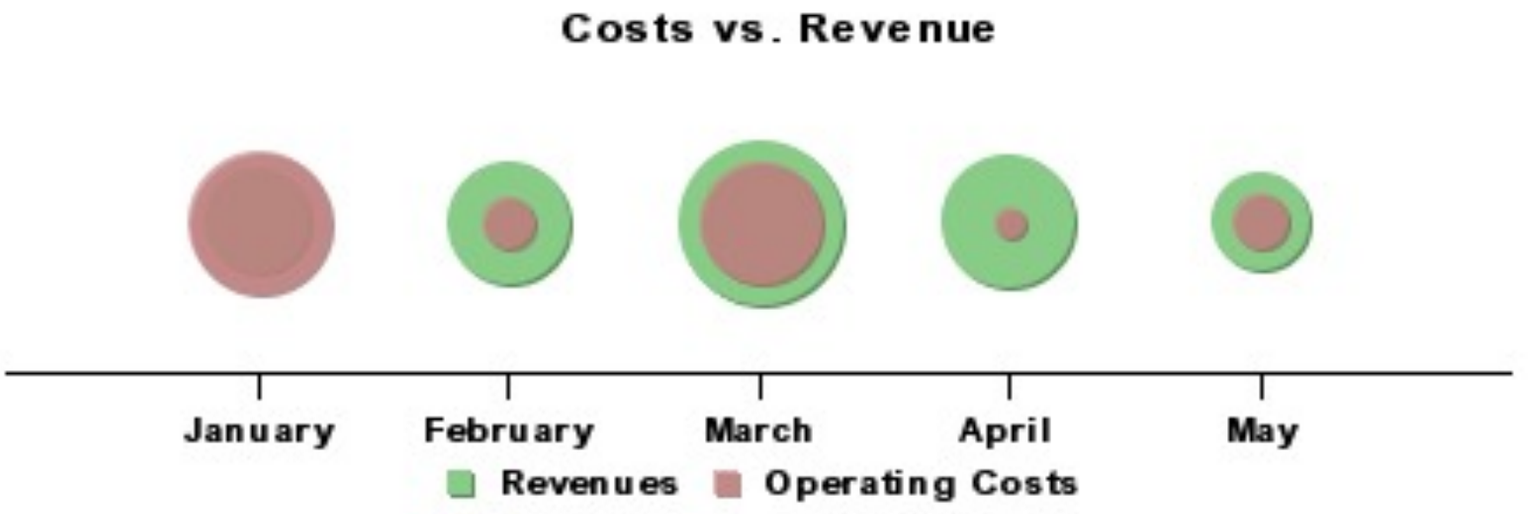
Example: Revenue by sales channel and quarter



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.)

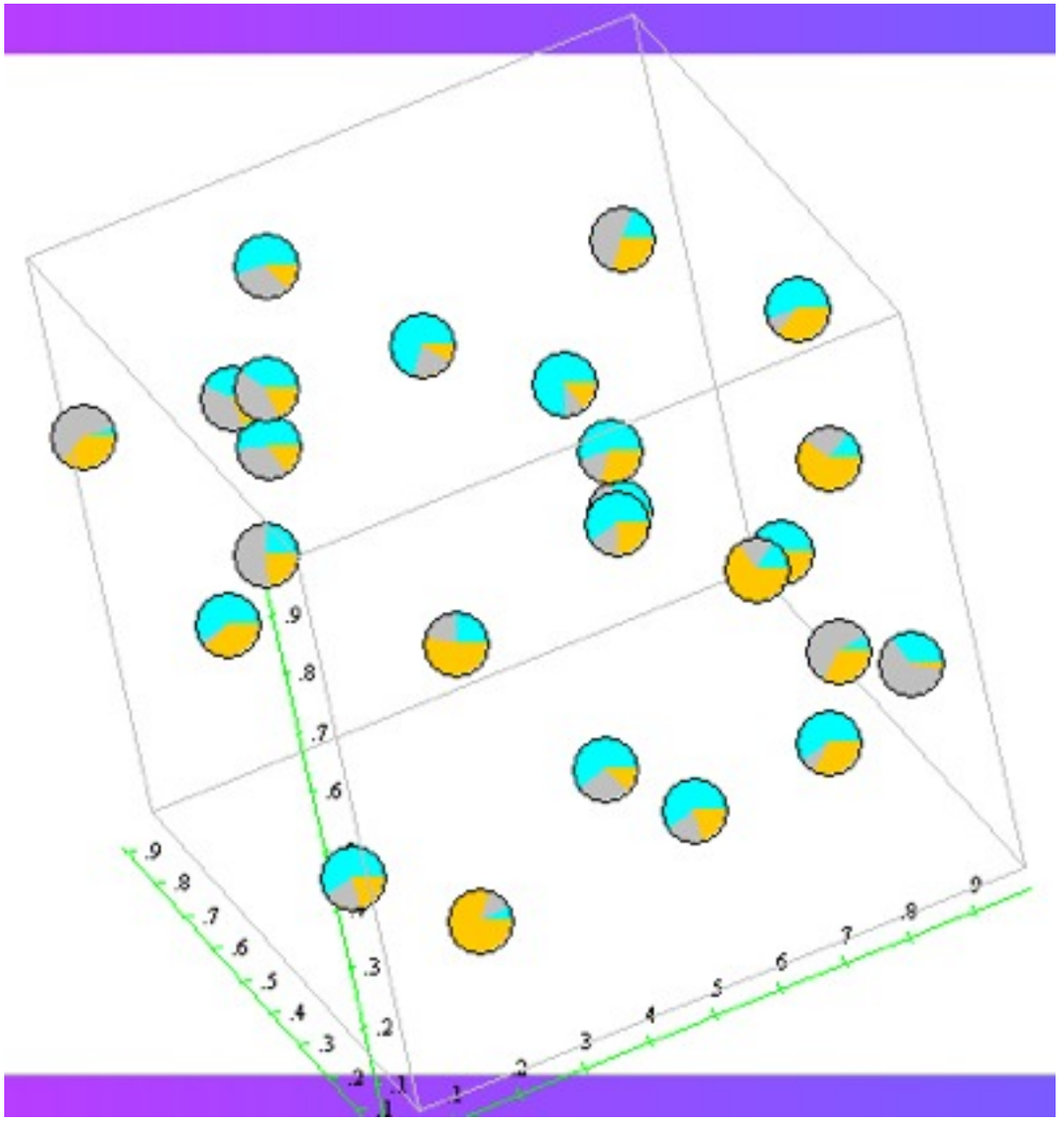
Example: Revenue vs. costs per month



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!

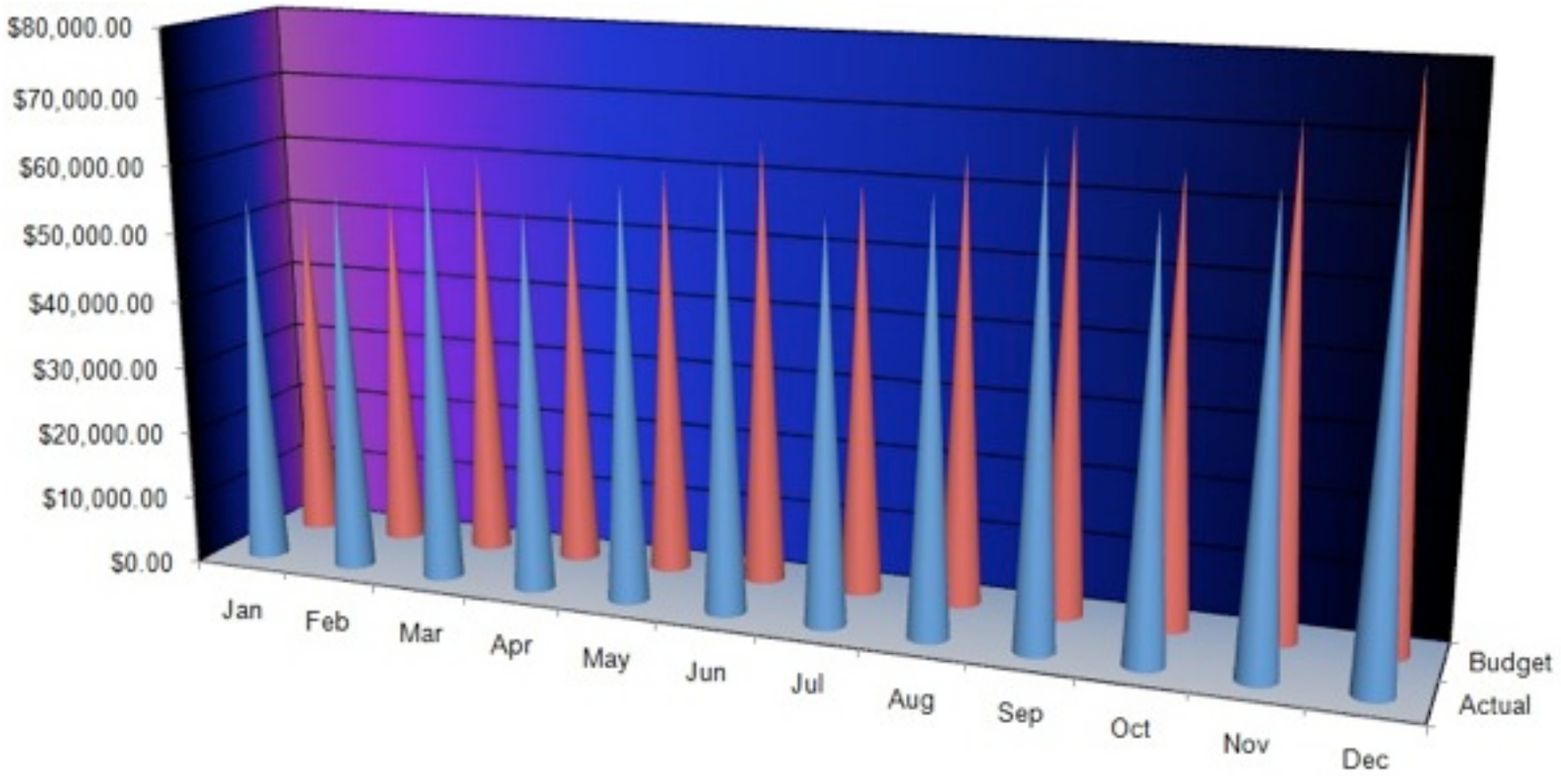
(Source: Website of Visualize, Inc.)

Tables 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

Poor graphs can be transformed.

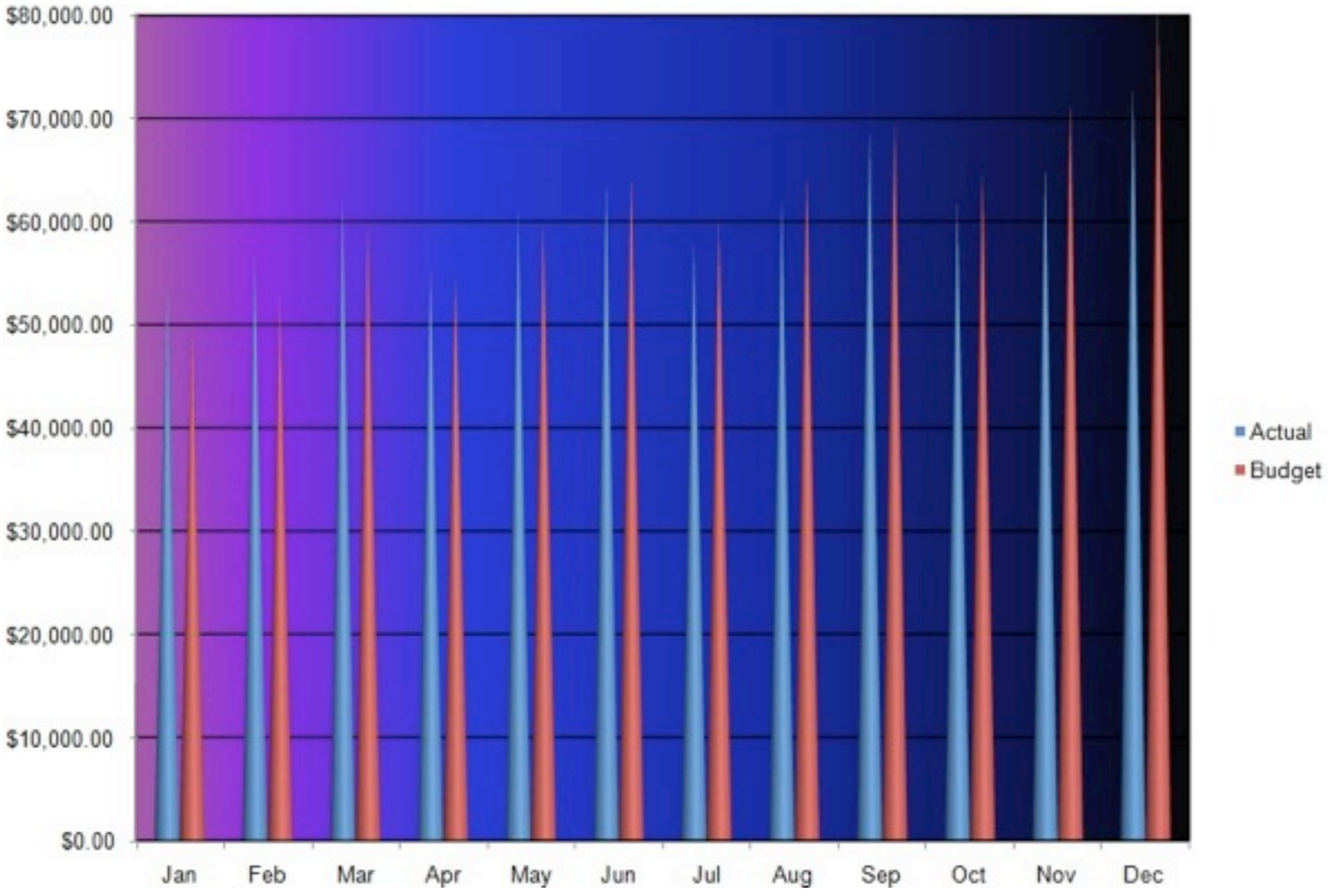


This is the kind of graph that software products, including Excel, encourage us to create. They give us an infinite selection of poorly-designed graphs from which to choose. What we really need, however, is a small selection of graphs that really work.

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.

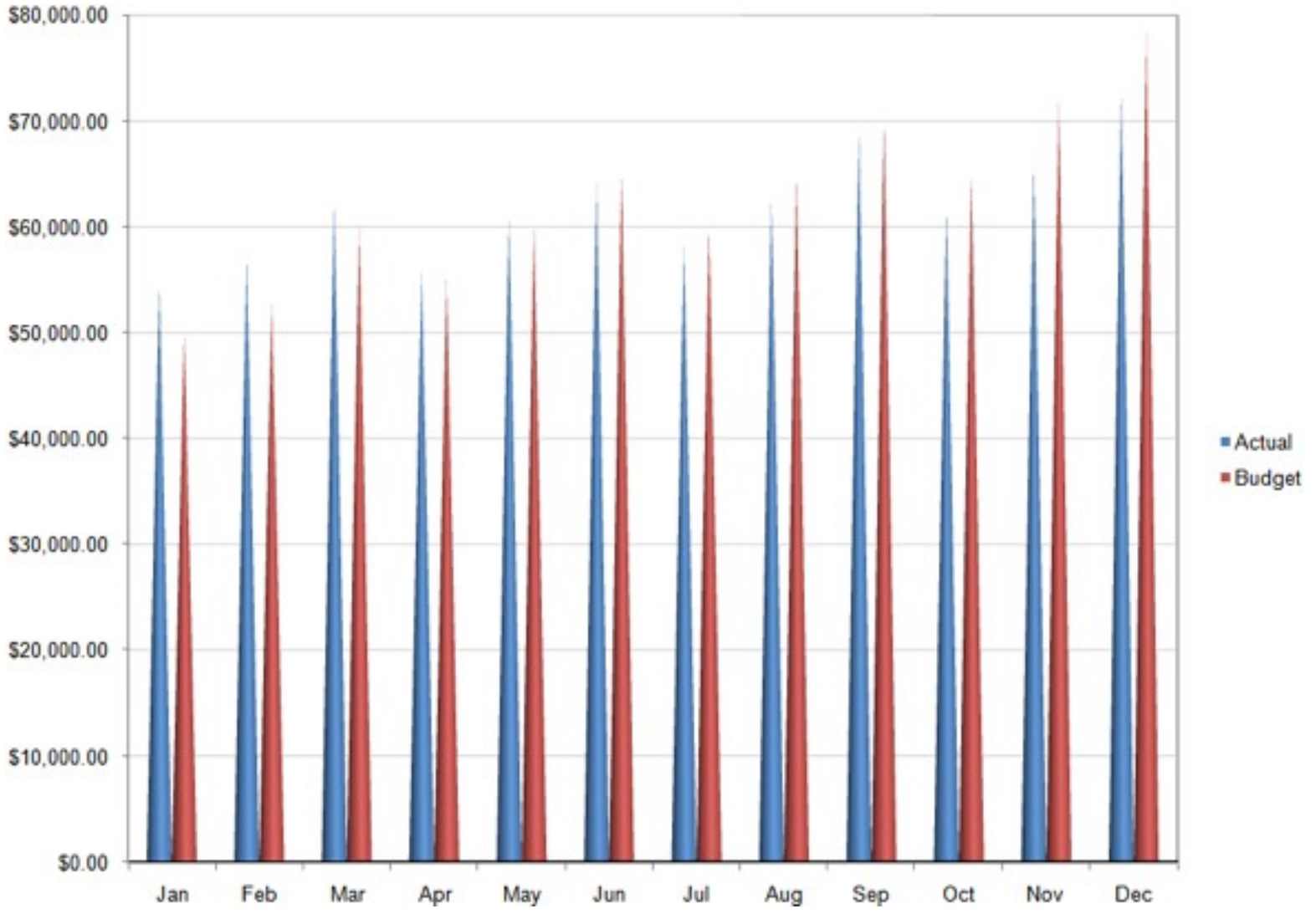
Let's transform this graph into one that communicates.

Poor graphs can be transformed.



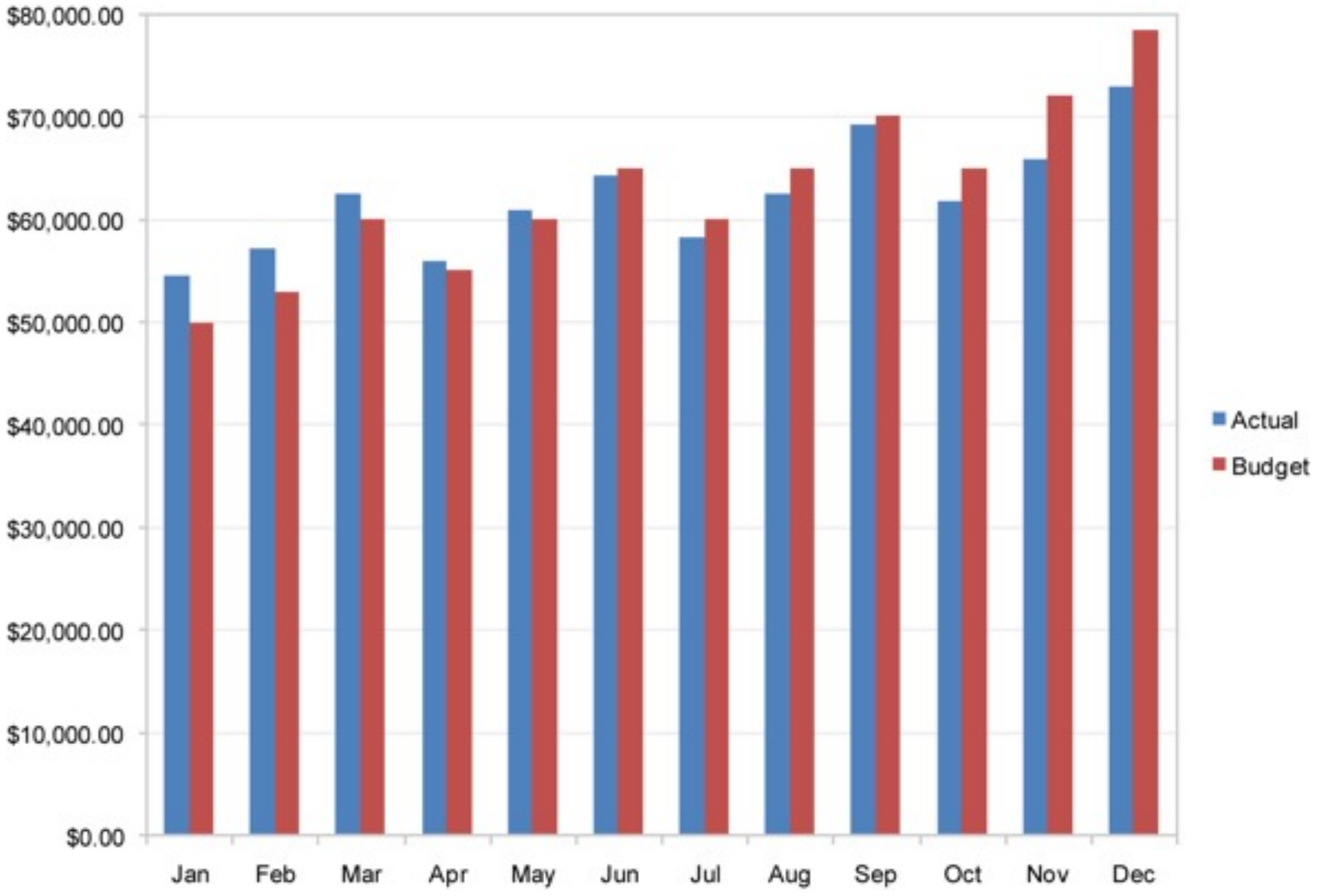
We have now removed the useless 3-D effects and angle, which makes the data easier to read.

Poor graphs can be transformed.



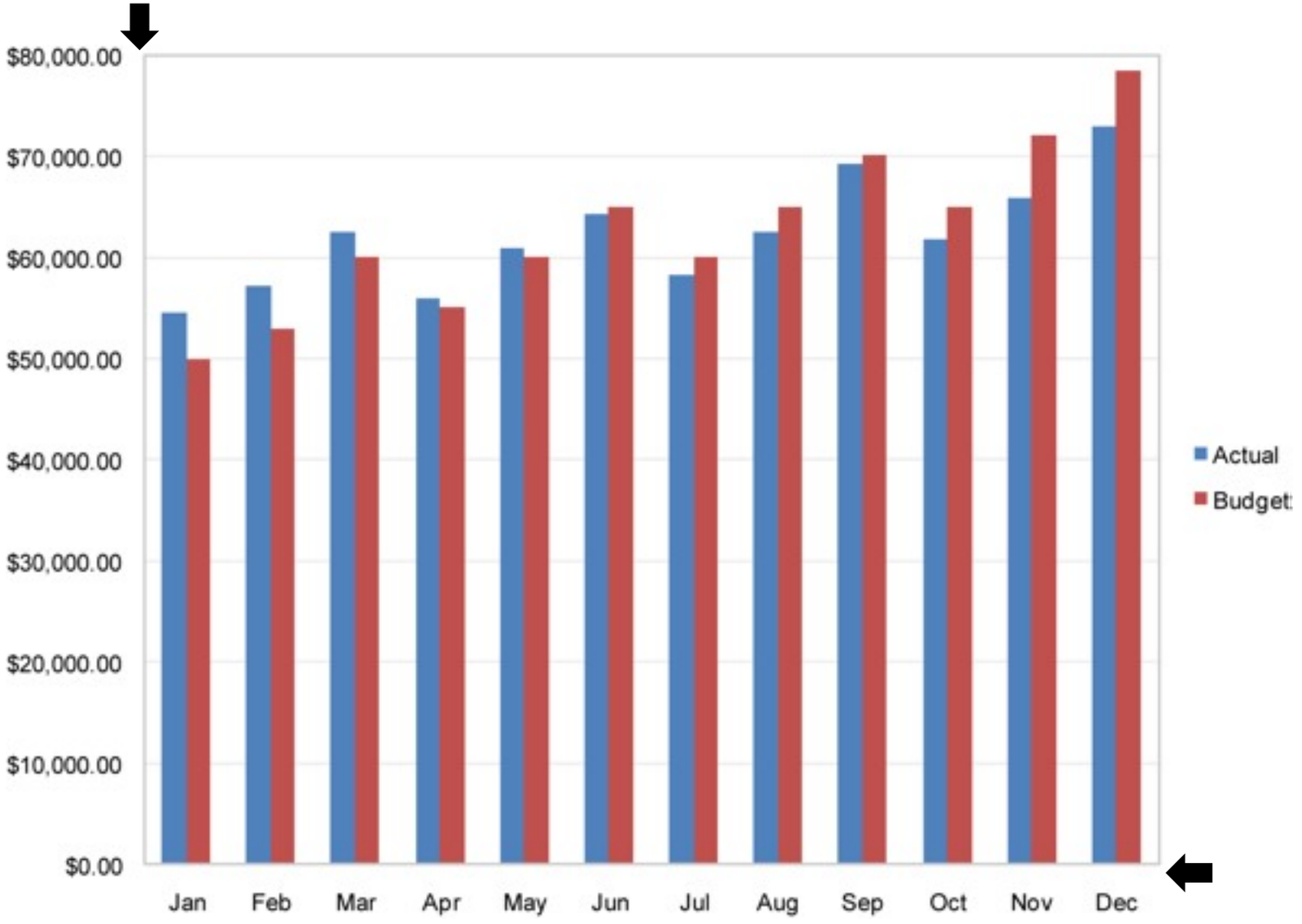
We have now removed the background fill color.

Poor graphs can be transformed.



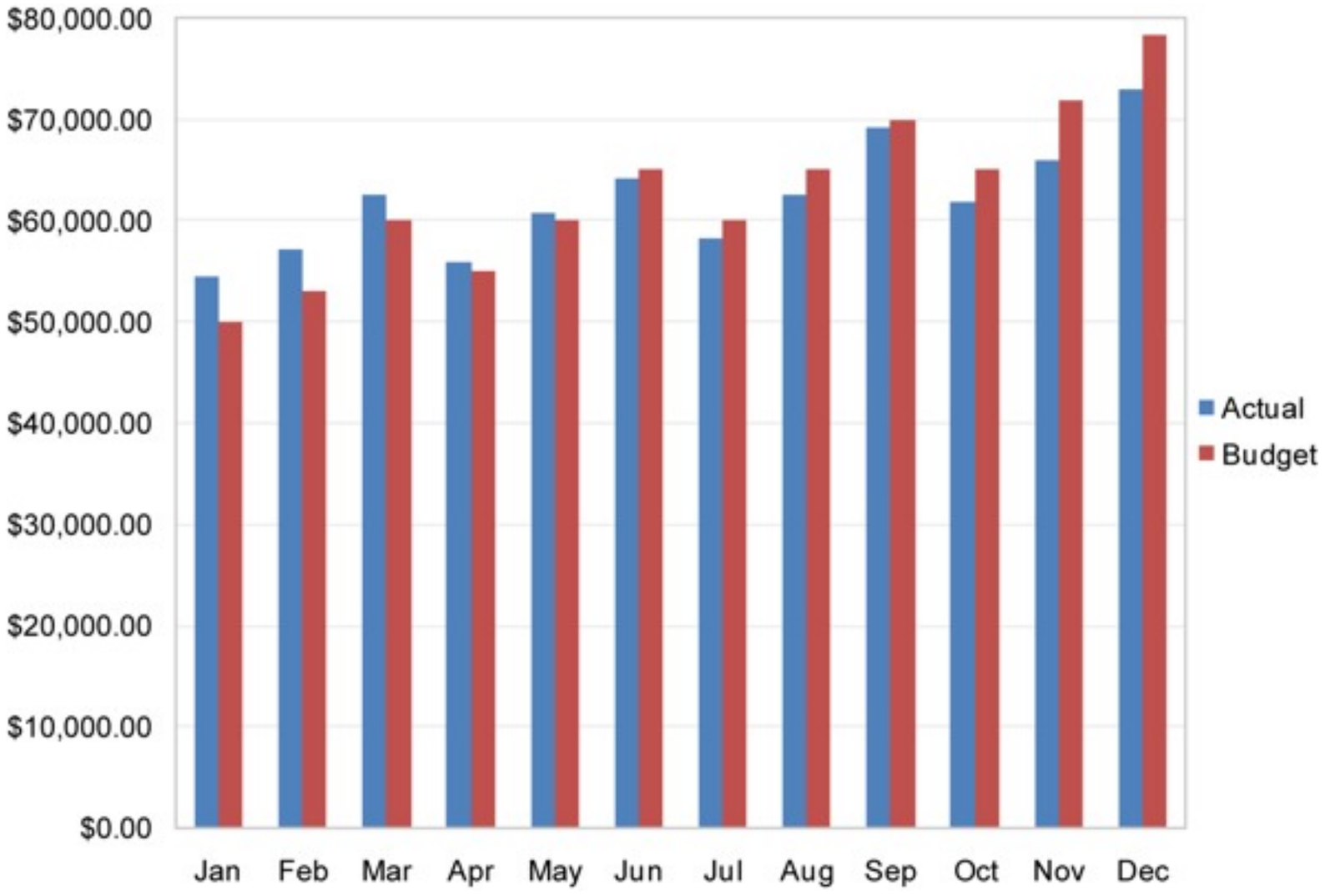
We have now replaced the silly cones with regular bars.

Poor graphs can be transformed.



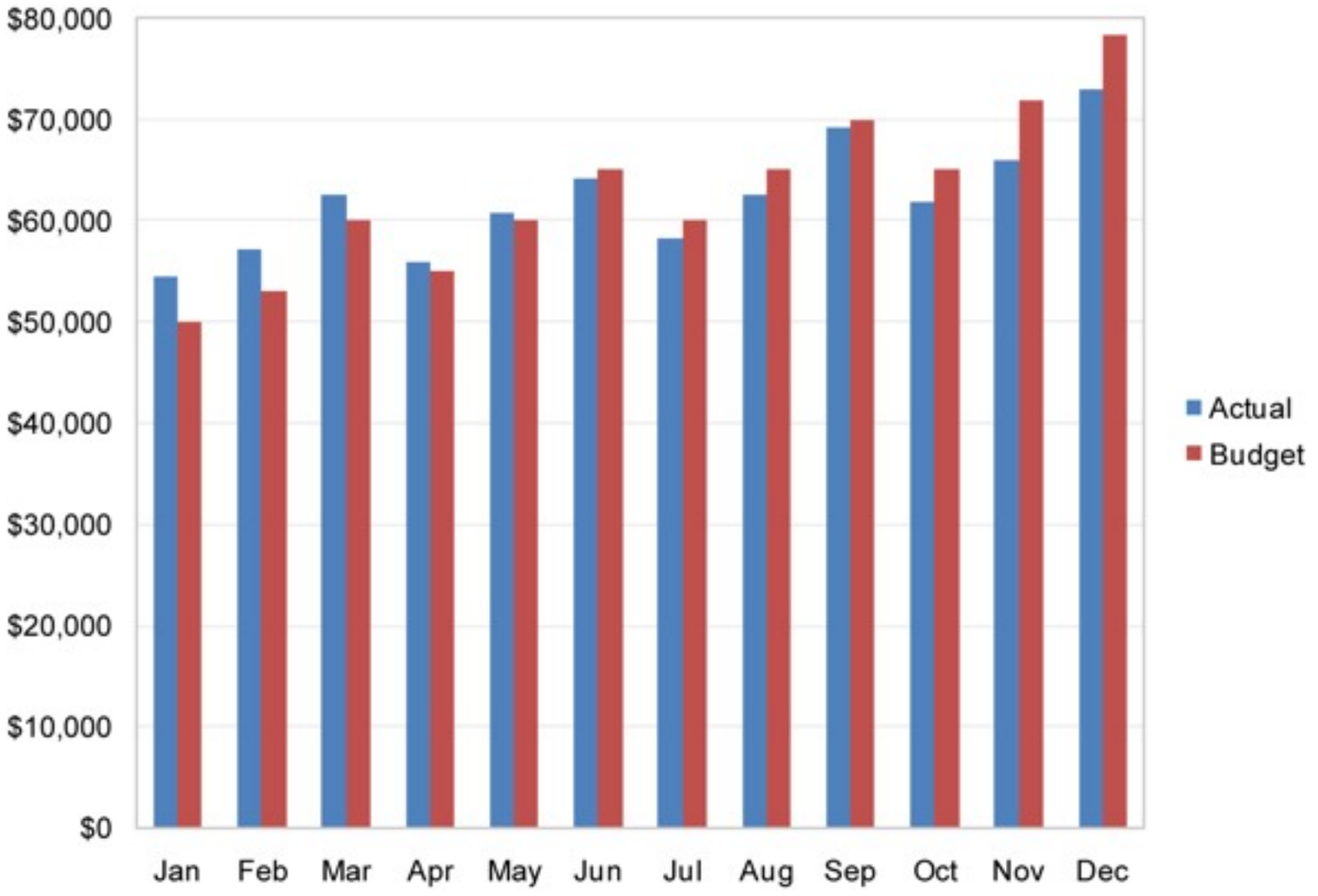
We have now removed the tick marks, which aren't necessary. Tick marks are not needed to separate the months along the X-axis and because horizontal grid lines are being displayed, there is no need for tick marks on the Y-axis either.

Poor graphs can be transformed.



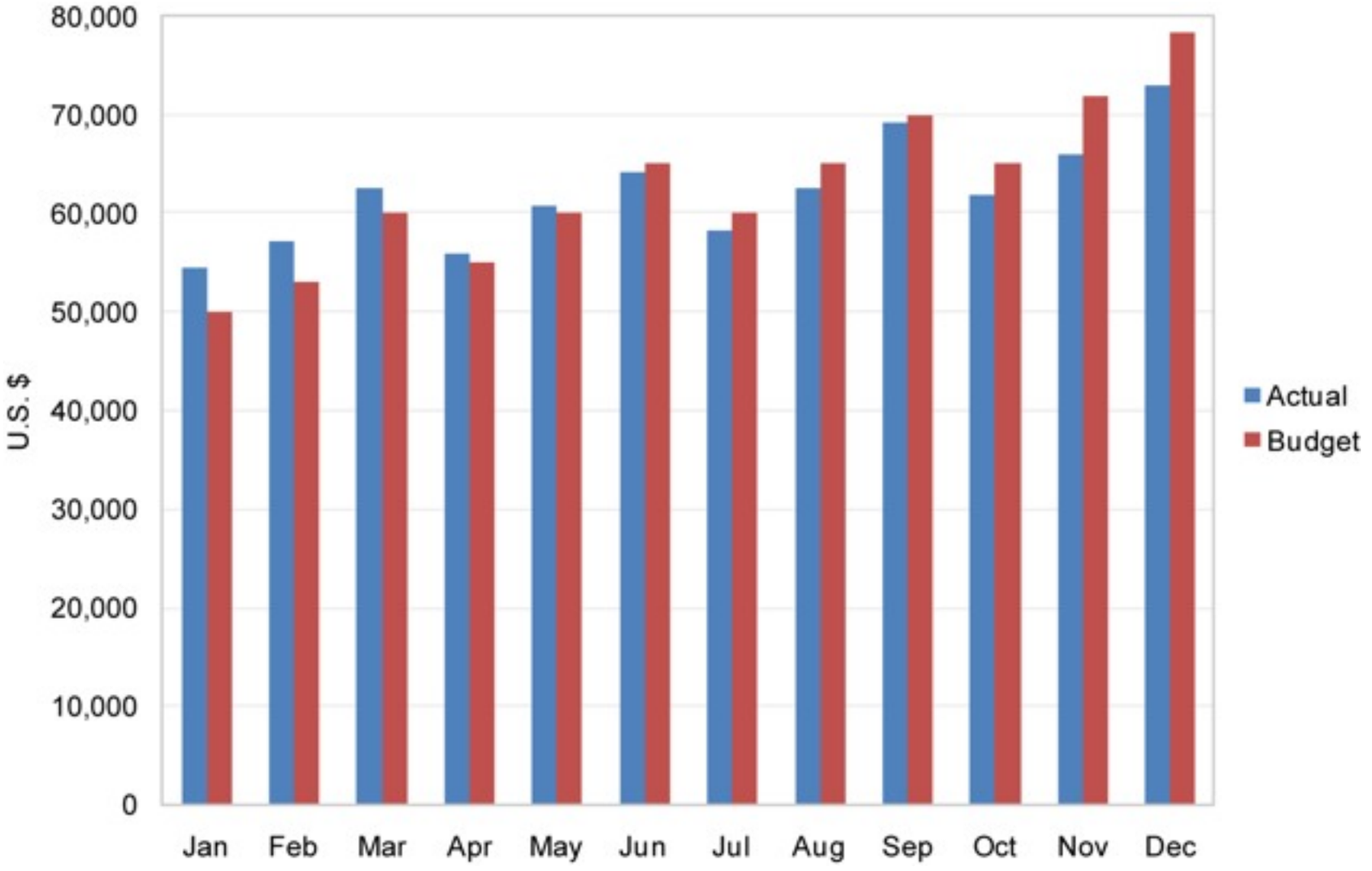
We have now enlarged the text, making it easier to read.

Poor graphs can be transformed.



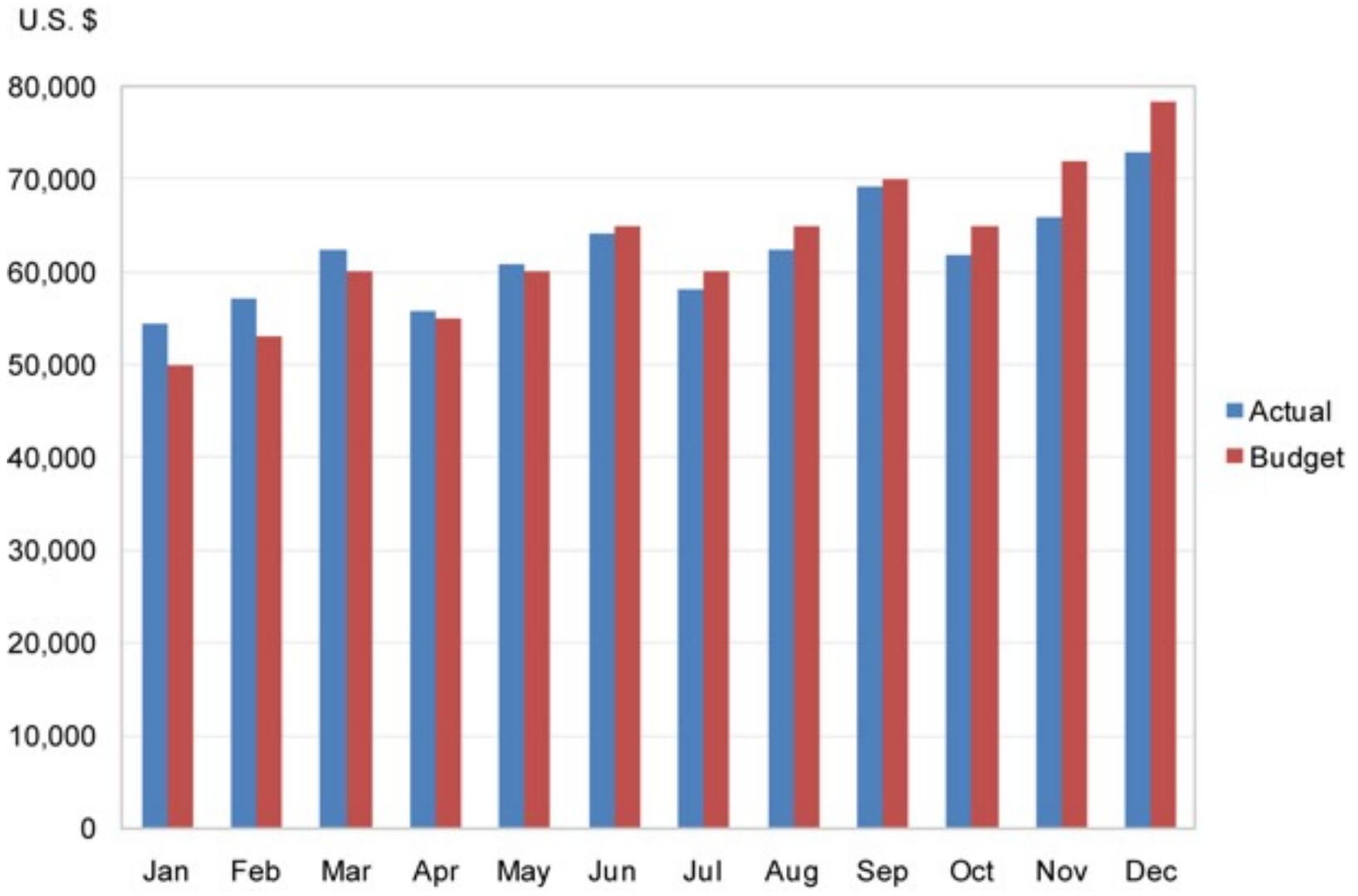
We have now removed the unnecessary decimal places in the dollar amounts along the Y-axis.

Poor graphs can be transformed.



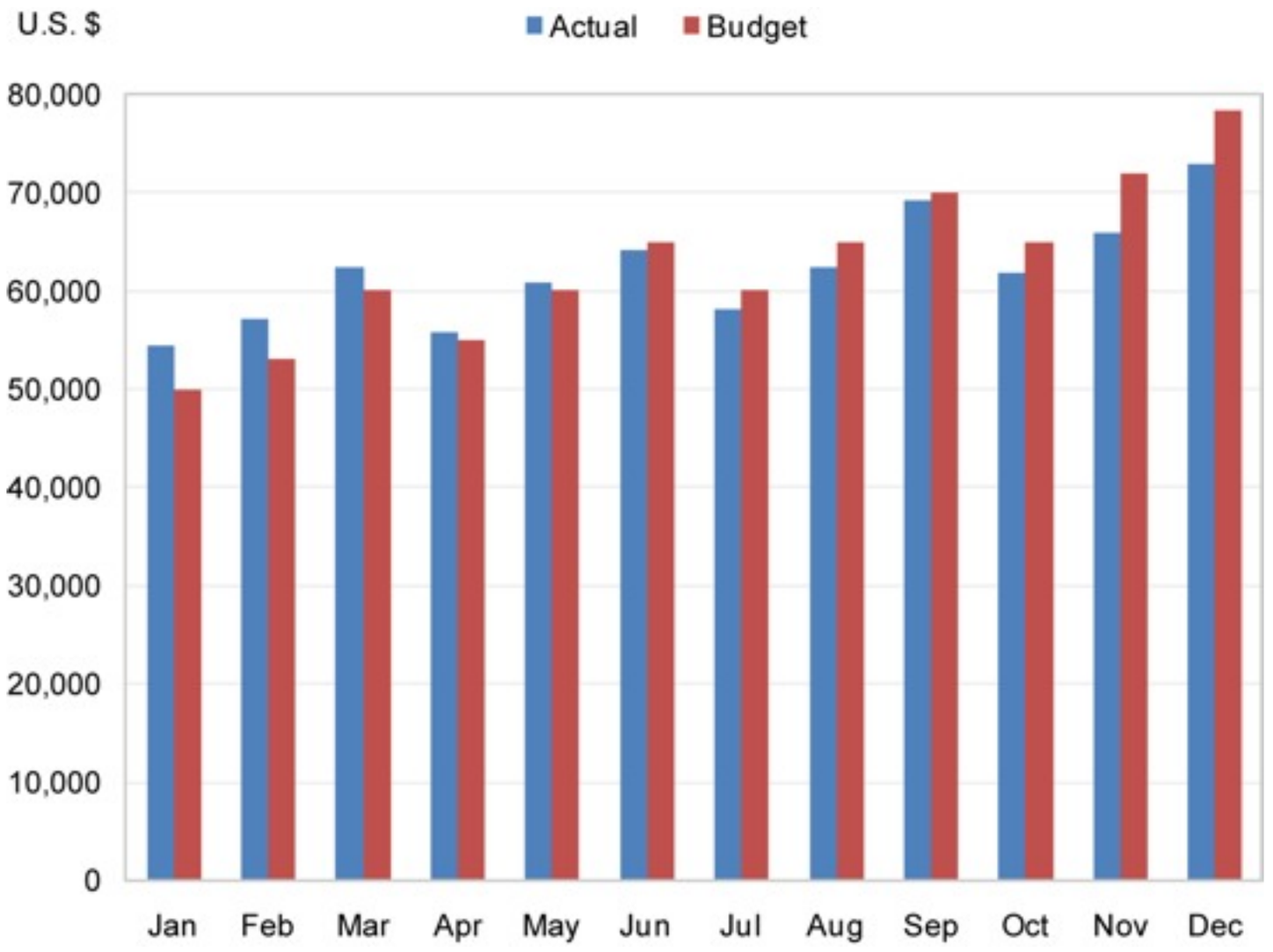
We have now removed the redundant dollar signs and labeled the unit of measure (U.S. $) clearly.

Poor graphs can be transformed.



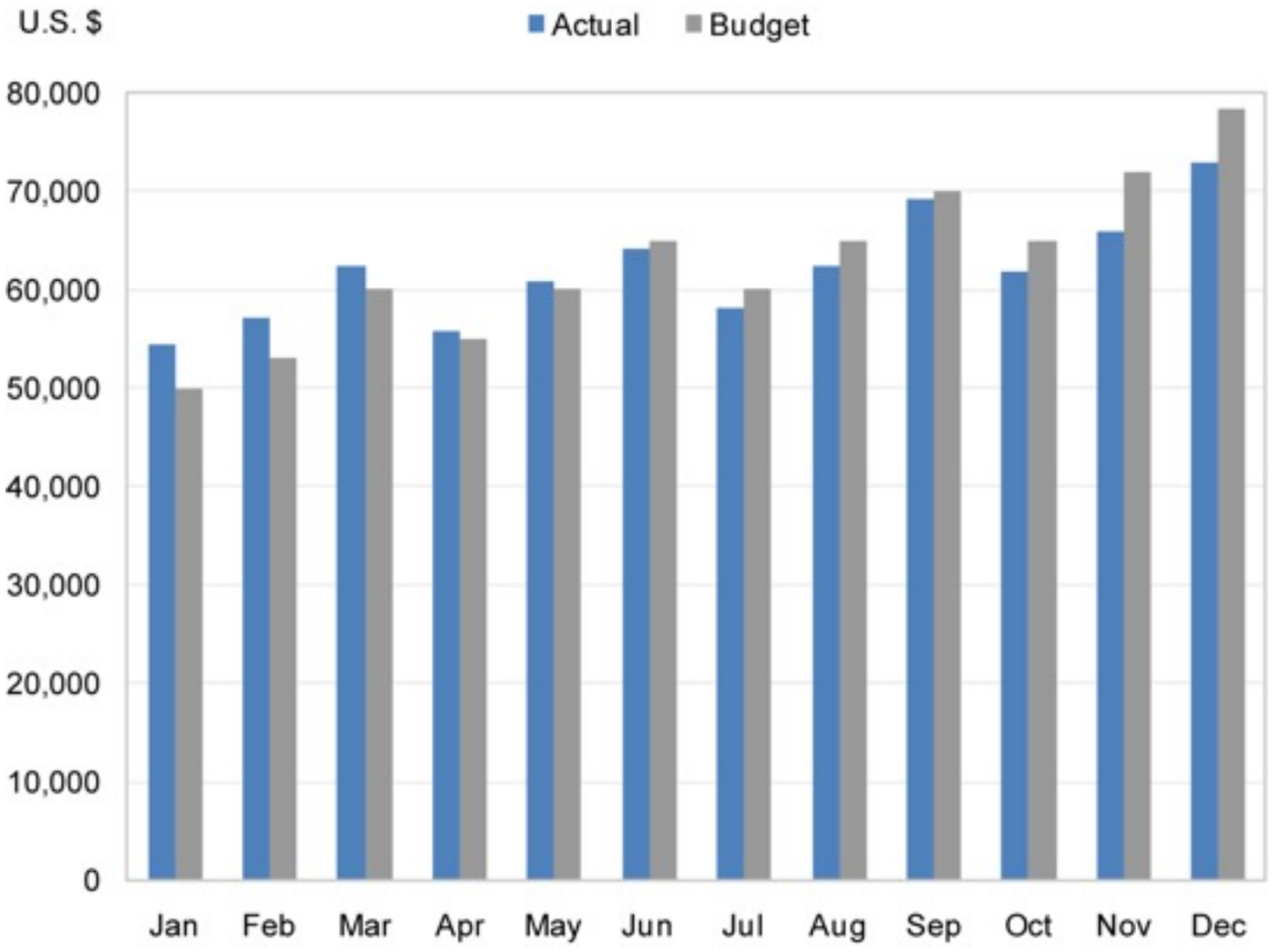
We have now reoriented the Y-axis label to the horizontal and placed it above the axis to make it easier to read.

Poor graphs can be transformed.



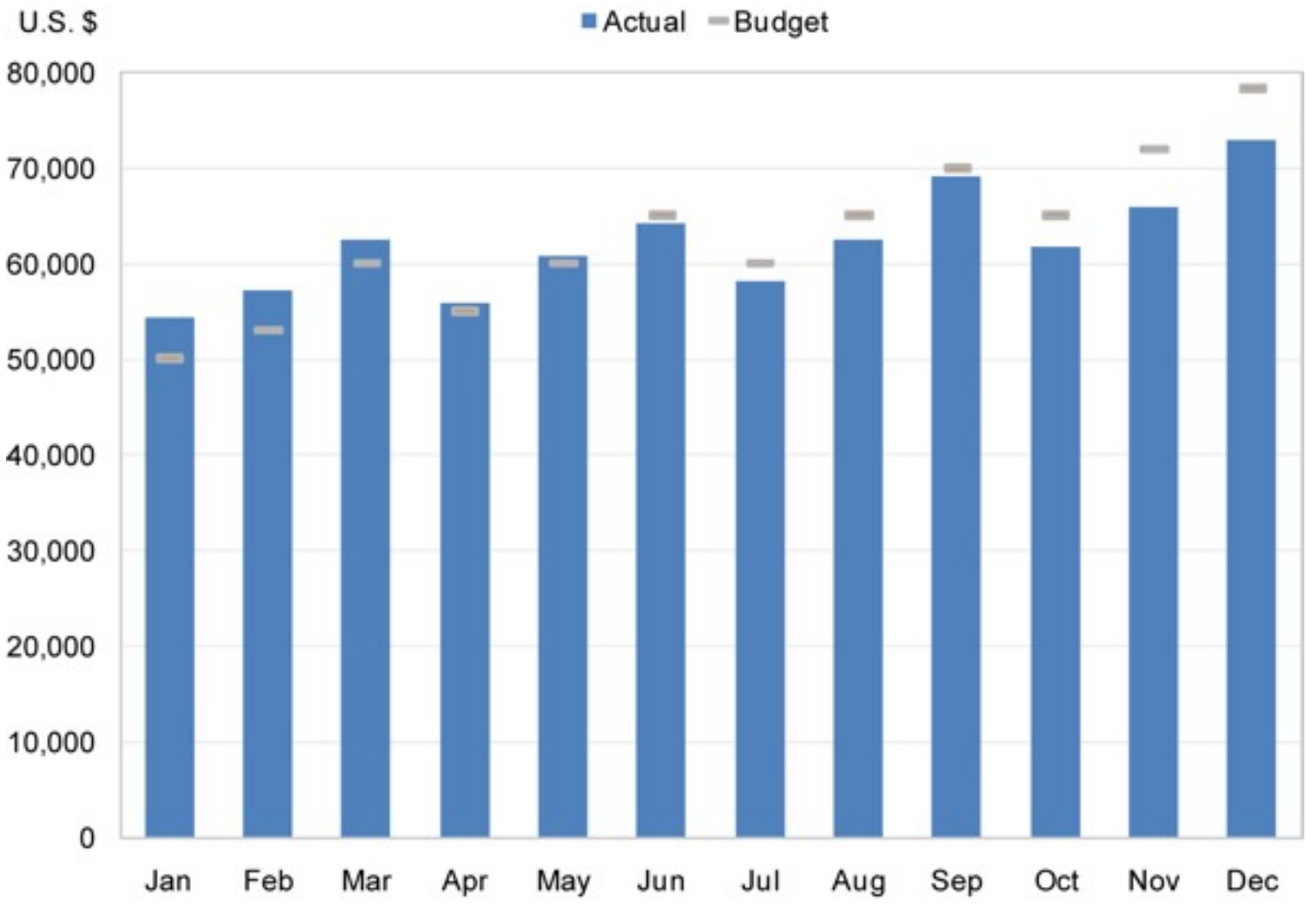
We have now reoriented and repositioned the legend to make it easier to associate it with the data bars.

Poor graphs can be transformed.



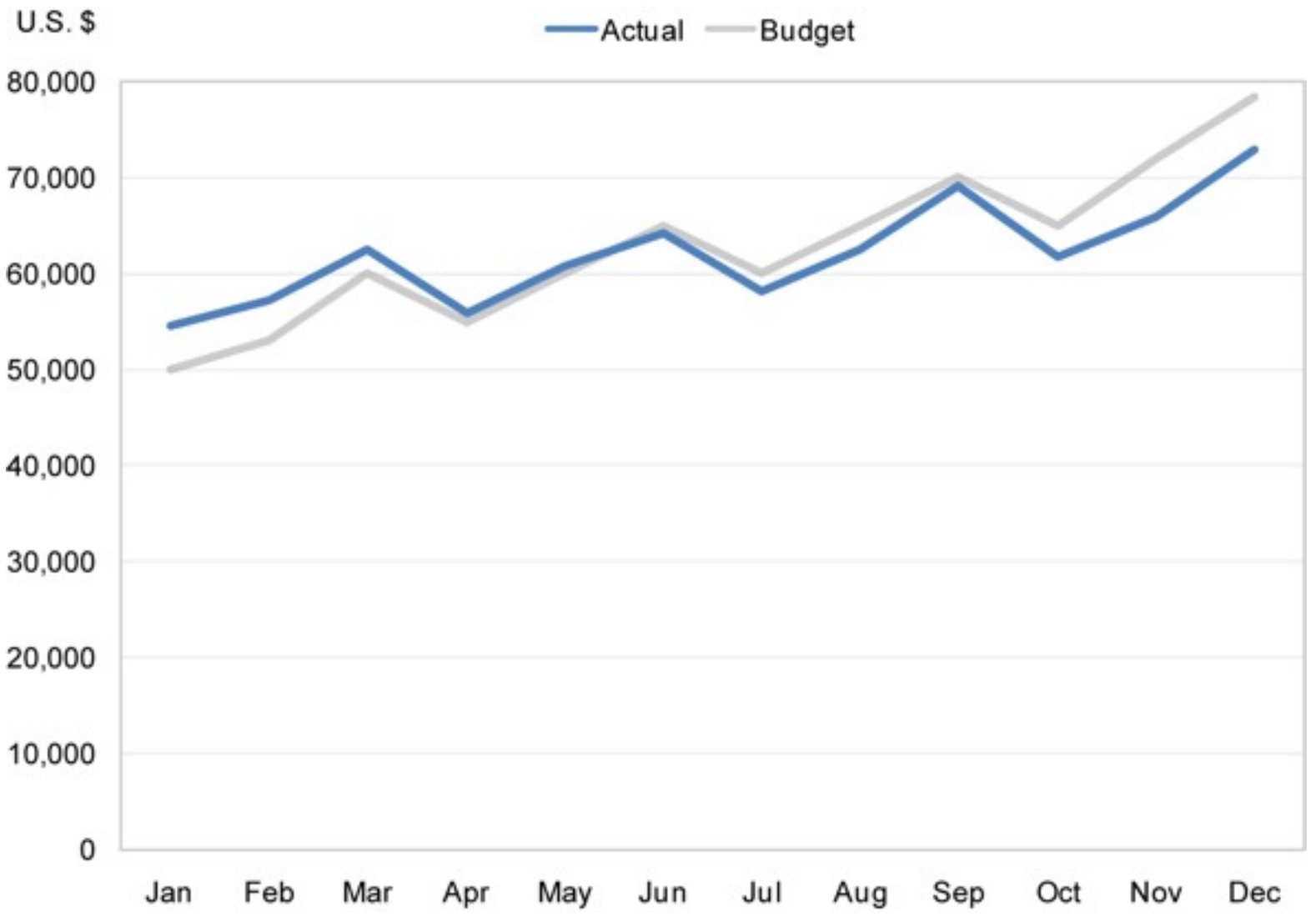
We have now changed the color of the Budget bar to be more visually pleasing in relation to the blue Actual bars. Changing from the color red also removed the possibility people interpreting the data as something bad or a warning, which red is often used to represent.

Poor graphs can be transformed.



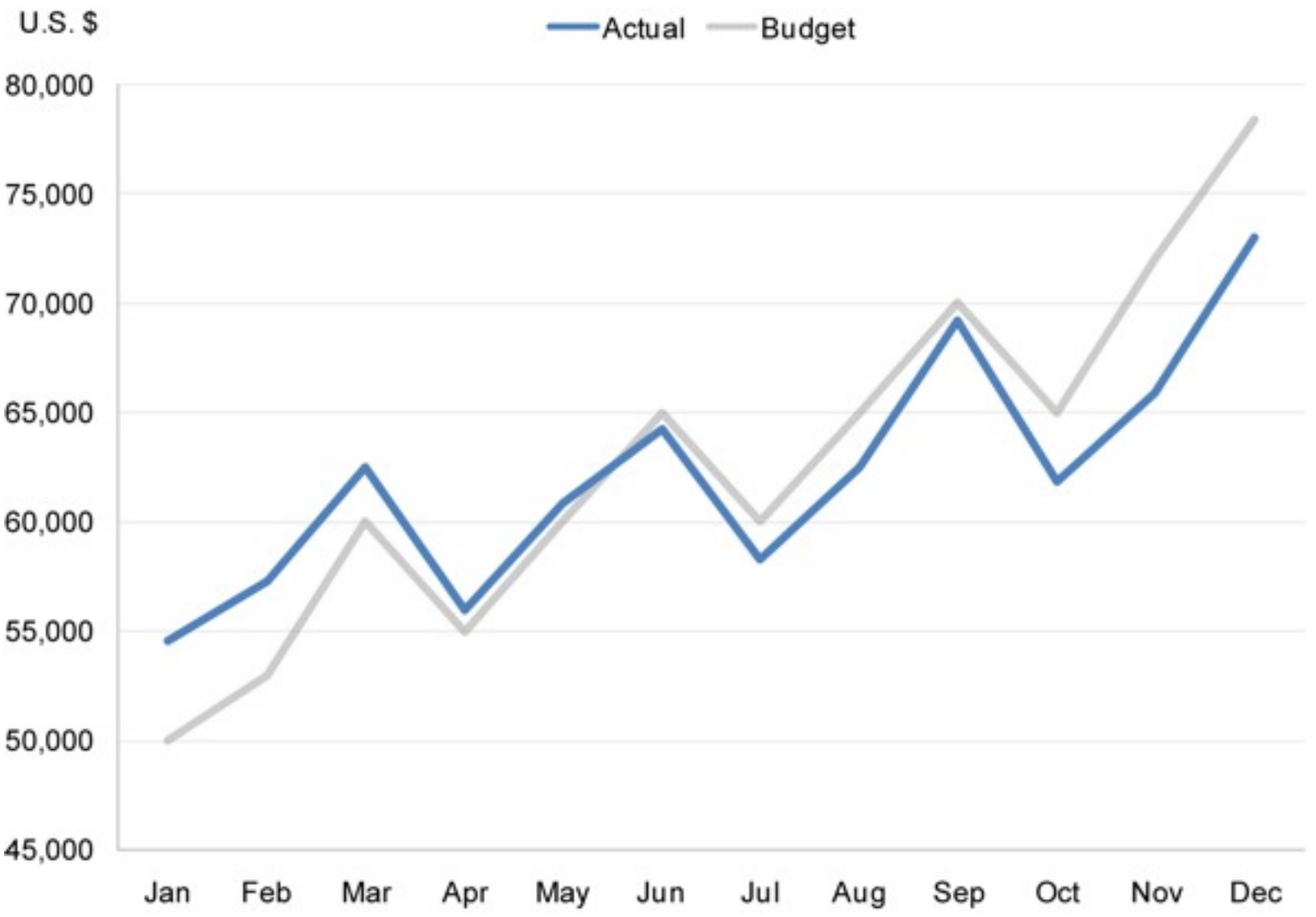
We have now reduced the visual salience of the Budget values, because they are less important that the Actual values, and have done so in a way that reduced clutter.

Poor graphs can be transformed.



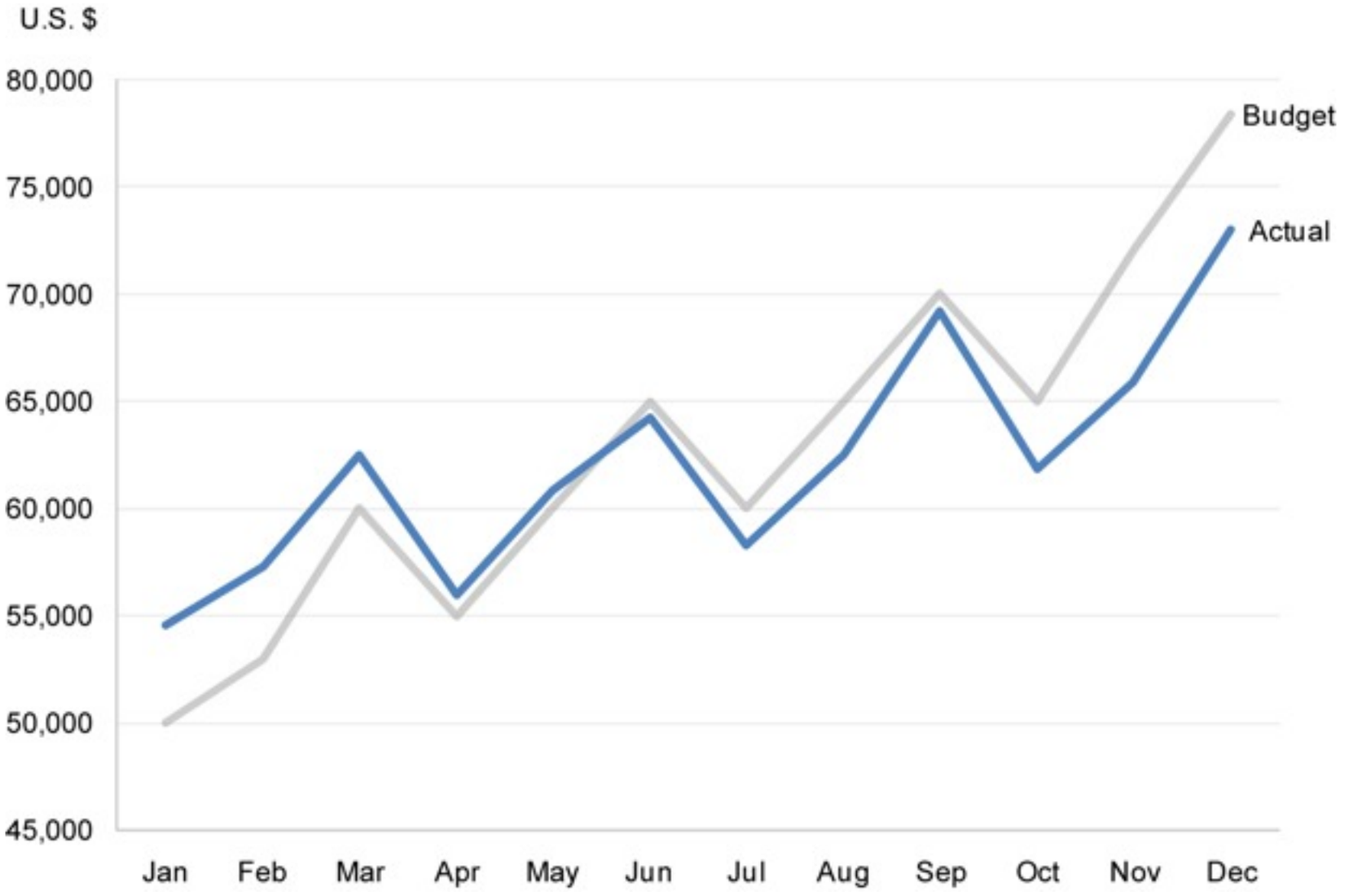
We have now made it much easier to see the pattern of change through time by using lines rather than bars to represent the data.

Poor graphs can be transformed.



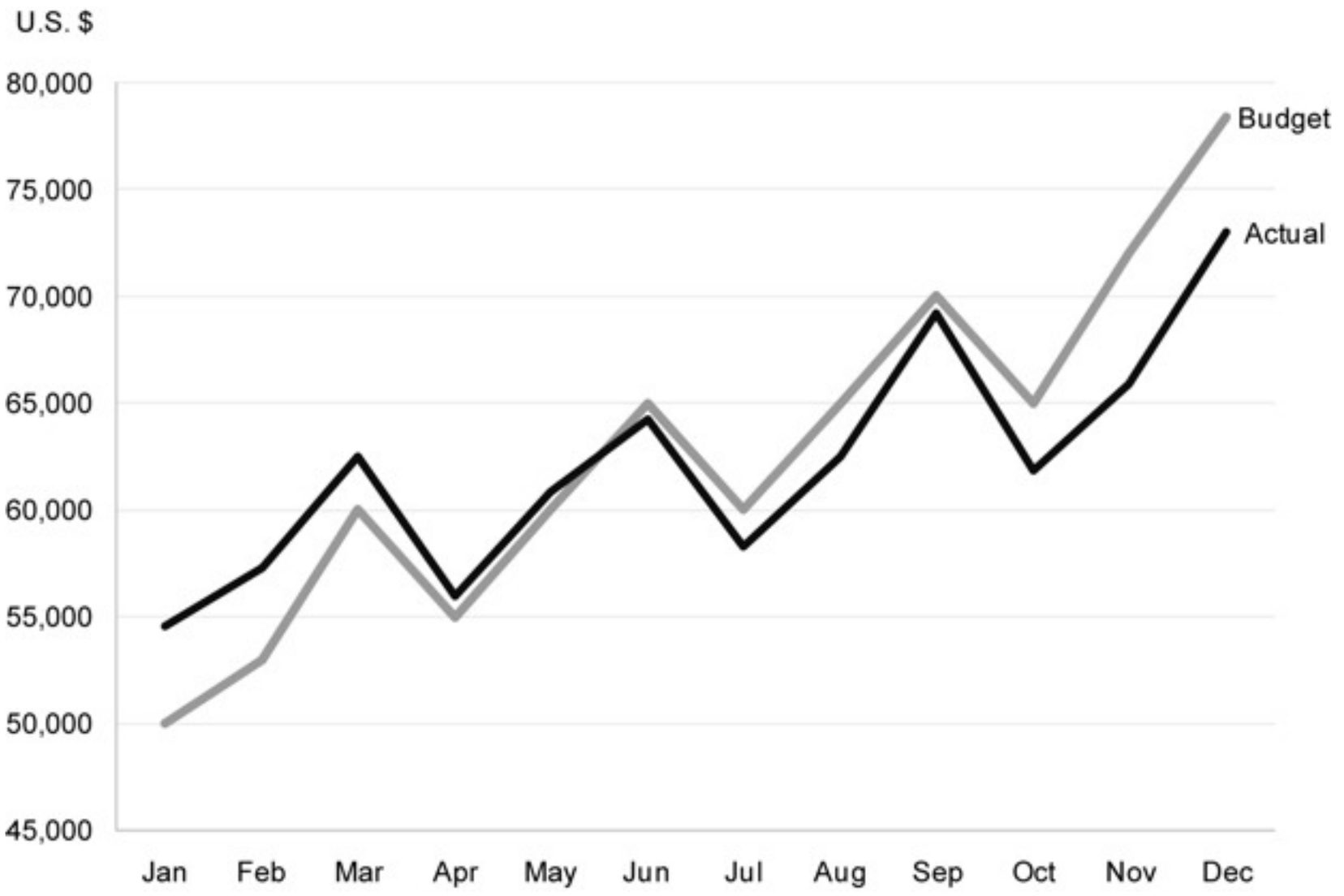
We have now made it much easier to examine the differences between actual expenses and the budget by spreading them across more space.

Poor graphs can be transformed.



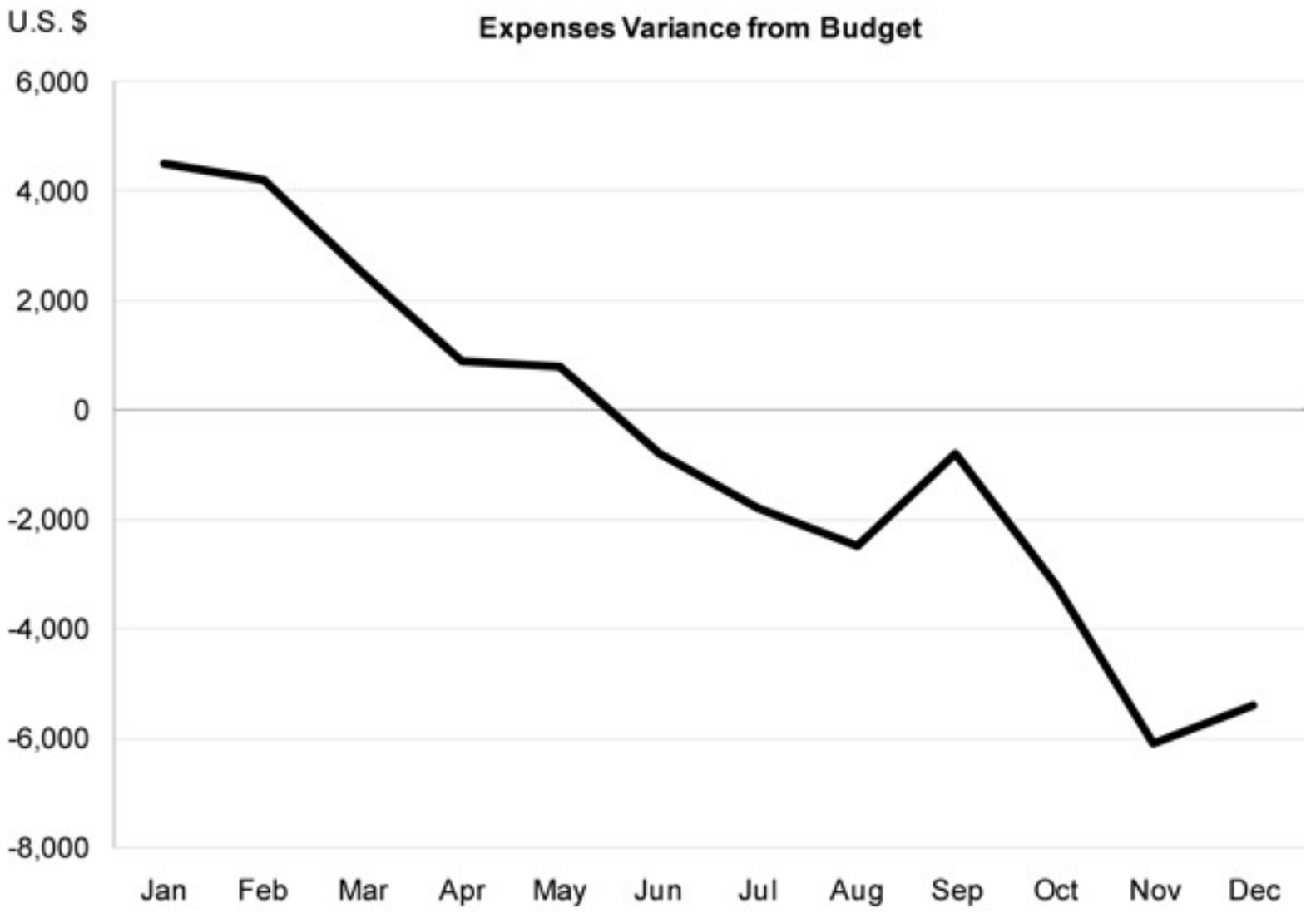
We have now labeled the lines directly, removing the need for a legend.

Poor graphs can be transformed.



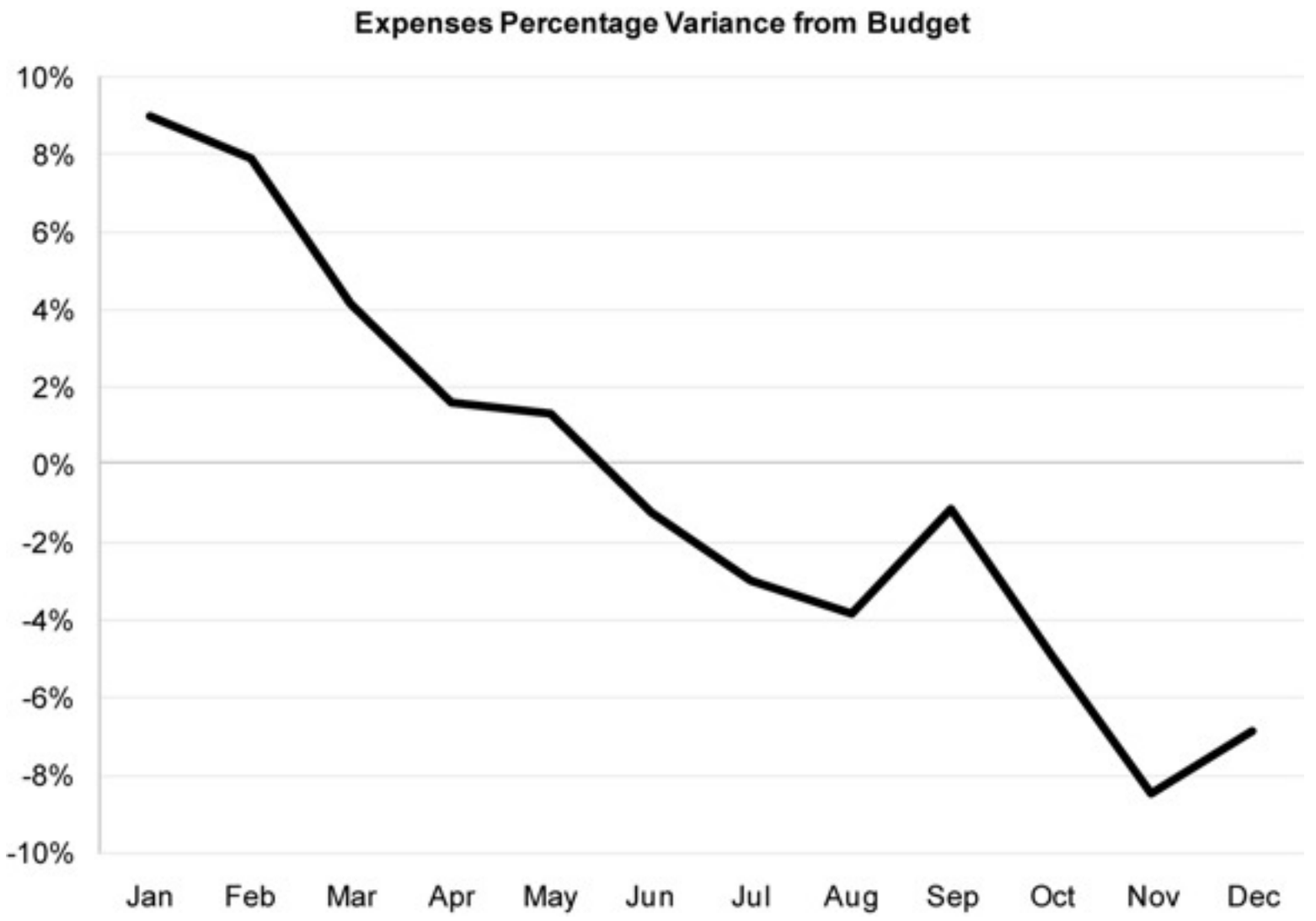
We have now changed the lines to two shades of gray to guaranty that even if the graph is printed on a black-and-white printer or photocopier, they will still look distinctly different from one another.

Poor graphs can be transformed.



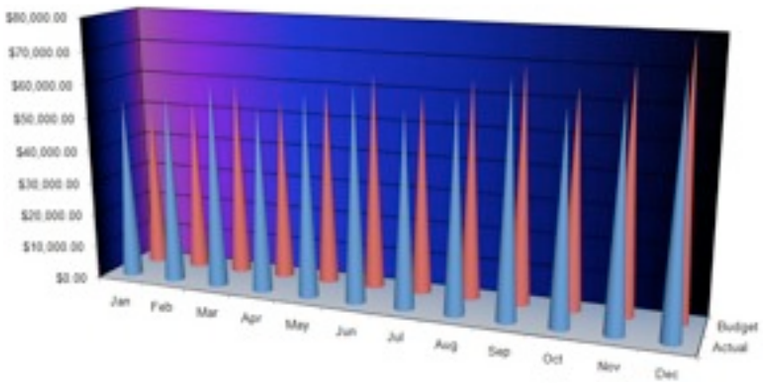
We have now represented the variance of actual expenses from the budget directly, as a single line.

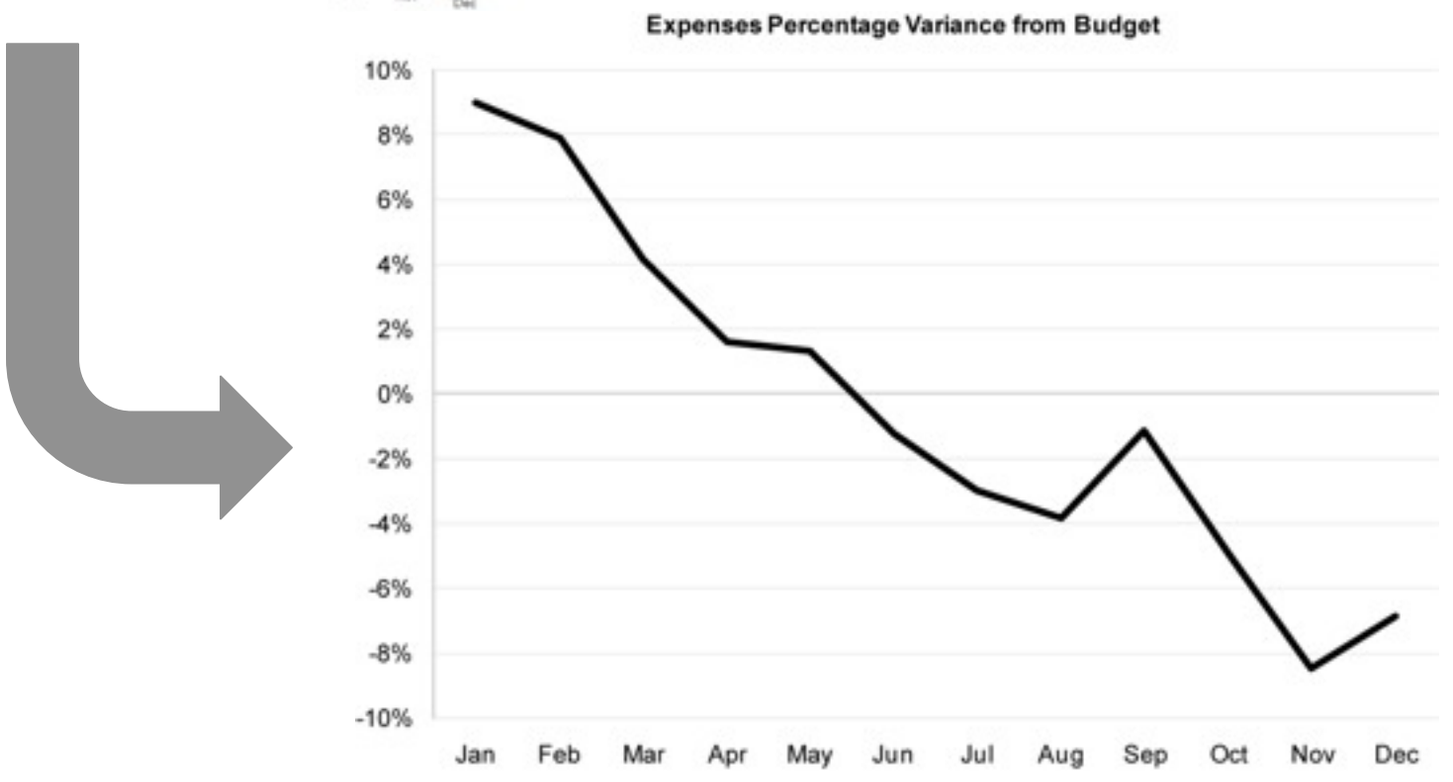
Poor graphs can be transformed.



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.

Data-Ink Ratio

Data Ink

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

Non-Data Ink

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

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.

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

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

987349702756479021947286240924060370804702890727
803208029007305901270238008374082078720272008083
247802602703793715709701379706674620970941027806
927979709123097230919592750927309272197873497260

9873497027**5**6479021947286240924060370804702890727
8032080290073**0****5**901270238008374082078720272008083
24780260270379371**5**709701379706674620970941027806
92797970912309723091**9****5**927**5**0927309272197873497260

How many fives are in this list of numbers?

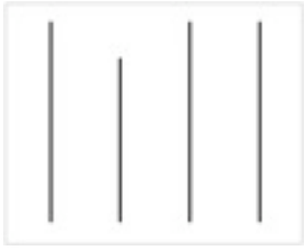
Text, the written form of verbal language, must be processed serially. Because the top list above consists of digits without spaces to group them into separate multi-digit numbers, you must read them one digit at a time.

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

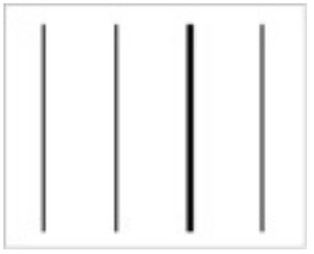
Some visual attributes are easier to see and distinguish than others. The most powerful of these are called *pre-attentive attributes* because we perceive them immediately, without conscious thought.

Pre-attentive attributes of visual perception

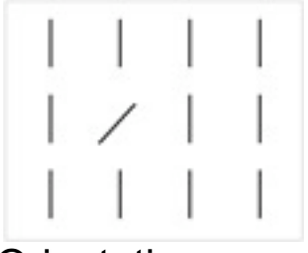
Form



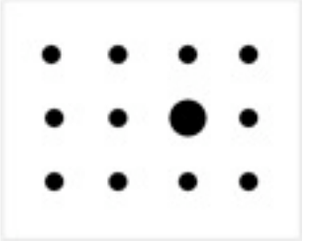
➡Length



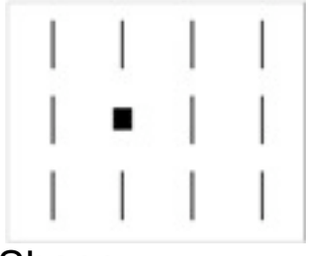
⇨Width



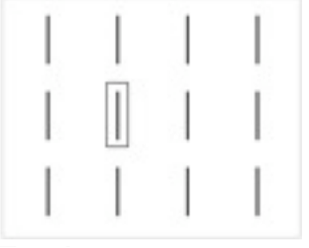
Orientation



⇨Size

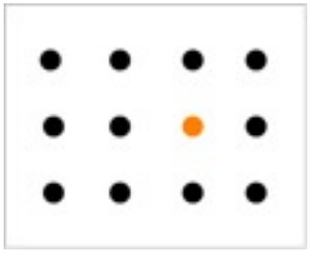


Shape

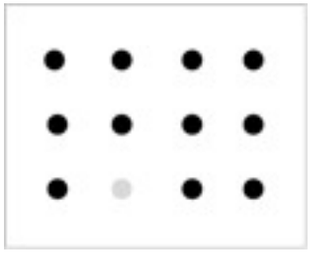


Enclosure

Color

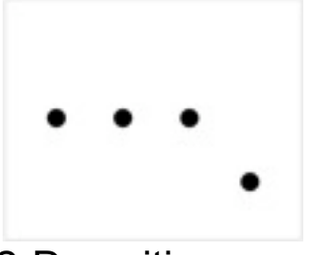


Hue



⇨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.

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

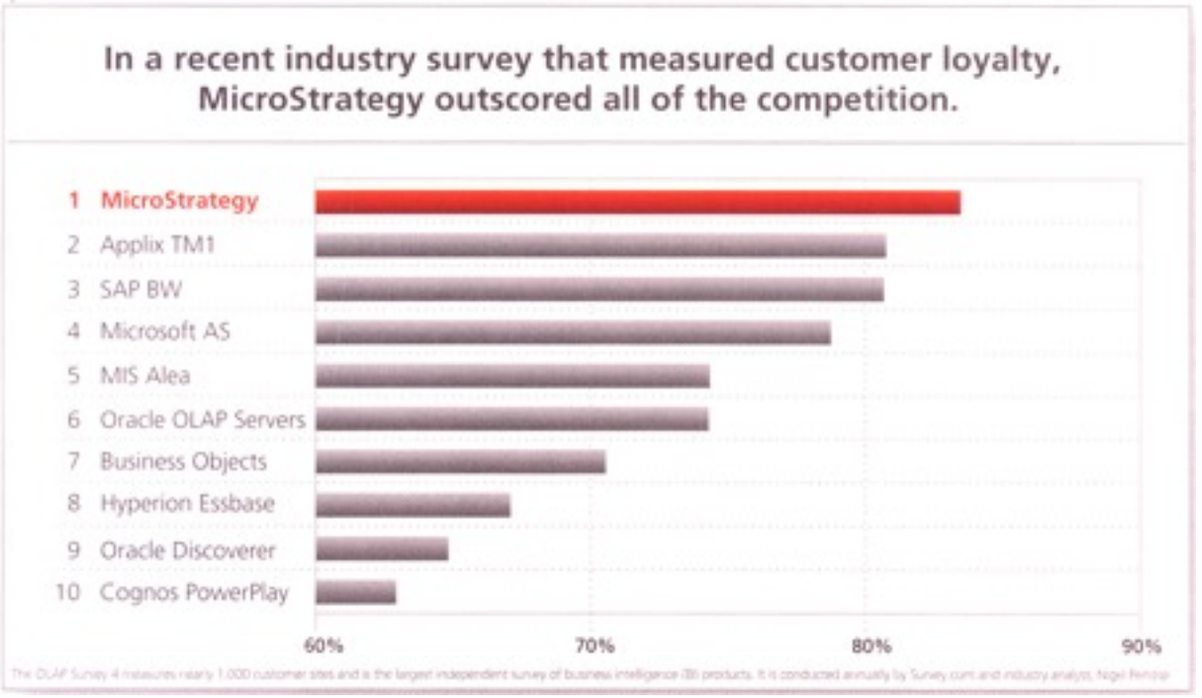
Text should be left aligned. Text includes numbers that do not represent quantitative data, such as the product and customer codes in these examples. The only exception to the left-alignment rule is when the text values in a given column all contain the same number of characters and that number of characters 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

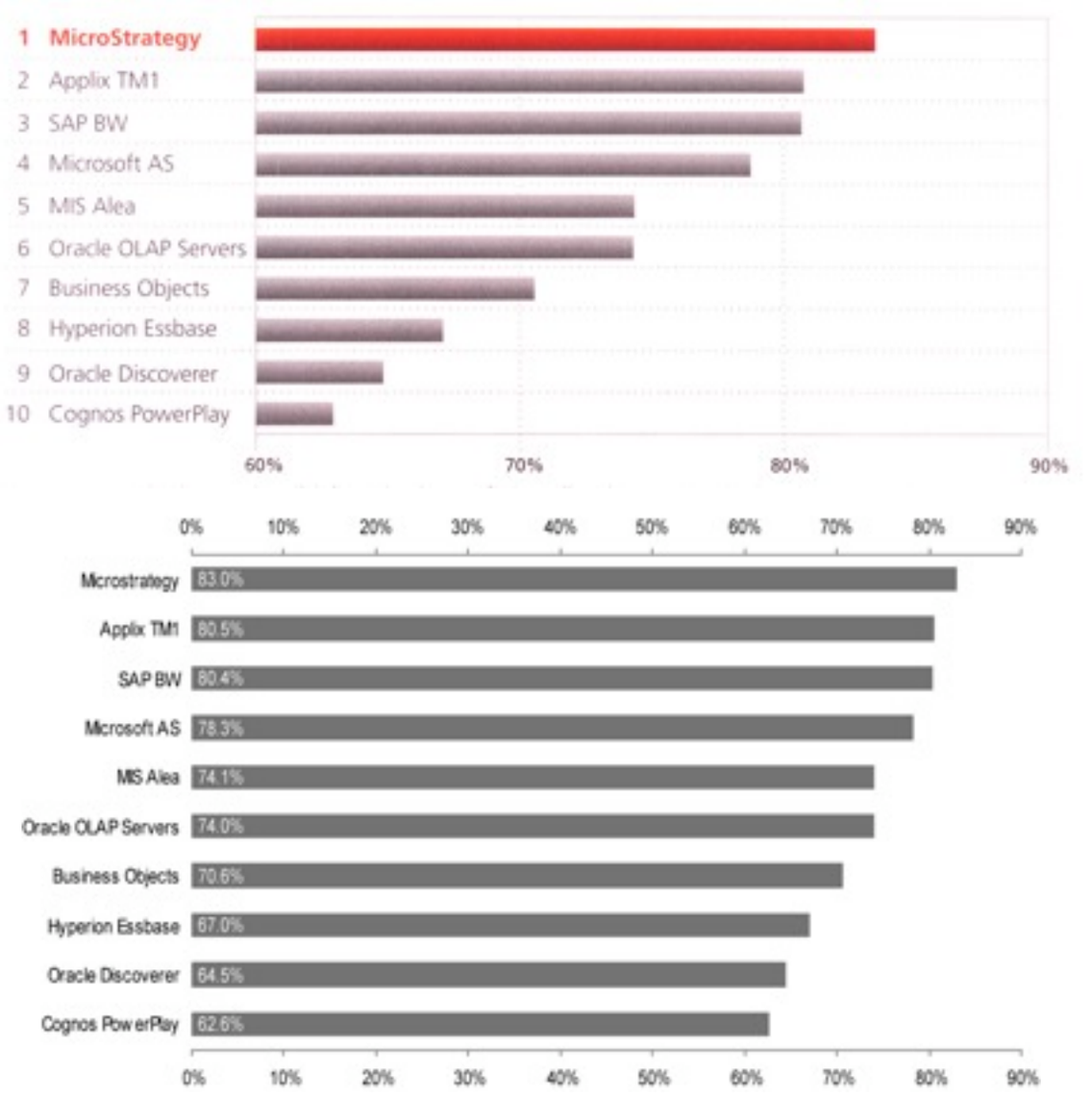
Whether dates are aligned to the left, right, or center is not important. What is important is that all date that appear in the same column contain a consistent number of digits to express the month (that is, two), a consistent number of digits to express the day (that is, two), and a consistent number 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.

What's wrong with this bar graph?



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

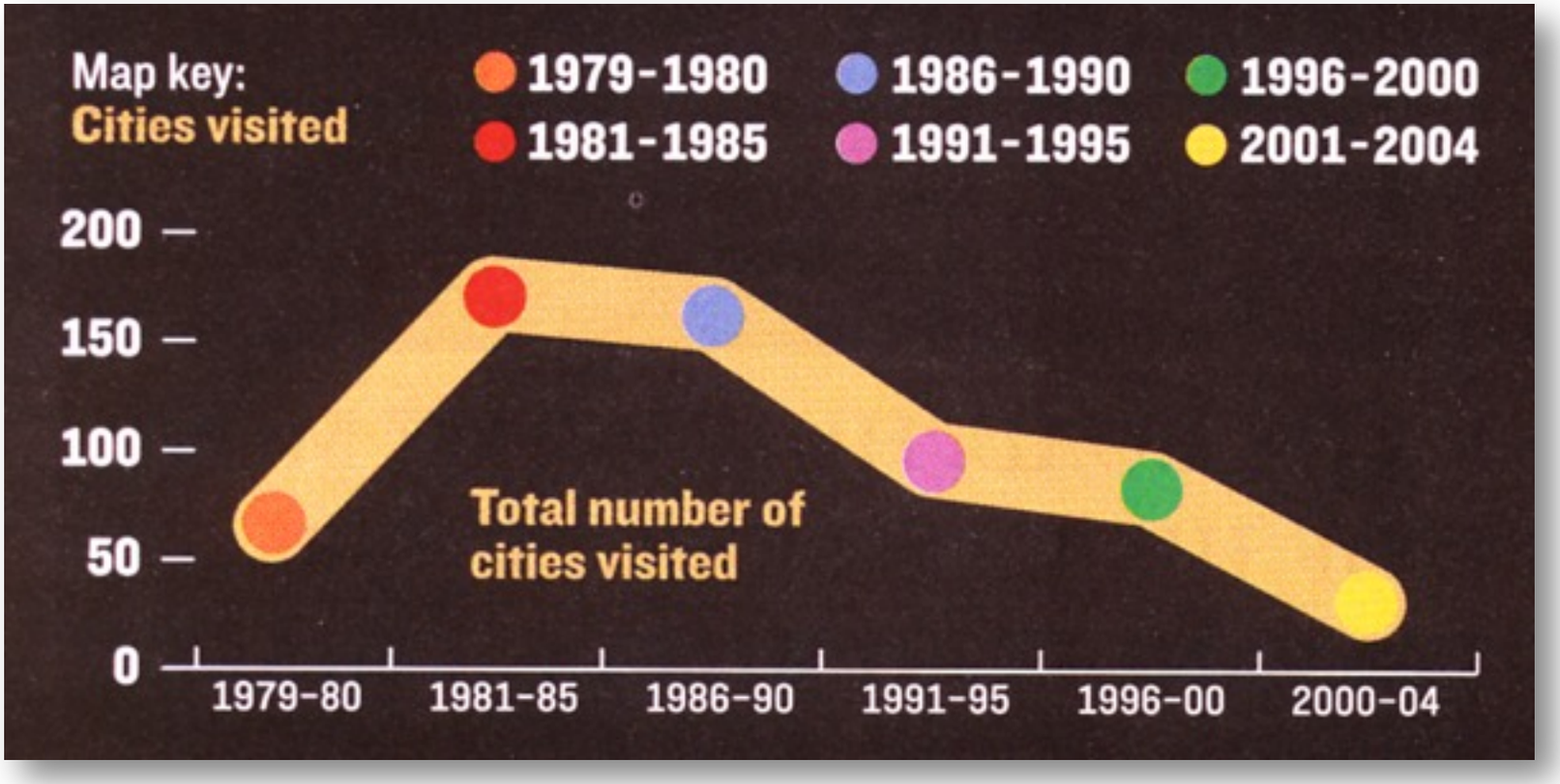
Bars should always begin at zero.



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.

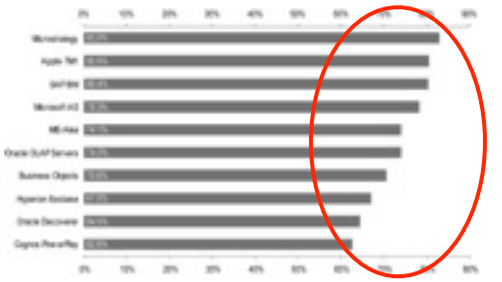
MicroStrategy's customer service score is only 33% greater than the lowest scoring product, Cognos PowerPlay, but the difference in the lengths of these two bars graph is 650%. That's a lie factor of 617% (650% / 33% = 617%). Notice how different the data looks when encoded accurately in the bottom graph.

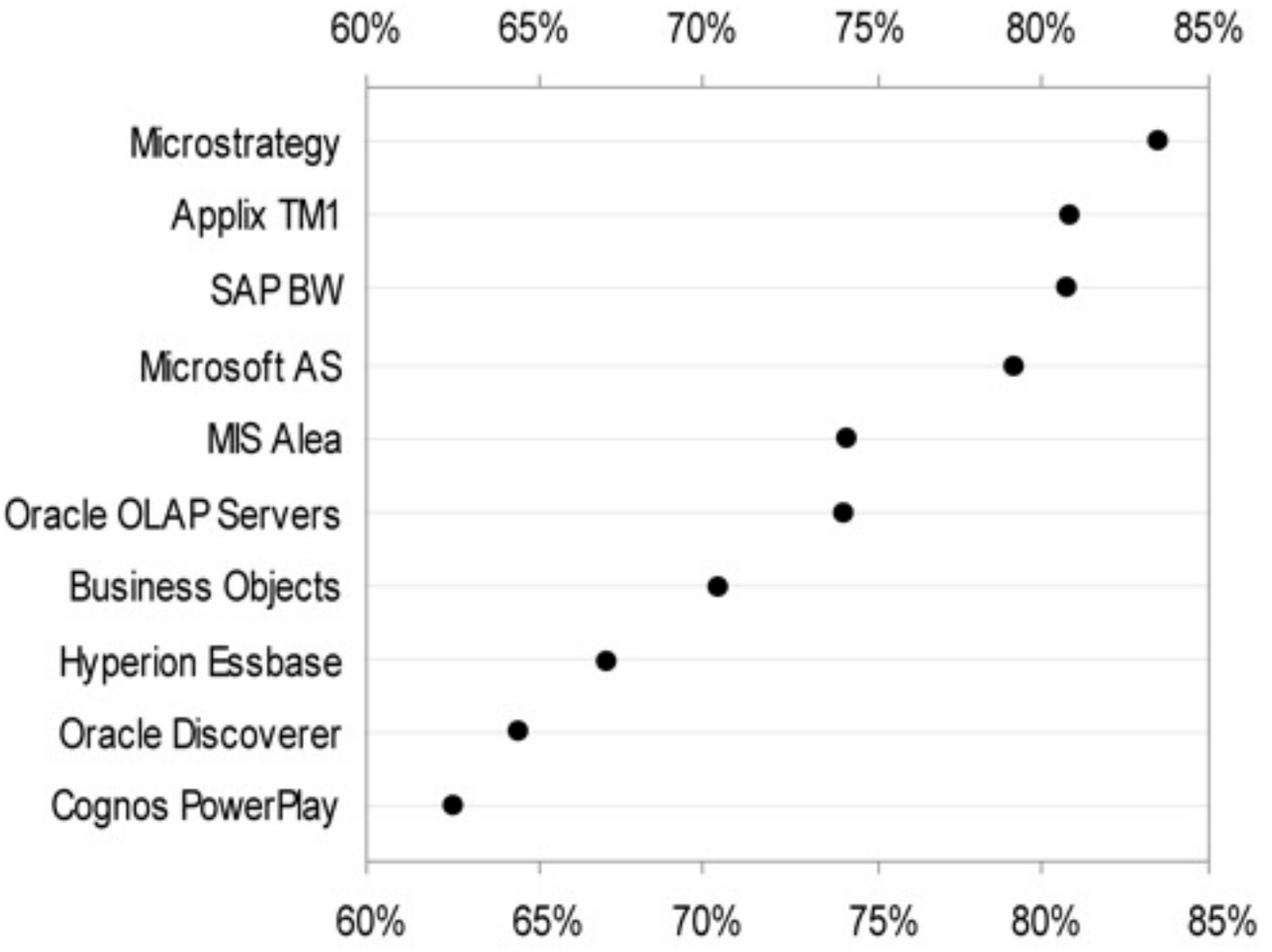
What's wrong with this graph?



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.

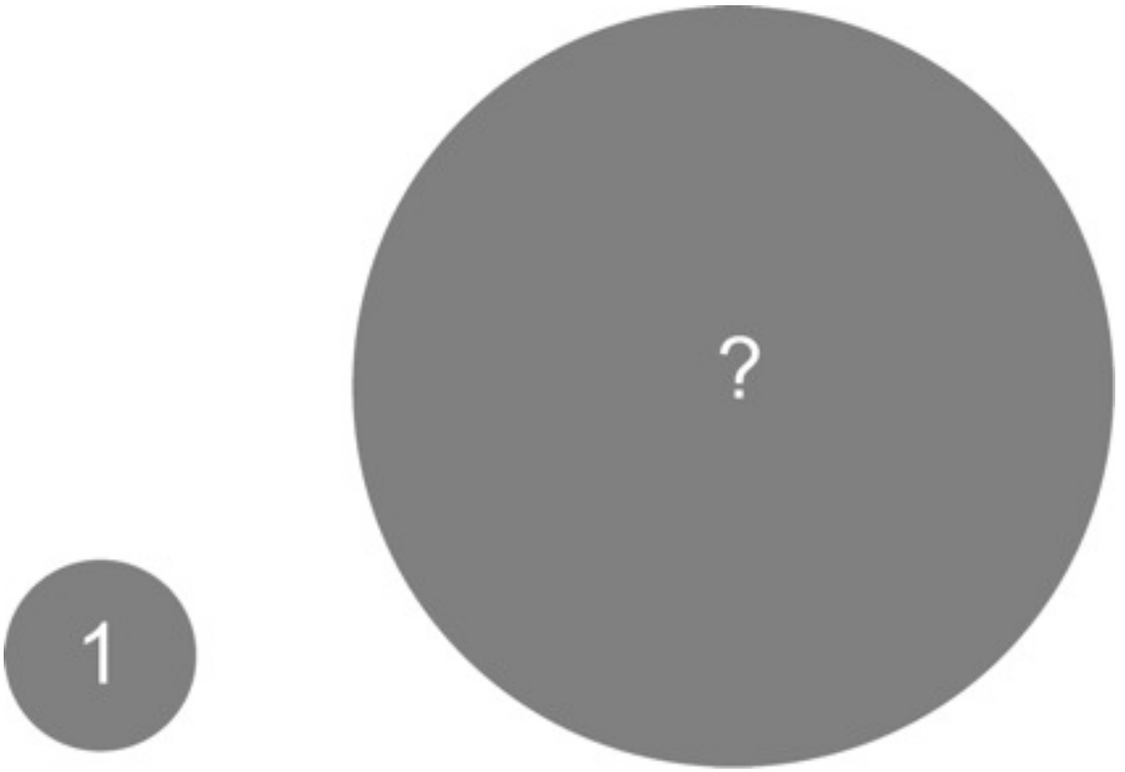
Another strength of points





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.

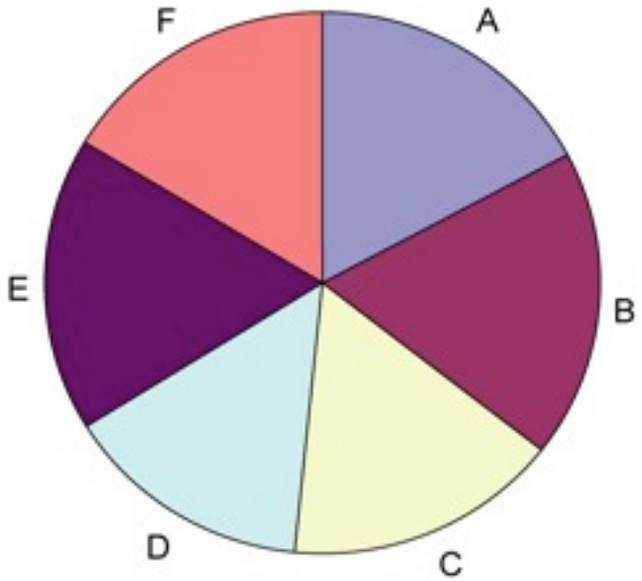
Why not encode data as 2-D areas?

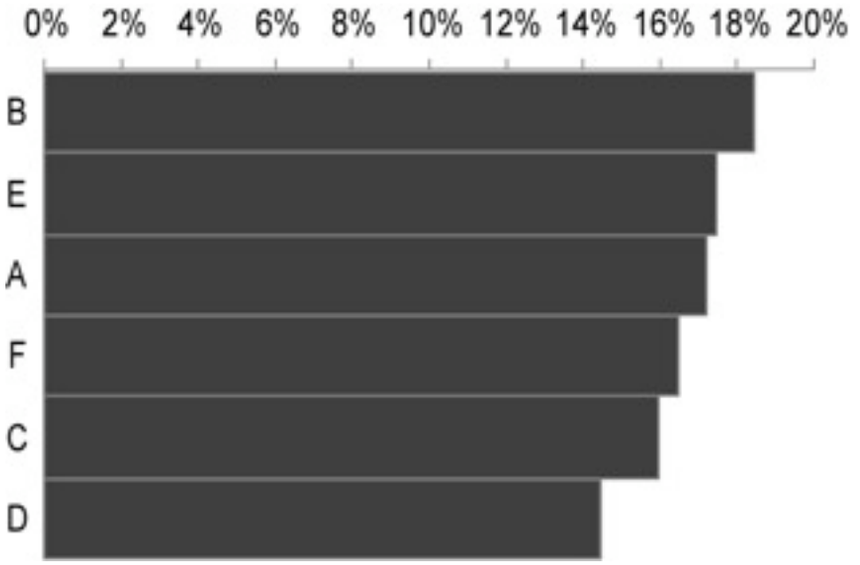


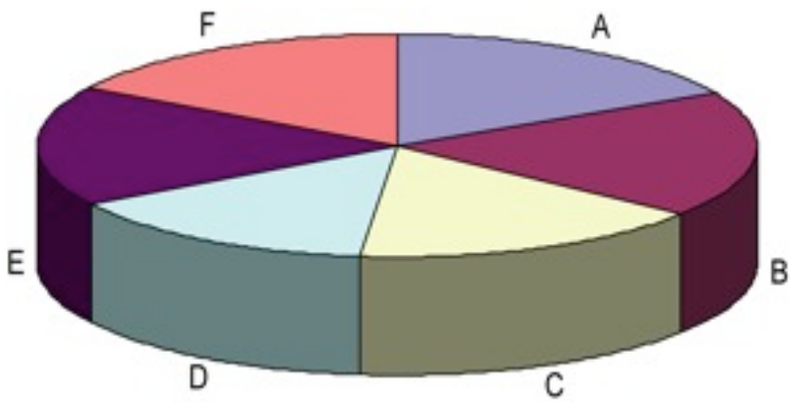
How big is the large circle? 16

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 an 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.*

Save the pies for dessert!

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.

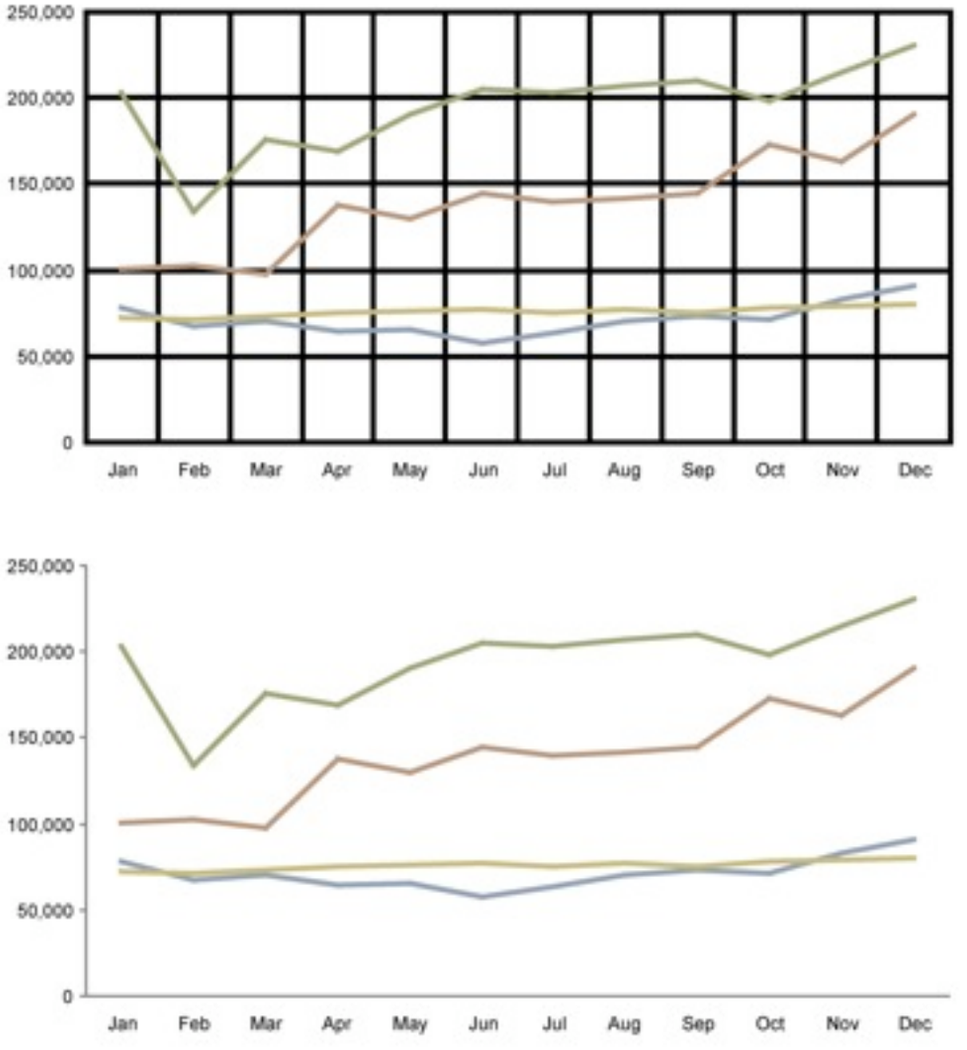
2. Enhance the data ink.

- 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.

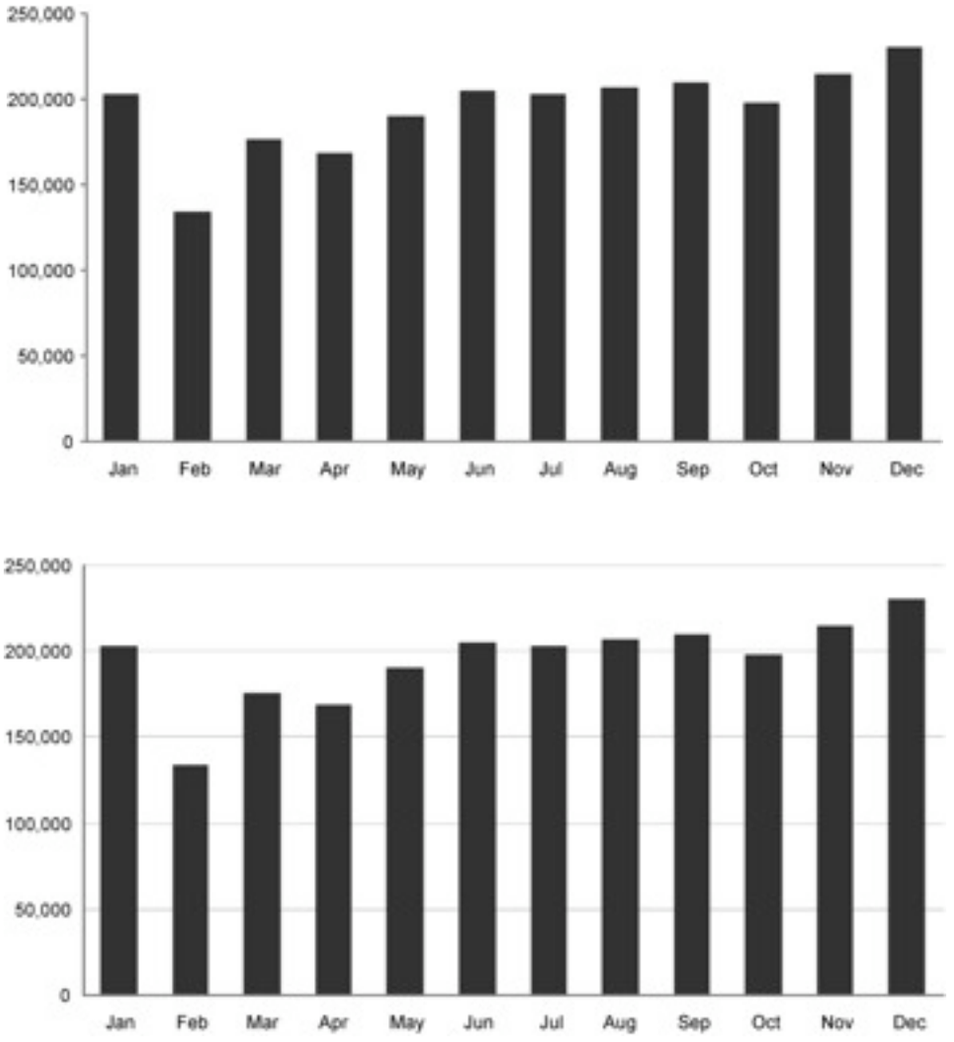
Now let's apply these steps to the design of graphs.

What about grid lines in graphs?



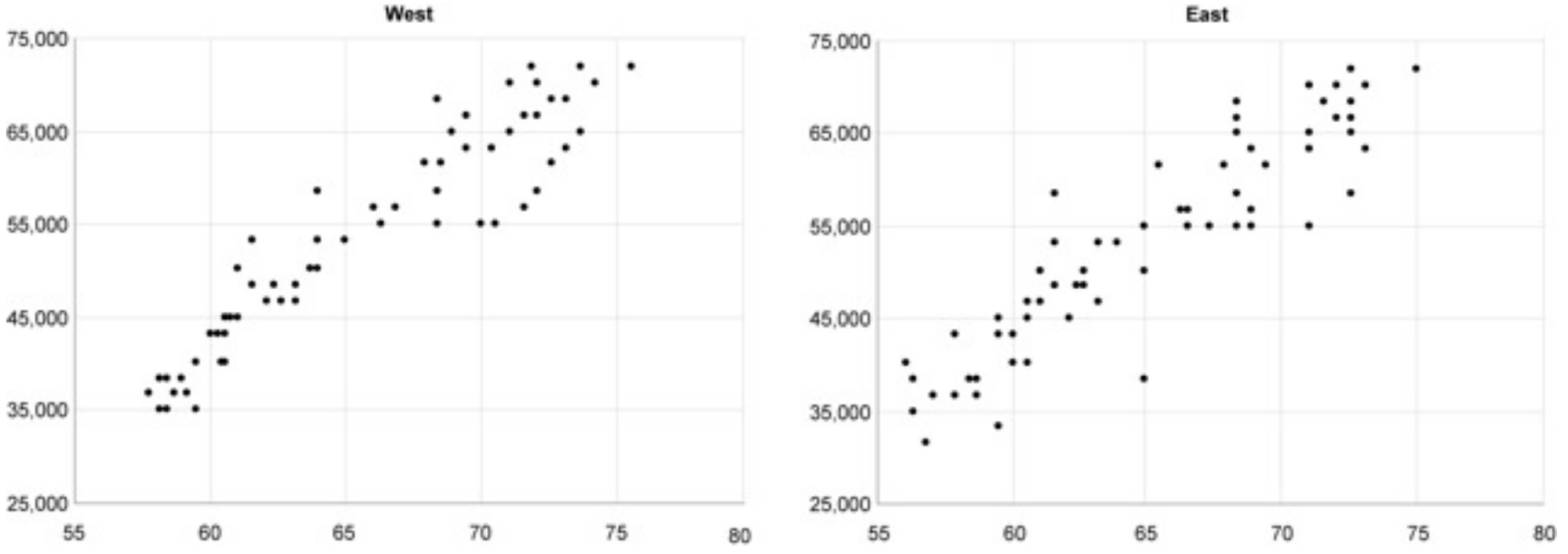
Grid lines are rarely useful and dark grid lines are never useful. They make it very difficult to pick out the shape of the data imprisoned behind the grid lines.

Grid lines clarify subtle differences.



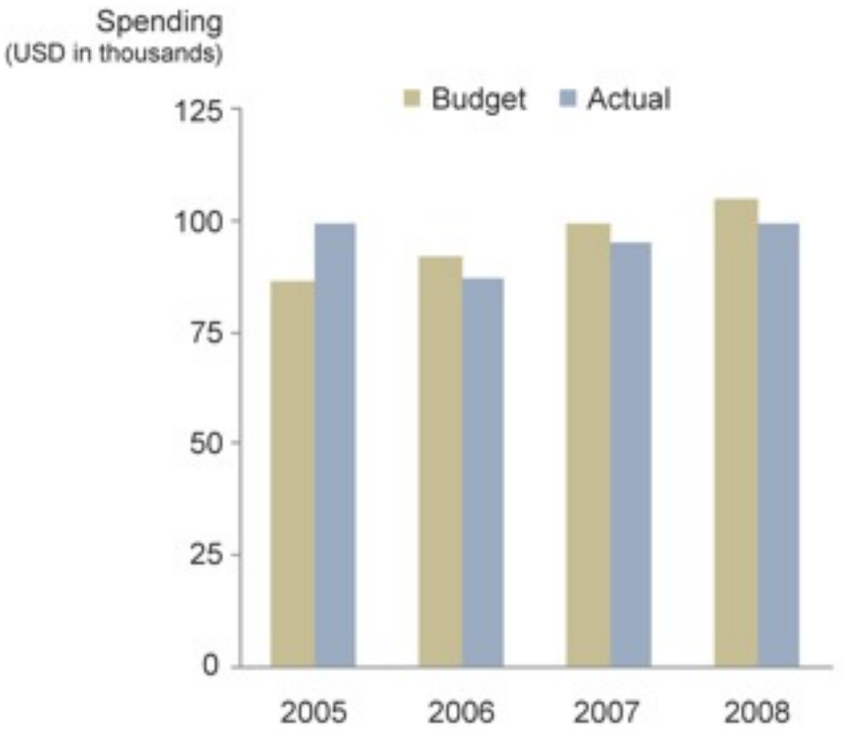
Light lines along the quantitative scale assist in making subtle distinctions.

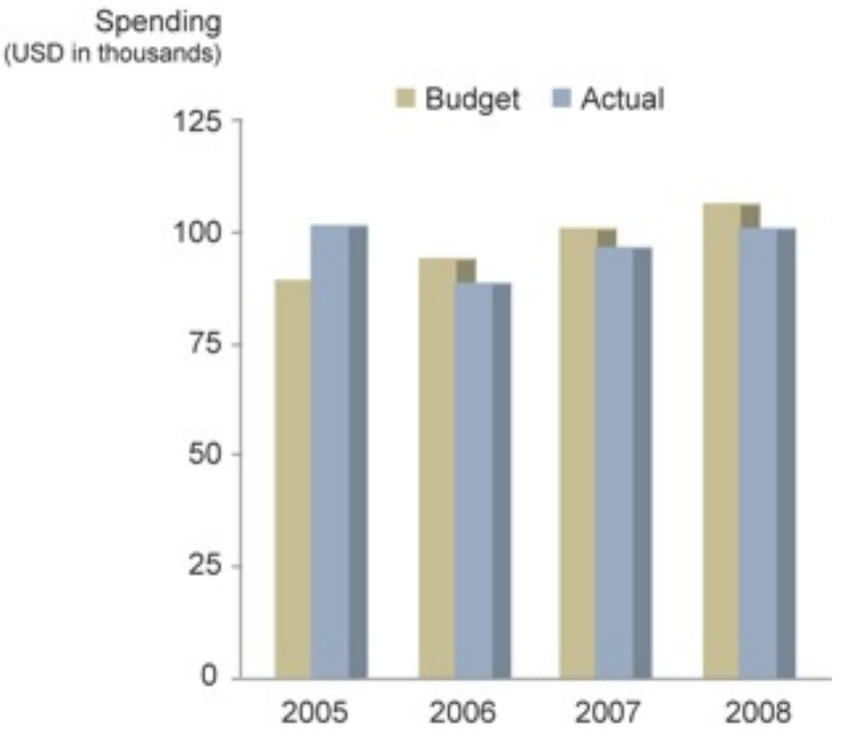
Grid lines delineate sections.



Light grids assist in reading and comparing subsections of graphs.

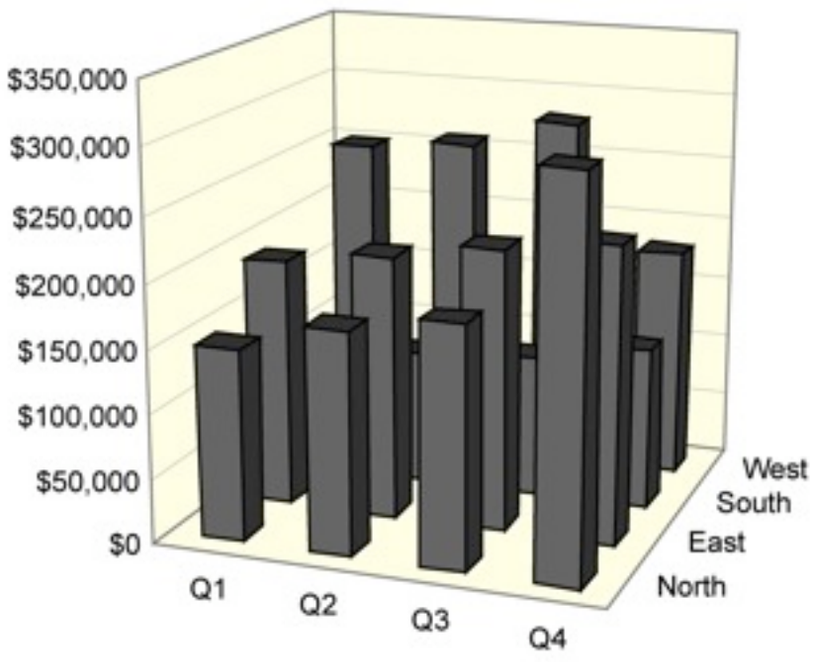
What about 3-D? It’s so cute!!!

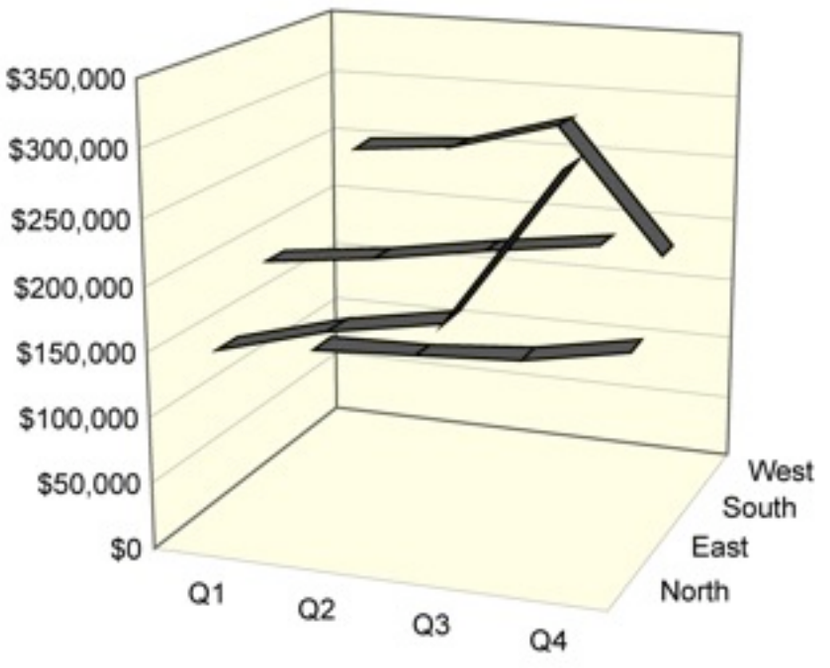




Adding a third dimension of depth to the bars on the right without adding a corresponding third variable is not only meaningless, it makes it more difficult to decode the data.

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?

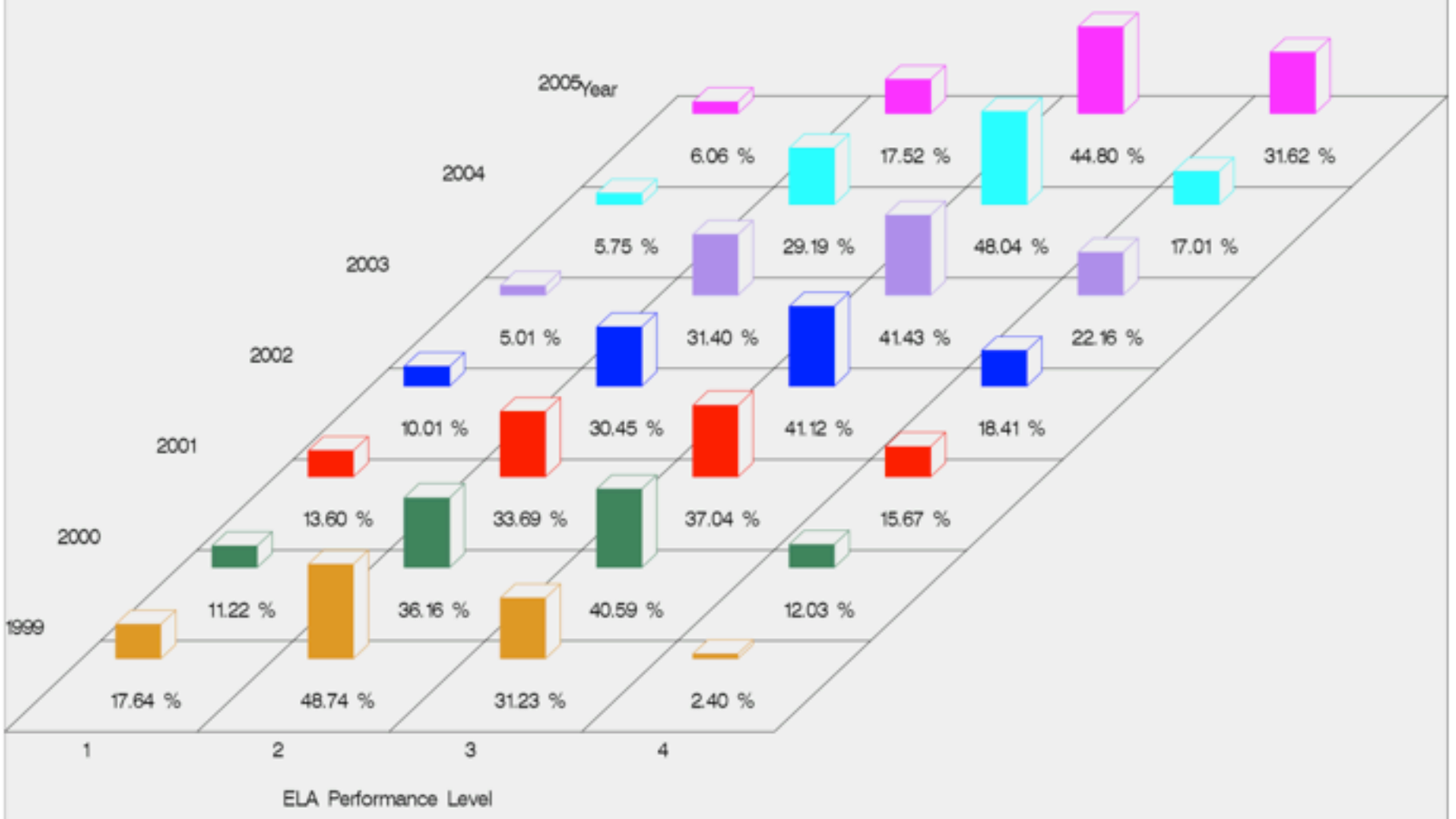
A third dimension with a corresponding variable is too hard to read.

YONKERS PUBLIC SCHOOLS

ELA Grade 4, 1999 through 2005

Percent By Performance Level

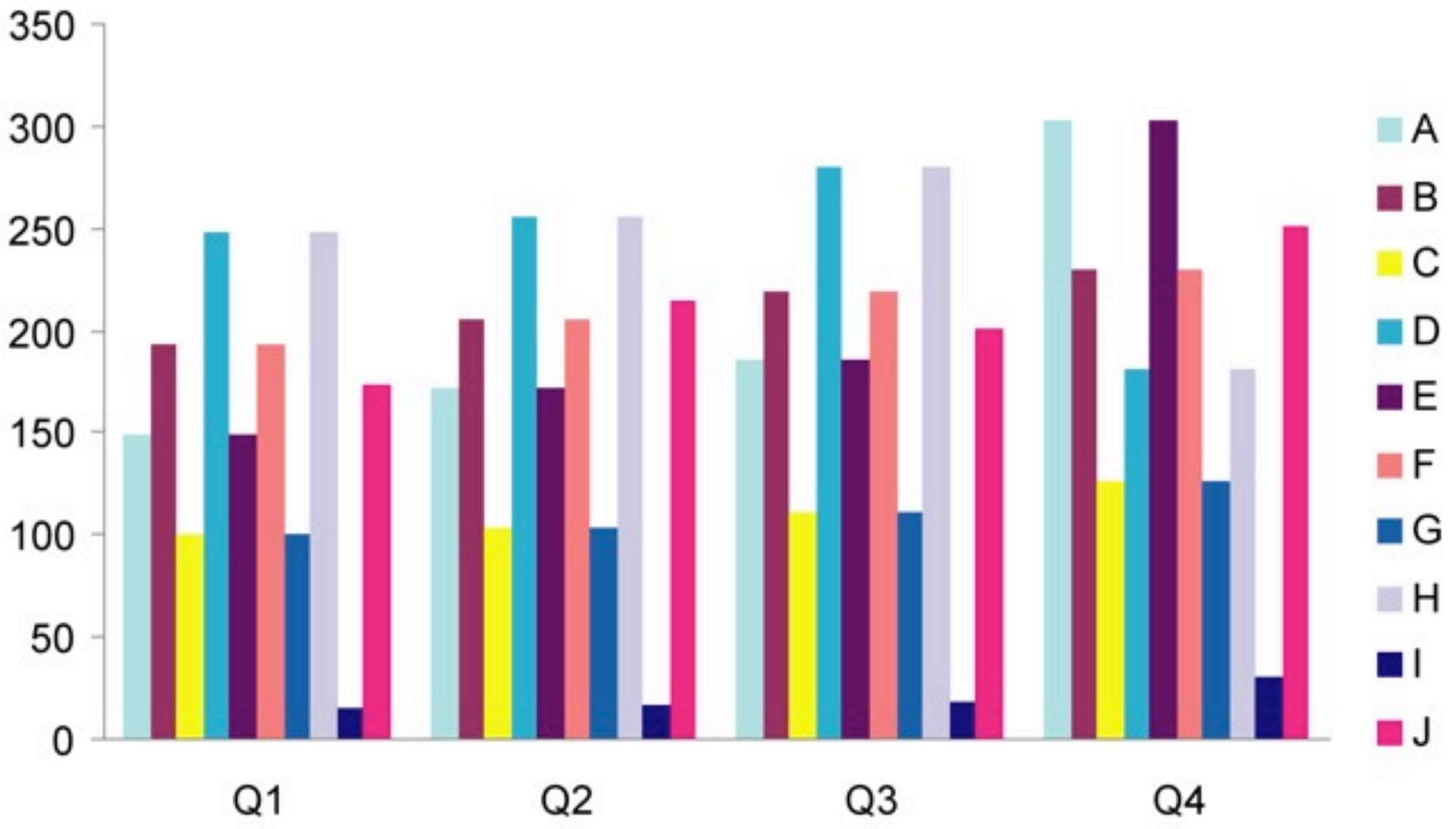
All Students, Districtwide, by Year



PERFORMANCE LEVEL = SCALE SCORE RANGE: 4=692–800 3=645–691 2=603–644 1=455–602

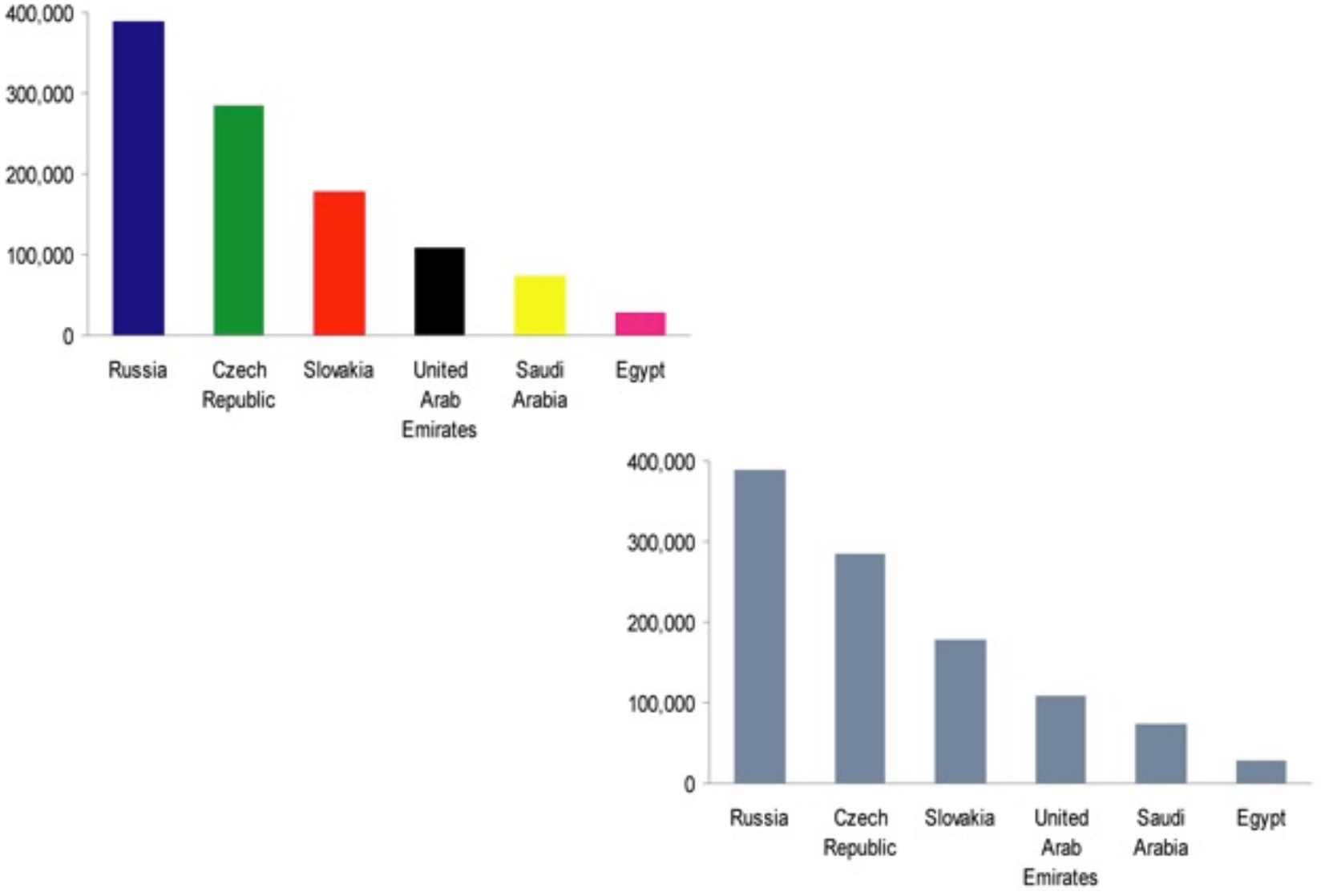
AT/ABOVE PROFICIENCY = PERFORMANCE LEVELS 3 & 4 BELOW PROFICIENCY = PERFORMANCE LEVELS 1 & 2

Can you ever show too much?



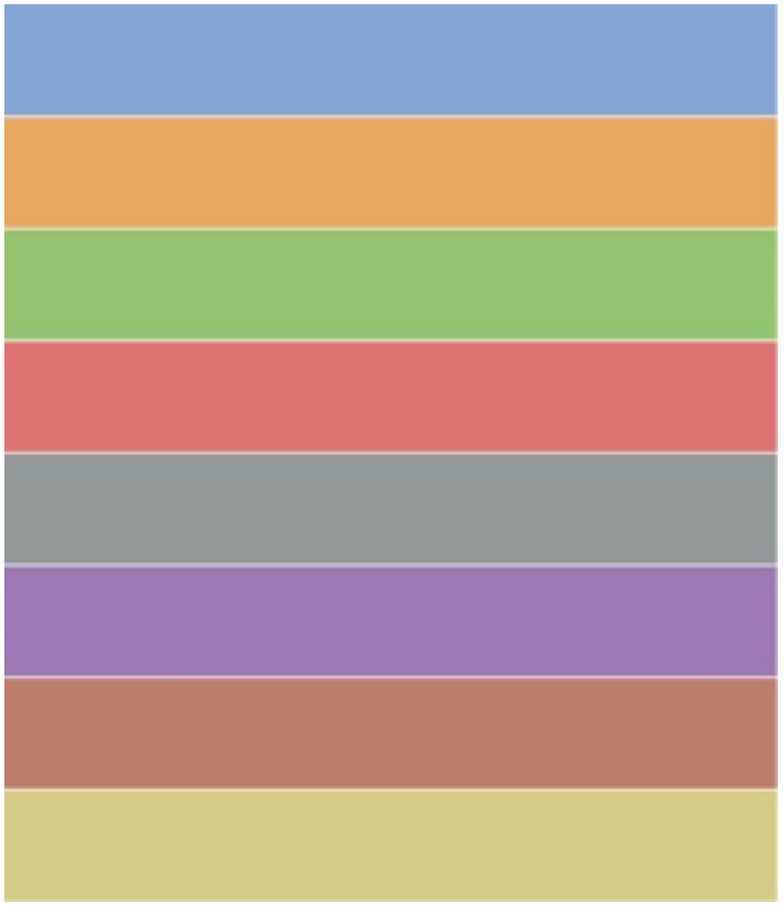
Limit the number of categorical subdivisions to around five, except when encoded by position along an axis. Short-term memory can only maintain about four chunks of information at a time.

Color choices make a difference.

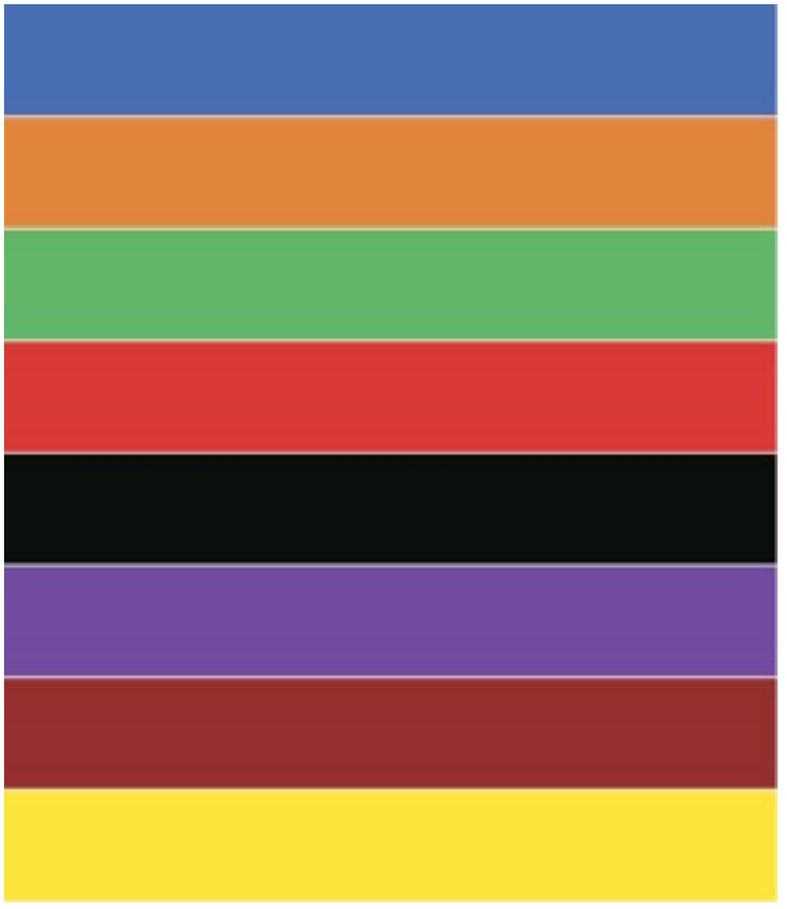


The top graph varies the colors of the bars unnecessarily. We already know that the individual bars represent different countries. Varying the colors visually separates the bars by making them look different from one another, but we want them to look alike to encourage people to compare them and to see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

Standardize on a good palette of colors.



Soft, natural colors for normal use

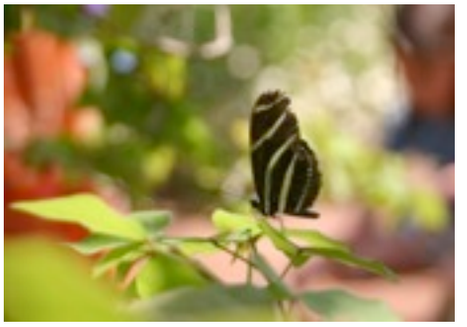


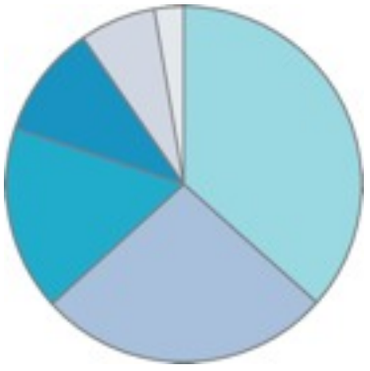
Fully saturated, dark colors for emphasis only

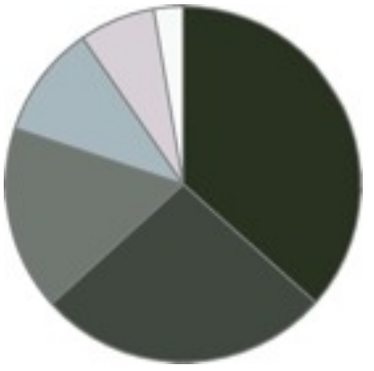
Soft, natural earth tones work best for everything except data that needs to stand out above the rest. Use colors that are fully saturated or dark only for highlighting data. If your software allows you to customize your color palette, it will definitely save you time to do this once, then rely on those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

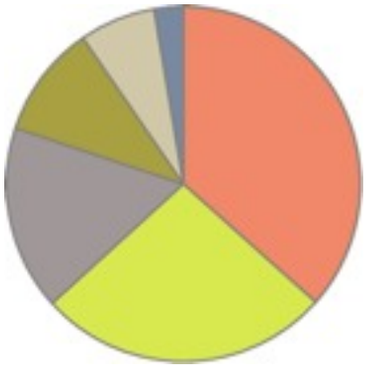
One of the best resources for selecting effective colors for data visualization is the free Color Brewer application that was developed by Cynthia Brewer for use on maps, which can be found at www.colorbrewer.org.

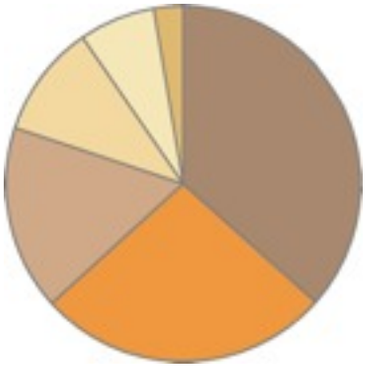
Nature’s colors are polite.









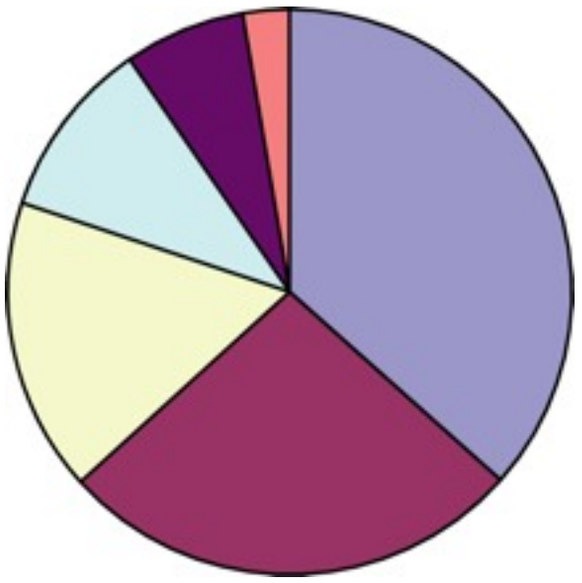


Fully saturated colors are found less often in nature than you might assume. The bright, saturated colors that appear in the bars of this graph cannot be found anywhere in these four photographs of the natural world.

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.

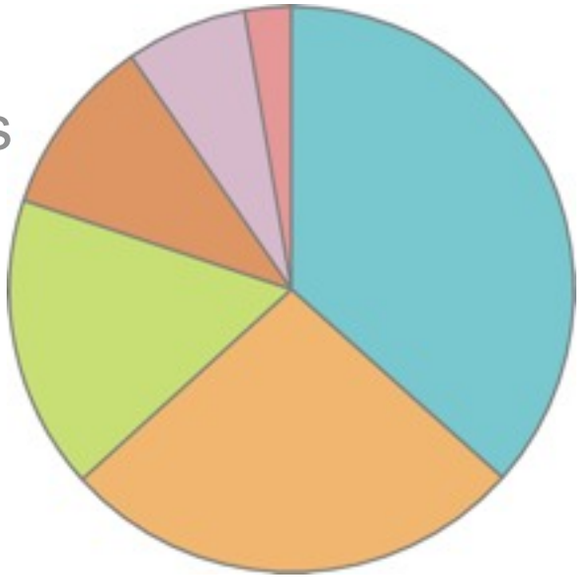
By the way, by displaying these palettes of colors in the form of pie charts, I am not suggesting that you should use this type of graph. A pie chart just happens to provide a nice means to show several color together, but it doesn’t provide an effective means of displaying data.

Which would you rather look at?



Excel 2003 defaults

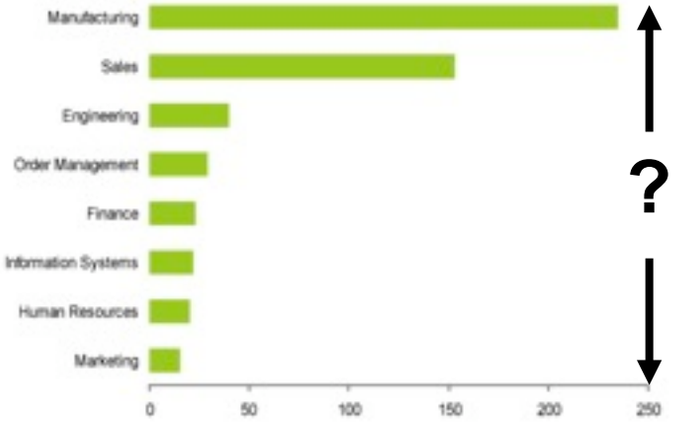
Fully saturated

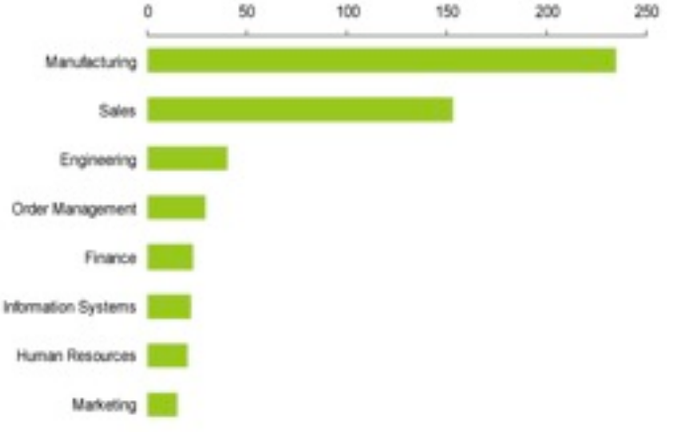


Natural

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.

Best location for the quantitative scale





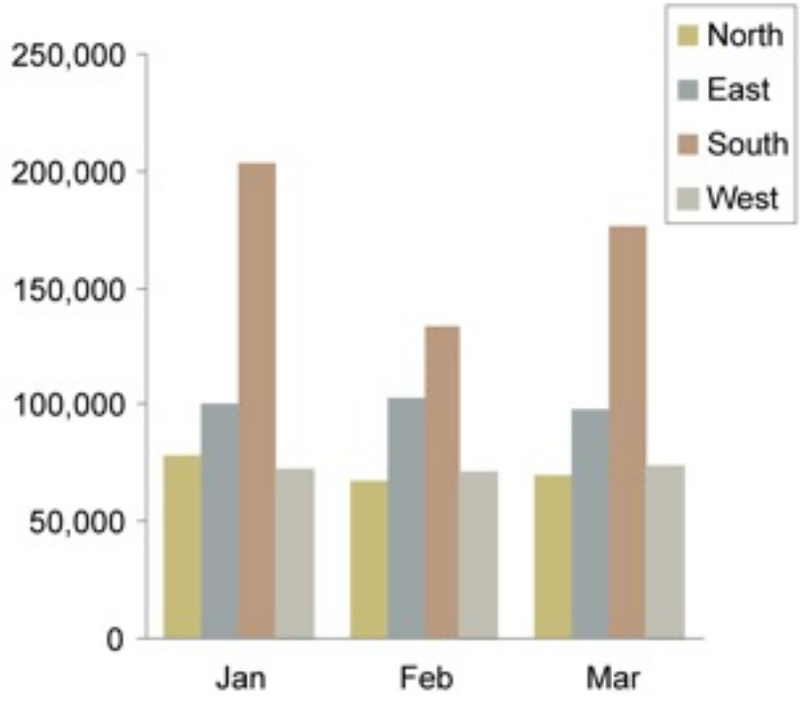
Quantitative scales are usually placed on the left or bottom axis, but there is no reason that they can't be placed on the right or top axis if that offers an advantage. As a general rule, place the quantitative scale on the axis that is closest to the most important data in the graph. For instance, 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).

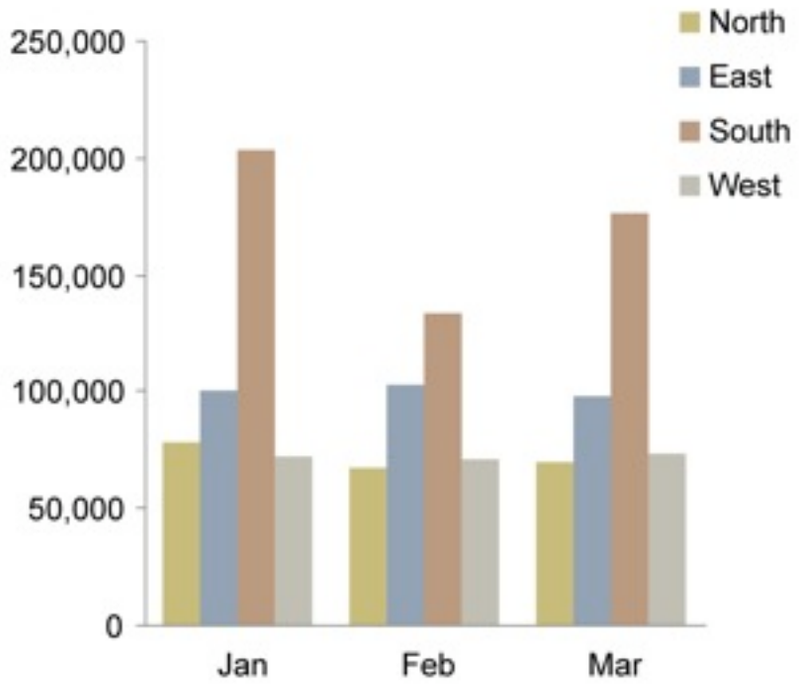
Best range for the quantitative scale

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.

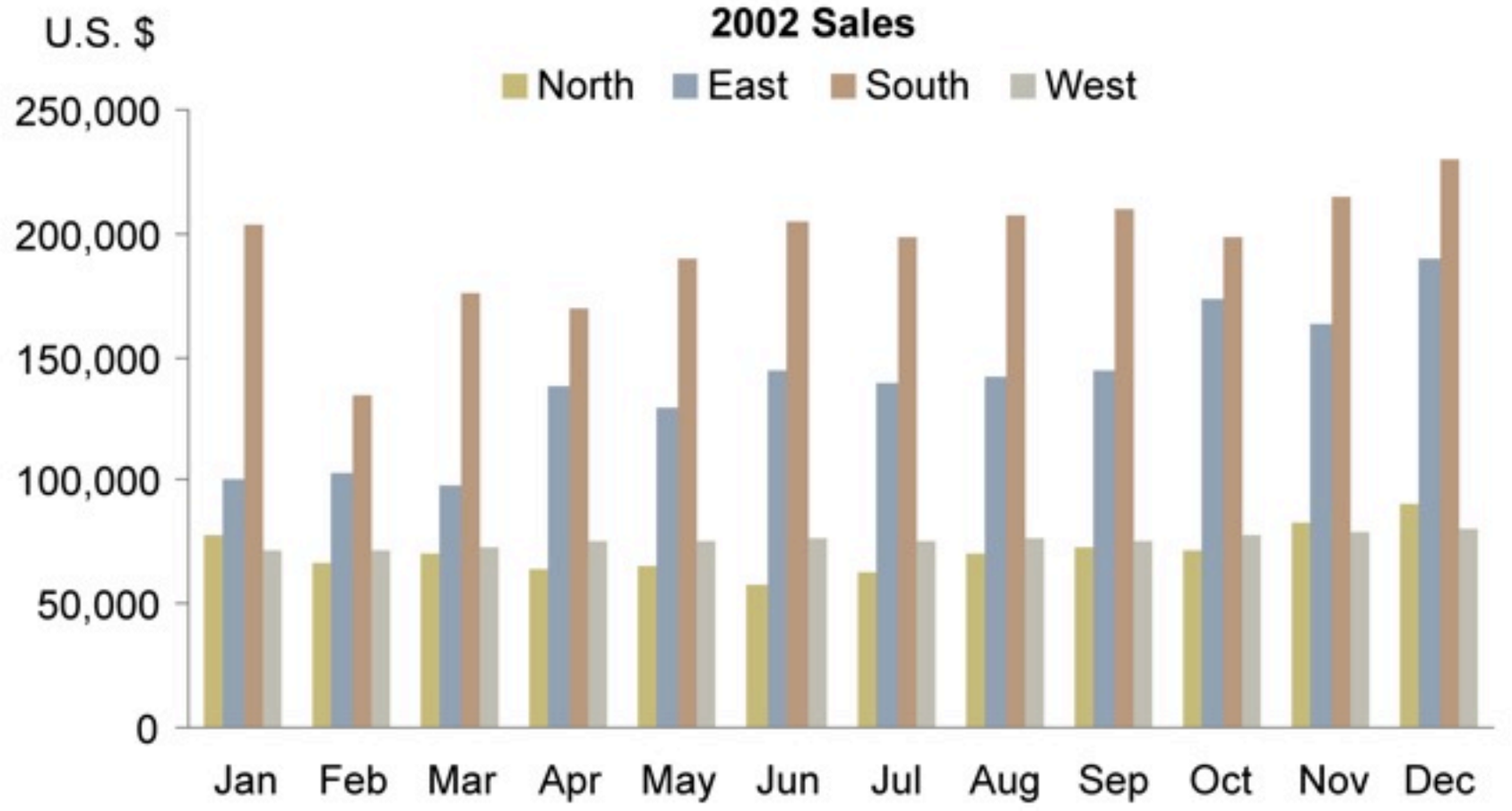
How can you minimize problems associated with legends?





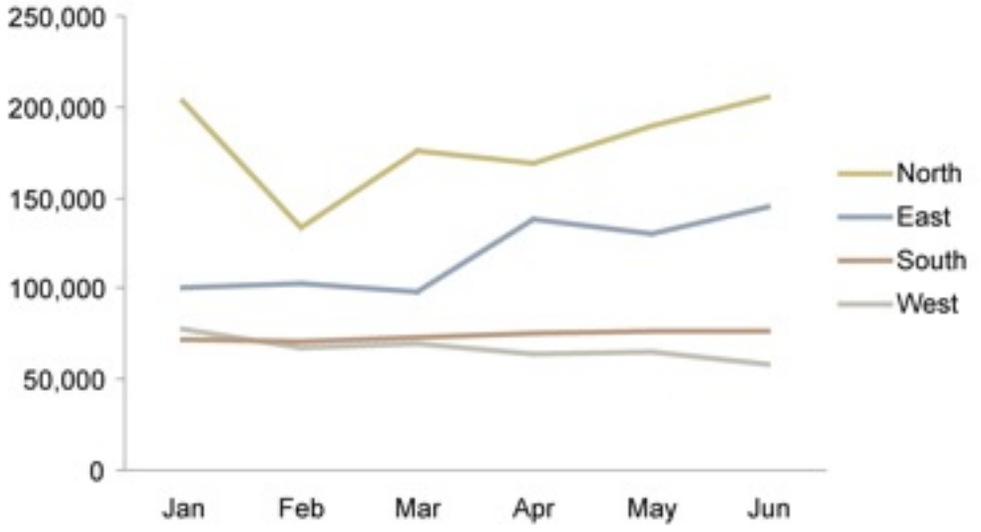
Borders around legends rarely add value.

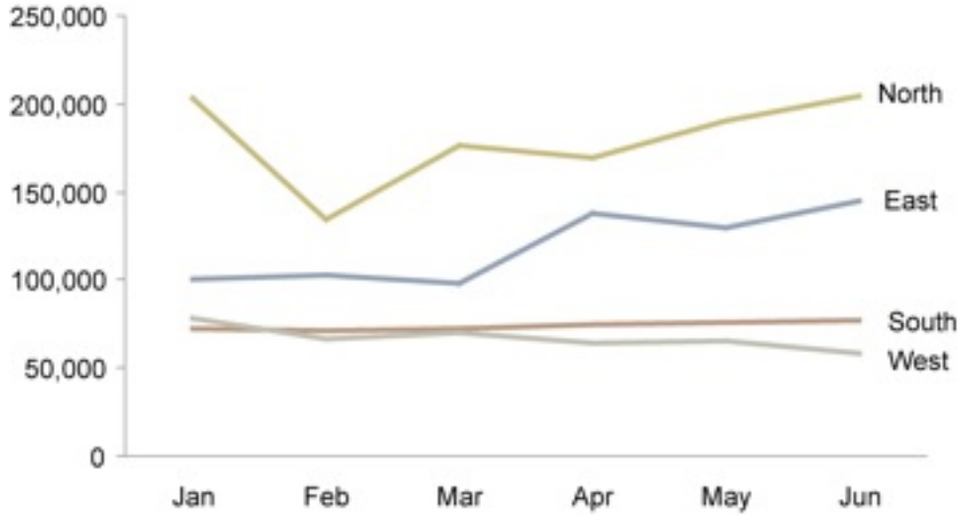
How else can you improve legends in bar graphs?



You can arrange the labels horizontally in the same order as the bars. This links the legend more tightly with the bars.

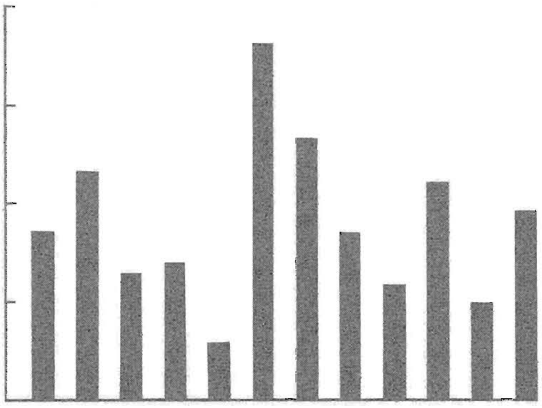
What about legends for line graphs?



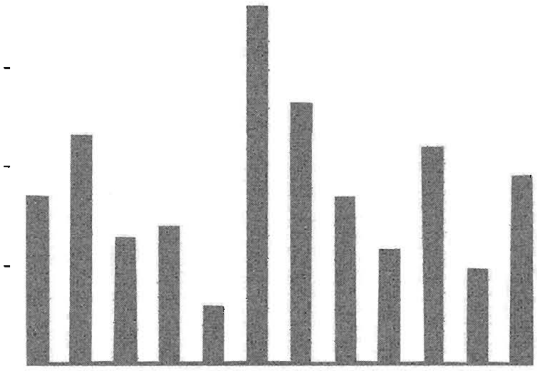


Place the labels next to the corresponding lines whenever possible. This saves the reader considerable time.

The box can be erased:



And the vertical axis, except for the ticks:



Even part of the data measures can be erased, making a *white grid*, which shows the coordinate lines more precisely than ticks alone: