A Picture is worth a Thousand words

How Graphs can be used to Cheat!

Allen Gapara

Table of Contents

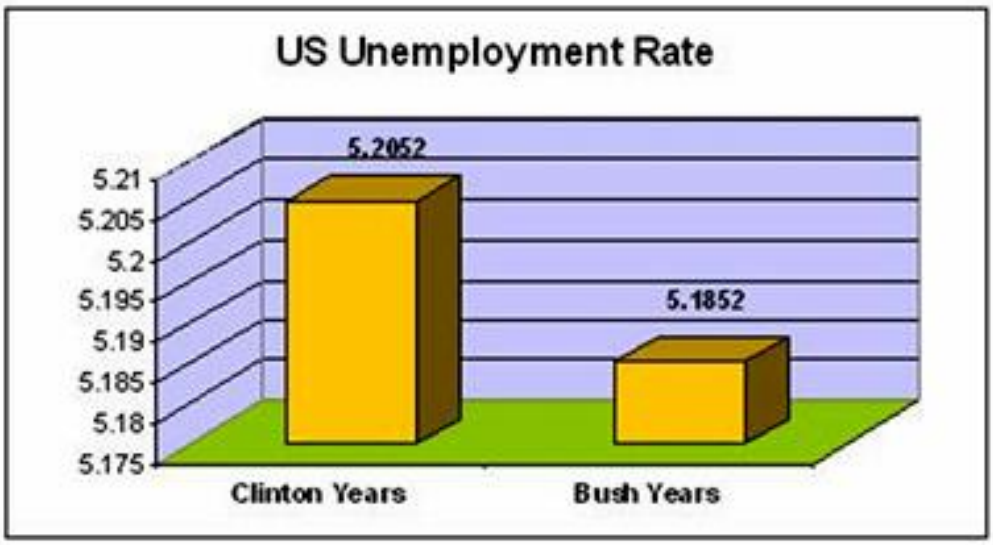
Introduction

Good graphs are extremely powerful tools for displaying large quantities of complex data; they help turn the realms of information available today into knowledge. But, unfortunately, some graphs deceive or mislead. This may happen because the designer chooses to give readers the impression of better performance or results than is actually the situation. In other cases, the person who prepares the graph may want to be accurate and honest, but may mislead the reader by a poor choice of a graph form or poor graph construction.

Scale Not Starting at Zero

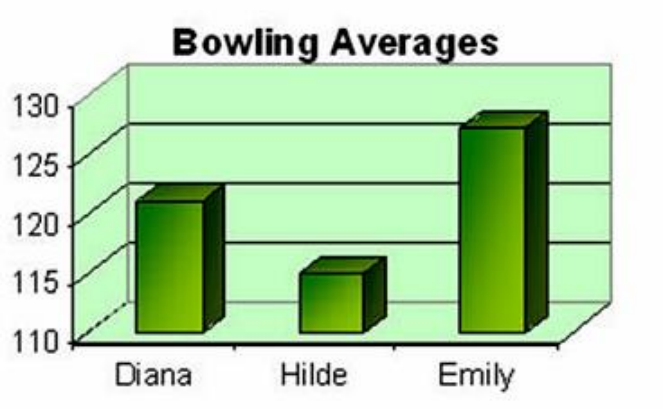
Here is another American Unemployment graph that has made a tiny difference look huge by using small scale units, and drawing the bars in 3D.

The most common trick with graphs is to not make the scale start at zero. This makes small differences look much bigger.



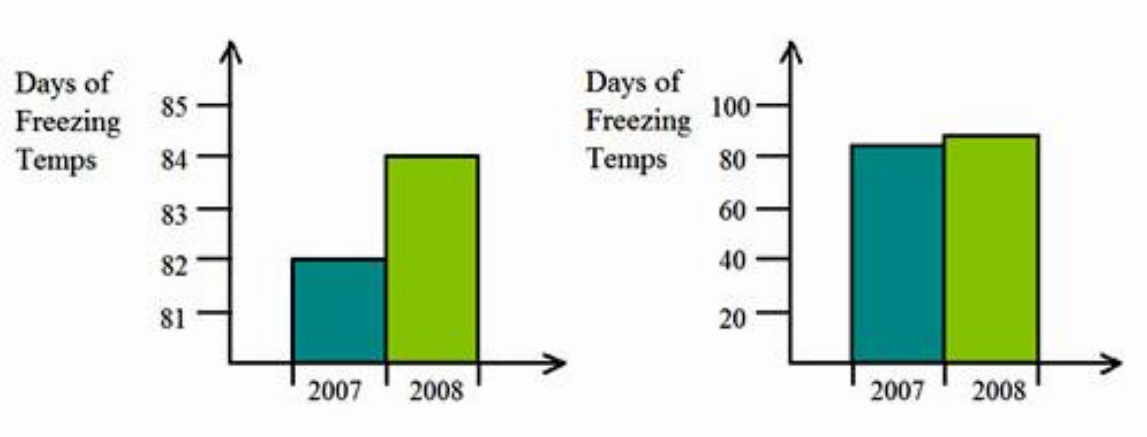
In this next graph, the bowling average difference between Emily and Diana is only 5, but is made to look much larger.

Sometimes the scale does not start at zero, and the graph is also put in a 3D perspective which confuses the eyes.



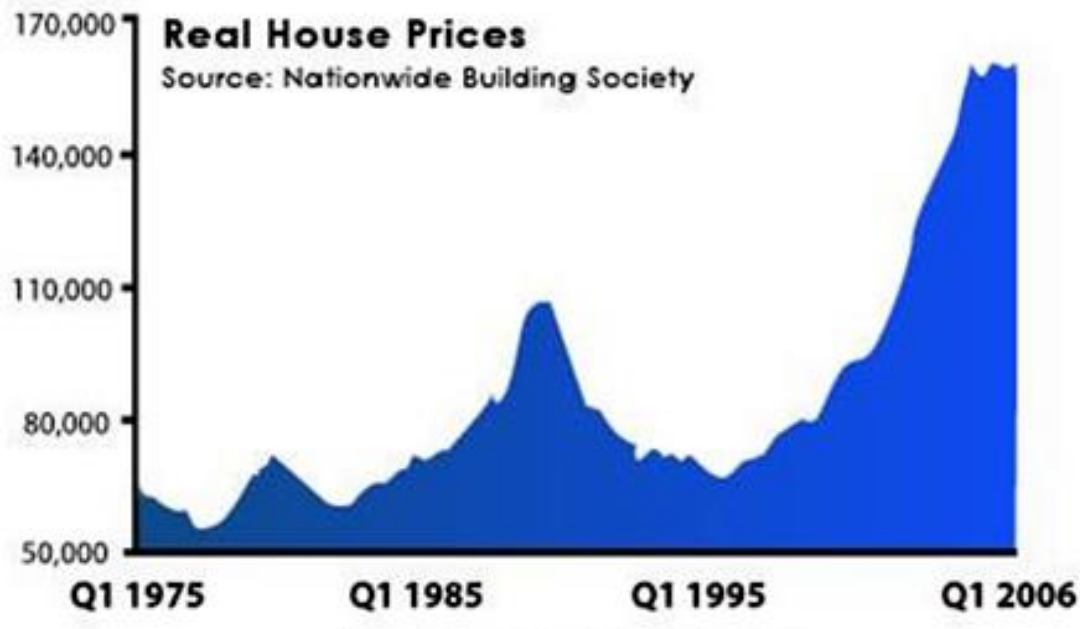
This next graph makes it look like there were twice as many freezing days in the USA in 2008 compared to 2007, when there were in fact only two extra freezing days in 2007.

Both show exactly the same data. However, the graph on the left makes the change appear to be much larger than it really is because the numbers on the vertical axis do not start at 0. Each vertical mark on the left graph represent 1 and each mark on the right represent 20 (the scale changes)



This next graph makes the House Prices in 2006 look huge compared to 1975.

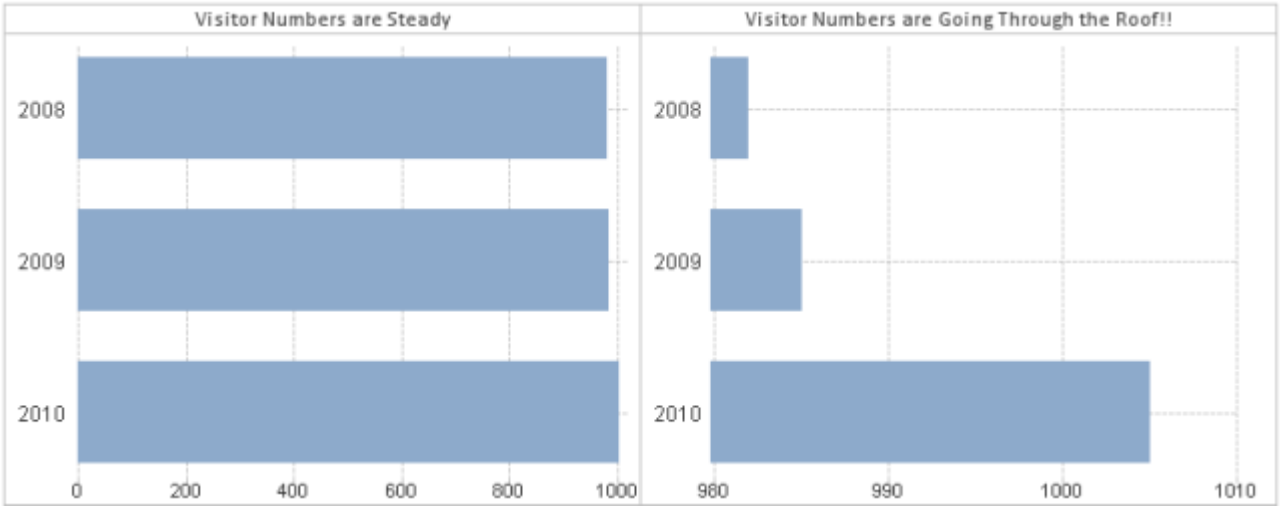
Also the next year after 1995 on the horizontal scale should in fact be 2005, and not 2006.



The scale does not start at zero, and makes the house price in 2006 look about seven times higher than 1975. When it is really only about two and a half times higher. (e.g. 160000 in 2006 is about 2.5 times 60000 in 1975)

To Zero or Not to Zero

Here we're looking at exactly the same basic dataset covering Visitor numbers per year, the chart on the left seems very placid and visitor numbers appear to be relatively steady year on year; not bad but not great, whilst the chart on the right looks much more positive; visitor numbers took a huge leap in 2010; cue extra investment and pay rises all round! The charts show identical data so what's going on?



The ONLY difference between these 2 charts is that the one on the left has the Properties > Axis > 'Forced 0' checked for the Y-Axis whilst the other doesn't; one check box and we're completely changing the message of this chart. Now of course if you spend a little time looking at this chart and read off the numbers it becomes clear that the increase isn't all that marked, however; from my experience people don't tend to read fine detail from charts; they simply look for generalizations; 'Visitors are up a lot', 'Sales are trending down' – the very definition of a dashboard, so in light of that the chart on the right can be very dangerous…or great if you're after a pay rise.

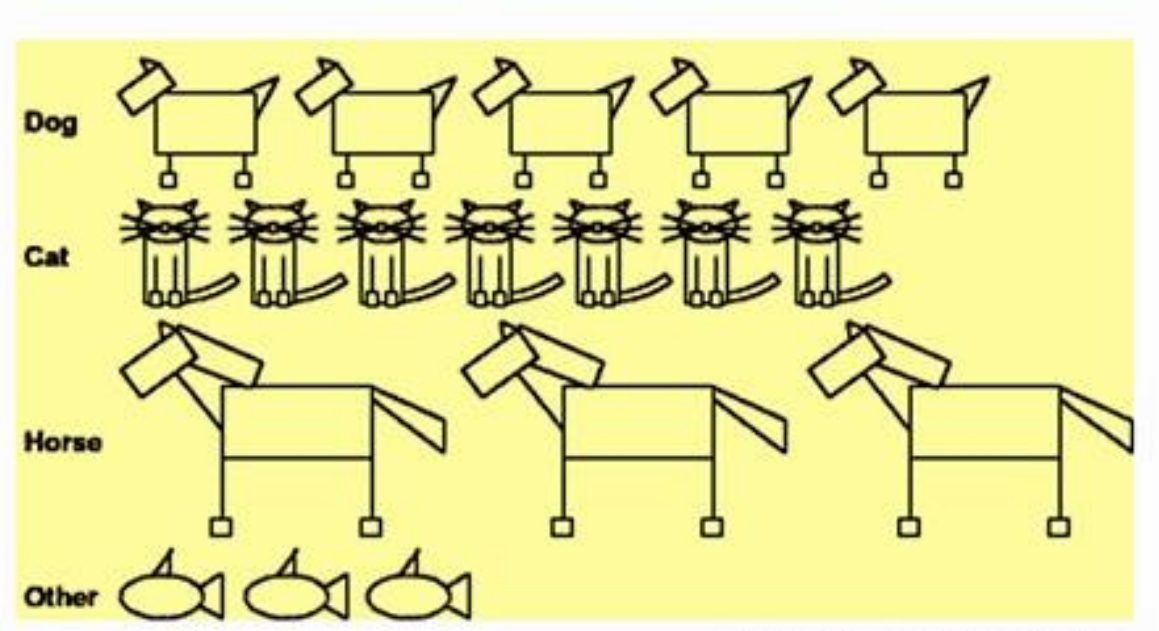
To accentuate the effect even further you can use the 'Static Min/Max' settings to zero in even more on the difference making the smaller value seem even smaller and the increase all the larger.

Misleading Picture Graphs

In Picture Graphs, (also called "Pictographs"), all of the image icons used to represent the different items should be made the same size.
By making these items different sizes, we can try and trick people who are looking at the Graph.

In the graph shown below, the Dogs have been drawn much bigger that the Cats, to make it look like there are more dogs.

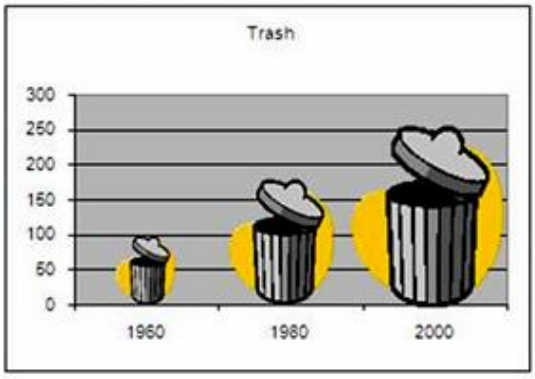
However, if we count the icons, there are actually only 5 dogs compared to 7 cats.

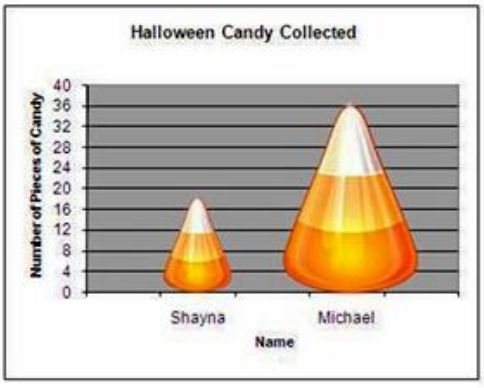


In Picture graphs, the items are sometimes all made different sizes to trick our eyes into thinking we have more of the bigger sized items.

Oversized Volumes on Graphs

In these type of graphs, images are used to replace normal vertical bars, and some of the volumes or sizes of these images are made much bigger than they should be.
In this picture Graph, the year 2000 is about eight times bigger than the 1960 image. However when we check on the vertical scale, year 2000 is only two and half times bigger. (2.5 x 100 = 250)





Here is another misleading graph, where the volume of the second image is about one and a half times bigger than it should really be.

Trick or Treat? In this picture graph, there is definitely a trick being done! From the vertical scale Michael collected about twice as much candy as Shayna, but his picture is about three times bigger.

Incorrect Pie Charts

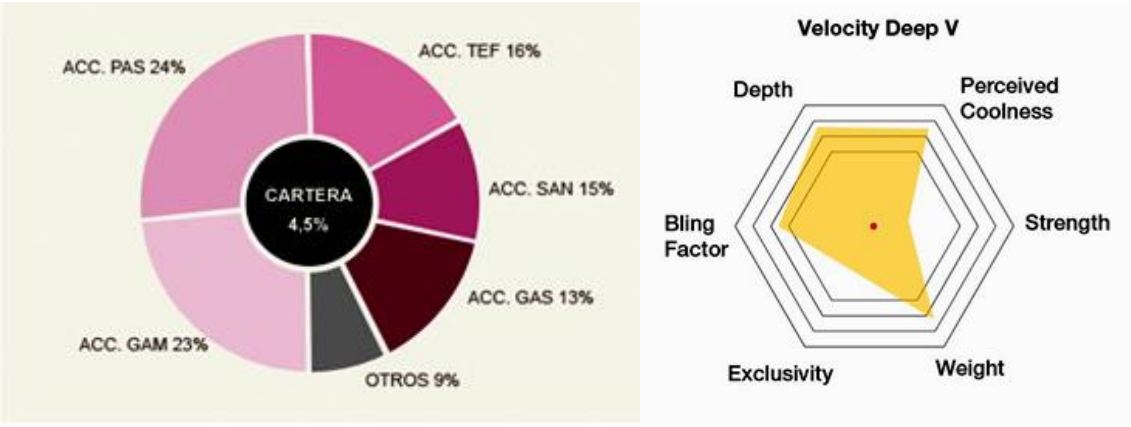
In these graphs, the pieces of pie sections are made the wrong size.

In the graph below, the right hand side pieces add up to 145 million, which is not equal to the 250 million piece on the left hand side. Also the 10% piece has been made way too big.

This pie chart is totally incorrect. The sizes of the pieces of pie have not been drawn the right sizes.

Non Standard Graphs Used

This non-standard pie chart is interesting. If “Cartera” was drawn properly, then it would only be a very thin slither, and not nearly as big looking as the big circle drawn for it here.



Here is another non-standard type of graph that is very hard to understand. In this non –standard chart, it is very hard to work out what point they are trying to make or explain.

Cheating with pie charts

Pie charts are some of the most commonly found graphical devices, although they can be difficult to read and are often misleading. (Several commentators suggest they are always misleading, and that, because they only make visual sense for visualising small data sets, it is often better just to use a numerical table.)

So what actually are they used for? Pie charts are charts that are used to represent the distribution of ‘proportions of a whole’. For example, if you conduct a survey of 100 people, you might use a pie chart to display how they answered a question of the form ‘choose only and exactly one item from the following list’, such as ‘which brand did you buy in your most recent purchase of a mobile phone?’ However, if you then went on to ask an optional, ‘yes/no’ question that only 27 of the 100 people were prepared to answer, representing the results from just those respondents in a pie chart would potentially be misleading – a reader might assume that the results applied to the whole survey population of 100. So in that case it might be better to show a chart with three sectors – one for ‘yes’, one for ‘no’, and one for ‘did not answer’.

Changing the size of the whole referred to in different charts in the same report is one way of potentially misleading the reader of a report. But it is also possible to mislead readers in their perception of a single chart. For example, in the pie charts in Figure 10, which sport has the biggest proportion? Which has the smallest?

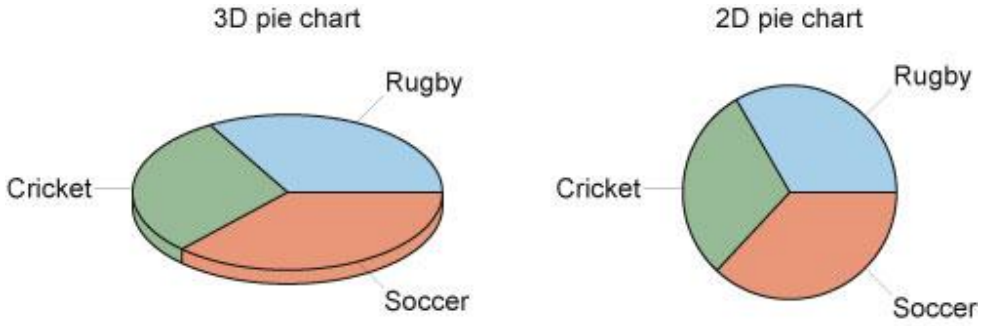


Figure 10 Two pie charts

Long description
The actual distributions are: soccer 100, rugby 90 and cricket 80 (in a situation where 270 people were asked to choose their favourite among these three sports). In this case, the 3D chart does manage to suggest this, although the differences are harder to spot than in the 2D chart. However, it is also possible to orientate the 3D chart so as to make one sector appear larger or smaller than another, similarly sized one. And colour can also have an effect on how we perceive the relative sizes. A full consideration of the perceptual effects that can be exploited to highlight particular results (or even to attempt to mislead a reader) when designing a chart will not be given here.
If you want to be able to read actual data values, a table may be more appropriate than a visual representation. Several people suggest that pie charts are always misleading, and that, because they only make visual sense for visualising small data sets, it is often better just to use a numerical table.

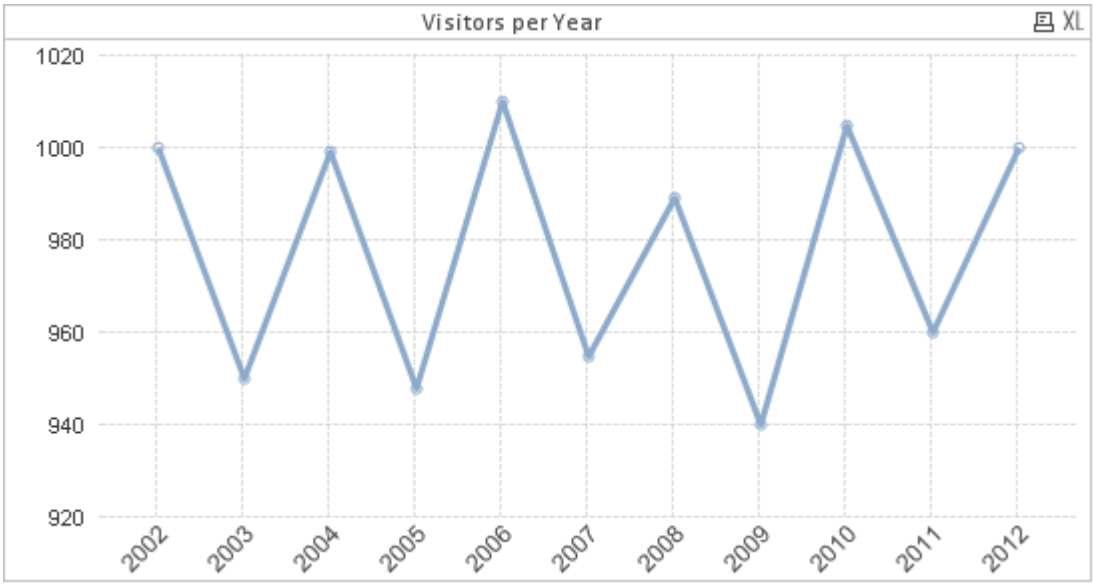
So what actually are they used for? Pie charts are charts that are used to represent the distribution of ‘proportions of a whole’. For example, if you conduct a survey of 100 people, you might use a pie chart to display how they answered a question of the form ‘choose only and exactly one item from the following list’, such as ‘which brand did you buy in your most recent purchase of a mobile phone?’ However, if you then went on to ask an optional, ‘yes/no’ question that only 27 of the 100 people were prepared to answer, representing the results from just those respondents in a pie chart would potentially be misleading – a reader might assume that the results applied to the whole survey population of 100. So in that case it might be better to show a chart with three sectors – one for ‘yes’, one for ‘no’, and one for ‘did not answer’.

Charts; they’re like photographs: they never lie, they just show the data as it is right? – Wrong. In the wrong hands they can be some of the most cunning and conniving creations ever conceived by man, they’re regularly put to work to lead the gullible astray and hide the all-important information held within. What I’m going to show; the same data in different ways to convey an entirely different meaning.

Flat or Erratic? – 'I'll take erratic please'

In a very similar way to the example above we can also use our chart detail choices to make data seem flat or erratic simply be altering the axis:

Flat Steady Visitor Numbers



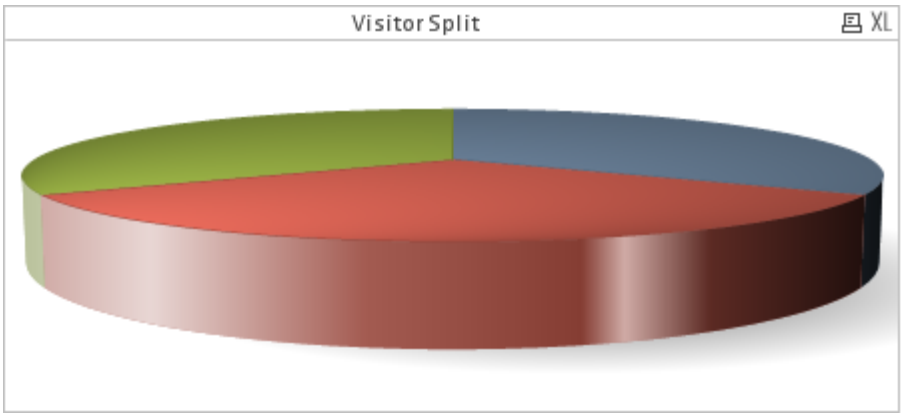
Seemingly Eratic Visitors

As before all we're doing is switching between having 'Forced 0' applied (Top) and having it not applied (Bottom), again the choice completely alters the story that this chart tells and again the data isn't changing just the viewer's perception of it.

These first 2 examples are classic uses of chart mal-manipulation employed by media and Government to make things seem better or worse as the situation, story or agenda dictates.

The Growing Pie Slice

It sounds like a cake lover's dream but in this instance it's a dashboard nightmare and it's our old friend the 3D Pie Chart at play again. Here we have 2 Pie Charts, the only difference; one is 3D and one is 2D; every segment is of equal size or at least it's supposed to be because in the 3D version the Red slice looks larger (and is by pixel count) than the others; which is clearly wrong.

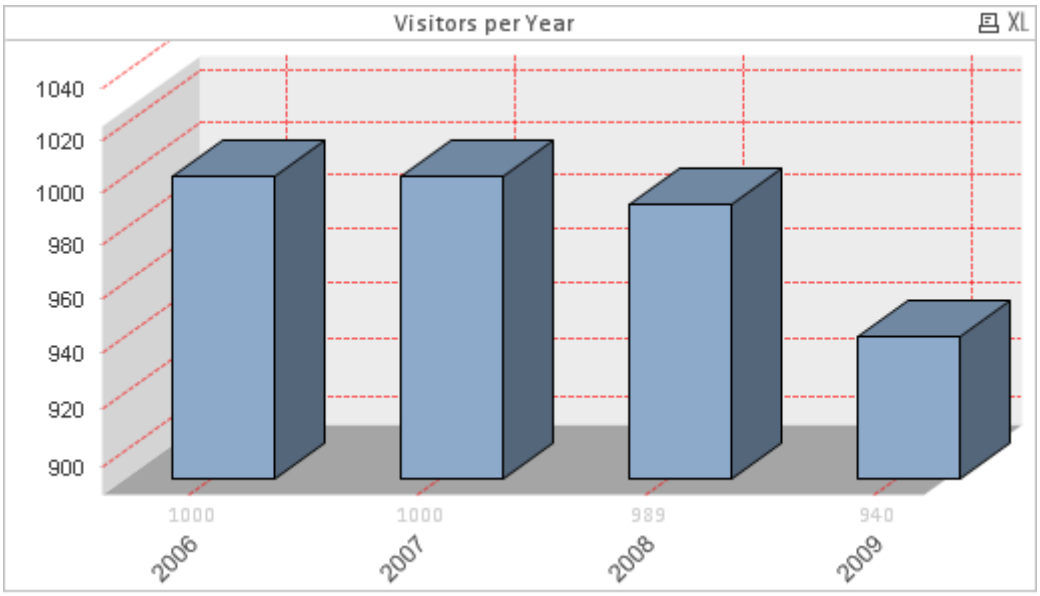


All Kinds of Wrong

In my mind there's no reason to ever use 3D Pie Charts and certainly 'The Users like them' is not a valid reason.

I Can Read You Like a Book…a Book in Another Language

Sticking with a 3D theme; it's not just Pie Charts that cause confusion the same can be said of 3D Bar Charts. Take a look at the example below, without looking at the 'Text on Axis' actual values what would you say the number of Visitors was in 2007?

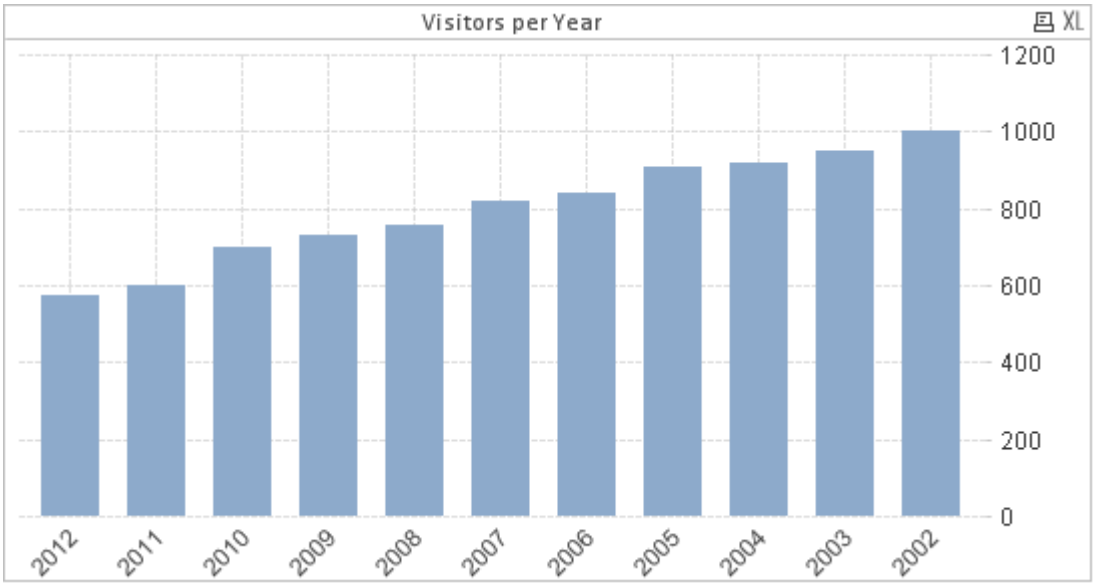


My initial reading would be about a 1,005 if I read from the top-front edge of the bar and about 992 if I use the top-back edge of the bar; so which is right; neither. What we are in fact meant to do is follow the top-side edges back until they hit the rear 'wall' of the chart and then read off from there!? Due to the way the

chart is rendered it appears that the 2007 bar never breaks through the '1,000' Grid Line so at first glance the bar appears lower than it actually is.

Visitors are down…We Can't Have That!

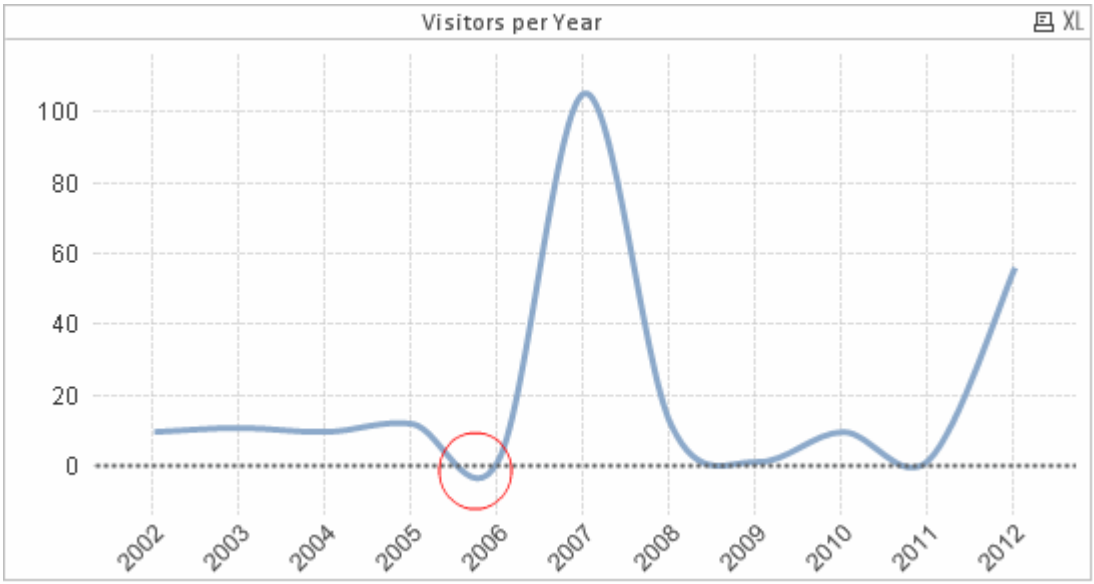
This is a blatant 'Don't do this under any circumstances' but I've seen it several times both intentionally used in the media and (probably) unintentionally used in apps, it plays on our view that a time series progress' from left to right. Look at the chart below; after a first quick glance you'd be forgiven for thinking that Visitor numbers are going up and up…



…you'd be wrong; look at the Dimension axis; it's been reversed, Visitor numbers are in fact going down year after year. This is achieved simply by reversing the sort order on the Dimension axis; easily done on purpose or by accident, whatever the reason we've completely changed the initial impression the chart gives.

Negative Visitor Numbers, Are You Sure?

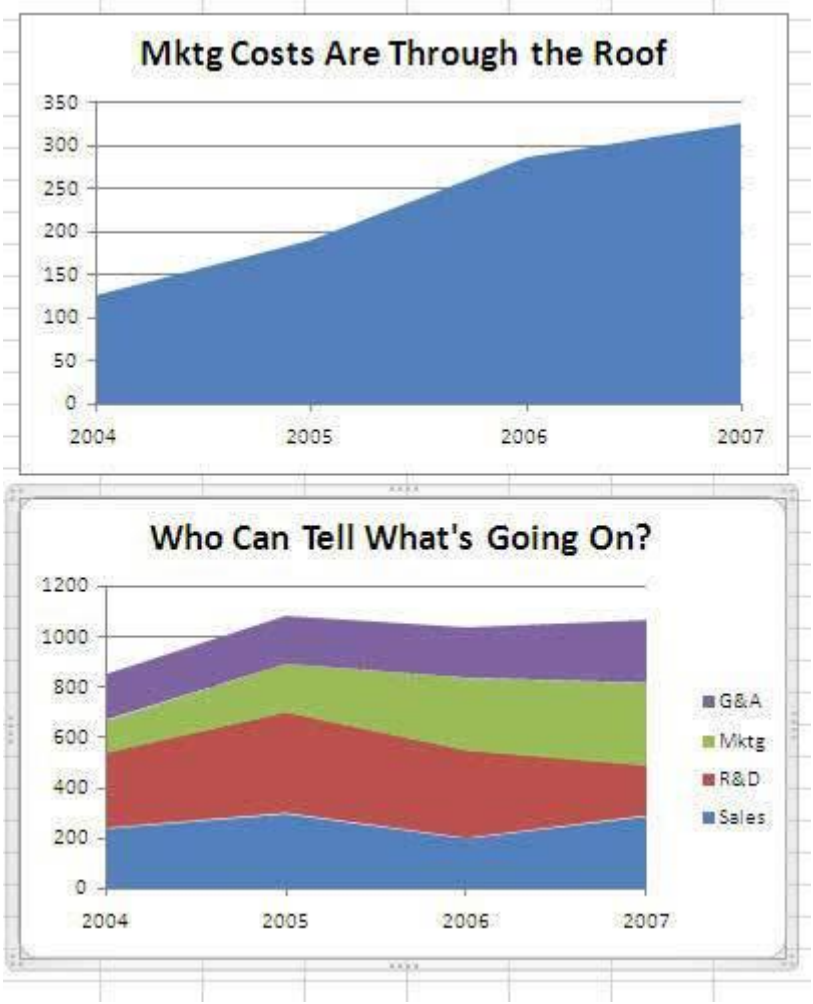
This is quite common; a Line Chart (or Combo Chart) showing negative values when negative values are in fact impossible, or where they've never occurred.



From this chart it appears as though there were a negative number of visitors around 2006 which was never the case; the chart is telling lies. Due to the way the Smooth Line is drawn it can 'peak or trough' over or below the values around it creating at best a misleading impression and at worst in cases like the one above an illusion that the impossible has happened.

How Can I Hide my Departments Rising Costs?

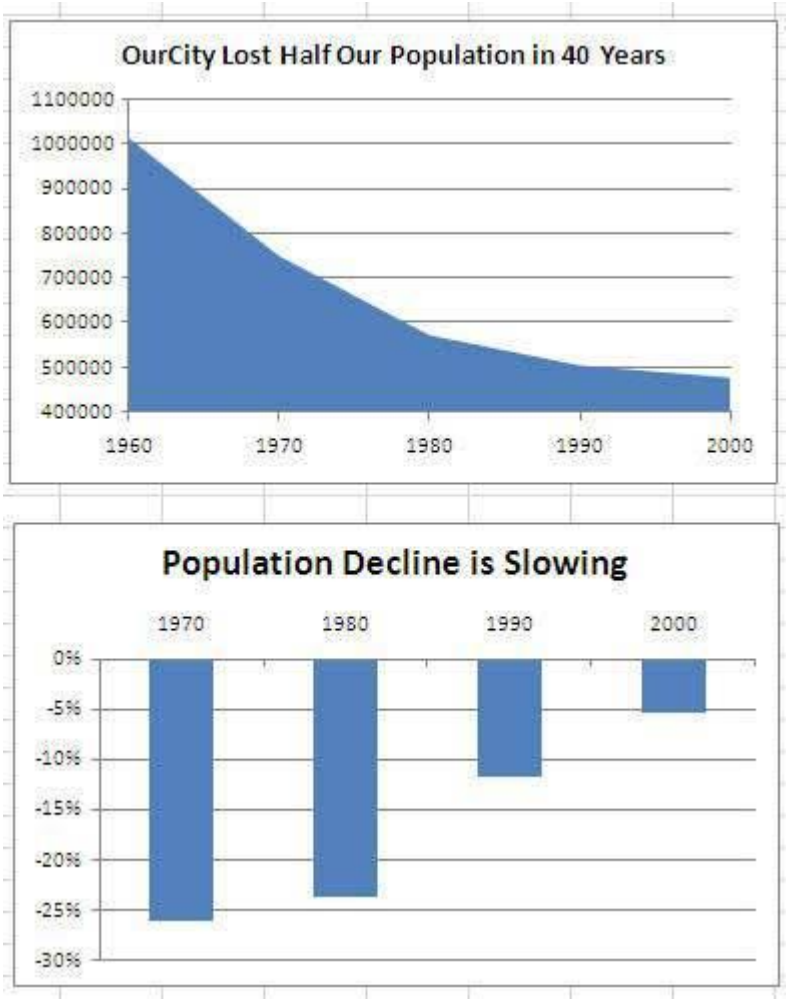
The Marketing costs are clearly going up and up; who would want to show that to their boss? So to get round this 'problem' all that has to be done is to bury the Marketing costs in a stack of other more positive data and there we have it; at first glance it even looks as though Marketing costs are falling!



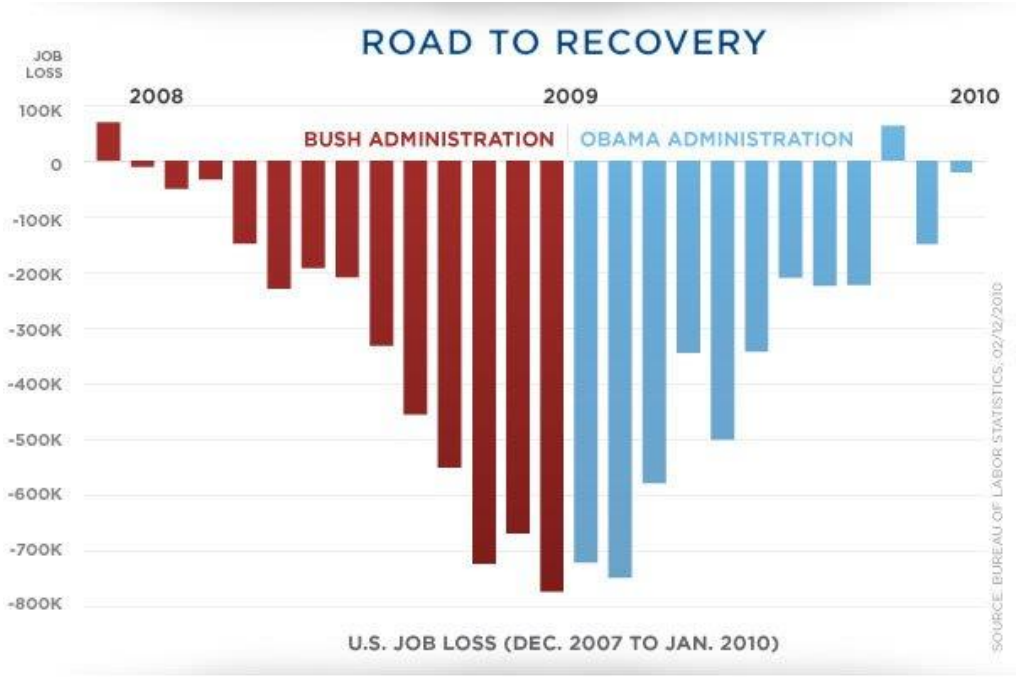
This is all thanks to the R&D costs falling at a greater rate than the increase in the Marketing Department. Yet again; we aren't changing the data; take the time to read the values and you'll get exactly the same information out of both charts but on initial viewing one is bad and one is good.

Like a Good Politician; Tell a Different Story

This one is another classic employed to make a bad situation appear better. In the top chart it looks as though there's been a crash in Population over the period (this drop is accentuated by using a non-zero Y-Axis) whilst in the bottom one things look to be getting decidedly better – how can that be? Simple; the 2 charts are driven by the same data but are showing 2 completely different things; the top one shows actuals and the bottom shows percentage change. As the rate of population decrease is lessening over time if we show this in a chart (Bottom) it looks positive thus radically changing the message taken from the data.



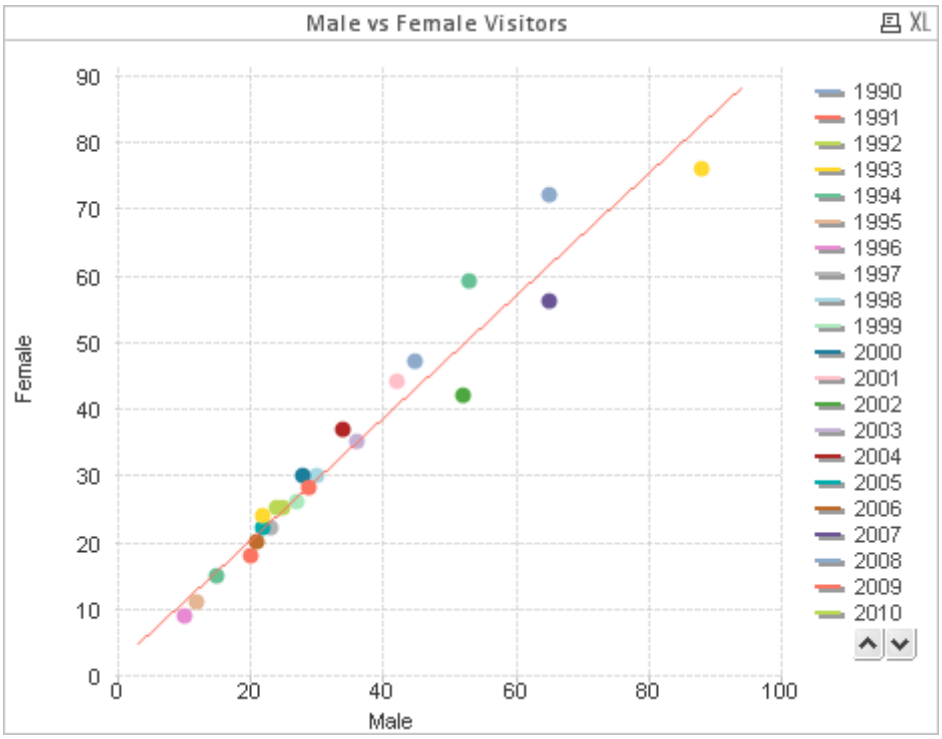
This technique is commonly trotted out for Unemployment and Inflation figures, the below is an example put out by the Obama Administration and appears to show that Employment is getting better…but it isn't; the situation may be better in a sense (unemployment is increasing at a slower rate) but month after month (bar one) unemployment continues to rise. The chart and technique are perfectly valid as long as there's no hidden agenda to deceive; show the same data on a simple line chart and it would be stoically downward; not something a politician wants to associate with, as with all these examples it's the same data but a completely different message.



A Good Chart but it Tells 1 Message and Ignores Another

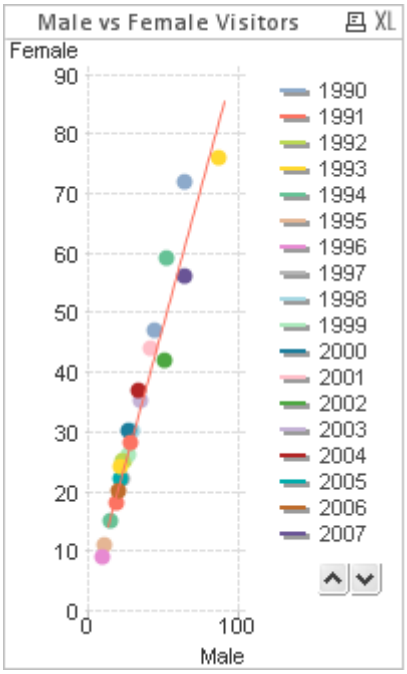
Even Scatter Charts Can Get in on the Act Too

Take a look at the first Scatter Chart showing yearly numbers of Male & Female visitors, everything looks fine, we can see for instance that once we get above around 50 men and 50 women the data points start to move away from the line of best fit; no issues here.



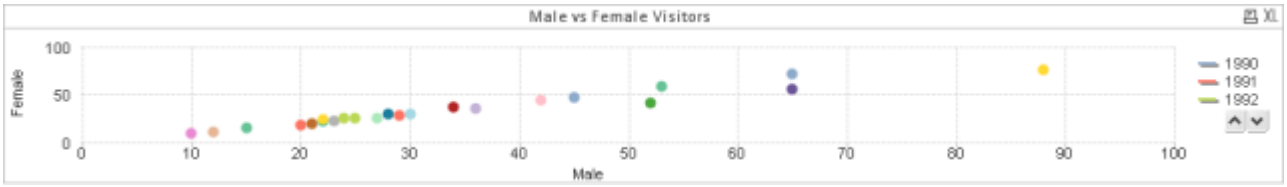
Let's make some changes to the chart (not the data) and change the impression.

Firstly we'll shorten one of the Axis – I've exaggerated this to make the effect more pronounced but I have seen Scatter Charts like this deployed in the wild:



It's now far less apparent that the values over 50/50 are moving away from the line of best fit; they seem to be much more 'normal'. This issue is documented by Stephen Kosslyn in his book 'Graph Design for the Eye and Mind':*http://www.amazon.co.uk/Graph-Design-Mind-Stephen-Kosslyn/dp/0195311841*

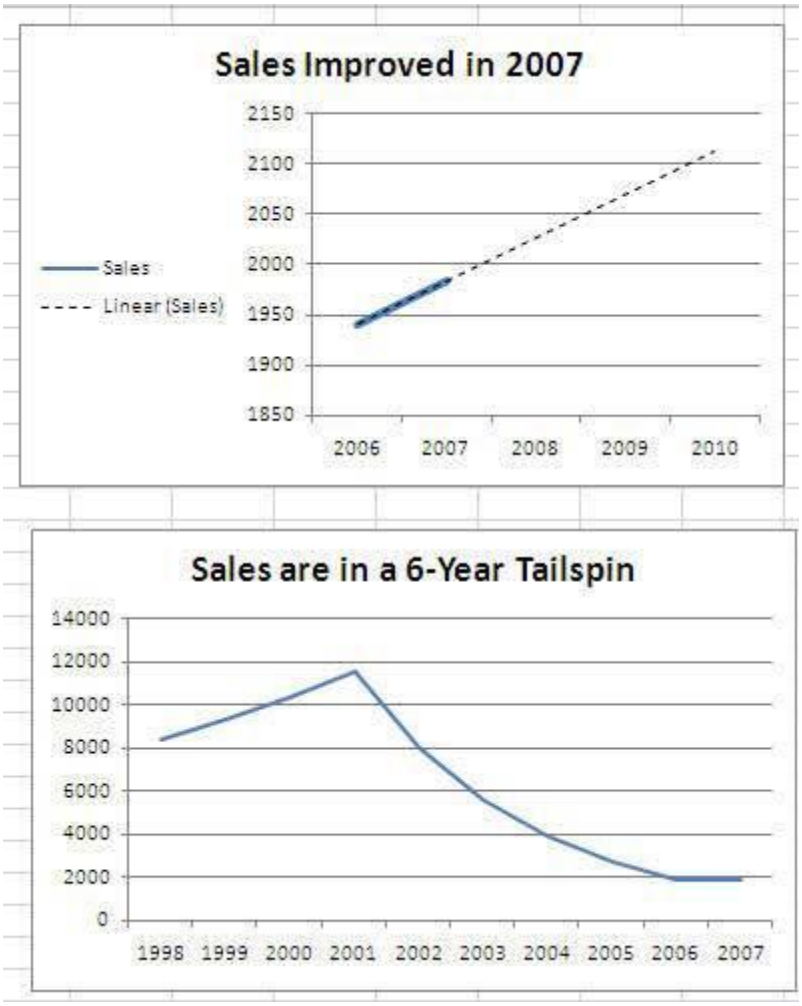
Next we'll lengthen one of the axis.



This time we've created a mis-leading impression that the values along the X-Axis are greater than those along the Y-Axis; they aren't they both run from 0-100. The vast majority of data points are very nearly equal yet looking at this chart you'd think that there were more Male visitors per year than Female.

As a rule I try to make Scatter and Grid chart Axis equal in scale – if they're both '100' of the same measure (e.g. Visitors) then they should be the same length, if one is double the other then the axis should technically be double the length to avoid creating a false impression.

Data You Don’t Like? – Just Ignore It



Again the data hasn’t changed at all but through a combination of a non-zero axis and a limited sub-set of the data the message is completely changed.

Excessive usage

The use of graphs where they are not needed can lead to unnecessary confusion /interpretation. Generally, the more explanation a graph needs, the less the graph itself is needed. Graphs do not always convey information better than tables.

Biased labelling

The use of biased or loaded words in the graph's title, axis labels, or caption may inappropriately prime the reader.
Pie chart

Main article: Pie chart

- Comparing pie charts of different sizes could be misleading as people cannot accurately read the comparative area of circles.
- The usage of thin slices which are hard to discern may be difficult to interpret.
- The usage of percentages as labels on a pie chart can be misleading when the sample size is small.

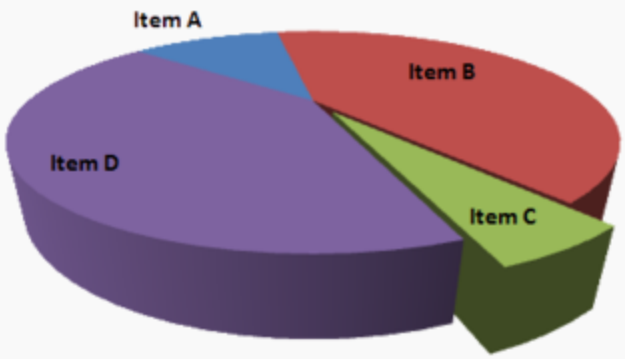
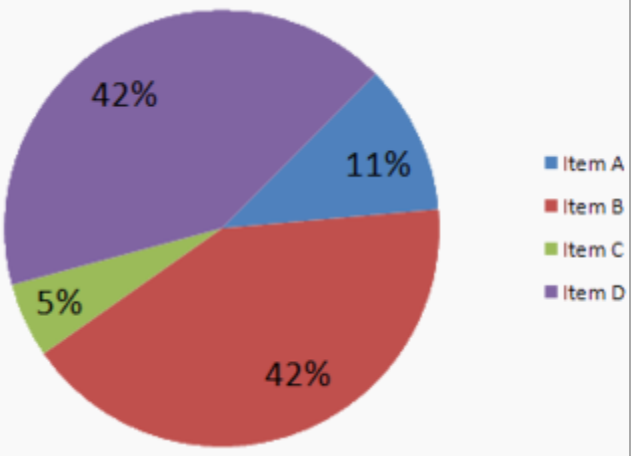
- Making a pie chart 3D or adding a slant will make interpretation difficult due distorted effect of perspective. Bar-charted pie graphs in which the height of the slices is varied may confuse the reader.

More Examples

3D Pie chart slice perspective

A *perspective* (3D) pie chart is used to give the chart a 3D look. Often used for aesthetic reasons, the third dimension does not improve the reading of the data; on the contrary, these plots are difficult to interpret because of the distorted effect of perspective associated with the third dimension. The use of superfluous dimensions not used to display the data of interest is discouraged for charts in general, not only for pie charts. In a 3D pie chart, the slices that are closer to the reader appear to be larger than those in the back due to the angle at which they're presented.

Misleading Pie Chart Comparison

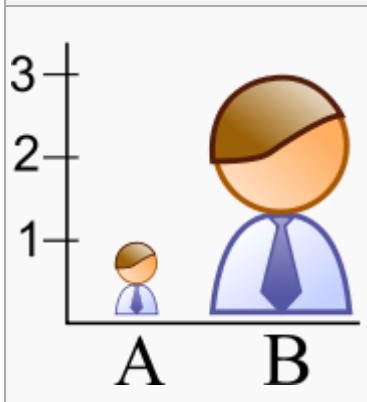
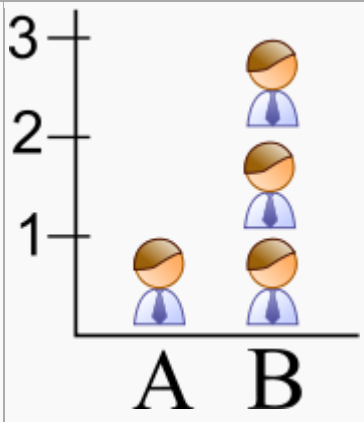
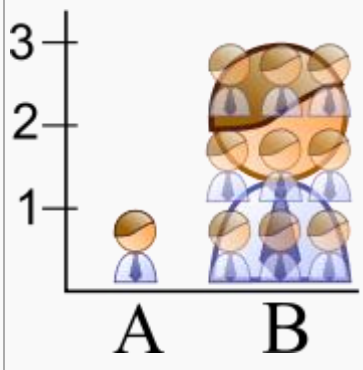
Misleading Pie Chart	**Regular Pie Chart**
	
In the misleading pie chart, Item C appears to be at least as large as Item A, whereas in actuality, it is less than half as large.	

Tables are preferable to graphics for many small data sets. A table is nearly always better than a dumb pie chart; the only thing worse than a pie chart is several of them, for then the viewer is asked to compare quantities located in spatial disarray both within and between pies - Given their low data-density and failure to order numbers along a visual dimension, pie charts should never be used.

Improper scaling

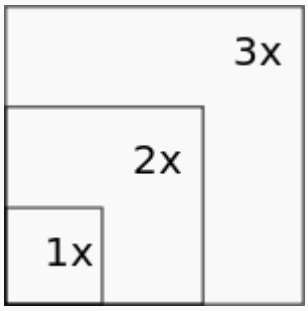
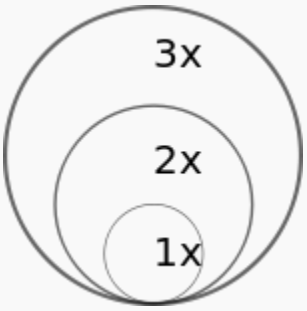
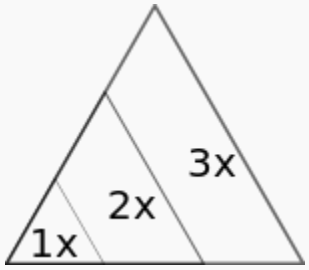
When using pictogram in bar graphs, they should not be scaled uniformly as this creates a perceptually misleading comparison. The area of the pictogram is interpreted instead of only its height or width. This causes the scaling to make the difference appear to be squared.

Improper scaling of 2D pictogram in bar graph

Improper Scaling	Regular	Comparison
		

Note how in the improperly scaled pictogram bar graph, the image for B is actually 9 times as large as A.

2D shape scaling comparison

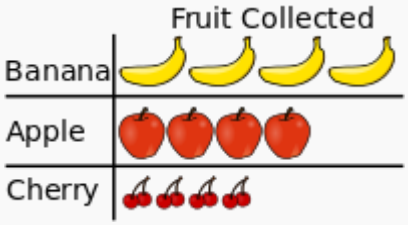
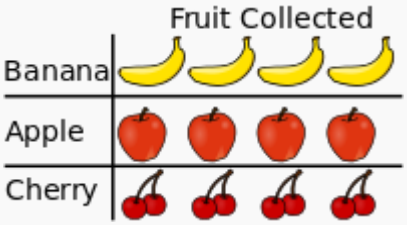
Square	Circle	Triangle
		

Note how the perceived size increases when scaling.

The effect of improper scaling of pictogram is further exemplified when the pictogram has 3 dimensions, in which case the effect is cubed.

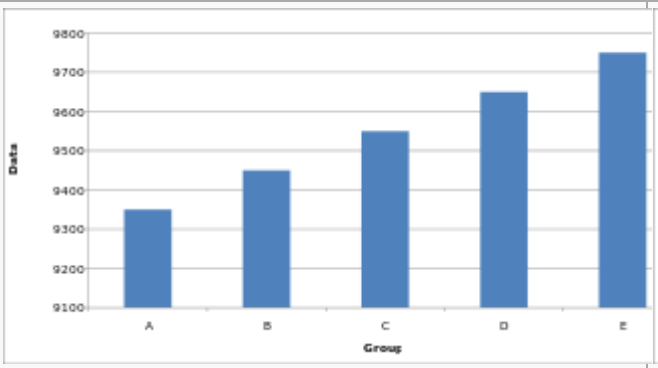
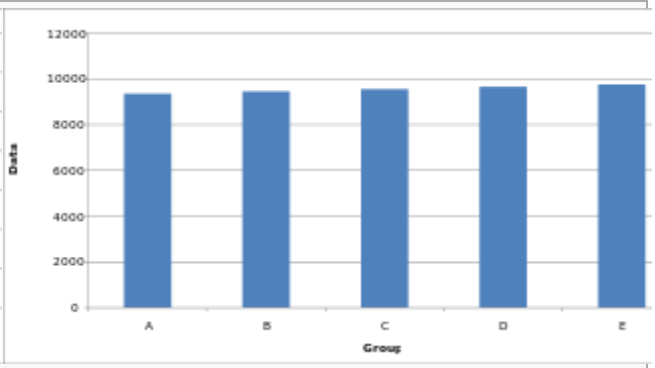
	**Note how the usage of improper scaling of a three-dimensional pictogram in this fictitious graph. It appears that home sales have gone up significantly in 2001 over the previous year. Additionally, because no frequency axis is supplied, readers are unable to quantify the change, and are only left with a misleading perception of the change. The scaling, which is 2x, causes the change to appear to be 2^3 or 8 times as large.**

Additionally, an improperly scaled pictogram may leave the reader with the sense that the item itself has actually changed in size.

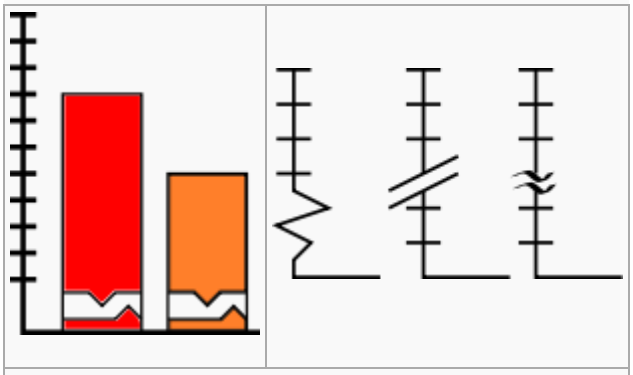
Misleading	Regular
	
Assuming the pictures represent equivalent quantities, note how in the misleading graph, there appears to be more bananas because the bananas occupy the most area and are furthest to the right.	

Truncated graph

A truncated graph (also known as a torn graph) has a y-axis that does not start at 0. These graphs can create the impression of important change where there is relatively little change.
Truncated graphs are useful in illustrating small differences. Graphs may also be truncated to save space. Commercial software such as MS Excel will tend to truncate graphs by default if the values are all within a narrow range, as in this example.

Truncated bar graph	
Truncated bar graph	**Regular bar graph**
	
Note that both of these graphs display *identical data*; however, in the truncated bar graph on the left, the data *appear* to show significant differences, whereas in the regular bar graph on the right, these differences are hardly visible.	

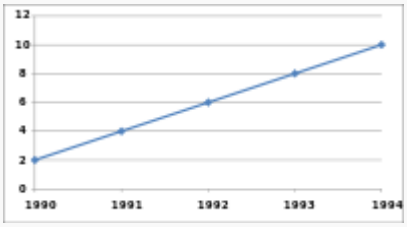
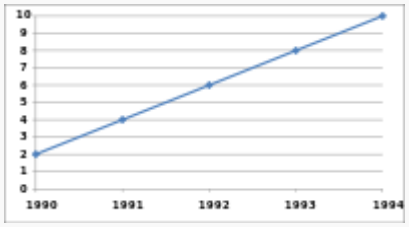
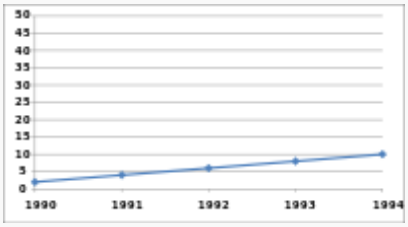
Indicating a y-axis break



There are several ways to indicate a y-axis break.

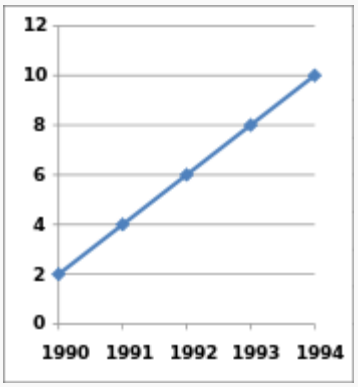
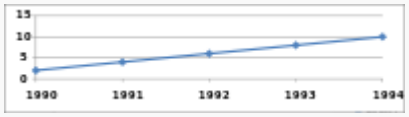
Axis changes

Changing y-axis maximum

Original graph	Smaller maximum	Larger maximum
		

Changing the y-axis maximum affects how the graph appears. A higher maximum will cause the graph to appear to have less-volatility, less-growth and a less steep line than a lower maximum.

Changing ratio of graph dimensions

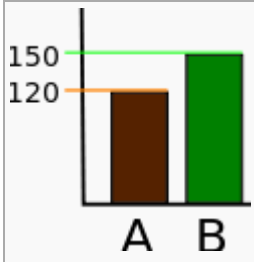
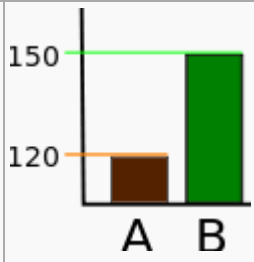
Original graph	Half width, twice height	Twice width, half height
		

Changing the ratio of a graph's dimensions will affect how the graph appears.

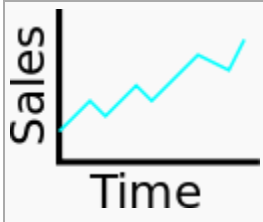
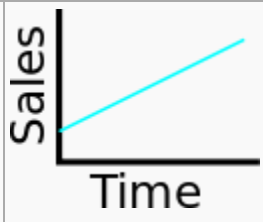
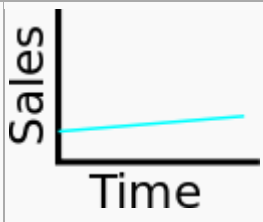
No scale

The scales of a graph are often used to exaggerate or minimize differences.

Misleading bar graph with no scale

Less difference	**More difference**
	
Note the lack of a starting value for the y-axis, which makes it unclear if the graph is truncated. Additionally, note the lack of tick marks which prevents the reader from determining if the graph bars are properly scaled. Without a scale, the visual difference between the bars can be easily manipulated.	

Misleading line graph with no scale

Volatility	**Steady, fast growth**	**Slow growth**
		
Though all three graphs share the same data, and hence the actual slope of the (x,y) data is the same, the way that the data is plotted can change the visual appearance of the angle made by the line on the graph. This is because each plot has different scale on its vertical axis. Because the scale is not shown, these graphs can be misleading.		

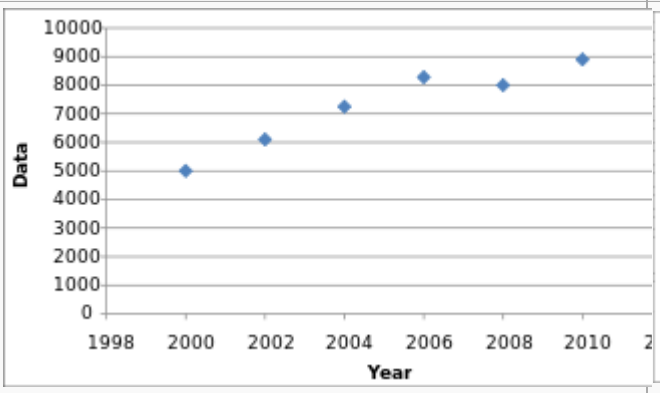
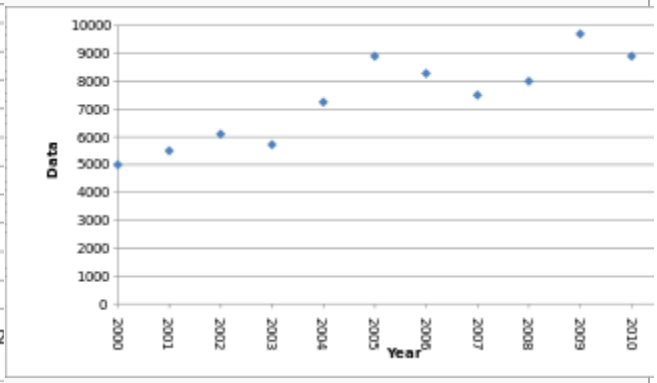
Improper intervals/units

The intervals and units used in a graph may be manipulated to create or mitigate the expression of change.

Omitting data

Graphs created with omitted data remove information from which to base a conclusion.

Scatter plot with missing categories

Scatter plot with missing categories	Regular scatter plot
	

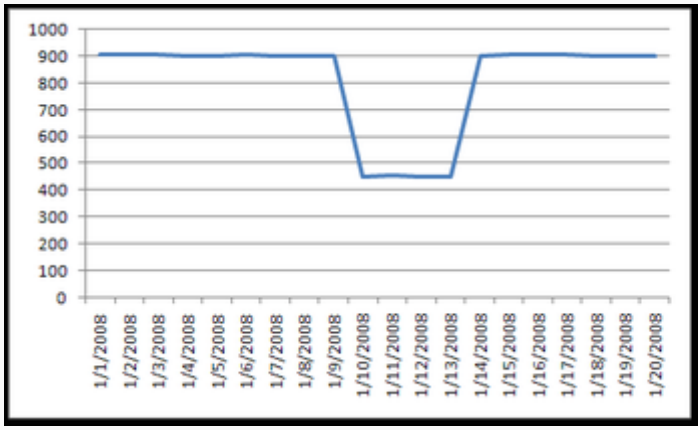
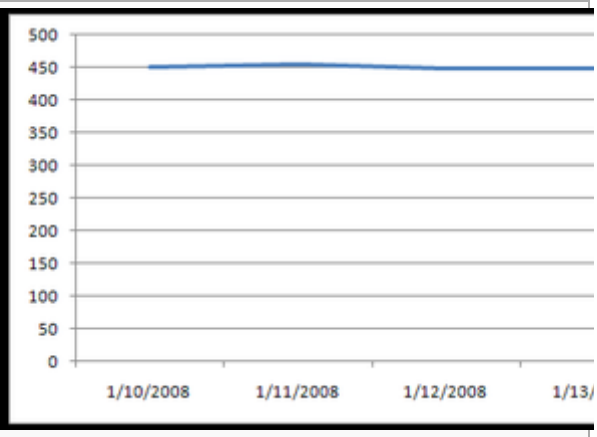
Note how in the scatter plot with missing categories on the left, the growth appears to be more linear with less variation.

In financial reports, negative returns, or data which does not correlate a positive outlook may be excluded to create a more favourable visual impression.

In engineering applications, the omission of data can be fatal. In the Space Shuttle Challenger disaster, engineers failed to properly display data.

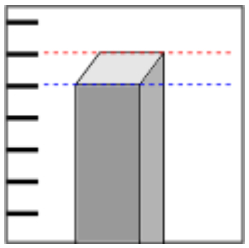
Improper extraction

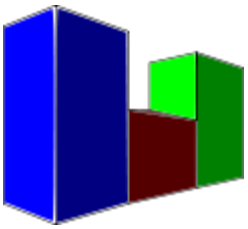
Graphs based on other graphs should be representative in their presentation.
Extraction has valid uses when searching for anomalies.

Original Graph	Extracted Graph
	

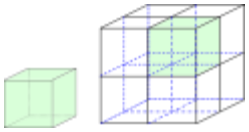
Note how the extracted graph does not accurately represent the original graph.

3D: The use of a superfluous third dimension which does not contain information is strongly discouraged as it may confuse the reader.





The third dimension may confuse readers. Note how the blue bar in the front appears larger than the green bar in the back, despite having the same value, due to perspective.



When scaling in three dimensions, the effect of the change is cubed.

Complexity
Graphs are designed to allow for easier interpretation of statistical data. However, graphs with excessive complexity can obfuscate the data and make interpretation difficult.

Poor construction
Poorly constructed graphs can make data difficult to discern and thus interpret.

Summary: Ways to Make a Graph Misleading

Vertical scale is too big or too small.
Vertical axis skips numbers, or does not start at zero.
Graph is not labelled properly.
Graph does not have a Title to explain what it is about.
Data is left out.
Scale not starting at zero.
Scale made in very small units to make graph look very big.
Scale values or labels missing from the graph.
Incorrect scale placed on the graph.
Pieces of a Pie Chart are not the correct sizes.
Oversized volumes of objects that are too big for the vertical scale differences they represent.
Size of images used in Pictographs being different for the different categories being graphed.
Graph being a non-standard size or shape.

Pie chart Problems

They represent much of what is wrong with the poor design of many websites and software applications. They're also ineffective, misleading, and inaccurate. Using a pie chart as your graph of choice to visually display important statistics and information demonstrates either a lack of knowledge, laziness, or poor design skills.

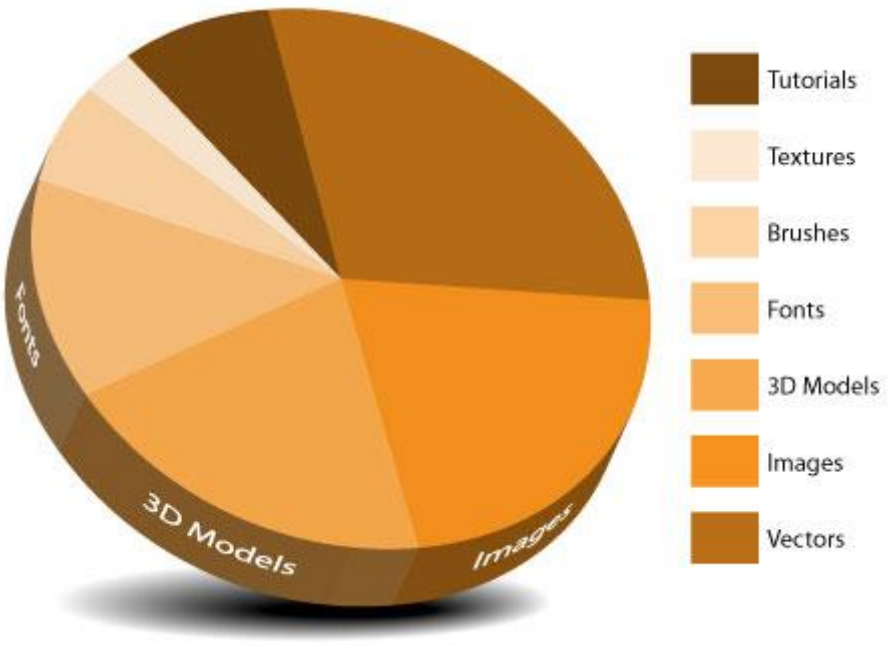


Figure 1: A floating, tilted, 3D pie chart with shadow trying (poorly) to show usage statistics within a graphics application.

Of course, pie charts in and of themselves are not evil. This blog is really about designers making poor decisions for all the wrong reasons. In order for a pie chart to appear on a web page, somebody chose it over the other alternatives, and probably thought they were doing the right thing. They weren't. Using a pie chart is almost always a bad design decision.

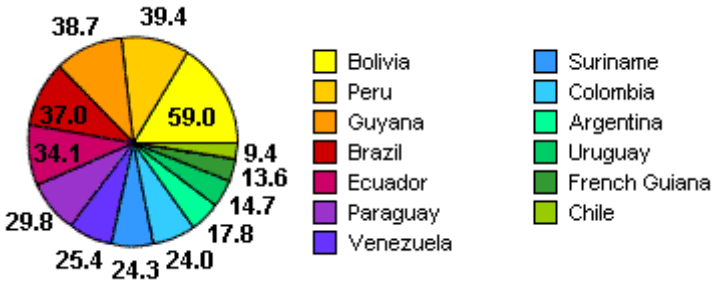


Figure 2: Pie Chart from an Oracle Reports User Guide

A pie chart does not do the job of effectively displaying information in an elegant visual form. Being circular, they use up too much space while not allowing their labels to line up. Bar charts, line charts, and tables do a much better job. Expert designers, statisticians, and business analysts have documented their many failings, and strongly urge software and report designers not to use them. It's obvious to them that the pie chart has too many inherent defects to ever be used effectively.

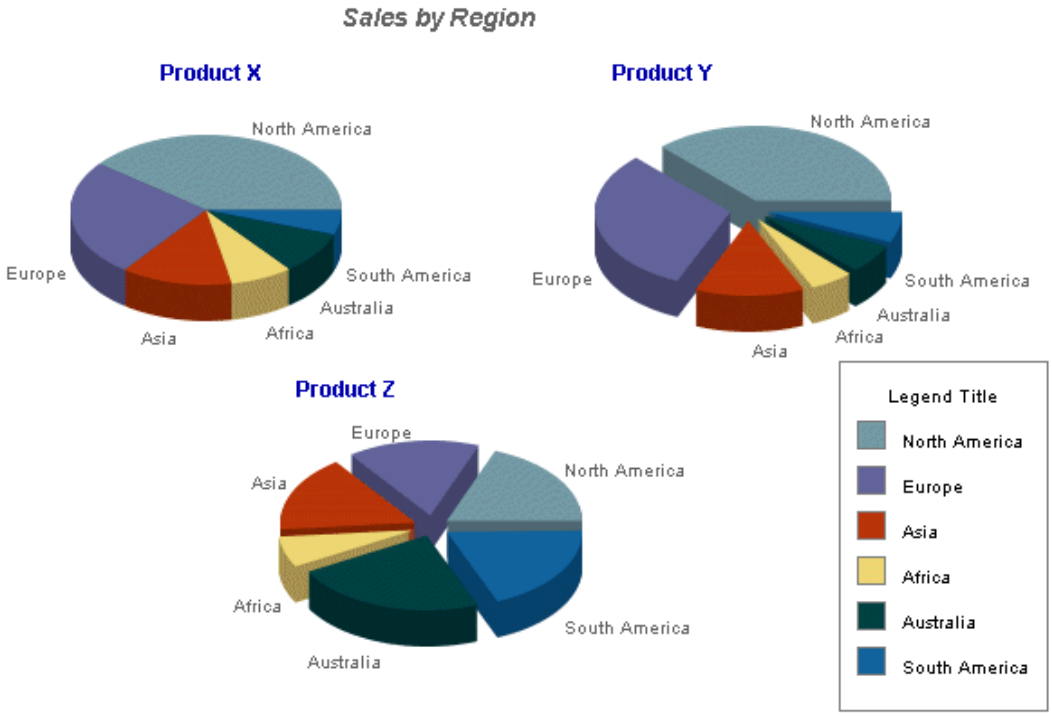
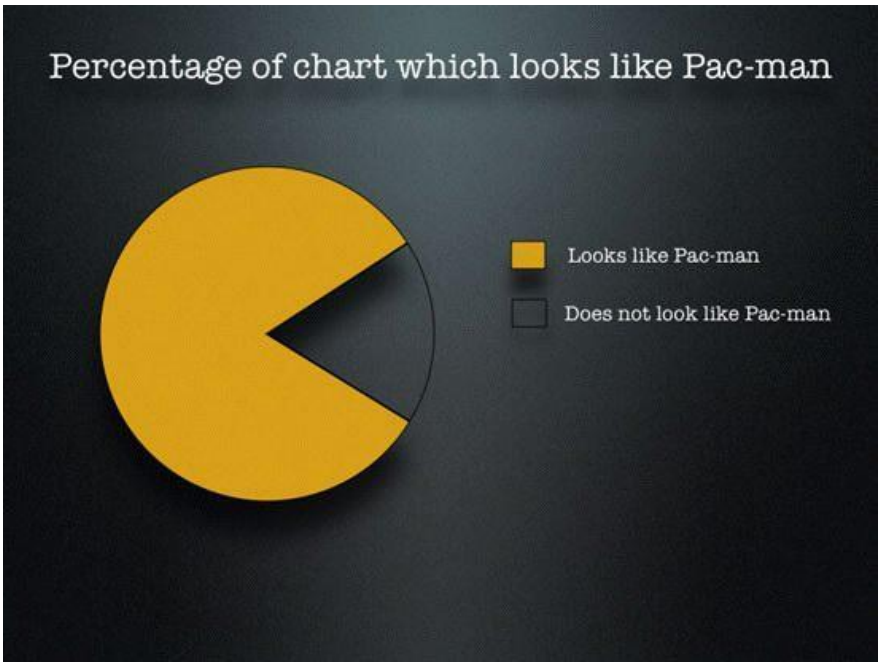


Figure 3: Demonstration of how comparing data between multiple pie charts is difficult.



Yet pie charts are still used frequently in today's software applications, financial reports, and websites, often on the opening page as a symbol of how the data inside is represented. In an attempt to get a flashy colourful graphic to break up boring text, designers will often settle for a pie chart that looks like Pac man, a coloured spinning wheel, or a 3D floating alien space ship.

10 reasons why you should never, ever user a pie chart

Number 10 - Pie Charts Just Don't Work When Comparing Data
Number 9 - You Have a Better Option: The Sorted Horizontal Bar Chart
Number 8 - The Pie Chart is Always Round
Number 7 - Some Genius Will Make It 3D
Number 6 - Legends and Labels are hard to Align and Read
Number 5 - Nobody Has Ever Made a Critical Decision Using a Pie Chart
Number 4 - It Doesn't Scale Well to More Than 2 Items
Number 3 - A Pie Chart Causes Distortions and Errors
Number 2 - Everyone Else Uses Them: Debunking the "Urban Legend" of Pie Charts
Number 1 - Pie Charts Make You Look Stupid and Lazy