Populations, Samples, and Sampling Bias:

An important area in inferential statistics is the idea of taking a sample from a population and using measurements from the sample to make predictions about the entire population.

In order for this process to be effective, the sample has to be representative of the population.

A sampling bias is a flaw in the sampling procedure that prevents the sample from being representative of the population.

In general, sampling is considered fair/unbiased if every item in the population has an equal chance of being selected. If a survey is involved, the phrasing and asking of the questions must be fair.

In the following examples, determine the population being studied, determine the sample taken, and explain any possible sampling biases.

1. A research group wants to determine the percentage of voters in Harris County in favor of a new flood control tax. A sample of 1,000 phone numbers from the Houston phonebook are selected, and the people answering the phone are asked about the matter.

2. The teaching performance of a professor is to be evaluated based on his students' opinions. The students' opinions will be measured by taking a sample of student evaluations. The professor chooses one of his five classes as the sample, he hands out the evaluation forms, and he remains in the room to answer any questions while the students fill out the forms.

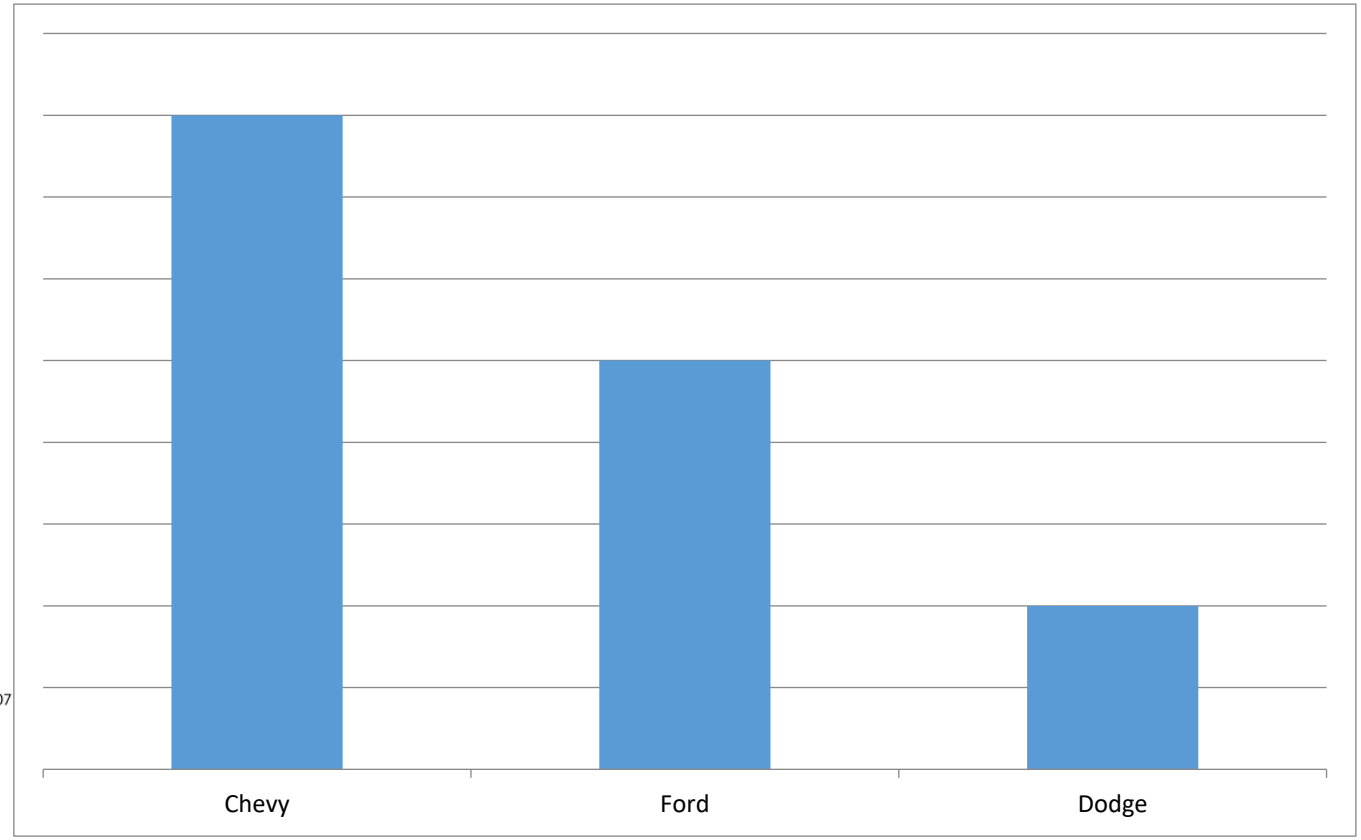
3. A biologist wants to estimate the number of fish in a lake. As part of the study, 250 fish are caught, tagged, and released back into the lake. Later, 500 fish are caught and examined. 18 of the captured fish are tagged, so the biologist reasons that the fraction of fish that are tagged in the lake is $\frac{250}{\text{fish population}}$, and this should be approximately the same as the fraction of tagged fish in the sample, $\frac{18}{500}$. $\frac{250}{\text{fish population}} \approx \frac{18}{500} \Rightarrow \text{fish population} \approx \frac{250 \times 500}{18} \approx 6{,}944$, so this is the prediction of the biologist for the number of fish in the lake.

Misleading Graphs:

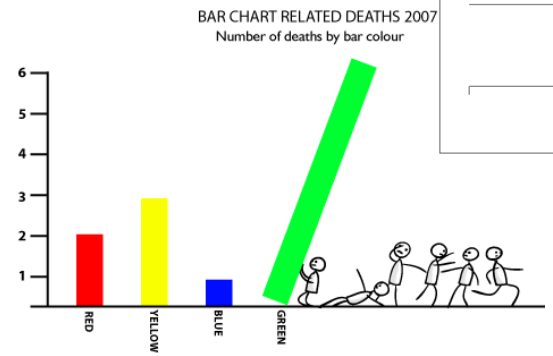
Statistical graphs can be misleading. Sometimes the deception is accidental, but sometimes it's intentional.

Misleading Bar Graph

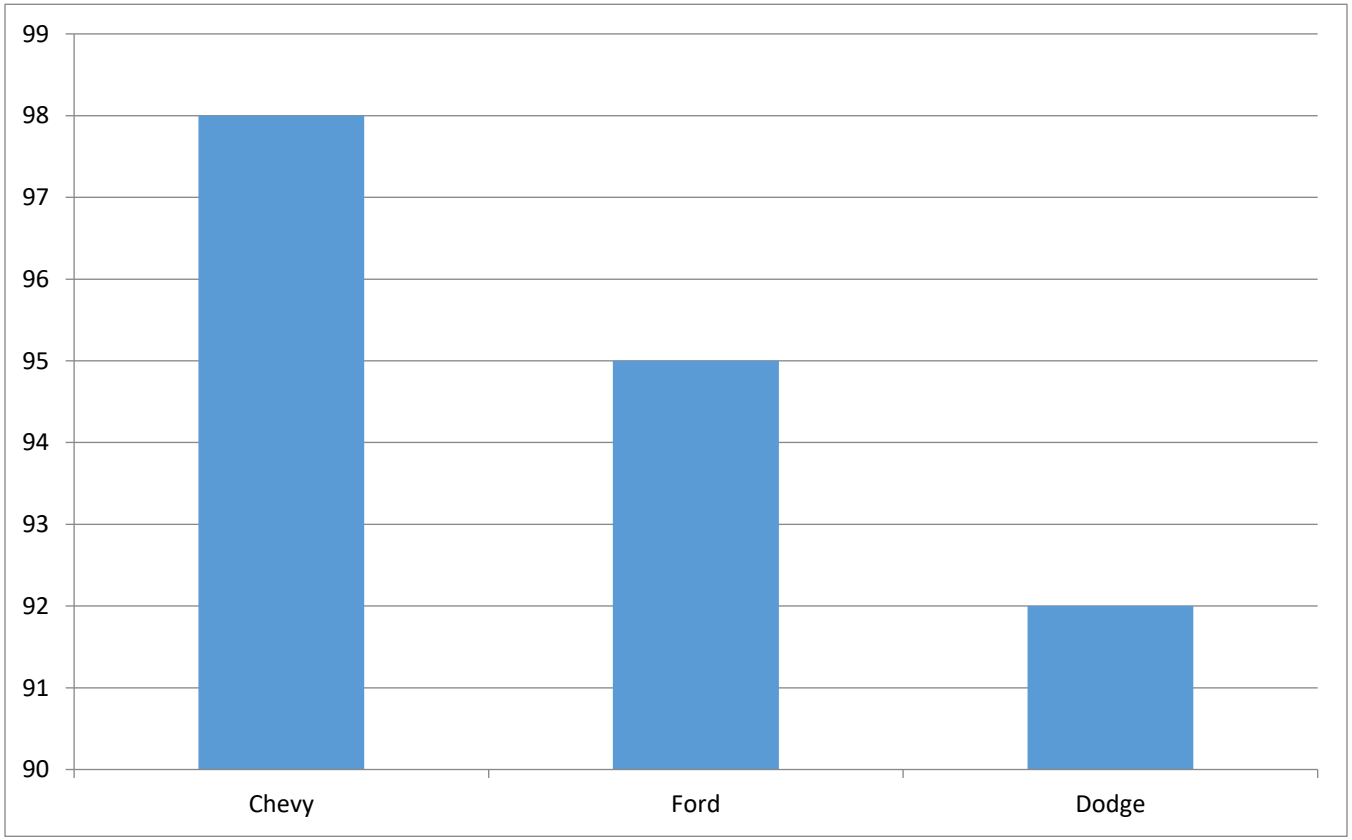
Percentage of Pickup Trucks Still on the Road After Ten Years



The vertical axis obscured.

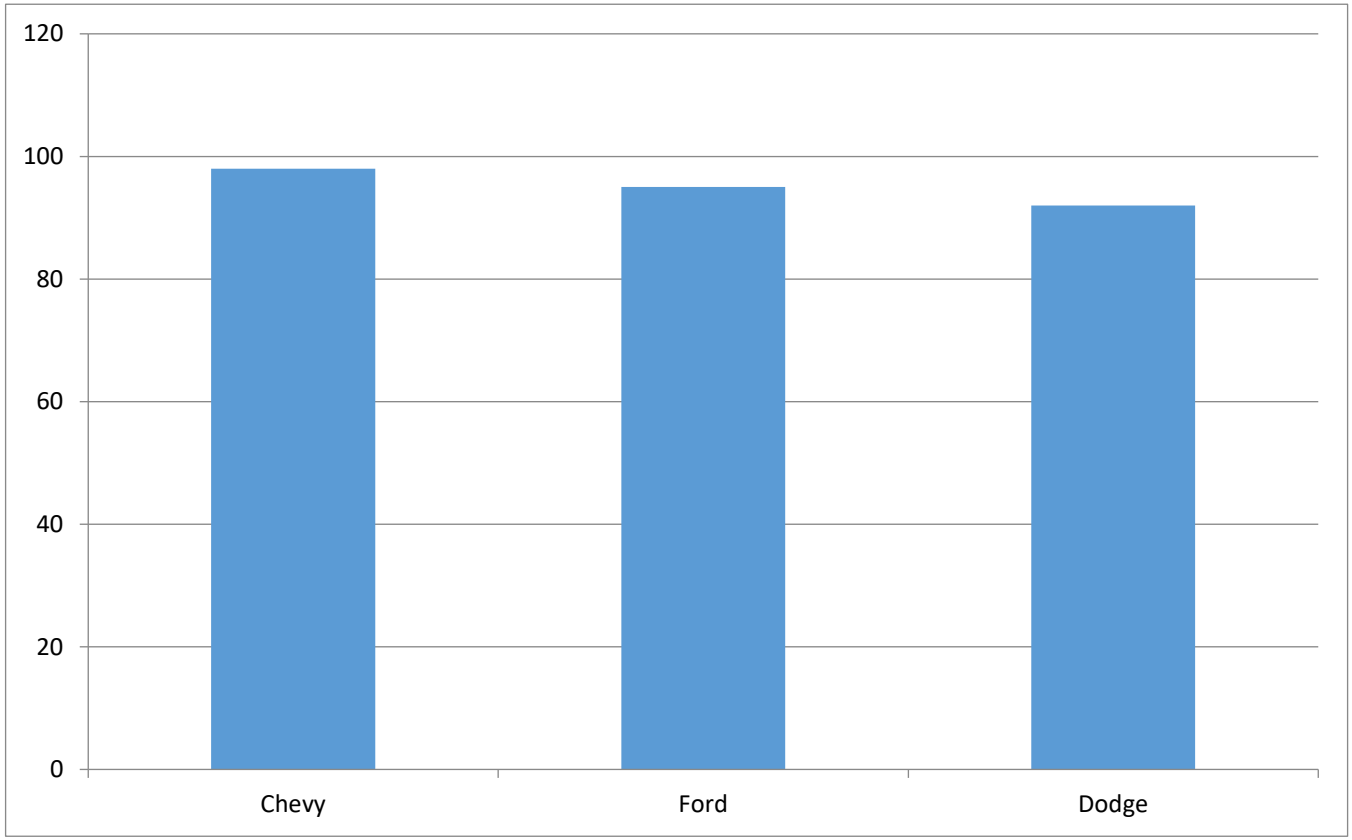


Percentage of Pickup Trucks Still on the Road After Ten Years

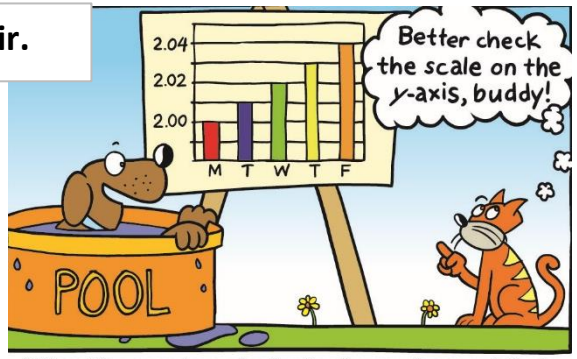


The vertical axis is visible, but it doesn't start at zero.

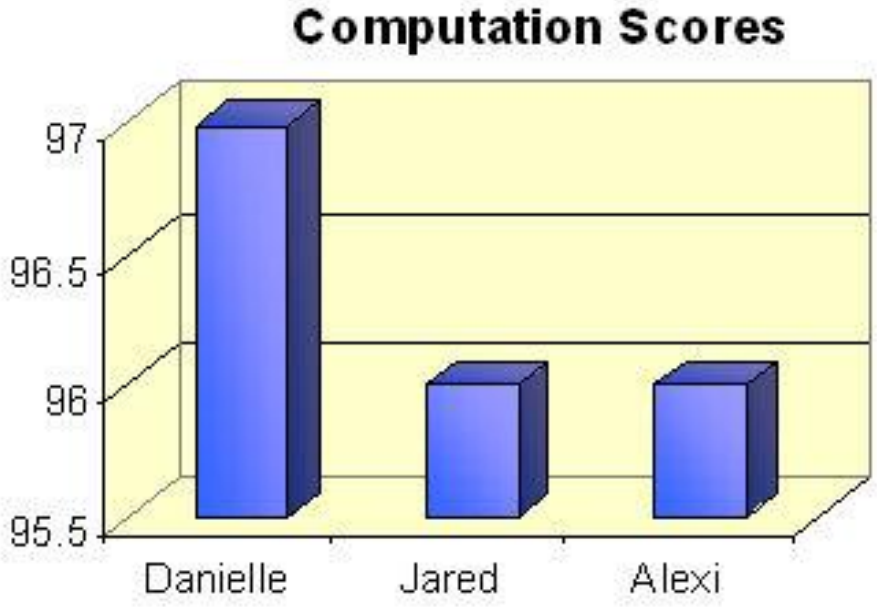
Percentage of Pickup Trucks Still on the Road After Ten Years



Starting the vertical axis at zero makes comparisons fair.

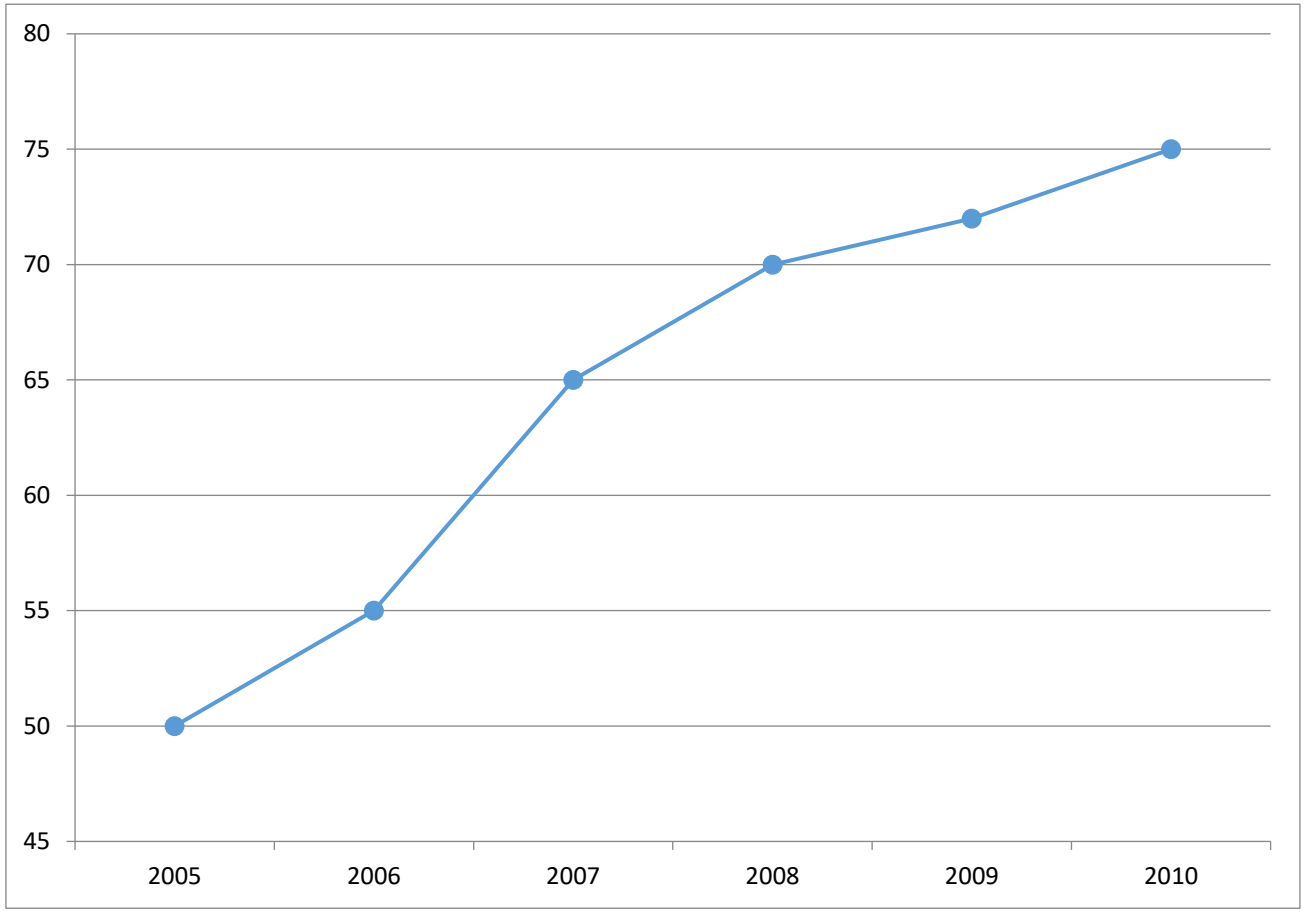


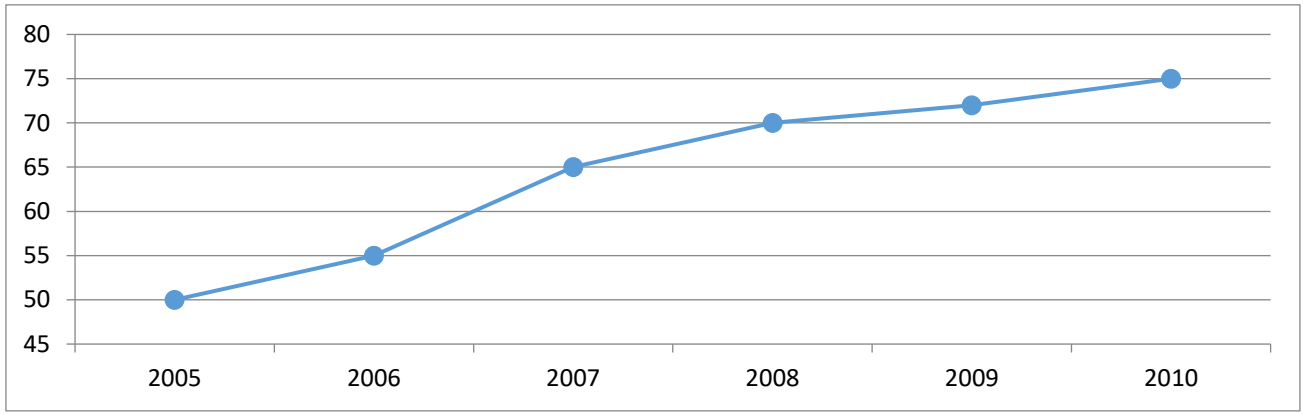
"Wow. The number of minutes I can dog paddle is growing like crazy!"


Computation Scores
97
96.5
96
95.5
Danielle
Jared
Alexi


UNDER PRESIDENT OBAMA,
MORE STUDENTS ARE EARNING THEIR HIGH SCHOOL DIPLOMAS THAN EVER BEFORE
HIGH SCHOOL GRADUATION RATE
75%
75%
78%
79%
80%
81%
82%
2007-08
2008-09
2009-10
2010-11
2011-12
2012-13
2013-14
#LeadOnEducation

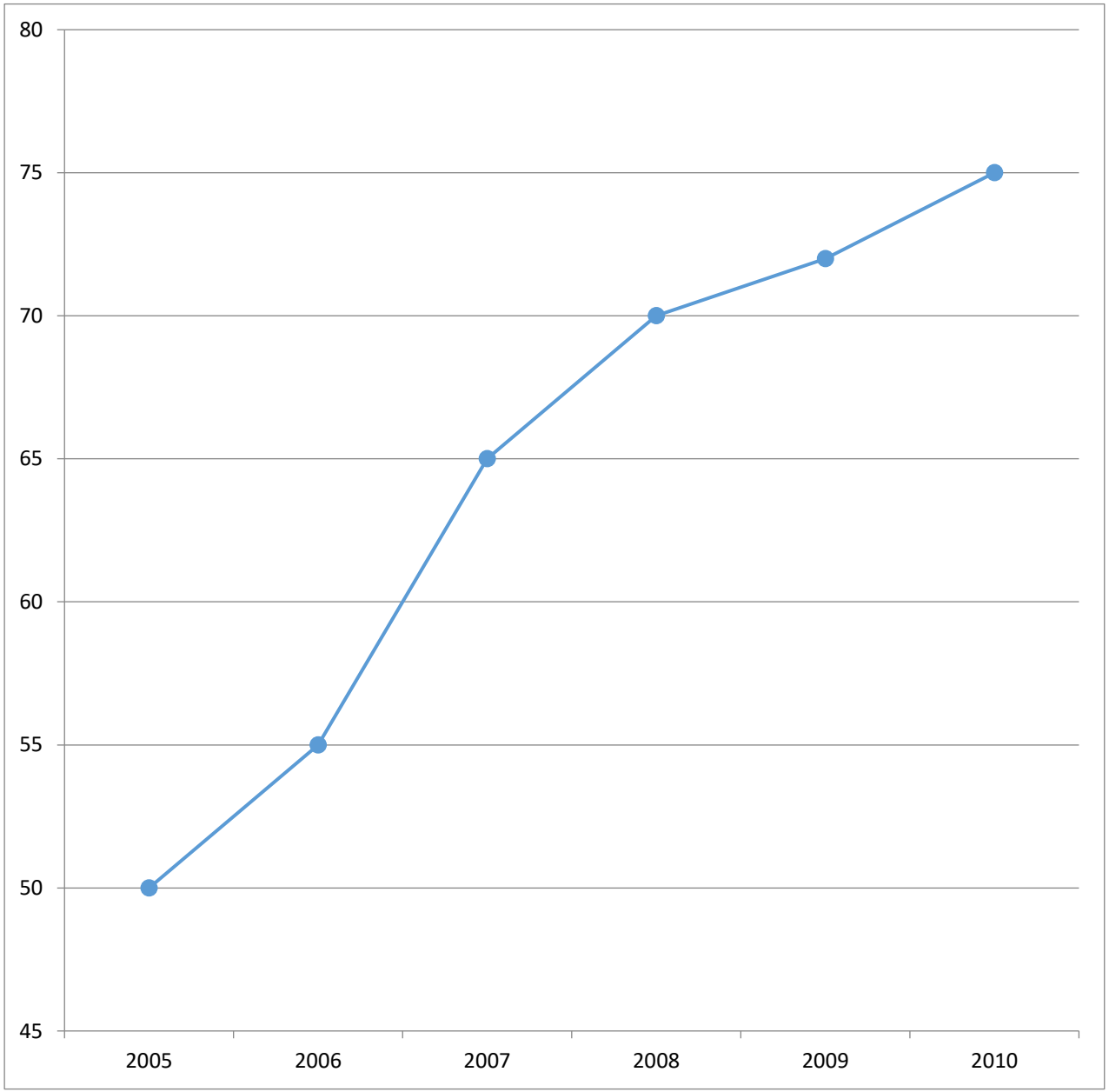
Misleading Line Graph



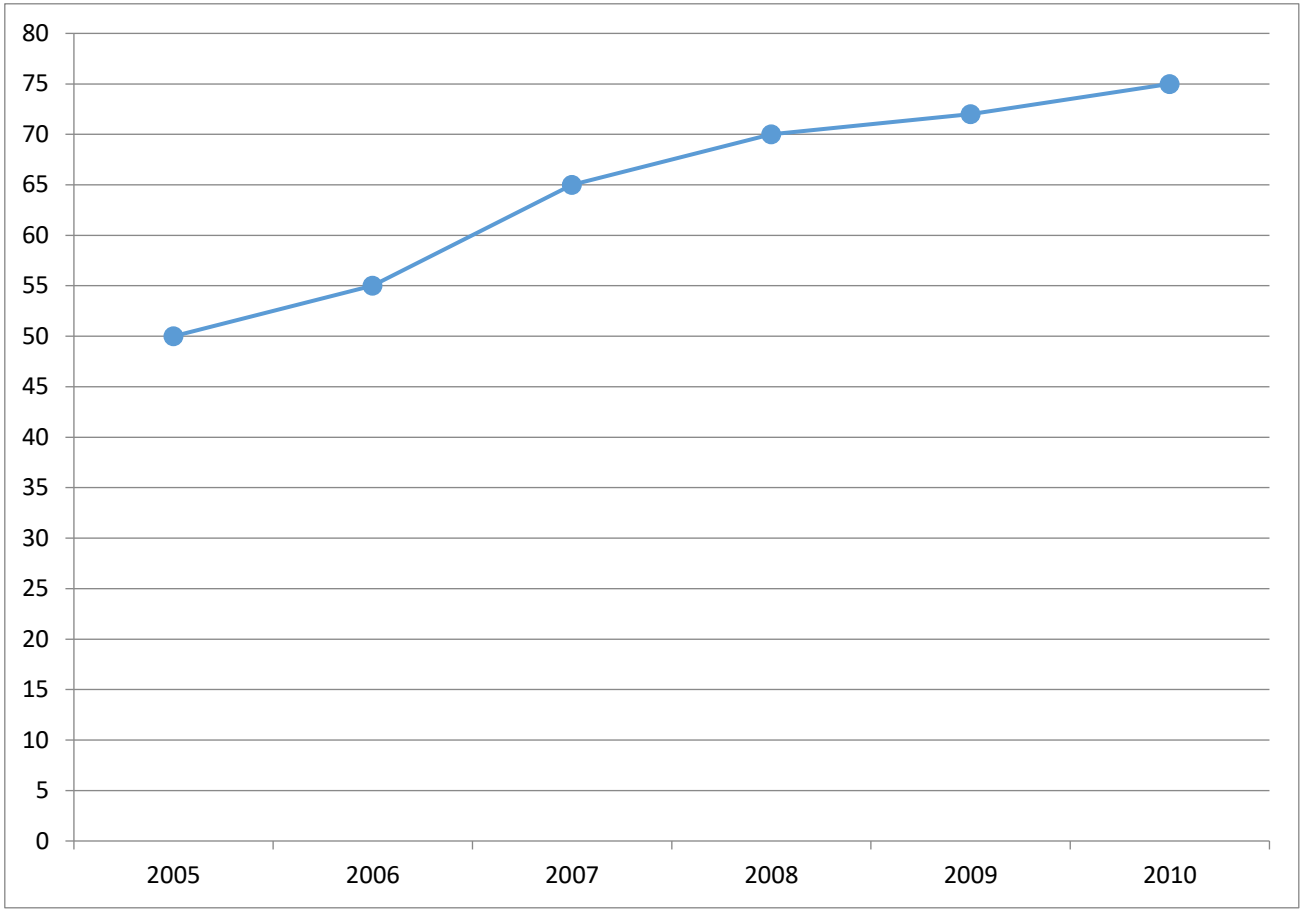


Shortening the vertical axis minimizes the upward trend.

"As you can see here, there were a few down-ticks, followed by a few upticks, finishing off with some antics."

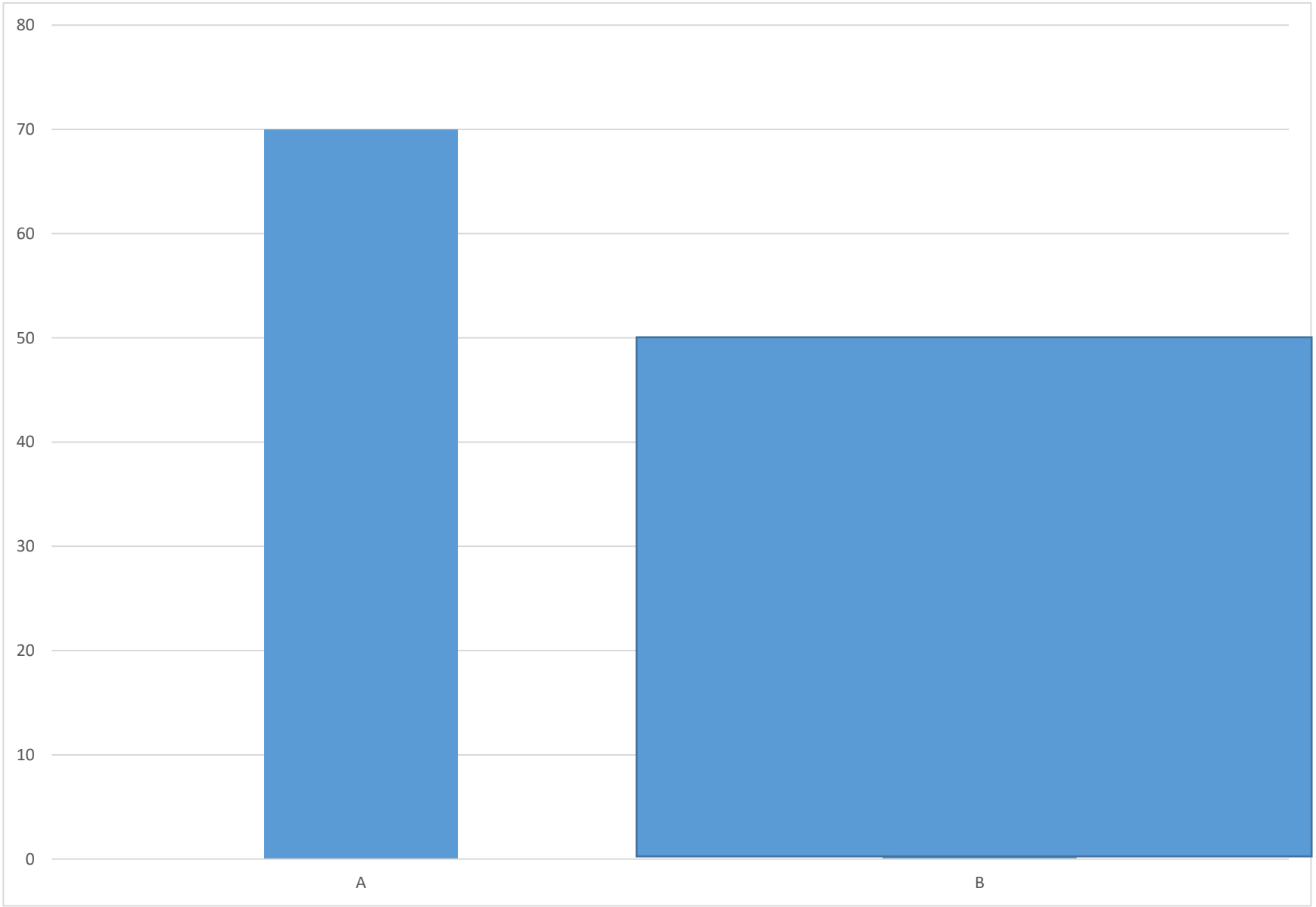


Lengthening the vertical axis enhances the upward trend.

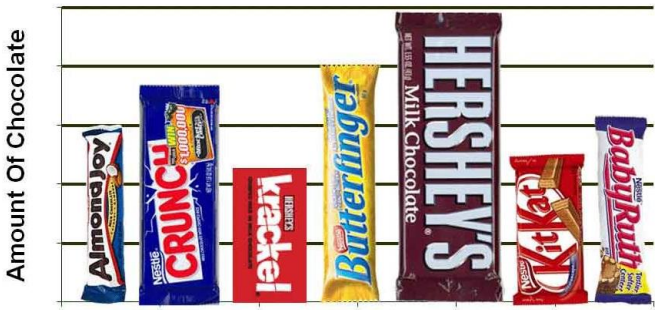


Starting the vertical axis at zero also minimizes the upward trend, and it makes comparisons fair.

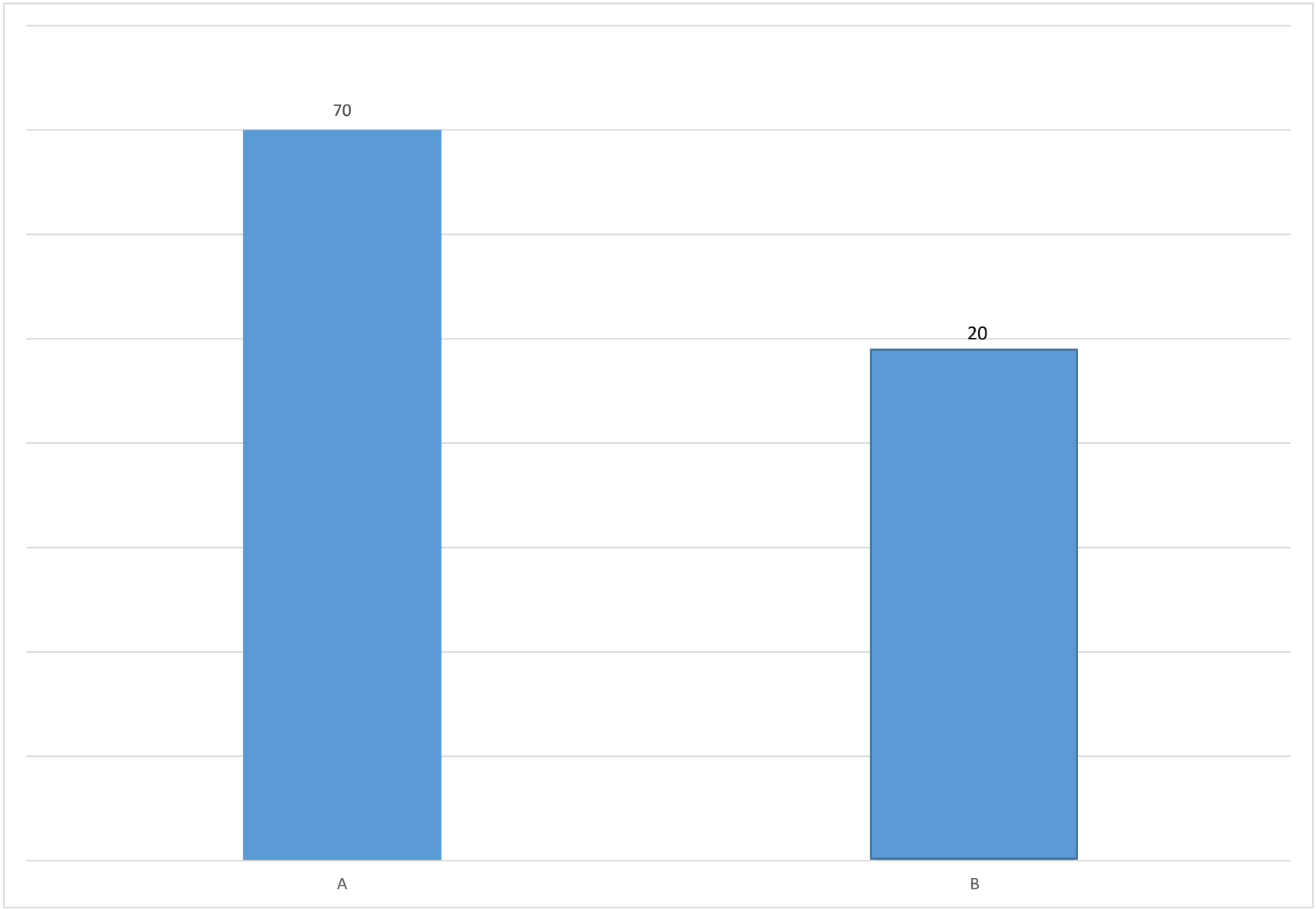
Uneven Bars



Candy Bar Chart

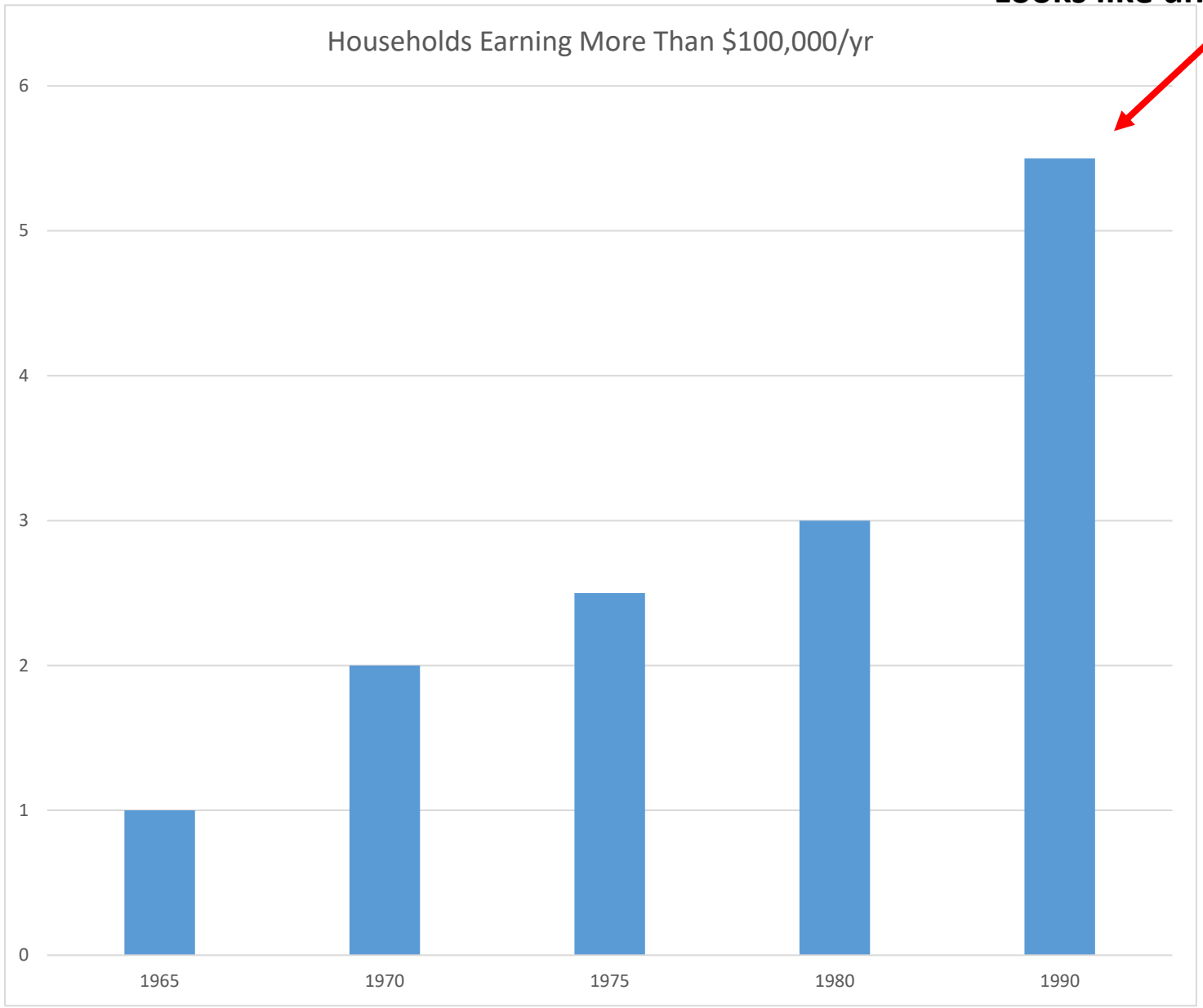


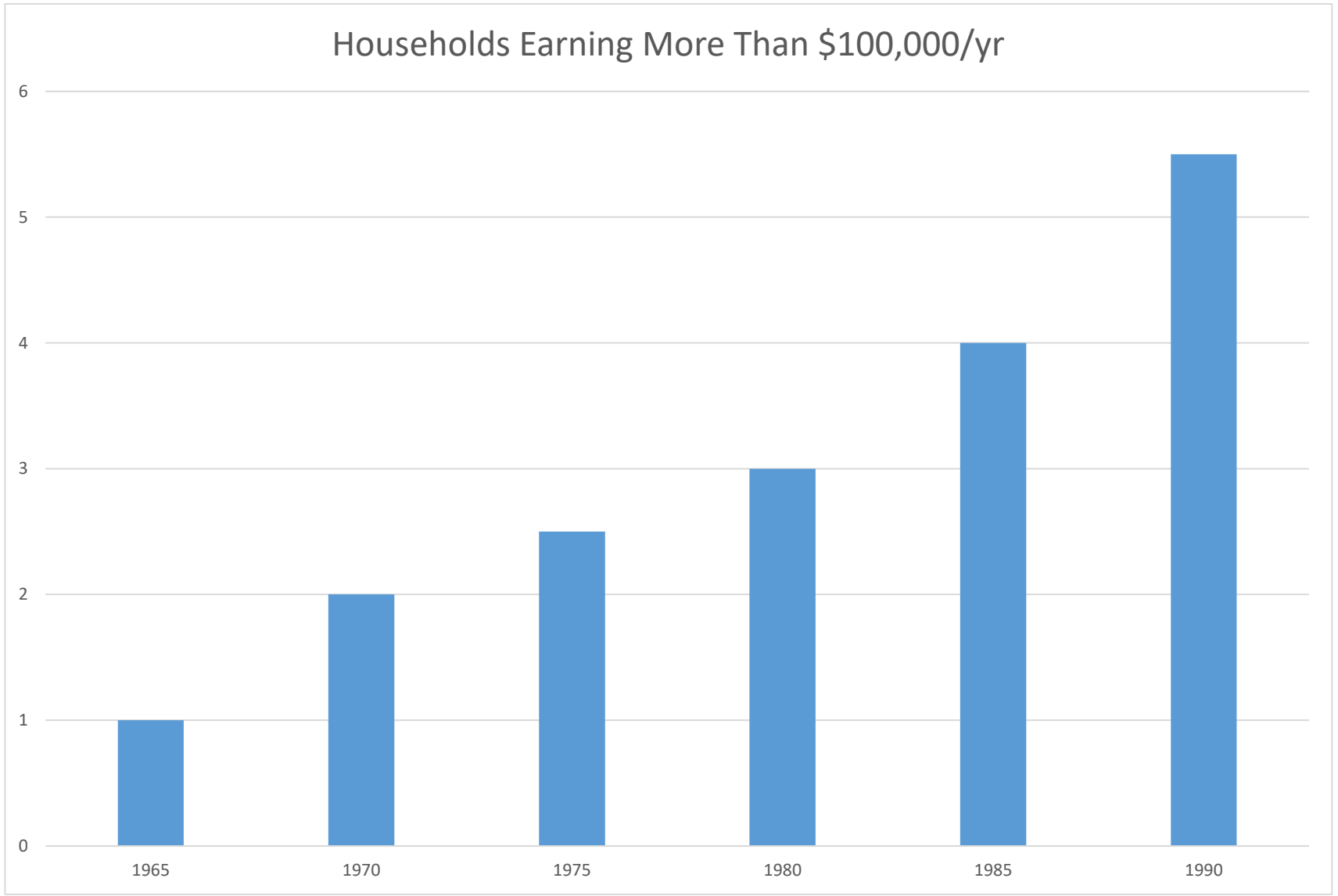
Inconsistent Vertical Scale



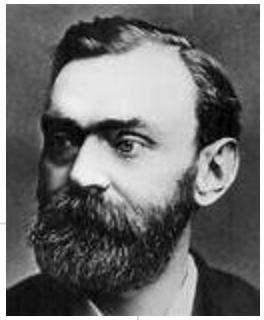
Inconsistent Horizontal Scale

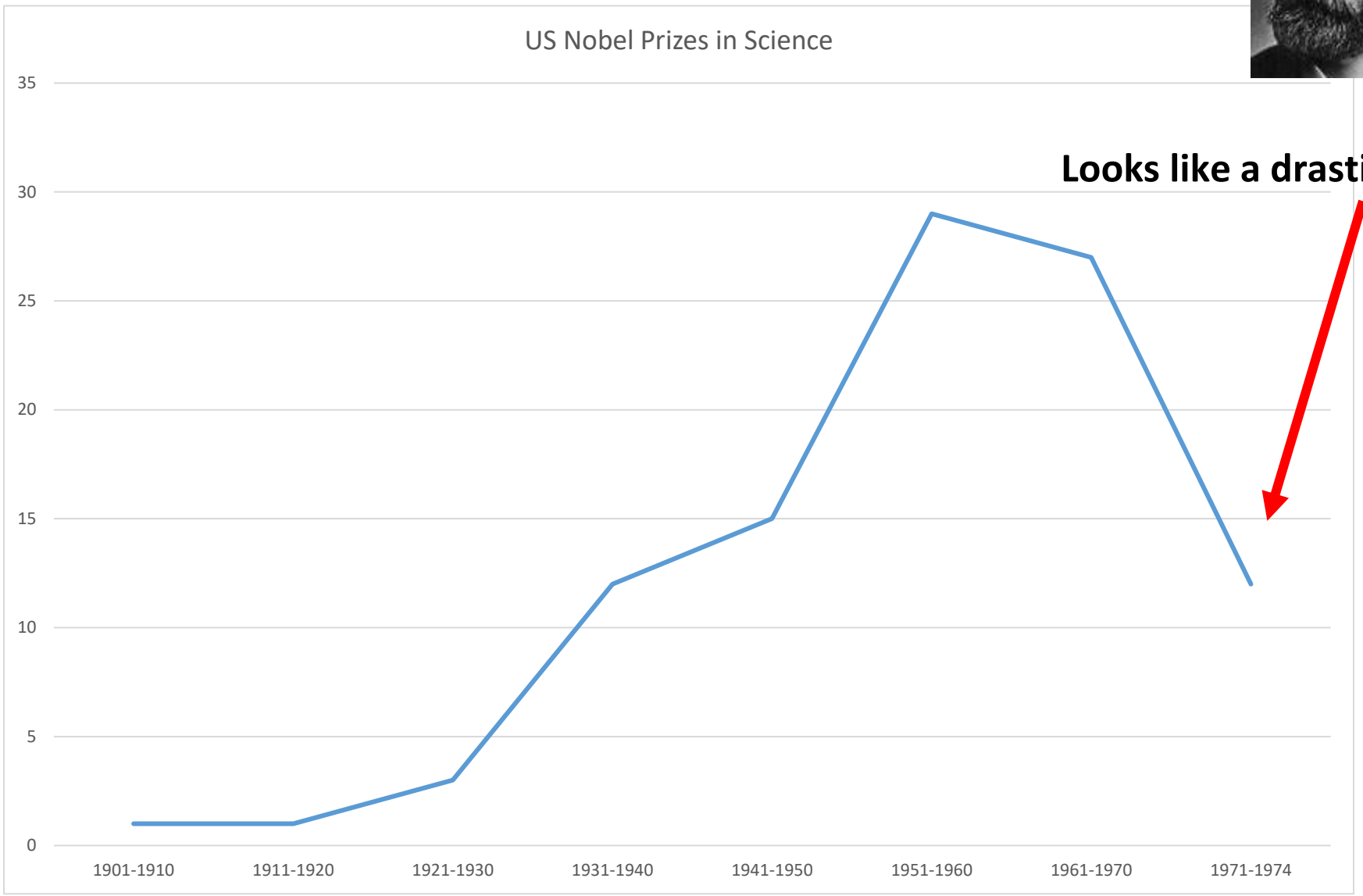
Looks like unusually large growth!



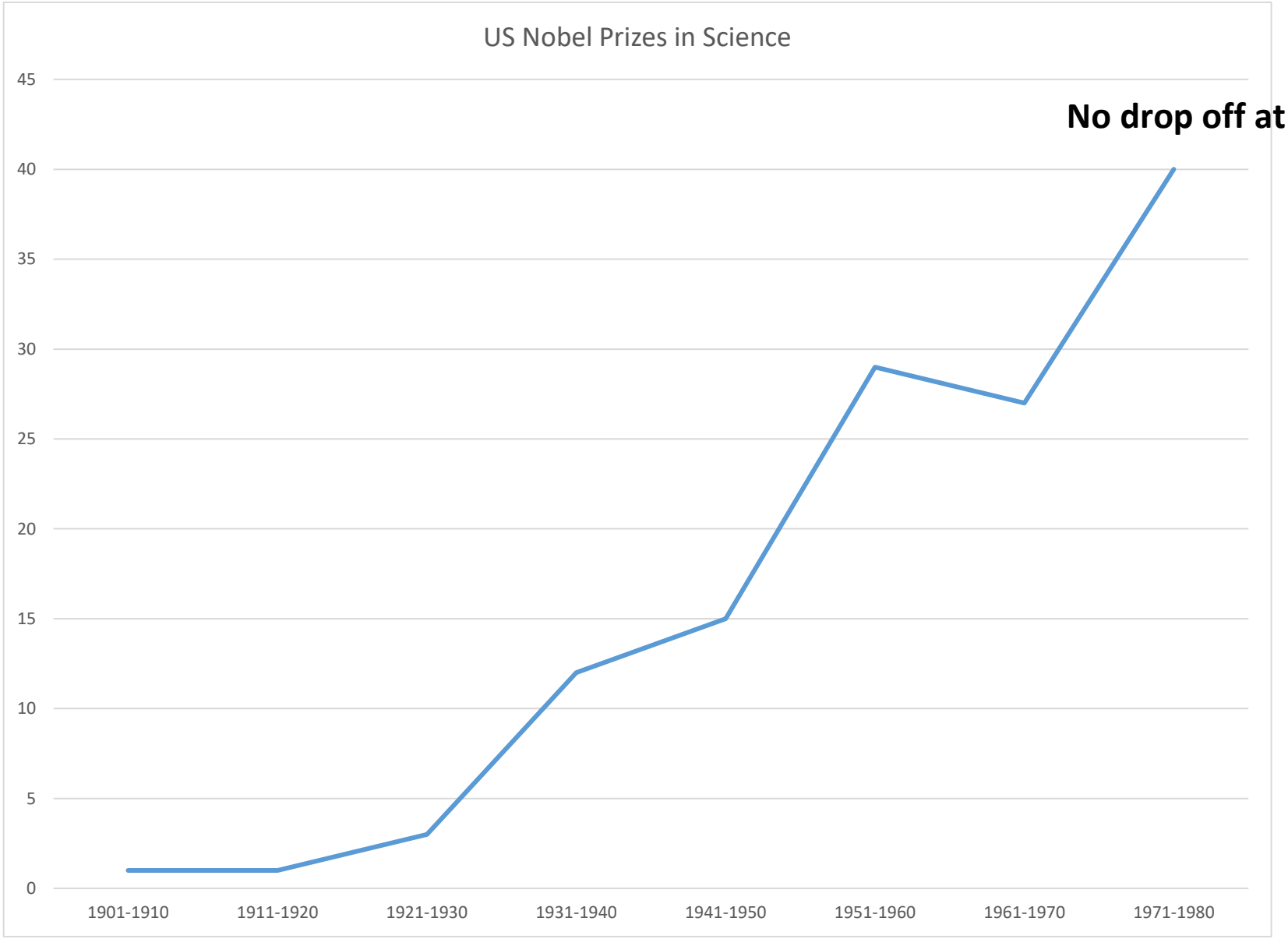

Households Earning More Than $100,000/yr
6
5
4
3
2
1
0
1965
1970
1975
1980
1985
1990

Inconsistent Classes

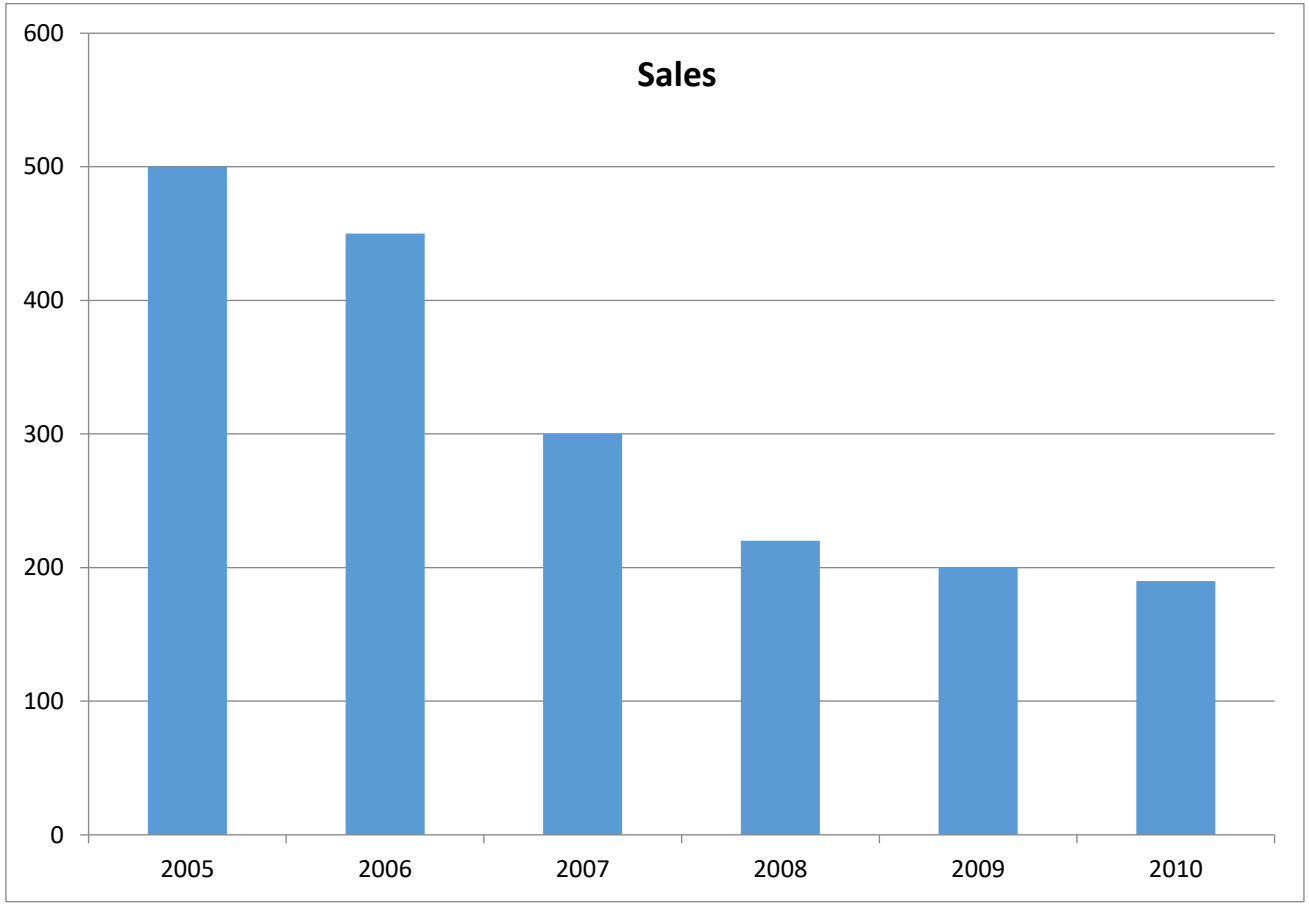




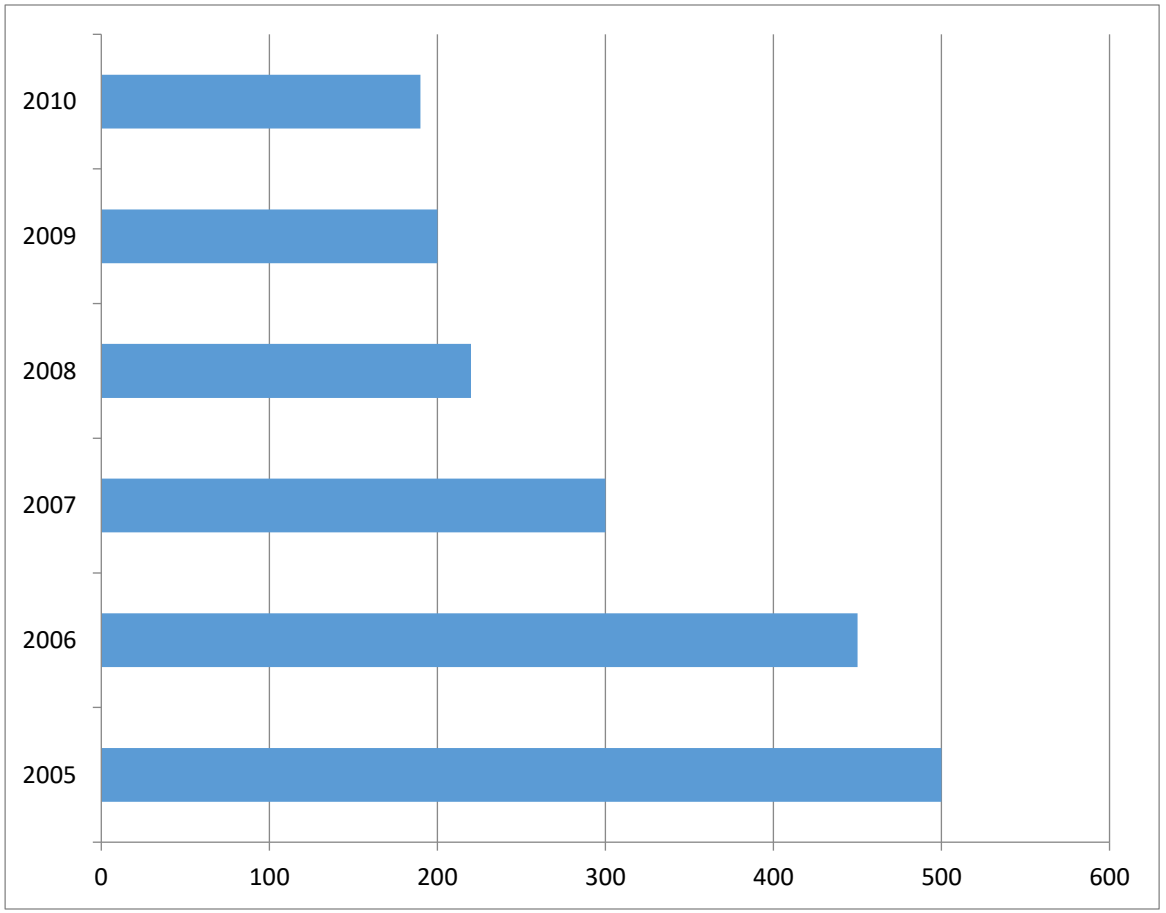
Looks like a drastic drop!


US Nobel Prizes in Science
45
40
35
30
25
20
15
10
5
0
1901-1910
1911-1920
1921-1930
1931-1940
1941-1950
1951-1960
1961-1970
1971-1980
No drop off at all!

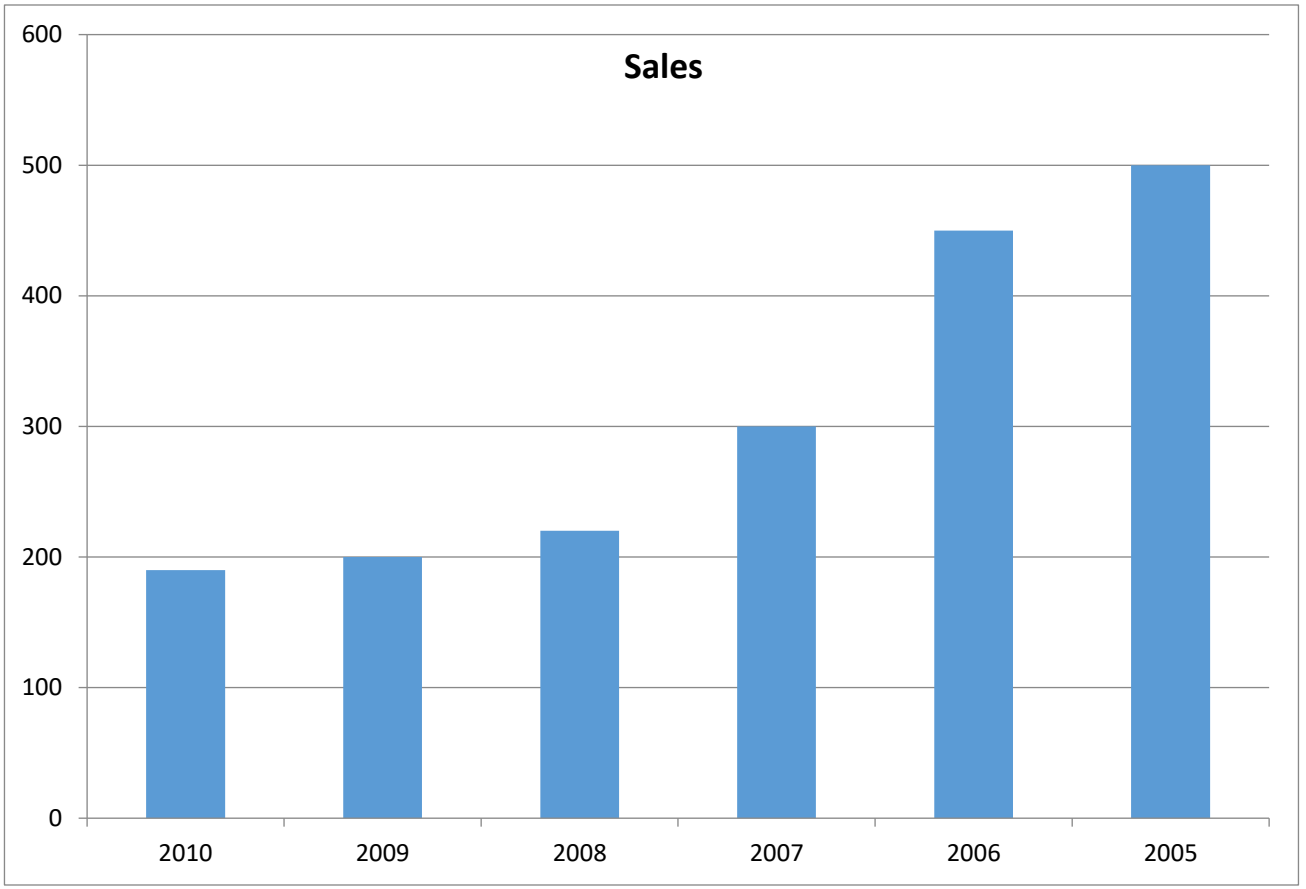
Misleading Trend



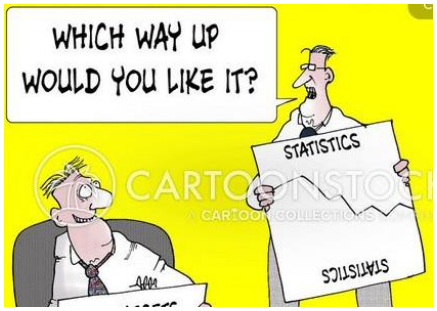
There's a definite downward trend.

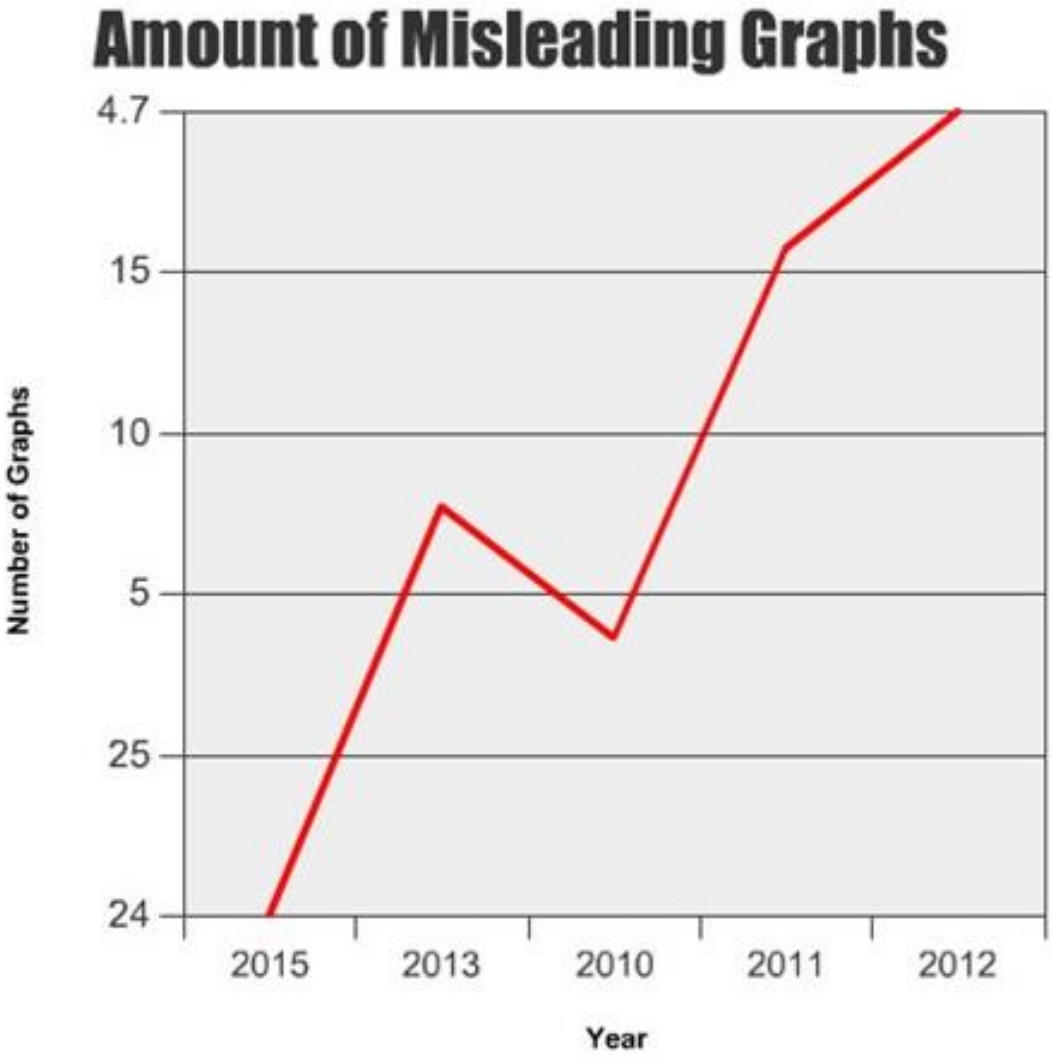


Trends are harder to detect in a vertical arrangement, so switch the axes.

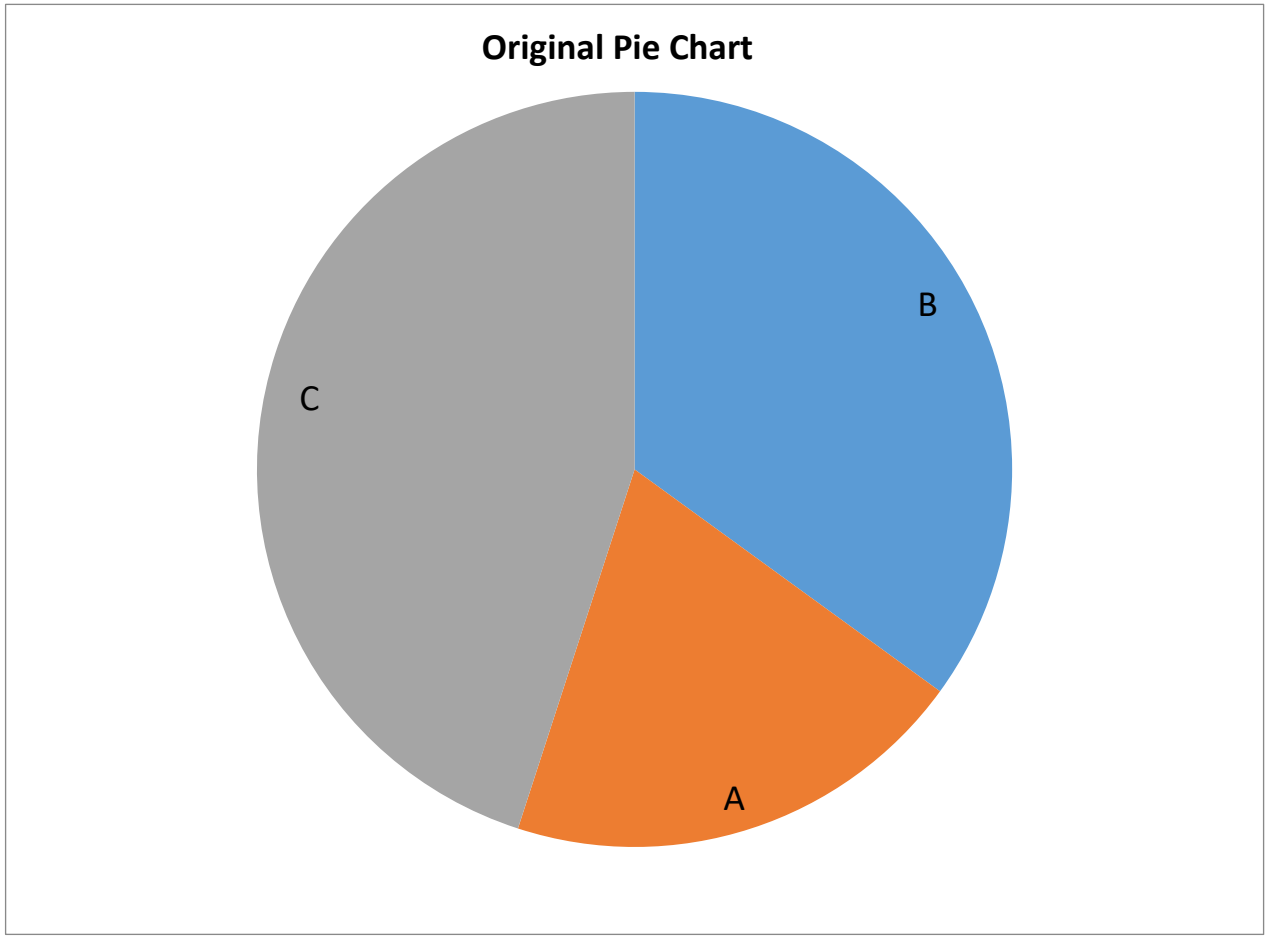


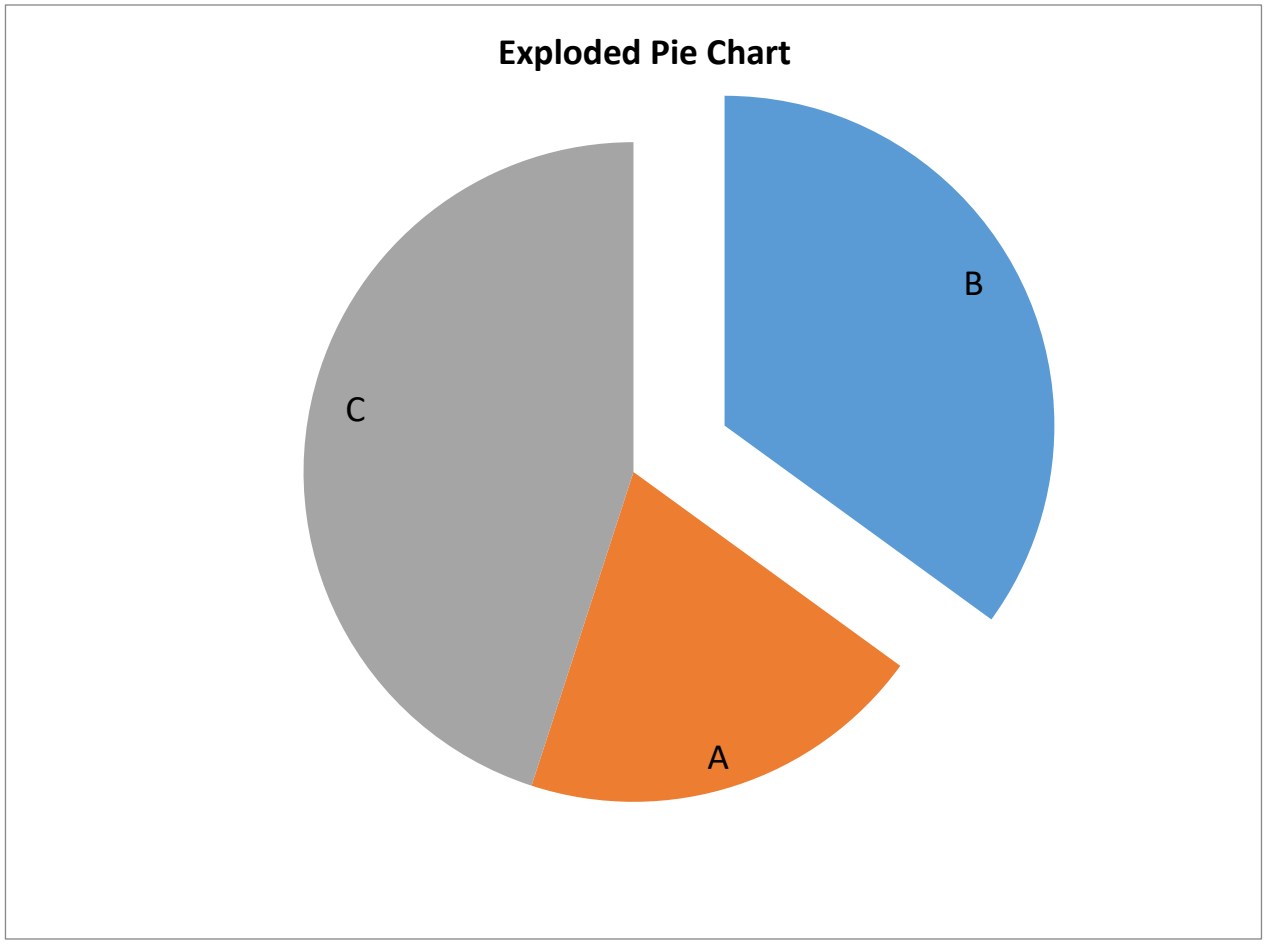
Reversing the horizontal axis turns a downward trend into an upward trend.



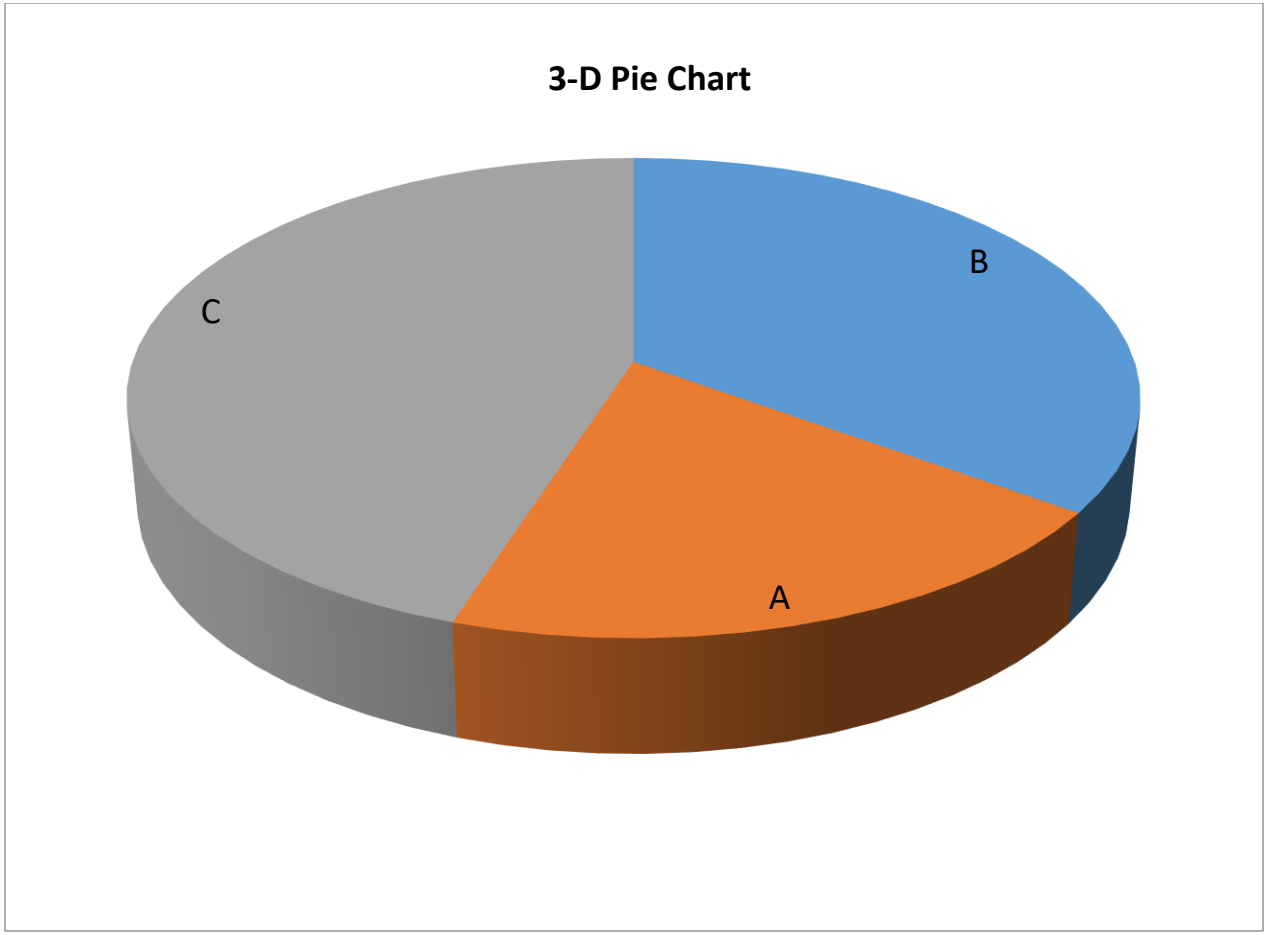

Amount of Misleading Graphs
4.7
15
10
5
25
24
Number of Graphs
2015
2013
2010
2011
2012
Year

Misleading Pie Chart

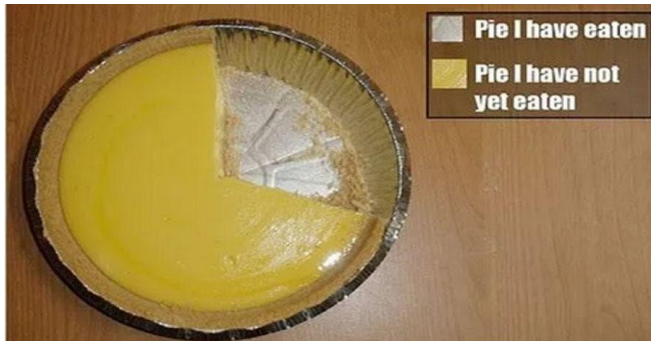




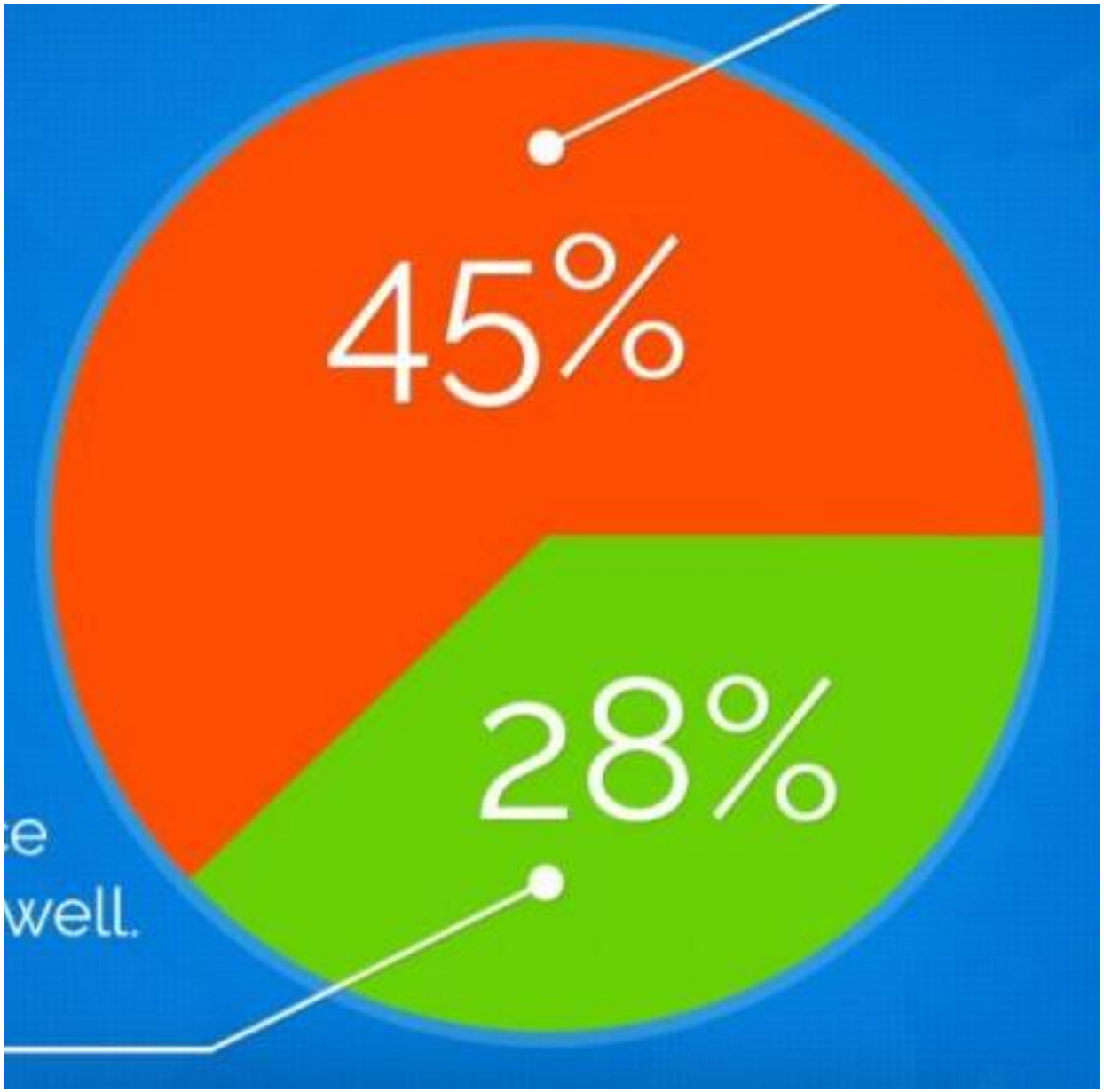
Emphasis is placed on the exploded piece.



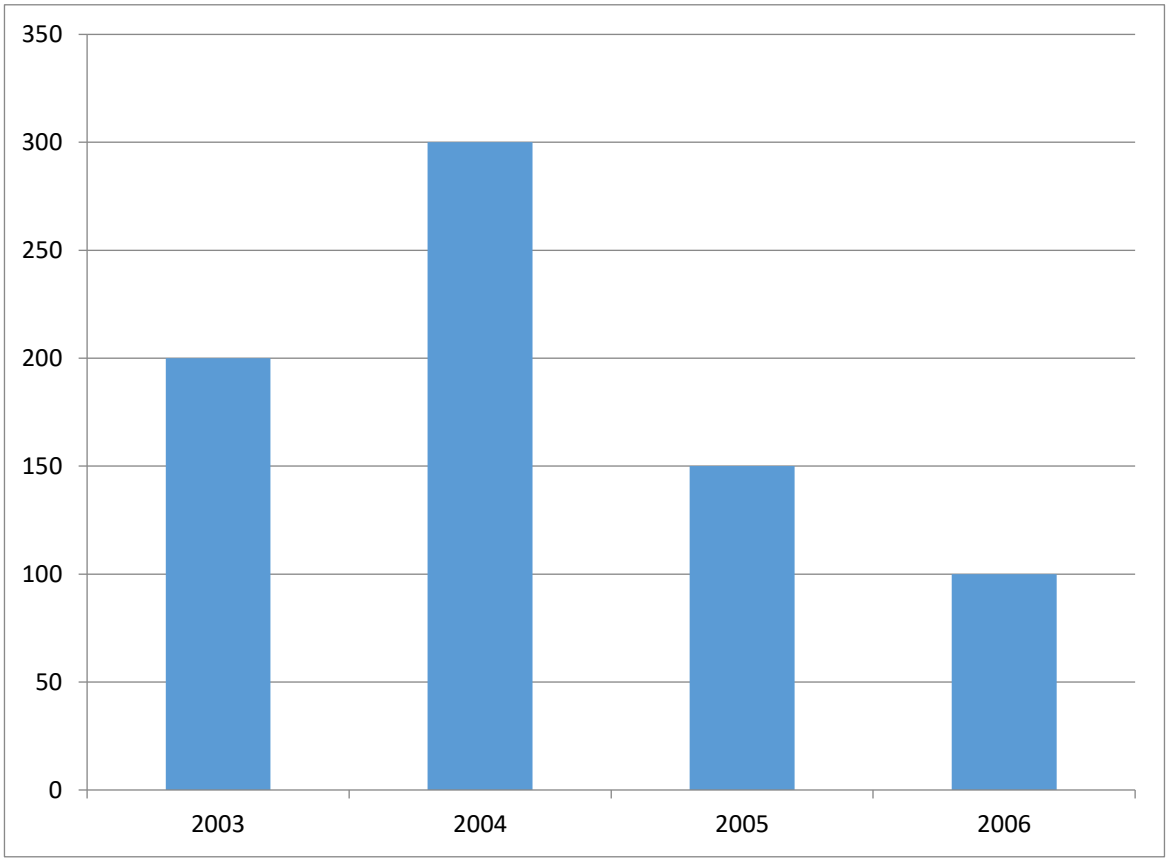
3-D can make comparisons more difficult.

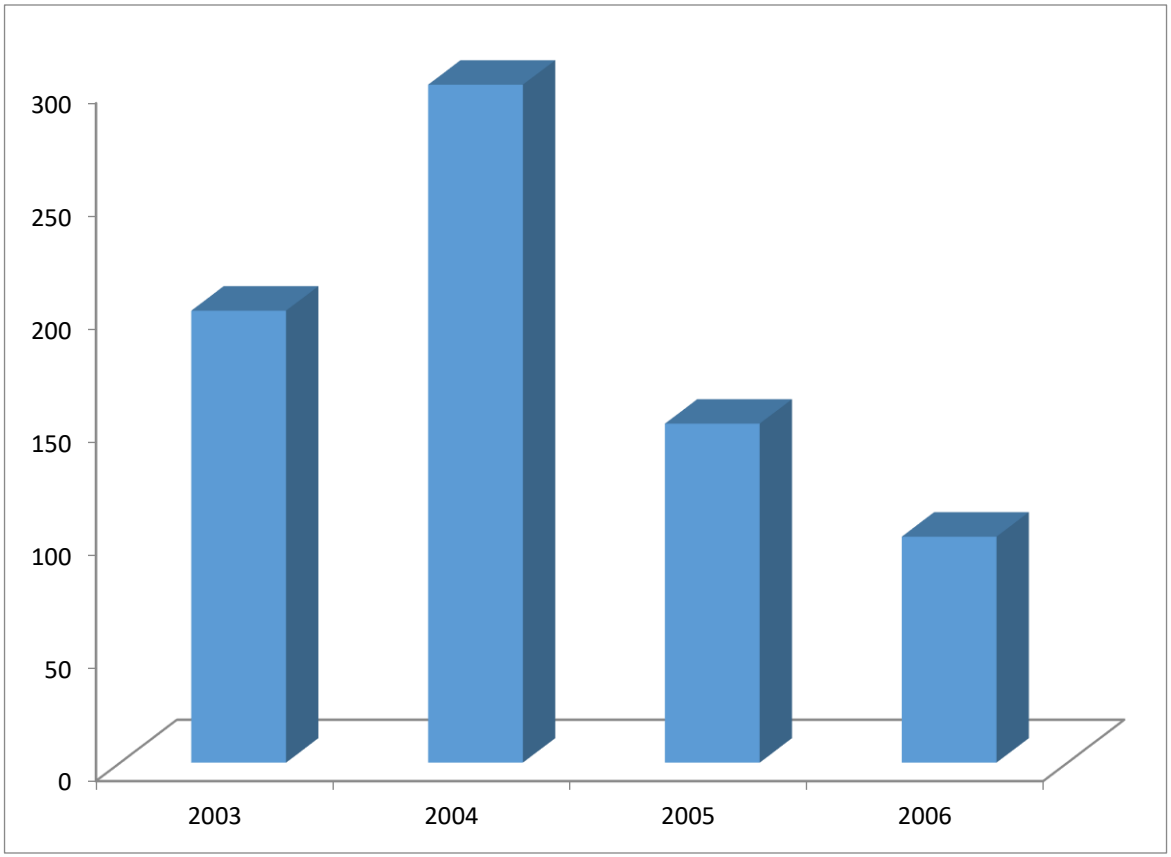



WHAT GROUP PROJECTS TAUGHT ME
COMMUNICATION
RESPONSIBILITY
COLLABORATION
TEAMWORK
TRUST NO ONE

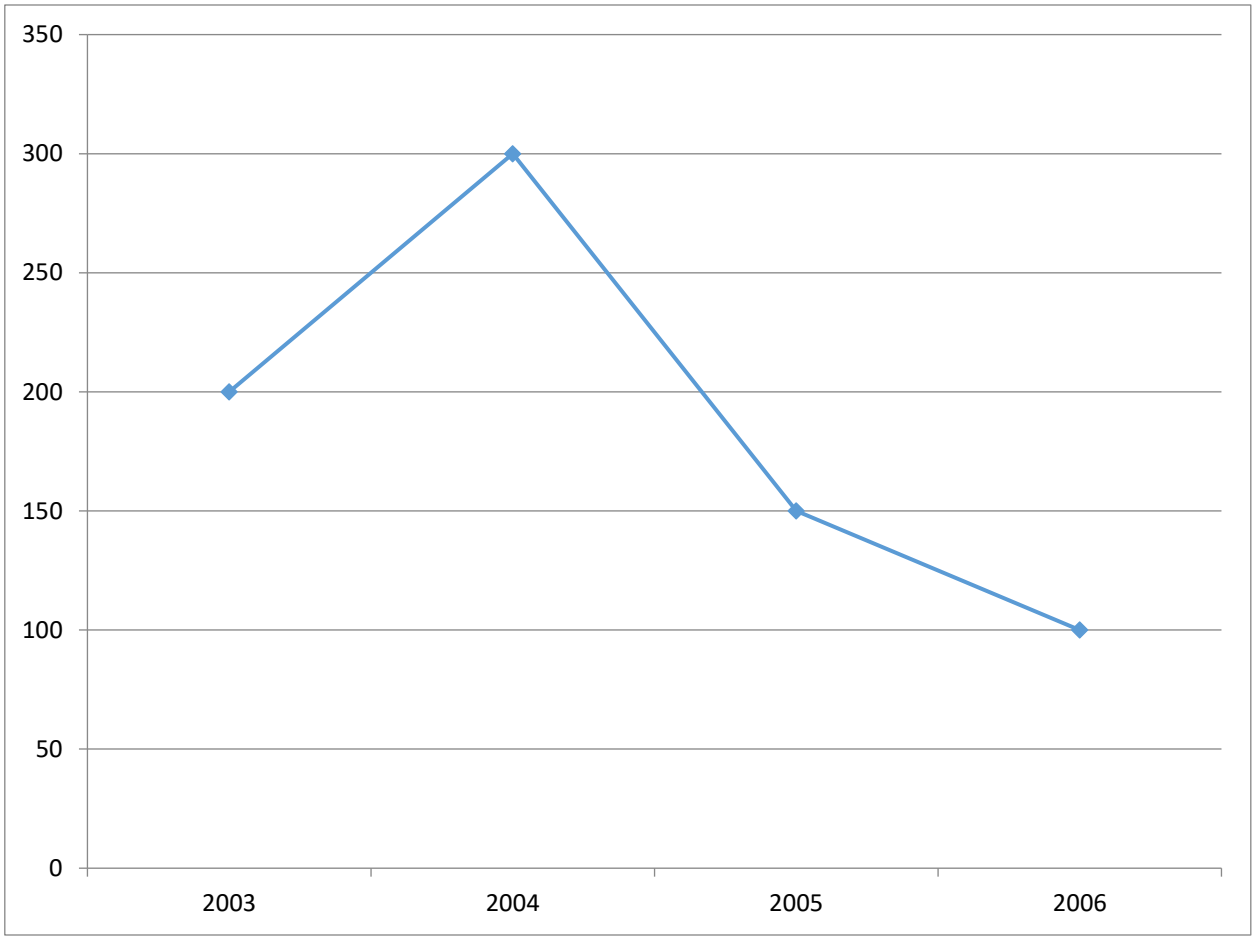

45%
28%
well.

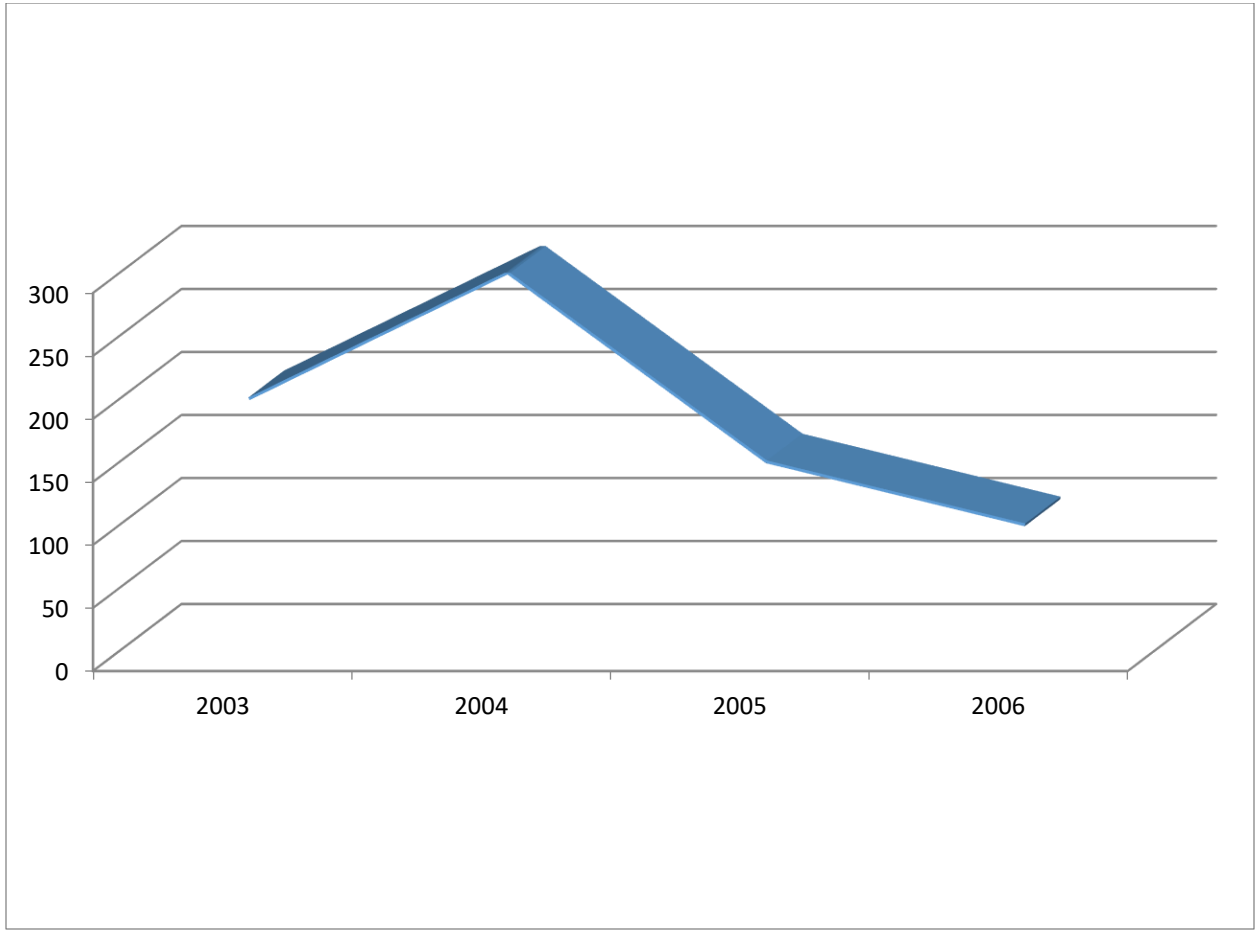
3-D Effects





The 3-dimensional bars make it more difficult to read the actual values.


350
300
250
200
150
100
50
0
2003
2004
2005
2006



The surface makes it more difficult to read the actual values.

Our Company's Sales Have Soared!