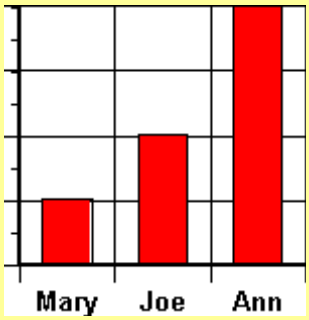
How to Lie and Cheat with Statistics

Ok, this is what you have been waiting for. How can you lie with statistics? Actually, the purpose of this page is NOT to teach you how to lie and cheat with statistics. Rather, I hope you will learn how it is possible to be misled and how to spot "statistical abuse." You can find poor use of statistics everywhere: magazines, newspapers, polls, TV, even research papers. I do not want to hear of any of you readers using these poor methods.

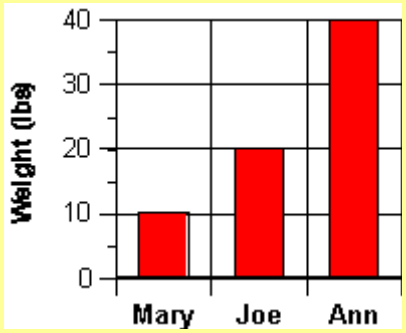
Games with Graphics

Misusing and abusing graphics are easy ways to mislead people. People like to see graphs for a quick way to evaluate a set of numbers. But BEWARE! Make sure you are not fooled. Let's use pumpkins grown in the gardens of Mary, Joe and Ann. Here is the first graph:



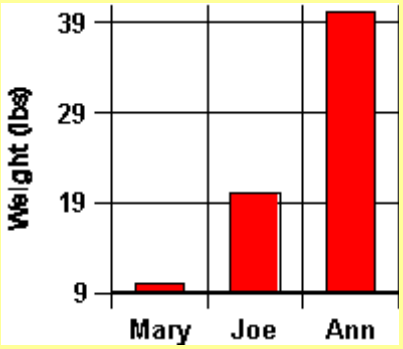
This graph does not say very much because there is no scale on the y-axis. Does the graph represent the weight, volume, width or height of the pumpkins? It does not say.

Here is a graph that is much better:



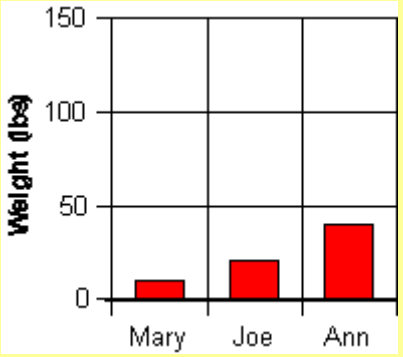
Now we know that the graph refers to the weight of the pumpkins and we know how much each pumpkin weighs because the numbers are given. This is a fair graph.

What if you wanted to convince people that Ann's pumpkin was bigger than Mary's and Joe's pumpkin. Look at this graph:



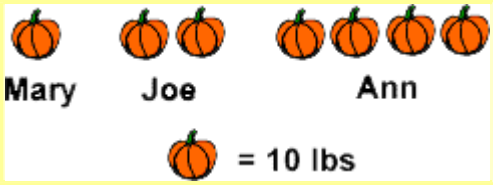
The numbers are the same, but the y-axis has been changed. Now it *appears* that Ann's pumpkin is much bigger than the other two.

What if you wanted to convince people that all the pumpkins were about the same size. Look at this graph:



The numbers are the same, but the y-axis has been changed again. This time the y-axis has expanded. Now it *appears* that there there is only a small difference in weight.

Often a picture is used to represent data. Here is a fair way to show the difference in the weights of the pumpkins using a picture:



In this graphic, each pumpkin represents 10 lbs. It clearly and fairly shows the difference in the weights of pumpkins from the different gardens.

However, let's try to show that Ann's pumpkin is much bigger than the rest:

This graphic distorts the data. To show the differences in the weights, this picture changes the height of each pumpkin to represent pumpkin weight. Joe's pumpkin (20 lbs) is twice as high as Mary's (10 lbs.); Ann's (40 lbs.) is twice as high as Joe's (20 lbs.) and four times as high as Mary's (10 lbs). Is this fair? I think not! The reason is because as the height of the pumpkin is increased, the WIDTH of the pumpkin increased. Therefore, while the heights are in proportion the

AREAS of the pumpkins are not. Remember, the formulas to determine area:

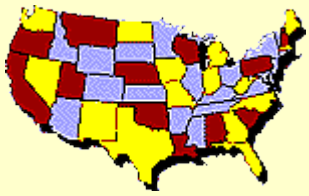
Area of a rectangle = Height X Width

Area of a circle = πr^2

So this picture makes it look like Ann's pumpkin is much larger than Mary's and Joe's. I also used different sized letters for the different pumpkin weights to give the impression that Ann's pumpkin was larger.

Meaningless Graphics

Newspapers and magazines like to use colorful pictures to represent public opinion and survey responses. However, often times the pictures are too simple to give meaningful information. Take this example:



This map shows how people in different states of the US like pizza. (I just made up these data). The code for the state color is:
Red States = People Love Pizza
Yellow States = People Like Pizza
Purple States = People Hate Pizza

That's all the information we have. The map really doesn't say very much. We don't know how it was determined that people like pizza...were people asked if they liked pizza? Were people asked how much pizza they ate in a week? a month? a year? Was the number of pizzas purchased at stores in different states counted? Was the number of pizza restaurants in different states counted?

We also do not know if there are any real differences between how much people like pizza in the different states. How much do people love pizza in California? What is the difference between how much people love pizza in Utah compared to how much they like pizza in Nevada? There are no scales or measurements to indicate any of this information. Although this type of graphic gives almost no information, it is used frequently in many popular magazines.

For more ways to misuse statistics, there are two interesting books:

1. Darrell Huff, *How to Lie with Statistics*, W.W. Norton & Co., New York, 1954 (reissued in 1982 and 1993).
2. Cooper B. Holmes, *The Honest Truth About Lying With Statistics*, Charles C. Thomas, Springfield, 1990