Graphing, Correlations And Hurricanes

Find on TCA!

Another issue we all grapple with is good graphing. More than that, good graphing interpretation!

Year	Sunspot Number
1950	76.0
1951	58.3
1952	29.6
1953	13.6
1954	4.4
1955	38.1
1956	126.1
1957	165.9
1958	175.1
1959	149.5
1960	103.8
1961	49.1
1962	31.4
1963	24.5
1964	10.2
1965	14.6
1966	43.8
1967	95.8
1968	98.2
1969	96.0
1970	108.5
1971	73.5
1972	72.0
1973	39.3
1974	34.0

YEAR	Tropical Systems	Hurri-canes
1974	11	4
1973	8	4
1972	7	3
1971	13	6
1970	10	5
1969	18	12
1968	8	4
1967	8	6
1966	11	7
1965	6	4
1964	12	6
1963	9	7
1962	5	3
1961	11	8
1960	7	4
1959	11	7
1958	10	7
1957	8	3
1956	8	4
1955	12	9
1954	11	8
1953	14	6
1952	7	6
1951	10	8
1950	13	11

On the first, I am a stickler for good graphing technique. My standards are that the graph axes should be labeled, with units of measurement shown (if more than just numbers or discrete categorizations). The graph should have a title WHICH SHOULD NOT INCLUDE THE AXES TITLE WORDS but should be like a subject line in an email message. So, if your graph is temperature versus height, perhaps the title should be something like "Balloon thermometer readings at various levels of the atmosphere."

Another issue is graphing scale. I insist that the graph should fill the paper. When I say the graph should fill the paper, I really mean that whatever the maxima and minima are, they should be near the top and bottom, and left and right borders—magnify the graph so we can see trends! Just leave enough other space around the axes' descriptions and interval labels, and similar margins around the other two sides. There is an opposing view—some people may think that if you have the same experiment but with different objects (like the positions of Jupiter's four big moons) then they should all have a common scale so we can see differences between the graphs. For students struggling with graph interpretation, I would rather they put the dots as far apart as possible. Graphs with a *lot* of white space make it hard to see trends and patterns.

Now, on to correlations. Here we mean that if there is a trend, e.g. Y may increase linearly with increases in X, or decreases with X, or even has a curved trend with X, then point it out. I don't mean there has to be an exact function or that if you connect the dots you get a smooth curve—with real observations that almost never happens! But if there is a trend, then you have some kind of correlation.

If you do NOT have a correlation, then the graph will be either a horizontal line—Y is essentially the same with all values of X—or a vertical line—all Y values occur at the same value of X. Another non-correlation is when there is NO apparent line you can make, the dots are too scattered around the graph such that a computer could make a line but with an uncertainty so high as to make the slope of the line meaningless.

It is easy to find values that correlate, linearly like adiabatic cooling, or with a curve like the relationship between periods and distances in Kepler's Third Law. Here's an example of a non-correlation that's in the news—hurricanes versus sunspot counts. Research question: do we get more hurricanes (or fewer hurricanes) when sunspot numbers are high? Above are two sets of data—annualized sunspot numbers for 25 years of data, and the numbers of hurricanes for the same years. Have them plot the number of hurricanes on the Y axis and the sunspot numbers on the X. Warning! Too often, despite the specific instructions, students could not get past the frequent use of plotting the YEAR along the X and putting in two plots of data on the same graph!

You should get no visible correlation at all, though sometimes it appears that there is a negative correlation (storms get fewer with higher sunspot numbers) in some plots I received.