A Statistical Tale

In the last couple of centuries, the basis of our rational thought has changed from being mostly anecdotal to mostly statistical. When used properly, each viewpoint has its purpose and value. Each person knows his own personal experience, and statistics suggest how common that particular experience is. Statistics were used successfully in shaping both business and government policies for a long time, but something gradually changed within the intellectual community. The balance between the two viewpoints has shifted until anecdotal evidence is ridiculed while statistical evidence is worshipped. Statistics now trump everything else, despite the narrow education and focus of those in charge of the statistics. Lately, the statistical tale has been wagging the dog.

The subject of Global Warming will serve as a good example. The charts at the end of this essay compare sunspot activity, temperature, carbon dioxide, and solar output from 1700 to 2010. Although validity of some data can be debated, the charts are good enough for common sense analysis. The following implications are pretty clear.

1. Sunspot numbers fluctuate on an eleven-year cycle and correspond to the level of total solar irradiation rather well.
2. Atmospheric carbon dioxide has increased steadily since 1860; and at an increased rate since 1965. The use of fossil fuels is a plausible cause. Deforestation may be another.
3. There is no correlation between temperatures in central England and Aberdeen, North Dakota.
4. There is no correlation between sunspot activity and temperature because temperatures lack an eleven-year cycle, and relatively high and low periods of sunspot activity and temperature don’t match most of the time.
5. There is no correlation between levels of carbon dioxide and air temperature because the recent sustained increase in carbon dioxide is not matched by a corresponding sustained increase in temperature.

Statistics can only be used to suggest that a trend is present or that a correlation exists. However, the results of statistical analysis are not absolute. They are right most of the time but are also wrong some of the time. Statistics can't prove anything, and they certainly never provide a specific cause and effect! Nevertheless, statistics have been used to draw two incorrect conclusions about temperatures.

About a hundred years after the recording of temperatures and sunspots began, the prolonged cold spell in England ended around the same time as an extended period of low sunspot activity. Logically, more solar radiation should result in higher temperatures. This direct relationship appeared even more credible as the temperature and sunspot numbers declined in concert for four decades starting about 1730. Unfortunately, that was the end of the apparent relationship and subsequent data was rather ambiguous. Even so, it's tempting to revive the theory after temperatures "unexpectedly" dipped in recent years while sunspot activity dropped to levels not seen in a long time.

In the second case, a general increase in temperatures from 1970 to 2003 occurred while the concentration of atmospheric carbon dioxide was also increasing. It seemed plausible to many that the carbon dioxide could be creating a greenhouse effect thus raising temperatures. However, the temperature generally decreased after 2003. In addition, temperatures generally decreased from 1930 to 1969 while the carbon dioxide was also increasing, albeit at a slower rate. When all this is considered, the global warming theory in its popular form is hardly tenable.

When statistics come from a limited amount of data and are used to determine relatively small trends, the results are unreliable. Sure, it's possible to present the data in a way to make a very small and somewhat ambiguous trend appear very important, but that doesn't change the truth. The visual impact of graphs is often increased by showing trend lines rather than actual data and by only showing a selected portion of the chart. The latter is like showing the undulating surface of a valley while ignoring the neighboring mountain. In the limited context, the valley appears very uneven: In the full context it appears level.

Even worse is using projected data to extend a chart. Whether the projected data comes out of the air or out of a computer model makes no difference. No fortune-teller, economist, weatherman or climatologist really has the gift of seeing into the future, and their forecasts should not be commingled with historical data.

Early last September, I was down at the Jersey Shore listening to the surf in the dark. The frequency of the waves varied and there were even occasional periods of absolute silence. The volume and quality of the sounds changed from wave to wave. Sometimes a succession of very similar waves came in, and sometimes not. While I naturally searched for a pattern and seemed briefly on the verge of success, the only sustained pattern was that of randomness. After looking at the shape of the temperature data for a while, a moment of insight occurred. It was just another excellent example of randomness. Just as I tried to find a pattern in the waves and was repeatedly misled, others have tried to find a pattern in the temperatures and were misled as well.

This doesn't negate cause and effect or prove that sunspots and carbon dioxide have no effect on temperature. It just shows that in both cases statistics have been used to exaggerate the importance of a theory, and more significantly, that a man convinced by statistics has a very hard time relinquishing his beliefs despite additional statistics.

The patterns of ocean waves and temperatures result from very complex causes. Some of the causes happened way in the past, some a while ago, and others nearly in the present. The resulting randomness in the data makes detecting weak causes and effects nearly impossible. Moreover, conclusions become dangerous when only short-term data is used, and especially dangerous when future projections are used. This is exactly what a gambler with a betting system does, and he usually loses! Future projections are usually flawed because future causes can't be predicted with accuracy. Although carbon dioxide may continue to rise based on its consistent behavior over the last 150 years, we don't know for sure, and its actual effect on temperature is ambiguous. Solar output may change dramatically, but no one can say when or by how much.

Statistics have crept into all aspects of our lives. No "serious" intellectual research can be published without them. Lawyers, accountants, and administrators swear by them. They've even reached areas like the study of literature. Our educational system has always been at the center of this. In the beginning statistics were successful in dispelling myths, but now that the shortcomings of statistics have been forgotten, statistics are often used to create myths in support of social and political agendas.

Recently, I wondered when and how the change began. For me, it happened at public school in about 4th grade. Around that time, the daily saying of the Lord's Prayer and the teacher's reading of a selected passage in the Old Testament ended. Being non-religious, it didn't matter much to me. Was it a coincidence that the idea of individual responsibility began to fade just as religion was banished from public schools? Nearly simultaneously, New Math was introduced with great fanfare. It relied heavily on Set Theory where individual items were grouped into sets and subsets which could be manipulated in their entirety. Emphasis was on the collective, not the individual. New Math didn't impress me much either. In 3rd grade I'd already done some work in an old 6th grade math textbook and found set theory a bit elementary and contrived. At first I thought New Math was a scheme to sell new textbooks to schools or retraining to teachers, and later that it was also intended to alienate parents from their children's education. Today I am even more suspicious.

Traditional statistics were compiled by recording data from every member of the group. The person doing the study gained great familiarity with the individual members. Statistics answered questions like "how many members are in the group?" or "how much did they produce or consume?" When this information was combined with the broad knowledge gained from doing the study, it was possible to rationally plan for the future.

Today's statistics are much lazier. Only a small sample of the group is examined, usually by a third party. The person doing the study never meets a single individual member in person, and this prevents him from acquiring any common sense on the subject. Within a room with no windows, the statistician creates a mythical being based on statistics. And then, he confidently predicts the future.

A more insidious aspect of Set Theory is the division of people into statistical subsets according to superficial attributes. Segregation in the South is an early example of this and coincidently came along shortly after a revival of Set Theory in the 1870s. We are now often managed and manipulated according to what subsets we belong to. Politicians and advocacy groups are especially guilty of this, often even pitting one subset against another. Policy targets statistical stereotypes of groups such as "working families", the "gay community" or "the rich" without regard for the diverse individual situations within each group. The resulting legislation may be needed and welcomed by some in the group, but it is unnecessary or even injurious for many others. So much time is spent catering to special interests that legislation to "promote the general welfare" is overlooked.

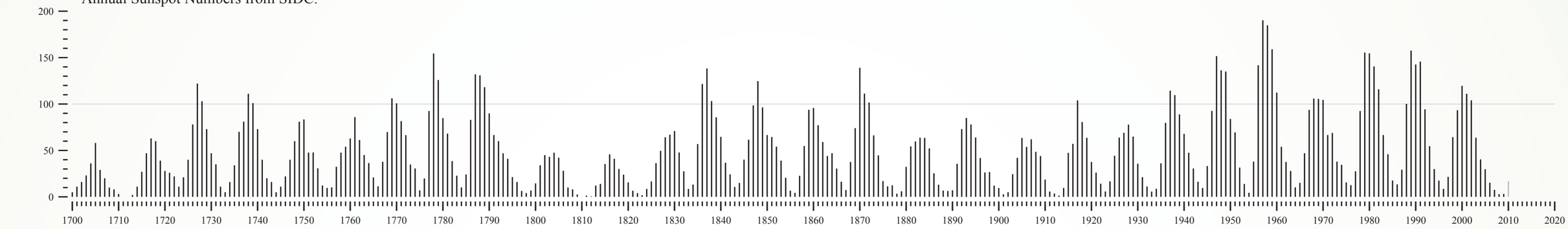
We are becoming increasingly uncomfortable in our statistical boxes as we see that new legislation helps our own situations less and less. We reach out to people in other boxes and find a surprising amount of agreement on the issues that really matter to us. The politicians sit in windowless rooms and cater to their imagined constituents. They don't have a clue!

Eric A. Neubauer
December 22, 2010

Solar output, temperature, and atmospheric carbon dioxide charts

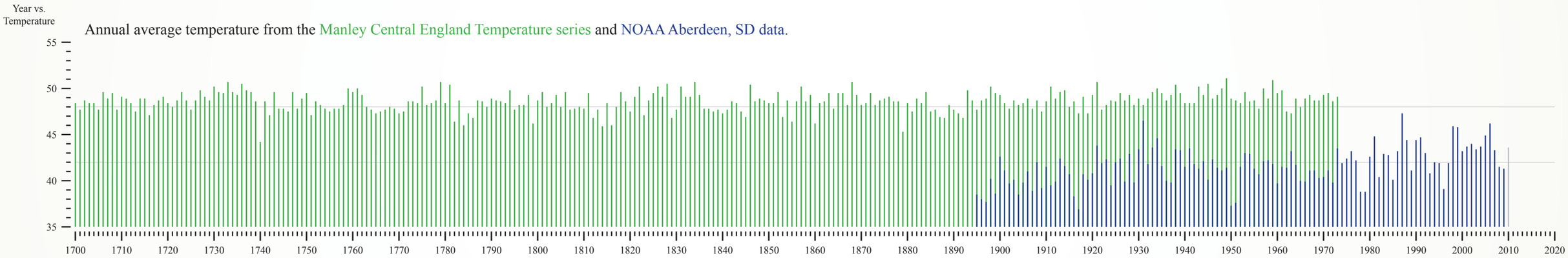
Year vs.
Sunspots

Annual Sunspot Numbers from SIDC.



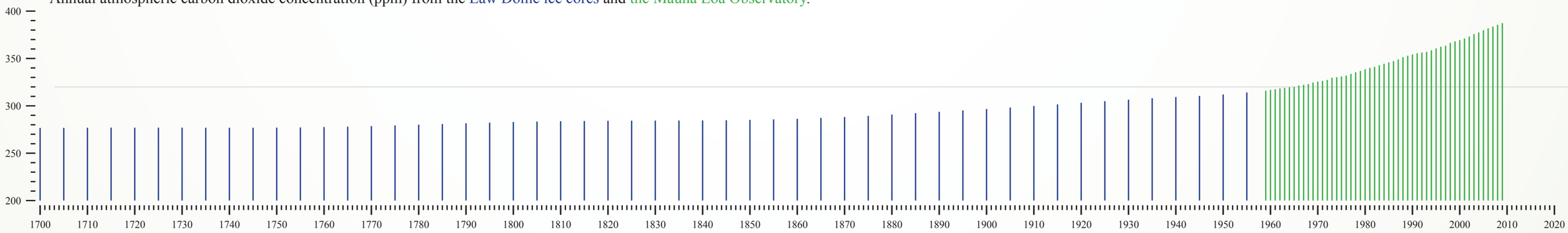
Year vs.
Temperature

Annual average temperature from the Manley Central England Temperature series and NOAA Aberdeen, SD data.



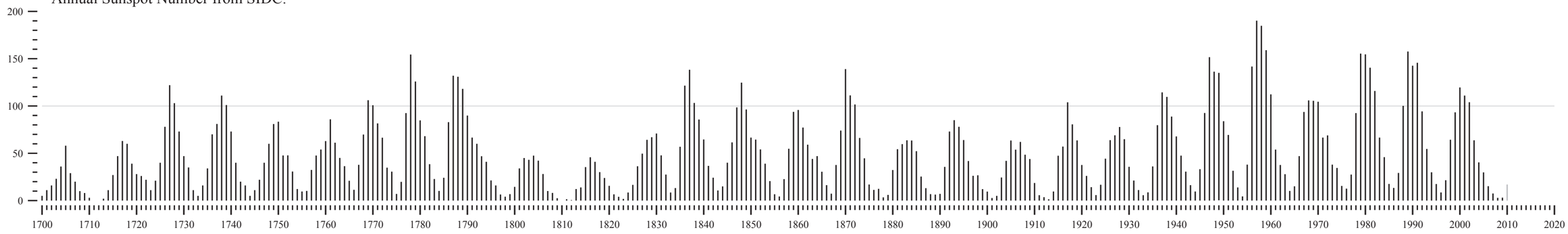
Year vs.
Carbon Dioxide

Annual atmospheric carbon dioxide concentration (ppm) from the Law Dome ice cores and the Mauna Loa Observatory.



Year vs. Sunspots

Annual Sunspot Number from SIDC.



Year vs. Irradiance

Average annual Total Solar Irradiance (W/m^2) from C. Fröhlich satellite data composite.