Representing and Presenting Data

The Role of Data in the Social Sciences

To understand why most studies in the social sciences use data, we have needed to understand the role of theory. Data are used to see whether a theory is useful. That means that the theory predicts some aspects of behaviour. But how we use data to assess our theory is an art in itself. Sometime the evidence is statistical, sometimes it is simply observational. It depends on what your theory needs. Regardless of the details of your theory, you will need to organize and characterize your data. It is to this we now turn.

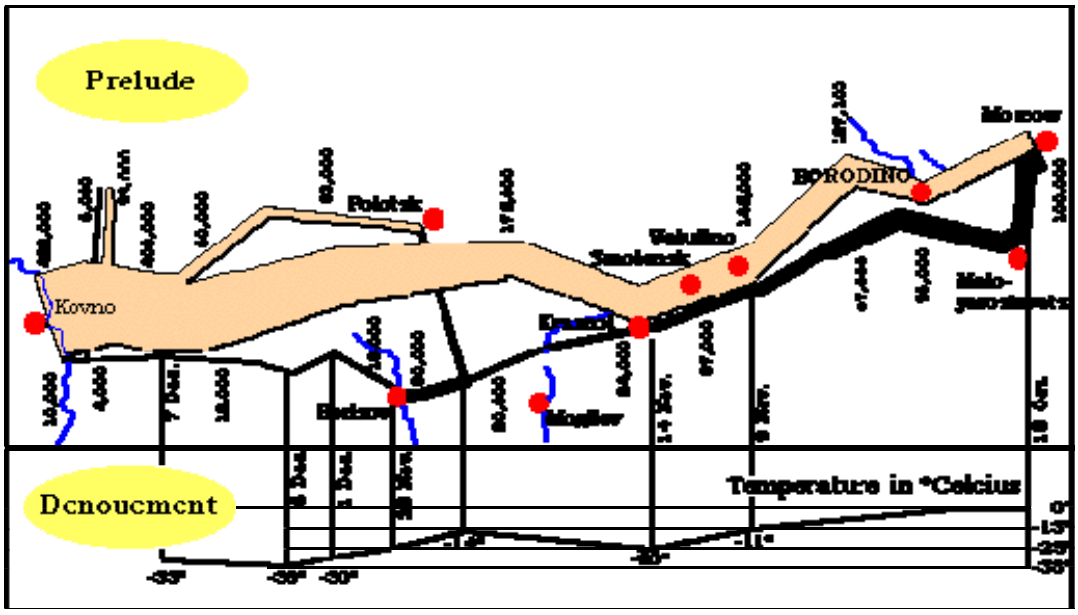
Organizing Data

There are many ways to organize information. Not every way will be appropriate for your problem, but if you have data, you need to know how to present it. In what follows, I hope that it gives you ideas. There are an infinite number of ways to organize information and since your task is to help the audience to understand your information in the context of your problem, you will have to select the displays that help understand best.

Let us look at some of the basic ways in which information can be organized.

It is sometime said that a picture is worth a thousand words. Sometime that is surely true. Napoleon's march on Russia is a classic case.

http://uts.cc.utexas.edu/~jrubarth/gslis/lis385t.16/Napoleon/index.html May 8, 2006



“The map, based on the 1869 chart by Minard, graphically illustrates (both literally and figuratively) how the size of the French army dwindled during the march into Russia and was reduced to almost nothing on the wretched rout back into Poland. The map can be read in several ways. The size of the peach colored bar indicates the relative strength of the French army during the march on Moscow. The black bar shows the dwindling French army during the retreat. In the lower portion of the map, the temperature in degrees Celcius is shown, along with dates during the retreat.”

The picture in this case gives a highly informative characterization of the way in which Napoleon’s army was devastated from its beginning as a massive body of men in xxxx to its final return a year later as a devastated trickle.

Another basic way is a table. Although basic, a table must be organized. The dimensions of the table must convey useful information. Suppose we believe that crime is universal so that the crime rate should be the same in every country. Obviously, one thing we might want to look at is whether in fact crime rates are the same in different countries.

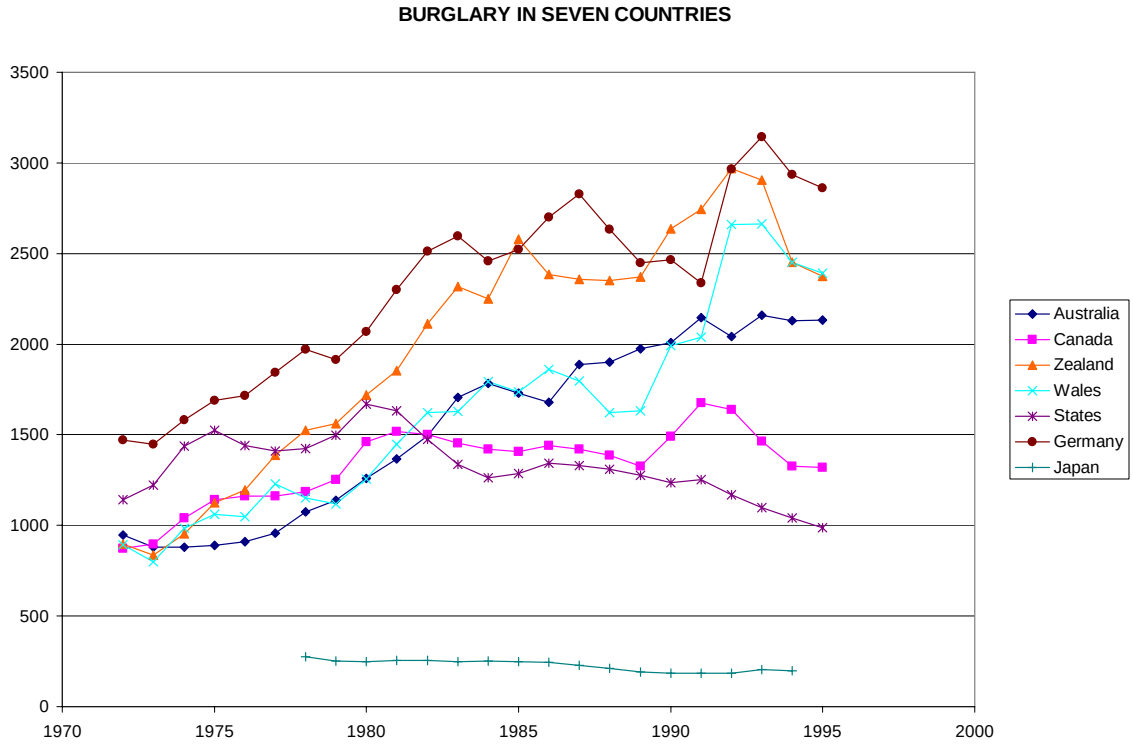
One way to see this is in a table:

			Burglary: Rate per 100,000 of population				
Year	Australia	Canada	New Zealand	England Wales	United States	Germany	Japan
1972	948	874.6	895.5	894.8	1140.8	1472.1	
1973	880.4	896.3	837.5	799.9	1222.5	1448.1	
1974	880.9	1039.6	952.6	984.2	1437.7	1581	
1975	890.9	1143.2	1126.6	1061.6	1525.9	1689.4	
1976	910.2	1161.1	1194.1	1048.9	1439.4	1716.3	
1977	958.7	1162	1386.6	1229.7	1410.9	1844.6	
1978	1075	1185.9	1525.6	1151.7	1423.7	1972.4	274.3
1979	1137.2	1252.3	1560.7	1118	1499.1	1915.8	250.5
1980	1260.7	1462.3	1721.2	1255.9	1668.2	2068.7	247.7
1981	1367.8	1518.2	1855.3	1448.3	1632.1	2299.3	255.3
1982	1496.3	1501.5	2114.3	1623.8	1475.2	2512.7	254.4
1983	1706.5	1456	2316.1	1629.7	1337.7	2597.1	249.3
1984	1782.2	1421	2251.9	1793	1263.7	2457.3	251
1985	1730.1	1407	2580.3	1736.9	1287.3	2523.2	247.8
1986	1680.6	1440	2385.8	1860.3	1344.6	2699.9	244.1
1987	1888.7	1421	2356.3	1797.3	1329.6	2829.4	228
1988	1901.1	1386	2349.7	1623.2	1309.2	2632.9	211.4
1989	1974.1	1328	2372.1	1633.5	1276.3	2450.3	190.7
1990	2009.9	1490.9	2635.1	1991	1235.9	2464.8	184.3
1991	2146.5	1674.8	2745.5	2040	1252	2337	183.8
1992	2041.1	1640.4	2967.7	2659.7	1168.2	2966.1	183.7
1993	2160.7	1465.7	2906.4	2662.5	1099.2	3144.1	204.4
1994	2130.5	1326.2	2450.7	2452	1041.8	2936	198.1
1995	2131.9	1321	2374.6	2392.1	987.6	2862	

Source: XXXXX

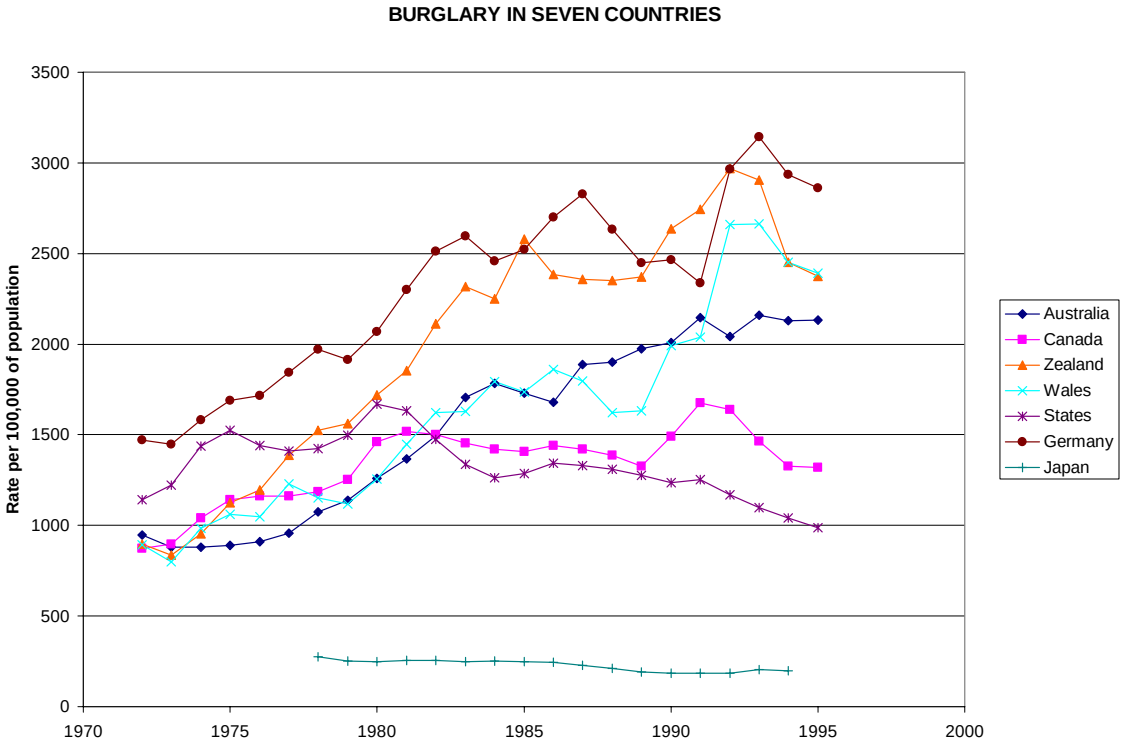
Now on the face of it, these crime rates differ. At this point we do not have to get into sophisticated definitions of what do we mean by differ although you can see that this might be an issue in a statistical sense.

The problem that I have with the table is that while it may illustrate the point, there is a lot of information in the table. Perhaps there is too much information. It is hard to see whether countries differ by very much. In this case, perhaps a picture would help:



For me, this is a much better way to represent the data. Why? From the picture it is much easier to see "what distinguishes the countries". Notice that Germany is generally higher than the other countries. It is easy to see that Japan is much lower. We can also more easily observe where other countries fit in. But this is not such a good picture. The axes are not properly labeled. The years are pretty clear, but we should have an indicator on the vertical axis that tells us what the units of measurement are.

Consequently, a better figure would look like:

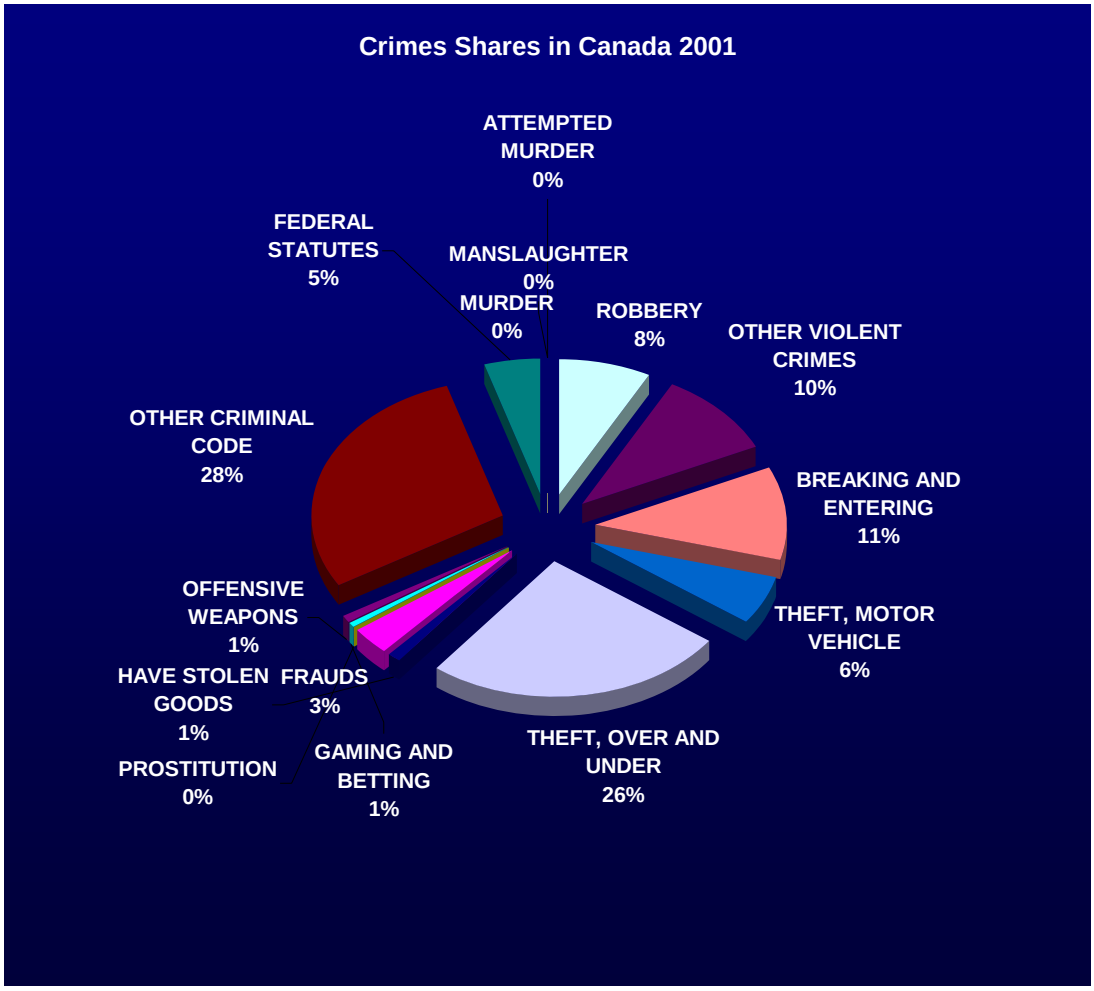

BURGLARY IN SEVEN COUNTRIES
Rate per 100,000 of population
3500
3000
2500
2000
1500
1000
500
0
1970
1975
1980
1985
1990
1995
2000
Australia
Canada
Zealand
Wales
States
Germany
Japan

What else can we do to characterize the information in the table? One of the things that we can do is to characterize the averages, again in a table

Burglary per 100,000 of population: 1972-1995

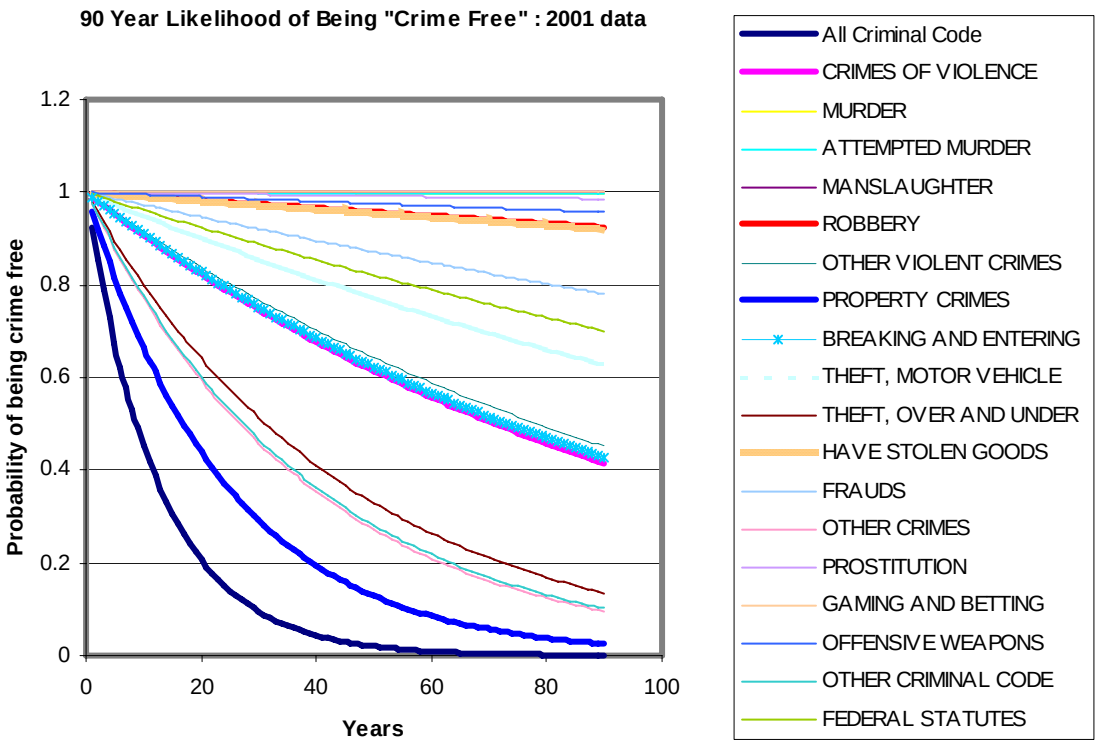
	Australia	Canada	New Zealand	England Wales	United States	Germany	*73-94 Japan
average	1545	1332	1994	1620	1325	2309	227
high	2161	1675	2968	2663	1668	3144	274
low	880	875	838	800	988	1448	184
trend	0.057	0.023	0.075	0.076	-0.006	0.043	-0.017

There are many other ways to characterize information in pictures and graphs. For example we may want to represent all the crime that takes place in Canada for a particular year in a pie graph:



This leads to a pretty clear representation of the information and gives a sense of which crimes are relatively important in terms of numbers without seeing the actual values. Unlike previous data it represents a cross-section at a moment in time whereas the earlier figures referred to time series: values over time.

Suppose we know the total number of crimes and the population. We might be interested in knowing how likely it is that an "average" person will be crime free over his or her lifetime. If there are 30 million Canadians and 3 million crimes per year, then a person has roughly a one in ten chance of being "exposed" to a crime each year. This means that you have a nine in ten chance of not being a victim each year. In two years, you have a probability of not being victimized of (9/10)x(9/10)=81/100 chance of not being victimized. In three years (9/10)x(9/10)(9/10), etc. Or, in a figure:



Lifetime Exposure to Crime

Your homework:

Choose some data that you believe bears on your topic and devise some graphs or tables to represent it. What implicit theory lies behind your presentation?