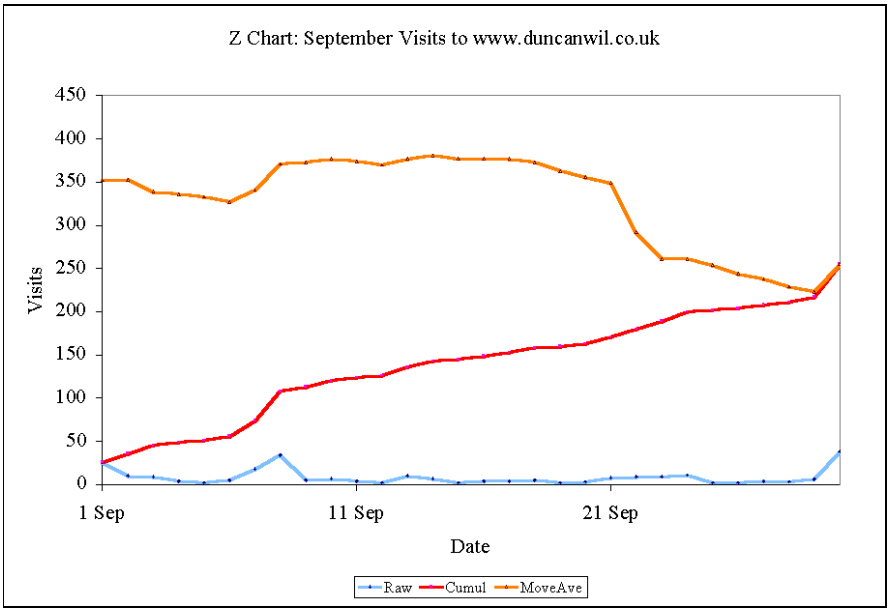
Different Charts: Same Data

A series of charts based on the same data: the number of visitors to www.duncanwil.co.uk for the period 1 August – 31 October 2001



This is the third set of graphs I have presented and it will probably be my last for quite a while!

© Duncan Williamson

4 November 2001

The aim of this paper is simply to demonstrate how a simple data series can be represented in a number of ways. The data series comprises the visitor statistics to http://www.duncanwil.co.uk for the period 1 August – 31 October 2001.

The following charts are here

- scattergraph
- bar chart
- histogram: two versions of these
- ogive
- radar chart
- doughnut chart
- logarithmic chart
- output gap type chart
- Z chart

In the **Appendix** you will find the raw data and the descriptive statistics that Excel offers as part of its Data Analysis Tools

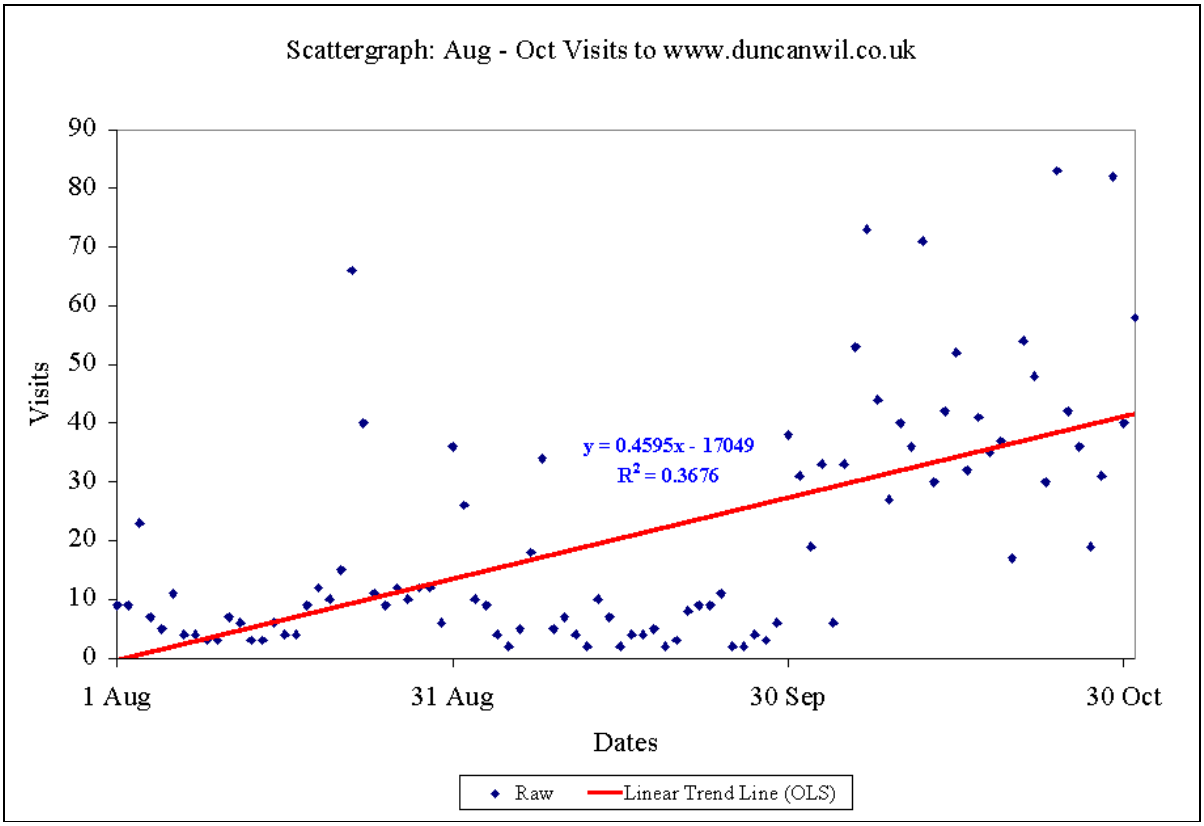
Acknowledgement: Thanks to Steven Wheat, Chris Rodda, Phil Bastow who all pointed me in the right direction of how Z Charts are constructed ; and a big thanks to Steven who also sent me some specimen data for me to relearn all about Z charts.

All of the graphs in this document were prepared with the help of Microsoft Excel 2000.

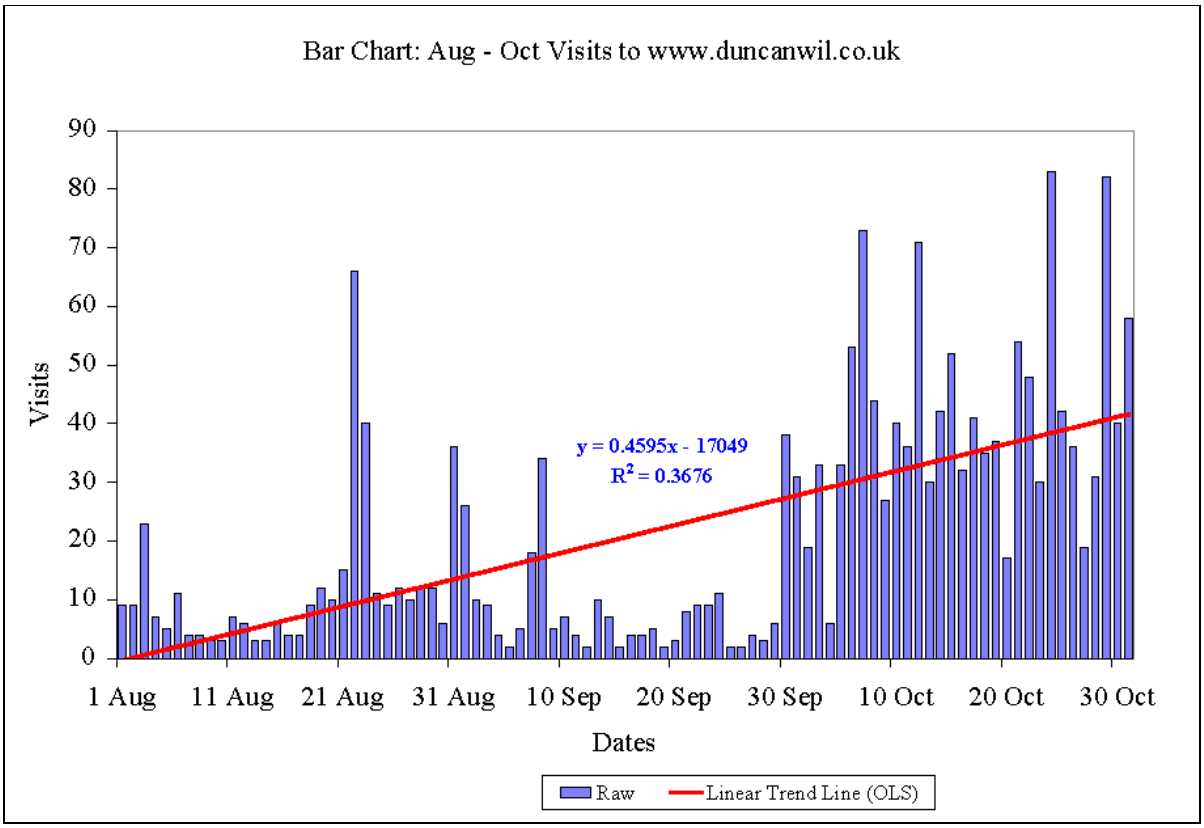
This paper can be downloaded free of charge from

http://www.duncanwil.co.uk/pdfs/charts.pdf

The Scattergraph: perhaps the simplest form of all graphs. This scattergraph consists of two axes, the X and Y (Dates and Visitors respectively) and the data series. I have added a trend line (based on ordinary least squares and put there by using Excel's Add Trendline utility). Note that the r^2 value is very low, at 0.3676, meaning that the trend line equation is unreliable and cannot be used to predict with accuracy the number of visitors in the future. The standard deviation is high, too, at around 20.2.



The Bar Chart: I think this chart is more useful than the scattergraph because it provides a bar for each day in the series and it help to focus us in on the fact that it is a day be day series. The scattergraph is less focused in this respect. I added a trend line here too and, no surprise, it's exactly the same as for the scattergraph.

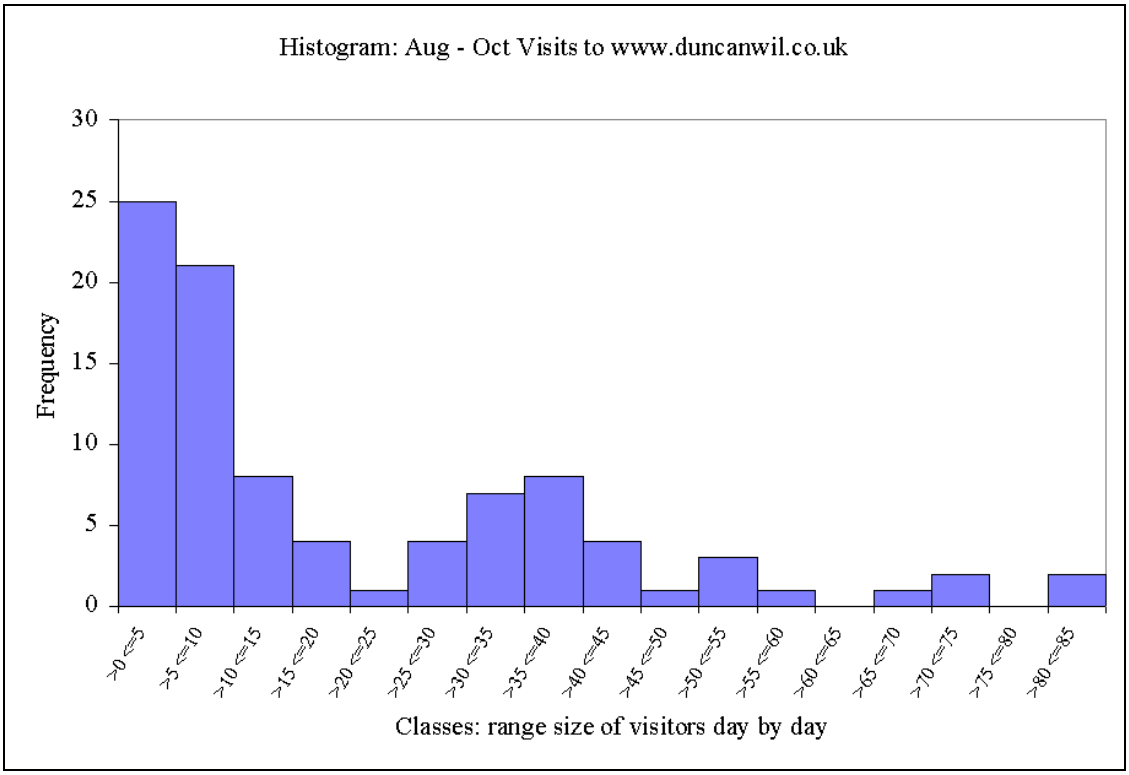


Histogram: a histogram should demonstrate relative proportions. Histogram 1 below is based solely on the raw data: sorted into various classes. This is the version that Microsoft Excel presents us with automatically when we use its Tools/Data Analysis/Histogram utility to sort the data for us into the various classes we choose for it.

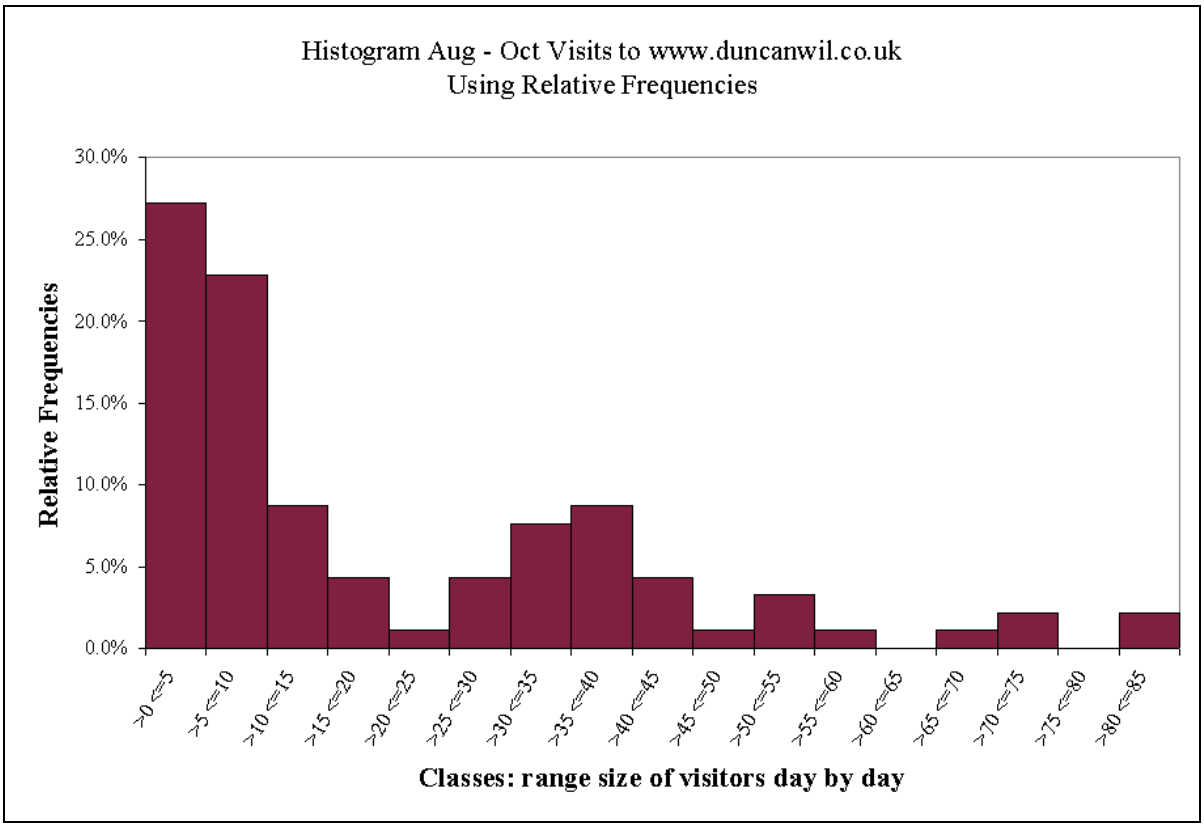
Technically speaking, the area under a histogram should be 1. That is, 1 or 100% or unity: see the diagram Histogram 2 below where that is exactly what happens. Excel doesn't really do that. Nevertheless, compare the two histogram diagrams and **spot the difference! There isn't a difference really, so Excel's not that bad after all, except for the purist.**

I find histograms very useful as they provide us with a detailed view of the likelihood that a given number of visitors have been to this site.

What the histogram loses, however, is the time series element of the data. Showing that on 25 days between 0 and 5 visitors arrived doesn't tell me whether this was at the beginning of the period, in the middle of the period or randomly distributed throughout the period.



Histogram 1 Excel's version using sorted raw data

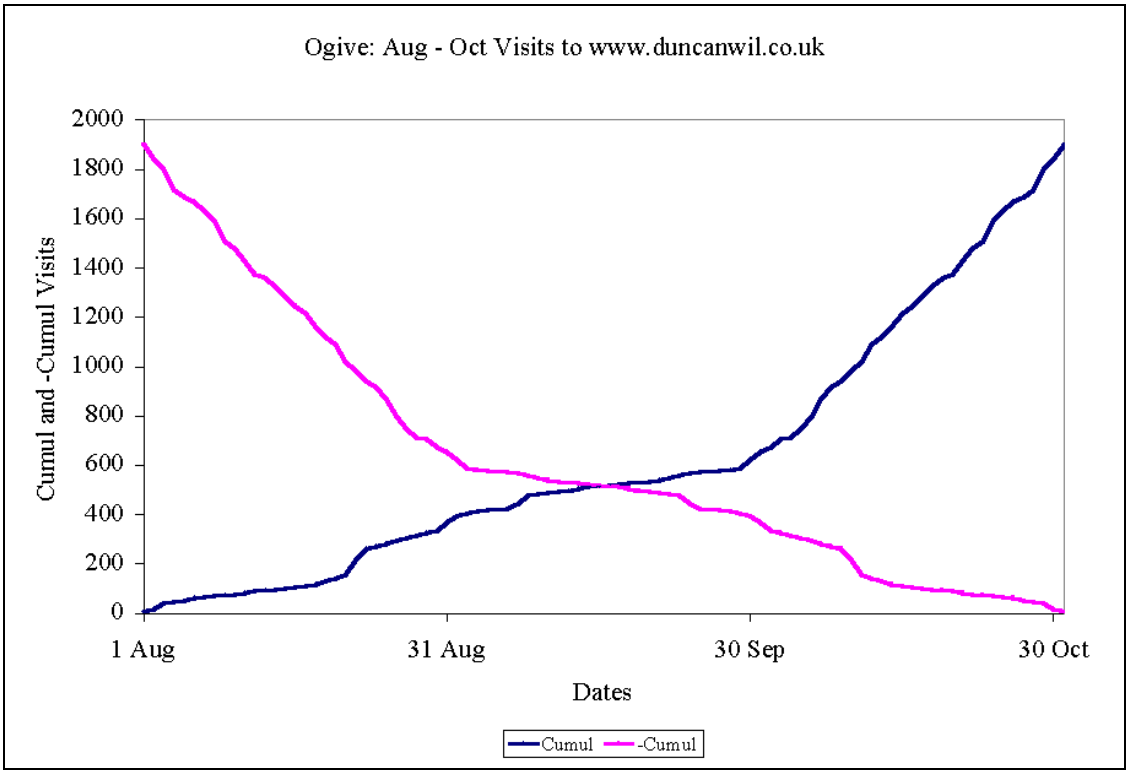


Histogram 2 Traditional version using relative frequencies

The Ogive: I have to confess that I hadn't drawn an Ogive for years and years until I started preparing this paper. An ogive is a greater than and less than chart. In this case, the first series shows the cumulative number of visitors starting from zero at the beginning of the period and ending up at 1,901 at the end of the period.

The second series shows exactly the same data but working backwards … from 1,901 to zero.

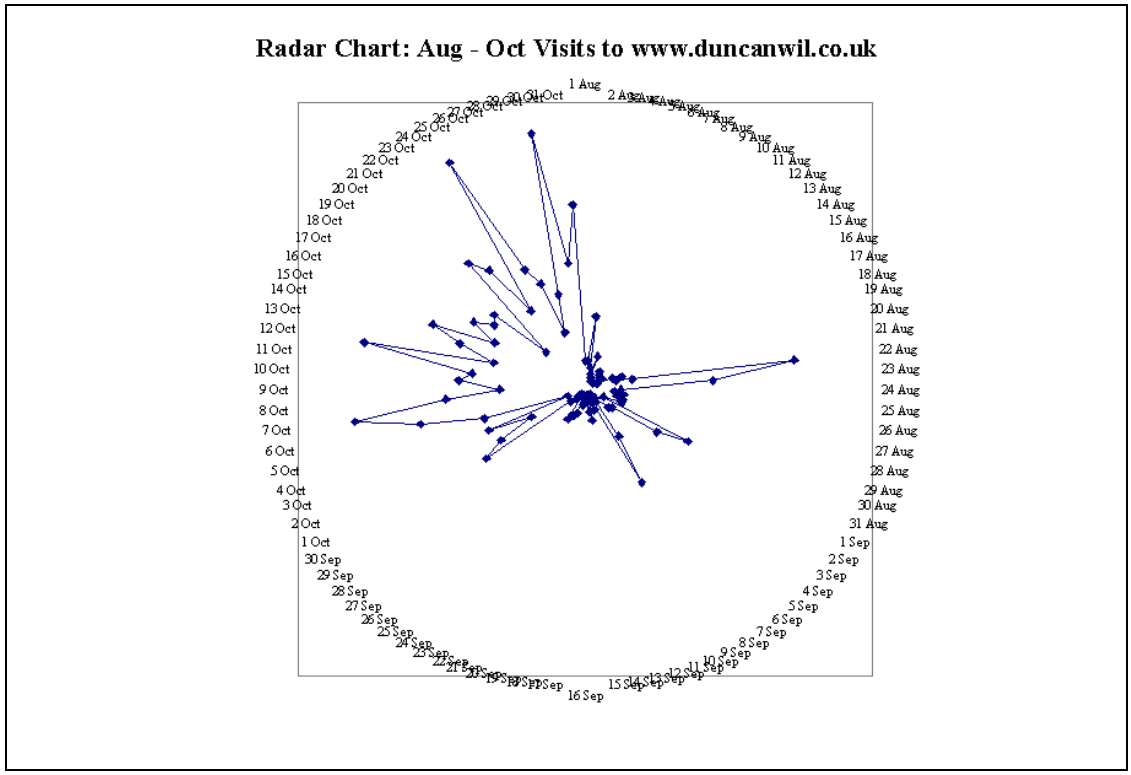
Why is it called an ogive? Apparently, an ogive is the shape of certain features found in medieval (?) church architecture. Let me know if that's wrong.



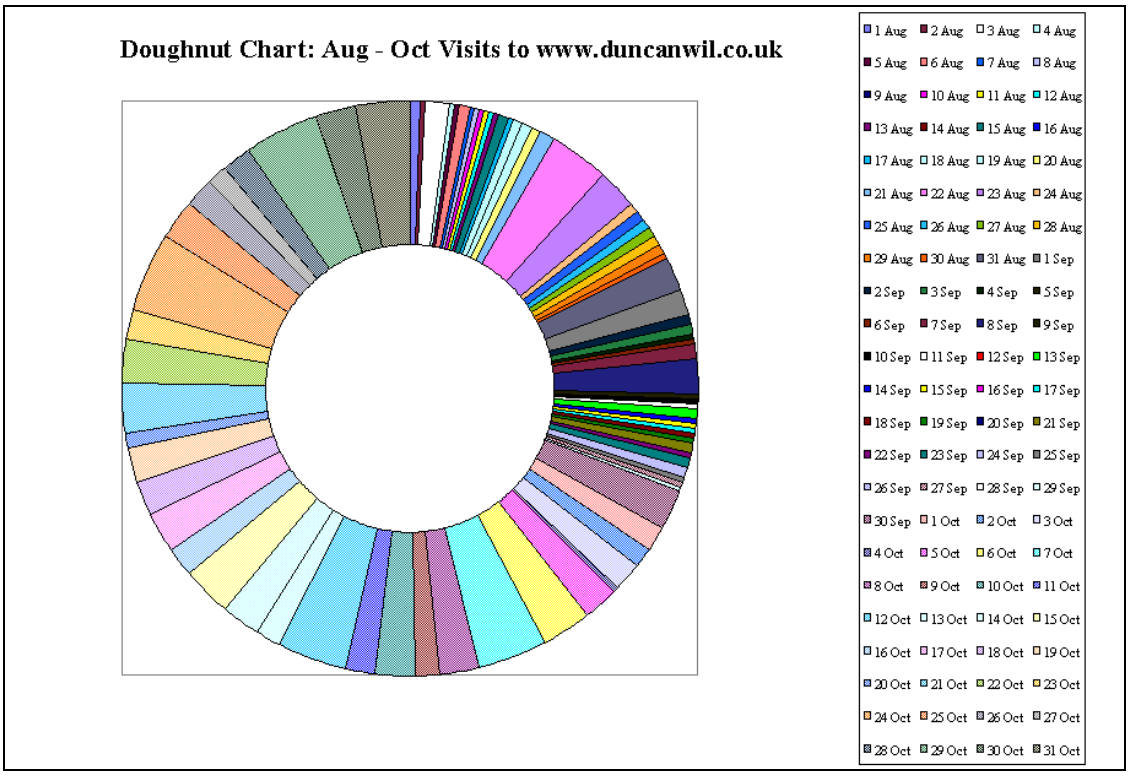
A Radar Chart: I didn't initially think of plotting a radar chart but having done so I think it is very effective. What the radar chart shows is the clustering effect of data … or lack of it. As we can see here, there is a major clustering of the data, and therefore the visitors, in October, with a blip near the end of August.

If this were a business site, we would need to follow through what happened at the clusters: are they random?; did they happen as a result of a marketing campaign?; is it really due to visitors or do the data contain other elements?

Note the dates on this chart: do you like them or do they get in the way? Well, there are lots of them but they are largely legible: on balance, though, I would prefer it if there were, say, one third of the dates shown rather than all of them.



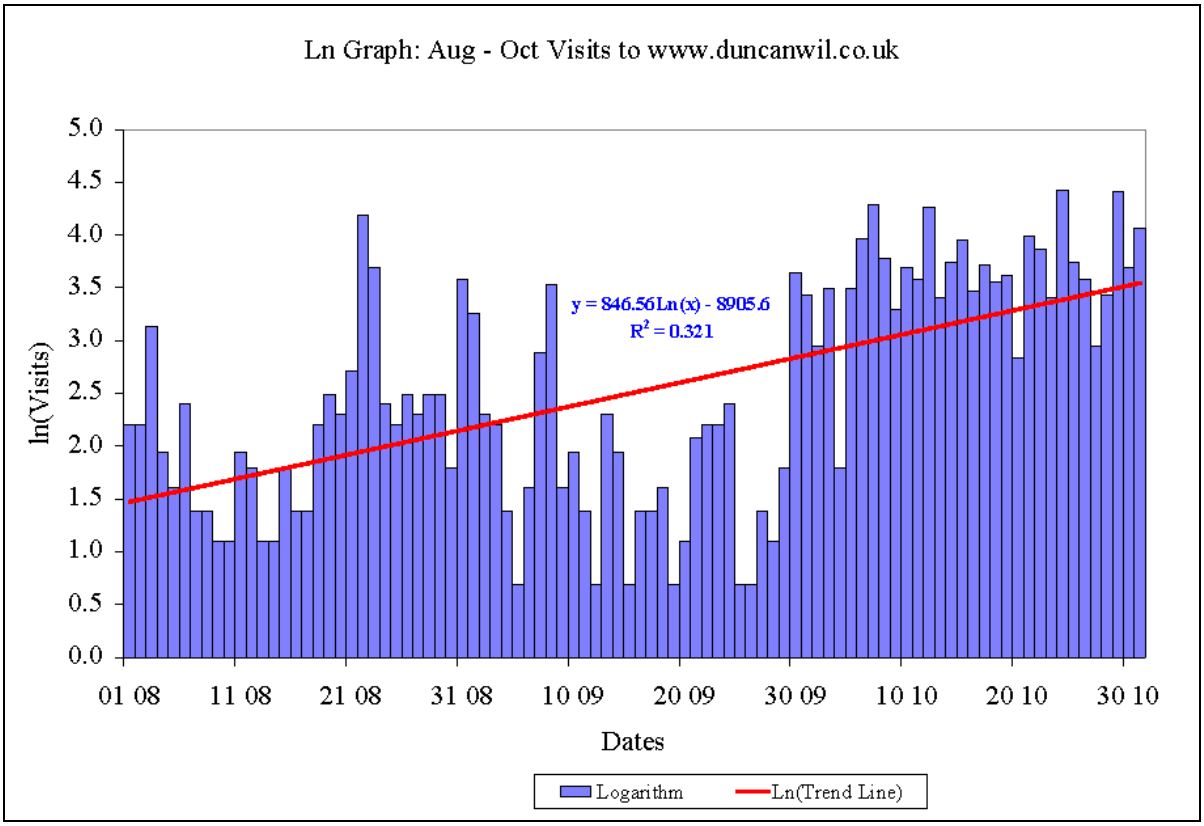
Doughnut Chart: isn't this awful? However, awful as it is, I included it to show that this kind of chart is not appropriate in the form shown here and for the data underlying it. Each slice of the doughnut is for one day … look at the Legend that has to accompany it!



Logarithmic Graph: log graphs can be very useful. This graph is based on Natural (Naperian) Logarithms. The essence of the log graph is that it can transform data from non linear to linear, making the analysis of those data much more straightforward.

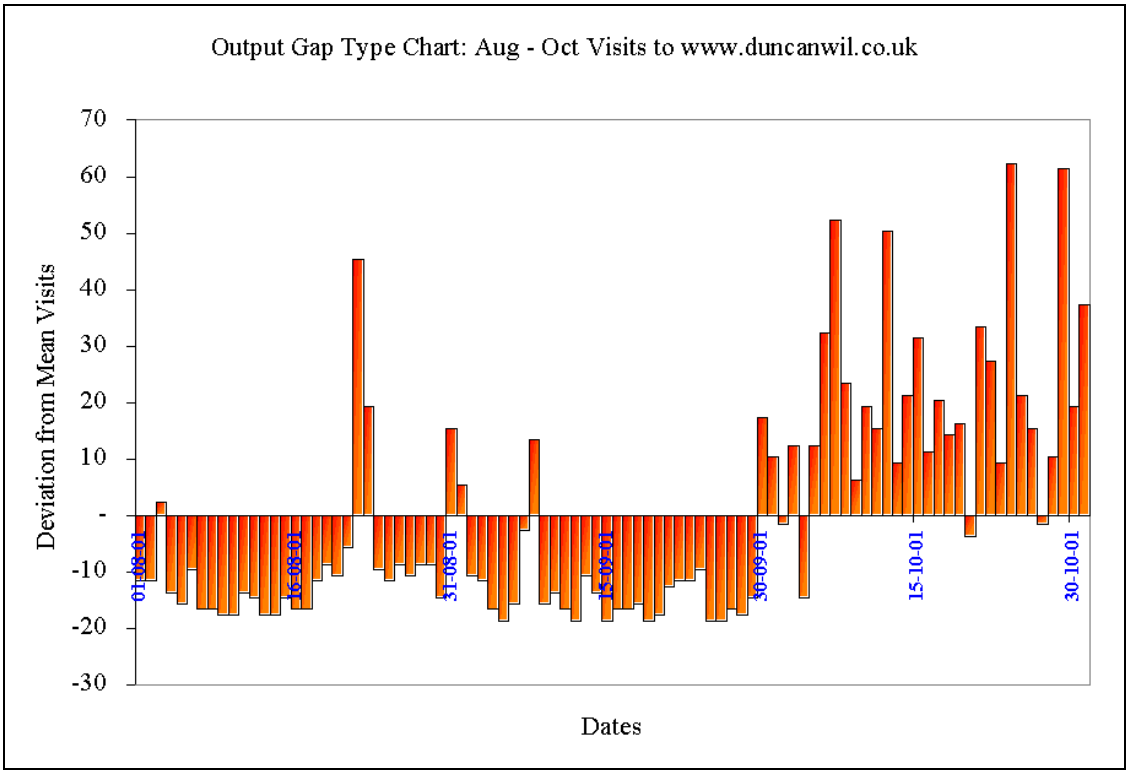
That is, if we transform non linear data by logs, plot them on a graph and carry out regression analysis, as we essentially have here, it might give us a function, or formula, that we can apply to the logs which, when transformed back into their base format, provides us with useful predictions.

An r^2 value of 0.321 is very low so even with logarithms, these data are not susceptible to the logarithmic transformation as we see here.



Output Gap Chart: I included this here following a discussion on an Economics/Business Studies List that I take part in. An economist asked about drawing an Output Gap chart and a small discussion followed. I thought it might be useful to demonstrate an Output Gap type chart with data other than National Income, or GDP, data.

All this chart shows is that the average is not a reliable indicator of any trend in these data since virtually every part of this analysis deviates significantly from it. Note the turning point, however, right at the end of September where, apart from the outliers found towards the end of August, the gaps almost all become positive having almost all been negative up to that point.

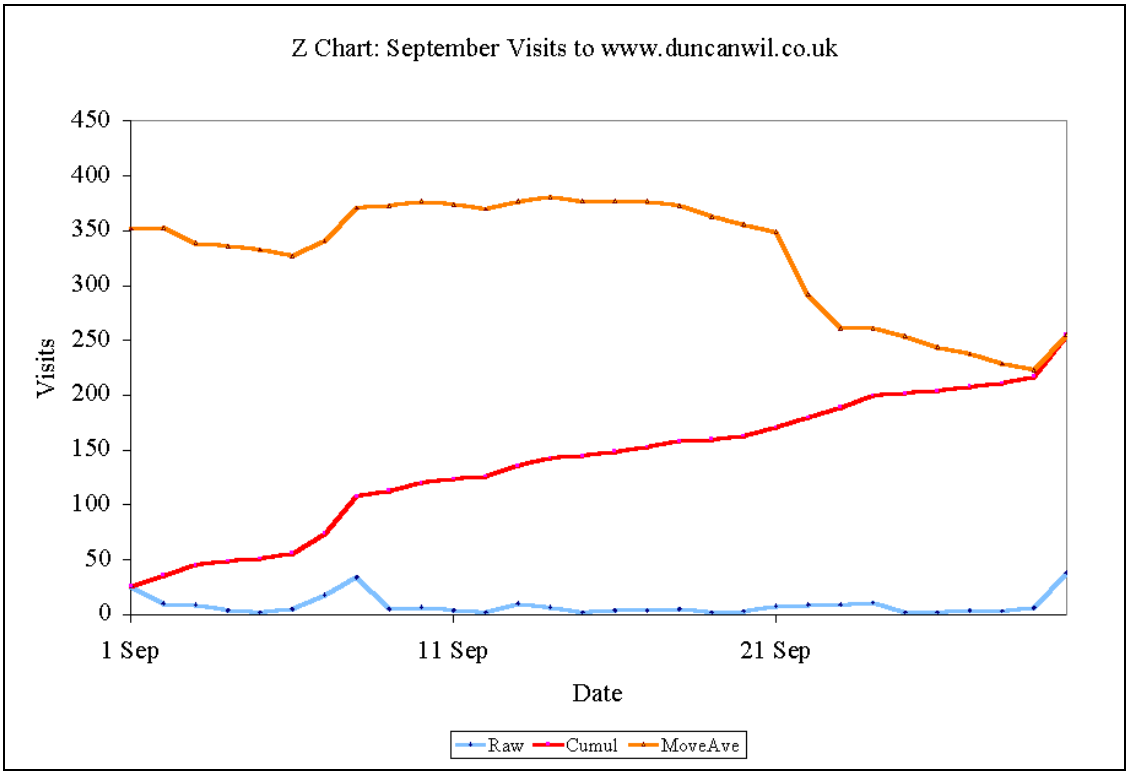


The Z Chart: I said in the Newsletter in which I announced this Z Chart that it was a special. The reason it's special is that I last drew a Z chart when I last drew an ogive, I think. Consequently, I could only remember how to derive two out of three of the curves: I couldn't remember what the third curve was; and trying to find a Z chart on the internet proved impossible.

The three curves that make up *any* Z chart are:

- raw data
- cumulative raw data (the same as we find in the ogive)
- moving average of the data (based on a whole month)

Take a look at the chart below and it should be obvious why Z charts have that name, although the Z chart for October from the data presented here isn't as *Z friendly* as the September version and the one based on the data that Steven Wheat sent me.



Thanks to Steven Wheat, Chris Rodda, Phil Bastow who all pointed me in the right direction and a big thanks to Steven who also sent me some specimen data for me to relearn all about Z charts.

Appendix: raw visitor data and descriptive statistics

Day	Visitors	Day	Visitors	Day	Visitors
01-Aug	9	01-Sep	26	01-Oct	31
02-Aug	9	02-Sep	10	02-Oct	19
03-Aug	23	03-Sep	9	03-Oct	33
04-Aug	7	04-Sep	4	04-Oct	6
05-Aug	5	05-Sep	2	05-Oct	33
06-Aug	11	06-Sep	5	06-Oct	53
07-Aug	4	07-Sep	18	07-Oct	73
08-Aug	4	08-Sep	34	08-Oct	44
09-Aug	3	09-Sep	5	09-Oct	27
10-Aug	3	10-Sep	7	10-Oct	40
11-Aug	7	11-Sep	4	11-Oct	36
12-Aug	6	12-Sep	2	12-Oct	71
13-Aug	3	13-Sep	10	13-Oct	30
14-Aug	3	14-Sep	7	14-Oct	42
15-Aug	6	15-Sep	2	15-Oct	52
16-Aug	4	16-Sep	4	16-Oct	32
17-Aug	4	17-Sep	4	17-Oct	41
18-Aug	9	18-Sep	5	18-Oct	35
19-Aug	12	19-Sep	2	19-Oct	37
20-Aug	10	20-Sep	3	20-Oct	17
21-Aug	15	21-Sep	8	21-Oct	54
22-Aug	66	22-Sep	9	22-Oct	48
23-Aug	40	23-Sep	9	23-Oct	30
24-Aug	11	24-Sep	11	24-Oct	83
25-Aug	9	25-Sep	2	25-Oct	42
26-Aug	12	26-Sep	2	26-Oct	36
27-Aug	10	27-Sep	4	27-Oct	19
28-Aug	12	28-Sep	3	28-Oct	31
29-Aug	12	29-Sep	6	29-Oct	82
30-Aug	6	30-Sep	38	30-Oct	40
31-Aug	36			31-Oct	58

Descriptive Stats: Visits	
Mean	20.663
Standard Error	2.1099
Median	10.5
Mode	4
Standard Deviation	20.2372
Sample Variance	409.5446
Kurtosis	0.949
Skewness	1.2605
Range	81
Minimum	2
Maximum	83
Sum	1,901
Count	92
5^{th} Largest	66
5^{th} Smallest	2