

Ruby and Scientific/Financial Data

© 2005-07 Protected by International Copyright law. All rights reserved worldwide.

Version: 16 October 2007

This document remains the property of Red Centre Software Pty Ltd and may only be used by explicitly authorised individuals who are responsible for its safe-keeping and return upon request.

No part of this document may be reproduced or distributed in any form or by any means - graphic, electronic, or mechanical, including, but not limited to, photocopying, recording, taping, email or information storage and retrieval systems - without the prior written permission of Red Centre Software Pty Ltd.

Ruby and Scientific/Financial Data

Ruby is a graphical cross tabulator, by which we mean data can be viewed, manipulated and analysed as either charts or tables with equal facility.

Though primarily a tool for the analysis of surveys, Ruby can also be used for any type of quantitative data, or any type of categorical data which can be represented as numeric codes.

This document covers some of the sorts of things which can be done using primarily quantitative data from four fields of study:

- A. Dendrology
- B. Climatology
- C. Paleoclimatology
- D. Financial Indices

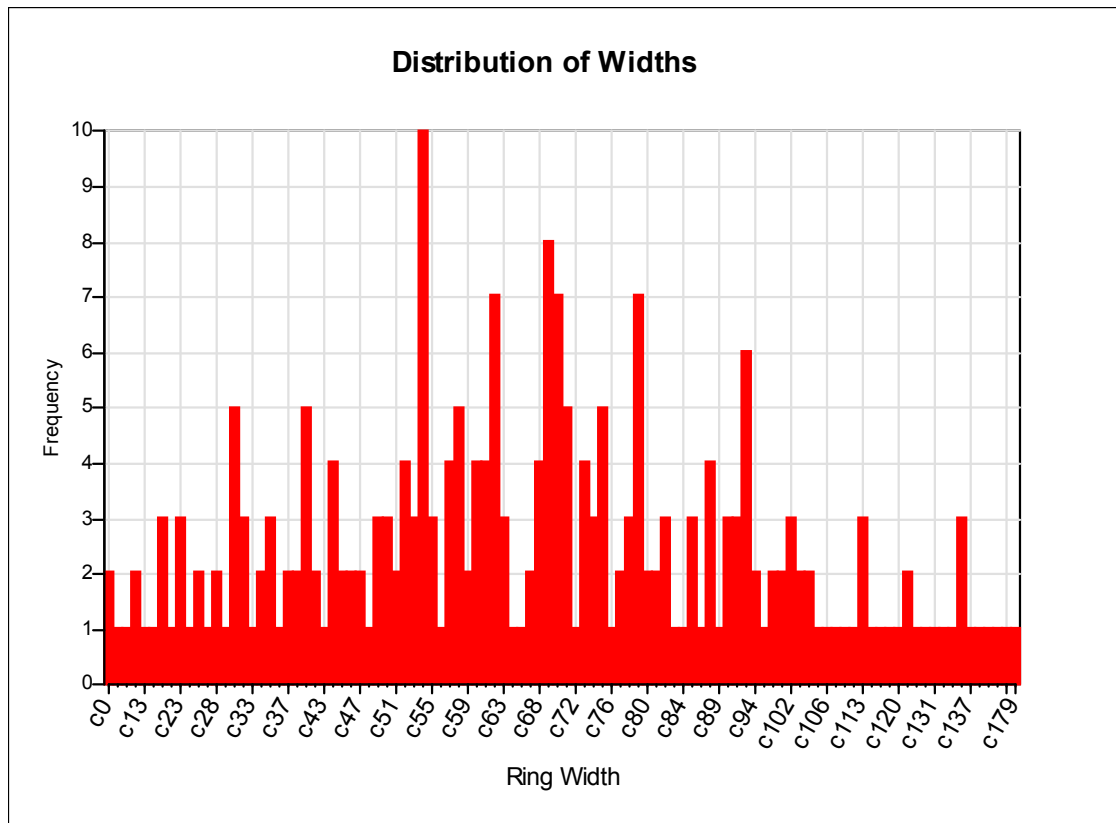
A. DENDROLOGY	3
B. CLIMATOLOGY	7
Melbourne/Sydney Comparison	7
Melbourne High/Low Differences	8
Hots Days in Melbourne	9
Melbourne Monthly Net Averages	10
Australia Wide Totals.....	10
Very Large Data Sets – Utah TMAX and TMIN.....	11
C. PALEOCLIMATOLOGY: ICE CORE CHARTS	14
D. FINANCIAL INDICES	16

A. DENDROLOGY

Here is a sequence of charts and tables showing several points of view on a single tree ring data set from the Franklin River, Tasmania. The set was obtained from

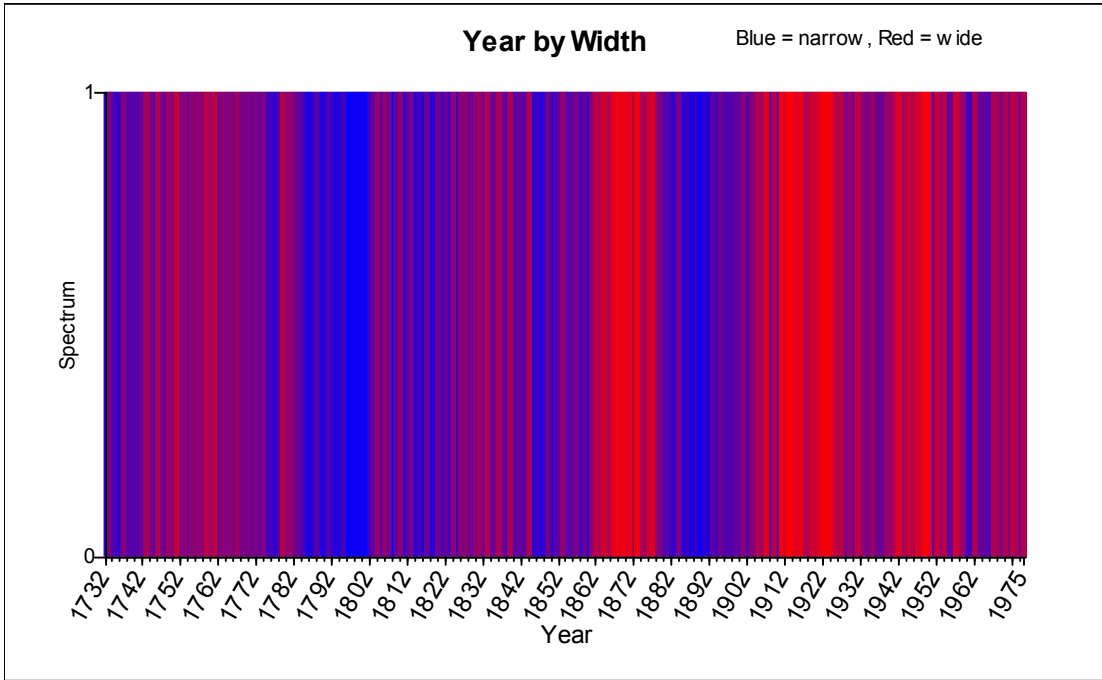
<ftp://ftp.ncdc.noaa.gov/pub/data/paleo/treering/measurements/australia/ausl013.rwl>

series FRV021.



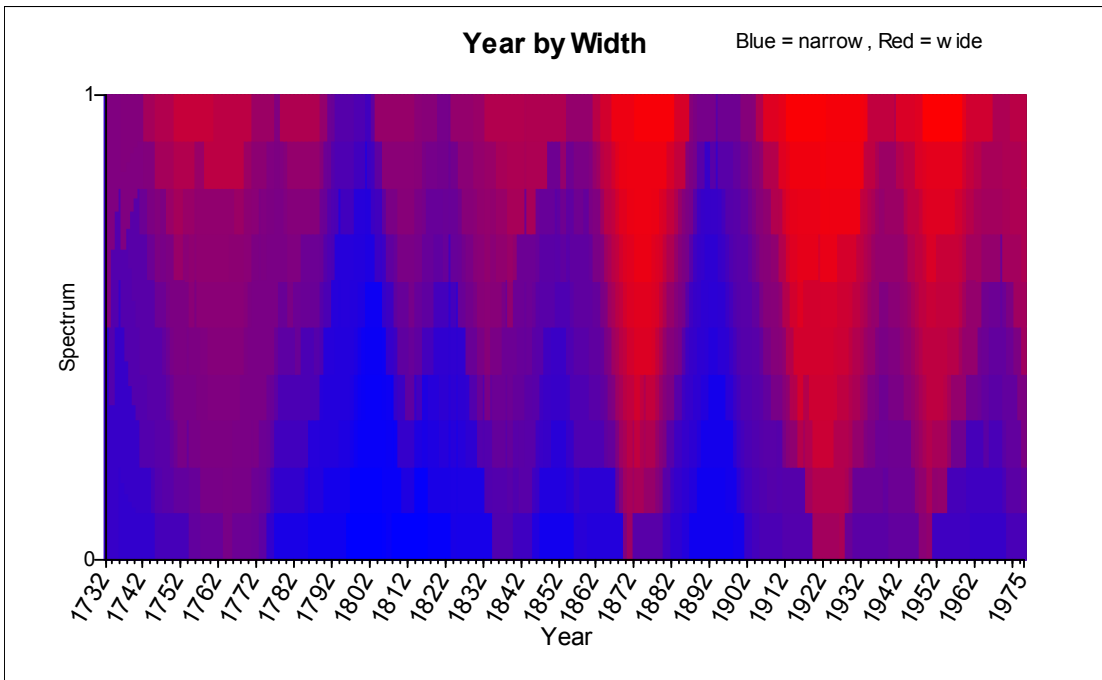
The widths are self-coded, indicated by the leading 'c' on the x axis labels. The chart tells us that there were two years with zero width (at the far left), the most common width is 54 with ten occurrences, widths 137 to 179 occurred once, and so on.

Colour gradations can be applied to get spectral plots.

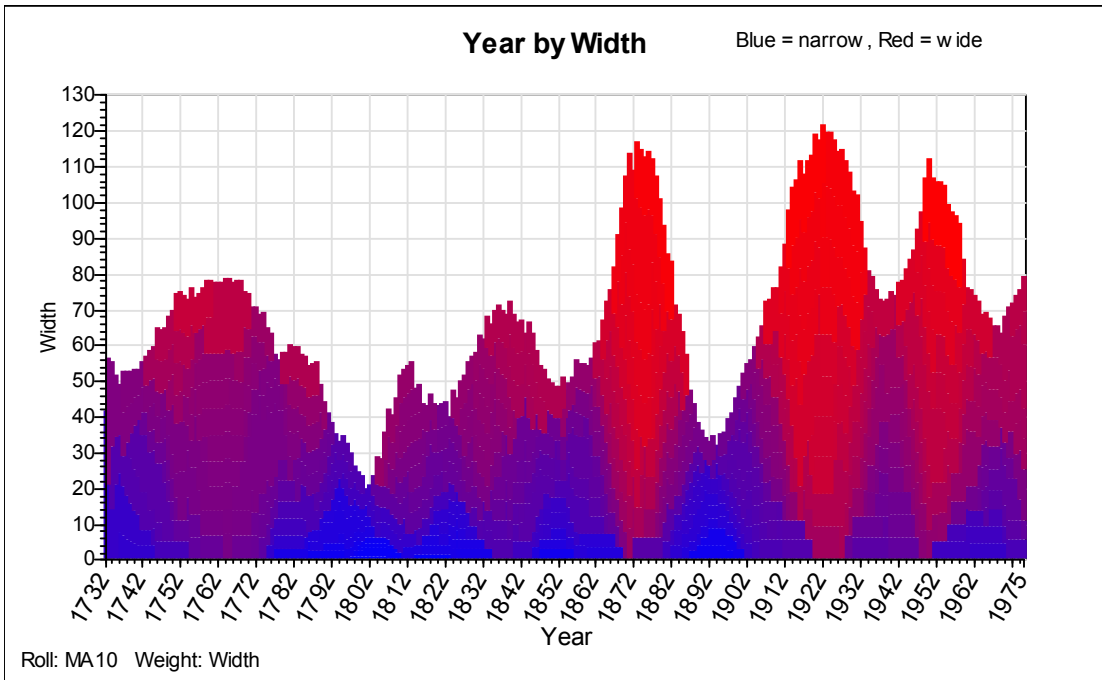


So it looks like the 1790s and 1880s were adverse years, and the 1870s, 1910s and 1920s, and 1940s were prolific years.

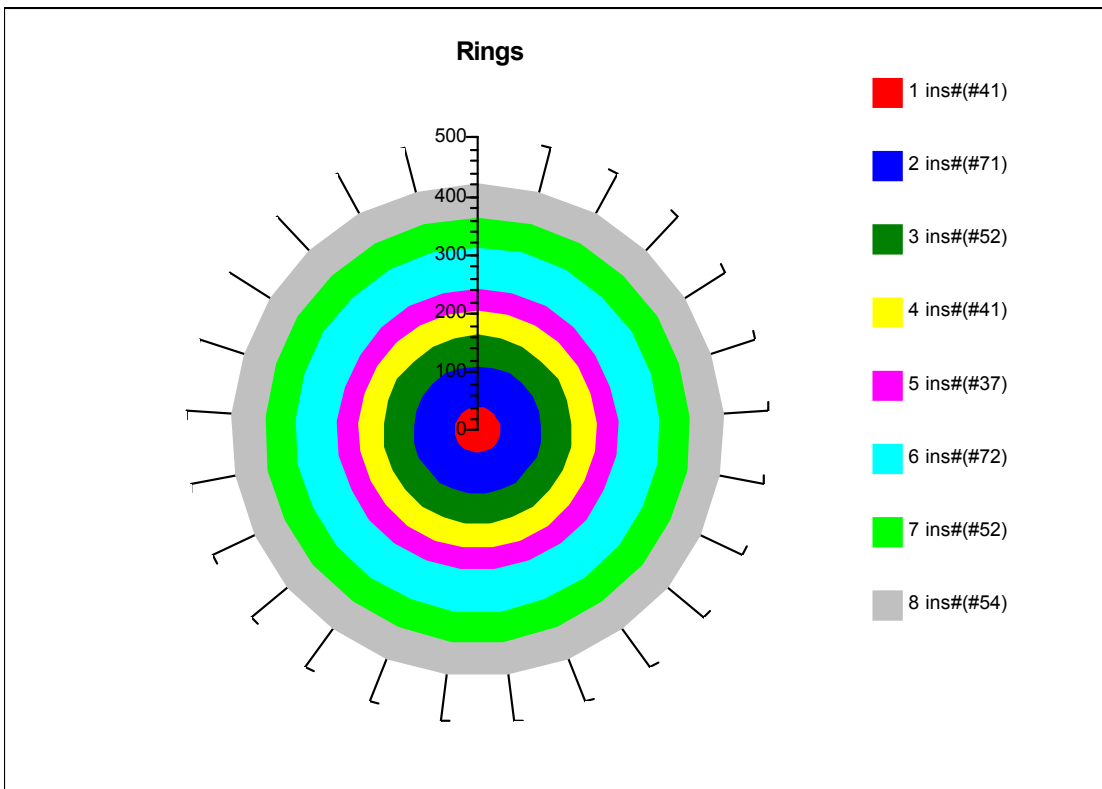
Rolling (apply a moving average) can make the patterns clearer. At MA10:



To indicate the actual widths as well, the chart can be weighted by width.



The obvious thing is to reproduce the rings themselves as concentric circles. Inserting each width as a series, and showing the first 8 only, the chart is



Ruby can plot thousands of series, so the number of rings presents no difficulty.

Which years had the greatest growth? Which years the least? A sorted table answers both questions.

Top: Count
 Side: Year
 Weight: Width (All) SideSort: col 1 decreasing

Frequencies		Count
		c1
Year	1949	179
	1913	163
	1873	162
	1922	157
	1923	143
	1914	137
	1867	134
	1871	134
	1924	134
	1869	133
	1916	132

Top: Count
 Side: Year
 Weight: Width (All) SideSort: col 1 increasing

Frequencies		Count
		c1
Year	1796	
	1808	
	1797	4
	1814	7
	1799	12
	1801	12
	1800	13
	1798	16
	1786	17
	1887	17
	1889	17

B. CLIMATOLOGY

The following charts examine some aspects of Australian temperature data.

El Nino years are from

http://conserwater.melbournewater.com.au/content/conserved/el_nino_years.htm

Temperature data was downloaded from

<ftp://ftp.bom.gov.au/anon/home/bmrc/perm/climate/temperature/daily/>

Data is available from approximately 700 stations across Australia. The Daily set we used comprises the 133 stations with relatively large continuous slabs of observations.

Under \daily there are two subdirectories, \analysis and \qc, which contain the same file set. We have used the data in \qc.

Resolution is daily from 1Jan1859.

Most stations have data from the 1940s. Data ends at 30Apr1999 for many stations with recent data, giving 51,254 cases.

In addition, some cases have -999 for both max and min, and so are effectively missing. Many other files have a lot of missing data.

Each data file has the same structure. The first few lines of Sydney.txt are

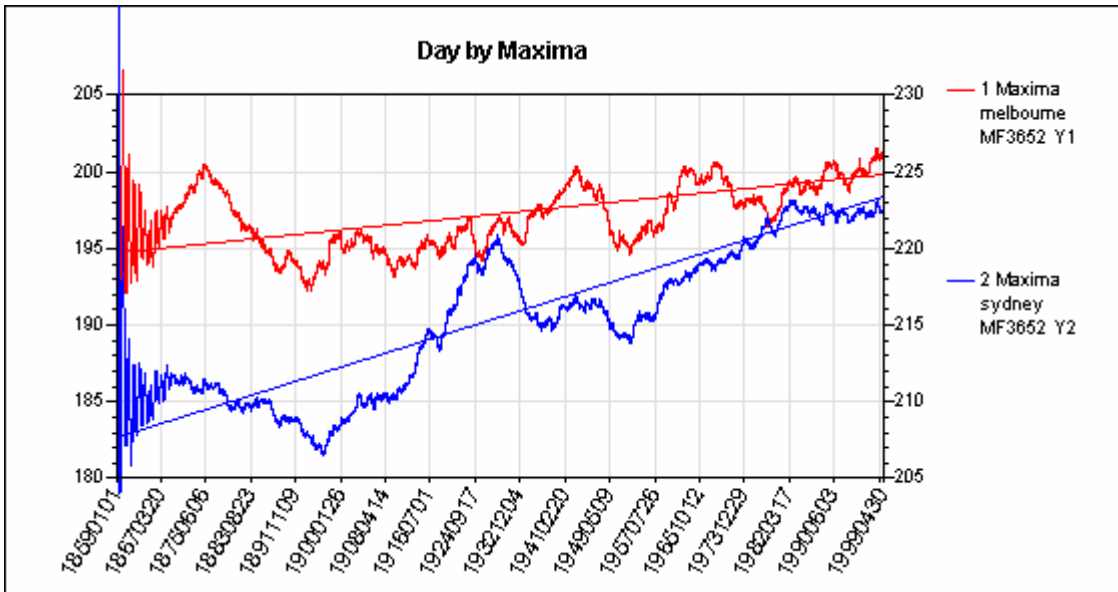
```
06606218590101 244 145
06606218590102 244 157
06606218590103 242 153
```

where 066062 is the station ID (columns 1/6), 18590101 is the date in YYYYMMDD format, 244 is the maximum temperature in tenths of degrees, and 145 is the minimum. So the first line says "in Sydney in 1Jan1859 the minimum was 14.5 and the maximum was 24.4".

These files were converted to Ruby format using an automation script to align all the data with respect to true calendar periods.

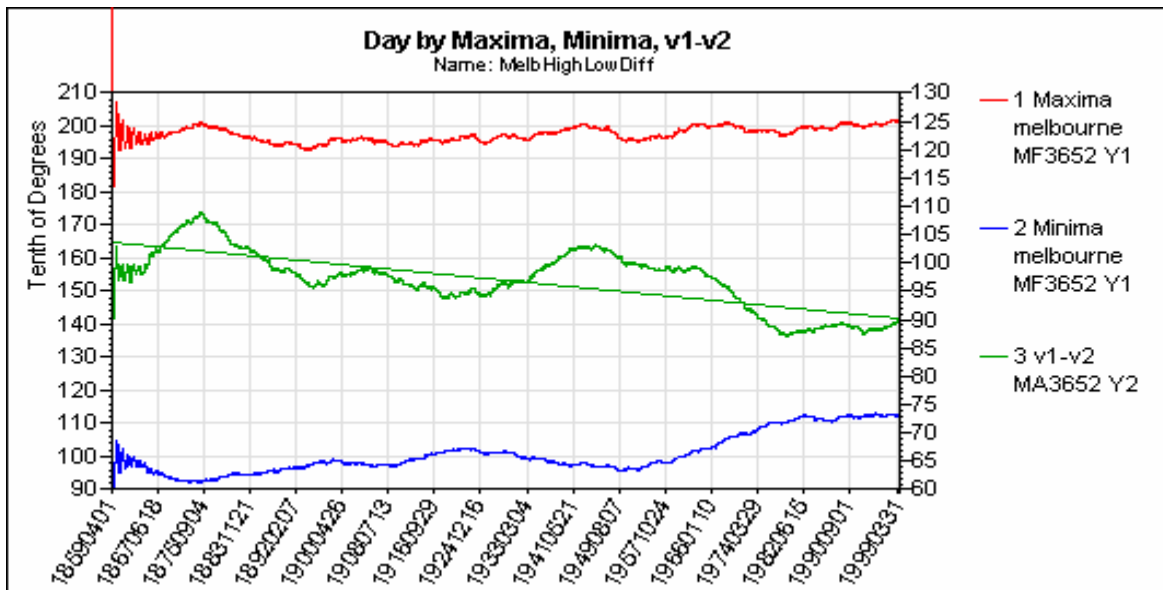
Melbourne/Sydney Comparison

This chart shows the daily maxima from 1859 to 1999 on a ten year roll, that is, the moving average period is 3652 days. Ruby can apply a moving average of any length virtually instantly. Manipulations performed on a series and all necessary information for interpretation can be automatically incorporated into the legend text.



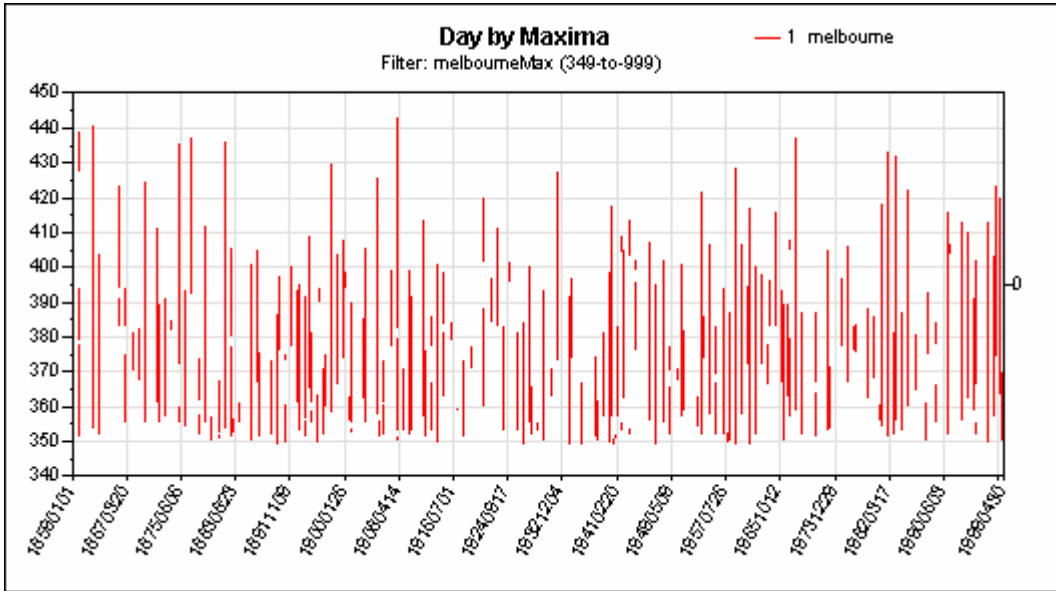
Melbourne High/Low Differences

Ruby can do most of the sorts of things for which you would usually turn to a spreadsheet application. Here, the calculation is simply maxima minus minima, and it shows that the gap is decreasing.

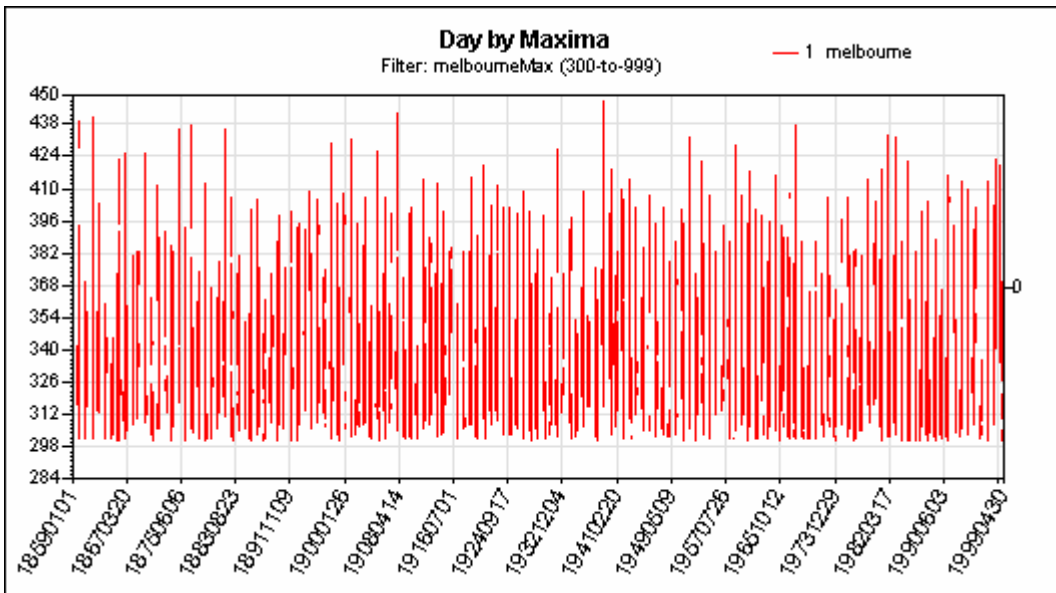


Hots Days in Melbourne

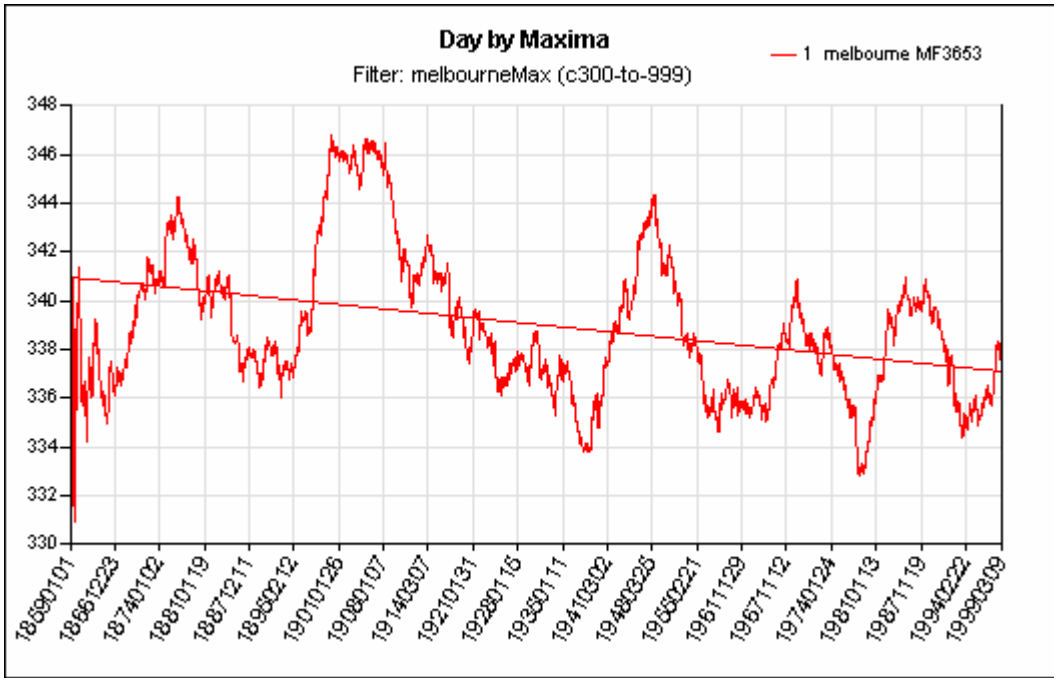
Data can be interactively filtered in many ways. Each vertical line is a string of days over 35C.



Melbourne Days Over 30 shows a denser distribution, though no obvious pattern yet.

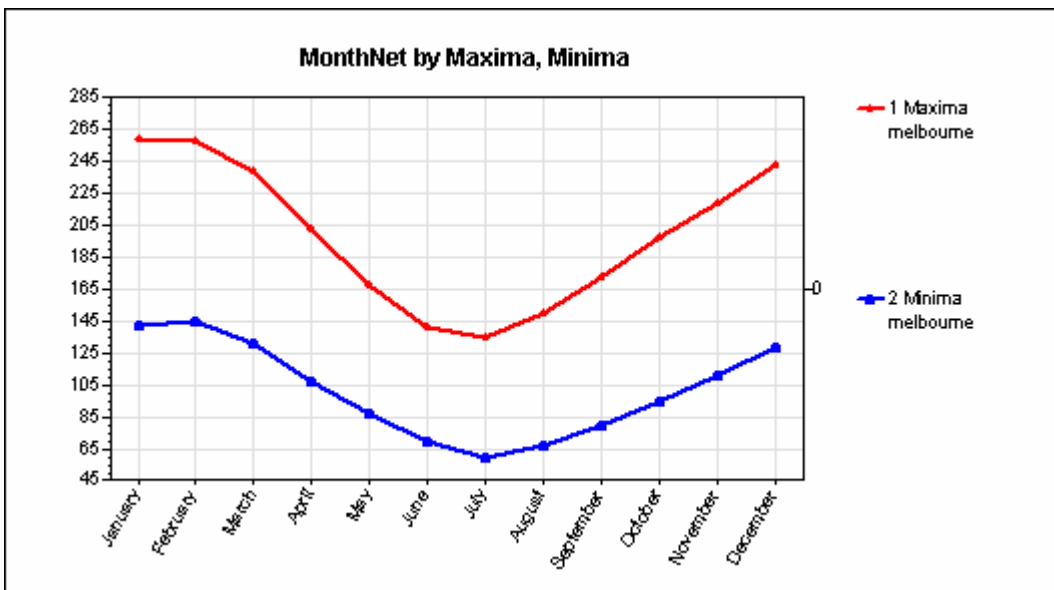


Closing the gaps, rolling at MA3652 and applying a trend line shows a decrease.



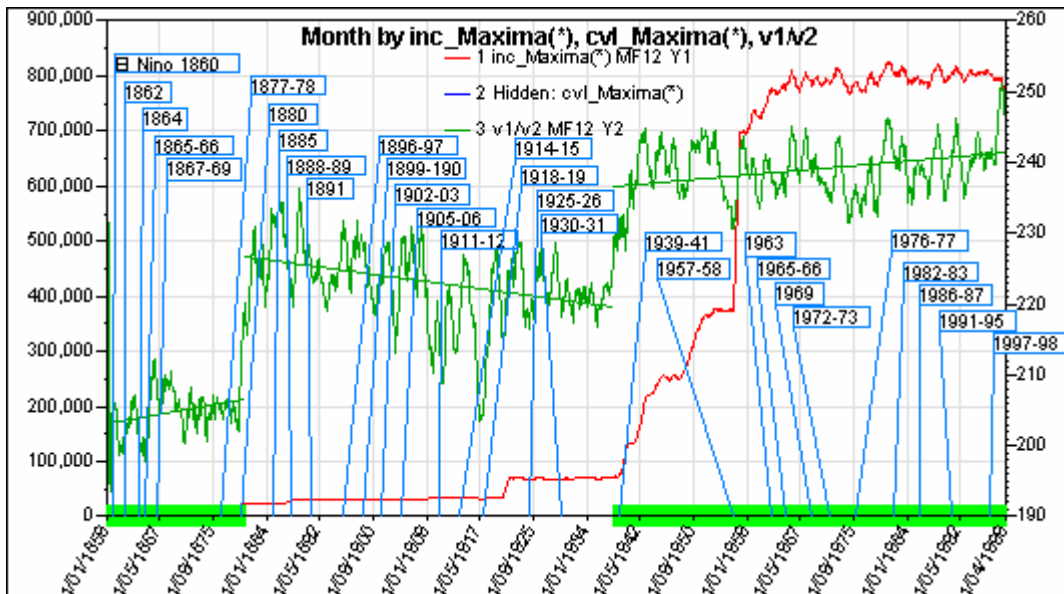
Melbourne Monthly Net Averages

Data can be netted in many different ways to provided aggregates and overviews.



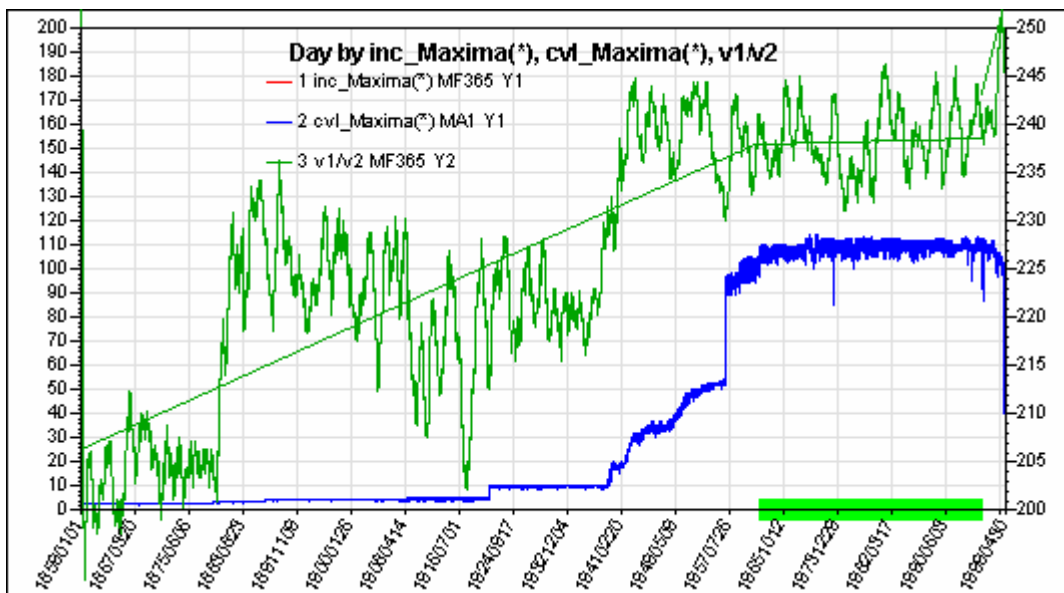
Australia Wide Totals

This chart at monthly resolution shows the total aggregate maxima across all stations on Y1, and the aggregate divided by the number of non-missing observations on Y2. The basic message is that the average temperature in Australia is 24C. El Nino events are also displayed. The El Nino events are stored in Ruby, and can be shown against any period axis.



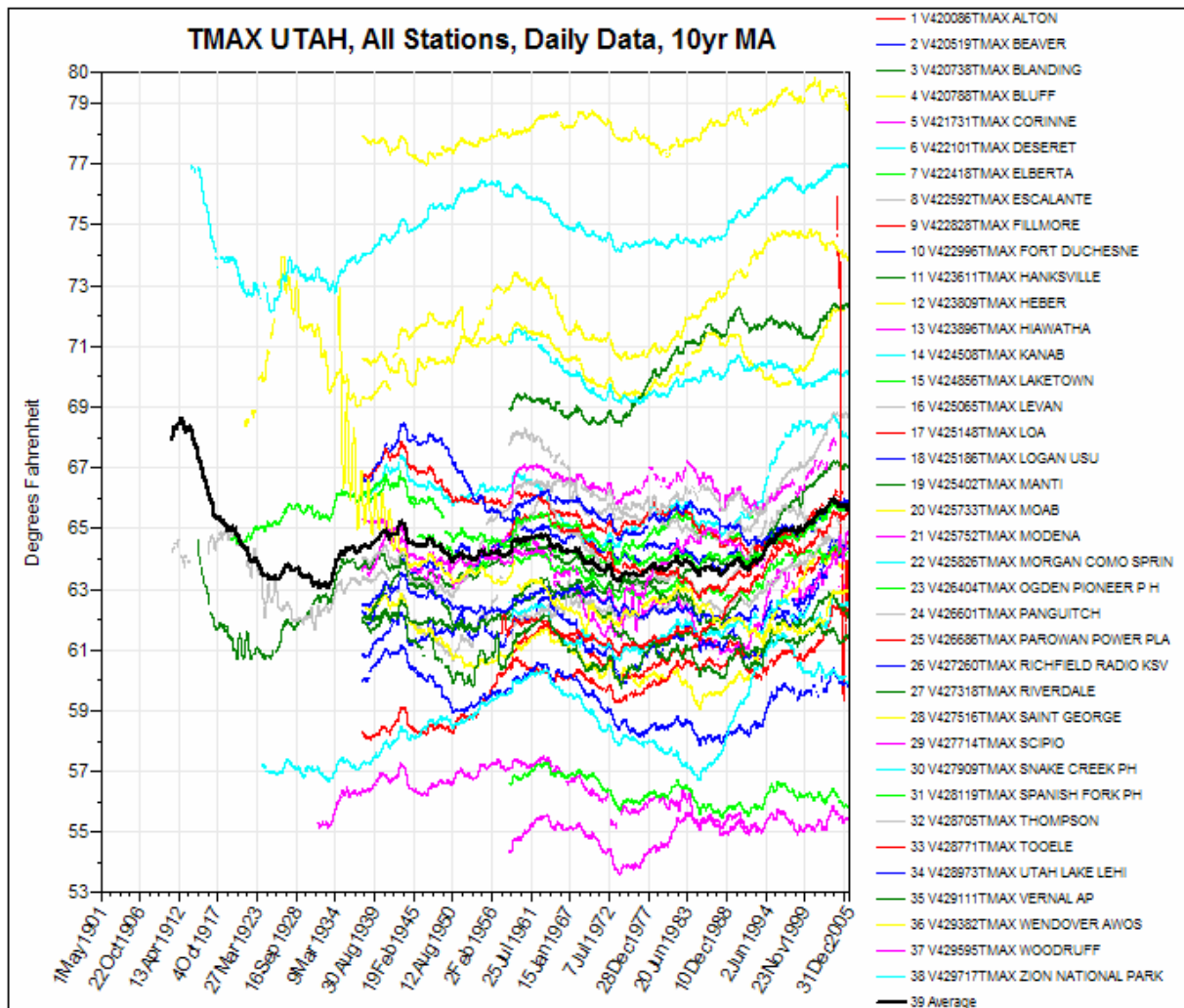
Note the piece-wise trending. The x axis selection determines where the trend lines start/end, and these can be adjusted interactively.

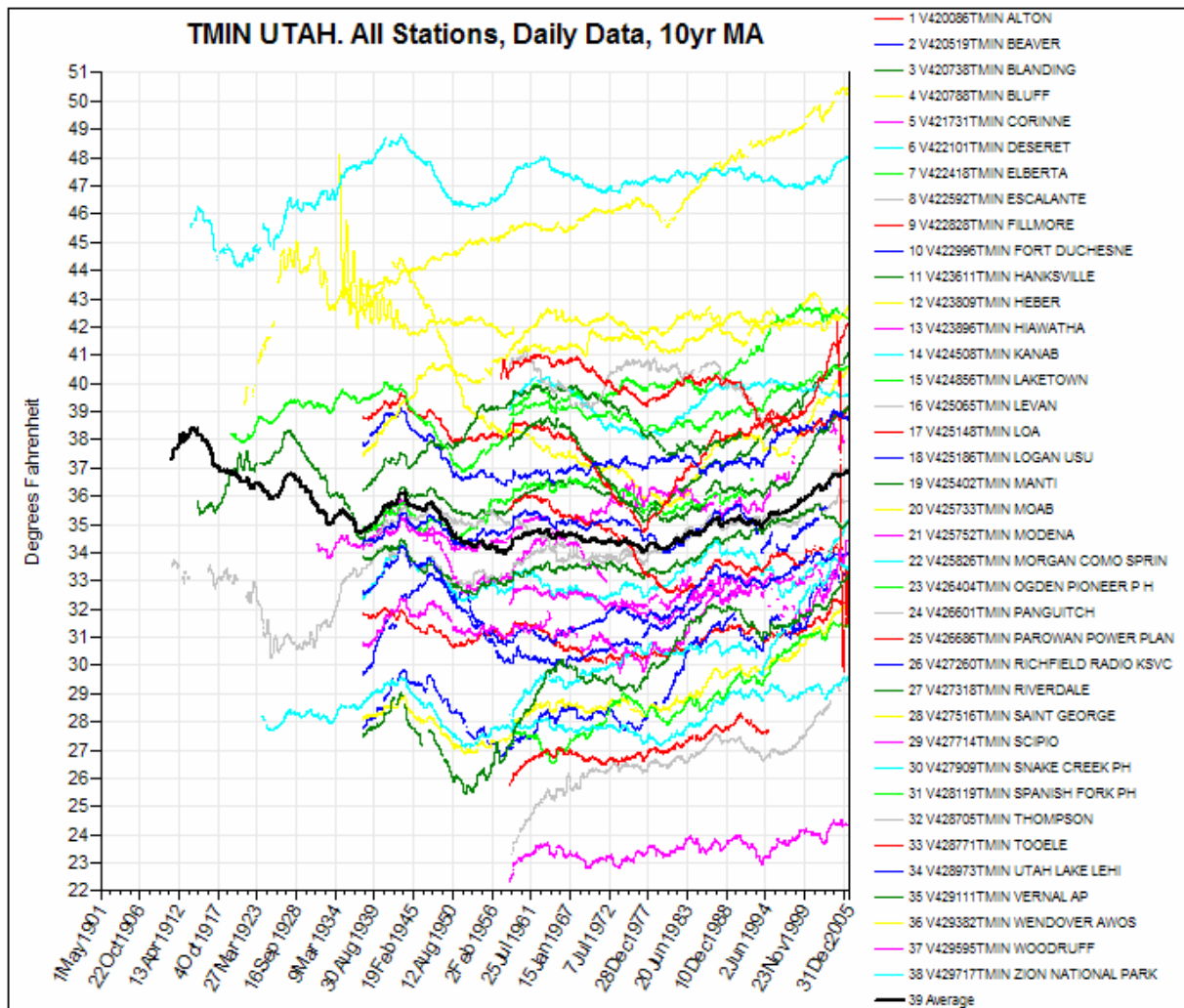
The next chart is as above, but at daily resolution, and additionally shows the number of stations reporting an observation for each day. Note that the trend line, when adjusted for the period of maximum station coverage, is virtually flat. The strong peak at the end of the Averages series is due to the sudden drop towards 1999 in the number of reporting stations, which biases the sample. The first series, the sum of all reported temperatures (in itself meaningless), is way off the Y1 scale, and cannot be seen unless scaled down.



Very Large Data Sets – Utah TMAX and TMIN

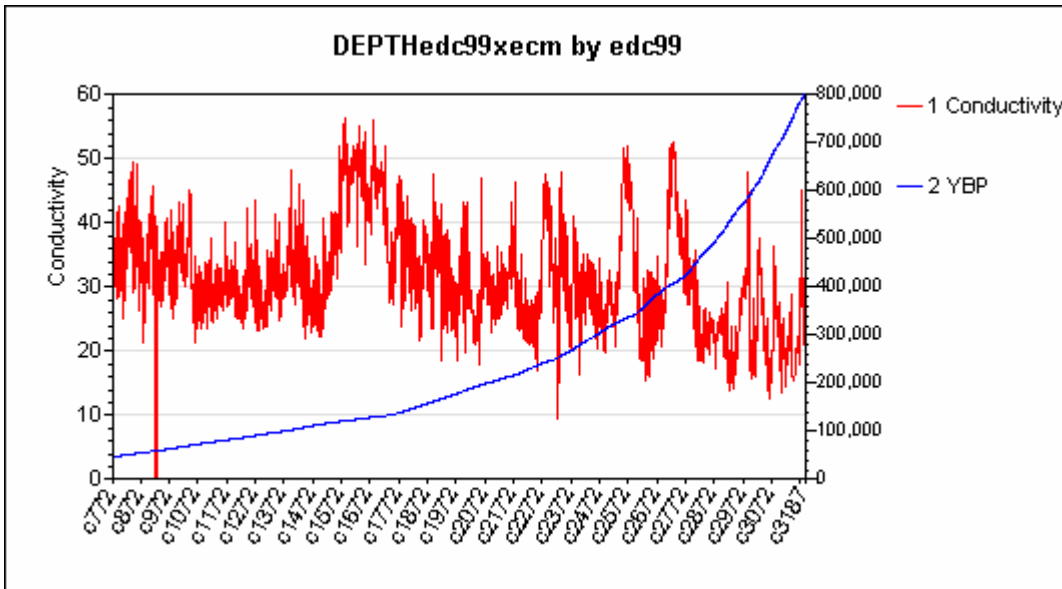
Ruby makes it easy to visually examine huge data sets. The following two charts comprise over 100 years of daily data, rolled at MA3652, making ten years. The average is calculated for all displayed series, and takes missing data into account.



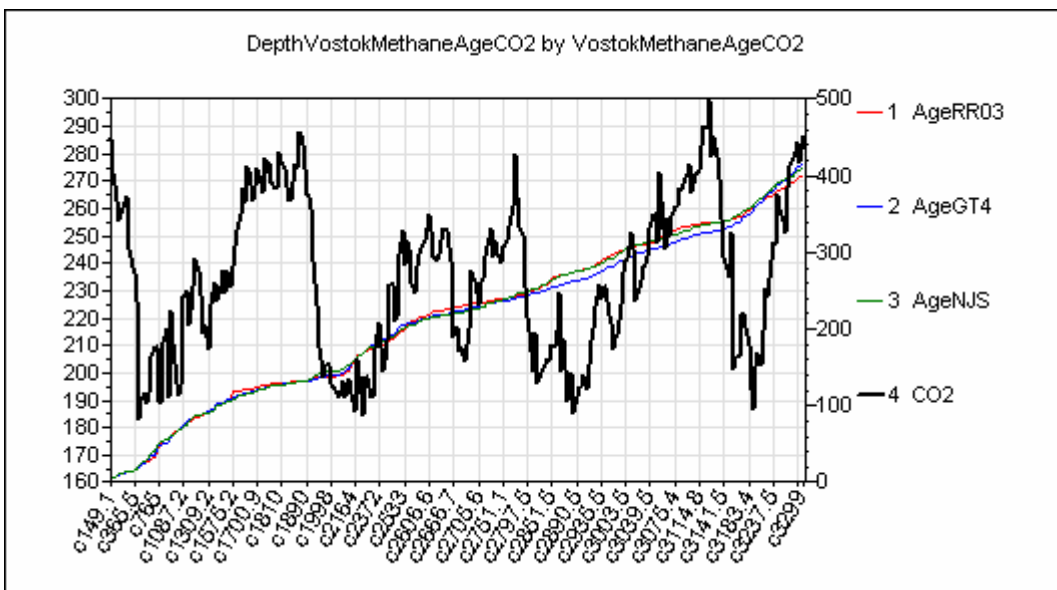


C. PALEOCLIMATOLOGY: ICE CORE CHARTS

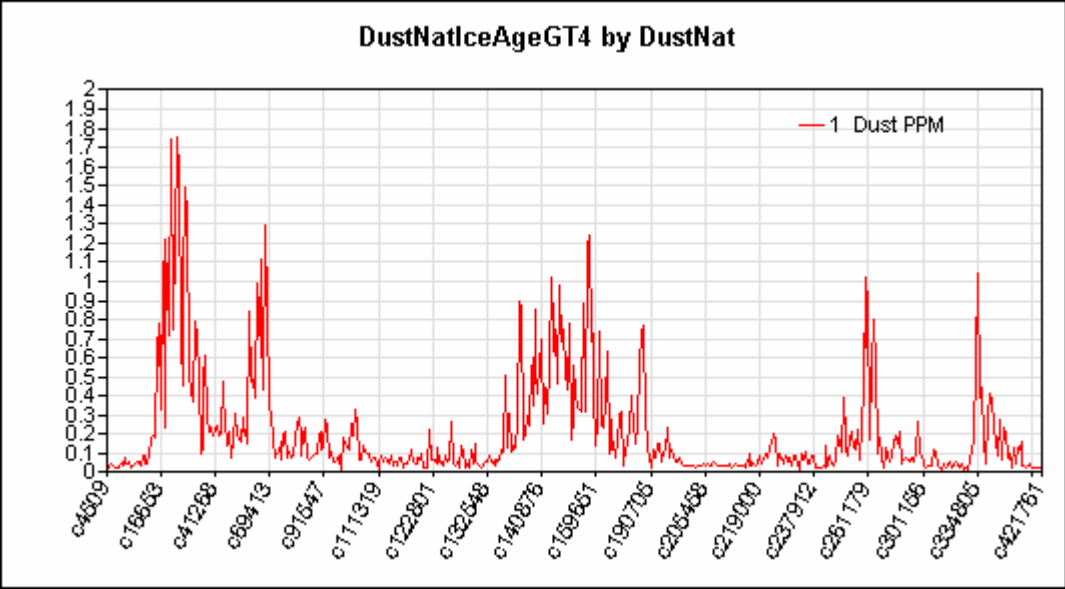
This chart shows the electrical conductance of the DomeC core, at a depth of 772 metres to 3187 metres. YearsBeforePresent is plotted on Y2 as a reference point for the estimated age at the depths on the X axis. Each peak is an interglacial. There are nine interglacials, the last five being somewhat compressed towards the right hand side.



The next chart shows the concentration of CO₂ from the Vostok core on Y1, and the three contenders for the best Age scale on Y2. The X axis gives the depth in metres. Y2 is in 1,000s of years, so the record extends for 500,000 years, and clearly shows the relationship between CO₂ and the last four ice ages.

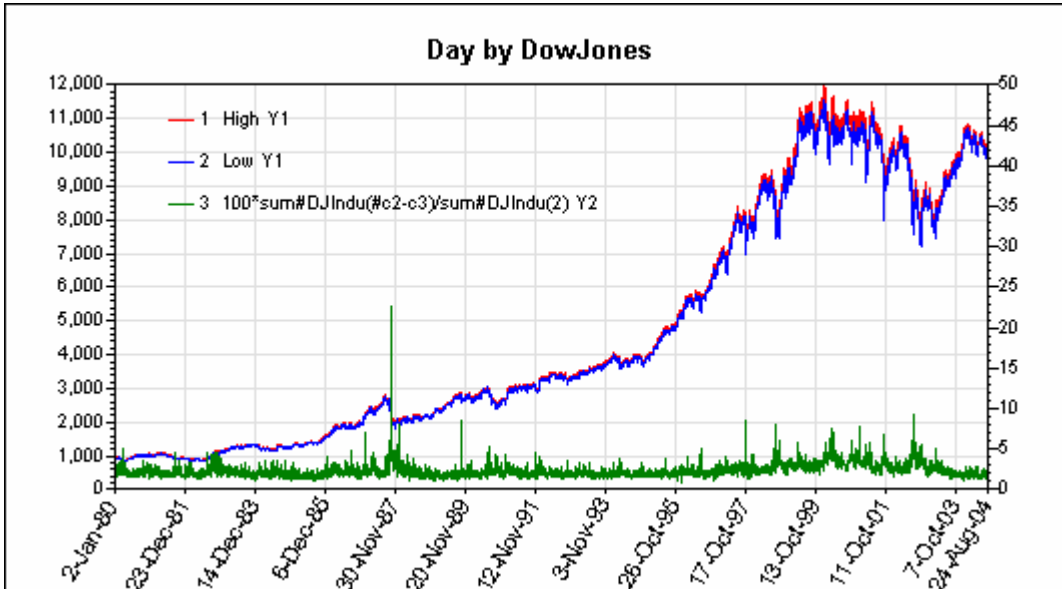


This chart shows the natural dust concentrations from the Vostok core. X axis is YBP using the GT4 scale.

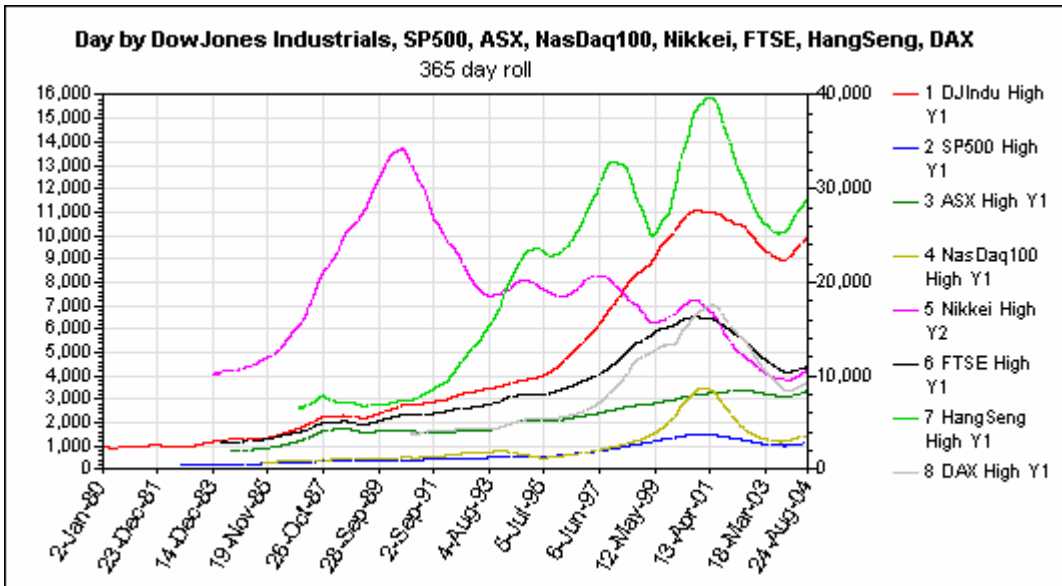


D. FINANCIAL INDICES

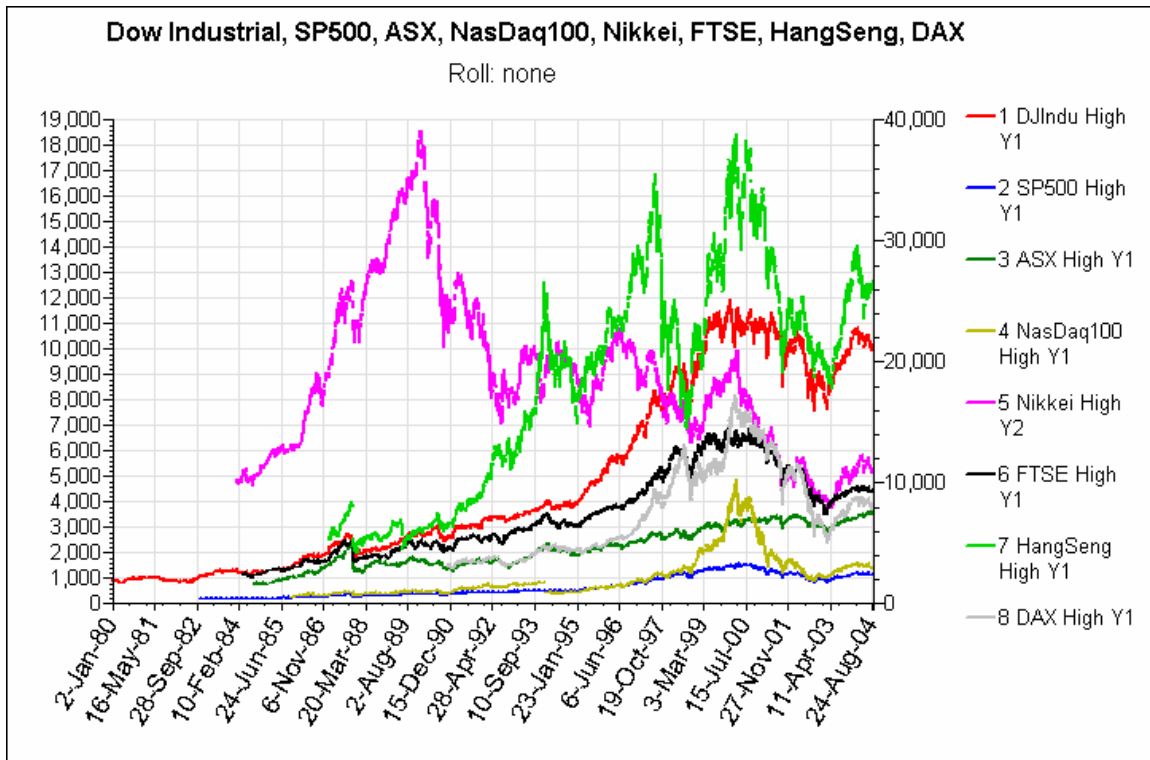
This chart shows the percentage difference in each day of trading for the Dow Jones index. Note that the peak difference is at the 1987 crash.



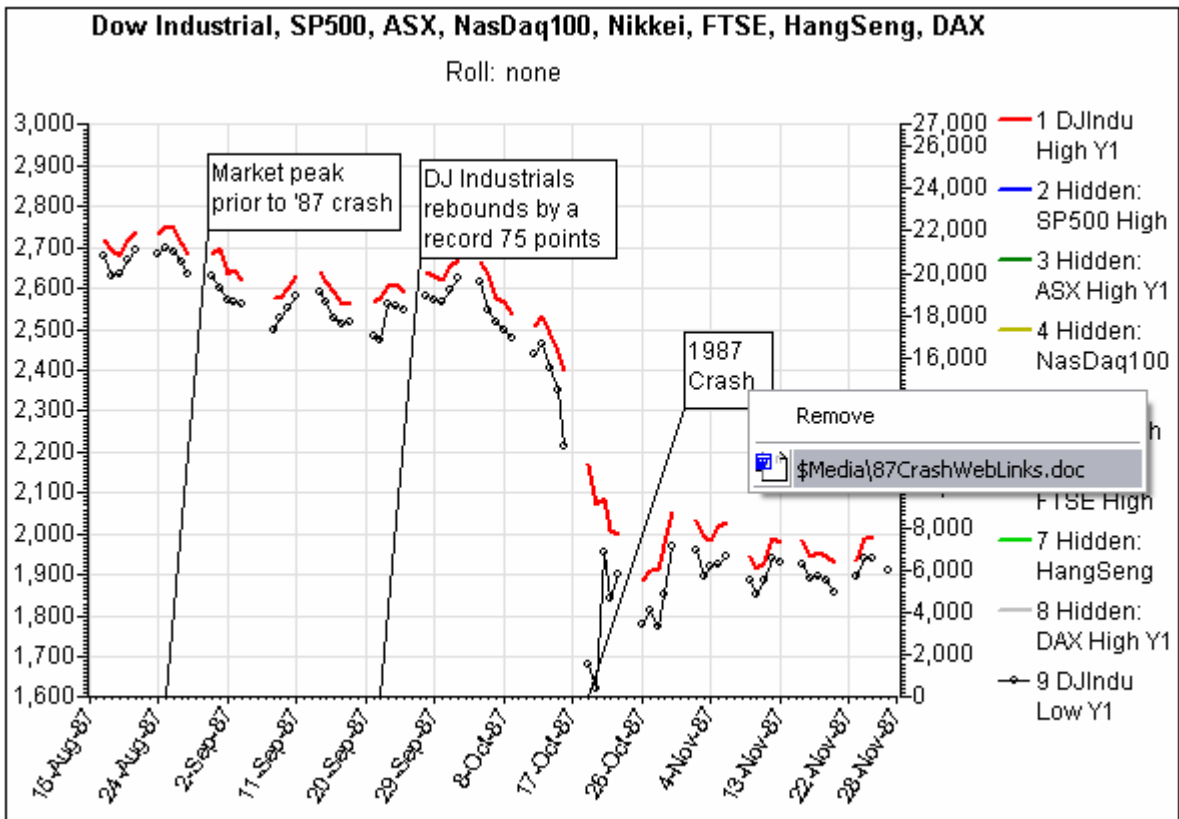
The next chart shows all major stock indices on the same calendar time axis, plotted simultaneously at daily resolution since 1980 on a 365 day roll. The Nikkei is plotted against Y2, others against Y1. This shows that the ASX is least affected by global trends.



To see this unrolled and with all out-of-field periods displayed takes a single click on a popup menu (to unroll) and a single click on the status bar (to unhide empty points).



The entire Highs data set, at daily resolution, is contained in this chart. To see any bit of it up close, such as the Dow Jones Industrials during the 1987 crash, you can solo the series (that is, hide all others) with a right click and then zoom in with a simple drag. There are some entries in the Events database which would be nice to see too. These can be added from the toolbar with a few clicks. Any annotation to an event can carry any number of file attachments. The crash event itself has an attached document of live web links which can be accessed by a right click and then activated with a left click. And for high/low comparison, adding a series for the Dow Jones Industrials Daily Lows takes a right click and a few left clicks. The result, prepared in just the several seconds it takes to do the mouse work, is



[end of document]