DATA TO INSIGHT: AN INTRODUCTION TO DATA ANALYSIS
THE UNIVERSITY OF AUCKLAND

THE UNIVERSITY OF AUCKLAND
NEW ZEALAND
Te Whare Wānanga o Tāmaki Makaurau

WEEK 8
INTRODUCING TIME SERIES DATA by Chris Wild

Welcome back. This week, we'll look at data collected over time, where we're interested in looking at changes over time. Statisticians call it "Time-series data". This is a stand-alone module that can be studied at any time after Week Four.

People are often fascinated by time-series data because it can help them understand the past but, even more, to help them predict the future.

This video will introduce you to plotting data against time and some patterns that that can reveal. In subsequent videos, we'll move on to estimating seasonal differences, forecasting, and comparing related series. We'll be dealing only with the simple case, in which each series has only a single observation at each time point, and our times are equally spaced, as in the following example.

Visitor arrivals by country of residence all travel modes (Qrtly-Mar/Jun/Sep/Dec)

Time	Australia	China PR	Japan	Rep Korea	Germany	UK	Canada	USA
...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...
1999Q1	128851	4516	45130	10325	18173	65127	12878	59117
1999Q2	107972	5237	25209	6791	4506	23418	4572	32473
1999Q3	130391	5332	33904	9568	4538	24982	4846	30767
1999Q4	156214	8156	43102	16550	19026	54744	11000	58524
2000Q1	143198	8077	43422	20778	20136	71694	12210	65078
2000Q2	119461	7140	26237	11753	5712	30641	4765	38235
2000Q3	139935	7116	35714	14113	5188	27941	4295	35536
2000Q4	171268	11169	46000	19937	20415	69974	11701	56932
2001Q1	162073	13732	49558	24855	22527	81372	14629	64848
2001Q2	138051	10878	30005	15712	5348	30072	5219	37370
2001Q3	161923	13685	37753	22615	5749	30857	5157	34812
2001Q4	168502	14879	31769	23985	18858	69345	11689	50351
2002Q1	165370	20306	49969	30966	19327	96363	16746	73588
2002Q2	128454	16240	30852	20314	4623	31521	4936	36047
2002Q3	156386	18342	38992	25602	5641	32678	5085	35636
2002Q4	182260	21646	53754	33054	19360	76424	12902	60018
2003Q1	171254	23538	49118	36347	20817	99406	16747	74095
2003Q2	145567	7185	21524	16029	5372	41008	5180	38091
2003Q3	172015	12603	32365	26881	6532	38116	5011	35652
2003Q4	213326	22663	47844	33401	19813	86289	13002	63786
...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...

| Time	Australia	...	UK	Canada
...	...	...	...	...
...	...	...	...	...
2000Q1	143198		71694	12210
2000Q2	119461	...	30641	4765
2000Q3	139935	...	27941	4295
2000Q4	171268	...	69974	11701
2001Q1	162073	...	81372	14629
2001Q2	138051	...	30072	5219
2001Q3	161923	...	30857	5157
2001Q4	168502	...	69345	11689
...	...	...	...	...
...	...	...	...	...

Source: Statistics New Zealand

Here's a portion of a set of data on visitor arrivals in New Zealand. It's quarterly data, which means it's reported four times a year, covering periods of three months. Notice how the time variable is represented with the year, and then the quarter of the year (Q1 to Q4) for which the figures are given.

Arrivals are reported separately for eight countries -- presumably, the eight biggest sources of visitors. Data recorded over time like this is called "Time-series data". The time format is the one used by Statistics New Zealand. That and several slightly different variants are used all around the world.

What are we going to do with data like this?
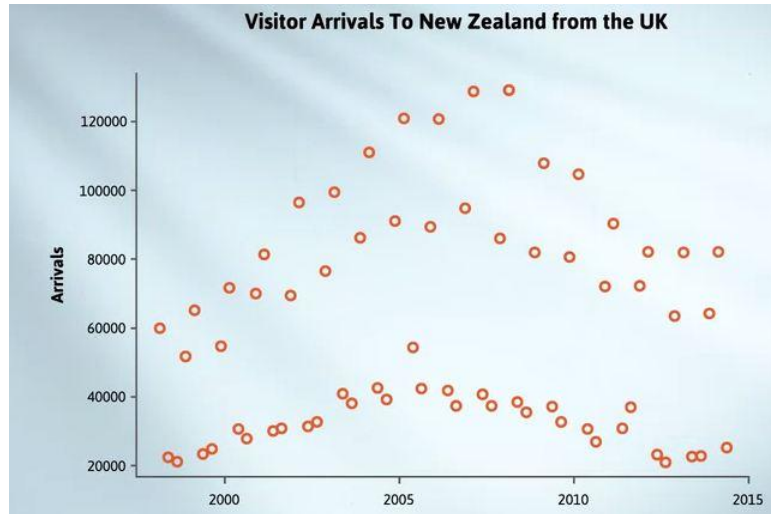
Well we could just do a scatterplot of arrivals versus time. I can see some patterns. I see three bands of points, and I see some sort of up and then down again trend. But people don't plot time series data like this. Typically the data is plotted against time with the points joined up by lines. Why?

Visitor Arrivals to NZ from the UK
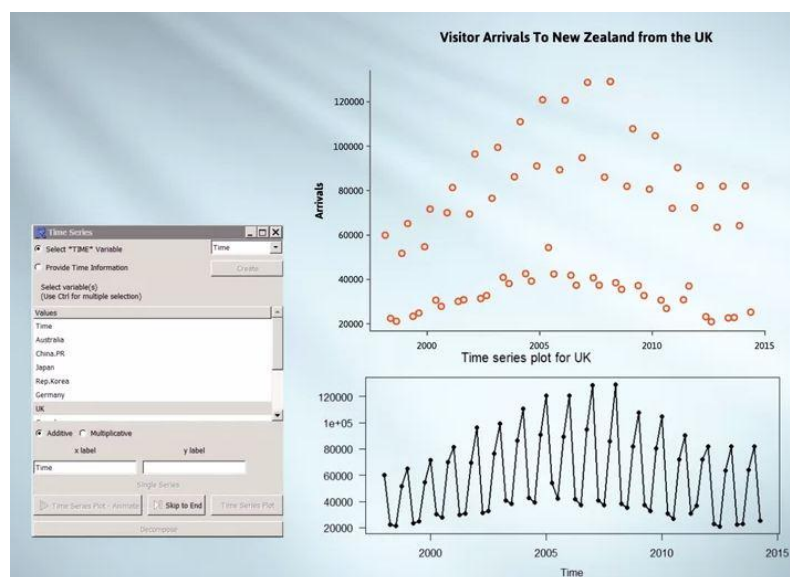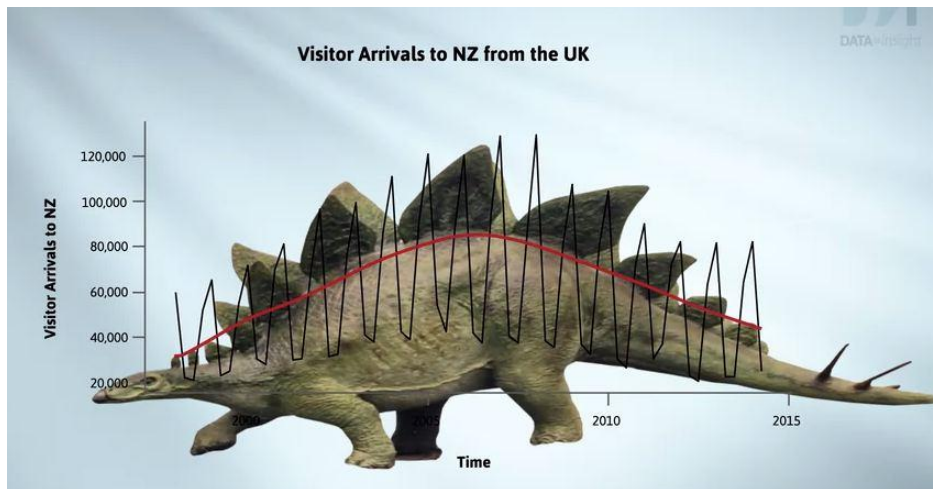
That's why. The saw-tooth pattern we couldn't see in the regular scatterplot jumps out at us as the points get connected up. The major things we see now are an overall trend and the saw-teeth. I'll draw in a vertical line at the position of each year.

Visitor Arrivals to NZ from the UK

There's a basic pattern that repeats every year. We call such patterns "seasonal patterns".

Seasonal plot for UK (Quarterly data)
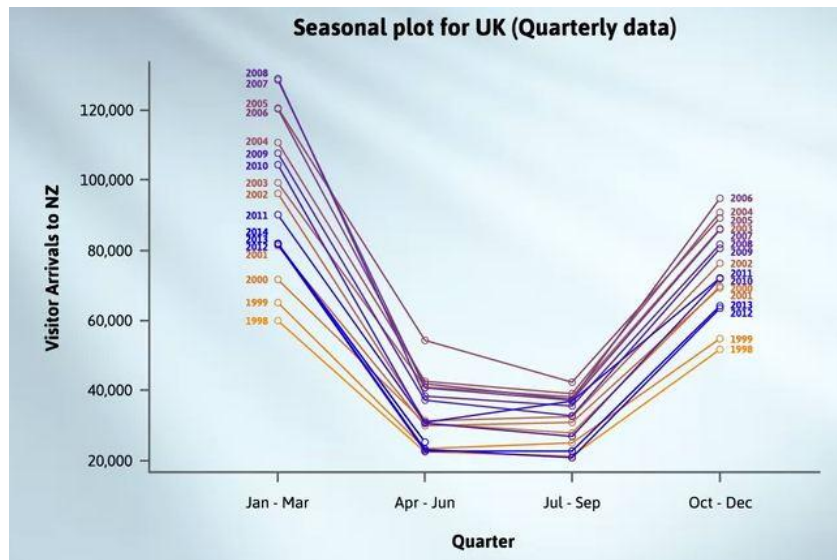
We can see it better here, where we plotted the data against quarter, with a separate line for each year. Every year, the visitor numbers are biggest in the January to March quarter (New Zealand's summer months) and lowest in the April to June and July to September quarters (New Zealand's winter months).

Now I suspect that the October to December figures are high because in New Zealand, December is a warm month, and contains the Christmas holidays. We can check on this because Statistics New Zealand also publishes a monthly version of this series.

The big months are January, February, and December. You may also have noticed there's something odd going on here and here. (Anyway, I digress. Back to this.) Seasonal patterns are common in time series data. There are often patterns that are repeated across hours of the day, days of the week or, as here, months or quarters of the year.

And joining the points by lines helps us to see these patterns. A series with the seasonal pattern is described as a "seasonal series". They are common in social and economic data, particularly that published by government agencies, and in some sorts of biological data.

Why do people worry about identifying and estimating patterns, like the ones we've been seeing?

Mainly because they want to take them and project them into the future. They want to use them for forecasting what's going to happen next so that they can plan for

this anticipated future, how to adapt to it or take advantage of it. This might be having the right levels of resources in place, perhaps budgeting to employ additional staff to cover peak periods. And so on. In the next video, we'll discuss breaking a seasonal series into component parts, with a particular emphasis on the seasonal components. A process called decomposition. We'll also talk about forecasting.

Some Interesting Links:

- http://www.bloomberg.com/graphics/2015-whats-warming-the-world/
- https://public.tableau.com/en-us/s/gallery/evolution-global-temperature
- http://xkcd.com/1732/