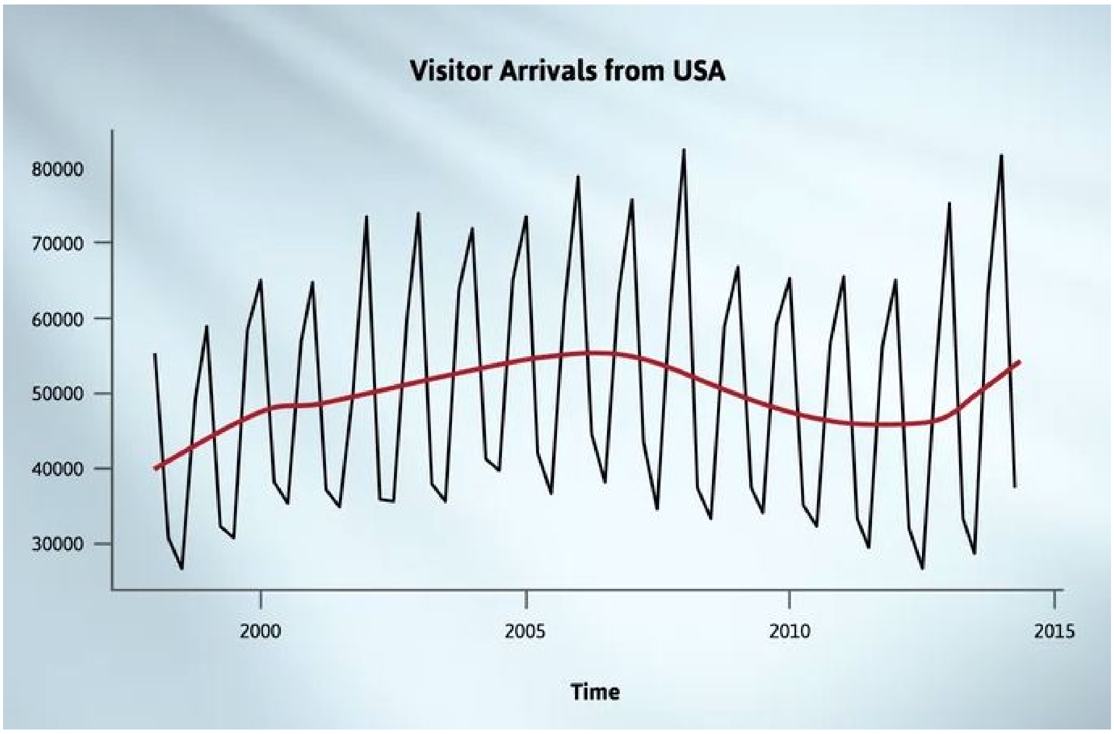
DATA TO INSIGHT: AN INTRODUCTION TO DATA ANALYSIS

THE UNIVERSITY OF AUCKLAND

WEEK 8

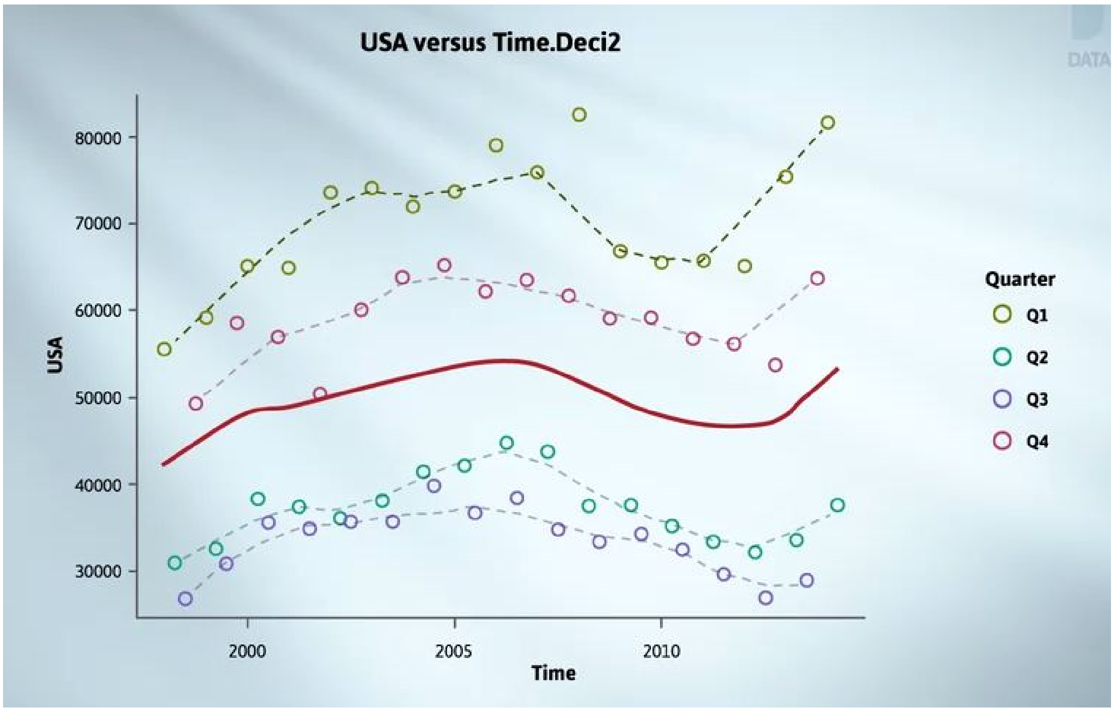
SEASONAL DECOMPOSITION AND FORECASTING, PART I by Chris Wild

Hello again. Last time we showed how plotting time series data with points joined up by lines helps us see seasonal patterns. We'll now move on to discuss decomposing a seasonal series into component parts. We'll emphasise investigating seasonal differences.



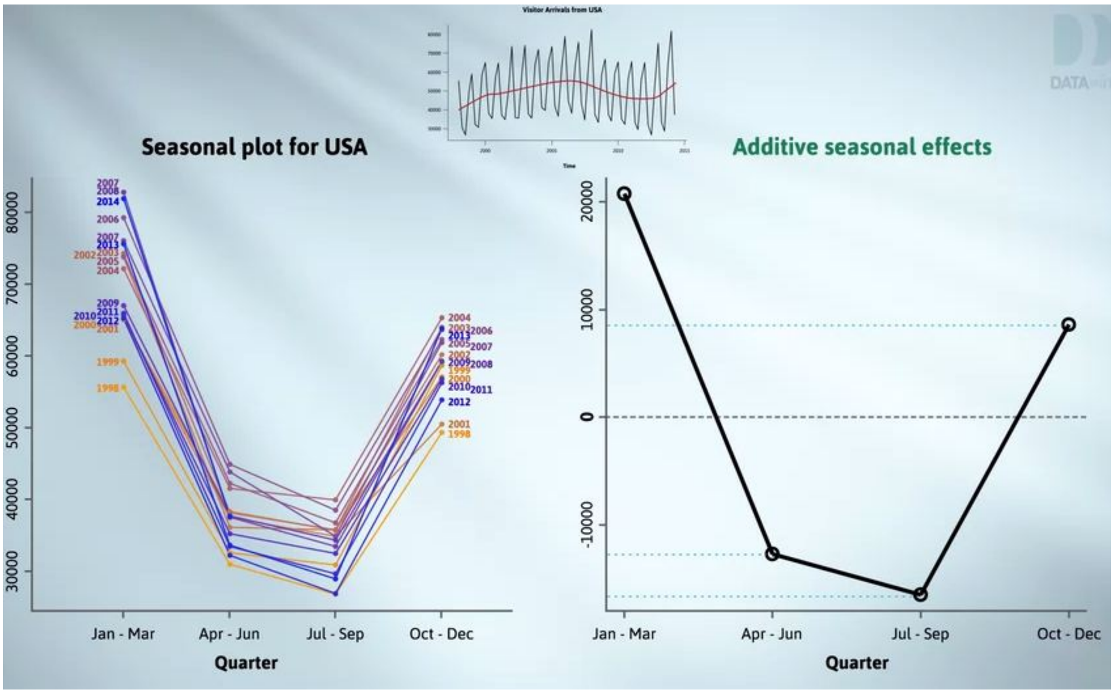
Still from animation

We'll start by looking at the visitor arrivals from the US. (A basic time series plot with a smoother added, summarising the trend.) We're thinking of what we see in terms of an underlying trend with a seasonal pattern superimposed on top of it. There are various ways of getting the trend.



Still from animation

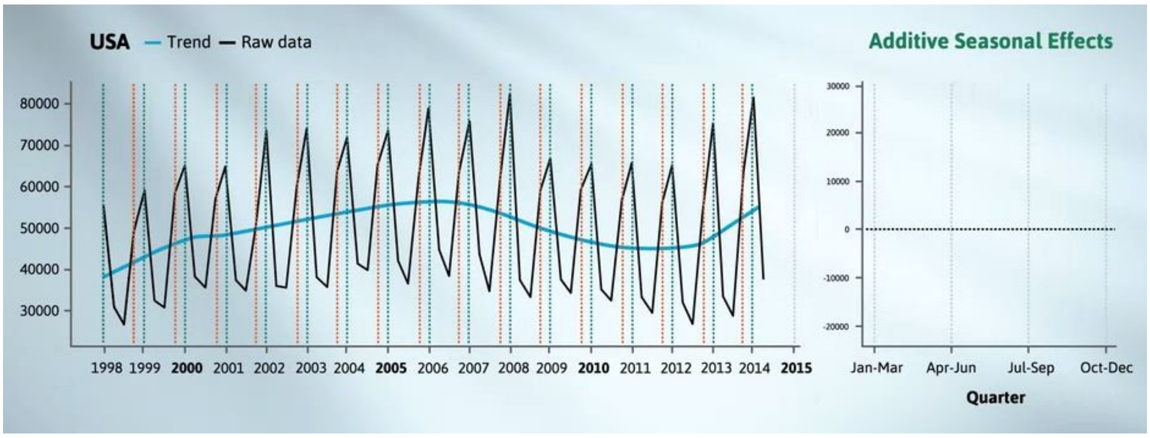
This is just a smoother applied to the scatterplot. Or we could smooth the series for each quarter separately and then average the smooths.



Still from animation

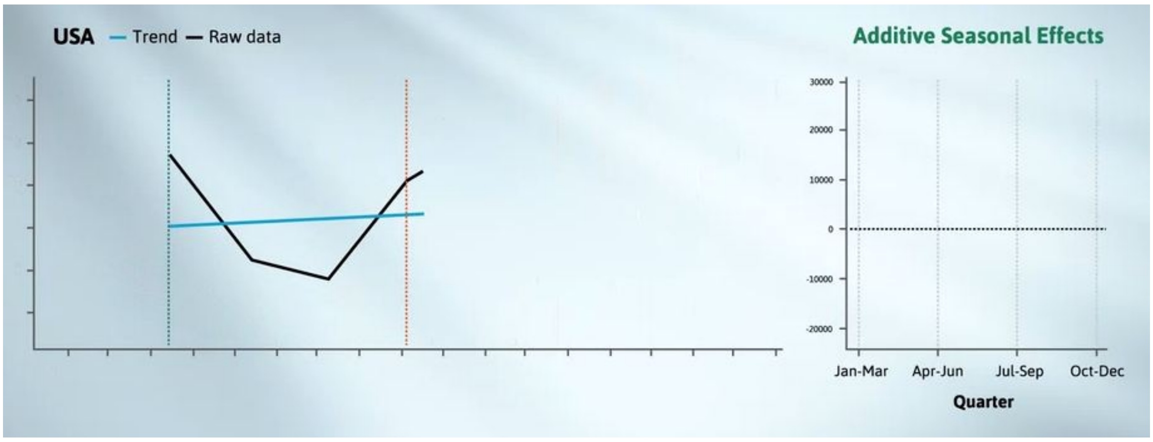
Here's a seasonal plot for USA. We've seen things like the left hand panel before. We plot against quarter and draw a separate profile for each year. But what about the right hand panel? It's the average seasonal effect. (We'll deal with the annotations later.) So what's the right hand picture and how is it made?

It goes basically like this.



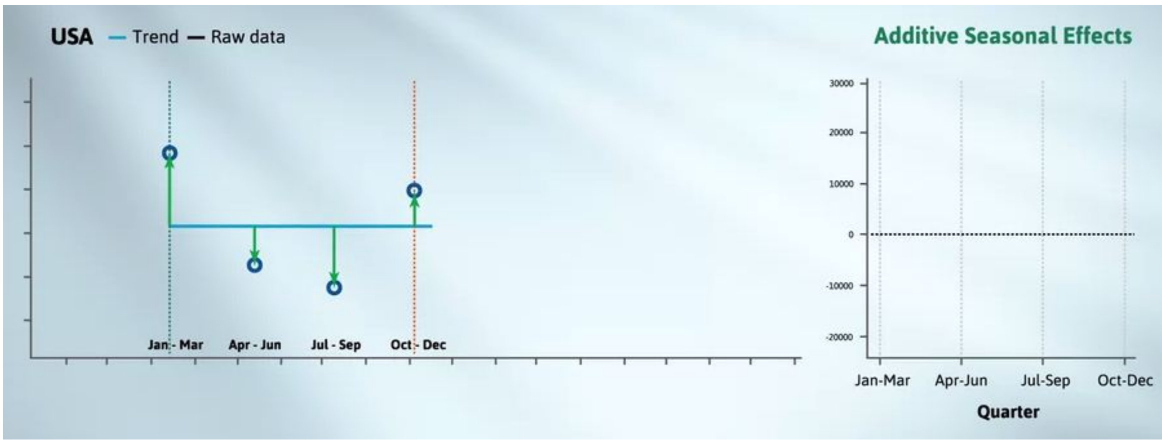
Still from animation

The first quarter of each year is marked with a blue green line; the last quarter with an orange line.



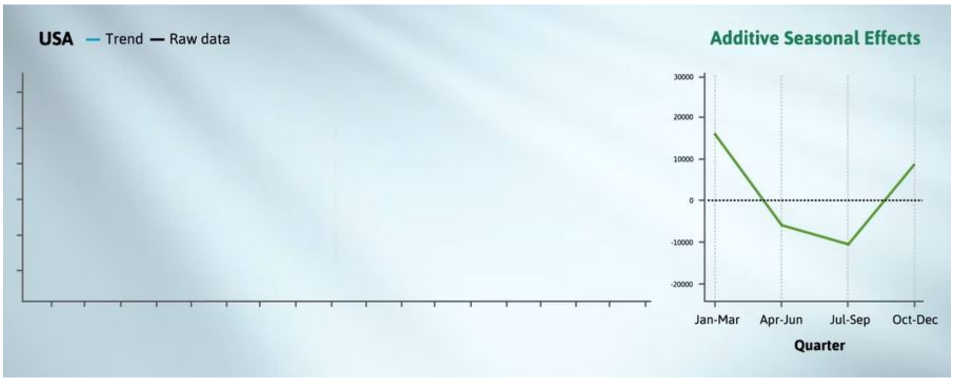
Still from animation –stretched out

Let's focus on the first year and stretch it out.



Still from animation

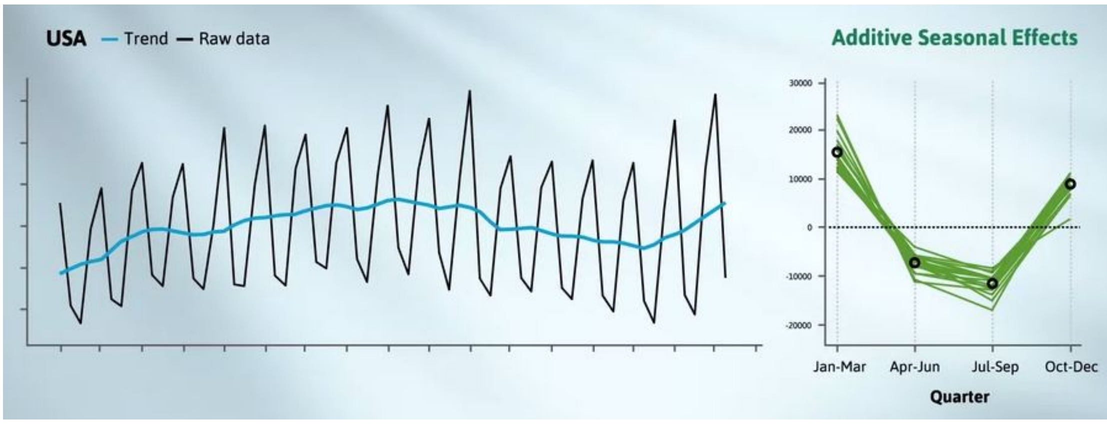
We'll put the actual points back on and take away the lines; put on the distances the points are from the trend; put it on a level playing field so that the distances are all we're seeing;



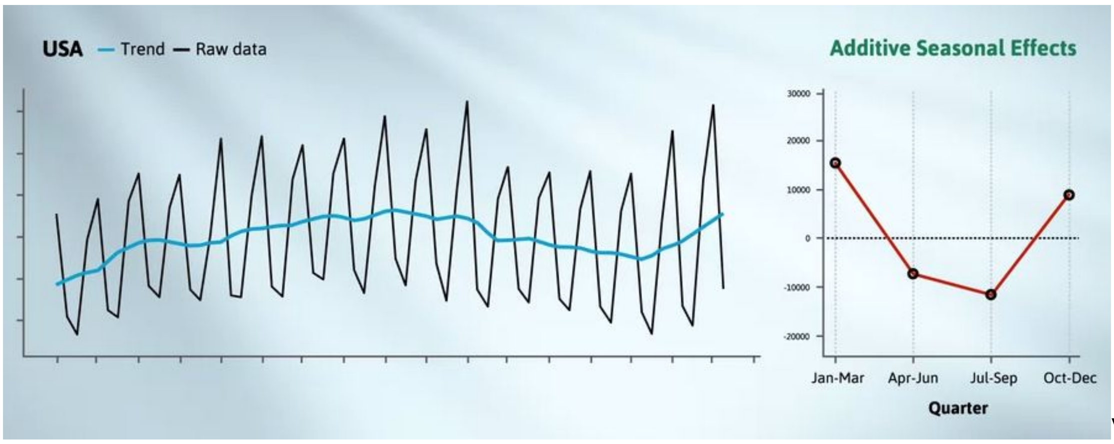
Still from animation

put the lines back on; and store it away in the right hand window.

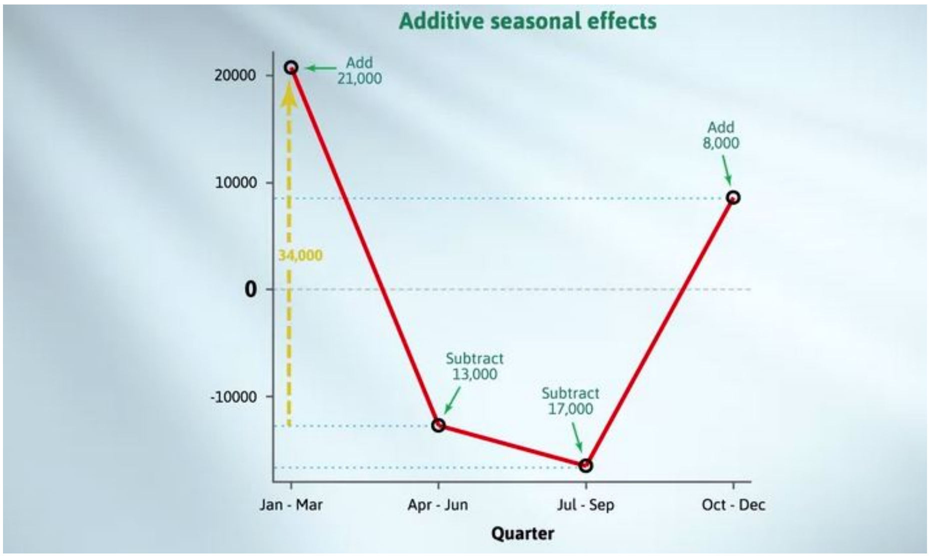
Now do this for every year.



Take the average at each quarter. Connect it up. That's your average seasonal effect. But don't worry, you'll never have to actually do any of this. Software does it for you.



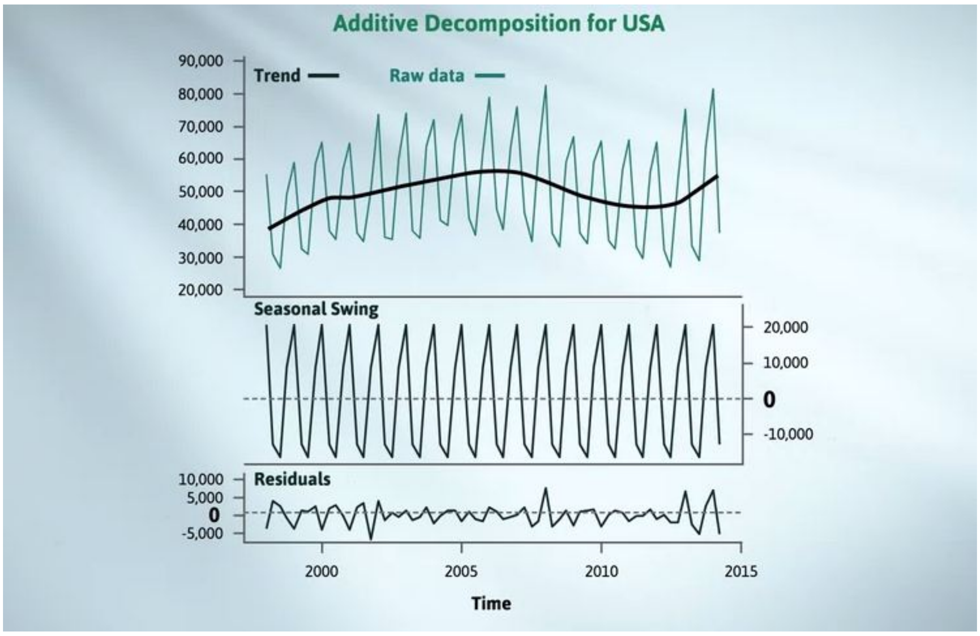
We simply showed it to you to help you understand what the red line represents.



So if we think the seasonal effects are fairly constant over time, it makes sense to use this and think that, generally, the January to March figures will be 21,000 visitors above what you'd expect from the trend. The April to June figures are about 13,000 below. July to September is about 17,000 below. And October to December is about 8,000 above what you'd expect from the trend. (Often the trend is fairly flat over a single year, so that we could think of these as deviations from the yearly average.)

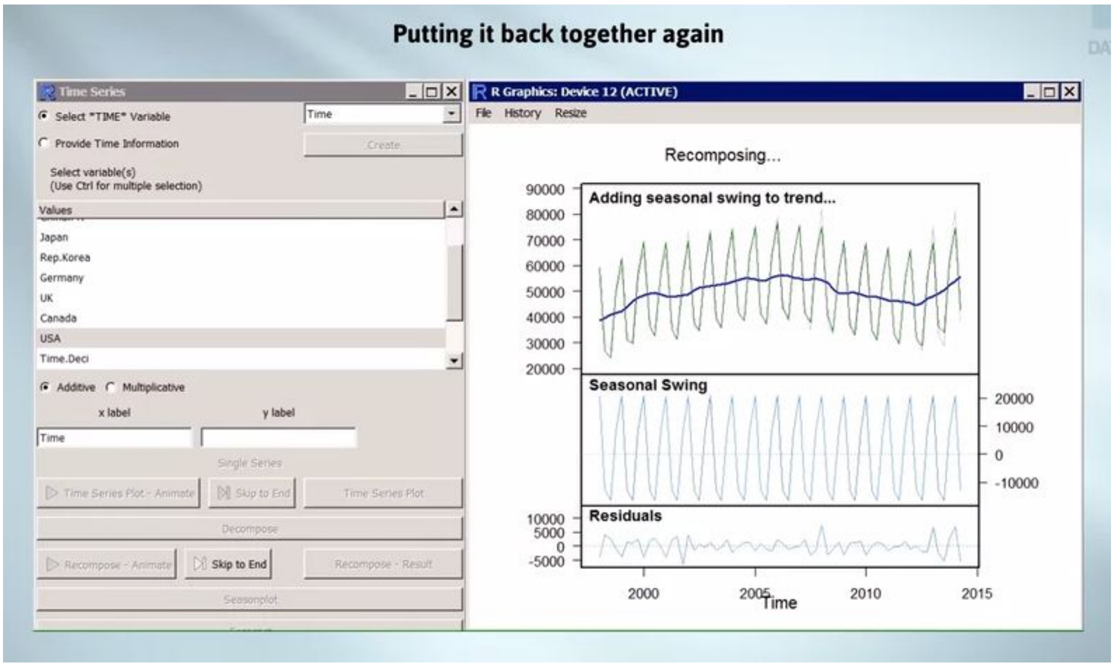
We can also look at differences between quarters in the obvious way. For example, we expect the January to March quarter to have about 21 plus 13, which equals 34,000 more visitors from the US than the April to June quarter.

Finally, note the horizontal line on this plot at zero. The seasonal terms are up or down adjustments to be made to the trend. Zero is the no-change value because adding zero makes no change. We'll need this idea later when we contrast additive and multiplicative seasonal effects.



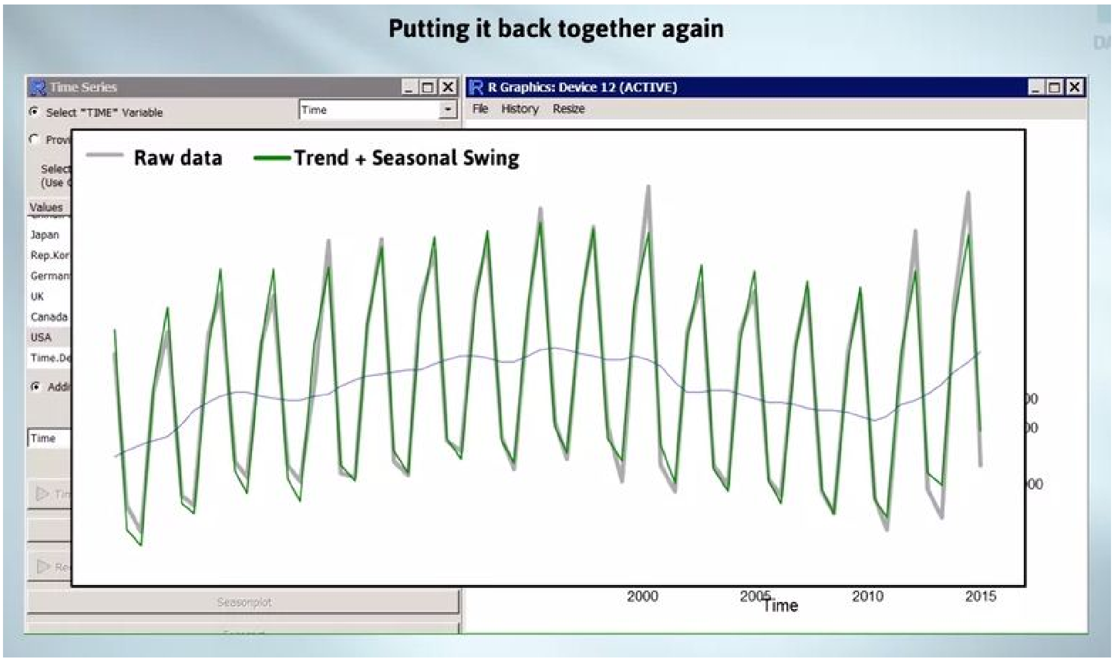
This is a decomposition plot from iNZight, assuming constant additive seasonal effects. The top panel has the basic plot repeated. The second panel is capturing the seasonal swings, assuming that they're all the same apart from purely random differences. So the seasonal averages are just repeated over and over again.

Why? We're hoping to approximate the series well, in terms of simple, stable components that we can project into the future to form forecasts. If constant seasonal swings give us good approximations, that's something which is easy to project forwards. We'll ignore the residuals in the third panel for now.

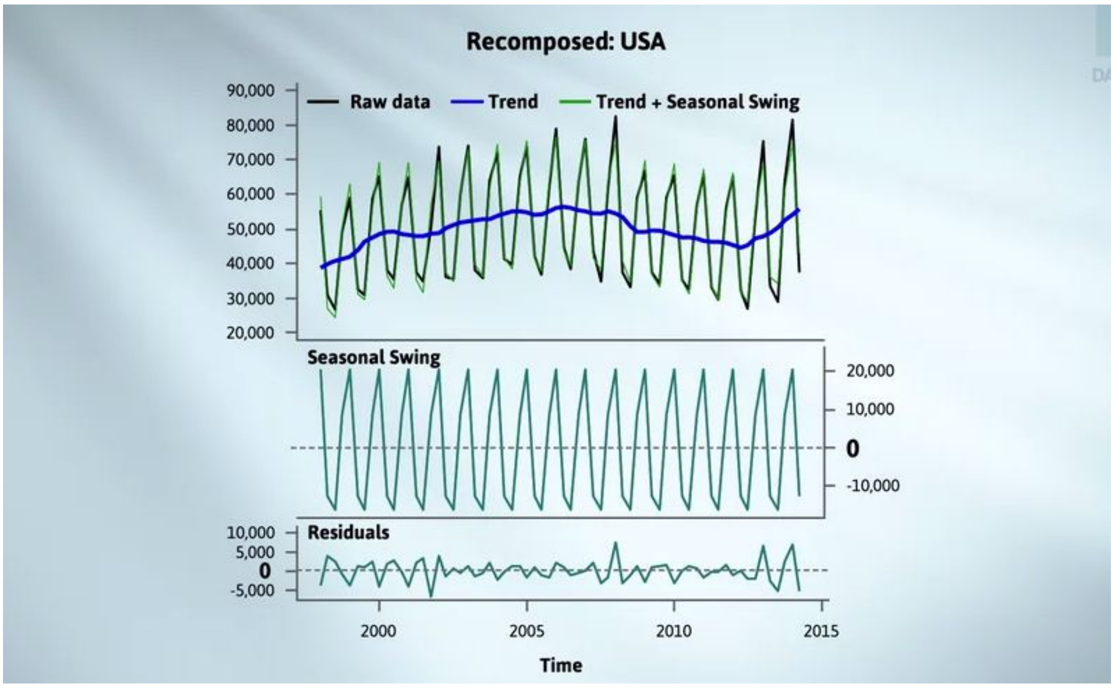


Here we're putting the pieces back together again.

We're unwrapping the seasonal swings around the trend. We've now completed drawing trend+season onto plot in green. If you look carefully, you'll see the original series in behind in grey. We'll go in and look at this more closely.



Trend+season is almost, but not quite, recreating the original series.



The residuals (in the bottom panel) are the difference between the original series and what trend+season gives us. Here are the residuals on now, giving us the original series.

For this series, the residuals are small compared to the movement of the trend (which ranges from about 40,000 to 60,000) and the seasonal swings (which go between about plus or minus 20,000).

Now this is what's called an "additive decomposition". It assumes the underlying seasonal swings are the same every year and they add to the trend value, as we've seen. They're only useful if the seasonal swings look similar.

Next, we'll show an example where the seasonal swings are obviously not constant. This will lead us to another type of decomposition, "multiplicative decomposition".

But even this example, if we look very critically at what trend plus additive seasonal swing is giving, there's a slight suggestion that the swings are too wide at the left hand end where the series-trend is low. And not wide enough in the places where the series-trend is high.