Coronavirus modelling:

Hey guys, I've been doing a bit of analysis with the data from the corona virus (after all, I AM a data scientist!), and I just wanted to share a bit of it with you all. I have refrained from pursuing full-blown ML/AI or mathematical modelling of the corona virus epidemic, as there are a lot of uncertainties and unknowns and some prediction can create least a sense of unease for people (myself included).

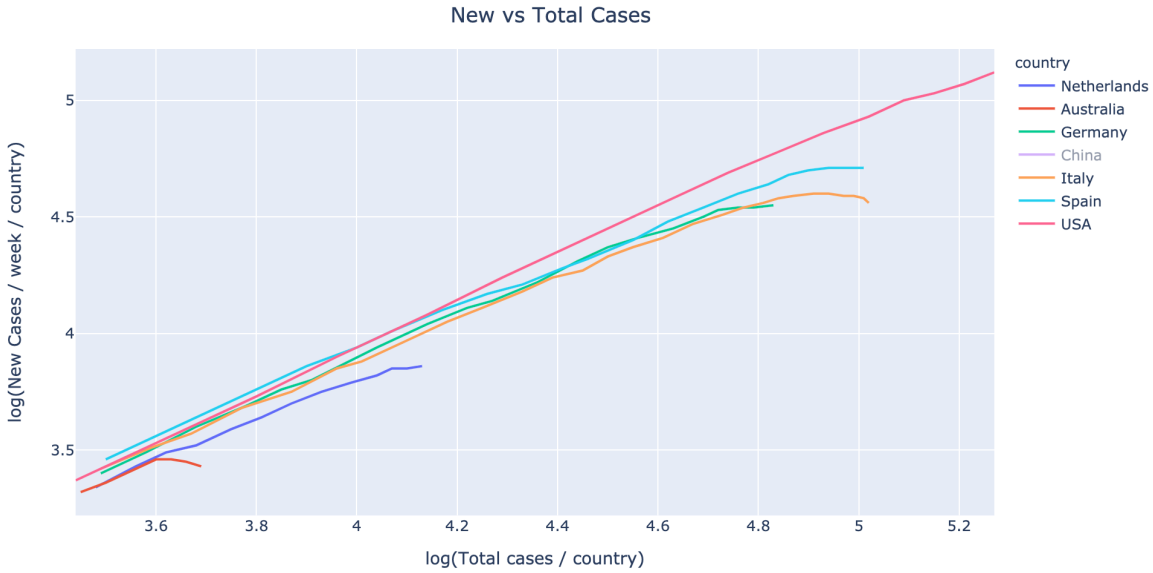
So for this reason I have taken a bit of a different tact and instead analysed the cases in log-space. This might seem quite technical, but it's actually quite simple. Epidemics usually follow what's called *exponential growth*. It is called this because it can be described by the exponential family of functions, including the natural number, *e*. If one takes a logarithm, say log10 — called log base 10 — then exponentials become straight lines. In this situation exponential growth becomes something like y=x and the flip side — exponential decay — becomes something like y=-x. I have used these types of plots for many years, though usually it was to look at the cooling times of electrons and positrons in giant molecular clouds in different parts of our Galaxy, and not to analyse pandemics...

There are several benefits to using logarithms in this instance: firstly, noise is smoothed out in log-space and secondly, trends are more readily apparent in log-space. So for this reason, analysing trends using this method can be more robust to outside influences. Having said that, these types of things still *do not rectify problems like under-testing, differences in testing regimes, bad reporting of numbers*, etc., so I make no claims for the following analysis being more accurate that anyone else's, least not those carried out by the RIVM and other associated national epidemic modelling bodies. However, I do claim that my way of doing things are more robust than just the raw numbers.

Analysis:

So for today, April 01 2020, you can see that I have plotted, for several countries, the log of the total cases that that country has reported against the log of the new cases over the past week for that

country: this gives some invariability to local noise of reported cases, such as all cases not being reported that day. As you can see, each country follows a straight line and then most of them curve over. The best cases are Italy and Australia, where they have probably started to move into an exponential decay scenario. Spain is also looking like they will follow any day now, which given the sheer number of cases there. The Netherlands has also seen some "flattening" of the curve, which in these plot types, looks like a turn-over. Unfortunately, the USA doesn't seem to be following this pattern yet.

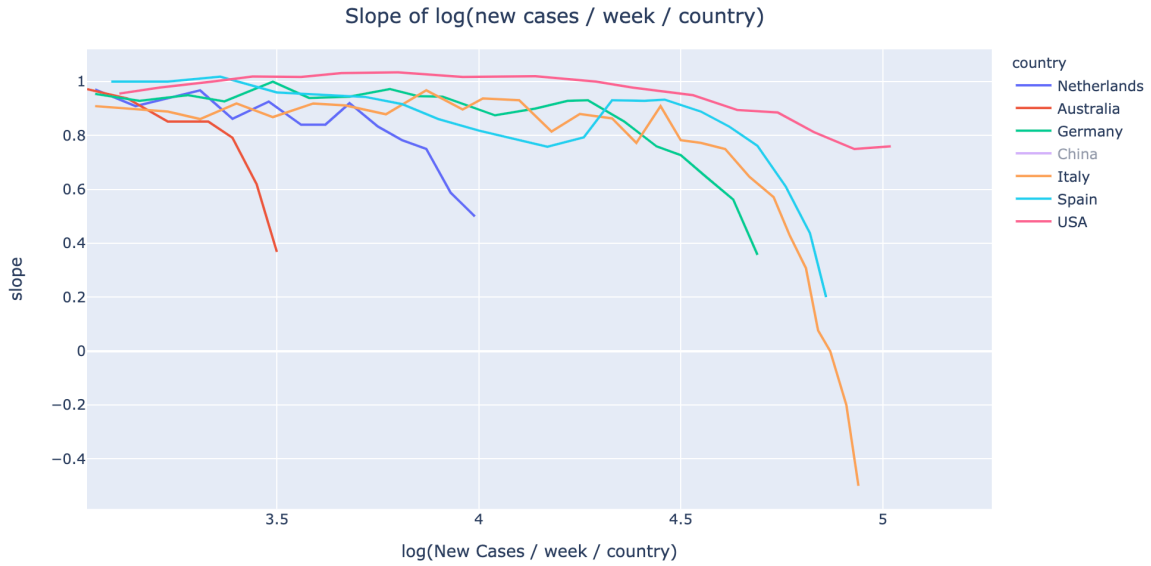


A small note/disclaimer on the data for this plot is that most data is accurate to what is reported on April 1, expect for the USA (due to them still being in the morning of April 1) and Italy (which reports its cases at 6pm their time). So this data is for cases reported daily up until 31 March, 2020.

One question people have is: how can one tell that a country will soon fall into exponential decay? The answer to this is that we can look at what is called *the derivative* of the curve and find when it goes through zero: this denotes turnover. One can think of the derivative in one sense as the *tangent* line to the curve. That is, a line that just grazes our line at the point we are looking at. When this "tangent line" is positive, the line will be increasing (exponential growth in our example); and when it is negative, it will be decreasing (exponential decay); and when it is zero, then the curve is flipping.

So, if we calculate the derivative for these curves, what does it look like? Well, it looks like the plot below. This shows the derivatives for

the same countries as above (I can't show China as it makes a nice plot like this for the other countries impossible). It shows that most countries come from a slope of between 0.8 and 1, which is exponential growth when the total number of cases are > 1000, where log(N) =3. The data go higher than this earlier, but is noisier than this because of the low numbers — and even in the plot below, I have averaged this over the previous 8 data points to make it much less noisy.



What can be seen here is that most countries are doing better: even the USA seems to be doing a little better (despite the up-tick of today). However, one can see that the slope for Italy has become negative: this suggests that it is *already* into exponential decay. So hopefully this means that we will see many fewer cases from Italy in the next few days! But one can also see significant (in that the changes showing now are much lower than any noise that is obviously present in the plot) lowering of the rate of increase for almost every other country: it this trend continues, Spain should be in exponential decay before next week, and Germany, Netherlands and Australia shortly thereafter. Hopefully this is also true for the USA.

The pleasing aspect is *there is no amount of averaging that I can do for these plots that does not show SOME progress* in the fight against this virus. I will be updating these plots in the coming days, so please let me know if you have any questions or want to know what is happening on these plots. Also, conversely, please also let me know if you do not like it. The *last* thing that I want to do is to make people feel bad, so if you do, then I will, of course stop this. But that said, my

colleagues in the Advanced Analytics team have very much enjoyed these plots and our surrounding discussions, which is why I've opened it up to you guys as well.

Stay safe and wash your hands a lot!
David