Prediction Using COVID-19 Data from Johns Hopkins

Bill Kayser

April 11, 2020

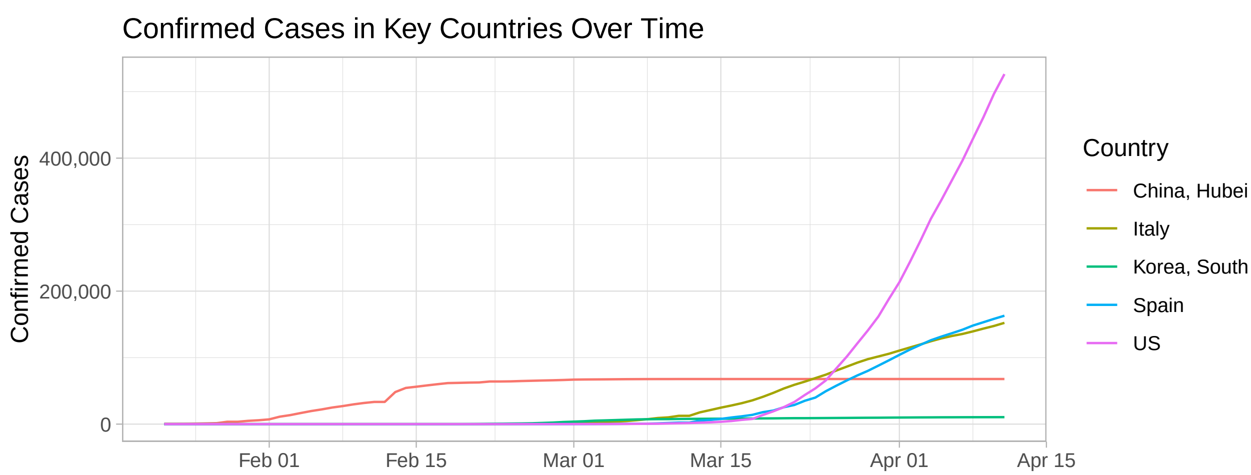
Contents

Introduction

This is an attempt to predict when recovery from the virus begins in different regions. Data in this report was pulled from several aggregators including primarily the 2019 Novel Coronavirus Visual Dashboard operated by the Johns Hopkins University Center for Systems Science and Engineering. I also pulled from the site CovidTracking.com.

I did a lot of exploration to see if there were any variables that I could normalize and model for prediction. I looked at many charts of Confirmed Cases and Deaths as well as their first and second derivatives. Data about Deaths is sparse, and the data about confirmed cases is unreliable.

Confirmed Cases is the total number of confirmed cases since the start of the outbreak. It does not take into account recoveries or deaths, and it never decreases. The time series chart of Confirmed Cases looks like the following.



You can see this is increasing relentlessly in the US, Spain, and Italy. In China where Wuhan is and in South Korea the cases have leveld off. This seemed to beg the question…

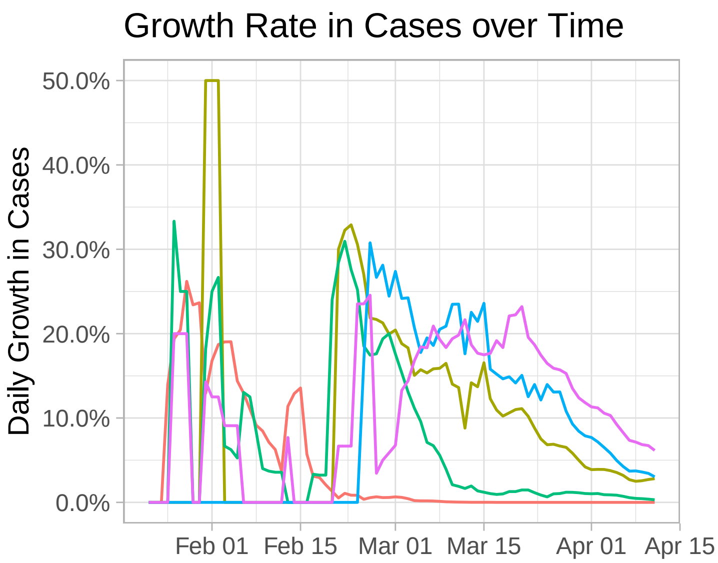
Can we predict when cases will level off in other regions?

I kept digging through the data until the one thing that stuck out was the rate of growth of new cases. I derived this parameter from the of total Confirmed Cases shown in the plot above. I called it "Growth" and calculated it by taking the average number of new cases in the last three days and dividing it by the current total number of cases.

For example, if in Texas I had 300 new cases in the last three days, that would work out to an average of 100 cases per day. If the total number of cases is already 1000, then my growth rate is 10%. If the number of new cases holds steady until we get to 10,000 total cases in Texas, then the number of new cases–100 per day–is a growth rate of only 1%.

If the rate holds steady at 10%, then when we get to 10,000 cases in Texas, we'll be seing about 1,000 new cases per day. This is the kind of growth we are generally seeing. But what I noticed is that the growth rate seems to have peaked just about everywhere and is on the decline. What we were calling exponential growth no longer seems exponential.

Here's what that chart looks like.



You can see this rate trending toward zero. In China and South Korea i'ts already pretty much leveled off at a point that represented a turnaround in both countries. It was late February in China and the second week of March for South Korea. The other countries seem to be on their way to zero.

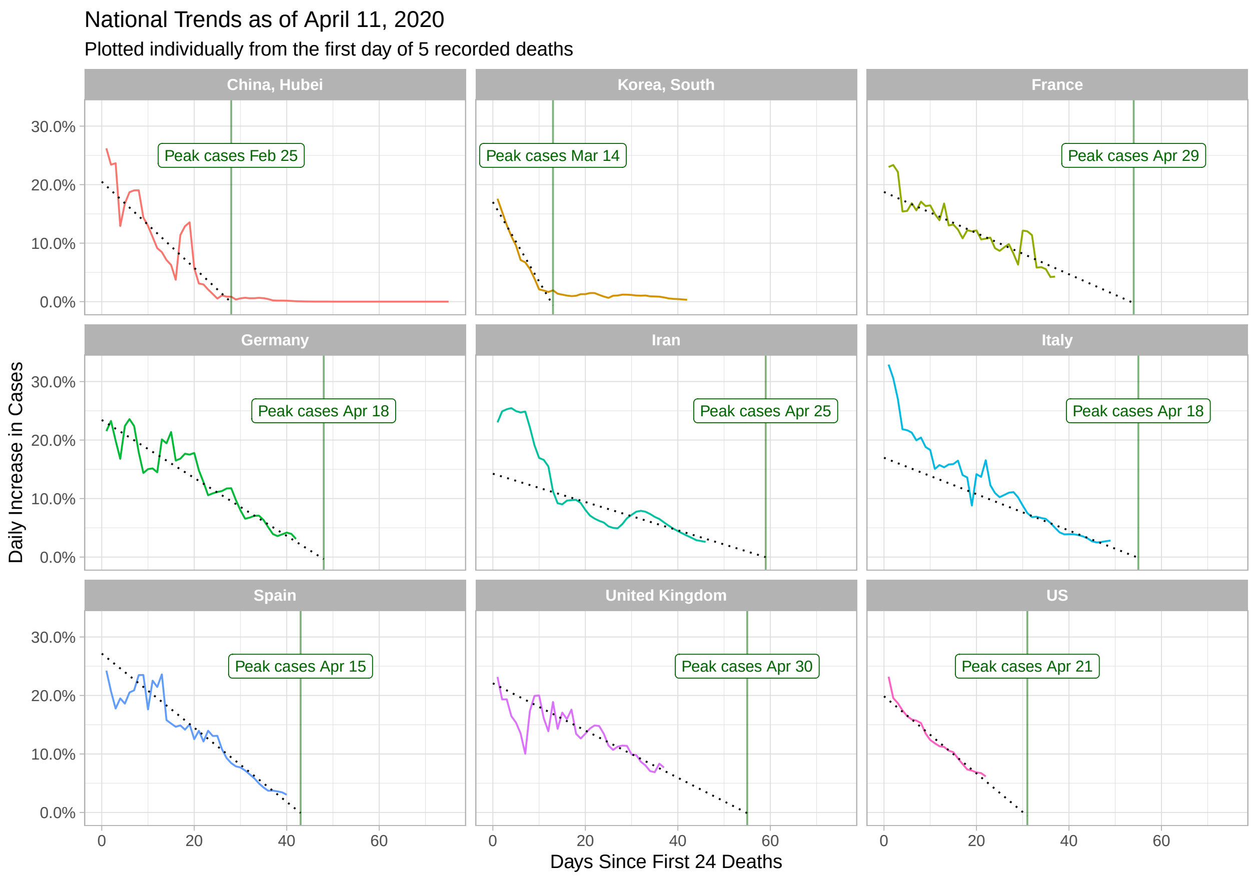
What struck me about this was the last leg of these charts. The decline in the growth rate seemed pretty steady, and almost linear. So I thought, let's use a simple linear extrapolation over the last 14 days and use that to predict when the growth of new cases essentially stops.

International Predictions

I started by analyzing several countries. I took a cue from a chart in the Financial Times to compare these charts side by side starting from where the peak in growth rate occurs. I plotted the growth rate for each of these countries and then plotted a linear model for the previous 14 days. For Hubei and South Korea I omitted the history once their rates bottomed out.

Then I used the linear model to predict on what date these countries will stop increasing new cases.

I believe these dates represent when things will be at their worst. There will be a relatively small number of new cases and the recoveries will have already kicked in, reducing the occupancy rates in hospitals.

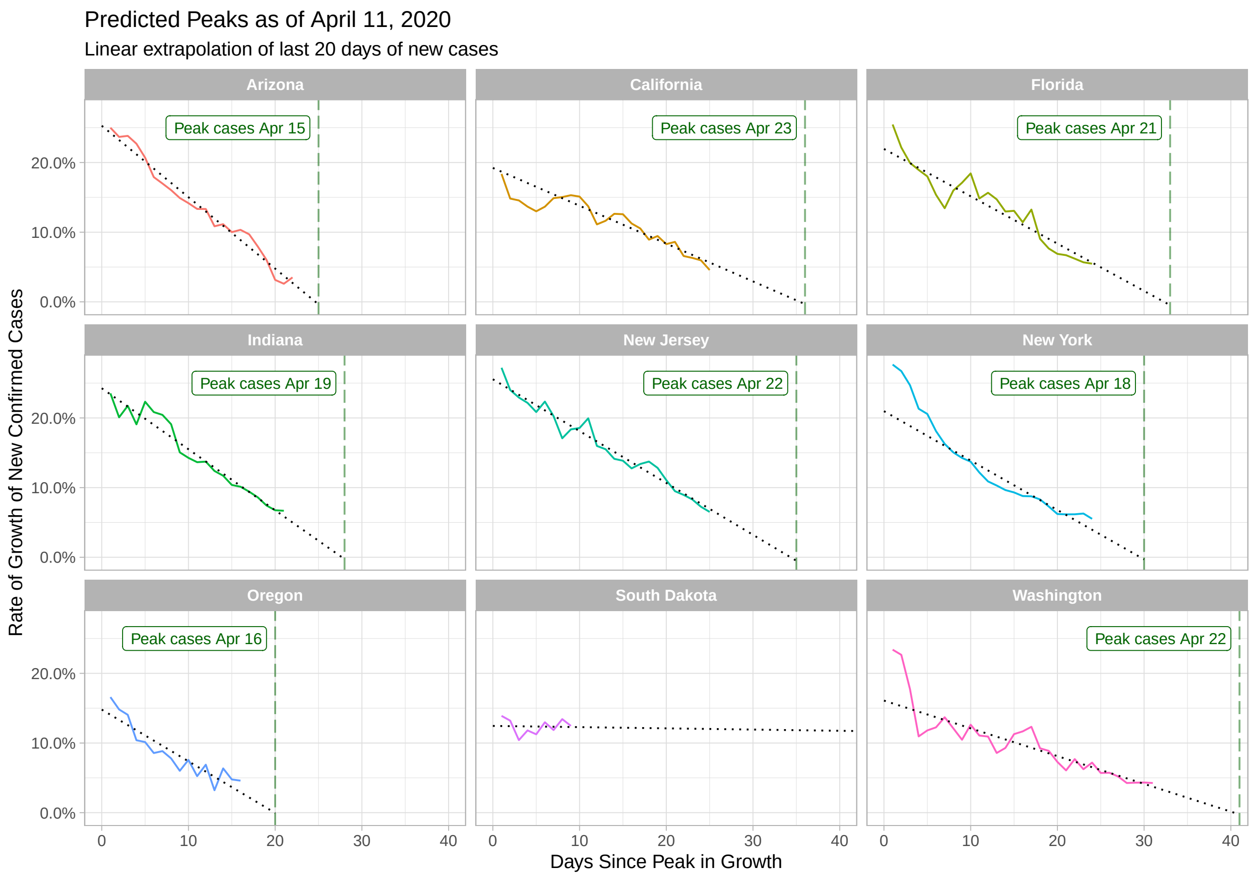


What's key to note is that the prediction for China and South Korea is pretty much right on. The prediction line for other countries looks like a pretty good fit, and is particularly striking in France, Spain and the US.

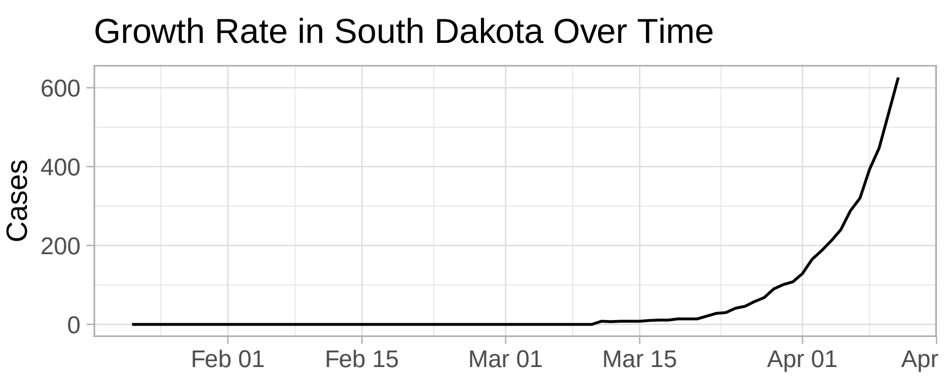
Country	Date	Days.From.Now
China, Hubei	February 25	-47 days
Korea, South	March 14	-29 days
France	April 29	17 days
Germany	April 18	6 days
Iran	April 25	13 days
Italy	April 18	6 days
Spain	April 15	3 days
United Kingdom	April 30	18 days
US	April 21	9 days

State by State Predictions

Then I turned to the US to see if I could go state by state.



In South Dakota, things are so early that the rate of increase is unstable.



Finally, I ran a calculation on each state and put the results in a table.

State	Date	Days.From.Now
Alabama	April 24	12 days
Alaska	April 19	7 days
Arizona	April 15	3 days
Arkansas	April 14	2 days
California	April 23	11 days
Colorado	April 20	8 days
Connecticut	April 24	12 days
Delaware	May 10	28 days
District of Columbia	May 11	29 days

State	Date	Days.From.Now
Florida	April 21	9 days
Georgia	April 27	15 days
Hawaii	April 16	4 days
Idaho	April 14	2 days
Illinois	April 23	11 days
Indiana	April 19	7 days
Iowa	April 26	14 days
Kansas	April 20	8 days
Kentucky	April 16	4 days
Louisiana	April 23	11 days
Maine	April 18	6 days
Maryland	May 9	27 days
Massachusetts	April 23	11 days
Michigan	April 18	6 days
Minnesota	April 21	9 days
Mississippi	April 19	7 days
Missouri	April 18	6 days
Montana	April 15	3 days
Nebraska	April 25	13 days
Nevada	April 25	13 days
New Hampshire	April 15	3 days
New Jersey	April 22	10 days
New Mexico	April 21	9 days
New York	April 18	6 days
North Carolina	April 22	10 days
North Dakota	April 21	9 days
Ohio	April 19	7 days
Oklahoma	April 17	5 days
Oregon	April 16	4 days
Pennsylvania	April 25	13 days
Rhode Island	May 6	24 days
South Carolina	April 23	11 days
South Dakota	March 16	703 days
Tennessee	April 19	7 days
Texas	April 29	17 days
Utah	April 18	6 days
Vermont	April 18	6 days
Virginia	May 2	20 days
Washington	April 22	10 days
West Virginia	April 26	14 days
Wisconsin	April 20	8 days
Wyoming	April 15	3 days

What's Next?

If I am going to keep developing this I'll probably try to mix in data on recoveries to see if I can predict the next milestone. If this predicted milestone is when things are at their worst, the next milestone is when we've recovered enough to where restrictions can be loosened.