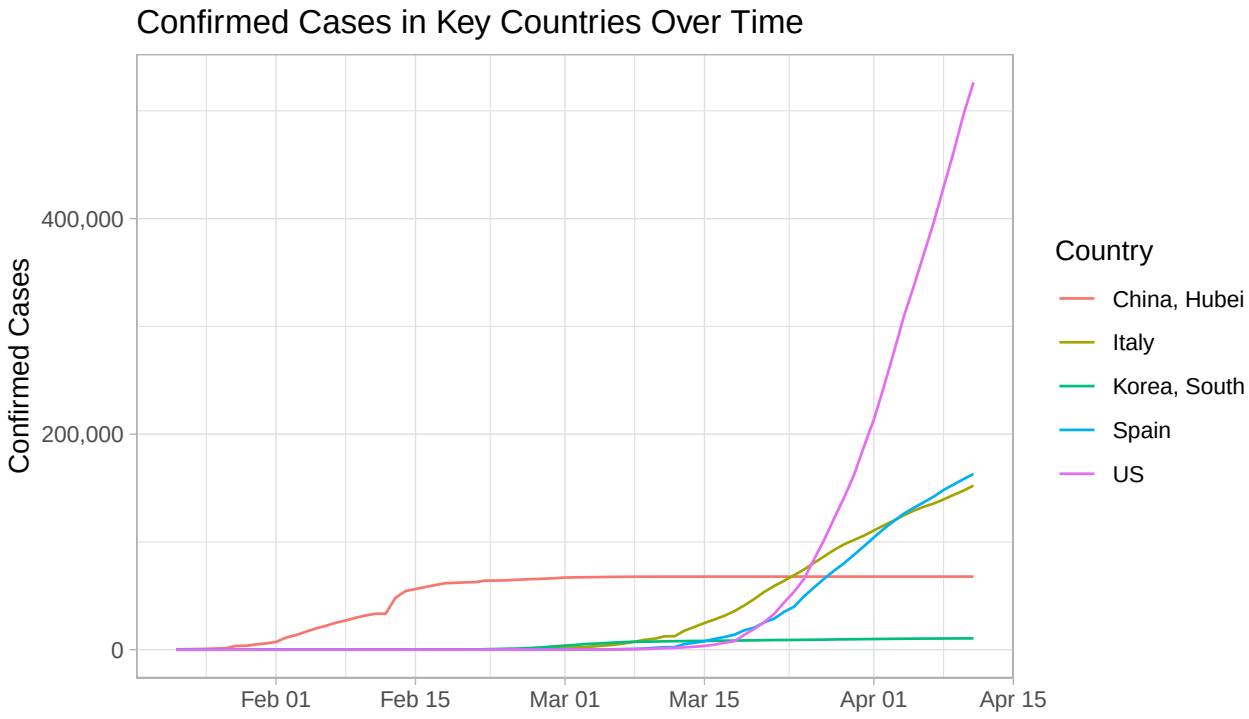
Prediction Using COVID-19 Data from Johns Hopkins

Bill Kayser

Contents

This is an attempt to predict when recovery from the virus begins in different regions. 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?

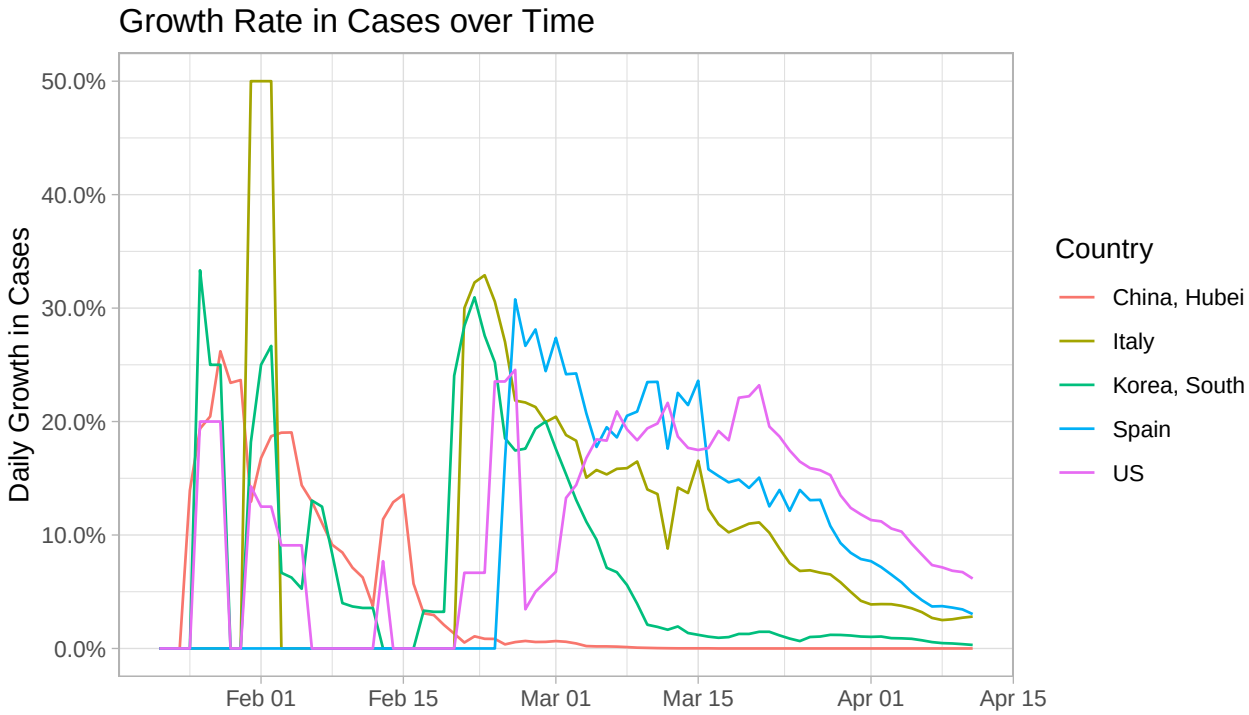
I kept digging through the data until the one thing that stuck out was the rate of growth of new cases. I derived this parameter from the of total Confirmed Cases shown in the plot above. I called it "Growth" and

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.



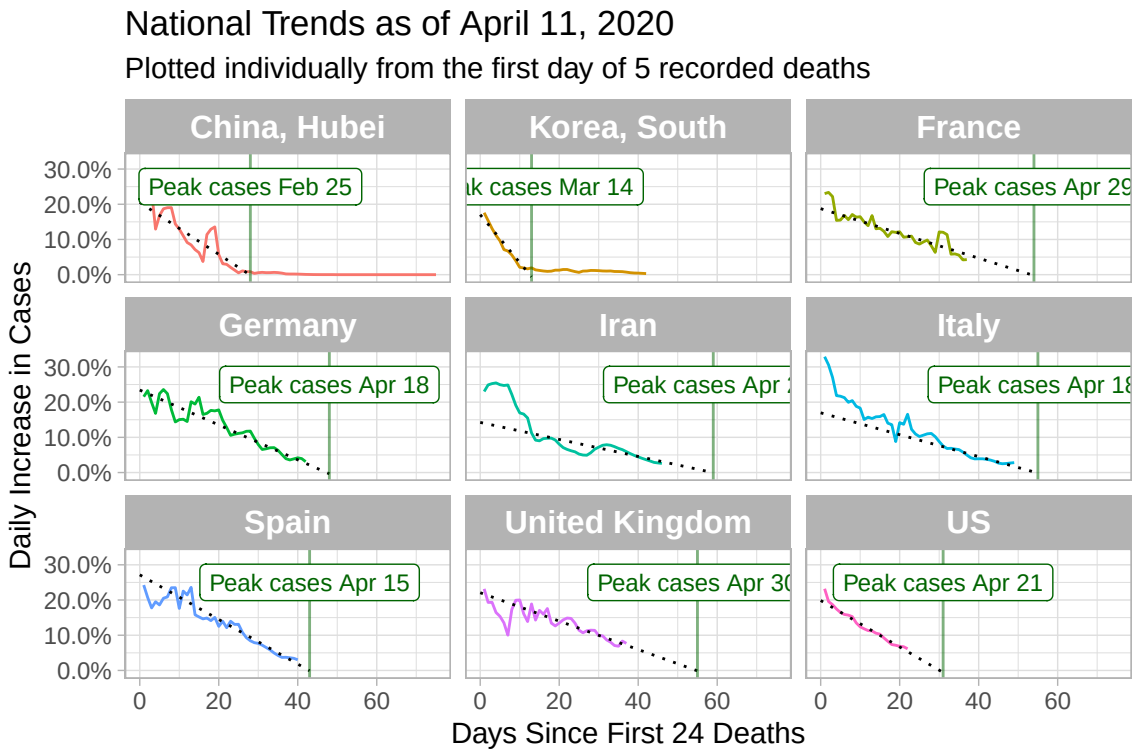
You can see this rate trending toward zero. In China and South Korea i'ts already pretty much leveled off at a point that represented a turnaround in both countries. It was late February in China and the second week of March for South Korea. The other countries seem to be on their way to zero.

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

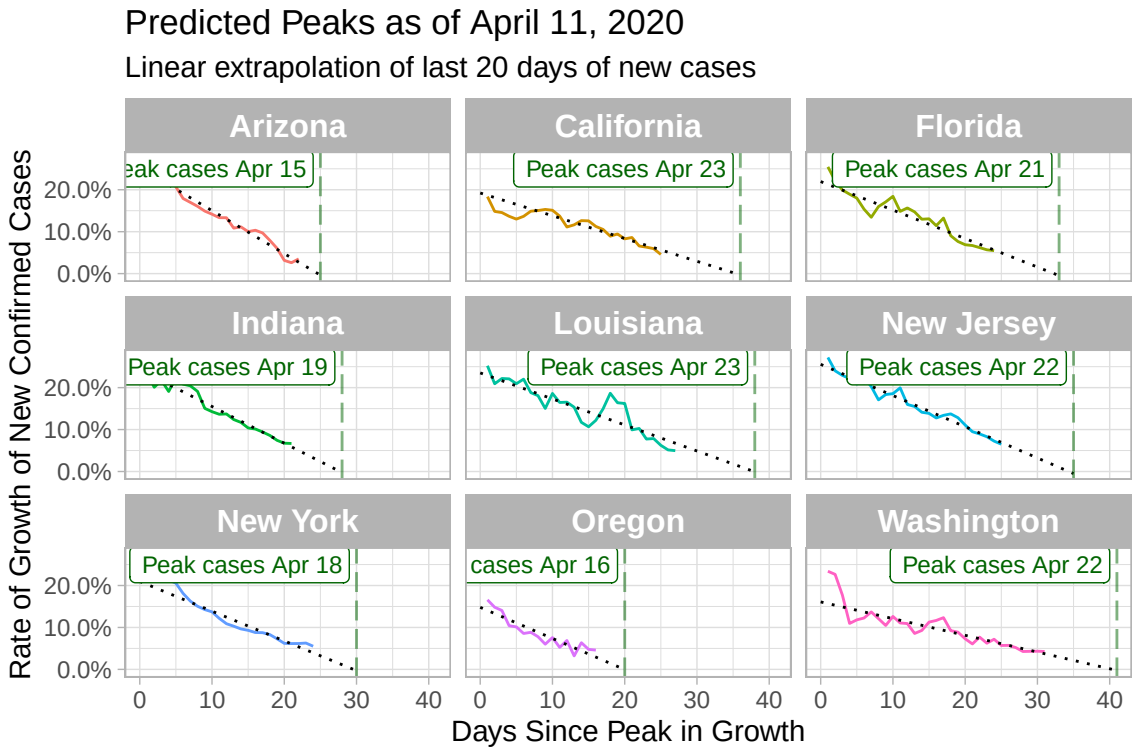
I started by analyzing several countries. I took a cue from a chart in the Financial Times to compare these charts side by side starting from where the peak in growth rate occurs. I plotted the growth rate for each of these countries and then plotted a linear model for the previous 14 days. For Hubei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



State by State Predictions

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



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

State	Date	Days.From.Now
Alabama	2020-04-24	13 days
Alaska	2020-04-19	8 days
Arizona	2020-04-15	4 days
Arkansas	2020-04-14	3 days
California	2020-04-23	12 days
Colorado	2020-04-20	9 days
Connecticut	2020-04-24	13 days
Delaware	2020-05-10	29 days
District of Columbia	2020-05-11	30 days
Florida	2020-04-21	10 days
Georgia	2020-04-27	16 days
Hawaii	2020-04-16	5 days
Idaho	2020-04-14	3 days
Illinois	2020-04-23	12 days
Indiana	2020-04-19	8 days
Iowa	2020-04-26	15 days
Kansas	2020-04-20	9 days
Kentucky	2020-04-16	5 days
Louisiana	2020-04-23	12 days
Maine	2020-04-18	7 days
Maryland	2020-05-09	28 days
Massachusetts	2020-04-23	12 days
Michigan	2020-04-18	7 days
Minnesota	2020-04-21	10 days
Mississippi	2020-04-19	8 days
Missouri	2020-04-18	7 days
Montana	2020-04-15	4 days
Nebraska	2020-04-25	14 days
Nevada	2020-04-25	14 days
New Hampshire	2020-04-15	4 days
New Jersey	2020-04-22	11 days
New Mexico	2020-04-21	10 days
New York	2020-04-18	7 days
North Carolina	2020-04-22	11 days
North Dakota	2020-04-21	10 days
Ohio	2020-04-19	8 days
Oklahoma	2020-04-17	6 days
Oregon	2020-04-16	5 days
Pennsylvania	2020-04-25	14 days
Rhode Island	2020-05-06	25 days
South Carolina	2020-04-23	12 days
South Dakota	2022-03-16	704 days
Tennessee	2020-04-19	8 days
Texas	2020-04-29	18 days
Utah	2020-04-18	7 days
Vermont	2020-04-18	7 days
Virginia	2020-05-02	21 days
Washington	2020-04-22	11 days
West Virginia	2020-04-26	15 days
Wisconsin	2020-04-20	9 days
Wyoming	2020-04-15	4 days