

Corona Virus: Answering Questions I have

Ritwik Sinha

3/24/2020

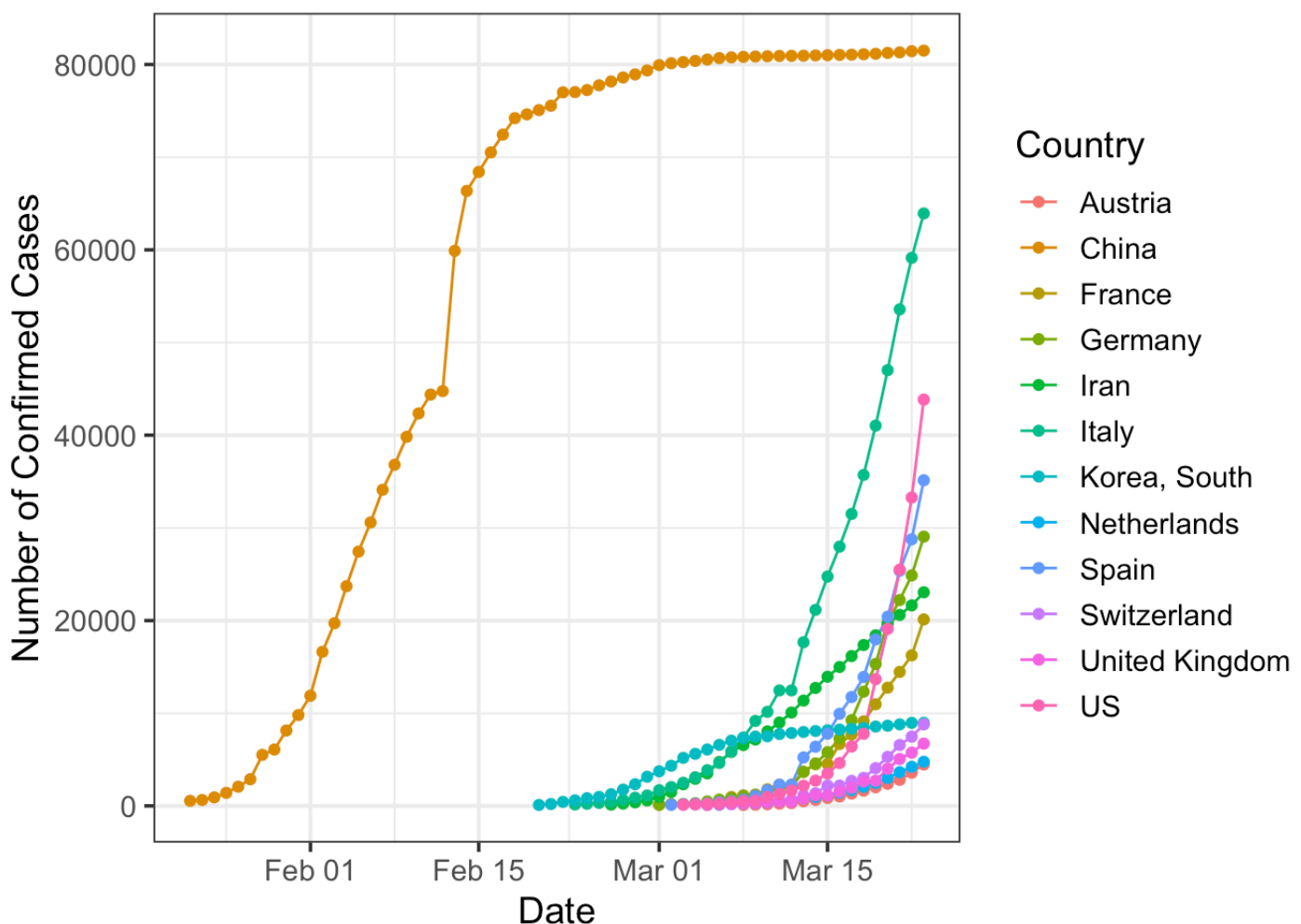
Let's import the data

This is based on the data that is stored and updated daily on this Git Repository

<https://github.com/CSSEGISandData/COVID-19> (<https://github.com/CSSEGISandData/COVID-19>). Please clone this repo, and get the data locally, if you want to rerun this code.

Confirmed Cases

Let's look at the daily counts by country. Below are the counts in their natural scale.

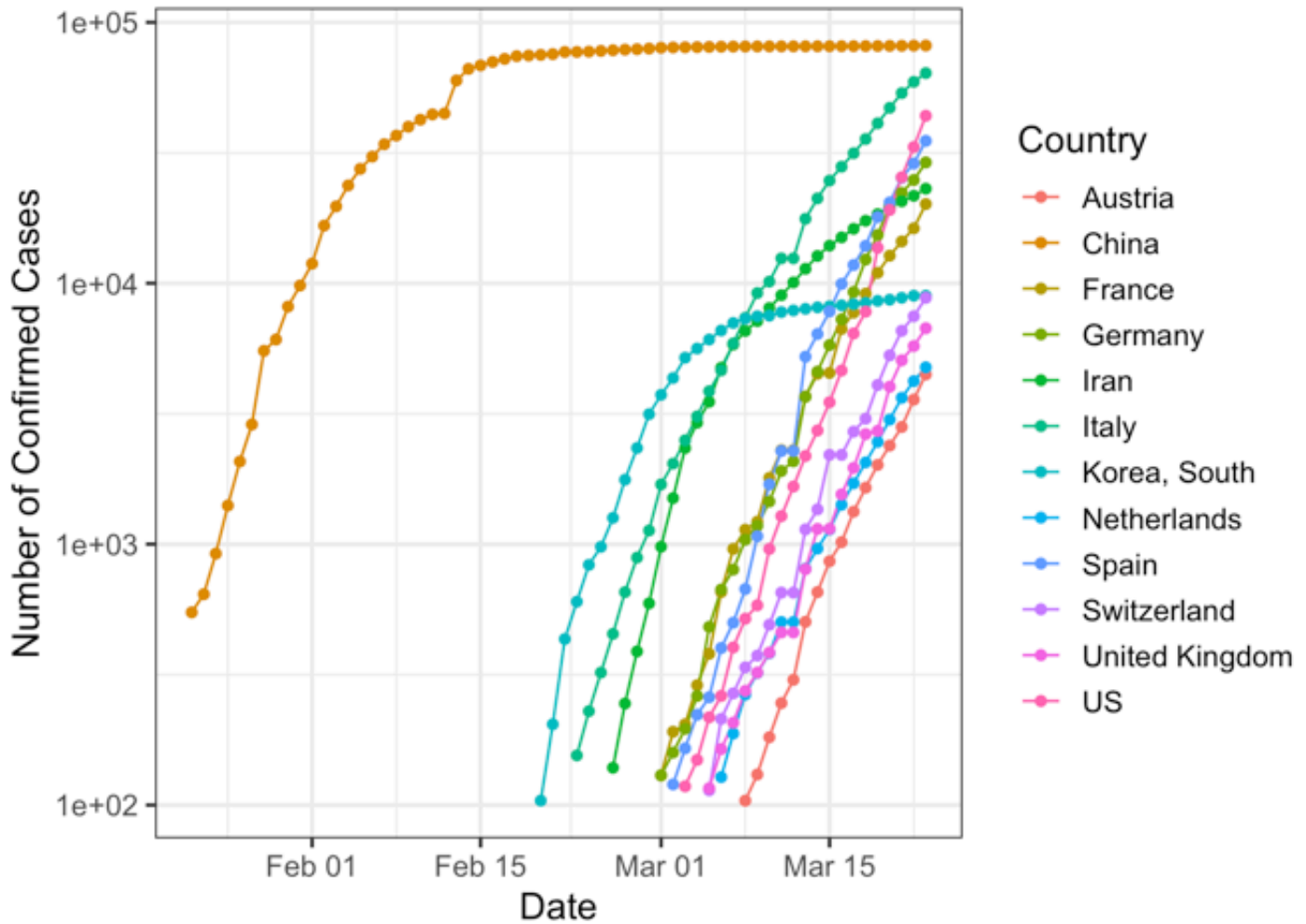


Below are the top countries by cases as off today, 2020-03-24.

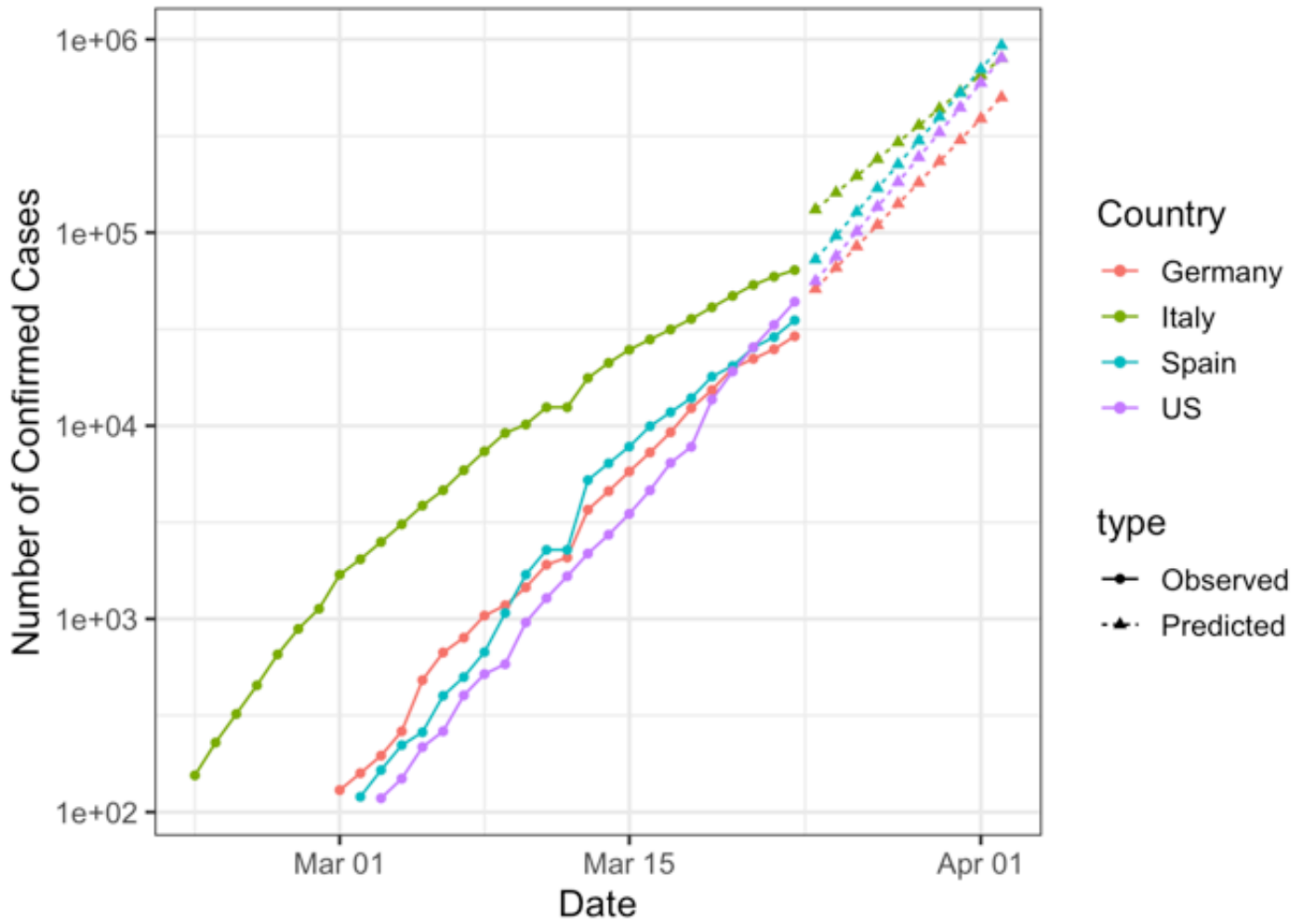
The top countries by cases

Country	total
China	81498
Italy	63927
US	43847
Spain	35136
Germany	29056
Iran	23049
France	20123
Korea, South	8961
Switzerland	8795
United Kingdom	6726
Netherlands	4764
Austria	4474
Belgium	3743
Norway	2621
Canada	2088

Next, let's look at the data on the logarithmic scale. Apart from a few countries, like China, South Korea, Iran, and even Italy, all other countries seem to be very close to a straight line. Suggesting an almost perfect exponential growth in cases.

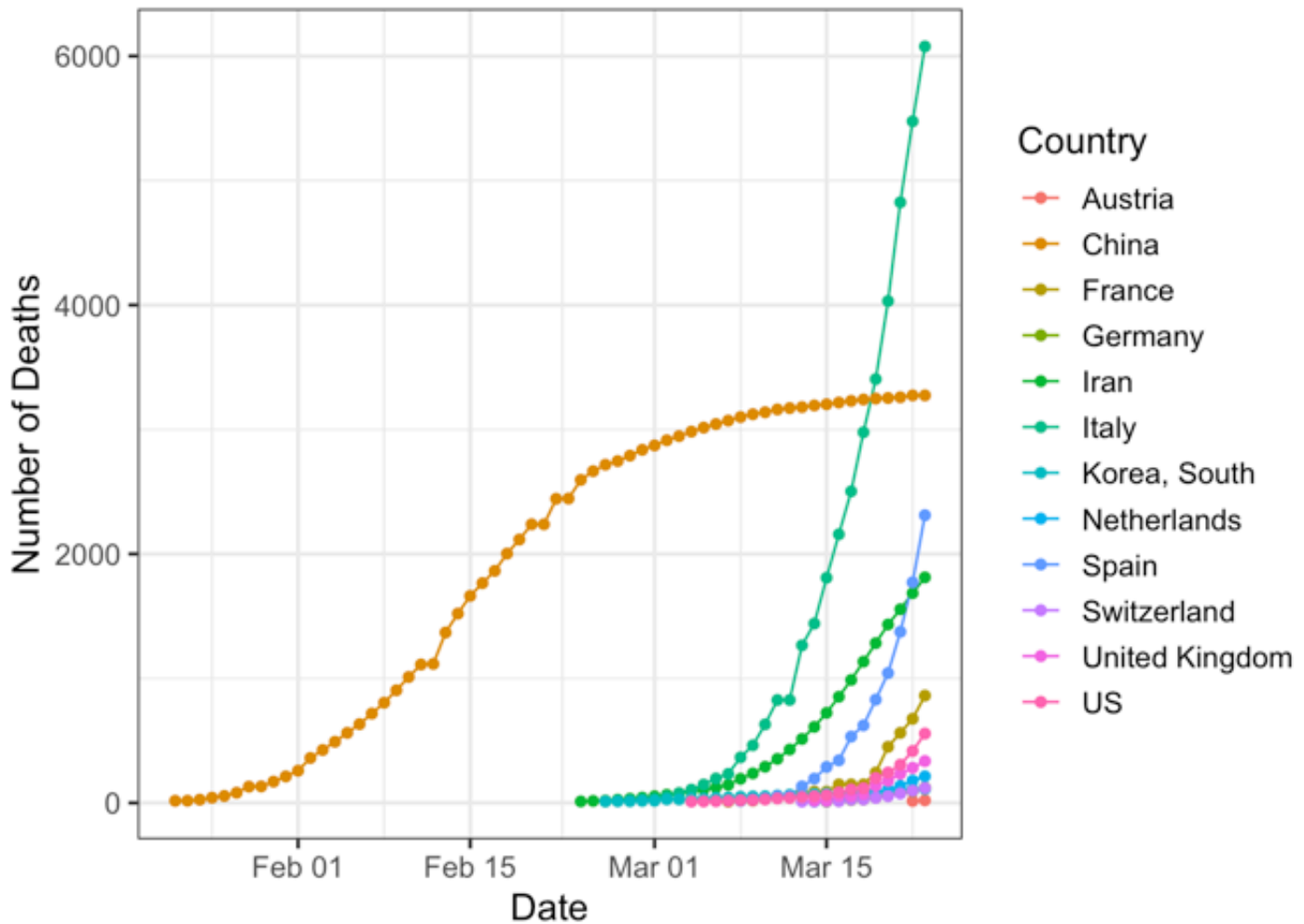


Given how good the exponential fit seems to be, I next perform a regression to predict the future number of cases by country. This is a very simple model, a country specific regression, where we fit a linear regression to the logarithm of the number of cases. In the absence of significant efforts, like the lockdowns that are currently in place, each of the countries below are likely to see million plus cases, in just 10 days.



Deaths

Let's look at the daily counts deaths by country. Below are the counts in their natural scale.



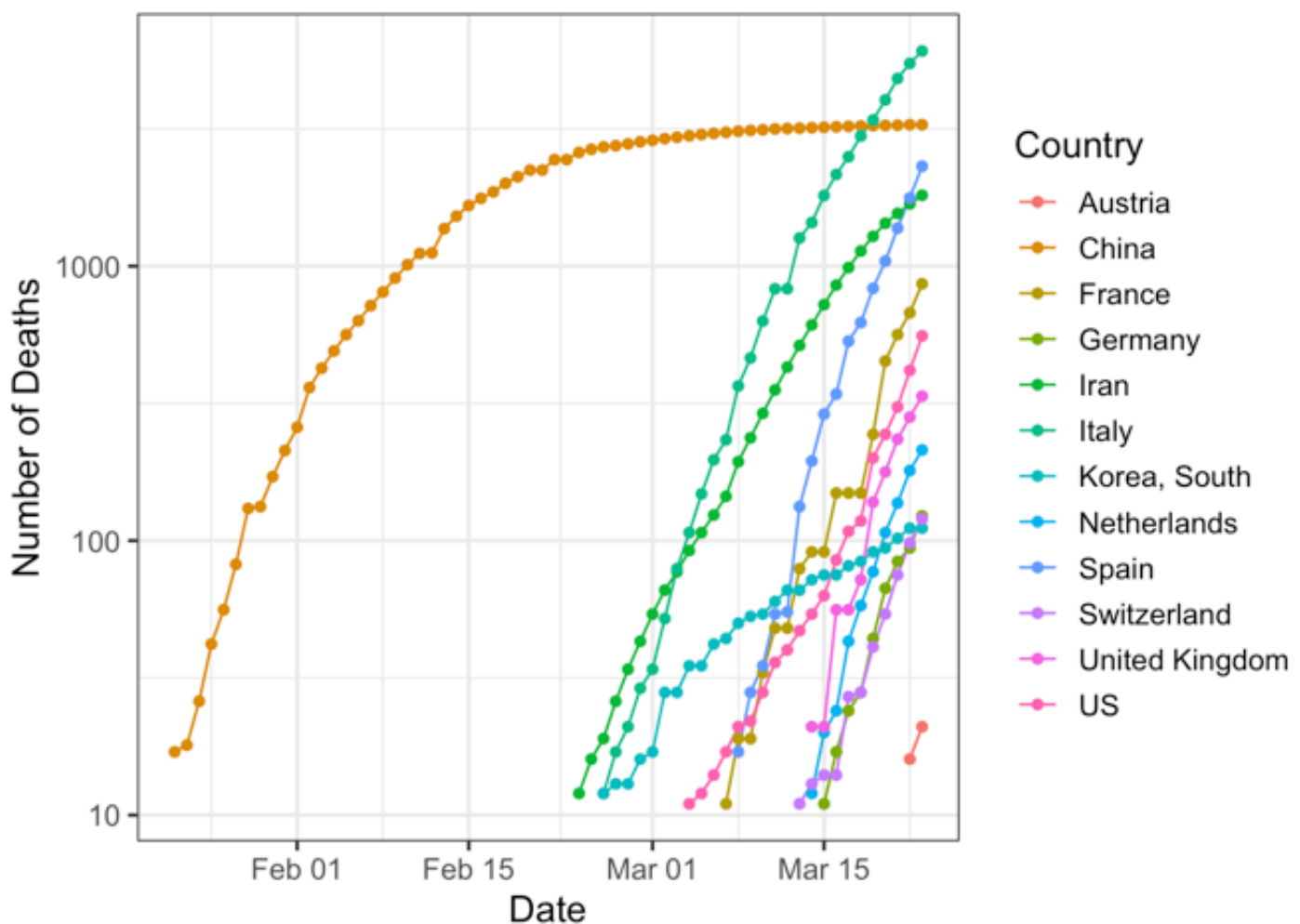
Below are the top countries by deaths as off today, 2020-03-24.

The top countries by cases

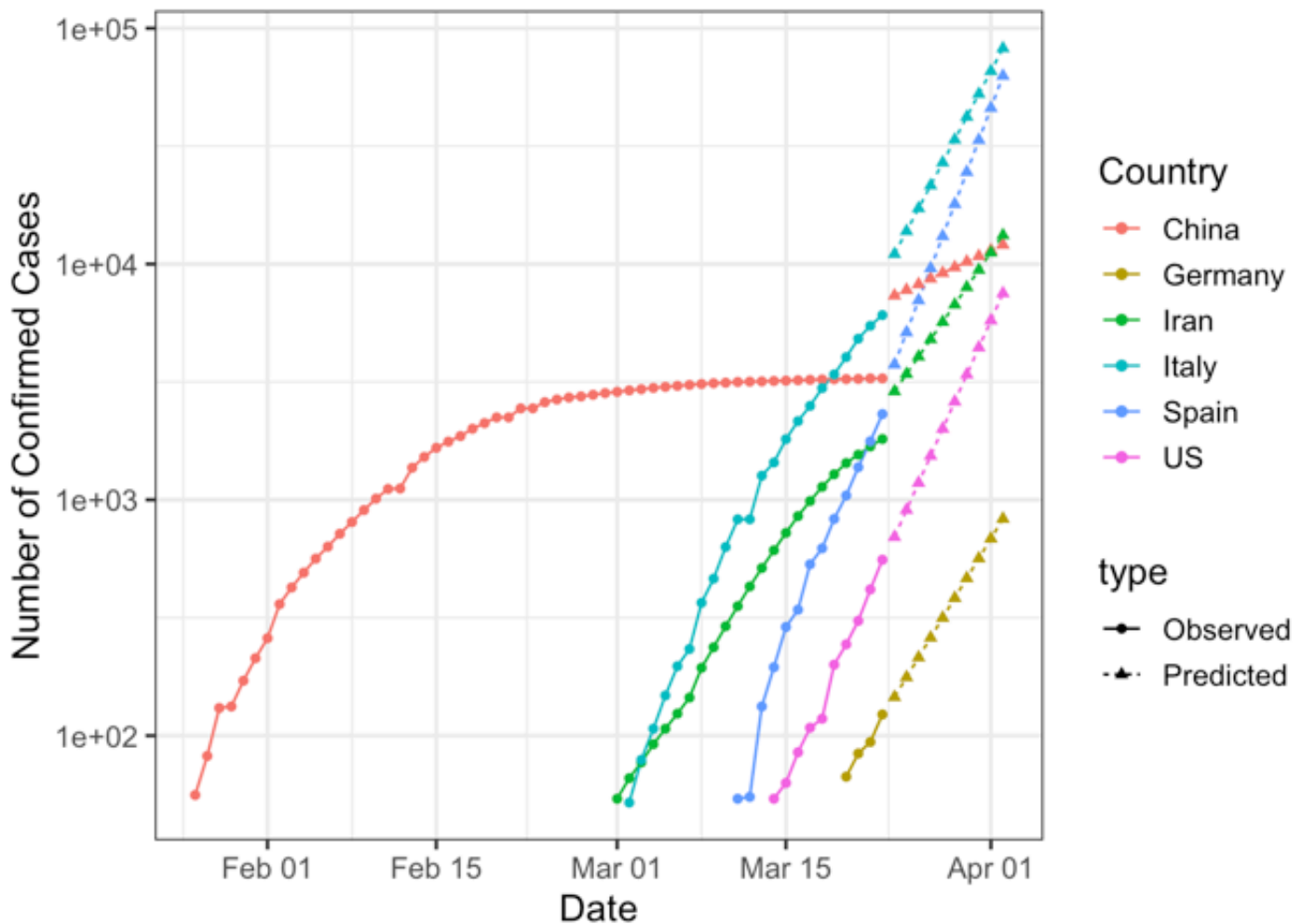
Country	total
Italy	6077
China	3274
Spain	2311
Iran	1812
France	862
US	557
United Kingdom	336
Netherlands	214

Germany	123
Switzerland	120
Korea, South	111
Belgium	88
Indonesia	49
Japan	42
Turkey	37

Let's look at the same plot on the logarithmic scale.



We next perform the same exercise as before, fitting a simple linear regression and extrapolating it. These fits are not good for China, and Iran, because these countries have flattened the curve. Even Italy seems to be doing better. But at this rate, the US will have **10,000** deaths within 10 days.



Deaths across all countries

Next, let's analyze the deaths for all countries. Again, I fit a simple model to capture the pattern. I fit a degree 3 polynomial to capture the pattern thus far, and extrapolate it for the next 15 days. Of course, this does not capture the effect of the social distancing efforts that are currently on-going all over the world. But if nothing were done, this could be a very likely scenario. The scary thing with this fit is that the regression has an Adjusted R-squared of 0.9986, showing how good a fit this simple regression model is to the data.

