COVID 19- Predictions

Facebook's Prophet library

The goal of the poject was fairly simple, if we can model successfully model and predict the growth of cumilative covid-19 infections. I documentation for the project is simple too as the R Script is failry self explainatory with proper comments. I will add some plots for this documentation alongside some summary statistics of the model.

Downloading the file

Anyone can easily download the dataset to replicate this project by downloading the data set in csv format from here.The datset is updated everyday, otherwise a cran package called COVID19 can be used as well. Also, I will try to add the screenshots from Johns hopkins COVID tracker to see if our predictions are close.

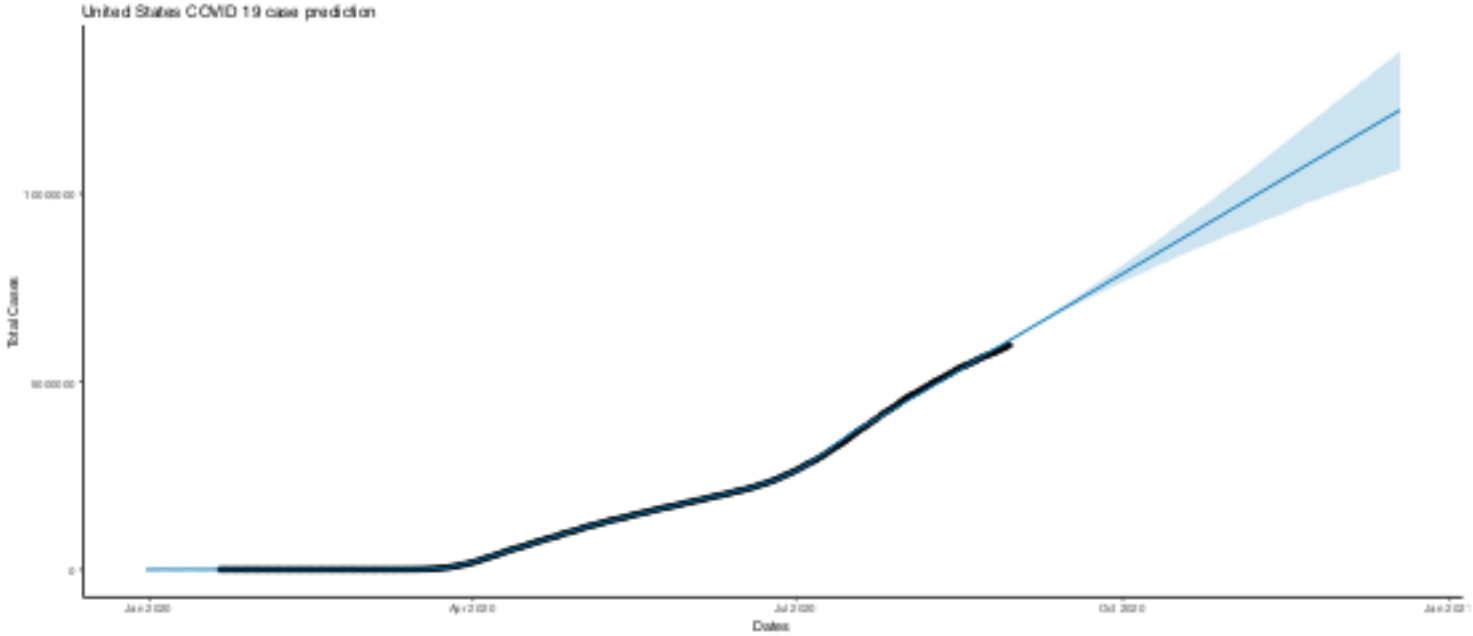
```
## Download and use the package in R, do not support older versions of R
install.packages("COVID19")
library(COVID19)
```

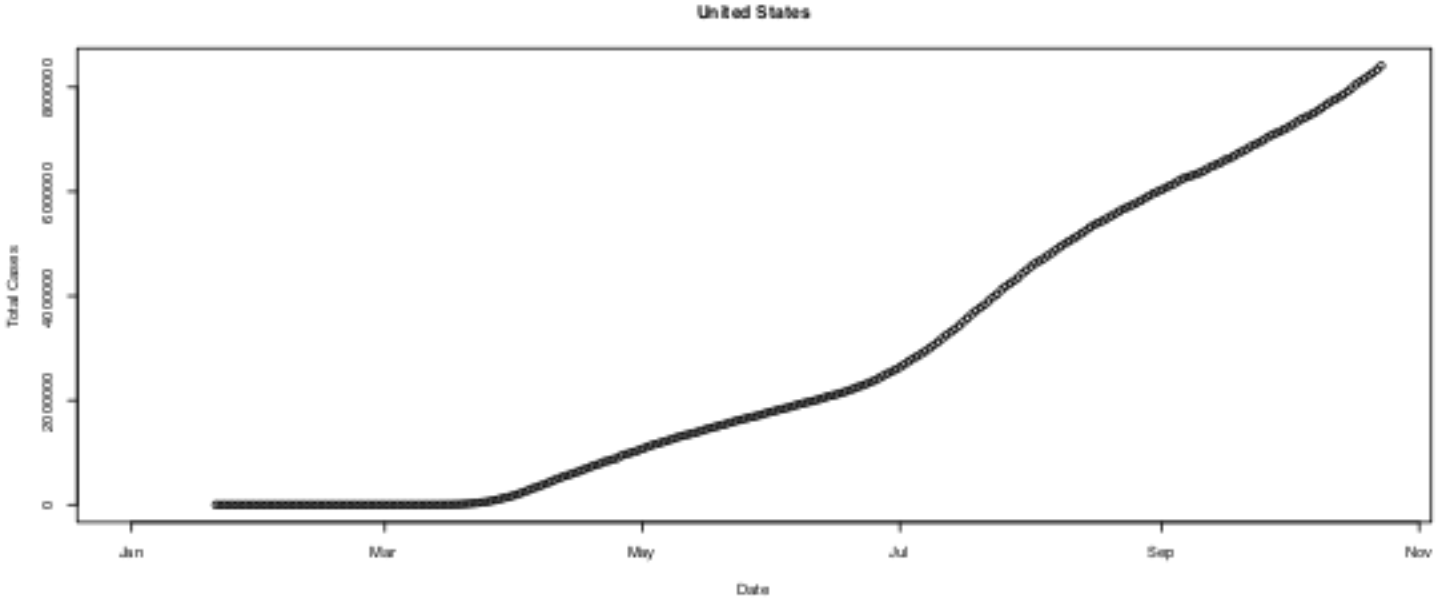
I used 3 countries for the purpose of this project, United States, United Kingdom and India. The reason behind this was to also measure if the models perform differently when applied on different countries. The first figure for each of the prediction shows the predictions with a confidence band, followed by the real numbers and a plot of actual vs predicted values.

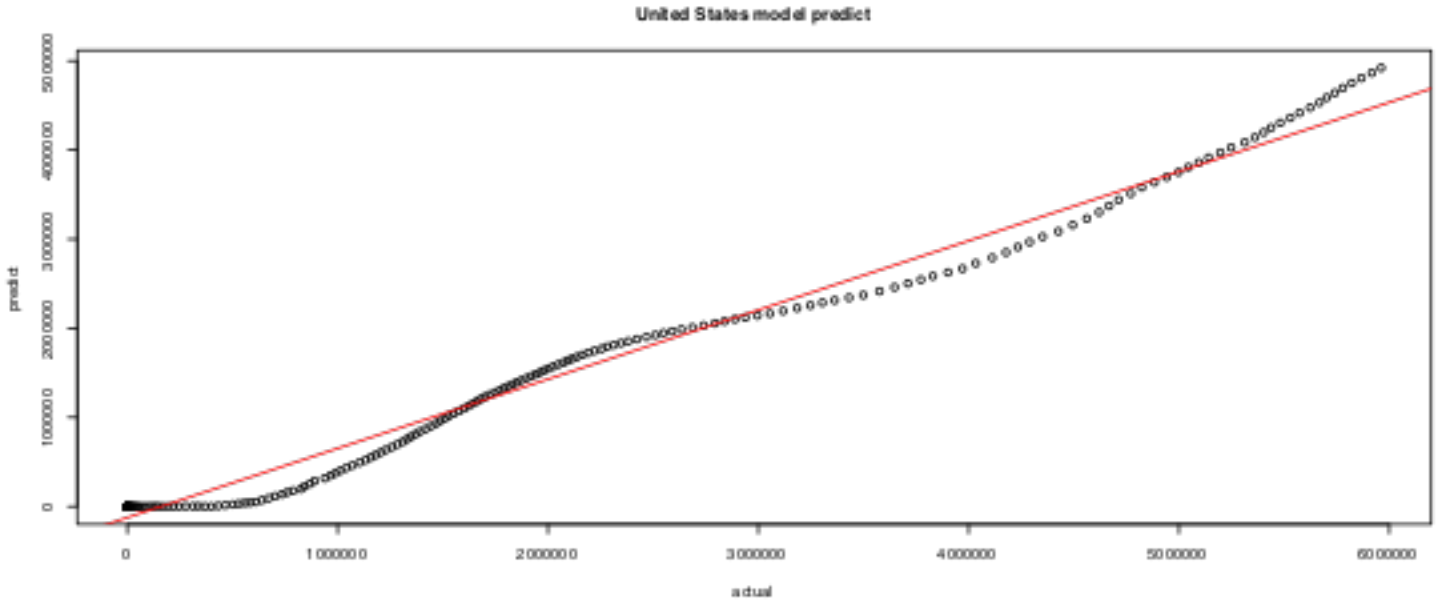
For the purpose of the analysis I created a subset of the dataset ending in August of 2020, so that we can see accuracy of the model.

United States

The data predicts that the number of total cases would hover around 8 million total infections and so does the next plot which is the real recorded cases in October,2020. Then I also plotted the actual values and the predicted values to see if the we have a linearity betwen actual and predictors, plot 3 shows that the model is very good as the plots are so close to the red line which is the abline.

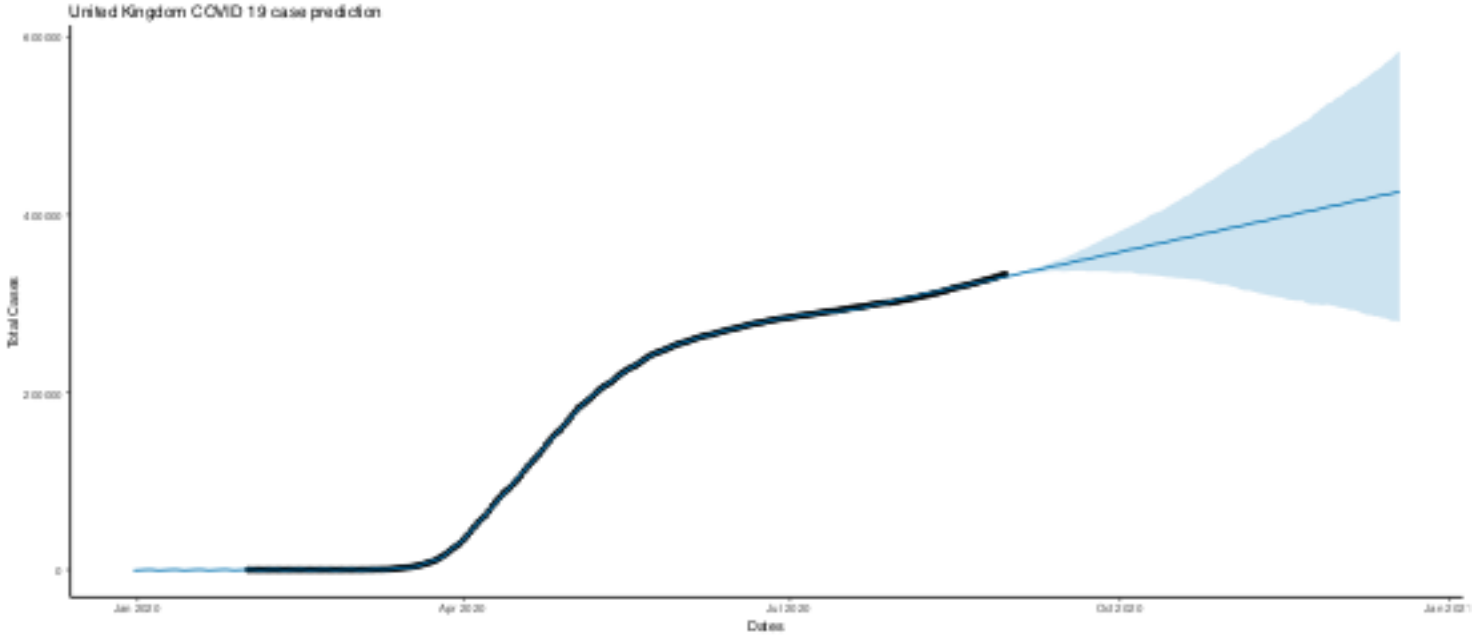

United States COVID 19 case prediction
Total Cases
Dates

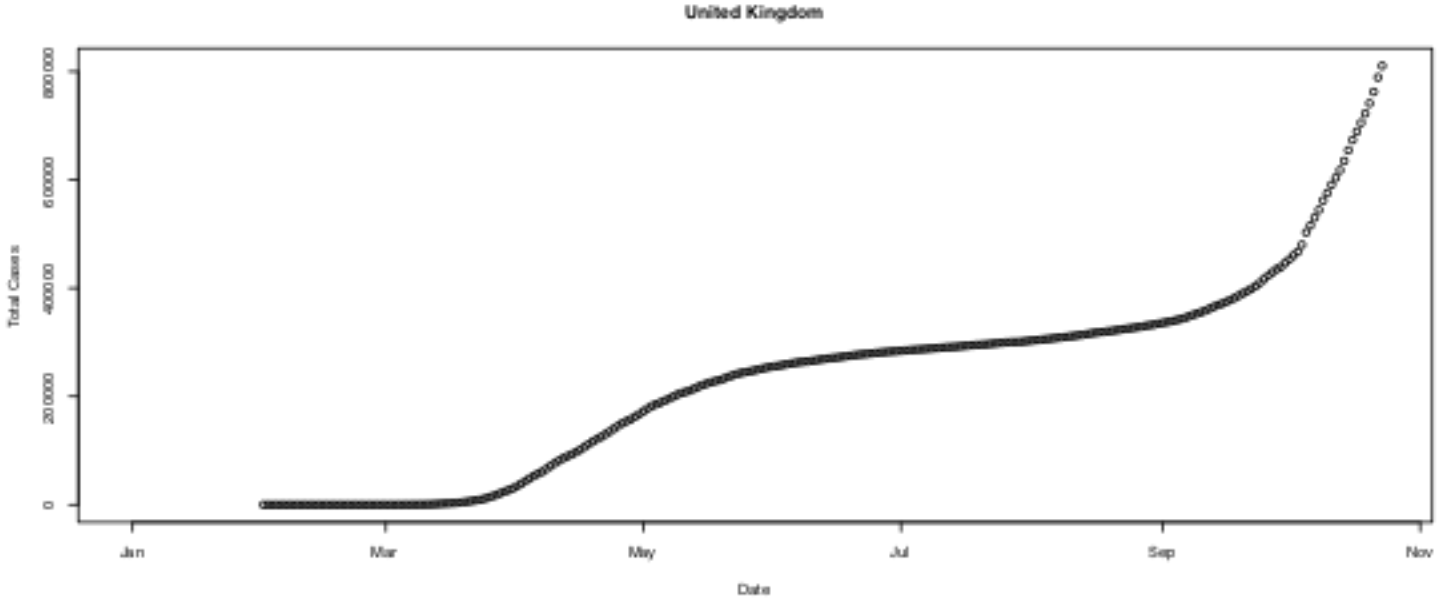

United States
Total Cases
0
2000000
4000000
6000000
8000000
Jan
Mar
May
Jul
Sep
Nov
Date

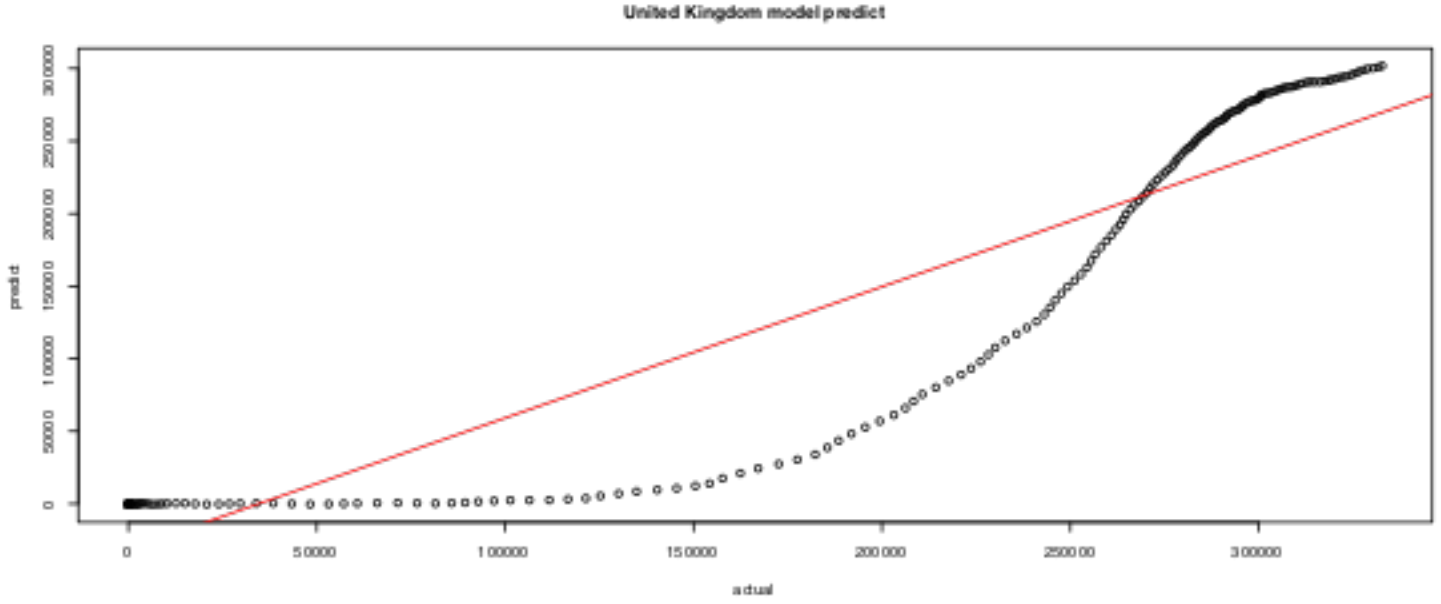


United Kingdom

On the other hand we can see that the confidence bad for the prediction model for United Kingdom is farely wide. We also see from the second plot that United Kingdom managed to flatten the curve very well. Hence, the prediction is shows a wider confidence band compared to the United States model. Also, when the real data shows a much higher number of infections compared to our model. Moreover, the the third plot which plots actual and predictions shows that the model is not linear at all.

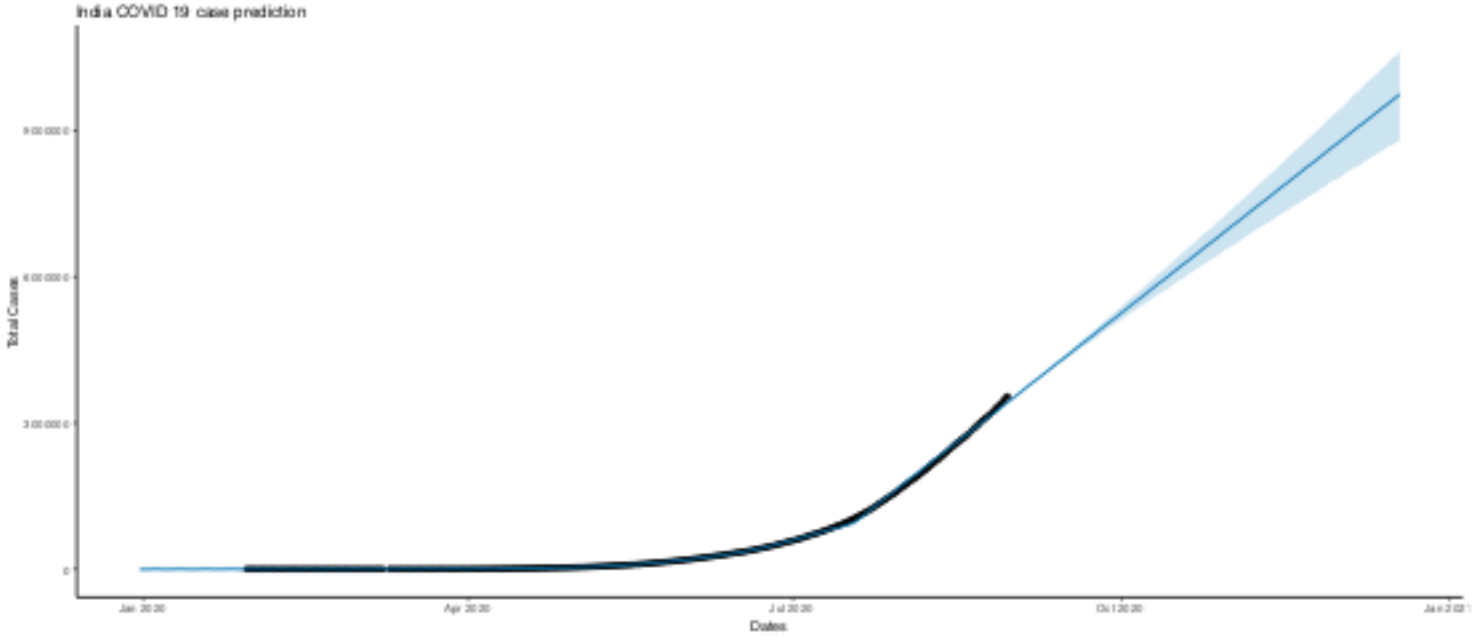


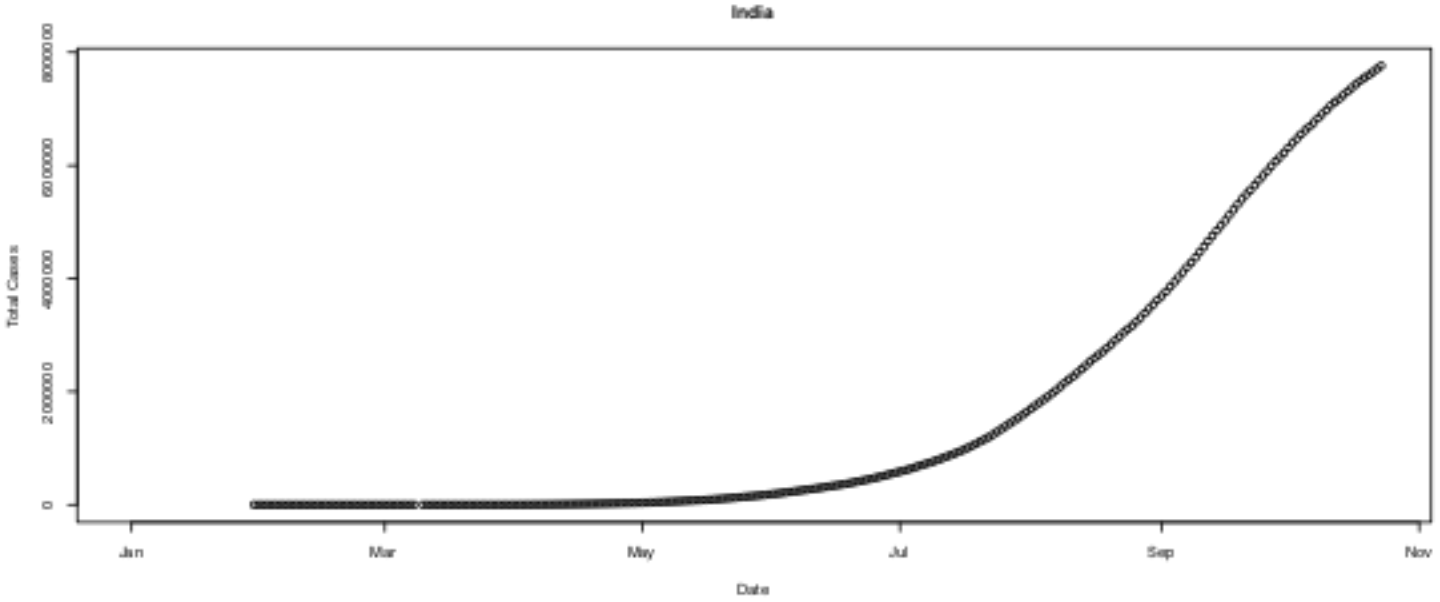


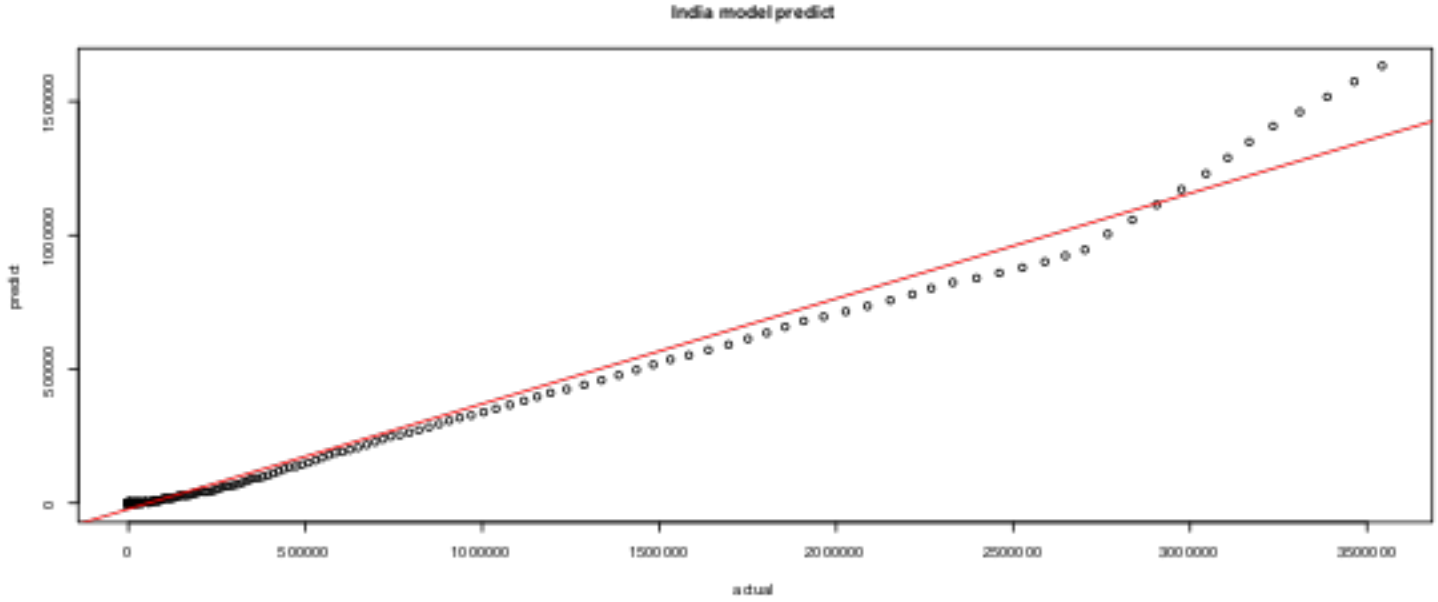


India

This model is similar to the US model.


India COVID 19 case prediction
Total Cases
Dates


India
Total Cases
0
2000000
4000000
6000000
8000000
Jan
Mar
May
Jul
Sep
Nov
Date



You can also check summary statistics of these models using the following pieces of code.

```
summary(lm(predict~actual))
```

I will be updating this at regular intervals.