Trend Analysis and Forecasting COVID-19

INTRODUCTION

This Covid 19 Data Analysis project aims to determine the reason and justify the causes by using different statistical techniques. This study provides a broad overview of the time series data of covid confirmed cases and fatality rates globally and compares reported cases in India with other Countries computes the actual growth rate and daily changes for specific locations by estimating the global trends per location. Also, the SARS Cov-2 trend and forecasting for the next 28 days up from Dec 23. 2021 to Jan 19, 2022, is also predicted by applying different statistical tools and techniques

PROBLEM STATEMENT

The novel coronavirus reported at the end of 2019 has impacted almost every aspect of life. This project mainly focuses on analyzing, forecasting, visualizing, and investigating the current coronavirus spread in different countries.

OBJECTIVE

1. To analyze the time-series data of covid confirmed cases and fatality rates globally
2. To compute the actual growth rate and daily changes for specific locations
3. To estimate the global trends per location
4. To compare reported cases in India with other Countries
5. To Predict the Pandemics Trend and forecast its data
6. To visualize the data by using Single Trend (for single location),itrend (interactive trends),mtrend (multiple locations)
7. To perform analysis for Vaccination & Testing and Genomics data

DATASET USED

A Live data set package from R studio named "covid19.analytics," which fetches current covid data from the JHU Github repository, is used for the analysis.

JHU's CCSE repository

https://github.com/mponce0/covid19.analytics

Focused Attributes - (Province_State,Country_Region,Confirmed, Deaths,Recovered and Active cases)

Total Observation – 4006

Total Variables – 40

TOOL DESCRIPTION(LIBRARIES, PACKAGES AND SOFTWARE USED)

Statistical Software used – R Studio

Libraries Used :

- library(covid19.analytics)
- library(plotly)
- library(lubridate)
- library(ggplot2)
- library(prophet)
- library(dplyr)

Techniques implemented:

- Summary Statistics

- ONE -WAY ANOVA
- Regression Analysis
- **SIR** MODEL (Susceptible-Infected-Recovered)
- Pandemics Trend Prediction and forecasting(time in day wise)
- Interactive visulaizations

CODE:

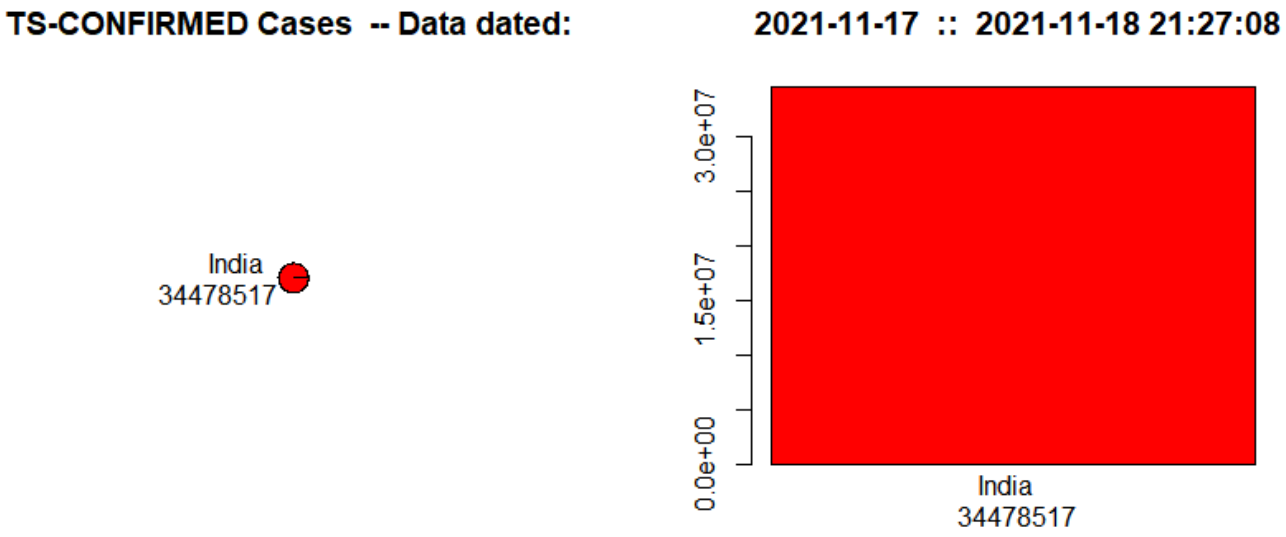
```
library(covid19.analytics)
library(plotly)
library(lubridate)
library(ggplot2)
library(prophet)
library(dplyr)

options(scipen = 6)

covidData <- covid19.data()
## obtain time series data for "confirmed" cases ##
confirmed_cases <- covid19.data(case = "ts-confirmed")
# India Confirmed Cases

confirmed_cases_in <-
  confirmed_cases %>% filter(Country.Region == "India")
print(confirmed_cases_in)


## Death Count / reads time series data for casualties ##
death_cases <- covid19.data(case = "ts-deaths")
# India Death Cases
death_cases_in <- death_cases %>% filter(Country.Region == "India")
print(death_cases_in)

## obtain time series data for "recovered" cases ##.
recovered_cases <- covid19.data(case = "ts-recovered")
# India Recovered Cases
recovered_cases_in <-
  recovered_cases %>% filter(Country.Region == "India")
print(recovered_cases)
```

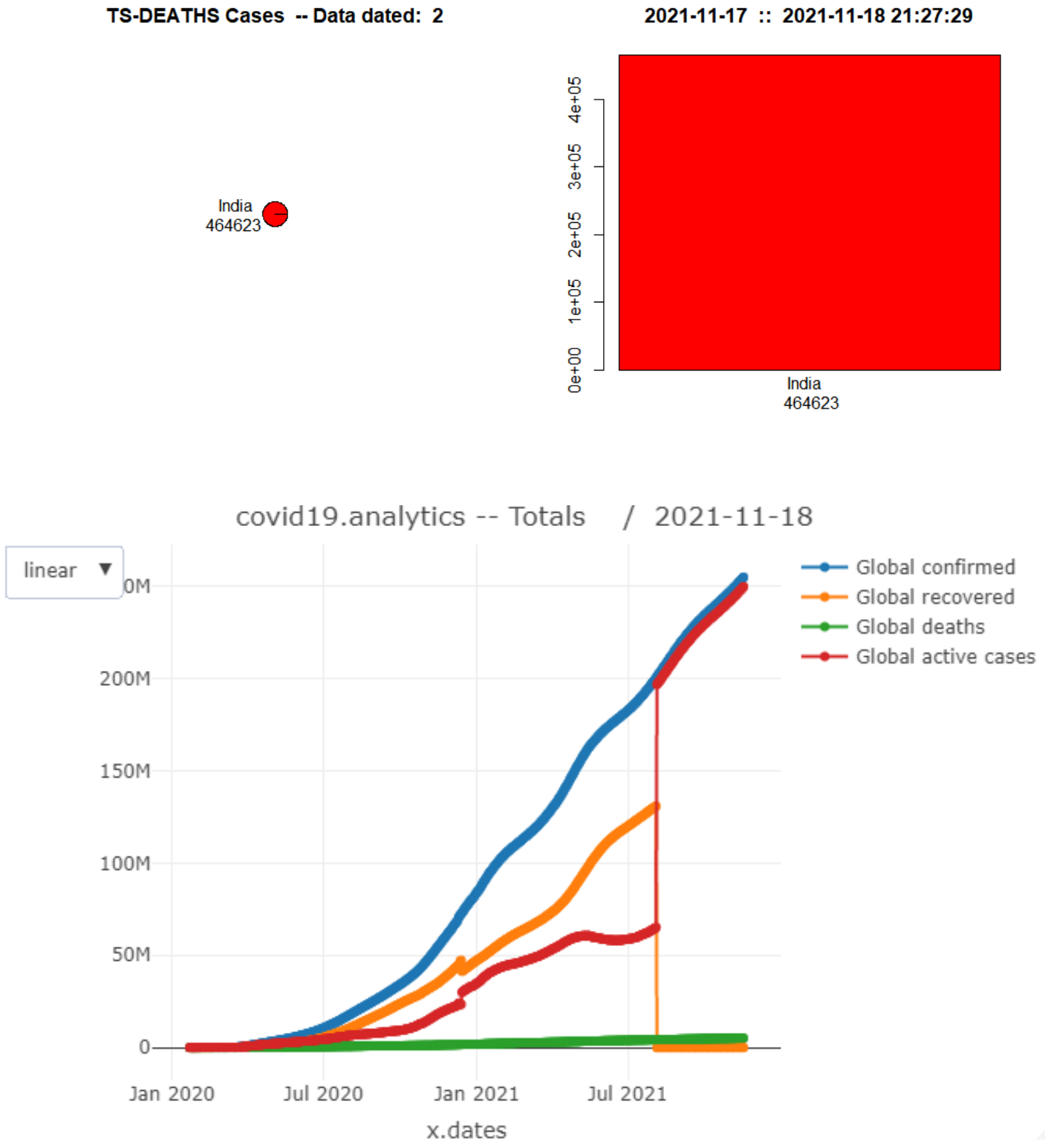
```r
#Summary Statistics

#GLOBAL
report.summary(Nentries = 5,
          graphical.output = T)

report.summary(Nentries = 5,
          graphical.output = T,
          geo.loc = "India")



#one way anova
one.way = aov(Confirmed ~ Deaths, data = covidData)
summary(one.way)


# Graphs and Visualization

total_ts <- covid19.data(case = "ts-ALL")
totals.plt(total_ts)
#totals per location  //Regression Analysis
tots.per.location(confirmed_cases, geo.loc = "India")

# growth rates
growth.rate(confirmed_cases, geo.loc = c('US', 'India'))
growth.rate(confirmed_cases, geo.loc = c('Brazil', 'India'))
#SIR modelling
SIR = generate.SIR.model(confirmed_cases, 'India', tot.population = 34478517)

#livemap
live.map(confirmed_cases)
##############################################################################

# Pandemics Trend Prediction (time in day wise)
tsc <- covid19.data(case = 'ts-confirmed')
tsc <- tsc %>% filter(Country.Region == 'India')
tsc  <- data.frame(t(tsc))
tsc <- cbind(rownames(tsc), data.frame(tsc, row.names = NULL))
colnames(tsc) <- c('Date', 'Confirmed')
tsc <- tsc[-c(1:4), ]
tsc$Date <- ymd(tsc$Date)
str(tsc)
tsc$Confirmed <- as.numeric(tsc$Confirmed)

#Plot
qplot(Date, Confirmed, data = tsc, main = 'Covid19 confirmed cases in India')
```

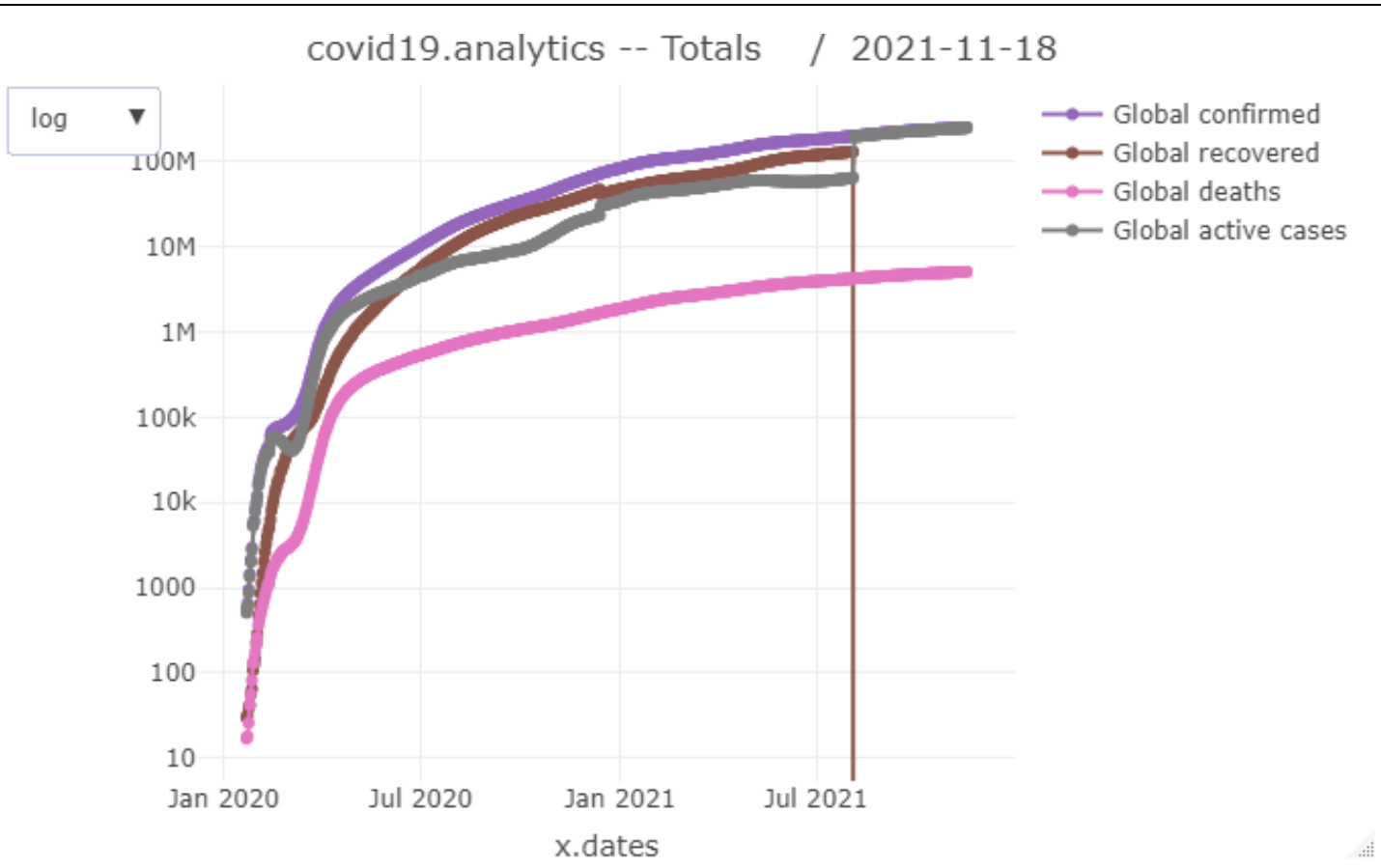
```r
ds <- tsc$Date
y <- tsc$Confirmed
df <- data.frame(ds, y)

# Forecasting
m <- prophet(df)

# Prediction
future <- make_future_dataframe(m, periods = 28)
forecast <- predict(m, future)

#Plot forecast
plot(m, forecast)
dyplot.prophet(m, forecast)

#Forecast components
prophet_plot_components(m, forecast)

#Model Performance

pred <- forecast$yhat[1:121]
actual <- m$history$y
plot(actual, pred)



#Single Trend for India

indiaData <-
  confirmed_cases[confirmed_cases$Country.Region == "India" , ]
single.trend(indiaData)

#Multiple Location


mtrends (confirmed_cases, geo.loc = c ("US", "India"))
#Interactive Locations trend daily cases
itrends (covid19.data("ts-confirmed") , geo.loc = "India")
itrends (covid19.data("ts-confirmed") , geo.loc = "ALL")

#TESTING
covidTest <- covid19.testing.data(tgt = "testing", disclaimer = TRUE)
#covidTest<-covidTest %>% filter(Country.Region == "India")
print(covidTest)

#VACCINATION
covidvaccine <-
```

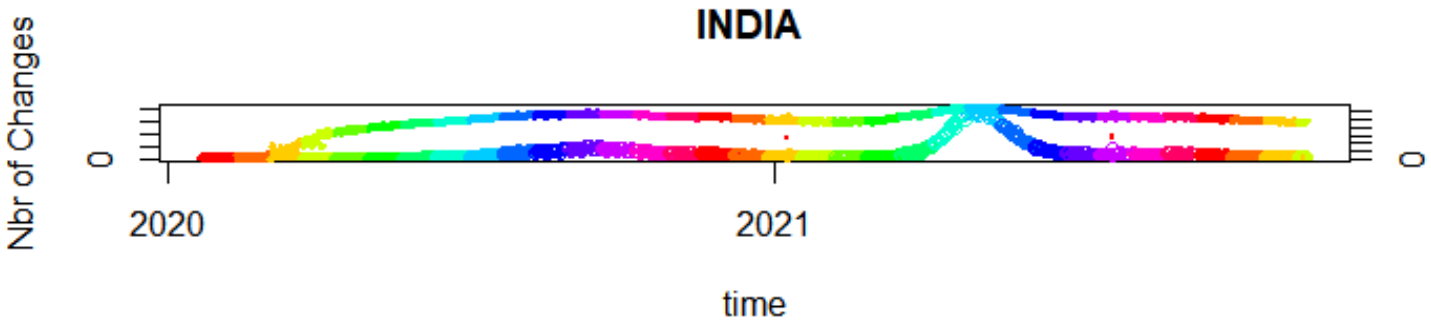
```
 covid19.vaccination(tgt = "global",
                data.fmt = "orig",
                disclaimer = TRUE)
print(covidvaccine)

#GENOMIC
covidgenomic <- covid19.genomic.data(
 type = "genome",
 src = "livedata",
 graphics.ON = TRUE,
 accOnly = TRUE
)
print(covidgenomic)
```

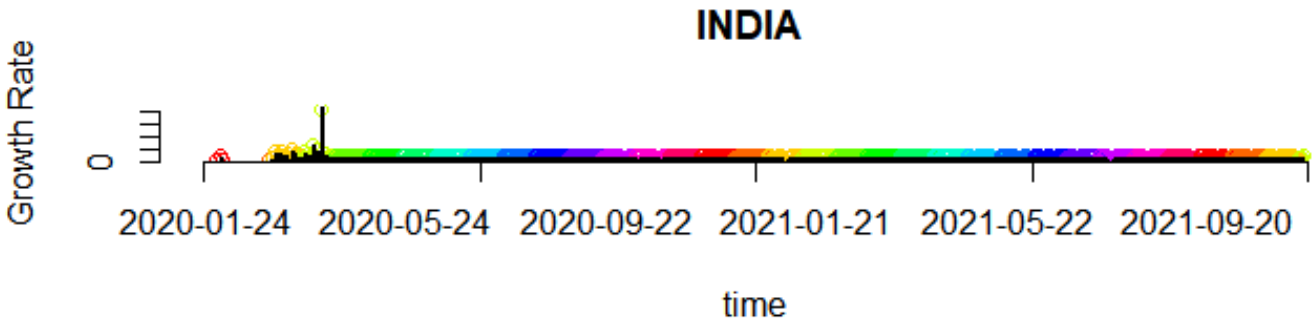
OUTPUT

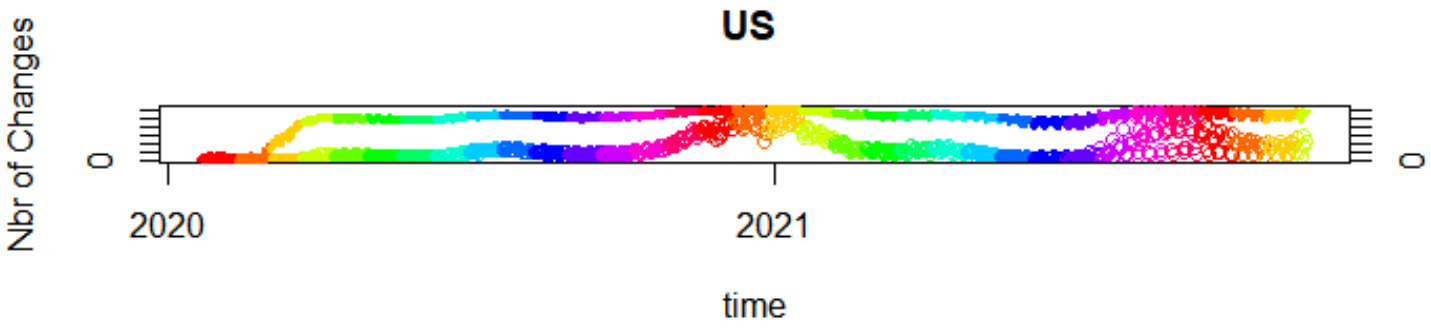


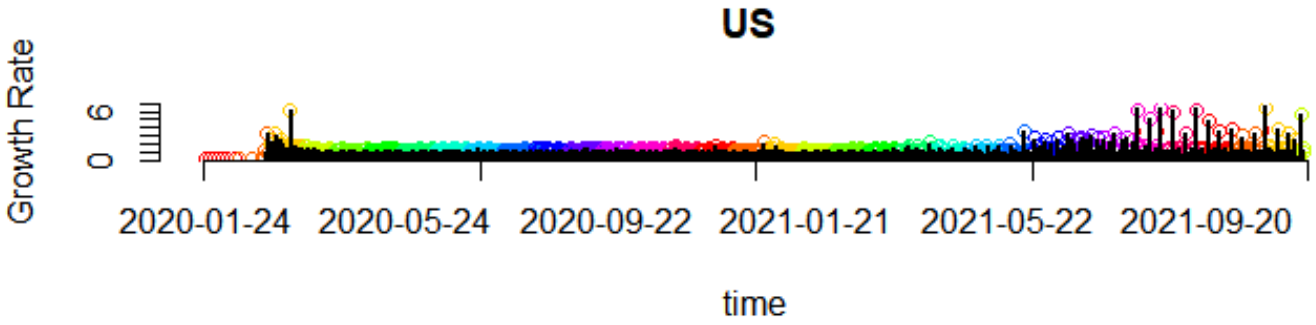

TS-DEATHS Cases -- Data dated: 2
2021-11-17 :: 2021-11-18 21:27:29
India
464623
0e+00
1e+05
2e+05
3e+05
4e+05
India
464623
covid19.analytics -- Totals / 2021-11-18
linear
0M
200M
150M
100M
50M
0
Jan 2020
Jul 2020
Jan 2021
Jul 2021
x.dates
Global confirmed
Global recovered
Global deaths
Global active cases

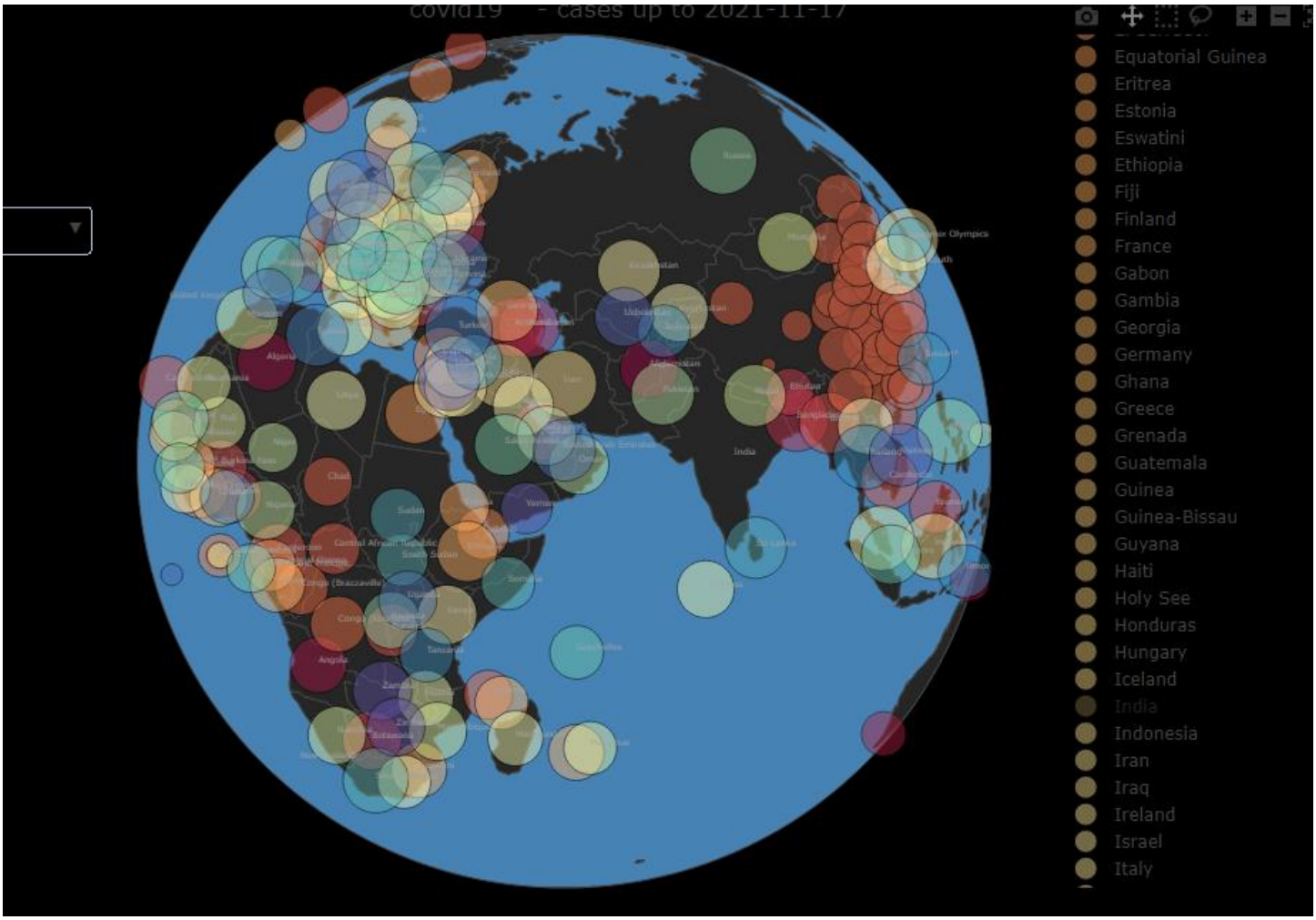

covid19.analytics -- Totals / 2021-11-18
log
Global confirmed
Global recovered
Global deaths
Global active cases
100M
10M
1M
100k
10k
1000
100
10
Jan 2020
Jul 2020
Jan 2021
Jul 2021
x.dates

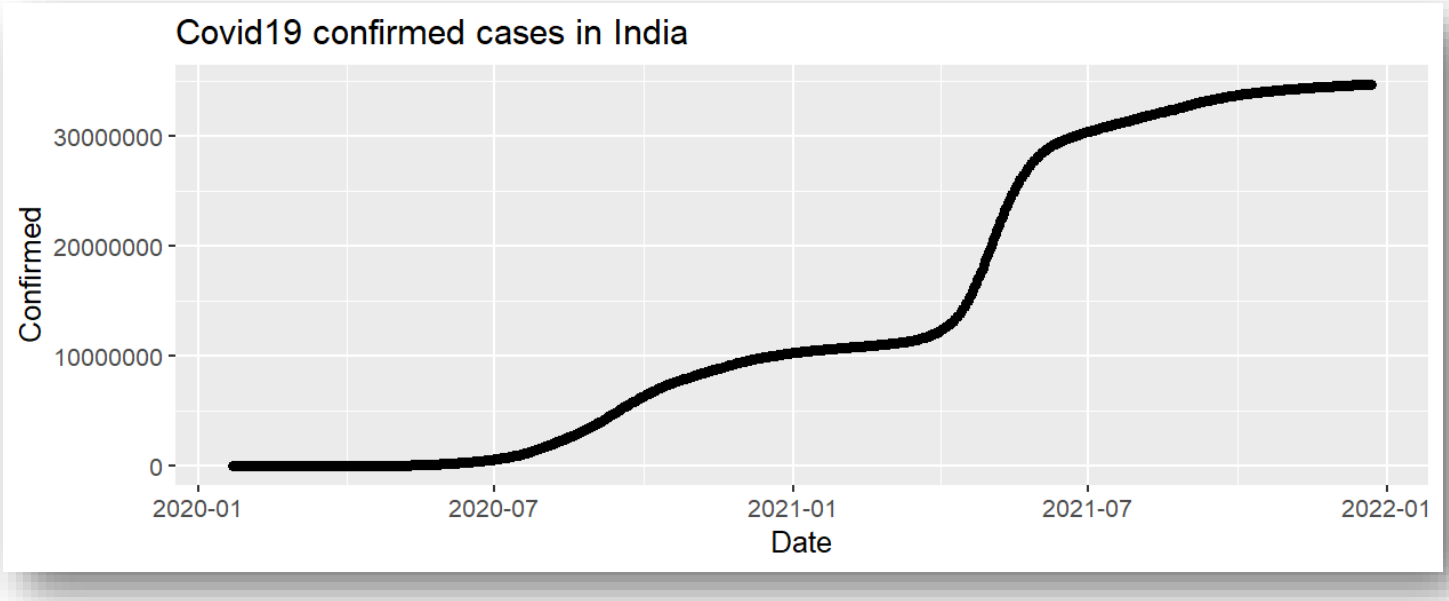

INDIA
Nbr of Changes
0
0
2020
2021
time

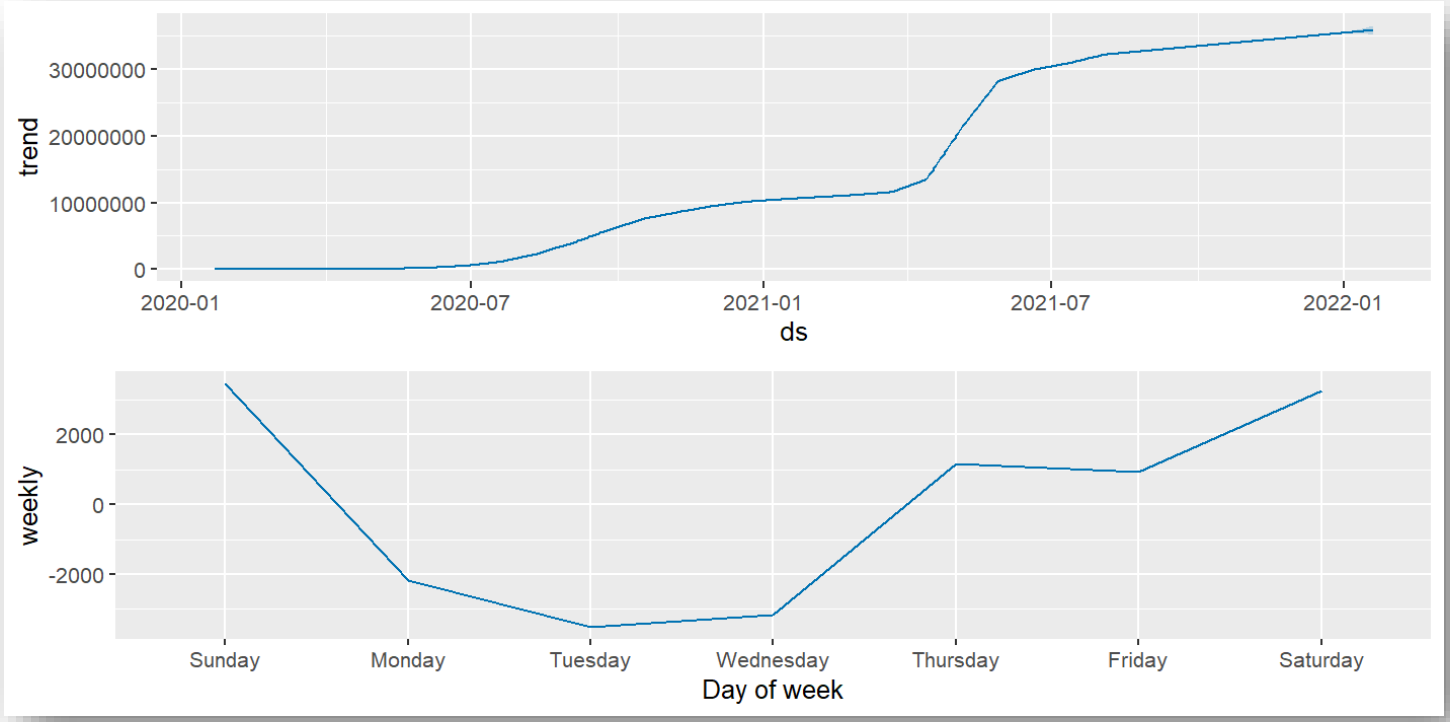

INDIA
Growth Rate
0
2020-01-24
2020-05-24
2020-09-22
2021-01-21
2021-05-22
2021-09-20
time

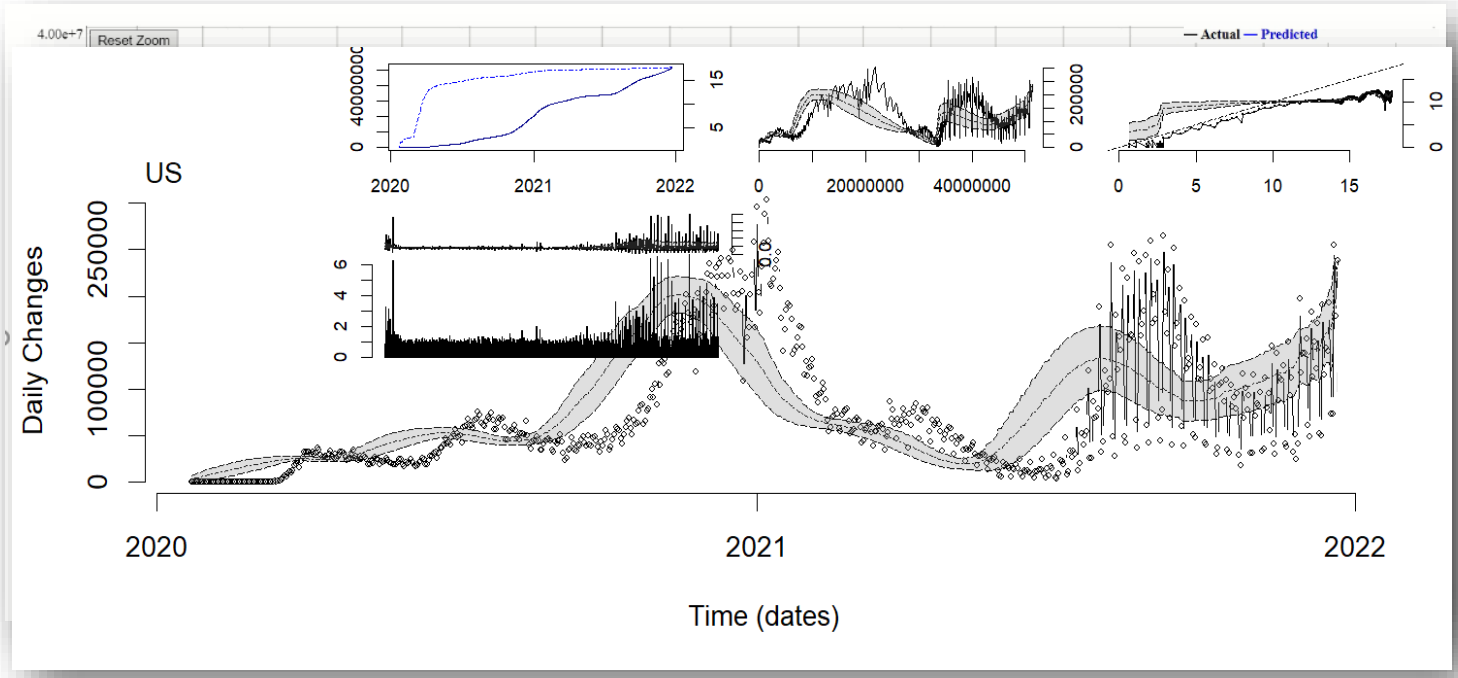

US
Nbr of Changes
0
0
2020
2021
time

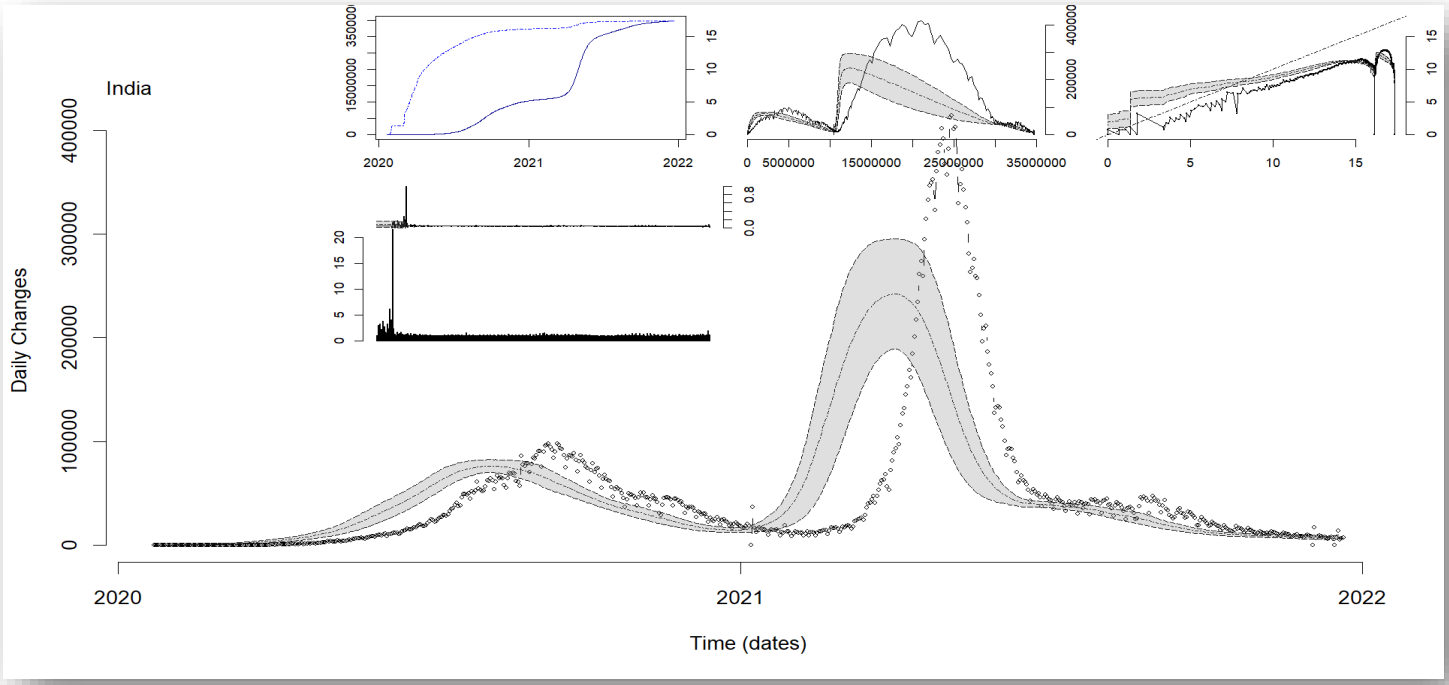

US
Growth Rate
0
6
2020-01-24
2020-05-24
2020-09-22
2021-01-21
2021-05-22
2021-09-20
time

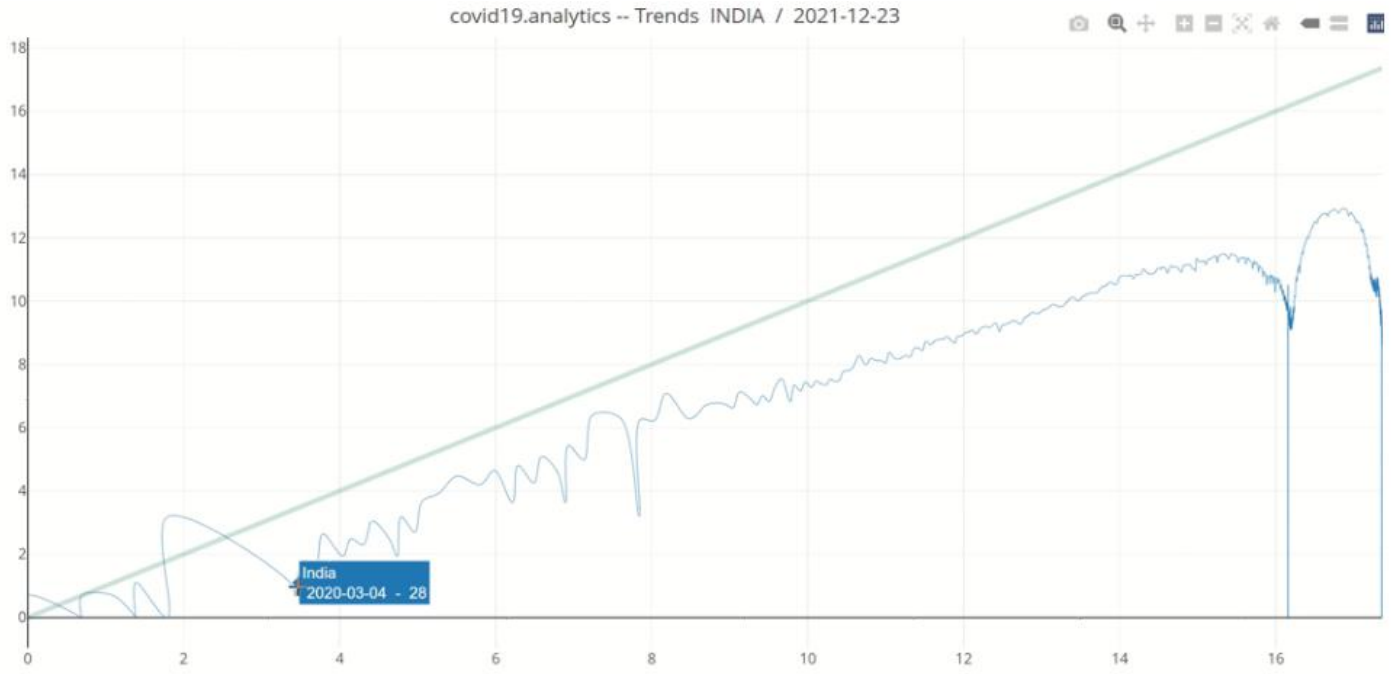

Equatorial Guinea
Eritrea
Estonia
Eswatini
Ethiopia
Fiji
Finland
France
Gabon
Gambia
Georgia
Germany
Ghana
Greece
Grenada
Guatemala
Guinea
Guinea-Bissau
Guyana
Haiti
Holy See
Honduras
Hungary
Iceland
India
Indonesia
Iran
Iraq
Ireland
Israel
Italy

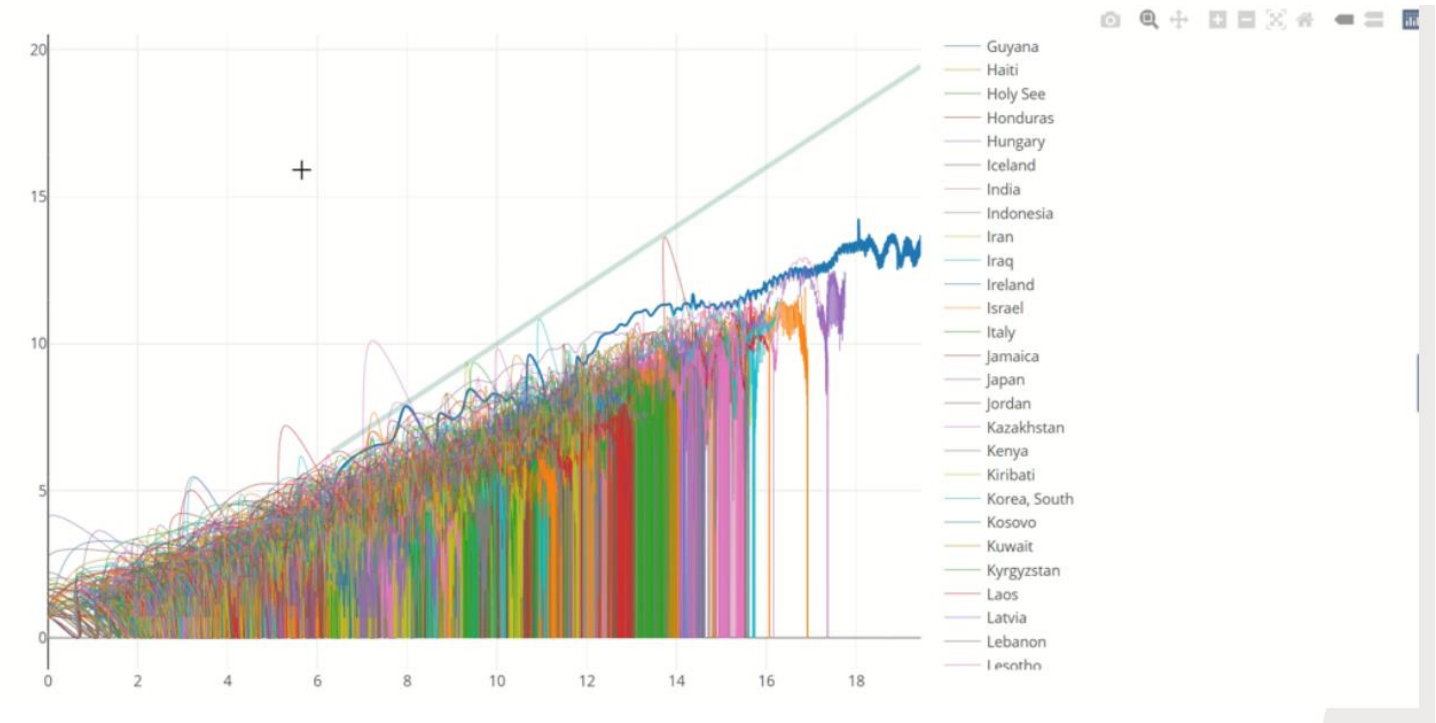

Covid19 confirmed cases in India
Confirmed
30000000
20000000
10000000
0
2020-01
2020-07
2021-01
2021-07
2022-01
Date

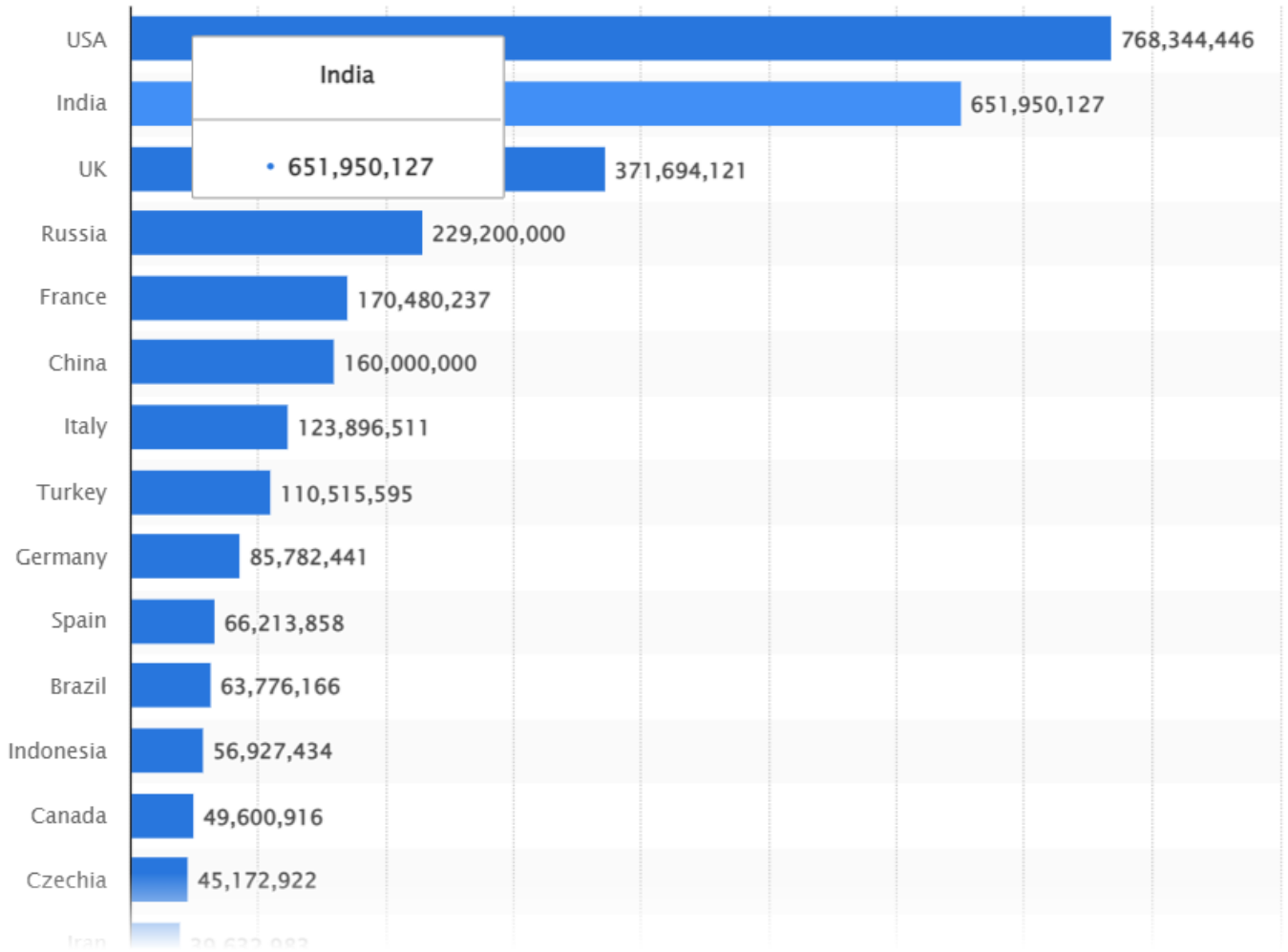

trend
30000000
20000000
10000000
0
2020-01
2020-07
2021-01
2021-07
2022-01
ds
weekly
2000
0
-2000
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Day of week


4.00e+7
Reset Zoom
— Actual — Predicted
US
Daily Changes
0
100000
250000
2020
2021
2022
Time (dates)

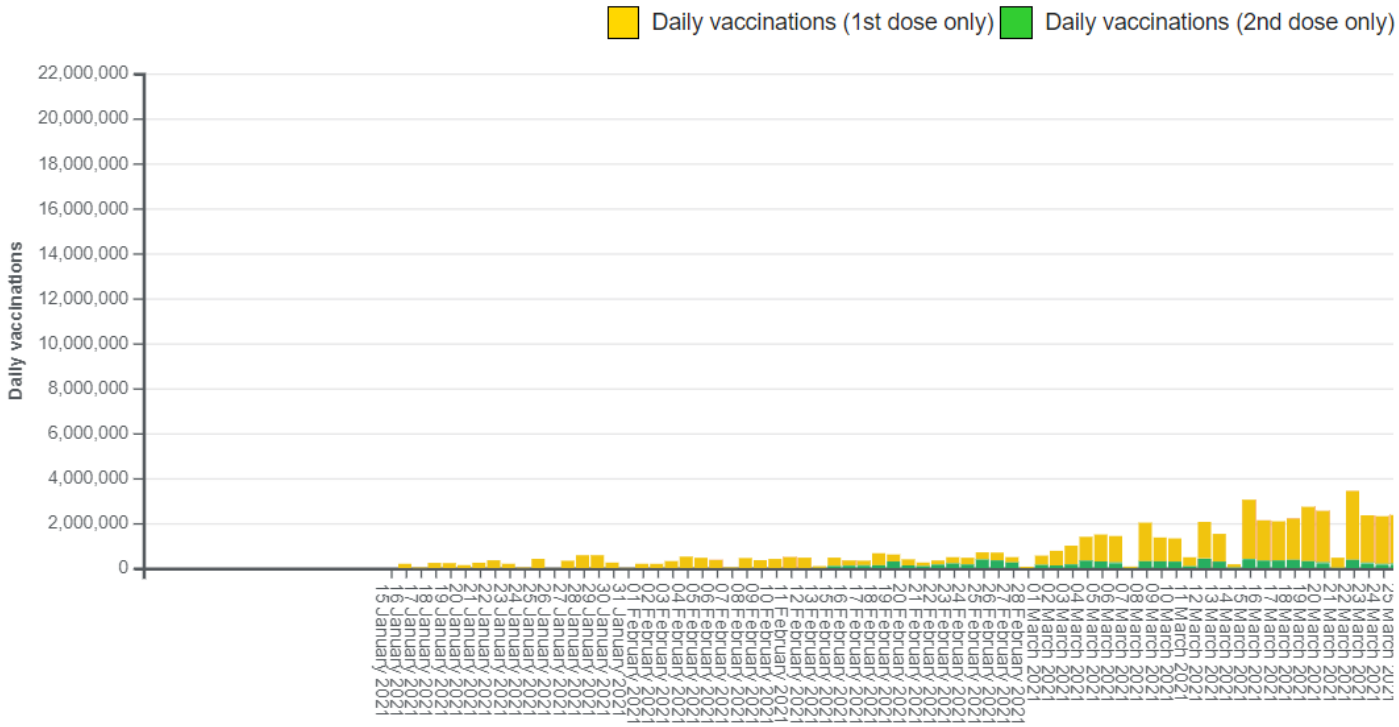

India
Daily Changes
0
100000
200000
300000
400000
2020
2021
2022
Time (dates)

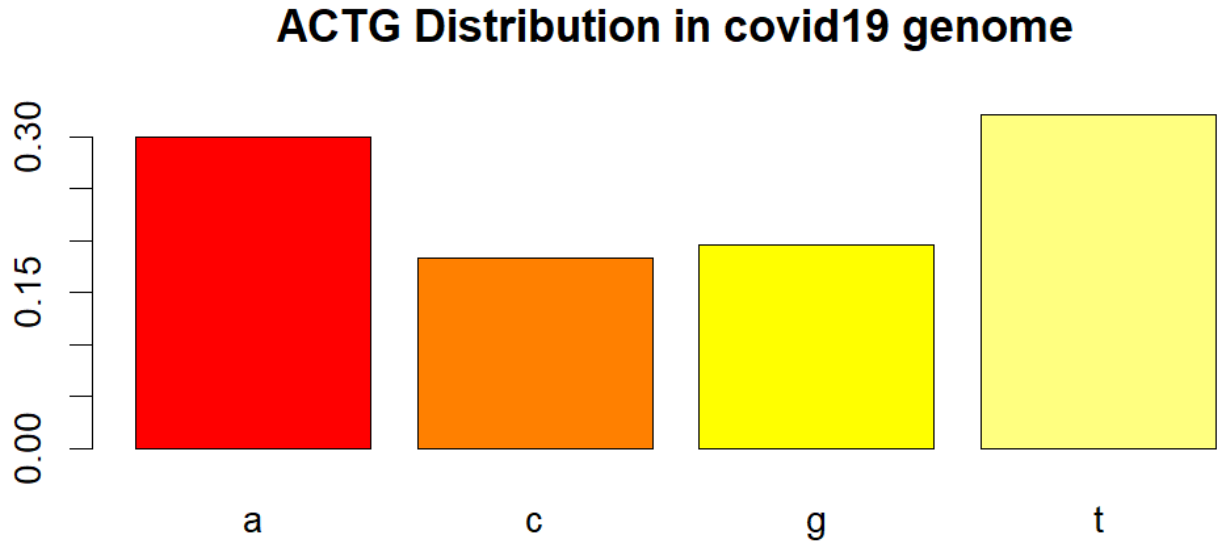

covid19.analytics -- Trends INDIA / 2021-12-23
India
2020-03-04 - 28


Guyana
Haiti
Holy See
Honduras
Hungary
Iceland
India
Indonesia
Iran
Iraq
Ireland
Israel
Italy
Jamaica
Japan
Jordan
Kazakhstan
Kenya
Kiribati
Korea, South
Kosovo
Kuwait
Kyrgyzstan
Laos
Latvia
Lebanon


USA
India
UK
Russia
France
China
Italy
Turkey
Germany
Spain
Brazil
Indonesia
Canada
Czechia
768,344,446
651,950,127
371,694,121
229,200,000
170,480,237
160,000,000
123,896,511
110,515,595
85,782,441
66,213,858
63,776,166
56,927,434
49,600,916
45,172,922
India
651,950,127

Graph of daily doses administered across the country





ANALYSIS REPORTS AND INFERENCE:

- Comparing Confirmed and Death cases globally shows that the two entities are significantly different as the F value is positive. We reject the Null Hypothesis. Infers that confirmed and instances of death vary, indicating that the confirmed cases are more than the death cases.
- Compared to India, brazil has more Covid cases
- Thus with the SIR model, the covid infected cases are getting increased in India as the number of persons who are prone covid is more
- Forecasting for the next 28 days predicts that the covid cases will increase concerning the actual data (up to Jan 19, 2022).
- It infers that the SARS COV-2 gets evolved in different variants by mutating its genome in a way by changing its characteristics and traits to spread even more vigorously, thus resulting in the rise of COVID cases globally

PROOF for TRUTHFULNESS OF THE FORECASTING DATA

17-11-2021

```
##################################################################
  ##### TS-CONFIRMED Cases  -- Data dated:  2021-11-17  ::  2021-1
##################################################################
  Number of Countries/Regions reported:  196
  Number of Cities/Provinces reported:  88
  Unique number of distinct geographical locations combined: 280
------------------------------------------------------------------
  Worldwide ts-confirmed  Totals: 255004381
------------------------------------------------------------------
  Country.Region Province.State   Totals GlobalPerc LastDayChange
1             US                47420114      18.60        111106
2          India                34478517      13.52         11919
3         Brazil                21977661       8.62         11977
4 United Kingdom                 9675058       3.79         37868
5         Russia                 9027163       3.54         35415
------------------------------------------------------------------
  Global Perc. Average:  0.36 (sd: 1.54)
  Global Perc. Average in top  5 :  9.61 (sd: 6.48)
------------------------------------------------------------------
==================================================================
```

23-12-2021

```
------------------------------------------------------------------
  Worldwide ts-confirmed  Totals: 277161199
------------------------------------------------------------------
  Country.Region Province.State   Totals GlobalPerc LastDayChange
1             US                51545991      18.60        238378
2          India                34765976      12.54          7495
3         Brazil                22222928       8.02          3451
4 United Kingdom                11647473       4.20        105330
5         Russia                10114983       3.65         25038
------------------------------------------------------------------
```

CONCLUSION:

JUSTIFICATION:

Though Herd immunity is developing among the people to fight against the novel coronavirus, it mutates itself by changing its characteristics and traits, resulting in a widespread virus.

SOLUTION TO THIS PROBLEM:

- The sequencing of the Genomes and testing can be done to prevent the spread of the mutated virus
- The initiation of vaccination can increase among the people
- People can maintain the Covid restrictions put forth by the Government to bring the covid in control