

COVID19_project2

Jose Rosario

2022-09-23

R Markdown

After we import the libraries, check the urls are correct and go download the csv

```
url_in <- "https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_data"
file_names <-
c("time_series_covid19_confirmed_global.csv",
  "time_series_covid19_deaths_global.csv",
  "time_series_covid19_confirmed_US.csv",
  "time_series_covid19_deaths_US.csv",
  "time_series_covid19_recovered_global.csv")
urls <- str_c(url_in, file_names)
urls
```

```
## [1] "https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_data"
## [2] "https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_data"
## [3] "https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_data"
## [4] "https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_data"
## [5] "https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_data"
```

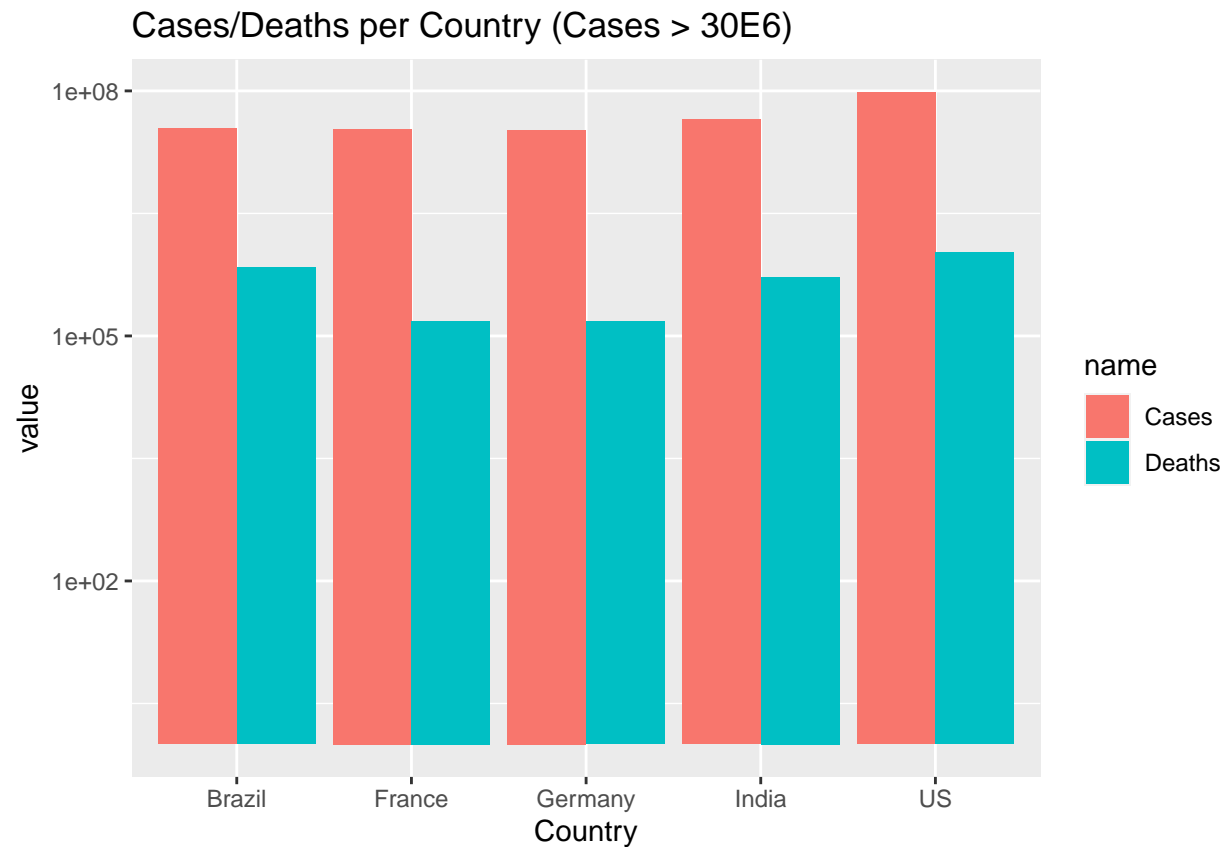
let's read in the data and see what we have

```
global_cases <- read_csv(urls[1])
global_deaths <- read_csv(urls[2])
US_cases <- read_csv(urls[3])
US_deaths <- read_csv(urls[4])
```

after downloading the raw data, it is time to analyze and visualize, in this case we are only using global data for this part of the project

Visualization 1

```
globalMaxCases <- data.frame(Country = global_cases$`Country/Region`,
                             Cases = global_cases[ncol(global_cases)],
                             Deaths = global_deaths[ncol(global_cases)]) %>% rename(Cases = X9.24.22,Deaths = X9.24.22)
g <- globalMaxCases %>% filter(Cases > 30000000)
g %>%
  pivot_longer(-Country) %>%
  ggplot(aes(fill=name, y = value, x=Country)) + ggtitle("Cases/Deaths per Country (Cases > 30E6)") +
  scale_y_log10() +
  geom_bar(position = "dodge", stat = "identity")
```



thanks for the data, it is important to know that this are the countries that has more than 30,000,000 cases in that time, so now it's time to choose one to compare with US

Visualization 2

```
GC <- global_cases %>%
  pivot_longer(cols = -c('Province/State', 'Country/Region', Lat, Long),
               names_to = "date",
               values_to = "cases")%>%
  select(-c(Lat, Long))

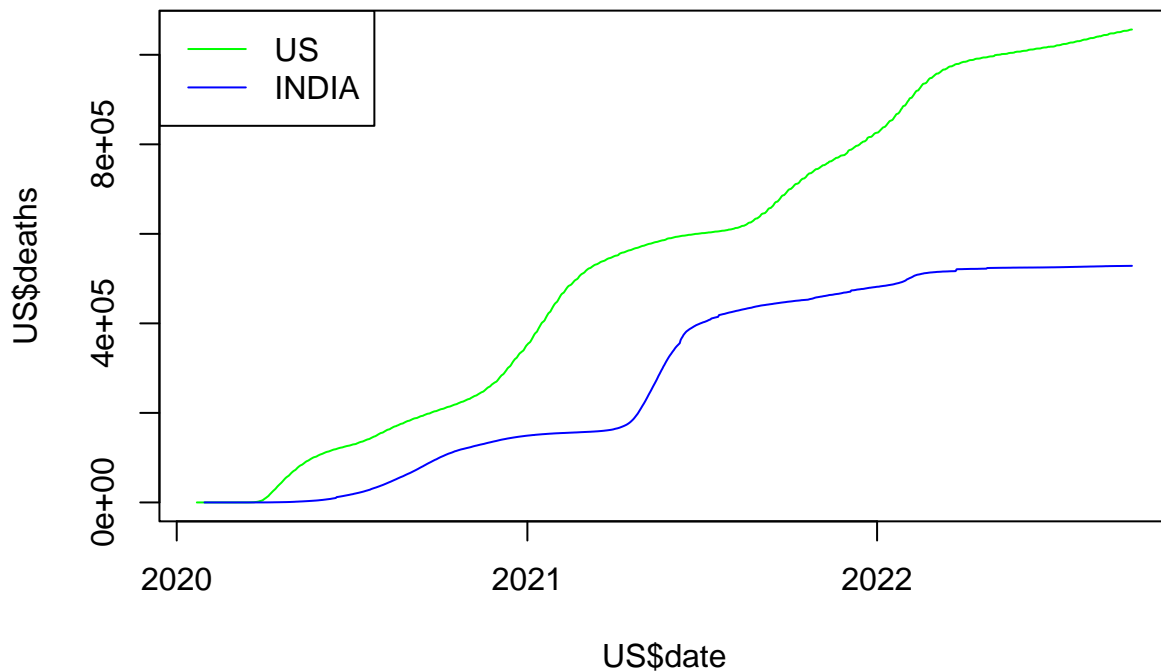
GD <- global_deaths %>%
  pivot_longer(cols = -c('Province/State', 'Country/Region', Lat, Long),
               names_to = "date",
               values_to = "deaths")%>%
  select(-c(Lat, Long))

global <- GC %>%
  full_join(GD)%>%
  rename(Country_Region = 'Country/Region',
         Province_State = 'Province/State')%>%
  mutate(date = mdy(date))
global <- global %>% filter(cases > 0)

US <- global %>% filter(Country_Region == 'US')
US$Province_State = NULL
```

```
INDIA <- global %>% filter(Country_Region == 'India')
INDIA$Province_State = NULL

plot(US$date,US$deaths,type="l",col="green")
lines(INDIA$date,INDIA$deaths,col="blue")
legend("topleft",legend=c("US", "INDIA"),
      col=c("green", "blue"), lty=1)
```



this plot can show how hard covid hit both countries but even if India had approximately the same number of cases as US, in US there are more deaths

```
##Model
```

```
mod1 <- lm(deaths ~ cases, data = US)
summary(mod1)
```

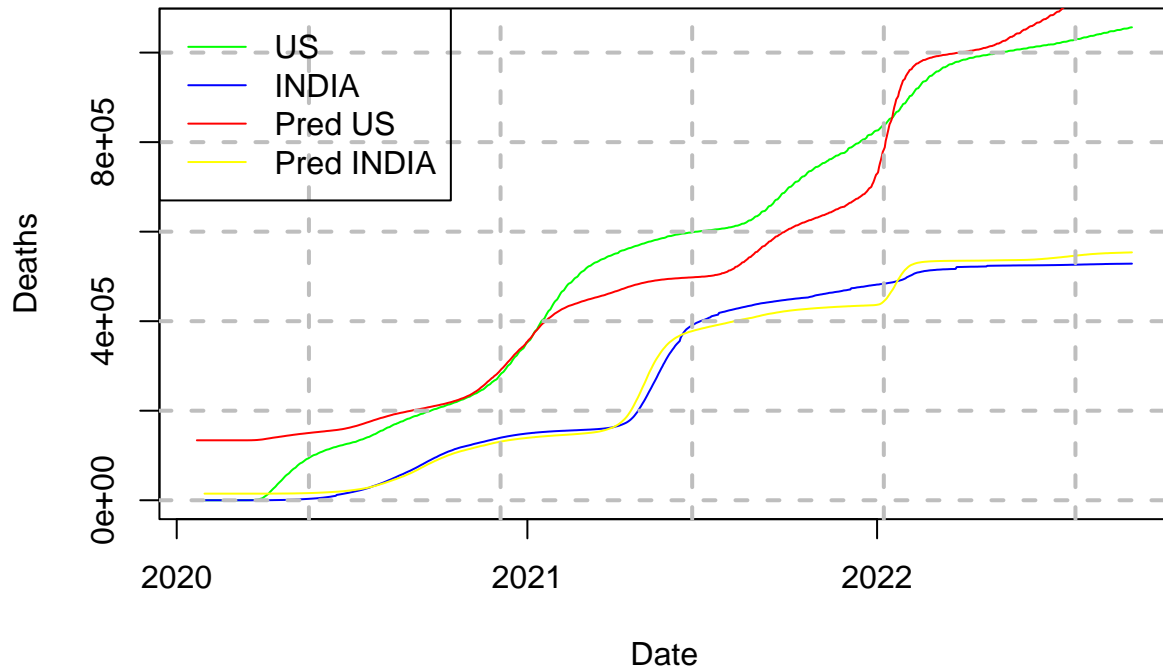
```
##
## Call:
## lm(formula = deaths ~ cases, data = US)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -133891  -50717  -10814   83826  123428
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) 1.339e+05 3.897e+03 34.35 <2e-16 ***
## cases      1.081e-02 7.872e-05 137.28 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 77770 on 975 degrees of freedom
## Multiple R-squared:  0.9508, Adjusted R-squared:  0.9508
## F-statistic: 1.885e+04 on 1 and 975 DF, p-value: < 2.2e-16
```

```
mod2 <- lm(deaths ~ cases, data = INDIA)
summary(mod2)
```

```
##
## Call:
## lm(formula = deaths ~ cases, data = INDIA)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -43268 -14742  -1035    9750   45176
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.474e+04  1.046e+03   14.09  <2e-16 ***
## cases      1.209e-02  3.724e-05   324.75  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19640 on 967 degrees of freedom
## Multiple R-squared:  0.9909, Adjusted R-squared:  0.9909
## F-statistic: 1.055e+05 on 1 and 967 DF, p-value: < 2.2e-16
```

```
US<-US %>% mutate(pred = predict(mod1))
INDIA<-INDIA %>% mutate(pred = predict(mod2))
plot(US$date,US$deaths,xlab="Date", ylab="Deaths",type="l",col="green")
lines(INDIA$date,INDIA$deaths,col="blue")
lines(US$date,US$pred,col="red")
lines(INDIA$date,INDIA$pred,col="yellow")
grid(nx = NULL, ny = NULL,
      lty = 2,      # Grid line type
      col = "gray", # Grid line color
      lwd = 2)      # Grid line width
legend("topleft",legend=c("US", "INDIA","Pred US","Pred INDIA"),
      col=c("green", "blue","red","yellow"), lty=1)
```



So in this model, with the information given thanks to the CSSEGISandData, it is imperative that both countries should decrease their number of deaths by investing more in health and care, and use the authorities to stop spreading the disease. Also, are model is predicting that India had faster action in dealing with it. Thinks that the government should do: *Increase of medical staff and equipment* better isolation *Authorities should be hard against people threatening their lifes by spreading the disease