

# Covid 19 in the world

Duc Huynh

2024-06-22

Packages used: tidyverse, lubridate, forecast, zoo

## Covid-19 cases worldwide:

```
#Input data set:
library(tidyverse)
library(lubridate)
url_in <- "https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_cov
global_cases <- read.csv(url_in)

#Tidy and transform data:
global_cases <- global_cases %>%
  pivot_longer(cols = -c("Province.State", "Country.Region", "Lat", "Long"),
               names_to = "Date",
               values_to = "Cases") %>%
  mutate(Date = as.Date(sub("X", "", Date), format="%m.%d.%y"))
head(global_cases,5)
```

```
## # A tibble: 5 x 6
##   Province.State Country.Region   Lat   Long Date       Cases
##   <chr>          <chr>      <dbl> <dbl> <date>    <int>
## 1 ""            Afghanistan  33.9  67.7 2020-01-22     0
## 2 ""            Afghanistan  33.9  67.7 2020-01-23     0
## 3 ""            Afghanistan  33.9  67.7 2020-01-24     0
## 4 ""            Afghanistan  33.9  67.7 2020-01-25     0
## 5 ""            Afghanistan  33.9  67.7 2020-01-26     0
```

```
tail(global_cases,5)
```

```
## # A tibble: 5 x 6
##   Province.State Country.Region   Lat   Long Date       Cases
##   <chr>          <chr>      <dbl> <dbl> <date>    <int>
## 1 ""            Zimbabwe    -19.0  29.2 2023-03-05 264127
## 2 ""            Zimbabwe    -19.0  29.2 2023-03-06 264127
## 3 ""            Zimbabwe    -19.0  29.2 2023-03-07 264127
## 4 ""            Zimbabwe    -19.0  29.2 2023-03-08 264276
## 5 ""            Zimbabwe    -19.0  29.2 2023-03-09 264276
```

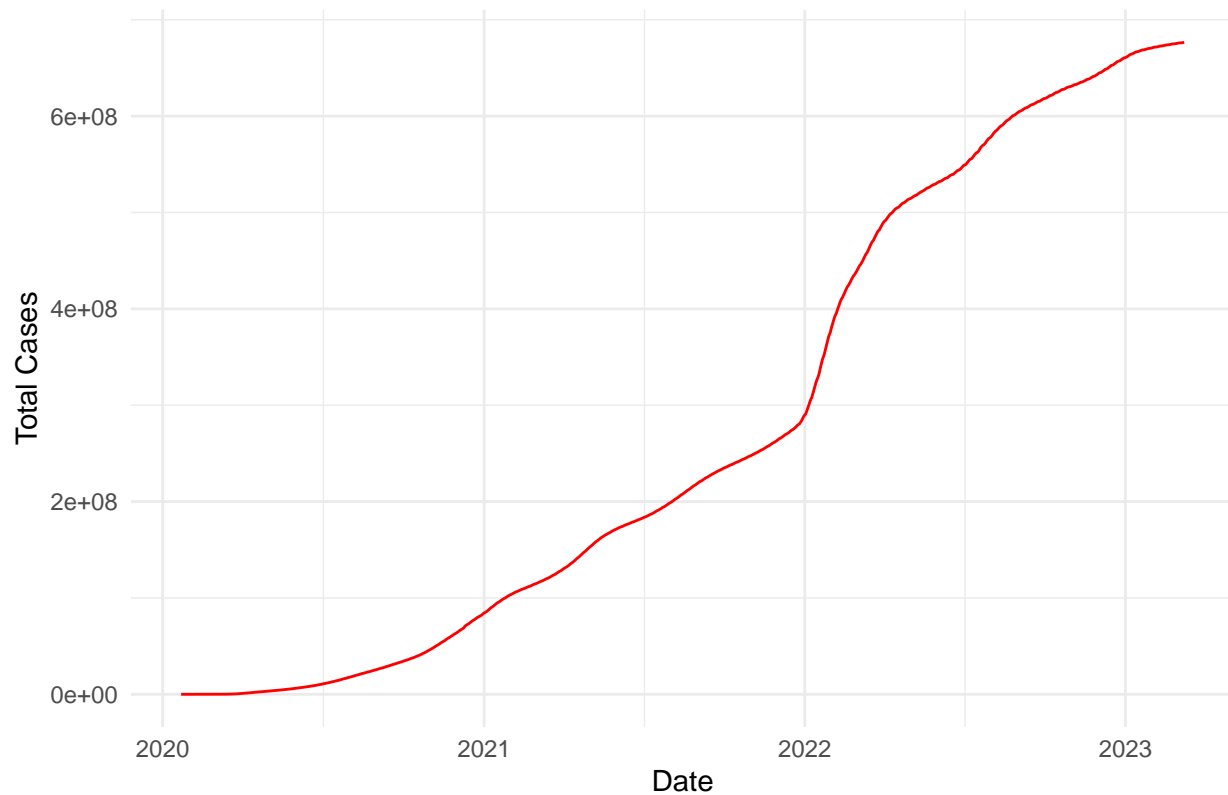
## Global Covid-19 cases through the years:

```
#Calculate the sum of cases per time:
global<- global_cases %>%
  group_by(Date) %>%
  summarise(TotalCases = sum(Cases))
global
```

```
## # A tibble: 1,143 x 2
##   Date      TotalCases
##   <date>      <int>
## 1 2020-01-22      557
## 2 2020-01-23      657
## 3 2020-01-24      944
## 4 2020-01-25     1437
## 5 2020-01-26     2120
## 6 2020-01-27     2929
## 7 2020-01-28     5580
## 8 2020-01-29     6169
## 9 2020-01-30     8237
## 10 2020-01-31     9927
## # i 1,133 more rows
```

```
#Make a graph of covid-19 cases through the years:
ggplot(global, aes(x = Date, y = TotalCases)) +
  geom_line(color = "red") +
  labs(title = "Global COVID-19 Cases from 2020 to 2023",
       x = "Date",
       y = "Total Cases") +
  theme_minimal()
```

## Global COVID-19 Cases from 2020 to 2023

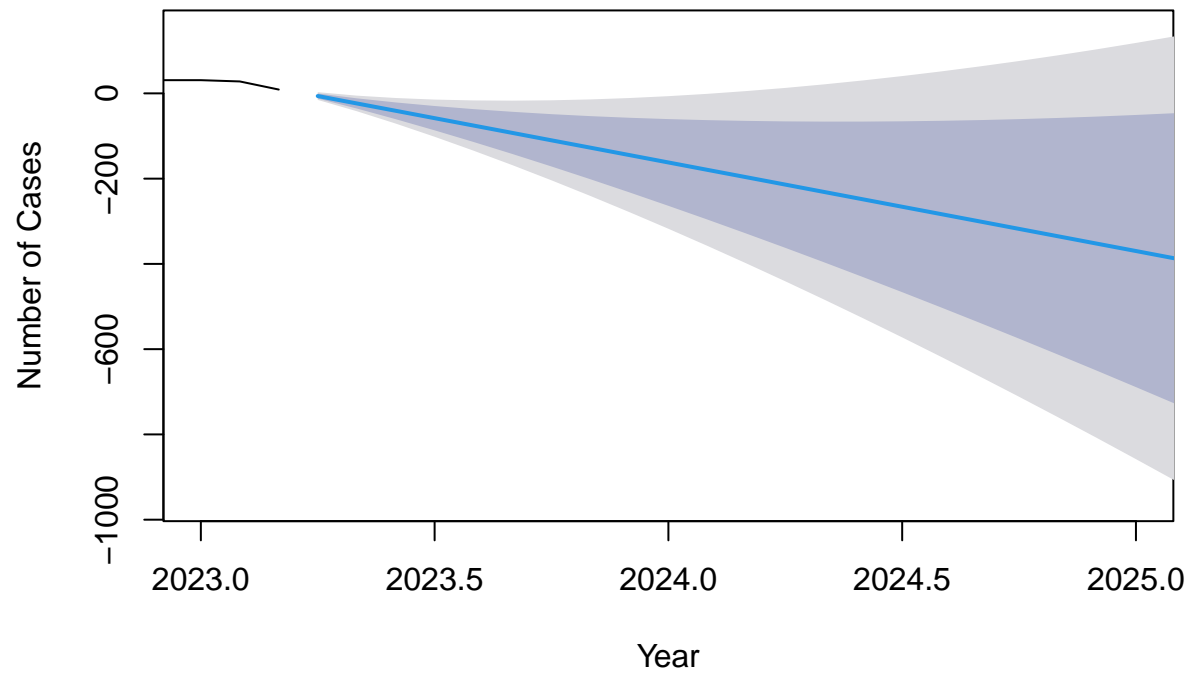


### Modeling Data:

I will try to make a prediction on global number covid cases from 2023 to 2025

```
library(forecast)
library(zoo)
#Add case count:
global$TotalCases <-1
aggYears<-aggregate(global$TotalCases,by=list(global$Date),sum)
aggMonth<-aggYears %>%
group_by(month = lubridate::floor_date(Group.1,"month")) %>%
summarize(summary_variable = sum(x))
#Convert to data frame:
frame<-as.data.frame(aggMonth)
tsdata<-as.ts(read.zoo(frame,FUN=as.yearmon))
prediction<-forecast(tsdata)
#Graph the prediction:
plot(prediction,xlim=c(2023,2025),
main="Forecast of global covid-19 cases",
xlab="Year",ylab="Number of Cases")
```

### Forecast of global covid-19 cases



From the graph we can see that the number of covid-19 cases decrease over the years.