COVID-19

2024-02-05

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(lubridate)
library(ggplot2)
```

COVID-19 Repository

The data set for this project has been collected from John Hopkins. The repository has plenty of information about the daily cases and deaths across the United States and Worldwide. However today the repository ceased collecting data.

```
url_in <- "https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covi

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")

urls <- str_c(url_in,file_names)
```

```
global_cases <- read_csv(urls[1])
```

```
## Rows: 289 Columns: 1147
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr    (2): Province/State, Country/Region
## dbl (1145): Lat, Long, 1/22/20, 1/23/20, 1/24/20, 1/25/20, 1/26/20, 1/27/20,...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
global_deaths <- read_csv(urls[2])
```

```
## Rows: 289 Columns: 1147
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr    (2): Province/State, Country/Region
## dbl (1145): Lat, Long, 1/22/20, 1/23/20, 1/24/20, 1/25/20, 1/26/20, 1/27/20,...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
US_cases <- read_csv(urls[3])
```

```
## Rows: 3342 Columns: 1154
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr    (6): iso2, iso3, Admin2, Province_State, Country_Region, Combined_Key
## dbl (1148): UID, code3, FIPS, Lat, Long_, 1/22/20, 1/23/20, 1/24/20, 1/25/20...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
US_deaths <- read_csv(urls[4])
```

```
## Rows: 3342 Columns: 1155
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr    (6): iso2, iso3, Admin2, Province_State, Country_Region, Combined_Key
## dbl (1149): UID, code3, FIPS, Lat, Long_, Population, 1/22/20, 1/23/20, 1/24...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

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

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

```
global <- global_cases %>%
    full_join(global_deaths) %>%
    rename(Country_Region = `Country/Region`,
           Province_State = `Province/State`) %>%
    mutate(date = mdy(date))
```

```
## Joining with `by = join_by(`Province/State`, `Country/Region`, date)`
```

```
global <- global %>% filter(cases > 0)
```

```
US_cases %>%
  pivot_longer(cols = -(UID:Combined_Key),
               names_to = "date",
               values_to = "cases")
```

```
## # A tibble: 3,819,906 x 13
##         UID iso2  iso3  code3  FIPS Admin2  Province_State Country_Region   Lat
##       <dbl> <chr> <chr> <dbl> <dbl> <chr>   <chr>          <chr>          <dbl>
##  1 84001001 US    USA     840  1001 Autauga Alabama        US              32.5
##  2 84001001 US    USA     840  1001 Autauga Alabama        US              32.5
##  3 84001001 US    USA     840  1001 Autauga Alabama        US              32.5
##  4 84001001 US    USA     840  1001 Autauga Alabama        US              32.5
##  5 84001001 US    USA     840  1001 Autauga Alabama        US              32.5
##  6 84001001 US    USA     840  1001 Autauga Alabama        US              32.5
##  7 84001001 US    USA     840  1001 Autauga Alabama        US              32.5
##  8 84001001 US    USA     840  1001 Autauga Alabama        US              32.5
##  9 84001001 US    USA     840  1001 Autauga Alabama        US              32.5
## 10 84001001 US    USA     840  1001 Autauga Alabama        US              32.5
## # i 3,819,896 more rows
## # i 4 more variables: Long_ <dbl>, Combined_Key <chr>, date <chr>, cases <dbl>
```

```
US_cases <- US_cases %>%
  pivot_longer(cols = -(UID:Combined_Key),
               names_to = "date",
               values_to = "cases") %>%
  select(Admin2:cases) %>%
  mutate(date = mdy(date)) %>%
  select(-c(Lat, Long_))
```

```
US_deaths <- US_deaths %>%
  pivot_longer(cols = -(UID:Population),
               names_to = "date",
               values_to = "deaths") %>%
  select(Admin2:deaths) %>%
  mutate(date = mdy(date)) %>%
  select(-c(Lat, Long_))
```

I will plot all the historic deaths of global data. However here is a bias, since, it's a lot of information, the visualization isn't great. I think there may be improvements to my code.

```
global_deaths_plot <- ggplot(global, aes(x=date, y=deaths, group=Country_Region, color=Country_Region))
  geom_line(size=0.5, alpha=0.8) +
  labs(title = "Historic Deaths from COVID-19 Worldwide",
       x = "Date",
       y = "Number of Deaths",
       color = "Country/Region") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

```
## Warning: Using ‘size‘ aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use ‘linewidth‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```

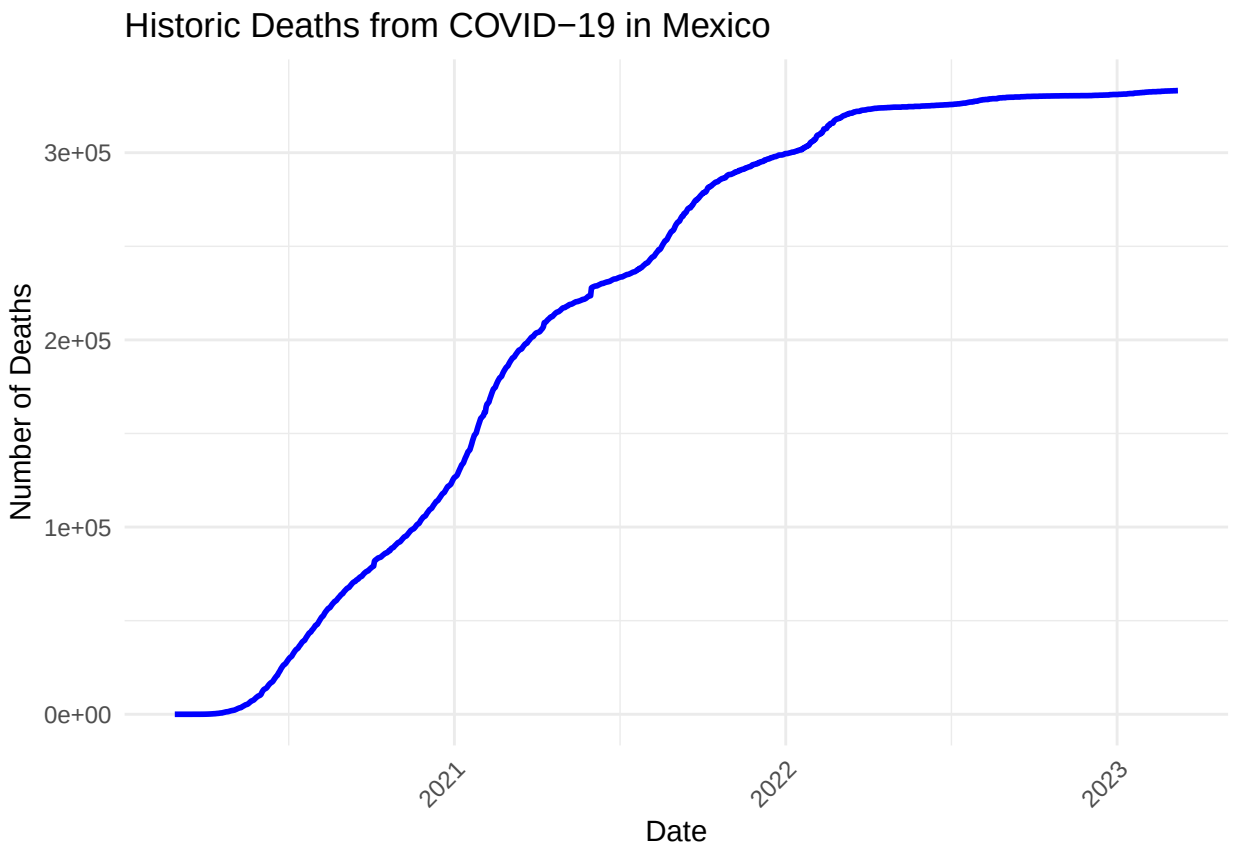
```
print(global_deaths_plot)
```

This is the second plot, in which I plot the historic data of Mexico. Its impressive to see that by 2022 it got stabilized th number of deaths.

```
mexico_data <- global %>%
  filter(Country_Region == "Mexico")

mexico_deaths_plot <- ggplot(mexico_data, aes(x=date, y=deaths)) +
  geom_line(size=1, color="blue") +
  labs(title = "Historic Deaths from COVID-19 in Mexico",
       x = "Date",
       y = "Number of Deaths") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

print(mexico_deaths_plot)
```



Here I propose a linear prediction model compared to the historic data of Mexico.

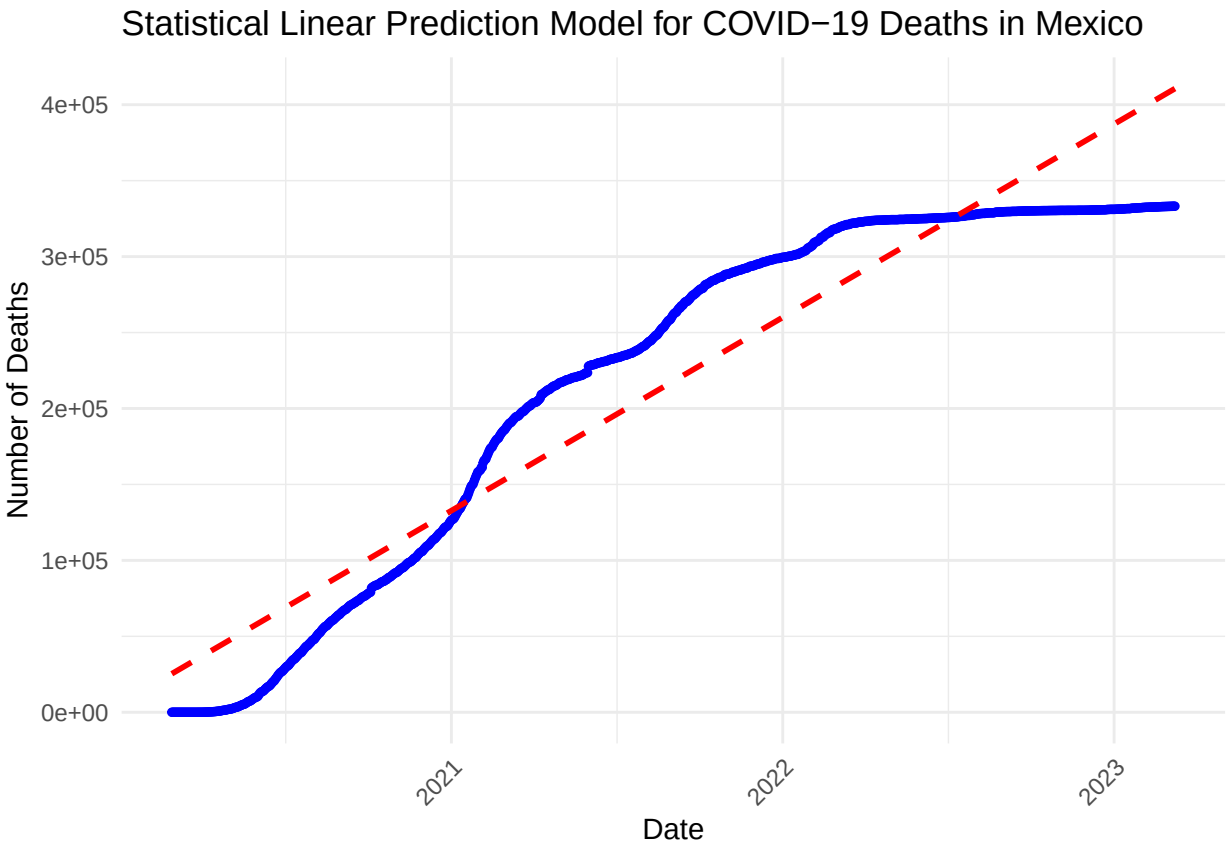
```
mexico_data <- global %>%
  filter(Country_Region == "Mexico")

model <- lm(deaths ~ date, data = mexico_data)

mexico_deaths_plot_with_model <- ggplot(mexico_data, aes(x=date, y=deaths)) +
  geom_point(size=1, color="blue") +
  geom_smooth(method=lm, se=FALSE, color="red", linetype="dashed") +
  labs(title = "Statistical Linear Prediction Model for COVID-19 Deaths in Mexico",
       x = "Date",
       y = "Number of Deaths") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

print(mexico_deaths_plot_with_model)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



Conclusions

The Covid-19 pandemic was a health problem worldwide, and with this analysis I can see that in Mexico occurred a stabilization process. However, at the beginning of each new year, it got elevated. This is because there were parties like new year's eve. As far as I know, the stabilization process began when there were vaccination for all the population. Bias I encountered were, firstly, my first visualization since it's very difficult to put all the countries into one analysis. This may be misleading. Secondly, it changes the analysis city from city here in Mexico. This may be a process of geographical economy and politics. And finally, the prediction should be compared with new cases for the year 2024.