COVID-19 Data Analysis

David Forero Botia

2024-02-25

Introduction

In this document, we will analyze the COVID-19 data Worldwide, presenting information of interest and obtaining relevant insights to answer the questions:

1. What has been the impact of these diseases in different countries?
2. Which ones have been the most and least affected?
3. What is the current state of cases?

Importing and Transforming the Data

Importing the Data

First, we will need to import the Data provided by the Johns Hopkins Coronavirus Resource Center. We will import directly from the repository so we are allowed to reproduce the analysis by anyone.

We will import the Data for the Number of cases as `global_cases`, Number of Deaths as `global_deaths` and Population for each country until 2023 as `pops`.

```
url_in<-'https://raw.githubusercontent.com/CSSEGISandData/COVID-
19/master/csse_covid_19_data/csse_covid_19_time_series/'
file_names<-
c('time_series_covid19_deaths_global.csv','time_series_covid19_confirmed_glob
al.csv')
urls<-str_c(url_in,file_names)

global_cases<-read_csv(urls[2])
global_deaths<-read_csv(urls[1])


UID_data <- 'https://raw.githubusercontent.com/CSSEGISandData/COVID-
19/master/csse_covid_19_data/UID_ISO_FIPS_LookUp_Table.csv'
pops <- read_csv(UID_data)
```

Transforming the Data

To transform the data into something that we can analyze we will follow the next steps:

1. Convert the `global_cases` and `global_deaths` tables from wide to long

2. Joining the tables `global_cases` and `global_deaths` and filter out the rows with cases equals to 0.

3. Joining the previous table with the `pops` table.

```
## # A tibble: 6 × 6
##   Province_State Country_Region date       cases deaths Population
##   <chr>          <chr>          <date>     <dbl>  <dbl>      <dbl>
## 1 <NA>           Afghanistan    2020-02-24     5      0   38928341
## 2 <NA>           Afghanistan    2020-02-25     5      0   38928341
## 3 <NA>           Afghanistan    2020-02-26     5      0   38928341
## 4 <NA>           Afghanistan    2020-02-27     5      0   38928341
## 5 <NA>           Afghanistan    2020-02-28     5      0   38928341
## 6 <NA>           Afghanistan    2020-02-29     5      0   38928341
```

4. We will group the previous table by Country and add a new variable where we get the deaths per million people.

5. We will add the New number of cases and the New number of deaths per Country to the new table.

```
## # A tibble: 6 × 8
## # Groups:   Country_Region [1]
##   Country_Region date       cases deaths deaths_per_mill Population
new_cases
##   <chr>          <date>     <dbl>  <dbl>           <dbl>      <dbl>
<dbl>
## 1 Afghanistan    2020-02-24     5      0               0   38928341
NA
## 2 Afghanistan    2020-02-25     5      0               0   38928341
0
## 3 Afghanistan    2020-02-26     5      0               0   38928341
0
## 4 Afghanistan    2020-02-27     5      0               0   38928341
0
## 5 Afghanistan    2020-02-28     5      0               0   38928341
0
## 6 Afghanistan    2020-02-29     5      0               0   38928341
0
## # ℹ 1 more variable: new_deaths <dbl>
```

6. We will create a new table where we display the maximum number of cases and deaths in the whole time frame for each country.

```
## # A tibble: 6 × 6
##   Country_Region      deaths  cases Population cases_per_thou
deaths_per_thou
##   <chr>                <dbl>  <dbl>      <dbl>          <dbl>
<dbl>
## 1 Afghanistan           7896 209451   38928341           5.38
0.203
```

```
## 2 Albania                3598 334457    2877800         116.
1.25
## 3 Algeria                6881 271496   43851043           6.19
0.157
## 4 Andorra                 165  47890      77265         620.
2.14
## 5 Angola                 1933 105288   32866268           3.20
0.0588
## 6 Antigua and Barbuda     146   9106      97928          93.0
1.49
```

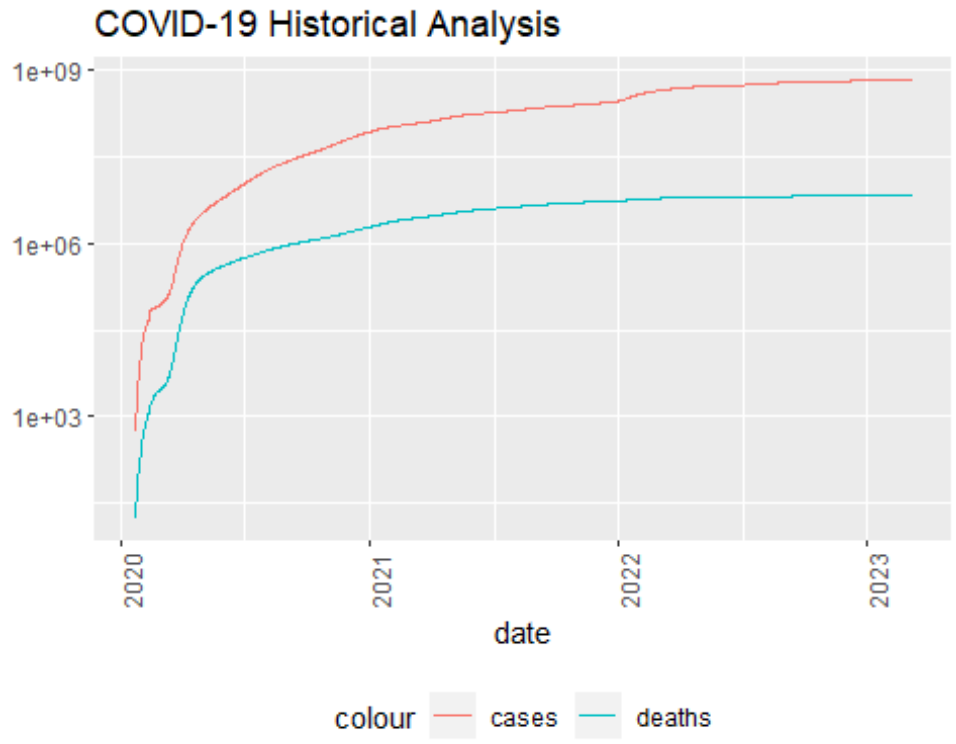
Analyzing the data

1. First, we will see the complete number of cases and deaths caused by the pandemic:

```
The total number of cases since the beginning of the pandemic until the end
of 2023 are: 316,910,296,319

The total number of deaths caused by the pandemic until the end of 2023 are:
4,419,815,836. This is equal to a 1.39% fatality rate.
```

This can be displayed historically in the next graph:

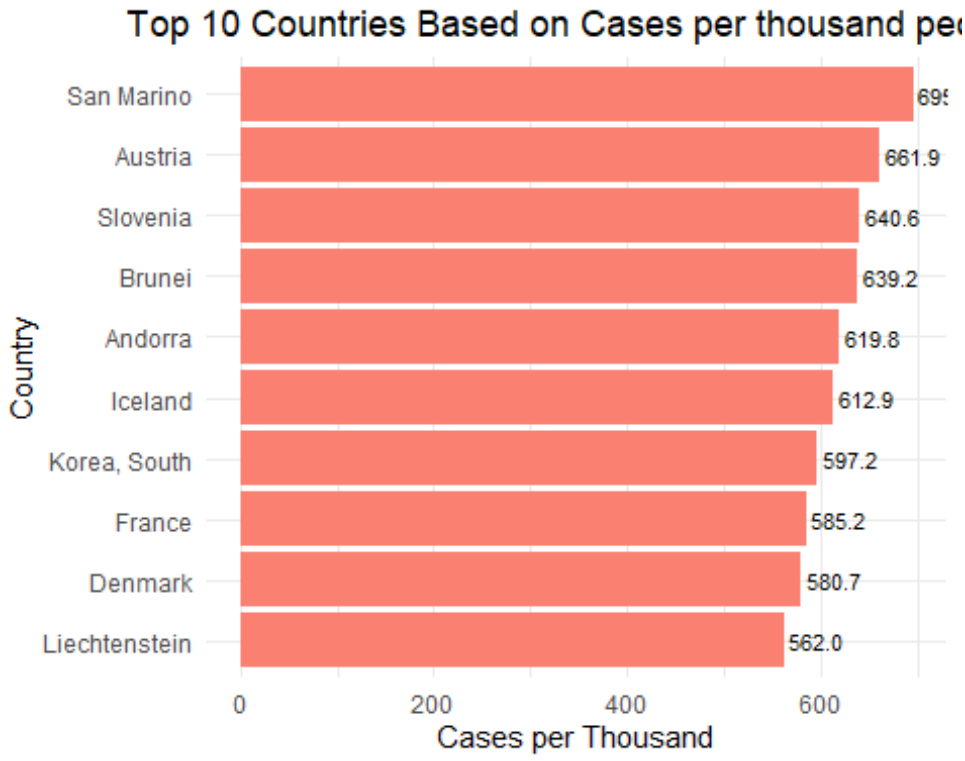


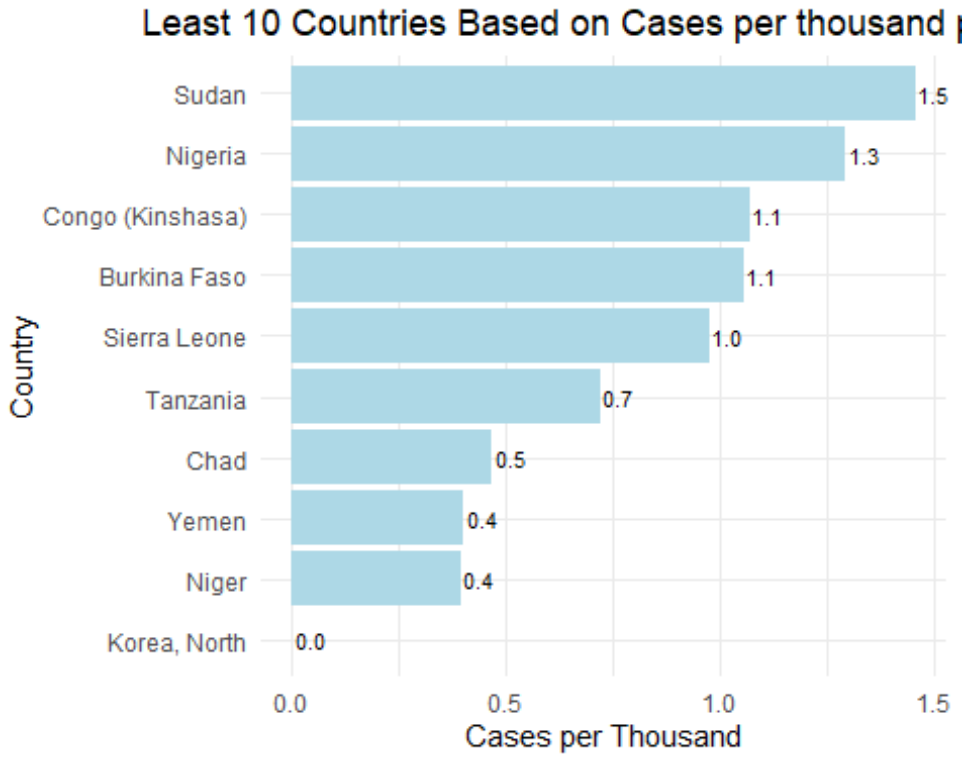
We can see that the rate of cases and deaths increased rapidly during 2020 and 2021 but after 2022 the number of new cases reduced making the graph flat in his latest part, this can be related to the application of the vaccine to the general population.

We can see better the deceleration of new cases during the years in the following graph, where is displayed the new number of cases and deaths historically.

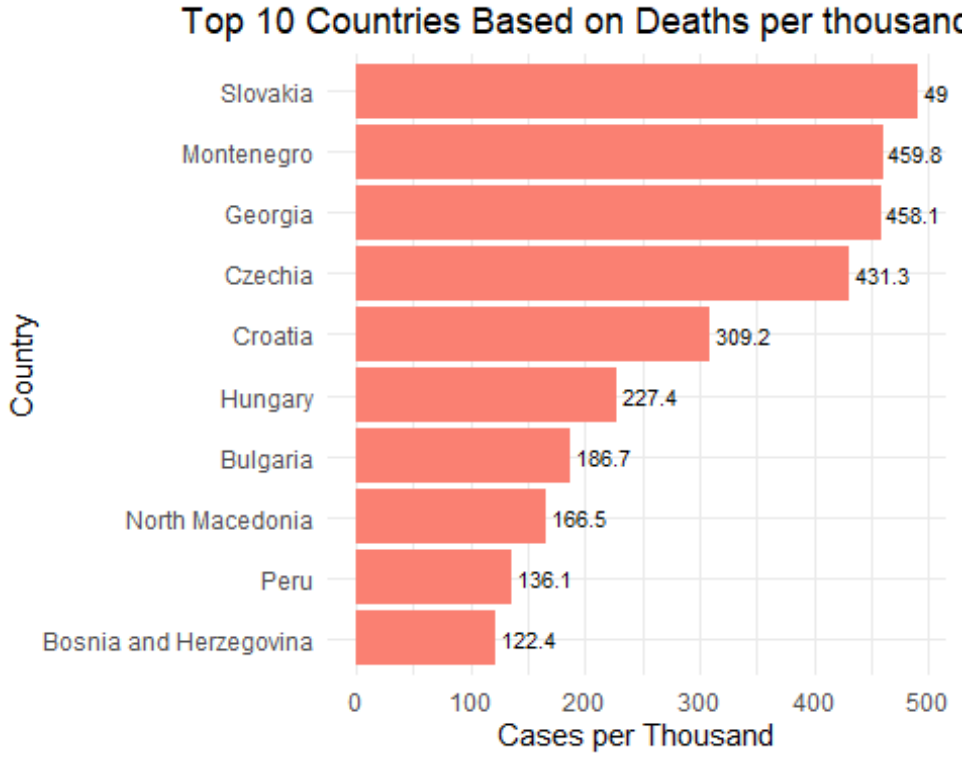
2. Now we will analyze the impact of the Pandemic on each country. To do this we will use a new variable, which is the number of cases per thousand people, this way we can normalize the numbers based on each population, and that way we can compare the real impact of COVID-19 on each country.

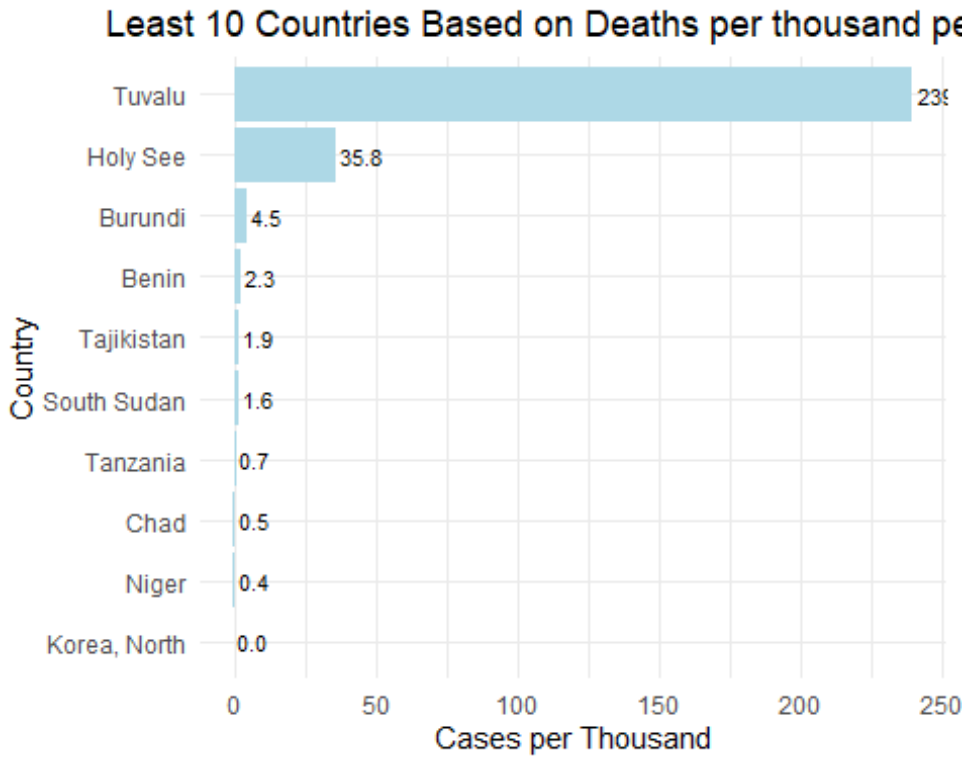
In the next graphs, I will present the Top 10 countries with more cases per thousand people and the 10 countries with the least cases per thousand people:


Top 10 Countries Based on Cases per thousand pe
San Marino
Austria
Slovenia
Brunei
Andorra
Iceland
Korea, South
France
Denmark
Liechtenstein
695
661.9
640.6
639.2
619.8
612.9
597.2
585.2
580.7
562.0
Country
0
200
400
600
Cases per Thousand


Least 10 Countries Based on Cases per thousand
Sudan
Nigeria
Congo (Kinshasa)
Burkina Faso
Sierra Leone
Tanzania
Chad
Yemen
Niger
Korea, North
1.5
1.3
1.1
1.1
1.0
0.7
0.5
0.4
0.4
0.0
Country
0.0
0.5
1.0
1.5
Cases per Thousand

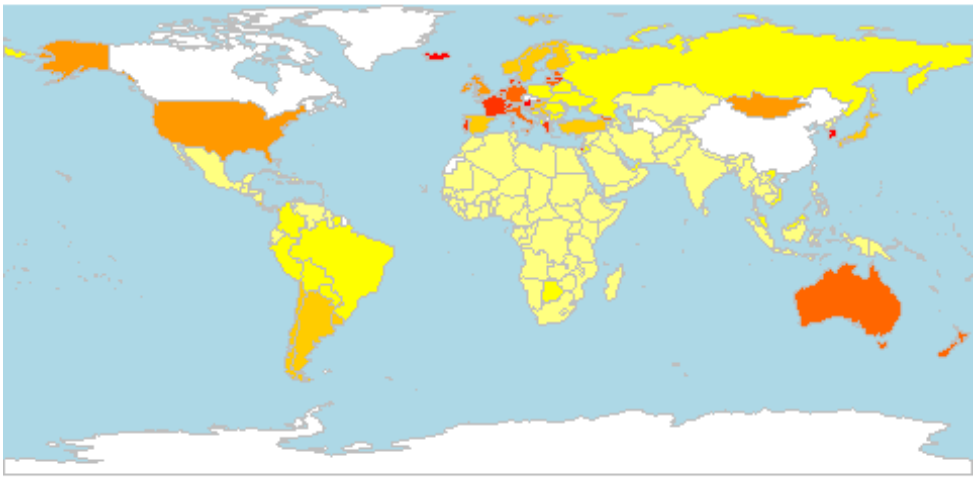
Below are the top 10 countries and 10 countries with the least deaths per thousand people:



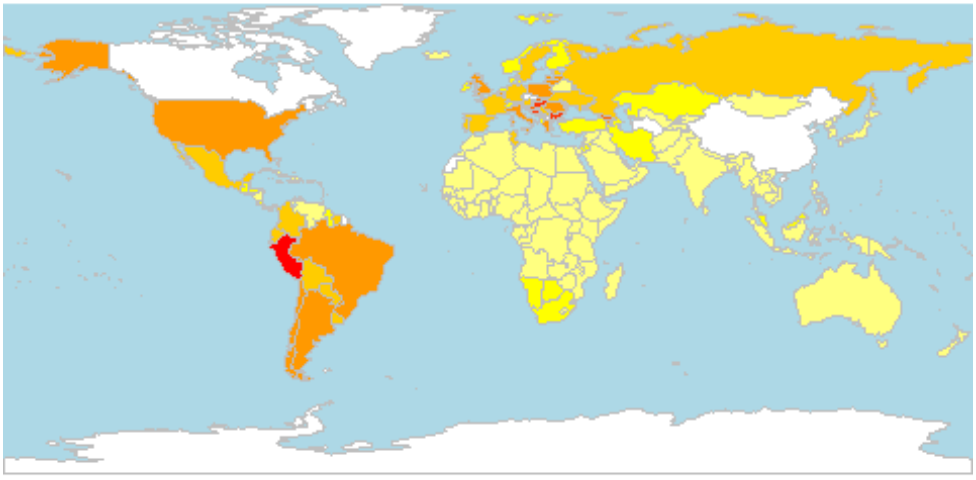


Below we can see the data plotted on a map, allowing us to see more clearly the location of each country and its impact:

COVID-19 Cases per Thousand People



COVID-19 Deaths per Thousand People



Conclusion

This analysis delves into the widespread effects of the COVID-19 pandemic, highlighting how each country tackled its unique challenges. Globally, nearly 300k million cases were reported, with a relatively low fatality rate of 1.39%, peaking between 2020 and 2021 before widespread vaccination efforts in 2022.

To provide a clearer picture, we examined cases and deaths per thousand individuals, identifying the top 10 and bottom 10 countries in terms of impact. This approach reveals that countries with high case rates didn't always have the highest death rates per thousand. Notably, Western Europe saw the highest case burden, while Eastern Europe and Latin America had the highest death rates per thousand.

Bias

It's important to note the limitations of this analysis. Data availability varied across countries, with some withholding information, like China, where the pandemic began. Others had unreliable data, potentially skewing our findings.

Despite these limitations, this analysis offers valuable insights into the global impact of the pandemic, helping us understand how different countries fared and providing a basis for comparison.