Exploratory Analysis of Global COVID-19 Cases
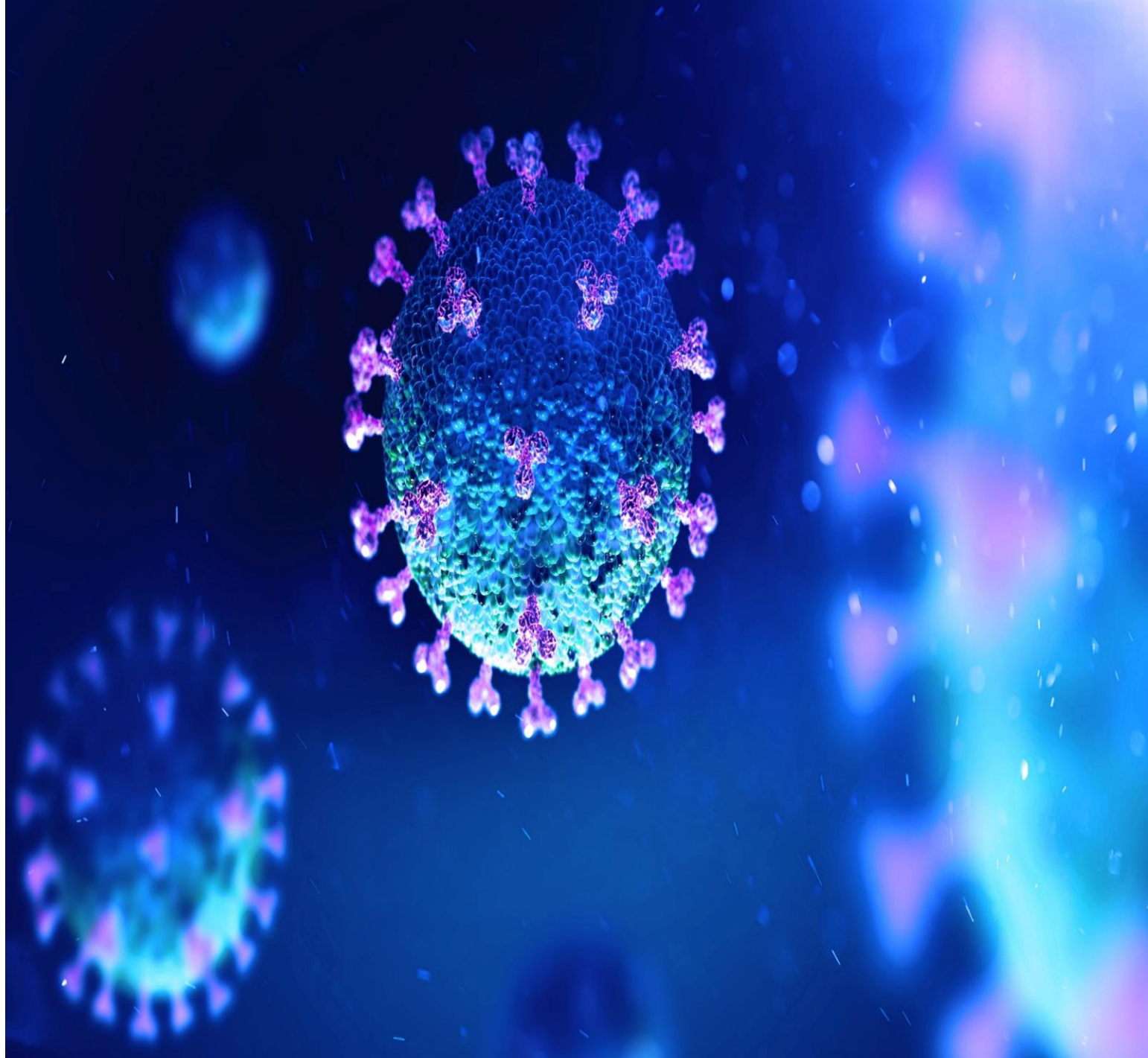
Case of John Hopkins University Dataset

By Ebunoluwa Tella
August, 2023

Introduction

The raw data used in this analysis was obtained from several sources and provided by the John Hopkins University.

This analysis pertains to all COVID instances worldwide. The dataset on confirmed cases, death cases, and recovered cases are included. I downloaded the csv file to my computer and imported the report into Power Query since I wanted to improve on the excel and power query skills I had been learning for this assignment.

These are the few points I wanted to find out about from the dataset:

1. The total number of confirmed cases reported globally?
2. The total number of deaths cases reported globally?
3. The annual trend in COVID cases I was curious if there had been a decrease or increase in cases since the pandemic began?
4. Which country has had the most confirmed cases?
5. Which countries have the highest number of death cases?
6. Which country has the least number of confirmed cases?
7. The number of confirmed cases each year.
8. The number of death cases each year.
9. The average number of confirmed cases each year.
10. The average number of deaths cases each year.

Process for Cleaning Data

- To check for inconsistent data, the dataset had to be cleaned up. This was achieved by importing the dataset into Power Query and checking for consistency in the row headers were in the first rows of each table.
- The dataset was checked for consistency as it did not include any duplicates. The columns which contained rows including dates and records for each day recorded in the dataset had to unpivot.
- Using the left join method, the three tables containing the cases were joined into a single table. The steps used to clean the datasets in a power query are depicted in the image below.
- Furthermore, the date function was to be separated into days, months, and years, but owing to the columns' inconsistent format, the text to column function was used with the (=text(value, format)) functions to separate them into days, months, and years.

Applied steps to cleaning the data using power query

Total Cases

Display Options ▾ ❓

```
1   let
2       Source = Table.NestedJoin(Confirmed, {"Province/State", "Country/Region", "Date"}, Deaths, {"Province/State", "Country/
            Region", "Date"}, "Deaths", JoinKind.LeftOuter),
3       #"Expanded Deaths" = Table.ExpandTableColumn(Source, "Deaths", {"Death Rates"}, {"Death Rates"}),
4       #"Merged Queries" = Table.NestedJoin(#"Expanded Deaths", {"Province/State", "Country/Region", "Date"}, Recovered,
            {"Province/State", "Country/Region", "Date"}, "Recovered", JoinKind.LeftOuter),
5       #"Expanded Recovered" = Table.ExpandTableColumn(#"Merged Queries", "Recovered", {"Recovered Cases"}, {"Recovered Cases"})
6   in
7       #"Expanded Recovered"
```

✔ No syntax errors have been detected.

Analyze Data

The cleansed data was used to create pivot tables and charts that summarised the merged datasets into more manageable groupings. The charts were then used to build a dashboard.

The overall insights from our research and visualisation that are provided below can help with certain questions.

603M+
Total Number of Global Confirmed Cases

6M +
Total Global Death Cases

The top five(5) countries with the lowest number of confirmed cases are Kiribati (3,430), Nauru (4,611), Winter Olympics 2022 (535), Summer Olympics 2020 (865), Diamond Princess (712).

76,653
Average number of Global Death Cases in 2020

86.30% →

559,782
Average number of confirmed cases in 2021

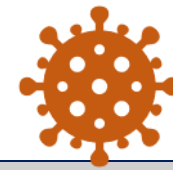**2,381**
Average Number of Death Cases in 2020

78.46% →

11,057
Average Number of Death Cases in 2021

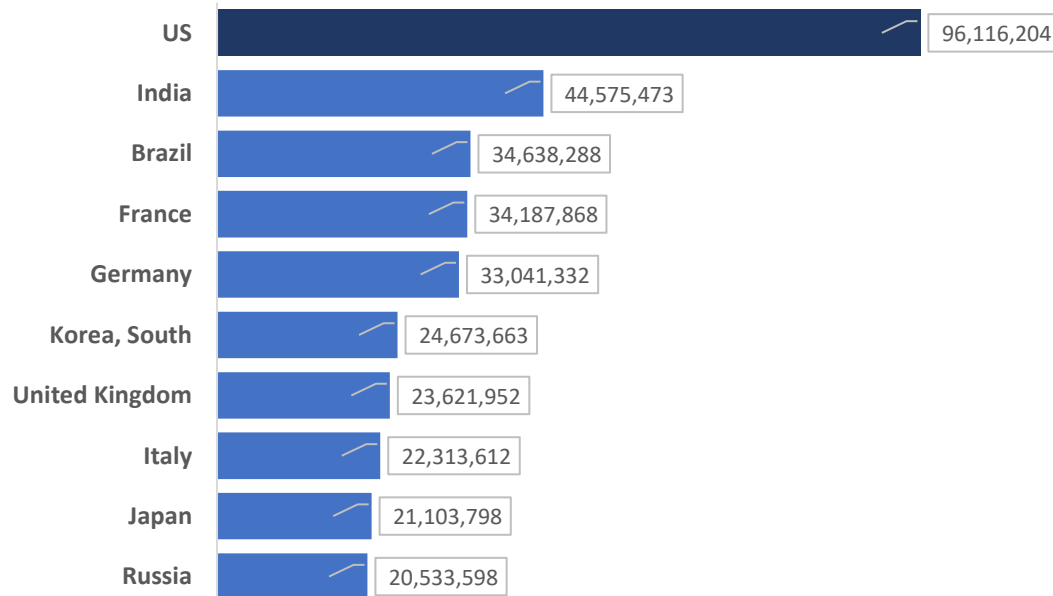# The top countries 10 with the most confirmed and death cases

The United States of America accounts for 15.93 percent of global population, with 96,116,204 confirmed cases, followed by India (44,575,473), Brazil (34,638,288), France (34,187,868), and Germany (33,041,332).
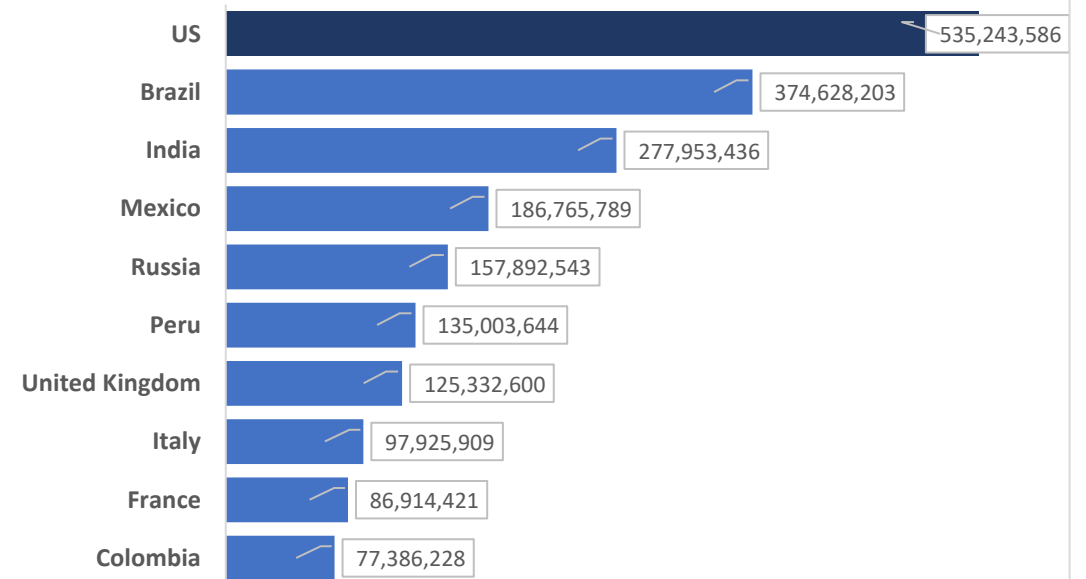
The United States of America has the highest number of deaths (1,056,789), accounting for 16.29 percent of the total population, followed by Brazil (685835), India (582,562), Russia (379,019), and Mexico (330,046).

The top ten nations with the highest number of deaths account for 59.94% of the total population. Find below the graphics that shows the top 10 countries.

Top 10 Countries with the Highest Confirmed Case

Country	Confirmed Cases
US	96,116,204
India	44,575,473
Brazil	34,638,288
France	34,187,868
Germany	33,041,332
Korea, South	24,673,663
United Kingdom	23,621,952
Italy	22,313,612
Japan	21,103,798
Russia	20,533,598

Top 10 Countries with the Highest Death

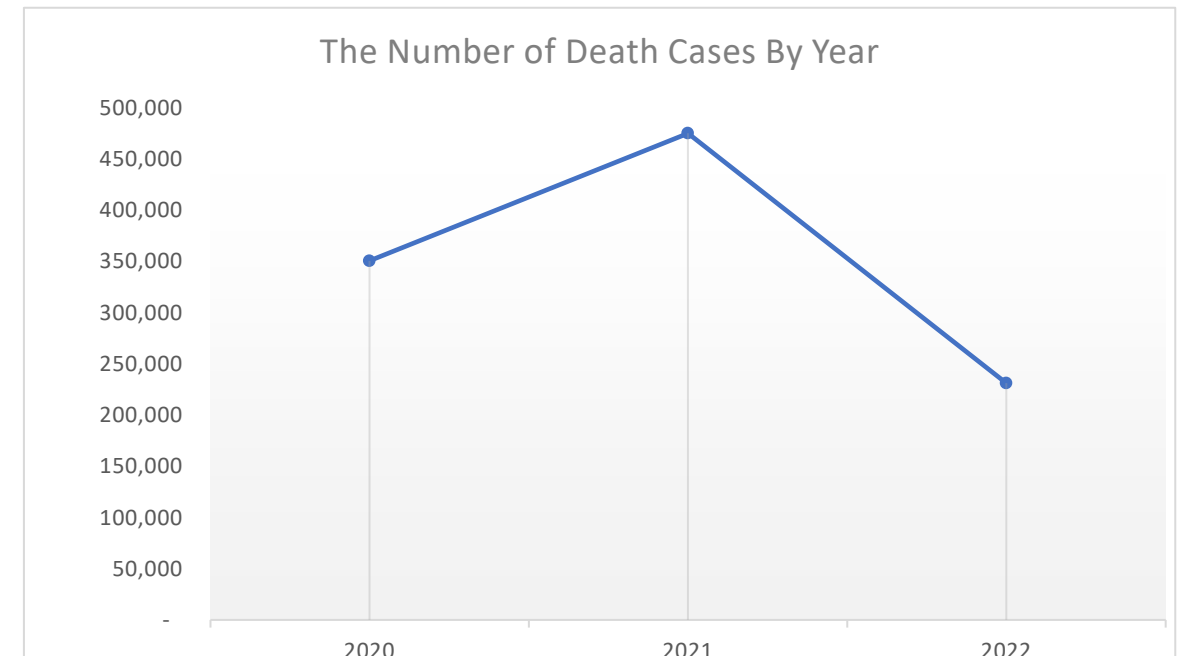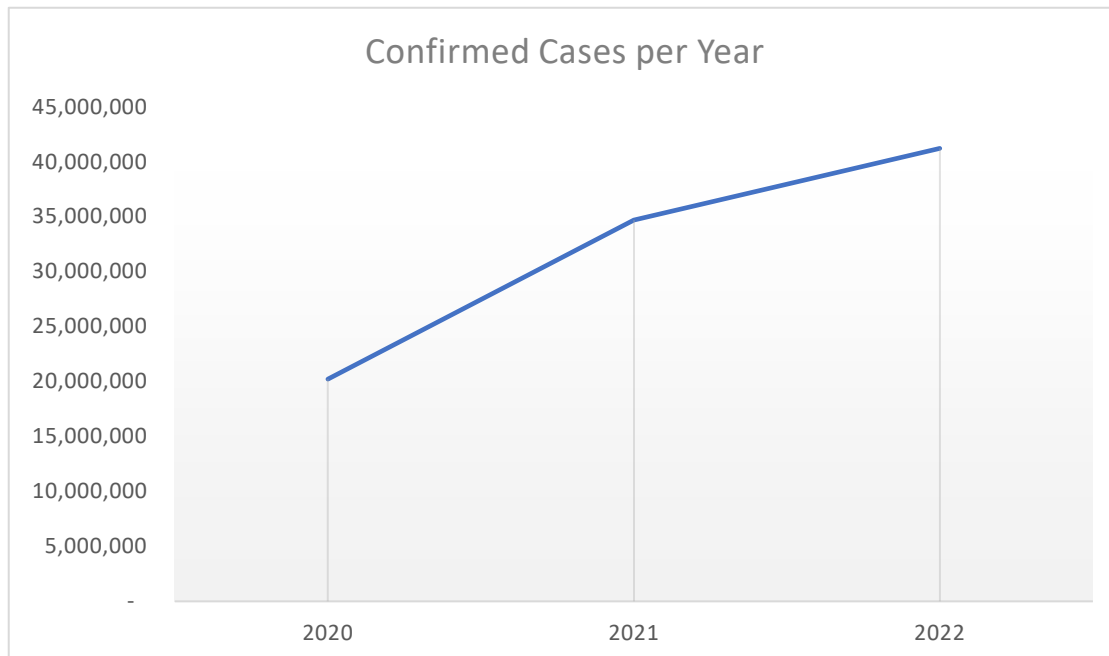| Country	Deaths
US	535,243,586
Brazil	374,628,203
India	277,953,436
Mexico	186,765,789
Russia	157,892,543
Peru	135,003,644
United Kingdom	125,332,600
Italy	97,925,909
France	86,914,421
Colombia	77,386,228

The average number of Confirmed and Deaths cases each year.

According to the graph above, the number of confirmed cases is growing on a daily basis. The number of confirmed cases in 2020 was 20,221,639; in 2021, it grew 34,690,608, representing a 41.61% rise between 2020 and 2021.

This information is updated daily, however as of September, the total number of instances documented had grown to 41,203,957.The number of recorded deaths cases has decreased over time. The year 2021 saw the most death instances, with 475061. The number of instances increased by 26.21 percent between 2020 and 2021.

Confirmed Cases per Year

The Number of Death Cases By Year

- Covid-19 is still common, and people are becoming infected with it. Brazil, America, and India, are among the top three (3) nations in terms of death and confirmed cases.
- These countries should continuously make efforts to flatten the curve and minimize the number of confirmed cases reported daily. Due to the rise of asymptomatic symptoms of Covid, there has been an unfounded assumption that Covid is no longer around us
- Citizens of certain countries should follow the local health ministry's instructions to reduce the high number of confirmed cases in certain countries.
- Know all the facts and take proper steps to protect themselves and people around them. Check with their local health authority for the most relevant recommendations for their location.

To curb some of the issues mentioned above, information needs to be disseminated to residents regularly on the ongoing dangers of the virus as its spread is still sweeping across the countries.

As the COVID-19 situation continues to escalate, we should endeavor to find a solution. The number of deaths has decreased dramatically, which we may attribute to more individuals getting vaccinated against the virus and the enforcement of guidelines by local Health Regulators.