Covid 19 Disease: Data Analysis and Impact

Lucas Jakin

2025-04-01

COVID-19: Data Analysis and Impact

The Covid-19 pandemic has profoundly affected the world, leading to significant health, economic, and social challenges. The global response has included lockdowns, travel restrictions and mass vaccination campaigns.

Overview of COVID-19

Coronavirus disease 2019 is a contagious disease caused by the coronavirus **SARS-CoV-2**.The first known case was identified in Wuhan, China and it quickly spread worldwide, resulting in the COVID-19 pandemic.

Impact on global health

COVID-19 has led to millions of infections and deaths worldwide. The pandemic has also overwhelmed healthcare systems and led to a long-term health complications for many survivors.

Data Analysis of Covid-19 Cases

Analyzing COVID-19 data helps us understand its spread and impact. The following COVID-19 dataset contains information about the numbers of Confirmed, Death and Recovered cases across the globe. The dataset was released in 2020, so the data has not been updated for the past 4 years. This serve us just as an example. Let's explore some COVID-19 related data. Let's explore some COVID-19 related data.

```r
library(dplyr)
library(tidyverse)
library(gt)
library(knitr)
# Importing the COVID-19 dataset from kaggle.
covid_data <- read.csv("country_wise_latest.csv")
#head(covid_data)
```

Since the dataset contains a lot of attributes(columns), the whole dataframe structure cannot be printed on the screen, so let's print out the dataframe as tibble:

```r
as_tibble(covid_data)
```

```
# A tibble: 187 x 15
   Country.Region      Confirmed Deaths Recovered Active New.cases New.deaths
   <chr>                   <int>  <int>     <int>  <int>     <int>      <int>
 1 Afghanistan             36263   1269     25198   9796       106         10
 2 Albania                  4880    144      2745   1991       117          6
 3 Algeria                 27973   1163     18837   7973       616          8
 4 Andorra                   907     52       803     52        10          0
 5 Angola                    950     41       242    667        18          1
 6 Antigua and Barbuda        86      3        65     18         4          0
 7 Argentina              167416   3059     72575  91782      4890        120
 8 Armenia                 37390    711     26665  10014        73          6
 9 Australia               15303    167      9311   5825       368          6
10 Austria                 20558    713     18246   1599        86          1
# i 177 more rows
# i 8 more variables: New.recovered <int>, Deaths...100.Cases <dbl>,
#   Recovered...100.Cases <dbl>, Deaths...100.Recovered <dbl>,
#   Confirmed.last.week <int>, X1.week.change <int>, X1.week...increase <dbl>,
#   WHO.Region <chr>
```

Total cases by month

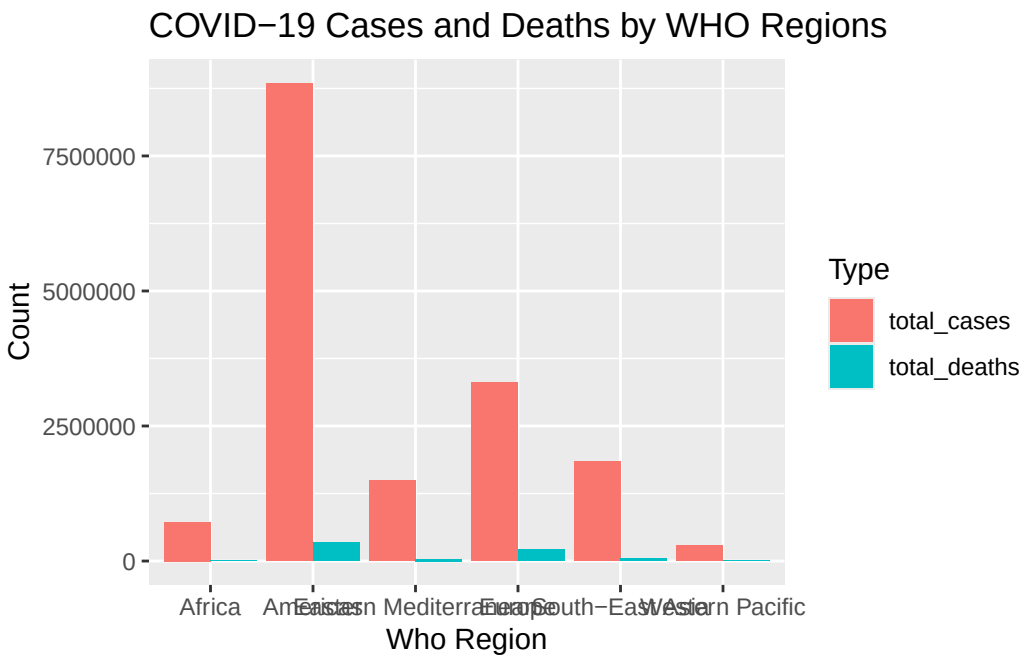
The selected dataset is very large and contains a huge amount of information. In order for us users to understand it, we should manipulate the data to extract the most meaningful insights from it. The example below shows the summary of all COVID-19 confirmed cases and deaths in all countries that are present in the **World Health Organization**:

```r
covid_summary <- covid_data %>%
  group_by(WHO.Region) %>%
  summarise(total_cases = sum(Confirmed), total_deaths = sum(Deaths))
covid_summary
```

```
# A tibble: 6 x 3
  WHO.Region            total_cases total_deaths
  <chr>                       <int>        <int>
1 Africa                     723207        12223
2 Americas                  8839286       342732
3 Eastern Mediterranean     1490744        38339
4 Europe                    3299523       211144
5 South-East Asia           1835297        41349
6 Western Pacific            292428         8249
```

```r
covid_long <- covid_summary %>%
  pivot_longer(cols=c(total_cases, total_deaths), names_to = "Type", values_to = "Count")

ggplot(covid_long, aes(x=WHO.Region, y = Count, fill = Type)) +
  geom_bar(stat = "identity", position = "dodge")+
  labs(title = "COVID-19 Cases and Deaths by WHO Regions",
       x = "Who Region", y = "Count", fill = "Type")
```



The previous data about total cases and total deaths in the WHO Regions, can be seen in a clearer way in the table Table 1, shown below:

```
covid_summary %>% gt()
```

Table 1: Total cases and deaths in all WHO regions

WHO.Region	total_cases	total_deaths
Africa	723207	12223
Americas	8839286	342732
Eastern Mediterranean	1490744	38339
Europe	3299523	211144
South-East Asia	1835297	41349
Western Pacific	292428	8249

COVID-19 Symptoms

Typical **COVID-19 symptoms** often show up 2 to 14 days after contact with the virus.

Symptoms can include:

- *Dry cough*
- *Shortness of breath*
- *Loss of taste or smell*
- *Extreme tiredness, calles fatigue*
- *Digestive symptoms such as upset stomach, vomiting or loose stools, calles diarrhea*
- *Pain, such as headaches and body or muscle aches*
- *Fever or chills*
- *Cold-like symptoms such as congestion, runny nose or sore throat*

COVID-19 numbers

Forecasting cumulative infected case numbers

The single, simple formula that researchers used in this study is an application of the Verhulst-Pearl logistic function, stated as:

$$N = \frac{PopulationCountry^{*}}{1 + be^{-ct}}$$

where N represents the cumulative number of infected cases at time t, e represents a mathematical constant approximately equal to 2.71828, c is the constant of integration, and b is the exponential function base.

Predicting number of deaths

In the case of the number of deaths in a given country caused by Covid-19, denoted as N(t), we have the following formula:

$$N(t) = \frac{N_f}{1 + \beta e^{-kt}}$$

The formula and its description can be found in the article Athanassios (2019).

Figure 1 depicts the total number of deaths as a function of time after the day that 25 deaths had occured, for the epidemics in Spain, Germany, Italy and the UK.

```
knitr::include_graphics("images/image2.jpg")
```

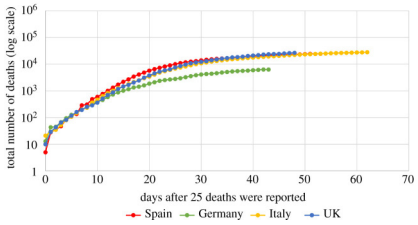


Figure 1: Estimate of infection fatality rate

For more detailed information, you can visit the World Health Organization website.Moreover I personally suggest the book Haseltine and Patarca (2023) if you want to get more in touch with the topic.

Athanassios, S. Fokas. 2019. "Covid-19: Predictive Mathematical Formulae for the Number of Deaths During Lockdown and Possible Scenarios for the Post-Lockdown Period." *Land Use Policy*, May, 390–402. https://doi.org/10.1016/j.landusepol.2019.02.004.

Haseltine, A. William, and Roberto Patarca. 2023. "The COVID-19 Textbook - Science, Medicine, and Public Health." {SSRN} {Scholarly} {Paper}. Rochester, NY. https://books.google.si/books?id=IuTiEAAAQBAJ&printsec=frontcover&hl=sl&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false.