Visualisation and Analyzation of The Novel Coronavirus (2019-nCoV) Dataset

Abstract

The novel coronavirus (COVID-19) that was first reported at the end of 2019 has impacted almost every aspect of life as we know it. Since, having methods to visualize and visually analyse a big data makes it easier to comprehend the data, never before has visualisation been more important than in the COVID-19 pandemic. In this project, we present a big data visualization for visualizing and analysing the COVID-19 epidemiological data.

1. Introduction

SARS-CoV-2 is a virus that is the cause of a serious life-threatening disease known as COVID-19. It was first noted to have occurred in Wuhan, China in November 2019 and the WHO reported the first case on December 31, 2019. The outbreak was declared a global pandemic on March 11, 2020 and by May 30, 2020, a total of 5 899 866 positive cases were registered including 364 891 deaths.

The Covid-19 vaccine has changed the course of the pandemic. In total, 162 doses have been given for every 100 people around the world. In countries that reached the highest levels of vaccination, death rates plummeted. Over a year into the biggest vaccination campaign in history, more than 12.7 billion doses have been administered across 184 countries.

However, the COVID-19 patterns are sometimes contradictory and this is where visualizations come in. This project focuses on covid cases at different time stamps and vaccination status worldwide.

2. Dataset Description

From the R package **coronavirus,** I have worked with the following 2 datasets:

1. Coronavirus Dataset:

   The dataset contains the daily summary of Coronavirus cases (confirmed, death, and recovered), by state/province

   The dataset has the following columns:

   1. **date** - The date of the observation, using Date class

2. **province** - Name of province/state, for countries where data is provided split across multiple provinces/states
3. **country** - Name of country/region
4. **lat** - The latitude code
5. **long** - The longitude code
6. **type** - An indicator for the type of cases (confirmed, death, recovered)
7. **cases** - Number of cases on given date
8. **uid** - Country code
9. **province state** - Province or state if applicable
10. **iso2** - Officially assigned country code identifiers with two-letter
11. **iso3** - Officially assigned country code identifiers with three-letter
12. **code3** - UN country code
13. **fips** - Federal Information Processing Standards code that uniquely identifies counties within the USA
14. **combined_key** - Country and province (if applicable)
15. **population** - Country or province population
16. **continent_name** - Continent name
17. **continent_code** - Continent code

2. Vaccination Dataset

The dataset contains data about population in selected countries and territories worldwide that had received a COVID-19 vaccination as of December 14, 2020

This dataset includes the following columns:

1. **country_region** - Country or region name
2. date - Data collection date in YYYY-MM-DD format
3. **doses_admin** - Cumulative number of doses administered. When a vaccine requires multiple doses, each one is counted independently
4. **people_partially_vaccinated** - Cumulative number of people who received at least one vaccine dose. When the person receives a prescribed second dose, it is not counted twice
5. **people_fully_vaccinated** - Cumulative number of people who received all prescribed doses necessary to be considered fully vaccinated
6. **report_date_string** - Data report date in YYYY-MM-DD format
7. **uid** - Country code
8. **province_state** - Province or state if applicable
9. **iso2** - Officially assigned country code identifiers with two-letter
10. **iso3** - Officially assigned country code identifiers with three-letter
11. **code3** - UN country code
12. **fips** - Federal Information Processing Standards code that uniquely identifies counties within the USA
13. **lat** - Latitude
14. **long** - Longitude
15. **combined_key** - Country and province (if applicable)
16. **population** - Country or province population

17. **continent_name** - Continent name
18. **continent_code** - Continent code

3. Graphical presentation of Key variables

Please view the dashboard for the plots: https://paulanwesha.shinyapps.io/covid/

4. Summary of Analysis

I have plotted the following:

1. Total Covid Cases Region-wise
2. Relationship Between cumulative Active v/s cumulative Death Cases and cumulative Active v/s cumulative Recovered Cases
3. Relationship Between cumulative Recovered and cumulative Death Cases
4. Confirmed Cases Worldwide
5. Active Cases Worldwide
6. Recovered Cases Worldwide
7. Death Cases Worldwide
8. cumulative COVID-19 cases in the 5 most affected countries
9. Partially Vaccinated Ratio Plot of different Countries
10. Fully Vaccinated Ratio Plot of different Countries

The summary (i.e., descriptive statistics) of every plot has been shown in the dashboard.

5. Conclusion

1. We observed that initially death cases increase more rapidly as compared to recovered cases whereas rate of recovery is slower initially compared to rate of increase in the active total.
2. Also, we observe a linear trend while we plotted the cumulative Recovered and cumulative Death Cases.
3. The current increase in the number of confirmed cases over time can be modelled by polynomial regression of degree 2.
4. However, new models have to be developed to reliably predict the future number of confirmed cases considering continuation and relaxation of lockdowns and other aspects.