A PROJECT REPORT
ON
Timeline Analysis of Covid-19

MASTER IN COMPUTER SCIENCE
SPECIALIZATION IN
DATA SCIENCE

SUBMITTED BY

MANJIRI HARISHCHANDRA SAWANT

UDCSDS 312

Date 26th April 2021

MUMBAI UNIVERSITY

Department of Computer Science,
University of Mumbai
Ranade Bhavan, B-Wing, Ground Floor,
Vidyanagari Campus, Kalina,
Santacruz (East), Mumbai-400098.

ACADEMIC YEAR: 2018 – 2020

Abstract

At the end of 2019 a novel virus, severe acute respiratory syndrome **Coronavirus 2 (SARS-CoV-2)**, causing severe respiratory syndrome expanded globally from **Wuhan, China**. In March 2020, the World Health Organization declared the SARS-Cov-2 virus a **global pandemic**.

COVID-19(Coronavirus) pandemic has created surge demand for essential healthcare equipment, medicines along with the requirement for advanced information technologies applications.

There have been a lot of previous studies that employed machine learning and statistical approaches to capture the patterns and trends of a number of varying events related to infectious diseases. Time series analysis is popular used to forecast different diseases such as SARS, Ebola, pandemic influenza and dengue.

For as long as we have been recording data, time has been a crucial factor. In **time series analysis**, time is a significant variable of the data. Time series analysis helps us to study our world and learn how we progress within it.

Data visualization is a quite new and promising field in computer science. It uses computer graphic effects to reveal the patterns, trends, relationships out of datasets.

The **objective** is to perform **exploratory data analysis** to identify interesting pattern and possible reasons helping **Covid-19 spread** with basic as well as advanced charts.

Acknowledgement

I take this opportunity to express my profound gratitude and indebtedness to **Prof. (Smt.) Jaya Bharti** for giving me the opportunity to accomplish this project.

I express my sincere thanks to **Prof. Dr. Jyotshna Dongardive, Head of Department, & University Department of Computer Science,** to encourage me and to provide me the opportunity to prepare the project.

Last but not least, my family is also an important inspiration for me. So with due regards, I express my gratitude's to them.

Table of Contents

Introduction

There exist many devastating epidemics that affected not only human health but also the economic condition of the countries, viz. American polio epidemic (1916), Spanish Flu (1918 - 1920), Asian Flu (1957 -1958), AIDS pandemic and epidemic (1918-present day), H1N1 Swine Flu pandemic (2009–2010), West African Ebola epidemic (2014–2016), Zika Virus epidemic (2015-present day), etc.

But now, the whole world is facing the most dangerous disease **COVID-19** due to Coronavirus. The mortality rate due to Coronavirus is different in different countries.

This virus is impacting people of all the ages, but it would be more severe for those people who are older and already affected by some diseases like asthma, diabetes, heart disease, etc. This virus is newly developed and discovered for the first time in **2019**. This virus is unique, that's why it will take time to develop its medicine and vaccine.

The symptoms of this disease are sore throat, cough, headaches, etc. Coronavirus enters the body of the human through the mouth, nose, and eyes, so we should avoid touching these body parts. Social distancing, covering the mouth, and wearing gloves are made mandatory to reduce the contagious pandemic of COVID-19.

Currently, COVID-19 is a **green research topic** as the whole world is suffering and struggling in this time due to this Pandemic disease. As there is no medication and strategy to get rid of this global crisis, it is important to investigate data sets to discover patterns, to spot anomalies for future preparedness. This work is useful for to get clear idea of what the **information means by giving it visual context**. This makes data more natural for human mind to comprehend and therefore makes it easier to support several social entities and organizations, viz. hospitals, pharmaceuticals, NGOs, Government bodies, etc for their readiness to combat.

The dataset is large, and the analysis can be done using different techniques and perspectives.

Implementing Visualization

There have been a lot of previous studies that employed **exploratory data analysis** and **data visualization** approaches to capture the patterns and trends of a number of varying events related to infectious diseases.

Human has a long history with basic data visualization, and data visualization is still a hot topic today. The history of visualization was shaped to some extent by available technology and by the pressing needs of the time, they include: primitive paintings on clays, maps on walls, photographs, table of numbers (with rows and columns concepts), these are all some kind of data visualization – although we may not call them under this name at that time.

Why do we do Data Visualization?

To see and understand pictures is one of the natural instincts of human, and to understand numerical data is a years training skill from schools, and even so, a lot of people are still not good with numerical data.

While data visualization is a way that present and display information in a way that encourages appropriate interpretation, selection, and association. It utilizes human skills for pattern recognition and trend analysis, and exploits the ability of people to extract a great deal of information in a short period of time from visuals presented in a standardized format.

Getting information from pictures is far more time-saving than looking through text and numbers – that's why many decision makers would rather have information presented to them in graphical form, as opposed to a written or textual form.

Effective visualization for big data techniques need to go beyond just painting pretty pictures for management. Experts say enterprises can improve their results by considering layout, designing iteratively, engaging users and understanding business needs.

Keep the user in Mind

- Use color, form, size and placement to inform the design and use of visualization. Aspirent uses color to draw attention to the aspects of the analysis that it wants users to focus on.

- Size effectively communicates quantity, but too much use of different sizes to communicate information can become confusing. Form determines the shapes used to present an analysis: for example, whether to use line or bar charts to present certain types of information.

Tell a coherent story

- Speak to the audience and keep design simple and focused. Minute details like colors to the number of charts can help ensure that a dashboard tells a coherent story. A dashboard, much like a book, needs design elements that keep the reader in mind and does not simply force fit all the data one has access to.

Prepare to design iteratively

- Data exploration sparks new ideas and questions over time, and making it more pertinent over time and over adoption makes users smarter. Building a quick proof of concept, getting feedback quickly and iterating tends to lead to a better result, faster.

Personalize Everything

- Make sure that the dashboard reveals personalized information to the end user, and make it relevant. Ensuring that the visualizations are responsive in design to the devices they're on and offering offline access to end users will take it a long way.

- Engage your audience, and propagate a data culture by using **well-designed and interactive visualizations** to make analytics engaging and fun.

Start with the analysis objective

- Ensure that the data type and analysis objective informs what visualization type is chosen. A simple table or bar chart may sometimes be most effective for visualization for big data projects.

Keep governance in mind

- This might take work, but it's important that end users trust the data. Gather all the help one's need from a technology, process and people standpoint to ensure that the data is vetted and accurate.

Empathize with the viewers

- Every situation invites a different approach to visualization. For example, many data visualizers categorically forbid the use of pie charts , since the human eye and mind can much more readily measure differences between lengths or positions than between angles.

- But turning a pie chart into a bar chart can make it less obvious that the different parts add up to a single whole. It is also harder to recognize certain fractions. Consider variations like a donut chart, which is just a pie chart missing the middle for quickly showing patterns like a 75%, 20% and 5% breakdown.

Understand the business

- Spend some time with business users to understand what they want to achieve from the visualization for big data product, and what data needs to be used to provide them with the required insights. If necessary, invest in some tools or techniques to profile and transform the data. We are dealing with tons of data in big data -- hence, empowering the user to gain from this scale is important.

Connect Visualization

- Ensure that different visualizations on the dashboard are connected and can be linked instantly to complete the story.

- Make sure visualizations for big data can be updated and interrogated in real time. A **static display**, or one that doesn't reveal where its underlying data comes from, is not helpful in analyzing the fast-moving, varied streams of big data that enterprises must process.

Simplify where Possible

- Most leading visualization for big data tools are feature-rich, which often results in a tendency by analysts to build visualizations that are dense and overly complex. This can make it difficult to glean actionable insights. A good analyst will simplify visualizations to address only the hypothesis at hand.

Information of Dataset

The complete COVID-19 dataset is a collection of the COVID-19 data maintained by Our World in Data (Coronavirus Source Data - Our World in Data). This data is updated daily and includes data on confirmed cases, deaths, hospitalization, testing and vaccinations as well as other variables of potential interest.

Data Source

- *Confirmed Cases and deaths:* This data comes from COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU).
- *Hospitalizations and intensive care unit (ICU) admissions:* This data comes from European Centre for Disease Prevention and Control (ECDC) for a select number of European countries; the government of the United Kingdom; the Department of Health & Human Services for the United States; the COVID-19 Tracker for Canada. But they are unable to provide data on hospitalizations for other countries.
- *Testing for COVID-19:* This data is collected by the Our World in Data team from official reports. The testing data is updated around twice a week.
- *Vaccinations against COVID-19*: This data is collected by the Our World in Data team from official reports.
- *Other variables:* This data is collected from a variety of sources (United Nations, World Bank, Global Burden of Disease, Blavatnik School of Government, etc.)

This data is used to **study timeline analysis of COVID-19** and employed to **perform data visualization in tableau.**

Before, dive into the visualizations, **exploratory data analysis** is an approach to data analysis that uses basic data exploration like –

- head of the dataset
- the shape of the dataset
- info of the dataset
- summary of the dataset

Data Analysis

Exploratory data analysis is one of the best practices used in data science today. Data Analysis is the statistics and probability to figure out trends in the data set. It is used to show historical data by using some analytics tools. It helps in drilling down the information, to transform metrics, facts, and figures into initiatives for improvement.

I will explore a Data set and perform the exploratory data analysis. The major tasks to be covered are below:

- Handle Missing Values
- Removing Duplicates

World's COVID – 19 dataset is really large dataset and it is increasing rapidly country by country dramatically. This real world data has a lot of missing values. The cause of missing values can be data corruption or failure to record data.

In statistics, missing data, or missing values, occur when no data value is stored for the variable in an observation. Missing data are a common occurrence and can have a significant effect on the conclusions that can be drawn from the data.

Quick classification of missing data

There are three types of missing data:
MCAR: Missing Completely At Random. It is the highest level of randomness. This means that the missing values in any features are not dependent on any other features values. This is the desirable scenario in case of missing data.

MAR: Missing At Random. This means that the missing values in any feature are dependent on the values of other features.

MNAR: Missing Not At Random. Missing not at random data is a more serious issue and in this case, it might be wise to check the data gathering process further and try to understand why the information is missing.

This Data contains huge numbers of values which are Completely Missing at Random. This effectively implies that causes of the missing data are unrelated to the data.

What to do with the missing values?

- Ignore the missing values
- Drop the missing values
- Case deletion
- Imputation
- Imputation by Mean/Mode/Median

I have identified the missing values in a data, so next to check the extent of the missing values to decide to further course of action.

Data is coming from; different countries with wide range of complexities so missing value should better represented by the number zero. Replacing missing values with zeros is accomplished using **Imputation.**

Imputation is the process of substituting the missing data by some statistical methods. Imputation is useful in the sense that it preserves all cases by replacing missing data with an estimated value based on other available information. But imputation methods should be used carefully as most of them introduce a large amount of bias and reduce variance in the dataset.

In summation, handling the missing data is crucial for a data science project. However, the data distribution should not be changed while handling missing data. Any missing data treatment method should satisfy the following rules:

- Estimation without bias — any missing data treatment method should not change the data distribution.

- The relationship among the attributes should be retained.

Duplicate records are not encountered in a data set. Data quality is fine with accurate records**. Data quality** is crucial – it assesses whether information can serve its purpose in a particular context of data analysis.

The purpose of data analysis is to gain insights; data is much more valuable when it is visualized. Visually-displayed data is much more accessible, and it's critical to promptly identify the weaknesses of that difficulty.

Overview of Tableau Software

Tableau was founded in 2003 as a result of computer science project at Stanford that aimed to improve the flow of analysis and make data more accessible to people through visualization. This includes making machine learning, statistics, natural language, and smart data prep more useful to augment human creativity in analysis.

Tableau Features

Tableau provides solutions for all kinds of industries, departments, and data environments. Following are some unique features which enable Tableau to handle diverse scenarios.

- *Speed of Analysis* − As it does not require high level of programming expertise, any user with access to data can start using it to derive value from the data.

- *Self-Reliant* − Tableau does not need a complex software setup. The desktop version which is used by most users is easily installed and contains all the features needed to start and complete data analysis.

- *Visual Discovery* − The user explores and analyzes the data by using visual tools like colors, trend lines, charts, and graphs. There is very little script to be written as nearly everything is done by drag and drop.

- *Blend Diverse Data Sets* − Tableau allows you to blend different relational, semistructured and raw data sources in real time, without expensive up-front integration costs. The users don't need to know the details of how data is stored.

- *Architecture Agnostic* − Tableau works in all kinds of devices where data flows. Hence, the user need not worry about specific hardware or software requirements to use Tableau.

- *Real-Time Collaboration* − Tableau can filter, sort, and discuss data on the fly and embed a live dashboard in portals like SharePoint site or Salesforce. User can save view of data and allow colleagues to subscribe to your interactive dashboards so they see the very latest data just by refreshing their web browser.

- *Centralized Data* − Tableau server provides a centralized location to manage all of the organization's published data sources. User can delete, change permissions, add tags, and manage schedules in one convenient location. It's easy to schedule extract refreshes and manage them in the data server.

Results of an Experiment

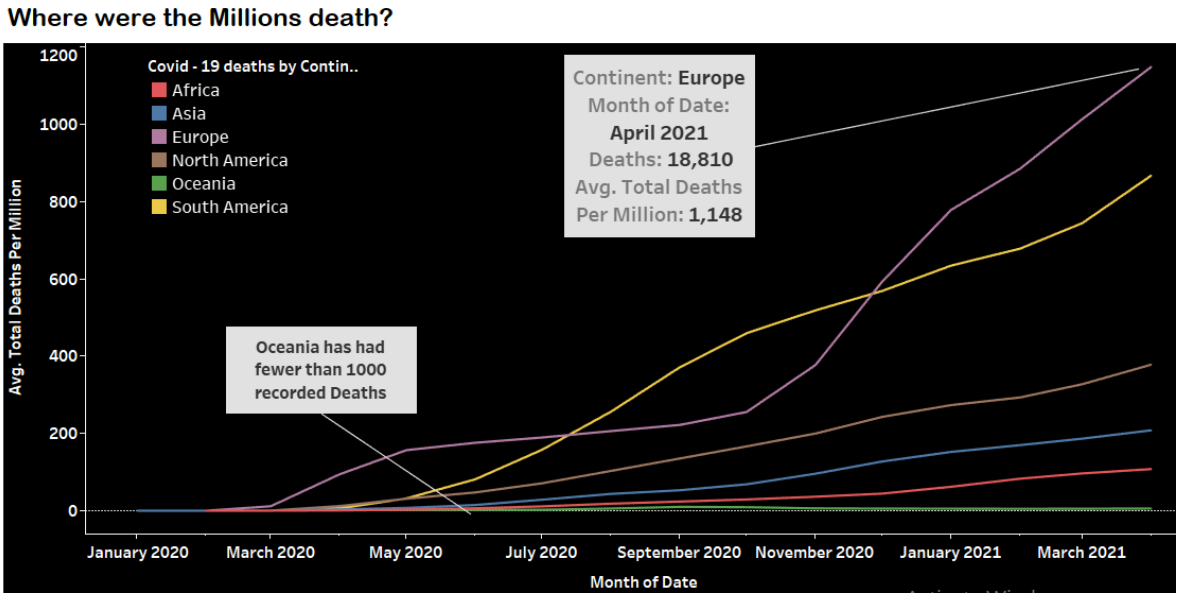
Why Tableau is So Popular?

It is fair to say that Tableau is an extremely powerful tool, and has no equal in data visualization capabilities and performances. Not only for its speed, scalability, but also for offering the best tools to create visual answers to most of the business questions, from bar charts to more complex visualization, including rich maps with tons of customizations. And best of all, Tableau is **free to use** for the **Tableau Public version.**

I have selected **Tableau Software** for data visualization for the understanding of Timeline analysis of COVID – 19.

The COVID-19 pandemic, also known as the coronavirus pandemic, is an **ongoing global pandemic** of coronavirus disease 2019 (COVID-19).

The ongoing COVID-19 pandemic has caused worldwide socioeconomic unrest, forcing governments to introduce extreme measures to reduce its spread. Data visualization is having a big moment during the COVID-19 pandemic. Social media feeds are inundated with infection heat maps and charts depicting transmission patterns.

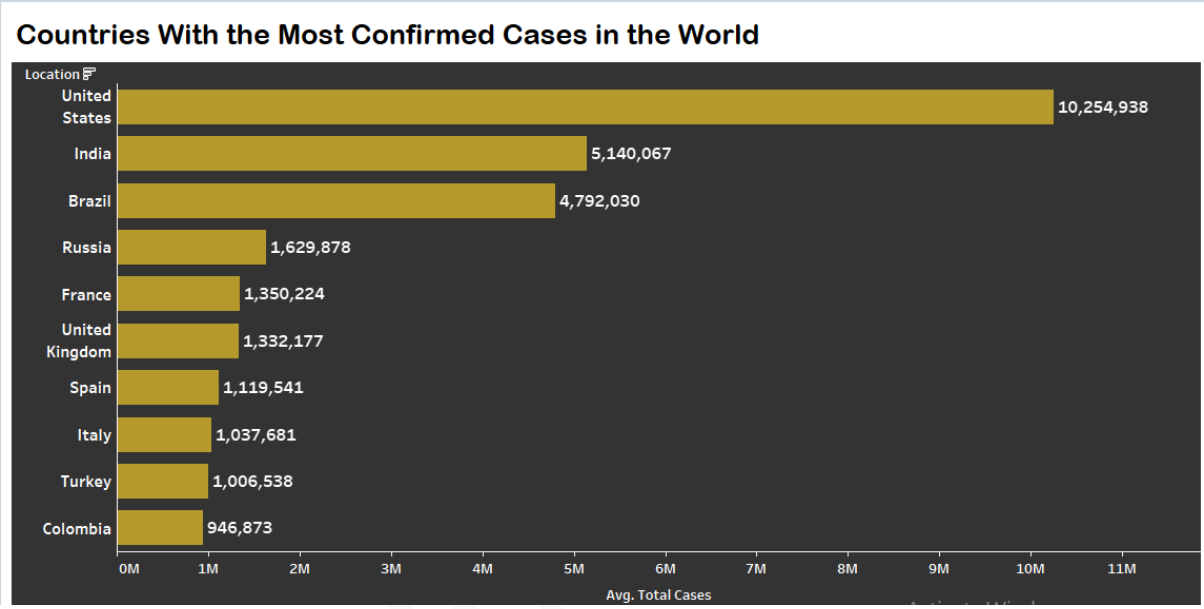


(Source: OWID, data to 24 Apr Fig 1.1)

The number of people who have died worldwide in the Covid-19 pandemic has surpassed **three million**, according to Johns Hopkins University.

India - experiencing a second wave - recorded more than 230,000 new cases. Almost 140 million cases have been recorded since the pandemic began. **The US, India and Brazil - the countries with the most recorded infections -** have accounted for more than a million deaths between them, according to Johns Hopkins University.

Fig 1.1 shows where the COVID – 19 spreading fast by Continent. The Europe has recorded 1,148 Avg Total Deaths per million by the month April 2021. After second wave of cases it appears to be on rise again now.

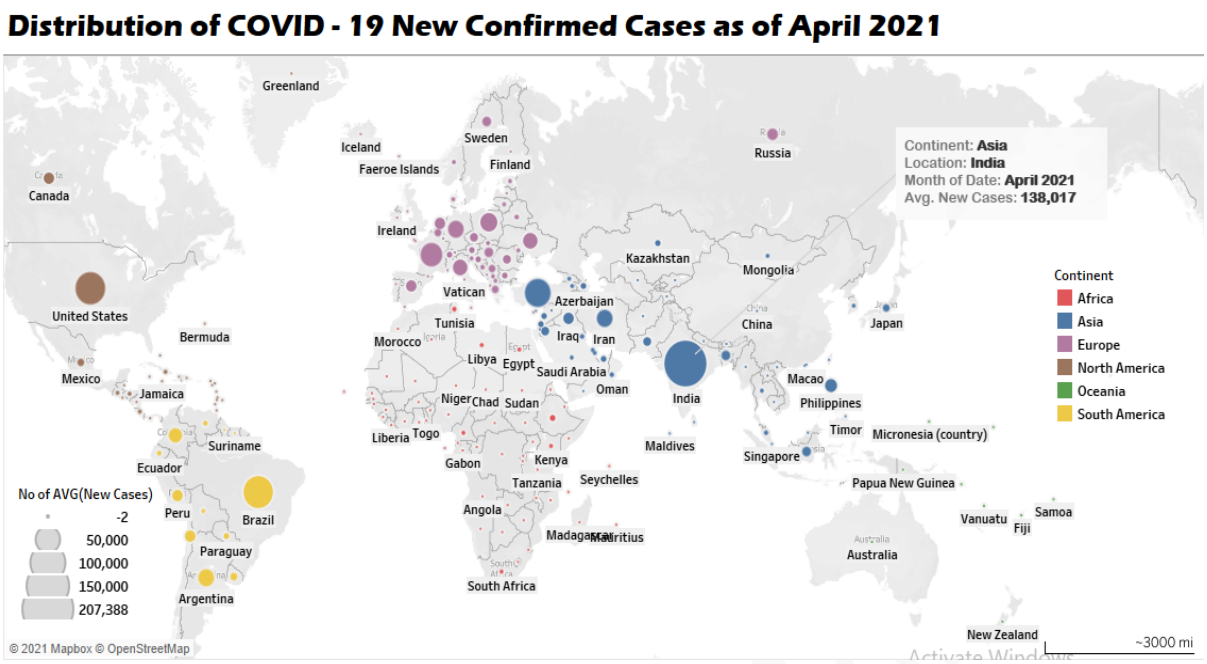


(Source: OWID, data to 24 Apr Fig 1.11)

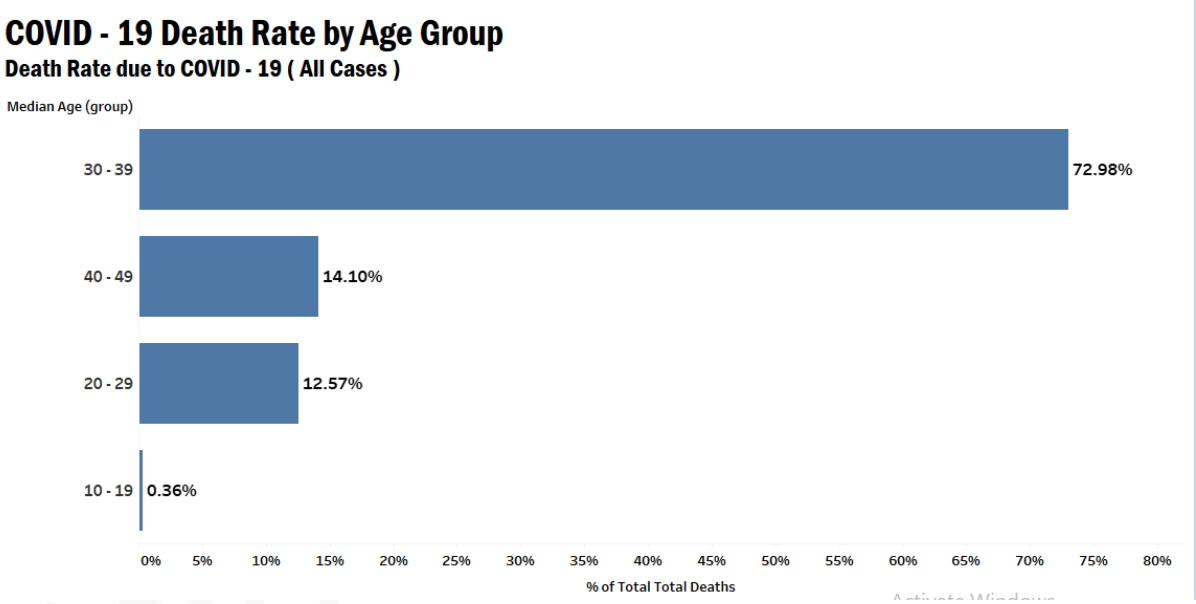
The coronavirus has been spreading fast in India, with the country recording about around 300 thousand cases a day in the end of April. Confirmed infections in India have reached sixteen million - the second-highest after the US.

Brazil has the highest number of deaths in Latin America and has recorded more than 14 million cases, the third highest in the world. Elsewhere in the region, newly confirmed infections are also rising quickly in Argentina, which now has more than 2 million cases. Even Russia has recording new confirmed cases is again at the highest peak.

In **Fig 1.2,** this map tracks the novel coronavirus outbreak in each country worldwide. Asia, South America, North America & Europe are beginning to see rise in daily cases compared to Africa and Oceania. Clearly we can observe from the map that **India**, **United State** and **Brazil** are now new **coronavirus hotspots**.



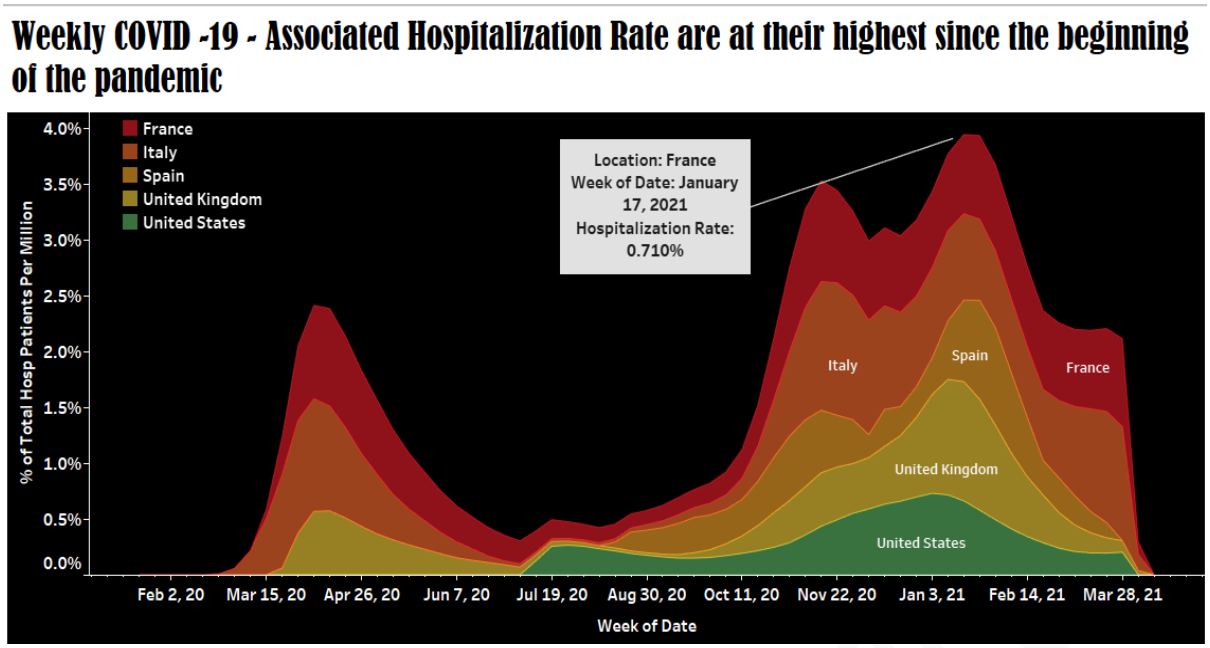
(Source: OWID, data to 20 April Fig 1.2)



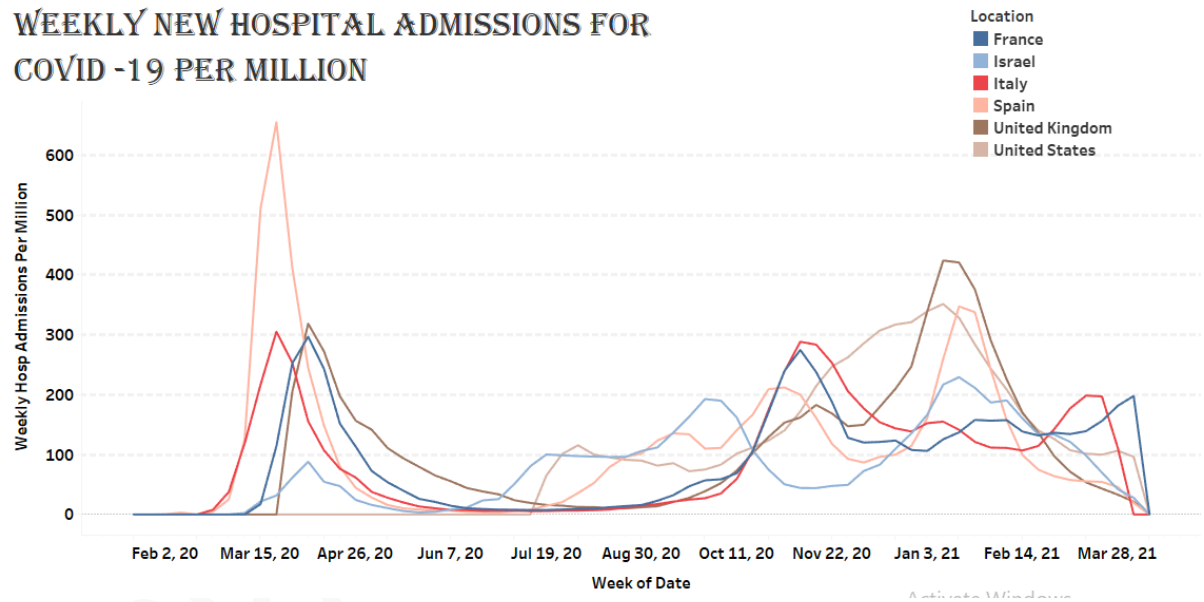
(Source: OWID, data to 20 Apr fig 1.3)

Early in the pandemic, COVID-19 incidence was highest among older adults. A fresh analysis of age-based statistics offers solid evidence that **COVID-19 is more widespread among young adults** (fig 1.3).

COVID-19 hospitalization rates rise exponentially with age. According to **The Royal Society Publishing**, the incidence of COVID-19 hospitalizations, in a number of countries, consistently doubles with every 16 years of age (R2 = 0.98 for top three countries (fig 1.4). (https://royalsocietypublishing.org/doi/10.1098/rsif.2020.0982)



(Source: OWID, data to 20 Apr fig 1.4)



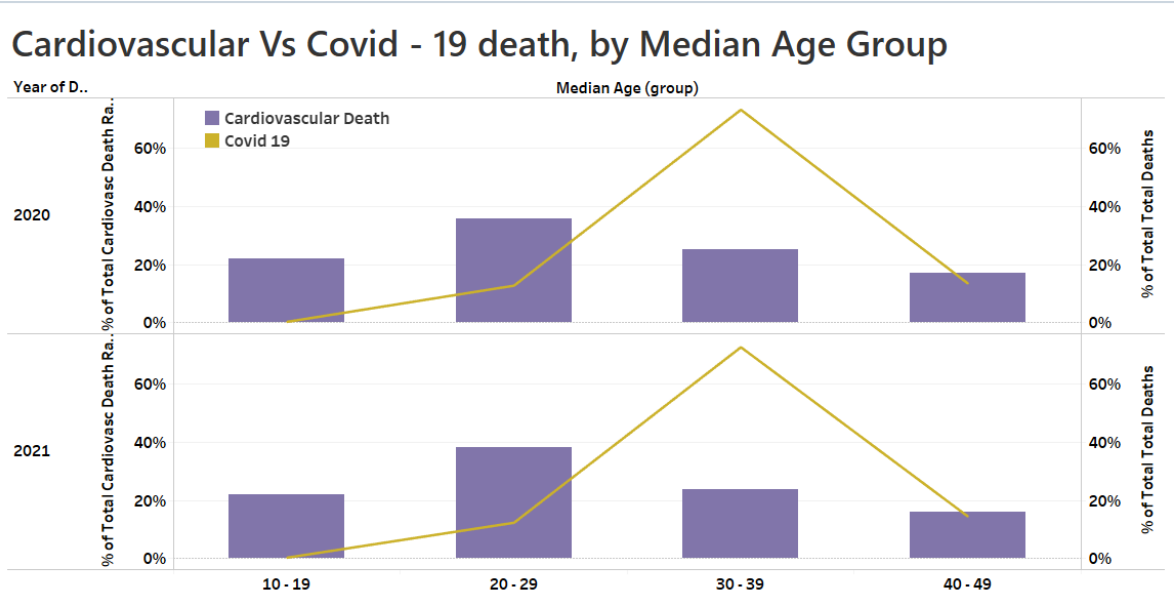
(Source: OWID, data to 21 Apr fig 1.41)

Fig 1.41, chart shows the weekly new admissions for COVID – 19 per million only provides the statistics only for **selected number of European Countries, United Kingdom and United State.** Unfortunately **source data on hospitalization on other countries are not available at present.** COVID- 19 Hospitalization data of India is not included currently.

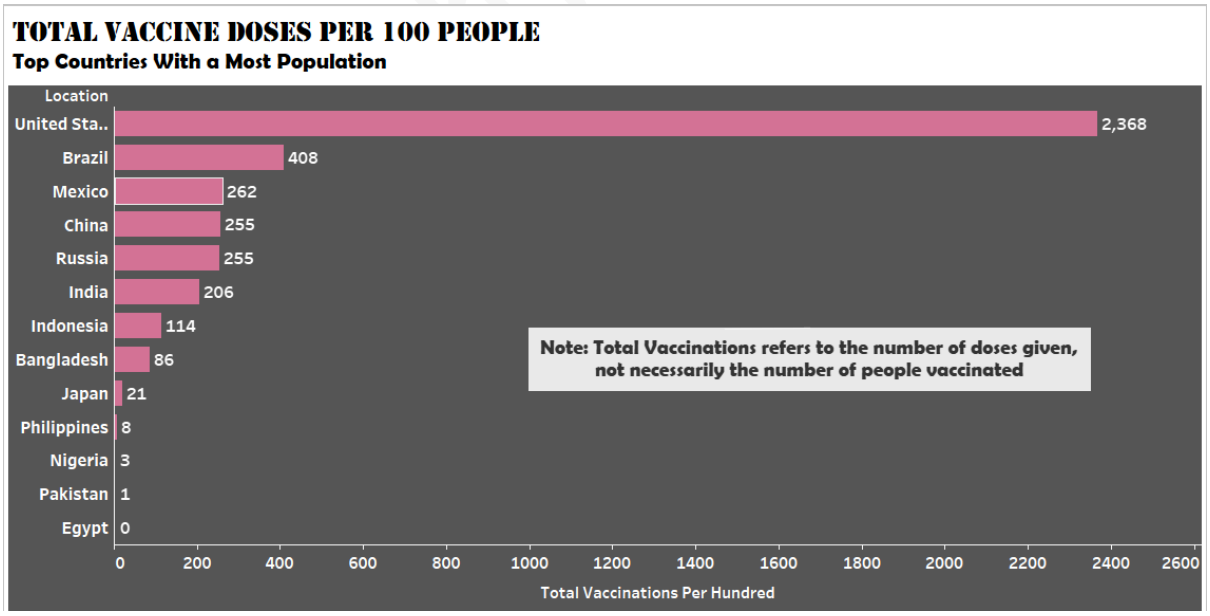
As COVID-19 cases rose, fewer people were being diagnosed with cardiovascular disease, likely putting many at higher risk for cardiac events like a heart attack.

As cardiovascular disease diagnoses went down, experts found that mortality rates from heart disease increased. Experts say many people likely avoided hospitals during COVID-19 surges even though they may have been having a cardiac event.

COVID-19 is a systemic disease, and not only did people who had underlying cardiovascular disease were more likely to get infected, their outcomes were worse (Fig 1.5).



(Source: OWID, data to 21 Apr fig 1.5)



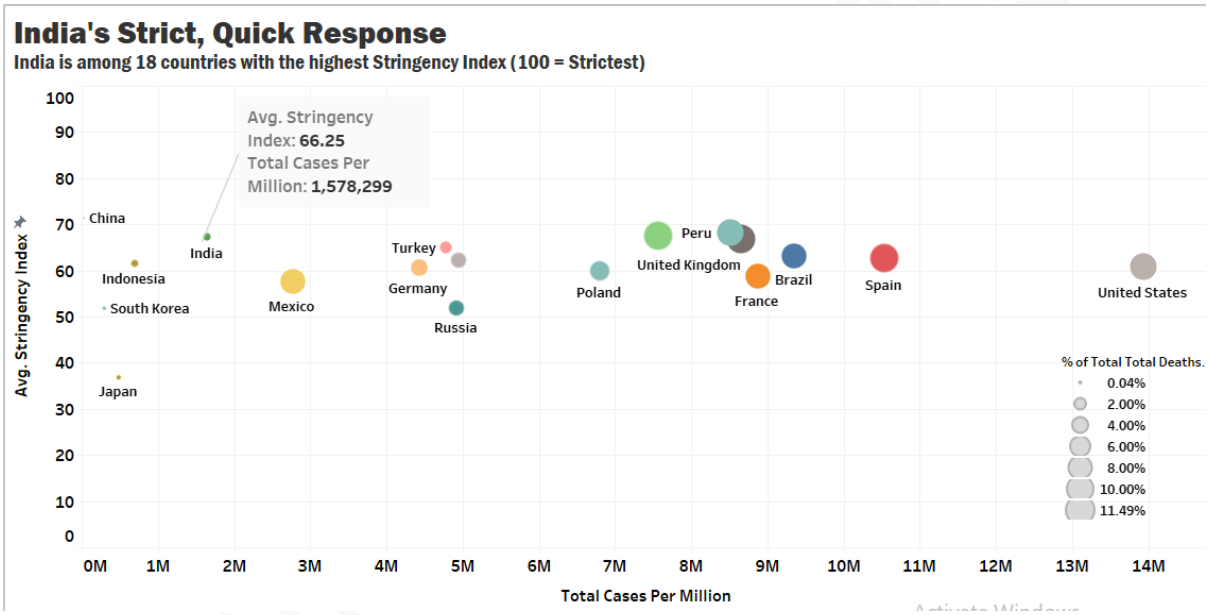
(Source: OWID, data to 24 Apr fig 1.6)

Fig 1.6 chart shows vaccine Inequality looks like. There is **massive inequality in vaccination rate globally**. Global data depicts that about 75% vaccines have gone to only 10 countries globally.

The countries that have managed to get a lot of people vaccinated — the U.S., the United Kingdom, India — all happen to have manufacturing plants that are producing the vaccines. They also have had export restrictions that have meant their own citizens have been at the front of the line to get immunized.

Important regional players such as South Africa have fully vaccinated only half of 1% of their population. In the Philippines it's less than 0.1%. Even wealthy nations in Europe such as Germany, Spain, Italy and France haven't gotten above 7%.

If the virus continues to spread and mutate for several more years, there's a good chance that a variant could emerge to which the vaccines provide no protection. Expert says, **“Getting the whole world immunized "is an investment in our own self-interest”.**



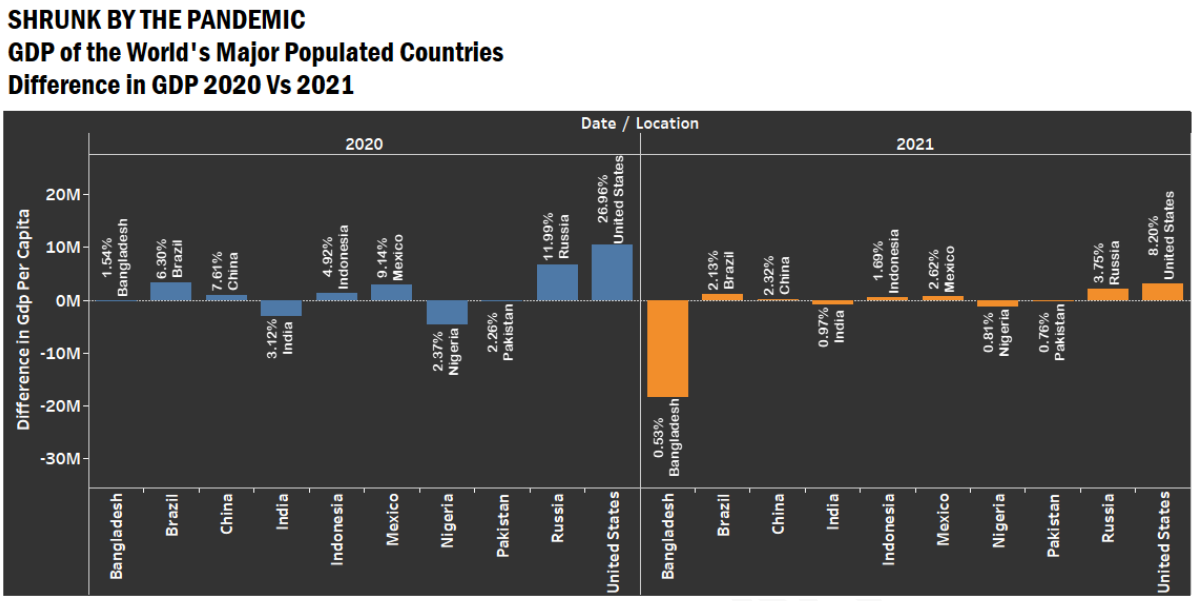
(Source: OWID, data to 24 Apr fig 1.7)

India enforced one of the strongest lockdowns at an early phase of case growth. Now, an index created by the University of Oxford quantifies that (Fig 1.7).

The Government Response Stringency Index is a score assign to countries by researchers at **the University of Oxford based on seven indicators such as closure of schools and workplaces, public events and restrictions on travels**, **stay at home policies**, etc. **The Stringency Index is a number from 0 to 100** that reflects these indicators. A higher index score indicates a higher level of stringency.

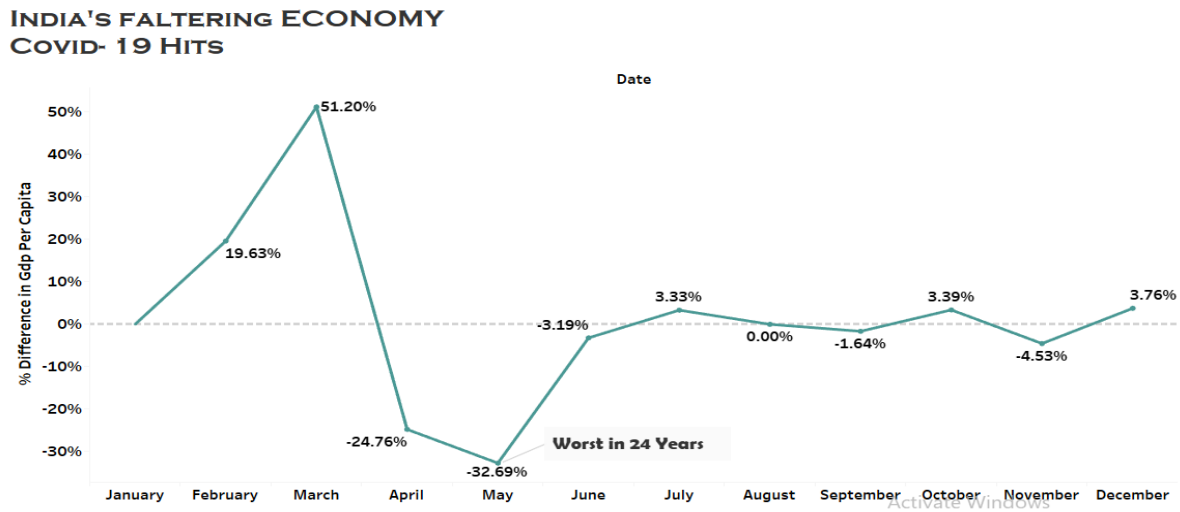
As China pulled stronger measures, its death curve plateaued.

In countries such as the UK, the US, and India, the Oxford graphs find that the death curve has not flattened after strictest measures were enforced.



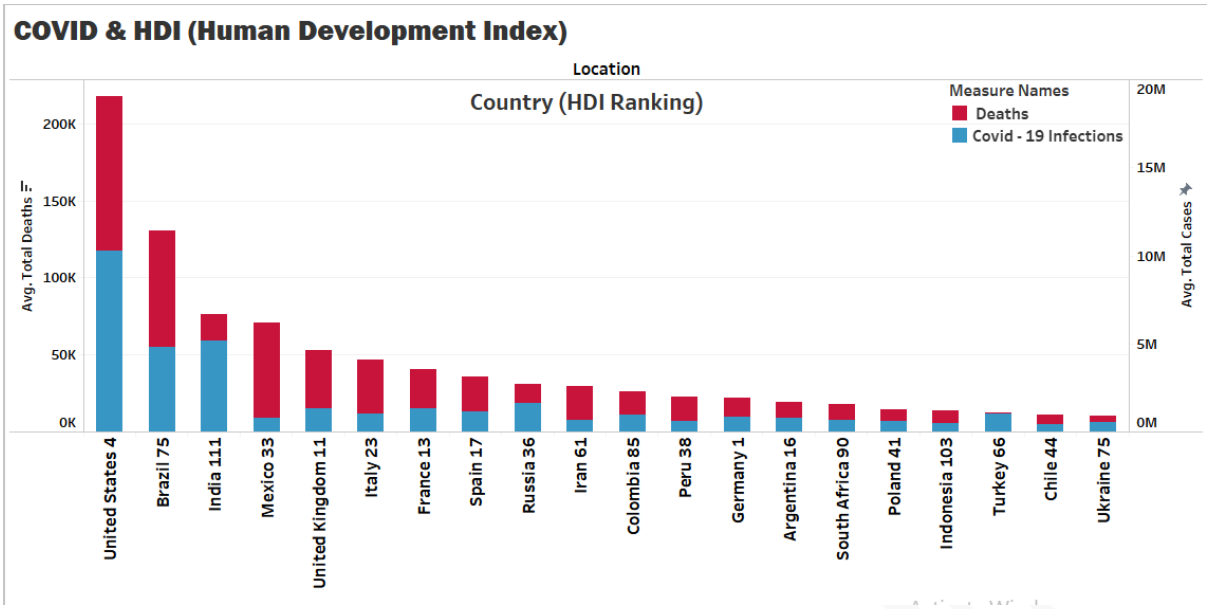
(Source: OWID, data to 24 Apr fig 1.8)

Top Economics have suffered worst GDP fall due to COVID-19. Barring China, the world's second-largest economy, all other major economies have felt the negative impact of coronavirus pandemic (Fig 1.8).



(Source: OWID, data to 24 Apr fig 1.81)

India's economic growth suffered its worst fall on record in the April-June quarter, with the gross domestic product (GDP) contracting **32.69 per cent**. The coronavirus-related lockdowns mainly weighed on the already-declining consumer demand and investment. The numbers are the worst since India started reporting quarterly data in 1996.



(Source: OWID, data to 24 Apr fig 1.9)

The COVID-19 pandemic has not only affected economic thinking on the path of GDP growth, but also on the path of **human development** (Fig 1.9). It has impacted one of the core foundations of the concept of human development i.e. a healthy and long life.

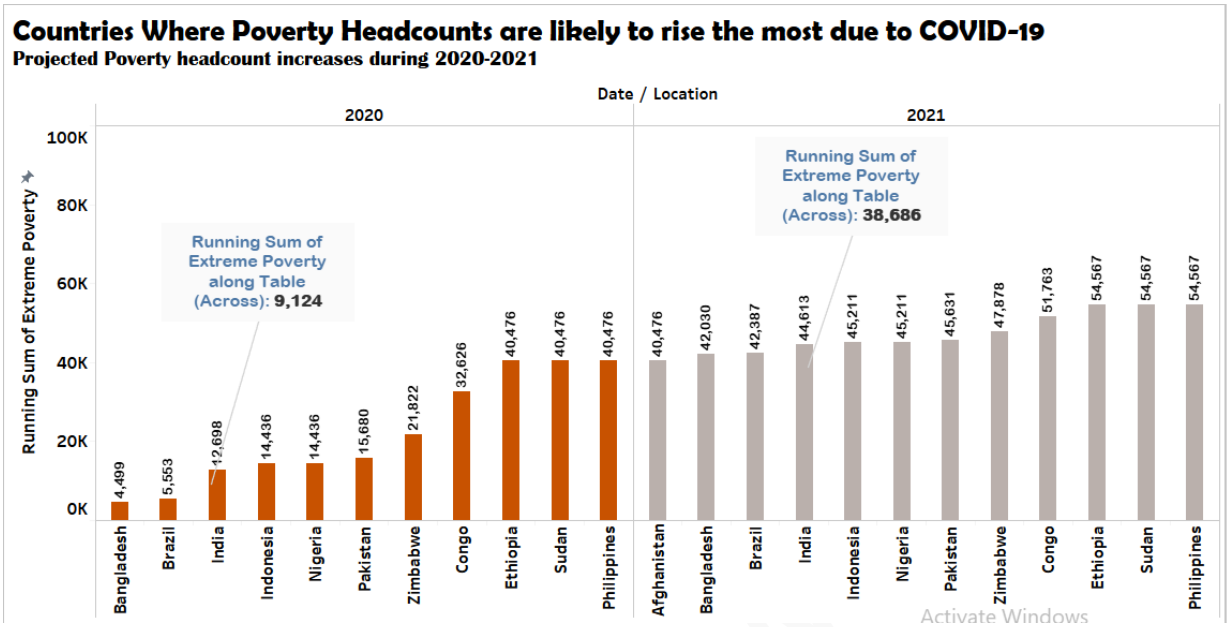
Human development is measured using three core indicators of life expectancy (at birth), knowledge (mean years of schooling) and standard of living (Gross National Income per capita). It is a critique of the 'economic growth before people' model of development. The c**urrent model of calculating human development has not been perfect** because of a lack of emphasis on epidemic prevention mechanisms within public health, while measuring performances of nations.

It is observed that though countries like Germany, Sweden, USA, Italy and the UK rank very high on the HDI and have the highest life expectancies around the world, they are oddly the worst affected. As of 2 May 2020, these countries topped the charts of coronavirus deaths worldwide per million populations.

The Human Development Index is an important index as it measures a country on its social and economic dimensions. However, the index introduced in 1990, needs to broaden its hemisphere by incorporating new variables. Coronavirus has thus given a wake-up call **to rethink the HDI in terms of pandemic preparedness** for the future.

Global extreme poverty is expected to rise in 2021 for the first time in over 20 years as the disruption of the COVID-19 pandemic compounds the forces of conflict and climate change, which were already slowing poverty reduction progress, the World

Bank said. COVID-19 has seriously affected these trends in ways that are still not clear-cut. The fig (2.0) below try to identify the most seriously affected countries.



(Source: OWID, data to 24 Apr fig 2.0)

The above chart shows 12 countries that are likely to see an increase in poverty of over 1 million people in 2020 as a result of COVID-19. They are in Asia and Africa, with Brazil as the sole exception. India and Nigeria stand out as likely to add 10 million and 8 million to the poverty rolls in 2020. In all these countries, COVID-19 has demonstrated the vulnerability of people who have only recently been able to escape poverty.

COVID - 19 Hitting African Poor Countries the hardest

Location	% of Total Gdp Per Capita	% of Total Total Tests	New Confirmed Cases	New Deaths	Total Vaccinations Per Hundred
Zambia	22.90%	34.48%	90,532	1,230	0
Mozambique	6.98%	15.84%	69,002	794	0
Madagascar	8.75%	2.30%	30,207	527	0
Democratic Republic of ..	5.11%	0.00%	28,769	745	0
Malawi	6.61%	2.34%	33,902	1,134	15
Rwanda	11.63%	34.88%	23,744	321	44
Lesotho	15.18%	0.00%	10,709	315	0
Togo	9.19%	10.15%	12,214	117	3
Guinea-Bissau	9.44%	0.00%	3,706	66	0
Burundi	4.22%	0.00%	3,424	6	0

AVG(Total Cases)
817 23,959

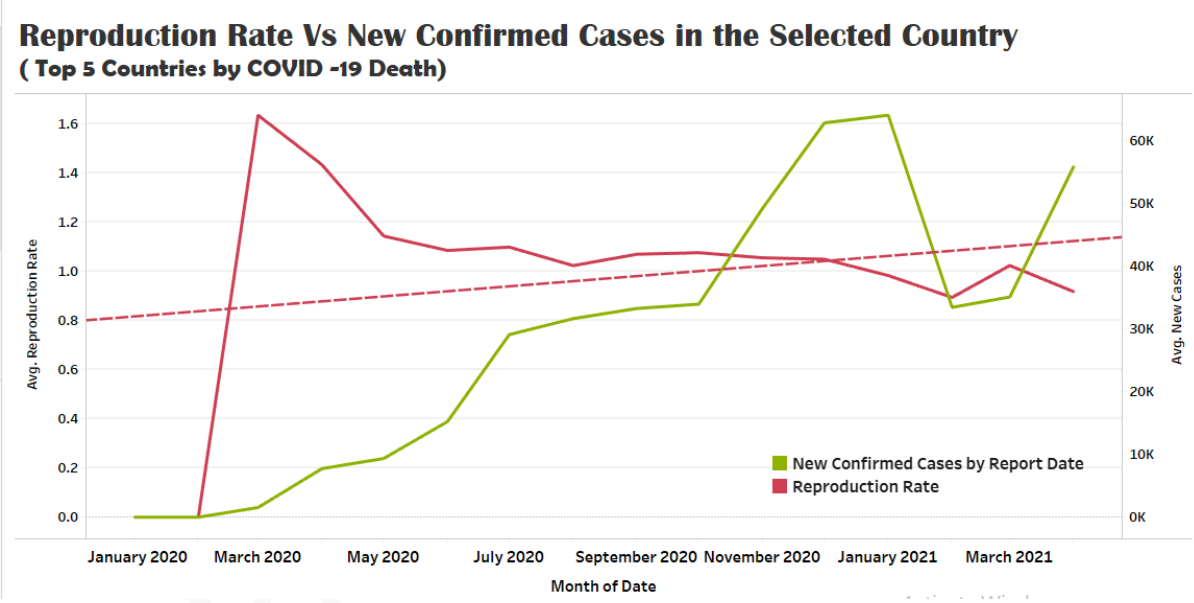
(Source: OWID, data to 24 Apr fig 2.1)

Social distancing is hard to apply or enforce in the slums of many developing country cities, and the safety net is not sufficiently well developed to allow people to stay at home without working and still feed their families.

As many as **nine out of 10 extremely poor people in the world currently live in Africa** and the novel coronavirus disease (COVID-19) will push an additional 5-29 million below the extreme poverty line (fig 2.1).

If the impact of the pandemic is not limited by 2021, an additional 59 million people could suffer the same fate, which would bring the total number of extremely poor Africans to 514 million people, warned the **Economic Commission for Africa (ECA).**

The ECA team presented the Integrated Planning and Reporting Toolkit, designed to help African countries facilitate the integration of the SDGs and Agenda 2063 into their national development plans and the monitoring of their implementation.



(Source: OWID, data to 24 Apr fig 3.0)

The COVID-19 reproduction rate from EpiForecasts is the average number of secondary infections generated by each new infectious case. In other words, it shows how easily the virus is spreading. The reproduction rates are calculated based on existing COVID-19 data and models, not observation samples (fig 3.0)

A reproduction rate of **higher than one** indicates that **the number of cases is likely to increase** exponentially. A reproduction level of **lower than one** indicates that the **virus will eventually peter out.**

The Countries that have 'ZERO' coronavirus cases

Location	Total Confirmed Cases	% of Total Life Expectancy	% of Total Cardiovasc Death Rate	Population	Total Vaccinations Per Hundred
Anguilla	0	6.62%	0.00%	1,020,136	193
Bermuda	0	8.94%	44.57%	5,666,843	549
Cayman Islands	0	11.08%	0.00%	7,294,920	1,407
Faeroe Islands	0	7.39%	0.00%	3,762,605	444
Falkland Islands	0	5.81%	0.00%	208,980	280
Gibraltar	0	9.03%	0.00%	3,200,645	8,348
Greenland	0	6.74%	55.43%	4,484,988	101
Guernsey	0	0.00%	0.00%	4,961,848	274
Hong Kong	0	12.62%	0.00%	937,123,500	249
Isle of Man	0	8.23%	0.00%	7,227,720	2,235
Jersey	0	0.00%	0.00%	7,883,694	740
Macao	0	6.51%	0.00%	42,207,230	280
Montserrat	0	4.59%	0.00%	259,948	89
Northern Cyprus	0	0.00%	0.00%	0	57
Saint Helena	0	4.89%	0.00%	309,621	167
Turks and Caicos Islands	0	7.54%	0.00%	3,058,722	63

(Source: OWID, data to 24 Apr fig 4.0)

220 Countries and Territories around the world have reported **a total of 147,886,536** confirmed cases of the coronavirus COVID-19 that originated from Wuhan, China, and a **death toll of 3,124,726 deaths**, according to **Worldometer.**

But some countries have staved off the coronavirus pandemic since it began in early 2020, reporting zero cases as of March, 2021, according to the World Health Organization.

Twelve of the 14 countries and territories reporting no cases are islands in the Pacific and Atlantic oceans, and are likely experiencing the benefit of bordering only the sea. But their strict travel policies may also be responsible for the islands' success, as other island nations and territories have not fared as well.

Although these countries are reporting zero cases that do not rule out that some may fly under the radar.

As **The Associated Press notes**, two countries in particular, **North Korea and Turkmenistan**, have drawn skepticism from the international community about the accuracy of their health data.

Conclusion

This **study investigates the applicability of different data visualizations to analyze COVID-19 data** for world as well as India during lockdown and unlock. There have never been so many line charts, bar charts and choropleth maps occupying the news, as simple data visualisations have become key to communicating vital information about the coronavirus pandemic to the public.

At this time of massive global crisis, it feels almost trivial to write about visual representations of data, but they play a significant role, and the public's ability to make sense of them has never been more important.

For governments and researchers looking to communicate public health information, finding out how simple data visualizations influence the public is now more pressing than ever.

But **Visualization is merely a process**. What analyst actually does when they make a good chart is get at some truth and move people to feel it—to see what couldn't be seen before. To change minds. To cause action.

Future Enhancement

The future extension to this work is **to forecast the effect of pandemic in World as well as in India using Forecasting Model such as ARIMA model** on the COVID-19 data would allow Governments to alter their policies accordingly and plan ahead.

Analyzing the outcome of condition should be the need of the hour research domain. In future work, will likely to explore various time series models for the prediction of positive cases.

Auto-Regressive Integrated Moving Average (ARIMA) model will going to employ for analyzing the unlocking effect and predicting the incidence of 2019-nCov disease.

The main future activity is **to deploy machine learning model into web application** with frameworks and tooling, software which can take in an input and return an output that can be used in making practical business decisions.

References:

Coronavirus Source Data - Our World in Data
https://link.springer.com/article/10.1007/s40031-021-00585-7
https://www.cs.uic.edu/~kzhao/Papers/00_course_Data_visualization.pdf
https://www.sciencedirect.com/science/article/pii/S2352340920310684
https://ieeexplore.ieee.org/abstract/document/9298390
https://www.tableau.com/learn/articles/time-series-analysis
https://www.worldbank.org/en/news/press-release/2020/10/07/covid-19-to-add-as-many-as-150-million-extreme-poor-by-2021
https://royalsocietypublishing.org/doi/10.1098/rsif.2020.0982
https://epiforecasts.io/covid/posts/global/
https://www.uneca.org/
https://www.worldometers.info/coronavirus/

Appendix

(Program Code for Data Cleaning written in Python Jupyter Notebook)

Enclosed herewith: Python Program
https://github.com/ManjiriSDS/COVID---19/blob/main/Clean%20data%20of%20covid%20-%2019.ipynb

Enclosed herewith: Tableau Dashboard
Time Series Analysis of COVID - 19 - Manjiri Sawant | Tableau Public