Analyzing the Impact of Demographic and Economic Factors on COVID-19 Outcomes:

A Comparative Analysis of USA and Global Trends in Cases and Deaths

Esl Kim, Bora Kayak, Tomomi Setowaga, Max Li

UCLA Extension - COM SCI X 450.2 (Winter 2023)
Exploratory Data Analysis and Visualization (March 16th, 2023)

Table of Contents

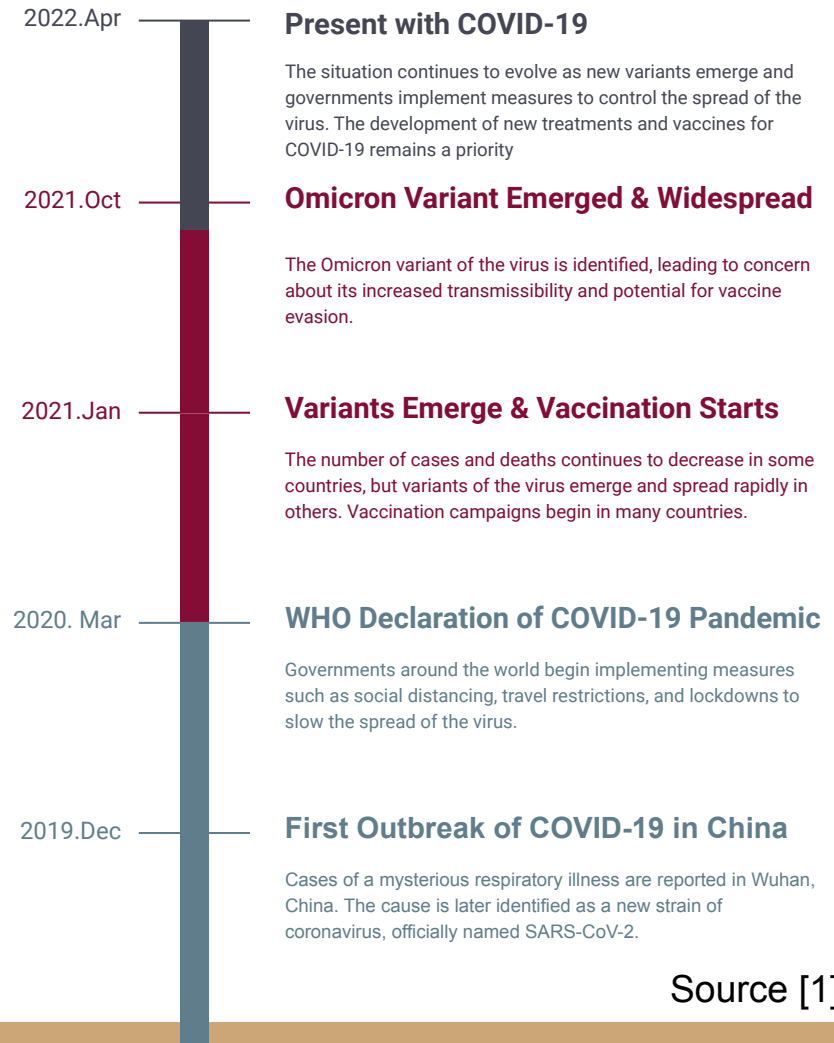
I. Introduction
II. Research Question
III. Data Set and Source
IV. Data Cleaning
V. Summary Statistics
VI. Results and Findings
VII. Conclusions
VIII. Future steps
IX. References

I. Introduction

Coronavirus disease (COVID-19): an infectious disease caused by the SARS-CoV-2 virus

Although most people infected with the virus will experience mild to moderate respiratory illness and recover without requiring special treatment, Older people and those with underlying medical conditions like cardiovascular disease, diabetes, chronic respiratory disease, or cancer are more likely to become seriously ill and require medical attention.

COVID-19 Cases	681,965,905
COVID-19 Deaths	6,814,875
Unemployment (2020)	144 Millions



Source [1][2]

II. Research Questions

Questions: How do demographic and economic factors impact on COVID-19 outcomes or what is the relationship between demographic/economic factors and COVID-19 outcomes

Purpose: To examine the relationship between demographic/economic factors such as age, gender, income, health status and GDP per capita and COVID-19 outcomes including cases, deaths, vaccination and reproduction rates in the USA and worldwide. By examining these factors, the project aims to identify patterns and correlations that can inform public health policies and interventions to mitigate the impact of the COVID-19 pandemic.

III. Data Set and Source

1st Data Set: COVID-19 in major continents and countries (except Antarctica) from the outbreak of disease to the Feb-27, 2023

Data set comes from: Kaggle (https://www.kaggle.com/datasets/sandhyakrishnan02/latest-covid-19-dataset-worldwide)

Data Source: Based on the Nature Scientific Data article, Hasell, J., Mathieu, E., Beltekian, D. et al. A cross-country database of COVID-19 testing. Sci Data 7, 345 (2020) and which continues to be updated, can be downloaded from a public GitHub repository (https://github.com/owid/covid-19-data/tree/master/public/data/testing)

2nd Data Set: total confirmed cases of COVID-19 per million people depending on their location and their GDP per capita in 2017

Data set comes from: Our World in Data (https://ourworldindata.org/coronavirus)

Data Source: World Health Organization (WHO), which updates its dataset weekly (up to 2023-02-28)

III. Data Set and Source

3rd Data Set: COVID-19 related illnesses and deaths with their certain group and also provides deaths by Pneumonia and Influenza within the United States

Data set comes from: Center for Disease Control and Prevention (CDC) (https://data.cdc.gov/NCHS/Provisional-COVID-19-Deaths-by-Sex-and-Age/9bhg-hcku)

Data Source: National Center for Health Statistics (NCHS)

4th Data Set: Average Temperature of States in the USA by Date

Data set comes from: National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental information, Climate at a Glance: Statewide Time Series (https://www.ncei.noaa.gov/access/monitoring/climate-at-a-glance/statewide/time-series)

Data Source: U.S. Climate Divisional Database, which have data from 1895 to the present

IV. Data Cleaning

1st Data Set: owid-covid-data.csv

- Filtering out unnecessary column and rows
- year: Convert date in form of MM/DD/YYYY to year
- season: Create a season column using imported library "hydroTSM" and its function time2season
- Replace NA to 0 as all NA columns was data point per each day

2nd Data Set: total-confirmed-cases-of-covid-19-permillion-people-vs-gdp-per-capita.csv

- Entity: Change Column Name into Location
- Filtering out unnecessary column and rows
- Code: Change name into iso_code
- year: Change name into Year and 2020, 2021, 2022, 2023
- Seasons: Create a Seasons column using “Day” (character) column and group it by Seasons
- Dec - Feb : Winter, March - May: Spring, Jun - Aug: Summer, Sep-Nov: Autumn
- GDP per Capita: Gross domestic product at purchasing power parity (constant 2011 international dollars), calculate GDP mean of the allocated time period

IV. Data Cleaning

3rd Data Set: Provisional_COVID-19_Deaths_by_Sex_and_Age.csv

- Remove Footnote column
- Put all NAs into 0 and do not remove any values

4th Data Set: 2020-2023_Temp_state.csv

- No NAs or empty string for all 3 variables
- Merged with the 3rd after selecting States within the USA mainland + Alaska

V. Summary Statistics - 1st Data Set
COVID-19 Outcomes, Health Risks and GDP per capita (2021) Worldwide

Dataset Features	Numbers
Observations	260,567
Total Variables	67
Character + Date Variables	5
Numerical Variables	62

Important Variables from the 1st data set:

- continent: Name of the continent to which the country belongs
- location: Name of the country
- season: Season of every entry
- total_cases: Number of allocated positive cases
- total_deaths: Total number of COVID-19 deaths
- total_vaccinatons: Total number of vaccination
- reproduction_rate: measurement of the transmissibility of infectious agents (COVID-19)
- GDP per capita : GDP per capita (2021)

V. Summary Statistics - 1st Data Set
COVID-19 Outcomes, Health Risks and GDP per capita (2021) Worldwide

Dataset Features	Numbers
Observations	260,567
Total Variables	67
Character + Date Variables	5
Numerical Variables	62

Dataset Features	Numbers
Observations	260,567
Total Variables	38
Character + Date Variables	5
Numerical Variables	33

V. Summary Statistics: 2nd Data Set
COVID-19 Cases with GDP per Capita (2017) Worldwide

Dataset Features	Numbers
Observations	246,505
Total Variables	8
Character Variables	4
Numerical Variables	4

Important variables from the 2nd dataset:

- Entity: Name of the country
- Year: Year of data entry
- Day: Date of every entry
- Total confirmed cases of COVID-19 per million people
- Continent
- GDP: GDP per capita, PPP (constant 2017 international $)

V. Summary Statistics: 2nd Data Set

COVID-19 Cases with GDP per Capita (2017) Worldwide

Dataset Features	Numbers
Observations	246,505
Total Variables	8
Character Variables	4
Numerical Variables	4

Dataset Features	Numbers
Observations	3,105
Total Variables	6
Character Variables	4
Numerical Variables	2

V. Summary Statistics: 3rd Data Set

COVID-19 Deaths with other diseases by Demographic factors in the USA

Dataset Features	Numbers
Observations	118,422
Total Variables	15
Character Variables	7
Numerical Variables	8

Important variables from the 3rd dataset:

- Sex: Gender of the person.
- State: State which this event happened.
- Age Group: Selection a group of people around the same age.
- COVID-19 Deaths : People who died from Covid-19.

V. Summary Statistics: 3rd Data Set

COVID-19 Deaths with other diseases by Demographic factors in the USA

Dataset Features	Numbers
Observations	118,422
Total Variables	16
Character Variables	8
Numerical Variables	8

Dataset Features	Numbers
Observations	83,592
Total Variables	17
Character Variables	7
Numerical Variables	10

V. Summary Statistics: 4th Data Set
Average Temperature of States in the USA by Date

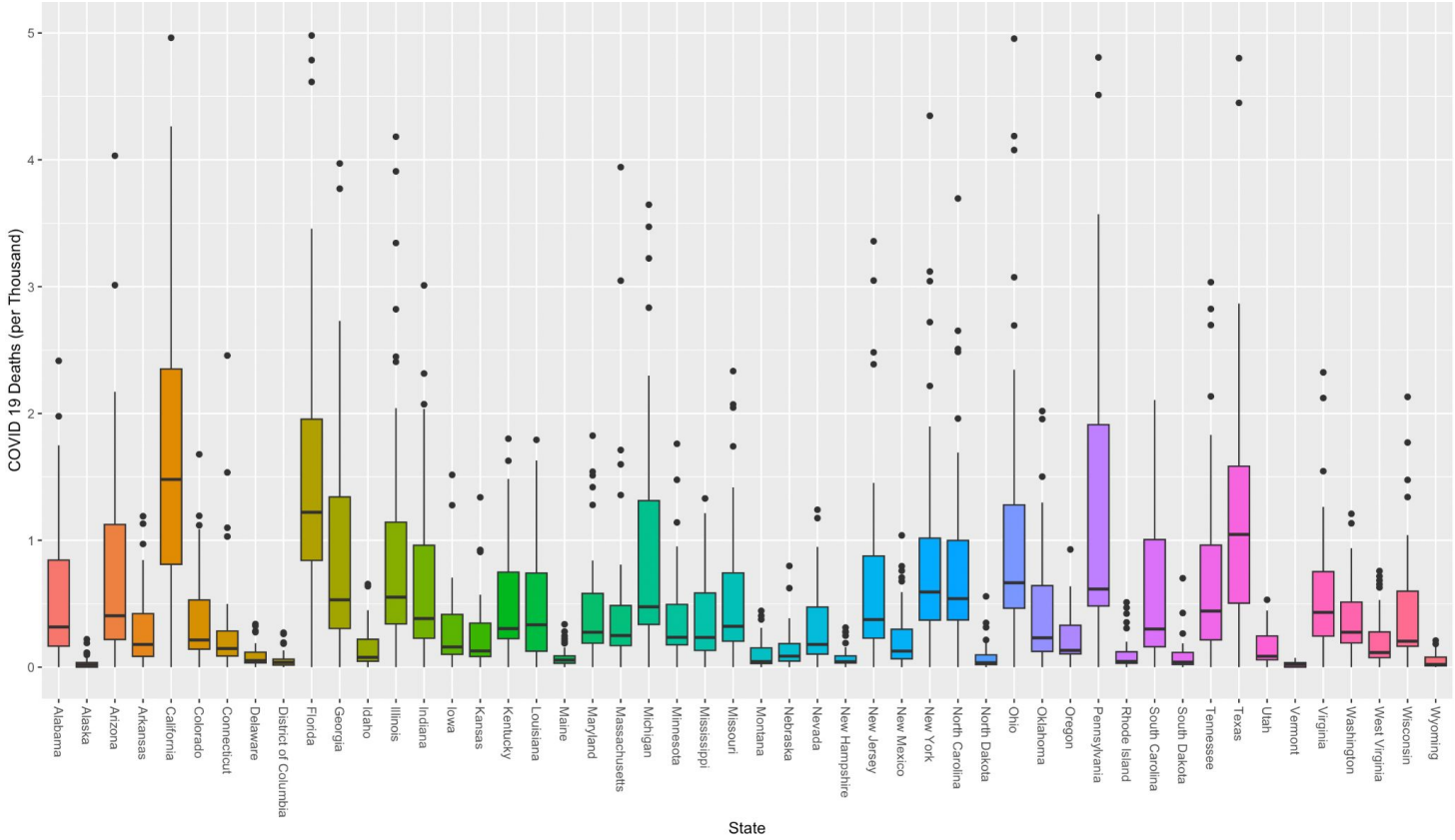
Dataset Features	Numbers
Observations	1850
Total Variables	3
Character Variables	1
Numerical Variables	2

Important Variables from the 4th dataset:

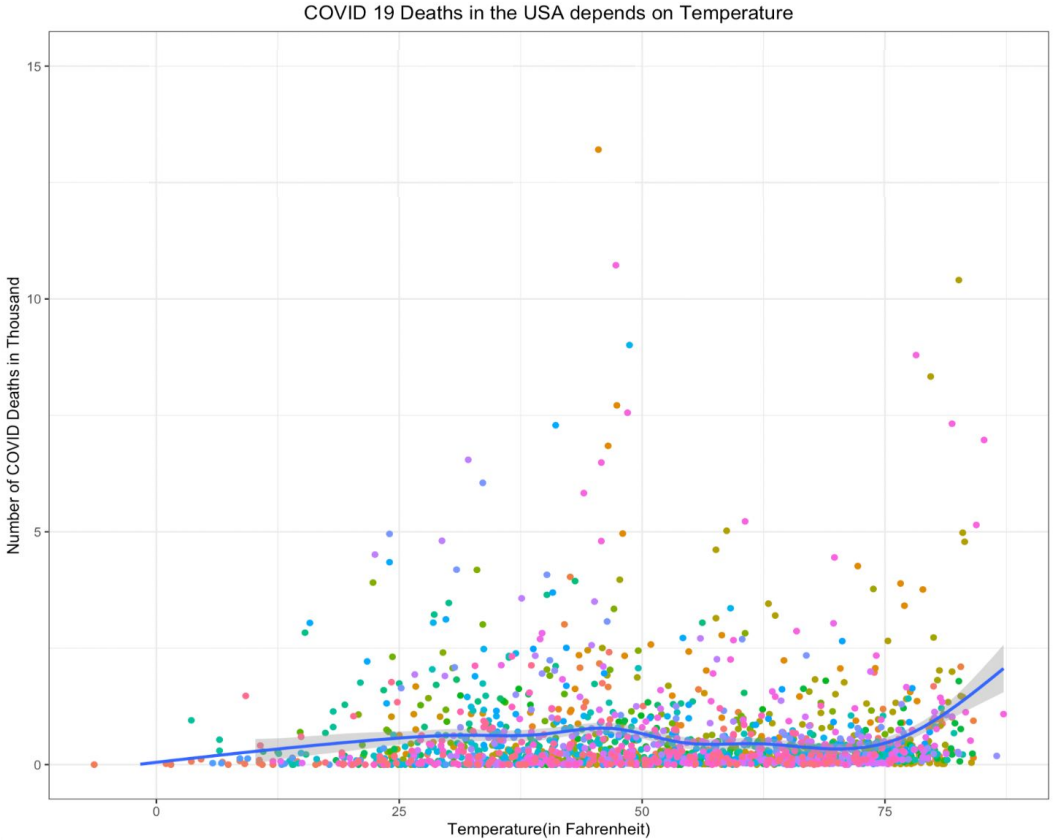
- Date : integer variable with format of YYMM (e.g. 2001 for 2020.Jan) from 2020. January to 2023. January
- Avg. Temp. F. : Average Temperature of a State per Month
- StateName: Name of the States in the USA

There was no N/A used with 3rd Data Set by merging them

VI. Results and Findings:
Number of Deaths due to COVID-19 by States in the USA including Alaska

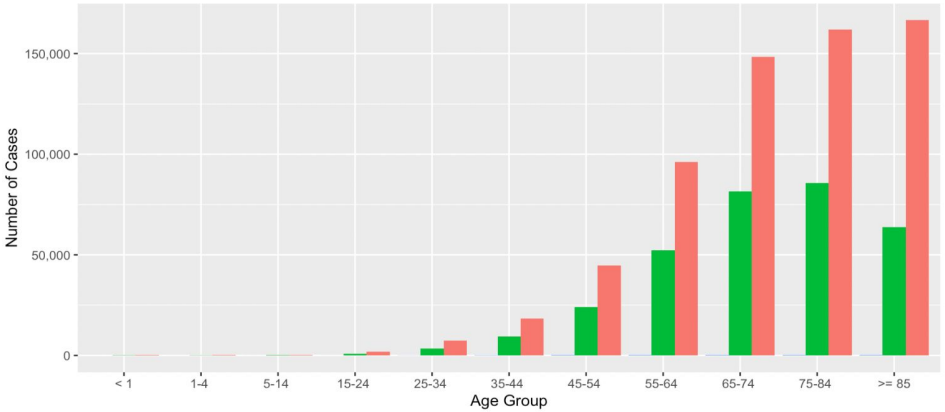


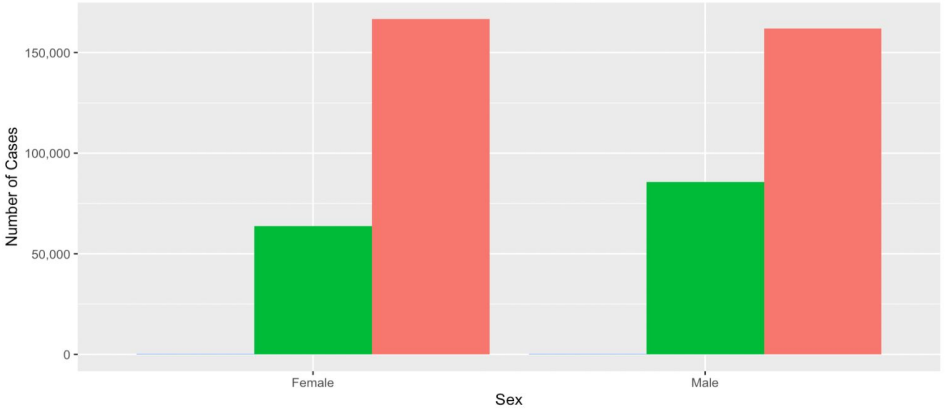
VI. Results and Findings:
Relationship between COVID-19 Deaths & Temperature in the USA



- According to the WHO: The COVID-19 virus can spread in hot and humid climates
- Assumption: Lower **temperature leads to** lower **movement** among population would have less infection and less deaths
- Results: The graph shows the tendency of increasing the number of deaths as the temperature increases. However, not highly correlated or significant relationship between temperature and deaths

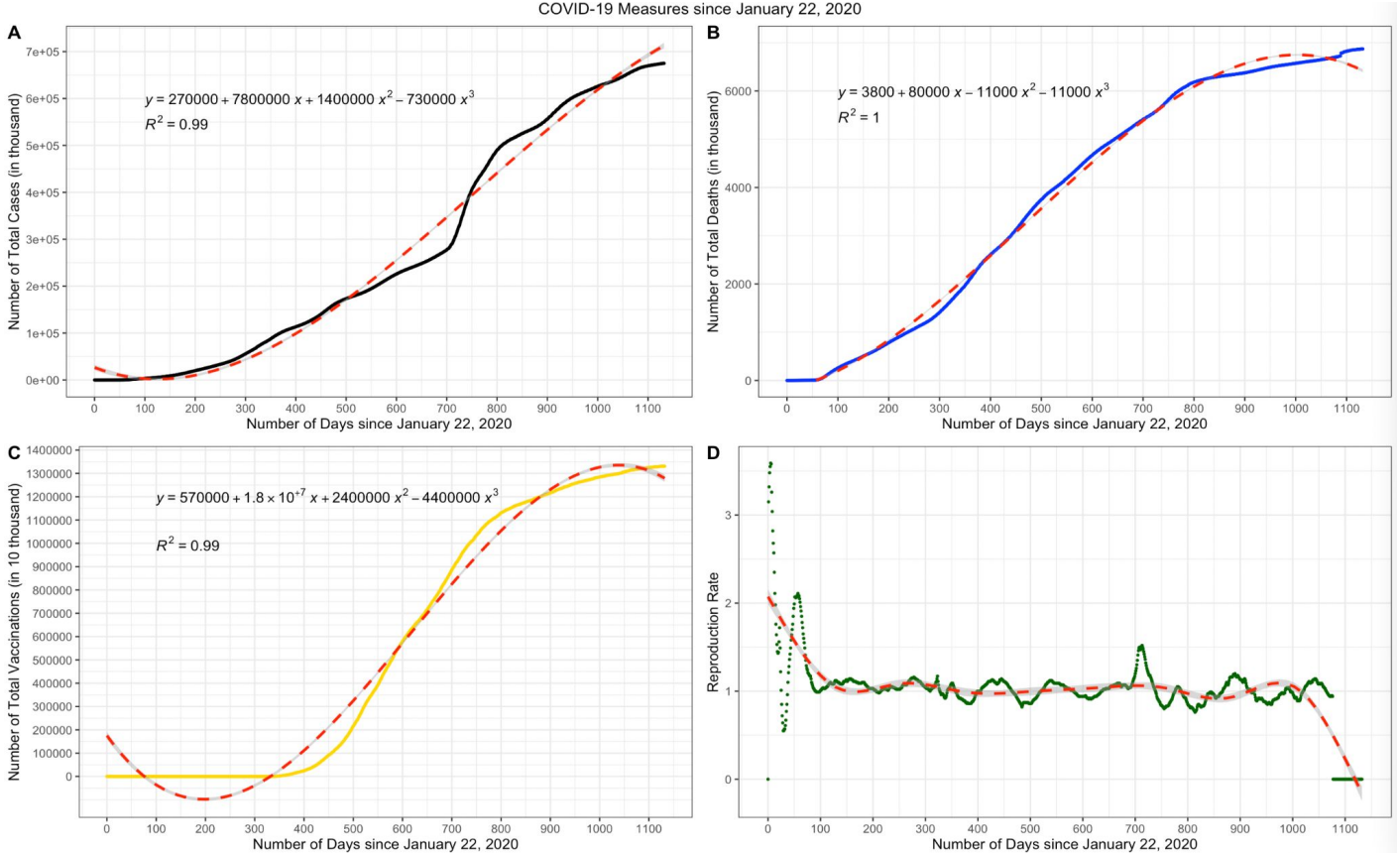
VI. Results and Findings:
COVID-19 Deaths with different diseases in the USA





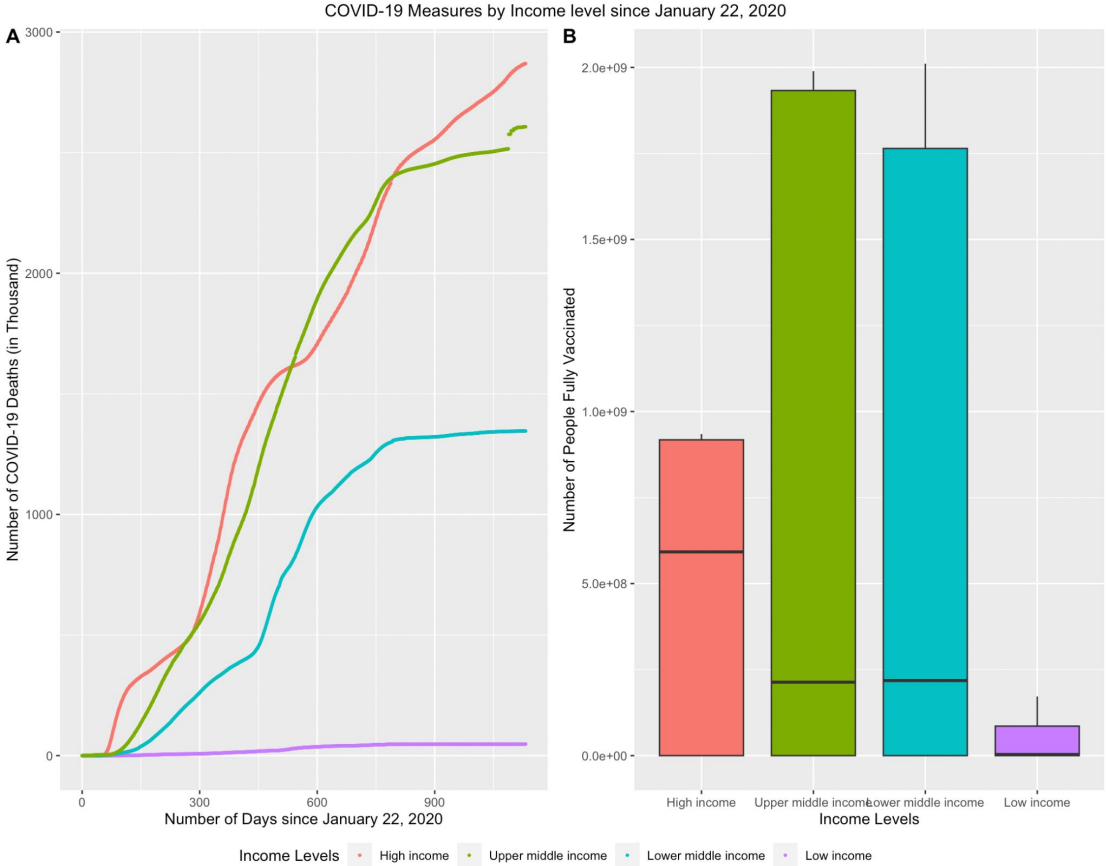
- Deaths caused by COVID-19 was outnumbered than other deaths caused by 2 other diseases
- Having Influenza(flu) and COVID-19 at the same time caused the least deaths in the USA among 3 groups
- As the number of deaths with having Pneumonia and COVID-19 was greater than that with Influenza and COVID-19, it shows that the function of lung is critical to the people who go through COVID-19
- As Ages getting older, higher number of deaths in COVID-19 and other disease
- More Female died due to COVID-19 only while more Male died due to both Pneumonia and COVID-19

VI. Results and Findings: Worldwide COVID-19 Outcomes



- COVID-19 Cases and Deaths shows **Logarithmic Growth** as Reproduction rate was very high at the beginning of the COVID-19 outbreak but decreased and stabilized after WHO declares the Pandemic
- COVID-19 Vaccination rate also shows **Logarithmic Growth**

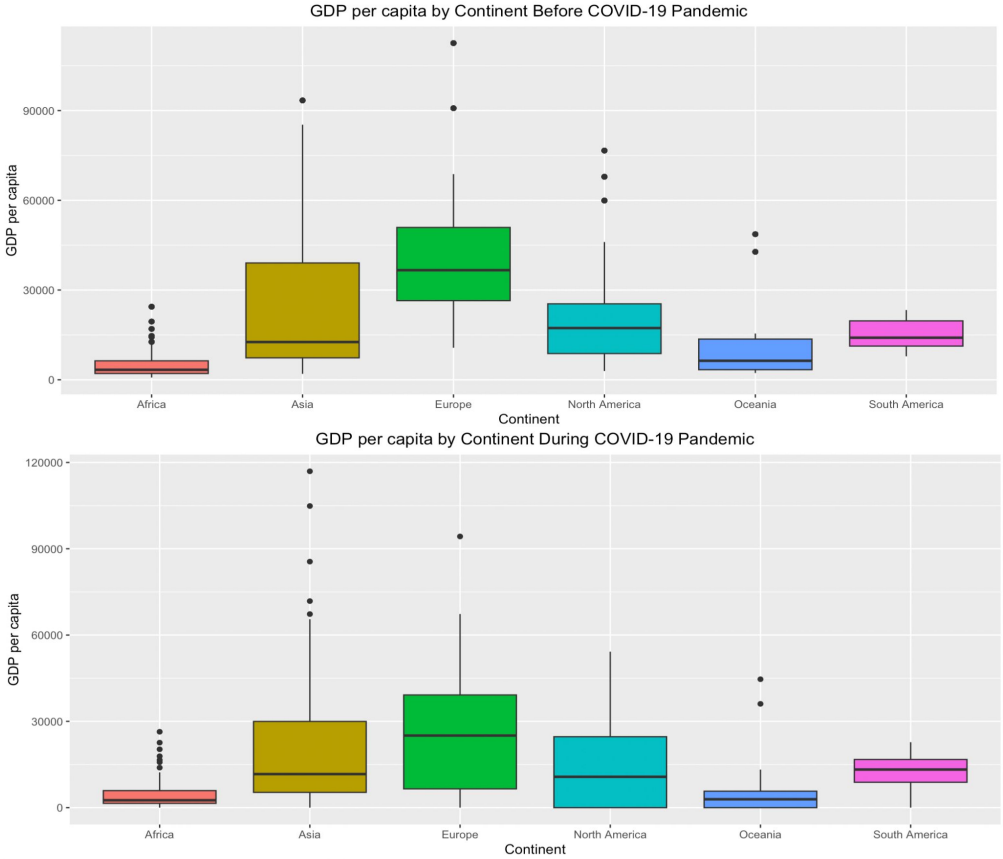
VI. Results and Findings:
Worldwide Income Level related to COVID-19 Outcomes



- Graph A shows that Number of COVID-19 deaths among Income level. Higher the income is higher number of COVID-19 Deaths

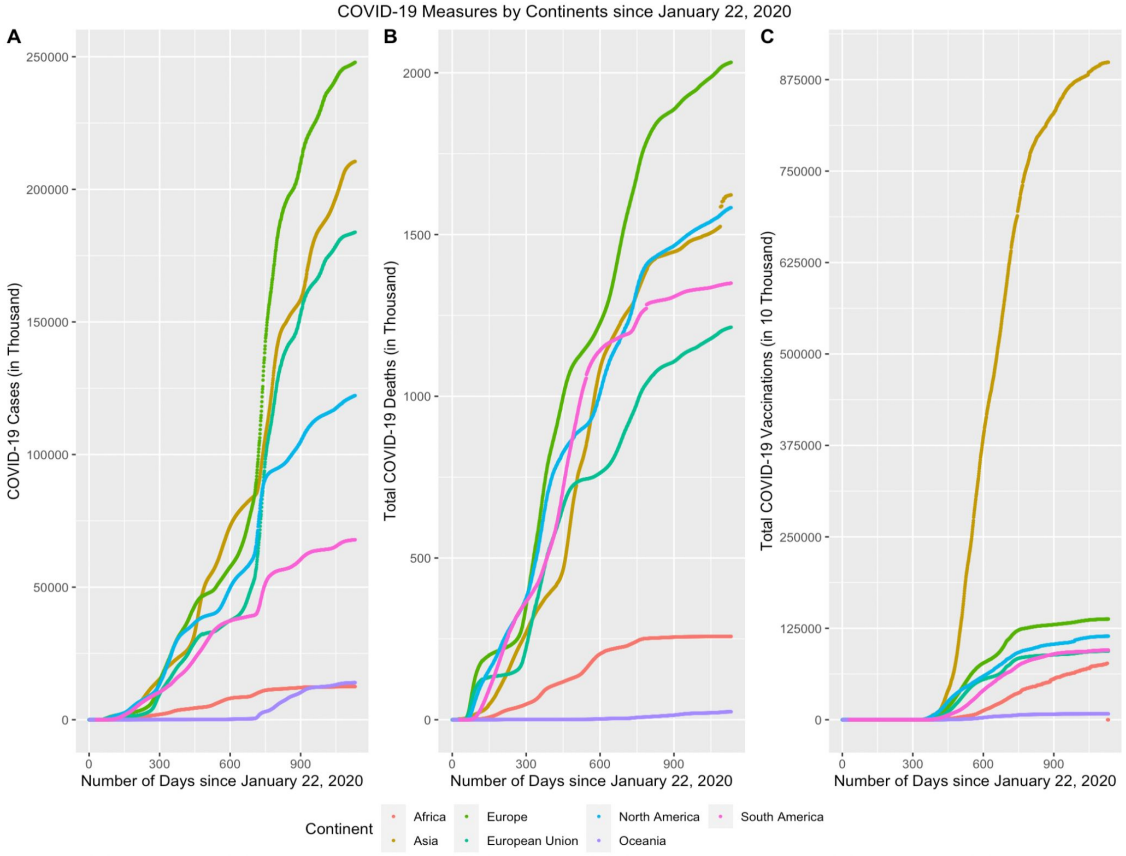
- Graph B shows the number of people fully vaccinated depends on the income level. Upper middle income has the highest number of people full vaccinated

V. Data Analysis and Visualization:
Average GDP per capita by Continents and Time



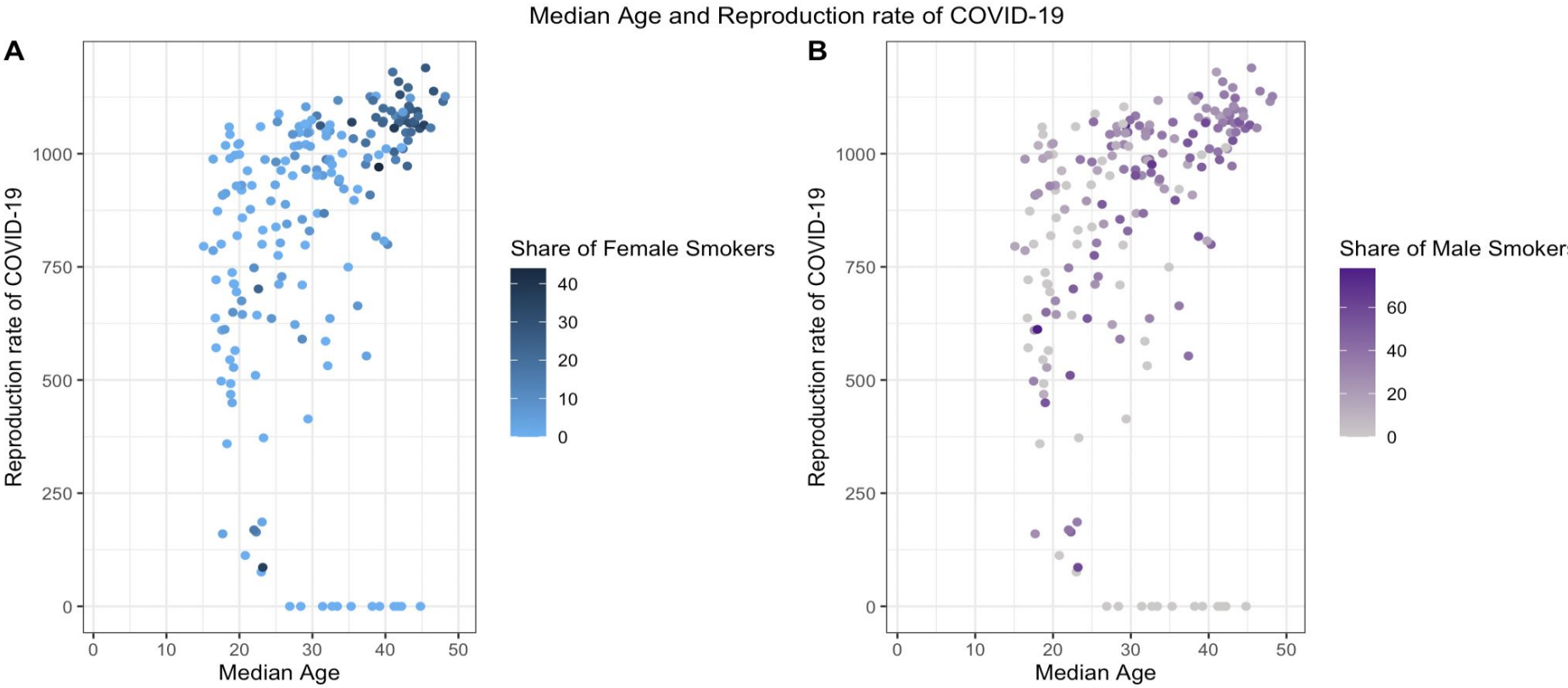
- Top Graph: GDP per capita before COVID-19 Pandemic (2017) by Continent
- Bottom Graph: GDP per capita during COVID-19 Pandemic (2021) by Continent
- Overall, **median** of GDP per capita was decreased during the pandemic, represented that there was global economic shrink due to the COVID-19 with restrictions on economic activities and failures in income support and consumer spending

V. Data Analysis and Visualization:
COVID-19 Measurements by Continents since Jan 22, 2020

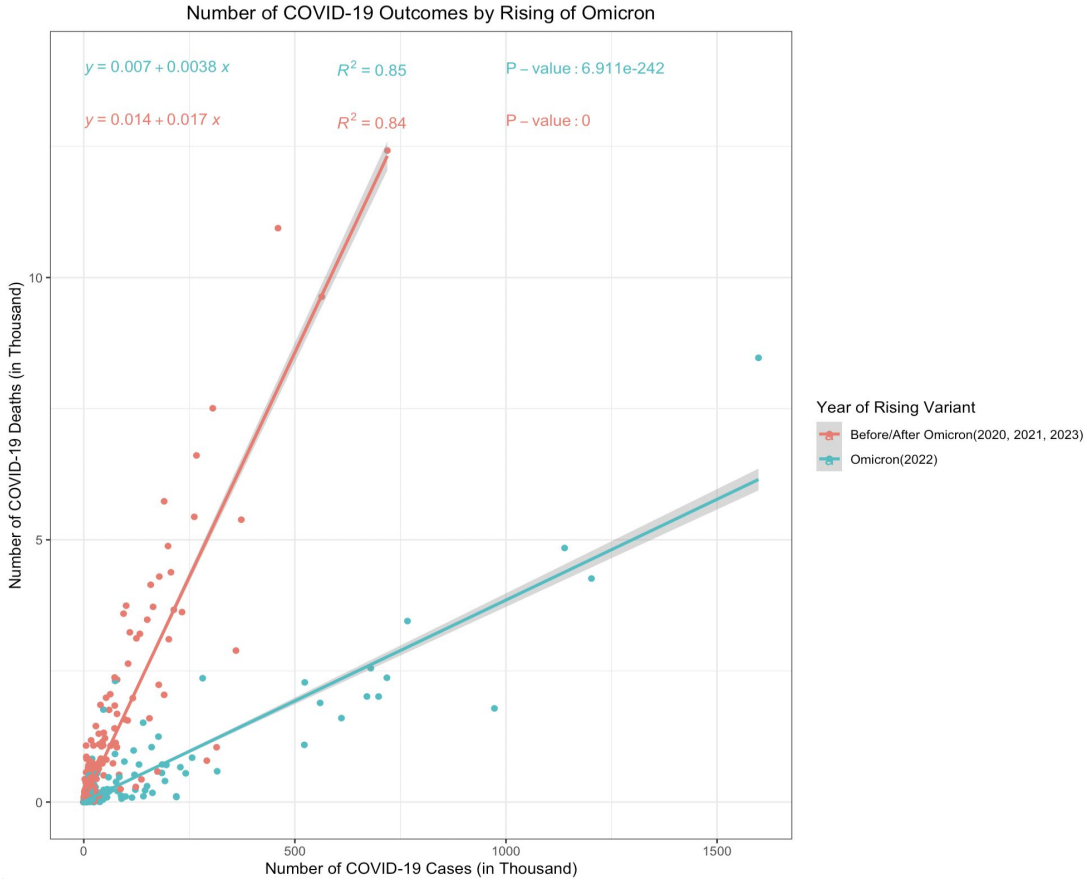


- **Graph A, B, and C** show the Number of COVID-19 Cases, Deaths and Vaccinations, respectfully since Jan 22, 2020
- **Graph A and B** show Europe (green) had highest number of COVID-19 cases and deaths
- While Africa (pink) has low COVID-19 cases, it has high COVID-19 deaths comparing to North America (blue)
- **Graph C** shows Asia had the most high number of Vaccinations comparing to the other continents. It shows that number of people fully vaccinated including the boosters in Asia is the higher than the other continents as the population of Asia is the largest among the continents

V. Data Analysis and Visualization:
Reproduction rate of COVID-19 by Median Age with the Share of Smokers

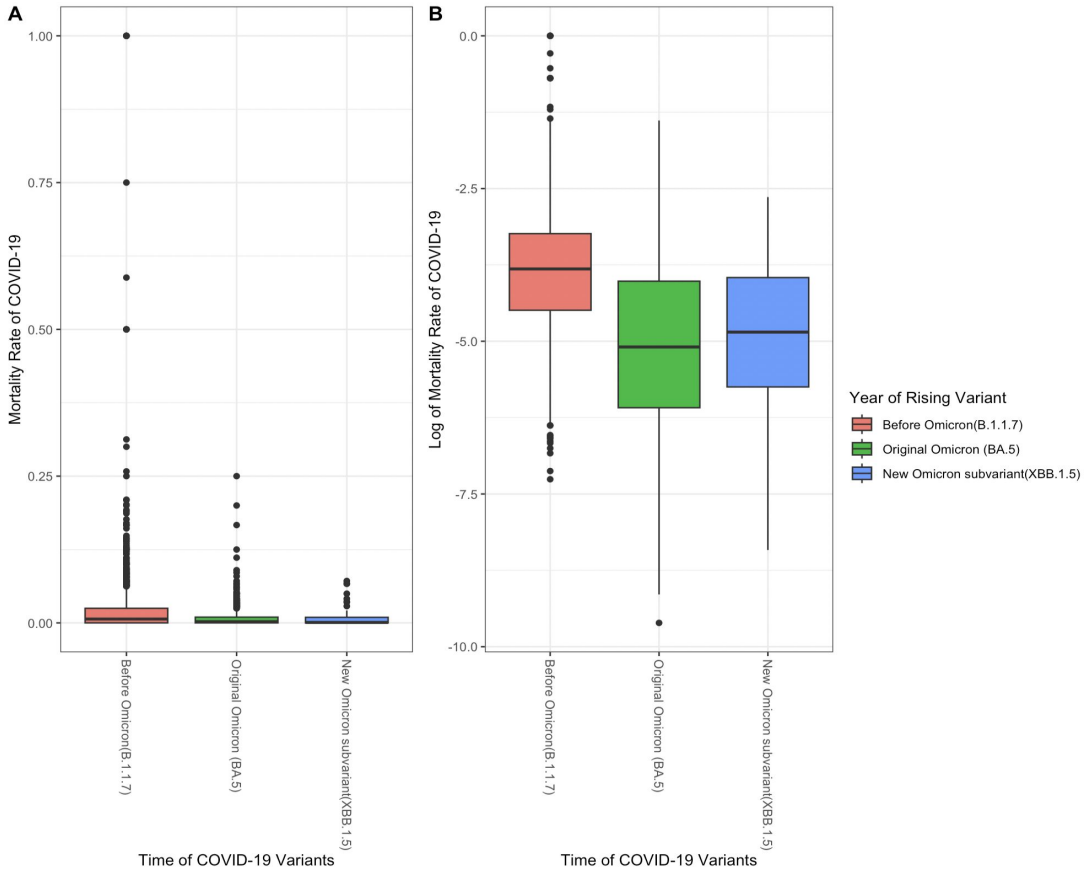


V. Data Analysis and Visualization:
Number of COVID-19 Outcomes by Rising of Omicron



- Omicron was emerged in late 2021 and widely spread over the world in 2022
- The mortality rate of COVID-19 (slope) during the time Before Omicron was emerged and in 2023 was higher than that of in 2022 when the Omicron is widely spread.
- Shows that Omicron spread more but less deaths comparing to other variants.

V. Data Analysis and Visualization:
Number of COVID-19 Outcomes by Rising of Omicron



- **Graph A and B** show the mortality rate of COVID-19 by Year of major rising variant.
- **Graph B** shows the Log of the mortality rate to easily visualize the mortality rate by the time.
- Before Omicron is emerged, mortality rate was the highest among other Omicron related variants.

VII. Conclusions

Through the COVID-19 outcomes in the USA and Worldwide relating to the Demographic and Economic factors:

- Temperature and income-level were not highly correlated with the deaths due to the COVID-19.
- The median of GDP per capita was decreased during the pandemic, indicating a global economic shrink due to restrictions on economic activities and failures in income support and consumer spending
- Health status (such as having pneumonia or smoking) are more closely related and impacted on the deaths due to COVID-19 and the reproduction rate of the disease.
- Rise of the Omicron impacted the mortality rate of COVID-19 by decreasing it significantly.

VIII. Limitations (Technical Issues between OS)

The “Day” column that shows date for that specific day is a “chr” type and also formatted in a different style than what we need

```
$ Day                                          : chr  "24/02/2020" "25/02/2020" "26/02/2020" "27/0
#changing the chr type date to a regular date format
mainDf$Day <- strptime(as.character(mainDf$Day), "%d/%m/%Y")

#changing the regular date format to a POSIXct in order to use it later
#(You cant change it directly into a POSIX format thats why we change it to a date format first)
mainDf$Day <- as.POSIXct(mainDf$Day, format = "%Y-%m-%d %H:%M:%S")
```

During formatting the date in the second dataset we realized when R studio opens it comes with different settings in Mac than Windows. In windows format needed to be changed 2 times in order to use it later in the code but in Mac when the document was first loaded it already comes pre formatted and ready to use.

VIII. Limitations and Future Steps

- During the data analysis and visualization, Worldwide Low income countries or Low GDP per capita countries data were not fully updated or had limited informations within the countries such as North Korea. Thus, the datasets could not fully represented COVID-19 Cases, Deaths, and Vaccinations for those countries while High income countries or High-Mid GDP per capita countries contains very specific data points.
- Some health risks such as diabetes prevalence, cardiovascular disease have same values over continents.
- For the Temperature and COVID-19 cases relationship, if humidity and longitude and latitude information were provided, the relationship between climates and COVID-19 cases may show stronger correlation as survival of viruses, including human coronaviruses, is reduced when the relative humidity is in the 40–60% range

VIII. Limitations and Future Steps

- Although many COVID-19 restrictions and measures are now released, there are still people who suffer under long COVID or Post-COVID symptoms. If we can get better demographic informations, health risks including cancer and respiratory diseases and types of variants of COVID-19, we may characterize the variants of COVID-19 and people who under going the post-COVID symptoms to develop vaccines and treatment.
- If there is a datasets for COVID-19 restrictions for worldwide (there are by country in their languages), we may can predict the effective restrictions and measurements to prevent the spread of disease in the future and also investigate on differences between countries which implemented lockdown policies and others did not.
- We can also visualize the COVID impact on various industries in economics way. If there is a dataset of profit of industries for each country, we could have more analysis on industries benefit from COVID (e-learning, entertainment) and which suffer the most (food services).

IX. References

[1] CDC Museum COVID-19 Timeline. David J. Sencer CDC Museum: In Association with the Smithsonian Institution. Retrieved from: https://www.cdc.gov/museum/timeline/covid19.html
[2] Archived: WHO Timeline - COVID-19. Retrieved from: https://www.who.int/news/item/27-04-2020-who-timeline---covid-19
[3] Edouard Mathieu, Hannah Ritchie, Lucas Rodés-Guirao, Cameron Appel, Charlie Giattino, Joe Hasell, Bobbie Macdonald, Saloni Dattani, Diana Beltekian, Esteban Ortiz-Ospina and Max Roser (2020) - "Coronavirus Pandemic (COVID-19)". Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/coronavirus [Online Resource]
[4] Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real-time. Lancet Inf Dis. 20(5):533-534. DOI: 10.1016/S1473-3099(20)30120-1
[5] National Center for Health Statistics. Provisional COVID-19 Deaths by Sex and Age. Date accessed [February 23, 2023]. Available from https://data.cdc.gov/d/9bhg-hcku.
[6] NOAA National Centers for Environmental information, Climate at a Glance: Statewide Time Series, published February 2023, retrieved on March 3, 2023 from https://www.ncei.noaa.gov/access/monitoring/climate-at-a-glance/statewide/time-series
[7] New World Bank country classifications by income level: 2022-2023. Date accessed [March 1, 2023]. Available from https://blogs.worldbank.org/opendata/new-world-bank-country-classifications-income-level-2022-2023