

Project Based Exam (7CCSMSDV)

(Student ID: 1919414)

Background: The 2019 Coronavirus (COVID-19) has become a global epidemic and has brought unprecedented challenges to the international community. We need to understand the state of the disease and coordinate each nation's policies for future trajectories which rely heavily on data on the spread (Q1) , mortality (Q2), recovery cases (Q3).

Part 1. Analytics

a. Research questions.

Q1: Analyse the spread trend of this virus all over the world. What is the spread over time?

Q2: It is crucial to know how the death rate for COVID-19 change and if there are any other factors can contribute more to it. The second question concerns the death cases and rate grouped by age group and gender to figure out whether these factors have impact on it and how much do they matter.

Q3: There is currently no treatment for this virus. Recovery rate is really hard to predict because we have many unknown and inadequate information about COVID-19. Therefore, the third question mainly concerns how many people around the world have recovered over the time. Later on, comparing it with confirmed cases and death cases over the period and observe the gap among them.

b. Explain type of data. & c. Explain if and how datasets, described in b.

Concerning the first question, we can also use data[1] containing the latitude and longitude which can generate 2D area type of data visualization. From the graph, it is clear and straightforward to compare relative position of countries in virus spread. What's more, data[2] like daily increase rate of confirmed cases can show how the rapidly outbreak compared to other countries. It is more intuitive to show things would get worse or better over the time than original confirmed cases. To be more specific, the chart will allow you to compare how rapidly the number of confirmed cases increased after the outbreak reached a similar stage in each country. High growth rate can be evident to increase the numbers in a short time.

As the coronavirus continues to spread, the number of new infection around the world as well as the number of deaths were reported in the global media. As for the question2, visualizing the data[3] including the number of deaths, the age group and gender at the same time is helpful to figure out how each group spreads COVID-19 and which group may lead to a higher risk of death after infected by virus. If the number of death rate or cases in a specific group is higher than other groups, it means these kind of group at more risk of COVID-19. In this case, we could pay more attention to protect those vulnerable people from risk and support them.

Many people cannot get tested for infection and we have limited data[4] about recovered cases during the period. Given the third question, the number of people recovering from the COVID-19 is not accuracy and may change over time but it does exist. We can compare the recovered cases with confirmed (Q1) and death cases (Q2) to find out the distance between

them. When people get enough tested, the distance may become closer which could suggest a good signal.

[1] <https://datahub.io/core/covid-19>

[2] <https://datahub.io/core/covid-19>

[3] sfortracker:<https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/datasets/deathsinvolvingcovid19englandandwalesmarch2020>

[4] The data is compiled by the Johns Hopkins University Center for Systems Science and Engineering (JHU CCSE) from various sources. <https://github.com/CSSEGISandData/COVID-19>

Part 2. Design and Discussion

Visualization and Design rationale (Figure1):

The first 2D area type of graph can describe how the virus spread over all the world until now. Also, this type of figure can be used to compare the confirmed cases/recovery cases/death cases in a different region but at the same time. It would be better to add a timeline on the graph to show the process of development. Every part has represented a country or region and it will change color to be red from origin white if the situation gets worse. With time going by, the number of cases is cumulative and thus the color will be a darker red than before. It is straightforward to show which part of the world is deeply desperate for now and what is the situation around it. From each part, you can see more details about cases like specific numbers and dates of time, etc.

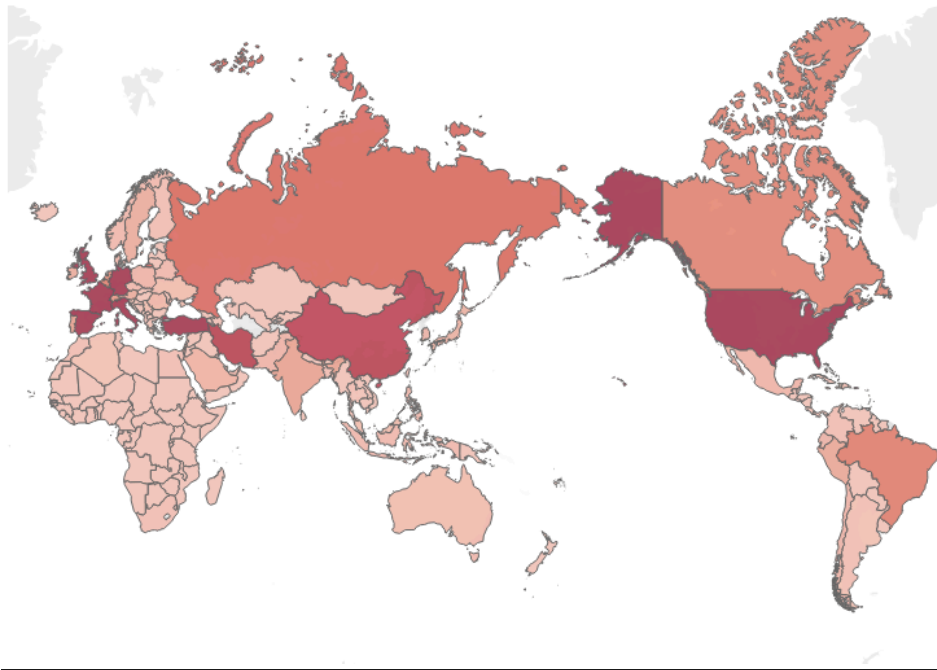


Figure 1

Visualization and Design rationale(Figure2):

The second chart reveals how the age group and sex affect the number of death cases. This can be applied in some questions like how death cases disperse in different age group or sex. If females and males in the same age group with a similar percentage of death, it will be hard to say which one is better. To improve this, put them together in one axis and add some degree will be clear to compare. For each age group, the pink one is the female percentage of death while the blue one is males. The percentage of death goes up, increase each row's width. This graph has shown a range of age groups. Therefore, the new cases will add the corresponding row's width. From the final result, we can get which female age group or male age group will be more vulnerable and give them support and care more.

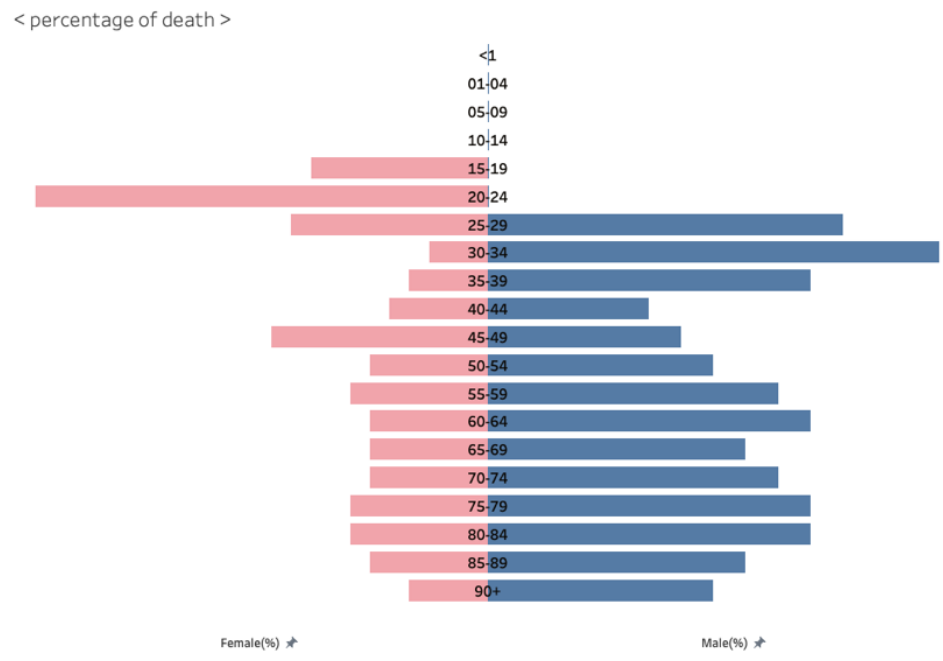
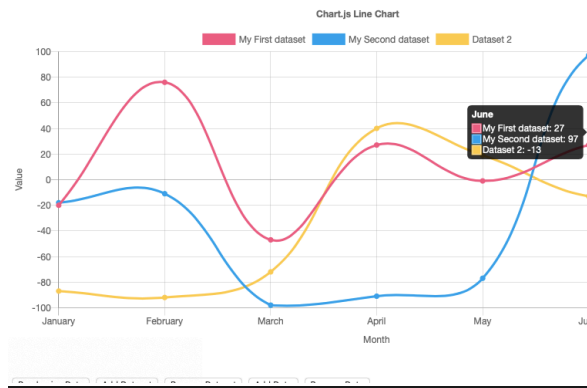


Figure 2

Visualization and Design rationale(Figure3&4):

The third line chart can be used to illustrate the changes in the number of new recoveries, new cases, and new deaths over the period. I will narrow the range of the y-axis to be more completely display the changes of cases. Every time records the number with different categories (recoveries/new/death) in different colors (in this instance, new death is red, new cases is orange, new recoveries is blue). There are gaps among three lines. As time goes by, the lines will show a trend whether there appears new cases, decrease or remain unchanged and gap change with it. If the number goes up, the line will reach a new height. An array of numbers can be turned into a narrative and be expressed in the language which everyone can understand.



Newly Infected vs. Newly Recovered

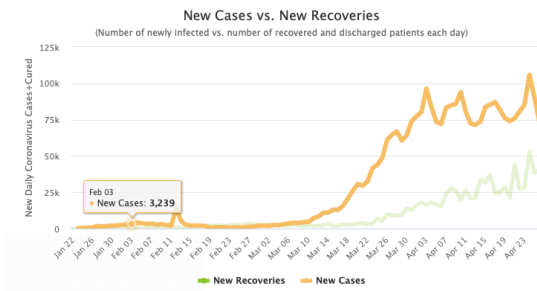


Figure 3&4

Note: There may occur “Cross origin requests are only supported for HTTP” error for Part3.

Please open the html file through IDE like WebStorm or change the COR settings of browser.