COVID 19 : Analysis of climatic conditions and relating factors

Shubham Rathore

Introduction

A new coronavirus designated 2019-nCoV was first identified in Wuhan, the capital of China's Hubei province. People developed pneumonia without a clear cause and for which existing vaccines or treatments were not effective.The virus has shown evidence of human-to-human transmission. Transmission rate (rate of infection) appeared to escalate in mid-January 2020. The world currently is struggling for survival. Lives have disrupted, families have been destroyed, people are left jobless. There is no end to the brutalities done to the humans by this virus. The world economy itself is at stake. Internet and the television are the sole sources of information about the current scenario. Death counts, conspiracy theories, economy collapsing etc. are some of the areas that media is putting light to. But there are some facts that are still undisclosed by the media.

The human brain thinks quite rationally and it is habitual for us to take out correlations amongst the factors visible. The aim of this project is to find correlations and analyze some factors related to the COVID-19. The correlations have been talked by media but again there is no proof for those. Analyzing data over a period of time doesn't necessarily prove a fact/hypothesis but still gives the hypothesis some more confidence. Thus, it becomes quite important to analyze data. Coming to some of the un-answered questions. Does Humidity affect transmission? Is the transmission dependent on the temperature of the area? How does the transmission vary with the population density ? Are the coasts the most affected? How is the air quality index changing with time? First thing that comes to the mind after reading this is that the data needs to be analyzed over a period of time, it isn't a static analysis.

There isn't any dashboard till now that analyses these factors. There exists a confusion among the masses right now related to these factors. Various media sources and publications beg to differ on these attributes. I plan to publicly host the dashboard and personally feel that this would ease the confusion currently pertaining in the minds of the crowd.

Dataset Description

This project requires the concatenation of various datasets as there isn't a dataset that takes into account all the factors that are to be analyzed. The project aims to weekly analyze data over a period of ~3 months. Some attributes are dynamic (people affected, temperatures, humidity etc.) while the other are static (area, coastline etc.).

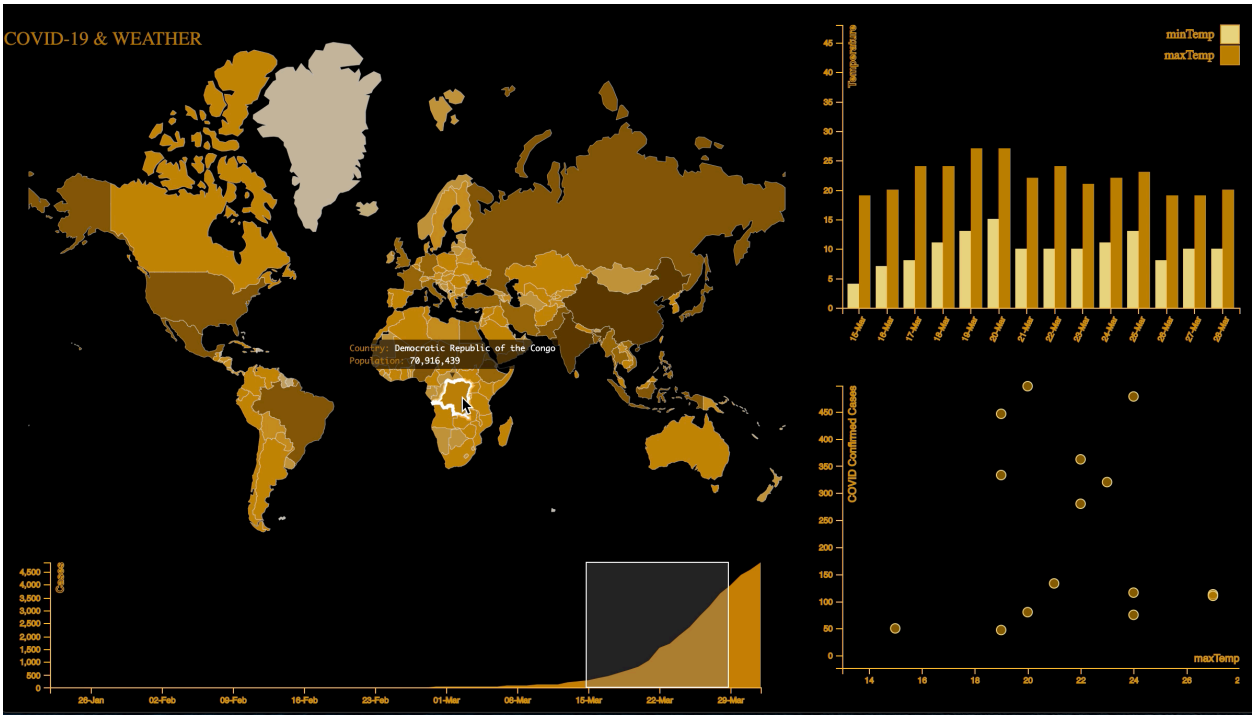
1: Countries of the World: https://www.kaggle.com/fernandol/countries-of-the-world .This dataset provides us with attributes like country size, population, and coastline.

2: Country-wise weather data for covid19: https://www.kaggle.com/ksudhir/weather-data-countries-covid19. This dataset provides us with the attributes like minTemp, max Temp and Humidity of all the countries over a span of 10-12 weeks.

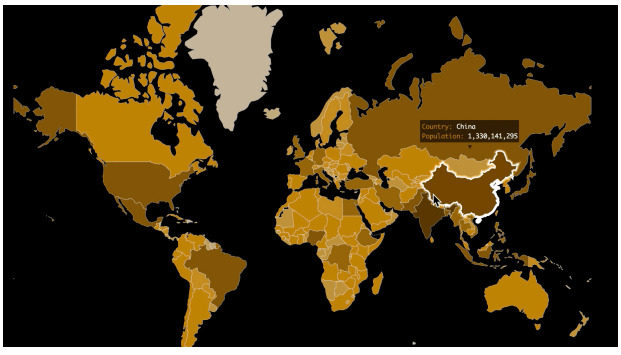
3: COVID-19 Dataset : https://www.kaggle.com/imdevskp/corona-virus-report. This dataset provides us with the demographics; no. of people affected, cured etc. It covers all countries over a span of 10-12 weeks.

Data Processing was a major task in this project. There was a need to account for the missing values, curate data for processing and moreover standardize/ normalize it for operations.

Dashboard

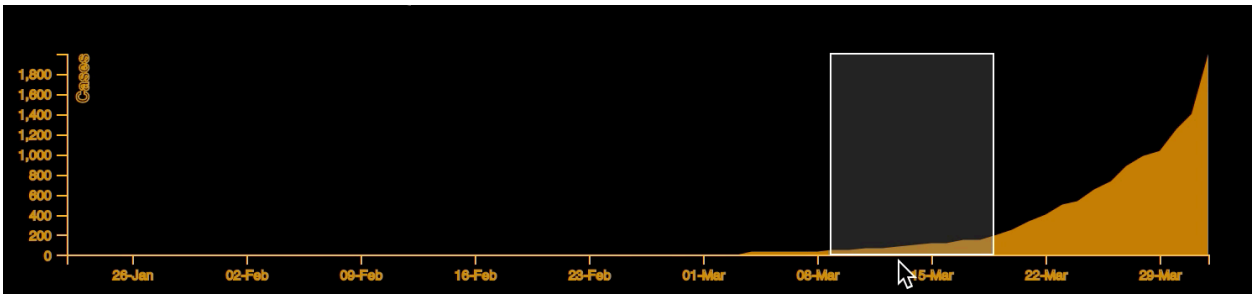


The dashboard is built using d3.js with flask supporting it at the backend. It is divided into 4 major sections which include map chart, area chart, grouped bar chart and scatter plot. The area chart is brushable which in turn is linked to bar chart and the scatterplot. The dashboard layout is mentioned above.

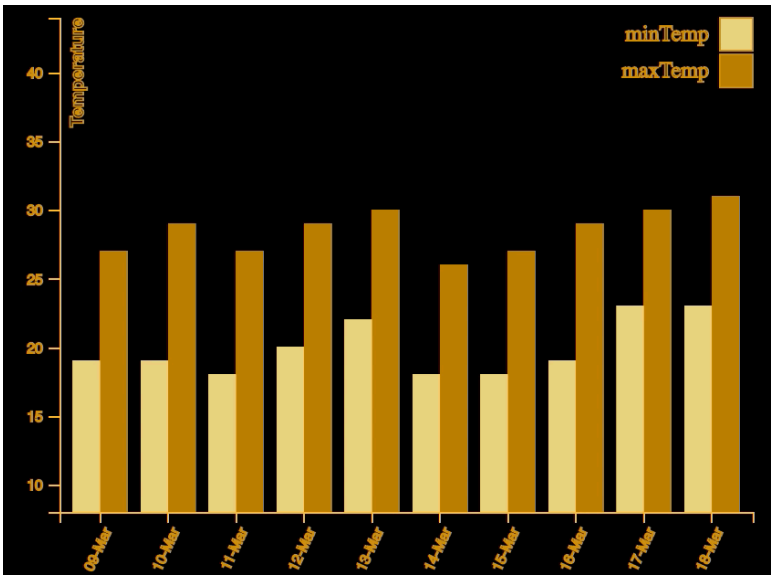


1) Map Chart :The map chart is shaded on the basis of population. Darker the shade, more is the population of that country. This is done mainly to provide user an insight if he/she wants to analyze the population factor with the COVID-19 cases. Clicking on any country paves the path for an area chart that

represents the number of COVID-19 confirmed cases with increasing time.

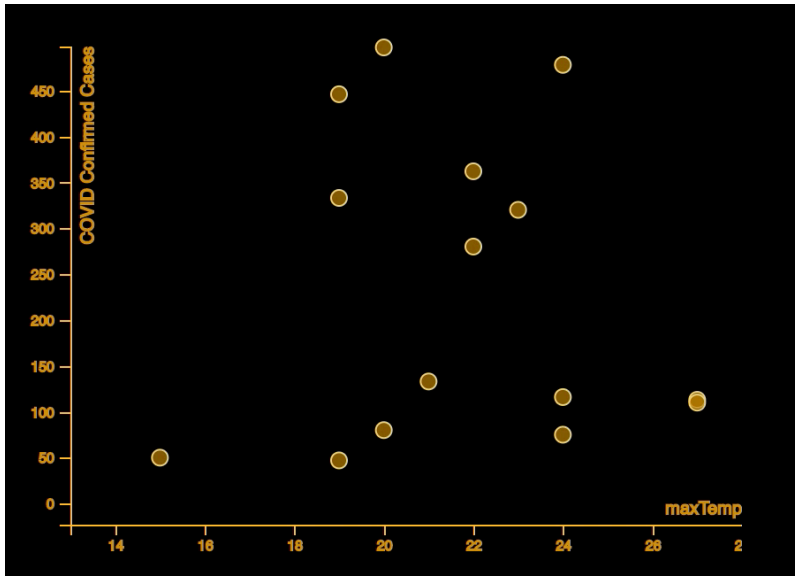


2) Area chart: Above is how the area chart is represented in the dashboard. x-axis comprises of the timeline and y-axis represents the cases. The area chart is brushable. The window can be chosen of any dimension along the x-axis and once chosen, can even be dragged throughout the axis. By brushing the user basically selects the domain of times for which he/she needs to analyze data. The brushed window is then linked to a grouped bar chart.



3)Bar chart: The bar chart's domain is linked to the brushed window in the area chart. The bar chart is grouped for min temperatures and max temperatures for the selected country and time range. It is a user interactive chart. Clicking on any bar (minTemp/ maxTemp) links to the scatter plot for that particular variable.

4) The last section of the dashboard is the scatterplot window which analyzes the correlation between Covid-19 cases and the temperature factor. It is also linked to the brushed domain. The data changes accordingly as the user changes the brushed window. The x axis denotes the parameter selected in the bar chart and the y axis denotes the number of confirmed cases.



Observations

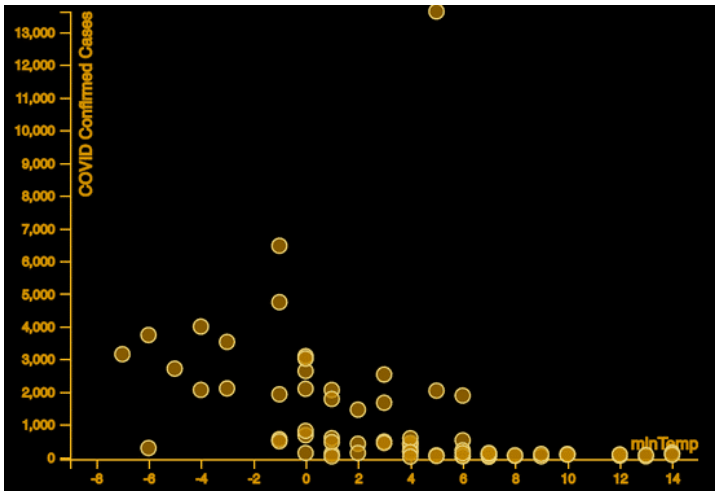
Some of the factors analyzed were:

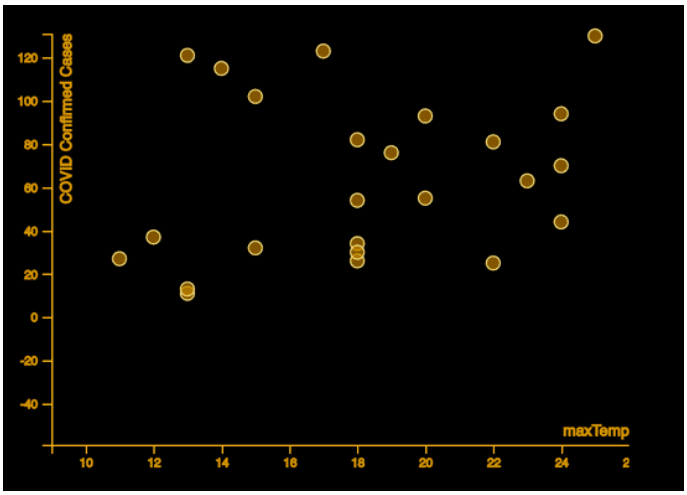
1.Correlation of COVID-19 cases with cold climatic conditions

2.Correlation of COVID-19 cases with hot climatic conditions

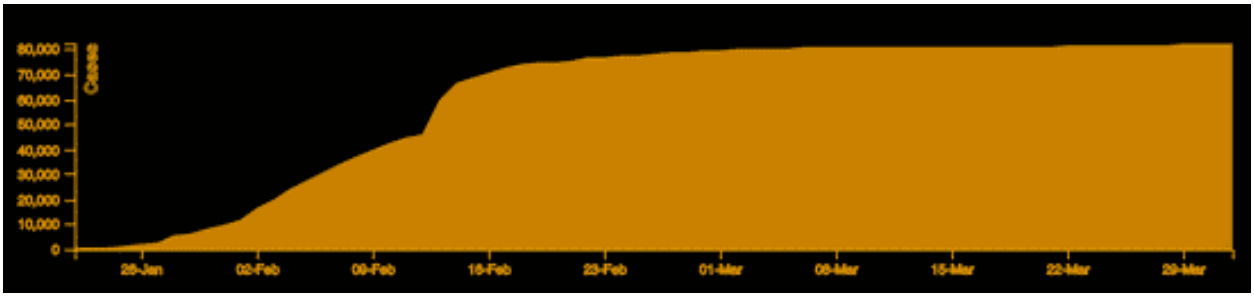
3.Changes in the weather over a period of time

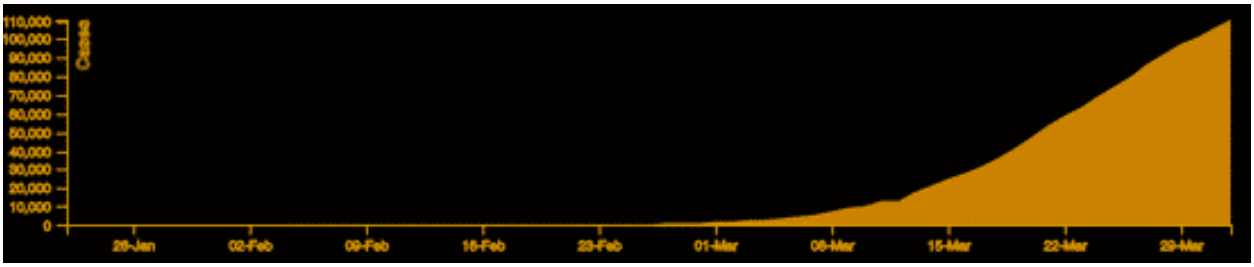
4. Flattening of the curve with the lockdown imposed early and late





Talking of the first two factors (1&2), analysis for China as shown in the above two scatterplots shows that the virus spread was more when the average weather condition was hot compared to when it was cold. Again, there weren't any prominent trends for all countries and this is just an inference that we can make out by looking at the data. The virus spread is not solely dependent on 2-3 factors, it is dependent on multiple other factors too. In general there were no





patterns observed using which we could prove or disprove the hypothesis stating the relation between weather and Covid spread.

Now lets talk about the 4th factor i.e. is the flattening of curve. It was observed that the countries which imposed lockdown early, flattened the curve early. Above are the graphs shown for China and Italy. It can be clearly seen that the graph for China flattened before Italy as the lockdown was imposed quite early.

Further Work

I have already prepared data for some other factors like humidity , coastline and air quality index which I aim to incorporate into the dashboard in the near future. This is an open ended topic without definite inferences, but adding factors would strengthen the hypothesis. The current data spreads from Jan 22 to April 1. I also plan to update the data once the present datasets become available.

Conclusion

It was observed that the corona spread was more when the climatic conditions were hot. Although the hypothesis can't be proved true with these observations but it's a step forward towards a finite conclusion. The current studies by Harvard and MIT suggest that from the evidence so far, the COVID-19 virus can be transmitted in ALL AREAS, including areas with hot and humid weather. The countries who imposed lockdown earlier flattened the curve before other countries. As such there seems to be no prominent correlation between temperature and the COVID19 cases which implies that other factors need to be

considered for the hypothesis to be proved true. Another observation is that weather/climate and COVID-19 positive cases share a symbiotic relation. They both are interrelated to each other. This analysis serves as a base to add other environmental factors and get some more solid inferences. As mentioned in the current studies about weather and Covid, there wasn't any noticeable trend over the timeline. Adding more and more factors to the analysis may somewhat help us reach a definite inference.