Early Outbreak Analysis of COVID-19 Epidemic: China and Global Health Perspectives

Samrat Kumar Dey (✉ sopnil.samrat@gmail.com)

Dhaka International University <https://orcid.org/0000-0002-7999-8576>

Md. Mahbubur Rahman

Military Institute of Science and Technology

Umme Raihan Siddiqi

Shaheed Suhrawardy Medical College and Hospital

Arpita Howlader

Patuakhali Science and Technology University

Research Article

Keywords: COVID-19, outbreak, data analysis, coronavirus, visualization, SARS-CoV-2, China

Posted Date: February 22nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-44328/v2>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Purpose: Globally, there is an obvious concern about the fact that the evolving 2019-nCoV coronavirus is a worldwide public health threat. The appearance in China at the end of 2019 of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2; previously provisionally labeled as 2019 novel coronavirus or 2019-nCoV) disease (COVID-19) caused a major global outbreak and right now is a major community health issue. As of 8 March 2020, World Health Organization (WHO) data showed that more than 105 500 confirmed cases were reported in over 100 countries/regions, with > 75% of cases being detected in China and >24% of cases detected globally. COVID-19 outbreak is evolving so rapidly; therefore, the available epidemiological data are essential to direct strategies for situational awareness and intervention.

Methods: This article will present a visual exploratory data analysis (V-EDA) approach to collect and analyze COVID-19 data on epidemiological outbreaks. Various open data sources on the outbreak of COVID-19 provided by the World Health Organization (WHO), the Chinese Center for Disease Control and Prevention (CDC), the National Health Commission (NHC), Johns Hopkins University Interactive Dashboard and DXY.cn have been used in this research.

Results: Therefore, an Exploratory Data Analysis (EDA) with visualizations has been designed and developed in order to understand the number of different cases reported (confirmed, death, and recovered) in different provinces of China and outside of China between 22 January 2020 to 4 March 2020. Various open data sources on the outbreak of COVID-19 provided by the World Health Organization (WHO), the Chinese Center for Disease Control and Prevention (CDC), the National Health Commission (NHC), Johns Hopkins University Interactive Dashboard and DXY.cn have been used in this research.

Conclusion: In all, this is extremely important to promptly spread information to understand the risks of this pandemic and begin containment activities.

1. Introduction

The latest outbreak of pneumonia in Wuhan, China has put the 2019 Novel Coronavirus (2019-nCoV) closely into our sight. The first cases of pneumonia were identified in Wuhan City, the capital of Hubei Province, China in early December 2019 was later determined as a non-SARS novel coronavirus by the Chinese Center for Disease Control and Prevention (CDC) (Huang et al. 2020). Coronaviruses are non-segmented positive-sense RNA viruses that belong to the *Coronaviridae* family and the *Nidovirales* order and are widely distributed in humans and other mammals (Ksiazek et al. 2009). The pathogen has been described as a novel enveloped RNA *betacoronavirus* (Lu et al. 2020) commonly referred to as severe acute respiratory coronavirus 2 syndrome (SARS-CoV-2), having a phylogenetic resemblance to SARS-CoV (Zhu et al. 2020). With a mortality rate of 10% for SARS-CoV and 37% for MERS-CoV, (WHO 2003; WHO 2019) the epidemics of the two beta coronaviruses, severe acute respiratory syndrome coronavirus (SARS-CoV), (Drosten et al. 2003; Kuiken et al. 2003) and Middle East respiratory syndrome coronavirus (MERS-CoV), (Zaki et al. 2012; Groot et al. 2013) have caused more than ten thousand cumulative cases

in the last two decades. Recently the World Health Organization (WHO) declared coronavirus disease 2019 (COVID-19) an international public health emergency. As of March 8, 2020, a total of ($Confirmed_{global}=105586$) laboratory-confirmed cases have been documented globally and among them, China has ($Confirmed_{china}=80859$) cases with a death rate of ($Deaths_{china}=3100$) patients (WHO Situation Report-48). The world has witnessed a rapid increase in different cases of COVID-19 caused by SARS-CoV-2 viruses in the last week or so. Although this article analyzes all the data sources till 4 March 2020, the coronavirus COVID-19 is affecting 104 countries and territories around the world and 1 international conveyance (the Diamond Princess cruise ship harbored in Yokohama, Japan) till to date. However, till 8 March 2020 several new cases were found globally including 1492 new cases were confirmed in Italy; 743 new cases in Iran; 272 new cases in South Korea; 240 new cases in Germany; 177 new cases in France; 88 new cases in Spain; 64 new cases in the United Kingdom; 42 new cases in Sweden; 41 new cases in Japan; 23 new cases in Austria; 16 new cases in Norway; 12 new cases in Singapore; 10 new cases in San Marino, and in Vietnam; 8 new cases in Denmark; 7 new cases found in Greece, Portugal and in the Netherlands; 6 new cases observed in India and in Malaysia (1 recovery also observed); 5 new cases reported in Iceland, and in Czech Republic; 4 new cases found in Israel, Philippines, Saudi Arabia, Finland, Hong Kong, Lebanon, Slovenia and in Costa Rica; 3 new cases found in Bangladesh, Canada, Russia, Qatar, and in Dominican Republic; 2 new cases found each in Iraq, Ireland, Romania, Kuwait, Hungary, Poland, Australia, Indonesia and in Bulgaria; 1 new case in Brazil, Kuwait, Latvia, South Africa, and in Mexico. Several deaths cases were also been reported worldwide including 133 new deaths in Italy on a single day; 1 new death observed in United Kingdom, Switzerland, Egypt, Hong Kong, Japan, and in Algeria; 2 new deaths in the Netherlands, and in South Korea; 3 new deaths in France; 7 new deaths in Spain and 49 new deaths in Iran. Regarding China, 44 new cases, 27 new deaths (all in Hubei) and 1,661 new discharges occurred on March 7, as reported by the National Health Commission (NHC) of China. Considering COVID-19's rapid spread, we have decided that an updated case review with outbreak analysis across worldwide and in China may help identify the epidemiological characteristics and severity of the disease. As the outbreak of COVID-19 is expanding rapidly in China and beyond, threatening to become a global pandemic, therefore we aim to describe the number of different cases caused by SARS-CoV-2 and visualize the epidemiological data through visual exploratory data analysis (EDA) approach. We hope our results from the study can inform the global community of this COVID-19 outbreak and create public awareness.

2. Materials And Methods

We used the Novel Corona Virus 2019 open dataset provided by Johns Hopkins University; they built an exceptional dashboard using the data of the affected cases to date (Dong et al. 2020). In addition, by providing the data accessible in google sheets format, they also provide an opportunity for the data analyst and researchers to understand the pattern of the different COVID-19 cases around the globe. Day-wise knowledge about the people affected can, therefore, provide some interesting insights when it is made available to the wider data science community. This dataset provides daily-level information on the number of confirmed cases, deaths cases and recovery from 22 January 2020 to 4 March 2020. The

study presented here, using the dataset given by John Hopkins University, the World Health Organization (WHO), the Chinese Center for Disease Control and Prevention (CDC), the National Health Commission (NHC), and DXY.cn, focused on exploratory data analysis (EDA) and visual exploratory data analysis (V-EDA). Nonetheless, all the datasets were pre-processed and cleaned properly and made suitable for experimentation. We have used NumPy's python-based library (<https://numpy.org>) and pandas (<https://pandas.pydata.org>) to store and analyze the data. Matplotlib (<https://matplotlib.org>), Plotly (<https://plot.ly>), Seaborn (<https://seaborn.pydata.org>), and Folium (<https://python-visualization.github.io/folium/>) were also used for interactive visualization of the highlighted results. All the analysis with the dataset was performed using Jupyter Notebook (<https://jupyter.org>) support on a Linux-based local computer platform, using Python Language in the research laboratory of Dhaka International University (DIU).

2.1 Dataset Description

We have used various dataset sources for data analysis and visualizations for this research work. Mainly, three different data sources were used including the 2019 Coronavirus dataset (January-February 2020) that monitors the spread of the 2019-nCoV, COVID-19 (nCOV-19) Corona Virus Spread dataset consisting of the number of confirmed, recorded deaths and recovered cases, and the 2019 Novel Corona Virus dataset that manages the day-level information on affected cases in 2019-nCoV (Dey et al. 2020).

2.2 Exploratory Data Analysis (EDA) Approach

We analyzed the datasets using various methods of exploratory data analysis and visualization methods to provide ample evidence of the COVID-19 outbreak worldwide. Three different data sources have been used including 2019 Coronavirus dataset (January-February 2020), COVID-19 (nCOV-19) Corona Virus Spread Dataset and Novel Corona Virus 2019 Datasets for exploratory data analysis. All the analysis and visual exploration are based on various datasets from 22 January 2020 to 4 March 2020. Nevertheless, as compared to the rest of the world, a large number of cases are recorded in China (Table 1). Predictably, with no surprise in China most of the reported cases are from a particular Hubei province. It is no surprise, because the capital of Hubei is Wuhan, where the first cases are registered (Table 2). Until 4 March 2020, COVID-19 propagated 84 countries worldwide and 31 states or provinces in China.

Table 1: Tabular representation of world wide reported cases caused by SARS-CoV-2 till 4 March 2020 (Countries_{confirmed} = 84). China reported highest number of cases with a mortality rate of ($MR_{\text{china}} = 3.71\%$).

<i>World wide data of different cases from 22 January 2020 to 4 march 2020 (based on highest proportion)</i>					
Country/Region	Confirmed	Deaths	Recovered	Mortality Rate (%)	Recovery Rate (%)
China	80271	2981	49955	3.71	62.23
South Korea	5621	35	41	0.62	0.73
Italy	3089	107	276	3.46	8.93
Iran	2922	92	552	3.15	18.89
Japan	331	6	43	1.81	12.99
France	285	4	12	1.40	4.21
Germany	262	0	16	0.00	6.11
Spain	222	2	2	0.90	0.90
US	153	11	8	7.19	5.23
Singapore	110	0	78	0.00	70.91
Hong Kong	105	2	37	1.90	35.24
Switzerland	90	0	3	0.00	3.33
UK	85	0	8	0.00	9.41
Norway	56	0	0	0.00	0.00
Kuwait	56	0	0	0.00	0.00
Bahrain	52	0	0	0.00	0.00
Australia	52	2	11	3.85	21.15
Malaysia	50	0	22	0.00	44.00
Thailand	43	1	31	2.33	72.09
Taiwan	42	1	12	2.38	28.57
Netherlands	38	0	0	0.00	0.00
Sweden	35	0	0	0.00	0.00
Iraq	35	2	0	5.71	0.00
Canada	33	0	6	0.00	18.18
Austria	29	0	0	0.00	0.00
India	28	0	3	0.00	10.71

United Arab Emirates	27	0	5	0.00	18.52
Iceland	26	0	0	0.00	0.00
Belgium	23	0	1	0.00	4.35
San Marino	16	1	0	6.25	0.00
Vietnam	16	0	16	0.00	100.00
Israel	15	0	1	0.00	6.67
Oman	15	0	2	0.00	13.33
Lebanon	13	0	1	0.00	7.69
Algeria	12	0	0	0.00	0.00
Denmark	10	0	0	0.00	0.00
Croatia	10	0	0	0.00	0.00
Ecuador	10	0	0	0.00	0.00
Macau	10	0	9	0.00	90.00
Greece	9	0	0	0.00	0.00
Qatar	8	0	0	0.00	0.00
Czech Republic	8	0	0	0.00	0.00
Belarus	6	0	0	0.00	0.00
Finland	6	0	1	0.00	16.67
Ireland	6	0	0	0.00	0.00
Portugal	5	0	0	0.00	0.00
Pakistan	5	0	0	0.00	0.00
Mexico	5	0	1	0.00	20.00
Senegal	4	0	0	0.00	0.00
Brazil	4	0	0	0.00	0.00
Romania	4	0	1	0.00	25.00
Azerbaijan	3	0	0	0.00	0.00
Saint Barthelemy	3	0	0	0.00	0.00
Russia	3	0	2	0.00	66.67
New Zealand	3	0	0	0.00	0.00

Georgia	3	0	0	0.00	0.00
Philippines	3	1	1	33.33	33.33
Estonia	2	0	0	0.00	0.00
Indonesia	2	0	0	0.00	0.00
Hungary	2	0	0	0.00	0.00
Egypt	2	0	1	0.00	50.00
Ukraine	1	0	0	0.00	0.00
Armenia	1	0	0	0.00	0.00
Chile	1	0	0	0.00	0.00
Andorra	1	0	0	0.00	0.00
Tunisia	1	0	0	0.00	0.00
Sri Lanka	1	0	1	0.00	100.00
Cambodia	1	0	1	0.00	100.00
Argentina	1	0	0	0.00	0.00
Dominican Republic	1	0	0	0.00	0.00
Saudi Arabia	1	0	0	0.00	0.00
Poland	1	0	0	0.00	0.00
North Macedonia	1	0	0	0.00	0.00
Nigeria	1	0	0	0.00	0.00
Nepal	1	0	1	0.00	100.00
Morocco	1	0	0	0.00	0.00
Monaco	1	0	0	0.00	0.00
Luxembourg	1	0	0	0.00	0.00
Lithuania	1	0	0	0.00	0.00
Liechtenstein	1	0	0	0.00	0.00
Latvia	1	0	0	0.00	0.00
Jordan	1	0	0	0.00	0.00
Faroe Islands	1	0	0	0.00	0.00
Gibraltar	1	0	0	0.00	0.00

Afghanistan	1	0	0	0.00	62.23
-------------	---	---	---	------	-------

2.3 Visual Exploratory Data Analysis (V-EDA) Approach

This segment will analyze the various time series data using some methods of visual exploratory data analysis. We have built a global map and provided awareness of how SARS-CoV-2 spread around the globe from 22 January 2020 to 4 March 2020. Folium has been used to spot the diverse regions or countries on the Map which is affected by SARS-CoV-2 viruses within China and rest of the world. OpenStreetMap (<https://www.openstreetmap.org/>) and Leaflet (<https://leafletjs.com/>) are also used with Folium to represent a specific region on the Map. We have designed seven different map representation including the cases reported across the globe, cases reported in China, countries with confirmed cases, countries with deaths reported, date wise spread outside China, date wise spread inside China and lastly pointed out the regions of Diamond princess cruise ship case. It is evident from the representation of the map that China registered the highest confirmed cases with a high number of ($Confirmed_{china}=78498$) patients. Similarly, China recorded the highest number of deaths and it was up to March 4, 2020 ($Deaths_{china}=2744$). We also analyze time-series data using visual EDA to provide a clear and comprehensive outcome about the severity of COVID-19 outbreak. It is understandable that processing these data in real-time is extremely useful in documenting this serious disease's epidemic behavior. Our research team suggests that this way of data analysis can certainly increase situational awareness and combat strategies. All the models of data analysis and visualization we have developed for this research article, like EDA and V-EDA, are available in this URL (http://samratdey.me/analysis_covid19.html).

Table 2: Tabular representation of case reported by different regions of China caused by SARS-CoV-2 until 4 March 2020 ($China_Region_{confirmed} = 31$). Hubei (Hubei's capital is Wuhan, where the first cases are reported), reported highest number of cases with a mortality rate of ($MR_{hubei} = 4.02\%$).

<i>Chines Province wise latest Data (Till 4 March 2020)</i>						
Province/State	Confirmed	Deaths	Recovered	Affected	Mortality Rate (%)	Recovery Rate (%)
Hubei	65596	2641	23383	39572	4.02	35.65
Guangdong	1347	7	890	450	0.51	66.07
Henan	1272	20	1068	184	1.57	83.96
Zhejiang	1205	1	932	272	0.08	77.34
Hunan	1017	4	804	209	0.39	79.06
Anhui	989	6	792	191	0.60	80.08
Jiangxi	934	1	754	179	0.10	80.73
Shandong	756	6	387	363	0.79	51.19
Jiangsu	631	0	498	133	0	78.92
Chongqing	576	6	401	169	1.04	69.62
Sichuan	534	3	321	210	0.56	60.11
Heilongjiang	480	13	270	197	2.70	56.25
Beijing	410	5	248	157	1.21	60.49
Shanghai	337	3	276	58	0.89	81.90
Hebei	317	6	274	37	1.89	86.44
Fujian	296	1	228	67	0.33	77.03
Guangxi	252	2	161	89	0.79	63.89
Shaanxi	245	1	195	49	0.40	79.59
Yunnan	174	2	150	22	1.14	86.21
Hainan	168	5	131	32	2.97	77.98
Guizhou	146	2	112	32	1.36	76.71
Tianjin	136	3	102	31	2.20	75.00
Shanxi	133	0	107	26	0	80.45
Liaoning	121	1	93	27	0.82	76.86
Jilin	93	1	67	25	1.07	72.04
Gansu	91	2	81	8	2.19	89.01
Xinjiang	76	2	43	31	2.63	56.58

Inner Mongolia	75	0	43	32	0	57.33
Ningxia	72	0	68	4	0	94.44
Qinghai	18	0	18	0	0	100.00
Tibet	1	0	1	0	0	100.00

3. Results

This research analyzed different data sources of COVID-19 outbreak and enlisted significant results obtained from the analysis. However, as this exploration took place in between 22 January 2020 to 4 March 2020, therefore all the exploratory data analyses are based on the specific and authentic data sources provided by different organizations. From 22 January 2020 to 4 March 2020, COVID-19 propagated 84 countries worldwide (Figure 1) and 31 states or provinces in China (Figure 2). Moreover, 2 new deaths and 34 new confirmed cases were reported in the United States. Also, 4 new confirmed cases were reported in Canada (British Columbia), including a woman in her 80s who is in critical condition at Vancouver General Hospital. Globally, till 4 March 2020, several cases were reported in different region of the world including 1 new case in Chile, 1 new case in Brazil (São Paulo), 3 new cases in Ecuador, 1 new case in Croatia, first case in Slovenia and in Poland, 4 new cases in Ireland, 587 new cases and 28 new deaths in Italy. Apart from these data, we also observed first 2 cases in Hungary, 22 new cases in Sweden, 63 new cases and 1 new death in Spain, 59 new cases in Germany, 73 new cases in France, 36 new cases in the UK, 23 new cases in Norway, 12 new cases in Iceland, 2 new cases and 1 new recovery in Romania, 15 new cases in the Netherlands, 5 new cases in Denmark and in San Marino, 3 new cases in the Czech Republic, 10 new cases in Belgium, 2 new cases in Portugal, Belarus, Greece and Scotland, 35 new cases in Switzerland, and lastly 5 new cases in Austria were reported. In terms of Middle East, 1 new case in the United Arab Emirates, 1 new case in Saudi Arabia, 3 new cases in Oman, 2 new deaths and 3 new cases in Iraq, 2 new cases in Lebanon, 3 new cases in Bahrain, 586 new cases, 15 new deaths, and 117 new recoveries in Iran, and 3 new cases in Israel were reported.

However, in Africa region only 2 new cases in Senegal and 9 new cases in Algeria were reported till 4 March 2020. Oceanian region confirmed it first case in New Zealand whereas 1 new death in Australia and 14 new cases were reported till date. Most alarming region Asia has confirmed huge number of cases including 2 new cases in Singapore and in Hong Kong, 38 new cases in Japan, 14 new cases in Malaysia, 22 new cases in India, 119 new cases, 38 new deaths (37 in Hubei) and 2,652 new discharges occurred in China on March 3, as reported by the National Health Commission (NHC) of China, and lastly 435 new cases and 2 deaths in South Korea were reported on 4 March 2020.

A visual exploratory data analysis on different datasets till 4 March 2020 provides significant outbreak information (C=Confirmed, D=Deaths, R=Recovered) regarding COVID-19 inside China (Figure 3). China confirmed (C=67 332) cases for the cause of SARS-CoV-2 virus between 22 January 2020 to 4 March

2020. However, the numbers of death cases (D=2871) with an increasing number of recovery growth (R=38 557) reported within China. From the analyzed data it is evident that, the mortality rate (MR=4.26%) increases gradually in comparison with previous cases. Surprisingly, recovery rate of SARS-Cov-2 affected patients increases dramatically with a recovery rate of (RR=57.27%) till 4 March 2020.

Regarding to global concern, Figure 4 also provides a concrete idea how globally number of peoples are affected with this novel corona virus (SARS-CoV-2). Outside China there were (C=13 995) confirmed cases till 4 March 2020 which indicates a rapid increase of COVID-19 outbreak worldwide. There were (D=264) deaths with a mortality rate of (MR= 1.89%) which was much less comparing with the mortality rate of China till 4 March 2020. With a recovery rate of (RR=8.50%), around (R=1190) cases of recovered patients were reported outside china which indicates a gradually slow recovery around the globe.

For better understandings of different cases (C=Confirmed, D=Deaths, R=Recovered) we have designed and analyzed the data of each day from 22 January 2020 to 4 March 2020 within China and Outside of China. Significant deviation of different new cases were found on 13 February 2020, where highest number of confirmed cases were reported in different provinces in China (Figure 5). China experienced highest number of confirmed cases till now on COVID-19 Outbreak with a ($C_{\text{new_case}} = 15\,133$) patients on a single day. However, it also observed a highest number of deaths ($D_{\text{new_case}} = 252$) with a recovery cases of ($R_{\text{new_cases}} = 1134$) on 13 February 2020. All the date by date analysis and visualization results have been made available for mass access and public awareness (http://samratdey.me/analysis_covid19.html).

In terms of global health concern on the issue of new cases reported, on 3 March 2020 world has observed a rapid number of confirmed cases all around the globe. With almost ($C_{\text{new_cases_globally}} = 2409$) new cases were reported which is treated as highest number of reported confirmed patients between 22 January 2020 to 4 March 2020. Moreover, on the very next day, highest numbers of deaths were reported on a single day worldwide with ($D_{\text{new_cases_globally}} = 58$) patients. In general, an increasing trend of different cases were found from the result of the visualization model. On 4 March, 2020 ($R_{\text{new_cases_globally}} = 391$) patients were recovered from SARS-CoV-2 virus around the world, which was also the highest number of patients who recovered on a single day in between 22 January 2020 to 4 March 2020 (Figure 6).

Figure 7 emphasizes on the different countries (except the China) proportional cases analysis from 22 January 2020 to 4 March 2020. We have categorized each case with three different variables; (A=Affected, D=Deaths and R=Recovered). For calculating affected rate for any specific region or country the following equation is used in this data analysis.

$$\text{Affected}_{\text{countries/regions}} = [\text{Confirmed}_{\text{countries/regions}}] - [\text{Deaths}_{\text{countries/regions}}] - [\text{Recovered}_{\text{countries/regions}}] \dots \dots \dots (1)$$

South Korea ($Confirmed_{SK} = 5621$), a country of Asia, also a neighboring country of China were severely affected with SARS-CoV-2 viruses. Around ($Affected_{SK} = 5545$) patients were affected till 4 March 2020 which considered as the highest number of affected cases reported by any countries after the China. With a mortality rate of only ($MR_{SK} = 0.62\%$) around ($Deaths_{SK} = 35$) deaths were confirmed till 4 March 2020. However, only ($Recovered_{SK} = 35$) patients recovered fully in South Korea. Countries like Iran ($Confirmed_{IR} = 2922$) and Italy ($Confirmed_{IT} = 3089$) also experienced grave situation in terms of health management due to the cause of COVID-19 outbreak. Italy ($Affected_{IT} = 2706$, $Death_{IT} = 107$, $Recovered_{IT} = 276$, $MR_{IT} = 0.55\%$, $RR_{IT} = 8.93\%$) and Iran ($Affected_{IR} = 2278$, $Death_{IR} = 92$, $Recovered_{IR} = 552$, $MR_{IR} = 3.14\%$, $RR_{IR} = 18.89\%$) experienced rapid number of different cases suddenly for the last 3 weeks (till 4 March 2020).

Apart from this, we also analyzed each single data based on different provinces of China and illustrated them in Figure 8. With no surprise Hubei has ($Affected_{Hubei} = 25\ 904$) affected patients, the greatest number of confirmed ($Confirmed_{Hubei} = 65596$) cases reported till date by any province in China. However, the number of death rate were also very alarming and it reaches almost to ($Hu_{death} \sim 3000$); and till 4 March 2020 it reached to ($Death_{Hubei} = 2871$) with a mortality rate of ($MR_{Hubei} = 4.37\%$) and a high recovery rate of ($RR_{Hubei} = 58.77\%$).

Figure 9 enlist the data of comparative analysis (Confirmed=C, Recovered=R, and Deaths=D) of Hubei, other provinces of China and the rest of the world till 4 March 2020. This representation demonstrates that Hubei has endured the largest number of infected patients ($C_{hubei} = 67332$). However, Hubei has also maintained a significant recovery rate of ($R_{hubei} = 38557$) patients along with the mortalities of ($D_{hubei} = 2871$) persons. On the other hand, rest of the provinces in China has confirmed ($C_{other_china_province} = 12939$) patients infected by SARS-COV-2 virus till 4 March 2020. Like Hubei, other provinces in China also showed a dramatic recovery rate of ($R_{other_china_province} = 11398$) patients along with confirmed deaths of ($D_{other_china_province} = 110$) persons. As of 4 March 2020, data from the different sources showed that there was a total of ($C_{rest_of_world} = 14147$) confirmed cases of COVID-19 worldwide. Among them, ($D_{rest_of_world} = 267$) deaths have been reported globally with a steady recovery rate of ($R_{rest_of_world} = 1206$) patients. From the observation, it is apparent that there has been a steady rise in the daily total number of COVID-19 cases globally, both within and outside China till 4 March 2020.

Finally, we performed a ratio analysis for three different cases analysis including number of deaths to 100 confirmed cases, number of recovered to 100 confirmed cases and number of recovered to 1 death cases (Figure 10). This analysis shows that during the first few weeks of this epidemic there were more deaths reported per day than recovered cases. However, over the time this trend has changed drastically. Although the death rate has not come down in significant level, the number of recovered cases has definitely increased with the passage of time. Till 4 March 2020, we have observed 54.2% of death cases in each 100 confirmed cases, 15.75% of recovery rate of each death cases, and 3.4% of recovered to hundreds of confirmed cases.

4. Discussion

We report here all the different cases (confirmed, death and recovered) with laboratory test confirmed caused by SARS-CoV-2 viruses across the world and in China between 22 January 2020 to 4 March 2020. The number of cases is rising very rapidly. As of 8 March 2020, according to the situation report (https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200308-sitrep-48-covid-19.pdf?sfvrsn=16f7ccef_4) provided by World Health Organization (WHO), 8 new countries/territories/areas (Bulgaria, Costa Rica, Faroe Islands, French, Guiana, Maldives, Malta, Martinique, and Republic of Moldova) and overall 100 countries now have reported laboratory-confirmed cases of COVID19 in the past 24 hours. When an outbreak like this happens, readily available data and information are equally important for beginning the assessment needed to understand the risks and start containment of outbreak activities. These information includes initial reports of countries with a confirmed, death and recovered case ratio, as well as how countries outside of China are affected, how the Chinese province are struggling to deal with the COVID-19 situation and, more significantly, the ratio analysis of these real-world data, as well as information obtained from different regions of the world from past outbreaks. Knowledge and understanding of the consequences are needed to help improve the risk assessment as the epidemic progresses and to ensure the best care of patients is achieved. Much of this information comes up in real-time, challenging our comprehension, while improving our responses. Currently, there is an apparent need to consider the implications of SARS-CoV-2 viruses not only in China but globally to be mindful of ourselves for days to come. Consequently, this is a small initiative of analyzing and visualizing real-world time series data in such a way that people around the globe have a better understanding of its serious nature. The undesirable occurrence of this SARS-CoV-2 virus is still being observed and, to date, 8 March 2020, the number of death cases reported was ($Deaths_{china}=3100$) in China and ($Deaths_{rest_of_china}=484$) death cases reported outside China. That is extremely alarming not only for China but also for the rest of the world. This research focuses on the issues of COVID-19 outbreak analysis and thus we have enlisted the most reported cases with their mortality and recovery rate in China, outside of China and in various provinces in China. We also analyzed the number of affected countries with reported confirmed, deaths and recovered cases. Besides that, we developed a visualization tool with Map view and Treemap view to examine the COVID-19 epidemiological outbreak in China and around the world.

5. Conclusion

COVID-19 epidemic has become a clinical threat to the general population and healthcare workers around the world. However, awareness on this novel virus (SARS-CoV-2) is still minimal. The different data sources we have used in this exploration can be useful to provide appropriate knowledge on this emerging outbreak of COVID-19. This research also investigated the mortality and recovery rate for several reported cases both in China and outside of China. We believe public authorities of each country across the globe should keep monitoring the situation in every moment. Our research team believes this is an early data analysis and visualization approach of a situation that is changing rapidly across the

globe. Therefore, as the more we learn about this SARS-CoV-2 virus and its associated outbreaks, the better we can respond. We will continue to monitor the epidemiological data of COVID-19 outbreak in upcoming days using data from official sources.

Declarations

Funding

None

Declaration of Interests

All authors declare no competing interest

Ethical Approval

Not required

Acknowledgment

Authors would like to acknowledge the Johns Hopkins University for open sourcing their dataset. Also, like to acknowledge World Health Organization (WHO) for their timely situational reports on 2019-nCoV. We are also highly grateful to CDC, NHC and DXY for making the data available in first place.

References

- Dey SK, Rahman MM, Siddiqi UR, Howlader A (2020) Analyzing the epidemiological outbreak of COVID-19: A visual exploratory data analysis approach. *J Med Virol* 92:632–638 . <https://doi.org/10.1002/jmv.25743>
- Dong E, Du H, Gardner L (2020) An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis* 20:533–534 . [https://doi.org/10.1016/S1473-3099\(20\)30120-1](https://doi.org/10.1016/S1473-3099(20)30120-1)
- Drosten C, Günther S, Preiser W, van der Werf S, Brodt H-R, Becker S, Rabenau H, Panning M, Kolesnikova L, Fouchier RAM, Berger A, Burguière A-M, Cinatl J, Eickmann M, Escriou N, Grywna K, Kramme S, Manuguerra J-C, Müller S, Rickerts V, Stürmer M, Vieth S, Klenk H-D, Osterhaus ADME, Schmitz H, Doerr HW (2003) Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome. *N Engl J Med* 348:1967–1976 . <https://doi.org/10.1056/NEJMoa030747>
- Groot RJ de, Baker SC, Baric RS, Brown CS, Drosten C, Enjuanes L, Fouchier RAM, Galiano M, Gorbalenya AE, Memish ZA, Perlman S, Poon LLM, Snijder EJ, Stephens GM, Woo PCY, Zaki AM, Zambon M, Ziebuhr J (2013) Commentary: Middle East Respiratory Syndrome Coronavirus (MERS-CoV): Announcement of the Coronavirus Study Group. *J Virol* 87:7790–7792 . <https://doi.org/10.1128/JVI.01244-13>

Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G, Jiang R, Gao Z, Jin Q, Wang J, Cao B (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* 395:497–506 . [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5)

Ksiazek TG, Erdman D, Goldsmith CS, Zaki SR, Peret T, Emery S, Tong S, Urbani C, Comer JA, Lim W, Rollin PE, Dowell SF, Ling A-E, Humphrey CD, Shieh W-J, Guarner J, Paddock CD, Rota P, Fields B, DeRisi J, Yang J-Y, Cox N, Hughes JM, LeDuc JW, Bellini WJ, Anderson LJ, Group the SW (2009) A Novel Coronavirus Associated with Severe Acute Respiratory Syndrome. In: <http://dx.doi.org/10.1056/NEJMoa030781>. https://www.nejm.org/doi/10.1056/NEJMoa030781?url_ver=Z39.88-2003&rfr_id=ori%3Arid%3Acrossref.org&rfr_dat=cr_pub++0www.ncbi.nlm.nih.gov. Accessed 3 Jul 2020

Kuiken T, Fouchier RA, Schutten M, Rimmelzwaan GF, Amerongen G van, Riel D van, Laman JD, Jong T de, Doornum G van, Lim W, Ling AE, Chan PK, Tam JS, Zambon MC, Gopal R, Drosten C, Werf S van der, Escriou N, Manuguerra J-C, Stöhr K, Peiris JSM, Osterhaus AD (2003) Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome. *The Lancet* 362:263–270 . [https://doi.org/10.1016/S0140-6736\(03\)13967-0](https://doi.org/10.1016/S0140-6736(03)13967-0)

Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N, Bi Y, Ma X, Zhan F, Wang L, Hu T, Zhou H, Hu Z, Zhou W, Zhao L, Chen J, Meng Y, Wang J, Lin Y, Yuan J, Xie Z, Ma J, Liu WJ, Wang D, Xu W, Holmes EC, Gao GF, Wu G, Chen W, Shi W, Tan W (2020) Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet* 395:565–574 . [https://doi.org/10.1016/S0140-6736\(20\)30251-8](https://doi.org/10.1016/S0140-6736(20)30251-8)

WHO (2003). Summary of probable SARS cases with onset of illness from 1 November 2002 to 31 July 2003. In: WHO. https://www.who.int/csr/sars/country/table2004_04_21/en/. Accessed 20 April 2020

WHO (2020). Middle East respiratory syndrome coronavirus (MERS-CoV). In: WHO. <http://www.who.int/emergencies/mers-cov/en/>. Accessed 20 April 2020

WHO Situation Report-48 (2020). Coronavirus disease 2019 (COVID-19) Situation Report-48 .In: WHO. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200308-sitrep-48-covid-19.pdf?sfvrsn=16f7ccef_4. Accessed 8 March 2020

Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus ADME, Fouchier RAM (2012) Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia. *N Engl J Med* 367:1814–1820 . <https://doi.org/10.1056/NEJMoa1211721>

Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P, Zhan F, Ma X, Wang D, Xu W, Wu G, Gao GF, Tan W (2020) A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* 382:727–733 . <https://doi.org/10.1056/NEJMoa2001017>

Figures

Number of Countries to which COVID-19 spread over the time

Figure 1

Number of Countries/Regions to which COVID-19 spread over the time. From 22 January 2020 to 4 March 2020 COVID-19 propagated in 84 countries.

Number of Regions of China to which COVID-19 spread over the time

Figure 2

Number of Provinces of China to which COVID-19 spread over the time. Number of Countries/Regions to which COVID-19 spread over the time. From 22 January 2020 to 4 March 2020 COVID-19 propagated in 31 provinces of China

Cases in China

Figure 3

Different cases (confirmed, deaths, and recovered) reported inside China from 22 January 2020 to 4 March 2020

Cases Outside China

Figure 4

Different cases (confirmed, deaths, and recovered) reported worldwide from 22 January 2020 to 4 March 2020

Number of new cases reported inside China

Figure 5

Number of news cases reported from 22 January 2020 to 4 March 2020 inside China

Number of new cases reported outside China

Figure 6

Number of news cases reported from 22 January 2020 to 4 March 2020 globally

Number of Cases globally

Figure 7

Number of different cases based on countries reported from 22 January 2020 to 4 March 2020 globally. Surprisingly South Korea, Italy and Iran confirmed highest number of cases globally.

Number of Cases inside China

Figure 8

Number of different cases based on provinces/regions reported from 22 January 2020 to 4 March 2020 inside China. With no surprise Hubei has highest number of confirmed patients caused by SARS-CoV-2 virus.

Comparative case analysis

Figure 9

Comparative analysis of different cases reported by Hubei, other provinces of China, and the rest of the world till 4 March 2020. Hubei has confirmed (Confirmedhubei=67332) infected patients whereas other provinces in China and the rest of the world confirmed (Confirmedother_provinces=12939) and (Confirmedworld_wide=14147) cases respectively.

Recovery and Mortality rate over the time

Figure 10

Ratio analysis based on Recovery and Mortality rate over the time within China and Worldwide.