Received: 23 February 2020 | Accepted: 1 March 2020

DOI: 10.1002/jmv.25743

RESEARCH ARTICLE

JOURNAL OF MEDICAL VIROLOGY WILEY

Analyzing the epidemiological outbreak of COVID-19: A visual exploratory data analysis approach

Samrat K. Dey[1] | Md. Mahbubur Rahman[2] | Umme R. Siddiqi[3] | Arpita Howlader[4]

[1]Department of Computer Science and Engineering, Dhaka International University (DIU), Dhaka, Bangladesh

[2]Department of Computer Science and Engineering, Military Institute of Science and Technology (MIST), Mirpur Cantonment, Dhaka, Bangladesh

[3]Department of Physiology, Shaheed Suhrawardy Medical College (ShSMC), Dhaka, Bangladesh

[4]Department of Computer and Communication Engineering, Patuakhali Science and Technology University (PSTU), Dumki, Bangladesh

Correspondence
Samrat K. Dey, Department of Computer Science and Engineering, Dhaka International University (DIU), Dhaka 1205, Bangladesh.
Email: sopnil.samrat@gmail.com

Abstract

There is an obvious concern globally regarding the fact about the emerging coronavirus 2019 novel coronavirus (2019-nCoV) as a worldwide public health threat. As the outbreak of COVID-19 causes by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) progresses within China and beyond, rapidly available epidemiological data are needed to guide strategies for situational awareness and intervention. The recent outbreak of pneumonia in Wuhan, China, caused by the SARS-CoV-2 emphasizes the importance of analyzing the epidemiological data of this novel virus and predicting their risks of infecting people all around the globe. In this study, we present an effort to compile and analyze epidemiological outbreak information on COVID-19 based on the several open datasets on 2019-nCoV provided by the Johns Hopkins University, World Health Organization, Chinese Center for Disease Control and Prevention, National Health Commission, and DXY. An exploratory data analysis with visualizations has been made to understand the number of different cases reported (confirmed, death, and recovered) in different provinces of China and outside of China. Overall, at the outset of an outbreak like this, it is highly important to readily provide information to begin the evaluation necessary to understand the risks and begin containment activities.

KEYWORDS

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

1 | INTRODUCTION

Recent pneumonia outbreak in Wuhan, China, has brought closely into our sight the 2019 novel coronavirus (2019-nCoV). This new coronavirus, named 2019-nCoV, belonging to the *Orthocoronavirinae* subfamily, distinct from Middle East respiratory syndrome-coronavirus and severe acute respiratory syndrome coronavirus (SARS-CoV), was described by Zhu et al.[1] The first case of an unexplained new pneumonia origin was detected on 12 December 2019 and was later determined by the Chinese Center for Disease Control and Prevention (CDC) as a non-SARS nCoV. The *Coronaviridae* family consists of a group of large, single, and plus stranded RNA viruses isolated from multiple species, and it is known to cause the common cold and diarrheal diseases in humans.[2,3] In 2003, the SARS outbreak was associated with a new coronavirus, that is, SARS-CoV.[2,3] However, a number of cases of unknown cause of pneumonia occurred in Wuhan, Hubei, China, in December 2019, with clinical presentations closely resembling viral pneumonia.[4] A total of 1975 cases of pneumonia have been confirmed in China so far (according to the state council information office in Beijing, China's capital, 26 January 2020).[5,6] Yet the virus has tended to spread out of China, since one case in Thailand, one case in Japan, and two cases in Korea had been sequentially reported since 15 January 2020.[7] Surprisingly, the transmission of animals to humans is considered the origin of epidemics, as in November, many patients reported to have visited a local fish and wild animal market in Wuhan. Apart from this, recently, evidence has been gathered for the animal to the human and interhuman transmission of the virus.[6,8] The situation is getting serious day by day and for further prevention and control, it is imperative that we have a better understanding of its pandemic nature. On 30 January 2020, World Health Organization (WHO) declared that COVID-19

TABLE 1 Tabular representation of different data sources of 2019-nCoV

Dataset	Description	Columns
2019 Coronavirus dataset (January-February 2020)		
2019_nC0v_20200121_20200126-SUMMARY.csv	This file is an aggregated version of the 2019-nCoV dataset collected by Johns Hopkins University.	Province/state, country, date last updated, confirmed, suspected, recovered, deaths
COVID-19 (nCOV-19) coronavirus spread dataset		
time_series_2019-ncov-Confirmed.csv	Data about the number of confirmed cases	Province/state, country lat, long, 1/22/20, 1/23/20
time_series_2019-ncov-Deaths.csv	Data about the number of Death cases	Province/state, country lat, long, 1/22/20, 1/23/20
time_series_2019-ncov-Recovered.csv	Data about the number of recovered cases	Data about the number of recovered cases
Novel coronavirus 22019 dataset		
2019_nCoV_data.csv	Daily level information on the number of 2019-nCoV affected cases across the globe	Sno, date, province/state province, country, last update, confirmed, deaths, recovered
time_series_2019_ncov_confirmed.csv	Time-series data of confirmed cases	Province/state, country Lat, Long, 1/22/20, 1/23/20
time_series_2019_ncov_deaths.csv	Time-series data of death cases	Province/state, country Lat, Long, 1/22/20, 1/23/20
time_series_2019_ncov_recovered.csv	Time-series data of recovered cases	Province/state, country Lat, Long, 1/22/20, 1/23/20

Abbreviation: 2019-nCoV, 2019 novel coronavirus; Lat, latitude; Long, longitude.

outbreak as the sixth public health emergency of international concern, following H1N1 (2009), polio (2014), Ebola in West Africa (2014), Zika (2016), and Ebola in the Democratic Republic of Congo (2019).[9] Meanwhile, on 11 February 2020, the WHO announced a new name for the epidemic disease caused by 2019-nCoV: coronavirus disease (COVID-19). Until 24 February 2020, 2019-nCoV has affected more than 79 331 patients in 29 countries/regions and has become a major global health concern. On the basis of situational report-35 by WHO, China has confirmed 77 262 cases with 2595 deaths reported until 24 February 2020. However, the rest of the world has confirmed 2069 cases (300 new cases) with 23 deaths reported so far. Till now, China reported 415 new confirmed cases, with 150 new deaths. In comparison, there were 300 new confirmed cases reported outside of China with the number of new deaths being 6 (https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200224-sitrep-35-covid-19.pdf?sfvrsn=1ac4218d2). With regard to the virus itself, the International Committee on Virus Taxonomy has renamed 2019-nCoV as SARS-CoV-2.[10] As the outbreak of the novel SARS-CoV-2 is expanding rapidly in China and beyond, threatening to become a global pandemic, epidemiological data need to be analyzed in a way so that the exploratory data analysis (EDA) methods and visualization model will increase the situational awareness among the mass community in upcoming days. Health workers, governments, and the public, therefore, need to cooperate globally to prevent its spread.

2 | MATERIALS AND METHODS

We used the open dataset of 2019-nCoV provided by the Johns Hopkins University; they made an exceptional dashboard using the affected cases data to date.[11] Apart from this, they also provide an opportunity for data analyst and researcher by providing the data available in Google sheets format. Therefore, daily level information on the affected people can provide some interesting insights when it is made available to the broader data science community. This dataset has daily level information on the number of affected cases, deaths, and recovery on the cases of 2019-nCoV. As this is a time-series data and so the number of cases on any given day is the cumulative number. The data are available from 22 January 2020. Our research team believes that these epidemic data should be openly available and easily accessible for all health professionals and data scientists. This dataset would serve as a starting point for people to gather more data about epidemics, not just statistics, but also new stories, government responses, and so on.

2.1 | Description of the dataset from different sources

For this study, various sources of the dataset have been used for our analysis and visualizations. Mainly, we have used three different sources of dataset including 2019 coronavirus dataset (January-February 2020), which tracks the spread of 2019-nCoV, COVID-19 (nCOV-19) coronavirus spread dataset, which consists of number of confirmed, death, and recovered reported, and 2019-nCoV dataset, which handles the day level information on 2019-nCoV affected cases. Table 1 provides insight on each dataset and their respective data files with their column description. Moreover, we also developed Table 2 for providing the data analyst with a comprehensive knowledge of every column for the used dataset. We have enlisted each distinguished column from all

TABLE 2 Columns description of 2019-nCoV datasets from different sources

Columns description of 2019-nCoV datasets from different sources	
Sno	Serial number
Date	Date and time of the observation in MM/DD/YYYY HH:MM: SS
Province/state	Province or state of the observation
Country	Country of observation
Last update	Time in UTC at which the row is updated for the given province or country.
Confirmed	Number of confirmed cases
Deaths	Number of deaths
Recovered	Number of recovered cases
Lat	Latitude
Long	Longitude
1/22/20	No. of deaths reported till this day, No. of recovered reported till this day, and No. of suspected reported till this day
02/15/2020	No. of deaths reported till this day, No. of recovered reported till this day, and No. of suspected reported till this day

Abbreviations: 2019-nCoV, 2019 novel coronavirus; UTC, coordinated universal time.

three different sources of the dataset and assemble it according to their appearances in the dataset.

2.2 | Exploratory data analysis

We analyzed our datasets with different EDA methods and visualize those data to provide a sufficient consciousness regarding the outbreak of COVID-19 all over the globe. Our exploit data performed with the 2019 coronavirus dataset (January-February 2020), COVID-19 (nCOV-19) coronavirus spread dataset, and 2019-nCoV datasets. Here, we present an effort to visualize and analyze data between 22 January 2020 and 16 February 2020. However, a massive number of cases are reported in China compared to the rest of the world, and interestingly, the next few affected countries are the neighbors of China. Moreover, even in China, most of the cases reported are from a particular province Hubei. It is no surprise because Hubei's capital is Wuhan, where the first cases are reported. Till now, COVID-19 propagated almost 29 countries worldwide and 31 states or provinces in China. Outside China, as expected, there was not much death due to COVID-19 recorded. Only five deaths outside China are reported until 16 February 2020. There are, however, more cases of recovered than death, and in comparison, with 1770 deaths, there were recovery cases of 10 865 patients. We also provide a map representation of different countries with confirmed cases and death reported respectively till 16 February 2020 based on their longitude and latitude. We have enlisted all 29 affected countries outside China and the number of confirmed cases of different Chinese provinces for providing a vibrant depiction of this intense SARS-CoV-2 (Table 3).

2.3 | Visual exploratory data analysis

This section will discuss different time-series data by using some visual exploratory data analysis (V-EDA) methods. We designed a worldwide map and provides a knowledge of how SARS-CoV-2 spread from 22 January 2020 to 16 February 2020 all around the globe. Each map segment represents a region, by using visual data analytics it helps the individuals to understand the epidemiological nature of COVID-19. From the map representation, it is apparent that China reported the highest confirmed cases with a high number of 70 446. Likewise, China reported the highest number of death, and it was 1765 (till 16 February 2020). We also examine the time-series data using V-EDA to provide a strong and understandable outcome of this extreme outbreak of COVID-19. It is obvious that analyzing these data in real-time is extremely useful in capturing an epidemic behavior of this severe disease. We believe this method of analyzing data will certainly increase the understanding of the situation and inform interventions. All the data analysis and visualization models that we have analyzed for this study including EDA and V-EDA are available at this URL (http://samratdey.me/visualization.html).

3 | RESULTS

This study analyzed three different categories of data including confirmed, death, and recovered cases inside China for a time period of 22 January to 16 February 2020. This will also provide a comparative analysis of all the cases reported inside and outside of China. However, we present different cases worldwide to comprehend the specific numbers of cases reported for a specific time period. After analyzing, there were 58 182 confirmed cases of COVID-19 on 16 February 2020. However, the highest number of deaths reported in China on 16 February 2020 was 1696. Surprisingly, there was a significant number of recovered cases reported till 16 February 2020 and it was around 6639 (Figure 1).

Another investigation on different cases of COVID-19 including confirmed, death, and recovered reported outside China also provided in this study. This also contains three different data representations of confirmed, death, and recovered cases outside China for a time period of 22 January 2020 to 16 February 2020. After analyzing the data, there were 399 confirmed cases of COVID-19 on 16 February 2020 globally. However, the highest number of deaths reported outside China on 16 February 2020 was 5. Surprisingly, there was also a significant number of recovered cases till 16 February 2020, and it is more than 100 according to the dataset we examined (Figure 2).

Table 3 represents all the confirmed cases reported in China provinces between 22 January 2020 to 16 February 2020. For a better understanding of the scenario, we designed a tree map and visualize the data according to different criteria. Our designed tree

TABLE 3 Countries with confirmed, deaths, and recovered reported till 16 February 2020 and different China provinces with confirmed cases

Country/region	Confirmed	Deaths	Recovered
Australia	15	0	8.0
Belgium	1	0	0.0
Cambodia	1	0	1.0
Canada	7	0	1.0
China	70 466	1765	10 748.0
Egypt	1	0	0.0
Finland	1	0	1.0
France	12	1	4.0
Germany	16	0	1.0
Hong Kong	57	1	2.0
India	3	0	3.0
Italy	3	0	0.0
Japan	59	1	12.0
Macau	10	0	5.0
Malaysia	22	0	7.0
Nepal	1	0	1.0
Philippines	3	1	1.0
Russia	2	0	2.0
Singapore	75	0	18.0
South Korea	29	0	9.0
Spain	2	0	2.0
Sri Lanka	1	0	1.0
Sweden	1	0	0.0
Taiwan	20	1	2.0
UK	9	0	8.0
US	15	0	3.0
United Arab Emirates	9	0	4.0
Vietnam	16	0	7.0

Different China provinces with confirmed cases (till 16 February 2020)				
Hubei-58182	Guangdong-1316	Henan-1231	Zhejiang-1167	Hunan-1004
Anhui- 962	Jiangxi- 925	Jiangsu- 617	Chongqing-551	Shandong-537
Sichuan-481	Heilongjiang-445	Beijing-380	Shanghai-328	Hebei-300
Fujian-287	Guangxi-237	Shaanxi-236	Yunnan-171	Hainan-162
Guizhou-144	Shanxi-129	Tianjin-124	Liaoning-121	Gansu-90
Jilin-89	Xinjiang-71	Inner Mongolia-70	Ningxia-70	Qinghai-18
Tibet-1				

map contains all the provinces that confirmed the presence of SARS-CoV-2 patients in China. According to the tree map, it is evident that Hubei province reported the highest number of confirmed cases and the number was 58 182. The next province is Guangdong with most 1316 confirmed cases reported. Alternately, we also visualize the tree map for the number of deaths and recovered cases reported between 22 January 2020 to 16 February 2020. The highest number of deaths reported in Hubei province in China. The greatest number

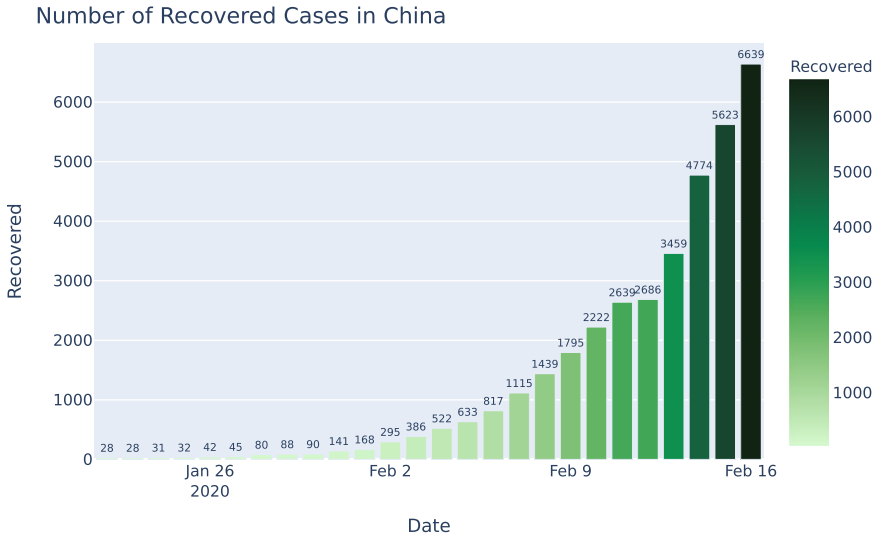

FIGURE 1 Exploratory data analysis of the number of recovered cases in China with data visualization. Initially, till 28 January 2020, the number of recovered patients in China was 23, but surprisingly it increases gradually and lastly till 16 February 2020, the number of recovered patients was 6639

of deaths reported in Hubei was 1696 till 16 February 2020 and the next highest number found in Henan with 13 deaths reported till the mentioned date. However, 6639 successful recovered cases were reported in Hubei and 465 were in Guangdong, and 464 were in Hunan (Table 3). Moreover, a tree map representation for all the countries that have reported confirmed, deaths, and recovered cases globally (except China) also provided in this exploration. Singapore reported the highest number of confirmed cases of COVID-19 immediately after China with a number of 75 and Japan with a confirmed case of 59 is in the next position after Singapore. In terms of deaths reported globally except China, all the five countries (France, Japan, Hong Kong, Taiwan, and the Philippines) reported one death case between 22 January 2020 to 16 February 2020 (Table 3). However, Singapore reported the highest number of recovered cases of 2019-nCoV with a number of 18 and Thailand with a recovered case of 14 is in the next position after Singapore in terms of recovery.

Figure 3 enlists 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 16 February 2020. This representation demonstrates that Hubei has endured the largest number of infected patients (C = 58 182). However, Hubei has also maintained a significant recovery rate of (R = 6639) patients along with the mortalities of (D = 1969) persons. On the other hand, rest of the provinces in China has confirmed (C = 12 264) patients infected by SARS-COV-2 virus till 16 February 2020. Like Hubei, other provinces in China also showed a dramatic recovery rate of (R = 4109) patients along with confirmed deaths of (D = 69) persons. As of 16 February 2020, data from the different sources showed that there was a total of (C = 425) confirmed cases of COVID-19 worldwide. Among them, only five deaths have been reported globally with a steady recovery rate of (R = 117) 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

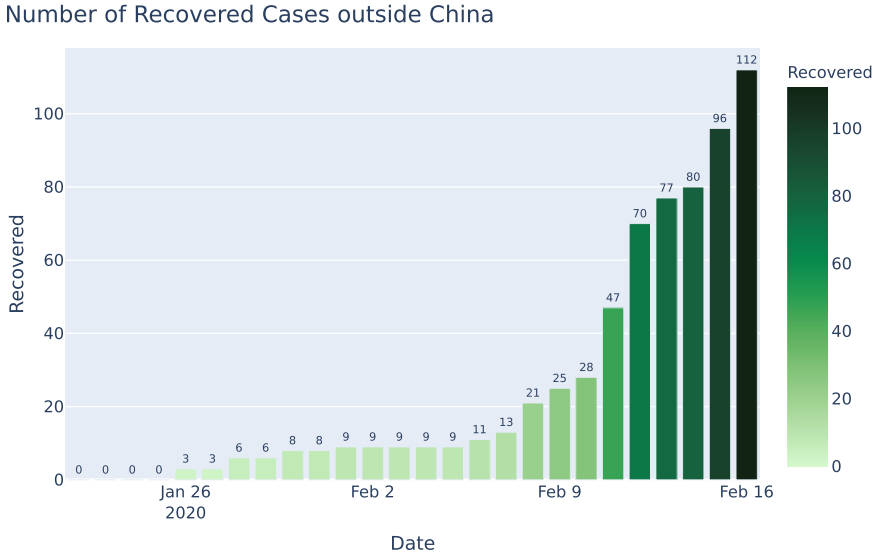

FIGURE 2 Exploratory data analysis of the number of recovered cases outside China with data visualization. It depicts that the rate of recovery outside China also increases regularly and till 16 February 2020, the number of recovered patients was 112 worldwide

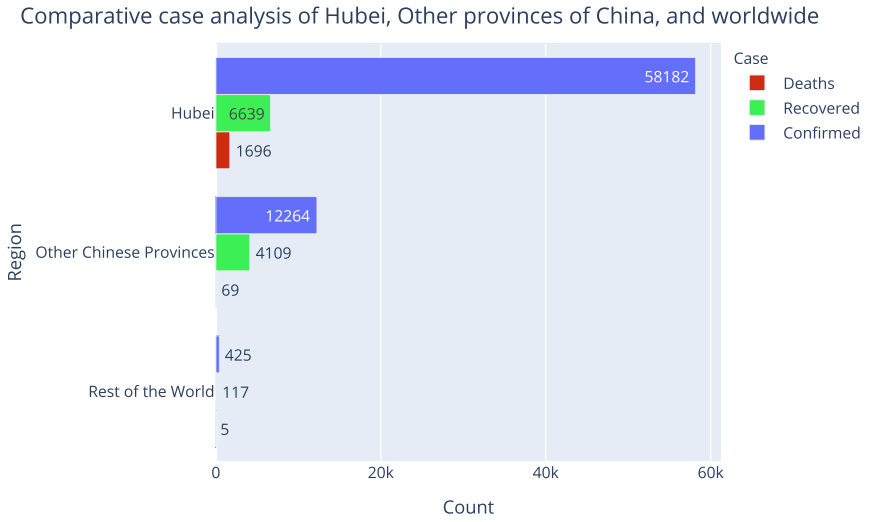

FIGURE 3 Comparative analysis of different cases reported by Hubei, other provinces of China, and the rest of the world till 16 February 2020. Hubei has confirmed 58 182 infected patients, whereas other provinces in China and the rest of the world confirmed 12 264 and 425 cases, respectively

China till 16 February 2020. Regarding new cases of COVID-19, till 24 February 2020, both within and outside China, there has been a dramatic increase in the number of new cases. It was reported that China has confirmed (n = 415/77 262 confirmed) new cases, while the rest of the world has confirmed (n = 300/2069 confirmed) caused by the SARS-COV-2 virus.

4 | DISCUSSION

Another example of the importance of animal-human interface infections is the current outbreak of the COVID-19 disease. And the issues resulting from the advent of a newly identified organism as it spreads through individuals and across national and international frontiers. At the advent of an outbreak, such as this, readily available data and information are equally important to begin the evaluation needed to understand the risks and start containment outbreak activities. Such information includes initial reports of countries with confirmed, death, and recovered cases ratio; also, how the countries outside china are affecting, how the province of China are struggling to handle the situation of COVID-19, and more importantly, ratio analysis of these real-world data, as well as information obtained from specific regions of globally from past outbreaks. Information and understanding of the consequences are needed to help us to refine the risk assessment as the outbreak continues and to ensure that patients are best managed. Much of this information emerges in real-time, challenges our understanding, and yet refines our responses. The analysis presented here based on EDA and V-EDA with the help of the dataset provided by John Hopkins University, WHO, CDC, National Health Commission, and DXY. However, we have preprocessed and cleaned the dataset information according to our needs. For storing and analyzing the data, we have used the python-based library of NumPy (https://numpy.org/) and pandas (https://pandas.pydata.org). Matplotlib (https://matplotlib.org), Plotly (https://plot.ly), Seaborn (https://seaborn.pydata.org), and Folium were also used to visualize the highlighted data in an interactive manner. All the experiment with the dataset has been made by using the support of Jupyter Notebook (https://jupyter.org) in a Linux based local machine platform by using Python Language in the Machine Learning Research Lab at the Dhaka International University. We report here each and every single detail of different cases of COVID-19 between 22 January 2020 to 16 February 2020. Currently, there is an obvious urgency to understand the consequences of SARS-CoV-2 viruses not only in China but also worldwide to aware of ourselves for upcoming days. Therefore, this is a minor initiative of analyzing real-world time-series data and visualize them in such a manner so that people around the globe have better understandings of its severe nature. We are still observing the undesirable prevalence of this SARS-CoV-2 virus, and to date 24 February 2020, the number of death cases reported was 2618, and among them, only 23 death cases reported outside China. This is extremely alarming not only to China but to the rest of the world as well. In this study, we have enlisted the most reported cases in China, outside China, and in different provinces in China. We also analyzed the number of affected countries with reported confirmed, deaths, and recovered cases. Apart from that, we designed map view and tree map's view with an appropriate number to analyze the epidemiological outbreak of the COVID-19.

5 | CONCLUSION

In conclusion, the dataset we have used for our experiment 2019 coronavirus dataset (January-February 2020), COVID-19 (nCOV-19) coronavirus spread dataset, and 2019-nCoV dataset can be useful to monitor the emerging outbreaks, such as 2019-nCoV. Such activities can help us to generate and disseminate detailed information to the scientific community, especially in the early stages of an outbreak, when there is a little else available, allowing for independent assessments of key parameters that influence interventions. We observe an interesting different case reported based on the different datasets of 2019-nCoV, which helps us to understand that it needs more epidemiological and serological studies. We also investigated early indications that the response is being strengthened in China and worldwide on the basis of a decrease in the case of detection time and rapid management of internationally identified

travel-related cases. As a caveat, this is an early data analysis and visualization approach of a situation that is rapidly evolving. To the best of our knowledge, this is the very first attempt on COVID-19, which focuses on the V-EDA based on different data sources. However, knowledge about this novel SARS-CoV-2 virus remains limited among general people around the globe. Raw data released by various sources are not adequately capable to provide an informative understanding of COVID-19, caused of SARS-CoV-2. Therefore, A user-friendly data visualization model will be more effective to understand the epidemic outbreak of this severe disease. Visualization model like map view and tree map view provides an interactive interface and visualize each and every raw fact in a comprehensive manner. Hopefully, in the coming weeks, we will continue to monitor this outbreak's epidemiology data that we have used in this study and from other official sources.

ACKNOWLEDGMENTS

The authors would like to thank the Johns Hopkins University for open sourcing their dataset. They would also like to thank the World Health Organization for their timely situational reports on 2019-nCoV. Atlast, they would like to thank CDC, NHC, and DXY for making the data available in the first place.

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

AUTHOR CONTRIBUTIONS

SKD and MdMR had the idea for and designed the study and had full access to all the data in the study and take the responsibility for the data and accuracy of the data analysis with their visualization. URS and AH contributed to the writing of the study. MdMR contributed to critical revision of the report. All the visualization and data presentation methods were developed by SKD and MdMR. All authors contributed to data acquisition, data analysis, and reviewed and approved the final version.

ORCID


Samrat K. Dey http://orcid.org/0000-0002-7999-8576

REFERENCES

1. Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med*. 2020;382:727-733. https://doi.org/10.1056/NEJMoa2001017
2. Drosten C, Günther S, Preiser W, et al. Identification of a novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med*. 2003;348:1967-1976.
3. Chen Y, Liu Q, Guo D. Emerging coronaviruses: genome structure, replication, and pathogenesis. *J Med Virol*. 2020;92:418-423. https://doi.org/10.1002/jmv.25681
4. WHO. Novel Coronavirus—China January 12, 2020. http://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/. Accessed 19 January 2020.
5. Hui DS, I Azhar E, Madani TA, et al. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health—the latest 2019 novel coronavirus outbreak in Wuhan, China. *Int J Infect Dis*. 2020;2020(91):264-266.
6. Lu H, Stratton CW, Tang YW. Outbreak of pneumonia of unknown etiology in Wuhan China: the mystery and the miracle. *J Med Virol*. 2020;92:401-402. https://doi.org/10.1002/jmv.25678
7. Centers for Disease Control and Prevention. 2019 Novel Coronavirus (2019-nCoV), Wuhan, China. 2019. https://www.cdc.gov/coronavirus/2019-nCoV/summary.html
8. Ji W, Wang W, Zhao X, Zai J, Li X. Cross-species transmission of the newly identified coronavirus 2019-nCoV. *J Med Virol*. 2020;92:433-440. https://doi.org/10.1002/jmv.25682
9. Yoo JH. The fight against the 2019-nCoV outbreak: an arduous march has just begun. *J Korean Med Sci*. 2020;35:e56. https://doi.org/10.3346/jkms.2020.35.e56
10. Gorbalenya AE, Baker SC, Baric RS, et al. Acute respiratory syndrome-related coronavirus: the species and its viruses—a statement of the Coronavirus Study Group [published online ahead of print February 11, 2020]. *bioRxiv*. 2020. https://doi.org/10.1101/2020.02.07.937862
11. Lauren G. Coronavirus COVID-19 Global Cases by Johns Hopkins CSSE January 23, 2020. https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html. Accessed 16 February 2020.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Dey SK, Rahman MM, Siddiqi UR, Howlader A. Analyzing the epidemiological outbreak of COVID-19: A visual exploratory data analysis approach. *J Med Virol*. 2020;92:632–638. https://doi.org/10.1002/jmv.25743