COVID19$^{\alpha}$: Interactive Spatio-Temporal Visualization of COVID-19 Symptoms through Tweet Analysis

Biddut Sarker Bijoy*
Syeda Jannatus Saba*
biddut12,syeda06@student.sust.edu
Shahjalal Univ. of Science & Tech.
Sylhet, Bangladesh

Souvika Sarkar
szs0239@auburn.edu
Auburn University
Alabama, US

Md Saiful Islam
saiful-cse@sust.edu
Shahjalal Univ. of Science & Tech.
Sylhet, Bangladesh

Sheikh Rabiul Islam
shislam@hartford.edu
University of Hartford
Connecticut, US

Md. Ruhul Amin
mamin17@fordham.edu
Fordham University
New York, US

Shubhra Kanti Karmaker
SKS0086@auburn.edu
Auburn University
Alabama, US

ABSTRACT

In this demo, we focus on analyzing COVID-19 related symptoms across the globe reported through tweets by building an interactive spatio-temporal visualization tool, i.e., COVID19$^{\alpha}$. Using around 462 million tweets collected over a span of six months, COVID19$^{\alpha}$ provides three different types of visualization tools: 1) *Spatial Visualization* with a focus on visualizing COVID-19 symptoms across different geographic locations; 2) *Temporal Visualization* with a focus on visualizing the evolution of COVID-19 symptoms over time for a particular geographic location; and 3) *Spatio-Temporal Visualization* with a focus on combining both spatial and temporal analysis to provide comparative visualizations between two (or more) symptoms across time and space. We believe that health professionals, scientists, and policymakers will be able to leverage this interactive tool to devise better and targeted health intervention policies. Our developed interactive visualization tool is publicly available at https://bijoy-sust.github.io/Covid19/.

CCS CONCEPTS

• **Human-centered computing → Geographic visualization**.

ACM Reference Format:
Biddut Sarker Bijoy, Syeda Jannatus Saba, Souvika Sarkar, Md Saiful Islam, Sheikh Rabiul Islam, Md. Ruhul Amin, and Shubhra Kanti Karmaker. 2021. COVID19$^{\alpha}$: Interactive Spatio-Temporal Visualization of COVID-19 Symptoms through Tweet Analysis. In *26th International Conference on Intelligent User Interfaces (IUI '21 Companion), April 14–17, 2021, College Station, TX, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3397482.3450715

1 INTRODUCTION

COVID-19 virus has created a widespread sense of uncertainty, stigma, anxiety, and insecurity across the globe [10]. As a consequence, people are constantly seeking and posting updates about the pandemic on social media like *Twitter*, *Facebook* etc, creating piles of user-generated contents [11]. Researchers have proposed data visualization techniques for modeling the spread and prediction of the COVID-19 virus, analyzing the demographic data, and enhancing awareness about the pandemic [2–4, 12]. In contrast to the existing approaches, we focus on building an interactive spatio-temporal visualization tool to help epidemiologists and policymakers to better monitor COVID-19 symptoms across the globe reported through tweets.

In this demo, we focus on analyzing COVID-19 related symptoms across the globe reported through tweets by building an interactive spatio-temporal visualization tool, i.e., COVID19$^{\alpha}$. For the dataset of this demonstration, we collected around 462 million tweets between March 19, 2020, and September 15, 2020. COVID19$^{\alpha}$ consists of three main visualization components as follows:

(1) **Spatial Visualization:** This component focuses on visualizing COVID-19 symptoms across different geographic locations in an interactive fashion. COVID19$^{\alpha}$ provides two types of spatial visualizations: 1) Comparative WordCloud Visualization between two geographic locations, and 2) Multi-level granularity based spatial distribution of COVID-19 symptoms through an interactive map interface. For more details, see section 2.1.
(2) **Temporal Visualization:** This component presents the interactive time-series visualization of COVID-19 symptoms for a particular geographic location. COVID19$^{\alpha}$ provides users with options to create visualizations for a particular subset of symptoms depending on their interests allowing them to see those symptoms' evolution patterns over time (refer to section 2.2).
(3) **Spatio-Temporal Visualization:** The third and last component of COVID19$^{\alpha}$ allows users to generate visualizations, according to their choice and preference, to observe the change in symptoms' mentions across time and space jointly. Specifically, we provide two types of comparative visualizations in this case: 1) Given two geographic locations, we create dynamic visualizations for the temporal evolution of different symptoms; and 2) Given two well-known COVID-19 symptoms, we create dynamic visualizations to demonstrate their severity across different geographic locations (More details in section 2.3).

The primary goal of these visualizations is to to help epidemiologists and policymakers to better monitor and analyze COVID-19

*Both authors contributed equally to this research.

IUI '21 Companion, April 14–17, 2021, College Station, TX, USA

ACM ISBN 978-1-4503-8018-8/21/04.
https://doi.org/10.1145/3397482.3450715

related symptoms through interactive visualizations. The paper also includes some interesting insights from our experiments.

2 DEMO

Data-Set: To create the dataset, we used the data-set introduced by [9], a publicly available data-set comprising more than 462 million tweet IDs. For those tweet ID's, we fetched the text of tweets using Twarc [6], Hydrator [5] and Twitter Developer API[1] for the duration of Mar 19,2020 to Sep 15, 2020.

Data Preprocessing: For our demo, we focused only on the tweet attributes 'text', 'created_at' (time and date of post), 'user_location' (set by user), 'coordinates' (only appears in the geotagged tweets). We applied standard data cleaning and preprocessing techniques for processing the JSON data corpus. We cleaned the tweet text by removing all types of smileies, emojis, mentions, reserved words, and URLs. Furthermore, we used the tweets-preprocessor [14] module for this task as Twitter users tend to use some abbreviations that are not used in regular languages, such as 'lol', 'lmao', 'btw', 'ty'. We replaced a number of those abbreviations with proper words for a more accurate analysis.

Because only 0.04% of the collected tweets contained precise coordinates, we used the user-provided 'used_location' attributes, and the GeoPy API [8] to extract coordinates, full address, and country name for each tweet with valid 'user_location'. Furthermore, for symptom tagging, we used the symptom list provided by [13] containing 45 symptoms mentioned in the tweets posted by COVID-19 affected people. We consider a symptom mentioned in a tweet if the tweet contains any of its associated phrases. We used one-hot encoding to tag the tweets with those 45 symptoms.

2.1 Visualizing Symptoms Across Space

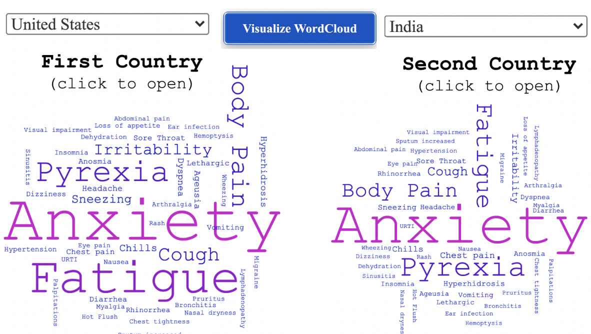


Figure 1: Comparative Word-Cloud Visualization of Symptoms between United States and India

2.1.1 **Twitter Word Cloud.** Our first spatial visualization compares the most frequent symptoms for any two given countries, chosen by the user. For this purpose, we considered all the tweets from March 19th to September 15th. After removing stop-words, we calculated the frequency of each symptom words in all tweets. Furthermore, using the most advanced amCharts [1] charting library, COVID19$^{\alpha}$ visualization tool creates the interactive word cloud side by side allowing a user to observe the differences. Figure 1 shows two juxtapositioned word clouds for two countries selected by a user. Both word clouds consist of the top 45 symptoms based on their frequencies observed in the tweets. We observed that the word clouds become a little bit different for different countries worldwide. We also noticed that *anxiety* is the most frequent symptom among all countries. Although most countries have *anxiety*, *fatigue*, and *pyrexia* as the highest frequent symptoms, we found *cough* as the most frequent symptom in *Zimbabwe*.

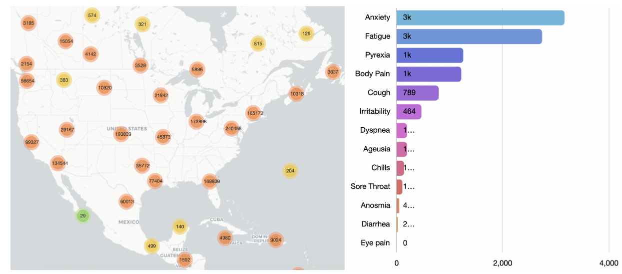


Figure 2: Clustered Symptom Map: Symptom-wise Tweet Frequency across Different Geographic Locations

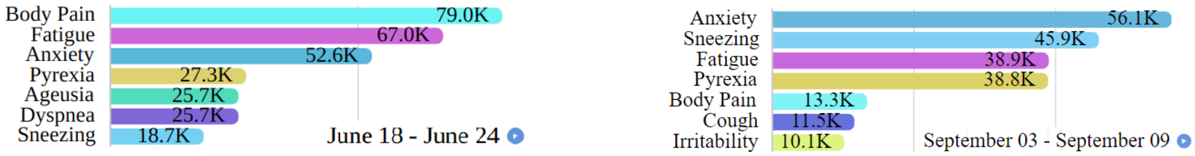
2.1.2 **Clustered Symptom Map.** Figure 2 shows our second spatial visualization—clustered symptom map. *Folium* [7], along with the *markercluster* plugin, was used to produce this map. It shows the frequency distribution of COVID-19 symptom-related tweets around the world. Tweets mentioning one of the top thirteen different symptoms and precise user locations are included in this map. We selected the top 13 symptoms in order to keep the user's cognitive load manageable. Furthermore, this map provides dynamic clustered representations of 'symptom-tweets' (tweets with at least one symptom) on different zoom levels. The clusters are divided into multiple sub-clusters or merged into a bigger cluster upon zooming in or out. This functionality enables the user to observe the distribution of symptom tweets on various locations at different granularity selected by the user, e.g., continent level, sub-continent level, country-level, or even the exact location of the tweet.

2.2 Visualizing Symptoms over Time

2.2.1 **Dynamic Temporal Simulation.** The temporal visualization uses the week-level aggregated data, to dynamically visualize the symptom frequencies hinged on a symptom list provided by [13]. With this visualization, the user can understand the evolution of worldwide popular symptoms over time. For example, anxiety, pyrexia, fatigue, cough, body pain, irritability, and sneezing are the most discussed symptoms on *Twitter* over the period of the collected data-set. In addition, we found that symptoms such as body pain, pyrexia, fatigue became widespread later in June (Refer to Figure 3a), while anxiety and sneezing became widespread in September (Refer to Figure 3b). Users are also able to observe the dynamic changes in symptom frequencies over time.

2.2.2 **Time-series Map.** Our second kind of temporal visualization shows country-level time-series analysis of tweets. Along with the change of the virus's strain, its effect or symptoms are also changing. We intend to capture this transfiguration of COVID-19 symptoms in a particular country through this visualization (Figure 4). Each location marker on the maps denotes a single country, island, or sea location. Information about country-level aggregated tweets are visualized in this map. The user can hover over the

[1]https://developer.twitter.com/en/products/twitter-api



(a) In June **(b) In September**

Figure 3: Dynamic Temporal Simulation

marker to find the total number of tweets posted about COVID-19 in that specific location. Upon clicking on a marker, a popup appears depicting the detailed time-series.

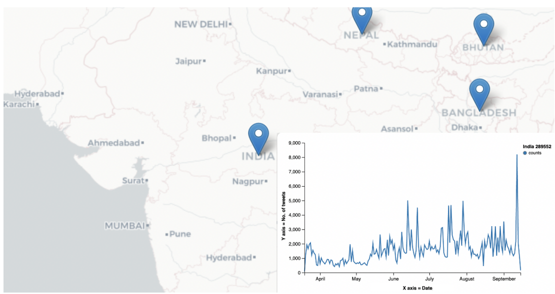


Figure 4: Time-series Map to Visualize Symptom Evolution.

We also provide two more similar time-series visualization maps with slightly different types of data. First, time-series map for only 'symptom-related' tweets instead of all COVID-19 related tweets, which shows how the number of COVID-19 symptom-related tweets varies over time. Second, a more customizable time-series map, where a user can select a set of symptoms from a drop-down menu to create a custom time-series of user's choice. It renders a clear idea about the fatality or dominance of a particular set of symptom over any time-span for a specific country.

2.3 Visualizing Evolution of Symptoms across Time and Space jointly

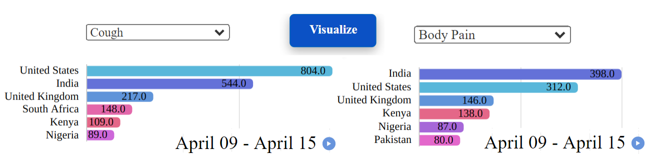


Figure 5: Visualization of Spatio-Temporal Evolution between Two Symptoms

This visualization focuses on the evolution of symptoms over time and space jointly. We created 2 types of visualizations for this purpose. 1) Given two symptoms of user's choice, the system will generate a juxtapositioned view of how these 'symptom-related' tweet counts changed over time and across geographic locations simultaneously, through dynamic bar charts (Refer to Figure 5). 2) A similar visualization like Figure 5, however, now with two geographic locations, selected by the user from a drop down menu, showing dynamic changes in symptoms over time between the two selected locations.

3 DISCUSSIONS AND CONCLUSION

We built an interactive demo tool to visualize the spatio-temporal evolution of COVID-19 related symptoms through self-reported and user-generated tweets. COVID19$^{\alpha}$ can help epidemiologists and policy makers quickly perform interesting comparative analysis through spatio-temporal visualization of symptom related tweets. For example, our visualization tool reveals that symptoms like anxiety, fatigue, pyrexia, cough, and body pain are common among different parts of the world. However, a few symptoms like irritability and sneezing are more prominent in the American subcontinent (e.g., United States, Brazil). In contrast, chest pain is more prominent in the Indian subcontinent (e.g., India, Pakistan). Additionally, although tweets on anxiety were continuously dominant in the beginning, it interchanged position with other symptoms many times in July and afterward. Beginning September 2020, we see an upward trend of sneezing, cough, and sore throat related symptom-tweets. Some of these even took place in the top five symptoms list many times. Other symptoms worth mentioning are body pain and irritability with a lot of fluctuations over time and space.

REFERENCES

[1] amcharts. [n.d.]. *amcharts4.* https://www.amcharts.com/ (Accessed on 04/10/2020).

[2] Rachel Atherton. 2021. "Missing/Unspecified": Demographic Data Visualization During the COVID-19 Pandemic. *Journal of Business and Technical Communication* 35, 1 (2021), 80–87.

[3] Baoquan Chen, Mingyi Shi, Xingyu Ni, Liangwang Ruan, Hongda Jiang, Heyuan Yao, Mengdi Wang, Zhenghua Song, Qiang Zhou, and Tong Ge. 2020. Data visualization analysis and simulation prediction for covid-19. *arXiv preprint arXiv:2002.07096* (2020).

[4] Ram A Dixit, Stephen Hurst, Katharine T Adams, Christian Boxley, Kristi Lysen-Hendershot, Sonita S Bennett, Ethan Booker, and Raj M Ratwani. 2020. Rapid development of visualization dashboards to enhance situation awareness of COVID-19 telehealth initiatives at a multihospital healthcare system. *Journal of the American Medical Informatics Association* 27, 9 (2020), 1456–1461.

[5] Documenting the Now. [n.d.]. Hydrator [Computer Software]. https://github.com/DocNow/hydrator (Accessed on 15/09/2020).

[6] Documenting the Now. [n.d.]. Twarc [A command line tool (and Python library)]. https://github.com/DocNow/twarc (Accessed on 15/09/2020).

[7] Filipe, Martin Journois, Frank, Rob Story, James Gardiner, Halfdan Rump, Andrew Bird, Antonio Lima, Joshua Cano, Juliana Leonel, Tim Sampson, Ben Welsh, Jon Reades, Oleg Komarov, Jason Baker, Qingkai Kong, odovad, Raphael Dumas, George Harris, Alex Crosby, kenmatsu4, Tales Paiva Nogueira, Nat Wilson, Daisuke Kato, andrew giessel, soymsk, Rich Signell, Justin Duke, Anand Patil, and FabeG. 2019. *python-visualization/folium: v0.9.1.* https://doi.org/10.5281/zenodo.3229045 (Accessed on 03/10/2020).

[8] GeoPy. [n.d.]. GeoPy [Geocoding library for Python]. https://github.com/geopy/geopy (Accessed on 25/09/2020).

[9] Rabindra Lamsal. 2020. Coronavirus (COVID-19) Tweets Dataset. https://doi.org/10.21227/781w-ef42

[10] Betty Pfefferbaum and Carol S North. 2020. Mental health and the Covid-19 pandemic. *New England Journal of Medicine* (2020).

[11] Alessandro Rovetta and Akshaya Srikanth Bhagavathula. 2020. Covid-19-related web search behaviors and infodemic attitudes in italy: Infodemiological study. *JMIR public health and surveillance* 6, 2 (2020), e19374.

[12] Anit N Roy, Jais Jose, Aswin Sunil, Neha Gautam, Deepa Nathalia, and Arjun Suresh. 2020. Prediction and Spread Visualization of Covid-19 Pandemic Using Machine Learning. (2020).

[13] Abeed Sarker, Sahithi Lakamana, Whitney Hogg-Bremer, Angel Xie, Mohammed Ali Al-Garadi, and Yuan-Chi Yang. 2020. Self-reported COVID-19 symptoms on Twitter: An analysis and a research resource. *medRxiv* (2020). https://doi.org/10.1101/2020.04.16.20067421

[14] vasisouv, alextsil, idimitriadis. [n.d.]. Tweets Preprocessor [Twitter preprocessor module]. https://github.com/vasisouv/tweets-preprocessor (Accessed on 01/10/2020).