Visual Analysis of COVID-19 Community Surveillance

Huan He *, Liwei Wang, Andrew Wen, Ming Huang, Yanshan Wang, and Hongfang Liu †

Division of Digital Health Research, Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA

ABSTRACT

The fast spread of the coronavirus disease 2019 (COVID-19) significantly impacts the lives of people in all regions of the globe. Timely community surveillance in this COVID-19 context is crucial for both regional disease prevention and policymaking. However, due to the large volume of the constantly generated COVID-19 data, it is challenging to track trends in specific regions. To address this challenge, we develop a visual analytic system to explore and monitor the regional trends of the COVID-19 pandemic.

Keywords: COVID-19, case doubling time, visualization.

Index Terms: Human-centered computing - Visualization - Visualization application domains - Visual analytics

1 INTRODUCTION

The fast spread of the coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), led to a worldwide pandemic affecting lives of millions of people [1]. To facilitate the COVID-19 data sharing, multiple surveillance systems have been built to help the public to obtain data and understand the situation [2, 3, 4].

One of the critical challenges faced in understanding the risk factors and distributions in the COVID-19 outbreak is the large volume of data from various surveillance systems, which poses many challenges for researchers and administrators while delivering data-driven solutions and decision making. Moreover, data is updated continuously, so that static reports have limited ability to fully reflect the changes in multiple aspects and their temporal mutual influence.

As the volume of COVID-19 data continues to grow and the factors affecting infections continue to change (e.g., shelter-at-home, social distancing, community spreading, etc.), there is a need for tracking and exploring how the COVID-19 outbreak evolves in different regions. Therefore, to address this issue, we use visual analytic methods to support community surveillance in this COVID-19 context. Our main contribution is the design and implementation of a COVID-19 tracking dashboard, which is open source and daily updated.

2 TASK ANALYSIS

Due to the rapid evolution of the COVID-19 pandemic in regions and countries, users' needs are continually changing over time. For example, in December 2019 when infections originated in Hubei province of China, the demand for relevant data came mainly for research purposes. From March 2020, the pandemic spread globally, and the number of confirmed cases in the US grew exponentially. More agencies and organizations are becoming concerned about the pandemic and the overall trend. As of 22 July 2020, more than 15 million cases have been reported across 188 countries and territories, resulting in more than 610,000 deaths [1]. At present, the researchers and the public became concerned about the COVID-19 in their local communities.

Therefore, we extracted a common task from these evolving needs: how the COVID-19 pandemic changes in different regions over time? To address this task, we propose using a visualization dashboard to show the metrics from multiple perspectives.

* e-mail: He.Huan@mayo.edu

† e-mail: Liu.Hongfang@mayo.edu

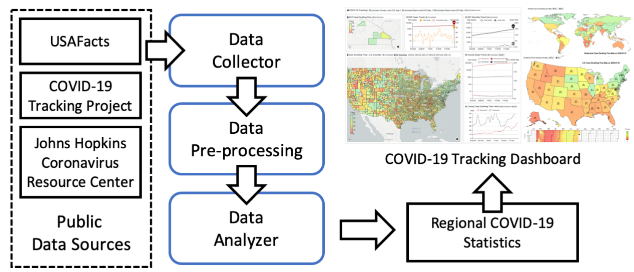

Figure 1: Data pipeline for COVID-19 tracking dashboard.

3 DATA ABSTRACTION

There are three data sources used in our process to generate visualization dashboard: (1) USAFacts, which provides county-level data on confirmed cases and deaths; (2) the COVID-19 tracking project, which provides the numbers on tests and hospitalizations from US states and territories; and (3) Johns Hopkins Coronavirus Resource Center, which provides globe-level COVID-19 information.

As shown in Fig. 1, we built a pipeline to reduce repetitive manual works. This pipeline consists of three major modules: data collector, data pre-processing, and data analyzer. First, in the data collector module, the data scraper downloads latest data from our data sources. Then, the data files are combined and fixed to resolve data issues such as data missing and duplication. Last, in the data analyzer module, we combine several Python and R packages to calculate metrics which reflect the current situation of COVID-19 in counties, regions, states and countries. The output data of this pipeline will be used in our COVID-19 tracking dashboard (source code at https://github.com/OHNLP/covid19tracking).

The COVID-19 tracking dashboard provides users with multiple coordinated interactive views to track the COVID-19 situation of communities of interest. The web interface is built on web front-end techniques, including D3.js, EChart.js and Plotly

3.1 Metrics

In addition to directly collected metrics from data sources, such as the number of cumulative cases, daily cases, deaths, N-day smoothed mortality rate, and N-day smoothed test positive rate, etc., we also use case doubling time and reproduction factor to capture short-term and long-term trends of COVID-19.

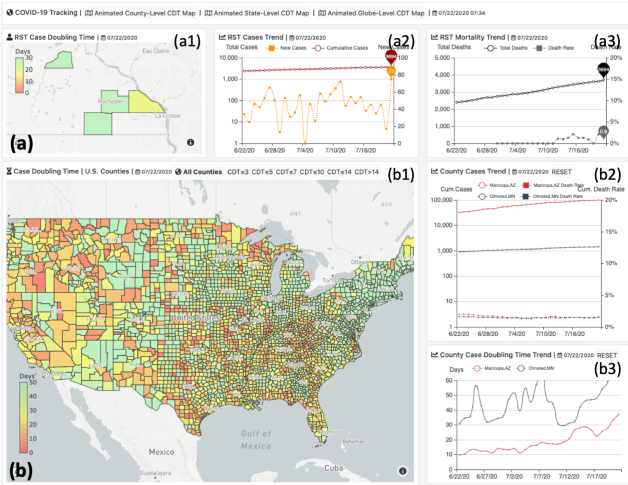

Figure 2: The screenshot of the region summary view of COVID-19 tracking dashboard that shows: (a) the metrics in user selected regions; (b) the overview of all counties and the trend of metrics.

Case doubling time (CDT). This metric measures the number of days taken for the number of coronavirus cases to double. The shorter the time frame, the steeper the curve and the faster the growth. The CDT values are calculated based on today's data comparing to the data of four days ago to provide more reliable estimation. The formula is as follows:

$$CDT_{today} = 4 \times \frac{\log(2)}{\log((N_{cases\ today} + 0.5)/N_{cases\ 4\ days\ ago})}$$

Reproduction factor (Rt). This metric describes how many people one person infects on a given day t. It let us estimate how many secondary infections are likely to occur from a single infection in a specific area. Ideally, we expect this value to be below the critical quantity of 1, which implies that the disease would gradually diminish.

4 Visualization Design

The main interface of our COVID-19 tracking dashboard consists of two modules, including a region summary view (Fig. 2) and a region trend view (Fig. 3).

The region summary view combines county-level choropleth map (Fig. 2(a1)) and line charts (Fig. 2(a2, a3)) to show the metrics in selected regions. We use a color scale to encode the CDT. The minimum, median and maximum values of CDT range from red to yellow and green respectively. Red indicates confirmed cases are increasing exponentially, while green indicates relatively lower increasing speed. Moreover, to help user to check and compare the situation in other regions, we add an overview of all counties (Fig. 2(b1)) and coordinated line plots (Fig. 2(b2, b3)). To provide flexible region observation, this view also supports region customization according to the URL parameters.

The region trend view combines a choropleth map and a calendar heatmap to show geographical and temporal changes of CDT. We design three maps with different geographic levels to show the pandemic in different regions simultaneously, including globe level (Fig. 3(a)), US state level (Fig. 3(c)) and US county level (Fig. 3(b)). To facilitate interactive exploration, the choropleth map is coordinated with the date cell in the calendar heatmap.

5 Discussion

Since the pandemic began to spread in the United States, we have been using this prototype system to observe the changes in metrics

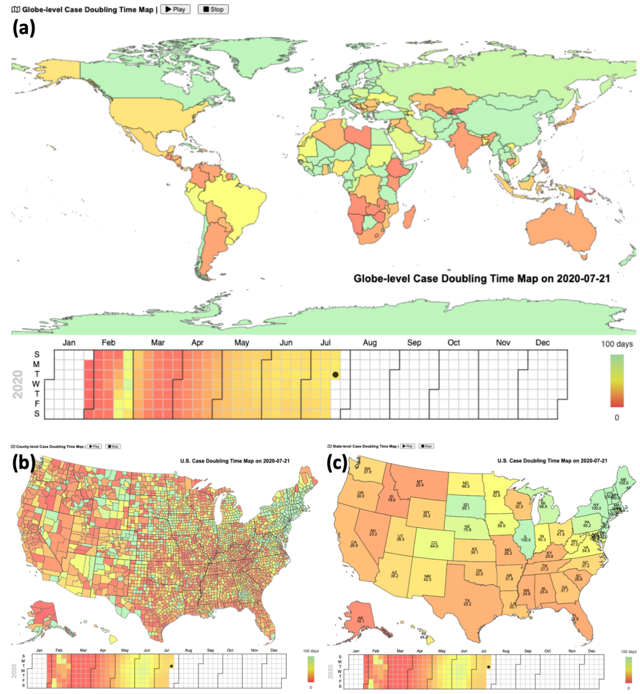

Figure 3: The screenshot of the region trend view of COVID-19 tracking dashboard that shows: (a) the global-level CDT trend map; (b) the county-level CDT trend map; and (c) the state-level trend map.

of COVID-19, such as cumulative confirmed cases, newly confirmed cases, and mortality rates. As shown in the calendar heatmap of Fig. 3(b), the number of cases in the United States increased exponentially in March, and then the situation began to moderate between April and June. However, at the end of June, the increase accelerated again.

By observing the day-by-day changes on the region trend view, we can also find the spread from a few counties to the entire country in the United States.

6 Conclusion and Future Works

In this work, we present a visualization system to facilitate COVID-19 community surveillance based on open platforms. It shows the pandemic's distribution from multiple geographical levels, which helps users monitor the trend of regional COVID-19 situation. Since the outbreak is not yet fully contained, we will continue to track, improve and expand our systems.

References

[1] E. Dong, H. Du, and L. Gardner, "An interactive web-based dashboard to track COVID-19 in real time," *Lancet Infect Dis*, vol. 20, no. 5, pp. 533–534, May 2020.

[2] B. D. Wissel *et al.*, "An interactive online dashboard for tracking COVID-19 in U.S. counties, cities, and states in real time," *J Am Med Inform Assoc*, vol. 27, no. 7, pp. 1121–1125, Jul. 2020.

[3] R. A. Dixit *et al.*, "Case Report: Rapid Development of Visualization Dashboards to Enhance Situation Awareness of COVID-19 Telehealth Initiatives at a Multi-Hospital Healthcare System," *J Am Med Inform Assoc*.

[4] S. K. Dey, Md. M. Rahman, U. R. Siddiqi, and A. Howlader, "Analyzing the epidemiological outbreak of COVID-19: A visual exploratory data analysis approach," *J Med Virol*, Mar. 2020.