Visual Analysis of Multi-scale Trends of COVID-19

Huan He * Sijia Liu † Liwei Wang ‡ Andrew Wen § Ming Huang ¶ Yanshan Wang ‖ Hongfang Liu **

Division of Digital Health Research, Department of Health Sciences Research, Mayo Clinic

ABSTRACT

The fast spread of the coronavirus disease 2019 (COVID-19) significantly impacts people's lives in all regions. Timely identifying the regional trend of COVID-19 pandemics is crucial for both local disease prevention and policymaking. However, due to the large volume of continuously generated COVID-19 data, it is challenging to capture and explore specific regions' trends. To address this challenge, we proposed new indicators to describe both the current status and trend. Moreover, we designed a trend chart for showing our proposed indicators and developed a visual analytic dashboard to track and analyze the regional COVID-19 pandemics. Based on our dashboard, we validated and identified some geographical and temporal patterns in the COVID-19 data.

Index Terms: Human-centered computing—Visualization—Visualization techniques; Human-centered computing—Visualization—Visualization systems and tools

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 [2, 20]. To facilitate the COVID-19 data analysis, many visual dashboards have been developed to help the public and researches to understand the situation and analyze the trends, such as CDC COVID Data Tracker [3] and New York Times Coronavirus Map [16].

However, as the rapid evolution of the COVID-19 pandemic, the needs for data analysis are continually changing over time. Existing dashboards are inadequate for more complex analytical needs. One of the critical needs is to capture the regional trends of the COVID-19 outbreak in the continuously updated data, which poses many challenges for researchers and administrators while delivering data-driven solutions and decision making. Basic charts and maps have limited ability to fully reflect the changes in multiple aspects. Moreover, as the regional differences in the COVID-19 pandemic increase and the factors affecting infections continue to change (e.g., executive orders, physical distancing, personal protection, etc.), there is a need for exploring and comparing how the COVID-19 outbreak evolves in different regions.

Therefore, to address these challenges, we propose to use visual analytic methods to explore the regional trends in this COVID-19 context, which include designing new chart to visualize the change and developing a visual analytic system. Moreover, we conducted case studies on the real dataset to explore both the geographical and temporal patterns of the COVID-19 for validating our design. The major contributions of this paper are as follows:

- Three new indicators for capturing the short-term and long-term trends of regional pandemics;
- A novel trend chart design for showing the trends of regional pandemics based on our proposed new indicators;
- An interactive visual dashboard which integrates our trend chart and data pipeline to explore the multi-scale trends in the COVID-19 data.

This paper is structured as follows: Section 2 summarize related work of COVID-19 indicators and visual dashboards. Section 3 analyzes the tasks and design requirements. Section 4 presents the new indicators. Seciton 5 introduces the system architecture of our visual dashboard. Section 6 demonstrates the visual design of each view. Section 7 illustrates the pattern we validated with our system. Section 8 discusses the comments from users. Section 9 conclues our work and provides insights into future work.

2 RELATED WORK

Since the beginning of the COVID-19 pandemic, there has been a rapid growth in the relevant literature, some of which is related to the overall trends and visualization of the pandemic. In this section, we first review the literature related to indicators and trend analysis on COVID-19 pandemics. Next, we review the literature related to COVID-19 dashboards and visualizations. Finally, we summarize these works and compare them with our work.

2.1 COVID-19 indicators and trend analysis

Similar to the analysis of other infectious diseases, to measure the spread of the COVID-19 epidemic, some fundamental indicators were used at the beginning of the pandemic, such as the cumulative number of confirmed cases, the cumulative number of deaths, the cumulative number of recovered, active cases, etc [6–8]. As the epidemic intensifies further around the world, more indicators are becoming known to the public, such as incidence rate, case fatality rate, positive test rate, etc [23, 25]. Based on these indicators, a variety of studies can be conducted, such as trend analysis [6], predictive modeling [19], policymaking [33], syndromic surveillance [12], etc.

Dey et al. [6] analyzed outbreak information on COVID-19 based on the several open datasets to understand the number of different cases reported in different regions and risks. Marvel et al. [18] proposed a pandemic vulnerability index (PVI) to support counties and municipalities by integrating baseline data on relevant community vulnerabilities with dynamic data on local infection rates and interventions. synthesizes county-level vulnerability indicators, enabling their comparison in regional, state, and nationwide contexts. In addition, reproduction number (Rt) is also widely used in studies [13], which describes how many people one person infects on a given day and estimates how many secondary infections are likely to occur from a single infection in a specific area.

*e-mail: He.Huan@mayo.edu
†e-mail: Liu.Sijia@mayo.edu
‡e-mail: Wang.Liwei@mayo.edu
§e-mail: Wen.Andrew@mayo.edu
¶e-mail: Huang.Ming@mayo.edu
‖e-mail: Wang.Yanshan@mayo.edu
**e-mail: Liu.Hongfang@mayo.edu

2.2 COVID-19 dashboards and visual analysis

To facilitate data exploration and keep the public informed, a large number of dashboards or visualization tools are developed and extensively used during the COVID-19 pandemic to help the public understand the pandemics, such as World Health Organization COVID-19 Dashboard [21], CDC COVID Data Tracker [3], Johns Hopkins COVID-19 dashboard [5], COVID-19 HealthMap [4], USAFacts coronavirus map [28], New York Times Coronavirus Map [16], Rt COVID-19 [25], etc.

In the academic field, there are also many studies on visual analysis of the COVID-19 data. Dong et al. [8] developed an online interactive dashboard to visualize and track reported cases of the COVID-19 in real-time. This dashboard also shared the datasets of the location and number of confirmed COVID-19 cases, deaths, and recoveries for all affected countries. Wissel et al. [30] created an online resource platform to inform the public of coronavirus disease 2019 outbreak in regions of multiple levels, including countries around the world as well as states in the United States. Bowe et al. [1] discussed the application of various visualization techniques in the analysis of COVID-19 pandemics and suggested that the creation of meaningful COVID-19 visualizations needs to consider the scales and dimensions of the pandemic. Yang et al. [32] developed a COVID-19 tracking and visualization platform that pinpoints the dynamics of the COVID-19 worldwide to show and compare the trends of COVID-19 at multi-grained levels. Ronquillo et al. [24] proposed applying biomedical informatics and data visualization tools to several public datasets to predict, manage and balance public health needs through all stages of the COVID-19 pandemic. Yang, Ang and Wang [31] designed various graphs to show the path of transmission of the virus between different patients in relation to time. Ulahannan et al [27] developed an open dashboard based on local data sources to help people in the region stay up to date on the local status of the COVID-19 pandemic.

In addition, there are some studies focusing on applying visual analysis to other aspects related to the COVID-19. The Nextstrain team [9, 11] proposed using novel visualization approach to show the spread and evolution of viral pathogens by integrating sequence data with other data types such as geographic information, serology, or host species. Thorlund et al. [26] created a dashboard of collecting all clinical trials related to the COVID-19 to help clinicians to assesse the efficacy and safety of clinical candidate interventions to treat COVID-19. Dixit et al. [7] proposed to apply a situation awareness model and user-centered design approach to rapidly develop dashboard visualizations to support COVID-19 telehealth operations for different types of users.

Most of the existing dashboards focus on showing the multi-level geographical distribution and the current status of the COVID-19 pandemic, which provide the public and researchers with current situation of the pandemics and guidance. However, these dashboards supports only a pre-define set of indicators and limited analysis of historical data, it's hard to get more insights into the trend and patterns. Compared to existing works, we expanded new analytical indicators and designed novel trend chart to show the overall trend of regional COVID-19 pandemics, which are integrated in our visual analytic system.

3 Task Analysis and Design Requirements

Due to the rapid evolution of the COVID-19 pandemic in regions and countries, the needs for data analysis are continually changing over time. In December 2019 when infections were reported, the needs for relevant data came mainly for research purposes, such as decision making [14], rapid case identification [19], and clinical care [2]. Since February 2020, the pandemic spread globally, and the number of confirmed cases grew exponentially. More agencies and organizations are becoming concerned about the pandemic and the overall trend. As of September 15 2020, more than 29 million cases

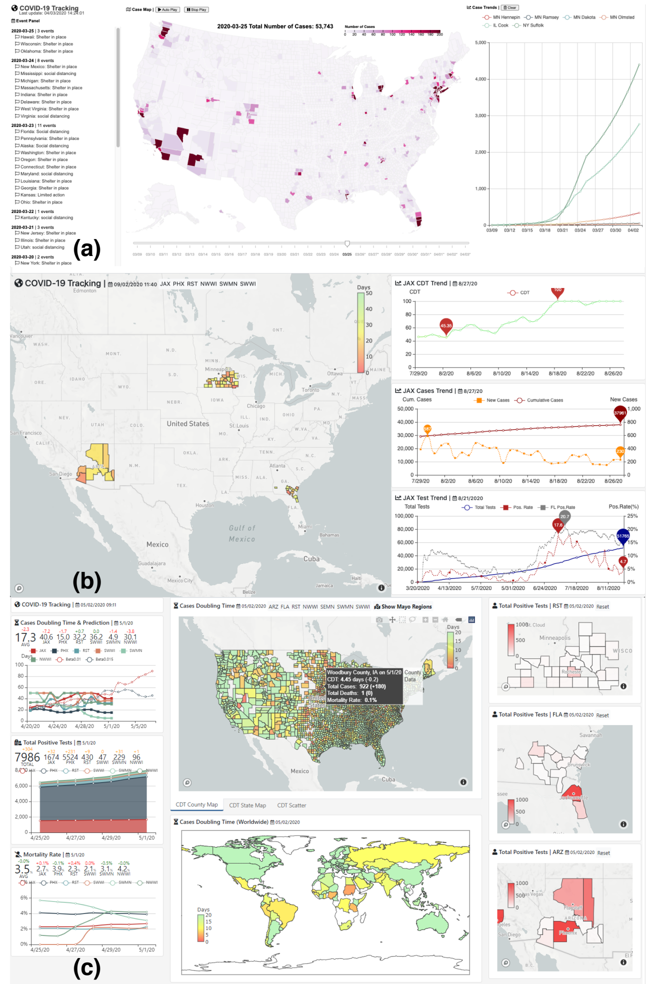

Figure 1: The prototype systems, including (a) the first prototype system showing the county-level distribution of the confirmed cases; (b) the second prototype system showing more indicators of a specific region; and (c) the third prototype system showing more regions for comparison.

have been reported across 188 countries and territories, resulting in more than 910,000 deaths [21]. As the pandemic spreads deeper into smaller regions, the public and researchers became more concerned about the potential trends of the COVID-19 spreading in their local communities.

To address this evolving needs, we collaborated with domain experts from our clinic and gathered user feedbacks to improve the design. During this process, we developed several prototype systems to meet the evolving needs at different times. As shown in Fig. 1, we designed and developed three prototype systems to validate the needs. These prototype systems were able to reflect different perspectives on the need for data analysis at the time and further helped us to understand the development of the pandemic and guided our design direction. Moreover, we built a data pipeline to automate data collection, cleansing, computation, and format conversion, which serves as a backend for each prototype system and our production.

3.1 Task analysis

In late February 2020, we began to focus on the spread of the pandemic in the United States and the policy response in each region.

Since the total number of cases was relatively small at the time, the only indicators we focused on were the total number of confirmed cases and the number of new cases per day in the United States. As shown in Fig. 1(a), we combined choropleth map and line charts to show the overall spread of COVID-19 in the United States.

Since the COVID-19 pandemic spreads further across the United States in April, counties are experiencing confirmed cases, and in some regions, the number of cases is snowballing. As a result, researchers need to get a complete picture of the outbreak status, such as the positive test rate, survival rates, and death rates in specific regions (e.g., several counties, several states). Therefore, in addition to the choropleth map, we combined multiple chart views to present the changes in multiple metrics for a given area over time (Fig. 1(b)).

As the outbreak of COVID-19 intensified further at the end of June, both the total number of confirmed cases and the number of daily confirmed cases increased to unprecedented levels. Moreover, due to various factors, the differences between regions were significantly increased (e.g., the number of confirmed cases may vary more than tenfold between states). In order to visually compare the pandemics in different regions, we designed several corresponding views, in which each view presenting the same indicator for multiple regions (Fig. 1(c)).

With the validation of these prototype systems and several rounds of communications with our domain experts, we concluded three core task requirements in these evolving needs:

- **T.1 What is the geographic distribution of the COVID-19 pandemic?** The spread of a pandemic is a gradual process that spreads in multiple ways across a region, with specific regions as the context for analyzing pandemic data. Therefore, visualizing a pandemic's geographic distribution is essential to support the study of pandemic transmission pathways, population prediction, and policymaking. This task is also confirmed during the development of our prototype systems shown in Fig. 1.

- **T.2 How does the regional COVID-19 pandemic change over time?** The pandemic's actual situation and the results presented by the prototype system reveal that the spread of the pandemic has shown specific trends over time. These trends not only reflect the state of the pandemic but may also reflect the role of local pandemic-related factors (e.g., pandemic prevention policies, special events, etc.). Presenting changes in pandemic indicators over time not only helps to monitor the pandemic status and trends in the region but also helps to analyze the potential factors influencing the pandemic.

- **T.3 What is the difference between COVID-19 outbreaks in different regions?** The state of the COVID-19 pandemic in different regions is influenced by a variety of factors such as the natural environment, social demographics, and local policies of regions. These factors make the pandemic state of different regions diverse. Even though some regions may have similar pandemic states at present, they may have different trends and influences.

3.2 Design requirements

Based on these three tasks described above and the experience gained in developing those prototype systems, we identified the following design requirements and applied them in our development:

- **R.1 Multi-scale exploration of regional COVID-19 pandemic state.** The pandemic is spreading globally, from country to country, state to state, and city to city. As a result, the data generated by this process is naturally geographically hierarchical (T.1 and T.3). To explore such hierarchical data, we need to design multi-scale maps to visualize the data at various levels and help users analyze the data at different levels.

- **R.2 Exploration of temporal trends of regional COVID-19 pandemic state.** The status of the COVID-19 pandemic is continuously changing and influenced by various factors (T.2). Our system should visualize these changes to help users understand how pandemics spread and analyze potential influencing factors.

- **R.3 Exploration among different regions.** Users need to compare different regions' pandemic status to explore potential influences and spread patterns (T.3). Our system needs to visualize these differences to help users compare pandemics in different regions.

- **R.4 Interactive exploration.** Long-term pandemic data contain multiple-level geographic information inter-correlated with each other by their geographic relations (T.1 and T.3). Our system should provide an interactive interface to check the pandemic states between regions and periods.

4 Data Abstraction

As the COVID-19 pandemic spreads, COVID-19-related data are generated continuously. In order to present and analyze a complete picture of the COVID-19 pandemic in each region, we search for relevant data sources and collecting data constantly during the development of prototype systems. At present, there are four data sources used in our system to support visual analysis:

- *The USAFacts: Coronavirus in the United States* [28], which provides county-level data of confirmed cases, deaths and population;

- *CDC COVID Data Trakcer: Pandemic Vulnerability Index* [18], which provides county-level PVI data and other demographic information;

- *The COVID-19 Tracking Project* [10], which provides the numbers on tests and hospitalizations from US states and territories;

- *Johns Hopkins Coronavirus Resource Center* [30], which provides globe-level COVID-19 information;

These data sources update COVID-19 data daily at different times of the day. To make full use of these data sources and reduce repetitive manual works, we built a data pipeline to automate data updates and perform basic data analysis. As shown in Fig. 2(a), the pipeline consists of three major steps: check for data updates, data pre-processing, and calculation of indicators. First, the data watcher module automatically detects updates of the data source at specified intervals and downloads the latest data if it has been updated. Then, in the data pre-processing module, the data files from different sources are combined by their geographic identifier (i.e., Federal Information Processing Standards), and fixed to resolve data issues such as value missing and duplication. Last, we combine several Python packages to calculate metrics which reflect the current situation of COVID-19 in different regions. The analysis results of this pipeline will be stored as JavaScript Object Notation (JSON) files, which can be used for both our COVID-19 tracking dashboard (source code at https://github.com/OHNLP/covid19tracking) and our COVID-19 map for the public.

4.1 Indicators

In addition to directly collected metrics from data sources, such as the cumulative number of confirmed cases, daily cases, and deaths, we further calculate some indicators that show the COVID-19 status, such as N-day smoothed mortality rate and N-day smoothed positive test rate.

Moreover, we propose three new indicators to capture short-term and long-term trends of COVID-19: case doubling time (CDT), 7-day smoothed average case rate per 100k capita (Cr7d100k), and the Cr7d100k ratio. These indicators provide a better way to compare the COVID-19 situation between counties, states, and countries.

4.1.1 Case doubling time

This indicator measures the number of days taken for the number of coronavirus cases to double. The CDT values are calculated based on today's data compared to four days ago to provide a more reliable estimation. The CDT of a given day d is calculated as follows:

$$CDT_d = 4 \times \frac{log(2)}{log(N_d + 0.5)/N_{d-4}} \tag{1}$$

In the above equation, N_d is the cumulative number of confirmed cases until day d, and 0.5 is added to avoid the computational error caused by $N_d = N_{d-4}$.

4.1.2 Cr7d100k

This indicator measures the increase in new cases in the last seven days and reflects the short-term trend of COVID-19. The $Cr7d100k$ values are calculated based on the daily new cases for the last seven days, which reduces the impact of the frequency of data updates. The $Cr7d100k_d$ of a given day d is calculated as follows:

$$Cr7d100k_d = \frac{1}{7} \times \frac{100,000}{Population_{region}} \times \sum_{i=d-7}^{d} n_i \tag{2}$$

In the above equation, n_i is the number of new confirmed cases in a specific region on day i.

4.1.3 RW_Cr7d100k and CrRW status

Based on $Cr7d100k$, we propose using the ratio of two $Cr7d100k$ from two adjacent weeks to measure the trend of COVID-19 in recent two weeks, namely $RW\_Cr7d100k$. This indicator reflects the long-term trend of COVID-19 in a specific region since it contains the information of recent two weeks. The $RW\_Cr7d100k_d$ of a given day d is calculated as follows:

$$RW\_Cr7d100k_d = \frac{Cr7d100k_d}{Cr7d100k_{d-7}} \tag{3}$$

By using $Cr7d100k$ and $RW\_Cr7d100k$ in combination, we define the CrRW status of a region to represent the current status of the pandemic as well as recent trends with the following thresholds:

- The GREEN status: if $Cr7d100k < 15$ and $RW\_Cr7d100k < 1$ for the past seven days, it would be on the relative safe;
- The RED status: if $Cr7d100k > 30$ or $RW\_Cr7d100k > 1$ for the past seven days, the current pandemic situation is more severe, i.e., showing a significant increase or a high number of new cases;
- The ORANGE status: the current pandemic situation is unstable, and it may turn to the GREEN status or the RED status.

5 System Overview

As shown in Fig. 2, the architecture of our proposed COVID-19 dashboard consists of four modules: data watcher, data pre-processing, data calculation, and data visualization.

To collect data from various data sources and meet automated data updates requirements, we design a flexible data watcher module to monitor data sources. This module defines a unified interface to check whether the data source is updated for the specified date and downloads the data file of a specific date. This interface is implemented for each data source. The scheduler invokes these

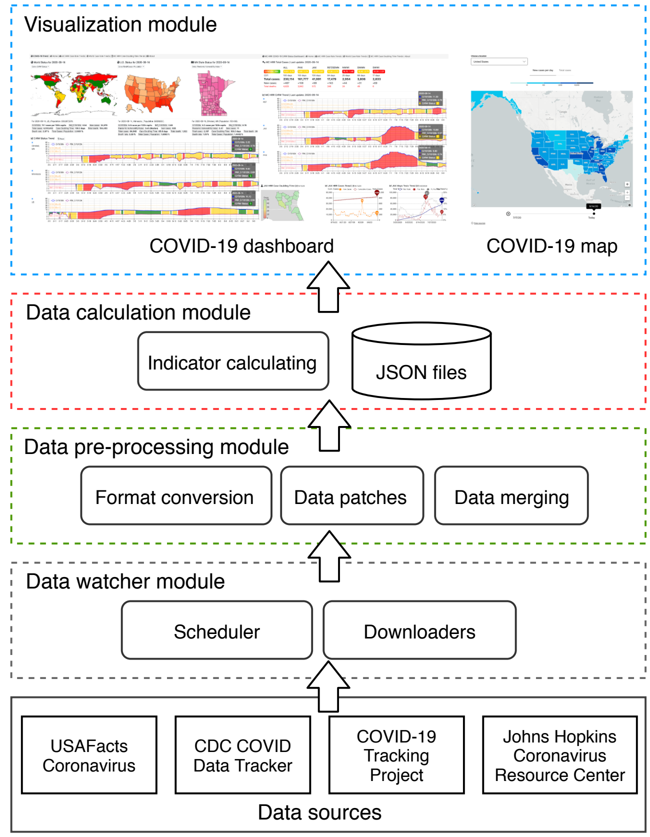

Figure 2: The architecture of our proposed visual dashboard, which consists of four modules: data watcher, data pre-processing, data calculation, and data visualization.

interfaces as scheduled. Also, the scheduler caches the checking results to avoid placing additional load on the data source.

The data preprocessing module includes several sub-modules to parse the raw data from different data sources. The data format of different data sources is different, which cannot be used directly in the subsequent calculation of indicators. For example, dates are stored in mm/dd/yy format in some data sources while yyyy-mm-dd format in others. In addition, we create a number of "data patches" for regions where data is missing or incorrect in a specific period, which could minimize the errors in calculations due to data anomalies.

In the data calculation module, we implement several functions based on Python packages such as Pandas, NumPy, and SciPy to calculate all indicators on the pre-processed data. The results of the calculations are saved as JSON files at different scales as required for visualization.

The visualization module includes two web-based applications for different users. The COVID-19 map [1] is for the public to get the latest trends in counties and states. The COVID-19 dashboard is for researchers to monitor and analyze the COVID-19 trends. To facilitate data exploration, we also publish the visualization module as static web pages on GitHub Pages [2]. The front-end of this module is implemented based on web visualization techniques, including Vue.js [29], Plotly [22], Mapbox [17] and ECharts [15].

[1] https://www.mayoclinic.org/coronavirus-covid-19/map
[2] https://ohnlp.github.io/covid19tracking/

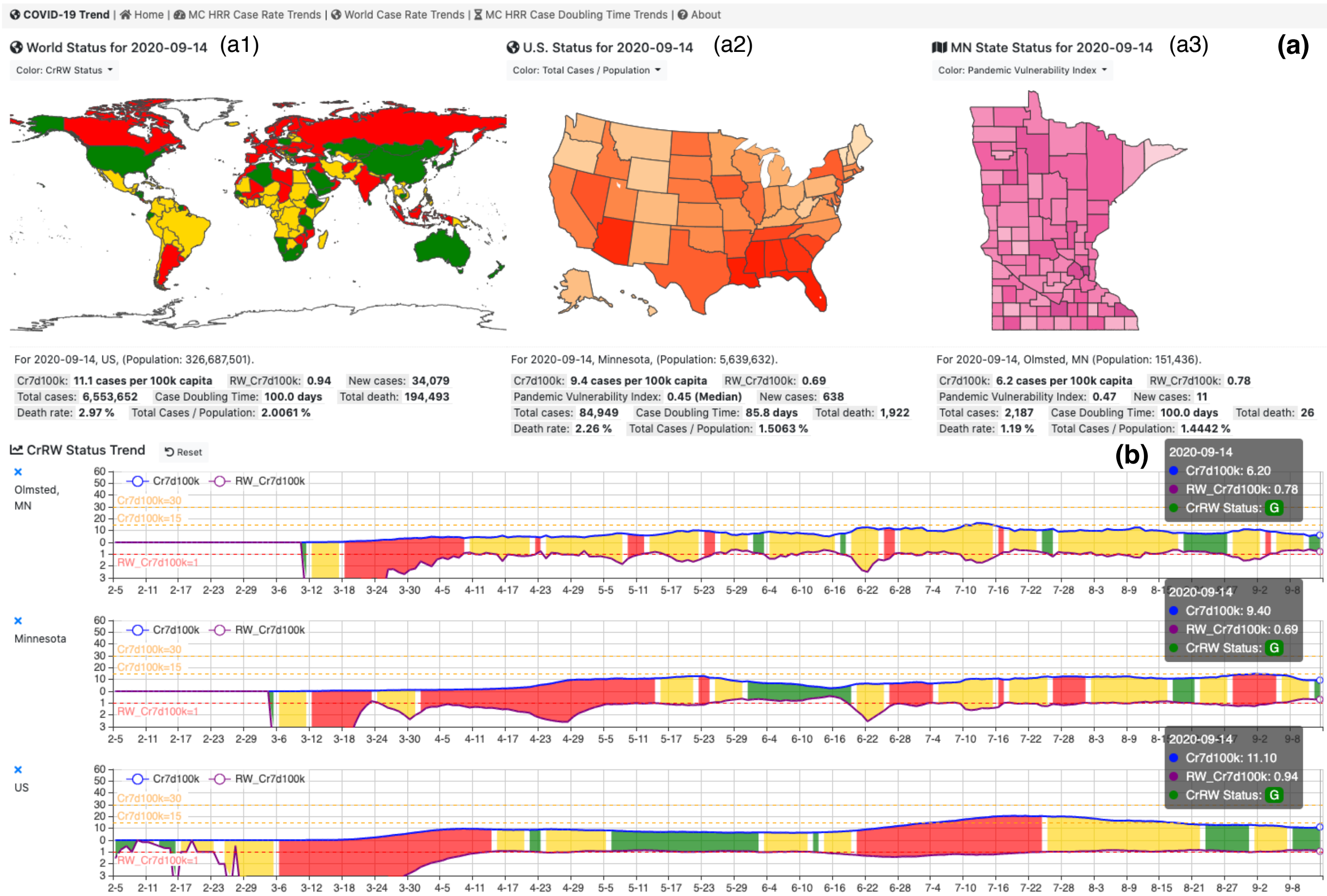

Figure 3: Demonstrating the COVID-19 dashboard shows: (a) the map view showing the multi-scale geographical distribution of pandemics in specific color-encodings; (b) a trend view depicting the overall trend of pandemics in the selected regions.

6 Visual Design

This section illustrates the visual design of our COVID-19 dashboard and the interaction among the views. As shown in Fig. 3, there are two coordinated views for exploring the indicators of the COVID-19 pandemic from different perspectives: map view and trend view. Besides, to help users quickly view the regions of interest, we design several tabs to switch to a specific region summary view (Fig. 4).

6.1 Map View

As discussed in the above task analysis in Section 3.1, to help users check the geographic distribution of the COVID-19 pandemic (T.1), we design three sub-views to show the COVID-19 indicators on maps of different levels in this view (Fig. 3(a)). Based on the data available and the need for analysis, COVID-19 outbreaks are shown in this view at the country (Fig. 3(a1)), state (Fig. 3(a2)), and county (Fig. 3(a3)) levels, respectively (R.1). Similar to the design used in other studies [24, 27], we also used choropleth maps to present the geographic distribution of the COVID-19.

We design different color encoding for different indicators so that users can identify different indicators on each map easier. For example, as shown in Fig. 3(a1), the CrRW status is displayed in three colors, as defined in Section 4.1.3. In addition, the detailed COVID-19 information of the selected region is shown under the map when users click on each region.

In addition to the basic map view, we combine the choropleth map and a calendar heatmap to show geographical and temporal changes of CDT (Fig. 5) at different geographic levels. The choropleth map is linked with the date cell in the calendar heatmap to provide better interactive exploration.

6.2 Trend View

The spread of the COVID-19 is a long-term process. To show the trends of the pandemic (T.2), we design a trend chart to show our proposed indicators simultaneously overtime (R.2). As shown in Fig. 3(b), each trend chart shows the daily values of 3 indicators over a period of time: $Cr7d100k$, $RW\_Cr7d100k$, and CrRW status. The $Cr7d100k$ is shown as a solid blue line in the upper area of the chart, while the $RW\_Cr7d100k$ is the solid purple line in the lower area. The area between two lines is filled according to the CrRW status on each day. To help identify the CrRW status, there are 3 dashed lines as references in the chart, representing $Cr7d100k = 15$, $Cr7d100k = 30$, and $RW\_Cr7d100k = 1$. In this chart, it can be easily seen that these two lines form a band, and the fill color in this band is related to its height. When the height of the band is small for a continuous period, it appears green. When the band's height increases continuously and exceeds the corresponding threshold value, it appears yellow or red.

When the mouse hovers a chart, the detailed indicator values for that region are displayed. In the meantime, the detailed indicator values of all charts on the same day automatically appear to make it easier to compare different regions (R.3 and R.4).

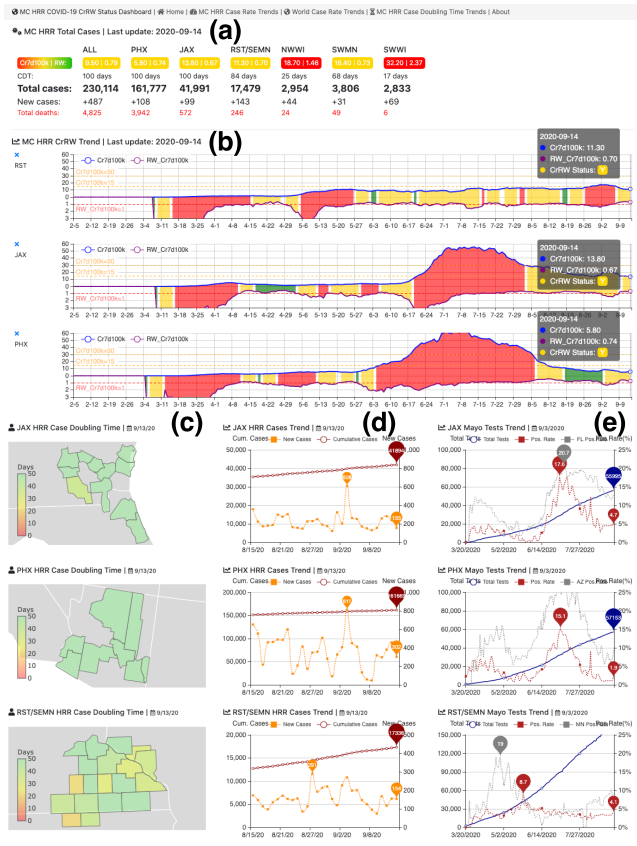

Figure 4: The screenshot of the region summary view that includes: (a) a region summary panel showing the latest values of key indicators; (b) a trend view showing the trend of $Cr7d100k$, $RW\_Cr7d100k$, and CrRW status; (c) a map view showing the geographic distribution of the specific region; (d) a case chart showing the number of confirmed cases; and (e) a positive test rate chart showing the values of test-related indicators.

6.3 Region summary view

In routine data analysis, users need to check the current COVID-19 situation of the regions of interest (T.1 and T.3) frequently, so we design this view as a separate tab integrating previous views to present the COVID-19 status and trends in multiple specified regions (R.3). This view consists of five parts, including a region summary panel (Fig. 4(a)), trend view of specific regions (Fig. 4(b)), map view of specific regions (Fig. 4(c)), case charts of specific regions (Fig. 4(e)), and positive test rate charts of specific regions (Fig. 4(e)).

The region summary panel (Fig. 4(a)) shows the latest values of main indicators in each region, such as $Cr7d100$, CDT, total cases, etc.; the trend view shows the values of $Cr7d100k$, $RW\_Cr7d100k$, and CrRW status over time, which is the same as described above; the map view shows the geographic distribution of the COVID-19 of the specific region on a county-level choropleth map; the case chart shows the cumulative number of confirmed cases and the number of daily new cases; the positive test rate chart shows the number of tests in a specific region, the positive test rate of both specific regions, and the reference region.

7 Case Study

As the COVID-19 pandemic spreads, we kept evaluating our visual designs and the effectiveness of our dashboard, including the prototype systems and the latest one. All of our systems are deployed in our clinic's intranet and can be accessed by our clinicians and researchers. In addition, the public version of the visualization module is uploaded to the GitHub pages and updated daily. We validated and identified some COVID-19 trends and patterns with our domain experts during this period.

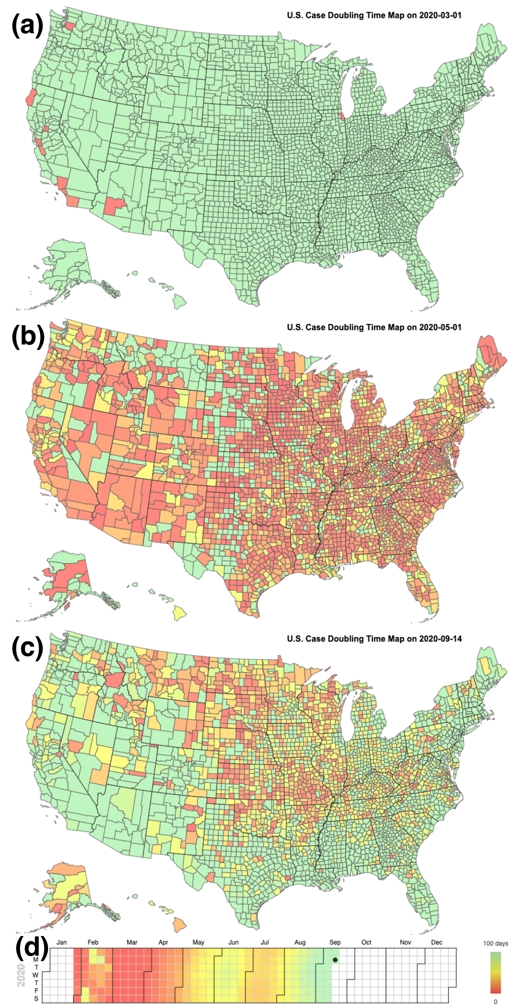

Figure 5: The geographical distributions of the COVID-19 in the United States over time, which includes: (a) the outbreaks in a few counties on March 1; (b) the outbreaks in most of the counties on May 1; (c) the outbreaks begin in some counties in the Midwest on September 14; and (d) a calendar heatmap for selecting the date for display.

7.1 Geographical distribution of the COVID-19

To help users understand the spread and distribution of the COVID-19 pandemic between regions, a calendar heatmap (Fig. 5(d)) is linked the map view is to play an animation by date. As shown in Fig. 5, each subfigure presents a CDT distribution of by county for the United States on March 1 (Fig. 5(a)), May 1 (Fig. 5(b)), and September 14 (Fig. 5(c)), respectively.

The Fig. 5(a) shows that at the beginning of the pandemic, the number of cases increased rapidly in only a few counties on the West Coast of the U.S., with no cases in most areas. Two months later, cases were appearing in most counties and growing rapidly (Fig. 5(b)). Four months later, the number of cases was no longer exponential in most counties, except for some central and northern counties (Fig. 5(c)). However, we also found that higher CDT values do not guarantee that regions are completely safe. In these high CDT

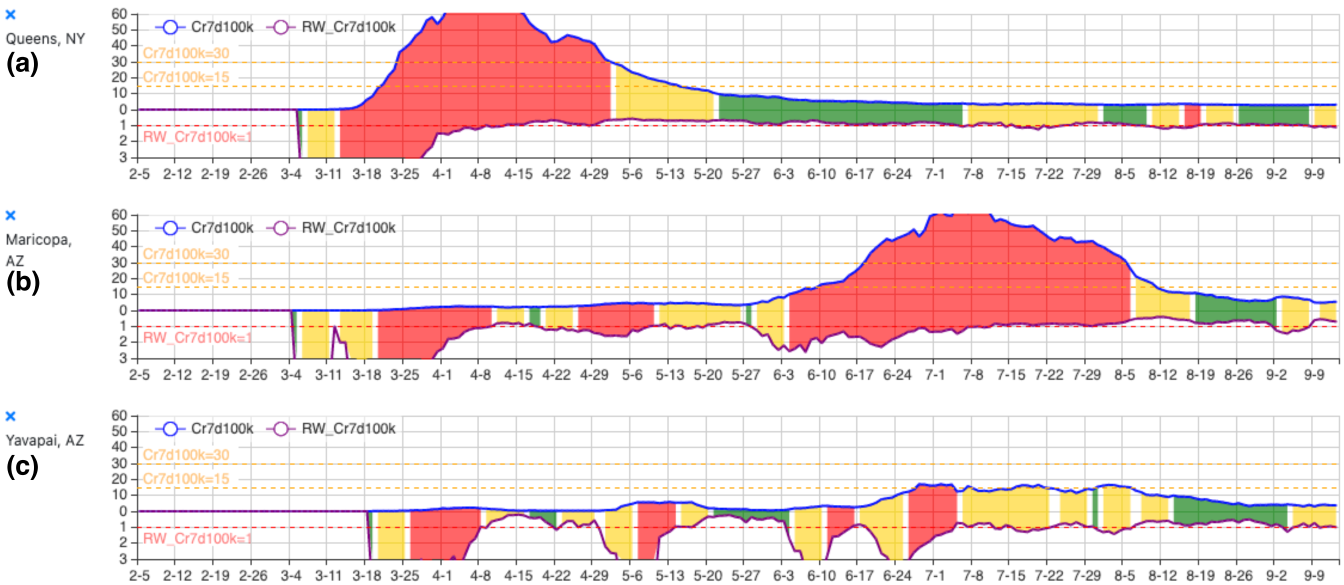

Figure 6: Typical cases of COVID-19 pandemic temporal patterns, which includes (a) a peak occurring at the beginning of the pandemic; (b) a peak occurring after a period of steady state; and (c) an overall stable pandemic state, which is in a neighboring county in the same state.

areas, the number of new cases per day may still be high but not exponentially increasing.

7.2 Temporal patterns of the COVID-19

With the trend view in our dashboard, we found that the COVID-19 pandemic may show similar temporal patterns but different characteristics in each region.

First, regions with severe outbreaks show a prominent peak that exceeds the thresholds ($Cr7d100k$=30), and this peak usually lasts for 1 to 2 months followed by a flattening trend (e.g., the peak during March to April in Fig. 6(a) and the peak during June to July in Fig. 6(b)). Although each peak is different in height and width (i.e., the max number of daily confirmed cases and the duration of the outbreak), the peak itself indicates that the COVID-19 outbreak in that region is spreading quickly and that the number of daily new confirmed cases is high, requiring increased protective measures. Besides, the peak may be followed by a gradual transition to a relatively stable status. These peaks can be further analyzed concerning other factors such as executive orders, physical distancing, and personal protective measures by combining with other data sources such as social media and news.

Second, even within the same area, sub-area COVID-19 pandemics can vary significantly. As shown in Fig. 6(b and c), two neighboring counties in Arizona state show the different shapes of bands between May and August. Although the band in Fig. 6(c) also shows an increasing trend, it is significantly smaller than that in Fig. 6(b). This phenomenon could also be seen in other states, indicating that regional factors can significantly influence regional COVID-19 pandemics. Therefore, it may imply that the COVID-19 pandemic can be controlled through improved regional pandemic control measures.

8 Discussion

Since the COVID-19 pandemic began to spread, we have been developing and improving our dashboard to monitor global and regional COVID-19 pandemics changes. In this process, we collected some comments and summarized them as follows.

Firstly, selecting appropriate indicators is essential to describe the current status of the COVID-19 pandemic accurately. The status of the COVID-19 pandemic varies tenfold or even a hundredfold from region to region. There are counties with tens of thousands of cases, while some counties with only hundreds. As a result, it is challenging to visualize these indicators, such as the cumulative number of confirmed cases, making visual distinctions difficult. The CDT is sensitive to regional changes in the pandemic at an early stage, but as the number of confirmed cases increases to a certain level, CDT is not sensitive to the same number of new cases as before. Therefore, we could use CDT to identify those regions where the number of cases is likely to snowball. When the number of cases is large, the CDT can be used as a secondary reference indicator. Using $Cr7d100k$ and $RW\_Cr7d100k$ could reduce the impact of population size and short-term anomalies (e.g., data missing and wrong values), but it is necessary to set appropriate thresholds according to the region's population size.

Second, the combination of multiple indicators in visualizing the COVID-19 pandemic is more useful for data analysis. Each indicator has its requirements and reflects a particular characteristic of the pandemic. Presenting a combination of indicators helps users gain a more comprehensive understanding of the pandemic's status and trends in a region. However, presenting multiple indicators in one chart also challenges a user's understanding. It is hard to track too many different variables on one chart. Our designed trend chart provides an intuitive visual representation of regional pandemic changes for researchers, but we need to pay attention to avoid information overload when adding more data to this chart.

Third, the COVID-19 pandemic and the needs of its data analysis change over time, posing the requirement of including other data sources in the visual analysis. In the current data analysis process, all of our data came from public data sources, from the county-level data to the country-level data. As a result, the outcome of the analysis obtained is also at the corresponding levels. While our clinic is also generating a large amount of fine-grained data, how to visualize these data to support the local COVID-19 pandemic analysis should also be further investigated.

9 Conclusion and Future Work

In this work, we present a COVID-19 visual dashboard to facilitate COVID-19 data analysis and community surveillance based on open data sources. We proposed several new indicators to capture the regional trends of COVID-19 pandemics. Based on these new indicators and our data pipeline, we create multiple views to show the geographical distributions and temporal trends of regional COVID-19 pandemics, which helps users explore and analyze the regional COVID-19 situation. We validated and identified some COVID-19 trends and patterns with our domain experts with the supports from our prototypes and the current dashboard.

As the pandemic continues to spread, we will keep tracking and improving our dashboard. In the future, we plan to extend our dashboard with other data sources. For example, we consider exploring the relationship between the regional socio-demographic factors and the trend of COVID-19 spreading to analyze the general patterns and provide supports for trend prediction and policymaking.

Acknowledgments

Research reported in this publication was supported by the National Center for Advancing Translational Science of the National Institutes of Health under award number U01TR002062. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

[1] E. Bowe, E. Simmons, and S. Mattern. Learning from lines: Critical COVID data visualizations and the quarantine quotidian. *Big Data & Society*, 7(2):2053951720939236, July 2020. Publisher: SAGE Publications Ltd.

[2] J. Budd, B. S. Miller, E. M. Manning, V. Lampos, M. Zhuang, M. Edelstein, G. Rees, V. C. Emery, M. M. Stevens, N. Keegan, M. J. Short, D. Pillay, E. Manley, I. J. Cox, D. Heymann, A. M. Johnson, and R. A. McKendry. Digital technologies in the public-health response to COVID-19. *Nature Medicine*, 26(8):1183–1192, Aug. 2020. Number: 8 Publisher: Nature Publishing Group.

[3] CDC. Coronavirus Disease 2019 (COVID-19) in the U.S. `https://www.cdc.gov/covid-data-tracker`, Mar. 2020.

[4] CDC. HealthMap | COVID-19. `https://www.healthmap.org/covid-19/`, 2020.

[5] J. H. C. R. Center. COVID-19 Map. `https://coronavirus.jhu.edu/map.html`, 2020.

[6] S. K. Dey, M. M. Rahman, U. R. Siddiqi, and A. Howlader. Analyzing the epidemiological outbreak of COVID-19: A visual exploratory data analysis approach. *Journal of Medical Virology*, Mar. 2020.

[7] R. A. Dixit, S. Hurst, K. T. Adams, C. Boxley, K. Lysen-Hendershot, S. S. Bennett, E. Booker, and R. M. Ratwani. 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*, 2020.

[8] E. Dong, H. Du, and L. Gardner. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet. Infectious Diseases*, 20(5):533–534, May 2020.

[9] G. epidemiology of novel coronavirus. Genomic epidemiology of novel coronavirus. `https://nextstrain.org/ncov/global`.

[10] T. A. M. Group. The covid tracking project. `https://covidtracking.com/`, 2020.

[11] J. Hadfield, C. Megill, S. M. Bell, J. Huddleston, B. Potter, C. Callender, P. Sagulenko, T. Bedford, and R. A. Neher. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*, 34(23):4121–4123, Dec. 2018. Publisher: Oxford Academic.

[12] N. K. Ibrahim. Epidemiologic surveillance for controlling Covid-19 pandemic: types, challenges and implications. *Journal of Infection and Public Health*, Aug. 2020.

[13] A. J. Kucharski, T. W. Russell, C. Diamond, Y. Liu, J. Edmunds, S. Funk, R. M. Eggo, F. Sun, M. Jit, J. D. Munday, N. Davies, A. Gimma, K. van Zandvoort, H. Gibbs, J. Hellewell, C. I. Jarvis, S. Clifford, B. J. Quilty, N. I. Bosse, S. Abbott, P. Klepac, and S. Flasche. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The Lancet Infectious Diseases*, 20(5):553–558, May 2020.

[14] K. Kupferschmidt and J. Cohen. Can China's COVID-19 strategy work elsewhere? *Science*, 367(6482):1061–1062, Mar. 2020. Publisher: American Association for the Advancement of Science Section: In Depth.

[15] D. Li, H. Mei, Y. Shen, S. Su, W. Zhang, J. Wang, M. Zu, and W. Chen. ECharts: A declarative framework for rapid construction of web-based visualization. *Visual Informatics*, 2(2):136–146, June 2018.

[16] N. Y. T. C. W. Map. New york times covid world map. `https://www.nytimes.com/interactive/2020/world/coronavirus-maps.html`.

[17] Mapbox. `https://www.mapbox.com/`.

[18] S. W. Marvel, J. S. House, M. Wheeler, K. Song, Y. Zhou, F. A. Wright, W. A. Chiu, I. Rusyn, A. Motsinger-Reif, and D. M. Reif. The COVID-19 Pandemic Vulnerability Index (PVI) Dashboard: Monitoring county-level vulnerability using visualization, statistical modeling, and machine learning. *medRxiv*, p. 2020.08.10.20169649, Sept. 2020. Publisher: Cold Spring Harbor Laboratory Press.

[19] C. Menni, A. M. Valdes, M. B. Freidin, C. H. Sudre, L. H. Nguyen, D. A. Drew, S. Ganesh, T. Varsavsky, M. J. Cardoso, J. S. El-Sayed Moustafa, A. Visconti, P. Hysi, R. C. E. Bowyer, M. Mangino, M. Falchi, J. Wolf, S. Ourselin, A. T. Chan, C. J. Steves, and T. D. Spector. Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nature Medicine*, 26(7):1037–1040, July 2020. Number: 7 Publisher: Nature Publishing Group.

[20] R. Muhareb and R. Giacaman. Tracking COVID-19 responsibly. *The Lancet*, 0(0), Mar. 2020. Publisher: Elsevier.

[21] W. H. Organization. WHO Coronavirus Disease (COVID-19) Dashboard. `https://covid19.who.int/`.

[22] Plotly. `https://plotly.com/`.

[23] D. D. Rajgor, M. H. Lee, S. Archuleta, N. Bagdasarian, and S. C. Quek. The many estimates of the COVID-19 case fatality rate. *The Lancet. Infectious Diseases*, 20(7):776–777, July 2020.

[24] J. G. Ronquillo, W. T. Lester, and D. M. Zuckerman. Using informatics to guide public health policy during the COVID-19 pandemic in the USA. *Journal of Public Health*.

[25] K. Systrom, T. Vladek, and M. Krieger. Project title. `https://github.com/rtcovidlive/covid-model`, 2020.

[26] K. Thorlund, L. Dron, J. Park, G. Hsu, J. I. Forrest, and E. J. Mills. A real-time dashboard of clinical trials for COVID-19. *The Lancet Digital Health*, 2(6):e286–e287, June 2020. Publisher: Elsevier.

[27] J. P. Ulahannan, N. Narayanan, N. Thalhath, P. Prabhakaran, S. Chaliyeduth, S. P. Suresh, M. Mohammed, R. E, S. Joseph, A. Balakrishnan, J. Uthaman, M. Karingamadathil, S. T. Thomas, U. Sureshkumar, S. Balan, and N. N. Vellichirammal. A citizen science initiative for open data and visualization of COVID-19 outbreak in Kerala, India. *medRxiv*, p. 2020.05.13.20092510, July 2020. Publisher: Cold Spring Harbor Laboratory Press.

[28] USAFacts. Usafacts: Coronavirus stats and data. `https://usafacts.org/issues/coronavirus/`, 2020.

[29] Vue.js. `https://vuejs.org/`.

[30] B. D. Wissel, P. J. Van Camp, M. Kouril, C. Weis, T. A. Glauser, P. S. White, I. S. Kohane, and J. W. Dexheimer. An interactive online dashboard for tracking COVID-19 in U.S. counties, cities, and states in real time. *Journal of the American Medical Informatics Association*, 27(7):1121–1125, July 2020. Publisher: Oxford Academic.

[31] L. Yang, H. C. Ang, and S. Wang. Customized visualizations for audience-oriented communication in covid-19 tracking story. `https://osf.io/mvkx7`, Aug 2020.

[32] Z. Yang, J. Xu, Z. Pan, and F. Jin. COVID19 Tracking: An Interactive Tracking, Visualizing and Analyzing Platform. *arXiv:2008.04285 [cs]*, Aug. 2020. arXiv: 2008.04285.

[33] Z. Yang, Z. Zeng, K. Wang, S.-S. Wong, W. Liang, M. Zanin, P. Liu, X. Cao, Z. Gao, Z. Mai, J. Liang, X. Liu, S. Li, Y. Li, F. Ye, W. Guan, Y. Yang, F. Li, S. Luo, Y. Xie, B. Liu, Z. Wang, S. Zhang, Y. Wang, N. Zhong, and J. He. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *Journal of Thoracic Disease*, 12(3):165–174, Mar. 2020.