Visualizing and Communicating COVID-19 Data Effectively

A Capstone Report
presented to the faculty of the
School of Engineering and Applied Science
University of Virginia

by

Evan Bernard

with

Eddie Moder
Matthew Hoffman

November 9, 2020

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

Evan Bernard

Capstone advisor: Yanjun Qi, Department of Computer Science

Visualizing and Communicating COVID-19 Data Effectively

Matthew Hoffman
Department of Computer Science
University of Virginia
Charlottesville, VA, United States
mfh5uw@virginia.edu

Evan Bernard
Department of Computer Science
University of Virginia
Charlottesville, VA, United States
emb2ug@virginia.edu

Eddie Moder
Department of Computer Science
University of Virginia
Charlottesville, VA, United States
ecm6yd@virginia.edu

ABSTRACT

The COVID-19 pandemic has dramatically affected all US citizens, and is front page news every day, but keeping up to date on relevant information in one's area can still be difficult. Providing the public with efficient access to current data, especially in areas not typically covered by national media, will allow individuals to make informed decisions on a daily basis.

Some applications have sought to achieve this by providing public access to general COVID-19 statistics for specific cities and states. However, encouraging users to actively engage with the data has typically fallen by the wayside. Our project seeks to remedy this issue by giving users the ability to receive notifications when specified statistical thresholds are met in their region, or to opt for digest messages providing tailored pandemic updates based on user preferences. The system will also allow users to visualize and explore relevant data via an interactive map.

Increased awareness of area-specific data will directly contribute to public safety as the pandemic continues.

INTRODUCTION

In the expanding field of computer science, many professions have turned to digital technology as a means of automation, statistical analysis, and increased workplace efficiency. Unsurprisingly, the healthcare industry has been at the forefront of this push; hospital systems have sought to implement software for error mitigation, data interpretation, and diagnostics[1]. During the global COVID-19 pandemic of 2020, improved access to computer technologies may have aided in the public's understanding and containment of the virus[2]. However, not all web applications for public health purposes are inherently valuable. Data must be engaging and succinct, otherwise a large proportion of the population may not have the capacity to effectively understand it[3]. Our proposed web application seeks to bridge demands for data access and comprehension by designing an online data visualization and notification platform for COVID-19 statistics.

1 Extracting Meaning from Data

For public data distribution to effectively support the containment of a virus, developers must first consider the technical literacy of their audience. Evidence has suggested that much of the adult populace struggles with information and communication technology (ICT) skills, which may inhibit one's ability to effectively find and decipher information online[3]. Thus, the chief error of many health departments is the exhibition of data in strictly tabular form. These impersonal visualizations "carry little meaning without proper context and framing", as users often cannot "interpret, relate, and organize data"[4]. Instead, our group will propose an interactive user interface centered around a heat map of United States COVID-19 metrics. The user experience flow would begin by offering breakdowns at national and state levels, to provide as specific and relevant data as possible to each distinct user. Users then have the option to explore a variety of statistics, including cases, deaths, and recovery rates. Finally, a customizable text or email notification system would enable users to receive updates when new thresholds are met, or when updates are released by local authorities. The consolidation of simple data display tools with individualized notifications seeks to establish a dependable alert system for users, especially those who may not actively follow news coverage of the outbreak or who struggle with ICT. The data itself will be primarily sourced from *The Atlantic*, which has compiled a public repository of free COVID-19 statistics[5].

Ideally, with our proposal enacted, individual citizens will be better informed of COVID-19 trends and guidelines and be equipped to make safe choices for themselves and their families. Recent evidence lends credence to this expectation; consider Germany, where regional health officials have tailored "national guidelines and recommendations to local needs", to great effect[6]. Conveying these instructions to residents is crucial for their health, and our notification system may provide the missing link.

Perhaps a more impactful long-term goal of the project is future adaptation for subsequent waves of COVID-19 or novel pandemics. Amaryllis Mavragani, a faculty member of the Department of Computing Sciences and Mathematics at the University of Sterling, expressed confidence that "web-based data [may] inform public health and policy" moving forward. By facilitating comprehension and engagement with relevant data, regional management of epidemics becomes straightforwardly forecasted and handled[7].

BACKGROUND

For this project, a prototype of the proposed application was developed using the Django web framework. Python, HTML, CSS, and JavaScript were used to develop the application, along with a SQLite database. The Google Maps API was used to create the application's heat map, which uses data retrieved from *The Atlantic*'s public API. The prototype was deployed on Amazon Web Services, using the EC2 service with an autoscaling group. Utilizing a cloud host allows for flexibility and scalability to be built into the deployment, which enables the application to handle increased traffic. The Twilio API was used in order to send notifications to users programmatically.

RELATED WORK

When viruses rapidly spread across a population, it is imperative that modern technology informs citizens swiftly about potential dangers and regulations. Currently, countries most effectively dealing with COVID-19 outbreaks, such as Australia and New Zealand, have relied heavily on digital public reporting and visualization tools. These methods have succeeded by presenting information in an easily digestible format to the public[8]. Furthermore, Kelly Servick, a cognitive scientist from *Science* magazine, suggested that even tools as simple as mobile applications have been lauded by worldwide health officials as means of halting the spread of COVID-19[2].

The World Health Organization provides a similar heat map implementation as that proposed in this project. Many other pandemic dashboards exist, including those maintained by state governments. However, these instances lack the notification capacity outlined in our paper. For example, the Virginia Department of Health's COVID-19 dashboard simply displays numerical values for several statistical categories, but does not offer opportunities for interaction or personalization. Adding notification capabilities would eliminate some of the friction preventing users from staying informed about the pandemic, and would increase user interaction with important data that directly impacts public health.

SYSTEM DESIGN

1 Tool Architecture

This system involves several key pieces. First, the data is gathered by periodically querying *The Atlantic*'s API, requesting updated statistics for each U.S. state. This information is stored locally for future access in order to improve performance and reduce the number of API queries that are performed. Once the updated data is available, it is processed, and the values of several statistics are extracted for each state. Then, the heat map is updated to reflect those new values.

A web form is available for users to sign up to receive notifications from the application. When this form is filled out, user preferences are stored in a database for future use. When new data is received from *The Atlantic*'s API, the stored user preferences are checked to determine whether messages should be sent or not. Whether a message is sent or not depends on the values of the new data, and the specifics of what each user has chosen.

The application is deployed on AWS EC2 as part of an autoscaling group. This allows the number of instances running the application to automatically increase as traffic increases. When demand goes back down, some instances will automatically be shut down in order to avoid wasting resources. This also prevents extended downtime for the application; if an instance shuts down unexpectedly, another will automatically start up to replace it. As a result, our application will be highly available. The instance(s) used to deploy the application are part of a security group which restricts access in order to secure the application and its data. The security group exposes only the minimal number of ports necessary, and requires an SSH key in order to access the instances.

At a larger scale, the application could be deployed across multiple AWS availability zones, and could also make use of a content delivery network such as AWS CloudFront in order to reduce latency and increase performance.

2 Risk Analysis

This tool relies on data from third party sources, meaning that the quality of data provided to the application can be viewed as a limiting factor. However, the main goal of the application is to maximize the potential of that existing data through effective presentation and communication. Additionally, since the application will store personal information such as email addresses and phone numbers, security is an important concern. As previously mentioned, built-in AWS security measures are used in order to protect the application. Finally, performance of the application is important. A slow, unresponsive application would likely result in less usage and engagement from users. The efficiency of displaying the heat map is the most important consideration, as that is the dominant user-facing piece of the application. Preprocessing and locally storing the data which will be displayed on the map will contribute to improved performance, compared to retrieving the data on demand and performing computation on it in real time. The performance of the process which sends messages to users is not as critical, as this is a background job. However, it is still desirable for this piece of the application to be as efficient as possible, so as to not waste computing resources.

3 Broader Impact

The overarching goal of this application is to contribute to public health by better informing the population. However, the tool can only be effective if it can be used. Thus, the issue of access to technology and the internet is a concern. If members of the public do not have internet access, or are unable to use a computer or mobile phone, they will be unable to access the application. As a result, these individuals would not be positively impacted by the tool. Additionally, groups will only be able to use this tool to remain informed about the status of the pandemic in their area if data is accurately collected for that area in the first place.

PROCEDURE

The application allows users to view a COVID-19 heat map, which shows a map of the United States and represents the value of certain statistics across the country. The map allows users to toggle between several statistics, such as new positive tests and current hospitalizations, and choose which one to display.

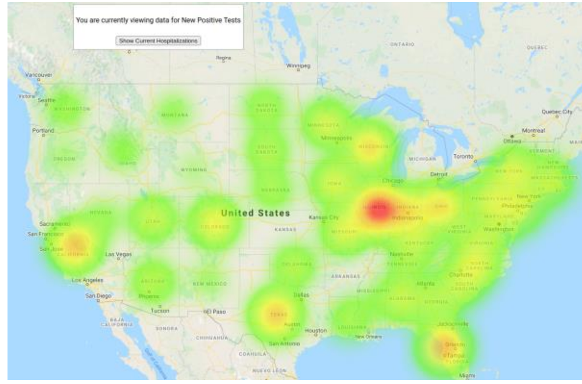

Figure 1: Heap map displaying new positive COVID-19 tests at a state level.

The application also includes a web form, through which users can sign up to receive notification messages. Users can select whether they would like to receive a text message or an email, and then provide the appropriate contact information. Users also choose which state's data they are interested in, and which statistic they wish to track. Finally, the form allows users to choose whether they would like to receive daily messages regarding the selected statistic, or only receive a message when that statistic reaches a specified value. If the latter is chosen, users indicate which value they would like to use as a threshold. These user preferences are stored in a database. Periodically, when new information is retrieved from *The Atlantic*'s API, the stored user preferences are checked to determine whether a message should be sent or not. If users have opted to receive daily digests for their chosen statistic, a message is always sent. Otherwise, a message will only be sent if the statistic the user has chosen has indeed crossed the user-specified threshold.

When the user receives a message, it will contain several pieces of information. The message will indicate that it is a notification from our application, and will include the date on which it is being sent. Next, it will inform the user what the value of the statistic they have chosen to receive is and what state the value pertains to. If this is not a digest message, meaning that the statistic has just crossed a threshold specified by the user, this will be stated in the message as well.

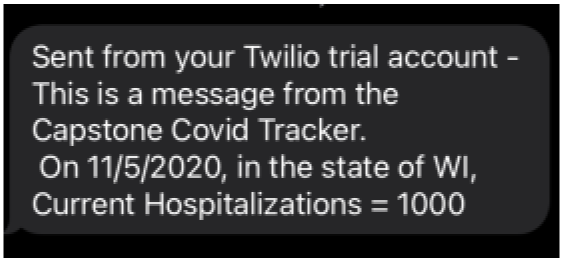

Figure 2: Example personalized notification featuring COVID-19 hospitalization data.

RESULTS

Users responded positively to the ability to receive notifications containing relevant pandemic information based on their preferences. After the user is signed up to receive notifications, no further action is required of them in order to remain informed of constantly changing information. This saves the user time, and increases the chances of them keeping up to date on pandemic information in their area.

While the concept of a heat map is not a new idea, offering this map in the same place where users sign up to receive notifications reduces the amount of time and clicks necessary for users to perform both of these actions. Users agreed that it was convenient to be able to accomplish these tasks in one place.

CONCLUSIONS

The intent of this project was to provide everyday users with the ability to interact with important statistics regarding the spread of the novel coronavirus during a global public health crisis. The interactive heat map and related notification system accomplish this goal by reducing friction between users and information. Public interaction with COVID-19 data has always been the mission of this tool, and hosting it publicly on AWS EC2 is an important step in accomplishing this goal.

By developing a system in which users receive messages about data in an area, we allow greater interaction with said data in order to promote an overall increase in public safety. This system also eliminates the need for users to repeatedly search for data in which they may have a vested interest, again reducing friction between the end user and the information that they seek. As awareness of the potency of this deadly pandemic increases, so too will the health of the general public. For this reason, the notification system can be regarded as the most important function of this tool.

Global strife and hardship are not new concepts to the human race, but the increasing capabilities of modern technology arm us with the knowledge to fight back against deadly diseases and pandemics. As engineers, it is our responsibility to continue to use new and emerging technologies to develop tools, such as the one outlined in this paper, for the betterment of society and public safety.

FUTURE WORK

Given the nature of the CS Capstone curriculum for the Fall 2020 term, there are numerous features that could still be added to our proposed tool if more time and resources were allocated to its development. The goal of this tool is to inform the general public about the spread of the coronavirus and potentially save lives. Adding features to make this tool more impactful, accessible, and accurate is paramount to its success.

First and foremost, it would be extremely beneficial to have the ability to reconcile COVID-19 data from several sources with one another in order to provide the most accurate information possible. Sources such as *The Atlantic*'s COVID Tracking Project and the World Health Organization can provide such data, as well as new and emerging APIs that seek to provide the public with access to comprehensive statistics about the coronavirus[5,9]. Using these sources in conjunction, the American people would have the ability to make the most informed decisions for their safety and the safety of their loved ones.

While the data currently available is useful, it can be hard for individuals to relate state-based data and statistics to themselves or their specific region. As data collection and testing methods improve with regards to the coronavirus, it would be useful to include more statistical data points in this tool, and to include them for smaller partitions of the country. The data currently available to the public is based on states alone, but we would hope to implement a method of narrowing statistics down to include regions, counties, and even specific cities within each state. Users would then be able to zoom in or toggle areas of the map in order to indicate regions of the country about which they would like to receive information. Data related to mortality rates, positivity rates, total cases, total deaths, and more could be added, and a feature could even be implemented to notify users when a new government ordinance has been passed that affects their selected area of interest. These statistics are not only useful, but are necessary to employ in an effort to inform the public of the dangers of COVID-19 and its rapid spread across the country.

The United States has suffered greatly during this pandemic but it is important to note that, by definition, countries all around the world are facing severe hardships due to the spread of the coronavirus as well. This tool has the potential to save countless lives worldwide, and its use should not be restricted to the United States. Many countries do not have the resources or the capability to report COVID-19 data accurately on a large scale, but future versions of this tool should still seek to include as many countries as possible in its reporting. According to the BBC, Italy saw record highs in new cases of COVID-19 in October, registering 10,925 and 11,705 new cases on consecutive days[10]. This brings the total number of recorded cases in Italy to 414,000 since the pandemic began, with 36,500 deaths having been recorded in total[10]. France similarly saw record numbers of new cases as recently as October 17, recording 32,427 in a single day[10]. This new surge of positive cases comes after an initial onslaught of cases in early 2020, but has since exceeded the severity of the first wave of cases. According to Cohen, "The level of confirmed infections is presently much higher than in March and April in many European countries"[11]. In addition to this, Cohen explains that the rise in positivity rates "suggests the intensity of the spread of the contagion is strengthening"[11]. The worldwide surge in new cases of COVID-19 is a clear indicator that this tool is needed in far more areas of the world than just the United States.

As horrific as the results of the coronavirus pandemic have been, this is likely not the last pandemic the world will face. For this reason, our tool can and should be applied to the spread of other dangerous diseases going forward, not just COVID-19. If history has taught us anything, it is that humanity is often dangerously underprepared for the consequences of an epidemic- or pandemic-level threat. The outbreak of severe acute respiratory syndrome (SARS) in 2003 was a precursor to the current pandemic, since both are caused by forms of the coronavirus[12]. In terms of scale, however, the current coronavirus pandemic is most commonly compared to the Influenza pandemic of 1918. The CDC estimates that roughly 500 million people were infected, resulting in about 50 million deaths[13]. The extreme severity of such a pandemic is sufficient evidence for the necessity of a tool to inform and protect the public with regards to highly infectious and deadly diseases. As worldwide awareness of communicable diseases increases, so too should the availability of technologies designed to monitor and mitigate their spread.

Finally, as part of our goal to make this tool accessible to as many people as possible and to keep the public informed about the data and statistics surrounding the spread of the novel coronavirus, we would hope to deploy a mobile version of this application in the future. According to Kelly Servick of *Science* magazine, "Recent modeling by infectious disease epidemiologist Christopher Fraser… predicted that if about 56% of a population used an app, it alone could reduce the virus' reproduction number—how many people catch the virus from each infected person—enough to control the outbreak," regarding the use of phone apps to disseminate information about the coronavirus[2]. In the face of such statistics, a mobile version of the application is elevated from possibility to necessity. This would allow people to receive updates about statistics in their area more frequently, it would allow for even greater interaction between users and data, and it would increase the amount of traffic to the site, since over 52% of internet traffic comes from mobile devices[14].

REFERENCES

[1] Eric J. Topol, 2019. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 25, 1, (2020), 44-56. DOI: https://doi.org/10.1038/s41591-018-0300-7
[2] Kelly Servick, 2020. Can phone apps slow the spread of the Coronavirus? *Science*, 368, 6497, (2020), 1296-1297. DOI: https://doi.org/10.1126/science.368.6497.1296
[3] Irvin R. Katz, 2005. Beyond technical competence: Literacy in information and communication technology. *Educational Technology* 45, 6 (2005), 44-47.
[4] Jofish Kaye (Ed.), 2016. Proceeding of the 2016 CHI Conference on Human Factors in Computing Systems. *ACM Press*. DOI: https://doi.org/10.1145/2858036
[5] The Atlantic, 2020. The COVID Tracking Project. *The Atlantic*.
[6] Lothar Wieler, Ute Rexroth, and René Gottschalk, 2020. Emerging COVID-19 success story: Germany's strong enabling environment. *Our World in Data* (June 30, 2020).
[7] Amaryllis Mavragani, 2020. Tracking COVID-19 in Europe: Infodemiology approach. *JIMR Public Health and Surveillance*, 6, 2, (2020). DOI: http://doi.org/10.2196/18941
[8] Sheree Lloyd, Sue M. Walker, and Ani Goswami, 2020. Health information: Applications and challenges in the COVID-19 pandemic. *Asia Pacific Journal of Medicine*, 15, 3, (July 29, 2020), 23-28. DOI: https://doi.org/10.24083/apjhm.v15i3.473
[9] The World Health Organization, 2020. The global health observatory. *WHO*.
[10] The BBC, 2020. COVID-19: Italy tights rules after coronavirus cases surge. *BBC* (October 19, 2020).
[11] Joshua Cohen, 2020. Second wave of coronavirus intensified across Europe [update]. *Forbes* (October, 2020).
[12] Centers for Disease Control and Prevention, 2017. SARS basic fact sheet. *CDC* (December 6, 2017).
[13] Centers for Disease Control and Prevention, 2018. History of 1918 flu pandemic. *CDC* (March 21, 2018).
[14] Eric Kazda, 2018. 5 key ways to grow mobile website traffic. *Quantum Dynamix*. (February 28, 2018).