Joshua Shelley
Joshua Divaker
Lyssa Buissereth
Neha Cherian
Michael Russo

ECES 450/650 Final Report

Abstract

Since the start of year 2020, COVID-19 has taken the world by storm. The resulting coronavirus disease rooted from the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has managed to impose a death toll of over 250,000 and infected millions worldwide to date. As these figures continue to rise, the ability to detect and track COVID-19 becomes increasingly imperative to understanding how the virus is spreading and transitioning.[1] Because the virus is transmitted between humans (rather than an animal or parasite), tracking individuals who have become exposed or infected have become vital to the efforts to mitigate and respond to outbreaks. To aid those tracking these outbreaks, their scale, the populations they affect, and so forth, thousands of researchers around the world are compiling records for the infected and uploading them to online databases. One such database, GISAID, has made its content freely-available to researchers to analyze and visualize, the latter of which is the focus of this project. To assist in visualizing the trends of outbreaks, we have built a website which accesses the database and overlays the positional data of those infected upon a world map. As a preprocessing step, entropy-based analysis is implemented and yields mutational differences in SARS-CoV-2 sequences. In doing so, sites of high variation are characterized between the viral genomes of affected individuals. Informative Subtype Markers (ISMs)

---

[1] Jin, B. (2020, March 16). *Live tracker: How many coronavirus cases have been reported in each U.S. state.* Retrieved from Politico :
https://www.politico.com/interactives/2020/coronavirus-testing-by-state-chart-of-new-cases/

characterize these sites of variation, serving as a useful tool in tracking the spread of a particular strain geographically and over time.

As a secondary processing step, the positional data is binned around epicenters and merged into a single statistic which represents all those infected in the region. Through binning, we are able to reduce the amount of data points on screen at any given time, which makes the information more comprehensible and simplifies the user interface so that clicking on a marker will give a summary statistic. Said summary statistic will aim to provide both a pie chart, to emphasize genetic variability and a time series chart, to help visualize which strains may be the most infectious.

We were able to create a site that is able to accept user input for geographic locations or regions of interest and display markers which represent the infected in each country for which there is collected data. There were some shortcomings in the plotting of the data, but this was largely due to a lack of time to develop robust solutions. The site, if it is ever to become public and sees further development, will perhaps become a valuable tool for providing visitors with an easy-to-use snapshot of the impact of the outbreak for their region.

Background/Literature

Interactive COVID-19 data visualization is a fundamental tool as it not only informs the public health and research communities but also easily conveys relevant SARS-CoV-2 metadata to a wider audience, whose cooperation and understanding of the situation is essential to mitigation of spread. Development of such a system was done by Berry, et al,[2]

---

[2] Isha Berry, Jean-Paul R. Soucy, Ashleigh Tuite, and David Fisman. "Open Access Epidemiologic Data and an Interactive Dashboard to Monitor the COVID-19 Outbreak in Canada." CMAJ. CMAJ, April 14, 2020. https://www.cmaj.ca/content/192/15/E420.

wherein an open access epidemiological database and interactive dashboard were developed to help monitor the SARS-CoV-2 outbreak in Canada. Said epidemiological database consisted of individual-level case data including location, report date, travel history, exposure source, time series deaths, and recoveries.[2] This data is updated daily into publicly available .csv and Google sheets files and are fed into an interactive dashboard, enabling users to follow the outbreak. Data was sourced from both government health authorities and commercial media sources.[2] This data is entered manually and tries to corroborate all entries against the Public Health Agency of Canada. Berry et al. claims this method is more timely and detailed than federally aggregated data.[2] The interactive dashboard was developed with shinyapps.io, a protected R development environment.[3] An example of the database visualization from Berry, et al. is included in *Fig. 1* below. This data sharing and visualization tool has been highly implemented, informing further evidence based control strategies and aiding Canada's public health response.
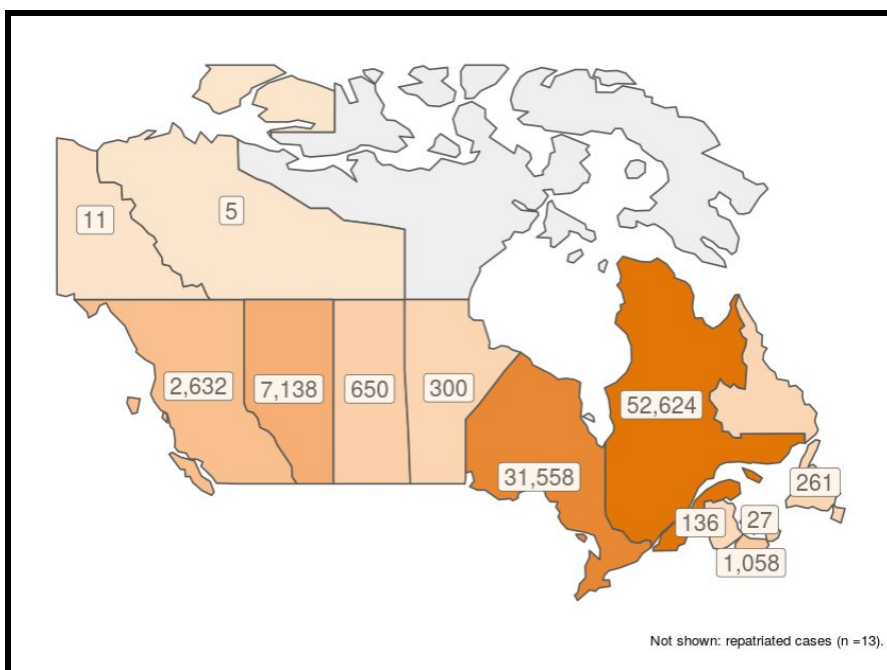
---

[3] RStudio. *Shinyapps.io*, www.shinyapps.io/.

Figure 1. Reported cases for both confirmed and presumptive positive cases.[1]

The map of Canada in *Fig. 1* lists total confirmed and presumptive cases per each providence. Aside from this map, more information for each province is supplemented within the website. For example, additional graphs show trends for daily cases and deaths per each providence over time on a logarithmic scale, similar to our time series maps. Supplementary graphs included in their website, and not in our own, include time series trends for reported cases, mortality, recovery, testing, and a global travel map. These informative maps should be considered in further iterations of our web interface.

The outbreak, which was first thought to be pneumonia, spread to every mainland in China and dozens of counties/regions around the world. As occurrences increasingly dominated countries, it was clear the world needed rapid methods to trace the transmission of the

COVID-19 and consequencing cases. The second interface to be evaluated was developed by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. The community has created an interactive web-based dashboard to track reported cases of COVID-19 in real time as shown below in *Figure 2*:
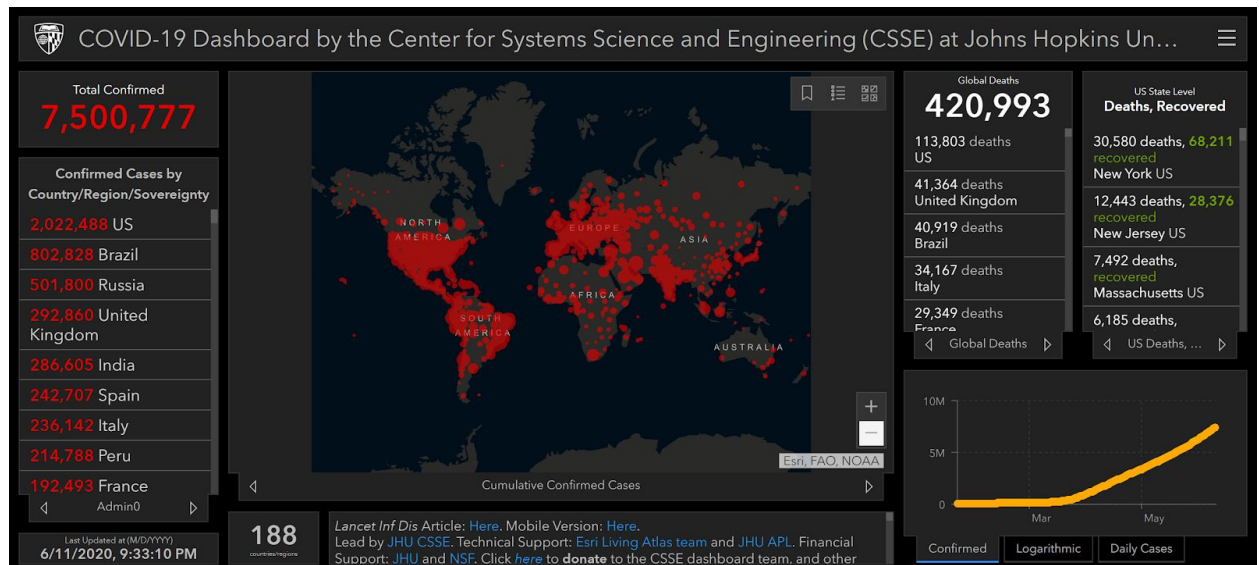

Figure 2. COVID-19 web-based dashboard by CSSE of Johns Hopkins University.

The dashboard displays the number of deaths, confirmed cases and recoveries at each location it lists. The dashboard is able to report said information at the province level in China, at the city level in the United States, Australia and Canada, and at the country level for the remaining included countries. To accomplish coverage at the province level in China and country level otherwise, the online platform DXY is used as the primary data source. DXY is run by members of the Chinese medical community to provide updated occurrences near real time (every 15 minutes). Because DXY cannot keep up with the speed of COVID-19 prevalence, manual entry is used to supplement its lag. However, manual entry being required also means

additional time is needed for quality/accuracy check. Manual entry cases are acquired via Twitter feeds, online news services, etc., and double checked local and regional health departments. As for the United States, Australia and Canada, data is provided by the CDC, the Australian Government Department of Health and the government of Canada. The data used can now be found on GitHub; at first, the interface manually updated the dataset twice a day but because of the increase in volume of cases, a semi-automated living data stream strategy has been employed.

The interactive dashboard has proven to be so efficient that, with the exception of Australia, Hong Kong and Italy, it has been able to describe new COVID-19 cases before WHO. And in the cases of Australia, Hong Kong and Italy, the CSSE interface is able to reach the accuracy of WHO's report within a few hours' time span. Precise recording of epidemiological information, geological spread and transmission rates are crucial for experts to have access to. Without such reports, strains and mutations of strains cannot be traced. Without the detailed information, the unique characteristics of the virus cannot be attained. And without this thorough analysis treatments and preventative measures cannot be created. [4]

A separate strategy was taken in developing our mapped visualization tool. Instead of corroborating infection data, information was initially taken from GISAIDto gain a better understanding of variability in COVID-19 genotypic strains. To process the different genetic strains, GISAID data was processed by Rosen et al,[5] wherein entropy-based analysis was utilized to identify different subtypes of COVID-19. This analysis allows for the interpretation of genotypic variation among specific nucleotides of COVID-19 within different individuals.

---

[4] University, J. H. (2020). An interactive web-based dashboard to track COVID-19 in real time. The Lancet , 534-44.
[5] Rosen, Gail. "EESI/ISM." GitHub, June 3, 2020. https://github.com/EESI/ISM.

Genetically different COVID-19 sequences are mapped per geographical location and are displayed in both a pie chart, to emphasize genetic variability and a time series chart, to help visualize which strains may be the most infectious. Our webpage aims to implement more interactivity by displaying subsequent graphs when the user clicks on them. This grants users a better understanding of the diversity, infectiousness, and spread of particular strains. Below, *Fig.3* illustrates a conceptual overview of our proposed interactive map, while *Fig.4* displays an overall timeline on what milestones must be completed on a weekly basis.
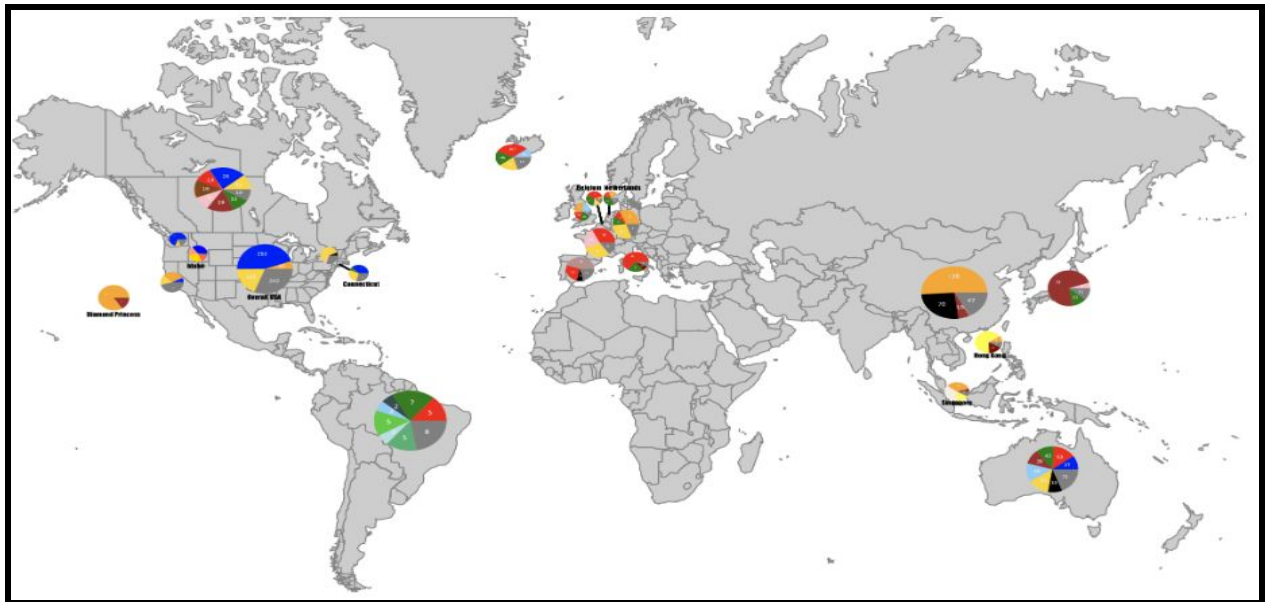

Figure 3. Conceptual overview of our proposed interactive map with corresponding pie charts.[3]

Week 4:-	Week 5:-	Week 6	Week 7	Week 8	Week 9
• Discuss potential topics with Dr. Rosen	• Obtain remote server from Drexel IT • Develop environment	• Gather missing data and latitude and longitude • Create Python dictionary	• Visualisation using geojson and augmented Mapbox Js • Start binning different genetic sequence • Label with colors	• Pop-up animation • Use django to pull genetic sequnece from 10 day interval	• Final project Presenetation

Figure 4. Week-to-week overview of project achievements and goals.

Both aforementioned literature methodologies have advantages and disadvantages in their design and function. Berry et al's larger database is advantageous as it allows for a greater analysis of additional variables, like recovery time or mortality, and subsequently allows for their time series graph's tracking of these numerous variables. To maintain this functionality, the large dataset is updated and unkept by their team. This logistically is not entirely feasible given the time and resources of our project. The CSSE developed dashboard suffers a similar disadvantage in which because the vast datasets cannot be processed quickly enough through DXY, manual entry is required for supplementation. At the same token, the interface provides narrow statistics on a city level (in certain cases). Our visualization map's key advantage is the convenient interactivity of clickable locations and display of the according pie chart, along with time series charts of different strains. This feature can make understanding the map easier to a

wider audience. Both of these advantages and disadvantages should be considered for both

methodologies as they can potentially improve the quality of output for each site. General direct

comparisons among these web-based strategies are mapped below in *Table 1*:

Table 1. Comparison Among Mentioned Interface Approaches			
	Canadian Interface	John Hopkins Interface	Our Interface
Dataset	Publicly available	Publicly available	Publicly available
Data Collection and Processing	Completely manual	Semi-automated (DXY) and manual	GISAID, python dictionary
Coverage	Canada - province level	China – province level USA, Australia, Canada – city level Others – country level	City level
Limitation	Less interactivity	Lag in data updates thus requiring manual entry	Limited dataset
Advantage	Large dataset with corresponding graphs	Accuracy	Interactivity
Future work	Interactivity within map	Establish a standing tool to monitor and report on future outbreaks	Automated python script to update large datasets; incorporation of ISMs

These interfaces are imperative for public safety and wellness. The everyday citizen

does not hold expertise in statistics, disease transmission or epidemiology. But many have

access to internet-accessing technology. Therefore it is the responsibility for those with this

knowledge to create tools to translate the information. Informative models such as graphs and tables make sense of the data in a manner it can be digested and understood widely. The more the population is informed, the better the situation may be managed. And as well early public health decision-making can be done early enough to prevent as much detriment as possible. Going forward, all informative tools must be increasingly optimized to be able to withstand the massive load of cases at an appropriate speed as time and spread continues. The speed of prevalence requires frequent monitoring; iif not the data is quickly rendered outdated and thereby inapplicable, which loses its impact to be useful in the outbreak's resolution.

Materials & Methods

Materials

For this project, a variety of resources were used to create the website. Aside from the immaterial "materials" of knowledge of web development and coding, there was of course the GISAID database which was accessed to gather COVID-19 infection data and metadata. Access to this database necessitated (1) a verification that our team members are students of a university and (2) an agreement that we will act in good faith to protect any sensitive information and (3) that the information to which we will be granted access cannot be shared in its entirety given copyright restrictions. Upon completion of the signup and verification steps, our team was granted full access to the COVID-19 databases hosted by GISAID. In addition, access was needed to gain access to the EpiCoV databases which log metadata that corresponds to COVID-19 infections logged in the GISAID databases.

Finally, our team utilized characteristic Informative Subtype Markers (ISMs) provided by Dr. Gail Rosen and Zhengqiao Chen which are used to understand the geographical bounds of

different strains of COVID-19. This genetic analysis tool is able to catch and categorize even the smallest of genetic difference among strains amid the dataset. These categories refer to strains with common entropy and can therefore be identified as a variation of COVID-19. In labeling these categories, the ISM allows for visual modelling of prevalence in relation to each strain. The current state of the interface does not integrate the ISM's, however is pending for future work.[6]

As for physical materials, our project required a server to host the site and manage the information interchange between the users and the COVID-19 data. The server is hosted by Drexel University, and needs to be accessed via VPN.

Methods

Because our project was a website, it had multiple deliverables with individual design requirements and "methods" by which we met these requirements. The site's main components can be understood as three unique interdependent constituents: the user interface, the intermediary exchanger, and the database interface. The user interface consists of the components of the site which the users can interact with, including the elements that accept user input and display the output of their queries. The intermediary exchanger is named this because it exists between the two interfaces (the user interface and database interface) and converts user inputs into a format for database queries and converts database values into a format which can be displayed in the user interface (including the binning of metadata). Finally, the database interface takes the output from the intermediary exchanger, queries the GISAID database with this information, then accesses all relevant metadata, which is passed back to the

[6] Zhao, Z., Sokhansanj, B., & Rosen, D. G. (2020). Characterizing geographical and temporal dynamics of novel coronavirus SARS-CoV-2 using informative subtype markers. BioRxiv, 1-32.

intermediary exchanger. A graphical representation of the project is presented below in *figure* as a flowchart schematic which shows how information is passed between the many components of the system.
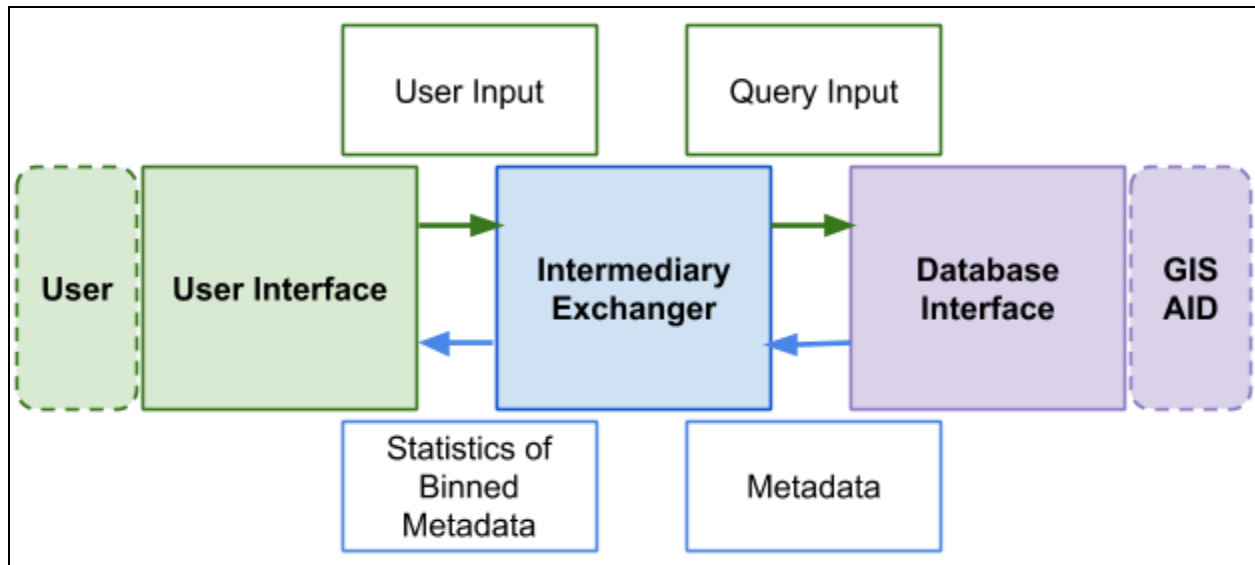

Figure 5. Map of designed graphical functionality.

User Interface

The user interface had to feature an interactable map so the user can pan and zoom to get a better view of their area of interest. It also necessitated the ability to translate mouse click coordinates into latitude and longitude values which would be the user's input sent to the intermediary exchanger. This is accomplished by a combination of MapBox's map-making API's to record the values in the geojson format. The MapBox map itself is written in NodeJS and integrated into the site using Javascript. The returned information from the intermediary exchanger is handled by MapBox and goes on to be plotted on the interactive map. As for the

site which hosts the user interface, it is coded with a variety of languages including HTML, CSS, JS.

Intermediary Exchanger

The intermediary exchanger is the backbone of our project as it allows for the communication between the user and the COVID-19 data and converts that data into a presentable manner. Its first task is taking the coordinate inputs from the user, which are in a geojson format, and passing it to the database interface so they can serve as search queries. It's second function is to, of course, take the output from the database interface, which is the infection data and metadata that was retrieved upon the user's query, and pass it back to the MapBox to plot on the interactive map for the user. During this information handling stage, the virus strains in the information to be plotted are categorized by entropy using characteristic Informative Subtype Markers (ISMs) to better understand the populations of the COVID-19 virus.

Database Interface / Website Vision

The database interface is the component of our system which accesses the GISAID databases containing COVID-19 data and the associated EpiCoV database' infection metadata. Because these two databases are hosted separately, they must be correlated prior to being queried, a process which is handled by pandas - a Python data handling and manipulation package. The correlated databases are then uploaded to an SQL database managed with Django ORM - a Python-based toolkit designed to handle SQL databases. Furthermore, because these databases are accessed only through trusted individuals, there is currently no API in existence which can directly query these databases. Due to this restriction, our team

must therefore download the databases for preprocessing with pandas and then upload them again into Django ORM.

Results

We developed an individual-level data set of confirmed and presumptive positive cases of COVID-19 in the world including demographic characteristics, location,and entropy analysis of the COVID-19 gene. Time series of the entropy was also recorded. This data set feeds into an interactive website, which enables users to view the data and follow the outbreak. Our primary data source is GISAID, an online platform, which promotes the rapid sharing of data from coronavirus causing COVID-19 and the ISM analysis was adapted from prior work done by Dr. Gail Rosen and Zhenqiao Zhao.

The database interface uses Django ORM. Once the data from GISAID was matched with the metadata from epicov using python pandas, the data frame was stored in a SQL database. This SQL database was scanned using Django's manage.py inspectdb command which helped the Django model to work with the database. Since the metadata did not include geographical coordinates, latitude and longitude fields were added to the Django model. In future versions, ISM data must also be added to the Django model. This should be a simple process.

Once the model was built, a platform to display the geographical transportation of the virus according to time series data and ISM. It proved difficult to get geographical data from every city region or location that occurs in the metadata. This led us to the decision to scale down the project to just use the more generalized latitude and longitude of countries instead of individual cities. This greatly reduces the implications of the final analysis because it fails to

represent different portions of the country. However, in the future, if we want to increase the resolution to city or county level, it will be as simple as writing a script to translate city/region to latitude and longitude. There are some services out there that can do this.

After mapbox was properly configured with javascript, a map can be implemented on the web page. This interactive map is written in NodeJS and comes with many built-in functions. For this project, we drew colored circles on the map for each record in the GISAID covid-19 database. At first glance in its current state, it does not appear to be many circles. That is because all of the circles are on top of eachother that belong to a single country. However, once the latitude and longitude fields are updated for each record, the dots will spread out.

Another task in the works is to use the ISM tool provided by the TA in order to assign a specific identifier based on entropy of each strain. By using the entropy, it will be possible to color code each circle on the map according to its ISM identifier. Additionally, user interface tools such as date/time fields can be added to the map in order to perform custom queries to see the status of reports at discrete moments in history. Using Django ORM this feature is very easy to put together, and since the GISAID data comes with date/time already built in, there won't be any data-cleaning necessary.

The date/time filter will set the course for the final vision of the tool, which is to animate the progression of COVID-19 infections across the planet based on ISM and location. This visualization tool can be used to help determine which strains came from where, and perhaps if some strains are more contagious than others.

One difficulty to consider moving forward with this website is ongoing maintenance. Since GISAID does not provide a link or an API to download the raw data, it is necessary to

manually log in, download two separate files and append them to your database. Moreover, the metadata has moved from being hosted on github to directly at GISAID's site. This makes maintenance difficult to predict since the source of the data changes. Also the formatting of the data is not guaranteed to stay the same either.

A proper solution for a system where data could be easily accessed for visualization would need to have a standardized, central location for everyone in the world to provide case reports for any disease. This platform would be open source, and would have the proper APIs and protocols to provide data to developers and tinkerers to analyze the data.

Figure 6 shows a screenshot of the current state of the website's map. As can be seen, the dots are centered on the country in which they are reported. The dots are all stacked on top of each other until we can update the latitude and longitude coordinates for each individual city. We have already started this process by creating a list of unique city, region, location found in the raw data, and we have begun to search for a service which is provided to do this as rapidly as possible. One such tool is called geocoders.
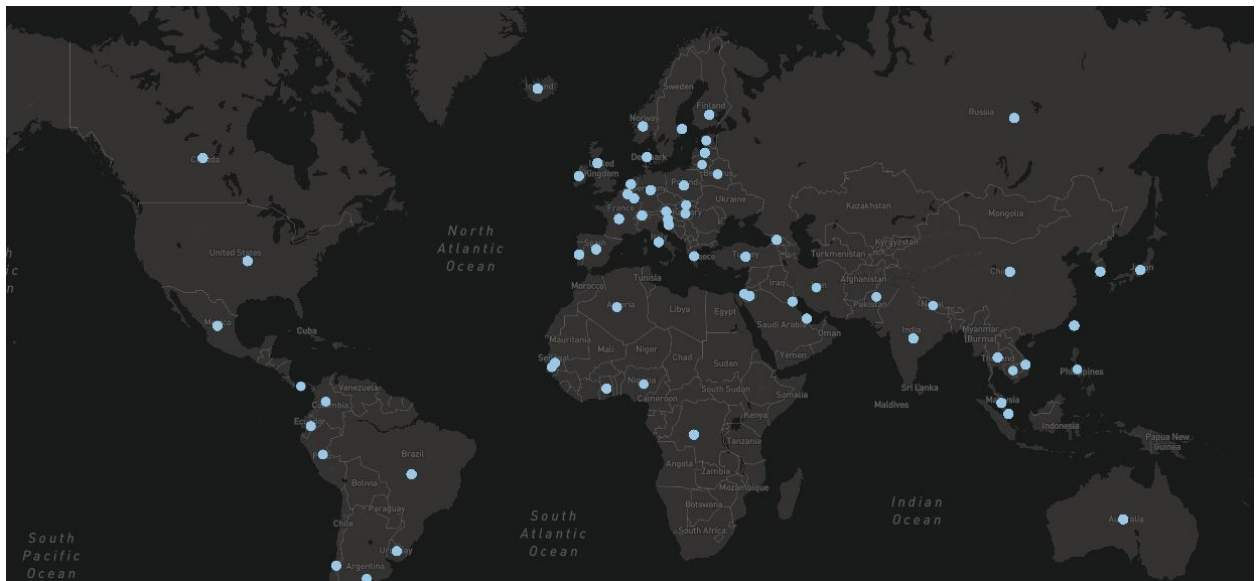
Figure 6. Newly developed interactive interface reports COVID019 information at city-level.

To ensure the validity of our website, we used other data visualization tools, namely Dash and plotly. Dash is an open source Python library for creating reactive,web-based applications. By help of Dash, we can create charts or graphs for data analysis, visualizations , reporting and so on. By using Dash, we were able to create an interactive map which provided the locational details of the respective countries which were affected by the virus according to the GISAID data.

Next, ISM analysis was done as per the guidance of Dr Gail Rosen and Zhenqiao Zhao with that of the data of GISAID and Nextstrain. Json files for the region pie chart and region time series were successfully uploaded. But we were unable to plot these onto the server due to an error in our code, which we hope in the future we will be able to resolve it and then upload to both the server as well as our main website.

Discussion

As of this time in development, we were not able to obtain all of the outcomes outlined in our plans. Our biggest shortcoming was in the plotting of infections on the interactable map. We were unable to plot the ISM pie charts around the different epicenters as visualized in *Fig. 3*., but we were able to plot the markers, which could be replaced with the pie charts at a later date. Another problem with the mapping is that the cities' coordinates were not included in the datasets and we did not have the time to aggregate the locations of each of the cities for each of the thousands of rows of data. Also, instead of designing some code to fetch these coordinates from the web or displaying infections at randomly generated coordinates, it was decided to simply set each infection's coordinates to those of their country (as seen in *Fig. 6)*. While this is not exactly the goal which we had intended for, this is not so bad of an outcome. In addition, we were impeded by an inability to bin the disease markers around epicenters or regions of interest. This too was deemed impossible without the explicit coordinates of each city. Because the binning algorithms would have relied on some distance metric to calculate bins and sort infections, and because distance measures are typically found as the distance between two coordinates, a lack of the coordinate data meant that binning was impossible. If a more liberal interpretation of binning is used, it can be said that all of the infection data for any given country is binned to one marker.

For future directions, we would intend on making the site display more information about the outbreaks, those infected, and treatment facilities so that users of the site would get a better understanding of both the impact of the virus and containment efforts. As for information about the outbreaks themselves, it may be of value to show the daily or weekly spread of the virus around an epicenter with an animated circle whose size encompasses the infections found within a given area. Furthermore, given the importance of the daily and weekly changes in the

number infected, healed, or dead, perhaps the location marker's hue or saturation could be tied to these rates, and a GUI element would allow the user to switch between which of those aforementioned rates impacts the color of the marker. Regarding conveying containment efforts, it may be beneficial to chart local hospitals, doctors offices, treatment centers, and makeshift diagnosis centers around the location of the user's chosen location. While it may seem of value, an effort such as this represents a widening of scope and a magnification of the work needed to be done as well as an increase in time commitment to add these functionalities. For these reasons, this possibility remains a future direction. Another potential future direction is a collaborative effort to create an API for GISAID's databases so that our site, and others like it, are able to easily grab information from their databases live (as compared to the current method of caching each update and then accessing that local copy). Finally, there is always the possibility of summarizing more metadata for the site's users to peruse. This will allow visitors and researchers alike to get a better and more detailed understanding of who the virus affected and how it affected them.

For the long-term impacts of this project, it can be assumed (or at least hoped) that the coronavirus pandemic will subside at some point or affect so few individuals that a tracking effort is no longer needed or justifiable. While this means that the world will be in a better place, our site will inevitably lose importance as its functionality is tied to an impermanent occurrence. At that time, there will be two things which we can do with the site and the framework that we have built that will serve to educate users and provide information for future researchers. The first of these options is to keep the server online so that anyone is able to access the data even after COVID-19 becomes non-threatening. This would allow anyone who is curious to crawl through the data and understand the scope of the issue in hindsight, and for this reason the site

may have value as an educational tool or as a database for those interested in historical research. In addition, keeping this database active will allow bioinformaticians, disease researchers, government agencies, and the like to understand how the virus spread, the scope of the pandemic, and the effects different measures had. This will help them so they can better prepare for future pandemics, learn how to mitigate transmission, and inform more timely and effective policies for crowd control, social distancing, and the reopening of businesses. This first option is more research oriented so that others may learn from what we did right and what we did wrong when addressing a pandemic in such a globalized society. The second option for our efforts is to repurpose and reutilize the site structure and means of data collection and representation for future pandemics. For those outbreaks, data capture and analysis will be as important for mitigation and treatment as it is now, and therefore methods for accessing this data and presenting it in a comprehensible manner will again be of utmost importance for those efforts. Because pandemics will happen and these tracking efforts will be needed, it would show a high degree of foresight to make the site's structure and code as generalizable as possible so that it can access whatever necessary databases and provide a similar user experience. This would allow for a new tracking site to go live almost as soon as pandemic data starts being recorded without the need to invest time, money, and resources into making a new tracker from scratch.

The biggest constraint in understanding the pandemic through plotted data on a site such as ours is the lack of consistency in metadata collection and logging. While there are universally-established means of collected metadata, there is much variation about how metadata is collected and choosing which gets uploaded. While this is not always a problem, there is sometimes not enough data in a category to form a meaningful pattern or notice a

correlation between factors. In addition, there is a more problematic impediment to learning and understanding pandemics - the presumption that some metadata is irrelevant and not worth collecting. By choosing to analyze, for example, only the ages, locations, and sex of those infected, researchers by default undervale the importance of other factors which prevents them from understanding the impact of other factors.

To be able to understand and appreciate the impact of this site, the pandemic must first be subdued so that our efforts may be evaluated in hindsight. Therefore, investigations about what information would be useful or even vital to mitigating and curtailing the outbreaks can only be conducted at some future date. However, even during the pandemic, researchers can begin to understand exactly what data they want to analyze, but without the benefit of hindsight, they may become stuck between choosing to record and process too little metadata or too much metadata. In the instance of too little metadata, some key factors about disease spread and prevention may be lost because they were deemed irrelevant or there is simply not enough data to make statistically-relevant conclusions about those factors. In the case of too much metadata, research efforts are slowed by the time taken to record all of the data, process it into a database, and then have that information transferred. Those efforts are further slowed by the amount of time taken to make projection models and predictions, run statistical tests, and rule out correlations and confounding effects. Because of these reasons, it is reasonable to assume that the most impactful investigations will occur after the pandemic has been curtailed.

References:

1. "COVID-19 Situation Reports." World Health Organization. World Health Organization. Accessed May 5, 2020.

https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports.

2. Grubaugh, N. D., Ladner, J. T., & al, e. (2019). Tracking virus outbreaks in the twenty-first century . *Nature Microbiology* , 10-19.

3. Isha Berry, Jean-Paul R. Soucy, Ashleigh Tuite, and David Fisman. "Open Access Epidemiologic Data and an Interactive Dashboard to Monitor the COVID-19 Outbreak in Canada." CMAJ. CMAJ, April 14, 2020. https://www.cmaj.ca/content/192/15/E420.

4. Jin, B. (2020, March 16). *Live tracker: How many coronavirus cases have been reported in each U.S. state.* Retrieved from Politico : https://www.politico.com/interactives/2020/coronavirus-testing-by-state-chart-of-new-cases/

5. Rosen, Gail. "EESI/ISM." GitHub, June 3, 2020. https://github.com/EESI/ISM.

6. RStudio. *Shinyapps.io*, www.shinyapps.io/.

7. University, J. H. (2020). An interactive web-based dashboard to track COVID-19 in real time. The Lancet , 534-44.

8. Zhao, Z., Sokhansanj, B., & Rosen, D. G. (2020). Characterizing geographical and temporal dynamics of novel coronavirus SARS-CoV-2 using informative subtype markers. BioRxiv, 1-32.

9. Dash for Python, https://dash.plotly.com/