Visualizing the Spread of Pandemics

Jose Figueroa and Liam Donohoe
Department of Computer Science
Rensselaer Polytechnic Institute
Troy, NY, USA

Abstract—In this work, we propose and implement a new visualization of the spread of pandemics through the world. Using R, we created a heat map of countries, representing the number of cases in any given country at any given time. The goals are to be able to gain a better understanding of the speed at which pandemics spread, as well as their severity over time.

I. Introduction

A. Motivation, Audience and Goals

The goal of this project is to create a visualization of the progression of past epidemic or pandemic viruses throughout their lifetimes. In doing this, we hope to be able to provide a better understanding of not only the rate at which these types of viruses can spread throughout the population, but also the time it takes for scientists to develop and distribute a vaccine or cure for the spreading virus. Our target audience are both casual audiences and policy makers. With the current situation of Coronavirus in the world, this is a big topic of discussion today. This kind of information could help anyone gain a better understanding of how the virus is spreading, and perhaps even a better idea of how long it may last.

B. Data and Expected Challenges

Datasets we are considering include an H1N1 influenza virus dataset provided by World Health Organization (WHO), 2019-nCov Coronavirus data provided by John Hopkins, and 1918 influenza data provided by the CDC. The main challenge we anticipate is creating a visualization that both provides a unique perspective to this data and is also easy to analyze and intuitive to interact with. That said, the data will likely be in a time series format, so highlighting the time frame for interaction may be the way to remedy this. Other important visual components will be the color scheme, as we must be able to convey risk, danger etc immediately, as well as an appropriate way to convey different forms of "safety" be it lower mortality, lower rates of transmission, more hospital capacity, or some aggregate.

C. Discussion Board Feedback

General approval of the goals of the ideas, though a few people bring up the challenge of collecting data with integrity. This comes from many causes, such as questionable self reporting from countries, and difficulty to pin causality with the stock market example. There was also concern over the potential large scope of the assignment, with feedback suggesting we pick a manageable subset of the assignment, focusing on the visualization aspect.

D. Hypothesis

Going back to our motivation, our central research question is "What can we learn from previous respiratory virus outbreaks to handle the current pandemic?" We hypothesize that given previous respiratory illness outbreaks move in similar ways around the globe, and severity of outbreaks can be linked to how each country handles said outbreak.

E. Demonstration

To demonstrate this visualization, the most basic example would be to start it up with a data set, and choose a date during the pandemic. Moving forward from this would be to have the visualization run from the start date to the end date on its own. Beyond this, testing would include changing the data set, and changing the scale and position of the map. Past this, we would ideally want to conduct some form of a user study. We don't just want to ensure that this visualization is not only functional, but informative and effective as well. We want to make sure that our hypothesis holds, and that people can effectively read and analyze the data represented in the visualization in order to draw the appropriate conclusions about the data.

II. The Visualization

A. Interactive Heat Map

For this project, we have created an interactive map, that displays the number of cases, and fatalities from the chosen virus over time. The number of cases will be represented on the map in one of two ways, depending on how much accuracy we can obtain in the locations of new cases. If we are able to obtain more precise location information, perhaps to the city or town level of a state or country, then we can show tall bars coming out of that location representing the number of cases. If we can only have precision down to the state level, then we will instead use color, or an expanding circle based on the number of cases. This option could also be used with a higher location precision, but might lose some of the accuracy with more data.

B. Adaptability

With the way we have implemented this project, we are able to handle different data sets from different pandemics. This only involves formatting the new data into the desired format using existing code we have already written. With this, we are able to view various different diseases, and potentially be able to draw additional conclusions about pandemics.

C. Design and Implementation

In our implementation, we decided on using the R programming language. This decision was made based on the data processing power of R, combined with the many tools and libraries included that would facilitate the creation of this visualization. To create the interactive world map, we took advantage of Leaflet and Geojson to draw each country's shape, and be able to represent the data on each country. The Shiny package was utilized in order to package everything together into an applet that was hostable on a server. Lastly we improved the aesthetics of the entire interface using javascript with Bootstrap CSS. Each of these packages combined together provided the ideal framework in R to create this visualization.

D. Distribution of Labor

In this project, we divided up the tasks as equally as we could. Jose was far more familiar with the language and frameworks, and so took the lead on a lot of the implementation specifics. Liam had to spend some additional time learning and getting familiar with the language to begin. Jose was responsible for setting up the main framework and processing the data for the project. Liam did work in implementing interactive components of the map, and additional work in visualizing the data.

III. Results

A. Initial Design and Plan

Our initial plan for the design of this visualization was to have a potentially three dimensional world map, with a directly overlaid heat map for each country. The color of each country would be representative of the number of cases in that country at the current date. Upon loading the visualization, the user would be able to start the progression of time, to see the progression of the number of cases over time. Additionally, they would be able to manually select any date range they wish to view to more accurately analyze the data at any given time. If the data is readily available, we hoped to be able to toggle between various pandemics in order to compare between different data sets to analyze the range of severity of pandemics. Finally, to coincide with the visualization, it would help the analysis of the data to have a line graph representing the number of cases in a selected country over time. This would allow the user to have a more accurate idea of the actual number of cases over time to help gauge the severity. Below is included a rough idea of what we would plan for this visualization to look like in a finished state.

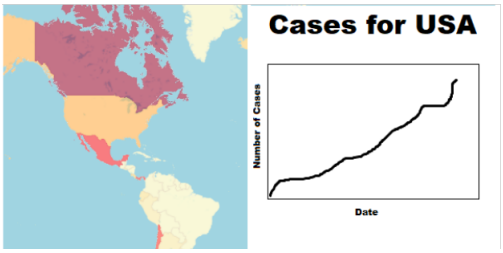

Fig. 1: Original design idea for the visualization. On the left is the world heat map of the number of cases in the world at the current date. The right half represents the number of cases over time for the currently selected country. Together, we have a more complete idea of the spread and severity of the pandemic.

B. Final Visualization

The final version of our visualization is very close to what we envisioned. We created a world heat map of countries, representing the number of cases in each country during any given weekly time interval. One last modification we had to make was to normalize the data based on the size of the country. It became clear that countries with higher populations would naturally have more cases than less populated countries. To prevent this from skewing the data, we utilized the closest census data to the time period of the pandemic in order to normalize the number of cases per country based on the population in that country. This gave us a cases per million metric, rather than a raw case count for each country. This allowed us to gain a better idea of the severity that the pandemic affected each country individually, as well as determine where it had the largest effect. Due to the time it took to render the map with each update of the timeline, we were unable to make an active progression of time in this visualization. Instead, we created a radio button system to toggle through each available date in which the pandemic was spreading. Included below are some screenshots of the visualization being used.

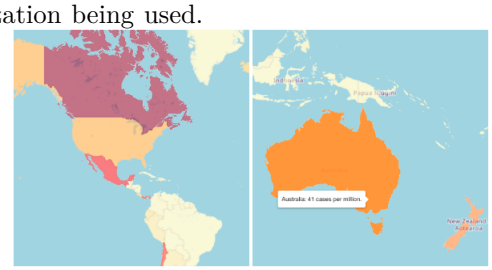

Fig. 2: Heatmap of the spread of the H1N1 pandemic represented with country shapes, dated weeks ending in 5/30/09 and 6/20/09 respectively. The maps highlight the early peaks of the pandemic in North America, Oceana, and Western parts of South America.

Also included with this report is a separate screen shot of the full sized visualization. Due to the size of the screen shot, it will not fit properly in this paper while remaining

easily readable. Instead, we have included it as a separate attachment.

Fig. 3: Full size screen shot of the Visualization. Clicking on a country will pan and zoom to scale in order to show the chosen country in the screen. The scale is shown on the right of the visualization representing the number of cases per million people in that country. On the top of the screen, is shown the given date ranges that can be shown for the given pandemic. Also included is an option to show the total cases for the current pandemic over the full time range.

C. Effectiveness

In analyzing the effectiveness of our visualization, it would be best to be able to conduct a user study. With this we would be able to ask individuals to use the visualization in order to answer questions regarding the pandemic details. These might include asking people to determine the number of cases at a given date in a given country, or determine at which date a certain country was affected the most by the pandemic. The results of this would be able to tell us if we have created an effective visualization that has accomplished our goal. With our limited time frame, we are not able to conduct any kind of user study for our project. If we were to pursue continuing this project, a user study would certainly be an important part of the timeline of its future.

IV. Risks and Limitations

A. Scope

This is a project that has nearly unlimited possibilities, so we had to stay away from planning too much. It would be very easy to come up with several ideas that would make a very interesting and useful visualization, but we must limit ourselves to a feasible scope. This project does not exactly have any real limitations, besides acquiring the data, and having enough time to accomplish the desired tasks. With these limitations in place, we designed our project in a way that we can create a working visualization that we can expand upon, and implement as many features as we can in the time we have.

B. Data Acquisition

The other possible limitation or risk in this project is finding accurate and complete data to visualize. We were able to find a data set for H1N1 that is accurate, but full of holes. Because of the nature of this data, it is hard to find uniform data that covers every individual day for each country, that is still accurate. To combat this, we decided to aggregate the data by a weekly basis. This way, we did not have any empty data points, where some countries would have an entry but others would not. Although we had to sacrifice some of the precision of the data, we overall gained a more informative and accurate visualization. After using this fixed data in our prototype visualization, we found a new problem with the data. Because of the variation in the population sizes of each country, the scale for the number of cases became skewed by more populated countries. Normalizing the data based on the population of each country was an effective way to represent all of the data in a balanced format.

V. Related Work and References

This type of visualization, especially today with a pandemic in the world, is a very popular and common visualization. The following are similar types of visualizations that we took inspiration from to create our own. MortalityMinder utilizes map based heatmaps to represent trends in mortality in the United States. This served as inspiration to use leaflet as a tool to represent a combination of time and location data. “A pandemic influenza modeling and visualization tool” by Maciejewski, Ross, et al also gave a similar representation for representing pandemic data with maps. GLEaMviz presented by Wouter Van den Broeck, gave inspiration as to what additional features we can add to the map given less time constraints, such as additional features as layers on the map.

VI. Future Work

A. Additional Features

With the shorter time frame and unusual working situation, we had to decrease the scope of the project to some extent. We have many additional features that we either originally planned and ran out of time to implement, or foresaw as long term goals in the first place. We would have liked to find additional data sets for other pandemics to visualize and compare to each other. With this we could have gone on to create a parallel visualization of the rate of new cases in a line graph over time, juxtaposed with the heat map of the same data. If we had the data available to us, we also envisioned being able to analyze Covid-19 using past pandemic data, and gain an idea of how it compares. After reading the DimpVis paper earlier in the semester, we had the idea to potentially attempt to utilize this tool in our visualization to improve the navigation of the time-based data.

B. Future Applications

Beyond just visualizing the data in itself, there are several additional applications we could apply. By using machine learning techniques, we could potentially use the existing data we have found to then create and predict future pandemic data based on various factors. We could compare infectivity, starting location, as well as many others to gauge the spread and severity of a new pandemic in the world. Moving beyond the scope of pandemics alone, we could combine this visualization with another time-based data set to potentially find correlation with pandemics. This could be the sales figures of certain types of products, or in a broader sense, change in stock prices.

References

[1] B. Kondo and C. Collins, "Dimpvis: Exploring time-varying information visualizations by direct manipulation," IEEE Transactions on Visualization and Computer Graphics, vol. 20, no. 12, pp. 2003–2012, 2014.

[2] J. K. Medema, Y. F. Zoellner, J. Ryan, and A. M. Palache, "Modeling pandemic preparedness scenarios: health economic implications of enhanced pandemic vaccine supply," Virus Research, vol. 103, no. 1, pp. 9 – 15, 2004, proceedings of the First European Influenza Conference. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0168170204001054

[3] R. Maciejewski, P. Livengood, S. Rudolph, T. F. Collins, D. S. Ebert, R. T. Brigantic, C. D. Corley, G. A. Muller, and S. W. Sanders, "A pandemic influenza modeling and visualization tool," Journal of Visual Languages Computing, vol. 22, no. 4, pp. 268 – 278, 2011, part Special Issue on Challenging Problems in Geovisual Analytics. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1045926X11000292

[4] W. V. d. Broeck, C. Gioannini, B. Gonçalves, M. Quaggiotto, V. Colizza, and A. Vespignani, "The gleamviz computational tool, a publicly available software to explore realistic epidemic spreading scenarios at the global scale," BMC Infectious Diseases, vol. 11, no. 1, p. 37, Feb 2011. [Online]. Available: https://doi.org/10.1186/1471-2334-11-37