**COVID-19 and the Weather: A data visualization**

Nick P Palacio

Department of Computer Science, University of Nebraska at Omaha

Capstone Project Proposal

Dr. Brian Dorn, Dr. Rex Cammack, Dr. Brian Ricks

**COVID-19 and the Weather: A data visualization**

On January 9th, 2020 the World Health Organization (WHO) announced that a coronavirus related pneumonia had been spreading in Wuhan, China. The US confirmed its first coronavirus case on January 21st, 2020 (American Journal of Managed Care, 2021). By March 11th, 2020 the WHO had declared the COVID-19 pandemic.

Since then the US has experienced several waves of increased infection rates that have varied in severity across the country. According to the CDC, COVID-19 can spread from human to human via respiratory droplets in the air. Specifically, the virus is known to spread more easily indoors where there is less air ventilation (CDC, 2020).

There is precedence to think that the COVID-19 virus spreads more easily in certain weather conditions. Influenza is another respiratory illness that is spread via respiratory droplets in the air. It is well established that influenza spread is influenced by the weather (Huang et al., 2017; Roussel et al., 2016). Specifically, temperature, humidity, and daily variation of both have been found to have a significant impact on influenza infection rates (Roussel et al., 2016). While the coronavirus is certainly not the same thing as the flu it does spread in a very similar manner. This makes the relationship between COVID-19 transmission and the weather worth exploring.

There has been some research published already exploring the relationship between weather and COVID-19. However, results from these studies have been mixed. One literature review published in the International Journal of Environmental Research and Public health analyzed the current available literature on the association between weather and COVID-19 incidence. This study concluded that weather does in fact play a significant role in COVID-19 transmission. Specifically, it said temperature and humidity were found to be the most common weather factors associated with COVID-19 transmission (McClymont & Hu, 2021). Another study published in the same journal highlighted an issue with the existing research on COVID-19 and weather. It said that existing research on this association only considers weather variables during analysis. This study concluded that weather on its own was a non-influential factor in COVID-19 transmission. Instead, it said that other factors such as urban density and mobility of the population influenced COVID-19 transmission much more than weather (Jamshidi et al., 2020). One limitation of both studies is the data that they had to work with. The first research article discussed was received for peer review in November 2020. The second article was received in September 2020. This means that both articles were working with limited COVID-19 data, specifically missing out on spikes that were seen in the United States during the November - January time frame. These articles highlight the need for continued exploration of the relationship between COVID-19 and the weather.

Existing research has been aimed at proving or disproving weather's effect on the pandemic. My project aims to allow a user to explore this relationship on their own as opposed to establishing whether one exists or not. My proposed project is a web application that would allow a user to explore the relationship between weather and COVID-19 in different parts of the United States by interacting with a map and several charting widgets that would plot weather and COVID-19 infection data side by side. This will be a map-based web application that allows a user to go to an area of interest in the United States and view COVID-19 data and weather data together for a given date range. A user would be able to view the weather and COVID data using charts that would be linked together. This visualization tool would have several configurable features that a user can set in order to visualize the COVID-19 and weather data, such as: county of interest, date range, lag between weather and COVID-19 data and which weather data points to view alongside the COVID-19 data (temperature, absolute humidity, relative humidity). (**Thinking I should provide a mockup or something of it here?)**

The intended user for my project would be a middle school scientist. I have chosen this population of users for several reasons. The effects of the pandemic have not been limited to any subset of the population. Students across the United States had to abruptly switch to remote learning as the pandemic began. My project aims to allow students to link what they are learning in school to what is happening in the world today. My tool would help facilitate an activity in the classroom, virtual or otherwise, that would allow the students to investigate the hypothesis that weather does, or does not, influence COVID-19 transmission in the United States. According to Nebraska standards for science education, by the 7th grade students should be able to understand evidence for how different factors contribute to the weather and climate. Students should also understand the scientific process for asking questions and carrying out investigations by gathering evidence (<https://cdn.education.ne.gov/wp-content/uploads/2017/10/Nebraska_Science_Standards_Final_10_23.pdf>). My project will help to reinforce these ideas by investigating the question of if weather affects COVID-19 transmission by following the scientific process. Students would be expected to think critically about other factors besides weather that might drive infection rates as well. An activity like this could help drive STEM education by showing students what they are learning in school is related to what is going on in our world today.

**Related Work**

In this section I will compare some existing COVID-19 data visualizations to highlight work that is currently out there as well as some gaps in that work. I looked at COVID-19 data visualizations from John Hopkins University of Medicine (John Hopkins, 2021), the COVID Tracking Project at the Atlantic (The COVID Tracking Project, 2021) and the Institute for Health Metrics and Evaluation (IHME) at the University of Washington (Institute for Health Metrics and Evaluation, 2021). I compared the visualizations that these organizations offered along several dimensions. Specifically, I looked at the following: How granular is the COVID-19 data? Which COVID-19 data points are visualized? Does it offer a spatial view? How configurable are the visualizations?

The results of this comparison can be seen in Table 1.

**Table 1**

*Comparison of Existing COVID-19 Data Visualizations*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Organization** | **Granularity of COVID-19 Data** | **COVID-19 Data Points** | **Any Spatial View** | **Configurability of the Visualizations** |
| **John Hopkins** | County, State and Country | Confirmed Cases, Deaths, Tests, Hospital Use | Yes, map of US with counties | Minimal, can toggle the data point plotted |
| **COVID Tracking Project** | State and Country | Confirmed Cases, Deaths, Tests, Hospital Use | Yes, map with hospital use data, a few cartograms | Moderate, can set date range and if data is normalized |
| **IHME** | State and Country | Confirmed Cases, Deaths, Tests, Hospital Use | Yes, most data points can be viewed on a map | Moderate – High, can set date range, if data is normalized and if data should be 7-day rolling averages |

Now I will summarize my findings and discuss how this relates to my visualization. Only one of the organizations, John Hopkins, offered COVID-19 data at the county level in the US. Given that my visualization will show weather and COVID-19 data together, the location granularity of this data becomes more important. Weather in any state can vary greatly across different locations in that state. This is why my visualization will use county level COVID-19 data. All of the organizations offered the same COVID-19 data points in their visualizations (cases, deaths, etc.). For my purposes of allowing a user to compare COVID-19 infection rates to weather patterns I will only be using confirmed COVID-19 case counts.

Since weather and COVID-19 infection rates both have a spatial dimension, a spatial view for my visualization is warranted. This is consistent with the existing visualizations I have looked at, all 3 provided some sort of spatial view for the COVID-19 data. This is why I will be displaying a map to the user that they can interact with in order to view data at their location of interest. These organizations offered a variety of levels of configurability in their visualizations. Given that the purpose of my visualization is to allow a user to explore the data on their own I will offer a high level of configurability in my visualization in order to allow a user to visualize the data in a few different ways.

**Methods**

**Data Sources**

For my COVID-19 data source I will be using one of the datasets generated and maintained by the New York Times hosted on GitHub (The New York Times, 2021). Specifically, I will be using the us-counties.csv dataset. This dataset contains a full history of cumulative COVID-19 cases and deaths by county by day in the US.  
For my weather data I will be using an API from Weather Source. Weather Source is a technology company that provides a suite of products that help businesses leverage weather and climate data. On March 16th, 2020 Weather Source opened their API for free to any researchers exploring the relationship between weather and the COVID-19 pandemic. Their Weather History API exposes many different weather data points that can be queried with a date range along with latitude and longitude, or zip code. Data can be returned in an hourly or daily format. For my purposes I will be retrieving average temperature, average relative humidity and average absolute humidity in a daily format.

**References**

America Journal of Managed Care. (2021, January 1). A Timeline of COVID-19 Developments in 2020. <https://www.ajmc.com/view/a-timeline-of-covid19-developments-in-2020>

CDC. (2020, October 28). COVID-19 and Your Health. Centers for Disease Control and Prevention. <https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/how-covid-spreads.html>

Huang, X., Mengersen, K., Milinovich, G., & Hu, W. (2017). Effect of Weather Variability on Seasonal Influenza Among Different Age Groups in Queensland, Australia: A Bayesian Spatiotemporal Analysis. The Journal of Infectious Diseases, 215(11), 1695–1701. <https://doi.org/10.1093/infdis/jix181>

Institute for Health Metrics and Evaluation. (2021, March 20). IHME | COVID-19 Projections. Institute for Health Metrics and Evaluation. <https://covid19.healthdata.org/>

Jamshidi, S., Baniasad, M., & Niyogi, D. (2020). Global to USA County Scale Analysis of Weather, Urban Density, Mobility, Homestay, and Mask Use on COVID-19. International Journal of Environmental Research and Public Health, 17(21), 7847. <https://doi.org/10.3390/ijerph17217847>

John Hopkins. (2021, March 20). Coronavirus Resource Center. Johns Hopkins Coronavirus Resource Center. <https://coronavirus.jhu.edu/>

McClymont, H., & Hu, W. (2021). Weather Variability and COVID-19 Transmission: A Review of Recent Research. International Journal of Environmental Research and Public Health, 18(2), 396. <https://doi.org/10.3390/ijerph18020396>

Roussel, M., Pontier, D., Cohen, J.-M., Lina, B., & Fouchet, D. (2016). Quantifying the role of weather on seasonal influenza. BMC Public Health, 16, 441. <https://doi.org/10.1186/s12889-016-3114-x>

The COVID Tracking Project. (2021, March 20). Charts. The COVID Tracking Project. <https://covidtracking.com/data/charts>

The New York Times. (2021). Coronavirus (Covid-19) Data in the United States. Retrieved March 7, 2021, from <https://github.com/nytimes/covid-19-data>.