

COVID-19 and the Weather: A data visualization

Nicholas P Palacio

Department of Computer Science, University of Nebraska at Omaha

Thesis Equivalent Project Final Report

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 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.

There is precedent to think that the coronavirus, also known as COVID-19, spreads more easily in certain weather conditions. According to the Centers for Disease Control and Prevention, CDC, COVID-19 can spread from human to human via respiratory droplets in the air, and the virus is known to spread more easily indoors where there is less air ventilation (CDC, 2020). Dr. Fauci, who serves as the director of the US National Institute of Allergy and Infectious Diseases, spoke about the potential connection between COVID-19 and the weather in April 2020 on ABC's Good Morning America saying:

There is precedent with other infections like influenza and some of the common more benign coronaviruses that when the weather gets warmer that the virus goes down, that its ability to replicate, to spread, it doesn't like warm, moist weather as much as it likes cold, dry weather.

But having said that, one should not assume that we are going to be rescued by a change in the weather. (AP, 2020)

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, which Dr. Fauci alludes to above (Huang et al., 2017; Roussel et al., 2016). Roussel et al. (2016) studied the role of weather on seasonal influenza spread in France. Their study found 2 groups of 3 climatic variables that had a significant impact on seasonal influenza spread at the intra-annual scale. The first group of variables was average temperature, absolute humidity, and daily variation of absolute humidity. The second group of variables

was sunshine duration, relative humidity, and daily variation of relative humidity. The impact of these groups of variables on seasonal influenza spread was found to be relatively low, between 3% – 6%.

While the coronavirus is certainly not the same thing as the flu it does spread in a similar manner. This motivates exploring the relationship between the weather and COVID-19 transmission.

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 (McClymont & Hu, 2021). This literature review looked for relevant studies on COVID-19 and weather by searching PUBMED, Web of Science and Scopus databases. The 23 articles selected for this review were epidemiological studies that evaluated the relationship between weather variables and COVID-19 transmission up to October 1st, 2020. All 23 articles included temperature in their study. 18 of the 23 studies reported a significant correlation between temperature and COVID-19 incidence. However, of these 18 studies 11 reported a negative correlation while the remaining 7 reported a positive correlation. 16 of the 23 articles included humidity in their assessment. Of these 16, 12 reported significant associations between humidity and COVID-19 incidence. However, of these 12, 4 reported a positive correlation, 6 reported a negative correlation and 2 reported an optimal range of humidity for new cases.

Another study published in the same journal highlighted an issue with the existing research on COVID-19 and weather. Jamshidi et al. (2020) noted that existing research on this association only considers weather variables during analysis. In this study instead of just looking at weather variables and their impact on COVID-19 transmission they looked at other important factors such as mobility, homestay, population, and urban density. For their weather variable they used equivalent temperature which is a combination of temperature and humidity. The study evaluated the impact of equivalent temperature on COVID-19 transmission using different scales such as global, regional, US state and US

county. At the global scale, this study found contradictory patterns between equivalent temperature and COVID-19 infection rates. From January to July 2020 USA, Italy and India showed a positive correlation between the two while China, Brazil and Australia had a negative correlation. At the US county scale equivalent temperature was found to have a contributing factor of <3%. This study recommended using finer scale weather data when incorporating it into a study given how much weather can vary across a country or region. They 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.

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 2020 – January 2021 time frame.

There is very limited recent research on weather's role in COVID-19 transmission. Most of the research currently out there is working with data from early to mid-2020. One study that works with relatively newer data studied weather and environmental air quality's effect on COVID-19 transmission and deaths in Istanbul, Turkey. Shahzad et al. (2021) looked at weather parameters, temperature and humidity, alongside environmental quality parameters, such as air quality pollutants, ozone, nitrogen dioxide, etc., in order to study their effects on COVID-19 transmission and deaths in Istanbul, Turkey. They used data from March 11th, 2020 – February 8th, 2021. They found that humidity was not correlated to COVID-19 transmission or deaths while temperature was found to be positively correlated to COVID-19 transmission and deaths.

One recent study took a slightly different approach than just looking at weather variables like temperature and humidity. Karapiperis et al. (2021) instead looked at ultra violet (UV) radiation. They hypothesized that UV radiation levels might give COVID-19 transmission a seasonality pattern. In their study they looked at data from 12 countries, selected for their similar characteristics. The data used from these countries was all in 2020, through November. They used machine learning techniques to investigate the relationship between UV radiation, population mobility and COVID-19 incidence. They found that UV radiation levels were strongly associated with COVID-19 incidence. This study provides a different perspective on weather's role in the pandemic. However, it is similar to previous studies mentioned in that there was a relationship found between a weather variable, UV radiation levels, and the pandemic.

This more recent research on weather's role in COVID-19 transmission is in line with some of the previous work that found a positive correlation between temperature and COVID-19 transmission. However, this can hardly be called conclusive due to the inconsistent findings in previous work. That notwithstanding, all of these articles highlight the fact that there is an ongoing debate right now in the scientific community around weather's role in the COVID-19 pandemic.

In this project, I built a visualization tool that can be used to explore this relationship. The intended user for my project is a middle school scientist because this open debate in the scientific community presents a unique opportunity to engage students. Middle school scientists were selected as the intended users over high school or even post-secondary students because this is the youngest population that should be ready to make this kind of scientific evaluation and I have limited resources. According to the Nebraska Department of Education (2017), 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. Given the right tools, teachers could engage students in the scientific process by tasking them

to perform their own investigation into the same question of weather's role in the COVID-19 pandemic. My visualization equips a teacher with a tool that students could use to explore this relationship. An activity like this would provide the students the opportunity to think critically and ask questions about the data and what conclusions can, or cannot, be drawn.

In 2018, Lee and Wilkerson studied data use by middle and secondary school students. One of the things they looked at was how teachers can best support students working with data. One of their recommendations for teachers' use of data in the classroom was that data should be leveraged in the context of meaningful scientific pursuits. My project falls in line with this guidance because students would be asked to participate in an open debate in the scientific community and draw their own conclusions using evidence they gather using the tool.

A study by Linn et al. (2006) found evidence that visualization technologies can improve student learning outcomes while they learn scientific concepts. From a high level, this study compared assessment results for two groups of students who received different curriculum. One group received a normal curriculum while the other group received curriculum that included visualizations of scientific phenomena in order to help illustrate them. They found that both groups of students performed equally well on multiple choice assessment questions. However, the group that received the curriculum that included the visualizations performed significantly better on assessment questions that required the student to provide their own explanations. Questions that require the student to provide their own explanations are better able to discriminate varying levels of knowledge integration, making these findings significant. While my visualization does not try to explain any particular scientific phenomena like heat transfer or a chemical reaction it does provide students a visual representation of a couple scientific phenomena, disease spread and weather.

Existing research has been aimed at determining 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 project is a web application that allows 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 will plot weather and COVID-19 infection data side by side.

Related Work

In this section I compare some existing COVID-19 data visualizations to highlight current work as well as some gaps in that work. I also highlight some existing literature on how students interpret graphs as well as best practices for presenting graphs to students. The visualizations I selected for evaluation were found by doing my own research. I wanted to find visualizations on the Internet that came from a trustworthy organization and provided views into similar data points that I wanted to use, specifically confirmed cases by location. I collected information about 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 |

Only one of the organizations, John Hopkins, offered COVID-19 data at the county level in the US. Given that my visualization is designed to 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 (Jamshidi et al., 2020). Therefore, my visualization uses county level COVID-19 data. All 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 only use 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 display 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 offer a high level of configurability in my visualization in order to allow a user to visualize the data in multiple ways.

These existing COVID-19 data visualizations are limited in a couple ways. They are designed for a general audience, presumably the public. My visualization will be designed for middle school scientists for use in a classroom activity. As I will discuss later, visualization for learning requires special design considerations. Another limitation of existing visualizations is they are not designed for exploring the relationship between weather and COVID-19 transmission. My visualization is designed for exploring this relationship.

There has been research done investigating how students interpret graphs as well as best practices for providing graphs to students for their interpretation. A literature review by Hoeffner and Shah (2002) looked at the cognitive literature on how people understand graphs. This paper looked at 3 factors that influence a viewer's understanding of a graph: the visual characteristics of the graph, a viewer's knowledge about graphs, as well as a viewer's knowledge about the data in the graph. The paper synthesizes these findings into recommendations for how to best present graphs to students. One of their recommendations was to represent the same data in multiple formats. This helps students' understanding when there are multiple quantitative facts to communicate about the data. Another recommendation from this paper was to be careful about the density of the data points, specifically for scatterplots, because users often mentally exaggerate how correlated 2 variables are in a scatterplot that is very dense with data points. A graph can become denser by either adding data points or shrinking its size.

I have 4 quantitative facts about the data I wish to communicate for a given US county and date range: the trend of COVID-19 infections, the trend of several weather data points, the covariance of

COVID-19 infections with each weather data point and the population density of the county. Given this, I provide a scatterplot graph that communicates the covariance of COVID-19 infections and a weather data point that a user could select from a predefined list. I also provide individual line graphs of COVID-19 infections and each weather data point that communicates the trend of each variable on its own. I do not limit the number of data points on the scatter plot in the tool. This is driven by the date range a user configures. However, I do provide the correlation coefficient, specifically the Pearson Correlation, on the scatterplot in order to provide another indicator for how correlated COVID-19 infections are with the selected weather data point. In order to communicate the population density of the counties, I am symbolizing the counties on the map based on population density.

Methods

Data Sources

For my COVID-19 data source I am using one of the datasets generated and maintained by the New York Times hosted on GitHub (The New York Times, 2021). This data source provides several datasets that can be downloaded via GitHub. There is also documentation about the datasets that can be viewed on GitHub to understand how they are structured. I am 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 going all the way back to January 1st, 2020. I evaluated two other sources for COVID-19 data before selecting the New York Times dataset. One of them came from the COVID Tracking Project published by *The Atlantic*. This data source provided an API as well as files you can download. However, it only had COVID-19 data at the state level. Given the location sensitive nature of both weather and COVID-19 data, state level data will not suffice. Weather in any state can vary greatly depending on location so I wanted county level data. The other data source I evaluated came from the Center for Systems Science and Engineering at Johns Hopkins University. This data source was also hosted on GitHub where the

dataset files can be downloaded. This data source is very similar to the New York Times data source in that it provides case counts by county in the US. It also provides good documentation. This data source would work for my project as well. In the end, selecting one of these was arbitrary so I went with the New York Times.

For my weather data I am 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 am retrieving average temperature, average relative humidity and average absolute humidity in a daily format.

User Stories

Table 2 below contains the list of user stories for my project. User stories are descriptions of a feature of a piece of software told from the perspective of the user who desires the feature. Their typical format is 'As a <type of user>, I want <some feature> so that <some reason>'.

Table 2

User Stories

| | User Stories |
|-------------|---|
| US.1 | As a student, I can search for a county of interest on a map in order to begin to investigate weather's role in the pandemic for this county. |
| US.2 | As a student, I can search for a county of interest by name and view a list of matches in order to begin to investigate weather's role in the pandemic for this county. |
| US.3 | As a student, I can select my county of interest on a map and view COVID and weather data for that county. |
| US.4 | As a student, I can view the number of daily COVID cases in my county of interest over a date range in order to understand the trend of COVID-19 cases for this date range for this county. |
| US.5 | As a student, I can view daily average temperature in my county of interest over a date range in order to understand the trend of average temperature for this date range for this county. |

| | |
|--------------|--|
| US.6 | As a student, I can view daily absolute humidity in my county of interest over a date range in order to understand the trend of absolute humidity for this date range for this county. |
| US.7 | As a student, I can view daily relative humidity in my county of interest over a date range in order to understand the trend of relative humidity for this date range for this county. |
| US.8 | As a student, I can view a scatter plot of the number of daily COVID-19 cases in my county alongside any of my weather data points over the date range in order to visualize the relationship between the 2 data points for this date range. |
| US.9 | As a student, I can view the correlation coefficient with the scatterplot in order to understand the strength of the relationship between the 2 variables in the scatterplot. |
| US.10 | As a student, I can adjust the date range for all data points. |
| US.11 | As a student, I can view all the same daily data points rolled up to weekly averages in order to smooth out the data. |
| US.12 | As a student, I can view all the same daily data points converted to 7 day rolling averages in order to smooth out the data. |
| US.13 | As a student, I can share/save my current views of the data with another person by sharing my current URL in order to share evidence of my conclusions. |

A couple user stories to point out are US.11 and US.12. Enabling the user to aggregate the data to weekly averages and 7 day rolling averages help to smooth out the data from large variations that can occur day to day in the COVID-19 dataset. This is important because there can be large spikes that occur in the COVID-19 datasets due to fluctuations in tests performed for a particular county on any given day. Aggregating helps to smooth out these spikes.

Another user story to point out is US.13. Enabling the user to share a link that contains their current application configuration will allow them to gather evidence they can use to support their conclusion. The URL would contain the current selected county, date range, and all the chart settings they have selected.

UI Design

Figures 1 – 6 contain mockups of the UI design. These mockups were used when I was building out the tool. Screenshots from the final product are shown later in the Final Result section.

Figure 1

Home page

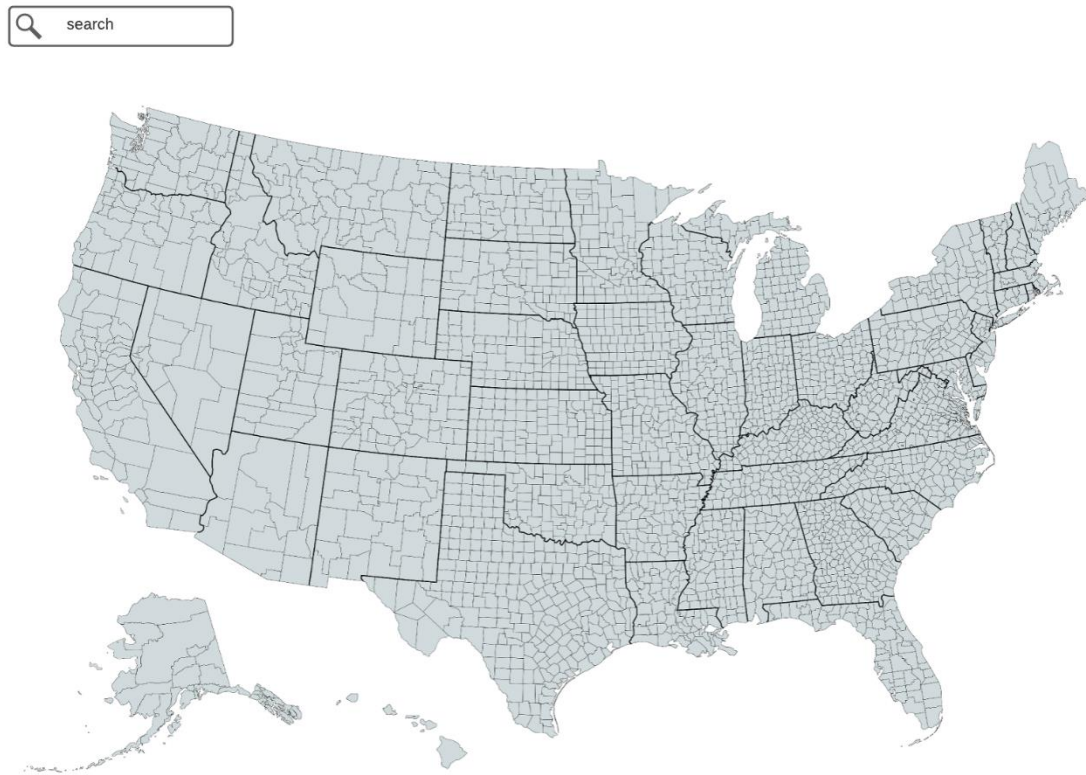


Figure 1 shows a mockup of the home page of the application. This is a map that the user can interact with in order to zoom to or search for their county of interest.

Figure 2

County Selected

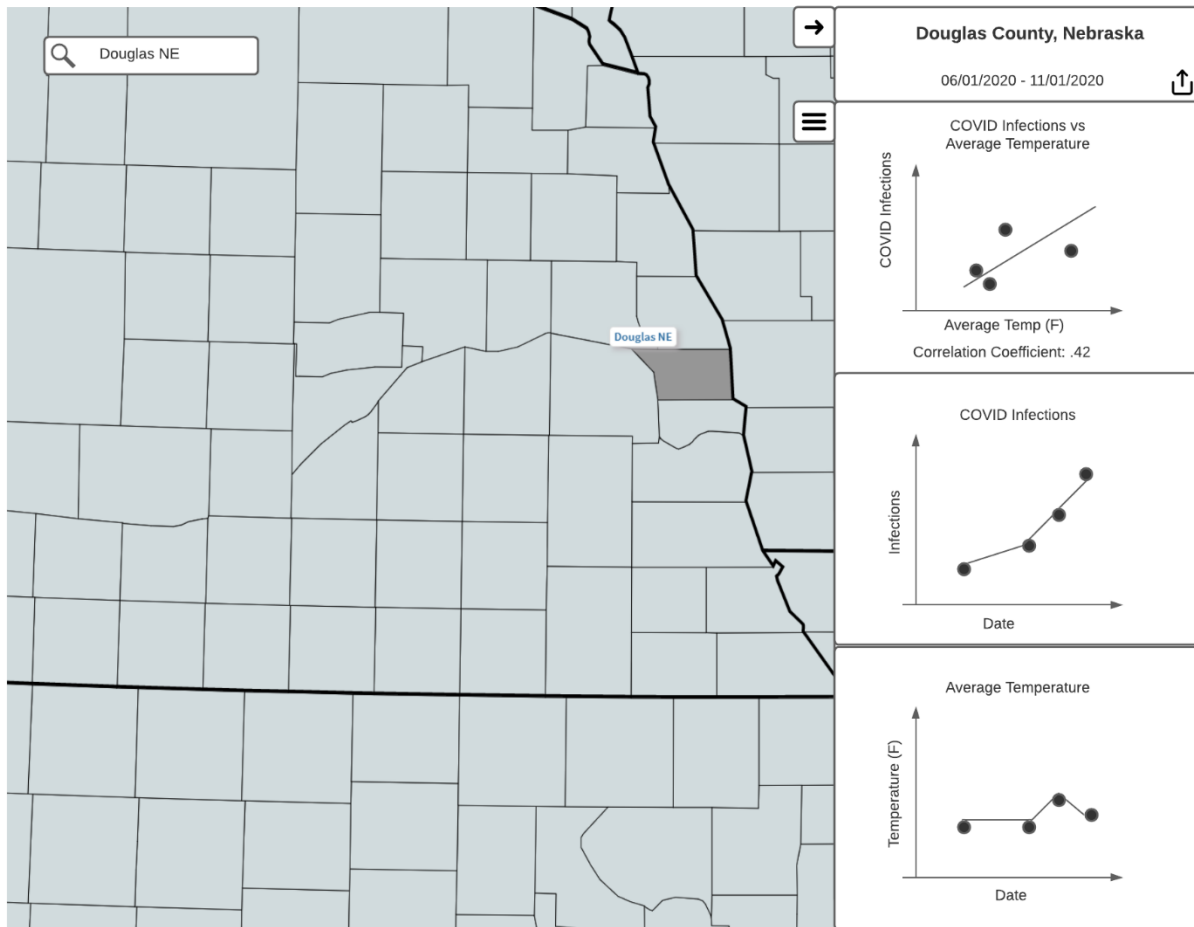


Figure 2 shows a mockup of the panel on the right that would show up once a county is selected, either through the search bar or by clicking on the map. This figure is related to US.3, US.4, US.8, and US.9.

Figure 3

Chart Settings Open

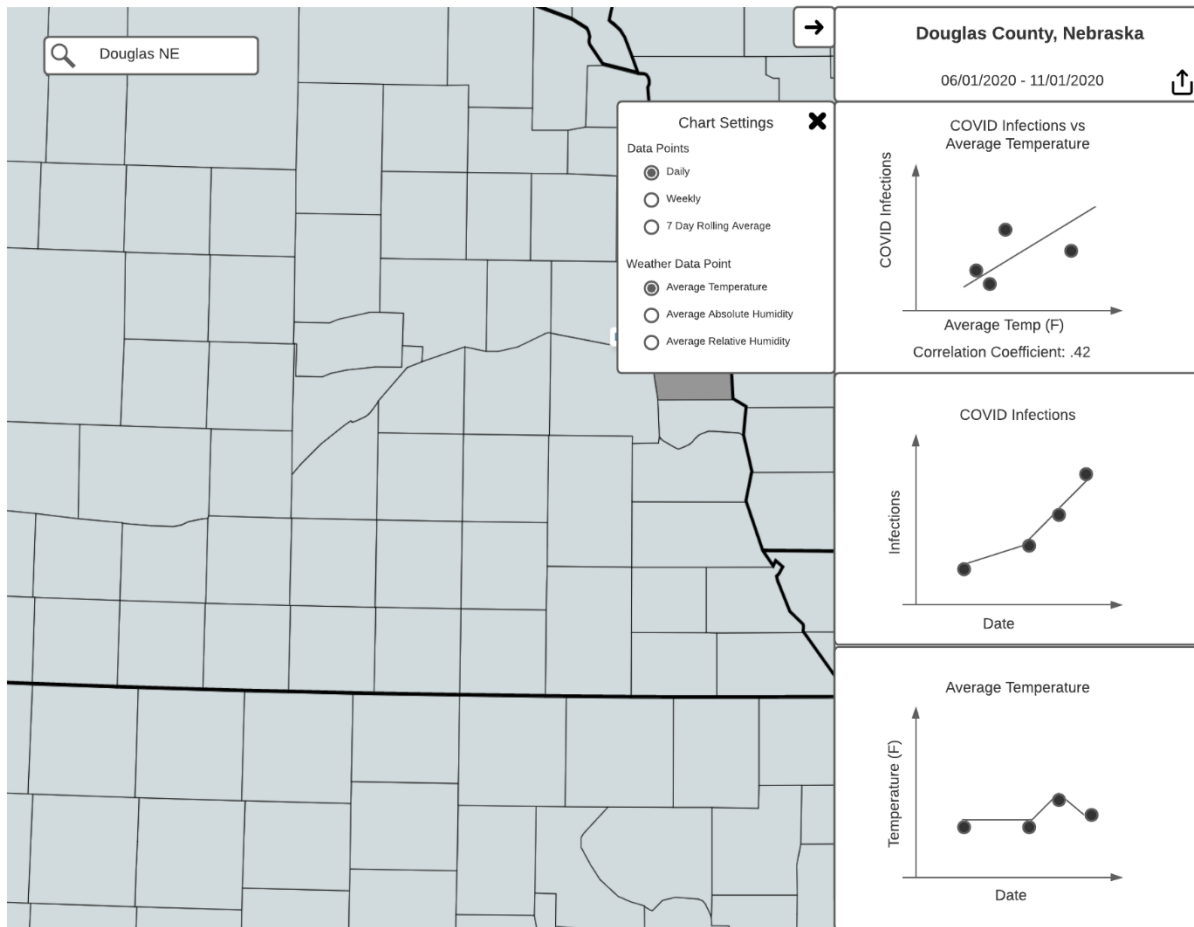


Figure 3 shows a mockup of the chart setting menu. There are options there to change the data points from daily to weekly averages or 7-day rolling averages. This is related to US.5, US.6, and US.7.

Figure 4

Weather Data Point Changed

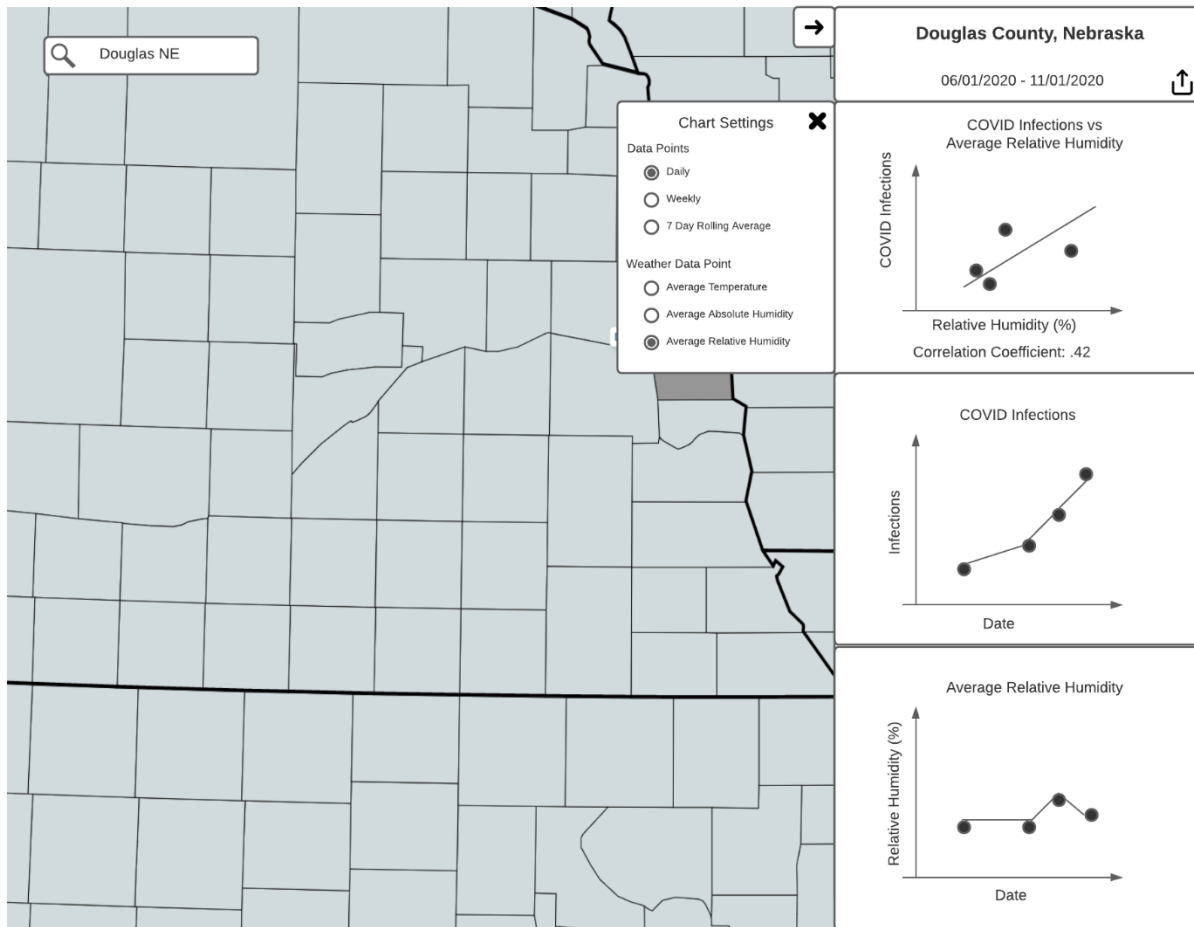


Figure 4 shows a mockup of what happens when a user changes the 'Weather Data Point' chart setting from 'Average Temperature' to 'Average Relative Humidity'. Changing this setting updates the bottom chart as well as the scatterplot on top.

Figure 5

Share Button Clicked

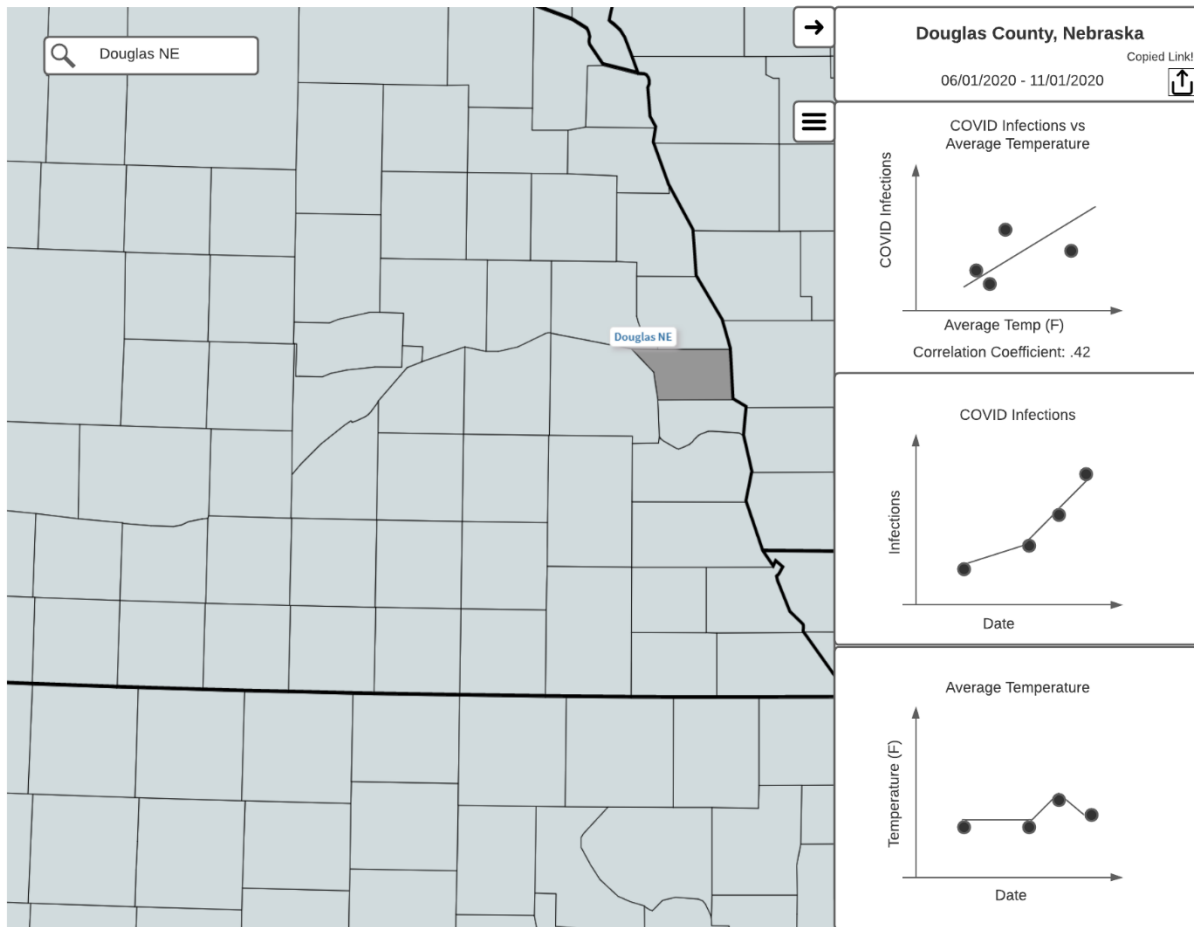


Figure 5 shows a mockup of the sharing feature. Students are able to share URLs of the application in different configurations in order to show someone else what they are seeing.

Figure 6

Chart Panel Closed

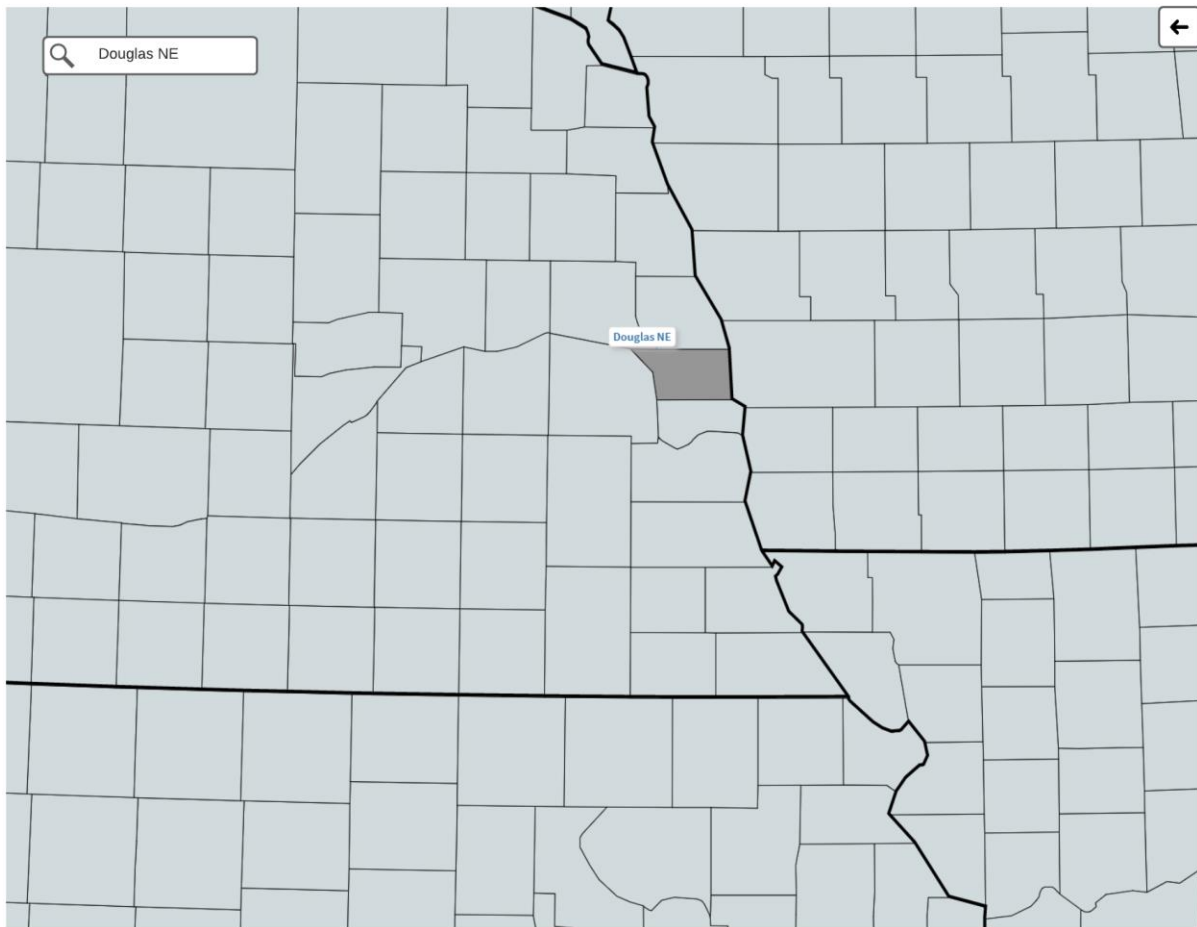


Figure 6 shows a mockup of what the screen looks like after a user closes the panel on the right.

Data Preparation

As previously mentioned, my data source for COVID-19 infections provides cumulative cases by county by day. The first thing I need to do is convert the cumulative cases by day into new cases by day. This can be done by taking the cumulative count for one day and subtracting the cumulative count for the previous day. In order to implement US.11 and US.12 from Table 2 I needed to transform this data. For US.11 I needed to convert the daily case count for a county over a date range into a weekly average case count for that county over that same date range. This was done by averaging the case count every

seven days in order to get the average case count for the week that the 7 days represents. For US.12 I needed to convert the daily case count for a county over a date range into a 7-day rolling average. This can be defined as:

$$7DayAverage[x] = \frac{\sum_{n=x-7}^{x-1} dataByDay[n]}{7} \quad (1)$$

Where x is the day we are calculating for and *dataByDay* is an array of the confirmed cases in a county by day. The same transformations detailed above is used for the weather dataset where I start with daily weather data points that can be rolled up to weekly and 7-day rolling averages.

Resources

Figure 7

Architecture Diagram

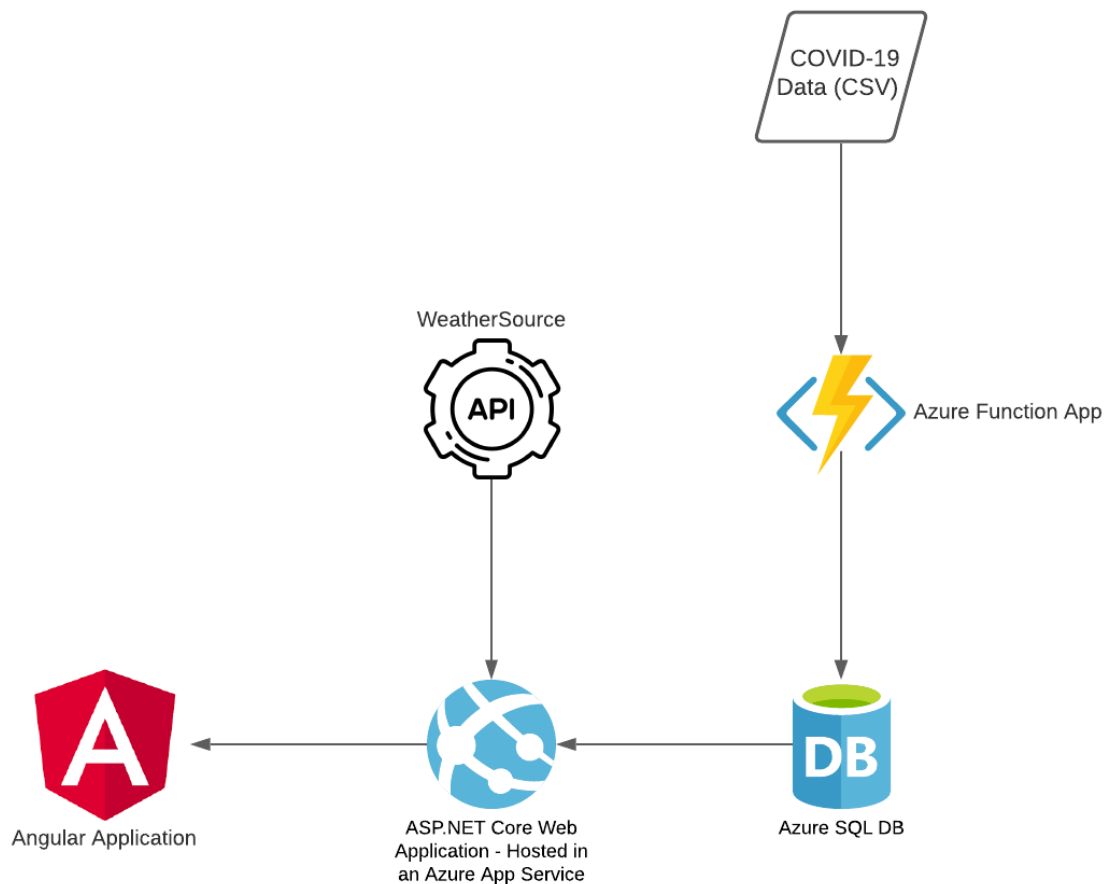


Figure 7 shows an architecture diagram of my project. I created an ASP.NET Core Web Application that serves an Angular frontend application. This ASP.NET Core Web Application is hosted in an Azure App Service via a free student account. The ASP.NET Core application also provides the API endpoints for the Angular frontend. The weather endpoints in my application use the WeatherSource API. The COVID-19 endpoints in my application use an Azure SQL Database that is loaded nightly with the latest CSV dataset from the New York Times. This nightly load of the CSV data into an Azure SQL database is done with an Azure Function App setup with a scheduled trigger.

For the ASP.NET Core Web Application I am using .NET 5. For the Angular Application I am using Angular 11. For the Azure Function App I am using .NET Core 3.1 and Azure Functions runtime version 3. The Azure SQL Database is a fully managed platform as a service and would be using the latest stable version of SQL Server.

I am leveraging quite a few different libraries in each part of my system that I have built. My ASP.NET Core Web Application uses: EntityFrameworkCore¹ as an object relational mapper, AutoMapper² for mapping between different types of objects, NUnit³ as a framework for writing unit tests, FakeItEasy⁴ as a library for mocking dependencies in unit tests. My Angular application uses: Observable-Store⁵ for state management, Font-Awesome⁶ for icons, NG Bootstrap⁷ for several widgets, Chart.js⁸ for creating charts, jStat⁹ for calculating the Pearson Correlation Coefficient, ngx-spinner¹⁰ for my loading icons, and the ESRI Javascript API¹¹ for the mapping functionality. My Azure Function App uses: EntityFrameworkCore as an object relational mapper, MoreLINQ¹² for extended LINQ functions in C#, and CsvHelper¹³ for reading the COVID-19 CSV data. A full list of dependencies can be found in the source code located here: <https://github.com/npalacio/covid-and-weather-data-visualization>.

¹ <https://docs.microsoft.com/en-us/ef/core/>

² <https://automapper.org/>

³ <https://nunit.org/>

⁴ <https://fakeiteasy.github.io/>

⁵ <https://github.com/DanWahlin/Observable-Store>

⁶ <https://fontawesome.com/>

⁷ <https://ng-bootstrap.github.io/#/home>

⁸ <https://www.chartjs.org/>

⁹ <https://github.com/jstat/jstat>

¹⁰ <https://www.npmjs.com/package/ngx-spinner>

¹¹ <https://developers.arcgis.com/javascript/latest/4.19/>

¹² <https://github.com/morelinq/MoreLINQ>

¹³ <https://joshclose.github.io/CsvHelper/>

Final Result

In this section I will walk through what the final application looks like and how it can be used. As of August 5th, 2021 the live site is hosted here: <https://covid-weather-visualization.azurewebsites.net/>. The code repository can be found here: <https://github.com/npalacio/covid-and-weather-data-visualization>.

Figure 8

Home Page

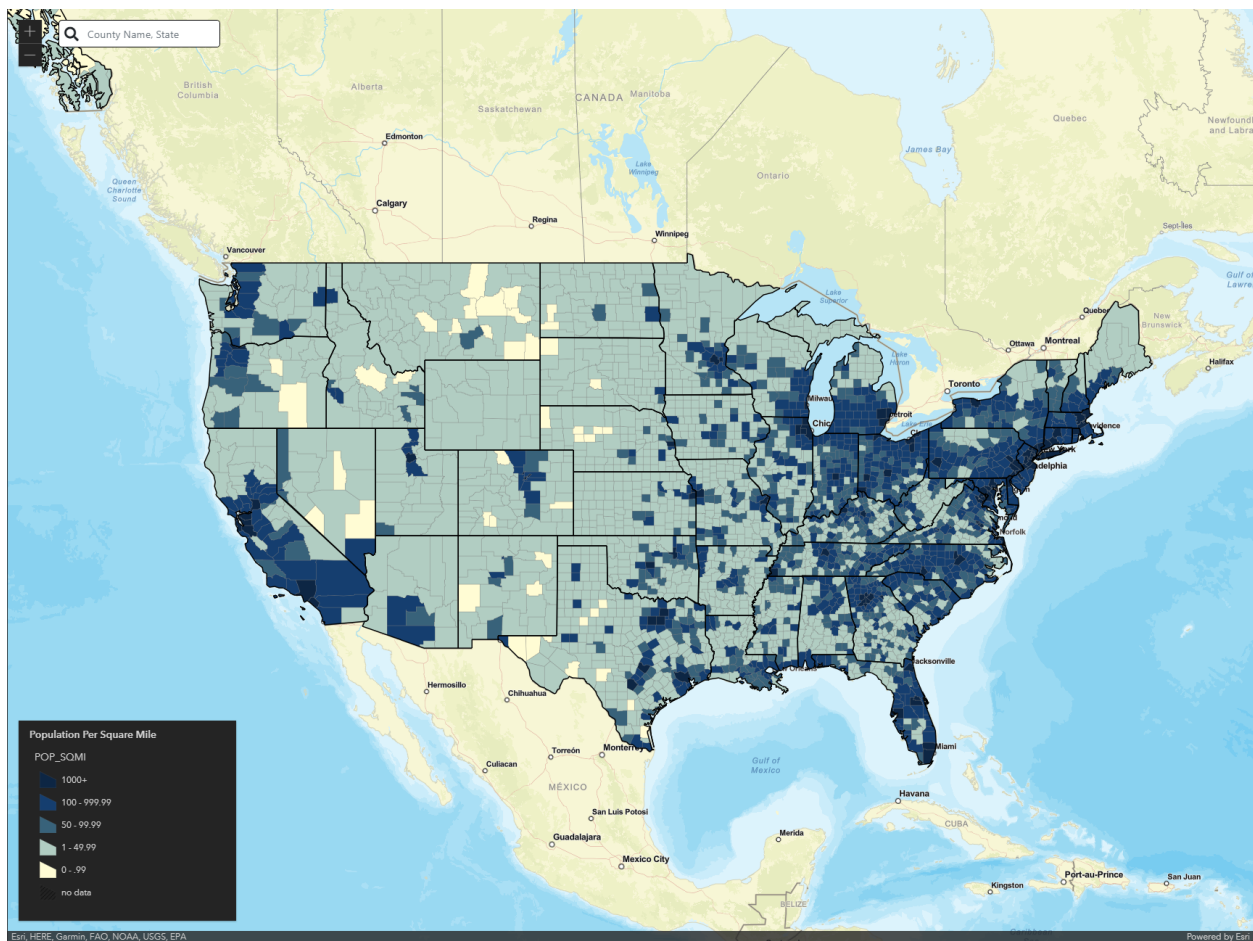


Figure 8 shows the home page of the application. This is a map of most of the United States. There are state and county outlines. The map is symbolized based on the population density of each county with a legend in the bottom left. Symbolizing the counties on population density adds some visual interest to the map. It allows users to compare and contrast how the data in the charts shown

later can differ based on the population density of the county they select. Users might also notice that counties with very low population densities have less COVID-19 infection data to plot. This feature was not in the original list of user stories but instead was a good recommendation from my graduate committee. Once a user gets here, they need to select a county. This can be done one of two ways.

Figure 9

County Selected with Mouse Click

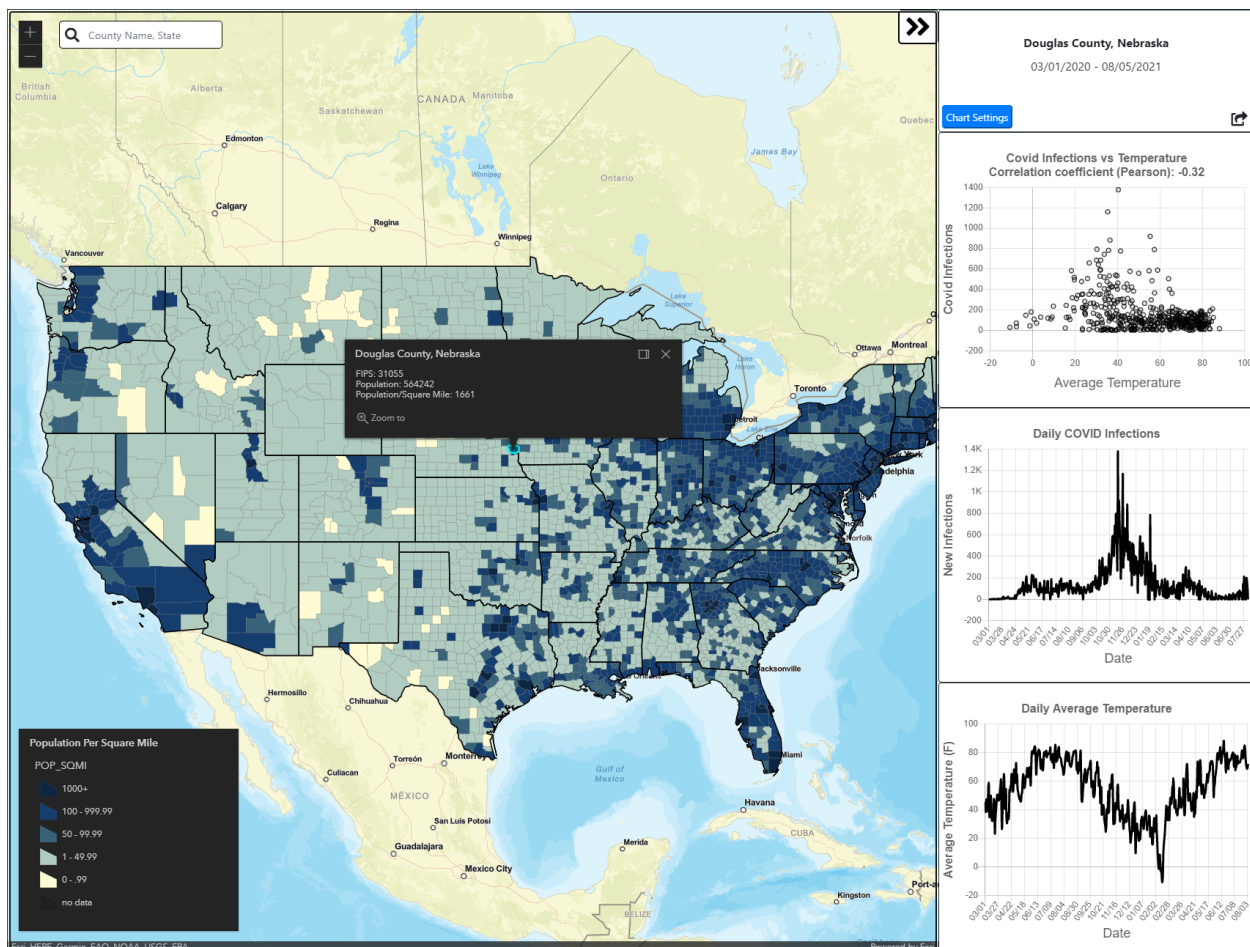


Figure 9 shows what the application looks like after a user selects a county by clicking on one on the map. The panel on the right starts showing up with the 3 charts previously discussed. The top of the panel shows the current selected county and current date range.

Figure 10

Searching for County

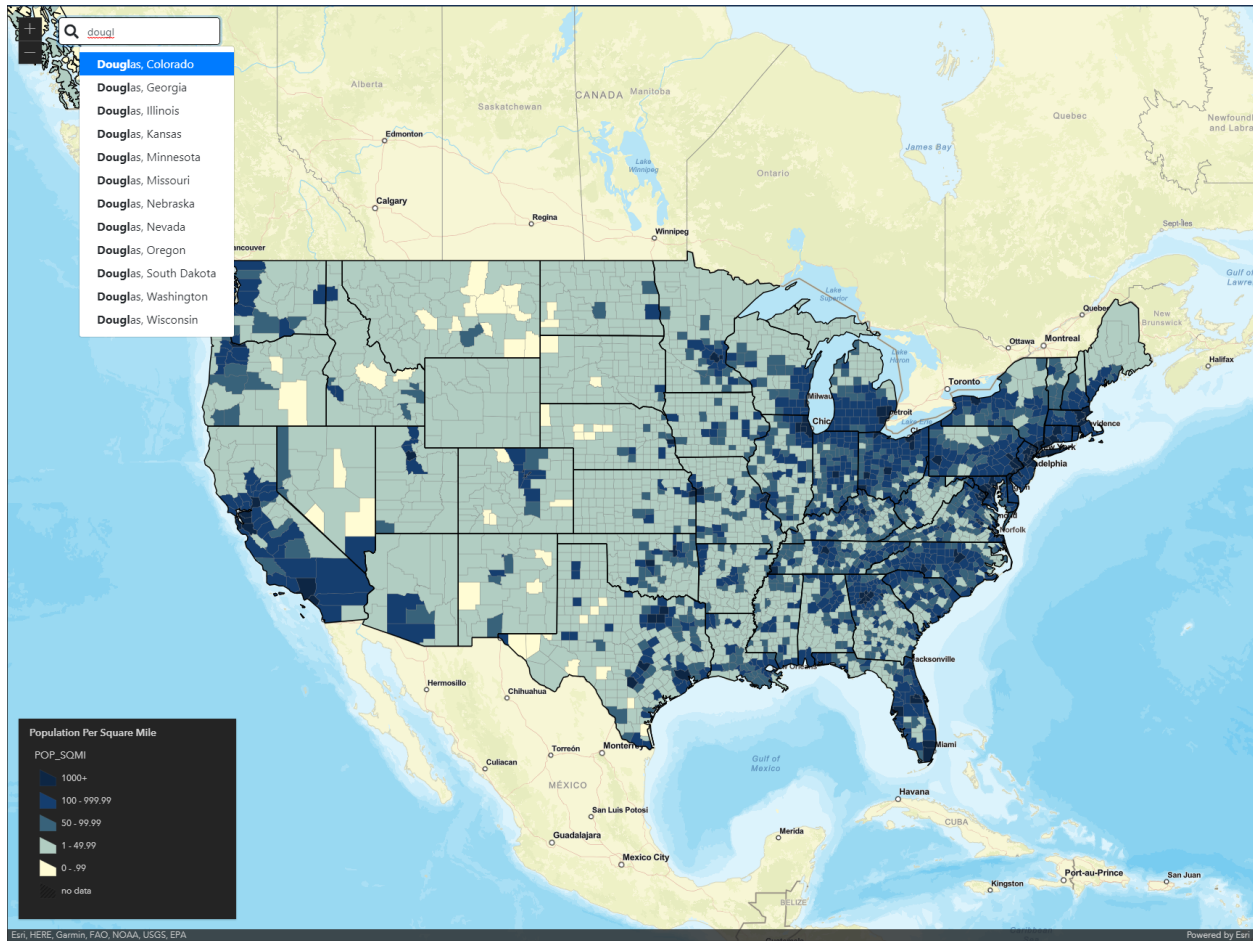


Figure 10 shows the search feature. An alternative way of selecting a county is to search for it by name. The search feature provides auto complete in order to show the user a potential list of counties they are interested in without having to type in the full name. Once a user selects an item from the auto complete list, they would see what is being shown in Figure 9.

Figure 11

Chart Settings Menu

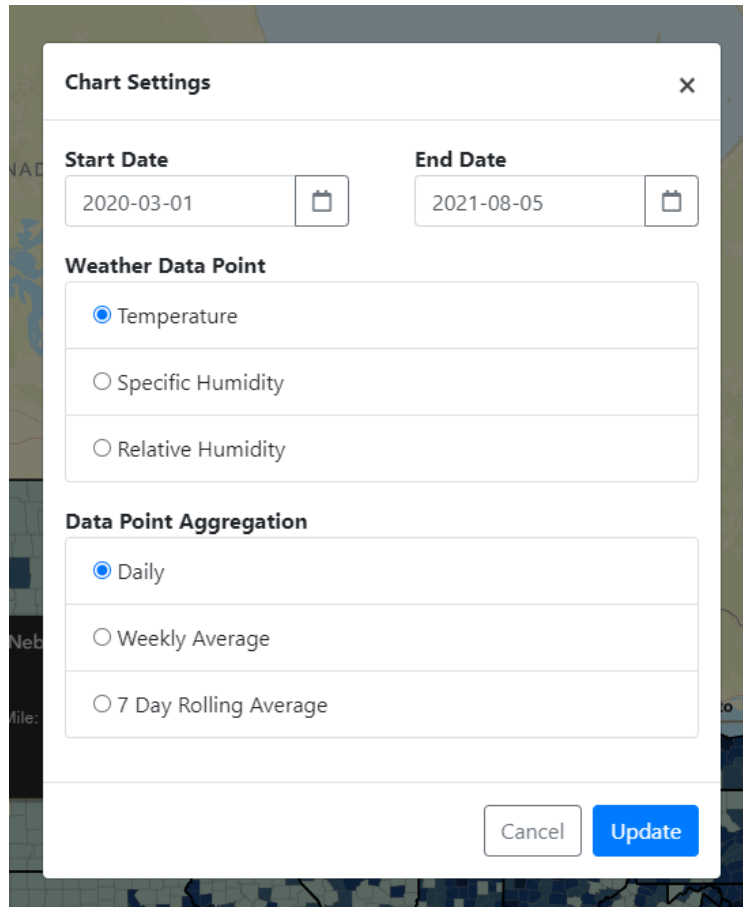


A screenshot of a 'Chart Settings' dialog box. The dialog has a title bar with 'Chart Settings' and a close button (X). It contains three sections: 'Start Date' and 'End Date' with date pickers showing '2020-03-01' and '2021-08-05' respectively; 'Weather Data Point' with radio buttons for 'Temperature' (selected), 'Specific Humidity', and 'Relative Humidity'; and 'Data Point Aggregation' with radio buttons for 'Daily' (selected), 'Weekly Average', and '7 Day Rolling Average'. At the bottom right are 'Cancel' and 'Update' buttons.

Chart Settings X

Start Date 2020-03-01 

End Date 2021-08-05 

Weather Data Point

☒ Temperature

☐ Specific Humidity

☐ Relative Humidity

Data Point Aggregation

☒ Daily

☐ Weekly Average

☐ 7 Day Rolling Average

Cancel Update

Figure 11 shows the chart settings menu. This allows a user to change 3 different settings: the date range for data, which weather data point to use, and how to aggregate the data.

Figure 12

Panel Comparison After Updating Chart Settings

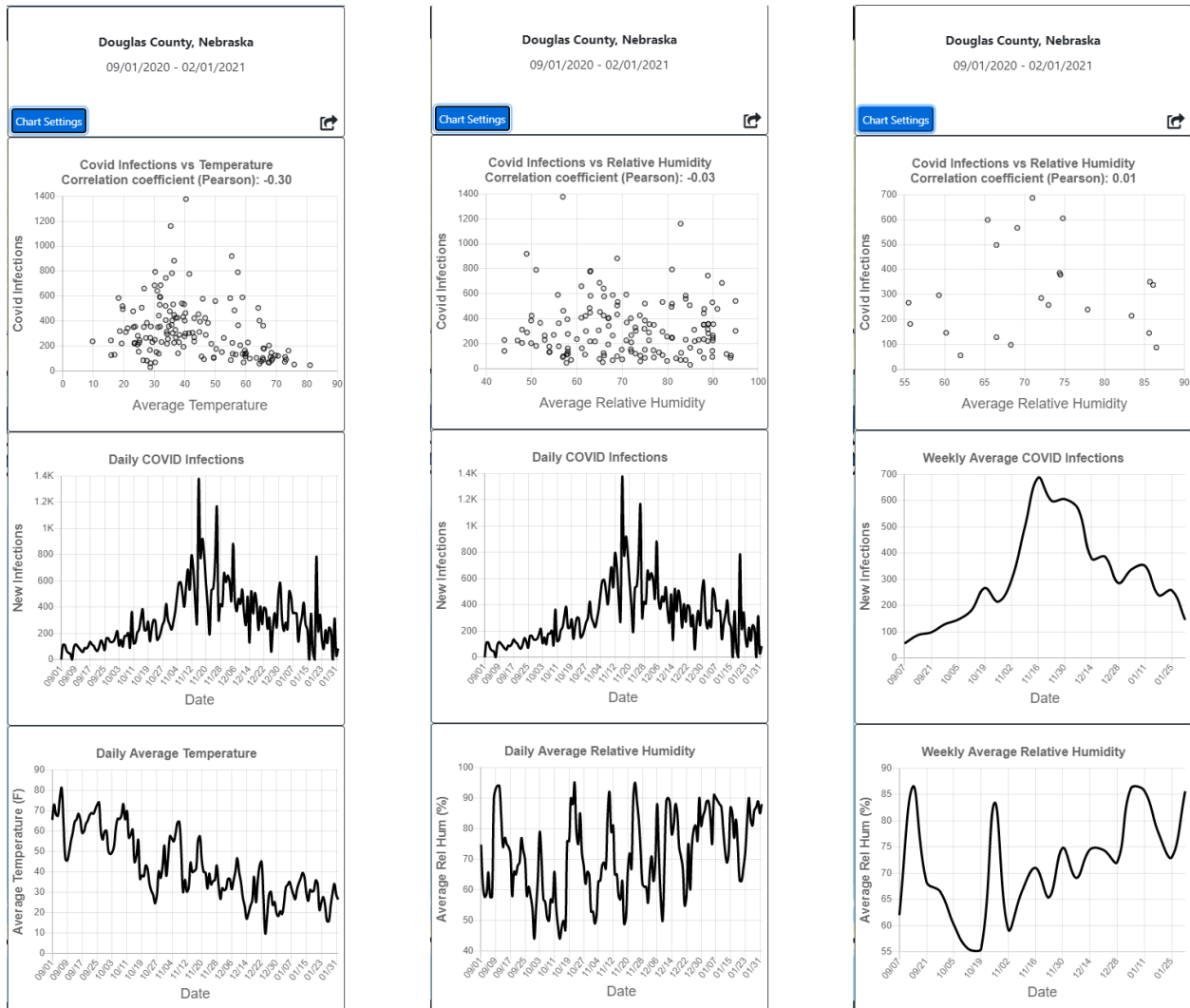


Figure 12 shows a side-by-side comparison of what the panel on the right side of the application looks like when updating chart settings. On the left it shows what the panel looks like when updating the date range. The date range in the top part of the panel reflects what the user has selected and all 3 charts dynamically update to show data for that new date range. In the middle it shows what the panel looks like when updating the selected weather data point from temperature to relative humidity. The top and bottom charts both update to reflect this change. On the right it shows what the panel looks like

when updating the data point aggregation from daily to weekly averages. All 3 charts update to reflect this change.

Figure 13

Sharing Feature

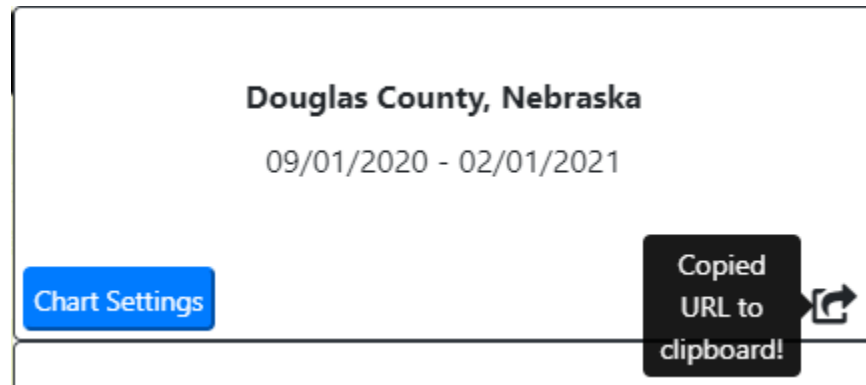


Figure 13 shows the sharing feature. There is a share icon in the top part of the panel that a user can click in order to share their work. This feature copies the user's current URL to the clipboard. The current application configuration (selected county, date range, weather data point, data point aggregation) is all encoded in the URL so that a user can show what they are seeing to someone else. A user could give that link to someone else and they would open up the application and see exactly what that user is seeing.

Code Repository

All of the code written for this project can be found in this public GitHub repository:

<https://github.com/npalacio/covid-and-weather-data-visualization>. The source code can be found

broken out in the following folders under *Source/*:

- *CovidAndWeatherVisualization*: This contains the ASP.NET Core Web Application as well as the Angular Application.
- *CovidDataLoad*: This contains the Azure Function Application.
- *Database*: This contains all the database objects created in the Azure SQL Database.

When building this project, I used a feature branch workflow in git. The only long living git branch was main. When I was picking up a new story, I would create a feature branch off of main, develop the code and then create a Pull Request (PR) on GitHub to review my changes before merging them into main.

Deployment

Continuous integration and continuous deployment (CI/CD) pipelines were setup using GitHub Actions for the Web Application and the Azure Function App. These pipelines can be found under the *.github/workflows* folder. These pipelines build the source code as well as run some additional checks such as unit tests and linting before deploying the applications to Azure. These pipelines were configured to run whenever a PR was created for the main branch as well as after a PR was completed and code was merged into main. Code would only be deployed to Azure after being merged into main, i.e. the PR completed successfully. This helped to catch mistakes and errors before deploying changes to the live site.

I did not set up a CI/CD pipeline for my Azure SQL Database. The only tool I was familiar with for doing this was RedGate Change Automation which I could not get free student access to. Due to time constraints and the fact that I only have a handful of database objects I made the decision not to try and learn and setup a new tool for this. As a compromise I manually checked in my database objects to source control in the *Source/Database* folder.

Future Work

Evaluation

The next step in this project would be getting it in front of students. There are some questions that need answered about the usability of the application, such as:

- Are students able to navigate the interface?
- In what ways are students able to interact with the application in order to select different counties?
- Do students pick up on the meaning of the Pearson correlation coefficient shown?

In order to try and answer these questions there would need to be some sort of classroom study done. This usability test should include tasks for the students where they are asked to do things like:

- Select a county
- Change the date range for the data being shown
- Change the selected weather data point from temperature to relative humidity
- Change the data point aggregation from daily to weekly averages
- Select a different county
- Save and share evidence they have found in the tool

A usability test with these tasks would help to highlight issues in the current implementation. This would help to inform improvements that could be made to things like: how a user selects a county, how a user updates chart settings, and how a user saves and shares their work.

It might also be useful to perform an eye tracking study with users in order to see how they reason about the data. A study like this would provide insights into things such as:

- What are users looking at in the application?
- How do users visually parse the page?

Based on the results of a study like this, improvements could be made to the current implementation's visualization of COVID-19 and weather data.

Additional Feature Work

The current design of the chart settings feature needs to be reworked. With the current design users are forced to click a button first in order to open up the chart settings modal. A better design would allow a user to update chart settings without that extra click. A better design might also place those controls closer to the charts that they update as opposed to a separate modal. This would make it more apparent to the user that these controls are related to the charts.

Another important next step would be to better capture the spatial component of this data. Currently you can only view data in the application one county at a time. This is a significant limitation because it makes it difficult to understand how the data differs by location. Allowing users to view data for multiple counties at a time would be a good next step.

Another limitation with the current implementation related to location is that weather data is continuous by nature but I represent it as a discrete value by fetching weather data using the centroid of

the county polygons. This is an approximation of weather in that county. This becomes a much less accurate approximation as the county gets larger because weather can vary more.

Adding US cities to the map would be another good addition to the application. Currently the map shows state and county borders. Adding US cities would enable a user to better orient themselves in the map. In order to increase the interpretability of the scatter plot a trendline should be added. This trendline would be another visual indicator as to how correlated the selected data sets are.

Conclusion

Overall, the implementation of this project went according to plan. I was able to build out all the user stories that I had proposed. The biggest challenges I ran into were normal software development challenges. By that I mean running into errors that take a while to track down, spending time reading documentation in order to learn how to leverage a new tool, etc.

I had to make tough decisions throughout the project to not solve some problems due to time constraints. For instance, my Azure Function App always fails the first time it runs every day due to some Azure SQL Database connection error. However, simply retrying the function right after it fails works just fine. I had to make a judgement call to not spend any more time looking into that error and simply setup a retry policy on the function so that if it fails it will retry immediately. This got it to a point where it was successfully running once a day. I created a bug card on my GitHub project board for this issue. I also had to decide not to spend time setting up CI/CD for my Azure SQL Database so that I could work on other more important parts of the project.

My goal with this project was to build a visualization tool that could be leveraged in a classroom in order to allow students to participate in an open scientific debate related to weather's role in the COVID-19 pandemic. I accomplished this goal. I have a fully functioning web application that shows

weather and COVID-19 data and allows a user to visualize their relationship. This application is ready for the next steps of usability studies and additional enhancements.

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>

AP. (2020, April 9). Dr. Fauci: Don't assume coronavirus fades in warm weather. ABC7 New York.

<https://abc7ny.com/6089537/>

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/>

Karapiperis, C., Kouklis, P., Papastratos, S., Chasapi, A., Danchin, A., Angelis, L., & Ouzounis, C. A. (2021). A Strong Seasonality Pattern for Covid-19 Incidence Rates Modulated by UV Radiation Levels.

Viruses, 13(4), 574. <https://doi.org/10.3390/v13040574>

Lee, Victor R, and Michelle H Wilkerson. "Data Use by Middle and Secondary Students in the Digital Age: A Status Report and Future Prospects," n.d., 43.

Linn, M., Lee, H.-S., Tinker, R., Husic, F., & Chiu, J. (2006). Teaching and Assessing Knowledge Integration in Science. *Science* (New York, N.Y.), 313, 1049–1050. <https://doi.org/10.1126/science.1131408>

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>

Nebraska Department of Education. (2017). Nebraska's College and Career Ready Standards for Science. https://cdn.education.ne.gov/wp-content/uploads/2017/10/Nebraska_Science_Standards_Final_10_23.pdf

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>

Shah, P., & Hoeffner, J. (2002). Review of Graph Comprehension Research: Implications for Instruction. *Educational Psychology Review*, 14(1), 47–69. <https://doi.org/10.1023/A:1013180410169>

Shahzad, K., Farooq, T. H., Doğan, B., Zhong Hu, L., & Shahzad, U. (2021). Does environmental quality and weather induce COVID-19: Case study of Istanbul, Turkey. *Environmental Forensics*, 1–12. <https://doi.org/10.1080/15275922.2021.1940380>

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>.