Weather-Respiratory Diseases Correlation Explorer

Team MMI: Juanyu Zhang, Yunyi Ru

2024-12-19

The project **Weather-Respiratory Diseases Correlation Explorer (click to see our github repo)** investigates the relationship between weather conditions and outbreaks of two major respiratory diseases: Influenza and Respiratory Syncytial Virus (RSV).

Both of these viruses present significant public health challenges, causing significant morbidity and mortality globally each year, particularly in vulnerable populations such as children, the elderly, and immunocompromised individuals. The seasonality and spread of these viruses are thought to be influenced by various climate and weather factors, including temperature, sunshine, precipitation, and wind patterns (Carter-Templeton, et al., 2022; Linssen, et al., 2021). Understanding these relationships is crucial for predicting potential outbreaks and informing public health strategies.

This project aims to integrate weather data and respiratory disease incidence data to explore potential correlations and build interactive dashboards that visualizes these relationships. Disease incidence data are scrapped from CDC RESP-NET API (CDC RESP-NET, CDC RESP-NET API Socrata) and weather data are scrapped from OpenMeteo API (OpenMeteo). CDC RESP-NET data contains weekly disease incident data from 15 different states (Utah, Tennessee, Oregon, New York, New Mexico, Minnesota, Michigan, Maryland, Georgia, Connecticut, Colorado, California, North Carolina, Ohio, and Iowa) and 7 flu seasons (2018 to 2024) for two viruses (Influenza and RSV). Weather data of these 15 states between 2018 to 2024 are scrapped from OpenMeteo, including daily weather conditions like maximum temperature, minimum temperature, precipitation, sunshine duration, daylight duration, wind speed, etc.

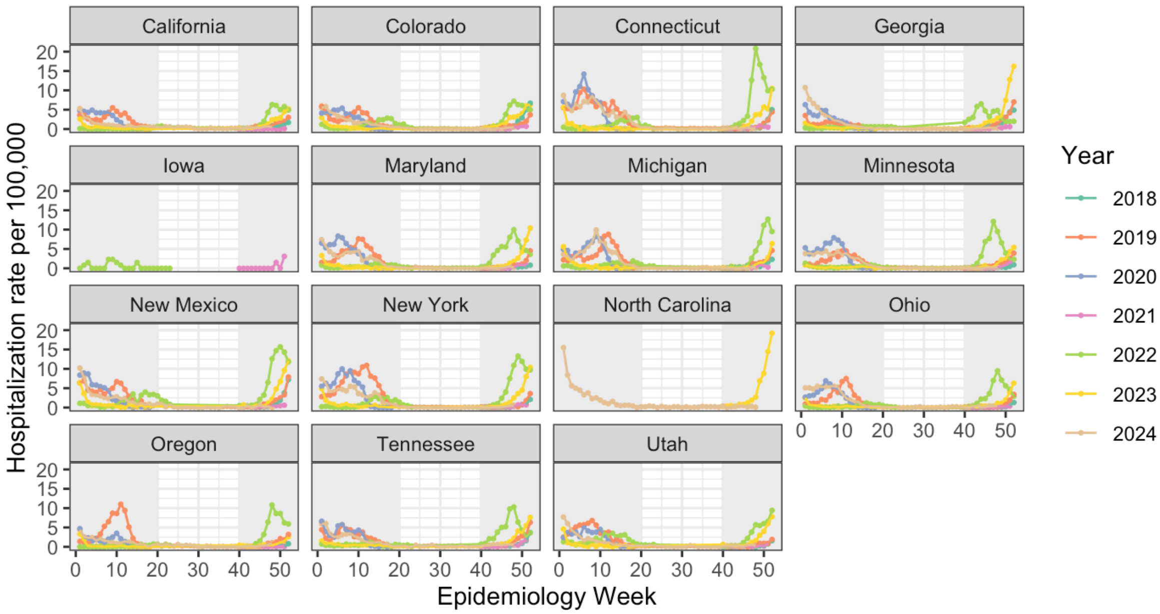
Our final product includes -

1. **Data Analysis**:

- Graphs showing the weekly rates of Influenza-associated and RSV-associated hospitalization by state between 2018 and 2024.

Weekly Rates of Influenza-Associated Hospitalizations by State, from 2018 to 2024

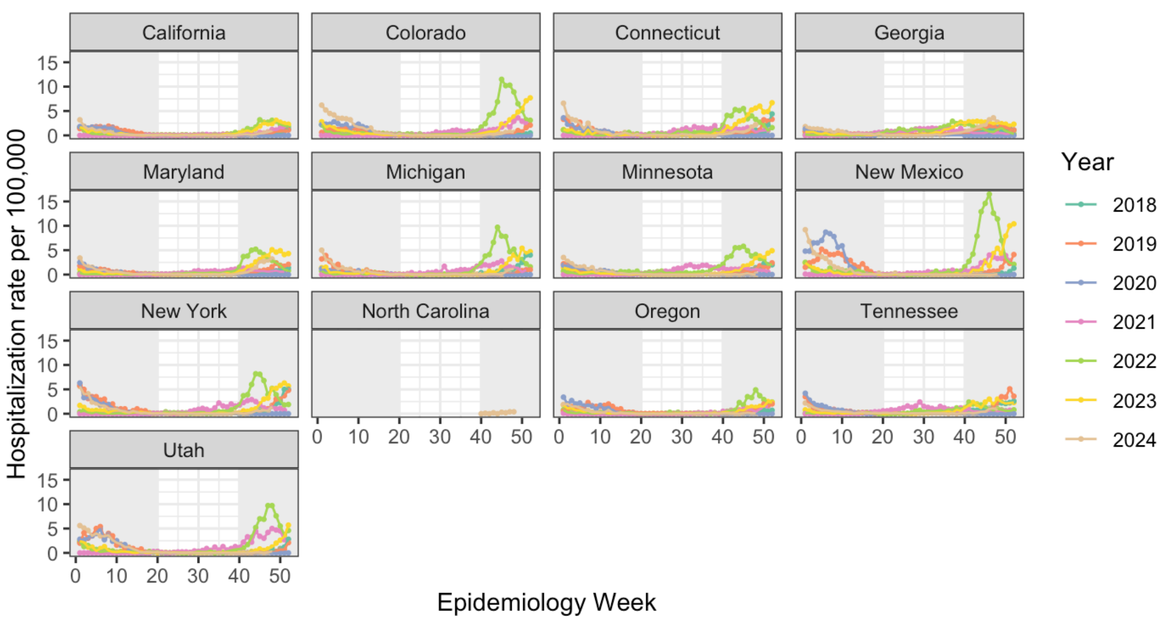
Data Source: CDC RESP-NET



Biostats 140.777 Project 4, Team MMI

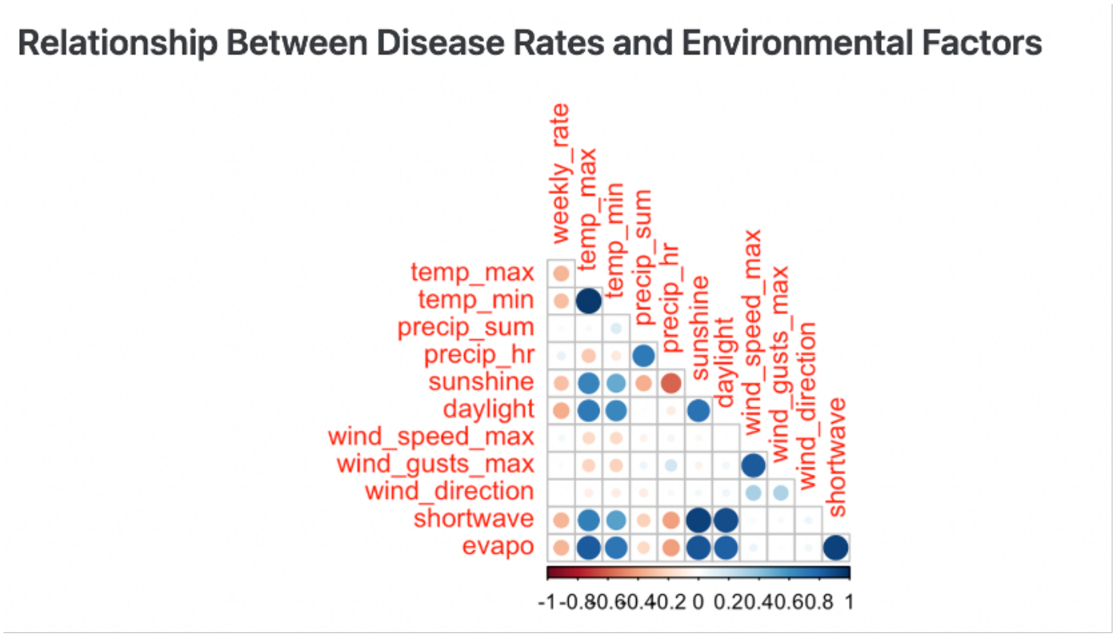
Weekly Rates of RSV-Associated Hospitalizations by State, from 2018 to 2024

Data Source: CDC RESP-NET



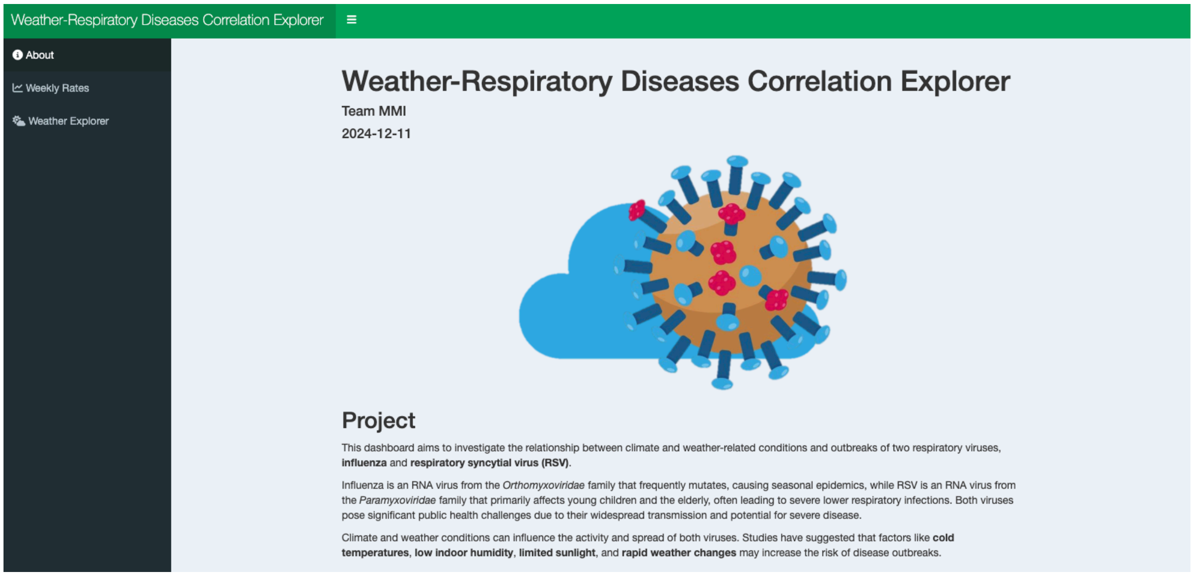
Biostats 140.777 Project 4, Team MMI

- A correlation heatmap showing the relationship between Influenza/RSV and environmental factors.

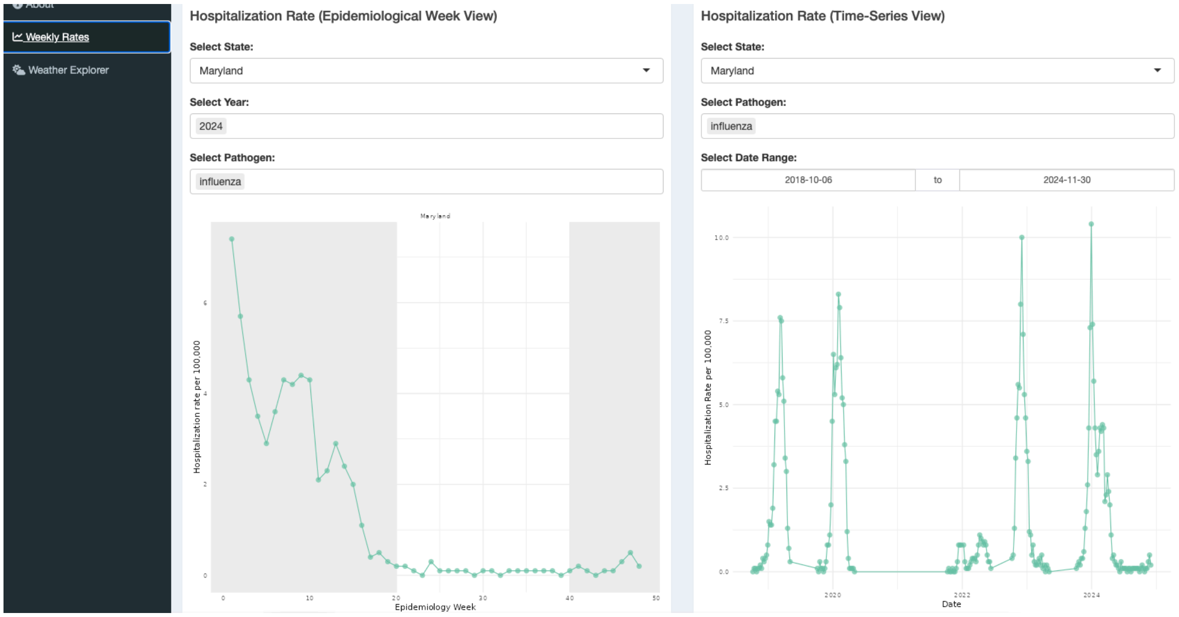


2. **Weather-Respiratory Diseases Correlation Explorer** made by R Shiny, deployed to shinyapps.io, which includes

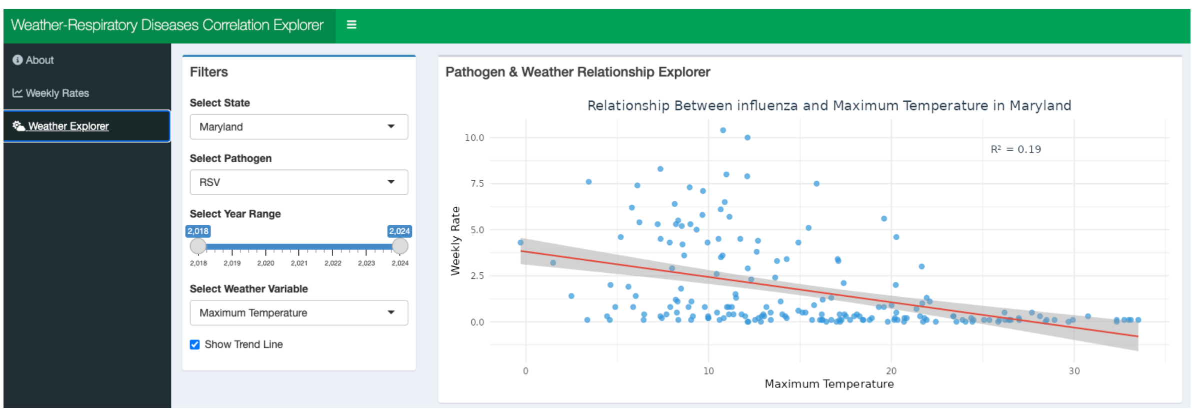
- An 'about' page which summarizes all the important information of this project, including motivation, key features, data source, data dictionary, contact information, and link to source code.



- An interactive dashboard of weekly rates of Influenza/RSV-associated hospitalization, allowing audience to view the weekly hospitalization rate of either or both viruses in certain state and year(s).



- An interactive dashboard showing the correlation relationship between Influenza/RSV and weather conditions, allowing the audience to view the relationship of between weather conditions and either virus in a selected state over a time period of their choice.



During the development of the Explorer, we faced several technical challenges, particularly in data access, integration, and processing. Initially, we attempted to use the National Weather Service (NWS) API for weather data retrieval but found it difficult to work with in R, so we switched to the OpenMeteo API, which provided more accessible and R-friendly data. Open-Meteo imposed a rate limit on API calls, but we addressed it by storing the weather data in

.RDS format locally and checking for existing data before making new requests. Since the OpenMeteo API requires geographic coordinates for data retrieval, we used the geometric center coordinates of each state to query the data. One particularly challenging aspect was handling wind direction, which is circular (0° and 360° represent the same direction). We solved this by using trigonometric functions to compute a correct weekly average. Additionally, dealing with inconsistent time data from both disease and weather sources was made easier with lubridate, which helped us standardize and manipulate dates and weeks for proper alignment. These solutions allowed us to integrate weather and disease data effectively, enabling meaningful correlation analysis and facilitating the development of our interactive dashboard.

Throughout the project, we learned valuable lessons in several areas. First, the process of cleaning and transforming raw data from multiple sources highlighted the importance of strong data wrangling techniques. We became proficient in using R libraries like dplyr, lubridate, and ggplot2 for tasks such as data transformation, integration, and visualization. Also, we learned how to effectively integrate data from APIs, a crucial skill for handling real-time or large-scale datasets. Working with time-series data was another significant learning experience, as we deepened our understanding of temporal correlations and the complexities of aligning multiple time-indexed datasets, such as weekly disease incidence rates and daily weather data. Additionally, we gained hands-on experience in performing correlation analysis, using correlation matrices to visualize relationships between weather and diseases. Data visualization was another critical takeaway. We created both static plots, like line plots and heatmaps, and interactive dashboards, which taught us the importance of clear labeling (title, labels for x- and y- axis, etc.), and user-friendly design in order to communicate complex data with audience. Finally, developing the interactive dashboard using Shiny taught us how to communicate scientific data effectively, reinforcing the significance of making data accessible to both professional and non-professional audiences.

This project effectively integrates two key programming paradigms, functional programming and programming in the command Line. We effectively applied functional programming paradigms to improve coding efficiency and reduce repetition, especially when working with the OpenMeteo API and developing the R Shiny application. For the weather data retrieval, we defined reusable functions to retrieve and process the data, which optimized the code and made it easier to work with by avoiding repetitive tasks. This functional approach not only minimized redundancy but also ensured that each task was modular, making the code easier to debug and extend. Similarly, in the development of the R Shiny application, we used functional programming paradigm to define both the user interface (UI) and server logic. By organizing UI components and server-side calculations in separate functions, we improved the application's structure and flexibility. In addition to functional programming, we also employed programming in the command line to manage version control and facilitate collaboration. Using GitHub for version tracking, we relied on command-line operations like git add, git commit, git push, and git pull to coordinate our work and maintain an efficient workflow. By combining these two paradigms, we were able to create an organized, efficient, and collaborative environment for the project.

This project provides valuable insights on how weather conditions might influence the incidence of influenza and RSV outbreaks. Understanding these relationships allows public health professionals to predict viral activity more accurately and to implement targeted prevention strategies, such as vaccination campaigns or health advisories, in a timely manner. The interactive dashboard also offers wet-lab scientists a tool to identify key environmental factors for further investigation *in vivo* and *in vitro*. By analyzing the correlation between weather conditions and viral activity, researchers can determine which variables to prioritize when designing animal model studies. However, while the study reveals correlations between weather and virus activity, it is important to note that correlation does not imply causation. Numerous other factors—such as human behavior, immunity levels, population density, healthcare infrastructure, and viral mutations—can also influence viral transmission, making it difficult to isolate the direct impact of weather conditions on viral spread. Furthermore, the weather conditions displayed in the dashboard may not fully capture the complexity of real-world settings. For example, colder weather may not only lower temperatures but also lead to increased indoor crowding, both of which can contribute to the spread of viruses, especially respiratory viruses.

Future research should focus on in vivo and in vitro experiments in controlled laboratory settings to investigate the correlation between respiratory disease incidence and weather conditions, using animal models and cell cultures to isolate the impact of single environmental factor. Integrating social and behavioral factors, such as population density and immunity levels, would provide a more complete understanding of how weather conditions affect viral transmission. Additionally, adding viral genomic evolution as a factor could increase the accuracy on revealing how environmental conditions influence viral mutations and spread. Finally, developing a predictive machine learning model that combines weather data, social factors, and viral genetics could enhance disease outbreak forecasting, allowing more effective public health emergency response.

Reference

Carter-Templeton, Heather, Gary F. Templeton, Leslie H. Nicoll, Latrice Maxie, Theresa S. Kittle, Susan A. Jasko, Eric E. Carpenter, and Karen A. Monsen. 2022. "Associations Between Weather-Related Data and Influenza Reports: A Pilot Study and Related Policy Implications." Applied Nursing Research: ANR 67 (October): 151413. https://doi.org/10.1016/j.apnr.2021.151413.

Linssen, Rosalie S., Bibiche den Hollander, Louis Bont, Job B. M. van Woensel, Reinout A. Bem, and null On Behalf Of The Pice Study Group. 2021. "The Association Between Weather Conditions and Admissions to the Paediatric Intensive Care Unit for Respiratory Syncytial Virus Bronchiolitis." Pathogens (Basel, Switzerland) 10 (5): 567. https://doi.org/10.3390/pathogens10050567.