Final Project Proposal: *Weather-Disease Correlation Explorer*

Team: Juanyu Zhang, Yunyi Ru, Yuzi Zhang

2024-11-15

Table of contents

Using quarto, write up a project proposal andsubmit the PDF. The final project proposal should include the following information:

1. The title of your project and the team members names

(As shown above)

2. You should describe a research or data analysis question and explain its importance

We want to investigate the correlation between climate and weather-related conditions and outbreaks of two respiratory viruses, influenza and respiratory syncytial virus (RSV).

Influenza is an RNA virus from the *Orthomyxoviridae* family that frequently mutates, causing seasonal epidemics, while RSV is an RNA virus from the *Paramyxoviridae* family that primarily affects young children and the elderly, often leading to severe lower respiratory infections.

Both viruses pose significant public health challenges due to their widespread transmission and potential for severe disease.

3. You should summarize work that already exists (if it does)

Climate and weather conditions can influence the activity and spread of both viruses. Studies have suggested that factors like cold temperatures, low indoor humidity, limited sunlight, and rapid weather changes may increase the risk of outbreaks (Carter-Templeton et al. 2022) (Chong et al. 2022) (Fdez-Arróyabe et al. 2021) (Linssen et al. 2021).

4. You should outline the work you plan to do

In this project, we plan to integrate weather data and disease incidence data together to perform correlation and regression analysis, and to build an interactive dashboard showing

those relationships. If possible, we also want to build a machine learning model to predict potential disease outbreaks based on weather data.

5. You should demonstrate you have access to the data, describe the data, and propose how you will collect the data

Weather data will be collected through the National Weather Service (NWS) API (US Department of Commerce n.d.) or Openweather API ("Weather API - OpenWeatherMap" n.d.).

Disease incidence data will be collected from the CDC National Notifiable Diseases Surveillance System (NNDSS) ("Notifiable Infectious Disease Tables | CDC" 2023).

All of the data sources mentioned above are open to the public and downloadable from the website.

6. You should describe the programming paradigms you plan to use and why it makes sense to combine them for your project

We plan to use a combination of functional programming and object-oriented programming (OOP) for this project. This paradigm is well-suited for our project because it emphasizes immutability and reusable functions, which align with our need to process data repeatedly across different subsets and analyses. OOP will be used for structuring the machine learning model and data pipeline.

7. You should describe any packages and/or software you plan to use

Data Collection :

- `httr` and `jsonlite`: For fetching weather data via APIs.

Data Analysis:

- `dplyr` and `tidyr`: For advanced data manipulation.
- `stats` and `lmtest`: For performing correlation and regression analyses.
- `lubridate`: For handling temporal data efficiently.

Visualization:

- `ggplot2`: For static and interactive visualizations of weather and disease patterns.

8. You should briefly describe the data analytic product you plan to build.

The final product will be an interactive dashboard, which will include:

- Weather-Disease Correlation Explorer: Users can select specific weather variables (e.g., temperature, humidity) and see their correlation with disease incidence over time.
- Seasonal Trend Analysis: Visualizations of trends for specific locations, highlighting high-risk periods.

- Customizable Reports: Users can generate downloadable summaries of findings in PDF or HTML format.

9. You should describe a tentative timeline for the proposal

- Nov. 18 - Dec. 5: Collect and clean data
- Dec. 5 - Dec. 10: Visualize data

10. You should describe how the tasks will be split amongst the team members

- Data Collection: Yuzi
- Data Analysis and Visualization: Yunyi and Juanyu

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.

Chong, Ka Chun, Yu Chen, Emily Ying Yang Chan, Steven Yuk Fai Lau, Holly Ching Yu Lam, Pin Wang, William Bernard Goggins, et al. 2022. "Association of Weather, Air Pollutants, and Seasonal Influenza with Chronic Obstructive Pulmonary Disease Hospitalization Risks." *Environmental Pollution (Barking, Essex: 1987)* 293 (January): 118480. https://doi.org/10.1016/j.envpol.2021.118480.

Fdez-Arróyabe, Pablo, Alberto Marti-Ezpeleta, Dominic Royé, and Ana Santurtún Zarrabeitia. 2021. "Effects of Circulation Weather Types on Influenza Hospital Admissions in Spain." *International Journal of Biometeorology* 65 (8): 1325–37. https://doi.org/10.1007/s00484-021-02107-y.

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.

"Notifiable Infectious Disease Tables | CDC." 2023. https://www.cdc.gov/nndss/data-statistics/infectious-tables/index.html.

US Department of Commerce, NOAA. n.d. "API Web Service." Accessed November 18, 2024. https://www.weather.gov/documentation/services-web-api.

"Weather API - OpenWeatherMap." n.d. Accessed November 18, 2024. https://openweathermap.org/api.