Data 512: Part 2 - Extension Plan

Introduction

In Part 1 of the Wildfire Analysis Project, we focused on assessing the potential wildfire smoke impact on Mesa, Arizona, spanning from 1964 to 2024. For this, we utilized the US Geological Survey (USGS) to gather historical wildfire information, which then allowed us to estimate smoke exposure in Mesa over the past 60 years. We then validated these smoke estimates against local Air Quality Index (AQI) measurements obtained from the US Environmental Protection Agency (EPA) Air Quality Service (AQS) API. This foundational analysis enabled us to develop a predictive model to forecast the potential smoke estimates in Mesa from 2021 to 2050 (since we were missing data from 2021 to 2024). Now, in this part, the objective is to inform Mesa's city council, city manager/mayor, and residents about the potential future impacts of smoke on their community over the next 25 years. This phase will extend the previous analysis by examining the specific health impacts of wildfire smoke on the Mesa community. We will utilize data from the CDC and the Arizona Department of Health Services (ADHS) to analyze trends in respiratory disease mortality and healthcare utilization. Through this analysis, we aim to provide insights into how wildfires affect public health and healthcare systems in Mesa.

Specifically, the extension plan's goal is to focus on chronic respiratory illnesses and mortality rates, thereby guiding public health strategies and policies for the coming decades. In the following sections, we will explore the problem statement driving this analysis, emphasize the critical role of healthcare in our study, and explain the relevance of our chosen data sources. We will also identify any unknown factors or dependencies that could affect our results and outline a detailed timeline for completing this extension.

Motivation/Problem Statement

As I delved into the health impacts of wildfire smoke on urban communities, I was struck by the growing evidence linking smoke exposure from wildfires to respiratory illnesses. In Mesa, the challenge is particularly pressing due to the increasing frequency of wildfires in the region over the last few decades. From part 1 of this project, we could see that the smoke estimates had steadily increased over the years, and that combined with the increase in the sheer number of wildfires, led me to question how these smoke patterns might be affecting the health, especially respiratory, of the individuals living in Mesa, AZ.

I am motivated to conduct this analysis because it addresses a critical need for actionable insights into how wildfire smoke affects public health in Mesa. By examining data from CDC Wonder and the Arizona Department of Health Services (ADHS), I aim to uncover trends in respiratory disease mortality and healthcare utilization. I will be specifically focusing on asthma-related hospitalizations, tuberculosis cases, the number of respiratory procedures, and mortality due to respiratory issues. My hypothesis is that the higher the smoke estimates (a result of wildfires), the higher the proportion of all aforementioned issues. I think that this analysis is not only scientifically intriguing but also practically essential for informing public health strategies. I believe this work is critical to guide Mesa's city council, city manager/mayor, and residents in planning for potential future health challenges. By focusing on the relationship between wildfire smoke impact and chronic respiratory illnesses and mortality rates, I hope to provide evidence-based recommendations that can help protect vulnerable populations during peak wildfire seasons, given my hypothesis is true.

Overall, through this extension I hope to understand better the effects of wildfire on respiratory health and help the people living in Mesa, AZ, to prepare for and adapt to the health challenges posed by increasing wildfire smoke exposure over the next 25 years. I believe that this analysis will be particularly valuable for healthcare providers and policymakers in developing strategies to help protect residents and ensure that the healthcare systems are adequately prepared for any future challenges.

Impact Focus

For this extension, I have chosen to focus on the healthcare impacts of wildfire smoke in Mesa, Arizona. My analysis aims to examine how smoke exposure affects respiratory health outcomes by analyzing several key indicators like mortality rates due to respiratory diseases, asthma-related healthcare patterns (including both inpatient hospitalizations and emergency room visits), tuberculosis rates, and the volume of respiratory-related medical procedures performed annually. I think this healthcare-focused approach is fundamentally human-centered, as it directly addresses how increasing wildfire smoke exposure affects our community's health and well-being.

Through my analysis, I aim to answer the following questions about the impact of wildfire smoke on health:

1. How, if at all, have respiratory disease mortality rates in Mesa (based on Maricopa County) been influenced by wildfire smoke exposure, and what patterns might emerge through 2050?

2. What is the relationship between smoke exposure and healthcare utilization, specifically examining:
   a. Trends in asthma-related inpatient hospitalizations.
   b. Patterns in emergency department visits for respiratory issues.

3. How has the number of respiratory-related medical procedures changed over time with increased smoke exposure? Is there a relationship between the two?

4. Is there a correlation between wildfire smoke exposure and tuberculosis cases? (Note: Recent studies have shown that residential exposure to wildfire-associated ambient air pollution is associated with an increased risk of active tuberculosis diagnosis).

Data

In Part 1 of the project, we utilized the below two data sources for our analysis and forecasting. We would also need to use that data, for this extension. The sources from Part 1 are:

- **USGS Wildland Fire Data:** This dataset provides comprehensive information on wildfires across the United States from the 1800s to 2020. We specifically use data from 1964 to 2020 for our analysis. It is publicly available and not subject to copyright restrictions. It is marked with "CC0 1.0 Universal." ([Link Here](#))

- **EPA Air Quality Index (AQI) Data:** This data was accessed through the EPA's Air Quality System API, and contained information on the pollutants and aqi. For the analysis of my city, Mesa, we found data from 1965 to 2024. This data is in the public domain and not subject to domestic copyright protection under 17 U.S.C. § 105. ([Link Here](#))

In this extension, I incorporate several additional datasets that provide insights into mortality rates due to respiratory diseases, asthma-related healthcare patterns (including both inpatient hospitalizations and emergency room visits), tuberculosis rates, and the volume of respiratory-related medical procedures performed annually.

CDC Wonder Mortality Data
- This dataset contains detailed information about mortality from 1999 - 2022 and provides the options to filter it by county, cause, year, month, etc. It has the ability to create datasets on a combination of conditions and I will be using this to get data about mortality for Mesa, AZ based on the county 'Maricopa' and the causes filter to respiratory issues. I will also use the system to ensure that I get data by month.

- A description of the various elements in the data and the schema is available [here](#).
- Fields of Interest: County, Year, Month, Cause of Death, Age Group, Gender
- Link: https://wonder.cdc.gov/mcd.html
- License and data restrictions:

"The Public Health Service Act (42 U.S.C. 242m(d)) provides that the data collected by the National Center for Health Statistics (NCHS) may be used only for the purpose for which they were obtained; any effort to determine the identity of any reported cases, or to use the information for any purpose other than for health statistical reporting and analysis, is against the law. Therefore users will:

- *Use these data for health statistical reporting and analysis only.*
- *Do not present or publish death counts of 9 or fewer or death rates based on counts of nine or fewer (in figures, graphs, maps, tables, etc.).*
- *Make no attempt to learn the identity of any person or establishment included in these data.*
- *Make no disclosure or other use of the identity of any person or establishment discovered inadvertently and advise the NCHS Confidentiality Officer of any such discovery."*

The above data restriction was taken as is from the CDC Wonder website. Additional conditions for access can be found [here](#).

Arizona Department of Health Services

The Arizona Department of Health Services (ADHS) is the state's public health agency responsible for overseeing a wide range of health-related programs and data collection efforts. For this analysis, I will utilize several datasets provided by ADHS to explore the healthcare impacts of wildfire smoke on respiratory illnesses in Mesa. Specifically, I will gather data on asthma-related inpatient discharges and emergency visits (2000-2021), medical procedures categorized by type (2000-2021), and tuberculosis cases by county (2009-2023). These datasets will help provide a comprehensive view of how respiratory health outcomes have evolved in response to environmental factors such as smoke exposure due to wildfires.

1. **ADHS Asthma-Related Inpatient Discharges and Emergency Visits Data:**
- This dataset includes information on asthma-related inpatient discharges (2000-2021) and emergency visits (2003-2021) across Maricopa County (where Mesa is located). It provides demographic breakdowns by race/ethnicity, gender, age group, and county of residence.
- Fields of Interest: County, Year, Gender, Age Group, Number of Discharges, Number of Emergency Visits
- Link: https://pub.azdhs.gov/health-stats/hip/index.php?pg=asthma
- Note: This data also contains information about whether Asthma was the first-listed diagnosis. I am not sure about if and how I will us yet.

2. **ADHS Medical Procedures Data:**
- This dataset includes county-level information about the number of inpatient discharges by different medical procedures performed annually from 2000 to 2021 by procedure category. It includes procedure categories such as respiratory therapy, which can be used to understand whether wildfires are leading to an increase in respiratory procedures.
- Fields of Interest: County, Year, Procedure Type, Number of Procedures
- Link: https://pub.azdhs.gov/health-stats/hip/index.php?pg=procedure

3. **Tuberculosis Cases by County Data:**

- This dataset tracks tuberculosis cases in Maricopa County from 2009 to 2023. It contains information on the number of cases as well as the incidence rate per 100,000. This will be useful to correlate if an increase in wildfire is causing an increase in tuberculosis cases.
- Fields of Interest: County, Year, Number of Cases, Incident Rate
- Link:https://www.azdhs.gov/preparedness/epidemiology-disease-control/disease-integration-services/tb-control/#tb-control-data-reports

For all data from ADHS, the license requirements are that it is publicly available for non-commercial use with appropriate attribution to ADHS.

Note: The Arizona Department of Health Services (ADHS) does not explicitly provide a specific license page for its data. However, it is stated on various health data pages that ADHS makes its health data publicly available for non-commercial use with the condition that proper attribution is given.

Unknown and Dependencies

1. **Incomplete Data Coverage Across Years:** The healthcare data from the Arizona Department of Health Services (ADHS) does not cover all years consistently across all datasets. For example, the Asthma-related inpatient discharges are available from 2000 to 2021, but the emergency visits data is only available from 2003 to 2021 and the Tuberculosis data is available from 2009 to 2023. This inconsistency limits the ability to conduct a long-term analysis of trends across all health outcomes. Some years may lack data for certain conditions, making it difficult to observe continuous trends or correlations with wildfire smoke exposure over time.

2. **Data Format Issues (PDF vs. CSV/Excel):** Some of the data (depending on the year) provided by ADHS is available in PDF format rather than structured formats like CSV or Excel. For example, some Tubercolosis data is only accessible in PDF reports. Due to this, converting PDF data into a usable format (such as CSV or Excel) requires additional manual processing, which could introduce errors or delays in the analysis process.

3. **Changing Categories Over Time:** The categorization of healthcare data (e.g., procedures, and diagnoses) may change over time due to updates in medical coding systems (such as ICD-9 to ICD-10) or changes in how ADHS reports certain health outcomes. These potential changes in categorization may result in inconsistencies when comparing data across different years. This could affect the ability to draw reliable conclusions about the long-term impacts of wildfire smoke on healthcare utilization.

4. **Data Suppression for Privacy Reasons:** In some cases, ADHS suppresses data when counts are low (e.g., fewer than 10 cases) to protect individual privacy. This suppression can result in gaps for certain years or demographic groups, limiting our ability to analyze the full scope of health impacts across different populations.

5. **COVID-19 Pandemic Impact:** The COVID-19 pandemic (2020-2022) had a significant impact on healthcare systems and respiratory health outcomes worldwide. Therefore, the data from these years may be skewed due to the pandemic's effect on hospitalizations and healthcare utilization for respiratory conditions unrelated to wildfire smoke.

Overall, it is also important to note that correlation is not causation so for example any correlations we might observe like an increase in asthma-related hospital visits along with an increase in the number of wildfires do not directly mean that wildfire is the cause for that increase. There are still multiple factors like wind direction or potential other causes of health issues like pollen that

I am not considering. Additionally, given that I have limited data for both wildfires and health, there is a possibility that the model may not perform as expected and might provide skewed results. To avoid that, I am still trying to find additional data.

Timeline to Completion

1. **Data Collection and Processing** *(Deliverable: Nov 10)*
   - Gather CDC Wonder mortality data for respiratory diseases
   - Collect AZDHS data for asthma-related hospitalizations and emergency visits
   - Obtain tuberculosis rates from AZDHS
   - Gather data on respiratory-related medical procedures from AZDHS
   - Clean and preprocess all datasets for consistency, including handling missing values.
   - Merge all the datasets to create one single healthcare dataset

2. **Model Development, Testing and Prediction** *(Deliverable: Nov 22)*
   - Analyze correlations between smoke exposure and health outcomes (feature engineering)
   - Experiment with different statistical models and tune them
   - Test models and validate results to find the best one
   - Create predictions for health impacts through 2050
   - Document methodology and assumptions

3. **Visualization and Results** *(Deliverable: Nov 24)*
   - Develop charts showing relationships between smoke exposure and health outcomes
   - Prepare final visualizations for presentation
   - Visualize the correlations and results of the models

4. **PechaKucha Presentation** *(Deliverable: Nov 27)*
   - Complete the PechaKucha presentation and submit it.

5. **Documentation and Presentation** *(Deliverable: Dec 4)*
   - Complete documenting all code and analysis methods
   - Write up findings and recommendations
   - Format and organize the GitHub repository
   - Complete the final report with all components

References

CDC Wonder - https://wonder.cdc.gov/

Arizona Department of Health Services - https://www.azdhs.gov/index.php

USGS Wildland Fire Data - https://www.usgs.gov/data/combined-wildland-fire-datasets-united-states-and-certain-territories-1800s-present

EPA Air Quality Index (AQI) Data - https://aqs.epa.gov/aqsweb/documents/data_api.html

Wilgus M-L, Merchant M. Clearing the Air: Understanding the Impact of Wildfire Smoke on Asthma and COPD. Healthcare. 2024; 12(3):307. https://doi.org/10.3390/healthcare12030307