Extension Plan

Wildland Fire Analysis for Stockton, CA

- Sushma Vankayala

Motivation/Problem Statement

Wildland fires create many challenges for communities near fire-prone areas, affecting air quality, public health, safety, and the environment. In places like Stockton, CA, where wildfires have become more common, the effects of smoke range from breathing problems to mental health, the economy, and overall community well-being. Local leaders, healthcare workers, and residents need to understand these risks fully. This analysis will look at how wildfire smoke impacts the health of people in Stockton, focusing mainly on respiratory diseases and cancer rates. The findings will help inform public health programs and resource distribution.
From a scientific viewpoint, studying how air quality affects health is vital. The results of this analysis could contribute to our understanding of environmental health and lead to better health advisories, improved readiness for smoke events, and targeted support for vulnerable groups, such as children and older adults. Ultimately, the aim is to identify the link between exposure to wildfire smoke and the rates of respiratory diseases and cancers in Stockton, highlighting the urgent need for preventive actions and policy changes to protect public health.

Impact Focus

This project will specifically examine the health effects of wildfire smoke, particularly concerning respiratory diseases and cancer. Understanding these impacts is crucial for addressing health inequalities in the community. Exposure to wildfire smoke, especially fine particulate matter (PM2.5), is linked to various respiratory illnesses such as asthma, chronic obstructive pulmonary disease (COPD), and lung infections. Additionally, PM2.5 has been associated with increased risks of cancers, including lung cancer and certain types of throat cancer. By analyzing health data in conjunction with the wildland fire data, this study aims to provide useful information for public health officials and community leaders in Stockton.

A few questions I wish to answer through this analysis are -

- How does wildfire smoke exposure impact mortality rates in Stockton, CA, particularly concerning respiratory illnesses and cancers?
- Is there a correlation between hospitalization data and wildfires - considering both short-term and long-term effects of exposure to smoke?

Data to Be Used

In the common analysis conducted in Part 1, I used data from the US Geological Survey and Air Quality Index to do a basic study on wildland fires and create an estimate for the wildfire smoke. For the next part, I'll be using health-related data to analyze how wildfire smoke affects the health of the residents of Stockton City. Below is a list of data sources I plan to use for my analysis

1. **Multiple Causes of Death - CDC**
   The Multiple Cause of Death database provides mortality and population statistics for all U.S. counties, based on death certificates of U.S. residents. Each certificate includes a primary cause of death, up to twenty additional contributing causes, and demographic information.

   - **Link:** CDC - Multiple Causes of Death Data

   - **Data Description:**
     The dataset contains fields with detailed information about most of the metrics related to deaths in the USA. For my analysis, I wish to use the fields related to location (State, county), demographics(gender, age, race), and cause of death(underlying cause of death and multiple causes of death).
     The tool allows users to filter on various conditions and aggregate on different columns. I will need to use the filter capability to fetch the data related to respiratory illnesses and cancer, only in the San Joaquin county
     A detailed description of the complete schema is present here: https://wonder.cdc.gov/wonder/help/mcd.html

   - **Data Use 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:
     i. *Use these data for health statistical reporting and analysis only.*
     ii. *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.).*
     iii. *Make no attempt to learn the identity of any person or establishment included in these data.*

iv. *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.*

This data on multiple causes of death allows us to aggregate various demographical and health-related fields. I believe that'll help understand how the severity of respiratory illnesses and cancer is changing with time, and how the number of deaths changed over time.

2. **Healthcare Cost and Utilization Project - AHRC**

   The Agency for Healthcare Research and Quality (AHRQ) is focused on improving the quality, safety, efficiency, and accessibility of healthcare in the U.S. It conducts research and provides resources aimed at enhancing healthcare systems and outcomes, with special attention to patient safety, care effectiveness, and reducing healthcare disparities. The Healthcare Cost and Utilization Project (HCUP) is a comprehensive source of hospital data across the U.S., providing insights into hospitalizations, healthcare utilization, and costs.

   - **Link:** HCUPnet Data Tools

   - **Data Description:**
     This dataset contains details like Average length of stay, Rate of discharges per 100k population, and average hospital costs. The data seems to be aggregated to provide annual counts and does not directly have a column stating the diagnosed condition. To gather all the data I'll potentially use for my analysis, I'll need to fire multiple queries, i.e., per year per diagnosed condition.

   - **Terms of Use:**
     The AHRQ Confidentiality Statute prohibits the use of AHRQ HCUP data to identify any person (including, but not limited to, patients, physicians, and other health care providers) or establishment (including, but not limited to, hospitals).
     Users of data on the HCUPnet website must agree to the following terms:
     i. *I will make no attempts to identify individuals, including by the use of vulnerability analysis or penetration testing. In addition, methods that could be used to identify individuals directly or indirectly shall not be disclosed, released, or published.*
     ii. *I will make no attempts to identify establishments directly or by inference.*
     iii. *I will not use deliberate technical analysis to discover or release information on small numbers of observations ≤10.*

iv. *I will not attempt to link this information with individually identifiable records from any other source.*
v. *I will not attempt to use this information to contact any persons or establishments in the data for any purpose.*

To adapt these datasets for the analysis, I will perform data cleaning and merging, focusing on aligning the temporal aspects of smoke exposure data with health outcome data. A regression model may be employed to analyze the correlation between smoke exposure and health outcomes.

Unknowns and Dependencies

Several factors may impact my ability to complete this analysis:

- **Data Accessibility:** While the datasets mentioned are publicly available, there may be delays in accessing specific data points or obtaining the necessary permissions for use.
- **Variability in Data:** The accuracy of health outcomes may vary based on reporting practices and the completeness of the datasets. Both the datasets seem to have had schema changes over the years, hence there will be an effort required in the data cleaning and merging steps. Furthermore, the granularity of the datasets is different, and this might pose a challenge in the analysis.
- **Data Availability:** The CDC data on multiple causes of death is available from 1999 to 2024, but the AHRC data is available for the years 2011 to 2020. Care is needed while processing the data to ensure relevancy. Since all the data available is at the county level, we'll need to analyze for the county and assume the city has similar trends. This assumption might lead to slight inaccuracies.
- **Accuracy of Smoke Estimates:** The smoke estimate calculated in Part 1 of this project was based on the area of the wildfire and its distance to Stockton. Considering that the wind direction was not considered in the calculation, we might need to adjust the estimates for our analysis of the wildfire smoke on the health of the residents.

Timeline to Completion

The following timeline outlines key milestones and tasks needed to complete the analysis:

1. **Data Collection (Nov 6 - Nov 10)**
    - Gather data from Multiple Causes of Death - CDC
    - Access hospitalization data from HCUPnet.
2. **Data Cleaning and Preprocessing (Nov 11 - Nov 16)**
    - Clean and merge datasets to ensure consistency.
    - Verify data integrity and completeness.
3. **Model Development, Testing and Analysis (Nov 17 - Nov 22)**

    - Explore relevant models in current literature.
    - Select and adapt a simple model for the analysis.
    - Test the model and analyze results for trends and correlations.
    - Conduct sensitivity analyses to understand the impact of assumptions.
4. **Visualization of Results (Nov 23 - Nov 24)**
    - Create visualizations to effectively communicate findings.
    - Prepare charts illustrating correlations and predictive outcomes.
5. **Documentation and Repository (Nov 25 - Dec 2)**
    - Compile findings into a comprehensive report. Reuse the content from previous homework where applicable and relevant.
    - Prepare a presentation for stakeholders. (**PechaKucha - Nov 27th**)
    - Review all materials for clarity and completeness before submission. Make sure that the GitHub repository adheres to the requirements. **(Final Submission - Dec 2nd)**

References:

1. Stockton, CA health dashboard: https://www.cityhealthdashboard.com/CA/Stockton/city-overview?metricId=29&dataPeriod=2021
2. Lung cancer and air pollution: https://www.crick.ac.uk/news/2022-09-10_scientists-reveal-how-air-pollution-can-cause-lung-cancer-in-people-who-have-never-smoked
3. Colorectal cancer and PM2.5: https://www.hsph.harvard.edu/news/press-releases/outdoor-air-pollution-may-increase-non-lung-cancer-risk-in-older-adults/#:~:text=Boston%2C%20MA%E2%80%94Chronic%20exposure%20to,%2C%202023%2C%20in%20Environmental%20Epidemiology.
4. PM2.5 and its impact on health (California Air Resources Board): https://ww2.arb.ca.gov/resources/inhalable-particulate-matter-and-health#:~:text=For%20PM2.5%2C%20short%2D,symptoms%2C%20and%20restricted%20activity%20days.
5. Wildfire smokes in California: https://oehha.ca.gov/climate-change/epic-2022/impacts-human-health/wildfire-smoke