Part 4 - Project Report

-Ananya Bajaj

Introduction

This project investigated wildfires within a specific region, focusing on those that occurred within a 1250-mile radius of Pueblo West County, Colorado. The first phase included a thorough examination of the wildfire's extent, as well as a quantitative estimation of smoke production. The project's subsequent stages include the incorporation of socioeconomic factors, providing a human-centered perspective that directly impacts Pueblo West County (such as health ramifications and correlations).

The significance of this analysis lies in its contribution to our understanding of wildfires and their repercussions on surrounding areas, particularly in the context of the specified region. It addresses the question of why this analysis is interesting or important by emphasizing the crucial role of comprehending the dynamics of natural disasters, which are frequently exacerbated by climate change.

The motivation for asking and answering these questions stems from a desire to understand the complexities associated with wildfires. The study sheds light on the spread of wildfire smoke, which is important for understanding the broader effects of climate change on local environments and communities. The study aims to provide insights into the real problems and challenges faced by communities affected by wildfires by examining socioeconomic impacts, thereby addressing unresolved research questions.

This project has significant real-world implications because it delves into issues related to air pollution caused by wildfire smoke. The analysis provides valuable data-driven insights into the socioeconomic vulnerabilities caused by wildfire smoke exposure. This directly addresses the question of whether the project solves a real problem by providing a thorough understanding of the health ramifications, including respiratory diseases such as Asthma, and potential long-term effects on the affected population.

The broader significance of this research is found in its contribution to the development of resilient communities. The analysis becomes a tool for increasing public awareness about potential impacts, fostering a sense of responsibility among residents, and promoting proactive measures to reduce wildfire risk and mitigate its impact by providing data-driven insights into socioeconomic vulnerabilities. In essence, the project

is important because it not only advances our scientific understanding but also has practical implications for community well-being and environmental sustainability.

1. Background/Related Work

A research paper that talks about wildfire risk transmission in similar areas is “Wildfire Risk Transmission in the Colorado, Front Range, USA” by Jessica R. Haas, David E. Calkin, Matthew P. Thompson. Using probabilistic fire simulation modeling, this paper identifies wildfire risk transmission pathways and high-exposure areas for human populations during severe weather events on Colorado's Front Range. Similarly, my project, which is focused on wildfires in a specific region, includes a thorough analysis of wildfire extent, smoke estimation, and the incorporation of socioeconomic factors. Both initiatives help to understand and quantify exposure, which helps to identify high-priority areas for wildfire risk mitigation and the development of resilient communities. The previously mentioned paper focuses on spatial risk assessment, whereas my project takes a broader approach by incorporating socioeconomic factors for a more complete understanding of wildfire impact.

The research presented in the provided paper informs the hypotheses, analysis, and system design of my project by offering valuable insights into wildfire risk transmission pathways, exposure assessment, and the impact on human populations, particularly in severe weather conditions along the Front Range of Colorado. The identification of high-exposure areas and the use of probabilistic fire simulation modeling in the paper contribute to the understanding of spatial risk dynamics.

Hypothesis: Increased socio-economic vulnerabilities in areas with high exposure to wildfires contribute to a greater impact on human populations.

Research Questions:

- What is the nature and strength of the link between estimated smoke levels in Pueblo West, Colorado, and hospital admissions for asthma-related complications? *This investigation aims to dissect the data in order to determine whether increased smoke exposure is a significant factor driving the rates of severe asthma incidents requiring hospitalization.*

- How do the fluctuations in the Air Quality Index (AQI) in Pueblo West relate to the variations in smoke estimates measured over the past decades? *This question*

aims to discern if there is a quantifiable, time-bound correlation that could inform public health advisories and environmental policies.

- Utilizing the trove of historical environmental data, can we develop a predictive model that accurately forecasts the trajectory of air quality trends and the consequent health impacts, particularly respiratory illnesses, in Pueblo West? *The focus is on leveraging statistical and machine learning techniques to anticipate future scenarios and guide proactive community health planning.*

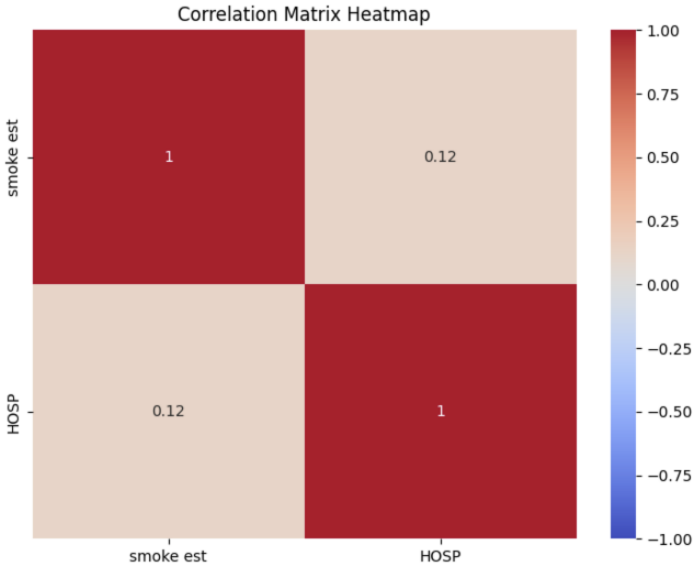
For my project, I'm leveraging the Community Level Estimates provided by the Colorado Department of Public Health and Environment (CDPHE). These estimates, spanning from 2004 to 2018, focus specifically on hospitalization due to asthma. They are derived from statistical models designed to generate health condition and risk behavior estimates for smaller geographic units known as census tracts.

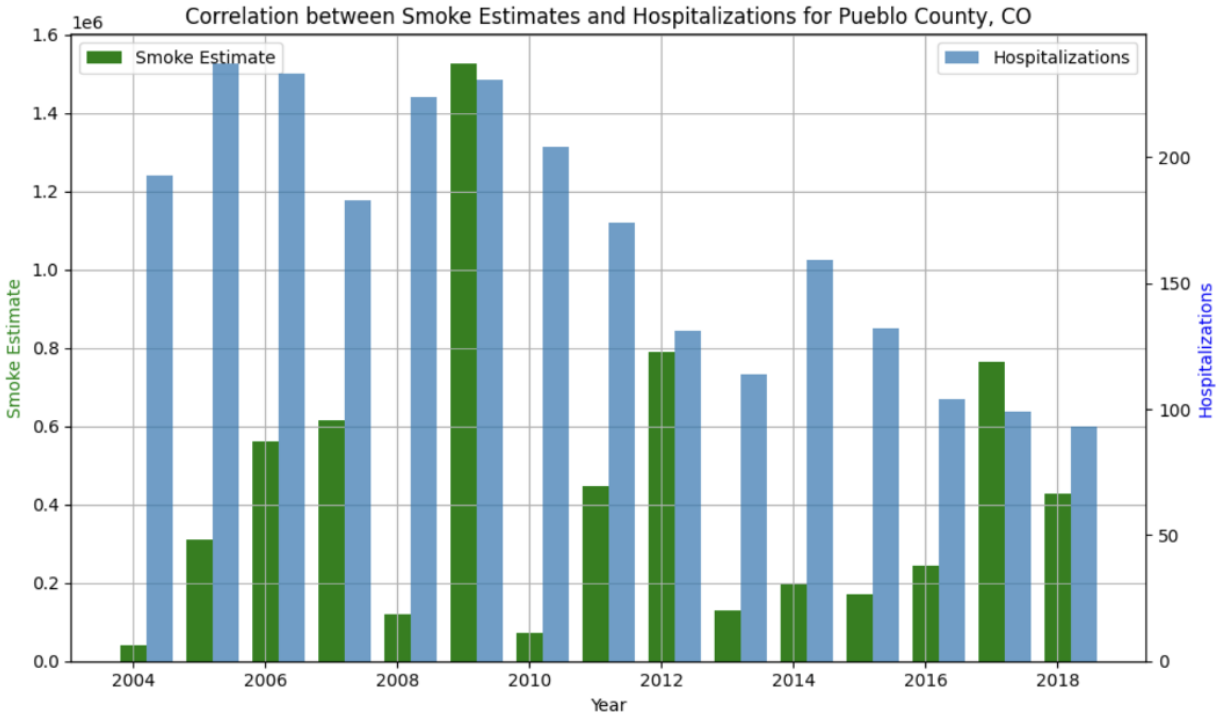
Due to sample size constraints, reporting estimates for census tracts are frequently limited in the traditional landscape of public health surveillance. The Community Level Estimates address this issue by implementing a multilevel model. This model incorporates not only individual responses from the Colorado Behavioral Risk Factor Surveillance System (BRFSS) survey, but also socio-demographic and contextual data from the United States Census, specifically the American Community Survey.

Individual survey responses are nested within geographic boundaries, namely counties, in the multilevel model. This nested structure allows the model to derive insights from both individual (demographic) characteristics and broader sociodemographic characteristics. By doing so, it estimates the likelihood of individuals having a specific health condition, such as asthma, or engaging in a risk behavior at the census tract level.

This dataset gives my project a unique opportunity to investigate the prevalence of asthma within smaller geographic units, providing a more nuanced understanding of health conditions at the community level. This data is critical as I investigate the complex relationships between wildfire smoke exposure, socioeconomic factors, and the resulting health impacts on the population in the specified region.

I used a correlation plot to map the correlations between asthma hospitalizations and the smoke estimate.





Link to the dataset:
https://www.cohealthmaps.dphe.state.co.us/cdphe_community_level_estimates/

2. Methodology

Data Extraction:

The dataset utilized in this analysis is the Combined Wildland Fire Datasets for the United States and certain territories, spanning from the 1800s to the present (combined wildland fire polygons). This dataset has been curated and compiled by the US Geological Survey. The extraction and processing of data are carried out using Python modules and diverse data processing techniques.

Data Filtering and Preprocessing:

The phase of data filtering and preprocessing involves the selection of fire perimeters within a 1250-mile radius of Pueblo West. This targeted approach enables the extraction of pertinent fire data for subsequent analysis.

Smoke Estimation Process:

The smoke estimation process entails the calculation and normalization of the smoke impact associated with each fire event, utilizing predefined parameters. The resulting estimated smoke impacts undergo further processing and correlation with the available Air Quality Index (AQI) data.

Smoke Estimate Equation:

weight * ((fire_size)**2 / (distance))

Correlation Analysis:

Correlation analysis delves into exploring the relationship between the estimated smoke impacts and the Air Quality Index data. This exploration provides valuable insights into the correlation between fire smoke and air quality in Pueblo West.

Predictive Model:

For the development of a predictive model for future smoke estimates in Pueblo West, the project employs the DARTS library and uses the exponential smoothing algorithm. This model facilitates the projection of potential smoke impacts over the next 25 years.

Data Visualization:

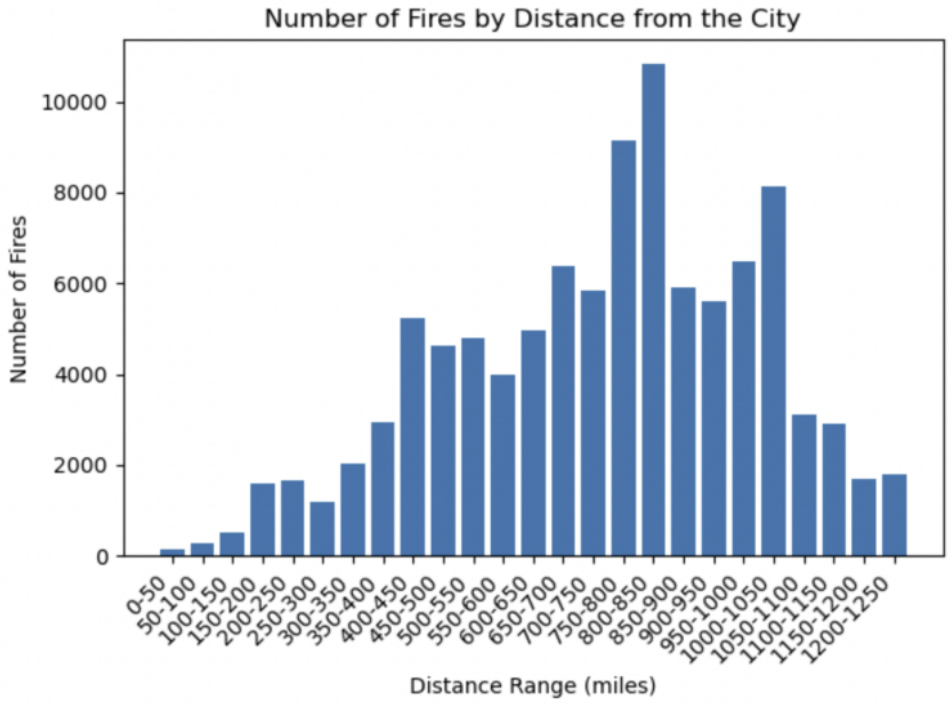
The outcomes of the data analysis are visually conveyed through a variety of time series graphs and histograms. This visual representation offers a comprehensive overview of the wildfire impact on Pueblo West.

Further Analysis on Wildfire Impacts:

Asthma Hospitalizations in Pueblo Country and their correlation with wildfires in the area: I leveraged the Community Level Estimates provided by the Colorado Department of Public Health and Environment (CDPHE). These estimates, spanning from 2004 to 2018, focus specifically on asthma. I further plotted a correlation matrix and also a two scale smoke estimate vs Hospitalization chart to view the relationship.

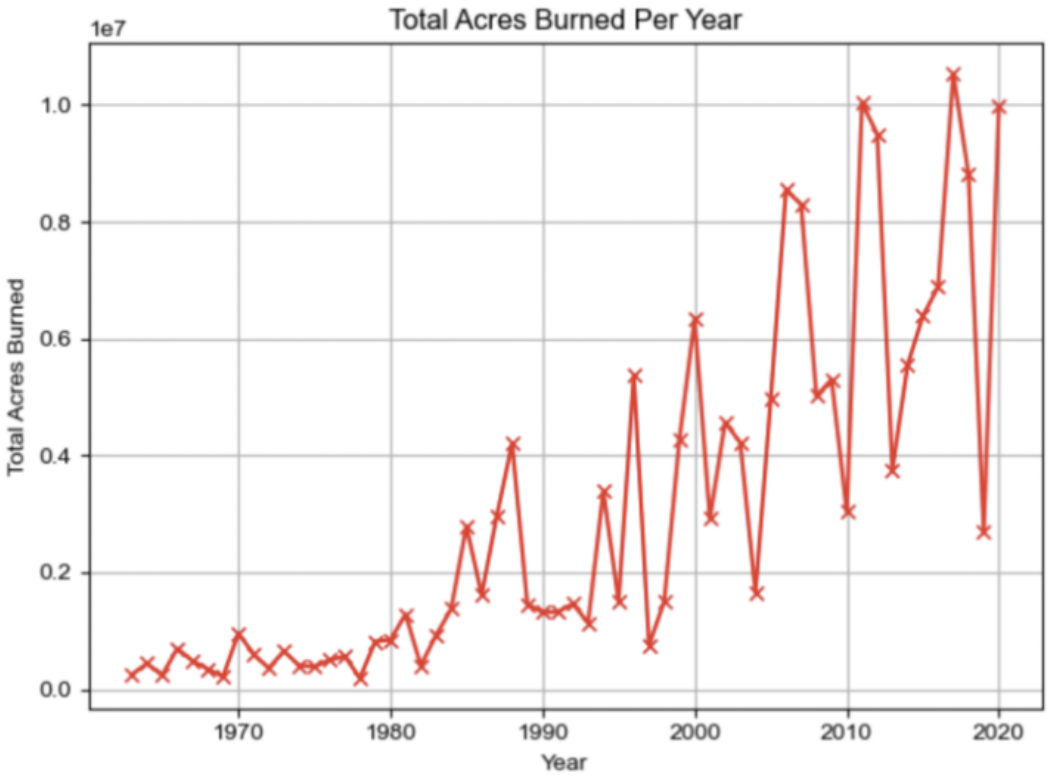
3. Findings

1.How far from Pueblo West are wildfires occurring?



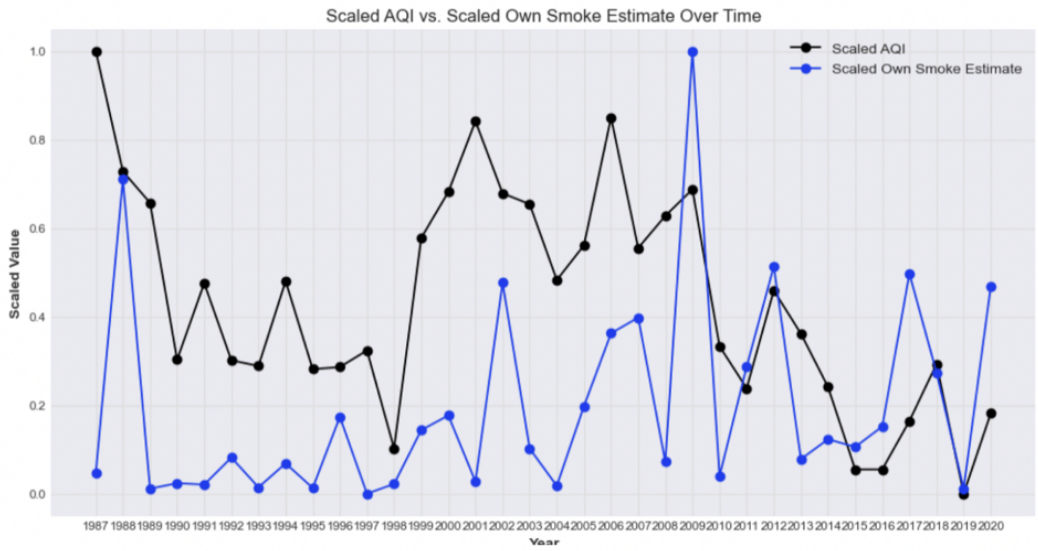
This histogram illustrates the distribution of wildfires based on their proximity to a specific city. The x-axis denotes distance from the city in miles, divided into 50-mile intervals, while the y-axis represents the count of wildfires.

2. What is the extent of land affected by wildfires?



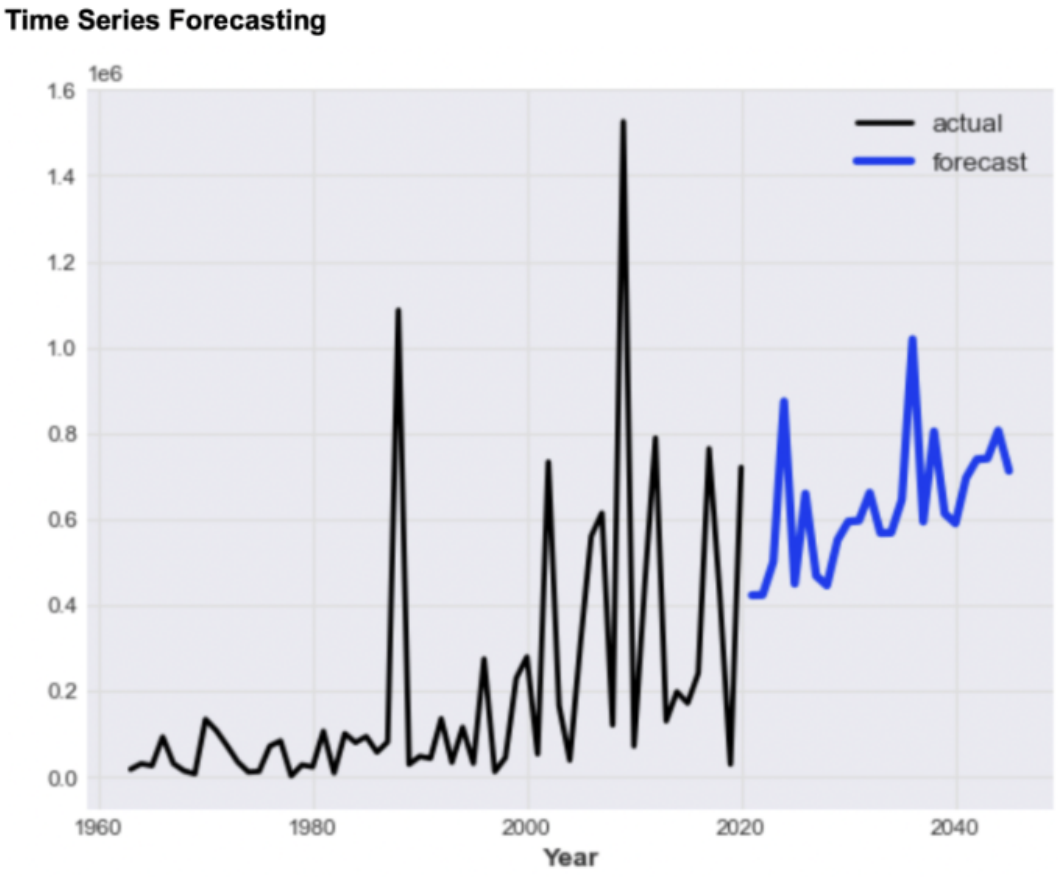
There is a clear upward trend in the total acres burned per year, especially noticeable from the 1980s onward. The early 1970s show relatively low numbers of acres burned. There are several notable peaks in the data where the total acres burned in a year are particularly high. These peaks occur more frequently and with greater magnitude in the latter half of the dataset, particularly from the 2000s onwards. The data suggests an increase in the frequency and/or severity of fires over the last five decades.

3. How does smoke impact correlate with the Air Quality Index?

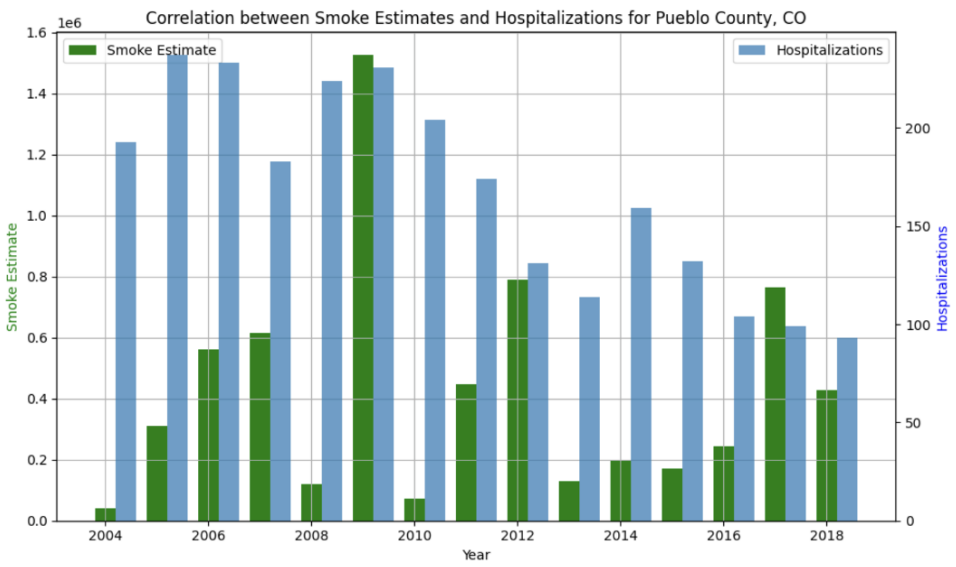


Time series data for the Air Quality Index (AQI) and smoke estimates show significant variability over time, with no smooth or consistent pattern. This irregularity highlights the numerous factors that influence these measurements. When the trends are examined, both the AQI and smoke estimates show distinct peaks and troughs, indicating fluctuations in air quality and smoke levels over time. Notably, both variables reach their apex around 1990, the early 2000s, and again between 2010 and 2020. There are times when the AQI peaks coincide with the smoke estimates, implying a possible correlation, but this connection does not hold consistently over time. More recently, particularly after 2010, the data indicates a general upward trend for both the AQI and smoke estimates, with some of the highest recorded values in the series. The relationship between the two variables waxes and wanes throughout the timeline, with periods of divergence, where the correlation appears tenuous, and convergence, where the correlation appears stronger, indicating a complex interaction between air quality and the presence of smoke.

What could the smoke impact in future years look like?



This diagram depicts a time series forecasting model that compares actual historical data to projected future values. The black line represents historical data, which shows a fluctuating pattern with several spikes indicating periods of high values spanning from 1960 to just before 2020. The forecast, denoted by the blue line, starts where the historical data ends and extends until 2040. The forecast suggests that the variability seen in the historical data will be generalized, with a rising trend initially followed by fluctuations. However, the forecast does not replicate the extreme spikes observed in the historical data, indicating that the model may smooth out such anomalies. It is essential to note that while the forecast is based on past patterns, it inherently carries uncertainty, especially as the time horizon extends further into the future. This forecast could be for any time-dependent data such as financial markets, weather patterns, or, as the previous charts might suggest, environmental measures like AQI or wildfire occurrences.

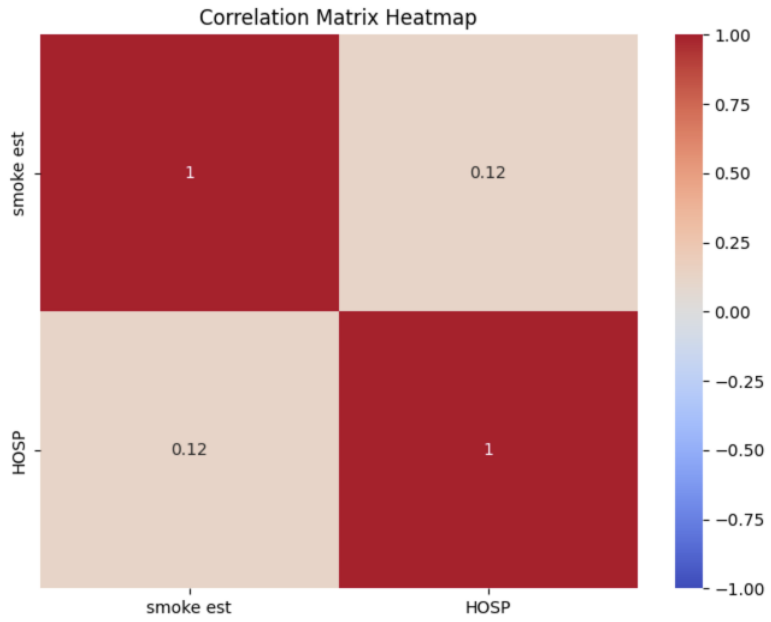


The chart depicts the correlation between hospitalizations due to asthma and smoke estimates for Pueblo County, Colorado, over various years from 2004 to 2018. There are two sets of bars representing each variable:

> **Smoke Estimate**: This is shown in green bars, and it seems to represent some form of measurement related to the presence or estimation of smoke in the environment. The unit of measurement is not provided, but it appears to be on the left y-axis.
>
> **Hospitalizations**: The blue bars represent the number of hospitalizations due to asthma. These seem to be directly correlated to the smoke estimate, implying that as the smoke estimate increases, the hospitalizations due to asthma also increase. The number of hospitalizations is read off the right y-axis, with values ranging from 0 to above 200.

The x-axis represents years from 2004 to 2018, with data for each even year presented.

From the visual representation, it appears that the years with higher smoke estimates also have higher rates of hospitalizations due to asthma. This suggests that there might be a positive correlation between the amount of smoke in the environment and the incidence of asthma hospitalizations in Pueblo County, CO.



The chart shows a correlation coefficient of 0.12. This value indicates a positive but very weak correlation between smoke estimates and hospitalizations. Perhaps if the data were available, we could account for doctor visits in addition to hospitalizations.

4. Discussion/Implications

The findings from the various charts underscore the critical relationship between environmental factors, such as air quality and smoke estimates, and their potential impact on public health and safety. These observations are particularly important for city planning and public policy due to the following reasons:

1. **Public Health Concerns**: The link between smoke estimates, AQI, and hospitalizations for respiratory conditions such as asthma suggests that air quality has a direct impact on residents' health. According to the data, days with poor air quality are associated with an increase in hospital visits, which could strain healthcare systems and lead to public health crises.
2. **Environmental Impact and Preparedness**: The rising number of acres burned each year indicates an increased risk of wildfires, which could cause property damage, loss of life, and further deterioration of air quality. Understanding these trends is critical for disaster preparedness and environmental conservation.

3. **Informing Policy and Investment**: Forecasting models predict that the variability and potentially harmful conditions may continue or worsen. This information is crucial for the city council and mayor to invest in sustainable environmental practices, improve emergency response infrastructure, and develop long-term urban planning that mitigates these risks.

In response to these findings, the following actions are suggested:

- **City Council and Mayor**: They should form a task force to develop an action plan over the next six months to a year. This plan should prioritize emergency services, green infrastructure investment, and the implementation of policies that reduce environmental risks, such as strict emission regulations and support for renewable energy sources.
- **City Residents**: Residents should be encouraged to take personal measures to reduce air pollution, such as taking public transportation, and to prepare for poor air quality days, such as staying indoors and using air filters, through public awareness campaigns.
- **Data-Driven Decision Making**: Human-centered data science principles have informed the recommendations by emphasizing the importance of using data to make decisions that prioritize the well-being of residents, balancing the needs of various stakeholders, and considering the long-term sustainability of the city.

The city officials must engage with data scientists to understand the nuances of the data and ensure that the interpretations are accurate and the forecast models are robust. It is also important to communicate these findings and the associated uncertainties clearly to the public to foster trust and cooperation in any initiatives undertaken.

5. Limitations

- **Exclusion of Doctor Visits**: Hospitalization data may not capture the full spectrum of health impacts related to poor air quality since many individuals affected by asthma may visit doctors without being hospitalized. This limitation could underestimate the true health burden and the sensitivity of public health to smoke and air quality.
- **Cleaning the asthma data:** Required a lot of preprocessing to change the level of granularity given in the dataset and also add a lot of filtrations to get the desired data.

- **Forecasting Uncertainty**: The forecasting model's projections are inherently uncertain, especially over long time horizons. The further the forecast extends into the future, the greater the potential for error due to unforeseen events and changes in underlying patterns.
- **Missing Gaseous data:** The absence of gaseous data for certain intervals within the study period poses a limitation, potentially leading to gaps in the analysis and affecting the accuracy of trend assessments and forecasts

6. Conclusion

Research Questions/Hypotheses (in short):

-Is there a correlation between smoke estimates and hospitalizations due to asthma in Pueblo County, CO?
-Does the Air Quality Index (AQI) correlate with smoke estimates over time?
-Can we forecast future trends in air quality and related health impacts based on historical data?

Summary of Findings:

- A weak positive correlation was identified between smoke estimates and asthma hospitalizations in Pueblo County, suggesting a relationship but not strong enough to predict hospitalization rates solely based on smoke levels.
- The scaled AQI and smoke estimates showed variability and certain years with synchronized peaks, hinting at a possible correlation, though it was not consistently strong across the entire timeline.
- Forecasting models indicated a continuation of the observed variability in air quality, with a rising trend in the recent past, suggesting potential worsening conditions.

Reflection:

This study employs human-centered data science by concentrating on data that have a direct impact on human health and well-being. It makes use of historical data to inform future decisions, takes a proactive approach to environmental conditions that affect community health, and identifies areas where city management can intervene for public benefit. Despite its limitations, the identification of correlations between environmental factors and health outcomes provides a foundation for informed policy-making that prioritizes human safety and quality of life. It also emphasizes the importance of

comprehensive data collection and robust analytical methods in order to ensure that the insights obtained are as accurate and actionable as possible, reflecting the human-centered ethos of serving community needs through data.

7. References

https://onlinelibrary.wiley.com/doi/epdf/10.1111/risa.12270
https://www.sciencebase.gov/catalog/item/61aa537dd34eb622f699df81
https://academic.oup.com/toxsci/article/186/1/149/6449455

8. Data Sources

https://www.cohealthmaps.dphe.state.co.us/cdphe_community_level_estimates/
https://www.sciencebase.gov/catalog/item/61aa537dd34eb622f699df81https://drive.google.com/drive/folders/1OJktGAx86hvMtirCUkGnS292r-FpPvLo