INTRODUCTION

Wildfires that burn near populated areas can have significant impact on the environment, property, livestock and human mortality and morbidity depending on the size, speed and proximity to the fire, and whether the population has advanced warning to evacuate.

Premature mortality is defined as all deaths where the deceased is younger than 75 years of age. 75 years of age is the standard consideration of a premature death according to the CDC's definition of Years of Potential Life Loss.

Emissions from wildfires worsen air quality and can adversely impact human health. Wildfire smoke is a mixture of air pollutants of which particulate matter (PM) is the principal public health threat. Wildfire smoke is associated with premature deaths in the general population, and can cause and exacerbate diseases of the lungs, heart, brain/nervous system, skin, gut, kidney, eyes, nose and liver. It has also been shown to lead to cognitive impairment and memory loss. Firefighters and emergency response workers are also greatly impacted by injuries, burns and smoke inhalation, particularly at high concentrations.

More interdisciplinary research is warranted to understand the latent and long-term health effects of wildfire exposure on vulnerable populations (children, older people, pregnant people, chronically ill people), particularly for geographic areas enduring repeated and cyclical exposure to these wildfire events.

PROBLEM STATEMENT

The objective of this study was to investigate and quantify the impact of wildfire smoke on premature deaths within the city of Benicia, California. The research aimed to assess the correlation between exposure to wildfire smoke and the incidence of premature mortality in the region. The findings will contribute to a better understanding of the public health implications of wildfire smoke exposure in a localized context, informing strategies for mitigation and public health interventions in Benicia.

MOTIVATION

The motivation behind this study lies in recognizing the increasingly evident threat that wildfire smoke poses to public health, particularly in regions prone to recurrent wildfires such as Benicia, California. With the rise in the frequency and intensity of wildfires, understanding the specific impact on premature deaths in a localized setting becomes crucial for effective public health planning and intervention. By exploring the correlation between wildfire smoke exposure and premature mortality in Benicia, the study aims to provide actionable insights for policymakers, healthcare professionals, and emergency responders. This research is motivated by the need to quantify the health risks associated with wildfire events, taking into account the unique characteristics of Benicia's population and environmental factors. Furthermore, the study addresses a gap in understanding the long-term health effects of

wildfire smoke exposure, especially in vulnerable populations. By shedding light on the specific health risks and disparities within the community, the research can guide the development of targeted strategies to mitigate the impact of wildfire smoke and protect the health of residents in Benicia. Overall, the motivation is rooted in the broader goal of safeguarding public health in the face of increasing wildfire risks.

BACKGROUND AND RELATED WORK

There have been several studies that have been conducted on the acute health effects immediately following a wildfire, such as reduced lung function, and exacerbation of chronic lung disease; as well as neurological and cardiovascular effects and increased mortality. Research has also been conducted on long-term health effects of wildfire and wildfire smoke exposure including morbidity and mortality related to burns or smoke inhalation, mental health effects, healthcare expenses related to wildfires. These studies informed my research question and hypothesis.

Research Question: What is the correlation between exposure to wildfire smoke in Benicia, California, and the incidence of premature mortality in the region?

Hypothesis: Increased exposure to wildfire smoke in Benicia, California, is associated with a higher incidence of premature mortality, highlighting the public health impact of such environmental factors.

The datasets to be used were published by FRED, which is a database of over 267,000 economic time series from 80 sources. This dataset was created using data that has been published by the Centers for Disease Control and Prevention which is the national public health agency of the United States.
The dataset contains the following columns:

1. Date in YYYY-MM-DD format
2. Age adjusted death rate

Description of the Columns:

Age-adjusted death rates are weighted averages of the age-specific death rates, where the weights represent a fixed population by age. They are used to compare relative mortality risk among groups and over time. An age-adjusted rate represents the rate that would have existed had the age-specific rates of the particular year prevailed in a population whose age distribution was the same as that of the fixed population. The unit for this value is the number of premature deaths per 100,000 individuals. Premature death rate includes all deaths where the deceased is younger than 75 years of age. 75 years of age is the standard consideration of a premature death according to the CDC's definition of Years of Potential Life Loss.

The Date column contains the date in the YYYY-MM-DD format. However, for the analysis, only the year will be relevant. So, the year must be extracted from the Date column as part of the preprocessing.

Poverty must be controlled for and so, data about Personal Income per capita was taken from US Census Data (FRED). The following column was added to the final dataset:

1. Per capita Personal Income

Description of the Columns:

Personal income is the income that is received by persons from all sources. It is calculated as the sum of wages and salaries, supplements to wages and salaries, proprietors' income with inventory valuation and capital consumption adjustments, rental income of persons with capital consumption adjustment, personal dividend income, personal interest income, and personal current transfer receipts, less contributions for government social insurance. This measure of income is calculated as the personal income of the residents of a given area divided by the resident population of the area.

License: Custom - https://fred.stlouisfed.org/fred-addin/legal.html

METHODOLOGY

The first phase of the project involved creating an annual wildfire smoke estimate for the city of Benicia, CA. Having reviewed the metadata and looked over what the features in the dataset look like,it was concluded that a reasonable smoke estimate could be derived from the following factors:

1. The GIS calculated hectares of the fire polygon calculated by using the Calculate Geometry tool in ArcGIS Pro - GIS Hectares. This value is multiplied by 0.00386 to convert it to Acres.
2. The average distance from the fire boundary and city of interest ie Benicia, CA - Distance

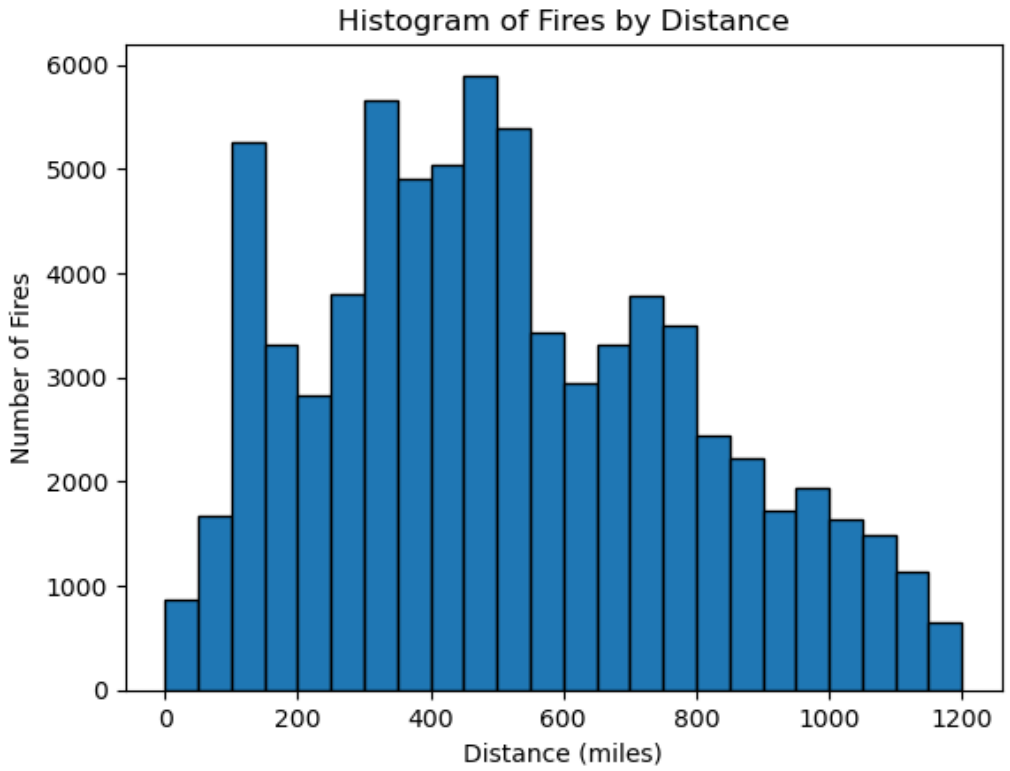
It is a reasonable assumption that a fire spread over a larger area will generate more smoke. Similarly, a fire burning further away from the city will have less impact on the city in terms of smoke. Based on these two assumptions, the formula that was used for the smoke estimate was:

Smoke Estimate = (GIS_Hectares * 0.00386)/Distance

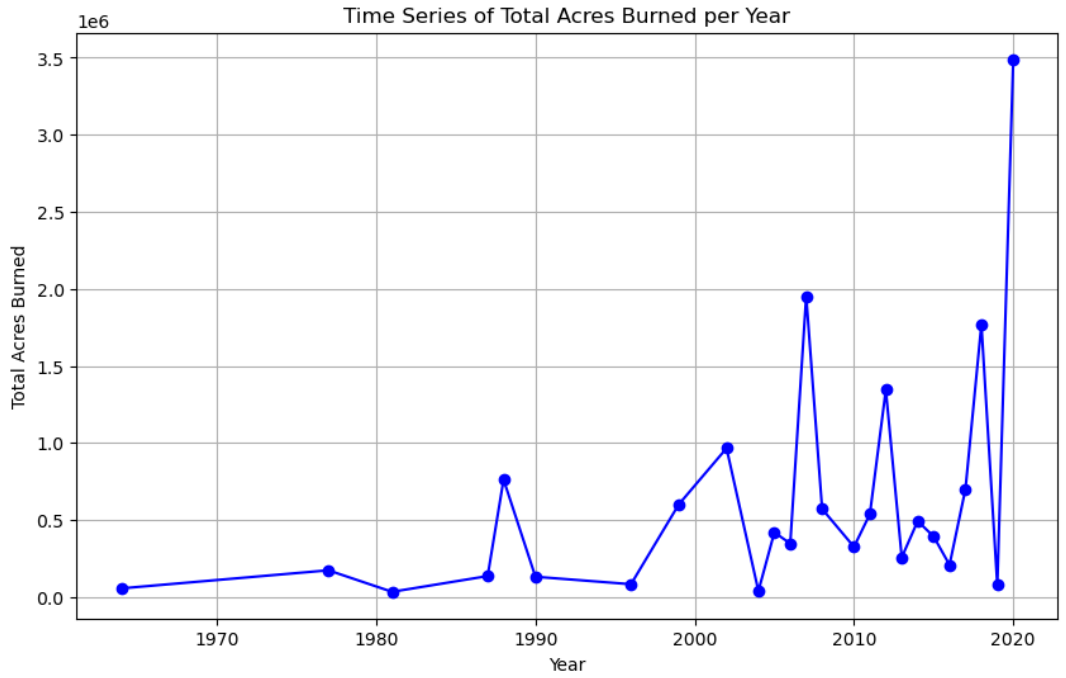
Exploratory Data Analysis:

The graph below is a histogram showing the number of fires occurring every 50 mile distance from Benicia, California up to the max specified distance. The x-axis represents the distance in miles from the assigned city. The distances are divided into 50-mile intervals. Each interval is represented by a bar on the histogram. The y-axis represents the count or frequency of fires that occurred within each 50-mile distance interval. It shows how many fires were recorded in each specific range. Each bar in the histogram represents one of the 50-mile distance

intervals. The height of each bar corresponds to the number of fires that occurred in that distance range.



The graph below is a time series graph of total acres burned per year for the fires occurring in the specified distance from Benicia, California. The x-axis represents the years during which the fires occurred. Each year is marked along the x-axis, allowing viewers to follow the progression of time. The y-axis represents the total acres burned by fires each year. The units on the y-axis indicate the scale of the total acres burned, which is the variable being measured.

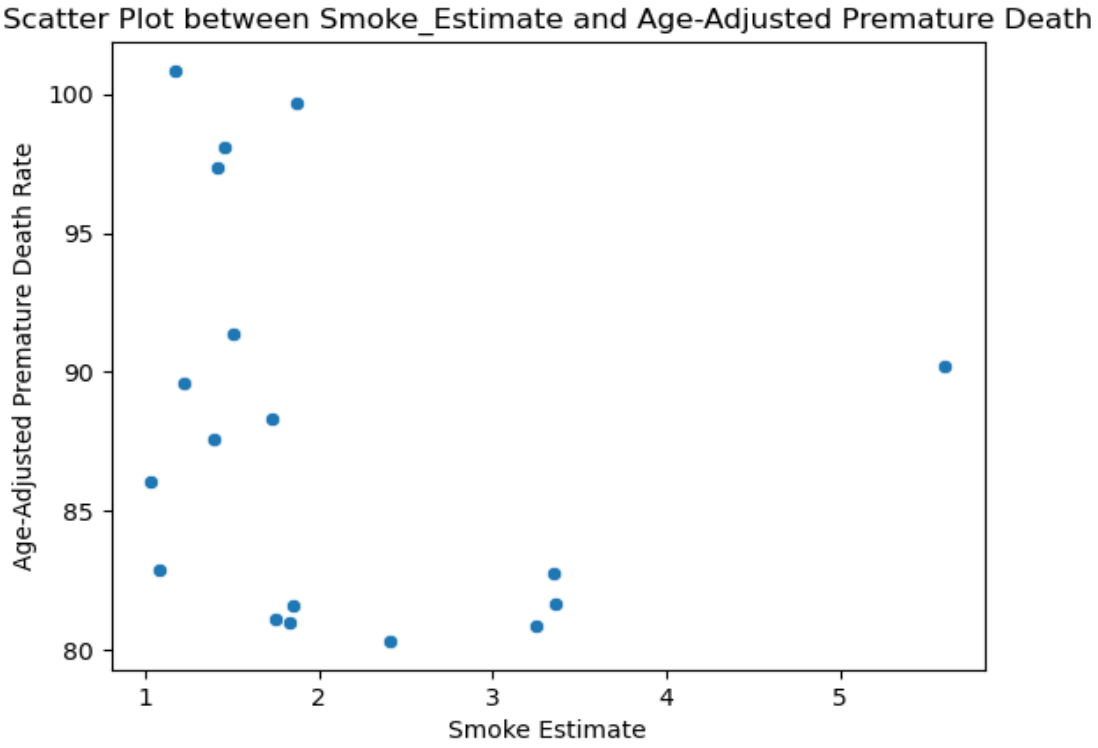


The second phase of this project involved extending the above analysis to observe and quantify the impact of wildfire smoke on premature deaths in the city of Benicia, CA. For the purpose of my analysis, I chose to use a linear regression model in order to study the relationship between wildfire smoke impact and premature death rate. I selected this model because it would allow me to assess the strength and direction of the correlation between wildfire smoke presence and premature deaths while considering influential factors such as personal income. I could also extend this to estimating expected premature mortality across different levels of smoke exposure, offering valuable insights for planning and implementing public health interventions.

There were also several human-centered aspects to this study that had to be considered. Firstly, controlling for personal income in the model was essential. This is because personal income serves as a socioeconomic indicator that is often associated with various determinants of health, including access to healthcare, living conditions, and lifestyle choices. It was important to ensure that this study was inclusive and addressed issues of equity. The data used was completely anonymous and did not contain any information that could potentially identify an individual and affect the way that they were treated as a result of this study. This ensured the privacy of the involved individuals. Lastly, the project aimed to communicate the results of the analysis in an accessible and transparent manner such that it could be of use to both, the scientific community as well as the general public and policymakers. It was critical that findings and implications were communicated clearly along with the limitations of this study allowing for the best possible planning by decision-makers.

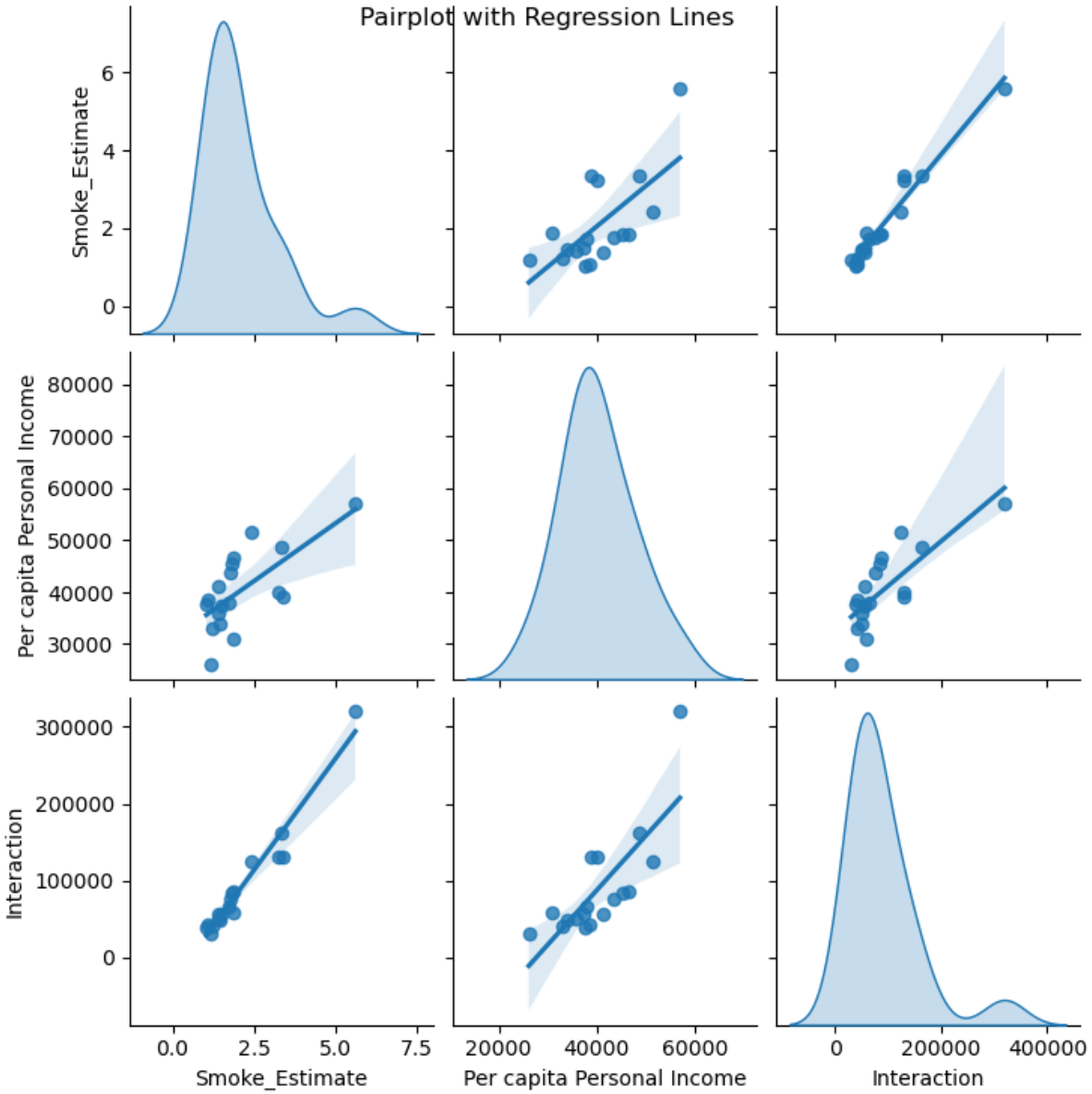
FINDINGS

When visualizing the relationship between the smoke estimate and the premature deaths, it could be observed that there was no direct linear relationship between the two variables. However, since controlling for income could potentially address the issue of linearity, it was



added as a predictor to the linear regression model. In order to further account for the linearity issue, an interaction term which was the product of the two predictors was added.

From the summary of this model, we could see that the coefficient for the smoke estimate predictor, per capita personal income predictor and the interaction term were -16.357, -0.0015 and 0.0004. However, we could see that all 3 values could be considered significant as the p-values were less than the conventional significance level of 0.05. The model seemed to explain a substantial amount of variance in the dependent variables (~75%), but it is important to interpret individual coefficients cautiously, considering both their magnitudes and statistical significance. Additionally, the presence of multicollinearity would need further investigation.



DISCUSSION AND IMPLICATIONS

These findings call for additional data gathering and research into the impact of wildfire smoke on the health and mortality of the residents of Benicia. In order to develop a more concrete understanding of the impact of wildfire smoke on premature deaths, data on several other variables is necessary. These variables include respiratory illnesses, hospitalizations, maternal health, genetic and chronic disease prevalence and residents' access to healthcare.
Further data gathering can be conducted in the following ways:

a. Air quality monitoring systems must be strengthened to provide real-time information on smoke levels.
b. Comprehensive health surveys can be conducted within the community to gather detailed information on residents' health conditions, respiratory symptoms, and any pre-existing medical conditions as well as socioeconomic indicators such as employment, education and housing. This data can provide a baseline for understanding health vulnerabilities.

Moreover, in order to develop an equitable analysis that is accessible to all residents, further research must be conducted on the disproportionate effect of wildfire smoke on vulnerable populations such as senior citizens, low-income individuals, pregnant women and children.

Over the course of my analysis, human centered principles informed each of the decisions that I made during the design and implementation of the study as well as in the recommendations that were made to the council based on the results of this study. In order to ensure privacy and data protection, I ensured that the data being used was sourced via ethical sources and that it was completely anonymized. The data being used in this study cannot be traced back to a single individual and will not be used in order to treat them differently. In order to account for accessibility and to ensure that the findings in this study would include and be useful to people with diverse abilities and needs, I controlled for personal income in my model as it serves as a socioeconomic indicator that would indicate an individual's access to healthcare, general lifestyle and overall health. Moreover, I recommended to the council that further data about various other socioeconomic indicators that would further strengthen the accessibility of this analysis. In order to ensure that the results of the study that informed these recommendations were transparent and understandable to the council in the form of a presentation with visualizations and clear and concise explanations of the study and the findings. I tried to ensure to the best of my ability that my study was ethical and human centered and that it placed a strong emphasis on understanding and addressing the needs, perspectives, and values of the individuals of Benicia, CA.

LIMITATIONS AND FUTURE WORK

The specific causes of death are not provided in the dataset. Knowing the leading causes of death for each year can help interpret the age-adjusted death rates more comprehensively. Moreover, trends in public health, medical advancements, and socio-economic factors

affecting mortality rates over time other than personal income were not explicitly included. Investigating such trends can help explain changes in death rates. Further, the data being used for this study only covers the last 20 years and has been restricted to the city of Benicia, CA. Collecting more data and expanding the study to explore patterns and trends across other cities would strengthen this analysis. It should also be noted that the data collected on premature deaths and per capita personal income was on a county level as opposed to just the city due to the lack of data on a city level. While county-level data provides a broader perspective and is designed to represent the characteristics of its cities, it may not be perfectly representative of a specific city within that county.

For the next phases of this study, I would like to extend the research to include the variables discussed above. I would also like to explore other models that could better account for the relationship between the smoke estimate and premature death rates.

CONCLUSION

The objective of this study was to investigate and quantify the impact of wildfire smoke on premature deaths within the city of Benicia, California. The hypothesis was that increased exposure to wildfire smoke in Benicia, California, is associated with a higher incidence of premature mortality. The findings of this study showed that while there was no direct correlation between the two, further research including various other predictors was critical for a proper understanding of the relationship. The considerations of fairness, privacy, accessibility and transparency that were explored in this study will inform the reader's understanding of human-centered data science.

DATA SOURCES

1. For Premature Death Rate: https://fred.stlouisfed.org/series/CDC20N2UAA006095
2. For Per Capita Personal Income: https://fred.stlouisfed.org/series/PCPI06095

REFERENCES

1. Emily Grant, Jennifer D. Runkle, Long-term health effects of wildfire exposure: A scoping review, The Journal of Climate Change and Health, Volume 6, 2022, 100110, ISSN 2667-2782, https://doi.org/10.1016/j.joclim.2021.100110. (https://www.sciencedirect.com/science/article/pii/S2667278221001073)

2. C.E. Reid, M.M. Maestas, Wildfire smoke exposure under climate change: impact on respiratory health of affected communities Curr Opin Pulm Med, 25 (2019), pp. 179-187, 10.1097/MCP.0000000000000552