Analysis of Wildfire Smoke Impact on Health Factors: A Report for Policy Makers

Executive Summary

This report analyzes the current and possible future impact of wildfire-induced smoke in Prescott City, Arizona. I analyzed the correlation between wildfire smoke estimates and various health-related factors. The analysis, conducted using historical fire data over the last 60 years and several health factors collected from multiple different sources, focuses on the impact of wildfire smoke on asthma-related deaths, asthma incidents, women's birth rate, and premature death rate.

Introduction

The intersection of environmental factors and public health has always been pertinent. Public access to regional data regarding various health factors has opened up new avenues for analysis and data-driven policy-making. It has become increasingly essential for policymakers to leverage data science techniques to improve the quality of life of the constituents of their communities. The city of Prescott, Arizona, is no exception.

My motivation for this extension plan stems from a deep concern for the health and well-being of the Prescott community, especially in the context of smoke exposure due to regional wildfires. The increasing frequency and intensity of wildfires, particularly in the western United States, have prompted a growing concern about their impact on urban environments and public health. This report focuses on Prescott City, a valley region increasingly affected by wildfire smoke.

The analysis I conducted focuses on the impact of smoke on critical health indicators, including the **incidence rate of asthma**, the **death rate due to asthma,** the **fertility rate in females,** and the **premature mortality rate**. The rationale behind focusing on these specific health markers is twofold. Firstly, some of these indicators, such as the incidence rate of asthma and death rate due to asthma, are among the most sensitive to changes in air quality and can provide early warnings about the broader health impacts of forest fires. Additionally, I aimed to explore the impact of smoke on tertiary factors like fertility rate that might have been ignored thus far. Secondly, they encapsulate immediate and long-term health outcomes, offering a comprehensive view of the public health landscape.

This analysis is not only scientifically intriguing but also practically essential. It seeks to answer a critical question: “How is the smoke from wildfires within a 1250-mile radius of Yavapai County impacting the health of Prescott’s residents?” This analysis yields a plethora of information that would help policymakers make informed decisions to mitigate the effects of smoke in their jurisdiction and plan for the future.

I have developed a smoke estimate model that incorporates fires' size and distance from Prescott City and analyzes the impact of the smoke on various health factors. This analysis aims to enable policymakers to understand smoke's current and future health impacts on the community. This understanding is crucial for data-driven decision-making, policy development, and resource allocation to mitigate the adverse effects of smoke exposure.

Ultimately, this analysis is not just about numbers and data but about people. It is about understanding the hidden costs of environmental challenges and taking proactive steps to safeguard the health and future of the Prescott community. The implications of this study extend beyond academic interest, offering tangible benefits and insights that could shape public health strategies and improve the quality of life for Prescott’s residents.

Background/Related Work

Previous research has extensively documented the detrimental effects of wildfire smoke on air quality and health [12] [13]. However, there still needs to be a gap in understanding the specific impacts on individual cities and their unique demographic and environmental contexts. My analysis builds upon these studies, focusing on the specific case of Prescott City. The analysis uses a 1250-mile radius around the city to quantify the impact of smoke. This approach allows for a more targeted and relevant assessment of the implications of wildfire smoke for local policy-making.

Models

1. Pearson Correlation

Pearson correlation is a statistical measure quantifying the linear relationship between two variables, indicating the relationship's strength and direction. It produces a value between -1 and 1, where 1 implies a perfect positive correlation, -1 indicates a perfect negative correlation, and 0 means no correlation. Pearson correlation assumes that both variables are normally distributed and have a linear relationship, and it is sensitive to outliers, which can significantly affect the correlation coefficient. It is widely used in various fields for preliminary data analysis, helping identify potential relationships for further investigation. It is important to note these assumptions, and in cases where my data is not normally distributed or has no linear relationship, my confidence in this model declines. I utilized this model to quantify the impact of smoke on various health factors.

Interpretation of values [13]:

Correlation Coefficient		**Dancey & Reidy (Psychology)**	**Quinnipiac University (Politics)**	**Chan YH (Medicine)**
+1	−1	Perfect	Perfect	Perfect
+0.9	−0.9	Strong	Very Strong	Very Strong
+0.8	−0.8	Strong	Very Strong	Very Strong
+0.7	−0.7	Strong	Very Strong	Moderate
+0.6	−0.6	Moderate	Strong	Moderate
+0.5	−0.5	Moderate	Strong	Fair
+0.4	−0.4	Moderate	Strong	Fair
+0.3	−0.3	Weak	Moderate	Fair
+0.2	−0.2	Weak	Weak	Poor
+0.1	−0.1	Weak	Negligible	Poor
0	0	Zero	None	None

Table 1: Interpretation scale for correlation value

2. Vector Autoregressive Model

The Vector Autoregressive (VAR) model is a statistical model used in econometrics and other disciplines that deal with time series data. It is particularly popular in macroeconomic analysis for forecasting systems of interrelated time series and analyzing the dynamic impact of random disturbances on a system of variables. The VAR model captures the relationship between multiple quantities as they change over time. I utilized this model to predict the future impact of smoke on various health factors.

3. Smoke Estimation Model

My custom smoke estimate model relies on the following formula:

smoke_estimate_value = sigmoid(0.1* area_of_fire + distance_to_city^{-2} + 0.1 * is_wild_fire + 0.05 * is_prescribed_fire)

A Sigmoid function is a mathematical function that has a characteristic S-shaped curve. I used it in the analysis to contain the value of the impact of each fire between 0 and 1. The final smoke estimate per year is taken as the sum of all impacts of each fire.

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Figure 1: Sigmoid Function formula

Methodology

My methodology involved analyzing historical wildfire data from the Combined Wildland Fire Datasets (US Geological Survey) [15] and health-related data from various sources mentioned in a different section below. I employed statistical methods to determine correlations between wildfire smoke intensity (estimated based on fire proximity and size) and health outcomes in Prescott City. Human-centered data science principles guided my analysis, ensuring that my methods and interpretations remained focused on the real-world impact on the city's residents. I focused on introducing a qualitative component to the analysis, looking beyond the numbers to interpret results in a human-centric fashion, and considering factors outside the model's features.

Findings

1. Asthma-Related Deaths and Smoke Impact (Correlation Value -0.8)

My analysis revealed a significant negative correlation between smoke impact and asthma-related deaths. This counterintuitive finding suggests that other factors, such as advancements in medical technology and public health initiatives, might play a more significant role in reducing asthma-related deaths than the direct impact of smoke.

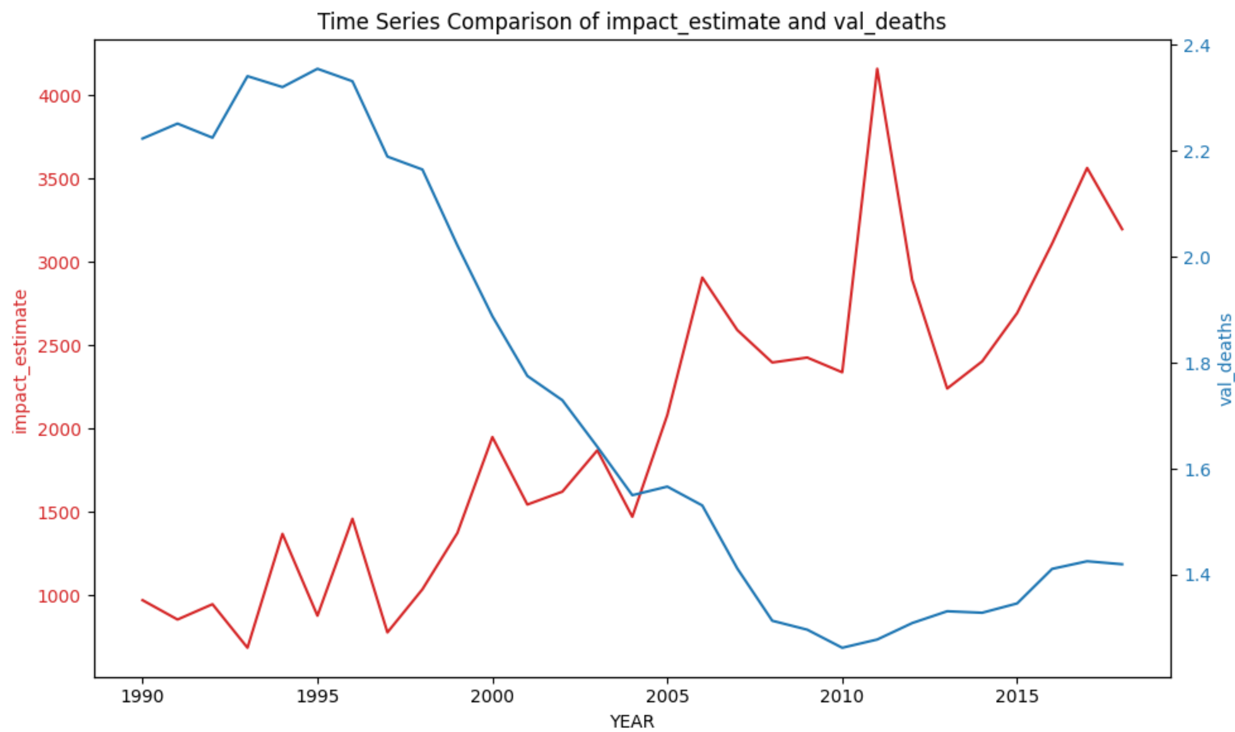

Figure 2: Comparision of my smoke impact estimate and the rate of deaths caused due to asthma

2. Asthma Incidents and Smoke Impact (Correlation Value 0.5)

I observed a moderate positive correlation between smoke impact and asthma incidents, suggesting that higher smoke levels might contribute to an increase in asthma cases.

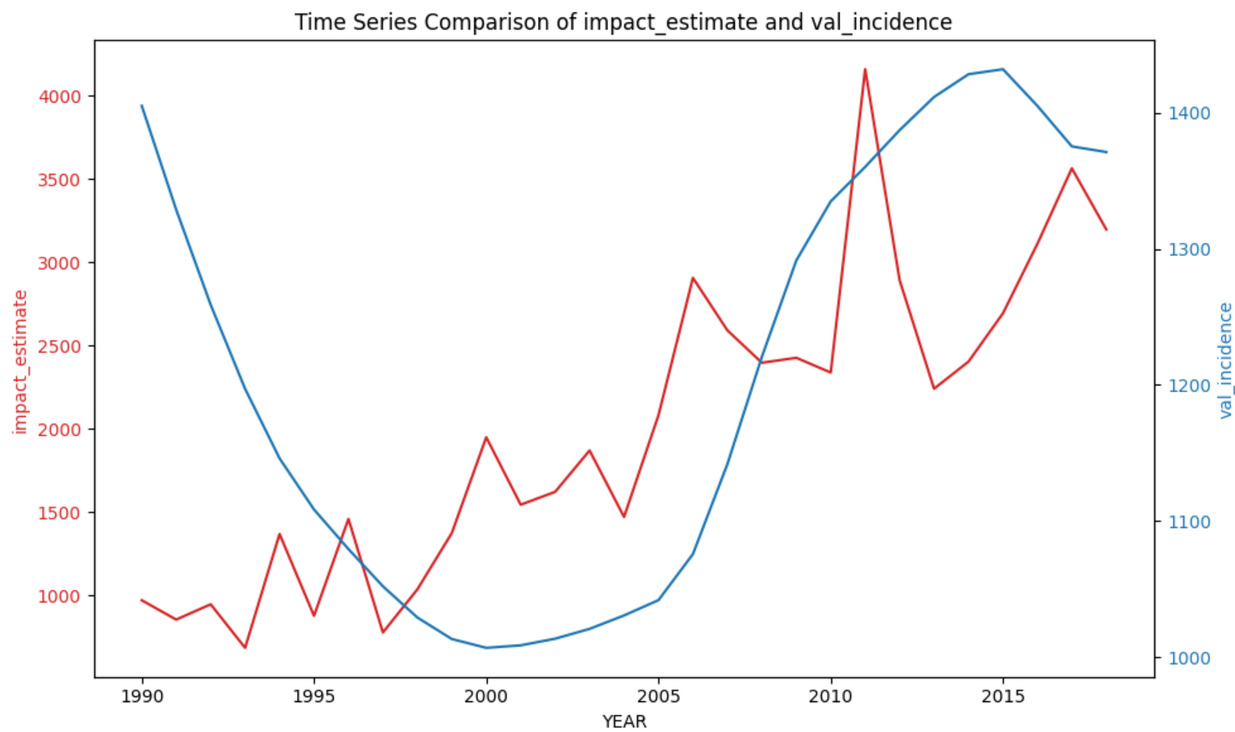

Figure 3: Comparision of my smoke impact estimate and the rate of asthma incidents

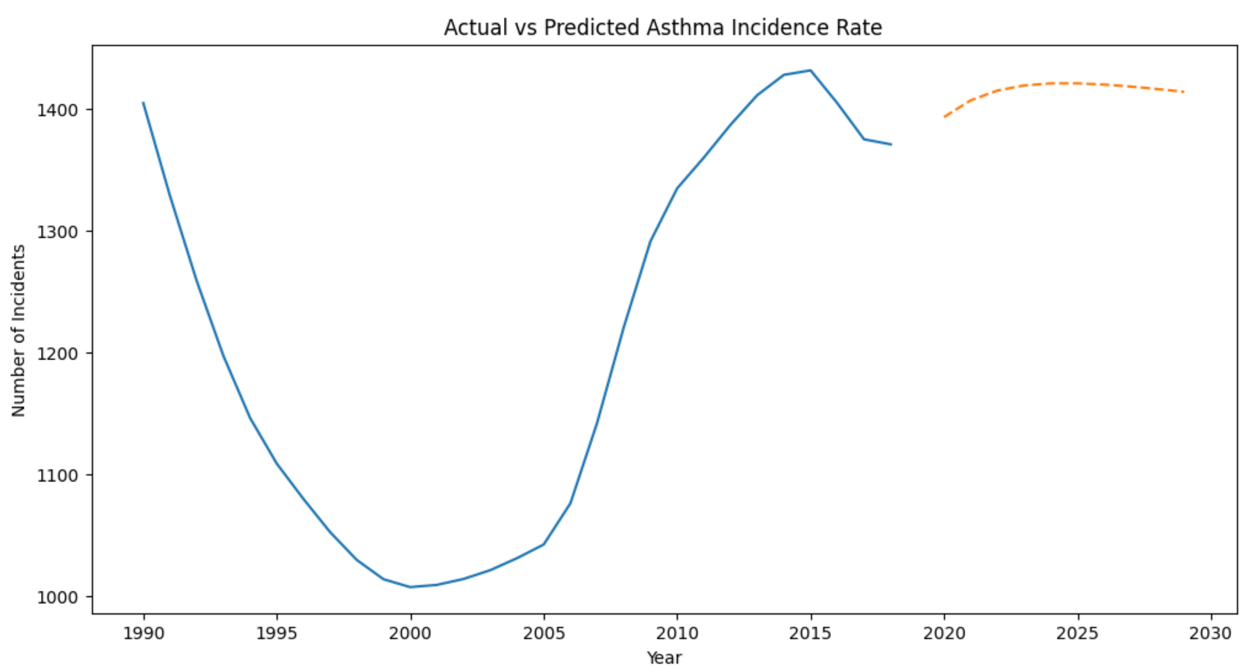

Figure 4: Predicted rate of asthma incidence using the smoke impact and historical incidence rates as features

3. Women's Birth Rate and Smoke Impact (Correlation Value 0.1)

The correlation between smoke impact and women's birth rate was found to be weak, indicating that factors other than smoke, possibly socio-economic or healthcare-related, have a more substantial impact on birth rates.

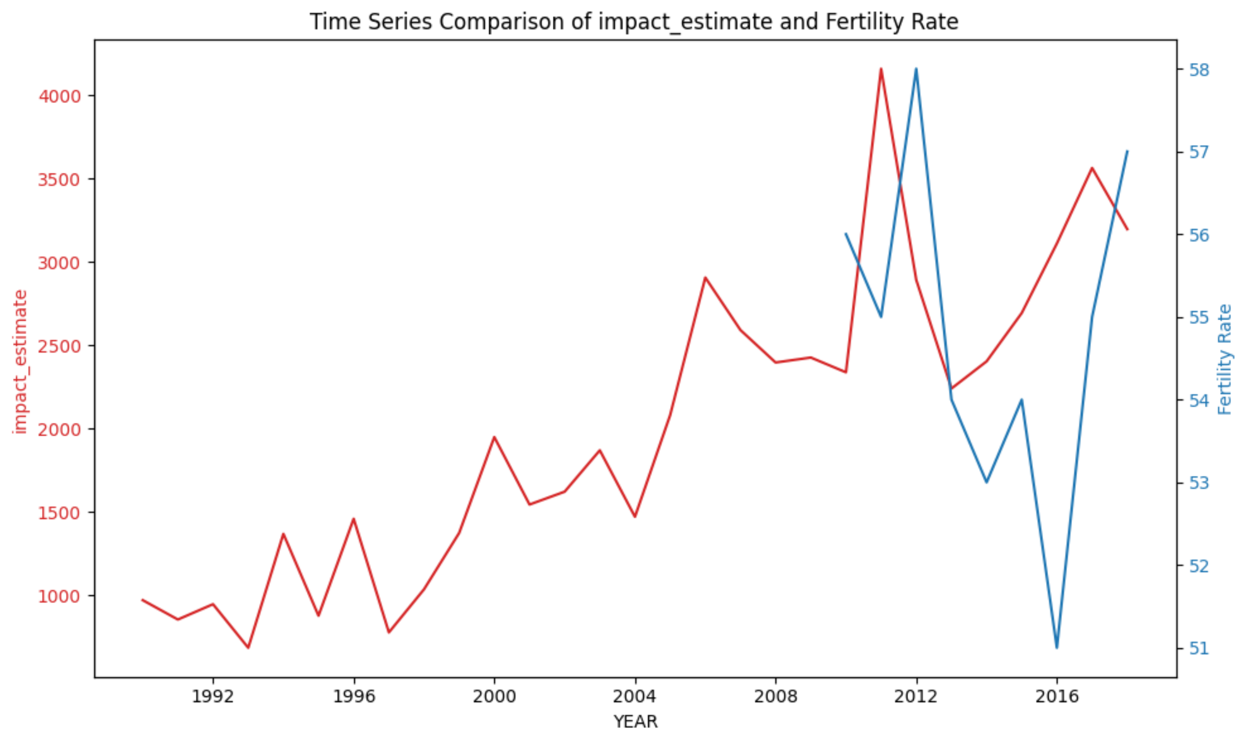

Figure 5: Comparision of my smoke impact estimate and fertility rates for women between the ages of 11-50

4. Premature Death Rate and Smoke Impact (Correlation Value 0.4)

There was a moderate positive correlation between smoke impact and the premature death rate, indicating a significant health concern that warrants immediate attention.

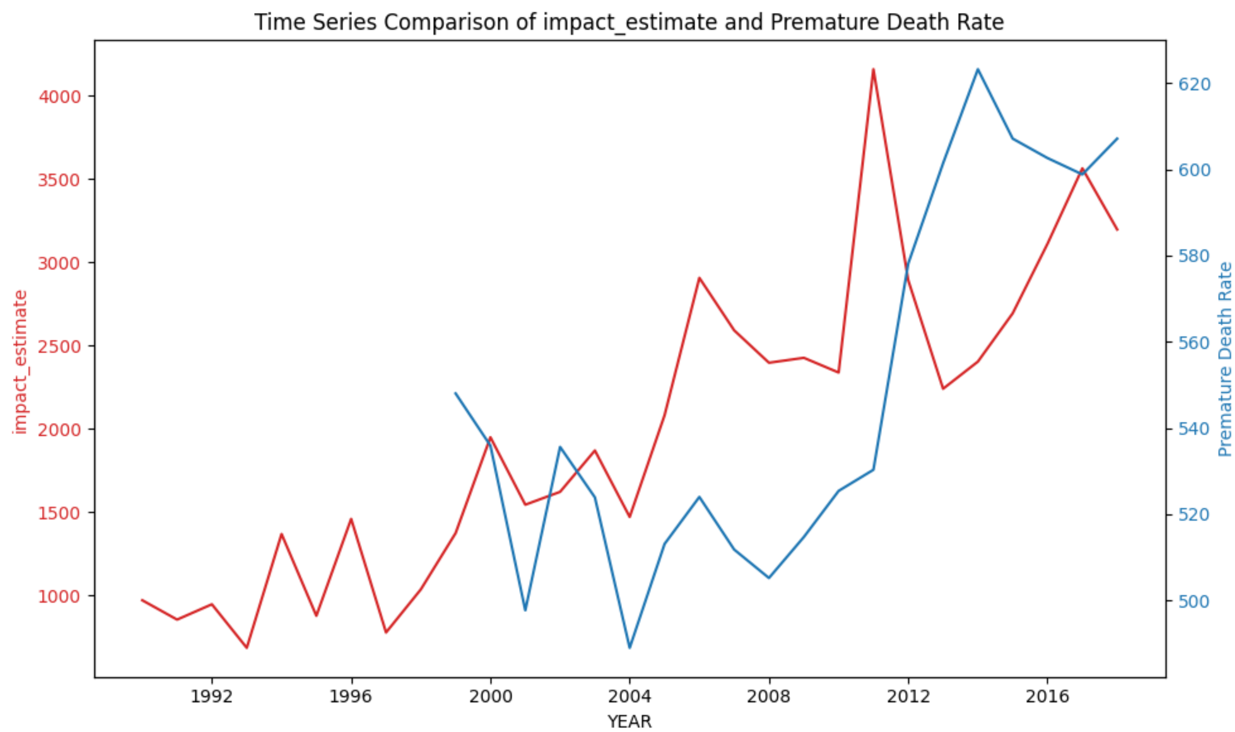

Figure 5: Comparision of my smoke impact estimate and premature deaths

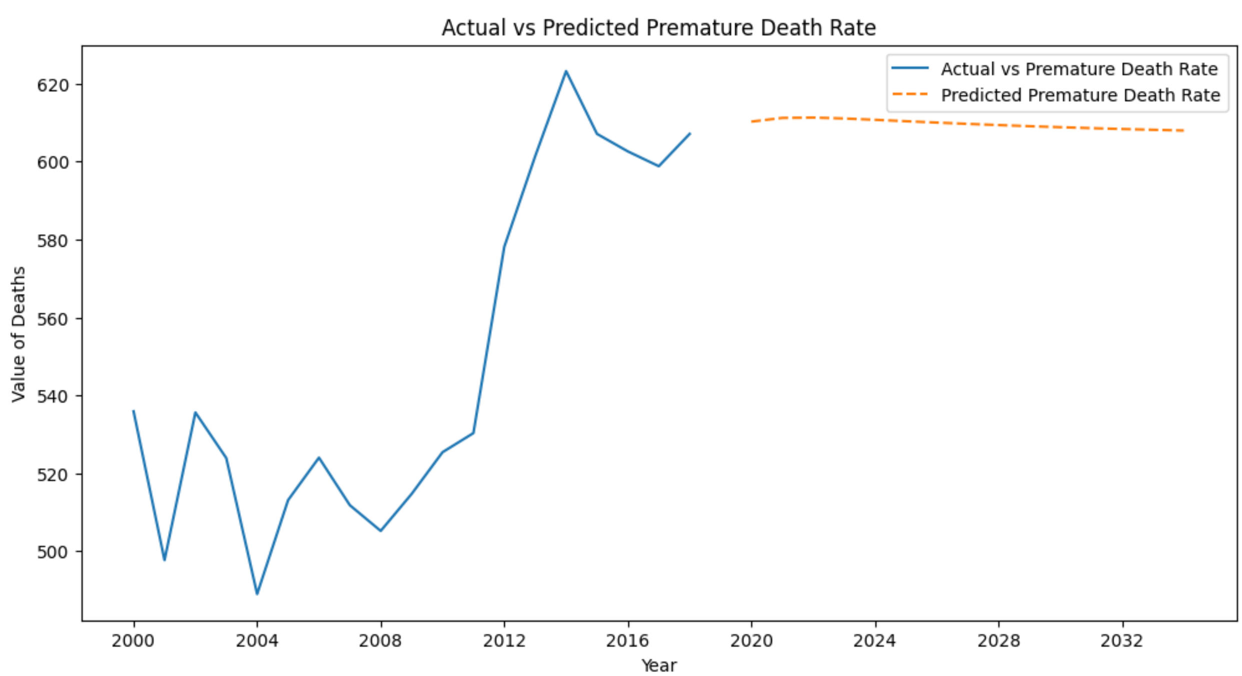

Figure 6: Predicted rate of premature death rate using the smoke impact and historical premature death rate as features

Discussion/Implications

The correlations observed in this study underscore the necessity of addressing wildfire smoke as a public health issue. We see either an increase or a plateau in the incidence rates of asthma and premature death rate which is undesirable. Immediate actions, such as enhancing healthcare infrastructure and conducting public awareness campaigns, are crucial. Long-term strategies, including urban planning and collaboration with environmental experts, are also vital. My findings should inform the city council, city manager/mayor, and city residents in their efforts to develop effective strategies to mitigate the impact of wildfire smoke.

Immediate Actions

- Enhance Healthcare Infrastructure: Strengthen healthcare services, especially during peak wildfire seasons, to manage potential increases in respiratory ailments.
- Public Awareness Campaigns: Educate residents about the health risks of wildfire smoke and effective personal protection measures.

Long-Term Strategies

- Urban Planning and Green Spaces: Invest in urban green spaces to improve air quantity. This can act as a natural buffer against the spread of smoke within city limits
- Collaboration with Environmental Experts: Work closely with environmental agencies to monitor and predict wildfire occurrences and smoke dispersion patterns
- Emergency Response Plans: Develop and regularly update emergency response plans that include provisions for evacuation, medical emergencies, and public shelters with air filtration systems.

Research and Monitoring

- Ongoing Research: Encourage and fund research to explore the long-term health impacts of repeated exposure to wildfire smoke.
- Monitoring Air Quality: Invest in city air quality monitoring systems to provide residents with real-time data and warnings.

Policy and Regulatory Measures

- Building Codes and Regulations: Implement building codes that require air filtration systems in public buildings and residential complexes, especially in high-risk areas.
- Environmental Policies: Advocate for and implement environmental policies to reduce the likelihood and severity of wildfires.

Limitations

This study has various limitations, including the reliance on historical data, which may only account for some variables influencing health outcomes. Additionally, the method of estimating smoke impact based on fire data may not capture the full complexity of smoke dispersion and its health effects. The principal assumption is that my chosen health markers are primarily influenced by smoke exposure, setting aside other potential external factors for focused analysis. Another assumption that my correlation metric relies on is normally distributed data and a linear relationship between the features. Since we only used a few years of data with a negative or positive trend (asthma rates going down or population increasing yearly), we cannot be too sure about my normality assumption. The investigation of a linear relationship is out of the scope of this investigation.

Another significant limitation of my current analysis is the exclusion of various external factors that could significantly impact health outcomes, such as medical advancements and regional features that affect health. For instance, improvements in medical treatments and healthcare accessibility could reduce the number of deaths due to asthma. Another example is the effects of lifestyle on fertility rates, wherein factors such as stress, nutrition, and exercise could skew the perceived impact of smoke exposure one way or the other. Such omissions could lead to an oversimplification of the complex interplay between smoke exposure and health outcomes and, in extreme cases, even produce incorrect results.

Additionally, my initial smoke impact estimate was created without a target metric and is therefore unreliable and does not conclusively model any actual relationship between the area of a fire and the distance of the fire from Prescott City with the impact of smoke coming from these fires. There are many additional features of interest that my model ignores. For instance, it does not consider the direction of the wind that carries the smoke.

Moreover, the uncertainty inherent in any data analysis exercise is also present here. Potential biases in the datasets, variabilities in data collection methods, and the reliability of predictive models based on historical data all introduce possible elements of errors to my analysis. Moreover, I discarded the distinction between correlation and causation for this assignment and my downstream analysis. This is generally a big mistake but is overlooked due to the limited scope of this assignment and the short period given for examination.

Understanding this analysis's shortcomings and baseline assumptions is essential before deriving any actionable insights.

Conclusion

The study highlights the urgent need for a comprehensive approach to mitigate the impact of wildfire smoke on public health in Prescott City. Proactive policies, community engagement, and continuous research are crucial in safeguarding the community's well-being against the increasing threat of wildfires.

The scope of our initial analysis was to investigate the relationship between smoke and various health factors. We successfully established quantitative and qualitative correlations. Additionally, the study informs the users about the human-centric aspect of quantitative analysis by highlighting the impacts of smoke, the importance of understanding the assumptions, the limitations of the study, and possible next steps towards mitigating the possible impact of smoke that our study has predicted for the future. All in all, we hope that shit study helps policymakers make informed decisions to avoid the ill effects of smoke in the future,

Data Sources

1. **Combined Wildland Fire Datasets (US Geological Survey)**
   The Combined Wildland Fire Datasets, developed by the U.S. Geological Survey, is a comprehensive collection of wildfire data for the United States and certain territories from the 1800s to the present. It integrates data from 40 wildland fire sources, each with varying spatial scales, resolutions, and periods, into a unified dataset. This dataset provides a single set of polygons with a singular fire boundary for each fire, aiming to create a more comprehensive and accurate representation of fire data while reducing duplication. It includes detailed attributes for each fire

event. It is intended for a wide range of applications, from environmental research to policy-making, offering a valuable resource for understanding wildfires' spatial and temporal patterns.

2. **American Community Survey 1-Year Data (2005-2022) [1]: Female Fertility [2]**
   The American Community Survey (ACS) is an ongoing survey that provides yearly data, giving communities the current information they need to plan investments and services. The ACS covers various topics about the U.S. population's social, economic, demographic, and housing characteristics. This data is aggregated and publically hosted by the US Census Bureau [3]. US Census Bureau data is available for public use [12].

   Within this dataset of over 600 columns, I focused on three:

   - S1301_C04_001E: Estimate of Women with births in the past 12 months per 1000 for ages 15 to 50 years
   - S1301_C04_001M: Margin of Error Women with births in the past 12 months per 1000 for ages 15 to 50 years
   - YEAR: Year of data collection.

3. **Centers for Disease Control and Prevention Data[4]: Premature Mortality Rate [5]**
   The crude death rate is the number of deaths reported each calendar year divided by the population multiplied by 100,000. 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 premature death according to the CDC's definition of Years of Potential Life Loss. Although the CDC collects the original dataset, an aggregated version of this dataset, compiled by the Federal Research Bank of St. Louis [6], will be used for this analysis. CDC data is available for public use [11].

   This dataset is relatively straightforward and contains only two columns:

   - CDC20N2U004025: Premature mortality rate
   - DATE: Date of data collection

4. **Institute for Health Metrics and Evaluation Global Burden of Disease: Asthma Data [7]**

The Global Burden of Disease (GBD) study provides a comprehensive picture of mortality and disability across countries, time, age, and sex. It quantifies health loss from hundreds of diseases, injuries, and risk factors to improve health systems and eliminate disparities. [7] For the academic scope of this analysis, this data is available free of charge under a non-commercial user agreement [10].

After subsetting the data for the relevant disease, I used the following data:

- Rate of deaths caused due to Asthma.
- Rate of Ashtma incidence.

Which will be derived from the following columns:

- measure_name: This indicates the type of data or metric being reported, such as incidence and mortality rate,
- location_name: Refers to the geographical area or region to which the data applies, ranging from global to country-specific levels.
- sex_name: Specifies the biological sex (male, female) for which the data is reported, or 'both' if the data encompasses all sexes.
- age_name: Denotes the age group or range for which the data is relevant, such as 0-5 years, 15-49 years, or all ages.
- cause_name: Identifies the specific health condition, disease, or cause of death being studied or reported on. In my case, this will be ‘Asthma.’
- metric_name: Refers to the specific measurement used in the dataset, such as rate, number, or proportion.
- year: Indicates the calendar year to which the data corresponds.
- val: Represents the primary value or measurement reported for the given parameters.
- upper: Gives the upper bound or the higher estimate of the confidence interval for the reported value.
- lower: Provides the lower bound or the lower estimate of the confidence interval for the reported value.

References

1. Bureau, U.C. (2020). American Community Survey 1-Year Data (2005-2019). [online] The United States Census Bureau. Available at: https://www.census.gov/data/developers/data-sets/acs-1year.html.

2. Explore Fertility Census Data. [online] Available at: https://data.census.gov/table/ACSST1Y2022.S1301?t=Fertility
3. Explore Census Data. [online] Available at: https://data.census.gov/table
4. CDC (2018). Data & Statistics. [online] Centers for Disease Control and Prevention. Available at: https://www.cdc.gov/datastatistics/index.html
5. HDPulse - Health Disparities Resources. [online] Available at: https://hdpulse.nimhd.nih.gov/
6. Centers for Disease Control and Prevention, Premature Death Rate for Yavapai County, AZ [CDC20N2U004025], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/CDC20N2U004025, November 15, 2023.
7. Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2020. Available from https://vizhub.healthdata.org/gbd-results/
8. Institute for Health Metrics and Evaluation. (2021). IHME free-of-charge non-commercial user agreement. [online] Available at: https://www.healthdata.org/Data-tools-practices/data-practices/ihme-free-charge-non-commercial-user-agreement.
9. www.cdc.gov. (2019). Data Access - Public-Use Data Files and Documentation. [online] Available at: https://www.cdc.gov/nchs/data_access/ftp_data.html
10. Bureau, U.C. (n.d.). License Information and Disclaimer. [online] Census.gov. Available at: https://www.census.gov/data/software/x13as/disclaimer.html
11. Akoglu, Haldun. 2018. "User's Guide to Correlation Coefficients." Turkish Journal of Emergency Medicine 18 (3): 91–93. https://doi.org/10.1016/j.tjem.2018.08.001.
12. Wayne E. Cascio, "Wildland fire smoke and human health," Science of The Total Environment, Volume 624, 2018, Pages 586–595, ISSN 0048-9697, https://doi.org/10.1016/j.scitotenv.2017.12.086.
13. Reid, C.E., Brauer, M., Johnston, F.H., Jerrett, M., Balmes, J.R. and Elliott, C.T. (2016). Critical Review of Health Impacts of Wildfire Smoke Exposure. Environmental Health Perspectives, [online] 124(9), pp.1334–1343. doi:https://doi.org/10.1289/ehp.1409277.
14. Black, C., Tesfaigzi, Y., Bassein, J.A. and Miller, L.A. (2017). Wildfire smoke exposure and human health: Significant gaps in research for a growing public health issue. Environmental Toxicology and Pharmacology, [online] 55, pp.186–195. doi https://doi.org/10.1016/j.etap.2017.08.022.
15. Combined wildland fire datasets for the United States and certain territories, 1800s-Present (combined wildland fire polygons) - ScienceBase-Catalog. [online]

Available at: https://www.sciencebase.gov/catalog/item/61aa537dd34eb622f699df81.