Health Impacts of Wildfire Smoke in Farmington, New Mexico (NM)

DATA 512 AU 2023 – Course Project

Adithyaa V

Table of Contents

1. Introduction

New Mexico has been at the forefront of a disturbing trend in the western United States - the increasing frequency and severity of wildfires. In 2022, the state witnessed one of its most devastating wildfire seasons [1], fuelled by factors such as climate change and historical land management approaches. These wildfires have not only wreaked havoc on the environment and infrastructure but have also raised significant health concerns due to the smoke they generate.

Figure 1. Farmington's Location

The city of Farmington, with its 46,422 residents as of 2021, finds itself directly affected by these environmental disasters. The health implications of the extensive smoke produced by nearby incidents, such as the significant blaze in San Juan County [2], are a source of growing concern. This study is thus critical in understanding how wildfire smoke impacts health outcomes in a moderately sized urban population, shedding light on potential public health crises and necessary interventions.

This research is driven by the urgent need to understand and quantify the health impacts of wildfire smoke in Farmington. By analyzing historical wildfire data, assessing smoke penetration, and forecasting future trends, the study aims to provide vital information that can guide public health interventions and policy decisions. The goal is to not only document the current health impacts but also anticipate future challenges. By doing so, it seeks to inform local authorities, healthcare providers, and the general public, providing them with vital information to strategize effective responses and health measures against the backdrop of increasing wildfire smoke exposure.

2. Background

Our analysis of the health impacts of wildfire smoke in Farmington draws from a range of pivotal studies. The New York State Department of Health highlights immediate health effects like eye and respiratory irritation and more severe impacts on lung and heart function from smoke inhalation, underscoring the necessity of our investigation into both immediate and chronic health effects in Farmington [3]. Complementarily, Environmental Health Perspectives link long-term exposure to particulate matter from wildfire smoke to an increased risk of cardiovascular diseases and mortality, reinforcing our focus on prolonged health impacts [4]. Research in Fertility and Sterility illuminates the potential relationship between smoke exposure and reproductive health, particularly fertility rates, shaping our inquiry into reproductive health consequences [5]. Additionally, a study in the Journal of Forest Economics connects particulate matter exposure from wildfire smoke to cardiovascular diseases, guiding our exploration of cardiovascular health risks [6]. These findings based on different demographics collectively deepen our understanding of smoke-related health risks, directing our attention to immediate, long-term, and specific health concerns like cardiovascular and reproductive health in Farmington, NM. Informed by the above, our study poses three critical questions:

1. What are the key health factors impacted by wildfire smoke?
2. Is the impact of wildfire smoke immediate, or does it exhibit a time-lagged effect?
3. How do future health trends appear considering continued wildfire smoke exposure?

To answer our research questions about the health impacts of wildfire smoke in Farmington, NM, we began by examining the occurrence and nature of smoke events in the area. We utilized USGS data to assess wildfire frequency, scale, and proximity to Farmington, factoring in variables such as distance from the city, acres burned, and wildfire type.

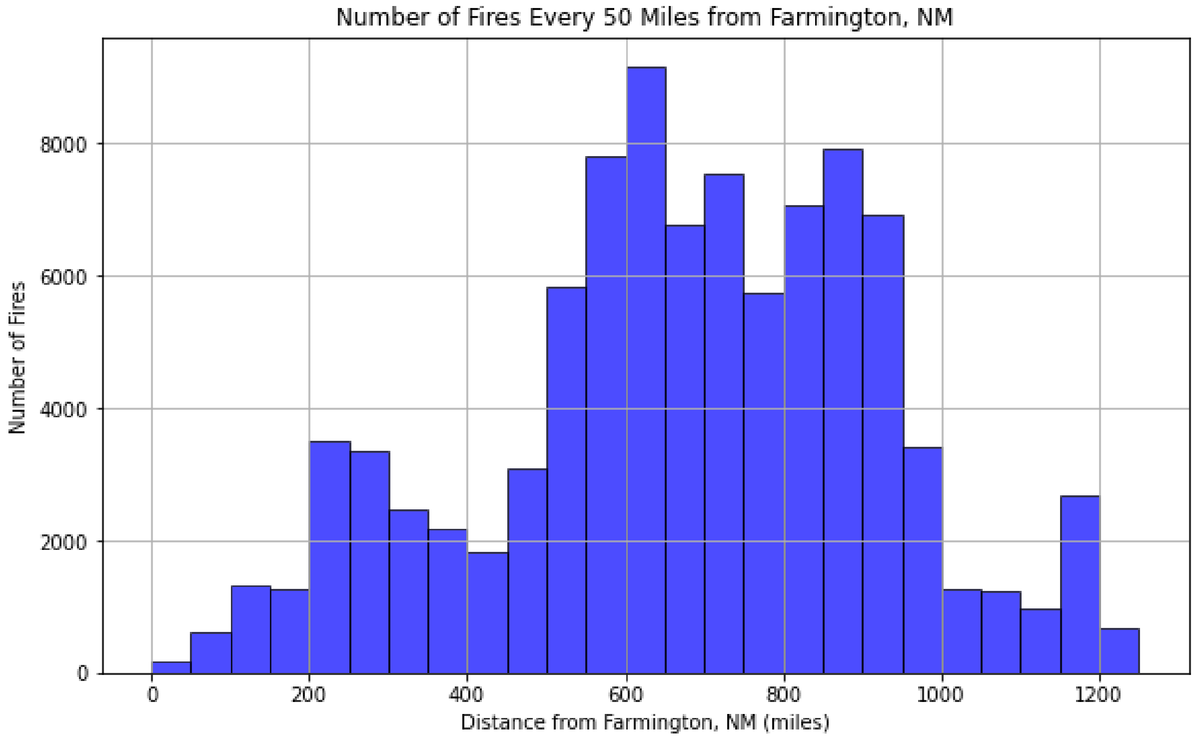

Figure 2. Number of fires Bucketed within Incremental Distances of 50 from Farmington

The histogram displays the distribution of wildfires within incremental 50-mile distances from Farmington, NM, from 1963 to 2020. Each bar signifies a 50-mile interval, with its height indicating the number of fires in that range. The x-axis, marked in 50-mile segments, spans from 0 to over 1200 miles, while the y-axis tracks the number of fires, peaking above 8000. The analysis shows notable fire occurrences within a 1,250-mile radius, with distinct peaks in certain intervals, suggesting varied but significant fire activity at different distances from Farmington.

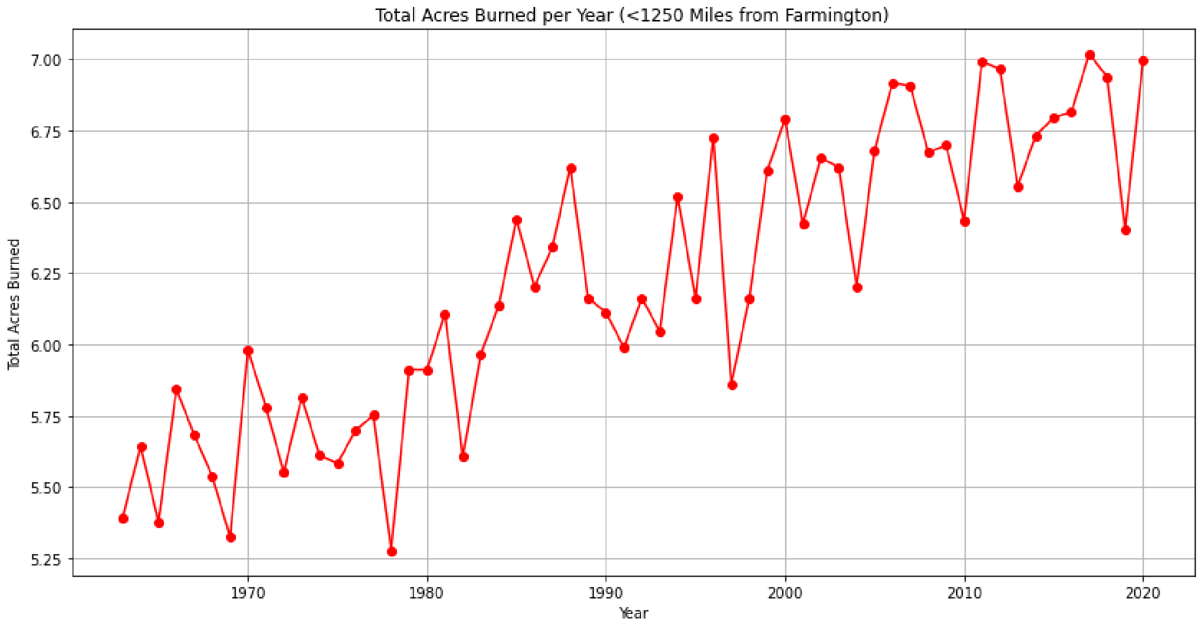

Figure 3. Total Acres Burned per Year for Fires within 1250 Miles of Farmington over a time

The line graph showcases the Total Acres Burned annually near Farmington from 1963 to 2020. Time is plotted on the x-axis, with markers every decade and finer lines for individual years. The y-axis, with a logarithmic scale, measures acres burned in millions. Red dots represent yearly totals, linked by a line to depict trends. This graph reveals a fluctuating yet increasing trend in acres burned, hinting at complex influences like climate change and land management practices, and forms the basis for linking wildfire events to health impacts in Farmington.

To quantify wildfire smoke impact in Farmington, NM, we developed a Smoke Estimate, incorporating factors like fire size, and proximity to Farmington. The model's foundation is based on EPA research [7], suggesting various influencing factors on smoke production.

The formula for our Smoke Estimate per Fire is illustrated in Figure 4.

$$\frac{Size\ of\ the\ Fire}{Proximity\ to\ Farmington}$$

Figure 4. Smoke Estimate Formula

This captures how smoke impact decreases with distance from Farmington, and increase with the size of the Fire. Annually, we calculate an average smoke impact, offering a balanced view of overall exposure, accounting for both numerous small fires and fewer large fires.

To verify the accuracy of our smoke estimates, we utilized air quality data from the United States Environmental Protection Agency (EPA). This data is accessible through the EPA's Air Quality System (AQS) API, which provides the Air Quality Index (AQI) data from numerous air quality monitoring stations across the United States. This AQI data is crucial for comparing against our smoke estimates to ensure their reliability. It is important to note that the AQI data is restricted to the fire season that runs from May to October every year.

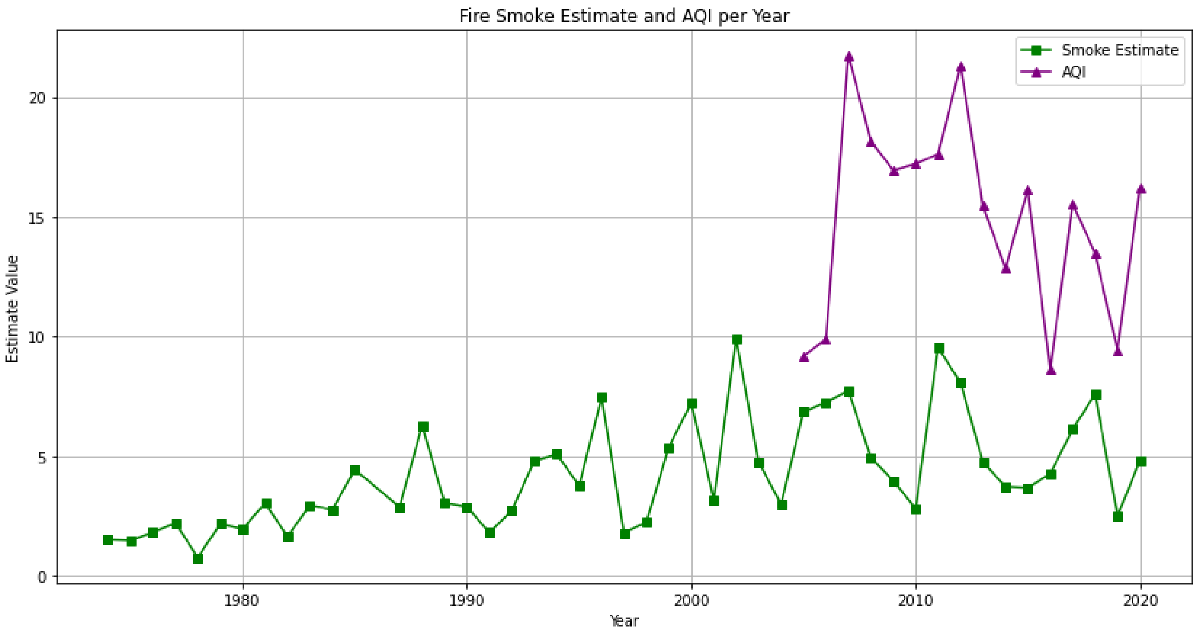

Figure 5. Comparison of Smoke Estimate with the Air Quality Index

The graph compares the Smoke Estimate with the Air Quality Index (AQI) over a set time. The Smoke Estimate is plotted with a green line and square markers, while the AQI is represented by a purple line and triangular markers. Both metrics show variability over time. Peaks in the Smoke Estimate suggest increased wildfire activity or larger fires, while spikes in the AQI could indicate the presence of smoke or other pollutants affecting air quality. The small correlation between the two lines might suggest a relationship between estimated smoke from fires and recorded air quality, validating the Smoke Estimate against actual air quality measurements.

To understand the trends of future smoke estimates, we use an ARIMA model [8] to predict the smoke estimate for 2021-2045.

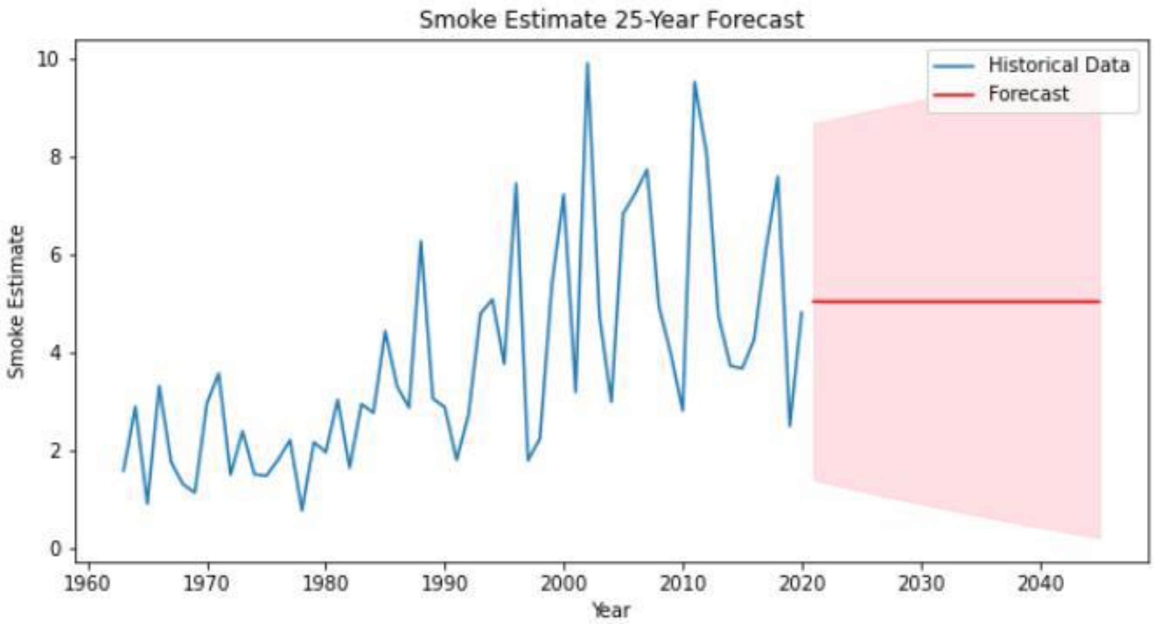

Figure 6. Smoke Estimate Forecast for 2021-2045

Figure 6 presents the future trajectory of smoke impact due to wildfires around Farmington up to the year 2045, using the model's results. The historical smoke estimates are depicted with a blue line, while the forecasted estimates are shown within a pink-shaded area. This area signifies a 95% confidence interval, within which the true smoke estimates are expected to fall, indicating a significant anticipated smoke penetration. This analysis was conducted to proactively understand and address the potential widespread health impacts of smoke in the future, highlighting the necessity for targeted health impact studies and public health interventions.

Understanding the future health impacts of smoke exposure necessitates a detailed analysis of vital health parameters. We will concentrate on mortality rate, fertility rate, and Medicare enrollees data as these metrics will provide a comprehensive view of the health repercussions within the community. These parameters were selected for their relevance to the health concerns raised by smoke exposure, allowing for a focused investigation into the direct and indirect effects on public health in Farmington. Here is a summary of the new data:

Fertility Data (US Census):

- Source: American Community Survey (ACS) by the U.S. Census Bureau.
- Coverage: Farmington City, NM, from 2010 to 2021.
- Key Columns: Label (Name of the parameter), Estimate (Fertility Rate %).

Mortality & Medicare Enrollees Data (Dartmouth Atlas):

- Source: Dartmouth Atlas Data Website.
- Coverage: Farmington City, NM, from 1999 to 2019.

- Key Columns: HSA Name (City name), Total Mortality Rate (%), Medicare Enrollees.

3. Methodology

To answer the research questions, a combination of statistical methods, including correlation analysis, lag analysis, and regression analysis, was employed. These methods were chosen for their ability to reveal relationships between variables and to assess the temporal dynamics of these relationships.

Correlation Analysis:

The correlation analysis aimed to examine associations between Smoke Estimates and various health-related factors such as Medicare Enrollees, Mortality Rate, and Fertility Rate, along with AQI. This step, crucial in identifying the most significantly impacted health factors by wildfire smoke, was approached with an emphasis on community relevance and data integrity.

- **Method:** Using Python's seaborn and matplotlib libraries, the correlation matrix and specific correlation values with Smoke Estimate were visualized, ensuring transparency and methodological clarity.
- **Rationale:** This method helps in understanding potential relationships between smoke exposure and different health outcomes, guiding further in-depth analysis with a focus on the ethical implications of these findings.

Lag Analysis:

Given the established impact of Smoke Estimates on certain health parameters, Lag Analysis was conducted to determine whether these impacts are immediate or delayed, keeping in mind the need for timely and effective public health interventions.

- **Method:** Correlations for different time lags (up to 10 years) were calculated to explore the temporal dynamics of smoke exposure effects on health outcomes, prioritizing the community's long-term well-being and health.
- **Rationale:** This analysis is crucial in understanding the immediate versus long-term health impacts, a key consideration in the development of public health strategies.

Regression Analysis:

To understand current trends and project future impacts of smoke estimates on health outcomes, regression analysis using Ordinary Least Squares (OLS) [9] was performed.

- **Method:** After data cleaning to remove null values, regression analyses for each health-related feature against Smoke Estimate were conducted, ensuring the study's findings are reflective of the actual community scenario.
- **Rationale:** This analysis allows exploration of the relationship between smoke exposure and health outcomes, providing insights into the severity and direction of these future impacts, with an overarching goal of informing health policymakers and practitioners.

Human-Centered and Ethical Considerations:

In every step of this methodology, careful consideration was given to ensure that the study remains directly relevant and respectful to the Farmington community. By prioritizing reliable and ethically sourced data, the study maintains a high standard of accuracy and integrity. The transparency in methodological choices, combined with an awareness of the potential ethical implications of the findings, underscores a commitment to respecting community well-being and privacy. The primary focus on public health implications reflects the goal of providing actionable insights for health policymakers and practitioners, underlining the human-centered approach of this research.

4. <u>Findings</u>

The analysis conducted for this study provided important insights into the health impacts of wildfire smoke in Farmington, NM. The key findings are summarized below:

Correlation Analysis:

The correlation analysis yielded the following significant results:

- **Mortality Rate:** There was a strong positive correlation of 0.80 between Smoke Estimates and Mortality Rate. This indicates a substantial association where increased smoke exposure correlates with higher mortality rates. Figure 7 vividly illustrates this.
- **Fertility Rate:** A notable negative correlation of -0.66 was found between Smoke Estimates and Fertility Rate, suggesting that increased smoke exposure may be linked to lower fertility rates. This inverse relationship is depicted in Figure 7.
- **Yearly Average AQI and Medicare Enrollees:** These variables showed relatively lower correlations with Smoke Estimates, suggesting less pronounced associations.

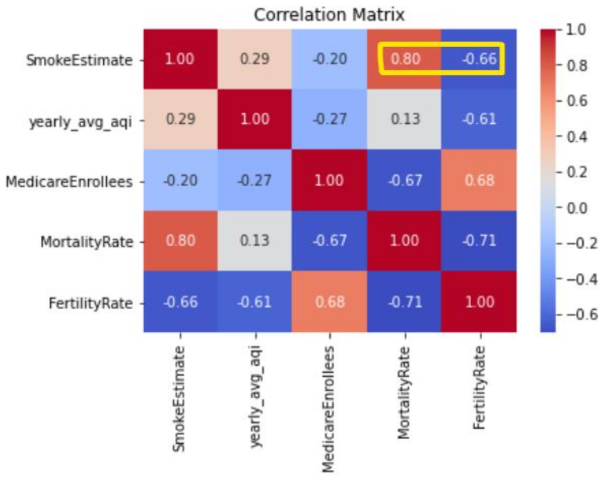

Figure 7. Correlations Heat Map

Lag Analysis:

The Lag Analysis revealed:

- The most significant impact was observed immediately (lag 0) with correlations of 0.80 for the Mortality Rate and -0.66 for the Fertility Rate, indicating pronounced immediate effects of smoke exposure. This is critical for understanding the urgency of addressing smoke-related health issues.
- The varying impacts over lags of up to 10 years are presented in Figure 8, showcasing the changing strength of correlation over time and highlighting the complex nature of long-term health impacts.

	MortalityRate	FertilityRate
0	0.801985	-0.659865
1	-0.080179	-0.239979
2	-0.323467	0.390893
3	-0.128740	0.201009
4	-0.170621	-0.153845
5	-0.017739	-0.379966
6	0.217751	-0.284827
7	-0.323966	0.042035
8	-0.621439	0.319870
9	-0.088049	0.037827
10	-0.410423	-0.178749

Figure 8. Lagged Correlation Values

Regression Analysis:

The regression analysis further elucidated these relationships:

- **Mortality Rate:** The regression analysis with Smoke Estimate showed an R-squared of 0.899, indicating that approximately 89.9% of the variance in Mortality Rate is explained by Smoke Estimate. This suggests a strong linear relationship, as depicted in the regression graph.

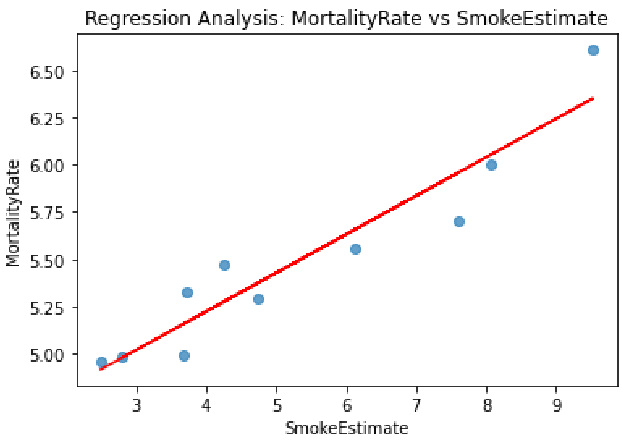

Figure 9. Trend Plot of Smoke Estimates and Mortality Rates

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          MortalityRate   R-squared:                       0.899
Model:                            OLS   Adj. R-squared:                  0.886
Method:                 Least Squares   F-statistic:                     70.83
Date:                Sat, 09 Dec 2023   Prob (F-statistic):           3.03e-05
Time:                        19:27:46   Log-Likelihood:                 4.3520
No. Observations:                  10   AIC:                            -4.704
Df Residuals:                       8   BIC:                            -4.099
Df Model:                           1
Covariance Type:            nonrobust
=================================================================================
                    coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------
const             4.4060      0.140     31.450      0.000       4.083       4.729
SmokeEstimate     0.2042      0.024      8.416      0.000       0.148       0.260
==============================================================================
Omnibus:                        0.429   Durbin-Watson:                   2.672
Prob(Omnibus):                  0.807   Jarque-Bera (JB):                0.481
Skew:                           0.144   Prob(JB):                        0.786
Kurtosis:                       1.965   Cond. No.                         15.0
==============================================================================
```

Figure 10. Mortality Rate Regression Analysis Result

- **Fertility Rate:** The regression analysis for Fertility Rate revealed an R-squared of 0.458, implying that around 45.8% of the variability in Fertility Rate is explained by Smoke Estimate, indicating a moderate correlation. The negative coefficient of -0.6947 suggests a decrease in Fertility Rate with an increase in Smoke Estimate.

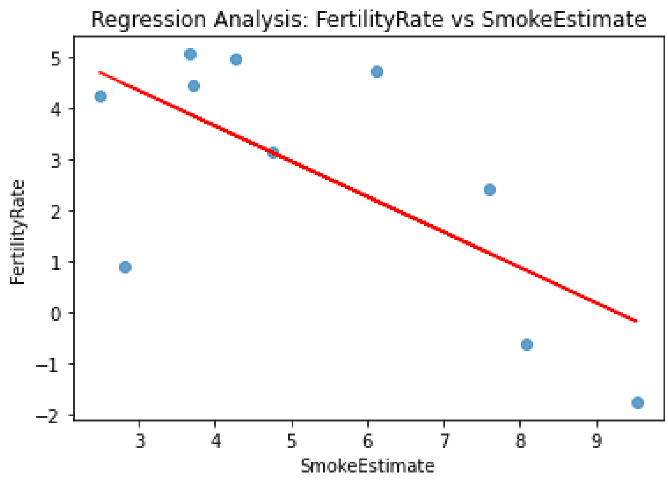

Figure 11. Trend Plot of Smoke Estimates and Fertility Rates

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          FertilityRate   R-squared:                       0.458
Model:                            OLS   Adj. R-squared:                  0.390
Method:                 Least Squares   F-statistic:                     6.750
Date:                Sat, 09 Dec 2023   Prob (F-statistic):             0.0317
Time:                        19:27:46   Log-Likelihood:                -19.644
No. Observations:                  10   AIC:                             43.29
Df Residuals:                       8   BIC:                             43.89
Df Model:                           1
Covariance Type:            nonrobust
=================================================================================
                    coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------
const             6.4325      1.544      4.167      0.003       2.873       9.992
SmokeEstimate    -0.6947      0.267     -2.598      0.032      -1.311      -0.078
==============================================================================
Omnibus:                        1.153   Durbin-Watson:                   0.426
Prob(Omnibus):                  0.562   Jarque-Bera (JB):                0.617
Skew:                          -0.571   Prob(JB):                        0.735
Kurtosis:                       2.577   Cond. No.                         15.0
==============================================================================
```

Figure 12. Fertility Rate Regression Analysis Result

5. Discussion

The findings from this study, highlighting the impact of wildfire smoke on public health in Farmington, NM, reveal a scenario that demands immediate attention and action. The strong positive correlation observed between smoke exposure and increased mortality rates, along with the inverse relationship with fertility rates, underscores a critical public health challenge. These results are not merely statistical indicators; they represent a tangible threat to the well-being of the Farmington community. Importantly, the regression analysis not only confirms these trends but also projects them into the future, suggesting that if no action is taken, these adverse health impacts are likely to persist or even intensify.

The implications of this research extend to all stakeholders in Farmington. City officials, including the city council and mayor, are faced with the urgent task of developing a comprehensive response to mitigate these health risks. The immediacy of the impact of smoke exposure, as shown in the study, suggests that planning and implementation of any action should ideally commence within the next few months, certainly before the onset of the next wildfire season. Potential strategies might include enhancing the city's air quality management, implementing public health initiatives focused on respiratory and reproductive health, and launching educational campaigns to raise awareness about the risks of smoke exposure. For residents, adopting personal protective measures such as using air purifiers and wearing masks during high-risk periods could be lifesaving. Moreover, engaging in community-driven environmental initiatives could amplify the collective effort to combat the effects of wildfire smoke.

Reflecting on the project, the principles of human-centered data science played a pivotal role in guiding the research methodology and interpretation of results. The study was designed with a strong emphasis on the relevance and applicability of its findings to the local community. By focusing on specific health outcomes pertinent to the residents of Farmington and employing methods that provide actionable insights, the study aligns closely with the ethos of human-centered data science. This approach ensured that the research was not only statistically sound but also meaningful and beneficial to the community it aimed to serve. The ethical implications of the findings, particularly in terms of public health and privacy, were carefully considered, ensuring that the study's conclusions respect and prioritize the well-being and interests of the Farmington community.

6. Limitations

These limitations span across data sources, methodological approaches, and potential external influencing factors:

1. **Historical Data Limitations:** The USGS dataset's focus post-1963 may not fully represent long-term wildfire trends. Pre-1980 data is less reliable, likely underreporting earlier wildfire occurrences.
2. **Data Accuracy and Completeness:** Challenges include standardizing data from multiple sources, underreporting of fire occurrences, and fire classification uncertainties, potentially affecting the study's accuracy.
3. **Methodological Constraints:** Limitations in fire distance calculations and the exclusion of certain fire shapes could omit relevant data. The Smoke Estimate model's assumptions might not fully represent real-world conditions.
4. **Air Quality Data Limitations:** Reliable AQI measurements are available only from 2005, and limiting data to the fire season may not capture full air quality impacts.

5. **Health Data Coverage and Quality:** Limited time coverage of Fertility, Mortality, and Medicare Enrollee data restricts longitudinal analysis. Sampling errors, especially during events like the COVID-19 pandemic, could introduce biases.

6. **Correlation and Causation:** The study's reliance on correlation analysis does not establish causality, and the absence of topography and weather variables could reduce accuracy.

7. **Handling of Missing Values:** In the data cleaning process, missing values were addressed by simply removing these entries (NaNs), which could potentially skew the results and limit the depth of analysis. This approach does not account for the underlying reasons for missing data or explore alternative methods of handling such gaps, like imputation or analysis of patterns in missingness, which might have provided a more comprehensive view of the data set.

8. **Generalizability and Specificity:** Findings specific to Farmington, NM, may not be applicable to other regions. The study also does not address potential confounding variables such as socioeconomic factors and healthcare access.

9. **Model Predictions and Assumptions:** The use of predictive models like ARIMA carries inherent limitations, and assumptions used in statistical methodologies may not encapsulate real-world complexities.

10. **Secondary Data Source Dependency:** Reliance on secondary data sources like USGS and EPA limits control over data quality and completeness.

11. **Data Integration and Resolution:** Challenges in integrating data from various sources and the temporal resolution of datasets may not capture short-term fluctuations in wildfire occurrences and air quality.

12. **Public Health Data Limitations:** The availability and resolution of public health data may not fully represent the impacts of wildfire smoke on specific health conditions.

13. **Policy and Implementation Feasibility:** The study does not fully address the feasibility of policy implementation, considering economic, political, and social constraints.

14. **Environmental Change Considerations:** Long-term environmental and climatic changes that could alter wildfire patterns in the future are not accounted for in the study.

15. **Predictive Reliance on Historical Trends:** Projections heavily based on historical data may not account for future changes in climate patterns or public health interventions.

By acknowledging these limitations, the study underscores the need for cautious interpretation of its findings and highlights areas for future research enhancements. Despite these constraints, the research provides valuable insights into the public health implications of wildfire smoke exposure, offering a foundational understanding that can inform policy development and community health initiatives.

7. <u>Conclusion</u>

The study embarked on a journey to unravel the impact of wildfire smoke on the health of residents in Farmington, NM, driven by specific research questions that sought to understand the key health factors impacted by wildfire smoke, the immediacy of these impacts, and the projected future trends in health outcomes considering continued exposure. The findings from the comprehensive analyses – encompassing correlation, lag, and regression analyses – provide substantial insights into these queries.

The correlation analysis revealed a strong positive correlation between Smoke Estimates and Mortality Rate and a notable negative correlation with Fertility Rate. These associations suggest that increased smoke exposure from wildfires significantly correlates with higher mortality rates and reduced fertility rates among the Farmington population. The immediacy of these effects was further highlighted in the lag analysis, where the most substantial impacts were observed without delay, underscoring the urgency of addressing these health issues. Moreover, the regression analysis not only corroborated these findings but also projected the continuation of these trends into the future, painting a concerning picture for the long-term health landscape of Farmington if current conditions persist.

This study significantly contributes to the understanding of human-centered data science, primarily through its methodological approach and the relevance of its findings to the community of Farmington. By employing data science techniques that directly address the health concerns of a specific population, the research epitomizes the essence of human-centered data science, which prioritizes the well-being, needs, and contexts of people at the heart of its inquiry. The study also demonstrates how data-driven insights can be effectively translated into actionable knowledge that informs public health policies and interventions, emphasizing the practical application of data science in real-world scenarios.

In conclusion, the research provides compelling evidence of the adverse health effects of wildfire smoke exposure on a community, highlighting the need for immediate and long-term strategies to mitigate these impacts. These findings not only contribute to the broader discourse on environmental health but also underscore the critical role of human-centered data science in addressing pressing public health challenges. The study's outcomes urge policymakers, healthcare providers, and the general public to recognize the seriousness of wildfire smoke as a public health issue and to take collaborative, informed actions to safeguard the health and well-being of the Farmington community.

8. References

1. New Mexico Wildfires, The New York Times. https://www.nytimes.com/interactive/2022/06/01/climate/new-mexico-wildfires.html
2. New Mexico Wildfires, AP News. https://apnews.com/article/floods-wildfires-mountains-fires-new-mexico-1b4102ee2d1ab5c5a0af304df0cbf720
3. Exposure to Smoke from Fires. New York State Department of Health. https://www.health.ny.gov/environmental/outdoors/air/smoke_from_fire.htm
4. Reid, C. E., Brauer, M., Johnston, F. H., Jerrett, M., Balmes, J. R., & Elliott, C. T. (2016). Critical Review of Health Impacts of Wildfire Smoke Exposure. Environmental Health Perspectives, 124(9), 1334-1343. https://ehp.niehs.nih.gov/doi/10.1289/ehp.1409277
5. Kornfield, M. S., Rubin, E. S., Parker, P. B., Gibbins, K. J., Krieg, S. A., & Baldwin, M. K. (2022). Wildfire Smoke Exposure and Pregnancy Loss. https://doi.org/10.1016/j.fertnstert.2022.09.199
6. Kochi, I., Champ, P. A., Loomis, J. B., & Donovan, G. H. (2012). Valuing mortality impacts of smoke exposure from major southern California wildfires. Journal of Forest Economics, 18(1), 61-75. http://dx.doi.org/10.1016/j.jfe.2011.10.002
7. Environmental Protection Agency. (n.d.). Retrieved from https://www.epa.gov/
8. StatsModels. ARIMA Model Documentation. StatsModels Documentation. Accessed from https://www.statsmodels.org/stable/generated/statsmodels.tsa.arima.model.ARIMA.html
9. StatsModels. OLS Regression Model Documentation. StatsModels Documentation. Accessed from https://www.statsmodels.org/stable/regression.html

9. Data Sources

1. USGS, “Combined wildland fire datasets for the United States and certain territories, 1800s Present (combined wildland fire polygons)”, https://www.sciencebase.gov/catalog/item/61aa537dd34eb622f699df81
2. EPA, “Air Quality System (AQS) API”, https://aqs.epa.gov/aqsweb/documents/data_api.html
3. American Community Survey (ACS). "Fertility Data." U.S. Census Bureau. Accessed from https://data.census.gov/table/ACSST1Y2022.S1301

4. The Dartmouth Atlas Project. "Mortality and Medicare Enrollees Data." The Dartmouth Institute for Health Policy and Clinical Practice. Accessed from https://data.dartmouthatlas.org/mortality/#by-year