DATA 512: Final Project Report

Table of Contents

Introduction

In recent years, the western United States has been experiencing increasingly severe wildfires, resulting in widespread smoke that affects multiple states. While the causes of these wildfires are debated, including factors like climate change and forestry policies, their impact is undeniable. In this report, I will explore the consequences of smoke exposure on student learning outcomes in schools with the objective of providing valuable insights to policymakers, city administrators, and local authorities.

The motivation behind this analysis stems from the escalating frequency and severity of wildfires, a trend expected to intensify with ongoing climate warming. Existing literature has extensively explored the social and economic impact of air pollution. However, studies focusing on the impact on a particular state in the United States of America remains a gap which this study aims to address.

This report focuses on the city of Gillette and the state of Wyoming. According to the US Forest Service, since 1980, Wyoming has experienced a notable increase in both the number of fires each year and the total annual area burned. This alarming trajectory underscores the urgency for a detailed understanding of the potential future impacts of smoke on the community, making it a critical dimension of the analysis. In addition, Wyoming's ranking as the second state on the "disasters-are-getting-worse" list (Sherwood), further indicates the gravity of the situation. By comprehensively exploring the potential effects of wildfire smoke and worsening air quality, this analysis aims to equip local authorities and residents with the knowledge required to prepare for and mitigate the consequences effectively. This proactive approach is essential for fostering resilience in the face of escalating wildfire challenges and ensuring the long-term well-being of Gillette, Wyoming.

Background/Related Work

According to research, air pollution has a wide-ranging impact. Studies have explored links between air pollution exposure and neuroinflammation, risks for diseases like Alzheimer's and Parkinson's (Calderón-Garcidueñas, L. et al), as well as impacts on cognitive performance (Künn, S. et al) and student test performance (Marcotte, D. E). This report draws from the work of Wen J. et al which investigates the impact of wildfire smoke exposure on learning outcomes in the United States using standardized test scores from 2009 to 2016 across nearly 11,700 school districts. The research reveals that wildfire smoke exposure during the school year reduces test scores by approximately 0.15% of a standard deviation compared to a year with no smoke. The adverse effects are more pronounced among younger students and persist across various levels of economic disadvantage and racial/ethnic composition. The study estimates that smoke exposure in 2016 led to a reduction of nearly $1.7 billion in discounted future earnings, with approximately 80% of these costs affecting disadvantaged districts. These findings highlight an unaccounted-for social cost of wildfires, which is expected to escalate with climate warming.

This report aims to explore and identify comparable trends in social costs associated with wildfires in Gillette, Wyoming. The goal is to assess if the trends align with those documented in more extensive research and to understand the potential implications for the community in Gillette, particularly concerning the escalating effects of climate warming.

Therefore, this report investigates two pertinent research questions:

1. What are the historical smoke impacts on Gillette, Wyoming over the last 60 years?
2. What is the impact of air quality on student learning outcomes in Gillette, Wyoming?

My initial hypothesis was that smoke exposure has worsened in Gillette, Wyoming over the last 60 years and is associated with lower academic performance of students.

To investigate the impact of wildfire smoke exposure on test scores, Wen J. et al used the model below:

$$\begin{aligned}\text{Score}_{igy} &= \beta_1 \text{SmokePM}_{iy}^{\text{school}} + \beta_2 \text{SmokePM}_{iy}^{\text{non-school}} \\ &\quad + f(\mathbf{X}_{iy}) + \eta_i + \gamma_{yg} + \epsilon_{igy}\end{aligned}$$

The researchers also conducted secondary analyses to investigate whether the effects of smoke differed across different grade levels and a combination of economic disadvantage and race/ethnicity. While my approach did not replicate the regression model above, this research served as a guiding framework to inform and shape the methodology. This contributes to the credibility of this study and also ensures a systematic and informed exploration of the nuanced relationship between smoke exposure and student performance in the context of Gillette, Wyoming.

To conduct the analysis, 3 different data sources were used. The first contained a list of wildfires collected and aggregated by the US Geological Survey. The second contained information about the air quality index for Gillette, Wyoming provided by the United States Environmental Protection Agency (EPA). The third dataset contained educational data from the National Assessment of Educational Progress (NAEP) for the state of Wyoming. The NAEP is a comprehensive and nationally representative assessment that evaluates student performance in various subjects across the United States every two years. The NAEP provides standardized and comparable measures of student achievement, offering a robust foundation for evaluating the potential effects of wildfire smoke on learning outcomes.

The dataset for wildfires contained the following information:

Variable	Description
USGS_Assigned_ID	A unique ID assigned by the USGS creators of the dataset to the focal fire.
Fire_Year	The calendar year when the dataset creators determined the fire occurred.
GIS_Acres	The GIS calculated acres of the fire polygon calculated by using the Calculate Geometry tool in ArcGIS Pro.
Distance	Shortest distance in miles of the perimeter of the fire to Gillette.

The NAEP dataset contained the following information:

Variable	Description
Subject	Subject of the assessment – in our case, either Math or Reading
Grade	Grade of assessment – in our case, either Grades 4 or 8
Year	Year the assessment was conducted – from 1990 – 2022
Average Scale Score	Score on the assessment with scale range from 0-500
Jurisdiction	The jurisdiction from which data is collected – in our case, Wyoming

An additional dataset was considered for this analysis but eventually excluded. This was the performance of Campbell County where Gillette is located on Wyoming state learning assessments. This data focusing on Campbell, would have provided a targeted lens into the localized effects of smoke exposure, offering insights into the unique dynamics of the community. However, due to the limited temporal scope of available data spanning only from 2016 to 2022, it became difficult to assess the long-term trends and

patterns in smoke exposure effects on the community. While the decision to omit this data impacts this analysis, it underscores the importance of acknowledging data limitations in drawing conclusive insights about the localized impacts of smoke exposure in Campbell.

Methodology

The study initiated by compiling a list of wildfires occurring within 1250 miles of Gillette, Wyoming. Subsequently, air quality data from monitoring stations near Gillette was acquired, and an annual average was computed. While the typical annual fire season spans from May 1st to October 31st, this study used a wider timeframe of analysis. In alignment with the methodology employed in the Wen J. et al paper and considering that NAEP assessments typically occur between January and March, the study opted to utilize air quality index (AQI) data for the entire year rather than solely during the fire season. This approach ensured consistency with previous research practices.

Using the collected data, an estimate of wildfire smoke impact was calculated for each fire within 1250 miles of Gillette. This calculation factored in two key elements: distance of the fire from the city and area burned. The methodology assigned a higher weight (90%) to distance and a lower weight (10%) to area burned, based on the observed correlation with extracted Air Quality data. Subsequently, the annual wildfire smoke impact was determined by computing the average smoke impact across all fires that occurred within a given year. While recognizing the inherent limitations and approximations, this methodology provides a quantitative basis for assessing the annual impact of wildfire smoke.

In the next part of the analysis, a predictive model was developed for Gillette based on the fire data and smoke estimates. The objective of this model was to forecast smoke estimates for each year in the future, spanning from 2021 to 2049. Due to practical time constraints, an Autoregressive Integrated Moving Average (ARIMA) model was employed for this predictive task. This is because ARIMA models are well-suited for time series forecasting and offer a balance between accuracy and computational efficiency, making them a fitting choice within the project's temporal constraints.

In the final phase of the analysis, the investigation into the relationship between smoke impact, air quality, and test scores was conducted by calculating correlations among the variables, specifically math and reading scores. This examination aimed to discern patterns and associations between smoke impact estimates, air quality data, and academic performance indicators.

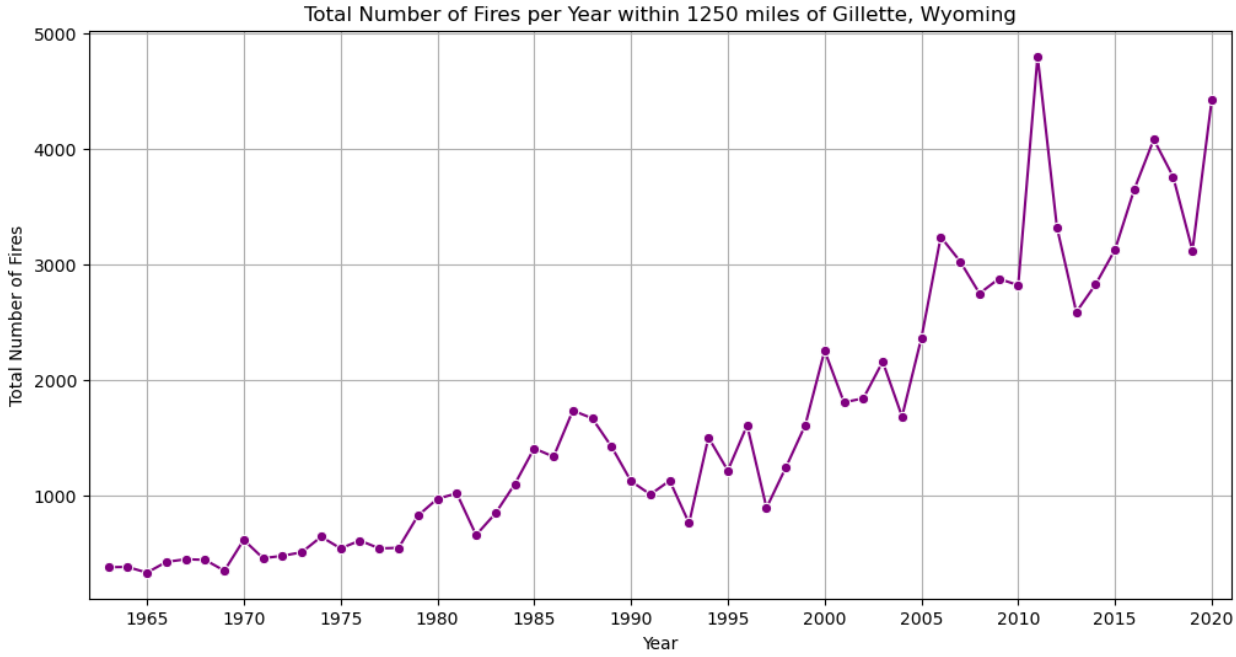
The methodology for this report accounted for human-centered considerations through various key practices. Firstly, all utilized data was sourced from publicly available datasets, and adherence to appropriate licenses ensured that data ownership and access rights were respected. This approach not only promotes transparency but also upholds ethical standards related to data usage. Secondly, when incorporating student learning outcomes into the analysis, a conscientious effort was made to prioritize privacy. No identifiable information of participants was utilized, thereby safeguarding their privacy, and complying with ethical standards in educational research.

In addition, transparency was a key aspect of the methodology. Throughout the various parts of the project, steps were taken to be transparent about the methods employed for data collection, analysis, and interpretation. The documentation of each step in the research process was made accessible to others via Github, fostering validation and understanding. This transparency contributes to the credibility of this analysis and allows for scrutiny and verification by the wider research community.

Findings

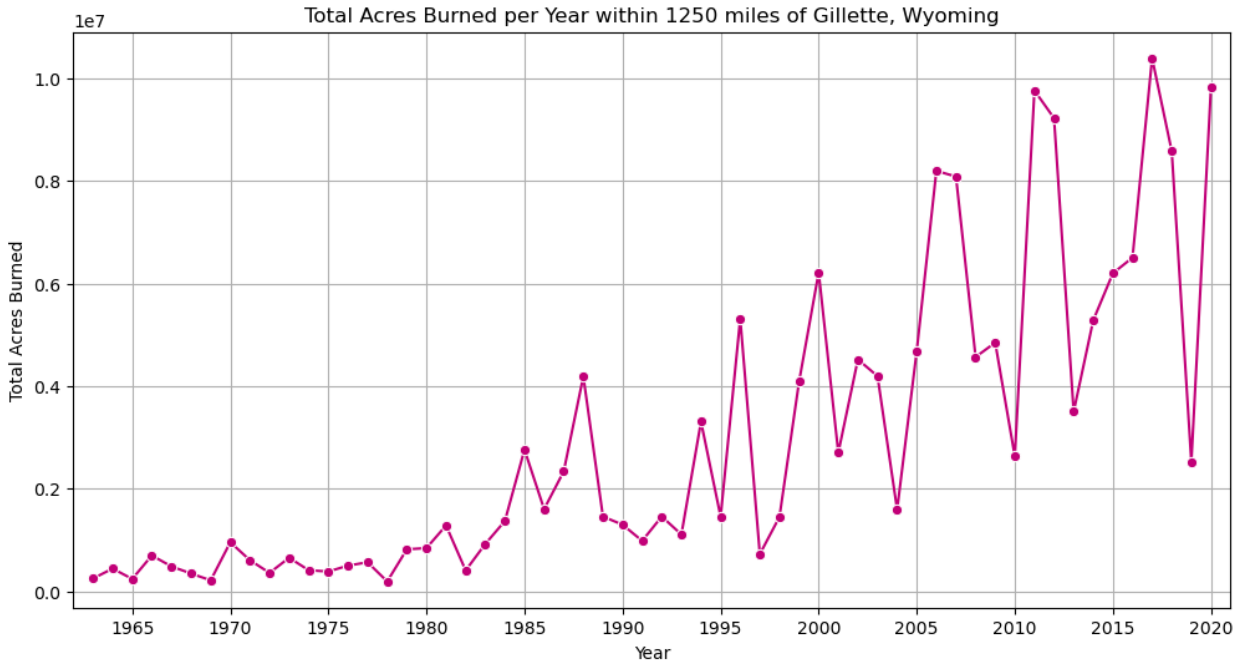
Consistent with findings from the US Forest Service, the data analysis for this report revealed a substantial increase in both the number of fires and the total annual area burned in Wyoming since 1980. Figure 1 below shows a rising trajectory in the number of fires occurring within 1250 miles from Gillette.

Figure 1: Number of Fires per Year within 1250 miles of Gillette, Wyoming



In addition to the upward trend in the number of fires, Figure 2 below provides further insight, indicating a noticeable increase in the total number of acres burned per year.

Figure 2: Total Acres Burned per Year within 1250 miles of Gillette, Wyoming

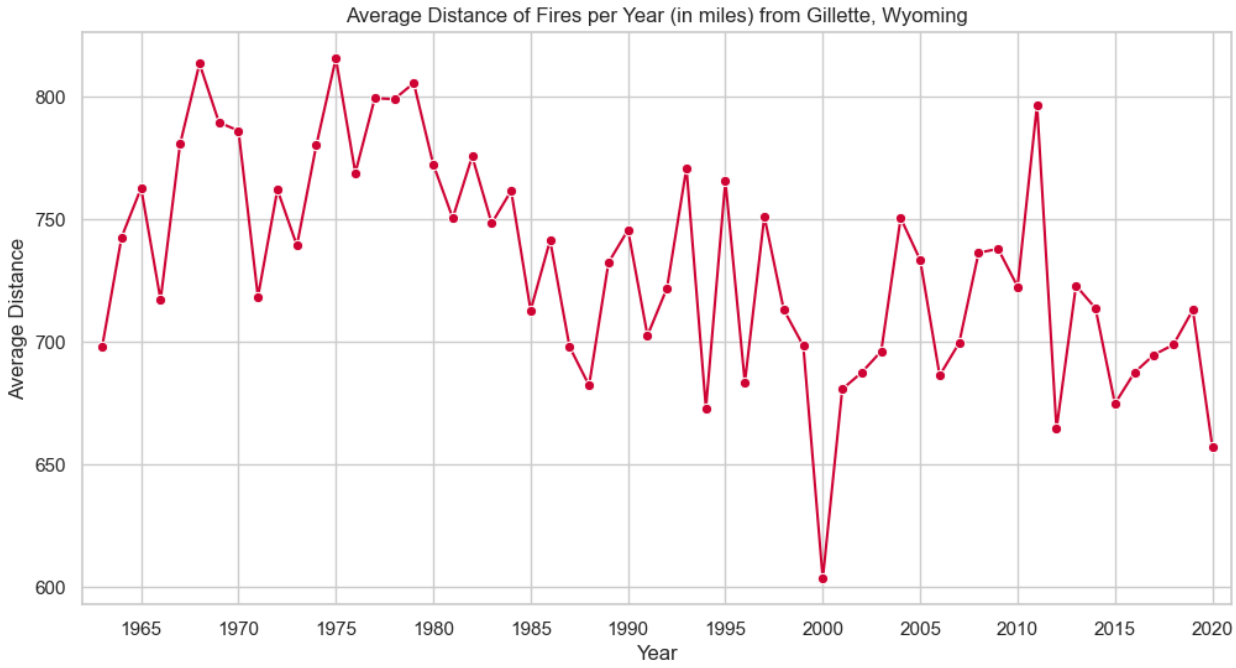


While the total acres burned displays a general upward trend over the analyzed years, the data exhibits distinct periods of increase and decrease in wildfire impact. During some years, such as 2011 and 2017, there are noticeable spikes in the total acres burned, signifying periods of elevated wildfire activity. Conversely, there are years, like 2010 and 2019, when the total acres burned decrease, indicating a reduction in wildfire impact.

This variability strongly suggests that Gillette, Wyoming, is not consistently prone to being near large fires. Instead, the region experiences intermittent periods of heightened wildfire risk, with some years witnessing more significant wildfire events while others enjoy relative respite.

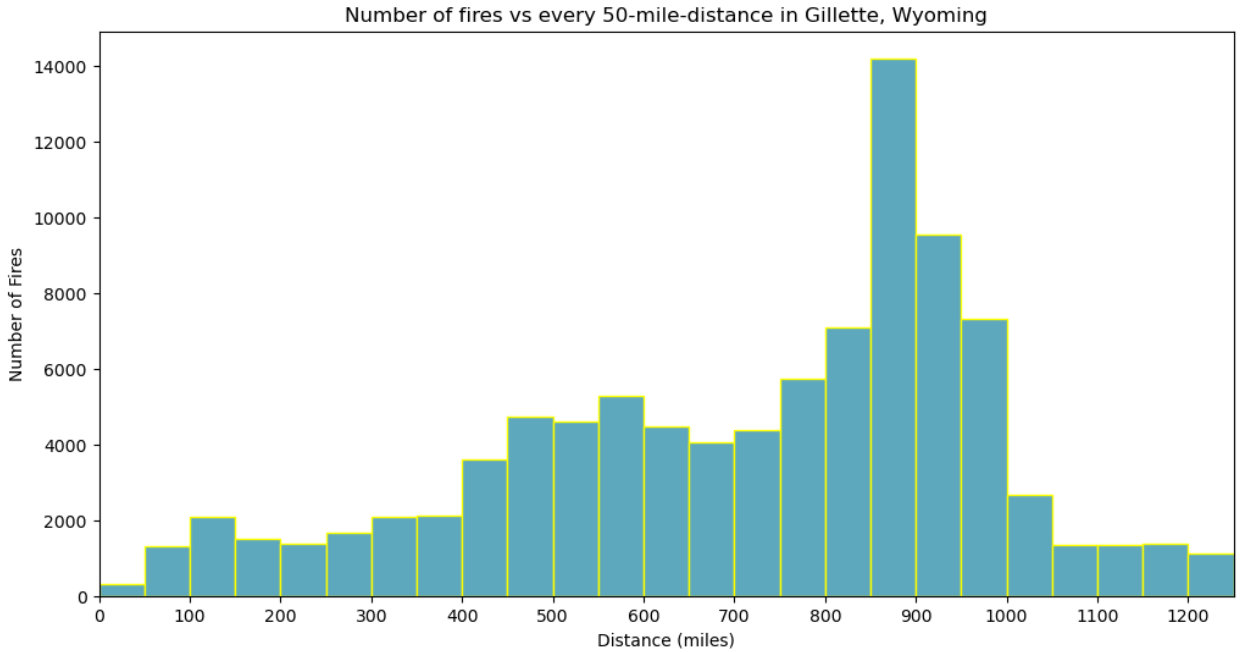
In addition to the trends highlighted by the US Forest Service, a noteworthy observation emerged from Figure 3 below, indicating a decrease in the average distance of fires per year from the city of Gillette. This distinctive pattern suggests a spatial shift in the proximity of wildfires to the city over time. This insight underscores a potential geographical impact on the community, raising concerns about the closer proximity of wildfires and the associated implications for air quality, public health, and overall safety.

Figure 3: Average Distance of Fires per Year within 1250 miles of Gillette, Wyoming



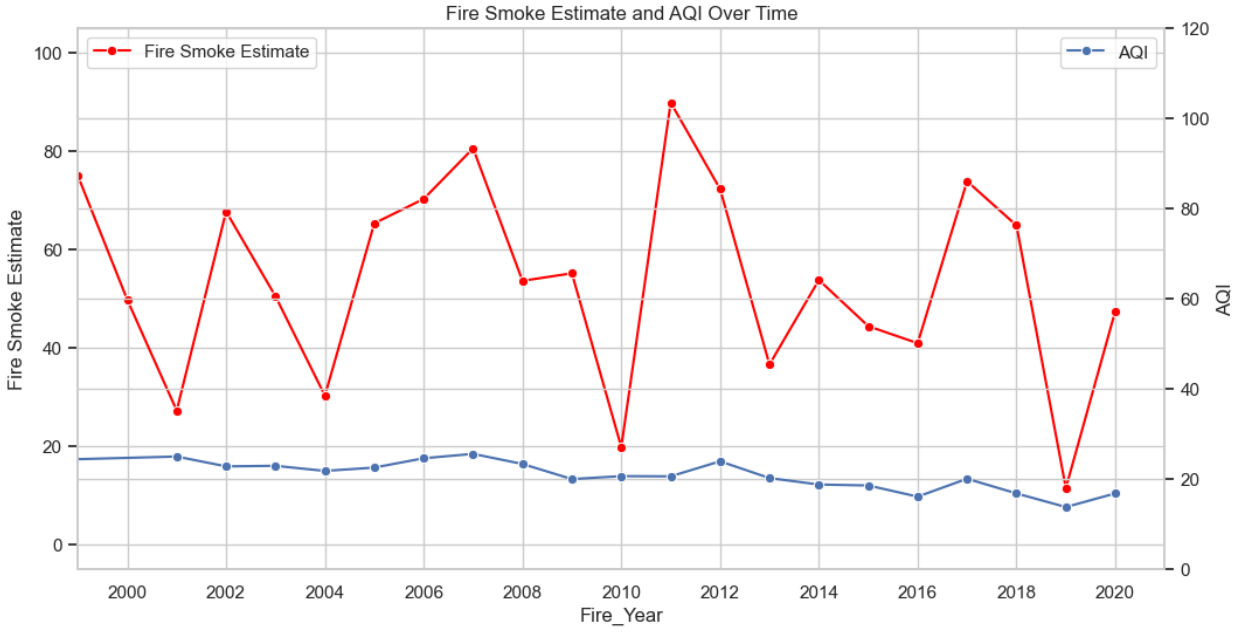
To investigate this further, the study looked at the number of fires occurring at every 50-mile-distance from Gillette, Wyoming up to a maximum specified distance of 1250 miles (Figure 4 below). From the histogram, it is evident that the most significant concentration of fires, numbering approximately 14,000 incidents, occurred in regions situated around 850 to 900 miles away from Gillette, Wyoming. In contrast, areas within a 100-mile radius of the city have experienced fewer than 2,000 fires in the last 60 years. After 1,000 miles, the number of fires declines again. This indicated a possibility that certain fire-prone regions of the country are located about 800-1000 miles from Gillette which contributes to the spike we see on the histogram.

Figure 4: Number of fires vs every 50-mile-distance in Gillette, Wyoming



Following the observation of escalating wildfire trends and changing spatial dynamics, this study turned its focus to comprehending the impact of these wildfires on air quality. The study employed the Fire Smoke Estimate, derived from both distance and area burned, as a metric to assess the impact of wildfires. Figure 5 below shows the relationship between the Fire Smoke Estimate and AQI over time.

Figure 5: Fire Smoke Estimate and AQI Over Time

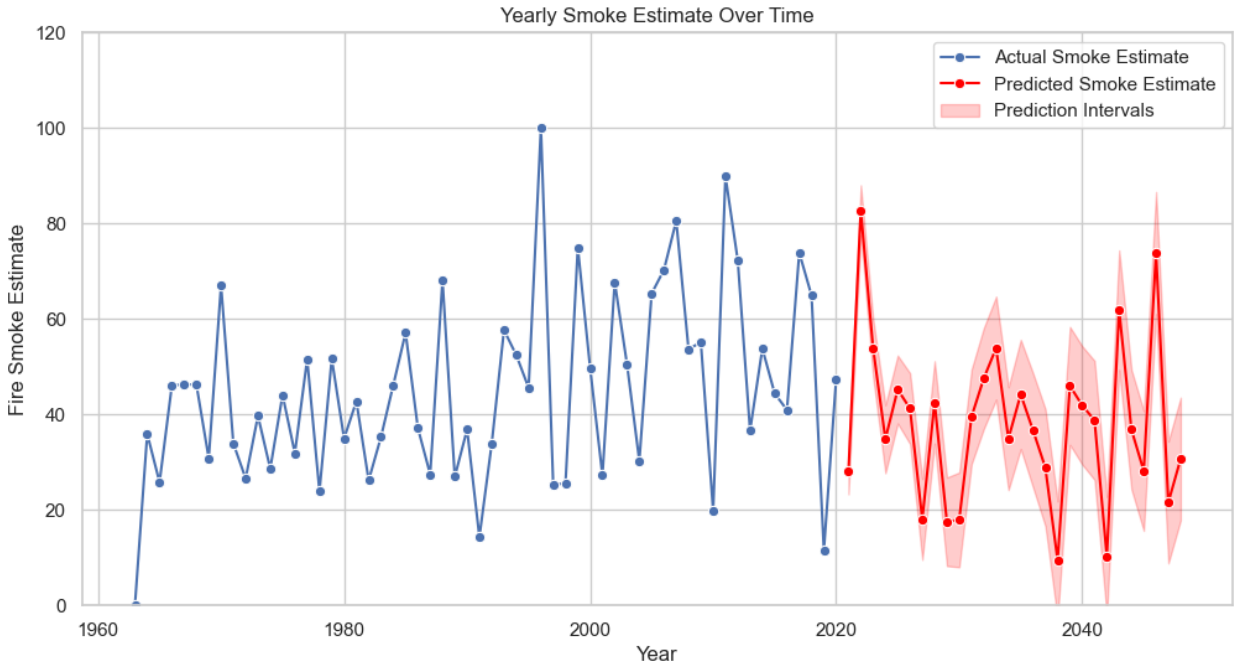


As is evident, the Fire Smoke Estimate showed fluctuations from year to year, indicating variability in the extent of smoke impact on the city of Gillette. In contrast, the Air Quality Index (AQI) for the city

demonstrated a relatively stable trend over the years, with a subtle decline noticeable from around 2012 onwards. A correlation of 0.37 between the Fire Smoke Estimate and AQI suggests a discernible but moderate relationship. It is important to acknowledge that this correlation accounts for limited information as crucial factors influencing smoke impact, such as wind direction, fire intensity, and duration, were not available for analysis.

Recognizing the importance of anticipating future trends to inform recommendations, this study extended its analysis to predict smoke impact for upcoming years, as depicted in Figure 6.

Figure 6: Yearly Smoke Estimate Over Time

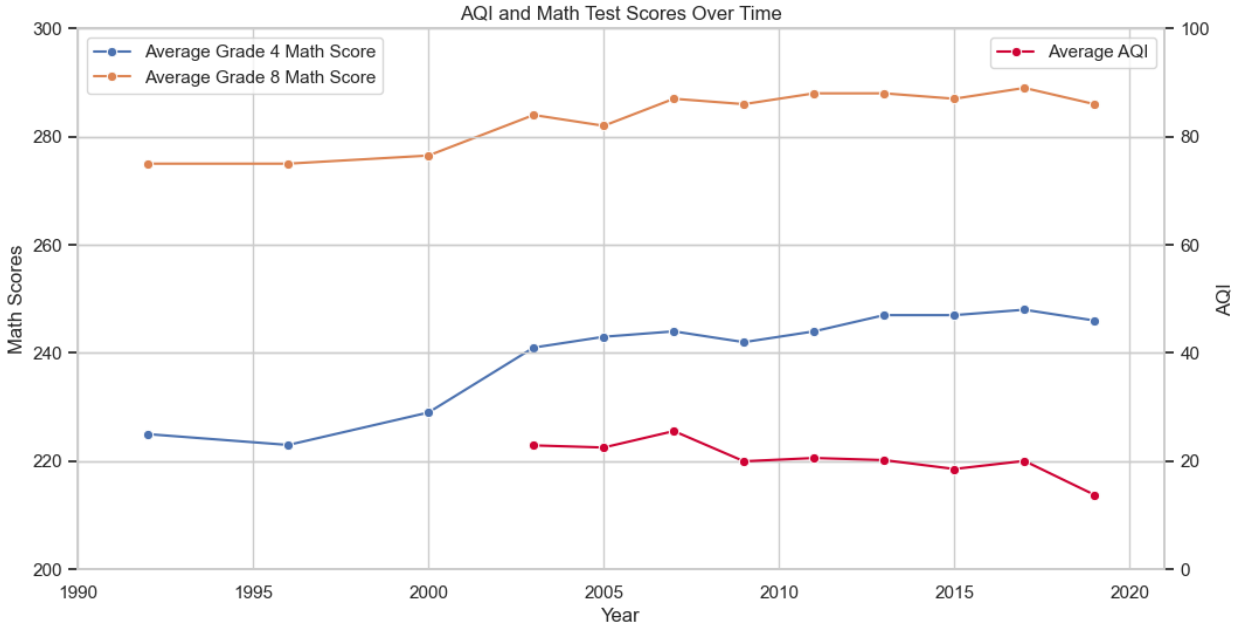


While acknowledging that the predictive methodology may not be as robust as desired, the forecasted smoke estimates reveal a noteworthy trend. The projections indicate a maintenance of current smoke levels and, in some instances, an increase in impact. This insight underscores the imperative for proactive planning and preparedness.

To address the research question regarding the estimated impact of air quality on student learning outcomes in Gillette, Wyoming, the study investigated the relationship between smoke impact, air quality (AQI), and test scores. Specifically, correlations were calculated for both math and reading scores. In the context of math scores, the correlation with the smoke estimate was found to be less than 0.2, suggesting a weak association between smoke impact and math performance.

Figure 7 below shows the relationship between the AQI and Math test scores for Grade 4 and 8. When considering the Air Quality Index (AQI), Grade 4 exhibited a correlation of -0.47, indicating a moderate negative association, while Grade 8 showed a correlation of -0.2, denoting a relatively weaker negative relationship. These correlations suggest that as AQI worsens, there is a negative association with math scores, with Grade 4 being more strongly affected than Grade 8. This is consistent with the results from Wen J. et al that showed that adverse effects on learning outcomes from smoke exposure are more pronounced among younger students.

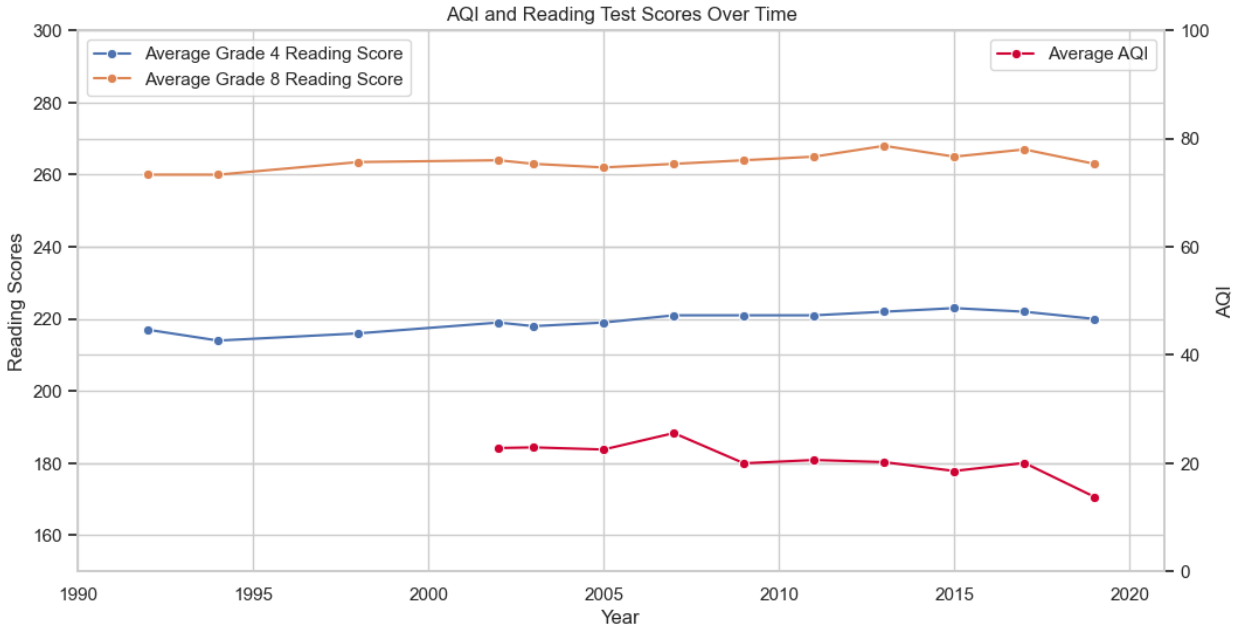
Figure 7: AQI and Math Test Scores Over Time



For reading scores, surprisingly, the correlation with the smoke estimate was found to be positive. This unexpected positive correlation suggests a counterintuitive relationship between smoke impact and reading performance, indicating that, in this specific context, an increase in smoke impact is associated with higher reading scores. It is crucial to interpret such findings cautiously and consider potential confounding factors or nuances in the data that may contribute to this unexpected result. Most importantly, it is imperative to acknowledge that the methodology employed to estimate smoke impact in this study comes with inherent limitations. These limitations underscore the need for further scrutiny and refinement to enhance its robustness.

Figure 8 below shows the relationship between the AQI and reading test scores for Grade 4 and 8. The correlation analysis reveals that Grade 4 exhibits a correlation of -0.29, indicating a moderate negative relationship between AQI and reading scores for this grade level. In comparison, Grade 8 shows a correlation of -0.19, denoting a relatively weaker negative association. These correlations suggest that as the AQI worsens, there is a negative association with reading scores, with Grade 4 being more strongly affected than Grade 8 which is also consistent with Wen J. et al's study.

Figure 8: AQI and Reading Test Scores Over Time



Discussion/Implications

The findings of this study hold significance for the city council, city manager/mayor, and residents of Gillette, Wyoming, as they shed light on the relationship between wildfire smoke impact, air quality, and student learning outcomes. The identification of worsening trends in the number of fires, total acres burned, and the decreasing average distance of fires from the city emphasizes the urgency of proactive planning and mitigation efforts. The predictions of future smoke impact underscore the need for timely and effective strategies to address the potential escalation of air quality issues, which could have repercussions on public health and academic performance.

Given the observed negative associations between air quality and reading/math scores and the extensive body of research on the negative impacts of air pollution, the city council and stakeholders should consider implementing measures to mitigate the impact of wildfire smoke on air quality. This could involve enhancing monitoring systems, improving air quality in schools, having clean-air public spaces, and developing awareness campaigns. Moreover, collaborative efforts with environmental and educational experts could contribute to comprehensive strategies. Despite the moderate to weak correlations within the specific context of this study, the broader evidence base suggests that addressing air quality concerns remains crucial. Considering the escalating trends of wildfires observed, the city council should consider this issue within the broader context of public health and well-established connections between environmental factors and overall community well-being. Therefore, they should prioritize the development of a concrete action plan in the next 5 years.

Informed by human-centered data science principles, this project prioritized transparency, privacy, and data quality. Utilizing publicly available and open datasets with appropriate licenses ensured ethical data usage, respecting privacy by avoiding identifiable information in educational outcomes analysis. Transparency was upheld by documenting each step in the research process, making it accessible for

validation. Additionally, the recognition of data limitations and the call for a more robust methodology for smoke impact estimation reflect a commitment to data quality.

Most importantly, the approach taken in the analysis, which prioritized simple visualizations and straightforward correlation analyses, aligns with human-centered principles by recognizing the diverse audience of council members and residents who may not have specialized expertise in data science. The emphasis was on producing easily understood representations of data, such as the increase in the number of wildfires per year or acres burned, that would be accessible to a broad audience and empower them to make informed decisions. By prioritizing clarity and accessibility, the analysis not only serves the purpose of uncovering trends but also fulfills the human-centered objective of effectively communicating complex data to a lay audience.

Limitations

This study is subject to several limitations that should be acknowledged to ensure a nuanced interpretation of the results. Firstly, the smoke impact estimate created in this project was quite simple. Replicating the approach taken by Wen J et al., required a more intricate and time-consuming regression model which was not feasible in the time available. Consequently, the smoke impact estimation in this analysis was more straightforward, potentially limiting the depth of its correlation with student learning outcomes.

In addition, the limited timeframe of the available data on student learning outcomes, spanning from 1990 to 2022 for NAEP and an even shorter duration for Wyoming Department of Education data, posed a constraint on the depth and historical breadth of the analysis. Similarly, AQI data was only available from 2000 onwards. This limitation impacts the ability to draw long-term conclusions about the relationship between wildfire smoke exposure and student performance.

Moreover, it's important to acknowledge that student learning outcomes are multifaceted, influenced by several factors beyond the scope of this analysis. Variables such as teaching quality, school funding, and other socio-economic determinants contribute significantly to educational outcomes. Recognizing these factors is imperative for a nuanced interpretation of the findings and for avoiding overgeneralization.

Lastly, due to the limited availability of specific data for the city of Gillette, it is important to acknowledge that the conclusions drawn from this analysis are inherently more applicable to a broader geographic scale. However, despite this limitation, the analysis is valuable as it provides insights into the relationship between air quality and student learning outcomes at a regional level. While the specificity to Gillette may be constrained, this broader perspective still offers relevant information for policymakers, educational authorities, and residents across the region.

Conclusion

This study aimed to investigate the estimated smoke impacts in Gillette, Wyoming, over the last six decades and assess the potential impact of wildfire smoke on student learning outcomes. The analysis revealed an increasing trend in the number of wildfires and acres burned within 1250 miles of Gillette, aligning with broader trends reported by the US Forest Service. Correlations between AQI and math and reading test scores, while moderate to weak, were found to be negative indicating that lower air quality is associated with lower test scores. Considering the extensive body of evidence about the wide ranging impact of worsening air quality, this report recommended the city council members and residents of Gillette, Wyoming to take a proactive approach in addressing air quality concerns.

This study contributes to the understanding of human-centered data science by exemplifying the application of data-driven insights to address real-world challenges. Through a combination of data exploration, visualizations, correlation analyses, and predictive modeling, this report translated complex environmental and educational data into accessible information for city council members and residents. The emphasis on simplicity in visualizations and straightforward correlation analyses reflected a human-centered approach, recognizing the diverse audience, including non-experts in data science. While the study identifies limitations, it underscores the importance of considering available evidence and acting proactively, aligning with human-centered principles in data-driven decision-making for community well-being.

References

Calderón-Garcidueñas, L. et al. Long-term air pollution exposure is associated with neuroinflammation, an altered innate immune response, disruption of the blood-brain barrier, ultrafine particulate deposition, and accumulation of amyloid β-42 and α-synuclein in children and young adults. Toxicol. Pathol. 36, 289–310 (2008).

Innes, Robin J. 2019. Fire regimes of Wyoming big sagebrush and basin big sagebrush communities. In: Fire Effects Information System, [Online]. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Missoula Fire Sciences Laboratory (Producer). Available: https://www.fs.usda.gov/database/feis/fire_regimes/WY_basin_big_sagebrush/all.html

Künn, S., Palacios Temprano, J. F. & Pestel, N. Indoor Air Quality and Cognitive Performance Discussion Paper No. 12632 (IZA, 2019).

Marcotte, D. E. Something in the air? Air quality and children's educational outcomes. Econ. Educ. Rev. 56, 141–151 (2017).

Sherwood, Nicole. *Wyoming Ranks High for Worsening Natural Disasters*. 24 August 2023. Available: https://laramielive.com/wyoming-ranks-high-for-worsening-natural-disasters/

Wen, J., Burke, M. Lower test scores from wildfire smoke exposure. Nat Sustain 5, 947–955 (2022). https://doi.org/10.1038/s41893-022-00956-y

Data Sources

A list of links to the relevant data sources that you used.

- Wildfire dataset This dataset was collected and aggregated by the US Geological Survey.
- FIPS The Federal Information Processing Standard Publication (FIPS) is a five-digit Federal Information Processing Standards code which uniquely identified counties and county equivalents in the United States, certain U.S. possessions, and certain freely associated states.
- NAEP Performance of Wyoming on the National Assessment of Educational Progress (NAEP) which is a comprehensive and nationally representative assessment that evaluates student performance in various subjects across the United States every two years.