DATA 512: Final Project Report

Mehjabeen Zameer

University of Washington

11th December 2023

Table of Contents

[Introduction 1](#_Toc153215457)

[Background/Related Work 2](#_Toc153215458)

[Methodology 3](#_Toc153215459)

[References 8](#_Toc153215460)

[Data Sources 9](#_Toc153215461)

# 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.

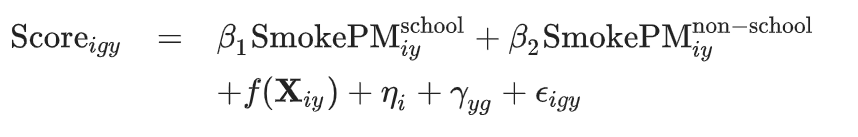
This report aims to explore and identify comparable trends in social costs associated with wildfires in Gillette, Wyoming. The goal is to assess if the trends align with those documented in more extensive research and to understand the potential implications for the community in Gillette, particularly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



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 this model was the inspiration for this report, due to shortage of time, it was not possible to replicate this analysis. As a result, for this study, a correlation analysis was conducted to assess trends in data for Wyoming.

To conduct the analysis, 3 different datasets 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.

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

Not just your analytical methods, but also, why you chose them, and how human-centered considerations such as ethics informed the way you designed your study.

Findings

What did you find? Use words and figures, don’t just point to code.

A graph with a line graph

Description automatically generated

A graph of a number of tires

Description automatically generated

A graph with lines and numbers

Description automatically generated

A graph showing the average distance of a company

Description automatically generated with medium confidence

A graph showing the growth of the number of people

Description automatically generated with medium confidence

A graph of a number of people

Description automatically generated with medium confidence

A graph with red lines and blue lines

Description automatically generated

A graph showing the average smoke

Description automatically generated with medium confidence

Discussion/Implications

Why are your findings important or interesting; What should the city council, city manager/mayor, and city residents do to address your findings? How long do they have to make a concrete plan?

This section should include a thoughtful reflection that describes the specific ways that human centered data science principles informed your decision-making in this project.

Limitations

This is a required section for your report. There are often many, many limitations for any study. If you honestly tried to list them all, this might end up being the longest section. You should prioritize and list the ones that are most likely to have a significant impact on your results. Specific license issues could be a limitation, depending on what data sources you used. Flaws in the data, data cleaning techniques, potential assumptions and/or how you handled missing values could be a limitation. Statistical techniques often have specific assumptions and preconditions; if you’re not certain all of the data meets those requirements - this is a good place to make that clear.

There are several factors that affect the ability of this analysis to answer the supplementary research question. Firstly, 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, poses a constraint on the depth and historical breadth of the analysis. This limitation may impact 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.

Additionally, the smoke impact estimate created in Part 1 of the project was quite simple. Replicating the approach taken by Wen J et al., will require a more intricate and time-consuming regression model which is not feasible in the time available. Therefore, this analysis may need to rely on a more straightforward methodology, potentially limiting the depth of the smoke impact estimation and its correlation with student learning outcomes. These considerations underscore the importance of managing expectations and ensuring the analysis remains realistic and feasible within the given constraints.

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 may inherently be more applicable to a broader geographic scale, such as Campbell County or the state of Wyoming. However, despite this limitation, the analysis remains valuable for several reasons. The findings can still provide insights into the potential relationships between wildfire smoke exposure and student learning outcomes at a regional level. This broader perspective still offers relevant information for policymakers, educational authorities, and residents across the region. While the specificity to Gillette may be constrained, the analysis can still contribute meaningful insights that inform decision-making and preparedness strategies for areas facing similar wildfire-related challenges in the broader geographic vicinity.

Conclusion

Restate your research questions/hypotheses and summarize your findings. Explain to the reader how this study informs their understanding of human centered data science.

# 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](https://www.sciencebase.gov/catalog/item/61aa537dd34eb622f699df81) This dataset was collected and aggregated by the US Geological Survey.
* [FIPS](https://www.census.gov/library/reference/code-lists/ansi.html) 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](https://www.nationsreportcard.gov/profiles/stateprofile/overview/WY?cti=PgTab_OT&chort=1&sub=MAT&sj=WY&fs=Grade&st=MN&year=2022R3&sg=Gender%3A%20Male%20vs.%20Female&sgv=Difference&ts=Single%20Year&tss=2022R3&sfj=NP) Performance of Wyoming on the National Assessment of Educational Progress (NAEP) which is a is a comprehensive and nationally representative assessment that evaluates student performance in various subjects across the United States every two years.