SMOKE AND MIRRORS: Will Wildfire Smoke Make Pueblo Cough Up a Lung?

Daniel Vogler

Introduction

In recent years, wildfires have become an increasingly urgent issue in the western United States, with important implications for public health, environmental sustainability, and public policy. This analysis focuses on Pueblo, Colorado, a midsize city located in a region that has seen an uptick in wildfires and their attendant public health risks, particularly stemming from smoke pollution. Wildfire smoke carries fine particulate matter and other harmful pollutants that pose significant health risks, especially for people with respiratory conditions like asthma.

This project aims to quantify the intensity of wildfire smoke in Pueblo by developing a novel smoke intensity estimate (SIE) metric. By comparing this metric to the widely used Air Quality Index (AQI), the study provides insights into how closely wildfire smoke exposure is associated with an established notion of air quality. Understanding this association is important because wildfire smoke is just one factor that affects air quality; quantitative analysis can help us understand the magnitude and directness of this influence and sense-check the output of the SIE computation. Additionally, using SARIMA time series modeling, the project forecasts the SIE for the next 25 years.

Finally, this study explores the relationship between wildfire smoke and asthma hospitalization rates in Pueblo. In doing so, it addresses a research question at the intersection of environmental science and public health that researchers have studied in other geographies, but not in Pueblo. By evaluating the public health implications of wildfire smoke openly and reproducibly, this analysis contributes to human-centered data science efforts. The findings could inform targeted interventions by local governments, enabling more effective health responses during wildfire events and better safeguarding at-risk populations.

This report begins by reviewing related work from other regions, which produced conclusions inspiring the hypotheses tested in this study. It describes how the smoke intensity estimate is computed and the reasoning behind the metric. Then, it summarizes the results of statistical analyses and data visualizations aimed at understanding the relationship between the smoke intensity index and 1) air quality as well as 2) asthma hospitalization rates in Pueblo County.

I find a moderate, positive association between smoke intensity and air quality. This suggests, – but does not definitely establish – that the smoke intensity estimate is reasonable, since it has the

kind of moderate positive relationship with an established notion of air quality that we'd expect from a sensible metric for smoke intensity. Furthermore, I find that yearly data *do not* support the hypothesis that there is a positive relationship between estimated smoke intensity and asthma hospitalizations. I conclude by discussing possible explanations for this surprising result, acknowledging the most important limitations in the dataset and methodologies used in this study.

Background

The Centers for Disease Control notes that "exposure to wildfire smoke has been shown to lead to several negative health outcomes and is a known asthma trigger" (CDC 2024). During a recent "wildfire smoke event" in New York state, researchers affiliated with the agency found that "asthma-associated [emergency department] visits increased over 80% on the day with the highest exposure to wildfire smoke" (Meek et al, 2023). Importantly, this study does not establish this link with a metric of smoke intensity; instead, it uses the concentration of particulate matter in the air as a proxy (ibid.).

Given these results, and the fact that wildfires affect communities across the United States, it is important for policymakers to understand whether they hold more broadly, equipping them with information to guide policies that can support vulnerable population during wildfire smoke events. The population of Pueblo specifically is likelier to be low-income or disabled than the average American, meaning it has slightly more people in groups the CDC identifies as "populations of concern" for climate-related health impacts.

The results of the CDC studies cited above raise at least two general research questions related to the relationship between wildfire smoke intensity and asthma risk:

1. Relatedly, how closely are smoke intensity and air quality indices (which are strongly influenced by particulate matter concentrations) associated?
2. Does the same relationship between smoke intensity and asthma hold in Pueblo if the dependent variable is a direct measure of smoke intensity, rather than the concentration of particulate matter?

To investigate these qualitative research questions, I formulated more specific empirical hypotheses based on the data that is available. Specifically, I hypothesized that:

1. Yearly levels of the SIE and AQI indices will be positively, linearly associated
2. Yearly levels of the SIE and the rate of asthma hospitalizations in Pueblo County will be positively, linearly associated

I model these relationships using ordinary least squares (OLS) linear regression and techniques from time series analysis, including seasonal autoregressive integrated moving average (SARIMA) models and de-trending. I chose the simple, accessible method of linear regression as the foundational tool for testing these relationships, while acknowledging that the traditional assumptions of OLS do not always hold perfectly, to ensure that my results are as easy as possible for policymakers to interpret, understand, and act upon.

Finally, I investigated the possibility of including a wider set of public health outcomes beyond asthma hospitalizations in this analysis, given that asthma is not the only condition that can be triggered by wildfire smoke. After reviewing the complete set of public health variables published by the Colorado Environmental Public Health Tracking service, I concluded that only the asthma data was sufficiently frequent, high-quality, and relevant to merit inclusion in this analysis.

Other datasets used in this analysis are the "Combined wildland fire datasets for the United States and certain territories, 1800s-Present," provided by the US Geographic Survey, which contained the acreage of each wildfire and enabled computation of the distances between the fires and Pueblo; and the US Environmental Protection Agency (EPA) Air Quality Service (AQS) API, from which I accessed the variables used to compute the Air Quality Index.

Methodology

The analytical methods and data selection practices in this study were chosen to ensure both analytical rigor and to uphold important principles of human-centered data science like prioritizing the public good, communicating accessibly, conducting research transparently, and showing respect for individual privacy. In the interest of maximum transparency and adherence to the hallowed principles and traditions of human-centered data science, the notebooks documenting each step of data gathering and analysis are published in a GitHub repository.

From a technical perspective, I constructed my Smoke Intensity Estimate (SIE) by applying an inverse-square law to combine the acreage burned and the distance of each fire to Pueblo, CO into a single, tractable metric. This method accounts for the fact that smoke intensity decreases with distance. It assumes that the impact of wildfire smoke declines proportionally to the square of the distance between Pueblo and the fire's location and increases proportionally to the acreage burned. By combining these two contributors to smoke intensity, I generated a custom metric capturing wildfire smoke exposure in Pueblo.

Subsequently, I conducted a SARIMA time series analysis to explore trends and seasonal variations in the SIE, making it possible to construct a rough prediction of its trajectory over the next 25 years.

Finally, to examine the public health implications, I used ordinary least squares (OLS) linear regression to investigate the relationship between the SIE and asthma hospitalizations in Pueblo. These methods were chosen for their ability to model temporal patterns and test my hypotheses about associations between environmental and health data, while remaining interpretable for policymakers.

The following section documents my findings, namely that neither of these two hypotheses were supported by data *at the yearly level*. The **Discussion & Limitations** section highlights why it may be problematic that I am forced to use yearly data to test the qualitative hypotheses that SIE, AQI, and asthma hospitalizations are positively associated.

However, technical considerations are just one part of designing this study. I also made every effort to ground this study in principles of human centered data science. Fundamentally, I chose to study a question that has obvious and actionable implications for public health and human welfare. I chose aggregated, anonymized datasets to preserve individuals' privacy. Beyond that, I prioritized transparency and reproducibility in developing the SIE and exploring its relationship to other variables. My original datasets (or procedures for extracting them), inverse-square law approach, and all related calculations and analysis are clearly documented and openly published, enabling external validation. Additionally, I took care to ensure that the metrics and models did not inadvertently stigmatize vulnerable populations by focusing solely on environmental drivers of health outcomes rather than incorporating any data about identity or socioeconomic status in my study[1].

After generating the SIE, I continued my attempts to align the analytical work in this project to human-centered design values continued as I examined the connection with asthma hospitalizations. By connecting wildfire smoke metrics (SIE) to asthma hospitalizations, I aimed to prioritize community health concerns and ensure that my findings contained insights that would be *actionable* for local governments. Furthermore, I chose interpretable statistical methods in order to communicate effectively with stakeholders outside the academic community. At all stages, I tried to

---

[1] This is not to say that public health studies should not consider socioeconomic status and identity-related traits, since these are generally known to influence public health outcomes. But since thoughtfully incorporating these data requires background knowledge about community context that I do not possess, considering them in this particular study could be irresponsible.

select the simplest model with sufficient analytical power to test my hypotheses; this was an explicit choice aimed at making this work as broadly accessible as possible.

Findings

This section reviews the most important findings in this analysis related to three major phases of work in this study:

1. Trends over time in the Smoke Intensity Estimate
2. The relationship between the SIE and the AQI
3. The relationship between the SIE and asthma hospitalizations

Smoke Intensity over Time

My estimate of smoke intensity is a function of two variables: the acreage burned by a fire and the distance from its center to Pueblo, CO. I will begin by pointing out important trends in the constituent variables individually, then cover trends in the SIE which combines them.

Most documented wildfires occur more than 650 miles from Pueblo, which served as the cutoff for inclusion in the smoke intensity estimate.[2] Up to this cutoff, the frequency of wildfires increases linearly with distance from Pueblo – wildfires very close to Pueblo are rare **(Figure 1)**.
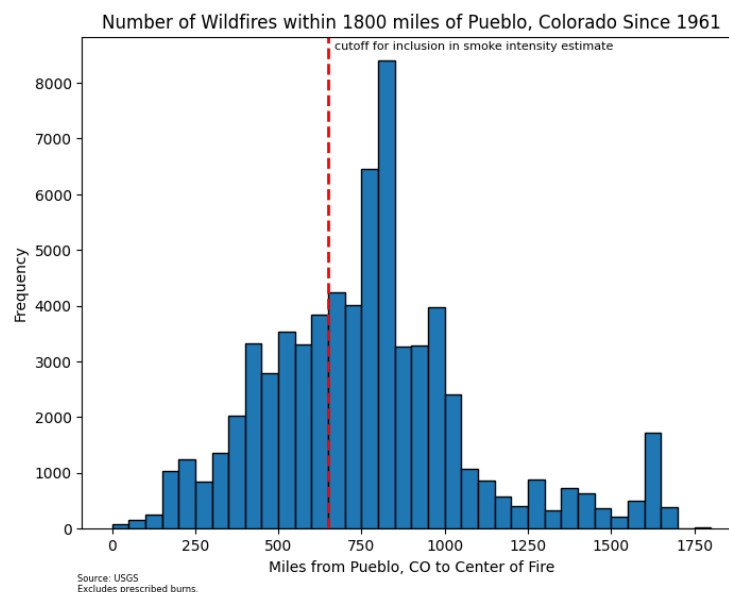

Figure 1

---

[2] This is because most wildfires in the US occur in California; most locations in California are more than 650 miles from Pueblo.

Moreover, the acreage burned by fires near Pueblo each year exhibits two important trends: 1) the average level of acreage burned increases noticeably after 2000 and 2) year-to-year variability is very high. This is true whether we define 'near Pueblo' extremely liberally – within 1800 miles, covering most of the country **(Figure 2)** – or quite conservatively – considering only a 100mile radius, roughly the distance from Pueblo to Denver **(Figure 3)**.

Considering acreage burned around Pueblo by drawing a 100-mile radius around the city helps filter fires most likely to affect the city, but it is still a crude method. This treats every fire within 100 miles of Pueblo equally, whether it is 99 miles or 0.9 miles away. Clearly, closer fires would influence the city more. Defining the smoke intensity metric using an inverse-square law attempts to smoothly adjust for the relationship between smoke intensity and distance. Therefore, SIE is calculated according to the following simple formula, where *A* is the acreage burned, and *d* is the distance from the fire to Pueblo:
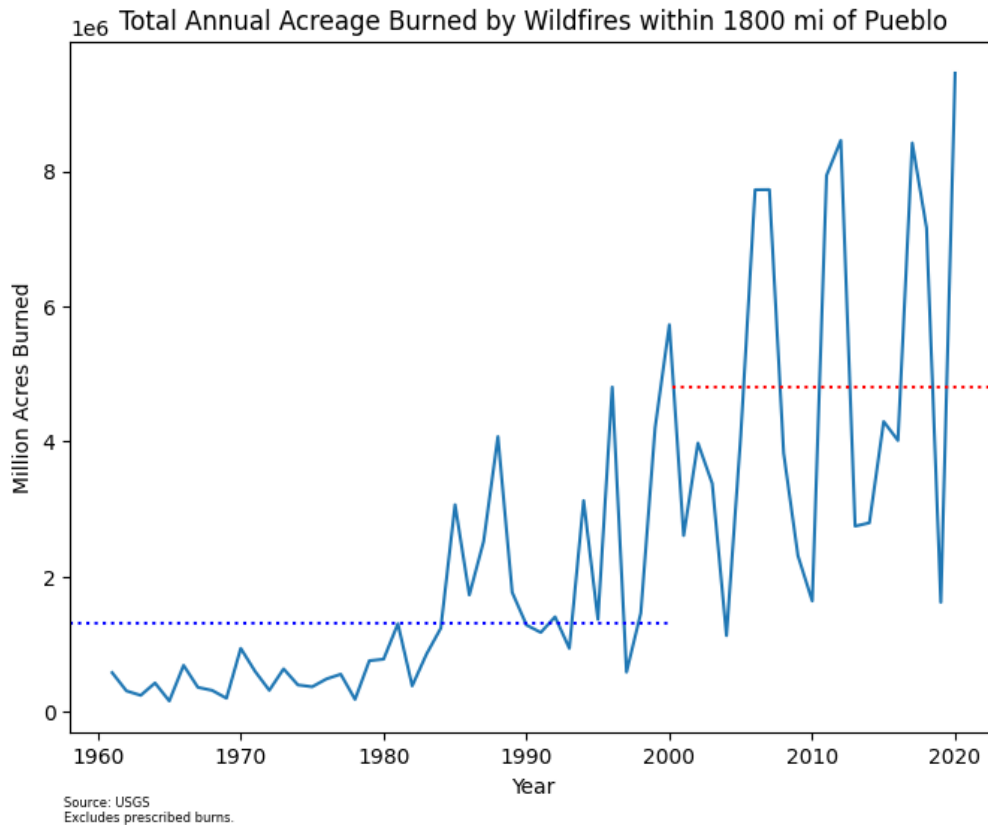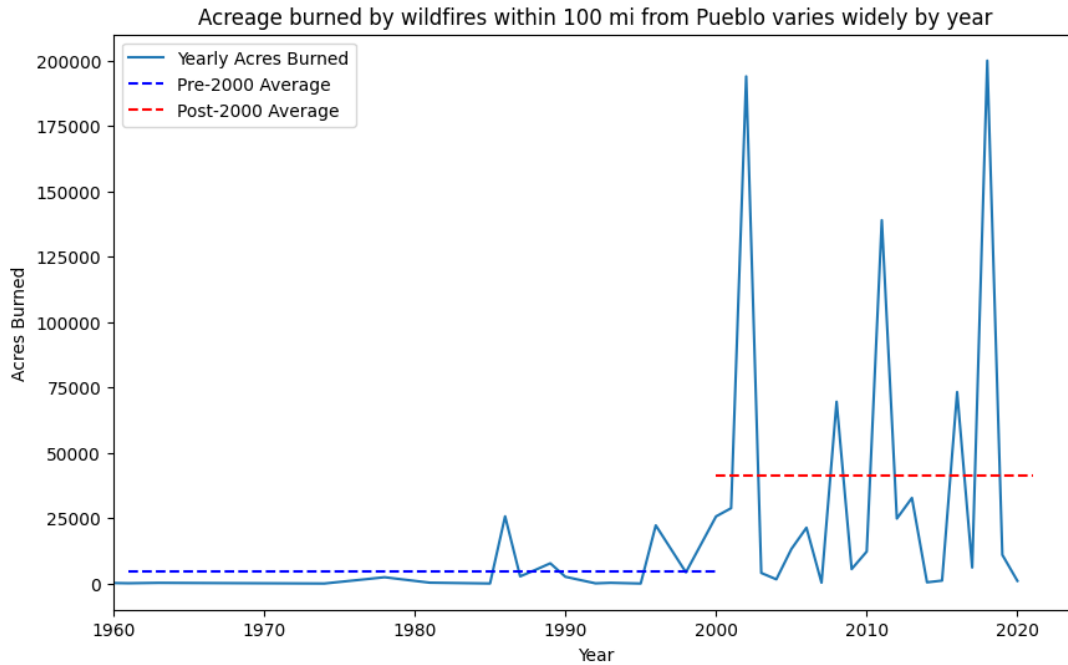
$$SIE = \frac{A}{d^2}$$

Figure 2

Figure 3

Applying the formula for the smoke intensity estimate produces a metric that is usually less volatile than the simple acreage burned within 100 miles from Pueblo, as evidenced by the fact that SIE peaks less intensely than acreage during 2002, 2008, and 2011 in **Figure 4**.
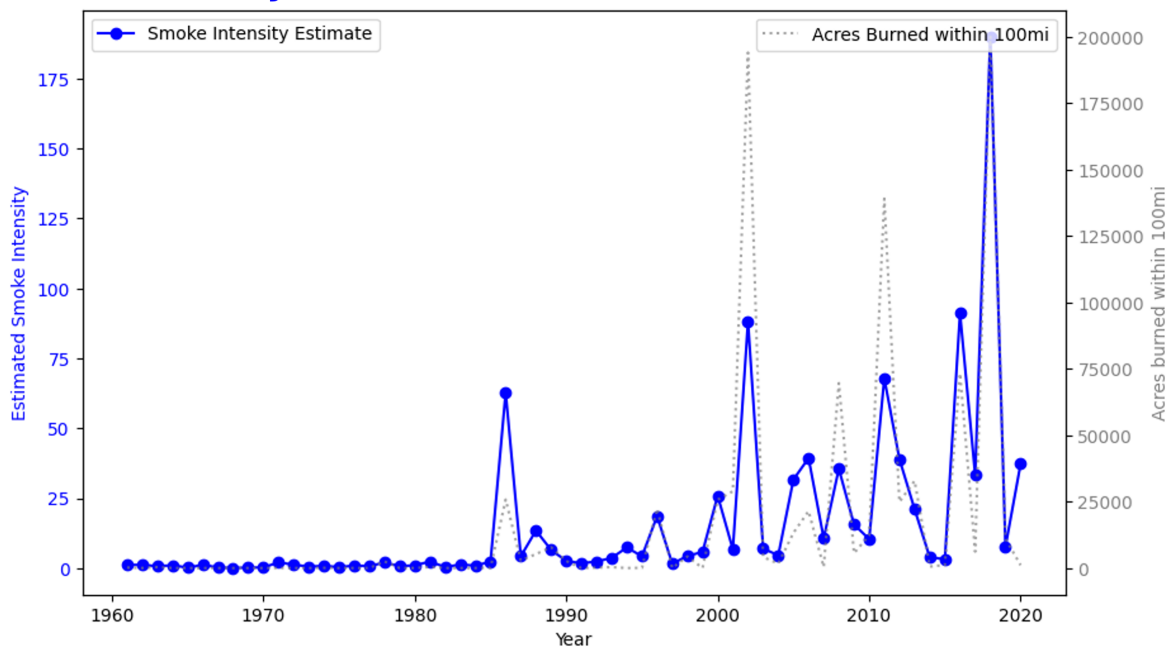

Figure 4

At the same time, SIE does not depart drastically from the acreage burned within 100 miles of Pueblo – as we should expect for a metric that weights fires close to Pueblo much more than fires far away. SIE exhibits the same trends as the acres burned: high volatility and a significant increase in average level after 2000.

After producing a historical time series of SIE, I turned to forecasting its future values. Observing that major peaks in SIE tend occur roughly 15 years apart – possibly indicating some seasonality in the data – I used a seasonal autoregressive integrated moving average (SARIMA) model with 15-year period parameter to forecast SIE 25 years into the future. Unfortunately, the noisiness of the historical data means that a 95% confidence interval (pink shading in **Figure 5**) is extremely wide, indicating very little confidence in yearly projections and the timing of peaks.
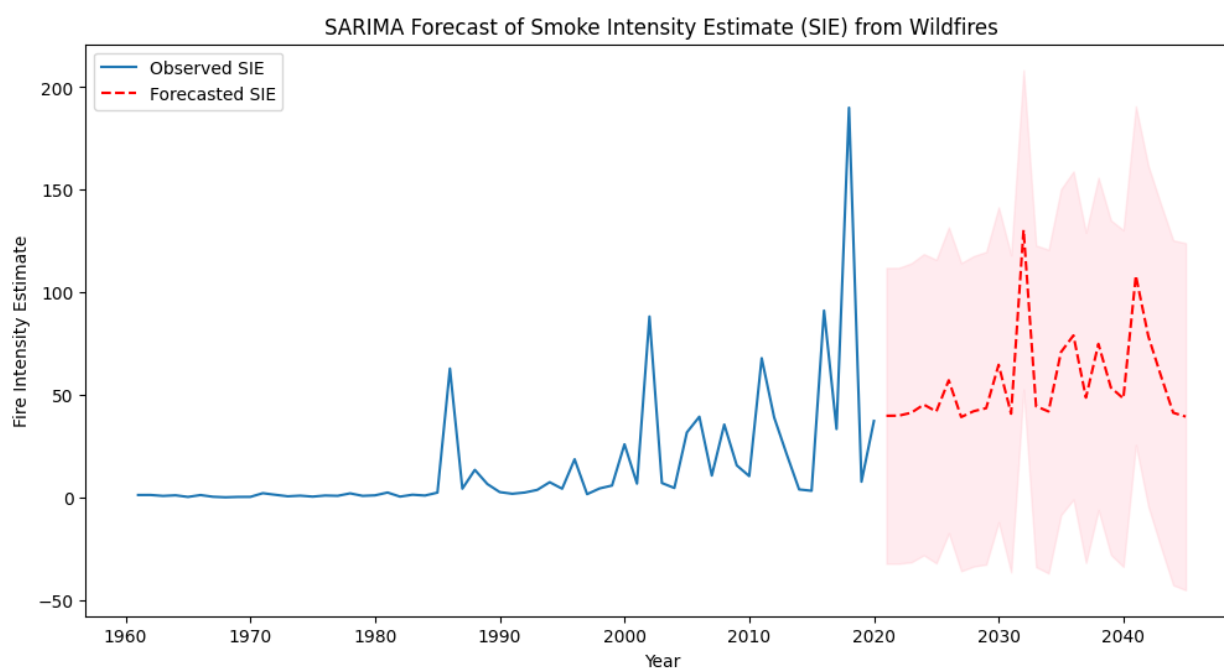

Figure 5

SIE's Relationship to AQI

The smoke intensity metric is a constructed metric meant to capture the severity of wildfire smoke in Pueblo. One way to sense-check this metric is to compare it to a separate notion of air quality. Wildfire smoke is one factor that can worsen air quality, but it is also one of many factors that exerts an influence. Therefore, I expected to find a moderate positive association between my smoke intensity estimate and the Air Quality Index, an index used by agencies like the EPA that *increases* as air quality *worsens*.

To examine the relationship visually, I superimposed the smoke intensity estimate and the air quality index on a single plot, shown in **Figure 6.**
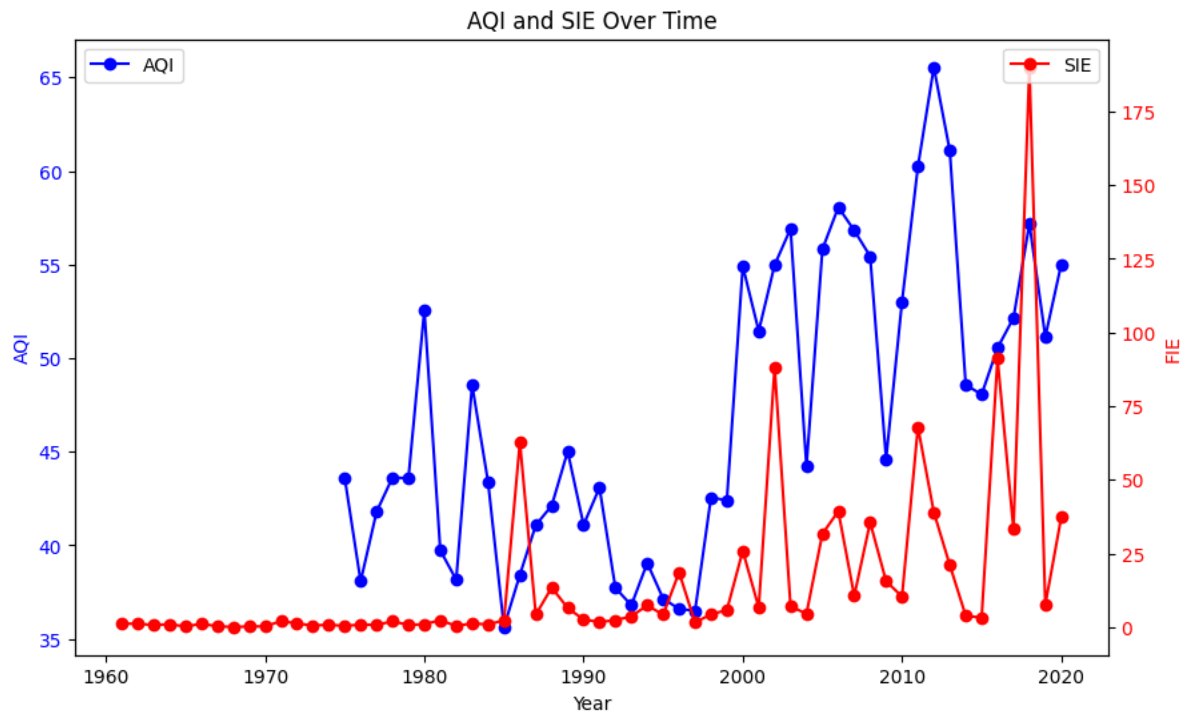

Figure 6

Visually, it seems there is some relationship – AQI and SIE both spike around 2001 and 2011, for example. The relationship becomes clearer when AQI and SIE values are placed on a scatterplot (Figure 7), demonstrating a moderate (R = 0.46) and positive (but nonlinear) relationship.
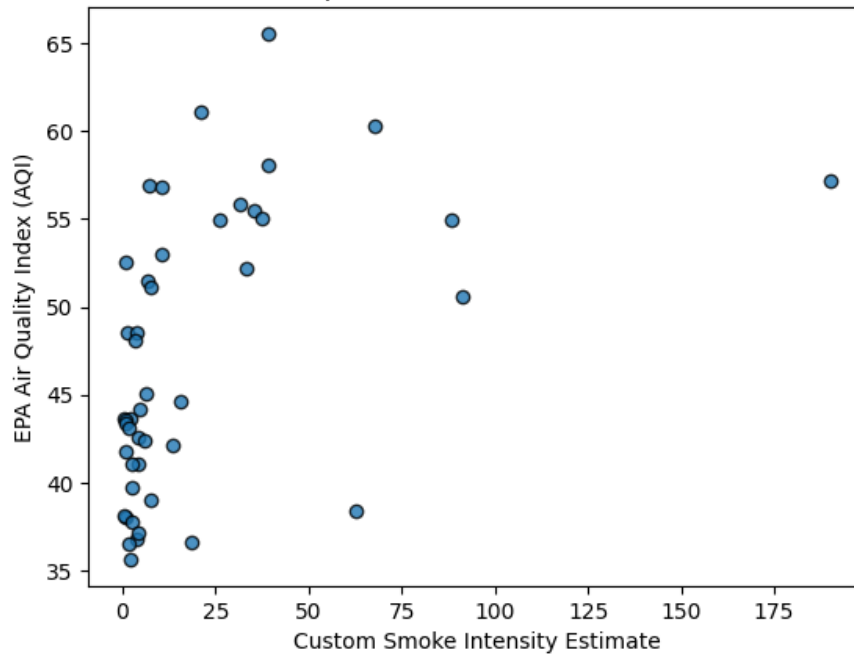
Figure 7

As expected, smoke intensity and the AQI are positively associated (again, a higher AQI means *worse* air quality), but not closely associated, as wildfire is one factor out of many that influence air quality.

SIE's relationship to Asthma Hospitalizations

Having developed a measure of smoke intensity and confirmed that it is moderately associated with air quality, I moved on to examining its relationship to an important wildfire-related public health outcome: asthma hospitalizations.

I expected to find a positive association between estimated smoke intensity each year and the average age-adjusted rate[3] of asthma hospitalizations in that same year. Surprisingly, I found that these two variables are *negatively* correlated (R = -0.32). Plotting the two time series on the same plot (Figure 8) suggests why – asthma hospitalizations are in steady decline since 2004, regardless of the level of smoke intensity.

---

[3] Colorado publishes age-adjusted and age-specific rates of asthma hospitalization per 100K people, since people in different age groups tend to be hospitalized for asthma symptoms at different rates. I used the age-adjusted figure because this makes it easier to compare values over time; if the mix of age groups in Pueblo's population changes over time, failing to include an age-adjustment could suggest that other factors may be driving the change, when really (some part of) the change is driven simply by a changing population.
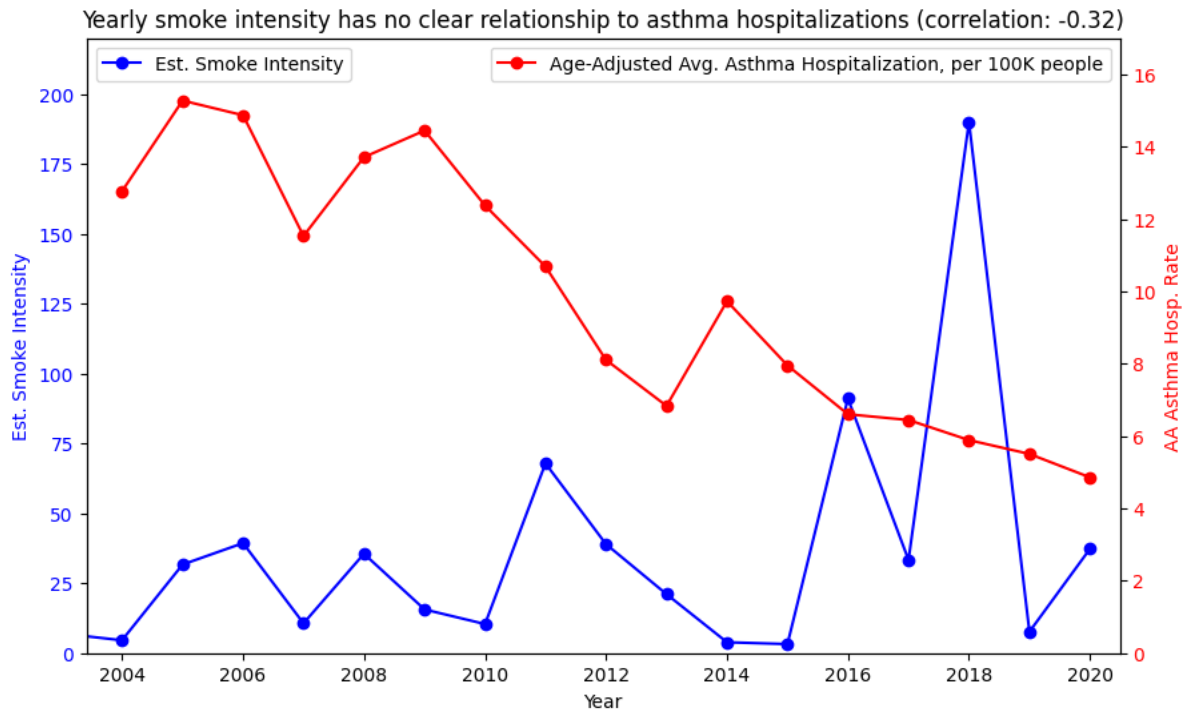
Figure 8

Clearly, there is a multiyear *trend* of declining asthma rates in Pueblo unrelated to wildfire smoke. This is great news for the residents of Pueblo, but it may stand in the way of properly testing my hypothesis. For example, effective public health interventions could be driving the consistent decline in asthma rates. If that effect is the predominant force affecting asthma rates, the impact of wildfire smoke could be drowned out by that steady decrease – even through wildfire smoke *could* still be causing spikes *relative* to that declining trend.

To test whether this could be the case, I *de-trended* the asthma hospitalization data. (Since smoke intensity does not exhibit a clear trend between 2004-2020, I did not detrend this series.) My modified hypothesis became that *de-trended* asthma hospitalization rates would be positively associated with estimated smoke intensity.

I applied two methods to de-trend asthma hospitalizations: in the first, I fit a linear model to the asthma data, with time as the independent variable. I treated this linear decline as the trend component of asthma hospitalizations; by subtracting this from the actual level of asthma hospitalizations, I inferred the residual (de-trended) rate of asthma hospitalizations each year. In the second method, I took the *difference* between asthma rates in successive years (starting in the second year of data) as the de-trended asthma hospitalization rate and compared it to the smoke intensity in each. Regardless of the methodology used to arrive at the de-trended asthma

hospitalization rate, the conclusion is the same: there is no relationship between yearly average smoke intensity and asthma hospitalization rates after adjusting for the general decline in asthma rates over time **(Figures 9, 10).**
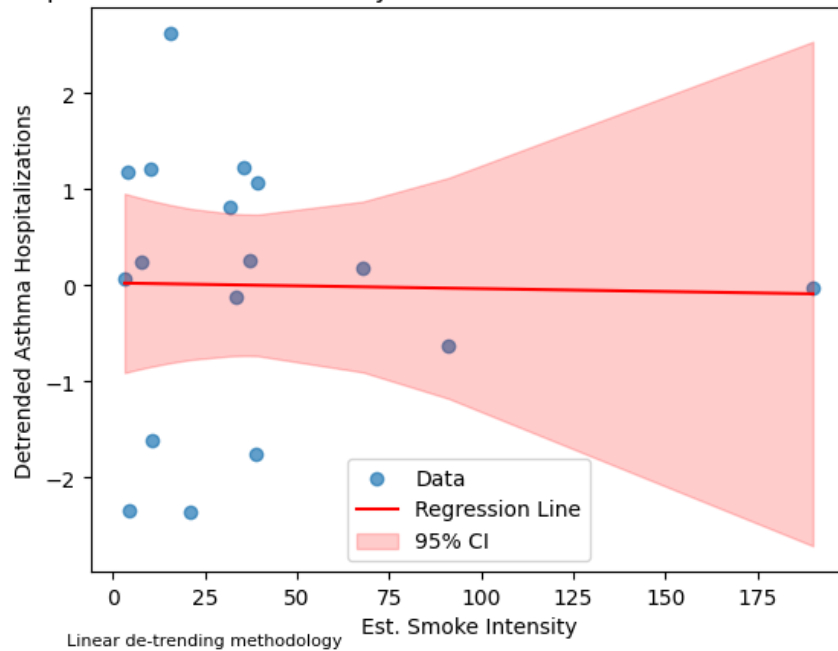
Figure 9

Figure 10

Discussion and Limitations

Whether or not the data on asthma hospitalizations in Pueblo is adjusted for long-term trends, the available evidence *does not* support the hypothesis that asthma rates rise as wildfire smoke intensity increases. Accounting for this surprising result, wh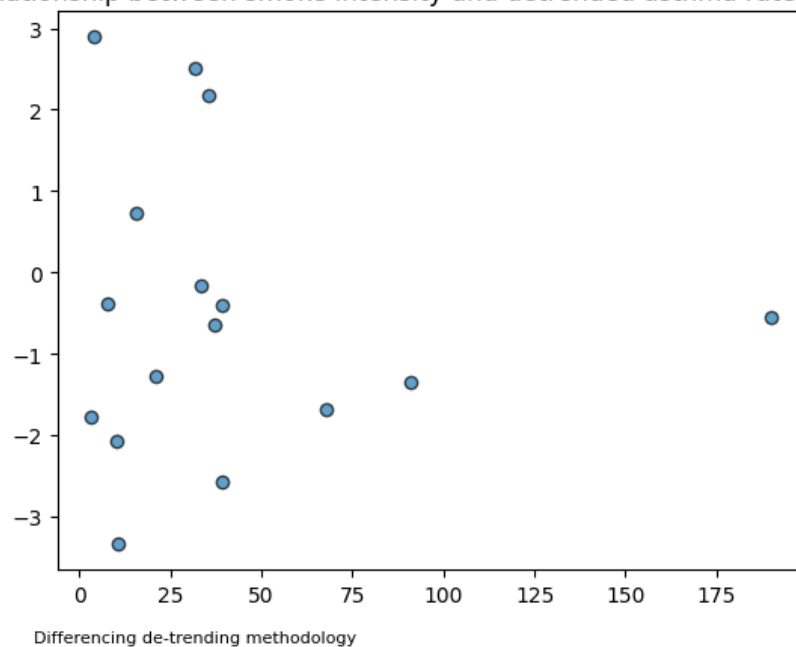ich contrasts with the CDC's findings about *daily* emergency room visits in New York (Meek et al, 2023), requires acknowledging a major limitation of this study: the fact that the available data is aggregated at a yearly resolution.

Unfortunately, we are limited to yearly resolution in the smoke intensity estimate, since the smoke intensity estimate is reliant on "fire polygon data [which] only (reliably) provide a year for each fire - [they do] not (reliably) provide specific start and end dates for the fire" (McDonald 2024). This limiting factor means that, even though Colorado publishes *monthly* data on asthma hospitalizations, it is only possible to *compare* these two variables at a yearly granularity. Therefore, any association between the smoke intensity estimate and the incidence of asthma is limited to yearly resolution. Wildfire data runs until 2020; this means there are 17 years in which I have a smoke intensity estimate AND asthma hospitalization data. Since we are limited to yearly granularity, this also means we're restricted to 17 data points. In general, small sample limit the conclusions we can draw from linear regression and time series analyses.

But in this case, it is plausible that there is an even more specific problem from performing linear regression on yearly data to make inferences about relationships between variables: aggregation bias (Nguyen 2020). An area typically feels the acute effects of wildfire smoke over a period of time much closer to days than years. It is very likely that, if the datasets in this study were available at a daily resolution, asthma hospitalization rates would be positively associated with the wildfire smoke intensity after adjusting for long-term trends and seasonality in the datasets. But when daily data is averaged to produce yearly rates, these patterns can be obscured, especially in an area like Pueblo when wildfires are not exceptionally common or intense.

Of course, there are alternative explanations for why there is no measurable association between my smoke intensity estimate and the rate of asthma hospitalizations. Perhaps these phenomena are not causally linked in Pueblo, although prior public health research suggests this is unlikely. Or perhaps the smoke intensity metric is simply not a very good metric – it does not accurately capture the level of wildfire smoke intensity in Pueblo in the way it was designed to do.

The only path to understanding of which explanation is supported by the evidence is more data, hinting at an implication of this work for Pueblo policymakers: they should expand data collection to ensure that wildfire data is available at least monthly, and ideally daily. This could mean coordinating with federal agencies like the US Geographic Survey to access and validate reliable daily

or monthly data on wildfire intensity near Pueblo. If this isn't possible, Pueblo city and county authorities may want to consider collecting daily data on wildfires themselves.

Understanding this relationship is important to the human-centered goals of this project and would be valuable to Pueblo policymakers. If better data supported the presence of a positive association between the intensity of wildfire smoke and the incidence of asthma, Pueblo policymakers could leverage *projections of smoke intensity in the future*, using the (hypothesized) positive association between smoke intensity and asthma incidence to prioritize people living with asthma for various public health interventions, like distributing masks or ensuring adequate numbers of hospital beds are available when wildfires strike. Currently, since smoke intensity cannot be linked to asthma in the historical yearly data, the forward-looking projections of smoke intensity generated in this study are not useful to Pueblo policymakers to support targeted public health interventions, since they don't imply any actionable information. By investing in gathering frequent, high-quality data, Pueblo policymakers could dramatically increase their predictive power to support an at-risk population.

A human-centered approach to data science is critical to surfacing issues like the ones highlighted above and providing transparent, public calls to action aimed at addressing them. In reflecting on how human-centered data science principles influenced this study, the focus on the public good and accessible communication of results were paramount. Adhering to principles of transparency and reproducibility ensures that all aspects my analysis – from the construction of the smoke intensity estimate to the analysis and visualizations of relationships to other variables, like asthma hospitalizations – are made accessible to the public.

This is not merely a nominal or theoretical benefit. By tracing the analytical procedures I followed to reach the surprising conclusion that smoke intensity and asthma hospitalizations are not related at a yearly level, researchers can fully appreciate the need for higher-granularity data and assess my claim that aggregation bias explains this result in full context, rather than simply getting the generic 'call for better data' that accompanies many empirical studies.

By running the fully reproducible notebooks that I am publishing with an open-source license accompanying this report, researchers can also challenge, modify, and improve my methodology for developing the smoke estimate; they can critique the methodologies I use to discern (the lack of) relationship between variables; and they can extend the work, testing the same hypothesized relationships I examine in this study when more granular or higher-quality data becomes available.

The principles of human-centered design not only guide my analysis, but the implications I suggest in light of its conclusions. The outcomes of this research should be actionable and ultimately serve the well-being of the Pueblo community. The problems with data resolution in this study should encourage more granular data collection aligned with the specific needs of vulnerable populations. By publishing these results transparently, I hope to make it evident to the Pueblo city and county governments that investing in better data-gathering is worthwhile.

Conclusion

In conclusion, this project provides important insight into the relationship between a novel metric for wildfire smoke intensity and asthma hospitalizations in Pueblo, CO. It explains how the smoke intensity estimate (SIE) is calculated from a comprehensive dataset on wildfires provided by the United States Geographic Survey (USGS). Furthermore, it evaluates SIE's association with air quality and asthma hospitalization rates in Pueblo. While the SIE shows a moderate positive association with the Air Quality Index (AQI), suggesting it is a plausible indicator of wildfire smoke's contribution to air quality degradation, no significant relationship was found between the SIE and asthma hospitalizations.

The absence of a measurable association between wildfire smoke and asthma hospitalizations in Pueblo is a surprising result that contrasts with the CDC's findings in a related study conducted in New York state. However, this discrepancy probably reflects the limitations of the available data, specifically the fact that it is only available at a yearly resolution. The lack of reliable daily or monthly granularity in wildfire data leaves no choice but to use yearly data, which may have an aggregation bias that obscures the short-term impacts of wildfire smoke on people with asthma. Future studies with daily- or monthly-resolution data are needed to address this limitation and refine our understanding of the relationship between wildfire smoke and asthma rates in Pueblo.

Despite the inconclusive analytical result, policymakers in Pueblo can draw several lessons from this study. First, the SIE represents a useful step toward better quantifying the impact of wildfire smoke on local air quality. Second, while this study does not find evidence of a direct link between wildfire smoke and asthma hospitalizations, the findings underscore the importance of investing in better data collection practices. Specifically, coordinated efforts to collect and validate daily or monthly wildfire intensity data would enhance their ability to detect short-term public health impacts and enable more targeted interventions.

In summary, this study highlights the complexity of assessing wildfire smoke's health effects and the need for highly granular data to fully understand its impacts. By expanding data collection,

Pueblo can better prepare to address the challenges posed by increasing wildfire activity and better serve its most vulnerable residents.

References

Centers for Disease Control. Climate and Health: Regional Health Effects – Northeast. June 3,
       2024. https://www.cdc.gov/climate-health/php/regions/northeast.html. Accessed Dec 4,
       2024.

McDonald, D. Part 1 – Common Analysis Description. DATA 512, Autumn 2024. Course at the
       University of Washington.

Meek HC, Aydin-Ghormoz H, Bush K, et al. Notes from the Field: Asthma-Associated Emergency
       Department Visits During a Wildfire Smoke Event — New York, June 2023. MMWR Morb
       Mortal Wkly Rep 2023;72:933–935. DOI: http://dx.doi.org/10.15585/mmwr.mm7234a6.

Nguyen, M. (2020). A Guide on Data Analysis. Bookdown.
       https://bookdown.org/mike/data_analysis/

US Census Bureau. Census Quickfacts: Pueblo, CO.
       https://www.census.gov/quickfacts/pueblocitycolorado. Accessed Dec 4, 2024.

Data Sources

- **Combined wildland fire datasets for the United States and certain territories, 1800s-Present** – **US Geographic Survey**
- **EPA Air Quality System (AQS) API** – United States Environmental Protection Agency
- **Asthma Dataset, Colorado Environmental Public Health Tracker** – Colorado Department of Public Health and the Environmental