Norman Wildfire Smoke Analysis

Alex Netzley

12/4/2024

Introduction

A bustling college town home to the University of Oklahoma, the city of Norman Oklahoma is situated 20 miles south of the Oklahoma state capitol, Oklahoma City. Located in the center of the Great Plains, Norman's geographical location is famous for its open, expansive lands and extreme weather, especially in the form of high winds and tornadoes. However, in recent years, these high winds have more frequently been bringing in extreme weather of another variety: wildfire smoke. Across the globe, wildfires have been on the rise[1], and central Oklahoma and the city of Norman are no exception. Despite the growing concern around the rising wildfires, there has been limited research into the potential direct impacts of wildfire smoke on the community of Norman. In this analysis, I will investigate the state of wildfires and wildfire smoke around Norman, as well as potential relationships between wildfire smoke and asthma rates in the community. Specifically, this investigation will attempt to answer the following two questions:

1. How has the intensity of wildfire smoke in and around Norman changed in the last six decades?
2. What is the state of asthma in and around Norman, and is there any relationship between the rates of asthma and the wildfire smoke?

By better understanding the current state and potential impacts of future wildfire smoke, I hope to provide government agencies and communities the necessary building blocks towards making informed decisions in resource allocation and overall preparing for the consequences of further wildfire smoke.

Background

Asthma is one of the most common chronic diseases, impacting 1 in 12 individuals in the US[2,3]. Causing coughing, wheezing and shortness of breath, asthma has a broad range of negative impacts on not only the individual by decreasing their quality of life, but also on a community, by increasing medical expenses and loads on hospitals. In fact, one study found that from 2008-2013, asthma cost medical systems in the US an estimated $50.3 billion dollars[4]. Although asthma can be caused by a variety of sources, one of the primary external causes of asthma is air pollutants, specifically fine particulate matter ($PM_{2.5}$), including those generated from wildfire smoke. Therefore a variety of previous studies have linked increases in $PM_{2.5}$ particulate matter from wildfire smoke to increases in asthma[5,6].

To determine the state of wildfire smoke around Norman and any potential regional relationships between wildfire smoke and asthma, I analyzed the following three datasets.

USGS Wildland Fire Historical Fire Data

To obtain historical wildfire data, I downloaded the full USGS_Wildland_Fire_Combined_Dataset.json, from the USGS government website[7], which contains a detailed government repository of all fires recorded (over 130,000) starting in the early 1900s. For each fire, the dataset contains a variety of metadata detailing attributes such as fire size, fire type, and date, as well as geographical coordinates detailing the location of the "ring" of the fire. A detailed overview and data schema are provided on the website reference above.

US Environmental Protection Agency Air Quality Service Historical AQI Data

As part of the analysis, I extract and process historical AQI to supplement the wildfire smoke estimate. To extract this data, I use the US Environmental Protection Agency (EPA) Air Quality Service (AQS) API[8]. This database contains historical air quality measurements from a variety of monitoring stations throughout the US, including in and around Norman, OK. Specific API usage documentation is provided on the website reference above.

Oklahoma State Department of Health Statistics Survey Asthma Data

To better understand the state of asthma in and around Norman, I extracted historical asthma data from the Oklahoma State Department of Health Statistics "OK2SHARE" website[9]. This website contains a variety of medical, hospital, and survey records from across the Oklahoma state medical system. For my analysis, I specifically extracted the Behavior Risk Factor Surveillance System (BRFSS) crosstab dataset associated with the Central region of Oklahoma, which reports survey results with the number and proportion of respondents with asthma as well as their current smoking status. This combined survey data was available from the years 2000 and 2003-2023. The usage of this data was in line with the license documented on their website referenced above, as I am "...monitoring the health of the people of Oklahoma".

Methodology

Processing the Wildfire Data

After extracting, loading, and parsing the raw wildfire data from the USGS government website, I performed a multi-step process to filter the dataset to only the relevant fires for my analysis. First, I wanted to keep only the data from the past six decades as data prior to this was much less reliable, so I filtered the dataset to fires occurring between 1961 and 2021. Next, for a fire to have the potential to create smoke in Norman, it must be located geographically close. For this general threshold, I chose 650 miles to be the absolute cutoff, and excluded all fires outside of this distance. To compute the distance from the fire to Norman, I used the average distance of all points on the geographical "ring" of the fire to estimate the fire centroid, as this provides a more accurate representation of the distance of the fire rather than selecting the closest distance point. However, as I will discuss further in the next section, the actual cutoff distance methodology ends up not having a large impact, as fires near the border of the 650 radius are scaled down in the smoke estimate calculation.

Computing the Smoke Estimate

Although the size of any given fire in the dataset can be directly extracted, understanding the impact that a fire might have on smoke contribution to Norman is also related to the distance the fire was away. Therefore, rather than purely aggregating the total fire acreage for a given year, a formula was devised to weight fires by their size and distance. Equation 1 describes the formula derived to estimate the contribution of any particular fire to wildfire smoke in Norman.

$$SC = D * A \qquad \text{(Eq. 1)}$$

Here, *SC* is the Smoke Contribution for a particular fire, *D* is the Distance Impact of the fire, and *A* is the Area Impact. The Distance Impact and Area Impact are computed by Equation 2 & 3, respectively.

$$D = (650 - \textit{Miles from Norman})/650 \qquad \text{(Eq. 2)}$$
$$A = (\textit{This Fire Acreage})/(\textit{Max Fire Acreage}) \qquad \text{(Eq. 3)}$$

Both *D* and *A* can take on theoretical values between 0 (when the fire is 650 miles away from Norman, or 0 acres) to 1 (when the fire is Norman, and is the largest in the dataset). Because of this, their product, *SC*, also takes on a value between 0 and 1. To compute the total Smoke Impact estimate for a given year, I then summed the *SC* (Smoke Contributions) from all the fires in a given year. By computing this impact metric, I am able to better estimate the overall smoke impact on Norman for a given year in a scaled, simplistic manner. Of course, there are a variety of strong assumptions and limitations to this methodology, which will be discussed further in the "Limitations" section below.

Extracting and Analyzing the AQI Data

Next, the AQI data was incorporated to supplement and validate our smoke estimate. The AQI data was extracted using the US Environmental Protection Agency (EPA) Air Quality Service (AQS) API described in the "Background" section above. From the API call, the sensor data was extracted from Cleveland County, home to Norman. However, upon initial investigation, only four monitoring stations were found, all of which were full of incomplete and missing data. Therefore, the search of monitoring stations was expanded to include those within a 25 mile radius and one promising monitoring site located in nearby Oklahoma City was found. This monitoring station contained a much more robust dataset, with daily air quality values for a variety of different particulates, including $PM_{2.5}$, PM_{10}, CO, NO2, and O2, for roughly the past 30+ years. To estimate the yearly average AQI value, I then computed the max of the averages across all particulate types, and averaged across all days of the fire season (May 31 to November 1), which is a simplified version of a suggested AQI formula presented by the EPA[10]. After computing this yearly AQI estimate, it was then compared to the yearly Smoke Estimate to validate the Smoke Estimate calculations.

Developing the Smoke Estimate Forecast Model

After completing the calculation of the Smoke Estimate, a predictive model was developed to attempt to forecast Smoke Estimate predictions for the years 2025-2050. Specifically, the Holt-Winters exponential smoothing time-series forecasting model from the statsmodels python package with the additive trend[11] was used. The Holt-Winters model is a modified exponential smoothing model with trend and seasonality components[12], making it particularly suitable for our time-series forecasting problem, as preliminary EDA revealed the wildfire smoke is clearly on the rise and appears to have some sort of yearly cyclical behavior. The model is also able to estimate confidence intervals with its associated predictions, which is especially valuable in our scenario to account for the rather noisy data.

Accounting for the Impact of Tobacco Smoking Status on Asthma Rates

After completing the Smoke Estimate forecasting, I then turned my attention towards investigating the rates of asthma, where the data was extracted from the Oklahoma State Department of Health Statistics website. Before directly extracting the reported asthma rates on the population as a whole, I wanted to isolate the impacts of other potential confounders, such as smoking. As discussed previously, air pollutants from wildfire smoke are a leading contributor to asthma. However, another leading external trigger of asthma is tobacco smoking[13], and tobacco smoking has seen a drastic decrease in the past six decades[14]. Therefore, I wanted to be sure that this decrease in tobacco smoke wasn't confounding the overall rates of asthma, so I compared the rates of asthma for smokers and non-smokers respectively. EDA showed clear differences in asthma rates between smokers and non-smokers. Therefore, in the future model development work, I specifically analyze the impact of wildfire smoke on asthma for non-smokers.

Developing the Asthma Models

Lastly, after extracting the yearly non-smoker asthma data, I investigated the relationship between the asthma data and the Smoke Estimate. After plotting both datasets on a scatter plot, a simple linear regression model predicting asthma from the Smoke Estimate was developed as a general linear trend was observed and confirmed by the Pearson correlation coefficient. Although simple linear regression does have a variety of assumptions and limitations, it was selected primarily for its simplicity and interpretability. After developing the simple linear regression model, we then used the forecasted Smoke Estimate predictions as inputs to the linear model to predict asthma rates from 2025-2050.

One additional challenge to aggregating the Smoke Estimate forecast model with the Asthma linear regression model was the compounding of prediction uncertainties. To estimate the confidence intervals, I took a conservative approach, using the 95% upper bound from the linear regression output from the 95% upper bound Smoke Estimate forecast value (and opposite for the lower bound). This approach is conservative in that it gives us wider confidence intervals than using either model's uncertainty measure in isolation, but likely does a better job in representing the uncertainty of the prediction.

Human-Centered Consideration

Throughout this project, I made a proactive effort to keep human-centered design making at the forefront of analysis choices. First, I organized and documented my full project analysis, including data, code, and findings into a public and open-access github repository to enable reproducibility. Next, when analyzing the potentially sensitive health data with asthma, I followed all licensing requirements and did not attempt to re-identify any individuals. Additionally, I extracted all three data sources from reputable government websites (the USGS, Oklahoma State Department of Health Statistics, and the EPA), each of which are held to high standards of data integrity and privacy.

Findings

Wildfire Trends

Figure 1 shows the distributions of the number of fires and their distance from Norman, from 1961-2021. Additionally, the dotted red line calls out the 650 mile cutoff threshold that was used in the analysis

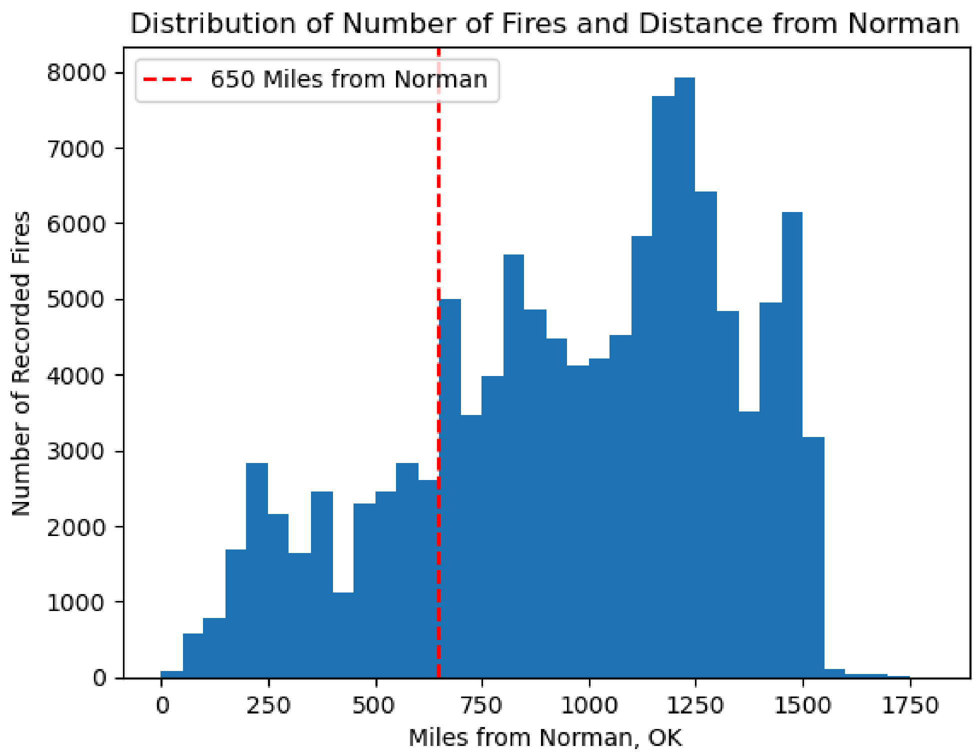

Figure 1: Histogram of the number of fires and their distance from Norman

Overall, we see that in general, there are more fires as we move further away from Norman. However, we should note that although the radial distance increases at a constant rate, the total square miles increases drastically as we get further from Norman, so we actually can't conclude anything about fire density. Second, it would appear that there is a distinct drop off of fires past 1550 miles, however, as Norman, OK is located near the geographic center of the US, beyond 1550 miles essentially puts us beyond the Continental United States.

Next, we can visualize the total fire acreage to get a sense of the yearly trends of wildfires, shown below in Figure 2.

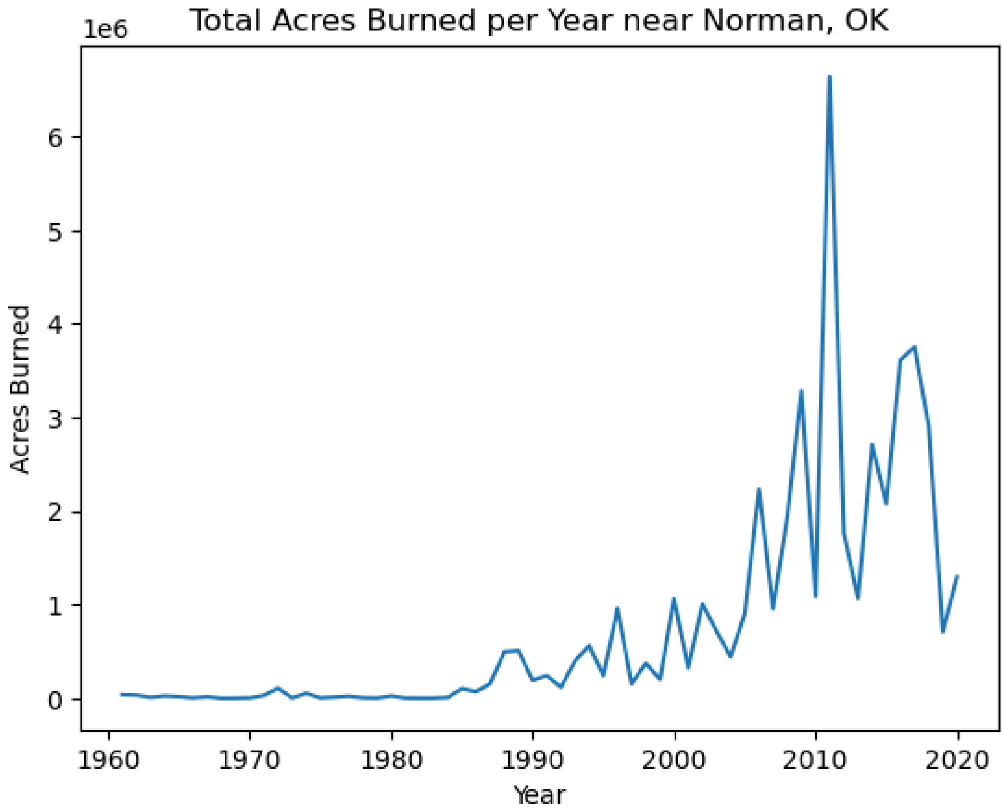

Figure 2: Yearly total acreage burned in near Norman, OK

Figure 2 displays the total number of acres burned per year within a 650 mile radius of Norman, OK, covering the years 1961 to 2021. Here, we can clearly see that wildfires are worsening in and around Norman, slowly in the long term but much more dramatically in recent years. Additionally, it is evident that the wildfires for any specific year display a higher amount of variability.

Smoke Estimate and AQI

After computing the Smoke Estimate and extracting the AQI estimate, we then sought to compare the two to validate our Smoke Estimate calculation. Figure 3 shows the year-over-year comparison of these two metrics. Notably, we only have AQI data going back to 1985.

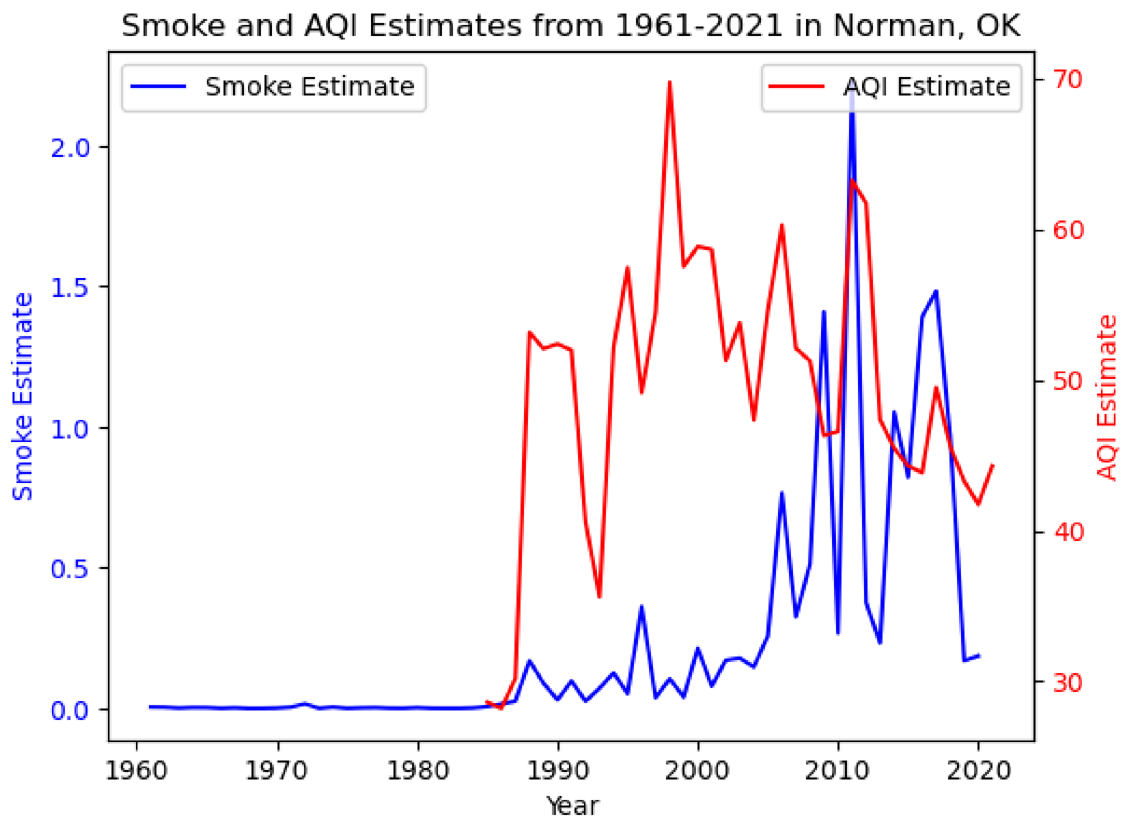

Figure 3: The calculated yearly Smoke Estimate and AQI estimate in Norman, OK

Overall, we see a few key takeaways from the graph. First, the Smoke Estimate has been rising dramatically, especially in recent years, and follows a similar trend to the wildfire acreage in Figure 2. However, the AQI Estimate has been generally decreasing over the past three decades. This could partially be explained by the fact that wildfires are not the only contributors to higher AQI, other pollutants from different sources (i.e. manufacturing) could also play a large role. Perhaps the Clean Air Act[15] accounted for more of the overall decline in AQI that can effectively balance out the impact of increased wildfire smoke. On the whole, although the two do show opposite long-term trends, we do see a bit of concordance in spikes from corresponding years, indicating that there is perhaps still some relationship present.

Modeling the Smoke Estimate and Forecasting

After computing our Smoke Estimate, we then developed a Holt-Winters exponential smoothing model to forecast the Smoke Estimate (Figure 4).

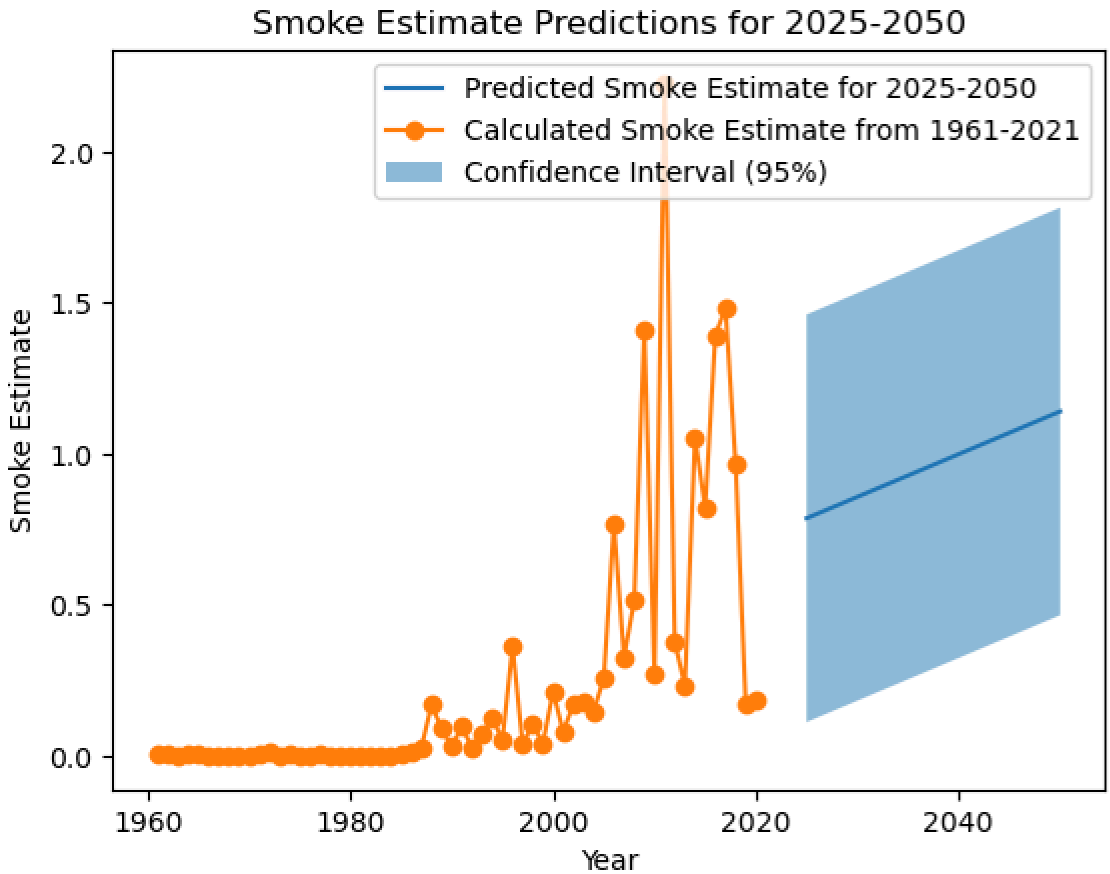

Figure 4: The forecasted Smoke Estimate predictions for 2025-2050

As we can see, the Smoke Estimate is expected to keep increasing in future years. However, the large confidence intervals indicate a high amount of uncertainty in predicting any specific year's Smoke Estimate, so we may expect to see some years having Smoke Estimates near the upper band, while others more moderate smoke years, near the lower band.

Exploratory Data Analysis of the Asthma Data

After extracting the survey asthma data, I performed EDA to attempt to identify whether smoking status played an impact on asthma rates. Figure 5 shows the reported smoking rates, while Figure 6a and 6b show the asthma rates for smokers and non-smoker, respectively.

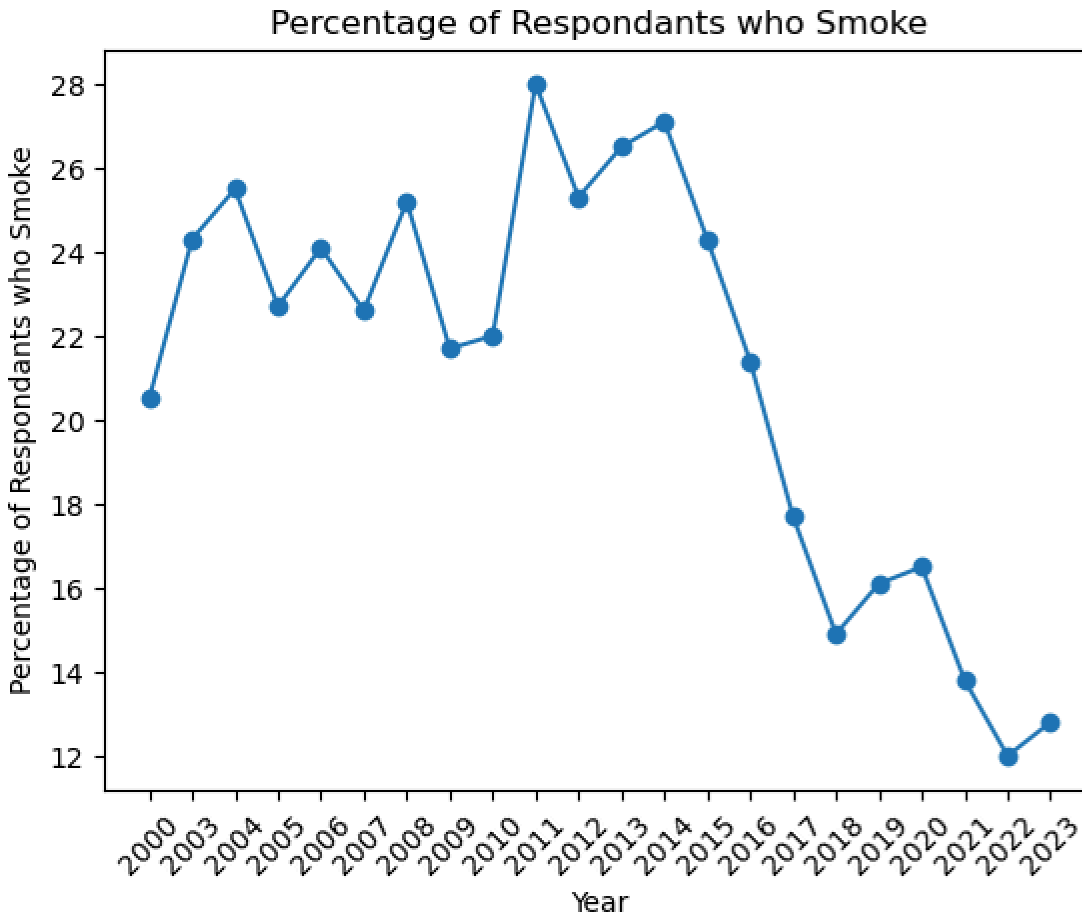

Figure 5: The yearly percentage of smokers in Central Oklahoma

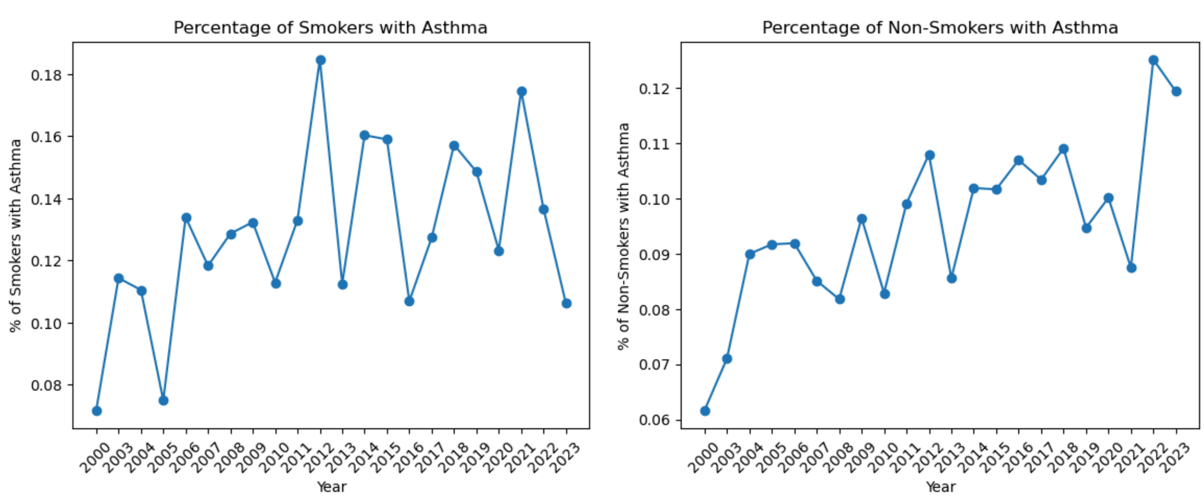

Figure 6a and 6b: The yearly percentage of smokers and non-smokers with asthma in Central Oklahoma

As we can see, the rates of smoking have dropped consistently and drastically over the past 20 years. Additionally, we see that the rates of asthma, although increased for both smoking and non-smoking groups, are consistently lower for the non-smoking group, and have increased more dramatically. Since the total number of smokers is declining heavily and nearing zero, we can expect that this trend will continue into future years. Therefore, we chose to look at the impact of wildfire smoke on non-smoking asthma rates, as non-smoking asthma rates are likely to make up an ever-increasing majority of the population in future years.

Modeling Asthma and Forecasting

After extracting the yearly non-smoker asthma data, we compared this to the Smoke Estimate data (Figure 7).

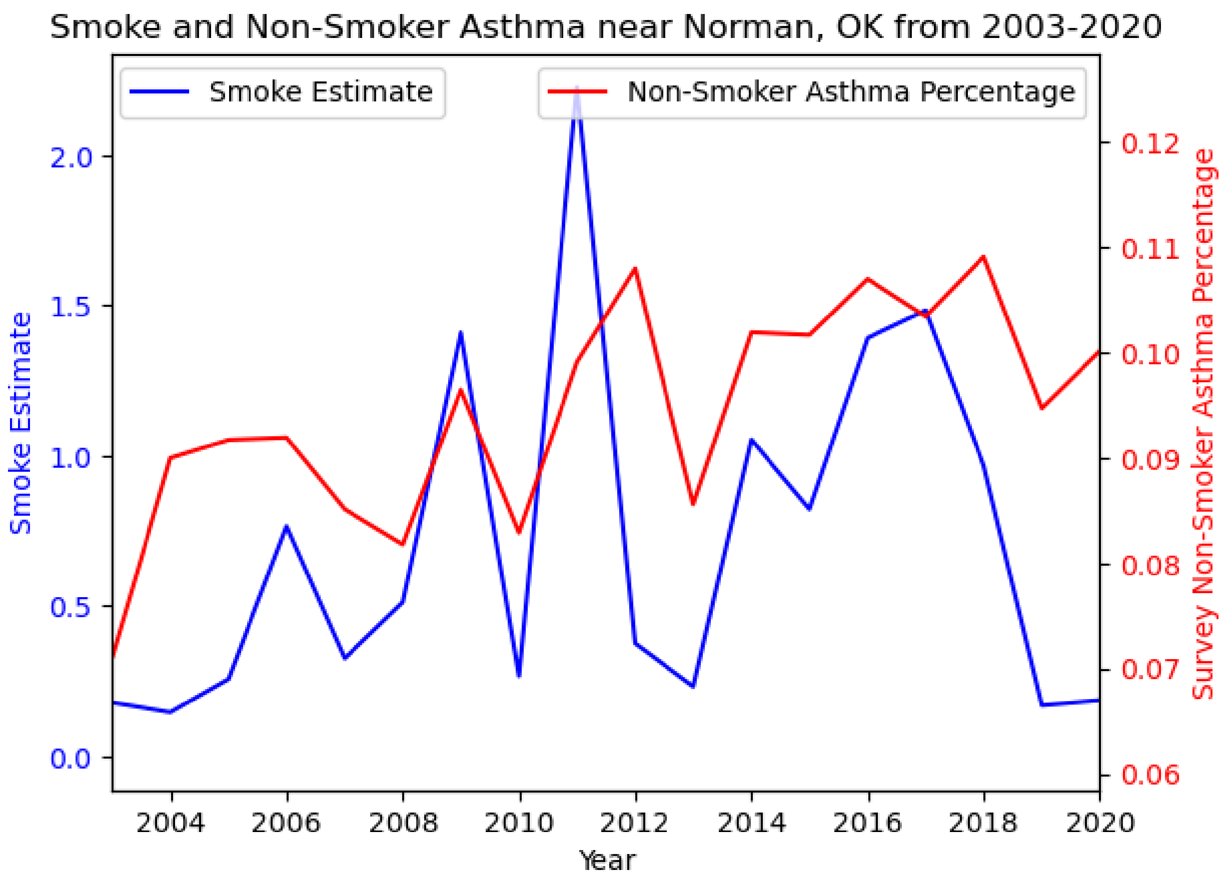

Figure 7: Smoke Estimate and asthma data yearly from 2003-2020

This figure shows a clear visual relationship between the two, with spikes happening in generally consistent years. To supplement this initial observation, I converted the data to a scatter plot, and computed a pearson correlation coefficient of 0.516 (P-value of 0.028), indicating a moderate positive correlation. Next, a simple linear regression model was developed to predict asthma from the Smoke Estimate. Figure 8 below shows the scatter plot and linear regression line of best fit.

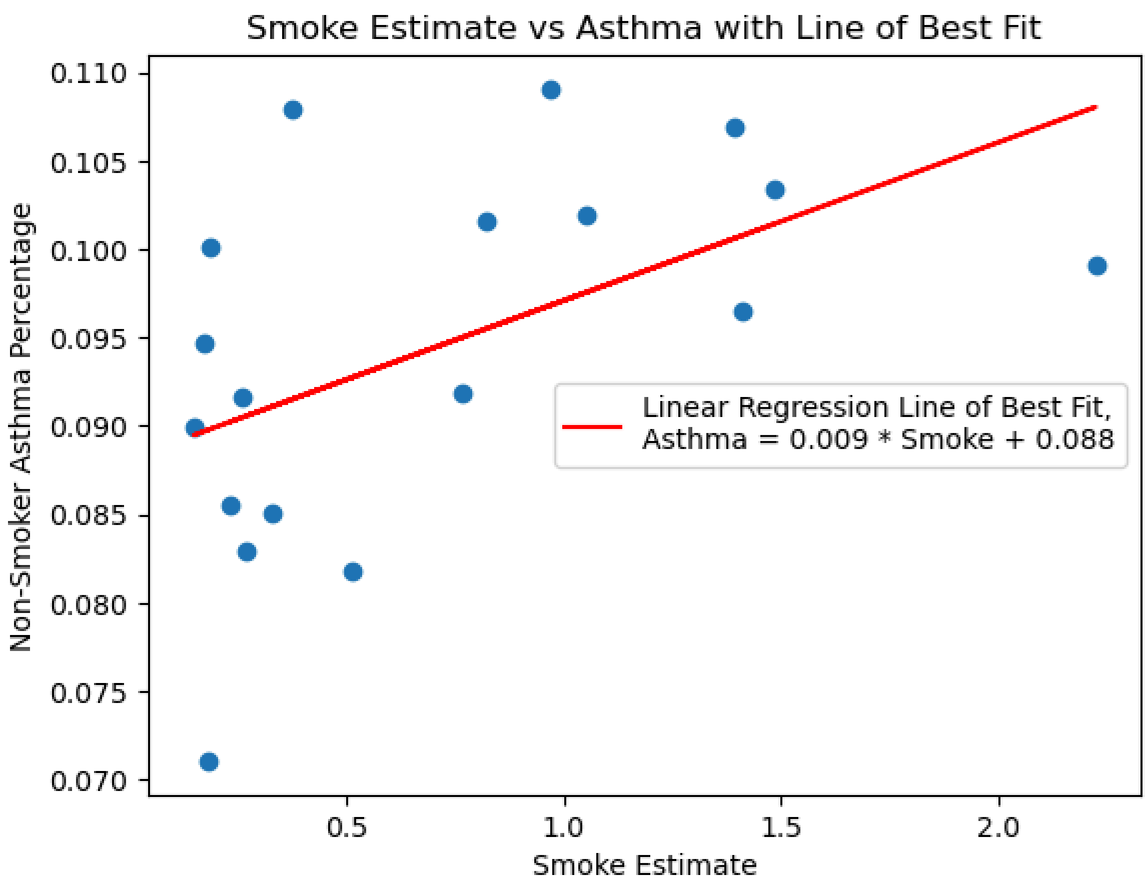

Figure 8: Scatter plot of Non-Smoker Asthma Percentage and Smoke Estimate, with the linear regression line of best fit

The linear regression analysis returned a linear regression LOBF estimate of:

$$Asthma = 0.009 * Smoke + 0.088 \qquad \text{(Eq. 4)}$$

The model had 16 degrees of freedom, and had an R-squared of 0.266, indicating the 26.6% of the variance in non-smoker asthma percentage could be explained by the smoke estimate.

After developing the linear regression model, we then used this model, in conjunction with the forecasted Smoke Estimate predictions, to forecast the asthma rates form 2025-2050 (Figure 9).

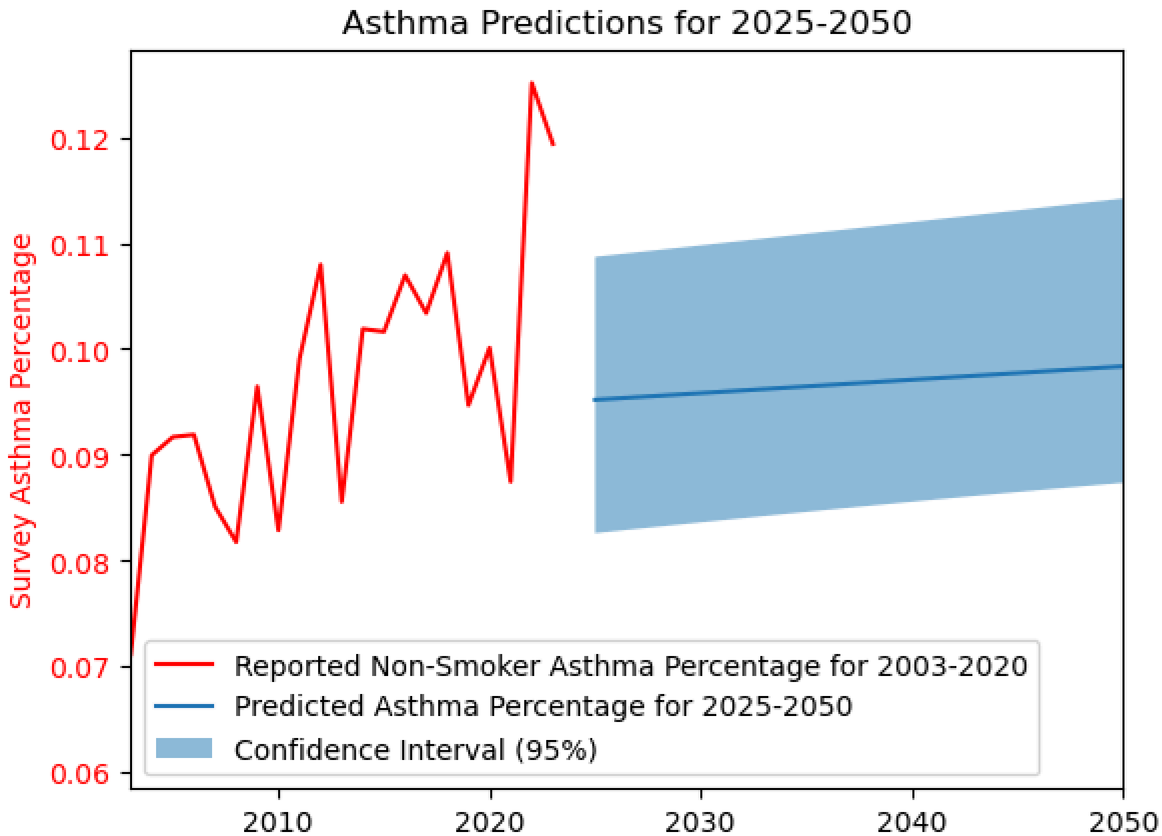

Figure 9: Forecasted asthma percentage from 2025-2050

As we can see, similar to the Smoke Estimate forecast, we predict a steady increase in asthma percentage in the long run, with a fairly high uncertainty in any specific year.

Discussion

After completing the analysis, we revisit the initially proposed research questions. First, *how has the intensity of wildfire smoke in and around Norman changed in the last six decades?* Clearly, we see that wildfires in and around Norman have dramatically increased, especially in the past few decades. Our forecasts estimate that this trend will continue, resulting in even worsening wildfire smoke on the community. Second, *what is the state of asthma in and around Norman, and is there any relationship between the rates of asthma and the wildfire smoke?* Asthma rates are also on the rise in Norman, even when accounting for the potential confounding factor of tobacco smoking. Additionally, there appears to be a moderate positive correlation between asthma rates and wildfire smoke, so similar forecasting methods anticipate increased asthma rates in the Norman community in the upcoming years.

Overall, city government officials and community members should take these results seriously. I recommend the following two main courses of action for the future. First, in the short-term, government officials should advocate for and invest in resources to support healthcare workers and hospital systems. With a likely increase in asthma, we will also see a corresponding increased burden placed on medical infrastructure. Investing in infrastructure improvement now will ensure that the medical systems will be able to handle the influx of patients when they arrive, instead of scrambling to improve infrastructure once the need is there, potentially compromising the health and well-being of the citizens.

Second, in the longer-term and as a more holistic goal, city officials should attempt to collaborate with other government entities nationwide to advocate for increases in research funding for wildfire smoke impact assessments such as this analysis. Specifically, incorporation of additional data sources, and more advanced and robust modeling techniques would likely be able to give a greater understanding of the potential impacts of wildfires to all parties involved. Additionally, this could likely involve a sort of collaboration and sharing of information across the globe – we might not be able to fund every community with the necessary resources to improve data collection and analysis methods, but publicly publishing analysis findings from any community will likely allow other communities to draw parallels and reap similar benefits.

Human-Centered Consideration Reflection

Human-centered design decisions were prevalent in my choices throughout this project. Beyond the engineering best practices of reproducibility and open-sourcing described previously, I also made a few conscious decisions to help reduce bias in my analysis. First, in data collection for both the AQI and asthma data, I specifically isolated my results to regions located near Norman. This ensured that the data I was drawing inferences from was consistent with Norman itself, and not biased by the demographics and measurement instruments from far away communities. Additionally, throughout the analysis and corresponding report, I made accessibility-focused design choices. For example, I kept technical jargon to a minimum, and used clearly labelled and referenced figures. I also walked through the analysis step-by-step and provided ample links to resources to enable understanding for a wide audience, regardless of technical background or domain experience in wildfire, smoke, or asthma.

Limitations

This study contained a large number of limitations due to its high-degree of complexity. From the Smoke Estimate calculations, some strong assumptions were made due to a lack of high-resolution data. For example, I did not take into account wind patterns when computing the Smoke Contribution, but rather assumed that fires a certain distance away impacted Norman the same, regardless of direction. However, we know that prevailing wind patterns (general prevalence of Westerly winds) mean that wildfires occurring west of Norman would likely have a greater impact on the smoke that Norman experiences. Another limitation is simply the lack of yearly data that is available on both the AQI and asthma side. For asthma, we could only go back in consecutive years to 2003. This limited time frame severely restricted any ability to do time-series forecasting, as we only had 17 years of data. To address this limitation, the two-pronged approach combining the forecasted Smoke Estimate and linear regression model was devised. However, the linear regression model itself comes with a variety of assumptions and limitations. First and foremost, by treating the asthma and smoke estimate data as a true linear relationship, we are removing the temporal yearly component, and quite directly violating the independence of observations assumption of linear regression. Another fundamental limitation of the asthma data was its survey nature. Although the Oklahoma State Department of Health Statistics did their best to reduce sampling bias, there are other forms of bias that can be present in surveys. Since asthma was not assessed and recorded by a medical practitioner, it was left to the individual to self-report. This also applies to the smoking status, which due to its taboo nature, may have complicated the analysis by being underreported.

Conclusion

In conclusion, this study sought to identify trends in wildfire smoke in and around Norman, and uncover a potential relationship between smoke and rates of asthma. Ultimately, the study showed that wildfires and wildfire smoke are on a dramatic rise in and around Norman. Additionally, asthma rates are also on the rise, and display a moderate positive correlation with wildfire smoke. Lastly, predictive models forecast further increases in wildfire smoke, leading to corresponding increases in asthma. Community leaders should invest in healthcare infrastructure to better prepare for an increase in asthma hospitalization, as well as in further research to better understand the potential impact of increased wildfire smoke on community health.

This study also showcases some key components of human centered data science. The study follows a high-degree of organization, clarity, and documentation, allowing for easy reproducibility and explanation. Additionally, this analysis took specific steps to help mitigate bias in the selection of specific regions to isolate demographics for a community.

References

[1] MacCarthy, J., Richter, J., Tyukavina, S., Weisse, M., & Harris, N. (2024). The latest data confirms: Forest fires are getting worse. https://www.wri.org/insights/global-trends-forest-fires

[2] National Center for Health Statistics. NHIS Child Summary Health Statistics. Data accessed September 19, 2024. Available from https://data.cdc.gov/d/wxz7-ekz9

[3] National Center for Health Statistics. NHIS Adult Summary Health Statistics. Data accessed September 19, 2024. Available from https://data.cdc.gov/d/25m4-6qqq

[4] Nurmagambetov, T., Kuwahara, R., & Garbe, P. (2018). The economic burden of asthma in the united states, 2008–2013. Annals of the American Thoracic Society, 15(3), 348–356. https://doi.org/10.1513/AnnalsATS.201703-259OC

[5] Wilgus, M.-L., & Merchant, M. (2024). Clearing the air: Understanding the impact of wildfire smoke on asthma and copd. Healthcare, 12(3), 307. https://doi.org/10.3390/healthcare12030307

[6] Noah, T. L., Worden, C. P., Rebuli, M. E., & Jaspers, I. (2023). The effects of wildfire smoke on asthma and allergy. Current Allergy and Asthma Reports, 1–13. https://doi.org/10.1007/s11882-023-01090-1

[7] Welty, J., & Jeffries, M. (2021). Combined wildland fire datasets for the United States and certain territories, 1800s-Present. U.S. Geological Survey. https://doi.org/10.5066/P9ZXGFY3

[8] https://aqs.epa.gov/aqsweb/documents/data_api.html

[9] Homepage. (n.d.). Retrieved December 4, 2024, from https://www.health.state.ok.us/index.shtml

[10] Technical assistance document for the reporting of daily air quality | airnow. Gov. (n.d.). Retrieved December 4, 2024, from

https://www.airnow.gov/publications/air-quality-index/technical-assistance-document-for-reporting-the-daily-aqi/
[11] Josef Perktold, Skipper Seabold, Kevin Sheppard, ChadFulton, Kerby Shedden, jbrockmendel, j-grana6, Peter Quackenbush, Vincent Arel-Bundock, Wes McKinney, Ian Langmore, Bart Baker, Ralf Gommers, yogabonito, s-scherrer, Yauhen Zhurko, Matthew Brett, Enrico Giampieri, yl565, … Yaroslav Halchenko. (2024). Statsmodels/statsmodels: Release 0. 14. 2 [Computer software]. Zenodo. https://doi.org/10.5281/ZENODO.593847
[12] Ariton, L. (2024, February 2). A thorough introduction to holt-winters forecasting. Analytics Vidhya. https://medium.com/analytics-vidhya/a-thorough-introduction-to-holt-winters-forecasting-c21810b8c0e6
[13] Bellou, V., Gogali, A., & Kostikas, K. (2022). Asthma and tobacco smoking. Journal of Personalized Medicine, 12(8), 1231. https://doi.org/10.3390/jpm12081231
[14] Association, A. L. (n.d.). Tobacco trends brief: Overall tobacco trends brief. Retrieved December 4, 2024, from https://www.lung.org/research/trends-in-lung-disease/tobacco-trends-brief/overall-smoking-trends
[15] US EPA, O. (2013, February 22). Summary of the clean air act [Overviews and Factsheets]. https://www.epa.gov/laws-regulations/summary-clean-air-act

Data Sources

1. USGS Wildland Fire Historical Fire Data
2. US Environmental Protection Agency Air Quality Service Historical AQI Data
3. Oklahoma State Department of Health Statistics Survey Asthma Data