Analysis of Wildfire Smoke and Its Impact on Crime Rate in Cedar City, Utah

Jenny Wong

University of Washington

1 Introduction

The increasing prevalence of wildfires in the western United States, particularly during the summer months, has led to widespread smoke affecting numerous states. This phenomenon has been attributed to various factors, including climate change, United States Forestry policy, heightened public awareness, etc. Regardless of its origins, the impact of these wildfires is widespread, impacting health, tourism, property, and various other societal elements. This comprehensive analysis aims to lay the groundwork for further investigations.

The first part of this analysis focuses on an in-depth examination of the wildfires around Cedar City, Utah, utilizing data from the United States Geological Survey (USGS) presented in both ArcGIS and GeoJSON formats. I will also develop the smoke estimate and compare it with the AQI data obtained from the US Environmental Protection Agency (EPA).

Then I will examine the potential impact of wildfire smoke on the economic development and public health of Cedar City, Utah. More specifically, I will focus on the aspect of crime, as it provides insights into its social impact while informing public health and economic policies. The overall goal of this analysis is to help inform our city council and residents about the potential future impact of smoke on their communities.

2 Background

2.1 Crime as a Potential Factor

The relationship between pollution exposure and various health consequences has been established across numerous academic fields. There has been evidence showing that long-term air pollution exposure could lead to chronic disease, respiratory problems, and mortality outcomes (Guan et al., 2016). Besides, there is also more recent evidence suggesting that these similar mechanisms also have short-term effects on cognitive behaviors associated with violent activities and anxiety (Power et al., 2015).

To dive deeper into the relationship between wildfire pollution exposure and crime behaviors, I found this study that examined the effect of air pollution exposure on crime in the US over eight years. The study found a robust significant relationship between air pollutants and violent crimes, specifically assaults. This study concluded that a 10% increase in PM2.5 exposure was associated with a 0.14% increase in violent crimes, indicative of aggressive behavior. A similar increase in ozone exposure led to a 0.3% rise in violent crime and a 0.35% increase in assaults (Burkhardt et al., 2019).

Wildfire smoke, a complex mixture of gaseous and particulate matter, significantly contains pollutants such as PM2.5 and ozone. These pollutants, capable of deep lung and bloodstream penetration, are linked not only to significant health risks but also to an increase in violent behaviors. This connection suggests that the physical and psychological stress induced by

exposure to wildfire smoke can exacerbate aggressive tendencies, leading to a notable increase in crimes in areas affected by wildfire smoke (Power et al., 2015). Such findings emphasize the need for comprehensive community response plans that address both health and socio-economic concerns.

2.2 Data

To get an initial exploration of the wildfire and air quality situation of Cedar City, I used the wildfire dataset provided by USGS and AQI data from EPA and computed a smoke estimator for future model building. To validate this relationship in Cedar City, I decided to incorporate the same dataset as the study mentioned above. The United States Federal Bureau of Investigation (FBI) maintains a crime database named the National Incident-Based Reporting System (NIBRS), which provides incident-level crime reports across the US. The study uses all crime data from this database to fit the model as the response variable.

Wildfire Data

This dataset is provided by the USGS. The original data is in GeoJSON format, which is a standard format for encoding a variety of geographic data structures, and it contains more than 30 attributes. For this analysis, I am only going to use the following attributes from the dataset:

- Fire_Year: the year when the fire occurred
- GIS_Acres: the size of the fire in acres
- geometry: a list of geometric coordinates for calculating the average distance between the fire and Cedar City

Since I am only focusing on the wildfire that happened around Cedar City, I applied several filters to the dataset. The filtered data only contains the last 60 years of wildland fires (1963 - 2023), and it only takes fires that are within 1250 miles of Cedar City. Since the original dataset does not include the distance but only the location coordinates, I first converted the geometric coordinates to format EPSG:4326, then calculated the average distance in miles from the boundary of fire to Cedar City. This calculated distance is saved as an attribute in the dataset as “Avg_Distance”. To ensure the data readability and combination, I converted the filtered data to a CSV file. More specific code could be found in the notebook in the Part 1 folder of the repository.

Air Quality Index Data

This dataset is obtained from the US EPA. This dataset is obtained by making API requests for all the monitors within a 150-mile range from Cedar City, Utah. The data is in JSON format, and it has over 30 attributes for each object. Only the following attributes are used in this analysis:

- date_local: the date when the AQI is collected
- parameter: the specific type of the parameter of AQI
- aqi: the air quality index value

Since we only have the yearly data of wildfires, here I also aggregated the AQI values by year. To make the data combination with other datasets easier, I converted the data to a CSV file. More specific code could also be found in the notebook in the Part 1 folder of the repository.

NIBRS Crime Data

The crime data from NIBRS managed by the FBI is used for examining the wildfire impact on the economic development of Cedar City, Utah. This comprehensive dataset covered 87 cities in Utah, which includes our city, Cedar City, and it includes crime data from 1993 to 2022 for Utah. For this analysis, I am going to look at the data around Cedar City only. This dataset does not provide specific information regarding the license or terms of use.

Since the data is reported on a yearly basis, the data cleaning and preprocessing progress is complicated. More specific code can be found in the notebook in the Part 2 folder of the repository. After cleaning and aggregating the data, a CSV file is generated to be used in future analysis named "nibrs_crime_data.csv" which also can be found in the repository. The dataset contains the following attributes:

- data_year: the year when the incident happened
- offense_code: the code representing a specific type of offense
- offense_name: the name of the offense corresponding to the offense_code
- offense_count: the total count of the corresponding type of offense

Population Data

The population data of Cedar City, Utah is from the US Census Bureau. This data is used to calculate the crime rate with the aggregation data obtained above. This is a simple dataset with only two attributes:

- date: the year when the population was
- Cedar City Count_Person: the population

3 Methodology

3.1 Modeling Variables

3.1.1 Smoke Estimator

Using the wildfire data obtained from Part 1 Common Analysis, I built an initial version of the smoke estimation model. This model is a simple linear regression model consisting of 3 variables: GIS_Acres (the size of the fire), Avg_Distance (average distance between the fire and Cedar City), and Fire_Year (the year when the fire happened). The model is as follows:

$$SmokeEstimator = 0.4 * GIS\_Acres - 0.4 * Avg\_Distance - 10 * (2023 - Fire\_Year) + e$$

In this smoke estimator model, the size of the fire has a positive relationship with the estimated amount of smoke, while both distance and the year when the fire happened have a negative relationship with smoke estimates. That is, I expect the smoke estimates to increase as the size of the fire increases, and to decrease as the distance increases. I also incorporate the time factor, that the sooner the wildfire happened, the less the smoke estimator will be subtracted. A standard error is added to indicate the uncertainty in the model. After the calculation, I applied a scaler to scale the estimator with a range from 0 to 500.

3.1.2 AQI of Ozone and PM2.5

Among all the AQI data obtained from the EPA, I decided to use the AQI of Ozone and PM2.5 in my model, as the reference paper suggests that these are the two main air pollutants affecting violent crimes significantly. The following graph represents the smoke estimates of my calculation and all the AQI data available around Cedar City. Most of the AQI information was not available until the 1990s, which could lead to some inaccuracies for future analysis.

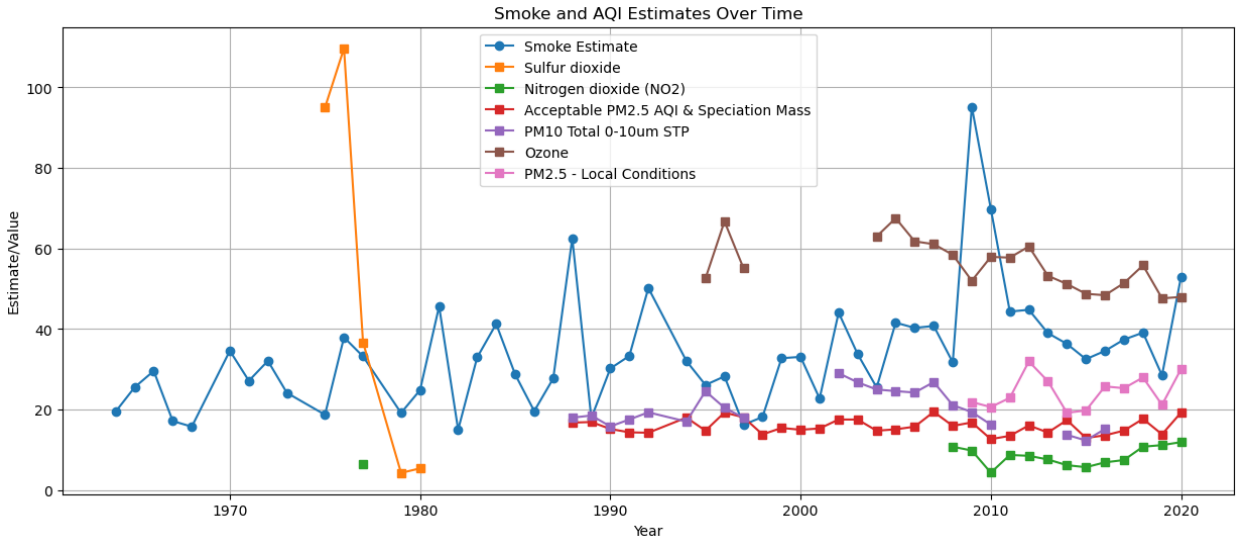

Figure 1. Smoke Estimate and AQI Estimates Over Time

Apparently, the data points for the smoke estimate and AQIs do not follow a consistent pattern, indicating variability in smoke and air quality, while the pattern of each AQI is somehow consistent.

3.1.3 Crime Rate

The other variable I used is the crime rate. This number is calculated using two datasets, crime data and population data. The crime rate is calculated as below:

$$CrimeRate = offense\_count/population$$

The graph below shows the trend of the smoke estimate, crime rate, and chosen AQI over time from 2004 to 2020. The reason to choose this time range is that even though the NIBRS gives crime data since 1993, there is not enough data for the region around Cedar City recorded. It is actually hard to see a clear trend here among the 4 variables. The trend tends to be consistent from 2017, but there is a steep increase in crime rate in 2020.

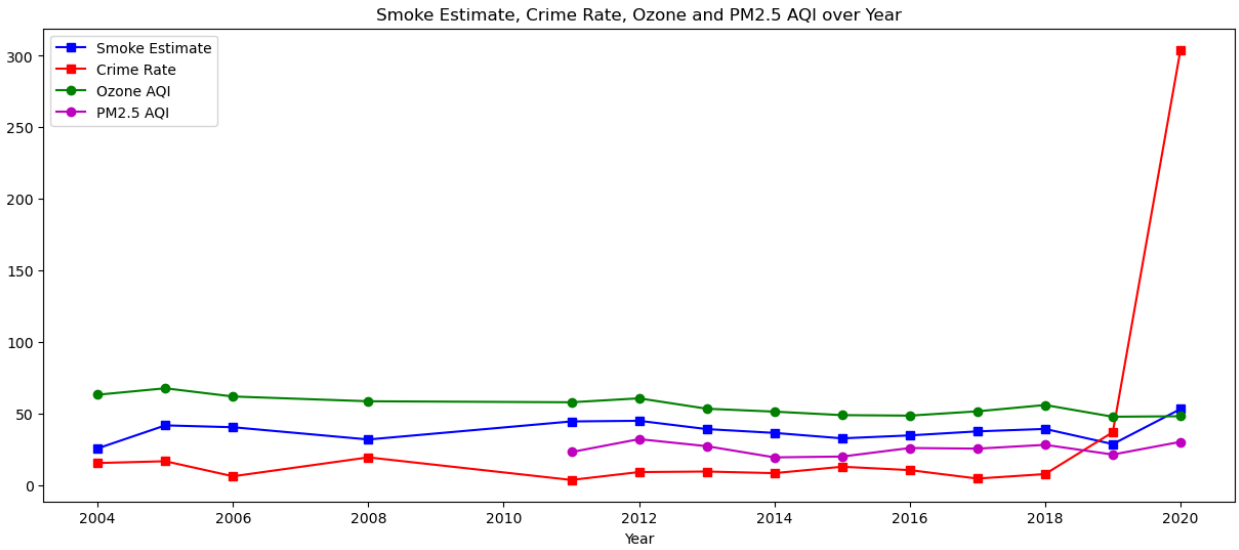

Figure 2. Smoke Estimate, Crime Rate, and AQI Estimates Over Time

3.2 Use of Models

In order to analyze the impact of wildfire smoke on the crime rate of Cedar City, I tried to build models to take the smoke estimator and AQIs to predict the crime rate and see how the models perform. The two models I decided to use are the simple linear regression model and the random forest model. All the code of the model implementation can be found in the repository.

3.2.1 Simple Linear Regression

The reason that I chose this model is because of its simplicity and the ease with which it can be implemented as a foundational baseline model. The model takes three variables, which are the smoke estimate, AQI of Ozone, and AQI of PM2.5. To analyze the performance of this model, I used two metrics, which are the mean squared error (MSE) and R-squared (R^2).

3.2.2 Random Forest

The random forest model is known for its ability to handle non-linear relationships and a diverse array of data types. The model's proficiency in identifying significant predictors from multiple variables is especially important in my multivariate contexts. Also, the capability of performing time series data makes it a good fit since all the data I used for this analysis is time series data. Similar to the simple linear regression, I used the same set of variables here. Besides, I created lag values for all variables with time interval = 1. To compare the performance of the random forest and simple linear regression models, I matched the metrics to the MSE and R^2 as well.

4 Findings

The results of the two models are as follows:

Metrics	Simple Linear Regression	Random Forest
R^2	-18.01	-0.17
MSE	3.60e-05	0.0001

Table 1. Model Results for Simple Linear Regression and Random Forest

Since I did not get a good linear correlation trend among my variables as Figure 2 shows, the poor linear regression model result is to be expected. Even though the MSE is very small, that is because of the low value of the crime rate. The R^2 values show that both models are not performing well.

5 Discussion and Implications

5.1 Reflection on the Findings

The analysis of the impact of wildfire smoke on crime rates in Cedar City, Utah, reveals that while the effect is present, is it not overwhelmingly significant. This suggests that immediate actions may not be necessary, but it still highlights the importance of proactive and forward-thinking planning. The findings are interesting as they add a new dimension to the indirect effects of environmental factors on social issues, even if the effects are not immediately pronounced.

5.2 Recommendations for City Council and Residents

Given the findings, the recommendations for the city council and the residents of Cedar City are as follows:

1. ***Long-Term Environmental Planning:*** Develop long-term strategies to prevent potential air pollution and manage the risks associated with wildfire smoke. Since Cedar City is beside Zion National Park, it is still vulnerable to wildfires. This could include collaboration with forestry and environmental protection agencies.
2. ***Community Awareness Program:*** Implement community education programs to raise awareness about the potential impacts of environmental factors on public health and safety, albeit the current impact is moderate.
3. ***Continued Monitoring and Research:*** Maintain ongoing monitoring and research efforts to understand the evolving nature of the relationship between environmental factors like wildfire smoke and social issues, including crime.

5.3 Human-Centered Data Science Approach

The approach to this analysis was deeply rooted in human-centered data science principles. The decision-making process was guided not just by the data but by a commitment to understanding the broader human context of the findings. Even though the direct impact of wildfire smoke on

crime rates in Cedar City is not substantial, the analysis and subsequent recommendations were formulated with an eye toward the future and the potential for changing conditions. This perspective ensures that the city's plans are resilient, adaptable, and most importantly, centered around the well-being and safety of its residents.

6 Limitations

As presented in the Findings section, both models are not performing well. The model results are far from the conclusion presented by the reference paper. Thus, I came up with some potential limitations to my model.

6.1 Other Possible Factors

6.1.1 Health and Mental Health

One limitation would be that there could be other possible factors that could have a more significant impact on the crime rate than wildfire smoke. For example, it could be mental health issues induced by COVID-19, as we may observe from Figure 2 that there is a steep increase in crime rate in 2020. The pandemic has led to a rise in mental health problems such as anxiety, depression, and stress-related disorders. The psychological impact of prolonged isolation can be profound. The mental health challenges may in some cases, lead to increased impulsivity, substance abuse, and domestic violence, all of which are likely to lead to criminal behavior.

6.1.2 Economy

There also could be other economic factors impacting the crime rate. For instance, income inequality and the unemployment rate could be a major factor in property crimes (Raphael & Winter‐Ebmer, 2001). Economic hardships and disparities can lead to increased financial strain and relative deprivation, potentially pushing individuals toward criminal activities.

6.1.3 Drug and Alcohol Abuse

The illegal drug trade itself is a significant contributor to crime. Drug trafficking, production, and distribution are inherently illegal activities that often involve other criminal acts, such as violence, use of weapons, and involvement with organized crime groups. It is noticeable that drug-related crime has increased, especially in recent years around Cedar City, and the ratio of drug-related crimes over the total number of crimes reached over 30% in 2019.

6.2 Lack of Data Support

Another limitation would be the lack of data support. The research that I referenced for choosing crime as my factor takes the crime and wildfire data for the entire US to examine the relationship. The data I used in this analysis is only for a small city with a low population. This crime and wildfire data are more likely to be biased, thus it could be harder to conclude a solid result compared to the reference study.

7 Conclusion

This report aimed to explore the impact of wildfire smoke on economic development and crime rates in Cedar City, Utah. The primary research question focused on understanding how exposure

to wildfire smoke, particularly pollutants like PM2.5 and ozone, correlates with changes in crime rates. Our analysis, using data from the United States Geological Survey, the US Environmental Protection Agency, and the FBI's National Incident-Based Reporting System, revealed that while there is a relationship between wildfire smoke and crime according to a rigorous reference study, this link is not as significant as initially hypothesized in Cedar City, Utah. The linear regression and random forest models used in the study showed poor correlation, indicating that other factors might play a more substantial role in influencing crime rates.

This study contributes to the field of human-centered data science by emphasizing the importance of considering a range of human factors and environmental influences when analyzing complex societal issues. The findings underscore the necessity of holistic approaches that integrate environmental, health, and socio-economic data to fully understand and address community challenges. The decision to focus on crime as a potential impact of wildfire smoke was driven by the goal of understanding not just the physical but also the social repercussions of environmental changes. This approach reflects the core principles of human-centered data science, where data analysis is deeply intertwined with human welfare and societal contexts.

The outcomes of this study highlight the need for further research to explore other contributing factors to crime rates, such as economic conditions, mental health issues, and substance abuse. It also suggests that policymakers should consider a broader range of influences when developing strategies to enhance public safety and health. While the direct impact of wildfire smoke on crime may not be as pronounced, this analysis provides a foundation for more comprehensive studies and informs city planning and public health strategies with a nuanced understanding of environmental and social interconnections.

References

Burkhardt, J., Bayham, J., Wilson, A., Carter, E., Berman, J. D., O'Dell, K., Ford, B., Fischer, E. V., & Pierce, J. R. (2019). The effect of pollution on crime: Evidence from data on Particulate Matter and ozone. *Journal of Environmental Economics and Management*, *98*, 102267. https://doi.org/10.1016/j.jeem.2019.102267

Guan, W.-J., Zheng, X.-Y., Chung, K. F., & Zhong, N.-S. (2016). Impact of air pollution on the burden of chronic respiratory diseases in China: Time for urgent action. *The Lancet*, *388*(10054), 1939–1951. https://doi.org/10.1016/s0140-6736(16)31597-5

Power, M. C., Kioumourtzoglou, M.-A., Hart, J. E., Okereke, O. I., Laden, F., & Weisskopf, M. G. (2015). The relation between past exposure to fine particulate air pollution and prevalent anxiety: Observational cohort study. *BMJ*. https://doi.org/10.1136/bmj.h1111

Raphael, S., & Winter‐Ebmer, R. (2001). Identifying the effect of unemployment on crime. *The Journal of Law and Economics*, *44*(1), 259–283. https://doi.org/10.1086/320275

Data Sources

USGS Wildfire Data: https://www.sciencebase.gov/catalog/item/61aa537dd34eb622f699df81

US EPA AQI Data: https://aqs.epa.gov/aqsweb/documents/data_api.html#monitors

NIBRS Crime Data: https://cde.ucr.cjis.gov/LATEST/webapp/#/pages/downloads

Census Bureau Population Data: https://www2.census.gov/programs-surveys/popest/tables