Peter Mohammadi
Data 512 Project Report
12/11/2023

The Health Implications of Wildfire Smoke in Larimer County, CO

Introduction

Historically the residents of Larimer County have had relatively low amounts of wildfire smoke exposure. This is due to the large majority of wildfires occurring hundreds of miles away. That was the case until 2020, when one of the largest fires in Colorado history, The Cameron Peak Fire, occurred less than 50 miles away from Larimer. This fire burned more than 200,000 acres of land and exposed the population of Larimer to heavy amounts of particulate matter. This analysis focuses on the health effects of wildfire smoke in Larimer County and the Front Range of Colorado. By learning more about how wildfire smoke affects the residents of Larimer, we can more effectively protect the population in case of future major wildfires. One of the most fascinating aspects of this occurrence is the fact that people are still looking to buy homes and move into Larimer, even after this major natural disaster. Thus this analysis will also dive into historical home prices, as they relate to wildfire smoke exposure.

Background/Related Work

Over the past decade, there have been numerous papers written on the health effects of wildfire smoke, and a handful written specifically for the Front Range of Colorado. Researchers at CSU published a paper focusing on hospitalizations in said area, which was the "first study in the United States that links acute wildfire smoke (WFS) exposure with death" (Magzamen 1). This paper led me to a large amount of public resources related to Colorado mortality data. It also provided further knowledge on historical hospitalization data, which is not open source, yet asserts that wildfire smoke does induce more asthma hospitalizations (Magzamen 4). Yet this paper only used data from 2010 - 2015, whereas my analysis uses data from 2010 - 2019.

In this paper, I have two hypotheses: 1) There is a positive correlation between wildfire smoke exposure and respiratory mortality. Meaning the more severe wildfire smoke, the more mortalities. 2) Wildfire smoke exposure does not affect the price of homes. These two questions will sufficiently argue that even though wildfire smoke is dangerous, United States residents do not take it seriously.

A handful of datasets were used in this analysis other than the AQI estimates and the historical fire counts. One of these datasets was from a researcher named Jason Vargo, who created: *Time Series of Potential US Wildland Fire Smoke Exposures* dataset. This dataset contains county-level wildfire smoke exposure metrics for the United States from 2010 - 2019. This dataset was used to perform an analysis of house prices over the entire United States. The other dataset used is data from the US Census. This data is yearly median house prices per County from 2010 - 2019.

Methodology

House Prices and Weighted Smoke

I used the weighted smoke dataset from Jason Vargo (for the house price analysis specifically) instead of the EPA's AQI data for two reasons 1. Jason Vargo's dataset accounts for all of the United States and 2. Both the EPA and Vargo sourced the smoke data from the same place. Vargo's smoke data originally came in the form of three values per county (and per year): light, medium, and heavy, all on the binary scale. Light means PM2.5 0-10, medium means 10-21, and heavy is 22+. I merged these values into a 'weighted' smoke value, to get a single number to analyze, and I was hoping it would be easier to understand. Weighted smoke was calculated like so: weighted = (medium * 2) + (heavy * 3) + light) / 3. I also created a new column in the historical median US house prices dataset called "net_price_change". This value represents the change in price of a single county's average house cost year-to-year. I created this value because it better represents the change in prices, because now you can see if it was positive or negative. The merging of the two datasets was based on the fips codes and the year, and then I simply plotted the values together. On top of plotting the values, I also calculated the correlation between net_price_change and the weighted smoke, to get a single value to represent the relationship.

Estimated Smoke and Mortality

My estimated smoke value was based upon the summarization of the area of each fire divided by the square root of the distance from Larimer County. All the fires are within a 1250-mile radius and are marked as 'wildfires' in the USGS dataset. I could have included code to also use other types of wildfires, but I did not because this analysis is specific to wildfires, purposeful, human-made fires. I used an ARIMA (AutoRegressive Integrated Moving Average) model to forecast smoke severity over 10 years. ARIMA is a statistical method often used in time series forecasting, it combines autoregressive and moving average models. It also incorporates differencing to make the time series stationary. I chose an ARIMA 5,1,1 meaning: autoregressive with order 5, thus the current value is based upon the last 5. Integrated order 1 means that the data needs to be differenced once to achieve stationarity. And finally moving average of degree 1 meaning I am smoothing out the noise in the data. I believe this method best models times series data that shows patterns of non-stationarity (like fire season) and 'random' fluctuations. I then used these predictions, merged them with the historical mortality data, and incorporated them all into a regression model to predict future respiratory mortality. This respiratory mortality data also included other mortalities like CVD, despair, and all combined mortalities, which were also used. I chose a linear regression model for the main reason that it is a simple model that works well with time series data.

Findings

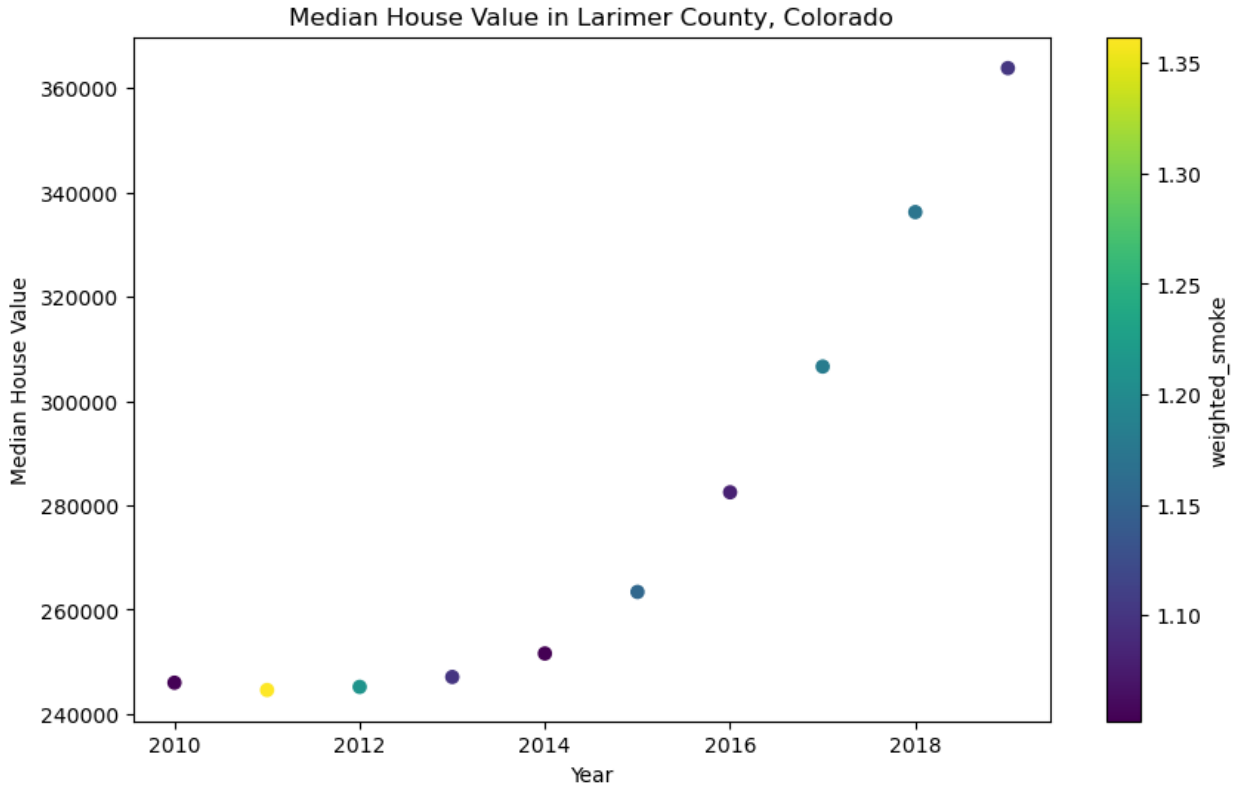

Figure 1. Median House Value in Larimer County

As shown in Figure 1, the price of homes in Larimer County has been increasing drastically for around a decade. This means that more and more people want to move and live in Larimer County, despite 2020 being one of the worst fire years to date. This graph makes sense when you also look at the entire state of Colorado when it relates to weighted wildfire smoke and house prices.

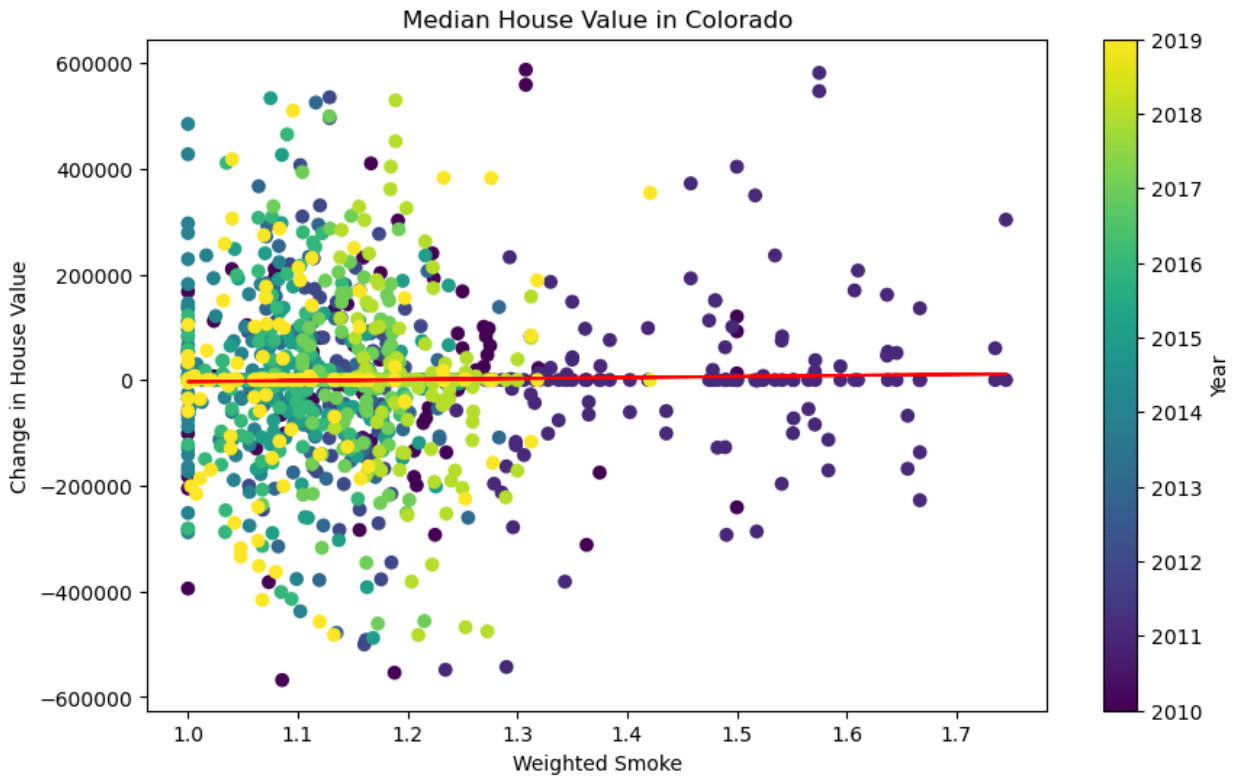

Figure 2. Median Change in House Value in Colorado

Figure 2 backs up our claims in Figure 1, showing that as wildfire smoke gets worse and worse, the price of homes does not change. The correlation coefficient between the change in house value and weighted smoke was shown to be: 0.018, not significant. Comparing the average estimated smoke levels for Larimer County in the years 2010 to 2015 (0.14) against 2016 to 2021 (0.85), we can see the estimated smoke values are over **six times** worse than in 2010 - 2015. Using estimated smoke values from the years 2020 - 2030, we can help predict estimated respiratory mortalities in the Front Range of Colorado.

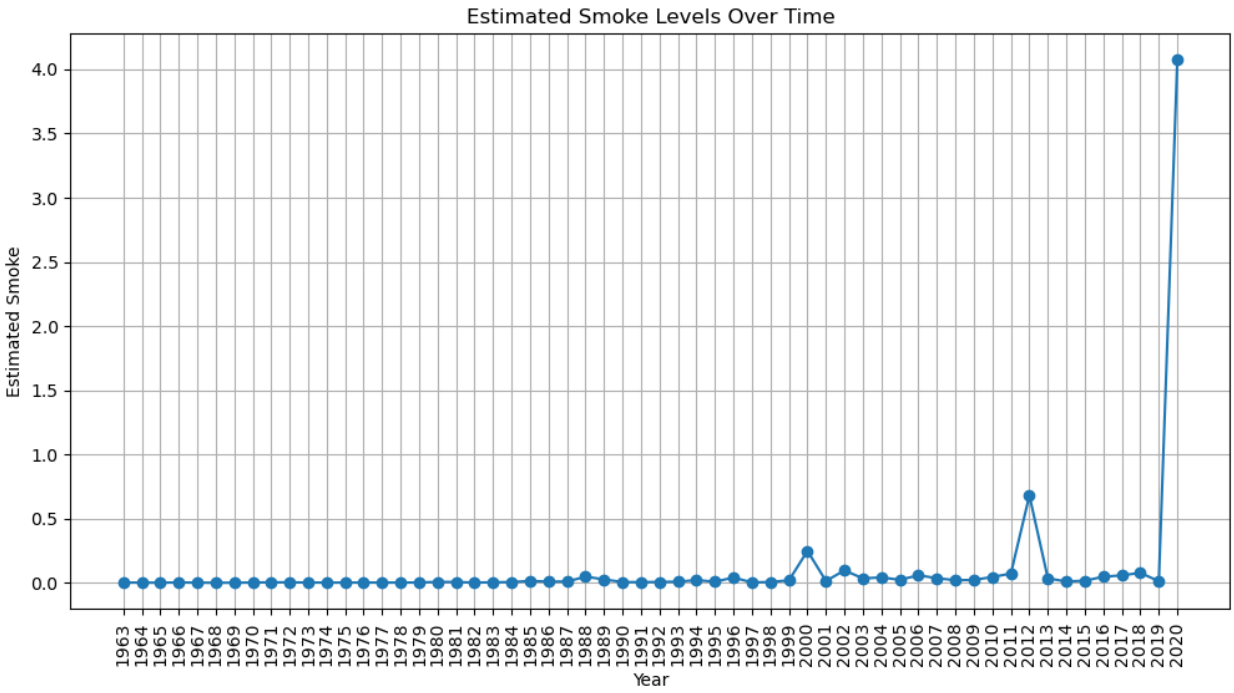

Figure 3. Estimated Smoke Levels in Larimer County over Time

We can see 2020 was a very bad year for smoke, due to the Cameron Peak Fire, but we can also see in 2000 and 2012 were some bad years too. This shows how these 'bad' years are becoming more and more often, thus wildfire smoke is getting worse in Larimer County.

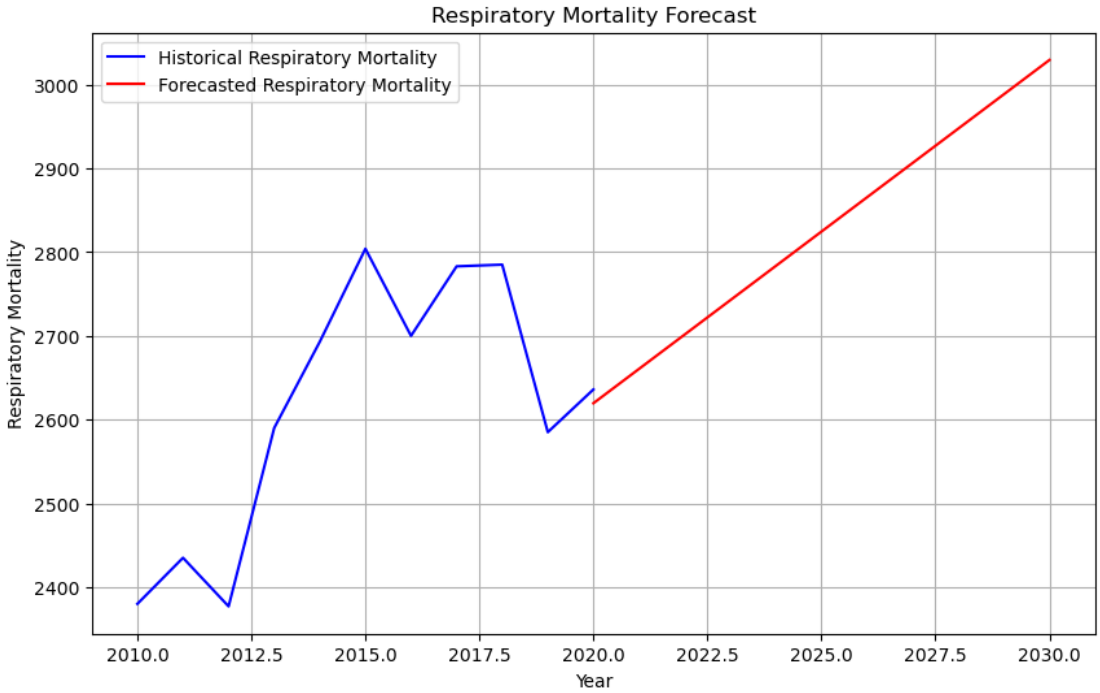

Figure 4. Forecasted Respitory Mortality in the Front Range of Colorado

In Figure 4, we can see that respiratory mortality was already increasing in the Front Range of Colorado (in which Larimer is located). And now that we can show 'future' respiratory mortality, we can see the mortalities continue to increase. These predicted mortalities are not visualized to give concrete estimates on how many people I think may die, but to visualize that the trend is pointing upwards. I would take these estimated values with a grain of salt. Predicting any type of data, especially one as sporadic as wildfire data is very complicated. Please locate the **Limitations** section for further insight into the shortcomings of this analysis.

Discussion/Implications

As mentioned in the background section, many recent papers have connected mortalities and hospitalizations with increased PM2.5 counts, but there has not been a specific study made for Larimer County. Meaning, that we can confirm (at the least) that smoke levels are getting worse, meaning the population is at risk of this smoke. I propose a very easy, actionable change that could protect a lot of people: change recommended AQI levels. On the Larimer County website (https://www.larimer.gov/naturalresources/trails/air-quality) they post daily AQI levels, I propose we change 'good' to 0 - 40, 'moderate' to 41 - 80, and anything above 80 to 'not recommended'. This could protect more people from wildfire smoke by recommending they go outside a little less.

I followed human-centered data science principles in this analysis. I did not present any information was not completely confident in. That is why I am not stating how many people I think may get asthma or even die, I simply know the AQI is getting worse year by year. I also only chose data from sources that I specifically trust and believe were created fairly without any sampling bias. Finally, I did not use models I did not see applied in other papers, because I wanted to create a project that was both easy to understand and follow in the footsteps of real scientists, thus I only used regression models.

Limitations

Smoke Estimation

My yearly smoke estimation is the summation of (Area_Of_Fire / Distance_From_Loveland ^ 2) per year. Unfortunately, the data used does not contain any wind data, meaning we do not know if the smoke does reach Larimer County, I am simply assuming it does based on the distance. Furthermore, this dataset combines fires if they occurred near the same time and place, thus the merging of these fires could result in the loss of data on smaller fires.

House Prices vs Weighted Smoke Visualizations

The weighted smoke data (via Jason Vargo) was originally day-by-day, but due to the Census collecting house prices yearly, the smoke data needed to be averaged by year. This is an issue

because smoke levels are normally low throughout the year, but are then very high for a small portion, meaning a year average of smoke levels does not represent the data well. Furthermore, the weighted smoke data is only from 2010 to 2019, thus we do not have a lot of historical data.

Mortality Data

My mortality data contains mortalities for the entirety of the Front Range of Colorado, which is the collection of a dozen different (large) counties in Colorado. Thus Loveland in Larimer County is a small subsection of this data. Furthermore, this mortality data only goes back to 2010, meaning I have a lack of historical data.

Conclusion

In this analysis we have shown the more severe wildfire smoke becomes, the more respiratory mortalities for the affected populations. This is shown by the increase in respiratory mortalities in the Front Range of Colorado and the increasing wildfire smoke estimation values. We have also shown that the price of homes does not change, even when the weighted smoke in said area is high. This is especially true in Larimer County, where both the smoke severity and house prices have been increasing. These two hypotheses connect to the main idea that people do not take wildfire smoke seriously, even though severe health effects could be caused by the existence of it. I proposed to the City Council of Larimer County to change the AQI recommendations on their 'air quality' website page to heighten the requirements for 'good', 'medium', and 'not recommended' days.

Throughout every step of the analysis, I have given directions on how to repeat all my steps. From locating and loading the required data, programming the data cleaning and merging, creating the visualizations, and finally the analysis. I have provided a great deal of detail on how and why I chose each method, so much detail that even the 'non-technical' would be able to follow the analysis. I have provided sources for all my data and gave credit where it is due. I have also highlighted weaknesses in my analysis. This is all to make the job of a reviewer, researcher, or even a colleague/friend to understand my analysis. This is the crux of human-centered data science, to create data science projects with the goal of understandability, reproducibility, and honesty at the forefront of the analysis.

References

Magzamen, S., Gan, R. W., Liu, J., O'Dell, K., Ford, B., Berg, K., Bol, K., Wilson, A., Fischer, E. V., & Pierce, J. R. (2021). Differential cardiopulmonary health impacts of local and long‐range transport of Wildfire Smoke. *GeoHealth*, *5*(3). https://doi.org/10.1029/2020gh000330

Martenies, S. E., Wilson, A., Hoskovec, L., Bol, K. A., Burket, T. L., Podewils, L. J., & Magzamen, S. (2023). The COVID-19-wildfire smoke paradox: Reduced risk of all-cause mortality due to wildfire smoke in Colorado during the first year of the COVID-19 pandemic. *Environmental Research*, *225*, 115591. https://doi.org/10.1016/j.envres.2023.115591

Vargo, Jason. (2020). Time Series of Potential US Wildland Fire Smoke Exposures. Frontiers in Public Health. 8. 126. 10.3389/fpubh.2020.00126.

Data Sources

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

US Census Data for Median Household Values:
https://www.census.gov/data/developers/data-sets.html

Yearly mortality data for the Front Range of Colorado:
https://www.sciencedirect.com/science/article/pii/S0013935123003833

County-wise wildfire smoke levels in the US:
https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/CTWGWE

US EPA AQI levels: https://aqs.epa.gov/aqsweb/documents/data_api.html

US FIPS CODES: https://github.com/kjhealy/fips-codes