Wildfire Smoke Project – Final Report

Introduction

Wildfires are a growing problem in the U.S. These wildfires emit large amounts of smoke, which causes significant direct effects on the population, from wheezing and trouble breathing to strokes, heart attacks, and deaths. While these direct health effects are already well-documented, an increasing area of study looks to shine a light on the indirect effects of wildfire smoke, such as on various socio-economic issues.

One of these socioeconomic issues is labor profits/wages across the United States. Research into the association between wildfires and labor profits was spearheaded by a group of researchers at the University of Illinois at Urbana-Champaign (UIUC). After looking into the association between the two at the national level, and accounting for the noise created by other factors, they concluded that there was a significant link between the two, with more wildfire smoke being associated with a decrease in wages. The goal of this project is to look at this relationship between wildfire smoke and labor profits and examine whether it can be localized for a specific town in the United States: Clovis, New Mexico.

Clovis is a small town on New Mexico's border with Texas. It can often face a varying amount of annual wildfire smoke coverage, which can be significant for years where wildfires rise in the surrounding areas. This project will draw from a variety of data sources to conduct an analysis on the relationship between wildfire smoke and labor profits in Clovis, and seeks to establish whether an association seems to exist (and how strong it is, if it does).

Background/Related Work

As explained in the introduction, the background for this project mostly comes from a prior study from researchers at the University of illinois at Urbana-Champaign, or UIUC. Published in 2021 by Mark Borgschulte, David Molitor and Eric Yongchen Zou, this piece looks into whether a relationship between wildfire smoke coverage and labor profits can be found on a national scale. The piece used data for all of the counties across the U.S. and aggregated its findings, while trying to account for other variables that could contribute to noise in the association between the two.

For labor profit data on all U.S. counties, the paper mainly used the U.S. Census Bureau's Quarterly Workforce Indicator (QWI) dataset, which offers county-level labor profit data. To track wildfire smoke, the paper used two different datasets - data from the National Oceanic and Atmospheric Administration's Hazard Mapping System (HMS), which uses satellite imaging to track how much each U.S. county is covered in wildfire smoke, as well as data from the EPA's Air Quality System API - specifically particulate AQI (Air Quality Index) data - which helps tracks particulate matter in the air, such as that emitted by wildfire smoke, as detected by sensors the EPA has placed across the nation.

After accounting for confounding factors and calculating the effect of a "smoke day", IE a day in which a specific county was mostly covered by wildfire smore the researchers found that, on average, each additional smoke day a U.S. county experienced was associated with a 0.097% loss in per-capita labor earnings. This indicates that wildfire smoke can have a large cumulative effect on any level of the American economy. In detailing the effects on the national economy, the researchers used the example of 2010, when 160 million workers made a total of $6.4 trillion, coming out to $40,000 per capita. After finding that the average U.S. county

experienced 20.2 smoke days that year, and applying the 0.097% decrease in per capita labor profits, the researchers showed that this indicated that wildfire smoke was associated with a loss of $125.4 billion in labor profits nationwide that year. This clearly shows that wildfire smoke is associated with a huge weight on the national economy.

Methodology

This section will have two subsections that cover the scope of the whole project. The first subsection will cover the preliminary work completed, which includes a common analysis phase that focused on understanding wildfire smoke in the vicinity of Clovis without considering its interaction with additional socioeconomic factors (also referred to as “Part 1”/”Common Analysis”), and then the development of an extension plan, where I looked into possible avenues of additional research into the effects of wildfire smoke on Clovis and decided to look into how it connects to labor profits (also referred to as “Part 2”/“Extension Plan”). The second subsection will cover the main analysis into the association between wildfire smoke and labor profits in Clovis; this largely consists of the implementation of what was proposed in the Extension Plan and how it was used, including the development of a presentation directed towards the city council of Clovis to initiate a discussion on the relation between wildfire smoke and labor profits in Clovis (“Part 3”/“City Council Presentation”) and the final work done to present in this report (“Part 4”/“Final report work”).

In describing the Common Analysis and Extension Plan, the preliminary work section will use a large part of the language contained in the Extension Plan itself, which already contained an overview of the work done in Part 1 in addition to documenting the work done in Part 2. However, it will be edited for clarity and expanded to give more detail on the thought process behind the two parts, as well as better place them in the context of the whole project.

Preliminary Work

In Part 1 of the Project, my colleagues and I conducted an analysis of wildfire smoke data on different target cities (my target city being Clovis, New Mexico). In this part of the project, we took in a large GeoJSON dataset from the United States Geological Survey (USGS). USGS's dataset had a list of every wildfire in the US that they had a record of, stretching back to 1835.

For Part 1 of our work, we focused in on the wildfires from the last 60 years (IE from 1963 onward) that fell within 1250 miles of our target city, and were tasked with using the attributes of the wildfires we were given in this USGS dataset to come up with a measure that would generally indicate how much impact the wildfire smoke for that year would cause on our city, relative to other years. We then compared this initial measure to data from the EPA's Air Quality System (AQS) API, which would give us an indication of what the EPA's systems detected in terms of air pollution for the area. As for all of Part 1, our initial measure was not intended to relate wildfire smoke to any other measure or socioeconomic indicator; it was purely meant to be a general measure that could indicate which years had more or less smoke. For my smoke impact measure, I theorized that smoke impact would be greater for larger wildfires, for wildfires that were closer to my target city, and for a series of smaller wildfires rather than one large wildfire, even if they added up to be same size (with the idea that the constant smoke emitted by the smaller fires is less likely to have its impact diminished by possible environmental conditions that are not accounted for in the USGS dataset, like gusts of wind that could blow away some or all of a fire's smoke).

After completing Part 1, I developed my extension plan by considering the way I would expand on my Common Analysis under ideal circumstances. Specifically, because my wildfire smoke indicator showed little to no correlation for the closest available AQI (Air Quality Index,

IE specific readings from the EPA's AQS) data, the first idea I proposed was to look into ways that my indicator model could be modified to more directly predict AQI readings in the future. I then began looking for past research into socioeconomic factors that both interested me and had a potential association with wildfire smoke. I was lucky to quickly find the work of the UIUC researchers mentioned in the "Background/Related work" section of this paper. After reading through their paper, I decided that looking into whether the association they found could be localized to Clovis was an interesting and engaging way to continue this project, and made the rest of my proposed methodologies for my extension plan.

The first part of this was finding data from the Census Bureau's Quarterly Workforce Indicators (which among other statistics tracks how much workers in an area are making), specifically for Curry County, New Mexico which is mostly just the city Clovis (according to demographic data, approximately 38,000 of Curry's 48,000 residents live in Clovis). I hoped to then use that data to see labor profits for Clovis and examine how their association with wildfire smoke, both through the lens of the data from the AQS and a second smoke dataset used by the UIUC researchers- the Smoke Product data from the National Oceanic and Atmospheric Administration's Hazard Mapping System (HMS), which tracks when U.S. counties are covered in varying levels of wildfire smoke (again, using Curry as a stand-in for Clovis). This dataset was used by the researchers to define one of their central metrics, which was the number of "Smoke Days" for a county, IE the number of days that a county was completely covered in Smoke. By looking for correlations between Curry's QWI data and Curry's AQI data and HMS data, I hoped to find whether the aforementioned association could be localized for Clovis.

Finally, I proposed that, if time permitted, a complete replication of the UIUC researchers' methodologies could be made to establish how much of a reduction in labor profits

an average smoke day in Clovis was associated with. A national version of this statistic was the central finding of the UIUC research paper (as explained in the "Background/Related work" section). While a localized version of this statistic would be an ideal way to cap this project, I was able to recognize at the time that time constraints would likely preclude me from being able to replicate their work, as they stated in their ReadMe that the full runtime of their code, on a computer with similar specs to mine, was around 24 hours, without accounting for modifying the code to only look at local data. Thus, I explained in my Extension Plan that this last proposal would, possibly, be unable to happen.

Main Analysis

After my Extension Plan was made and I moved to begin making analyses for Part 3, I immediately noticed that the quality of the AQS data available for Clovis was a much larger problem than I anticipated while working on the Plan. Specifically, it quickly became apparent that it would be difficult to obtain sufficiently high quality data on the Particulate AQI (Air Quality Index) for this city (this is further discussed in the Limitations section of this report), due to a lack of reasonably close EPA sensors. When looking over my work for Part 1, this became a problem when considering the association between my wildfire smoke indicator and the closest available AQS data - graphing the two together showed no clear correlation, and when theorizing ways to improve my model so that it would directly predict AQI readings, it was difficult to say whether my initial indicator and AQI readings' lack of noticeable association was due to wildfire smoke being only one possible influence on AQI readings, or because the AQS sensors were too far away to give any meaningful insight into Clovis.

Realizing that I could not meaningfully trust the AQS data I had available, I looked for other datasets that would serve as a substitute for readings from the far-away AQS Sensors. After

not finding a reasonable substitute, I decided to mostly use the HMS data I had for Clovis instead of trying to use a mix of the two, so that I could then work on Part 3 and present my findings to the Clovis city council. I realized that I was quite lucky to have also considered the HMS data during my Extension Plan, or else I would not have been able to meaningfully implement any of my Extension Plan's proposals.

Unfortunately, when considering what I could present to the city council, I realized that I would have to scrap my plans to improve my model to predict AQI readings for Clovis. Instead, I focused on whether I could find an association between wildfire smoke, both through the HMS's Smoke Days as well as the amount of wildfires within range of Clovis, and the yearly labor profits for the town. I go into more detail on how I accomplished this below.

First, I started by looking at the QWI data by itself. I proceeded to look at how it changed over time, and graphed a short time series of it to present an overview of the dataset to the city council in the City Council Presentation. The graph showed that while labor profits are clearly growing over time, the rate of growth has greatly varied. I then did the same for the HMS data, to show the number of smoke days varies year-to-year in Clovis, finding that the yearly amount greatly varies.

After doing this, I calculated the yearly increase in labor profits for each year I had data for (up from the previous year) and graphed this against both the number of wildfires within 1250 miles of Clovis for each corresponding year, as well as the number of smoke days for each corresponding year.

After this yielded no significant correlation or association between the two, I considered other forms a relationship could take. I considered that, maybe, the effect of wildfire smoke on labor profits could be confined to the workers who are most vulnerable to the health effects of

wildfire smoke mentioned in the introduction. After remembering that the elderly are generally associated with higher vulnerability to these same symptoms, I filtered out the QWI data to only include labor profits for workers with an age of 65 years or older. This immediately produced a closer association with both the annual number of wildfires within 1250 miles of Clovis, as well as the number of smoke days Clovis experienced, with a higher amount of both being associated with a drop in the rate of change for elderly workers' labor profits. I presented all of the work up to this point in my City Council Presentation.

After presenting to the city council, I had a discussion with a colleague who had a chance to attend my presentation. He mentioned that he suspected my graphs on the labor profits of workers aged 65 and older, while showing a stronger correlation with the wildfire smoke metrics than the labor profit data for all of Curry County, still looked sparse around their trendlines and seemed potentially unreliable. This conversation inspired me to evaluate the correlation between the number of smoke days/the number of wildfires in range and the labor profits for elderly workers aged 65 and older.

I decided that the best way to do this was to take both correlations and submit them to a Pearson test, which would both give me a p-value associated with the correlation. Using a standard confidence level of 95%, I can use the p-value, if it ends up being less than 0.05, to reject the idea that there is no correlation between labor profits for workers 65 years old or older in Curry County and either the number of annual smoke days Curry experiences or the number of wildfires within 1250 miles of Clovis (IE the null hypothesis, which represents a case of no correlation). If I am able to reject my null hypothesis, I can also construct a 95% confidence interval for the true value of the correlation coefficient for both relationships, which gives me a range of likely values for the true correlation coefficient.

Findings

After taking the steps outlined in the "Methodology" section and reviewing the results, I can conclude that I was not able to find a meaningful relationship between wildfire smoke and the labor profits for Clovis, New Mexico. No decisive evidence was found that could meaningfully indicate that the two are related, even if the possible existence of a relationship was also not decisively disproven.

I place special emphasis on the results of the Pearson test described at the end of the "Methodology" section. Both of the p-values for the two possible correlations were between 0.4 and 0.5, much higher than the 0.05 that would have been needed to reject the null hypothesis at a confidence level of 95%. Additionally, a hypothetical confidence interval for both correlations yielded a range that roughly went from -0.7 to 0.35 for both, including the possibilities of strong negative correlation, moderate positive correlation, or no correlation at all. In other words, we were unable to eliminate any possibility.

Therefore, this project was unfortunately unable to uncover any interesting trends or correlations related to wildfire smoke in Clovis, New Mexico.

Discussion/Implications

The potential for discussions this project can inspire are limited by the fault of conclusive findings. However, the project did uncover potential future avenues of researchers for finding an association in the future. Specifically, filtering for elderly workers in Curry did seem to strengthen the correlation between metrics of wildfire smoke and labor profits, even if this correlation ended up not being significant. Looking for ways to combine this approach with

datasets that address the issues listed in the "Limitations" section could uncover an association that was unable to be found here.

Ultimately, however, it is quite possible that wildfires are simply not a large problem for Clovis, New Mexico. While not explicitly proven, this idea is certainly supported by the results of this project.

Limitations

One of the biggest limitations in trying to research the association between wildfire smoke and labor profits for Clovis, New Mexico is the lack of good air quality data for Clovis. For most of the U.S., one can find good air quality data using the API for the EPA's Air Quality System (AQS). The EPA's Air Quality System has a large number of sensors nearby that read in Air Quality Index (AQI) scores associated with various features of the atmosphere. Wildfire smoke pollution is usually measured by the AQS's readings for PM2.5 AQI, or the AQI from small air particles with a diameter of less than 2.5 micrometers. PM2.5 AQI data was used in the UIUC paper to measure wildfire-related pollution, and it was part of AQS data my colleagues and I used when conducting a Common Analysis on wildfire smoke in proximity to our target cities.

The issue when trying to analyze the effects of wildfire smoke on Clovis, unfortunately, lies with the proximity (or lack thereof) of the nearest AQS sensors. The way to find the nearest sensors using the AQS data is to ask the API to list all of the sensors within a range of acceptable latitude-longitude values. By centering such a range on the latitude and longitude coordinates for a particular city (again, Clovis in our case) one can gradually increase the latitude and longitude of the range until the API returns the location for a sensor, which would naturally be the nearest

one. After doing this for Clovis, it becomes apparent that the closest sensors to Clovis are in Amarillo, Texas and Roswell, New Mexico, which are roughly 70 miles away from Clovis.

If the closest sensors to Clovis are at this distance, it becomes clear that their readings are unlikely to provide any meaningful insight into pollution in that city. Thus, any research into the effects of wildfire smoke on Clovis will likely need to be conducted without knowing what levels of pollution these wildfires cause for Clovis.

The other major limitation is that the most local level of detail offered by the main datasets used in this project is the county level, which makes it difficult to completely localize for Clovis. Although Clovis makes up the majority of Curry County, it is unavoidable that the rest of the county will have a non-negligible effect on the data. While Curry county data is still largely applicable to Clovis, unlike the data from Amarillo and Roswell offered by the AQS, the mismatch between the intended scope of the project and the county-level data will likely have an influence.

Conclusion

This project hypothesized that wildfire smoke would be negatively correlated with labor profits in Clovis, New Mexico. While this project was unable to support its hypothesis, it did not preclude the existence of such an association and identified potential future avenues of research and mapped out significant limitations that similar work centered on Clovis will likely need to address.

References

Borgschulte, Mark, et al. “Air Pollution and the Labor Market: Evidence from Wildfire Smoke.” *SSRN Electronic Journal*, 2021, https://doi.org/10.2139/ssrn.4141995.

Bureau, US Census. “Quarterly Workforce Indicators (QWI) (Time Series: 1990 - Present).”

Census.gov, www.census.gov/data/developers/data-sets/qwi.html. Accessed 12 Dec. 2023.

Operations, US Department of Commerce; NOAA; National Environmental Satellite Data and Information Service; Office of Satellite and Product. "NOAA's Office of Satellite and Product Operations." *Www.ospo.noaa.gov*, www.ospo.noaa.gov/Products/land/hms.html#about. Accessed 12 Dec. 2023.

United States Environmental Protection Agency. "Particulate Matter (PM) Basics." *US EPA*, United States Environmental Protection Agency, 2018, www.epa.gov/pm-pollution/particulate-matter-pm-basics.

US EPA. "Health Effects Attributed to Wildfire Smoke." *US EPA*, 13 Aug. 2019, www.epa.gov/wildfire-smoke-course/health-effects-attributed-wildfire-smoke.

Data Sources

The following links point to the data sources that were used in this project:

- AQS API documentation
- HMS data sheet for all U.S. counties
- USGS wildfire GeoJSON dataset
- QWI data aggregated at the county level for all US counties
- QWI data separated by age group for all U.S. counties