Data 512 Project – Extension Plan

# Introduction and Recap

In Part 1 of the Project, we 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 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. Our initial measure was not intended to relate wildfire smoke to any other measure, such as its impact on health, tourism, or the local economy; 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 producing a smoke estimate and comparing it to the EPA’s data, this part of the project will focus on producing a way to tie this smoke estimate to a concrete societal problem that will impact the target city. The goal for this will be to produce concrete evidence of the harm that wildfire smoke produces, as a way to make recommendations to policymakers. In order to do this, the problem that I have chosen to look into is the association between wildfire smoke and labor profits for my target city.

# Association of Wildfire Smoke with Labor Profits

The association between wildfire smoke and labor profits was studied in 2019 by researchers at the University of Illinois Urbana Champaign. [Their paper](https://docs.iza.org/dp15373.pdf) presents an analysis that combines Labor Market data from two different main sources (the Quarterly Workforce Indicators, or QWI, dataset from the Census Bureau, which was the more central source of the two, as well as data from the Local Area Unemployment Statistics program of the US Bureau of Labor Statistics), as well as pollution data that also comes from the EPA’s AQS API (specifically particulate AQI data), combined with data from the National Oceanic and Atmospheric Administration’s Hazard Mapping System (HMS) to locate wildfires to line up with the pollution data (similar to how the Common Analysis portion of this project used the USGS data) and weather data from the Global Historical Climatology Network of the National Climatic Data Center and the North American Regional Reanalysis (NARR) of the National Centers for Environmental Information, to help account for temperature/precipitation and wind direction respectively. After controlling for these factors and calculating the effect of a “smoke day”, IE a day in which a specific county was covered by wildfire smore, on the Labor Market data, the researchers from UIUC found that each smoke day resulted in a 0.097 percent loss in per-capita labor earnings.

To illustrate what that means on the national level, the researchers used the example of 2010, when 160 million workers made a total of 6.4 trillion USD, or 40,000 USD per capita. The researchers went on to note that, in 2010, the country had an average of 20.2 smoke days. Applying the 0.097 percent decrease for that amount of days, this means that those 160,000,000 workers that made up the 2010 national labor force collectively lost $125.4 billion dollars. This shows that lost earnings due to smoke represent a significant economic blow to workers nationwide.

My project will focus on establishing that a relationship between lost labor force profits and wildfire smokes holds when focused on the target city, and what the potential loses could look like for this city. In other words, this project will focus on proving whether the relationship the UIUC researchers observed holds at a local scale, and if it appears to it will also focus on approximating exactly what this means for the residents of Clovis, New Mexico.

# Proposed Methodologies

The first step in exploring this relationship between wildfire smoke and labor force profits will be to refine my existing model. The model currently only predicts aggregate values for the smoke estimate equation I developed myself, and only does so on a yearly basis, without providing any sort of estimation on the number of wildfires the target city can expect in a year.

In order to align it better with the trends described in the UIUC paper, I plan to attempt to tweak my model to both try to predict pollution more directly, as well as to produce more precise date ranges. The way I will attempt to introduce more specific date boundaries is to make better use of the “Listed\_Fire\_Dates” field that can be found in most of the individual wildfire JSONs. I hope that by doing so, I will be able to produce a metric similar to the “Fire Days” articulated by the UIUC researchers, and be able to approximate how much of a yearly wildfire season could result in large smoke coverage for my target city.

I also plan to try to change my model to directly use the EPA’s AQI data as a dependent variable, enabling it to help predict differences in AQI data that may be due to wildfires. I plan to do this by mapping wildfires to both absolute AQI readings from the EPA’s sensors, as well as creating a trendline that AQI shows the direction the readings go over time and mapping wildfire data to its residuals. I hope that this will help isolate the influence that wildfires specifically have over AQI readings, even if other factors, such as EPA programs or local environmental policy produce effects that improve air quality over time despite an increasing number of wildfires producing a negative effect on air quality.

After further defining this relationship between wildfires and smoke pollution, I hope to also establish a link between labor profits for my target city and air quality. As discussed earlier, the primary source for labor profit data that the UIUC researchers used was the Quarterly Workforce Indicators Dataset from the Census Bureau. The Census Bureau maintains [a set of endpoints](https://www.census.gov/data/developers/data-sets/qwi.html) that allow for retrieving data for a specific quarter-year range (IE 2012 Q1 to 2022 Q3, to provide an example range) at the county level. This would normally not be helpful to provide information on a specific city; however, it can be useful in this specific case. Curry County in New Mexico is home to my target city, Clovis. [Demographic data for Clovis,](https://www.city-data.com/city/Clovis-New-Mexico.html) and [Curry as a whole](https://www.city-data.com/county/Curry_County-NM.html), shows that Curry County is largely dominated by Clovis, with 38,000 of Curry’s 48,000 residents living in Clovis. For simplicity’s sake, I will assume that most of the QWI’s trends for Curry County generalize to Clovis specifically. Using the QWI data for Curry’s local economy, I will conduct a similar analysis on its relationship to air pollution, trying to understand how much changes in air pollution seem to impact Curry’s workforce indicators. As with my plans for modeling the relationship between wildfire smoke and air pollution, I plan to model the relationship between observed air pollution and the relevant indicators, both in absolute terms and for the residuals that might help show the relationship between the two even if other trends push these indicators in the opposite direction.

Finally, if time permits, I also hope to replicate the analysis produced by the UIUC researchers and run their code using only data for my target city, to better understand how much a “smoke day” can impact Curry’s/Clovis’s local economy specifically. The code needed to replicate the UIUC paper’s analysis can be found in the personal website for one of the researchers, David Molitor. He maintains [a list of all of the papers and publications he’s worked on](https://www.davidmolitor.com/research), and under each one he keeps a link for all of the data and code used to produce the results for that work. The materials for the paper on wildfire smoke and its impact on the labor market can be found [here](https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/CSGEJI). While this has a high chance of producing useful insights into the effect of wildfire smoke on Curry’s/Clovis’s local labor profits, the Readme for this code lists that the expected runtime for all of the code is around 48 hours (on a computer with very similar specs to my laptop), and uses Stata, a piece of software I am not currently familiar with. Because just running the code is expected to take a significant amount of time, without factoring in the time commitments of understanding the code or becoming familiar with Stata, replicating the setup for the UIUC paper will need to be deprioritized to prevent this project from simply becoming a replication study. If time allows for it, however, it will likely help produce a much deeper understanding of the effects of wildfire smoke on Clovis.

# Conclusion

To summarize, I am hopeful that the work outlined in this extension plan will help show the impact that wildfire smoke can produce on Clovis, and will serve as a call to action to address the problems that result from it. Although wildfire smoke is already known to produce general harm, I hope this work will show concrete evidence of one aspect of its effects.