Daniel Vogler

DATA 512

Visualizations and Reflections

Number of Wildfires within 1800 miles of Pueblo, Colorado Since 1961

Source: USGS
Excludes prescribed burns.
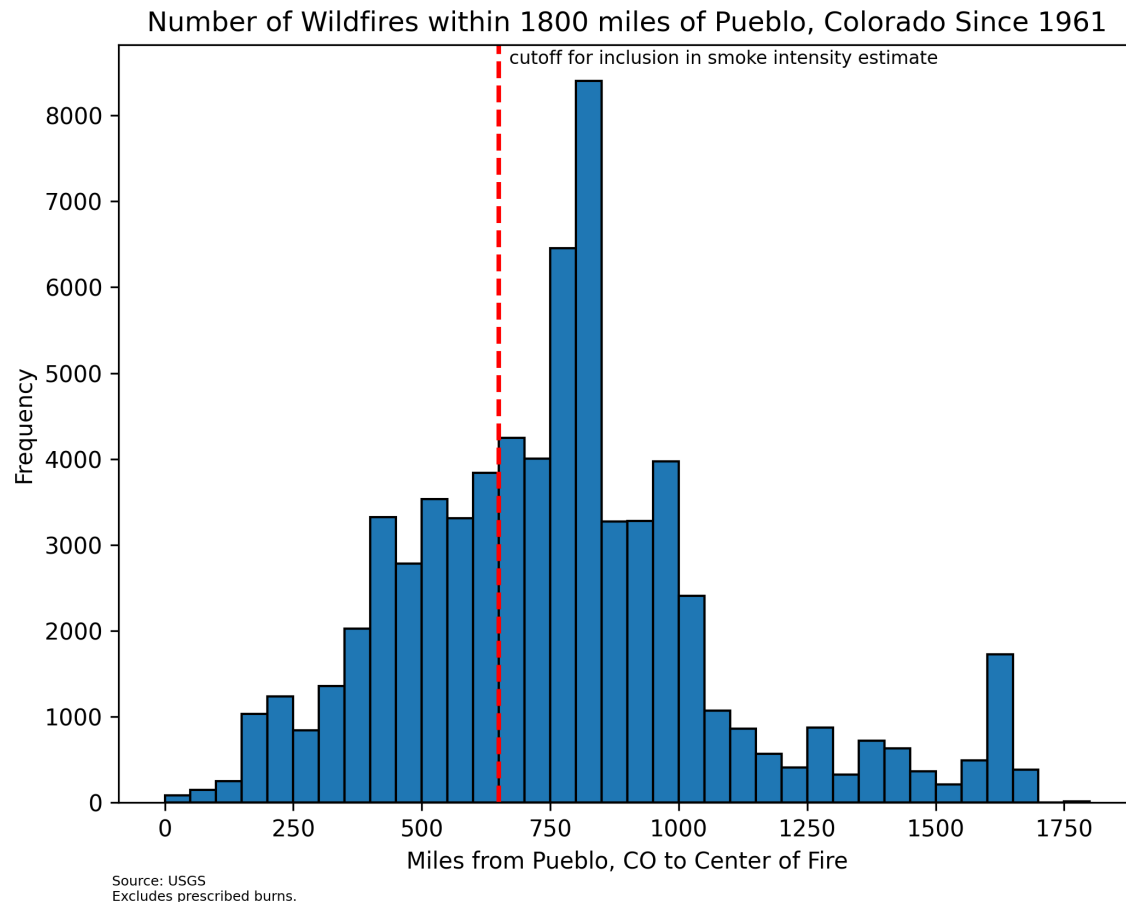
Reflection

This figure shows the frequency of wildfires since 1961 (during the last 60 years of available data) at each increment of 50 miles from Pueblo, CO. The axes are frequency (y-axis) and miles from Pueblo, CO to the center of a given fire (x-axis) A viewer reads this by interpreting the relative height of each bar as indicating the frequency with which wildfires occur at that distance from Pueblo. The underlying data is provided by the United States Geographic Survey. It was processed by calculating distances from each file to Pueblo, CO; filtering out the distances that exceed 1800 miles; and dropping prescribed fires to leave only wildfires.

Although Pueblo, CO is located in a region that is at high risk of wildfires, relatively few wildfires occur in its immedate vicinity (<250 mi). However, a greater number occur relatively close to it (<650 mi). This part of the country is densely forested in some areas and extremely sparse in others; this combination of factors could allow smoke to travel long distances and impact air quality in Pueblo, even if the fire doesn't occur very close by.

Total Annual Acreage Burned by Wildfires within 1800 mi of Pueblo
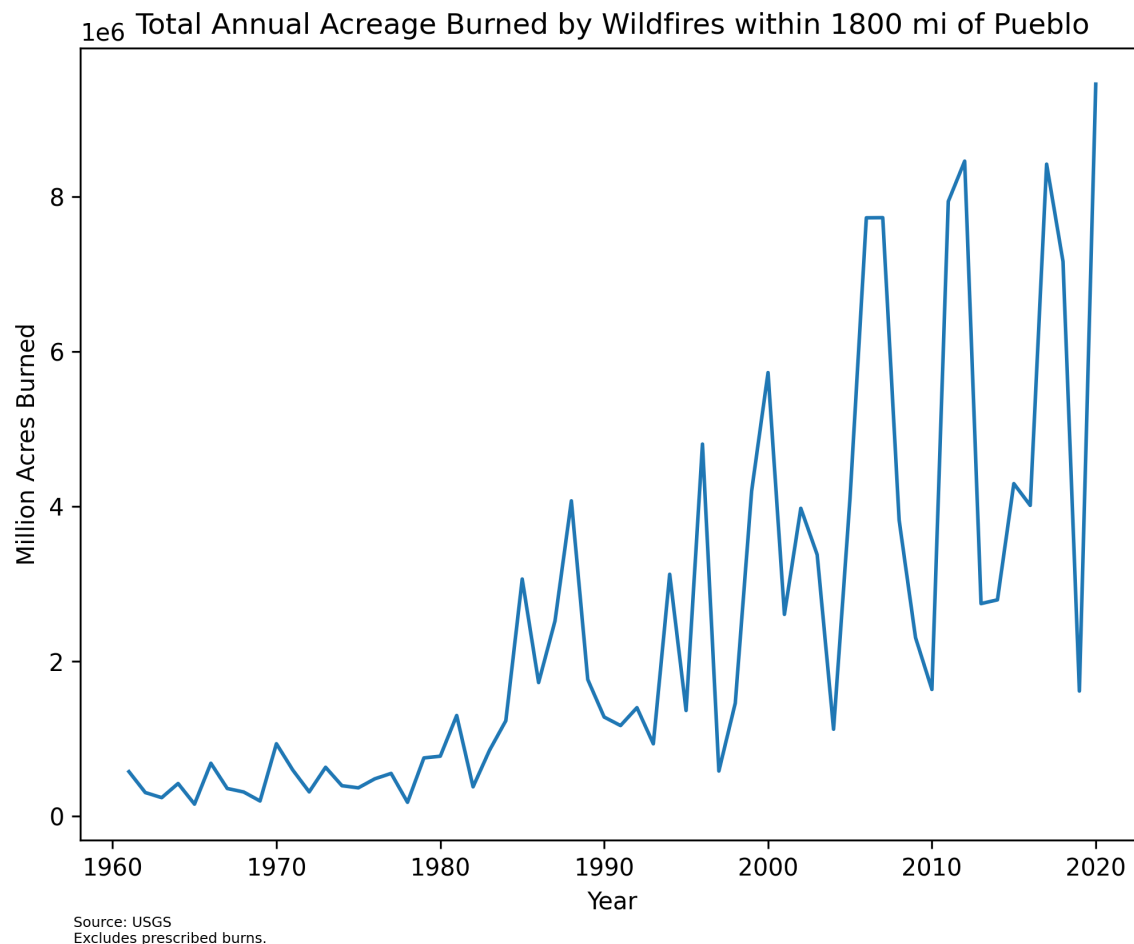
This figure shows the acreage burned by wildfires since 1961 (during the last 60 years of available data) within 1800 miles of Pueblo, CO. The axes are acres burned (y-axis) and time, measured in years and starting at 1961. A viewer reads this by interpreting the relative height the line as indicating the total number of acres burned by wildfires within 1800 miles of Pueblo in each year. The underlying data is provided by the United States Geographic Survey. It was processed by calculating distances from each file to Pueblo, CO; filtering out the distances that exceed 1800 miles; and dropping prescribed fires to leave only wildfires.

This chart has a few interesting features. First, there is a clear trend: over time, the acreage burned by wildfires tends to go up. However, the acreage burned each year is highly volatile. There seem to be discernible peaks in acreage burned every 3-7 years (roughly

1982, 1986, 1988, 1996, 2000, 2008, 2012, 2018, 2021). This suggests some multi-year sesonality in the data.

AQI and SIE Over Time
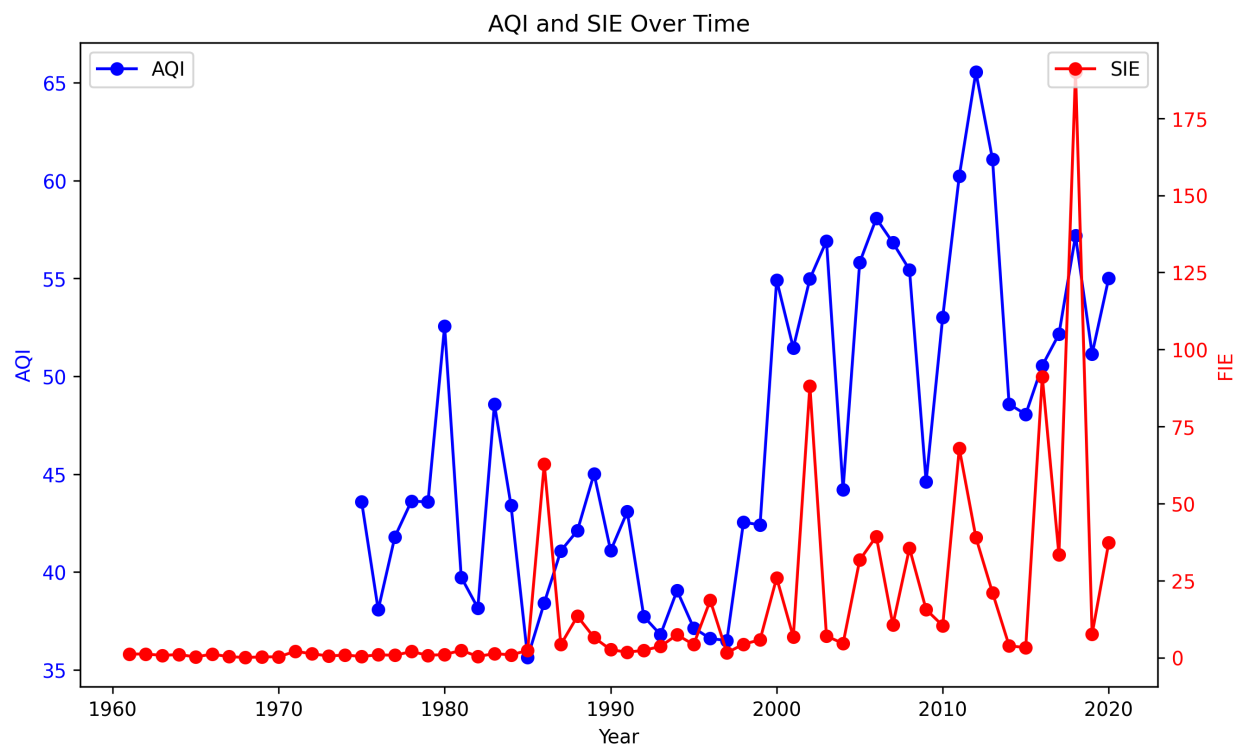
Reflection

This figure shows the air quality index (AQI) since 1974 (the period during which the EPA has been conducting the measurements that produce the AQI) and a custom scoring metric of called the Smoke Intensity Metric (SIE). These latter metric is constructed by considering fires within 650 miles of Pueblo, CO and assigning a score based on the acreage burned and the distance from Pueblo.

This plot as a dual y-axis. The blue line, representing the AQI, is indexed by the left-hand axis. The red line, representing the SIE, is indexed by the right-hand axis. This format allows us to superimpose two time series whose values lie on different scales and see how they both vary with time, the variable on the x-axis. A viewer reads this by interpreting the relative height each line as indicating the level of the respective metrics, and comparing the patterns in the two lines – do they move together over time? Is there any lag? Are there notable times they come apart?

The underlying data is provided by the United States Geographic Survey and the EPA. The SIE was constructed by calculating distances from each file to Pueblo, CO; filtering out the distances that exceed 1800 miles; and dropping prescribed fires to leave only wildfires; then the SIE was computed by multiplying the acreage burned by each wildfire by the inverse-square of the distance and summing the contribution of each fire over an entire year. The AQI was pulled on a daily basis from the US EPA; it was processed by averaging the daily figures to produce yearly estimates.

This chart displays interesting trends and relationships in each of the underlying datasets and between them. First, the absolute value of both scores tends to increase over time, possibly indicating worsening air quality. It also illustates that while there is some relationship between the two variables, they do not necessarily move up and down together. Sometimes, as in 2001, 2010, and 2018, peaks coincide. Other times, as in 2016, one metric sharply peaks while the other does not. The correlation between same-year observations is 0.47, indicating a moderate relationship between the two metrics displayed in this chart.

Collaboration Reflection Statement

In the first phase of this project, I collaborated with others in the course to make important decisions about the most efficient tools to work with the data, how to filter it and pre-process it, designing my smoke intensity metric, and how to approach predictive modeling.

Collaboration helped me select the right tools for different tasks required to complete this assignment. For example, Apoorva Sheera and Manya Chadha advised me to use geopandas because this package is optimized for dealing with geoJSONs with geometry objects. Taking this advice allowed me to efficiently work with dataframes, which simplified the work required to merge datasets across different tables.

I also worked with others to resolve perceived ambiguities in the assignment specification and make judgements about the most effective ways to pre-process and filter the data. For example, I discussed with Ed Seryozhenkov and Sid Gurajala whether to filter down the dataset to only wildfires or to include prescribed burns.

Furthermore, Ed, Sid and I collaborated to develop out smoke intensity metrics. We discussed whether to use centroid or edge distances; models that assumed a Gaussian distribution of smoke around the point source; and justifications for settling on an inverse-square law. This was particularly helpful. Even as we each implemented out metrics in slightly different ways, the act of discussing them forced us to clarify our thoughts, express them in clear equations, and justify one modeling choice over another.

Finally, it was helpful to consult collaborators for the predictive modeling aspect of this assignment. Sid Gurajala and Alex Netzley shared with me that traditional, non-seasonal ARIMA models had yielded poor results in their trials. This led me to scrutinize trends in my smoke intensity score and examine whether a *seasonal* model might be more appropriate. This in turn led me to choose a SARIMA model.

In conclusion, the collaborative efforts in this project significantly enhanced the depth and quality of our analyses, and forced me to clearly communicate my reasoning for certain analytical decisions. By engaging in discussions with my peers, I was able to navigate the complexities of data wrangling and predictive modeling more effectively. The insights

shared by Apoorva, Manya, Ed, Sid, and Alex not only influenced the tools I selected but also helped me develop a clearer and better-justified smoke intensity metric and predictive model.