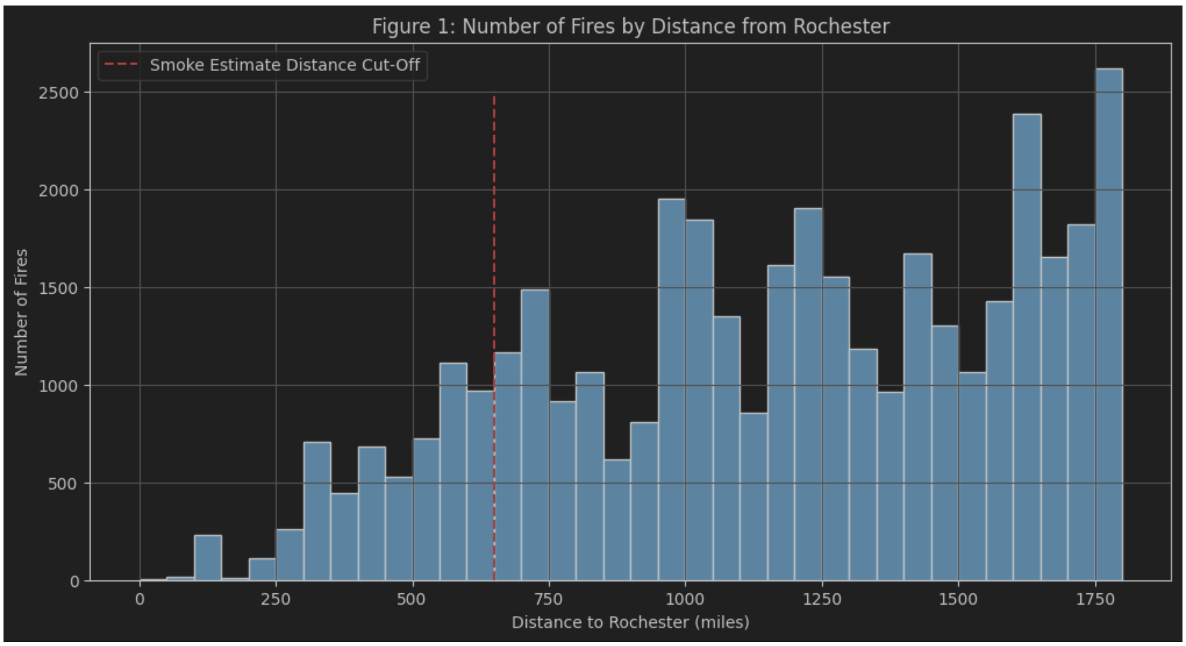
Project Part 1 – Common Analysis

Visualizations and Reflections

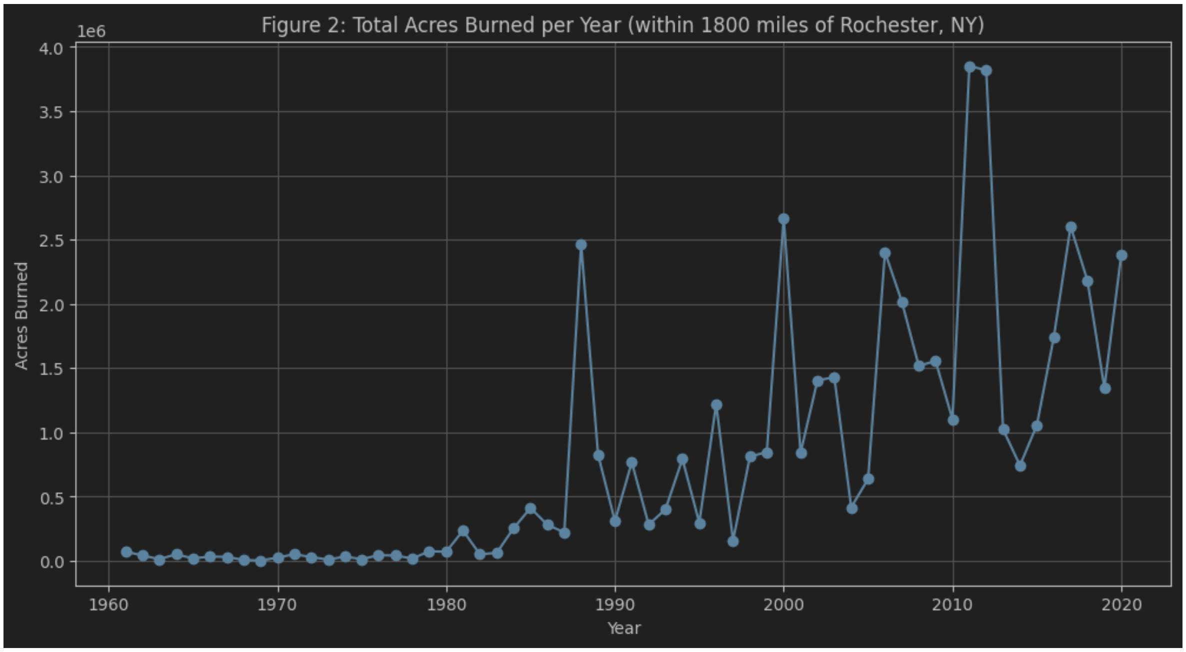
Ted Liu

Visualizations:



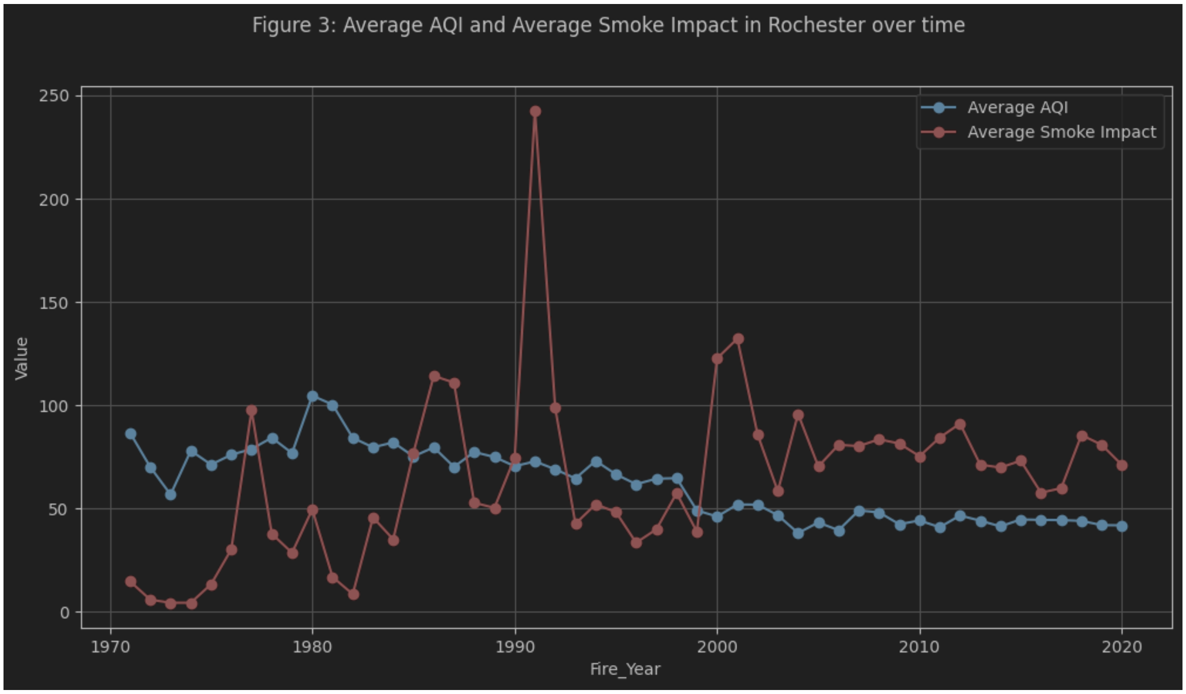
This histogram describes the distribution of wildfires within 1800 miles of Rochester, NY. The histogram shows that many fires occurred further away from the city, with most fires occurring more than 1000 miles away. The x-axis on this plot represents the distance buckets. Each bucket represents a 50-mile range distance from Rochester, so the first bucket represents fires that occurred within 0-50 miles of Rochester and the second bucket represents fires that occurred within 50-100 miles. This pattern continues until the maximum specified distance of 1800 miles. The y-axis represents the number of fires that occurred in each 50-mile bucket. We can therefore use the x-axis and y-axis to find out the total number of fires for a given distance bucket.

The data source of this histogram was the USGS wildfire data from sciencebase.gov. We processed the combined dataset, filtering for fires that were within 1800 miles of Rochester and occurred between 1961 and 2021.



This time series plot displays the total acres burned annually by fires within 1800 miles of Rochester, NY. The x-axis represents the year, and the y-axis represents to total acres burned in units of 1e6. We can therefore use the x-axis and y-axis to find out the total acres of land burned for a given year.

The data source of this time series data was also the USGS wildfire data from sciencebase.gov. We processed the combined dataset, once again filtering for fires that occurred within 1800 miles of Rochester and occurred between 1961 and 2021. We then took the fire data and aggregated it by year, summing up the acres burned to get a total acre burned value for each year.



This time series plot displays a comparison of the Average Smoke Impact estimate we created and air quality index (AQI) data from the EPA AQS API. The x-axis represents the year. The y-axis represents the numeric value of the AQI and the Smoke Impact estimate. We specifically designed our smoke estimate to be within the same range as the AQI.

The data source for this time series plot is both the USGS wildfire data from sciencebase.gov and AQI data from the EPA AQS API. We calculated a smoke estimate for each fire in the USGS wildfire combined data set using the formula:

$$Smoke\ Impact = \frac{Acres\ Burned\ \times D}{Distance\ to\ City\ in\ Miles}, where\ D\ is\ a\ scaling\ constant$$

By trial and error, we set $D = 100$. We also limited the smoke impact to values between 1 and 500, so if the smoke impact's value ended up being greater than 500 or less than 1, we simply set it these limits appropriately. We then performed an aggregation by year to get the Average Smoke Impact estimate per year.

We then pulled data from the AQS API for AQI data. We pulled AQI data for the parameters codes which correspond to common pollutants like CO_2, O_3, PM2.5, etc. We then performed an aggregation on these pollutants associated AQI, taking the max AQI for a given date. Before merging this AQI data onto our smoke estimate data, we performed another aggregation to get the average AQI for each year. We then merged the average smoke impact estimate per year with the average AQI per year and plotted it.

Reflection:

This project tasked us with answering the question:

> "What are the estimated wildfire smoke impacts on your assigned city each year for the most recent 60 years of wildfire data?"

In my analysis, I discovered that, for my assigned city of Rochester, NY, most of the relevant fires (within 1800 miles) occurred outside of the 650-mile threshold we defined for our smoke estimate. This is interesting because this means that the smoke estimate we created does not actually use data from most fires that occurred within the specified distance from the actual city. We did observe in the second figure that the total acres being burnt by fires each year seems to be trending upwards as time progresses – perhaps suggesting an ever-increasing amount of fires or increasing severity of fires. We also observed that our basic smoke impact estimate is quite inaccurate (at best) when compared to actual AQI data. This perhaps suggests the importance of external factors, such as wind data, when estimating the smoke impact from fires.

The possibility of collaboration in this portion of the project was certainly welcome, but did not impact my approach or thinking. While I did go out of my way to avoid collaboration, I hesitated to provide other students with even snippets of code as I wanted to minimize the risk of any plagiarism policies. Despite this, I found discussions about the assignment with my peers to be insightful. It was interesting to learn about all the various approaches other students took to process the data as well as their approach to coming up with a smoke estimate. I specifically took inspiration from my discussions with my peers – Parvati Jayakumar, Chakita Muttaraji, and Himanshi Naidu – on the topic of coming up with a smoke estimate, where the idea of the ratio between the acres burned and distance initially came up.