# Data 512- Part1 Common Analysis of Wildfires

-Saikripa Mohan

## Part1 : Visualization analysis

In this section I am going to describe and discuss the graphs plotted as part of visualization.
**Plot 1: Number of fires by distance from city**

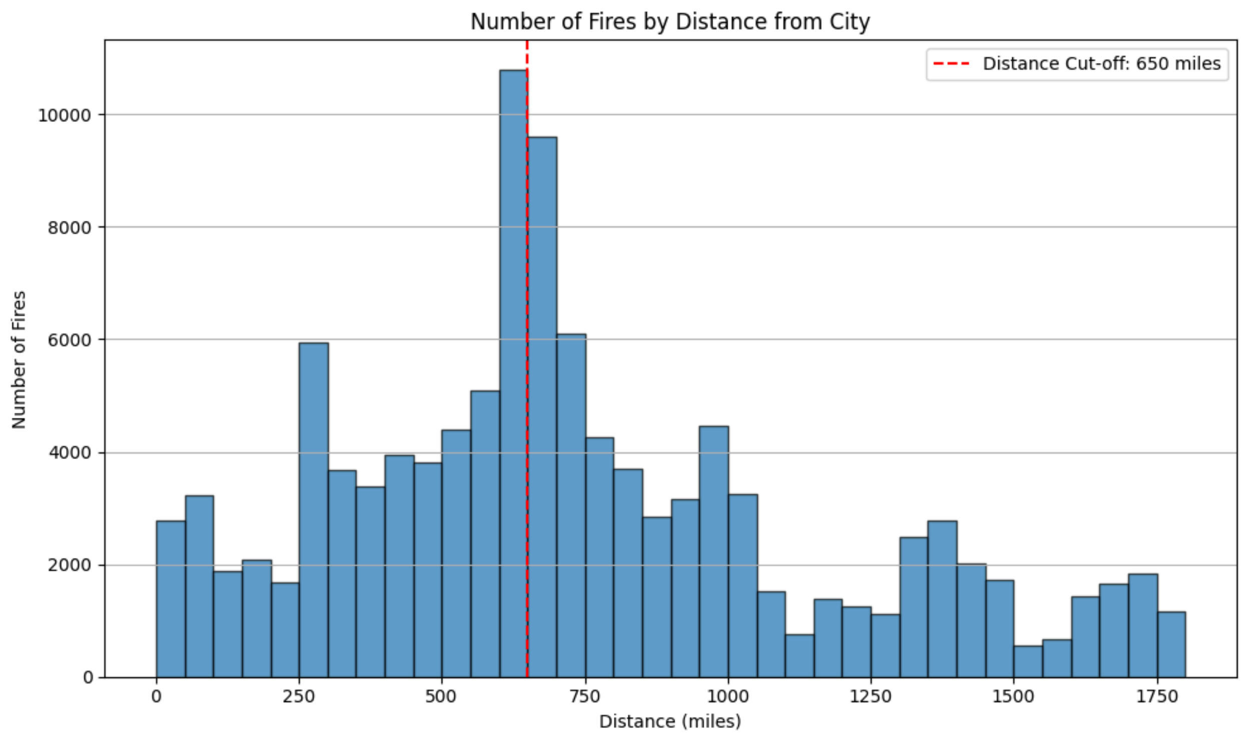


*Figure 1: Number of fires by distance from city*

In this histogram, I'm showing the distribution of fires at various distances from the city of Rialto, which helps reveal where fires are most commonly occurring relative to the city. The x-axis represents distance in miles from Rialto, split into equal-width bins, while the y-axis shows the number of fires within each distance bin. I've marked a red dashed line at the 650-mile distance, which serves as a cut-off point for analysis.
From the histogram, we can see that fire occurrences peak between 500 and 750 miles from Rialto, with the number of fires exceeding 10,000 in this range. This suggests that fires are particularly frequent within this band, while their frequency drops both closer to and further from this range. There are also smaller peaks at other distances, indicating additional clusters of fires, although these are less pronounced than the main peak.

To create the plot, I aggregated the fires into distance bins to make the pattern clearer. This approach allows me to visually highlight how fire occurrences are distributed around Rialto and to identify potential hotspots or areas with higher fire frequencies.

**Plot 2: Total Acres Burned per year**

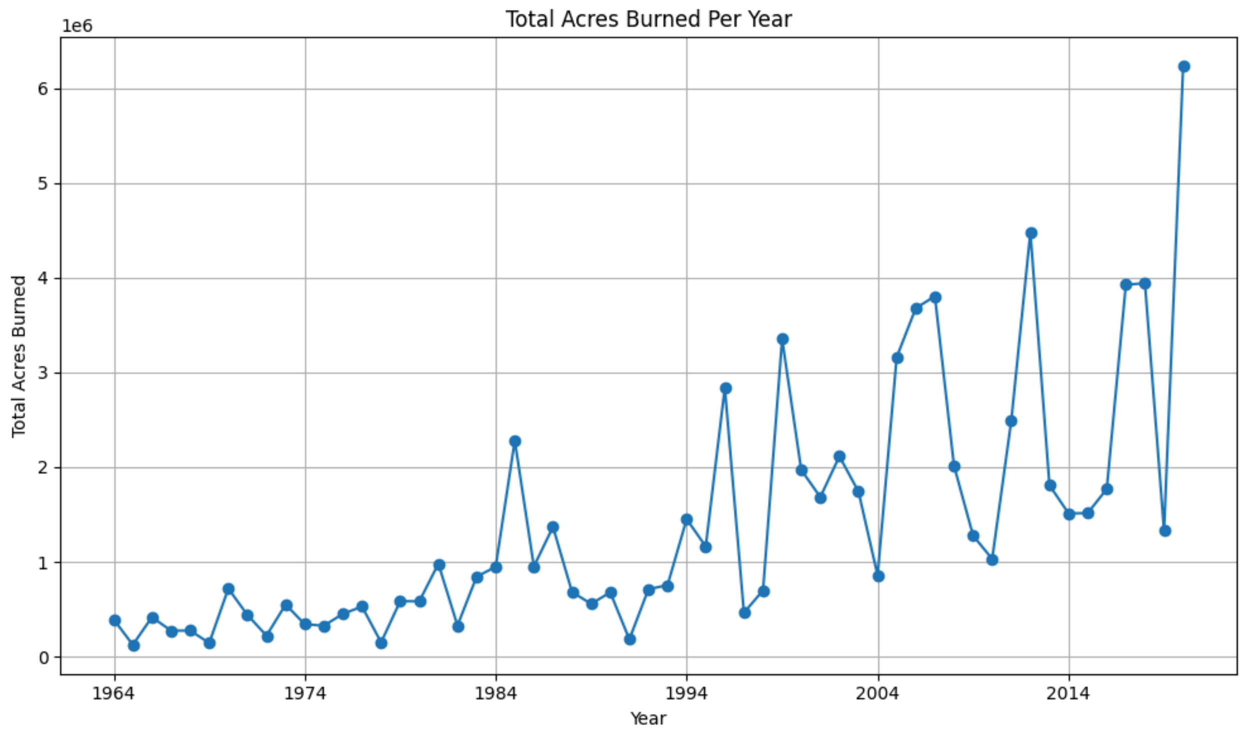


*Figure 2: Total Acres burned per year*

In this line plot, I'm showing the total acres burned by wildfires each year, allowing us to track trends in wildfire impact over time. The x-axis represents the years, spanning from 1960 to recent years, while the y-axis shows the total acres burned, measured in millions. Each data point represents a single year's total burned acreage, and the line connecting them helps to visualize changes over time.

The plot reveals an upward trend in wildfire impact, with some years marked by sharp peaks in burned acreage. Notable spikes occur in the early 1980s, mid-2000s, and a significant peak in the most recent year, where the total burned area exceeds 6 million acres. This trend suggests that wildfires have generally become more severe or more widespread over the decades, particularly in recent years.

To create the plot, I aggregated these records by year to give a clear, year-by-year view of wildfire impact. By visualizing this data over time, I can highlight the increasing scale of wildfires, which may reflect changes in climate, land management practices, or other environmental factors influencing fire frequency and intensity.

**Plot 3: AQI and Smoke Estimate over time**

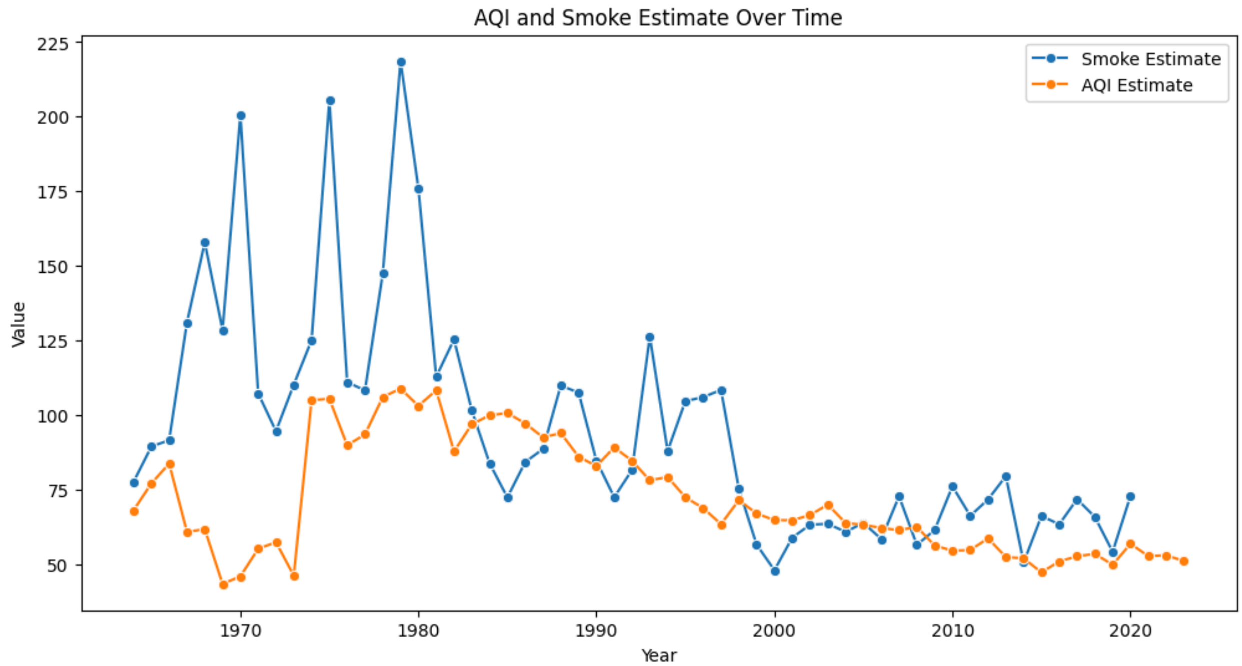


*Figure 3 AQI and Smoke estimates over time*

This line plot illustrates the trends of Smoke Estimate and AQI (Air Quality Index) Estimate over time. The x-axis represents the year, while the y-axis shows the values for both metrics. The Smoke Estimate, shown in blue, fluctuates significantly with high peaks in the 1960s to early 1980s, reaching values above 200, then gradually declines with fewer and lower peaks in recent years. The AQI Estimate, represented in orange, generally remains lower than the Smoke Estimate and shows a relatively steady pattern with a gradual decrease over time. This plot suggests that while smoke levels were initially high and variable, both smoke and air quality estimates have improved in recent years.

From the graph we can see a positive correlation between the Smoke Estimate and AQI Estimate over time, especially in the earlier years when both metrics tend to move in tandem. When smoke levels are high, AQI values generally increase as well, indicating poorer air quality associated with higher smoke levels. However, the correlation weakens in later years as the Smoke Estimate shows a more pronounced decline, while the AQI Estimate remains relatively stable with less fluctuation. This shift suggests that although smoke levels have decreased over time, other factors may now be influencing air quality independently of smoke, or air quality management efforts may be mitigating some impacts of smoke on AQI.

# Reflection

In this assignment, I engaged deeply with the impact of wildfires on air quality, particularly through the lens of the Air Quality Index (AQI). This exploration led to a twofold learning experience: a nuanced understanding of the relationship between wildfires and air quality metrics, and technical proficiency in processing geoJSON data and utilizing pooling methods for efficient data processing. A significant part of my learning involved understanding smoke estimates and their actual calculation, which are crucial for assessing the impact of wildfires on air quality.

**Learning from the Research Question**

The research question prompted an investigation into how wildfires contribute to fluctuations in the AQI. I learned that air quality is significantly affected by the particulate matter released during wildfires, which can lead to increased health risks in nearby populations.

In addition to conceptual insights, I developed practical skills in data processing. Working with geoJSON data was particularly challenging, as I learned to extract relevant information to create my DataFrames. This experience enhanced my ability to handle complex data structures effectively and allowed me to incorporate vital geographic information into my analysis. Moreover, I utilized a multiprocess pool method for pooling, which enabled me to read and process large datasets quickly. This approach significantly improved the efficiency of my data handling and analysis, allowing me to focus on drawing meaningful conclusions from the data.

**Collaboration and Brainstorming**

The collaborative aspect of this assignment proved invaluable. Engaging with my classmates allowed me to refine my understanding of the project requirements and explore optimization techniques in coding. For example, during discussions about their smoke estimation formulas, I gained insights into different methodologies for calculating the impact of wildfires. These discussions helped me appreciate the various approaches to solving the same problem.

I would like to provide Attribution to Raagul Nagendran and Manasa Shivappa

I collaborated with Raagul N, who shared his efficient coding techniques for processing geoJSON data. The method of organizing the data effectively streamlined my DataFrame creation process. Similarly, I learned from Manasa about effective handling of AQI API data, which improved my approach to integrating real-time air quality measurements into my data analysis.This assignment deepened my understanding of the intricate relationship between wildfires and air quality while enhancing my technical skills in data processing. The collaborative environment not only clarified project requirements but also opened avenues for optimizing my approach to problem-solving.