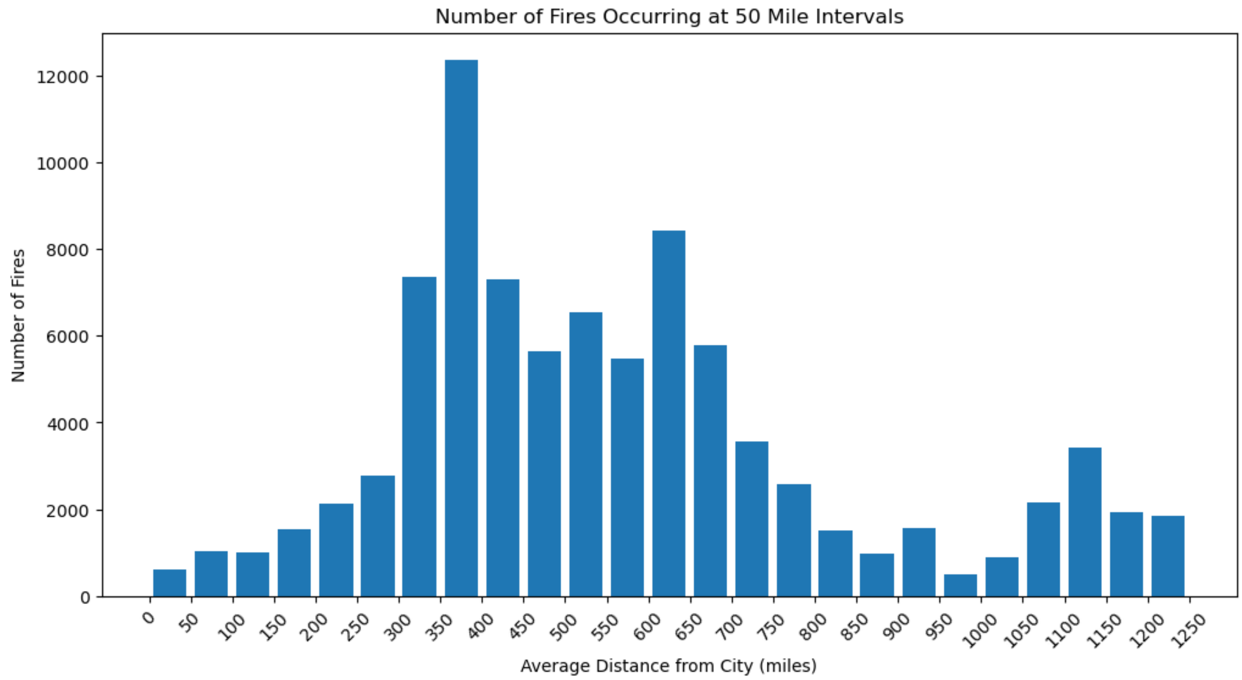
Visualizations

1. The histogram shows the number of fires occurring every 50-mile distance from your assigned city up to the maximum specified distance.

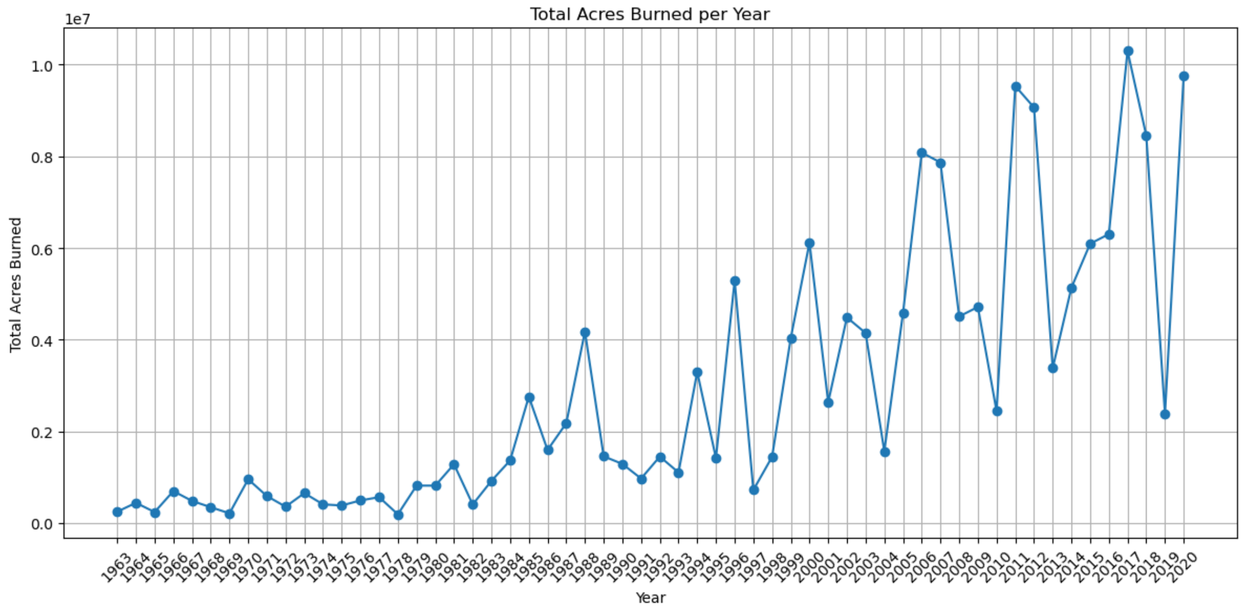


This histogram shows a graphical representation of the distribution of the number of fires. The x-axis represents the average distance from a city in miles, and the data is grouped into 50-mile intervals. The y-axis represents the number of fires, which is the frequency of the fires within each interval on the x-axis.

This histogram shows a non-uniform distribution of fire occurrences relative to the distance from a city. There are peaks at certain intervals, for example around 350 miles and 600 miles. There are fewer fires at distances closest to 0-200 miles, and the number increases as the distance range increases to around 350 miles, then fluctuates.

The underlying data for this histogram is a set of records of fires, each with a recorded distance from Cedar City. To make this histogram, the data was preprocessed with calculation to sort the distances into 50-mile intervals. The number of fires within each interval was counted and displayed in the histogram.

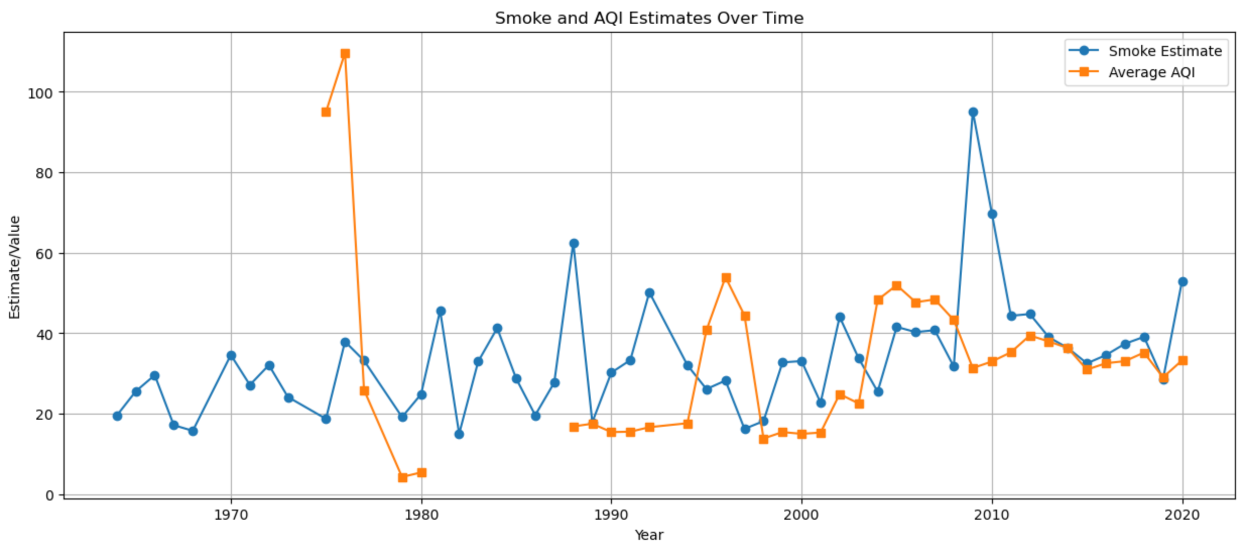
2. The time series graph of total acres burned per year for the fires occurring in the specified distance from Cedar City, Utah.



This time series graph displays a trend over time of the total acres burned by wildfires each year. There has been an overall increasing pattern over the years, but it has been a big fluctuation, especially since 1988. To read the figure, the viewer can look at the x-axis to identify the year and then see the corresponding point on the y-axis to determine the total acres burned in that year. The x-axis represents the year, ranging from 1963 to 2020. The y-axis represents the total acres burned, and it uses scientific notation to indicate the scale, with the numbers representing multiples of 10 million acres.

The underlying data consists of the year and size of the fires. The data was processed by summing up the total area burned each year to produce the annual points on the graph. Lines are drawn connecting consecutive years' data points to visualize the trend of acres burned over time.

3. The time series graph containing your fire smoke estimate for your city and the AQI estimate for Cedar City, Utah.

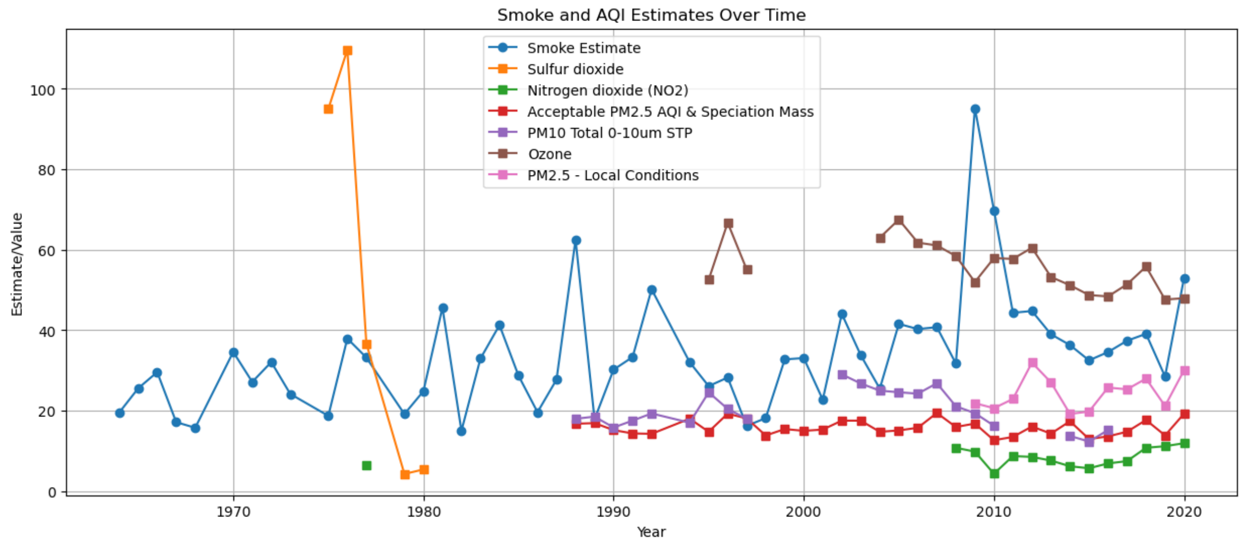


This time series graph presents 2 sets of data over time, from 1963 to 2020. The blue line represents the smoke estimate generated in the previous section. The orange line represents the average AQI generated from the notebook `epa_air_quality_history.ipynb`. The x-axis represents the year, and the y-axis represents both the smoke estimate index and air quality index since the smoke estimate and AQI have the same range as I rescaled in Section **Estimate the Smoke**.

The viewer can read the graph by looking at the year on the x-axis and looking at the corresponding estimated values for both smoke and AQI on the y-axis. The lines connecting the data points help to visualize how these estimates and AQI have changed over the years.

The data points for both series do not follow a consistent pattern, indicating variability in smoke and air quality, and the average AQI has missing data for some of the years. The pattern of smoke estimate and average AQI has a relatively consistent pattern in the most recent decade.

(continue from the previous page)



Similar to the previous time series graph, this time series graph presents 2 sets of data over time, from 1963 to 2020. The blue line represents the smoke estimate generated in the previous section. All other lines represent different pollutant AQIs as shown in the legend in the graph.

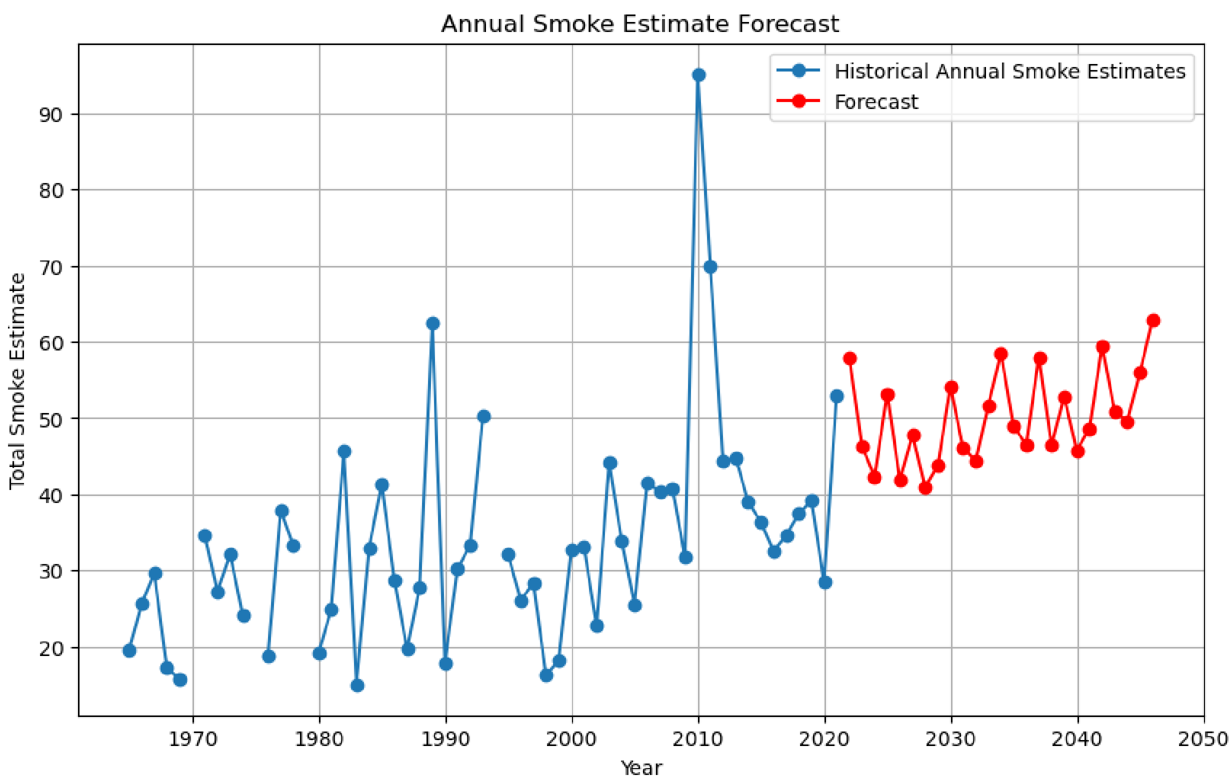
The x-axis represents the year, and the y-axis represents both the smoke estimate index and air quality index since the smoke estimate and AQI have the same range as I rescaled in the previous section.

The data points for the smoke estimate and AQIs do not follow a consistent pattern, indicating variability in smoke and air quality, while the pattern of each AQI is consistent. The pattern of smoke estimate and average AQI has a relatively consistent pattern in the most recent decade.

Reflection Statement

While working on this assignment about the impact of wildfires on a city's air quality, I've learned a lot more than I expected. One thing I learned is that wildfire smoke is not the predominant factor affecting air quality. It turns out that the reality is a lot more complicated. There could be other factors contributing to the impact – like the weather, where the city is located, what other kinds of pollution are already there, etc. Even though the smoke estimator is simply defined, we can see that the pattern between the AQI and smoke estimate is not very consistent.

However, there is still an overall pattern that we can discover. Even though the AQI and smoke estimate fluctuate aggressively over the year, there is an overall increasing trend in smoke estimate. The predictive model also presents an increasing smoke estimate forecast for the future 25 years. As the graph below shows, the red part represents the prediction with the model, while the blue line represents the smoke estimates. The prediction shows an increasing fluctuation for the next 25 years. Combined this trend with the graph Total Acres Burned per Year in the Visualization section, there is also an increasing pattern in the total acres burned by the wildfires.



The collaboration enriched my understanding and approach to the research problem significantly. I was only able to focus on my city initially, but collaboration opened up new perspectives. Working with other students, I was able to take a look at the pattern for other cities in the Northwestern part of the US. When requesting air quality monitoring stations, I started with the county-wise approach and did not realize the shortcomings of this method. For example, Tejal Kolte is analyzing the city of Longview in Washington state. This is a city located on the border of Washington and Oregon. If we only requested stations using county FIPs code, then the monitors located in Oregon would not be retrieved, and the AQI data would be incomplete, which can negatively impact the analysis result. This discovery inspired me a lot, even though Cedar City is not located on the border, it is still important to try different approaches such as Bounding Box to collect data from as many sites as possible within a reasonable distance.