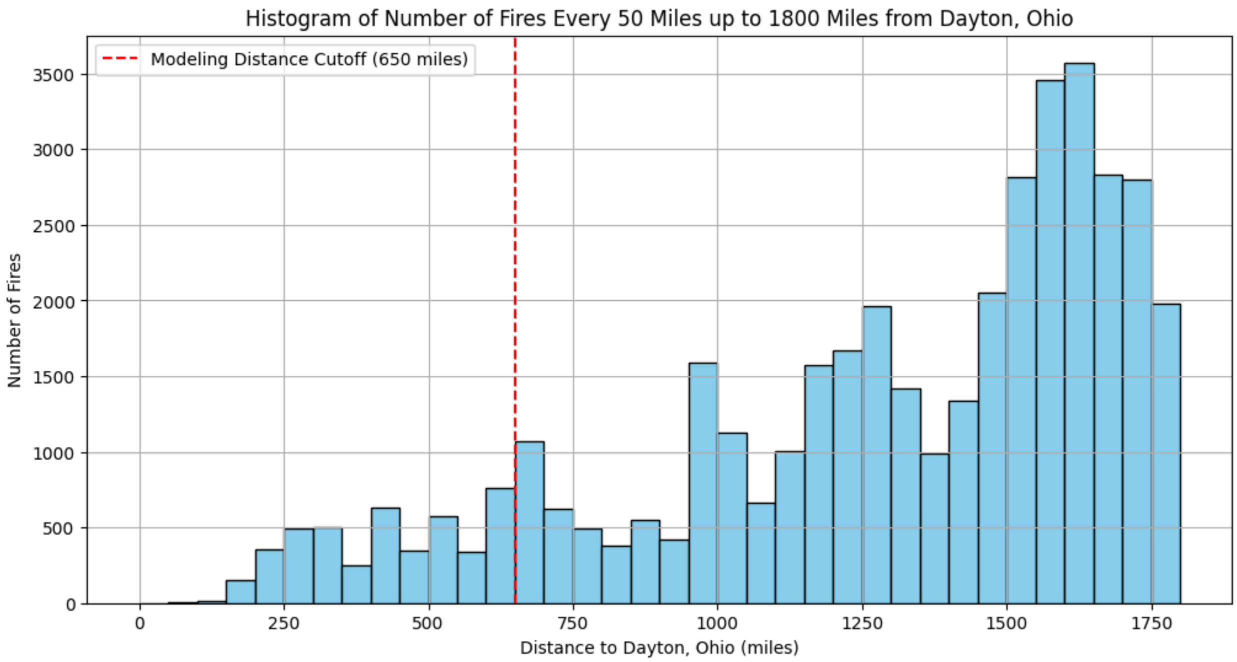
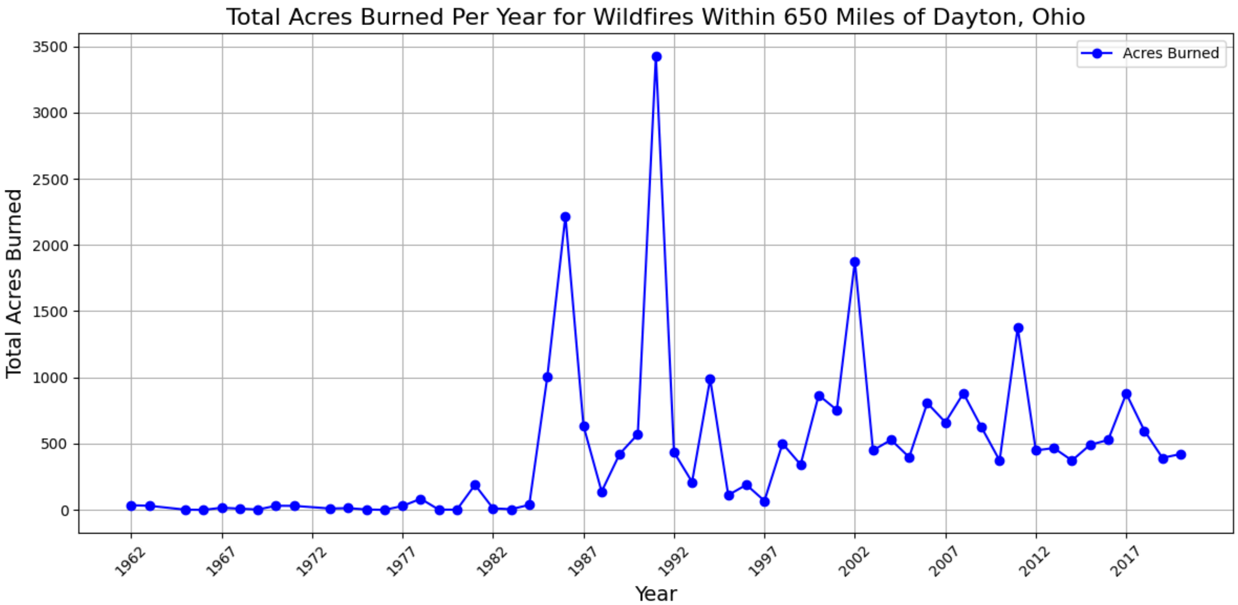
Isha Gokhale

Common Analysis Part 1 Reflection

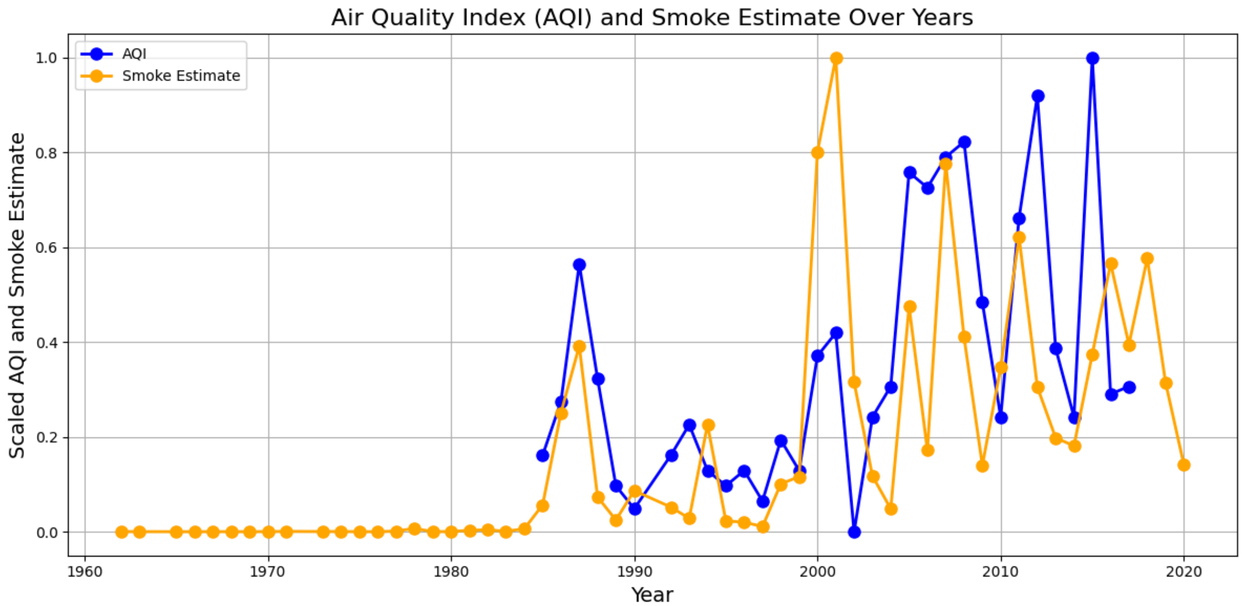


The figure above shows the number of fires every 50 miles from Dayton, Ohio. The red line indicates the distance limit for the rest of my analysis (650 miles). The X axis is the number of fires and Y axis is the distance from Dayton, Ohio in miles. Each bar is the number of fires that occurred within a given 50 mile threshold (as specified on the X axis). The viewer should read the figure from left to right and observe how the number of fires increases as the distance from Dayton increases. The underlying data is wildfire data from 1061 to 2021, which includes the distance from Dayton. To process the data, I had to extract the distance for each year of wildfire data I had and filter out any wildfires that were more than 1800 miles away. Then, I had to create 50 unit bins from 0 to 1800 and place the fires in each bin accordingly, giving us the histograms we see above. The data was collected and aggregated by the US Geological Survey. The dataset provides fire polygons in ArcGIS and GeoJSON formats. The link to the data is here. This dataset was used as the source for all my visualizations (seen below).



The next visualization is the Total Acres Burned per Year for Wildfires within 650 Miles of Dayton, Ohio. The x axis is the Year and y axis is the total number of acres burned. The user should follow the data chronologically from left to right to see how the total acres burned per year throughout the yearly data we have available. The data was gathered through the US Geological Survey mentioned above and was processed by extracting the acres burned per year by iterating through all wildfires within a 650 mile radius of Dayton, Ohio. For each year, if there were multiple wildfires, I aggregated the total acres burned, ensuring that the graph reflects the cumulative impact of these events.

Notably, the graph reveals a significant increase in burned acres in 1991, indicating a particularly active wildfire season with several large incidents within the 650-mile radius. This uptick suggests that environmental conditions or other factors may have contributed to more intense fire activity that year. However, the period from 1962 to 1984 appears relatively calm, with acres burned consistently near zero. This could suggest a time when either fewer wildfires occurred or conditions were less conducive to fire spread, highlighting a notable contrast in wildfire activity over the decades.



The last visualization is the AQI index and Smoke Estimate from 1961 to 2021. The x axis is the year and the y axis is the scaled AQI and Smoke Estimate. Since my smoke estimates were very small (on the order of around 0.0001) and my AQI estimates were in the 10s, I decided to scale both those columns to be between 0 and 1. For example, to scale a value x I followed the following equation: *(x - min) / (max - min)* to get my scaled value.

With these adjustments, the graph reveals a clear trend showing how smoke estimates and AQI move together over time. It's interesting to see that they follow each other pretty closely, which suggests a relationship between smoke levels and air quality. We have to keep in mind there is no AQI data before 1985 which is why there is missing AQI data for that time on the graph. Looking at the data, you'll notice a noticeable spike in both AQI and smoke estimates around 2004. This could indicate a significant event or change in environmental conditions during that time. After that, both metrics show a drop around 2020, suggesting a decrease in smoke or improvement in air quality. Overall, as you read through the graph from left to right, you can clearly see how smoke trends have changed over the decades, highlighting important shifts in air quality and environmental health.

Working on this assignment taught me some really valuable skills, especially in accessing and handling large databases. I learned how to access a large database and extract the information necessary to make graphs and conduct visualizations. This made me appreciate how crucial it is to ensure data accuracy when you're dealing with a lot of information. I also dove into modeling time series data, which was a great learning experience. Understanding the underlying assumptions of time series analysis—such as stationarity, seasonality, and autocorrelation—helped me to refine my approach. I realized that verifying these assumptions is essential to ensure the validity of the models used for forecasting and visualization. Through this process, I became proficient in using libraries like StatsModels and Matplotlib for time series analysis and visualization, enabling me to create informative graphs that clearly conveyed the trends and patterns in the data.

Collaboration was a huge part of this learning journey. Sharing ideas with my peers exposed me to different perspectives and techniques. For example, some of my peers were considering using interactive visualizations through the Plotly package. Although I did not incorporate that into my final plots, it was refreshing to learn about different visualization techniques and packages. Even though we sometimes had different opinions on interpreting the data, those discussions helped me think more critically and consider other viewpoints. Overall, this assignment was a great learning experience.