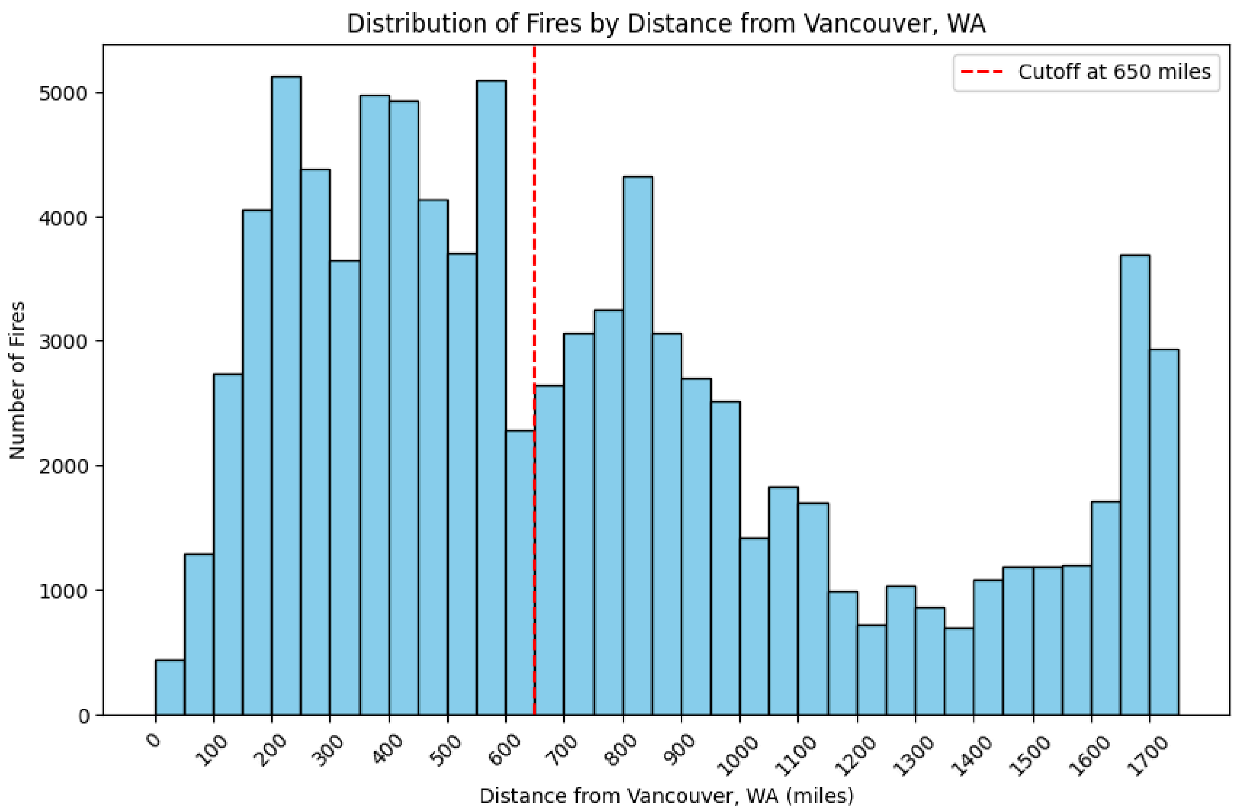
Project Report

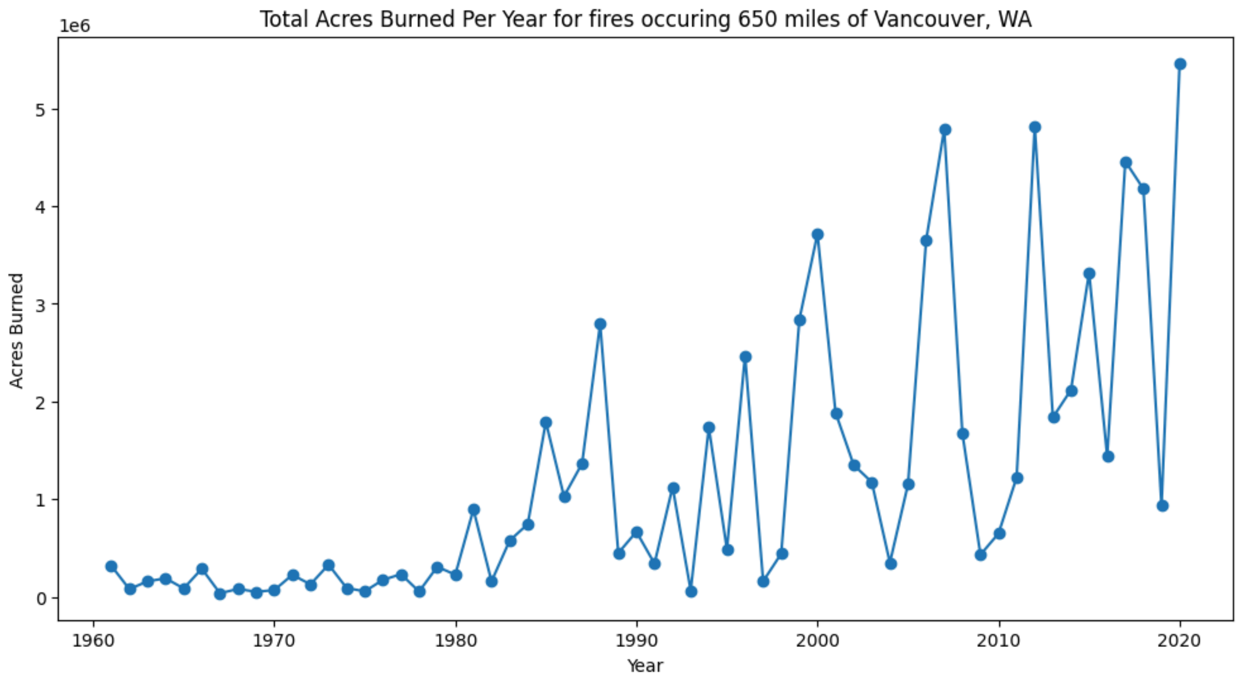
Question 1: Explanation for each visualization

Visualization 1: Distribution of Fires by Distance from Vancouver, WA



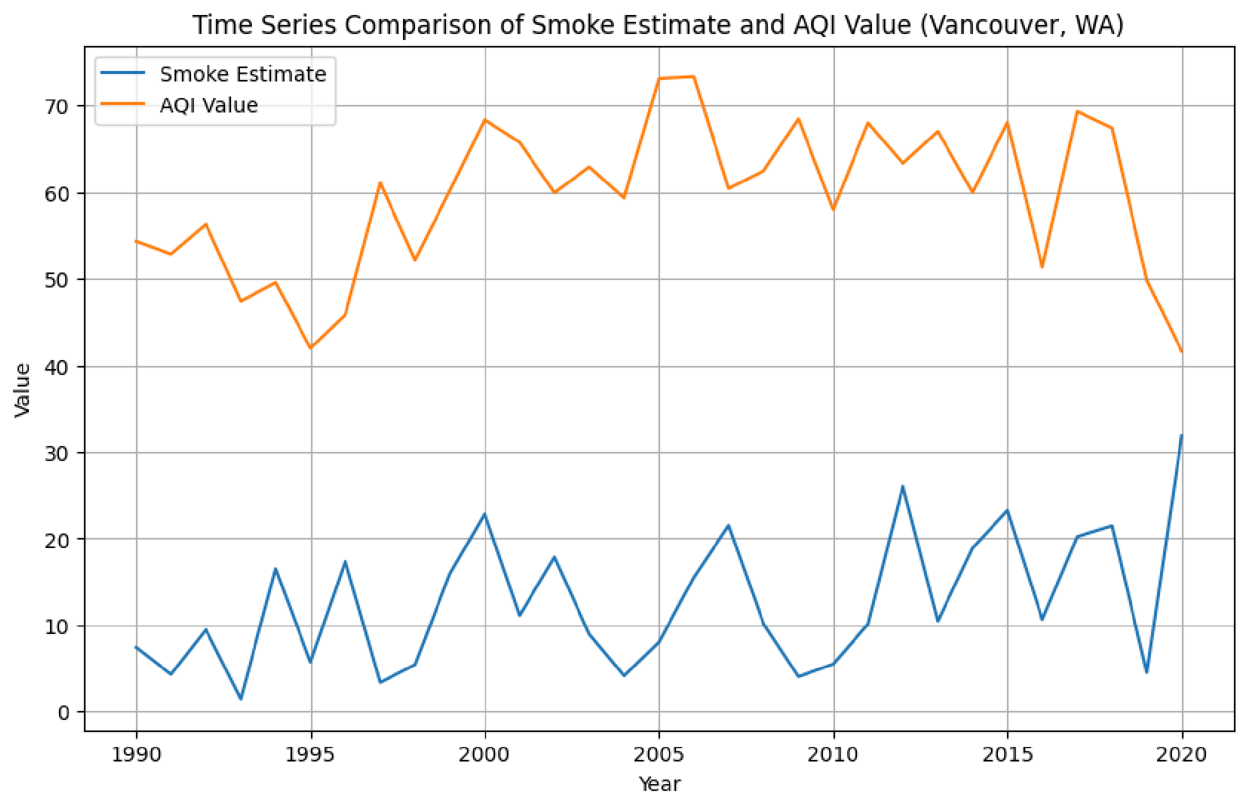
This histogram visualizes the distribution of fires based on their distance from Vancouver, WA. The x-axis represents the distance from Vancouver in miles, and the y-axis represents the number of fires observed at each distance range. To interpret the figure, the viewer can examine the height of each bar. A taller bar indicates a higher number of fires at that specific distance range. For instance, the tallest bar, located around the 100-mile mark, suggests that the highest number of fires occur within 100 miles of Vancouver. The underlying data consists of the average distance from Vancouver, WA to the occurrence of fire for all dates. The distances were grouped into 50-mile intervals, and the number of fires in each interval was counted to create the histogram distribution. The highest frequency of fires occurs within 100 miles of Vancouver, suggesting that the immediate vicinity of the city is most susceptible to fires. The histogram shows a general trend of decreasing frequency with increasing distance. This indicates that the number of fires decreases as the distance from Vancouver increases, which could be attributed to factors like fuel availability, topography, and human activities. The vertical dashed line at 650 miles represents a cutoff or limit in the data collection or analysis. The data beyond this distance was excluded or not considered in the analysis. The distribution of fires by distance from Vancouver, WA, somewhat appears to be right-skewed (though we have a lot more after the 1400 miles mark). This means that there is a long tail to the right, indicating that a majority of the fires occur closer to Vancouver, with a smaller number of fires occurring at greater distances. This distribution aligns with the expected pattern of fire occurrence. Typically, wildfires are more frequent in areas with higher population density, more human activity, and abundant fuel sources. As the distance from Vancouver increases, these factors generally decrease, leading to a lower frequency of fires. We have a lot more fires after the 1400 mark, probably referring to cities like LA in California where there are more fires reported.

Visualization 2: Total Acres Burned Per Year for Fires Occurring 650 Miles of Vancouver, WA



This line plot visualizes the total number of acres burned per year due to fires occurring within a 650-mile radius of Vancouver, WA. The x-axis represents the year, and the y-axis represents the total number of acres burned. To interpret the figure, the viewer can examine the position of the data points and the line connecting them. A higher data point indicates a larger number of acres burned in that particular year. The line connecting the points helps visualize the trend in the number of acres burned over time. The data consists of a dataset containing information about wildfires, including their distance from Vancouver, WA, size, the year they occurred, and various other factors for all dates. The data was processed to identify fires within a 650-mile radius of Vancouver, WA in the initial step. The total number of acres burned was calculated for each year, and these values were plotted on the graph. The graph shows significant fluctuations in the number of acres burned from year to year. This suggests that wildfire activity varies considerably over time, likely influenced by factors such as weather conditions, fuel availability, and human activities. While there are periods of high and low fire activity, there appears to be a general upward trend in the number of acres burned, especially in recent years. This could indicate an increase in the severity and frequency of wildfires. There are a few years with extremely high values, representing years with significantly larger wildfires. These outliers could be attributed to extreme weather events, such as droughts or heatwaves, or large-scale wildfires.

Visualization 3: Time Series Comparison of Smoke Estimate and AQI value for Vancouver, WA



This line plot visualizes the trends in smoke estimates and Air Quality Index (AQI) values for Vancouver, WA, over time. The x-axis represents the year (ranging from 1990 - 2020; we only have particulate and gaseous data from 1990), and the y-axis represents the value of both smoke estimates and AQI. To interpret the figure, the viewer can examine the position of the data points and the lines connecting them. A higher data point indicates a higher value for either smoke estimate or AQI. The lines help visualize the trends and fluctuations in these values over time. The underlying data consists of two datasets: one containing historical smoke estimates for Vancouver, WA, and another containing AQI values for the same period. The data was processed to align the time series and ensure consistency in the units of measurement and were grouped by year for consistency. Both datasets were then plotted on the same graph to facilitate comparison. The plot suggests a general co-variation between smoke estimates and

AQI values. This is expected, as higher smoke levels often correlate with poorer air quality and higher AQI values. I also computed the Spearman correlation to compare between both and it showed a weak positive correlation. Both smoke estimates and AQI values exhibit fluctuations over time. These fluctuations could be attributed to factors such as wildfire activity, meteorological conditions, and changes in emissions sources. There are a few years with significantly higher values for both smoke estimates and AQI. These outliers might correspond to years with severe wildfires or other extreme events that impacted air quality.

Question 2: Project Reflection

Working on this project was a great learning experience for me. I gained a better understanding of how wildfires affect air quality, especially how wildfire smoke impacts communities like Vancouver, WA. During this assignment, I improved my data analysis skills and learned how important teamwork is. I realized that sharing ideas with others can really help us think more critically.

Key Learnings from the Assignment

One major takeaway from this project was learning how to use time series models, especially the ARIMA model, to forecast environmental impacts. Before this assignment, I only knew about ARIMA in theory and had little practical experience with it. By actually working with the model and tuning its parameters, I realized how important it is to choose the right parameters (p, d, q) for making accurate predictions. The articles cited in the scripts helped me understand model evaluation and validation techniques better. Going through different parameter combinations showed me how statistical methods can improve predictions and influence policy decisions regarding air quality based on estimated smoke levels.

Creating smoke estimates also taught me about the assumptions needed in data modeling and their limitations. The challenge of accurately representing wildfire smoke made me aware of the simplified assumptions I initially missed. Using concepts from "What are the values p, d, q, in ARIMA?", I learned to consider factors like distance and fire type that can affect the relationship between wildfires and air quality. This helped me think more critically about the complexities researchers face when analyzing environmental data.

The Impact of Collaboration

Collaboration during this project significantly influenced my understanding and approach to solving the research question. The different viewpoints shared by my friends created richer discussions about the implications of fire data analysis and environmental science. When we talked about our methods and data choices, I saw how important it was to cross-reference our findings, which helped me understand how our individual projects fit into the broader environmental context. Insights from my peers often prompted me to reconsider my initial assumptions, especially regarding how to interpret data correlations, perform statistical analysis and model fitting.

Another example: in our discussions, we often highlighted the contextual factors that could affect air quality beyond the smoke estimates we were calculating. Hearing others share their thoughts on public health implications and different modeling methods made me realize the need for a more comprehensive analysis.