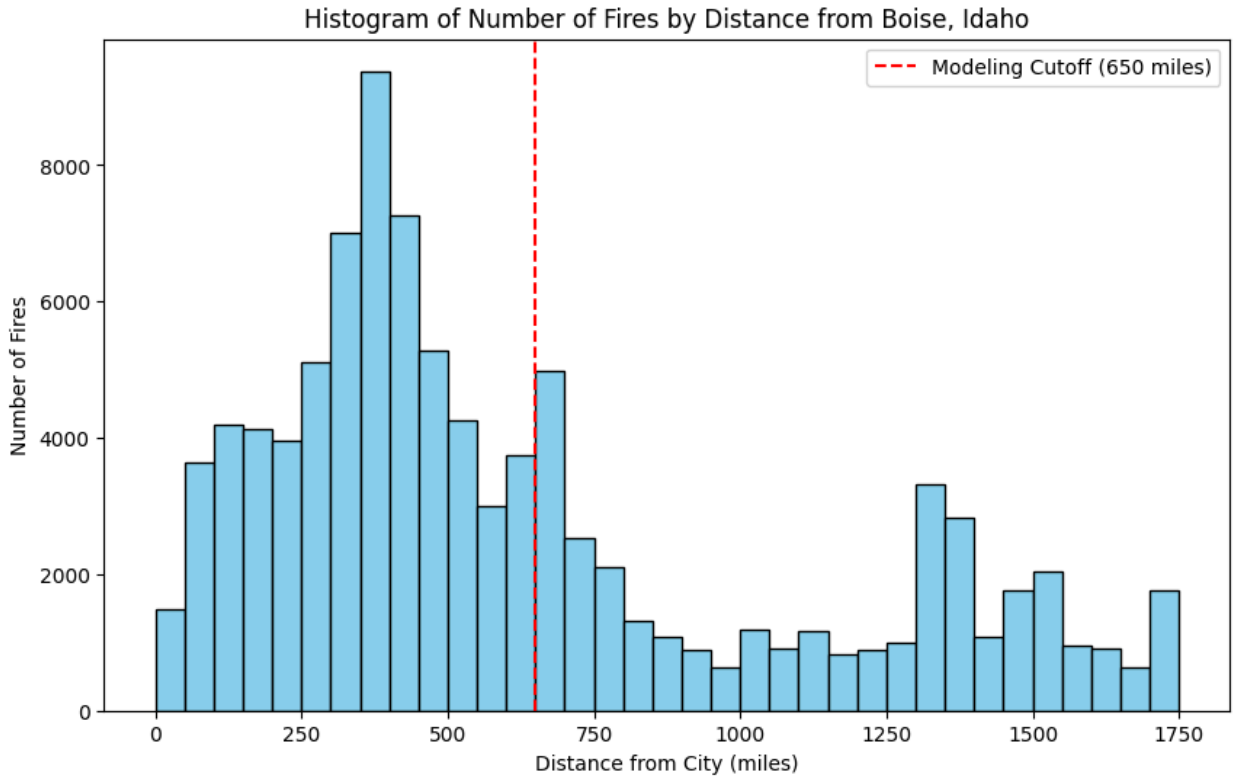
DATA 512
10/29/2024
Joobee Jung

Part 1 - Common Analysis

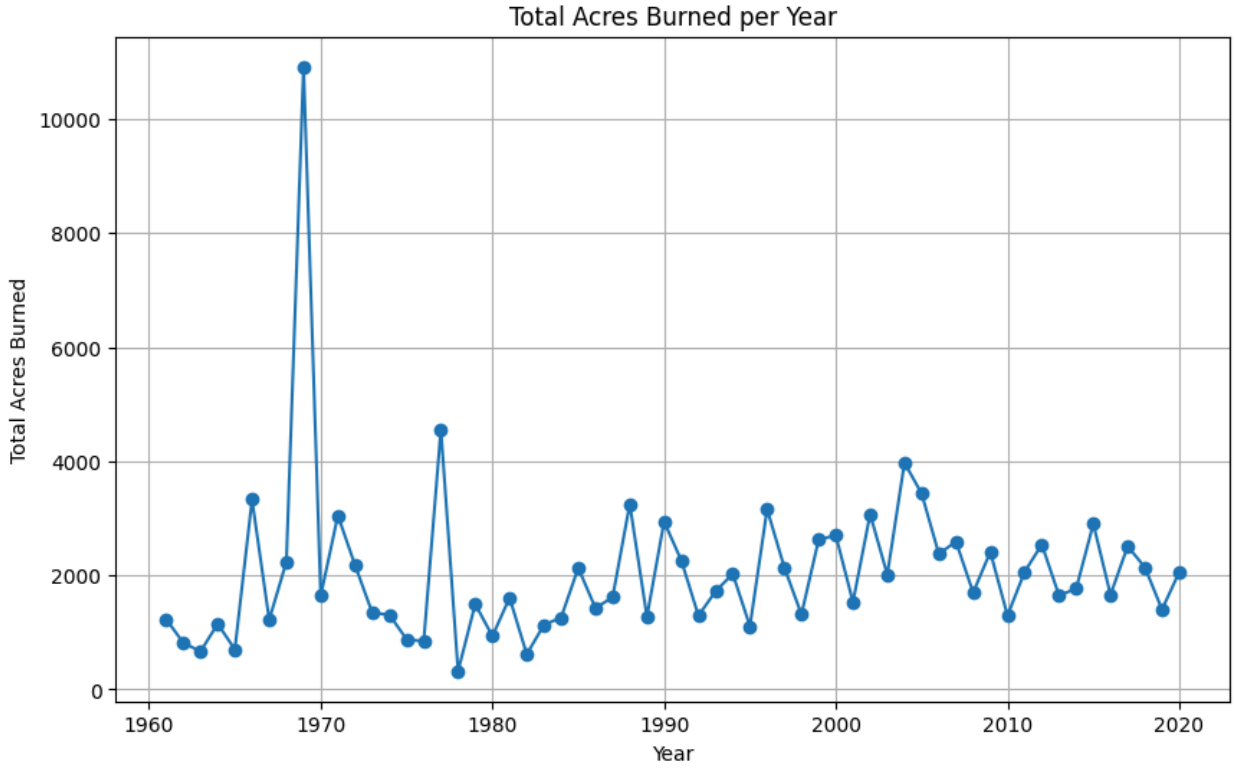
1. Data Visualization

a. Histogram of Number of Fires by Distance



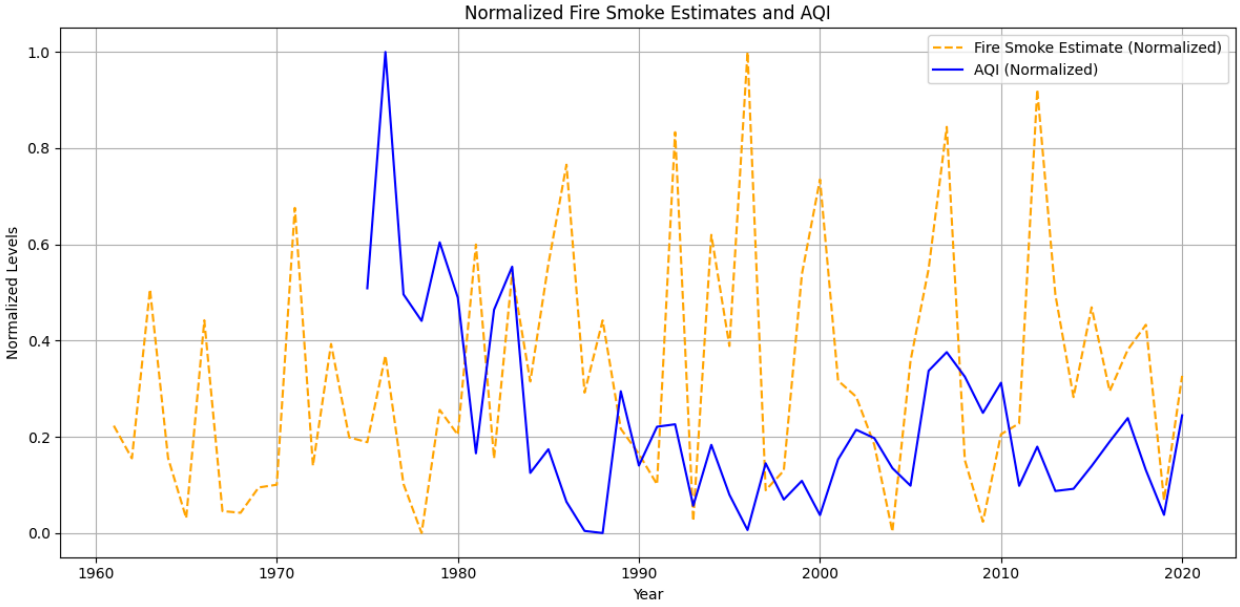
This histogram illustrates the distribution of wildfires based on their distance from Boise, Idaho, with a cutoff of 650 miles. The x-axis represents the distance from Boise, measured in miles, while the y-axis indicates the number of fires recorded. The data used to create this histogram was processed by binning distances into 50-mile intervals, allowing for a clear visual representation of how many wildfires occurred at various distances from the city. By analyzing this histogram, viewers can easily identify that most fires occur closer to Boise, emphasizing the importance of monitoring wildfire activity in proximity to urban areas.

b. Time Series of Total Acres Burned per Year



This time series plot displays the trend of total acres burned annually by wildfires. The x-axis represents the years, while the y-axis shows the average total acres burned. The data for this visualization was derived from grouping annual wildfire records to sum the acres burned each year. This allows viewers to observe fluctuations in wildfire activity over time.

c. Time Series of Fire Smoke Estimates and AQI Estimates



This analysis estimates smoke levels from wildfires in Boise, Idaho, using a Linear Regression model that identifies correlations between fire type and size by encoding categorical types as numeric values. I normalized the coefficients to clarify each fire type's influence, establishing a direct relationship with potential smoke output. Smoke estimates were calculated by incorporating the average distance to wildfires and converting fire size from acres to GIS hectares.

To estimate smoke levels, I developed a formula that combines the area affected by wildfires in GIS hectares with the average distance from Boise to the wildfire. This formula also includes a normalized weight for different fire types. The resulting smoke estimates were then merged with AQI data for further analysis and comparison to evaluate the model's accuracy.

The resulting time series plot compares the normalized fire smoke estimates with actual AQI values over time. The x-axis represents the years, while the y-axis shows the normalized levels of both fire smoke estimates and AQI. Normalizing the data allowed for a direct comparison of trends between the two variables, with fire smoke estimates based on distance from the city, fire type, and fire size, while AQI values were sourced from air quality monitoring stations.

2. Reflection

Through the collaborative activities in this assignment, I gained significant insights into the complexities of modeling smoke impacts on air quality. A key takeaway was the importance of integrating diverse datasets, such as fire data and air quality measurements, to develop a comprehensive understanding of how various environmental factors interact. This holistic approach allowed me to appreciate how elements like fire size, type, and distance contribute to changes in air quality, ultimately enhancing my ability to make informed predictions.

Collaboration played a crucial role in deepening my understanding of this multifaceted problem. Engaging in discussions with my classmate Kelly Wang about methodologies for estimating smoke levels clarified concepts such as feature engineering and the significance of data normalization in comparative analyses. For instance, the normalization process applied in the visualizations was inspired by our conversations about best practices for ensuring fair comparisons between disparate datasets. These collaborative dialogues not only enhanced my grasp of the techniques but also fostered an environment where ideas and strategies were freely exchanged, significantly improving my modeling approach.

I leveraged several coding techniques and methods from previous assignments, particularly in the implementation of linear regression for estimating smoke levels. Examining example code from notebooks like `wildfire_geo_proximity_example.ipynb` and `epa_air_quality_history_example.ipynb` from Dr. David W. McDonald proved invaluable in focusing my efforts on building a robust smoke estimation model.

Overall, this assignment underscored the value of collaboration in tackling complex research questions. It encouraged knowledge sharing and diverse perspectives, ultimately leading to more robust findings. By integrating insights gained from my peers and refining my coding techniques, I was able to navigate the complexities of the analysis and contribute to a richer understanding of how wildfires affect air quality in Boise, Idaho.