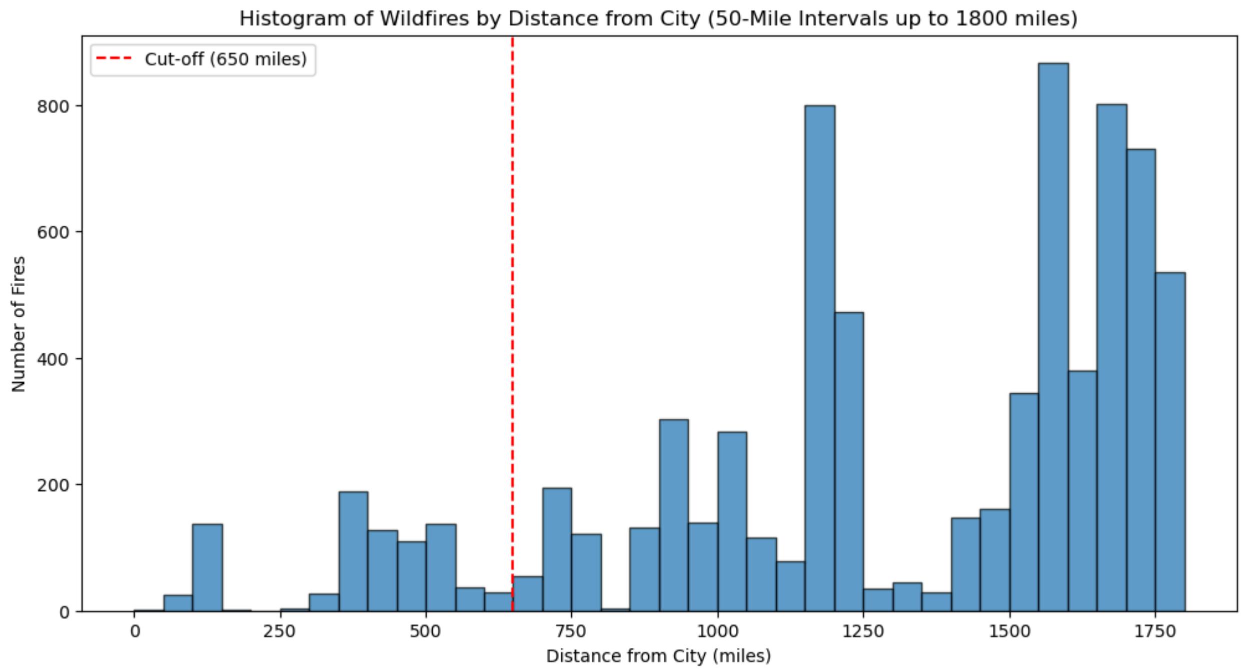
Trisha Prasant
10/30/2024
DATA 512

# Common Analysis Visualizations & Reflection



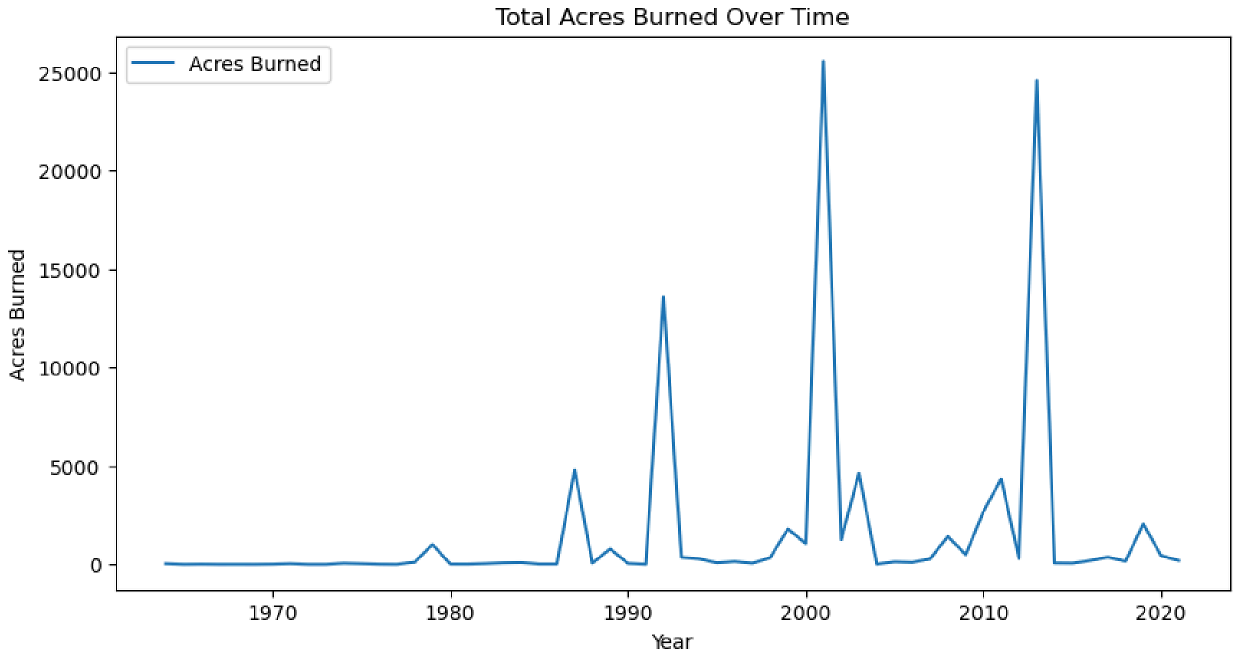

1. **Histogram of Wildfires by Distance from City (50-mile Intervals up to 1800 miles)**

This histogram displays the distribution of wildfire events based on their distance from the assigned city. The x-axis represents distance intervals in miles, divided into 50-mile bins, going up to 1800 miles. The y-axis shows the number of wildfires that occurred within each distance bin. A vertical red dashed line at 650 miles marks the distance cut-off used in the model for wildfire impact analysis, indicating that only wildfires within this range were considered for the model.

The histogram provides insights to the audience into how wildfire frequency varies with distance. Most wildfires appear to be concentrated in specific distance bands, most occurring beyond the 650-mile cutoff. This could suggest a geographic concentration of wildfire-prone areas (forests) at further distances from the city.

**Data Processing:**

The data for this visualization was filtered to include wildfires up to 1800 miles from the assigned city. The histogram was then generated by binning the distances into 50-mile intervals and counting the occurrences within each bin. This visualization provides a clear understanding of the distribution of wildfires by proximity to the city, highlighting the cutoff distance used for modeling.



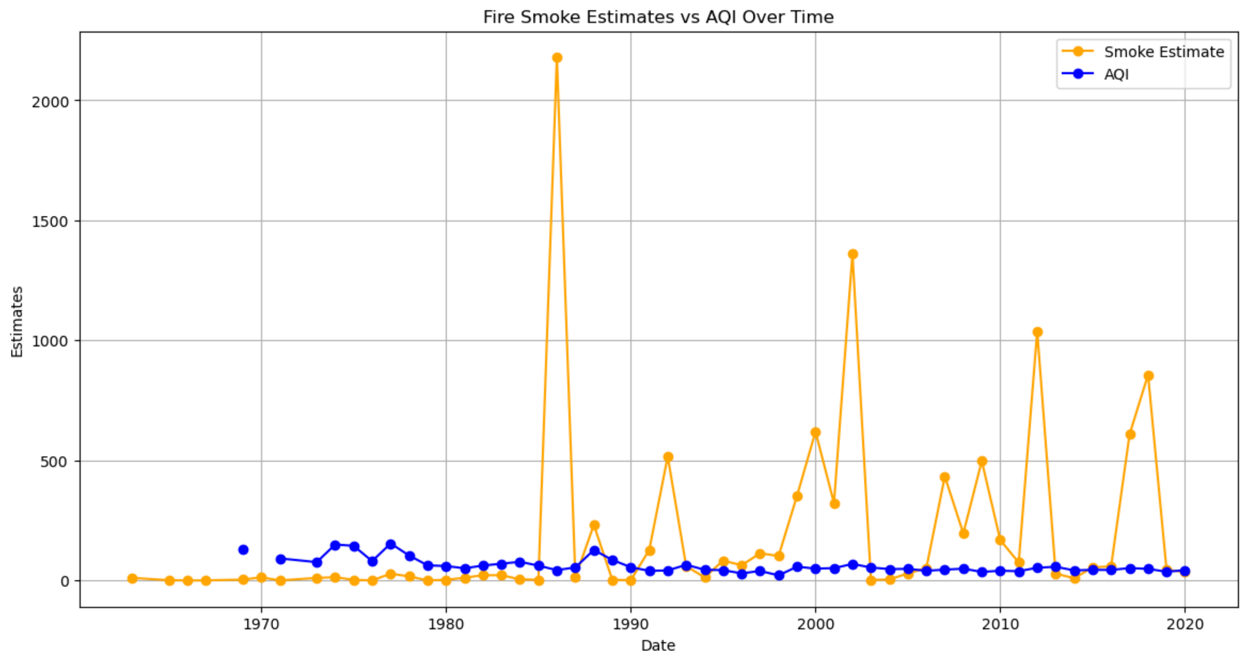

2. **Time Series of Total Acres Burned per Year (Fires within 650 Miles)**

This line plot shows the annual total acres burned by wildfires that occurred within 650 miles of the assigned city. The x-axis represents the years spanning the timeframe of the dataset. The y-axis shows the total acres burned each year. The line indicates the fluctuations in burned acreage over time, with sharp peaks indicating specific years with unusually high wildfire activity.

The visualization allows the viewer to identify trends or patterns in annual burned acreage over the years. Some notable spikes in specific years include 1993, 2002, and 2014, when large wildfire events occurred. The steady periods between peaks suggest years with relatively lower fire activity.

**Data Processing:**

The data was filtered to include only wildfires within 650 miles of the city. The burned acreage for each fire was summed by year to obtain annual totals. This aggregation enables my audience to compare total burned areas year by year, emphasizing high-activity years and long-term trends in wildfire impact within the 650 miles from Hartford, CT.




### 3. Fire Smoke Estimates vs. AQI Over Time

This time series graph compares the smoke estimates derived from wildfire data with the US EPA Air Quality Index (AQI) recorded for the city over the years. The x-axis represents time (or the discovery date of the wildfire), while the y-axis represents the magnitude of the estimates. The orange line represents the modeled smoke estimates based on wildfire activity, while the blue line represents the average AQI values recorded by AQI monitors near Hartford.

This graph provides a visual correlation between wildfire smoke and AQI levels, allowing the audience to observe how changes in wildfire smoke estimates align with differences in air quality. Peaks in the smoke estimate line correspond to high wildfire activity, which might correlate with higher AQI readings — indicating poorer air quality, possibly due to smoke.

**Data Processing:**

The smoke estimates were derived using a formula that considers each fire's acreage (a) burned, distance from Hartford (b), and duration of the wildfire (c): (a + c / b). These estimates were aggregated to match the dates of AQI readings, and both were plotted together for comparison. This visualization highlights the relationship between wildfire activity and air quality in the city, emphasizing the potential impact of nearby wildfires on air pollution.

**Reflection Statement**

In the Common Analysis portion of my final project, I aimed to analyze wildfire data and predict future smoke impact on my assigned city — Hartford, CT, using various computational and and forecasting techniques. This analysis involved obtaining and filtering the data from the USGS (United States Geological Survey), calculating distances using GIS systems, constructing a smoke impact estimate, retrieving air quality data via US EPA API, and developing an ARIMA model to forecast future impacts in the next 25 years. Through collaboration and consultation with my peers, as well as guidance from my professor's resources and online research, I was able to implement each step effectively.

Collaboration with my peers was extremely valuable in navigating the initial stages of the project. I got a clearer understanding of how to accurately wrangle our provided dataset and how to approach the geodetic distance computation portion of this assignment. One of my peers and I discussed some of the logic that should be used when calculating geodetic distances based on latitude and longitude coordinates, which helped me set up my distance filtering function. This insight improved my data filtering by ensuring I only included relevant wildfires within the specified distance, strengthening the foundation for the rest of the analysis. I also got insights as to how geographic data is structured and how I can wrangle it to suit my needs. These initial conversations saved me a lot of time and stress at the beginning stages of the assignment.

Additionally, while calculating smoke impact estimates, I engaged in discussions with classmates to decide on a reasonable formula. These conversations helped me finalize an equation that factored in the acreage burned, the distance from the city, and the number of days the wildfire was burning, which produced a more meaningful representation of the smoke's potential impact on air quality. Although I didn't directly share code, this exchange of logic and methodology guided my approach and made me feel more confident in my smoke impact estimate calculation.

For the other parts of the analysis, particularly in obtaining air quality data using the US EPA API, I relied on the example API code provided to our class by Dr. McDonald. This example code outlined the structure of API requests and data extraction, simplifying the gathering of relevant AQI data for my city. The professor's code was an essential resource, allowing me to focus on interpreting the data rather than getting caught up in the technical jargon of API integration.

The forecasting component of the project required additional outside research, particularly when choosing a suitable model. I explored various online resources, which gave me a starting point for effectively implementing ARIMA and SARIMA models. I consulted with online sources, listed in the *Outside Sources* section, which clarified the assumptions made during ARIMA forecasting, seasonality handling, and how to plot my results with confidence intervals.

In addition to this outside research, I discussed other possible time series forecasting methods with my peers, including exponential smoothing and Prophet from Meta. While I ultimately decided on the SARIMA approach due to its fit with my dataset's seasonal trends, these conversations broadened my understanding of time series modeling.

The collaborative aspect of this project greatly enhanced my understanding of several technical aspects. Discussing logic and techniques with my peers pushed me to think more critically about my choices, specifically in calculating smoke impact and selecting a forecasting model. The exchanging of ideas provided me with more options and more diverse perspectives, allowing me to refine my approach beyond what I might have done by myself. Also, having access to example code and online resources allowed me to fill in knowledge gaps without relying solely on trial and error.

This project showed me the value of collaboration in technical problem-solving. While I worked on and developed my solution independently, my discussions with peers provided me with clarity and strengthened my understanding of the data and methods, resulting in a stronger and more informed analysis.