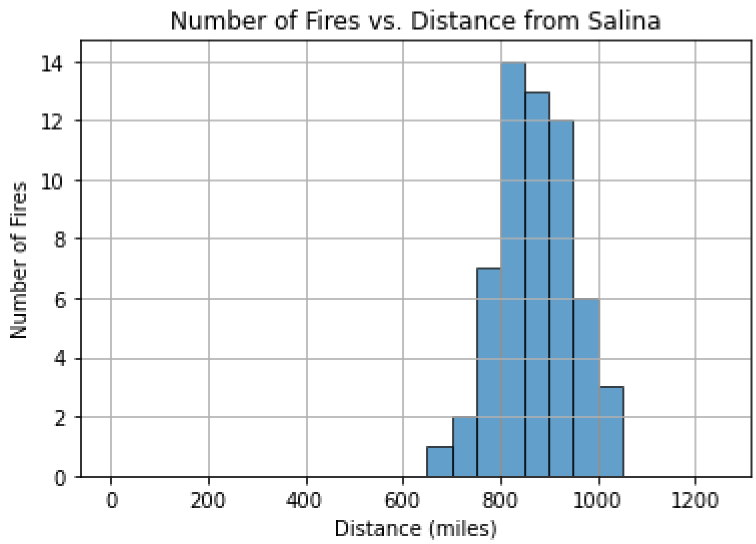
Figure Descriptions



The histogram gives a visual overview of how many fires happened within each 50-mile zone starting from Salina and extending up to a maximum distance of 1250 miles.

When you're looking at the graph, you naturally scan it from left to right, following the x-axis, which shows distances measured in miles. Each vertical bar in the histogram corresponds to a 50-mile span away from Salina. The height of each bar tells you how many fires occurred within that specific distance range. A taller bar indicates a greater number of fires, whereas a shorter bar means fewer fires in that area.

The horizontal x-axis serves as the "Distance (miles)," providing a sense of how far away these locations are from Salina, all measured in miles. Meanwhile, the vertical y-axis, labeled "Number of Fires," provides a numerical count of how many fires were reported within each 50-mile distance zone.

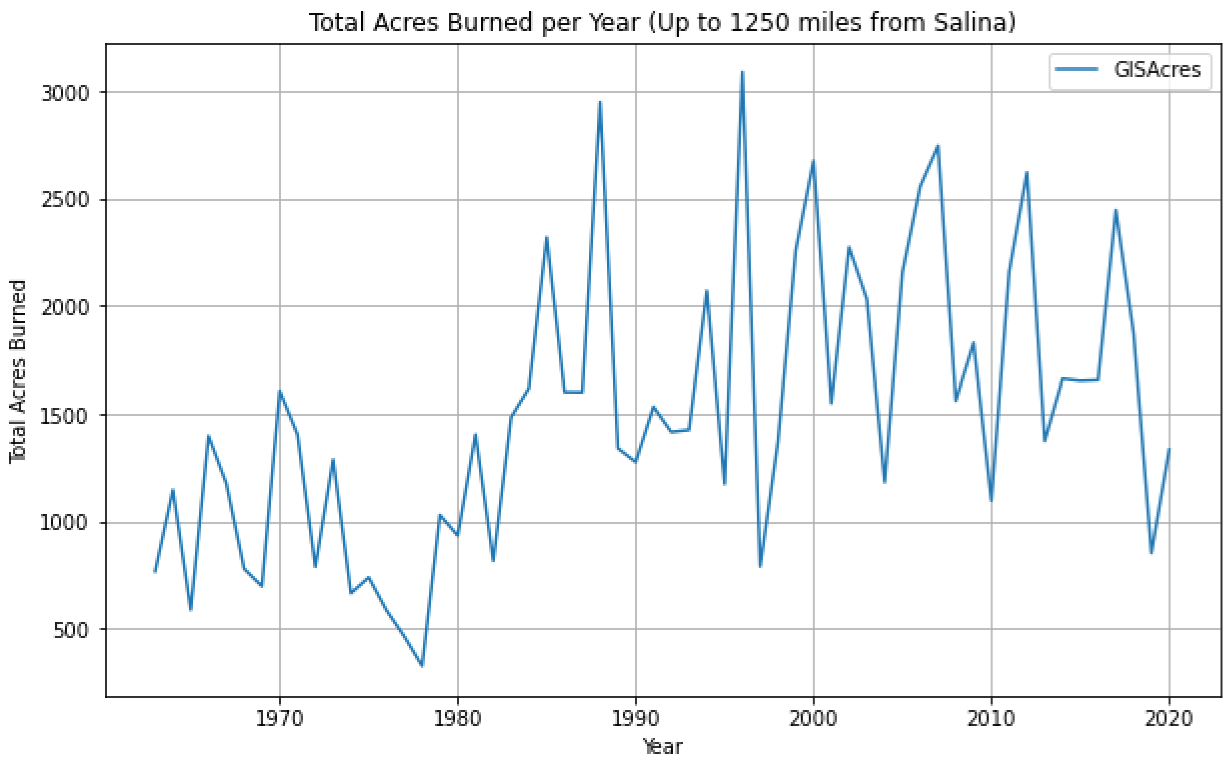
This histogram doesn't follow the usual bell-shaped distribution seen in many datasets. Instead, it has a peak around the 800-mile mark from Salina. It's noteworthy that most fires were concentrated between 800 and 950 miles from Salina. Interestingly, even the closest fires were located more than 600 miles away. This pattern might be influenced by the presence of the nearby state of Oklahoma. According to data from the state fire marshal's office, there were 17,499 reported wildfires in Oklahoma between 2000 and 2007.

The data used here is a combination of two sources:

1. Monthly Air Quality Index (AQI) estimates for Salina, sourced from the US Environmental Protection Agency (US EPA). This dataset provides year-wise records of air quality in Salina, including columns for the year, month, and the corresponding AQI values.

2. Information regarding the estimated impact of smoke from wildfires. This dataset contains details such as the number of acres burned, the shortest distance from the city, and a smoke impact estimate. It is derived from a comprehensive USGS dataset that consolidates wildfire information from approximately 40 different sources. The specific focus is on the period after 1963 and within a radius of 1250 miles from Salina.

To enhance data completeness and accuracy, imputation techniques are applied to fill any missing AQI values. Additionally, the dataset is segmented into 50-mile distance intervals to analyze the spatial distribution of wildfires concerning their proximity to Salina.



The graph visually represents the yearly total acreage consumed by fires within a designated distance from Salina, with the maximum limit set at 1250 miles.

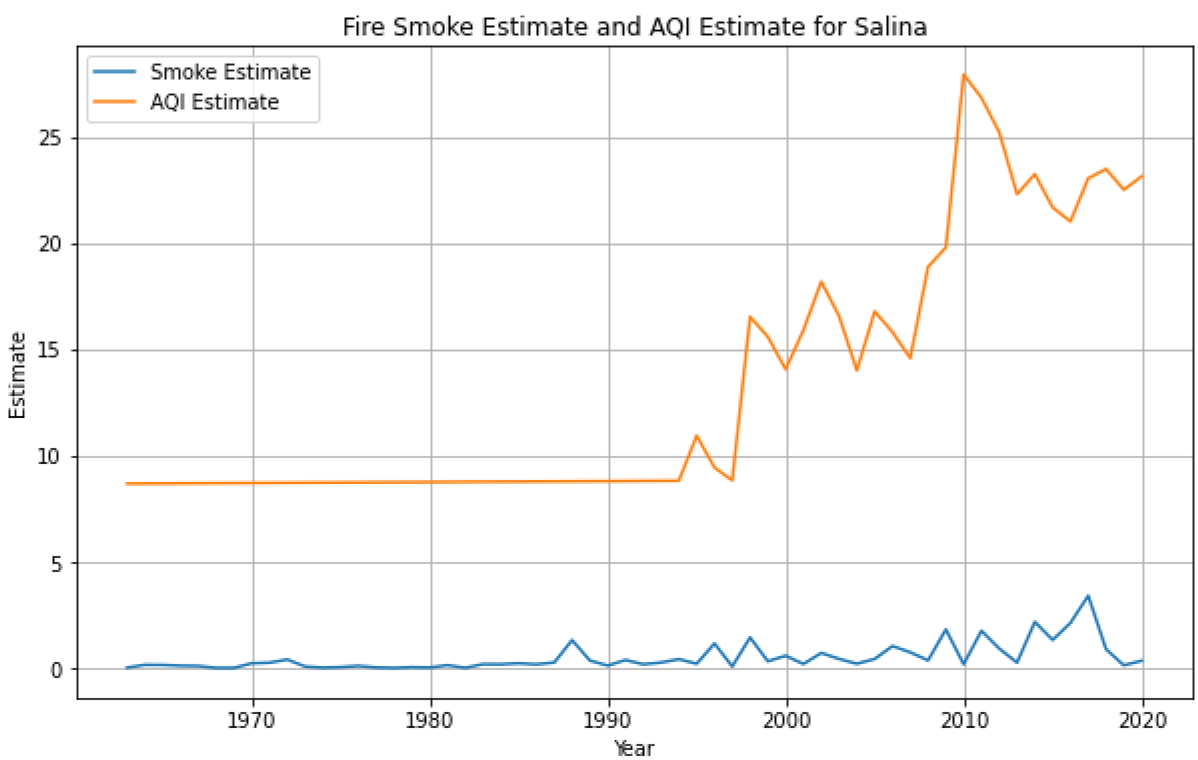
To understand the graph, you follow the horizontal timeline on the x-axis, denoting the years, and progressing from left to right. Each data point on the graph corresponds to a specific year, and the vertical axis measures the cumulative acreage burned. As you examine the graph, you can track how the total burned acreage changes from one year to the next within the specified distance range.

The horizontal axis represents "Year," marking the passage of time. The vertical axis quantifies the "Total Acres Burned," indicating the combined area affected by fires within the defined bounds.

Now, as you analyze the graph, you'll notice that the data points show fluctuations in the total burned acreage. Some years witness a relatively smaller impact, with acreage falling within the range of 1000 to 2500 acres. However, in certain years, the situation can be quite different. For instance, let's say you come across the year 1996. In that specific year, the graph reveals a peak that exceeds 3000 acres. This tells you that 1996 was a particularly challenging year when wildfires caused substantial land damage in the vicinity of Salina.

One additional point of note is that the graph's accuracy may be somewhat limited for years before 1980. During that period, the technology and methods available for recording and mapping wildfires were not as advanced as they are today. Consequently, data from that era may not provide a complete or entirely accurate picture of the actual acreage burned in wildfires.

The base data used here is the same as the earlier graph. To reiterate, the data combines Monthly Air Quality Index (AQI) estimates from the US Environmental Protection Agency (US EPA) and wildfire impact data from the USGS. The wildfire data focuses on fires within a 1250-mile radius of Salina after 1963. We've filled in missing AQI values, and the data is organized into 50-mile distance intervals for analysis. Further, the total acres burned per year were then aggregated, and calculated as the sum of burned acres for each year. The resulting dataset is used to generate the time series graph, providing a visual representation of the yearly variation in acres burned within the specified proximity to the city.



In this figure, we've got a time series graph showcasing two important estimates: the "Fire Smoke Estimate" and the "AQI (Air Quality Index) Estimate" for Salina. This graph helps us see how these estimates change over time.

Now, to make sense of this, take a look at the horizontal x-axis – that's the one running from left to right and representing the years. On the other hand, the vertical y-axis measures the "Estimate" values. You'll notice two lines on the graph, each marked with a legend: "Smoke Estimate" and "AQI Estimate." As you follow these lines, you can compare the trends and fluctuations in these two estimates over the years.

The x-axis, which represents the "Year," provides us with a timeline of events. Meanwhile, the y-axis tells us about the "Estimate" values, helping us gauge the Fire Smoke and AQI Estimates for Salina. The two lines you see on the graph correspond to these specific estimates, with the legends showing which line represents which estimate.

At first glance, there doesn't seem to be a strong positive correlation between these estimates on the graph. However, it's worth considering further analysis to explore any potential relationships between these two estimates more comprehensively.

This code calculates the total acres burned each year, focusing on fires within a specified distance of Salina. The data is processed by grouping the filtered dataset by year and summing the 'GISAcres' column for each year. Once again, we have used the same data that has monthly AQI estimates and smoke impact data that has undergone transformations.

The data processing steps, intermediate data, and data extraction steps have been elaborated on much further in the notebooks in the src folder as well as the README.md for this project's repository.

Reflection Statement

Throughout this assignment, I had some significant takeaways that I'd like to share. Firstly, I learned the importance of managing my time effectively. This project was a bit overwhelming at first, so I decided to tackle it step by step, dedicating a bit of time each day. For instance, I spent two whole days on documentation early on. This approach turned out to be much better than leaving everything to the last few days. It took me a total of eight days to complete the project, but this way, I felt more in control.

Additionally, I discovered the power of creative problem-solving. Even seemingly simple tasks, like loading a JSON file or extracting AQI data, presented their own challenges. I had to dig into research and scripting to get things right. Another interesting part was coming up with a smoke estimate. With limited data, it was a puzzle to find a meaningful metric. But the process of figuring it out was quite enlightening.

Collaboration with my peers played a crucial role. Especially when it came to data extraction, it was beneficial. For instance, I initially attempted to load the data from the JSON file without referring to Professor David's example with my friends Rhea and Shweta. That turned out to be a mistake, and his example proved to be a lifesaver. Also, discussing issues related to extracting data about gaseous and particulate matter with others helped me make informed decisions. This led me to analyze only one type of data, which made more sense given the limited data I had to work with. It was reassuring to find that my friend Nizan was pursuing a similar approach.

In a nutshell, this assignment taught me valuable lessons in time management and creative thinking. The collaborative aspect of the project had its ups and downs, but it was a big help, especially in handling data extraction challenges. Overall, it was an enriching experience, and it reinforced the importance of structured work methods and teamwork when tackling complex tasks.