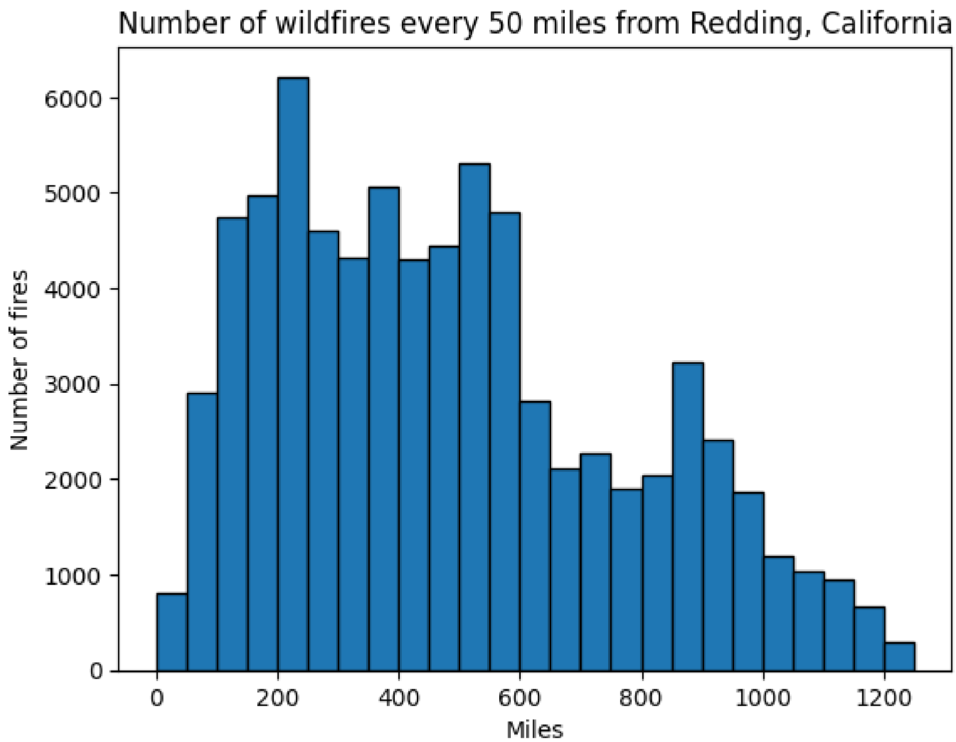
Visualizations

1. **Produce a histogram showing the number of fires occurring every 50-mile distance from your assigned city up to the maximum specified distance.**



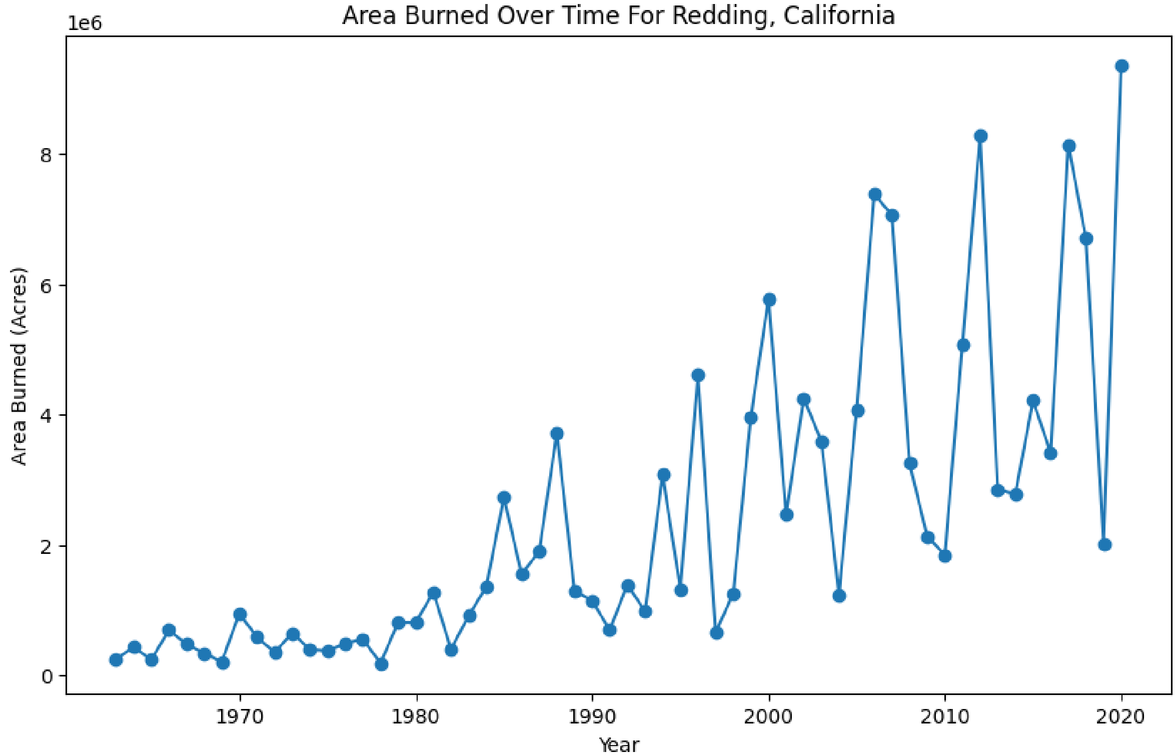
The figure is a bar chart that shows the frequency of wildfires within various distances from Redding, California. To read the figure, one would look at the bars and note their height, corresponding to the number of fires occurring within each 50-mile increment from Redding.

The horizontal axis (x-axis) represents the distance from Redding, measured in 50-mile increments. The vertical axis (y-axis) represents the number of wildfires, with the height of each bar corresponding to the number of fires in each distance range.

The underlying data represents the count of wildfires observed or recorded over a period (1963-2020) and has been binned according to distance from the shortest point of the fire ring. A ring is a geometrical polygon of the area that was impacted by the fire. The shortest distance from Redding to any point of the ring is considered as the distance. The data processing involved collecting individual records of wildfires, determining their distance from Redding, and then tallying them into these 50-mile bins to visualize the distribution of wildfire frequency relative to Redding.

In summary, the chart provides a visual distribution of wildfire occurrences. It shows variation in wildfire frequency with distance from Redding, with certain distances showing higher frequencies than others.

2. Produce a time series graph of total acres burned per year for the fires occurring in the specified distance from your city.



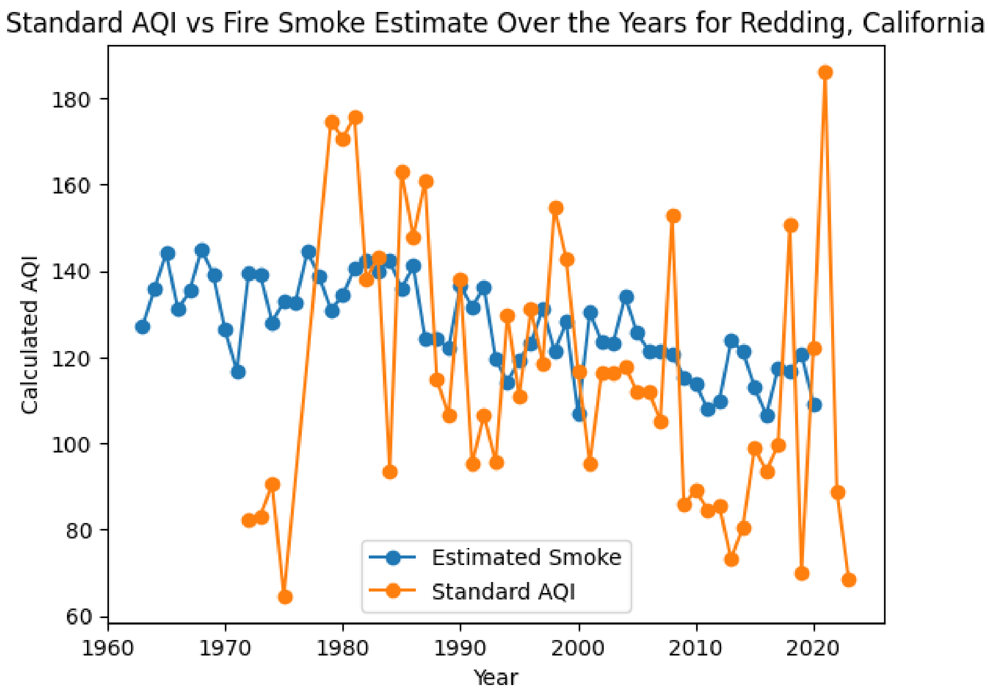
The figure is a line graph depicting the area burned by wildfires over time in Redding, California. The horizontal axis (x-axis) represents time, with years from around 1963 to 2020. The vertical axis (y-axis) represents the area burned, which is measured in acres as indicated by the "1e6" notation, suggesting the data is in the millions of acres.

To read the figure, the viewer would trace the line with the dots, where each dot represents the area burned in a specific year. The height of the dot above the horizontal axis corresponds to the total area burned that year.

The underlying data comes from records of wildfires in the region around Redding, California. The data is collected annually and then summed up to give the total area burned each year.

The line graph provides a visual representation of trends in wildfire activity over time, showing fluctuations and potentially allowing for the identification of any patterns or changes in the frequency or severity of wildfire events in the area. The sharp peaks suggest years with huge wildfires, while the lower points indicate years with less area burned.

3. Produce a time series graph containing your fire smoke estimate for your city and the AQI estimate for your city.



This graph compares the standard Air Quality Index (AQI) and estimated smoke presence in Redding, California from 1963 to 2023. The blue line represents "Estimated Smoke," and the orange line represents "Standard AQI." Peaks indicate higher AQI or estimated smoke, while dips indicate lower values. The standard AQI data is from AQS EPA API, measuring various pollutants, while estimated smoke data is from the USGS Wildfire Dataset. For AQI calculation, daily data is converted to yearly values. The process involves considering multiple sensors, up to 8 pollutants, and averaging the worst 5 days' maximum AQI per year.

For estimating smoke, the data is derived from the USGS Wildfire Dataset. The process involves assessing each fire feature, considering factors such as distance from Redding (capped at 1250 miles) and hectares burnt. The distance is normalized to a range of 0-1, penalizing the overall score based on proximity. To address extreme values in hectares burnt, the 75th percentile value (204) is used. This method provides a more accurate estimation of the fire smoke's effect on air quality, especially considering the extreme events.

The processed data is then utilized for visualizations, such as the line graph comparing standard AQI and estimated smoke over time, offering insights into the relationship between wildfires and air quality in Redding, California.

This graph provides insights into the correlation between AQI and estimated smoke, especially during years with significant smoke peaks, potentially linked to wildfires.

Collaboration Reflection

Working together on this assignment wasn't just about ticking off tasks; it genuinely transformed how I tackle research questions and problem-solving. The collaborative journey revealed the profound impact of diverse perspectives on analytical methodologies. Discussing ideas with peers not only challenged my initial assumptions but also broadened the analytical scope, bringing fresh angles to the table.

A pivotal learning moment came from delving into David McDonald's work, where I uncovered advanced techniques for data extraction and geojson calculation. This set the tone for the entire project. Collaborative discussions with Aaditya and Professor David McDonald deepened my understanding of processing daily pollutant data and formulating a strategy for deriving a yearly Air Quality Index (AQI) estimate. The decision to incorporate the worst 10-day maximum AQI approach directly stemmed from these fruitful collaborations.

The collaborative environment fostered not just knowledge-sharing but also a culture of efficient coding practices. I meticulously documented specific attributions to David McDonald's contributions within the codebase, ensuring transparency and giving credit where it's due. Additionally, I borrowed the logic and data model for processing yearly AQI data from Aaditya, further enhancing the efficiency of our coding endeavors.

In the quest for a formula to estimate fire smoke, a discussion with Neel became a defining moment. Initially, I toyed with percentages for both factors, but it felt too limited in capturing the nuances of different years. Collaborating with Neel led to the realization that using area as a base and distance as a penalizing factor offered a more nuanced and effective approach. This collaborative brainstorming and validation process refined the formula significantly.

However, collaboration wasn't all seamless. Coordinating schedules introduced its own set of challenges. This became particularly evident when attempting to commit a large file on GitHub. During a discussion with Aaditya, I learned that GitHub had a file size limitation of 100MB, and when I tried to commit a 200MB JSON file, the snag in the process prompted a clone, file extraction, and a fresh commit—an episode underscoring the pivotal role of effective communication and organization in collaborative endeavors.

In response to collaborative suggestions, integrating Git Large File Storage (LFS) emerged as a practical solution to circumvent the challenge of handling large files on GitHub. Additionally, exploring Git BFG Repo-Cleaner as an alternative for managing large files significantly streamlined the version control process. These collaborative solutions not only unraveled technical challenges but also showcased the collective problem-solving prowess within the team. While the issue had already been resolved with the new clone, the collaboration was invaluable in learning about potential pitfalls and solutions.

In reflection, collaboration played a transformative role, not just enhancing the quality of the analysis but reshaping my entire approach to problem-solving. The amalgamation of diverse perspectives borrowed methodologies, and efficient coding practices stemming from collaborative interactions significantly fortified the robustness of our research. This experience underscored the paramount importance of open communication, meticulous documentation, and the synergistic exchange of ideas in collaborative projects.