Visualizations and Reflection

Aditi Kharkwal

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

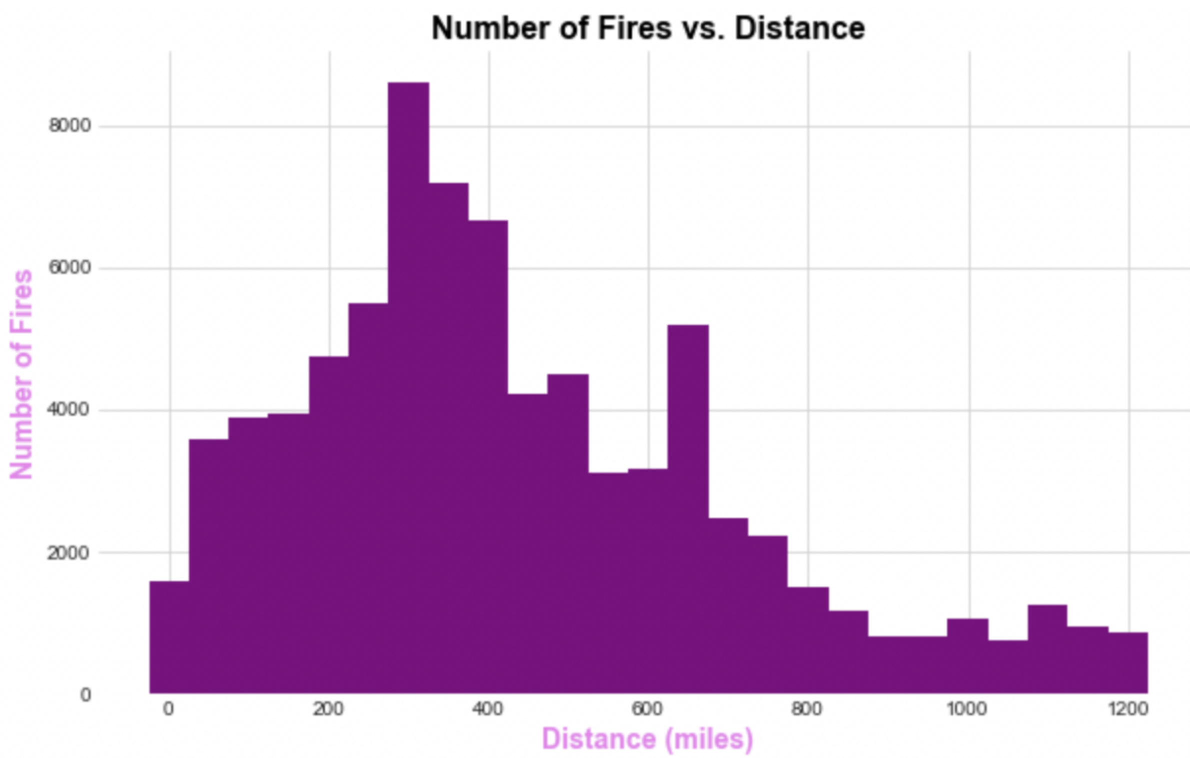

Figure Description: This histogram visualizes the distribution of wildfires based on their distance from the assigned city, Cladwell. The x-axis represents distance in miles, divided into bins of 50-mile intervals. The y-axis shows the number of fires that occurred within each distance range. What I observe from this graph is that the number of fires around my city - 'Cladwell', have peaked at around 300-450 miles.

Interpretation: To read this figure, start from the left side of the x-axis, which represents the distance from Cladwell in miles. Each bar in the histogram corresponds to a specific range of distances (e.g., the first bar represents fires occurring between 0 and 50 miles from Cladwell). The height of each bar indicates the number of fires that occurred within that distance range. This visualization provides insights into the spatial distribution of wildfires relative to Cladwell. For example, if there is a peak in the histogram at a particular distance range, it suggests that a significant number of fires occurred within that distance from the city.

Underlying Data and Processing: We are using the wildfire data to estimate the impact fire size vs distance. It counts the number of fires within different distance ranges from a specific city. Here's an explanation of the underlying data and how it was processed:

- Wildfire Feature Data: The underlying data consists of wildfire features, each containing attributes like fire year (wf_year), GIS acres burned (wf_size), assigned fire type (fire_type), and geometry data (ring_data) representing the fire's perimeter. These features are part of a dataset containing information about various wildfires.
- Smoke Impact Estimation: The calculate_smoke_impact function computes an estimate of smoke impact based on the fire size, distance from the city, and fire type. The impact is calculated using the formula: smoke_impact = fire_size / distance.

- Annual Smoke Impact Estimates: A dictionary named annual_smoke_impact is defined to hold the annual smoke impact estimates. For each wildfire feature, if it meets certain conditions (e.g., within the last 60 years, during the fire season), the smoke impact is calculated and added to the corresponding year's list in annual_smoke_impact.

- Fire Counts and Distance Ranges: A dictionary named fire_counts is created to count the number of fires occurring within specific distance ranges from the city (in 50-mile intervals). For each wildfire feature, the distance from the city is calculated, and the corresponding distance range is determined. If the distance range is within the predefined range (up to 1250 miles), the count for that range is incremented.

- Time Tracking: The code includes time tracking using the time module to measure the elapsed time for processing features.

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

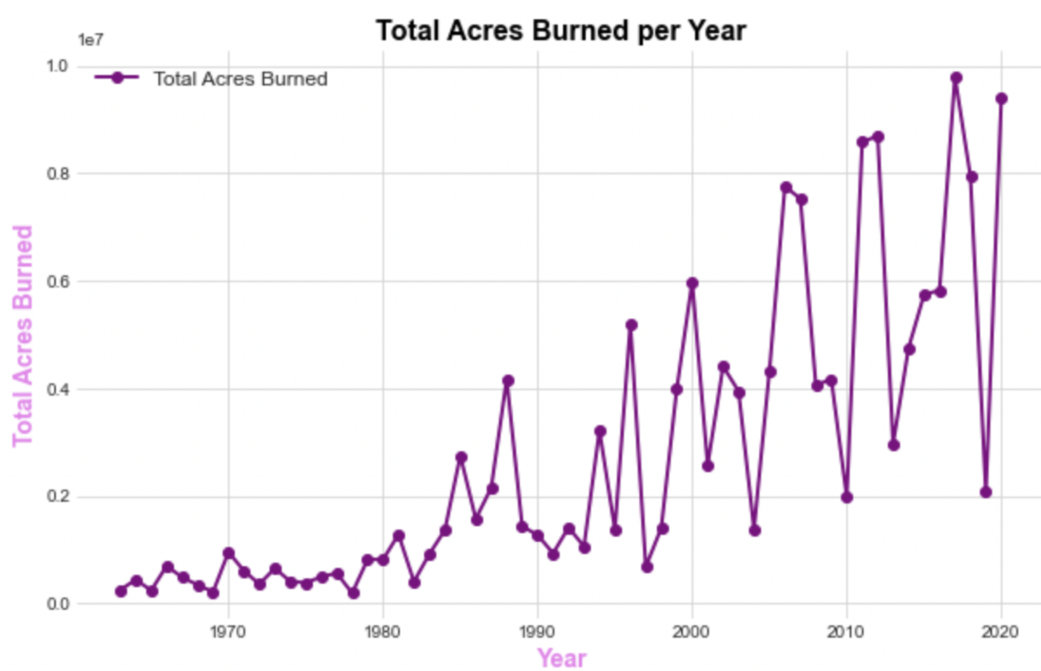

Figure Description: This time series graph presents the total number of acres burned by wildfires per year for fires occurring within the specified distance from Cladwell. The x-axis represents the years, while the y-axis indicates the total acres burned (we need to multiply the numbers on y axis by y-axis * 10000000). We can clearly see that with time the total acres burned has increased from this data for the city which was assigned to me - 'Cladwell'.

Interpretation: To interpret this figure, observe the x-axis, which denotes the years, and the y-axis, which represents the total acres burned. Each point on the graph corresponds to a specific year, showing the cumulative acres burned for fires within the specified distance from Cladwell. This visualization provides insights into the temporal trends and patterns of wildfire activity in terms of total acres burned per year within the specified radius from Cladwell.

Underlying Data and Processing: The data for this visualization the following variables were used: wf_feature['attributes']['Fire_Year']: The year in which the wildfire occurred. wf_feature['attributes']['GIS_Acres']: The number of acres burned by the wildfire. wf_feature['attributes']['Assigned_Fire_Type']: The assigned type or category of the wildfire. wf_feature['geometry']['rings'][0]: The geometry data representing the perimeter of the wildfire.

The data processing involves the following steps:

- Iterating Through Wildfire Features: The code iterates through a list of wildfire features (feature_list) to access information about each individual wildfire event.

- Filtering by Year and Fire Season: For each wildfire feature, the code checks if the fire year (wf_year) falls within a specified range (between 1963 and 2023) and if it occurred within the fire season.
- Calculating Distance from City: The code uses the function average_distance_from_place_to_fire_perimeter to compute the shortest distance from the specified city (Cladwell) to the perimeter of the wildfire.

- Checking Distance Threshold: It ensures that the distance to the fire is within 1250 miles, as specified in the code.

- Calculating Smoke Impact: The calculate_smoke_impact function computes an estimate of the smoke impact based on fire size, distance, and fire type. This impact is calculated using the formula smoke_impact = fire_size / distance.

- Aggregating Data for Visualization: The code maintains two main data structures: annual_smoke_impact: A dictionary that holds lists of smoke impacts for each year. If the year is not present in the dictionary, a new entry is created; otherwise, the impact value is appended to the existing list. acres_burned_annual: A dictionary that keeps track of the total acres burned per year. If the year is not present, a new entry is created with the acres burned value; otherwise, the value is added to the existing total.

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

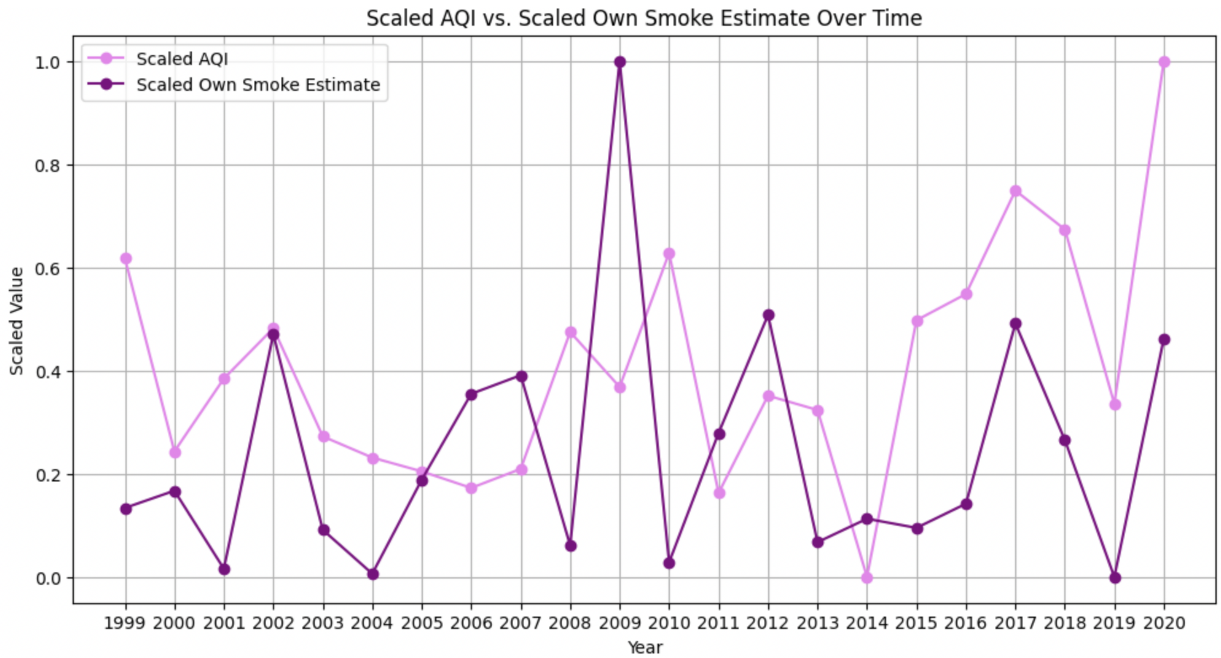

Figure Description: X-axis: Represents the years from the dataset and Y-axis: Represents the scaled values of both the AQI and the own smoke estimate. This line chart visually represents how the scaled AQI and the scaled own smoke estimate vary over the years. We will look more into the interpretation below.

Interpretation: The x-axis represents the passage of time, ranging from the earliest year to the most recent. On the y-axis, values have been transformed into a standardized unit, allowing for a meaningful comparison between the two datasets. The violet line, denoting the scaled AQI values, demonstrates fluctuations over time, indicating changes in air quality. Simultaneously, the purple line, representing the scaled own smoke estimate values, also exhibits its own pattern of variation. Interestingly, there appears to be a noticeable degree of correspondence between the two datasets, suggesting a potential correlation. This observation is further supported by a calculated correlation coefficient. The positive correlation coefficient indicates a tendency for the scaled AQI and smoke estimate values to move in the same direction over time. Overall, this graph provides valuable insights into the relationship between air quality and smoke estimates, potentially offering crucial information for understanding the impact of smoke on the environment.

Underlying Data and Processing:

Following are some of the main parts of the way I generated this visualization:

- **Correlation Coefficient Calculation:** The np.corrcoef() function from the NumPy library is used to calculate the correlation coefficient. This function takes two arrays (scaled_aqi_values and scaled_smoke_estimate_values) as input and returns a 2x2 matrix containing correlation coefficients. In this case, we're interested in the coefficient at position [0, 1], which represents the correlation between the first variable (scaled AQI values) and the second variable (scaled smoke estimate values). This coefficient quantifies the strength and direction of the linear relationship between the two variables.

- **Line Chart Visualization:** A figure is created with a specific size using plt.figure(figsize=(12, 6)). Two lines are plotted on the figure using plt.plot(): The first line represents scaled AQI values (scaled_aqi_values). It is labeled 'Scaled AQI', uses circular markers (marker='o'), a solid line (linestyle='-'), and is colored violet (color='violet'). The second line represents scaled own smoke estimate values (scaled_smoke_estimate_values). It is labeled 'Scaled Own Smoke Estimate', uses circular markers, a solid line, and is colored purple. The x-axis is labeled as 'Year' with plt.xlabel('Year'), and the y-axis is labeled as 'Scaled Value' with plt.ylabel('Scaled Value'). The title of the plot is set to 'Scaled AQI vs. Scaled Own Smoke Estimate Over Time' using plt.title(). A legend is included to distinguish between the two lines using plt.legend(). Gridlines are added to assist in reading the values accurately with plt.grid(True). Finally, the plot is displayed using plt.show().

Writing and Reflection:

This particular assignment allowed for the students to collaborate on the project which proved to be really helpful to us. Firstly, it allowed for a broader understanding of the problem and different ways to approach it. By sharing code snippets, statistical approaches, and visualization techniques, I was able to learn from my classmates' perspectives and insights. It increased efficiency and productivity of each student. Instead of struggling with a particular aspect of the assignment, I could discuss with my classmates on how they approached these problem areas and tackled the situation. This was good because different classmates may have different approaches or techniques to solve a problem. Seeing alternative methods helped me gain a more comprehensive understanding of the material. Collaboration also provided opportunities for validation and clarification. If I was unsure about a specific code snippet or statistical approach, I could discuss it with my classmates to ensure its correctness. After this project I can proudly say that collaboration fosters a sense of community and mutual learning. It encourages open communication, knowledge sharing, and a supportive environment for all students to succeed. I only have good things to say about collaboration because I believe that's how good projects are completed.

I have discussed multiple times throughout the project with different classmates about different steps in the project. I have mentioned the parts where I discussed with my classmates below:

1) First and foremost I have referred to and taken the initial code snippets from the code Dr McDonald provided. This helped me a lot in thinking in the right direction and then starting to work on my own code. The specific files were provided by Dr. McDonald in his private shared google drive.

2) I had to discuss and clarify if I was thinking in the right direction when writing the smoke estimation code. All the classmates had different ideas and approaches. But overall a few of us discussed a similar way to approach the problem. But the code is still mine.
3) First and foremost I have referred to and taken the initial code snippets from the code Dr McDonald provided.
4) The AQI estimation was tricky to understand. I was not able to understand the specific steps of that part of the project. 2 of my classmates explained to me briefly what exactly needs to be done and how to approach that. Once I confirmed if I understood everything correctly I was able to write my own code cohesively.
5) I also had to take help on the writing and reflection part because I have never been active on github and I could not understand the difference between what needs to be put in the readme file and the write and reflection document.

For the research question posed - "What are the estimated smoke impacts on your assigned city for the last 60 years? " I have concluded the following things from the visualizations that I have produced for my city - 'Cladwell':

Based on the analysis of three visualizations, it is evident that wildfires have been most prevalent at distances ranging from 150 to 650 miles from my assigned city, Caldwell. This geographic range has consistently experienced the highest intensity of wildfire activity. Additionally, the total acres burned per year have exhibited a consistent upward trend, reaching a peak in 2020. Despite occasional fluctuations, this upward trajectory suggests a concerning escalation in the frequency and extent of wildfires, with potential implications for air quality and public health. Notably, when we examine the Air Quality Index (AQI) trends depicted in the visualizations, a disconcerting pattern emerges. Over the last six decades, the AQI in the vicinity of Caldwell has demonstrated a consistent and concerning decline. This decline signifies a deterioration in air quality, with potentially detrimental effects on the health and well-being of the community. The sustained decrease in AQI underscores the urgency of addressing the underlying factors contributing to this decline, which may include a combination of local industrial emissions, regional air circulation patterns, and, notably, the influence of wildfires at varying distances from the city.

In conclusion, the analysis of wildfire data and its impact on air quality over the past 60 years unequivocally indicates a growing concern for the city of Caldwell. The increasing frequency and size of wildfires, coupled with the declining air quality, highlight the need for proactive measures to mitigate the potential health and environmental consequences. Addressing this issue necessitates a multi-faceted approach, including enhanced wildfire management strategies, improved air quality monitoring, and targeted interventions to safeguard the well-being of the community and the surrounding environment.