DATA 512: Part 1 - Common Analysis

This document provides an explanation of the visualizations created as part of the assignment. Each visualization is designed to display important data in a clear and meaningful way. The following sections will describe what each figure shows, how to read them, and the significance of the axes and underlying data. This is followed by a reflection statement to discuss the collaborative activities involved in this assignment. This reflection will highlight the key lessons I learned from working together and how collaboration influenced my understanding of the research question.

Visualizations

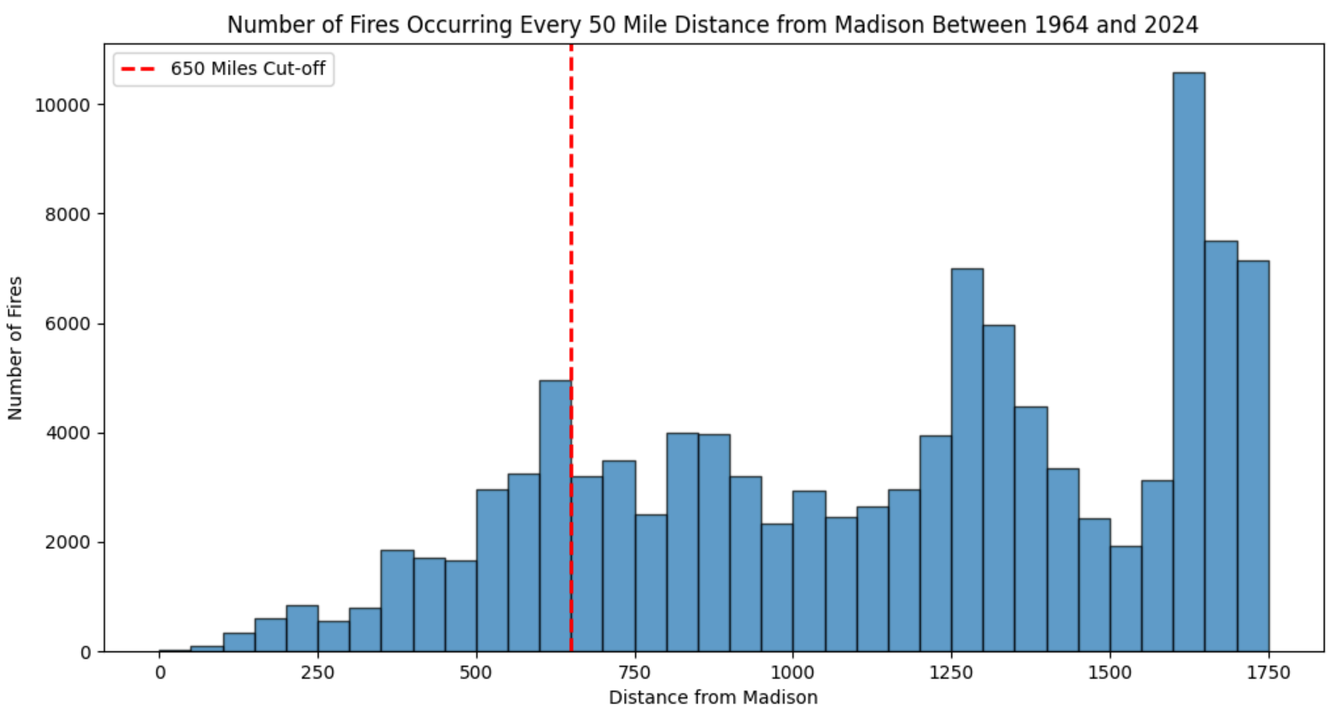

Figure 1: Histogram showing the number of fires occurring every 50 miles from Madison.

This histogram shows the number of fires that occurred at various distances from Madison between 1964 and 2024. The x-axis measures distance(shortest distance) from Madison in 50-mile increments, up to a maximum of 1800 miles, while the y-axis indicates how many fires were recorded within each distance interval. A red dashed line marks a 650-mile cut-off point, helping to highlight local fire activity and the scope of the project.

The count of fires was obtained from wildfire dataset - [Combined wildland fire datasets for the United States and certain territories, 1800s-Present (combined wildland fire polygons)] generated by US Geological Survey. Each fire record has the coordinates of the fire polygon perimeter. This was used to calculate the shortest distance from the fire area to Madison. The accuracy of fire perimeter data is not guaranteed to be 100%, and it becomes less precise, especially before the year 1980. This is a limitation to consider while using this graph. Although the dataset has different types of fires such as wildfire, prescribed fire , unknown etc. For analysis all fires have been included in the graph.

The histogram reveals that, within the cut off line, the fire activity is concentrated within 500 to 600 miles of Madison, suggesting that areas with dense forests, probably in the Northern Plains and parts of the upper Midwest, are more prone to wildfires. We see a linear trend within the cut off distance indicating more wildfires as we move farther away from Madison. But it is worthy to note that distant fires can affect air quality in Madison due to smoke being transported by winds, even if the fires themselves are far away. We also notice that a significant number of fires occurred beyond 1250 miles. This indicates that while nearby regions experience frequent fires, more distant areas, particularly in the western states, contribute more to overall fire count.

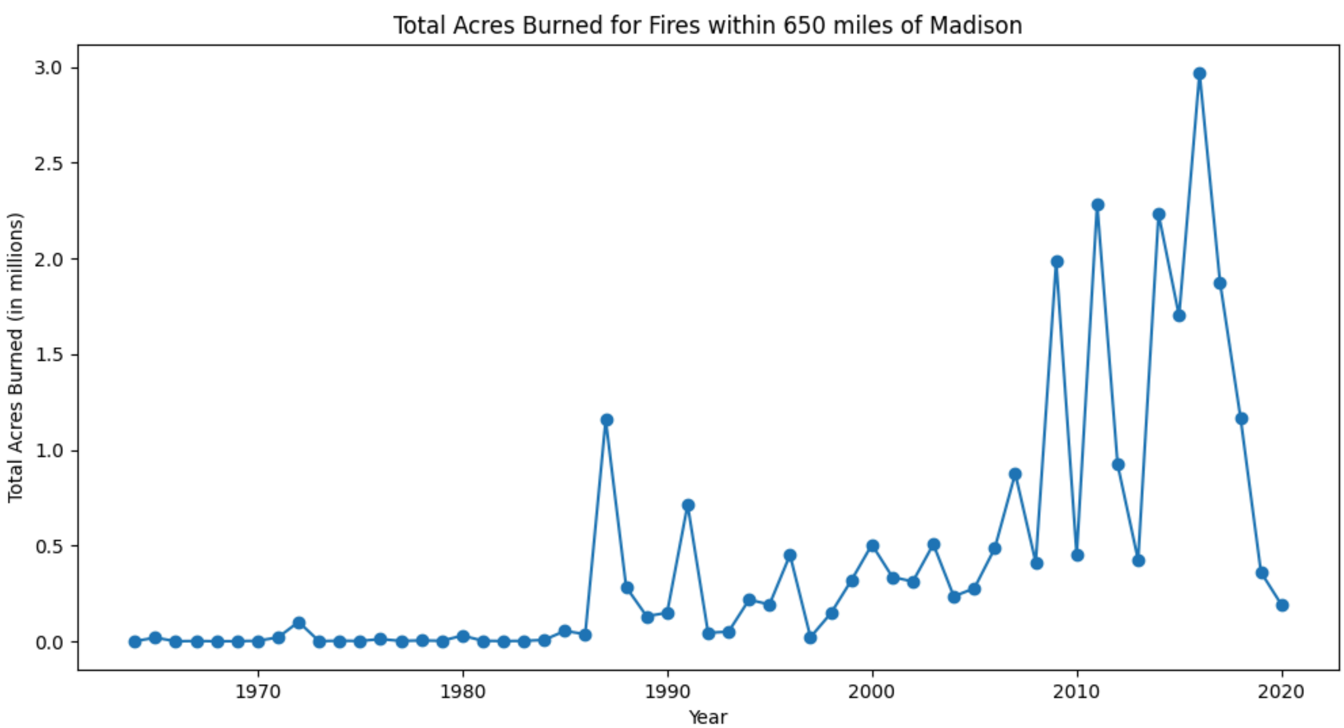

Figure 2: A time series graph of total acres burned from the fires near Madison

This line chart displays the total acres burned in fire incidents within a 650-mile radius of Madison from 1964 to 2024. The x-axis shows the years, while the y-axis indicates the total acres burned, measured in millions. This chart provides insight into how the scale of burned land has changed over time, highlighting important but concerning trends over the years.

The total acres burned data was obtained from 'GIS_Acres' of the same wildfire dataset. The burned area value of individual fire is summed up to obtain the total value for each year. For easy interpretation, values are scaled and displayed. One of the assumptions as per original dataset is that a particular area can only burn once per year, even with multiple fires. However for analysis, the 'GIS_Acres' is used irrespective of single or multiple fires for the given year.

From the chart, sharp increases in the total acres burned are observed starting in the late 1990s and early 2000s, peaking around the mid-2010s. This pattern aligns with a global rise in more intense wildfires, which have been linked to longer fire seasons and drier conditions caused by climate change. Additionally, year-to-year fluctuations in burned acres are evident, with some years showing significant spikes that may be attributed to extreme weather events, such as droughts or heatwaves. Recently, a slight decline in the total acres burned has been noted after the peak around 2011, 2016. This decrease may be due to improved fire management practices or natural variations in weather conditions. Factors such as extended fire seasons, changes in forest management, and natural climate variability contribute to the complexity of fire activity in the region.

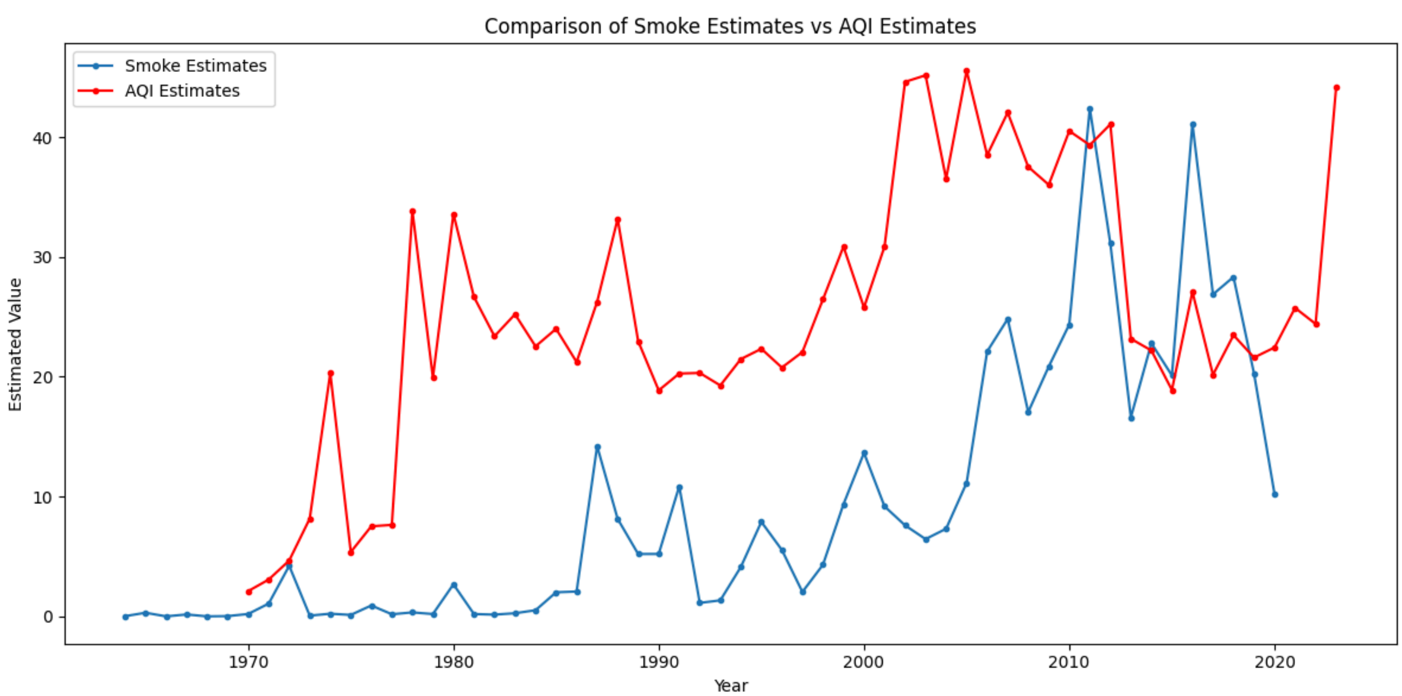

Figure 3: Comparison graph of Smoke estimates and AQI estimates

The above figure compares the estimated smoke levels and estimated average Air Quality Index (AQI) trends annually from 1964 to 2024. The plot has years on the x-axis, estimated values on the y-axis. Blue line indicates the trend in smoke estimates and red line indicates the trend in AQI estimates both marked with data points. This layout allows us to see year-by-year variations in the smoke and AQI estimates with noticeable peaks and declines pointing to significant changes over time. Additionally it also allows us to see any correlation or divergence between smoke and AQI levels

The AQI estimates were obtained from monitoring stations located in Dane county, WI which measure different types of gaseous and particulate pollutants and provide AQI values. Data was collected using the US Environmental Protection Agency (EPA) Air Quality System (AQS) API from To ensure accuracy, missing AQI data was recalculated using the available pollutant concentration data. The annual AQI estimates were calculated by averaging the daily AQI values collected during the fire season, which runs from May 1st to October 31st. This was to ensure a fair comparison with our smoke estimates, as the impact of smoke is mainly calculated for this period. To estimate smoke from wildfires, several key factors influencing smoke production were considered from the wildfire dataset. The area burned, derived from GIS_Acres, was used, as it directly relates to fuel consumption. Distance was factored in using the shortest_dist to the fire, since smoke concentration decreases with distance from the source. Fire intensity was estimated based on the Assigned_Fire_Type, while duration was inferred from the Shape_area of the fire, with larger fires expected to last longer. Wind effects were assessed through the ratio of Shape_Length to Shape_Area, indicating how easily smoke might disperse, and the circularity of the fire's shape was measured using Circleness_Scale, affecting smoke containment. By combining these elements, a comprehensive smoke estimate was formulated.

Over the years, both Smoke estimates and AQI estimates have generally increased, especially since the late 1990s, likely due to a rise in global wildfire activity linked to climate change factors like longer droughts and higher temperatures. After 2000, noticeable fluctuations in smoke and AQI values were observed, with few peaks in Smoke corresponding to major wildfires in North America during the 2000s and 2010s. The major peaks in smoke estimate in 2011 and 2016 can be attributed to the large number of wildfires that occurred in that year within 650 miles and highest area burned as seen in above two graphs. There are instances when AQI levels increased without a rise in smoke and vice versa, indicating that while wildfire smoke significantly impacts air quality, other factors such as industrial emissions, urban pollution, and agricultural activities also play a role.

Reflection

In analyzing fire incidents and their impact around Madison, Wisconsin, I have learned a lot about the complex relationship between wildfires, climate change, air quality etc but most importantly the chemistry behind the Smoke production and math behind the AQI calculations that we see in weather apps almost everyday. The visualizations created in this study were as part of the analysis but some of these trends are to be feared. They remind us of the real life consequences of the numbers being analyzed. The quality of the air we are breathing and the suffering of future generations if the numbers continue the way we are seeing in the graphs.. Climate change is a pressing reality that cannot be ignored.

The collaborative activity during this assignment has greatly improved my approach to data analysis. I had a couple of brainstorming sessions with some of my cohort members when faced with data challenges where we shared our ideas and refined our thoughts. The session outcomes helped me to a large extent in overcoming a few of the limitations of the data. We discussed how to calculate different metrics that could potentially be used for final data analysis and checking with each other on interpreting the data correctly. Below are some of the techniques I used from sessions.

Technique 1:
Attribution - Navya Eedula <needula@uw.edu>

The circleness_scale of the fire can improve the smoke estimates. A more circular shape tends to contain the smoke more effectively, leading to higher concentrations of smoke in the immediate area, as it limits the surface area exposed to wind. Conversely, a less circular shape allows for greater smoke dispersion leading to lower concentrations in the area. Navya did an extensive research on this attribute and it was her idea to include this factor in the numerator to improve smoke estimate.

Technique 2:
Attribution - Navya Eedula <needula@uw.edu>

It was very challenging to understand and translate the AQI calculation from the technical assistance document. My discussion with Navya enhanced my understanding of the topic and she probed me into getting more details on some of the aspects that I had ignored earlier. The final formula used to generate AQI using pollutant concentration for records with missing AQI was a combined effort from both of us which was later converted to code individually.

This Collaboration with her saved me a lot of time because I was able to learn techniques that were based on thorough research conducted by them. Instead of spending hours figuring things out on my own, I benefited from their insights and expertise, which allowed me to focus on applying what I learned rather than starting from scratch

In addition to the techniques mentioned earlier, I had general discussions about the GeoJSON libraries that I was using for the first time. These libraries helped me work with geographic data, which was a new experience for me. I also explored ways to make the distance calculation process faster. This was important because reducing the time taken for calculations can make the overall analysis more efficient. Additionally, I shared some of my ideas and approaches with them, and their feedback was valuable for validating my data and results. Their perspectives allowed me to see my work from different angles, which improved my understanding and enhanced my analysis. Overall, the experience of creating new data analyses and solving problems together was very educational. I learned practical skills that will be useful in future projects, and I also developed a deeper appreciation for teamwork. Teamwork is crucial in research because it allows for sharing knowledge and ideas, leading to better outcomes.