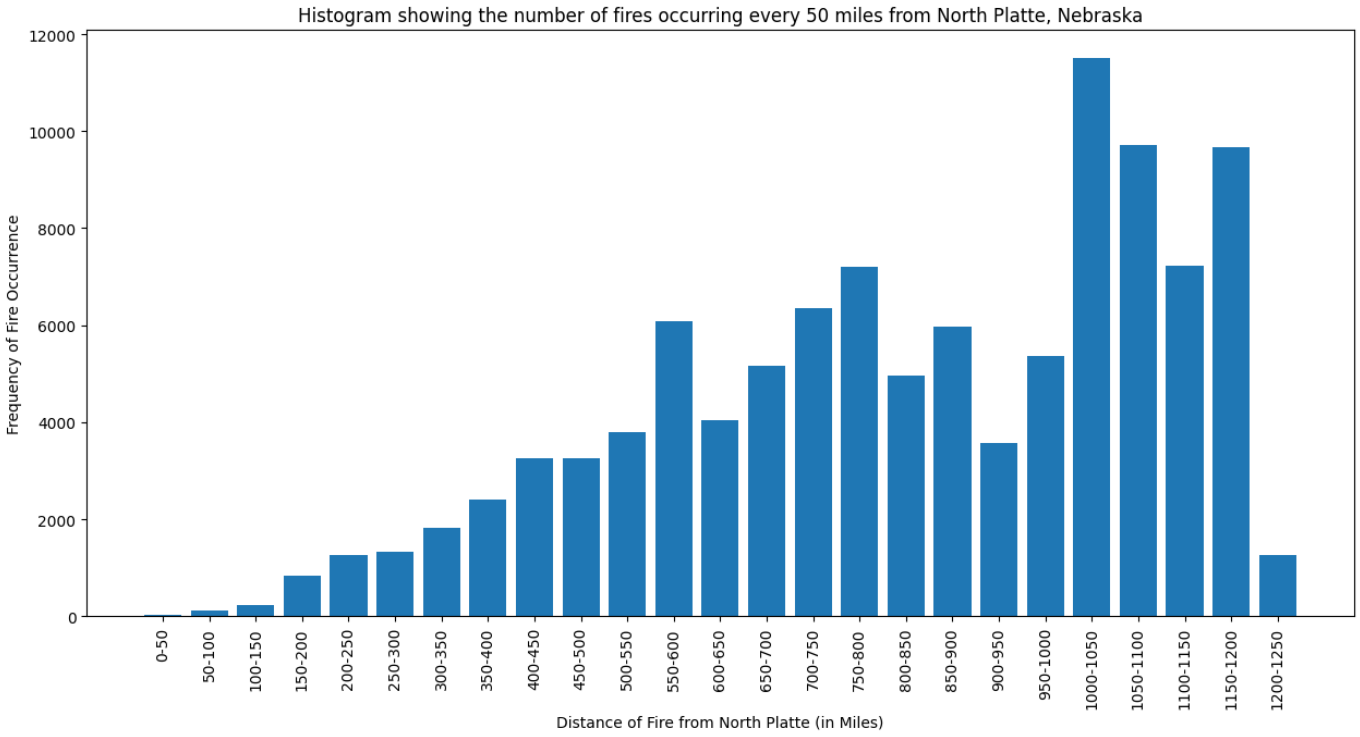
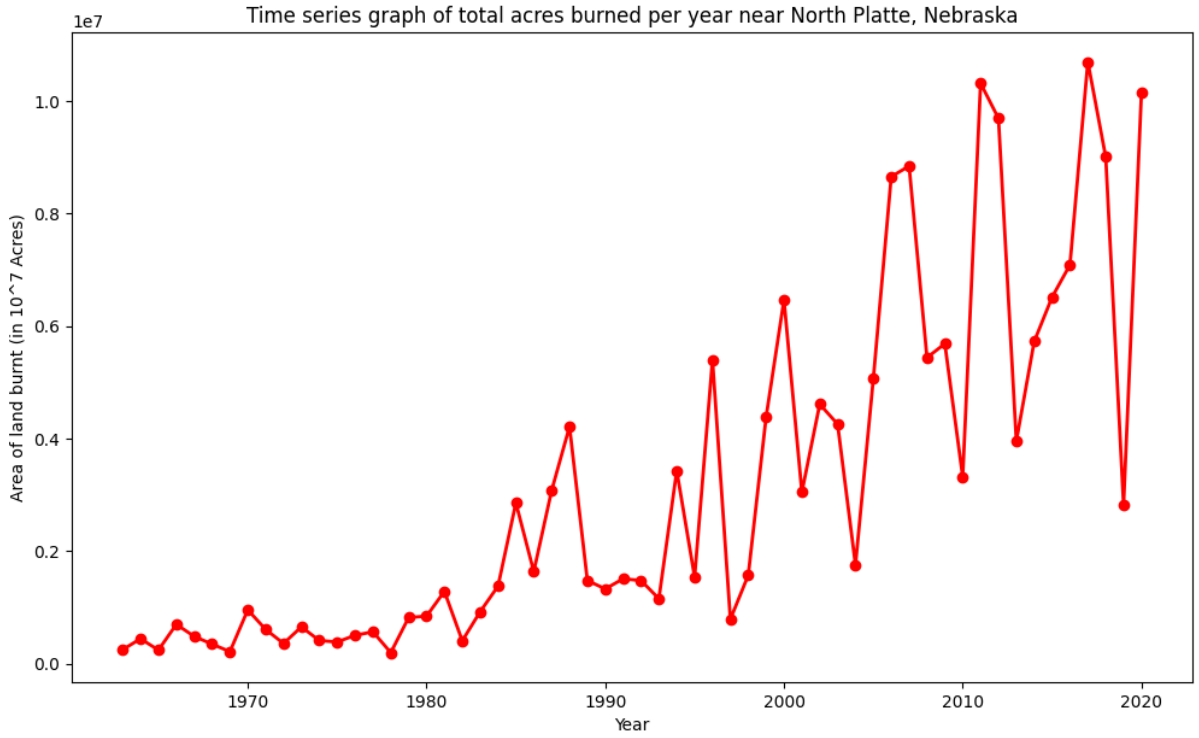
VISUALIZATION EXPLANATION

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



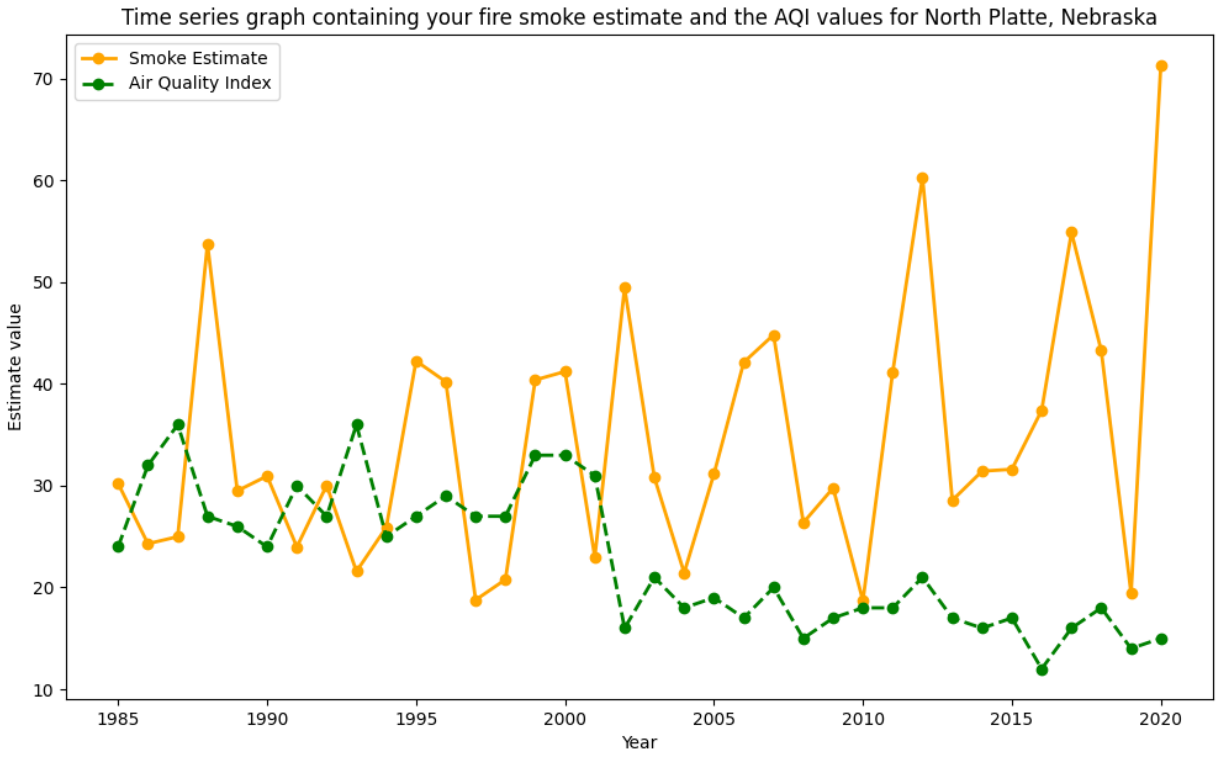
This histogram represents the frequency of occurrence of fires, arranged distance wise in bins of 50 miles, from North Platte, Nebraska. The data came from the filtered GeoJSON from USGS which includes the fire date, distance from city, smoke estimate, area burnt, fire type and other such parameters. The features used for this visualization is a count of the number of fire occurrences within each 50 mile distance bucket. The X axis represents the distance from North Platte in miles. The y axis represents the count of the number of wildfires. We can see that it is fairly well distributed, with more fires being present farther away from the city. The maximum number of fires are at a distance of 1000-1500 miles from the city. I can infer that North Platte, Nebraska is not a big wildfire center because there are not a lot of fires closer to the city.

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



This time series graph represents the area burnt by fires close to North Platte. The data came from the filtered GeoJSON from USGS which includes the fire date, distance from city, smoke estimate, area burnt, fire type and other such parameters. The features used for this visualization is a summation of the total area burnt by the fire for each year. The X axis represents the years of the fires (1963-2020). The y axis represents the total area of land burnt in 10^7 acres for better visualization. We can clearly see that the amount of land area being burnt is increasing with time. The 2000s had a way higher burn impact than the 1900s. I can infer that the impact of fires around North Platte, Nebraska has grown significantly. This kind of correlates with the increased number of fires that we saw in the first graph, and probably fire intensity has also increased substantially, which does not bode well for North Platte city in general.

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



This time series graph represents the values of smoke estimates generated by me and the air quality indices, across years. The data came from the filtered GeoJSON from USGS which includes the fire date, distance from city, smoke estimate, area burnt, fire type and other such parameters. It also included the AQI data from the API, which has been aggregated by year to give one single value. The features used for this visualization are smoke estimate, aqi, fire year. The X axis represents the years of the fires (1985-2020). The y axis represents the estimated values. The orange solid line shows my smoke estimate and the green dotted line shows the actual AQI values. We can see some sort of correlation between the two lines charted. In the initial years, the smoke estimate and aqi overlap significantly. Later on the AQI drops a bit (which is great! While the smoke estimate actually goes up. The smoke estimate going up can be attributed to the higher number of fires and higher area burnt, as we say in the first two graphs. The discrepancy might arise from the fact that we are considering fires from upto 1250 miles away, but we are considering air quality only upto 200 miles away. Maybe the number of fires did increase, but the city still succeeded in maintaining a good air quality!

COLLABORATION AND RESEARCH REFLECTION

This assignment, despite being the first part of the project, in itself represented an end to end data science project lifecycle. It involved data collection, feature engineering, data wrangling, creating visualizations and generating insights. The USGS survey JSON is a massive file with 2.8GBs of data. This was really tough to load and work with. However, the reader module provided by Dr. McDonald made it easier to quickly read in the data, because the python geojson library was failing and crashing. Next up, working with GeoJSON data, calculating geodesic distances and filtering data for specific locations was also made easier by Dr. McDonald's example notebook. I loved the part where I had to come up with a logic to calculate smoke estimate that had some semantic sense to it, along with good exception handling for the big dataset. Then, I moved onto the AQI extraction. Once again, the example notebook benefitted me enormously. I am glad that this AQI data is available open source so geospatial analytics enthusiasts can easily use it! Here again, I had to write a lot of logic to handle missing AQI data. For my city, North Platte, there were not a lot of weather stations close by, hence I had to expand my search radius to 200 miles. Even then, different weather stations had different years of operation, hence 3 different stations had to be considered. Finally I ran the API call per year and aggregated by mean, a lot of coding work! The final part involved visualizations, which were fairly straightforward, but gave me great perspectives into the impact of the fires in terms of area burnt, and also revealed some trends in fire occurrences. I noticed that fires are occurring more in recent years and burning down more acres of land. This is an alarming trend and if not arrested, does not bode well for the sustainability of forest areas.

The collaboration aspect of this assignment was also a fun process and enabled me to gain different perspectives of my classmates, which often differed from my own! Right from whether the right data file had been downloaded, to the time taken to load the data, to the formulation of smoke estimate calculations, there was support through the process. It felt good to know that others are facing similar issues, or relief that my results are similar to others! The code I wrote was my own, however I did refer to some Medium articles, stackoverflow posts, API documentations or even our very own DATA 512 Slack discussions to get a better understanding of how to proceed with my analysis. The collaboration aspect helped me streamline my workflow, but I agree that code or solution sharing is simply wrong to those who work hard. I have worked on this assignment for the better part of a week, but if someone starts on the last day and uses my code under the name of 'collaboration', that would be unfair. Hence, collaboration must be well defined and carried out upholding UW principles and moral ethics.

Overall, I enjoyed this assignment and putting in efforts for this, and look forward to what future parts of the project are going to entail!