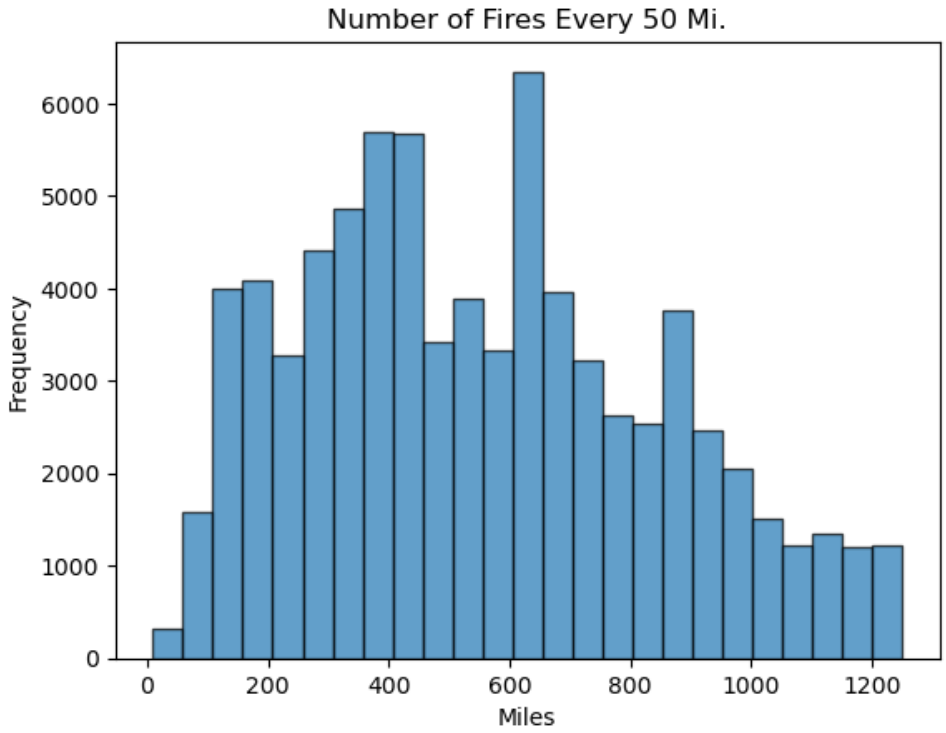
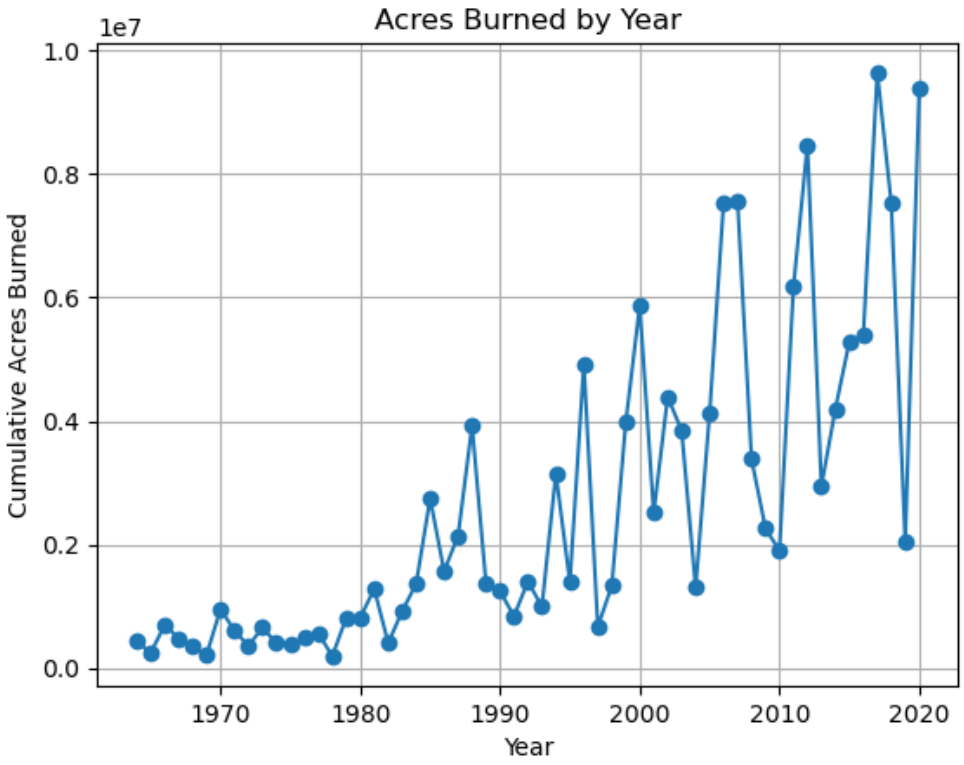
Reflection of Common Analysis

Part 1 – Visualizations



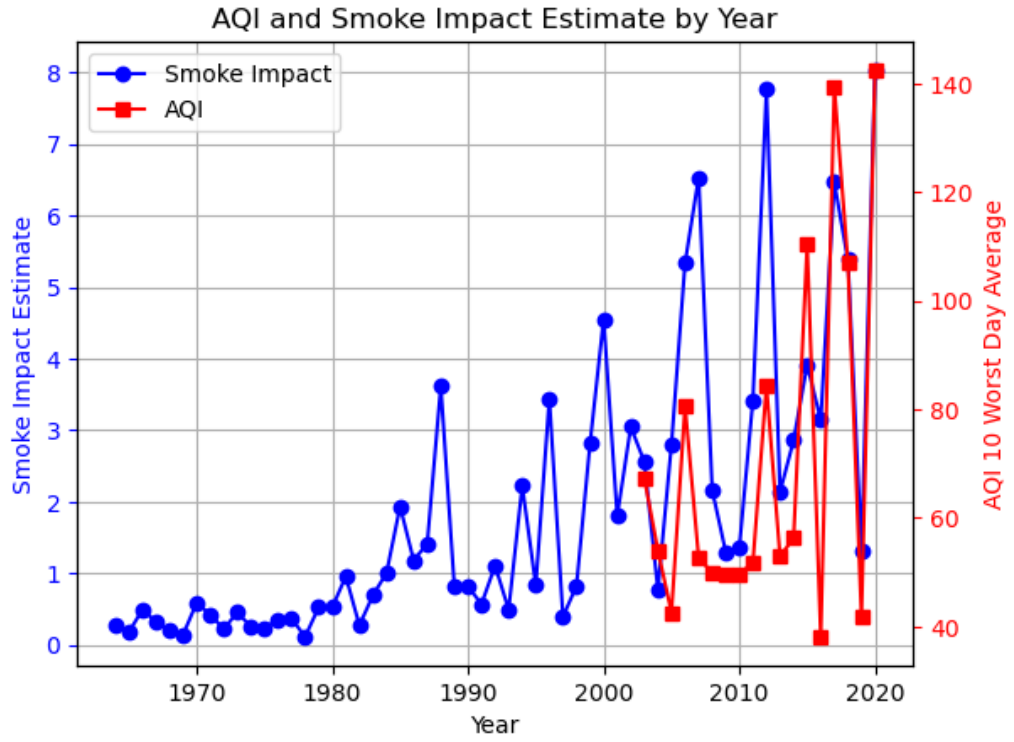
Prompt: Produce a histogram showing the number of fires occurring every 50 mile distance from your assigned city up to the max specified distance.

This figure demonstrates how many fires have fallen within the 50 mile ring surrounding Pullman, WA since 1963. Each bin of the histogram represents 50 miles. The x-axis is the miles from Pullman, WA and the y-axis is the number of times fires have occurred that number of miles away from Pullman, WA. The underlying data was processed from the combined wildfire dataset found here. The geo markers for each fire in the dataset were processed to find the closest point to Pullman, and if that fell within the 1250 mile radius, it was added to the subset of the data and the distance of the fire was saved as well. The histogram was then created using these fires and their calculated distances.



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

This figure demonstrates how acres were burned each year by fires within a 1250 mile radius of Pullman, WA since 1963. This time series graph has the year on the x-axis and the cumulative acres burned on the y-axis. The underlying data was processed from the combined wildfire dataset found here. The geo markers for each fire in the dataset were processed to find the closest point to Pullman, and if that fell within the 1250 mile radius, it was added to the subset of the data. The target fires had associated years. The fires were grouped by their year and the acres they impacted were added together to calculate the cumulative sum of damage.



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

This figure demonstrates a few things. First, in blue, it has my smoke impact estimate of fires within a 1250 mile radius of Pullman, WA since 1963. Second, in red, it has the AQI estimates for Pullman since collection began in 2003. The x-axis is the year of interest. The y-axis on the left and in blue represents the smoke impact factor for each fire is calculated by the formula $e^{-d+a} * a$ where d is the normalized distance and a is the acreage burned by the fire. The smoke impact factor for the year is the sum of all the individual fire's smoke impact scores. The y-axis on the right and in red represents the average of the 10 worst air quality index scores for each year. I took the average of the 10 worst days to better capture the accumulation of bad fires throughout the year. If the smoke impact score follows the general trends of the AQI average score, it could indicate that the impact score is accurately capturing the amount of smoke in Pullman for a given year.

Reflection

I learned a couple things while completing the research questions. First, I learned how AQI is calculated, the different measurements that go into it, and what measurement captures wildfires the best. I think it is interesting that two places with the same AQI score could have drastically different air qualities. For instance, if one location has an Ozone AQI of 300 and a PM10 AQI of 301 while another location has an Ozone AQI 0 and a PM10 AQI of 301, they will both be reported as having AQI of 301. Additionally, particulate matter 10 is most representative of wildfire pollution because the particles are bigger than those produced in engine combustion (PM2.5). Second, I learned a little more about SARIMA models for predicting time series data with seasonality. This is great for our needs in this assignment because every 5 years or so the fires appear to go through a cycle of more acres burned and less acres burned. Thus, I learned some things about wildfires and air quality as well as data science methodology.

Regarding collaboration, it had pros and cons. To begin, the idea to use the SARIMA model came from Zach Price and appeared to have more reasonable results than a linear regression or something of that ilk. Thus, that was helpful. On the other hand, hearing what other people said in passing about processing the data confused me more than helped me. Maybe this is because we had different requirements given our target cities. Outside of this, I did not get any code or ideas from others. I prefer to develop on my own first and then iterate in the future if better methodology presents itself. Thus, collaboration helped me understand the better way to predict smoke impact.

The assignment was inherently a little confusing. Getting feedback from others that I was not alone was helpful for morale. I am still a little skeptical of the value of predicting smoke impact from previous smoke impact values which already have dubious interpretability. I was especially unsure of how to propagate this error if I used a linear regression model. Thus, Zach's suggestion to use SARIMA which had a built in CI calculation was very useful.