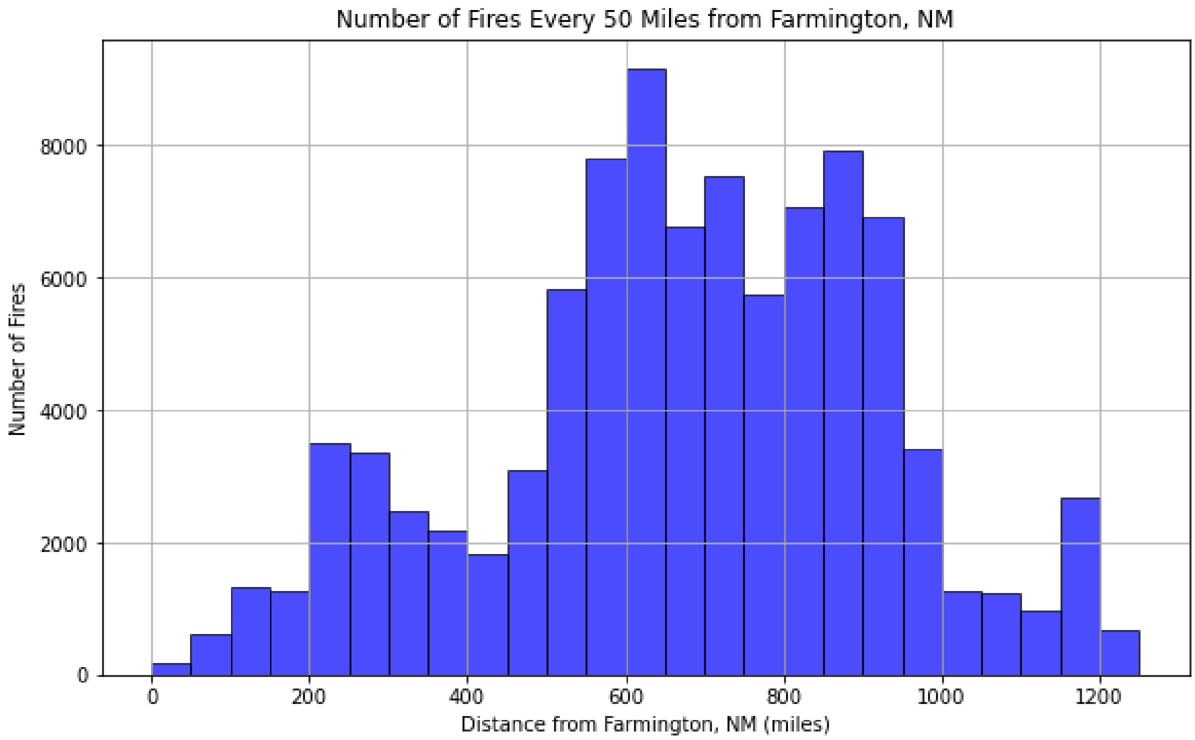
Step 3: Write and reflect

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 histogram above visualizes the data for the number of fires within incremental distances of 50 from Farmington, New Mexico. This data is for the period from 1963 to 2020. Ideally, our analysis would extend to 2023, but the dataset available to us concluded with the year 2020.

Each bar represents a 50-mile distance interval and the height of each bar corresponds to the number of fires recorded in that interval. The x-axis is labeled "Distance from Farmington, NM (miles)" and is divided into equal intervals of 50 miles. The scale starts from 0 and extends beyond 1200 miles. The y-axis represents the "Number of Fires," with tick marks at intervals that appear to be every 2000 fires, starting from 0 and going up beyond 8000. Observing the bars, we can comment on several notable aspects:

- The first few intervals from 0 to around 200 miles have lower frequencies, with the number of fires increasing modestly within this range.
- There is a significant rise in the number of fires that seems to begin after the 400-mile mark, with a noticeable peak forming in the 500 to 700-mile range.
- The peak of this histogram occurs between approximately 550 miles and 650 miles from Farmington, NM. The tallest bar within this peak suggests this is where the most fires were recorded, reaching a height above the 8000 mark on the y-axis.
- After this peak, there is a marked decrease in the number of fires as the distance increases from 700 miles and beyond, with occasional smaller peaks but none reaching the height observed in the central region of the graph.

- The number of fires dips significantly as we move towards the 1000-mile mark and continues to remain relatively lower beyond this distance.

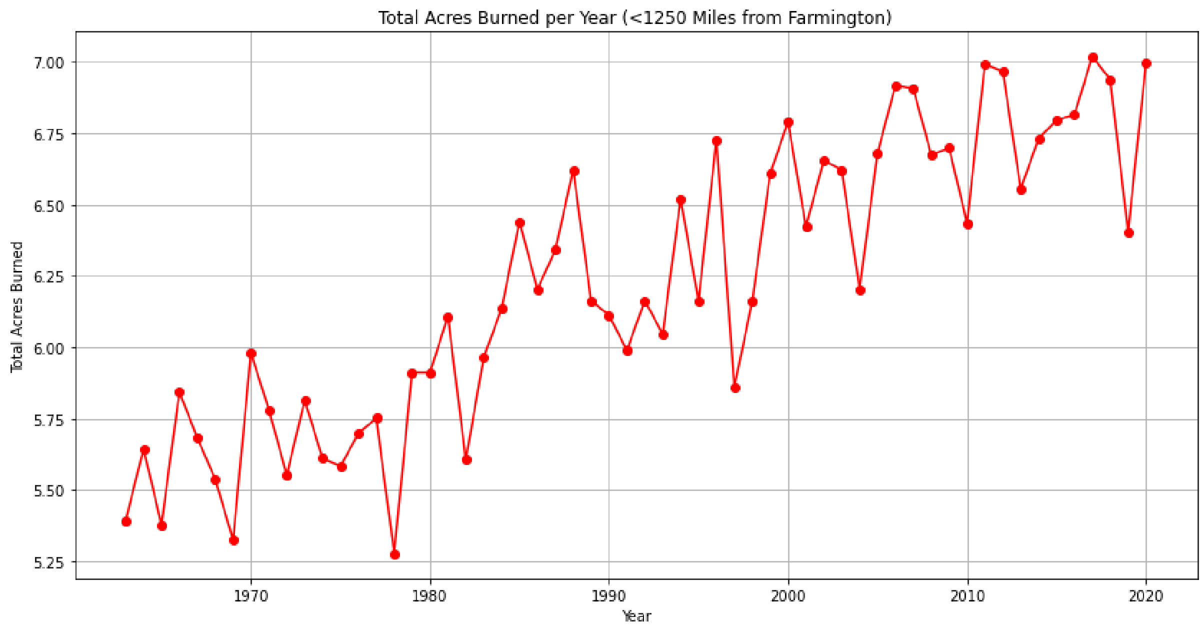
From this histogram, one might infer that there is a particular zone, roughly between 500 and 700 miles from Farmington, NM, where fire incidents were particularly high during the above-mentioned time.

Talking about the data, it originates from the USGS and was generated by merging 40 different wildland fire datasets that have been published. To narrow down the raw data, we filtered it to only include fires that occurred within a 1250-mile radius of Farmington (as indicated by the upper limit on the x-axis of the histogram) and those that occurred after the year 1963, inclusive.

There are several factors within the USGS data that are likely to have an impact on our analysis. Firstly, the accuracy of fire perimeter data is not guaranteed to be 100%, and it becomes less precise, especially before the year 1980. Additionally, the creators of the data assume that a particular area can only burn once per year, which is usually true. Therefore, if an area is recorded as having burned twice in one year, it is counted as a single burn. As per the documentation, all fires that were identified as wildfires or prescribed fires were categorized accordingly. We included all types of fires in our tally of fires and smoke estimates.

Our present approach to measuring the distance between fires and Farmington assumes a "ring" shape for fires. Although the majority of fires conformed to this assumption, 36 fires displayed a "curved ring" shape and were thus not included in the dataset. This omission represents less than 10% of the data that was eliminated due to considerations of geometric shape.

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



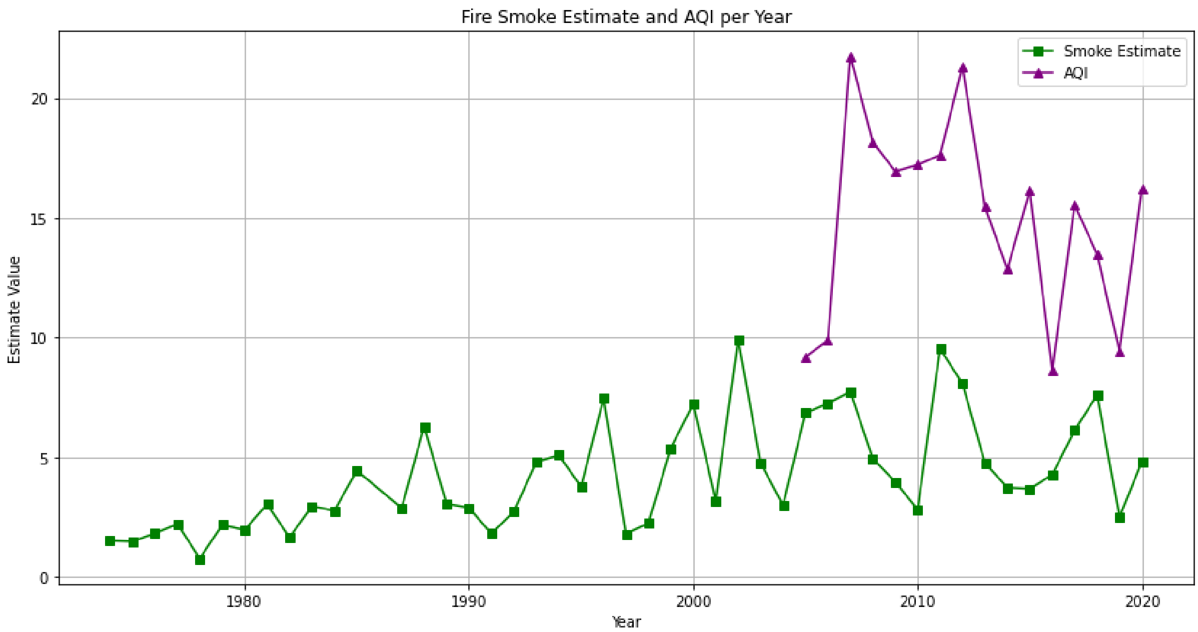
The second visual is a line graph depicting the Total Acres Burned per Year for fires within 1250 Miles of Farmington over a time frame that spans from 1963 to 2020. The x-axis represents time, specifically years, ranging from a little before 1970 to a bit past 2020. The marks are every 10 years, with finer lines indicating the individual years. The vertical axis quantifies the total acres burned, measured in millions. The scale is logarithmic, to quantify for high values (earlier in 10 million acres) with markers at every quarter. The Red dots represent the total acres burned for each year, and these points are connected by a red line. This helps in visualizing trends and fluctuations over the years. The line graph shows quite a bit of variability, with some years experiencing significantly more acres burned than others. For instance:

- From the beginning of the graph till the end, there's a fluctuating but general trend of increase in acres burned.
- A sharp increase in acres burned can be observed at several points, notably around the early 1980s, the mid-1990s, and during the 2000s.
- The line graph shows that the early 2010s had a peak, followed by a sharp decrease and then another increase towards the end of the graph.

I feel that this graph could be useful in studying the effects of climate change, land management practices, or other factors that influence wildfire intensity and frequency. It shows that there is no simple, steady trend, but rather a complex pattern that may require more in-depth analysis to fully understand.

The data we use to create this visual is the same data that we used for visualization 1. Hence, all the points mentioned above about the data should hold good for this visualization as well.

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



The third visual is a dual-line graph that depicts the Fire Smoke Estimate and AQI per Year showing the trends of smoke estimate and air quality index (AQI) over several decades, from 1963 to 2020. However, it is to be noted that the data for AQI starts only in 2005 for Farmington, NM.

This x-axis represents the timeline, with years marked at each decade and finer lines indicating individual years. The vertical axis measures the smoke estimate value, in acres/miles units. It ranges from 0 to above 20, with markers likely at every 5 units. Smoke Estimate (Green line with squares) represents the estimated amount of smoke produced by fires each year. The line fluctuates, with notable peaks and troughs throughout the observed period. The AQI signifies the Air Quality Index each year, marked by triangles linked by a purple line. The AQI data shows significant variability, with sharp increases and decreases, suggesting years with poorer air quality interspersed with years with relatively better air quality. Both lines show variability over the years, but there doesn't seem to be a consistent correlation between the two datasets at a glance. This graph was created to be utilized to study the relationship between the amount of smoke produced by fires and the overall air quality in a given area. However, the apparent lack of a consistent direct correlation suggests that air quality is influenced by a variety of factors, not just fire smoke and that these factors can vary from year to year. It is important to remember that the Smoke estimate and the AQI are not the same.

In our approach to calculating the annual smoke impact for Farmington, NM, we take into account two critical factors for each fire event: the size of the fire (measured in GISAcres) and its proximity to the city (measured by shortest_dist). Fire size is directly proportional to the amount of smoke generated, while distance from the city inversely affects the smoke's impact. To account for the dispersion and dilution of smoke with distance, we incorporate an inverse distance weighting (IDW) function as a representation of the impact decay with distance. Rather than summing impacts (which could overstate or understate effects), we calculate the average smoke impact over the year. The formula for estimating smoke impact for each fire involves the fire's size and distance, with a decay exponent (α) determining the rate of impact reduction with distance; we initially set $\alpha=1$ for simplicity. Finally, the annual smoke impact for Farmington is derived as the average of the smoke estimates per fire, providing a comprehensive view of the city's yearly air quality impact.

We derived our AQI estimate from sensors located in San Juan County. These sensors measure multiple types of particulate matter, so we selected the highest AQI value recorded at that station across various gases or particles. Furthermore, for our annual AQI estimates, we computed the average of daily AQI values recorded during the fire season, which spans from May 1st to October 31st. We intentionally excluded AQI estimates from months outside the fire season to ensure a more equitable comparison with our smoke estimate, as smoke impact primarily occurs during the fire season.

Reflection Statement:

In undertaking the task of this analysis of fire incidence and impact around Farmington, NM, I have gained valuable insights into the intricate relationship between wildfires, environmental management, and climate in general. The visualizations created serve not only as analytical tools but also as stark reminders of the tangible realities these numbers represent – the scarring of the earth and the air we breathe. Climate change is real after all?!

The collaborative aspect of this project significantly improved my data analysis approach. Working together with my colleagues provided a diverse perspective on the problem we were tackling. Our discussions mainly revolved around methods for calculating various metrics and approaches to data reading. These exchanges of ideas greatly clarified my understanding of these metrics, as we shared and refined our thoughts. We all had unique insights, serving as a sort of reality check to ensure we were on the right track. It is important to note that these discussions were invaluable in terms of encouraging us to critically examine all aspects, ensuring the robustness of our conclusions. Even though we collaborated, we maintained a strict policy of not directly copying code or ideas from one another.

Throughout this process, we relied heavily on the Python programming language, utilizing libraries that I've never used before such as geojson. Furthermore, the code and information provided by the Professor provided me with a strong basis to start my work.

In conclusion, the combined experiences of crafting new data analysis and engaging in collaborative problem-solving have been profoundly educational. They have not only equipped me with practical skills but also fostered a deeper appreciation for the collective pursuit of knowledge and solutions in environmental research.