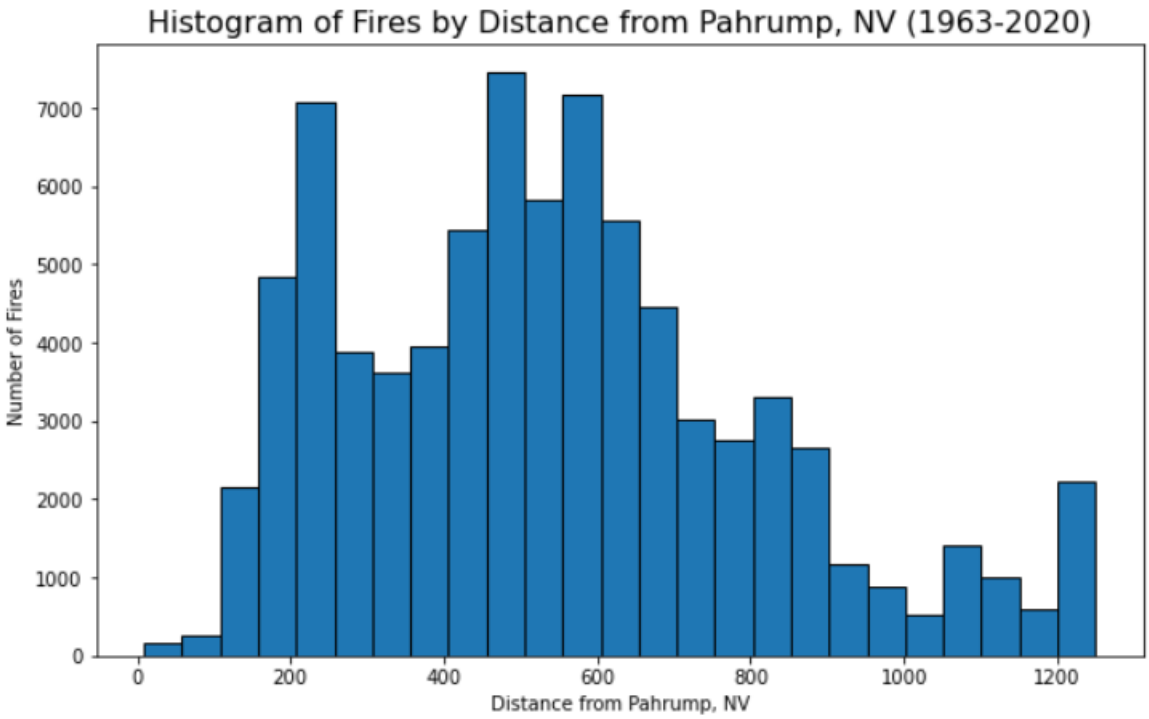
Step 3: Writing and Reflection



Visualization 1 depicts a histogram of the number of fires occurring every 50 miles from Pahrump, NV within the years 1963-2020. While we would have preferred to track fires until 2023, our source data terminated in the year 2020. The x-axis represents the distance of the fires from Pahrump, increasing from left to right. Specifically, distance is measured from the center of Pahrump (estimated visually on Google Maps) to the closest edge of the fire. The distance considers the curved surface of the earth and is the length radially out from the center of the city "as the crow flies". The y-axis represents the number of fires. To read this visualization, consider each "bar" the number of fires contained within a 50 mile "bucket" from Pahrump. As an example, the third bar from the left of the chart shows there were ~2000 files between 100-150 miles from Pahrump in the years 1963 - 2020.

As we can see from the figure above, fires are not normally distributed over distances from Pahrump. It would appear that our distribution has 2 major peaks occurring at 450-600 miles and 200-250 miles. There may also be a smaller peak in the number of fires at approximately 1050-1100 miles from Pahrump. Geographically, we contain much of the US Southwest in our peak fire radii, with most of California, Nevada, Utah, and Arizona contained by the 500 mile radius. When I think of states commonly in the news for wildfires, this radius of fires aligns with those states, especially California.

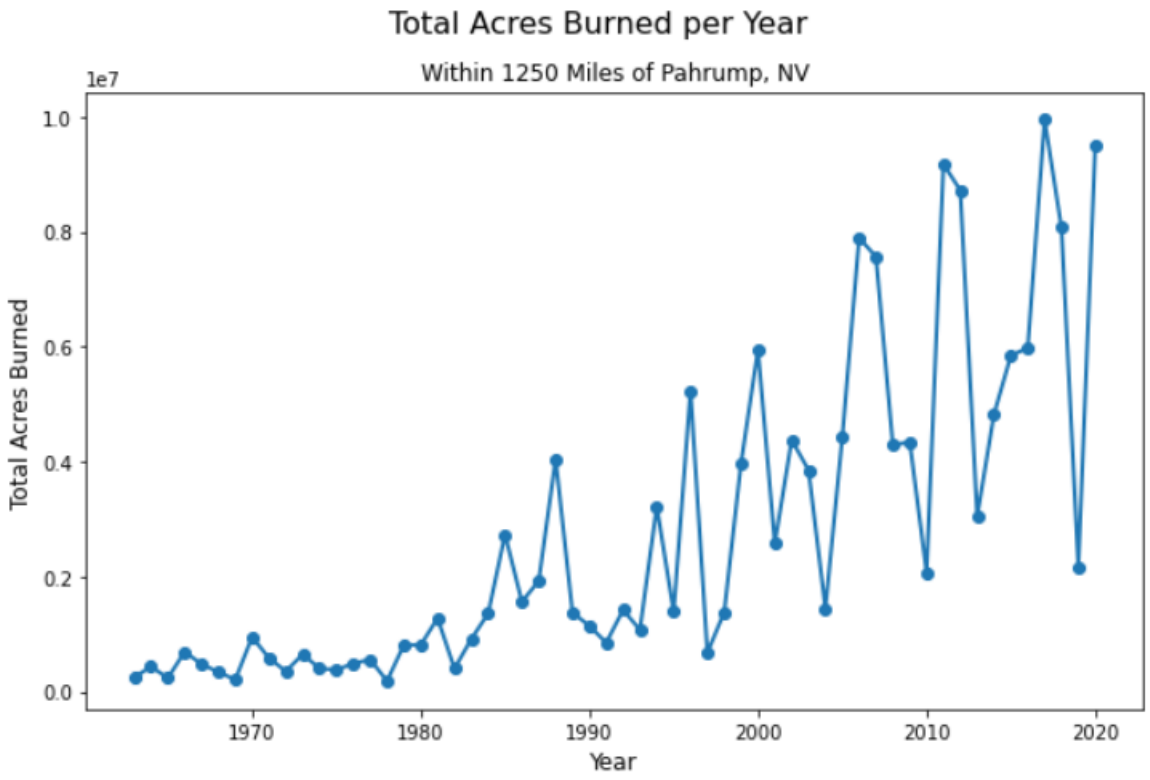
The underlying data comes from the USGS and was created by combining 40 different published wildland fire datasets. We used the "combined" dataset which avoids duplicate fires as much as possible. We reduced the raw data file to contain only fires within 1250 miles of Pahrump (reflected in the upper bound on the x-axis of the histogram) and those which happened after 1963, inclusive. There are a number of considerations in the USGS data which likely impact our analysis. First, fire perimeter data is not 100% accurate, and becomes less accurate prior to 1980. Data before 1980 underestimates the actual number of fires, but it is impossible to estimate by how much. In general, the boundaries used to calculate distance should be assumed to be approximate area burned.

Further, the data creators assume land can only burn once per year (usually true), thus if an area is listed as burned twice in one year it is counted as one burn. Fires which burned in the same year within 500 meters of each other were also counted as the same fire, potentially reducing total fire count. Wildfires without years were removed from the dataset by the creators as much as possible.

Per documentation, "All fires that were clearly labeled as wildfires or prescribed fires were labeled as such. Remaining fires in wildfire datasets were labeled as "likely wildfire". Fires from the Monitoring Trends in Burn Severity Dataset marked as "Unknown" were labeled as "Unknown - Likely Wildfire" if a wildfire report existed for these fires. Otherwise, they were labeled as "Unknown - Likely Prescribed Fire". We consider all kinds of fires in our count of fires and smoke estimates.

Our current method to calculate the distance between fires and Pahrump assumes a fire “ring” geometry. While most fires conformed to this expectation, 35 had a “curve ring” geometry and were removed from the dataset. This represents <10% of data omitted on the basis of geometric shape.

For more information on the underlying data, intermediate data files, and processing steps please see the README located in our wildfire directory on GitHub.



Our next visualization shows the total acres burned per year for fires occurring with 1250 miles from Pahrump and burning between 1963 and 2020. The x-axis of the figure illustrates the year of the fire(s) increasing from right to left, and the y-axis shows the total number of acres burned for all fires in that year. It is important to note that the units of the y-axis are 10,000,000 acres (10 million acres). To read the figure, identify the year of interest and visually navigate to the dot above that year. By finding the dot’s corresponding position on the y-axis, the reader can understand the total number of acres

burned by fires in their selected year. As an example, in 1980 ~1,000,000 acres were burned (or 0.1 as it's noted on the y-axis). Lines connecting yearly measures were included to draw the reader's eye to the variability in the chart.

We can see that the total acres burned per year is variable, but overall increasing at a linear or exponential rate. This aligns with our perception of fire coverage in the news over the past few years. However, it is worth noting that fire estimates prior to 1980 under-count the number of acres burned, so it's possible the trend may not be as exponential as it appears here.

As mentioned in visualization 1, the underlying data comes from the USGS and was created by combining 40 different published wildland fire datasets. We used the "combined" dataset which avoids duplicate fires as much as possible. We reduced the raw data file to contain only fires within 1250 miles of Pahrump and those which happened after 1963, inclusive. We know that the fire perimeter data is not 100% accurate and becomes less accurate prior to 1980. Data before 1980 underestimates the actual number of fires, and thus the acreage burned, but it is impossible to estimate by how much. In general, the boundaries used to calculate distances to fires should be assumed to contain the approximate area burned. It is possible that our metric of total acres burned (calculated by summing all burned acres per fire occurring 1963-2020 within 1250 miles of Pahrump, NV, as noted in the "GIS_Acres" field of the data) may over-estimate or under-estimate the actual burned acreage, though it is impossible to say by how much, and in which direction.

While we included all kinds of fires in our calculation (wildfires, prescribed fires, likely wildfire, Unknown - Likely Wildfire, Unknown - Likely Prescribed Fire), fires with a non-ring geometry (35 fires, <10% of data) were omitted from our data and burned area calculation. The data creators assume land can only burn once per year (usually true) so if areas are listed as burned twice in one year it is counted as one burn. Fires which burned in the same year within 500 meters of each other were also counted as the same fire, potentially reducing total fire count, and impacting the number of acres burned. Wildfires without years were removed from the dataset by the creators as much as possible.

For more information on the underlying data, intermediate data files, and processing steps please see the README located in our wildfire directory on GitHub.

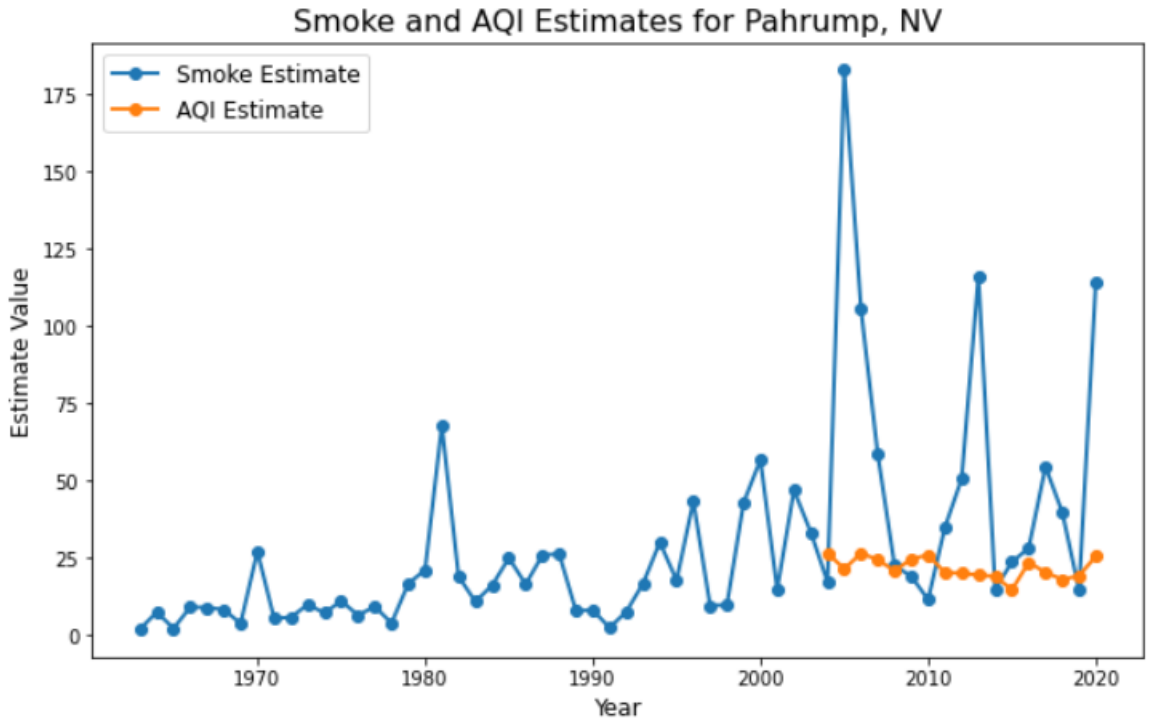


Figure 3 shows the Smoke and AQI estimates for Pahrump, NV between 1963 and 2020. The x-axis displays years (increasing from left to right), and the y-axis contains the estimate value. Users should note that AQI and Smoke Estimates are not on the same scale, thus it's their movement relative to each other which is interesting, not their absolute values. To read the figure, identify the line of interest using the legend in the upper left. Trace a point on the line from the desired year and find its corresponding value on the y-axis.

This figure shows that there does not appear to be a strong correlation in movement or intensive between the AQI measurements taken within Nye County (home of Pahrump, NV) and our smoke estimate. After normalizing the smoke estimate and AQI we find a correlation of 0.13 (see epa_comparison script for more information and graph). This correlation isn't very strong, but we are heartened to see that both estimates are moving in the same direction (e.g., positively correlated). Because we have detailed the raw data and its processing extensively in the previous two figure descriptions, we will focus on how the smoke estimate and AQI estimates were created.

The following section pulls from documentation in the data_processing script. For more information, please reference that file. The smoke estimate is built from features considered most relevant by the author, specifically type of fire, acres burned, distance to town, recency of other fires in the same area, and year of fire. If given more time, it would also be interesting to pull vegetation for burned areas (e.g., arid climates may have less to burn and thus less smoke) as well as weather conditions (e.g., windy conditions could disperse smoke more readily than stagnant conditions).

After consulting literature and estimating the various factors' impacts (see data_processing script for more detailed information and sources) we decided prescribed burns produce half the smoke of wildfires, there is a linear relationship between acres burned and smoke produced, there is a 1/n^2 relationship between distance from fire and smokiness, fires which attempt to burn in areas burnt within 2 years produce 20% as much smoke as they otherwise would, and data for fires prior to 1984

leads to a smoke underestimation by 50%. These factors combine to create a smoke estimate in units of acres/mile^2. We found that not all entries of fires with previously burned acres were able to be parsed uniformly, so given the percentage of fires with unreadable pre-burned acres was less than 10%, we assumed no pre-burned acres for those fires. If the percentage of fires with unreadable pre-burned acres was >10%, the data_processing program would have printed a warning message to the user.

Because we created smoke estimates at the annual level not monthly level, an "amortization" of smoke throughout the year did not feel necessary. Further, parsing fires by sub-year dates was unreliable due to multiple dates recorded from merging datasets (even years is somewhat unreliable, see USGS metadata for more information). We also assumed most fires are contained to "fire season", making it unlikely that smoke from 1 year will bleed into the next year. Annual smokiness estimates were averaged across all individual fire smokiness estimates that occurred within the year. We chose to do this because when comparing with other estimates of pollution (e.g., AQI) we felt the average of values would be most appropriate.

Our AQI estimate was gathered from sensors in Nye County. While AQI is typically the maximum AQI measured across several gasses or particles, our county sensors only measured one kind of particle, so it was the only variable included in the AQI measurement. Additionally, we chose to create annual AQI estimates by averaging the daily AQIs captured over the fire season (May 1st - Oct 31st) in the county. We chose not to include AQI estimates from the months outside of fire season to ensure a more equal comparison with our smoke estimate which occurs primarily during fire season.

Collaboration Reflection Statement

In this assignment I learned how many variables can impact smokiness estimates. It seemed obvious that acres burned and distance from fire would be relevant factors, but the more I read through the data and external sources, the more variables I found. Ultimately, I added in multiplicative factors to account for kind of fire, recency of other fires in the same area, and year of fire data collection (see data_processing script for more details on the math and reasoning behind these variables). If I had more time and data, I would have liked to include weather information, fire fuel information, and topographic information. Overall, I found this project very interesting, and it gave me good experience in creating and documenting a unique metric.

The possibility of collaboration had little impact on how I thought about the problem. I don't tend to work with others on homework assignments to get the coding practice. In general, I find that coding has the largest impact on my time to complete an assignment, rather than logically solving or structuring a problem. Unfortunately, I discovered that this assignment was very time intensive for me to complete alone (likely more than 30 hours). Collaborating with others could have allowed me to move through sections more quickly by dividing the work. Additionally, I could have learned new insights from my classmates.

I did discuss with others how they were creating their smoke estimates. Most of my friends and I chose to use acres/distance^2, with some adding other factors. I also talked through how to calculate AQI with other members of the class. In future assignments I may try to collaborate more (where allowed/appropriate) to learn and share new ideas with my peers and make large projects more manageable.