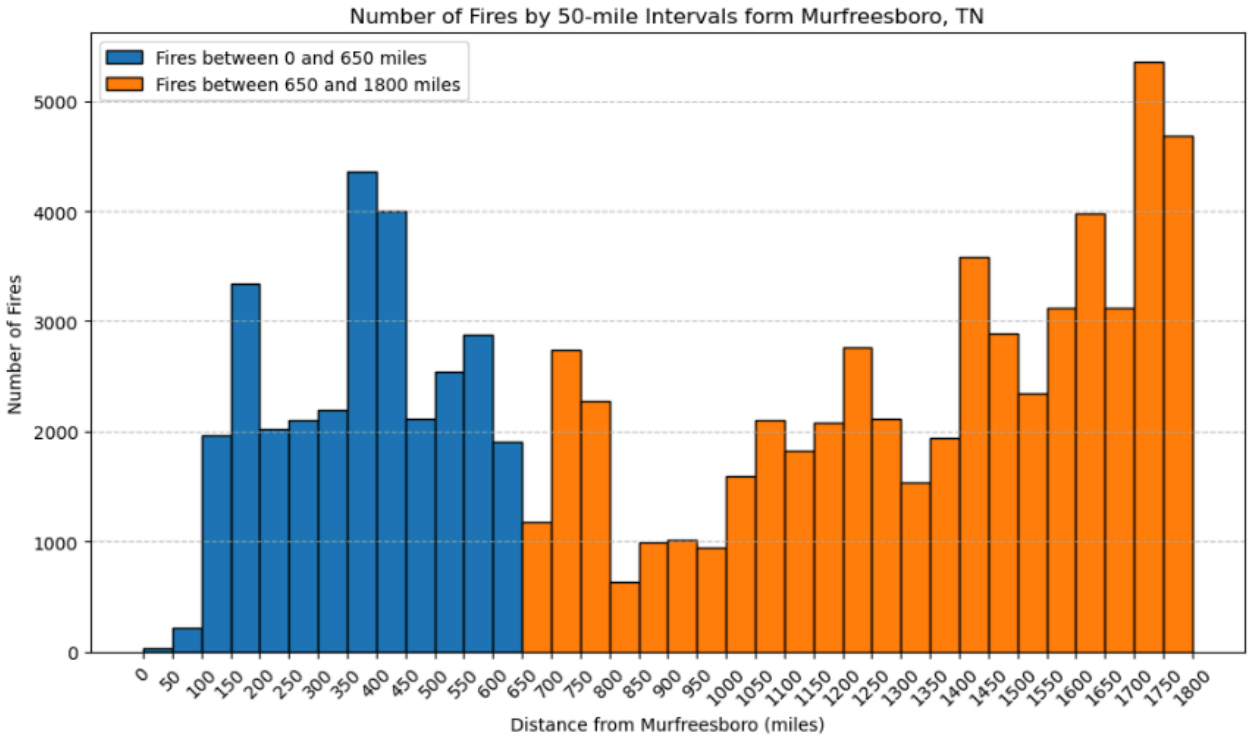
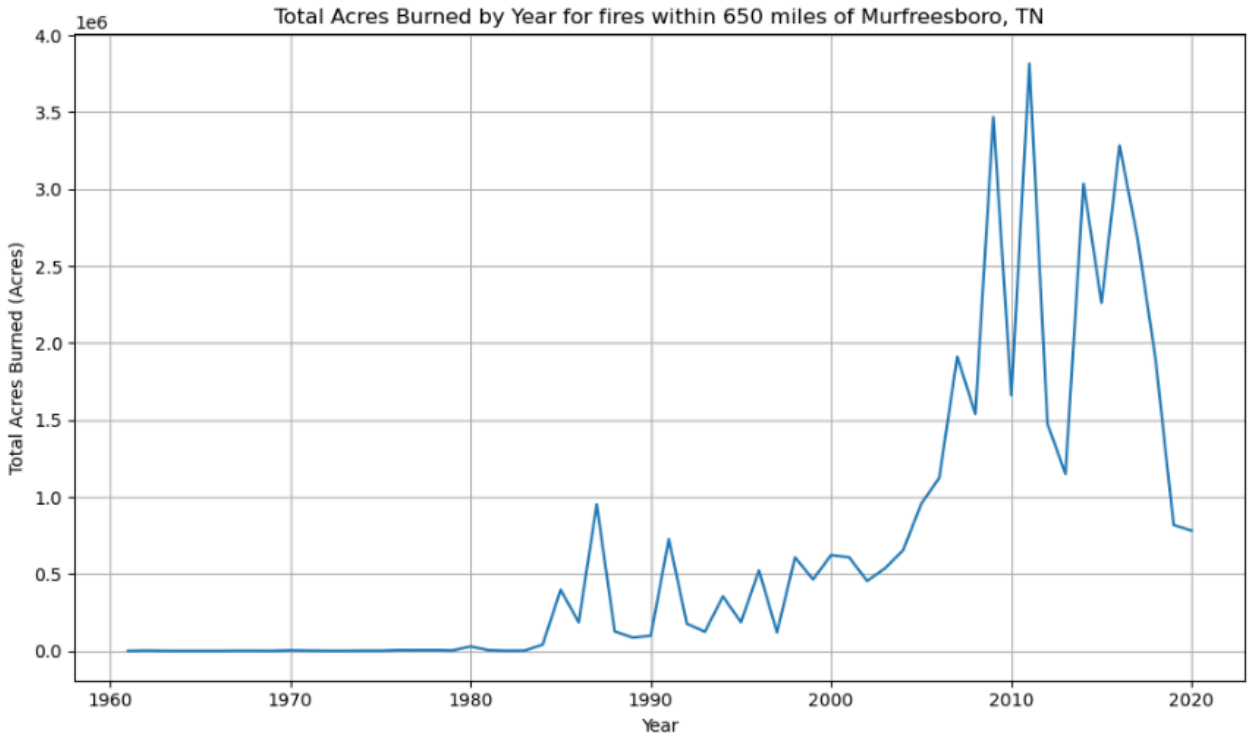
Visualizations:



This first visualization is a histogram of the number of wildfires within 1800 miles of Murfreesboro, TN from 1961- 2021. The fires are binned in intervals of 50 miles for the histogram. The y-axis represents the number of fires and the x-axis represents the distance of the fire in miles from Murfreesboro. The data in blue are the fires that are within 650 miles of the city, and are the fires that were used for my predictive model . The data in orange are fires that are between 650 and 1800 miles from the city.

A user can “read” this graph by looking at the height of the bars and the corresponding on the y-axis to see how many fires occurred at the distance. The distance can be found by looking at the x-axis to see what bin the bar is in. The bins represent intervals of 50 miles. A histogram is useful to show users the distribution of the number of fires across different distances.

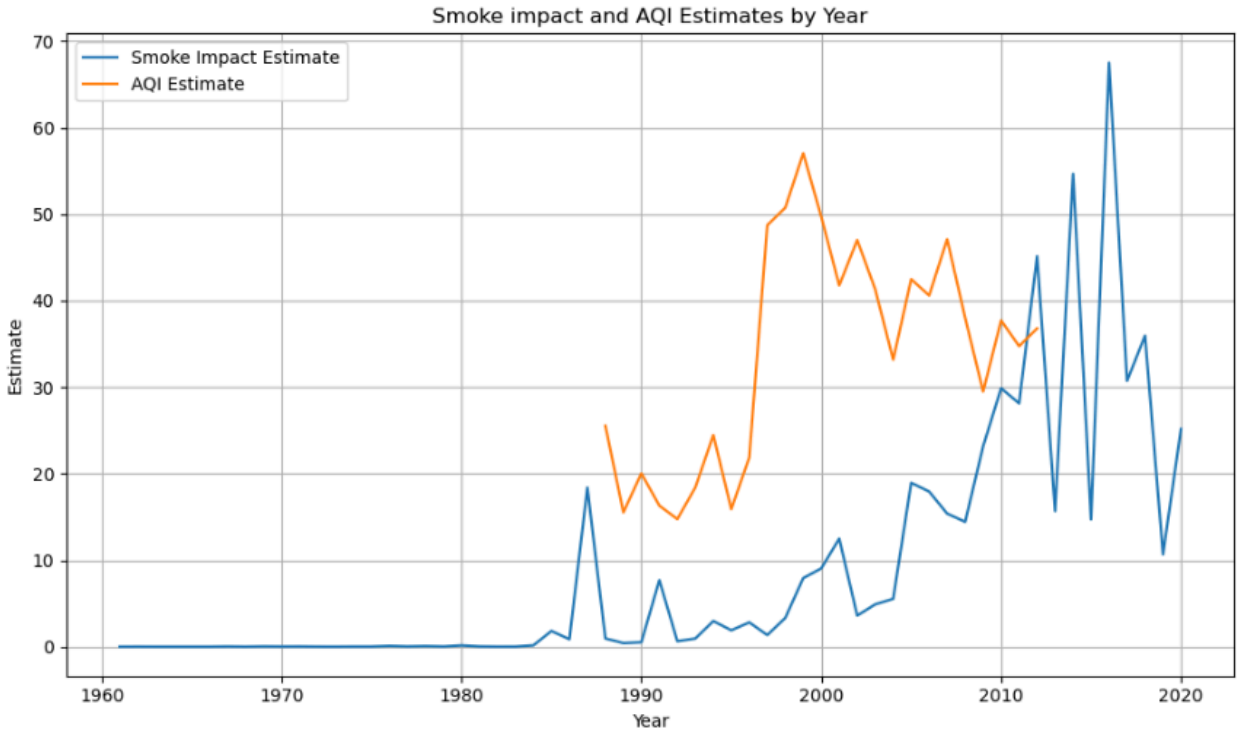
The data for this graph came from the wildfire dataset and was read in and filtered through for fires that occurred in the years 1961 to 2021. Every fire in the dataset has a ring of geospatial coordinates. The distance is calculated as the average of the distances between Murfreesboro and every point in the fire ring. For all of these calculations, the coordinates have to be converted from the ESRI:102008 coordinate system to the EPSG:4326. Additionally, meters have to be converted to miles. Then the fires were filtered for ones that were within 1800 miles and 650 miles.



This time series graph shows the total number of acres burned annually during fire season, May 1 to October 31, within a 650-mile radius from Murfreesboro, TN. The x-axis represents the year from 1961 to 2021, while the y-axis shows the total acres burned for each year. Each data point on the line represents the total acres burned across all fires for that year. The line shows the overall trend of the total acres burned by wildfires from 1961 to 2021.

A user can “read” the graph by looking at the trend line to see how fire size has changed from the years 1961 to 2021. A peak in the graph shows the years with the most total acres burned by wildfires and the lowest points indicate years with the lowest total acres burned. A user could imply that the years with the most acres burned had a more severe fire season than those with a lower amount of total acres burned.

The data for this graph came from the wildfire dataset and was read in and filtered through for fires that occurred in the years 1961 to 2021. The fires were also filtered for ones that were within 650 miles of Murfressboro, TN. The distance calculations were calculated as the average of the distances between Murfreesboro and every point in the fire ring which is a ring of geospatial coordinates. For all of these calculations, the coordinates have to be converted from the ESRI:102008 coordinate system to the EPSG:4326. Additionally, meters have to be converted to miles. Then for each fire meeting this criteria, the acres of all fires for each year were summed up to get the total acres burned for each year.



This time series graph presents wildfire smoke impact estimates and AQI estimates for Murfreesboro, TN, from 1961 to 2021. The x-axis represents time, where each point is a year between 1961 and 2021, while the y-axis shows the smoke impact estimate and AQI values, indicating levels of smoke intensity and air quality, respectively. The graph's two trend lines allow users to observe changes over time, as well as to compare the two and see the relationship between smoke impact estimates and AQI. The blue line is the smoke impact estimate and the orange line is the AQI estimate.

Users can follow the two trendlines over time to see estimates at any given year or see any potential relationship between the smoke and AQI estimates. While both trends show a slight upward movement over the years, they differ in their peaks and dips, suggesting that other factors also contribute to the estimates. The computed correlation of 0.28 suggests a weak positive relationship. While smoke may influence AQI, the metrics differ in how they were computed and the components of them.

The smoke impact estimate was calculated using the Inverse Square Law, accounting for the reduced effect of smoke intensity with increased distance from the fire. To ensure calculations avoided division by zero, a distance of one mile was added, resulting in the formula:

$$\text{Smoke Impact} = \text{intensity} \times \left(\frac{1}{(\text{distance} + 1)^2} \right)$$

For each fire meeting criteria (occurring between 1961-2021 and within 650 miles of Murfreesboro), the smoke impact was calculated, then summed by year to produce an annual smoke impact estimate. Distances were derived as averages across each fire's "fire ring" coordinates, which were converted from ESRI:102008 to EPSG:4326, and from meters to miles.

The AQI estimate represents an annual average for Murfreesboro, TN, based on daily data from a single monitoring station capturing SO_2, NO_2, and O_2 levels from 1988-2012. For each day, the maximum AQI across pollutants was selected, and an annual AQI average was calculated. Due to limited pollutant data available in Rutherford county, this AQI estimate may not fully reflect smoke specific impacts. Oxygen is not really impacted by wildfire smoke, but sulfur levels can be if the fire is burning vegetation with a lot of sulfur. NO2 however, is a part of what makes up wildfire smoke, so these levels should be directly correlated with wildfire smoke levels. If I had a monitor that had particulate values like PM2.5, my AQI estimate would be a better detection of wildfire smoke.

Collaboration:

The collaborative aspect of this assignment was very beneficial due to the open ended nature of answering this research question. There were many aspects of this part of the project that required us, as researchers, to make decisions and interpret the given instructions. Overall, I do think that collaboration in this environment is helpful. However, before I get into the benefits that I had, I do want to take a brief second to acknowledge the ways that collaboration on an open ended project like this can be slightly limiting. I believe that when there are so many decisions to be made while processing data and analyzing it, if you discuss ideas with people who are far ahead of you in the process, it can be easy to just listen to what decisions and processes they have done and just follow them. I think that collaboration can be most beneficial when you have already done your own work and research so that way you have some sort of idea on what path you want to take and have reasons to back it up. This allows collaboration to be truly collaborative and allow both sides to explain where they were coming from in their thought process rather than one simply following the other.

I had 2 positive collaboration stories that I would like to share. The first is that the task of estimating the smoke impact was very open ended, and I was not sure which direction to go and once I had a direction, I was not sure if I was thinking about it correctly. After doing some research on formulas for how substances disperse, I found the inverse-square law. This was a great fit in my opinion because the data points that I had and thought that I should use would be the acres burned and the distance of the fire from the city. I discussed this idea with Sarah K. after she had also chosen a method and computed her smoke impact estimates. Sarah also based her estimate off the inverse square law and we both discussed how this made sense, especially since the distance was squared. The dispersion of smoke should not just be a linear decrease as you move away from the fire. Sarah explained to me that she did distance + 1 in her denominator of her smoke impact equation which is something that I did not do. She explained that she did this to avoid dividing by 0 in the denominator if the fire was 0 miles away from the city. This is very unlikely and not present in my dataset, but when thinking about developing a formula for the estimate, I thought it was a really good idea to account for these edge cases. Especially because we will be expanding this project, so I wanted it to be robust. Therefore, I added 1 to my distance in my smoke impact calculation.

The second collaboration story that I would like to share is about the predictive model. My initial thought for the predictive model was to just use the smoke impact estimates as the feature for my time series data. After looking into some time series model options, I was debating between using ARIMA and Exponential Smoothing. I discussed with Manya, Apoorva, and Sarah N. about how they were going about their predictive model. After discussing ARIMA and exponential smoothing, and the different use cases for both of them, we decided that ARIMA was probably the better choice due to the fact that it handles long term time series better and Exponential Smoothing does a better job of predicting short term. The data we are using for our time series model is over many years, 60 years of

historical data and 25 years at least of forecasting. Another aspect of the predictive modeling that we discussed was what features to use in the predictive model. I came into the discussion wanting to use only the smoke impact estimate values that I had calculated using my formula discussed above. Some of the other people I was discussing with were thinking about including other features of the fire data, like acres and distance, burning time, in their models. However, as they discussed this, I did not see the point in adding these because my smoke impact estimate was calculated using the distance and acres already, so including acres and distance as separate features in the model would result in them being autocorrelated. Additionally, adding in a feature like burning time, would also most likely be correlated to the acres burned because the longer a fire burns, the more likely it is to burn more acres. I am not sure what everyone ended up doing for their models, but I ended up using ARIMA with smoke impact estimate as my feature.

These two collaborative experiences were different but still very helpful because I had my own opinions and ideas going into them. The first where I was discussing the smoke impact estimate, I was introduced to a new idea I had not thought about, and I ended up incorporating it. In the second experience, I was introduced to a new idea, but the discussion we had around only reinforced my initial stance.