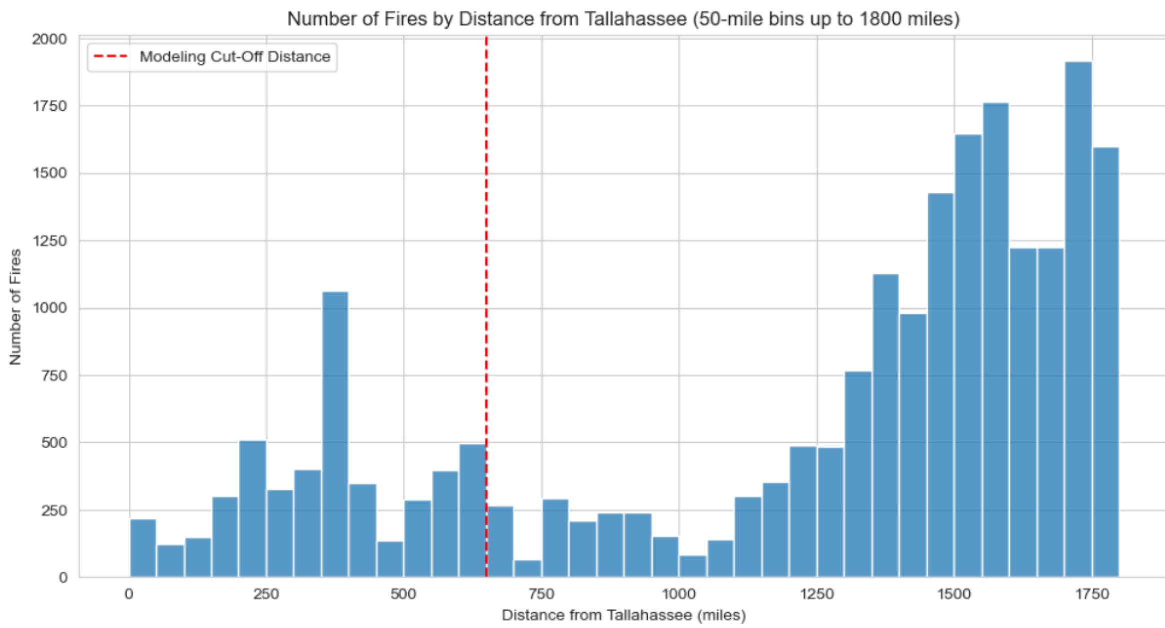
Part 1 Write Up

Visualization Descriptions

Number of Fires by Distance from Tallahassee:



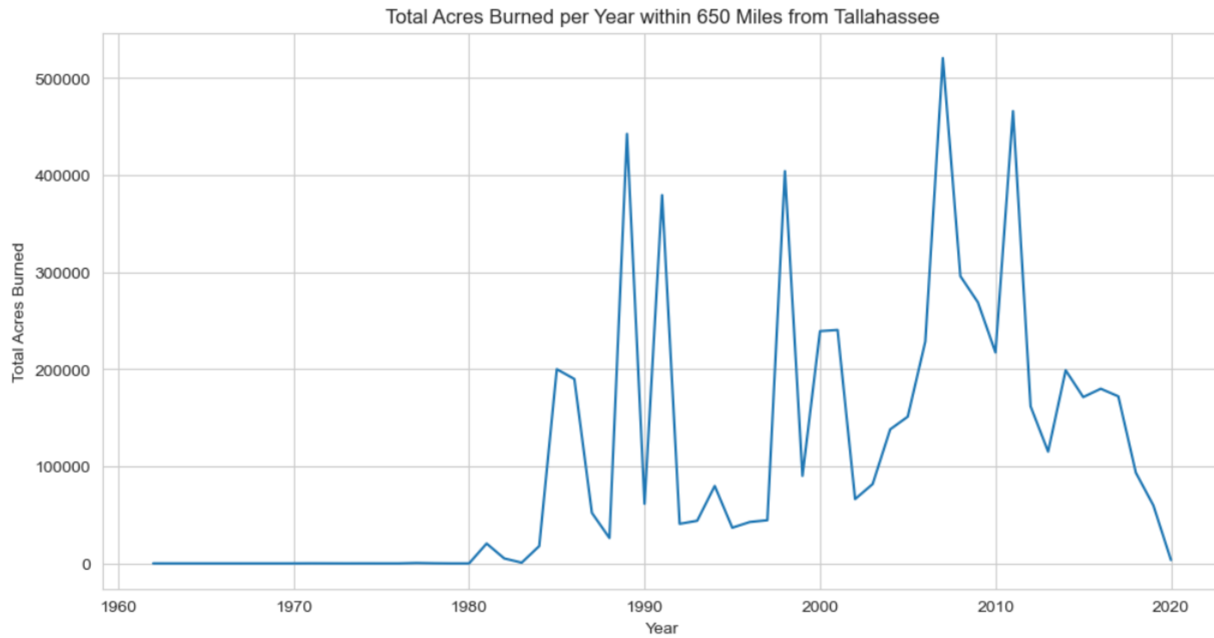
Description: This histogram visualizes the distribution of wildfires by their distance from Tallahassee, with distances binned every 50 miles up to a maximum of 1800 miles. This data was derived from the wildfire GeoJSON distance calculations (Subset Wildfire Data and Calculate Distances.ipynb), and represents the closest point on a fire's perimeter

Axes: The x-axis represents the distance from Tallahassee in miles, and the y-axis shows the number of fires recorded within each 50-mile bin.

Data Interpretation: The histogram provides insights into where wildfires occur relative to Tallahassee, with a higher concentration of fires at certain distances. The red dashed line marks the 650-mile cut-off, indicating that only fires within this distance are considered for the primary analysis. From the plot you can see that there are many more fires that occur outside of a 650-mile radius of Tallahassee, which makes sense because a large portion of that radius projected into water, and for the areas that are on land, the majority of the fires in the United States occur in the western states.

Purpose: This visualization helps establish the scope of fire activity around Tallahassee and gives context for estimating smoke impact on the city.

Total Acres Burned per Year within 650 Miles from Tallahassee



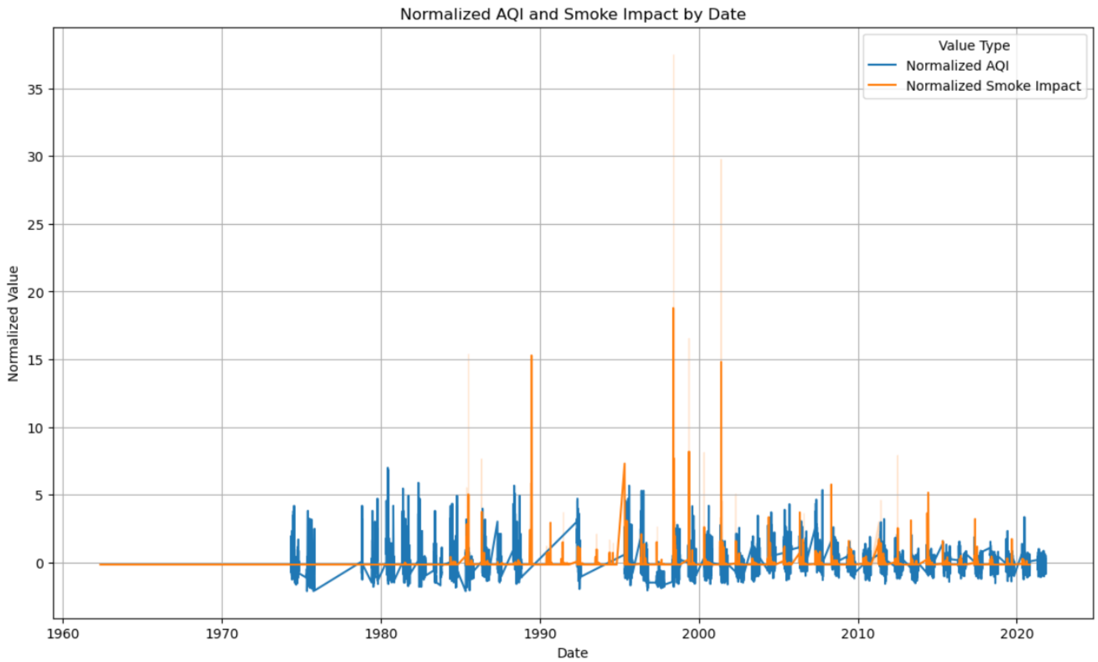
Description: This timeseries displays the total acres burned annually within 650 miles of Tallahassee from 1960 to 2021, aligning with the analysis's 60-year time frame. The underlying data is metadata provided in the dataset, but subset by the distance calculations (fires with a closest perimeter point within 650 miles of Tallahassee) that were computed in Subset Wildfire Data and Calculate Distances.ipynb

Axes: The x-axis represents years, and the y-axis shows the total acres burned each year within the specified distance.

Data Interpretation: Spikes in the graph indicate years with significant fire activity. This historical trend provides insights into the frequency and severity of fires that could contribute to smoke impact on Tallahassee.

Purpose: This visualization highlights annual variations in fire size and intensity, which are crucial for understanding patterns in smoke exposure and air quality over time.

Normalized AQI and Smoke Impact by Date



Description: This time series compares the normalized Air Quality Index (AQI) values and the estimated smoke impact from wildfires near Tallahassee over the same period.

Axes: The x-axis represents the date, while the y-axis shows normalized values for both AQI (blue line) and smoke impact (orange line).

Data Interpretation: Peaks in the smoke impact line typically coincide with elevated AQI values, indicating a potential relationship between fire activity and air quality in Tallahassee. If you look closely you will see that increases in the Smoke Impact slightly precede increases in the AQI, indicating there may be some lagging effect for smoke to impact AQI.

Data Normalization: The AQI and smoke impact values were normalized to allow for direct comparison on the same scale. Since AQI and smoke impact are measured in different units and can vary greatly in magnitude, normalization was necessary to reveal patterns and correlations over time. By bringing both datasets onto a common scale, it becomes easier to observe whether peaks in smoke impact align with increases in AQI, helping to validate the relationship between wildfire activity and air quality in Tallahassee.

Purpose: This chart serves to validate the smoke impact estimate by comparing it against AQI data, providing a reference for how wildfire smoke may have affected air quality historically.

Reflection Statement

As a part-time student who is new to this cohort, I had limited opportunities to collaborate with other students on this assignment. However, I was able to consult with some of my coworkers at the Allen Institute for Neural Dynamics, who offered valuable suggestions on the modeling aspect. Although I frequently work with neural time series data in my job, the nature of this dataset posed unique challenges. The data we handle at work lacks seasonality and is collected at regular intervals, which differs significantly from the wildfire data in this assignment, where irregular intervals and seasonal factors are prominent.

These differences required additional effort in data cleaning and model selection, particularly when accounting for seasonal trends in wildfire occurrences. My coworkers' insights were helpful in understanding how to approach modeling, even though adapting those techniques to a dataset with seasonal components and irregularities was challenging. This experience underscored the importance of customizing analytical approaches to the specific characteristics of each dataset.

Additionally, while I could theoretically select the perfect model for my data, the act of actually implementing that model was a large hurdle to overcome as well.