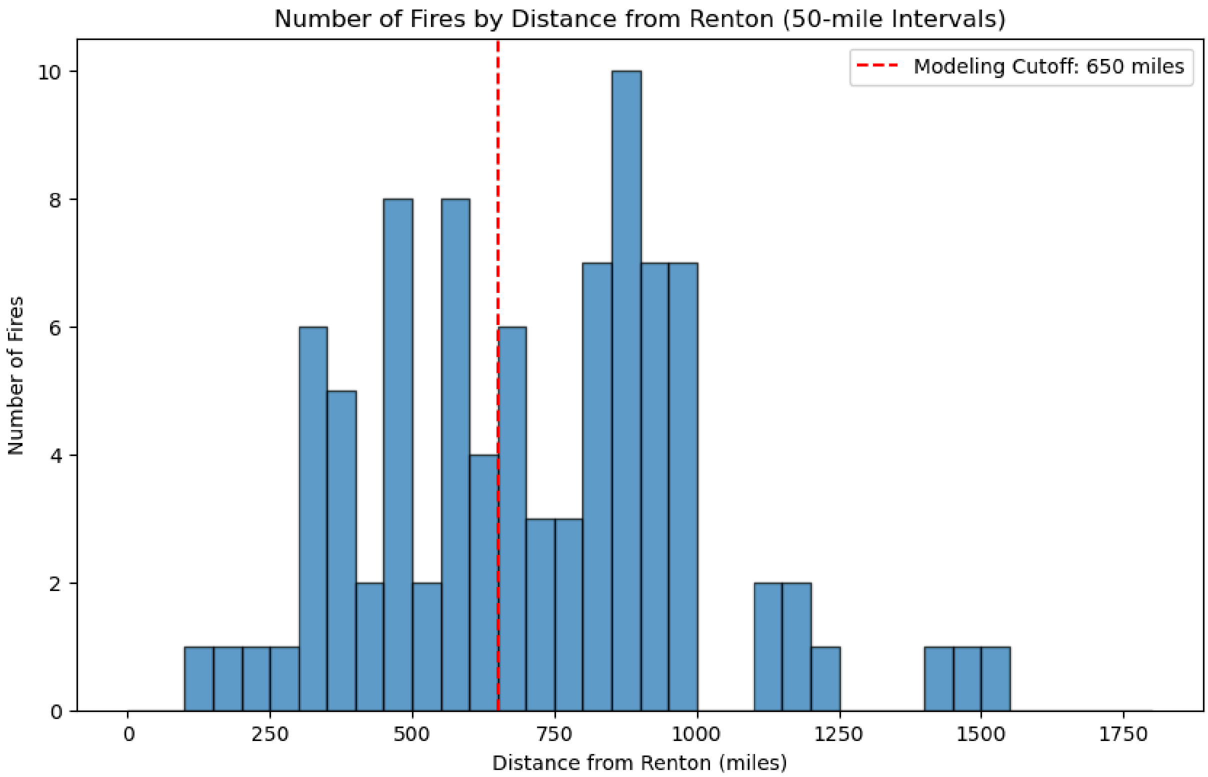
Salah Elbakri
Common Analysis Part 1, Write and reflect

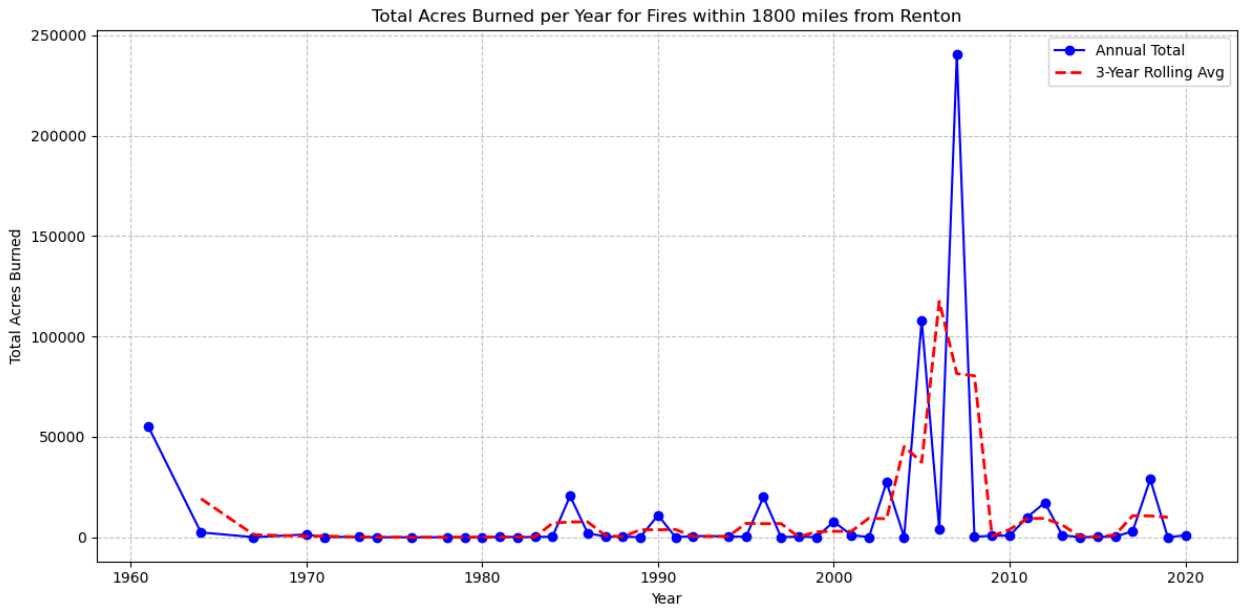
Visualization 1: Histogram of Fires by Distance from Renton, WA



This histogram represents the distribution of wildfires by their proximity to Renton, Washington. The x-axis shows distance in 50-mile increments up to a maximum of 1800 miles, while the y-axis depicts the count of wildfires within each distance range. This plot was created to identify how frequently fires occur relative to Renton and provides insight into how distance affects the potential smoke impact on the city. The data for this visualization was derived by calculating the distance from each fire's coordinates to Renton, WA, using the coordinates that surround the area of the wildfire. We then filtered fires within 1800 miles to ensure relevancy to Renton's air quality impact.

Reading this chart involves examining each bar's height within a given distance bin to gauge the frequency of fires within that range. Notably, there is a higher concentration of fires around 500 miles away, suggesting that Renton is more affected by regional fires. This pattern aligns with the geography of fire-prone areas within the Pacific Northwest.

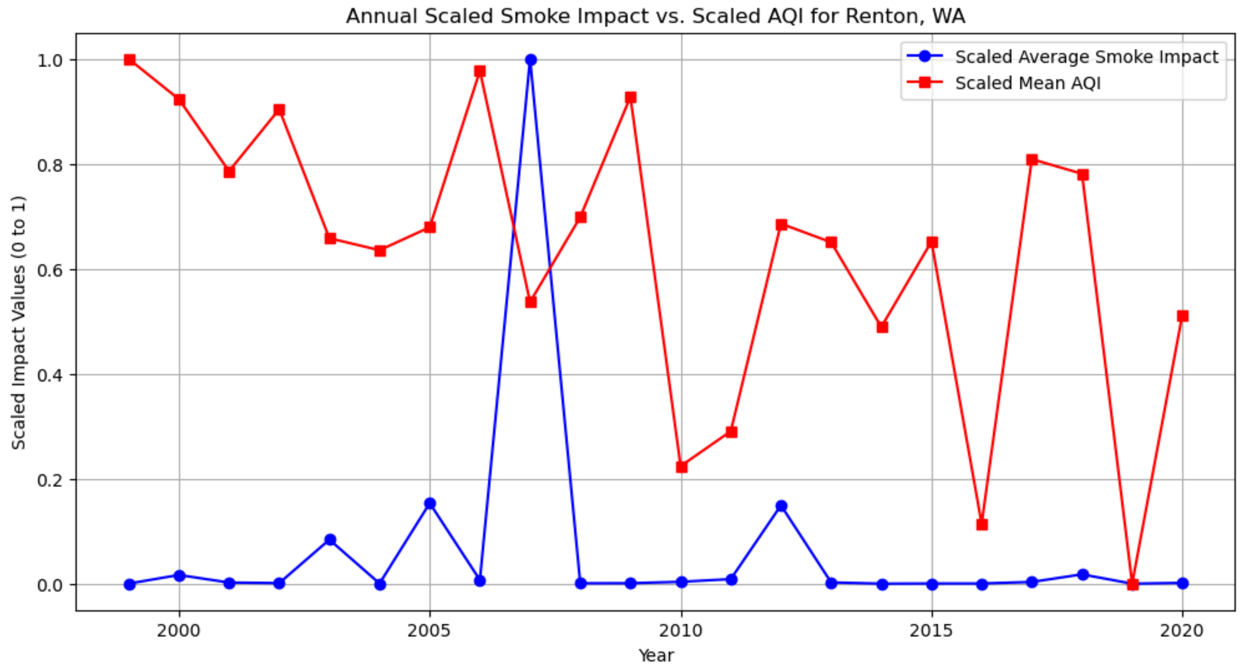
Visualization 2: Time Series of Total Acres Burned Per Year within 1800 Miles of Renton



This line graph illustrates the total acres burned annually by wildfires within 1800 miles of Renton, from the earliest recorded data through 2020. The x-axis represents the year, and the y-axis represents the cumulative acres burned in each year. To construct this visualization, we filtered fires within the 1800-mile range and aggregated their acreage by year. The graph offers insights into the long-term trend of wildfire acreage around Renton, providing a contextual background for assessing smoke impacts and AQI.

The line graph shows fluctuations in burned acreage over the decades, with notable peaks that likely correlate with particularly severe fire seasons. This trend points to the potential influence of climate patterns and changing wildfire intensity, which in turn impacts air quality. The viewer can easily identify years with high wildfire acreage, which are relevant for assessing spikes in Renton’s smoke impact.

Visualization 3: Time Series Comparison of Smoke Impact vs. AQI for Renton, WA



This time series chart compares the annual smoke impact estimate (blue line) and the AQI (Air Quality Index, in red) for Renton from 1985 to 2020. The x-axis shows the years, while the y-axis shows the magnitude of the smoke impact and AQI. The smoke impact score was calculated by weighting the fire acreage by its inverse distance to Renton, adjusted by the fire's severity when possible, while the AQI data was gathered from the EPA's database for particulate matter (PM2.5), which is heavily influenced by wildfire smoke.

This visualization allows viewers to observe the alignment (or divergence) between wildfire activity and air quality in Renton. A strong overlap between spikes in smoke impact and AQI trends would indicate that wildfire smoke significantly influences local air quality. Notably, in some years, AQI remains steady even when smoke impact fluctuates. This could indicate other AQI-influencing factors, suggesting that smoke impact alone may not fully account for AQI variability. Consequently, it underscores the need for refined smoke impact modeling, potentially incorporating additional environmental or atmospheric variables.

Reflection

The project started with the USGS wildfire dataset, which includes detailed records of wildfires across the United States. The initial steps included filtering the data to include only fires within a 650-mile radius of Renton to focus on those most likely to influence local air quality. This cutoff was specified in the assignment and guided by the observation that fires beyond this distance tend to have limited direct smoke impacts.

Data preparation also involved processing the wildfire coordinates, date of occurrence, fire size, and other attributes necessary for calculating smoke impacts. Notably, I spent considerable time troubleshooting the loading process for the large JSON file only to discover that it was corrupted during download. After multiple attempts, I eventually requested a classmate to send a copy, which allowed me to proceed with the analysis.

Fire Size: Fires with a larger area are assumed to produce more smoke, making fire size a significant component of the impact score.

Distance from Renton: Fires closer to the city are assigned a greater impact score since smoke concentration decreases with distance.

In the early stages, I experimented with a Severity Metric “attributes_Fire_Polygon_Tier” that incorporated both fire size and fire type, hypothesizing that certain fire types or larger fires would produce more smoke. This metric applied greater weight to fires flagged as severe, aiming to capture their presumed higher impact on air quality. However, the Severity Metric added complexity without yielding clear, actionable distinctions. I ultimately removed it in favor of a simpler, more direct approach based on distance-weighted impact.

The Smoke Impact Score is an initial estimate that will evolve with future parts of the project. Currently, the score does not account for factors like wind direction, wind speed, or other meteorological conditions that can drastically influence smoke travel. Adding these elements in future work will help make the score more reliable.

To model and quantify the impact of wildfire smoke on Renton’s air quality, I tried several approaches before settling on a final model. Here’s a summary of the models I tested and the reasoning behind my final choice:

1. **Multiple Linear Regression**: The first model I tried was a multiple linear regression, incorporating factors such as distance, fire size, and fire severity to predict a potential impact score. However, the results were mixed, as the linear regression model did not adequately capture the diminishing influence of fires at greater distances. The linearity assumption simplified the relationships too much, leading to inaccurate smoke impact estimates.
2. After experimenting with several approaches, Random Forest emerged as the most effective model for this dataset. The initial attempts—linear regression and a custom severity metric— fell short in capturing the full complexity of smoke impact factors. The Random Forest model, in contrast, delivered better results by accommodating non-linear relationships, which are essential because fire characteristics and air quality effects do not follow simple patterns.

This project was mostly completed independently. One of the primary challenges was working with the large JSON dataset, which required significant troubleshooting to read properly. After realizing that the file had become corrupted during the download, I reached out to a classmate who provided a working copy. This collaboration helped me proceed with the analysis without additional delay.