Amit Peled
DATA 512: Human Centered Data Science
30 October 2024

Part 1 - Common Analysis: Peoria, AZ

Data Visualizations

1. Histogram of Fire Occurrences by Distance (up to 1800 miles)

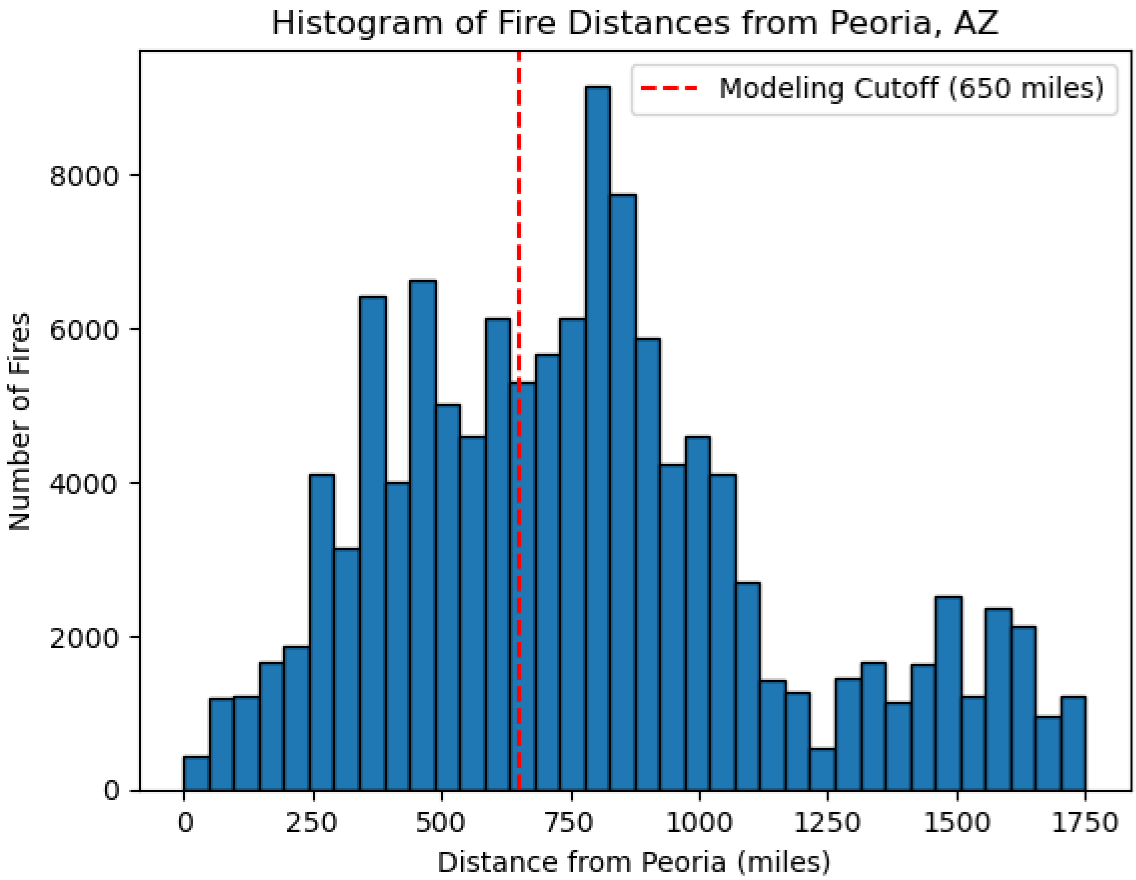

Figure 1. This histogram illustrates the distribution of wildfires at different distances from Peoria, AZ. The fires are grouped into 50-mile bins, up to a maximum distance of 1800 miles away from Peoria. The red line at 650 miles signifies the distance cutoff for modeling.

Axes: The x-axis represents distance from Peoria in miles, while the y-axis shows the count of fires within each distance bin. The red line at 650 miles marks the distance cutoff used in this analysis for estimating smoke impact.

How to Read: Note where the majority of fires are clustered, especially within the first few hundred miles. Peaks in the histogram highlight distances where fire events are most frequent, indicating the proximity of wildfire risk to Peoria.

Underlying Data and Processing: The wildfire data was filtered to include only events within a maximum radius of 1800 miles, from 1961 to 2021. Fires were counted by their distance from Peoria in increments of 50 miles.

2. Time Series: Smoke Impact and AQI Comparison

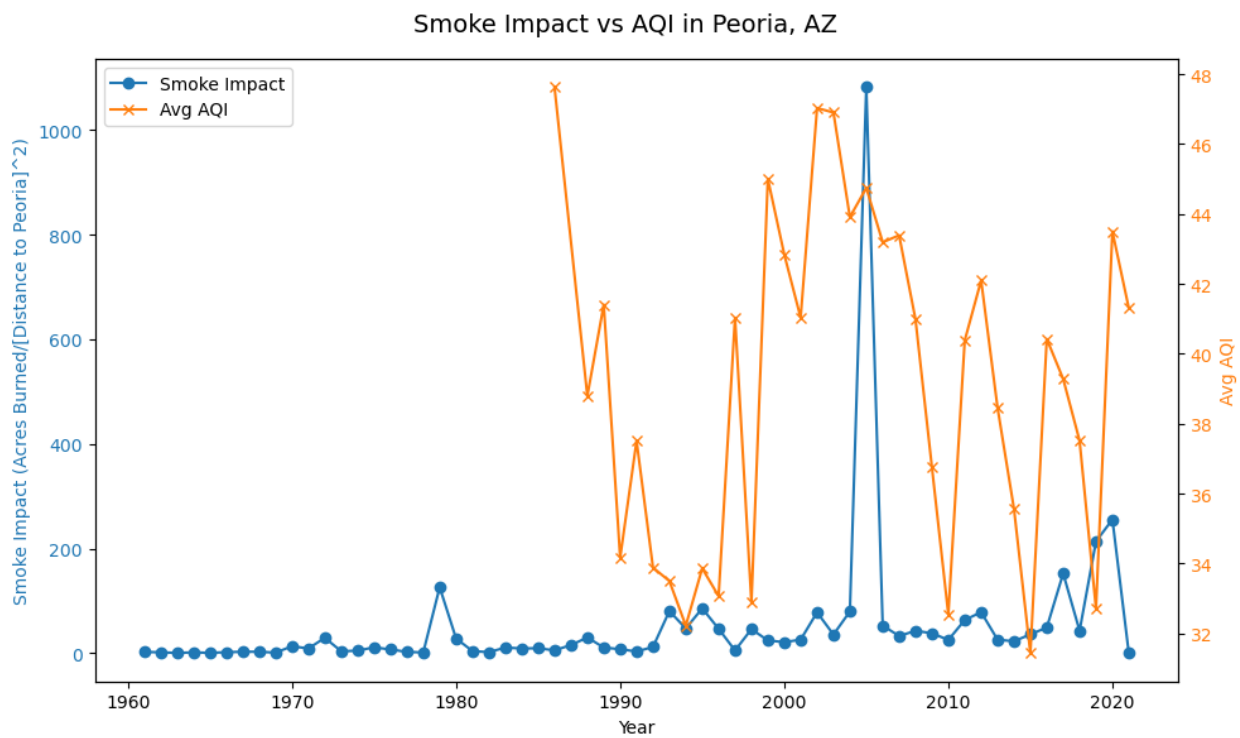

Figure 2. This time-series plot compares the wildfire smoke impact estimates for Peoria with the annual average AQI (Air Quality Index) over time, using dual y-axes to distinguish between metrics.

Axes: The x-axis shows the years from 1961 to 2021. The left y-axis (in blue) represents smoke impact estimates, while the right y-axis (in orange) shows AQI values.

How to Read: Try and observe any correlation between smoke impact and AQI, such as simultaneous peaks or sustained high values across both metrics, which could suggest a relationship between wildfire activity and air quality. Any divergence between smoke impact and AQI trends might indicate other factors influencing air quality.

Underlying Data and Processing: Smoke impact was estimated using proximity and fire size. The AQI data was aggregated from EPA monitoring stations, calculating average AQI for each fire season.

3. Time Series: Total Acres Burned per Year

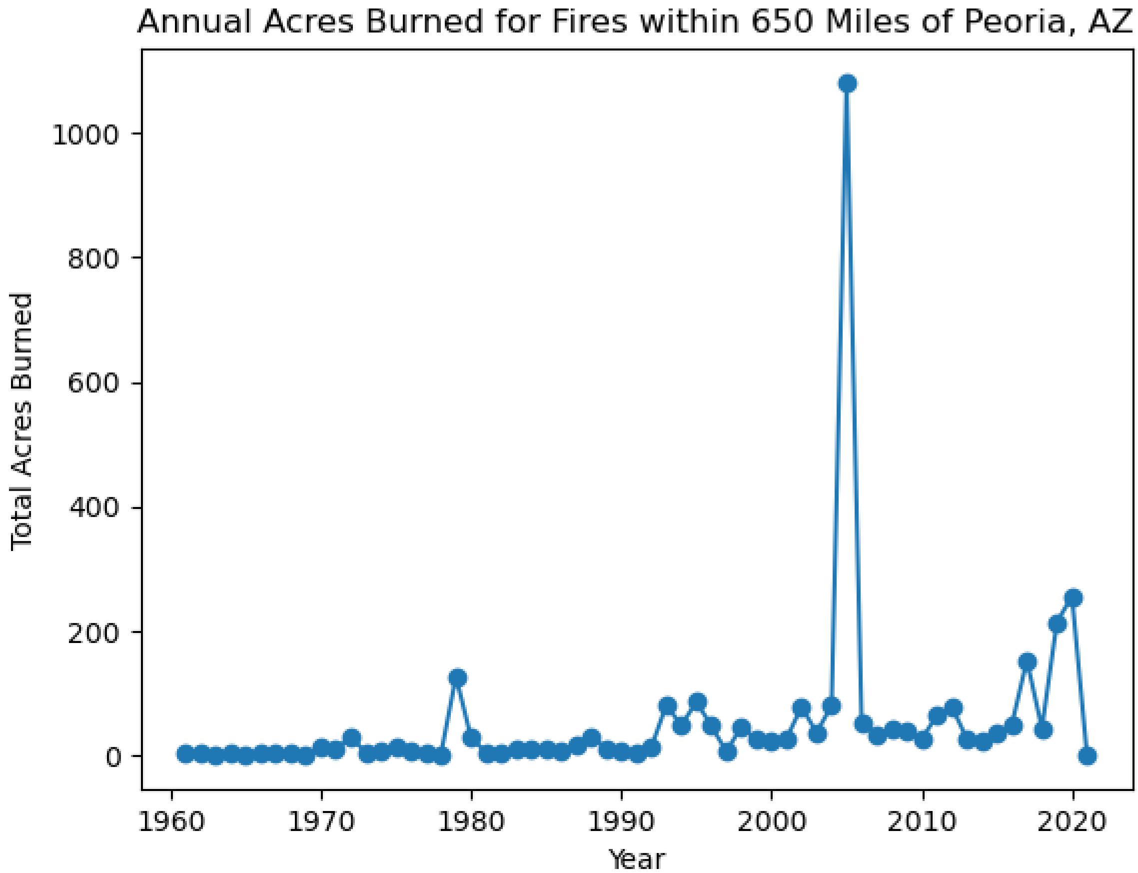

Figure 3. This time series figure tracks the total acreage burned per year within 650 miles of Peoria, showing changes in fire severity and impact over time.

Axes: The x-axis spans years from 1961 to 2021, while the y-axis shows the total acres burned within the specified range.

How to Read: Look for trends, such as years with high peaks indicating severe fire years, or any increases in the acreage burned over time. Periods with sustained high or low acres burned may suggest changes in fire frequency, climate impact, or other environmental factors.

Underlying Data and Processing: The annual acreage burned was calculated by summing all fires occurring within the defined distance and during the fire season (May 1–October 31) for each year.

4. Forecasted Wildfire Smoke Impact for Peoria, AZ (2025-2050)

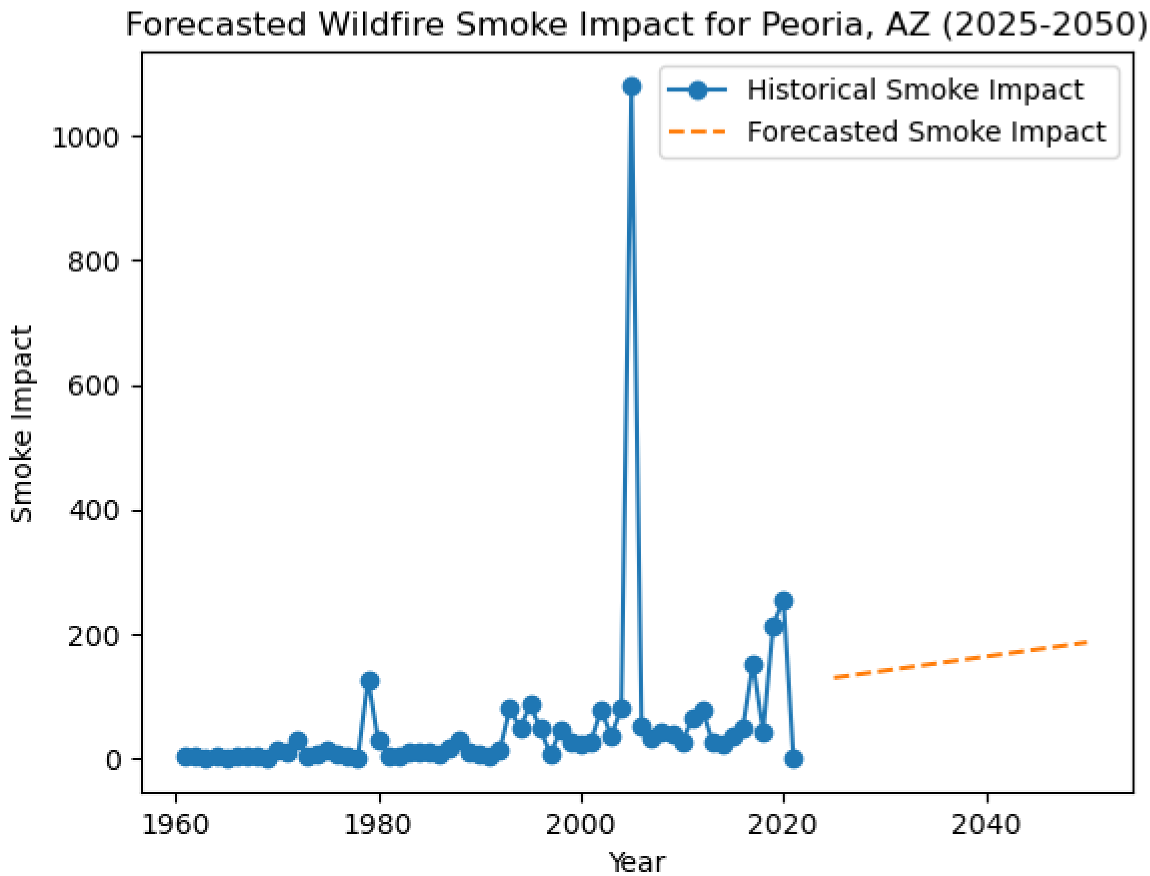

Figure 4. This forecast chart shows projected wildfire smoke impacts in Peoria from 2025 to 2050 (dashed orange line), based on historical trends which are also shown in the figure (solid blue line).

Axes: The x-axis represents years with historical data and future years with forecasted data, as it spans years from 1961 to 2050, and the y-axis shows the historical and predicted smoke impact values.

How to Read: Try and observe the projected trajectory and any upward or downward trends, helping to visualize potential future wildfire impacts. Dashed lines may indicate uncertainty, emphasizing that these predictions are based on past data trends and could vary.

Underlying Data and Processing: A linear regression model was used to predict future smoke impact, based on historical smoke impact estimates from 1961 to 2021.

Reflection

Working on this project has been an eye opening experience that deepened my understanding of environmental data analysis, especially in using wildfire and AQI data to model and interpret smoke impacts. One of the most significant aspects I learned was how to handle large, geospatial datasets effectively, particularly in calculating distance-based filtering for fires relative to Peoria. Initially, I underestimated the complexity of accurately determining the impact of wildfires based on proximity alone and this experience showed me the importance of carefully chosen distance metrics and filtering thresholds in achieving reliable estimates. And I did this for only one city in the United States!!

Using the geodetic distance calculations, based on Dr. David W. McDonald's Wildfire Proximity Computation Example, was essential in accurately estimating distances from fire perimeters to Peoria. This example code provided key functions and conceptual insights, like handling distance projections using the Pyproj library, that greatly aided the project's calculations. Collaborating with classmates proved highly valuable and led me to think of and apply a more refined approach for the distance calculations which were initially inspired by approaches we discussed in class and that were outlined in Dr. McDonald's example notebook. After consulting with Jake Flynn, he suggested that a dual axis approach for the plot, *Time Series: Smoke Impact and AQI Comparison*, would be a good way to show how the two information's compared. Hence after this consultation I went home and implemented the approach, making my data much more readable, allowing for viewers to extract insights in an easier manner.

Integrating EPA's Air Quality System (AQS) API data for AQI comparison was another challenging and informative experience. Given the lack of EPA AQI data for early years in our analysis range, I initially struggled to bridge gaps in the data. The US EPA Air Quality System API Example by Dr. McDonald was very helpful here, providing clear examples of retrieving and structuring data from the API in a manageable way. The code snippets for requesting data with pagination, county-based, and bounding-box search options, as well as dealing with varying AQI parameters, saved significant time and effort in handling and validating the data.

The project also emphasized the complexities in air quality modeling as a whole. While AQI data gives a snapshot of general air quality, correlating it directly to wildfire smoke is complex due to additional factors like wind patterns and pollutants from other sources. By comparing my smoke impact estimates to EPA AQI data, I was able to identify limitations in how AQI alone reflects wildfire smoke effects. The process made me more aware of the nuances in environmental data and how such data can be influenced by various factors that are not immediately apparent.

Specific Attributions:

- I adapted proximity calculations and geospatial functions from Dr. McDonald's Wildfire Proximity Computation Example, which provided key methods using Pyproj and GeoJSON libraries, essential for calculating accurate fire distances.

- The code structure and API handling methods from the US EPA Air Quality System API Example were reused and modified to access daily summary AQI data for Peoria, which significantly streamlined the process of data retrieval and preprocessing.
- Thought processes for various methodologies of calculations, modeling, and visualizations were discussed with my classmates, but all code and final code are specific to my own work.