Sarah Kilpatrick
Common Analysis Part 1 - Writeup

Visualizations

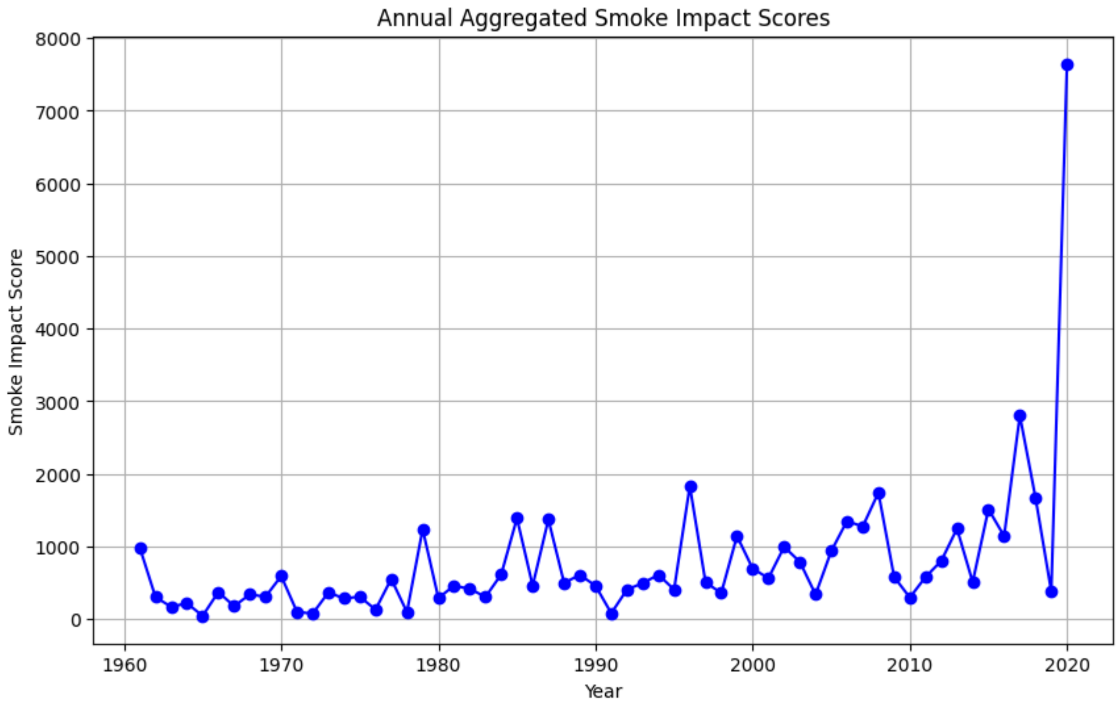


Figure 1. *Annual Aggregated Smoke Impact Scores*

Description: This figure displays the evolution of total smoke impact estimates between 1961-2020. The viewer reads the figure by observing the increasing variability as time goes on. The most notable extrema occurring in 2020, a "record-setting year" in California. According to Wikipedia, "8,648 fires burned 4,304,379 acres". Note that there was no wildfire data for 2021, even when attempts were made to search for it in the original USGS wildfire dataset.

Axes:

X-axis: ["Year"] – represents the temporal scope of this analysis.
Y-axis: ["Smoke Impact Score"] – measures the cumulative impact scores for each year.

Data and Processing: The underlying data originates from aggregating a custom function developed for this assignment, as seen below. Smoke estimates found via the same formula.

$$\sum_{day=1}^{365} \frac{\text{Acreage Burned} \times \text{Weighted Cardinal Direction from Fresno}}{(\text{Distance in Miles from Fresno} + 1)^2 / 5}$$

Image 1. Formula for total smoke estimate per year.

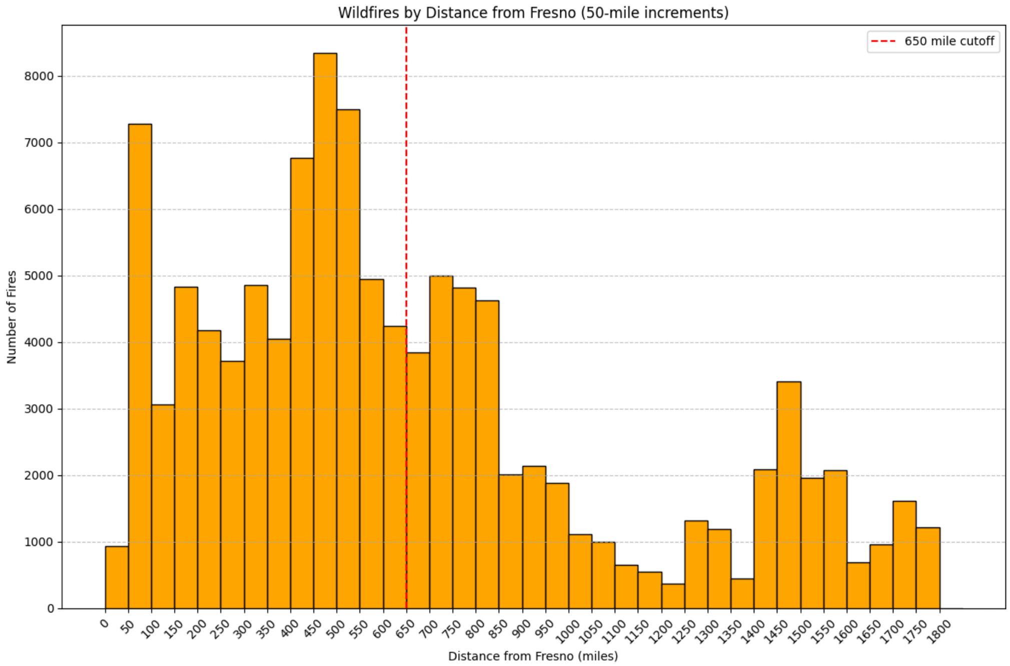
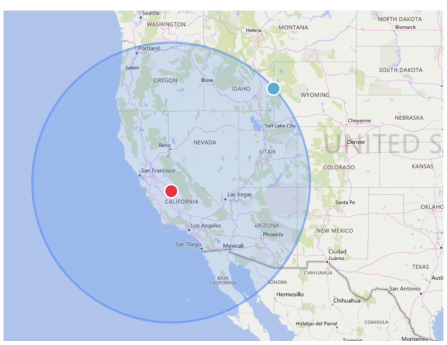


Figure 2. Count of Wildfires by Distance from Fresno, in 50-Mile Increments

Description: This figure illustrates the bulk of wildfires during fire season (May 1 - Oct. 31) between 1961-2020 occur within 650 miles of Fresno. The viewer can interpret this conclusion by noting the red dashed line representing the 650 mile cutoff.



Axes:

X-axis: ["Distance from Fresno (miles)"] – categorizes wildfires by their closest distance to Fresno, indicating proximity
Y-axis: ["Number of Wildfires"] – displays the amount of wildfires to provide a sense of scale

Image 2. 650-mile radius about Fresno, CA

Data and Processing: The data reflects the sum of all wildfires, with each bar representing the amount of wildfires within that mile range. For example, the first bar represents the count of all fires between 0-50 miles from Fresno. Processing involved iterating through the list of fires from the USGS dataset, finding their closest distance to Fresno using Professor McDonald's function, and categorizing them based on the closest distance to Fresno.

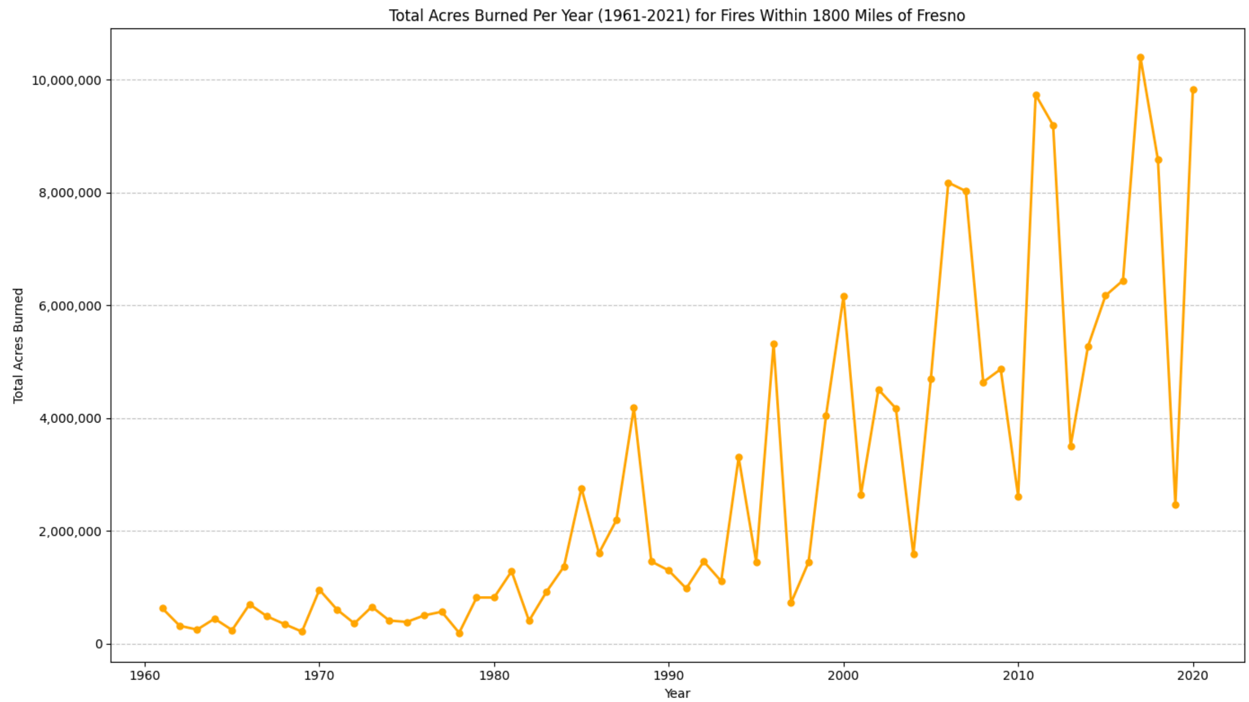


Figure 3. Total Acres Burned per Year for Fires within 1800 Miles of Fresno
Description: This figure shows the increasing variability of burned acreage over time. It's read by comparing the low-variability and low-burned acreage of the 1960s through 1980s to the highly extreme and variable burned acreage of the 21st century.

Axes:

X-axis: ["Year"] – represents the temporal scope of this analysis.
Y-axis: ["Total Acres Burned"] – quantifies the smoke impact, facilitating a visual analysis of the influence of distance on the smoke score.

Data and Processing: Derived from the original USGS dataset, the data was processed by analyzing the fire year and the acres burned.

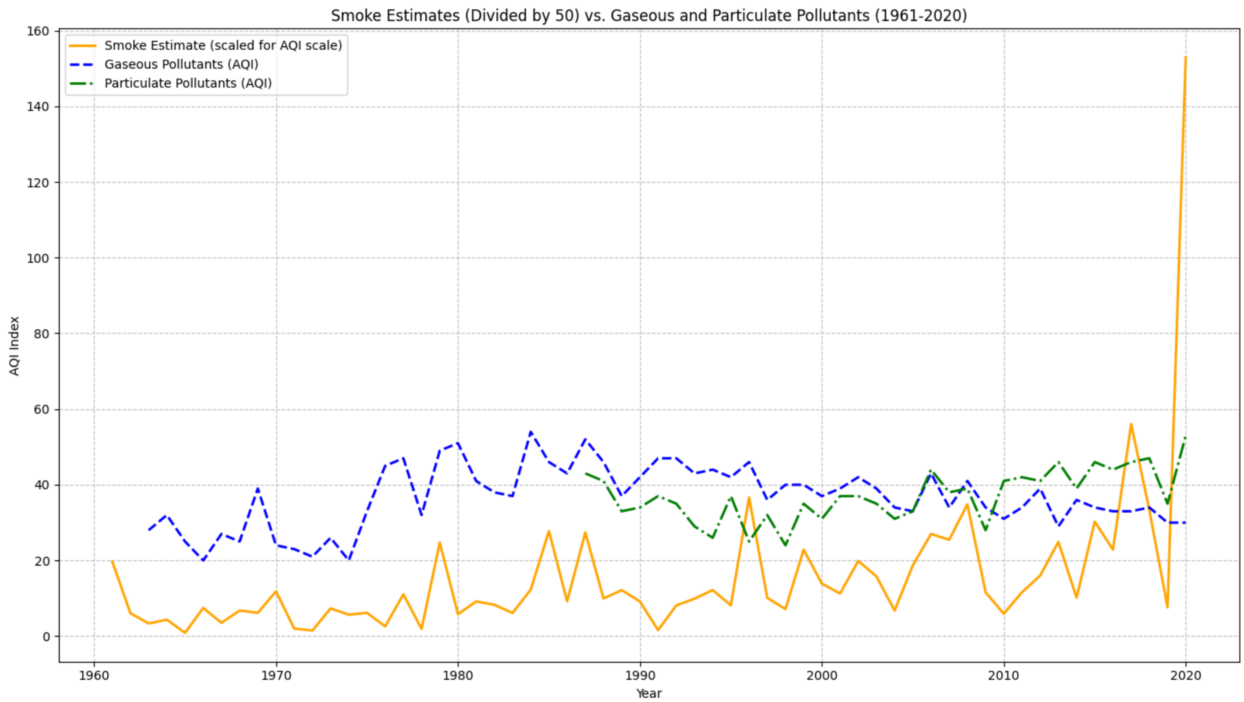


Figure 4. Fire Smoke Estimates vs. Gaseous and Particulate Pollutants, Scaled for Viewability

Description: This figure presents a comparison between the evolution of this project's smoke estimate and the EPA-provided air quality data. This helps the viewer recognize where the smoke estimate succeeds and where it fails by observing smoke estimate trends and changes mirrored by the EPA AQI trends.

Axes:

X-axis: ["Year"] – represents the temporal scope of this analysis.
Y-axis: [“AQI Index”] - shows all trend lines on a familiar scale used for the Air Quality Index

Data and Processing: Sourced from both the smoke estimate by year dictionary, and the aggregate gaseous and particulate pollutant data by year from the EPA's API. Processing involved dividing the smoke estimates by 50 to make them better scaled to the AQI data.

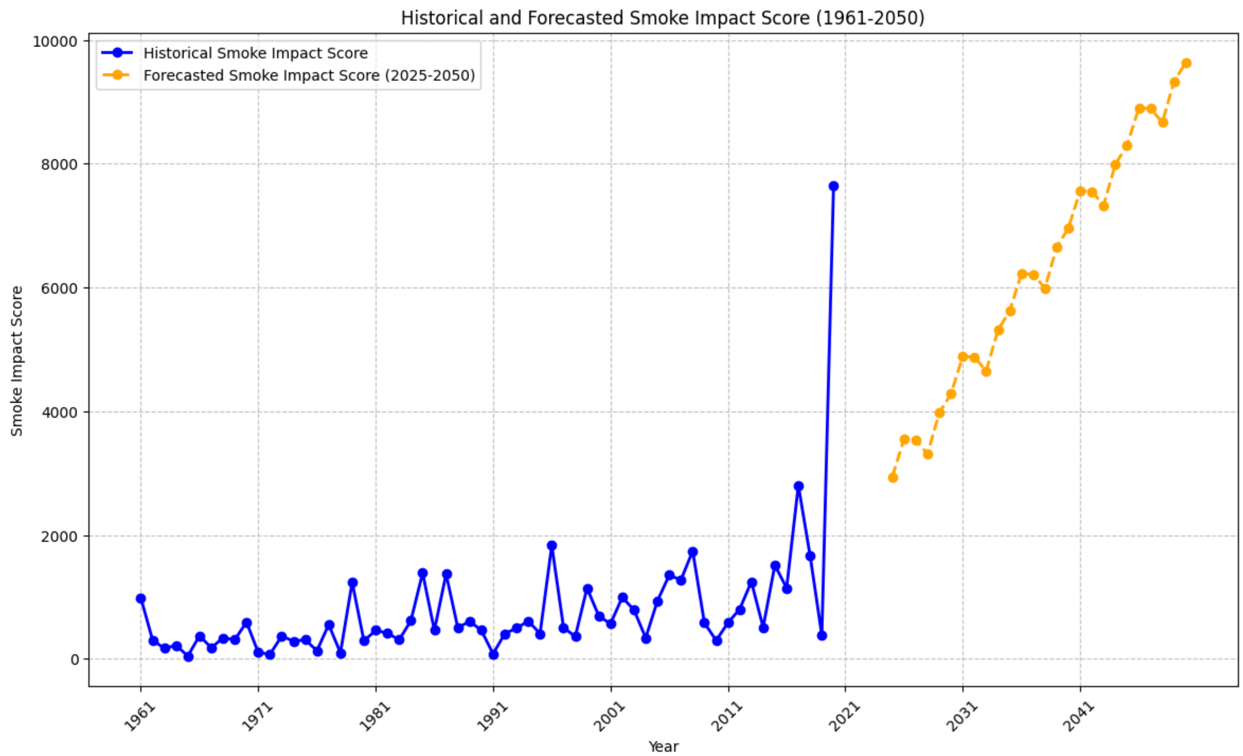


Figure 5. Historical and Forecasted Smoke Estimates

Description: This graph presents a possible forecast for future smoke estimates between 2025-2050. It does so by taking into account the historical trends of the previous few decades. Viewers can draw conclusions that smoke estimates are expected to rise in the coming decades, if these estimates are to follow the trajectory of the previous decades.

Axes:

X-axis: ["Year"] – represents the temporal scope of this analysis.
Y-axis: [“AQI Index"] – represents the AQI scale

Data and Processing: This visualization uses smoke estimate data, aggregated by year. The forecasted smoke estimates are done using the Holt-Winters Seasonal model. This model is also known as triple exponential smoothing. It is well-suited for forecasting annual wildfire smoke impact because it smoothly handles noisy data, filtering out short-term anomalies. However, it has limitations in forecasting information when this information is impacted by irregular or nonlinear dynamics.

Sarah Kilpatrick

Reflection

Through Part 1 of this project I gained a deeper understanding of how to interpret air quality data in relation to wildfires over time. Specifically, I learned the importance of using publicly available data from sources like the EPA's Air Quality Index (AQI) database. Using the API example code taught me how much information is accessible for studying the large-scale, cumulative impacts of climate change on Fresno, CA.

Using the AQI data helped quantify air quality during wildfire seasons over many decades. This revealed broader trends indicating the escalating influence of wildfires on population centers. For example, by comparing annual average AQI to an independently-generated smoke estimate, I identified potential worsening patterns, such as increased particulate matter during specific months in recent years. This could signal that wildfires are becoming more frequent and intense.

Throughout the visualizations I started noticing patterns that weren't immediately obvious from just looking at the raw AQI data. For example, some years showed a mismatch between high wildfire smoke estimates and moderate AQI values, possibly due to prevailing winds or other factors that dispersed the smoke before it reached Fresno. While AQI provides a good amount of information, adding more localized estimates could refine our understanding of smoke impact variations.

The collaborative aspect of this project also significantly shaped my approach. It encouraged me to make my code more modular and reproducible. Knowing that other people might build on this analysis encouraged me to break down each data processing step into clear, reusable functions. For example, instead of hardcoding specific parameters or steps, I created generalized functions that could accept parameters like date ranges or pollutant types, making the code adaptable to various analyses.

When discussing the smoke estimate with classmates, I noticed that many were using a similar formula. However, no one I talked to had experimented with tuning it by adding a variable to capture the effect of seasonal wind patterns on smoke dispersion. This prompted me to add an additional factor in my smoke estimate formula to account for these variations. The collaborative exchange inspired me to go beyond a "standard" approach and explore creative representations of the smoke estimate.

Sarah Kilpatrick
Common Analysis Part 1 - Writeup

Attributions

1. Cardinal Direction Variable in Smoke Estimate
Basis provided by the Wind & Weather Statistics page from Fresno Yosemite International Airport.

2. Holt Winter's Exponential Smoothing
Theory and basis for code provided by Jason Brownlee in his article "11 Classical Time Series Forecasting Methods in Python"