## Extended Figure Captions

1. **Fire Distances up to 1800 Miles from Gresham, OR**
   This histogram breaks down the number of wildfires that happened at various distances from Gresham, OR, all the way up to 1800 miles, with each bin covering 50 miles. The x-axis shows how far the fires were from Gresham, while the y-axis gives us the number of fires within each range. There's a red dashed line at 650 miles, which marks the distance cutoff we used for modeling smoke impacts on air quality. This visualization gives us a snapshot of how fire activity varies by distance, showing that most fires occur within a few hundred miles of the city.
2. **Yearly Total Acres Burned within 650 Miles of Gresham**
   This time series tracks the total acres burned each year from wildfires within 650 miles of Gresham. Years are on the x-axis, and the total acreage burned (in millions) is on the y-axis. It highlights fire patterns over time, showing low activity during some years and big spikes during others. The recent uptick suggests climate or other factors might be fueling these changes. Overall, this graph helps us understand wildfire trends near Gresham, which could help us anticipate future air quality impacts.
3. **Comparing Fire Smoke Impact and AQI in Gresham**
   This graph lines up annual fire smoke impact estimates with Air Quality Index (AQI) values for Gresham, scaling both metrics between 0 and 1 for an easy, side-by-side comparison. Scaling helps us see how well these two trends match over time, even though smoke impact values are small (like 0.002), while AQI can reach much higher (e.g., 76+). The x-axis covers the years, and the y-axis shows the scaled values for each. Seeing both trends together lets us spot years when high smoke impacts seem to coincide with poor air quality, hinting at a potential link between wildfire smoke and AQI in Gresham.

## Reflection Statement

This project taught me a lot about working with environmental data and combining different data sources to answer complex questions. I learned about GeoJSON formats and working with spatial data, which was initially challenging. GeoJSON, which is used for encoding geographic data, was essential for mapping and analyzing fire locations in relation to Gresham. Although I was unfamiliar with it at first, I now have a much better handle on how to process and use geospatial data in analysis due to Dr. McDonald's examples his wildfire_geo_proximity_example.ipynb.

Working with GIS data and doing spatial analyses was a big learning curve. I had to figure out how to calculate the distance between fire locations and Gresham, and then incorporate that information into time series and regression models. This was all pretty new territory for me, and it was helpful that a few other classmates had worked with GIS data before and could offer some pointers. For example, the distances took approximately 20 minutes to calculate. A classmate told me that geopandas.GeoSeries.buffer gives you a GeoSeries of geometries representing all points within a given distance of each geometric object in minutes. While I did not implement it because I had discussed this with a classmate after I had finished running the distances, this would have been very helpful as I had to run the calculations multiple times.

I also learned more about air quality metrics and how the AQI is measured. I explored how agencies monitor air quality and calculate AQI, which was particularly interesting given the impact of wildfire smoke on public health. Understanding the AQI and doing the research to figure out how it was calculated, which pollutants were relevant to wildfires, and in general what it represents added real-world context to my analysis, showing me the importance of understanding your domain to make quality analysis.

One of the main takeaways from the collaborative aspect of this assignment was realizing the value of reusing code and learning from shared methods. For instance, I adapted some external methods for handling time series data and setting up ARIMA and VAR models for forecasting. This was one of my first times applying a time series forecasting model like this, so it was interesting to learn about the different types, like the Moving Average, Autoregressive, and ARIMA. I ultimately chose an ARIMA model due to its flexibility and interpretability. I used ChatGPT's guidance for ARIMA modeling, where it helped me choose models and understand implementation specifics, such as data requirements and transformations (noted in my common_analysis.ipynb). A lot of collaboration, whether by directly working with others or learning from shared resources (like others suggesting different types of formulas or methods on how they calculated the smoke estimate), helped me tackle areas where I felt a bit out of my depth, especially with more advanced statistical modeling. Additionally, I learned the struggles of interacting with files so big and had to figure out best practices on batching to make coding more efficient and how to upload these types of large files to GitHub.

Overall, this project expanded my understanding of handling big data and environmental data science, especially with integrating spatial and temporal data.