Writing and Reflection

Visualization Explanations

Graph 1

Produce a histogram showing the number of fires occurring every 50 mile distance from your assigned city for all fires ranging up to 1800 miles away from your assigned city. Your histogram should indicate the distance cut-off for your modeling work.

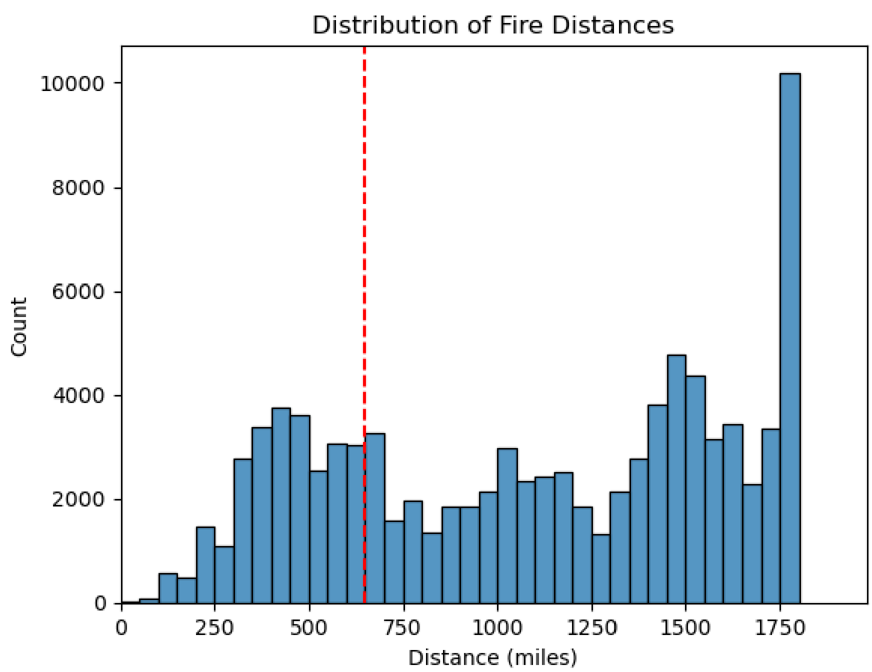

Figure 1: Histogram for Distribution of Fire Distances

This histogram shows the frequency of fires occurring at various distances from Indianapolis, IN, organized into 50-mile bins. The data comes from the combined USGS ArcGIS GeoDatabase, which includes attributes like fire date, distance from the city, smoke estimates, area burned, fire type, and additional details.

For this visualization, we used the number of fire occurrences within each 50-mile distance interval, up to 1800 miles. The x-axis displays the distance from Indianapolis in miles, and the y-axis shows the number of wildfires. For additional detail, we have also divided each bar in the chart by its distribution of fire type.

The distribution appears fairly even, with a greater concentration of fires at further distances from the city. The highest number of fires occurs around the 1800 mile distance from Indianapolis, suggesting that Indianapolis itself is not a major wildfire hotspot, as fewer fires are observed closer to the city.

Finally, the dotted red line indicates the distance threshold (=650) that is used to perform the modeling and analysis.

Graph 2

Produce a time series graph of total acres burned per year for the fires occurring in the specified distance from your city.

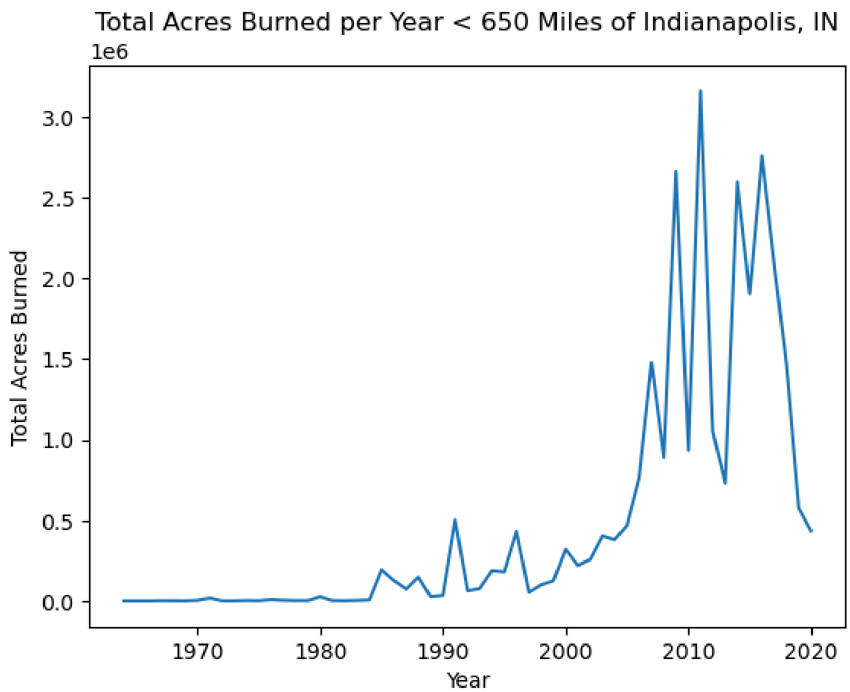

Figure 2: Time Series Graph of Total Acres Burned Per Year (within 650 miles)

This time series graph shows the area burned by fires near Indianapolis, IN. The data originates from the combined USGS ArcGIS GeoDatabase, which includes details such as fire date, distance from the city, smoke estimate, burned area, fire type, and other relevant parameters. For this visualization, we used the yearly total area burned by fires. The x-axis represents the years from 1963 to 2020, and the y-axis displays the total burned area in units of 10^6 acres to improve readability. It's evident that the area affected by fires has been increasing over time, with a significantly higher burn impact in the 2000s compared to the 1900s. Since the late 2000s, the trends have fluctuated massively, but on average, been quite high. Overall, this suggests that the impact of fires around Indianapolis has grown substantially over the last few decades.

Graph 3

Produce a time series graph containing your fire smoke estimates for your city and the AQI estimates for your city.

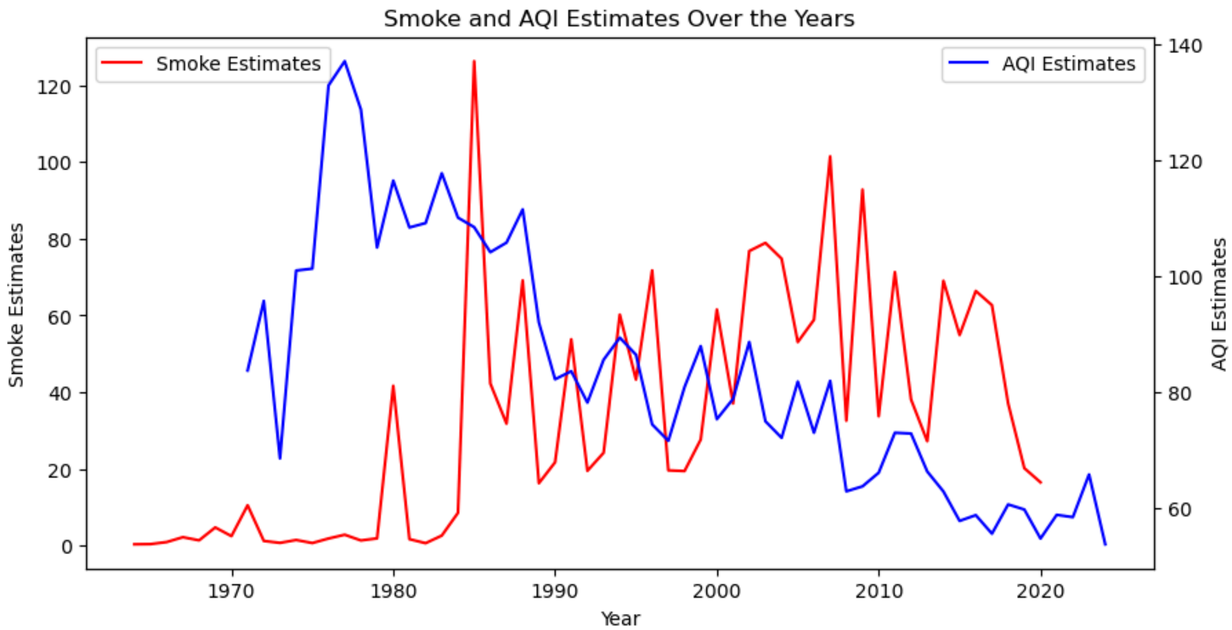

Figure 3: Time Series Graph for Fire Smoke Estimates and AQI Estimates

This time series graph represents the smoke estimates generated for Indianapolis, IN, alongside the air quality index (AQI) values over the years. The data comes from the combined USGS ArcGIS GeoDatabase, which includes parameters such as fire date, distance from the city, smoke estimate, area burned, fire type, and other relevant details. Additionally, AQI data was obtained from an API and aggregated annually to provide a single AQI value per year. The key features used in this visualization are smoke estimates, AQI, and fire year. The x-axis shows the years of fires (1970-2024), and the y-axis displays the estimated values. The red line represents the smoke estimate, while the blue line shows actual AQI values.

Interestingly, we observe a negative correlation between the two lines over time. In the initial years, the smoke estimate is quite low while the AQI values are relatively high. Over time, this trend reverses, with smoke estimates increasing and AQI values declining. This negative correlation suggests that, even as the prevalence and intensity of fires grew, Indianapolis's air quality improved, potentially due to effective air quality control measures implemented in Indiana. Consequently, despite an increase in smoke-generating fires, Indianapolis has managed to maintain or even enhance local air quality over the years.

Reflection on Collaboration Activities

This Part 1 of the project involved extensive data collection, feature engineering, data wrangling, visualization, and insight generation. The project involved dealing with USGS data, a large file in GeoJSON format. Right off the bat, I recognized potential challenges with loading it directly as JSON or a Pandas DataFrame. While I was aware that Professor David McDonald had provided a reader notebook for this, I decided to approach it with GeoPandas. This is owing to my previous internship experience where I utilized GeoPandas and PostGIS to deal with Geospatial analysis. Also, I had been informed of issues that several of my colleagues were facing with the distance calculation logic of the example notebooks, which I knew would be handled very easily by GeoPandas.

However, when I tried to load the GeoJSON into GeoPandas, I began to face several errors. It seems like there were certain records in the GeoJSON file that were not formatted in a manner compatible with GeoPandas. This was not too surprising as it turns out, the authors of this dataset had actually utilized ArcGIS for creating this dataset, and had later converted the data to GeoJSON. Thus, the GeoJSON file still had a lot of attributes that were more relevant to ArcGIS. Hence, after several hours (> 3 hours) of debugging, I eventually figured out the solution: Import the ArcGIS Geodatabase file as a GeoPandas dataframe, by utilizing a different driver available in GeoPandas. Once the data loading was done, all I had to do was figure out the appropriate Coordinate Reference System, and all the distance and time based filtering were easily done. All in all, my entire code of Geospatial data loading and filtering was reduced to around 10 lines. This code also runs within 1 minute, a radical improvement over the original code, that could take well up to 4 hours for a few students. I am thus, if allowed, willing to share this code to make the lives of future students easier, so that they can spend more time with analysis.

Once I had the wildland fire data, I focused on coming up with a smoke estimate formula, which I do believe would benefit from further improvements in the future. Then, I moved onto the AQI extraction. This time, I utilized the example notebook provided by Professor David McDonald, which benefitted me enormously. I ran the API call per year and aggregated by mean, a lot of coding to get the final AQI estimates. The final part involved visualizations, which were fairly straightforward, but do seem to serve as excellent preliminary insights into how I should be going forward with this analysis in the coming weeks.

The collaboration aspect of this assignment was mostly restricted to discussions on smoke estimate calculation and forecasting. Owing to the fact that I took a completely different approach to data loading and filtering, there was not a lot I could discuss with my colleagues, but it was nice to discuss this with the Professor after lecture. The code I have written, however, is entirely my own work, which mostly is thanks to my previous internship experience where I was extensively involved in GIS and data analysis.