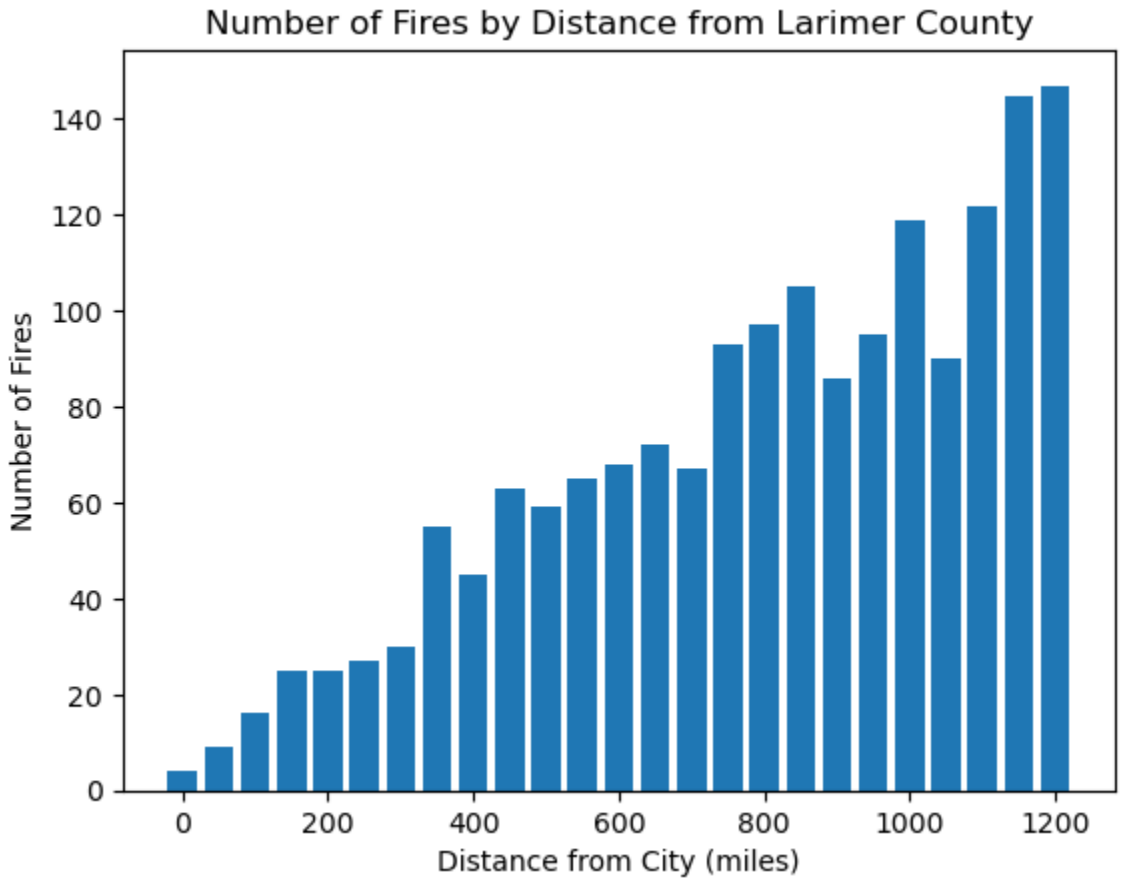
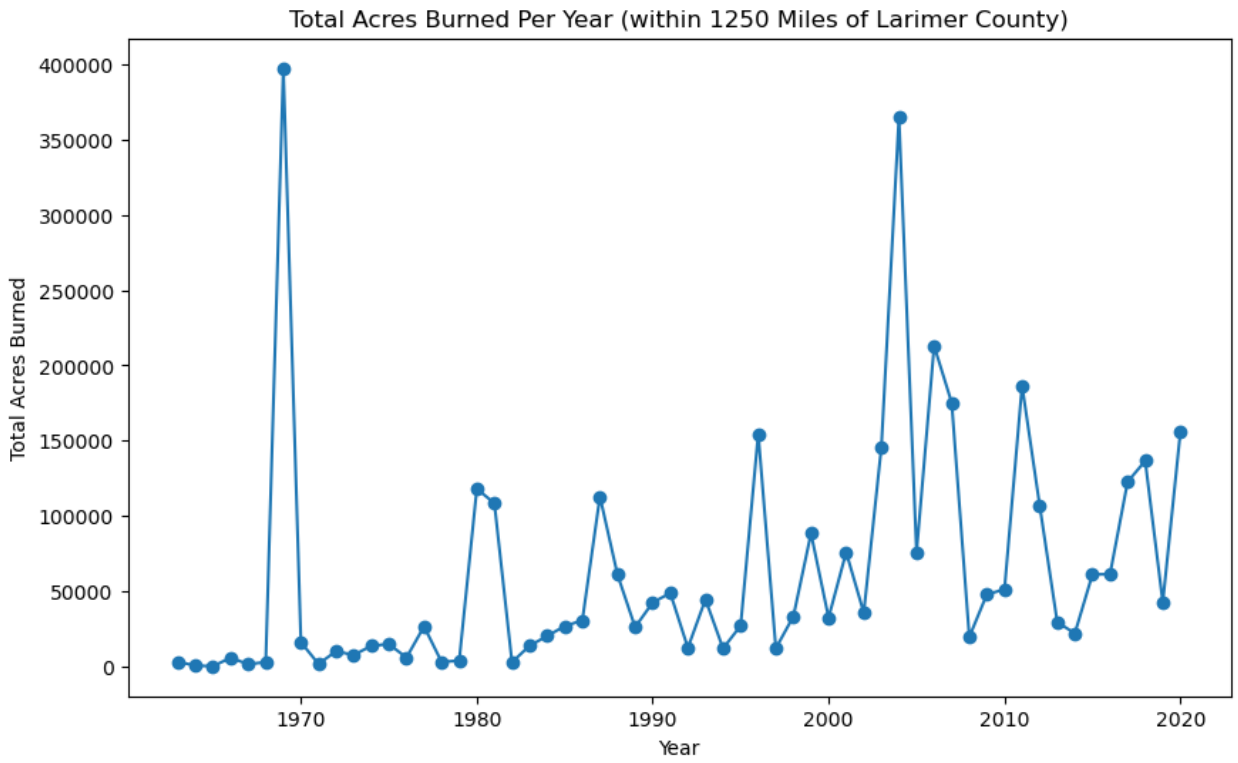
Peter Mohammadi



Visualization Explanation 1: Fires by Distance from Larimer County

This visualization represents the total number of Wildfires within 1250 miles of Larimer County. The x-axis is counted by miles, while the y-axis represents the total number of fires. Each bar is the total number of fires within a 50-mile block (ie: there were nearly 100 wildfires between 1200 and 1250 miles from Larimer). The data came from the USGS[1], where their people have combined all known fire data into one large dataset. This data was cleaned to only count wildfires if: it was from 1963 - 2023, if it was during fire season, if we are sure (or almost sure) it was a wildfire, and if it was within 1250 miles of our city. We counted a fire within 1250 miles of Larimer County based upon its centroid, not the outer ring. As we can see, the large majority of fires are actually over 600 miles away from our city. This is because a lot of fires originate in California, so our graph is picking up on those fires.

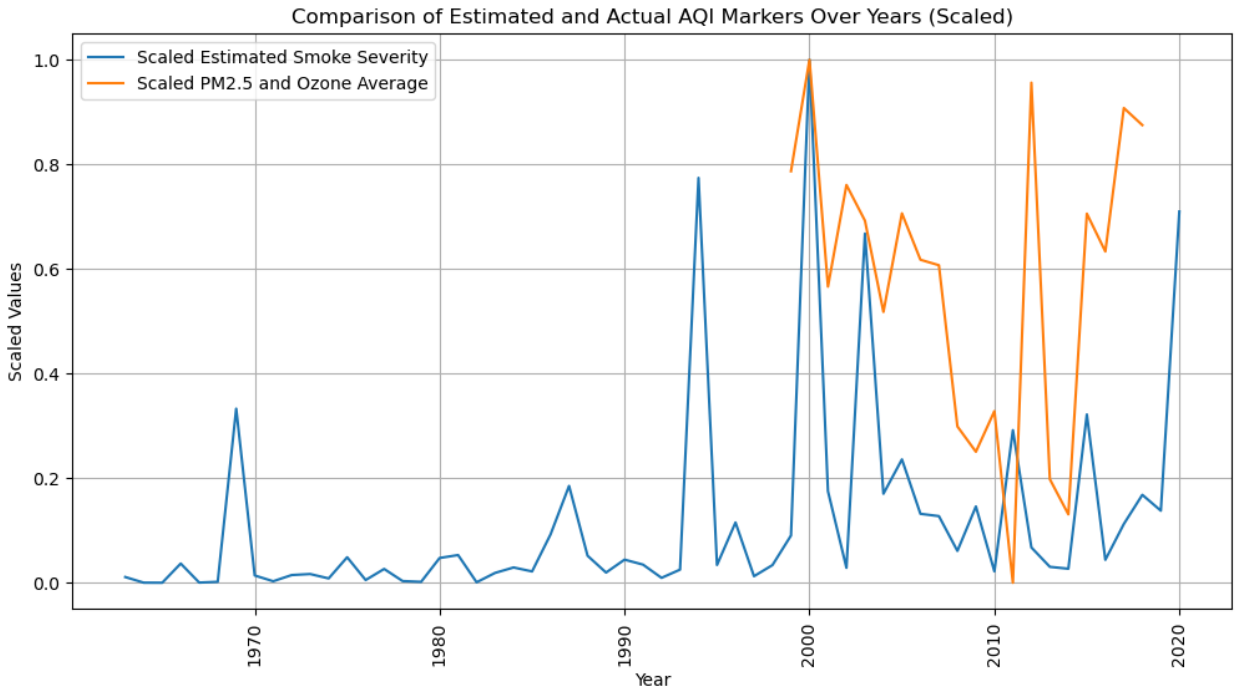
[1] https://www.sciencebase.gov/catalog/item/61aa537dd34eb622f699df81



Visualization Explanation 2: Total Acres Burned Per Year (within 1250 miles of Larimer County)

This visualization represents the total number of acres that were burned because of wildfires within 1250 miles of Larimer County. The x-axis is counted by years, while the y-axis represents the total number of acres burned. On the line plot, each dot represents a different year. The data came from the USGS[2], where their people have combined all known fire data into one large dataset. This data was cleaned to only count wildfires if: it was from 1963 - 2023, if it was during fire season, if we are sure (or almost sure) it was a wildfire, and if it was within 1250 miles of our city. In the dataset, there is a variable that represents the total number of acres burned by said fire, so the calculations were not complex. As we can see, this graph is very sporadic. Some years a lot of acres are burned because of wildfires, while other years are somewhat tame.

[2] https://www.sciencebase.gov/catalog/item/61aa537dd34eb622f699df81



Visualization Explanation 3: Comparison of Estimated and Actual AQI Markers Over Years

This visualization compared my wildfire severity index to the average AQI index of Larimer, only representing wildfires within 1250 miles of Larimer County. The x-axis is counted by years, while the y-axis represents the scaled values. I created a variable called the wildfire severity index, which (theoretically) represents how severe the wildfire season was for Larimer County. This was calculated by taking the total miles burned by said fire divided by the square root of the distance from Larimer County. This value was summed up for every fire in a given year. I collected the AQI values from the EPA's AQI API[3]. I grabbed PM2.5, PM10, and CO values from the closest weather monitoring station. I then got the average of these values, resulting in the average AQI value per year. Scaling the two values makes them easier to compare. As we can see there are times when the lines meet up and follow similar patterns (for example in 2000), yet in other times (like 2011) the lines do opposite things. The correlation coefficient was recorded as: 0.38, not very strong.

[3] https://drive.google.com/file/d/1bxl9qrb_52RocKNGfbZ5znHVqFDMkUzf/view

Collaborative Reflection

I want to start off with this reflection with this statement: I did not share or receive any code from a member of the class. Frankly, I wish I did, it probably would have made my analysis not take as long, yet I started on this assignment too early, so I was not willing to go back and change my code for no reason. If I could go back, I would have asked fellow students how to efficiently pull the wildfire data from the geoJSON file, it took me way too long. My collaboration consisted of me telling fellow students how long it took me to make the visualizations and pull the data. With that being said, let's move on to what I learned:

One of keep forgetting and re-learning when participating in research is how hard it is to find quality data. I have participated in research papers dealing with Wildfires before, and most, if not all of that work dealt with finding and cleaning massive, messy data. This analysis was almost the opposite, the data was pre-cleaned, and pre-processed, and we were even given a Reader.py file to load the JSON files. This allows us students to complete this large analysis in under a week's time when in reality, it could take over a month without the cleaned data. This project reminded me of the appreciation I have for quality data like this.

Another thing I learned throughout this common analysis was how much domain knowledge one needs to have to work on data like this. Someone could have all the computer science knowledge in the world, but it wouldn't matter when dealing with AQI values. A good researcher, and data scientist, must have sufficient domain knowledge to both understand and correctly analyse the given data. Or else a lot of mistakes could occur. For example, I did not know all the types of particulate matter or chemicals wildfires produce. I knew that wildfires produce some sort of PM, and CO, but not how much. If I had not properly read about these AQI measurements, it's possible I could have completely messed up my analysis by introducing measurements not related to wildfire smoke.