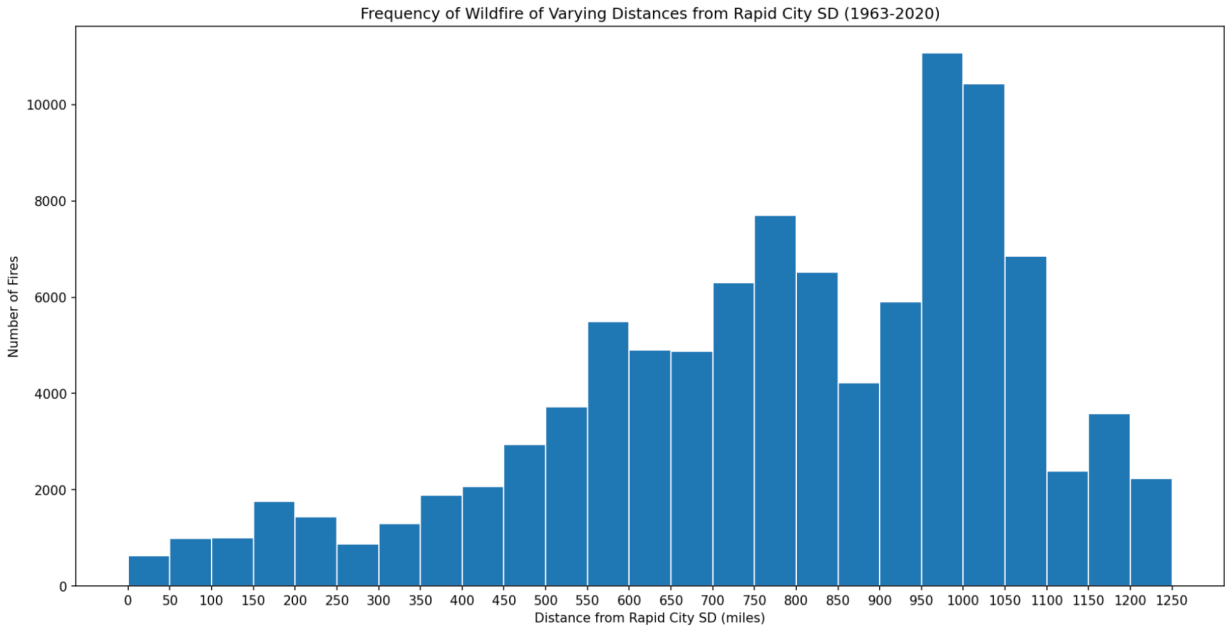
**Plot 1: Produce a histogram showing the number of fires occurring every 50 mile distance from Rapid City, SD to the max specified distance (1250 miles).**



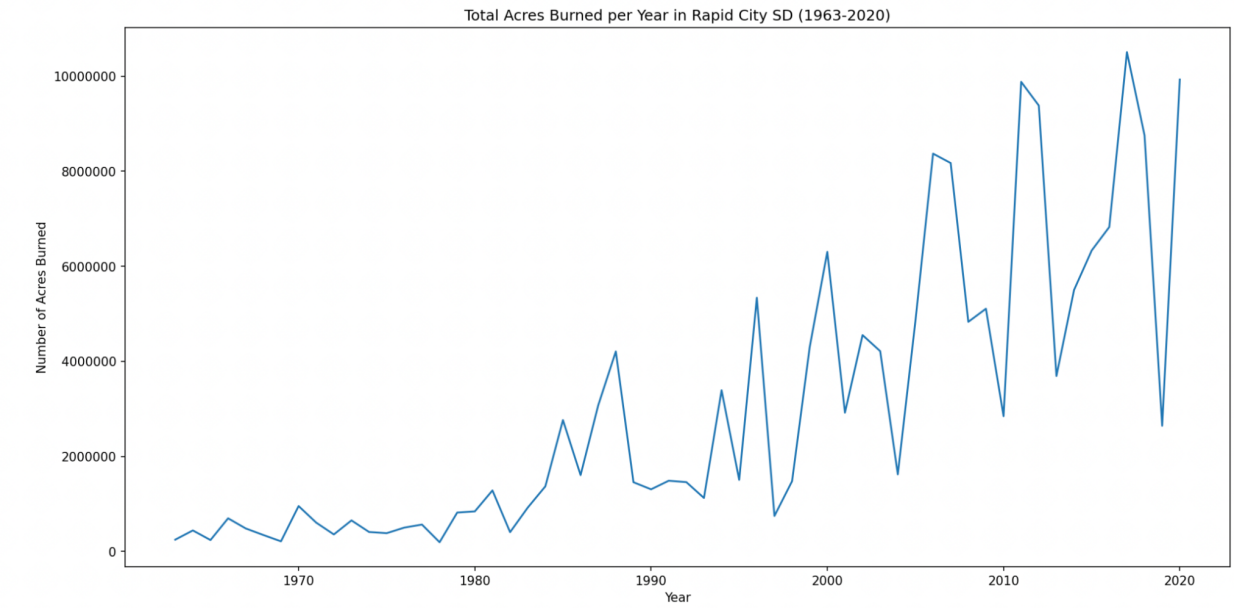

In the above figure, we are looking at the distribution of wildfire occurrences within certain distances from Rapid City, South Dakota. In other words, the above plot is showing how many wildfires have occurred within 0-50 miles of Rapid City, as well as 50-100 miles, 100-150 miles, and all other buckets up to 1250 miles. The x-axis represents the previously mentioned 50 mile buckets, each marked with a tick for ease of understanding. The y-axis represents the number of fires that occurred within the specific range of distance.

Examining the plot further, we find a quite prevalent left skew in the distribution, indicating that there are far more fires that are farther away from Rapid City, than there are fires quite close to Rapid City. It appears that wildfires occur most frequently between 950-1050 miles from Rapid City. One interesting feature in this plot is the dip in wildfire frequency occurring 850-950 miles from Rapid City, especially in comparison to the number of fires before and after. This may indicate that there is some sort of infrastructural or

environmental boundary existing in that location, or simply that the data holds some random noise.

To create this figure, very little processing was required. We essentially group the data by 'Fire_Year', sum up the count of fires per year, and plot.

**Plot 2: Produce a time series graph of total acres burned per year for the fires occurring within 1250 miles of Rapid City.**



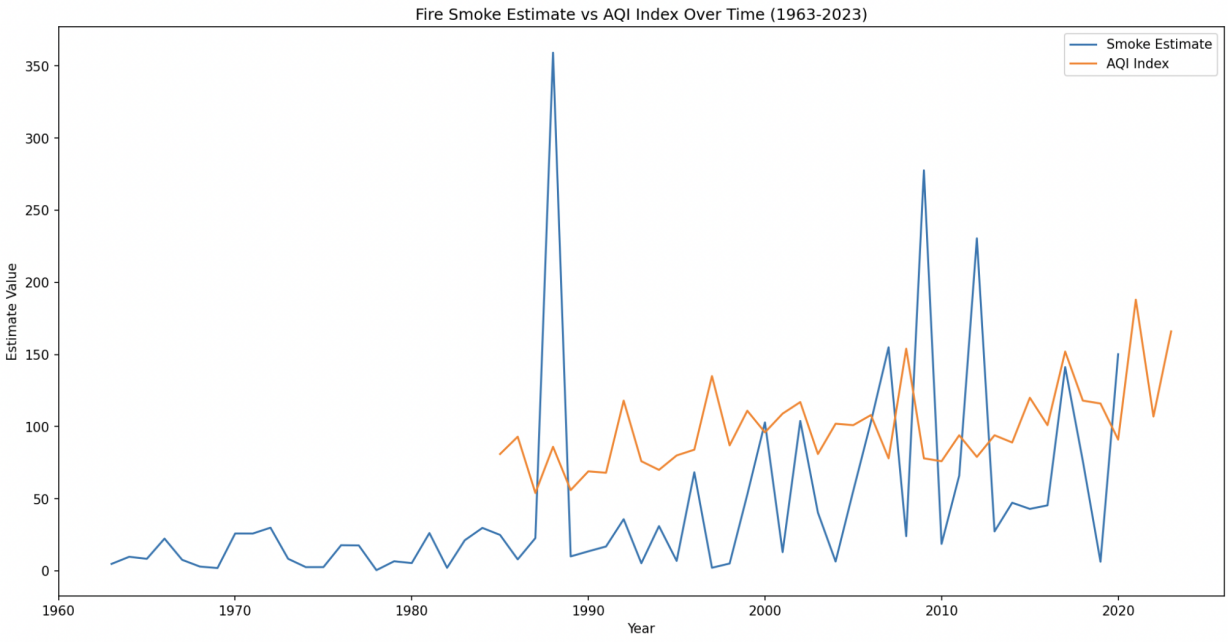

In this plot, we examine the total sum of acres burned by wildfires over the entirety of the data (from the year 1963 to 2020). This is a time-series format plot in which the x-axis represents the year and the y-axis shows the total number of acres that all the wildfires from that year had affected. We leave the y-axis in plain style (rather than abbreviating as 1e7, for example) to assist with ease of understanding and comparison.

What we find in the above plot is that the total number of acres burned has undoubtedly increased through the years. There are also some interesting oscillations with lengths of about ~2 years that become increasingly prevalent after the year ~1998, in which the total number of acres burned drops and rises wildly. I am curious to see how this plot compares

to simply the number of wildfires per (or the average size of wildfires over time), to see whether there might be some indication that the number of acres burned isn't *entirely* caused by an increase in wildfires. I am also curious as to how this trend compares to the increase in population over the years, and whether there might be any causation.

To create this plot, we simply summed up the acreage of all fires for a given year and plotted it against the time series 1963-2020.

**Plot 3: Produce a time series graph containing your fire smoke estimate for your city and the AQI estimate for your city.**




The above plot compares the fire smoke estimate we created in the notebook *smoke_estimate_viz.ipynb* to the AQI (Air Quality Estimate) created by the EPA. To create the fire smoke estimate, we of course consider that miles burned would play a large role in the ecological and health effects of the wildfire-- especially since it is by far the most common

way to measure a fire. But given that we are looking at wildfires up to 1250 miles away from Rapid City, South Dakota, we should definitely take into consideration the fact that even the fires that are burning thousands of acres would have little negative consequence on a city 1000+ miles away. As such, the smoke estimate metric used takes into account both the direct and inverse relationships that miles burned and distance have on wildfire severity, respectively, as the final smoke estimate to later be used in the predictive model.

As such, the resulting yearly averaged smoke estimate (in blue) holds the same variability and oscillation we were seeing in Plot 2 previously. And when plotted against the AQI estimate, we can see that the AQI fails to capture the same rise and fall, but does still show the same upwards movement. But, we should consider the whether a comparison is really appropriate between these two metrics— while the smoke estimate takes into consideration only the miles burned and the distance from Rapid City, the AQI estimate is an estimate of air quality that incorporates "ground-level ozone, particle pollution, carbon monoxide, sulfur dioxide, and nitrogen dioxide", which are not all directly related to wildfire (AQI Technical Assistance).

To create this visualization, we simply find the smoke estimate for each year by averaging those of all the individual wildfires, and then plot again the time series 1963-2023 (with EPA AQI data existing only from 1985 onwards).

Reflection on Collaboration

Via this data exploration, I learned much about not only wildfire prevalence, but also how they are measured and recorded. I have very little experience with GIS data, so it was interesting to see how to best manipulate it and use the ring coordinates to find each fire's distance from Rapid City, SD. Also, I previously knew very little about the AQI index, so I was excited by the opportunity to read the AQI Technical Assistance doc and see how exactly the index is calculated. Most importantly, I feel that the experience of doing the first part of this project was greatly enhanced by the introduction of a collaborative environment.

I think that the collaborative nature of this assignment most closely parallels the type of projects that me and my peers will be working on once we are employed upon graduation, so it's nice to take advantage of being able to bounce ideas off each other and solve disagreements as we would in the real world. For this project specifically, which was entirely new territory for me and several of my peers in terms of familiarity, domain, and data type, it was very useful to get different perspectives on the problem. For example, to gain a more personal perspective on the wildfire data, it helped that most of us in the program come from and have experienced living in varying parts of the country– this, in some sense, helped to understand and validate the results we were seeing. It also helps that we all have different levels of expertise and background on different parts of this assignment– while I am more familiar with data cleaning and manipulation, it was helpful that some of my peers had much experience requesting and loading large datasets such as those provided by USGS.

I did not directly consult or borrow code from anyone other than David McDonald, nor did I use any online resources. However, me and my peers did discuss the smoke estimate index in class, and agreed that we would need some combination of a positive relationship with acreage and an inverse relationship with distance from the city.