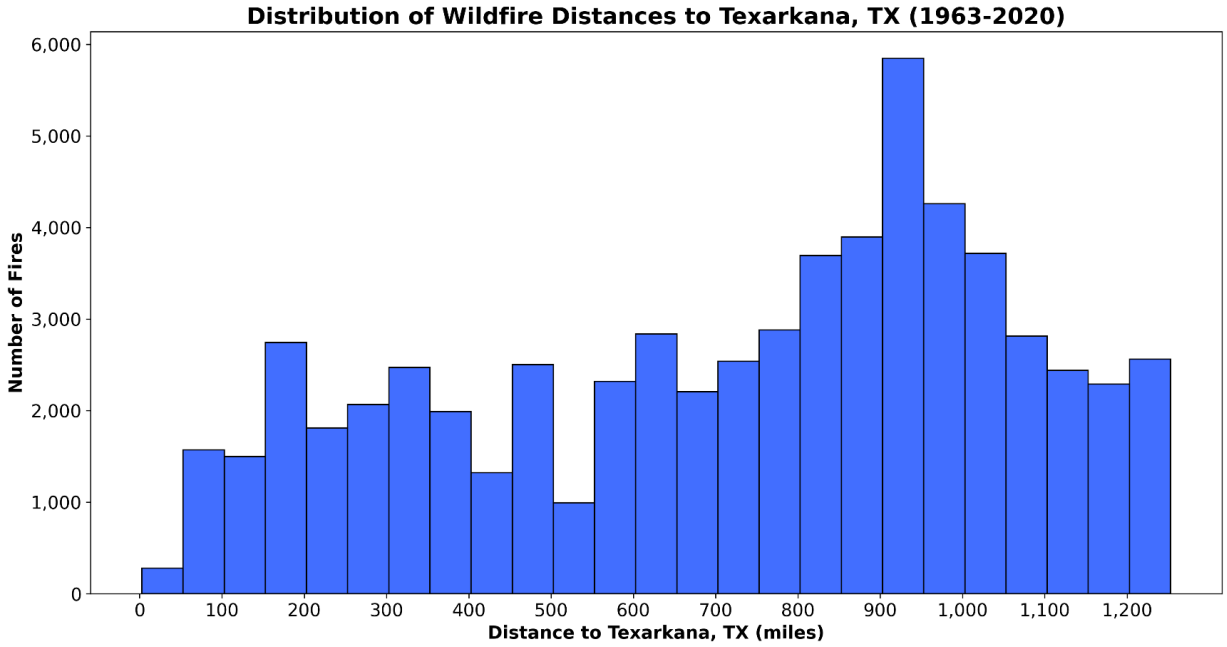
Part 1 - Common Analysis

DATA 512 | November 8, 2023 | John Michael

Visualizations

Figure 1



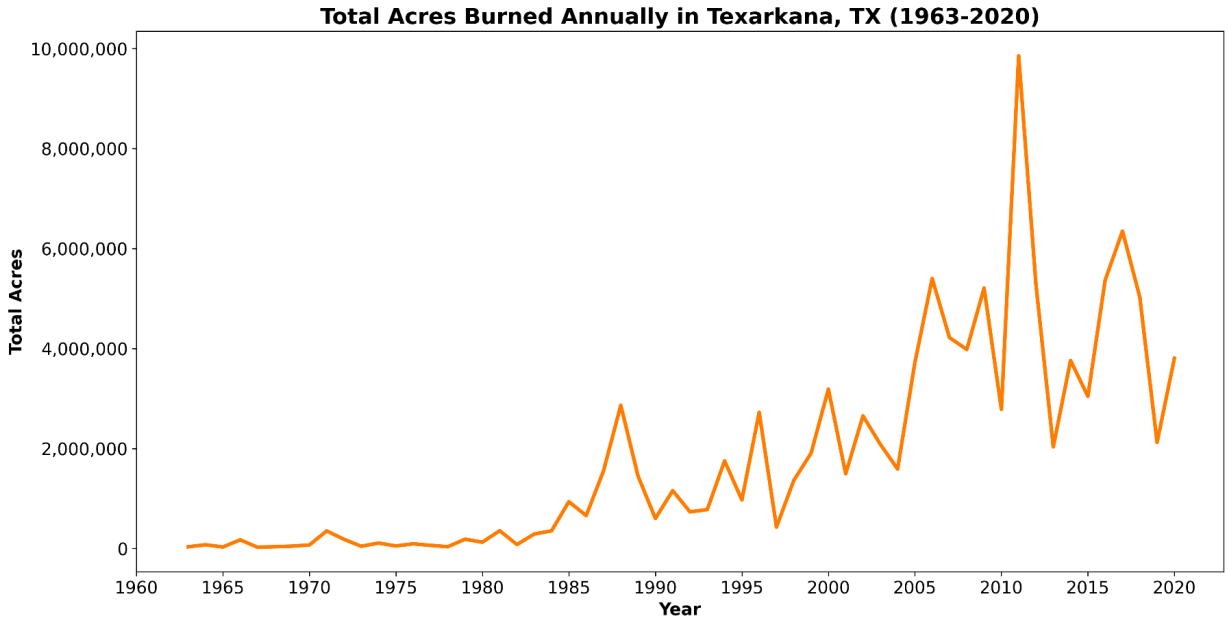
Source: USGS, “Combined wildland fire datasets for the United States and certain territories, 1800s-Present (combined wildland fire polygons)”, available at https://www.sciencebase.gov/catalog/item/61aa537dd34eb622f699df81 (last accessed November 8, 2023).

Figure 1 shows the distribution of distances of wildfires in the United States to the city of Texarkana, TX. The figure includes wildfires of up to 1,250 miles from Texarkana that occurred from 1963 to 2020. There were a total of XX fires that make up this figure. The y-axis shows the number of fires while the x-axis shows the distance to Texarkana in miles. The x-axis utilizes bins of 50 miles.

There are 135.061 wildfires in the USGS dataset. The dataset has an entry for every wildfire that occurred with each year. I calculated the distance between each wildfire and the center of Texarkana by computing the shortest distance from the perimeter of each wildfire to the city center. I then filtered the fires to include those with less than or equal to 1,250 miles of distance, as well as to those that occurred from 1963 onwards. These two filters resulted in a list of 63,528 wildfires that go into this figure.

The figure shows a roughly long-tailed distribution. The 50-mile bin with the highest frequency of wildfires is the 900-950 mile range. Most fires are in the 800 to 1,100 mile range.

Figure 2

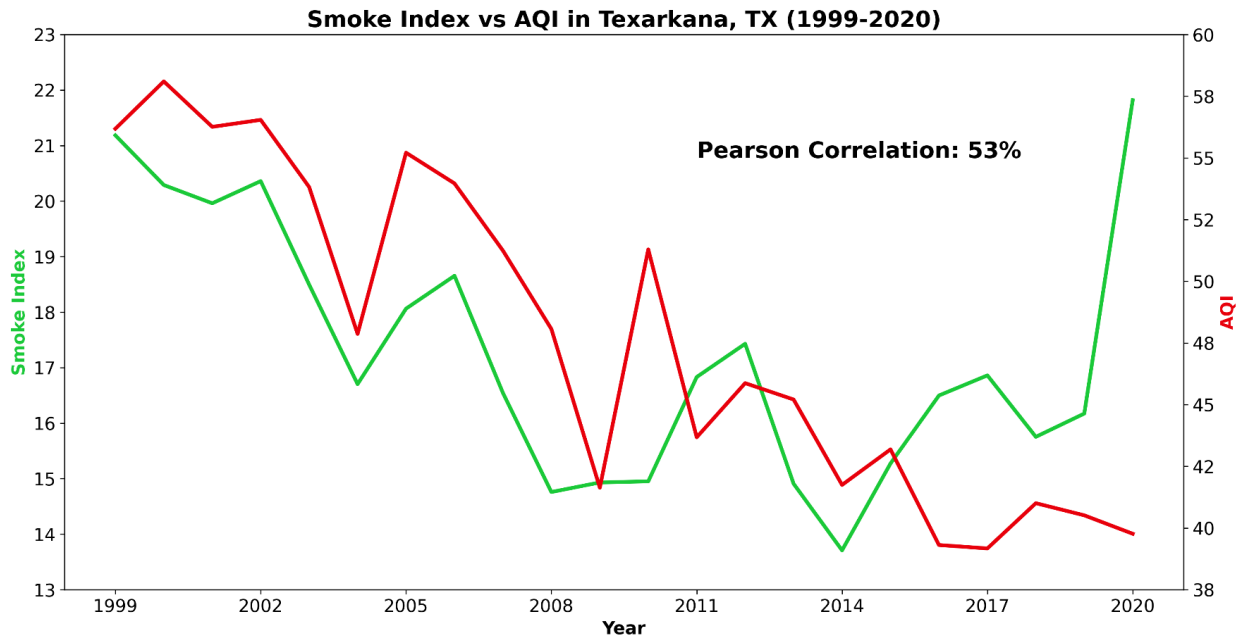


Source: USGS, "Combined wildland fire datasets for the United States and certain territories, 1800s-Present (combined wildland fire polygons)", available at https://www.sciencebase.gov/catalog/item/61aa537dd34eb622f699df81 (last accessed November 8, 2023).

Figure 2 shows the total acres burned annually by wildfires that are within 1,250 miles of Texarkana, TX. The data in this figure are the same as the data in Figure 1. The y-axis whos the total acres burned while the x-axis shows the years from 1963 to 2020. We can notice a cyclical pattern to the peaks and troughs of burned acres, with a slightly upward trend starting in the early 2000s. The spike in 2011 is likely due to the 2011 Texas wildfires[1].

[1] Rebecca Hennes, "A look back at the catastrophic 2011 Texas Tri-County Fire", Chron - Magnolia News, September 12, 2019, available at https://www.chron.com/neighborhood/magnolia/news/article/2011-Tri-County-Texas-wildfires-photos-anniversary-14435506.php (last accessed November 8, 2023).

Figure 3



Sources:
USGS, "Combined wildland fire datasets for the United States and certain territories, 1800s-Present (combined wildland fire polygons)", available at https://www.sciencebase.gov/catalog/item/61aa537dd34eb622f699df81 (last accessed November 8, 2023).
Environmental Protection Agency, "Air Quality System API", available at https://aqs.epa.gov/aqsweb/documents/data_api.html (last accessed November 8, 2023).

Figure 3 compares the smoke estimate index that I have calculated and the air quality index (AQI) obtained from the Air Quality Service (AQS) API. The left axis shows the smoke index, the right axis shows the AQI and the x-axis shows the years. I show only years 1999 to 2020 because these are the only years in which the AQI data for Texarkana and the wildfire data overlap.

The smoke index of each wildfire uses the formula outlined below:

$$SmokeIndex = Acres/Distance^2 + FireType^2$$

There are five unique categories of fire types. They are "Wildfire", "Likely Wildfire", "Prescribed Fire", "Unknown - Likely Wildfire" and "Unknown - Likely Prescribed Fire". I have assigned these categories the values of 5 to 1 respectively. The intuition behind the smoke formula is that we expect the size of fires and the fire type to be positively correlated with smoke impact while the distance of the fire to the city should have a negative correlation with smoke impact.

The wildfire data for this figure is the same as the ones used in Figure 1 and Figure 2. I collapsed the smoke index to an annual estimate by simply taking the average smoke index of all wildfires within each year.

I obtained the annual AQI estimates by first taking the maximum AQI for each station and each day of data. I then take the average AQI across all stations within a day. To align with the fire season of May 1 to October 31, I filter the data to only include those months in each year before averaging the AQI across the days in the fire season.

The Pearson correlation between the AQI and my smoke estimate is roughly 53%. The smoke estimates do not take into account factors such as weather and the length or time of occurrence of the wildfires. These are important factors that likely affect the smoke impact of the wildfires. The AQI can also be erratic depending on how the quality of monitoring has changed over time.

Reflection

One thing I learned from this exercise is the difficulty of estimating the impact of wildfires on air quality. From researching the topic of wildfires and discussing with classmates, I learned that there are so many factors we must take into account to fully capture the smoke impact. These factors go beyond features of the wildfires and include aspects such as weather conditions, wind patterns and the topography of cities and their surroundings.

From working with the AQS API, I also learned about the methods underlying air quality measurements and the challenges that come with them. Air quality monitoring can be erratic depending on location and how monitoring facilities have evolved over time. Working with air quality monitoring data effectively requires an understanding of the development of air quality monitoring over the past few decades.

I found the notebooks that Dr. David McDonald prepared very helpful in getting familiar with the wildfire data and the AQS API. My discussions with fellow classmates such as Yash Manne, Mark Qiao and Evan Yip also helped me refine my smoke impact estimation. I initially came up with a relatively simple formula that combines acres burned, distance to my city and the wildfire type. After discussion with classmates, I landed on a more sophisticated formula that involved using powers. This increased the correlation between my smoke estimates and the AQI from 11% to 53%.