DATA 512 Project Part 1 – Common Analysis: Visualizations and Reflection

Logan O’Brien

Visualizations:

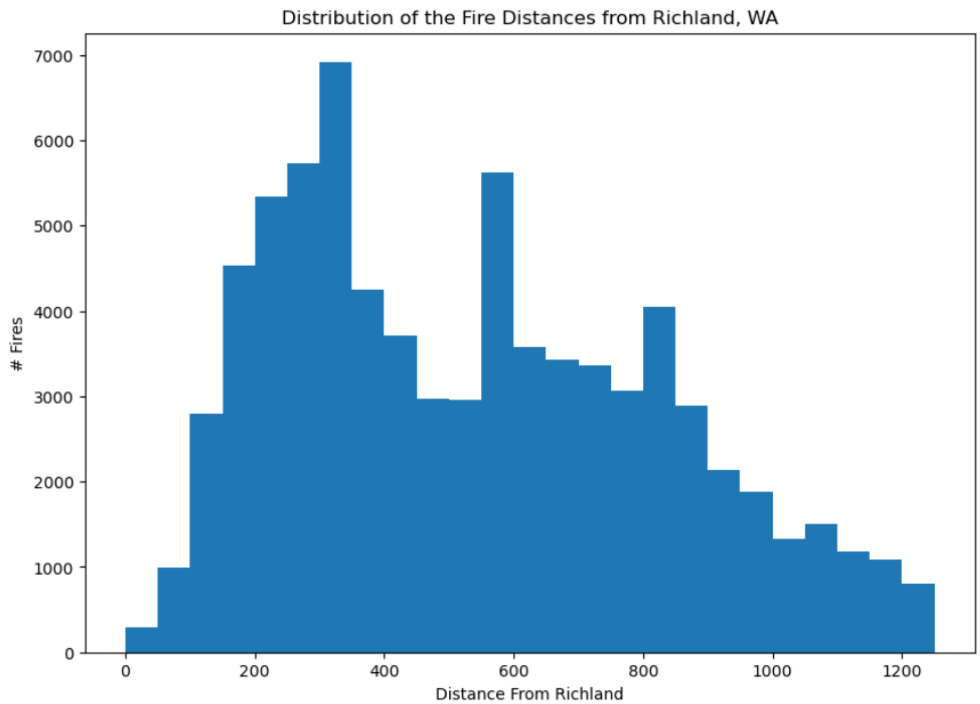

Figure 1

Figure 1 Explanation:

This histogram depicts the distribution of the wildfires within 1250 miles of Richland Washington. It shows that the majority of the fires occurred within 100 and 1000 miles from Richland, and that a significant portion of the fires were between 150 and 450 miles away. To read this chart you must understand the following: the y axis represents the number of fires in each “bucket”, while the x axis represents the different “buckets” of data. What do I mean by “bucket”? Well, each segment of the x-axis represents a group of fires that occurred in the corresponding distance from Richland. In this case, each “bucket” is a multiple of 50, so the first “bucket” contains all fires that occurred from 0 – 50 miles, the second 50 – 100 miles and so on, all the way up to 1250 miles.

The data behind this chart was processed by exporting it from sciencebase.gov and filtering it to fires that occurred between 1963 and 2023 and were within 1250 miles of Richland, WA. Then, the data was converted to a DataFrame for ease of use and the distance field was extracted and plotted.

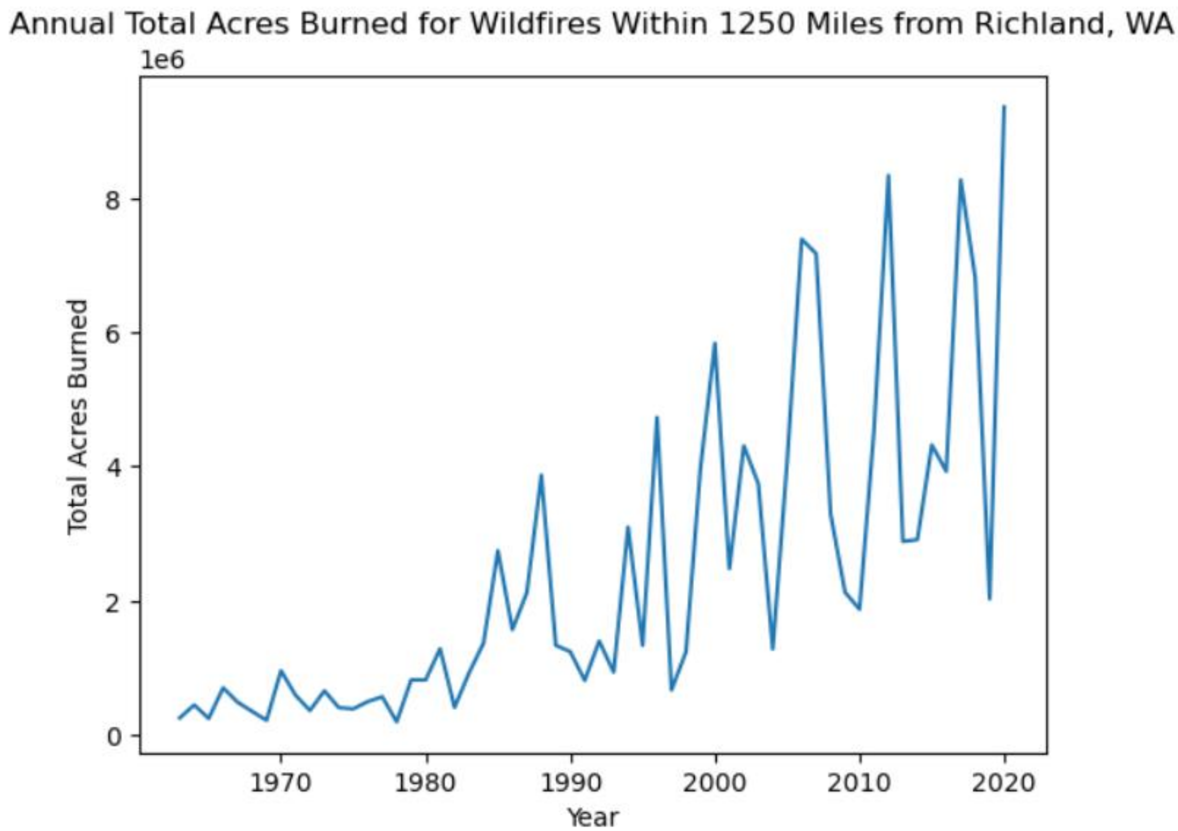

Figure 2

Figure 2 Explanation:

In this visualization, we can see the total acres burned from fires each year, within 1250 miles from Richland, WA. The y axis represents the number of acres burned and the x axis represents the years. So, by looking at a specific year on the x axis, we can then see where the line in the chart is above that spot and look to the y-axis to see how many acres burned that year. From this figure, we see that there is quite a bit of variability in the acreage burned, but that there is a significant general increasing trend.

To prep the data for this view, I took the wildfire data stored as a DataFrame and grouped it by year while adding up the acres field to get the annual total acres burned each year. Then I plotted a time series chart of the acreage burned vs time in years.

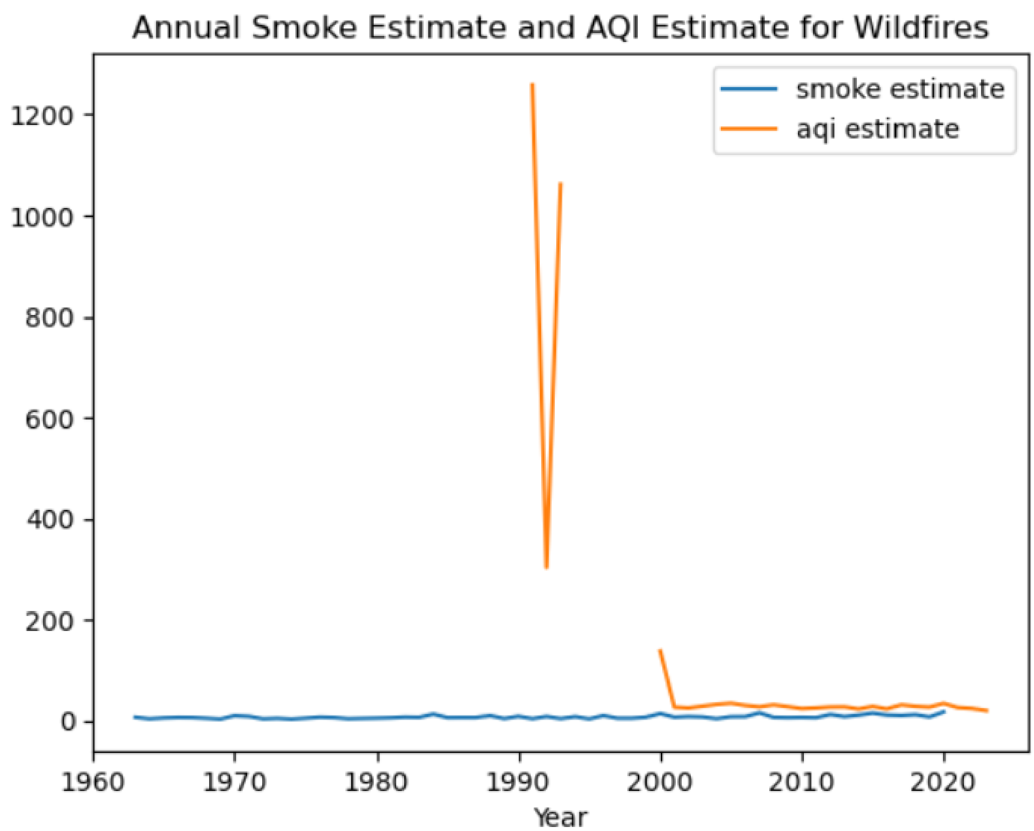

Figure 3

Figure 3 Explanation:

Lastly, my third figure depicts the annual smoke and AQI estimates for the wildfires. The x axis represents time, in years, and the y axis is the numeric value of the air quality estimate. To read this figure, you simply select a point in time on the x axis and draw a line vertically to the lines representing the air quality estimates – the higher that part of those lines, the higher the value of the estimates. In short, by looking at this chart, we see that the AQI estimate is not available until around 1990, and that the two estimates are wildly different at a few points, but in the same order of magnitude for the others.

To create this view, I created my own smoke estimate and applied it to every fire in the collection of relevant fires. Then, I included the smoke estimates into my DataFrame of wildfire info. I grouped that DataFrame by year and took the average of all smoke estimates for each year to get the annual smoke estimates. In contrast, I acquired annual AQI estimates from the EPA and through a series of steps arrived at AQI measurements for locations close to Richland, WA. The AQI estimates were initially daily measurements, so I had to group them by year and average the estimates in each year to get the annual estimates. Then, I simply plotted the smoke and AQI estimates.

Reflection:

For this part of the project, we were given a research question to answer, namely "What are the estimated smoke impacts on your assigned city for the last 60 years?" Through my analysis, I found that the total acreage burned each year from wildfires within 1250 miles of Richland Washington varies a lot from year to year, but there is a definite upward trend (Figure 2). Additionally, as seen in Figure 3, the air quality (based on my smoke estimate) does not appear to be increasing – at least not very much. So from this rudimentary analysis of the data, it appears that wildfires might be getting slightly worse over the years.

To be honest, the possibility of collaboration did not impact my approach for working through the assignment all that much. The way I understood the assignment, we had permission to collaborate according to limitations specified, however, collaboration was not a requirement. Thus, I tried to do the assignment on my own, in part to avoid any risks of collaborating or attributing someone's help in a way that was against the assignment policy. However, I do see the benefit of discussing possible approaches with my peers, as this assignment was very open – ended and a lot of times there was not a right answer. For example, Dr. McDonald made this very point when he talked after class with me and several other students about how we should design our predictive model and what the teaching staff was looking for. I appreciated the challenge of navigating the ambiguity of this assignment because it makes it closer to what the real world is like.

While I sought to not rely on the assistance of others, there were a couple cases where I would like to credit my peers with ideas that I implemented in my code.

1. I had a discussion with Sagnik regarding the processing times for reading in the original file and then filtering it to the desired results. As I remember it, I did not change my approach based on what he told me, but our discussion did give me confidence in knowing what to expect. He also told me about a neat tool in Python, tqdm(), that allows you to track the system's progress when iterating on an iterable. I put this to good use multiple times in the project.
2. Emily Rolen suggested I select the max value when comparing multiple AQI values from different pollutant types – which I incorporated into my approach for retrieving the AQI estimates.
3. I spoke with several other students about various aspects of the project, which helped to grow my confidence in my approach, but I don't remember the discussion changing my solution approach significantly – so I don't believe they need to be credited.

References:

- Visuals come from my Jupyter Notebook 'Project Part 1 - Common Analysis-s3-Comparison-and-Visualization.ipynb'