

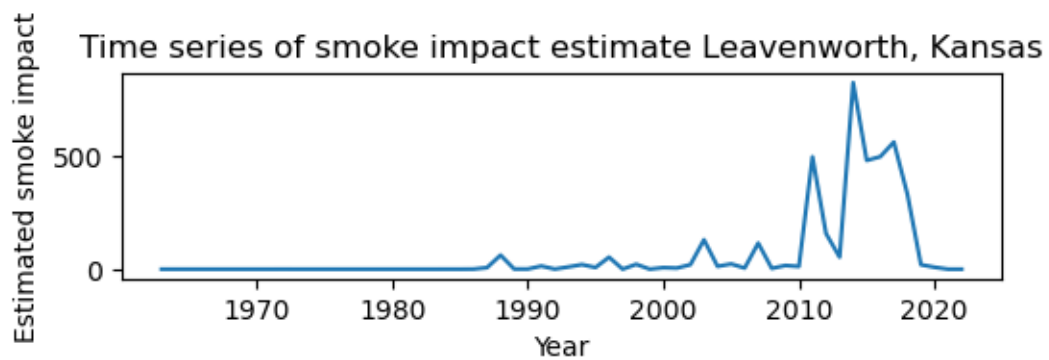
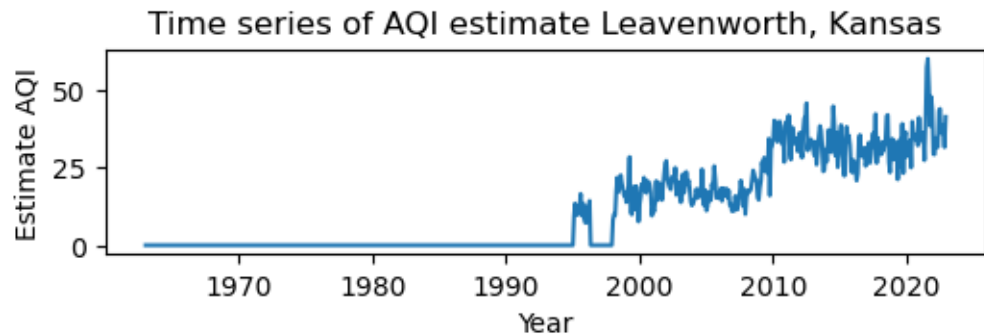
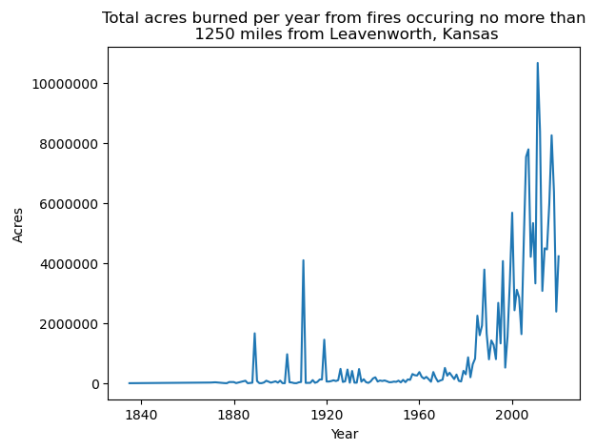
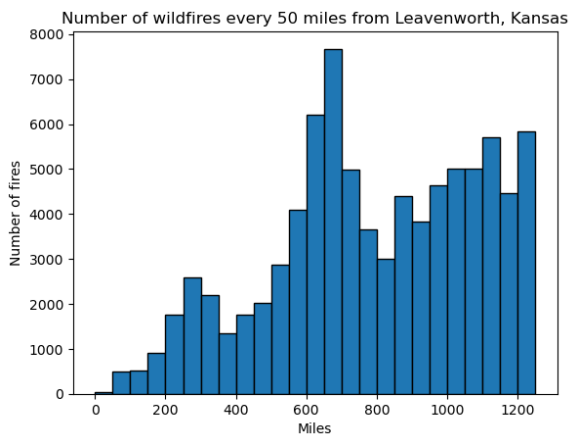
Zachary Bowyer

Data 512

11/5/2023

Part 1 Reflection

### Visualizations:



### **Visualization 1:**

The first visualization shown above represents how many fires have occurred in each 50-mile interval from Leavenworth, Kansas up to 1250 miles. The main takeaway here is that most wildfires occur greater than 600 miles away. This is an interesting result for our analysis because it can be used to suggest that our AQI and smoke estimates should be quite low, as most fires occurred very far away, which means smoke should be heavily dispersed before it reaches Leavenworth. The viewer can read this visualization by counting each bar as a segment of 50 miles and then interpreting the height of the bar as the number of fires that occurred for that mile segment. The X axis represents the various 50-mile range increments, and the Y axis represents the number of fires that occurred. There are no special units to be concerned about. This specific visualization was built using the GeoJSON data. In particular, geodetic distance calculations had to be performed to get the distance between each point in a fire polygon and Leavenworth, Kansas. The final distance between a wildfire and Leavenworth is the average distance between all points in the fire polygon to Leavenworth. Specifically, the “rings” attribute of the GeoJSON data was used.

### **Visualization 2:**

This visualization represents the total number of acres burned per year from wildfires occurring no more than 1250 miles away from Leavenworth, Kansas. In this case, the “GIS\_Acres” field from the GeoJSON data was used and summed annually. To be specific, GIS acres are calculated using mapping software, and it can be assumed that areas from the past were calculated with current technology rather than stored as a different format historically. The viewer can read this figure by looking at where the line is during a specific year, and from there determining how many acres burned that year. The X axis represents the year being considered. The Y axis represents the number of acres burned by wildfires within 1250 miles of Leavenworth, Kansas. One thing to note is that there is little to no data before 1860.

### **Visualization 3:**

Comprised of two graphs to avoid scaling issues, this visualization represents a comparison between our developed smoke estimate and estimated AQI for Leavenworth, Kansas. Overall, one can see that there is an increasing trend in AQI estimates while the smoke estimate varies. One might argue there is some similarity between the two graphs between 2010 – 2018. However, that does not prove any explicit relationship exists or even that the estimates are good. With that being said, these visualizations used both GeoJSON data and the AQS API. Specifically, the first graph used the AQS API, and the second graph used the “Shape\_Area” and “rings” attributes of the GeoJSON data. The viewer can read these figures by looking at where the line is during a specific year, which represents the estimate metric for that year/month. One important distinction is that estimate AQI data is monthly, while estimated smoke impact data is annual. The X axis represents the year being considered while the Y axis represents the value of the estimated metric. One thing to note is that data was sparse or nonexistent before 1980 for both graphs.

### **Thoughts on collaborative activities and other reflections:**

In this work, while there was an option to collaborate with other students, which gave students the option to share code snippets, statistical approaches, and visualizations techniques, I did not personally participate in such collaborative activities. However, I did use code examples provided by Dr. David W. McDonald in multiple places. For example, I used Dr. McDonald's code in the following ways: Using a GeoJSON reader module to read in data, converting esri:102008 coordinates to epsg:4326 coordinates, calculating the average distance from a fire polygon to a set of target coordinates, and using the AQS API [1] to access and operate on historical AQI data. When it comes to whether or not the possibility of collaboration helped, hindered, or changed my perspective about the problem, I can confidently say that my perspective did not change, as I usually opt to do projects by myself in academic settings. However, if the use of Dr. McDonald's code falls under the distinction of "collaborative", then I will say that the collaboration I did partake in was immensely helpful in saving me time.

In terms of what I've learned after completing this assignment and attempting to answer the research question, there are two main things. First, I realize that working with real world data is incredibly difficult for a large variety of reasons. In this work there are many examples of how I struggled with the datasets and had to end up making bad estimates to accommodate for this. Some examples include: Having to write code in multiple places that excluded fires that had no associated polygons, averaging only particulate AQI estimates because there was no way of knowing the relevance level of all AQI measurements to our smoke estimate, having to include stations that were not in Leavenworth county in my AQI estimates as the stations in Leavenworth county did not track particulates, and having to pick the largest AQI estimate for all of the stations as my final monthly estimate as some stations had no data for certain time-periods. Second, I have learned that making time-series projections requires knowing information about models called 'ARMAX', 'ARIMAX', 'ARMA', 'AR' and 'MA'. Something valuable I learned is that the 'ARMAX' model combines both linear regression and single variable forecasting to make a forecast based on the trends of both the predictors and the target variable. However, with that being said there are still many things about time forecasting I do not know, but I am happy that I was exposed to some of the methods used.

Overall, I believe the common analysis section of the final project went decently well. In terms of implementation of the AQS API as well as estimating smoke, I believe that I was able to generally follow the guidelines. In terms of difficulties encountered, a few things to note is that I found the AQS API a bit difficult at first to follow, my smoke estimate has many documented issues, my predictive models are not good, and some parts of my code take a very long time to run. With this work completed, I believe that I should be equipped to work on the next part of the project.

### **References:**

1. [https://aqs.epa.gov/aqsweb/documents/data\\_api.html](https://aqs.epa.gov/aqsweb/documents/data_api.html)