Zachary Bowyer

Data 512

11/5/2023

Part 1 Reflection

**Visualization 1:**

asd

**Visualization 2:**

asd

**Visualization 3:**

asd

**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 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, then I will say that the collaboration I did partake in was immensely helpful.

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 very 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>