Analysis of air quality index and its impacts on public health

Muhammad Abdullah (BSCS20015)

Session: Spring 2023 *Supervisor: Dr. Adnan Siddique* *Course: Spatial Data Science*

1. Introduction

The idea of project is analyze the weekly trend of AQI values in the state of California, USA and its impacts on the number of cases of Influenza-like-illness diseases.

2. Problem Statement

- Which counties of California have to face the poorest air quality.
- Visualizing the weeks in which California has to face the poorest air quality.
- Identify the relationship between poor air quality and incidence of ILI cases in 5 major regions of California.

3. Data

USA Counties Shapefile: A shapefile of all USA Counties.

AQI Dataset: Daily AQI values of USA Counties from 2017 to 2020.

Influenza-like-illness Cases Data: Number of positive ILI cases in 5 major regions of California. This dataset consists of weekly data from 2017 to 2020.

Preprocessing: The preprocessing of the dataset consisted of:

Converting the Daily AQI data into Weekly AQI data by taking the max AQI value of each week. I accomplished this by dividing all the days of a year into 52 weeks and based on available daily data, assigning a collection of days to a week. The day with the highest AQI Value in a given week is taken for the WEEKLY_MAX_AQI column. The dataset is then categorized into 5 regions, the same regions as in the ILI dataset. This dataset is then merged with ILI dataset based on the dates and region names. I then removed all duplicate values and forward filled the missed AQI Values. Lastly, I added the geometries of each county of California from the USA shapefile.

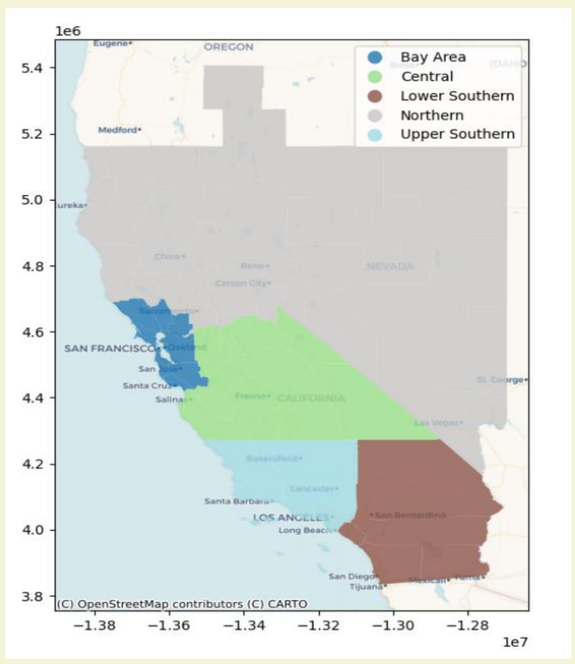

Fig. 1: Plot of regions of California

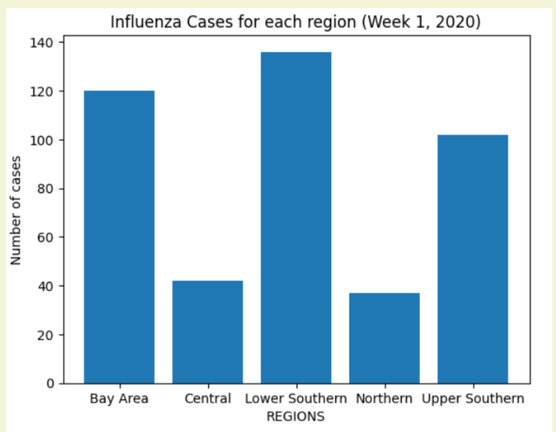

Fig. 2 Bar Plot of Influenza Cases of each region of week 1,2020

4. Methodology

- Used adaptive kernel weights to estimate AQI values for missing counties in California.
- Spatial autocorrelation analysis was conducted using Moran I to identify potential positive autocorrelation.
- To align weekly AQI values with available ILI cases data, a low-to-high resolution conversion was performed by averaging weekly AQI values for all counties in a given region.

5. Results & Discussion

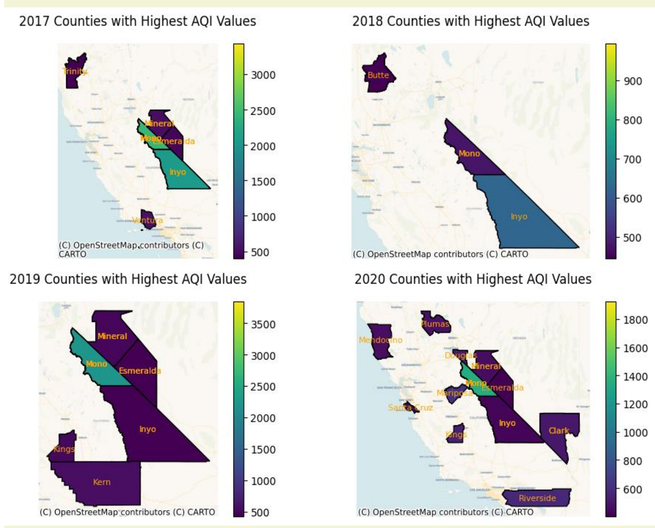

Fig. 3: Counties with Hazardous Air Quality

Counties with Hazardous Air Quality:
Here we can see the counties that have seen consistently seen poor air quality.
In 2017, there were 6 counties that faced hazardous air quality and this number increased to 12 in 2020. Noticeably, the county “Mono” can be seen as the one with the worst air quality from 2017 to 2020.

Spatial Autocorrelation:
Here we can clearly see positive spatial autocorrelation. The Moran I’s constant comes to be 0.71 for this particular week. Counties with poor air quality are together. This autocorrelation may vary week to week.

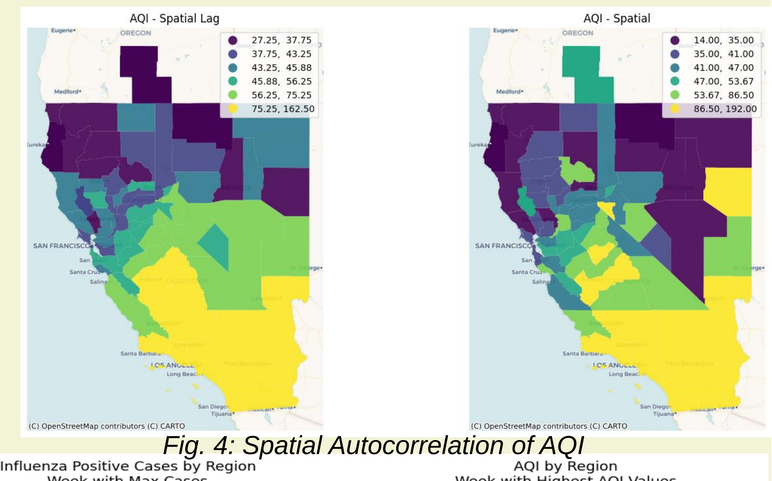

Fig. 4: Spatial Autocorrelation of AQI

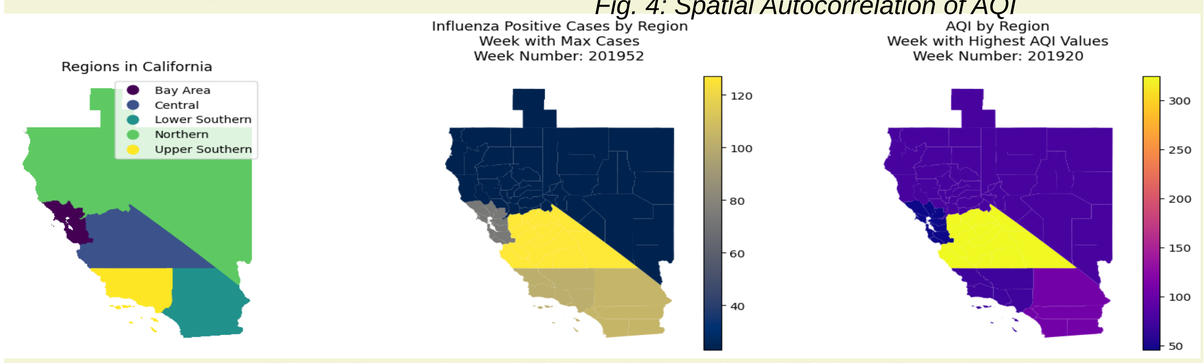

Fig. 5: Comparison of Highest AQI week and Highest Cases Week

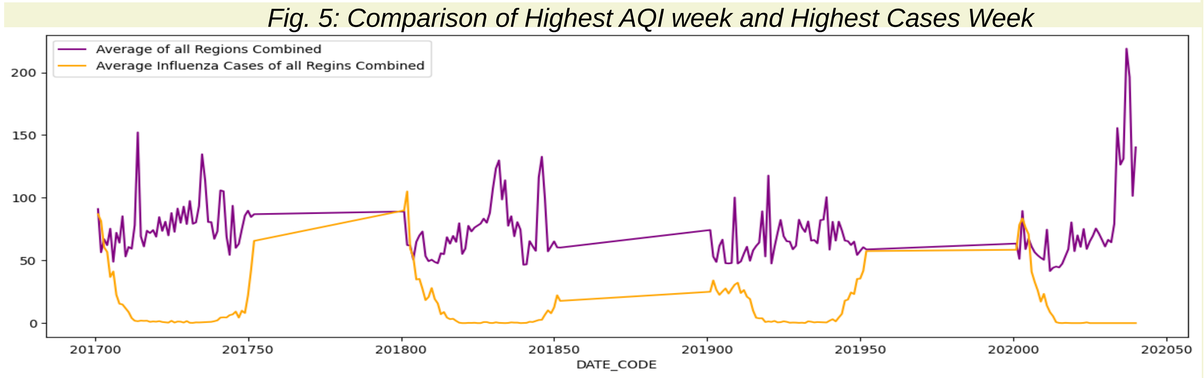

Fig. 6: Trend of average AQI and ILI Cases in California from 2017 to 2020

Conclusion:
The highest mean AQI week did not have the highest count of ILI cases (Figure 5), as confirmed by trends in Figure 6. No correlation was found between Influenza Cases and AQI data, possibly due to limitations of influenza cases data being regionally-based and high resolution, while AQI data was only available at the county level.

Acknowledgments: Please add acknowledgements here.

Remote Sensing & Spatial Analytics

INFORMATION TECHNOLOGY UNIVERSITY