COVID-19 Data Analysis

Exploring the Impact of the Spread of COVID-19 on Air Quality

Jingyi Lu (rebeccalu99@ucla.edu), Kaixin Wang (kaixinwang@ucla.edu)

UCLA DataFest 2020 – Team R++

Introduction

The level of PM2.5 in the air is associated with the everyday transportation as well as industrial production. The worldwide outbeak of the COVID-19, a large portion of the population is under self-quarantine, which has led to a decreased amount of daily transportation and industrial production.

The goal of this project is to explore if the spread of the COVID-19 has an impact on the level of PM2.5 in the air. To do this, we collected the data of COVID-19 (01/2020-05/2020) and the level of PM2.5 in the air (01/2019-05/2020), selected three cities from China (Beijing, Shanghai, and Wuhan) and three cities from the U.S. (Los Angeles, New York, and Santa Clara) as the sample, and conducted statistical analysis.

Methods

Since the data we have are time-series data, and because we want to see if the spread of the pandemic has an influence on the level of PM2.5 in the air, we decided to conduct the Granger-Causality Test. A time-series X is said to Granger-cause Y if it can be shown, usually through a series of t-tests and F-tests on lagged values of X, that those X values provide statistically significant information about future values of Y. Our hypothesis and methodology are shown in Figure 1.

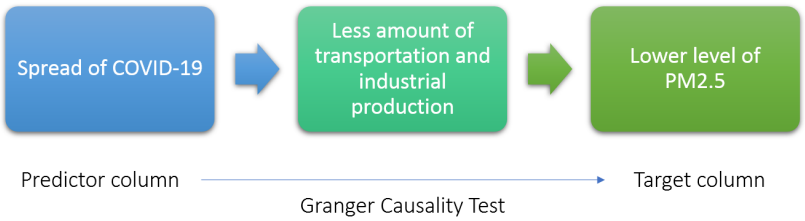

Figure 1: Hypothesis and Methodology

Results

Via conducting the Granger-Causality test on all six cities, we obtained the following results, as shown in Table 1:

Table 1: Number of Lags and P-values from Granger-Causality Tests

Statistics	Beijing	Shanghai	Wuhan	Los Angeles	New York	Santa Clara
Number of Lags	5	2	3	6	3	2
ssr based F test	p = 0.1163	p = 0.1853	p = 0.0102	p = 0.2993	p = 0.0042	p = 0.0415
ssr based χ^2 test	p = 0.0668	p = 0.1640	p = 0.0049	p = 0.1775	p = 0.0008	p = 0.0312
likelihood ratio test	p = 0.0812	p = 0.1694	p = 0.0071	p = 0.2072	p = 0.0019	p = 0.0350
parameter F test	p = 0.1163	p = 0.1853	p = 0.0102	p = 0.2993	p = 0.0042	p = 0.0415

Based on the p-values obtained from the Granger-Causality tests, we see the spread of the pandemic has a statistically significant impact on the air quality in Wuhan, New York, and Santa Clara.

Furture Steps

Although we have seen that the spread of the pandemic *Granger-causes* the change in the level of PM2.5 in the air in some of the cities we sampled, we know the pandemic actually implicitly influences air quality through affecting the amount of transporation and industrial production. The next step is to collect relevant data regarding the flux of transporation (e.g., number of flights, trains, or buses per day) and data that reflects the intensity of industrial production (e.g., certain indicies in industry), and delve into further investigation.

Datasets

The datasets that we used in the project come from three main resources, as shown in Table 2:

Table 2: Sources of the Datasets

Dataset	URL
Statistics for COVID-19	https://github.com/CSSEGISandData/COVID19/tree/master/csse_covid_19_data/csse_covid_19_time_series
PM2.5 Level in cities in the U.S.	https://www.epa.gov/outdoor-air-quality-data/download-daily-data
PM2.5 Level in cities in China	https://quotsoft.net/air/