IS415 GEOSPATIAL ANALYTICS PROJECT | RESEARCH POSTER

GEOVID: A GEO-VISUALIZATION TOOL FOR SPATIAL ANALYSIS & GIS OF CORONOVIRUS-2019 IN CHINA

Ivy HAU Jia Yi | Jasky ONG Qing Hao | Shermin TAN

1 PROBLEM & MOTIVATION

Coronavirus disease (COVID-19) is an infectious disease caused by a newly discovered coronavirus. Most people infected with the COVID-19 virus will experience mild to moderate respiratory illness and recover without requiring special treatment. The COVID-19 virus spreads primarily through droplets of saliva or discharge from the nose when an infected person coughs or sneezes. With such similar symptoms to a common "cold", this virus outbreak has spread to more than 200 countries and regions, affecting more than 60,000,000 people around the world. In addition, it has also swept out more than 1,400,000 deaths, as of November 2020.

Presently, there are no vaccines available globally to protect against this virus, and the best way to prevent and slow down transmission is to be well informed about the COVID-19 virus, the disease it causes and how it spreads. In due of that, it will be significant to analyse the spread of COVID-19, understand where, when, and how it occurred by studying the spatio-temporal patterns of confirmed, recovered and death cases.

In addition, by performing an in-depth analysis of the spatio-temporal patterns of the disease spread, it will greatly improve our understanding and deter the size of this threat.

2 KEY OBJECTIVES

Particularly, this project will focus on the following objectives:

1. Visualize amount/density of the cases by performing relevant exploratory data analysis
2. Conduct global and local spatial analysis to uncover spatial correlation patterns and its influencing factors
3. Create a web-based interface through R Shiny with relevant user inputs to display relevant data analysis

The application will comprise(s) of the following analysis requirements:

- Graphical/Geographical visualization framework that can display the amount/density of the COVID-19 cases overtime
- Map visualization framework that supports macro and micro views
- Customizable analysis based on-demand through user's input

PACKAGES USED

R | Leaflet | plotly | Shiny | tidyverse | ggplot2 | lubridate | dplyr

3 APPROACH

Data Collection

- Extracted from ***Harvard Dataverse***
- Aspatial Data
  - Confirmed, Recovered & Death Cases in China, based on provincial-cities level
  - Population Data
- Geospatial Data of China cities

Data Cleaning & Wrangling

- Extracting daily & cumulative count of cases from January to September 2020
- Remove duplicates & NA data

Exploratory Data Analysis

- Line graph
- Choropleth mapping
- Box map

Spatial Autocorrelation Analysis

- Local Moran's I
- Global Moran's I
- Localized Indicator of Spatial Association (LISA)

Report Findings & Conclusion

- Understanding of the virus spread geographically
- Identifying clusters

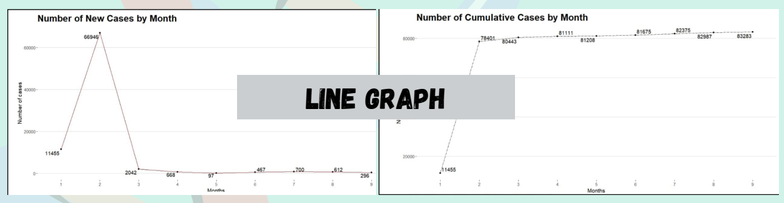
Geo-Visualization Tool

- R-Shiny Application: On-demand Graphical/Geographical Visualization tool
- Allowing users to analyze results based on their input

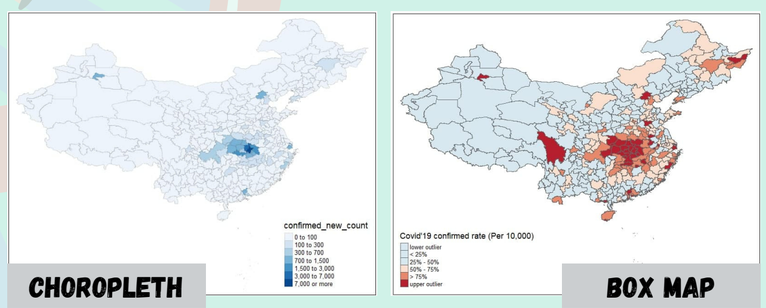
4 WHAT CAN THE APPLICATION ANALYZE?

EXPLORATORY DATA ANALYSIS

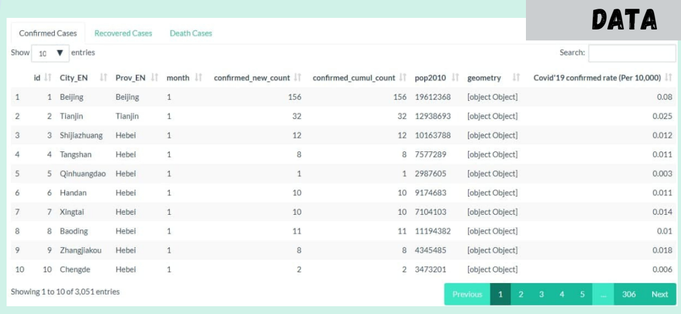
- Line Graph is used to analyze the daily and cumulative count of COVID-19 cases over the period of months from January to September 2020.



- Choropleth Mapping is used to view the geographical distribution of the spread at provincial-cities level, whereas Box Map is used to visualize the outliers using the interquartile range of the COVID-19 rate based on the total number of cases and the total number of population in each city.



- Interactive view of data in a table format where the user will be able to view the raw data that is used for the analysis.

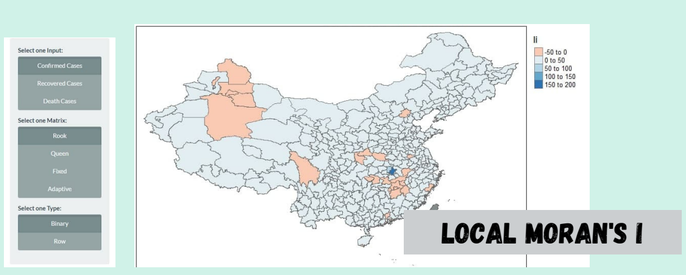


id	City_EN	Prov_EN	month	confirmed_new_count	confirmed_cumul_count	pop2010	geometry	Covid'19 confirmed rate (Per 10,000)
1	Beijing	Beijing	1	156	156	19612368	[object Object]	0.08
2	Tianjin	Tianjin	1	32	32	12938693	[object Object]	0.025
3	Shijiazhuang	Hebei	1	12	12	10163788	[object Object]	0.012
4	Tangshan	Hebei	1	8	8	7577289	[object Object]	0.011
5	Qinhuangdao	Hebei	1	1	1	2987605	[object Object]	0.003
6	Handan	Hebei	1	10	10	9174683	[object Object]	0.011
7	Xingtai	Hebei	1	10	10	7104103	[object Object]	0.014
8	Baoding	Hebei	1	11	11	11194382	[object Object]	0.01
9	Zhangjiakou	Hebei	1	8	8	4345485	[object Object]	0.018
10	Chengde	Hebei	1	2	2	3473201	[object Object]	0.006

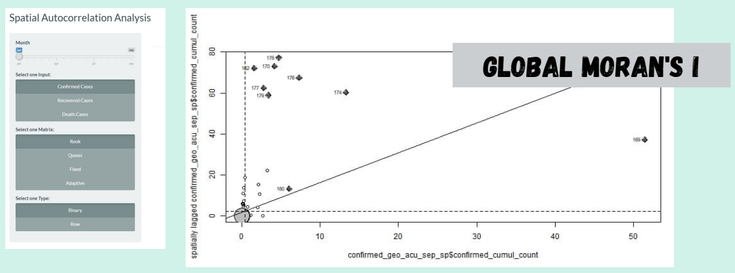
Showing 1 to 10 of 3,051 entries | Previous | 1 | 2 | 3 | 4 | 5 | ... | 306 | Next

SPATIAL AUTOCORRELATION ANALYSIS

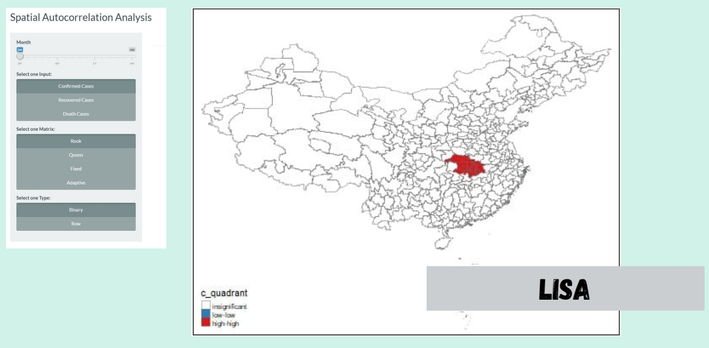
- Local Moran's I is used to compute the relationship between each region to its neighbour and identifies concentrations of high values, concentrations of low values, and spatial outliers.



- Global Moran's I measures spatial autocorrelation based on both feature locations and feature values simultaneously. Given a set of features and an associated attribute, it evaluates whether the pattern expressed is clustered, dispersed, or random.

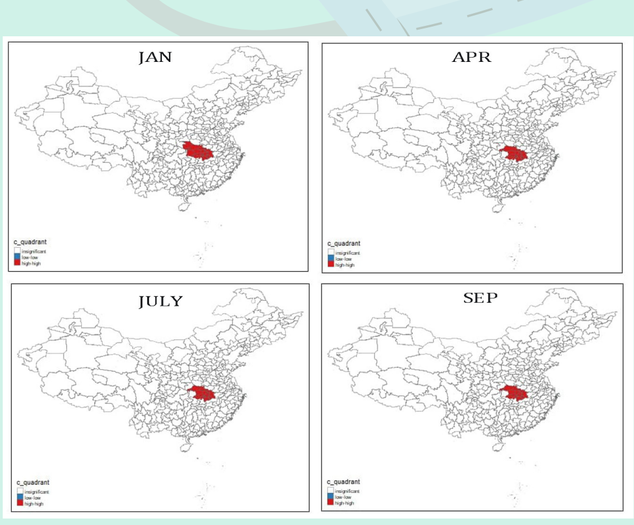


- Localized Indicators for Spatial Association (LISAs) are a class of local statistics that measure spatial association in sub-regions of the study region. They are also used to identify spatial outliers and local clusters.



5 INSIGHTS

In our initial review of the exploratory data analysis, it can be seen that this virus have affected many cities in China, which indirectly making us to assume that there could be many spatial clusters based on the general geographical view of the spread of the virus. However, after performing the spatial autocorrelation analysis, it can be seen that there is only one cluster which is the Wuhan.



Based on the LISA results above, we can see that from January to September 2020, most of the confirmed cases are clustered only in Wuhan, with the majority of the rise of cases in February. This indicates that necessary action should be taken on treating and detering the threat in the most seriously affected areas.

With regards to this analysis, there are limitations such as incomplete data resulting from human error with no inputs. This causes a weaker accuracy and missing areas when we plot the geographical distribution at provincial-cities level.

6 FUTURE WORK

If we can acquire more data, we will be able to perform a more in-depth analysis of the spread and suggest more valuable recommendations to deter the situation.
Data such as:

- Demand and supply of medical facilities in each areas
- Factors affecting the spread

Acquiring bigger scale data could allow forecasting on the potential areas that would be affected by the spread of Covid-19 cases.

CONCLUSION

As Wuhan was the main cluster affected by Covid-19 in our analysis, we can presumably expect this result in due of the China's mitigation strategy in performing a pandemic lockdown of Wuhan and increasing medical facilities immediately right at the outbreak area.

SMU SINGAPORE MANAGEMENT UNIVERSITY | School of Information Systems