

Geospatial-Temporal Analysis of the COVID-19 Pandemic in Malaysia

Group 1

Chew Zheng Xiong (A0167003R)

Lim Zhu An (A0224991X)

Shermaine Neo Xuan Min (A0172500W)

Presentation Outline

- ◎ Background
- ◎ Data Description
- ◎ Exploratory Data Analysis
- ◎ Exploratory Spatial Data Analysis
- ◎ Spatial Data Analysis
- ◎ Spatial-Temporal Analysis
- ◎ Discussion

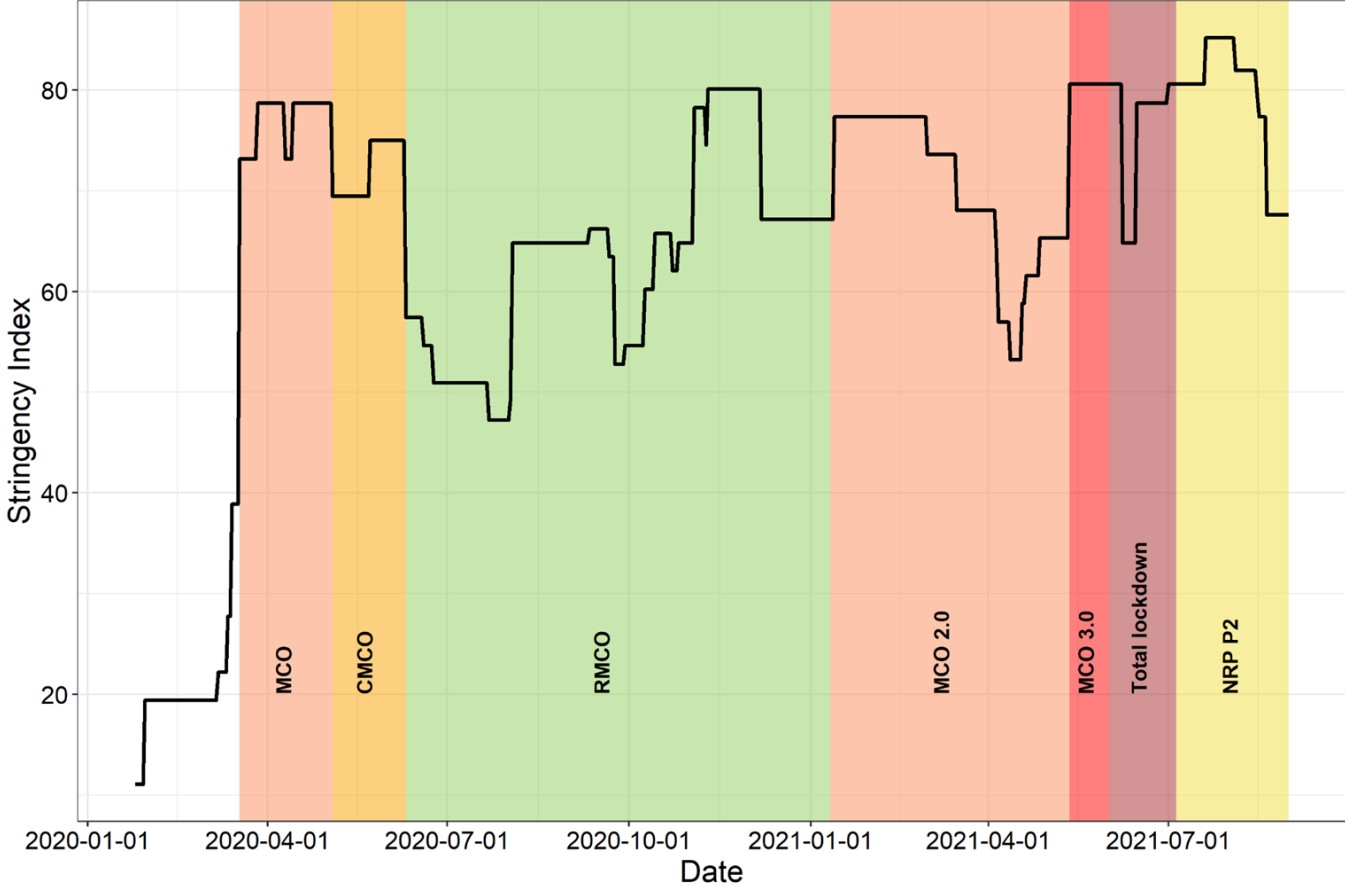
A decorative network diagram in the top-left corner, consisting of various sized nodes (some solid grey, some hollow white) connected by thin grey lines, forming a complex web structure.

1.

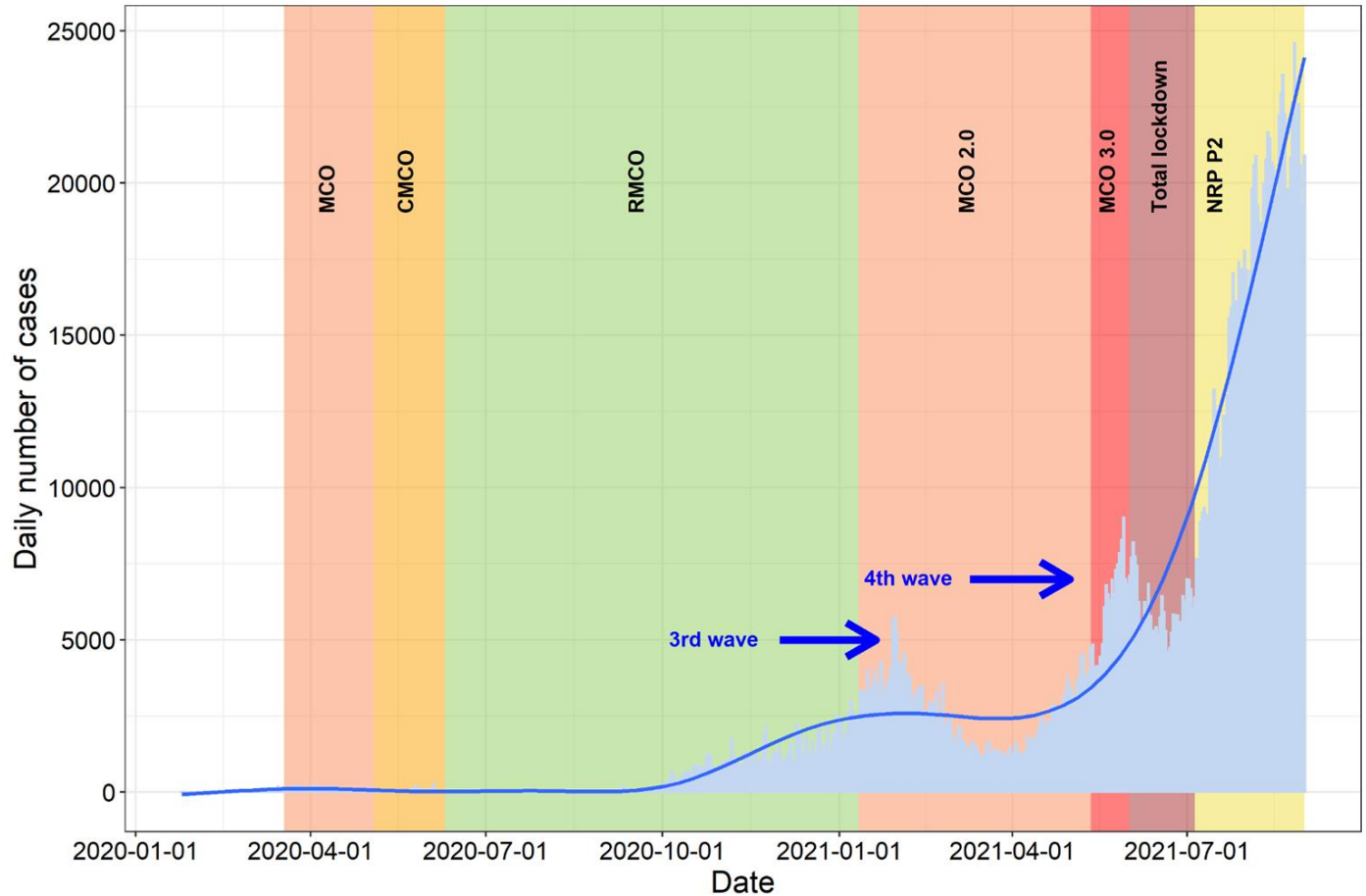
Background

Background (1)

Governmental Restrictions



Background (2) Governmental Restrictions in Relation to COVID-19 Cases



Hypothesis:

We hypothesize that the **3rd wave of COVID-19** and **population dynamics** are essential in determining the outbreak of the 4th wave of COVID-19 cases and its spatial distribution over time.



A decorative network diagram in the top-left corner, consisting of various sized nodes (some solid, some hollow) connected by thin lines, forming a complex web structure.

2.

Data Description

Datasets used

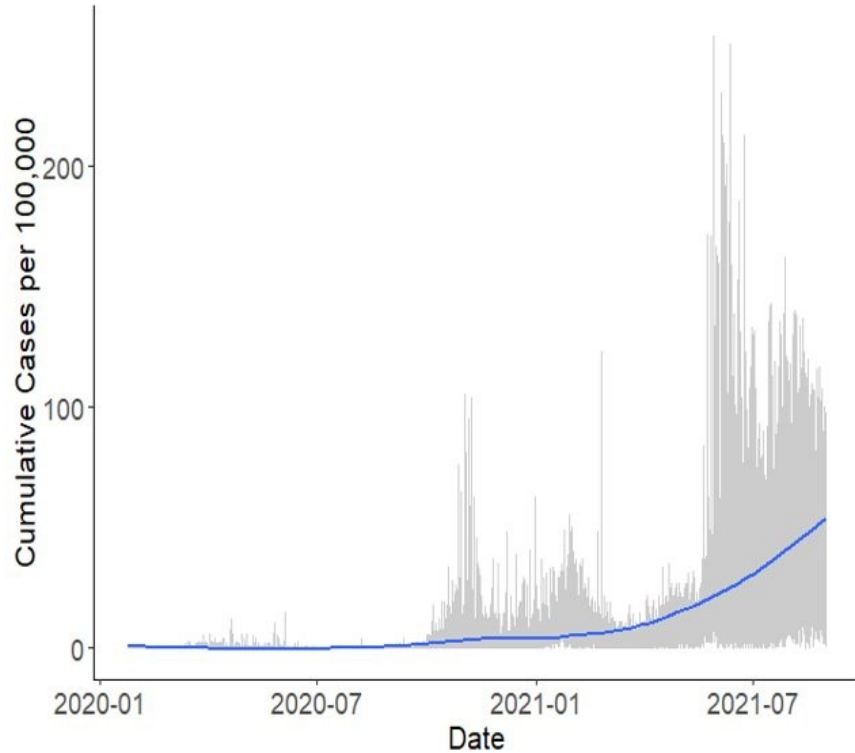
Dataset	Source
COVID-19 Data	Ministry of Health, Malaysia
Socioeconomic Data	Department of Statistics, Malaysia
Migration Data	Department of Statistics, Malaysia
Government Response Indices	Oxford COVID-19 Government Response Tracker



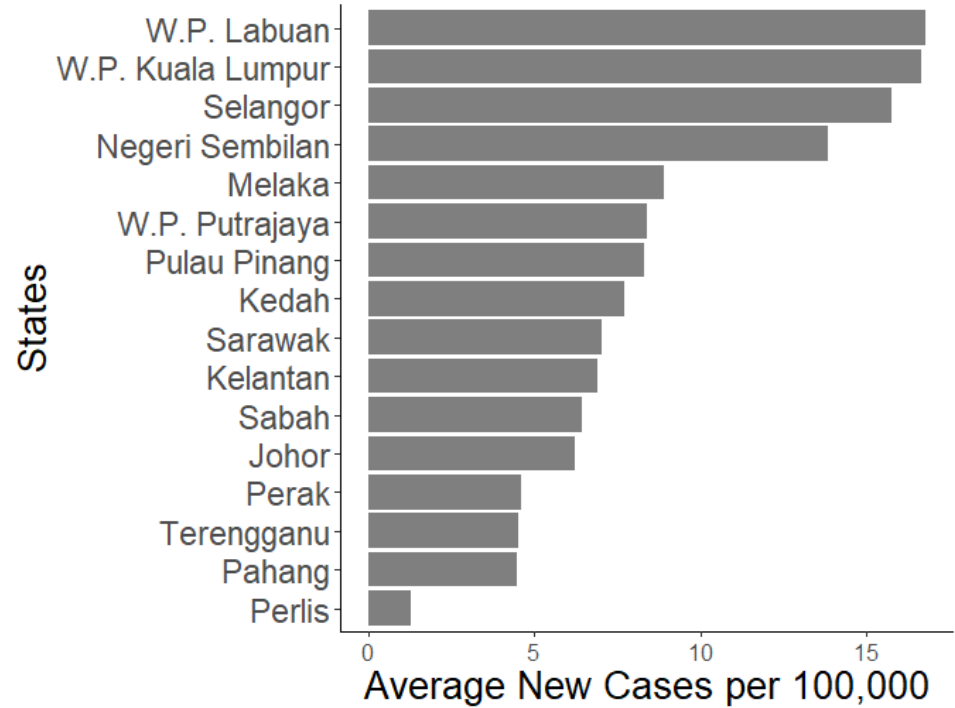
3.

Exploratory Data Analysis

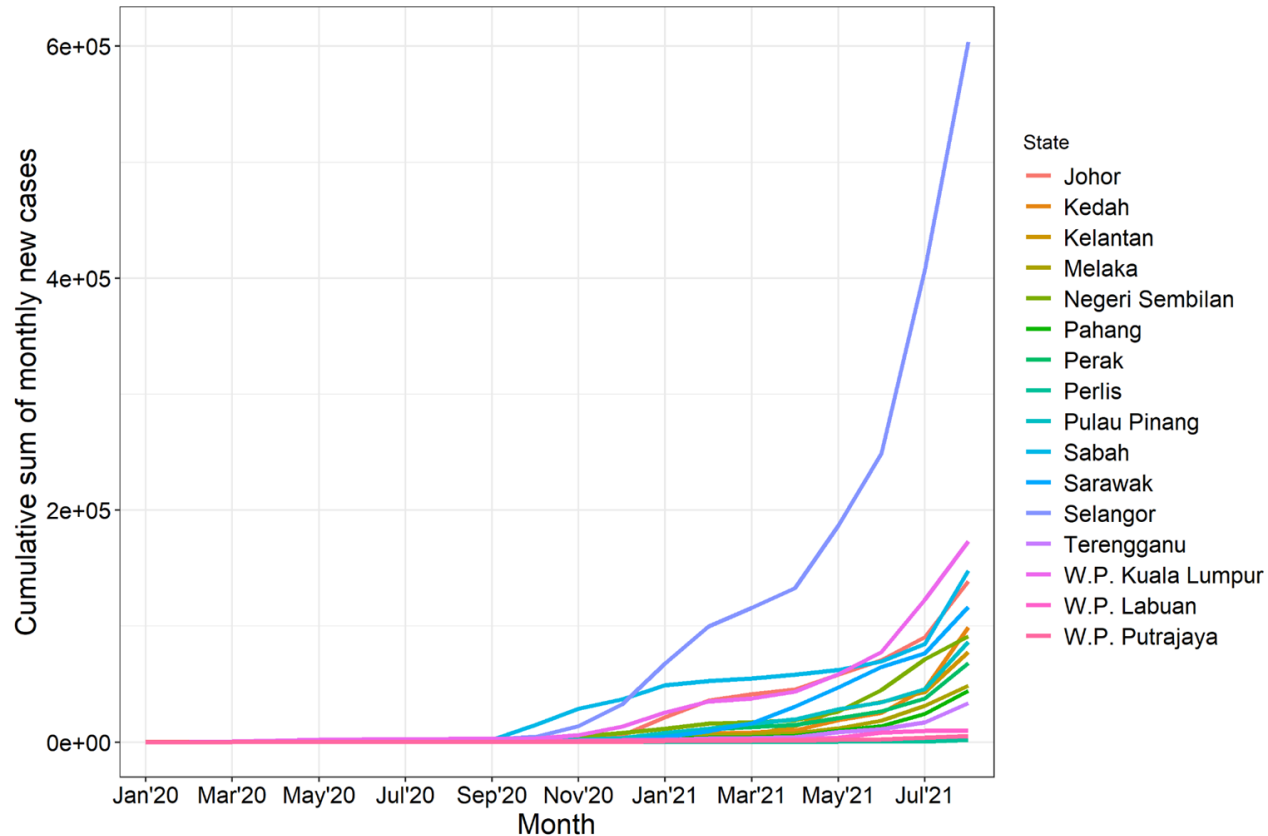
Empirical Spatial Mean



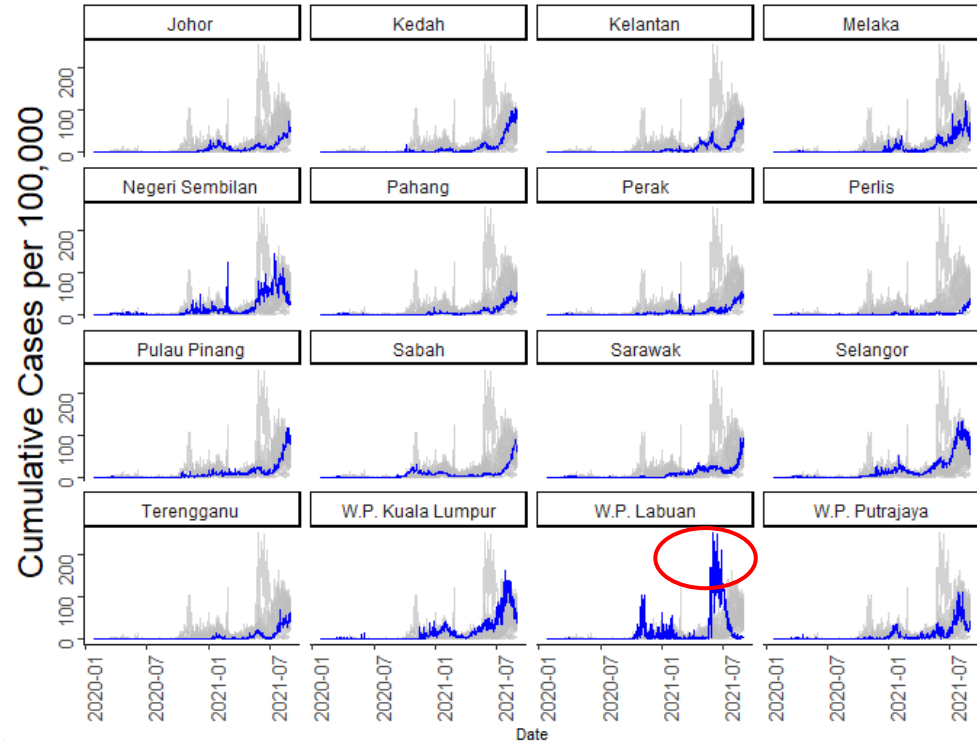
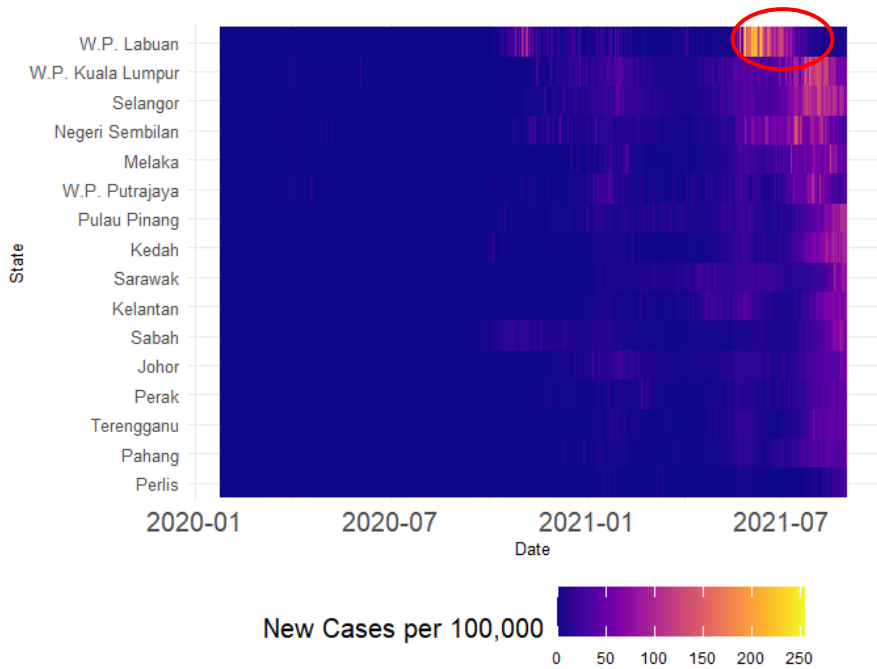
Empirical Temporal Mean



Cumulative Sum of Monthly New Case Rate per 100,000



Further temporal analysis based on state

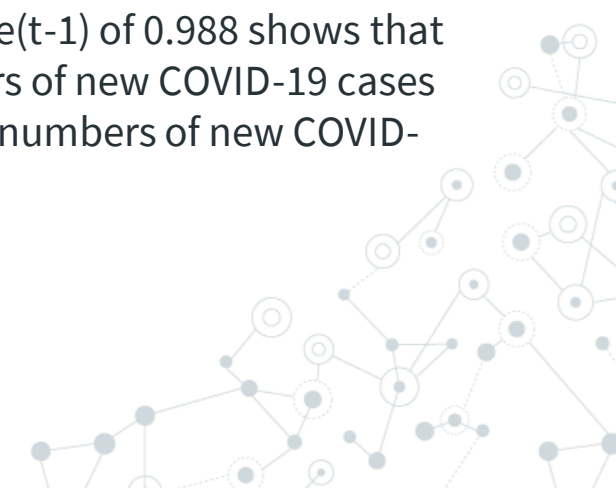




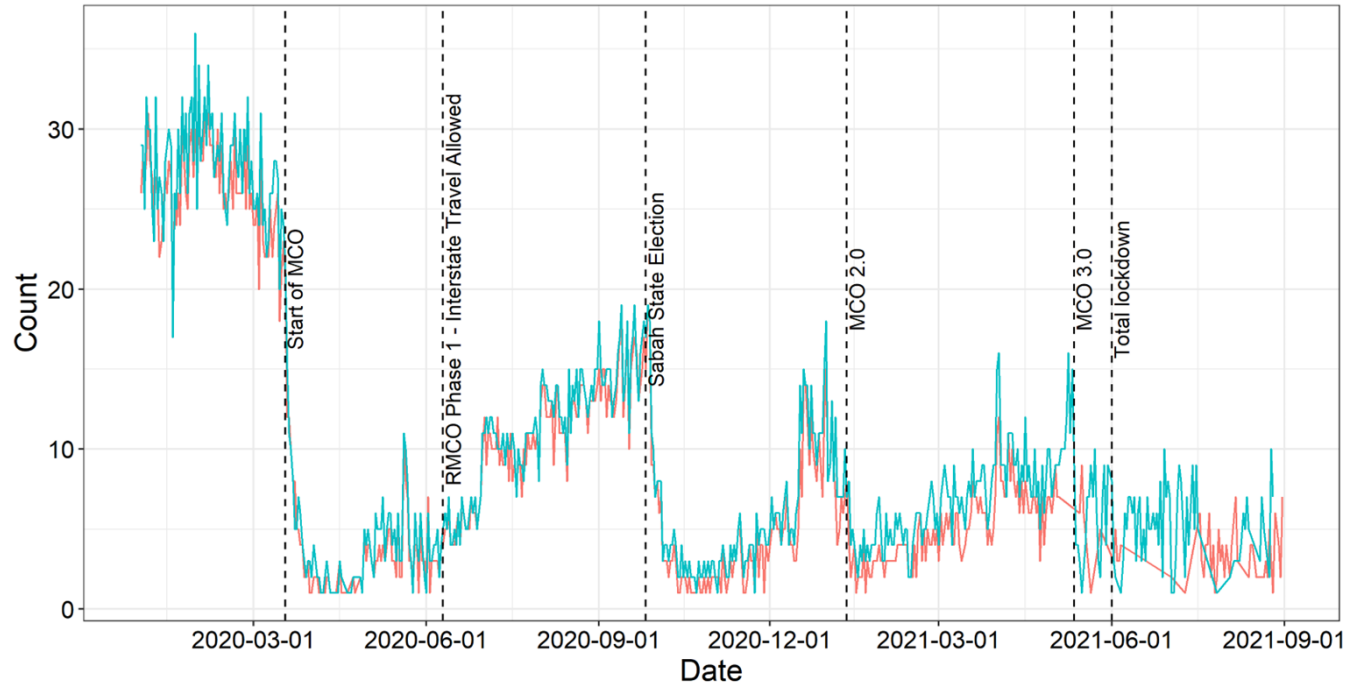
Temporal Dependence (Autocorrelation)

	Covid_new_cases	lag_new_cases
Covid_new_cases	1.0000000	0.9882381
lag_new_cases	0.9882381	1.0000000

Correlation coefficient between time (t) and lag measured at time(t-1) of 0.988 shows that there is high positive autocorrelation; i.e. high (low) past numbers of new COVID-19 cases per 100,000 people tend to correlate with higher (lower) present numbers of new COVID-19 cases



Malaysia Domestic Travel Flight Data



Route

- Kota Kinabalu International Airport-Kuala Lumpur International Airport
- Kuala Lumpur International Airport-Kota Kinabalu International Airport



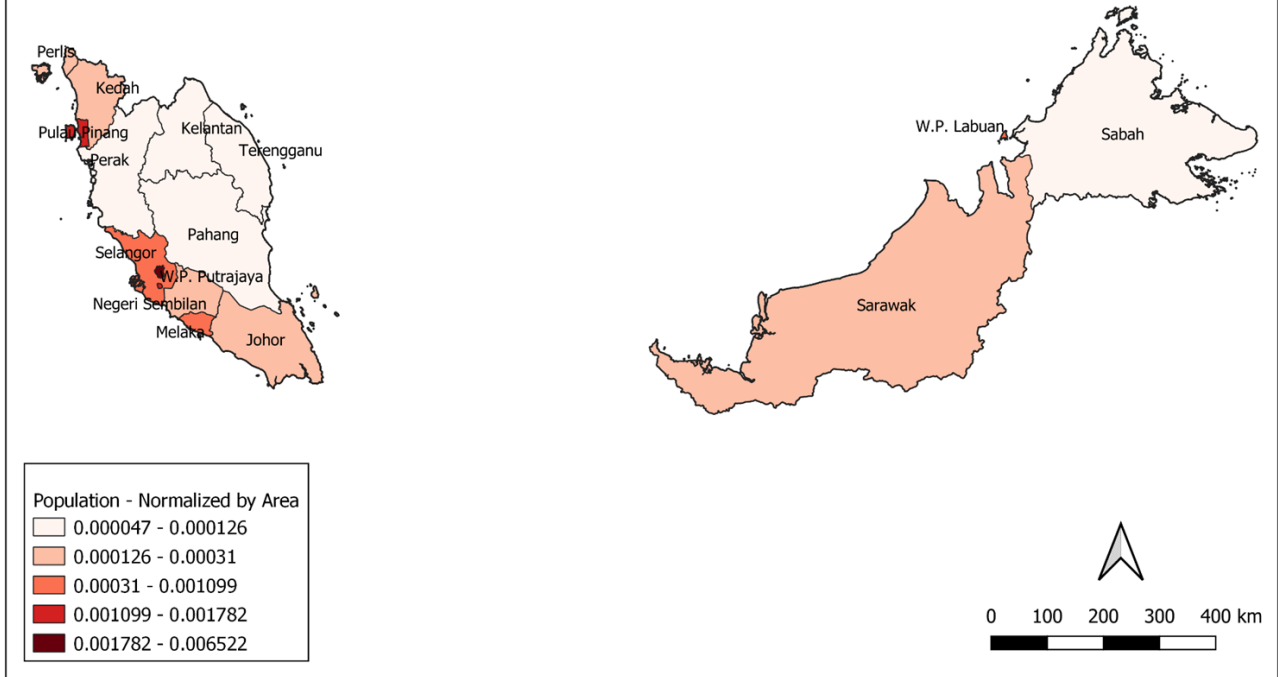
4.

Exploratory Spatial Data Analysis

States with the highest population density:

1. Kuala Lumpur
2. Pulau Pinang
3. Putrajaya
4. Melaka
5. Selangor

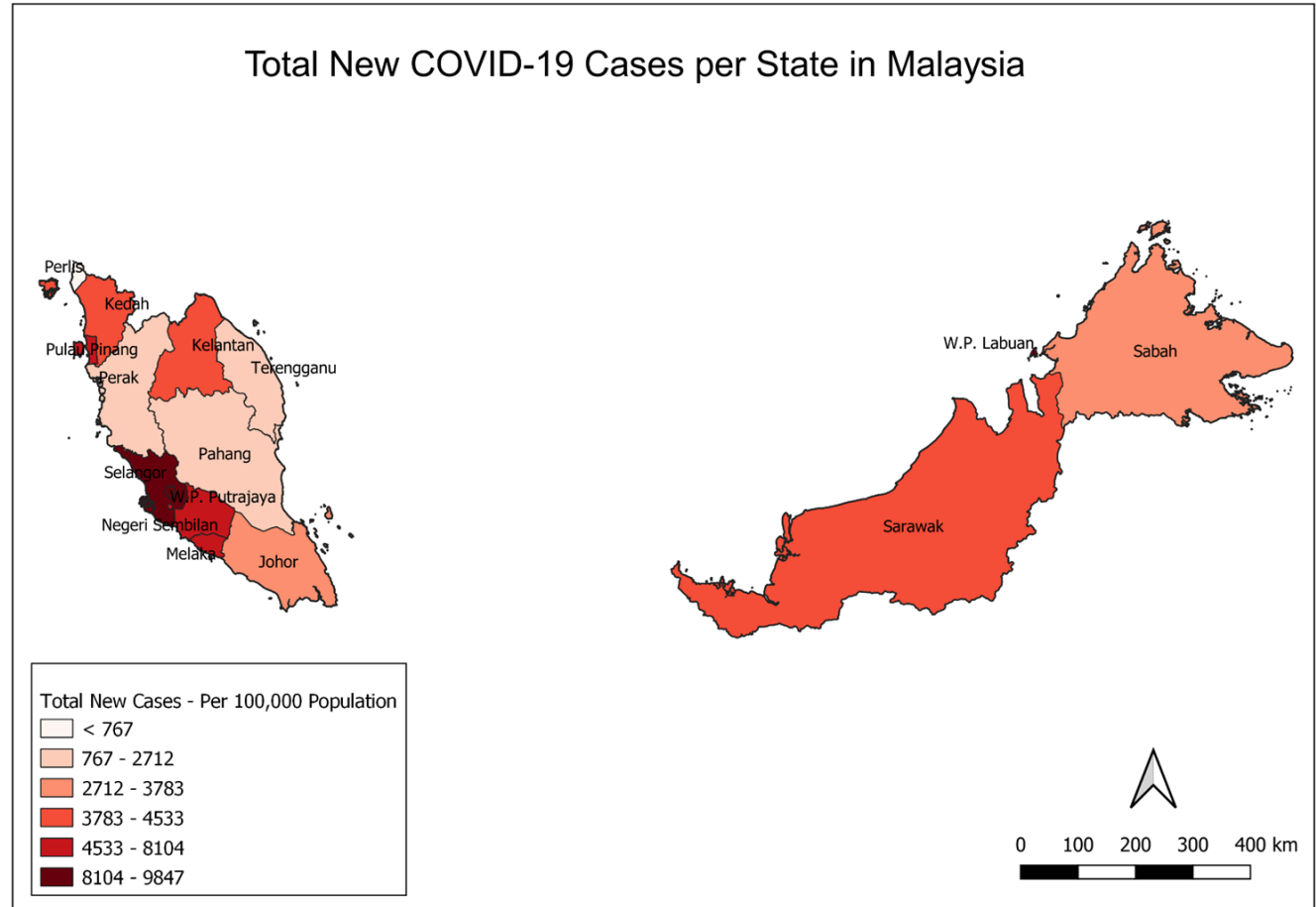
Population Density per State in Malaysia



States with the highest COVID-19 Cases per 100,000

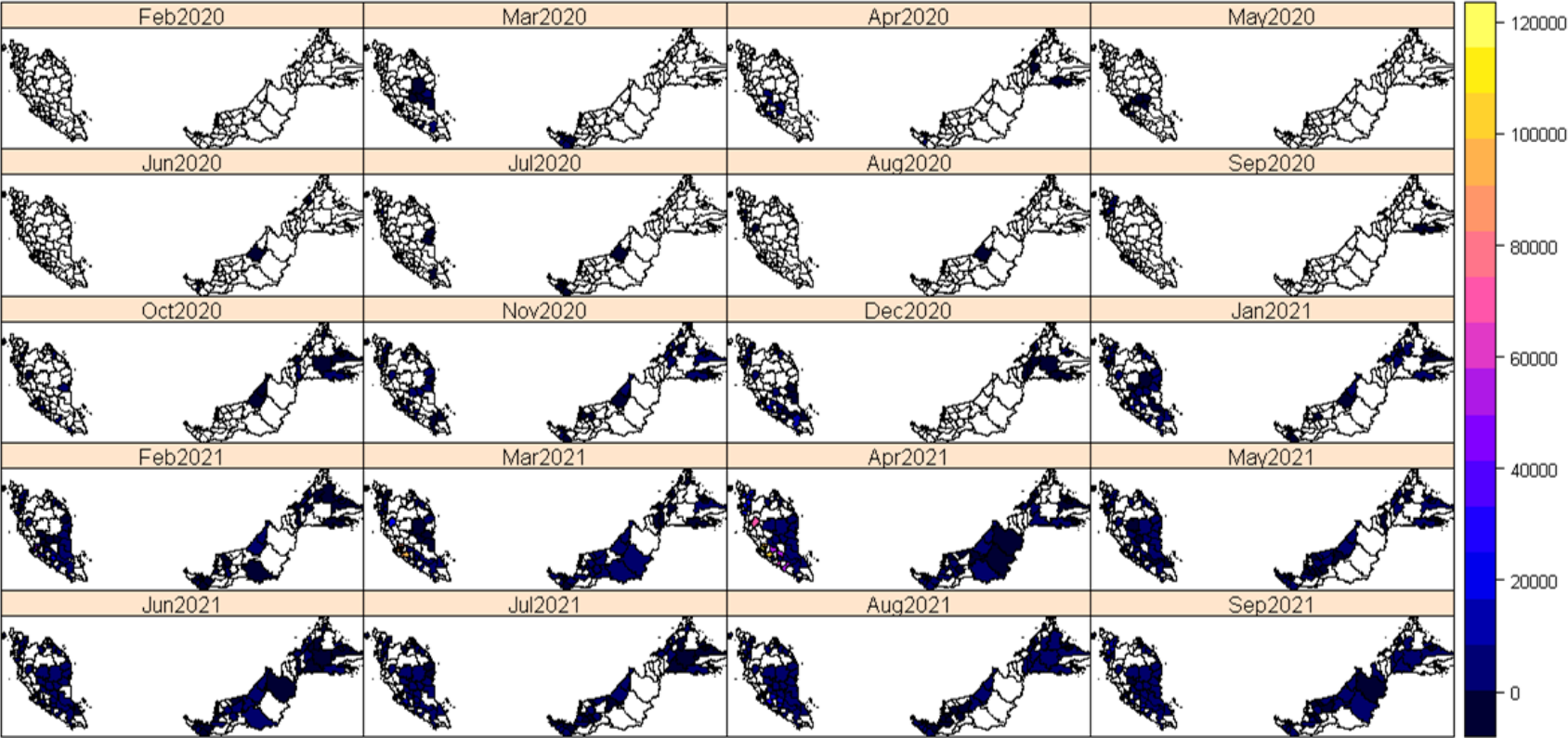
1. Labuan
2. Kuala Lumpur
3. Selangor
4. Negeri Sembilan
5. Melaka

Total New COVID-19 Cases per State in Malaysia



Spatio-Temporal Visualisation of COVID-19 cases at the district level

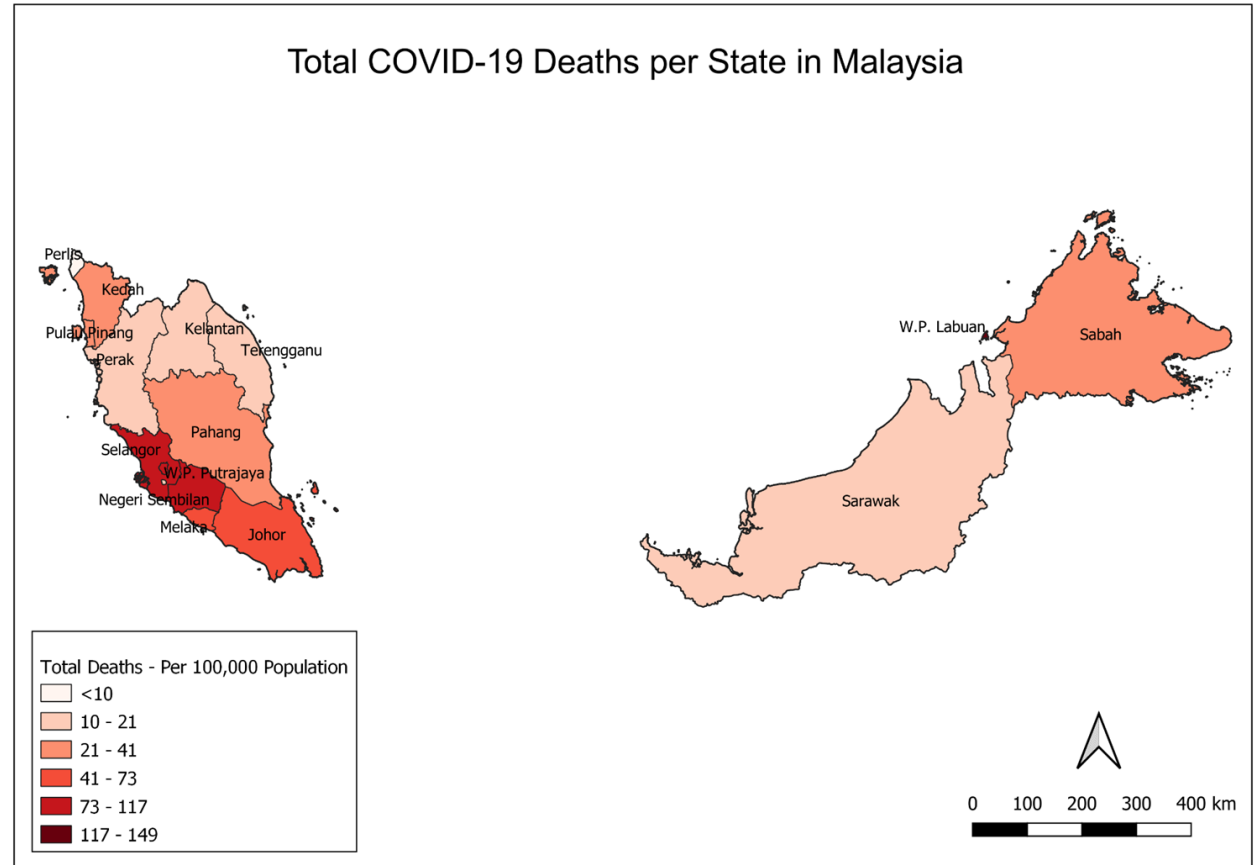
Malaysia COVID-19 Cases by District



COVID-19 Cases per month

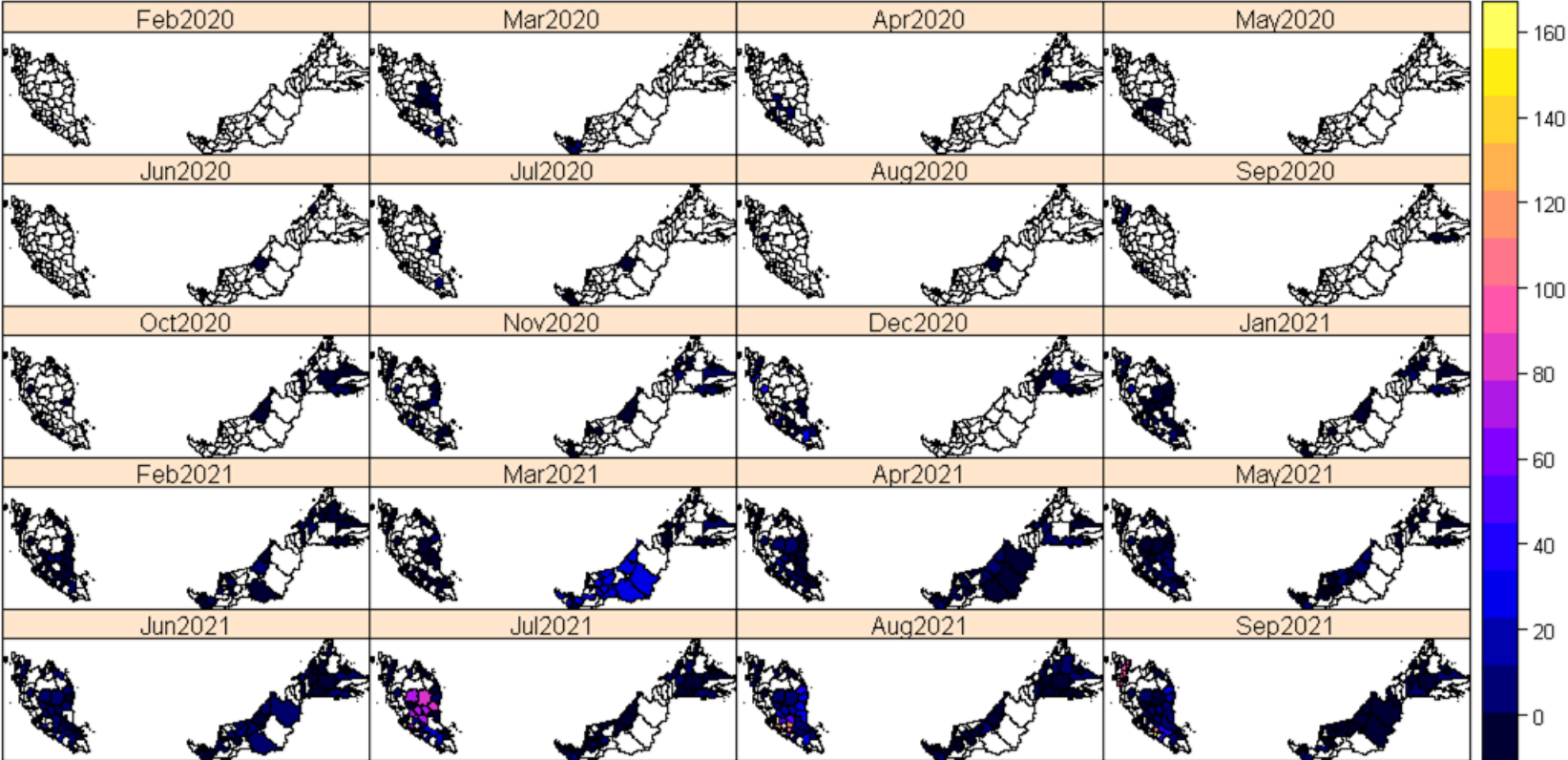
States with the highest COVID-19 Deaths per 100,000

1. Labuan
2. Kuala Lumpur
3. Selangor
4. Negeri Sembilan
5. Melaka



Spatio-Temporal Visualisation of COVID-19 deaths at the district level

Malaysia COVID-19 Deaths by District



COVID-19 Deaths per month



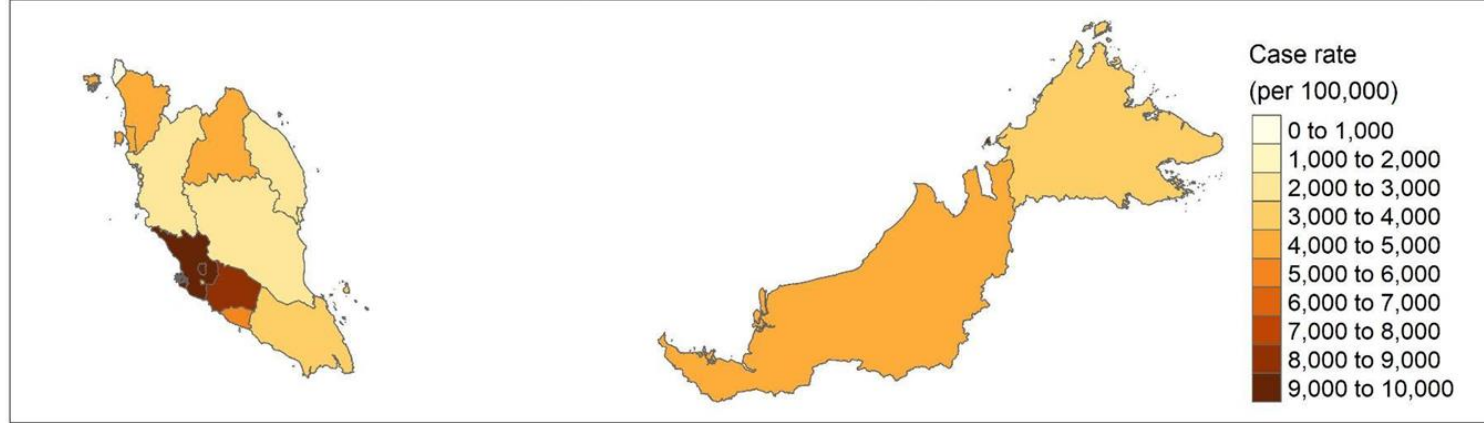
5.

Spatial Data Analysis

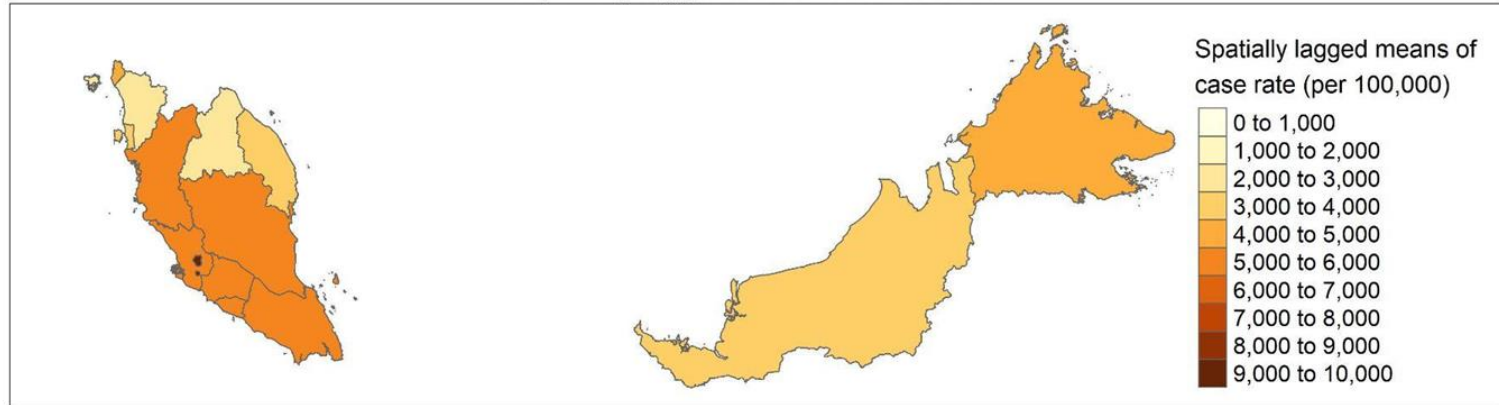
- Moran's I

Spatial Clustering of cases at the national scale

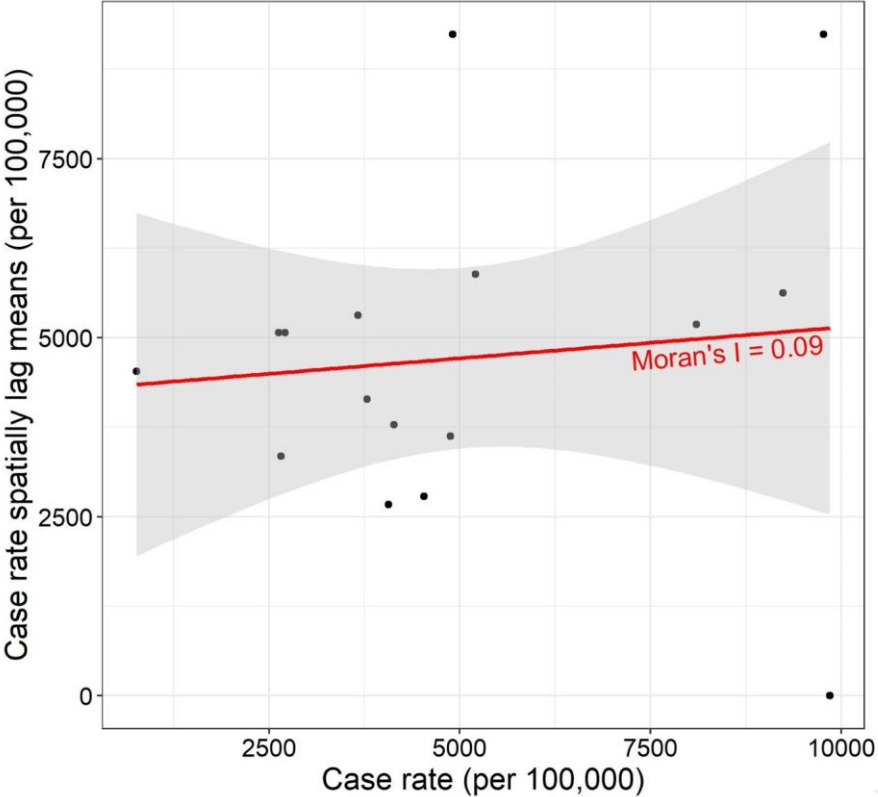
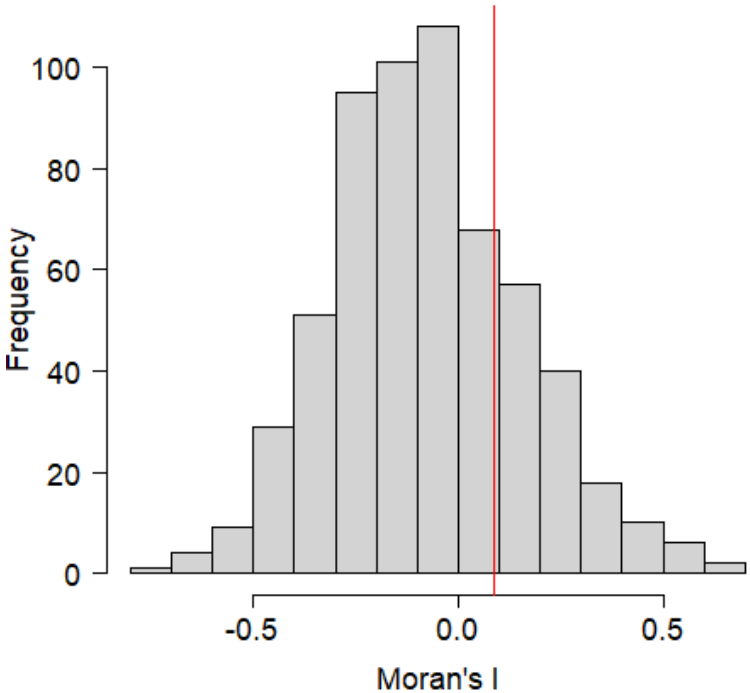
COVID-19 case rate in Malaysia



Spatially lagged means of case rate

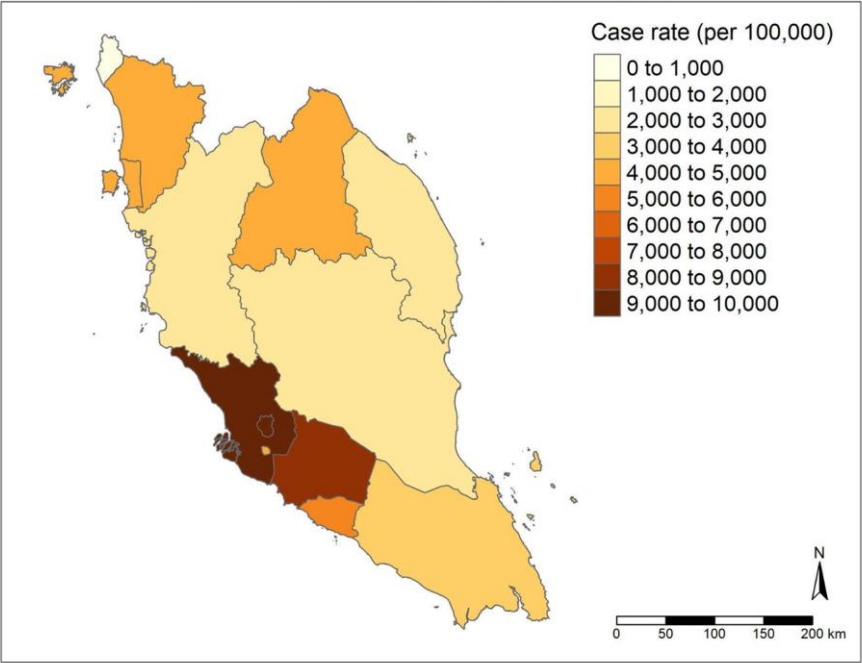


Spatial Clustering of cases at the national scale

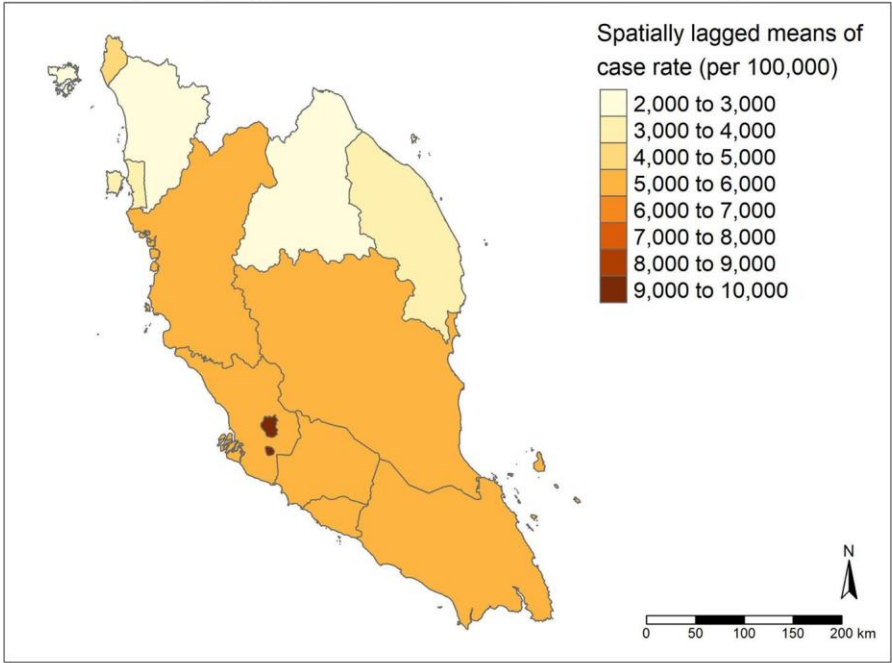


Spatial Clustering of cases in Western Malaysia

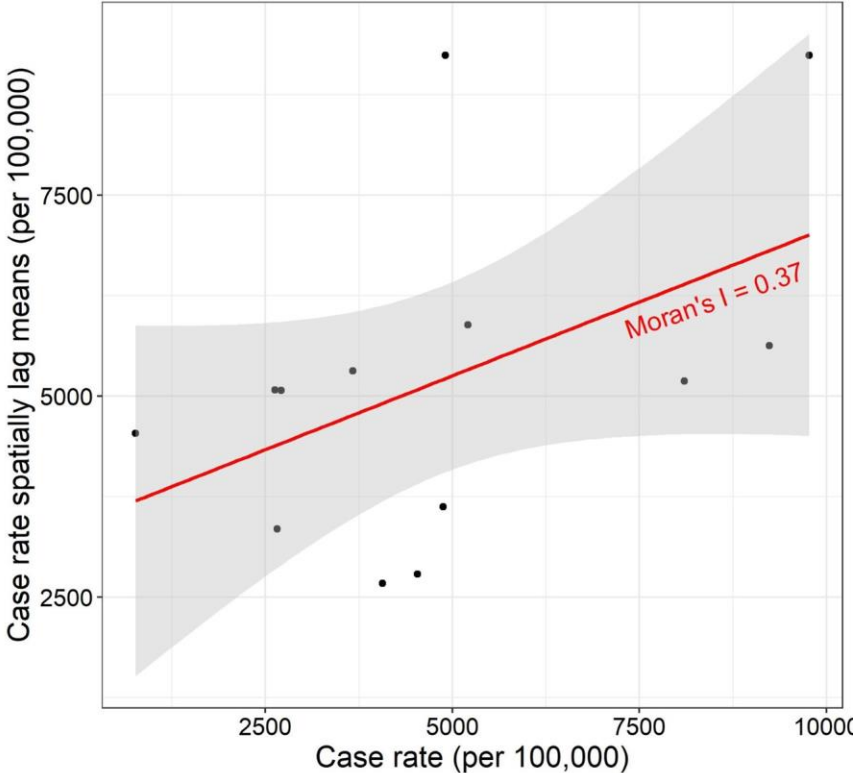
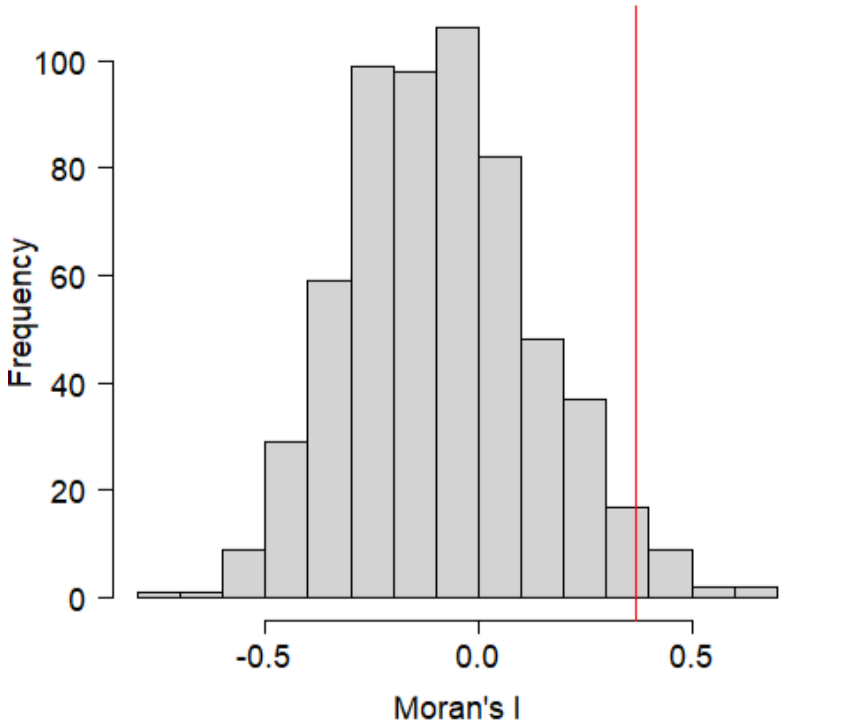
COVID-19 Case rate in West Malaysia



Spatially lagged means of case rate in West Malaysia

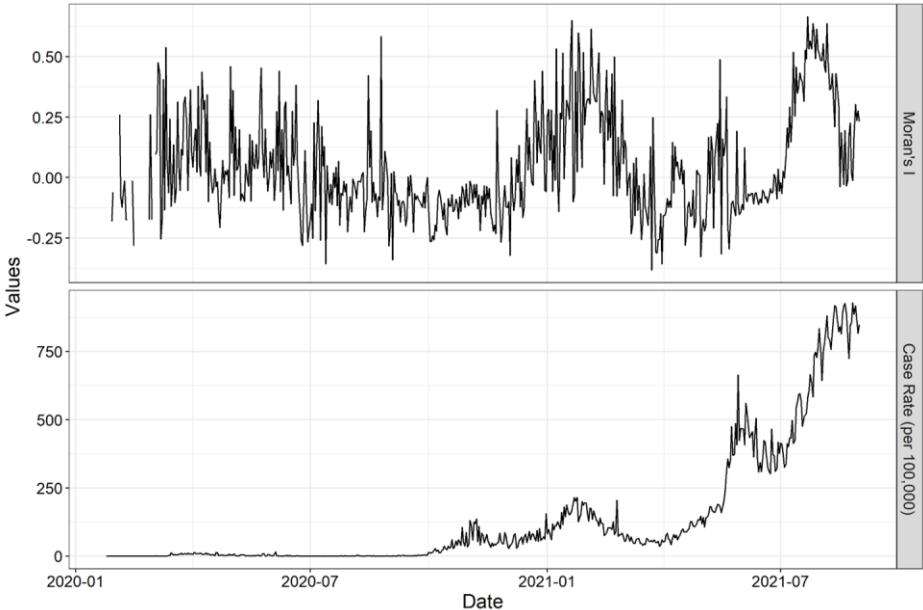


Spatial Clustering of cases in Western Malaysia

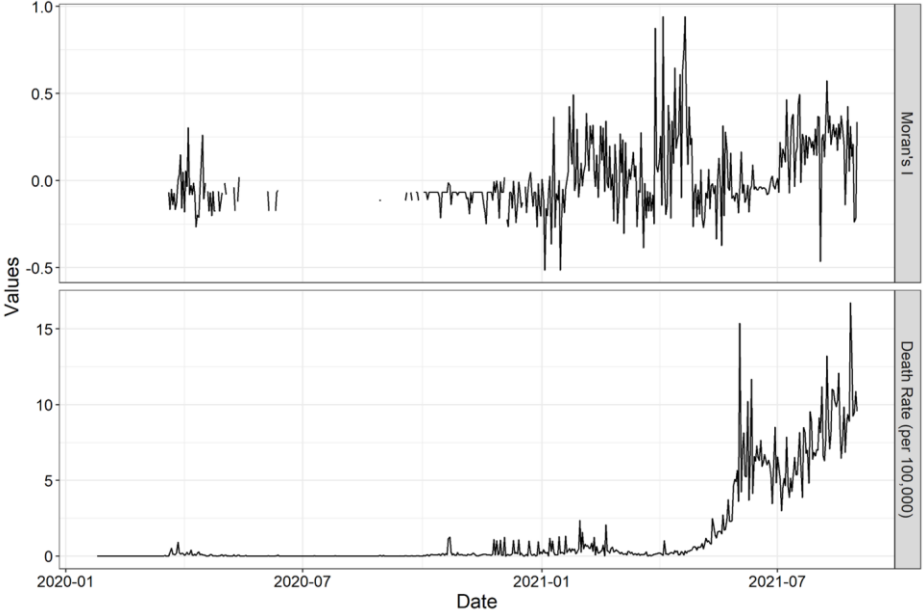


Temporal trend of Global Moran's I

Case rate



Death rate

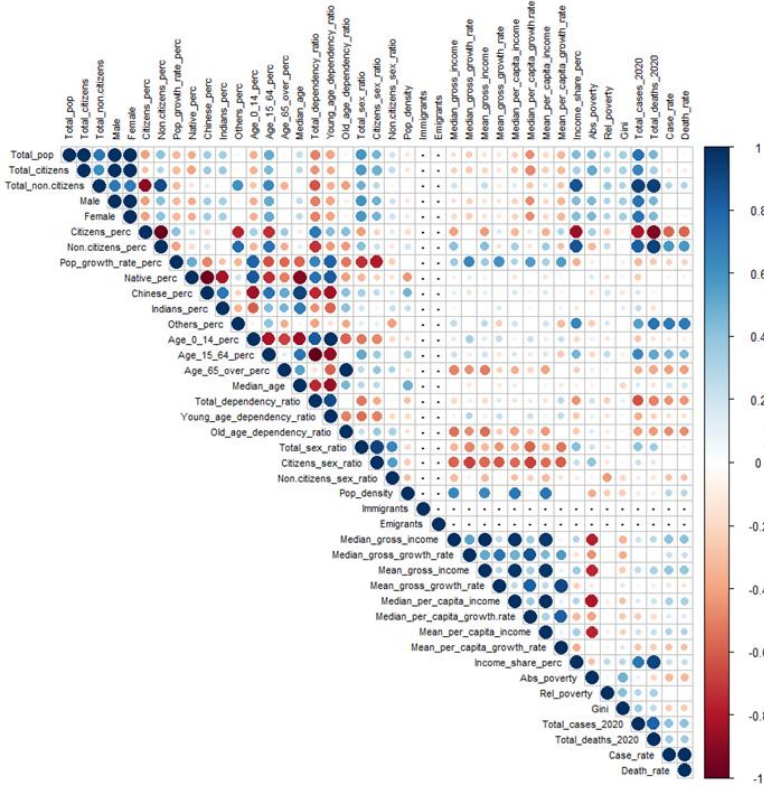




6.

Spatial Data Analysis - Spatial-Autoregression Models

Correlation matrix of chosen predictor variables



Ensured there is no multi-collinearity (> 0.7)

	Others %	Old age dependency	Sex ratio	Immigrants	Gini	Income share %	Abs poverty	Rel poverty	Vacc rate	Bed rate	ICU rate
Indians %	-0.35	0.41	0.19	0.28	0.11	-0.13	0	0.05	0.11	0.28	-0.16
Others %	-0.39	0.13	-0.06	0.01	0.66	-0.25	0.18	-0.06	0.08	0.52	
Old age dependency		0.22	-0.13	0.06	-0.27	0.13	-0.1	-0.36	-0.03	-0.31	
Sex ratio			0.45	0.05	0.43	0.19	-0.12	-0.4	0.33	-0.12	
Immigrants				0.37	0.33	0.36	0.47	-0.28	-0.19	-0.31	
Gini					0.27	0.48	0.41	-0.35	-0.04	-0.44	
Income share %						-0.29	0.25	-0.18	0	-0.09	
Abs poverty							0.06	-0.42	-0.24	-0.34	
Rel poverty								-0.33	-0.42	-0.17	
Vacc rate									0.31	0.44	
Bed rate										0.38	

Case rate spatial-autoregression results

*More minorities
= More cases*

*More males
= More cases*

Predictor variables	Estimate	Std. Error	z value	p-value
<i>Indians %</i>	7.0894e+01	2.7525e+01	2.5756	0.010007
<i>Others %</i>	1.2891e+03	1.8170e+02	7.0945	1.298e-12
<i>Sex ratio</i>	3.6720e+02	3.2530e+01	11.2878	< 2.2e-16
<i>Gini coefficient</i>	2.2046e+04	6.7802e+03	3.2515	0.001148
<i>Income share %</i>	-4.0484e+02	3.7302e+01	-10.8532	< 2.2e-16
<i>Rel poverty</i>	5.6470e+02	8.4012e+01	6.7217	1.796e-11
<i>Vacc rate</i>	1.0423e-01	4.9812e-03	20.9236	< 2.2e-16

Death rate spatial-autoregression results

Predictor variables	Estimate	Std. Error	z value	p-value
<i>Indians %</i>	-2.2769e+00	1.0535e+00	-2.1614	0.0306677
<i>Old age dependency</i>	8.5925e+00	1.5420e+00	5.5722	2.516e-08
<i>Immigrants</i>	3.4654e-01	1.1510e-01	3.0107	0.0026067
<i>Gini coefficient</i>	-6.9314e+02	1.7429e+02	-3.9769	6.982e-05
<i>Income share %</i>	-1.1359e+00	3.7091e-01	-3.0625	0.0021949
<i>Abs poverty</i>	2.9754e+00	7.0197e-01	4.2387	2.248e-05
<i>Relative poverty</i>	1.2916e+01	2.1267e+00	6.0732	1.254e-09
<i>Vaccination rate</i>	1.3924e-03	2.2198e-04	6.2725	3.552e-10
<i>Bed availability rate</i>	5.5877e-03	7.3157e-04	7.6380	2.198e-14
<i>ICU availability rate</i>	2.3527e-02	6.8924e-03	3.4134	0.0006415



7.

Spatial-Temporal Analysis - OLS Regression Model

$$y(s_i, t_j) = \beta_0 + \beta_k x(s_i, t_j) + e(s_i, t_j)$$



OLS Regression Model Features – New COVID-19 Cases Model

Independent Variables:

- State Polygon Centroid Coordinates
- Day
- Week
- Month
- Year
- Vaccination Rate per 100,000 population
- COVID-19 Test Rate per 100,000 population

Dependent variable = new
COVID-19 cases per 100,000
population



OLS Regression Model Results Summary (Cases)

```
Residuals:
  Min       1Q   Median       3Q      Max
-61.722  -3.964  -1.079   1.659  237.567

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.283e+04  7.351e+02 -31.051 < 2e-16 ***
X             1.484e-01  2.688e-02   5.523 3.43e-08 ***
Y            -3.674e-01  1.044e-01  -3.519 0.000436 ***
day          -6.460e-02  7.062e-02  -0.915 0.360318
week         1.041e+00  4.810e-01   2.165 0.030448 *
month       -3.887e+00  2.093e+00  -1.857 0.063373 .
year         1.129e+01  3.639e-01  31.031 < 2e-16 ***
n_vax_r      1.526e-02  7.213e-04  21.155 < 2e-16 ***
n_covtest_r  1.002e-01  2.785e-03  35.999 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.95 on 9351 degrees of freedom
Multiple R-squared:  0.518,    Adjusted R-squared:  0.5176
F-statistic: 1256 on 8 and 9351 DF,  p-value: < 2.2e-16
```

Vaccination
Rate



COVID-19
Testing Rate






OLS Regression Model Features – Death Model

Independent Variables:

- State Polygon Centroid Coordinates
- Day
- Week
- Month
- Year
- New COVID-19 Cases per 100,000 population
- Vaccination Rate per 100,000 population
- Hospital bed Rate per 100,000 population
- ICU bed Rate per 100,000 population
- COVID-19 Test Rate per 100,000 population

Dependent variable = COVID-19
Deaths per 100,000 population



OLS Regression Model Results Summary (Deaths)

```
Residuals:
  Min      1Q  Median      3Q      Max
-2.2221 -0.0278  0.0049  0.0373  6.0134

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.982e+02  1.557e+01  19.161 < 2e-16 ***
X             -3.004e-03  5.441e-04  -5.522 3.44e-08 ***
Y            -1.213e-03  1.863e-03  -0.651  0.515
day           4.664e-04  1.246e-03   0.374  0.708
week        -8.887e-03  8.489e-03  -1.047  0.295
month        2.981e-02  3.694e-02   0.807  0.420
year        -1.475e-01  7.698e-03 -19.159 < 2e-16 ***
n_covid19_r  1.152e-02  1.904e-04  60.514 < 2e-16 ***
n_vax_r       6.261e-05  1.381e-05   4.534 5.85e-06 ***
n_beds_r      2.400e-04  3.773e-05   6.361 2.09e-10 ***
n_icu_r       2.180e-02  1.454e-03  14.996 < 2e-16 ***
n_covtest_r  -3.860e-04  5.247e-05  -7.357 2.04e-13 ***
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2461 on 9348 degrees of freedom
Multiple R-squared:  0.4904,    Adjusted R-squared:  0.4898
F-statistic: 817.7 on 11 and 9348 DF,  p-value: < 2.2e-16
```

- New COVID-19 Cases
- Vaccination Rate
- Hospital Bed Rate
- ICU Bed Rate
- COVID-19 Test Rate

Future steps and considerations for Spatio-Temporal Model...

- A key issue to note is that the dependent variable (deaths/cases) is a count and is highly skewed, and the Vaccination, ICU bed, hospital bed data all has no data for earlier portion of the time period in the dataset
- Consider lagged variables, some relationship between covariates might have delayed effect.
- More socio-economic variables might be instrumental in building a more robust model, provided they are important features.

A decorative network diagram in the top-left corner, consisting of various sized circles (nodes) connected by thin lines (edges). Some nodes are solid grey, while others are hollow with a grey outline. The network is dense and irregular, extending from the top-left towards the center of the page.

8. Discussion


COVID-19 Propagation

Population Density

- Densely populated areas as a catalyst for 3rd and 4th Wave
- Politically linked territories but considerations to be made a smaller scales - district/states
- Aligning with governmental restrictions at the time for analysis - Urban Planning*

Hospital Capacity

- Hospital Bed Space and ICUs as important indicators
- Positive Estimates for Bed Space/ICU predictors in spatial autoregression model and spatio-temporal model
- Spatial Data Accuracy on ICU and Bed space availability



9. Conclusion

1. Importance of spatially accurate data for further analysis
1. Socio-economic factors such as population density are important factors for consideration
1. Local community levels of implementation to enhance adherence to government restrictions