

ECON 414: Final Project

Sergio Cornejo, Angelina Chang, Claire Szilagyi, Hanna Fei

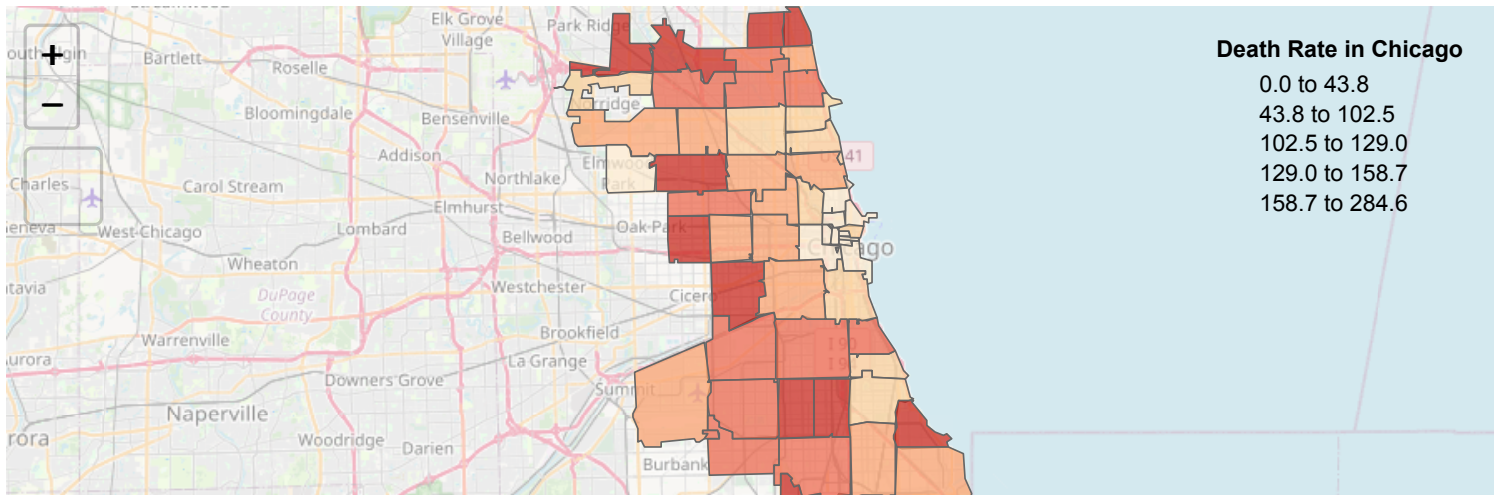
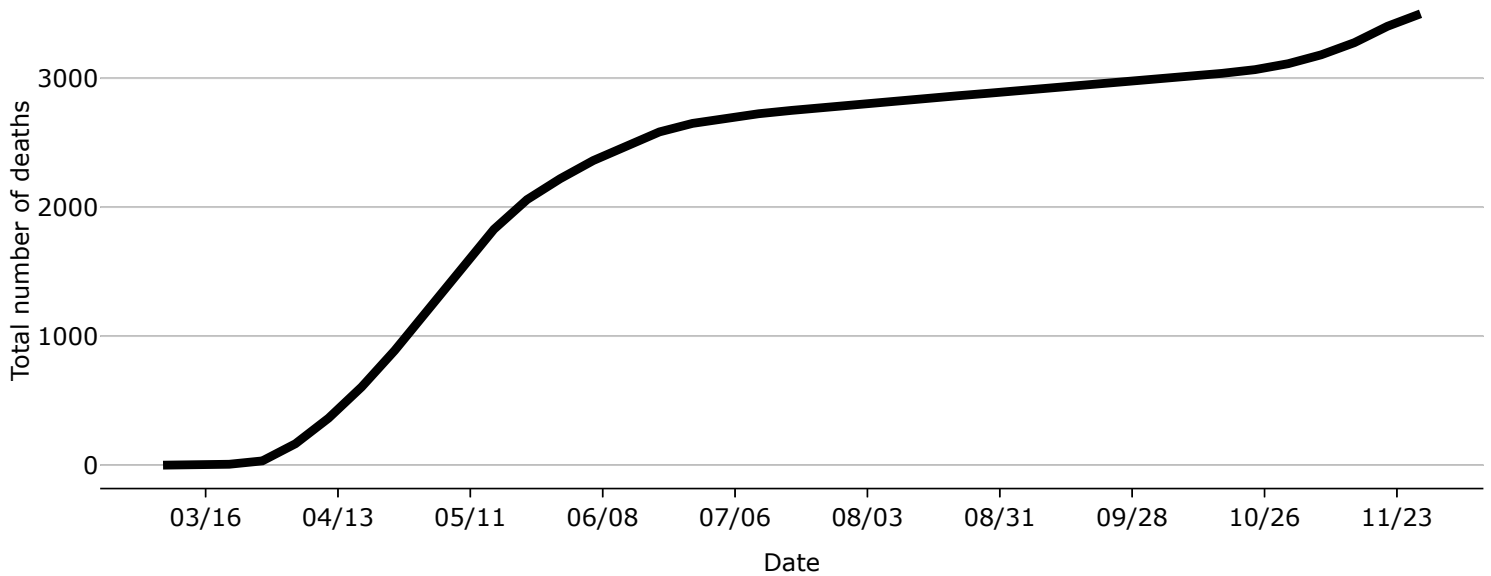
12/19/2020

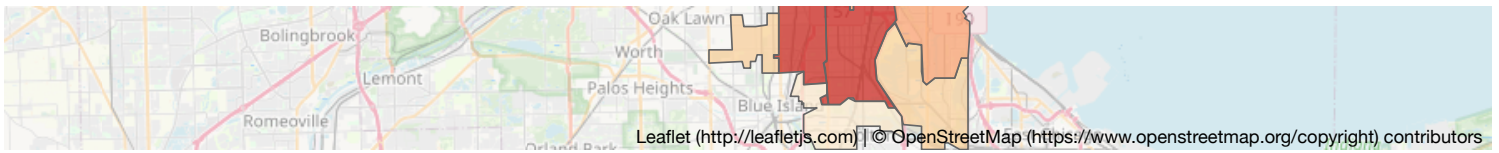
- 1 Introduction
- 2 ESDA
 - 2.1 Moran's I and LISA map for Death Rates
 - 2.2 Moran's I and LISA map for Case Rates
- 3 Regression Analysis
- 4 Final remarks

1 Introduction

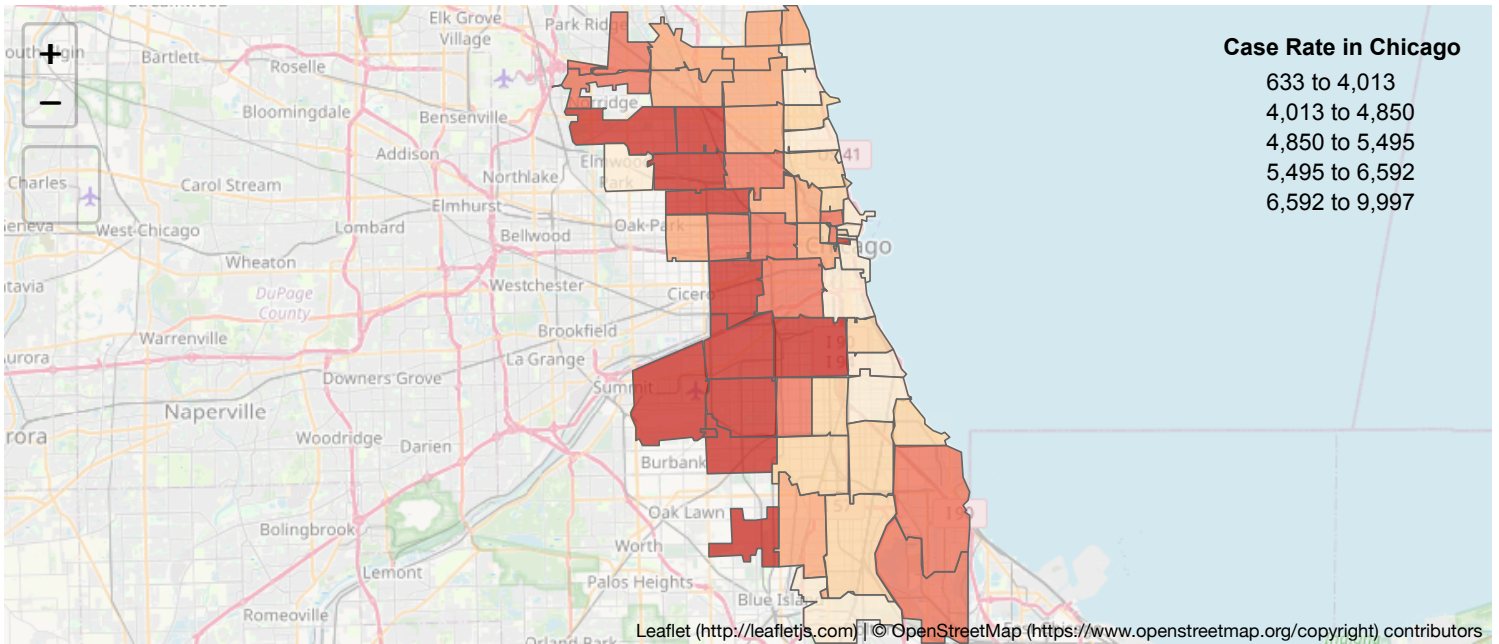
According to the World Health Organization (WHO), as of May 15, there have been more than 4 million 300 thousand coronavirus cases worldwide, and more than 290,000 deaths. We are interested in investigate the factors that might cause the increase of confirmed coronavirus cases and deaths. By doing that, we first make the confirmed cases and deaths in Chicago areas with some possible factors that might have casual relationship with in detail. By doing this we will know where the death and cases take place and observed if there is a spatial pattern among coronavirus in Chicago.

The graph below illustrates that the slope of the total number of deaths was at the highest between dates 03/16 and 06/08. The death since 06/08 has still been positive as more people are dying from coronavirus in Chicago.





From this map of death rate in Chicago, one can conclude that death rates are higher in neighborhoods that are further from downtown.

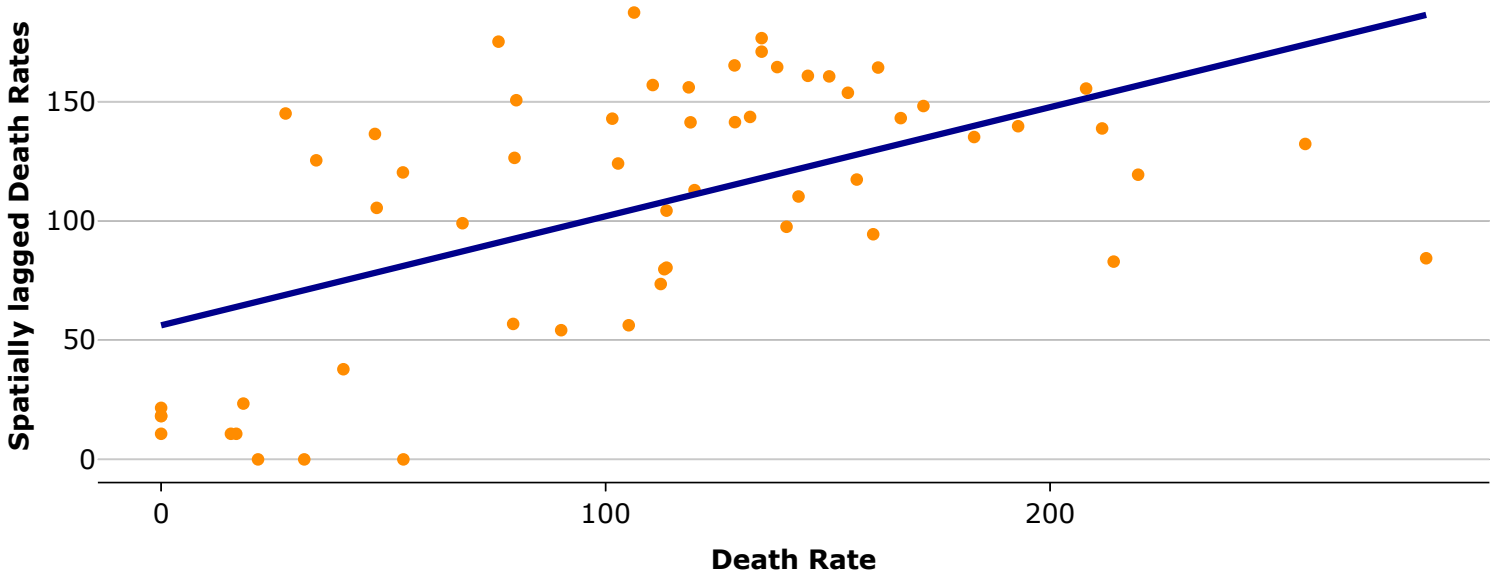


The map of case rates in Chicago illustrates that neighborhoods that are on the west side have higher case rates than neighborhoods that are either further north or further south.

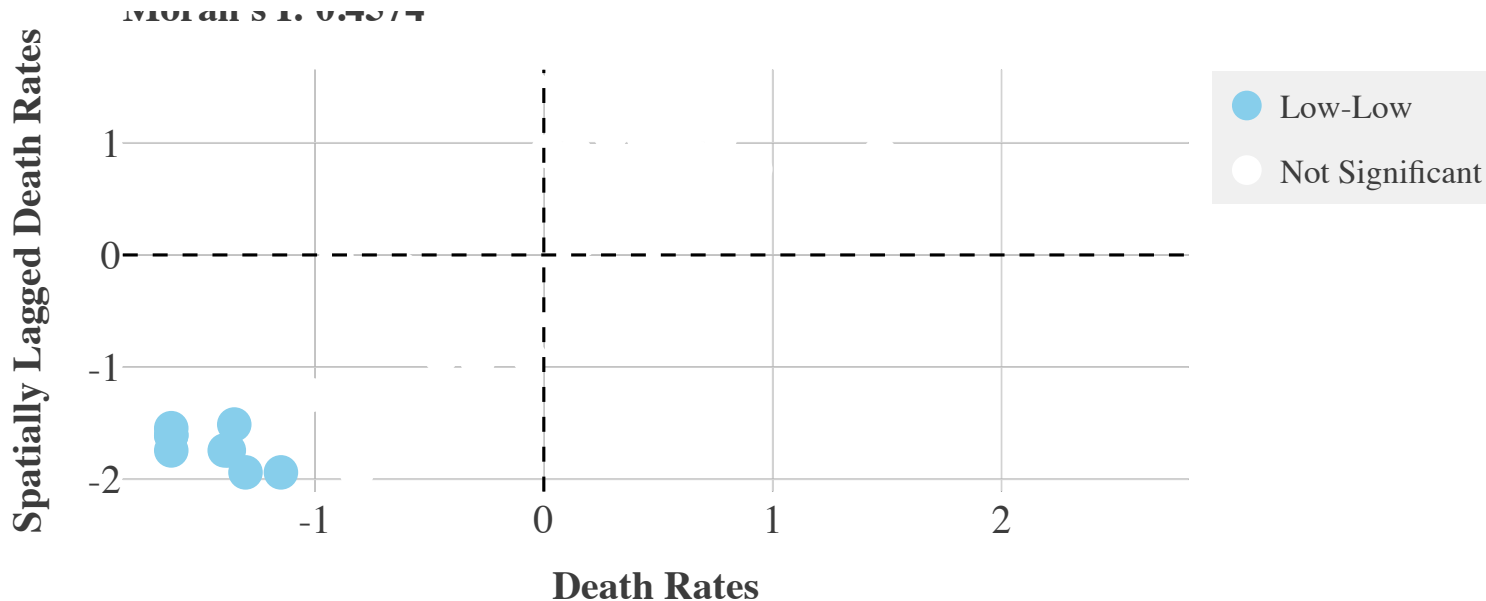
2 ESDA

2.1 Moran's I and LISA map for Death Rates

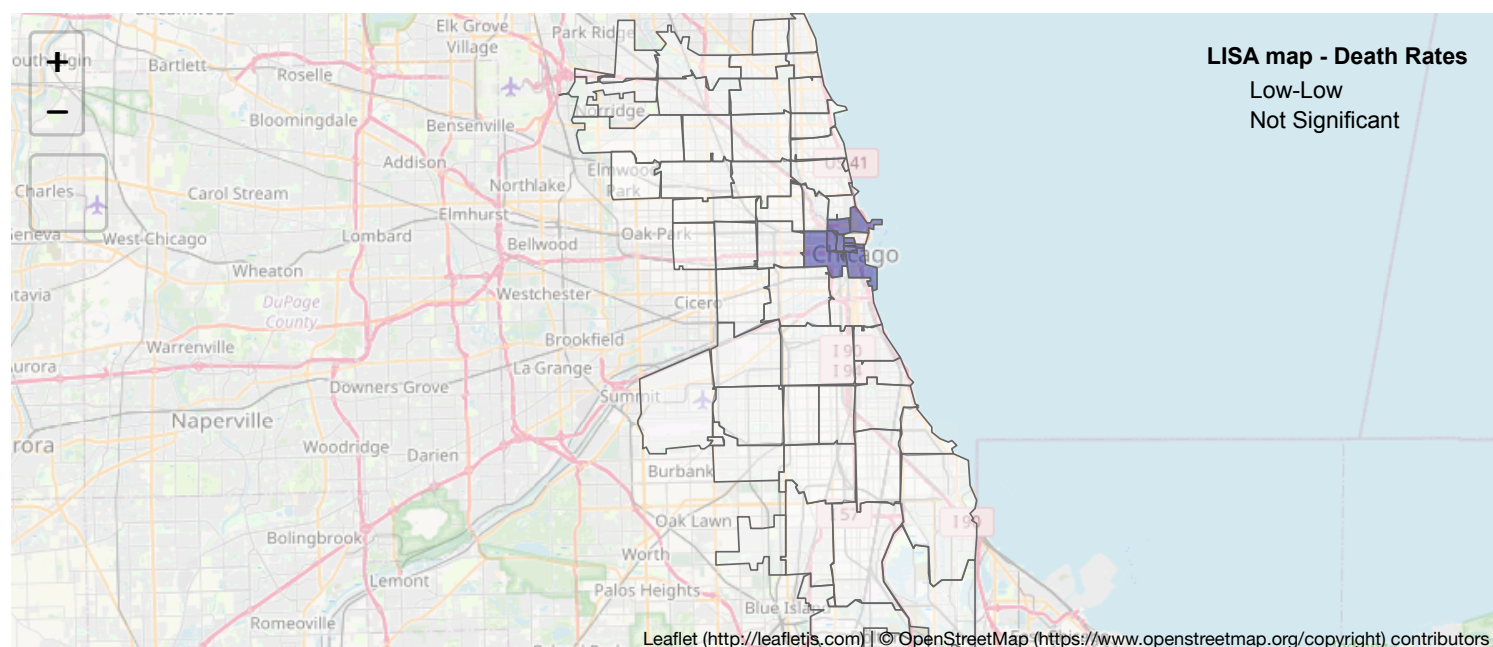
This is the scatter plot of death rate and spatially death rates. From the plot conduct we learned that there is a positive relationship between them. There also appears to be a strong and positive autocorrelation.



Moran's I: 0.1574

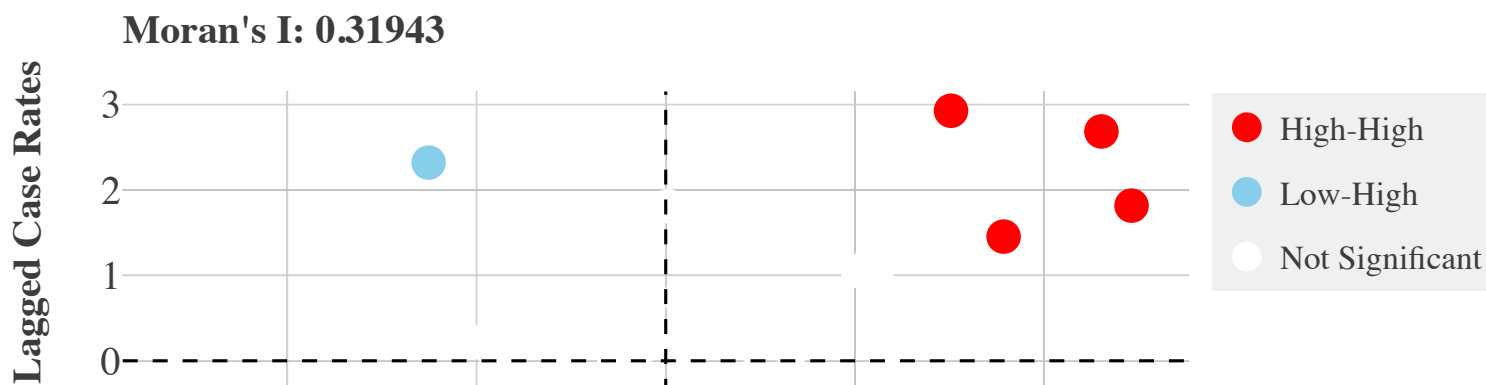


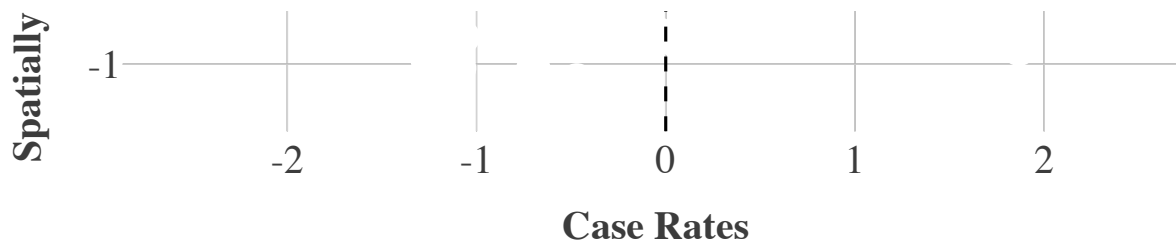
The Moran's I from the dataset is 0.4574. According to this result, 0.4574 was inside the critical region and one can reject the null hypothesis. In conclusion, the data is not randomly distributed over space.



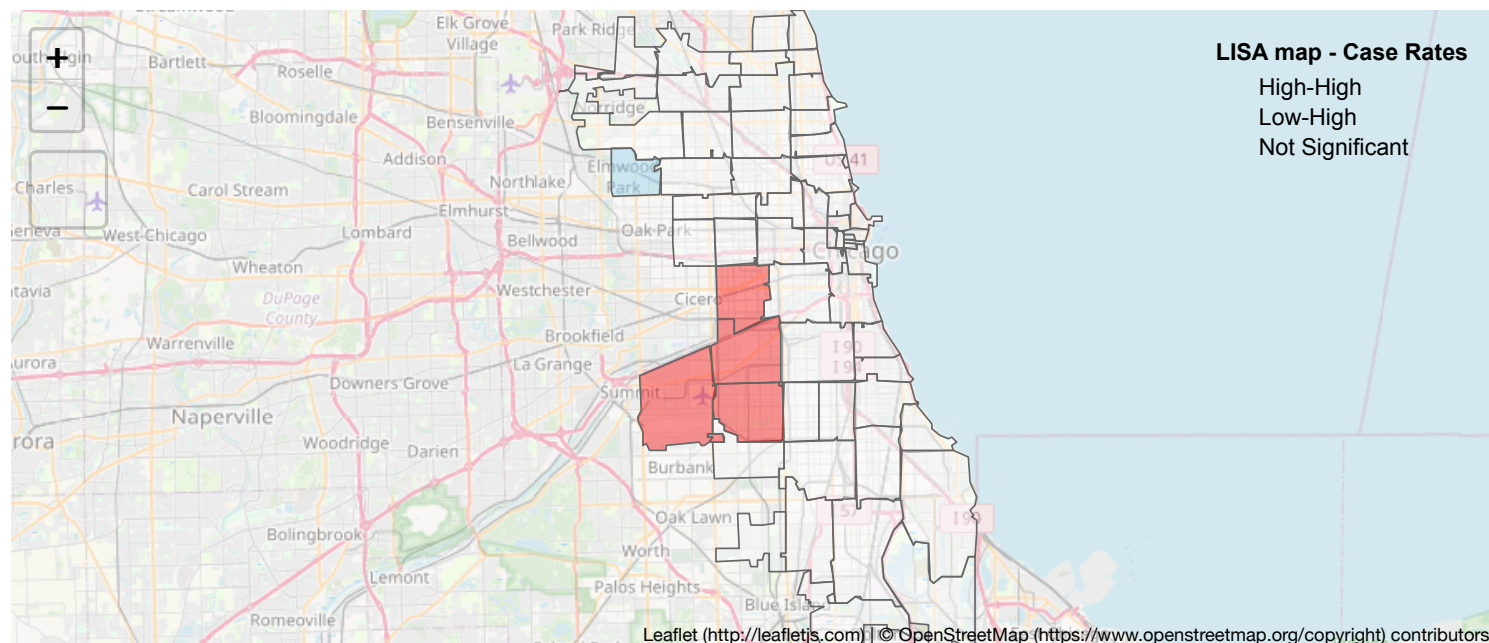
All the neighborhoods with p-values higher than .05 will be considered "Not significant". In this LISA map, the neighborhoods with "low-low" are all around downtown Chicago.

2.2 Moran's I and LISA map for Case Rates





The Moran's I from the dataset is 0.31943. According to this result, 0.4574 was inside the critical region and one can reject the null hypothesis. In conclusion, the data is not randomly distributed over space.



This LISA map explains that the most significant neighborhoods for case rates is around the central west side of Chicago

3 Regression Analysis

Here is where I write my model

$$\text{Death rate}_i = \beta_0 + \beta_1 \text{popdens}_i + \beta_2 \text{above60}_i + \beta_3 \text{crowdedHH}_i + \beta_4 \text{HICov}_i + \beta_5 \text{belowpov}_i + \epsilon$$

where popdens_i is the population density in zip code i , above60_i is the share of people 60 or older in zip code i , crowdedHH_i is the share of houses with 4 or more people in zip code i , and HICov_i is the share of people with health insurance in zip code i .

$$\text{Case rate}_i = \beta_0 + \beta_1 \text{testrate}_i + \beta_2 \text{above60}_i + \beta_3 \text{crowdedHH}_i + \beta_4 \text{HICov}_i + \beta_5 \text{belowpov}_i + \beta_6 \text{popdens}_i + \epsilon$$

where testrate_i is the number of tests per 100,000 people in zip code i .

Death Rate Regression

```
summary(reg)
```

```
##
## Call:
## lm(formula = Deathrate ~ above60 + popdens + crowdedHH + HIcov +
##     belowpov, data = covid19_chitown@data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -91.748 -30.546  -2.974   28.636  144.004
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.536e+02  2.143e+02   1.184   0.2419
## above60      3.183e+02  1.235e+02   2.578   0.0128 *
## popdens      2.297e-03  2.297e-03   1.000   0.3218
## crowdedHH     2.085e+02  1.005e+02   2.075   0.0429 *
## HIcov        -3.117e+02  2.084e+02  -1.495   0.1408
## belowpov      1.583e+02  7.941e+01   1.994   0.0514 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 50.32 on 52 degrees of freedom
## Multiple R-squared:  0.4885, Adjusted R-squared:  0.4393
## F-statistic: 9.931 on 5 and 52 DF, p-value: 1.043e-06
```

The regression for death rates illustrates that variables, population density, crowded households, share of people with health insurance, and share of people below the poverty line are insignificant because they are higher than alpha is 0.05. This regression also illustrates that there is a positive relationship between people above 60 and death rates.

Case Rate Regression

```
summary(reg)
```

```
##
## Call:
## lm(formula = Caserate ~ above60 + popdens + crowdedHH + HIcov +
##     belowpov + test_rate, data = covid19_chitown@data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2954.1  -772.7    -9.4    487.9   4134.4
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.834e+03  4.845e+03   0.378   0.707
## above60      -5.057e+02  2.816e+03  -0.180   0.858
## popdens      -3.059e-04  5.136e-02  -0.006   0.995
## crowdedHH     1.571e+04  2.399e+03   6.546 2.84e-08 ***
## HIcov        -2.450e+03  4.664e+03  -0.525   0.602
## belowpov     -2.521e+03  1.779e+03  -1.417   0.162
## test_rate      4.899e-02  8.550e-03   5.730 5.41e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1125 on 51 degrees of freedom
## Multiple R-squared:  0.661, Adjusted R-squared:  0.6211
## F-statistic: 16.57 on 6 and 51 DF, p-value: 1.739e-10
```

The regression for death rates illustrates that variables, population density, share of people with health insurance, people over 60, and share of people below the poverty line are insignificant because they are higher than alpha is 0.05. This regression also illustrates that there is a positive relationship between crowded households and test rates with death rates.

4 Final remarks

In conclusion, data is not randomly distributed in Chicago. Death and case rates seem to be more significant in states that are west of downtown. Some possible variables that can explain the situation could be the share of people that are over 60 years old and the share of houses with four people or more.