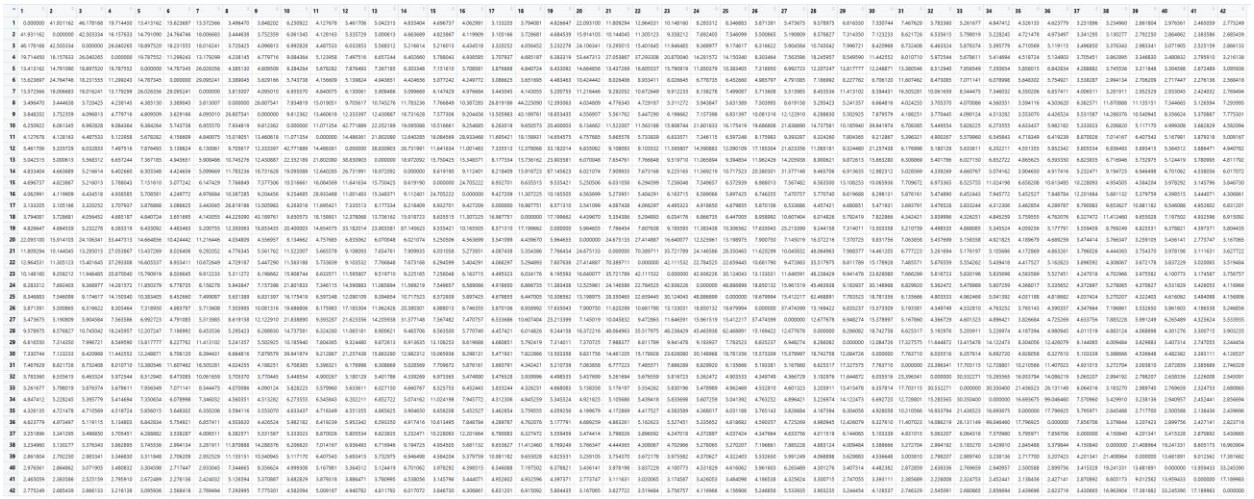
**INFSCI 2809: Spatial Data Analytics**

**Project 3 Report**

Student: Zechen Wang (zew20)

## I. Autocorrelation

The spatial weights are (42*42 matrix):



According to the formula:

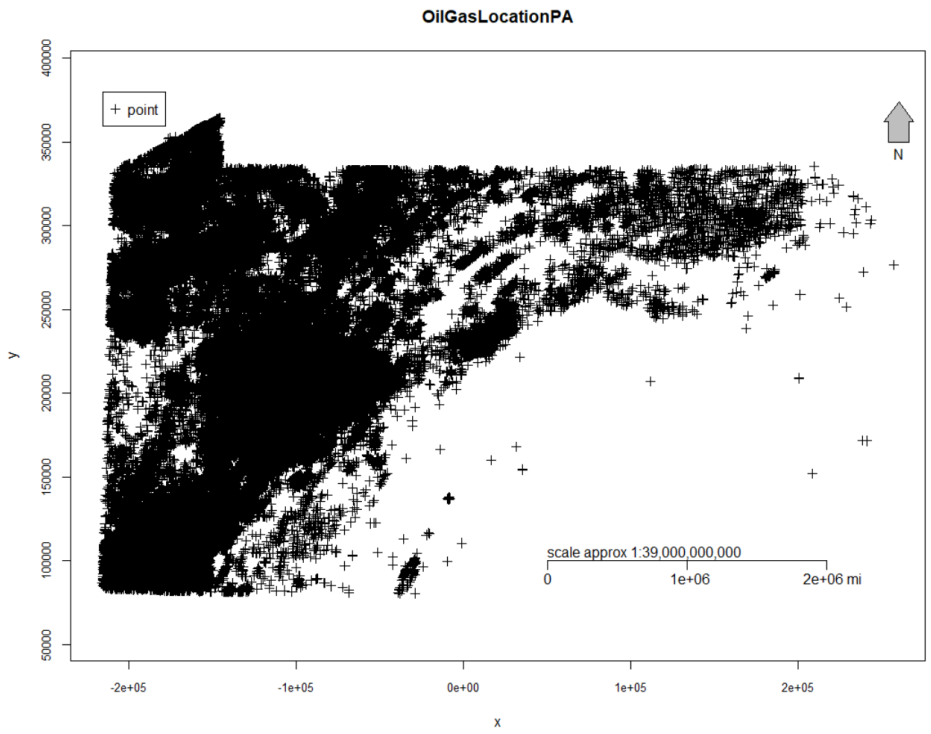
$$I = \left[\frac{n}{\sum_{i=1}^{n}(y_i-\bar{y})^2}\right] \times \left[\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i-\bar{y})(y_j-\bar{y})}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}}\right]$$

Moran's I for the attribute (PM25) is 0.1536012, a positive value, which indicates the data are positively autocorrelated, and most pairs of adjacent locations have values on the same side of the mean. What's more, the index score is less than 0.3, which indicates it is not a relatively strong autocorrelation.

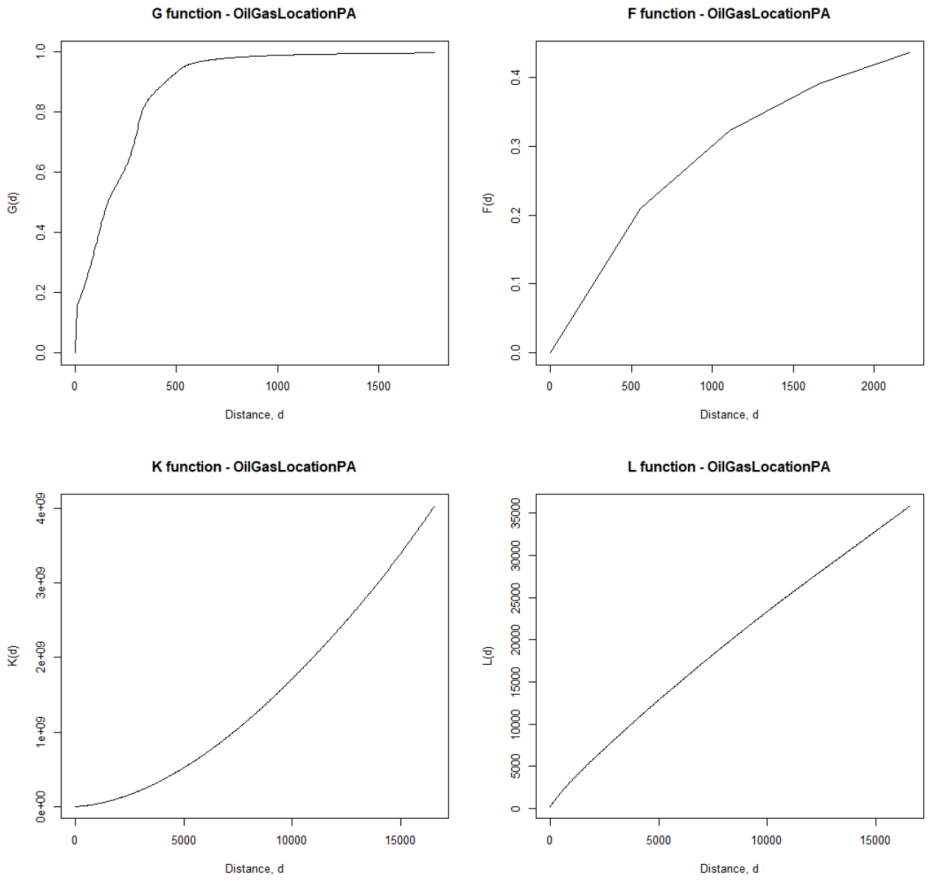
## II. Distance-Based Techniques

a. OilGasLocationPA

The map for the dataset:




$G$、$F$、$K$、$L$ functions:



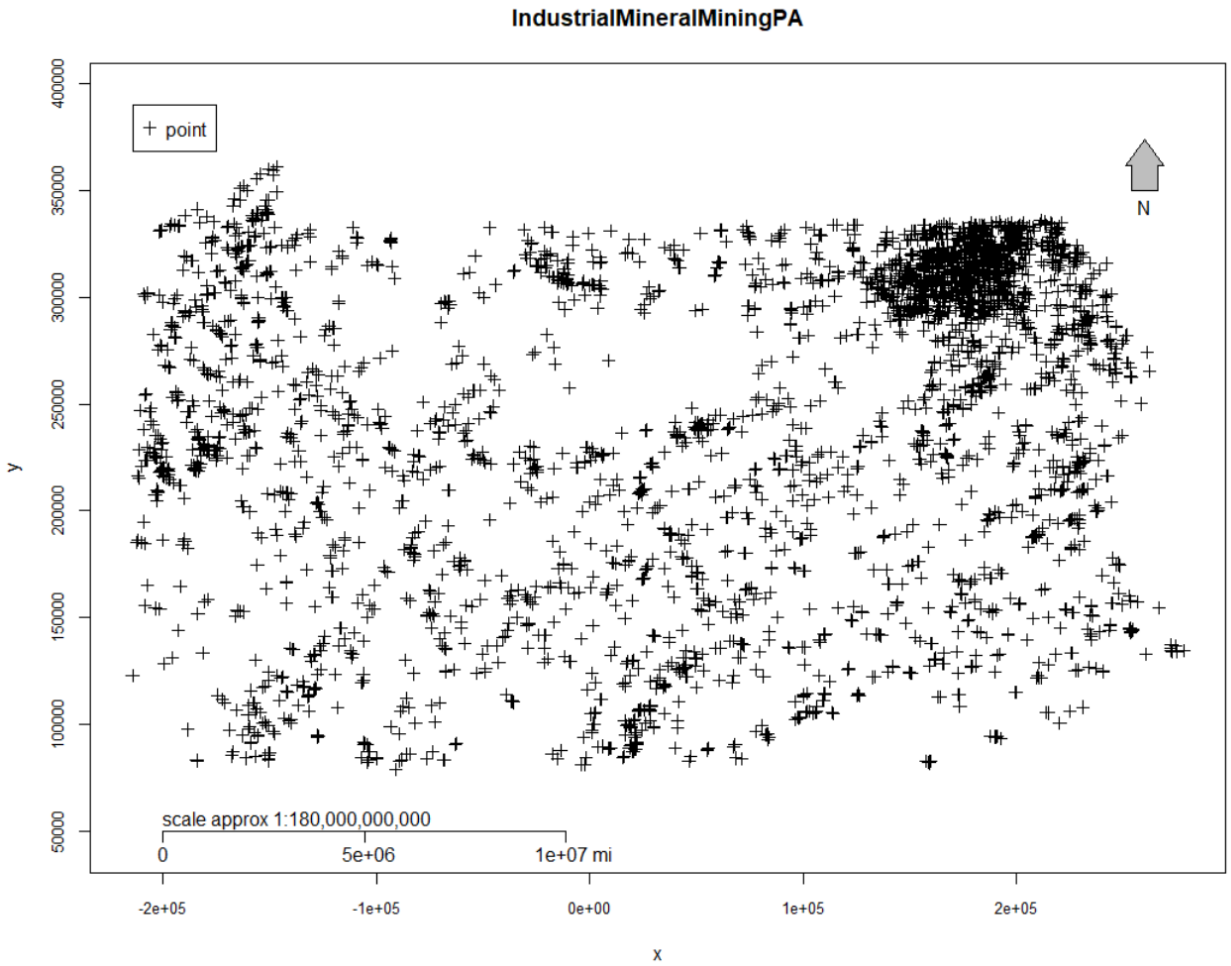
Result:

From the plots, *G* function rises sharply and much faster than *F* function, which indicates that the data set is clustered. In *K* function plot, $K(d) > \pi d^2$, and in *L* function, $L(d) > 0$, both indicate that the data set is clustered.

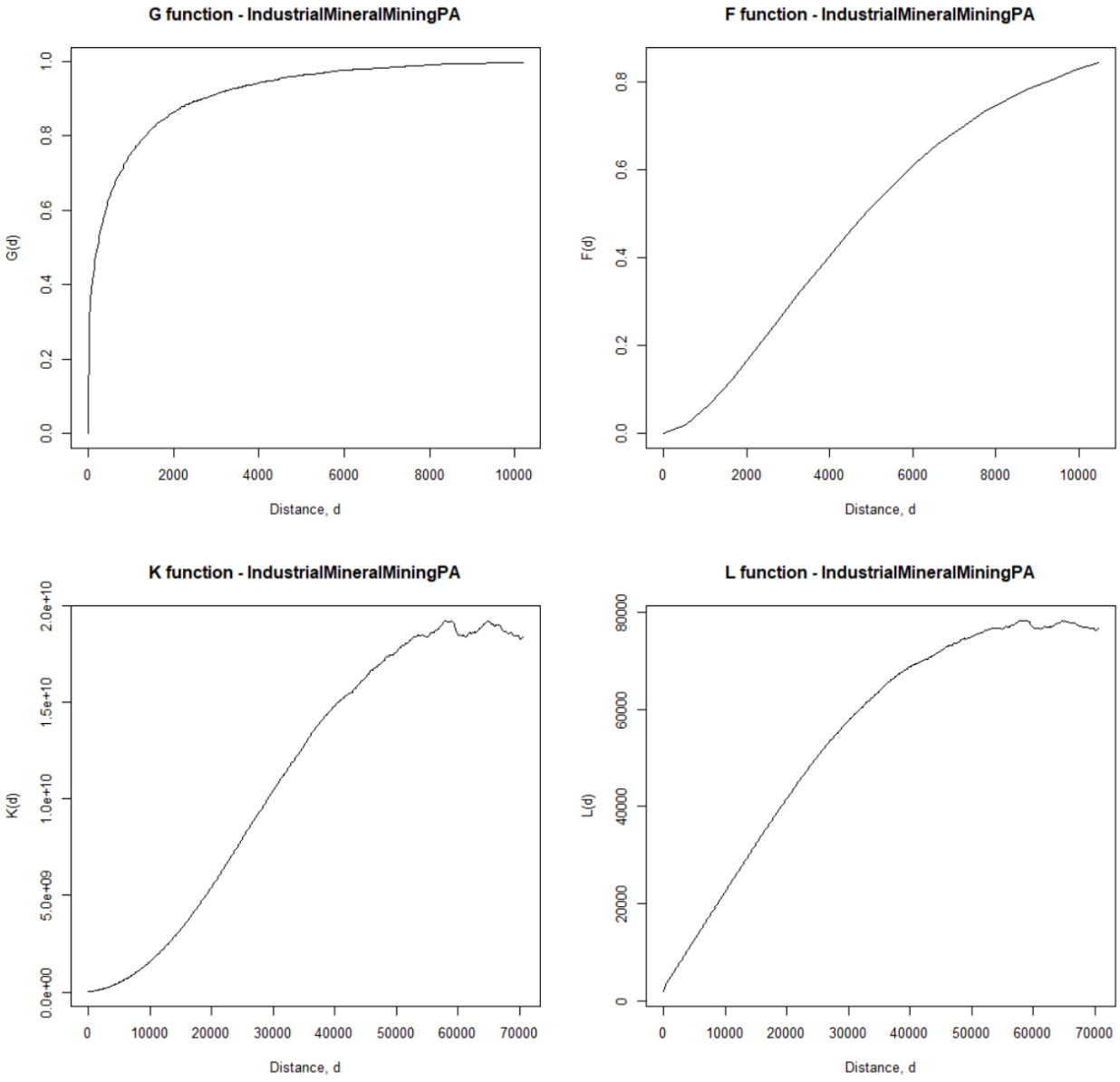
Therefore, the data set is clustered.

b. IndustrialMineralMiningPA

The map for the dataset:



$G$、$F$、$K$、$L$ functions:




Result:

From the plots, $G$ function rises sharply and much faster than $F$ function, which indicates that the data set is clustered. In $K$ function plot, $K(d) > \pi d^2$, and in $L$ function, $L(d) > 0$, both indicate that the data set is clustered.

Therefore, the data set is clustered.

c. Compare the results between the two data sets

From the plots, all $G$、$F$、$K$ and $L$ functions in 'OilGasLocation' rise faster than those in 'IndustrialMineralMiningPA', which indicates the data in 'OilGasLocation' are much more clustered than the data in 'IndustrialMineralMiningPA'.