Submitted by: Charu Sreedharan(chs263)

Project 3

Note: The R file is submitted separately.

Part - 1

Report:

```
Moran's I value is:  0.1536012
```

Weak Positive spatial auto-correlation

Expected value of Moran's I value, E(I) = -1/(N-1) = -1/41 = -0.0243

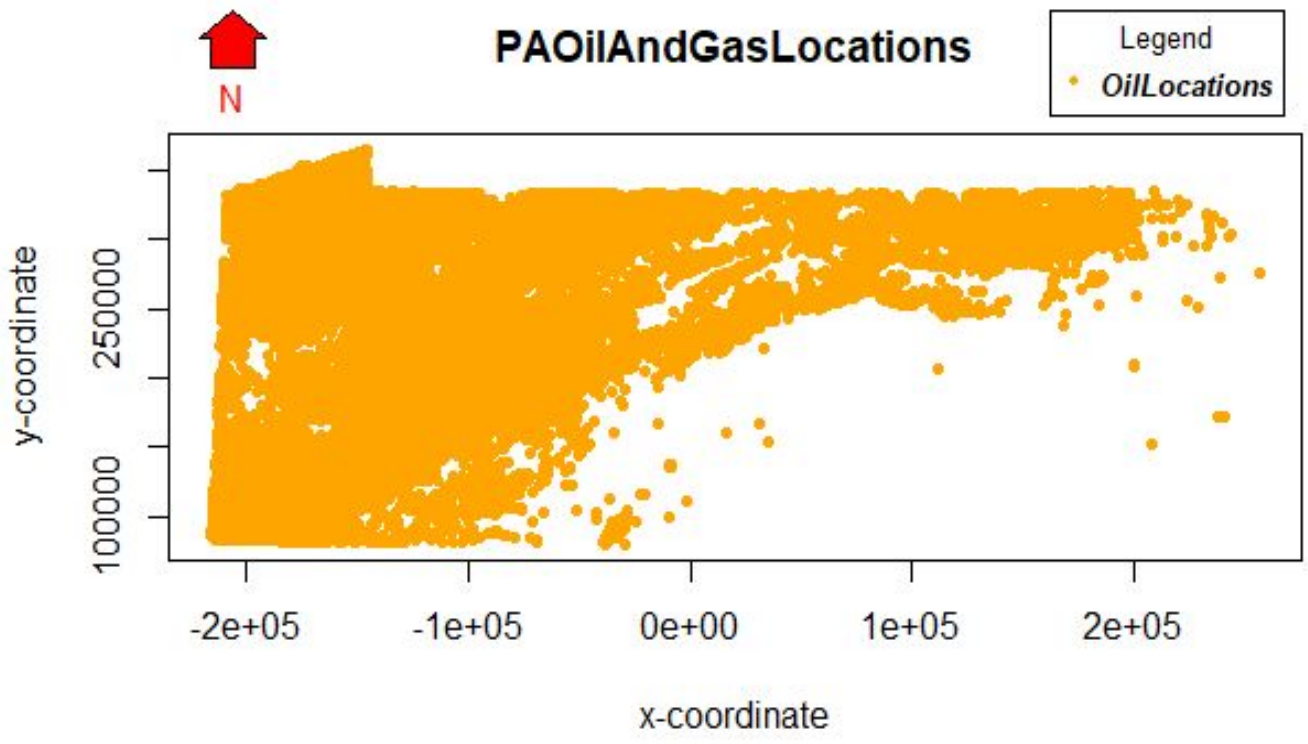
I got a positive spatial auto-correlation of **0.1536012**. Values of Moran's I usually range from −1 to +1. Values significantly below -1/(N-1), where N is the number of events indicate negative spatial autocorrelation and values significantly above -1/(N-1) indicate positive spatial autocorrelation. As the Moran's I value I got of 0.1536 is greater than the E(I) of -0.0243, I can confidently say that the attribute (PM25) exhibits positive spatial auto-correlation. So, it's obvious that the PM25 attribute values from locations near one another in space are more likely to be similar than data from locations remote from one another.

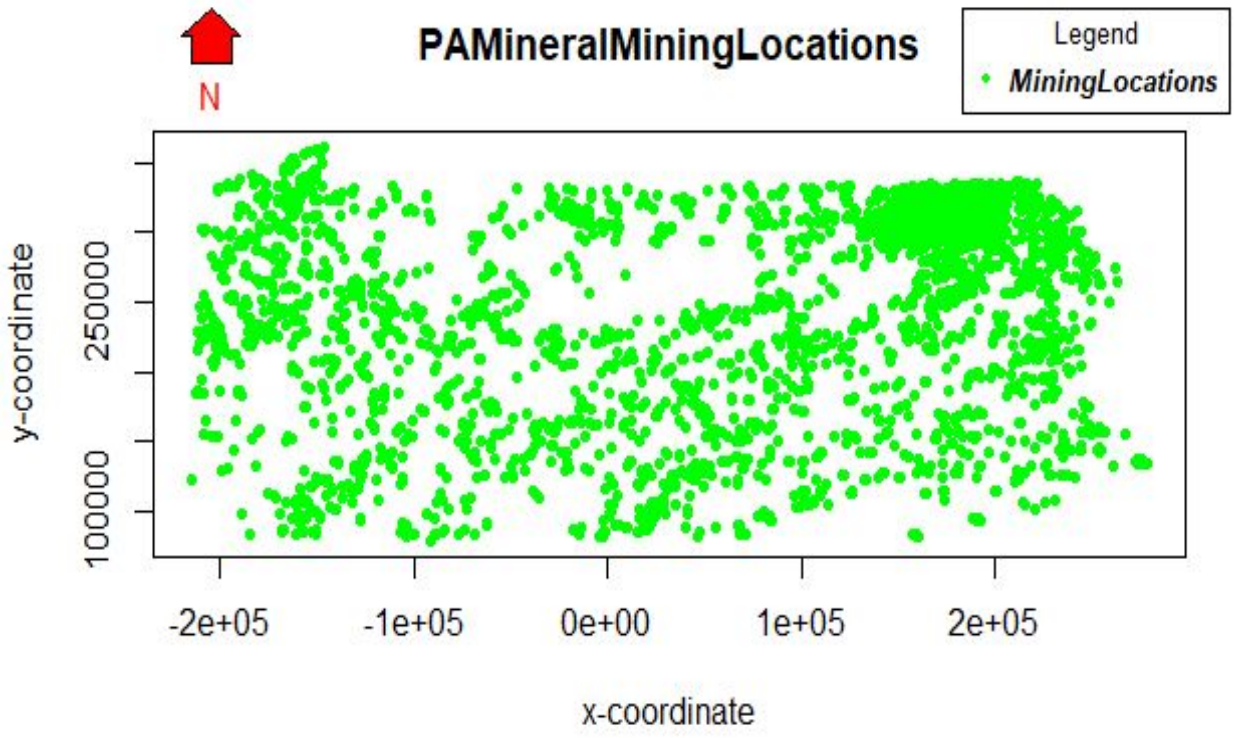
Moran's I has a positive value because most pairs of adjacent locations have values on the same side of the mean. An index score of 0.3 or more, or of 0.3 or less, is an indication of relatively strong autocorrelation. Since our score is <0.3, we can say that the auto-correlation is weak positive.

I also verified this I-value with the value obtained after applying the R library function Moran.I(). The observed I value after applying the Moran.I() library function is: 0.1276045. These 2 are sufficiently close.

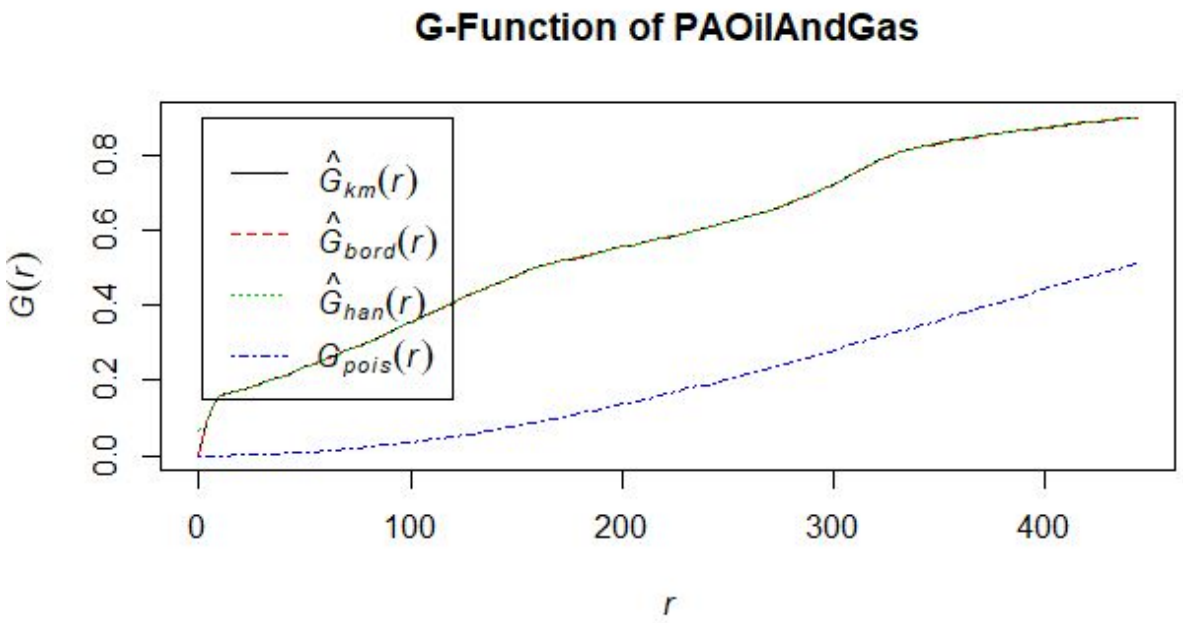
Part - 2

Maps:-

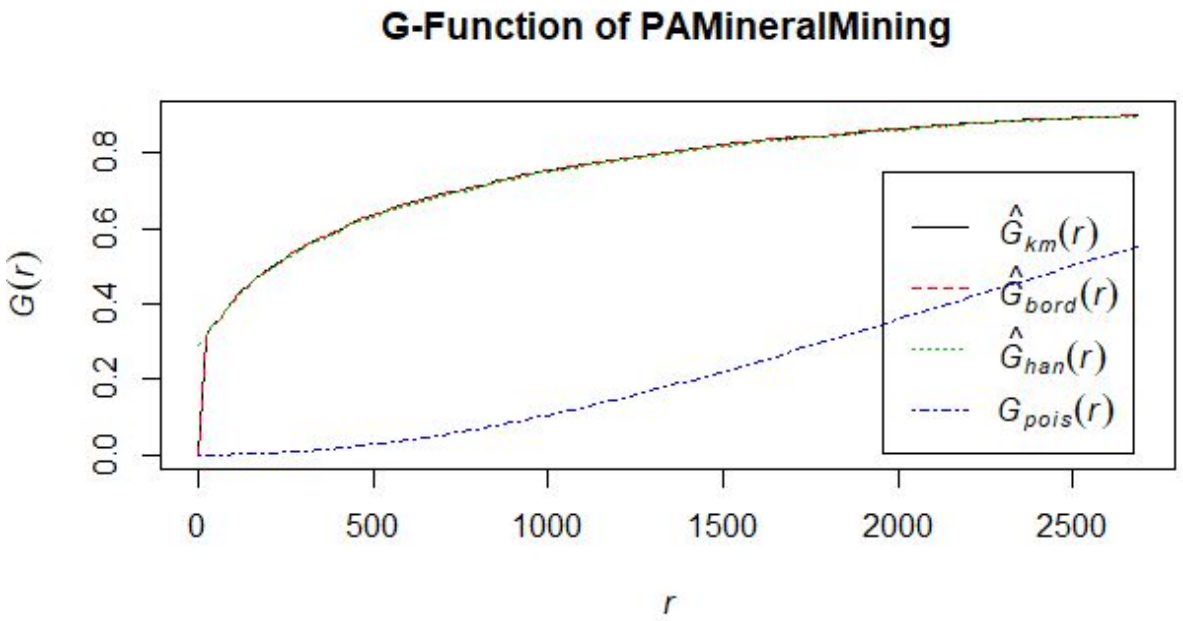



PAMineralMiningLocations
N
Legend
MiningLocations
y-coordinate
250000
100000
-2e+05
-1e+05
0e+00
1e+05
2e+05
x-coordinate

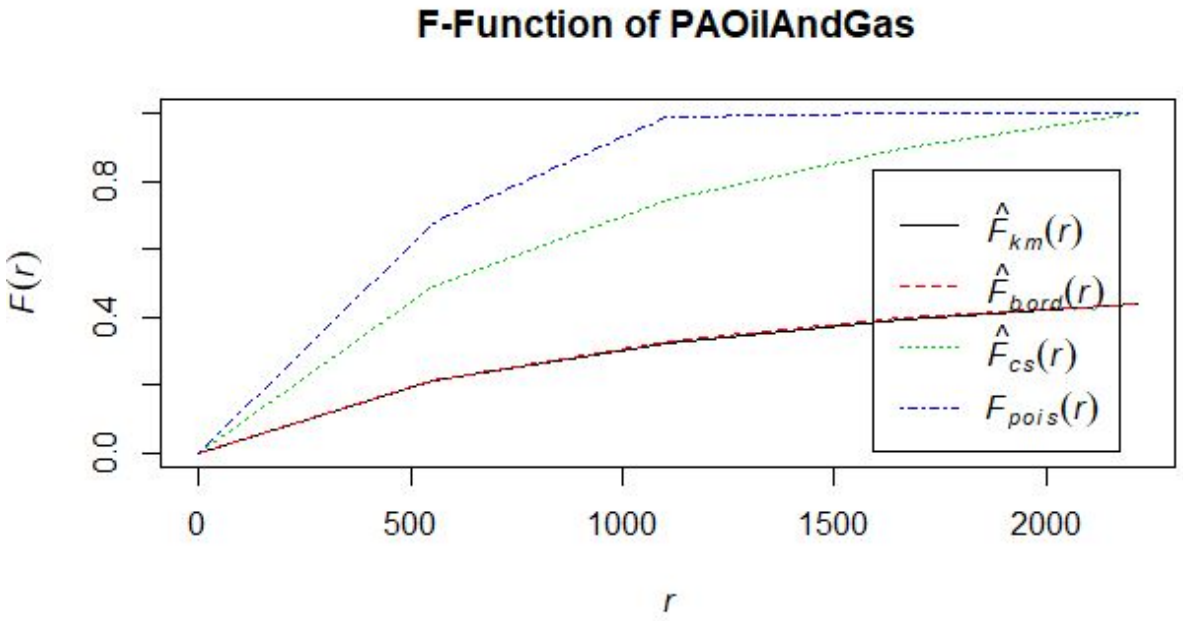
Plots and Report:-



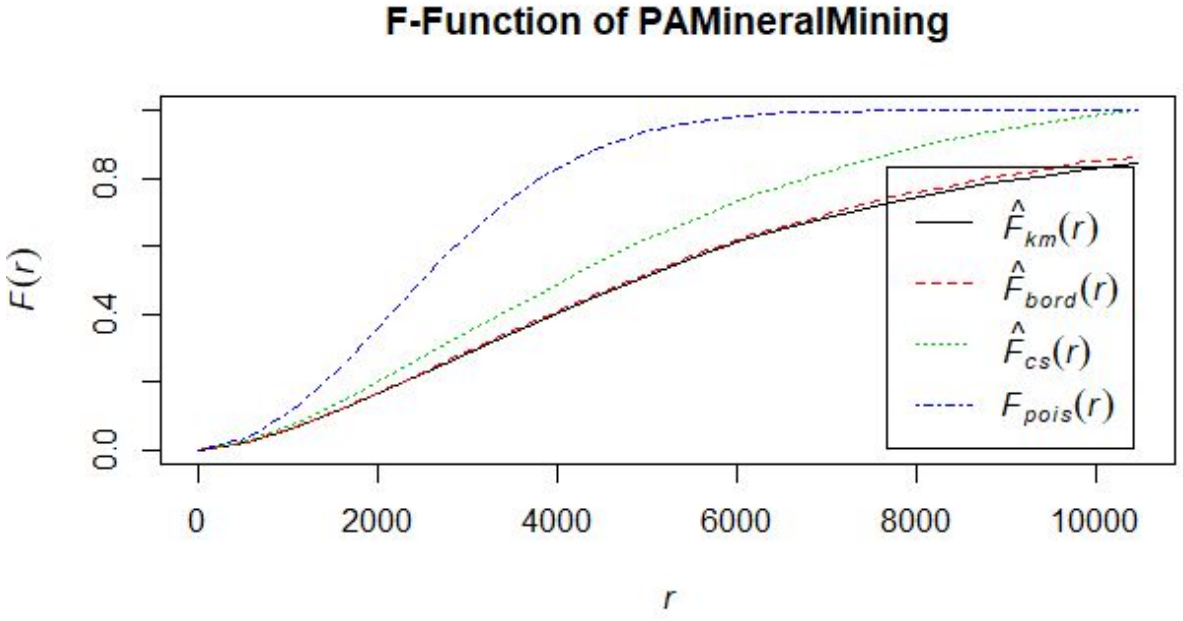
As can be observed, in the case of PAOilAndGas dataset, the G function rises rapidly at short distances and tapers off later. This is indicative of clustering in this dataset, as many events have a very close nearest neighbor. For a clustered pattern, the G function reveals that events in the pattern are closer together than expected.



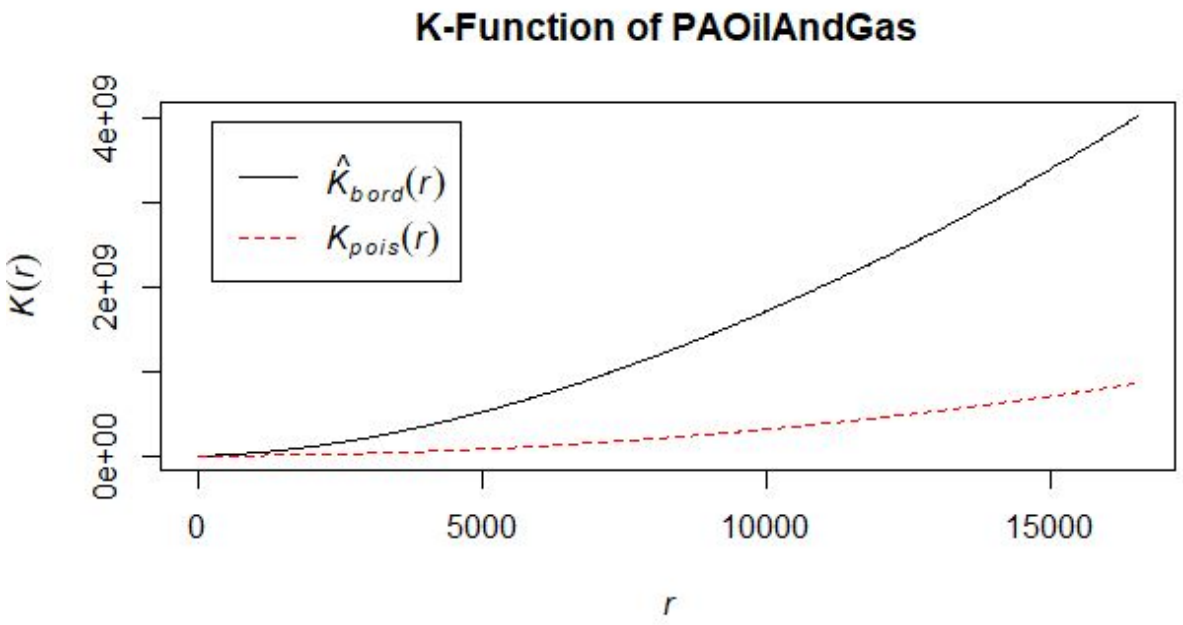
As can be observed, in the case of PAMineralMining dataset, the G function rises rapidly at short distances and tapers off later. This is indicative of clustering in this dataset, as many events have a very close nearest neighbor. For a clustered pattern, the G function reveals that events in the pattern are closer together than expected.



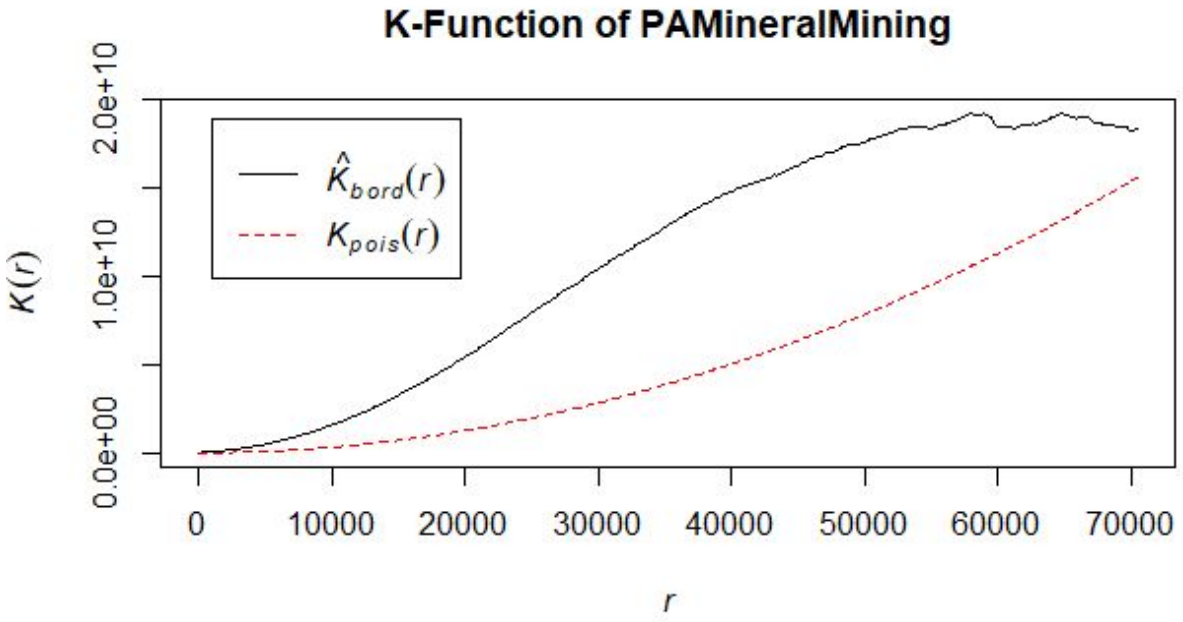
In contrast to the G-function, the F function here rises steadily across a range of distances and rises slowly at first. This is indicative of clustering in this dataset. For a clustered pattern, the F function shows that typical locations in the study region are farther from any event in the pattern than would be expected (because they are empty).



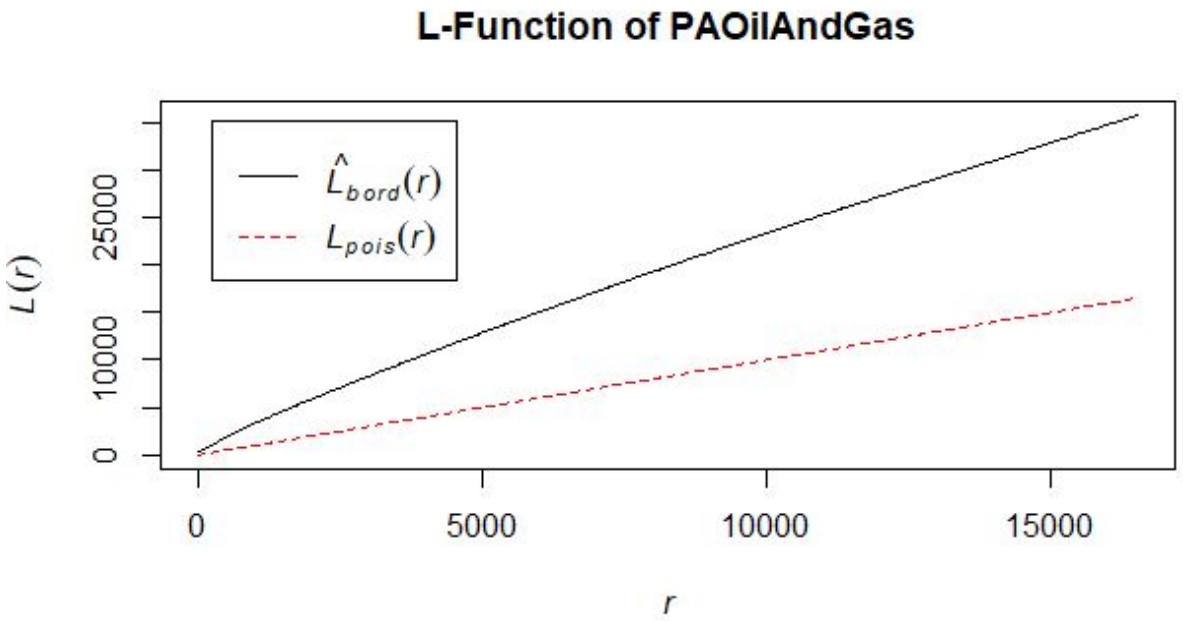
In contrast to the G-function, the F function here rises steadily across a range of distances and rises slowly at first. This is indicative of clustering in this dataset. For a clustered pattern, the F function shows that typical locations in the study region are farther from any event in the pattern than would be expected (because they are empty).



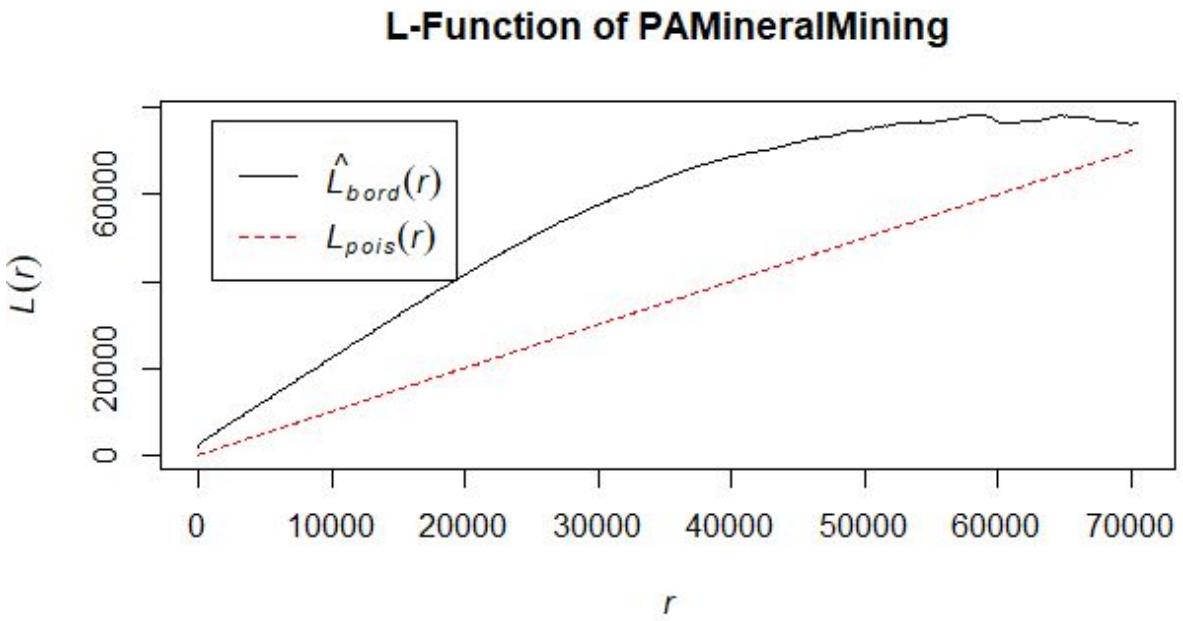
K(r) describes the average number of events inside a circle of radius r centered on an event. If clustered, smaller values of r will result in a large number of events being captured inside the circle of radius r. So, the K function is likely to rise at small distances. This is the case in the above plot. So, we can conclusively say that this data set is clustered.



K(r) describes the average number of events inside a circle of radius r centered on an event. If clustered, smaller values of r will result in a large number of events being captured inside the circle of radius r. So, the K function is likely to rise sharply at small distances and then smoothly over larger distances and gradually taper off. This is the case in the above plot. So, we can conclusively say that this data set is clustered.



When L(r) is above zero, it means that there are more events at the corresponding spacing than would be expected. Where as, if it is below zero, it means that there are fewer events than expected. In this case, L(r) > 0 for r values across the whole range of plotted values, indicating that there are more events at these spacings than expected. This is clearly indicative of clustering in the PAOilAndGas dataset.



When L(r) is above zero, it means that there are more events at the corresponding spacing than would be expected. Where as, if it is below zero, it means that there are fewer events than expected. In this case, L(r) > 0 for r values across the whole range of plotted values, indicating that there are more events at these spacings than expected. L(r) rises sharply at first and then smoothly over larger r and gradually tapers off. This is clearly indicative of clustering in the PAMineralMining dataset.

One failing of the G and F functions is that they only make use of the nearest neighbor for each event or location in a pattern. This can be a major drawback, especially with clustered patterns where nearest-neighbor distances are very short relative to other distances in the pattern, and can mask other structures in the pattern. A relatively simple way around this problem is to make use of K and L functions based on all the distances between events in S. K and L functions provide much more information about a data set than G and F functions.

After a thorough analysis of the above results, we can conclusively say that both PAOilAndGas and PAMineralMining datasets are **clustered**.