Identifying the spatial patterns of housing distribution in Johor Bahru through spatial autocorrelation

N A Mohd Sairi\(^1\), B Burhan\(^2\) and E E Mohd Safian\(^3\)

\(^1,2,3\)Department of Real Estate and Facilities Management, Faculty of Technology Management and Business, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Johor, Malaysia

asyikinsairi@gmail.com\(^1\), burhaida@uthm.edu.my\(^2\), edie@uthm.edu.my\(^3\)

Abstract. Everything that has a location in geographic space will naturally generate a spatial pattern either clustered, dispersed or random. A testable assumption in the concept where geographic location matters is Tobler’s First Law of Geography whereas one means for quantifying the law of geography by Tobler is through the measures of spatial autocorrelation. Hence, this study aims to identify the spatial clustering patterns of housing distribution in Johor Bahru by applying spatial autocorrelation methods. Through global spatial autocorrelation analysis, the results depict a high clustering within the housing distribution as the value of Moran’s I is 0.993852 which is highly positive and near to 1. Next, the LISA cluster map had successfully identified individual clusters where the housing characteristics and location characteristics in Skudai and Pasir Gudang are similar to its nearby housing units. However, the housing units in Johor Bahru have dissimilar characteristics of housing and locations with its nearest housing units. Based on this analysis, although the housing distribution indicates a clustering pattern overall, the type of clustering is however, locally different. Apart from that, the results also reflects buyer’s preferences of housing choices through locational characteristics whilst suggesting for possibility of conducting distance matrix for further analysis.

1. Introduction
Housing is considered as a special commodity possessing different characteristics such as design, neighborhood condition, accessibility and other attributes which affect consumer’s consideration in order to attain maximum utilisations [1]. In addition, housing remains as one of the fundamental growth engines for developing economies as it contributes to urbanization and infrastructure development [2]. Apart from that, houses are typically the largest component of household wealth, the key collateral for bank lending and play a central role for long-run trends in wealth-to-income ratios and the size of the financial sector [3]. Cecchini \textit{et al}. [4] claimed real estate as one of the largest economic activities worldwide influencing the arrangement of settlement as well as shaping built environments. According to Mohamad \textit{et al}. [5], three qualities from a house that are often emphasized by home buyers are the building quality, location of the property and also neighborhood conditions where these three factors could also affect the housing prices. Basically, housing data
which contains location information can be visually map hence providing opportunity for further research of spatial analysis.

According to Scott [6] data associated with locations can be mapped, whereas mapping geographic data is an important first step in analysing spatial patterns. The next steps involve identifying, describing and measuring the spatial characteristics of data [6]. Bivand [7] agreed by stressing that in a case where data are located at geographical coordinates, mapping the data alone is not enough as it seems infelicitous not to examine the possibility of dependence. Furthermore, a unique feature of spatial data is that geographical location provides a key shared either exactly or approximately between data sets of different origins. Similarly, Er et al. [8] also claimed that in dealing with problems of space, the step beyond mapping was spatial analysis which in geographical research was the tool used to compare the spatial distribution of a set of features to a hypothetically-based random spatial distribution.

Klippel et al. [9] stated that everything that has a location in geographic space will naturally create or contributes a spatial pattern either clustered, dispersed or random. Besides, the spatial distributions or patterns, are of interest to many areas of geographic research because they can identify and quantify patterns of features in space so that the underlying cause of the distribution can be determined [8]. Chou [10] had defined the spatial pattern of a distribution as the arrangement of individual entities in space and the geographic relationships among them. Similarly, Laohasiriwong et al. [11] claimed that spatial pattern detection can be a useful tool for understanding the geographical distribution. As a matter of fact, a testable assumption in the concept where geographic location matters is Tobler’s First Law of Geography. In addition, one means of quantifying the law of geography by Tobler is through the measures of spatial autocorrelation.

The application of cluster analysis through spatial autocorrelation has received overflowing attention from scholars of various field of study including tourism, health, crime and urban studies. Through adaption of spatial autocorrelation, researchers are able to identify the spatial patterns of their specific data based on their field of studies. In an attempt to investigate the spatial patterns of household distribution in Tokyo, Monzur [12] had applied both global and local spatial autocorrelation in his study. The analysis used global spatial autocorrelation to identify the distribution type where local spatial autocorrelation were then used to understand the locational spatial patterns.

This study aims to identify the spatial clustering patterns of housing distribution in Johor Bahru area through cluster analysis using spatial autocorrelation. Both global and local spatial autocorrelation were used to achieve the aim of this particular research. An ample explanation on Tobler’s First Law of Geography was also provided along with in depth explanation on the methodology as well as the results and conclusions from this analysis.

2. Tobler’s First Law of Geography

Introduced by Waldo Tobler in 1970, the Tobler’s First Law of Geography (TFL) which stated that: “everything is related to everything else, but near things are more related than distance things” has become central to the core of spatial analytical techniques as well as to geographic conceptions of space [13]. According to Waters [14], the term first law in TFL does not indicate that it was the first law to be declared, instead Tobler was arguing that this was of great significance law in geography and that distance was the most important variable governing the influence of one entity on another. In other words, he was trying to capture something that is most fundamental in geography [15].

As a matter of fact, the concept of TFL is implied in the practice of spatial analysis as near and related are useful concepts at the core of spatial analysis and modeling [13]. Apart from that, TFL states that spatial data are defined by their spatial dependence or spatial autocorrelation [14]. Similarly, Miller [13] stressed that TFL is at the core of spatial autocorrelation statistics, which is quantitative techniques for analyzing correlation relative to distance or connectivity relationships. According to Nunung and Pasaribu [16], spatial autocorrelation refers to the pattern in which observations from nearby locations are more likely to have similar magnitude compared to those from distant locations.
Basically, from the TFL, we can notice that there were two facts highlighted by Tobler. Hence, to provide further understanding of TFL, it can be viewed from those two consecutive parts. The first part of TFL which is “everything is related to everything else” showed the existence of spatial dependence and spatial association. Association can also be defined as relationship, relatedness or interdependence which is what the term “everything is related to everything” implies. In the case of housing distribution in Johor Bahru, we take into account the facts featured by Klippel et al. [9] which is everything that has a location in geographic space will naturally create or contributes to a spatial pattern either clustered, dispersed or random. Therefore, we can test whether those data are really clustered or related as stated in TFL. On the other hand, the second part which is “near things are more related than distance things” suggest the importance of distance as well as location in regional science and geography. In other words, if a data is located nearer to each other, it tends to be more similar compared to a data which is located further from each other.

Chou [10] stressed that spatial autocorrelation which is equivalent to TFL is an important building block of spatial information theories because it represents the most fundamental structure of spatial relationships. Furthermore, Zhang et al. [17] explained that based on TFL, spatial association is deep-rooted in geographic data and analyses based on regular statistics are very likely to be misleading when working on spatial data. Hence, the prevalence of spatial dependence in the data for analysis requires the application of appropriate techniques of spatial statistics and spatial econometrics [18]. In addition, Anselin [18] stressed that methodologically, a proper treatment of space requires the recognition of the importance of the two-dimensional nature of spatial interaction (or spatial autocorrelation) and its implications for statistical analysis.

Therefore, according to TFL, it can be concluded that spatial association existed in geographic data. As the housing distribution in Johor Bahru has geographical location, the existence of spatial association as well as the clustering pattern can be measured by taking into account of TFL.

3. Methodology

3.1. Study Area
Located at the Southern tip of Peninsular Malaysia, Johor Bahru acts as the capital of Johor state. The Johor Bahru area covers 1,765.43 square kilometre and is the most developed district in Johor, either in economic or real estate development activities [19]. In addition, Johor Bahru is a commercially oriented area and give rise to the economy of the region and the nation in general [20].

3.2. Spatial Autocorrelation Analysis
This study applied spatial autocorrelation method to identify the clustering patterns of housing distribution in Johor Bahru. The global spatial autocorrelation was first computed to show the existence of clustering within the data set. Next, local spatial autocorrelation was then computed to show the individual clusters of the housing distribution.

According to Mazzulla and Forciniti [21], spatial association or spatial autocorrelation is the tendency of variables to display some degree of systematic spatial variation. Several measures of spatial association have been developed to examine the nature and extent of spatial dependence in spatial data. The global measures of spatial autocorrelation uses the complete data set to derive a single value for the entire study region as this measures emphasize average or typical characteristics of the complete data set. On the other hand, the local measures of spatial association examine spatial dependence in subsets (variously known as neighborhoods, windows or kernels) defined with respect to each data site in the complete data set. Such measures focus on the identification of variations, rather than regularities, in the nature of spatial association within the study region [22].

Fotheringham [23] had mentioned that spatial dependence has led to a large body of research into spatial autocorrelation. According to Wang et al. [24], spatial dependence exists when statistical values are correlated. Spatial autocorrelation signifies the interdependence of values of a variable at various geographic locations and measure the nature and strength of that interdependence [25].
other words, spatial autocorrelation is the dependence of a given variable’s values on the values of the same variable recorded at neighboring locations [26]. Apart from that, spatial autocorrelation is also used to describe spatial dependence through statistics such as Moran’s Index [27].

Griffith and Chun [28] emphasized that spatial autocorrelation exists because geography matters. The concept of spatial autocorrelation was successfully popularized by Cliff and Ord in 1967 [28]. Spatial autocorrelation is one component of pattern where it is limited to the clustering or dispersion of objects rather than measuring the geometric aspects of pattern [22]. There is a distinction that is made in all spatial autocorrelation measures using various spatial analysis methods [9]. This differentiation is among three kinds of patterns which are clustered, dispersed and random. Furthermore, spatial autocorrelation concept helps in analyzing the patterns, where it measures the relationship among values of a variable according to the spatial arrangements of the values and it finds whether the data is clustered, random or dispersed based on the similarity of the values and their spatial proximity [29]. Apart from that, spatial autocorrelation also indicates the extent to which the occurrence of one feature is influenced by similar features in the adjacent area [10].

Clusters occur in a geographic distribution when features are found in close proximity or when groups of features with similarly high or low values are found together (hot spots and cold spots) [30]. According to Griffith and Chun [28], spatial autocorrelation perspective focuses on the clustering of similar or dissimilar rather than random mixtures of phenomena in geographic space to form map patterns or the orderness of geographically distributed phenomena typified by TFL. In fact, spatial data are tend to be highly self-correlated [31]. Getis [32] agreed by stating that a single variable may be spatially autocorrelated that is, values of the variable are somehow connected or related spatially. Apart from that, Tsai et al. [33] defined spatial autocorrelation as the relation between the values of a single variable where this relation is caused by the geographic arrangement of areal units on a map. As a matter of fact, spatial autocorrelation can be used to identify the degree of spatial clustering. According to Wang et al. [24], the indicators for calculating spatial autocorrelation can be classified into two groups which are global spatial autocorrelation indicators and local spatial autocorrelation indicators.

Based on the previous discussions provided, in simpler explanation, it can be concluded that spatial autocorrelation or spatial association is used to measure the presence of spatial dependence within a spatial data set whereas cluster is the results of the occurrence of spatial dependence. As this analysis will conduct spatial autocorrelation analysis from the Johor Bahru housing distribution based on both global and local indicators, further discussions on both levels of spatial association will be provided in the sub subsection below.

3.2.1. Global Spatial Autocorrelation. Moran’s I which was developed by P. A. P. Moran in 1948 is a widely used measure of spatial autocorrelation [24, 25, 34]. Similarly, Ying [35] pointed out that Moran’s I is probably the best known test for spatial autocorrelation among most geographers and regional scientists. In technical terms, Moran’s I is a measure of global spatial autocorrelation which is an indication of whether similar values of a particular variable are closer together in space thus detecting the presence of the clustering of similar values [29, 36]. A positive coefficient indicates that nearby areas have similar values, whereas a negative coefficient reveals a dissimilar pattern of nearby attributes [35, 37]. Wang et al. [24] claimed that global Moran’s I is capable to explain the overall distribution of specific phenomena and whether the phenomena present clustering characteristics in a given space. In addition, Ying [35] explained that Moran’s I is also used to calculate the degree of dispersion or concentration of the features or spatial data.

According to Er et al. [8], Moran’s I measures spatial autocorrelation based on both feature location and feature values simultaneously. In addition, this analysis evaluates the spatial feature distribution pattern through the simultaneous observance of the location and characteristics of the feature [30]. Numerically, Moran’s I can be expressed as below [25]:

(1)
\[ I = \sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij} (x_i - \bar{x}) (x_j - \bar{x}) \]

\[ s^2 = \sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij} \]

where:
- \( n \) = the number of observation locations/sites
- \( x_i \) = the value at a point \( i \)
- \( x_j \) = the value at point \( i \)’s neighbor \( j \)
- \( \bar{x} \) = the mean of observed values in all the sites
- \( s^2 \) = the variance of \( x \) value
- \( w_{ij} \) = the weight, defined based on the spatial proximity between locations \( i \) and \( j \)

The value of Moran’s I ranges from -1 to 1. The Moran’s I is positive when the observed values of locations within a certain distance or their contiguous locations tend to be similar or in other words, clustering of similar values [38]. It is negative when they tend to be dissimilar (checkered pattern), and approximately zero when the observed values are arranged randomly and independently over space [37]. In other words, the negative values indicate negative spatial autocorrelations and positive values indicate positive spatial autocorrelations while a zero value indicates the existence of a random spatial pattern [34]. Valcu and Kempenaers [26] explained that spatial autocorrelation is positive when high values are associated with relatively high values at neighboring locations and oppositely, spatial autocorrelation is negative when high values correspond to relatively low values at neighboring locations.

Spatial autocorrelation may also be summarized using a Moran scatter plot which provides more information on the type of spatial autocorrelation [39]. This scatter plot, suggested by Anselin, plots the variable of interest on the horizontal axis against a spatial lag (the standardized spatial weighted average) on the vertical axis. According to Eck and Chainey [40], a spatially lagged variable or spatial lag is derived as the spatially weighted average of its neighboring values and act as an important part of the computation of spatial autocorrelation tests. Furthermore, the Moran scatter plot is presented by four different quadrants corresponding to the four types of local spatial association between each delegation and its neighbors [39].

Unfortunately, the global tests of autocorrelation are unable to specifically indicate clustering of high or low within the data set [24, 41]. As stressed by Boots [22], the global approaches yield a single value for the entire data set, while local approaches are capable of generating a local value for each data site in the data set. Apart from that, Zhang et al. [17] had suggested that it is advisable for us to take into consideration that the spatial autocorrelation level of different census area is not exactly the same hence, initiated the need for local indicator. Therefore, in order to identify the individual locations of the clustering within the whole data set, local Moran’s I statistics must be conducted.

3.2.2. Local Spatial Autocorrelation. In contrast to global Moran’s I which assumes homogeneity of the whole dataset, local Moran’s I, on the other hand is a local indicator of spatial association and shows the level of spatial autocorrelation at various individual locations within the data set [37]. The local Moran statistics which also referred to as Local Indicator of Spatial Association (LISA), are more often used to measure the local spatial concentration or clustering [39]. Local Moran’s I is therefore a decomposition of global Moran’s I where conversely, global Moran’s I is a summation of individual cross-products of local Moran’s I [42]. Similarly, Mazzulla and Forciniti [21] stated that LISA allows for the decomposition of global indicators, such as Moran’s I into the contribution of each individual observation.

The core of defining LISA as famously cited by many researchers are those by Anselin [43] where he operationally defined LISA as any statistic that satisfies two requirements. The first requirement is the LISA for each observation gives an indication of the extent of significant spatial clustering of similar values around that observation. As the name LISA itself is the local indicator of spatial
association, it should provide the indicators of the clustering or association. Apart from that, results from LISA enables the identification of locations with its neighbor as HH, LL, HL and LH which will be further explain in Table 1. The second requirement is the sum of LISAs for all observations is proportional to a global indicator of spatial association. As stated previously that LISA allows the decomposition of global indicators, hence, it is not questionable that the sum of LISA is proportional to the global indicator. In addition, the Moran’s I value for both local and global from the same dataset are usually of the same values.

The outcome of LISA differentiates between a statistically significant cluster of high values, cluster of low values, outlier in which a high value is bounded mainly by low values, and outlier in which a low value is encircled primarily by high values [25]. As a matter of fact, LISA are able to detect places with unusual concentrations of high or low values to be analyzed as hot or cold spots [11, 21]. Chaikaew et al. [44] described hotspot as a condition indicating some form of clustering in a spatial distribution while Rothlisberger et al. [45] characterized hotspot as an extreme form of spatial autocorrelation.

According to Zhang et al. [46], local Moran I does not range between -1 and +1. However, a positive value still implies positive spatial autocorrelation (clusters) and a negative value indicates negative spatial autocorrelation (outliers). According to Bhunia and Shit [25], the Local Moran’s I statistics of spatial association is computed as:

$$L_i = Z_i \sum W_{ij} Z_j$$

where:

- $Z_i$ = the association of values of a random variable at site $i$
- $W_{ij}$ = the weight, defined based on the spatial proximity between locations $i$ and $j$
- $Z_j$ = the association of values of a random variable at site $j$

Basically, the LISA cluster map classify those locations by the type of association either high values with high values, low values with low values or high value with low values [47]. According to Anselin [42], Franczyk and Chang [48] and Laohasiriwong et al. [11], the results from LISA cluster maps will specify information as shown in table 1 below:

| Color Schemes | Spatial Association | Details |
|---------------|---------------------|---------|
| Red           | High-high (H,H)     | Locations with high values of the phenomenon surrounded by neighbors with high value (clustering of similar value) |
| Blue          | Low-low (L,L)       | Locations with low values of the phenomenon surrounded by neighbors with a low values (clustering of similar value) |
| Light Blue    | Low-high (L,H)      | Locations with low values of the phenomenon surrounded by neighbors with a high value (potential spatial outliers) |
| Pink          | High-low (H,L)      | Locations with high values of the phenomenon surrounded by neighbors with a low value (potential spatial outliers) |
| White         | Random patterns     | Not significant |
Both global and local Moran’s I aimed at measuring the spatial autocorrelation. However, the most notable difference between both measures as pointed out by most scholars, is the way the spatial autocorrelation were computed. As global Moran’s I refers dataset as a whole while suggesting homogeneity, it failed to identify individual clustering hence calls for the needs of local measures as it is able to identify spatial autocorrelation at local level as well as identifying high and low clustering and also spatial outliers. Apart from that, by referring to Anselin et al. [47], the distinction they made between both measures were global Moran’s I is the test for clustering in the data set whereas local Moran’s I is design to identify the locations of the cluster. Therefore, in other to overcome the weakness of global Moran’s I, further analysis of local Moran’s I must be conducted as to obtain deeper and wider results.

4. Results and Discussion
The analysis involved measuring the degree of spatial autocorrelation from the simple housing distribution in Johor Bahru. The data entry contains a number of 393 043 housing data. The housing distribution which is the arrangement of housing in space can be further explored by taking into account of TFL. Through TFL, it can be concluded that nearest things are more related compared to distance things. Technically, spatial autocorrelation is the measure of similarity or correlation between nearby observations. Since both TFL and spatial autocorrelation emphasize on nearest distance, hence, spatial autocorrelation method is chosen to show the clustering patterns in this case study.

4.1. Global Moran’s I Scatter Plot
First of all, a global autocorrelation measures using global Moran’s I was applied to test for homogeneity and the existence of clustering within the whole housing data set. In other words, Moran’s I statistics is used to test whether the housing characteristics and location characteristics within Johor Bahru area are spatially correlated or randomly distributed whilst testing for clustering within the data set as a whole. Figure 1 below shows the result of the analysis through a Moran’s I scatter plot:

![Figure 1. Global Moran’s I Scatter Plot of Housing Distribution in Johor Bahru.](image-url)
For this analysis, the housing data was chosen as the variable of interest because we want to see whether the housing distribution mapped previously are dependent and clustered with each other or otherwise. Based on the analysis conducted through GeoDa software, the result from Moran’s I scatter plot showed a clustering result. The highly positive Moran’s I value of 0.993852 which is also near to the value 1, had proven the existence of clustering among the housing units in Johor Bahru area hence indicating a strong clustered pattern of housing units in Johor Bahru.

Even though global Moran’s I has successfully proven the clustering among Johor Bahru housing data, it fails to specifically identify whether the housing characteristics and location characteristics were clustered with similar values or dissimilar values with its neighbors. As global Moran’s I can suggest clustering without designating particular location as clustered, the next step is to identify the individual location of the clusters through Local Indicator of Spatial Association (LISA) whilst detecting the types of association between the structural characteristics and spatial characteristics of housing in Johor Bahru.

4.2. LISA Cluster Map
Based on the LISA cluster map portrayed in figure 2 below, the results showed specific locations with different colors of either red, blue, light blue, pink, grey or white. Basically, the red colored regions denote High-High (HH) value while the blue regions denotes Low-Low (LL) value where both represent spatial clustering patterns. On the other hand, the Low-High (LH) regions represented by a light blue and a High-Low (HL) regions colored with pink which are less noticeable in figure 2 below are the potential spatial outliers whereas the grey and white regions represent neighborless and insignificant data respectively.

![LISA Cluster Map of Housing Distribution in Johor Bahru.](image.png)

Majority of the red colored area can be found in Skudai and Pasir Gudang with lesser red areas in Iskandar Puteri. The red colored regions signified the clustering of similar values. From the map, it suggests that in Skudai, Pasir Gudang and few in Iskandar Puteri, the similar housing characteristics with its neighbors are being grouped together with similar characteristics of location with its neighbors.
hence forming a clustering patterns of similar characteristics. Apart from that, the analysis identified locations of about 67 164 housing of similar characteristics with its neighbors being clustered with similar location characteristics with its neighbors (HH).

Next, the clustering of dissimilar values which are shown through the blue colored regions can be seen mostly in the city center of Johor Bahru and lesser in Iskandar Puteri. Hence, it suggests that in the city center of Johor Bahru and few in Iskandar Puteri, the dissimilar housing characteristics with its neighbors are being clustered together with the dissimilar location characteristics with its neighbors which later portrays a clustering patterns of dissimilar characteristics. The analysis point out 62 176 locations where housing of dissimilar characteristics with its neighbors being clustered with dissimilar location characteristics with its neighbors (LL). It is important to note that the dissimilar location characteristics mentioned before does not mean that the housing are from different locations, rather, it point out the dissimilarity of the spatial characteristics of the location itself.

From the global Moran’s I, it can be concluded that the Johor Bahru housing data is spatially correlated with their neighbors as the Moran’s I indicates a value of 0.993852 which is a highly positive spatial autocorrelation as it is near to 1 hence confirming clustering of housing distribution in Johor Bahru. In addition, the LISA map had successfully portrayed the clustering patterns of housing distribution in Johor Bahru area as well as identifying the locations of the cluster. The results had led to an assumption that in Skudai and Pasir Gudang, the housing units share similar housing and locational characteristics with its nearest housing units. On the other hand, most of the housing units in Johor Bahru are potentially having different housing characteristics and locational characteristics with each other although they are located nearer to each other. The housing characteristics may include building quality, space arrangement and physical characteristics such as land size, building scale and housing age as well as attributes like having a garage as well as the number of bedrooms and bathrooms [49].

Apart from showing the clustering patterns of housing characteristics and location characteristics in Johor Bahru area, the results from LISA map also gives insight towards the buying preferences of each home owners through clustering of similar and dissimilar location characteristics. In Skudai and Pasir Gudang, which is the clustering of similar location characteristics, it suggest the existence of a hub or center which causes them to be clustered together. This fact instigate the possibility of distance matrix analysis to discover the hub or center in Skudai and Pasir Gudang that causes them to be clustered together and forming the red colored region in the LISA map.

Location characteristics of a residential community is connected with the accessibility to the targeted place which is related to the evaluation of transportation accessibility [1]. Based on the traditional view of location, accessibility is measured in terms of access to the Central Business District (CBD) [50]. In addition, location and accessibility plays a decisive role towards household choice of a house [51]. Taking into account of the information given, the existence of Johor Bahru area which is the city center within the blue colored region urge a challenge in this analysis as the clustering of location characteristics in Johor Bahru area are of dissimilar value hence suggesting that the CBD does not contribute to home owners buying preferences. Therefore, further analysis should be conducted to further explore the hub that affects the home buyer’s preferences in Johor Bahru city center.

Even though the results had led to the identification of locations with clustering of similar housing and location characteristics as well as clustering of dissimilar housing and location characteristics, it is important to take into account that the data used to compute Moran’s I in this study does not have the specific housing and location characteristics of each housing units in Johor Bahru. This has caused for failure to provide an accurate statement or elaboration of which specific housing characteristics are similar or dissimilar to its neighbor. The same goes to the location characteristics of the housing units. Therefore, further identification of both structural and spatial attributes need to be conducted. Apart from that, the clustering of similar and dissimilar location characteristics reflects the home owner’s preferences of buying the house in terms of location wise hence, stimulates for the possibility of distance matrix analysis to discover the center or hub that brings them to be clustered together.
5. Conclusion
The motivation to conduct spatial autocorrelation analysis from the simple housing distribution is by taking into account of Tobler’s First Law of Geography (TFL), which claims that “everything is related to everything else but near things are more related than distance things”. As housing distribution contains location of the housing and location itself is in fact a geography matter, hence, justifying the relationship between housing distribution and TFL. Apart from that, TFL indicates the existence of spatial dependence within a data set through relatedness and nearness. Therefore, there’s a need to test for the presence of spatial dependence as well as clustering among the Johor Bahru housing distribution. By using the most widely used measure of spatial autocorrelation which is Moran’s I, the results had shown a clustered pattern within the spatial distribution of housing in Johor Bahru area, hence, proven the existence of spatial dependence in the housing data set.

Unfortunately, the global tests of autocorrelation do not indicate the clusters of high or low values [41]. By taking into account of the information, a local test of spatial association which is known as LISA was conducted to identify individual clustering of hot or cold clusters in the Johor Bahru housing distribution. Through LISA map, the High-High, Low-Low, High-Low and Low-High of the housing characteristics with its neighbors as well as the housing locations with its neighbors in Johor Bahru were successfully identified. As the global Moran’s I aided in identifying the clustering of Johor Bahru housing data set, LISA analysis on the other hand, assist in identifying individual clusters.

Apart from successfully identifying the individual clustering, LISA also managed to identify the spatial outliers. Fu et al. [52] described spatial outliers as an obvious difference of the target value from the values of its surrounding locations or neighborhood. In Johor Bahru housing distribution, we manage to identify a number of six outliers where three of them consist of the dissimilar of housing characteristics with its neighbors were clustered with the highly similarity of housing locations with its neighbors (LH), while another three denotes (HL) where the high similarity of housing characteristics with its neighbors were grouped with the dissimilar of housing characteristics with its neighbors.

The first step in spatial autocorrelation analysis is to construct a spatial weight files which contains information on the neighborhood structure for each location [42]. According to Barreca et al. [53], generating a standardized weight matrix is important for the computation of spatial dependence statistics. In GeoDa software, users can identify the neighbors either through contiguity weight or distance weight. Contiguity weight offers two choices namely rook contiguity or queen contiguity. The issue is, the choices of the spatial weight matrix will affect the result of spatial autocorrelation. Although using the same data set but with different spatial weight, the results will convey slightly different results of spatial autocorrelation. Hence, before computing the spatial autocorrelation, it is vital to have significant justification on choosing the types of weight matrix for a genuine result.

In a nutshell, this paper had highlighted how both TFL and spatial autocorrelation emphasize on nearest distance which later used for analyzing the housing distribution theoretically through TFL and statistically through spatial autocorrelation. The results confirm a strong clustering value among housing data in Johor Bahru area through Moran’s I scatterplot as well as mapping the specific location of clusters through LISA map. Based on this analysis, although it denotes a clustering pattern overall, the type of clustering is however, locally different. The housing characteristics and location characteristics in Skudai and Pasir Gudang are similar to its nearby housing units. In contrast, the housing units in Johor Bahru have dissimilar characteristics of housing and locations with its nearest housing units. In addition, based on location characteristics, the results reflects that the residents in Skudai and Pasir Gudang share the same location preferences whilst Johor Bahru area have dissimilar locational preferences while buying the residential unit. Both results had motivated for further analysis on discovering the center or hub that brings the housing units clustered together with similar or dissimilar values of location characteristics through distance matrix analysis.
References

[1] Tam V W Y, Fung I W H, Wang J and Ma M 2019 International Journal of Construction Management 1-20
[2] Nor M I, Masron T A and Gedi S Y 2019 International Journal of Housing Markets and Analysis 1-26
[3] Knoll K, Schularick M and Steger T 2017 The American Economic Review 107 331-353
[4] Cecchini M, Zambon I and Salvati L 2019 Sustainability 11 1-23
[5] Mohamad M H, Nawawi A H and Sipan I 2016 Procedia - Social and Behavioral Sciences 234 452-460
[6] Scott L M 2015 International Encyclopedia of the Social & Behavioral Sciences 23 178-184
[7] Bivand R 1998 A Review of Spatial Statistical Techniques for Location Studies CEPR Symp on New Issues in Trade and Location (Lund: Sweden) pp 1-21
[8] Er A C, Rosli M H, Asmahani A, Mohamad Naim M R and Harsuzilawati M 2010 International Journal of Geological and Environmental Engineering 4 251-255
[9] Klippel A, Hardisty F and Li R 2011 Annals of the Association of American Geographers 101 1011-31
[10] Chou Y H 1995 Springer 988 365-376
[11] Laohasiriwong W, Puttanapong N & Singsalasang A 2018 Geospatial Health, 13 20-27
[12] Monzur T 2014 ESDA Techniques in Identifying the Spatial Structure of the Tokyo Metropolitan Area (Master’s Thesis: Ritsumeikan Asia Pacific University)
[13] Miller H J 2004 Annals of the Association of American Geographers 94 284-289
[14] Waters N 2018 The International Encyclopedia of Geography 1-15
[15] Sui D Z 2004 Annals of the Association of American Geographers 94 269-277
[16] Nunnung and Pasaribu U S 2006 Identifying Spatial Pattern Using Spatial Autocorrelation Int. Conf. on Mathematics and Natural Sciences (Bandung:Indonesia) pp 760-764
[17] Zhang D, Mao X and Meng L 2010 The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 38 273-278
[18] Anselin L 1998 Journal of Housing Research 9 113-133
[19] Bujang A A, Zarin H A and Agus M R Urban Housing Ownership: Factors Influenced the Problems Faced by the Bumiputera in the District of Johor Bahru, Johor, Malaysia Int. Real Estate Research Symp. (PWTC Kuala Lumpur, Malaysia) 2008 pp 1-18
[20] Abdulaziz H, Shuaibu A and Abdulaziz M A 2018 International Journal of Science, Environment and Technology 7 382 – 396
[21] Mazzulla G and Forcinitii C 2012 European Transport Research Review 4 217-233
[22] Boots B 2003 Journal of Geographical Systems 5 139-160
[23] Fotheringham A S 2009 Geographical Analysis 41 398-403
[24] Wang W C, Chang Y J and Wang H C 2019 International Journal of Geo-Information 8 1-20
[25] Bhunia G S and Shit P K 2019 Geospatial Analysis of Public Health 99-138
[26] Veleu M and Kempenaers B 2010 Behavioral Ecology 21 902-905
[27] Sarrion-Gavilan M D, Benitez-Marquez M D and Mora-Rangel E O 2015 Tourism Management Perspectives 15 29-45
[28] Griffith D A and Chun Y 2018 Reference Module in Earth Systems and Environmental Sciences, 75 25-37
[29] Bandyopadhyay M, Singh M P and Singh V 2012 International Journal of Advanced Research in Computer Science and Software Engineering 2 69-74
[30] Aghajani M A, Dezfulian R S, Arjoody A R and Rezaei M 2017 Transportation Research Procedia 25 2126-38
[31] Shekhar S, Evans M R, Kang J M and Mohan P 2011 Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 1 193-214
[32] Getis A 2015 International Encyclopedia of the Social & Behavioral Sciences 23 100-104
[33] Tsai P J, Lin M L, Chu C M and Perng C H 2009 BMC Public Health 9 1-13
[34] Das M and Ghosh S K 2017 *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **10** 2631-39

[35] Ying L G 2000 *Papers in Regional Sciences* **79** 75-89

[36] Brian B 2016 *Applied Geomatics* **8** 201-216

[37] Musakwa W and Niekerk A V 2014 *International Journal of Applied Geospatial Research* **5** 1-20

[38] Ma R, Gu C, Pu Y and Ma X 2008 *Sensors* **8** 6371-95

[39] Ayadi M and Amara M 2009 *Spatial Patterns and Geographic Determinants of Welfare and Poverty in Tunisia*. Economic Research Forum Working Paper No 478

[40] Eck J E, Chainey S, Cameron J G, Leitner M and Wilson R E 2005 *Mapping Crime: Understanding Hot Spots* (USA: National Institute of Justice)

[41] Khan D, Rossen L M, Hamilton B E, He Y, Wei R and Dienes E 2017 *Spat Spatiotemporal Epidemiol* **21** 67-75

[42] Anselin L 2003 GeoDa™ 0.9 User’s Guide (Urbana: Center for Spatially Integrated Social Science)

[43] Anselin L 1995 *Geographical Analysis* **27** 93-115

[44] Chaikaew N, Tripathi N K and Souris M 2009 *International Journal of Health Geographics* **8** 1-10

[45] Rothlisberger V, Andreas P Z and Keiler M 2017 *Science of the Total Environment* **598** 593-603

[46] Zhang C, Luo L, Xu W and Ledwith V 2008 *The Science of the Total Environment* **398** 212-221

[47] Anselin L, Syabri I and Kho Y 2006 *Geographical Analysis* **38** 5-22

[48] Franczyk J and Chang H 2009 *Water Recourses Management* **23** 755-774

[49] Chung Y S, Seo D and Kim J 2018 *Sustainability* **10** 1-18

[50] Kemunto M G and Nyangena W 2017 *American Journal of Economics* **1** 64-85

[51] Aluko O 2011 *Ethiopian Journal of Environmental Studies and Management* **4** 69-82

[52] Fu W J, Jiang P K, Zhou G M and Zhao K L 2014 *Biogeosciences* **11** 2401-09

[53] Barreca A, Curto R and Rolando D 2018 *Sustainability* **10** 1-20