Spatial Points Pattern Analysis With Geographically Weighted Correlation Anlaysis of Facilities and Amenities for Aged Population

Neo Jia Hui
Singapore Management University

Ong Chen Yu
Singapore Management University

ABSTRACT

Though Singapore has been a global example of a modern smart city and deemed as an urban planning success, it has one of the fastest aging populations in the world. As such, Singapore faces the challenge of planning and developing the city to accommodate the changing age structure. This paper examines some of the essential amenities and facilities for the aged population in Singapore to evaluate if these are enough to support the increasing number of elderly residents. By designing and developing Old but Gold, a visualisation tool that uses various analysis methods such as supply and demand analysis, spatial point patterns analysis, and geographically weighted correlation, provides valuable insights for policymakers to determine if there is a need for more amenities and facilities to be built and where it should be built.

1. KEYWORDS

Spatial Point Pattern, Kernel Density Estimation, 2nd Order Point Pattern, Geographically Weighted Correlation, Ageing Population, R shiny

2. INTRODUCTION

Singapore is experiencing a demographic change as the population age. Our population is aging rapidly, by 2030, the number of Singaporeans aged 65 and above is projected to double to 900,000. This means that one in four Singaporeans will be aged 65 and above. In the face of an aging population, the total resources required to meet their needs will continue to grow. Besides, the land is also a key factor when it comes to planning due to land space limitations. Hence, it is essential that relevant amenities and facilities are strategically built and made readily available for the seniors in this land-scarce Singapore.

Singapore aspires to be a home where our seniors can age confidently and gracefully. To realize the goal, it requires planning ahead of the expected demand. However, policymakers do not have a convenient way to determine if there are adequate amenities and facilities to meet the needs of the seniors. Our project aims to provide policymakers the means to determine the availability and the adequacy of essential amenities and facilities in order to support and empower the seniors.

In our work, amenities, and facilities included are CHAS clinics, community clubs, eldercare services, gyms, and residents committee. In a public consultation with the elderly in Singapore by the Ministerial Committee on Ageing (MCA) showed that eldercare centers, practitioner clinics, and polyclinics are a few of the amenities that they value most [@seoulsg]. In a recent article by The Straits Times, seniors are given free entry to all public gyms and swimming pools as an initiative to encourage the elderly to stay active, as sports and physical activities allow them to "sustain their health and mobility longer, and continue to be engaged in the community" [@straitstimes].

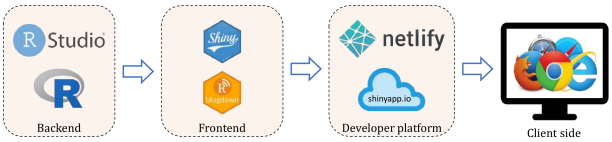
3. RELATED WORKS

Spatial analysis has been conducted by many researchers over the years and in a journal published by MDPI, the authors analysed the spatial distribution of urban service facilities, healthcare facilities in particular, in relation to the road network [@mdpi]. Three methods are used by the team in analysing the distribution of healthcare facilities and the relationship between healthcare facilities and road networks. They are namely, network Kernal Density Estimation (KDE), network K-function, and multiple centrality assessment model and correlation coefficient. As KDE is a frequently used technique to determine hotspots, it is applicable in our situation too.

In a work done by Chris Brunsdon in collaboration with Maynooth University in 2019, he made use of geographically weighted summary statistics to visualise the spatial patterns of human impact (a score to measure human impact) of Myroslava. By using the function gwss in the GWmodel package in R, he obtained the geographically weighted summary statistics, which he then proceeded to plot the various maps of geographically weighted mean, standard deviation, skewness as well as correlation [@gwss].

4. METHODS

4.1 System Architecture



4.2 Data Collection

All the data required for this project are readily available for download either from data.gov.sg or singstat.gov.sg. The data that we used are as follows:

shiny	shinythemes	DT
sp	rgeos	sf
rgdal	tidyverse	tmap
maptools	raster	spatstat
httr	rvest	GWmodel
lctools	RColorBrewer	shinycssloaders

Figure 1: System Architecture

4.3 Data Cleaning

CHAS clinics, community clubs, eldercare services, gyms, and residents committee datasets consist of locations of the facility and upon taking a closer look at the datasets, there are cases whereby both facilities share the same building and hence we did not remove the duplicated points except for the KDE analysis as the points used cannot be coincident.

4.4 Data Preparation

As our application focuses on the aged population in Singapore, we extracted the latest census data (2020) and calculated the population of residents aged 65 and above in each subzone from the CSV file obtained from Singstat.

4.5 Supply and Demand Analysis

To determine whether or not the facilities in a specific planning area are well-catered to the aged population, we conducted a supply and demand analysis. We join the points of interest to the census tract polygon and count them up for each polygon. By mapping out both the supply and demand map, users can then better visualise and compare the supply and demand of the facility, and determine if there is a need for more facilities to be built in the subzone.

4.6 First Order Spatial Point Pattern Analysis - Kernel Density Estimation

KDE is one of the most commonly used techniques to determine the concentration of point events. It estimates the local density of points in a non-parametric and continuous way by counting the number of events in a region that is centered at the location where the density is calculated. The use of KDE to examine the distribution of facilities in subzones enables users to see where facilities are sparsely distributed (dispersed), and where they are more intensely distributed (clustered). Therefore, we decided to use KDE to visualise the distribution of the chosen facility in a specific subzone.

4.7 Second-Order Spatial Point Pattern Analysis - G, F, L, & Ripley's K Functions

The second-order spatial point pattern analysis is a distance-based method to determine how the points are distributed relative to one another.

G-function tells us the way the events are spaced in a point pattern. If G(r) increases rapidly at a short distance, it means that there is a sign of clustering in that area, for that facility. With the envelope, can say that there is a sign of clustering if the G(r) function is above the simulated envelope.

We can also tell if the events are clustered or evenly-spaced through F-function. If F(r) rises slowly at first, but more rapidly at longer distances, they are clustered, while if F(r) rises rapidly at first and slowly at longer distances, they are evenly-spaced. With the envelope, we can say that there is a sign of clustering if the F(r) function is below the envelope, and the regular pattern if above the envelope.

Ripley's K-function is used to describe how point patterns are distributed over a region. We can determine if the point patterns are dispersed, clustered, or randomly distributed over the region. Compared to G and F-function, it uses more points while providing an estimate of spatial dependence over a wider range of scales, assuming isotropy over the region. With the envelope, we can say that there is a significant cluster pattern if the function is above the envelope, and significant regular if below.

L-function is a variance-stabilising transformation of the K-function. L(r) more than 0 indicated that the observed distribution is geographically concentrated, while L(r) less than 0 implies dispersion.

4.8 Geographically Weighted Correlation

Geographically weighted correlation measures the degree to which two variables are associated. The most commonly used measure of geographically weighted correlation is Pearson Correlation Coefficient, followed by Spearman's Rank Correlation Coefficient. The idea of geographically weighted correlation is to use a moving window approach to determine whether or not this degree of association varies geographically. In order to calculate the geographically weighted correlation, geographically weighted summary statistics are used. The team will be adopting the method used by Chris Brunsdon as mentioned.

5. APPLICATION

This section will highlight some of the application features of the Old but Gold application at different stages of a user's journey.

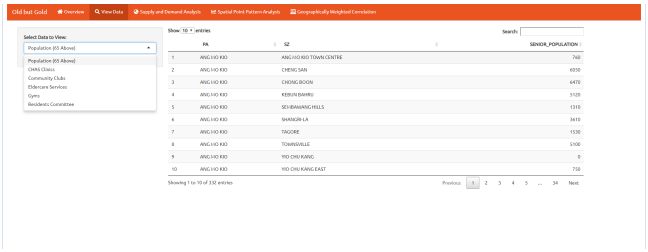


Figure 2: View Data Tab

View data tab allows the user to have a glimpse of all the datasets that are being used in the application. From the population dataset, we included the planning area, subzone, and the number of seniors in that specific subzone. For the rest of the location data, the name of the facility, postal code, and address are shown.

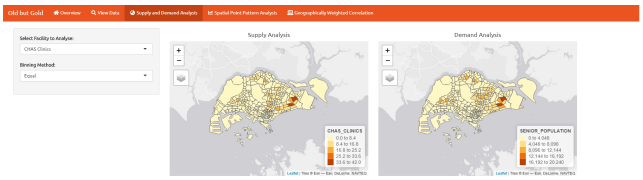


Figure 3: Supply and Demand Analysis

On the supply and demand analysis page, users will be able to select the facility they would like to analyse, as well as the binning method for the analysis from the side panel. On the main panel, the supply of the selected facility is then shown, with the demand map on the right. The demand is mapped using the senior population in the subzone, while the supply map is mapped using the number of facilities in the subzone.

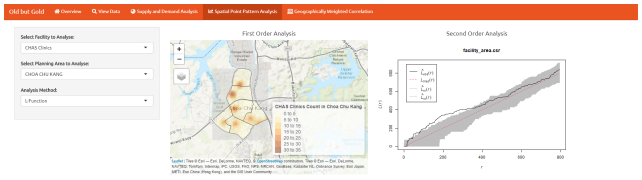


Figure 4: Spatial Point Pattern Analysis

Under the spatial point pattern analysis tab, users are able to select the facility to analyse, the planning area, and the analysis method for second-order analysis from the side panel.

On the main panel, it shows the KDE analysis map on the left and the second-order analysis on the right.

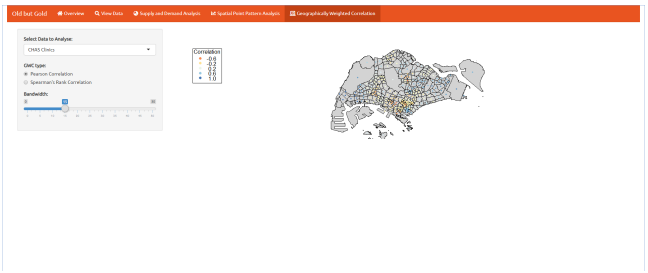


Figure 5: Geographically Weighted Correlation

Under the geographically weighted correlation tab, it shows the correlation between the selected facility and the population geographically. Users are able to select from the list of facilities, select the type of correlation, and the bandwidth of the statistics.

6. RESULTS

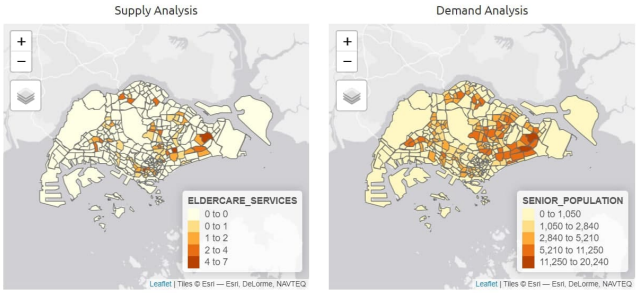


Figure 6: Demand and Supply Maps

Taking eldercare services as an example, we can see that from the demand map, as shown above, the aged population is more populated in the east, followed by the west and the north of Singapore, and the supply of eldercare services are largely located in the east, with some in the south and west, with lesser eldercare services in the north. These can be deduced from the shade of the colour of the map. From this analysis, we can derive that there could be more eldercare services in the north and west to cater to the aged population in those areas.

The observations are logical as the shaded regions on the map are mainly residential area, it would be inevitable that there will be some residents aged 65 and above living in the area.

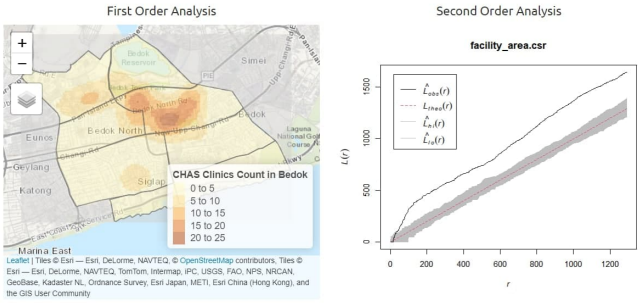


Figure 7: First and Second Order Analysis

Taking a closer look at the east side, we will take CHAS clinic in the area Bedok as an example. From the KDE analysis on the left, we can see that the CHAS clinics are more concentrated in the heart of Bedok as the colour is more intense in that area. L-function is used in this example and looking at the second-order analysis, the result indicates strong clustering at almost all distances, which corresponds to the observations made from the KDE plot.

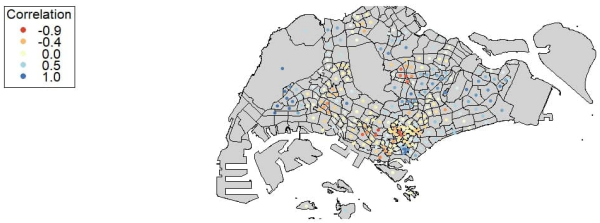


Figure 8: Geographically Weighted Correlation

Geographically weighted correlation measures the degree that two variables are associated and in the application, the selected facility is being plotted against the population of residents aged 65 and above, to determine the relationship between the two variables. In the example shown here, we selected the CHAS clinic variable and Pearson correlation with 10 as its bandwidth. As we can observe from the map, most of the regions are not correlated with CHAS clinics as most of them are yellow or red in colour. Regions such as the west, east and some parts in the south of Singapore are either marginally or highly correlated with the senior population in Singapore.

The objective of plotting geographically weighted correlation is to ensure that there are no areas that have an excess of facilities and too little aged population, or vice versa. From the map, as shown above, we can see that there are areas, mainly the central regions, that do not have a high correlation between the aged population and CHAS clinics. Hence, they are possible regions to build more CHAS clinics for the residents in those areas.

7. DISCUSSION

In light of the Covid-19 situation, townhall presentation is not being held this year. However, the team managed to present the application to a few close friends and family in order to gather some feedback and insights.

A common learning point from all the audience is that as a whole, it is easy to determine which areas require more facilities to cater to areas with more elderly in the supply and demand analysis. User studies proved that the KDE provides a clear visualisation of where the facilities are located within a specified region. Users also feedbacked that the geographically weighted correlation acts as a good indicator as to whether or not the facilities are well catered to the aged population in Singapore, though some users felt that the graph could have been plotted clearer.

Additionally, an observation that we made while the users are trying out the application, is that the users generally are able to figure out where to click in the application, which can also imply that our application is rather user friendly and intuitive, where users are able to navigate without much help.

Overall, most users that used the application find that the application would be useful for policymakers in planning and deciding which regions require more attention and which areas within the subzone need more facilities.

8. FUTURE WORK

A lot of efforts have been done by the government of Singapore in helping the elderly age gracefully and healthily, and this application that our team has done is merely scratching the surface of what can be done.

Currently, the facilities that our application includes are CHAS clinics, community clubs, eldercare services, gyms, and resident committees. However, there might be more facilities and amenities that the elderly in Singapore value a lot but not included here, such as traditional Chinese medicine clinics. Hence, if there is a chance, more facilities and amenities could be included.

Given the opportunity, we believe that we could allow users to upload their own CSV, SHP, or KML files to our system to model the analysis of the different facilities. Presently, our application does not include file uploads by users and is loading existing data to run the analysis.

Implementing the aforementioned, the application would provide more user interaction and variation, ultimately improving it.

9. ACKNOWLEDGMENTS

The authors would like to thank Dr. Kam Tin Seong, the Associate Professor of Information Systems from Singapore Management University for his kind support and guidance in the process of working on this project.

10. REFERENCES

Chia, N. (2019). AGE-FRIENDLY CITIES - Lessons from Seoul and Singapore, CLC Publications, March. Available at: https://www.clc.gov.sg/docs/default-source/boo ks/book-age-friendly-cities.pdf

Ni.J, Qian.T, Xi.C, Rui.Y, Wang.J (2016). Spatial Distribution Characteristics of Healthcare Facilities in Nanjing: Network Point Pattern Analysis and Correlation Analysis, MDPI, August. Available at: https://www.mdpi.com/1660-4601/13/8/833

Brunsdon.C (2019). Geographically Weighted Summary Statistics, R-Studio, July. Available at: https://rstudio-pubs-static.s3.amazonaws.com/503649_13027b8d4636466d93917435af74aea3.html