1 Motivation

In the last few years, there has been a strong push by the government to encourage cycling in Singapore by expanding the existing cycling network. This is in hopes that cycling becomes a viable option for people to get around, and by proxy, also aims to reduce our city-state's reliance on motor vehicles. This project aims to discover some of the factors that led to the existing placement of park connector networks.

2 Problems to solve

It is assumed that park connectors were not built in a random manner, but were placed in a way to benefit Singapore residents. Initially, we decided to look at the distribution of various amenities such as bicycle racks, hawker centres, supermarkets and shopping malls, to see if park connectors were placed to provide easy access to such amenities. We also analysed the distribution of park connectors with respect to population density, to find out if park connectors were serving more densely populated areas. Next, we decided to investigate if park connectors could possibly be designed as a last-mile solution to the MRT network. Lastly, we also wanted to statistically verify if park connectors truly connect different parks in Singapore.

3 Context of data

A note on the chosen data format: we initially wanted to perform analysis on line data consisting of the park connectors themselves, but quickly realised that this was difficult especially for the density-based and distance-based functions that we required. Because of this, we managed to find another dataset that provided just the access points to the park connector (PCAPs) - fortunately, this was all point data. Access points were also a more accurate proxy for accessibility to these park connectors, since people can only get on these park connectors through one of the access points, and not merely at any arbitrary point along the line of park connectors.

4 Data Description

4.1 Park connector access points

This dataset was retrieved from Data.gov.sg and it contains all access points to park connectors in Singapore. The file formats are KML and GeoJson and each PCAP is represented as a point.

4.2 Hawker centres, supermarkets, MRT exits and parks

From Data.gov.sg, we used datasets on hawker centres, supermarkets, MRT exits, bicycle racks which were in KML file format while the parks dataset was in GeoJson format.

They are represented by point data.

4.3 Bus stops and bicycle racks

Bus stops were retrieved from LTA's Datamall in SHP file format.

For bicycle racks, we used LTA bicycle racks which we retrieved from Data.gov.sg in GeoJson file format.

4.4 Planning Area

We downloaded Singapore's planning areas from Data.gov.sg as a KML file. This dataset is in vector format.

4.5 Singapore's population density

Lastly, we downloaded Singapore's population density as raster data in the tif file format.

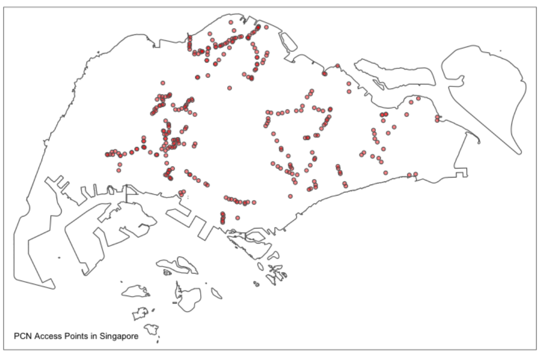
5 Descriptive Analysis

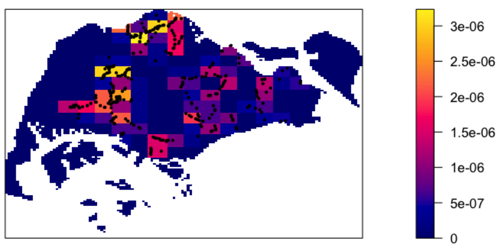
To achieve the above, we first generated visualisations plotting the PCAPs over all the underlying distributions we were interested in. This includes point data such as MRT stations, bus stops, amenities and parks, as well as population density raster data.

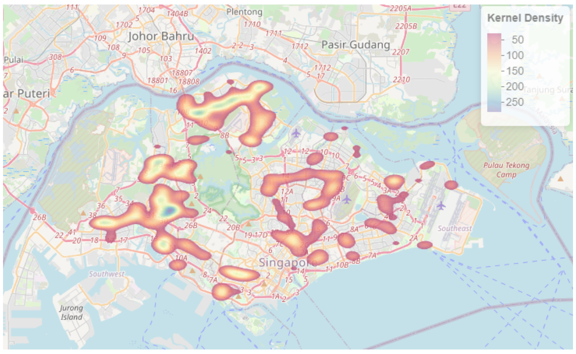
We tried various ways of visualising the data, including plotting the points themselves on a map, as well as using local density estimators such as quadrats or KDEs to better visualise areas with higher concentrations. We also used buffers to calculate the percentage of PCAPs within 500m of MRT stations, bus stops, amenities and parks.

5.1 Distribution of Park Connector access points (PCAPs)

First, we plot the PCN access points themselves. At first glance, it is apparent that these points cannot be randomly distributed, with noticeable clusters in the Choa Chu Kang/Bukit Batok area as well as the Woodlands/Sembawang area. The points also seem to occur as a "string" of points, which obviously stands to reason since these are access points to Park Connectors which are essentially lines. Also, large areas of Singapore do not contain any PCAPs at all, such as the northwestern region of Lim Chu Kang as well as the middle of the island, which is home to nature reserves like the Central Catchment Nature Reserve (home to Macritchie Reservoir) and the thickly forested Mandai area (home to the Singapore Zoo and Night Safari). By plotting them on a topological map, it is obvious that these areas are coloured green, meaning they are largely forested and not urbanised at all.



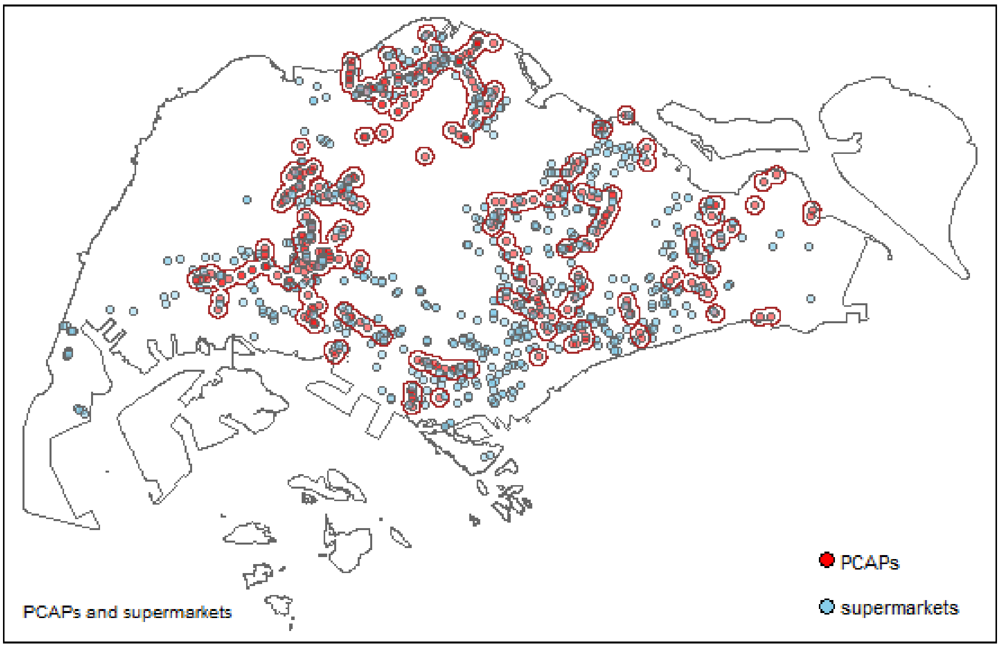




Next, we plotted the geographical distribution of the various amenities (supermarkets, hawker centres, etc.) on top of the locations of the PCN access points.

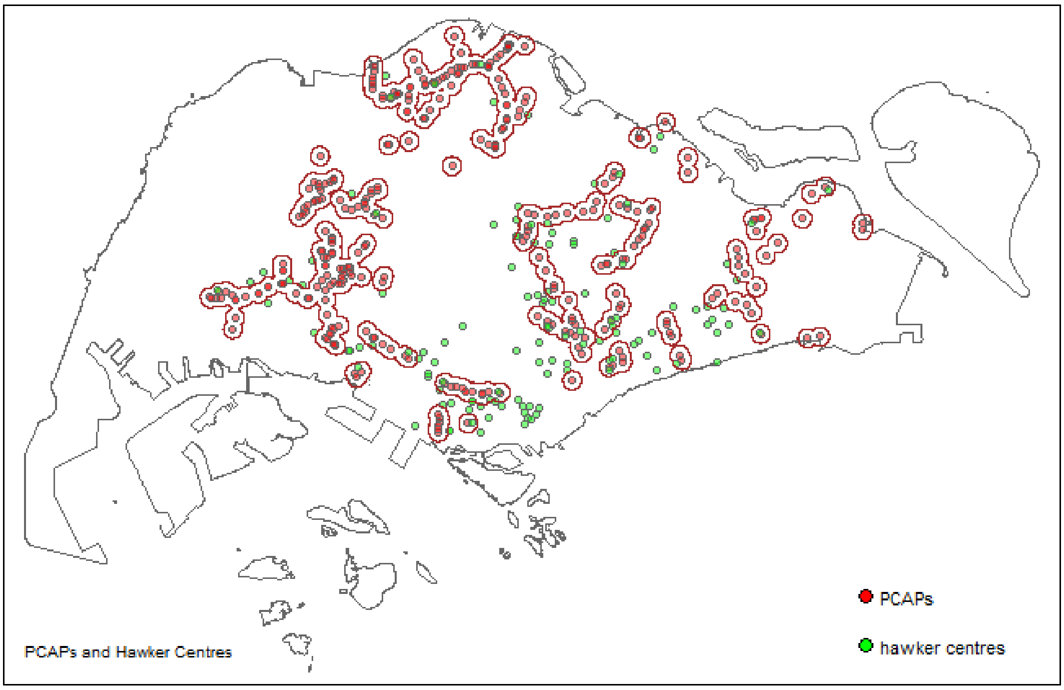
5.2 Supermarkets

We overlaid supermarket locations on top of PCAPs on a map of Singapore. We see that most access points are located conveniently to supermarkets, with 74.18% of access points being within 500m of at least one supermarket.



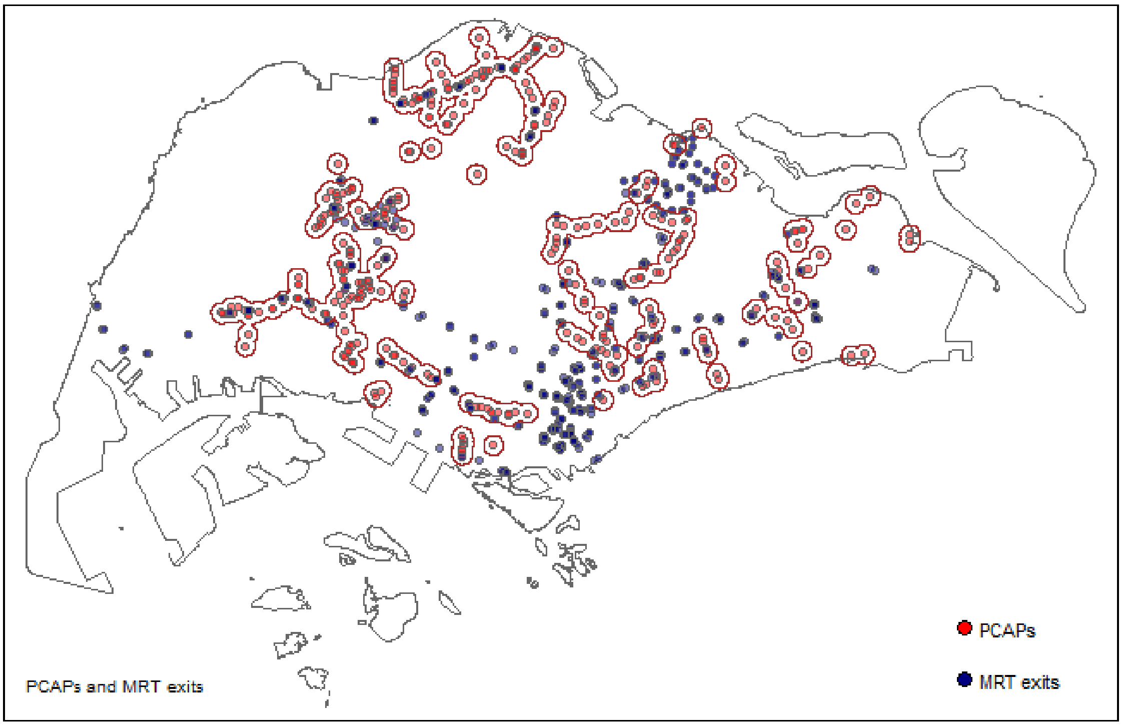
5.3 Hawker Centres

The global density of hawker centres is much lower than that for supermarkets, and as a result, only 23.20% of PCAPs are located within 500m of a hawker centre.



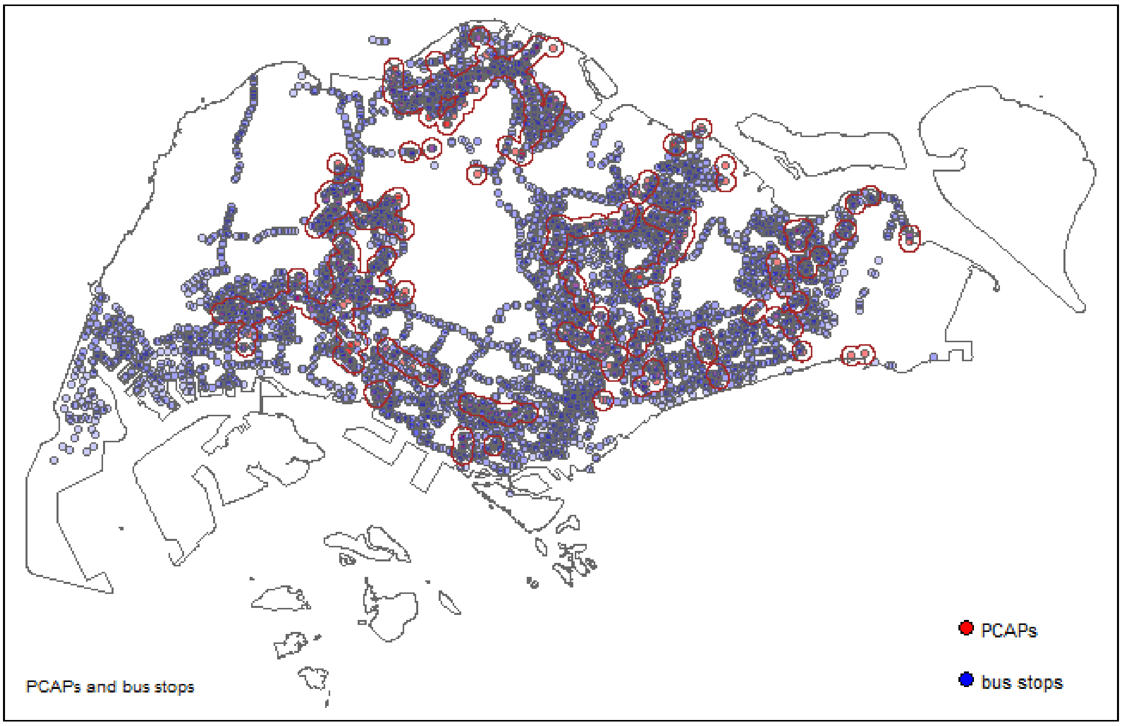
5.3 MRT exits

We notice that the Downtown Core area has a high concentration of MRT exits. This could be because many of the stations in that area have many exits, sometimes as many as 7 or 8. Unlike the MRT exits, we see that there are very few PCN access points in the Downtown Core area. This could be because of the lower priority assigned to cycling paths over mass rapid transit in a space-scarce area like the Downtown Core. 36.60% of PCAPs are within 500m of an MRT exit.



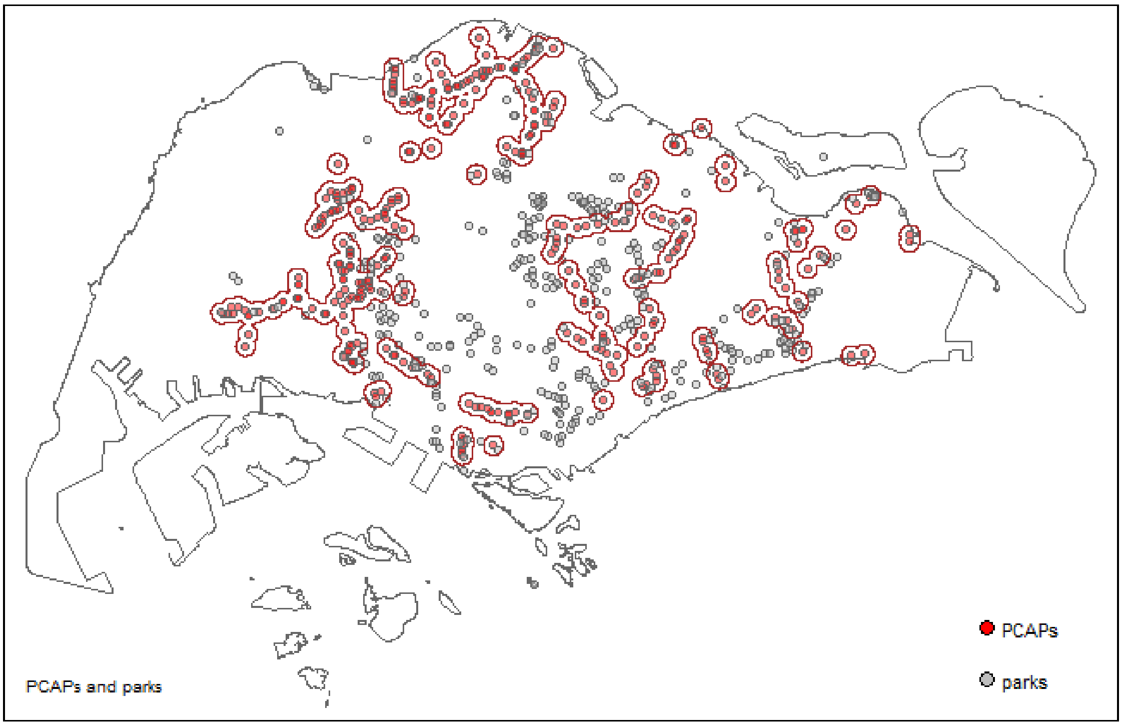
5.4 Bus stops

Singapore is saturated with over 5160 bus stops, hence it is not surprising that 98.69% of PCAPs are within 500m of a bus stop.



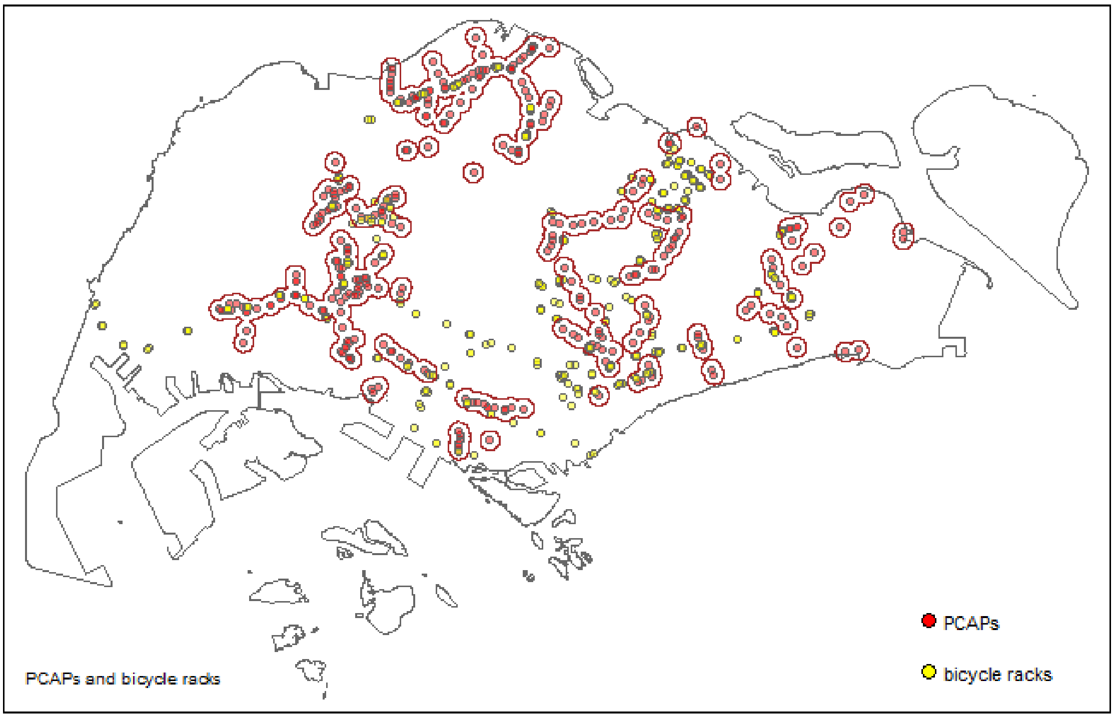
5.5 Parks

We also want to see if park connectors have any relationship with the parks in Singapore. We observed that 39.54% of PCAPs are within 500m of parks.



5.6 Bicycle racks

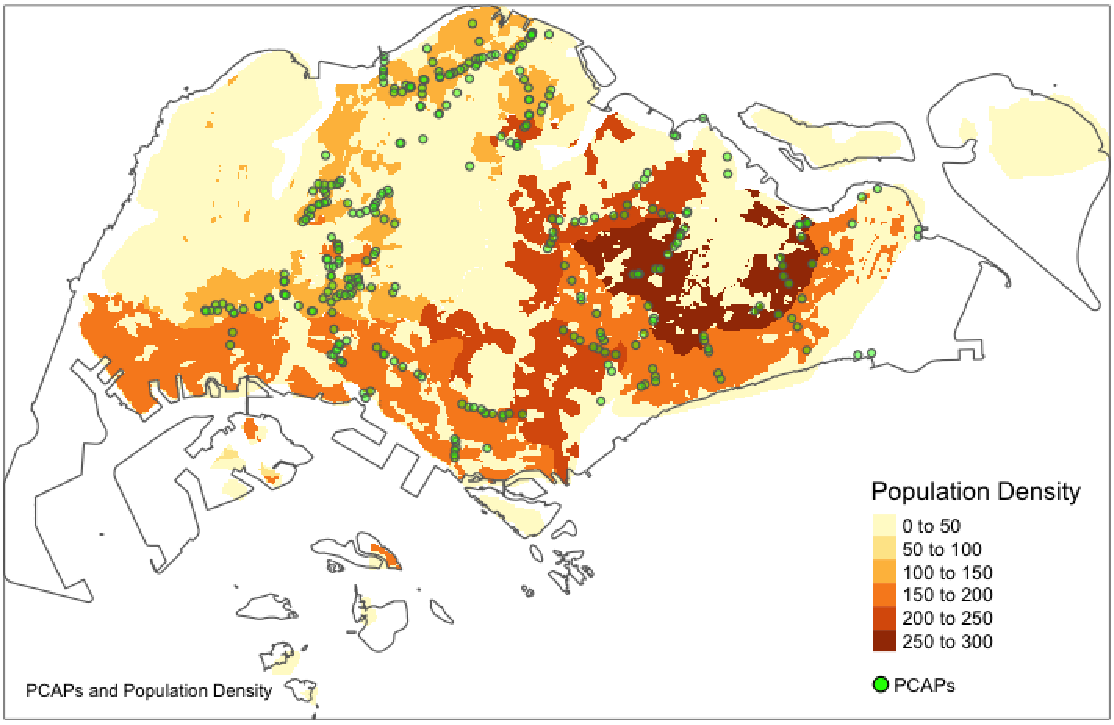
37.58% of PCAPs are within 500m of bicycle racks



5.7 Population Density

Interestingly, areas with the highest population density (Serangoon, Hougang and the areas surrounding Paya Lebar) seem to be relatively underserved by access points as compared to Western regions like Bukit Batok which has a much lower population density.

Standing out as well is the central part of Singapore, which has virtually zero population density. This region encompasses the MacRitchie Reservoir area and naturally is not home to any human population density. The same can be said of the northwestern region of Singapore - Lim Chu Kang. This corresponds with what we see on the topological map earlier when visualising the KDE for the PCAPs, where these thinly populated areas correspond to areas marked green, indicating thick forest, on the topological map. Understandably, there are very few PCAPs located in these areas as well. It might be useful to exclude these region in later analysis to prevent confounding.



After initial exploratory analysis, we calculated the local densities of these underlying distributions using local density functions such as Quadrat and Kernel Density Estimator (KDE). We plotted the distribution of PCAPs over these density functions, to check if these were correlated.

6 Spatial data analysis

6.1 Methodology

First, underlying distributions of various amenity types will be captured using density-based methods like quadrats. A thousand Monte Carlo simulations would then be run to generate many realisations of the PCAPs, using these underlying distributions to assign probabilities for the random generation of points. In the context of density, an area with a higher density value would have a greater probability of having a random point generated there in the Monte Carlo simulation. A hypothesis test would then be performed to check for significant evidence that the distribution of PCAPs indeed follows these underlying distributions. In this case, the hypotheses are as follows:

H_0: The PCAPs follow a distribution similar to that of the underlying point process

H_1: The distribution of PCAPs is not correlated with the underlying process

Here, unlike in usual hypothesis testing, the desired p-value would be one close to 0.5, meaning that the observed distribution of PCAPs follows a similar distribution as the simulations generated by the Monte Carlo on the underlying process.

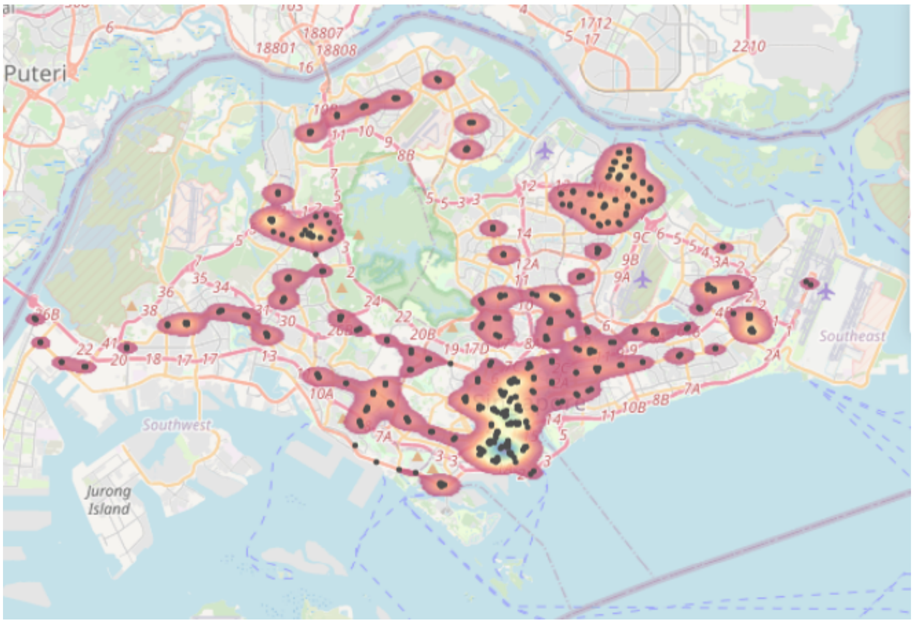
6.2 Density-based methods

In order to create an underlying distribution for each of these types of amenities (MRT exits, supermarkets, etc.) we first experimented with density based methods like KDE and quadrats to derive a locally varying concentration of each of these amenities across Singapore.

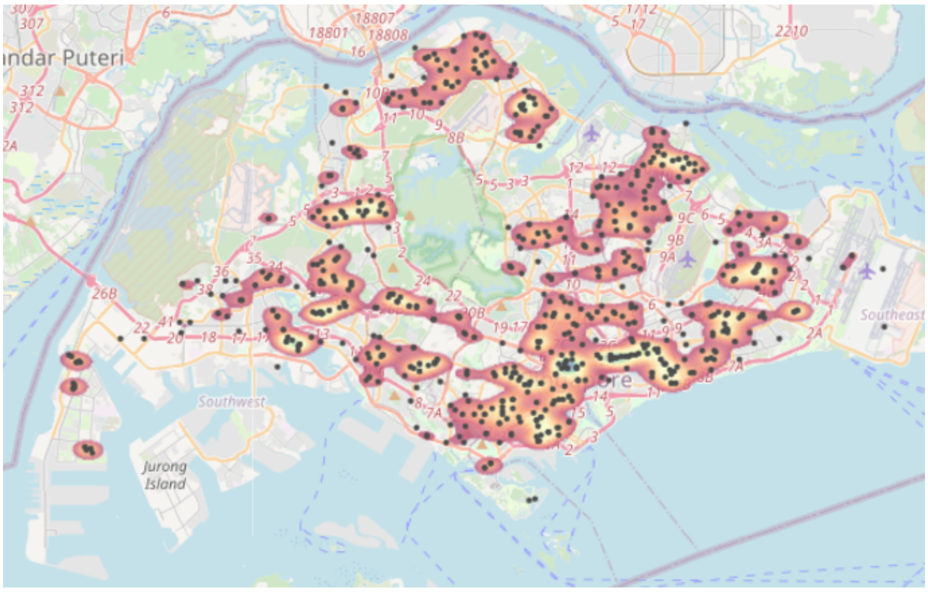
6.2.1 Kernel Density Estimate (KDE)

Using Gaussian (bivariate normal) kernels, KDEs were performed over the different amenities to observe densities over the map of Singapore. To do so, KDE values were plotted via heatmaps, with modes of the density distribution representing hotspots denoting locales in Singapore with high density. The bandwidth selection was done with respect to the observed output of these visualisations. Selecting too high a bandwidth resulted in over-smoothing with a single unimodal output that spanned across the entire analysis region. Thus, the bandwidth selection was hand-tuned with the bandwidth value being pared down until a satisfactory visualisation output was derived.

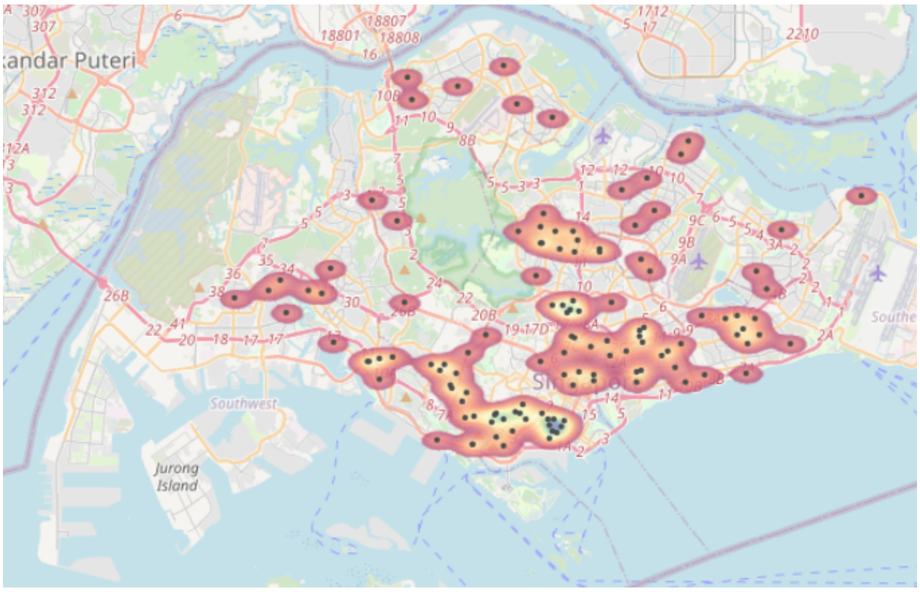
KDE - Park Connector Access Points



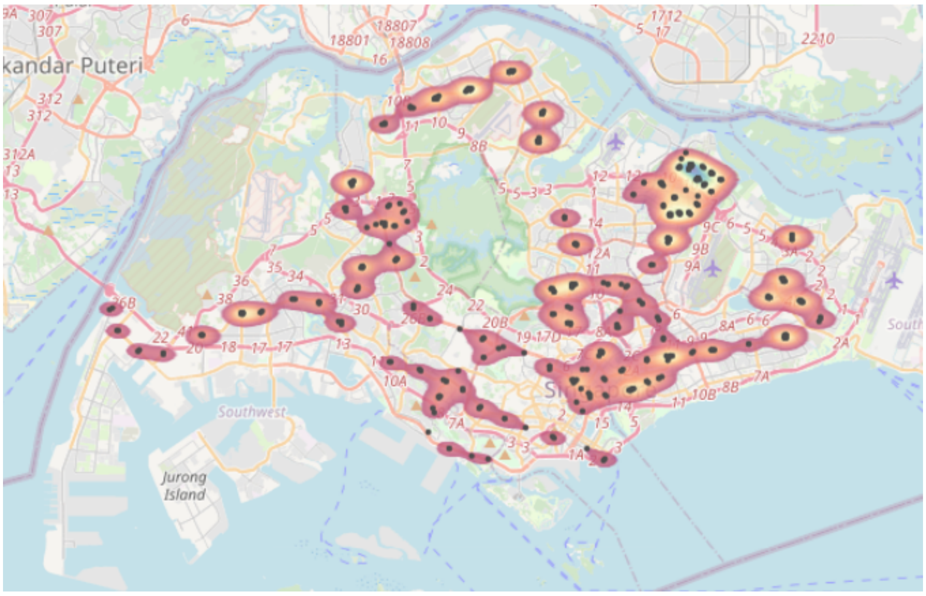
KDE - MRT Exits



KDE - Supermarkets



KDE - Hawker Centres



KDE - Bicycle Racks

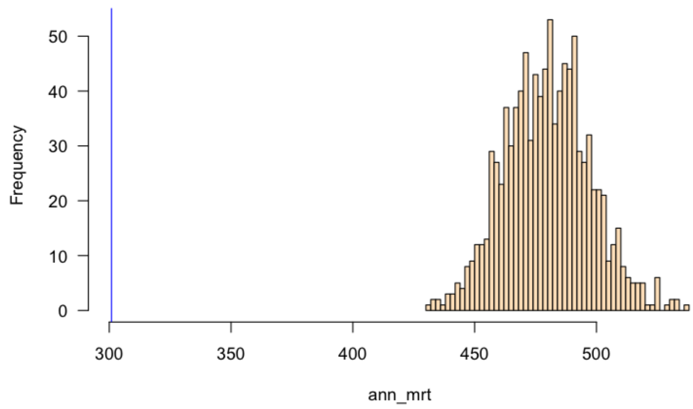
From the KDE analysis, it seemed there were promising correlations between the density distributions of PCAPs, MRT exits and other amenity types. From here, quadrat analysis was performed to investigate further.

6.2.2 Quadrat density

The map of Singapore was divided into quadrats, and the frequency of each amenity type was calculated for each quadrat. In our case, we worked with a map divided into 20 by 20 quadrats, giving us rectangles of around 2.687km by 1.725km to work with, which were suitable, being with reasonable cycling distance.

6.2.3 Monte Carlo on Density-based Underlying Distributions

Disappointingly, we found that the distribution of PCAPs did not seem to follow any of the density-based distributions we used. Across all amenity types, PCAPs were more clustered than expected given the density of that amenity type as an underlying distribution. All of them returned a p-value of zero, meaning that we could find enough evidence to say that the PCAPs followed these underlying distributions.



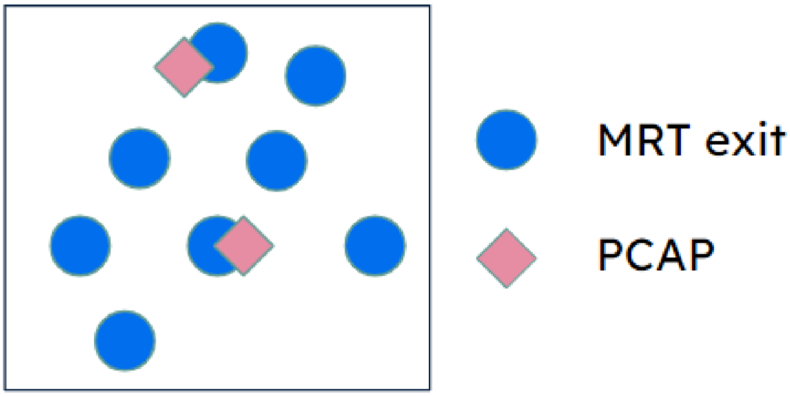
Monte Carlo distribution plot - MRT exit density

For example, the observed Average Nearest Neighbour using (using neighbour order, k = 1) distance was around 300 for the Monte Carlo simulation based on MRT exit density, whereas the Monte Carlo simulations had a mean of 480, and a standard deviation of 17.5, with the vast majority of realisations between 450 and 500. In fact, after the true distribution, the next most extreme simulation had an ANN distance of 430. This means that the distribution of PCAPs are highly unlikely to be correlated with density of MRT exits, and the same applies for supermarkets and parks.

6.3 Distance-based methods

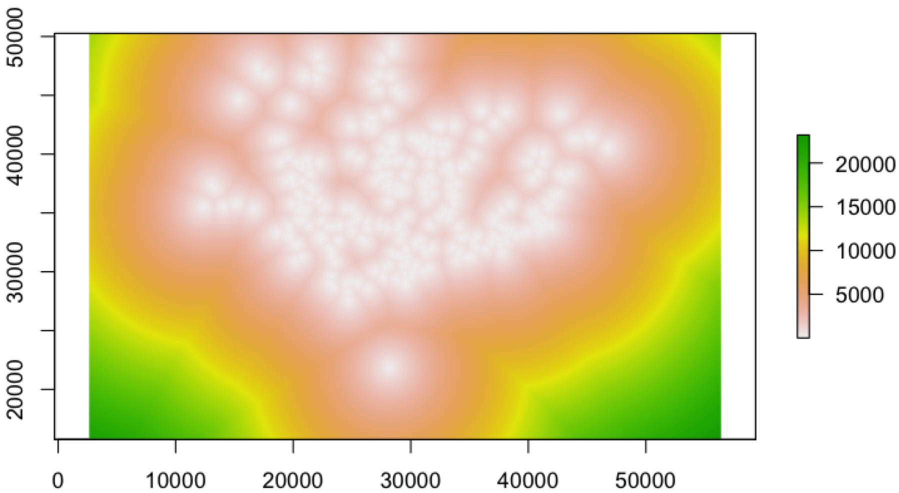
Finding no significance with density-based methods, we realised that density-based functions might not be appropriate for statistically testing our point data. What we wanted to know was how accessible these points were for people coming from a park connector. Take MRT stations as an example. Using the density function would lead us to conclude that if there were a correlation of densities of PCAPs and MRT stations across space, park connectors were indeed making access more easy to MRT stations. However, in reality, areas with low PCAP density but high MRT density may not mean that MRTs were less accessible - the few PCAPs in existence could still be arranged strategically in a way that provided easy access to the many MRT stations.

The diagram below illustrates how low PCAP density but high density MRT could provide easy access from the PCAPs to MRT exits. This is because the distance from PCAPs to the MRT exits is short. Hence, we want to use distance-based methods for analysis instead.



6.3.1 Raster Euclidean Distance function

Upon having this realisation, we pivoted our methodology to use the Euclidean Distance tool instead of the density functions. This enabled us to calculate the shortest distance to any of these amenities/MRTs/parks from any arbitrary point on the map. This was a better representation of how accessible these locations are from the nearest PCAP.

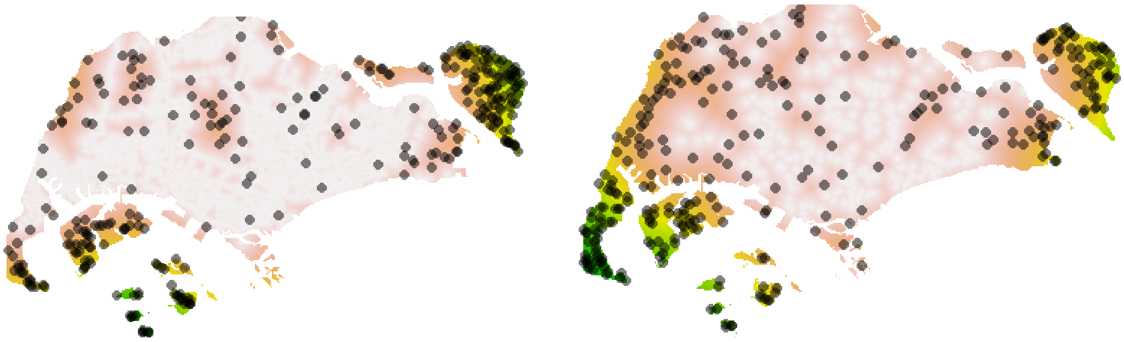


Global distance raster from nearest MRT exit

Using the new distance-based rasters as underlying distributions, we generated Monte Carlo simulations of PCAPs and used hypothesis testing to verify if the distributions PCAPs were truly related to key infrastructure/amenities like MRT stations.

6.3.2 Inverting the distance raster

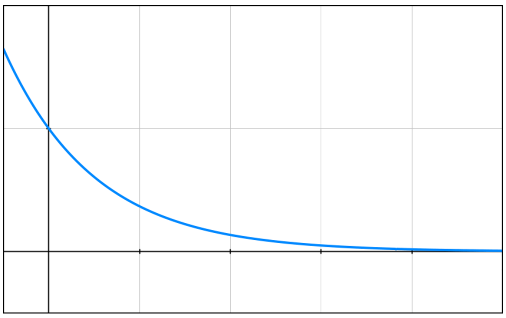
Initially, we directly fed the distance raster as an underlying distribution to the Monte Carlo simulation. However, this resulted in simulations that seemed to “avoid” areas where the distance was farthest away from the underlying points. In many of the simulations, the points seemed to be clustered in areas like Pulau Tekong in the northeast and even Jurong island on the southwest.



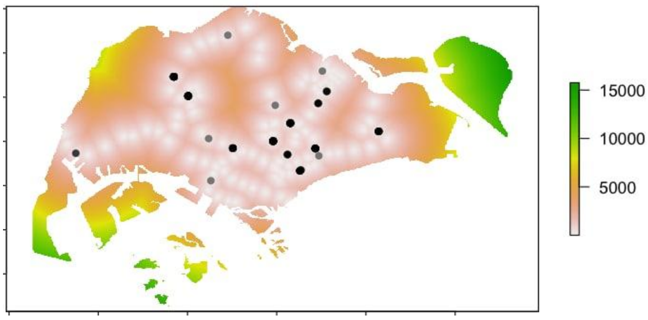
Some simulations generated from directly feeding the distance raster into the Monte Carlo simulation

We soon realised that the simulation was generating points at a higher probability in areas where the underlying distribution raster showed the highest values. This was the opposite behaviour of what we wanted, where we wanted points to be generated at a higher probability in areas where the distances are closer to the points used in the underlying distribution (eg. MRT exits, parks).

Because of this, we had to figure out a way to somehow invert the values in the distance raster. Initially, we tried simply negating the values, but the rpoint function in R could not take negative values as an underlying probability distribution. We also experimented with an exponential function - e^{-x}, which would assign lower values to values that used to be higher and vice versa. Unfortunately, this over-emphasised areas with the shortest distances, causing the points to be highly clustered around these areas. This corresponds to mathematical intuition, where the e^{-x} function assigns exponentially higher probability as distance decreases to zero. In the below distribution, 306 PCAPs have been condensed into merely what appears to be less than 20 black dots on the map, showing that the exponential function is so skewed that many of the points have been plotted over each other.

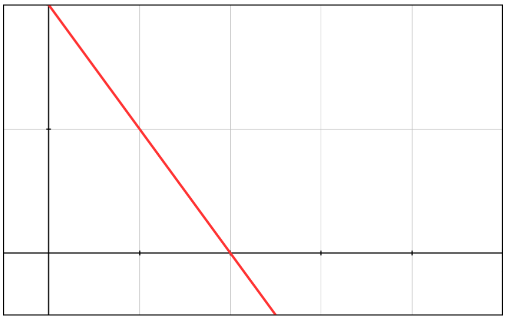


e^{-x} function

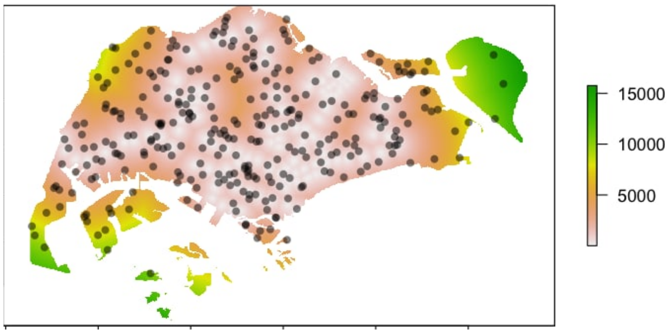


sample Monte Carlo simulation

We then changed the function to a negative linear function, y = -x + c where c, the intercept, was assigned the value of the maximum value in the distance raster, such that all points in the inverted distribution fall above zero. This preserves the linear relationship between distances, meaning probability of generating points varies in proportion to the distance.



y = -x + c



sample Monte Carlo simulation

This resulted in a much more evenly spread distribution, and was applied across all the Monte Carlo simulations we used.

Summary of Results - Islandwide

Underlying distribution	Density/ Distance-based	Mean ANN distance of MC simulation	SD of ANN distance of MC simulation	p-value
MRT exits	Density	479.90	17.48	0
MRT exits	Distance	848.61	30.89	0
Bus stops	Density	651.51	23.48	0
Bus stops	Distance	854.94	29.70	0
Parks	Distance	756.85	29.72	0
Bike racks	Distance	850.06	30.91	0
Population Density	Density	567.91	25.80	0

Disappointingly, our results for showed that the observed distribution was way more extreme than anything generated by our Monte Carlo distribution, meaning that we do not have enough evidence to say that the distribution of PCAPs follows that of any of these underlying distributions - MRT exits, Bus stops, bike racks, population density and even parks!

One thing we did observe was that the mean and standard deviation of the Average Nearest Neighbour distance changed when moving from a density-based analysis to a distance-based one. For the Monte Carlo simulation of MRT exits for example, the mean ANN distance for the distance-based simulation was almost twice that for the density based one. This means that if PCAPs truly followed the distance to the nearest MRT exits as its underlying distribution, PCAPs would be located an average of 848m away from each other. On the other hand, if PCAPs truly followed the density of MRT exits as its underlying distribution, PCAPs would be located just under 480m away from each other. This shows that density and distance-based analysis are highly incongruent, especially for our case when looking at accessibility measures.

We were not convinced that none of our chosen underlying distributions were statistically significant, so we decided to zoom in on certain planning areas of Singapore to see if the trend held true on a more localised basis.

6.4 Regional Analysis

Two particular regions were selected based on the visualisations from KDE, Jurong West and Geylang. The choice of Geylang was made after observing common clustering patterns of PCAPs and amenities in that region. By contrast, such a degree of clustering was not seen in the case of Jurong West.

These regions are derived from the planning areas as demarcated by Singapore. The same methods employed previously over the entire Singapore region are applied here, the only difference is the scope of analysis. To do this, we filtered the necessary data according to the planning area. Following are the results from running Monte Carlo simulation that compare the ANN of PCAPs in that region to the indicated amenity type in that region.

Summary of Results - Geylang

Underlying distribution	Density/ Distance-based	Mean ANN distance of MC simulation	SD of ANN distance of MC simulation	p-value
MRT exits	Density	776.32	322.14	0.646
MRT exits	Distance	977.76	308.85	0.415
Parks	Density	691.51	342.24	0.722
Parks	Distance	945.36	301.20	0.452
Bike racks	Density	605.95	283.23	0.834
Bike racks	Distance	956.55	294.09	0.446

Summary of Results - Jurong West

Underlying distribution	Density/ Distance-based	Mean ANN distance of MC simulation	SD of ANN distance of MC simulation	p-value
MRT exits	Density	63.008	15.245	1
MRT exits	Distance	467.45	63.50	0
Parks	Density	101.525	25.36	0.982
Parks	Distance	469.70	62.86	0
Bike racks	Density	75.180	21.687	1
Bike racks	Distance	466.37	64.92	0

From the above results, it is clear that regional analysis yielded much more meaningful information about the distribution of PCAPs and amenities. For the case of Geylang, the p-value of comparative ANNs show that the distribution of MRTs, parks and bike racks can be shown to have some significance in the PCAP distribution. In particular, the distribution trends of MRT exits and PCAPs are notably similar.

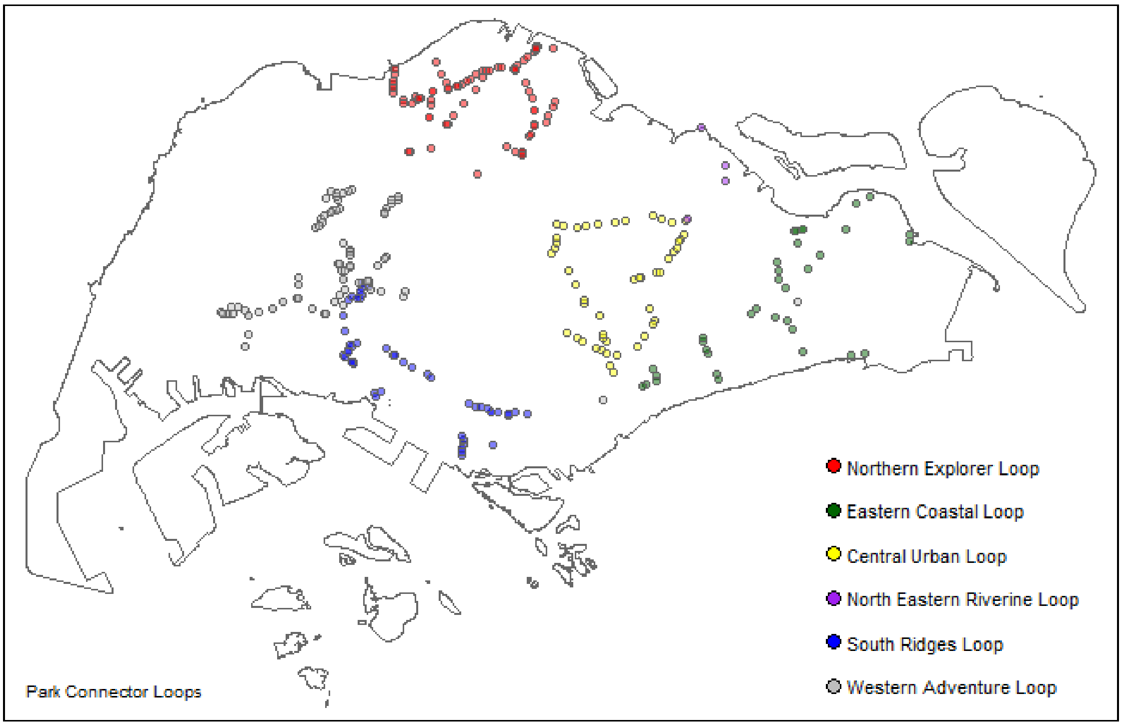
The Monte Carlo simulation done in Jurong West is more polarising in the results of the p-value, with results either showing a 0 or 1. This demonstrates a dissimilar distribution trend of MRT exits and amenities in that region.

These results are aligned with what was observable in the KDE and quadrat visualisations, bolstering the appropriateness of taking this analytical approach.

Conclusively, it can be seen that by narrowing the scope to a smaller area of analysis, we were able to derive more information that was better able to explain the phenomena of distributive patterns of PCAPs and amenities. One possible reason for this was that in the process of reducing the area under which we performed the Monte Carlo simulation, we are reducing the possibility of outlying samples being generated in locations which do not make sense. For example, a sample of a MRT exit being generated at Pulau Tekong would be an outlier which affects the calculation of ANN values.

6.5 Park Connector Loop Evaluation

As a conclusion to our findings, we also wanted to evaluate the 6 different park connector loops in Singapore, namely, the Northern Explorer Loop, Eastern Coastal Loop, Central Urban Loop, North Eastern Riverine Loop, South Ridges Loop, and the Western Adventure Loop. The diagram below illustrates how the different PCAPs are distributed across the 6 different loops.



In order to evaluate the park connector loops, we first count the number of occurrences of each amenity type within 500m for each loop using buffers. We then normalise the count by the number of PCAPs within the loop. We then give a score based on the ranking of the normalised count. For example, if the Eastern Coastal loop has the highest normalised count for MRT Exits it is given a score of 6, while the South Ridges loop is given a score of 1. The scores are tallied up to determine the final evaluation.

Loop	Northern Explorer (n=75)	Eastern Coastal (n=35)	Central Urban (n=48)	North Eastern Riverine (n=4)	South Ridges (n=49)	Western Adventure (n=77)
MRT Exits	Count = 24 Normalised count = 0.32 Score = 2	Count = 21 Normalised count = 0.60 Score = 6	Count = 21 Normalised count = 0.52 Score = 4	Count = 2 Normalised count = 0.5 Score = 3	Count = 13 Normalised count = 0.27 Score = 1	Count = 45 Normalised count = 0.58 Score = 5
Bus stops	Count = 338 Normalised count = 4.51 Score = 2	Count = 255 Normalised count = 7.29 Score = 6	Count = 336 Normalised count = 7 Score = 5	Count = 17 Normalised count = 4.25 Score = 1	Count = 254 Normalised count = 5.18 Score = 3	Count = 405 Normalised count = 5.26 Score = 4
Bicycle racks	Count = 41	Count = 23	Count = 25	Count = 4	Count = 11	Count = 51

	Normalised count = 0.55 Score = 3	Normalised count = 0.657 Score = 4	Normalised count = 0.52 Score = 2	Normalised count = 1 Score = 6	Normalised count =0.22 Score = 1	Normalised count =0.662 Score = 5
Parks	Count = 11 Normalised count = 0.15 Score = 2	Count = 25 Normalised count = 0.71 Score = 6	Count = 13 Normalised count = 0.27 Score = 4	Count = 0 Normalised count = 0 Score = 1	Count = 18 Normalised count = 0.37 Score = 5	Count = 16 Normalised count = 0.21 Score = 3
Supermarkets	Count = 48 Normalised count = 0.64 Score = 1	Count = 34 Normalised count = 0.97 Score = 4	Count = 56 Normalised count = 1.17 Score = 6	Count = 4 Normalised count = 1 Score = 5	Count = 43 Normalised count = 0.88 Score = 3	Count = 53 Normalised count = 0.69 Score = 2
Hawker centres	Count = 4 Normalised count = 0.05333333 Score = 3	Count = 7 Normalised count = 0.2 Score = 5	Count = 15 Normalised count = 0.3125 Score = 6	Count = 0 Normalised count = 0 Score = 1	Count = 9 Normalised count = 0.1836735 Score = 4	Count = 2 Normalised count = 0.02597403 Score = 2
Total scores	13	30	27	17	17	21

From the table above, we can conclude that the Eastern Coastal Loop performs the best while the Northern Explorer Loop comes in last. While our evaluation is relatively simple and only considers a few factors, it provides us with a rough idea of what different park connector loops are like and how accessible nearby transportation and amenities are.

For example, the Northern Explorer Loop performs relatively poorly based on our scoring system. However, this is to be expected as this loop takes cyclists through more natural areas such as coastal and forested areas.

7 Conclusion and discussion

One possible significant factor in the poor results from analysing Singapore as a whole is that our data could be inherently biased, because the MRT network has always been designed to bring people from the heartlands to work at the CBD. This radial structure has resulted in a much higher density of MRT exits (and also stations) than areas in the heartlands - refer to KDE/Quadrat map of MRT exits to prove this.

In contrast, park connectors were never meant to provide transport to and from the workplace, but rather are places of leisure meant to provide residents with respite from hectic urban life. Concerted effort is made to “green up” these park connectors to give people the illusion they are moving through a green oasis in the middle of the city. The NParks website describes park connectors as “ a diverse palette of recreational green opportunities, provides avid outdoor enthusiasts with enhanced accessibility to nature spaces across the island”[1] Therefore, by the very intent of the city planners, the park connectors were designed with recreational use in mind and not commuting.

Nevertheless, we believed the utility of these existing park connectors should still be observed. By changing the framing of data to that of regional instead of nation-wide, we were able to elicit information on the distributive trends and provide a framework to give measures of how useful the current placements of exits are. With this, the efficacy of the current placements can be scrutinised to improve the decisions in future construction of park connector exit points towards improving the everyday functionality of these park connectors.

8 Future work

Following our insight to taking the analysis to regional focus, a possible extension to our analysis of bicycle commuting in Singapore include cycling paths in our study. These are mostly *within* towns and not *between* towns, serving to connect different areas in the same neighbourhood (e.g. Yishun or Jurong). Cycling paths are not built randomly across Singapore, but rather certain towns are earmarked for construction of cycling paths, after which the paths are built to serve that town’s needs. Hence, in a similar vein to the regional analysis of PCAPs, it would be more accurate to look at the placement of cycling paths within a single region with existing cycle paths than across Singapore as a whole.

[1] https://www.nparks.gov.sg/gardens-parks-and-nature/park-connector-network

9 References

1. *Parks & Nature*. National Parks Board. (2022, August 26). Retrieved November 11, 2022, from https://www.nparks.gov.sg/gardens-parks-and-nature