Applications of Spatial Data Science

Comparative public bus transportation network and accessibility analysis: a case study for Greater London and Hong Kong

Due: 14 Jan 2022

Student number: 1931393

2500 words

1 Introduction

Cities are the heart of the world, being a nexus of the flow of goods and people. Transportation networks in cities are therefore an important core logistic for businesses and policymakers alike. It is crucial to get better insight into the interplay between urban form and the natural environment.

The aim of this essay is to develop an original, robust, population-aware method of network and accessibility analysis. It compares Greater London and Hong Kong, because their contrasting terrain and population distribution allows a comparison of how accessibility is distributed spatially, relative to the entire population, highlighting the method's flexibility. The network analysis gives insight to the urban structure and how the role of bus routes differs between the two cities. Cumulative accessibility analysis gives insight on how many people have what level of accessibility to buses, and where they are. It concludes that Hong Kong's high population density is hyper-local whereas London is more traditionally distributed like the concentric zone model. Finally, it reveals specific sites of spatial inequality, identifying places in particular need of support.

Hong Kong has a mountainous terrain with few habitable land. This particular spatial distribution of the population is reflected in the transportation network. However, studies like Deltoro Soto et al. (2018) and Prastacos et al. (2019) did not consider the background population distribution of the cities. Oba et al. (2008) and Deng et al. (2019) was aware of the background population distribution and identified specific places in the city that has room for improvement, but did not repeat that analysis for another city, so it is unclear whether it is generalizable. This essay improves on them by comparing two cities. Too many cities will hinder the ability to reveal unique spatial patterns: Song et al. (2017) and Zheng et al. (2012) compared 35 and 20 Chinese cities respectively, so the conclusion had to generalize over all cities, leaving little room to be specific.

2 Data and Methods

Table 1 shows the types of analysis used. The network design gives insight to the urban form of the city, and opportunities to discuss different interpretations. Cumulative accessibility analysis reveals how many people have what level of accessibility to bus transportation, and where those people are. The spatial distribution of accessible stops will be identified by quadrants on a scatterplot, similar to Arranz-López et al. (2019), but with a

Table 1: Methods used

Type	Measure	Presentation
Network design	Average length of bus routes	Scatterplot and histogram
	Average number of bus stops per bus route	
	Average distance between two consecutive stops in a bus route	Boxplot
Cumulative accessibility	Share of population living within a certain distances of a bus stop	Barchart and boxplot
	Spatial distribution of accessible stops relative to population	Scatterplot and map

different threshold, with vector points instead of raster cells, and comparing between two cities.

London is used as a "traditional" city to base the analysis on. The population and bus network of London is relatively uniformly distributed around a single centre. It is "traditional" because of its relative similarity to the concentric zone model, with the centre at the core, the suburbs surrounding the core, and the periphery in the boundary. The TfL fare zone system supports this assumption. Hong Kong is used to compare because its population is distributed very unevenly due to the terrain and the border with Mainland China. It is the differences not the similarities that are the most interesting.

Table 2: Sources of the datasets

Data	London	Hong Kong
Population-points	Facebook Connectivity Lab et al. (2020)	Facebook Connectivity Lab et al. (2019)
Bus network	Transport for London (2021)	Transport Department (2021)

Table 2 shows the datasets used. The bus network data are reliable because it is official data and not crowdsourced. Some code is based on the analysis used in the group project, because I was responsible for the code and analysis in the group project. The code from the group project is mostly cleaning up the data, such as excluding virtual bus stops for London.

For Hong Kong, ferries and trams are excluded because there is no comparable alternative in London. Minibuses are also excluded because some minibus routes are missing rows about intermediate stops, because they are paratransit. Falchetta et al. (2021) has encountered the same limitation as well, so focusing on buses only avoids this problem.

Cumulative accessibility uses the population dataset, which are attached to vector points roughly spaced every 30 m. The population was estimated based on census and population statistics, and is assigned to individual buildings based on high resolution (0.5 m) satellite images.

The population-points (PP) are better than the census because it is possible to calculate a distance matrix between every PP to every bus stop. In contrast, polygons of census tracts will contain multiple bus stops, and it will no longer be clear how to calculate the "distance" between a census tract and a bus stop. The dataset is available for 140 countries, but as an alternative, transforming the tracts into its centroids is a good approximation in return for lower spatial resolution.

Hong Kong has 109955 PPs and 4553 bus stops. Because of memory constraints, the data is chunked and the result incrementally saved into a binary numpy file on the disk. The entire array does not need to be loaded into memory. London has 1044718 PPs and 18919 bus stops. There is not enough storage, so 23247 random PPs are sampled and used to represent the full population dataset, so that the size of the two cities' matrix are roughly the same. It is ran multiple times and the figures are chosen for their rough similarity to the average result of all runs. All calculations involving London's PPs are thus estimates. Further research should either prepare more storage or average the results.

3 Results and Discussion

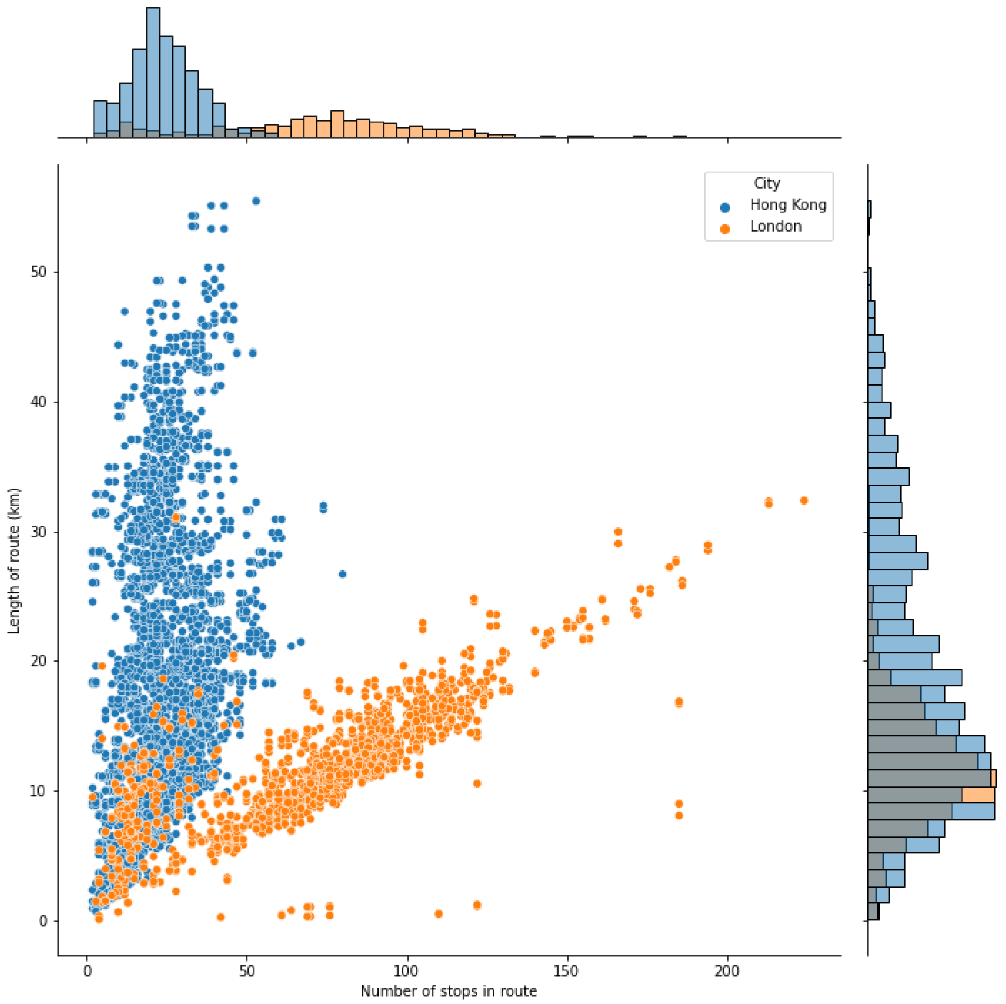

Figure 1: Scatterplot and histograms of the length and number of stops of bus routes

Hong Kong buses have to navigate around difficult terrain, resulting in a more varied role of buses, in contrast to London's more concentric urban form. Figure 1 shows that London's route lengths are quite normally distributed with a relatively symmetrical shape. In contrast, Hong Kong's route lengths are positively skewed with a wider spread, despite having fewer stops on average and a smaller land area than London. Figure 2 shows that the stop separation are also larger in Hong Kong. This suggests that Hong Kong has a more polycentric form. First, bus routes are longer because they navigate around terrain and water bodies. Second, longer routes connect towns surrounded by mountains to the core, the border crossings with China in the North, and the Airport in the West. Hong Kong has short routes just like London, but also has these longer routes.

Figures 11 and 12 support this: the network is more concentrated in central London and dispersed in the boundary, but Hong Kong has important transportation nodes in the

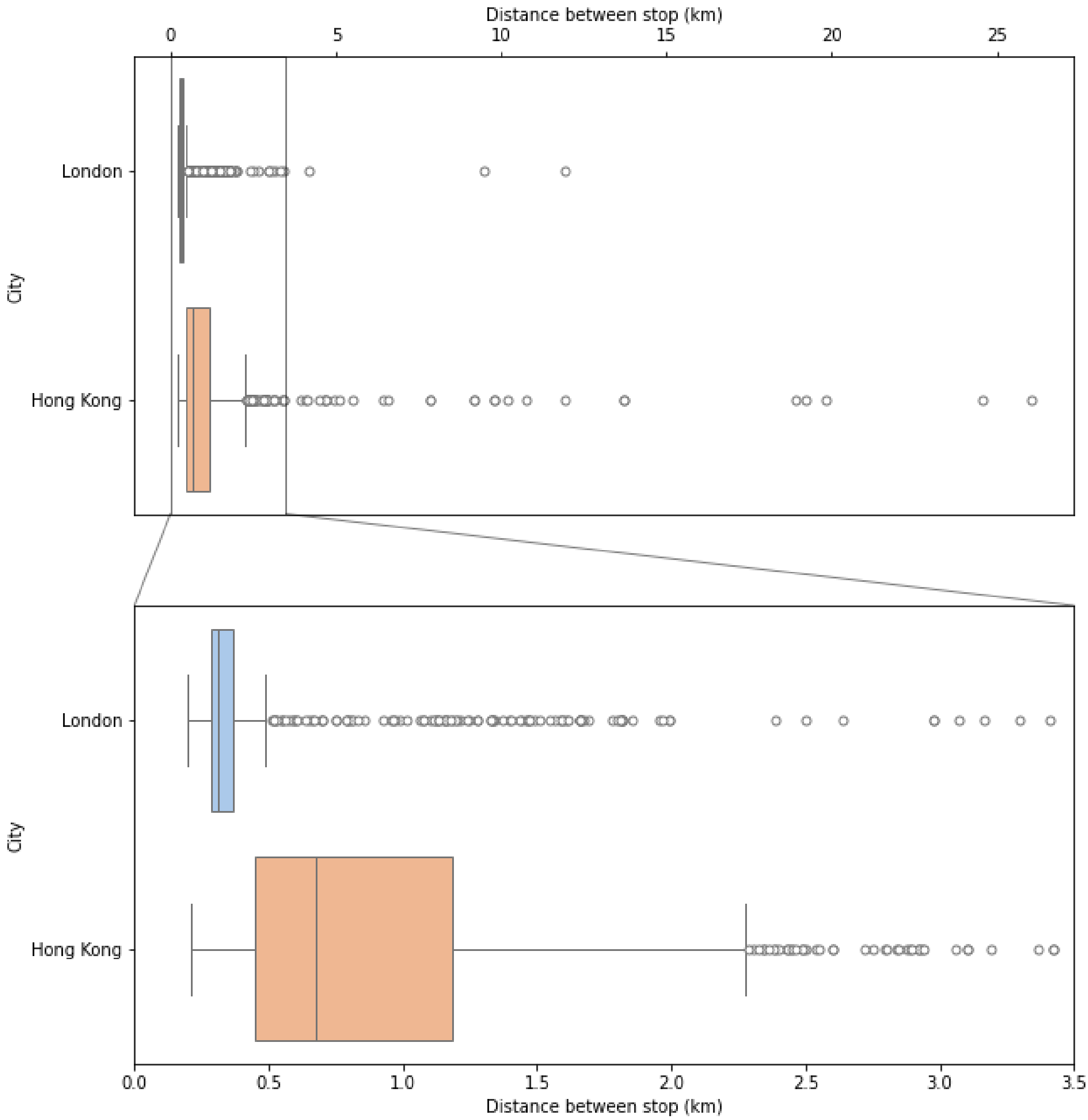

Figure 2: Boxplot of mean distance between two consecutive stops in the same bus route

northern border and scattered towns surrounded by terrain, necessitating a network with a more varied role that connects the core to settlements surrounded by terrain and to the border.

Those maps show straight-line-length, not by-road length, so the actual length is even longer. This is not too important for London, as it roughly approximate the road network. However, Hong Kong has many routes crossing terrain and water, so the by-road length would be larger, and the actual difference between London and Hong Kong is larger. This does not affect cumulative accessibility analysis, because the stop locations are still accurate.

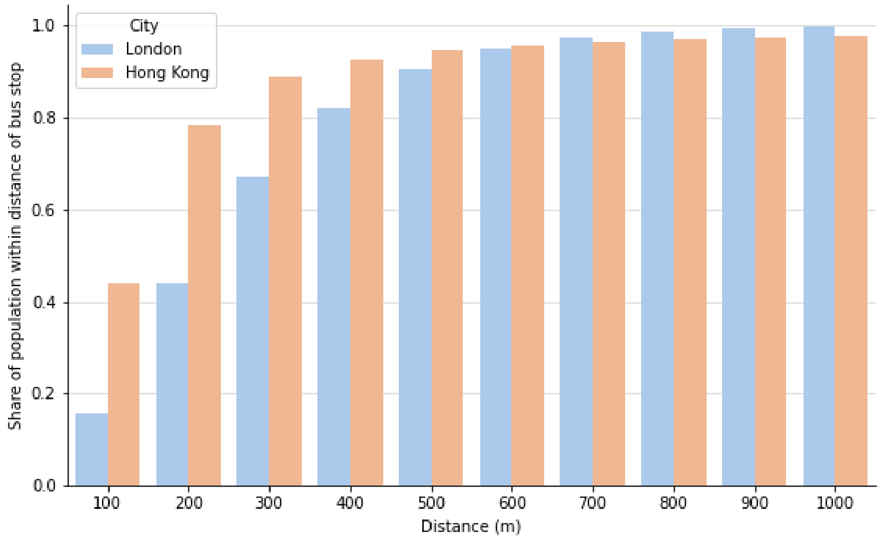

Figure 3: Cumulative share of population within a certain distance of a bus stop

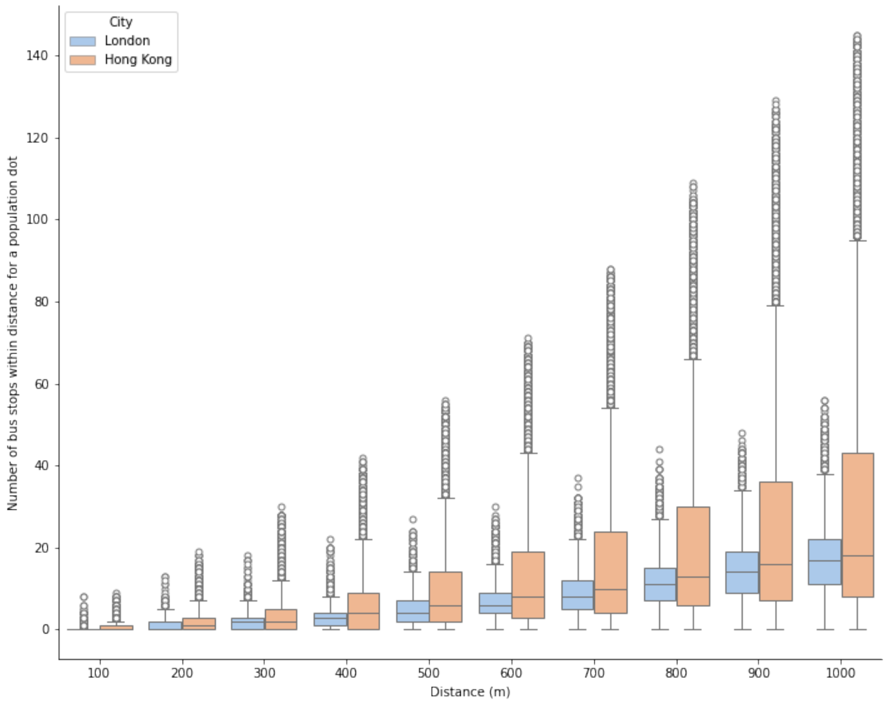

Figure 4: Boxplots of the number of stops within a certain distance of a PP

The density of stops within a route does not necessarily correspond to the density of stops within an area, so a population approach is used. Figure 3 shows that less than 20% of London's population lives within 100m of a bus stop, but more than 40% of Hong Kong's population lives within 100m of a bus stop. A vast majority of Hong Kong's population is covered by its bus network at 200m, while London's just above 40%. This illustrates how Hong Kong's selectively high density in flat terrain benefits bus network coverage. Diminishing returns for Hong Kong start to appear at around 400m, as 90% of the population is already covered. London actually exceeds Hong Kong's population coverage starting at

700m. This could be due to the outlying islands around Hong Kong that are not served by buses. London has no such terrain problems so is still able to increase its coverage. Indeed, Figure 4 shows that accessibility in Hong Kong is more uneven, with larger spread and outliers. Hong Kong's core is denser, but settlements within mountains or islands are less accessible. These numbers broadly agrees with Transport for London (2015) and Corran (2018).

The quality of routes are ignored in this analysis: only one stop within the threshold would suffice. Further research could limit to share of population within a certain distance of a bus stop with at least 2 routes. They can also distinguish between different priorities, for example focusing only on bus stops with routes to the airport or hospitals.

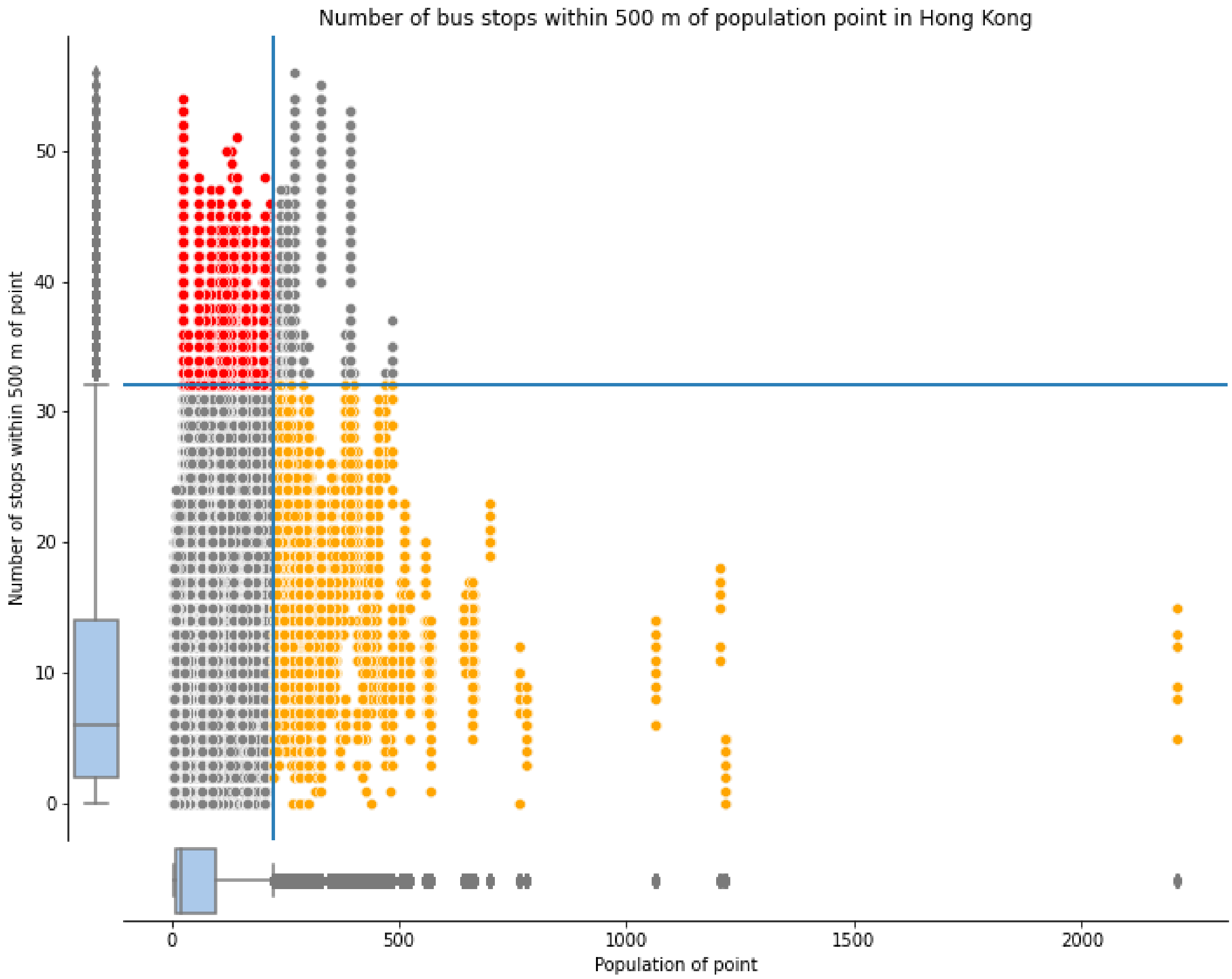

Figure 5: Number of stops within 500m of PPs versus their population in Hong Kong

Figure 5 shows a scatterplot of the Hong Kong PPs, assuming stops within 500m are accessible. Two lines, based on boxplot outliers, divide the scatterplot into four quadrants. The bottom-left and top-right quadrants are not surprising and coloured gray; they are points with a normal population and a normal number of accessible stops, and points with high population and a lot of accessible stops. Points with a normal population but lot of accessible stops are coloured red; these points are relatively over-served. Points with high population but just a normal number of accessible stops are coloured orange; these points are relatively under-served.

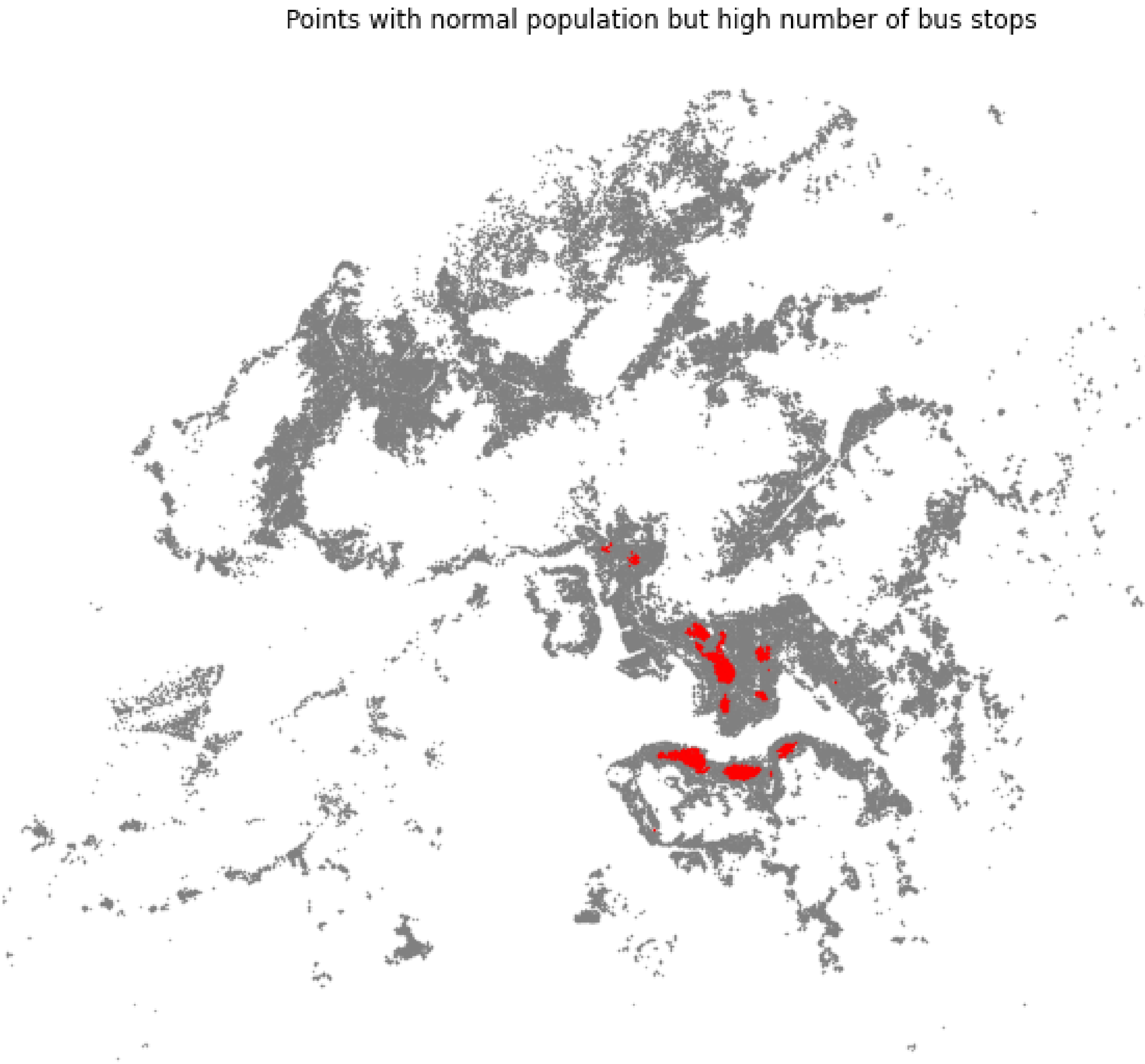

Figure 6: Map of PPs in Hong Kong with normal population but high number of bus stops within 500m

Figure 6 maps the red points. It clearly identifies Hong Kong's core areas. Hong Kong Island has three clusters. The western cluster includes Central and Admiralty; the middle cluster includes Wan Chai and Causeway Bay; the eastern cluster includes North Point. The clusters in Kowloon are dominated by spots on Nathan Road, including all of Tsim Sha Tsui, Prince Edward, Mong Kok, and Sham Shui Po.

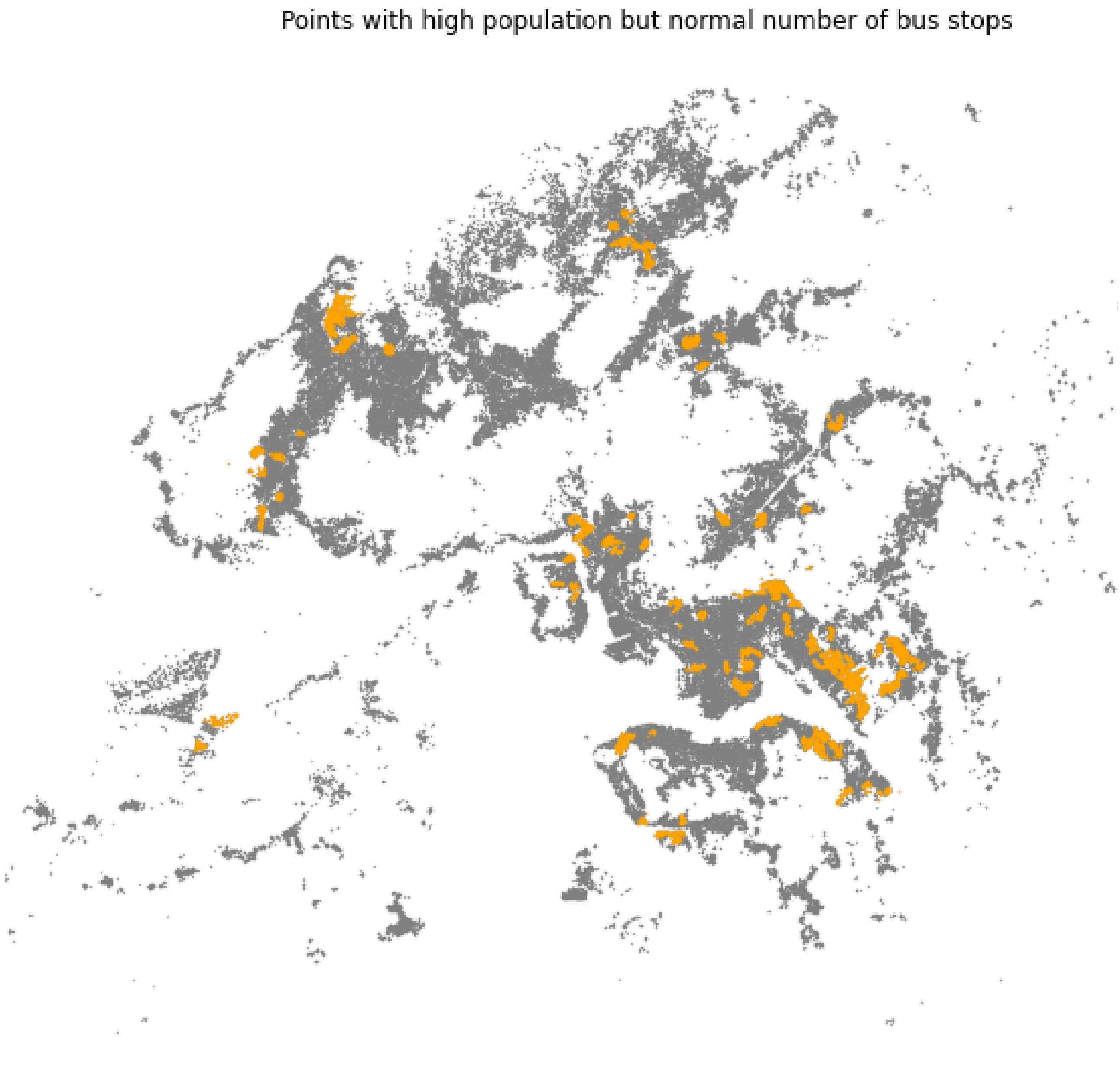

Figure 7: Map of PPs in Hong Kong with high population but normal number of bus stops within 500m

Figure 7 maps the orange points, which are relatively under-served. They mainly appear to be residential areas, which make sense because of their high population: the Sau Mau Ping area in East Kowloon is a highly populated residential area, but due to its difficult terrain, it does not have a metro station. It is thus concerning that the bus network is still under-serving it, when it is an area that needs buses more.

The gray points are spatially concentrated in a few areas. Most of Hong Kong's land area is simply uninhabitable, resulting in hyper-local density, so a global measure of population and stop density is misleading. It explains why bus routes are longer than London, but not the number of stops in the routes: bus routes are longer because routes transport people between uninhabitable terrain, increasing the average distance between consecutive stops.

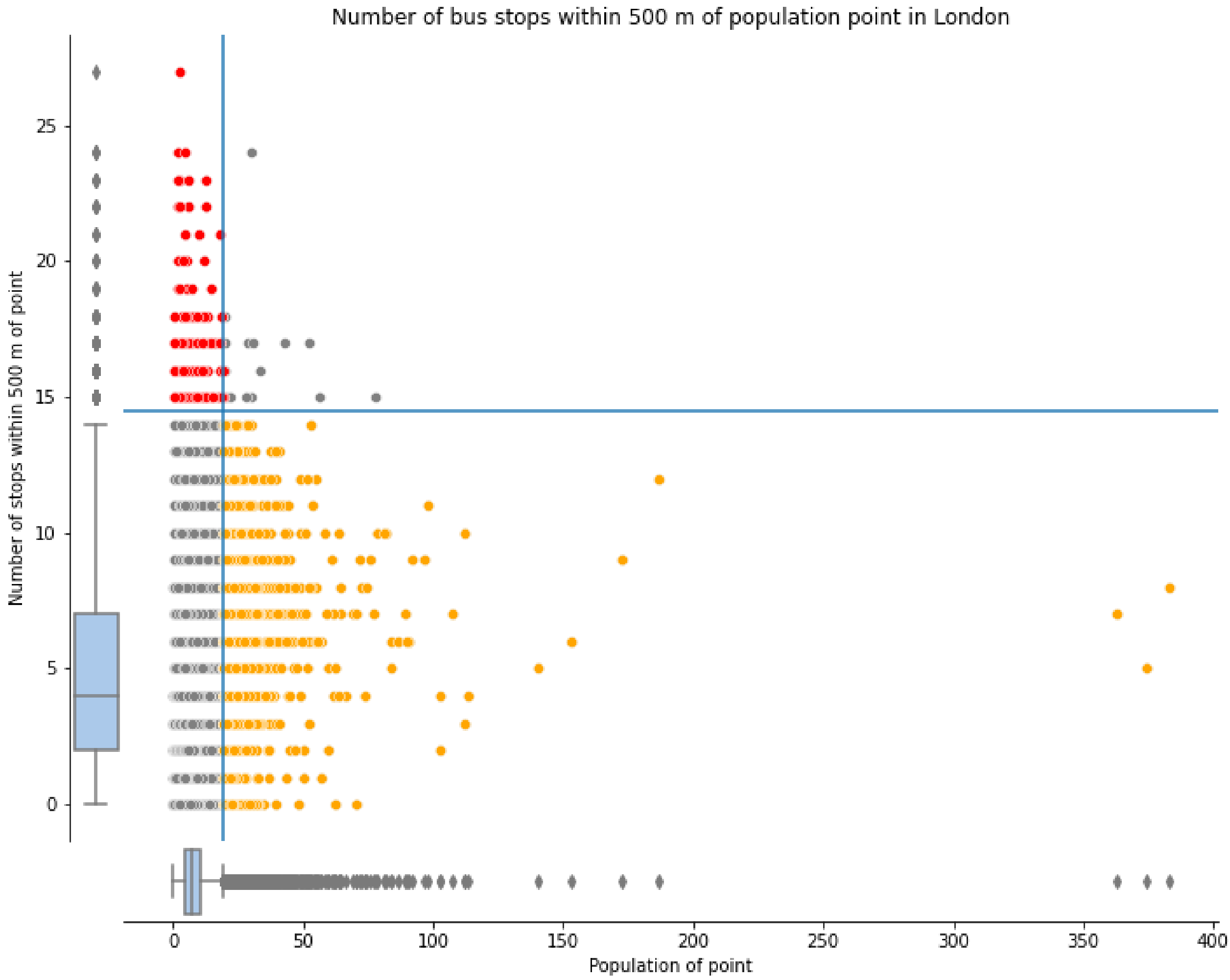

Figure 8: Number of stops within 500m of PPs versus their population in London

The analysis is repeated for London in Figure 8. Note that the scales are different: most of London's dots would be colored grey if they were in Hong Kong, because Figure 4 shows that Hong Kong has more stops within 500m. The population of individual PPs in Hong

Figure 9: Map of PPs in London with high population but normal number of bus stops within 500m

Kong are higher, despite having a lower total population, because PPs represents buildings, which are clustered closer in Hong Kong.

Figure 9 shows relatively under-served areas in London. They are mostly in Zone 2, in between the city core and the periphery. There is less clustering than Hong Kong, but some places do have a slightly higher concentration, such as Upton, West Kilburn, and Finsbury Park, pointing out a possible need for improvement work there.

Figure 10 maps the red points for London. It shows some clustering of core areas. It appears to be less clustered, but this is partially due to the sample size (see Figure 13). Comparing both maps to Hong Kong, the clusters are further apart and less concentrated in the centre of the city. While there are more clusters in the centre, there are some near the boundary, unlike Hong Kong. The gray points are also more evenly distributed around London, as people do not necessarily live near the city centre, unlike Hong Kong. London

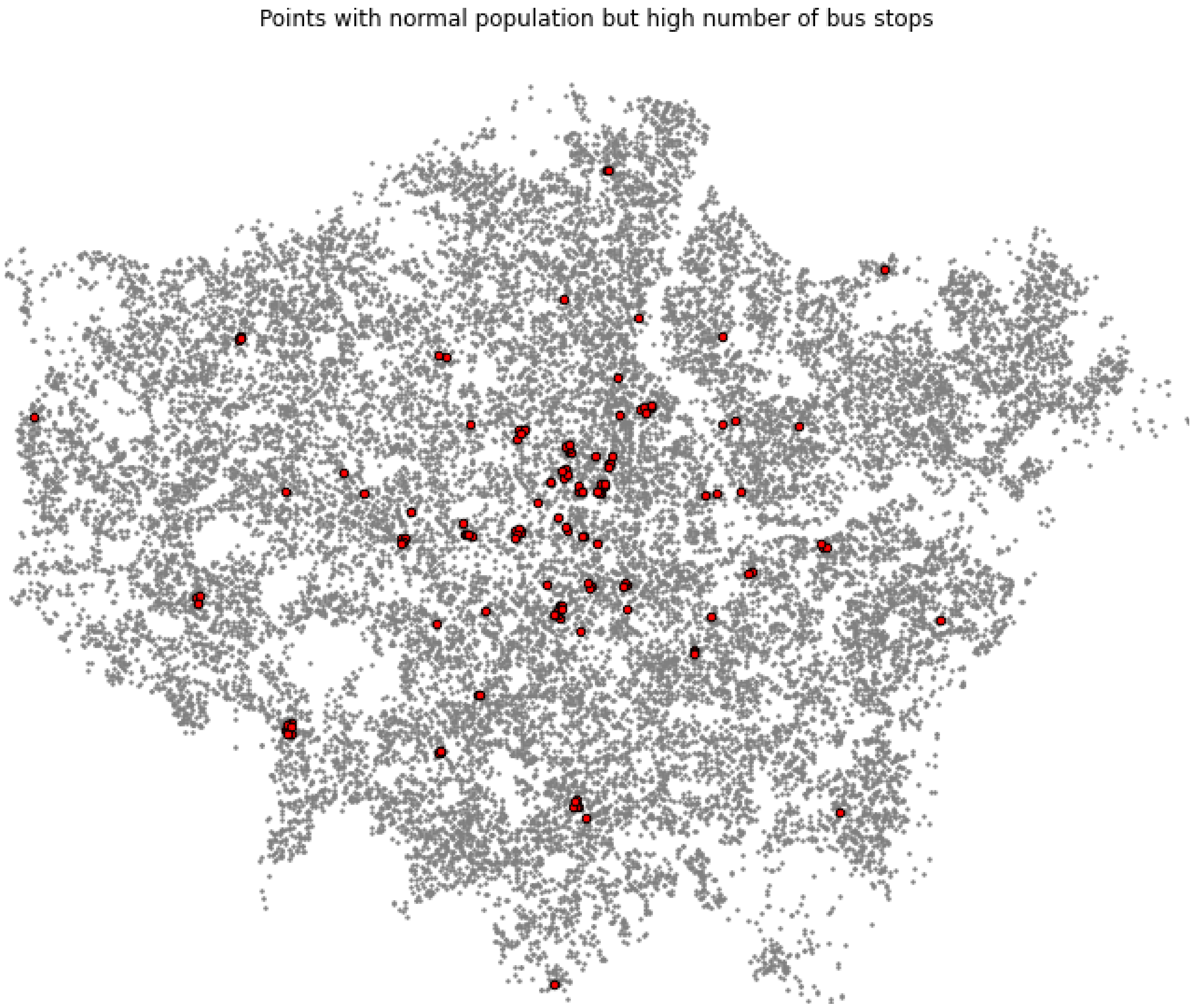

Figure 10: Map of PPs in London with normal population but high number of bus stops within 500m

does not have significant terrain, making it more spatially homogeneous. This comparison suggests the monocentric-polycentric dichotomy is misleading. Hong Kong's strong urban cores are in the centre of the city, but London has bus transport hubs near the boundary as well. London has lower spatial inequality in bus accessibility and its transportation hubs are more evenly spread around the city. Despite Hong Kong's population being spread apart unevenly by terrain, its core is still tightly clustered. The best conclusion is that Hong Kong's high density is spatially concentrated, showing the need for local measures over global measures.

The definition of "accessible" stops for these maps are defined as within 500m of a bus stop, but it is rather arbitrary. According to Figure 3, more than 80% of both cities are within such stops. This is a limitation of cumulative accessibility in general. A continuous gravity-based measure could be used, but it is still unclear how to determine whether a stop is accessible or not.

Furthermore, stops might be counted as accessible regardless of the actual road network. Two stops might be close together but separated by difficult terrain, so in practice they are not close together. The usage of straight-line-distances fail to consider this, and therefore inflate the number of accessible stops.

4 Conclusion

The aim of this essay is to analyse bus networks and their accessibility using a robust, population-aware method. To demonstrate its flexibility and sensitivity, the analysis draws a contrast between London and Hong Kong and found specific places in both cities that needs more investment in public transportation. London's network is more evenly distributed, with bus stop hubs in the centre and near the boundary. In the absence of terrain restrictions, the network was effective in reducing inequalities in accessibility to bus stops. In contrast, Hong Kong's population is limited by mountains, resulting in selectively high population density. The bus network reflects this by having longer routes and stops being spaced further apart. Its unique population distribution helped bus stops to cover 90% of the population within 400m, yet hindered it for hilly terrain and outlying islands, resulting in a larger variation in accessibility. Hong Kong also has a land border, revealing a more varied role that buses play in shaping urban form. The key contribution is to highlight that terrain and the natural environment is inseparable from infrastructure networks, encouraging analysis to be more local-scale, flexible, and generalizable.

Appendix

Source code for this report

`https://github.com/akazukin5151/network-and-accessiblity-analysis`

Source code for the group project

`https://github.com/akazukin5151/comparative-accessibility`

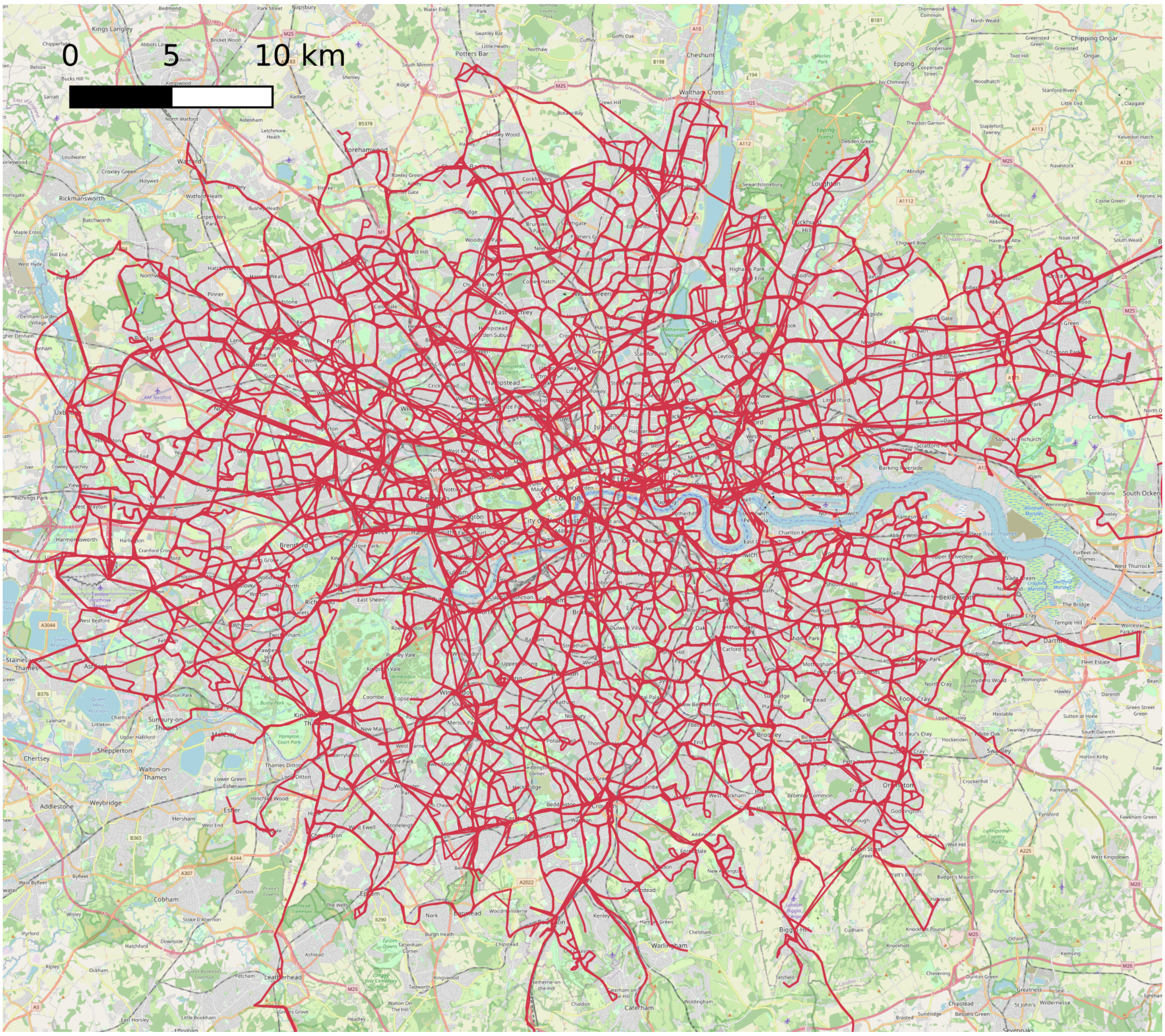

Figure 11: Edges of the London bus network

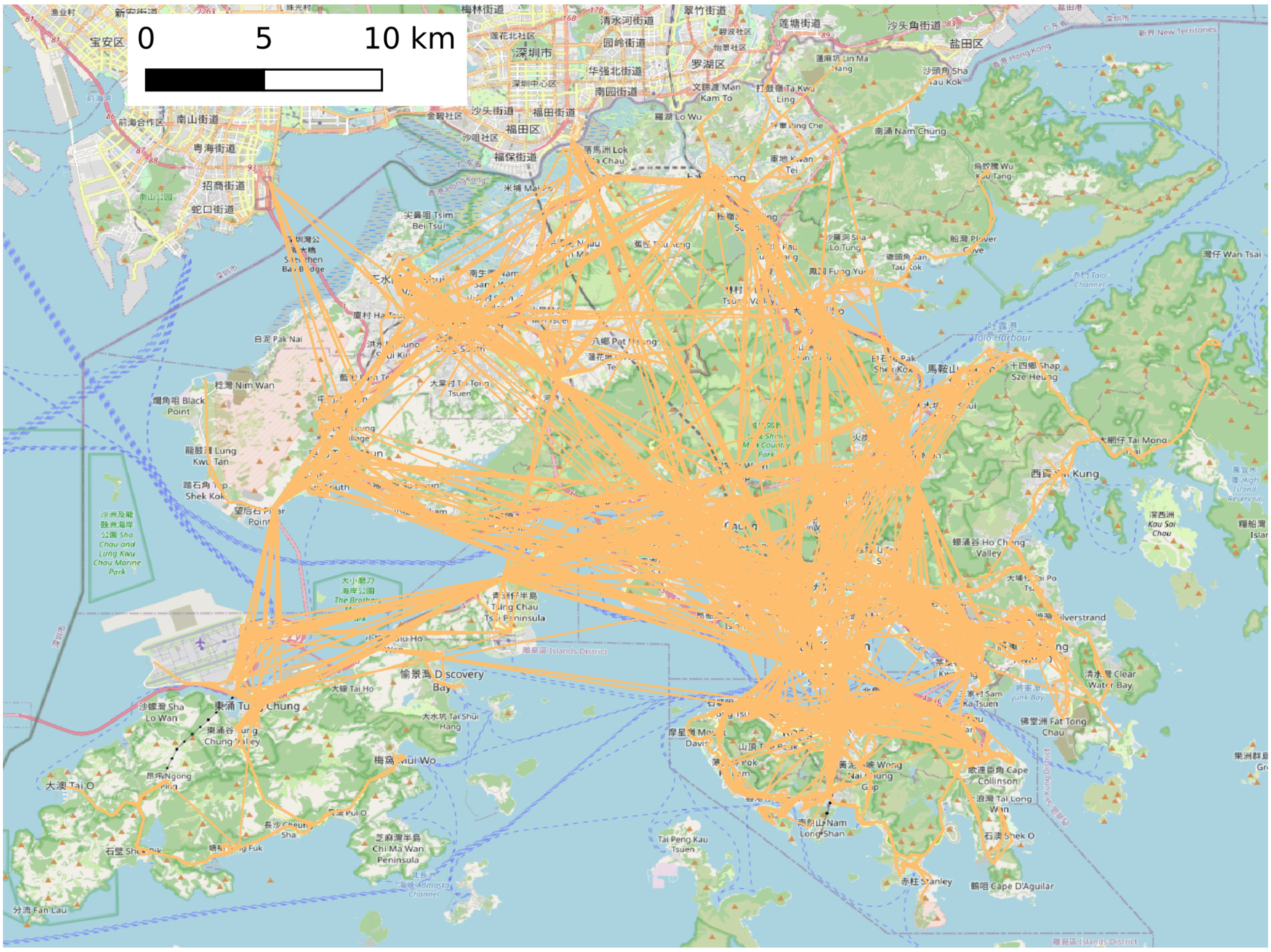

Figure 12: Edges of the Hong Kong bus network

Table 3: City profile. The land area (which excludes water area) is from the Lands Department (2020) and the Greater London Authority (2011), and the population density is calculated using the land area.

City	Population	Land area (km^2)	Population density (km^{-2})
Hong Kong	7,557,179	1110	6808
Greater London	9,079,712	1572	5776

Table 4: Basic network properties

Criteria	Hong Kong	Greater London
Number of routes	1438	774
Number of stops	4553	18872
Number of routes per km^2	1.30	0.49
Number of stops per km^2	4.10	12.0
Number of routes per capita	1.9×10^{-4}	85.2×10^{-6}
Number of stops per capita	6.0×10^{-4}	2.07×10^{-3}

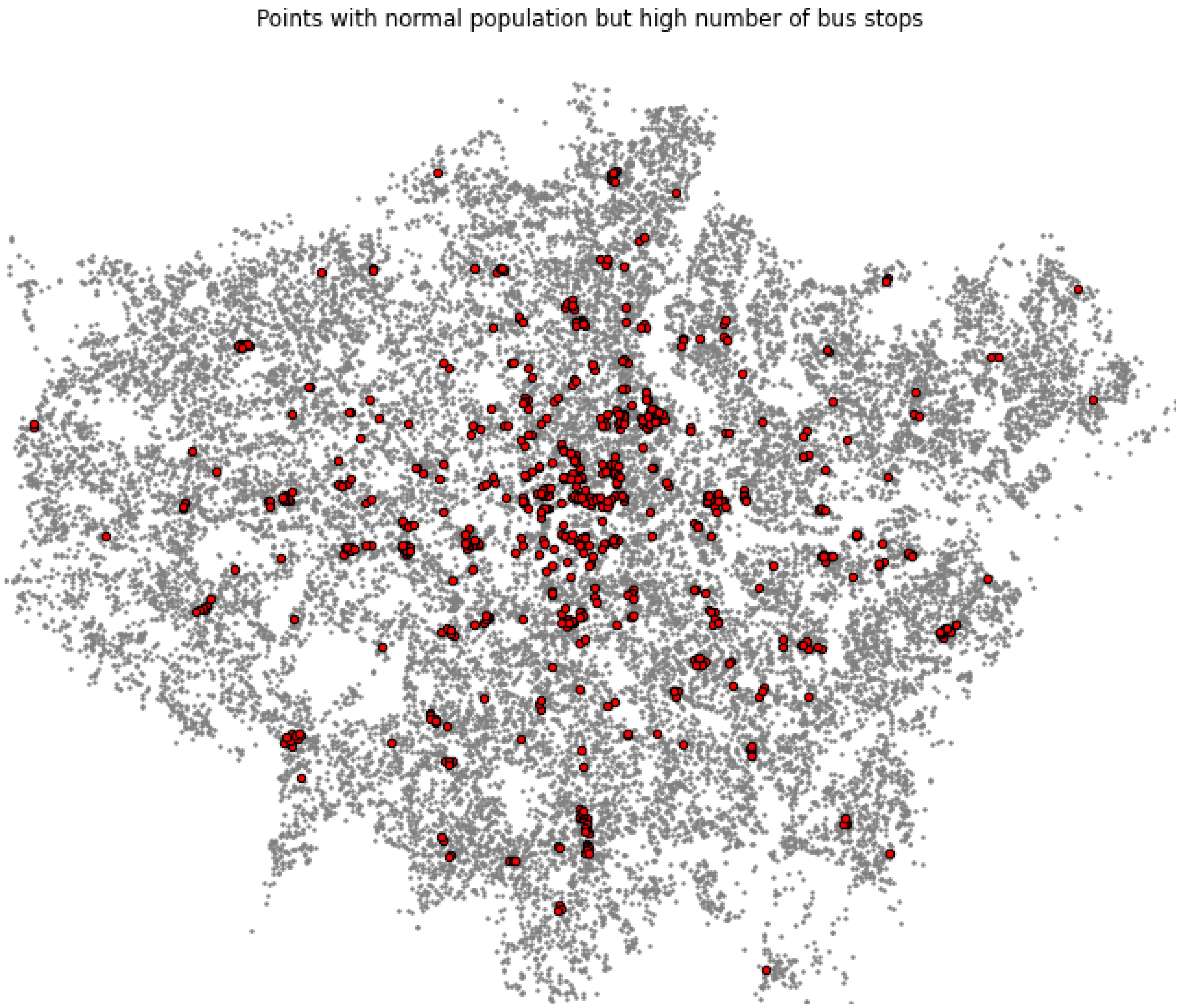

Figure 13: Map of PPs in London with normal population but high number of bus stops within 500m, from another random sample

References

Arranz-López, A. et al. (2019). "Measuring relative non-motorized accessibility to retail activities". *International Journal of Sustainable Transportation* 13.9, pp. 639–651.

Corran, P. A. (2018). *The everyday travel of older disabled Londoners mobility and wellbeing in an urban environment*. Kings College London. PhD Thesis. URL: https://kclpure.kcl.ac.uk/ws/files/116575063/2019_Corran_Philip_1682481_ethesis.pdf.

Deltoro Soto, J., C. Blasco Sanchez, and F. Martinez Perez (2018). "Evolution of the urban form in the British new towns". *24TH ISUF INTERNATIONAL CONFERENCE: CITY AND TERRITORY IN THE GLOBALIZATION AGE*. Ed. by Colomer, V. 24th International Seminar on Urban Form, Valencia, SPAIN, SEP 27-29, 2017, 533–542.

Deng, Y. et al. (2019). "Detecting Urban Polycentric Structure from POI Data". *ISPRS INTERNATIONAL JOURNAL OF GEO-INFORMATION* 8.6.

Facebook Connectivity Lab and Center for International Earth Science Information Network - CIESIN - Columbia University (2019). *Hong Kong: High Resolution Settlement Layer (HRSL)*. The Humanitarian Data Exchange. URL: https://data.humdata.org/dataset/hong-kong-high-resolution-population-density-maps-demographic-estimates.

— (2020). *United Kingdom: High Resolution Settlement Layer (HRSL)*. The Humanitarian Data Exchange. URL: https://data.humdata.org/dataset/united-kingdom-high-resolution-population-density-maps-demographic-estimates.

Falchetta, G., M. Noussan, and A. Hammad (2021). "Comparing paratransit in seven major African cities: An accessibility and network analysis". *Journal of Transport Geography* 94.

Greater London Authority (2011). *Land Area and Population Density, Ward and Borough*. URL: https://data.london.gov.uk/dataset/land-area-and-population-density-ward-and-borough?.

Lands Department (2020). *Hong Kong Geographic Data (Include Area by District Council)*. URL: https://www.landsd.gov.hk/en/resources/mapping-information/hk-geographic-data.html.

Oba, T. et al. (2008). "Effect of urban railroads on the land use structure of local cities". *URBAN TRANSPORT XIV: URBAN TRANSPORT AND THE ENVIRONMENT IN THE 21ST CENTURY*. Ed. by Brebbia, CA. Vol. 101. WIT Transactions on the Built Environment.

14th International Conference on Urban Transport and the Environment in the 21st Century, MALTA, 2008, pp. 437+.

Prastacos, P. and A. Lagarias (2019). "Urban Form and Transportation Infrastructure in European Cities". *DATA ANALYTICS: PAVING THE WAY TO SUSTAINABLE URBAN MOBILITY*. Ed. by Nathanail, EG and Karakikes, ID. Vol. 879. Advances in Intelligent Systems and Computing. 4th Conference on Sustainable Urban Mobility (CSUM), GREECE, MAY 24-25, 2018, 79–88.

Song, Y. et al. (2017). "The Relationships between Urban Form and Urban Commuting: An Empirical Study in China". *SUSTAINABILITY* 9.7.

Transport Department (2021). *Headway information of public transport services*. URL: `https://data.gov.hk/en-data/dataset/hk-td-tis_11-pt-headway-en`.

Transport for London (2015). *Bus stop accessibility doubles since 2008*. URL: `https://tfl.gov.uk/info-for/media/press-releases/2015/march/bus-stop-accessibility-doubles-since-2008`.

— (2021). *Our Unified API - Bus, Coach and River - Bus routes*. URL: `https://tfl.gov.uk/info-for/open-data-users/api-documentation#on-this-page-4`.

Zheng, Z. and Z. Bohong (2012). "Study on Spatial Structure of Yangtze River Delta Urban Agglomeration and Its Effects on Urban and Rural Regions". *JOURNAL OF URBAN PLANNING AND DEVELOPMENT-ASCE* 138.1, 78–89.