The future of community public libraries: How does the accessibility of public libraries within Greater London impact community socio-economic indicators?

Student name: Yicong Li

Date: 2024.08.27

Module title and code: Urban Spatial Science Dissertation (CASA0010)

Supervisor: Fulvio Lopane

Word Count: 8324

Statement: This dissertation is submitted in part requirement for the MSc in the Centre for Advanced Spatial Analysis, Bartlett Faculty of the Build Environment, UCL.

Abstract

The study investigates the relationship between public library accessibility and community socio-economic indicators, such as employment rates and the percentage of people with bachelor's degrees or higher in Greater London. Through spatial analysis and regression models, the research measures accessibility via travel times to libraries across various transportation modes, focusing on walking, public transit, driving, and cycling. The findings reveal a significant negative correlation between longer travel times across different transportation modes and lower employment rates and educational attainment. The study underscores the crucial role of public libraries in supporting community development and suggests the need for improved library accessibility to enhance socio-economic outcomes.

Contents

1. Introduction

1.1 The importance of public libraries

Public spaces and the public sphere, as forms of social infrastructure, play a crucial role in the functioning of society. The concept of the public sphere 'refers first of all to a realm of our social life in which something approaching public opinion can be formed. In principle, all citizens have access to the public sphere' (Audunson et al., 2019). Public libraries serve a significant function within this public sphere. Traditionally, they have been perceived as public buildings primarily dedicated to the collection of printed materials (Appleton and Hall, 2023). However, with the advent of the digital age and increased scholarly attention to public libraries, their influence on the broader community has become increasingly significant. This influence is particularly evident in various socio-economic indicators within the community.

Public libraries demonstrate their importance in various aspects. Traditionally, they serve as vital centers for knowledge acquisition within communities. Library public resources, including books and electronic resources, facilitate individual learning and personal improvement. The public library is considered an important resource for supporting families, especially in underserved communities. Thus, public libraries act as essential educational supplements in these communities (Campana et al., 2019). These educational resources have the potential to contribute to school education at different age levels, the attainment of academic qualifications, and future employment prospects. In the study that examines the relationship between accessibility score and socio-economic indicators, Guo et al. (2017) utilize Hong Kong as a case study. The study assessing spatial inequality in the distribution of libraries across various regions of Hong Kong using multiple methods found that certain socio-economic factors are significantly correlated with library accessibility scores. For example, areas with lower educational attainment levels are negatively correlated with local library accessibility (Guo et al., 2017).

From the perspective of public space, public libraries become more than just reading services, as traditionally perceived, but a major place for community engagement. This

place provides a safe and inclusive environment where community members can communicate and interact, and its community events benefit the community members. Research by Griffin et al. (2022) shows that older adults who participate in a writing project held by the public library conceive writing as a form of purposeful leisure. Healthy relationships (with both self and others) generated from this project have clear implications for combating SI (social isolation). In another study, Philbin et al. (2019) note that public libraries in San Diego, California, also partner with other associations to disseminate HIV/AIDS-related information and provide free computer training to 2500 people, which makes public libraries used as a space to provide targeted public health promotion and even healthcare.

Public libraries have a multifaceted impact on the communities they serve and the residents within them. Socially, public libraries are open to everyone, regardless of social status, economic background, or educational level. This openness symbolizes the public welfare and social equality that libraries offer. Vulnerable groups within the community are a primary focus for public libraries, which reflects their significant social value. In the UK, it is recognized that a cultural divide exists, with those from disadvantaged communities (such as the unemployed or those with limited education) being the least likely to engage in cultural activities (Summers, 2022). Many low-cost community activities and public resources are inaccessible to these individuals, and the advent of the digital age introduces an additional barrier—the "digital divide". Through community activities and engagement, public libraries can provide vital pathways for those most affected by digital, social, and cultural divides. For example, public libraries in Glasgow, Scotland, have maximized the provision of public Internet services, offering resources for searching, sending emails, and job finding (Anderson and Whalley, 2015). In the United States, the American Library Association (ALA) has worked to address digital equity by partnering with companies to offer training courses and providing online resources to public libraries in rural areas with poor internet access, in an effort to bridge the digital divide (Grimes and Porter, 2024). Public libraries play an active role as "places for the production, dissemination, and

adaptation of cultural capital" in helping individuals integrate into the community and broader society. And in the case of emergencies such as natural disasters or pandemics, the digital and online services provided by public libraries play a crucial role in community resilience and sustainable development (Lee, 2024).

However, it is concerning that, over the past two decades, budgets for public services in the Global North have been widely defunding (Schloffel-Armstrong, 2023). In the UK, the implementation of austerity policies in recent years, particularly the reduction of public library budgets in the public services sector, has resulted in systemic damage to the public library system. Consequently, library services have increasingly shifted towards community-led models (Paton and McMenemy, 2024). Many public libraries have been forced to suspend operations due to insufficient funding, a trend that raises significant concerns.

1.2 The concept of accessibility

The concept of accessibility refers to the degree to which transport systems enable people to reach desired activity locations (Neutens, 2015). From the perspective of an individual or community, this means how easy or difficult it is for them to reach a particular location or utilize a specific resource. Accessibility encompasses multiple dimensions, including geographic, economic, and social factors. It is typically related to transportation, time, cost, and spatial distance, which helps to unravel the complex web of causalities between the social impacts of transport systems, human activity, and land use policies (Giannotti et al., 2021). Specifically in the research field of public libraries, the study of this concept has different values for several subjects, such as governmental parts, library operators, and community residents.

For the government, public sectors, and policymakers, spatial accessibility and related indicators can be used to measure disadvantaged subpopulations with higher demand for library resources and services. Such evaluation is useful for urban planners and policymakers to identify under-provision areas, assess the effectiveness of existing urban services-provision policies, and advise on allocating scarce public facilities. For example, how does the choice of address for additional libraries affect spatial

accessibility? The results of spatial accessibility help planners and policymakers to consider the equity situation promptly in their decision-making process (Cheng et al., 2021).

For public library operators, spatial data analysis, such as spatial accessibility, is seen as the most direct link between Geographic Information System (GIS) and Library and Information Science (LIS) (Downey, 2007). In research, Loendorf (2009) extracts library cardholder address information from the Wake County Public Library system, uses data on the distance of these users relative to the library location as an indicator of accessibility, maps the distribution of library users, and assesses the library's market area. These findings can help public libraries to further optimize their operational models and service strategies, such as increasing the number of mobile libraries, extending operating hours, or offering online resources.

Research on public library accessibility is precious for community residents, as spatial accessibility directly affects their quality of daily life and cultural well-being. Such studies are particularly important for low-income residents or those with limited access to transportation. Rosichan (2020) points out that public libraries in low-income and low-education areas provide services that help community members lacking basic literacy skills and educational backgrounds to access essential learning resources and educational support, thereby effectively promoting social equity. Consequently, such research can drive policy and service improvements, ensuring all residents have equal access to public resources and services.

Specific approaches to accessibility measurement have been explored from different perspectives. Allen (2019) analyzes public libraries in Regina, Canada, and highlights that the accessibility of public libraries can vary substantially by time of day, day of the week, and available travel mode. Similarly, Tenkanen and Toivonen (2020) focus on the Helsinki region, where they present a spatial dataset called the "Helsinki Region Travel Time Matrix" and apply a so-called door-to-door approach. This method maintains consistent origins and destinations (OD pairs) across different travel modes, making the travel times calculated for different modes more comparable. The R5py

tool, mentioned in section 3.2.3 of this thesis, incorporates functionalities that are based on the "Helsinki Region Travel Time Matrix".

Given the critical importance of accessibility to public libraries, this study posits that examining the impact of spatial accessibility to public libraries on community socio-economic factors is of significant value. The aim is to draw attention from various stakeholders to the public library sector.

1.3 Research question

Public libraries have a significant impact on communities, making the ease of access for community members a critical issue. In a study on library site decisions, Adkins et al. (2014) note that "The location of a library affects who can access it and who cannot". Building on this insight, this study aims to refine the research focus further by addressing the research question: How does the accessibility of public libraries within the Greater London area impact community socio-economic indicators? This research question is broken down into two objectives. Objective 1 is to conduct a spatial analysis of public library accessibility within communities using different transportation modes. Objective 2 is to examine the relationship between accessibility and two specific socio-economic indicators: the employment rate and the proportion of individuals with a college degree.

For this study, in the process of research conduction, the research methodologies of Allen (2019), Tenkanen and Toivonen (2020) in section 1.2 are highly reproductive and closely related to the present study. Therefore, this research can draw on their approaches and apply them to the study of public library accessibility in Greater London. Also, in section 1.1, although the study area and methods of calculating accessibility differ, the finding of Guo et al. (2017) highlights a potential correlation between accessibility and socio-economic factors and naturally raises the question of whether a similar relationship exists between public libraries and local socio-economic indicators in Greater London, another global metropolis.

Thus, this thesis's study serves as a newly embedded piece in the puzzle of research exploring the correlation between public library accessibility and socio-economic

indicators. This research provides a more comprehensive perspective and deeper understanding of the entire field, thereby rendering the overall picture of the complex relationship between public library accessibility and socio-economic indicators clearer and more complete.

2.Methodology

2.1 Spatial analysis

2.1.1 Different methods used in accessibility

In recent research on spatial accessibility, researchers continuously update their methods for calculating accessibility, driven by advancements in GIS technology and computer science, and select the appropriate method according to the specific research topic. This study uses travel time as a measure of accessibility. In earlier studies on accessibility, such as the research by Kiniki (2004) on how Weber County libraries could better meet community needs through the spatial distribution of library services, the author uses a method that involves adding "buffers" to the map, which are concentric circles drawn at specified intervals from each library location, to determine the number of residents living within a certain distance of each library. Subsequently, Park (2012) argues that straight-line distance does not accurately reflect the actual travel routes of library users. Therefore, to measure library accessibility more realistically, distance measurement based on transportation or road networks should be applied. Park (2012) also uses network analysis to measure the driving distance from users' homes to public libraries and applies the Thiessen polygon method to analyze the influence area in the same thesis. However, when assessing travel distance and time for library users in a multi-destination travel environment, Jae Park (2012) finds that in diverse geographical contexts, travel time has a more significant impact on library access than travel distance.

Therefore, travel time derived from network analysis is used to measure accessibility in this study. When selecting OD pairs, the destinations are chosen to be all public libraries within the Greater London region, with the origins set at the centroids of the LSOAs (Lower Layer Super Output Areas). LSOAs are made up of groups of Output Areas (OAs), which are the lowest level of geographical area for census statistics. Each LSOA usually contains four or five OAs. LSOAs comprise between 400 and 1,200 households and have a usually resident population between 1,000 and 3,000 persons (Office for National Statistics, 2021). The choice of LSOA over the larger

MSOA (Middle Layer Super Output Areas) is made because in the urban context, as spatial units vary in size from smaller areas, such as census blocks, to larger areas, such as census tracts, the accessibility measured in smaller units is less affected by aggregation errors compared to larger units (Apparicio et al., 2017).

It is important to note that this study uses the nearest neighbor analysis method, which calculates the distance between each LSOA centroid and the nearest public library. Each LSOA centroid serves as the origin, and the OD pair is generated based on the minimum distance. The nearest neighbor analysis measures the distance between each feature's centroid and its nearest neighbor's centroid using Euclidean (straight-line) distance (O'Sullivan and Unwin, 2010) rather than actual travel distance or other complex path distances. Therefore, the OD pairs generated by nearest neighbor analysis represent a hypothetical scenario, as the geographically nearest library may not necessarily be the quickest to reach, which is particularly relevant in a complex and well-developed transportation network like the Greater London region, where the shortest straight-line distance does not always equate to the shortest network distance, and varying road conditions at different times can impact travel time. As a result, this assumption may not always be accurate. However, due to limitations in computational resources and time, the initial plan to calculate the actual travel time between each LSOA and every public library and then select the minimum value is not feasible. Therefore, for this step, this study opts to use the nearest neighbor analysis method to generate OD pairs for subsequent analysis.

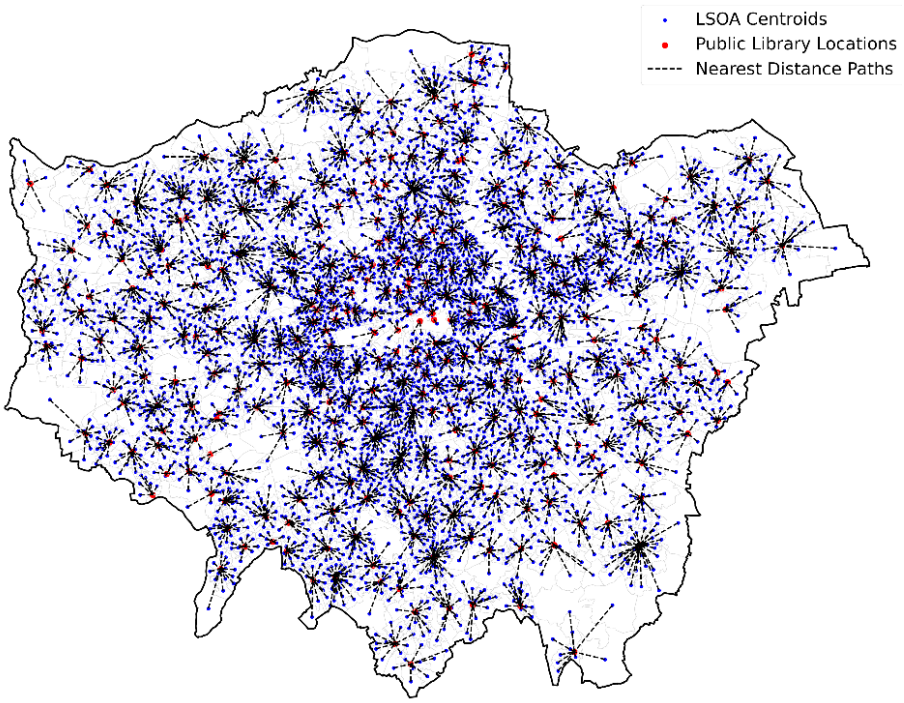


Figure 1: Nearest public library for each LSOA centroid in Greater London

(*Source:* Author's own creation)

2.1.2 The tool of spatial analysis: R5py

The R5py library, a Python tool for transportation network analysis, is used in this study. R5py can calculate accessibility and travel time across multimodal transportation networks, including walking, public transit, cycling, and driving. It serves as a Python interface for the R5 engine, an open-source transportation network engine developed by Conveyal. R5 is known for its efficiency and accuracy, especially when dealing with large-scale urban transportation networks (Fink et al., 2022).

The method for calculating travel time using the R5py library requires OSM (OpenStreetMap) files for network analysis, with the corresponding GTFS (General Transit Feed Specification) public transit schedules as an optional component, making it particularly suitable for studying accessibility in transportation contexts. These two datasets, combined into R5py-compatible variables, allow for a variety of parameter settings and method choices. The most crucial parameters are the origins and destinations, typically provided in the form of latitude and longitude coordinates.

Additionally, users must configure travel parameters, including different travel modes (such as walking, cycling, driving, and public transit) and time parameters (such as a specific departure time). Other optional settings include the time limit for accessibility calculations, distance limits, and the speed of the transportation modes (R5py, n.d.). Based on these inputs, R5py uses the R5 engine to calculate travel times from each origin point to each destination, generating a travel time matrix. From this matrix, users can further calculate accessibility metrics, such as the number of destinations reachable within a certain time threshold or the average travel time to reach facilities (e.g., libraries and hospitals).

The choice of R5py as the research tool is justified by its ability to rapidly calculate travel times, indicating accessibility within large transportation networks, making it a valuable tool for transportation network analysis. It supports the integration of multiple travel modes, allowing for more flexible and precise analysis, particularly in studies that incorporate both public transit and other modes of travel. The new R5py builds upon the ideas and concepts presented by Tenkanen and Toivonen (2020): its scope expands, the entire routing engine and methodology are revamped, and the modeling is based on new, more realistic measurements of walking, cycling, and driving speeds, all while maintaining backward compatibility for comparability (Fink et al., 2024). R5py simulates travel scenarios at specific times, such as during peak hours, providing more realistic travel time predictions. The path and time calculations offered by R5py are highly detailed and capable of simulating specific routes under particular travel modes and parameters (R5py, n.d.), with time accuracy down to the minute, making it ideal for research requiring fine-grained analysis.

2.1.3 Comparison with other methods

Compared to R5py, traditional static methods typically consider only straight-line or simple network distances for path calculations (O'Sullivan and Unwin, 2010), often neglecting real-world traffic conditions and the complexities of various travel modes. Models and methods derived from the gravity model of spatial interaction are also commonly used to calculate spatial accessibility. These models are relatively simple

and offer intuitive explanatory power. However, their effectiveness largely depends on the selection and calibration of parameters during actual model computation, with different parameter settings potentially leading to significant variations in results. For example, in the study by Stępniak and Rosik (2018), the choice of specific parameters for the distance decay function in the model can affect the extent of accessibility improvement. Depending on the parameters applied in the accessibility model, the same change in accessibility could lead to opposite conclusions. The greatest advantage of R5py over other accessibility calculation models is its ability to generate travel time matrices that incorporate multiple travel modes for any OD pairs. Notably, R5py allows for the integration of GTFS files, providing public transit schedules down to the exact year, month, day, hour, and minute. This advantage enables the consideration of factors such as transfers and waiting times across different modes, resulting in travel time calculations that are more explanatory and comprehensive compared to other models (Stępniak and Rosik, 2018).

The Two-step Floating Catchment Area (2SFCA) method is the most widely used and comprehensive method for assessing accessibility in recent years (McGrail, 2012; Pei et al., 2022). This method is an improvement over the traditional gravity model, designed to simultaneously consider both the supply of services (such as the number of service facilities) and the demand (such as the population requiring these services) to assess the accessibility of specific services (Higgs et al., 2013). The 2SFCA method is widely applied in evaluating the accessibility of healthcare facilities, and it can also be used to assess the accessibility of public services such as public libraries, schools, and parks.

As research progresses, scholars have introduced several enhanced versions to address the limitations of the original 2SFCA method. For instance, the Enhanced Two-Step Floating Catchment Area (E2SFCA) (Huhndorf and Działek, 2018) method incorporates a distance decay function, the Three-Step Floating Catchment Area (3SFCA) (Takyi et al., 2023) method factors in competition effects among service points, and the Weighted Two-Step Floating Catchment Area (W2SFCA) (Bozorgi et

al., 2021) method assigns different weights to service facilities based on differences in quality, scale, or capacity. These versions are suitable for different research needs and can help researchers more comprehensively understand and assess the spatial accessibility of service facilities (Kiani et al., 2021).

The primary reason for not using this method in the current study is that the 2SFCA and its improved models often incorporate a comprehensive set of socio-economic indicators, such as population density, income levels, and education levels, as parameters in the gravity model. This inclusion makes the calculated accessibility scores more reflective of public library accessibility. However, this study aims to examine the relationship between accessibility and these socio-economic indicators. Since the accessibility scores are partially calculated using the same socio-economic indicators, which also serve as dependent variables in the regression model, this could lead to high correlation coefficients between accessibility and these indicators. Such high correlations would result in model instability and reduce the explanatory power of the regression model presented in section 3.4.

2.2 Socio-economic indicators: “Employment Rate” and “Qualification”

Among various socio-economic indicators, this study focuses on an education-related indicator, "the proportion of the population with higher education degrees," and an employment-related indicator, "employment rate," as both indicators directly reflect the economic status of a region. (Kong et al., 2022) point out that although the global prevalence of higher education remains limited, it has been shown to play a significant role in driving economic growth. Additionally, (Zhang et al., 2023) and Burggraeve et al. (2015) highlight a positive feedback relationship between the employment rate and GDP. Therefore, examining the relationship between public library accessibility and these two indicators can most directly reflect the relationship between public library accessibility and the community's economic status.

Books, research databases, and quiet study areas are essential for school-aged children, so Bhatt (2010) studies the impact of library use on minors' reading habits

and academic performance, and the results of Bhatt's thesis indicate that families living closer to libraries are more likely to use them than those living farther away. Therefore, the accessibility of public libraries seems to have a positive impact on students' study habits and academic performance. The authors also suggest that local governments consider improving library accessibility to promote community education and enhance social welfare.

In terms of improving employment rates and educational attainment among adults, some scholars conduct multifaceted studies on the role of public libraries in the United States. Cole and Stenström (2021) find in their study of California public libraries that these libraries promote individual economic development by providing opportunities and support for skills development, job search assistance, connections to other social services, and small business growth. In studies of rural areas in the United States, Strover et al. (2020) point out that public libraries, as public infrastructure providing broadband and bridging the digital divide, are often closely linked to improved economic outcomes, such as employment growth and reduced unemployment rates. In research in the field of public libraries in the UK, Rooney-Browne (2011) uses the RIMS II model to measure the impact of libraries on employment and the local economy. She finds that the Suffolk Cooperative Library System (SCLS) expenditures create 1,200 jobs in the local economy. Additionally, public libraries significantly impact community education levels by supporting educational efforts, such as offering courses for individuals who have not obtained a high school diploma or equivalent.

As information technology and the internet become widely used in various workplaces, people's work and business methods are rapidly changing. On one side, this shift allows for greater flexibility and efficiency in how, when, and where people work; on the other side, it requires individuals to possess different skills than previous generations to achieve economic success (Taylor et al., 2012). Many public libraries have keenly recognized the importance of providing knowledge, skills training, and employment services through their institutions, and they offer these services to their communities (Bertot and Jaeger, 2012). Consequently, the accessibility of public

libraries within and around communities plays a crucial role in enhancing knowledge, skills, and employment opportunities for many community members in the new economic era. This subject of inquiry is also the objective and value this study aims to achieve by extensively exploring the relationship between public library accessibility and these socio-economic indicators.

3. Data sources and analysis

3.1 Research area: public libraries in Greater London

This research focuses on Greater London’s public libraries. London is not only the capital and largest city in England and the United Kingdom but also one of the largest cities in the world. Its metropolitan area is the largest in Western Europe, with a population of 14.9 million (CityPopulation, n.d.). The map below shows where London is located, and the area covered by the Greater London area.

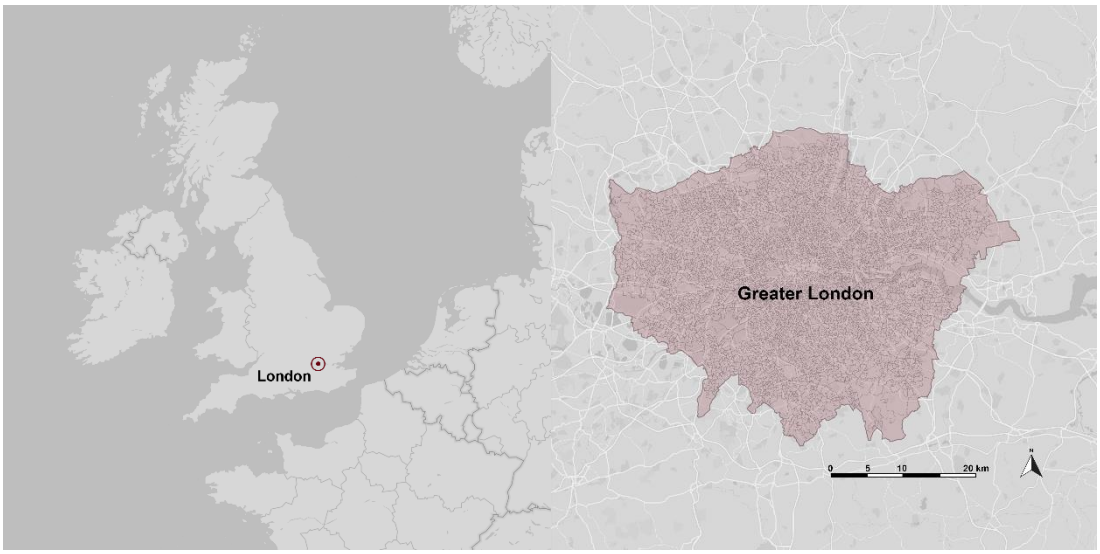


Figure 2: Research area: Greater London, UK (*Source:* Author’s own creation)

London's urban form is characterized by high-density development in the central areas, with concentrated commercial, office, and residential spaces, contrasted with low-density suburban areas dominated by residential neighborhoods and green spaces. This difference in density reflects the city's morphology, which gradually transitions from a dense core to more dispersed outer areas. Regarding transportation accessibility, London has a large and complex public transportation network, including the underground, trains, and buses. The extensive coverage of London’s underground and overground networks allows residents to move conveniently throughout the city. However, London's surface transportation has significant issues. London is one of the

cities where congestion has the greatest impact on automobile accessibility (Moya-Gómez and García-Palomares, 2017).

From the perspective of public libraries, the Cultural Infrastructure Map published by the Greater London Authority (GLA) (Greater London Authority, 2023) indicates that the number of public libraries in the Greater London area increased to 347 in 2022, up by two from 2019. The distribution of public libraries in London follows the general pattern observed in most cities: there are more library facilities in the metropolitan and densely populated boroughs of London compared to non-metropolitan and less populated areas (Adkins et al., 2014). The distribution of public libraries in the Greater London area is illustrated in Figure 3. Research on London's public libraries is not only relevant to the UK but can also be extrapolated to other global megacities.

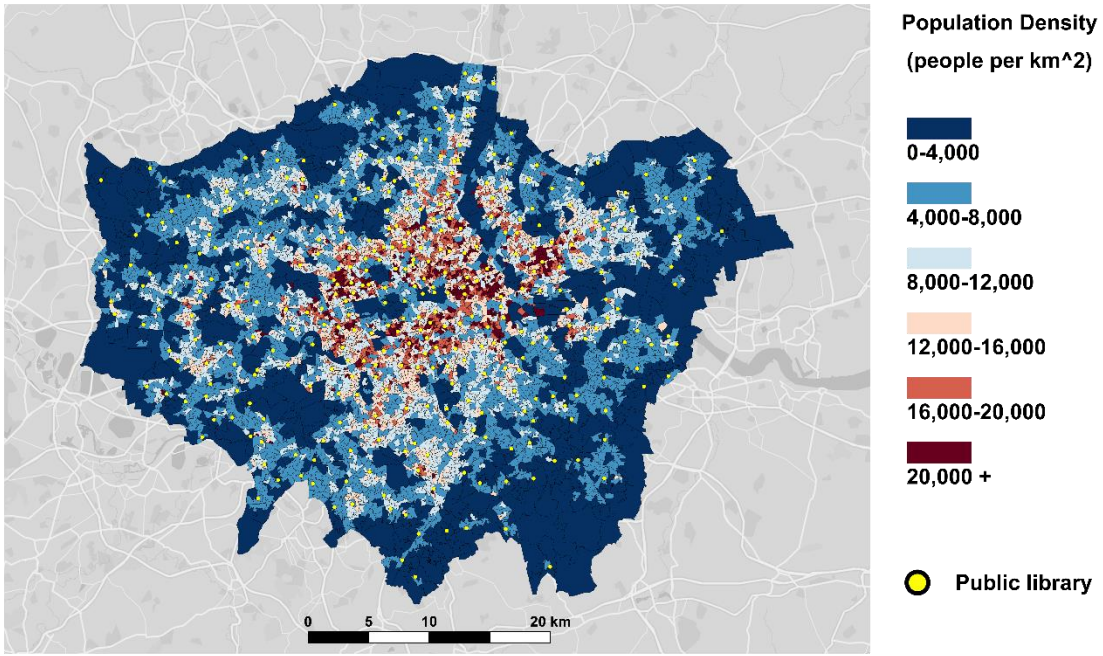


Figure 3: Public libraries and population density in Greater London

(*Source:* Author's own creation)

3.2 Data resources

This study utilizes data from various sources, as shown in Table 1. All files and input data used in this research are either open or derived from open data. The Python

package R5py and the underlying Java library R5 are both released under free and open-source licenses. The code for this study is written and executed using Anaconda and is published on GitHub (see Code Availability below). Therefore, the methodology of this study is highly portable, making it easy to replicate and apply to metropolitan areas in other cities.

To clarify, all data usages and tools in this study are publicly published; therefore, they are all low risk in ethical considerations. Using low-risk data is crucial to ensure the privacy and confidentiality of any personal data obtained from public resources or datasets, which includes anonymizing data where necessary and obtaining permission to use data where direct interaction is involved.

Data sources	**Publisher of public data sources**	**Website**
London LSOA boundaries	Office for National Statistics (ONS)	https://geoportal.statistics.gov.uk/
Public library location	Cultural Infrastructure Map-Greater London Authority (GLA)	https://data.london.gov.uk/dataset/cultural-infrastructure-map-2023
London Network	Geofabrik download server-OpenStreetMap (osm.pbf)	https://download.geofabrik.de/
GTFS	UK government official website-"Bus Open Data" service	https://www.bus-data.dft.gov.uk/
Population density	Office for National Statistics (ONS)-	https://www.ons.gov.uk
Socio-economic indicators	Office for National Statistics (ONS)-Census 2021	https://www.ons.gov.uk/datasets/create

Table 1: Data sources used in this study (*Source:* Author's own creation)

3.2.1 Map Data: LSOA geojson file

The boundary data for Greater London are derived from the Lower Layer Super Output Areas (LSOAs) within the whole of England, as updated by the Office for National

Statistics (ONS) on its official website (Office for National Statistics, 2024) in June 2024. After filtering the data, the original boundary data retain only the LSOAs within Greater London. LSOAs are one of the geographic units used by the ONS to collect and analyze statistical data and usually cover smaller communities or regions. Each LSOA encompasses approximately 1,000 to 3,000 residents. LSOAs are widely used in socio-economic research and government policy-making, particularly in census data collection, public service allocation, and regional inequality analysis. Once the filtering process is completed, 4,994 LSOAs are identified in Greater London, the total geographic area considered in this study.

3.2.2 Library sites

The geographic information data for public libraries in Greater London are sourced from the Cultural Infrastructure Map published by the Greater London Authority (GLA) (Greater London Authority, 2023). The specific dataset is included in the map's accompanying folder.

3.2.3 Network map (osm) and GTFS

In the R5py calculations, a class called "TransportNetwork" is constructed to estimate the travel time from origins (LSOAs centroids) to destinations (public library sites) using various modes of transportation. The "TransportNetworkis constructed based on data representing street network segments and transit schedules.

The network data is sourced from OpenStreetMap (OSM) and is stored in PBF format. It contains geographic information for the Greater London area, including roads, buildings, natural features, and more. Compared to government or other proprietary data sources, OSM data offers extensive global coverage and is continuously updated by a global community of users, often providing more timely and detailed information than government data.

Transit schedules are sourced from the UK government's "Bus Open Data" service (Department for Transport, 2024), provided in the General Transit Feed Specification (GTFS) format. GTFS is a standardized format used to describe public transit systems' routes, schedules, stops, and other information. Data for the Tube system is

included in the GTFS dataset, which is released alongside other public transit modes, such as buses and trains.

The generated "TransportNetwork" can be used with other parameters to calculate travel times and detailed itineraries under different modes of transportation—such as public transit, driving, cycling, or walking—at various times, from multiple origins to multiple destinations. For further details, refer to the R5py documentation and the open-source code on GitHub (R5py, 2024).

3.2.4 Socio-Economic indicators data

After calculating the accessibility metric "travel_time," this study links it to relevant socio-economic factors for regression analysis to explore the underlying relationships. The dataset for socio-economic factors is sourced from the Census 2021 data available on the ONS website (ONS, 2024). Census 2021 is the eleventh decennial census conducted by ONS, aimed at providing a comprehensive understanding of the socio-economic characteristics of the UK and its residents. The census collects detailed information on various aspects, including population, housing, education, employment, health, ethnicity, religion, and language. The data for the Greater London Metropolitan Area is crucial for this study.

The ONS website offers a "Create a custom dataset" service (ONS, 2024). For this study, data is selected by choosing "All usual residents" as the population type, "Lower Layer Super Output Areas" as the area type, and "England and Wales" as the coverage, focusing on the specific region, Greater London. Finally, qualifications and employment rate are chosen as the variables of interest for the research.

3.3 Spatial accessibility analysis

Table 2 and Figure 4 compare the differences in multiple travel modes, particularly between private cars, public transport, cycling, and walking. For this study, 14:00 on Tuesday is chosen as the specific time point for the R5py library to build the class ‘TravelTimeMatrixComputer’. This is because, normally, all public libraries are open at this time of the week, and public transport is fully operational (Allen, 2019).

Travel mode	% of population by 5 min intervals						
	0-5	5-10	10-15	15-20	20-25	25-30	30+
Walk	5%	15%	23%	22%	15%	9%	10%
Public transit	5%	15%	23%	23%	15%	9%	10%
Car	77%	21%	1%	< 1%	< 1%	< 1%	< 1%
Bicycle	27%	46%	19%	6%	< 1%	< 1%	< 1%

Table 2: Travel time (in minutes) to the nearest library comparing by travel mode.

(*Source:* Author's own creation)

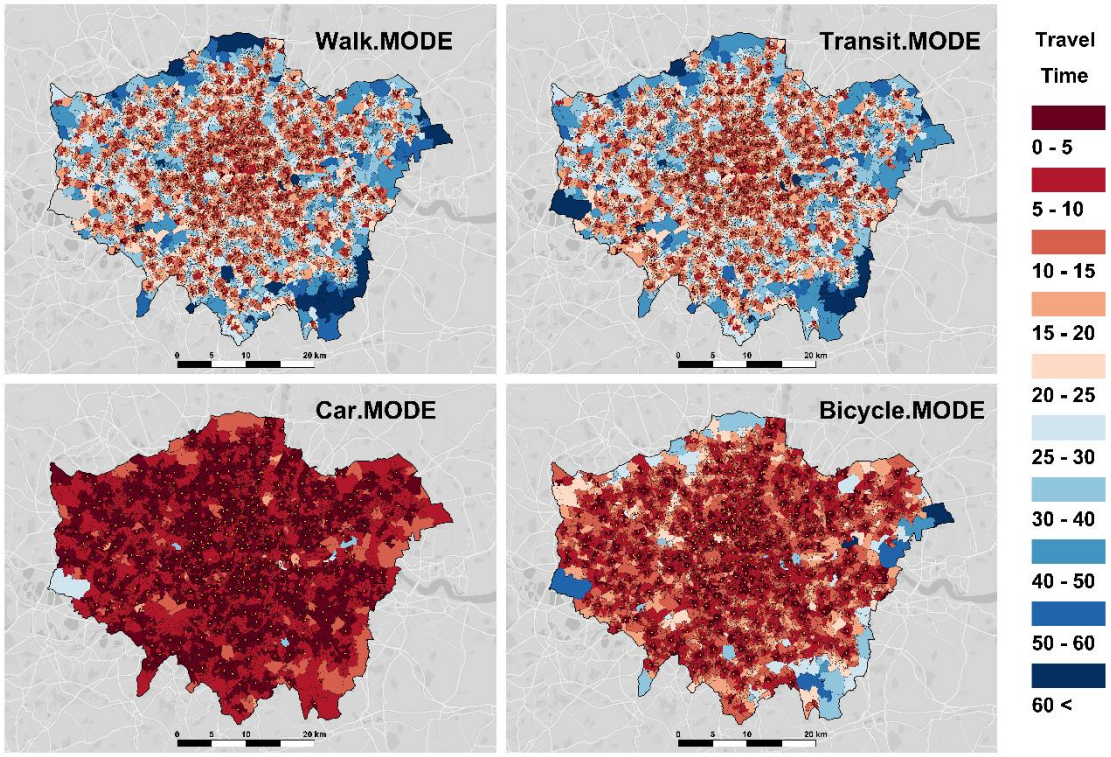


Figure 4: Comparing spatial accessibility (travel time) to public libraries by travel mode.

(*Source:* Author's own creation)

Table 2 shows the average travel time to the nearest public library using four different modes of transportation. The travel time for each mode is mapped in Figure 4, displaying the travel time from each LSOA to its nearest public library. The travel time for CAR.MODE is significantly lower than that for WALK.MODE and

TRANSIT.MODE, indicating that individuals with access to a car can reach public libraries more easily than those without a car.

Interestingly, cycling is more convenient than using public transit. This conclusion closely aligns with results from other studies on public library accessibility in urban areas (Allen, 2019). These figures clearly illustrate that people who rely on public transportation or walking have lower travel convenience than those with private vehicles. Moreover, the disparity in travel convenience is particularly pronounced for those who cannot drive or cycle (Allen, 2019).

Before calculating travel time, the Nearest Neighbor Analysis mentioned in Section 2.1.1 is used. In this study, this analysis method is not further employed to determine the distribution pattern of libraries, such as random, clustered, or uniform distribution. Instead, it is used to pair LSOA centroids with the nearest public libraries to OD pairs.

During the travel time calculation, an interesting phenomenon emerges: the travel times for WALK.MODE and TRANSIT.MODE appear nearly identical on the map. Analysis of the data frame reveals that out of the 4,994 OD pairs generated, only 217 have different travel times, while the rest are the same. The R5py library and its functions are verified to operate correctly, as demonstrated by the global accessibility analysis for a single library address. This analysis calculates the travel time for all LSOA centroids and finds the coverage area for TRANSIT.MODE is significantly larger than that for WALK.MODE under the same parameter setting of a maximum travel time of 120 minutes.

According to the R5py documentation, if reaching the destination directly (without using public transit) is faster, R5py just opts for WALK.MODE and calculates the time to walk directly to the destination. Since TRANSIT.MODE includes GTFS data, a so-called “door-to-door approach”, just shown in Figure 5, which is used when calculating the route between the origin and destination, considering every step of the journey—such as walking from the origin to the station, waiting for the bus, transferring between vehicles, and walking from the station to the destination (Tenkanen and Toivonen, 2020).

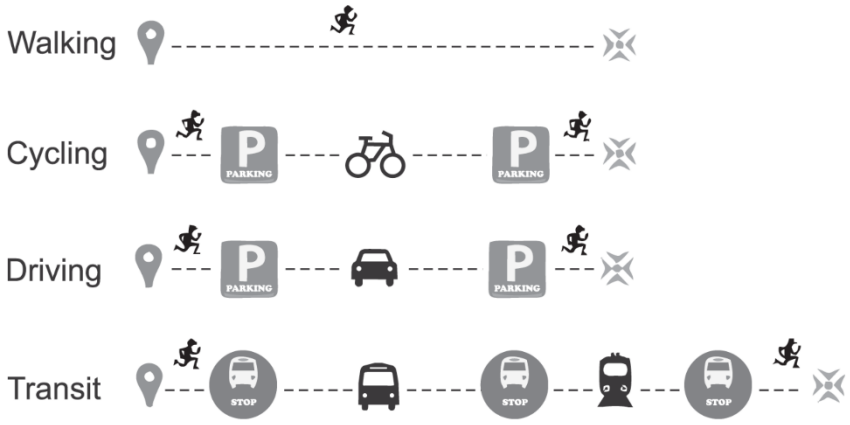


Figure 5. Door-to-door approach where every step of a journey in different travel modes is considered. *Source:* Adapted from Tenkanen and Toivonen (2020, p. 77).

The prior use of Nearest Neighbor Analysis results in OD pairs that are very close in distance, meaning that in most cases, the travel time calculated by TRANSIT.MODE is likely replaced by WALK.MODE, leading to the observed high similarity. This interpretation could also explain why BICYCLE.MODE generally shows lower travel times than TRANSIT.MODE in Figure 4. Consequently, this study ultimately merges WALK.MODE and TRANSIT.MODE, using their combined travel time as a single variable for further research.

Additionally, during the calculation of WALK.MODE and TRANSIT.MODE, the travel time for four LSOAs returns a "null" result, indicating that no travel time is obtained. To understand why this issue occurs, this study first calculates the estimated travel time using the straight-line distance obtained from the Nearest Neighbor Analysis and the default walking speed of 3.6 km/h in the R5py library. Subsequently, each of these four OD pairs is examined individually using Google Maps to calculate the travel time for the corresponding OD pairs. It is found that the travel times on Google Maps are significantly longer than the estimated travel times based on straight-line distance. This study hypothesizes that the issue arises because some "edges," which indicate "roads" in the real world in the Greater London network, are not connected, preventing R5py from calculating the travel time. Figure 8 provides a detailed

illustration of one such OD pair. The centroid of this LSOA is located within Heathrow Airport, with the estimated travel time based on the straight-line distance being 30 minutes, while Google Maps shows that "Sorry, we could not calculate the walking directions from 'Bedfont Library, Staines Rd, Middlesex TW14 8DB' to 'Heathrow Terminals 2 & 3, Hounslow TW6 1SD'". Since this study involves 4,994 OD pairs and the problematic LSOAs account for only four, representing 0.08% of the total, the impact on the overall results is minimal. Therefore, these four LSOAs are dropped from further analysis.

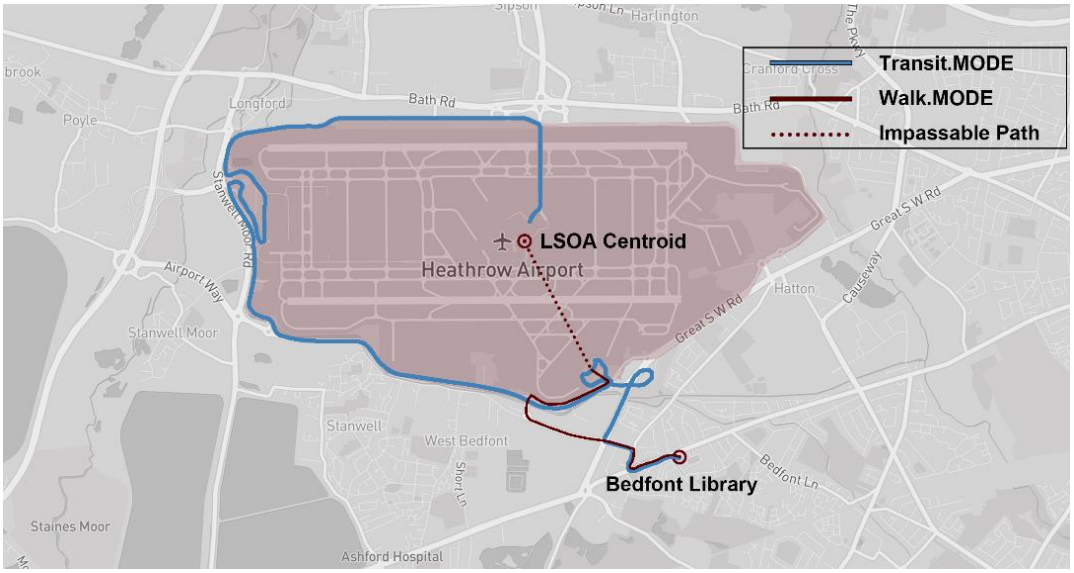


Figure 6: The Example of “null travel time” LSOA “Heathrow” (*Source:* Author’s own creation)

3.4 socio-economic indicators data analysis

Figure 7 shows the Pearson correlation matrix of the indicators used in this study, computed using Python's Pandas package. Tabel 3 shows the statistical distribution of employment rate and the proportion of residents with high qualifications (with a bachelor's degree or higher). The coefficients in the correlation matrix in Figure 7 indicate the extent of the correlation between each pair of variables. For example, the correlation coefficient between 'Walk and Transit travel time' and 'Car travel time' is 0.71, indicating a strong positive correlation. This coefficient means that car travel time usually increases when walking and public transport travel time increases.

Figure 7: Correlation Matrix (*Source:* Author's own creation)

Variable	Walk_and_Transit _travel_time	Car _travel_time	Bicycle _travel_time	Level_4: Qualification	Employment _rate
Unit of Measure	min (original data is int.)	min (original data is int.)	min (original data is int.)	%	%
Mean	18.28	4.26	8.62	46.79	48.18
Std. Dev.	9.77	2.51	5.22	13.85	7.43
Min	0	0	0	20.22	21.62
25%	12	3	5	35.34	43.46
50%	17	4	8	44.77	47.51
75%	23	5	11	57.21	52.08
Max	89	38	90	87.28	84.44
Obs.	4,991	4,994	4,994	4,994	4,994

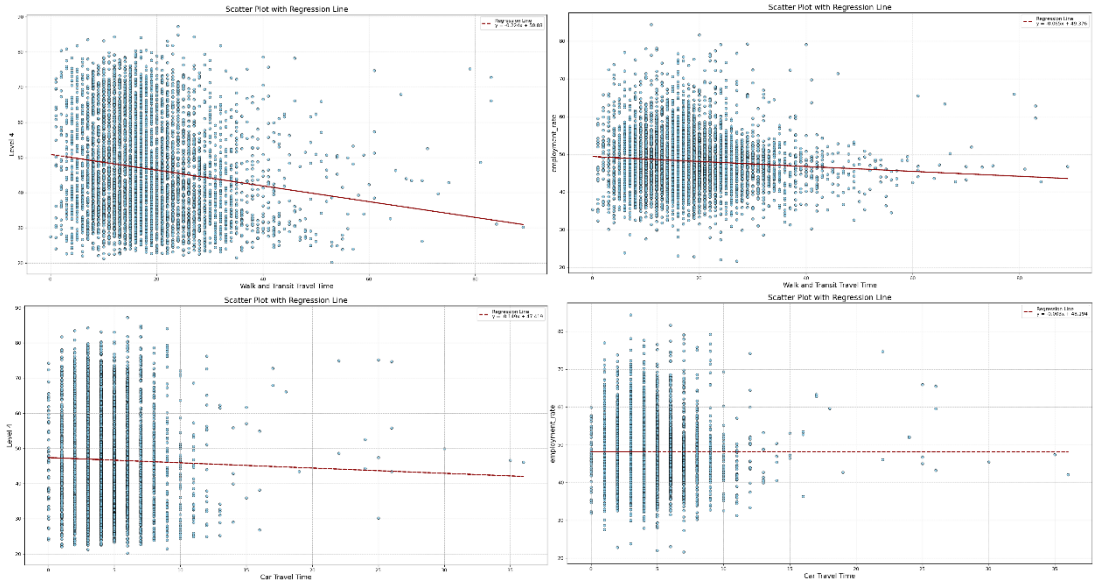
Table 3: Statistical description of variables (*Source:* Author's own creation)

4. Result of regression model

	(1)	**(2)**
Dependent variable:	Level_4 (Qualification)	Employment Rate
Walk and Transit	-0.2241***	-0.0655***
	(0.020)	(0.011)
	(3)	**(4)**
Dependent variable:	Level_4 (Qualification)	Employment Rate
Car	-0.1495*	-0.0034
	(0.080)	(0.043)
	(5)	**(6)**
Dependent variable:	Level_4 (Qualification)	Employment Rate
Bicycle	-0.2178***	-0.0663***
	(0.038)	(0.021)
Obs.	4,991	4,991

1. *** *p* < 0.01, ** *p* < 0.05, * *p* < 0.1

Table 4: Regression Results (*Source:* Author's own creation)



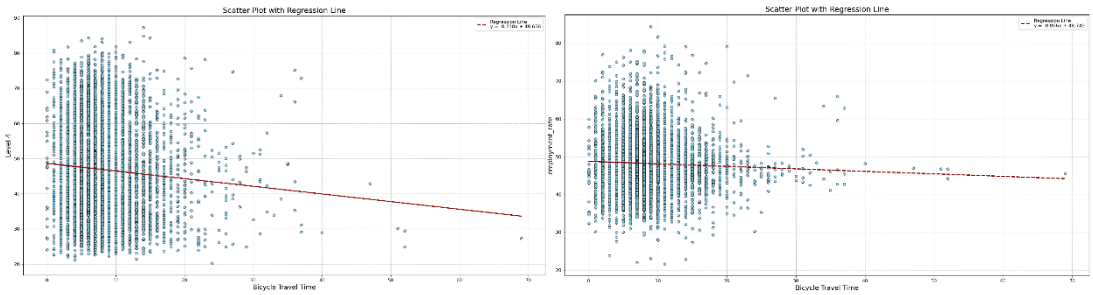


Figure 8: Linear Fit Plot of All Regression Models (*Source:* Author's own creation)

The independent variables are 'Walk and Transit travel time,' 'Car travel time,' and 'Bicycle travel time.' The dependent variables, analyzed separately in different models, are 'employment rate' and 'Level 4' (the percentage of people with bachelor's degrees and above). The regression model used in this study is the Ordinary Least Squares (OLS) regression model, which is used to estimate the parameters of the linear regression model. Its goal is to find a best-fit straight line that minimizes the sum of squares of the errors between the actual data points and that line. The results of the OLS model are shown in Table 4 and Figure 8.

Multiple linear regression is not used because the travel times for the different modes, which serve as independent variables, are highly correlated. The coefficient in regression analysis represents the magnitude of the impact of the independent variable on the dependent variable, indicating how much the dependent variable is expected to change with a one-unit increase in the independent variable. The p-value is used in statistical hypothesis testing to measure the probability of observing the results under the null hypothesis (usually the hypothesis of no effect). In regression analysis, the p-value is used to determine whether the independent variable has a significant effect on the dependent variable.

Based on the parameters of the six models shown in the figure, the following conclusions are drawn: walk and transit travel time, as well as bicycle travel time, have a significant negative impact on the dependent variables. The coefficients for these two variables are negative, and their p-values are both less than 0.05, indicating statistical significance. This result suggests that longer travel times for these modes are associated with lower educational attainment and employment rates. Specifically,

the analysis of the model coefficients reveals that, holding other conditions constant, for each additional minute in walk and transit travel time to the nearest public library, the employment rate decreases by 0.07%, and the proportion of residents with a bachelor's degree or higher decreases by 0.22%. A similar decrease is observed for bicycle travel time. In contrast, car travel time has no significant impact on the dependent variables. Although the regression coefficients for car travel time with employment rate and the proportion of residents with a bachelor's degree are negative, their p-values are greater than 0.05, indicating that the impact is not statistically significant at the standard confidence levels.

Overall, there is a significant negative correlation between walk and transit travel time, bicycle travel time, and both the employment rate and the proportion of residents with a bachelor's degree. In other words, shorter travel times by walking, public transit, and bicycle are associated with higher educational attainment and employment rates in an area. This suggests that better transportation conditions may contribute to enhancing a region's socio-economic level.

5. Discussion

5.1 Discussion on results

5.1.1 The interpretation of the results

Based on the conclusions drawn in the section 4, it is observed that there is a negative correlation between travel time from the LSOAs centroids to the nearest public library and both the employment rate and the proportion of residents with a bachelor's degree or higher in these areas within Greater London. This means that the shorter the travel time to reach a public library, the higher the employment rate and the proportion of highly educated individuals in the LSOAs.

It is important to note that the research results indicate a strong correlation between these factors without establishing a causal relationship. Carlozzi (2022) points out that low socio-economic status (SES) communities typically have lower library activity, which means that traditional physical library services do not reach low SES populations as effectively as higher SES populations. Public library surveys from that period note that “Public libraries serve the middle class, whether defined by occupation or economic status” (Berelson, 1949). In practice, traditional public library models may reinforce rather than alleviate social inequality, as they fail to effectively serve those most affected by these inequalities. Therefore, it is difficult to determine whether accessibility or socio-economic indicators cause the other; it is more likely that these factors interact, creating a positive feedback loop. This implies that people who are further away from public libraries are less likely to obtain higher qualifications and are less likely to obtain employment, indicating that they have a lower SES.

Based on the conclusions drawn, we can analyze some of the high correlations observed. From the perspective of the resources provided by public libraries, they offer community residents a wealth of educational resources, including books, journals, digital resources, educational courses, study spaces, and regular thematic lectures. Residents who live closer to a library have easier access to these educational resources and training opportunities, enabling them to improve themselves and stay updated with information and skills, which may lead to higher educational attainment

and employment rates. From the perspective of public libraries as venues for social activities, residents who frequently use libraries are more likely to participate in community activities and build social networks. These social networks can help residents access employment information and job opportunities, increasing employment rates. Moreover, the assembly of social capital may also encourage community residents to place a higher value on education, thus raising their level of educational attainment. Regarding the socio-economic characteristics of the areas where public libraries are located, public libraries are often situated in areas with better socio-economic conditions, which typically have a concentration of other public facilities such as schools and community centers. This planning makes these areas centers for education and career development, more likely to attract residents with higher education levels and employment rates. Therefore, areas close to libraries tend to have higher employment rates and a greater proportion of highly educated residents, reflecting the role of libraries as symbols of quality community resources, not just educational facilities. These analyses provide a level of data analysis to support the introduction of the importance of public libraries in section 1.1.

5.1.2 Comparison of regression models for different travel modes

The results in section 4 indicate that, unlike the other two modes of transportation, the regression coefficient for car travel time is not statistically significant, which may reflect fundamental differences between car travel and other modes of transportation.

As shown clearly in Table 3 of Section 3.4, more than 75% of LSOAs in Greater London have a travel time of no more than 5 minutes to the nearest public library. Therefore, travel time by car lacks distinction as an independent variable in the regression analysis. From Figure 4 in section 3.3, the comparison of travel times for different transportation modes across the Greater London area clearly shows that, compared to walking, public transit, and cycling, car travel offers greater flexibility, especially over longer distances. Therefore, in the context of this study, the accessibility of public libraries calculated using the nearest neighbor analysis may not be as relevant for car travel. Within the same time frame, individuals with access to a car may have greater

flexibility in choosing public libraries, which means that the accessibility of public libraries in car mode may have a minor direct impact on employment rates and educational attainment levels within specific LSOA areas, thereby explaining the lack of statistical significance in the coefficient.

Compared to car travel, modes of transport such as walking, public transit, and cycling are often associated with stronger community engagement. Residents living in walkable neighborhoods are more likely to know their neighbors, engage in politics, trust others, and actively participate in social activities (Leyden, 2003). As mentioned in Section 1.1, public libraries serve as significant venues for community involvement. Therefore, shorter walking times indicate easier access for community residents to the resources provided by public libraries, as discussed in Section 1.1. Additionally, Section 2.2 elaborates on how public libraries contribute to improving community employment rates and educational attainment. Thus, this physical proximity and reduced travel time to public libraries can directly promote higher employment rates and educational levels among community residents.

In summary, walking, public transit, and cycling travel times more accurately reflect the close connection between residents and their workplaces or educational resources, thereby making their impact more significant.

5.2 Discussion on methodology

5.2.1 The Limitation of LSOAs centroid selection

In section 2.1.1 of this study, the LSOAs centroids are chosen as origins. Further, in the research titled "Methods for Assessing the Potential Spatial Accessibility of Urban Healthcare Services," Adkins et al. (2014) provide a detailed discussion on spatial reference units and aggregation methods. When selecting centroids for regions, the most straightforward approach is to use the centroid of the area. Based on this approach, the study area can be subdivided into smaller spatial units to improve accessibility assessments. Further optimization involves adjusting the centroid's position within spatial units using land use maps, primarily because population distribution within an area is often uneven. In some cases, the population weight of

larger spatial units becomes negligible when subdivided. The most accurate method involves adjusting the block centroid based on the actual distribution of the residential population within the spatial unit (Apparicio et al., 2017).

Ideally, such step-by-step optimization can significantly reduce errors in accessibility calculations, especially in large suburban areas. However, this approach requires highly detailed spatial population distribution data and land use data at the LSOA level, as well as significant time and computational resources, which are unavailable for this study. Consequently, this study can only attempt to subdivide larger spatial units into smaller units as finely as possible and then select their centroids as a substitute. This limitation is acknowledged as a constraint of the current research, and future studies could focus on gradually optimizing this aspect to significantly reduce travel time calculation errors and obtain regression models with more substantial explanatory power.

5.2.2 The limitation of “datatime” selection in R5py

As mentioned in the paragraph introducing the R5py library in section 2.1.2, when calculating travel time as an accessibility indicator, the design of the R5py code requires the input of a specific "datetime," which includes the exact year, month, day, hour, and minute. This necessitates a very precise and specific time. According to research by other scholars, accessibility studies are not as straightforward as measuring distances on a map; they can vary significantly depending on the time of day, the day of the week, and the available modes of transportation (Allen, 2019). The initial research plan intends to categorize and discuss these factors in detail according to the time of day and the day of the week. However, the workload required for this task is unmanageable due to limitations in time and computational resources. This limitation is acknowledged as a constraint of the current research and represents an area that can be optimized in future studies.

Consequently, as noted in section 3.3, this study selects "2024, 7, 23, 14:00," meaning July 23, 2024, at 14:00. This time is chosen because it does not fall within peak hours, making commute times less likely to experience significant fluctuations due to traffic

congestion. Additionally, public transportation is typically operational at this time, and public libraries are generally open, rendering this time point general and representative.

5.3 Discussion on the future of public libraries

As noted in 5.1.1, public library accessibility is strongly correlated with important socio-economic indicators such as employment rates and the proportion of highly educated individuals. This research reinforces the understanding that public libraries play an irreplaceable and multifaceted role in public infrastructure. Beyond their traditional function as reading spaces, public libraries have evolved to provide public digital services and serve as community activity centers, becoming multifunctional hubs within communities.

According to the DCMS report (2017), the main funding source for public libraries in the UK is local government, whose budgets are largely composed of central government grants. Therefore, it is hoped that the UK government, as the largest funder of public libraries, recognizes the irreplaceable positive role that public libraries play for individuals, communities, and society as a whole and increases funding for public libraries, particularly in communities with low accessibility to these facilities or those with a higher concentration of disadvantaged groups, as mentioned in Section 1.1. If implementing these kinds of measures in the short term proves challenging, alternative methods are also hoped to be identified to ensure these communities benefit from public library services.

One proposed idea is to emulate the traditional "mobile library" model. Historically, when books are expensive and inaccessible to normal people, mobile libraries offer an affordable way to access books by charging fees or subscription costs (Pereyra et al., n.d.). The government could partner with private companies to establish "quasi-public library" activity centers, offering paid digital book services and internet access. This initiative allows local community residents to access resources that otherwise require special permissions while providing a venue for community activities. Ultimately, such centers can fulfill many of the ideal functions of a public library, thereby continuing to have an irreplaceable positive impact on individuals, communities, and society.

6. Conclusion

In section 2.1, the research question “How does the accessibility of public libraries within the Greater London area impact community socio-economic indicators?” is posed, which includes two research objectives. For Objective 1, a spatial analysis of public library accessibility within communities is conducted in the spatial analysis of section 3.3. According to Figure 4, the accessibility of London public libraries, measured by travel time, varies across different modes of transportation, with Car.MODE is the fastest, followed by Bicycle.MODE, and lastly, Walk and Transit.MODE. For Objective 2, a linear regression analysis is performed in section 4, concluding that the accessibility of public libraries significantly impacts socio-economic factors such as employment rates and the proportion of highly educated individuals within the community. The shorter the travel time from the LSOA centroids to public libraries by walking and public transit or by bicycle, the higher the proportion of highly educated individuals and the employment rate within the community.

However, the Discussion of section 5 also points out that future research can further optimize the model by refining the selection of origins (LSOA centroids) and conducting more detailed classifications in calculating accessibility times. These improvements aim to enhance the model’s explanatory power, thereby better illustrating the importance of public libraries.

From a broader perspective, on the one hand, as outlined in section 1.1, public libraries play a multifaceted role in communities. It is essential for the government, communities, and individuals to recognize this as soon as possible and fully utilize the resources public libraries offer in the digital age, according to their respective needs. On the other hand, although public libraries serve diverse functions and may alleviate many social issues on the surface, the underlying causes that fall outside the purview of public libraries also need to be addressed.

Code Availability:

https://github.com/EthanLi1922/Yicong_Li_CASA_Dissertation.git

Acknowledgements

In this research, I would like to thank my supervisors, Fulvio Lopane and Guido Pialli, who are very conscientious and responsible. They respond to my questions in emails promptly and give positive feedback in offline discussions. Their academic guidance in literature review, thesis writing, and advanced spatial analysis techniques have been beneficial to me. Special thanks go to my parents and my friends Robert Luo and Ghost Huang for their moral support.

I acknowledge the use of ChatGPT 4o (Open AI, https://chat.openai.com) to summarize my initial notes and to proofread my final draft and the work of coding.

References

Audunson, R., Aabø, S., Blomgren, R., Evjen, S., Jochumsen, H., Larsen, H., Rasmussen, C.H., Vårheim, A., Johnston, J. and Koizumi, M., 2019. Public libraries as an infrastructure for a sustainable public sphere: A comprehensive review of research. *Journal of Documentation*, *75*(4), pp.773-790.

Appleton, L. and Hall, H., 2022. The public library as public sphere: a longitudinal analysis. Journal of Documentation, 79(1), pp.112-126.

Campana, K., Mills, J.E. and Martin, M.H., 2019, March. The role of community data in helping public libraries reach and serve underserved communities. In *International Conference on Information* (pp. 247-253). Cham: Springer International Publishing.

Griffin, M., Harvey, K., Gillett, J. and Andrews, G., 2022. Writing as/about leisure: Connecting with oneself and others through creative practice. *Leisure Sciences*, *44*(7), pp.862-880.

Philbin, M.M., Parker, C.M., Flaherty, M.G. and Hirsch, J.S., 2019. Public libraries: A community-level resource to advance population health. *Journal of Community Health*, *44*, pp.192-199.

Summers, S., 2022. Participation, people and place: the role of public libraries in fostering connection and cultural capital in disadvantaged communities.

Anderson, G. and Whalley, J., 2015. Public library internet access in areas of deprivation: The case of Glasgow. *Telematics and Informatics*, *32*(3), pp.521-537.

Grimes, N.D. and Porter, W., 2024. Closing the digital divide through digital equity: the role of libraries and librarians. *Public Library Quarterly*, *43*(3), pp.307-338.

Lee, P.C., 2024. From community engagement to community resilience: the evolving role of public libraries. *Public Library Quarterly*, *43*(3), pp.339-366.

Schloffel-Armstrong, S., 2024. The public library and the futures of social infrastructure. *Dialogues in Human Geography*, *14*(2), pp.312-316.

Paton, C. and McMenemy, D., 2024. The influences of communitarian philosophy in public policy: mapping the discourse of Scottish public library strategy. *Journal of Documentation*, *80*(1), pp.73-100.

Adkins, D.C., Haggerty, K.C. and Haggerty, T.M., 2014. The influence of community demographics on new public library facilities. *Proceedings of the American Society for Information Science and Technology*, *51*(1), pp.1-8.

Allen, J., 2019. Mapping differences in access to public libraries by travel mode and time of day. *Library & Information Science Research*, *41*(1), pp.11-18.

Tenkanen, H. and Toivonen, T., 2020. Longitudinal spatial dataset on travel times and distances by different travel modes in Helsinki Region. *Scientific Data*, *7*(1), p.77.

Guo, Y., Chan, C.H. and Yip, P.S., 2017. Spatial variation in accessibility of libraries in Hong Kong. *Library & Information Science Research*, *39*(4), pp.319-329.

Neutens, T., 2015. Accessibility, equity and health care: review and research directions for transport geographers. *Journal of Transport Geography*, *43*, pp.14-27.

Giannotti, M., Barros, J., Tomasiello, D.B., Smith, D., Pizzol, B., Santos, B.M., Zhong, C., Shen, Y., Marques, E. and Batty, M., 2021. Inequalities in transit accessibility: Contributions from a comparative study between Global South and North metropolitan regions. *Cities*, *109*, p.103016.

Cheng, W., Wu, J., Moen, W. and Hong, L., 2021. Assessing the spatial accessibility and spatial equity of public libraries' physical locations. *Library & Information Science Research*, *43*(2), p.101089.

Downey, G., 2007. Human geography and information studies. *Annual Review of Information Science And Technology*, *41*(1), pp.683-727.

Loendorf, S.C., 2009. Catching Patrons with a Library 2.0 Net: Using Geographic Information Systems to Measure the True Market Area of the Wake County Public Library System.

Kinikin, J., 2004. Applying geographic information systems to the Weber County Library System. *Information Technology and Libraries*, *23*(3), p.102.

Park, S.J., 2012. Measuring public library accessibility: A case study using GIS. *Library & Information Science Research*, *34*(1), pp.13-21.

Jae Park, S., 2012. Measuring travel time and distance in library use. *Library hi tech*, *30*(1), pp.151-169.

Office for National Statistics (ONS), 2021. Census 2021 geographies: Lower Layer Super Output Areas (LSOAs). Available at: https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeographies/census2021geographies#lower-layer-super-output-areas-lsoas-(Accessed: 26 June 2024).

Apparicio, P., Gelb, J., Dubé, A.S., Kingham, S., Gauvin, L. and Robitaille, É., 2017. The approaches to measuring the potential spatial access to urban health services revisited: distance types and aggregation-error issues. *International Journal of Health Geographics*, *16*, pp.1-24.

O'Sullivan, D. and Unwin, D.J., 2010. New Approaches to Spatial Analysis.

Fink, C., Klumpenhouwer, W., Saraiva, M., Pereira, R. and Tenkanen, H., 2022. r5py: Rapid Realistic Routing with R5 in Python. *Computer software]. Zenodo. https://doi. org/10.5281/zenodo*, *7060438*.

Higgs, G., Langford, M. and Fry, R., 2013. Investigating variations in the provision of digital services in public libraries using network-based GIS models. *Library & Information Science Research*, *35*(1), pp.24-32.

Kiani, B., Mohammadi, A., Bergquist, R. and Bagheri, N., 2021. Different configurations of the two-step floating catchment area method for measuring the spatial accessibility to hospitals for people living with disability: a cross-sectional study. *Archives of Public Health*, *79*(1), p.85.

Kong, D., Zhang, B. and Zhang, J., 2022. Higher education and corporate innovation. *Journal of Corporate Finance*, *72*, p.102165.

Zhang, Y., Kumar, S., Huang, X. and Yuan, Y., 2023. Human capital quality and the regional economic growth: Evidence from China. *Journal of Asian Economics*, *86*, p.101593.

Burggraeve, K., de Walque, G. and Zimmer, H., 2015. The relationship between economic growth and employment. *Economic Review*, *1*, pp.32-52.

Bhatt, R., 2010. The impact of public library use on reading, television, and academic outcomes. *Journal of Urban Economics*, *68*(2), pp.148-166.

Cole, N. and Stenström, C., 2021. The value of California's public libraries. *Public Library Quarterly*, *40*(6), pp.481-503.

Strover, S., Whitacre, B., Rhinesmith, C. and Schrubbe, A., 2020. The digital inclusion role of rural libraries: social inequalities through space and place. *Media, Culture & Society*, *42*(2), pp.242-259.

Taylor, N.G., Jaeger, P.T., McDermott, A.J., Kodama, C.M. and Bertot, J.C., 2012. Public libraries in the new economy: Twenty-first-century skills, the internet, and community needs. *Public Library Quarterly*, *31*(3), pp.191-219.

Bertot, J.C. and Jaeger, P.T., 2012. Implementing and managing public library networks, connectivity, and partnerships to promote e-government access and education. In *Managing E-Government Projects: Concepts, Issues, and Best Practices* (pp. 183-199). IGI Global.

CityPopulation, n.d. World agglomerations. Available at: https://www.citypopulation.de/en/world/agglomerations/ [Accessed 29 April. 2024].

Moya-Gómez, B. and García-Palomares, J.C., 2017. The impacts of congestion on automobile accessibility. What happens in large European cities?. *Journal of Transport Geography*, *62*, pp.148-159.

Greater London Authority, 2023. *Cultural Infrastructure Map 2023*. Available at: https://data.london.gov.uk/dataset/cultural-infrastructure-map-2023 (Accessed: 18 May 2024).

Office for National Statistics (ONS), 2024. *Lower Layer Super Output Areas (December 2021) Boundaries UK BGC.* Available at: https://geoportal.statistics.gov.uk/datasets/b8263c2364e9452483a0e5783c6fdb53_0/explore?location=52.721275%2C-2.487028%2C6.62 (Accessed: 26 June 2024).

Department for Transport, 2024. *Bus Open Data Service*. Available at: https://www.bus-data.dft.gov.uk/ (Accessed: 4 July 2024).

R5py, 2024. *R5py: Rapid Realistic Routing with R5 in Python*. Available at: https://github.com/r5py/r5py (Accessed: 4 June 2024).

Office for National Statistics (ONS), 2024. *Home page*. Available at:

https://www.ons.gov.uk/ (Accessed: 5 May 2024).

Office for National Statistics (ONS), 2024. *Create a customized dataset*. Available at: https://www.ons.gov.uk/datasets/create/ (Accessed: 11 May 2024).

Carlozzi, M., 2022. If You Build It, They Might Not Come: The Effects of Socioeconomic Predictors on Library Activity and Funding. *Open Information Science*, *6*(1), pp.116-135.

Berelson, B.R., 1949. *The Library's Public: A Report of the (US) Public Library Enquiry*.

Pereyra, P.R., 2022. *The Circulating Library in Melbourne: its place in the book-trade and its influence on readers* (Doctoral dissertation, Monash University).

Fink, C., Willberg, E., Klein, R., Heikinheimo, V. and Toivonen, T., 2024. A travel time matrix data set for the Helsinki region 2023 that is sensitive to time, mode and interpersonal differences, and uses open data and novel open-source software. *Scientific data*, *11*(1), p.858.

McGrail, M.R., 2012. Spatial accessibility of primary health care utilising the two step floating catchment area method: an assessment of recent improvements. *International journal of health geographics*, *11*, pp.1-12.

Rooney-Browne, C., 2011. Methods for demonstrating the value of public libraries in the UK: a literature review. *Library and Information Research*, *35*(109), pp.3-39.

Leyden, K.M., 2003. Social capital and the built environment: the importance of walkable neighborhoods. *American journal of public health*, *93*(9), pp.1546-1551.

Takyi, S., Antwi, R.B., Ozguven, E.E., Alisan, O., Ghorbanzadeh, M., Mardis, M. and Jones, F., 2023. Measuring spatial accessibility of public libraries using floating catchment area methods: A comparative case study in Calhoun County, Florida. *Transportation research interdisciplinary perspectives*, *22*, p.100944.

Pei, X., Guo, P., Chen, Q., Li, J., Liu, Z., Sun, Y. and Zhang, X., 2022. An Improved multi-mode two-step floating catchment area method for measuring accessibility of urban park in Tianjin, China. *Sustainability*, *14*(18), p.11592.

Bozorgi, P., Eberth, J.M., Eidson, J.P. and Porter, D.E., 2021. Facility attractiveness and social vulnerability impacts on spatial accessibility to opioid treatment programs in

South Carolina. *International journal of environmental research and public health*, *18*(8), p.4246.

Stępniak, M. and Rosik, P., 2018. The role of transport and population components in change in accessibility: the influence of the distance decay parameter. *Networks and Spatial Economics*, *18*, pp.291-312.

Huhndorf, S. and Działek, J., 2017. Assessment of the walking accessibility of public libraries in Krakow using the E2SFCA approach. *Urban Development Issues*, *54*.

Department for Digital, Culture, Media and Sport (DCMS), 2017. Income generation for public libraries. [online] Available at: <https://dcmslibraries.blog.gov.uk/2017/06/15/income-generation-for-public-libraries/> [Accessed 23 August 2024].

Stokel-Walker C. Chatgpt Listed as Author on Research Papers. Nature. 2023;613(7945):620-U21. https://www.nature.com/articles/d41586-023-00107-z

Research log

Date (dd/mm/yyyy)	Discussion
13/03/2024	Discussion on the accessibility of the London Public Library project before choosing a dissertation topic and dissertation tutor.
09/04/2024	Having the kick-off meeting with Guido and Victoria.
23/04/2024	Have a team meeting (online) with Guido and Victoria to discuss the working plan and the first task to address.
16/05/2024	Discussion on literature review and what to do for the next step. Fulvio sends public library data in a Geojson file after the meeting by e-mail.
04/06/2024	Discussion on the methodology used and select the R5py library as the tool.
28/06/2024	Problems with R5py were resolved (import London GTFS for public transit analysis and deal with incompatibilities in Python libraries).
04/07/2024	(By e-mail) Discussion on choosing OD pairs in spatial analysis by using Nearest Neighbor Analysis.
26/07/2024	Updating the progress and solving the problem with R5py public transportation.
02/08/2024	Discussion on correlation analysis variables.
14/08/2024	Discussion on preliminary correlation results.
16/08/2024	Discussion on regression results and autocorrelation of travel times variables in OLS regression models.
23/08/2024	Discussion on the issue on the first version of the thesis draft.