East London Survey Data

Sari Aslam N[*1], Tzivanakis N[†1], Woodcraft S[‡1], and Kurbucz MT[§1]

[1]Institute for Global Prosperity (IGP), University College London, London, UK

CUPUM 2025

Summary

This study presents the East London Survey Data, collected from 4,093 households across 94 Output Areas (OAs) in five boroughs undergoing the East London Longitudinal projects. The data focuses on key aspects of prosperity, based on responses from 293 households representing diverse communities. It covers demographic attributes (age, ethnicity, gender), socio-economic factors (income, housing, employment), and community cohesion (trust in neighbours and local government). The ground-truth data for understanding long-term trends through future longitudinal studies aim to address inequalities and foster sustainable, resilient communities in East London.

KEYWORDS: Open Data Source, Household Surveys, East London Study, Inequalities

1 Introduction

The East London Longitudinal Study aims to evaluate the impact of regeneration projects on prosperity in East London. Promises of improved socio-economic outcomes often accompany regeneration efforts, yet there is limited evidence of their effectiveness, particularly for low-income and long-standing residents. This study seeks to fill this gap by collecting detailed, hyper-local survey data, referred to as the East London Survey Data (ELSD), to assess changes in prosperity across diverse neighbourhoods. The focus is on areas where regeneration is in different stages (complete or not), such as East Village in Newham and Hackney Wick, capturing the complex interactions between established and newer neighbourhoods. Moreover, these areas are characterised by East London's rich diversity in ethnicities, ages, housing types, tenures, and employment patterns, and they promise to provide further insight into urban planning.

This paper presents ELSD as an open data source (https://doi.org/10.5281/zenodo.14847390) for the first time to investigate robust, evidence-based insights to design and implement more like regeneration urban strategies that work for all, ensuring equitable and sustainable prosperity across communities. To eliminate sensitivity issues, the ELSD from 4,093 households is georeferenced to Output Areas (OAs) and Lower Layer Super Output Areas (LSOAs) for spatial precision. Second, a data dictionary is generated from 287 questions hierarchically: (i) possible attributes for urban studies, (ii) the code for questions to represent the households' responses. The data includes all details from households that either agreed to participate (attended) or did not participate (no household member was available, or no competent respondent was at home during the visit). The responses are kept in their original form (not encoded), allowing researchers to design their analysis.

2 Data and Method

This section addresses three key components: the study area, data collection methodology, and open data sources, particularly the East London Survey Data.

[*] n.aslam.11@ucl.ac.uk
[†] n.tzivanakis@ucl.ac.uk
[‡] saffron.woodcraft@ucl.ac.uk
[§1] m.kurbucz@ucl.ac.uk

2.1 East London as Study Area

The key locations of the study are the Royal Docks, Olympic Park, Tower Estate, and Heath in East London. These areas, situated across the boroughs of Barking and Dagenham, Hackney, Newham, Tower Hamlets, and Waltham Forest, provide a rich background and diverse context for understanding the impacts of urban regeneration. Figure 1 illustrates the geographical distribution of these areas within East London.

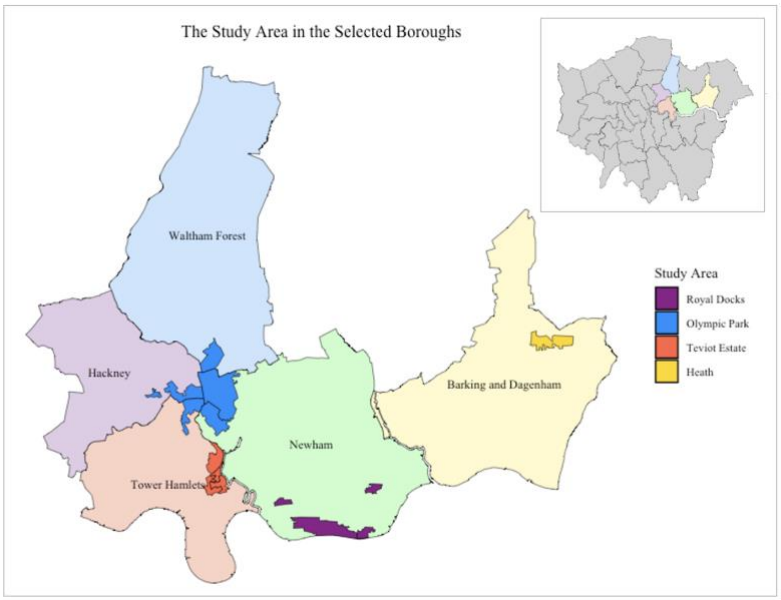

Figure 1: The study area for East London Survey Data

2.2 Method for data collection

The data collection methodology is divided into three sections: (1) Preparing the Questionnaire - The questionnaire was developed through an extensive literature review, expert consultations, and qualitative research to ensure it captured citizens lived experiences in defining prosperity. (2) Conceptual Framework - Based on the Citizen Prosperity Index (CPI), the framework identified key dimensions of prosperity. Pilot testing was conducted to refine question clarity and sequence, aligning them with respondents' realities. (3) Data collection - The final questionnaire was designed to minimise respondent bias by avoiding leading questions and providing balanced response options. Thus, ELSD was collected through face-to-face interviews with households in the study area, utilising the KOBO Toolbox to ensure consistency and efficiency. A stratified sampling strategy was employed to ensure representation across key population characteristics, such as age and gender, within the study area.

2.3 East London Survey Data (ELSD)

East London area is one of the deprived areas in the Great London Area (GLA) in terms of income, health, and other attributes (https://data.cdrc.ac.uk/dataset/index-multiple-deprivation-imd). The East London Survey Data (ELSD) contributes to the same areas, providing additional data sources as ground truth, and it has been made publicly available for the first time through this study. The ELSD supports quantitative and qualitative research from the residents and their definition of the good life. Some urban indicators/attributes are grouped from the dataset presented in Figure 2.

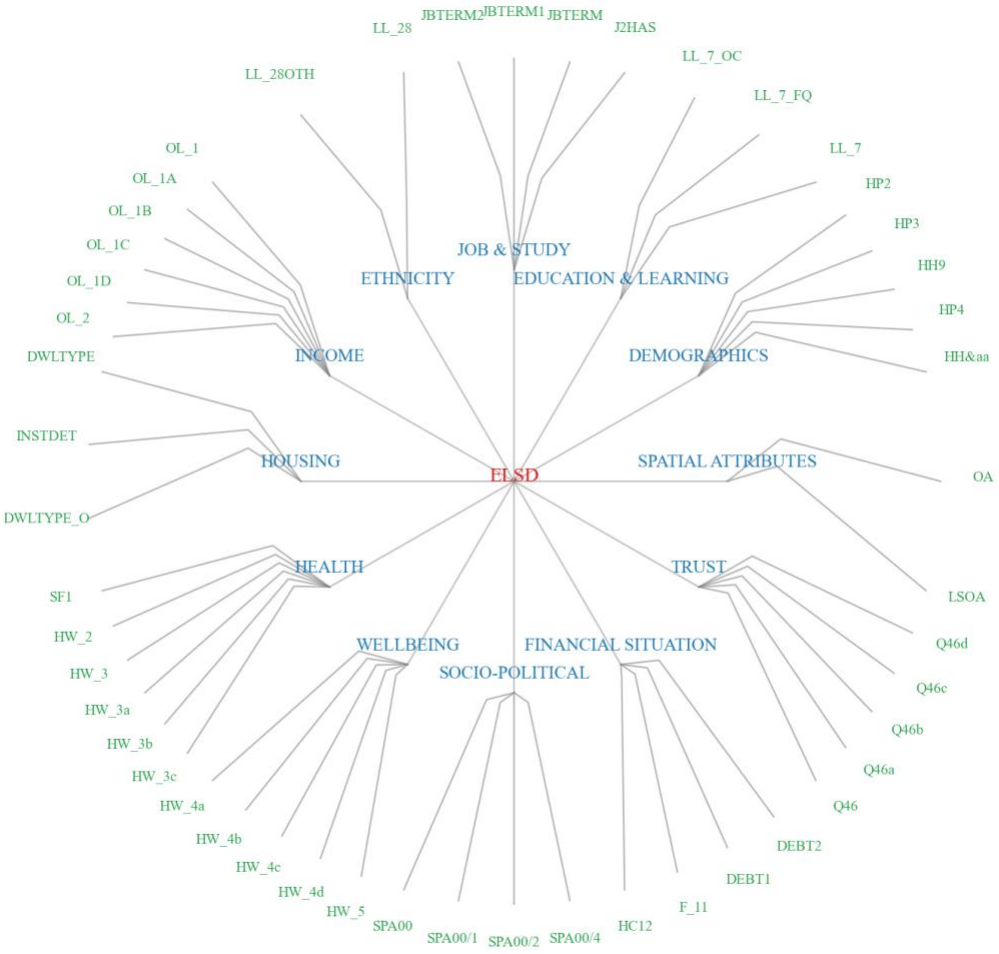

Figure 2 East London Survey Data (ELSD) with possible indicators and corresponding attributes. The outer layer represents the encoded attributes, e.g., DWLTYPE refers to 'type of housing'.

This figure illustrates the East London Survey Data (ELSD) hierarchy, demonstrating its potential attributes in a structured manner. The root node (level 0), 'ELSD,' branches out into possible key attributes (level 1) such as demographics, education & learning, and trust. The outer layer (level 2) illustrates the responses of the survey participants and can be found here (https://doi.org/10.5281/zenodo.14847390). Note that this figure is provided as an example; researchers can generate relevant indicators based on their research interests and the scope of their studies.

3 Results

3.1 Preliminary Data Analysis

HP2 – Gender, as an attribute under household demographics (Level 1), is selected to illustrate the population distribution by gender across selected boroughs, using a colour-coded scheme to represent different population counts. The three maps display the distribution for females, males, and non-binary (NB) individuals, respectively. The data suggests that female and male populations are more evenly distributed, with some areas having significantly higher counts, as indicated by the yellow and orange regions. In contrast, non-binary individuals are concentrated in fewer areas, primarily in dark blue and purple regions, indicating lower population densities in East London.

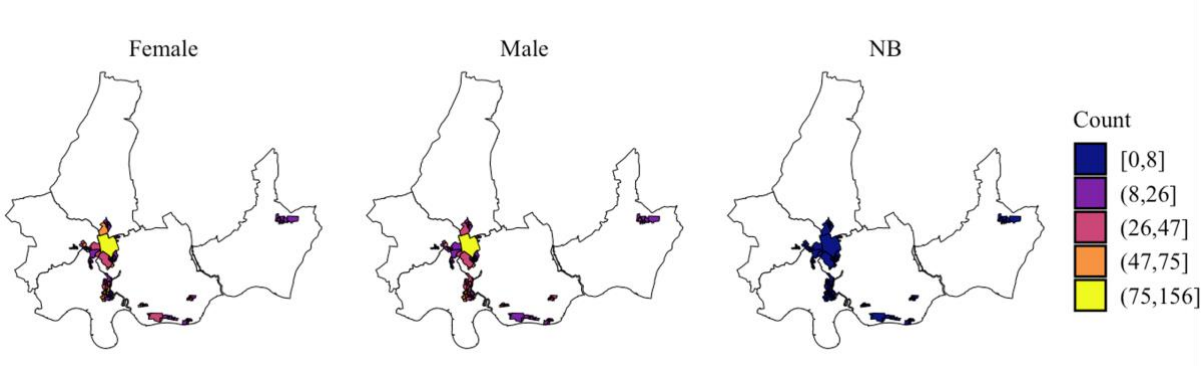

Figure 3: Spatial distribution of gender populations across selected boroughs using Jenks Natural Breaks classification

3.2 Application of East London Survey Data

A Citizen Prosperity Index (CPI) for the East London study is one of the main applications of this data, transferred using the methodology published by Melios et al. (2024). The combination of urban indicators is different in this application, reflecting the meaning of prosperity for the communities that live in East London.

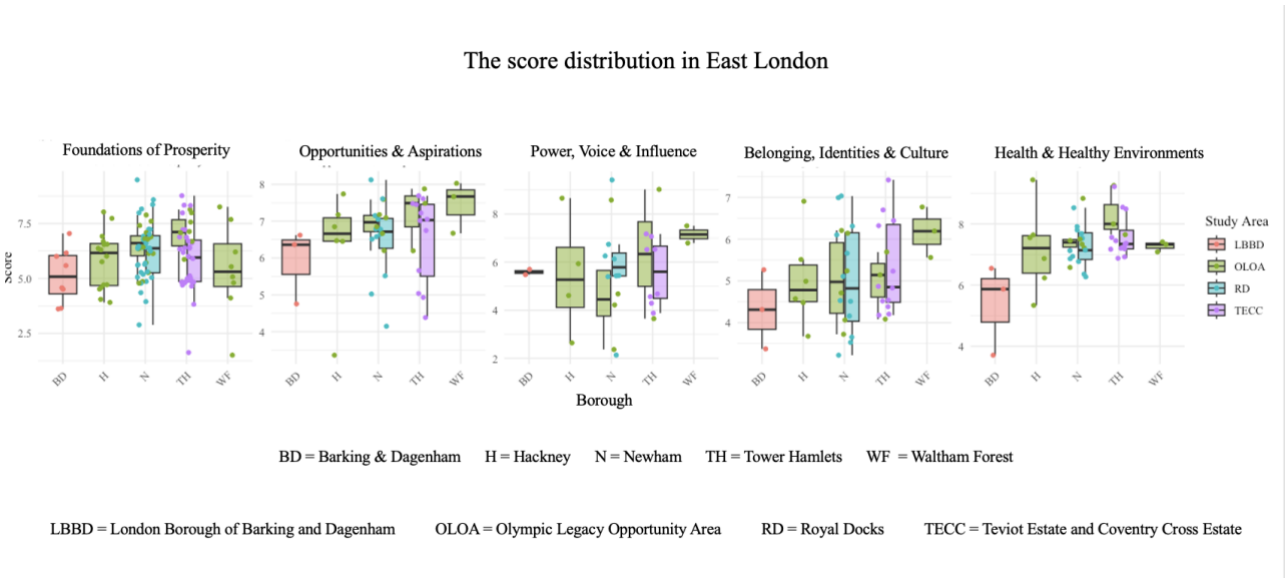

Figure 4 The distribution of scores from CPI in the study area.

Figure 4 presents the score distribution across various boroughs in East London for five CPI domains: Foundations of Prosperity, Opportunities & Aspirations, Power, Voice & Influence, Belonging, Identities & Culture, and Health & Healthy Environments. Each boxplot represents the variability and central tendency of scores for a specific borough, categorised by study areas: London Borough of Barking and Dagenham (LBBD), Olympic Legacy Opportunity Area (OLOA), Royal Docks (RD), and Teviot Estate and Coventry Cross Estate (TECC). The most significant finding from this figure is the difference in "Foundations of Prosperity" scores across East London boroughs. Tower Hamlets (TH) consistently shows the lowest scores in this category, while areas like Hackney (H) and parts of the Olympic Legacy Opportunity Area (OLOA) perform much better. This shows that Tower Hamlets faces serious disadvantages in basic prosperity conditions compared to nearby areas. While the other four prosperity categories show more similar performance across boroughs, this foundational gap stands out as the biggest inequality, indicating that improving basic prosperity conditions in Tower Hamlets should be a key priority for reducing differences between areas in East London.

3.3 Data Linkage with Secondary Data Sources

This data can be linked with additional secondary data sources to provide further insights into the communities in East London. The linkage can be generated using either spatial units such as OAs and LSOAs or the coordinates of locations obtained from various data sources. For instance, Point of Interest (POI) data (https://digimap.edina.ac.uk/) can be linked using the spatial attributes/coordinates to represent the distribution of urban functions such as shopping or eating locations to investigate spatial inequalities in the study area. Another possible example can be linking data with crime data (https://www.ukcrimestats.com/LSOA/) based on the spatial units to see the type of crimes in the area. Note that the dataset must be normalised before comparison using normalisation techniques. One common method, z-score normalisation, is provided as a reference below.

$$x_{z-score} = \frac{(x-\bar{x})}{\lambda_x} \quad (1)$$

x refers to a variable for the dataset, $\bar{x}$ and λ_x refer to the mean and standard deviation of the attribute, respectively.

4 Conclusion

The ELS dataset provides a comprehensive resource for analysing the East London area, with detailed information at the OA and LSOA levels. It offers potential for applications in urban planning, policymaking, and academic research by enabling localised assessments of demographic changes, housing trends, and the distribution of regeneration benefits. The data is generated from face-to-face household surveys and is presented in two parts: (i) a data dictionary with possible indicators (level 1) and the encoded attributes (level 2), and (ii) the response values from household surveys based on OAs and LSOAs. The potential preliminary analyses and applications, including data enrichment with secondary sources, are discussed. Future research will focus on enriching data with additional data sources to present further insight into long-term trends for the local communities in the study area.

References

Melios, G., Moore, H., Tzivanakis, N., Woodcraft, S., 2024. East London Citizen Prosperity Index Methodology. https://doi.org/10.2139/ssrn.4945628.

Acknowledgement

This research was jointly funded by London Prosperity Board members: Royal Docks, Lendlease, London Legacy Development Corporation, Hill Group, Poplar HARCA, the London Boroughs of Hackney, Waltham Forest, and Barking and Dagenham, and the Institute for Global Prosperity at UCL. This research was carried out in accordance with UCL's ethical guidelines and was approved by the relevant Ethical Committees (project ID 16659/002).

Biographies

Dr Nilufer Sari Aslam is a Research Fellow at UCL's Institute for Global Prosperity. Her research includes spatial-temporal data mining and machine learning (ML) applications in urban mobility and inequalities.

Dr Nikolaos Tzivanakis is a Senior Research Fellow at UCL's Institute for Global Prosperity. His research focuses on economic growth and development, political economy, and welfare, with policy design, inequalities, and prosperity.

Dr Saffron Woodcraft is a Principal Research Fellow and Director of Social Policy at the Institute for Global Prosperity. Her research explores lived experiences of community, shared prosperity, and urban

inequalities.

Dr Marcell Tamás Kurbucz is a Research Fellow at the Institute of Global Prosperity at UCL. He applies advanced quantitative and computational methods to investigate economic and social phenomena.