eMOTIONAL Cities

Mapping the cities through the senses of those who make them

DELIVERABLE 4.2

Mapping Urban Health

17/05/2022

UNIVERSITY OF CAMBRIDGE

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement nº945307. This document reflects only the author's view and the Commission is not responsible for any use that may be made of the information it contains.

Project Title	**eMotional Cities**
Deliverable	**Deliverable 4.2. – Mapping urban health across pilots and hotspots**
Work package	**WP4**
Task	**T4.1, T4.2, T4.3**
Number of pages	**69**
Dissemination level	**Public**
Leader	**University of Cambridge (CAMB)**
Contributors and peer-reviewers	**Instituto de Geografia e Ordenamento do Território da Universidade de Lisboa (IGOT); Faculdade de Medicina da Unversidade de Lisboa (FMUL)**
Date	**17/05/2022**
File name	**eMCities_WP4_D4.2**
Version	**V02**
Main author	**Prof. Elisabete A. Silva (CAMB), Dr. Haifeng Niu (CAMB), Ana Paula Seraphim (CAMB)**

Index

Index of Tables

Index of Figures

Executive Summary

This document is built on the previous Deliverable by extending the content and adding more in-depth analysis. The report first presents the portfolio of urban health variables and metrics with new inputs from reports and feedback. The following sections deliver the method and result of mapping urban health with traditional datasets (e.g., census and survey) and new datasets (e.g., points of interest data and Twitter data). Considering the data availability in the case studies, this report delivers the detailed result (with 350m hexagon grids) of urban health in London, covering health outcomes, physical environment, socioeconomic groups, and urban perceptions. The result presents the current landscape of urban health in London and provides an example for other case studies.

1. Introduction

1.1 Spatial data analysis for urban health

Studies about the impact of the urban fabric on the physical and emotional wellbeing of populations are raising an important need to include public health strategies in urban management and planning. The restorative environment generally adopted two schemes to assess the restorative qualities of specific places. One is to test changes in real indicators such as heart rate and blood pressure, but this method usually requires strict experimental conditions and professional equipment[1]. The other is to evaluate people's perceived restorative effects, which is mainly in the means of quantitative and qualitative data such as questionnaires, interviews, census, and social media data[2].

In these studies, spatial analysis is an essential part of a suite of **methods and metrics** to **map the physical and mental wellbeing of populations**[3]. It refers to "techniques that are used to analyse and acquire profound knowledge out of urban data"[4]. Conventional data in urban analysis contains census data, housing survey, land cover data and post office data which are sometimes expensive, uncertain, inaccurate, and time-consuming[5]. Nevertheless, the regular frequency of the collection and the methodological consistency are still one of the best attributes and can be the foundation of analytics for more refined data analysis for validation, calibration, but also for valuable time analysis and long-range trends[6]. Nowadays, numerous sensors, online posts, digital pictures, and other records in smart cities contribute to the explosion of big data with a vast potential to be used in health and wellbeing studies[7]. Big data and data analytics can also supply the conventional data to unfold the medium-term and long-term urban issues such as public participation, social segregation, individual perception of local amenities and uses[8]. Adding to this big data availability, new data analytics' methods

[1] Weber, A. M., & Trojan, J. (2018). The restorative value of the urban environment: A systematic review of the existing literature. Environmental Health Insights, 12, 1-13

[2] Nordh, H. (2012). Quantitative methods of measuring restorative components in urban public parks. Journal of Landscape Architecture, 7(1), 46-53

[3] Banerjee S. (2016). Spatial Data Analysis. Annual review of public health, 37, 47–60. https://doi.org/10.1146/annurev-publhealth-032315-021711

[4] Moustaka, V., Vakali, A., & Anthopoulos, L. G. (2018). A systematic review for smart city data analytics. *ACM* Computing Surveys (CSUR), 51(5), 1-41.

[5] Brito-Henriques, et al. 2018. 'Morfologia Da Cidade Perfurada: Padrões Espaciais de Ruínas e Terrenos Vacantes Em Cidades Portuguesas'. Finisterra-Revista Portuguesa de Geografia (108): 111–33

[6] Chen, Y, et al. (2020) 'Measuring policy debate in a regrowing city by sentiment analysis using online media data: a case study of Leipzig 2030', Regional Science Policy & Practice

[7] 2020 Hard and Soft Data Integration in Geocomputation: Mixed Methods for Data Collection and Processing in Urban Planning (chapter 3). In: Hand. on Planning Support Science. Ed. SG&JS, Edward Elgar Publishers.

[8] Hao, J., Zhu, J., & Zhong, R. (2015). The rise of big data on urban studies and planning practices in China: Review and open research issues. Journal of Urban Management, 4(2), 92-124.

and vast computation power, promote potential solutions more adjusted to local characteristics and variations.

1.2 WP4 Objectives and Relation to other Project WPs

The overall aim of the eMOTIONAL Cities project is to provide robust scientific evidence on how the natural and built urban environment shapes the neural system underlying human cognitive and emotional processing, with a perspective that also incorporates age, gender, and vulnerable groups' specificities. To achieve this aim, the project will adopt a systems approach, based on natural experiments and actual problems of selected case-study cities, designed to fully characterise, and understand the intensity, diversity, dynamism and complexity of urban health problems and inequalities. There are four different case studies across two continents (three in Europe and one in the US): London, Lisbon, Copenhagen, and Lansing.

In these four case studies, WP4 will apply spatial data analysis of health (e.g., the prevalence of diseases, medical prescriptions or admissions to hospital, healthcare system), demographic, socio-economical, the urban built environment and geotagged social-media data (Twitter). The outcome of this analysis will identify the driving factors affecting the perceptions of the urban environment; help subsequent stages as it will identify areas within the case studies; capture the baseline spectrum of cities' (geographical) features helpful to reduce urban health inequalities; and will be relevant when considering confounding in further spatial multivariate analysis.

The work of WP4 is divided in 5 tasks and 5 deliverables linked with those tasks (Figure 1):

- **Task 4.1 - Deploy a methodology for physical and mental urban health analysis** involves a review of spatial variables and metrics related to urban health (via government reports, data used by project partners, scientific literature reviews, data available in multiple platforms), that are used to develop our spatial analysis strategy for tasks 2 and 3.
    - D4.1. - A portfolio of metrics and methods for urban health
- **Task 4.2 - First snapshot mapping of the current urban health status across all cases studies** involves the mapping of the physical environment, socioeconomic environment, and health outcomes across all case studies and subsequent spatial analysis.
    - D4.2 - Quantitative/Qualitative Mapping of Urban health across the pilot studies and identified 'hot spots'
- **Task 4.3 - Characterization of cities based on emotions triggered by the built environment** involves the mapping of cognitive and emotional responses

triggered by the built environment in specific selected sites, building on the work of WP5.

  - D4.3 - Mapping of cities based on cognitive aspects and emotional responses triggered by the built environment

- **Task 4.4 - Customizing spatial health and well-being actions for cities and across specific populations** involves development of a framework based on the previous analysis, which will contribute to scenarios building and transferable methodologies.
  - D4.4 - A framework for the delivery of Urban Health and Well-being actions for cities and across specific populations
- **Task 4.5 - Set up a spatial analysis toolbox for health and well-being for cities** involves setting up a toolbox linking traditional and recent approaches of crowdsourcing data with qualitative/quantitative urban health calculation methods, customized to city and local analysis.
  - D4.5 - Spatial Analysis toolbox

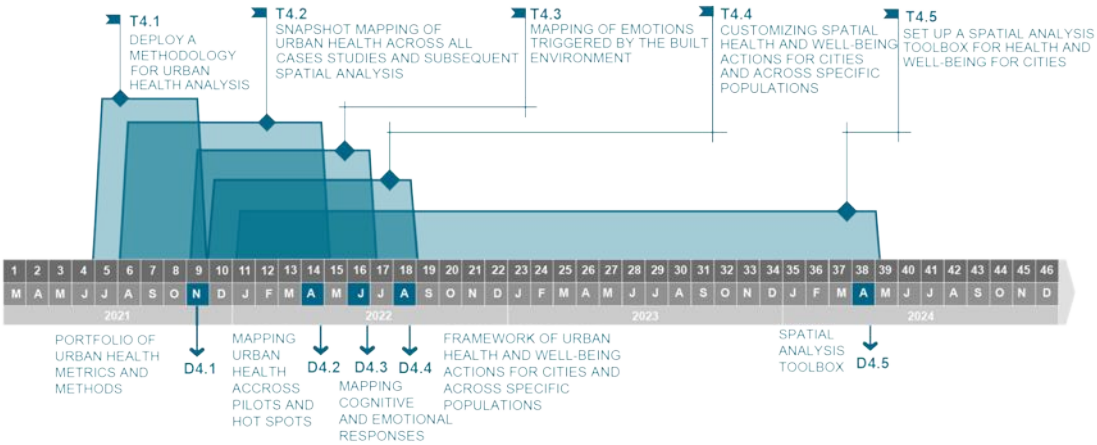


Figure 1 Timeline of deliverables and tasks for WP4

1.3 Report structure

We are already working on WP4 T4.1, T4.2 and T.4.3. Therefore, in this deliverable (D4.2.), instead of just presenting the results for T4.2 we will present the results of all our work so far in these 3 tasks (Please see Table 1).

The first section of the report **'Urban Health Variables and Metrics'** (Chapter 2) is linked with T4.1, and there we present the methodology and first results for the identification and prioritisation of variables and metrics based in a three-layer strategy composed of a bottom-up process, a science-to-policy process and a policy-to-science process. The second section '**Spatial Analysis of Urban Health**' explains the methodology for the spatial analysis for urban health mapping, which was built based on the results of T.4.1,

however considering the data (traditional data and social media data) availability in the case studies. The Chapter 4 '**Sentiment Analysis and Hotspot Identification**' then introduce the sentiment analysis for hotspot identification with Twitter geotagged data. The following section (Chapter 5-7) shows the results of mapping urban health, including results of **'Mapping Urban Health (Europe)'** and **'Mapping Urban Health (London)', which** are linked with T4.2. The end of this section also shows the results of mapping emotions with social media.

Table 1 Linkage between WP tasks and chapters in the report

WP Tasks	Chapter in the report
Task 4.1 - Deploy a methodology for physical and mental urban health analysis	Chapter 2: Urban Health Variables and Metrics Chapter 3: Spatial Analysis of Urban Health
Task 4.2 - First snapshot mapping of the current urban health status across all cases studies	Chapter 3: Spatial Analysis of Urban Health Chapter 5: Mapping Urban Health (Europe) Chapter 6: Mapping Urban Health (London)
Task 4.3 - Characterization of cities based on emotions triggered by the built environment	Chapter 4: Sentiment Analysis and Hotspot Identification Chapter 7: Mapping emotions with social media data

2. Urban Health Variables and Metrics

In this section, we are going to present the process through which we are identifying, structuring, and prioritising urban health variables. The main aim of this process is to recognise which are the urban environment and socioeconomic dimensions that most affect health outcomes. In the following section, the identified dimensions will be used to develop our spatial analysis strategy, considering the data availability.

In accordance with the project objectives (O4.1, O4.2, O4.3, and O4.4), we are considering 3 types of urban health variables: urban physical environment variables, socioeconomic variables, and health variables. The urban physical environment variables include both the natural and built environment physical characteristics, as for example shape, density, and configuration. The socioeconomic variables relate to the characteristics of the population, such as income, age, gender, and other vulnerable groups' specificities. And the health variables include both physical and mental health.

We are using a 3-layer strategy for the identification of variables and metrics, composed of a bottom-up, a science-to-policy, and a policy-to-science processes (Figure 2). The bottom-up process is based on the eMOTIONAL Cities members' input on what should be the priority variables and metrics for the WP4 to work on. The science-to-policy process is based on the WP2 Literature Review and in other scientific developments of WP4, which will feed subsequent work packages in the development of scenarios. And the policy-to-science process, is based on our review of policy documents to identify which are the urban health variables prioritised by policy. In the next sessions we are going to present the results of the bottom-up and policy-to-science process. The results for the science-to-policy can be found in the Deliverable 2.2 – Preliminary eMOTIONAL Cities conceptual framework of WP2.

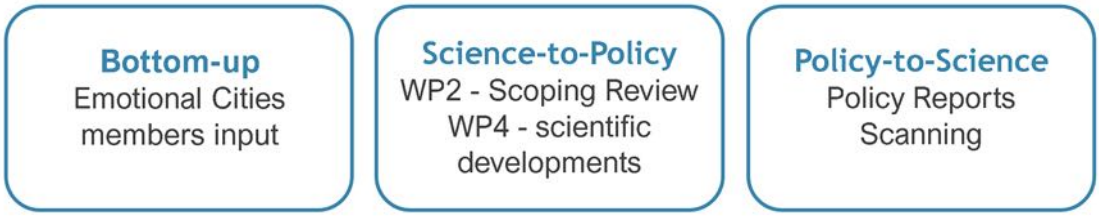


Figure 2 Three-layer strategy for the identification of urban health variables and metrics

2.1 Bottom-up Process – Emotional Cities Members Input

For the bottom-up process we forwarded a table that was filled by the eMOTIONAL Cities members with priority urban health variables and metrics, considering their experience. The table is composed of the following fields for each variable or metric:

- Name;
- Category, as a general theme in which it falls between urban physical environment, socioeconomic, and health;
- Type, between quantitative and qualitative;
- If it is spatial, in relation to being associated with a specific location in space;
- Level, which refer to the spatial extension coverage;
- A general description of what it represents and/or how it can be calculated;
- A description of how you should interpret the metric results;
- A description of the necessary data for calculating the metric;
- The expected computing time, considering the case studies; and
- A link to the methodological reference.

We had 36 inputs added to the table, mostly by IGOT and Michigan State University members. Some of the entries do not refer to an individual metric or variable, but to an index or academic paper that uses a variety of metrics and variables. Here we are going to present a summary of each input. The inputs are, grouped between the three categories of urban health variables (urban physical environment, socioeconomic, and health) and eleven sub-categories. The sub-categories green and blue spaces, heat stress and Urban Heat Island (UHI), air quality, noise pollution, and access to health services are related to the urban physical environment category. The sub-categories cardio-metabolic diseases, alcohol and drugs consumption, general comorbidities, heat stress, and mental health are related to the health category. Finally, the subcategory elderly groups are related to the socioeconomic category. The detailed information and references the members provided in the table can be found in Annex 1. Most of the table entries (25) are more closely related to the physical environment category. From that, 8 contributions are more related to *green and blue spaces* sub-category:

- **Green space metrics and health outcomes** - Normalised Difference Vegetation Index (NDVI) measures with different spatial resolutions and buffers from homes;
- **Green Spaces as an Indicator of urban health -** Availability and accessibility to green spaces, using land cover data from a random forest classifier;

- **Influence of urban green-blue spaces on human health** - green-blue spaces metrics, with a prevalence of land-use and vegetation metrics from moderate resolution satellite imagery, and buffer zones ranging between 30 and 5000 m;
- **Neighbourhood greenness** – Correlation between NDVI and psychologist rating of greenness;
- **Urban Green Infrastructure (GI) and Health Inequalities** - spatial analysis to identify priority areas for the implementation of GI. It combines land cover data from a biotope map, socioeconomic data based on income and education, and health data based on physical health (heart attacks rate, infections hospitalisation rate) and mental health (suicide attempts rate, open psychiatric treatments rate);
- **Urban Green Spaces (UGSs) and Human Health -** correlation between 6 characteristics of UGSs (availability, accessibility, shape complexity, mean distance of patches, patches cohesion, and NDVI) and 3 of morbidity (cardiovascular disease, mental disorders, and chronic respiratory disease);
- **Multifunctional Green Infrastructure (GI**) - spatial analysis to identify hot spots of GI function, based on 33 indicators in total: 5 indicators for the provision of natural resources (e.g. food production), 3 for water management (e.g. water provision and consumption), 12 for climate regulation (e.g. pollutants removal, carbon sequestration, cooling capacity, Leaf Area Index); 4 for health and wellbeing (population exposed to pollutants, high temperatures, noise, and flood risk), 5 for resilience (e.g. water retention capacity, surface runoff), 2 for tourism (e.g. accessibility to public parks), 1 for education (accessibility of parks from schools), and 1 for conservation benefits (cultural and natural heritage sites);
- **Health effects of the natural outdoor environment -** nature impact on health includes reduced levels of stress and aggression; cognitive restoration; reduced crime; increased physical activity and social interaction; improved sleep patterns, general health, mental health, and neural development; reduced cardiovascular, cancer and respiratory mortality and morbidity; reduced birth and pregnancy negative outcomes, and obesity. Also, the following socioeconomic groups are affected: age groups, socioeconomic status, and ethnic minorities.

Other 8 contributions are related to the *heat stress and UHI* sub-category*:*

- **Universal Thermal Climate Index (UTCI**) - thermo-physiological modelling of human comfort response to meteorological conditions data;
- **Heatwave Early Warning Systems (HEWS**) - warning system designed to reduce the avoidable health consequences of heatwaves through timely notification to vulnerable populations;

- **Urban structure and its implication on heat stress** - maximum likelihood classification method used to identify urban structures based on satellite image and classify their impact on heat stress (air temperature, mean radiant temperature, solar radiation intensity, relative humidity, and wind speed), based on ENVI-met[9] simulations;
- **Thermal infrared remote sensing of urban heat -** very high resolution (VHR), airborne thermal infrared (TIR) remotely sensed data to identify hotspots of Land Surface Temperature (LST);
- **Land-Use and Land-Cover Change, UHI, and Health -** correlates LST, NDVI extracted from satellite images, volatile organic compounds (VOC) and nitrogen oxides (NOx) emissions, and rates of cardiovascular and chronic lower respiratory diseases;
- **UHI implications on health -** UHI effect with excess heat mortality;
- **UHI and its impact on heat waves and human health -**correlation of UHI intensity measure difference between neighbouring neighbourhoods and excess heat mortality; and
- **Local Climate Zones (LCZ) -** dataset of publicly available LCZ information based on Copernicus Land Monitoring Service (CLMS).

Another 5 inputs are related to the *air quality* sub-category:

- **Air Quality Management Zones (AQMZs) -** spatial analysis to identify AQMZs based on urban form indicators, combining areas with decreased ventilation potential (building floor area ratio, building height, street canyon density, and tall vegetation density) and areas with increased potential to air pollution exposure (floor area ratio for residential and commercial, density of risk functions, cycling infrastructure density, parks density);
- **Public health impacts of urban air pollution -** short- and long-term effects of exposure to ozone, PM10, and PM2.5 on general mortality and hospitalisations;
- **Health-based assessment of particulate air pollution –** use the Cox's proportional hazard model to determine the effect of PM10 pollution exposure on morbidity and mortality rates, based on monitoring stations;

[9] ENVI-met is a three-dimensional microclimate model designed to simulate the surface-plant-air interactions in urban environments.

- **RAQ – Random Forest for Predicting Air Quality -** based on meteorological data, road information, real-time traffic status and point of interest (POI) distribution; and
- **Air Quality Indicators -** Spatial variability of air pollution based on satellite TROPOspheric Monitoring (NO2, CO, O3, and UVAI).

The 2 *access to health services* entries are:

- **Spatial Urban Health Equity Map -** Aggregating spatial accessibility criteria, density, and distribution of Health Care Medical Centres; and
- **Time-distance to health centres.**

Still in the physical environment category, we have 1 *noise pollution* entry, which is the **assessing urban soundscape,** based on publicly available environment noise maps.

We also have ten entries more closely related to the health category. From that, three entries were related to the *alcohol and drugs consumption* sub-category, which are **alcohol consumption rate, drugs consumption rate, and tobacco consumption rate**. Four entries were related to the *cardio-metabolic diseases* sub-category, which are **cardiovascular diseases prevalence, heart rate variability (HRV), obesity and overweight prevalence, and type II diabetes prevalence**. One entry is related to *comorbidities*, which means the existence of more than one chronic disease at the same time. One entry is related to *heat stress* (**Excess heat mortality**). And, finally, one is related to *mental health* (**Patient Health Questionnaire - PHQ**).

Moreover, we had 2 entries related to the socioeconomic category and the *elderly groups* sub-category, which are **ageing and people with more than 65 living alone.**

A summary of all the physical environment, health, and socioeconomic variables categories mentioned in this internal bottom-up process is show in figure 3. This figure clearly shows a focus in environmental exposures, that between the heat stress and UHI, air quality and noise pollution sub-categories, add up to 14 inputs. This is followed by a focus in the green and blue spaces sub-category with 8 inputs, and the in cardio-metabolic diseases sub-category, with 4 inputs.

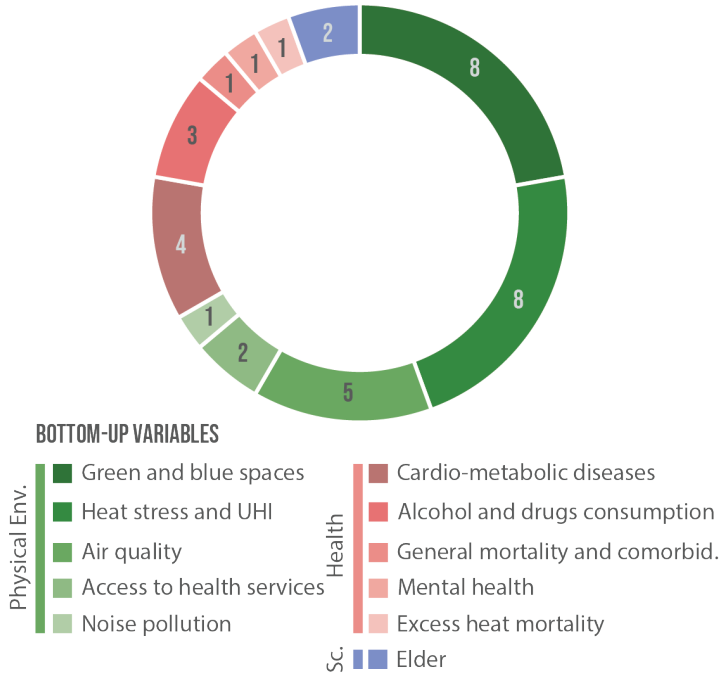


Figure 3: Bottom-up variables categories

All the heat stress and UHI, air quality and noise pollution sub-categories associated indicators are quantitative and linked with a specific location in space, although the spatial resolution and extent they cover vary greatly. The main types of data used were meteorological (e.g. air temperature, humidity), pollution (e.g., PM10, PM2.5), and urban structure (e.g., building floor area ratio and height), and urban land use (POI, traffic status) data. This data came from land monitoring stations; satellite imagery used to classify different urban structures; satellite sensors to identify LST and air pollution emissions; and land use data associated with areas of high air pollution exposure.

The green and blue spaces associated indicators are, also, all quantitative and spatial. The green areas are identified mainly by satellite imagery and sensors using random forest classification or NDVI. And a smaller amount was based on already available land-use and land-cover maps. This data is then used to measure a variety of indicators. The availability and accessibility to green spaces were the most common in the neighbourhood scale.

The indicators associated with the health variables include metrics related to cardiovascular disease, mental and cognitive disorders, chronic respiratory disease, diabetes, obesity, cancer, physical activity, social interaction, sleep patterns, alcohol and drugs consumption, general mortality and morbidity, and birth and pregnancy outcomes. The data sources are both from health statistical aggregated data and questionnaires. Finally, the indicators associated with socioeconomic variables, include age groups,

socioeconomic status, and ethnic minorities, from government statical aggregated data and questionnaires.

For all the variables the information about the spatial level – required data to calculate the indicator and computing time – are especially important for the development of our spatial analysis since this information can limit our capacity of calculating the indicators for the project case studies and orient the scale applicability.

2.2 Policy-to-science - Reports Scoping Review

The aim of the policy-to-science review process is to structure a framework of physical environment, health, and socioeconomic groups variables that have been highlighted by policy documents as having an important contribution for urban health. This will be done with an explicit spatial context in mind, making sure that as much as possible the resulting framework can be translated to spatial analysis and spatial planning strategies.

We are analysing policy documents in 3 levels: at the Global Level, at the Europe Level, and at the City Level. The Global level includes policy reports from intergovernmental agencies, such as the United Nations (UN) and the World Health Organization (WHO). The Europe Level includes reports from intergovernmental agencies in the European Continent region, such as the European Commission (EC). And the City level includes policy documents from both the national and local levels that reflect the region of the project case studies: London, Lisbon, Copenhagen, and Lansing. The Global, Europe, and City levels will also be used to develop our spatial analysis strategy.

In these levels we looked for the most updated versions of policy reports from 2010 forward which focused on the urban health field – addressing whether and how the physical and social urban environments impact population health. At the City level, since for some case studies there was not sufficient cross-sector documents that focused directly on urban health, we also considered sectorial reports that focused on urban and transport, health and social, and environmental policies, looking inside these documents for links between the urban environment and health.

To find the policy reports, we searched the webpages of relevant government agencies and also asked the eMOTIONAL Cities' members to identify relevant documents for the case studies they had a bigger familiarity with. From this initial result, we also expanded the results through citation searching.

Using this search strategy, we identified, at the Global level, 23 policy reports from the following agencies: WHO; UN; United Nations Human Settlements Programme (UN-

Habitat); and United Nations International Children's Emergency Fund (UNICEF). At the Europe Level we identified 9 reports from the EC and Eurocities.

Moreover, at the City Level, in the London region, we considered a total of 19 documents, from the National Health Service (NHS), Public Health England (PHE), Mayor of London, and Transport for London (TfL). In the Lisbon region we considered 15 documents, from the *Serviço Nacional de Saúde (SNS), Câmara Municipal de Lisboa, Agrupamento de Centros de Saúde de Lisboa, Área Metropolitana de Lisboa, Rede Social Lisboa, Ministério do Ambiente, Ordenamento do Território e Energia*. In the Copenhagen region we considered 15 documents, from Statistics Denmark and City of Copenhagen. And, in the Lansing region we are considering 2 reports from the Ingham County Health Department, and Healthy Capital Counties. In total we are considering 83 policy documents.

We have so far systematised the scanning results for the policy documents for London and Lisbon, Copenhagen, Worlds and Europe Levels which will be complemented by theLansing results in the next deliverable.

From the reports we extracted and charted variables similar to the ones of the bottom-up process. Four variables were considered: physical environment, healthy behaviour, and socioeconomic groups, and health outcomes.By extracting and charting these data and their relation to each other from the reviewed policy reports, we expect to construct an overview of how the urban health field is approached by policy in all levels of analysis and to highlight important conceptual links and gaps in this policy agenda.

These the physical environment, health behaviour, and socioeconomic variables are also seen by the policy reports as health determinant factors (HDF),, which are the conditions of our daily life and of our environment, not related to medical capacity and genetics, which influences our health. Among these variables, we identified nine common themes that were highlighted in the reports as priority areas for health promotion:: (i) healthy diet; (ii) drug, alcohol, and tobacco; (iii) social cohesion; (iv) physical activity; (v) personal safety; (vi) active travel; (vii) green infrastructure and open spaces; (viii) environmental exposures; and (ix) basic services. The entire review framework built in this report was structured around these HDF themes.

.

In total, we charted 40 physical environment variables related to the improvement or deterioration of the health determinant factors. These variables were categorized by type in 7 groups: (i) land Use, (ii) natural spaces, (iii) quality design and maintenance, (iv) mobility infrastructure, (v) urban structure, (vi) environmental exposures, and (vii) basic

infrastructure and services provision. Moreover, the healthy behaviour variables include individual habits that affects one`s risk of disease. We identified 10 healthy behaviours which are reported as being affected by any of the HDF themes.

The socioeconomic groups variables include communities with different demands, potentials, and limitations regarding health and wellbeing. Some socioeconomic groups are mentioned in the reports as of special interest for specific HDF themes. This special interest can be either due to the relative importance of an HDF theme to that socioeconomic group, or because that socioeconomic group is marginalized in relation to an HDF theme, or both. We charted 27 socioeconomic groups categorized by type in 5 groups: (i) age and life stages; (ii) gender; (iii) pre-existing health conditions and disabilities; (iv) minorities; and (v) special life circumstances.

Finally, the health outcomes variables include a variety of diseases and health conditions. We charted 38 health outcomes, mentioned as being affected by any of the HDF themes and categorized them in two groups: mental health outcomes and physical health outcomes. The mental health outcomes group includes one's emotional, cognitive, social, and behavioural well-being. And the physical health outcomes group includes the well-being of the body and the proper functioning of the organism of individuals. To standardize and group the numerous specific types of mental and physical health disorders mentioned in the reports, we decided to use the International Classification of Diseases and Related Health Problems (ICD-11), which is an international standard developed by WHO for recording and grouping conditions that influence health.

More detailed information on the analysis, each variable identified, discussion, and conclusion of this review can be found in the full article in Annex 2 – A Scoping Review of Current Urban Health Determinant Factors: A Policy-to-Science conceptual framework.

3. Spatial Analysis of Urban Health

3.1 Case studies

Figure 4 illustrates European countries' health and well-being status across cities, towns and suburbs, and rural areas. **In most countries, rural areas have a lower proportion of self-perceived health (above fair condition) than cities, towns and suburbs. Portugal sees a significant difference in terms of health outcome between rural areas and urbanised areas** while the United Kingdom almost have similar health outcome across areas with different degrees of urbanisation.

Figure 4 Proportion of self-perceived (above fair) health conditions by the degree of urbanisation in 2015 (Source: eurostat hlth_silc_18)

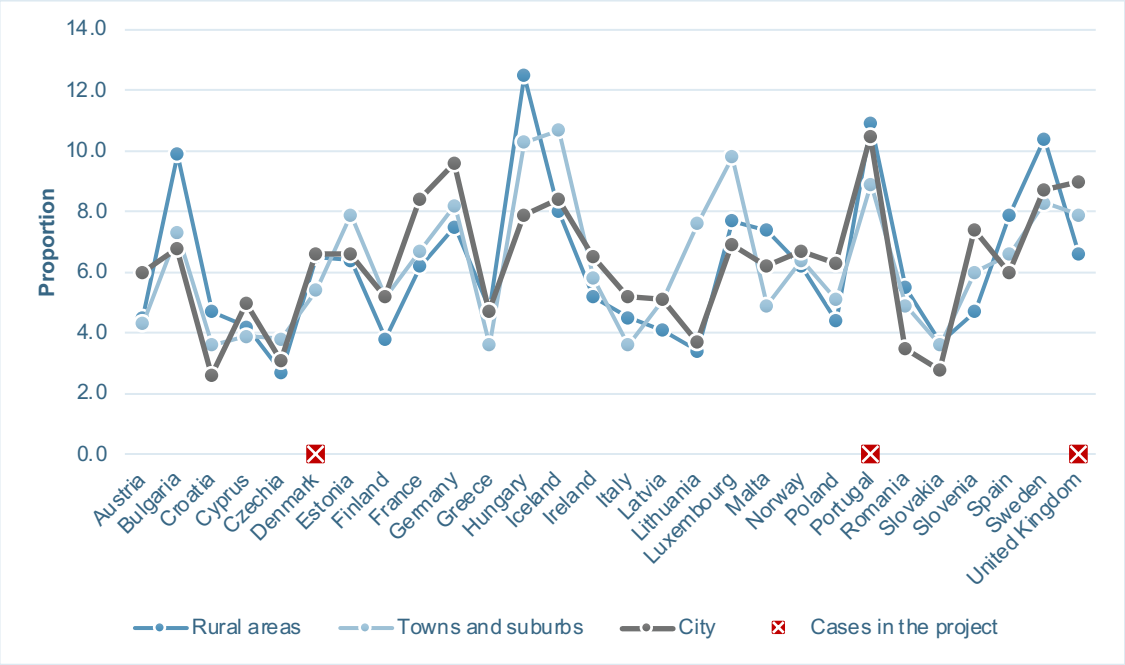


Figure 5 Proportion of depressive symptoms by the degree of urbanisation in 2014 (Source: eurostat hlth_ehis_mh1u)

By further looking into mental health by the degree of urbanisation (see Figure 5), **we found that urban dwellers in the United Kingdom have a higher proportion of depressive symptoms than people who live in towns and suburbs.** In the cases of Portugal, over 10% of people have developed depressive symptoms although no significant difference between city and rural areas was found.

The four pilot cities chosen in the project are London, Lisbon, Copenhagen, and Lansing. These four cities have different sizes in terms of urban population (see Figure 6 for cases in Europe). According to the Eurostat City Statistics in 2021, there are over 8.8 million people in Greater London, 1.8 million people in Lisbon metropolitan area and 0.5 million people in Copenhagen. The population of Lansing in 2019 is 117 thousand people[10]. Based on the feedback we have from the last report, we select central areas instead of metropolitan areas as study area boundaries. For example, we choose the boundary of inner London (13 London Boroughs) as the study area in London.

[10] Bureau, U., 2021. *American Community Survey 5-Year Data (2009-2019).* [online] Census.gov. Available at: <https://www.census.gov/data/developers/data-sets/acs-5year.html> [Accessed 29 October 2021].

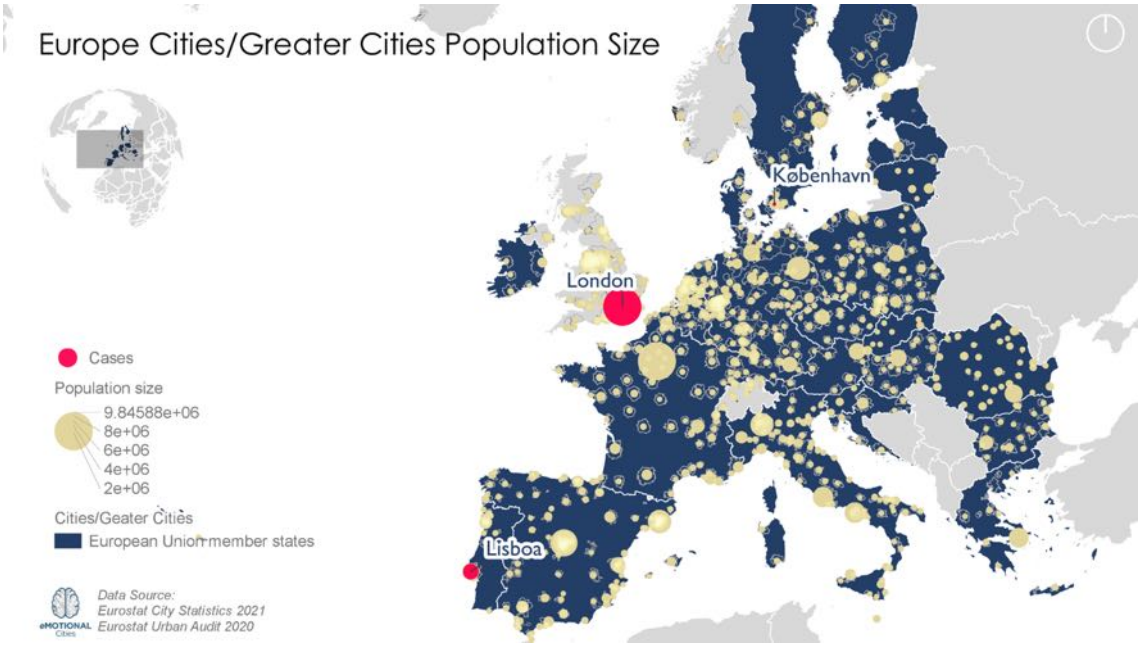


Figure 6 Selected cases in Europe (London, Lisbon, and Copenhagen)

3.2 The framework of spatial analysis considering the data availability

Spatial analysis of urban health heavily relies on the data availability in terms of coverage, granularity, consistency with all cases, etc. Thus, to build an operational spatial analysis framework, we first gather available data, including government data, public open data, and other data sets that can be accessed with licenses. This work is finished by searching databases and sending surveys to colleagues from all WPs[11].

We identify available data sets related to urban health in the region of Europe and the four case studies (currently focusing on London and Lisbon). We also identify the worldwide and Europe level data sets as a basic reference for the analysis. Urban health data sets identified in the project are mainly from the following sources and platforms.

- Worldwide datasets
    - *The Global Health Observatory*[12]
    - *Word Bank*[13]

[11] The survey of urban health datasets identification and data assemblage was sent on 15 April 2020 via email to all WPs.
[12] https://www.who.int/data/gho
[13] https://data.worldbank.org/

- Europe Level
    - *Eurostat*[14]
- City Level
    - *London Database* by Greater London Authority[15] and *Ordnance Survey*[16]
    - Lisbon Database (TBC)
    - Copenhagen Database (TBC)
    - Lansing Database (TBC)

The rationale of identifying urban health-related data

Aligning with the objectives of WP4 (Spatial analysis of urban health), the identified data sets include the stats of health outcomes and urban physical fabric, socioeconomic realities, and human emotional reactions. The review of WP2 has indicated the interactions between urban physical/socioeconomic environment and urban health. Thus, along with **health outcomes data**, **physical environment data** and **socioeconomic environment data** are also collected for spatial analysis.

Most of the above data sets can be found in our identified data sources and platforms (e.g., Eurostat and London Database), where census, surveys and questionnaires are publicly accessible. For instance, *Eurostat* provides most urban health-related data sets of European countries in the data collection of population and social conditions with sub-themes such as health, living conditions and welfare, demography, population and housing censuses, and sport. The statistics of health data covers health status, health determinants (e.g., BMI, physical activity, and alcohol consumption), health care, disability, mortality rates, health, and safety at work. Comparing the datasets at the European level, platforms at the city level have more available datasets regarding communities, transport, and physical environment. Taking London as an example, *London Database* provides such as active travel, land use data, access to open spaces that can be helpful in further understanding urban health. *Ordnance Survey* provides the footprints of buildings in London and points of interest data which can be used to calculate spatial metrics or variables describing the physical or socioeconomic characteristics of cities.

[14] https://ec.europa.eu/eurostat/data/database
[15] https://data.london.gov.uk/
[16] https://www.ordnancesurvey.co.uk/

On top of the above three categories of data sets, another necessary type of data is needed here is the **urban perception data,** which refers to people's experience and feeling in urban places[17]. The studies on urban perception start from the seminal work by Kevin Lynch, who explores the perceptual forms of the urban environment by crowdsourcing people's 'perception map' of cities[18]. In recent years, there has been an increasing number of studies focused on exploring urban perception from new forms of data, such as extracting emotions from social media data or measuring people's reactions to distinct types of street view images[19]. Moreover, studies also examined the relationship between the qualified urban perception with happiness[20], safety[21], and so on. One of the most common practices is to extract sentiment and emotions from social media data such as Twitter[22, 23, 24]. For instance, geotagged social media data link the individual opinions/reactions with the specific locations in cities. By conducting sentiment analysis of geotagged social media data, this helps to identify the prevalence of the positive/negative sentiment or specific emotion in local places and further identify the hotspot areas of different emotional reactions in cities. Thus, urban perception extracted from social media data will also be included in the spatial analysis. The detail of data collection, processing and analysis of geotagged social media data will be explained in Section 4.

In summary, both traditional data (e.g., census and survey data) and new urban big data (e.g., points of interest data and geotagged social media data) are included in the data identification of urban health-related data (Figure 7). As urban health is a complex and interdisciplinary topic, the required datasets for spatial analysis should include the features of the physical and socioeconomic environment and the interactions between

---

[17] Salesses, P., Schechtner, K., & Hidalgo, C. A. (2013). The collaborative image of the city: mapping the inequality of urban perception. *PloS one*, *8*(7), e68400.
[18] Lynch, K. (1964). *The image of the city*. MIT press.
[19] Dubey, A., Naik, N., Parikh, D., Raskar, R., & Hidalgo, C. A. (2016, October). Deep learning the city: Quantifying urban perception at a global scale. In *European conference on computer vision* (pp. 196-212). Springer, Cham.
[20] Quercia, D., Schifanella, R., & Aiello, L. M. (2014, September). The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city. In *Proceedings of the 25th ACM conference on Hypertext and social media* (pp. 116-125).
[21] Naik, N., Philipoom, J., Raskar, R., & Hidalgo, C. (2014). Streetscore-predicting the perceived safety of one million streetscapes. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 779-785).
[22] Plunz, R. A., Zhou, Y., Vintimilla, M. I. C., Mckeown, K., Yu, T., Uguccioni, L., & Sutto, M. P. (2019). Twitter sentiment in New York City parks as measure of well-being. *Landscape and urban planning*, *189*, 235-246.
[23] Gibbons, J., Malouf, R., Spitzberg, B., Martinez, L., Appleyard, B., Thompson, C., Nara, A., & Tsou, M.-H. (2019). Twitter-based measures of neighborhood sentiment as predictors of residential population health. *PLOS ONE*, 14(7), e0219550. https://doi.org/10.1371/journal.pone.0219550

[24] Jaidka, K., Giorgi, S., Schwartz, H. A., Kern, M. L., Ungar, L. H., & Eichstaedt, J. C. (2020). Estimating geographic subjective well-being from Twitter: A comparison of dictionary and data-driven language methods. *Proceedings of the National Academy of Sciences*, 117(19), 10165–10171. https://doi.org/10.1073/pnas.1906364117

residents and the environment they live (i.e., urban perception). Thus, there are four themes of data identified for the spatial analysis of urban health:

- Urban Health Data
- Physical Environment Data
- Socioeconomic Environment Data
- Urban Perception Data

Figure 7 Framework of spatial analysis of urban health

3.3 Selecting variables and metrics for spatial analysis

The rationale of selecting variables and metrics

After identifying the main themes of urban health-related data, we further select the variables/indicators/metrics for spatial analysis from the portfolio summarised in Section 2 with taking account of the data availability. As mentioned in the section, the strategy of the identification of variables and metrics is composed of bottom-up, a science-to-policy, and policy-to-science processes. Among the measures, those contributed from the eMOTIONAL Cities' members and previously examined in the scientific studies can be easily included in the spatial analysis as datasets are specifically identified and the methods/metrics are clearly explained. On top of this, city data platforms (e.g., *London Database*) also directly provides calculated variables or index with method explanation, which are also considered for spatial analysis. For some variables (e.g., availability of food growing spaces, quality design, and inclusive design) highlighted in policy reports, although their policy implication is obvious, they would not be included here due to the lack of appropriate data and the missing measurement.

List of selected variables and metrics

Figure 8 lists the selected variables and metrics that are included in the spatial analysis based on the availability of data sets[25]. The list follows a Theme – Sub-themes – Variables/Metrics structure. For instance, in the theme of Health Outcomes Data, there are two sub-themes (Physical and mental health). The metrics such as diabetes rates, cardiovascular disease rates, sports participation, childhood obesity and mortality rate are used to measure the physical health of a region or a neighbourhood. To measure the mental health in cities, we choose metrics such as suicide rate, wellbeing/happiness score, depression prevalence.

[25] The list has been shared and discussed among Principal Investigators (PIs) from the eMOTIONAL Cities.

Health Outcomes Data

Physical Health
Life Expectancy at birth, Mortality rate, Childhood Obesity, **Diabetes rates**, **Cardiovascular diseases**

Mental Health
Suicide rate, **Well-being/Happiness score**, Depression

Cognitive Health
Dementia rates

Health Behaviour
Sports Participation

Socioeconomic Env. Data

Deprived Community
Socioeconomic Status Index, Health insurance, Education level (alternative to SSI), Income level (alternative to SSI)

Age
Elderly people ratio

Gender
Gender ratio

Disabilities
Physical disability rate

Perception Data

Sentiment
Sentiment polarity

Emotion
Emotion

Physical Env. Data

Open Spaces
Greenery (Index/NDVI), Accessibility to green/blue infrastructure

Environmental Exposures
Exposure to pollutants (noise, air, and light), Heat island index, Flood vulnerability index

Active Travel
Walkability, Rail/road network, Bikeability, Car ridership, Connectivity, Visible Green Index

Crime Prevention
Safety index (Crime), Domestic violence and crime

Land Use
Land use mix, **Urban function mix**, **Building Density**, Building Age, Building Height

Healthy Food
Density of fast-food outlet, Community gardens allotment and greenspace productions

Services and Work Access
Accessibility to Health and community infrastructure

Urban Design
Street D/H Ratio, Street Façade (Continuity/Glazed openings/Aesthetics)

Others
Population density

Creator: eMotional Cities WP4

Figure 8 Measures lists for spatial analysis with urban health-related data. (Variables and metrics in bold are chosen for spatial mapping at the current stage.)

3.4 Analysis unit

To easily compare areas and combine different variables in the following spatial analysis, single uniform geography is adopted for the spatial analysis. In the case of London, we use a hexagonal grid, developed by Transport for London (TfL) as the analysis unit. The

distance from centre to centre in two near hexagons is 350 meters (Figure 9). The following are the advantages of using this hexagon grid:

- unified units with equal areas
- reduce sampling bias due to edge effects of the grid shape[26]
- less distortion when processing data on a large scale
- the circularity of a hexagon grid allows it to represent curves in the patterns of your data more naturally than square grids.
- several data sources provided by the public sector in London are based on these hexagon grids (e.g., transport data and green infrastructure)

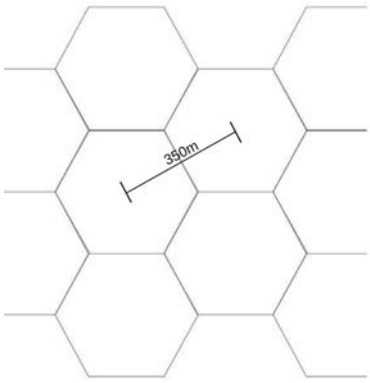


Figure 9 Hexagon grids (350 meters) as analysis unit for London

[26] Birch, Colin P.D., Oom, Sander P., and Beecham, Jonathan A. Rectangular and hexagonal grids used for observation, experiment, and simulation in ecology. *Ecological Modelling*, Vol. 206, No. 3–4. (August 2007), pp. 347–359.

4. Sentiment Analysis and Hotspot Identification

Ubiquitous geotagged social media have been increasingly critical to many practical and scientific studies in the urban context as it provides 'whereabouts' of people in cities. One important characteristic is that user-generated content of geotagged social media data contains individual expressions, opinions, and discussions, providing opportunities for detecting citizens' sentiment and its spatial distribution in cities. We collected social media data from Twitter to sense the sentiment in four project cases (London, Lisbon, Copenhagen, and Lansing). We choose Twitter as the data source because Twitter is one of the most popular social media platforms and also the main source that has been used in previous studies[27, 28]. In 2020, there are approximately 500 million tweets sent by millions of users every day, with an average rate of 0.83-3% of tweets being geotagged per day[29]. In the context of this project, geotagged social media data are used to detect urban perceptions of local urban places from the individuals.

4.1 Twitter API and settings

Two types of Twitter Application Programming Interfaces (APIs) are used for data collection. Previous studies mainly used the Twitter API - Standard v1.1 to collect the real-time streaming tweets. Although this type of API has no rate limit and upper limit on data response, it only returns 1% sampled tweets from the whole database[30]. Recently, Twitter released the Academic Research product track[31] that allows researchers to access the full archive of the database without exceeding the upper limit and rate limit (see Table 2). Two methods of data collection will be examined during the pilot study.

Table 2 Twitter APIs used for data collection

API	Sample ratio	Running	Upper limit	Rate limit	Monthly count

[27] Martí, P., Serrano-Estrada, L., & Nolasco-Cirugeda, A. (2019). Social Media data: Challenges, opportunities and limitations in urban studies. *Computers, Environment and Urban Systems*, 74, 161–174. https://doi.org/10.1016/j.compenvurbsys.2018.11.001

[28] Niu, H., & Silva, E. A. (2020). Crowdsourced data mining for urban activity: Review of data sources, applications, and methods. *Journal of Urban Planning and Development*, 146(2), 04020007. https://doi.org/10.1061/(ASCE)UP.1943-5444.0000566

[29] Internet Live Statistics. (2020). *Twitter usage statistics*. http://www.internetlivestats.com/twitter-statistics/

[30] https://developer.twitter.com/en/docs/twitter-api/v1/tweets/sample-realtime/overview

[31] https://developer.twitter.com/en/products/twitter-api/academic-research

Method 1	**Twitter API - Standard v1.1 (Streaming)**	Around 1%	Constantly running on PC/server	No (not mentioned in docs)	No limit for streaming	1.3 million without setting keywords (Greater London)
Method 2	**Twitter API v2 - Academic Research track (Full-archive search)**	Near 100%	Run monthly for several days	10 million per month	500 tweets per request; 300 requests per 15 mins;	Depends on case

4.2 Pilot result

We use both the abovementioned methods to conduct pilot studies of data collection and preliminary sentiment analysis in London, Lisbon, Copenhagen, and Lansing on 07-21-2021 (Wed) and 07-25-2021 (Sun). A summary is listed below.

- The scripts of data collection with both Twitter Standard API (Method 1) and Academic Research track API (Method 2) have been successfully tested for all four cases. Academic Research track API shows advantages in collecting data with few noises and setting specific parameters such as removing retweets and refining tweets by languages.
- The pilot results show that the rate limit for Academic API will not be a challenge during the data collection. Taking London as an example, the ratio of daily tweets count to API monthly cap is around 0.2%. The numbers are even lower in the other three cities.
- There are three types of geotags for tweets: point coordinates, admin coordinates and city coordinates. Only geotagged tweets with point coordinates give the accurate location of tweets that can be used in the hotspot analysis, which provide support in selecting city blocks or streets as experiment sites for WP5, neuroscience experiments.
- According to the word cloud extracted from tweets during the pilot, we can identify some words that are relevant to emotion. For text mining in the future, we need to improve the result by including stop words for non-English tweets.

4.3 Data collection and processing

Based on the pilot studies and discussions[32] with project PIs, we use the Academic Research track API to collect the data for sentiment analysis. The specifications of the data query are:

- Only collect tweets with geographic information tagged
- Set bounding box or point with radius to limit the extent of data collection
- Only keep original tweets by removing retweets
- For each case, only collect tweets in top two most used languages

Simultaneously, we also use the Twitter API - Standard v1.1 (Streaming) to collect data as the safety net.

Figure 10 Workflows of sentiment analysis

To explore the emotions of different demographic groups from tweets, we infer sociodemographic characteristics of social media users. In inferring demographics, we utilised user metadata (i.e., profile image, username, screen name and biography) for training a deep learning model that predicts demographic characteristics, including **age** and **gender** (Figure 10). The demographic inference was based on an open-source model – M3Inference – developed[33]. The M3Inference model is multimodal since it can integrate both text and image models. The output of gender inference is binary classification, and the output of age inference has four levels: $\leqslant$ 18, (18, 30), [30, 40) and [40, 99). It is worth noting that the model has been evaluated with the gender and age distribution of the European population dataset provided by the European Statistical Office (Lansing might be considered separately using US census).

4.4 Sentiment analysis (methodology)

As a popular technique from the domain of natural language processing, sentiment analysis explores insights from social media data. By implementing sentiment analysis,

[32] Meetings of sentiment analysis: 5th July (Methodology); 13th July (Methodology); 20 July (Data collection scripts).
[33] Wang, Z., Hale, S., Adelani, D. I., Grabowicz, P., Hartman, T., Flöck, F., & Jurgens, D. (2019, May). Demographic inference and representative population estimates from multilingual social media data. In *The world wide web conference* (pp. 2056-2067).

we reveal the positive, negative, or neutral tones and specific emotions of textual information in individual social media posts (i.e., tweets). With the attached geo-location of tweets, we can plot the individual emotions on a map and identify their local environment in the city. The methodology of sentiment analysis was discussed with project's PIs with the following agreements:

- Age and gender will be considered in sentiment analysis
- Sensitive keywords will be removed under the guidance of project PIs
- NRC Sentiment and Emotion Lexicon[34] will be used for emotion identification
- Hotspot analysis will be applied to identify the specific places for the following WP.

More detailed information on methodology, tools and analysis of sentiment analysis can be found in the full article in **Annex 3** – Urban Emotions Maps: Portraying Spatial and Demographic Patterns of Public Emotions with Large-scale Social Media Data [Unpublished manuscript].

[34] The NRC Emotion Lexicon is a list of English words and their associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive).

5. Mapping Urban Health (Europe)

To have a basic understanding of health outcomes in European countries, we plot the following maps (Figure 11 and Figure 12) by using life expectancy (to indicate physical health) and depressive symptoms (to indicate mental health). **Countries from Central Europe have lower life expectancy than those in Southern, Northern and Western Europe (over 80 years old).** When it comes to depressive symptoms[35], there is noticeable spatial variation among countries.

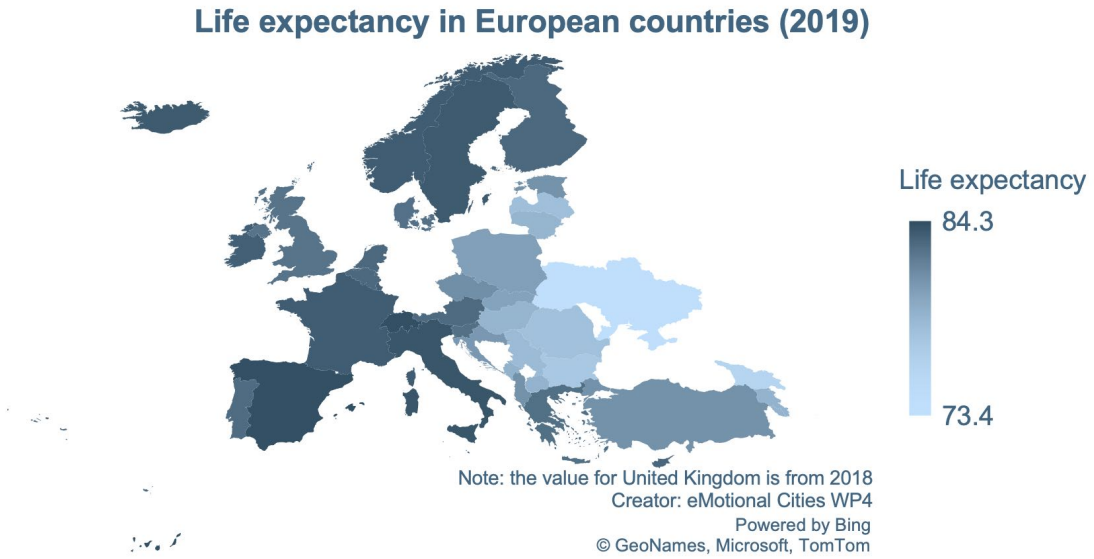


Figure 11 Life expectancy in European countries in 2019 (source: eurostat demo_mlexpec)

[35] Including 'major depressive symptoms' and 'other depressive symptoms'. Major depressive symptoms: if item MH1A or MH1B and five or more items MH1A to MH1H score at least 'more than half the days'. Other depressive symptoms: if item MH1A or MH1B and two, three or four items of MH1A to MH1H score at least 'more than half the days'

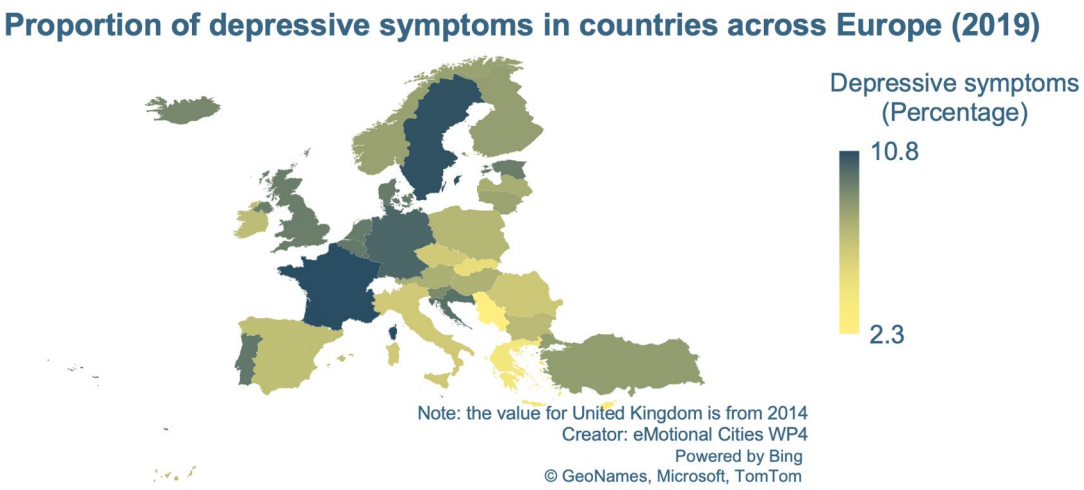
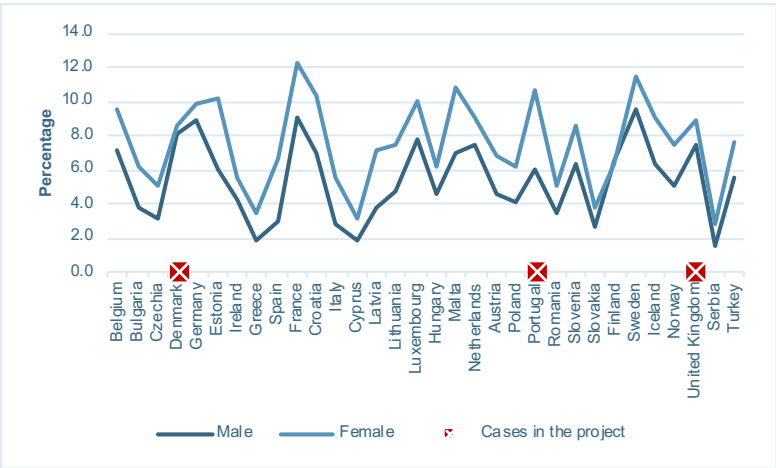


Figure 12 Proportion of depressive symptoms in countries across Europe in 2019 (source: eurostat hlth_ehis_mh1e)

According to the mental health data provided by Eurostat in 2019, countries such as France, Sweden, Germany and Portugal and Denmark have higher proportions of depressive symptoms. Around 8.5% of people in Portugal and around 8.3% of people in Denmark have depressive symptoms. As age and gender plays a very important role in health, we further look into the mental health condition in different European countries by social (i.e., age and sex) groups (See Figure 13 and Figure 14). **Across Europe, an overall difference appears between males and females in terms of having depressive symptoms.** Portugal shows the largest gender gap in depressive symptoms, where the number for females is 4.6% higher than that of males. In Denmark and the United Kingdom, the gender gap can also be identified, but they are not as large as in Portugal.



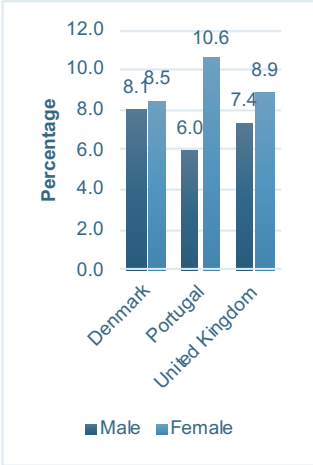


Figure 13 Proportion of depressive symptoms by sex in European countries (source: eurostat hlth_ehis_mh1i)

Regarding the age gap, there is no clear patterns in depressive symptoms across Europe. **In Denmark, people below 45 years old are likely to report having depressive symptoms.** The prevalence of depressive symptoms declines as people get older after 45. **When looking at the impact of aging, Portugal shows a clear trend that senior groups have higher proportion of people reporting depressive symptoms.** People over 64 have the second highest prevalence, around 14%, across all European countries. Differently, **the rates of depressive symptoms for Britain in their 50s and 40s are the highest, around 11.4% and 9.9%, respectively.**

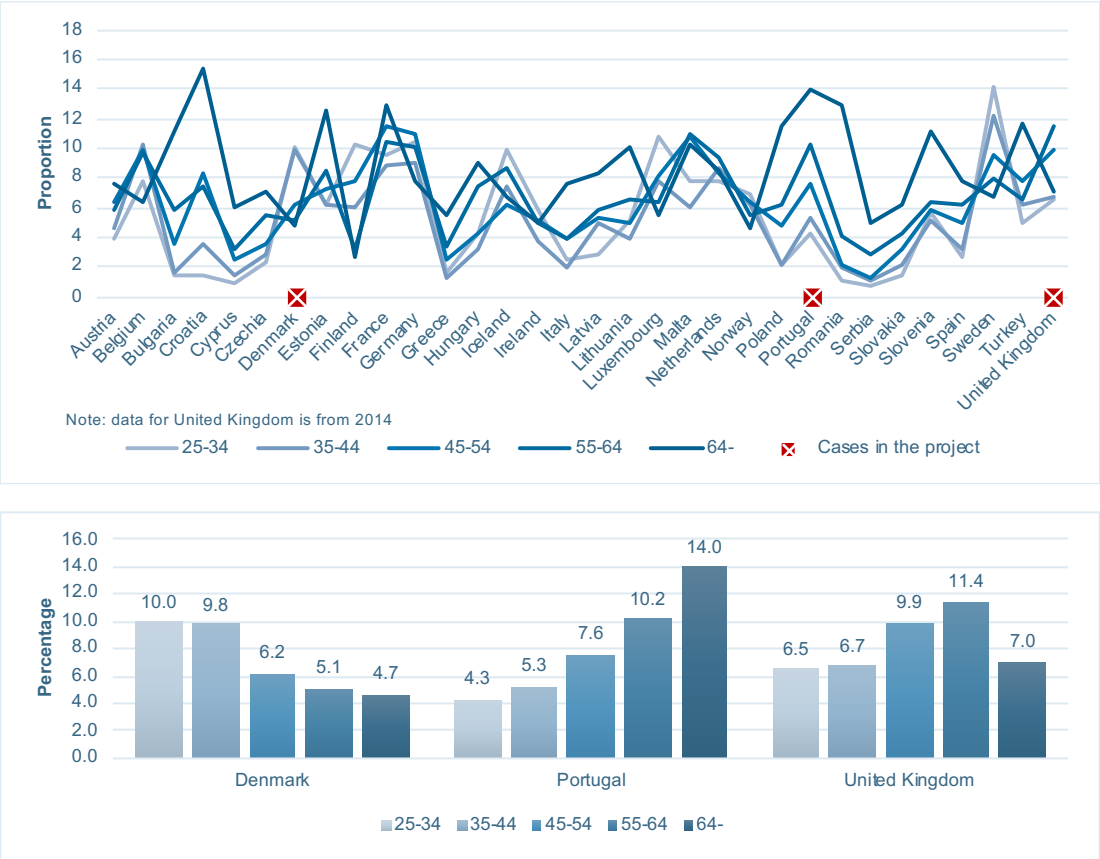


Figure 14 Proportion of depressive symptoms by age in European countries 2019 (source: eurostat hlth_ehis_mh1i]

6. Mapping Urban Health (London)

6.1 Maps of health outcomes

Based on the Public Health England data, we mapped the spatial distribution of diabetes prevalence, the mortality rate from cardiovascular diseases and physical activity levels in London boroughs. The examples indicate that health conditions are rarely the same across London areas - some have much higher health conditions than others.

- Western parts of Inner London - e.g., Camden, Kensington and Chelsea, City of Westminster and the City of London - have lower diabetes rates (Figure 15). Communities in the eastern part, especially in Newham, Tower Hamlets and Greenwich tend to have higher rates of diabetes.
- A hotspot of cardiovascular diseases related death is located at the East London, e.g., Barking and Dagenham, Newham, Greenwich along the Thames River (Figure 16). The mortality rate from cardiovascular diseases is particularly high in the borough of Barking and Dagenham.
- In terms of mental health in London, Figure 17, Figure 18 and Figure 19 show the maps of personal well-being with self-evaluated scores of happy feelings, life satisfaction and anxious feelings. Boroughs such as Kensington and Chelsea, Camden, Islington, and Hackney have lower scores of happy feelings and life satisfaction, compared with others within Inner London. In terms of anxiety, the City of Westminster, and Kensington and Chelsea have the highest scores of anxiety, along with South wark and Lewisham from the south side of the River Thames.
- It is surprising to see that some Inner London boroughs with relatively underperformed physical health outcomes, such as Tower Hamlets and Newham, have high scores of happy feelings and low scores of anxious feelings.

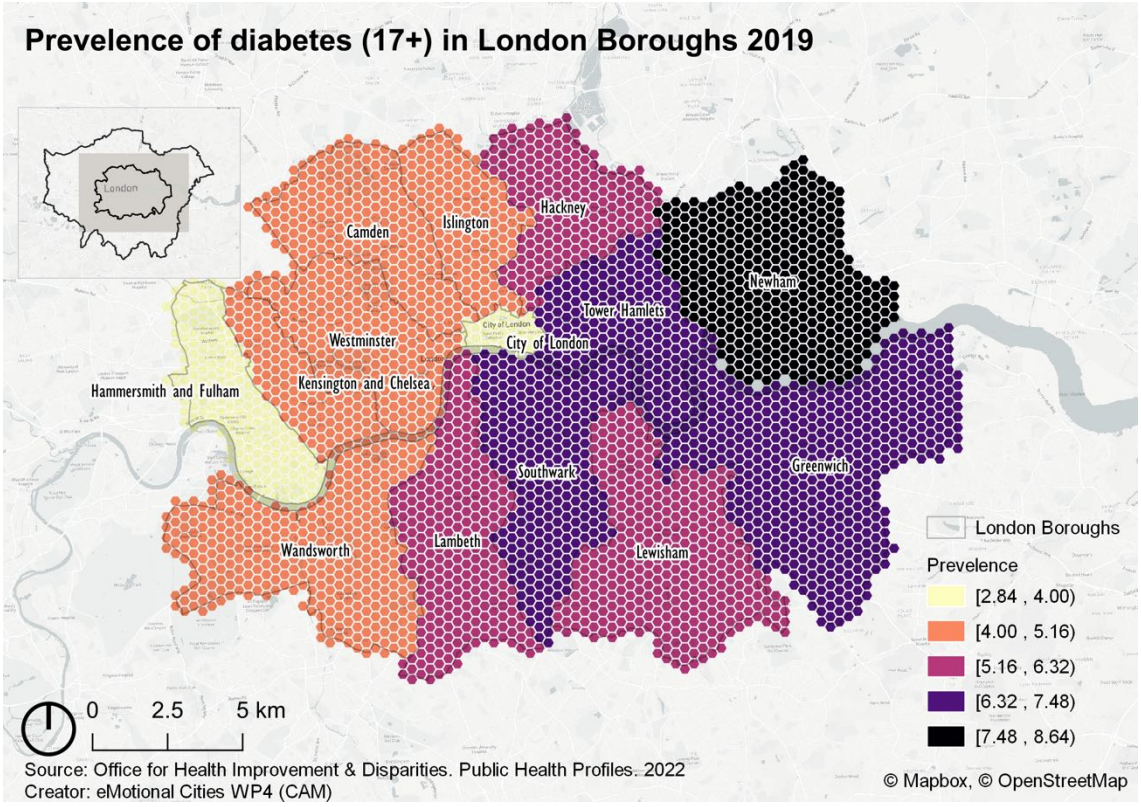


Figure 15 Diabetes prevalence in London Boroughs

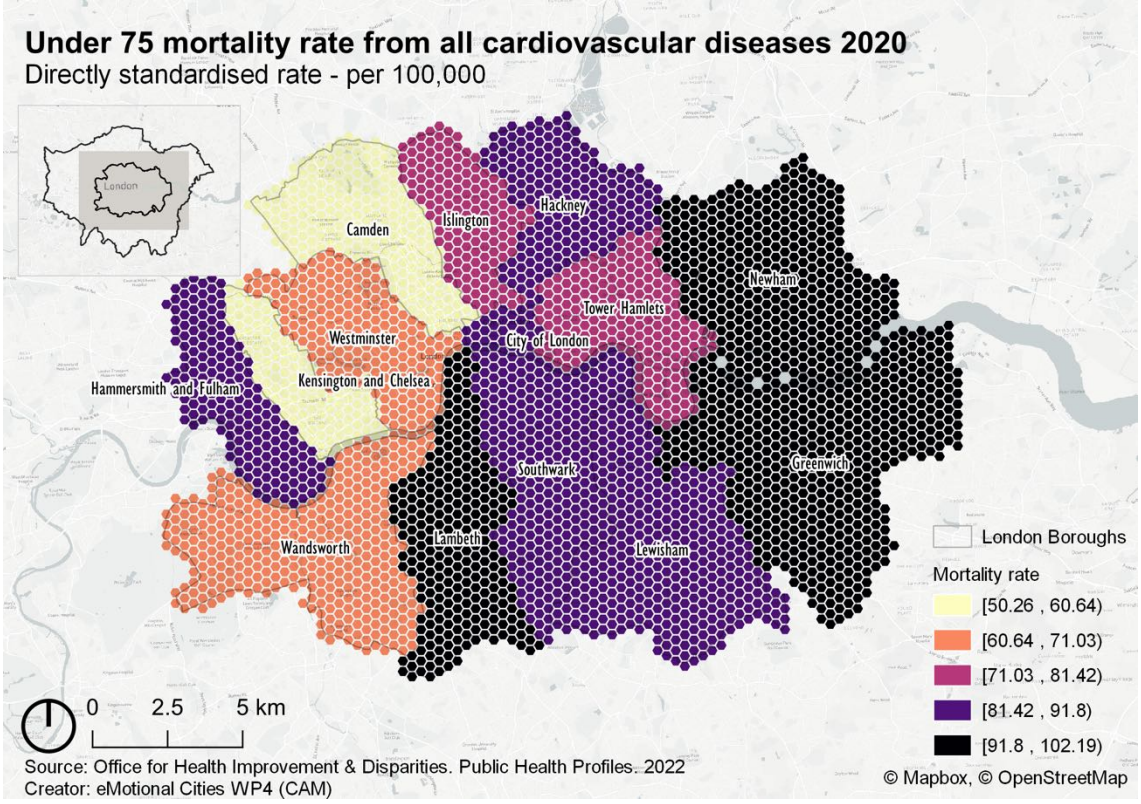


Figure 16 Mortality rate from all cardiovascular diseases in London Boroughs

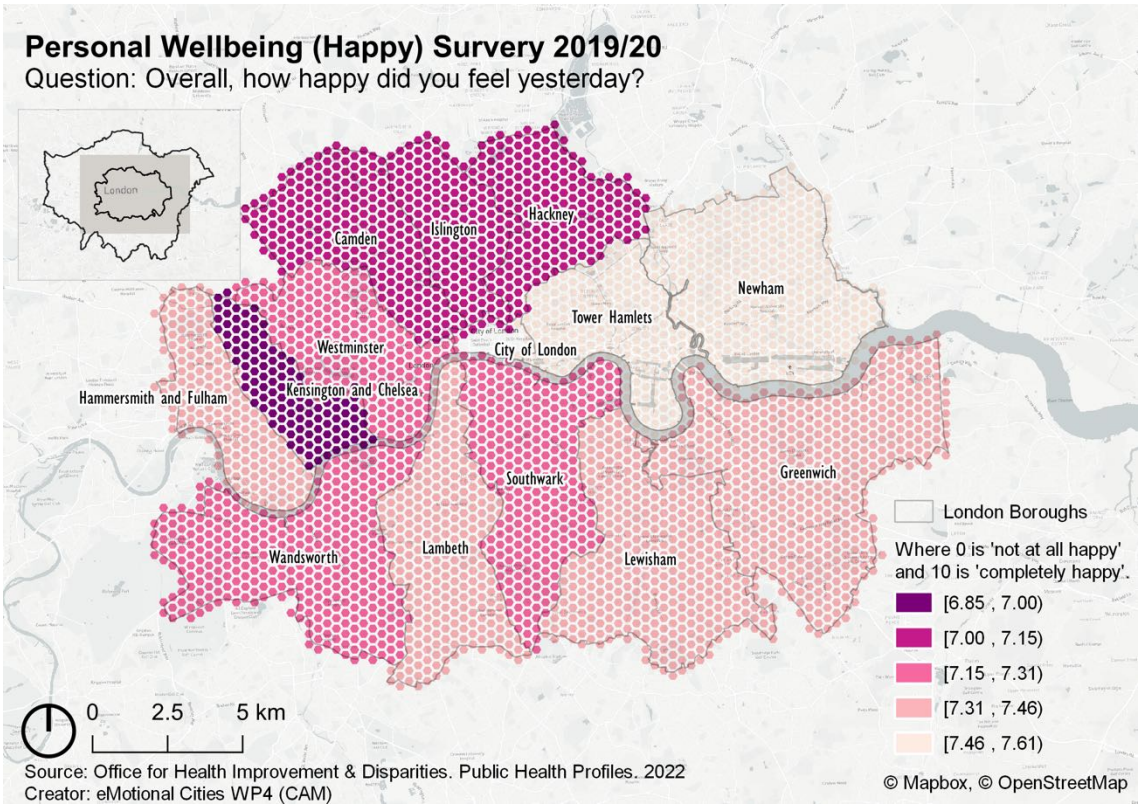


Figure 17 Personal wellbeing (happy) scores in London

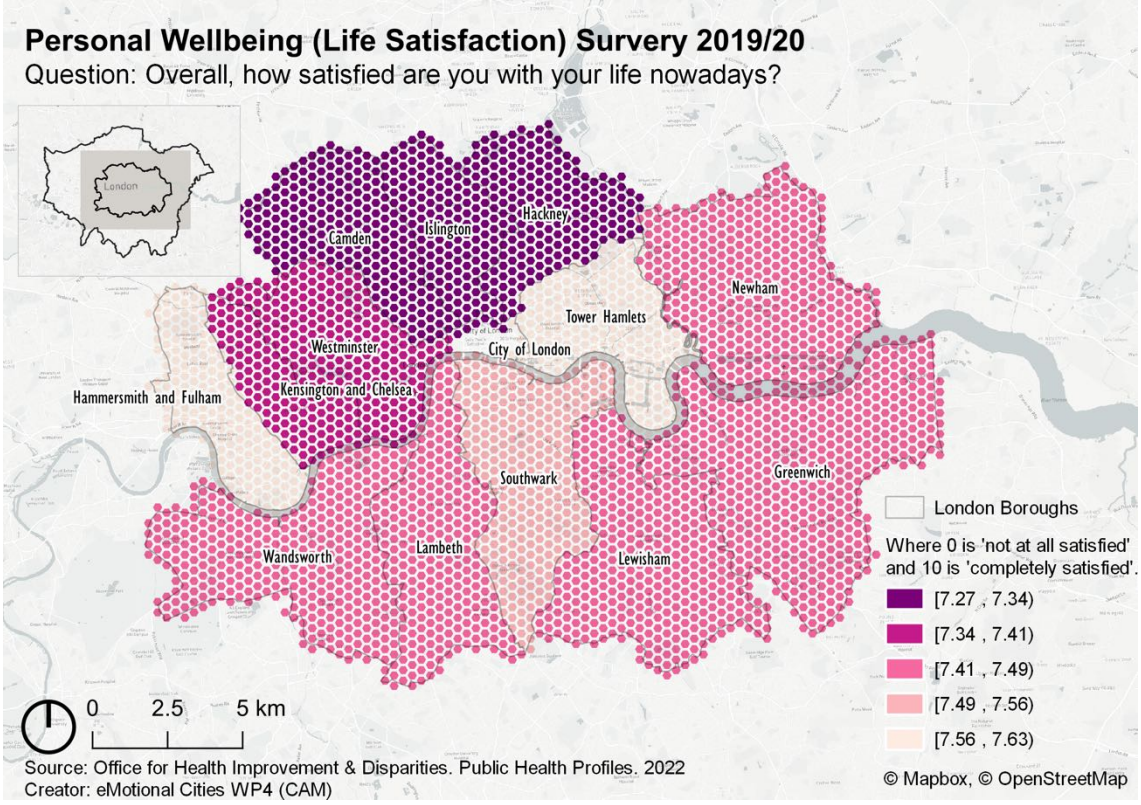


Figure 18 Personal wellbeing (life satisfaction) scores in London

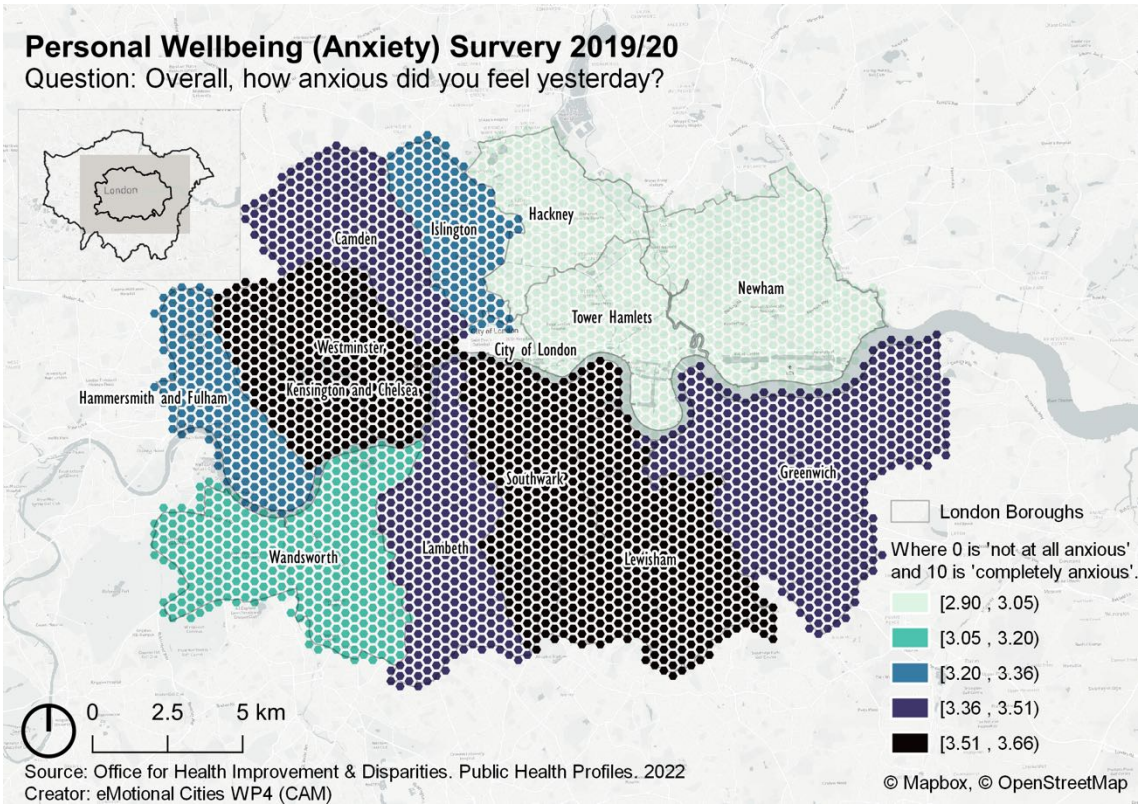


Figure 19 Personal wellbeing (anxiety) scores in London

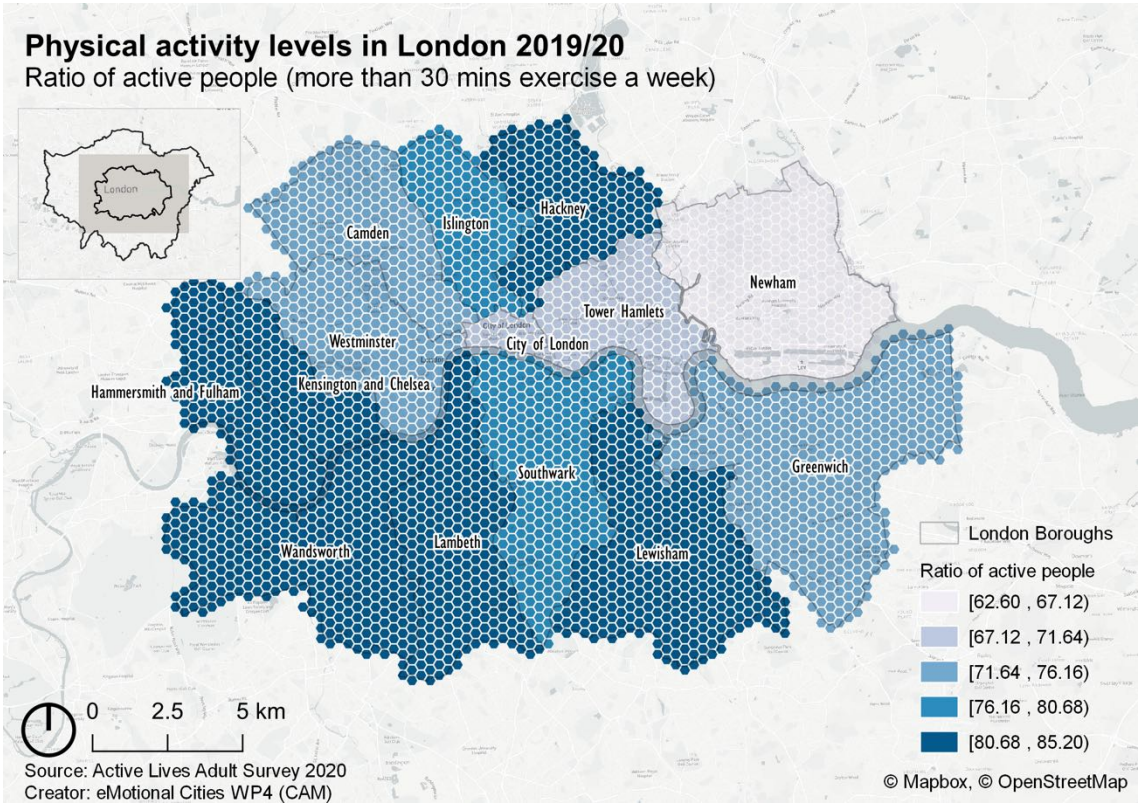


Figure 20 Physical activity levels in London Boroughs (2020)

6.2 Maps of the physical environment

- Overall, the access to public open spaces proportion in Inner London is high. By combining Figure 21 and Figure 22, we find the local places with a lower vegetable index or low accessibility of open space only exist around local places around Westminster and Mayfair and Piccadilly Circus.
- Environmental exposures (PM2.5 and PM10) are mainly concentrated in the densest areas such as the City of London and high streets in the city core. Areas outside of the city core are mainly impacted by rail noise.
- Areas near the Thames River and other rivers in Wandsworth, Tower Hamlets, and Lewisham are highly likely to be impacted by flooding. Among boroughs in Inner London, Newham has more areas with the risk of flooding, although the likelihoods are low.
- There is no clear geographical pattern of surface temperature during summer daytime in Inner London. One major heat spot is identified in Newham, where most areas in this Borough have a mean temperature of over 34°C during summer daytime. Surprisingly, the temperature in Canary Wharf is a Cool Island.

6.2.1 Open spaces

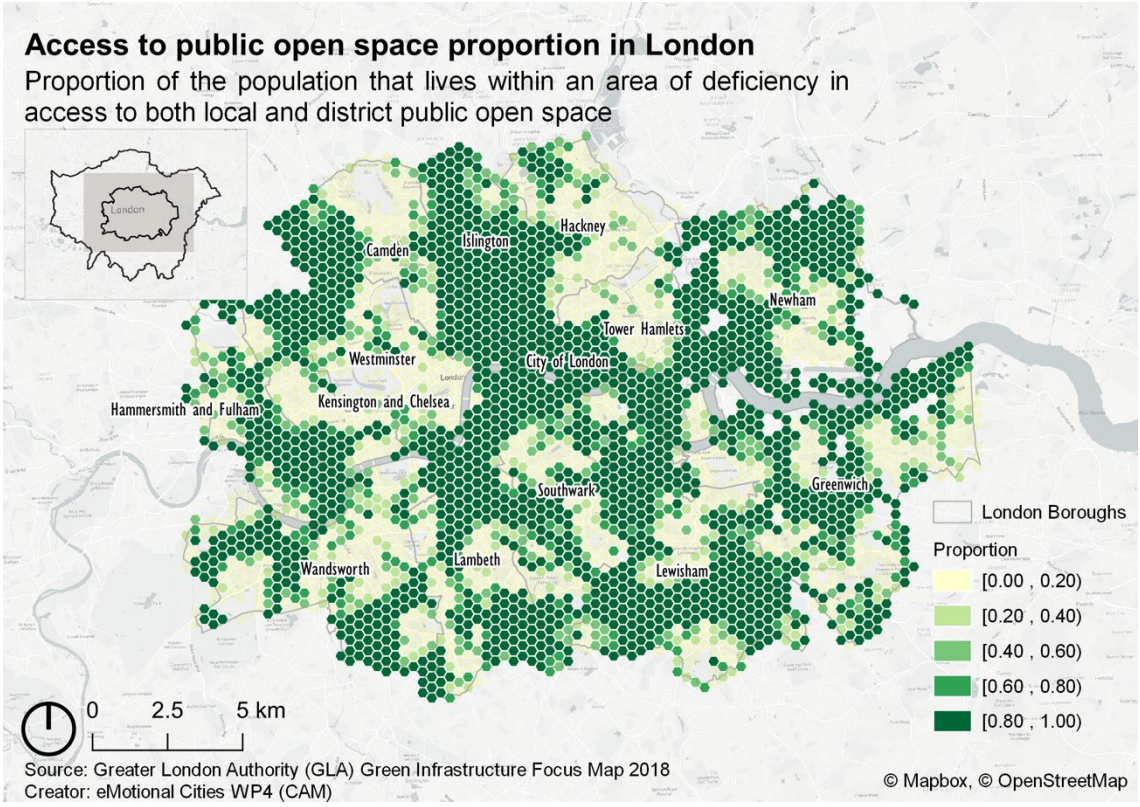


Figure 21 Proportion of population with access to public open space in London

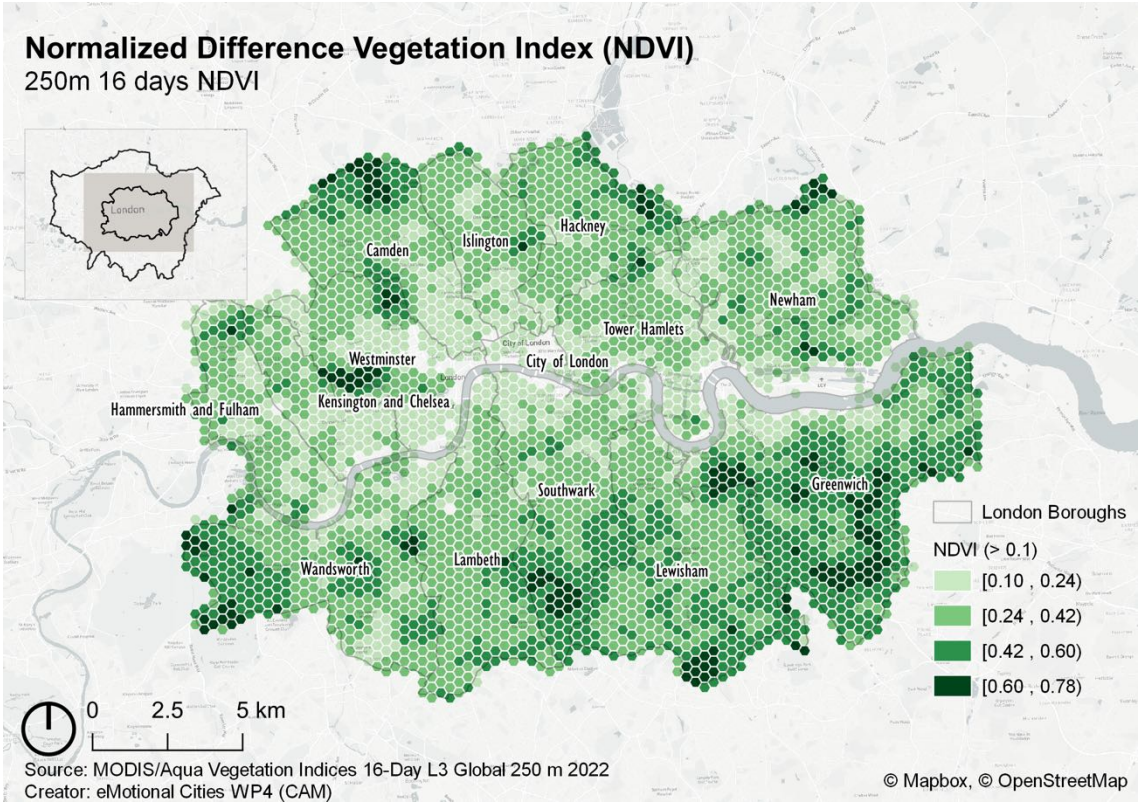


Figure 22 Normalized difference vegetation index (NDVI) in London

6.2.2 Environmental exposures

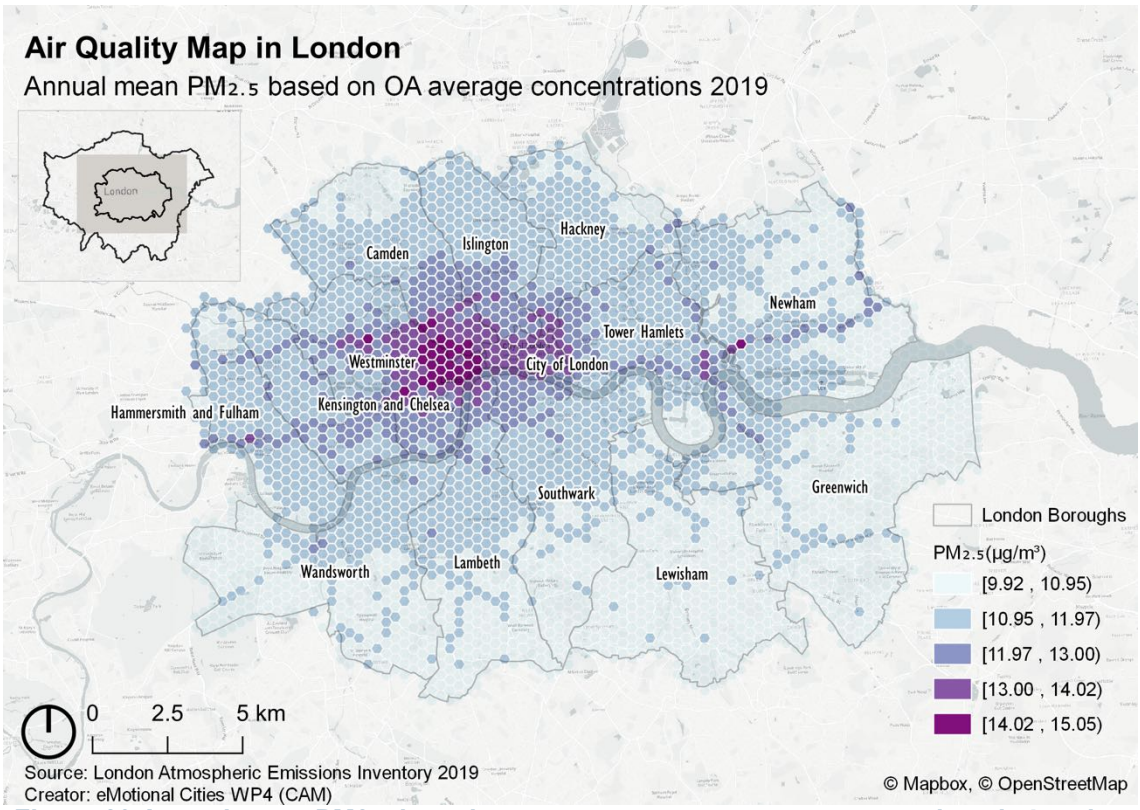


Figure 23 Annual mean PM2.5 based on output areas average concentrations in London

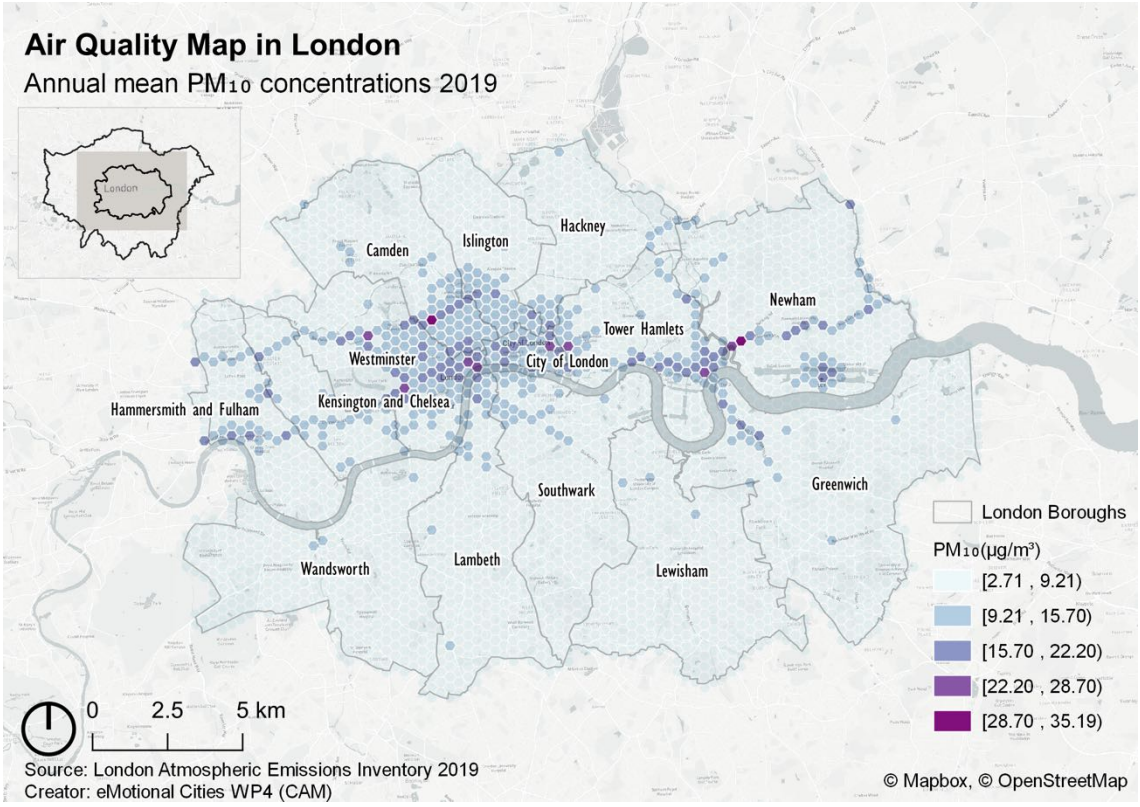


Figure 24 Annual mean PM10 based on output areas average concentrations in London

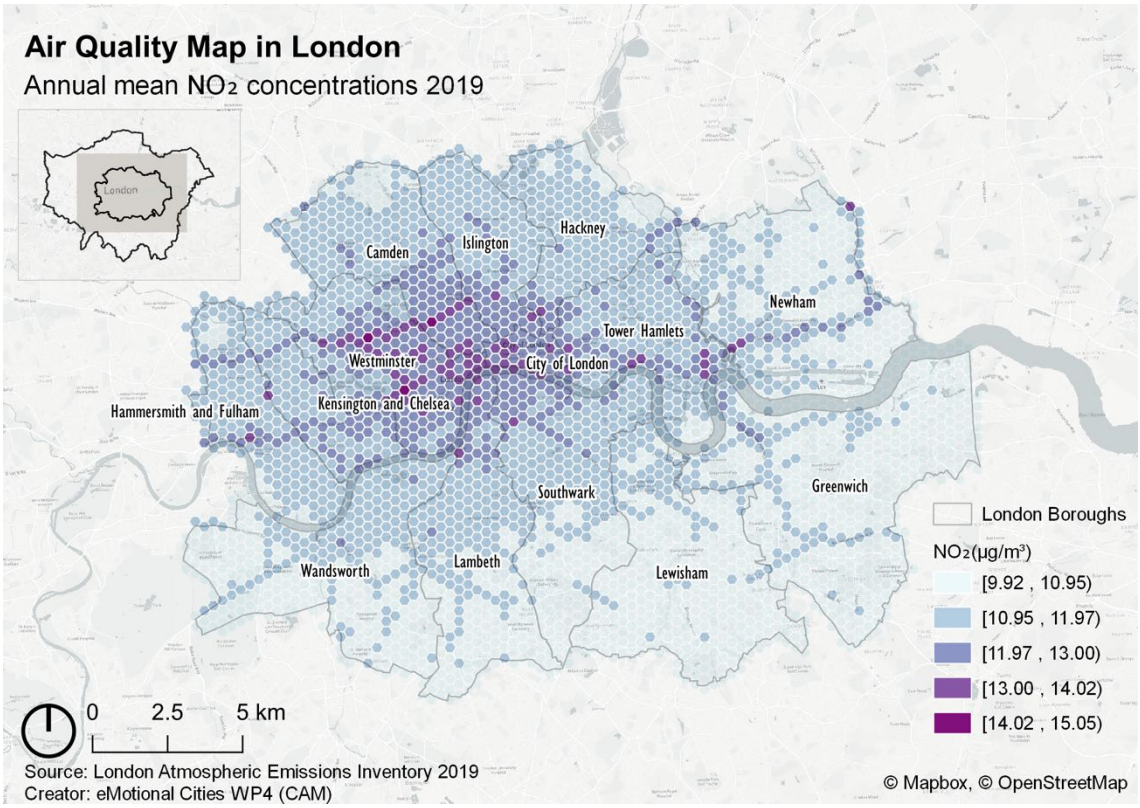


Figure 25 Annual mean NO2 concentration in London

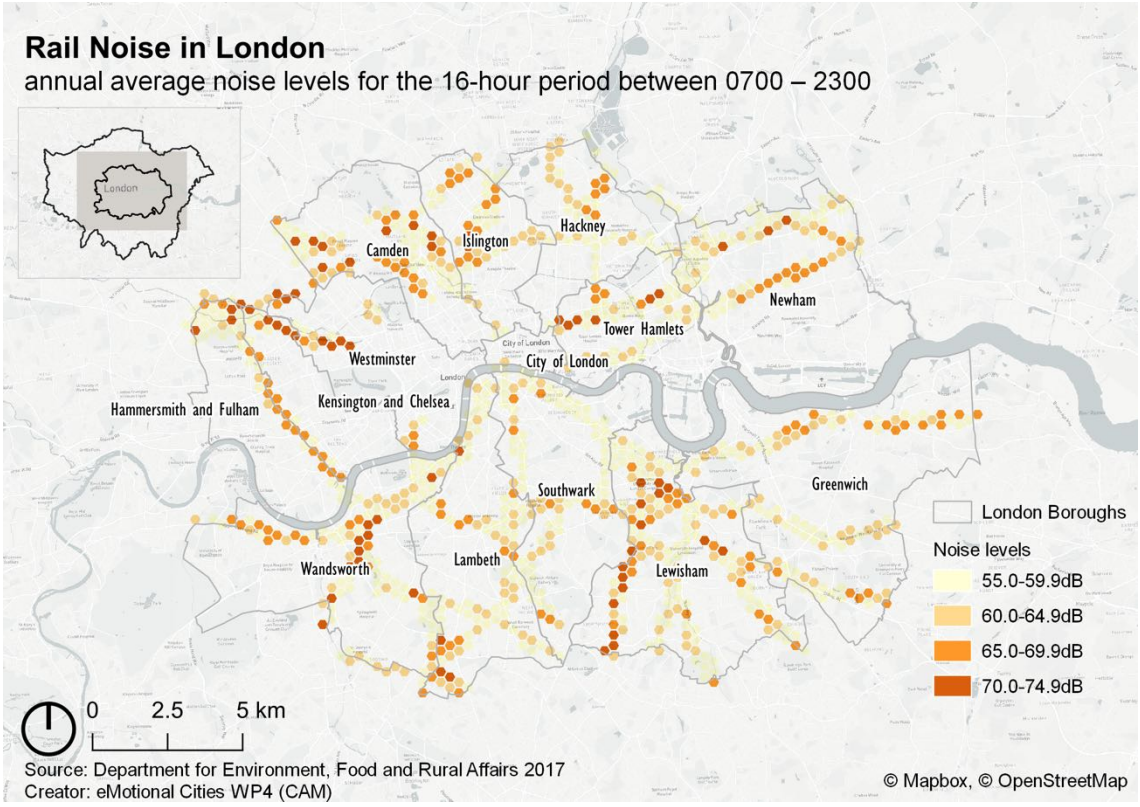


Figure 26 Annual average noise levels of rail noise in London

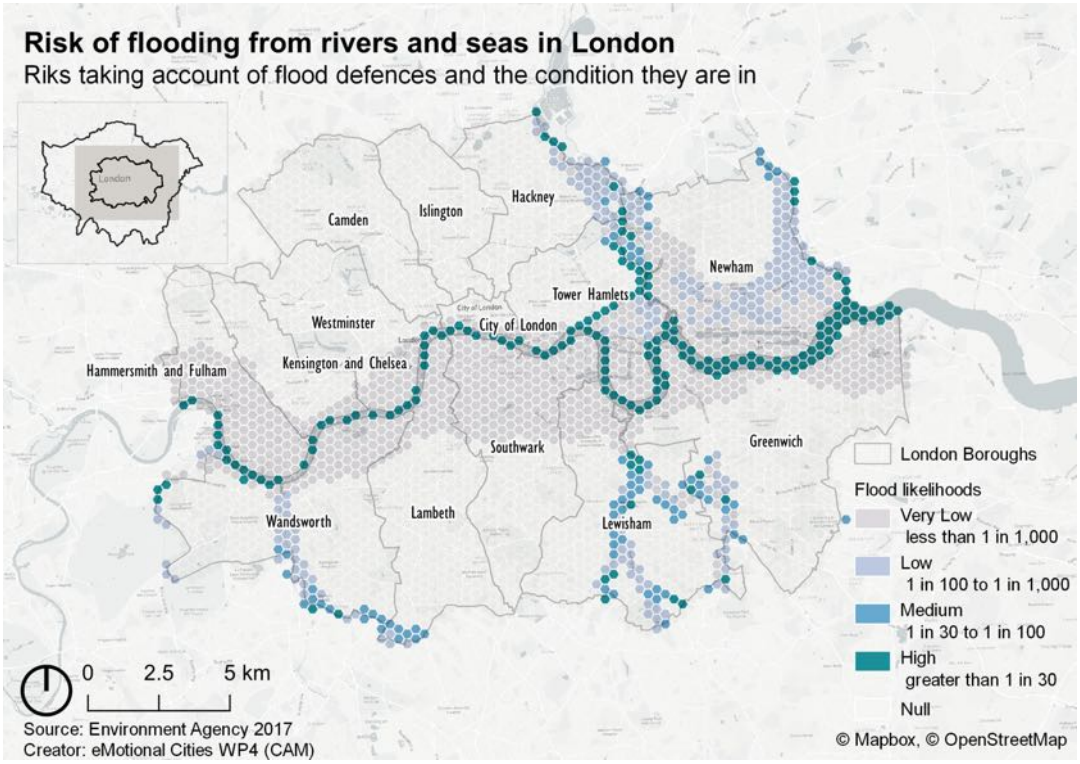


Figure 27 Risk of flooding from rivers and seas in London

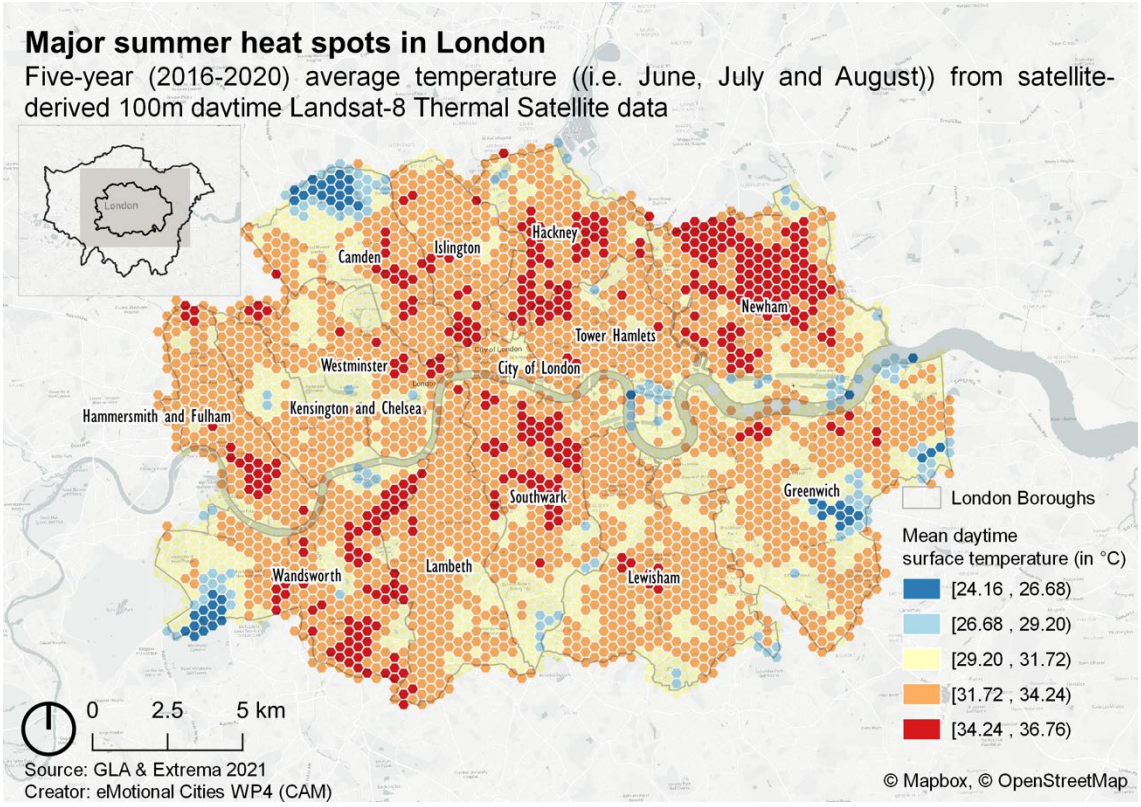


Figure 28 Major summer heat spots in London

6.2.3 Active travel

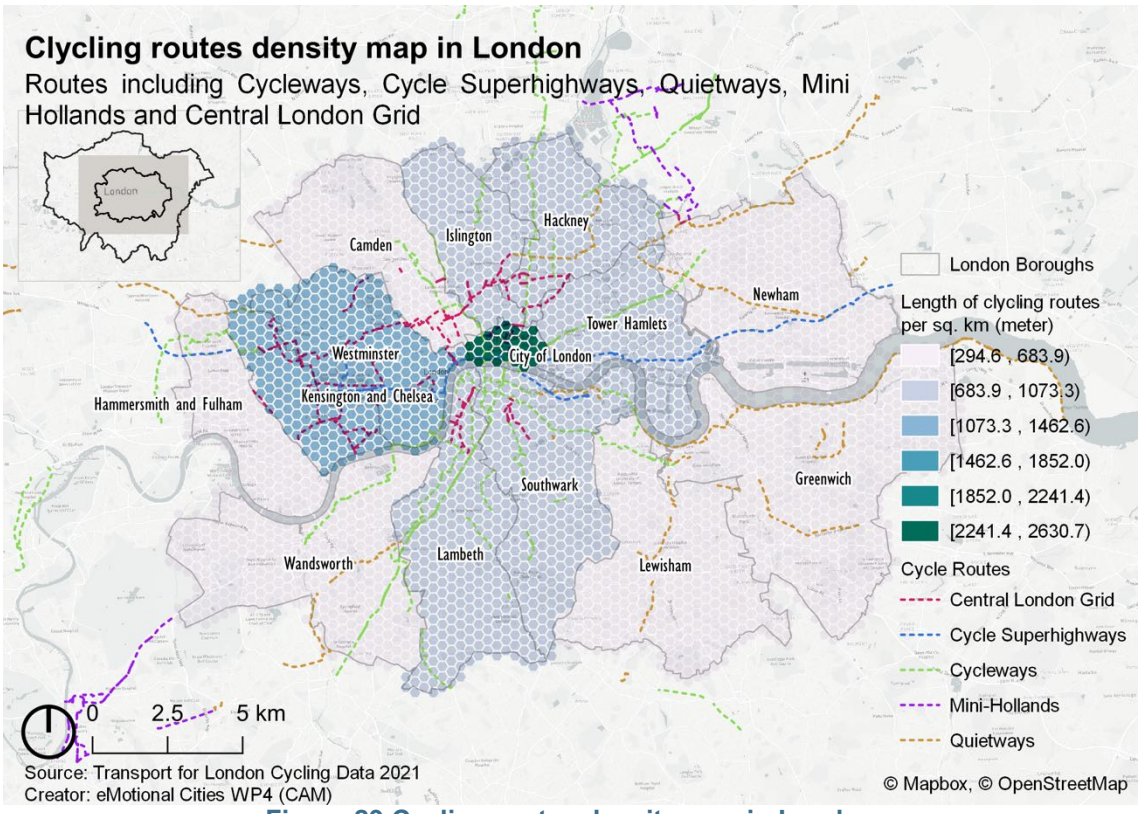


Figure 29 Cycling routes density map in London

6.2.4 Land use

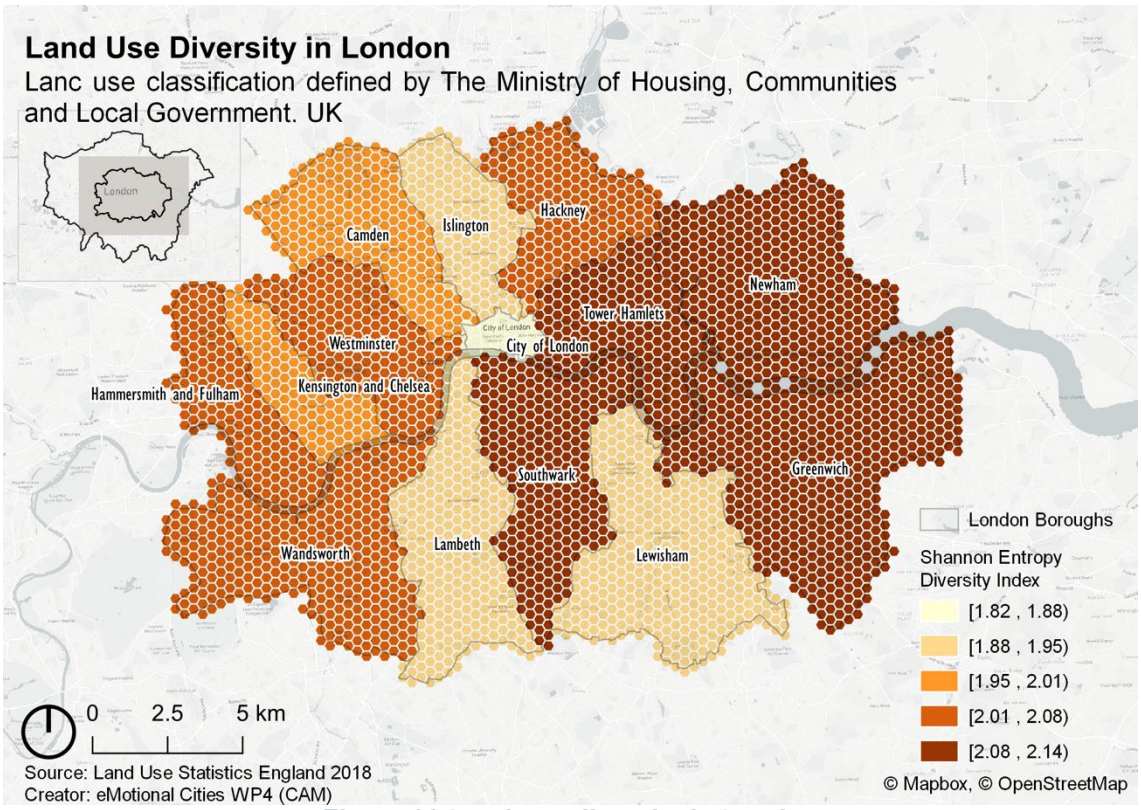


Figure 30 Land use diversity in London

6.2.5 Others (population density, etc.)

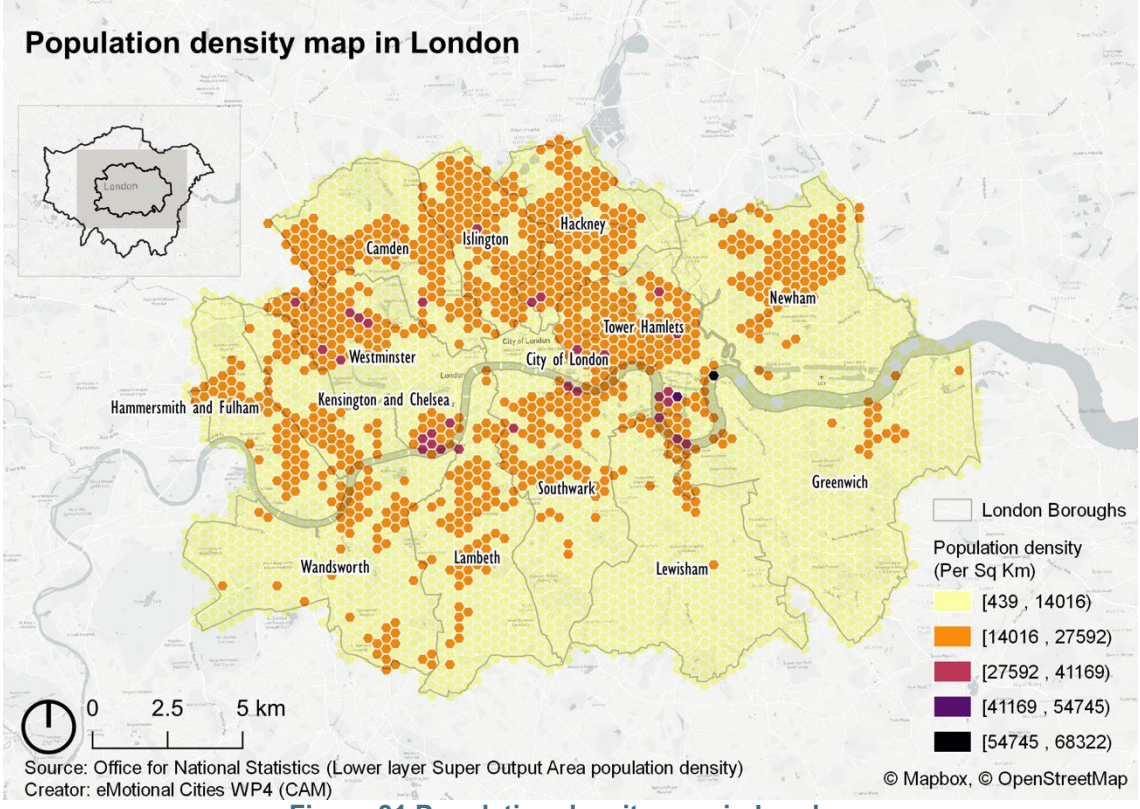


Figure 31 Population density map in London

6.3 Maps of the socioeconomic groups

- Almost half of the areas located within Inner London belong to the 50% least deprived areas of Great London. Deprived communities (in 20% most deprived) can be found in areas of Hackney, Newham and Lewisham.
- Spatial disparities in the elderly ratio exist within Inner London. London Boroughs such as Kensington and Chelsea, Hammersmith and Fulham, City of Westminster and the northern part of Camden, and Greenwich have a high ratio of people over 65 years old. Communities in Chelsea have the highest elderly ratio, up to 40%.
- A clear spatial pattern of gender groups could be found in London. Areas located on the northern bank of the Thames River, especially in the City of Westminster, City of London, and Canary Wharf tend to have a higher proportion of males. Outside of the city core, Newham and Wandsworth both have large gender gaps. Communities from the former have a higher ratio of males while those from the latter have a higher ratio of females.

6.3.1 Deprived community

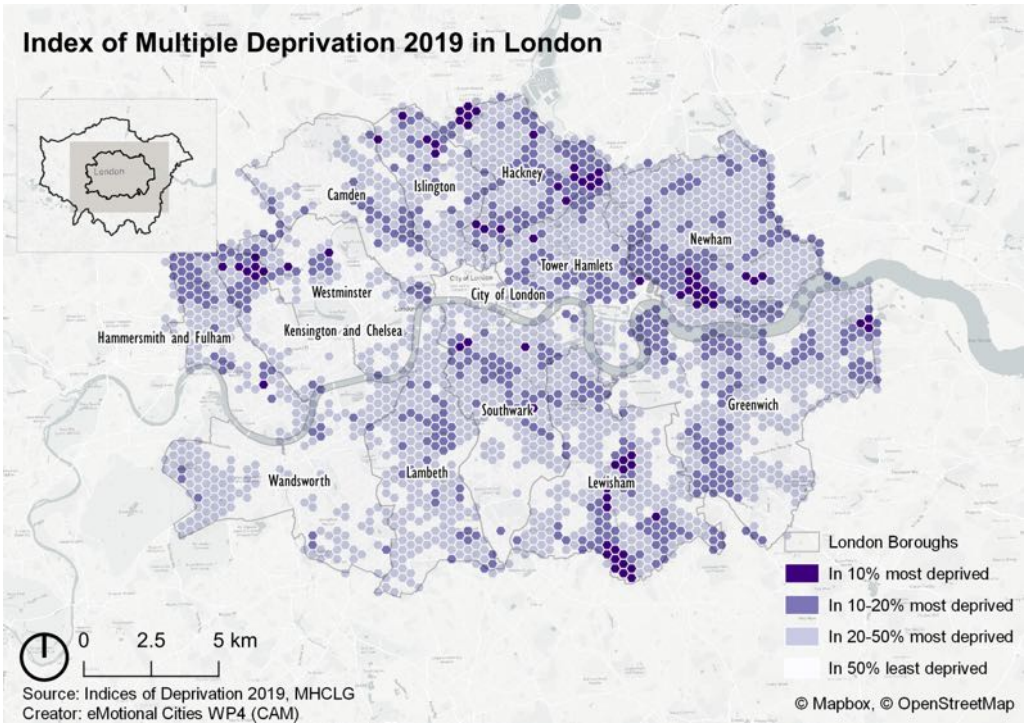


Figure 32 Deprived communities in London

6.3.2 Age

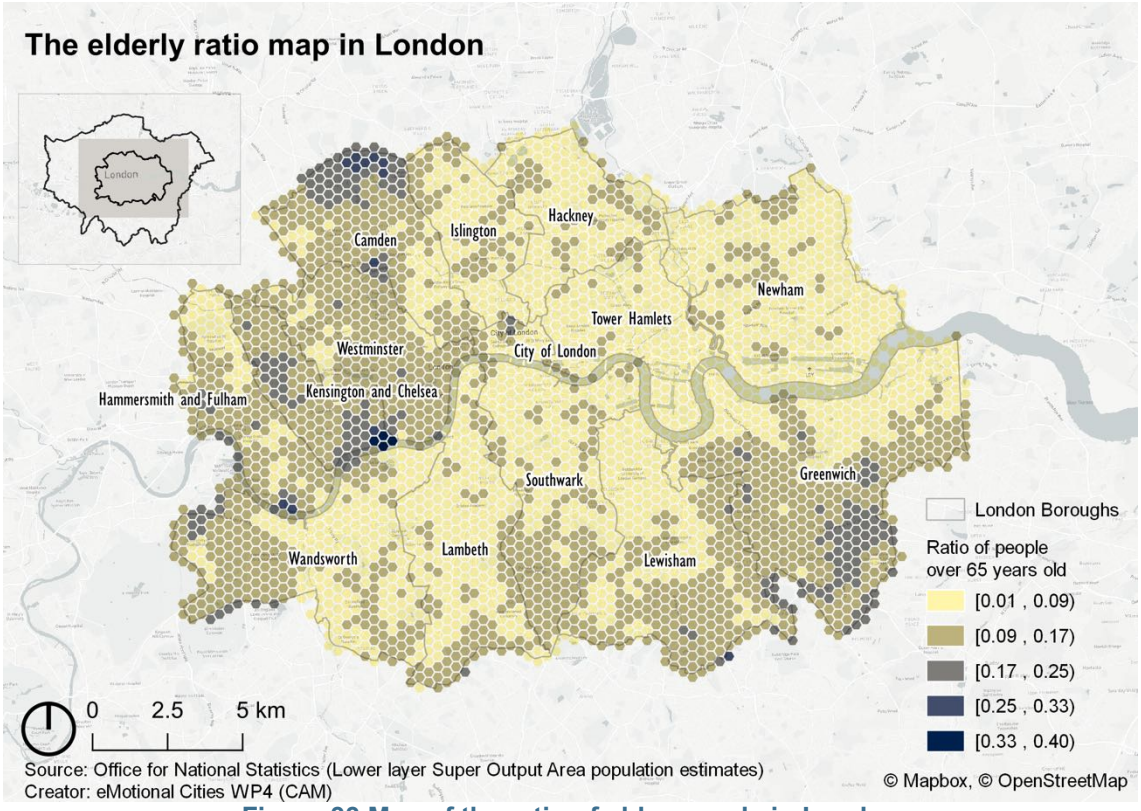


Figure 33 Map of the ratio of elder people in London

6.3.3 Gender

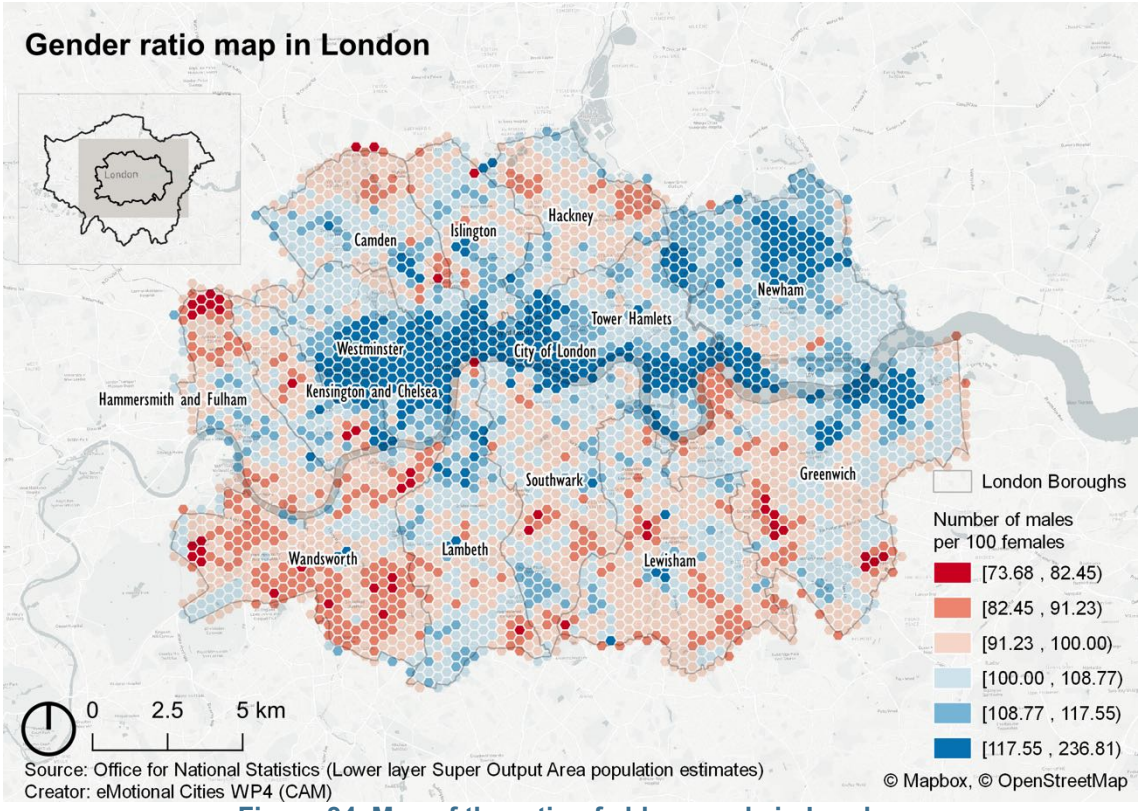


Figure 34 Map of the ratio of elder people in London

7. Mapping Emotions with Social Media Data

In this section, we present several examples of the preliminary results of urban health mapping in London and sentiment analysis in London and Lisbon.

7.1 Mapping of distribution of geotagged social media in London and Lisbon

The mapping shows the distribution of geotagged tweets during August in London and Lisbon. As we are currently at the preliminary stage of sentiment analysis, distribution of tweets with different types of sentiment polarity (positive, negative, neutral) will appear in the next deliverable.

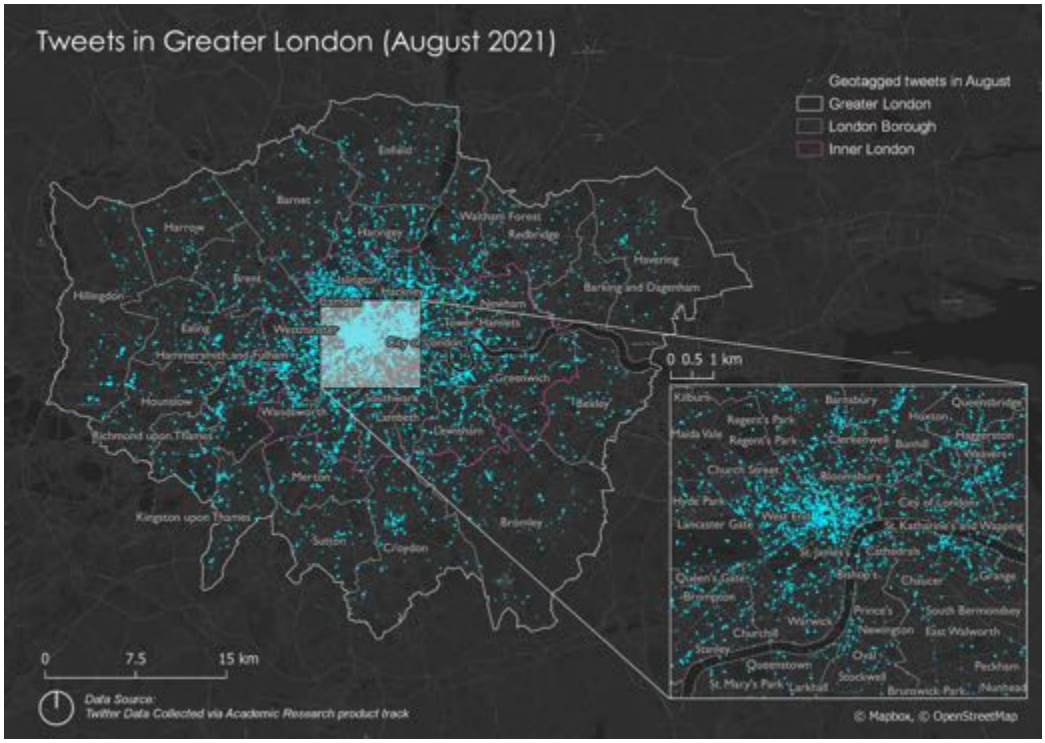


Figure 35 Spatial distribution of geotagged tweets in Greater London

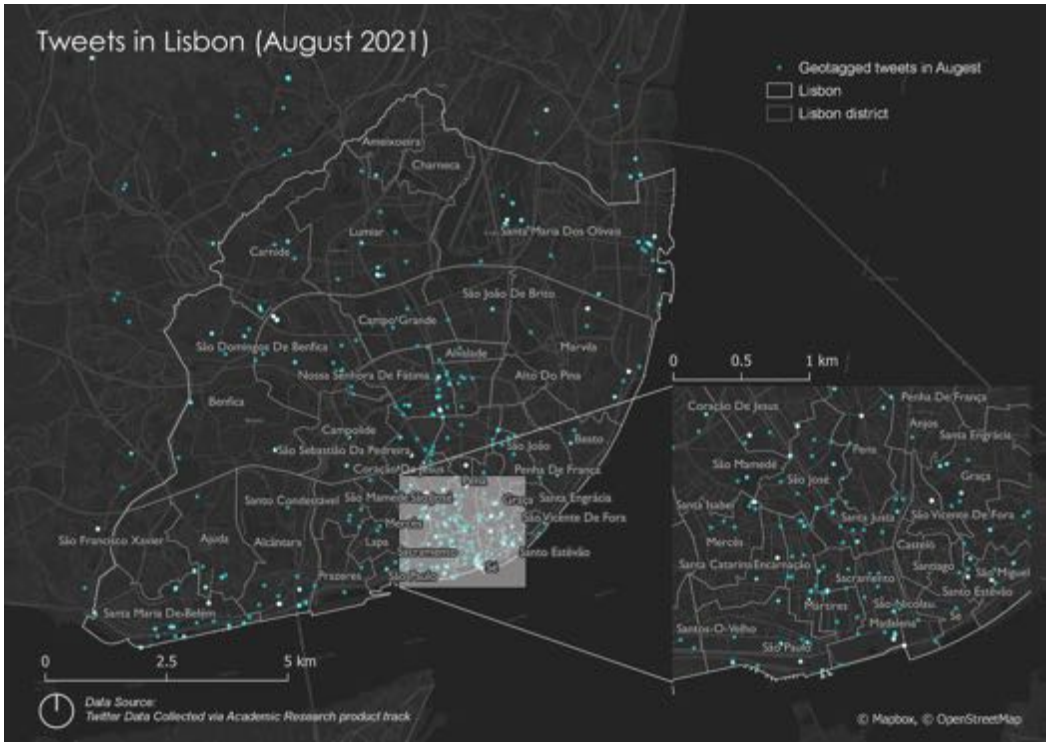


Figure 36 Spatial distribution of geotagged tweets in Lisbon

7.2 Sentiment and Emotions

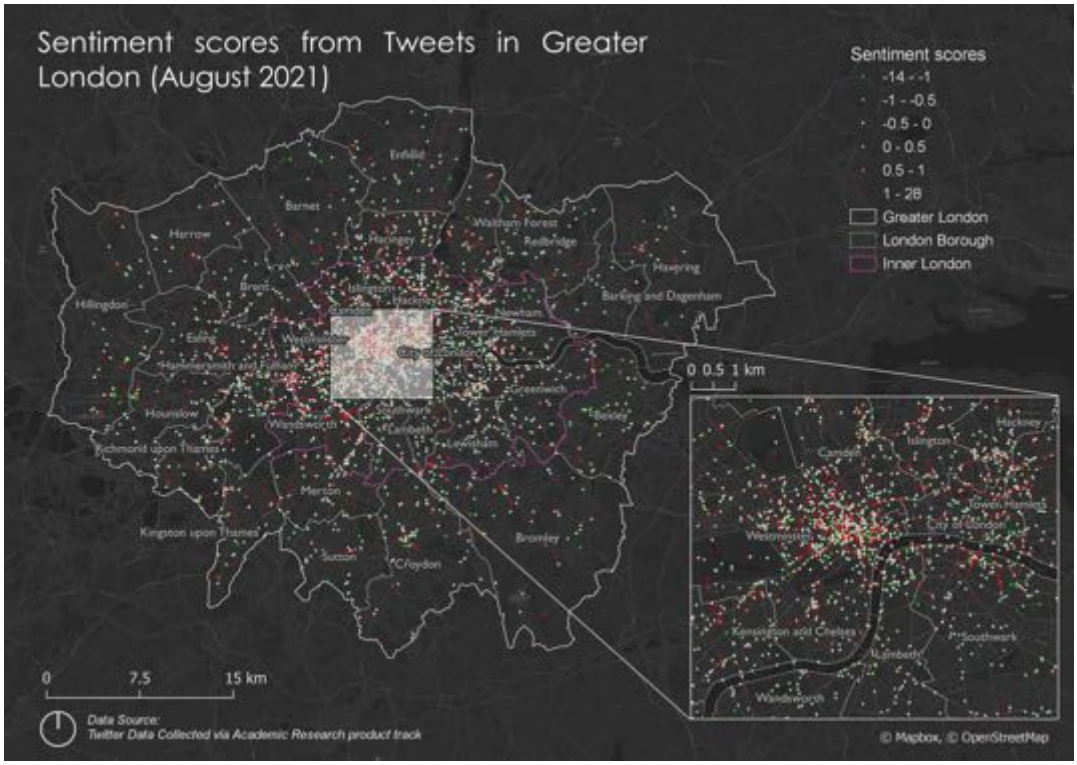


Figure 37 Sentiment polarity map in London

Source: extracted from Annex 3

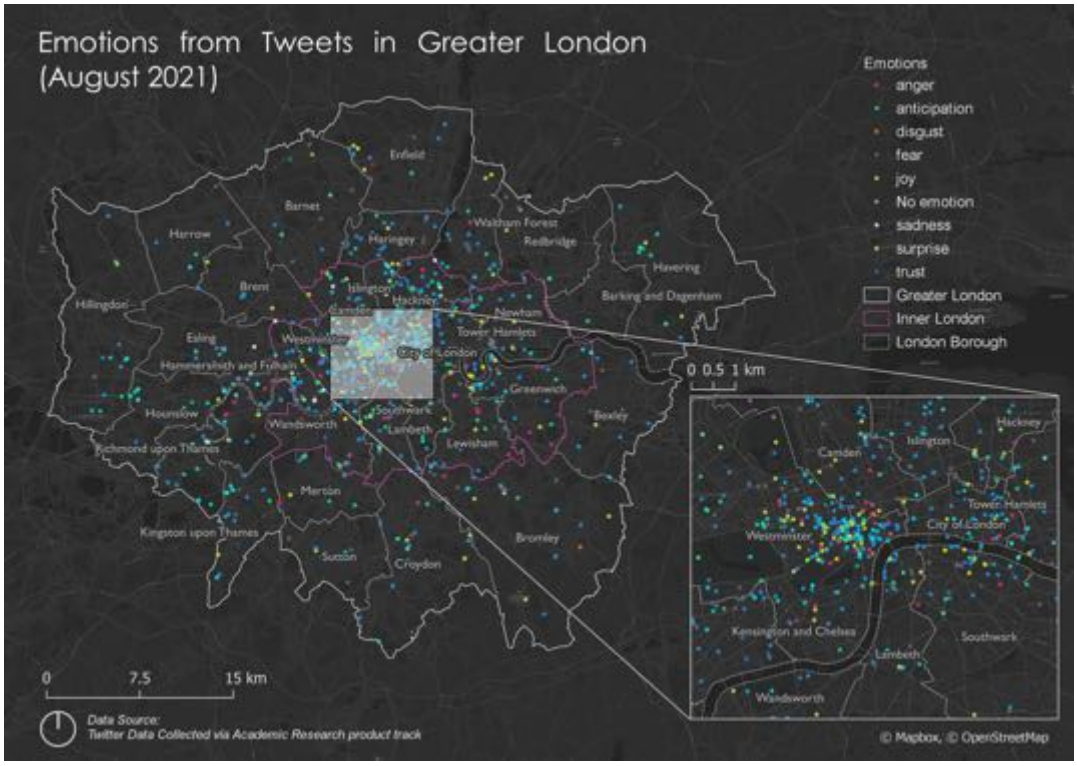


Figure 38 Emotion (eight types) Map in London

Source: extracted from Annex 3

8. Conclusion

This document presents the outcome of the work led by WP4 based on the previous deliverable. Throughout the document, the team first identifies variables and metrics related to the physical environment and socioeconomic groups that are most related to health and well-being outcomes (see Section 1 and Annex 2). Taking the data availability into consideration, the report then deploys the methodology of spatial analysis for urban health mapping and sentiment analysis for hotspot identification (see Section 3 and Annex 3). The report delivers the result of mapping urban health in Europe (Section 4) and London (Section 5).

Appendix 1: Bottom-up Process- Urban health variables and metrics

Name	Category	Category_1	Type	Is it Spatial ?	Level	General description	Results interpretation	Necessary data	Expected comput time	Methodological Reference
Alcohool consumption	Health	Alcohol and drugs consumption	Quantitative	No						
Drugs consumption	Health	Alcohol and drugs consumption	Quantitative	No						
Tabaco consumption	Health	Alcohol and drugs consumption	Quantitative	No						
Cardiovascular diseses prevalence	Health	Cardio-metabolic Diseases	Quantitative	No	European City	Prevalece of disease		statistical data		
Heart rate variability (HRV)	Health	Cardio-metabolic Diseases	Quantitative	No	Non european country - Local Scale	HRV is the variation in the time interval between consecutive heartbeats	Normally higher HRV are associated with relaxing and recovering activities and lower HRV with stress	Study Population measurements with a heart monitor		Influence of green spaces on environmental satisfaction and physiological status of urban residents
Obesity and overweight prevalence	Health	Cardio-metabolic Diseases	Quantitative	No	European City	Prevalece of disease		statistical data		
Type II diabetes prevalence	Health	Cardio-metabolic Diseases	Quantitative	No	European City	Prevalece of disease		statistical data		
comorbidities/difficulties	Health	Cormobidities	Qualitative	No						
Excess heat mortality	Health	Heat Stress	Qualitative	No	European City	Prevalece of disease		statistical data		
Mental Health Disorders - Patient Health Questionnaire (PHQ)	Health	Mental Health	Quantitative	No	General	The PHQ, is a self-administered questionaire version of the PRIME-MD, to assess depression, anxiety, alcohol, eating, and somatoform disorders based on personal symptoms of the past 2 weeks.	Scoring for the PHQ dependes on the type of disorder module questionaire, according to informations in the reference	Study population answers to the Patient Health Questionnaire (PHQ)		Instructions for Patient Health Questionnaire (PHQ) and GAD-7 Measures
Air Quality Management Zones (AQMZs)	Physical Environement	Air Quality	Quantitative	Yes	European City	formulate a theoretical framework for the management of urban ventilation potential and human exposure to air pollution and to 2) develop methods for its implementation by means of a geographic information system (GIS)	Modelled air quality maps	geospatial datasets		A framework for Air Quality Management Zones - Useful GIS-based tool for urban planning

Green space metrics and health outcomes	Physical Environement	Green and blue space	Quantitative	Yes	European City	In this paper, extended those green space analyses with three different satellite sensors in respective spatial resolutions of 2, 5 and 30 m, for buffer distances from home address of 50, 100, 250 and 500 m, to take into consideration of potential existence of unmapped trails to and visual effects from green space. Then compared whether these different estimates of green space influence the health outcomes observed in epidemiological studies. Hypothesize that NDVI derived from higher spatial resolution remote sensing data or greater buffer size would lead to larger health outcomes detection from green space exposure.	Differences in characterizing green space and health outcomes	Data included those of 2 m spatial resolution from WorldView2, 5 m resolution from RapidEye and 30 m resolution from Landsat.	hours	Associations of green space metrics with health and behavior outcomes at different buffer sizes and remote sensing sensor resolutions
Green Spaces as an Indicator of urban health	Physical Environement	Green and blue space	Quantitative	Yes	World-Wide - 28 cities	Mapped land covers of megacities using Landsat images and a random forest classifier running on Google Earth Engine and calculated the availability and accessibility of urban green spaces using the land cover maps and gridded population data.	Urban green spaces to indicate availability of urban green space in each megacity. The percentage of urban green space was calculated. It also used an accessibility indicator of urban green space proposed by the WHO. This indicator is relevant to public health, and is suitable for inter-city comparison.	Satellite images (Landsat)		Green Spaces as an Indicator of Urban Health: Evaluating Its Changes in 28 Mega-Cities
Influence of urban green-blue spaces on human health	Physical Environement	Green and blue space	Quantitative	No	World Wide	How spatial scale, datasets, methods, and analytics are currently applied in studies investigating the relationship between green and blue spaces and human health in urban areas.	Relationship between green and blue spaces and human health in urban areas	Review		Spatial dimensions of the influence of urban green-blue spaces on human health: A systematic review
Neighborhood greenness (NDVI)	Physical Environement	Green and blue space	Quantitative	Yes	Non european country - Local Scale	Greenness is calculated with the Normalized Differential Vegetation Index based on remote-sensing spectral data for a specific area of interest	Scores can range from −1 to 1, where −1 indicates no presence of vegetation and 1 indicates dense levels of vegetation	Satellite spectral imagery data	hours	Validation of the Normalized Difference Vegetation Index as a Measure of Neighborhood Greenness
Urban Green Infrastructure and Health Inequalities	Physical Environement	Green and blue space	Quantitative	Yes	European City	GIS-Based Approach to combine Health, Land-use, Socioeconomics and Ecosystem Services	Single and aggregated health indicators map	spatial datasets (land-cover, health and healthcare consumption, socioeconomics)		Identifying Optimal Locations for Urban Green Infrastructure to Reduce Health Inequalities

Public health impacts of urban air pollution	Physical Environement	Air Quality	Quantitative	Yes	European City	Exposure assessment - PM 2.5 and PM 10	Short and long term effects of exposure to ozone on mortality and hospitalizations.Short-term impacts of exposure to PM10 on hospitalizations. Long-term impacts of chronic exposure to PM2.5 mortality	geospatial datasets		
Health-based assessment of particulate air pollution	Physical Environement	Air Quality	Quantitative	No	Non european country - Local Scale	analysing mortality and morbidity effects of PM10 pollution based on statistical data and the epidemiological exposure–response function	Exposure–response function	statistical data		A health-based assessment of particulate air pollution in urban areas of Beijing in 2000–2004
RAQ – Random Forest for Predicting Air Quality,	Physical Environement	Air Quality	Quantitative	Yes	Non european country - Local Scale	Random forest approach for predicting air quality (RAQ) is proposed for urban sensing systems	The data generated by urban sensing includes meteorology data, road information, real-time traffic status and point of interest (POI) distribution	POI, traffic and road data	hours	RAQ–A Random Forest Approach for Predicting Air Quality in Urban Sensing Systems
Air Quality Indicators	Physical Environement	Air Quality	Quantitative	Yes	Non european country - Local Scale	TROPOspheric Monitoring Instrument on board Sentinel-5 Precursor, shows an annual mean of high-resolution maps of selected air quality indicators (NO2, CO, O3, and UVAI) of the MENA countries for the first time. The correlation analysis among the aforementioned indicators show the coherency of the air pollutants in urban areas	Spatial Variability and Properties of Aerosol over the Selected area	Multiple Satellites and AERONET Data	hours	Urban Health Related Air Quality Indicators over the Middle East and North Africa Countries Using Multiple Satellites and AERONET Data
Multifunctional Green Infrastructure (GI)	Physical Environement	Green and blue space	Quantitative	Yes	European City	Study enclosed 18 indicators, as well as identified hot and cold spots of selected GI functions and their multifunctionality	Spatial distributions of hot/cold spots of GI functions.	LULC from urban atlas Indicator framewok from several source	Hours	An Integrated Indicator Framework for the Assessment of Multifunctional Green Infrastructure
Health effects of the natural outdoor environment	Physical Environement	Green and blue space	Quantitative	No	European City	Epidemiological studies to examine long-term and short-term of the natural environment. Examining the underlying mechanism in the daily life setting.	Investigates the interconnections between natural outdoor environments and better human health and well-being.	Landuse planning management Quantitative and qualty characteristics.	hours	Positive health effects of the natural outdoor environment in typical populations in different regions in Europe (PHENOTYPE)

Appendix 2: Paper 1

Appendix 3: Paper 2

eMOTIONAL
Cities

Mapping the cities through the senses
of those who make them