Portfolio of Metrics and Methods for Urban Health
<table>
<thead>
<tr>
<th><strong>Project Title</strong></th>
<th>eMotional Cities</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Deliverable</strong></td>
<td>Deliverable 4.1. – Portfolio of metrics and methods for urban health</td>
</tr>
<tr>
<td><strong>Work package</strong></td>
<td>WP4</td>
</tr>
<tr>
<td><strong>Task</strong></td>
<td>T4.1, T4.2, T4.3</td>
</tr>
<tr>
<td><strong>Number of pages</strong></td>
<td>41</td>
</tr>
<tr>
<td><strong>Dissemination level</strong></td>
<td>Public</td>
</tr>
<tr>
<td><strong>Leader</strong></td>
<td>University of Cambridge (CAMB)</td>
</tr>
<tr>
<td><strong>Contributors and peer-reviewers</strong></td>
<td>Instituto de Geografia e Ordenamento do Território da Universidade de Lisboa (IGOT), Faculdade de Medicina da Universidade de Lisboa (FMUL), Michigan State University (MSU), Tallinna Tehnikaülikool (TT)</td>
</tr>
<tr>
<td><strong>Date</strong></td>
<td>04/01/2022</td>
</tr>
<tr>
<td><strong>File name</strong></td>
<td>eMCities_WP4_D4.1</td>
</tr>
<tr>
<td><strong>Version</strong></td>
<td>V02</td>
</tr>
<tr>
<td><strong>Main author</strong></td>
<td>Prof. Elisabete A. Silva (CAMB), Dr. Haifeng Niu (CAMB), Ana Paula Seraphim (CAMB)</td>
</tr>
</tbody>
</table>
Index

Executive Summary ........................................................................................................ 6

1. Introduction .................................................................................................................. 7
   1.1 Spatial data analysis for urban health ................................................................. 7
   1.2 WP4 Objectives and Relation to other Project WPs .............................................. 8
   1.3 Report structure ...................................................................................................... 9

2. Urban Health Variables and Metrics ......................................................................... 11
   2.1 Bottom-up Process – Emotional Cities Members Input ..................................... 11
   2.2 Policy-to-science - Reports Scanning ................................................................. 16
      2.2.1 Healthy food environment ........................................................................... 17
      2.2.2 Crime prevention environment ................................................................... 18
      2.2.3 Active Travel environment .......................................................................... 19
      2.2.4 Multifunctional open spaces ...................................................................... 21
      2.2.5 Environmental Exposures ........................................................................... 22
      2.2.6 Social, work, and housing infrastructure ............................................... 23
      2.2.7 Land Use ...................................................................................................... 24

3. Urban Health Mapping ............................................................................................... 26
   3.1 Case studies ........................................................................................................... 26
   3.2 The framework of spatial analysis considering the data availability .......... 26

4. Sentiment analysis and hotspot identification .......................................................... 31
   4.1 Twitter API and settings ..................................................................................... 32
   4.2 Pilot result ............................................................................................................. 33
   4.3 Data collection and processing ........................................................................... 33
   4.4 Sentiment analysis ............................................................................................... 34

5. Preliminary Results of urban health mapping and sentiment analysis ................. 35
   5.1 Examples of urban health mapping in London ................................................. 35
   5.2 Mapping of distribution of geotagged social media in London and Lisbon ... 38

6. Conclusion .................................................................................................................... 40
Index of Tables

Table 1 Twitter APIs used for data collection ..................................................................... 32
Index of Figures

Figure 1 Timeline of Deliverables and Tasks for WP4 ........................................... 9
Figure 2 Three-layer strategy for the identification of urban health variables and metrics .......................................................... 11
Figure 3: Bottom-up variables categories............................................................. 15
Figure 4 Selected cases in Europe (London, Lisbon, and Copenhagen)............... 26
Figure 5 Framework of spatial analysis of urban health....................................... 29
Figure 6 Measures lists for spatial analysis with urban health-related data. Variables and metrics in bold are chosen for spatial mapping at the current stage. ......................... 31
Figure 7 Workflows of sentiment analysis .......................................................... 34
Figure 8 Diabetes prevalence in London Boroughs (2020)................................. 36
Figure 9 Mortality rate from all cardiovascular diseases in London Boroughs (2020). 36
Figure 10 Physical activity levels in London Boroughs (2020)............................ 37
Figure 11 Spatial distribution of geotagged tweets in Greater London............... 38
Figure 12 Spatial distribution of geotagged tweets in Lisbon.............................. 39
Executive Summary

This document first presents the portfolio of metrics and methods to inform our analysis of urban health in selected case studies. The portfolio identifies variables and metrics related to the urban physical and socioeconomic environment that most affect physical and mental health outcomes. The identification process follows a three-layer strategy composed of eMotional Cities member inputs (data, reports and feedback to methodology), scientific research review and policy reports scanning. The following sections of this document then explain the methodology of spatial analysis for urban health mapping and sentiment analysis for hotspot identification, considering the data (traditional data and social media data) availability in the case studies. Preliminary results of both urban health mapping and hotspots identification are also presented.
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 equipment¹. 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 data².

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. It refers to "techniques that are used to analyse and acquire profound knowledge out of urban data"³. 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-consuming⁴. 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 trends⁵. 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 studies⁶. 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 uses⁷. Adding to this big data availability, new data analytics’ methods and vast computation power, promote potential solutions more adjusted to local characteristics and variations.

⁵ 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
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 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 (D4.1.), instead of just presenting the results for T4.1 we will present the methodology and first results of all our work so far in these 3 tasks.

The first section of the report ‘Urban Health Variables and Metrics’ 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 ‘Urban Health Mapping’ is linked with T4.2, and it explain the methodology of spatial analysis for urban health mapping, which was build based on the results of T.4.1, however considering the data (traditional data and social media data) availability in the case studies. The third section ‘Sentiment analysis and hotspot
identification’ is linked both with T4.2 and T4.3 and it explains the methodology for the sentiment analysis for hotspot identification using twitter geotagged data. Finally, the fourth section presents the preliminary results of both urban health mapping and hotspots identification.
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 (Deliverable 2.2 – Preliminary eMOTIONAL Cities conceptual framework), and in other scientific developments of WP4 (these 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.

![Figure 2 Three-layer strategy for the identification of urban health variables and metrics](image)

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;

GA 945307

Deliverable 4.1 | Metrics and Methods for Urban Health
December 2021
• 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 elder 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 (NDVI)** – 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: stress and aggression reduction; cognitive restoration; reduced crime; increase physical activity and social interaction; better 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 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, Urban Heat Island, 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;

● **Urban Heat Island (UHI) implications on health** - UHI effect with excess heat mortality;

● **Urban Heat Island (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;

● **RAQ – Random Forest for Predicting Air Quality** - based on meteorology 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. 1 entry is related to general comorbidities, 1 with heat stress (Excess heat mortality), and 1 with mental health (Patient Health Questionnaire - PHQ).

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

This internal bottom-up process 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.

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 statical aggregated data and questionnaires. Finally, the indicators associated with socioeconomic variables, include age groups, socioeconomic status, ethnic minorities, from government statical aggregated data and questionnaires.

For all the variables the information about the spatial level, necessary 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 Scanning

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 we will translate the spatial analysis and spatial action most of our work.

We are analysing policy documents in 3 levels: the World Level, the Europe Level, and the City Level. The world level includes policy reports from intergovernmental agencies at the global level, such as the United Nations (UN) and the World Health Organization (WHO). The Europe Level includes reports from intergovernmental agencies in the Europe Region, such as the European commission. 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 same levels of groups will later be used to develop our spatial analysis strategy.
In the review we considered any policy report from 2010 forward in the World, Europe and City Levels that approaches the impact of the urban physical environment on the health of different socioeconomic groups. To find the policy reports we searched the webpages of the relevant government agencies and 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 expended the results through citation searching.

We don’t have the final document selection yet for all levels and regions. Nonetheless, currently, at the world level we are considering 23 policy reports from the following agencies: World Health Organization (WHO); United Nations Human Settlements Programme (UN-Habitat); United Nations (UN), and United Nations International Children’s Emergency Fund (UNICEF). At the Europe Level we are considering 9 reports from the European Commission and Eurocities.

Moreover, at the City Level, in the London region, we are considering a total of 17 documents, from the National Health Service (NHS), Public Health England (PHE), Mayor of London, and Transport for London (TfL). In the Lisbon region we are considering 7 documents, from the Serviço Nacional de Saúde (SNS), Câmara Municipal de Lisboa, Rede Social Lisboa, and Rede Portuguesa de Cidades Saudáveis. In the Copenhagen region we are considering 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 73 policy documents.

We have so far systematised the scanning results for the policy documents for London and Lisbon, which will be complemented by the World, Europe, Copenhagen, and Lansing levels. From the reports we extracted the same 3 types of variables of the bottom-up process, which are: physical environment variables, health variables, and socioeconomic groups variables.

These 3 types of variables are organised in relation to 7 topics, identified in the reports as important urban characteristics regarding health improvement: (i) healthy food environment; (ii) crime prevention environment; (iii) active travel environment; (iv) multifunctional open spaces; (v) environmental exposures; (vi) social, work, and housing infrastructure; and (vii) land use. In this section we are going to present each of the 3 types of variables identified in the reports relating them with the 7 topic groups.

2.2.1 Healthy food environment

The healthy food environment topic is related to enabling more people access to healthier food options, leading to change in eating and food purchasing behaviour.
Two physical environment variables were identified in this topic:

- **Accessibility to healthy food and drink**, which is linked with ensuring that healthier food and drink options are physically accessible. Some examples of strategies for improving healthy food and drink access include: the installation of public drinking fountains; the diversification of the healthy food retail options; the avoidance of over-concentration of hot food takeaways; the limitation of alcohol and smoking outlets; and the limitation of advertising of food and drink high in fat, sugar, and salt across public infrastructure.

- **Green Infrastructure and Biodiversity**, which is linked with the opportunity of growing food in allotments, community gardens, and at home, leading to learning more about whole foods and gaining access to affordable vegetables and fruits. For that, cities can establish minimum requirements for food growing spaces, make use of spaces such as roofs, walls, and balconies, and consider edible landscaping.

The topic is related to the improvement or reduction of risk of 11 health variables: mental health; physical health; oral health; obesity; heart disease; stroke; cancers; type 2 diabetes; low birthweight; diet; and drugs, alcohol, and tobacco misuse.

And, regarding the socioeconomic groups, a healthy food environment has a greater impact on the eating behaviour of 4 groups: the younger, the elder, the less mobile, and deprived communities. Therefore, the prioritisation of healthy food options is of special relevance on high streets, deprived areas, and around schools.

### 2.2.2 Crime prevention environment

The **crime prevention environment** topic is related to Increasing safety, reducing crime and ‘fear of crime’, and diverting people at risk of offending.

Poor urban design can exacerbate crime and feelings of unsafety by creating under-used and isolated spaces, without natural surveillance. Seven physical environment variables were identified in this topic:

- **Public space permeability**, which allows people to move freely, with strong intervisiblility, and several access and exit points. Visual and physical barriers such as high walls and fences, gated communities, heavy traffic, and high traffic speeds, and, in some cases, dense vegetation should be avoided since they lead to people feeling insecure and difficult pedestrian mobility.

- **Control of construction and demolition sites**, to alleviate perceived unsafeness associated with those areas.
• **Natural surveillance**, which means that public spaces are overlooked by people using these spaces and by people inside buildings through glazed openings.

• **Effective lighting**, which is orientated towards pedestrian rather than motorised activity and with attention to the excessive use of light.

• **Good maintenance and upkeep**, since according to the broken window theory, even minor signs of deterioration can lead to further damage and increase in criminal activity. For this, it is important to make sure public spaces are free of rubbish, fallen leaves, potholes, and deteriorated street furniture. Good upkeep can also lead to more people using the space.

• **Attractive views and activities**, attractive and interesting views and activities – such as public art, iconic buildings, varied shop frontages, lighting installations, and opportunities for informal play and social activities – can lead to community pride, a sense of place, and ownership that leads to individual care and maintenance for those areas, discourages illegitimate uses, and lead to more people using the space.

• **Green Infrastructure and biodiversity**, well-cared vegetation, and water features can improve visual aesthetics, encouraging the use of spaces and improving the sense of place and ownership.

The topic is related to the improvement or reduction of the following health variables: mental wellbeing, anxiety, sleep deprivation, social isolation because of fear, active travel, children outdoor play, disability, victimisation.

And the following socioeconomic groups are more likely to feel unsafe and suffer from crime: women, younger and older people, people with disabilities, and residents of deprived areas.

### 2.2.3 Active Travel environment

The **active travel environment** topic is related to encouraging walking, cycling, and public transport in everyday journeys, alternatively to car use.

Twelve physical environment variables were identified in this topic:

• **Walking and biking infrastructure**, which is safe, connected, and convenient, linking key destinations – such as homes, jobs, services, cultural, and leisure spaces – and public transport in a direct way.

• **Public space permeability**, since visual and physical barriers can make it difficult for pedestrians and bike mobility.
- **Attractive views and activities**, which is attractive and interesting and offers amenities that support active travel along the walking and biking routes, such as public toilets, benches, and secure cycle parking and storage. This can encourage the use of public spaces by making them more pleasant and enjoyable.

- **Inclusive design**, with routes inclusive for all ages and abilities, including step free accesses, dropped kerbs, tactile paving, trip-free surfaces, and sufficiently wide to support different activities and mobility levels.

- **Good maintenance and upkeep**, which can aid in nurturing a sense of belonging and encourage the use of public spaces

- **Wayfinding strategies** include adding signage with route information, distances, and time to get to key destinations.

- **Effective lightning**, to increase safety and feeling of safety.

- **Natural surveillance**, to increase safety and feeling of safety.

- **Green Infrastructure and Biodiversity** can help in the ambience and safety of cycle and walk routes by acting as a sound and air pollution barrier, by providing shade, by improving the visual aesthetics, and by supporting interest through eye-catching flowers, leaf's colours, and wildlife (e.g., birds).

- **Road safety strategies**, which include mainly traffic calming measures – such as reducing speed limits, adding speed bumps, removing the centre line from streets, narrowing traffic lanes, and sharing space – and safe crossing measures, such as raising the carriageway and providing good lighting at crossways.

- **Public transport offer**, which is related to guaranteeing people can easily access effective and reliable public transport modes. This includes the quality and accessibility of public transport stations and of public transport services, making sure lines are frequent, reliable, and direct enough to provide a competitive alternative to car use and that is easy to change between different services modes, including cycling.

- **Traffic reduction strategies** include charges for private car use in specific zones or times, reducing car parking provision; implementation of car clubs; and regular street closures to motorised traffic in various times of day, or on special days, such as weekends.

The topic is related to the improvement or reduction of the following health variables: physical activity, type 2 diabetes, heart diseases, stroke, cancer, obesity, premature death, mental health, antisocial behaviour, road-related injuries, and deaths.

The group of people more likely to be affected by the underplay of active travel are disabled people and their carer-givers children, older people, and people living in
deprived communities, usually more dependent on walking and public transport for daily travel.

2.2.4 Multifunctional open spaces

The multifunctional open spaces topic is related to living near quality green and public open spaces. Seven physical environment variables were identified in this topic:

- **Public open space access and availability**, this include living, working, and studying around a short walking distance of an open space of any size; and ensuring the network of open spaces is an integral part of the active travel network, connected to walking, biking, and public transport routes.

- **Green Infrastructure and Biodiversity** should be included or preserved wherever possible in the design of urban open space to improve aesthetic and interest, help developing a sense of place and ownership, reduce air pollution, reduce flood risk, cool areas, store carbon, prevent soil erosion, improve the soundscape, give food growing opportunities, and attract investment. Green infrastructure includes natural and semi-natural areas and features, such as parks, gardens, playing fields, rivers, wetlands, lakes, woodlands, trees and vegetation, green roofs and walls, and sustainable drainage systems (SuDS).

- **Multifunctionality**, which includes designing for a range of play, sport, leisure, and cultural activities – such as running, sitting, chatting, eating, playing, and sports. This allows diverse groups of people to use the same spaces, contributing to inter-generational mixing and potentially reduces isolation.

- **Attractive views and activities** can make open spaces attractive and inviting.

- **Good maintenance and upkeep** can make open spaces attractive and inviting.

- **Inclusive design**, for all ages and abilities, which allows diverse groups of people to use the same spaces.

- **Effective lightning** to increase safety and feeling of safety.

The topic is related to the improvement or reduction of the following health variables: life expectancy, social connection, physical activity, stress, obesity, mental health, physical health, cardiovascular disease, mortality, anxiety.

Regarding the socioeconomic groups, availability of multifunctional open spaces has shown to reduce health inequalities between people living in the wealthiest and the poorest areas. Moreover, for children, play is vital for the enjoyment of childhood, as well as social, emotional, intellectual, and physical development.
2.2.5 Environmental Exposures

The environmental exposures topic includes a range of complex mixtures of chemical, biological, or physical substances found in air, water, food, or soil that may have a harmful effect on a person's health. The reports highlighted the following environmental exposures as especially detrimental in cities:

- **Air pollution**, which is considered the most pressing environmental threat to health. Two air pollutants are of specific concern in cities: particulate matter (PM10, PM2.5 and black carbon) and nitrogen dioxide (NO2).

- **Noise pollution**, which leads to adverse effects in health.

- **Water pollution**, which includes pollution of potable water sources and of other water bodies that can liberate bad odours and affect the interaction of the population with open spaces near rivers and streams.

- **Heat stress**, which can make homes, workplaces, and public transport uncomfortable and stressful for all and dangerous for the most at risk, and increase demand for cooling, which, in time, can increase Greenhouse Gases (GHG) emissions.

- **Flooding**, which can result in injury, pollution of soils and water bodies, and cause damage to public and private infrastructure.

- **Drought**, which can impact public water supply, agriculture, energy generation and industry.

Heat stress, flooding, and drought are at risk of worsening due to climate change, which also make climate change potentially a significant threat to public health. The physical variables that are characteristics of the urban structure related with the above-mentioned environmental factors are:

- **Control of construction and demolition sites**, which can lead to high volumes of dust and emissions from heavy machinery (PM, NO2, and GHG) and disturbance due to noise and vibration. Mechanisms should be put in place to control hours of construction, vehicle movements and pollution.

- **Active travel environment strategies** to reduce road traffic, which is one of the main causes for air pollution (PM and NO2), noise pollution and GHG emissions levels in cities. Active travel strategies encourage the use of public transport, cycling and walking, and discourage private car use. Moreover, reducing the space occupied by cars can give opportunities for more GI and SuDS, which can mitigate the impacts of a diversity of environmental exposures.

- **Rail traffic, commercial and industrial uses** are also responsible for noise pollution in cities, although these are much more spatially concentrated than
traffic noise. The impacts can be lessened by land use separation, controlling operation hours, orientation of residential units, and acoustic design.

- **Clean energy production and energy efficiency** have a significant impact on air pollution and GHG emissions. Strategies related to this include phase out fossil fuels in the transport system, heating, and other sources (e.g., large scale generators); increase energy efficiency of buildings and appliances; and increase renewable energy sources. Moreover, energy efficient buildings can also contribute to less overheating in summer, reducing the impacts of heatwaves.

- **Water and sanitation infrastructure**, an integrated infrastructure of potable water, sewerage, drainage, waste management, and flood defences can reduce the risks of contamination and flooding and mitigate air pollution, GHG emissions, and odour annoyance.

- **Green Infrastructure and Biodiversity** can mitigate the impacts of air pollution, noise pollution, water pollution, heat stress, flooding and drought. GI can improve air quality by PM deposition, reduce flood risk with the application of SuDS and conservation of wetlands and flood areas, cool areas during heat waves with shading and transpiration, store carbon, prevent soil erosion, and improve the soundscape of cities.

The discussed environmental exposures can impact the following health variables: physical health, mental health, neonatal complications, life-shortening chronic lung and heart conditions, stroke, asthma, cancer, diabetes, dementia, headaches, sleep disturbance, psycho-physiological problems, anxiety, hearing impairment, reduced performance and learning, heat stroke.

These health impacts fall disproportionately on the most disadvantaged communities, affecting the poorest, the youngest, the oldest, those with pre-existing health conditions and those from minority ethnic groups the most. Children are especially vulnerable, since there is evidence air pollution can limit lung development during childhood. Due to this, it is important to reduce exposure where the particularly vulnerable live, work, or study, such as schools, nurseries, other educational establishments, care homes, and hospitals.

### 2.2.6 Social, work, and housing infrastructure

The social, work, and housing infrastructure topic is related to three physical variables: the adequate provision of social infrastructure, jobs infrastructure, and housing infrastructure.

- **social infrastructure** includes health services, social care services, education and sports services, and early life support services (e.g., day care). Access to social infrastructure can directly impact health, increase employability and
earning capability, and divert those at risk of involvement in crime. Social Infrastructure should be supplied locally and be accessible to all, in terms of transport and access into a building. It is also important to co-locate services to improve the effectiveness of service delivery.

- **Jobs infrastructure**, employment and income is a key determinant of good health and wellbeing. Adequate provision of jobs infrastructure includes allocation of appropriate sites (e.g., affordable workspaces) and infrastructure provision (e.g., sites are easy to get to by public transport, walking and cycling networks) to facilitate attractive opportunities for businesses and ensure local jobs retention.

- **Housing infrastructure**, access to adequate and affordable diverse forms and types of housing are also related with health outcomes and increase employment and engagement with healthcare services.

This topic is related to the improvement or reduction of the following health variables: physical health, mental health, health inequalities, social cohesion, suicide, alcohol and drug misuse, physical activity, wellbeing.

People with mental health disorders and health disabilities are especially vulnerable to the lack of social, work, and housing infrastructure, since they are less likely to get the health and social treatment, and more likely to be unemployed and living in inadequate conditions. Also, the most deprived communities, children and the old are particularly vulnerable to the lack of social and work infrastructure.

### 2.2.7 Land Use

The **land use** topic is related to two urban physical variables:

- **Compactness and density**, which is related with higher street connectivity and higher residential densities.
- **Mixed-use design**, which is related with a diverse combination of commercial, residential, cultural, leisure, service, and industry land use.

This topic is a shared strategy among almost all the other 6 topics, with the exemption of open spaces. A compact and mixed-use design can improve the **active travel environment** topic, by increasing chances of people walking, biking, and using public transport, since key places, services and facilities are more conveniently closer together. This, in turn, can impact the **environmental exposures** topic, since the reduction of road traffic and used road spaces can reduce air and noise pollution and heat stress.

A compact and mixed-use design can also improve the **healthy food environment**, since mixed-use can bring a better mix of food ‘spaces. It is also associated with better social, jobs and housing infrastructure, since a compact and mixed-use design can
improve local economies, provide more local employment opportunities, and facilitate access to social infrastructure. And, finally, the topic is also related with a crime prevention environment, since mixed-use design, incorporating residential and commercial buildings, has the potential to provide day and night natural surveillance.

Since the land use urban physical variables impact all these topics, it also indirectly impacts the health variables and socioeconomic groups of these topics.
3. Urban Health Mapping

3.1 Case studies

The four pilot cities are London, Lisbon, Copenhagen, and Lansing. These four cities have different sizes in terms of urban population (see Figure 4 for cases in Europe). According to the Eurostat City Statistics (2021), there are over 8.8 million people in Greater London, 1.8 million people in Lisbon and 0.5 million people in Copenhagen. The population of Lansing in 2019 is 117,808.

![Image](image_url)

Figure 4 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.

---


9 The survey of urban health datasets identification and data assemblage was sent on 15 April 2020 via email to all WPs.

GA 945307

Deliverable 4.1 | Metrics and Methods for Urban Health

December 2021
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*\(^{10}\)
  - *Word Bank*\(^{11}\)

- **Europe Level**
  - *Eurostat*\(^{12}\)

- **City Level**
  - *London Database* by Greater London Authority\(^{13}\) and *Ordnance Survey*\(^{14}\)
  - 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

---

\(^{10}\) https://www.who.int/data/gho
\(^{11}\) https://data.worldbank.org/
\(^{12}\) https://ec.europa.eu/eurostat/data/database
\(^{13}\) https://data.london.gov.uk/
\(^{14}\) https://www.ordnancesurvey.co.uk/

Deliverable 4.1 | Metrics and Methods for Urban Health
December 2021
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.

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. 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 'mental map' of cities. 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 different types of street view images. Moreover, studies also examined the relationship between the qualified urban perception with happiness, safety, and so on. One of the most common practices is to extract sentiment and emotions from social media data such as Twitter (Gibbons et al., 2019; Jaidka et al., 2020). 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.


Deliverable 4.1 | Metrics and Methods for Urban Health
December 2021
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. 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 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 5 Framework of spatial analysis of urban health](image)

**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 the specific 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.

Selected variables and metrics for spatial analysis

Figure 6 lists the selected variables and metrics that are included in the spatial analysis based on the availability of data sets\textsuperscript{21}. The list follows a Theme – Sub-themes – Variables/Metrics structure. For instance, in the theme of Urban Health 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.

As we included such number of variables and metrics, we currently choose the administrative district level (e.g., boroughs in Greater London) as the unit for spatial analysis. For instance, the London boroughs have populations of around 150k to 300k, whereas those in Inner London are smaller in both population and area.

\textsuperscript{21} The list has been shared and discussed with PIs.

GA 945307

Deliverable 4.1 | Metrics and Methods for Urban Health

December 2021
**Figure 6** Measures lists for spatial analysis with urban health-related data. Variables and metrics in bold are chosen for spatial mapping at the current stage.

### 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 (Martí et al., 2019; Niu & Silva, 2020). 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 (Li et al., 2017; Internet Live Statistics, 2020). In the context of this project, geotagged social media data are used to extract urban perceptions of local urban places.

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. Recently, Twitter released the Academic Research product track that allows researchers to access the full archive of the database without exceeding the upper limit and rate limit (see Table 1). Two methods of data collection will be examined during the pilot study.

<table>
<thead>
<tr>
<th>Method</th>
<th>API Description</th>
<th>Sample ratio</th>
<th>Running</th>
<th>Upper limit</th>
<th>Rate limit</th>
<th>Monthly count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method 1</td>
<td>Twitter API - Standard v1.1 (Streaming)</td>
<td>Around 1%</td>
<td>Constantly running on PC/server</td>
<td>No (not mentioned in docs)</td>
<td>No limit for streaming</td>
<td>1.3 million without setting keywords (Greater London)</td>
</tr>
<tr>
<td>Method 2</td>
<td>Twitter API v2 - Academic Research track (Full-archive search)</td>
<td>Near 100%</td>
<td>Run monthly for several days</td>
<td>10 million per month</td>
<td>500 tweets per request; 300 requests per 15 mins</td>
<td>Depends on case</td>
</tr>
</tbody>
</table>

23 https://developer.twitter.com/en/products/twitter-api/academic-research
GA 945307

Deliverable 4.1 | Metrics and Methods for Urban Health
December 2021
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.

- 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 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.
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 7). The demographic inference was based on an open-source model – M3Inference – developed\(^{25}\). 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: \( \leq 18 \), \((18, 30)\), \([30, 40)\) 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.

### 4.4 Sentiment analysis

As a popular technique from the domain of natural language processing, sentiment analysis explores insights from social media data. By implementing sentiment analysis, 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 PIs with forming 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\(^ {26}\) will be used for emotion identification
- Hotspot analysis will be applied to identify the specific places for the following WP.

---


\(^{26}\) 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).

GA 945307

Deliverable 4.1 | Metrics and Methods for Urban Health

December 2021
5. Preliminary Results of urban health mapping and sentiment analysis

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

5.1 Examples of urban health mapping in London

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

- Central London - e.g., Camden, Islington, Lambeth, Southwark, Kensington and Chelsea, Westminster and the City of London - evidently have lower diabetes rates (Figure 8). Communities in Outer London especially in the western (e.g., Harrow, Brent, Ealing, and Hounslow), northern (e.g., Redbridge, Newham, and Barking and Dagenham) 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 9). The mortality rate from cardiovascular diseases is particularly high in the borough of Barking and Dagenham.

- Shown as Figure 10, boroughs of Hammersmith and Fulham, Kensington, Lambeth and Hackney have the higher ratio of active people (more than 30 mins physical activity a week).
Figure 8 Diabetes prevalence in London Boroughs (2020)

Figure 9 Mortality rate from all cardiovascular diseases in London Boroughs (2020)
Physical activity levels in London Boroughs, 2019
Ratio of active people (more than 30 minutes a week)

Source: Active Lives Adult Survey 2020

Figure 10 Physical activity levels in London Boroughs (2020)
5.2 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 11 Spatial distribution of geotagged tweets in Greater London
Figure 12 Spatial distribution of geotagged tweets in Lisbon
6. Conclusion

This document presents the first outcome of the work led by WP4. Throughout the document, the team identifies variables and metrics related to the urban physical and socioeconomic environment that most affect physical and mental health outcomes (preliminary results can be found in Annex 1 and Figure 6) and deploy the methodology of spatial analysis for urban health mapping and sentiment analysis for hotspot identification.

The outcome of this work helps subsequent stages as it will identify areas within the case study cities; capture baseline spectrum of cities’ (geographical) features helpful to reduce urban health inequalities; and will be relevant when considering confounding in further spatial multivariate analysis.
eMOTIONAL Cities
Mapping the cities through the senses of those who make them
<table>
<thead>
<tr>
<th>Name</th>
<th>Category</th>
<th>Category_1</th>
<th>Type</th>
<th>Is it Spatial?</th>
<th>Level</th>
<th>General description</th>
<th>Results interpretation</th>
<th>Necessary data</th>
<th>Expected comput time</th>
<th>Methodological Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcohol consumption Drugs consumption</td>
<td>Health</td>
<td>Alcohol and drugs consumption</td>
<td>Quantitative</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tabaco consumption</td>
<td>Health</td>
<td>Alcohol and drugs consumption</td>
<td>Quantitative</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiovascular diseases prevalence</td>
<td>Health</td>
<td>Cardio-metabolic Diseases</td>
<td>Quantitative</td>
<td>No</td>
<td>European City</td>
<td>Prevalence of disease</td>
<td>Statistical data</td>
<td></td>
<td></td>
<td>Influence of green spaces, on environmental satisfaction and physiological status of urban residents</td>
</tr>
<tr>
<td>Heart rate variability (HRV)</td>
<td>Health</td>
<td>Cardio-metabolic Diseases</td>
<td>Quantitative</td>
<td>No</td>
<td>Non european country - Local Scale</td>
<td>HRV is the variation in the time interval between consecutive heartbeats</td>
<td>Normally higher HRV are associated with relaxing and recovering activities and lower HRV with stress</td>
<td>Study Population measurements with a heart monitor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obesity and overweight prevalence</td>
<td>Health</td>
<td>Cardio-metabolic Diseases</td>
<td>Quantitative</td>
<td>No</td>
<td>European City</td>
<td>Prevalence of disease</td>
<td>Statistical data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type II diabetes prevalence</td>
<td>Health</td>
<td>Cardio-metabolic Diseases</td>
<td>Quantitative</td>
<td>No</td>
<td>European City</td>
<td>Prevalence of disease</td>
<td>Statistical data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type II diabetes comorbidities/difficulties</td>
<td>Health</td>
<td>Cormobidities</td>
<td>Qualitative</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Excess heat mortality</td>
<td>Health</td>
<td>Heat Stress</td>
<td>Qualitative</td>
<td>No</td>
<td>European City</td>
<td>Prevalence of disease</td>
<td>Statistical data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mental Health Disorders - Patient Health Questionnaire (PHQ)</td>
<td>Health</td>
<td>Mental Health</td>
<td>Quantitative</td>
<td>No</td>
<td>General</td>
<td></td>
<td>Study population answers to the Patient Health Questionnaire (PHQ)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Air Quality Management Zones (AQMZs)</td>
<td>Physical Environment</td>
<td>Air Quality</td>
<td>Quantitative</td>
<td>Yes</td>
<td>European City</td>
<td>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)</td>
<td>Modelled air quality maps</td>
<td>Geospatial datasets</td>
<td>A framework for Air Quality Management Zones - Useful GIS-based tool for urban planning</td>
<td></td>
</tr>
<tr>
<td>Topic</td>
<td>Environment</td>
<td>Quality Type</td>
<td>Scale</td>
<td>Methodology</td>
<td>Key Findings</td>
<td>Data Sources</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>-------------</td>
<td>--------------</td>
<td>------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public health impacts of urban air pollution</td>
<td>Physical</td>
<td>Air Quality</td>
<td>Quantitative Yes</td>
<td>European City</td>
<td>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.</td>
<td>geospatial datasets</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health-based assessment of particulate air pollution</td>
<td>Physical</td>
<td>Air Quality</td>
<td>Quantitative No</td>
<td>Non European country - Local Scale</td>
<td>Analysing mortality and morbidity effects of PM10 pollution based on statistical data and the epidemiological exposure–response function.</td>
<td>Exposure–response function statistical data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAQ – Random Forest for Predicting Air Quality</td>
<td>Physical</td>
<td>Air Quality</td>
<td>Quantitative Yes</td>
<td>Non European country - Local Scale</td>
<td>Random forest approach for predicting air quality (RAQ) is proposed for urban sensing systems.</td>
<td>The data generated by urban sensing includes meteorology data, road information, real-time traffic status and point of interest (POI) distribution.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Air Quality Indicators</td>
<td>Physical</td>
<td>Air Quality</td>
<td>Quantitative Yes</td>
<td>Non European country - Local Scale</td>
<td>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.</td>
<td>Spatial Variability and Properties of Aerosol over the Selected area. Multiple Satellites and AERONET Data.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multifunctional Green Infrastructure (GI)</td>
<td>Physical</td>
<td>Green and blue space</td>
<td>Quantitative Yes</td>
<td>European City</td>
<td>Study enclosed 18 indicators, as well as identified hot and cold spots of selected GI functions and their multifunctionality.</td>
<td>Spatial distributions of hot/cold spots of GI functions. LULC from urban atlas Indicator framework from several source.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health effects of the natural outdoor environment</td>
<td>Physical</td>
<td>Green and blue space</td>
<td>Quantitative No</td>
<td>European City</td>
<td>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.</td>
<td>Landuse planning management Quantitative and quality characteristics. Positive health effects of the natural outdoor environment in typical populations in different regions in Europe (PHENOTYPE).</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Green space metrics and health outcomes</td>
<td>Physical Environment</td>
<td>Green and blue space</td>
<td>Quantitative</td>
<td>Yes</td>
<td>European City</td>
<td>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.</td>
<td>Data included those of 2 m spatial resolution from WorldView2, 5 m resolution from RapidEye and 30 m resolution from Landsat.</td>
<td>Differences in characterizing green space and health outcomes</td>
<td>hours</td>
<td>Associations of green space metrics with health and behavior outcomes at different buffer sizes and remote sensing sensor resolutions</td>
</tr>
</tbody>
</table>

| Green Spaces as an Indicator of urban health | Physical Environment | 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. | 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 Environment | 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 Environment | Green and blue space | Quantitative | Yes | Non European country - Local Scale | 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 | Validation of the Normalized Difference Vegetation Index as a Measure of Neighborhood Greenness |

<p>| Urban Green Infrastructure and Health Inequalities | Physical Environment | 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 |</p>
<table>
<thead>
<tr>
<th>Spatial Urban Health Equity Map and Validating via Volunteered Geographic Information</th>
<th>Physical Environment</th>
<th>Health Services</th>
<th>Quantitative</th>
<th>Yes</th>
<th>Non european country - Local Scale</th>
<th>Health equity map generated from reference data to one generated by volunteered geographic information (VGI) by citizens.</th>
<th>Spatial Pattern of the Characteristics of UGSs and the Morbidity of Diseases</th>
<th>National Health Insurance Research Database</th>
<th>Developing an Approach for Generating a Spatial Urban Health Equity Map, and Validating via Volunteered Geographic Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time-distance to health cares Heatwave early warning systems (HEWS)</td>
<td>Physical Environment</td>
<td>Health Services</td>
<td>Quantitative</td>
<td>Yes</td>
<td>European City</td>
<td>Heat related early warnings</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Land-Use and Land-Cover Change, Urban Heat Island, and Health</td>
<td>Physical Environment</td>
<td>Heat Stress</td>
<td>Quantitative</td>
<td>Yes</td>
<td>Non european country - Local Scale</td>
<td>LULC changes have forced the development of a significant urban heat island effect at both the urban canopy and urban boundary layers as well as an increase in ground level ozone production to such an extent that Atlanta has violated EPA's ozone level standard in recent years. Using canonical were found to correlate strongly with volatile organic compounds (VOC) and nitrogen oxides (NOx) emissions, the two ingredients that form ozone by reacting with sunlight, but only weakly with the rates of cardiovascular and chronic lower respiratory diseases.</td>
<td>LULC - Unsupervised Iso data, Surface temperature extracted using quadratic regression model and vegetation using NDVI</td>
<td>Satellite images Landsat</td>
<td>Spatial Characteristics of Urban Green Spaces and Human Health: An Exploratory Analysis of Canonical Correlation</td>
</tr>
<tr>
<td>Local Climate Zones (LCZ)</td>
<td>Physical Environment</td>
<td>Heat Stress</td>
<td>Quantitative</td>
<td>Yes</td>
<td>European City</td>
<td>Local Climate Zones</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thermal infrared remote sensing of urban heat</td>
<td>Physical Environment</td>
<td>Heat Stress</td>
<td>Quantitative</td>
<td>Yes</td>
<td>Non european country - Local Scale</td>
<td>Thermal imagery for identifying hotspots, however this relies on the assumption that patterns in land surface temperature (LST) coincide with patterns in air temperature</td>
<td>High LST do represent air temperature hotspots</td>
<td>Thermal Infrared images</td>
<td></td>
</tr>
<tr>
<td>Universal Thermal Climate Index (UTCI)</td>
<td>Physical Environment</td>
<td>Heat Stress</td>
<td>Quantitative</td>
<td>Yes</td>
<td>General</td>
<td>UTCI is the equivalent temperature for the environment quantifying how a human would physiologically react to a given set of environmental conditions based on a thermo-physiological modeling of human response to meteorological conditions including the acclimatization issue. The numbers between 9 to 26 represent the no thermal stress zone. any number below or above this threshold can be translated to cold stress or heat stress respectively.</td>
<td>Air temperature, Humidity, wind speed, solar radiation</td>
<td>hours</td>
<td>UTCI Universal Thermal Climate Index</td>
</tr>
<tr>
<td>---------------------------------------</td>
<td>----------------------</td>
<td>-------------</td>
<td>--------------</td>
<td>-----</td>
<td>---------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------</td>
<td>-------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>urban heat island and its impact on heat waves and human health</td>
<td>Physical Environment</td>
<td>Heat Stress</td>
<td>Quantitative</td>
<td>No</td>
<td>Non european country - Local Scale</td>
<td>With global warming forecast to continue into the foreseeable future, heat waves are very likely to increase in both frequency and intensity. In urban regions, these future heat waves will be exacerbated by the urban heat island effect, and will have the potential to negatively influence the health and welfare of urban residents.</td>
<td>The UHI intensity of each site (ΔTi) is calculated by the temperature difference between the urban site and each suburban or exurban site as follows: ΔTi = Tmax0 - Tmaxi. While Tmax0 is the daily maximum temperature at the urban site, Tmaxi is the daily maximum temperature at the suburban or exurban site.</td>
<td>Data from weather stations (Point)</td>
<td>Data from weather stations (Point)</td>
</tr>
<tr>
<td>Urban Heat Island: Implications for Health in a Changing Environment</td>
<td>Physical Environment</td>
<td>Heat Stress</td>
<td>Quantitative</td>
<td>No</td>
<td>World Wide</td>
<td>The most direct effect on health from the Urban Heat Island is due to heat risk, which is exacerbated in urban areas, particularly during heat waves.</td>
<td>Review</td>
<td>-</td>
<td>The Urban Heat Island: Implications for Health in a Changing Environment</td>
</tr>
<tr>
<td>Urban structure and its implication of heat stress</td>
<td>Physical Environment</td>
<td>Heat Stress</td>
<td>Quantitative</td>
<td>Yes</td>
<td>Non european country - Local Scale</td>
<td>Research provides an accurate and efficient method for urban structure identification (maximum likelihood classification) and determines its effect on heat stress. (ENVI-met was applied to simulate the microclimate in six idealized models to confirm the urban structure’s heat stress impacts)</td>
<td>The outputs of the simulation were air temperature, mean radiant temperature, solar radiation intensity, relative humidity, and wind speed simulation data</td>
<td>Satellite images Landsat 8</td>
<td>Satellite images Landsat 8</td>
</tr>
<tr>
<td>Assessing urban soundscape</td>
<td>Physical Environment</td>
<td>Noise Pollution</td>
<td>Quantitative</td>
<td>Yes</td>
<td>European City</td>
<td>Environmental noise maps</td>
<td>Environmental noise maps</td>
<td>geospatial datasets</td>
<td>Environmental noise maps</td>
</tr>
<tr>
<td>Ageing</td>
<td>Socioeconomic</td>
<td>Age</td>
<td>Quantitative</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>People 65+ living alone</td>
<td>Socioeconomic</td>
<td>Age</td>
<td>Qualitative</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>