

WalkingStreet: understanding human mobility phenomena through a mobile application

Luís Rosa¹, Fábio Silva², and Cesar Analide¹

¹ University of Minho, ALGORITMI Center, Dep. of Informatics, Braga, Portugal,
id8123@alunos.uminho.pt, analide@di.uminho.pt

² CIICESI, ESTG, Politécnico do Porto, Felgueiras, Portugal,
fas@estg.ipp.pt

Abstract. Understanding human mobility patterns requires access to timely and reliable data for an adequate policy response. This data can come from several sources, such as mobile devices. Additionally, the wide availability of communications networks enables applications (mobile apps) to generate data anytime and anywhere thanks to their general adoption by individuals. Although data is generated from personal devices, if a relevant set of metrics is applied to it, it can become useful for the authorities and the community as a whole. This paper explores new methods for gathering and analyzing location-based data using a mobile application called WalkingStreet. The article also illustrates the great potential of human mobility metrics for moving spatial measures beyond census units, key measures of individual, collective mobility and a mix of the two, investigating a range of important social phenomena, the heterogeneity of activity spaces and the dynamic nature of spatial segregation.

Keywords: Geo-location Data, Human Mobility Metrics, Mobile Application, Smart Cities.

1 Introduction

Over the past decade smart cities have been enhanced with the introduction of new technological devices, increasing the data capture points generated from interaction with citizens. Along with this ubiquity of devices, there is growing concern about the impact human mobility phenomena have on the sustainability of cities [12]. Additionally, both academia and industry have introduced a term and definition related to human mobility in cities by Information and Communication Technologies (ICTs) which is smart human mobility [15]. Therefore, it is not surprising that the study of human mobility has contributed to solve the problems of smart cities.

Human mobility prototypes are proposed in the literature to measure the moving spatial beyond census units [11], investigate the heterogeneity of activity spaces [1], to predict movement patterns based on personality or spatio-temporal routines [16], to provide real time information of contrasting social and non-social

sources of predictability in human mobility [2], to calculate the travel time of any trip [17], and to monitor human mobility during big events [18]. These pilot studies illustrate the great potential of mobile phone methodology and the role of mobile-based innovation in phenomena related to people’s movements.

To promote more human mobility analysis, current strategies have used metrics to measure and discover daily mobility patterns taking advantage of mobile devices data such as smartphones. [14] uses a non-parametric Bayesian modeling method such as Infinite Gaussian Mixture Model (IGMM) to estimate the probability density and Kullback-Leibler (KL) divergence as the metric to measure the similarity of different probability distributions of daily mobility. In [20], the authors re-examined human mobility patterns via cell-phone position data recorded and considered four metrics to quantify the trajectories of individuals. However, this work revealed that Markov process can quantitatively reproduce the observed travel patterns at both individual and population levels at all relevant time-scales. Furthermore, [9, 10, 19] analyze data-driven human mobility metrics and their correlations with the transmissibility of COVID-19 using mobility data collected from mobile phone users. Thus, thanks to the availability of timely data and choosing appropriate metrics, mobile platforms can be developed for near-real-time mapped and predicted human mobility to help policy makers and the public (e.g, companies, communities and cities).

Inspired by existing metrics and mobile solutions, we aim to build a mobile application called WalkingStreet App. This novel human mobility app enables a visual interpretation of the phenomena related to people’s movements. It incorporates (1) different strategies for human mobility analysis that use a potentiality wide range of the Google Maps Android API like clustering— handles the display of a large number of points, heat maps — displays a large number of points as a heat map, poly encoding and decoding — compact encoding for paths and spherical geometry (i.e, distance, heading and area); (2) several mobility metrics and patterns from data, both at the individual and collective level (i.e, length of displacements and typical distance); (3) synthetic individual trajectories using standard mathematical models such as random walk models; (4) assessing the privacy risk associated with a mobility datasets, and (5) different geographical formats like geoshape (polygon), or geotrace (polyline), that can be associated with other types of information. However, our mobile application does not consider prediction algorithms for now, since that it requires access to a huge amount of human mobility data. Moreover, this work is a prototype mobile device application designed to use georeferenced datasets with different formats such as Keyhole Markup Language (KML) and GeoJSON data supported by an API infrastructure in background based on sensor data.

With the WalkingStreet App, we aim to investigate the geographical regularity and variability of human mobility in local public amenities; to quantify the extent of spatial and temporal regularity and variability in these places; to explore geometry mapping features for georeferencing human mobility trajectories; to measure the availability and accessibility of local urban amenities or Point-Of-Interests (POIs); and, to assess the correlation between human mobil-

ity and sociodemographic characteristics of Barcelos population, a city located in the North of Portugal. Normally, internet-based travel diary instruments or traditional survey approaches are used to solve these tasks. However, we draw a prototype to estimate human mobility based on GPS traces of two simple datasets (i.e, KML and GeoJSON data). This app also is a simpler deployment and scalable alternative. Moreover, it provides a suite for offline mapping and visualization of geo-referenced data from a soft data mining process and support for a basic infrastructure. These may be the main differences between this and other existing projects [3, 6].

The rest of this paper presents the WalkingStreet prototype and is organized as follows. Section 2 describes the geo-referenced datasets and human mobility metrics employed to analyze the relationship between local amenities and mobility flows. We also discuss our prototype. Section 3 presents the details of the graphical view interface incorporated in the prototype. In Section 4, we present and discuss the results and their implications for the prototype. Lastly, we summarize the key insights from our analysis and discuss potential avenues for future work in Section 5.

2 WalkingStreet Prototype

Based on the potential of georeferenced data, we define Barcelos City as our area of study, assigning a coordinate system on two datasets. We have also chosen a set of metrics to estimate human mobility. Finally, we present the architecture of a novel open mobile mapping tool that allow views, queries, and analyzes proposed datasets.

2.1 Geo-referenced Datasets

A pair of datasets is used to verify and validate the human mobility data and include geo-referenced information, i.e. placename and geospatial referencing. This data is collected based on innovative capture methods such as MyGeodata Converter, a converter for geospatial data [7]. Although this makes it difficult to analyze results obtained, since the idea is to propose a scope/prototype, for the time being without commercial purposes and they are easily delivered on the Internet and viewed in a free application, we believe that the innovative data capture model comparing with another modern sources is more adequate to achieve the objectives of this article.

In this work we use a KML dataset, an official Open Geospatial Consortium (OGC) standard. It is an XML-based format for storing geographic, associated content of several POIs of Barcelos City such as Tourist Spots, or user-specific places such Commercial Points. KML is a common format for sharing geographic data with non-Geographic Information System (GIS) users. Regarding KML elements, this dataset is composed of feature and raster elements including points, lines, polygons, and pictures. Whereas this kind of dataset is typically seen as separate and homogeneous elements (for example, point feature classes can only

contain points, rasters can only contain cells or pixels and not features), a single KML file can contain features of different types.

The second dataset includes Geodata data package, providing GeoJSON polygons of the Barcelos City. It defines several types of JSON objects and combines them to represent data about geographic features, their properties, and their spatial extents [8]. Finally, GeoJSON consists of the different parts (1) Geometry object is the location information; (2) Feature object is a geometry object which associates random ad hoc data and does not take into consideration what data is associated with the location information; and (3) FeatureCollection is a list of feature objects. Therefore, GeoJSON dataset typically consists of a FeatureCollection containing a list of data. In this work we use GeoJSON as a means to collect and share location data acquired from mobile devices. The structure of polygons and data collected will further developed in future work to account for data privacy as the objective is to determine flows of movement and not every single person's movement.

2.2 Human Mobility Metrics

We leverage a set of metrics to analyze the patterns of human mobility in the two datasets. One of them is trajectory preprocessing method. As any analytical process, mobility data analysis requires validation metrics. Thus, the trajectory preprocessing step allows the user to perform two metrics: stop detection and stop clustering. In these metrics, some points in a trajectory are called Stay Points or Stop Points. To detect them we apply spatial clustering algorithms to cluster trajectory points by looking at their spatial proximity. In other words, based on heatmaps of WalkingStreet app we find these visited points by a moving object. For example, we can identify the stops where the object spent at least minutes within a distance spatial radius, from a given point. In this work, our mobile application merges all the points that are closer than $0.5km$ from each other.

Another metric is origin-destination matrix. It represents the flow of objects between two locations based in specific column names and data types such as Origin, Destination and Flow. In human mobility tasks, the territory is covered by the bi-dimensional space using a countable number of geometric shapes (e.g., squares, hexagons) with no overlaps and no gaps. For instance, for the analysis of human mobility flows, we aggregate flows of people moving among locations.

Other metrics are useful to capture the patterns of human mobility at individual and collective levels. Individual measures summarize the mobility patterns of a single moving object, while collective measures summarize mobility patterns of a population as a whole. The WalkingStreet app provides individual synthetic trajectories corresponding to a single moving object, assuming that an object is independent of the others. It provides the most data volume observed during each time. With collective trajectories we can estimate spatial flows between a set of discrete locations and compute thematic maps of the territory. For example, we identify trips between neighborhoods, migration flows between municipalities

or freight shipments between states. Therefore, through these metrics, we can obtain several important human mobility patterns [4,5].

2.3 Design and Architecture

In this paper the WalkingStreet architecture is inspired in Layered Architecture. This approach combines the ideas of several other architectural approaches where (1) to be tested and audited; (2) not depend on UI; and (3) not depend on the database, external frameworks and libraries [13]. In addition, each layer is also independent while still being able to transmit information and data. Thus, our architecture (as illustrated in Fig. 1) is composed of three components: Presentation Layer, Domain Layer and Data Layer.

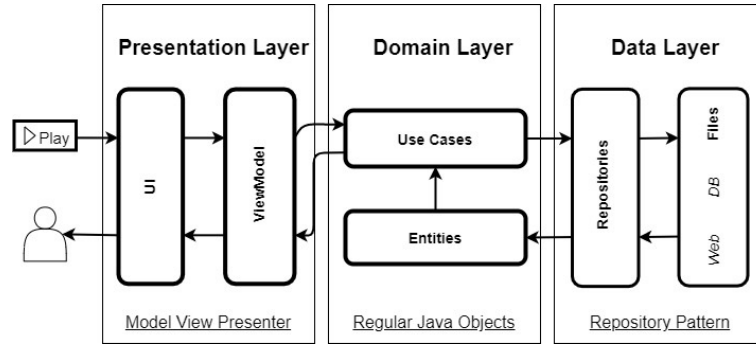


Fig. 1: WalkingStreet application Architecture.

The user event goes to the Presenter. In this layer, the user selects a metric and, then, it passes to the Use Case. Use Case of Domain layer makes a request to the Repository, located in the Data layer. Repository gets the data stored in GeoJSON and KML datasets, creates an Entity, passes it to Use Case. The Use Case gets all the Entity needs, being the place for all the domain processes and operations. Then, applying them and his logic, he gets the result, and passes it back to Presenter. And that, in turn, displays the result in the User Interface (UI). Finally, the user visualizes the map of selected metric with several points.

As we mentioned, since we are presenting a prototype, we do not consider it relevant to propose a concrete data source. In other words, in this phase of our project the data source level is not important and do not use any database located on a machine or the API of a service available on the internet, all available data is stored by GeoJSON and KML files.

3 Implementation and Graphical View

In this section, we present methods for the exploration of human movement patterns using the proposed tracking formats (i.e, KML and GeoJSON). But,

first we define the user flow of mobile app interface. This flow is the path that the user follows to access the several available maps, performed by selected metric (e.g, the user can see the map associated with the clustering metric). Then, we present some metrics that can be analysed based on mobility flow maps in the WalkingStreet prototype. These maps are used to extract flows from proposed datasets, rendering trajectory lines.

3.1 Mobility Presentation

The WalkingStreet app design has a focus on a user-friendly flow, promoting easy browsing through its contents. The process of designing the user flow also helped to prioritize content requirements, in terms of human mobility metrics that can be accomplished and put in the right place in order to achieve this in the most efficient way possible. Additionally, we avoid potential barriers in the navigation flow and implement quick routes to complete the intended actions.

Fig. 2 explains the process of a user getting the map that applies spatial clustering algorithms to cluster trajectory points by looking at their spatial proximity. At each step in the flow is shown the wireframe available to users.

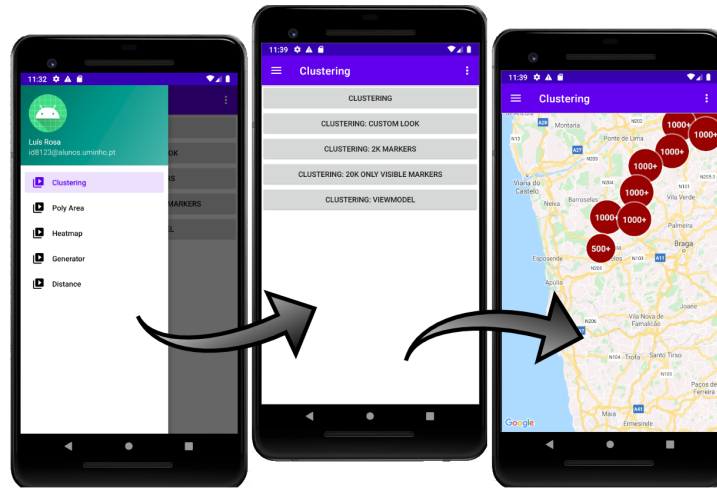


Fig. 2: Flow of the mobile application interface.

In this example, when a user selects an available metric in left menu the wireframe corresponds to the same app page, rather than representing different app pages. Each step clearly indicates the hotspots that connect to the next step in the task flow. An arrow is used to indicate the specific UI component where the user takes action such as a tap on a button and selecting a metric, and points to another wireframe image of what happens as a result of the interaction, i.e, a list of several map options associated with the selected human mobility metric.

The second "node" of that interaction shows the same page with the result of that interaction, e.g., display a large number of points as a heat map. Thus, the arrows clearly indicates the clickable "hotspots" that lead to the next step in the flow, in order to decrease the ambiguity in the wireflow.

3.2 Analytic Metrics

Understanding human mobility and how it manifests across temporal and spatial scales can be conceptually possible by implementing a set of metrics using GoogleMaps API. This section shows how our prototype drives important metrics to analyze the patterns of human mobility.

Trajectory Extraction This flow extraction is based on the georeference data and determine flow of the human mobility in specific time. In Fig. 3, the resulting flows can be visualized, for example, to explore the popularity of different paths of movement:

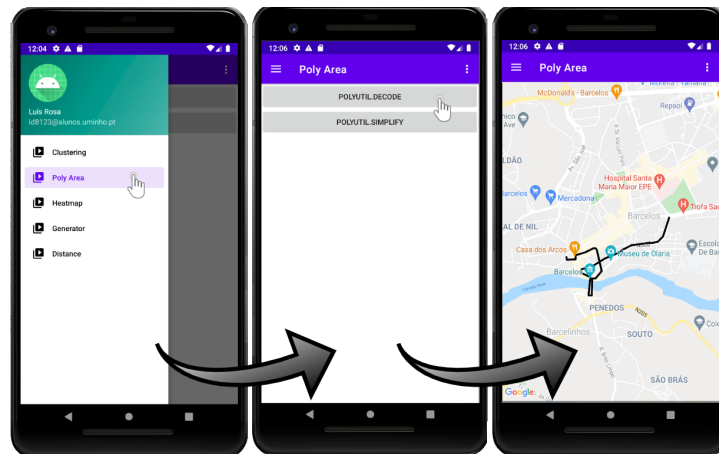


Fig. 3: Flow of extracting trajectory.

After the prototype has been computed, the flow algorithm computes transitions between points. An individual moving from point A to point B triggers an update of the corresponding flow. Additionally, the raw human mobility data records need to be converted into trajectories. Afterwards, each trajectory is processed independently ensuring (1) the distance to the match is below the distance threshold; (2) ensuring (2) the flow is retrieved or created in the prototype; and (3) and flow direction matches the record's direction.

This approach scales two datasets where the flow results and the trajectory have to be kept in memory for each iteration. However, this algorithm does not allow for continuous updates. Flows would have to be recomputed (at least

locally) whenever datasets changed. Therefore, the algorithm does not support the exploration of continuous data streams.

Generating trajectories This method explores origin-destination relationships in several trajectories. The GeoJSON and KML databases identify individual trips with their start/end locations and trajectories between them. Moreover, trip trajectories were generated by consecutive records into continuous tracks and then splitting them at stops. For example, in our paper, we extracted human mobility paths which meant that we also had to account for observation gaps when individuals had no access to wi-fi, no longer had wireless network access due to lack of mobile data or weak network. Fig. 4 explores trajectories for two users.

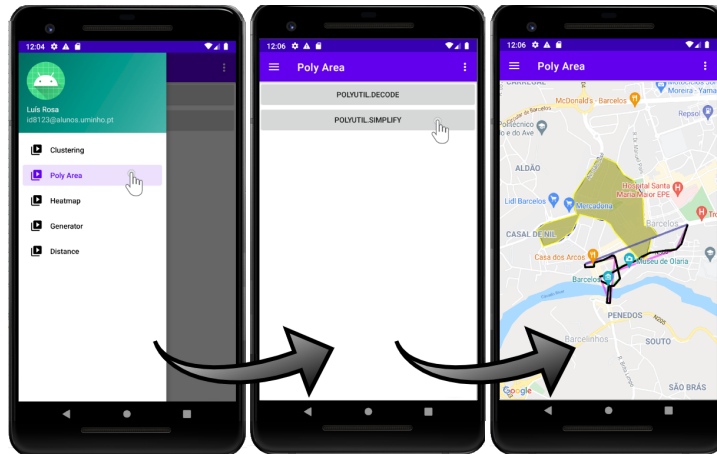


Fig. 4: Flow of generating trajectory.

Like extracting trajectory metric, the trajectory aggregation approach uses human trajectory data and the fact that operations only produce correct results if applied to a complete and chronologically sorted set of location records. This means that an aggregator needs to collect and sort the entire track on the map. Although the volume of datasets used in this prototype is reduced and we still don't have to worry about out-of-memory errors potentially being frequently encountered, it will be a challenge when the WalkingStreet app deals with large datasets.

Human mobility data aggregation Visualizations of point density maps called heat maps provide data exploration capabilities for reduced datasets. Although be limited in volume of data comparing with existing aggregation approaches to large datasets, these aggregation approaches can reveal movement patterns.

Using gespatial tools such as HeatmapTileProvider, it is quite straightforward to query timestamped location records. Thanks to the integration of HeatmapTileProvider with Google Maps for Android it is also possible to redraw the tiles with the available options such as Radius, Gradient and Opacity which can be equally visualized and explored. Although these setups get point maps and point density maps using an aggregation method, other important movement characteristics like speed and direction are not included.

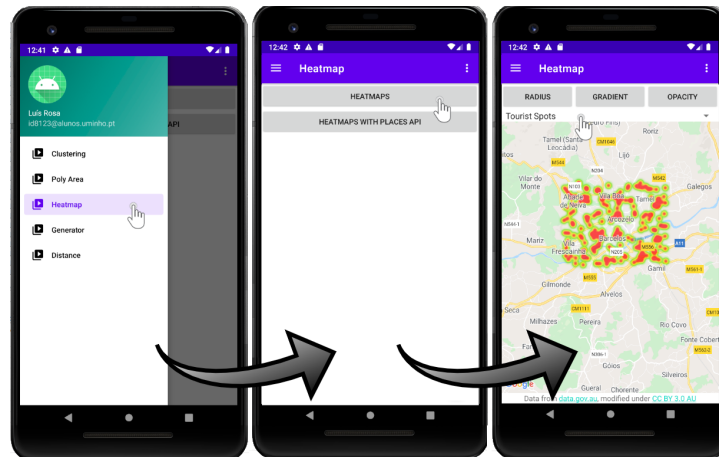


Fig. 5: Flow of the aggregating human mobility data.

Fig. 5 shows how our application explores human movement data. It gives a first impression of the spatial distribution of records. The real value becomes clearer when we zoom in and start exploring local patterns. Then we can discover more details about movements. Even though the points we use are rather distributed, the densities are clearly formed in some locations. We can see exactly where these densities are and human mobility there, without having to increase the grid resolution to impractical values. The marker size shows the concentration of records on the location and thus helps distinguish heavily traveled zones from minor ones.

4 Results and Discussion

As shown in this paper, to explore human mobility datasets we needed to aggregate the data. These aggregations helped to discover patterns and visual exploration of human movements. But, applied in large datasets can be computationally expensive and therefore slow to generate. On the other hand, density maps are readily available and quick to compute but they provide only very limited insight. However, these metrics are a starting point for a new approach

to exploring human mobility. Using raw location records, different forms of aggregation can be useful to learn more about GeoJSON and KML datasets:

1. aggregating raw location records to summarize human movements;
2. connecting consecutive records into continuous tracks to generate trajectories;
3. representing trajectory clusters by extracting flows.

Besides clever aggregation approaches, the human mobility datasets used in this article also require appropriate computing resources. To ensure that we can efficiently explore these datasets, we have implemented the aggregation steps in the WalkingStreet app. This enables us to run the computations on general-purpose computing clusters that can be scaled according to the dataset size.

However, during the development of WalkingStreet application problems were identified and some common statements were considered. In particular, human mobility data is sensitive since the movements of individuals can reveal confidential personal information, creating serious privacy risks. Or malicious actors can get to know a certain number of locations visited by an individual, but they do not know the temporal order of the visits. In addition, extracting trajectories from large datasets can also be challenging, particularly if the records of individual moving objects don't fit into memory anymore and if the spatial and temporal extent varies widely.

5 Conclusions and future work

This project has showcased how a human mobility app works/could work a human mobility application, using Google Maps API, GeoJSON, KML, and API in Python. The app helps to review the information about a zone and calculate a trajectory between points on maps from georeference previously defined in proposed datasets. As these datasets contain multiple trajectories from few users, the preprocessing methods automatically apply to the single trajectory and, when necessary, to the collective moving. In order to promote the interaction of the user with maps, different metric types can be selected, a few standard Google Maps controls are added, such as zoom controls; map scale control; and tooltips (e.g. place name). Additionally, the applicability of a set of metrics to facilitate their understanding until decision-making by the authorities should be as narrow as possible. Finally, choosing Clean Architecture is a good solution to this problem because it is not tied to a specific framework or database. In addition, each layer is also independent while still being able to transmit information and data.

In future work, we plan to introduce more users/individuals records in our research, investigating urban community patterns. Our aim is to avoid traditional methods to collect data and use self-reported data method. Self-reported data can provide additional value compared to traditional data since this data might be more spatially accurate, not outdated and with a frequent sampling time to make comparisons. Thus, in the next release, we will provide a way for

users to send geo-referenced data via the proposed application. Predicting human mobility is another crucial component of urban planning and management. We will predict the individual or collective behaviors over time based on the person's past trajectory and the geographical features of the area. The effectiveness of the prediction process (e.g, in terms of points of interest and trip distance) will be possible using a massive mobile phone location dataset along with an improvement in our API features in Python.

Acknowledgments

This work has been supported by FCT - Fundacao para a Ciencia e Tecnologia within the R&D Units Project Scope: UIDB/00319/2020. It has also been supported by national funds through FCT – Fundação para a Ciência e Tecnologia through project UIDB/04728/2020.

References

1. Bu, J., Yin, J., Yu, Y., Zhan, Y.: Identifying the daily activity spaces of older adults living in a high-density urban area: A study using the smartphone-based global positioning system trajectory in shanghai. *Sustainability (Switzerland)* **13**(9), 5003 (may 2021). <https://doi.org/10.3390/su13095003>
2. Chen, Z., Keltly, S., Welles, B.F., Bagrow, J.P., Menezes, R., Ghoshal, G.: Contrast-ing social and non-social sources of predictability in human mobility (apr 2021), <http://arxiv.org/abs/2104.13282>
3. Clouse, K., Phillips, T.K., Mogoba, P., Ndlovu, L., Bassett, J., Myer, L.: Attitudes toward a proposed gps-based location tracking smartphone app for improving engagement in HIV care among pregnant and postpartum women in South Africa: Focus group and interview study. *JMIR Formative Research* **5**(2), e19243 (feb 2021). <https://doi.org/10.2196/19243>
4. van Duynhoven, A., Dragičević, S.: Analyzing the effects of temporal res-olution and classification confidence for modeling land cover change with long short-term memory networks. *Remote Sensing* **11**(23) (dec 2019). <https://doi.org/10.3390/rs11232784>
5. Feng, X., Li, J.: Evaluation of the Spatial Pattern of the Resolution-Enhanced Thermal Data for Urban Area. *Journal of Sensors* **2020** (2020). <https://doi.org/10.1155/2020/3427321>
6. Frith, J., Saker, M.: It Is All About Location: Smartphones and Tracking the Spread of COVID-19. *Social Media and Society* **6**(3) (jul 2020). <https://doi.org/10.1177/2056305120948257>
7. GeoCzech, I.: MyGeodata Cloud - GIS Data Warehouse, Converter, Maps (2021), <https://mygeodata.cloud/>
8. IETF, I.E.T.F.: The GeoJSON Format (2016), <http://www.rfc-editor.org/info/rfc7946>
9. Kraemer, M.U., Bisanzio, D., Reiner, R.C., Zakar, R., Hawkins, J.B., Freifeld, C.C., Smith, D.L., Hay, S.I., Brownstein, J.S., Perkins, T.A.: Inferences about spatiotem-poral variation in dengue virus transmission are sensitive to assumptions about human mobility: a case study using geolocated tweets from Lahore, Pakistan. *EPJ Data Science* **7**(1) (dec 2018). <https://doi.org/10.1140/epjds/s13688-018-0144-x>

10. Kraemer, M.U., Yang, C.H., Gutierrez, B., Wu, C.H., Klein, B., Pigott, D.M., du Plessis, L., Faria, N.R., Li, R., Hanage, W.P., Brownstein, J.S., Layan, M., Vespignani, A., Tian, H., Dye, C., Pybus, O.G., Scarpino, S.V.: The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* **368**(6490), 493–497 (may 2020). <https://doi.org/10.1126/science.abb4218>
11. Liu, Y., Wang, Z., Rader, B., Li, B., Wu, C.H., Whittington, J.D., Zheng, P., Stenseth, N.C., Bjornstad, O.N., Brownstein, J.S., Tian, H.: Associations between changes in population mobility in response to the COVID-19 pandemic and socioeconomic factors at the city level in China and country level worldwide: a retrospective, observational study. *The Lancet Digital Health* **3**(6), e349–e359 (jun 2021). [https://doi.org/10.1016/s2589-7500\(21\)00059-5](https://doi.org/10.1016/s2589-7500(21)00059-5)
12. Louro, A., da Costa, N.M., da Costa, E.M.: Sustainable urban mobility policies as a path to healthy cities-the case study of LMA, Portugal. *Sustainability (Switzerland)* **11**(10) (may 2019). <https://doi.org/10.3390/su11102929>
13. Martin, R.C.: Clean Coder (2014), <https://blog.cleancoder.com/uncle-bob/2012/08/13/the-clean-architecture.html>
14. Qian, W., Lauri, F., Gechter, F.: A Probabilistic Approach for Discovering Daily Human Mobility Patterns with Mobile Data. In: *Communications in Computer and Information Science*. vol. 1237 CCIS, pp. 457–470. Springer (jun 2020)
15. Rosa, L., Silva, F., Analide, C.: Mobile networks and internet of things: Contributions to smart human mobility. In: *Advances in Intelligent Systems and Computing*. vol. 1237 AISC, pp. 168–178. Springer (2021). <https://doi.org/10.1007/978-3-030-53036-5-18>
16. Stachl, C., Au, Q., Schoedel, R., Gosling, S.D., Harari, G.M., Buschek, D., Völkel, S.T., Schuwerk, T., Oldemeier, M., Ullmann, T., Hussmann, H., Bischl, B., Bühner, M.: Predicting personality from patterns of behavior collected with smartphones. *Proceedings of the National Academy of Sciences of the United States of America* **117**(30), 17680–17687 (jul 2020). <https://doi.org/10.1073/pnas.1920484117>
17. Thomas, T., Geurs, K.T., Koolwaaij, J., Bijlsma, M.: Automatic Trip Detection with the Dutch Mobile Mobility Panel: Towards Reliable Multiple-Week Trip Registration for Large Samples. *Journal of Urban Technology* **25**(2), 143–161 (apr 2018). <https://doi.org/10.1080/10630732.2018.1471874>, <https://www.tandfonline.com/action/journalInformation?journalCode=cjut20>
18. Watson, J.R., Gelbaum, Z., Titus, M., Zoch, G., Wrathall, D.: Identifying multi-scale spatio-temporal patterns in human mobility using manifold learning. *PeerJ Computer Science* **6**, e276 (2020). <https://doi.org/10.7717/peerj-cs.276>
19. Yabe, T., Tsubouchi, K., Fujiwara, N., Wada, T., Sekimoto, Y., Ukkusuri, S.V.: Non-compulsory measures sufficiently reduced human mobility in Tokyo during the COVID-19 epidemic. *Scientific Reports* **10**(1), 18053 (dec 2020). <https://doi.org/10.1038/s41598-020-75033-5>
20. Zhao, C., Zeng, A., Yeung, C.H.: Characteristics of human mobility patterns revealed by high-frequency cell-phone position data. *EPJ Data Science* **10**(1), 1–14 (dec 2021). <https://doi.org/10.1140/epjds/s13688-021-00261-2>