

Changing Pattern of Human Movements in Istanbul During Covid-19

Ayse Giz Gulnerman(✉)

Ankara Haci Bayram Veli University, 06450 Ankara, Turkey
ayse.gulnerman@hbv.edu.tr

Abstract. Human trajectories provide spatial information on how citizens interact with the city. This information can explain the daily routine, economic and cultural cycle of citizens. In the last year, the pattern of human trajectories is expected to change over cities due to Covid-19. This paper investigates the change in human trajectories based on the social media data (SMD) before and after the Covid-19. This study aims to find out the differences between the years 2018, 2020, and 2021. Firstly, all accounts for each year are classified based on movement behaviors. Secondly, spatial distributions of the tweets in terms of classified accounts are visualized after the hierarchical clustering applied to each dataset. Lastly, the average step lengths (ASL) are calculated for each account and classified in terms of step length levels as no movement, neighborhood, district, inter-districts, inter periphery, and center, outbound. The number of tweets and distinct accounts decreased by 90% and 84% from 2018 to 2021. The decrease in the number of single tweeting accounts is 84%, it is 60% in stationary accounts, and 94% in moving ac-counts. The size of the spatial clusters also decreased for all types of accounts maps, however, some of the previously visited spatial points are disappeared while new ones appeared on maps of single tweeting and moving accounts. The ASL of moving accounts also confirms the human movement decrease. According to that, the max, mean, and median ASL decreased 22%, 13%, and 35%. Results point out outcomes vary in terms of accounts' movement behaviors. This study is expected to contribute the measuring the pandemic impacts on human movement with SMD.

Keywords: Human movement · Spatial trajectories · Social media analysis · Covid-19

1 Introduction

1.1 Human Trajectories and Mobility

A human is a moving object that interacts with the city [1–3] for diverse purposes and under diverse circumstances. Mobile technologies and apps have enabled the retrieval of this interaction as human trajectories. A trajectory is simply defined as an array of timestamped locations (a pair of latitude and longitude) with/-out attributes [4]. Trajectory data vary in terms of spatial and/or temporal frequency based on data collection settings

© Springer Nature Switzerland AG 2021
O. Gervasi et al. (Eds.): ICCSA 2021, LNCS 12957, pp. 220–230, 2021.
https://doi.org/10.1007/978-3-030-87013-3_17

(either pre-defined or controllable). Social media is one of the sources for crowdsourced trajectory data where each location pair is generated by geotagged social media posts [5]. In other words, social media is an unsystematic trajectory source since it has no pre-defined frequency settings and the data generation is entirely dependent on the users' instant preferences or un-consciously defined default settings.

Trajectory data with diverse data mining techniques provide diverse information. Zheng [6] defines the trajectory data mining flow in three bases; preprocessing (noise filtering, stay point detection, segmentation), indexing and retrieval, and processing (pattern mining, classification, anomaly detection). Buchin [7] categorizes the purpose of trajectory data analysis in terms of the number of trajectories (single, two, multiple). According to that, recurring pattern detection might be one purpose of single trajectory mining, while common or typical trajectory analysis could be performed with multiple trajectories.

Jurdak et al. [8] mention that there are diverse tracking systems (such as; mobile phones, GPS, and RFID) on human movement. However, these tracking systems have concerns such as; ethical breaching issues, data access problems, and low resolutions in data [5, 8]. Demšar [9] points out one of the big problems in movement science is modifiable areal unit problems (MAUP) that is different scales and shapes applied to movement data might return different insights. Dodge [5] also mentions responsible Computational Movement Analysis (CMA) is important to lead the vital studies that emerged and enhanced with the Covid-19 in this area.

Geospatial data analytics with trajectory analysis became a more important topic than ever before with the Covid-19 since the pandemic distort the usual human trajectories at any granularity (global and regional) [5, 10]. Jurdak et al. [8] demonstrated the social media (Twitter) is a reliable source for analyzing human movement patterns. So far, social media-based trajectory data is used to analyze human mobility patterns [11], to reveal trajectory patterns characteristic of traveling attendees of large-scale events [12], urban mobility patterns based on urban dynamics, and functional places [13, 14].

Social media is a wide data source that is used for the research on diverse topics such as; community detection [15], prediction of customers sentiments [16], and regulation of the use of social media to prevent disinformation during political campaigns [17]. In this study, social media data is assessed as a spatial trajectory source for monitoring human movement.

1.2 Aim of the Study

This study aims to demonstrate the change in human movement between the years (2018, 2020, 2021) based on social media (Twitter) data and search the impacts of Covid-19 on human movement within Istanbul city. In addition to the main aim, there are sub-objectives determined in order to reveal the possible indicators for monitoring the human movement change. The first sub-objective is finding out how the change in the number of tweets impacts the spatial distribution of tweets. The second objective is assessing the change in tweet distribution according to users' tweeting behavior. The third objective is revealing the change in step lengths of moving users under different circumstances. All these sub-objectives are aimed to define possible indicators to contribute algorithms built for human activity monitoring analytics based on social media data.

2 Data and Evaluation

This study investigates social media data (Twitter) for the years 2018, 2020, and 2021 collected via Twitter API. This API returns free data which is randomly selected 1% of all generated data in the meantime. Time interval is determined for each year in terms of the Covid-19 case declaration day (11th of March 2020) in Turkey. 15 days as one week before and after the 11th March accepted as the time interval and retrieved data for each year summarized in Table 1.

Table 1. Social media data details in terms of year 2018, 2020, and 2021.

* (number of)	2018	2020	2021
* tweets	459,237	71,138	48,952
* distinct accounts	25,836	7278	4192
* single tweeting accounts	9612	3233	1552
* stationary accounts	4434	2039	1763
* moving accounts	11,790	2006	877

Accounts are categorized in terms of movement behaviors i.e., single tweeting, stationary, and moving. Accounts having single tweets have no values for movement step length comparison however contribute to change in the spatial density and place semantic interpretation. Accounts having multiple tweets classified as stationary or moving accounts depending on the calculated bounding box (BBOX) area for each. In this determination, 100 m^2 is accepted as GPS locating error for mobile devices in all directions. According to that, accounts having less than 0.04 km^2 BBOX area of their tweets are determined as stationary, others are determined as moving accounts.

It is clearly seen that the total amount of data in 2018 is quite high in every field when it is compared to the data in 2020 and 2021 in Table 1. Although the % of moving users decreased dramatically as well as the total amount, the % of stationary users increased more than double from 2018 to 2021 (Fig. 1). The single tweeting % represents the most

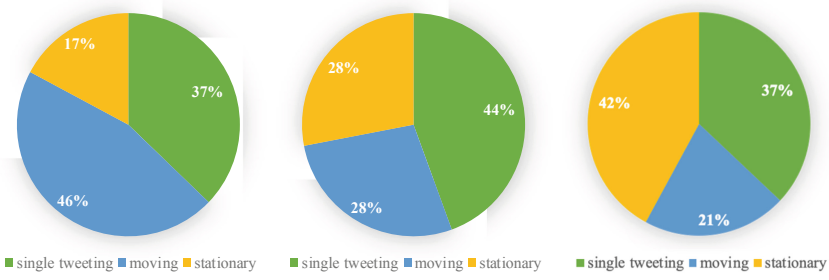

Fig. 1. Users' moving behavior charts for the year (a) 2018 (b) 2020 (c) 2021.



Fig. 2. The number of categorized accounts in terms of moving behaviors by day for the year (a) 2018 (b) 2020 (c) 2021.

stable part of this slicing comparison which might provide a valuable benchmark for further investigations.

Overall numbers of categorized accounts provide the information that the % of moving accounts and stationary accounts shifted between 2018 and 2021. In order to add a temporal fine-grained perspective, Fig. 2 shows the number of categorized accounts by day for the years 2018, 2020, and 2021. Trends for all category user counts are very similar in 2018 while in 2020, two of them (single tweeting and moving accounts) are similar in 2020. Also, none of the categories are showing a correlation in 2021. In 2018, single tweeting and moving users' counts are reaching a peak on the weekends. In 2020, single tweeting and moving users are reaching a peak in the weekends (7th-8th March) however after the Covid-19 declaration on 11th of March, the number of users for all kinds show decreasing trends even in the following weekends (14th–15th of March). In 2020, the number of moving accounts might be expected to decrease more with the 1st case declaration day however the partial precautions decision has been taken and announced on the late 12th of March for the later weeks. In 2021, all lines are fluctuating and the peaks seen in the former years' weekends are not seen this year.

Spatial footprints of the accounts for each categorization are visualized in Fig. 3 by year. A hierarchical clustering algorithm is used as an aggregation method to cluster spatial points (tweets). The evaluation process follows;

- geodesic distance calculation between spatial points with "distm" function from "geosphere" package [18]
- building hierarchical structure with "hclust" function from "stats" package [19]
- clustering points setting 5 km as distance threshold with the "cutree" function from "stats" package [19]
- calculating spatial center of each cluster with the "gCentroid" function from "rgeos" package [20]
- plotting each cluster centroid with "tm_bubbles" function from "tmap" package [21] defining size as point count facets by years

steps and applied to each set of tweets from single tweeting, stationary, and moving accounts. Since stationary accounts have tweets from the very same locations, an additional process is applied by grouping the tweets from each account beforehand.

Figure 3 shows the tweet density and distribution maps for each year in terms of movement behaviors of accounts. The overall decrease in the number of tweets between 2018, 2020, and 2021 is also spatially observed. Tweet clusters of single tweeting accounts, concentrated in the central part of the Istanbul city in 2018, dispersed to the outer parts and getting smaller in 2020, and get concentrated in the center again with a much smaller size in 2021 (Fig. 3 (a)). Tweet clusters of stationary accounts between years represent the most stable cluster distribution. These clusters emerged at very similar locations with the size reduction in years (Fig. 3 (b)). Tweet clusters of moving accounts also concentrated in the central for all years with the size reduction. Be-side the reducing size of the clusters, some of them disappeared in years in the outer parts and around the central parts of Istanbul (Fig. 3. (c)).

Human movement variation between years is analyzed referring to the moving accounts. Each step length and bearing angle is calculated between timely ordered tweets

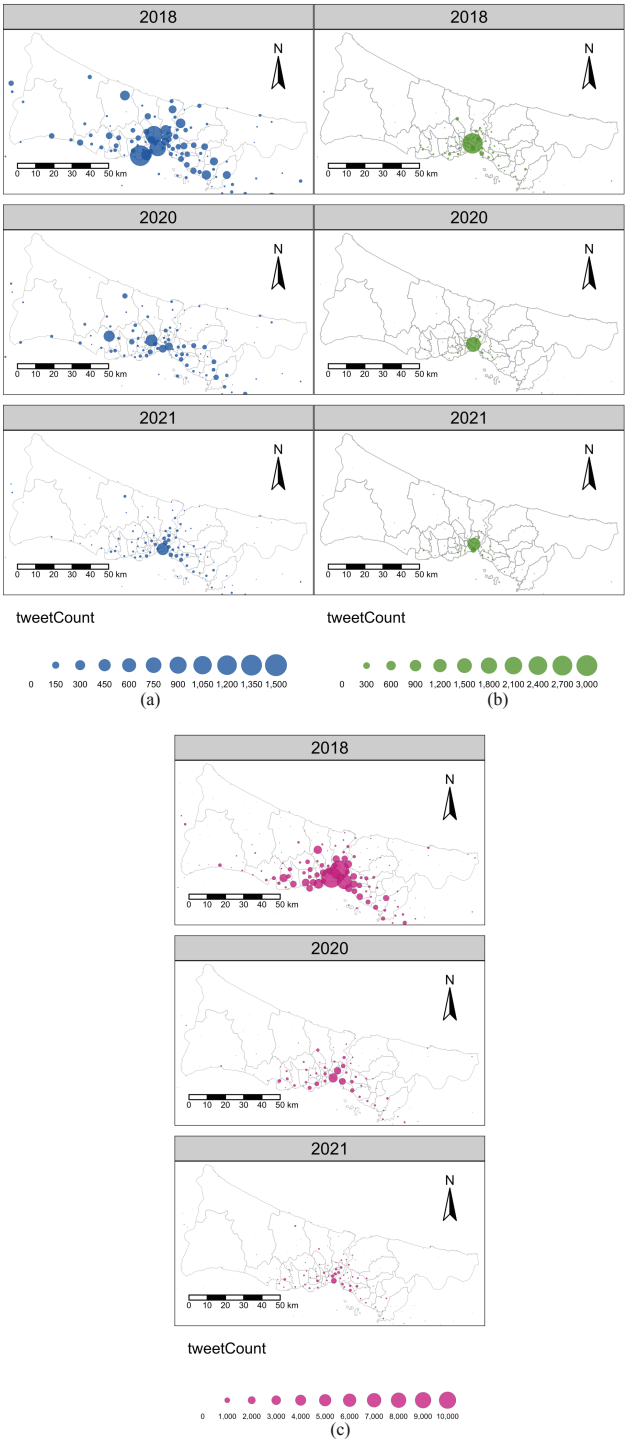

Fig. 3. Tweet distribution by year in terms of movement behavior of users (a) single tweeting users (b) stationary users (c) moving users.

for each user with the "moveHMM" R package [22]. A summary table for the movement data is given in Table 2. Mean step length and mean average step lengths are both decreased between years. In addition to that, median values are decreased more than mean values which means the number of tweets tends to right skewing in the step length distribution. Also, standard deviation increased in years and even higher than the mean value which represents the data widely distributed.

Table 2. Moving accounts movement details in terms of year 2018, 2020, and 2021.

(* km unit)	2018	2020	2021
Max. step length	123.98	132.24	165.12
Mean step length	9.36	9.06	8.91
Median step length	5.80	5.00	4.56
Std. dev. step length	11.00	11.78	11.94
Max average step lengths of an account	116.5	131.70	91.19
Mean average step lengths of accounts	10.83	9.81	9.41
Median average step lengths of accounts	8.24	6.08	5.35
Std. dev. of average step lengths of accounts	9.83	11.78	12.37

Furthermore, moving accounts steps are classified as to their step lengths and six levels are defined as no movement, neighborhood, district, inter-districts, inter periphery, and center, outbound. Break values are determined for the defined levels as 0.1 km, 1 km, 5 km, 10 km, 30 km, and above respectively considering the area of districts and distances to the central area. Figure 4 shows the number of steps with regards to moving length levels and years. No movement level steps are omitted in this representation. The number of steps in each level is ahead by far for the year 2018. However, the number of steps for each level decreased from 2018 to 2021, the decrease ratio between 2018 and 2020 is approximately three times higher than the de-crease ratio between 2020 and 2021.

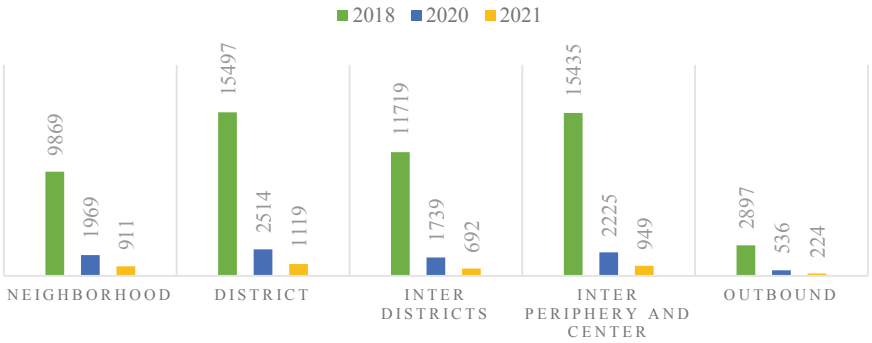

Fig. 4. The number of steps by each year in terms of step length levels.

Figure 5 represents the distribution of steps with regards to their length levels for the years 2018 (a), 2020 (b), and 2021 (c). While the neighborhood ratio is 18% in 2018, it is 22% in 2020 and 23% in 2021. Inter districts and inter periphery and center levels have a higher ratio in 2018 than in 2020 and in 2021. However, these ratio differences are not significant as in the number of differences seen in Fig. 4.

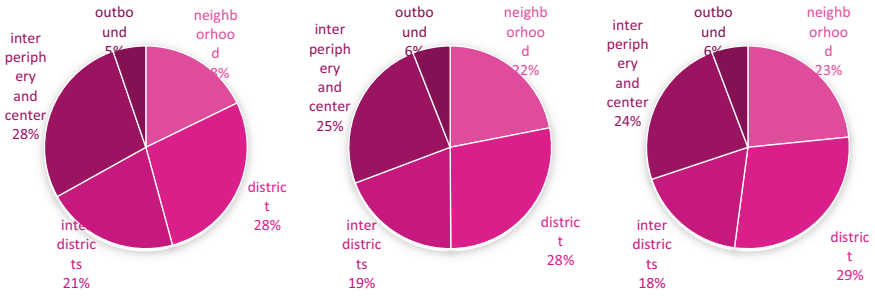

Fig. 5. The distribution of moving accounts step length levels for the year (a) 2018 (b) 2020 (c) 2021.

3 Results and Discussion

In the case study, three sets of data belong to years 2018, 2020, and 2021 are investigated and compared with regards to human movement behaviors. The first outcome of this investigation is the significant decrease in the number of tweets and the number of distinct accounts. This decrease is greater in the number of moving accounts while the least decrease seen proportionally in the number of stationary accounts. Therefore, omitting stationary users from the datasets in the pandemic estimation studies might not disrupt the results but benefit to reducing time cost. The change in number of accounts by day show while the all types of account movement behaviors show a similar trend after the pandemic declaration the number of single tweeting and moving accounts tend to decrease. However, the week after the Covid-19 declaration in 2020, the number of stationary accounts steadily tend to increase. Also, this second inference is supporting the stationary accounts can be omitted since it has no inferred change before and after Covid-19 declaration.

The spatial distribution of the stationary accounts is unchanged in years, however, the spatial cluster size changed proportionally between years. Since the number of tweets decreased, the size of the spatial clusters became small for all movement types. However, the single tweeting accounts map shows new locations become popular in the outer part of the Istanbul central area in 2020 and 2021. The map showing the clustered tweets of moving accounts shows that there is a significant decrease in cluster size and also many points are disappeared in 2020 and 2021. This reveals the certain change in moving accounts and the decreasing magnitude of its impact within the overall dataset. According to that, sets of tweets should be assessed separately based on accounts movement behaviors in order to see the real effects of the Covid-19.

Movement details of moving accounts show the shortening step lengths which means accounts travel to nearer locations than in 2018. Increasing standard deviations also reveals the growing disparities between accounts moving behaviors. Though the number of moving level decreases is higher between 2018 and 2020 than the decrease between 2020 and 2021, the decrease still continues. This might refer to adaptation to the pandemic normal but also refer to continuing migration from the city. As it is known mostly younger generations tend to share their locations and trajectories through social media. Millions of students are living in Istanbul and many of them immigrated back to their family houses in the other cities. So, the decrease in the number of moving accounts in each step length level is not surprising at all.

4 Conclusion

Geoinformatics has a leading role to manage the pandemics since Snow's Cholera Map [23]. Two centuries ago, data retrieval was a big problem, especially for wide areas. While several kinds of spatial data sources are available nowadays, the data need is still a big issue due to some reasons. Social media is a valuable continuous way of retrieving data, though it has data access limitations, representation and spatial bias, and uncertainty problems [24]. Therefore, social media is an open area to discover new geospatial data inferring techniques to diverse domains [25].

In this study, human movement change within Istanbul before and after Covid-19 is monitored with social media data. Tweets from classified accounts based on movement behaviors are investigated to reveal the change between years. According to that, the number of tweets and accounts decreased 90% and 84% between years. In a further look, this decrease in the number of accounts varies in terms of movement behaviors of accounts. While the number of single tweeting accounts decreased 84%, the decrease is 60% in stationary accounts and 94% in moving accounts. According to this, the most decrease was seen in the number of moving accounts.

The change in the spatial distribution of tweets is separately assessed with maps in terms of classified accounts. Generally, the size of the spatial clusters decreased in all maps of classified accounts. This unchanged spatial distribution is pointed out which might be omitted for further investigations to remove illusive weights on the spatial distribution. However, some of the spatial clusters are disappeared and new clusters emerged on maps of single tweeting and moving accounts. In addition to this spatial distribution, moving accounts add a new dimension to the results with the step length changes. According to that, human movement decrease is confirmed with the decrease in the max, mean, and median average step lengths of moving accounts which are 22%, 13%, and 35% respectively.

This study is limited with the data interval determined considering the first case of the Covid-19 declaration in Turkey. Only 15 days of each year is assessed and the outer regions of the Istanbul BBOX area discarded. The users that have only one tweet inside the BBOX of Istanbul and have other tweets outside are assessed within the single tweeting accounts. This study is planned to enhance the semantic segmentation of stay points of the single tweeting and moving users. In that way, the changing pattern of human movements is detailed with another dimension. In further studies, anomaly patterns of human behaviors are aimed to estimate the significant impacts of pandemics.

References

1. Ahas, R., et al.: Everyday space–time geographies: using mobile phone-based sensor data to monitor urban activity in Harbin, Paris, and Tallinn. Int. J. Geogr. Inf. Sci. **29**(11), 2017–2039 (2015)
2. Huang, Q., Wong, D.W.: Activity patterns, socioeconomic status and urban spatial structure: what can social media data tell us? Int. J. Geogr. Inf. Sci. **30**(9), 1873–1898 (2016)
3. Noulas, A., Scellato, S., Lambiotte, R., Pontil, M., Mascolo, C.: A tale of many cities: universal patterns in human urban mobility. PloS ONE **7**(5), e37027 (2012)
4. Buchin, K., Löffler, M., Popov, A., Roeloffzen, M.: Fréchet distance between uncertain trajectories: computing expected value and upper bound. In: 36th European Workshop on Computational Geometry (EuroCG 2020). Würzburg, Germany (2020)
5. Dodge, S., Gao, S., Tomko, M., Weibel, R.: Progress in computational movement analysis–towards movement data science (2020)
6. Zheng, Y.: Trajectory data mining: an overview. ACM Trans. Intell. Syst. Technol. (TIST) **6**(3), 1–41 (2015)
7. Buchin, K.: Trajectory Similarity. Winter School on Computational Geometry AUT, 11th Winter School on Computational Geometry. https://www.youtube.com/watch?v=GqOIzUEZgsA&list=LL&index=17. Accessed 03 May 2021
8. Jurdak, R., Zhao, K., Liu, J., AbouJaoude, M., Cameron, M., Newth, D.: Understanding human mobility from Twitter. PloS ONE **10**(7), e0131469 (2015)
9. Demšar, U., Long, J.A., Siła-Nowicka, K.: Integrated science of movement. J. Spat. Inf. Sci. **2020**(21), 25–31 (2020)
10. Minghini, M., Coetzee, S., Grinberger, A.Y., Yeboah, G., Juhász, L., Mooney, P.: OpenStreetMap research in the COVID-19 era. Editors **1** (2020)
11. Ma, D., Osaragi, T., Oki, T., Jiang, B.: Exploring the heterogeneity of human urban movements using geo-tagged tweets. Int. J. Geogr. Inf. Sci. **34**(12), 2475–2496 (2020)
12. Xin, Y., MacEachren, A.M.: Characterizing traveling fans: a workflow for event-oriented travel pattern analysis using Twitter data. Int. J. Geogr. Inf. Sci. **34**(12), 2497–2516 (2020)
13. Gabrielli, L., Rinzivillo, S., Ronzano, F., Villatoro, D.: From tweets to semantic trajectories: mining anomalous urban mobility patterns. In: Nin, J., Villatoro, D. (eds.) CitiSens 2013. LNCS, vol. 8313, pp. 26–35. Springer, Cham (2013). https://doi.org/10.1007/978-3-319-04178-0_3
14. McKenzie, G., Janowicz, K., Gao, S., Gong, L.: How where is when? On the regional variability and resolution of geosocial temporal signatures for points of interest. Comput. Environ. Urban Syst. **54**, 336–346 (2015)
15. Kumari, A., Behera, R.K., Shukla, A.S., Sahoo, S.P., Misra, S., Rath, S.K.: Quantifying influential communities in granular social networks using fuzzy theory. In: Gervasi, O., et al. (eds.) ICCSA 2020. LNCS, vol. 12252, pp. 906–917. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58811-3_64
16. Behera, R.K., Jena, M., Rath, S.K., Misra, S.: Co-LSTM: convolutional LSTM model for sentiment analysis in social big data. Inf. Process. Manag. **58**(1), (2021)
17. Slös Wogu, I.A.P., Njie, S.N.N., Katende, J.O., Ukagba, G.U., Edogiawerie, M.O., Misra, S.: The social media, politics of disinformation in established hegemonies, and the role of technological innovations in 21st century elections: the road map to us 2020 presidential elections. Int. J. Electron. Gov. Res. (IJEGR) **16**(3), 65–84 (2020)
18. Hijmans, R.J.: Introduction to the "geosphere" package (Version 1.5-10) (2019)
19. R Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2020)

20. Bivand, V.R., Rundel, C.: rgeos: Interface to Geometry Engine - Open Source ('GEOS'). R package version (2020)
21. Tennekes, M.: tmap: Thematic maps in R. J. Stat. Softw. **84**(6), 1–39 (2018)
22. Michelot, T., Langrock, R., Patterson, T.A.: moveHMM: an R package for the statistical modelling of animal movement data using hidden Markov models. Methods Ecol. Evol. **7**(11), 1308–1315 (2016)
23. Snow, J.: On the Mode of Communication of Cholera. John Churchill (1855)
24. Gulnerman, A.G., Karaman, H., Pekaslan, D., Bilgi, S.: Citizens' spatial footprint on Twitter—anomaly, trend and bias investigation in Istanbul. ISPRS Int. J. Geo Inf. **9**(4), 222 (2020)
25. Gulnerman, A.G., Karaman, H.: Spatial reliability assessment of social media mining techniques with regard to disaster domain-based filtering. ISPRS Int. J. Geo Inf. **9**(4), 245 (2020)