Assessing the Relationship Between Socio-Demographic Characteristics and OpenStreetMap Contributor Behaviours

Dominick Sutton
University of Glasgow
Glasgow, United Kingdom
2428886s@student.gla.ac.uk

Guy Solomon
University of Glasgow
Glasgow, United Kingdom
Guy.Solomon@glasgow.ac.uk

Xinyi Yuan
University of Glasgow
Glasgow, United Kingdom
x.yuan.1@research.gla.ac.uk

Merve Polat Kayali
University of Glasgow
Glasgow, United Kingdom
m.polat-kayali.1@research.gla.ac.uk

Zoe Gardner
University of Leicester
Leicester, United Kingdom
zg61@leicester.ac.uk

Ana Basiri
University of Glasgow
Glasgow, United Kingdom
Ana.Basiri@glasgow.ac.uk

ABSTRACT

'Volunteered Geographic Information' (VGI) has particular importance – in part – for its democratisation of geographic information. However, some recent research has suggested that despite being publicly open, several successful VGI platforms have under-representation of particular socio-demographic groups, which may lead to biases in the types of information contributed. This paper examines the relationship between demographic characteristics and user contributions to OpenStreetMap (OSM), one of the most successful examples of a project reliant on VGI. It demonstrates statistically significant differences in the information provided by users of different genders, ages, and education-levels. Differences between the demographic characteristics of OSM contributors and the underlying population are therefore likely to be reflected in the VGI contained in OSM.

CCS CONCEPTS

• **Social and professional topics** → *User characteristics*; • **Information systems** → **Geographic information systems**; • **Human-centered computing** → **Empirical studies in collaborative and social computing**.

KEYWORDS

Contributor Bias, Crowdsourcing, Volunteered Geographic Information, Gini-Simpson Index, OpenStreetMap

ACM Reference Format:
Dominick Sutton, Guy Solomon, Xinyi Yuan, Merve Polat Kayali, Zoe Gardner, and Ana Basiri. 2023. Assessing the Relationship Between Socio-Demographic Characteristics and OpenStreetMap Contributor Behaviours. In *1st ACM SIGSPATIAL International Workshop on Geocomputational Analysis of Socio-Economic Data (GeoSocial '23), November 13, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3615892.3628477

GeoSocial '23, November 13, 2023, Hamburg, Germany

ACM ISBN 979-8-4007-0354-6/23/11…$15.00
https://doi.org/10.1145/3615892.3628477

1 INTRODUCTION

The term 'Volunteered Geographic Information' (VGI) was first used by Goodchild [14] to describe the creation and distribution of geographic data drawn from a community of volunteers. VGI projects rely on participants to add and modify spatial data, and only require limited geographic expertise in order to contribute [15].

OpenStreetMap (OSM) has emerged as one of the leading examples of VGI and the crowdsourcing of geographic data [3, 33, 25]. However - as various authors [6, 7, 16, 17, 24, 4, 29, 28] have highlighted - although OSM (and other VGI projects) are typically open to anyone to add data and have no restrictions on access, there are significant demographic participation biases in contributors. This is important because biases in *contributors* might lead to biases in *contributions*, and could potentially result in unrepresentative data.

In this paper, we analyse the relationship between demographic characteristics and contributor behaviours. We analyse two different aspects of user contributions: firstly, the diversity of countries to which they contributed; and secondly, the types of changes made ('Creations', 'Modifications', and 'Deletions'). Whilst propensity for both the type and location of user contributions have been shown to vary between demographic groups [8, 13, 27], *where* people map and *how* people map may be driven by distinct skills, motivations, and interests.

We demonstrate that contributors with different gender, age, and educational backgrounds display different user behaviours – and by extension, that under-representation of particular combinations of these factors (relative to the underlying population) may cause bias in the information on which OSM is based. This may influence the usefulness of OSM (for example: in relation to navigation or the identification of points of interest), as well as any conclusions drawn on the basis of OSM data.

2 BACKGROUND

Technological developments have transformed the creation and use of spatial data [2]. Crowdsourcing in particular has become more organised with the widespread use of social media, and increasingly easy access by participants from anywhere in the world [18, 9]. OSM is one of the most successful projects to emerge from this new landscape. The rapid growth of OSM as a platform has in large part been thanks to the participation of (at time of writing) over 10 million registered members [23], and the accessibility of data in much of the world has created significant potential for the development and growth of OSM beyond its original scope [26].

Previous research has noted a strong gender skew; with approximately 96 percent of contributors reported as male [16, 25, 8, 29, 5]. Such biases may be compounded by the 'long tail effect', whereby a small number of users (who are not representative of the entire population) contribute the majority of information [32, 1, 17, 20].

Studies on the nature of contributions to OSM have placed an emphasis on associations between gender and user engagement or behaviour. Steinmann et al have highlighted the potential importance of 'social motives' and strict rules for the nature of contributions in driving female engagement, neither of which – they argue – are present in OSM [28]. Das et al found that men were comparably more likely to contribute to mapping in rural areas and women to urban ones, as well as men making comparatively more contributions to 'feminized spaces' and women to their 'masculinized' counterparts [8]. Gardner and Mooney, and Gardner et al found women were comparatively more likely to create objects, whereas by contrast men were more likely to edit them [11, 12, 13]. However, whilst Gardner et al [13] note that education level may play a role in user behaviours, there has been limited analysis of the extent to which interactions between gender, age, and education might influence the overall picture.

3 DATA

3.1 Data collection

This paper draws on an online survey of OSM contributors conducted by Gardner et al [12]. The 326 unique participants were asked to respond to six survey questions relating to their OSM username, gender, age, country of residence, nationality, and highest level of education. Although fewer women than men responded to the survey, women responded in greater proportion relative to OSM users – this provides good representation for the purposes of analysis.

Users who participated in the survey were linked to two further sources of data (on the basis of their self-reported username): their 'changesets', via the OSM API; and their user activity as summarised by 'How Did You Contribute to OpenStreetMap' (HDYC [22]), an online tool which provides an interpretation layer for some elements of this information.

An OSM changeset is a group of edits made by a particular user over a relatively short period of time. There is no formal limit on the geographical size of a changeset, but contributors are encouraged to restrict a single changeset to a small geographical area. In addition to information regarding the types of changes the user has made, the geographical extent of each changeset is identified by the latitude and longitude of the 'bounding box' for the group of edits [30]. This allows for identification of the geographic location and extent of each changeset contributed by the user.

Two related data sets are therefore obtained for each user: one which indicates the type of changes made (classified as either 'Created', 'Modified', or 'Deleted' by HDYC) and the object of the change ('Nodes', 'Ways', or 'Relations' [31]); and another which specifies the geographic range of a users' changes and consequently the total changes by country.

3.2 Data processing

Of the initial 326 respondents, 293 can be considered after accounting for duplication, lack of OSM activity, or irreconcilable usernames. Given the focus on the relationships between gender, age, and education with OSM contributions, users who selected 'Prefer not to say' to any of these questions were also excluded.

The respondents who could be matched to activity in OSM demonstrated a considerable range of engagement. In order to focus on established users with a relatively large body of OSM contributions, 7 users with a comparatively low number of changesets (defined as less than 1,000) were eliminated. Whilst this does mean the sample concentrates on a particular type of (strongly-engaged) user within the long tail, the survey responses do not provide sufficient representation of more sporadic users to enable confidence in analysis of this group. In addition, a single superuser with a disproportionately large number of changesets was removed to avoid skewing calculations involving mean averaging. This left a final user set of 246 individuals.

Changesets with missing geometry values were then removed – although these were relatively small in number, totalling 29,668 (or 1.4 percent of the original total). The centroids of remaining bounding boxes were then calculated. Using world country boundaries from ArcGIS Hub [10], a spatial intersection analysis with these centroids was conducted. In this way, the country in which each changeset was located could be extracted for all matched OSM users in the survey.

3.3 Sample characteristics

The demographic profile of the 246 individuals comprising the data set is given in table 1.

Table 1: Demographic breakdown of OSM users within analysis

		Age				
	Education level	18 - 29	30 - 39	40 - 49	50 or over	Total
Male	Not University	17	13	10	9	49
	University	27	45	28	19	119
	Postgrad - PhD	4	16	14	14	48
	Subtotal	48	74	52	42	216
Female	Not University	2	0	0	0	2
	University	10	3	1	0	14
	Postgrad - PhD	2	5	4	3	14
	Subtotal	14	8	5	3	30
Total		62	82	57	45	246

As can be seen from this table there is a significant imbalance between the genders (30 users who reported their gender as 'female' as opposed to 216 who reported their gender as 'male') – and with age and education also taken into account the disparities are magnified. Considering age alone, it can be seen that levels for men are of the same order of magnitude (with a peak at '30 - 39') whereas those for women drop off with age. As far as education is concerned, men are more likely to have been 'University' educated, with 'Not

University' and 'Postgrad - PhD' more-or-less equally likely. In contrast, only two women are in the 'Not University' category, with the rest equally divided between the other two categories. It is not known whether this is representative of the underlying OSM contributor population, but it does mean that trends which initially seem to be driven by gender may actually be a reflecting a combination of gender and educational attainment. Accordingly, ignoring educational level may risk introducing a confounder.

The total OSM edits for each of these 246 users, by main activity (Creation, Modification and Deletion) are given in table 2. The interaction between the demographic characteristics represented in table 1 and their OSM activity is shown in figure 1. This shows a hierarchical tree of total activity (Creations + Modifications + Deletions) by user demographic.

Table 2: OSM edits by type of activity, classified by gender

Gender	Created	Modified	Deleted	Total
Female	10,548,338	5,212,122	3,100,353	18,860,813
Male	180,918,464	62,105,321	31,005,329	274,029,114
Total	191,466,802	67,317,443	34,105,682	292,889,927

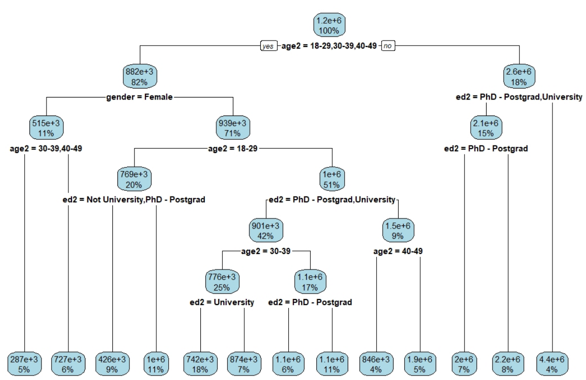

Figure 1: Hierarchical tree of key demographics and OSM activity

A breakdown of the proportions of the actions that are Creations, Modifications or Deletions is shown in table 3. In this case, the data is first normalised for each user to remove any scale distortions before the mean is calculated; and the results are split by gender. It demonstrates that men were comparatively more likely to modify objects, whereas women were slightly more likely to create them.

Table 3: OSM edits by type of activity, classified by gender

Gender	Created	Modified	Deleted
Female	0.685	0.218	0.097
Male	0.633	0.279	0.088

The disparity in activity between Female and Male users can be seen from the histograms of Creation activity by gender in figure 2, which represents the largest proportion of the edits (similar activity levels are found for both Modified and Deleted actions).

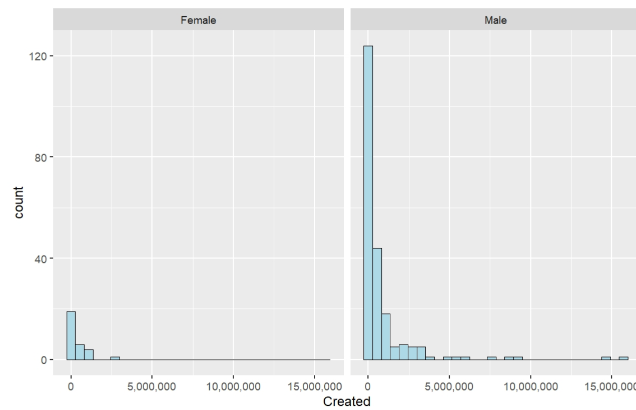

Figure 2: 'Created' activity classified by gender

A similar pattern is found if looking at activity classified into the editing of 'Nodes', 'Ways' and 'Relations', as can be seen from table 4. 'Nodes' represent a specific point-of-interest or location. By contrast, 'Ways' generally represent features such as roads or rivers (although can also represent boundaries). 'Relations' are used to represent connections between other elements, such as bus or cycle routes [31]. It can be seen that Nodes are the most common type of edit, with significantly lower levels of activity associated with Ways and Relations.

Table 4: Mean OSM edits by type, classified by gender

Type	Gender	Created	Modified	Deleted
Node	Female	262,707.9	117,954.2	78,185.8
	Male	807,684.1	231,240.0	359,333.5
Ways	Female	30,275.2	26,582.9	7,798. 9
	Male	88,676.8	94,293.8	12,047.6
Relations	Female	150.1	340.8	179.8
	Male	1,239.10	4,677.30	391.3
Total	Female	293,133.2	144,878.0	86,164.0
	Male	869,218.7	300,661.1	147,321.6

These tables show the differences in activity concealed by the overall edits data, even when split into Created, Modified and Deleted categories. Clearly, Node activity dominates OSM entries, with the lowest levels for both genders associated with Relations. Indeed, the low levels of female activity in Relations means that any tests carried out on these are unlikely to meet the standards for statistical validity. We expand on this in the following section.

4 ANALYSIS

As outlined in the previous section, the contributors within this sample are not gender balanced. However, as has already been highlighted, it is worth noting that OSM population has also under-representation issues. Thus, the sample is representative of the estimated underlying population of OSM – although true demographic characteristics are not known due to the anonymity of users.

In the data collection phase, the female contributors were over-sampled and therefore the sample actually contains a higher proportion of respondents identifying as female than would otherwise be expected. This leads naturally to an examination of possible differences in activity between users of the different gender, as well as a cross-examination of how gender interacts with two other key demographic measures: Age and Education Level.

4.1 Differences in geographic contributions

The first possible divergence is in the contribution of edits to different countries. To assess this, the Gini-Simpson's Index (D'), defined as one minus the Simpson's Index (D), is employed:

$$D' = 1 - D \text{ , and } D = \frac{\Sigma n(n-1)}{N(N-1)}$$

Where n refers to the number of contributions by a user for a single country, and N is the number of contributions summed across all countries. This index ranges from 0 to 1. The Simpson's Index was developed to evaluate the probability that two consecutive samples would be of the same type [19]; in this case that two randomly selected individuals would have the same country coverage.

The conversion to the Gini-Simpson version of the index (GSI) gives a more natural representation of diversity, where close to 1 represents higher and closer to 0, lower diversity. In this instance, diversity can be interpreted such that individuals with a high GSI score (near to 1) made contributions across many different countries in relatively even proportions; conversely, a low GSI score (near to 0) would indicate that user contributions were heavily concentrated in few countries. The mean GSI by gender across the users is given in table 5, which shows a higher GSI for women than men.

Table 5: Mean GSI by gender

Gender	Female	Male
Mean GSI	0.479	0.303

However, as can be seen from figure 3, there appears to be a complex, two-mode relationship between female gender and diversity; whereas for men diversity and density (the relative distribution of the data across the range of GSI measures) move in opposite directions. This means that whereas the contributions of men in the data set were most commonly concentrated in a few countries – although this was not a completely linear decline, with a smaller group of male contributors making geographically diverse contributions (demonstrated by the slight peak at the more diverse end of the scale). Women were more clearly divided into two groups; one which tended to concentrate on contributing in few countries, and another which mapped in a variety of places.

By contrast, both Age (figure 4) and Education (figure 5) show generally declining density as diversity increases for each group. Indeed, the shape of the distribution for each age category shows little variation, outside of 30-39 year olds being more heavily concentrated towards the non-diverse end of the scale. With regards to Education, all three categories demonstrate the heaviest heaping towards the non-diverse extreme, with smaller secondary concentrations at the other end of the scale – although this happens earlier for those in the 'Not University' category, and to a lesser extent for those in the 'University' category.

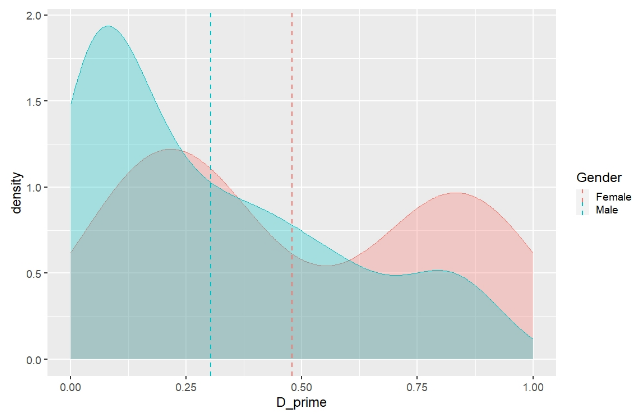

Figure 3: GSI density plot by gender

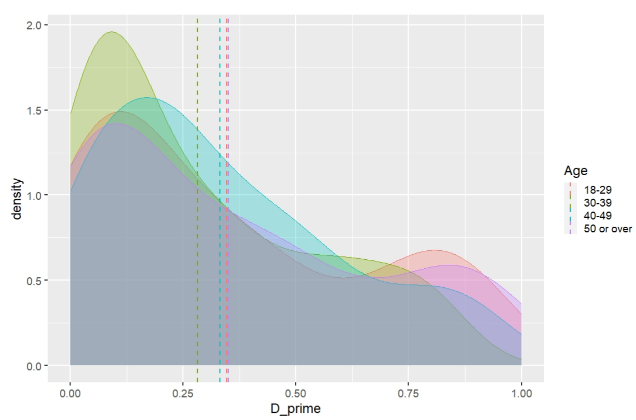

Figure 4: GSI density plot by age

Overall, this indicates that: women appear to be more likely to contribute more evenly to a range of countries than men, although this is at least partially driven by the presence of two (comparably sized) distinct groups; different age groups appear to have very similar patterns; and education level appears to exhibit a moderate effect, albeit one which does not seem to follow the ordinal nature of the categories.

4.2 Differences in types of contribution

The second element of this analysis is differences in Creations/Modifications/Deletions and Nodes/Ways/Relations behaviours. A series of Pearson's Chi-squared tests were carried out on these data to determine if there is a statistically significant difference between

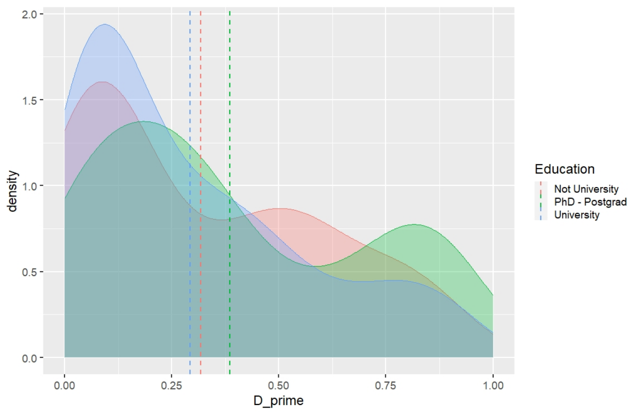

Figure 5: GSI density plot by education

the OSM activity of women and men. Pearson's Chi-squared is considered the most suitable test for these data owing to its non-parametric nature and the contingency table presentation of the data [21].

The results of these tests are given in table 6. They indicate that there are statistically significant differences between men and women in their average contributions to both types (Created, Modified, and Deleted) and target element (Nodes, Ways, and Relations) of changes.

Table 6: Results of Pearson's Chi-squared tests on gender-segmented data

Data	Variables (mean values)	Df	P-value
Total edits	Created, Modified, Deleted	2	$2.2e^{-16}$
Nodes edits	Created, Modified, Deleted	2	$2.2e^{-16}$
Ways edits	Created, Modified, Deleted	2	$2.2e^{-16}$
Relations edits	Created, Modified, Deleted	2	$2.2e^{-16}$
Created edits	Nodes, Ways, Relations	2	$2.2e^{-16}$
Modified edits	Nodes, Ways, Relations	2	$2.2e^{-16}$
Deleted edits	Nodes, Ways, Relations	2	$2.2e^{-16}$

In order to further investigate the apparent gender differences a series of Pearson's Chi-squared tests were carried out on the interaction of age and educational attainment. Owing to the small numbers in some age and educational categories (see table 1) the age categories were amalgamated to over/under 30 year of age, and the educational attainment to 'no university or undergraduate'/'postgraduate'. Even so, some categories were at the threshold of statistical validity and so these results need to be considered indicatory rather than definitive. The breakdown of these categories is given in table 7.

Using this data categorisation the following Pearson's Chi-square tests were undertaken: younger women v older women; younger men v older men; younger women v younger men; older women v older men; less highly educated women v more highly educated women; less highly educated men v more highly educated men; less highly educated women v less highly educated men; and more highly educated women v more highly educated men. The results of these analyses are given in table 8.

Table 7: Clustered breakdown of OSM users

		Age		
	Education level	18 - 29	30 or over	Total
Male	Not University & University	44	124	168
	Postgraduate - PhD	4	44	48
	Subtotal	48	168	216
Female	Not University & University	12	4	16
	Postgraduate - PhD	2	12	14
	Subtotal	14	16	30
Total		62	82	246

Table 8: Results of Pearson's Chi-squared tests on gender-segmented data

Data	Variables (mean values)	Df	P-value
Younger women v Older women	Created, Modified, Deleted	2	$2.2e^{-16}$
Younger men v Older men	Created, Modified, Deleted	2	$2.2e^{-16}$
Younger women v Younger men	Created, Modified, Deleted	2	$2.2e^{-16}$
Older women v Older men	Created, Modified, Deleted	2	$2.2e^{-16}$
Less highly educated women v More highly educated women	Created, Modified, Deleted	2	$2.2e^{-16}$
Less highly educated men v More highly educated men	Created, Modified, Deleted	2	$2.2e^{-16}$
Less highly educated women v Less highly educated men	Created, Modified, Deleted	2	$2.2e^{-16}$
More highly educated women v More highly educated men	Created, Modified, Deleted	2	$2.2e^{-16}$

Even with allowances made for the small sample sizes in some categories these results suggest that there can be statistically significant differences in the activity of OSM users depending on their gender, age and educational level. These results also match with the finding from the GSI data, which shows differences in density across gender, age and educational attainment. It means that the low representation of women - particularly older and/or less highly educated women - may result in a bias against OSM entries that are of particularly relevant to these cohorts in society.

5 LIMITATIONS AND FUTURE WORK

The biases in user behaviours have significant implications for understanding the extent to which VGI can be considered representative of underlying populations. Although we reaffirm the finding that gender effects are statistically significant, the evidence presented here demonstrates that this cannot be characterised solely as a male-female dichotomy. Instead, segmenting the data by gender also reveals important differences between users of different ages and with different educational attainment. Nonetheless, there are several limitations we intend to address in future work.

First, we intend to re-survey OSM contributors. The original survey took place in 2017, and although this does not present a problem for tracking the behaviours of identified users, it is possible that the socio-demographic profile of individual users and OSM in general may have shifted in the intervening years. Re-sampling therefore provides the opportunity to obtain updated demographic information, as well as include users who may have started mapping after 2017. This also provides the option to obtain further socio-demographic information – such as ethnicity or income – which may provide additional insights regarding contributor behaviour. A major extension of this exercise would be collecting data concerning the causes of differences in contributions.

Second, although we are able to compare the contributions of individual users, we presently do not account for changes in behaviour over time. It is plausible that user behaviour may evolve (for example: because a user develops additional confidence/expertise due to experience, or their personal circumstances alter their capacity to engage). Future iterations of this work must therefore incorporate a time-series analysis.

Finally, the Gini-Simpson index for location of changesets does not account for distance between countries. For example, a user mapping in three different countries on three different continents would not be differentiated from a user mapping in three countries on the same continent by this measure, so long as the number and ratio of changesets were otherwise equal. Whether this distinction is important is presently unknown, but requires further consideration. Given the practical differences between 'armchair mapping' and local knowledge, it likely that distance is not the only element in determining user propensity to contribute to the mapping of a location, but this relationship is important to evaluate.

6 AUTHOR CONTRIBUTIONS

Author contributions are mapped to the following categories:

DS: conceptualisation and scoping, data processing, methodology, analysis, visualisation, writing, editing and reviewing.

GS: conceptualisation and scoping, data processing, methodology, writing, editing and reviewing, organisation and supervision.

XY: conceptualisation and scoping, data processing, methodology, writing.

MPK: conceptualisation and scoping, literature review, data processing, methodology, writing.

ZG: conceptualisation and scoping, data collection, editing and reviewing.

AB: conceptualisation and scoping, methodology, writing, editing and reviewing, organisation and supervision, funding acquisition.

ACKNOWLEDGMENTS

The authors received the support from the UK Research and Innovation Future Leaders Fellowship "Indicative Data: Extracting 3D Models of Cities from Unavailability and Degradation of Global Navigation Satellite Systems (GNSS)", grant number MR/S01795X/2, the Republic of Turkey Ministry of National Education Scholarship Program, and China Scholarship Council.

REFERENCES

[1] Anahid Basiri et al. "Crowdsourced geospatial data quality: Challenges and future directions". In: *International Journal of Geographical Information Science* 33.8 (2019), pp. 1588–1593. DOI: 10.1080/13658816.2019.1593422.

[2] Michela Bertolotto, Gavin McArdle, and Bianca Schoen-Phelan. "Volunteered and crowdsourced geographic information: the OpenStreetMap project". In: *Journal of Spatial Information Science* 20 (2020), pp. 65–70. ISSN: 1948-660X.

[3] Filip Biljecki, Yoong Shin Chow, and Kay Lee. "Quality of crowdsourced geospatial building information: A global assessment of OpenStreetMap attributes". In: *Building and Environment* 237 (2023), p. 110295. ISSN: 0360-1323.

[4] Jonathan Bright, Stefano De Sabbata, and Sumin Lee. "Geodemographic biases in crowdsourced knowledge websites: Do neighbours fill in the blanks?" In: *GeoJournal* 83.3 (2018), pp. 427–440. ISSN: 03432521, 15729893. URL: http://www.jstor.org/stable/45118661.

[5] Nama Raj Budhathoki. "Participants' motivations to contribute geographic information in an online community". PhD thesis. University of Illinois at Urbana-Champaign, 2010.

[6] Noam Cohen. "Define Gender Gap? Look Up Wikipedia's Contributor List". In: *The New York Times* (Jan. 30, 2011). URL: https://www.nytimes.com/2011/01/31/business/media/31link.html?smid=url-share (visited on 08/28/2023).

[7] Alexis Comber et al. "Crowdsourcing: It Matters Who the Crowd Are. The Impacts of between Group Variations in Recording Land Cover". English. In: *PloS one* 11.7 (2016), e0158329–e0158329.

[8] Maitraye Das, Brent Hecht, and Darren Gergle. "The gendered geography of contributions to OpenStreetMap: Complexities in self-focus bias". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–14.

[9] Alexander Dunkel. "Visualizing the perceived environment using crowdsourced photo geodata". In: *Landscape and urban planning* 142 (2015), pp. 173–186. ISSN: 0169-2046.

[10] Esri. *World Countries Generalized*. https://hub.arcgis.com/datasets/esri::world-countries-generalized/. 2022.

[11] Zoe Gardner and Peter Mooney. "Investigating gender differences in OpenStreetMap activities in Malawi: a small case-study". In: *AGILE, Lund, Sweden*. 2018.

[12] Zoe Gardner et al. "Gender differences in OpenStreetMap contributor activity, editing and tagging behaviour". In: *25th Annual Geographical Information Science Research UK (GISRUK), Manchester, United Kingdom* (2017), p. 6.

[13] Zoe Gardner et al. "Quantifying gendered participation in OpenStreetMap: responding to theories of female (under) representation in crowdsourced mapping". In: *GeoJournal* 85 (2020), pp. 1603–1620. ISSN: 0343-2521.

[14] Michael F Goodchild. "Citizens as sensors: the world of volunteered geography". In: *GeoJournal* 69 (2007), pp. 211–221. ISSN: 0343-2521.

[15] Müslüm Hacar. "Analyzing the behaviors of OpenStreetMap volunteers in mapping building polygons using a machine learning approach". In: *ISPRS International Journal of Geo-Information* 11.1 (2022), p. 70. ISSN: 2220-9964.

[16] Mordechai Haklay. "How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets". In: *Environment and planning B: Planning and design* 37.4 (2010), pp. 682–703. ISSN: 0265-8135.

[17] Mordechai (Muki) Haklay. "Why is participation inequality important?" In: *European Handbook of Crowdsourced Geographic Information*. Ubiquity Press, 2016.

[18] Mokter Hossain and Ilkka Kauranen. "Crowdsourcing: a comprehensive literature review". In: *Strategic Outsourcing: An International Journal* 8.1 (2015), pp. 2–22. ISSN: 1753-8297.

[19] P. R. Hunter and M. A. Gaston. "Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity". English. In: *Journal of Clinical Microbiology* 26.11 (1988). Place: Washington, DC, Publisher: American Society for Microbiology, pp. 2465–2466. ISSN: 0095-1137. DOI: 10.1128/JCM.26.11.2465-2466.1988. URL: https://go.exlibris.link/lX4JS0v6.

[20] Wen Lin. "Revealing the making of OpenStreetMap: A limited account". In: *The Canadian Geographer/Le Géographe canadien* 59.1 (2015), pp. 69–81. ISSN: 0008-3658.

[21] Mary L. McHugh. "The chi-square test of independence". English. In: *Biochemia medica* 23.2 (2013). Place: Croatia, Publisher: Medicinska naklada, pp. 143–149. ISSN: 1330-0962;1846-7482; DOI: 10.11613/BM.2013.018. URL: https://go.exlibris.link/3TtfMK7m.

[22] Pascal Neis. *How did you contribute to OpenStreetMap?* Online; data downloaded 21-May-2023. 2023. URL: https://hdyc.neis-one.org/.

[23] Pascal Neis. *OSMstats: Registered OSM members*. Online; accessed 28-August-2023. 2023. URL: https://osmstats.neis-one.org/?item=members.

[24] Pascal Neis and Alexander Zipf. "Analyzing the Contributor Activity of a Volunteered Geographic Information Project — The Case of OpenStreetMap". In: *ISPRS International Journal of Geo-Information* 1.2 (2012), pp. 146–165. ISSN: 2220-9964. DOI: 10.3390/ijgi1020146. URL: https://www.mdpi.com/2220-9964/1/2/146.

[25] Manuela Schmidt and Silvia Klettner. "Gender and experience-related motivators for contributing to openstreetmap". In: *International workshop on action and interaction in volunteered geographic information (ACTIVITY)*, pp. 13–18.

[26] Sukhjit Singh Sehra, Jaiteg Singh, and Hardeep Singh Rai. "Assessment of OpenStreetMap data- A review". In: *International Journal of Computer Applications* 76 (2013), pp. 17–20.

[27] Hyesop Shin and Ana Basiri. "Geographic Biases in OSM Contributions: How do the Geographic Extent of Contributions Differ Among Demographic Groups?"

In: *30th Annual Geographical Information Science Research UK (GISRUK), Liverpool, United Kingdom*. 2022.

[28] Renate Steinmann et al. “Gender Dimensions in UGC and VGI: A Desk-Based Study”. In: $GI_Forum, Volume 1$. 2013.

[29] Monica Stephens. “Gender and the GeoWeb: divisions in the production of user-generated cartographic information”. In: *GeoJournal* 78 (2013), pp. 981–996. ISSN: 0343-2521.

[30] OpenStreetMap Wiki. *API v0.6 — OpenStreetMap Wiki*. Online; accessed 29-August-2023. 2023. URL: https://wiki.openstreetmap.org/w/index.php?title=API_v0.6&oldid=2589236.

[31] OpenStreetMap Wiki. *Elements — OpenStreetMap Wiki*. Online; accessed 28-August-2023. 2023. URL: https://wiki.openstreetmap.org/w/index.php?title=Elements&oldid=2582681.

[32] Harry Wood. *The Long Tail of OpenStreetMap*. Web Page. 2014. URL: https://harrywood.co.uk/blog/2014/11/17/the-long-tail-of-openstreetmap/.

[33] Yingwei Yan et al. “Volunteered geographic information research in the first decade: a narrative review of selected journal articles in GIScience”. In: *International Journal of Geographical Information Science* 34.9 (2020), pp. 1765–1791. DOI: 10.1080/13658816.2020.1730848. URL: https://doi.org/10.1080/13658816.2020.1730848.

Received 30 August 2023; revised 13 October 2023