Geographic Biases in OSM Contributions: How do the Geographic Extent of Contributions Differ among Demographic groups?

Hyesop Shin*[1] and Ana Basiri[1]

[1]School of Geographical and Earth Sciences, University of Glasgow, G12 8QQ, UK

Abstract

OpenStreetMap (OSM) is one of the most successful participatory mapping platforms for creating and editing geographic data. Despite being technically open and available to anyone to contribute, there is a significant demographic participation bias in the contributors of OSM, particularly from their spatial patterns on OSM. This study presents how geo-demographic biases of OSM contributions can be measured using the users' 'number of contributed countries' and their 'changesets'. We found that working-age male participants have a larger geographic extent of entries compared to their female counterparts. However, this once again varied significantly by the age groups. Both variables were employed as proxies to estimate the individual has a propensity to contribute locally or internationally. Future studies could add temporal aspects to compare the temporal patterns between demographic groups to give a multi-dimensional insight for VGI studies.

Keywords: OpenStreetMap, Spatial Bias, VGI, Simpson's diversity

1 Introduction

OpenStreetMap (OSM), as one of the most successful Volunteer Geographic Information (VGI) projects, has grown both in terms of the number of contributions and the number of contributing users, with 8 million contributors and 8 billion features (Neis, 2021). Despite such rapid growth in popularity and a large number of contributors, studies have questioned the representativeness of the contributors and the notion of the term "crowd". Most contributions in OSM are made by a demographically biased sample of crowds, who are most likely to be featured as the young, digitally educated, and male (Stephens, 2013; Gardner and Mooney, 2018; Das et al., 2019; Gardner et al., 2020).

While having access to geo-demographic characteristics of the VGI contributors are becoming openly available through apps such as Strava and Social Media platforms, these platforms have only identified the distinctions of temporal patterns between demographic groups but lacked the spatial information (Livingston et al., 2021). Contributing more internationally or domestically can be a possible representation to understand geospatial biases from the VGI literature.

*Email: `Hyesop.Shin@glasgow.ac.uk`

As a preliminary study, our study has two main objectives: 1) to explore the number of contributed countries to OSM between gender and age groups, 2) to quantify the diversity of OSM contributions between gender and age groups.

2 Methods

2.1 Data Collection and Cleaning

We collected two sets of data of which one is the participants' demographic information, and the other is the OSM features of the participants (Table 1). The demographic data were provided by Gardner et al. (2020). The variables include userID, gender, and age. The OSM features were retrieved from an open-source webpage 'How Did You Contribute to OpenStreetMap' (Neis, 2015). Each individual's information was cleaned by an online JavaScript parser then downloaded as a json file. The boundary box was assumed to be the length of the x-axis[1]. The number of countries is the observations of the countries that the user has contributed. Note the numbers of changesets are not considered for this study.

Table 1: List of Variables

Type	Variable	Description
Survey	userID	ID used in OpenStreetMap
	Gender	Men/Women
	Age	20s, 30s, 40s, >50
OSM dataset	userID	ID used in OpenStreetMap
	No. of Countries	Number of countries contributed by users
	No. of Changesets	Number of changesets made by users

We excluded the participants who contributed fewer than 5 countries contemplating those limited activities of participants may polarise the data even further i.e. 90-9-1 rule. This gave us a final selection of 218 participants, where 30 were female and 189 were male, see Table 2.

Table 2: Demographic structure of the participants

Age	Female	Male	Age Subtotal
20s	15	41	56
30s	10	62	73
40s	3	45	48
Over 50s	2	41	43
Gender Subtotal	30 (13%)	189 (87%)	219 (100%)

The number of countries were contributed by 219 OSM users (see Figure 1). Amongst 263 countries, the UK, the US, and Germany accounted for the highest proportion at around 20% respectively.

Breaking down the number of countries by the users' demographic groups, the average contributions in all male groups were around 40-60 countries. However, as shown in Figure 2, there is a wide gap

[1]https://resultmaps.neis-one.org/newestosmcreatefeed

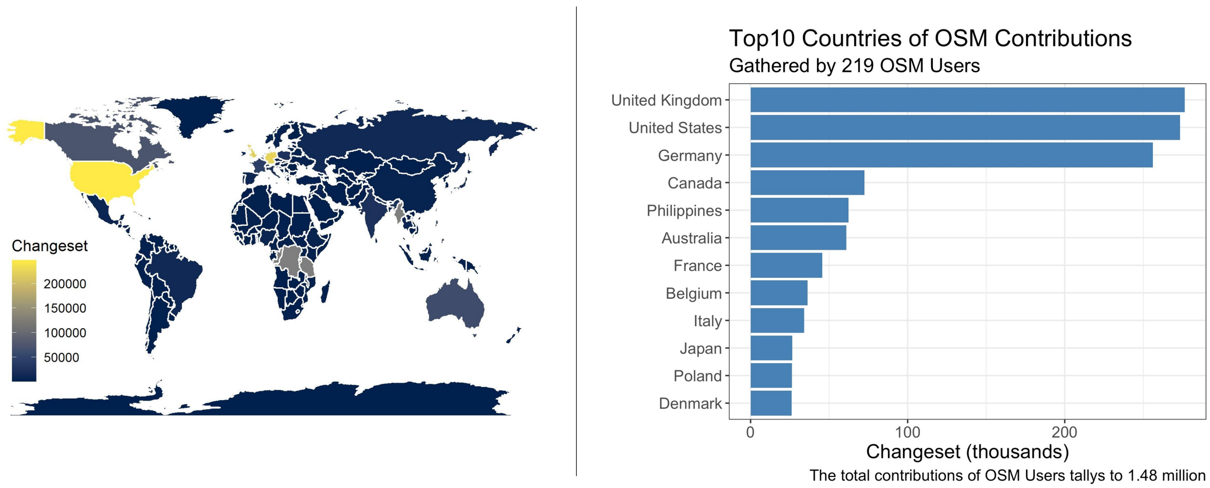

Figure 1: Top 10 countries contributed by the OSM users illustrated with a map and chart

in the distribution, where the tail is long towards the top of the plotting area, which reflects that there are a few extreme contributors for each age group. A similar polarised distribution was seen in the women in the 20s' group. The average was around 40 countries in the women's 20s', 40s', and over 50s' while only 18 in the 30s'.

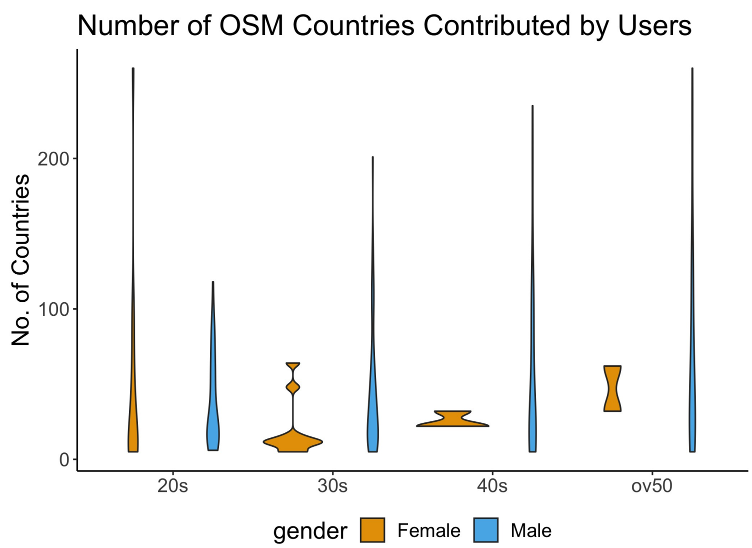

Figure 2: Distribution of the number of contributed countries by gender age groups

2.2 Measuring Spatial Diversity

This study uses Gini-Simpson Index (GSI) to quantify the level of spatial diversity of each participant. The idea of Simpson's index is that if we randomly select two individuals (here two changesets in OSM) and when replaced, they will represent the same type (same country in our context).

We employed this method because most of the users have contributed to less than 50 countries in total, with some concentrated in certain countries, although the sample contained a superuser who contributed to 216 countries. Thus, rather than devaluing the weight of the few countries to which they have contributed from the perspective of Shannon's Index (Gauvin et al., 2020), we used the idea of Simpson's index to allow the weighting of the dominance. The equation is described as follows

$$D = \frac{\sum n(n-1))}{N(N-1)} \tag{1}$$

where n denotes the number of contributions for a single country, and N denotes the number of contributions for the total countries contributed. The index ranges between 0 and 1. In the equation, the lower D value indicates more diversity while the higher value indicates less diversity, which does not sound intuitive (Daly et al., 2018). Instead, we reframe the equation, namely the Gini-Simpson diversity index, to complement the confusion, representing 1 as more diverse.

$$D' = 1 - D \tag{2}$$

3 Results

3.1 Spatial Diversity

To measure biases in spatial preferences between contributor groups, the gendered average of the Gini-Simpson Index (GSI) was 0.5 for the female users and 0.67 for the male users (see Figure 3). This means that male participants had more variability in regard to contributing to more countries and wide distribution than their female counterparts. However, two peaks of diversity patterns of the female users mean that the spatial preference within the group contrasts significantly. Meanwhile, male participants show a right-skewed distribution, implying that more men tend to create/edit more international areas.

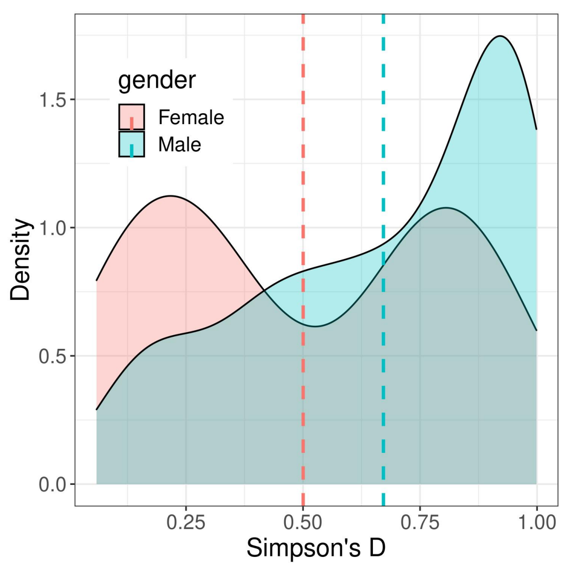

Figure 3: Simpson's Gini-Index Distribution between gender groups

Decomposing the gender groups by age (see Figure 4), male participants have systematically scored higher diversity measures across both age groups, and the average scores are less likely to differ between age groups. The scores from female participants are constantly lower than that of males, and it varies significantly by age group. This can infer those female participants have less spatially diverse contributions than their male counterparts. Also, women in the 40s' and 50s' have a higher deviation than the other groups due to a small number of participants.

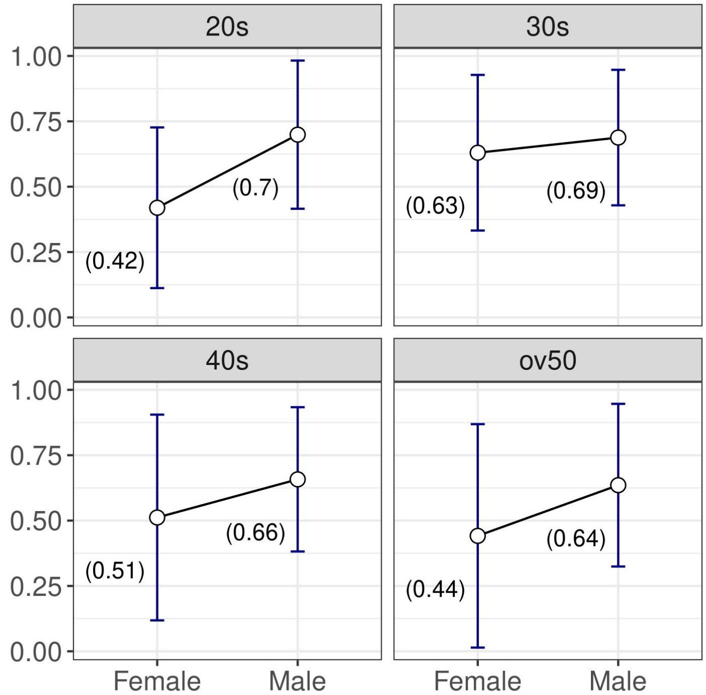

Figure 4: Simpson's Gini-Index by gender and age groups

4 Conclusion

This paper examined the geo-demographic biases using the spatially related features of OSM. We employed the number of countries and changesets contributed by users as proxies to estimate the individual has a propensity to contribute locally or internationally. We found that men in 20s' to over 50s' groups and women in the 20s' group had polarised contributions where the averages were at 40-50 countries but have extreme users who contribute to over 150 countries. Also, nearly 60% of the participants made changes to the UK, the US, and Germany. The Simpson's Gini-Index between the gender and age groups showed a score of 0.50 for women and 0.67 for men, meaning that men were more likely to be spatial diverse than their female counterparts. However, this once again varied significantly by the age groups. For future work, we could add temporal aspects to compare the temporal patterns between demographic groups to give a multi-dimensional insight for OSM studies.

5 Acknowledgements

The work presented in this paper has been funded by the UK Research and Innovation (UKRI) Future Leaders Fellowship MR/S01795X/2. We truly appreciate Dr. Zoe Gardner for providing the survey data.

References

A. J. Daly, J. M. Baetens, and B. De Baets. Ecological diversity: measuring the unmeasurable. *Mathematics*, 6(7):119, 2018.

M. Das, B. Hecht, and D. 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*, pages 1–14, 2019.

Z. Gardner and P. Mooney. Investigating gender differences in OpenStreetMap activities in Malawi: a small case-study. *AGILE, Lund, Sweden*, pages 17–20, 2018.

Z. Gardner, P. Mooney, S. De Sabbata, and L. Dowthwaite. Quantifying gendered participation in OpenStreetMap: responding to theories of female (under) representation in crowdsourced mapping. *GeoJournal*, 85(6):1603–1620, 2020. ISSN 1572-9893. doi: 10.1007/s10708-019-10035-z. URL `https://doi.org/10.1007/s10708-019-10035-z`.

L. Gauvin, M. Tizzoni, S. Piaggesi, A. Young, N. Adler, S. Verhulst, L. Ferres, and C. Cattuto. Gender gaps in urban mobility. *Humanities and Social Sciences Communications*, 7(1):1–13, 2020. ISSN 2662-9992.

M. Livingston, D. McArthur, J. Hong, and K. English. Predicting cycling volumes using crowd-sourced activity data. *Environment and Planning B: Urban Analytics and City Science*, 48(5): 1228–1244, 2021. ISSN 2399-8083.

P. Neis. How did you contribute to OpenStreetMap. *[Online]*, 2015.

P. Neis. OSMstats - Statistics of the free wiki world map, 2021. URL `https://osmstats.neis-one.org/?item=members`.

M. Stephens. Gender and the GeoWeb: divisions in the production of user-generated cartographic information. *GeoJournal*, 78(6):981–996, dec 2013. ISSN 0343-2521. doi: 10.1007/s10708-013-9492-z. URL `http://link.springer.com/10.1007/s10708-013-9492-z`.