Investigating gender differences in OpenStreetMap activities in Malawi: a small case-study

Zoe Gardner
University of Nottingham
School of Geography
University Park
Nottingham, UK
zoe.gardner@nottingham.ac.uk

Peter Mooney
Maynooth University
Dept. of Computer Science
Maynooth, Ireland
peter.mooney@nuim.ie

Abstract

This paper presents results from a selected sample of OpenStreetMap (OSM) contributors to the Malawian OSM dataset in response to a need to measure and understand the impact of participation biases in VGI, specifically by gender. Empirical evidence of the causal relationship between data quality and contribution modes in Volunteered Geographic Information (VGI) and demographics are required. For the purposes of a small case-study, from demographic data from a recent survey of globally distributed OSM users, Malawi was chosen as a dataset for the analysis of gendered mapping behaviours. The results are reflective of gender trends for the larger global OSM dataset which show that men are significantly more active than women in the volume of contributions made and number of days active, as well as demonstrating higher numbers of edits for the modification of existing data. Many of the edits in Malawi, for both male and female contributors, are the result of targeted mapping initiatives such as those for Humanitarian OSM (HOTOSM). Despite the reported increase of participation of women in these activities we found that the prescriptive process of targeted mapping tasks, impacts on the tagging results for both genders who demonstrate higher levels of tagging in the prescribed HOT task categories. Men are also more likely to label features and demonstrated a greater focus on geometric accuracy. Whilst this case-study is small, these results contribute to the discussion of participation biases in VGI and respond with a quantifiable impact of gendered mapping practices, showing that some local patterns of gendered editing are replicated at the global level.
Keywords: OpenStreetMap, OSM, Gender differences, Humanitarian OpenStreetMap, HOTOSM, volunteered geographic information, VGI

1 Introduction

1.1 Background: Gender and VGI

Within broader debates concerning data quality in the evolution in cartographic production, gender dimensions in volunteered geographic information (VGI) have been the subject of a modest but important research focus since its inception in mid-2000's. As the most successful and prolific example of VGI, this has often taken place in the context of the online mapping project OpenStreetMap (OSM) which currently registers 4.2million users. However, a strong male participation bias has been repeatedly observed (Budhathoki, 2010; Schmidt and Klettner, 2013; Stephens, 2013; Gardner *et al*. 2018). Scholars have taken a critical GIS approach to problematize this status quo on the grounds of a failure of crowdsourced mapping projects to represent the interests of the wider 'crowd', specifically that the interests of women are excluded by this ostensibly 'democratising' force in geospatial production (Elwood, 2010; Haklay, 2010; Leszczynski and Elwood, 2015;). These discourses of gender and VGI propose that the crowdsourced map is a reflection of the geospatial interests of the people that create it and, given the participation bias, that the interests of women are repeatedly excluded by the process. Based on observations of the gender participation, both within the OSM community and the wider VGI community calls have been made for the increased participation of women, a drive which often constitutes the focus of forum threads as well as the raison d'être of some groups. For example, the 'Geochicas' group based in South America, specifically aims to raise the contribution profiles of women as well as explore barriers to their participation in OSM[1]. However, these assertions appear to continue with no obvious empirical grounding with regard to the impact of the gender bias on the mapping of topographical features or the added value that democratising participation would bring to the vast geospatial database.

Extending an earlier analysis of gendered mapping behaviours (see Gardner *et al*. 2018), this study examines the specific differences in editing and detailed tagging behaviours in a specific geographical locale. As a small study it informs the potential of a larger analysis to provide empirical grounds on which a clear rationale for widening participation in OSM might be formed. Although previous studies have attempted to explore the relationship between demographics and VGI quality by inferring demographic characteristics on to the local mapping community (see Mullen *et al*. 2014) this study is unique in its attribution of specific user demographics to OSM edits. It also distinguishes itself from other analyses of OSM contributor survey data which have not linked profiles to mapping behaviours (for example Budhathoki, 2010).

1.2 The Malawi dataset

One response to the need to measure the value of gendered mapping is to compare the individual edits of

[1] See https://twitter.com/GeoChicas

men and women at a local (in this case national) level. To do this, the national OSM dataset for Malawi was selected. Informed by the global survey, which provided gender data on a specific subset of OSM contributors, 20 users known to have edited in Malawi were selected for this analysis. Given the relatively small size and *immaturity* of the Malawian dataset, it served as a 'blank canvas', enabling analysis of new 'creations' of data as well as modifications to existing edits. Malawi is also often the geographical focus of targeted mapping activities such as Humanitarian OSM (HOTOSM) tasks, which also provided the opportunity to explore the role of gender in this context and which were relevant to the results (see 4.2).

2 Methodology

The findings presented here resulted from a three-stage methodological approach: 1) an online demographic survey of OSM users; 2) the identification of a sample set of contributors (informed by the aforementioned survey data); and 3) an extraction of the Malawian OSM file from which the selected user's edits were extracted.

2.1 Global survey of OSM users

An online survey collected responses from OSM contributors over a 4-week period in August 2017. In addition to five demographic indicators (gender, age, educational background, location and nationality), participants were asked to provide their OSM username. This subsequently enabled each indicator to be directly assigned to their edits contained in the OSM Full History file. It also facilitated access to a range of further user information openly accessible via the 'how did you contribute to OSM' (HDYC) page curated by OSM affiliate Pascal Neis[2]. This data includes the range of countries' to which users have contributed as well the number of changesets (groups of edits) made to each country.

Using the HYDC data the national OSM datasets to which users had made edits were recorded. From here, a common country to which users had contributed at least 50 edits was selected. A further criterion, given the volume of data contained in each changeset, was the potential file size of the eventual dataset. As such, a small country was sought and Malawi was selected as the focus of the study. From this sample of users, the edits of the 10 male and 10 female with the most edits in Malawi were selected. Our choice for Malawi was also influenced by the need to find a country/region where contributors to OSM had the opportunity to supply a wide variety of new objects but also edit existing data.

2.2 Collecting individual edits

From the GeoFabrik OSM data download service[3] the OSM History file of all edits for Malawi was downloaded in Feburary 2018. All edits to points, lines and polygons (collectively called *ways* in OSM) were extracted from this history file and stored in a database. The file contains the characteristics of each edit such as its timestamp, tags applied to the object and changes to the object's geometry. This database subset allowed us to analyse the characteristics of the edits performed by the specific 20 users.

2.3 Statistical approach

For both gender groups, the full survey data revealed that participation (as measured by the total number of changesets and the number of days active) mimics the 'long tail effect' distribution of participation characteristic of crowdsourcing projects more widely (see Haklay, 2016). That is, that a small proportion of the respondents conduct the greatest proportion of the activity. Given this skewed distribution, non-parametric statistical tests were applied to the data to identify differences in mapping behaviours between genders.

2.4 Global survey: population characteristics

The survey garnered 326 responses, 29 of which were excluded on the grounds of duplication, invalid username, sabotage or absence of edits. Of the remaining 297 responses, 39 were from women and 258 from men. The median age groups were 35-39 for men and 25-29 for women. This differential in age may be reflective of recent efforts to encourage female engagement in Science, Technology, Engineering and Maths (STEM) activities (i.e. women in the older age groups may be less likely to have the skills/knowledge to contribute than their male counterparts where there is a longer tradition of participation in these interests). Both groups reflected high levels of educational attainment. The greatest proportion of male respondents were educated to degree (73%) or post-graduate (21%) level. For women these figures increased to 95% and 54% respectively. Unpublished survey data finds that women often experience hostile situations in male-dominated online mapping communities[4]. Thus, these figures might be interpreted as a reflection of the self-efficacy women require to participate in what is perceived as a male- dominated activity, an assertion supported by earlier studies (Steinmann *et al.*, 2013; Schmidt *et al.*, 2013). These population demographics repeat previous profiling results of OSM contributors which have collectively found a bias towards young, tech-savvy, men (Budhathoki and Haythornthwaite, 2013; Coleman, 2009; Stephens, 2013; and Schmidt and Klettner, 2013).

[2] http://www.hdyc.neis-one.org/

[3] http://download.geofabrik.de/africa/malawi.html

[4] Email correspondence: Yang, S., 2017, OSM and Gender Inclusion

Table 1: Raw editing data

User	Gender	Malawi Rank	Total Objects Edited	Total Changesets	Total Creates			Total mapping Days
					Nodes	*Ways*	*Total*	
1	Female	149	1627	122	0	1562	1562	16
2	Female	185	1283	113	0	910	910	18
3	Female	280	903	90	0	885	885	13
4	Female	779	365	40	0	364	364	8
5	Female	813	354	2	0	315	315	1
6	Female	1580	210	41	0	207	207	5
7	Female	1651	203	34	0	156	156	5
8	Female	6015	28	4	0	15	15	1
9	Female	6177	25	15	0	22	22	1
10	Female	7457	9	1	0	9	9	1
11	Male	2	39691	935	5	2069	2074	127
12	Male	47	5003	145	380	3175	3555	63
13	Male	64	3403	407	1	2661	2662	57
14	Male	78	3074	116	51	1124	1175	20
15	Male	85	2936	51	3	2807	2810	19
16	Male	117	2090	17	13	2070	2083	2
17	Male	165	1448	99	2	1343	1345	18
18	Male	247	975	70	199	458	657	10
19	Male	1584	210	23	0	132	132	7
20	Male	7301	11	5	0	6	6	1

Source: OSM History File (Malawi)

3 Gender differences in editing and tagging in Malawi

3.1 Demographic and editing profiles

The population of the Malawi sample (10 men and 10 women) mimicked the demographic characteristics of the global sample (outlined in 2.4) in age and educational background. The median age of women was 25-29 and each was educated to university degree level or above. The median age of the male group was 40-44 with 80% educated to at least degree level.

At the time of analysis, 8921 users had contributed to the Malawian dataset. Table 1 shows the range of ranks for the users in the sample demonstrates (as the ranks cover almost the entire range) that the sample is representative of the wider group of contributors to the Malawian database. Just over 36% of all contributors to the dataset have made over 100 edits. Of the sample 20 users, 16 feature in this group with just under half as women (43%). However, narrowing the focus further to consider those with 1000 edits or more (which constitutes the top 2% of contributors to Malawi), 9 (45%) of the sample feature in this group: 78% of these men and 22% women.

Although these results generally mimic the results for the global survey, in that men are significantly more active than their female counterparts (Gardner *et al*. 2018) there is an increased proportion of female users in the group of top editors to the Malawian dataset. This is discussed in 4.2.

3.2 Gender differences in editing

The results for comparative volumes of activity again replicate those observed for the global dataset, which found that men are statistically significantly more active than their female counterparts in the range of editing activities included in the analysis. Median values for the total number of objects edited in the Malawi dataset reveal that men have conducted statistically more (2090 compared to 282) than the female cohort at a 0.05 level of significance. The highest ranked male editor tallied over 24 times more 'total objects edited' than the highest ranked female editor (39691 compared to 1627. See Table 1). The same two users were separated by almost 8 times as many mapping days devoted to the Malawian dataset (127 compared to 16) which also reveals differences (although not statistically significant) in behaviour when this is measured by volume of activity. Statistically significant differences *were* observed however in the total number of tagged nodes created 'Total Creates: Nodes'. This is likely due to the fact that notably, none of the female cohort have created a single tagged node. In terms of creating new data (rather than modifying or correcting existing edits), the highest ranked male editor in this category has created over twice as many as the highest ranked female in this sample (3555 compared to 1562). See 3.3.

Statistically significant gender differences were also revealed in the *modification* of data. This was recorded in the editing of ways (see Table 2, 'Bdg Mod.') and three additional variables which demonstrated a male focus on geometry and

Table 2: Raw OSM data (tagging)

User	Gender	Bdg Create	Bdg Mod.	LU Create	LU Mod	Highway Create	Highway Mod.	Name Create	Name Mod.
1	Female	1478	15	0	0	81	47	0	0
2	Female	894	255	0	1	14	113	0	3
3	Female	882	9	0	**0**	3	9	0	0
4	Female	364	1	0	**0**	0	0	0	0
5	Female	210	**0**	39	**0**	47	38	2	3
6	Female	187	**0**	0	1	0	2	0	1
7	Female	0	**0**	0	**0**	147	45	0	2
8	Female	5	**0**	4	3	6	9	0	0
9	Female	7	**0**	0	**0**	13	3	0	1
10	Female	9	**0**	0	**0**	0	0	0	0
11	Male	337	7633	903	246	819	29329	0	335
12	Male	2210	257	164	154	460	725	331	247
13	Male	2467	483	0	14	190	227	1	18
14	Male	259	199	398	98	339	1277	0	9
15	Male	2720	35	1	6	76	73	0	7
16	Male	1936	0	16	0	31	6	19	3
17	Male	1322	61	0	0	16	34	0	0
18	Male	78	5	111	9	180	230	3	13
19	Male	124	68	0	2	8	6	0	1
20	Male	0	0	0	0	6	5	0	0

Source: OSM History File (Malawi)
Key: Bdg = building LU = landuse; Mod. = modified

Table 3: Raw OSM data (editing - modifying)

User	Gender	Geometry + tags	Geometry	Tags	Modify
1	Female	149	1627	122	16
2	Female	185	1283	113	18
3	Female	280	903	90	13
4	Female	779	365	40	8
5	Female	813	354	2	1
6	Female	1580	210	41	5
7	Female	1651	203	34	5
8	Female	6015	28	4	1
9	Female	6177	25	15	1
10	Female	7457	9	1	1
11	Male	2	39691	935	127
12	Male	47	5003	145	63
13	Male	64	3403	407	57
14	Male	78	3074	116	20
15	Male	85	2936	51	19
16	Male	117	2090	17	2
17	Male	165	1448	99	18
18	Male	247	975	70	10
19	Male	1584	210	23	7
20	Male	7301	11	5	1

Source: OSM History File (Malawi)

Key: *Geometry + Tags*: addition or removal of nodes and changes to the tags (added tags, deleted tags or changed the values of tag keys); *Geometry*: added or removed nodes only; *Tags*: changes to tags only; *Modify*: geometrical changes to an object but no additions or removal of nodes (e.g. positional)

an increased male focus on and potentially (if the data is accurate) instrumental role in the positional and 'shape' accuracy of nodes and ways. As discussed in section 2, there are several factors which may contribute to this result including the role of technical skills and knowledge, which have been proven to act as barriers to participation.

3.3 Gender differences in tagging

The results of the tagging data reveal substantially less activity than editing. As shown in Table 1, several of the tagging categories, particularly 'Landuse' (LU-C and LU-M) and 'Name' (N-C and N-M) contained no edits at all for several contributors and more evident in the female cohort. This data supports anecdotal evidence which has suggested a reluctance to tag and label in an African OSM context with a tendency instead towards the creation of nodes and ways, an assertion which is supported by this dataset. The Malawian dataset downloaded on 7th February 2018, contains 1,547,393 ways and 61,692 nodes (with tags). The sample users have created 61,287 of these objects, 60,197 of which are ways, which immediately demonstrates an emphasis on way creation (editing) over tagging. Prescriptive, targeted mapping activities (discussed in section 4) may account for the dearth of tagging data in some categories, i.e. Landuse and Name), as they often emphasise the geolocation of topographical features (namely buildings and roads) essential for improved disaster response and mitigation in health and welfare initiatives.

4 Humanitarian OSM in Malawi

As well as in the context of wider gender dimensions in VGI (discussed in section 2), these results can be considered in the context of targeted mapping initiatives, notably Humanitarian OSM (HOTOSM). For all but 3 of the users, their largest changeset was conducted for an HOTOSM task (9 women, 8 men). HOTOSM is a branch of OSM that creates maps for the purposes of supporting both the improved response to and the prevention of humanitarian crises including meeting the Sustainable Development Goals (SDGs). The World Bank classes Malawi as a Low Income country[5] and it features regularly on the tasking manager of HOTOSM. Specific mapping activities are tasked to HOTOSM by partner agencies involved in humanitarian initiatives, e.g. the UN and the Red Cross but also other aid agencies and NGOs.

4.1 The impact of prescriptive targeted mapping

HOTOSM tasks are usually prescriptive: tasks provide instructions on what to edit and how to tag. This might include the use of changeset hashtags, data sources and specific nodes, ways and tags to map. Given the socio- economic characteristics and often remote geographical locations that form the subject of tasks, and where society,

technology and education may preclude the participation of the local community, these tasks are often performed remotely using aerial imagery. The recent HOTOSM report confirms that at least 42% of its members were resident in the Global North (i.e. countries unlikely to be focus of HOTOSM activities) and therefore geographically remote from the subject they are mapping. This by its very nature precludes the attribution of amenity tags such as 'name=' and address tags such as 'house number=' which may not be visible in sources such as aerial imagery. The directed nature of HOTOSM tasks described above, may explain the emphasis on the creation of ways, over the addition of tags, particularly those in the land-use and name categories, editing of which may not have been requested by HOTOSM.

4.2 Humanitarian mapping and gender

Of the female cohort, Malawi was the users' highest or second highest edited country for 6 of the 10 contributors whereas this was only the case for 1 of the male users with Malawi featuring as one of often scores of countries mapped for the remaining 9 male editors. The aforementioned HOTOSM study reported a 28% female participation rate, which could be as much as 25% higher than comparative figures for the wider OSM project. Where these activities are targeted through local grants this figure increases to an average of 44% (HOTOSM, 2017). These figures hint at the role both intrinsic reward and personal investment may play in female participation in VGI.

Despite these increased numbers in female participation, women's lack of knowledge, skills, experience or confidence may affect the levels of detail and accuracy they are able to provide as reflected in the tagging results for the Malawian dataset. Earlier studies have found that competing demands on time (including differential caring responsibilities) serve as barriers to participation in VGI (Steinmann *et al*., 2013; Schmidt *et al*., 2013). However, it is important to note that given the status of Malawi as the frequent object of a range of targeted mapping initiatives[6], many of the edits under scrutiny here are the result of a prescribed set of mapping pathways and processes. Although users are free to contribute additional attribute data when carrying out HOTOSM tasks, the limitations of this parameter on the potential to observe differences in mapping behaviours when disaggregated from these prescribed mapping processes is acknowledged.

5 Conclusions and further work

This paper has presented the results of a small study of OSM edits in Malawi which sought to identify differences in mapping behaviours between men and women at a local level. This analysis was performed as a way of further exploring the impact of the male participation bias in OSM as well as contributing to discourses in gender dimensions in VGI more broadly. The results replicate those observed at the global level where men are significantly more active than their female counterparts. Men also demonstrated higher participation rates in the *modification* of data, suggesting an

[5] https://data.worldbank.org/country/malawi

[6] A HOTOSM Task Manager search returned 61 current tasks for Malawi, significantly more than its neighbours or comparatively sized African countries. https://tasks.hotosm.org/contribute?difficulty=ALL&text=Malawi searched 15/2/18

additional focus on the geometric accuracy of data. In Malawi, men are also more involved in tagging. Given the remote nature of the work, this may reveal an increased skill level in both the acquisition and interpretation of data sources enabling a greater degree of tagging contributions. However, these results must be considered in the context of targeted mapping initiatives (often with a humanitarian objective), which are prescriptive in their nature and may therefore mask mapping preferences and behaviours of men and women outside these activities.

The relative weighted participation of women in humanitarian mapping initiatives suggests a potential model for the recruitment of women to the activity more widely. However, the implied role of intrinsic reward suggests that this would need to be incorporated. In addition, the unique barriers to participation that impact on women's editing behaviours may not be removed by such an approach. The results reported here however, are suggestive of a deeper set of nuanced behaviours which will benefit from further investigation.

Replicating this analysis for a similar sized developed country with a more mature OSM database and in an urbanized context would also contextualize the analysis presented here. Research has shown not only a wide disparity in coverage between rural and urban datasets, but also in completeness, evidenced in this set of results.

References

Budhathoki, N. R. (2010) *Participants' motivations to contribute geographic information in an online community*, Urban and Regional Planning, Urbana-Champaign, University of Illinois.

Budhathoki, N. R. and Haythornthwaite, C. (2010) Motivation for open collaboration: crowd and community models and the case of OpenStreetMap, American Behavioural Scientist, 57, 548-575

Coleman, D. J., Georgiadou, Y. and Labonte, J. (2009) Volunteered Geographic Information: the nature and motivation of produsers, *International Journal of Spatial Data Infrastructures, 4, 332-358*

Elwood S (2010). Geographic information science: Emerging research on the societal implications of the geographical web. *Progress in Human Geography*, 34(3), 349-357.

Gardner, Z., Mooney, P., Dowthwaite, L. and Foody, G., (2018), Gender differences in OpenStreetMap contributor activity, editing and tagging behaviour, *GISRUK 2018: Proceedings of the Geographical Information Science Research UK Conference*, 17th-20th April, Leicester

Haklay, M. (2010) Neogeography and the delusion of democratisation, *Environment and Planning A*, 45(1) 55-69

Haklay M (2016). Why is participation inequality important? *European Handbook of Crowdsourced Geographic Information*, C Capineri, M Haklay, H Huang, V Antoniou, J Kettunen, F Ostermann and R Purves, London: Ubiquity Press, 35-44.

Humanitarian OpenStreetMap Team (HOTOSM), (2017), State of the Community Report 2017, HOTOSM, available at https://drive.google.com/file/d/1BskG72I0M1eOSbw6K5QvqpGwPKecJz7l/view

Leszczynski A and Elwood S (2015). Feminist geographies of new special media. *The Canadian Geographer*, 59(1), 12-28.

Mullen, W.F., Jackson, S.P., Croitoru, A., Crooks, A., Stefanidis, A., amd Agouris, P., 2014, Assessing the impact of demographic characteristics on spatial error in volunteered geographic information, *GeoJournal*, 80, 587-605

Schmidt M and Klettner S (2013). Gender and experience related motivators for contributing to OpenStreetMap, *AGILE*, 14-17 May, Leuven, Belgium.

Schmidt M, Klettner S and Steinmann R (2013). Barriers to contributing to VGI projects, *Proceedings of the 26th International Cartographic Conference*, 25-30 August, Dresden, Germany.

Steinmann R, Häusler E, Klettner S, Schmidt M and Lin Y (2013). Gender dimensions in UGC and VGI: A desk-based study, *GIS Forum 2013: Creating the GISociety*, T Jekel, A Car, J Strobl and G Griesebner, Berlin/Offenbach: Herbert Wichmann Verlag, 355-364.

Stephens M (2013). Gender and the GeoWeb: Division in the production of user-generated cartographic information, *GeoJournal*, 78(6), 981-996.