Gender differences in OpenStreetMap contributor activity, editing and tagging behaviour

Z. Gardner[*1], P. Mooney[†2], L. Dowthwaite[‡3] and G. Foody[§4]

[1] Nottingham Geospatial Institute, University of Nottingham
[2] Department of Computer Science, Maynooth University
[3] Horizon Digital Economy Research Institute, University of Nottingham
[4] School of Geography, University of Nottingham

January 12, 2017

Summary

Recognising a need to measure the impact of the male participation bias on crowdsourced mapping, this paper presents some preliminary results from an online survey of globally distributed OpenStreetMap users. The survey enabled access to 293 user's statistics collated on the 'how did you contribute to OSM' wiki page which were then analysed by gender. The results show that men are more active and contribute significantly more data in each category of editing and tagging than their female counterparts. Men are also more likely to *modify* existing edits and contribute to a wider range of tagging categories, with women demonstrating a preference to add new data concentrated in fewer categories.

KEYWORDS: Volunteered Geographic Information, Gender Biases, OpenStreetMap, OSM Crowdsourced Data

1. Background

Six years after the inception of the online mapping project OpenStreetMap (OSM), Budhathoki (2010) identified that 96.2% of contributors were men, hence a major participation bias. This imbalance was later supported by studies specifically exploring gender dimensions in crowdsourced mapping (see Schmidt and Klettner, 2013; Stephens, 2013). Since then, as part of a critical GIS approach, scholars have problematized this status quo on the grounds of a failure of crowdsourced mapping projects to represent the interests of the wider 'crowd' (Elwood, 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. However, these assertions appear to continue with no obvious empirical grounding with regard to the impact of the gender bias on which topographical data is and isn't 'volunteered'. This paper presents some preliminary results from a broader study designed to address this need.

2. Surveying OSM users

In August 2017, an online demographic survey of OSM contributors was conducted. The link to the survey was distributed globally via both the OSM user diaries and five English language talk mailing lists (the survey was written in English as the de facto language of OSM)[**]. Of the 326

[*] Zoe.Gardner@nottingham.ac.uk
[†] Peter.Mooney@mu.ie
[‡] Liz.Dowthwaite@nottingham.ac.uk
[§] Giles.Foody@nottingham.ac.uk
[**] For the OSM diary entry go to https://www.openstreetmap.org/user/Geospa_gal/diary

responses, 33 were excluded on the grounds of duplication, invalid username or user having no edits in OSM. Of the remaining 293 responses, 38 were from women and 255 from men. In addition to five demographic indicators, participants were asked to provide their OSM username. This subsequently enabled each indicator, here gender, to be assigned to a series of tagging and editing contributions, openly accessible via the 'how did you contribute to OSM' (hdyc) wiki page (seventeen variables are included here). In doing so, this study connects respondents with their stated gender and subsequently the summary of the full history of their contributions to OSM. The linking of gender to edits in this way sets the study apart from similar analyses (based on OSM contributor survey data).

Despite the difference in sample size, the median distribution curves for the number of days active and number of changesets for both men and women (as indicators of levels of activity) correspond to the 'long tail effect' distribution of general participation, characteristic of crowdsourcing projects (see Haklay, 2016). That is, that a small proportion of the respondents conduct the greatest proportion of the activity (see figures 1 and 2). This suggested that both sample groups were somewhat representative of the overall population, although there is currently no official means of quantifying this. Median ranking data displayed in Table 1 suggests that the male respondents are amongst some of the highest ranking contributors, whereas female respondents are ranked considerably lower. However, given a community of over 4 million users these rankings can be still deemed relatively high with median values in the top 1%. It is important to state that is not possible to state whether these are *the* most active female users as gender is unknown for the missing ranks. Based on earlier estimates of a 24:1 male to female participation ratio (Budhathoki, 2010) this self-selecting survey would appear to have attracted a greater response from the female OSM cohort with a 13:2 ratio of male to female responses. This could be explained and understood by the focus of the survey, or that the female OSM cohort are more likely to subscribe to mailing lists and message boards, although both of these assertions lead us into speculative territory.

Table 1 Median ranking values

	Men		Women	
	Median	Range	Median	Range
Nodes	5486	20 - 341070	21382	744 -398667
Ways	4696	7 - 270072	18556	657 - 350051
Relations	3762	2 - 277779	24211	665 - 354915

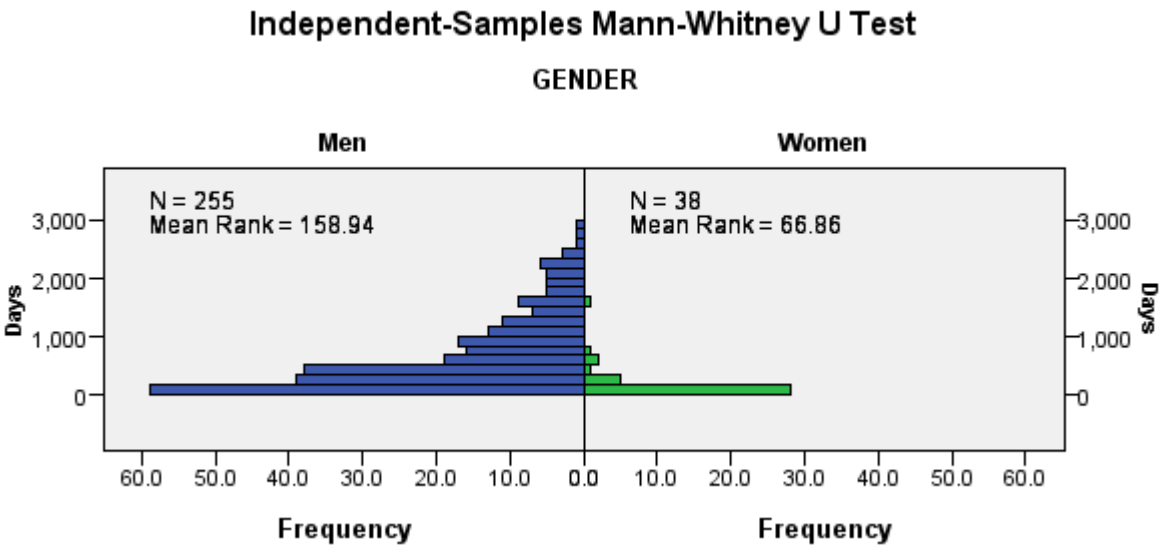

Figure 1 Male and female distribution curves for median 'days active' values

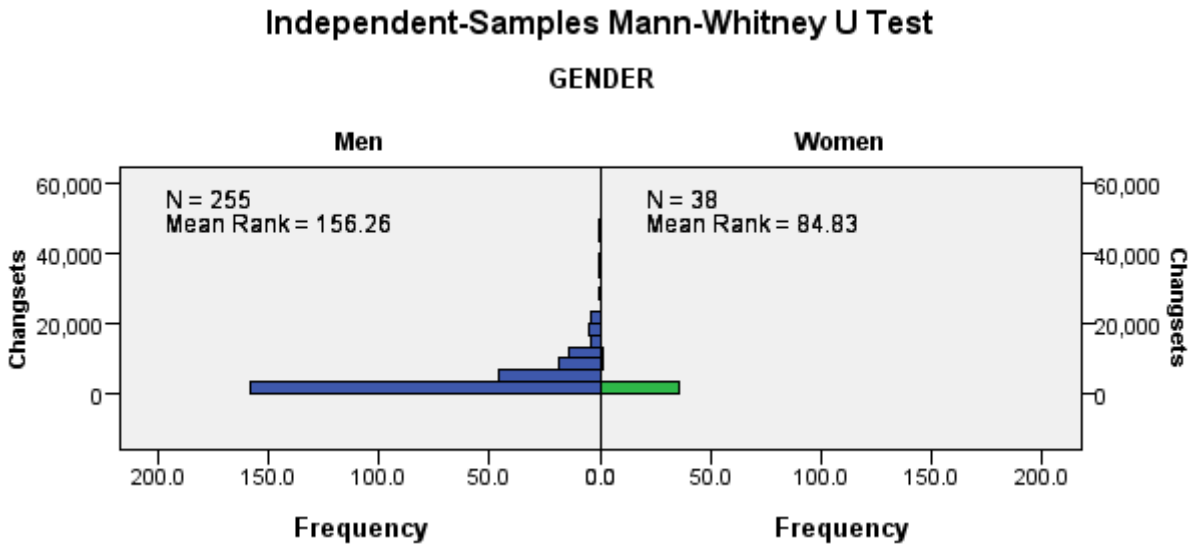

Figure 2 Male and female distribution curves for median 'number of changesets' values

3. Results

The results were broken down in to three different categories of contributor behaviours and actions: activity, editing and tagging. (i) 'Activity' comprised of two statistics: number of days active and overall number of changesets; (ii) 'editing' comprised of 9 separate statistics on users' contributions, deletions and modifications of nodes, ways and relations (i.e. three different 'modes' of editing and three different types of objects edited); and (iii) 'tagging' comprised of the users' statistics on contributions to 8 different labelling categories (Amenity, Building, Highway, Landuse, Leisure, Name, Natural and Address). For each of the seventeen variables the median values were calculated and analysed using a Mann Whitney U test with the independent variable 'gender', which has two levels, male and female.

3.1 Gender differences in Activity

Despite demonstrating similar distribution curves for each of the variables (figures 1 and 2), the median values for the volumes of activity for both category was statistically higher for men at a 0.05 level of significance: Median values show that men have statistically more 'days active' (158.94 compared to 66.86) and demonstrate statistically higher numbers of changesets than their female counterparts (i.e. those that demonstrate the greatest activity amongst the most active *female* cohort of the sample) (156.26 compared to 84.83). These values suggest that on the days that women are active they contribute a greater number of changesets, which may suggest that differences in time management play a role in gendered OSM contributing. Studies which have explored the criteria which influence participation in VGI, find that these include competing demands on time (including the difference in caring responsibilities) and the perceived necessity of advanced GIS skills (Steinmann *et al.*, 2013; Schmidt *et al.*, 2013).

3.2 Gender differences in Editing

A similar phenomenon was observed for the editing variables. In all nine of the categories, the hypothesis that there was no difference between male and female editing was rejected at a 0.05 significance level. Statistically, men made more edits in each category. Again, revealing that for this survey, men are more active as OSM editors of nodes, ways and relations. In addition to this result, some simple data visualisations reveal several similarities. For both groups, 'nodes'

dominated the objects of interest (over ways and relations) with approximately 85% of activity for both groups focused on this object (see figure 3). More broadly of the 9 practices, it was the *creating* of nodes which dominated activity for both groups with around three-quarters of each groups' activity focused on this type of editing. When considering further gender differences in editing preferences, it was the *mode* of editing that demonstrated the most difference between the groups. Men were much more likely to 'modify' objects, whereas women demonstrated higher values for 'creating' them (see figure 4). These results raise a series of questions about the role and influence of the technology and interfaces by which contributions to OSM are made, as well as the influence of contributor knowledge in terms of GIS/mapping/web skills and locale or domain knowledge, all of which require further exploration.

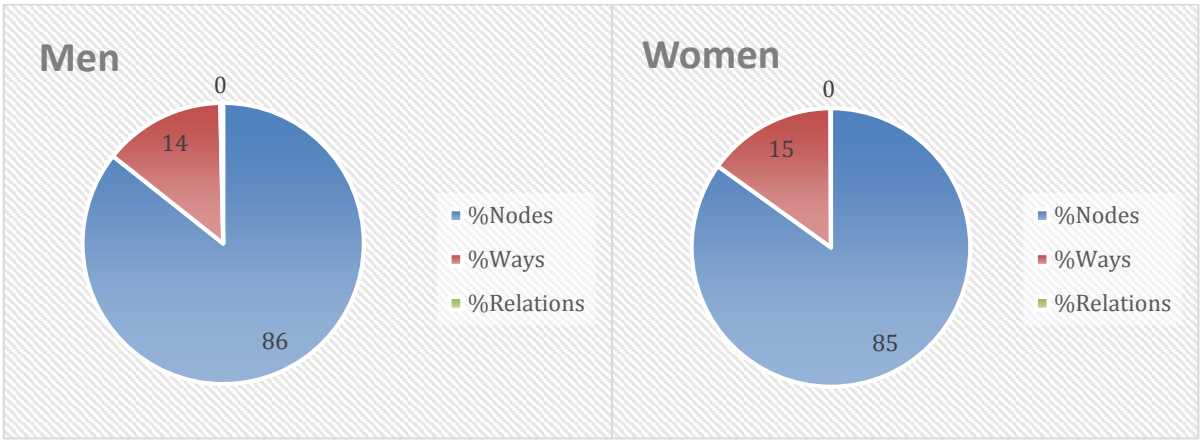

Figure 3 Object focus for male and female editing

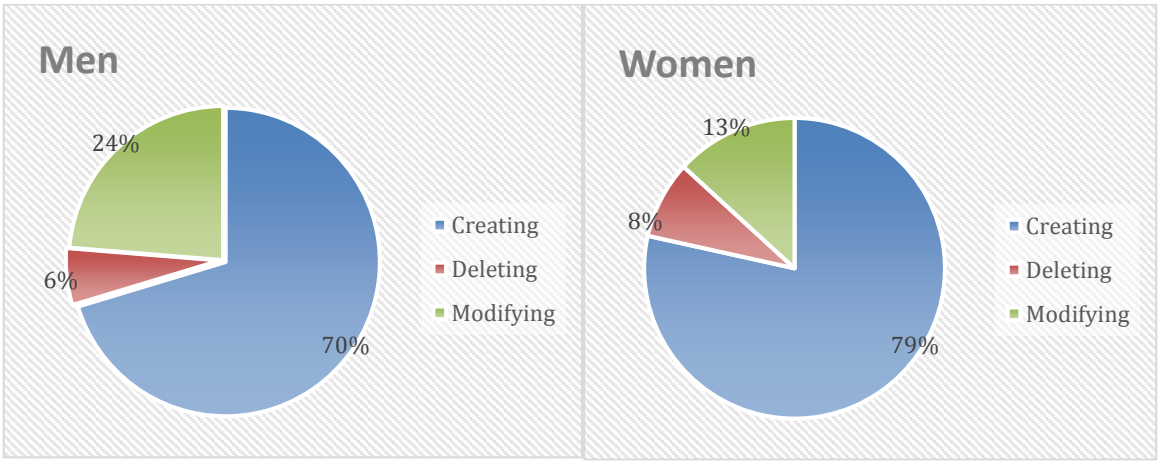

Figure 4 Median values for male and females 'modes' of editing

3.3 Gender differences in Tagging

When exploring to which tagging categories users are more likely to contribute, non-parametric testing revealed again, that men participate in statistically more tagging than women at a 0.05 significance level. However, this merely reinforces the results shown in 3.1, that men contribute significantly more edits overall. When charting the median values for tagging, more variation in the preferences of the two groups was revealed. Women were much more likely to add labels in the 'buildings' category than men (67% for women compared to 35% for men), whereas the greatest volume of men's tagging was in the 'highway' category (39% for men compared to 23% for women) (see figure 5). However, men and women demonstrated similarities in their

preference for these two categories over others: building and highway labelling comprised the majority of their tags at 74% (men) and 91% (women). The six remaining categories constituted only 9% of female tagging, whereas the same proportion for men totalled 26%, demonstrating a higher level of variance in the topographical features that men chose to edit. This could be related to overall activity levels as well as issues related to availability and competing demands on time as suggested in 3.1. It may also suggest that men are more invested in overall levels of coverage rather than mapping particular features, but again, this is speculative and more research is required to qualify such assertions.

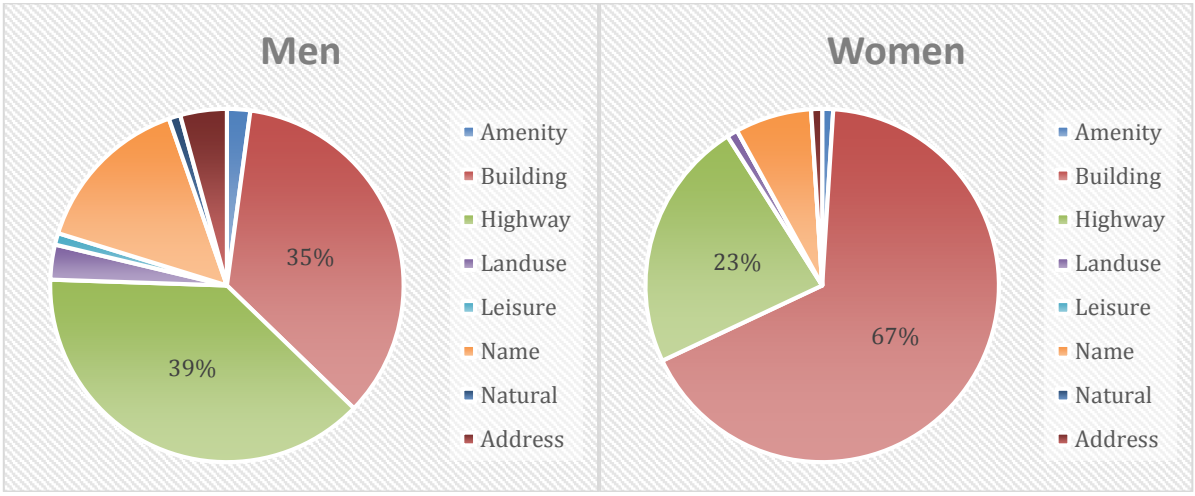

Figure 5 Breakdown of median values for male and female tagging

4. Conclusion and further work

This analysis has considered differences in the way men and women contribute geospatial data to the online mapping platform OSM, according to the statistics collated on the 'hdyc' website. The allocation of gender to real mapping behaviours makes it distinct from existing OSM contributor surveys. As well as demonstrating that men are significantly more active than their female counterparts in terms of the number of days active and volume of contributions made, this study reveals a difference in *modes* of editing, with men demonstrating higher levels of modifying than women. Differences are also observed in gendered preferences for tagging categories for topographical features. Further analysis on these empirically observed variations has been conducted in the context of gendered contributions in humanitarian mapping activities in Malawi revealing stronger male emphases on both geometric accuracy and tagging (Gardner and Mooney, 2018). A planned meta-analysis will extend this work and provide further insights in to the impacts of these phenomena, specifically on variations in types of topographical features mapped and levels of attribute data produced. The analysis here has initiated this process by eliminating broader level behaviours in which differences could be demonstrated. With more extensive analysis, the more specific nuances of gendered geospatial editing may be revealed, thereby allowing an informed and targeted mitigation strategy against the effects of gender bias, which could be applied to demographic biases in participation more widely.

Acknowledgements

This work has been jointly funded by the Engineering and Physical Sciences research Council and the University of Nottingham through a Daphne Jackson Fellowship. The authors are indebted to Pascal Neis for creating the 'How Did You Contribute to OSM' Open Access data and of course the work of millions of OSM contributors, who have created this incredible geospatial database.

Biographies

Zoe Gardner is a Post-Doctoral Researcher at the Nottingham Geospatial Institute and currently holds a Daphne Jackson Research Fellowship. With a research background in Human Geographies her research interests are inter-disciplinary, but fall broadly under the theme of 'Space, Society and Technology' currently focused on gender dimensions in VGI.

Peter Mooney is a tenure track Lecturer in Computer Science in Maynooth University, Ireland. He is interested in a wide range of spatial data handling topics, in particular: volunteered geographic information, quality of crowdsourced geospatial data, data mining approaches for quality analysis and citizen science spatial data generation.

Liz Dowthwaite is a Post-Doctoral Researcher in the Horizon Digital Economy Research Institute at the University of Nottingham. She has a multi-disciplinary research background grounded in Psychology and Human Factors. Her research interests surround attitudes, behaviour and motivation in online crowd systems, including crowdfunding, social media and crowdsourcing.

Giles Foody is Professor of Geographical Information Science at the University of Nottingham.

References

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

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 (forthcoming 2018). Investigating gender differences in OpenStreetMap activities in Malawi: a small case-study. *Proceedings of the AGILE Conference*, 12-15 June, Lund, Sweden

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

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.

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.