Learning from the community: Surveying OSM contributors
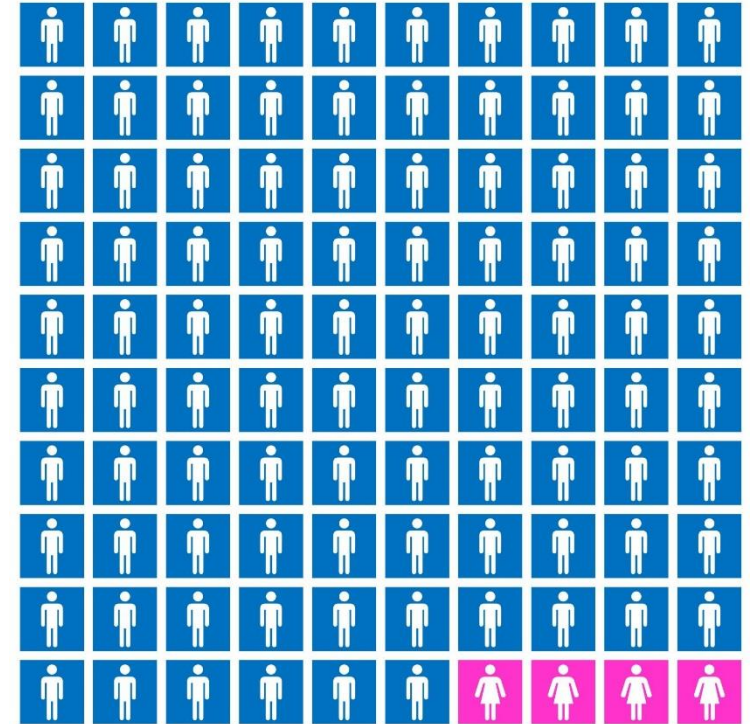

Dr Zoe Gardner
Dr Peter Mooney

28th – 30th July 2018, Milan

STATE OF
THE MAP
Milan, Italy 2018

University of
Nottingham
UK | CHINA | MALAYSIA

Background: Understanding the OSM community

- August 2017: Unique study (survey) of OSM users linking demographics to edits

- Previous studies: **96% of all edits in OSM are made by men (Budhathoki, 2010)**

- Emphasised a need to understand the community

- However little emphasis on the *impact* of the bias on the data

96%

4%

Previous research has found a strong gender bias towards <u>young, tech-savvy men</u> (92-95%)
(Budhthoki, 2010; Schmidt and Klettner, 2013; Stephens, 2013)

- Critiques about biased representation have been theoretical rather than empirically measured
- Only way to measure the impact is through linking demographics to actual edits
- <u>This means collecting data directly from users</u>

Surveying OSM users: Systematic process

20 July 2017:
2 weeks prior to launch

- Introduction to me and my research agenda via OSM User Diaries

3 August 17

- Survey launch: Free online survey tool www.onlinesurveys.co.uk
- Notification via user diaries and 5 English language OSM mailing lists

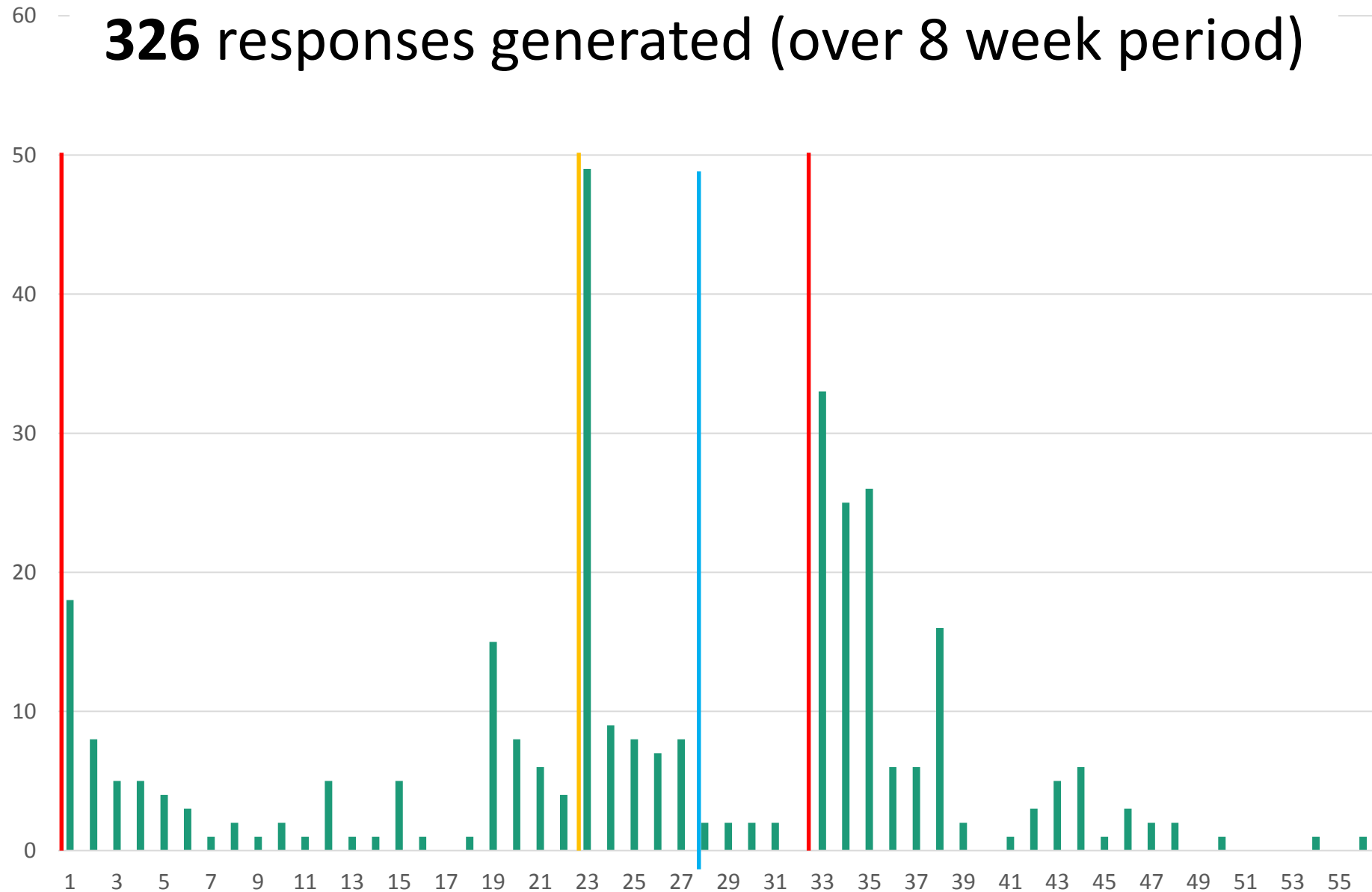**4 September 17:**
4 weeks after launch

- Reminders sent
- user diaries and Facebook

- Survey design: 6 questions: Gender, Age, Education, Country of Residence, Nationality AND, OSM username
- Incentivisation: Prize draw for 60 amazon vouchers

1. Access: You have to be on the mailing list to be able to send anything to it

2. Effectiveness: no information about numbers - how many people are we able to reach?

3. Which lists: how do we chose/decide which are valuable as routes of dissemination?

4. Language barriers: English correspondence via non-English speaking mailing lists?

Surveying OSM users: Response rate

326 responses generated (over 8 week period)

Surveying OSM users: Response rate

- **Overall sample n=297**
- **(326 – 33)** (duplication, unwillingness to provide username, user does not exist - incorrect username, typo, no account)

	Budhathoki 2010	Gardner 2017
Response rate	444/120k **(0.37%)** 444/33440 (1.33%)	326/4.3m **(< 0.01%)**
Sampling process	Direct messaging	User diaries + mailing lists
Personal data?	NO	YES

- **Causes**: Survey fatigue; many more 'lurkers', requirement of OSM identity, method of dissemination
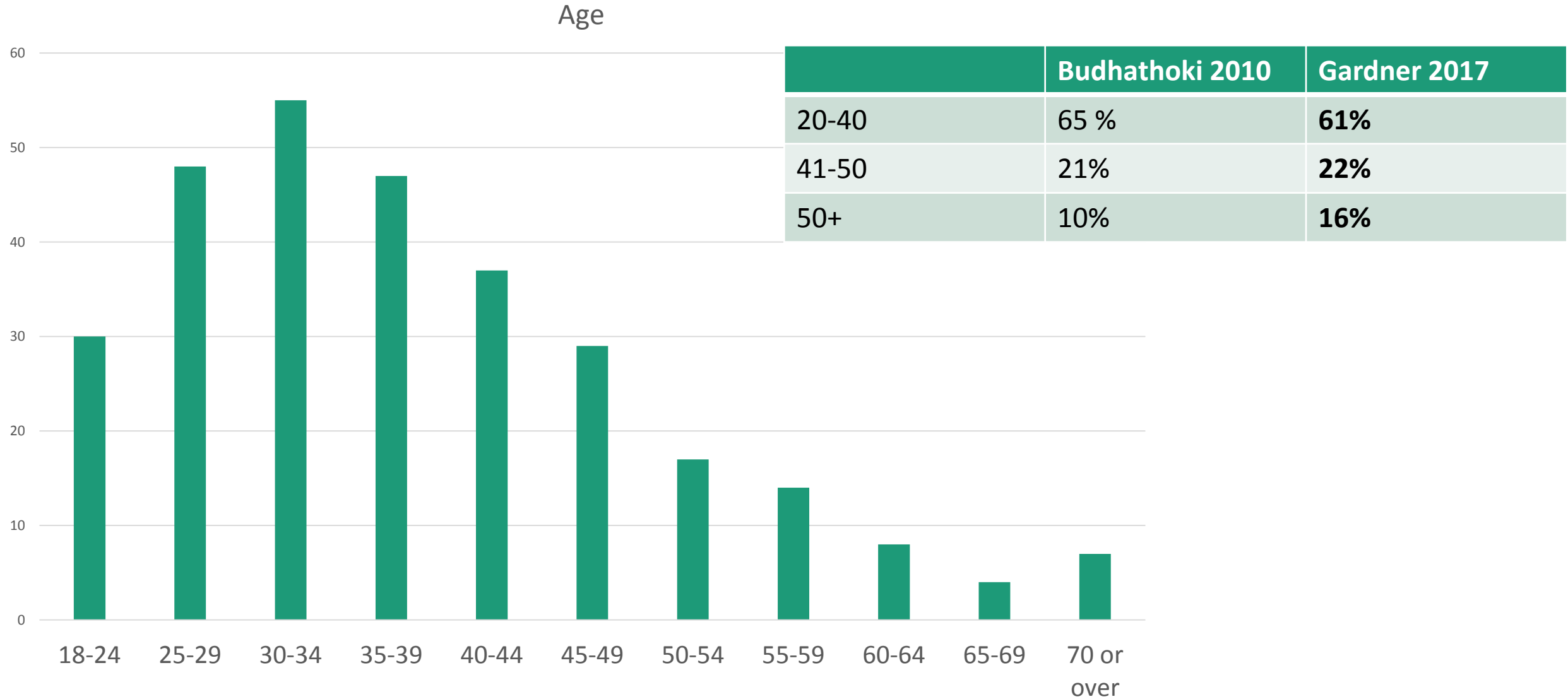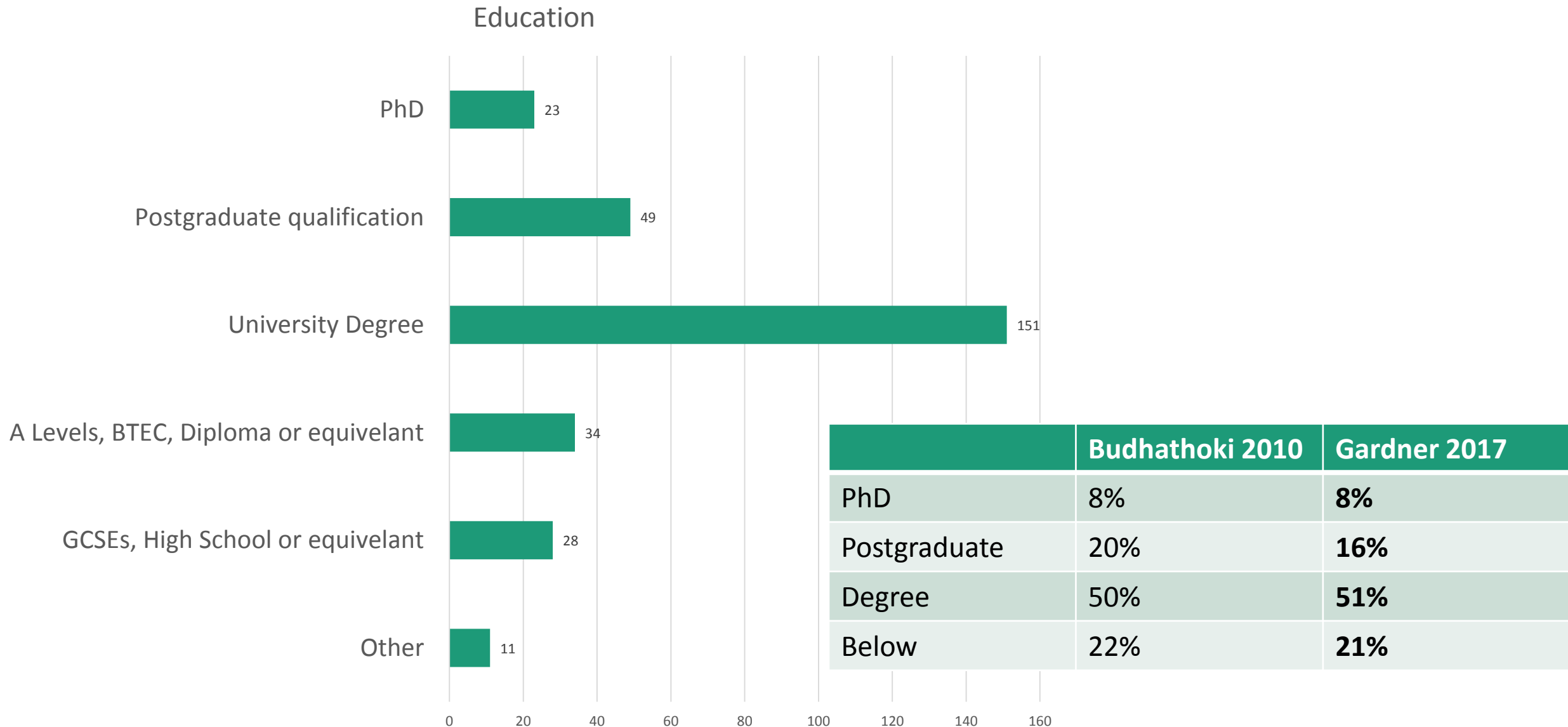
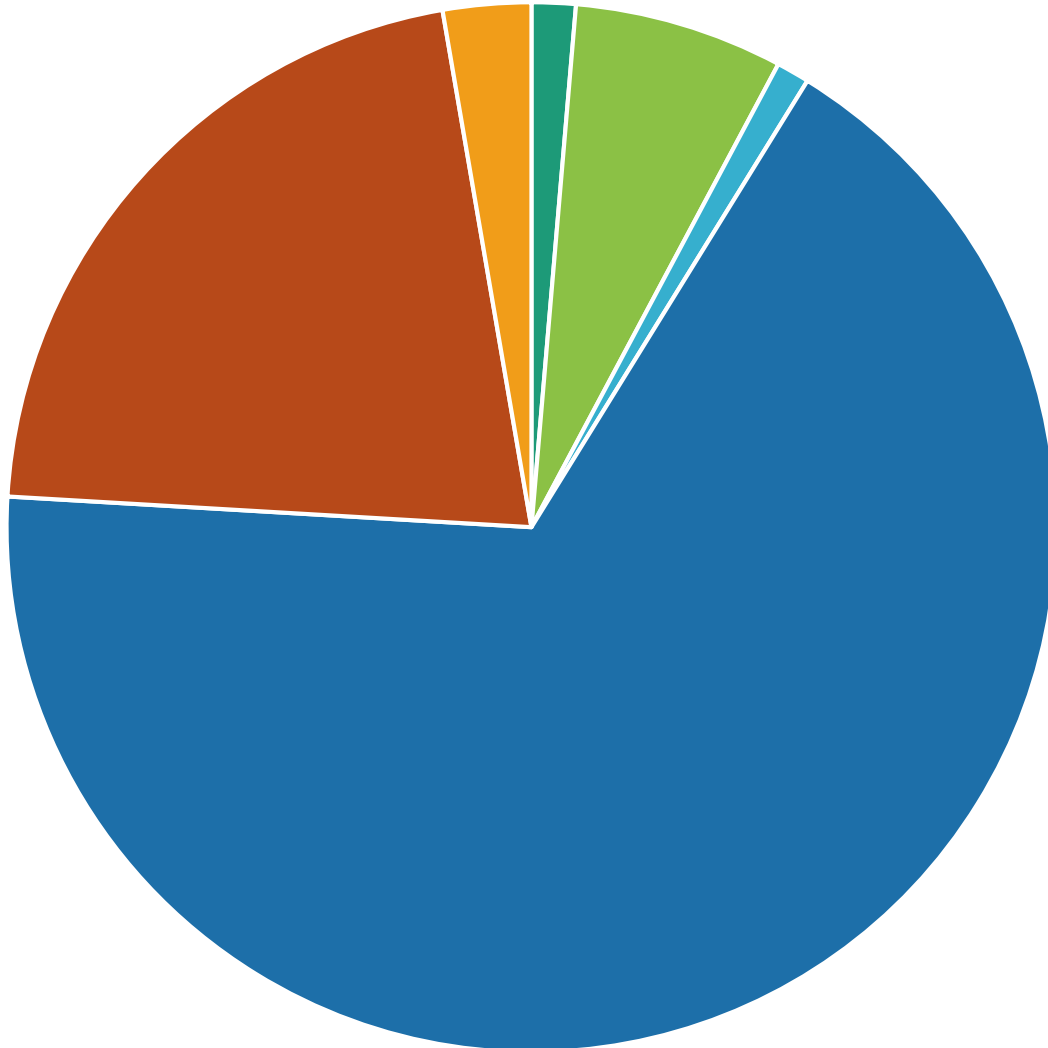# Surveying OSM users: Gender

87%

13%

Gardner *et al.*, 2017

Surveying OSM users: Age

Age

	Budhathoki 2010	Gardner 2017
20-40	65 %	**61%**
41-50	21%	**22%**
50+	10%	**16%**

Surveying OSM users: Education

Education

	Budhathoki 2010	Gardner 2017
PhD	8%	8%
Postgraduate	20%	16%
Degree	50%	51%
Below	22%	21%

Surveying OSM users: Location

Continental distribution

- Africa
- Australasia
- Australia
- Europe
- N.America
- S.America

	Budhathoki 2010	Gardner 2017
Europe	80%	**67%**
N. America	11%	**21%**
Africa	5%	**1%**
Australasia	3%	**7%**
S. America	<1%	**3%**

1. Identity, credibility and suspicion

"Not identifying yourself to us is troubling. All we have is your user name. Please identify yourself so we can learn more about you. Tell us about what some of the other research projects you may have been involved with".

"the general recommendation for scientists studying OSM is to get a decent amount of experience on the project before beginning the study. Your user account has zero edits at the moment - which makes your approach a bit like someone starting a study on a Japanese sociology topic without ever having been to Japan…"

2. Protectiveness: Challenging the need for the research

"Are there any studies concerning biases in non-citizen science … has anyone ever looked systematically at bias in non-crowdsourced geodata collections?"

"if your hypothesis is that crowdsourced geodata reproduces the bias in conventional geodata gathering to actually verify or falsify that you would need to know (a) what the nature of the conventional bias is and (b) what non-biased geodata looks like. Otherwise you'd end up with a relatively meaningless statement like "The Japanese do a lot of things in ways that are similar to the ways of the Americans"."

3. Misunderstanding the research

There is a misconception amongst men that there are no barriers to participation and that women are welcome to participate in a welcoming virtual community:

"OSM is an open community, there is no single obstacle for anyone to participate, regardless of their gender. If there are fewer females in OSM, it only reflects their free will (lack of will) to participate, nothing else. Equality is not about "having the equal number of these and those", it's about an equal opportunity to exercise own free will.

"Your work is awesome"

"That sounds like interesting research. Keep us posted"

"This is terrific - really beginning to cut through the generalities of the debate to some really worthwhile specifics"

1. **Sampling issues**: access to the survey (knowledge, language, mailing lists)

2. **Issues of trust**: getting the crowd on-board

3. **Issues of understanding**: clear articulation of the research agenda and the potential benefit to the community

1. How to we optimise sampling size?

2. **How do we engender trust and confidence?**

3. **How to we ensure effective understanding?**

4. [How] does understanding the crowd matter ?!