Open Source Software and the Community

Yu Huang

Vanderbilt University

yu.huang@vanderbilt.edu

What is open source?

- Open source software (OSS)
  - Software with source code that anyone can inspect, modify, and enhance.
- "Closed source" software
  - Only the person, team, or organization who created it—and maintains exclusive control over it—can modify.
- OSS examples
  - Linux
  - LibreOffice
  - Mozilla Firefox
  - …

What is open source?

- Promotes free distribution and access to a product design or ideas and implementation details

Open -> Collaboration is Open
Source -> Source is freely available

- Is open source software free?
  - Not really... Free here means "freedom", not in price
    - "Free" speech, not "free" beer
  - Most of them are, some do not qualify as free licenses (e.g., Open Watcom)
    - As far as we know, all existing released free software source code would qualify as open source. Nearly all open source software is free software, but there are exceptions.

Official Definition of Open Source Software

- Open source doesn't just mean access to the source code. The distribution terms of open-source software must comply with the following criteria:

    1. Free Redistribution
        - The license shall not restrict any party from selling or giving away the software as a component of an aggregate software distribution containing programs from several different sources. The license shall not require a royalty or other fee for such sale.
    2. Source Code
        - The program must include source code, and must allow distribution in source code as well as compiled form.
    3. Derived Works
        - The license must allow modifications and derived works, and must allow them to be distributed under the same terms as the license of the original software.

https://opensource.org/osd

Official Definition of Open Source Software

- Open source doesn't just mean access to the source code. The distribution terms of open-source software must comply with the following criteria:

    4. Integrity of the Author's Source code

    The license may restrict source-code from being distributed in modified form *only* if the license allows the distribution of "patch files" with the source code for the purpose of modifying the program at build time.

    5. No Discrimination Against Persons or Groups

    6. No Discrimination Against Fields of Endeavor

    7. Distribution of License

    The rights attached to the program must apply to all to whom the program is redistributed without the need for execution of an additional license by those parties.

    8. License Must Not Be Specific to a Product

    9. License Must Not Restrict Other Software

    10. License Must Be Technology-Neutral

https://opensource.org/osd

Open Source License Models

- Three popular open source licenses
  - Apache License
  - BSD license
  - Most common: MIT license

Open Source Software History

- ## 1950s and 1960s
  - Software and compilers were delivered as a part of hardware purchases without separate fees
  - Source code was generally distributed with the software, providing the ability to fix bugs or add new functionality
  - Many of the modifications developed by universities were openly shared
- ## 1970s
  - "Unbundling" of software from hardware [IBM]
- ## 1980s
  - Richard Stallman, formerly a programmer at the MIT AI Lab launched *the GNU project*
  - Computer Science Research Group (CSRG) of UC Berkeley working on *BSD UNIX system*

Open Source Software History

- 1990s

Open Source Software History

- 1990s
  - In Finland, Linus Torvalds, a student of computer science was implemented and released the first version of *Linux kernel*

Open Source Software History

- 1990s
  - In Finland, Linus Torvalds, a student of computer science was implemented and released the first version of *Linux kernel*
  - *GNU/Linux*
    - GNU is an operating system designed as a replacement for UNIX with many software programs; Linux is the Linux kernel. GNU/Linux refers to the system as a whole which is basically the GNU system, with Linux added.
  - 386BSD (Jolix)
    - Unix-like OS based on BSD (already discontinued)
  - BSD based OS: NetBSD, FreeBSD, OpenBSD

Open Source Software History

- 1990s
  - GNU/Linux distributions: Slackware, Debian, Red Hat, Suse, Mandrake…
  - Apache (web server)
  - Perl
  - GNOME and KDE (desktop environment)
  - Mozilla (funded by Netscape to build a web browser)
  - Open Source Initiative (OSI)
    - a non-profit corporation whose goal is to **promote the use of open source software in the commercial world**. To accomplish this goal, OSI maintains and promotes the Open Source Definition and offers the OSI Certified Open Source Software Certification Mark and Program

Open Source Projects

- Linux

Google runs on Linux-based Server

YAHOO! runs on Linux-based Server

amazon runs on Linux-based Server

facebook runs on Linux-based Server

YouTube runs on Linux-based Server

WIKIPEDIA runs on Linux-based Server

twitter runs on Linux-based Server

Open Source Projects

- Linux

- Apache
  - Developed and maintained by an open community of developers under the auspice of the Apache Software Foundation
  - The most popular web server on the internet since 1996

Open Source Projects

- Linux

- Apache
  - Developed and maintained by an open community of developers under the auspice of the Apache Software Foundation
  - The most popular web server on the internet since 1996

- MySQL
  - 1994: development starts
  - Most popular open source database
  - Used by Meta and Wikipedia, etc.

Open Source Projects

- PHP
  - A programming language designed for web development
  - "Personal Home Page" -> "Hypertext Preprocessor"
- Wordpress
  - Most popular blogging system
- VLC player
- Notepad++
- Ubuntu
- Android
  - A mobile OS based on Linux kernel, currently developed by Google
  - Source code is released by Google under open source licenses

Open Source Projects

- GIMP
  - GNU Image Manipulation Program
  - Free and open source graphics editor
- LaTex
  - A document preparation system for high quality typesetting (is also a PL)
- Arduino
  - Open source **hardware**!!

Open Source Platforms

Open Source Platforms

- Internet-based source code hosting facilities to support OSS development
  - GitHub
  - SourceForge
  - GitLab
  - Bitbucket
  - …

Name	Developer	Initial release	Free server?	Free client?	Associated collaborative development environment	Notes
Assembla	Assembla, Inc	2005	No	Unknown	Unknown	
Azure DevOps Services	Microsoft	2012[1]	No	No	Azure DevOps Services Microsoft Visual Studio	Most features are free for open source projects or teams of 5 members or less[2]
Bitbucket	Atlassian	2008	No	No	Atlassian BitBucket Server, JIRA and Confluence	Denies service to Cuba, Iran, North Korea, Sudan, Syria[3]
Buddy	Buddy, LLC.	2015	No	No	Unknown	Cloud version free for 1 project with no limit on size. Self-hosted version free up to 10 users with Fair Source license[4] applied.
CloudForge	CollabNet	2012	No	Unknown	Unknown	
Gitea	Gitea organization (open source community)[5]	2016-12[6]	Yes	Yes	Gitea	Gitea is an open-source software tool funded on Open Collective that is designed for self-hosting, but also provides a free first-party instance.[5]
GForge	The GForge Group,Inc.[7]	2006	Partial	Yes	Cloud version – free up to 5 users. On-premises version – free up to 5 users.	GForge is free for open source projects.
GitHub	Microsoft/GitHub, Inc	2008-04	No	No	Unknown	Denies service to Crimea, North Korea, Sudan, Syria[8] List of government takedown requests
GitLab	GitLab Inc.	2011-09[9]	Partial[10]	Yes[11]	GitLab FOSS – free software GitLab Enterprise Edition (EE) – proprietary	Denies service to Crimea, Cuba, Iran, North Korea, Sudan, Syria[12]
GNU Savannah	Free Software Foundation	2001-01	Yes	Yes	Savane	For use by projects with GPL compatible licenses, subject to staff approval. Code access review.[13]
Helix TeamHub	Perforce Software	1995	No	No	Cloud version – free up to 5 users. On-premises version requires a license.	Free cloud version has no limits on projects within 5gb storage limit. On-premises version has DevOps pipeline technology and free replicas.
Launchpad	Canonical	2004	Yes	No	Launchpad	Supports Bazaar and Git for version-controlled repository hosting.[14][15]
OSDN	OSDN K.K.	2002–04	Unknown	Yes	Unknown	For open-source projects only.[16] Ad-supported.
Ourproject.org	Comunes Collective	2002	Yes	Yes	FusionForge	For free software, free culture and free content projects.
OW2	OW2	2008	No	No	GitLab	Oriented on middleware technology.
Phabricator	Phacility, Inc	2010	Yes	Yes	Phabricator	End of life.[17]
SEUL	Unknown	1997-05	Unknown	No	Unknown	
SourceForge	BizX LLC	1999-11	Yes[18][19]	Yes	Apache Allura	For use by open-source projects.[20] Ad-supported. Subject to American export restrictions, so denies service to Cuba, Iran, North Korea, Sudan, Syria.[21]
Name	Manager	Established	Server side: all free software	Client side: all-free JS code	Developed or used CDE	Notes

https://en.wikipedia.org/wiki/Comparison_of_source-code-hosting_facilities

Name	Code review	Bug tracking	Web hosting	Wiki	Translation system	Shell server	Mailing list	Forum	Personal repository	Private repository	Announce	Build system	Team	Release binaries	Self-hosting
Assembla	Yes[22]	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes[23]	Yes	Yes	Yes	Unknown	Commercially[24]
Azure DevOps Services	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Commercially (Azure DevOps Server)
Bitbucket	Yes[25]	Yes[a]	Yes[26]	Yes	No	No	No	No	Yes	Yes[b]	No	Yes[27]	Yes	No[28]	Commercially (Bitbucket Server formerly Stash)[c]
Buddy	Yes	Yes	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes[d]	Yes	Yes	Yes
CloudForge	Unknown	Yes	Yes	Yes	No	No	No	No	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	No
GForge	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gitea	Yes	Yes	No	Yes	No	No	No	No	Yes	Yes	Unknown	3rd-party (e.g. Travis CI, Appveyor and others)[29]	Yes	Yes	Yes
GitHub	Yes[30]	Yes[31][e]	Yes[32]	Yes	No	No	No	No	Yes	Yes	Yes	Yes[33]	Yes	Yes	Commercially (GitHub Enterprise)
GitLab	Yes[34]	Yes	Yes[35]	Yes	No	No	No	No	Yes	Yes	Yes	Yes[36]	Yes	Yes[37]	Yes[f]
GNU Savannah	Yes[38]	Yes	Yes	No	No	Yes	Yes	No[39]	No	No	Yes	No	Yes	Unknown	Yes
Helix TeamHub	Yes[40]	Yes	No	Yes	No	No	Yes	Yes	Yes	Yes	No	Yes, with hooks. Jenkins, TeamCity, etc.	No	Yes	Yes
Kallithea	Yes	No	Yes	No	No	Unknown	No	No	Yes	Yes	No	No	Yes	Yes	Yes
Launchpad	Yes	Yes	No	No	Yes	No	Yes	No	Yes	Yes[g]	Yes	Yes[h]	Yes	Unknown	Yes
OSDN	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Yes	No	Yes	Yes	No
Ourproject.org	Unknown	Yes	Yes	Yes	No	Unknown	Yes	Yes	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Yes
Phabricator	Yes	Yes	Yes	Yes	Unknown	Yes	Unknown	Yes	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Yes
RhodeCode	Yes	No	Yes	No	No	Unknown	No	No	Yes	Yes	Yes	No	Yes	Yes	Yes
SourceForge	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes[i]	Yes	No	Yes	Yes	Yes
Name	Code review	Bug tracking	Web hosting	Wiki	Translation system	Shell server	Mailing list	Forum	Personal repository	Private repository	Announce	Build system	Team	Release binaries	Self-hosting

https://en.wikipedia.org/wiki/Comparison_of_source-code-hosting_facilities

Research on OSS

- Program analysis, testing, debugging analysis
- Unique for OSS
  - Commit history (CI/CD)
  - Issue recommendation
  - Sustainability
    - Diversity
    - Gender biases
    - **Social good community: OSS for Social Good (OSS4SG)**

Motivation

- Social Good projects are under-explored, but very important to OSS community, on broadening participation in OSS and solving societal issues

Motivation

- Social Good projects are under-explored, but very important to OSS community, on broadening participation in OSS and solving societal issues

REFUGE restrooms

Providing safe restroom access to transgender, intersex, and gender nonconforming individuals.

REFUGE is an effort to fill the void left by the now-defunct Safe2Pee website. It provides a free resource to trans* and queer individuals in need of gender neutral and other safe restrooms.

This project is open source. Feel free to contribute. We could use the help.

Motivation

- Social Good projects are under-explored, but very important to OSS community, on broadening participation in OSS and solving societal issues

REFUGE

Providing safe restr

REFUGE is an effo
queer individuals in

This project is open

ans* and

Little Window

Little window is a clever little cat chatbot that directs women to the information they are looking for as quickly as possible. Think of it like google search on turbo for all of Chayn's resources and those provided by our friends too. We want to drastically reduce the time women take to search for information which in many cases can save lives.

Currently in active development `build` `passing`

Little Window

I'm still learning how to do this. Tell the Chayn team what you think. How would you rate me (1 (not helpful) to 5 (very helpful))?

5

How can I improve?

provide more information on h

Submit

Motivation

- Social Good projects are under-explored, but very important to OSS community, on broadening participation in OSS and solving societal issues

REFUGE

Providing safe restr...

REFUGE is an effo...
queer individuals in...

This project is open...

Little Window

Little window is a clever little cat chatbot that directs women to the information they are looking for as quickly as possible. Think of it like google search on turbo for all of Chayn's resources and those provided by our friends too. We want to drastically reduce the time women take to search for information which in many cases can save lives.

Currently in active development `build passing`

trans* and

Little Window

I'm still learning how to do this. Tell the Chayn team what you think. How would you rate me (1 (not helpful) to 5 (very helpful))?

5

How can I improve?

provide more information on h| Submit

CommCare HQ

CommCare HQ is a server-side tool to help manage community health workers. It seamlessly integrates with CommCare J2ME, Android, and Web Apps as well as providing generic domain management and form data-collection functionality.

Key Components

- CommCare application builder
- OpenRosa compliant XForms designer
- SMS integration
- Domain/user/mobile worker management
- XForms data collection
- Case management
- Over-the-air (ota) restore of user and cases
- Integrated web and email reporting

CommCare

Digital Solutions for COVID-19 Response

Motivation

- Social Good projects are under-explored, but very important to OSS community, on broadening participation in OSS and solving societal issues
  - A greater "social good"
  - Increased interest
  - Lack of support

https://socialimpact.github.com/assets/images/GitHub_tCF_OSSInSocialSector_FINAL_updated.pdf

Motivation

- Social Good projects are under-explored, but very important to OSS community, on broadening participation in OSS and solving societal issues
  - A greater "social good"
  - Increased interest
  - Lack of support
- *"Help one, help all"*

https://socialimpact.github.com/assets/images/GitHub_tCF_OSSInSocialSector_FINAL_updated.pdf

Challenges

- No unified definition of "*Social Good*" in OSS
  - AI for Social Good (AI4SG)
    - *"Appropriate moral choices in AI and machine learning models"*

Challenges

- No unified definition of "*Social Good*" in OSS
  - AI for Social Good
    - *"Appropriate moral choices in AI and machine learning models"*

- Sources for projects
  - No labels or clear categorizations
  - No official platform

Challenges

- No unified definition of "*Social Good*" in OSS
  - AI for Social Good
    - *"Appropriate moral choices in AI and machine learning models"*

- Sources for projects
  - No labels or clear categorizations
  - No official platform

- No previous studies on "Social Good" in OSS from developers' perspectives
  - *We introduce the notion:* **Open Source for Social Good (OSS4SG)**

High-level Question

- **How to characterize the OSS4SG community? How can we support them?**

Research Questions

- **RQ1:** How do contributors define OSS for Social Good?
- **RQ2:** What motivations do contributors have to contribute to OSS for Social Good?
- **RQ3:** What factors do contributors consider to select an OSS for Social Good project?
- **RQ4:** What are the current challenges to work in OSS for Social Good?

Outline

- **Motivation**
- **Challenges**
- **Research questions**
- **Study design**
- **Results**
- **Implications**
- **Conclusions**

Study Design

Sequential mixed-methods study
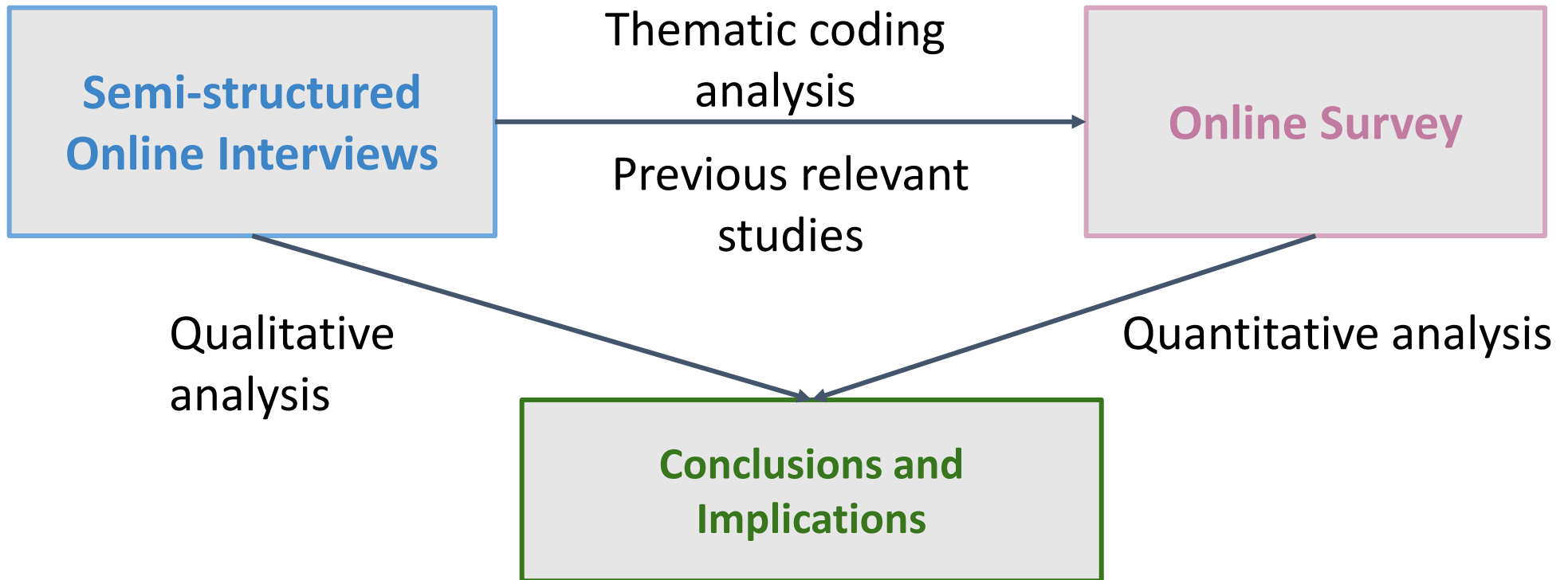
Semi-Structured Interviews

- 438 OSS4SG projects on GitHub
  - *OVIO*: 7 social impact labels

Semi-Structured Interviews

- 438 OSS4SG projects on GitHub
  - *OVIO*: 7 social impact labels
  - *Digital Public Goods*: 17 SDG labels
  - 14,256 contributors, 7500 valid

Semi-Structured Interviews

- 438 OSS4SG projects on GitHub
  - *OVIO*: 7 social labels
  - *Digital Public Goods*: 17 SDG labels
  - 14,256 contributors, 7500 valid

- Recruited 21* participants
  - Average OSS experience: 4.38 years
  - Average OSS4SG experience: 3.05 years
  - Coming from 11 different countries

- 45-55 minutes
  - 17 hours of audio data
  - Thematic coding analysis for survey design

ID	Gender	SG Exp	OSS Exp	Location of Contribution	Project Domains
P1	W	2	2	Mexico	Crypto, Security
P2	M	1	1	USA	Finance
P3	M	8	8	Germany	Education, Healthcare, Disaster Tracking
P4	W	1	1	UK	Charity, Domestic Violence
P5	M	1	3	India	Environment
P6	M	0.5	10	Turkey	COVID-19 Tracking
P7	M	0.5	0.5	India	Education, Environment
P8	M	4	5.5	Israel	Structurally-Safe Buildings
P9	M	8	8	Australia	Healthcare, Education
P10	W	2	2	India	Healthcare, Education
P11	W	0.5	0.5	India	Education
P12	M	2	2	USA	COVID-19 Tracking
P13	M	2	6	USA	Education, Non-profit Tools
P14	NB	8	8	Germany	Anti-Gentrification, Safe Restrooms
P15	M	10	10	Spain	eGovernment, Civil Participation
P16	M	0.5	0.5	India	Healthcare
P17	M	2	2	India	Education
P18	M	1.5	1.5	Romania	Local Administration
P19	M	0.5	1	India	Healthcare
P20	M	5	15	Canada	Management for Government and Charity
P21*	M	5	5	USA	Healthcare

Semi-Structured Interviews

- 438 OSS4SG projects on GitHub
  - **_OVIO_**: 7 social labels
  - **_Digital Public Goods_**: 17 SDG labels
  - 14,256 contributors, 7500 valid

- Recruited 21* participants
  - Average OSS experience: 4.38 years
  - Average OSS4SG experience: 3.05 years
  - Coming from 11 different countries

- 45-55 minutes
  - 17 hours of audio data
  - Thematic coding analysis for survey design

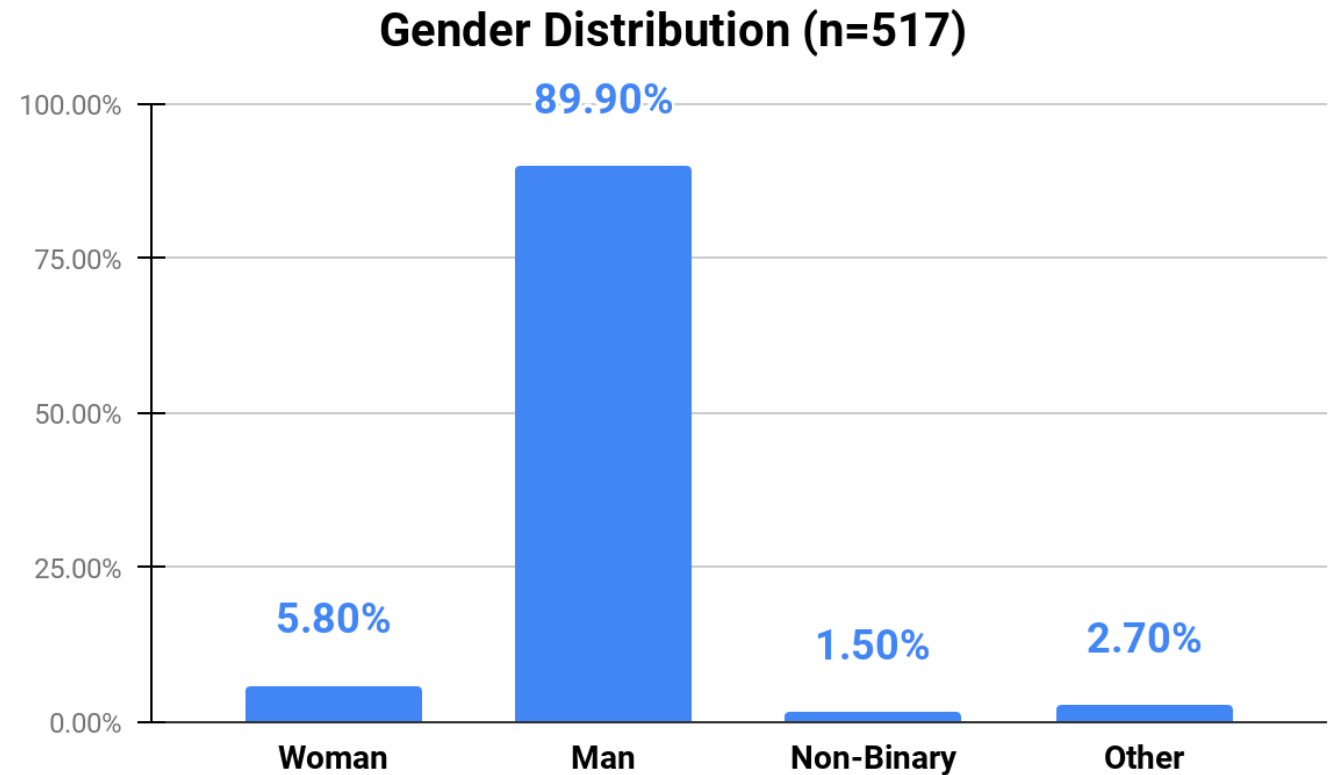| ID	Gender	SG Exp	OSS Exp	Location of Contribution	Project Domains
P1	W	2	2	Mexico	Crypto, Security
P2	M	1	1	USA	Finance
P3	M	8	8	Germany	Education, Healthcare, Disaster Tracking
P4	W	1	1	UK	Charity, Domestic Violence
P5	M	1	3	India	Environment
P6	M	0.5	10	Turkey	COVID-19 Tracking
P7	M	0.5	0.5	India	Education, Environment
P8	M	4	5.5	Israel	Structurally-Safe Buildings
P9	M	8	8	Australia	Healthcare, Education
P10	W	2	2	India	Healthcare, Education
P11	W	0.5	0.5	India	Education
P12	M	2	2	USA	COVID-19 Tracking
P13	M	2	6	USA	Education, Non-profit Tools
P14	NB	8	8	Germany	Anti-Gentrification, Safe Restrooms
P15	M	10	10	Spain	eGovernment, Civil Participation
P16	M	0.5	0.5	India	Healthcare
P17	M	2	2	India	Education
P18	M	1.5	1.5	Romania	Local Administration
P19	M	0.5	1	India	Healthcare
P20	M	5	15	Canada	Management for Government and Charity
P21*	M	5	5	USA	Healthcare

Survey: Participant Recruitment

- 438 OSS4SG projects
  - ***OVIO***
  - ***Digital Public Goods***
  - 14,256 contributors, 7500 valid

- 642 OSS projects
  - Randomly selected from GitHub
  - 17,978 contributors, 9409 valid

Same set of projects included in online interviews

Survey: Participant Recruitment

- 438 OSS4SG projects
  - ***OVIO***
  - ***Digital Public Goods***
  - 14,256 contributors, 7500 valid

- 642 OSS projects
  - Randomly selected from GitHub
  - 17,978 contributors, 9409 valid
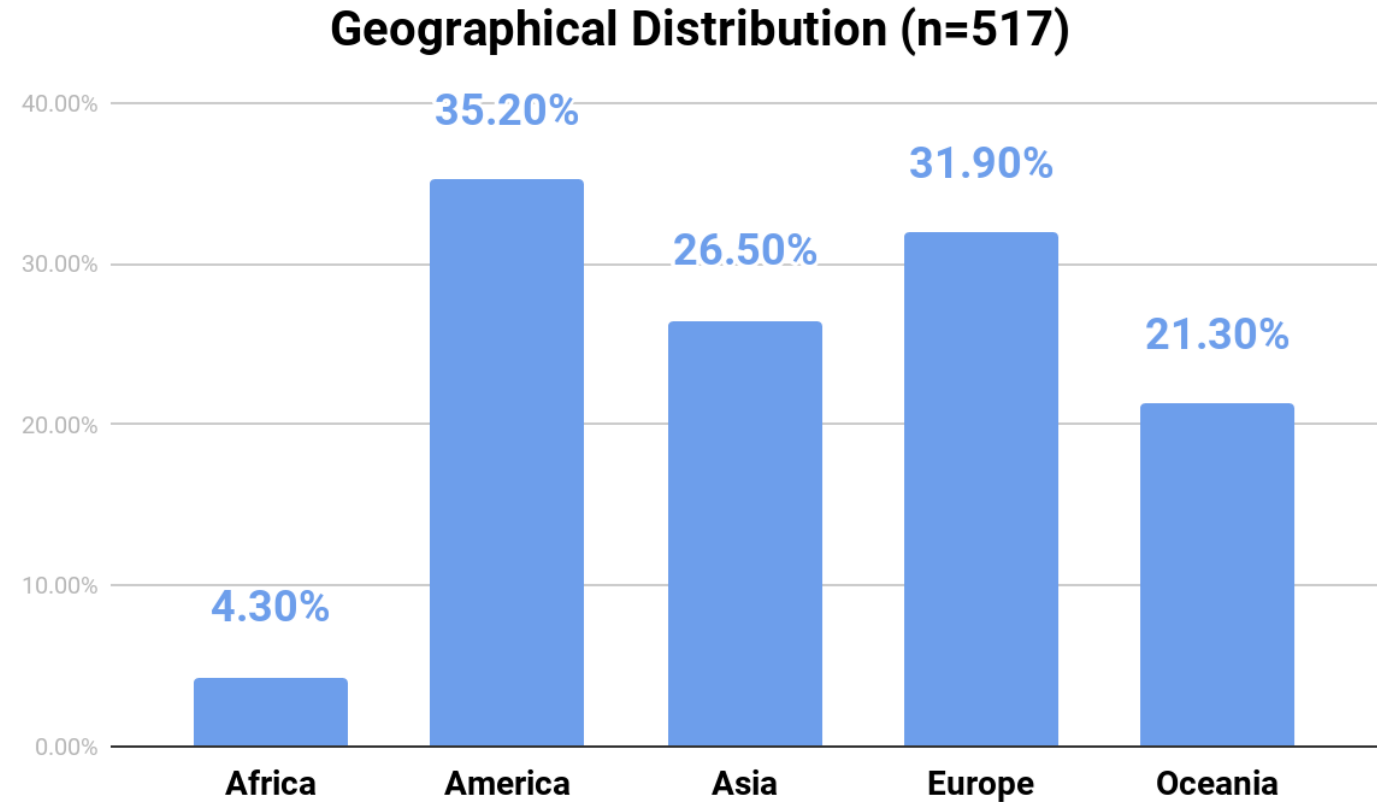
- 517 respondents (8.97%)

Survey: Participant Recruitment

- 438 OSS4SG projects
  - ***OVIO***
  - ***Digital Public Goods***
  - 14,256 contributors, 7500 val

- 642 OSS projects
  - Randomly selected from GitH
  - 17,978 contributors, 9409 val

- 517 respondents (8.97%)

Gender Distribution (n=517)

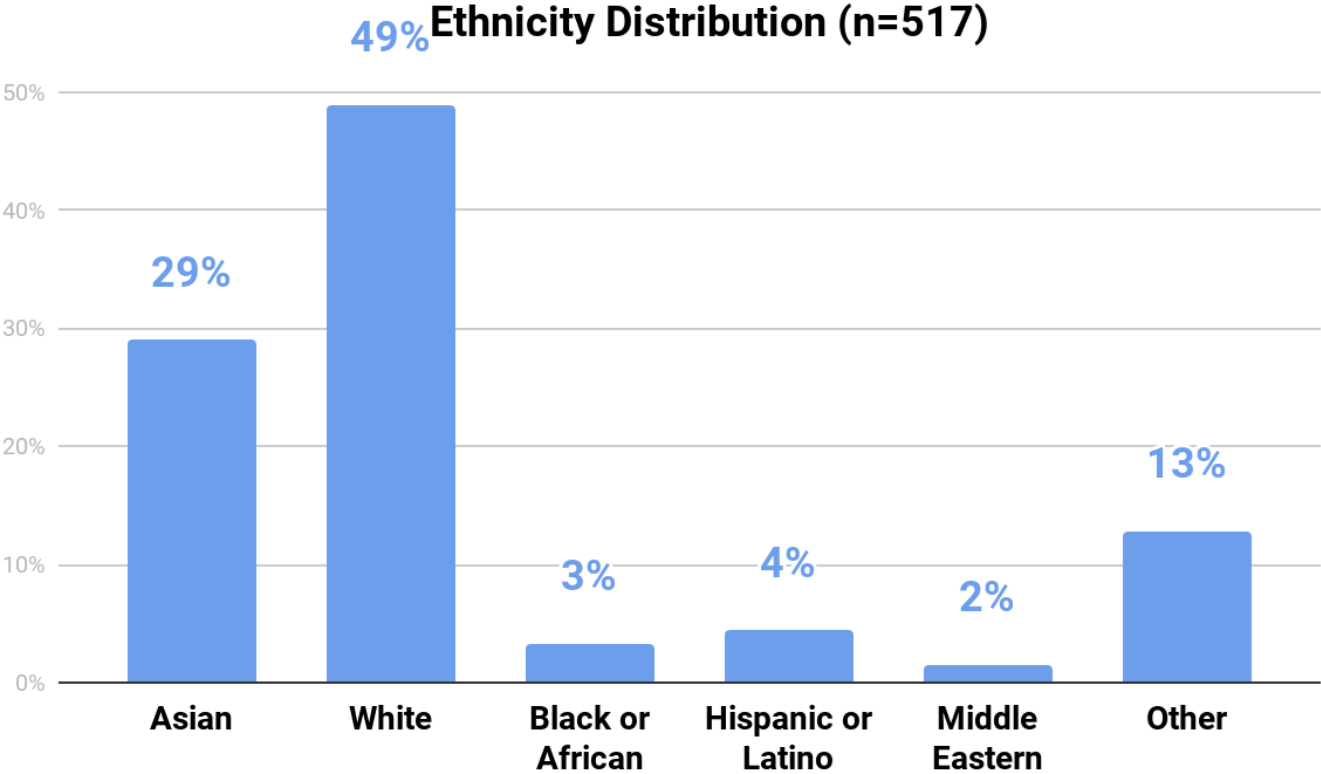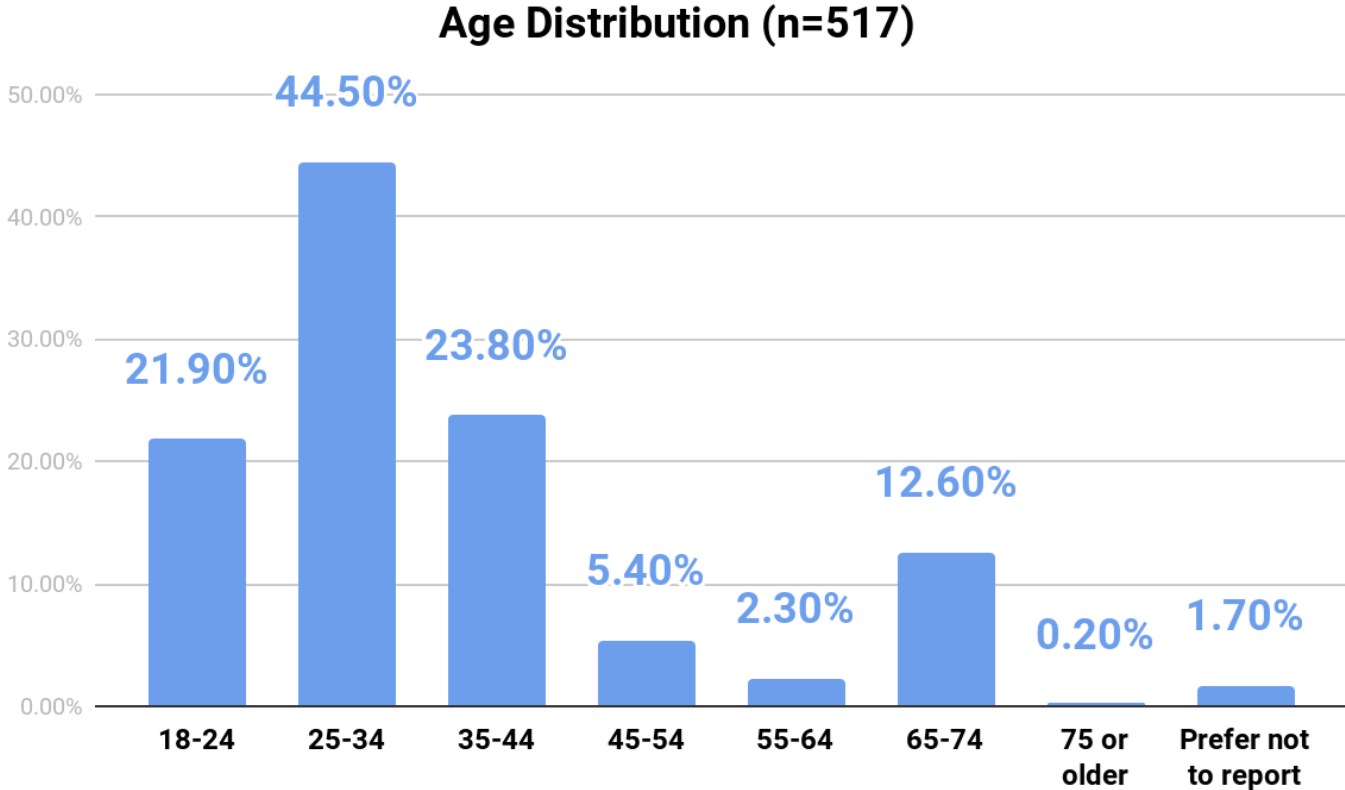|		
100.00%	89.90%	
75.00%		
50.00%		
25.00%	5.80%	1.50% 2.70%
0.00%	Woman Man Non-Binary Other	

Survey: Participant Recruitment

- 438 OSS4SG projects
  - *OVIO*
  - *Digital Public Goods*
  - 14,256 contributors, 7500 val

- 642 OSS projects
  - Randomly selected from GitH
  - 17,978 contributors, 9409 val

- 517 respondents (8.97%)

Geographical Distribution (n=517)

Region	Percentage
Africa	4.30%
America	35.20%
Asia	26.50%
Europe	31.90%
Oceania	21.30%

Survey: Participant Recruitment

- 438 OSS4SG projects
  - *OVIO*
  - *Digital Public Goods*
  - 14,256 contributors, 7500 vali

- 642 OSS projects
  - Randomly selected from GitH
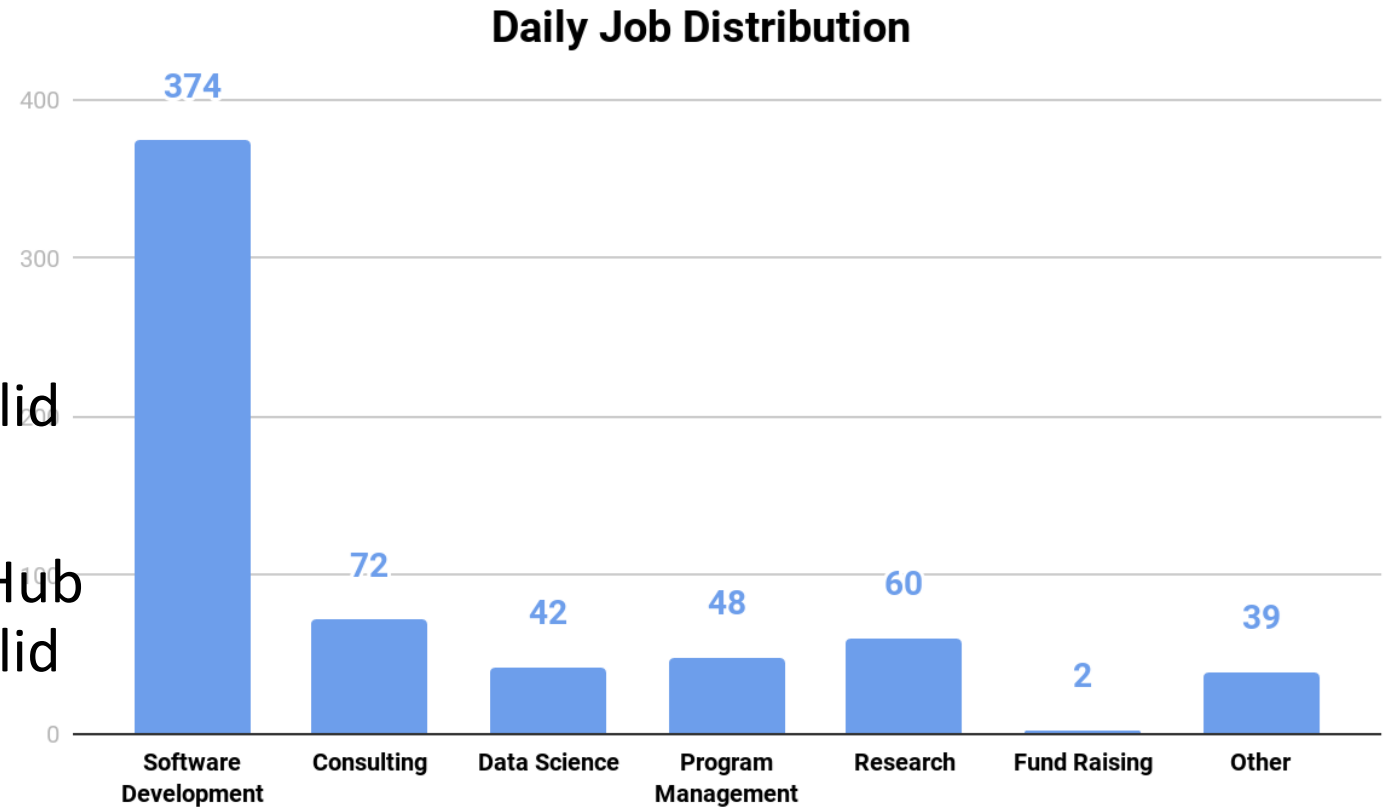 - 17,978 contributors, 9409 vali

- 517 respondents (8.97%)

Ethnicity Distribution (n=517)

49%	
29%	
50%	
40%	
30%	
20%	13%
10%	3% 4% 2%
0%	

Asian White Black or African Hispanic or Latino Middle Eastern Other

Survey: Participant Recruitment

- 438 OSS4SG projects
  - **OVIO**
  - **Digital Public Goods**
  - 14,256 contributors, 7500 va

- 642 OSS projects
  - Randomly selected from GitH
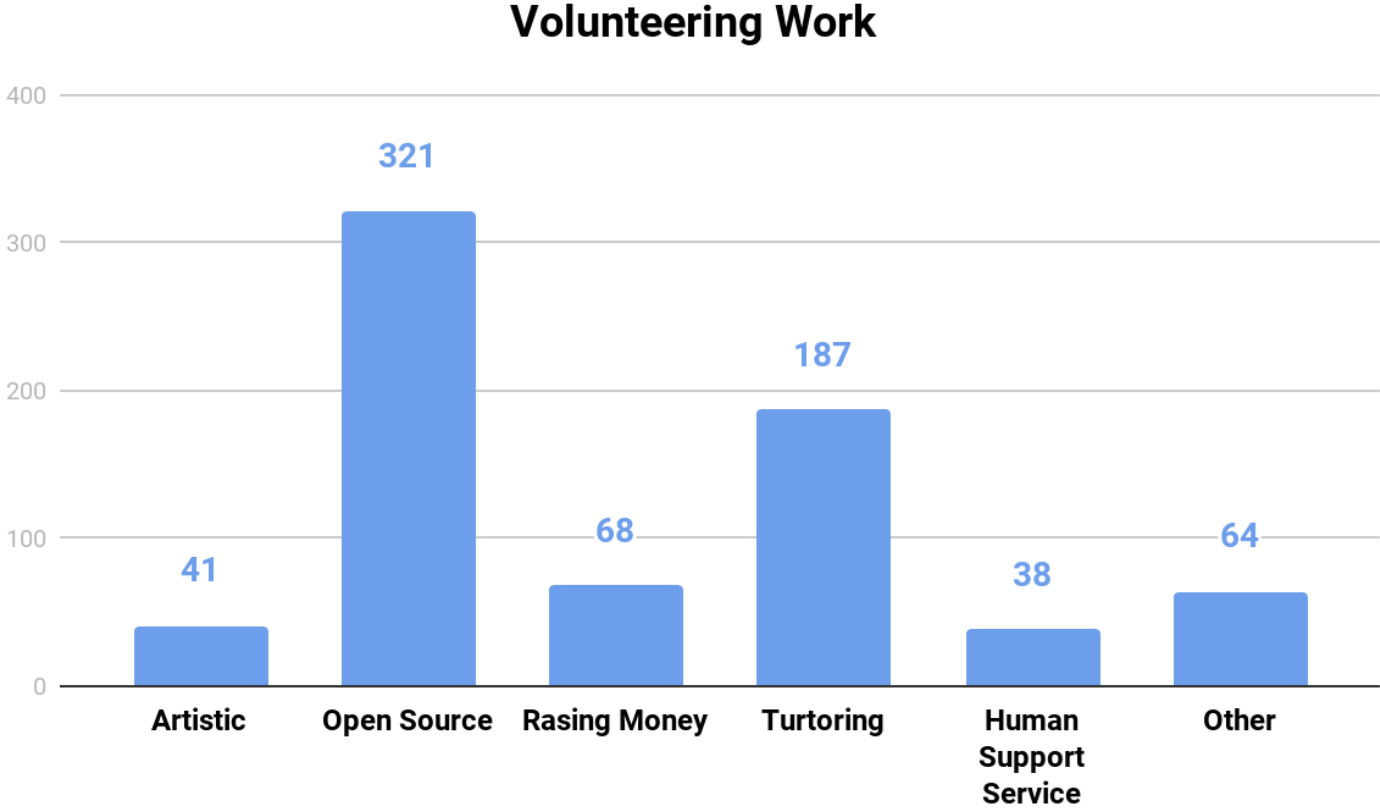 - 17,978 contributors, 9409 va

- 517 respondents (8.97%)

Age Distribution (n=517)

Survey: Participant Recruitment

- 438 OSS4SG projects
  - ***OVIO***
  - ***Digital Public Goods***
  - 14,256 contributors, 7500 valid

- 642 OSS projects
  - Randomly selected from GitHub
 - 17,978 contributors, 9409 valid

- 517 respondents (8.97%)

Daily Job Distribution

Category	Value
Software Development	374
Consulting	72
Data Science	42
Program Management	48
Research	60
Fund Raising	2
Other	39

Survey: Participant Recruitment

- 438 OSS4SG projects
  - ***OVIO***
  - ***Digital Public Goods***
  - 14,256 contributors, 7500 val
- 642 OSS projects
  - Randomly selected from GitH
  - 17,978 contributors, 9409 val
- 517 respondents (8.97%)

Volunteering Work

Survey: Participant Recruitment

- 438 OSS4SG projects
  - *OVIO*
  - *Digital Public Goods*
  - 14,256 contributors, 7500 valid

- 642 OSS projects
  - Randomly selected from GitHub
  - 17,978 contributors, 9409 valid

- 517 respondents (8.97%)

- Grouping participants into four sets, two pairs
  - P-OSS vs. P-OSS4SG
  - P-OSS+ vs. P-OSS4SG+
    - Actual OSS4SG experience
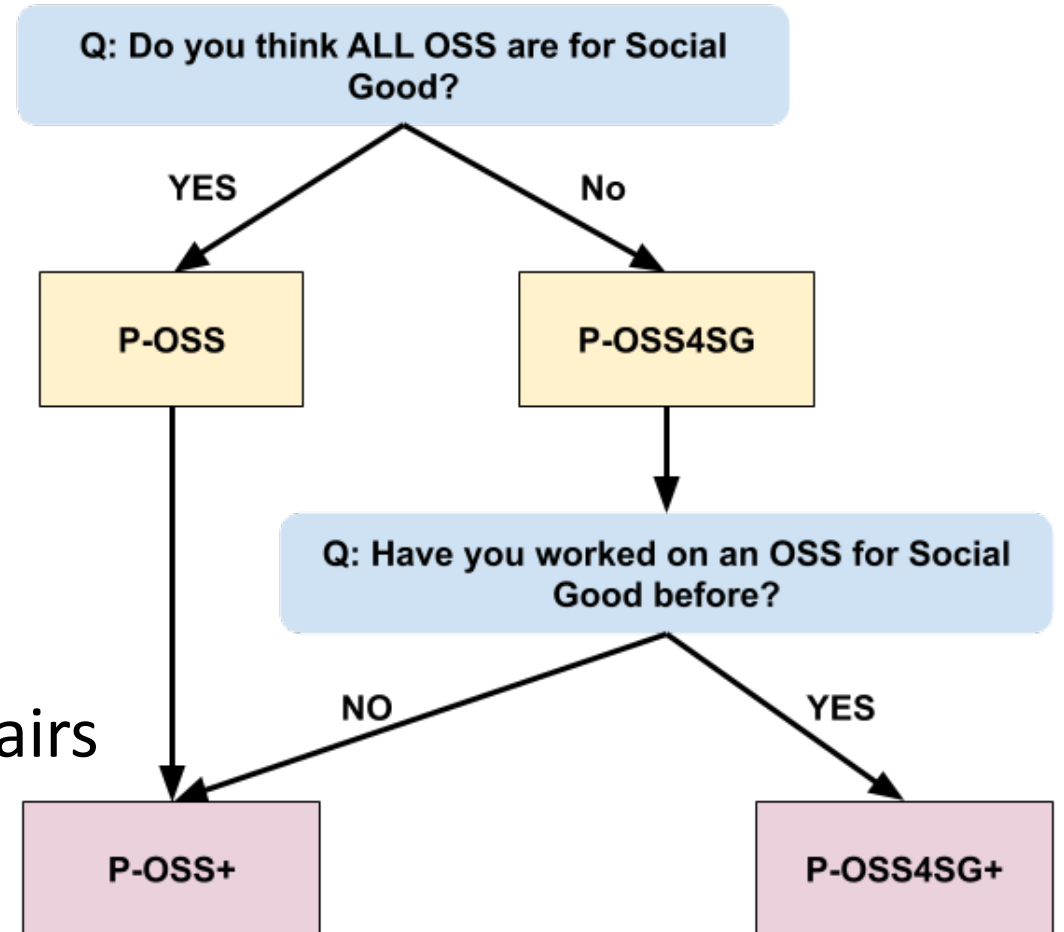
RQ1: Developers' Definition of OSS4SG (OSS4SG ?= OSS)

- **Interview highlights**
  - **ALL OSS are for Social Good (OSS4SG == OSS) (3/20)**
    - *"I think the fact that the project is open source and people can contribute their knowledge and can use it as they are required by the license, there's a social benefit that everyone can use from the project itself."*
    - *"I think almost all projects should come into social good because some way or another, you never know what the bigger project is. So I may be using another open source software for my project, my project being something for kids or for education or for some other social purposes, but the open source project I am working on may be just a tech-based project."*

RQ1: Developers' Definition of OSS4SG (OSS4SG ?= OSS)

- **Interview highlights**
  - **ALL OSS are for Social Good (OSS4SG==OSS) (3/20)**
  - **OSS4SG != OSS (17/20)**
    - **The project targets people or communities that need help**
      - *"It's just for the software that actually makes a difference on the access or on the usability of people who needs some help."*
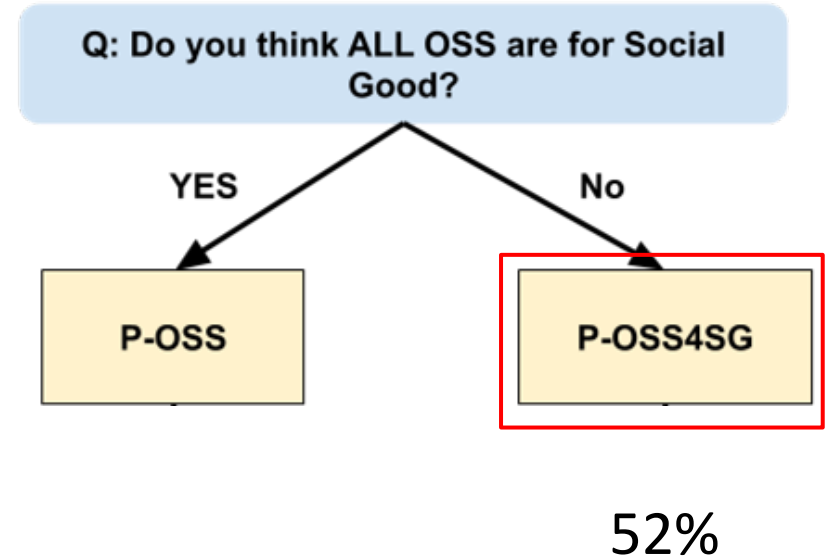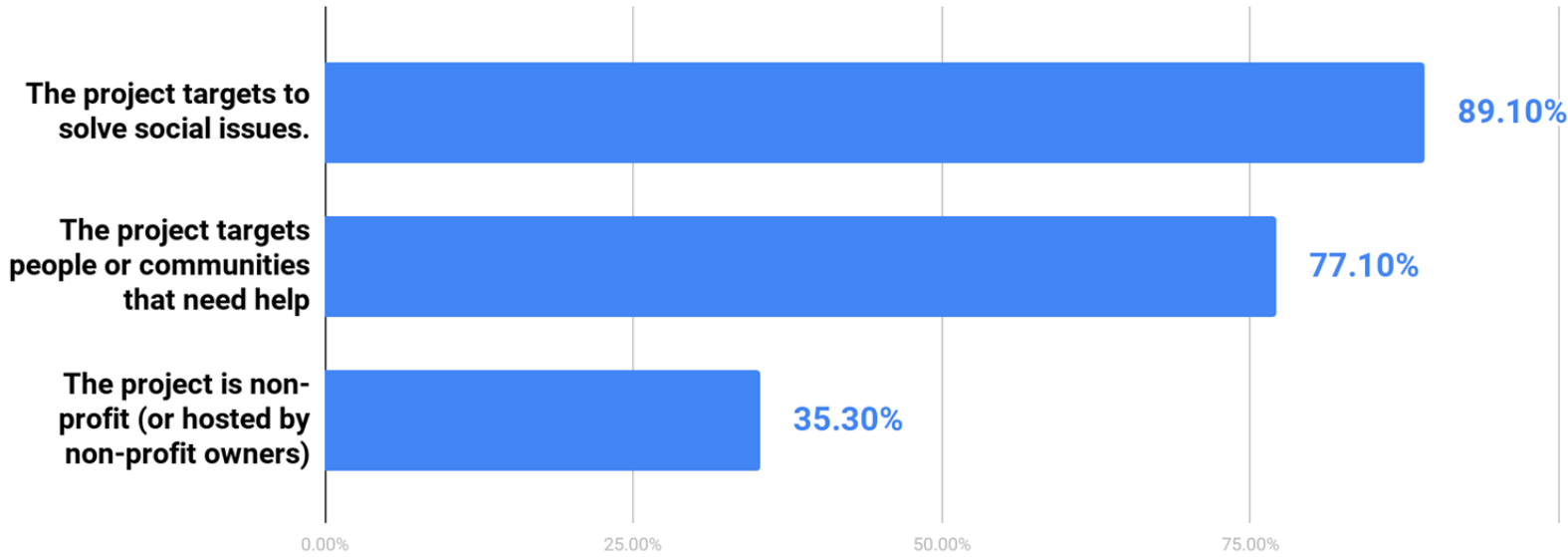   - **The project aims to solve some societal issues or provide social benefits:**
      - *"If we are talking about the social good term, it comes with any project or any initiative that is, like, that aims to provide any help to the social projects."*
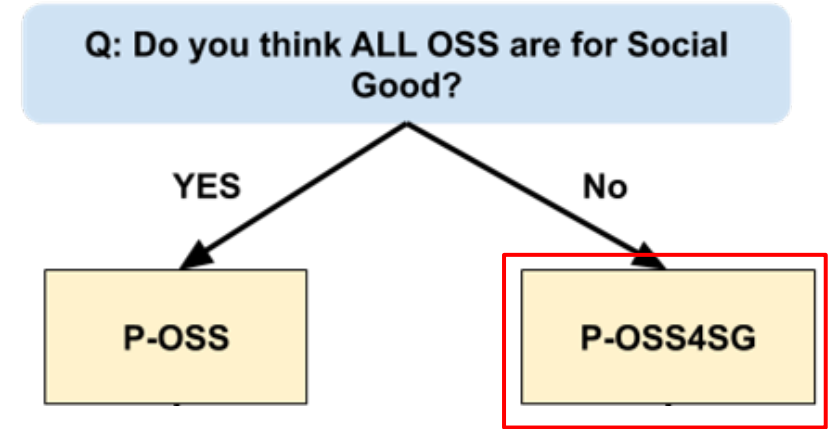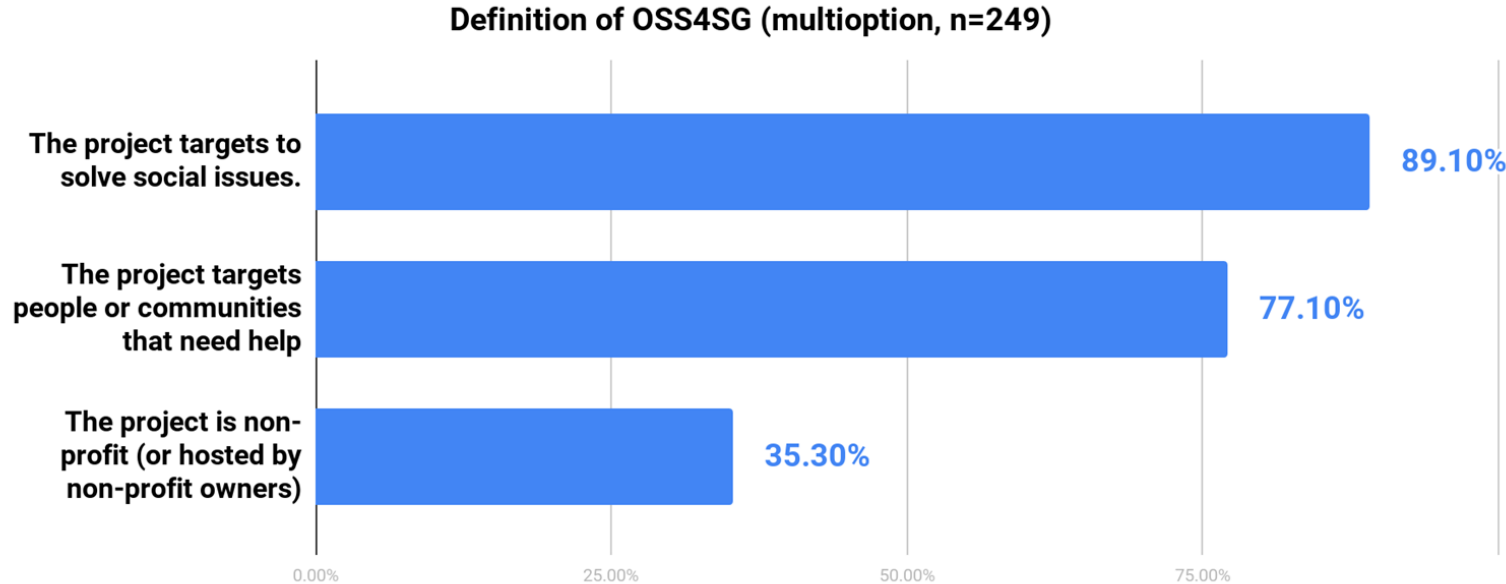   - **The project is non-profit (or hosted by non-profit owners):**
      - *"The goal is the software itself, not necessarily profiting off the software."*

RQ1: Developers' Definition of OSS4SG

Definition of OSS4SG (multioption, n=249)

- The project targets to solve social issues. — **89.10%**
- The project targets people or communities that need help — **77.10%**
- The project is non-profit (or hosted by non-profit owners) — **35.30%**

Q: Do you think ALL OSS are for Social Good?

YES → P-OSS

No → P-OSS4SG

52%

RQ1: Developers' Definition of OSS4SG

Definition of OSS4SG (multioption, n=249)

- The project targets to solve social issues. — 89.10%
- The project targets people or communities that need help — 77.10%
- The project is non-profit (or hosted by non-profit owners) — 35.30%

Q: Do you think ALL OSS are for Social Good?

YES → P-OSS

No → P-OSS4SG

OSS4SG are open source software projects where the outcome distinctly targets a community of people to overcome a societal issue.

RQ2: Motivations for OSS4SG contributors

- ## Interview highlights
  - ### To help those in need (13/20)
    - *"I'm so much more motivated to build products that I know have a good outcome for a group of people that is generally underserved"*
  - ### To become a better programmer (10/20)
    - *"when I contribute to that, it can definitely give me more experience."*
  - ### To have an impact on society (8/20)
    - ***"So, I think the main reason is because I want to make a difference on my life... make a fingerprint on the world."***
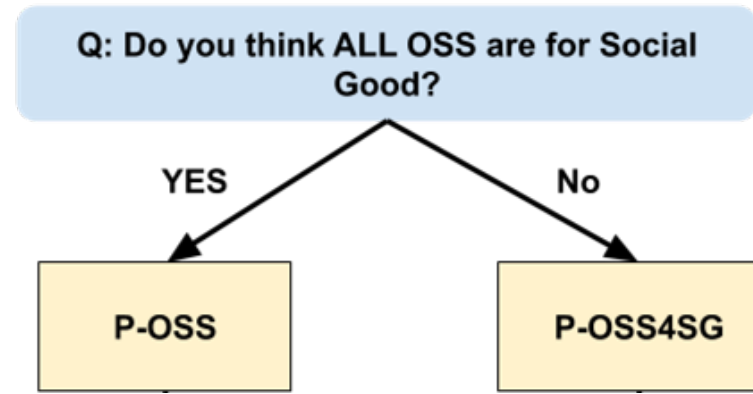 - ### For emotional fulfillment (7/20)
    - *"It gives a mental satisfaction that I'm working towards something good"*
  - ### To give back as I received (5/20)
    - *"I also feel like however much you take from something, you should give back."*

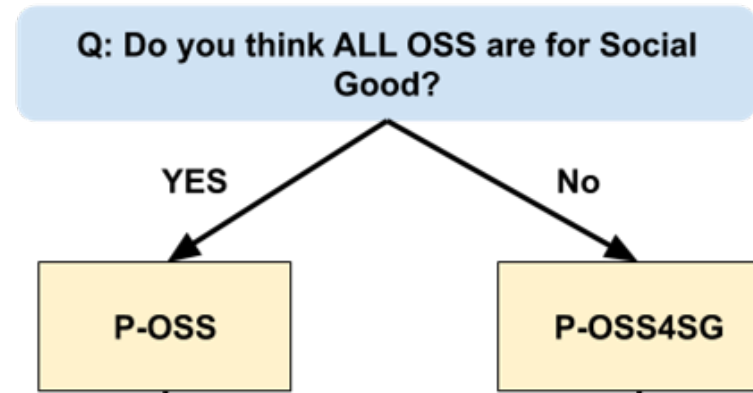# RQ2: Motivations

- **Survey Results**

Q: Do you think ALL OSS are for Social Good?

YES → P-OSS

No → P-OSS4SG

Motivation	P-OSS4SG (n=222)	P-OSS (n=198)	Delta (P-OSS4SG - P-OSS)
I want to help the target users. (M1)	69.4%	65.7%	3.7%
I want to give back. (M2)	66.2%	63.6%	2.6%
I want to have an impact on society. (M3)	64.4%	60.1%	4.3%
I want to help solve a societal issue. (M4)	63.5%	54.0%	9.5% **
It's my hobby. (M5)	63.5%	70.7%	-7.2%
I can learn or improve technology skills. (M6)	53.1%	72.7%	-19.6% ***
I want to help other contributors in the team. (M7)	51.8%	62.6%	-10.8%
It's my job. (M8)	51.8%	53.0%	-1.2%
It helps me to build my portfolio and reputation for my career. (M9)	39.2%	53.5%	-14.3% *
I need to improve this project for my work or school studies. (M10)	29.3%	39.9%	-10.6% ***
I want to meet new people. (M11)	17.6%	30.3%	-12.7% **
My organization/boss encourages me to work on this project. (M12)	15.3%	21.2%	-5.9%
I want to get paid. (M13)	12.6%	13.6%	-1.0%

Wilcoxon rank sum test significance codes: '***' $p < 0.001$, '**' $p < 0.01$, '*' $p < 0.05$

RQ2: Motivations

- **Survey Results**

Q: Do you think ALL OSS are for Social Good?

YES → P-OSS

No → P-OSS4SG

Motivation	P-OSS4SG (n=222)	P-OSS (n=198)	Delta (P-OSS4SG - P-OSS)
I want to help the target users. (M1)	69.4%	65.7%	3.7%
I want to give back. (M2)	66.2%	63.6%	2.6%
I want to have an impact on society. (M3)	64.4%	60.1%	4.3%
I want to help solve a societal issue. (M4)	63.5%	54.0%	9.5% **
It's my hobby. (M5)	63.5%	70.7%	-7.2%
I can learn or improve technology skills. (M6)	53.1%	72.7%	-19.6% ***
I want to help other contributors in the team. (M7)	51.8%	62.6%	-10.8%
It's my job. (M8)	51.8%	53.0%	1.2%
It helps me to build my portfolio and reputation for my career. (M9)	39.2%	53.5%	-14.3% *
I need to improve this project for my work or school studies. (M10)	29.3%	39.9%	-10.6% ***
I want to meet new people. (M11)	17.6%	30.3%	-12.7% **
My organization/boss encourages me to work on this project. (M12)	15.3%	21.2%	-5.9%
I want to get paid. (M13)	12.6%	13.6%	-1.0%

Wilcoxon rank sum test significance codes: '***' $p < 0.001$, '**' $p < 0.01$, '*' $p < 0.05$

RQ3: Factors of Evaluation and Searching Strategies

- ## **Interview highlights**
 - ### **Owners of projects**
    - *"If it's even a charity organization I go and look at who their sponsors are. And if it's a government I'm already like, no, it's not gonna happen. A political party maybe. But government is too far for me."*
  - ### **Searching for a project**
    - *"I can go to like a underground, political event every night and like meet activist, tech people all the time. It's like I can just talk to people."*
    - *"I went to view some of the large releases that they (owners) made...and I checked through their websites also."*
    - *"A friend talk about that or that other project, and you just have a feeling, hey, maybe find you to work on this."*
    - *"I was reading about Coronavirus and tracking the spread of it, and found that there are no mobile apps that provide this ability."*

RQ3: Factors of Evaluation

- **Survey Results**
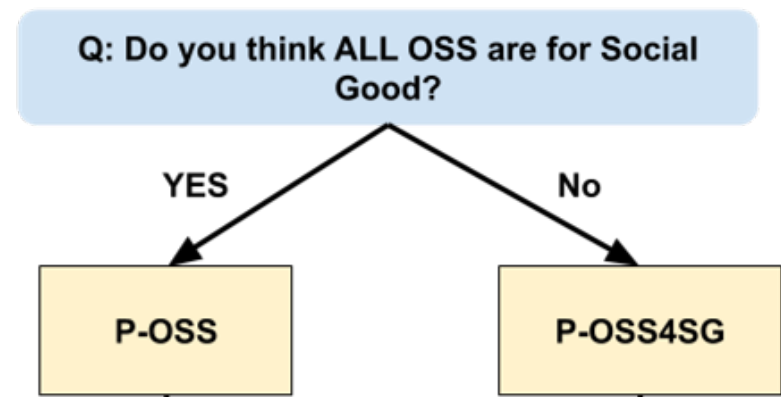
Factors to Consider When Selecting a Project	P-OSS4SG (n=226)	P-OSS (n=202)	Delta (P-OSS4SG - P-OSS)
I personally respect/care about the issue this project is trying to solve. (F1)	83.6%	75.7%	7.9%
I like the idea of this project. (F2)	83.2%	78.2%	5.0%
This project is active. (F3)	74.3%	72.3%	2.0%
The goal of this project meets some form of needs I care about. (F4)	73.0%	67.3%	5.7%
This project is welcoming. (F5)	67.7%	64.3%	3.4%
I fully understand the goal and value of this project. (F6)	65.9%	65.3%	0.6%
I trust the owner/organizer of this project. (F7)	63.5%	42.1%	21.4% ***
I feel confident in my skills to help with this project. (F8)	62.4%	65.8%	-3.4%
This project is well-maintained. (F9)	57.5%	64.8%	-7.3%
I can learn some new skills or enhance my skills in this project. (F10)	43.8%	62.4%	-18.6% ***
This project targets a lot of users. (F11)	24.8%	27.7%	-2.9%
This project has a diverse contributor team (e.g., gender, race, geography). (F12)	16.4%	18.3%	-1.9%
Someone else in my community is also working on this project. (F13)	15.9%	16.3%	-0.4%
This project is popular in the community. (F14)	15.5%	27.2%	-11.7% *

Wilcoxon rank sum test significance codes: '***' $p < 0.001$, '**' $p < 0.01$, '*' $p < 0.05$
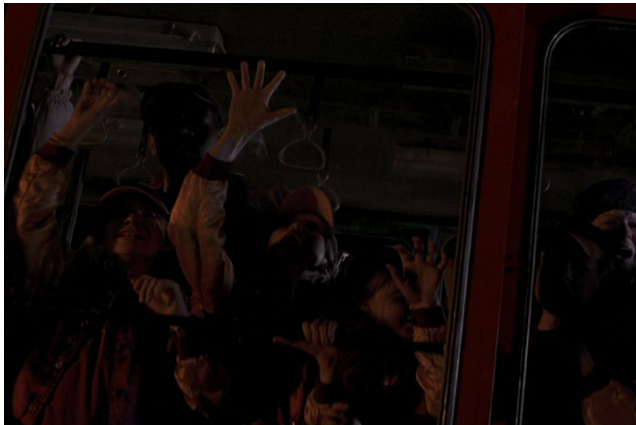
RQ3: Scale of Impact

- **Survey Results**

"Mary Jane, Spiderman, and the Trolley Problem" for Open Source Contributors:
What will you choose to contribute to?

RQ3: Scale of Impact

- **Survey Results**

"Mary Jane, Spiderman, and the Trolley Problem" for Open Source Contributors:
What will you choose to contribute to?

RQ3: Scale of Impact

- **Survey Results**

"Mary Jane, Spiderman, and the Trolley Problem" for Open Source Contributors:
What will you choose to contribute to?

RQ3: Scale of Impact

- **Survey Results: All participants**

"Mary Jane, Spiderman, and the Trolley Problem" for Open Source Contributors: What will you choose to contribute to?
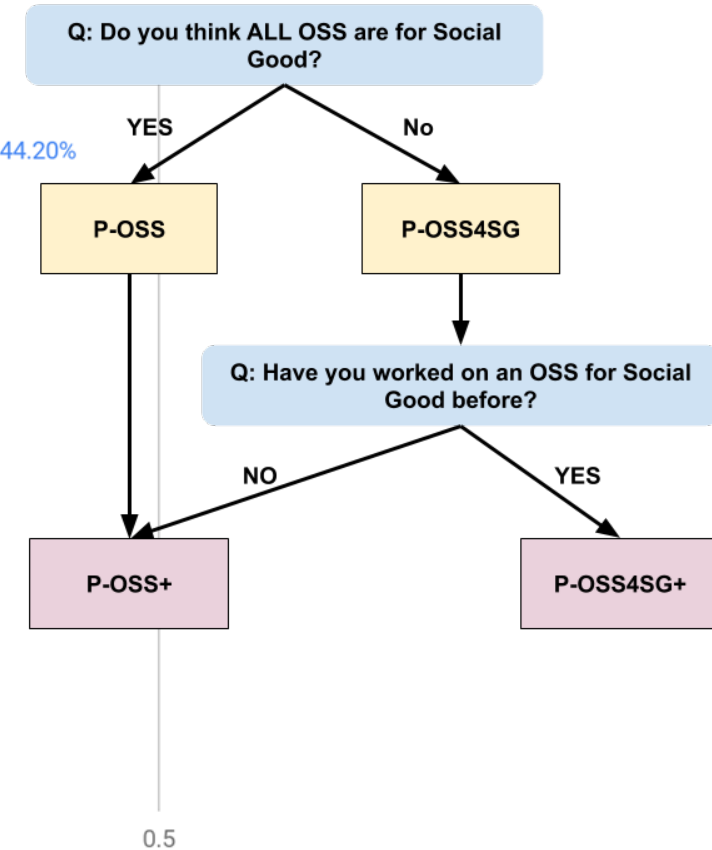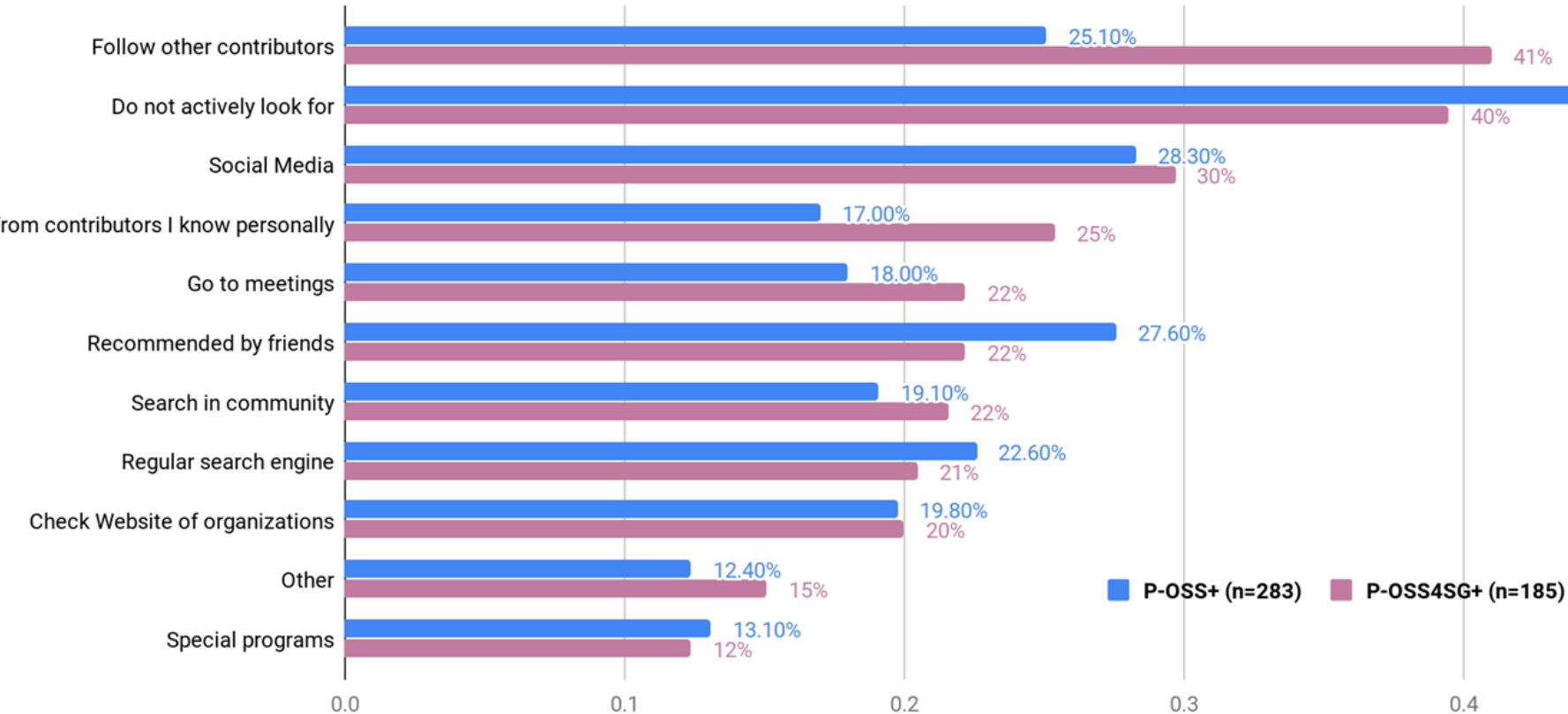
- Spatial Proximity

- Temporal Proximity

- Social Proximity

Project Selection Based on Scale of Impact	Percentage (n=404)
SPATIAL PROXIMITY	
A: A project that is needed globally (e.g., tracking pandemic issues like COVID19)	65.8%
B: A project that is needed only in my local area (e.g., tracking local health issue)	34.2%
TEMPORAL PROXIMITY	
A: A project that is beneficial in the long term (e.g., monitor global warming)	57.7%
B: A project that is beneficial now (e.g., monitor a recent flood disaster)	42.3%
SOCIAL PROXIMITY	
A: A project that a stranger needs (e.g., monitoring system for a health issue that does not affect my family)	31.9%
B: A project that someone I know personally needs (e.g., diabetes tracker my family, friend or myself can use)	68.1%

RQ3: Searching strategies

- **Survey Results**

Channels for Project Searching

Channel	P-OSS+ (n=283)	P-OSS4SG+ (n=185)
Follow other contributors	25.10%	41%
Do not actively look for	44.20%	40%
Social Media	28.30%	30%
From contributors I know personally	17.00%	25%
Go to meetings	18.00%	22%
Recommended by friends	27.60%	22%
Search in community	19.10%	22%
Regular search engine	22.60%	21%
Check Website of organizations	19.80%	20%
Other	12.40%	15%
Special programs	13.10%	12%

Q: Do you think ALL OSS are for Social Good?

YES → P-OSS

No → P-OSS4SG

P-OSS → Q: Have you worked on an OSS for Social Good before?

NO → P-OSS+

YES → P-OSS4SG+

RQ4: Challenges in OSS4SG

- **Interview highlights**
  - **Hard to find a OSS4SG project**
    - *"It is difficult to know where the projects are. Where the communities are. And getting involved in it. **There are many, many, many developers that might want to contribute, but they never get, you know, an announce or publication, a post**, something."*
  - **Lack of stable funding**
    - *"I honestly think the hardest thing about working on social good is very frequently they're funded by charities, so it's very hard to get people's full focus on it. Like, paid full focus on it."*
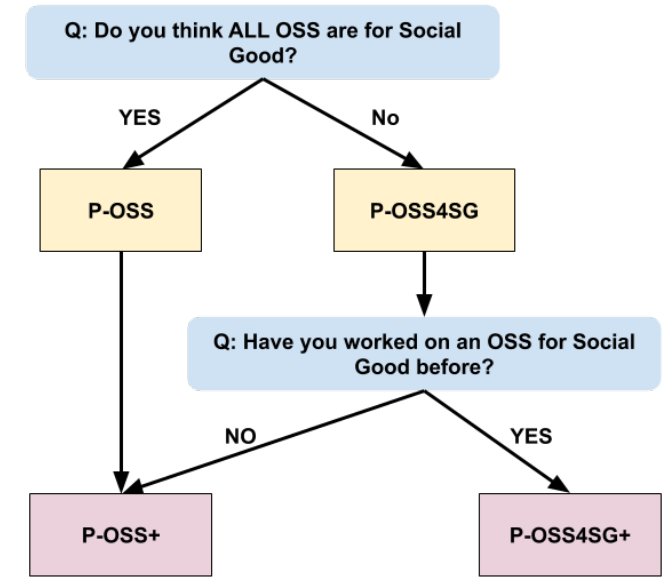 - **Communication with people from different backgrounds**
    - *"In a social good project, we would also have engineers and designers, so it'll take a lot of time to come to a conclusion because there are so many different skill perspectives on the table."*
  - **Lack of OSS concept**
    - *"For developers maybe it's quite common, but for normal people, it's not familiar with that concept [of open source], like that you can contribute something back and that is free, so you like blame the developer for error or something"*
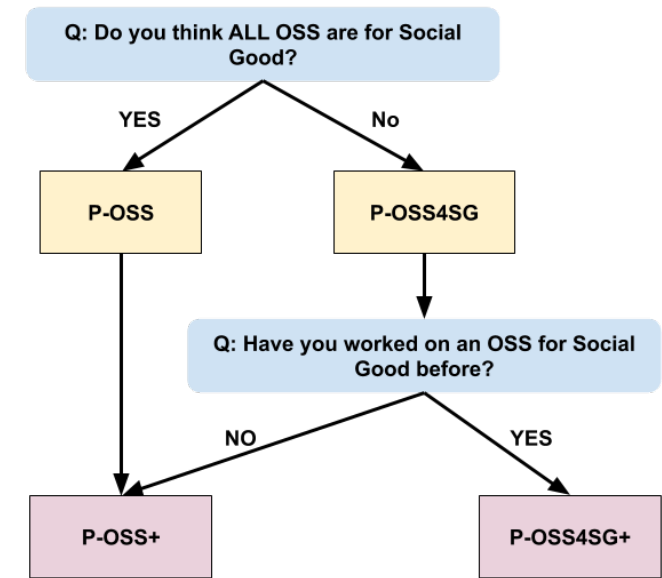
RQ4: Challenges in OSS4SG

- **Survey Results**

Challenges	P-OSS4SG+ (n=153)	P-OSS+ (n=228)	Delta (P-OSS4SG+ - P-OSS+)
It is hard for newcomers to understand how to contribute to the project. (C1)	69.9%	64.5%	4.5%
It is hard to understand what features my users need. (C2)	53.6%	49.1%	4.5%
Not knowing where to find good projects to work on. (C3)	49.7%	36.8%	12.9%
Needing more money to work on a project. (C4)	49.0%	43.0%	6.0%
Not understanding the direction of a project. (C5)	41.8%	34.2%	7.6%
Other contributors losing sight of direction of a project. (C6)	32.0%	35.5%	-3.5%
Working with people who do not understand open source. (C7)	32.0%	33.8%	-1.8%
The project is over-engineered. (C8)	24.8%	24.6%	0.2%
Stakeholders are unreasonable on feature requests. (C9)	22.2%	29.4%	-7.2% **
Too much time is spent on documentation in this projects. (C10)	14.4%	21.1%	-6.7% *

Wilcoxon rank sum test significance codes: '***' $p < 0.001$, '**' $p < 0.01$, '*' $p < 0.05$

RQ4: Challenges in OSS4SG

- **Survey Results**

Challenges	P-OSS4SG+ (n=153)	P-OSS+ (n=228)	Delta (P-OSS4SG+ - P-OSS+)
It is hard for newcomers to understand how to contribute to the project. (C1)	69.9%	64.5%	4.5%
It is hard to understand what features my users need. (C2)	53.6%	49.1%	4.5%
Not knowing where to find good projects to work on. (C3)	49.7%	36.8%	12.9%
Needing more money to work on a project. (C4)	49.0%	43.0%	6.0%
Not understanding the direction of a project. (C5)	41.8%	34.2%	7.6%
Other contributors losing sight of direction of a project. (C6)	32.0%	35.5%	-3.5%
Working with people who do not understand open source. (C7)	32.0%	33.8%	-1.8%
The project is over-engineered. (C8)	24.8%	24.6%	0.2%
Stakeholders are unreasonable on feature requests. (C9)	22.2%	29.4%	-7.2% **
Too much time is spent on documentation in this projects. (C10)	14.4%	21.1%	-6.7% *

Wilcoxon rank sum test significance codes: '***' $p < 0.001$, '**' $p < 0.01$, '*' $p < 0.05$

Insights from P21

- *Technical advisor*
  - No OSS programming experience
  - Profession: Pharmacy
  - Years of experience on collaborating with local government and software development teams

- Insights
  - **Special regulations may apply**
    - *"It takes a long time to set up the global communities around the software...The biggest challenge is to get a political consensus in the given countries that this is what we are going to do and formalize it and making sure that you have national level consensus and a plan, commitment"*
  - **Need a model to convert SG projects to open source**
    - *"So what we have been looking at as a model is [elided software name], or the open source application for district health information system. Because I think they have done a very good job in regards to both publishing the software, but also, creating a community around it that, of course, massively increases sustainability."*

What can we do?

Challenges	P-OSS4SG+ (n=153)	P-OSS+ (n=228)	Delta (P-OSS4SG+ - P-OSS+)
It is hard for newcomers to understand how to contribute to the project. (C1)	69.9%	64.5%	4.5%
It is hard to understand what features my users need. (C2)	53.6%	49.1%	4.5%
Not knowing where to find good projects to work on. (C3)	49.7%	36.8%	12.9%
Needing more money to work on a project. (C4)	49.0%	43.0%	6.0%
Not understanding the direction of a project. (C5)	41.8%	34.2%	7.6%
Other contributors losing sight of direction of a project. (C6)	32.0%	35.5%	-3.5%
Working with people who do not understand open source. (C7)	32.0%	33.8%	-1.8%
The project is over-engineered. (C8)	24.8%	24.6%	0.2%
Stakeholders are unreasonable on feature requests. (C9)	22.2%	29.4%	-7.2% **
Too much time is spent on documentation in this projects. (C10)	14.4%	21.1%	-6.7% *

Wilcoxon rank sum test significance codes: '***' $p < 0.001$, '**' $p < 0.01$, '*' $p < 0.05$

Implications

- Match contributors to OSS4SG projects
  - For OSS community: highlight OSS4SG projects
    - Instrumenting badges and labels: targeted users and social issues
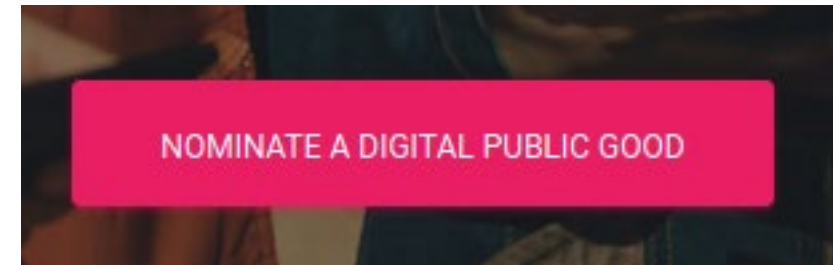
Implications

- Match contributors to OSS4SG projects
    - For OSS community: highlight OSS4SG projects
        - Instrumenting badges and labels: targeted users and social issues
        - Improving nomination guidelines for more effective nominations

Implications

- Match contributors to OSS4SG projects
  - For OSS community: highlight OSS4SG projects
    - Instrumenting badges and labels: targeted users and social issues
    - Improving nomination guidelines for more effective nominations
  - For OSS4SG advocates, project owners, organizers, and fund raisers: clarify and emphasize relevant OSS4SG information to attract contributors of interest
    - OSS4SG contributors care more about project owners, projects' goals, social impact and targeted users
    - project documentation or websites, easily accessible information

Implications

- Match contributors to OSS4SG projects
  - For OSS community: highlight OSS4SG projects
    - Instrumenting badges and labels: targeted users and social issues
    - Improving nomination guidelines for more effective nominations
  - For OSS4SG advocates, project owners, organizers, and fund raisers: clarify and emphasize relevant OSS4SG information to attract contributors of interest
    - OSS4SG contributors care more about project owners, projects' goals, social impact and targeted users
    - Project documentation or websites, easily accessible information

- Protect safety and privacy in OSS4SG
  - Training and reporting structures: sensitive societal issues, marginalized populations of users
    - Signals to potentially risky situations

Implications

- Match contributors to OSS4SG projects
  - For OSS community: highlight OSS4SG projects
    - Instrumenting badges and labels: targeted users and social issues
    - Improving nomination guidelines for more effective nominations
  - For OSS4SG advocates, project owners, organizers, and fund raisers: clarify and emphasize relevant OSS4SG information to attract contributors of interest
    - OSS4SG contributors care more about project owners, projects' goals, social impact and targeted users
    - Project documentation or websites, easily accessible information

- Protect safety and privacy in OSS4SG
  - Training and reporting structures: sensitive societal issues, marginalized populations of users
    - Signals to potentially risky situations
  - Special regulations to protect data privacy: non-traditional entities (gov, charity)
    - Equipped to resolve potentially harmful outcomes from violations of data privacy

Follow-up study

A Four-Year Study of Student Contributions to OSS vs. OSS4SG with a Lightweight Intervention

Zihan Fang
zihan.fang@vanderbilt.edu
Vanderbilt University
USA

Madeline Endres
endremad@umich.edu
University of Michigan
USA

Thomas Zimmermann
tzimmer@microsoft.com
Microsoft Research
USA

Denae Ford
denae@microsoft.com
Microsoft Research
USA

Westley Weimer
weimerw@umich.edu
University of Michigan
USA

Kevin Leach
kevin.leach@vanderbilt.edu
Vanderbilt University
USA

Yu Huang
yu.huang@vanderbilt.edu
Vanderbilt University
USA

- ***OSS4SG*** Pull Requests submitted ***by students*** are ***13%*** more likely to ***be accepted*** than Non-SG
- We checked all ***1.9 million*** Pull Requests from these ***443*** GitHub Projects, ***1.4 million*** Accepted Pull Requests
- We found ***no significant difference*** between the acceptance rates of Non-SG and SG projects in general ($p = 0.23$)

Table 2: Pull Request Acceptance Rate on GitHub

PRs	Non-SG	OSS4SG
Student-Specific	40%	53%
Non-Student-Specific	71%	72%