Gender Diversity Analysis in the
OpenStack Community

Contributors:
Daniel Izquierdo Cortázar, Bitergia
Nicole Huesman, Intel Corporation
Allison Price, OpenStack Foundation
Victoria Martinez de la Cruz, Red Hat

April/May 2018

Bitergia

Foreword

One characteristic of a strong and thriving open source community is an active effort to attract new talent, diverse skill sets, and fresh perspectives. Mentorship programs are especially important to help newcomers -- as well as those working in new areas -- to engage and contribute successfully. Mentors play a critical role in closing the gap in the underrepresentation of females and minorities across our collective community.

That's why I'm pleased to see in this edition of the OpenStack Diversity Report the wealth of mentorship programs serving the OpenStack community, including Outreachy, Google Summer of Code, Long Term Mentoring and Speed Mentoring programs, as well as the Upstream Institute. Similar to the ongoing diversity research and analysis conducted by Bitergia and sponsored by Intel, the aim of this report is to share key insights and best practices for use by the community. We hope you find this data informative and, more importantly, actionable.

Thank you to the OpenStack Foundation for continued support of this work and openness to recommendations derived from this research. Additional thanks to the leads of these mentorship programs; without them, this report would not be possible. Finally, I am grateful to the community for their interest and input. We welcome continued engagement to help improve this research in ways that will benefit the communities we participate in together.

Imad Sousou
Corporate Vice President and General Manager
Intel Open Source Technology Center

Table of Contents

Executive Summary

This report edition provides updated data through April 2018, pinpointing any considerable variances that have occurred since the date of the last report in October 2017. In the spirit of continual evolution of this research and analysis, this edition also highlights new data focused on mentorship programs that serve the community.

The topline numbers have remained relatively steady over the time of these reports. The most notable variance is the increase in female representation among elected members of the User Committee, from 0% to 20%. In response to community feedback, this report includes the representation of female session speakers alongside keynote speakers at the OpenStack Summits. Women comprised a healthy 12% of attendees over the last three summits, accounting for a 22% average of keynote speakers and a 12% average of session speakers. Given that data has remained consistent over time, the authors of this report recognize the need to pair qualitative with quantitative research to deliver greater insight into the numbers, and this work is noted within Further Work and Recommendations within this report.

This report marks a starting point for research and analysis of OpenStack mentorship programs, with the aim to identify and share best practices. It includes an overview of several of the programs, with a deeper dive into the Outreachy and Google Summer of Code (GSoC) programs. The research reveals that newcomers who enter the community through a mentorship program, such as Outreachy and GSoC, boast a higher retention rate at 13% than other newcomers in the community at lower than ten percent. In addition, mentees who progress through these programs contribute to the community for a far longer time period, averaging 15 months, than the duration of their program involvement of three months.

The majority of research analysis and results are based on publicly available data sources. Git and Gerrit repositories are used as the sources for code contributions, while leadership and governance data is derived from the OpenStack website and mailing lists. Event attendee and speaker representation was provided by the OpenStack Foundation, and mentorship program data was provided by the leads of the respective programs or derived from the program websites, with additional filtering through the aforementioned Git, Gerrit and mailing list sources.

1. Leadership & Governance

This section focuses on the gender diversity from a leadership and governance perspective in the OpenStack community. The most notable change in representation is among elected positions within the User Committee, from 0% to 20%, since the time of the last report. Other representation percentages have held steady over time. Females comprise a considerable percentage of OpenStack Ambassadors at 20%, followed by presence on Working Groups at 15.8%, elected Technical Committee positions at 15.4%, and Board of Directors at 13%. Conversely, female representation continues to lag among Project Team Lead (PTL) positions at 3%.

The sources for the data in this section include the OpenStack website and the Project Teams Governance file. These sources provide varying data--some include company associations for individuals, whereas others do not. Subsequently, some of the charts within this report include company affiliation, while others do not.

OpenStack Foundation Officers

The OpenStack Foundation's goal is "to serve developers, users, and the entire ecosystem." At the time of this report, 25% of OpenStack Foundation officers are women[1], consistent with the last report. Current officers include:

Officer	Job Title
Jonathan Bryce	Executive Director
Mark Collier	Chief Operating Officer
Lauren Sell	VP, Marketing & Community Services
Thierry Carrez	VP, Engineering

Table: OpenStack Foundation Officers

OpenStack Board of Directors

The Board of Directors "provides strategic and financial oversight of Foundation resources and staff."[2] At the time of this report, females comprise approximately 13% of Board membership, or 3 of 23 members, which has remained relatively steady over time. The list of members is as follows:

Member	Co. Affiliation	Member	Co. Affiliation

---

[1] https://www.openstack.org/foundation/staff/
[2] https://www.openstack.org/foundation/board-of-directors/

Alan Clark	SUSE	Joseph Wang	inwinSTACK
Allison Randal	Ericsson	Kenji Kaneshige	Fujitsu
Anni Lai	Huawei	Kurt Garloff	T-Systems International GmbH
Arkady Kanevsky	Dell EMC	Lew Tucker	Cisco Systems
Boris Renski	Mirantis	Mark McLoughlin	Red Hat Inc.
Brian Stein	Rackspace	Monty Taylor	Red Hat
ChangBo Guo	EasyStack	Prakash Ramchandran	Dell
Christopher Price	Ericsson AB	Robert Esker	NetApp
Egle Sigler	Rackspace	Shane Wang	Intel
Gnanavelkandan Kathirvel	AT&T	Steven Dake	Cisco Systems
Imad Sousou	Intel	Tim Bell	CERN
Johan Christenson	City Network Hosting AB		

Table: OpenStack Foundation Board of Directors

Technical Committee

The Technical Committee is a "fully-elected committee that represents the contributors to the project" whose primary purpose is "to provide technical leadership for OpenStack as a whole." This committee is totals 13 people, 2 of them females (15.4%)[3]. Current roster is as follows:

Member	Co. Affiliation	Member	Co. Affiliation
Chris Dent	VMware	Jeremy Stanley	OpenStack Foundation
Colleen Murphy	SUSE	John Garbutt	StackHPC
Davanum Srinivas	Huawei	Julia Kreger	Red Hat Inc.
Dean Troyer	Intel	Paul Belanger	Red Hat Inc.

---

[3] https://www.openstack.org/foundation/tech-committee/

Doug Hellmann	Red Hat Inc.	Sean McGinnis	Huawei
Emilien Macchi	Red Hat Inc.	Thierry Carrez	OpenStack Foundation
Flavio Percoco	Red Hat Inc.	-	-

Table: OpenStack Foundation Technical Committee

User Committee

The User Committee is "led by a core group of five individuals, who provide oversight and guidance to a number of working groups that target specific areas for improvement."[4] Currently, females account for 20% of this 5-person committee. The list of members is as follows:

Member	Co. Affiliation
Amy Marrich	Linux Academy
Dr Yih Leong Sun	Intel Corp.
Matt Van Winkle	Rackspace
Melvin Hillsman	Huawei
Saverio Proto	Switch

Table: OpenStack Foundation User Committee

Project Technical Leads (PTLs)

Project Team Leads (PTLs) are elected every six months to govern each OpenStack project, as specified in the Governance YAML file by the Foundation.[5] At the time of this report, of the total 65 PTLs, 2, or approximately 3%, are women. A similar percentage are of unknown gender. The following is the list of PTLs:

PTL Name	Project	PTL Name	Project
Ade Lee	Barbican	Graham Hayes	Designate
Masahito Muroi	Blazar	Petr Kovar	Documentation
Samuel Cassiba	Chef OpenStack	Omer Anson	Dragonflow

---

[4] https://www.openstack.org/foundation/user-committee/
[5] http://git.openstack.org/cgit/openstack/governance/tree/reference/projects.yaml

Jay Bryant	Cinder	Andrey Pavlov	Ec2-api
Christophe Sauthier	Cloudkitty	Saad Zaher	Freezer
Eric Kao	Congress	Erno Kuvaja	Glance
Zhipeng Huang	Cyborg	Rico Lin	Heat

PTL Name	Project	PTL Name	Project
Ivan Kolodyazhny	Horizon	Dean Troyer	OpenStackClient
Frank Kloeker	I18n	Monty Taylor	OpenStack SDK
Clark Boylan	Infrastructure	Ben Nemec	Oslo
Julia Kreger	Ironic	Javier Peña	Packaging-rpm
Chen Ying	Karbor	Matthew Edmonds	PowerVMStackers
Lance Bragstad	Keystone	Mohammed Naser	Puppet OpenStack
Jeffrey Zhang	Kolla	Lingxian Kong	Qinling
Daniel Mellado	Kuryr	Ghanshyam Mann	Quality Assurance
Sam Yaple	Loci	Andrey Kurilin	Rally
Spyros Trigazis	Magnum	Chris Hoge	RefStack
Tom Barron	Manila	Sean McGinnis	Release Management
Sampath Priyankara	Masakari	Matthew Thode	Requirements
Dougal Matthews	Mistral	Telles Mota Vidal Nóbrega	Sahara
Witold Bedyk	Monasca	Steve McLellan	Searchlight
Rong Zhu	Murano	Luke Hinds	Security
Miguel Lavalle	Neutron	XueFeng Liu	Senlin
Melanie Witt	Nova	Rong Zhu	Solum
Michael Johnson	Octavia	Tony Breeds	Stable Branch

			Maintenance
James Page	OpenStack Charms	Kota Tsuyuzaki	Storlets
Matt McEuen	OpenStack-Helm	John Dickinson	Swift
Jean-Philippe Evrard	OpenStackAnsible	Yong Sheng Gong	Tacker

PTL Name	Project	PTL Name	Project
Julien Danjou	Telemetry	Alexander Chadin	Watcher
Zhiyuan Cai	Tricircle	Claudiu Belu	Winstackers
Alex Schultz	Tripleo	Wang Hao	Zaqar
Zhao Chao	Trove	Feng Shengqin	Zun
Ifat Afek	Vitrage	-	-

Table: OpenStack Foundation PTLs

Working Groups

There are a number of working groups that the User Committee assists.[6] These working groups and their co-chairs are listed in the chart below. At the time of this report, of the total 6 co-chairs, 1, or 16.66%, are women.

Working Group	Co-Chairs
Fault-Genes Working Group	Nemat Bidokhti, Rochelle (Rocky) Grober
LCOO	Jamey McCabe
Public Cloud	Tobias Rydberg, Howard Huang
Diversity Working Group	Amy Marrich

Table: OpenStack Foundation Working Groups

---

[6] https://wiki.openstack.org/wiki/Governance/Foundation/UserCommittee#Working_Groups

Special Interest Groups

In addition to the working groups, there are various special interest groups[7] (SIGs) that "collaborate to bring unified discussions for all community members who share a common interest", from the initial stages of a new project to standards adoption, policy implementation, adjacent community work, and more. Of these groups, females comprise 10.5% of co-chairs, or 2 of 19 people.

---

[7] https://governance.openstack.org/sigs/

Special Interest Group	Co-Chairs
API	Chris Dent, Michael McCune, Ed Leafe
FEMDC	Adrien Lebre, Paul-André Raymond
First Contact	Kendall Nelson
K8s	Chris Hoge, David Lyle, Robert Morse
Meta	Melvin Hillsman, Thierry Carrez
Scientific	Blair Bethwaite, Stig Telfer, Martial Michel
Self-healing	Adam Spiers, Eric Kao
Upgrade	James Page, Lee Yarwood, Lujin Luo

Table: OpenStack Foundation Special Interest Groups

OpenStack Ambassadors

OpenStack ambassadors[8] are globally distributed and help "tie user groups together, and work with each one to mentor it to be the best it can be". Of the total 15 ambassadors, 3, or 20% are women. Current list of ambassadors:

Ambassador	Ambassador
Akihiro Hasegawa	Lisa-Marie Namphy
Akira Yoshiiyama	Lu Ye
Andy Botting	Marcelo Dieder
Christian Berendt	Márton Kiss
Erwan Gallen	Sajid Akhtar
Ilya Alekseyev	Shilla Saebi
Jaesuk Ahn	Stacy Véronneau
John Studarus	-

[8] https://groups.openstack.org/ambassador-program

2. Event Representation

Event Attendance

Over the last three OpenStack Summits and Project Team Gatherings (PTGs), female attendance has maintained a healthy average of 11-12%.

Speaker Representation

Over the last three OpenStack Summits, females have comprised a considerable percentage of speakers, mostly notable among keynotes at 19%, or 14 of a total 74. Among session speakers, female accounted for 11.66%.

OpenStack Summit	Attendance	Keynote Speakers	Session Speakers
Sydney Summit (Oct 2017)	13%	10%	9%
Boston Summit (May 2017)	12%	19%	12%
Barcelona Summit (Oct 2016)	11%	26%	14%

Table: OpenStack Summit women attendance and speaker representation

Project Team Gathering	Attendance
Dublin PTG (Feb 2018)	10%
Denver PTG (Sep 2017)	12%
Atlanta PTG (Feb 2017)	10%

Table: Openstack Project Team Gatherings women attendance

3. Mentorship Programs

This report incorporates a new section that focuses on membership programs, and their importance to healthy open source communities. The retention rate of newcomers who enter the OpenStack community through mentorship programs is particularly promising, as exemplified by the Outreachy program, exceeding that of other newcomers whose retention rate is lower than ten percent. In addition, mentorship program participants contribute to the community for a time period that far exceeds their 3-month program duration, averaging 15 months.

Outreachy

The Outreachy program[9] brings technology to underrepresented minorities, which helps to foster the adoption of open source technologies such as OpenStack. In the first 10 editions, 38 students from 13 different countries have contributed to OpenStack across 17 projects. This report studies the time period between Outreachy's inception and May/August 2017.
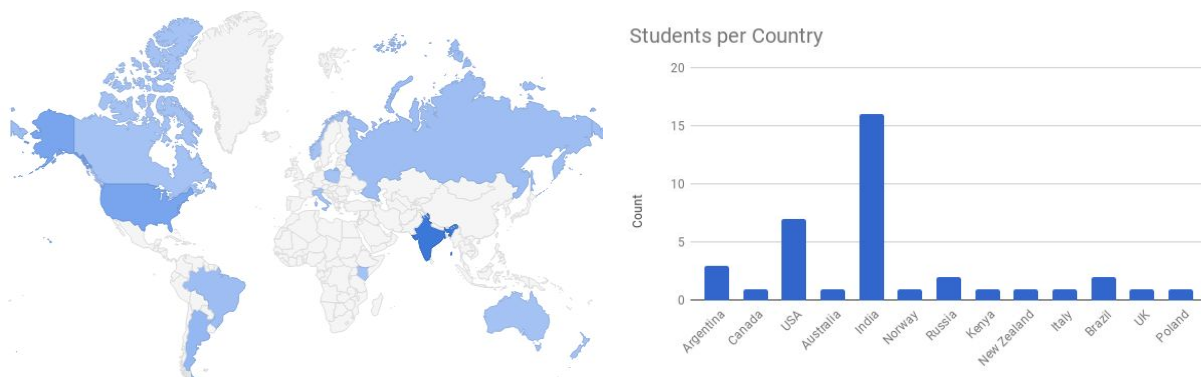

Figure: Distribution of Outreachy students who have contributed to OpenStack

Note: For the purposes of this research, a developer is defined as active if a single trace of her activity, including commits, changesets or emails, is located in the repositories, or dataset.

<u>Retention Rate</u>

In examining the retention rate of newcomers to the community, the Outreachy program shows a 13.15% retention rate, where five of 38 students are still active in the community. Average retention time is 18 months, while median retention time is 11.5 months, which indicates that developers tend to contribute far longer than the length of their duration in the program.

Note: The minimum activity period is 0 where no traces were detected across the maximum length of 65 months. The number of months is calculated as the difference between the time periods during which the first and last traces of activity are detected. This does not indicate the

---

[9] https://wiki.openstack.org/wiki/Outreachy

total number of consecutive months of activity, but instead, the time period during which the developer has contributed to the community and has since become inactive.

During the time period of analysis, the trend in retention rates of other newcomers typically falls below ten percent, as reflected in the median retention rates shown in the figure below.

Note: The retention rate is calculated as the number of developers currently active (within the last six months) compared with the total number of developers who entered the community during each of the Outreachy rounds. The months selected for each of the Outreachy rounds are the months during which the first traces of activity from the Outreachy participants is detected.
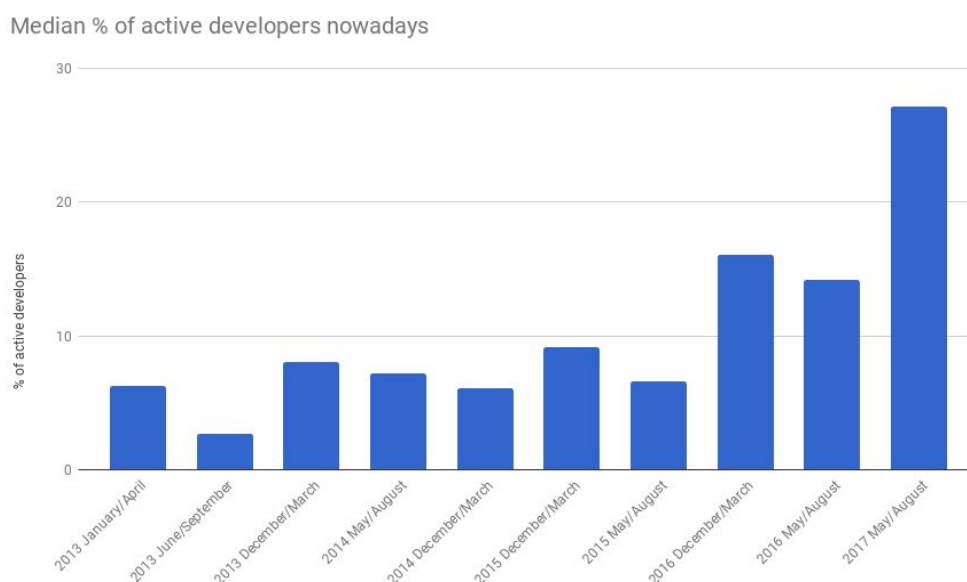
Median % of active developers nowadays

Figure: Median retention rates of all active newcomers (including Outreachy students)

The figure below shows the retention rates of Outreachy students, by round. The total retention rate of the Outreachy program for the time period analyzed exceeds 13%, higher than the general retention rate of other newcomers during this same time period.
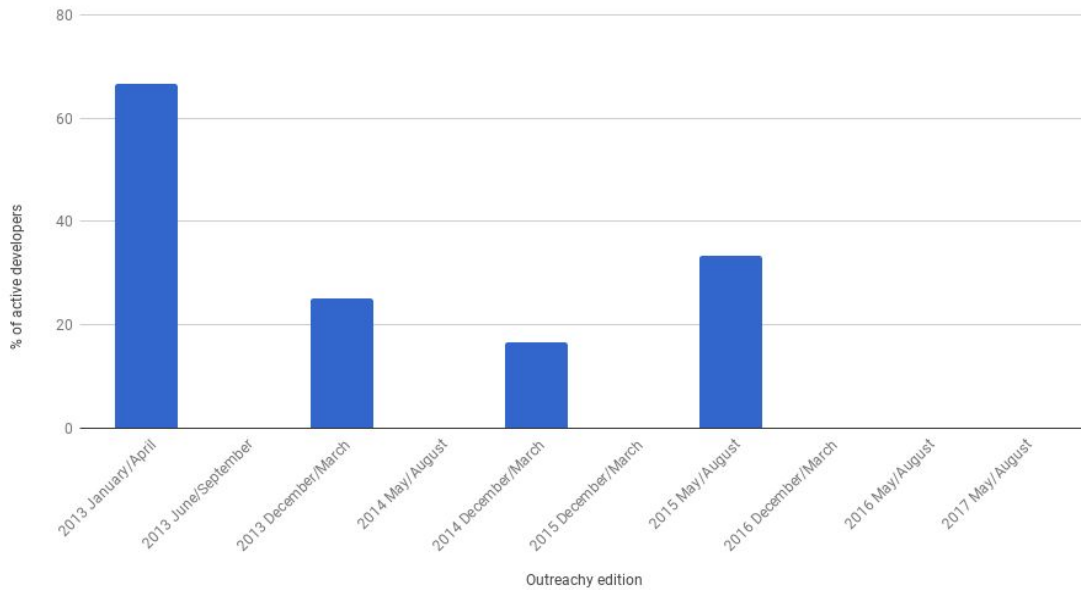
% of Active Developers nowadys

Figure: Percentage of currently active Outreachy developers from specified Outreachy rounds

Contributions

Since its inception, Outreachy participants have contributed extensively to the OpenStack community, across 17 projects, as shown in the chart below.
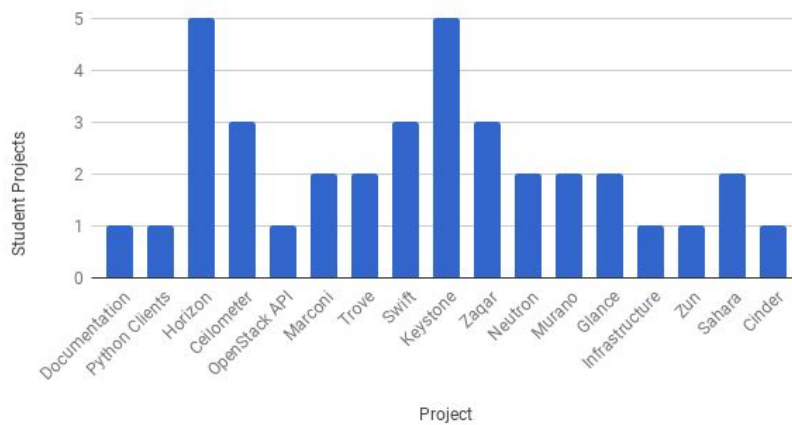

OpenStack projects involved in Outreachy

Figure: Number of students assigned to each of the OpenStack projects

While their activity level varies from year to year, the typical aggregate ratio of Outreachy participants' contributions, dependent on round and number of students, is fewer than 150 events, as measured by commits, changeset submissions and emails, shown in the chart below.

Naturally, those students in Outreachy's inaugural year have demonstrated the highest number of contributions, as they have had a longer time period to contribute. The remaining students register a lower level of aggregate activity at fewer than 50 events each.
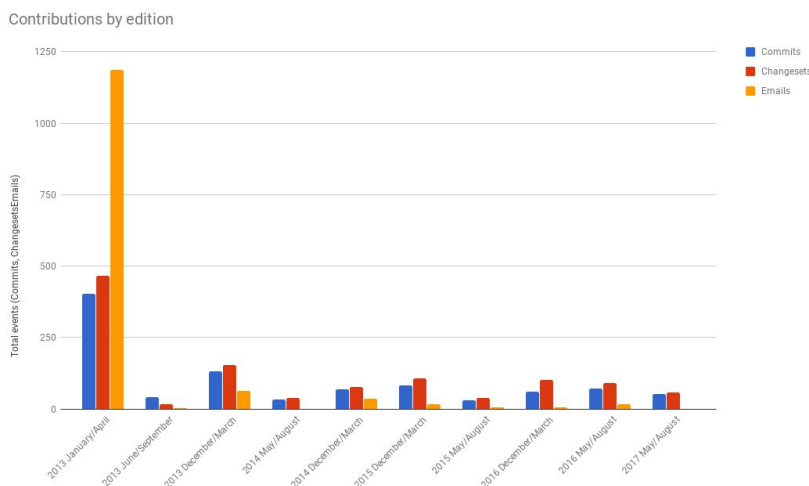

Figure: Aggregate contributions from Outreachy participants, between 2013 and 2017

Google Summer of Code

The Google Summer of Code[10] (GSoC) aims at "bringing more student developers into open source software development. Students work with an open source organization on a 3-month programming project during their break from school".

The OpenStack Foundation hosted six (6) GSoC mentees in 2014. Their activity totaled 185 commits, 204 code reviews, and 263 emails across four (4) OpenStack projects. Each student accounted for an average 31 commits, 34 open changests, and 44 emails. Median duration of their contributions (both inside and outside their 3-month program duration) was six (6) months, whereas the average duration of their contributions reached nearly 15 months, with one student still active in the community. This time length of contributions reflects an average retention rate between 5% and 7.5% and a median retention rate of 6.25%.

Note that there is slight duplication in these numbers as one of the students participated in both this program as well as the Outreachy program.

---

[10] https://wiki.openstack.org/wiki/Category:GSoC

No. of Students	No. of Projects	No. of Contributions	Retention Rate
6 (in 2014)	4	652 events (commits + changesets + emails)	5-7.5% average 6.25% median

Table: Summary of GSoC participation

Other Programs

In addition to the Outreachy and Google Summer of Code (GSoC) programs examined in this report, several other programs and activities exist to welcome newcomers into the OpenStack community. These include but are not limited to::

- **First Contact Special Interest Group[11]:** With the recognition that new contributors are the future of OpenStack and the surrounding community, this group aims to make it easier for these newcomers to engage in the OpenStack project and community by welcoming them and giving them information, advice and the tools they need to succeed.

- **Speed Mentoring Workshop[12]:** This workshop, hosted at each OpenStack Summit, brings together contributors with at least one year of experience in the community with newcomers or those interested in working in new areas of the community. Each workshop is comprised of three (3) lightning rounds of mentor/mentee interaction across career, technical and community tracks.

- **Long Term Mentoring[13]:** This program connects mentors with mentees, focusing in two main areas: (a) technical mentorship where mentors help newcomers in specific areas of the project, and (b) career mentorship where mentors help newcomers to define work they would like to do in the project. At the time of this report, this program is evolving and exploring the use of a cohort-style mentorship program.

- **Upstream Institute[14]:** Due to the size of its community, it is characterized by a large diversity in social norms and technical conventions, which can significantly hinder the speed at which newcomers successfully engage and participate. The Upstream Institute is a training program designed to share knowledge about the different ways to contribute to OpenStack, including providing new features, writing documentation, participating in working groups, and more. Hosted a day prior to OpenStack Summits and at select OpenStack Days, this live, 1.5-day class teaches students how to find information, and navigate the intricacies of the project's technical tools and social interactions in order to get their contributions accepted. Attendees are also given the opportunity to join a mentoring program to get further help and guidance on their journey to becoming active, successful members of the OpenStack community.

---

[11] https://wiki.openstack.org/wiki/First_Contact_SIG
[12] https://wiki.openstack.org/wiki/Mentoring
[13] https://wiki.openstack.org/wiki/Mentoring
[14] https://docs.openstack.org/upstream-training/

4. Code Contributions

Source Code Contributions

The analysis of source code contributions has remained relatively steady over time, along the following dimensions:

- Top projects, as measured by number of total female participants
- Percentage of commits authored by women, men and unidentified genders during the last year
- Percentage of population authoring such commits
- Evolution of female-authored over the last five years
- Evolution of female participation over the last five years
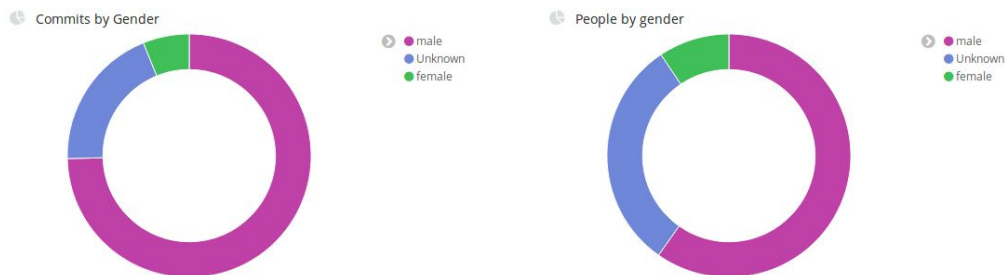

Figure: Left chart represents breakdown of commits by women (6.5%), men (74.62%) and unidentified over the last year. Right chart represents author breakdown of commits by women (9.5%), men (59.8%) and unidentified members of the community.
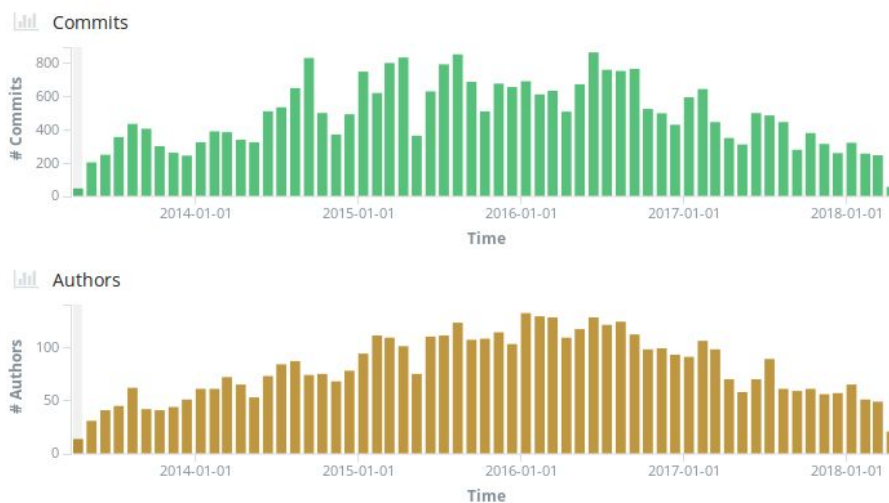
Figure: Number of female-authored commits (top chart) and female participants (bottom chart) over the last five years

Based on the number of female authoring commits over the last year, these are the projects with the greatest gender diversity:

Project ⇵ Q	Commits ⇵	Authors ⌄	Repositories ⇵
Infrastructure	1,242	49	99
cinder	205	33	4
QualityAssurance	195	28	5
tripleo	419	19	15
nova	51	18	4
Documentation	87	17	4
requirements	23	15	1
neutron	55	15	12
horizon	60	15	2
ironic	318	15	15

Figure: Top 10 projects, as measured by females authoring commits over the last year

Code Review Contributions

The analysis of code review contributions has also remained relatively steady over time, along these dimensions:

- Number of code review submissions by gender over time
- Number of developers who have submitted changesets, by gender over time
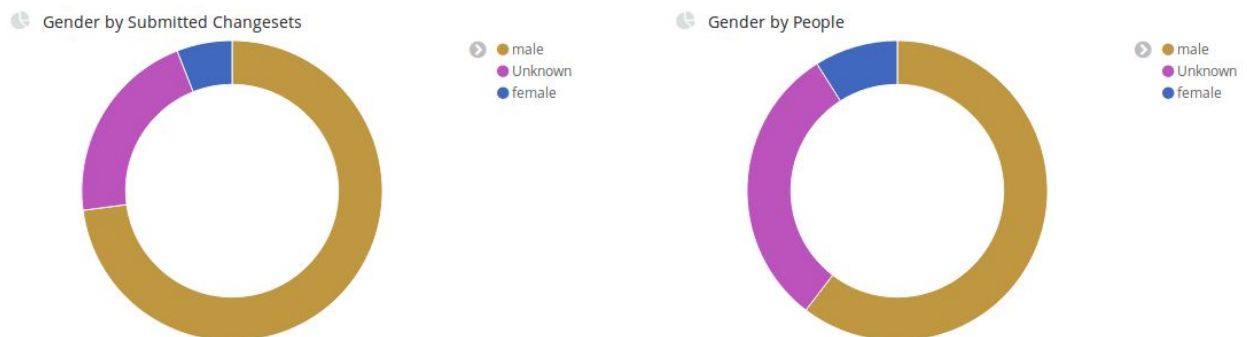
Figure: Left chart represents breakdown of changesets submitted by women (5.95%), men (73.9%) and unidentified over the last year. Right chart represents author breakdown of changesets by women (9%), men (60.5%) and unidentified members of the community.
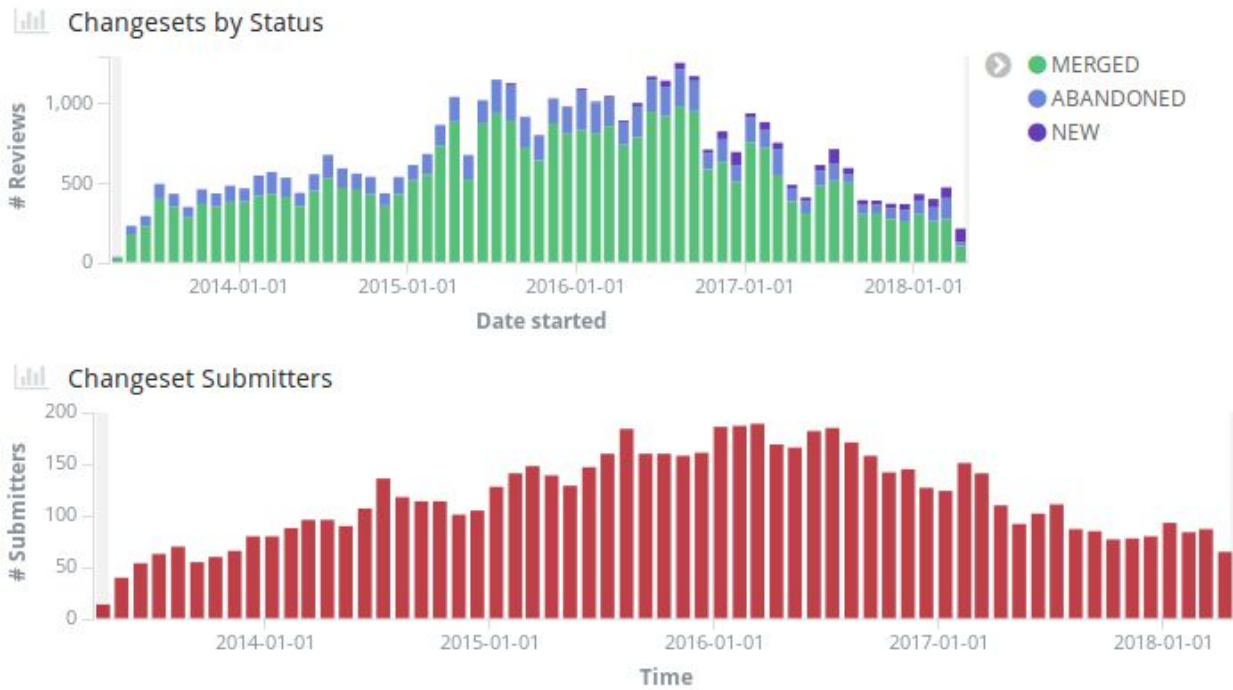
Changesets by Status

Changeset Submitters

Figure: Number of female-submitted changesets (top chart) and female participants (bottom chart) over the last five years

Based on the number of females submitting changesets over the last year, these are the projects with the greatest gender diversity:

Repositories

Repository Q	# Changesets	# Submitters
openstack-dev/sandbox	424	190
openstack/stackalytics	302	173
openstack/nova	1,219	169
openstack/openstack-manuals	2,853	137
openstack/neutron	944	119
openstack/horizon	1,153	114
openstack/cinder	944	111
openstack-infra/project-config	736	106
openstack/tempest	922	102
openstack-dev/devstack	332	90

Figure: Top 10 projects, as measured by females submitting changesets over the last year

5. Mailing List Representation

In addition to measuring female participation in leadership and governance, event attendance and speaker representation, and code contributions, this report also captures female representation within OpenStack community mailing lists, both in terms of activity level and population, as an indicator of their overall engagement. From this perspective, females comprise a healthy percentage of mailing list activity, as depicted in the charts below.
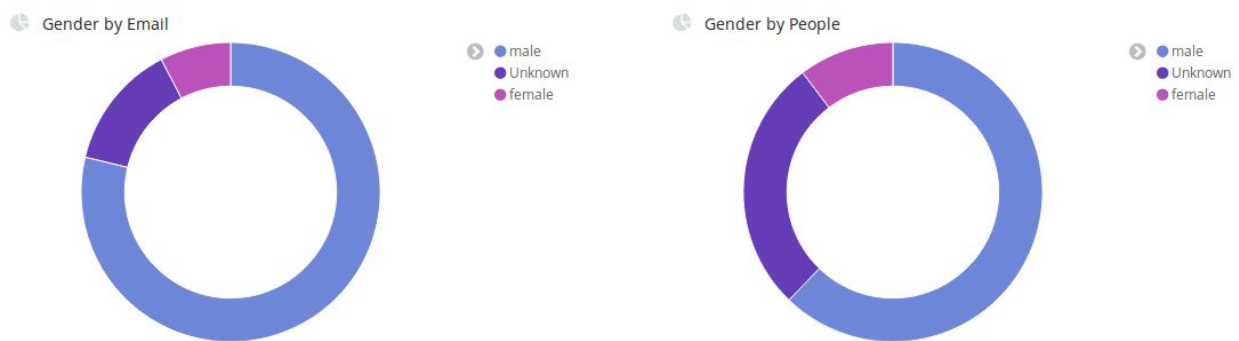

Figure: Left chart represents breakdown of emails authored by women (7.67%), men (78.76%) and unidentified over the last year. Right chart represents author breakdown of emails by women (10.3%), men (62.21%) and unidentified members of the community.
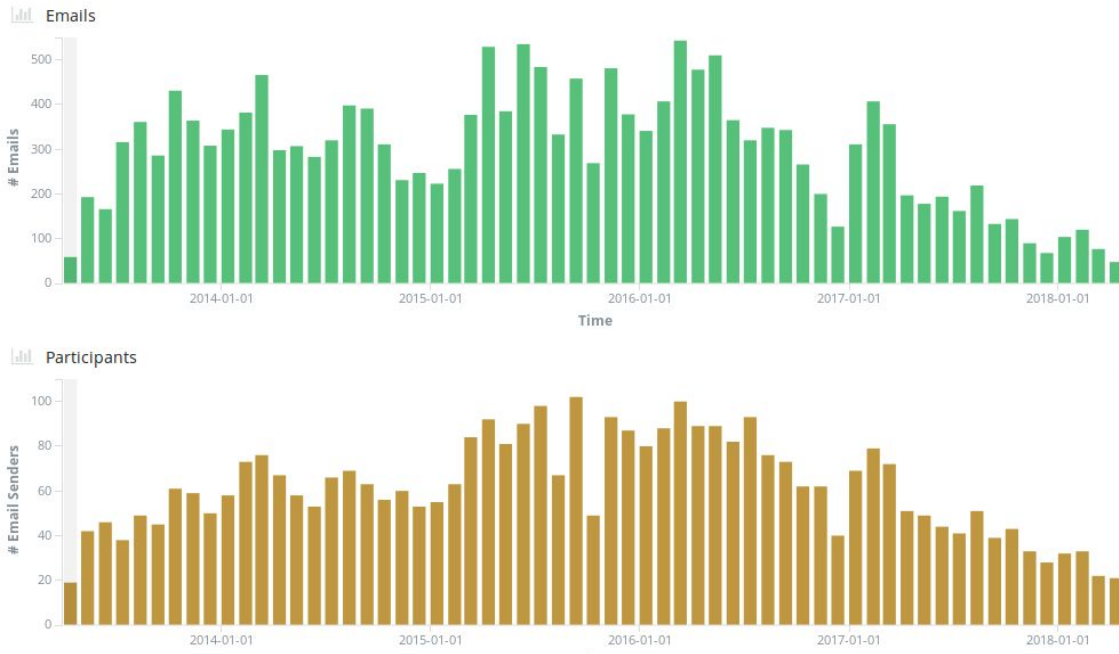
Figure: Number of female-authored emails (top chart) and female participants in mailing lists (bottom chart) over the last five years

6. Discussion and Further Work

Over its duration, the OpenStack gender diversity research been useful in bringing attention to the issue of underrepresented females, and in turn, other minorities. The report has led to fruitful discussions at several OpenStack Summits, as well as other venues including the Open Source Leadership Summit, the Open Source Summits in North America and Europe (formerly LinuxCon), and OSCON. It has also helped form the basis of the CHAOSS Project's Diversity & Inclusion Work Group, in part.

Through these discussions, we have received invaluable input that has enabled us to further this research. At its inception, this research incorporated both code and non-code contributions in recognition of their importance to a healthy, thriving community. Non-code contributions initially included leadership and governance, committees, working groups and ambassadors, as well as event attendance and keynote representation. We have continued to evolve the research parameters based on foundation and community input to include special interest groups, session speaker representation, and mailing list activity. In addition, this latest report places a spotlight on the importance of mentorship programs.

This analysis of mentorship programs aims to reveal insights into one of the entry points of any community, using the OpenStack community as an example. Due to the lack of data, the analysis focused solely on the Outreachy and Google Summer of Code (GSoC) and programs, yet this study has brought interesting results. The retention rate of newcomers who enter the community through a mentorship program appears to be higher than other newcomers, at 13% and less than ten percent, respectively. Though these numbers may be biased by the size of the groups analyzed, 5 of 38 Outreachy students continue to contribute to the OpenStack community today. This glimpse into the analysis of mentorship programs is intended as a starting point only; one of the next steps is to understand the rationale behind the numbers. A potential hypothesis may be that the creation of solid social networks is a key factor in the retention of new contributors.

The numbers have remained relatively consistent over time, providing us with a starting point against which to measure our collective progress in furthering diversity and inclusion across the open source community. Typically, females account for ten percent of open source communities. Their activity levels range from 7-8% of code contributions up to 20% of leadership and governance.

This consistency in numbers has brought an open discussion around the need to delve deeper into understanding them through the use of qualitative research. While the numbers have brought awareness, we must now truly comprehend what they mean to develop meaningful action plans that help to create environments that attract and retain diverse talent and unique perspectives.

CHAOSS Diversity & Inclusion Working Group

In September 2017, the Linux Foundation introduced the CHAOSS[15] (Community Health Analytics for Open Source Software) Project, focused on the ability to measure the health of open source communities. Since then, we spearheaded the project's Diversity and Inclusion Working Group[16], in part, thanks to this research, which is is comprised of several organizations and open source foundations, including the OpenStack Foundation. The mission of this work group is create a consistent, repeatable set of metric categories across open source communities by which to measure the level of diversity and inclusivity within a given community. Its current focus is solidifying specific goals using a goal/question/metric approach.

The ongoing OpenStack gender diversity research and this new work group share a symbiotic relationship. While this research, complemented by other community work such as that conducted by Mozilla, has helped to inform this working group, insights from this working group have, in turn, helped to evolve this research. Examples include the need for a qualitative understanding of the mentor-mentee relationship, current perceptions of inclusivity, and other areas of further work.

Complexities in Scaling the Research

While we are actively addressing the need to scale the research by identifying a consistent, repeatable set of metric categories through the CHAOSS Diversity & Inclusion Working Group, we have faced several complexities in scaling this research. These issues include the limitation of software capabilities to automate data collection and analysis, the introduction of additional data sources, the protection of sensitive information, and the polishing and curation of data.

Given the current size of the analysis with thousands of developers and hundreds of analyzed repositories, it is not scalable to review all of the members of the community. In addition, gender diversity is not a binary issue, complicating its analysis. In light of the privacy of the dataset, and the peculiarities of the data used, open discussion remains around the topic of proper handling of the data and allowing others to improve their own datasets. Moreover, compliance with several personal data laws, including the new GDPR law within the European Union, has created additional complexities.

Collaboration with Other Teams and Working Groups

In addition to continually evolving this research and better understanding the quantitative results through qualitative research, the authors of this study also view deeper collaboration with other teams and working groups, such as the OpenStack Diversity Work Group, as a next step. Through deeper collaboration, we can learn from each other, share insights and identify and disseminate best practices to benefit the entire open source community.

---

[15] https://chaoss.community
[16] https://github.com/chaoss/wg-diversity-inclusion

Appendix A: Technical Details and Limitations

This appendix outlines the sources and methodologies for the research within this report. It also identifies potential limitations of this research, and proposes ways to counter these limitations.

- The data for this report from Git repositories is retrieved from the governance file using the command below, and is parsed and stored in an ElasticSearch instance.:
  ```
  git log --raw --numstat --pretty=fuller --decorate=full --parents
  --reverse --topo-order -M -C -c --remotes=origin --all
  ```
- The data for this report from Gerrit repositories is retrieved from the governance file using an SSH interface, and is parsed and stored in an ElasticSearch database.
- The data for this report from both Git and Gerrit repositories is retrieved using Perceval[17], a 100% open source software tool under the GrimoireLab[18] umbrella.
- The code used to enrich and visualize the raw information in this report is provided by GrimoireLab that at the moment of this analysis produces by default gender-diversity analysis. In this report, gender identification is based on the individual's first name, which relies on the genderize.io API. However, this report also required manual manipulation of the datasets (e.g. manual updates and improvements), which proved to be extremely time-consuming. One of the goals of this project is to capture an increasingly more curated dataset, which will require assistance and support from the community. The need for protection of privacy of the data brings complexities, such as the need for handling of the data by a third party and the restriction of access to the dataset.
- In this report, code contributions are defined as commits, and code reviews. This is a starting point for this ongoing research, based on input from the community. The list of non-code related contributions are initially mailing lists, although other sources for measurement of female activity, engagement and contributions can be added in the future, such as wiki editions.
- The analysis in the Governance/Leadership section of this report relies on data in the OpenStack Foundation governance file[19] and the wiki sites that contain information about the Board of Directors, the Technical and User Committees and others. Therefore, the analysis of OpenStack governance/leadership is only as accurate and updated as these sources are.

---

[17] https://github.com/grimoirelab/perceval
[18] http://grimoirelab.github.io/
[19] http://git.openstack.org/cgit/openstack/governance/tree/reference/projects.yaml