

Gender Diversity Analysis in the OpenStack Community

Contributors:

Nicole Huesman, Intel Corporation
Daniel Izquierdo Cortázar, Bitergia
Allison Price, OpenStack Foundation

November 2017

Work commissioned by Intel Corporation





(cc) 2017 Bitergia, Intel. Some rights reserved.
This work licensed under Creative Commons Attribution-ShareAlike 4.0
Unported License.
To view a copy of full license, see
<http://creativecommons.org/licenses/by-sa/4.0>,
or write to Creative Commons, 559 Nathan Abbott Way, Stanford,
California 94305, USA.



Executive Summary

This new version report seeks to study and provide data about the gender diversity within the OpenStack community through October 1, 2017 adding two new areas of analysis: (1) attendees representation and (2) non-technical contributions. The other two previous areas of analysis were leadership and technical contributions.

All research and results are based on publicly available data sources except the data coming directly from the OpenStack Foundation records with respect to the attendees representation. The website is the main data source for the leadership/governance study. Git and Gerrit repositories are used for the analysis of the technical contributions. And the mailing lists are the base for the analysis of non-technical contributions.

From a governance/leadership standpoint (contained in Chapter 1), this report analyzes women that hold roles on the OpenStack Board of Directors, the Technical and User Committees, the Working Groups, the OpenStack Ambassadors, and the Project Team Leaders (PTL) positions. Numbers are in line with the previous report although it is remarkable the addition of two women to the Technical Committee that represents 15% of the total members. With respect to the other positions, women represent 21.4% of Ambassadors, 15.8% of the Working Groups, 12.5% of the Board of Directors, 3.2% of the PTLs. On the other hand there are not women participating in the User Committee.

This first part of the research adds in this new version the percentage of women attending the last 2 Summits and Project Teams Gathering (PTG) events (Chapter 2). PTGs attendance is in line with the Summit analysis with numbers over 10%. Women participating as keynote speakers represent a remarkable 30%.

The technical contributions (Chapter 3) are again in line with the previous analysis with numbers around 10% for the population and the activity for commits, code review submissions, and code review votes.

As new addition of this report, the mailing lists were added and counted as non-technical contributions (Chapter 4). Women participation is measured in values under 10% of the activity. While focusing on other mailing lists that are not the main development one (openstack-dev), that percentage increases with numbers over 13% of the total activity.

Table of Contents

Executive Summary	2
Table of Contents	3
1. Leadership & Governance	5
Board of Directors	5
Technical Committee	6
User Committee	7
Working Groups	7
OpenStack Ambassadors	8
Project Technical Leads (PTLs)	9
Other Considerations	11
2. Summits and PTGs Representation	11
Keynote Summit Representation	11
Summit and PTGs attendees and speakers	11
3. Technical Contributions	12
Top Projects	12
Source Code Contributions	14
Commits by Gender	14
Developers by Gender	15
Code Review Contributions	16
Code Review Submissions	17
Developers Submitting Changesets	18
Code Review Votes	19
Core Code Review Votes	21
4. Non-Technical Contributions	24
5. Further Work and Recommendations	29
Appendix A: Detailed Summary of Female Developer Activity	31
Appendix B: Technical Details and Limitations	34
Appendix C: Feedback retrieved in previous releases	36



1. Leadership & Governance

This section focuses on the analysis of gender diversity in the OpenStack community from a leadership and governance perspective. Of total membership, women comprise a sizable percentage at 21.4% of OpenStack Ambassadors, 15.4% of the Technical Committee, 15.8% Working Groups, and 12.5% of the Board of Directors. A quarter of the OpenStack Foundation officers are women, while nearly 30% of women have served as keynote speakers at the previous two OpenStack Summits in Barcelona and Boston. Besides, the mean attendance of women to the last two Summits and PTGs are over 10%. Alternatively, female representation lags among Project Team Lead (PTL) positions at 3.2%, and within the User Committee at 0%.

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.

Board of Directors

The Board of Directors "provides strategic and financial oversight of Foundation resources and staff."¹ At the time of this report, females comprise 12.5% of Board membership, or 3 of 24 members. The list of members is as follows:

Member	Co. Affiliation	Member	Co. Affiliation
Alan Clark	SUSE	Joseph Wang	inwinSTACK
Allison Randal	OSI, HPE, others	Junwei Liu	China Mobile Research Institute
Anni Lai	Huawei	Kavit Munshi	Aptira
Boris Renski	Mirantis	Kenji Kaneshige	Fujitsu
Brad Topol	IBM	Lew Tucker	Cisco Systems
Brian Stein	Rackspace	Mark McLoughlin	Red Hat Inc.
ChangBo Guo	EasyStack	Robert Esker	NetApp

¹ <https://www.openstack.org/foundation/board-of-directors/>

Christopher Price	Ericsson AB	Russell Bryant	Red Hat Inc.
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	Tristan Goode	Aptira

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.” At the time of this report, this committee is comprised of 13 people, 2 of them females (15.4%)².

Current roster is as follows:

Member	Co. Affiliation	Member	Co. Affiliation
Chris Dent	VMware	Jeremy Stanley	OpenStack Foundation
Colleen Murphy	HP, SUSE	John Garbutt	StackHPC-
Davanum Srinivas	Futurewei Tech, Mirantis	Julia Kreger	IBM
Dean Troyer	Intel	Paul Belanger	Red Hat Inc.
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

² <https://www.openstack.org/foundation/tech-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."³ At the time of this report, there are 5 members, all male. The list of members is as follows:

Member	Co. Affiliation
Edgar Magana	Workday
Matt Van Winkle	Rackspace
Melvin Hillsman	Huawei
Saverio Proto	SWITCH
Shamail Tahir	athenaHealth

Table: OpenStack Foundation User Committee

Working Groups

There are a number of working groups that the User Committee assists.⁴ These working groups and their co-chairs are listed in the chart below. At the time of this report, of the total 19 co-chairs, 3, or 15.8%, are women.

Working Group	Co-Chairs
App Dev Enablement Working Group	Patricia Montenegro, Christopher Aedo
Enterprise Working Group	Yih Leong Sun
Fault-Genes Working Group	Nemat Bidokhti, Rochelle (Rocky) Grober
LCOO	Jamey McCabe, Sundar Krishnamurthy
Large Deployment Team	Matt Van Winkle

³ <https://www.openstack.org/foundation/user-committee/>

⁴ https://wiki.openstack.org/wiki/Governance/Foundation/UserCommittee#Working_Groups

Massively Distributed Clouds	Adrien Lebre
Ops Tags Team	Tom Fifield, Edgar Magana, Jon Proulx, Shilla Saebi
Operators Telecom/NFV	Curtis Collicutt
Product Working Group	Yih Leong Sun, Shamail Tahir
Scientific Working Group	Stig Telfer, Blair Blethwaite, Martial Michel

Table: OpenStack Foundation Working Groups

OpenStack Ambassadors

OpenStack ambassadors⁵ 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 14 ambassadors, 3, or 21.4% are women.

Current list of ambassadors:

Ambassador	Ambassador
Akihiro Hasegawa	Kavit Munshi
Akira Yoshiyama	Lisa-Marie Namphy
Christian Berendt	Lu Ye
Erwan Gallen	Marcelo Dieder
Ilya Alekseyev	Márton Kiss
Jaesuk Ahn	Shilla Saebi
John Studarus	Stacy Véronneau

⁵ <https://groups.openstack.org/ambassador-program>

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.⁶ At the time of this report, of the total 63 PTLs, 2, or 3.2% are women with a similar percentage of unknown. The following is the list of PTLs:

PTL Name	Project	PTL Name	Project
Dave McCowan	Barbican	Michal Jastrzebski	Kolla
Masahito Muroi	Blazar	Antoni Segura Puimedon	Kuryr
Samuel Cassiba	Chef OpenStack	Spyros Trigazis	Magnum
Jay Briant	Cinder	Ben Swartzlander	Manila
Christophe Sauthier	Cloudkitty	Sampath Priyankara	Masakari
Eric Kao	Congress	Renat Akhmerov	Mistral
Zhipeng Huang	Cyborg	Witold Bedyk	Monasca
Graham Hayes	Designate	Rong Zhu	Murano
Petr Kovar	Documentation	Miguel Lavalle	Neutron
Omer Anson	Dragonflow	Matt Riedemann	Nova
Andrey Pavlov	Ec2-api	Michael Johnson	Octavia
Saad Zaher	Freezer	James Page	OpenStack Charms
Brian Rosmaita	Glance	Matt McEuen	OpenStack-Helm
Rico Lin	Heat	Jean-Philippe Evrard	OpenStackAnsible

⁶ <http://git.openstack.org/cgit/openstack/governance/tree/reference/projects.yaml>

Ying Zuo	Horizon	Dean Troyer	OpenStackClient
Frank Kloeker	I18n	ChangBo Guo	Oslo
Clark Boylan	Infrastructure	Thomas Goirand	Packaging-deb
Dmitry Tantsur	Ironic	Thomas Bechtold	Packaging-rpm
Chen Ying	Karbor	Mohammed Naser	Puppet OpenStack
Lance Bragstad	Keystone	Andrea Frittoli	Quality Assurance

Table: OpenStack Foundation PTLs (1/2)

PTL Name	Project	PTL Name	Project
Andrey Kurilin	Rally	John Dickinson	Swift
Chris Hoge	RefStack	Gongysh Gongysh	Tacker
Sean McGinnis	Release Management	Gordon Chung	Telemetry
Matthew Thode	Requirements	Zhiyuan Cai	Tricircle
Telles Mota Vidal Nóbrega	Sahara	Alex Schultz	Tripleo
Steve McLellan	Searchlight	Manoj Kumar	Trove
Luke Hinds	Security	Ifat Afek	Vitrage
Ruijie Yuan	Senlin	Alexander Chadin	Watcher
Monty Taylor	Shade	Claudiu Belu	Winstackers
Rong Zhu	Solum	Fei Long Wang	Zaqar
Tony Breeds	Stable Branch Maintenance	Hongbin Lu	Zun
Kota Tsuyuzaki	Storlets		

Table: OpenStack Foundation PTLs (2/2)

Other Considerations

OpenStack Foundation Officers

The OpenStack Foundation’s goal is “to serve developers, users, and the entire ecosystem.” At the time of this report, a quarter of OpenStack Foundation officers are women⁷.

Current List of Officers:

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

2. Summits and PTGs Representation

Keynote Summit Representation

During the previous two OpenStack Summits, in Barcelona and Boston, women comprised 27%, or 12 of a total 45, keynote speakers.

Summit and PTGs attendees and speakers

Barcelona Summit (October 2016)

- 11% of overall attendance were female
- 14% of speakers were female

Boston Summit (May 2017)

- 12% of overall attendance were female

⁷ <https://www.openstack.org/foundation/staff/>

- 12% of speakers were female

Atlanta PTG (February 2017)

- 10% of overall attendance were female

Denver PTG (September 2017)

- 12% of overall attendance were female

3. Technical Contributions

This report examines technical activity as measured by source code contributions and code review contributions. It also identifies which projects rank highest in gender diversity, as measured by population, or number of female contributors, and their respective activity.

The number of total OpenStack contributors peaked at the beginning of 2016 and has seen an overall decrease since then. Where the data presented in this report shows slight decreases in the number of female contributors within the OpenStack community, this trend is line with the overall decrease of total OpenStack contributors. The same is true when examining the level of activity. These trends may indicate the overall maturation of the OpenStack project itself.

Across both source code and code review contributions, female representation has remained relatively steady, as measured by population, or the number of female contributors, and their respective number of commits. While the level of source code activity among females over the last year slightly exceeds this activity in aggregate over the last four years, the level of code review activity among women over the last year falls slightly below its respective four-year aggregate.

Sources used in this section include Gerrit repositories, where code reviews take place, and Git repositories, where the pieces of code that have been accepted in Gerrit are merged. The report provides an analysis of all Git and Gerrit repositories available within the OpenStack Foundation governance file,⁸ which contains pointers to all of the projects, repositories and PTL names used in this report.

⁸ <http://git.openstack.org/cgit/openstack/governance/tree/reference/projects.yaml>

Top Projects

Of the approximately 60 projects within the OpenStack Foundation, the following tables summarize the projects with the greatest gender diversity, as measured by population, or number of female contributors, and their respective activity, or number of commits submitted by females.

Overall, the Infrastructure, Documentation, Nova, Quality assurance, and Neutron projects ranked highest, followed by the Horizon, Cinder, Oslo, Keystone, and Heat projects from a population perspective. At the activity level standpoint, the Infrastructure, Documentation, Ironi, Puppet and Murano projects ranked high.

Project	Authors	Commits	Ratio Authors (Women/Total)	Ratio Commits (Women/Total)
Infrastructure	145	2,638	11.79	5.82
Documentation	119	2,266	21.91	33.38
Nova	97	766	16.19	7.43
Quality Assurance	89	577	14.03	8.32
Neutron	89	706	16.03	6.96
Horizon	83	780	23.44	23.2
Cinder	69	489	14.83	10.26
Oslo	52	244	13.13	4.67
Keystone	52	572	18.05	10.71
Heat	51	206	16.39	3.71

Table: List of top 10 projects with the greatest gender diversity, as measured by population, or number of female contributors. Source: Git repositories.

Project	Authors	Commits	Ratio Authors (Women/Total)	Ratio Commits (Women/Total)
---------	---------	---------	-----------------------------	-----------------------------

Infrastructure	145	2,638	11.79	5.82
Documentation	119	2,266	21.91	33.38
<i>Ironic</i>	43	1,341	14.87	25.13
<i>Puppet OpenStack</i>	32	868	9.63	10.88
<i>Murano</i>	35	844	23.48	29.73
<i>Horizon</i>	83	780	23.44	23.2
<i>Nova</i>	97	766	16.19	7.43
<i>Neutron</i>	89	706	16.03	6.96
<i>OpenStack Charms</i>	8	637	7.33	3.64
<i>Tripleo</i>	35	596	12.68	5.05

Table: List of top 10 projects with the greatest gender diversity, as measured by the highest number of commits contributed by women. Source: Git repositories.

Source Code Contributions

This report provides an analysis of source code contributions along the following dimensions:

- Number of commits by gender over time
- Number of developers by gender authoring those commits over time

Commits by Gender

This section reflects activity by gender, and aggregates this data over the last four years. During this time period, females contributed 6.7% of total commits. Over the last year, this percentage is slightly smaller at 6.42%.

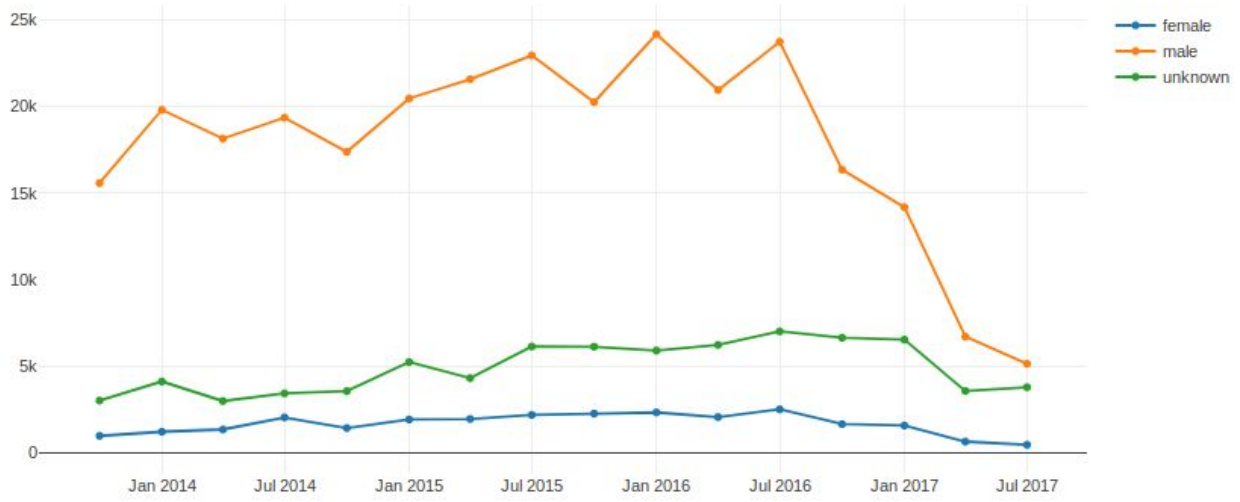


Chart: Number of commits by gender. Source: Git repositories.

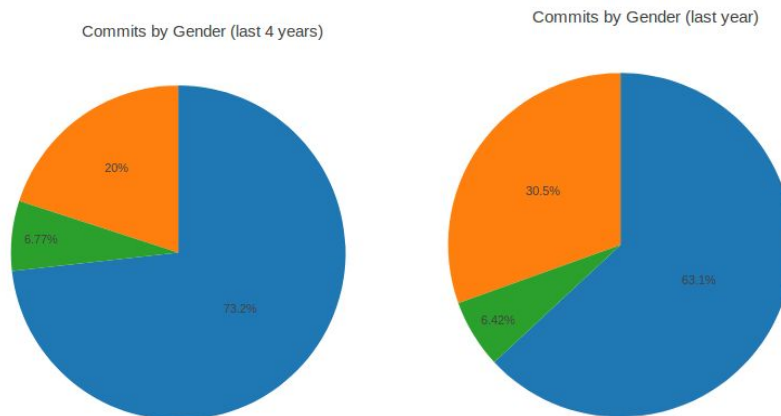


Chart: Number of commits by gender over the last 4 years (left) and over the last year (right). Blue denotes male developers, green denotes female developers, and orange denotes developers unaffiliated with a particular gender. Source: Git repositories.

Developers by Gender

This section provides an account of individuals who have authored commits, by gender, and aggregates this data over the last four years. During this time period, women comprised 10% of

the total population across OpenStack project teams. Over the last year, this percentage is slightly smaller at 9.87%.

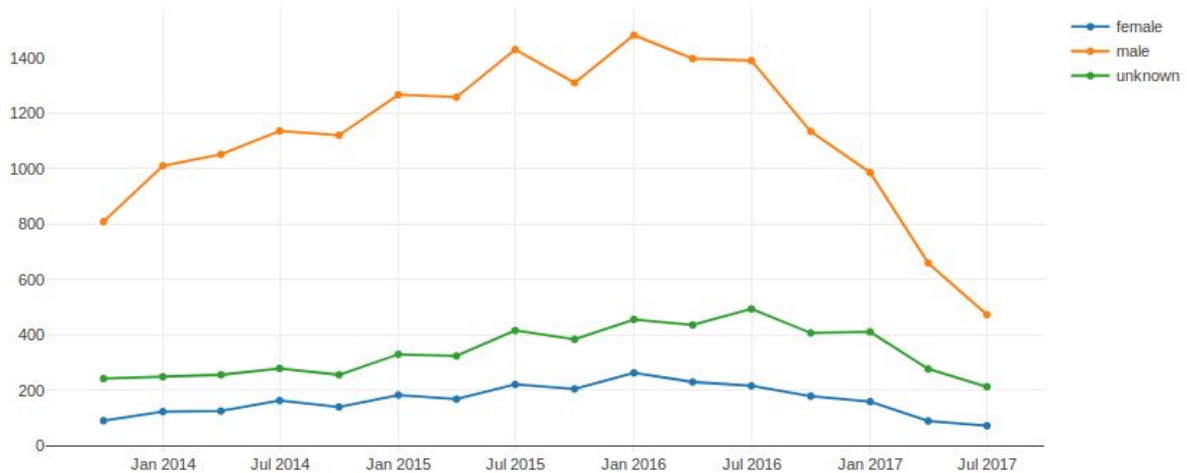


Chart: Number of individuals who have authored commits by gender. Source: Git repositories.

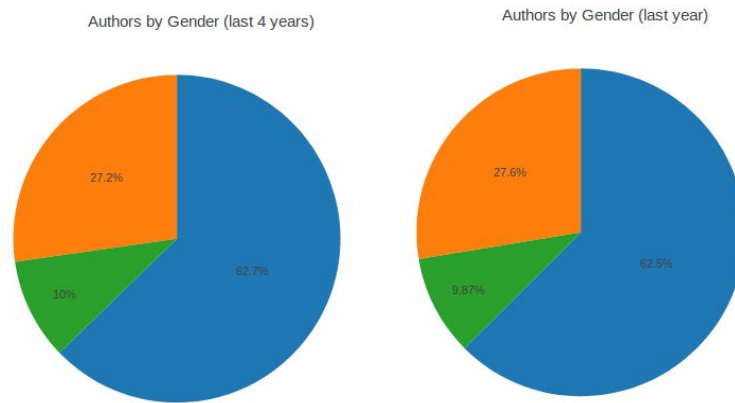


Chart: Number of individuals who have authored commits, by gender, over the last 4 years (left) and over the last year (right). Blue denotes male developers, green denotes female developers, and orange denotes developers unaffiliated with a particular gender. Source: Git repositories.

Code Review Contributions

This report provides an analysis of code review contributions along the following dimensions:

- Number of code review submissions by gender over time
- Number of developers who have submitted changesets, by gender over time
- Number of code review votes undertaken by developers, by gender over time
- Number of developers who have voted in a code review process, by gender over time
- Number of core code review votes undertaken by developers, by gender over time
- Number of core code review votes undertaken by developers, by gender over time
- Number of developers who have participated in a core code review process, by gender over time

Code Review Submissions

This section reflects the number of changeset submissions by gender, and aggregates this data over the last four years. These numbers do not imply that these changesets have been accepted, only that they have been submitted for review. Within the OpenStack community, 83% of changeset submissions are merged into code, while 17% are abandoned. Of these total submissions, those submitted by women in aggregate over the last four years has reached 8.47%. This percentage over the last year falls slightly below the aggregate at 7.24%. This trend is also reflected among male contributors, with a four-year aggregate of 66.5% compared to 69.8% over the last year.

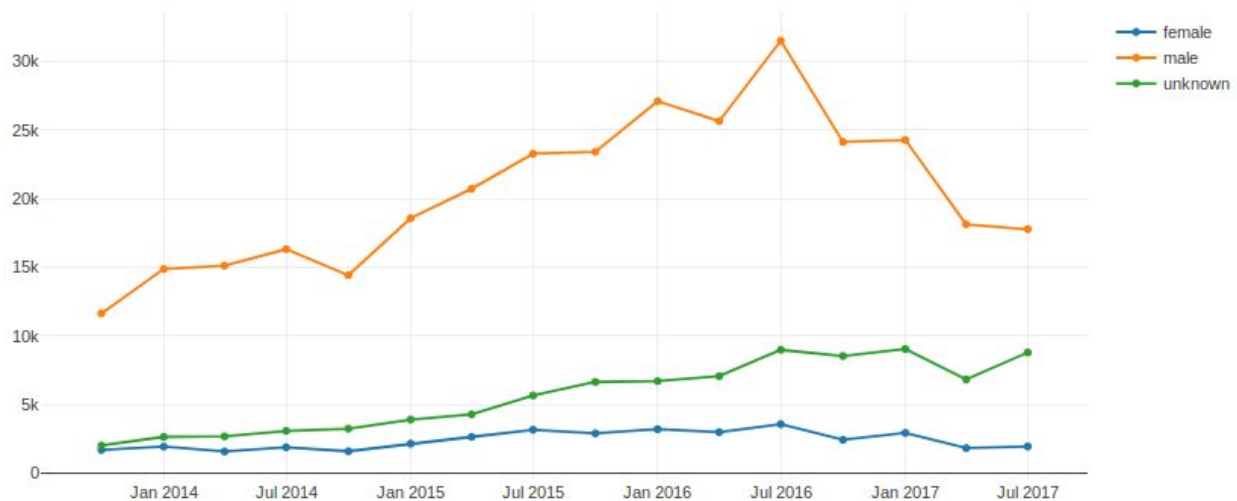
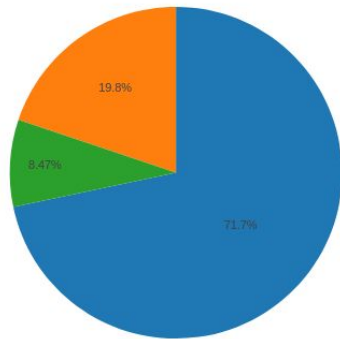


Chart: Number of changeset submissions by gender. Source: Gerrit repositories.

Changeset Submissions by Gender (last 4 years)



Changeset Submissions by Gender (last year)

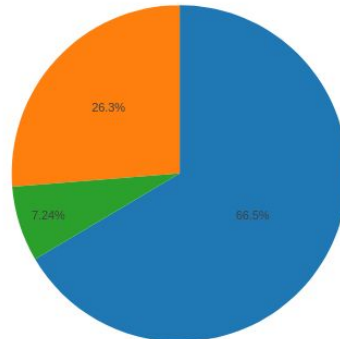


Chart: Number of changeset submissions by gender over the last 4 years (left) and over the last year (right). Blue denotes male developers, green denotes female developers, and orange denotes developers unaffiliated with a particular gender. Source: Gerrit repositories.

Developers Submitting Changesets

This section reveals the number of developers by gender who have submitted changesets in aggregate over the last four years. Of the total population who have submitted code for review over the last four years, women have represented 12.6% in aggregate. This percentage over the last year has fallen slightly under the aggregate at 11.7%.

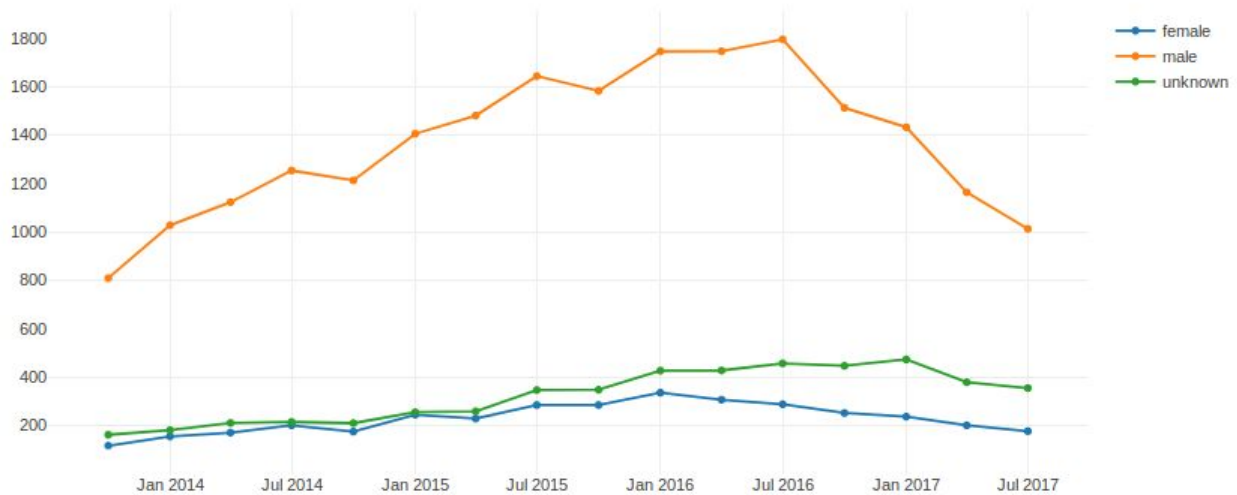


Chart: Number of developers by gender who have submitted changesets. Source: Gerrit repositories.

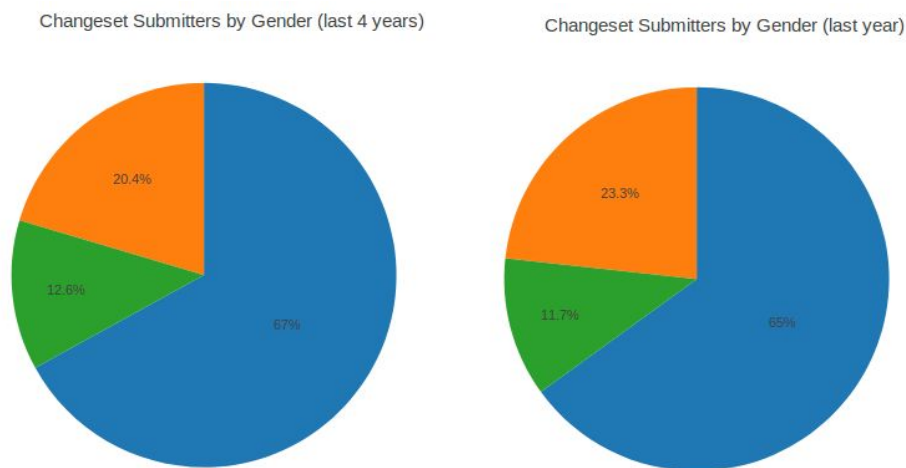


Chart: Number of developers by gender who have submitted changesets over the last 4 years (left) and over the last year (right). Blue denotes male developers, green denotes female developers, and orange denotes developers unaffiliated with a particular gender. Source: Gerrit repositories.

Code Review Votes

This section details the number of code review votes by gender, as well as the number of developers by gender who have voted in a code review process, and aggregates this data over the last four years. A code review is conducted by a developer when a -1, +1, -2 or +2 is provided as a response to a piece of code that has been submitted.

Of the total population who have submitted code review votes, the percentage of women has decreased a bit from 11.6% when comparing the four-year aggregate with that of the last year at 10.9%. When examining the male population using this same lens, the trend remains consistent, with approximately 1% variance, at 67.8% and 66.1%, respectively. In terms of the activity, the trend is pretty similar as being stable for women at 8.5% of the total votes when comparing the four-year aggregate with the last year activity.

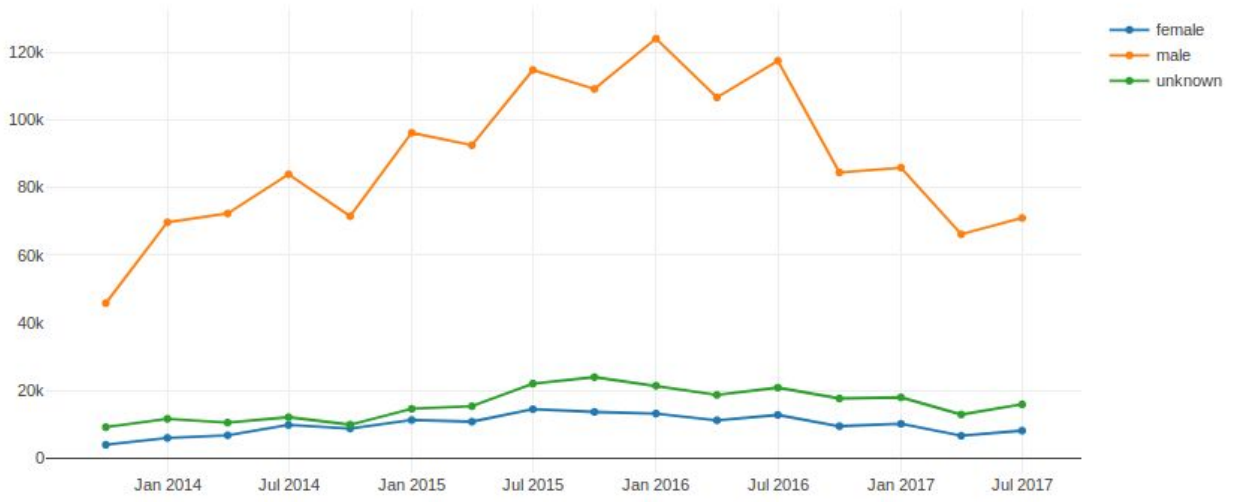


Chart: Number of code review votes by gender (votes as -2, +2, -1 and -1). Source: Gerrit repositories.

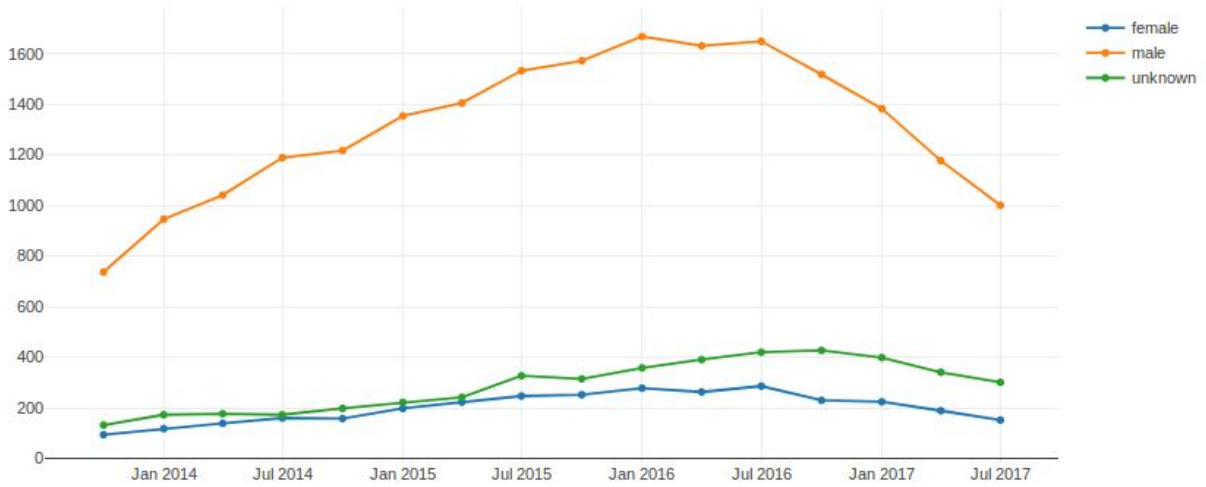


Chart: Number of developers by gender who have voted in a code review process (votes as -2, +2, -1 and -1). Source: Gerrit repositories.

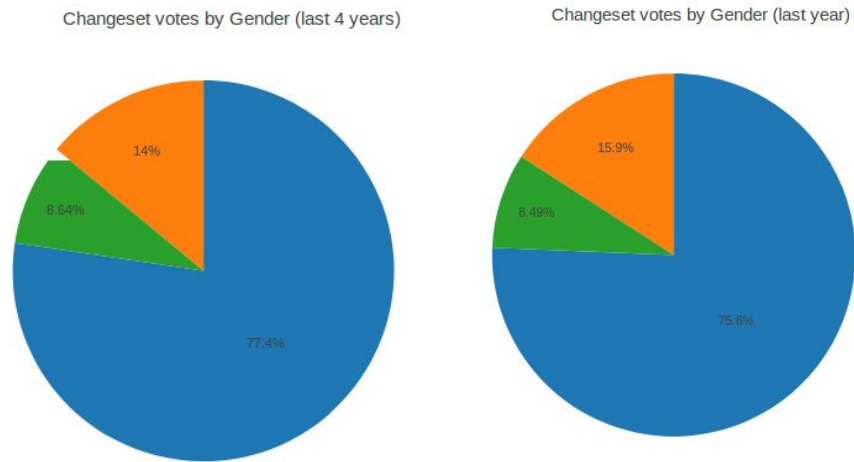


Chart: Number of code review votes by gender for the last 4 years (left) and number of people voting in a code review process as -2, -1, +1, +2 (right) for the last year. Blue denotes male developers, green denotes female developers, and orange denotes developers unaffiliated with a particular gender. Source: Gerrit repositories.

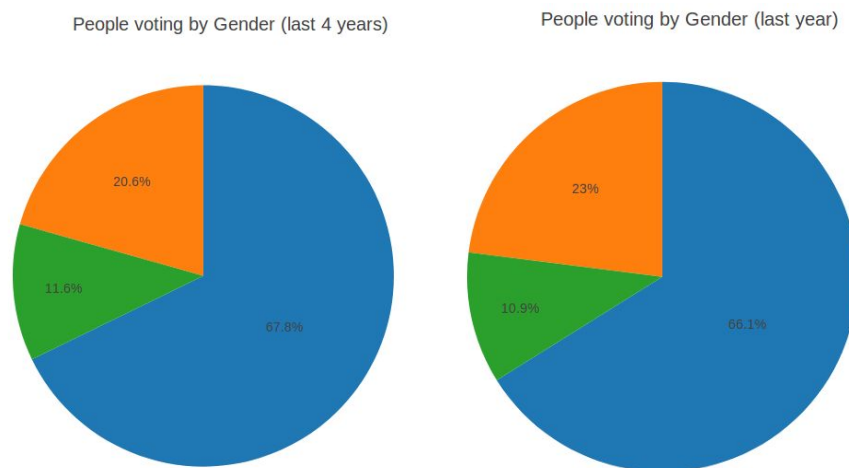


Chart: Number of people voting in a code review process by gender for the last 4 years (left) and number of people voting in a code review process as -2, -1, +1, +2 (right) for the last year. Blue denotes male developers, green denotes female developers, and orange denotes developers unaffiliated with a particular gender. Source: Gerrit repositories.

Core Code Review Votes

This section illustrates the number of core code review votes by gender, as well as the number of developers by gender who have participated in a core code review process, and aggregates this data over the last four years. A core code review is delineated as any vote in the Gerrit system, being +2 or -2. Only core reviewers are allowed to vote in core code reviews, and to allow the piece of code to be merged into the master branch or abandoned.

Of the total population who have submitted core code review votes, again, the percentage of women has remained steady at 11-12% (11.9% and 10.9%, specifically) when comparing the four-year aggregate with that of the last year. This trend mirrors that of the male population when making the same comparison, at 73.2% and 73.7%, respectively.

This trend holds steady when examining the level of activity, or number of core review votes submitted by women. This activity level has remained steady at 7.5% (7.58% and 7.48%, specifically) when comparing the four-year aggregate with that of the last year. Again, this trend is on par with the activity level of the male population using the same lens, at 81.8% and 80.6%, respectively.

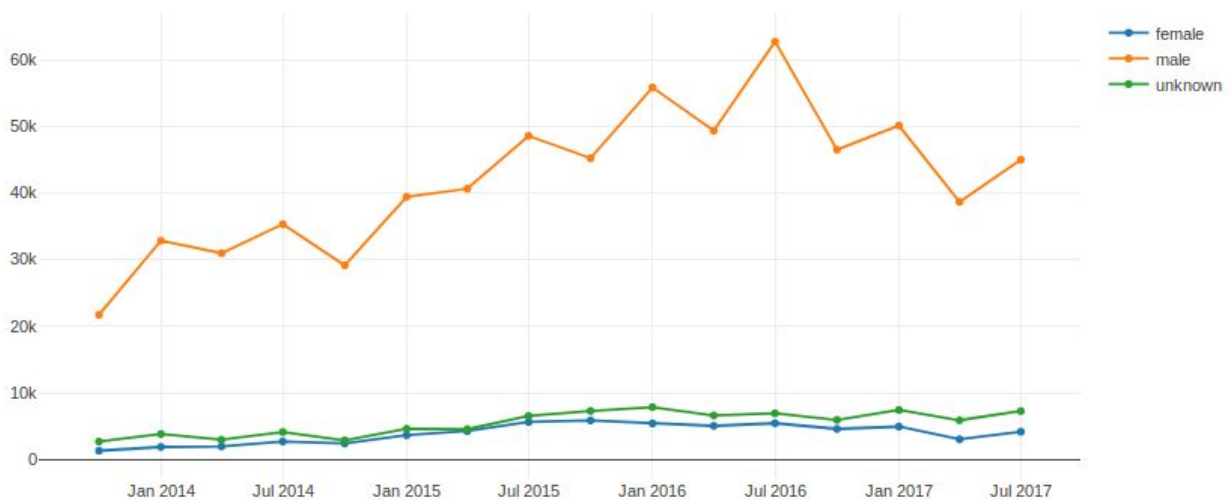
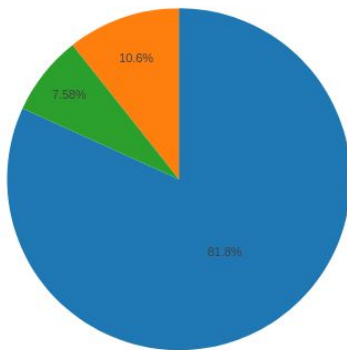


Chart: Number of core code review votes (votes as +2 or -2). Source: Gerrit repositories.

Core Reviews by Gender (last 4 years)



Core Reviews by Gender (last year)

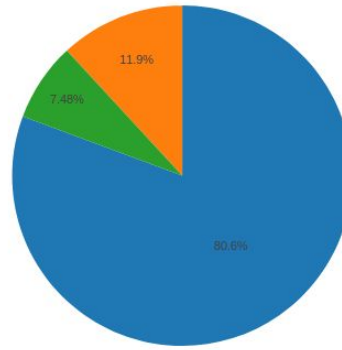


Chart: Core review activity (as -2 or +2) by gender. Last four years of activity (left) and last year of activity (right). Blue denotes male developers, green denotes female developers, and orange denotes developers unaffiliated with a particular gender. Source: Gerrit repositories.

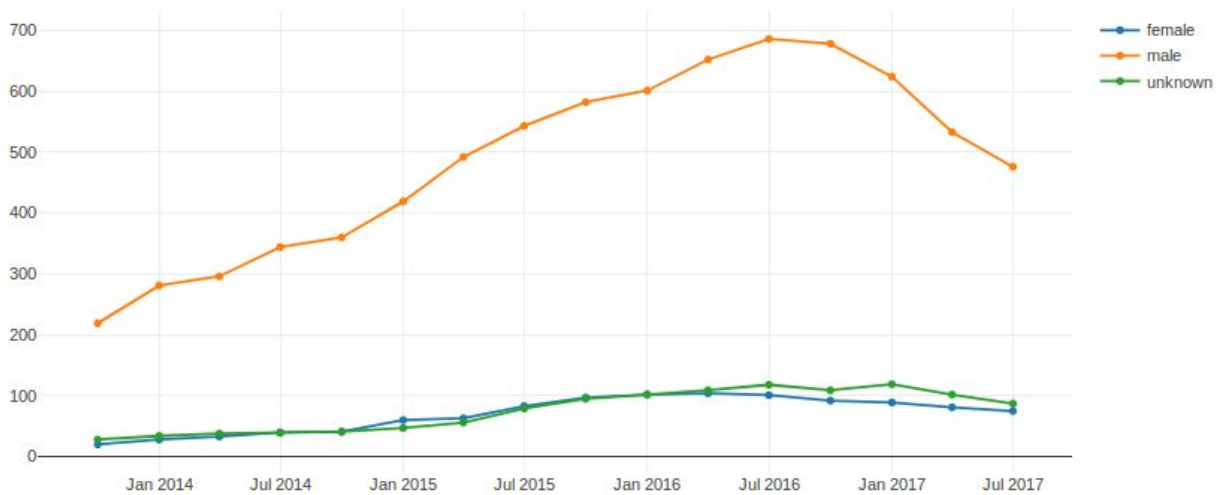


Chart: Number of core reviewers (those allowed to vote +2 or -2). Source: Gerrit repositories.

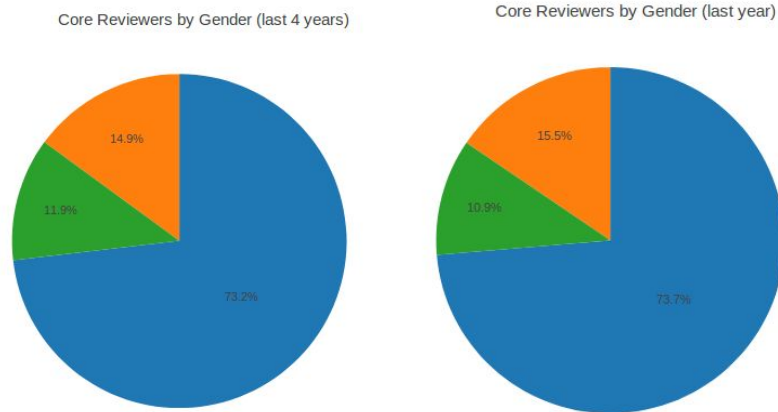


Chart: Core reviewer activity (as -2 or +2) by gender. Last four years of activity (left) and last year of activity (right). Blue denotes male developers, green denotes female developers, and orange denotes developers unaffiliated with a particular gender. Source: Gerrit repositories.

4. Non-Technical Contributions

The non-technical contributions are defined as any other contribution not focused on Git, Gerrit, or any CI system. With this respect the communication channels were a good starting point to depict this type of data sources. The whole set of mailing lists⁹ were analyzed with the goal of understanding the differences between this and purely development workflows.

In general there is a decrease of the activity across the several mailing lists. This activity is similar to other type of activities measured within the community (mainly commits and code review processes). In terms of diversity, women activity reaches a level of 8.67% of the total activity for the last four years of activity. The last year activity is a bit smaller reaching 7.49% of the total activity.

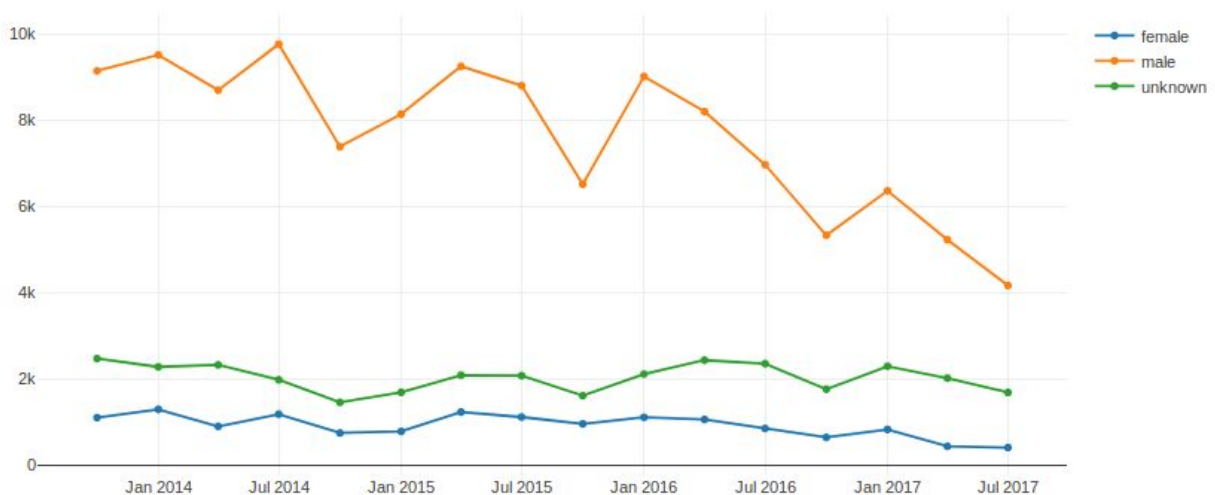


Chart: Number of emails by gender. Source: Mailing list repositories.

⁹ <http://lists.openstack.org/cgi-bin/mailman/listinfo>

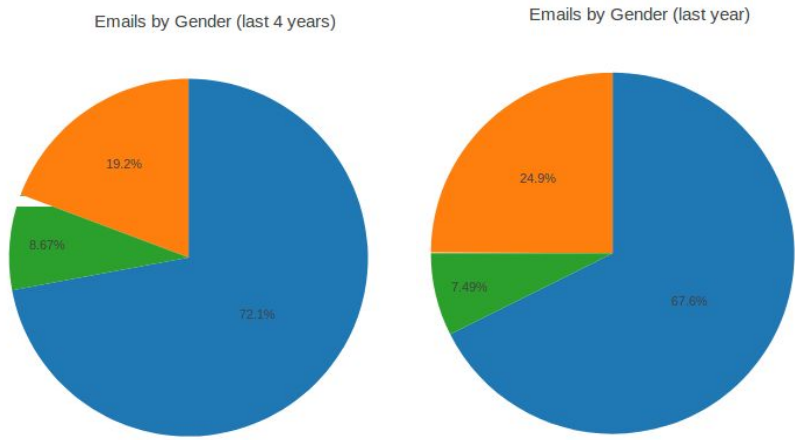


Chart: Emails activity by gender. Last four years of activity (left) and last year of activity (right). Blue denotes male developers, green denotes female developers, and orange denotes developers unaffiliated with a particular gender. Source: Mailing list repositories.

The following charts divide the activity by gender in the mailing lists by three types: the development mailing list (openstack-dev), the language-related mailing lists (all of those in other language as openstack-es or openstack-it and the i18n ones), and the rest of the mailing lists.

When focusing on the development mailing lists the evolution of the activity is similar to the aggregated ones. However the percentage of activity is much smaller down to 6%.

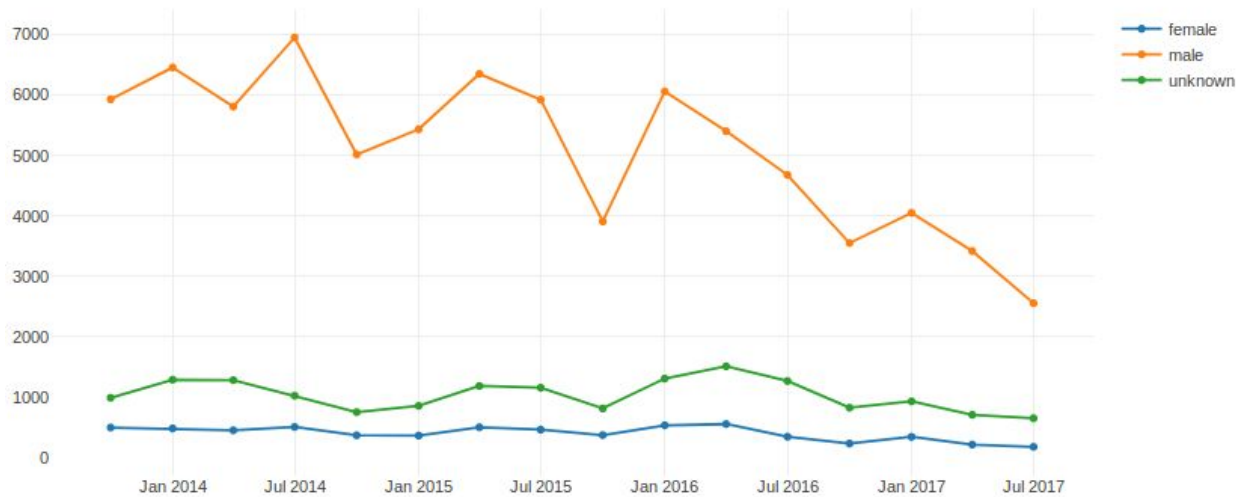
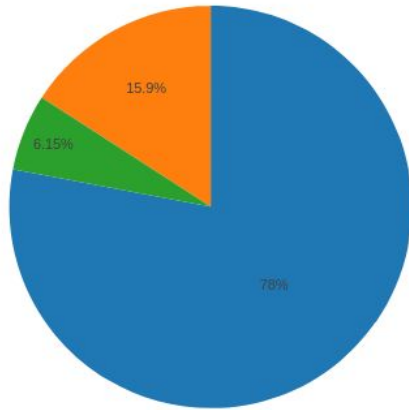


Chart: Number of emails by gender filtered by the development mailing list. Source: Mailing list repositories.

Openstack-Dev Emails by Gender (last 4 years)



OpenStack-Dev Emails by Gender (last year)

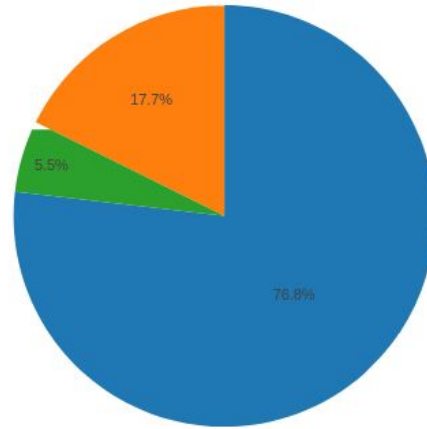


Chart: Emails activity by gender. Last four years of activity (left) and last year of activity (right). Blue denotes male developers, green denotes female developers, and orange denotes developers unaffiliated with a particular gender. Source: Mailing list repositories.

The evolution of the activity in the other two cases is similar, with a decrease, but there is a main difference in terms of participation by gender. The third group of mailing lists show numbers over 13%. These lists contains activity from the community, the technical committee or the operators among others.

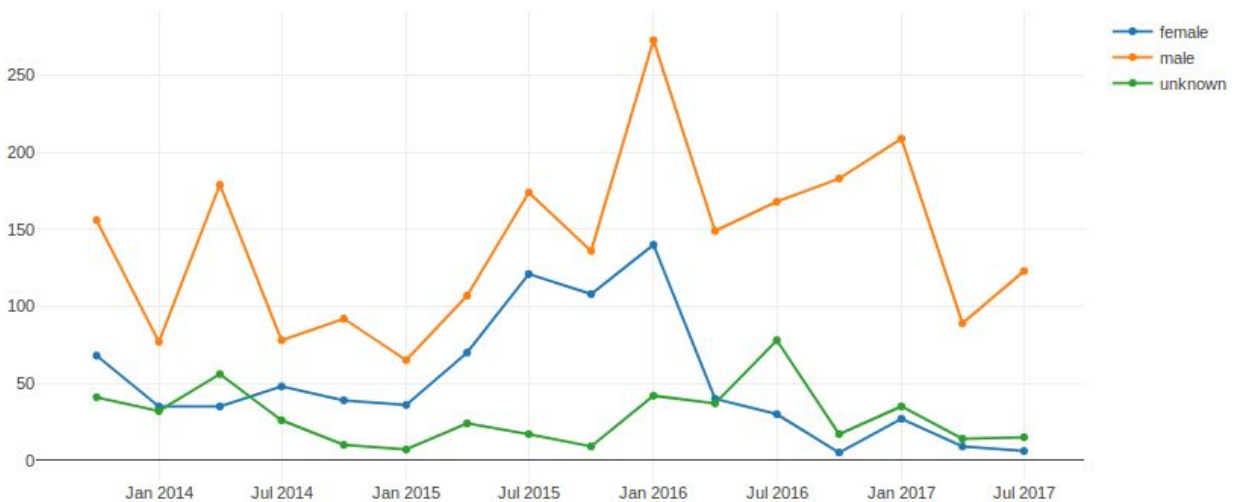
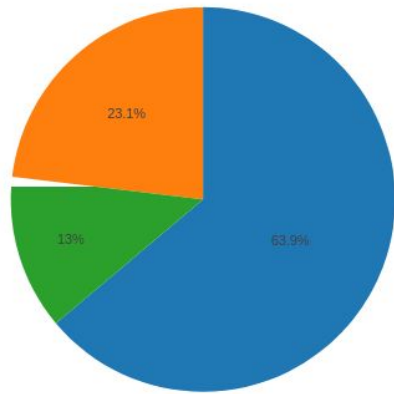


Chart: Number of emails by gender filtered by the language related mailing lists. Source: Mailing list repositories.

Lang-focused Emails by Gender (last 4 years)



Lang-focused Emails by Gender (last year)

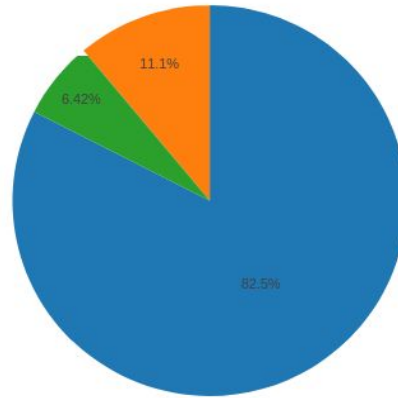


Chart: Emails activity by gender in the group 'language related'. Last four years of activity (left) and last year of activity (right). Blue denotes male developers, green denotes female developers, and orange denotes developers unaffiliated with a particular gender. Source: Mailing list repositories.

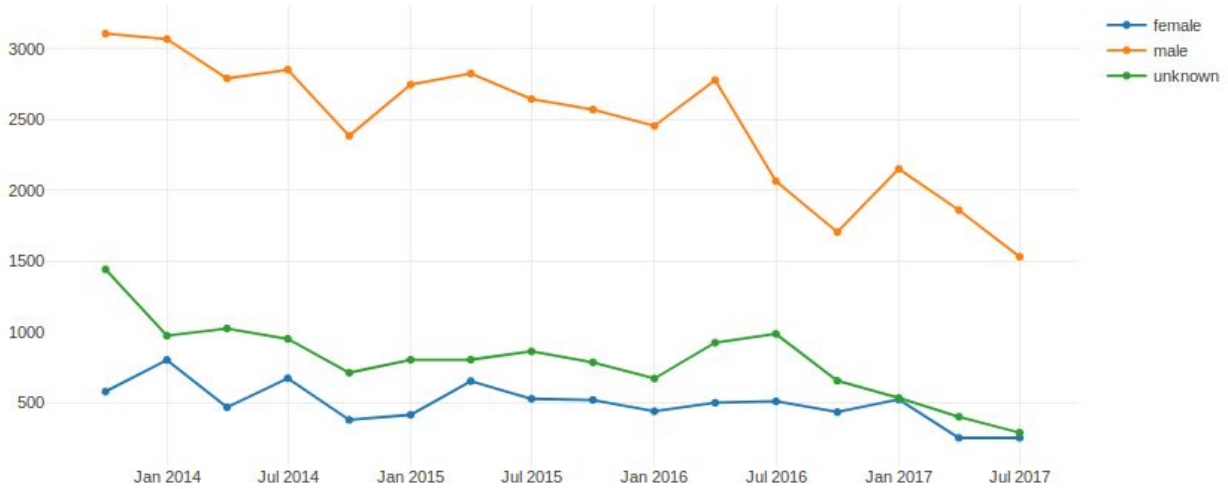
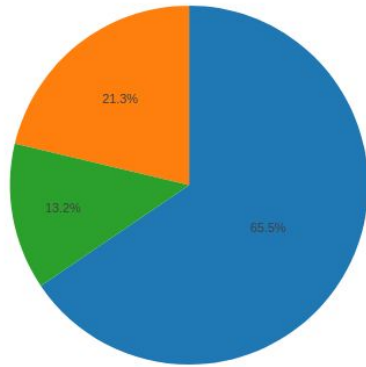


Chart: Number of emails by gender filtered in the group 'other mailing lists'. Source: Mailing list repositories.

Other MLS Emails by Gender (last 4 years)



Other MLS Emails by Gender (last year)

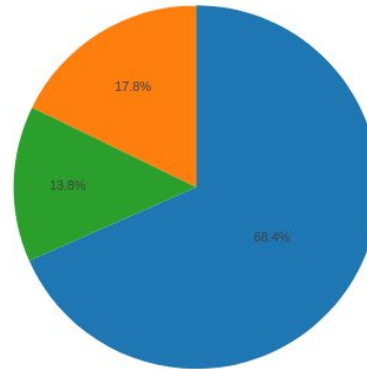


Chart: Emails activity by gender in the group 'other mailing lists'. Last four years of activity (left) and last year of activity (right). Blue denotes male developers, green denotes female developers, and orange denotes developers unaffiliated with a particular gender. Source: Mailing list repositories.

5. Further Work and Recommendations

This research report has examined gender diversity within the OpenStack community from four perspectives: (1) governance/leadership, (2) attendees representation, (3) technical contributions, and (4) non-technical contributions.


While diversity focuses on bringing women and underrepresented minorities into the community, inclusion emphasizes the importance of environments that welcome and support a diverse community. The data in this report suggests some recommendations as further work, divided into focuses and diversity, followed by inclusion.

Recommendations to increase **gender diversity**:

- Continue to track women's participation in governance/leadership and technical activities within the OpenStack community, while extending this tracking to include other forms of contribution such as marketing, community building, and participation in question-and-answer forums. All contributions--both technical and non-technical--must be recognized, and women often contribute more heavily in non-technical areas as seen in the analysis of the mailing lists and not focused on the main development one.
- Collaborate with the OpenStack project teams with the highest diversity (as measured by the highest number of female developers and highest number of activity from female developers, and other parameters) to collect, document and publicize best practices. The authors of the report have already been approached by some of them. We still need to work on further collaboration and provide updated data.
- Study the impact of specific, diversity-related policies and initiatives undertaken by the OpenStack Foundation to determine their impact on the pipeline and entrance of women and underrepresented minorities into the community.
- Work with the community to understand their needs more deeply, and build follow-on action plans based on this data to address these needs.
- Assist PTLs who express a desire for greater diversity within their project team with recruitment activities to achieve this stated objective.

Recommendations to foster **more inclusive teams and communities**:

- Track both the tenure and attrition of women in the OpenStack community, and study the impact of specific policies and initiatives undertaken by the OpenStack Foundation, such



as the Outreachy mentor program, to determine their impact on these factors. It is important to create an inviting and supportive environment for these individuals to land, once they have decided to join the community. Outreachy is indeed one of the main objectives for the next version of this report.

- Collaborate with the OpenStack project teams with the most inclusive environments (as measured by the highest tenures and retention rates among women and underrepresented minorities, and other parameters) to collect, document and publicize best practices.
- Invest in increasing the number of women and underrepresented minorities who participate in technical leadership, such as the Technical and User Committees, as well as PTL positions. Consider a mentorship program that aims to provide mentorship and shadowing opportunities to women with PTL potential. Consider extending invitations to these females and underrepresented minorities to attend and observe technical meetings, so that they gain first-hand experience and knowledge about how technical leadership teams work.
- Continue to support the Women of OpenStack program and associated onboarding activities, such as the Upstream University, Long-Term Mentoring and Speed Mentoring programs, and event speakerships, with diversity as a foundation. These programs have all been well accepted and are succeeding in increasing diversity and inclusion in the OpenStack community.
- Develop ways to ensure that the community is well aware of how important diversity and inclusion are to the leadership of the OpenStack Foundation and to the success of the OpenStack Project.
- Continue to enforce the OpenStack Foundation's code of conduct to reinforce the importance of diversity and inclusion within OpenStack project teams.
- Lastly, it has been shown that inclusive communities have good documentation, onboarding processes and mentors. Ensure all projects have these elements as a baseline.

Appendix A: Detailed Summary of Female Developer Activity

This appendix tracks the activity by female developers within each of the OpenStack project teams.

Each of the columns within the table below are computed as described below:

- **Ratio_Authors:** The number in this column denotes the ratio of 100 female developers for every 100 male developers for a given project.
- **Ratio_Commits:** The number in this column denotes the ratio of 100 commits submitted by female developers for every 100 commits submitted by male developers for a given project.
- **Authors:** The number in this column denotes the total number of individuals identified as females who have contributed to a given project, as tracked through the Git repositories.
- **Commits:** The number in this column denotes the total number of changes to the source code for a given project. A commit is usually submitted through a code review process.

Project	Authors	Commits	ratio_authors	ratio_commits
Infrastructure	145	2638	11.8	5.83
Documentation	119	2266	21.92	33.38
ironic	43	1341	14.88	25.13
Puppet OpenStack	32	868	9.64	10.89
murano	35	844	23.49	29.74
horizon	83	780	23.45	23.21
nova	97	766	16.19	7.43
neutron	89	706	16.04	6.96
OpenStack Charms	8	637	7.34	3.64
tripleo	35	596	12.68	5.05
Quality Assurance	89	577	14.04	8.33
keystone	52	572	18.06	10.71
cinder	69	489	14.84	10.26
OpenStackClient	37	386	20.22	21.04

OpenStackAnsible	26	350	14.94	2.97
magnum	27	334	18.75	22.77
congress	10	251	12.99	25
octavia	40	250	12.74	8.72
oslo	52	244	13.13	4.68

Table: List of female developer activity within OpenStack project teams (1/3)

Project	Authors	Commits	ratio_authors	ratio_commits
dragonflow	13	237	33.33	24.69
Telemetry	51	224	22.67	7.6
glance	49	217	20	10.93
manila	25	215	16.56	11.38
heat	51	206	16.4	3.71
swift	35	202	18.23	8.65
sahara	25	201	16.13	6.34
senlin	18	201	35.29	64.01
trove	30	201	21.13	12.61
vitrage	12	198	46.15	30.05
rally	32	191	15.76	9.04
monasca	21	179	15.56	4.97
zaqar	19	173	20.43	20.69
tacker	28	152	23.73	20.51
kolla	31	141	15.27	3.87
zun	13	132	39.39	51.16
mistral	13	127	13.13	8.35
Security	27	119	24.11	7.79
barbican	26	110	20.16	9.01
kuryr	10	109	20.41	29.07

Table: List of female developer activity within OpenStack project teams (2/3)

Project	Authors	Commits	ratio_authors	ratio_commits
requirements	40	103	12.66	5.01
shade	13	103	17.11	8.52
RefStack	7	66	18.92	16.79
designate	12	58	9.76	5.13
Release Management	19	58	11.8	1.82
storlets	3	53	16.67	18.21
Chef OpenStack	17	49	19.54	2.55
solum	6	45	11.54	5.43
watcher	12	44	20.69	9.17
tricycle	8	44	66.67	115.79
searchlight	7	35	16.67	9
winstackers	5	33	13.16	7.16
Packaging-rpm	2	28	6.45	2.99
l18n	1	18	7.69	16.82
freezer	7	14	12.5	1.57
karbor	8	12	36.36	5.97
blazar	2	9	9.09	15.79
cloudkitty	4	6	11.11	1.87
ec2-api	2	5	11.76	1.74

Appendix B: 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 Perceval¹⁰, a 100% open source software tool under the GrimoireLab¹¹ umbrella.
- The code used to enrich and visualize the raw information in this report is provided by Perceval, which is available in the GitHub repository of one of the authors of this study¹². Charts and tables can be visualized in the same repository¹³.
- 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, technical contributions are defined as commits, code reviews and code review vote actions. This is a starting point for this ongoing research, based on input from the community. Other sources for measurement of female activity, engagement and contributions can be added in the future, such as mailing lists, wiki editions, and more.
- The analysis in the Governance/Leadership section of this report relies on data in the OpenStack Foundation governance file¹⁴ and the wiki sites that contain information about the Board of Directors, the Technical and User Committees and others. Therefore, the


¹⁰ <https://github.com/grimoirelab/perceval>

¹¹ <http://grimoirelab.github.io/>

¹² <https://github.com/dicortazar/ipython-notebooks/tree/master/projects/openstack-diversity>

¹³ <https://htmlpreview.github.io/?https://github.com/dicortazar/ipython-notebooks/blob/master/projects/openstack-diversity/OpenStack%20Diversity%20Metrics.html>

¹⁴ <http://git.openstack.org/cgit/openstack/governance/tree/reference/projects.yaml>



analysis of OpenStack governance/leadership is only as accurate and updated as these sources are.

Appendix C: Feedback retrieved in previous releases

The feedback retrieved during previous releases has been focused mainly on the selected data sources, proper understanding of the report data, and details on the projects analyzed.

In first place there is a clear bias in the dataset as gender diversity is not a binary concept. This is a clear limitation of this study. However the aim of this report is not to produce exact and precise data, but to bring attention to the topic of gender diversity in open source communities and specifically in the case of the OpenStack Foundation. We believe that transparency is one of the key legs of open source communities and bringing attention to this topic is a way to take advantage of such key value in order to improve the community, or at least be aware that there is a real issue.

We have explored ways to improve the dataset as it is and from a quantitative point of view, but given the privacy of the information we are consulting, it is hard to find ways of improving this. First from a community perspective, and letting others to update their own gender, and secondly from a more broader perspective as the data could be potentially consumable.

The initial selected data sources were good as starting point. Feedback suggested to add others such as the PTGs attendees and speakers or data sources not focused in development activity as Git or Gerrit does. For this, this new release of the report focuses on the analysis of the attendees during the mid-term releases processes known as PTGs. And the addition of the OpenStack mailing lists as a first step to work on the analysis of the non-technical contributions.

Going to the details, the packaging-deb project seemed to be a potential problem in terms of forking and double counting commits in some cases. For this reason net numbers in the previous report could have been overestimated, while the relative ones should be pretty similar or exactly the same as for instance the same percentage of women are participating in Nova even if this project is being doubled counted.