

Summary Analysis of the 2017 GitHub Open Source Survey

R. Stuart Geiger
Berkeley Institute for Data Science
University of California, Berkeley
stuart@stuartgeiger.com

7 June 2017

1 Abstract

This report is a high-level summary analysis of the 2017 GitHub Open Source Survey dataset,¹ presenting frequency counts, proportions, and frequency or proportion bar plots for every question asked in the survey.

2 Overview

2.1 The 2017 Open Source Survey

This report analyzes the open dataset from the 2017 Open Source Survey, which was conducted by staff at GitHub, with help, support, and feedback from many others [Zlotnick et al., 2017a]. The survey was run in 2017, asking over 50 questions on a variety of topics. The survey’s designers explain the motivation, design, and distribution of the survey on the project’s website:

"In collaboration with researchers from academia, industry, and the community, GitHub designed a survey to gather high quality and novel data on open source software development practices and communities. We collected responses from 5,500 randomly sampled respondents sourced from over 3,800 open source repositories on GitHub.com, and over 500 responses from a non-random sample of communities that work on other platforms. The results are an open data set about the attitudes, experiences, and backgrounds of those who use, build, and maintain open source software." [Zlotnick et al., 2017b]

2.2 Purpose and goal of this report

The GitHub survey team presented analyses of some questions when releasing the survey [Zlotnick et al., 2017b], but there were many more questions asked that are relevant to researchers and community members. This report is an exploratory analysis of all questions asked in the survey, providing a basic summary of the responses to each question. This report presents and plots summary statistics – mostly frequency counts, proportions, then a frequency or proportion bar graph

¹<https://opensource-survey.org/2017/>

– of all questions asked in the survey. Most questions are presented individually, with panel questions grouped together as appropriate. There are no correlations, regressions, or descriptive breakouts between subgroups. Likert-style questions (e.g. Strongly agree <-> strongly disagree) have not been recoded to numerical, scalar values. There are no discussions or interpretations of results. This is left for future work.

The purpose of this report is to facilitate future research on this dataset by giving an overview of the kinds of questions asked in the survey, as well as provide a single, stable reference for citing broad claims in the data. This report is the PDF version of a Jupyter Notebook, which can be run to reproduce the results of the tables and graphs in this report. The Jupyter notebook and a copy of the data is public on GitHub ² and the Open Science Framework ³. Others are encouraged to extend it as they see fit, as this report and the notebooks are licensed CC-BY-4.0.⁴ The "Out[number]" notes before each table and chart are linked to the Jupyter notebook, so you can easily navigate to the notebook cell where the applicable code can be found. If you find this report useful, please cite it as [Geiger, 2017] as detailed in the bibliography at the end of this report.

2.3 Software used

This analysis was conducted in Python [van Rossum, 1995] version 3.6, using Pandas dataframes [McKinney, 2010] for data parsing and transformation, SciPy [Jones et al., 2001] and NumPy [van der Walt et al., 2011] for quantitative computations, and Matplotlib [Hunter, 2007] and Seaborn [Waskom et al., 2014] for visualization. It was conducted in Jupyter Notebooks [Kluyver et al., 2016] using the IPython kernel [Pérez and Granger, 2007], and nbconvert (also discussed in [Kluyver et al., 2016]) was used to convert the notebook into LaTeX for publication in this report.

3 Table of Contents

Contents

1 Abstract	1
2 Overview	1
2.1 The 2017 Open Source Survey	1
2.2 Purpose and goal of this report	1
2.3 Software used	2
3 Table of Contents	2
4 Analysis	5
4.1 Contributor identity	5
4.1.1 People participate in open source in different ways. Which of the following activities do you engage in?	5
4.1.2 How often do you engage in each of the following activities?	6
4.1.3 Employment status	7
4.1.4 In your main job, how often do you write or otherwise directly contribute to producing software?	8
4.1.5 How interested are you in contributing to open source projects in the future?	9
4.1.6 How likely are you to contribute to open source projects in the future?	10
4.2 Priorities and values	11

²<https://github.com/staeiou/github-survey-analysis>

³<https://osf.io/enrq5/>

⁴<https://creativecommons.org/licenses/by/4.0/>

4.2.1	When thinking about whether to use open source software, how important are the following things?	11
4.2.2	When thinking about whether to contribute to an open source project, how important are the following things?	13
4.2.3	How often do you try to find open source options over other kinds of software?	15
4.2.4	Open source software usability	16
4.2.5	Open source software security	17
4.2.6	Open source software stability	18
4.2.7	Identification with open source	19
4.3	Transparency vs privacy	20
4.3.1	Attribution	20
4.3.2	In general, how much information about you is publicly available online?	21
4.3.3	Do you feel that you need to make information available about yourself online for professional reasons?	22
4.3.4	General privacy practices	23
4.3.5	OSS privacy practices	24
4.4	Mentorship / Help	25
4.4.1	Have you ever received any kind of help from other people related to using or contributing to an open source project?	25
4.4.2	Thinking of the most recent case where someone helped you, how did you find someone to help you?	26
4.4.3	Which best describes your prior relationship with the person who helped you?	27
4.4.4	What kind of problem did they help you with?	28
4.4.5	Have you ever provided help for another person on an open source project?	29
4.4.6	Thinking of the most recent case where you helped someone, how did you come to help this person?	30
4.4.7	Which best describes your prior relationship with the person you helped?	31
4.4.8	What kind of problem did you help them with?	32
4.5	Open Source Software in Paid Work	33
4.5.1	Do you contribute to open source as part of your professional work?	33
4.5.2	How often do you use open source software in your professional work?	34
4.5.3	How does your employer's intellectual property agreement/policy affect your free-time contributions to open source unrelated to your work?	35
4.5.4	Which is closest to your employer's policy on using open source software applications?	36
4.5.5	How important do you think your involvement in open source was to getting your current job?	37
4.6	Demographics	38
4.6.1	Do you currently live in a country other than the one in which you were born?	38
4.6.2	Thinking of where you were born, are you a member of an ethnicity or nationality that is a considered a minority in that country?	39
4.6.3	Thinking of where you currently live, are you a member of an ethnicity or nationality that is a considered a minority in that country?	40
4.6.4	What is your gender?	41
4.6.5	Do you identify as transgender?	42
4.6.6	Do you identify as gay, lesbian, or bisexual, asexual, or any other minority sexual orientation?	43
4.6.7	How well can you read and write in English?	44
4.6.8	What is your age?	45
4.6.9	What is highest level of formal education that you have completed?	46
4.6.10	What is the highest level of formal education that either of your parents completed?	47

4.6.11	How old were you when you first had regular access to a computer with an internet connection?	48
4.6.12	Where did you first have regular access to a computer with internet connection?	49
4.6.13	Where was the respondent surveyed from?	50
4.7	Harassment / Inclusiveness of OSS	51
4.7.1	Have you ever observed any of the following in the context of an open source project?	51
4.7.2	Have you ever witnessed any of the following behaviors directed at another person in the context of an open source project? (not including something directed at you)	52
4.7.3	Have you ever experienced any of the following behaviors directed at you in the context of an open source project?	53
4.7.4	Thinking of the last time you experienced harassment, how did you respond?	54
4.7.5	How effective were the following responses? Response counts	55
4.7.6	How effective were the following responses? Proportions	56
4.7.7	As a result of experiencing or witnessing harassment, which, if any, of the following have you done?	57
5	Bibliography	58

4 Analysis

4.1 Contributor identity

4.1.1 People participate in open source in different ways. Which of the following activities do you engage in?

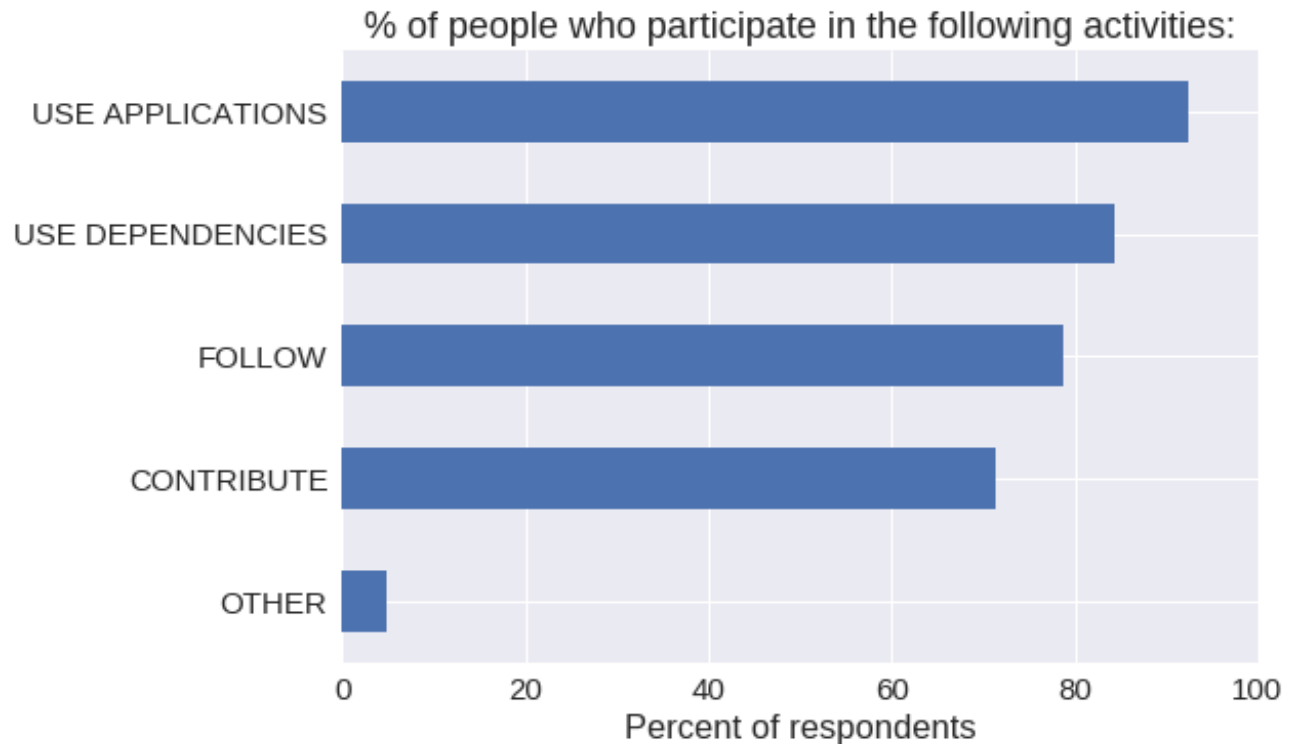
PARTICIPATION.TYPE.*

Out [29] :

	No	Yes
PARTICIPATION.TYPE.FOLLOW	1287	4742
PARTICIPATION.TYPE.USE.APPLICATIONS	454	5575
PARTICIPATION.TYPE.USE.DEPENDENCIES	946	5083
PARTICIPATION.TYPE.CONTRIBUTE	1722	4307
PARTICIPATION.TYPE.OTHER	5742	287

Out [30] :

	percent
PARTICIPATION.TYPE.OTHER	4.76%
PARTICIPATION.TYPE.CONTRIBUTE	71.44%
PARTICIPATION.TYPE.FOLLOW	78.65%
PARTICIPATION.TYPE.USE.DEPENDENCIES	84.31%
PARTICIPATION.TYPE.USE.APPLICATIONS	92.47%

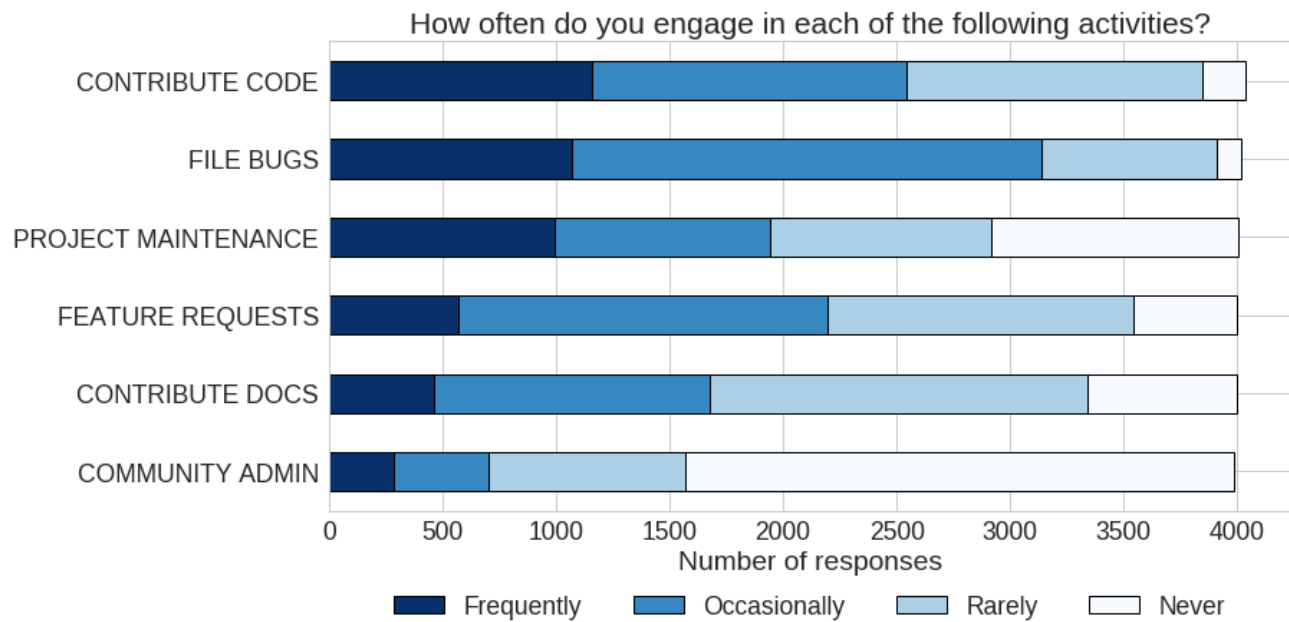


4.1.2 How often do you engage in each of the following activities?

CONTRIBUTOR.TYPE.*

Out [32]:

	Frequently	Occasionally	Rarely	Never
CONTRIBUTOR.TYPE.COMMUNITY.ADMIN	287	417	867	2412
CONTRIBUTOR.TYPE.CONTRIBUTE.DOCS	460	1214	1665	661
CONTRIBUTOR.TYPE.FEATURE.REQUESTS	573	1625	1346	451
CONTRIBUTOR.TYPE.PROJECT.MAINTENANCE	996	944	974	1090
CONTRIBUTOR.TYPE.FILE.BUGS	1067	2073	768	106
CONTRIBUTOR.TYPE.CONTRIBUTE.CODE	1160	1383	1301	189



4.1.3 Employment status

EMPLOYMENT.STATUS

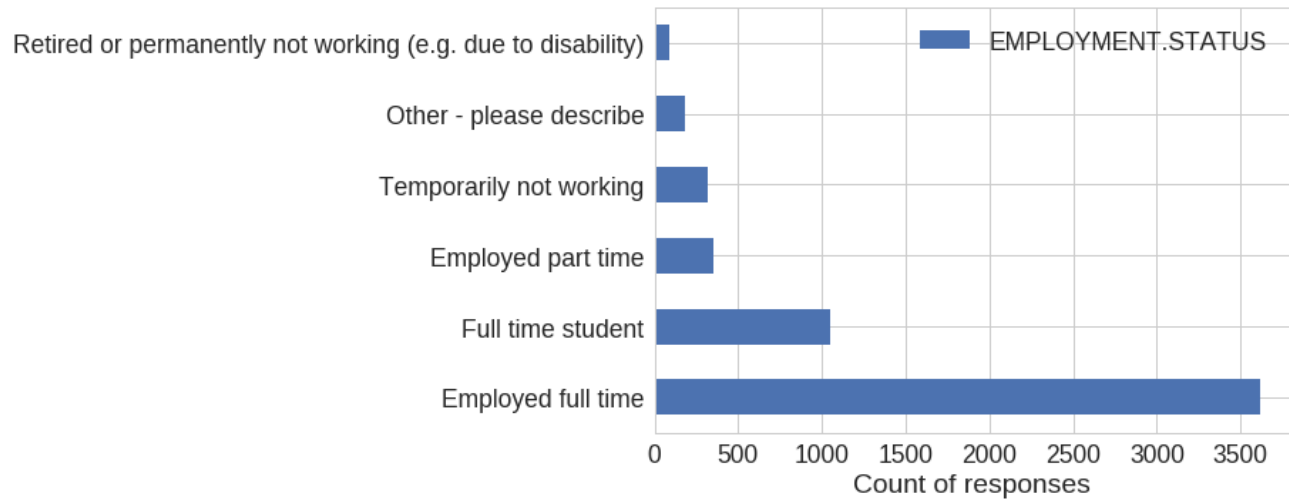
Out [34] :

	count
Employed full time	3615
Full time student	1048
Employed part time	349
Temporarily not working	314
Other - please describe	184
Retired or permanently not working (e.g. due to...	90

Out [35] :

	percent
Employed full time	64.55%
Full time student	18.71%
Employed part time	6.23%
Temporarily not working	5.61%
Other - please describe	3.29%
Retired or permanently not working (e.g. due to...	1.61%

Employment status



4.1.4 In your main job, how often do you write or otherwise directly contribute to producing software?

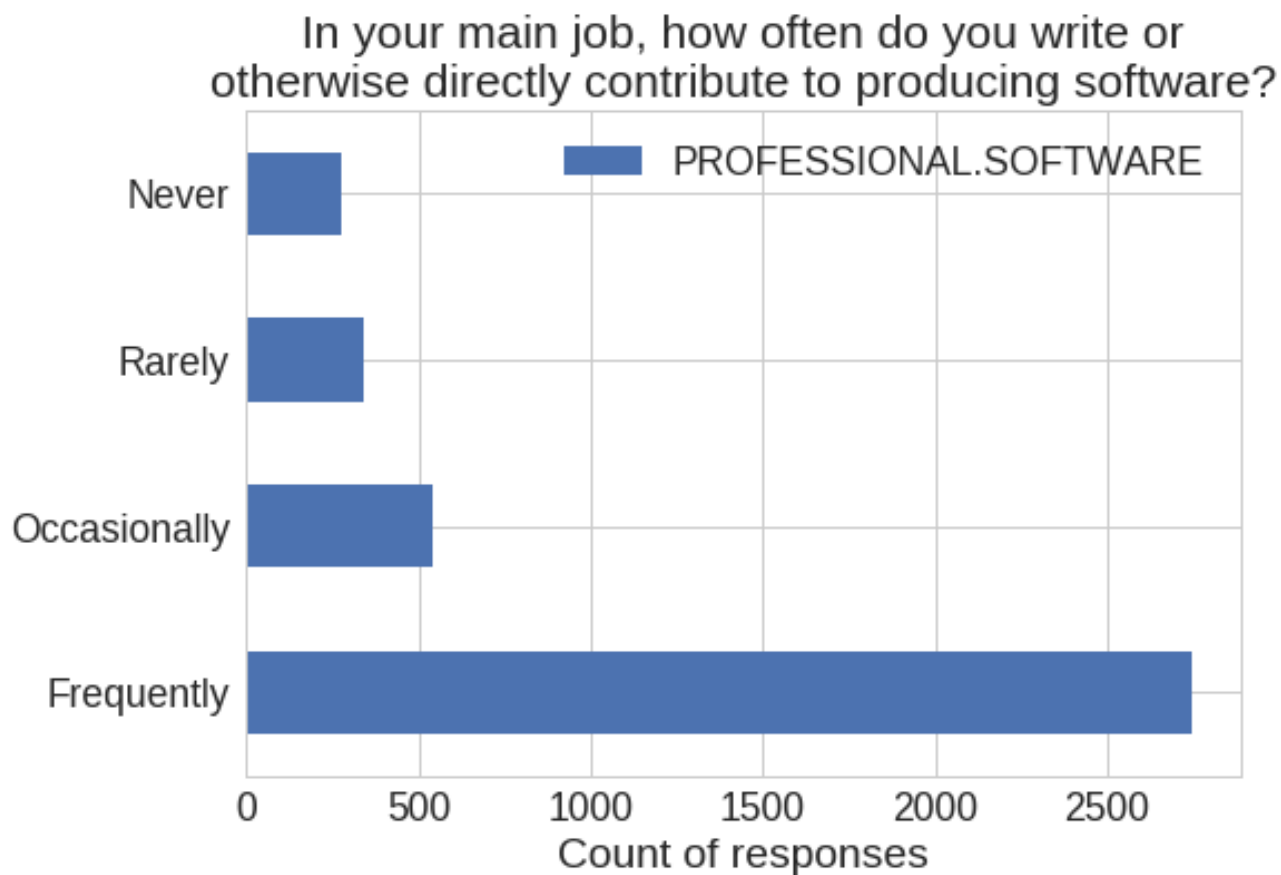
PROFESSIONAL.SOFTWARE

Out [37] :

	count
Frequently	2747
Occasionally	542
Rarely	339
Never	279

Out [38] :

	percent
Frequently	70.31%
Occasionally	13.87%
Rarely	8.68%
Never	7.14%



4.1.5 How interested are you in contributing to open source projects in the future?

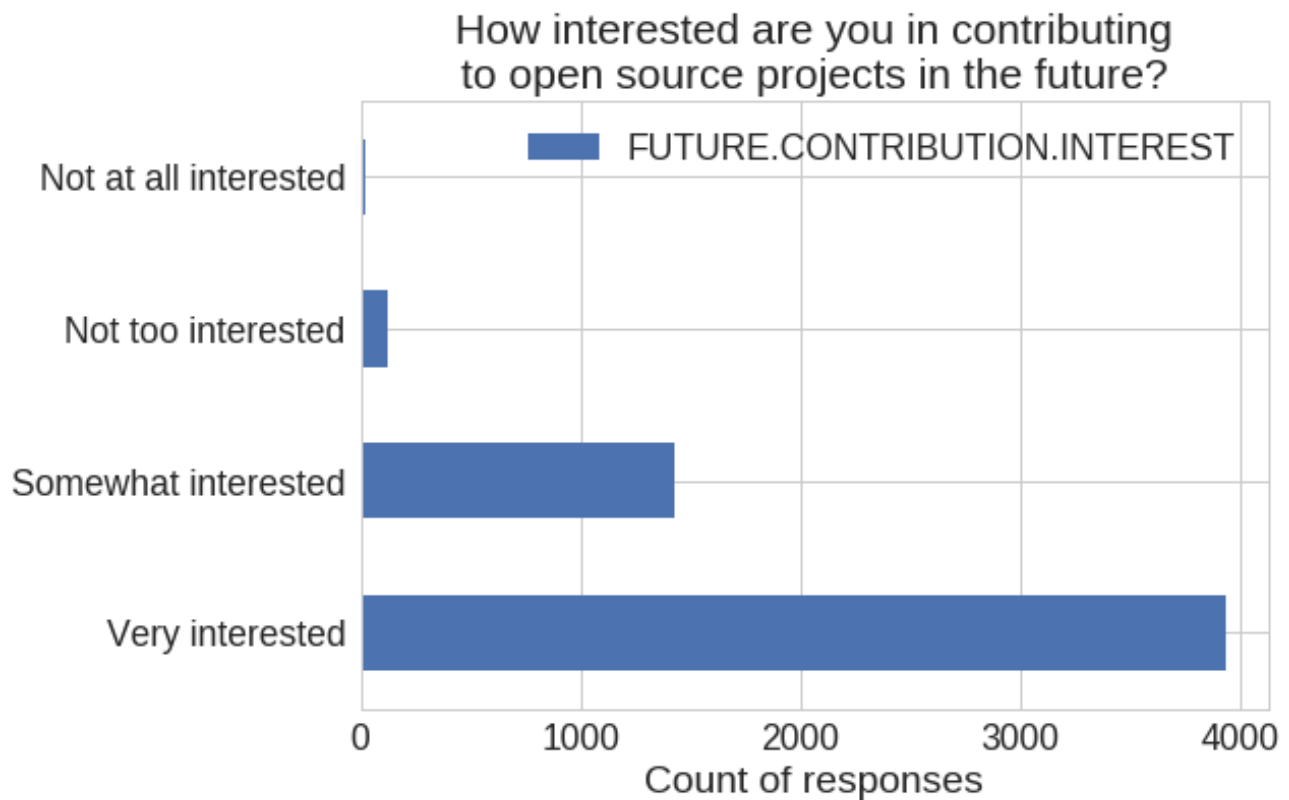
FUTURE.CONTRIBUTION.INTEREST

Out [40] :

	count
Very interested	3929
Somewhat interested	1430
Not too interested	125
Not at all interested	24

Out [41] :

	percent
Very interested	71.33%
Somewhat interested	25.96%
Not too interested	2.27%
Not at all interested	0.44%



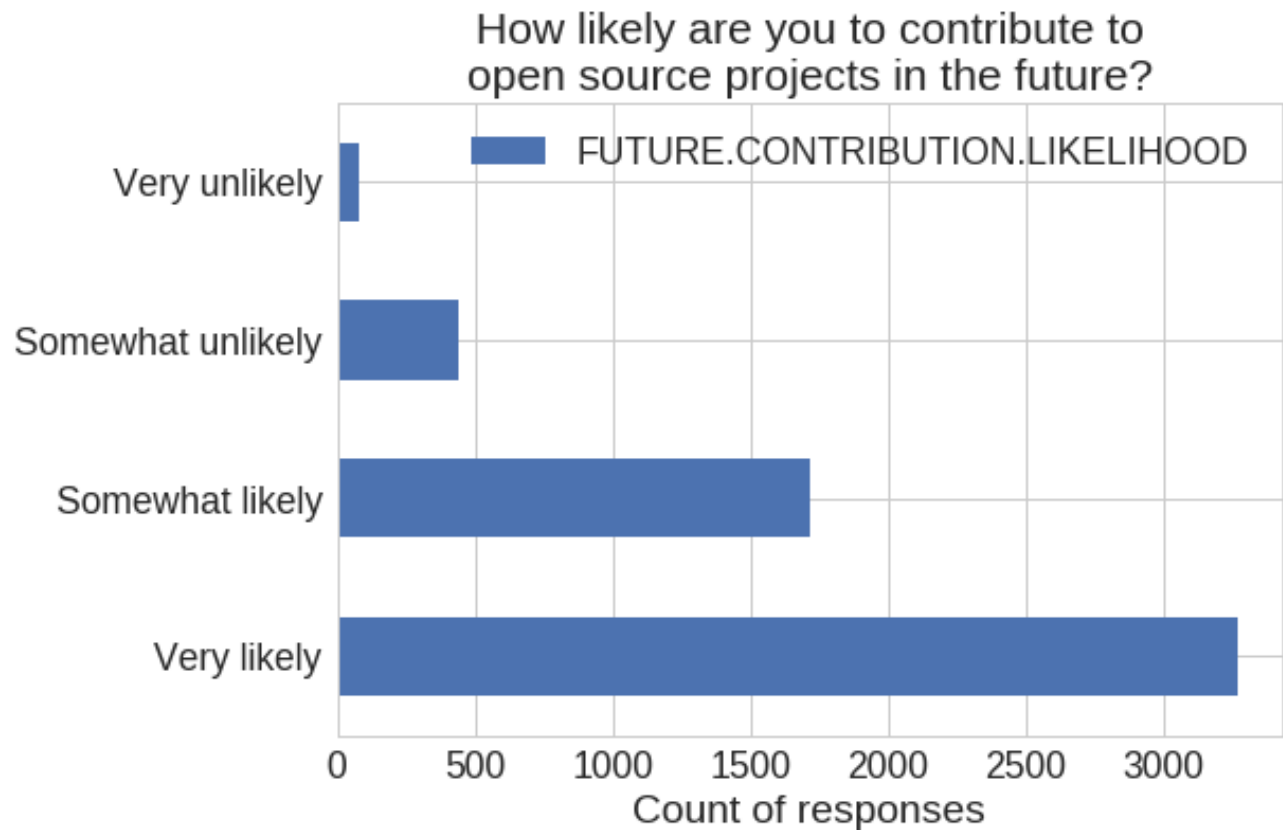
4.1.6 How likely are you to contribute to open source projects in the future?

Out [43]:

	count
Very likely	3271
Somewhat likely	1719
Somewhat unlikely	440
Very unlikely	81

Out [44]:

	percent
Very likely	59.35%
Somewhat likely	31.19%
Somewhat unlikely	7.98%
Very unlikely	1.47%



4.2 Priorities and values

4.2.1 When thinking about whether to use open source software, how important are the following things?

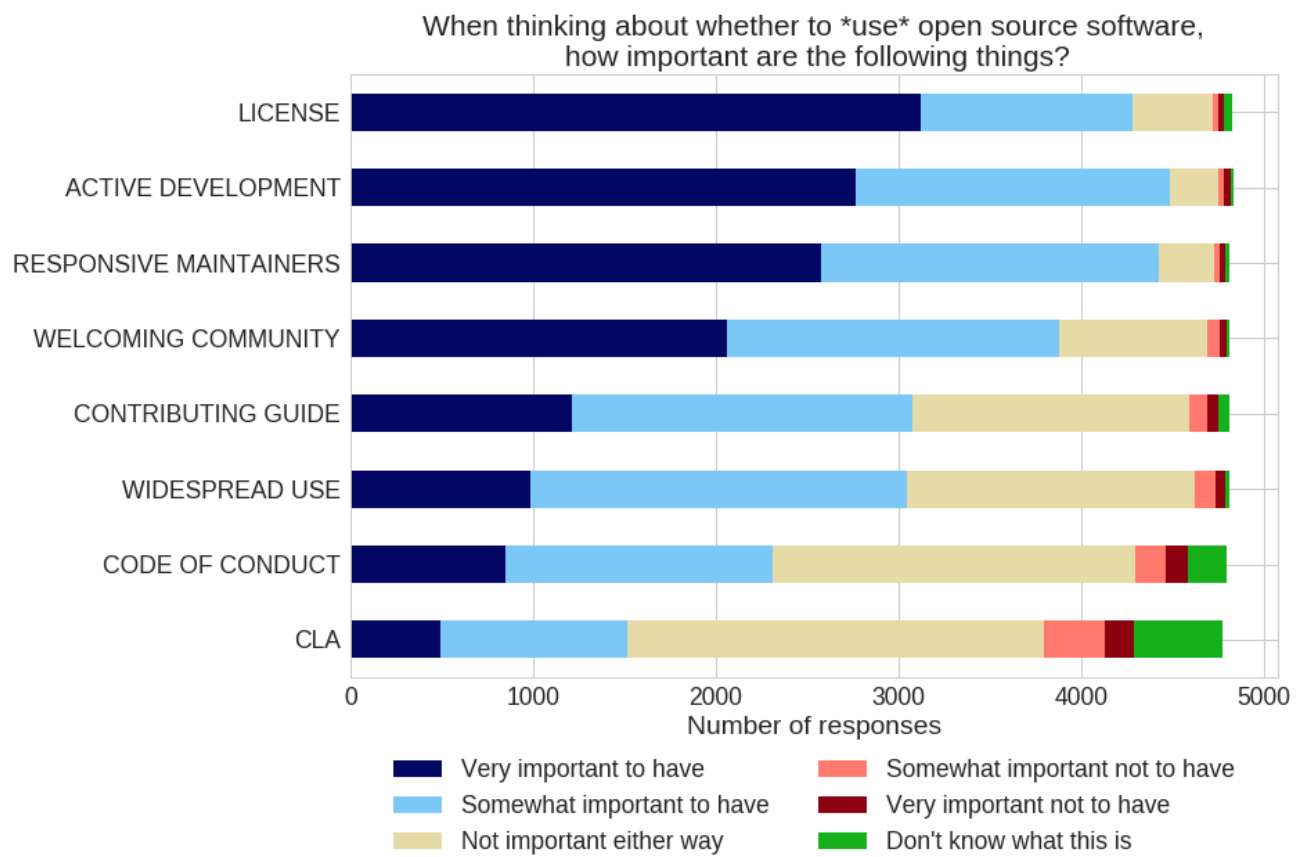
OSS.USER.PRIORITIES.*

Out [47] :

	Very imp to have	Somewhat imp to have	Neither	Somewhat imp to not have	Very imp to not have	Don't know
CLA	490	1024	2282	336	157	488
CODE.OF.CONDUCT	848	1461	1993	166	120	209
WIDESPREAD.USE	984	2067	1576	114	47	28
CONTRIBUTING.GUIDE	1212	1866	1516	95	62	62
WELCOMING.COMMUNITY	2062	1822	812	67	33	18
RESPONSIVE.MAINTAINERS	2575	1850	302	31	35	20
ACTIVE.DEVELOPMENT	2768	1722	267	30	31	16
LICENSE	3125	1160	435	31	33	47

Out [49] :

	Very imp to have	Somewhat imp to have	Neither	Somewhat imp to not have	Very imp to not have	Don't know
CLA	10.26%	21.44%	47.77%	7.03%	3.29%	10.22%
CODE.OF.CONDUCT	17.68%	30.46%	41.55%	3.46%	2.50%	4.36%
WIDESPREAD.USE	20.43%	42.92%	32.72%	2.37%	0.98%	0.58%
CONTRIBUTING.GUIDE	25.18%	38.77%	31.50%	1.97%	1.29%	1.29%
WELCOMING.COMMUNITY	42.83%	37.85%	16.87%	1.39%	0.69%	0.37%
RESPONSIVE.MAINTAINERS	53.50%	38.44%	6.27%	0.64%	0.73%	0.42%
ACTIVE.DEVELOPMENT	57.26%	35.62%	5.52%	0.62%	0.64%	0.33%
LICENSE	64.69%	24.01%	9.00%	0.64%	0.68%	0.97%



4.2.2 When thinking about whether to contribute to an open source project, how important are the following things?

OSS.CONTRIBUTOR.PRIORITIES.*

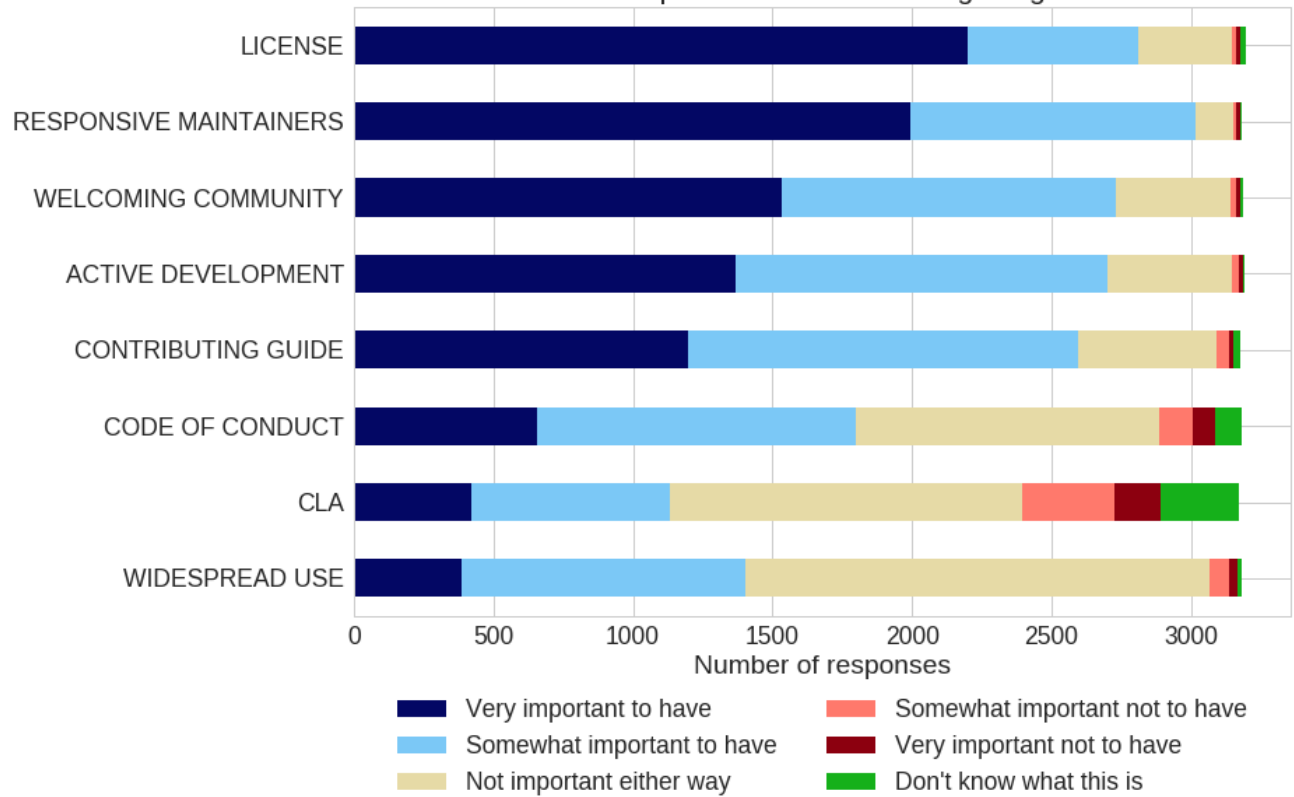
Out [52] :

	Very imp to have	Somewhat imp to have	Neither	Somewhat imp to not have	Very imp to not have	Don't know
WIDESPREAD.USE	387	1016	1666	70	30	12
CLA	419	712	1266	327	166	280
CODE.OF.CONDUCT	655	1145	1085	119	84	96
CONTRIBUTING.GUIDE	1198	1396	500	41	18	24
ACTIVE.DEVELOPMENT	1368	1333	448	21	18	5
WELCOMING.COMMUNITY	1533	1199	411	21	15	7
RESPONSIVE.MAINTAINERS	1994	1022	138	7	16	7
LICENSE	2199	610	337	16	15	18

Out [54] :

	Very imp to have	Somewhat imp to have	Neither	Somewhat imp to not have	Very imp to not have	Don't know
WIDESPREAD.USE	12.17%	31.94%	52.37%	2.20%	0.94%	0.38%
CLA	13.22%	22.46%	39.94%	10.32%	5.24%	8.83%
CODE.OF.CONDUCT	20.57%	35.96%	34.08%	3.74%	2.64%	3.02%
CONTRIBUTING.GUIDE	37.71%	43.94%	15.74%	1.29%	0.57%	0.76%
ACTIVE.DEVELOPMENT	42.84%	41.75%	14.03%	0.66%	0.56%	0.16%
WELCOMING.COMMUNITY	48.12%	37.63%	12.90%	0.66%	0.47%	0.22%
RESPONSIVE.MAINTAINERS	62.63%	32.10%	4.33%	0.22%	0.50%	0.22%
LICENSE	68.83%	19.09%	10.55%	0.50%	0.47%	0.56%

When thinking about whether to *contribute* to an open source project, how important are the following things?



4.2.3 How often do you try to find open source options over other kinds of software?

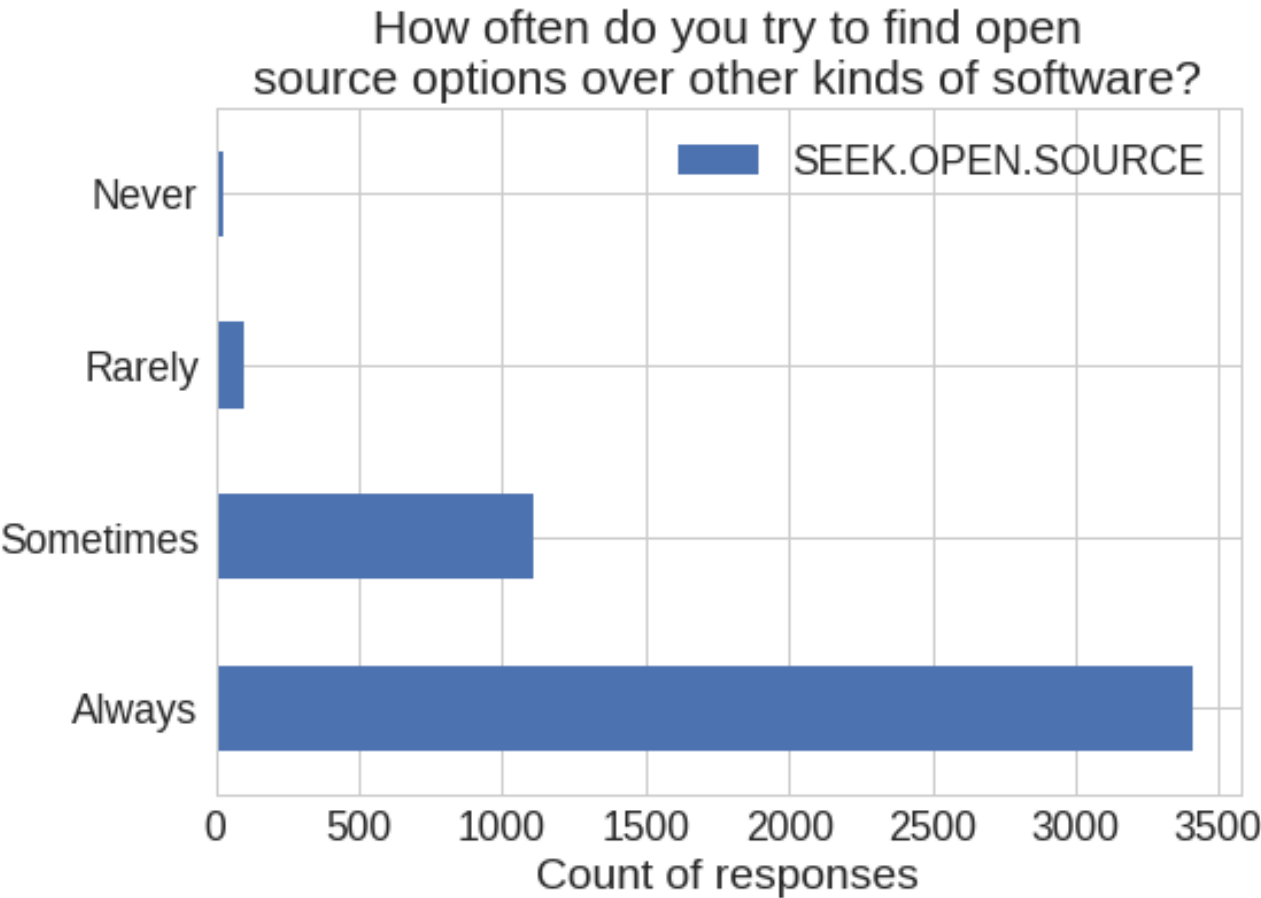
SEEK.OPEN.SOURCE

Out [56] :

	count
Always	3407
Sometimes	1111
Rarely	100
Never	25

Out [57] :

	percent
Always	73.38%
Sometimes	23.93%
Rarely	2.15%
Never	0.54%



4.2.4 Open source software usability

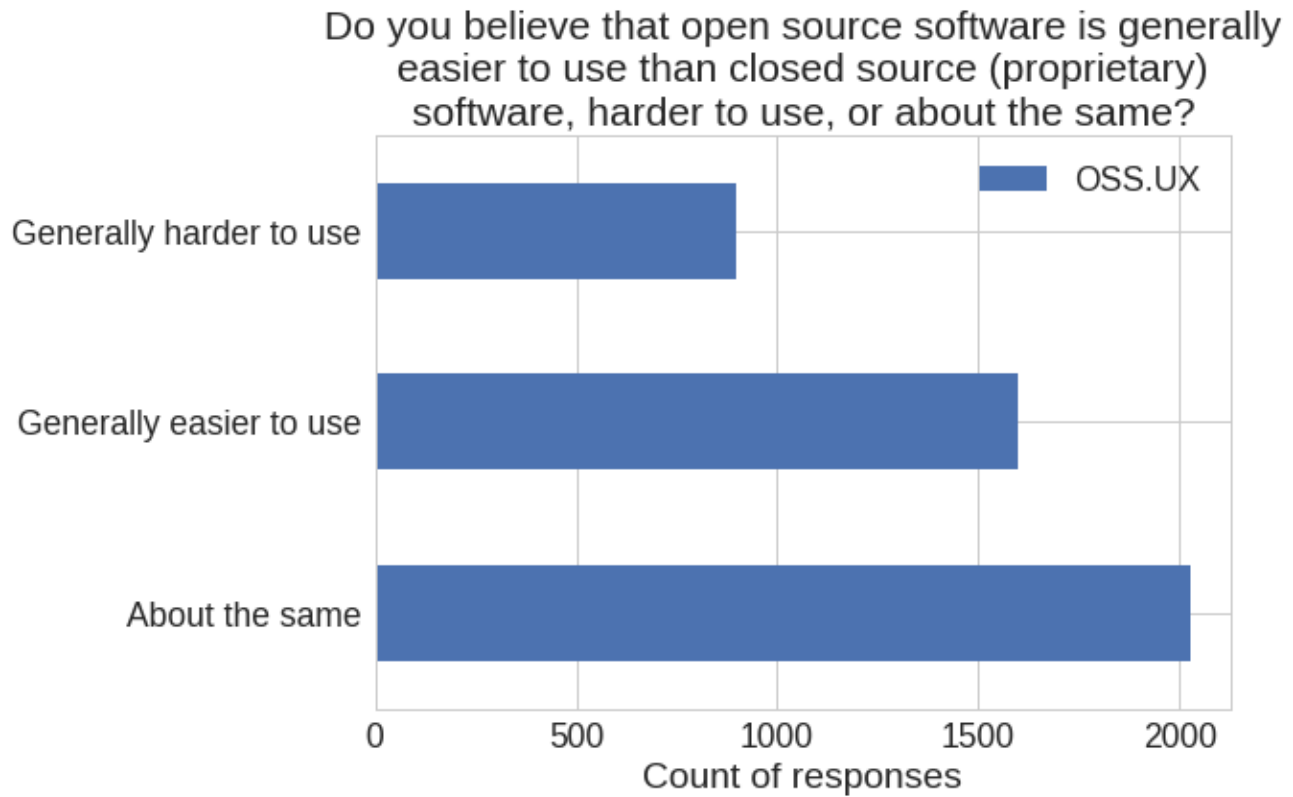
OSS.UX: Do you believe that open source software is generally easier to use than closed source (proprietary) software, harder to use, or about the same?

Out [59] :

	count
About the same	2027
Generally easier to use	1597
Generally harder to use	897

Out [60] :

	percent
About the same	44.84%
Generally easier to use	35.32%
Generally harder to use	19.84%



4.2.5 Open source software security

OSS.SECURITY: Do you believe that open source software is generally more secure than closed source (proprietary) software, less secure, or about the same?

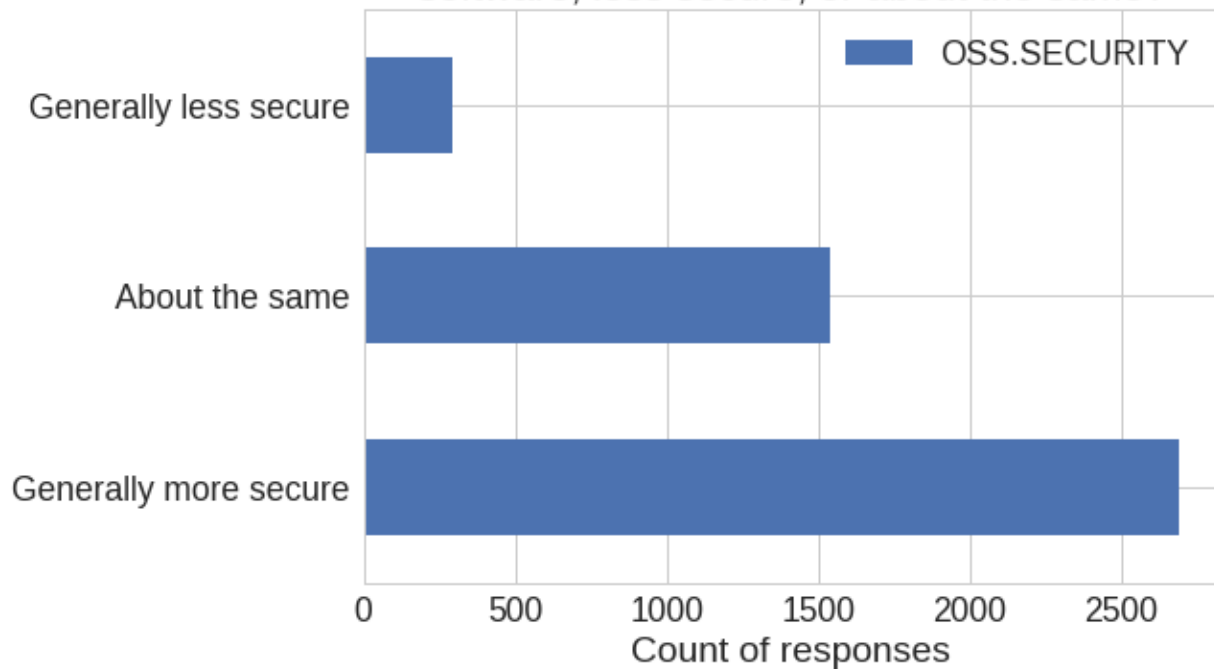
Out [62] :

	count
Generally more secure	2688
About the same	1537
Generally less secure	295

Out [63] :

	percent
Generally more secure	59.47%
About the same	34.00%
Generally less secure	6.53%

Do you believe that open source software is generally more secure than closed source (proprietary) software, less secure, or about the same?



4.2.6 Open source software stability

OSS.STABILITY: Do you believe that open source software is generally more stable than closed source (proprietary) software, less stable, or about the same?

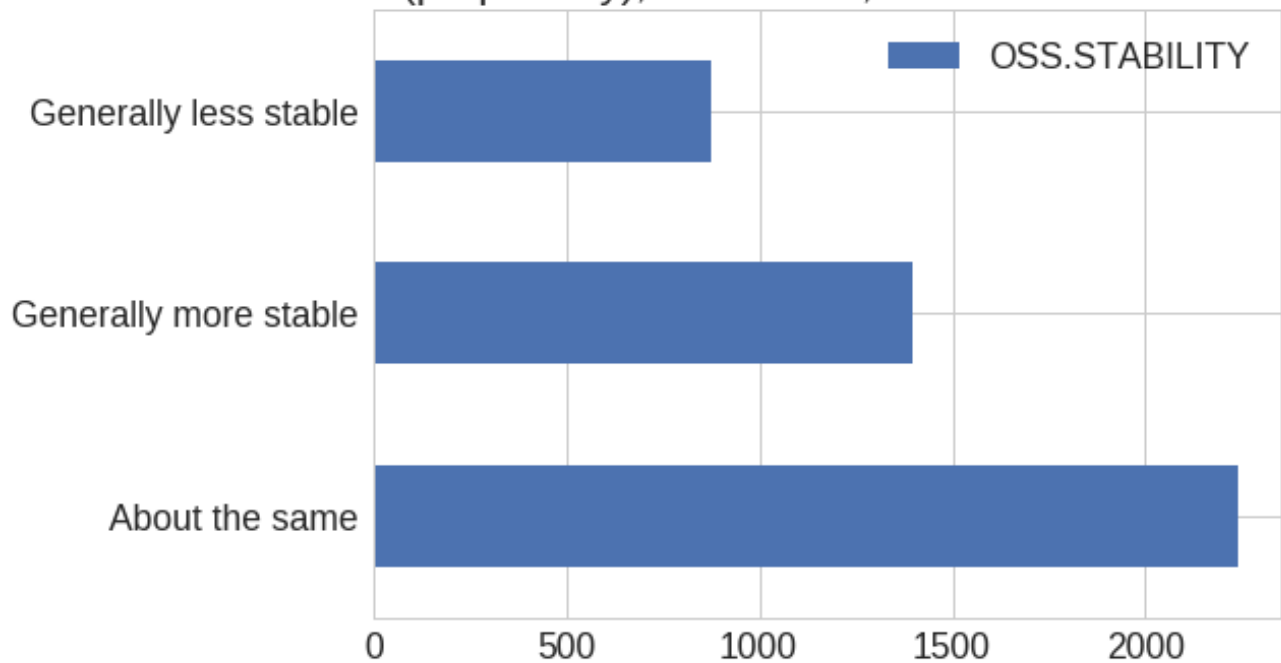
Out [65] :

	count
About the same	2240
Generally more stable	1399
Generally less stable	877

Out [66] :

	percent
About the same	49.60%
Generally more stable	30.98%
Generally less stable	19.42%

Do you believe that open source software is generally more stable than closed source (proprietary), less stable, or about the same?



4.2.7 Identification with open source

How much do you agree or disagree with the following statements:

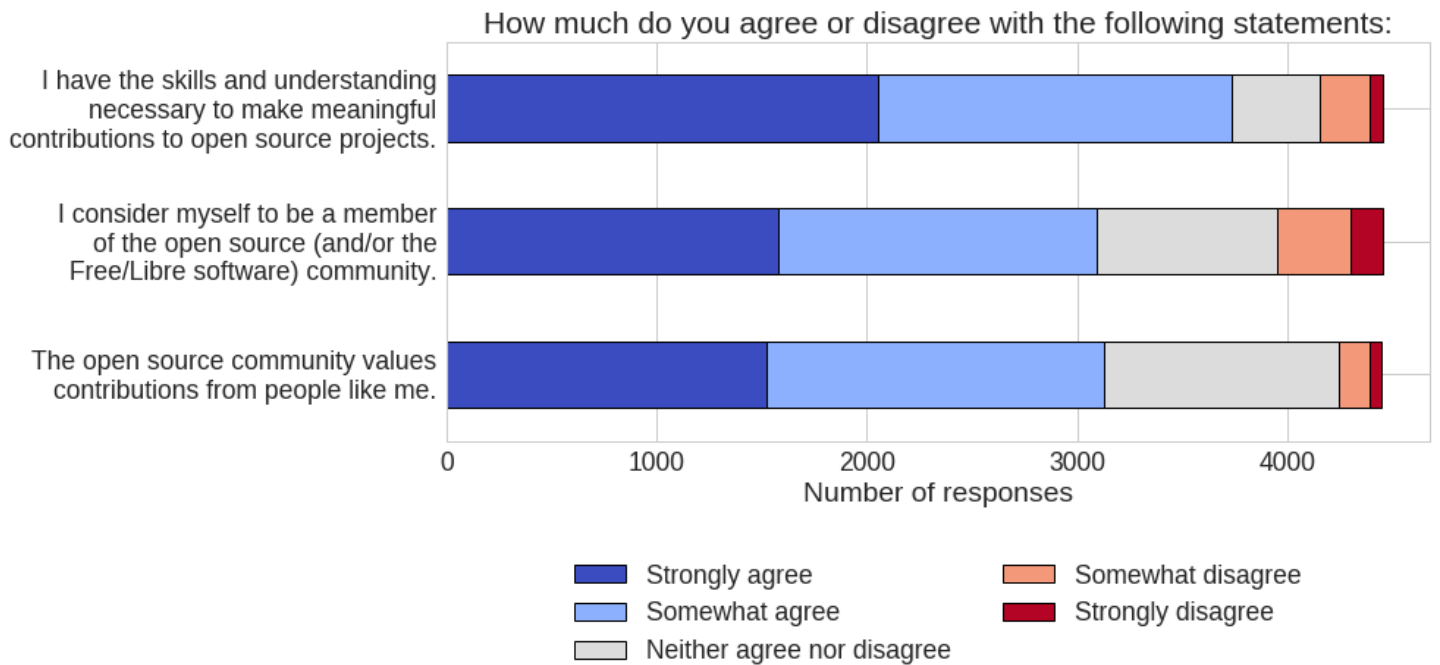
- EXTERNAL.EFFICACY: The open source community values contributions from people like me.
- INTERNAL.EFFICACY: I have the skills and understanding necessary to make meaningful contributions to open source projects.
- OSS.IDENTIFICATION: I consider myself to be a member of the open source (and/or the Free/Libre software) community.

Out [69] :

	Strongly agree	Somewhat agree	Neither	Somewhat disagree	Strongly disagree
EXTERNAL.EFFICACY	1518	1610	1116	150	58
OSS.IDENTIFICATION	1579	1513	863	351	150
INTERNAL.EFFICACY	2052	1685	418	240	62

Out [70] :

	Strongly agree	Somewhat agree	Neither	Somewhat disagree	Strongly disagree
INTERNAL.EFFICACY	9.38%	37.81%	5.38%	46.04%	1.39%
EXTERNAL.EFFICACY	25.07%	36.16%	3.37%	34.10%	1.30%
OSS.IDENTIFICATION	19.37%	33.95%	7.88%	35.44%	3.37%



4.3 Transparency vs privacy

4.3.1 Attribution

TRANSPARENCY.PRIVACY.BELIEFS: Which of the following statements is closest to your beliefs about attribution in software development?

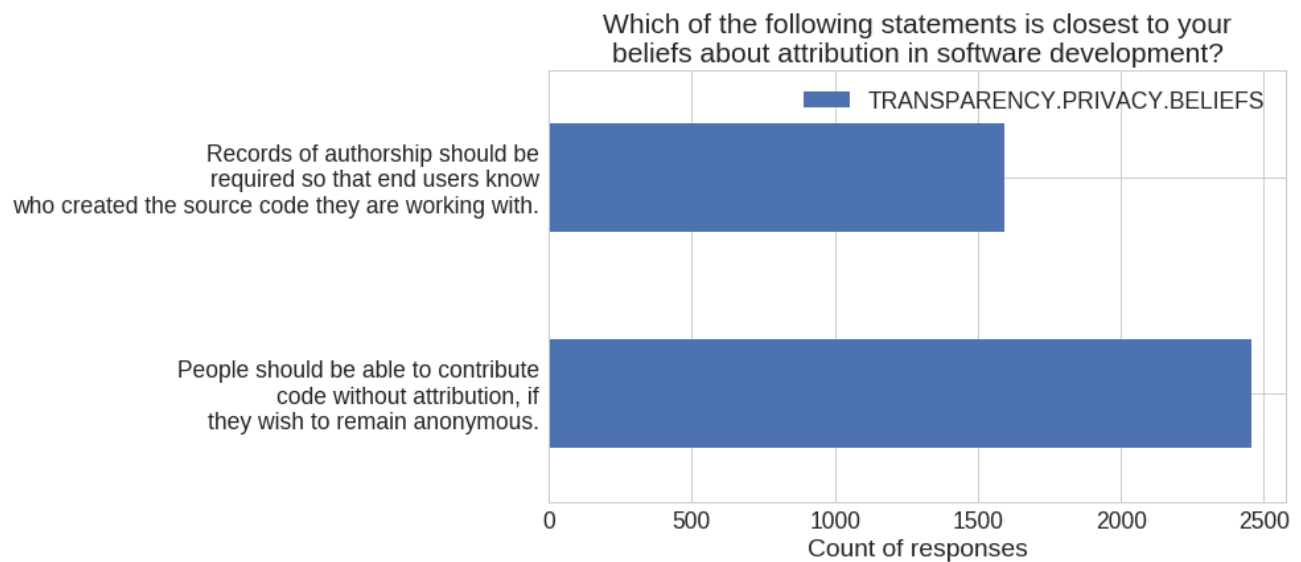
- Records of authorship should be required so that end users know who created the source code they are working with.
- People should be able to contribute code without attribution, if they wish to remain anonymous.

Out [72] :

	count
People should be able to contribute code without attribution, if they wish to remain anonymous.	2454
Records of authorship should be required so that end users know who created the source code they are working with.	1594

Out [73] :

	percent
People should be able to contribute code without attribution, if they wish to remain anonymous.	60.62%
Records of authorship should be required so that end users know who created the source code they are working with.	39.38%



4.3.2 In general, how much information about you is publicly available online?

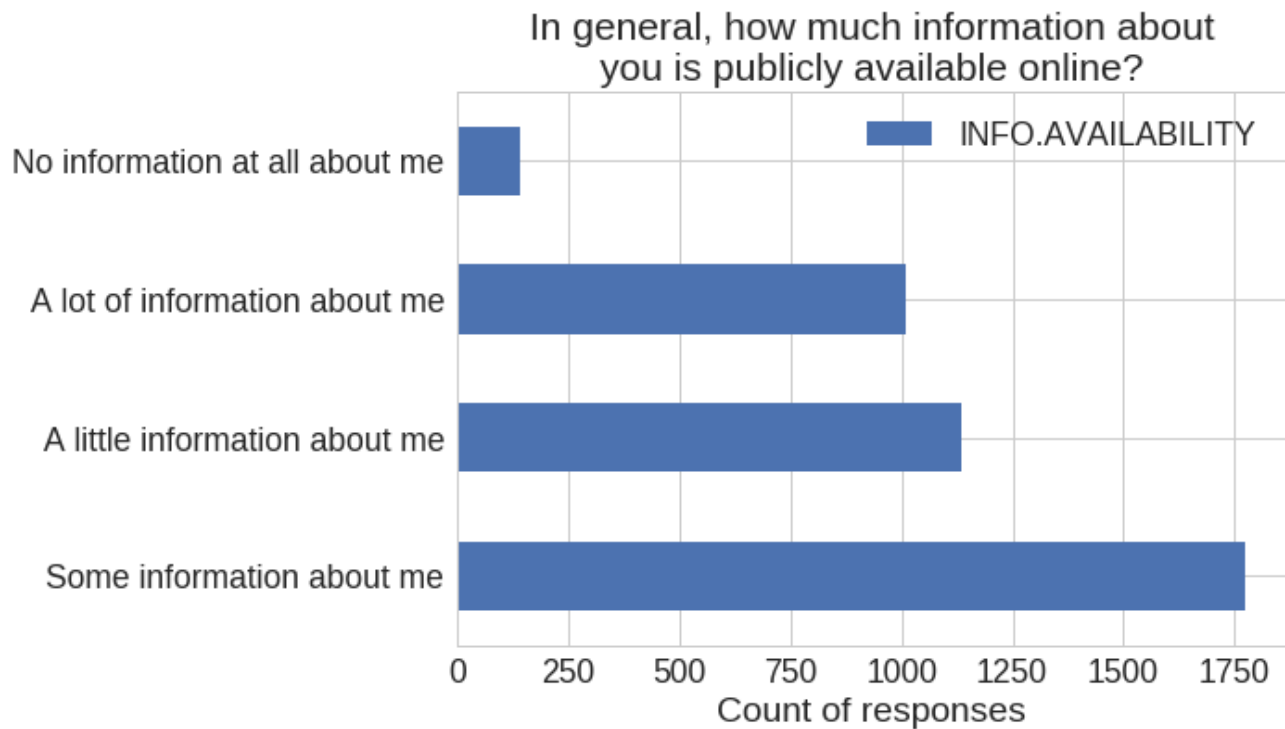
INFO.AVAILABILITY

Out [75] :

	count
Some information about me	1776
A little information about me	1133
A lot of information about me	1011
No information at all about me	140

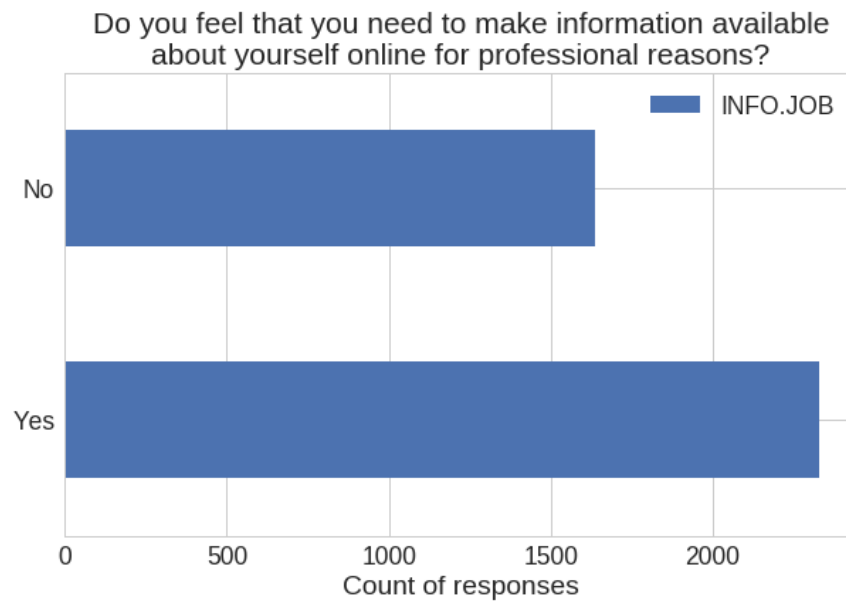
Out [76] :

	percent
Some information about me	43.74%
A little information about me	27.91%
A lot of information about me	24.90%
No information at all about me	3.45%



4.3.3 Do you feel that you need to make information available about yourself online for professional reasons?

INFO.JOB	
Out [78] :	
count	
Yes	2327
No	1638
Out [79] :	
percent	
Yes	58.69%
No	41.31%



4.3.4 General privacy practices

TRANSPARENCY.PRIVACY.PRACTICES.GENERAL

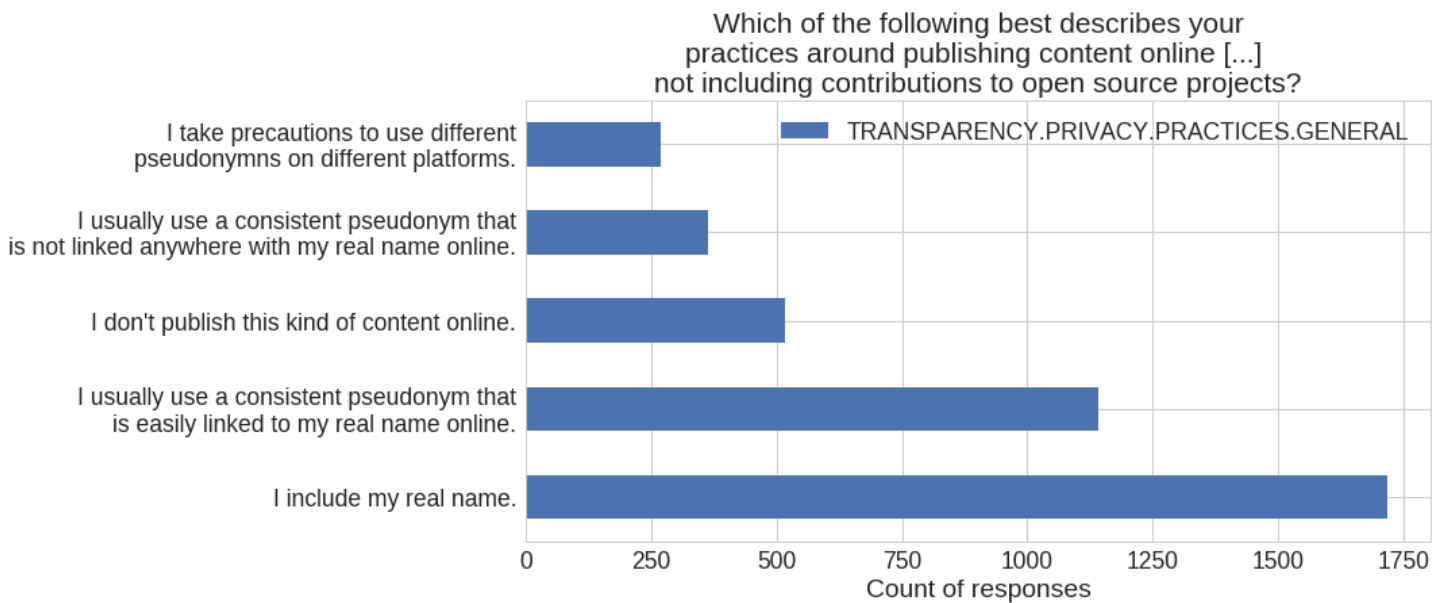
"Which of the following best describes your practices around publishing content online, such as posts on social media (e.g. Facebook, Instagram, Twitter, etc.), blogs, and other platforms (not including contributions to open source projects)?" (single choice)

Out [81]:

	count
I include my real name.	1718
I usually use a consistent pseudonym that is easily linked to my real name online.	1141
I don't publish this kind of content online.	517
I usually use a consistent pseudonym that is not linked anywhere with my real name online	363
I take precautions to use different pseudonymns on different platforms.	270

Out [82]:

	percent
I include my real name.	42.85%
I usually use a consistent pseudonym that is easily linked to my real name online.	28.46%
I don't publish this kind of content online.	12.90%
I usually use a consistent pseudonym that is not linked anywhere with my real name online	9.05%
I take precautions to use different pseudonymns on different platforms.	6.73%



4.3.5 OSS privacy practices

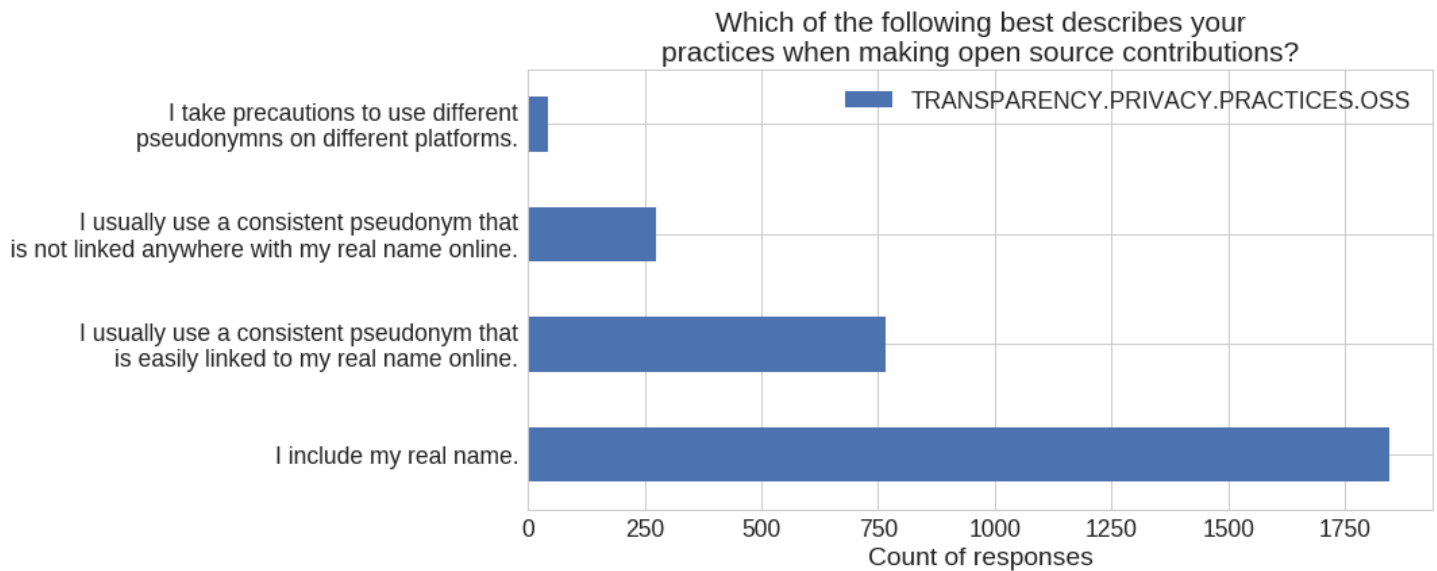
"Which of the following best describes your practices when making open source contributions?"

Out [85]:

	count
I include my real name.	1845
I usually use a consistent pseudonym that is easily linked to my real name online.	766
I usually use a consistent pseudonym that is not linked anywhere with my real name online	273
I take precautions to use different pseudonymns on different platforms.	42

Out [86]:

	percent
I include my real name.	63.06%
I usually use a consistent pseudonym that is easily linked to my real name online.	26.18%
I usually use a consistent pseudonym that is not linked anywhere with my real name online	9.33%
I take precautions to use different pseudonymns on different platforms.	1.44%



4.4 Mentorship / Help

4.4.1 Have you ever received any kind of help from other people related to using or contributing to an open source project?

RECEIVED.HELP

Out [89] :

count

Yes 2845

No 1064

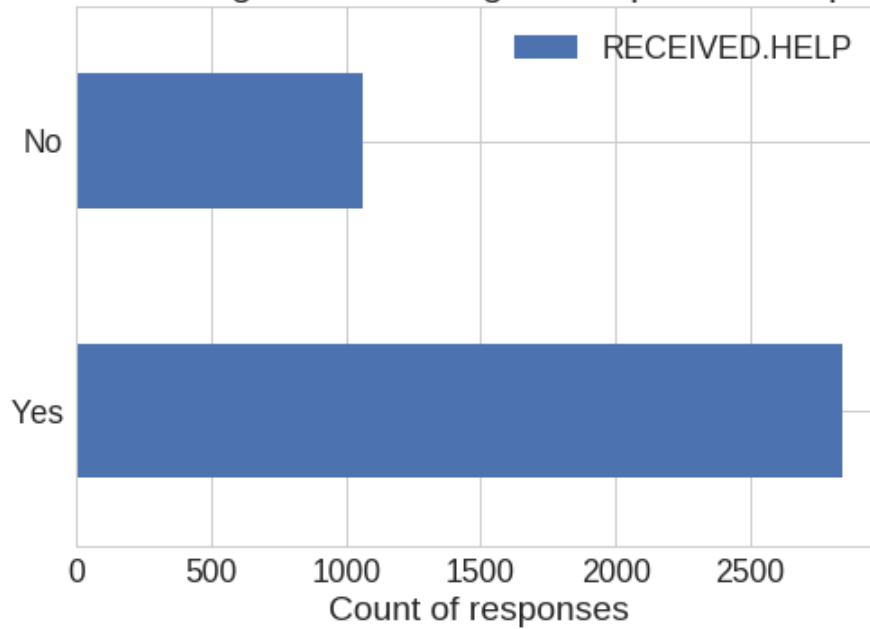
Out [90] :

percent

Yes 72.78%

No 27.22%

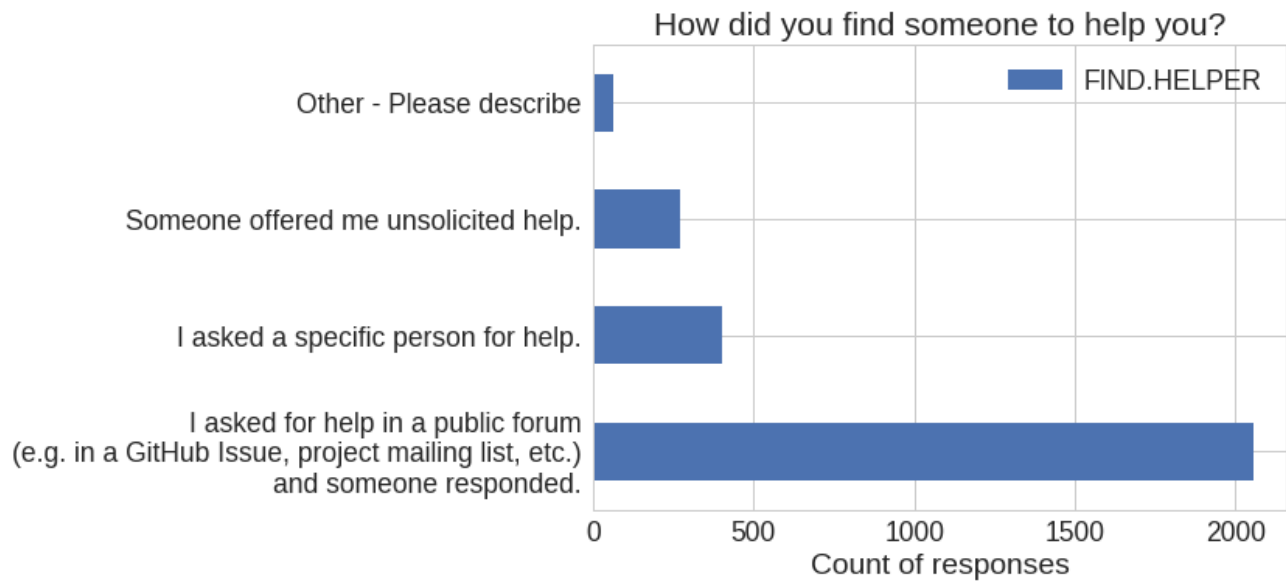
Have you ever received any kind of help from other people related to using or contributing to an open source project?



4.4.2 Thinking of the most recent case where someone helped you, how did you find someone to help you?

Out [92] :	
	count
I asked for help in a public forum	2057
I asked a specific person for help.	403
Someone offered me unsolicited help.	272
Other - Please describe	64

Out [93] :	
	percent
I asked for help in a public forum	73.57%
I asked a specific person for help.	14.41%
Someone offered me unsolicited help.	9.73%
Other - Please describe	2.29%



4.4.3 Which best describes your prior relationship with the person who helped you?

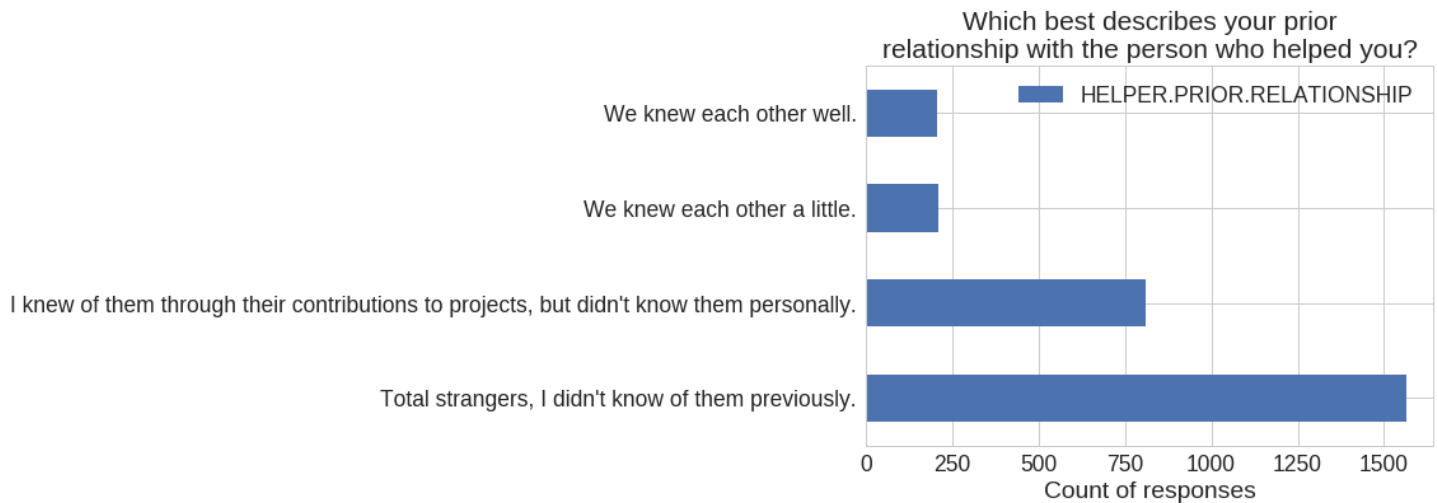
HELPER.PRIOR.RELATIONSHIP

Out [95] :

	count
Total strangers, I didn't know of them previously.	1565
I knew of them through their contributions to projects, but didn't know them personally.	809
We knew each other a little.	211
We knew each other well.	208

Out [96] :

	percent
Total strangers, I didn't know of them previously.	56.03%
I knew of them through their contributions to projects, but didn't know them personally.	28.97%
We knew each other a little.	7.55%
We knew each other well.	7.45%



4.4.4 What kind of problem did they help you with?

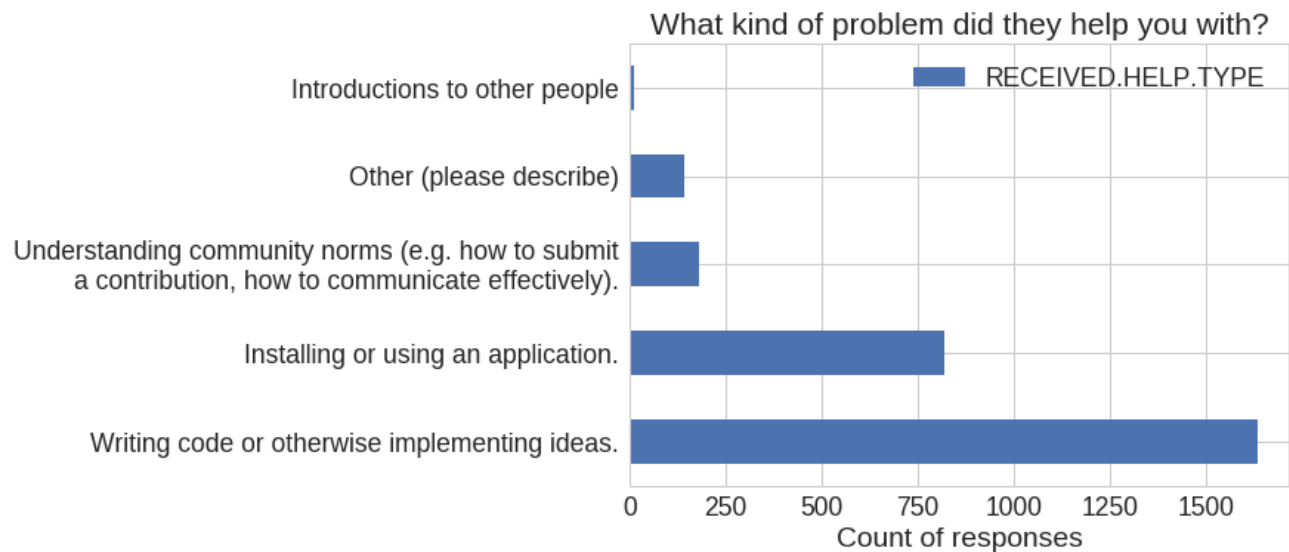
RECEIVED.HELP.TYPE

Out [98] :

	count
Writing code or otherwise implementing ideas.	1633
Installing or using an application.	820
Understanding community norms	181
Other (please describe)	142
Introductions to other people	13

Out [99] :

	percent
Writing code or otherwise implementing ideas.	58.55%
Installing or using an application.	29.40%
Understanding community norms	6.49%
Other (please describe)	5.09%
Introductions to other people	0.47%



4.4.5 Have you ever provided help for another person on an open source project?

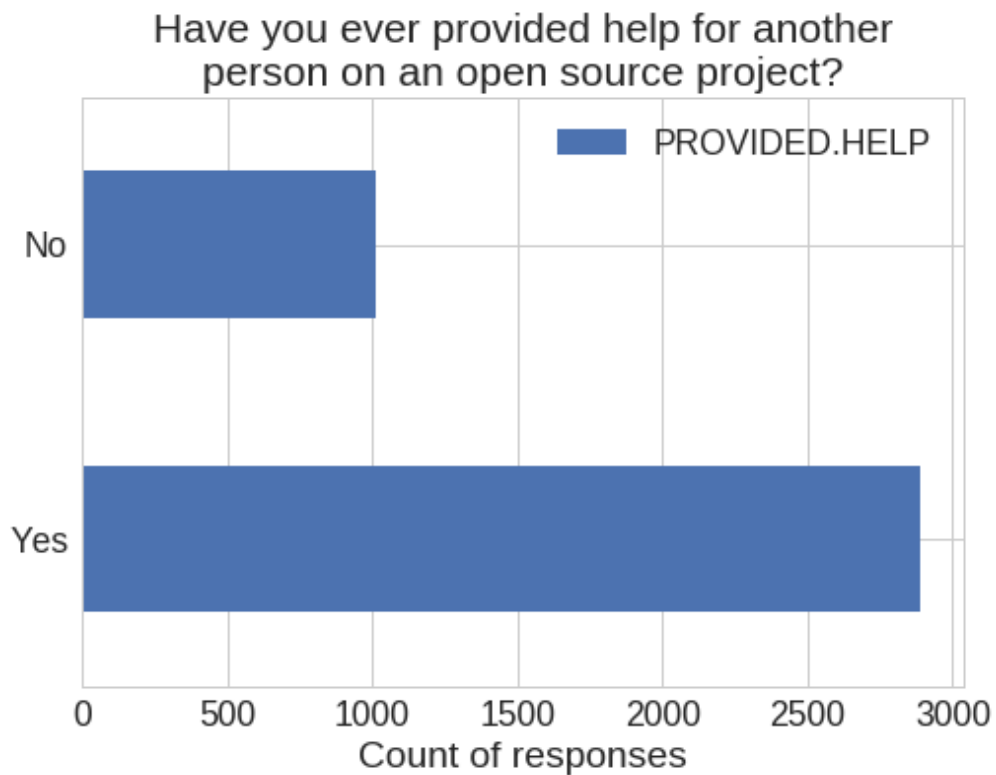
PROVIDED.HELP

Out[101]:

count	
Yes	2891
No	1013

Out[102]:

percent	
Yes	74.05%
No	25.95%



4.4.6 Thinking of the most recent case where you helped someone, how did you come to help this person?

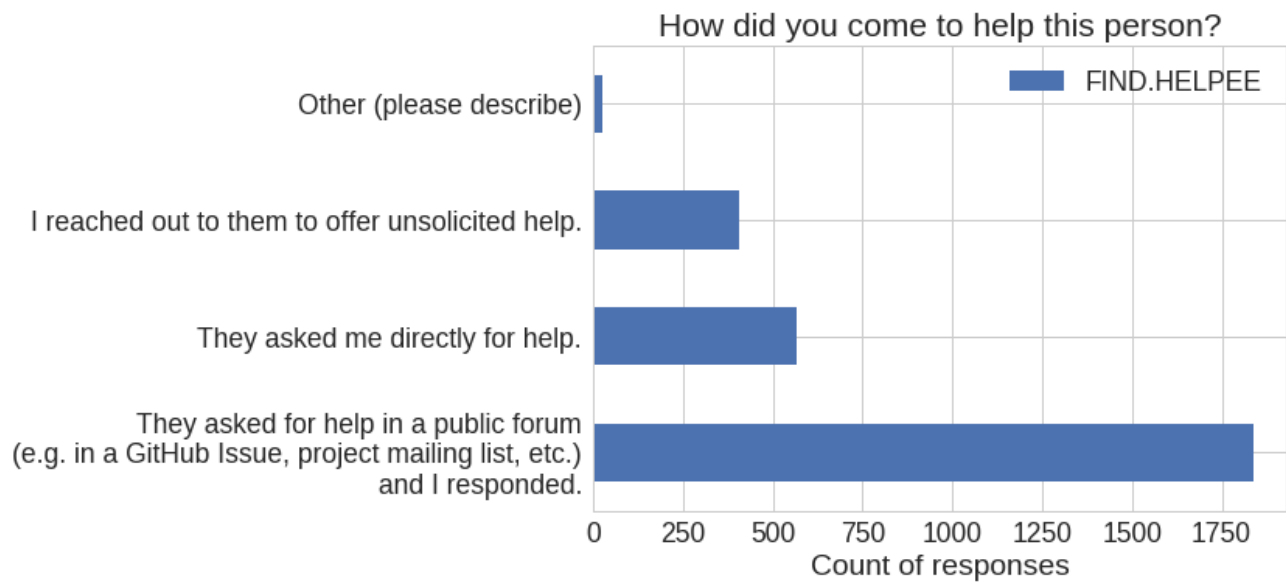
FIND.HELPEE

Out [104] :

	count
They asked for help in a public forum	1839
They asked me directly for help.	566
I reached out to them to offer unsolicited help.	405
Other (please describe)	28

Out [105] :

	percent
They asked for help in a public forum	64.80%
They asked me directly for help.	19.94%
I reached out to them to offer unsolicited help.	14.27%
Other (please describe)	0.99%



4.4.7 Which best describes your prior relationship with the person you helped?

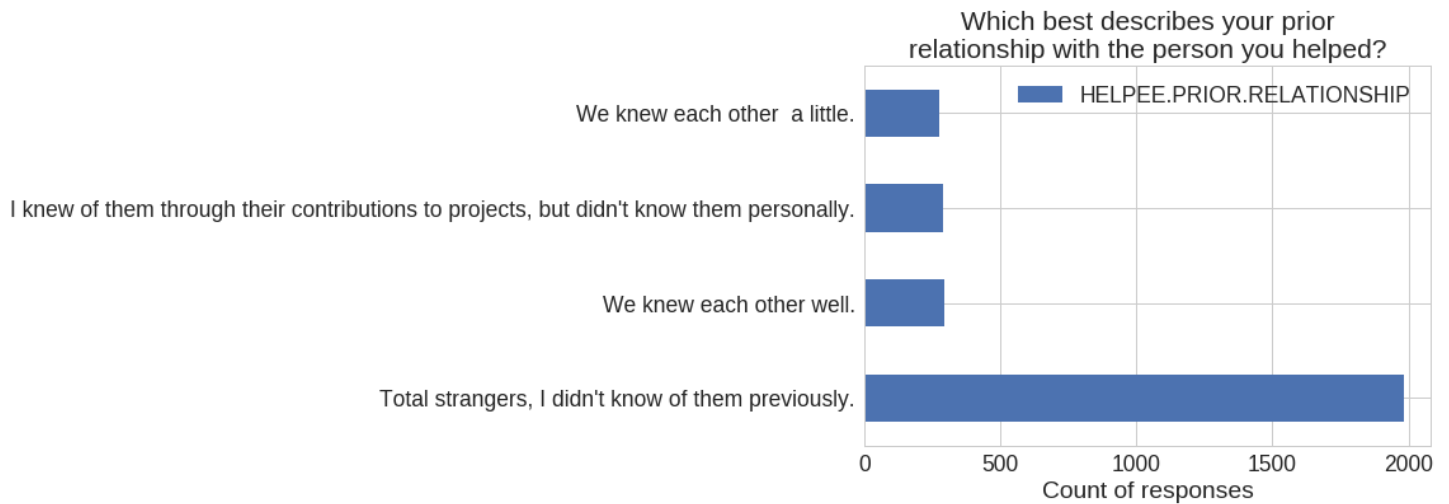
HELPEE.PRIOR.RELATIONSHIP

Out [107] :

	count
Total strangers, I didn't know of them previously.	1984
We knew each other well.	292
I knew of them through their contributions to projects, but didn't know them personally.	288
We knew each other a little.	275

Out [108] :

	percent
Total strangers, I didn't know of them previously.	69.88%
We knew each other well.	10.29%
I knew of them through their contributions to projects, but didn't know them personally.	10.14%
We knew each other a little.	9.69%



4.4.8 What kind of problem did you help them with?

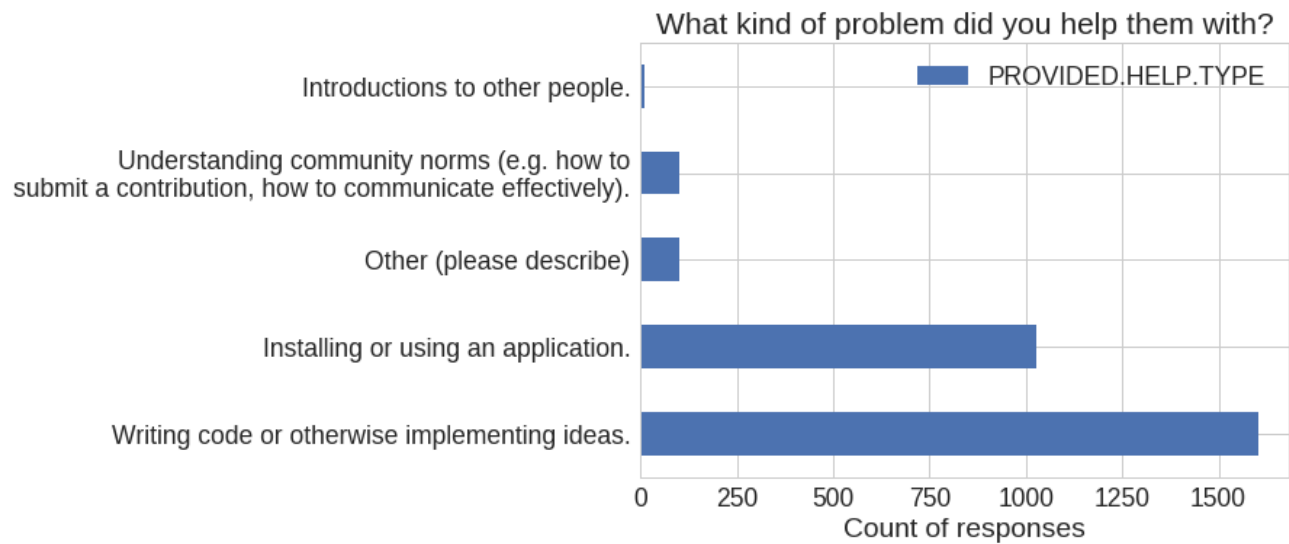
PROVIDED.HELP.TYPE

Out [110] :

	count
Writing code or otherwise implementing ideas.	1602
Installing or using an application.	1028
Other (please describe)	101
Understanding community norms	99
Introductions to other people.	8

Out [111] :

	percent
Writing code or otherwise implementing ideas.	56.45%
Installing or using an application.	36.22%
Other (please describe)	3.56%
Understanding community norms	3.49%
Introductions to other people.	0.28%



4.5 Open Source Software in Paid Work

4.5.1 Do you contribute to open source as part of your professional work?

OSS.AS.JOB: Do you contribute to open source as part of your professional work? In other words, are you paid for any of your time spent on open source contributions?

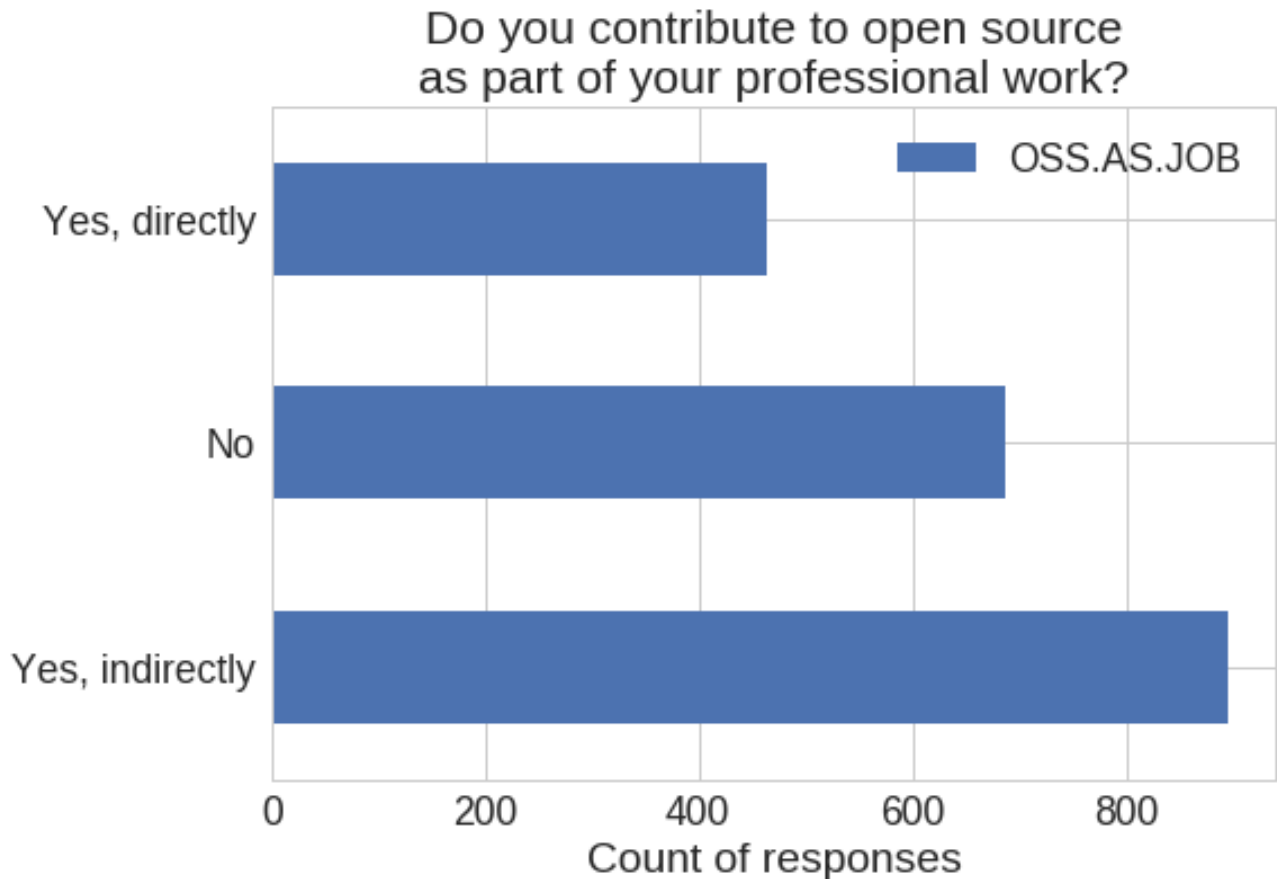
- Yes, indirectly- I contribute to open source in carrying out my work duties, but I am not required or expected to do so.
- No.
- Yes, directly- some or all of my work duties include contributing to open source projects.

Out [113] :

	count
Yes, indirectly	896
No.	687
Yes, directly	464

Out [114] :

	percent
Yes, indirectly	43.77%
No.	33.56%
Yes, directly	22.67%



4.5.2 How often do you use open source software in your professional work?

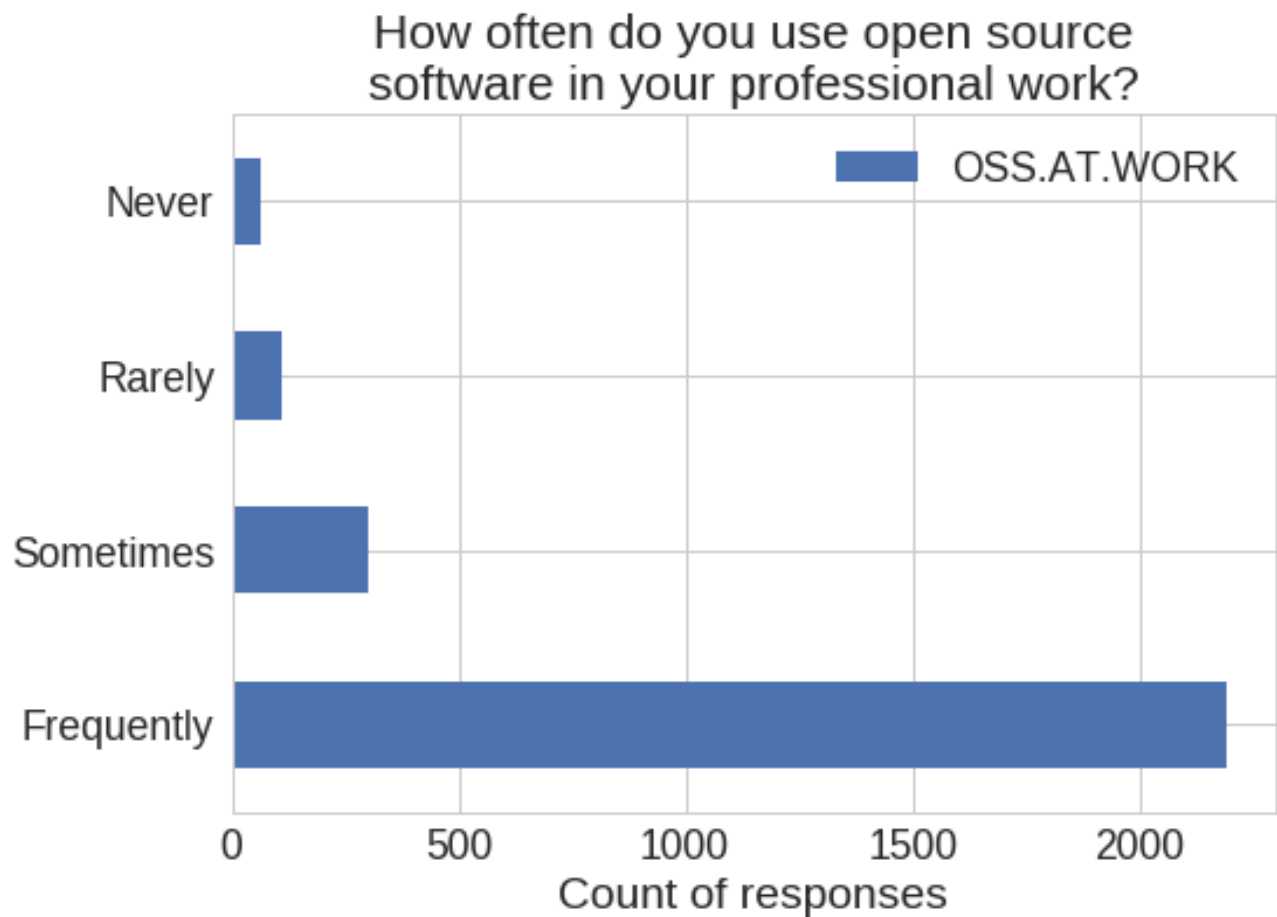
OSS.AT.WORK

Out [116] :

	count
Frequently	2191
Sometimes	300
Rarely	110
Never	65

Out [117] :

	percent
Frequently	82.18%
Sometimes	11.25%
Rarely	4.13%
Never	2.44%



4.5.3 How does your employer's intellectual property agreement/policy affect your free-time contributions to open source unrelated to your work?

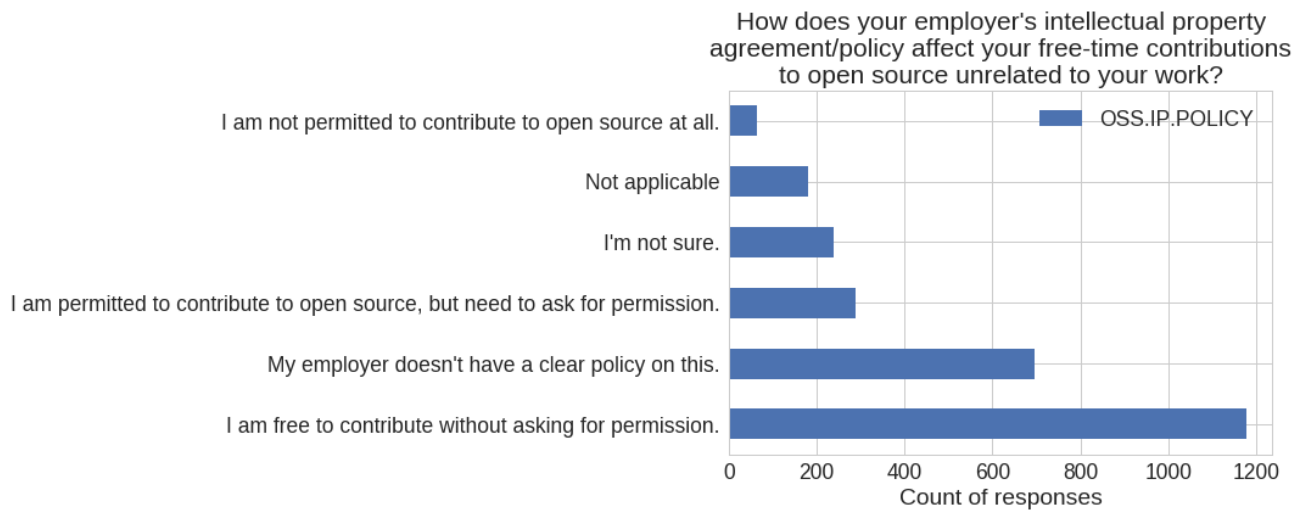
OSS.IP.POLICY

Out [119] :

	count
I am free to contribute without asking for permission.	1178
My employer doesn't have a clear policy on this.	695
I am permitted to contribute to open source, but need to ask for permission	287
I'm not sure.	238
Not applicable	180
I am not permitted to contribute to open source at all.	63

Out [120] :

	percent
I am free to contribute without asking for permission.	44.60%
My employer doesn't have a clear policy on this.	26.32%
I am permitted to contribute to open source, but need to ask for permission	10.87%
I'm not sure.	9.01%
Not applicable	6.82%
I am not permitted to contribute to open source at all.	2.39%



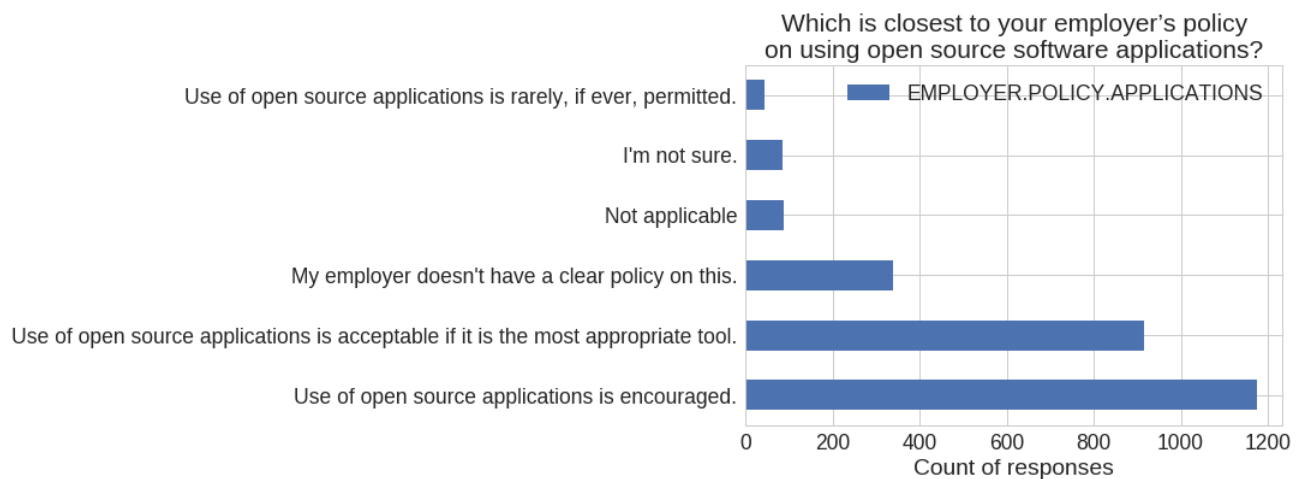
4.5.4 Which is closest to your employer's policy on using open source software applications?

Out [122] :

	count
Use of open source applications is encouraged.	1174
Use of open source applications is acceptable if it is the most appropriate tool.	916
My employer doesn't have a clear policy on this.	338
Not applicable	88
I'm not sure.	83
Use of open source applications is rarely, if ever, permitted.	42

Out [123] :

	percent
Use of open source applications is encouraged.	44.45%
Use of open source applications is acceptable if it is the most appropriate tool.	34.68%
My employer doesn't have a clear policy on this.	12.80%
Not applicable	3.33%
I'm not sure.	3.14%
Use of open source applications is rarely, if ever, permitted.	1.59%



4.5.5 How important do you think your involvement in open source was to getting your current job?

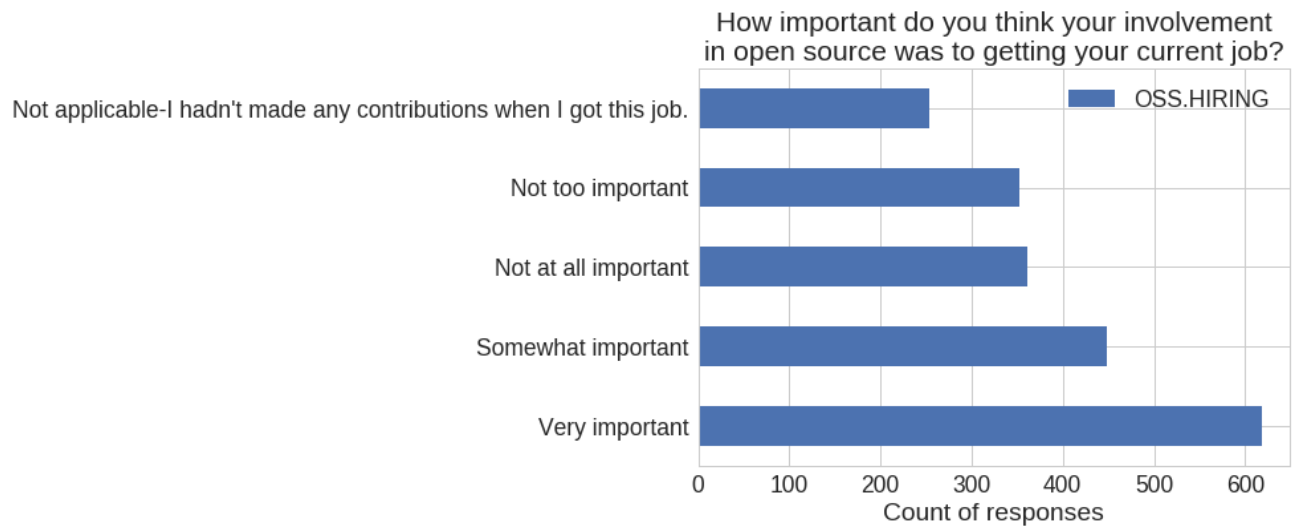
OSS.HIRING

Out [125] :

	count
Very important	618
Somewhat important	448
Not at all important	361
Not too important	352
Not applicable-I hadn't made any contributions	254

Out [126] :

	percent
Very important	30.40%
Somewhat important	22.04%
Not at all important	17.76%
Not too important	17.31%
Not applicable-I hadn't made any contributions	12.49%



4.6 Demographics

4.6.1 Do you currently live in a country other than the one in which you were born?

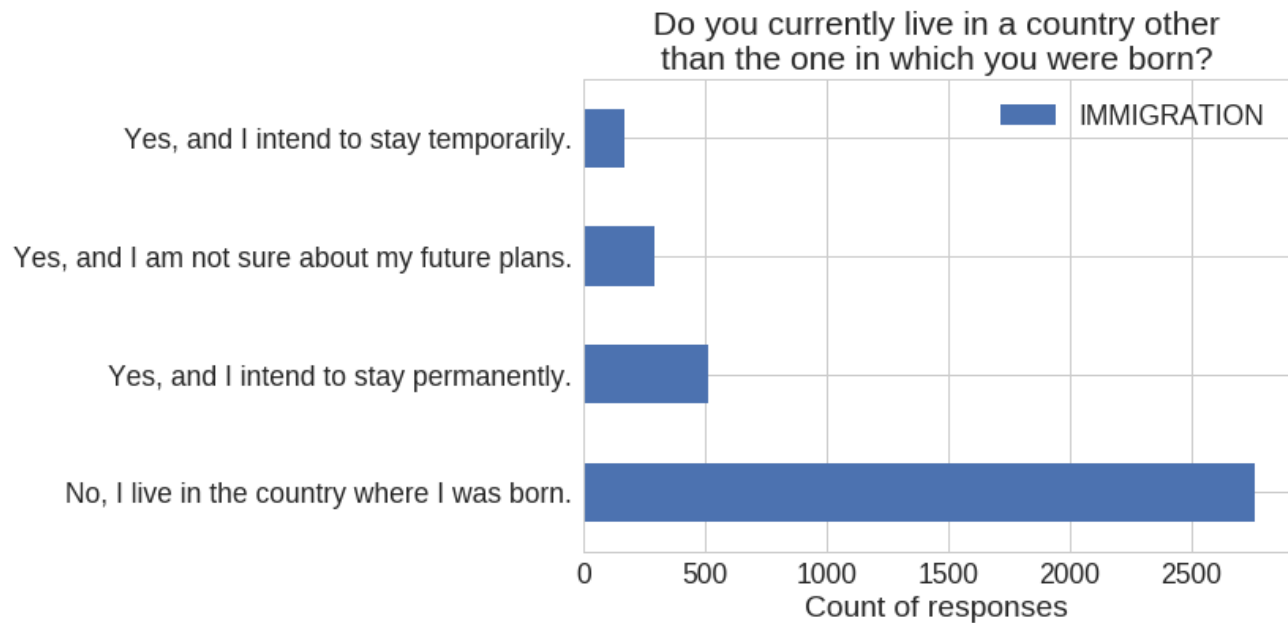
IMMIGRATION

Out[128]:

	count
No, I live in the country where I was born.	2764
Yes, and I intend to stay permanently.	513
Yes, and I am not sure about my future plans.	292
Yes, and I intend to stay temporarily.	165

Out[129]:

	percent
No, I live in the country where I was born.	74.02%
Yes, and I intend to stay permanently.	13.74%
Yes, and I am not sure about my future plans.	7.82%
Yes, and I intend to stay temporarily.	4.42%



4.6.2 Thinking of where you were born, are you a member of an ethnicity or nationality that is a considered a minority in that country?

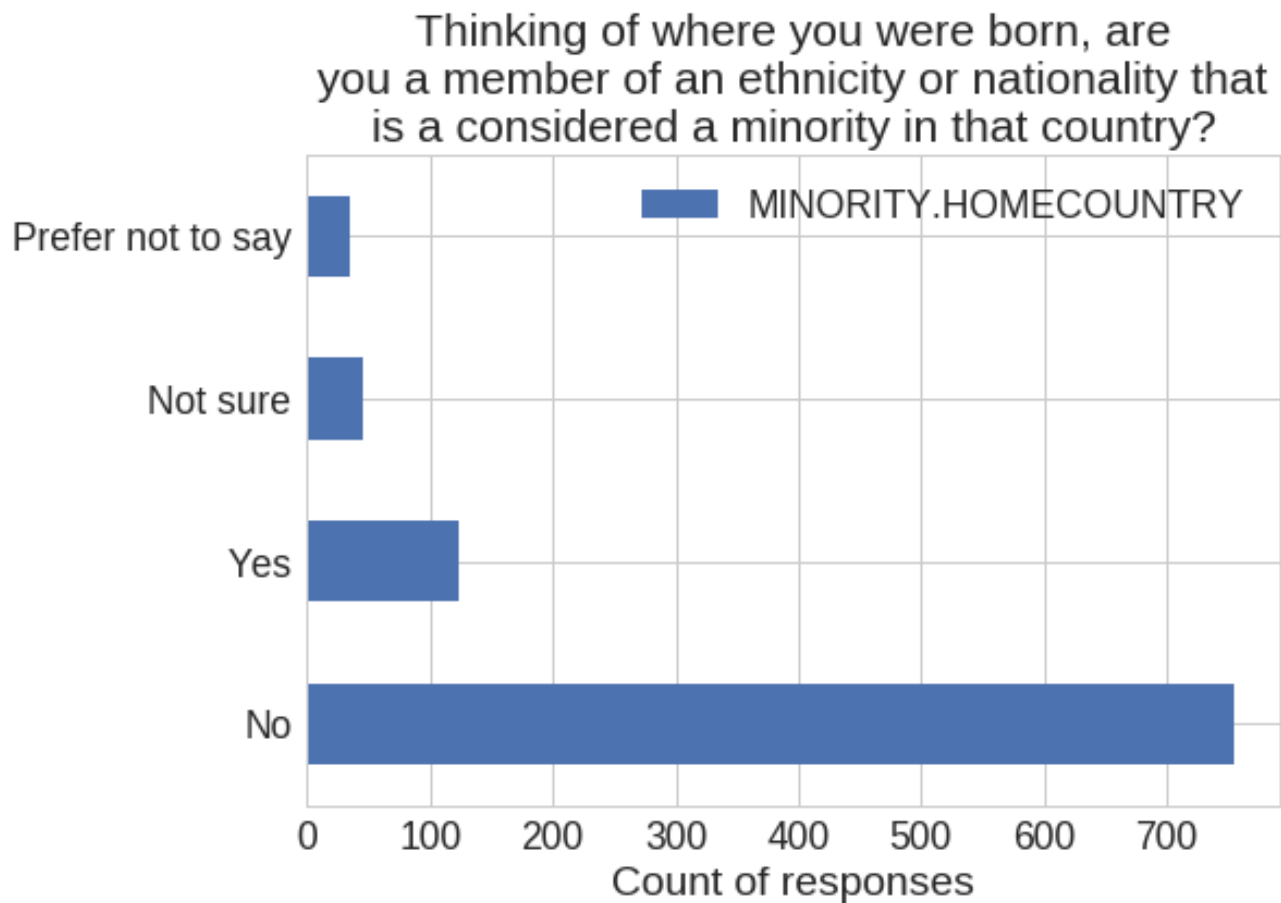
MINORITY.HOMECOUNTRY

Out [131] :

	count
No	754
Yes	124
Not sure	45
Prefer not to say	34

Out [132] :

	percent
No	78.79%
Yes	12.96%
Not sure	4.70%
Prefer not to say	3.55%



4.6.3 Thinking of where you currently live, are you a member of an ethnicity or nationality that is a considered a minority in that country?

MINORITY.CURRENT.COUNTRY

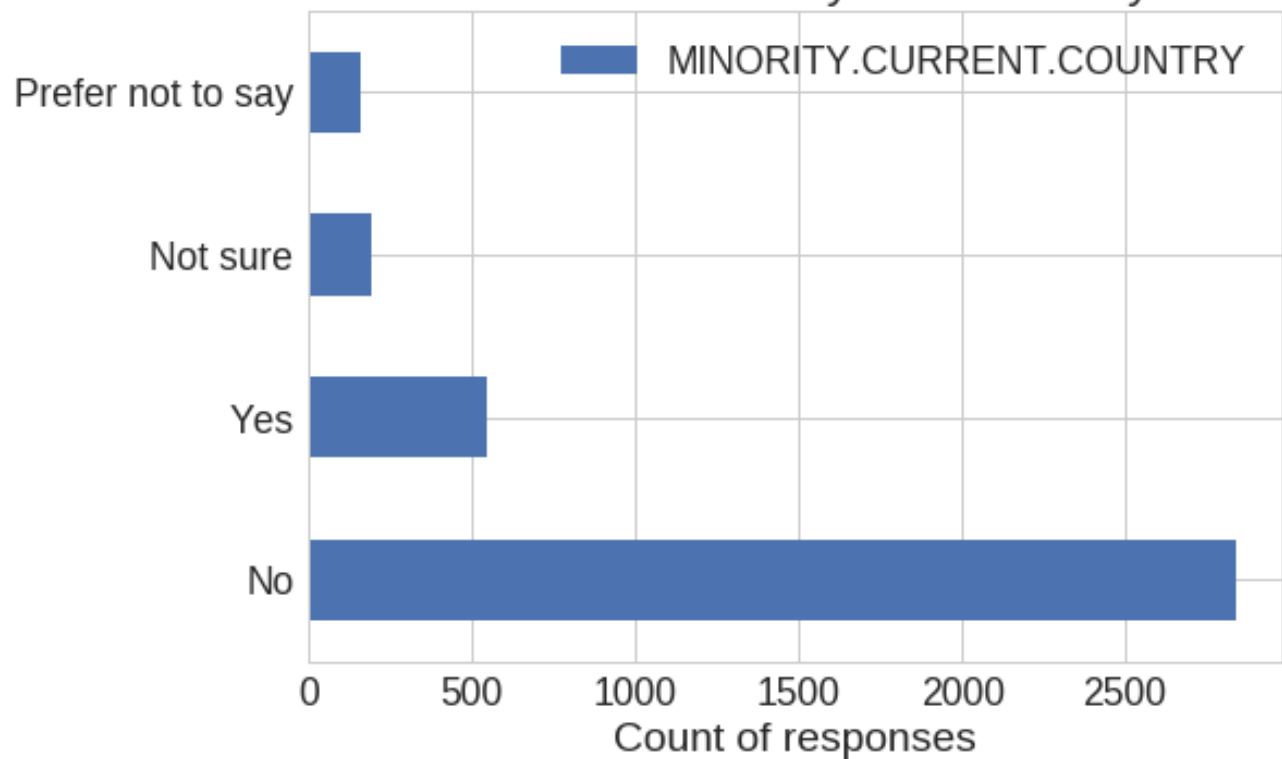
Out [134] :

	count
No	2837
Yes	546
Not sure	193
Prefer not to say	156

Out [135] :

	percent
No	76.02%
Yes	14.63%
Not sure	5.17%
Prefer not to say	4.18%

Thinking of where you currently live, are you a member of an ethnicity or nationality that is a considered a minority in that country?



4.6.4 What is your gender?

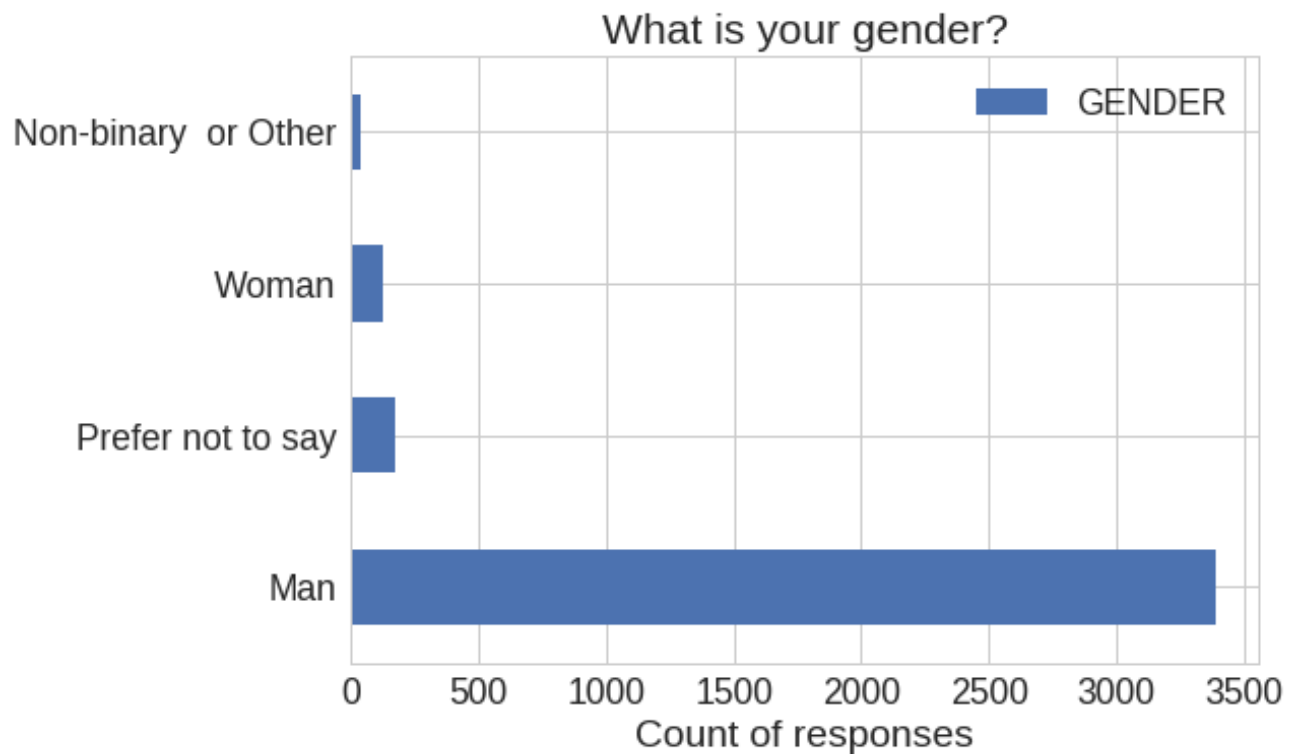
GENDER

Out[137]:

	count
Man	3387
Prefer not to say	173
Woman	125
Non-binary or Other	39

Out[138]:

	percent
Man	90.95%
Prefer not to say	4.65%
Woman	3.36%
Non-binary or Other	1.05%



4.6.5 Do you identify as transgender?

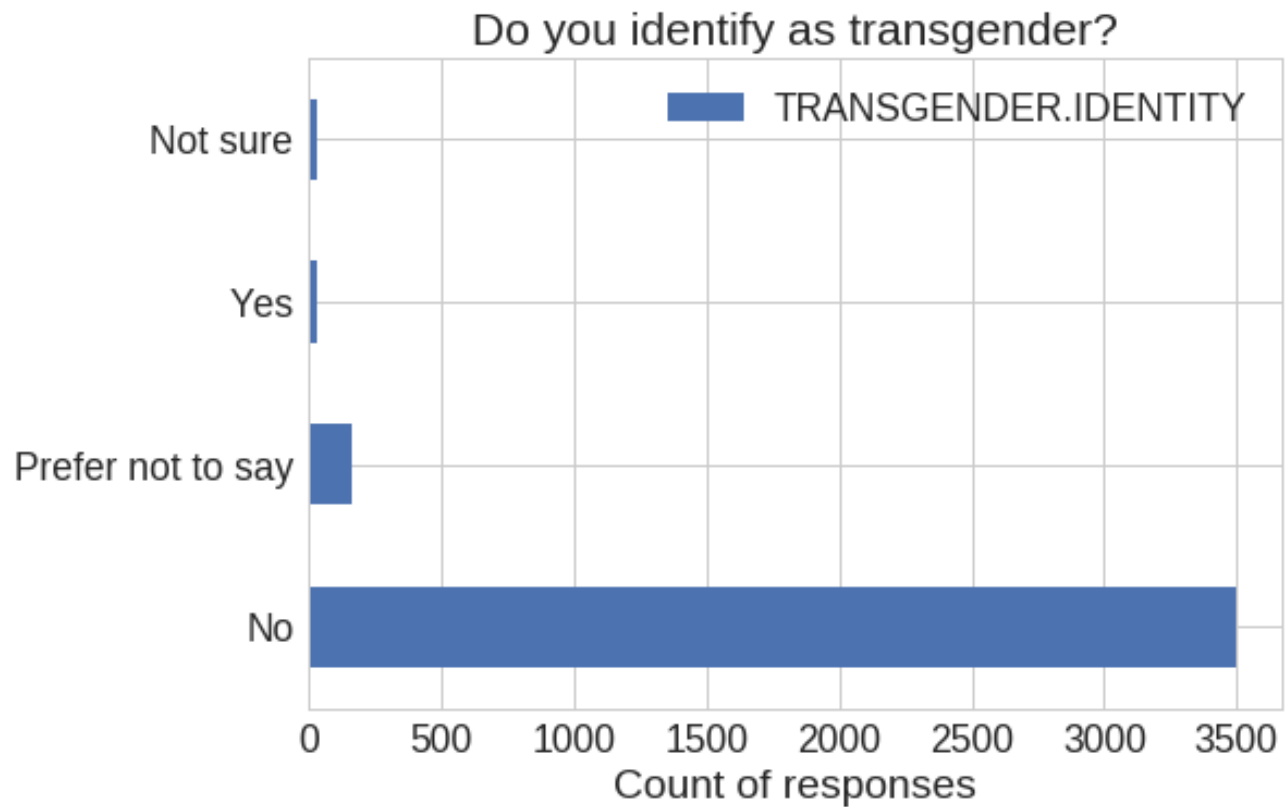
TRANSGENDER.IDENTITY

Out [140] :

	count
No	3494
Prefer not to say	158
Yes	33
Not sure	30

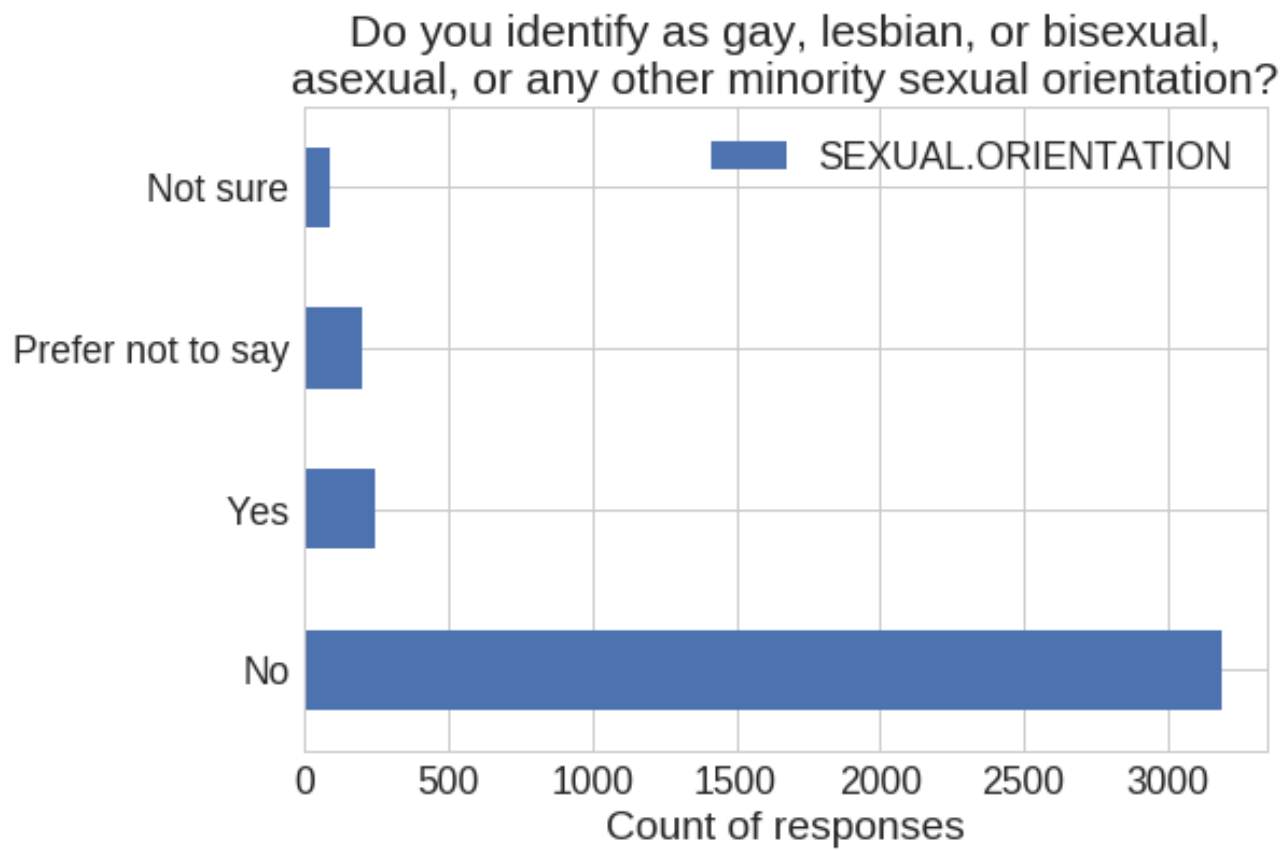
Out [141] :

	percent
No	94.05%
Prefer not to say	4.25%
Yes	0.89%
Not sure	0.81%



4.6.6 Do you identify as gay, lesbian, or bisexual, asexual, or any other minority sexual orientation?

SEXUAL.ORIENTATION	
Out [143] :	
count	
No	3187
Yes	246
Prefer not to say	201
Not sure	85
Out [144] :	
percent	
No	85.70%
Yes	6.61%
Prefer not to say	5.40%
Not sure	2.29%



4.6.7 How well can you read and write in English?

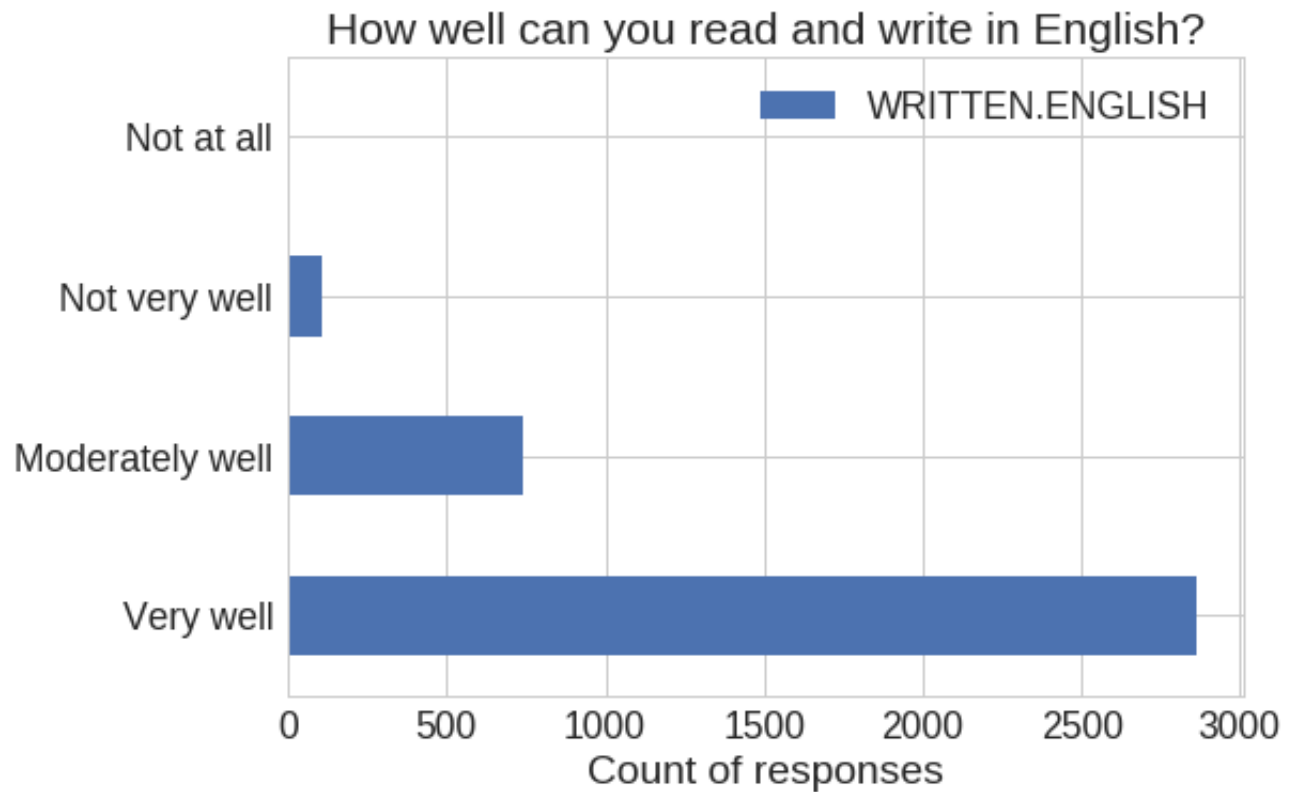
WRITTEN.ENGLISH

Out [146] :

count	
Very well	2865
Moderately well	742
Not very well	108
Not at all	6

Out [147] :

percent	
Very well	77.00%
Moderately well	19.94%
Not very well	2.90%
Not at all	0.16%



4.6.8 What is your age?

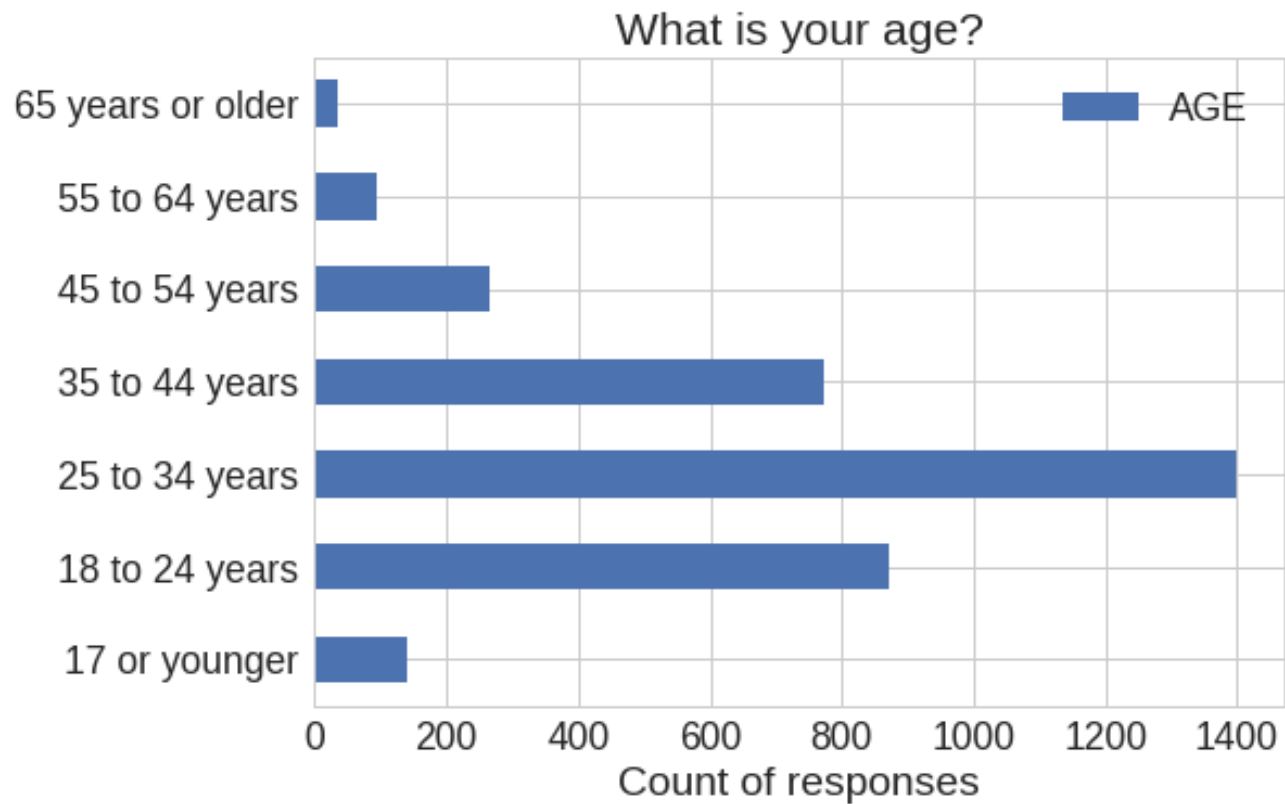
AGE

Out [149] :

	count
17 or younger	139
18 to 24 years	871
25 to 34 years	1400
35 to 44 years	772
45 to 54 years	267
55 to 64 years	93
65 years or older	36

Out [150] :

	percent
17 or younger	3.88%
18 to 24 years	24.34%
25 to 34 years	39.13%
35 to 44 years	21.58%
45 to 54 years	7.46%
55 to 64 years	2.60%
65 years or older	1.01%



4.6.9 What is highest level of formal education that you have completed?

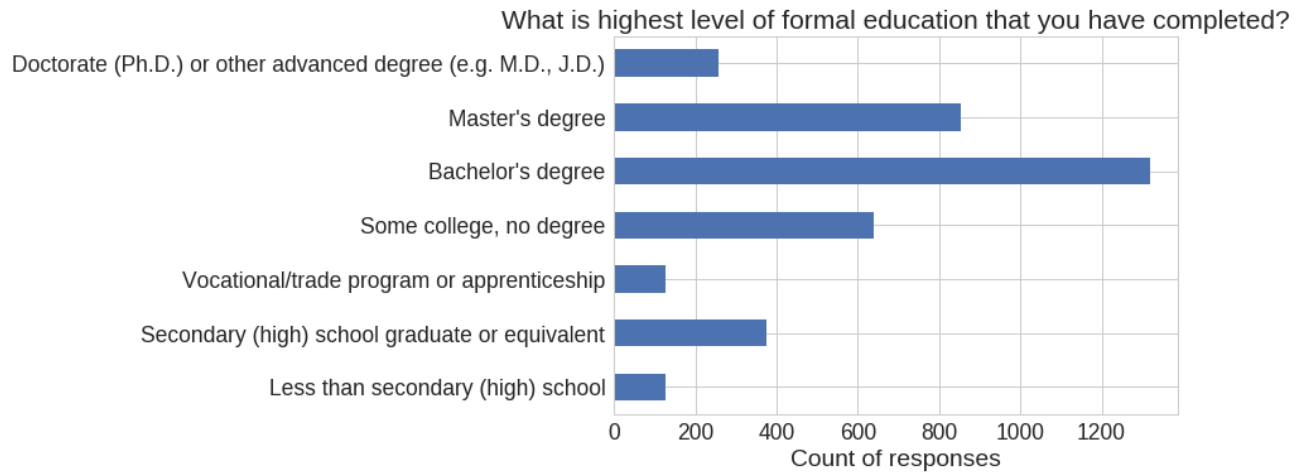
FORMAL.EDUCATION

Out [152] :

	count
Bachelor's degree	1321
Master's degree	852
Some college, no degree	640
Secondary (high) school graduate or equivalent	375
Doctorate (Ph.D.) or other advanced degree	256
Vocational/trade program or apprenticeship	127
Less than secondary (high) school	126

Out [153] :

	percent
Bachelor's degree	35.73%
Master's degree	23.05%
Some college, no degree	17.31%
Secondary (high) school graduate or equivalent	10.14%
Doctorate (Ph.D.) or other advanced degree	6.92%
Vocational/trade program or apprenticeship	3.44%
Less than secondary (high) school	3.41%



4.6.10 What is the highest level of formal education that either of your parents completed?

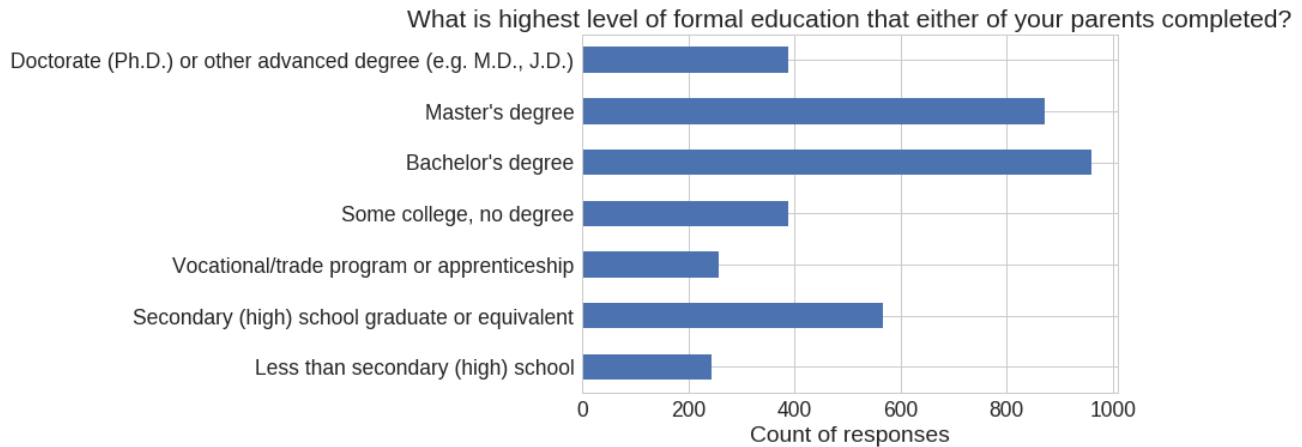
PARENTS.FORMAL.EDUCATION

Out [155] :

	count
Bachelor's degree	961
Master's degree	871
Secondary (high) school graduate or equivalent	566
Some college, no degree	388
Doctorate (Ph.D.) or other advanced degree	387
Vocational/trade program or apprenticeship	257
Less than secondary (high) school	243

Out [156] :

	percent
Bachelor's degree	26.16%
Master's degree	23.71%
Secondary (high) school graduate or equivalent	15.41%
Some college, no degree	10.56%
Doctorate (Ph.D.) or other advanced degree	10.54%
Vocational/trade program or apprenticeship	7.00%
Less than secondary (high) school	6.62%



4.6.11 How old were you when you first had regular access to a computer with an internet connection?

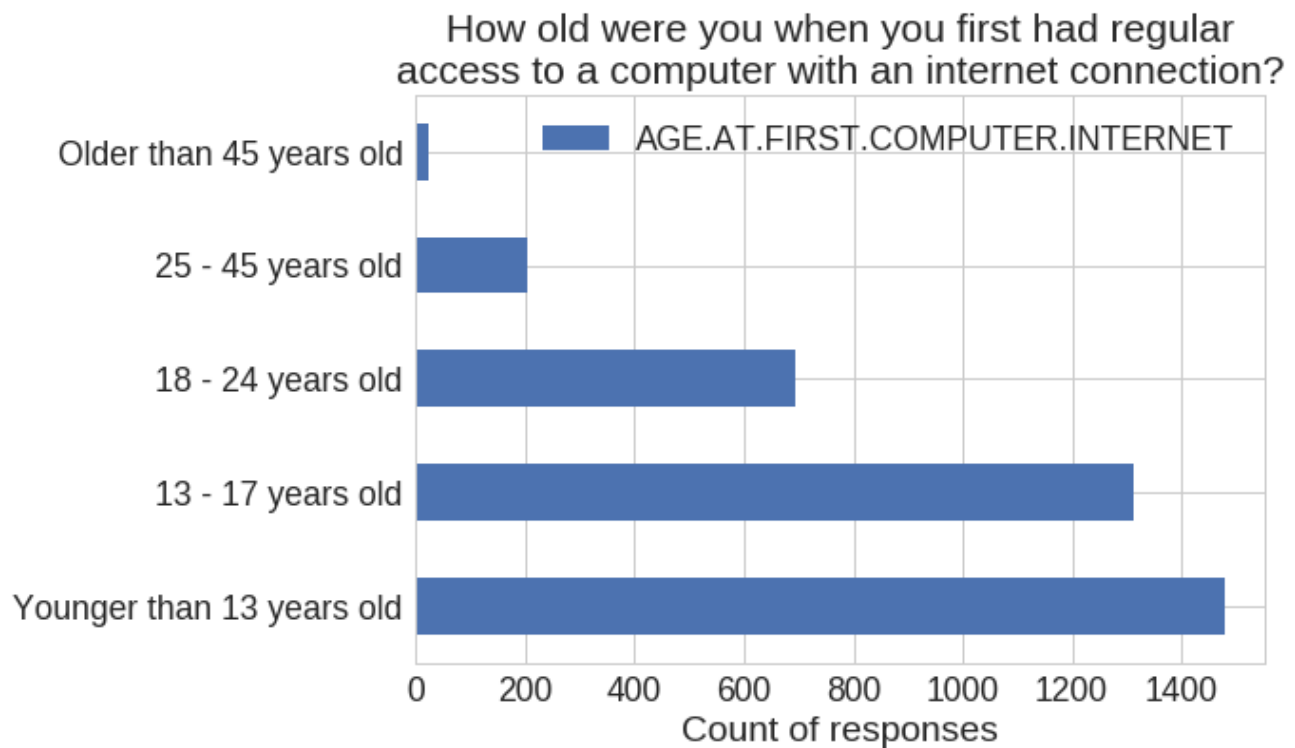
AGE.AT.FIRST.COMPUTER.INTERNET

Out [158] :

	count
Younger than 13 years old	1478
13 - 17 years old	1313
18 - 24 years old	695
25 - 45 years old	202
Older than 45 years old	23

Out [159] :

	percent
Younger than 13 years old	39.83%
13 - 17 years old	35.38%
18 - 24 years old	18.73%
25 - 45 years old	5.44%
Older than 45 years old	0.62%



4.6.12 Where did you first have regular access to a computer with internet connection?

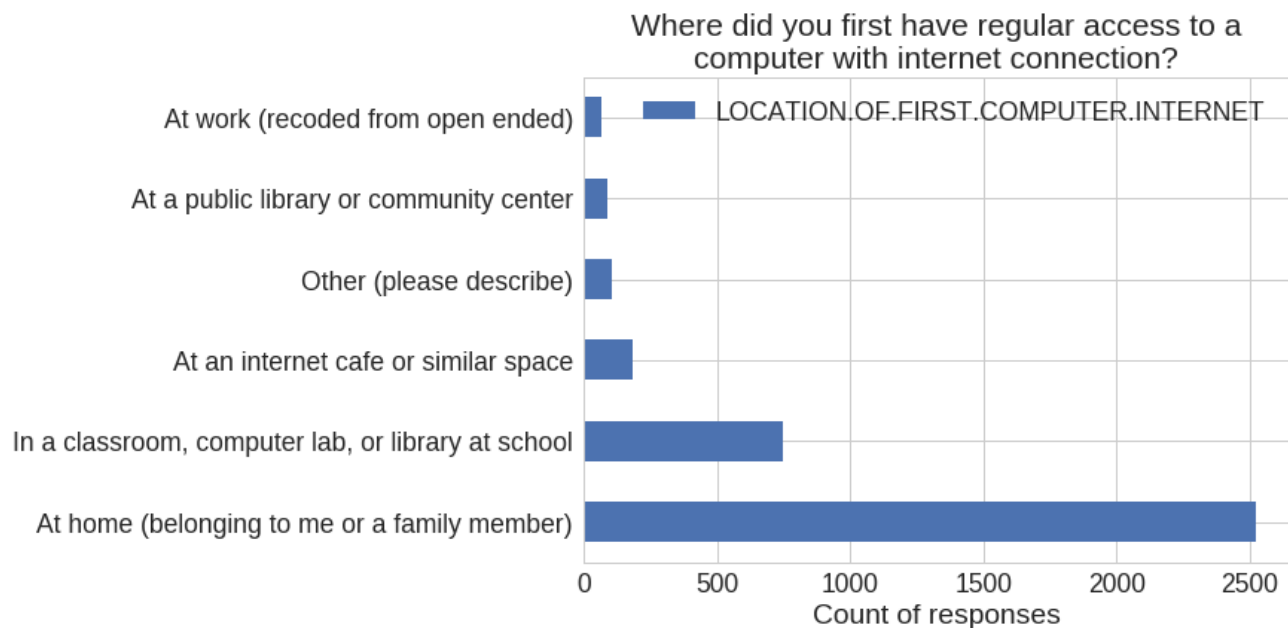
LOCATION.OF.FIRST.COMPUTER.INTERNET

Out [161] :

	count
At home (belonging to me or a family member)	2520
In a classroom, computer lab, or library at school	746
At an internet cafe or similar space	182
Other (please describe)	106
At a public library or community center	87
At work (recoded from open ended)	70

Out [162] :

	percent
At home (belonging to me or a family member)	67.91%
In a classroom, computer lab, or library at school	20.10%
At an internet cafe or similar space	4.90%
Other (please describe)	2.86%
At a public library or community center	2.34%
At work (recoded from open ended)	1.89%



4.6.13 Where was the respondent surveyed from?

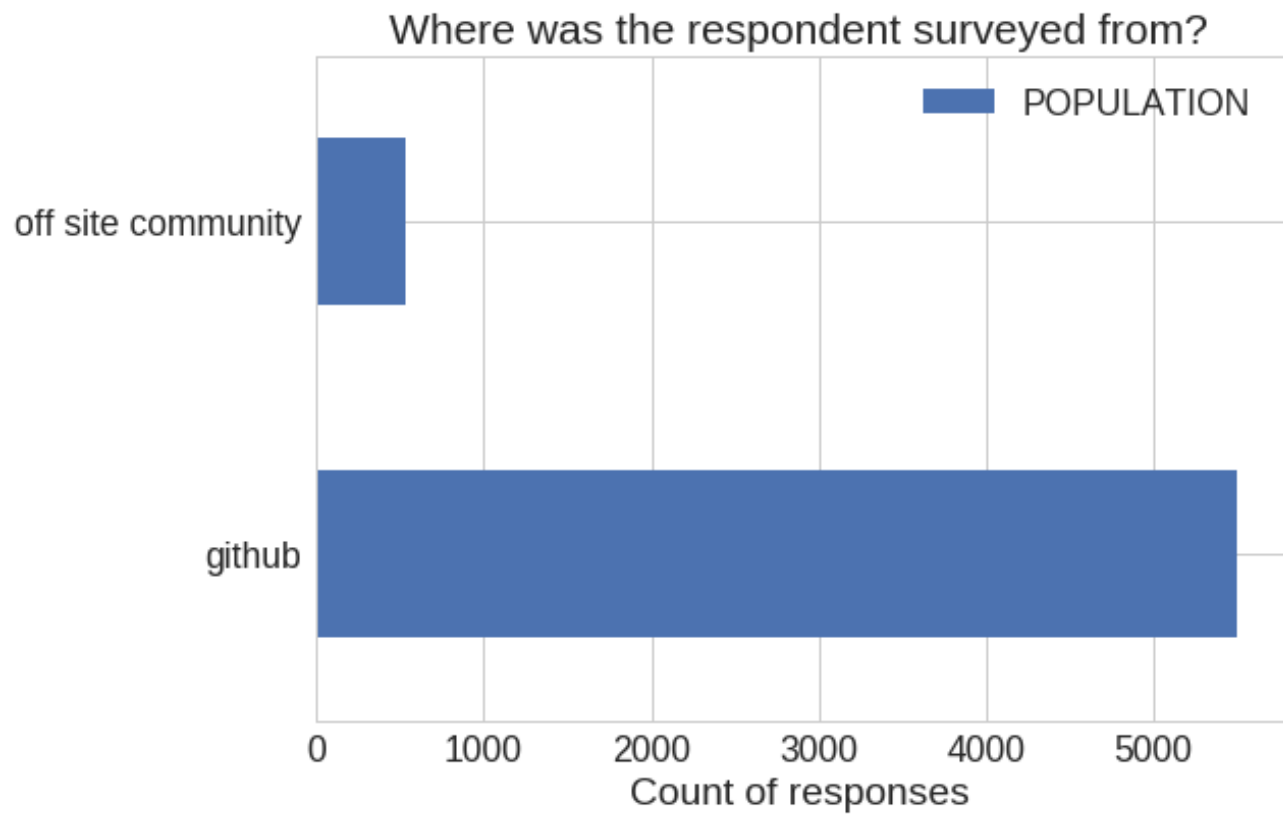
POPULATION

Out[164]:

count	
github	5495
off site community	534

Out[165]:

percent	
github	91.14%
off site community	8.86%



4.7 Harassment / Inclusiveness of OSS

4.7.1 Have you ever observed any of the following in the context of an open source project?

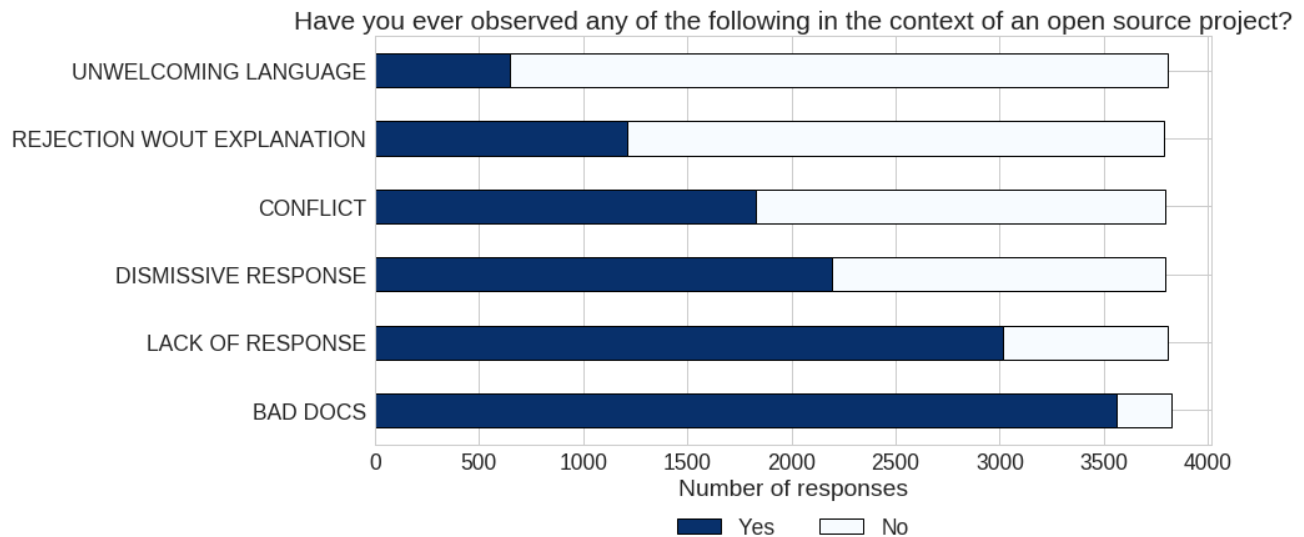
DISCOURAGING.BEHAVIOR.*

Out [167]:

	Yes	No
DISCOURAGING.BEHAVIOR.LACK.OF.RESPONSE	3017	792
DISCOURAGING.BEHAVIOR.REJECTION.WOUT.EXPLANATION	1210	2580
DISCOURAGING.BEHAVIOR.DISSMISSIVE.RESPONSE	2195	1598
DISCOURAGING.BEHAVIOR.BAD.DOCS	3559	263
DISCOURAGING.BEHAVIOR.CONFLICT	1830	1966
DISCOURAGING.BEHAVIOR.UNWELCOMING.LANGUAGE	649	3158

Out [168]:

	percent_yes
DISCOURAGING.BEHAVIOR.UNWELCOMING.LANGUAGE	17.05%
DISCOURAGING.BEHAVIOR.REJECTION.WOUT.EXPLANATION	31.93%
DISCOURAGING.BEHAVIOR.CONFLICT	48.21%
DISCOURAGING.BEHAVIOR.DISSMISSIVE.RESPONSE	57.87%
DISCOURAGING.BEHAVIOR.LACK.OF.RESPONSE	79.21%
DISCOURAGING.BEHAVIOR.BAD.DOCS	93.12%



4.7.2 Have you ever witnessed any of the following behaviors directed at another person in the context of an open source project? (not including something directed at you)

NEGATIVE.WITNESS.*

Out [172] :

	Yes	No
NEGATIVE.WITNESS.RUDENESS	1753	1911
NEGATIVE.WITNESS.NAME.CALLING	789	2875
NEGATIVE.WITNESS.THREATS	162	3502
NEGATIVE.WITNESS.IMPERSONATION	177	3487
NEGATIVE.WITNESS.SUSTAINED.HARASSMENT	237	3427
NEGATIVE.WITNESS.CROSS.PLATFORM.HARASSMENT	175	3489
NEGATIVE.WITNESS.STALKING	108	3556
NEGATIVE.WITNESS.SEXUAL.ADVANCES	136	3528
NEGATIVE.WITNESS.STEREOTYPING	423	3241
NEGATIVE.WITNESS.DOXXING	151	3513
NEGATIVE.WITNESS.OTHER	78	3586
NEGATIVE.WITNESS.NONE.OF.THE.ABOVE	1721	1943

Out [173] :

	percent_yes
NEGATIVE.WITNESS.OTHER	2.13%
NEGATIVE.WITNESS.STALKING	2.95%
NEGATIVE.WITNESS.SEXUAL.ADVANCES	3.71%
NEGATIVE.WITNESS.DOXXING	4.12%
NEGATIVE.WITNESS.THREATS	4.42%
NEGATIVE.WITNESS.CROSS.PLATFORM.HARASSMENT	4.78%
NEGATIVE.WITNESS.IMPERSONATION	4.83%
NEGATIVE.WITNESS.SUSTAINED.HARASSMENT	6.47%
NEGATIVE.WITNESS.STEREOTYPING	11.54%
NEGATIVE.WITNESS.NAME.CALLING	21.53%
NEGATIVE.WITNESS.NONE.OF.THE.ABOVE	46.97%
NEGATIVE.WITNESS.RUDENESS	47.84%



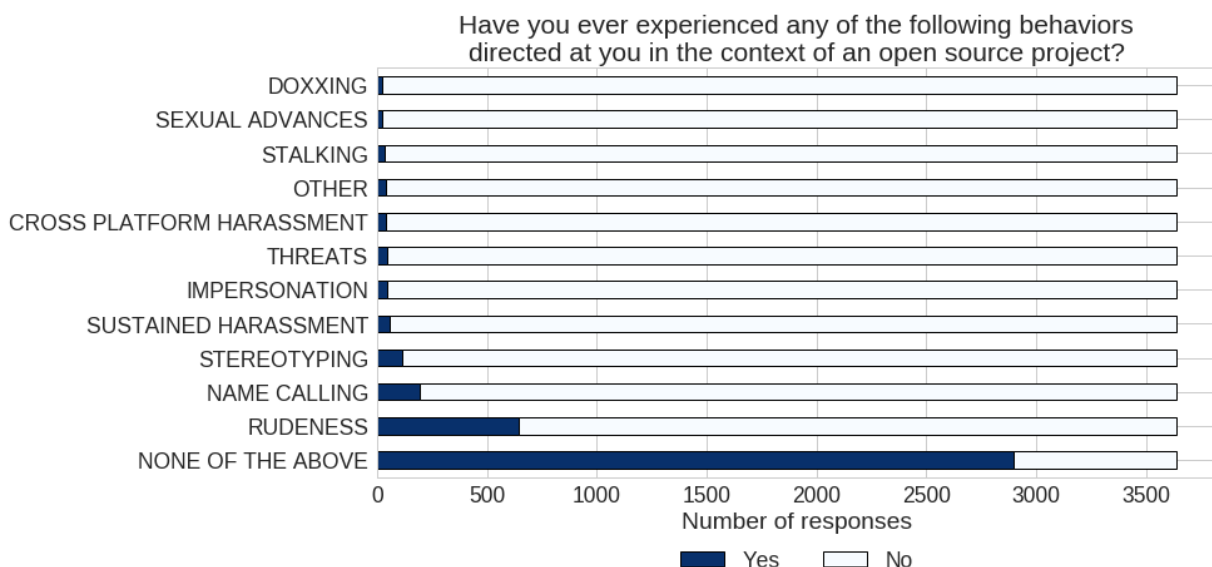
4.7.3 Have you ever experienced any of the following behaviors directed at you in the context of an open source project?

Out [177] :

	Yes	No
NEGATIVE.EXPERIENCE.RUDENESS	646	2992
NEGATIVE.EXPERIENCE.NAME.CALLING	192	3446
NEGATIVE.EXPERIENCE.THREATS	43	3595
NEGATIVE.EXPERIENCE.IMPERSONATION	45	3593
NEGATIVE.EXPERIENCE.SUSTAINED.HARASSMENT	55	3583
NEGATIVE.EXPERIENCE.CROSS.PLATFORM.HARASSMENT	42	3596
NEGATIVE.EXPERIENCE.STALKING	35	3603
NEGATIVE.EXPERIENCE.SEXUAL.ADVANCES	25	3613
NEGATIVE.EXPERIENCE.STEREOTYPING	114	3524
NEGATIVE.EXPERIENCE.DOXXING	23	3615
NEGATIVE.EXPERIENCE.OTHER	39	3599
NEGATIVE.EXPERIENCE.NONE.OF.THE.ABOVE	2900	738

Out [178] :

	percent_yes
NEGATIVE.EXPERIENCE.DOXXING	0.63%
NEGATIVE.EXPERIENCE.SEXUAL.ADVANCES	0.69%
NEGATIVE.EXPERIENCE.STALKING	0.96%
NEGATIVE.EXPERIENCE.OTHER	1.07%
NEGATIVE.EXPERIENCE.CROSS.PLATFORM.HARASSMENT	1.15%
NEGATIVE.EXPERIENCE.THREATS	1.18%
NEGATIVE.EXPERIENCE.IMPERSONATION	1.24%
NEGATIVE.EXPERIENCE.SUSTAINED.HARASSMENT	1.51%
NEGATIVE.EXPERIENCE.STEREOTYPING	3.13%
NEGATIVE.EXPERIENCE.NAME.CALLING	5.28%
NEGATIVE.EXPERIENCE.RUDENESS	17.76%
NEGATIVE.EXPERIENCE.NONE.OF.THE.ABOVE	79.71%



4.7.4 Thinking of the last time you experienced harassment, how did you respond?

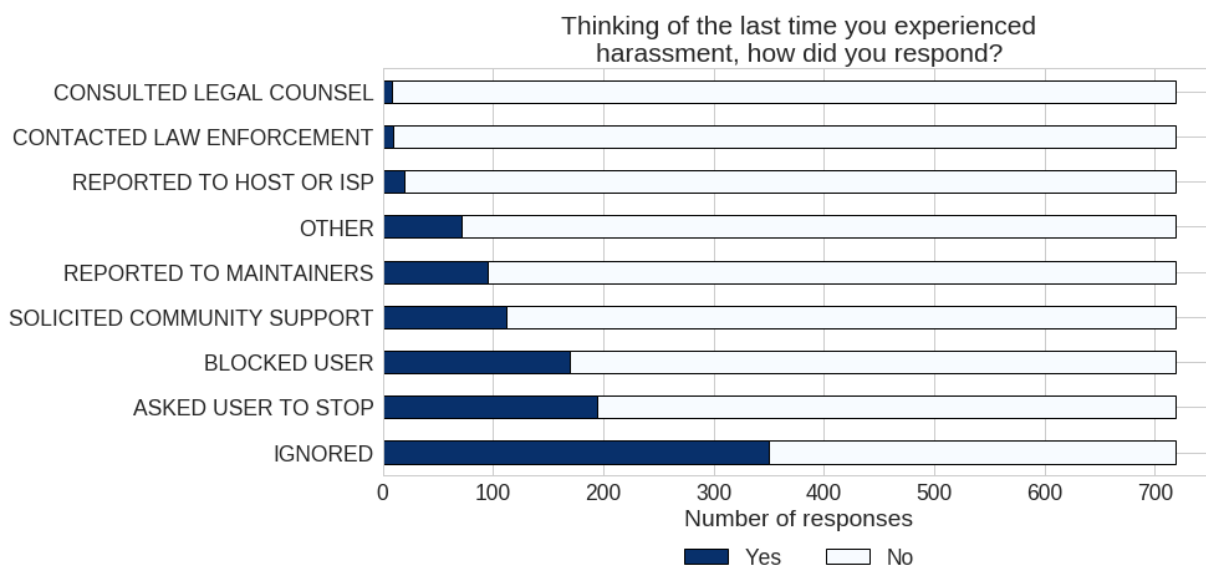
NEGATIVE.RESPONSE.*

Out [182] :

	Yes	No
NEGATIVE.RESPONSE.ASKED.USER.TO.STOP	194	525
NEGATIVE.RESPONSE.SOLICITED.COMMUNITY.SUPPORT	112	607
NEGATIVE.RESPONSE.BLOCKED.USER	170	549
NEGATIVE.RESPONSE.REPORTED.TO.MAINTAINERS	95	624
NEGATIVE.RESPONSE.REPORTED.TO.HOST.OR.ISP	20	699
NEGATIVE.RESPONSE.CONSULTED.LEGAL.COUNSEL	8	711
NEGATIVE.RESPONSE.CONTACTED.LAW.ENFORCEMENT	9	710
NEGATIVE.RESPONSE.OTHER	71	648
NEGATIVE.RESPONSE.IGNORED	350	369

Out [183] :

	percent_yes
NEGATIVE.RESPONSE.CONSULTED.LEGAL.COUNSEL	1.11%
NEGATIVE.RESPONSE.CONTACTED.LAW.ENFORCEMENT	1.25%
NEGATIVE.RESPONSE.REPORTED.TO.HOST.OR.ISP	2.78%
NEGATIVE.RESPONSE.OTHER	9.87%
NEGATIVE.RESPONSE.REPORTED.TO.MAINTAINERS	13.21%
NEGATIVE.RESPONSE.SOLICITED.COMMUNITY.SUPPORT	15.58%
NEGATIVE.RESPONSE.BLOCKED.USER	23.64%
NEGATIVE.RESPONSE.ASKED.USER.TO.STOP	26.98%
NEGATIVE.RESPONSE.IGNORED	48.68%

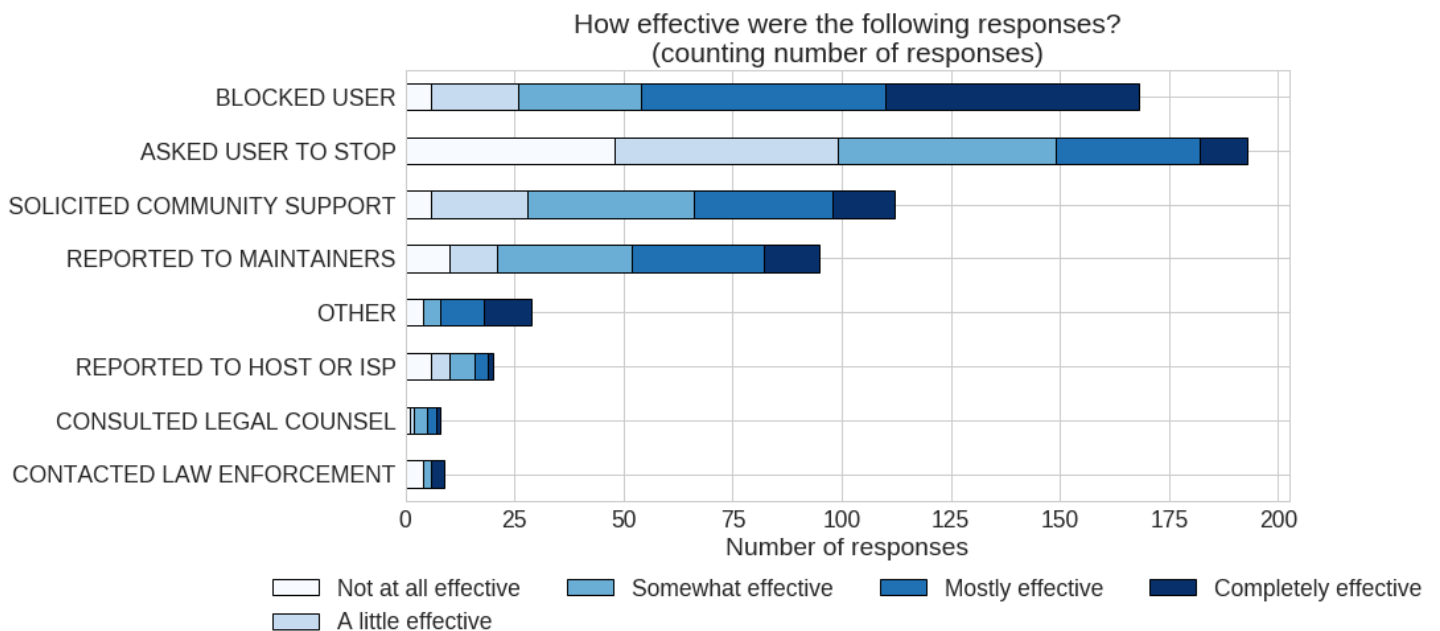


4.7.5 How effective were the following responses? Response counts

RESPONSE.EFFECTIVENESS.*

Out [186] :

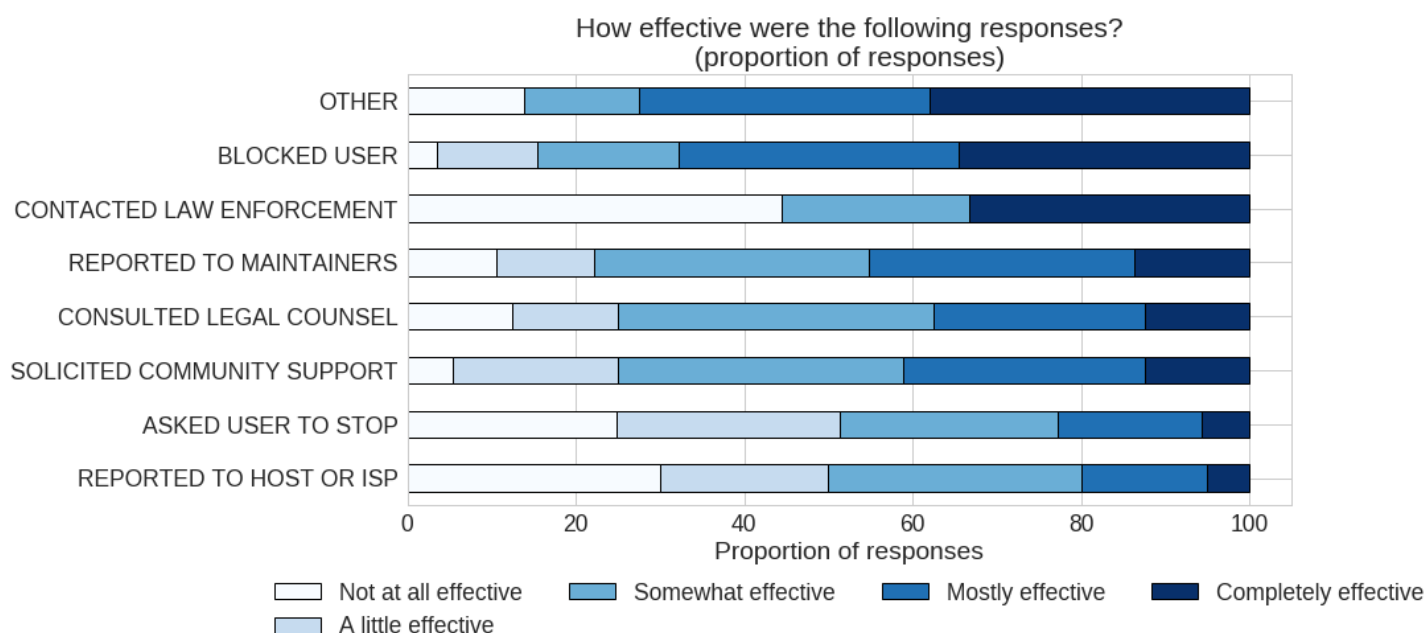
	Not at all effective	A little effective	Somewhat effective	Mostly effective	Completely effective
CONTACTED LAW ENFORCEMENT	4	0	2	0	3
CONSULTED LEGAL COUNSEL	1	1	3	2	1
REPORTED TO HOST OR ISP	6	4	6	3	1
OTHER	4	0	4	10	11
REPORTED TO MAINTAINERS	10	11	31	30	13
SOLICITED COMMUNITY SUPPORT	6	22	38	32	14
ASKED USER TO STOP	48	51	50	33	11
BLOCKED USER	6	20	28	56	58



4.7.6 How effective were the following responses? Proportions

Out [189] :

	Not at all effective	A little effective	Somewhat effective	Mostly effective	Completely effective
REPORTED TO HOST OR ISP	30.00%	20.00%	30.00%	15.00%	5.00%
ASKED USER TO STOP	24.87%	26.42%	25.91%	17.10%	5.70%
SOLICITED COMMUNITY SUPPORT	5.36%	19.64%	33.93%	28.57%	12.50%
CONSULTED LEGAL COUNSEL	12.50%	12.50%	37.50%	25.00%	12.50%
REPORTED TO MAINTAINERS	10.53%	11.58%	32.63%	31.58%	13.68%
CONTACTED LAW ENFORCEMENT	44.44%	0.00%	22.22%	0.00%	33.33%
BLOCKED USER	3.57%	11.90%	16.67%	33.33%	34.52%
OTHER	13.79%	0.00%	13.79%	34.48%	37.93%



4.7.7 As a result of experiencing or witnessing harassment, which, if any, of the following have you done?

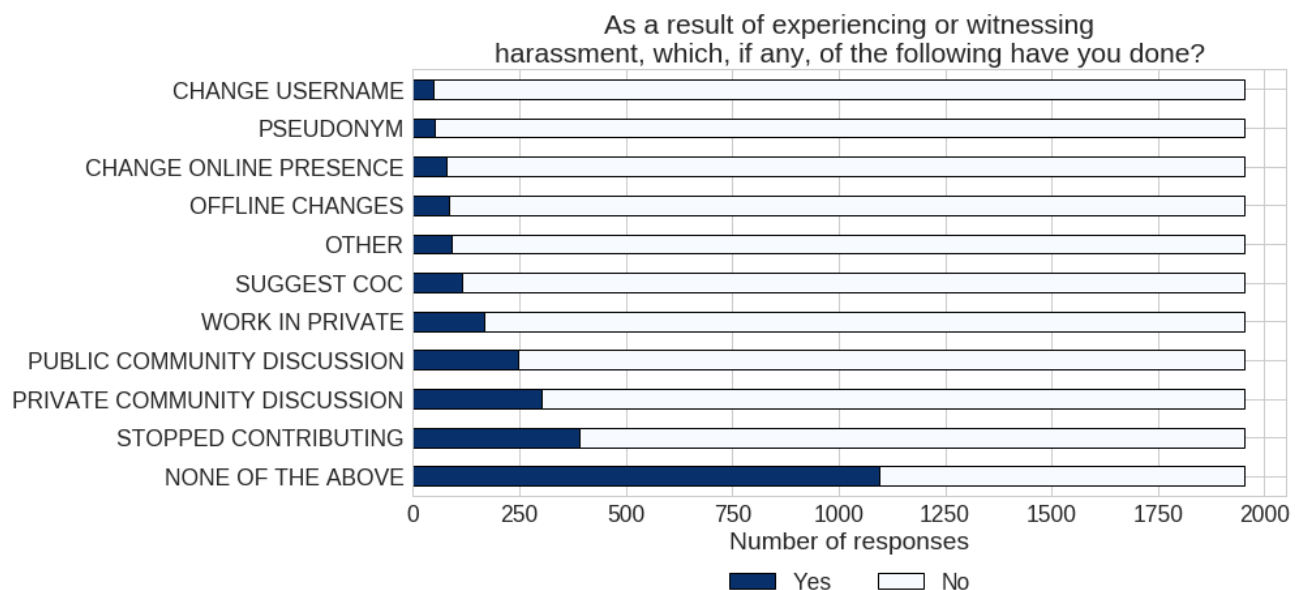
NEGATIVE.CONSEQUENCES.*

Out [193] :

	Yes	No
NEGATIVE.CONSEQUENCES.STOPPED.CONTRIBUTING	390	1563
NEGATIVE.CONSEQUENCES.PSEUDONYM	50	1903
NEGATIVE.CONSEQUENCES.WORK.IN.PRIVATE	166	1787
NEGATIVE.CONSEQUENCES.CHANGE.USERNAME	48	1905
NEGATIVE.CONSEQUENCES.CHANGE.ONLINE.PRESENCE	79	1874
NEGATIVE.CONSEQUENCES.SUGGEST.COC	116	1837
NEGATIVE.CONSEQUENCES.PRIVATE.COMMUNITY.DISCUSSION	301	1652
NEGATIVE.CONSEQUENCES.PUBLIC.COMMUNITY.DISCUSSION	248	1705
NEGATIVE.CONSEQUENCES.OFFLINE.CHANGES	85	1868
NEGATIVE.CONSEQUENCES.OTHER	90	1863
NEGATIVE.CONSEQUENCES.NONE.OF.THE.ABOVE	1094	859

Out [194] :

	percent_yes
NEGATIVE.CONSEQUENCES.CHANGE.USERNAME	2.46%
NEGATIVE.CONSEQUENCES.PSEUDONYM	2.56%
NEGATIVE.CONSEQUENCES.CHANGE.ONLINE.PRESENCE	4.05%
NEGATIVE.CONSEQUENCES.OFFLINE.CHANGES	4.35%
NEGATIVE.CONSEQUENCES.OTHER	4.61%
NEGATIVE.CONSEQUENCES.SUGGEST.COC	5.94%
NEGATIVE.CONSEQUENCES.WORK.IN.PRIVATE	8.50%
NEGATIVE.CONSEQUENCES.PUBLIC.COMMUNITY.DISCUSSION	12.70%
NEGATIVE.CONSEQUENCES.PRIVATE.COMMUNITY.DISCUSSION	15.41%
NEGATIVE.CONSEQUENCES.STOPPED.CONTRIBUTING	19.97%
NEGATIVE.CONSEQUENCES.NONE.OF.THE.ABOVE	56.02%
NEGATIVE.CONSEQUENCES.ANY.RESPONSE	100.00%



5 Bibliography

References

- R. Stuart Geiger. Summary analysis of the 2017 github open source survey. *SocArXiv Preprints*, 2017. doi: 10.17605/OSF.IO/ENRQ5. URL <https://osf.io/preprints/socarxiv/qps53>.
- J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science Engineering*, 9(3):90–95, May 2007. ISSN 1521-9615. doi: 10.1109/MCSE.2007.55. URL <http://ieeexplore.ieee.org/document/4160265/>.
- Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python, 2001. URL <http://www.scipy.org/>.
- Thomas Kluyver, Benjamin Ragan-Kelley, Fernando Pérez, Brian Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, Jessica Hamrick, Jason Grout, Sylvain Corlay, Paul Ivanov, Damián Avila, Safia Abdalla, Carol Willing, and Jupyter development team. Jupyter notebooks: a publishing format for reproducible computational workflows, 2016. URL <https://eprints.soton.ac.uk/403913/>.
- Wes McKinney. Data Structures for Statistical Computing in Python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 51–56, 2010. URL <http://conference.scipy.org/proceedings/scipy2010/mckinney.html>.
- Fernando Pérez and Brian E. Granger. IPython: a system for interactive scientific computing. *Computing in Science and Engineering*, 9(3):21–29, May 2007. ISSN 1521-9615. doi: 10.1109/MCSE.2007.53. URL <http://ipython.org>.
- S. van der Walt, S. C. Colbert, and G. Varoquaux. The numpy array: A structure for efficient numerical computation. *Computing in Science Engineering*, 13(2):22–30, March 2011. ISSN 1521-9615. doi: 10.1109/MCSE.2011.37. URL <https://arxiv.org/abs/1102.1523>.
- Guido van Rossum. Python library reference, 1995. URL <https://ir.cwi.nl/pub/5009/05009D.pdf>.
- Michael Waskom, Olga Botvinnik, Paul Hobson, John B. Cole, Yaroslav Halchenko, Stephan Hoyer, Alistair Miles, Tom Augspurger, Tal Yarkoni, Tobias Megies, Luis Pedro Coelho, Daniel Wehner, cynddl, Erik Ziegler, diego0020, Yury V. Zaytsev, Travis Hoppe, Skipper Seabold, Phillip Cloud, Miikka Koskinen, Kyle Meyer, Adel Qalieh, and Dan Allan. seaborn: v0.5.0 (november 2014), November 2014. URL <https://doi.org/10.5281/zenodo.12710>.
- Frannie Zlotnick, Arfon Smith, Mike Linksvayer, Anna Filippova, and The Open Source Initiative. The open source survey. *GitHub repository*, 2017a. URL <https://github.com/github/open-source-survey>.
- Frannie Zlotnick, Arfon Smith, Mike Linksvayer, Anna Filippova, and The Open Source Initiative. The open source survey website. *GitHub, Inc.*, 2017b. URL <https://opensourcesurvey.org/2017/>.