An Empirical Comparison of Ethnic and Gender Diversity of DevOps and non-DevOps Contributions to Open-Source Projects

Online Appendix

Weeraddana et al.

The Table of Content

Appendix A: Results for various diversity indices.

Table 1. Results of statistical analysis of ethnic diversity metrics for DevOps and Non-DevOps contributors

Diversity Metric	**DevOps (median)**	**Non-DevOps (median)**	**Wilcoxon, one-tailed, paired Effect size ($\alpha = 0.0023$)**	**Effect size (Cliff's $\lvert\delta\rvert$)**
Richness	2	3	$2.8 \times 10^{-47} \ll \alpha$ (less than test)	large
Evenness	0.3	0.5	$2.4 \times 10^{-31} \ll \alpha$ (less than test)	medium
Blau Score (a.k.a. Simpson/Diversity index)	0.22	0.25	$6.9 \times 10^{-7} \ll \alpha$ (less than test)	negligible
Prevalence Rankings and Diffusion Score	9.1	10.1	$0.006 > \alpha$ (less than test)	-

Table 2. Results of statistical analysis of gender diversity metrics for DevOps and Non-DevOps contributors

Diversity Metric	**DevOps (median)**	**Non-DevOps (median)**	**Wilcoxon, one-tailed, paired Effect size ($\alpha = 0.0023$)**	**Effect size (Cliff's $\lvert\delta\rvert$)**
Richness	2	2	$1.4 \times 10^{-25} \ll \alpha$ (less than test)	small
Evenness	0.2	0.2	$7.8 \times 10^{-20} \ll \alpha$ (less than test)	small
Blau Score (a.k.a. Simpson/Diversity index)	0.12	0.13	$1.6 \times 10^{-6} \ll \alpha$ (less than test)	small
Prevalence Rankings and Diffusion Score	6.25	7.18	$6.4 \times 10^{-5} \ll \alpha$ (less than test)	small

Appendix B: Results when all unknown genders are considered female.

The percentage of names identified as unknown genders by the gender-inferring tools is 15.2% (2,542) among DevOps contributors and 21.5% (20,130) among non-DevOps contributors. Following Vasilescu et al.'s approach [1], we examine what would happen if all unknown genders are perceived as women to evaluate whether biases persist even if the female representation is increased to an upper bound. Below, we report the results. Accordingly, we find that our results still holds in this upper bound scenario. For example, we find that, during the last ten years, the percentage of DevOps contributors who are perceptible as females ranges between 6%–16% in this upper-bound scenario.

(RQ1) Does the ethnic and gender diversity of DevOps contributors differ from ethnic and gender diversity of non-DevOps contributors?

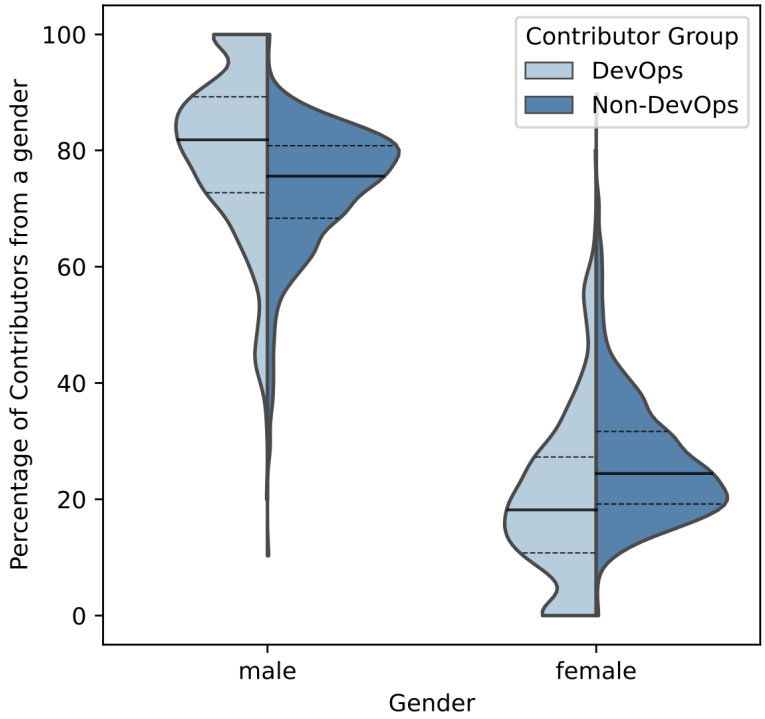


Fig. 1: Bean plots showing the distribution of percentages of DevOps and non- DevOps contributors from four perceptible genders (i.e., male and female). The solid lines represent the median percentages, and the dotted lines represent the first and third quantiles.

Table 1. Results of statistical analysis of gender proportions for DevOps and Non-DevOps contributors

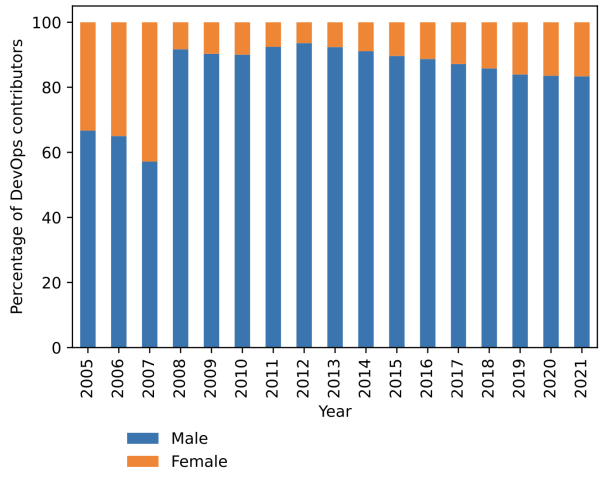
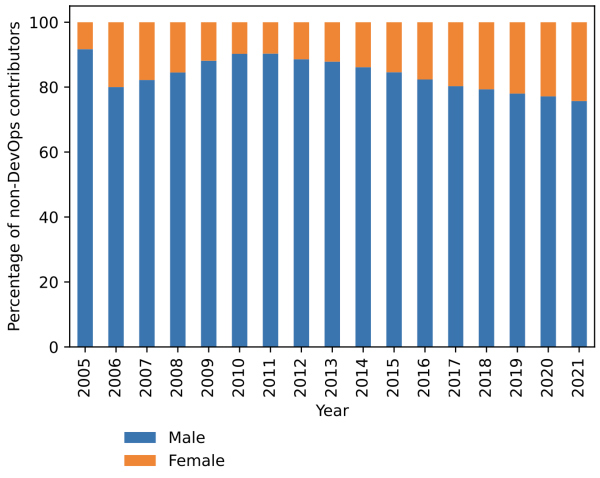
Gender	**DevOps (median)**	**Non-DevOps (median)**	**Wilcoxon, one-tailed, paired (α = 0.0023)**	**Effect size (Cliff's |δ |)**
Male	81.8%	75.6%	3.7×10^{-31} << α (greater than test)	medium
Female	18.2%	24.4%	3.7×10^{-31} << α (less than test)	medium

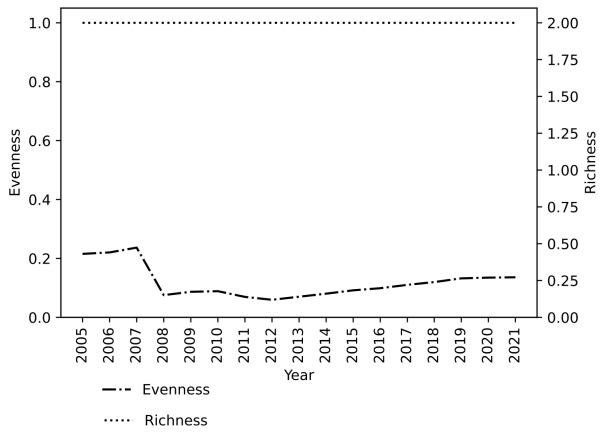
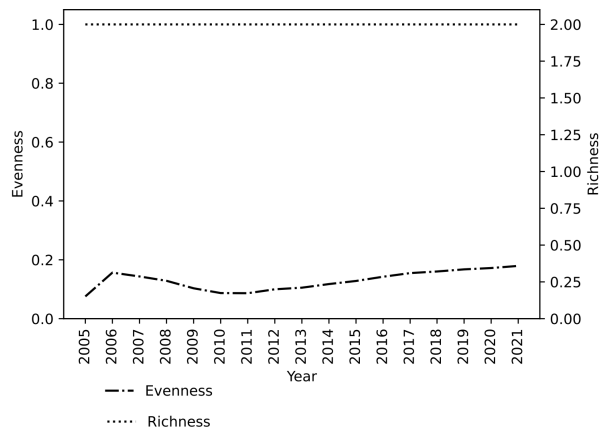
Table 2. Results of statistical analysis of gender diversity metrics for DevOps and Non-DevOps contributors

Diversity Metric	**DevOps (median)**	**Non-DevOps (median)**	**Wilcoxon, one-tailed, paired Effect size ($\alpha = 0.0023$)**	**Effect size (Cliff's $\lvert\delta\rvert$)**
Richness	2	2	$4.0 \times 10^{-13} \ll \alpha$ (less than test)	negligible
Evenness	0.41	0.53	$2.4 \times 10^{-57} \ll \alpha$ (less than test)	large
Blau (Simpson)	0.30	0.37	$1.2 \times 10^{-34} \ll \alpha$ (less than test)	medium
Prevalence Rankings and Diffusion Score	18.18	24.40	$7.1 \times 10^{-31} \ll \alpha$ (less than test)	medium

(RQ2) How does the distribution of ethnic and gender diversity change as projects age?

Table 3. (RQ2-b) Change in Perceptible Gender Diversity Over Time

DevOps Contributors	Non-DevOps Contributors
	
Percentages of perceptible genders of DevOps contributors over time.	Percentages of perceptible genders of non-DevOps contributors over time

	
Values of diversity metrics of DevOps developers over time: the left y-axis presents the evenness values, while the right y-axis presents the richness values.	Values of diversity metrics of non-DevOps developers over time: the left y-axis presents the evenness values, while the right y-axis presents the richness values.
Average growth of richness of DevOps contributors over time: 0.0 Average growth of evenness of DevOps contributors over time: 0.008487073120975443	Average growth of richness of non-DevOps contributors over time: 0.0 Average growth of richness of non-DevOps contributors over time: 0.00882762907177941

Appendix C: Results when the threshold number of DevOps file changes in the definition of a “DevOps contributor” is two.

We experiment with different thresholds of contribution counts to DevOps files. In this appendix, we report the results when the threshold number of DevOps file changes in the definition of a “DevOps contributor” is two. Below, we report the results. Accordingly, we do not observe substantial differences in the results between the threshold being two and threshold being one (as in our study).

(PA) Preliminary Analysis: Identifying DevOps and non-DevOps Contributors

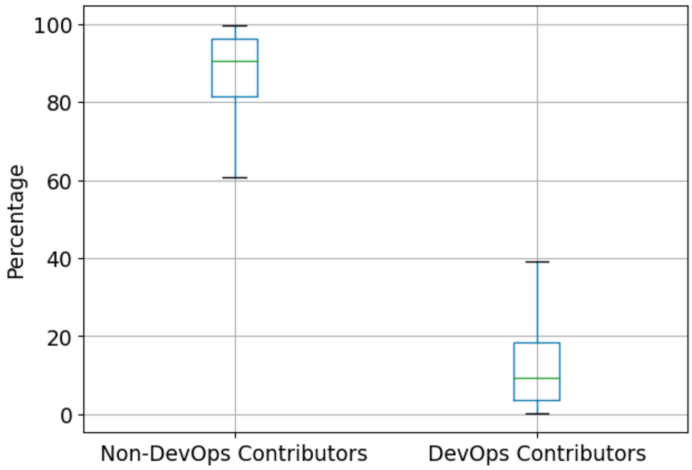


Fig. 1: Box-plots showing the percentages of contributors to non-DevOps files only and the percentages of contributors to DevOps files in the studied projects.

Median Non-DevOps Contributors per project: 90.605837%
Median DevOps Contributors per project: 9.394163%

(RQ1) Does the ethnic and gender diversity of DevOps contributors differ from ethnic and gender diversity of non-DevOps contributors?

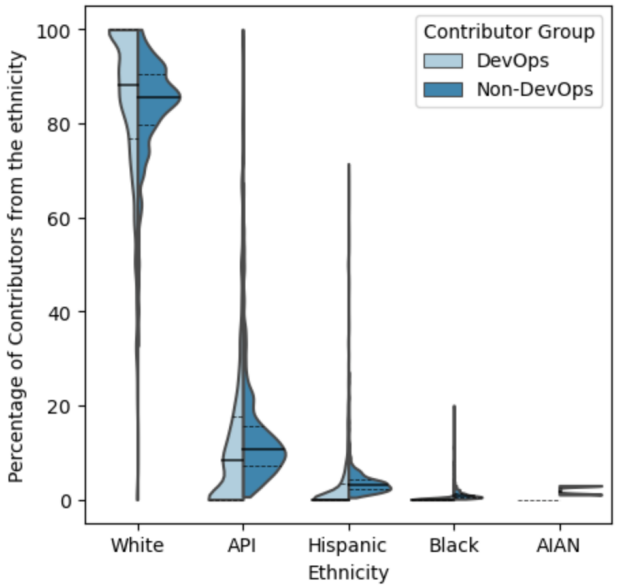


Fig. 2: Bean plots showing the distribution of percentages of DevOps and non-DevOps contributors from four perceptible ethnicities.

Table 1. Results of statistical analysis of ethnic proportions for DevOps and Non-DevOps contributors

Ethnicity	**DevOps (median)**	**Non-DevOps (median)**	**Wilcoxon, one-tailed, paired ($\alpha = 0.0023$)**	**Effect size (Cliff's $\lvert\delta\rvert$)**
White	88.2%	85.4%	$0.0006 < \alpha$ (greater than test)	small
API	8.3%	10.7%	$8.3 \times 10^{-5} << \alpha$ (less than test)	small
Hispanic	0.0%	3.1%	$1.5 \times 10^{-12} << \alpha$ (less than test)	large
Black	0.0%	0.7%	$1.9 \times 10^{-22} << \alpha$ (less than test)	large
AIAN	0.0%	2.0%	$0.25 > \alpha$ (less than test)	-

Table 2. Results of statistical analysis of ethnic diversity metrics for DevOps and Non-DevOps contributors

Diversity Metric	**DevOps (median)**	**Non-DevOps (median)**	**Wilcoxon, one-tailed, paired Effect size ($\alpha = 0.0023$)**	**Effect size (Cliff's $\lvert\delta\rvert$)**
Richness	2	3	$9.4 \times 10^{-59} << \alpha$ (less than test)	large
Evenness	0.3	0.5	$8.5 \times 10^{-48} << \alpha$ (less than test)	medium
Blau (Simpson)	0.2	0.2	$8.2 \times 10^{-12} << \alpha$ (less than test)	small
Prevalence Rankings and Diffusion Score	7.8	10.5	$0.0001 < \alpha$ (less than test)	small

Table 3. Results of statistical analysis of gender proportions for DevOps and Non-DevOps contributors

Gender	**DevOps (median)**	**Non-DevOps (median)**	**Wilcoxon, one-tailed, paired ($\alpha = 0.0023$)**	**Effect size (Cliff's $\lvert\delta\rvert$)**
Male	94.4%	92.8%	$1.0 \times 10^{-6} << \alpha$ (greater than test)	small
Female	5.5%	7.2%	$1.0 \times 10^{-6} << \alpha$ (less than test)	small

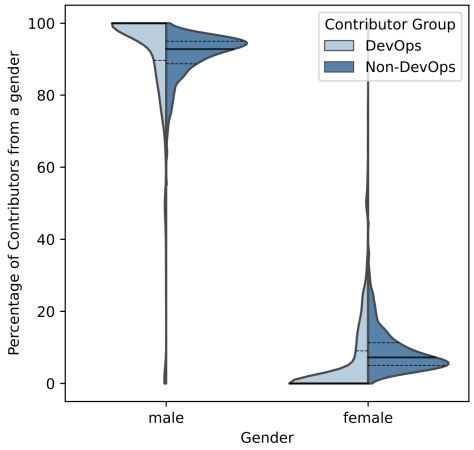


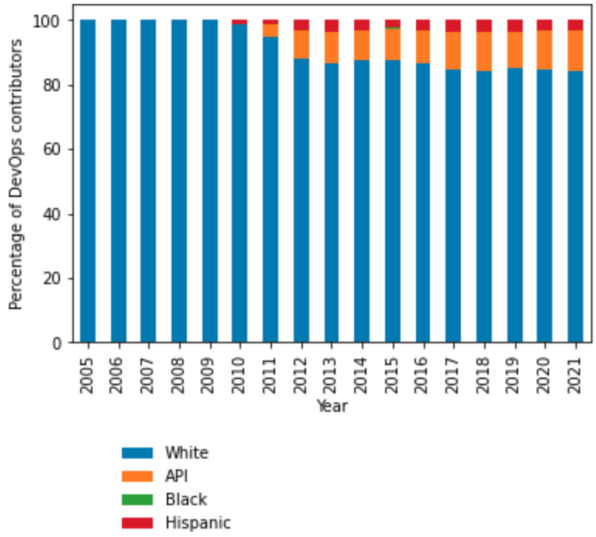
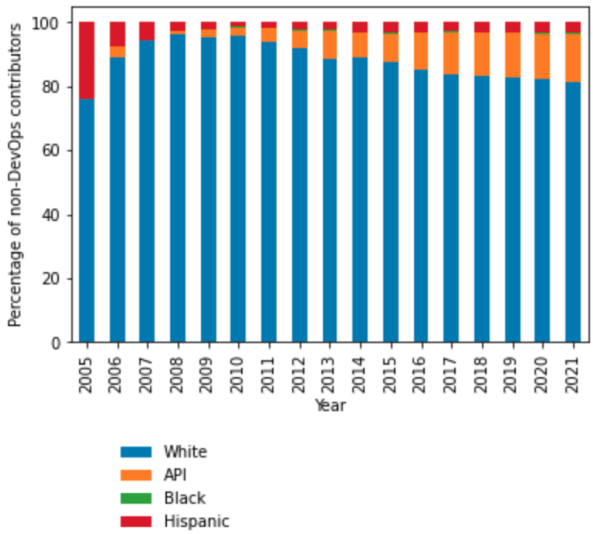
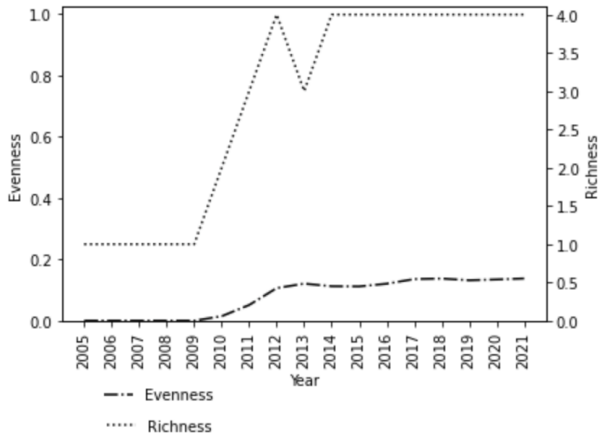
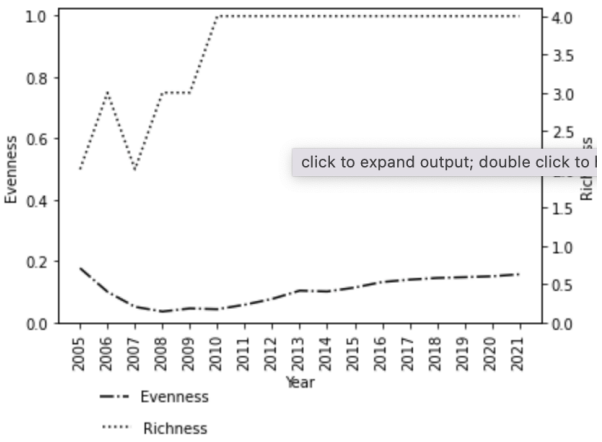
Fig. 3: Bean plots showing the distribution of percentages of DevOps and non- DevOps contributors from four perceptible genders.

Table 4. Results of statistical analysis of gender diversity metrics for DevOps and Non-DevOps contributors

Diversity Metric	**DevOps (median)**	**Non-DevOps (median)**	**Wilcoxon, one-tailed, paired Effect size ($\alpha = 0.0023$)**	**Effect size (Cliff's $\lvert\delta\rvert$)**
Richness	2	2	$8.2 \times 10^{-40} << \alpha$ (less than test)	medium
Evenness	0.17	0.24	$8.3 \times 10^{-28} << \alpha$ (less than test)	small
Blau (Simpson)	0.10	0.13	$6.7 \times 10^{-9} << \alpha$ (less than test)	small
Prevalence Rankings and Diffusion Score	5.5	7.2	$1.0 \times 10^{-6} << \alpha$ (less than test)	small

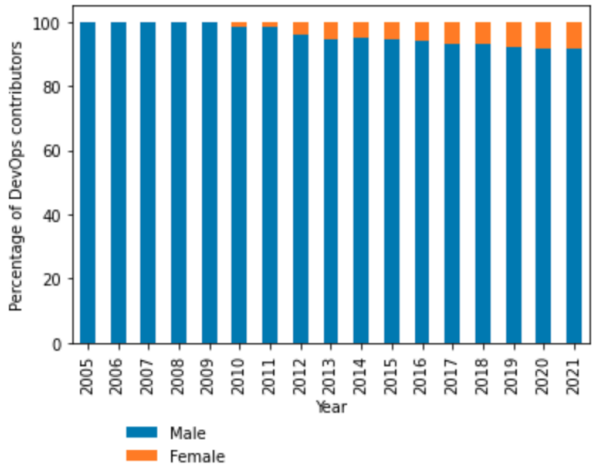
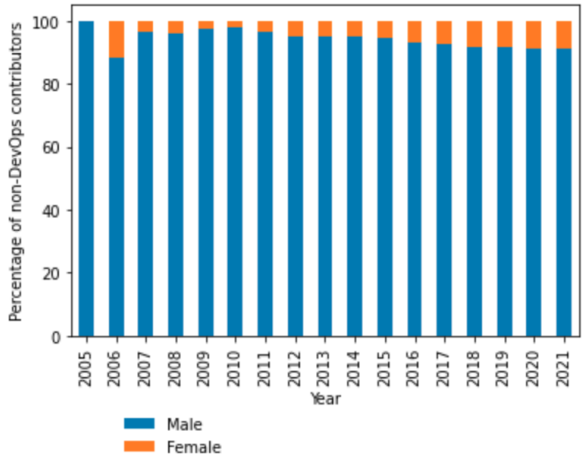
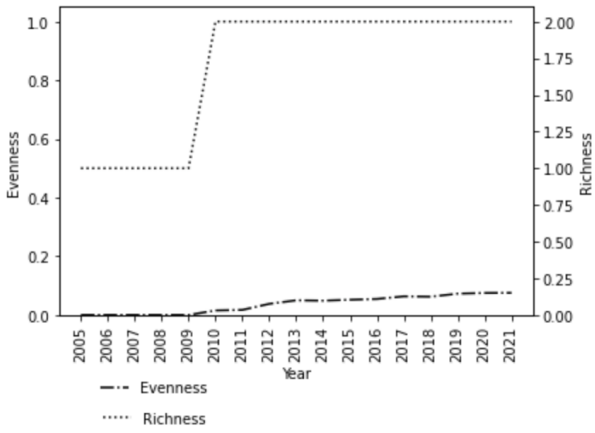
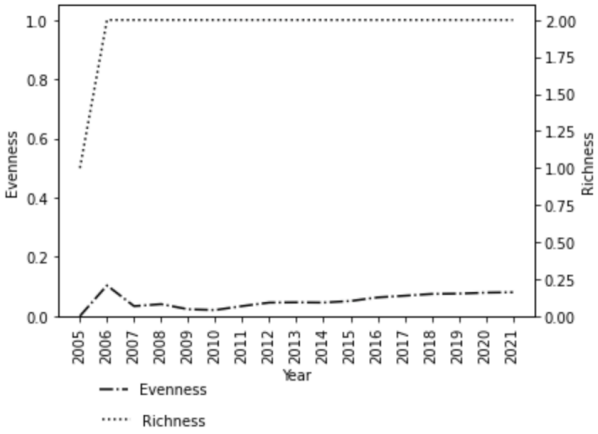
(RQ2) How does the distribution of ethnic and gender diversity change as projects age?

Table 5. (RQ2-a) Change in Perceptible Ethnic Diversity Over Time

DevOps Contributors	Non-DevOps Contributors
	
Percentages of perceptible ethnicities of DevOps contributors over time.	Percentages of perceptible ethnicities of non-DevOps contributors over time
	
Values of diversity metrics of DevOps developers over time: the left y-axis presents the evenness values, while the right y-axis presents the richness values.	Values of diversity metrics of non-DevOps developers over time: the left y-axis presents the evenness values, while the right y-axis presents the richness values.
Average growth of richness of DevOps contributors over time: 0.0 Average growth of evenness of DevOps	Average growth of richness of non-DevOps contributors over time: 0.0 Average growth of richness of non-DevOps

contributors over time: 0.0034762645148051653	contributors over time: 0.0089056524567327

Table 6. (RQ2-b) Change in Perceptible Gender Diversity Over Time

DevOps Contributors	Non-DevOps Contributors
	
Percentages of perceptible genders of DevOps contributors over time.	Percentages of perceptible genders of non-DevOps contributors over time
	
Values of diversity metrics of DevOps developers over time: the left y-axis presents the evenness values, while the right y-axis presents the richness values.	Values of diversity metrics of non-DevOps developers over time: the left y-axis presents the evenness values, while the right y-axis presents the richness values.
Average growth of richness of DevOps contributors over time: 0.0 Average growth of evenness of DevOps contributors over time: 0.004167960760444908	Average growth of richness of non-DevOps contributors over time: 0.0 Average growth of richness of non-DevOps contributors over time: 0.003874098372568772

Appendix D: Results when the threshold number of DevOps file changes in the definition of “DevOps contributor” is ten.

We experiment with different thresholds of contribution counts to DevOps files. In this appendix, we report the results when the threshold number of DevOps file changes in the definition of a “DevOps contributor” is ten. Below, we report the results. Accordingly, we do not observe substantial differences in the results between the threshold being ten and threshold being one (as in our study).

(PA) Preliminary Analysis: Identifying DevOps and non-DevOps Contributors

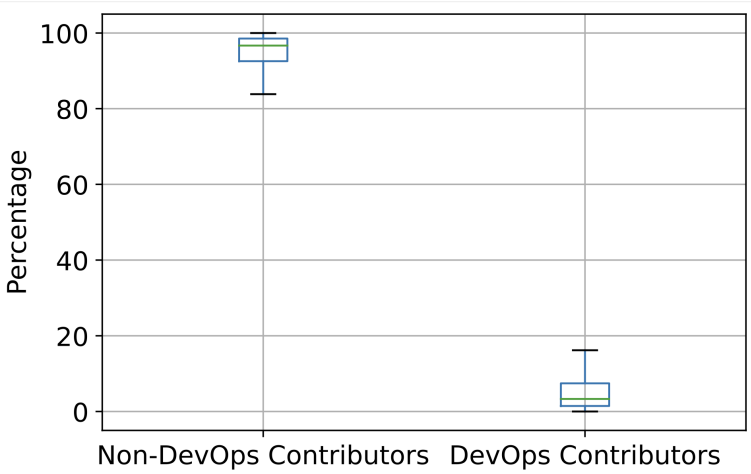


Fig. 1: Box-plots showing the percentages of contributors to non-DevOps files only and the percentages of contributors to DevOps files in the studied projects.

Median Non-DevOps Contributors per project: 96.684982%
Median DevOps Contributors per project: 3.315018%

(RQ1) Does the ethnic and gender diversity of DevOps contributors differ from ethnic and gender diversity of non-DevOps contributors?

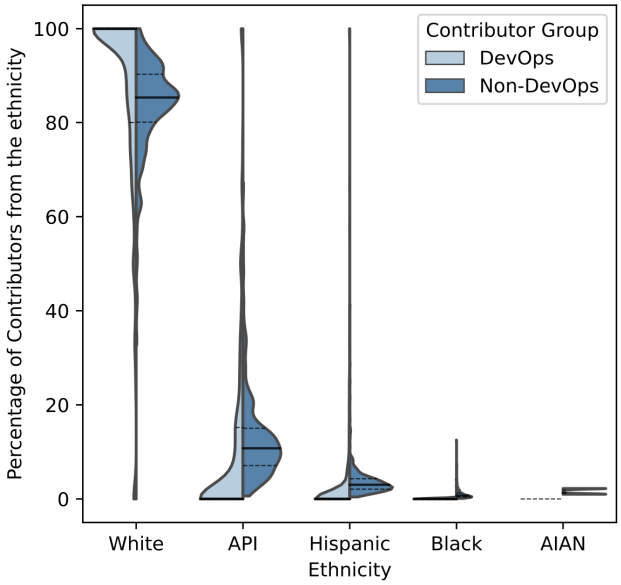


Fig. 2: Bean plots showing the distribution of percentages of DevOps and non-DevOps contributors from four perceptible ethnicities.

Table 1. Results of statistical analysis of ethnic proportions for DevOps and Non-DevOps contributors

Ethnicity	**DevOps (median)**	**Non-DevOps (median)**	**Wilcoxon, one-tailed, paired ($\alpha = 0.0023$)**	**Effect size (Cliff's $\lvert\delta\rvert$)**
White	100.0%	85.3%	$3.0 \times 10^{-11} << \alpha$ (greater than test)	median
API	0.0%	10.8%	$2.6 \times 10^{-11} << \alpha$ (less than test)	median
Hispanic	0.0%	3.0%	$2.6 \times 10^{-27} << \alpha$ (less than test)	large
Black	0.0%	0.7%	$6.2 \times 10^{-30} << \alpha$ (less than test)	large
AIAN	0.0%	1.6%	$0.25 > \alpha$ (less than test)	-

Table 2. Results of statistical analysis of ethnic diversity metrics for DevOps and Non-DevOps contributors

Diversity Metric	**DevOps (median)**	**Non-DevOps (median)**	**Wilcoxon, one-tailed, paired Effect size ($\alpha = 0.0023$)**	**Effect size (Cliff's $\lvert\delta\rvert$)**
Richness	1	3	$5.6 \times 10^{-69} << \alpha$ (less than test)	large
Evenness	0.0	0.5	$3.2 \times 10^{-63} << \alpha$ (less than test)	large
Blau (Simpson)	0.0	0.2	$1.4 \times 10^{-31} << \alpha$ (less than test)	medium
Prevalence Rankings and Diffusion Score	0.0	10.5	$8.5 \times 10^{-13} << \alpha$ (less than test)	medium

Table 3. Results of statistical analysis of gender proportions for DevOps and Non-DevOps contributors

Gender	**DevOps (median)**	**Non-DevOps (median)**	**Wilcoxon, one-tailed, paired ($\alpha = 0.0023$)**	**Effect size (Cliff's $\lvert\delta\rvert$)**
Male	100.0%	92.8%	$2.5 \times 10^{-14} << \alpha$ (greater than test)	medium

Female	0.0%	7.2%	$5.7 \times 10^{-19} << \alpha$ (less than test)	large

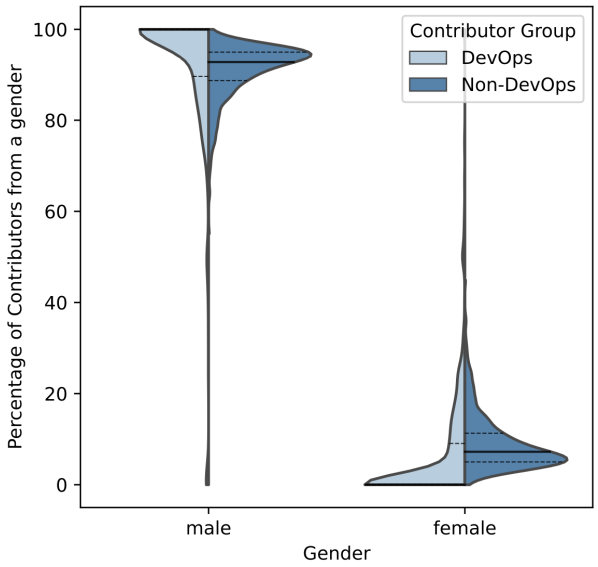


Fig. 3: Bean plots showing the distribution of percentages of DevOps and non- DevOps contributors from four perceptible genders.

Table 4. Results of statistical analysis of gender diversity metrics for DevOps and Non-DevOps contributors

Diversity Metric	**DevOps (median)**	**Non-DevOps (median)**	**Wilcoxon, one-tailed, paired Effect size ($\alpha = 0.0023$)**	**Effect size (Cliff's |δ |)**
Richness	1	2	$4.6 \times 10^{-63} << \alpha$ (less than test)	large
Evenness	0.0	0.2	$4.4 \times 10^{-52} << \alpha$ (less than test)	large
Blau (Simpson)	0.0	0.1	$1.4 \times 10^{-24} << \alpha$ (less than test)	medium
Prevalence Rankings and Diffusion Score	0.0	7.2	$7.0 \times 10^{-20} << \alpha$ (less than test)	large

(RQ2) How does the distribution of ethnic and gender diversity change as projects age?

Table 5. (RQ2-a) Change in Perceptible Ethnic Diversity Over Time

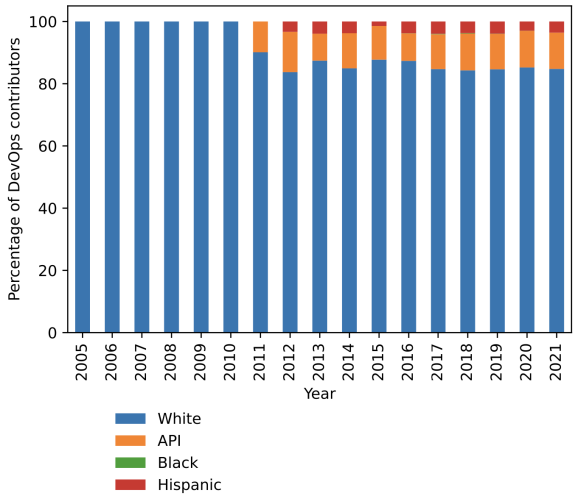
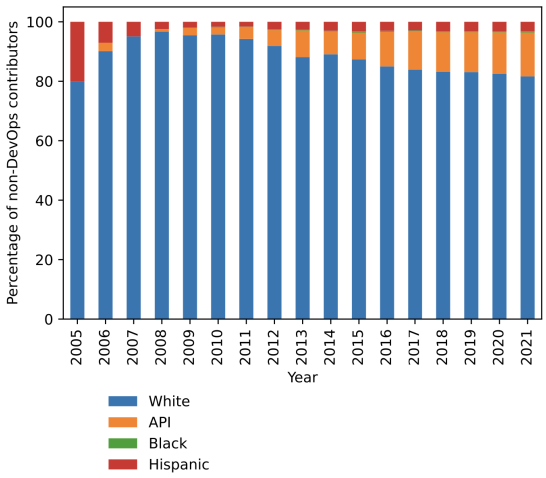
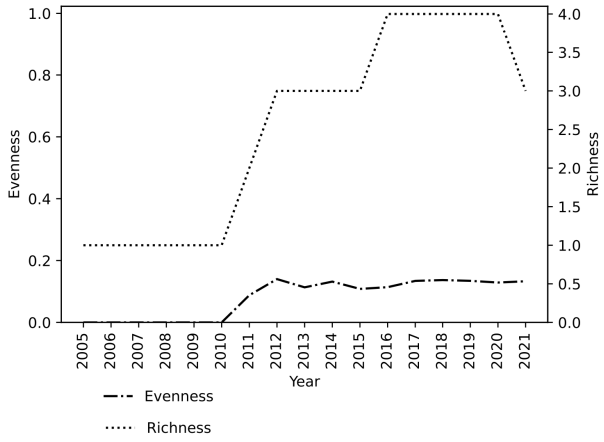
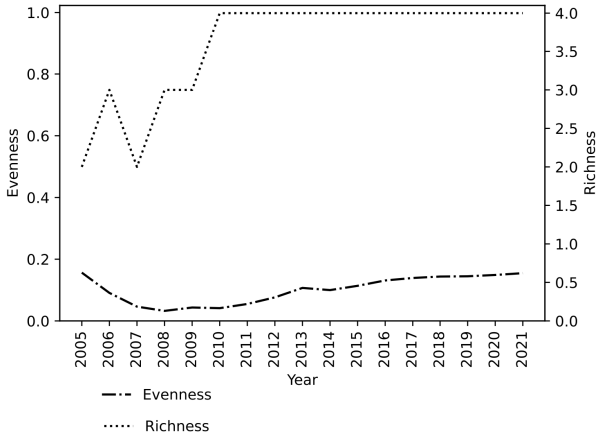
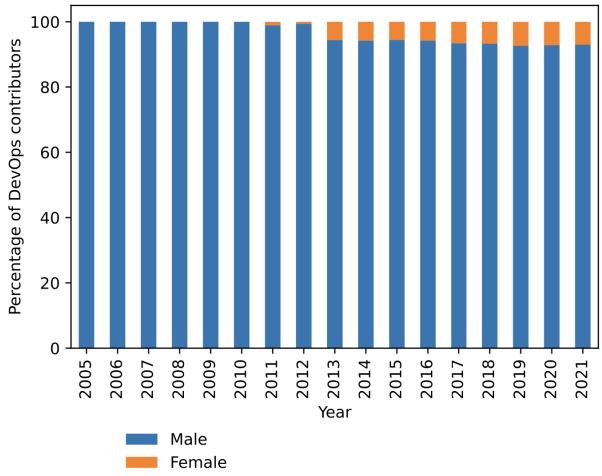
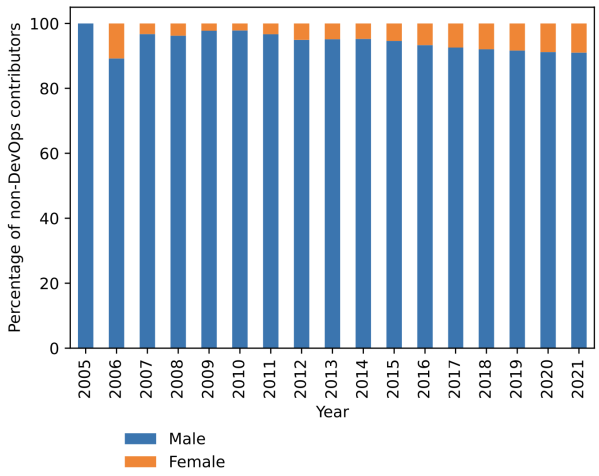
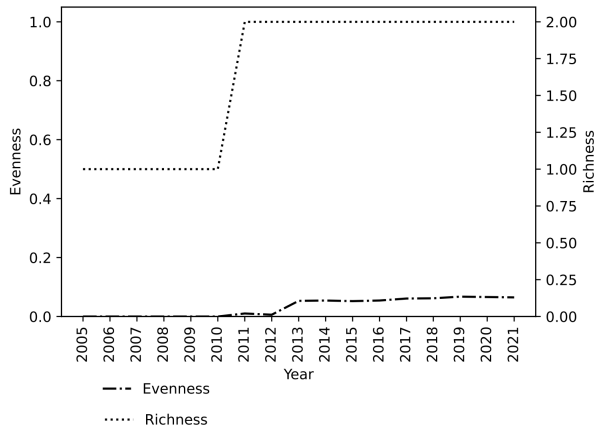
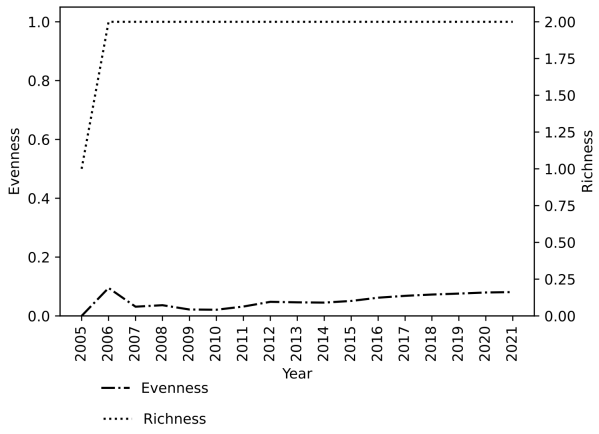
DevOps Contributors	Non-DevOps Contributors
	
Percentages of perceptible ethnicities of DevOps contributors over time.	Percentages of perceptible ethnicities of non-DevOps contributors over time
	
Values of diversity metrics of DevOps developers over time: the left y-axis presents the evenness values, while the right y-axis presents the richness values.	Values of diversity metrics of non-DevOps developers over time: the left y-axis presents the evenness values, while the right y-axis presents the richness values.
Average growth of richness of DevOps contributors over time: 0.0 Average growth of evenness of DevOps contributors over time: -0.0007777736935033906	Average growth of richness of non-DevOps contributors over time: 0.0 Average growth of richness of non-DevOps contributors over time: 0.008710427354668933

Table 6. (RQ2-b) Change in Perceptible Gender Diversity Over Time

DevOps Contributors	Non-DevOps Contributors
	
Percentages of perceptible genders of DevOps contributors over time.	Percentages of perceptible genders of non-DevOps contributors over time
	
Values of diversity metrics of DevOps developers over time: the left y-axis presents the evenness values, while the right y-axis presents the richness values.	Values of diversity metrics of non-DevOps developers over time: the left y-axis presents the evenness values, while the right y-axis presents the richness values.
Average growth of richness of DevOps contributors over time: 0.0 Average growth of evenness of DevOps contributors over time: 0.006530533220542128	Average growth of richness of non-DevOps contributors over time: 0.0 Average growth of richness of non-DevOps contributors over time: 0.0036803380321606283

Appendix E: Results when developers perceived as from MENA countries are considered as a separate ethnic group.

We re-compute the results by reclassifying individuals from the Middle-East and Northern Africa (MENA) region who are perceived as White from the White category to the non-White category. This allows us to consider this group of people as a separate ethnic group rather than as White. We modify the Threats to Validity section in the manuscript to include the reviewer's concern and our new analyses. Below, we describe the approach and results of our new analysis.

Approach.
We use the nationality-inferring API provided by the NAME-PRISM tool to identify perceptibly White contributors who are also perceptible as from MENA countries. NAME-PRISM's nationality-inferring API is based on a fine-grained 39-nationality taxonomy. For contributors who have the perceptible ethnicity as White, we check whether the perceptible nationality of that contributor is corresponding to a MENA country. If so, we label the ethnicity of the contributor as "MENA" instead of "White." Note that we consider such labelling only if NAME-PRISM's confidence rate is higher than 0.8.

Results. We re-compute the results of RQ1 and RQ2. With respect to RQ1, Figure 1 shows the proportions of different ethnicities among DevOps and non-DevOps contributors. We find our results still hold even after considering developers perceived as from MENA countries explicitly. For example, from the figure, we observe that perceptibly White contributors are the dominating ethnicity among the rest. With respect to diversity indices, we find that the richness and evenness among DevOps contributors is significantly less than (Wilcoxon, $p << \alpha = 0.0023$, one-tailed, paired) that among non-DevOps contributors with large and medium effect sizes, respectively.

(RQ1) Does the ethnic and gender diversity of DevOps contributors differ from ethnic and gender diversity of non-DevOps contributors?

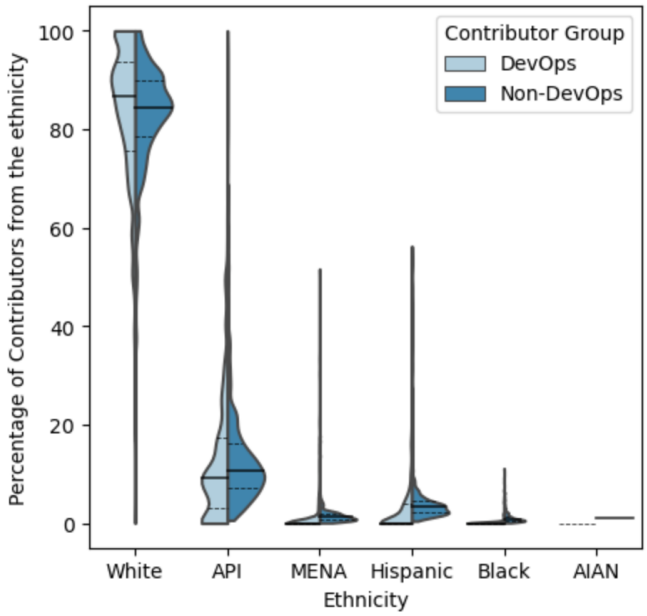


Fig. 1: Bean plots showing the distribution of percentages of DevOps and non- DevOps contributors from four perceptible ethnicities.

Table 1. Results of statistical analysis of ethnic proportions for DevOps and Non-DevOps contributors

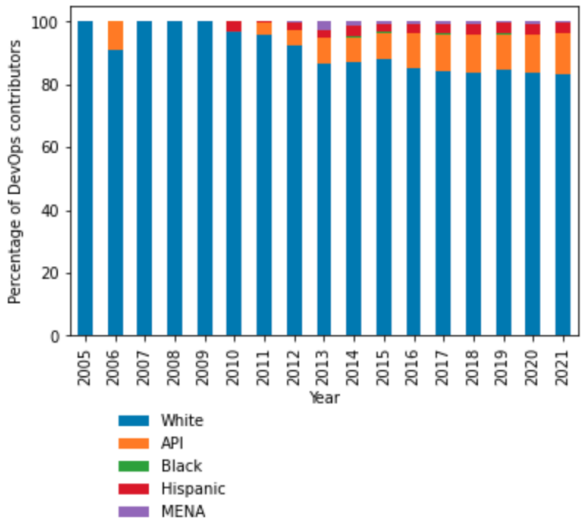
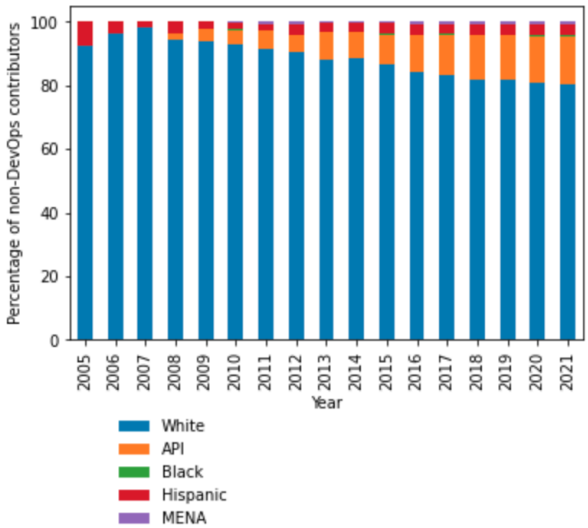
Ethnicity	**DevOps (median)**	**Non-DevOps (median)**	**Wilcoxon, one-tailed, paired ($\alpha = 0.0023$)**	**Effect size (Cliff's $\lvert\delta\rvert$)**
White	86.6%	84.5%	$0.002 < \alpha$ (greater than test)	negligible
API	9.1%	10.5%	$0.0008 < \alpha$ (less than test)	negligible
Hispanic	0.0%	3.2%	$7.9 \times 10^{-10} << \alpha$ (less than test)	medium
Black	0.0%	0.8%	$9.2 \times 10^{-19} << \alpha$ (less than test)	large
MENA	0.0%	1.2%	$3.0 \times 10^{-17} << \alpha$ (less than test)	large

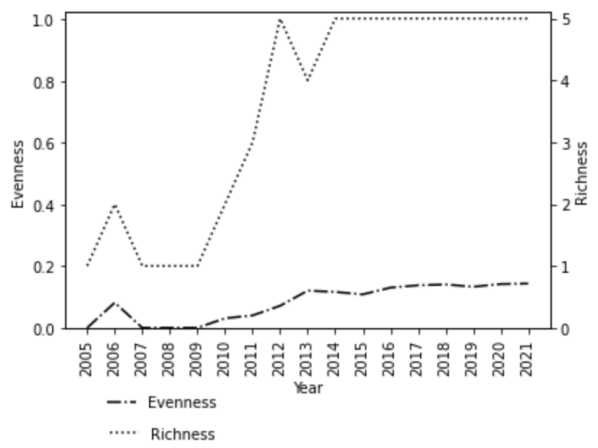
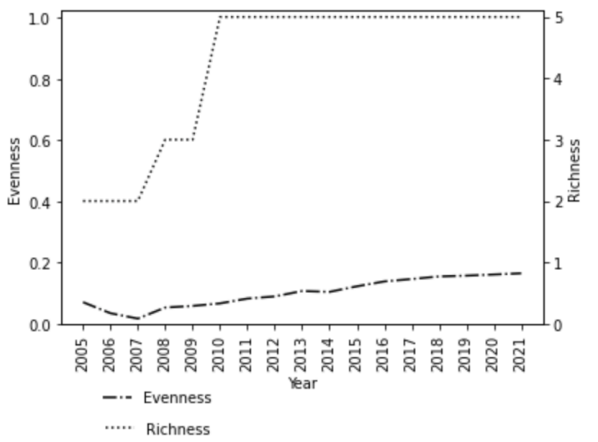
Table 2. Results of statistical analysis of ethnic diversity metrics for DevOps and Non-DevOps contributors

Diversity Metric	**DevOps (median)**	**Non-DevOps (median)**	**Wilcoxon, one-tailed, paired Effect size (α = 0.0023)**	**Effect size (Cliff's |δ |)**
Richness	2	4	3.1×10^{-53} << α (less than test)	large
Evenness	0.3	0.5	1.1×10^{-35} << α (less than test)	medium
Blau (Simpson)	0.24	0.27	1.9×10^{-7} << α (less than test)	negligible
Prevalence Rankings and Diffusion Score	9.1	10.2	0.006 > α (less than test)	-

(RQ2) How does the distribution of ethnic and gender diversity change as projects age?

Table 3. (RQ2-a) Change in Perceptible Ethnic Diversity Over Time

DevOps Contributors	Non-DevOps Contributors
	
Percentages of perceptible ethnicities of DevOps contributors over time.	Percentages of perceptible ethnicities of non-DevOps contributors over time

	
Values of diversity metrics of DevOps developers over time: the left y-axis presents the evenness values, while the right y-axis presents the richness values.	Values of diversity metrics of non-DevOps developers over time: the left y-axis presents the evenness values, while the right y-axis presents the richness values.
Average growth of richness of DevOps contributors over time: 0.1111111111111111 Average growth of evenness of DevOps contributors over time: 0.008023050476276687	Average growth of richness of non-DevOps contributors over time: 0.0 Average growth of richness of non-DevOps contributors over time: 0.008409261894237997

Appendix F: Results for the dataset of GitHub repositories using GH Actions

We use another dataset that contains commits of projects that adopt GitHub Actions[1] to check the generalizability of our conclusions. From these projects we randomly selected 100 projects filtered by similar criteria as for the CircleCI dataset: having more than 1,500 commits and more than 50 contributors. Our overall conclusions generalize to the repositories that use GitHub Actions. For example, the bean plots showing the distribution of percentages of DevOps and non-DevOps contributors from four perceptible ethnicities (i.e., White, API, Hispanic, Black) are shown Figure 2. From the figure, we see that non-White developers are less represented among DevOps developers compared to non-DevOps developers.

[1] https://github.com/features/actions

(PA) Preliminary Analysis: Identifying DevOps and non-DevOps Contributors

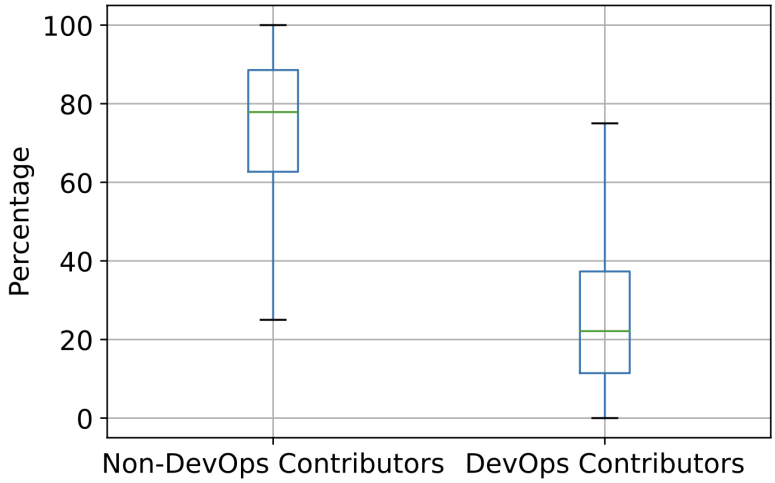


Fig. 1: Box-plots showing the percentages of contributors to non-DevOps files only and the percentages of contributors to DevOps files in the studied projects.

Median Non-DevOps Contributors per project: 77.87194%
Median DevOps Contributors per project: 22.12806%

(RQ1) Does the ethnic and gender diversity of DevOps contributors differ from ethnic and gender diversity of non-DevOps contributors?

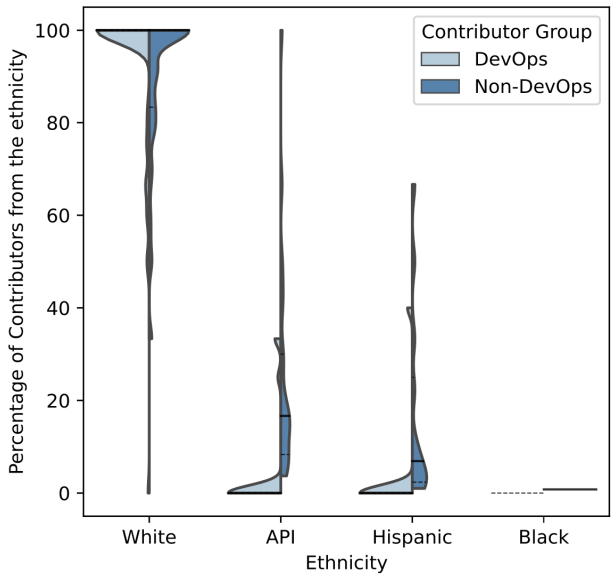


Fig. 2: Bean plots showing the distribution of percentages of DevOps and non- DevOps contributors from four perceptible ethnicities.

Table 1. Results of statistical analysis of ethnic proportions for DevOps and Non-DevOps contributors

Ethnicity	**DevOps (median)**	**Non-DevOps (median)**	**Wilcoxon, one-tailed, paired (α = 0.0023)**	**Effect size (Cliff's |δ |)**
White	100.0%	100.0%	0.03 > α (greater than test)	-
API	0.0%	16.7%	1.2×10^{-5} << α (less than test)	large
Hispanic	0.0%	6.9%	0.008 > α (less than test)	-
Black	0.0%	0.8%	0.5 > α (less than test)	-

Table 2. Results of statistical analysis of ethnic diversity metrics for DevOps and Non-DevOps contributors

Diversity Metric	**DevOps (median)**	**Non-DevOps (median)**	**Wilcoxon, one-tailed, paired Effect size (α = 0.0023)**	**Effect size (Cliff's |δ |)**
Richness	1	2	2.5×10^{-6} << α (less than test)	small
Evenness	0.0	0.0	0.002 < α (less than test)	small
Blau (Simpson)	0.0	0.0	0.01 > α (less than test)	-
Prevalence Rankings and Diffusion Score	0.0	0.0	0.004 > α (less than test)	-

Table 3. Results of statistical analysis of gender proportions for DevOps and Non-DevOps contributors

Gender	**DevOps (median)**	**Non-DevOps (median)**	**Wilcoxon, one-tailed, paired (α = 0.0023)**	**Effect size (Cliff's |δ |)**
Male	100.0%	95.8%	0.26 > α (greater than test)	-
Female	0.0%	3.9%	0.003 > α (less than test)	-

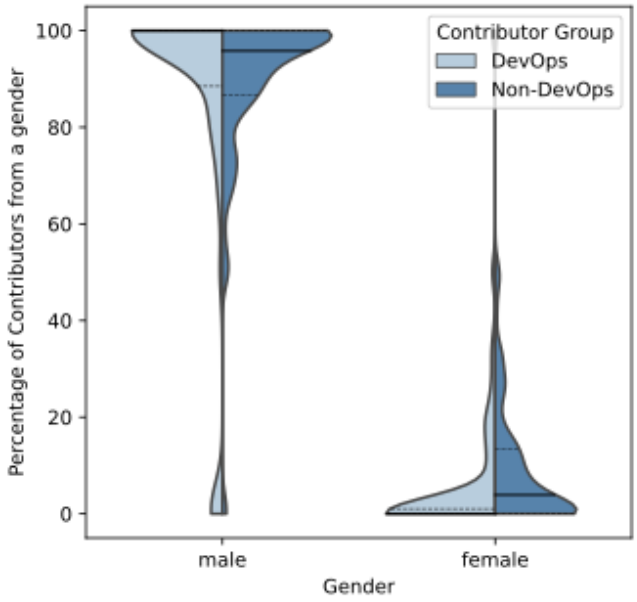


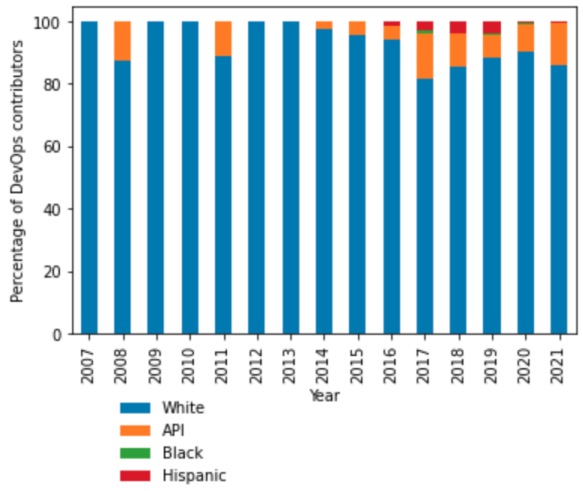
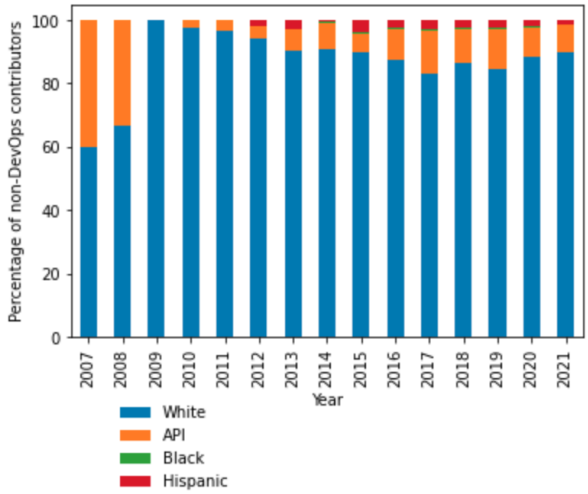
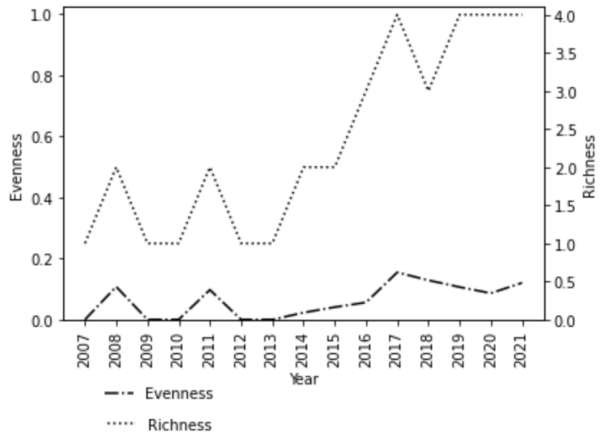
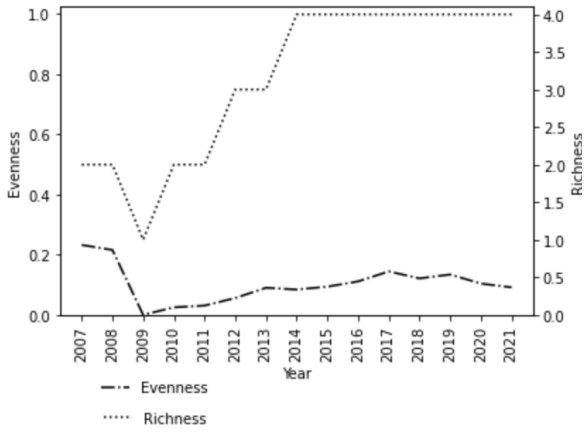
Fig. 3: Bean plots showing the distribution of percentages of DevOps and non- DevOps contributors from four perceptible genders (i.e., male and female). The solid lines represent the median percentages, and the dotted lines represent the first and third quantiles.

Table 4. Results of statistical analysis of gender diversity metrics for DevOps and Non-DevOps contributors

Diversity Metric	**DevOps (median)**	**Non-DevOps (median)**	**Wilcoxon, one-tailed, paired Effect size ($\alpha = 0.0023$)**	**Effect size (Cliff's $\lvert\delta\rvert$)**
Richness	1	2	$2.1 \times 10^{-8} \ll \alpha$ (less than test)	medium
Evenness	0.0	0.1	$1.7 \times 10^{-5} \ll \alpha$ (less than test)	small
Blau (Simpson)	0.0	0.05	$0.001 < \alpha$ (less than test)	small
Prevalence Rankings and Diffusion Score	0.0	3.5	$0.004 > \alpha$ (less than test)	-

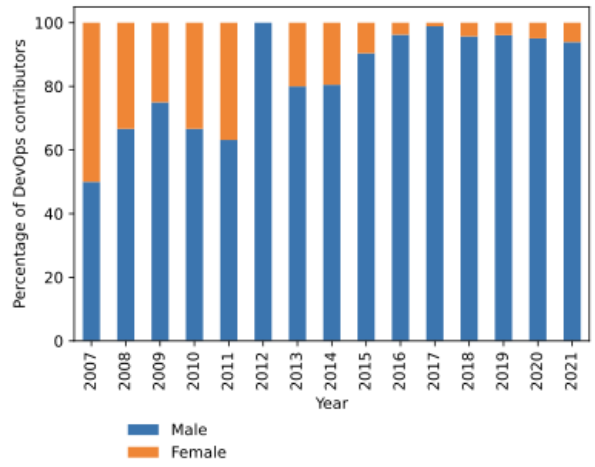
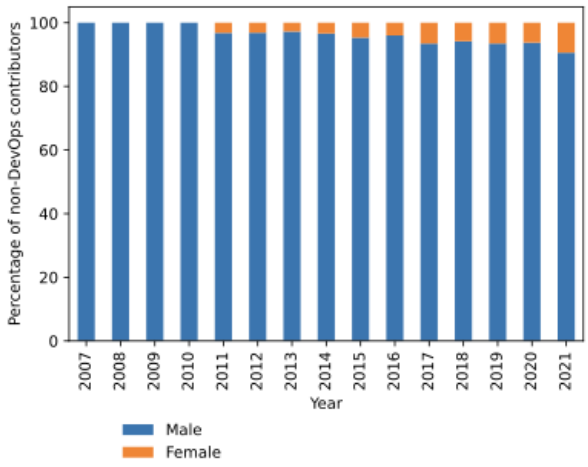
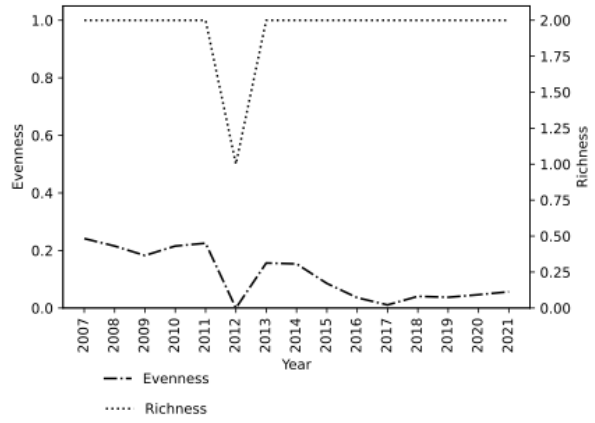
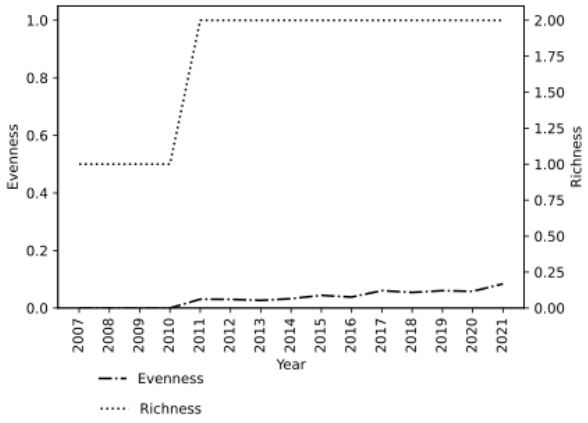
(RQ2) How does the distribution of ethnic and gender diversity change as projects age?

Table 5. (RQ2-a) Change in Perceptible Ethnic Diversity Over Time

DevOps Contributors	Non-DevOps Contributors
	
Percentages of perceptible ethnicities of DevOps contributors over time.	Percentages of perceptible ethnicities of non-DevOps contributors over time
	
Values of diversity metrics of DevOps developers over time: the left y-axis presents the evenness values, while the right y-axis presents the richness values.	Values of diversity metrics of non-DevOps developers over time: the left y-axis presents the evenness values, while the right y-axis presents the richness values.
Average growth of richness of DevOps contributors over time: 0.1111111111111111 Average growth of evenness of DevOps contributors over time: 0.013933884396320631	Average growth of richness of non-DevOps contributors over time: 0.0 Average growth of richness of non-DevOps contributors over time:

	-0.0008356508559769624

Table 6. (RQ2-b) Change in Perceptible Gender Diversity Over Time

DevOps Contributors	Non-DevOps Contributors
	
Percentages of perceptible genders of DevOps contributors over time.	Percentages of perceptible genders of non-DevOps contributors over time
	
Values of diversity metrics of DevOps developers over time: the left y-axis presents the evenness values, while the right y-axis presents the richness values.	Values of diversity metrics of non-DevOps developers over time: the left y-axis presents the evenness values, while the right y-axis presents the richness values.
Average growth of richness of DevOps contributors over time: 0.1111111111111111 Average growth of evenness of DevOps contributors over time: 0.006294225715650383	Average growth of richness of non-DevOps contributors over time: 0.0 Average growth of richness of non-DevOps contributors over time: 0.005986279440711415

Appendix G: Results after removing the contributors who change their names from one to another and there are slight variations in their ethnicities and genders inferred by the tools after the change.

We remove such contributors from our dataset and recompute the results, and below, we report the results. We find that our conclusions remain the same. For example, with respect to RQ1, Figure 2 shows the proportions of different perceived ethnicities among DevOps and non-DevOps contributors. We find our results still hold even after removing the contributors who are perceived as multiple ethnicities in our dataset. For example, we observe that perceptibly White contributors are the dominating ethnicity among the rest. With respect to diversity indices, we find that the richness and evenness among DevOps contributors is significantly less than that among non-DevOps contributors with medium and small effect sizes, respectively.

(PA) Preliminary Analysis: Identifying DevOps and non-DevOps Contributors

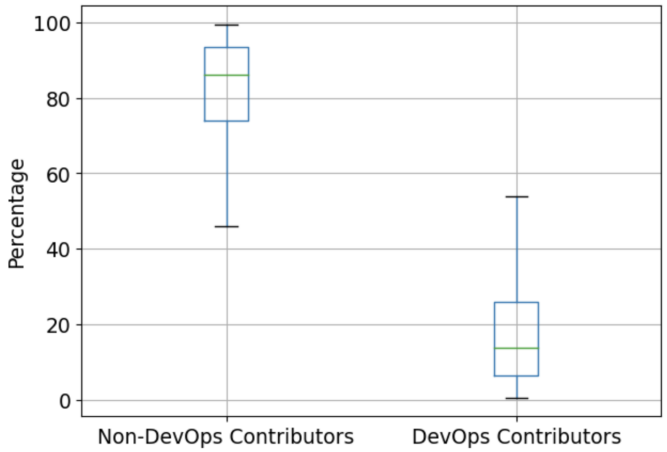


Fig. 1: Box-plots showing the percentages of contributors to non-DevOps files only and the percentages of contributors to DevOps files in the studied projects.

Median Non-DevOps Contributors per project: 86.095822%
Median DevOps Contributors per project: 13.904178%

(RQ1) Does the ethnic and gender diversity of DevOps contributors differ from ethnic and gender diversity of non-DevOps contributors?

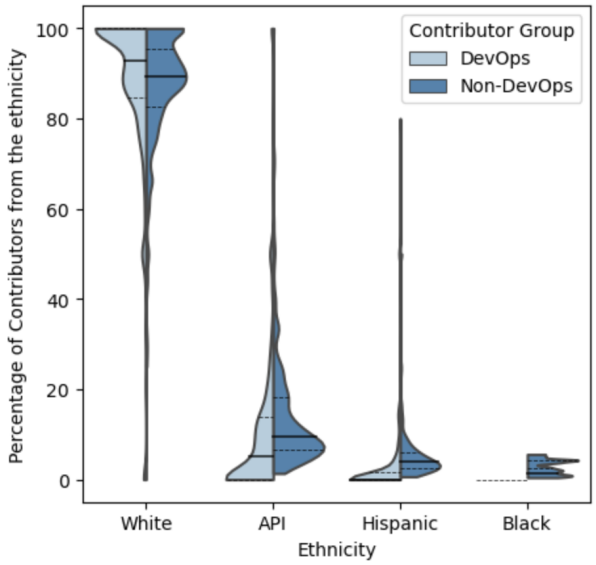


Fig. 2: Bean plots showing the distribution of percentages of DevOps and non- DevOps contributors from four perceptible ethnicities.

Table 1. Results of statistical analysis of ethnic proportions for DevOps and Non-DevOps contributors

Ethnicity	**DevOps (median)**	**Non-DevOps (median)**	**Wilcoxon, one-tailed, paired ($\alpha = 0.0023$)**	**Effect size (Cliff's $\lvert\delta\rvert$)**
White	92.8%	89.5%	$0.0003 < \alpha$ (greater than test)	small
API	5.1%	9.5%	$1.7 \times 10^{-11} << \alpha$ (less than test)	medium
Hispanic	0.0%	3.8%	$1.7 \times 10^{-15} << \alpha$ (less than test)	large
Black	0.0%	2.6%	$1.5 \times 10^{-5} << \alpha$ (less than test)	large

Table 2. Results of statistical analysis of ethnic diversity metrics for DevOps and Non-DevOps contributors

Diversity Metric	**DevOps (median)**	**Non-DevOps (median)**	**Wilcoxon, one-tailed, paired Effect size ($\alpha = 0.0023$)**	**Effect size (Cliff's $\lvert\delta\rvert$)**
Richness	2	2	$1.33 \times 10^{-25} << \alpha$ (less than test)	medium
Evenness	0.18	0.31	$4.55 \times 10^{-19} << \alpha$ (less than test)	small
Blau (Simpson)	0.11	0.19	$8.28 \times 10^{-9} << \alpha$ (less than test)	small
Prevalence Rankings and Diffusion Score	0.0	7.14	$0.0002 < \alpha$ (less than test)	small

Table 3. Results of statistical analysis of gender proportions for DevOps and Non-DevOps contributors

Gender	**DevOps (median)**	**Non-DevOps (median)**	**Wilcoxon, one-tailed, paired ($\alpha = 0.0023$)**	**Effect size (Cliff's $\lvert\delta\rvert$)**
Male	93.7%	92.9%	$2.0 \times 10^{-5} << \alpha$ (greater than test)	small
Female	6.2%	7.1%	$2.0 \times 10^{-5} << \alpha$ (less than test)	small

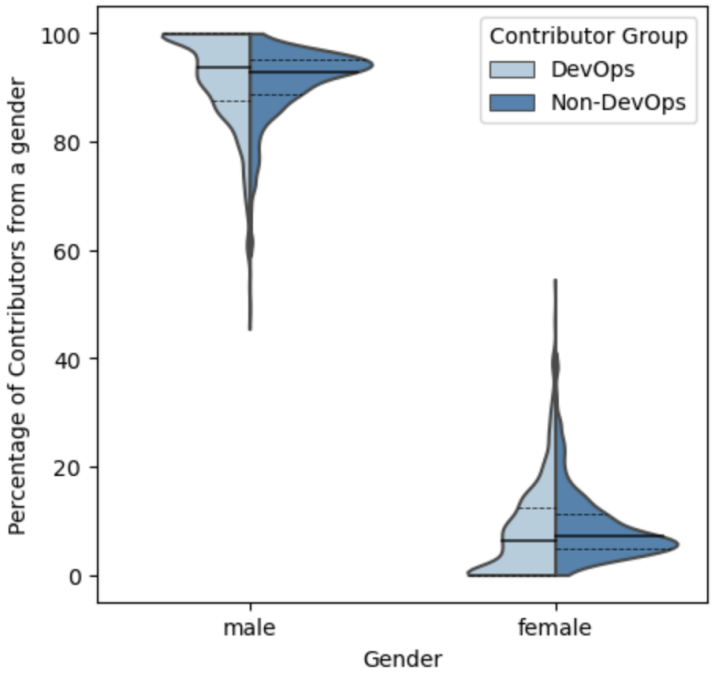


Fig. 3: Bean plots showing the distribution of percentages of DevOps and non- DevOps contributors from four perceptible genders (i.e., male and female). The solid lines represent the median percentages, and the dotted lines represent the first and third quantiles.

Table 4. Results of statistical analysis of gender diversity metrics for DevOps and Non-DevOps contributors

Diversity Metric	**DevOps (median)**	**Non-DevOps (median)**	**Wilcoxon, one-tailed, paired Effect size ($\alpha = 0.0023$)**	**Effect size (Cliff's $\|\delta\|$)**
Richness	2	2	$2.4 \times 10^{-26} \ll \alpha$ (less than test)	small
Evenness	0.18	0.24	$9.9 \times 10^{-21} \ll \alpha$ (less than test)	small
Blau (Simpson)	0.12	0.13	$4.01 \times 10^{-7} \ll \alpha$ (less than test)	small
Prevalence Rankings and Diffusion Score	6.25	7.13	$1.60 \times 10^{-5} \ll \alpha$ (less than test)	small

Similarly, for RQ2. we find similar results to the original study.

Appendix H: Triangulating with Canedo et al. [2]

Inspired by the findings of Canedo et al.'s study, we perform the following two analyses to better understand potential patterns for contributors made by different genders.

Analysis 1. In this analysis, we analyze commit activity (i.e., number of commits) of two groups: one consisting of randomly selected 1000 DevOps contributors who are perceptibly female (F), and another consisting of four randomly selected samples of 1000 DevOps contributors who are perceptibly male (M*). Additionally, we apply that same analysis to non-DevOps contributors.

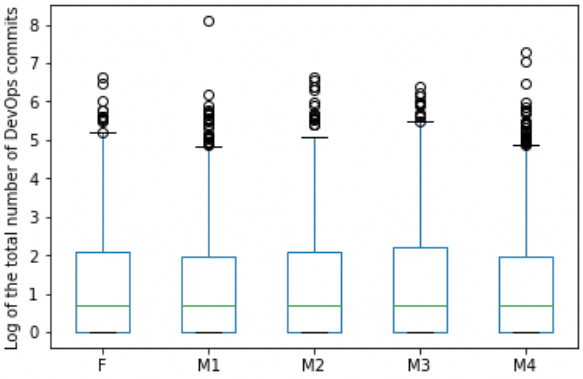


Fig. 1: Number of commits from the five sets of DevOps contributors. F stands for the random sample of perceptibly female DevOps contributors while M* stands for the random samples of male DevOps contributors

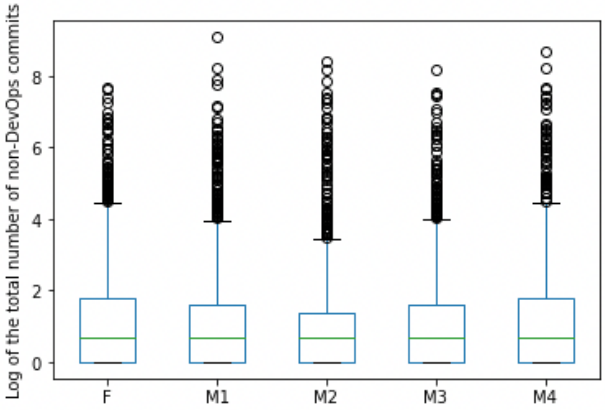


Fig. 2: Number of commits from the five sets of non-DevOps contributors. F stands for the random sample of perceptibly female non-DevOps contributors while M* stands for the random samples of male non-DevOps contributors

Results. Figure 1 illustrates the total number of DevOps commits across the sets (M1, M2, M3, M4, and F), with the y-axis represented on a logarithmic scale. It is worth noting that the results

of the Mann-Whitney U tests for the female contributors (F) and each sample of male contributors (M*) do not demonstrate any significant differences (p >> 0.0023).

Similarly, Figure 2 presents the total number of commits by non-DevOps contributors within the five samples, and our analysis using the Mann-Whitney U test did not reveal any significant differences (p >> 0.0023)

Analysis 2. To understand if there is any difference in the level of engagement of contributors from different genders, we analyze the frequency corresponding to the percentage of days the developers contribute to DevOps and non-DevOps artifacts. The frequency is defined as follows.

$$Frequency = \frac{distinct\ days}{intercal\ (min\ date, max\ date)} \times 100 \quad [3]$$

In particular, we compute the frequency of each contributor in one random sample of 1000 perceptible female DevOps contributors, and four random samples of 1000 perceptibly male DevOps contributors.

Results. Similar to Analysis 1, we do not observe any significant difference between perceptibly male and female contributors of their frequency of either committing to DevOps or non-DevOps artifacts (Mann-Whitney U; p>>0.0023).

Overall, our findings provide a complement to the study conducted by Canedo et al. [2]. While Canedo et al. focused on core developers within a project, we distinguished between DevOps contributors and non-DevOps contributors.

References

[1] B. Vasilescu, A. Capiluppi, and A. Serebrenik, “Gender, representation and online participation: A quantitative study,” *Interact. Comput.*, vol. 26, no. 5, pp. 488–511, 2014.

[2] E. D. Canedo, R. Bonifácio, M. V. Okimoto, A. Serebrenik, G. Pinto, and E. Monteiro, “Work practices and perceptions from women core developers in OSS communities,” in *Proceedings of the 14th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, 2020, pp. 1–11.

[3] W. et al., “An Empirical Comparison of Ethnic and Gender Diversity of DevOps and non-DevOps Contributions to Open-Source Projects (Anonymized).” Zenodo, Sep. 2022. doi: 10.5281/zenodo.7843085.