



# **Equity & inclusion as a way of developing diverse, resilient, and more relevant open source software**

Griselda Cuevas | Aizhamal Nurmamat kyzy

Airflow Summit  
July 6-17, 2020

# What we will discuss

- Common challenges for Open Source Communities
- State of Diversity in Open Source
- Diversity & Inclusion | Apache Software Foundation
- Diversity & Inclusion | Apache Airflow community



# **Work Overload** for Project Maintainers



# High Number of Open Bugs

# Slow Project

## Adoption





# Communities in Other Countries you haven't engaged



# Maybe...

# **Equity**

# **Diversity**

# **Inclusion**



# Diversity & Inclusion in open source

The process of employing/attracting a diverse team of people that's reflective of the society in which the technology or project exists and operates.

***“Diversity is a Metric, Inclusion is a process  
and Equity is an outcome.”***

@jesshmitchel

# The ASF 2020 Community Survey

## The Research Team



**Anita Sarma**  
Oregon State University



**Daniel Izquierdo**  
Bitergia



**Mariam Guizani**  
Oregon State University



**Georg Link**  
Bitergia



**Griselda Cuevas**  
The Apache Software  
Foundation

# Motivation

---

**Survey** design, data collection, and analysis aligned with the **ASF D&I strategy**.

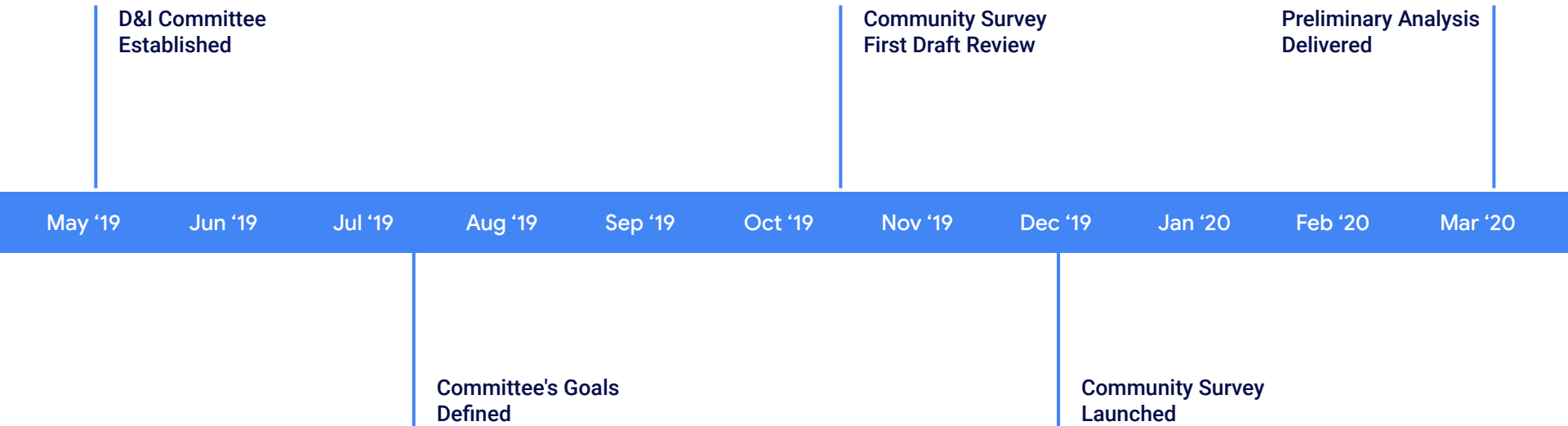
## 2019 short term goals

- **Gather scientific data** to study current status of Diversity and Inclusion at the ASF
- **Raise awareness** in our community about the importance of Diversity & Inclusion in the business, and in the open source industry
- **Find key indicators** to track over time

## 2020 medium term goals

- Instrument ASF's contributor funnel to **recommend a participation baseline** for underrepresented groups
- Create a toolkit to **address the top 3 entry barriers for new contributors** from under represented groups
- **Become a trusted thought partner** for PMCs when it comes to D&I

# The 2020 ASF Community Survey Timeline



## 3 Efforts to gather data about the ASF community



**The Community  
Survey**

**Project Metrics  
Analysis**

**New Contributor  
and Mentor  
Friction Logs**



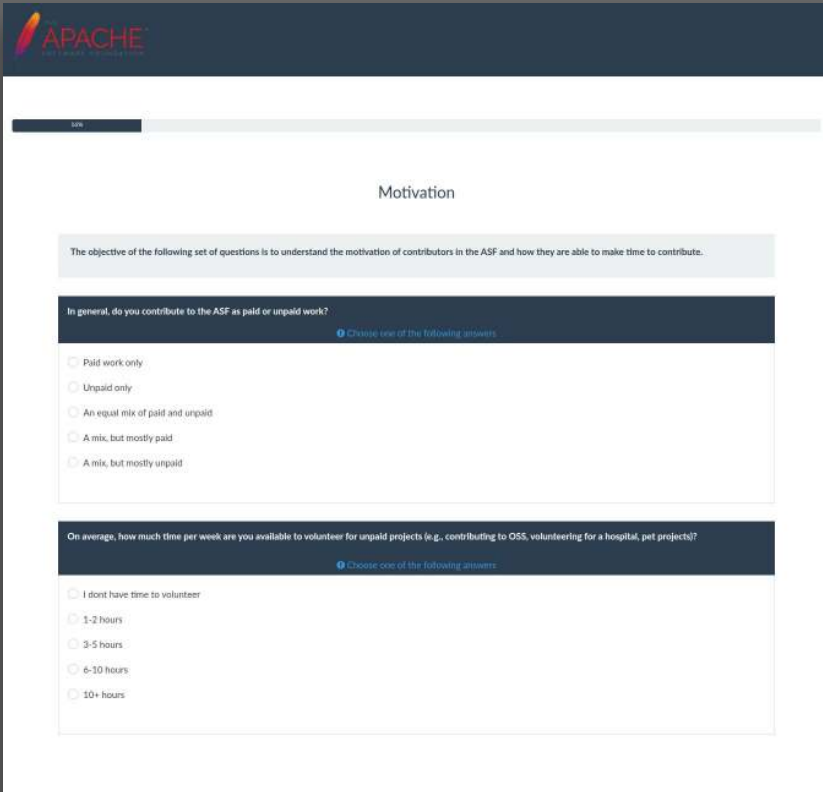
# Survey

---

**Survey\*** is split into:

1. Contributor Role and Tenure (4 questions)
2. Motivation (2)
3. Availability of Protocols / Guidelines (2)
4. Support for Newcomers (4)
5. Diversity and Inclusion (11)
6. Wrap up (3)

Each section contains a motivation introductory paragraph.



The screenshot shows the 'Motivation' section of the Apache survey. At the top is the Apache logo. Below it is a progress bar with 'Motivation' selected. The section title 'Motivation' is centered. A light blue box contains the objective: 'The objective of the following set of questions is to understand the motivation of contributors in the ASF and how they are able to make time to contribute.' The first question is 'In general, do you contribute to the ASF as paid or unpaid work?' with a link 'Choose one of the following answers'. It has five radio button options: 'Paid work only', 'Unpaid only', 'An equal mix of paid and unpaid', 'A mix, but mostly paid', and 'A mix, but mostly unpaid'. The second question is 'On average, how much time per week are you available to volunteer for unpaid projects (e.g., contributing to OSS, volunteering for a hospital, pet projects)?' with a link 'Choose one of the following answers'. It has five radio button options: 'I dont have time to volunteer', '1-2 hours', '3-5 hours', '6-10 hours', and '10+ hours'.

APACHE  
The Apache Software Foundation

Motivation

The objective of the following set of questions is to understand the motivation of contributors in the ASF and how they are able to make time to contribute.

In general, do you contribute to the ASF as paid or unpaid work?  
[Choose one of the following answers](#)

☐ Paid work only

☐ Unpaid only

☐ An equal mix of paid and unpaid

☐ A mix, but mostly paid

☐ A mix, but mostly unpaid

On average, how much time per week are you available to volunteer for unpaid projects (e.g., contributing to OSS, volunteering for a hospital, pet projects)?  
[Choose one of the following answers](#)

☐ I dont have time to volunteer

☐ 1-2 hours

☐ 3-5 hours

☐ 6-10 hours

☐ 10+ hours

The background image shows a laptop screen with a dark blue overlay. On the screen, there is a line graph with a fluctuating line and a pie chart. The text 'The ASF Community Readout Summary' is prominently displayed in the center in a large, white, sans-serif font.

# The ASF Community Readout Summary

# Analysis Dimensions

---

- Demographics:
  - Age, Gender, English fluency, Background culture
- Socio-economic aspects:
  - Education, Compensation, Time to volunteer
- Experiences in the ASF:
  - Tenure, mentorship, challenges

# Method

---



## Demographics Analysis

Population analysis that help draw a first approach to the ASF ecosystem. This section mixes several attributes to understand specific areas, including education, gender, migrants, newcomers and challenges.

## Profile Analysis

Identification of key profiles to be part of the *in depth interview process*. These are dependent of significant attributes important for ASF EDI working group such as gender, English speaking confidence, and others.

# Survey Results

■ New Visitor ■ Returning Visitor



# Average ASF contributor

---

## *Demographics*

40 Years old

Man

Confident in  
English

Born/lives in the  
USA

## *Socio-economic Aspects*

Bachelor's  
Degree

No  
compensation

1 or 2 hours for  
volunteering

## *Experience in the ASF*

5 Years in the  
community

Didn't have a  
mentor

Faced no  
challenges



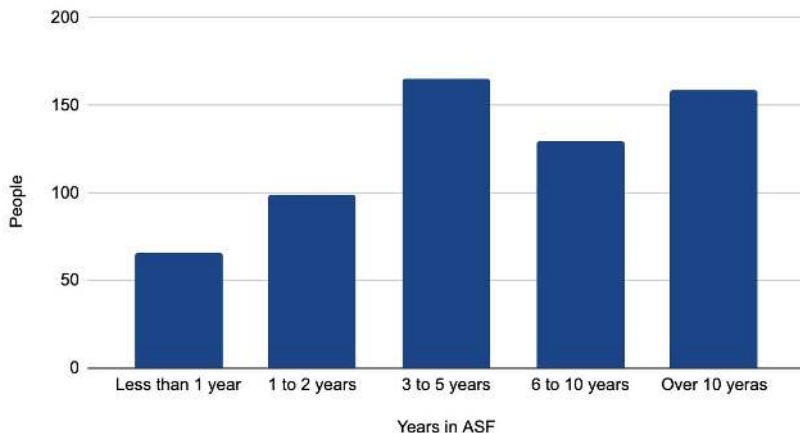
# Getting Experienced

People with less than 1 year of experience represents **10%** of the answers

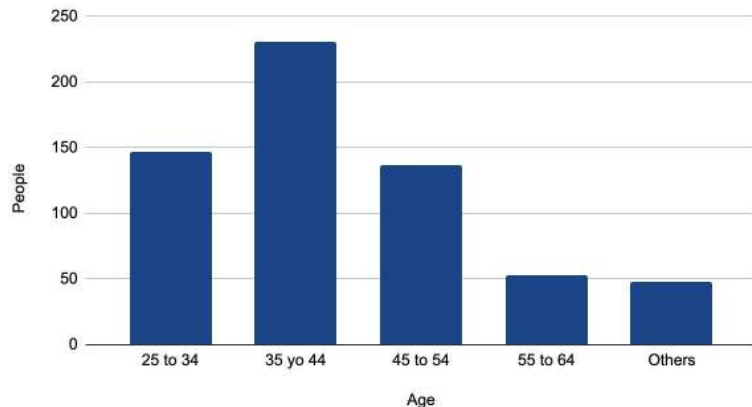
Participants under 24 represents less than a **5 %**

Are there barriers for newcomers or young adults?

People vs Years in ASF



People vs Age



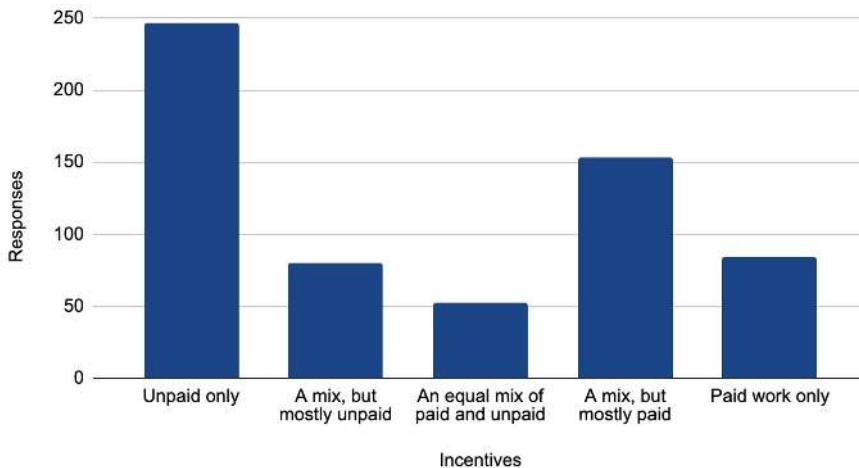
# Mostly Volunteers

40% work as volunteers

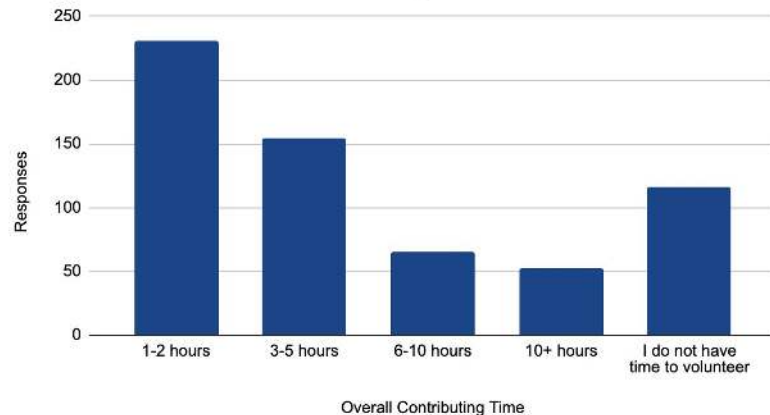
81% have less than 5 hrs/week  
for volunteering

Are there ways to be more  
efficient? To help those unpaid  
volunteers to contribute?

Responses vs Incentives



Responses vs Overall Contributing Time



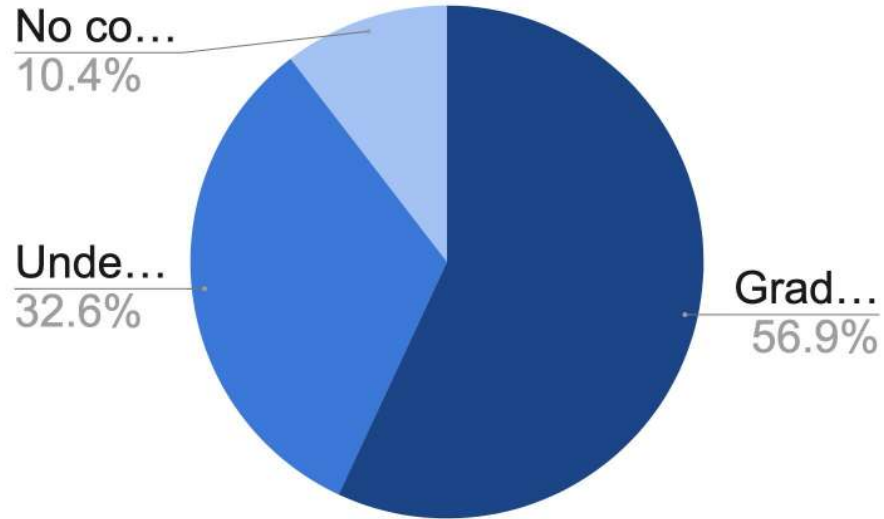
# Highly Educated

---

Around 90% of ASF contributors are college educated or higher

Do these contributors know about OSS at their schools?

Are the “no college” group aware of OSS at the same level?

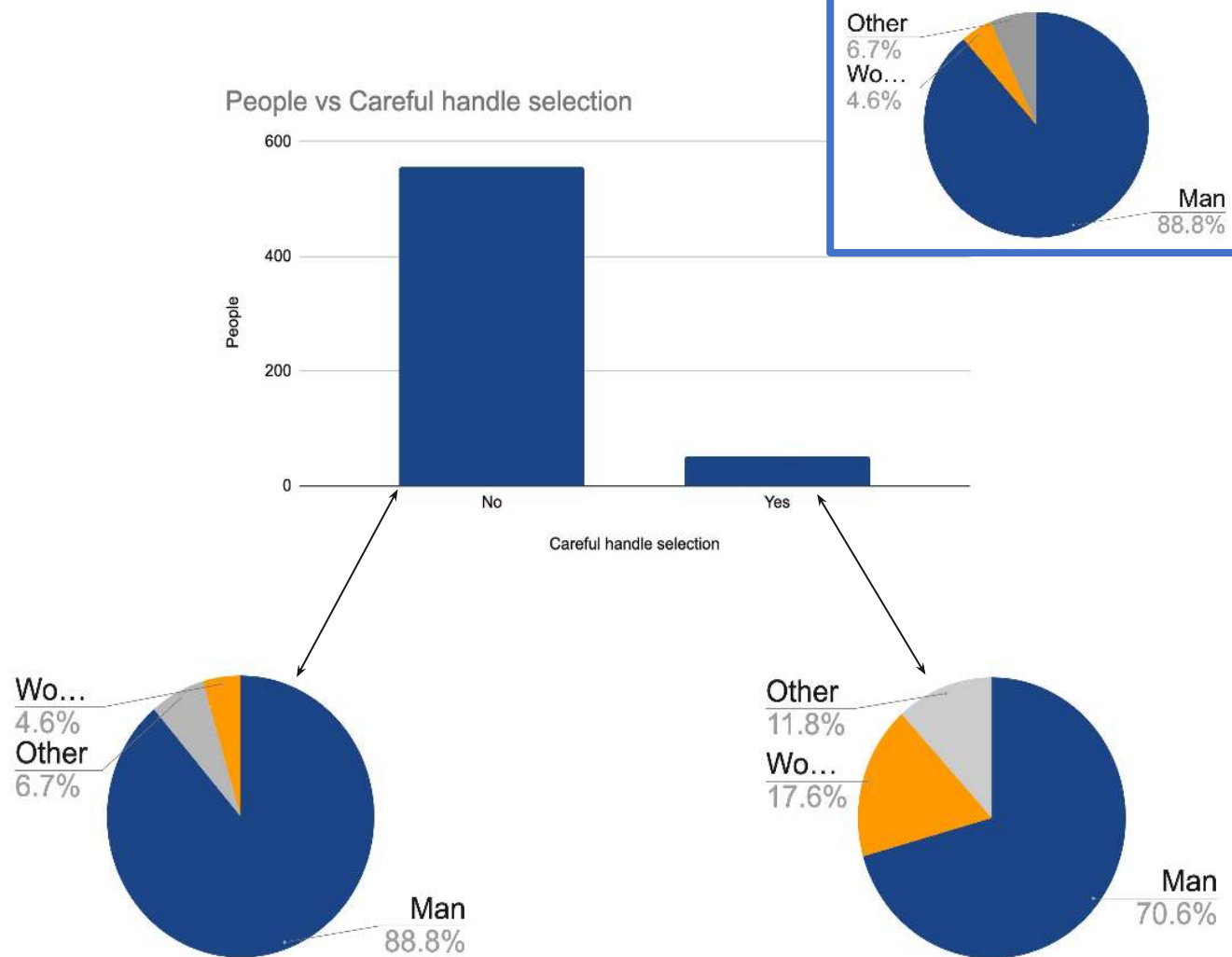


# Mostly Men

11% have self-defined as women or with another gender

8.4% have carefully think what gender their handles represent

Are there issues related to non-gender neutral handle?  
How can we improve the percentage of people coming from the Women and Others group?



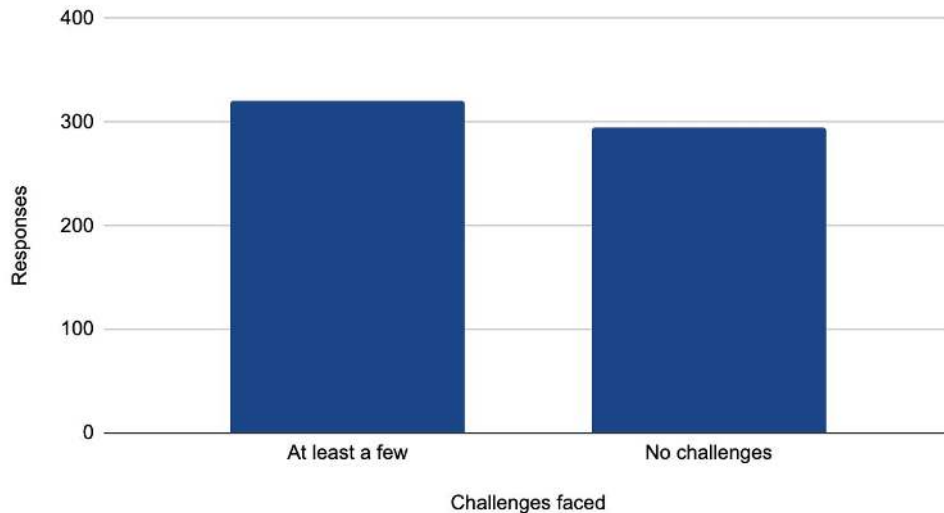
# There are Challenges

About 52% of ASF contributors have faced challenges

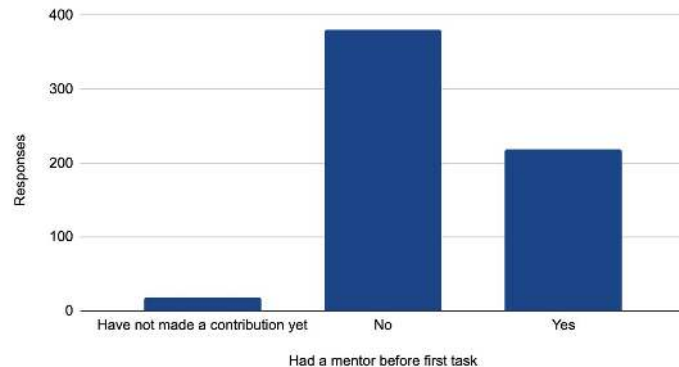
62% of contributors didn't have a mentor

How can we reduce these challenges to make the contributors path easier?

Responses vs Challenges faced



Responses vs Had a mentor before first task



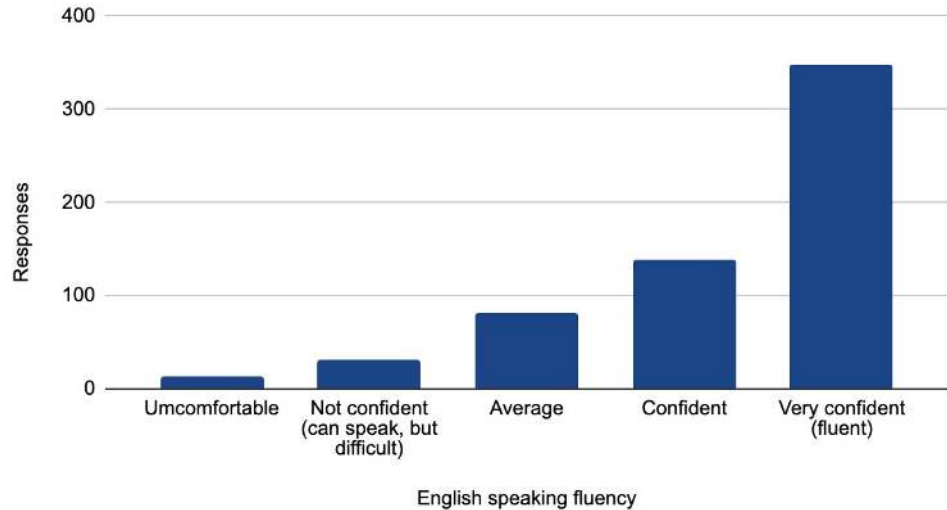
# 7% not English confident

---

Around 20% of contributors feel they have an average English level or less

Language seems to be a barrier for communication

Responses vs English speaking fluency





## Results: Demographics

Culture	# Responses	% Responses
United States	171	28.84%
Germany	69	11.64%
China	49	8.26%
France	32	5.40%
India	28	4.72%
United Kingdom	25	4.22%
Russian Federation	23	3.88%
Canada	18	3.04%
Spain	14	2.36%
Japan	11	1.85%

*Top 10 countries by Culture*

Residence	# Responses	% Responses
United States	230	38.27%
Germany	76	12.65%
China	37	6.16%
United Kingdom	28	4.66%
France	27	4.49%
Canada	18	3.00%
Netherlands	15	2.50%
Russian Federation	14	2.33%
Switzerland	12	2.00%
Japan	11	1.83%

*Top 10 countries by Residence*

# Hypotheses



# Hypotheses

---

Education impacts compensation

Those in minority face challenges

Men who face challenges have different demographics

The culture where you grew up makes a difference

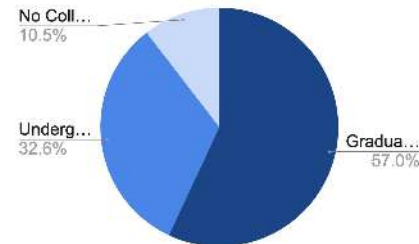
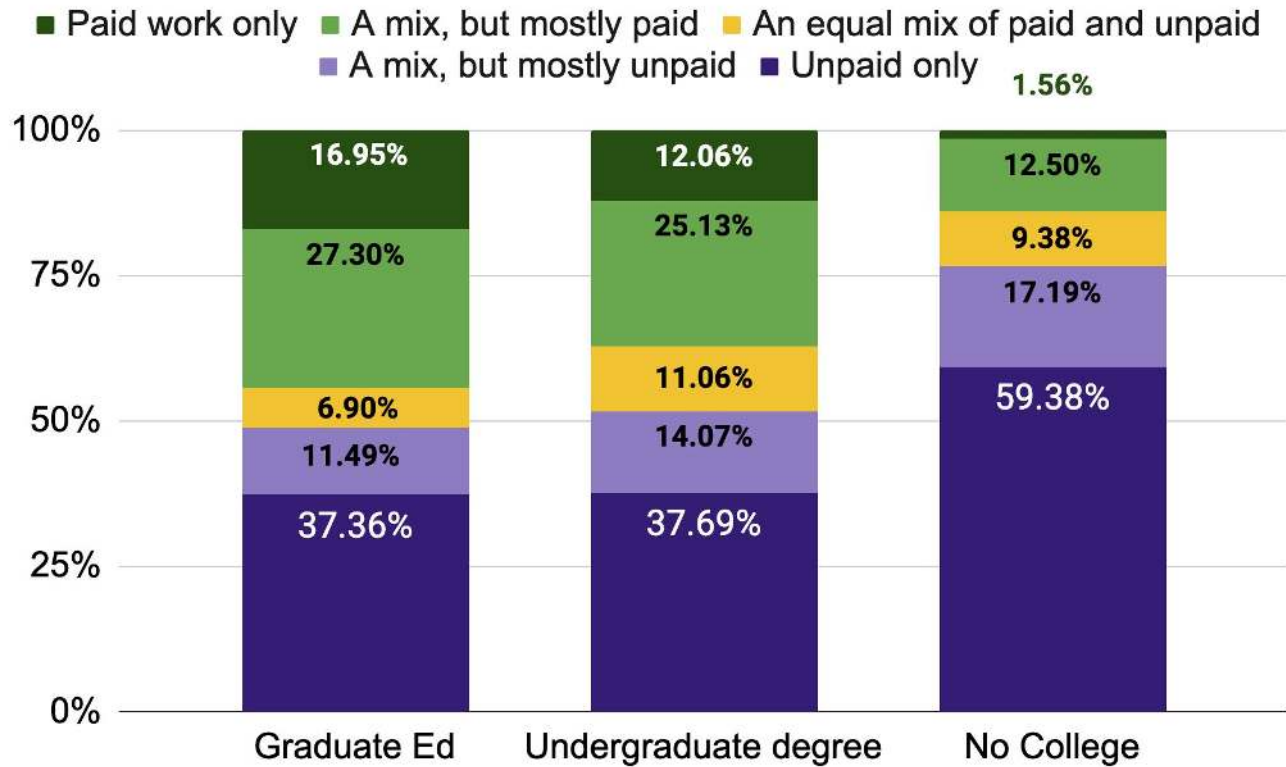
---

**Education impacts compensation**

# ↑ Education ↑ Compensation

Increase of Unpaid population  
for those with no college  
education

Willing to work as volunteer to  
gain experience?



---

**Those in minority face challenges**

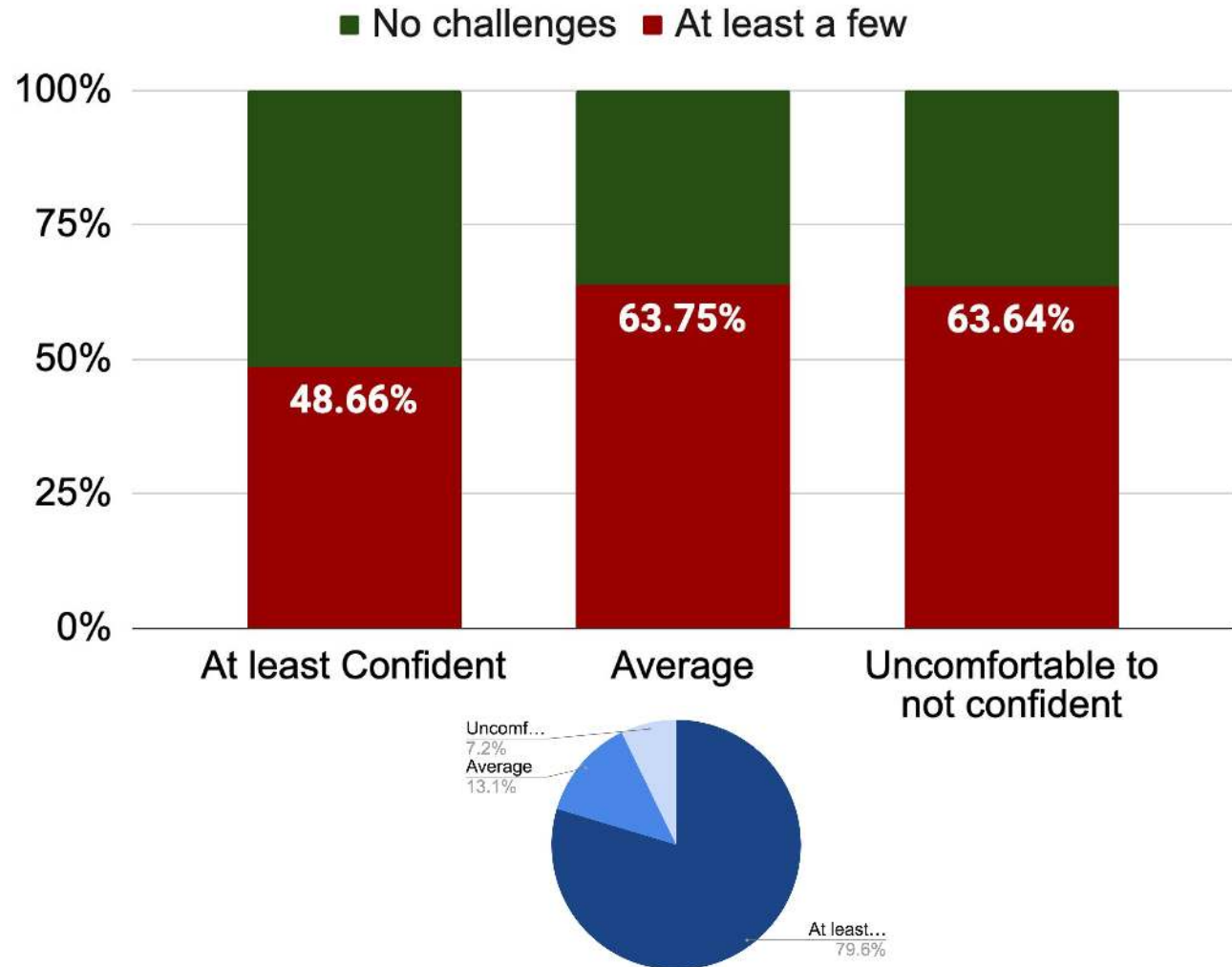
- English fluency**
- Gender**
- Having a mentor**



# Language

Those with an average English level or less faced more challenges

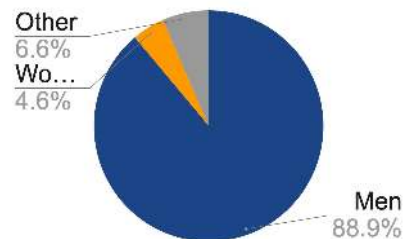
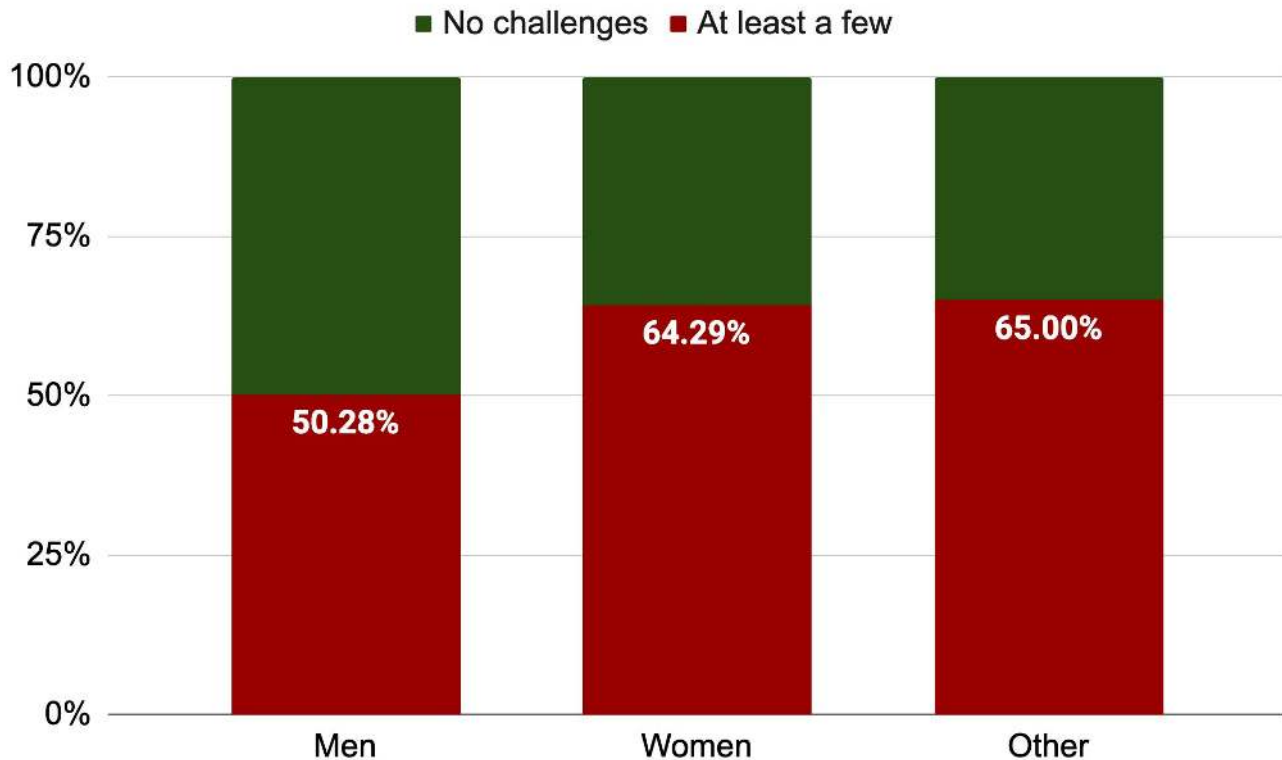
How can we lower the language barrier?



# Gender

Those that self-assigned as Women or other groups (not men) faced more challenges

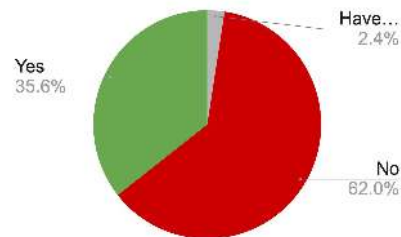
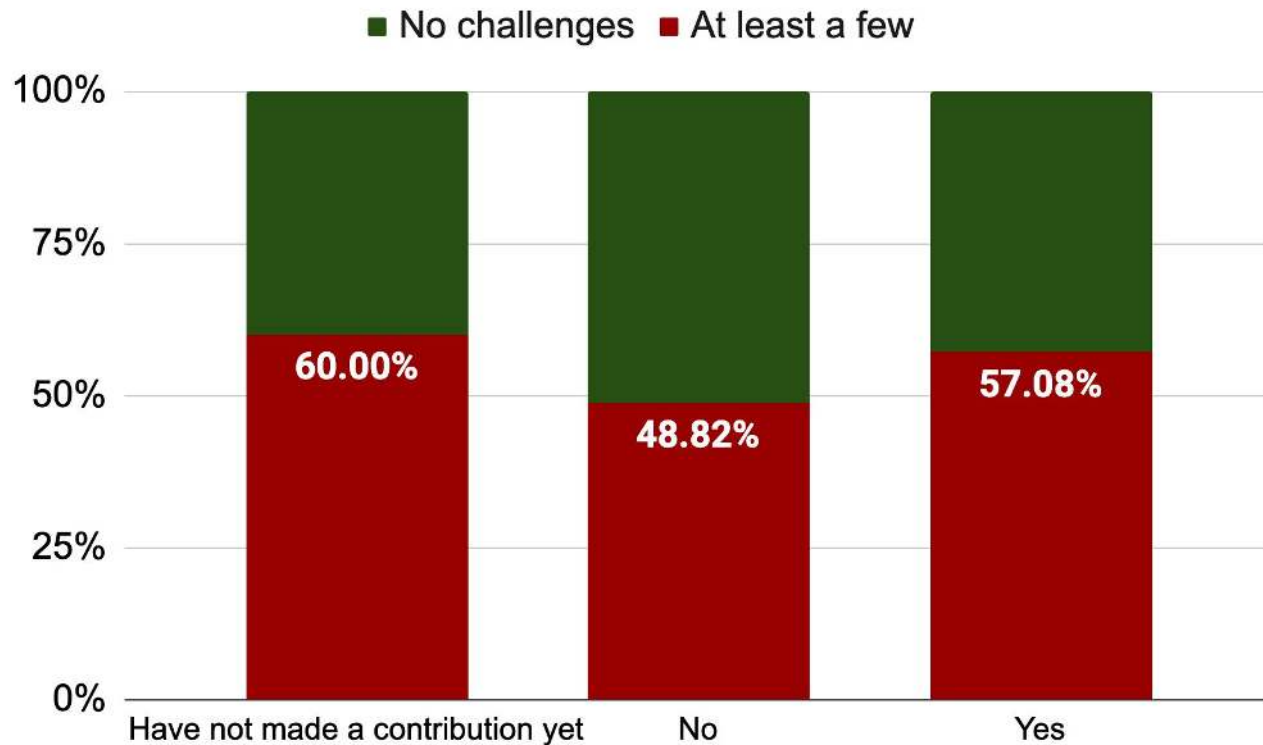
How can we reduce this barrier?



# Having a Mentor

Those with a mentor faced more challenges

How can we reduce this barrier?



---

# Men who face challenges have different demographics

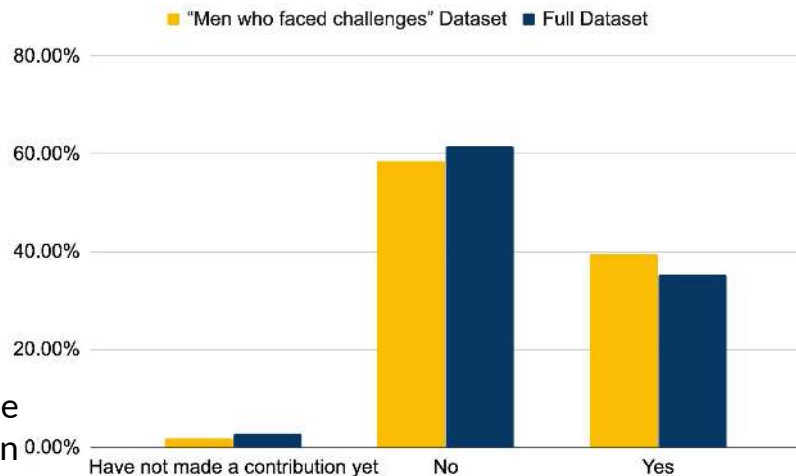
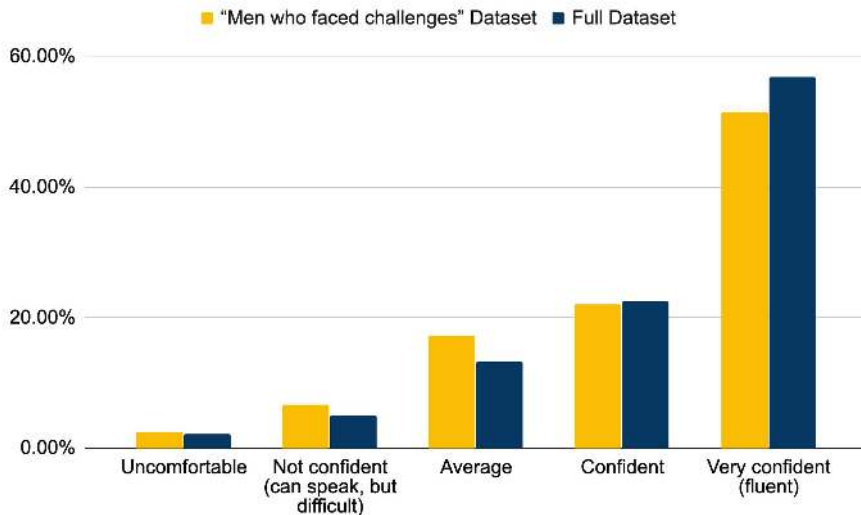
\* Population of 272 people

# No main differences

Men with certain challenges do not show a different distribution if compared to the full dataset

Less confident in English

Slightly higher distribution of those that had a mentor before their first contribution



Men that had a mentor before their first contribution

---

**Those who move from the country  
where they grew up have...**

**... different demographics**

**... challenges to contributing**

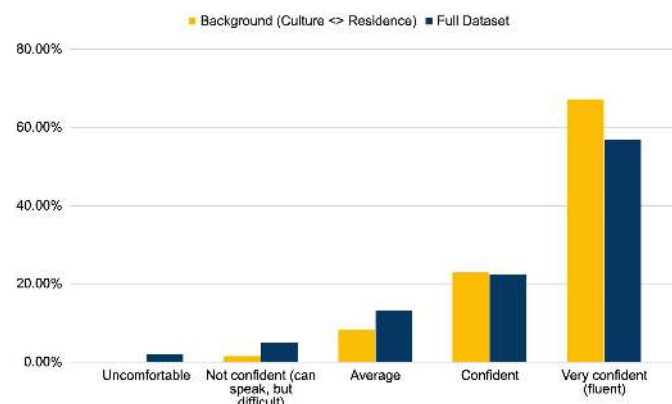
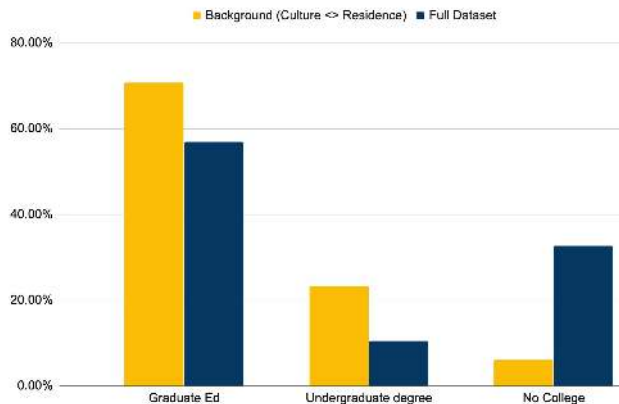
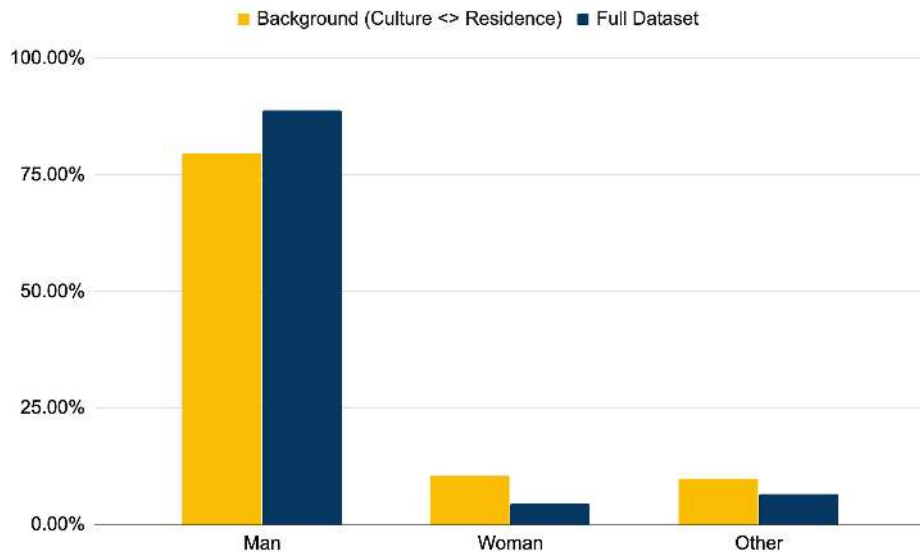
\* Population of 134 people

# Cultural Background

There are more women with a different cultural background and current residence

In general, those that moved to another country have a higher education and feel more English confident

And more are compensated for participating



# Recap

---

1. Education impacts compensation: yes ( $p < 0.05$ )\*
2. Those in minority face challenges
  - a. English fluency: Yes ( $p < 0.05$ )
  - b. Gender: Yes ( $p < 0.05$ )
  - c. Having a mentor: No ( $p = 0.05$ )
3. Men who face challenges have different demographics: No
4. Those who move country where they grew up...
  - a. ... different demographics: No
  - b. ... challenges to contributing: No ( $p > 0.05$ )

\*Chi<sup>2</sup> test for significance



---

**Question: what barriers  
newcomers face?**

\* Population of 18 people

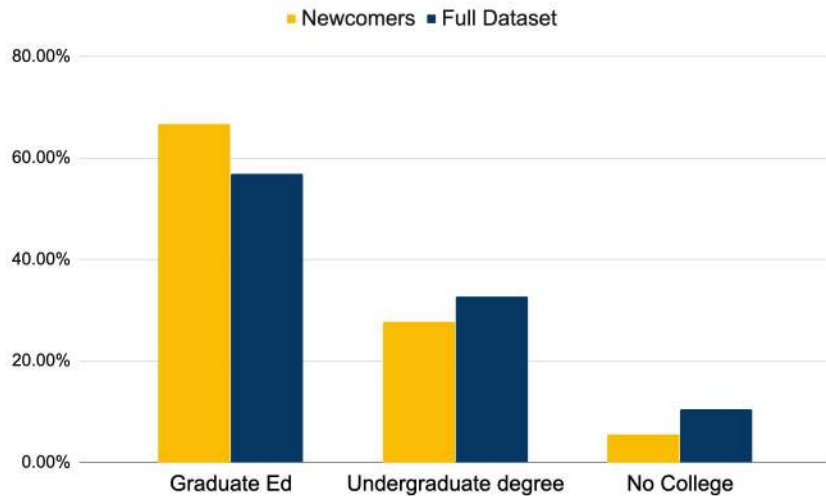
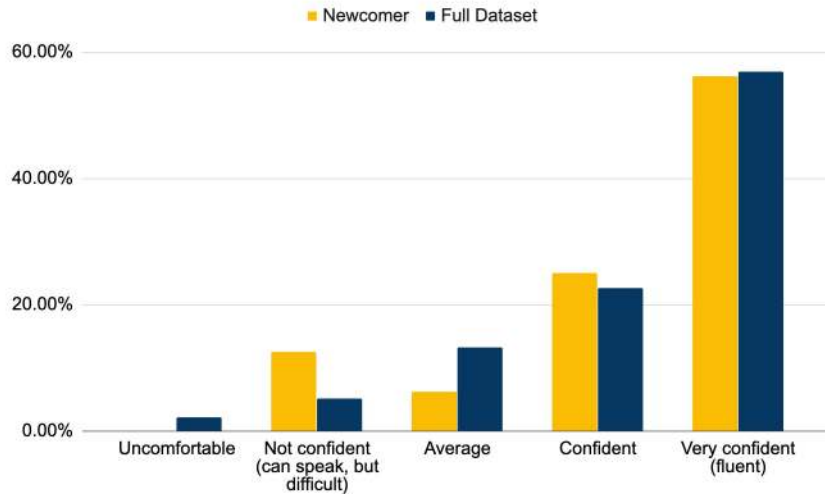
# Newcomers have barriers

A higher percentage of them face challenges

There are more women in percentage. Is gender a factor for this?

The language might be another barrier

And most of them are volunteering with no compensation



# Next Steps



# Next steps



## Interviews to validate hypothesis

- Interviewee identification distributed across studied dimensions
  - Minorities
  - Men with challenges
  - Average profile
- Additional interviews with groups of interest
  - Interns
  - Those who left

## Quantitative analysis

- Selection of projects to analyze
- Dashboard generation with gender info

---

---

# Diversity & Inclusion

— Apache Airflow Community —

---

---

# Apache Airflow Community

24 PMC members

31 committers

~700 Contributors in 2020

Used by 300+ companies

<https://s.apache.org/apache-way-for-everyone>

Oct 5, 2014 – Jul 6, 2020

Contributions: Commits ▾

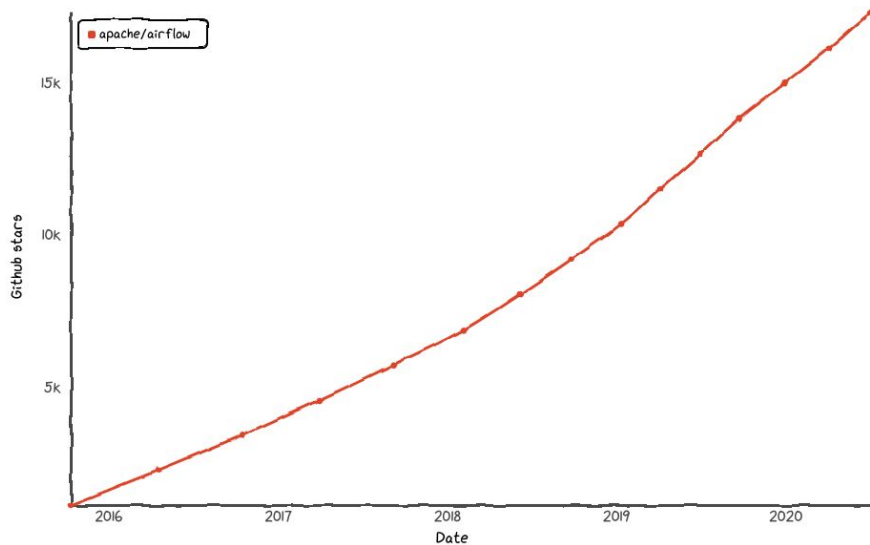
Contributions to master, excluding merge commits



apache/airflow

View star history

Star history



***“Diversity is a Metric, Inclusion is a process  
and Equity is an outcome.”***

@jesshmitchel

# Diversity metrics

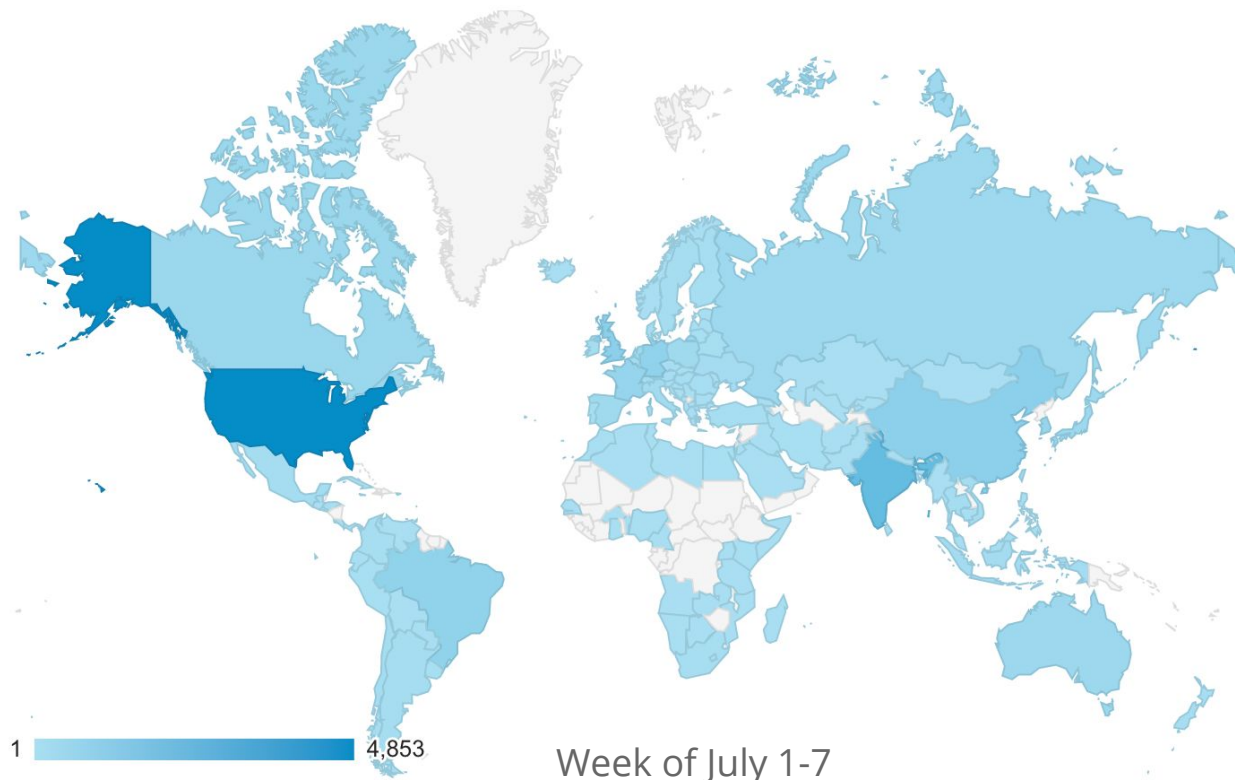
Geographic diversity of users and contributors

Diversity of vendors

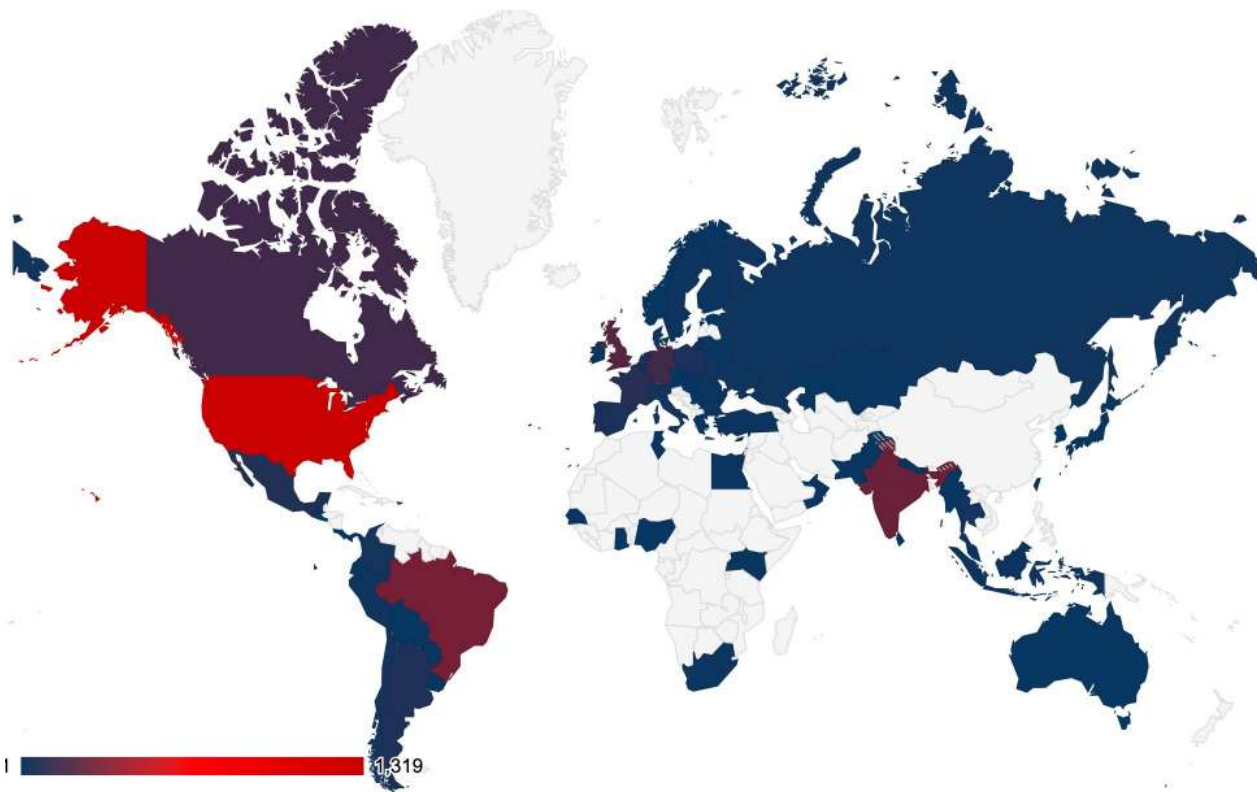
Diversity of contributions



# Airflow Website Visitors



# Airflow Summit Attendees



Aggregate geo locations of the attendees on the 1st day of Airflow Summit 2020

# Geographic Diversity of Contributors



# Inclusion Processes

- Asynchronous communication
  - Mailing list - all important discussion happen on the mailing list
  - Discussions are open for at least 72 hours
  - People are encouraged to use simple English
- ✓ Transparency - all emails are public and archived
- ✓ Allow global participation
- ✓ Knowledge sharing (no tribal knowledge!)

# Diversity of Vendors

## Contribute to Apache Airflow

- Astronomer
- Google
- Polidea
- Lyft & many more

## Use Apache Airflow: +300 companies

- AirBnB
- Lyft
- Reddit
- Tesla and many, many more:  
<https://github.com/apache/airflow>

# Inclusion Processes

- Business friendly license
  - Vendor-neutrality
    - We don't provide any statements that endorse, promote, or advertise the capability, credibility, or quality of any company's technology, product, or service.
- 
- ✓ Ensure long term sustainability of the project
  - ✓ Avoid vendor lock-in

# Diversity of Contributions

- Bug reports
- Issue triage
- Code and code reviews
- Documentation
- Design work
- Website
- Meetups
- Summits and many more

# Inclusion Processes

- Community > code
    - We value health of the community more than code
    - We strive to be inclusive to minorities
  - Recognition
    - User -> Contributor -> Committer -> PMC -> ASF member
    - All contributors earn individual merit, and merit never expires!
  - **Mentorship**
    - Google Season of Docs
    - Google Summer of Code
    - Outreachy
- ✓ Healthy interaction -> welcoming community -> new contributors
- ✓ Great communities create great software!



# Looking ahead

- ▶▶ Strive to have more contributors - diversity of perspectives, knowledge, experiences, background, etc.
- ▶▶ Create a friendlier environment for underrepresented groups
- ▶▶ Diversity analysis for Apache Airflow
- ▶▶ Please join [dev@airflow.apache.org](mailto:dev@airflow.apache.org) - and participate!



# Thank you