

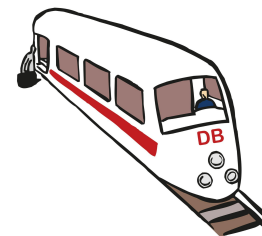
# How to increase diversity in open source projects

Dr Maren Westermann

# About me



- PhD in agricultural/ environmental science from University of Queensland
- 2 years of experience as Data Scientist
- 2 years of experience as Machine Learning Engineer
- Active contributor to open source since 2020, especially scikit-learn
- PyLadies Berlin co-organisier since 2021, especially monthly mentored “Open source hack nights”



# Diversity



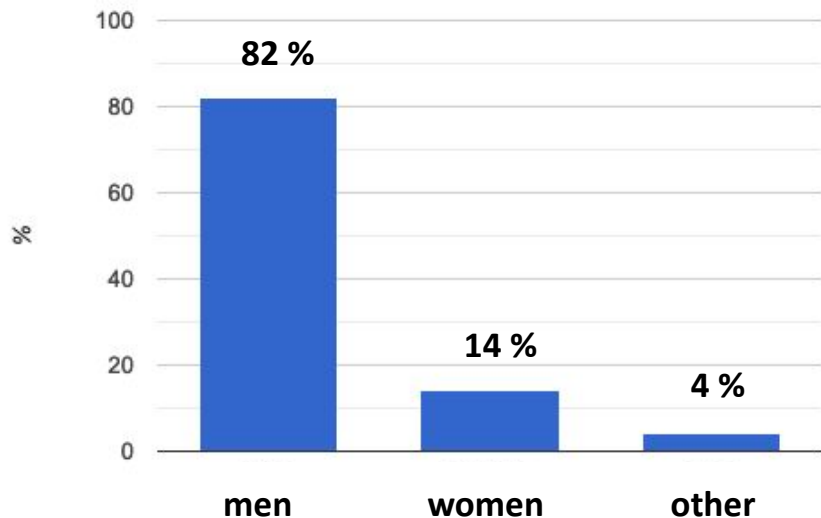
What is diversity?

“We define diversity within open source communities as a pluralism of any number of the following possibilities: gender identity and expression, race, ethnicity, sexual orientation, age, social class, caste, language, physical and neurological ability or attributes, religious beliefs, value systems, national origin, and political affiliation.”

*The Linux Foundation. Diversity, Equity, and Inclusion in Open Source, 2021*

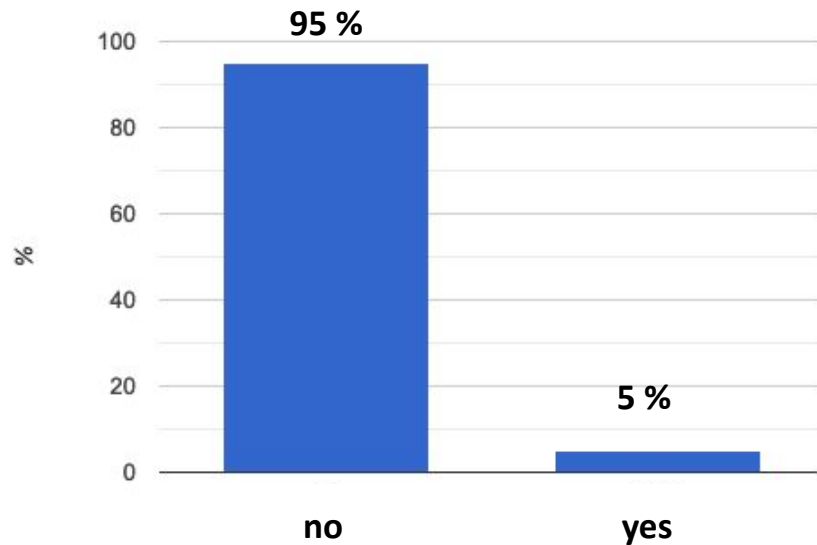
# Statistics

**Gender of contributors**



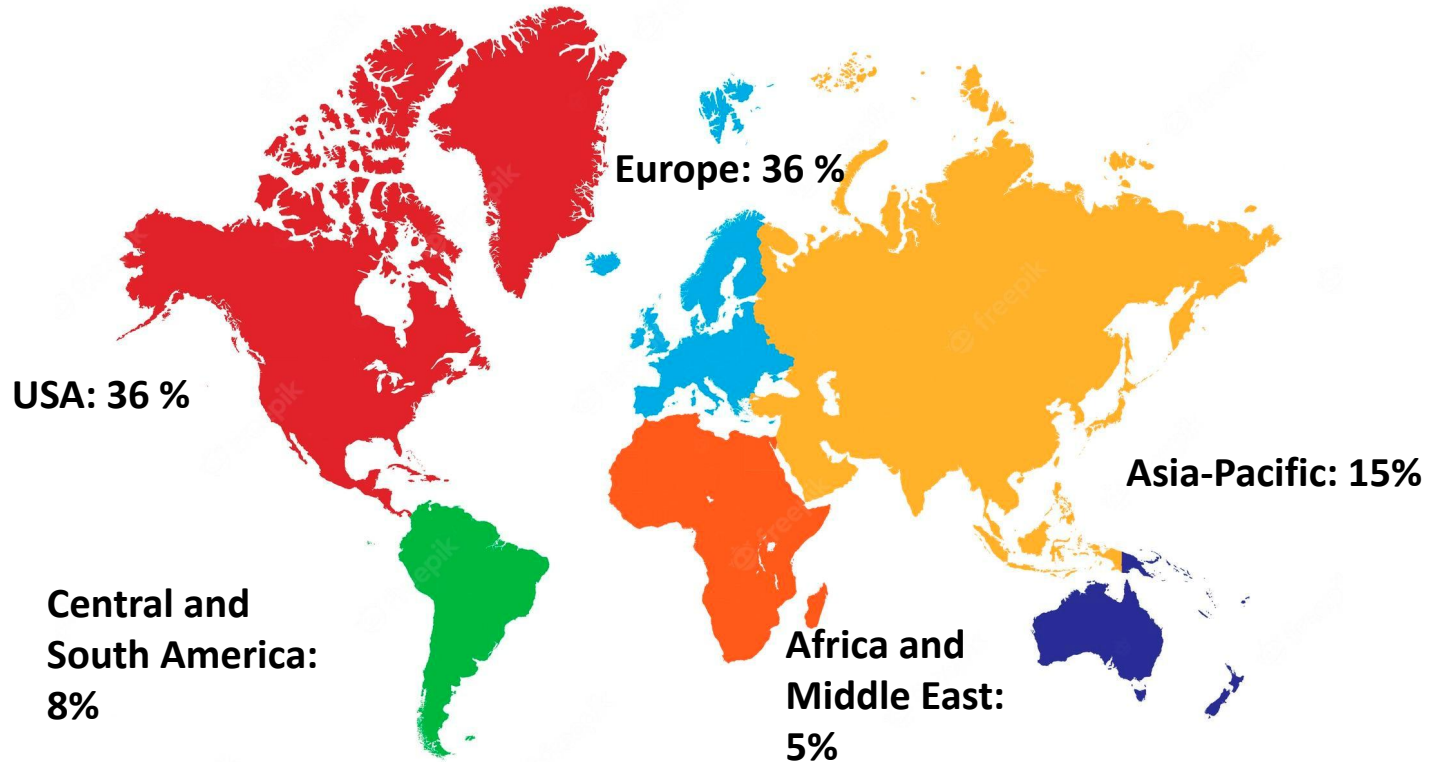
*The Linux Foundation, Diversity, Equity, and Inclusion in Open Source, 2021*

**Female core developer?**



*Trinkenreich et al. 2022*

# Statistics



# Consequences

## For underrepresented groups

- Loss of opportunities
  - Skill development
  - Career advancement

## For open source projects

- Lack of diversity of perspectives
  - Lower problem-solving capacity
  - Less robust technology
  - Lower productivity
  - Susceptibility to unhealthy work environments

# Challenges faced

## Mainly social challenges

- Lack of peer parity
- Non-inclusive communication
- Toxic culture
- Expectation issues
- Gender bias/ stereotyping
- Imposter syndrome
- Work life balance issues

# Challenges faced

## Other challenges

- Language access
- Educational access
  - Educational opportunities (hackathons, internships, courses)
- Geographic access
  - Time zones: ability to attend meetings and network
- Access to technologies (“digital divide”)
  - Reliable Internet access at home vs. elsewhere
- Economic and professional access
  - Compensation for open source contributions
  - Employers’ approval of contributing to open source projects
  - Sponsorships, mentorships



# Strategies

## Use inclusive language



“Guys”



“Master/ slave”



“Man hours”



“So simple your mother can do it”



Don’t assume someone’s gender



“Folks/ people/ everyone”



“Primary/ secondary”



“Engineer/ personal hours”



“User-friendly”




Ask for people’s pronouns

# Strategies

Have a good “Contributing to” documentation and keep it up to date

DOC: "Creating a Python environment" in "Creating a development environment" needs update #48060

Closed 1 task done marenwestermann opened this issue on Aug 12, 2022 · 25 comments · Fixed by #49914

 marenwestermann commented on Aug 12, 2022 · edited Contributor

**Pandas version checks**  
☒ I have checked that the issue still exists on the latest versions of the docs on [main](#) [here](#)

**Location of the documentation**  
[https://pandas.pydata.org/docs/dev/development/contributing\\_environment.html#creating-a-python-environment](https://pandas.pydata.org/docs/dev/development/contributing_environment.html#creating-a-python-environment)

**Documentation problem**  
My machine is a MacBookPro with chip Apple M1 Pro, I'm using conda, and I created an environment without Docker. I checked my version of `xcode-select` and it's 2395 which seems to be the latest version. I followed the instructions in "Creating a Python environment". When I ran `python -m pip install -e . --no-build-isolation --no-use-pep517` I got the following output with the following errors:  

```
(pandas-dev) + pandas git:(main) python -m pip install -e . --no-build-isolation --no-use-pep517
Looking in indexes: https://pypi.org/simple, https://maren.westermann:****@repos.mytaxi.com/artifactory/api/py/
Obtaining file:///Users/maren/open_source/pandas
Preparing metadata (setup.py) ... done
```

Assignees  
No one assigned

Labels  
[Docs](#) [good first issue](#)


Projects  
None yet

Milestone  
No milestone

Development  
Successfully merging a pull request  
Issue  
 **DOC: update running the t**  
marenwestermann/pandas

Add documentation about re-building extensions to "Installing the development version of scikit-learn" #25985

Open marenwestermann opened this issue 3 weeks ago · 8 comments

 marenwestermann commented 3 weeks ago Contributor

**Describe the issue linked to the documentation**  
Quite regularly when contributing to scikit-learn I realise that I need to re-build the Cython extensions. I realise this when I run `pytest` and get an `ImportError`. When I started contributing, I didn't know about this and I would have been rather lost when I faced this error the first time if @adrinjalali didn't tell me how to fix this. As many people, especially from underrepresented groups in tech, will face the same issue, a note about re-building the extensions should be put in the documentation. The pandas team for example now recommends to their contributors to do this after each fetch and merge from `upstream/main` : [https://pandas.pydata.org/docs/dev/development/contributing\\_environment.html#step-3-build-and-install-pandas](https://pandas.pydata.org/docs/dev/development/contributing_environment.html#step-3-build-and-install-pandas)

**Suggest a potential alternative/fix**

Assignees  
No one assigned

Labels  
[Documentation](#)

Projects  
None yet

Milestone  
No milestone

# Strategies

Have contributing documentation in different languages (e.g. Spanish)

## Ejemplo de contribución y envío de un Pull Request a [scikit-learn](#)

---



## Transcripción del Video

---

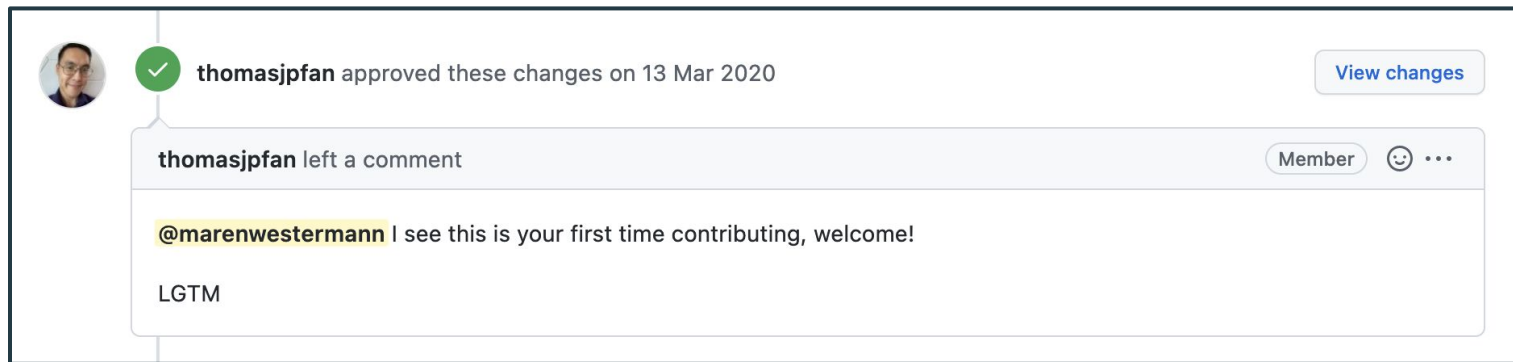
- Ponente: [Reshama Shaikh](#)
- Video: [Scikit-learn PR Contributing Example](#) (30 minutes)
- Transcritora: [Reshama Shaikh](#)

# Strategies



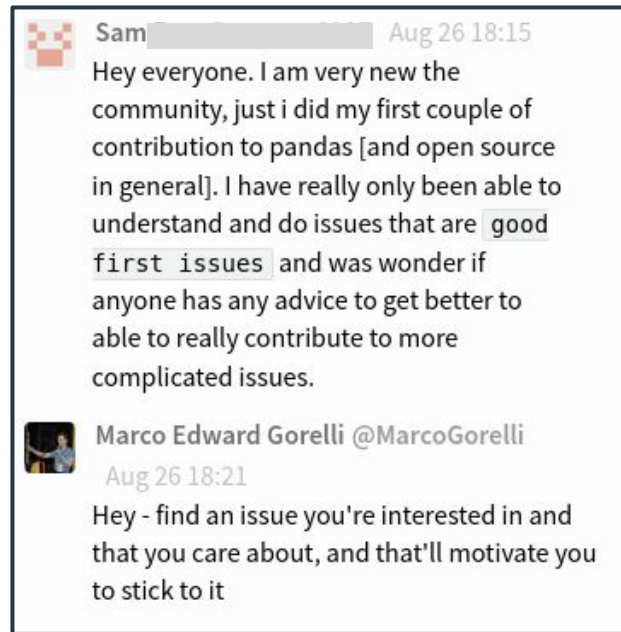
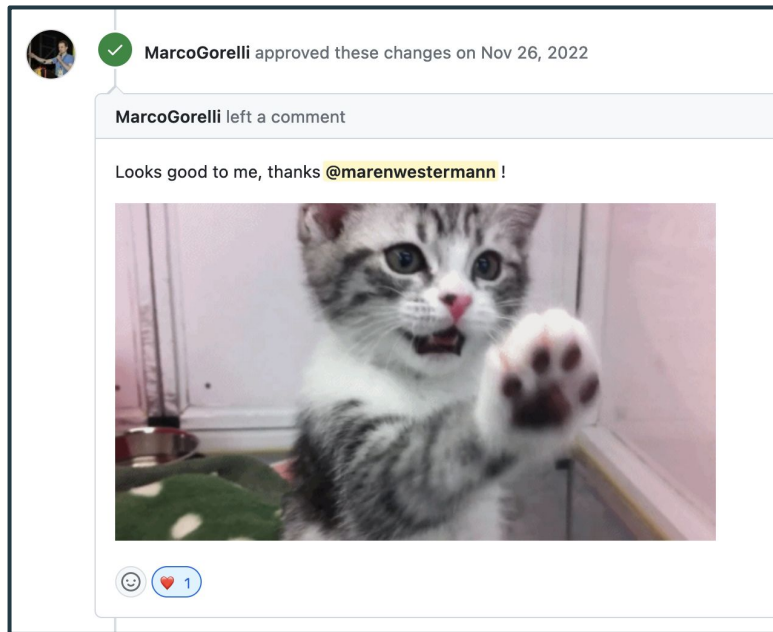
## Encourage and be welcoming

- “Thanks a lot for your contribution, this is a valuable addition to the project.”
- “We really value this kind of contribution.”



# Strategies

Encourage and be welcoming



# Strategies

## Promote specific groups and events

- Run events through specific groups, e.g.
  - scikit-learn: WiMLDS Paris sprint 2022
  - pandas: PyLadies London hack evenings
  - PyLadies Berlin: open source hack nights
- Social media: like and share
- Run events in schools and universities and give priority to people from underrepresented groups
- Run online events targeted at people from geographically underrepresented regions



# Strategies

## Promote awareness of the presence of peers

- At conferences: make sure you have speaker diversity
- Promotion on social media
- Write blog articles, e.g. about sprints, contributor interviews
- Recognize people's achievements



# Strategies

## De-stereotype the OSS contributor

- Don't automatically assign community building tasks to underrepresented groups; spread the workload



“women developers”



“developers”

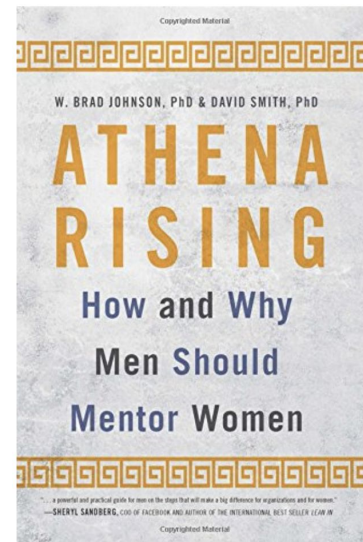


# Strategies

## Mentorship

“Invaluable to improving diversity and inclusion in open source communities” - *Mentorship program leaders at the Linux Foundation*

- Lower barriers to entry and onboarding
- Networking
- Help junior contributors being able to establish themselves
- Career development



# Strategies

## Mentorship



PyLadies Berlin • You

--

2w • 🌐

The [scikit-learn](#) team has an exciting mentored, paid internship opportunity coming up for people belonging to underrepresented groups in tech! 🥳 Check out this amazing opportunity which could be your path to becoming a core contributor to the project: ➡️ <https://lnkd.in/exT6zWp6>

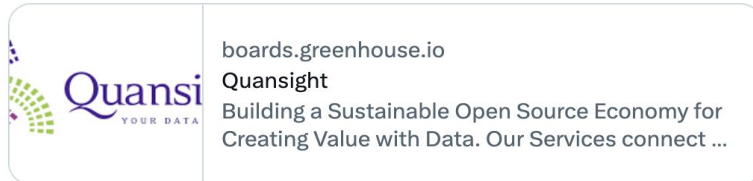
Key points:

- available to people from underrepresented groups in tech
- mentored internship
- duration: 5 months
- 2000 EUR gross payment per months
- location: Berlin, Germany



🌟 Tania Allard 🇮🇹🇬🇧 @ixek • Mar 10

I am happy to share what the applications for our paid [#opensource](#) internships at [@quansightai](#) are now open 🎉  
This is a great opportunity for students, recent graduates, career changers, and tech returners to work on [#opensource](#) projects for 3 months



💬 3

↺↻ 41

❤️ 46

📊 6,811



# Strategies

## Governance

- Have governance board
  - Should be representative of different demographics
- Transparency and accountability around decision making
  - How are leaders determined?
  - How are contributors onboarded, managed, and supported?
- Transparent career pathways
- Promote people from underrepresented groups to leadership roles
- value non-coding contributions to the project
- Try incremental changes over time

# Strategies

## Create and enforce a Code of Conduct

- A Code of Conduct
  - can signal to marginalized people that the open source community cares
  - provides guidance for newcomers
  - provides reference for core values and norms
- Have a dedicated CoC team or hire people who follow up on Code of Conduct breach reports

# Further Reading

Bianca Trinkenreich



ABOUT

PUBLICATIONS

AWARDS

EXPERIENCE

EDUCATION

SERVICE

## PUBLICATIONS

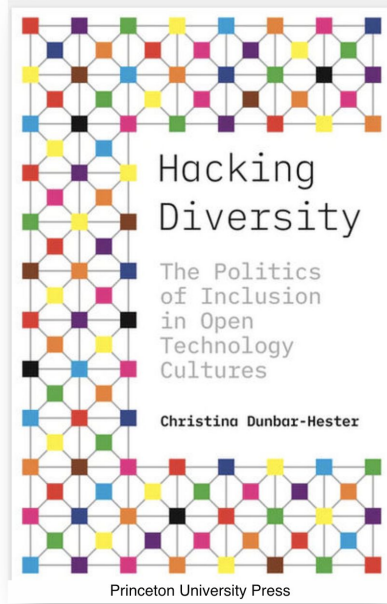
### JOURNAL PAPERS

- 📖 **Women's Participation in Open Source Software: A Survey of the Literature**  
Trinkenreich, B., Wiese, I., Gerosa, M., Sarma, A., Steinmacher, I.  
Accepted at *Transactions of Software Engineering and Methodology*, 2022.
- 📖 **Pots of Gold at the End of the Rainbow: What is Success for Open Source Contributors?**  
Trinkenreich, B., Guizani, G., Wiese, I., Conte, T., Gerosa, M., Sarma, A., Steinmacher, I.  
Accepted at *Transactions of Software Engineering*, 2021.
- 📖 **Being a mentor in Open Source projects**  
Steinmacher, I., Balali, S., Trinkenreich, B., Guizani, M., Izquierdo-Cortazar, D., Zambrano, G., Gerosa, M., Sarma, A.  
Accepted at *Journal of Internet Services and Applications*, 2021.
- 📖 **SINIS: A GQM+Strategies-based Approach for Identifying Goals, Strategies and Indicators for IT Services**  
Trinkenreich, B., Santos, G., M. P. Barcellos  
Accepted at *Information and Software Technology*, v. 100, p. 147-164, 2018.
- 📖 **Evaluation of measurement process on a Global Organization under the light of MR-MPS-SV maturity model (in Portuguese)**  
Trinkenreich, B., Santos, G.  
Accepted at *iSys Brazilian Journal of Information Systems*, v. 8, issue 2: p. 58-77, 2015

[biancatrink.github.io](https://biancatrink.github.io)

# Further Reading

Christina Dunbar-Hester



# References

- Trinkenreich, B. et al, 2022. Women's Participation in Open Source Software: A Survey of the Literature. Manuscript submitted to ACM
- Carter, H. et al, 2021. Diversity, Equity, and Inclusion in Open Source. The Linux Foundation
- Geiger, R. S., 2017. Summary analysis of the 2017 github open source survey
- Zlotnick, F. et al, 2017. The open source survey. GitHub repository, 2017
- Finley, K., 2017. Diversity in Open Source Is Even Worse Than in Tech Overall, Wired Magazine
- Weber Mendonça, M., 2022. Beyond the basics: Contributor experience, diversity and culture in Open Source Projects. Keynote speech at PyConDE and PyData 2022
- Steinmacher, I. et al, 2021. Being a Mentor in open source projects. J Internet Serv Appl 12, 7

# Thank you

Let's connect :)

- Mastodon: [@maren@fosstodon.org](mailto:@maren@fosstodon.org)
- Twitter: [@MarenWestermann](https://twitter.com/MarenWestermann)
- LinkedIn: [de.linkedin.com/in/dr-maren-westermann-0b8575144](https://de.linkedin.com/in/dr-maren-westermann-0b8575144)
- GitHub: [github.com/marenwestermann](https://github.com/marenwestermann)