Beyond the 10%
Analysis of the gender-diversity gap

Python DevRoom @ FOSDEM

Daniel Izquierdo Cortázar
@dizquierdo
https://speakerdeck.com/bitergia
All started in Tokyo...
Extension of previous analysis (focus on CPython)

Lack of existing data

Transparency and improvement

Learn from best practices

Ethical and business perspective

Involved in CHA OSS D&I
Python diversity statement:

“some of these attributes include (but are not limited to): age, culture, ethnicity, gender identity or expression, national origin, physical or mental difference, politics, race, religion, sex, sexual orientation, socio-economic status, and subculture. We welcome people regardless of the values of these or other attributes.”
11% - openstack
10% - Linux
8% - Hadoop
4% - Python

More data at speakderdeck.com/bitergia
$ perceval git
"https://github.com/chaoss/grimoirelab-perceval.git"

> [2019-02-01 23:36:59,096] - Sir Perceval is on his quest.

> [2019-02-01 23:36:59,102] - Fetching commits:
'https://github.com/chaoss/grimoirelab-perceval.git' git repository from 1970-01-01 00:00:00+00:00; all branches
"backend_name": "Git",
"backend_version": "0.10.2",
"category": "commit",
"data": {
    "Author": "Alvaro del Castillo <acs@bitergia.com>",
    "AuthorDate": "Mon Apr 25 00:47:22 2016 +0200",
    "Commit": "Santiago Dueñas <sduenas@bitergia.com>",
    "CommitDate": "Thu Apr 28 01:31:50 2016 +0200",
    "commit": "babb8f761a436ea00cac90957afc3d38f0bdc946",
    "files": [
        {
            "action": "M",
            "added": "22",
            "file": "perceval/backends/github.py",
            "message": "[github] Check and control rate limit in GitHub API queries\nThe backend will check the rate limit. When this rate is consumed, it will sleep until the rate is reset."
        }
    ],
    "origin": "https://github.com/chaoss/grimoirelab-perceval.git"
}
Name
Email
Nickname
Country
Other identities
Gender

https://github.com/chaoss/grimoirelab-sortinghat
Genderize.io as external API

$ curl https://api.genderize.io/?name=peter

> 

{"name":"peter","gender":"male","probability":1,"count":4373}
More context: 2004: ~0%

2014: 2.21%

Last year: 4.36%
Community last year activity

Summary

8,323 # Commits
872 # Authors
40 # Repositories
Women last year activity

Summary

316 # Commits
38 # Authors
22 # Repositories
Community evolution over time

Commits by gender over time

- Commit made by men
- Commits made by women
- Unknown Commits
Geographical distribution last year active people
Geographical distribution last year active women
Last year network analysis of github.com/python
Last year women network analysis of github.com/python
Women as newcomers in github.com/python
Gender is not binary

Focused on github.com/python

Small fraction of what diversity is

Only focused on quantitative data

Ethics and Accountability
Forget about the numbers! Clear issue in the industry

Glass ceiling of 10%

Diversity & Inclusion is a challenge

[Permanent & Updated] Data can help to be aware and lead a change

Data and tech. are just a tool to achieve our goals
Cross Foundations Initiative

Can we learn from others?

Recommendations from OpenStack

- Policies impact study
- Collaboration with tech. leaders
- Bring women to key positions
- Keep supporting the WOO group
- Enforce the CoC
Community Health Analytics for Open Source Software

GMD and Diversity and Inclusion working groups

Diversity is a key factor for healthy communities
OpenStack gender diversity report as example

Gender-diversity and mentorship: https://bit.ly/2O08p3j

Focus areas:

Event diversity
Contributor community diversity
Communication inclusivity
Recognition of good work
Leadership
Governance
Project places.
Join the **D&I Mailing List** and introduce yourself

Look through the D&I mail list archive for past discussions and self-introductions of people involved.

Add to your calendar our weekly meetings on Mondays at 7:30 am PT / 4:30 CET
To be continued ...
Let’s go for questions!

Daniel Izquierdo Cortázar
dizquierdo@bitergia.com
@dizquierdo