Social Diversity on GitHub

Bogdan Vasilescu & Vladimir Filkov • Computer Science • University of California, Davis
Alexander Serebrenik • Computer Science • Eindhoven University of Technology
vasilescu@ucdavis.edu • filkov@cs.ucdavis.edu • a.serebrenik@tue.nl

How does team diversity relate to team effectiveness on GitHub?

Types of Diversity

Gender diversity = mix women/men

Tenure (experience) diversity = mix junior/senior

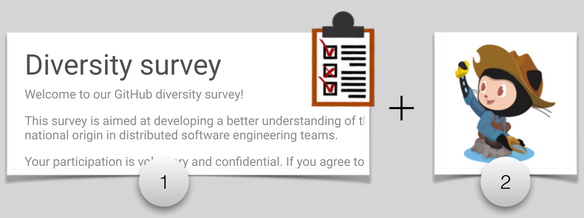
Cultural diversity = mix countries

Theory

Varied backgrounds and ideas provide access to broader information and enhanced problem solving skills.

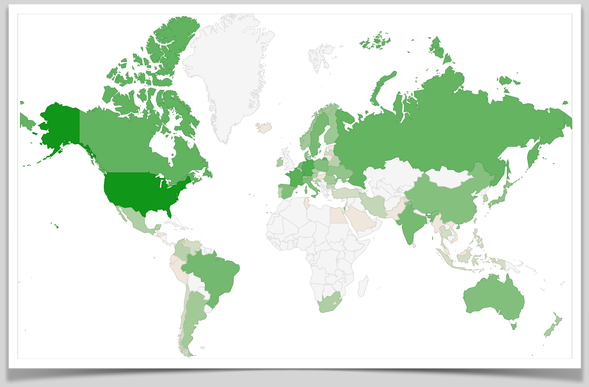
In more diverse teams, members are more likely to engage in stereotyping, cliquishness, and conflict.

Approach: Mixed Methods



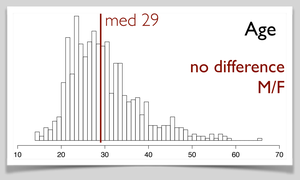
1. **Perceptions of Diversity on GitHub: A User Survey**. B. Vasilescu, V. Filkov, A. Serebrenik, *CHASE 2015*
2. **A Data Set for Social Diversity Studies of GitHub Teams.** B. Vasilescu,, A. Serebrenik V. Filkov. *MSR Data 2015*

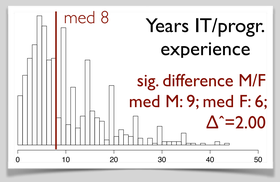
1 Survey: 4,500 invites, 816 responses



72 countries

24% female; 75% male; 1% other





Q: Whom do you consider part of your team?

- The repository owner and others who can push directly
- People who contribute code frequently
- People who work on my particular feature/branch

less inclusive ⇕ more inclusive

#1 (72%) • **Everyone who does something in this repository**

Q: Which of the following characteristics of your team members are you aware of?

- Programming skills 74%
- **Gender 48%**
- Real name 45%
- Social skills 42%
- Country of residence 40%
- Personality 39%
- Reputation as programmer 31%
- Ethnicity 30%
- Employment 30%
- GitHub experience 28%
- Educational level 26%
- Age 23%
- Hobbies 11%
- Political views 4%

Q: Experiences working in a diverse team

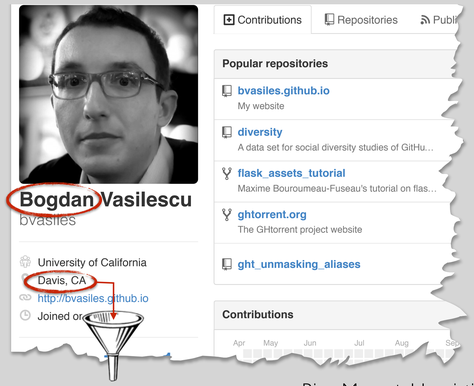
"code sees no color or gender" — Meritocracy; no effects of diversity

"diverse viewpoints often lead to **lively discussions and new ideas**" — Positive effects of diversity

"I have used a **fake GitHub handle** (my normal GitHub handle is my first name, which is a distinctly female name) **so that people would assume I was male**" — Negative effects of diversity

2 Mining GitHub

Infer Gender



USA + Bogdan → male

Bing Maps + Heuristics
http://github.com/tue-mdse/countryNameManager

Name frequency tables for 30 countries
http://github.com/tue-mdse/genderComputer

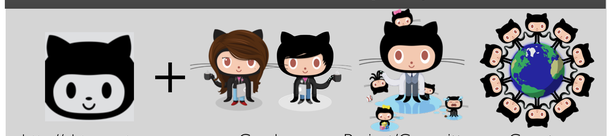
Merge Aliases

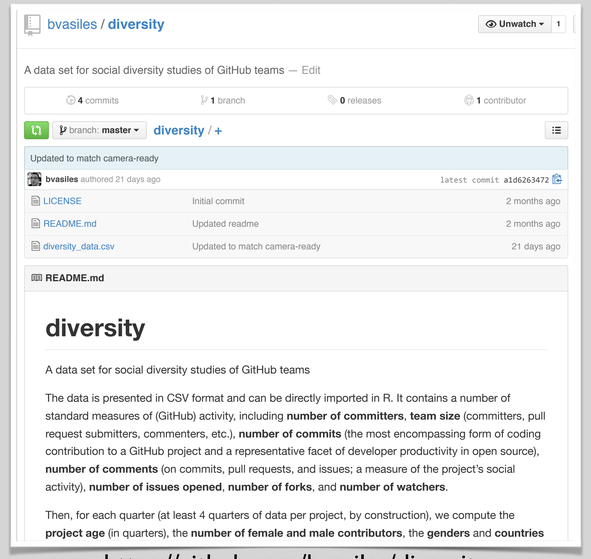
Leon Nathaniel Maurer — leon.maurer@gmail.com

Leon Maurer — leon.maurer@gamil.com

Leon Maurer — leonmaurer@leon-maurers-macbook-pro.local

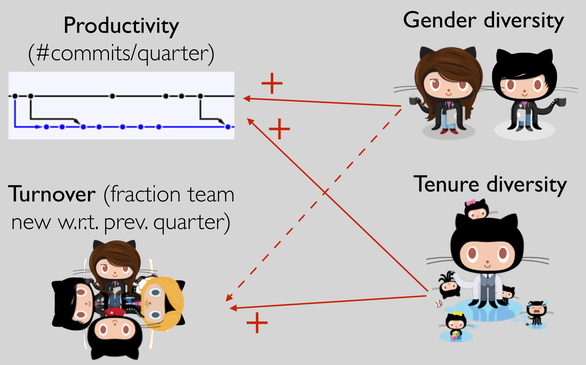
Data Set: 23,493 projects





https://github.com/bvasiles/diversity

Example: Multivariate Regression



References

- **Gender and Tenure Diversity in GitHub Teams**. B. Vasilescu, D. Posnett, B. Ray, M. vdBrand, A. Serebrenik, P. Devanbu, V. Filkov, *CHI 2015, pages 3789-3798. ACM*
- **Gender, representation and online participation: A quantitative study.** B. Vasilescu, A. Capiluppi,, A. Serebrenik, *Interacting with Computers 26, 5 (2014), 488–511*

Octocats from: https://octodex.github.com/