

Gender and GitHub

Alexander Serebrenik
@aserebrenik



Nicolaas Pietersz. Berchem (1620 - 1683). Harvest, KMSKB, Brussel

MS Windows post-release fault prediction	Precision: predicted & correct / predicted	Recall: predicted & correct / correct
Code churn	79%	80%
Code complexity	79%	66%
Code coverage	84%	55%
Code dependencies	74%	70%
Organizational structure	86%	84%
Socio-technical network	77%	71%

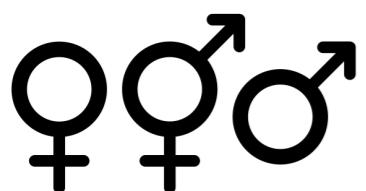
Nagappan, N., Ball, T.: Evidence-based failure prediction.

In: Oram, A., Wilson, G. (eds.) Making Software: What Really Works and Why We Believe it? O'Reilly 2011

Individual

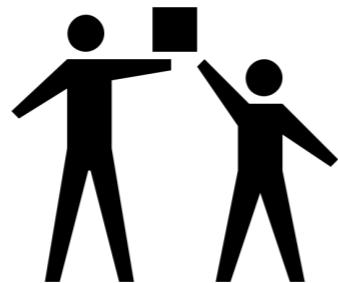


Created by Pham Duy Phuong Hung
from Noun Project
tenure



Created by Jonathan Li
from Noun Project
gender

Team

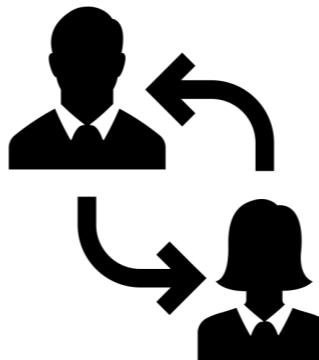


Created by Bruno Castro
from Noun Project
*tenure
diversity*

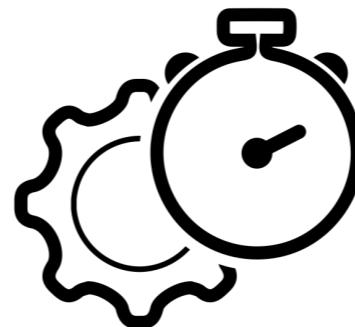


Created by Magicon
from Noun Project
*gender
diversity*

Process



Created by Gregor Cresnar
from Noun Project
turnover



Created by Delwar Hossain
from Noun Project
productivity



Created by Allen Wang
from Noun Project
*comm. smell:
black cloud*

Product



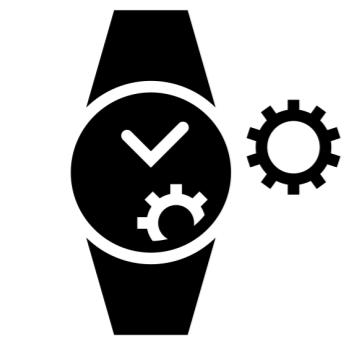
Created by mark
from Noun Project
*code smell:
long method*



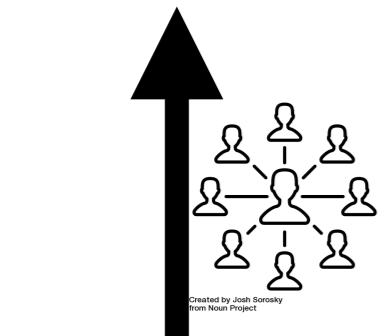
Diversity in Teams: Bad or Good?



Individual



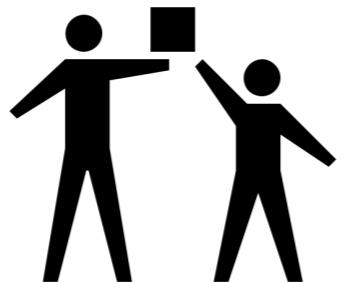
tenure



gender

Created by Jonathan Li
from Noun Project

Team

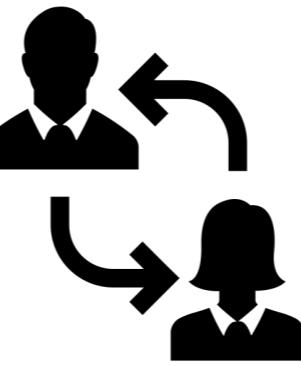


*tenure
diversity*

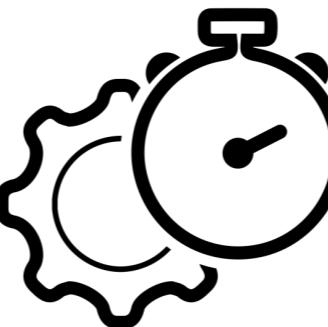


*gender
diversity*

Process



Created by Gregor Cresnar
from Noun Project
turnover

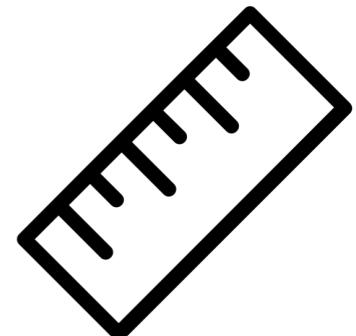


Created by Delwar Hossain
from Noun Project
productivity



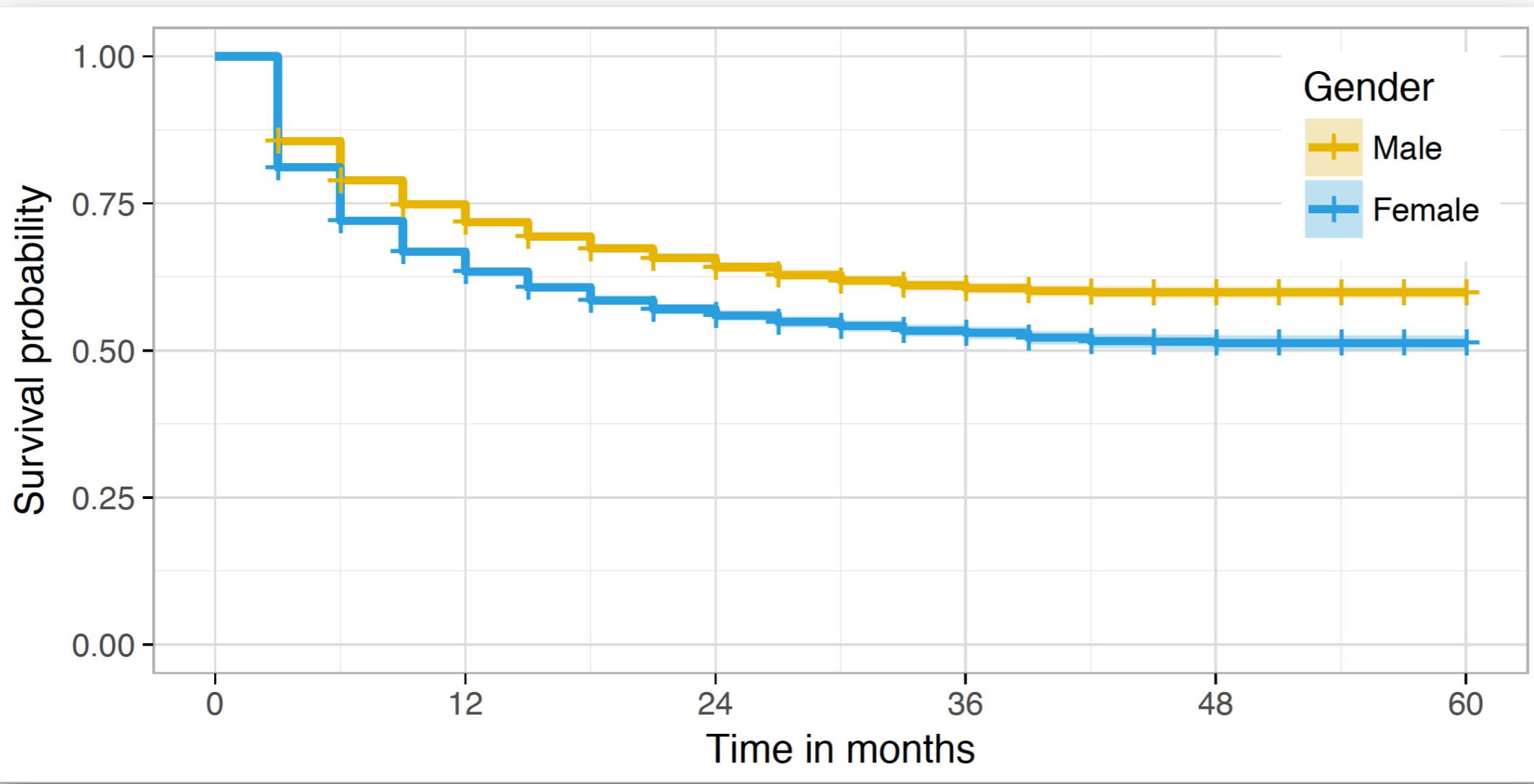
Created by Allen Wang
from Noun Project
*comm. smell:
black cloud*

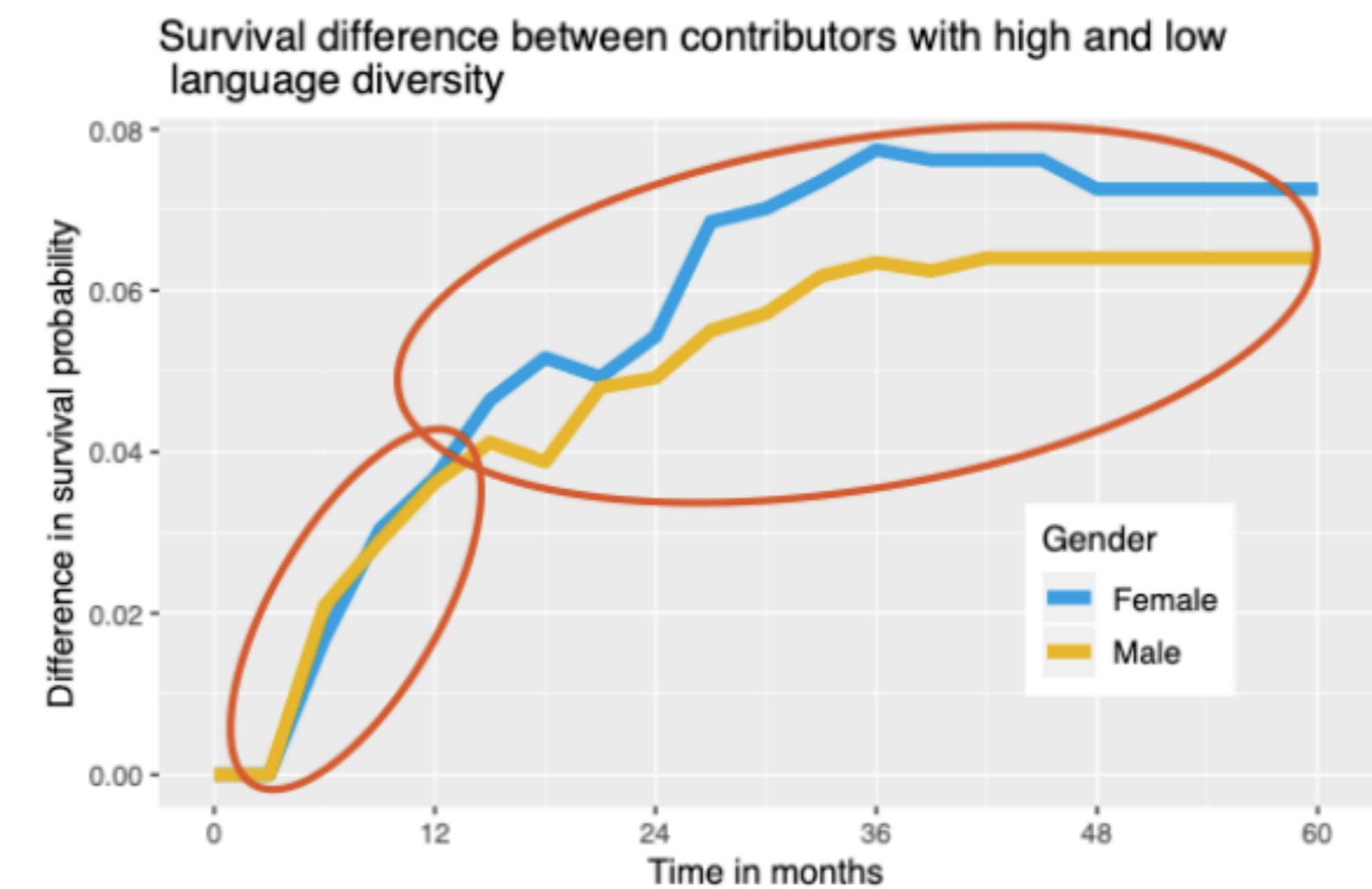
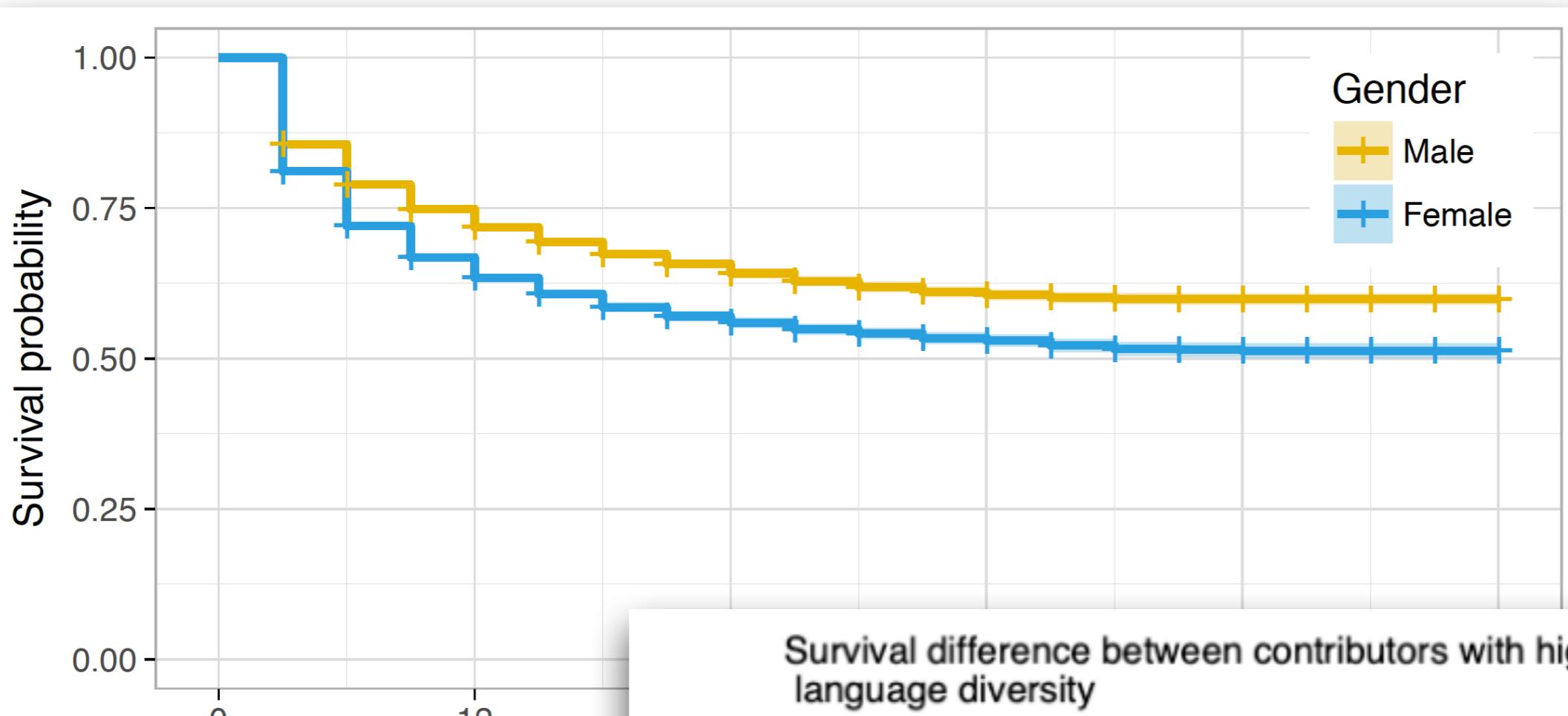
Product

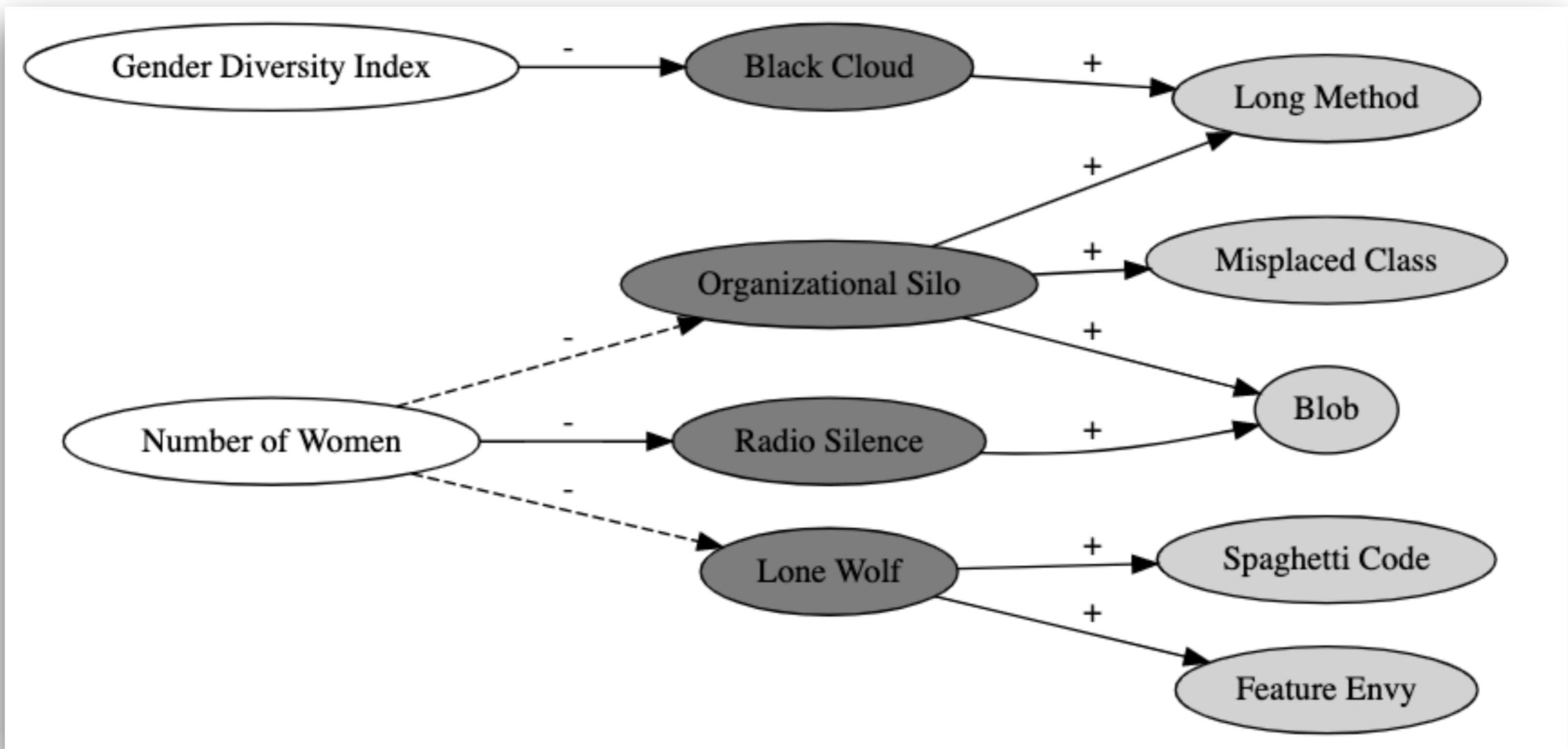


Created by mark
from Noun Project

*code smell:
long method*





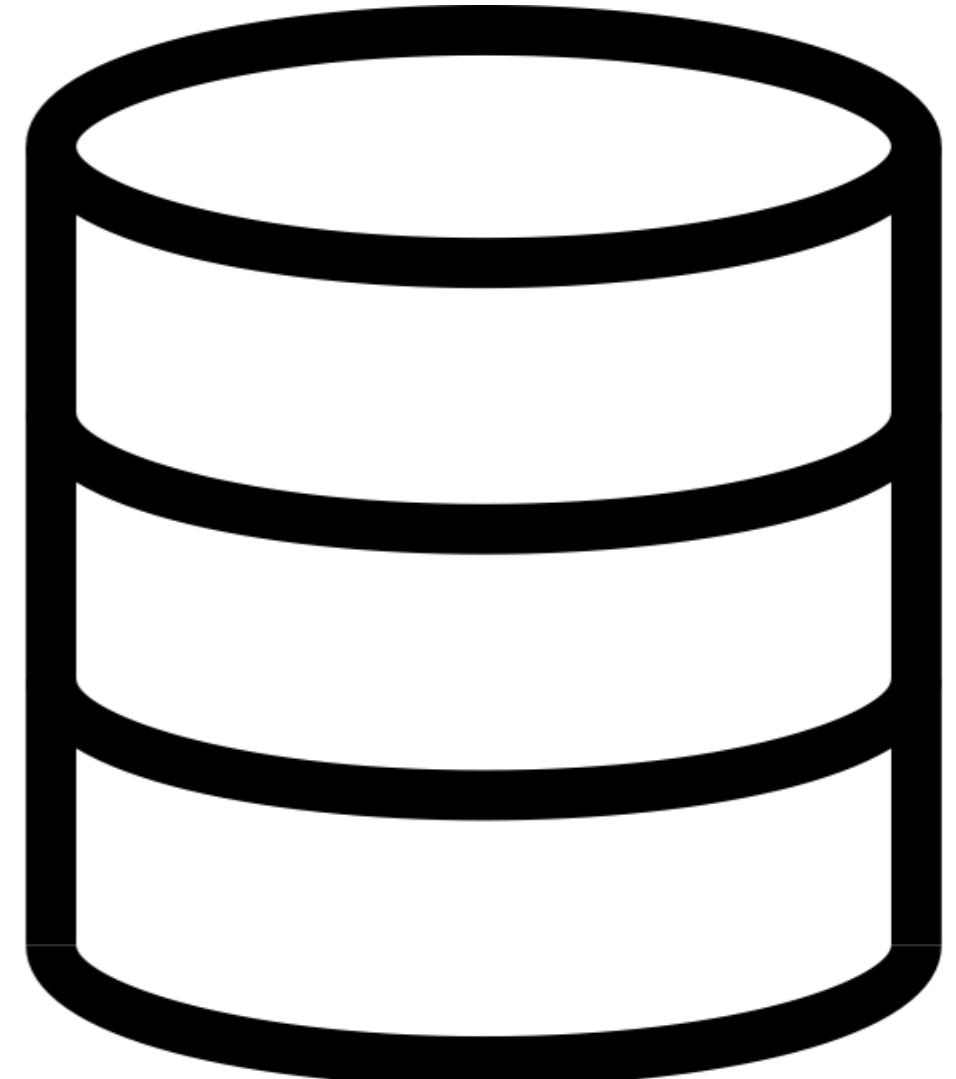


Catolino, Palomba, Tamburri, Serebrenik, Ferrucci. Gender Diversity and Women in Software Teams: How Do They Affect Community Smells? ICSE SEIS, 2019, pp. 11-20
 Palomba, Tamburri, Serebrenik, Zaidman, Arcelli Fontana, Oliveto. Beyond Technical Aspects: How Do Community Smells Influence the Intensity of Code Smells? IEEE Transactions on Software Engineering, 2018

How do we do it?

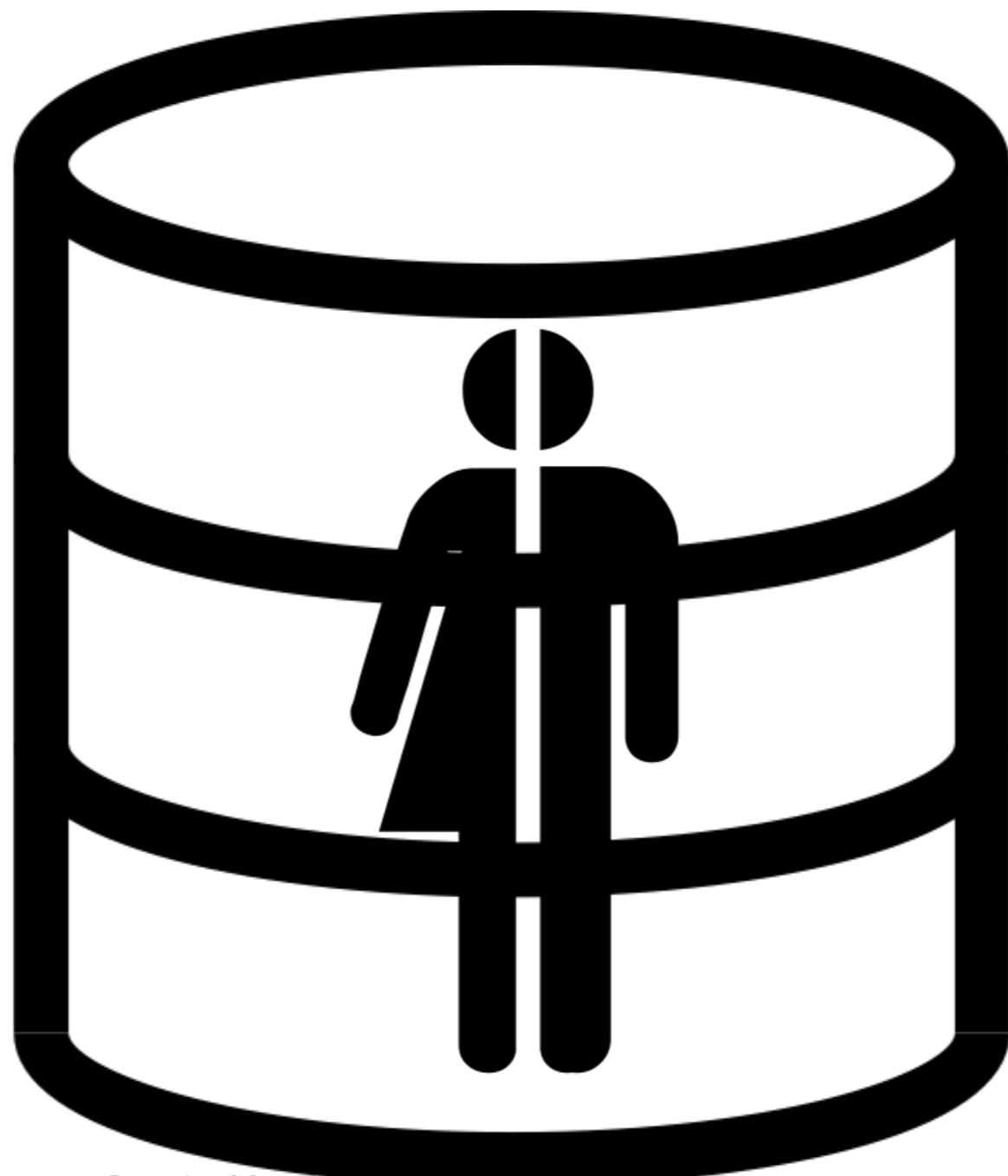


Created by Gan Khoon Lay
from Noun Project



Created by Evgenii Likhachov
from Noun Project

But... gender?



Created by Andrew J. Young
from Noun Project

Created by Evgenii Eikhachov
from Noun Project



2004

Asking Questions

The Definitive Guide to
Questionnaire Design—
For Market Research,
Political Polls, and
Social and Health
Questionnaires

What is your/(NAME's) sex?

Male Female

0.7-0.9%

0.7-0.9%

x²

A1. Are you ... ?

1 Male

2 Female

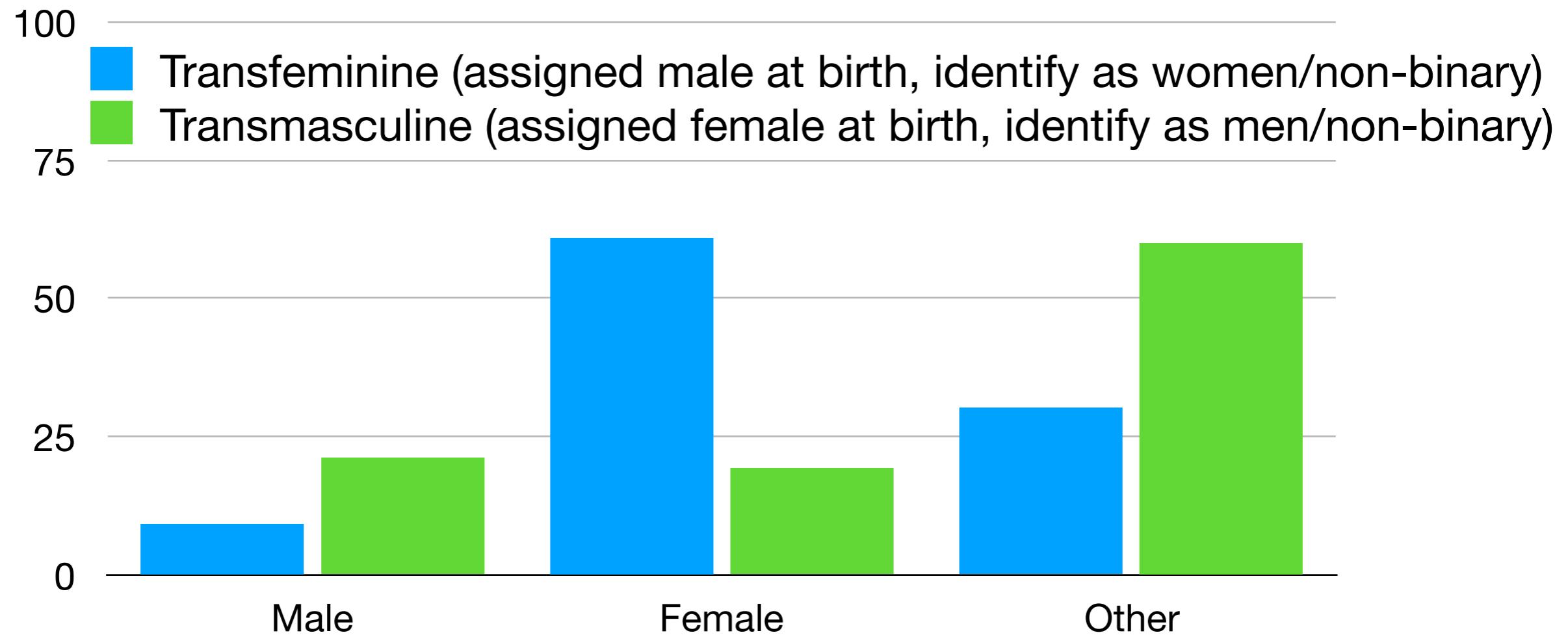
3 Something else, specify: _____

A1. Are you ... ?

1 Male

2 Female

3 Something else, specify: _____



Bauer GR. Making sure everyone counts: considerations for inclusion, identification, and analysis of transgender and transsexual participants in health surveys. In: Coen S, Banister E, editors. What a difference sex and gender make. Vancouver: Institute of Gender and Health, Canadian Institutes of Health Research; 2012. pp. 59–67.

Bauer GR, Braimoh J, Scheim AI, Dharma C (2017) Transgender-inclusive measures of sex/gender for population surveys: Mixed-methods evaluation and recommendations. PLoS ONE 12(5): e0178043.

Do you consider yourself to be transgender?

- Yes
- No
- Questioning

Do you consider yourself to be gender non-conforming, gender diverse, gender variant, or gender expansive?

- Yes
- No
- Questioning

Are you intersex?

- Yes
- No
- I don't know

Where do you identify on the gender spectrum (check all that apply)?

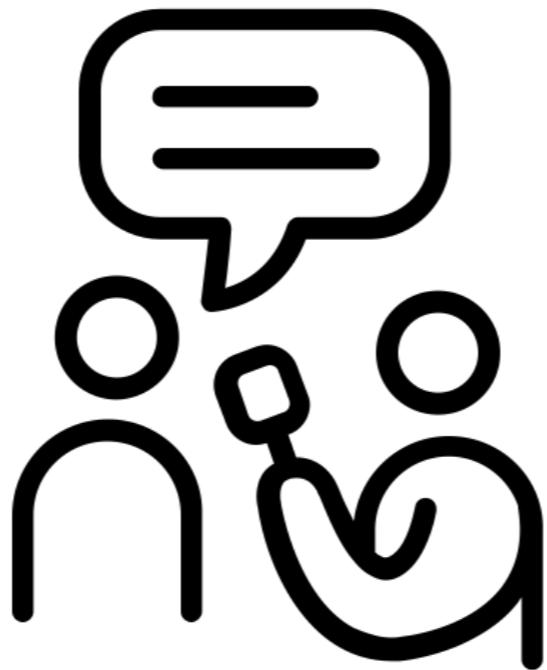
- Woman
- Demi-girl
- Man
- Demi-boy
- Non-binary
- Demi-non-binary
- Genderqueer
- Genderflux
- Genderfluid
- Demi-fluid
- Demi-gender
- Bigender
- Trigender
- Two-Spirit
- Multigender/polygender
- Pangender/omnigender
- Maxigender
- Aporagender
- Intergender
- Maverique
- Gender confusion/Gender f*ck
- Gender indifferent
- Graygender
- Agender/genderless
- Demi-agender
- Genderless
- Gender neutral
- Neutrois
- Androgynous
- Androgynie
- Prefer not to answer
- Self Identify:

Open demographics

Where do you identify on the gender spectrum?

Your answer

<https://www.morgan-klaus.com/sigchi-gender-guidelines>

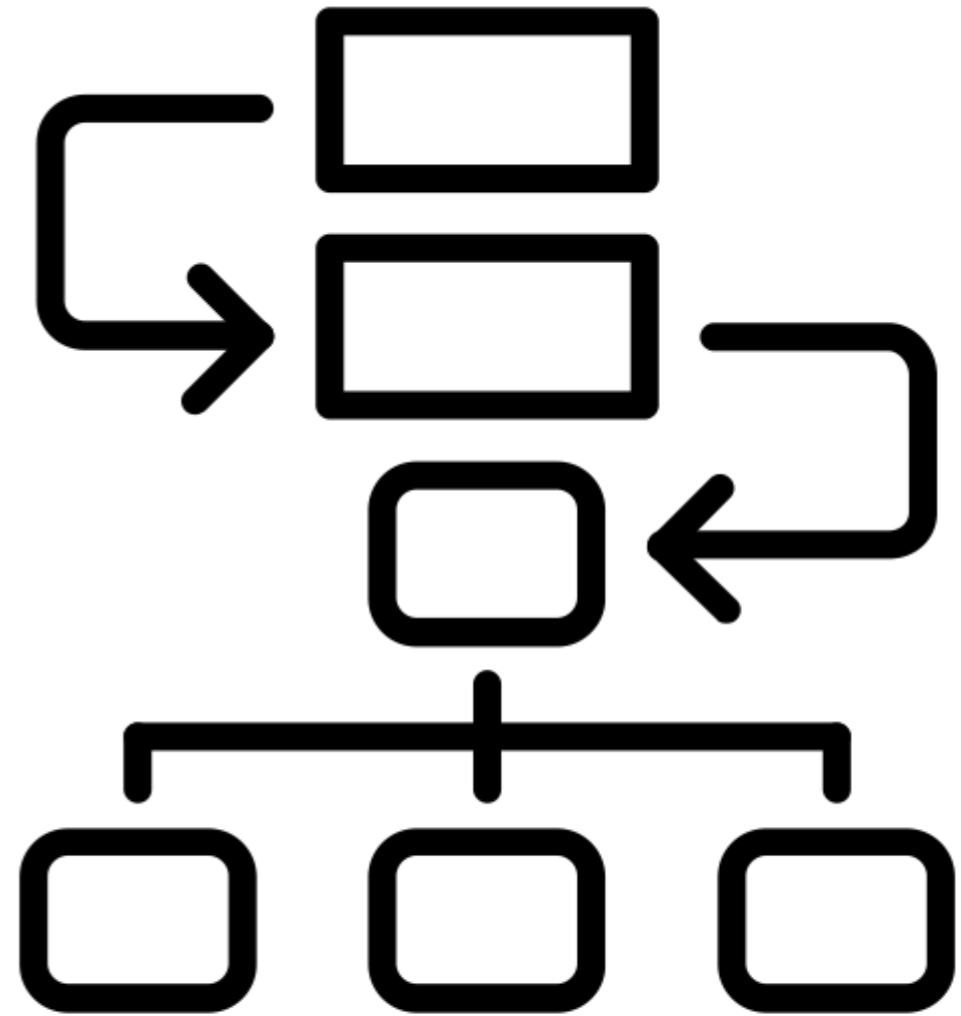


Created by Yu Luck
from Noun Project

10-20%



Created by Yu Luck
from Noun Project



Created by QualityIcons
from Noun Project

Self-presentation

+

Artefacts created

⇒

Gender









Andrea



Alexander Serebrenik top 12% overall

Associate Professor Software Evolution at Eindhoven University of Technology

Measuring and mining



3,222 REPUTATION

• 2 • 11 • 31

• Eindhoven University of Technology

• Netherlands

✉ a.serebrenik tue.nl

⌚ http://www.win.tue.nl/~aserebre/

137 answers 7 questions ~193k people reached

Eindhoven, Netherlands

aserebrenik

aserebrenik

win.tue.nl/~aserebre

Member for 7 years, 2 months

698 profile views

Last seen 5 secs ago

Visited 822 days, 2 consecutive



The screenshot shows a Google+ post. At the top left is the Google+ logo. To its right is a dark rectangular box containing the text 'Gender Guesser' with a star icon, and 'montanaflynn' with a created date of 'Created: March 2013'. Below this is a subtitle: 'Guess whether a first name is of male or female gender in the USA.' The main content of the post is a large, semi-transparent text box containing the words 'gender Computer' on top, followed by 'Josh Terren' and 'et al.' at the bottom.

Gender Guesser ★

montanaflynn Created: March 2013

Guess whether a first name is of male or female gender in the USA.

gender
Computer

Josh Terren
et al.

Different data sets require different techniques

Dr Andrea Capiluppi

Lecturer

Wilfred Brown Building 114

Email: andrea.capiluppi@brunel.ac.uk

Tel: +44 (0)1895 267485

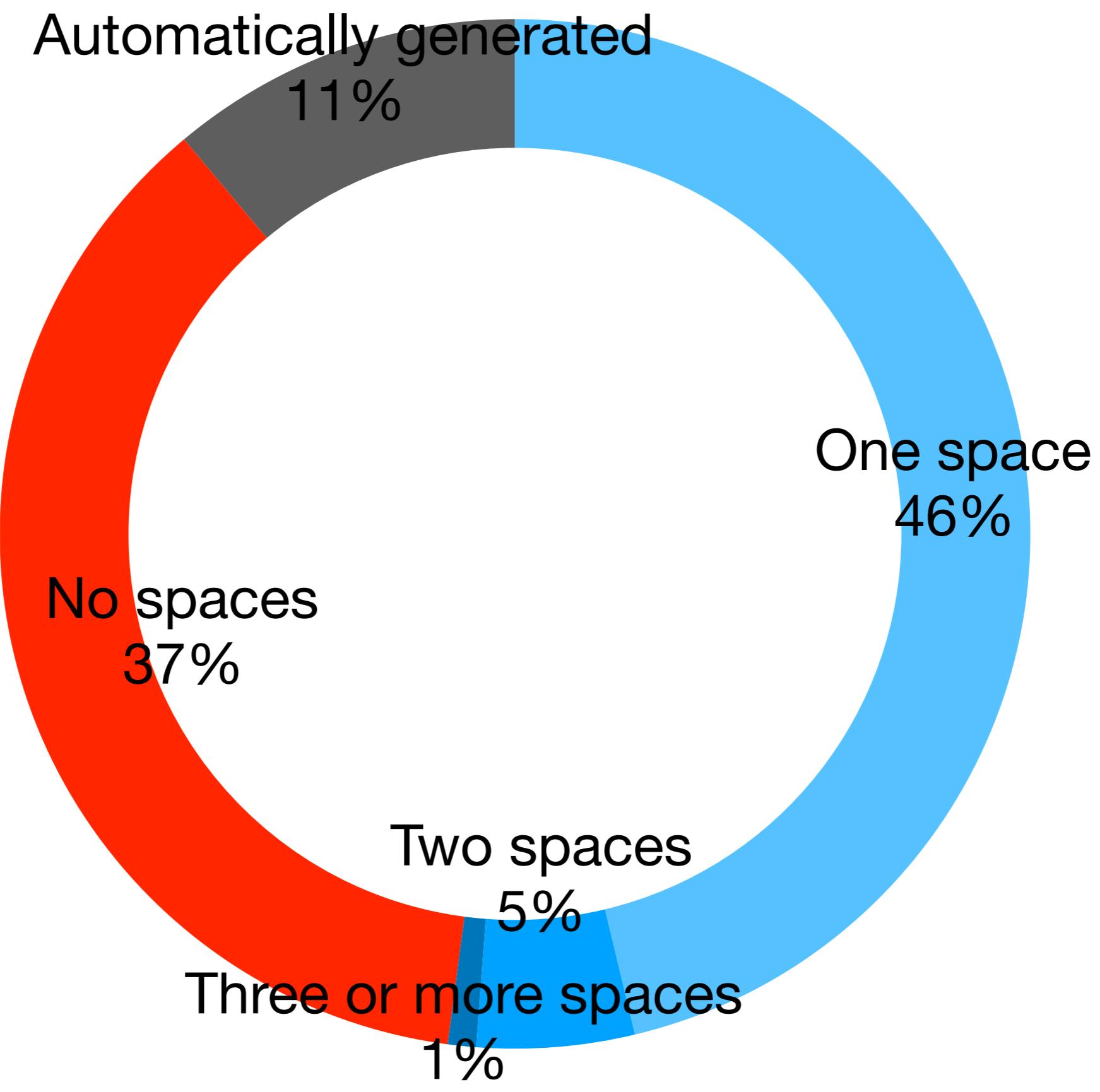


Dr Andrea Capiluppi

Lecturer

[NamSor](#) | [api doc](#) | [pricing](#) | [contact](#)Your list [dataset_20181002_11_34](#)[Lists operations](#)Download full results [Excel \(XLSX\)](#) | [CSV/Latin-1](#) | [TXT/UTF-8](#)[Page 0](#) [Next](#)

Input Name		Extract Gender		Extract Origin						Extract Diaspora		
Full/First	Last	G	S (-1..1)	Country Name	ISO2	Sub-Region	Region	Origin Score	Script	Ethnicity	Country (ISO2)	Diaspora Score
andrea	capiluppi	M	-0.50	Italy	IT	Southern Europe	Europe	5.72	LATIN	Italian	IT	5.07



Tends to
overreport
unknowns

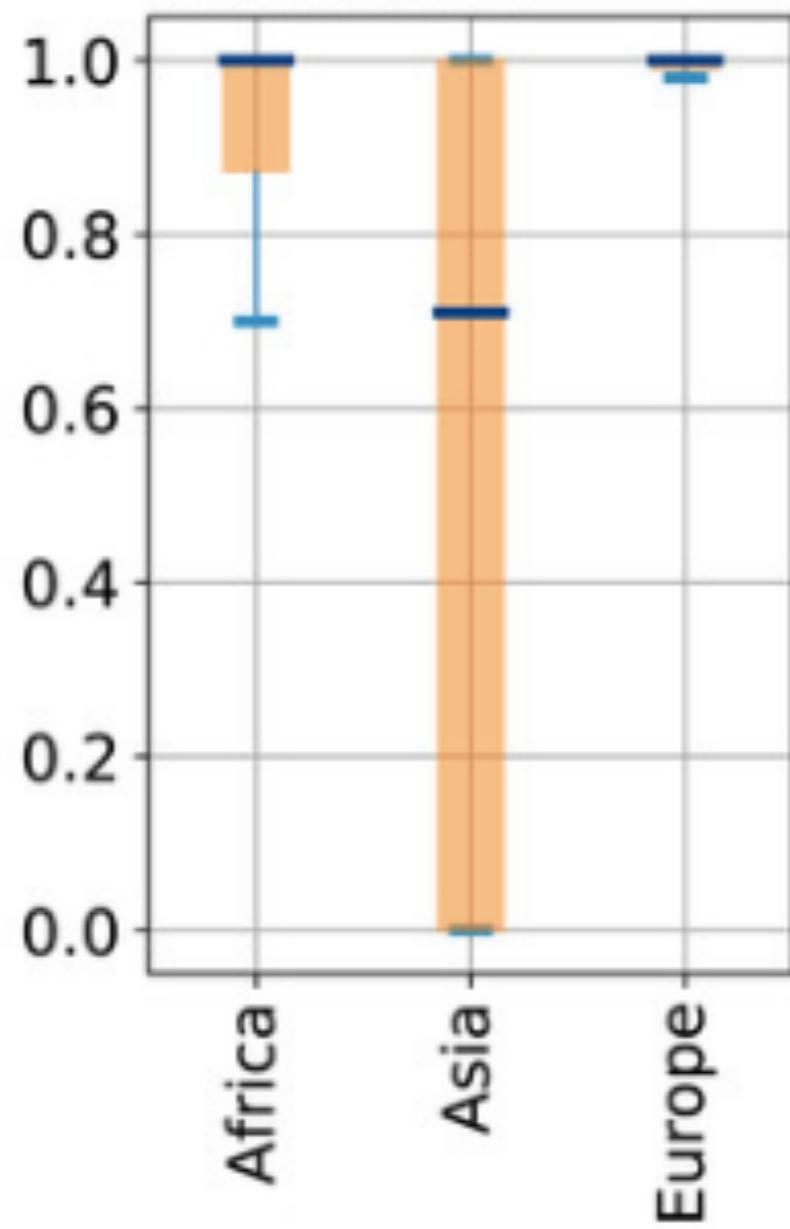
Tends to
overpredict
women as men

	errorCoded	errorCodedWithoutNA	errorGenderBias
Gender API	0.0789	0.0503	-0.0111
gender-guesser	0.2224	0.0264	0.0022
genderize.io	0.1428	0.0502	0.0222
NameAPI	0.1794	0.0342	0.0037
NamSor	0.1282	0.0429	0.0072

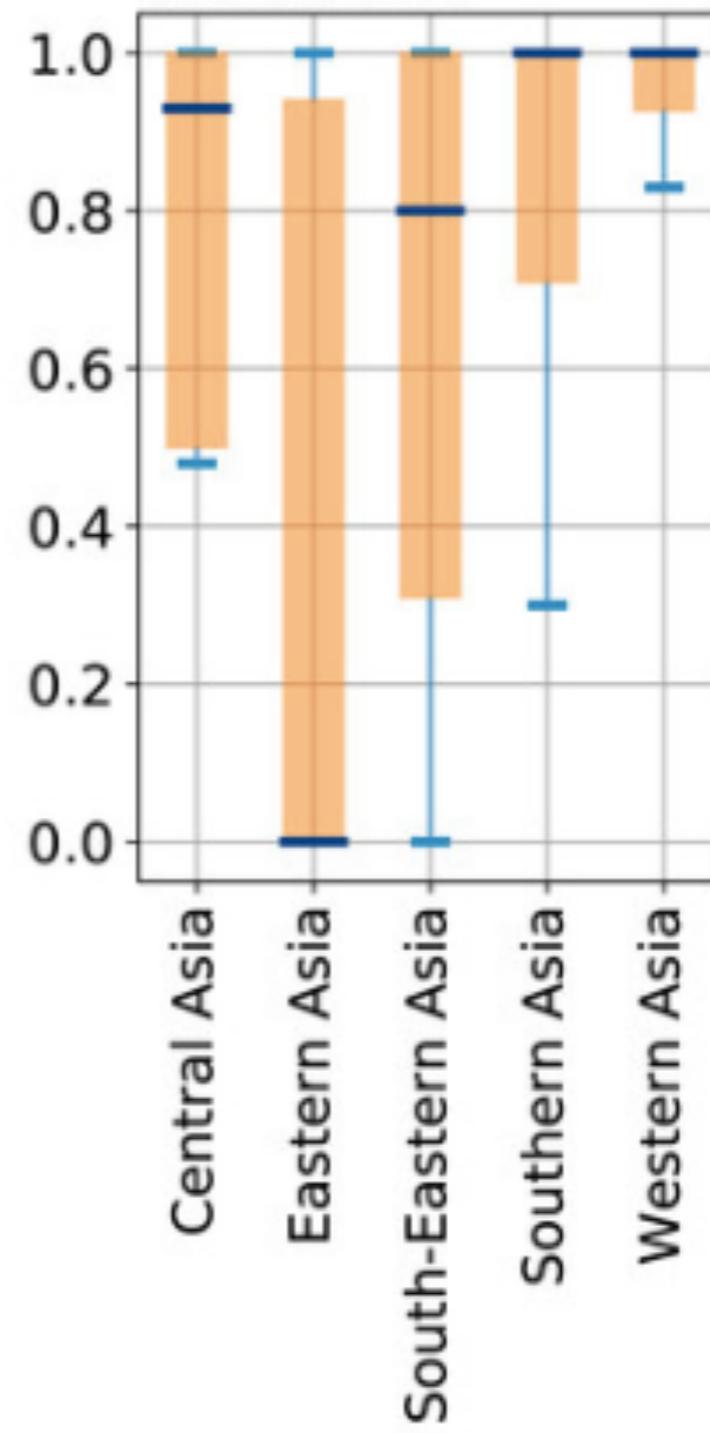
7,076 names
3,811 male,
1,968 female,
1,297 unknown

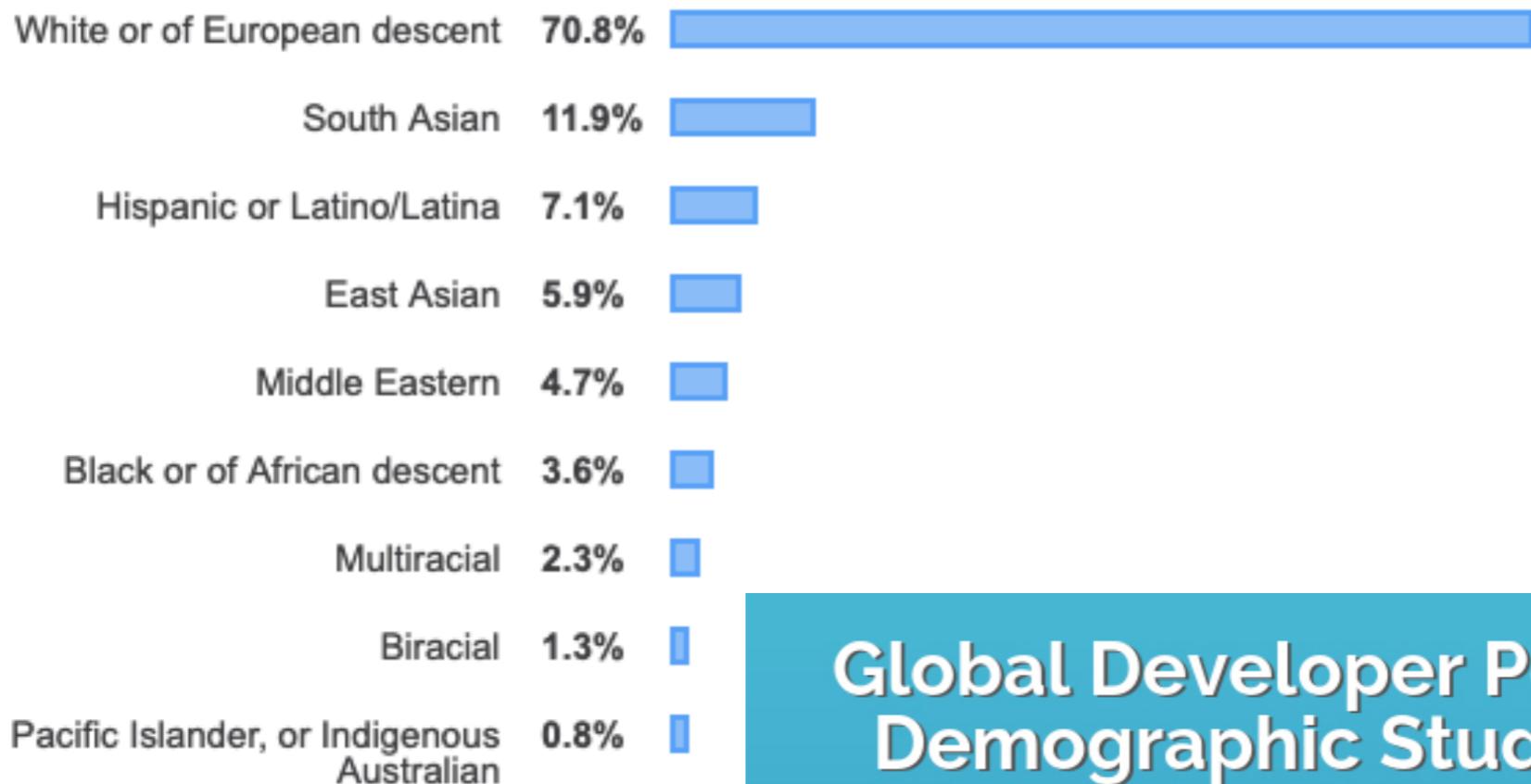
Tends to
overpredict men
as women

E. NamSor



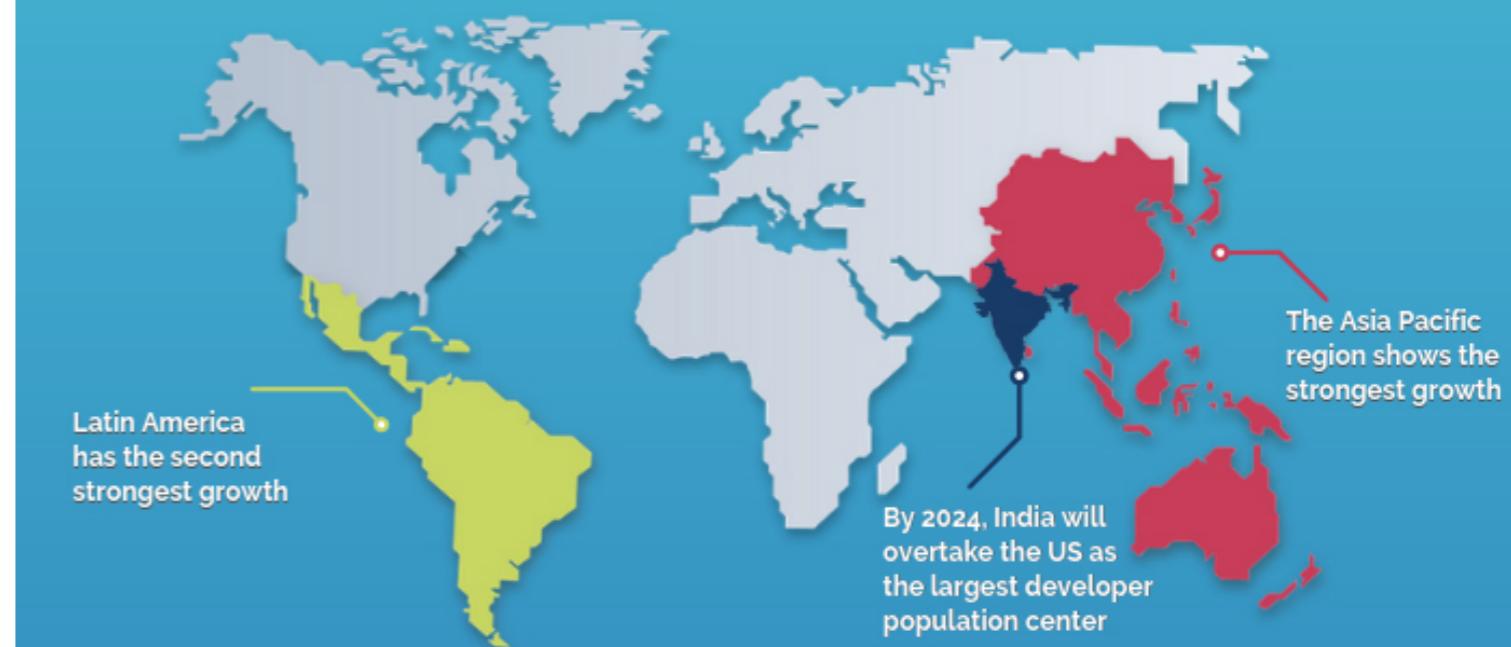
E. NamSor





76,668 responses; select all that apply

Global Developer Population and Demographic Study 2019, Vol 1



2019: 23.9 million developers
2024: 28.7 million developers

Inferring gender from names

<https://github.com/tue-mdse/genderComputer>



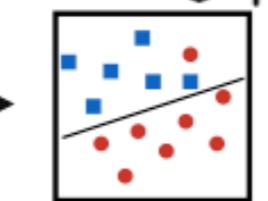
<https://www.namsor.com>



name features, e.g., the last two characters



Public name lists & celebrity names,
including 3,000 East Asian names



Naive Bayes
classifier



Binary gender
prediction

With special thanks to Huilian Sophie Qiu (CMU)

Inferring gender from names

<https://github.com/tue-mdse/genderComputer>



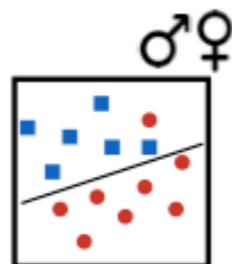
<https://www.namsor.com>



name features, e.g., the last two characters



Language	Accuracy		
	genderComp.	NamSor	Our classifier
Chinese	18%	7%	60%
Japanese	77%	27%	80%
Korean	19%	14%	68%
All	79%	74%	84%

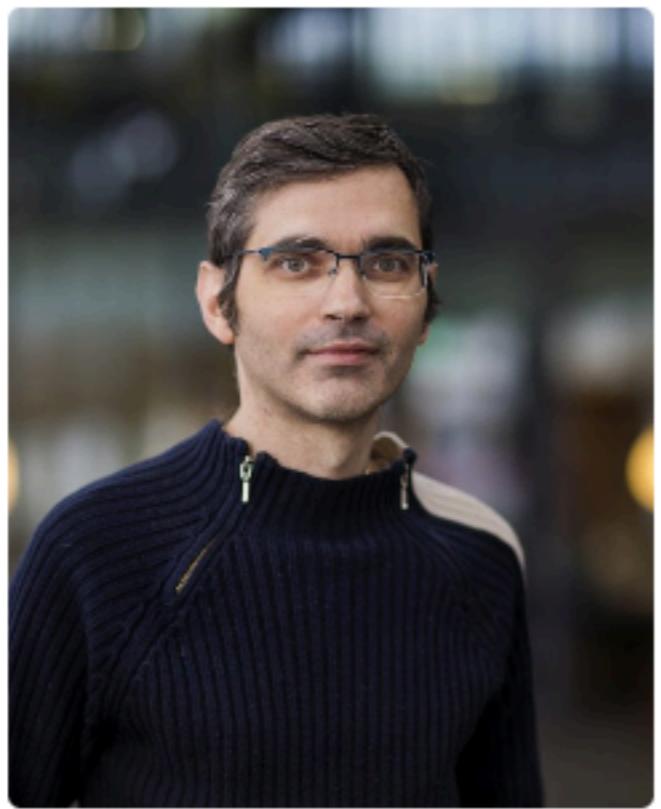


Naive Bayes classifier



Binary gender prediction

With special thanks to Huilian Sophie Qiu (CMU)



Redo

YOUR RESULT



Man



From 31 to 38 years
old

<https://www.facelytics.io/en/>

YOUR RESULT



Redo



Man



From 33 to 40 years
old

YOUR RESULT



Redo



Woman



From 27 to 34 years
old



~30%

autogenerated profile images

53% (479/900)





Table 1: Classification of Automated Gender Identification Approaches 2017 - 2019

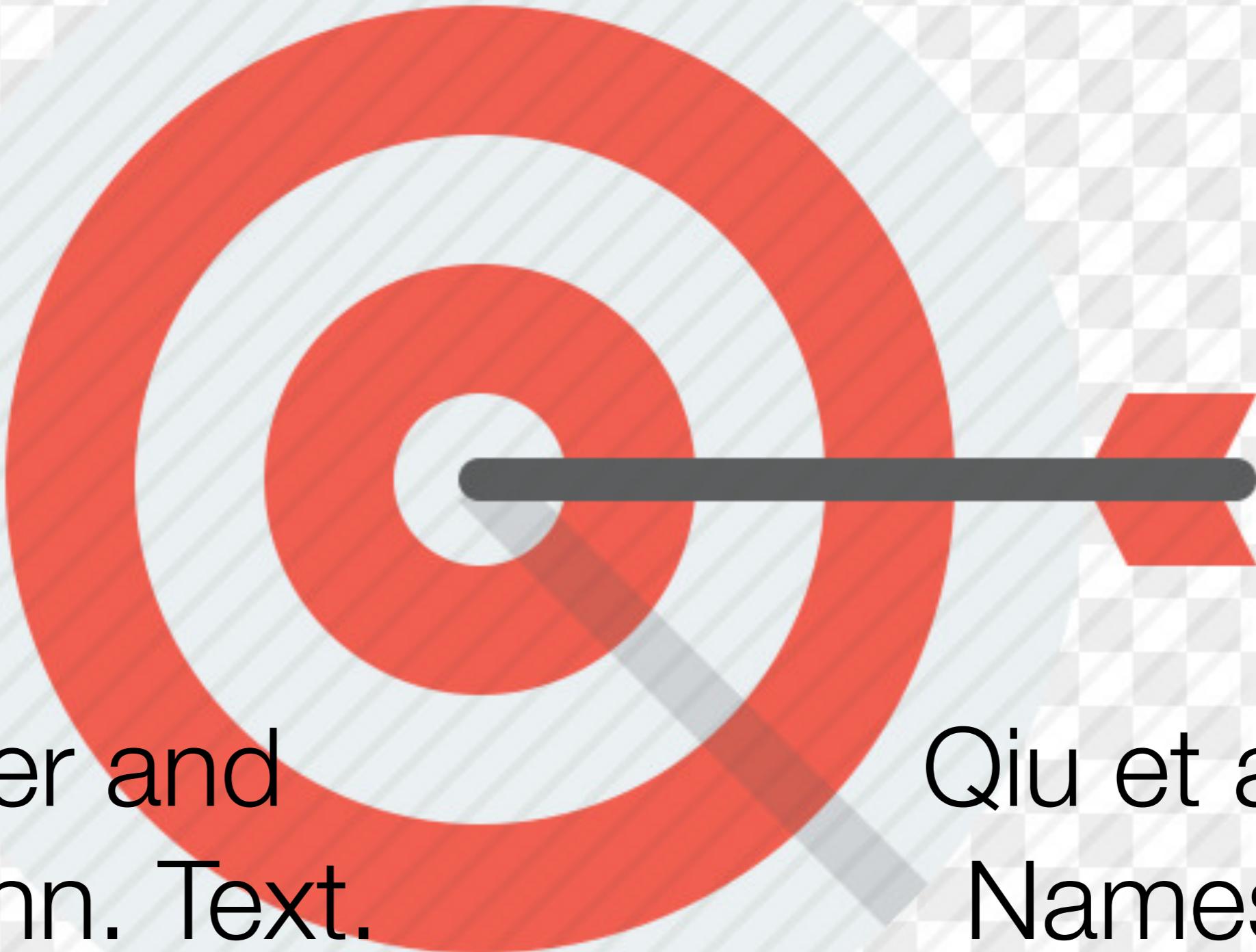
Approach	Training Data	Test Data	Best Reported Accuracy
Team nissim17 at PAN 2017	240,000 Arabic Tweets	160,000 Arabic Tweets	68.31%
	360,000 English Tweets	240,000 English Tweets	74.3%
	420,000 Spanish Tweets	280,000 Spanish Tweets	80.4%
Team miura17 at PAN 2017	120,000 Portuguese Tweets	80,000 Portuguese Tweets	85.75%
Tellez et al. [25] at PAN 2018	150,000 Arabic Tweets	100,000 Arabic Tweets	81.7%
Daneshvar and Inkpen [6] at PAN 2018	300,000 English Tweets	190,000 English Tweets	82.21%
	300,000 Spanish Tweets	220,000 Spanish Tweets	82.0%
Team deepCybErnet at IberEval 2017	4,319 Catalan Tweets	1,081 Catalan Tweets	48.6%
González et al. [7] at IberEval 2017	4,319 Spanish Tweets	1,081 Spanish Tweets	68.6%
Markov et al. [17] at RusProfiling 2017	Tweets from 300 Users Facebook Posts from 228 users Reviews from 776 authors Essays from 370 authors	Tweets from 200 Users	68.3%
Bhargava et al. [2] at RusProfiling 2017	Texts from 94 authors in Gender Imitation Corpus		93.4% 61.9% 78.4% 66%
Bsir and Zrigui [3]	240,000 Arabic Tweets	160,000 Arabic Tweets	79.2%
Bsir and Zrigui [4]	240,000 Arabic Tweets Facebook Posts from 4,444 users	160,000 Arabic Tweets	79% 62.1%
Sanchez-Perez et al. [20]	5,187 Spanish news articles (using cross-validation)		75.61%
Company and Wanner [5]	4,284 Journalistic Posts 48 Novels		89.97% 91.78%

Table 1: Classification of Automated Gender Identification Approaches 2017 - 2019

Approach	Training Data	Test Data	Best Reported Accuracy
Team nissim17 at PAN 2017	240,000 Arabic Tweets	160,000 Arabic Tweets	68.31%
	360,000 English Tweets	240,000 English Tweets	74.3%
	420,000 Spanish Tweets	280,000 Spanish Tweets	80.4%
Team miura17 at PAN 2017	120,000 Portuguese Tweets	80,000 Portuguese Tweets	85.75%
Tellez et al. [25] at PAN 2018	150,000 Arabic Tweets	100,000 Arabic Tweets	81.7%
Daneshvar and Inkpen [6] at PAN 2018	300,000 English Tweets	190,000 English Tweets	82.21%
	300,000 Spanish Tweets	220,000 Spanish Tweets	82.0%
Team deepCybErnet at IberEval 2017	4,319 Catalan Tweets	1,081 Catalan Tweets	48.6%
González et al. [7] at IberEval 2017	4,319 Spanish Tweets	1,081 Spanish Tweets	68.6%
Markov et al. [17] at RusProfiling 2017	Tweets from 300 Users Facebook Posts from 228 users Reviews from 776 authors Essays from 370 authors	Tweets from 200 Users	68.3%
Bhargava et al. [2] at RusProfiling 2017	Texts from 94 authors in Gender Imitation Corpus		93.4%
Bsir and Zrigui [3]	240,000 Arabic Tweets	160,000 Arabic Tweets	61.9%
Bsir and Zrigui [4]	240,000 Arabic Tweets Facebook Posts from 4,444 users	160,000 Arabic Tweets	78.4%
Sanchez-Perez et al. [20]	5,187 Spanish news articles (using cross-validation)		66%
Company and Wanner [5]	4,284 Journalistic Posts		79.2%
	48 Novels		79%
			62.1%
			75.61%
			89.97%
			91.78%

However...

Accuracy



Krüger and
Hermann. Text.

62%-93%

for different datasets

Qiu et al.
Names

60%-84%

for different kinds of names

Reliability

“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”**

Gender binary

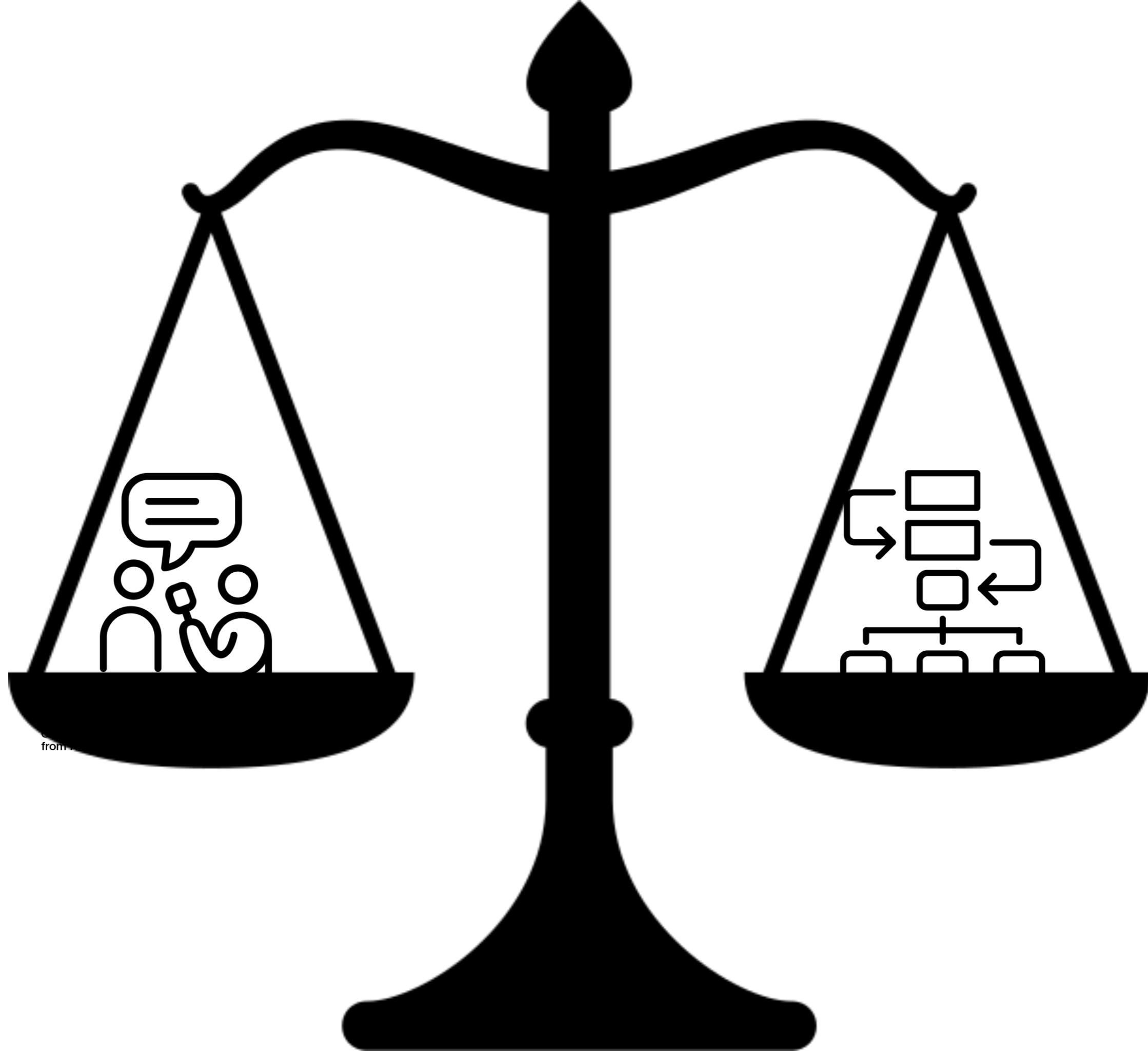
Krüger and Hermann. Text. 100%

Keyes. Face. 92.9-96.7%

Santamaría and Mihaljević. Names. 20%

Stefan Krüger, Ben Hermann. Can an Online Service Predict Gender? - On the State-of-the-Art in Gender Identification from Texts. Gender Equality Workshop ICSE 2019

Os Keyes. The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition. CSCW 2018



from:





First steps

Approach



Semi-Structured Interviews



3 Trans Software Engineers



Reviewed Recordings



Compared Themes

Interview Topics

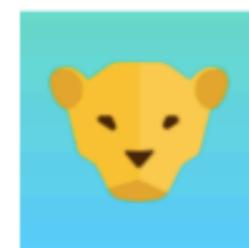
- Professional Identity Management
- Platform specific barriers
- Online safety
- Perceptions of inclusion

Control of Identity Disclosure: The desire to be seen as presented

“Stack Overflow has constrained expressions of identity. **It's up to you** what content you want to fill in. GitHub for a while it was required you expose your email address to the rest of the world.”

Petruzalek: The obvious drawback of not being passable is that you become an instant target. So passability is not only an **identity goal**, its also a mean of **self-preservation**

Top Hunters

**plan_rich**12 bounties solved
\$83,220 rewards**iidebyo**1 bounties solved
\$50,015 rewards**arigo**8 bounties solved
\$44,440 rewards**quickwritereader**8 bounties solved
\$29,400 rewards**Marcin Kościelnicki**14 bounties solved
\$24,700 rewards**ViralBShah**2 bounties solved
\$21,500 rewardsBitcoin-
com

Add CashShuffle support

 posted by:

abrkn

 created: 1 year ago  updated: 17 days ago

[!\[\]\(https://api.bountysource.com/badge/issue?issue_id=52741547\)](https://api.bountysource.com/badge/issue?issue_id=52741547) Add optional [CashShuffle](<https://cashshuffle.com/>) support when sending transactions. The lack of fungibility/privacy in Bitcoin

Economically Stable Work: Distance technical merits from identity



rust-lang/rust

Altivec/VSX support

 posted by:

lu-zero

 created: 1 year ago  updated: 21 days ago

Since llvm seems to have an almost viable intrinsics support now, is it possible to expose them in rust like the neon ones?

 11 comments

\$ 5,000 USD

You cannot tell from my technical profiles that I'm transgender. I don't make a big deal that in professional context. It's just not relevant

Search



Get CAN

Sign In



pian_rich

12 bounties solved
\$83,220 rewards



iidebyo

1 bounties solved
\$50,015 rewards



arigo

8 bounties solved
\$44,440 rewards



quickwritereader

8 bounties solved
\$29,400 rewards



Marcin Kościelnicki

14 bounties solved
\$24,700 rewards



ViralBShah

2 bounties solved
\$21,500 rewards



abrkn

created: 1 year ago updated: 17 days ago

 Add optional [CashShuffle](https://cashshuffle.com/) support when sending transactions. The lack of fungibility/privacy in Bitcoin

Economically Stable Work: Distance technical merits from identity



Altivec/VSX support

posted by:

lu-zero

created: 1 year ago updated: 21 days ago

Ross: “Technology has totally **leveled the playing field** for someone like me. I can get on the internet and watch tutorials. I have the drive to spend five hours a day to teach myself a skill.”

www.bountysource.com

10

A photograph of a woman with dark hair tied back, wearing a patterned top and jeans, sitting on a wooden dock. She is looking out over a sandy beach towards the ocean under a clear blue sky.

Control of Identity Disclosure:
The desire to be seen as presented

Economically Stable Work:
Distance technical merits from identity

Autonomy to Disengage or Reengage



We believe that **remote work** offers a mechanism of control for identity disclosure and empowerment of software developers from **any marginalized communities.**

With special thanks to Margaret Burnett (Oregon State University)



With special thanks to Margaret Burnett (Oregon State University)



