Introduction to gender research in the Wikimedia movement

Nazia Akhtar, Netha Hussain, Miriam Redi, Rosie Stephenson-Goodknight

October 20, 2023

WIKIWOMENCAMP

WIKI WOMEN
CAMP 2023

Background

Rosiestep

MAP UP, RISE UP!

Our approach

- **Research team members**:
  - Nazia Akhtar, Netha Hussain, Miriam Redi, Rosie Stephenson-Goodknight
- **Pre-WikiWomenCamp**:
  - Develop the Research Pre-Read Pack in order to acquaint you with some of the Wikimedia+gender research to date
  - Develop the Research Pre-Conference Survey to gain insight into your experiences and viewpoints with Wikimedia+gender research
  - Develop the research track sessions

Objectives

Day	Research session #	Title
Friday	1	Intro to research, gender, and Wikimedia
Saturday	2	Research discovery and centralization
	3	Research community
Sunday	4	Hubs+Research
	5	Building a research network
	(plenary)	Partner research

Scan this QR code to join →

AhaSlides

Research team

We develop **models and *insights*** utilizing **scientific methods**, and grow the research communities

to

Support technology/policy needs and advance the understanding **of the Wikimedia projects**.

16 staff, 5 contractors, 1 Research Fellow, **17 Formal Collaborators**

Knowledge Gaps

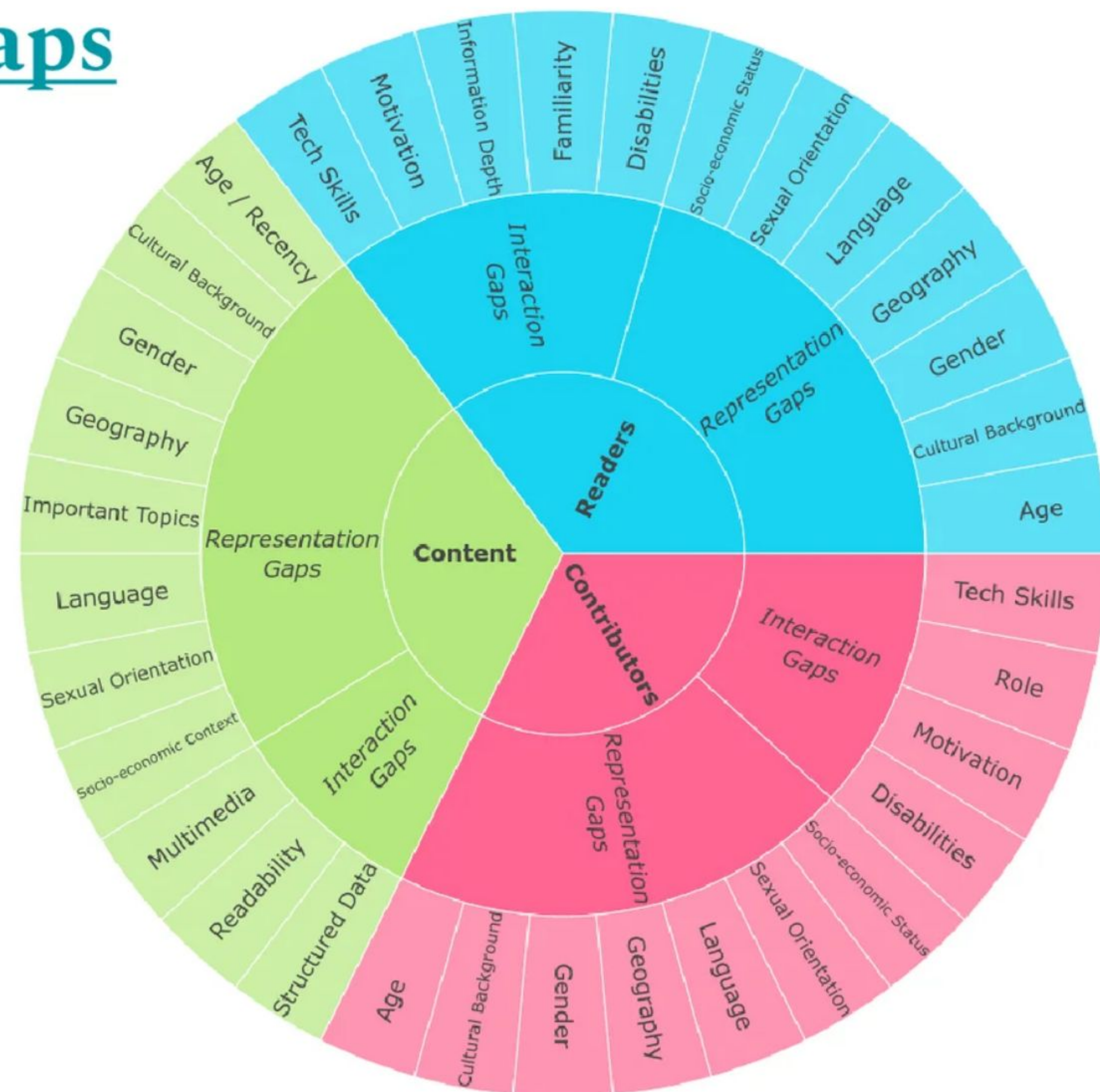**What** are knowledge gaps across readers, contributors and content?

How can we address them?

What are knowledge gaps?

Taxonomy of Knowledge gaps

- 200+ references
- Finding and classifying evidence of inequalities in Wikimedia projects.
- Grouping gaps into Content, Readers, and Contributors

Redi, Miriam, Martin Gerlach, Isaac Johnson, Jonathan Morgan, and Leila Zia. "A taxonomy of knowledge gaps for wikimedia projects (second draft)." *arXiv preprint arXiv:2008.12314* (2020).
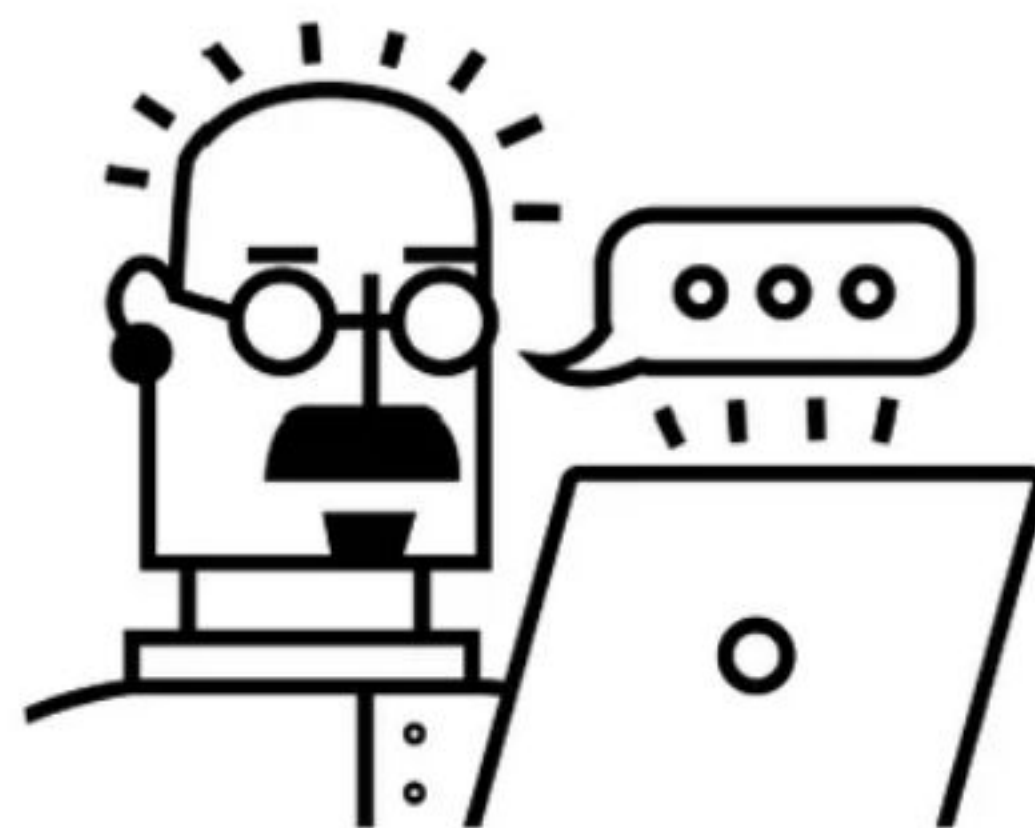
The Gender Gap

Content

81%

Of Wikipedia articles across languages are about men

From the Knowledge Gap Index

Contributors

80%

Of active editors self-identify as men

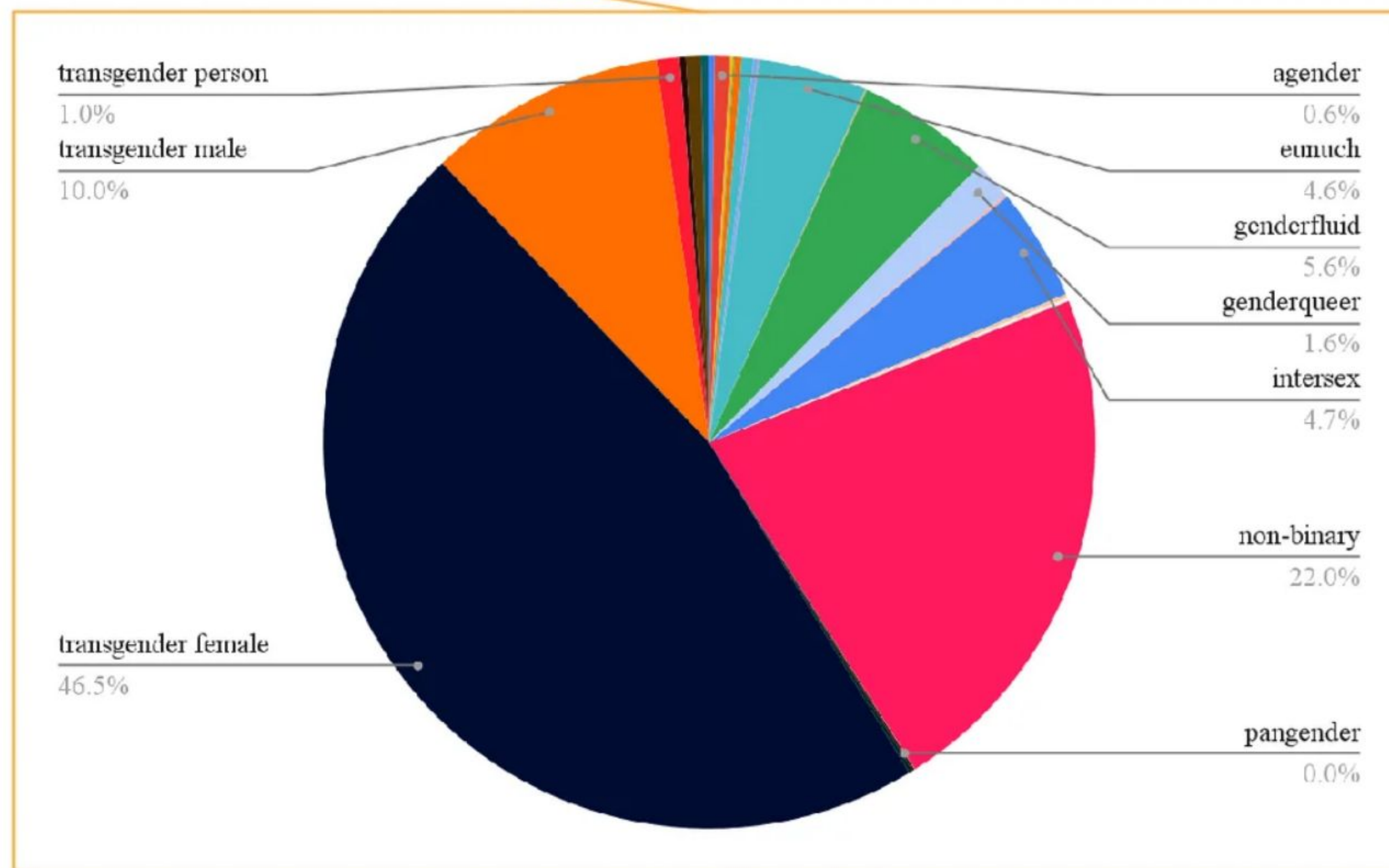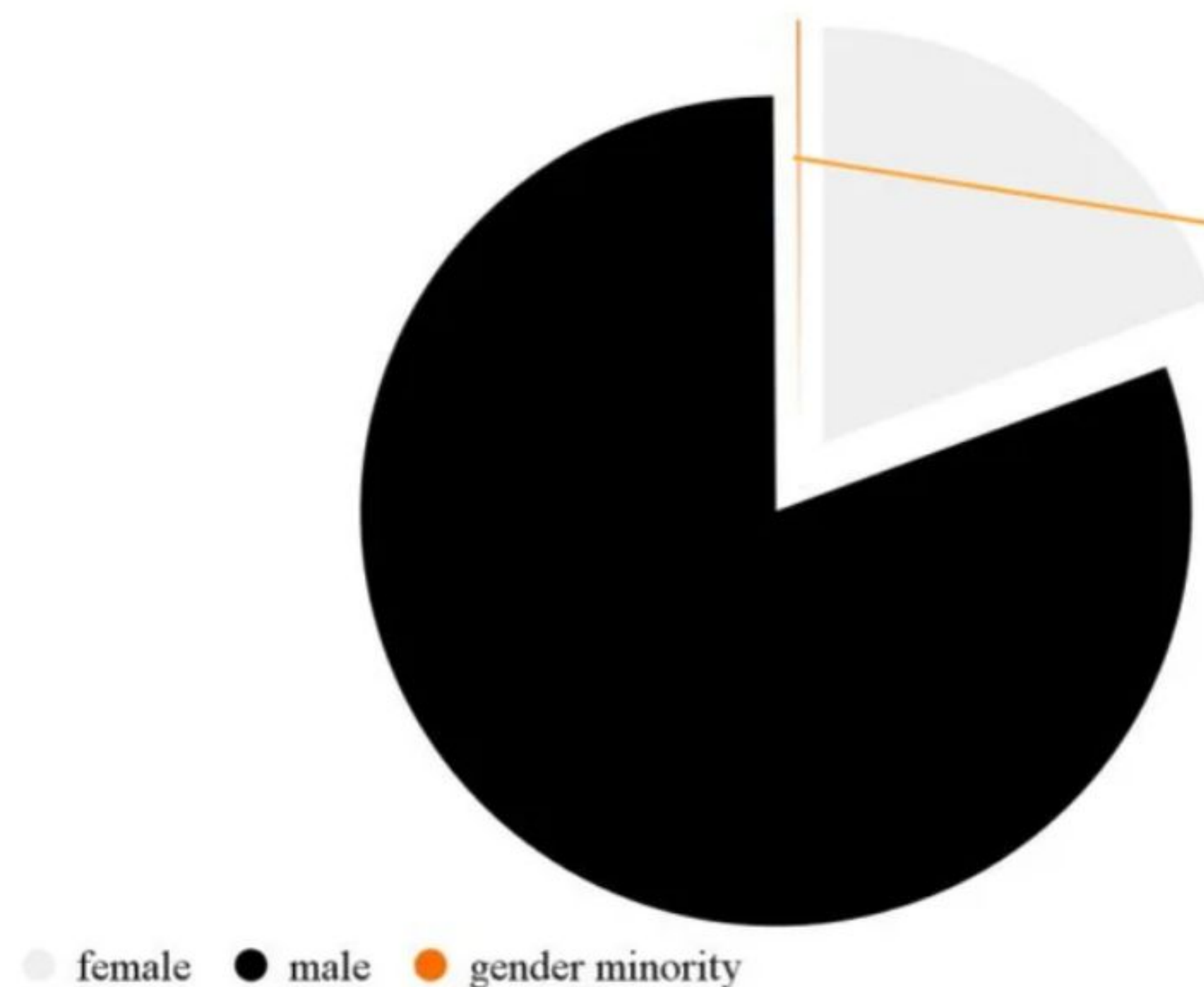From the '22 Community Insights Survey

Readers

72%

Of pageviews come from people who self-identify as men

From the '19 Reader Demographics surveys

The majority of biographies is about men ...

female   male   gender minority

Category	Percentage
transgender person	1.0%
transgender male	10.0%
transgender female	46.5%
agender	0.6%
eunuch	4.6%
genderfluid	5.6%
genderqueer	1.6%
intersex	4.7%
non-binary	22.0%
pangender	0.0%

Predicted article quality, 3 genders, over time
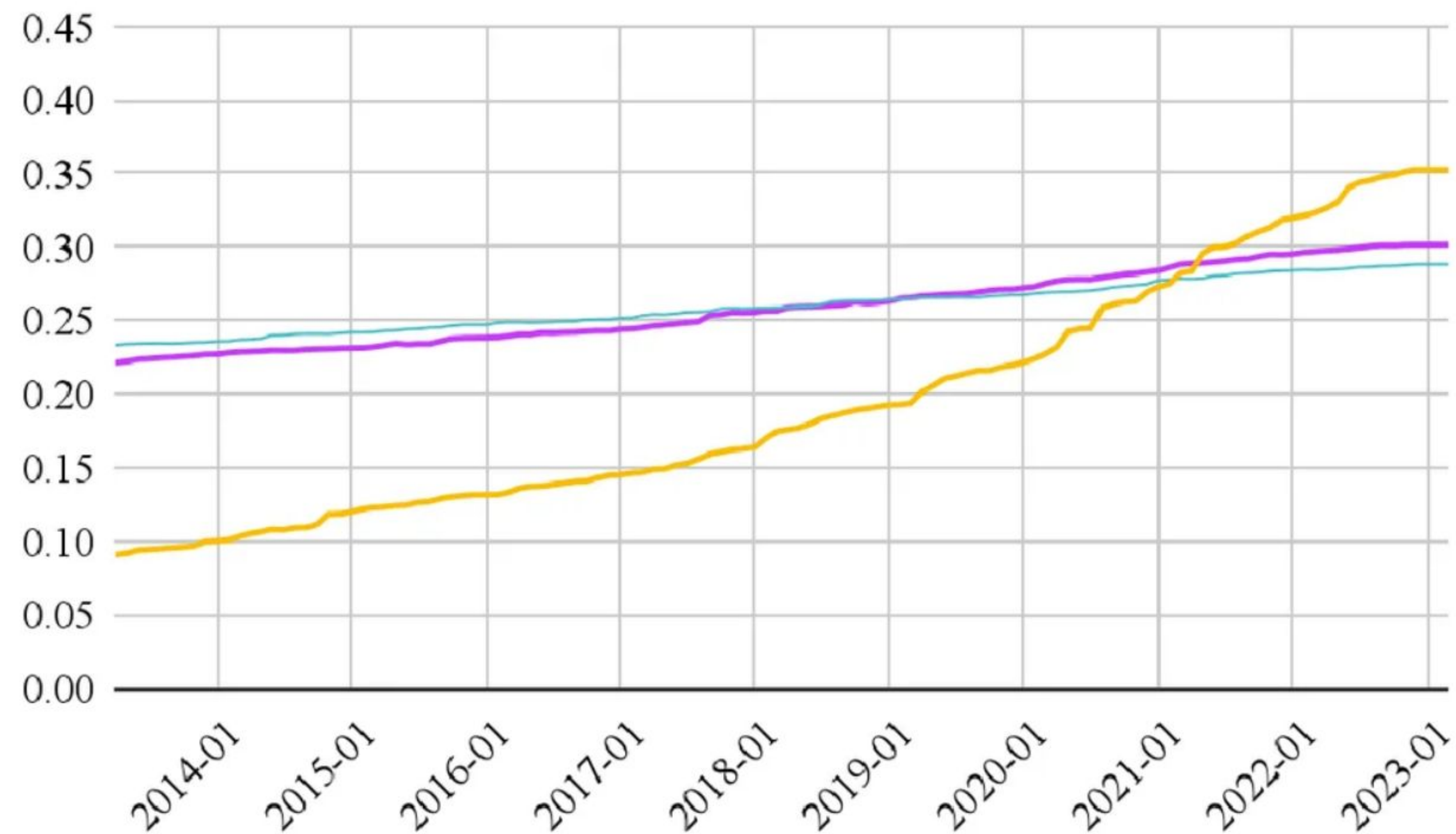
... but the quality **quality of articles** about **women and gender minorities** is on average **higher** than in men bios
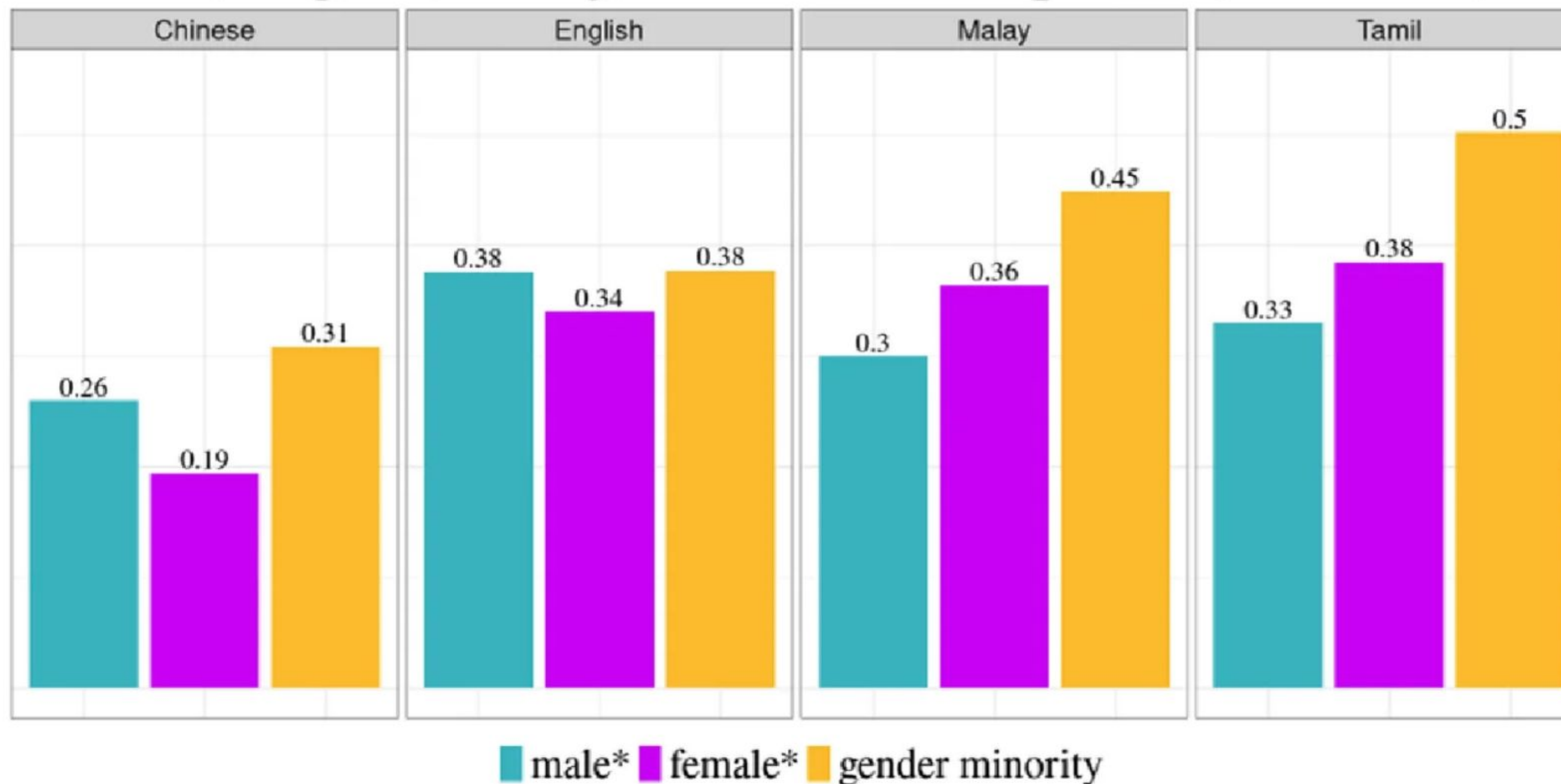
... and it **varies across language** editions

Average article quality scores, 3 gender categories:
Chinese, English, Malay, and Tamil Wikipedias (June 2023)

Chinese | **English** | **Malay** | **Tamil**

Chinese: male 0.26, female 0.19, gender minority 0.31
English: male 0.38, female 0.34, gender minority 0.38
Malay: male 0.3, female 0.36, gender minority 0.45
Tamil: male 0.33, female 0.38, gender minority 0.5

■ male* ■ female* ■ gender minority

* "male" and "female" here refer to non-gender-minority males and females

Knowledge Gap Metrics **Datasets**
Now Available! API and tools are coming!

Article **Quantity**

Article **Quality**

Monthly **Pageviews**

Monthly **Revisions**

✖

Gender of the article subject

Geography countries related to the article

Sexual Orientation of the article subject

Time the temporal setting of the article

Multimedia whether the article has image

Watch the full video on this:
https://meta.wikimedia.org/wiki/Research:Knowledge_Gaps_Index/Wikimania2023

The Gender Gap

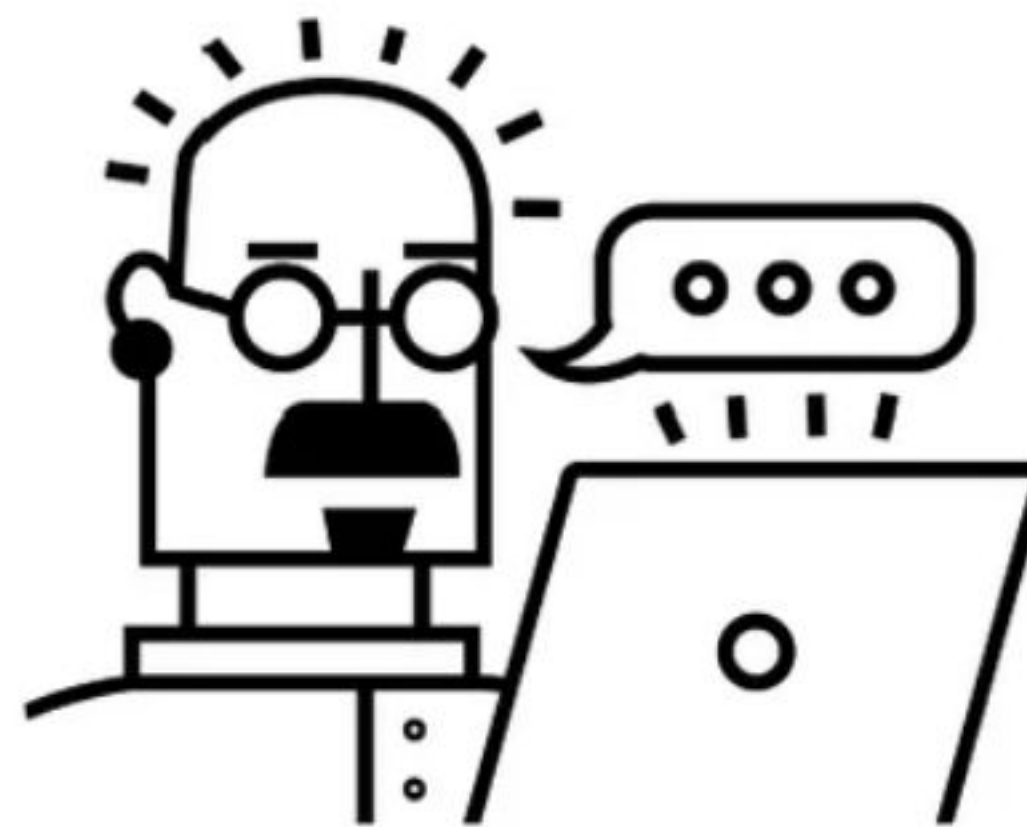## Content

81%

Of Wikipedia articles across languages are about men

From the Knowledge Gap Index

Contributors

80%

Of active editors self-identify as men

From the '22 Community Insights Survey

Readers

72%

Of pageviews come from people who self-identify as men

From the '19 Reader Demographics surveys

But women and gender diverse contributors are more likely to indicate having **felt unsafe on-wiki**.

"In the last 12 months, how often have you felt unsafe or uncomfortable contributing to Wikimedia projects online?"

Man	Woman	Gender Diverse	Prefer not to say

Man
- Never: 62%
- Once or more: 33%
- Unsure: 6%

Woman
- Never: 53%
- Once or more: 44%
- Unsure: 3%

Gender Diverse
- Never: 47%
- Once or more: 52%
- Unsure: 1%

Prefer not to say
- Never: 43%
- Once or more: 45%
- Unsure: 12%

The Gender Gap

Content

81%

Of Wikipedia articles across languages are about men

From the Knowledge Gap Index

Contributors

80%

Of active editors self-identify as men
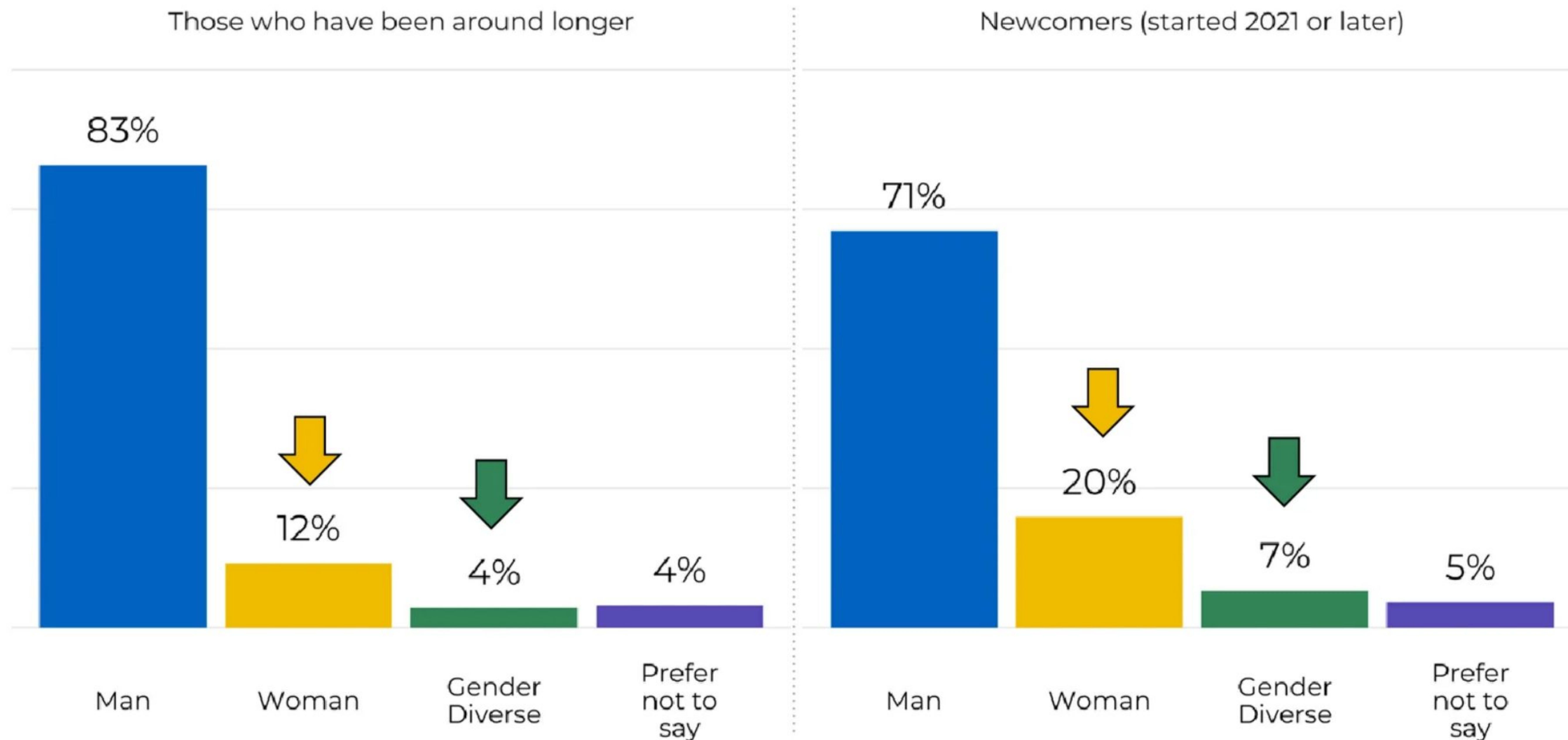
From the '22 Community Insights Survey

Readers

72%

Of pageviews come from people who self-identify as men
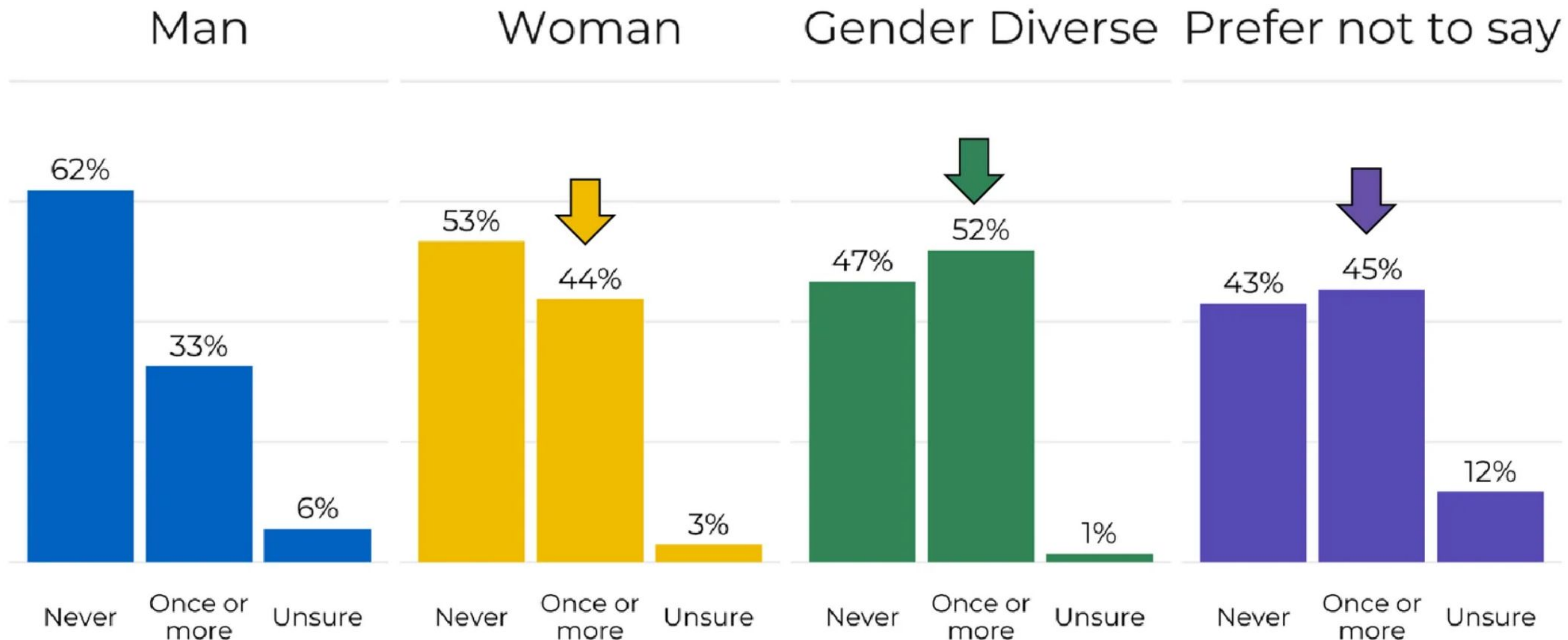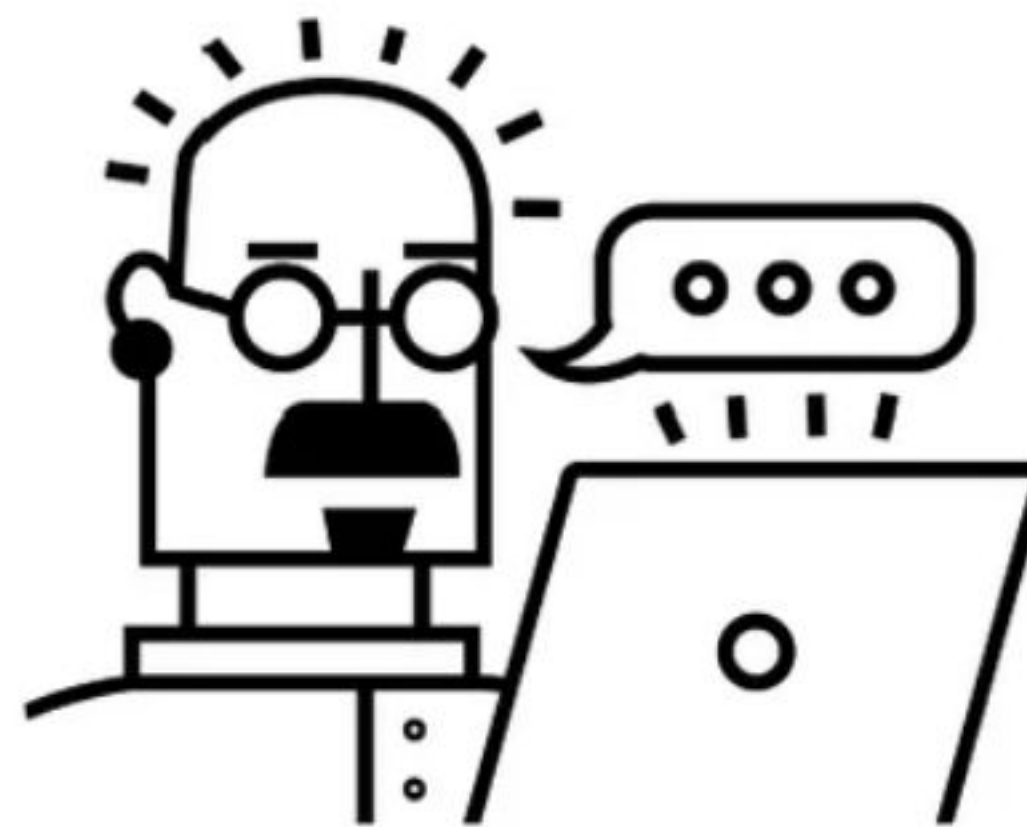
From the '19 Reader Demographics surveys

Some **languages** are a bit **more diverse**

Distribution of reader gender demographics

men   women and non-binary

romanian
ukranian
russian
hungarian
chinese
hebrew
spanish
french
german
english
norwegian
arabic
farsi

0%   25%   50%   75%   100%

Johnson, I., Lemmerich, F., Sáez-Trumper, D., West, R., Strohmaier, M., & Zia, L. (2021, May). Global gender differences in Wikipedia readership. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 15, pp. 254-265).
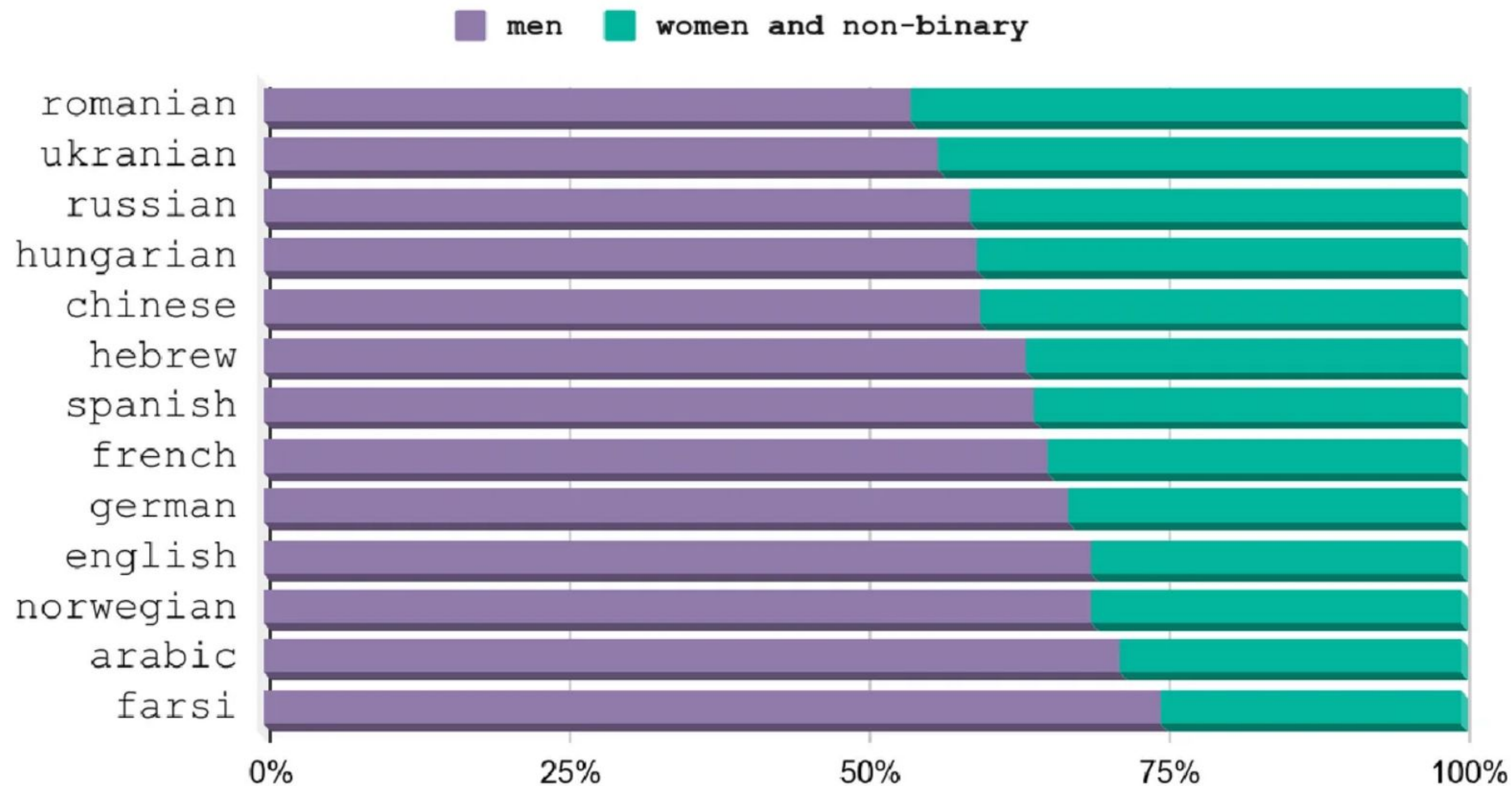
Topic preferences also vary by gender!

Readers tend to **read articles about people of similar demographics** (age and gender) to them

Legend:
- Men who read biographies of men
- Women who read biographies of men
- Men who read biographies of women
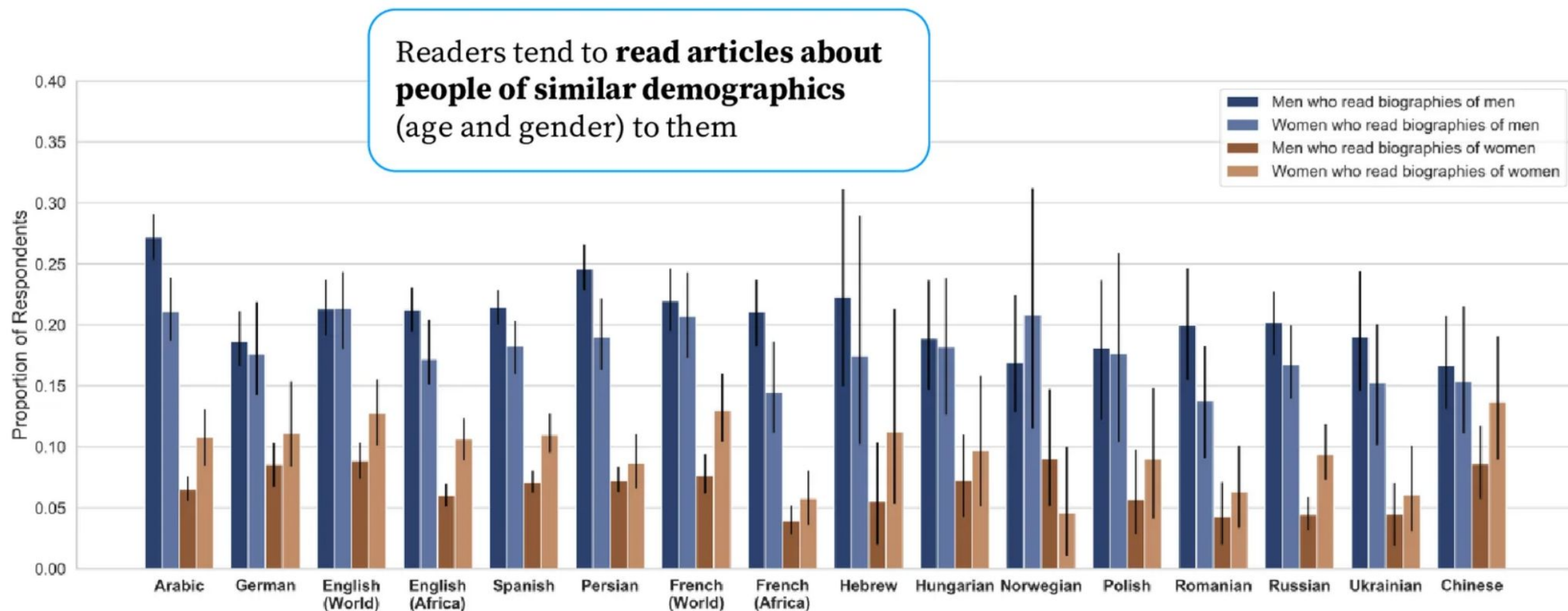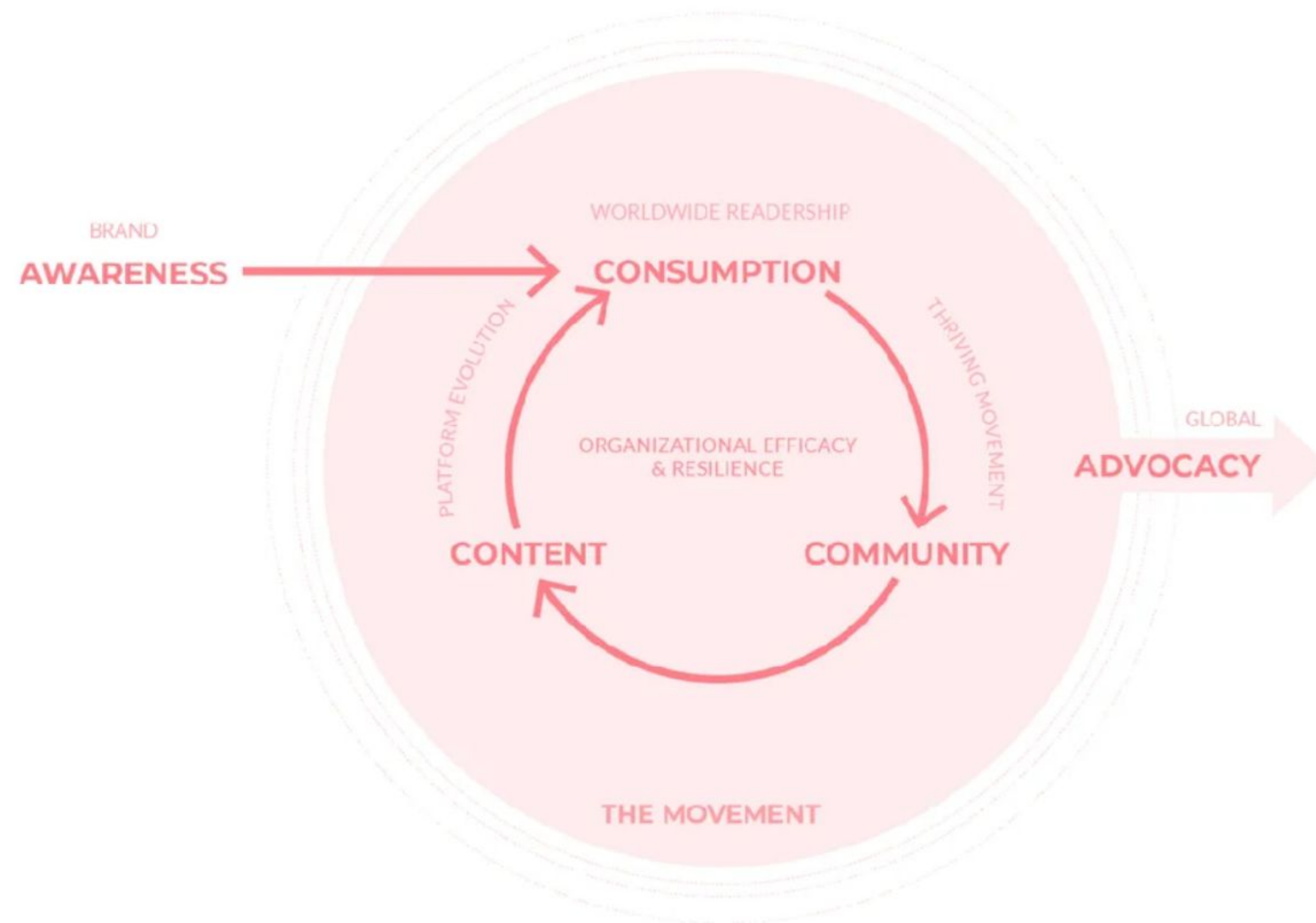- Women who read biographies of women

Figure 4. Interest in Wikipedia biographies about men and women by reader gender. The figure shows the proportion of men, resp. women, viewing at least one article with a biography about a man/woman. We can see that in most languages women read articles about women comparatively more often.

The Gender Gap propagates in the flywheel

Readers tend to read articles about people of similar demographics (age and gender) to them, suggesting a **relationship between content gaps and reader gaps.**

We cannot **address gaps** in our **editor population** without also **addressing gaps in our reader population.**

BRAND
AWARENESS

WORLDWIDE READERSHIP
CONSUMPTION

PLATFORM EVOLUTION

THRIVING MOVEMENT

ORGANIZATIONAL EFFICACY & RESILIENCE

GLOBAL
ADVOCACY

CONTENT

COMMUNITY

THE MOVEMENT

Reach out!

- *wiki-research-l@lists.wikimedia.org*
- **@WikiResearch**

25

WIKIWOMENCAMP

WIKI WOMEN
CAMP 2023

Participation/people gender gap research

Rosiestep

MAP UP, RISE UP!

Participants

- Editors
- Developers
- Organizers
- Leaders
- Social media influencers
- Partners
- Donors
- Researchers
- Readers

Editors

Percentage of Wikipedia editors who identify as women		
2010	12.6%	United Nations University study
2011	8.5%	Wikipedia Editor Survey
2018	9%	Community Engagement Insights survey
2019	11%	Community Insights/Community Insights Report
2020	15%	Community Insights/Community Insights Report
2021	15%	Community Insights/Community Insights Report
2022	13%	Community Insights/Community Insights Report

https://meta.wikimedia.org/wiki/Community_Insights/Community_Insights_2023_Report#Gender

Editors

- **Peer-reviewed**: Ferran-Ferrer, Núria; Castellanos-Pineda, Patricia; Minguillón, Julià; Meneses, Julio (September 6, 2021). "The gender gap on the Spanish Wikipedia: Listening to the voices of women editors". *El Profesional de la Información*. **30** (5). doi:10.3145/epi.2021.sep.16. S2CID 241442991.

- **WikiWorkshop**: Björn Ross, Marielle Dado, Maritta Heisel, Benjamin Cabrera, Gender Markers in Wikipedia Usernames, *Wiki Workshop*, April 2018, Lyon, France

- **Newsletter**: Wilson, Jason (10 February 2015). "Are misogynists running Wikipedia?". *Overland*.

- **Blogpost**: Gardner, Sue (20 February 2011). "Nine Reasons Women Don't Edit Wikipedia (in their own words)". *Suegardner.org*.

- **Thesis**: Meyer, Christine (May 2022). "'If You Want to Change the World, Edit Wikipedia': Mitigating the Gender Gap and Systemic Bias on Wikipedia". University of Idaho.

https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Women_in_Red/Research#2011

Readers

WMF Research: *Characterizing Wikipedia Reader Behaviour*:

- March 2019, small-scale pilot, WMF survey, run on English Wikipedia
- Survey respondents skewed heavily young and male.
- Including those who were under the age of 18, 70% of respondents were under the age of 30.
- Overall, 76% of the respondents identified as male.
  - Gender percentage was consistent across age groups.
  - United States: 67% male
  - United Kingdom: 75% male
  - India: 83% male

https://meta.wikimedia.org/wiki/Research:Characterizing_Wikipedia_Reader_Behaviour/Robustness_across_languages

Content

Netha Hussain

Coverage bias

- Approximately 20% of biographies on English Wikipedia are about women.
- Articles about women on Wikipedia are longer and well-covered than traditional encyclopedias.
- Women are less likely to have a Wikipedia article compared to their male counterparts. Women in Wikipedia are more notable than men.
- Biography categories are over-represented by men.
- Gender gap is larger in larger Wikidata communities, but is becoming narrower.

Structural bias

- Articles about people with the same gender tend to link to each other.
- Articles with high centrality tend to be predominantly about men. The centrality of women decreases faster than that of men with decreasing rank.
- Wikipedia pages intermingle personal and professional life events for females but not equally so for males.
- Biographies that land in articles for deletion conversations are more likely to be about women.
- Symmetrical counterparts describing the same concept for the other gender exist only in half of the articles having 'male' or 'female' in the title.

Lexical bias

- There is a social bias to assume male as the standard gender.
- Words like "married", "divorced", "children" or "family" are much more frequently used in articles about women.
- Career related words are more strongly associated with males than females.

Visibility bias

- Publications by women authors are cited less by Wikipedia than expected.
- Systematic overrepresentation of males in the images included in biographies on Wikipedia.
- Biographies of males tend to have more images compared to females.
- Biases shape our social perceptions of knowledge. Biases on Wikipedia can propagate beyond the platform.

Scan this QR code
to join →

In terms of women, what content gaps are most concerning for you as a Wikimedian?

Articles about women

Structural bias

Women's health

Mental health

Women's, Queer, Black, Indigenous, and Decolonial History

In all perspectives

Unbalanced content, deletions, less complete articles

That there is an implicit labelling of 'female' but none for 'male' etc male doctor

Rosie's point that if you're not a reader, you're not an editor

Coverage bias

Biography for female public feagures.

Harrasment

Women content more about victimsation and less about achievements

Sexual and reproductive health

women warriors

Structure

Absence of Representation

Biographies of women

Women's traditional interests such as fashion

Speaking in conferences

Coverage

Gender determination

Every if them

Policies and practices

NaziaAkhtar27

Hidden Biases: Notability and Miscategorization

- **Women's gender plays an important role in their perceived significance.** Biographies of women are more likely and more frequently targeted and **miscategorized** for deletion on grounds of **lack of notability ("too soon")**,
  - In spite of meeting guidelines for adequate, independent media coverage.
  - Relying on uneven notions and opinions about career stage.
- This only **aggravates a long chain of historical inequalities** and marginalizations in which women's qualifications and achievements are undervalued:
  - academic recruitment of women (esp. women BIPOC academics) → academic recognition → disproportionate media coverage → erased from Wikipedia

Hidden Biases: Notability and Miscategorization

- For editors: emotional labour, degrading intrusion and harassment, burnout. Impact:
  - **Exclusion:** at least two editors monitoring, questioning, and arguing against the deletion of their articles have been driven away.
  - **Self-Effacement:** Women editors prefer to write about obscure or dead women to avoid attracting attention.
  - The likely **erasure of notable women from cultural memory**.
- Activist editors, trainers, and gender gap campaigns have helped to address the asymmetry, although these are frequently besieged by
  - **Harassment**
  - **Large volume of data to get through**

Hidden Biases: Reliability Guidelines

- **"[T]he guidelines themselves are unreliable":** there is no academic rigour in definition of reliability and its historical context.

- The **role of interpretation** in the absence of rigorous definition and contextualization is largely overlooked.

- As a result, Reliability guidelines are very clear, but **do not account for the complexity of lived experiences of different communities**, as well as different domains and areas, which may call for a different way to think about reliability. They
  - reproduce western knowledge formations and **exclude contextualized/marginalized knowledges** that matter to marginalized communities.
  - homogenize and ahistoricize sources of knowledge and **assume forms of / claims to "objective" knowing are universal.**
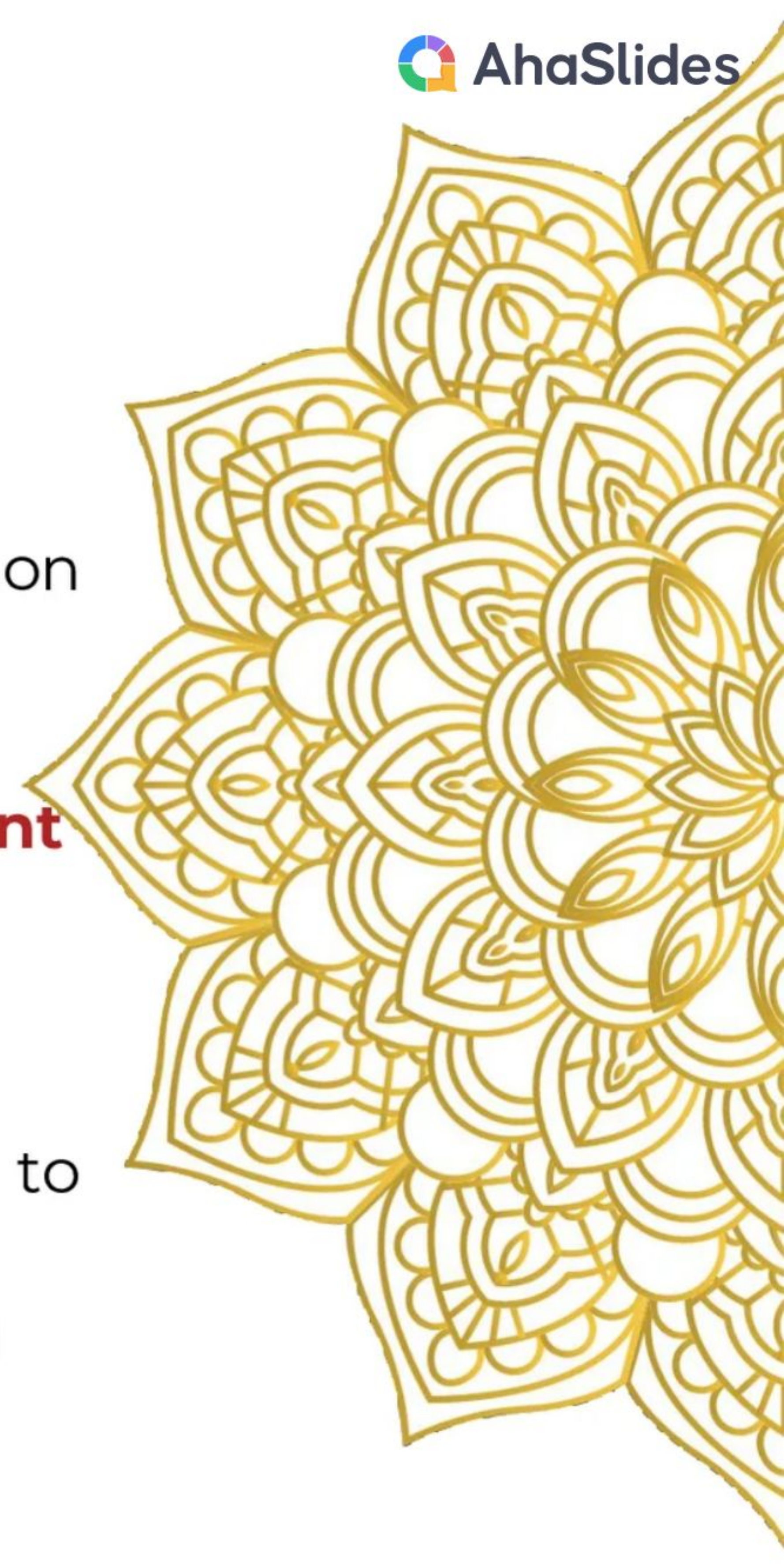
Hidden Biases: Reliability Guidelines

- Source authority on Wikipedia comes about through:
  - **"community consensus" based on silence** on the part of the majority of editors, who hesitate to interrupt the information architecture.
  - **only a small number of experienced, privileged editors then decide** what is reliable.

- **Implications:**
  - **Awareness of hidden biases is critical** to equitable application of Wikipedia's rules and guidelines.
  - **Wider implications of misrepresenting and/erasing women**: Wikipedia is playing an increasing and influential role in shaping and verifying knowledge as well as training AI on other platforms (Google, YouTube, Amazon Alexa, Twitter).

- **Recommendations:**
  - Weak-tie networks of editors to protect articles
  - New users/editors with trainers are more successful
  - Reliability guidelines and consensus processes should be revisited with more stakeholders.

What are some other policies/practices that affect women's participation in Wikimedia projects that you are concerned about?

defending existing projec

less internet access

no trust

wp casting aspertions

discussions

pdc ucoc

what harassment is

fear of re

rigidness of guidelines

women and non binary

majority over justice

not getting paid for work

voluntary work not paid

hiding misogyny

train the triners jobs

lack of agency

guidelines

harassment

confidence and trust

bias of content

fear of rebuke

reliability and notabiliy

low content knowledge

cybersafety

hire more trust and safet

a level of knowledge

reliable resources

missing allys

notability policies

not aware

moderation practices

imposter syndrome

cyber security

getting unwanted attentio

difficult to start on own

language barrier

lack of autonomy

Thanks!

Akhtar, N.; Hussain, N.; Redi, M.; Stephenson-Goodknight, R, *Introduction to gender research in the Wikimedia movement*, 20 Oct 2023, WikiWomen Camp 2023, New Delhi, India, CC BY-SA 4.0

@ Nazia_Akhtar27 | @ nethahussain | @ mad_astronaut | @ Rosiestep
@WikiWomenCamp

WIKI WOMEN CAMP 2023 | MAP UP, RISE UP!

Attributions

MAP UP, RISE UP!