

Characterizing Reader Behavior on Wikipedia

Me: Isaac (WMF) -- isaac@wikimedia.org

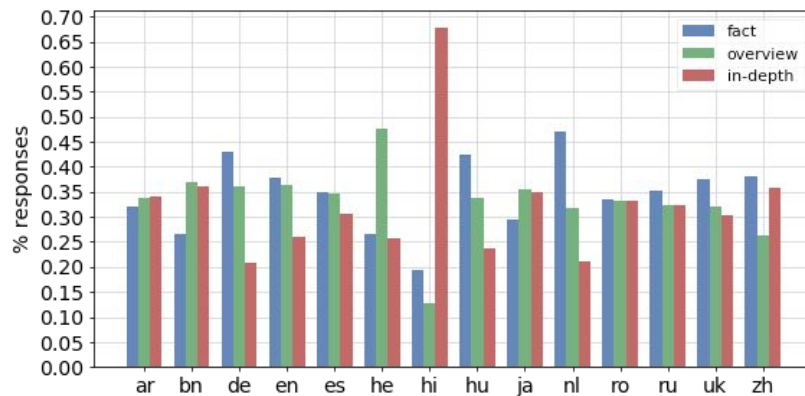
Collaborators: Florian Lemmerich (RWTH Aachen), Diego Sáez-Trumper (WMF), Markus Strohmaier (RWTH Aachen / GESIS), Bob West (EPFL), Leila Zia (WMF)

...with additional help from: Abbad (Volunteer), Amire80 (Volunteer), Yury Bulka (Volunteer), Adélaïde Calais (WMFr), Kaganer (Volunteer), Liang-chih ShangKuan (Volunteer), Jon Harald Søby (WMNO), Strainu (Volunteer), Tgr (Volunteer), Morten Warncke-Wang (WMF), Readers Web, Legal, and probably a few others that I am forgetting

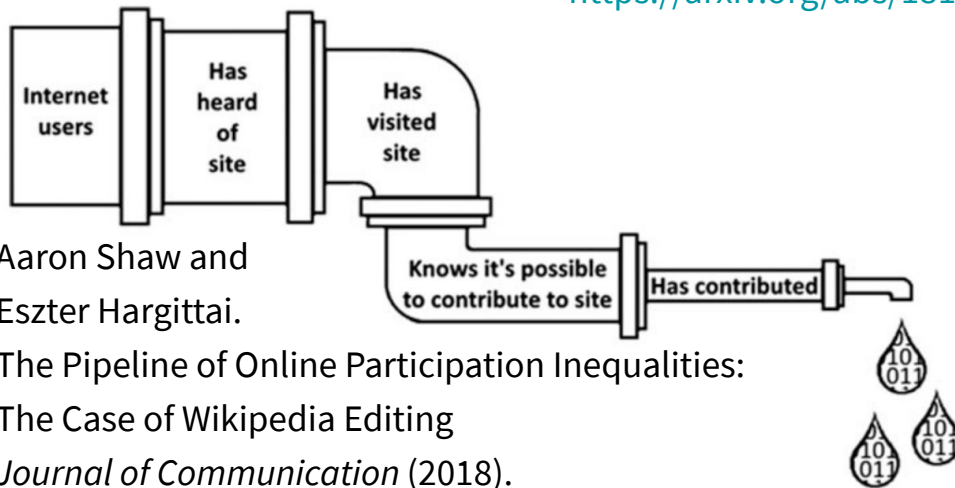
Overview

- Motivation / Narrative
- How does this survey work?
- Results
 - Age
 - Gender
 - Language
- Takeaways

Motivation / Narrative



Why the World Reads Wikipedia
<https://arxiv.org/abs/1812.00474>



Aaron Shaw and
 Eszter Hargittai.

The Pipeline of Online Participation Inequalities:

The Case of Wikipedia Editing

Journal of Communication (2018).

<https://doi.org/10.1093/joc/jqx003>

WIKIPEDIA

Gender bias on Wikipedia

Gender bias on Wikipedia reflects findings that a dominant majority of volunteer Wikipedia editors, particularly on the [English-language site](#), are male. Also, there are fewer and less extensive articles about women or topics important to women. It figures among the most frequent [criticisms of Wikipedia](#), sometimes as part of a more general criticism about [systemic bias in Wikipedia](#). The [Wikimedia Foundation](#), which runs Wikipedia, agrees with these criticisms and has made an ongoing attempt to increase female editorship of Wikipedia. [Edit-a-thons](#) have been held to encourage female editors and increase the coverage of women's topics. [\[1\]\[2\]](#)

en:Gender_bias_on_Wikipedia

Narrative -- Gender Example

1. Men and women have similar high-level behavior and needs.
2. We do, however, see different interests in specific topics.
3. This data raises a number of interesting hypotheses around the relationship between readership gaps and content (or contributor) gaps.

The Survey

1 Week

26 June - 1 July 2019

13 Languages

Arabic, German, **English**, Spanish, Persian, **French**, Hebrew, Hungarian, Norwegian, Romanian, Russian, Ukrainian, Chinese

3 Motivation and 5 Demographic Questions

motivation, information need, familiarity

age, gender, education, locale (urban/rural), native languages

64,343 Responses

including large focus on Africa through English, French, and Arabic

What do I mean by "readers"?

Daily readers? Weekly readers? Monthly readers? People who have visited Wikipedia at some point?

Example -- Pew Research 2010 Survey in the United States:

- "Do you read Wikipedia?" → 53% Men and 47% Women (~gender parity!)
- "Did you read Wikipedia yesterday?"
 - Only about a third of respondents say yes
 - 60% Men and 40% Women

Survey Bias -- Arabic Wikipedia

Post debiasing so that
survey results are more
representative



Reader Session Type	Random	Survey	%-Women
Overall	1000	1000	27.1% 24.2%
Logged-in	2	24 (10x)	12.9%
Read a single page	648	354 (0.6x)	29.1%
Read 2-4 pages	271	366 (1.4x)	24.4%
Read >5 pages	78	256 (3.3x)	18.1%
Viewed Main Page	8	37 (4.8x)	11.3%
Multilingual	64	112 (1.7x)	21.5%
Reading on mobile device	771	867 (1.1x)	24.9%

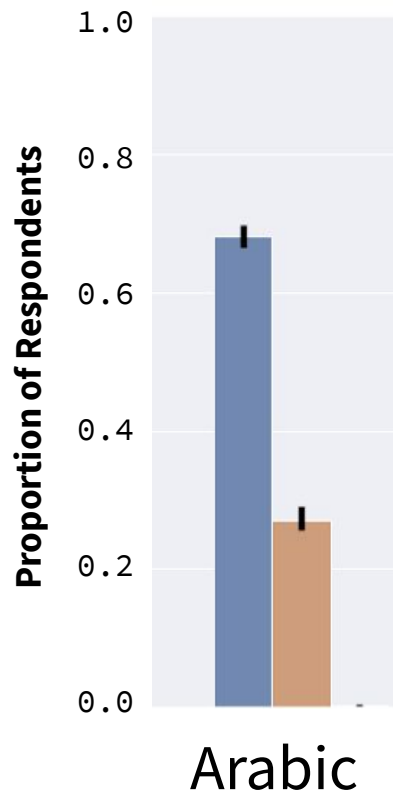
Motivation
Information Need
Prior Familiarity
Age
Gender
Education
Locale
Language

Full results: <https://w.wiki/CJg>

What is your gender?

- ⌘ Woman
- ⌘ Man
- ⌘ Prefer not to say
- ⌘ Other... [open-text response]

Gender of Respondents by Survey



Legend

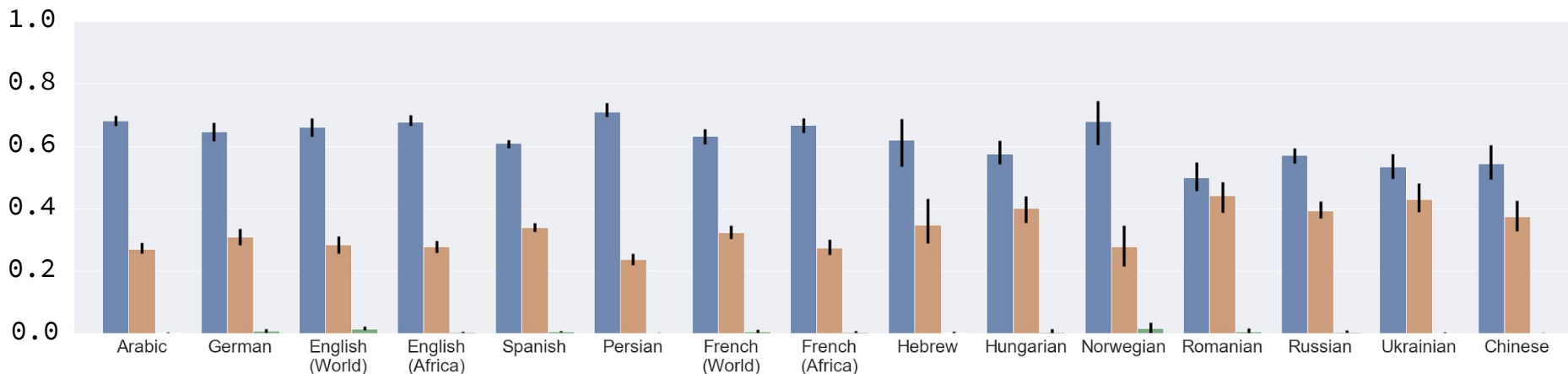


Gender of Respondents by Survey

Men

Women

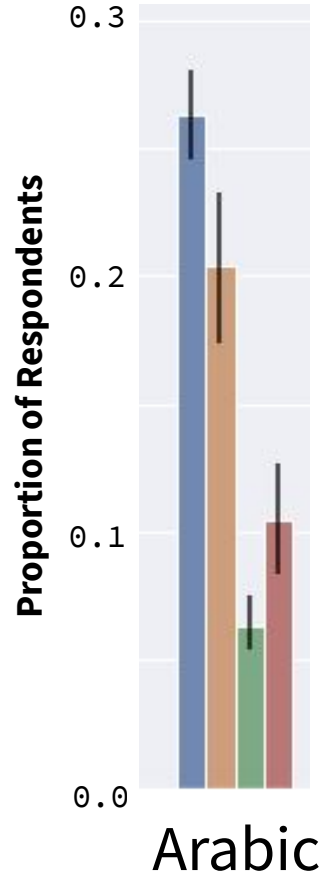
Self-Describe



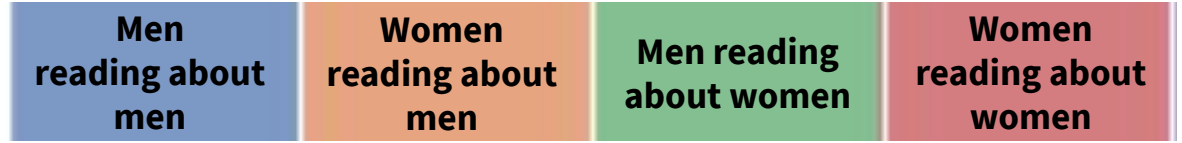
- The majority of respondents in almost every language identified as **men** (exception is Romanian)
- For many languages, younger readers are more balanced in gender. Also a lot of variance across countries.
- Small proportion of respondents chose to self-describe

Larger version: <https://w.wiki/CJg>

Gender of Readers vs. Gender of Biography

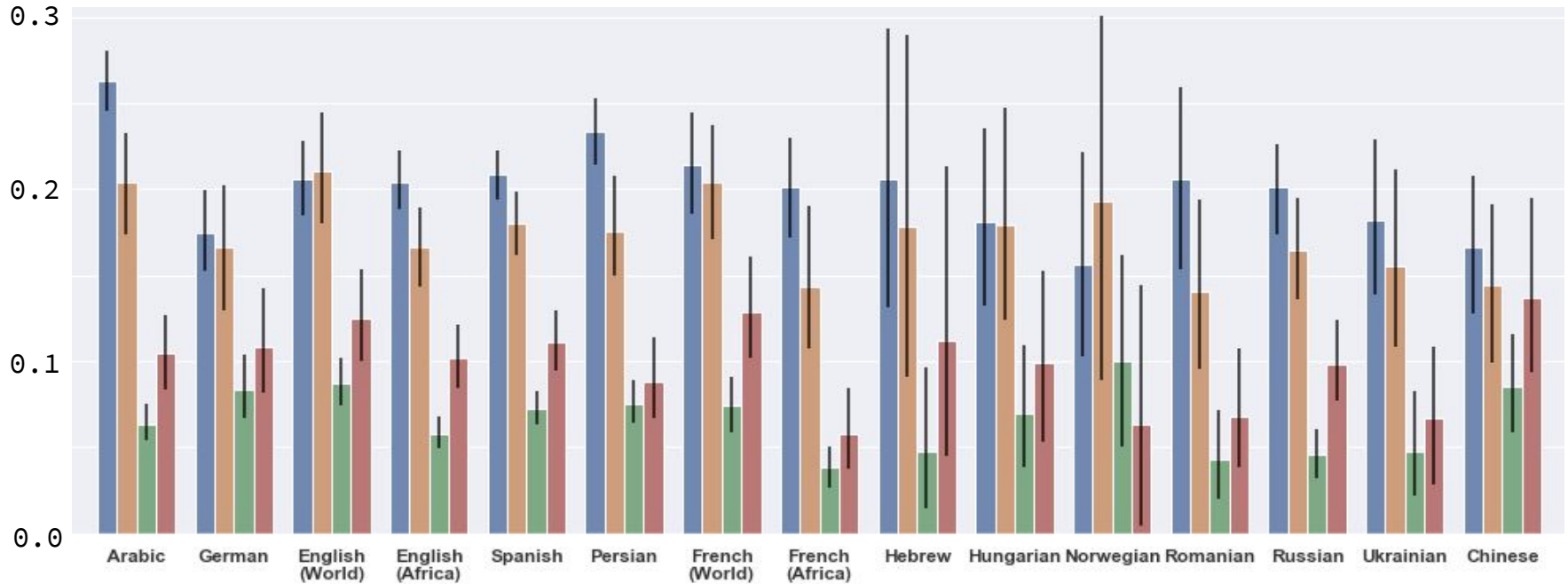


Legend



Gender of Reader vs. Gender of Biography

Proportion of pageviews



In many languages...

Men read relatively more biographies of men than **women do** and **women read relatively more biographies of women** than **men do**.

Topic vs. Gender

Out of the 15 surveys, example topics that trend towards...

Men Readers

Sports (12) -- e.g., 2019 Africa Cup of Nations, Copa América

Technology (10) -- e.g., YouTube, WhatsApp

Military/Warfare (10) -- e.g., Cold War, American Civil War

Women Readers

Medicine (9) -- e.g., Schizophrenia, Asperger syndrome

Broadcasting (8) -- e.g., Chernobyl (miniseries), Dark (TV series)

Biology (8) -- e.g., Oak processionary, Saccharum

Balanced Readership

Philosophy/Religion (15) -- e.g., Astrological sign, Seven deadly sins

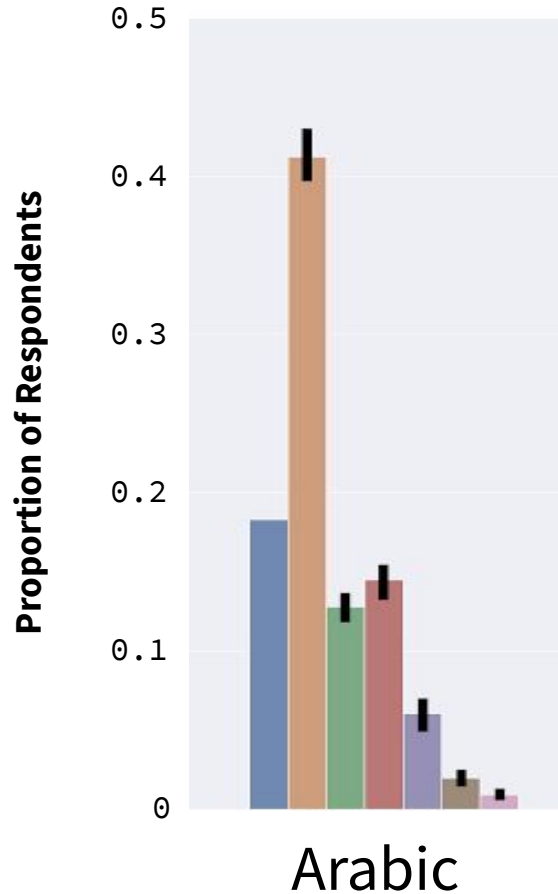
Geography/Americas (13) -- e.g., 2019 Copa América, United States

Biographies (12) -- e.g., Lisa Martinek, Deaths in 2019

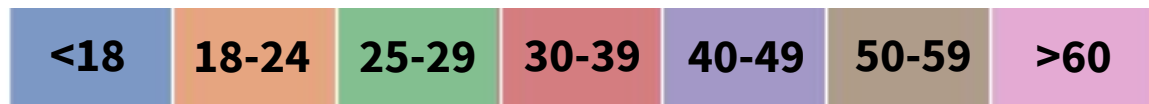
What is your age?

- α 18-24 years
- α 25-29 years
- α 30-39 years
- α 40-49 years
- α 50-59 years
- α 60 years and older
- α Prefer not to say

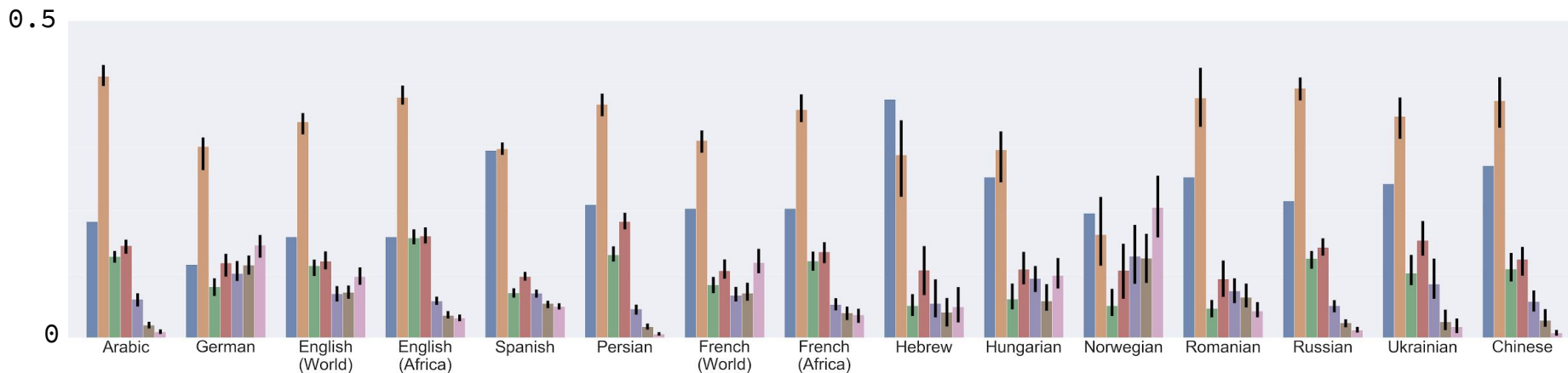
Age of Respondents by Survey



Legend



Age of Respondents by Survey



- The majority of respondents in almost every language were less than 25 years of age
- German and Norwegian are exceptions to this general trend
- Spanish and Hebrew -- school in session?

Larger version: <https://w.wiki/CJg>

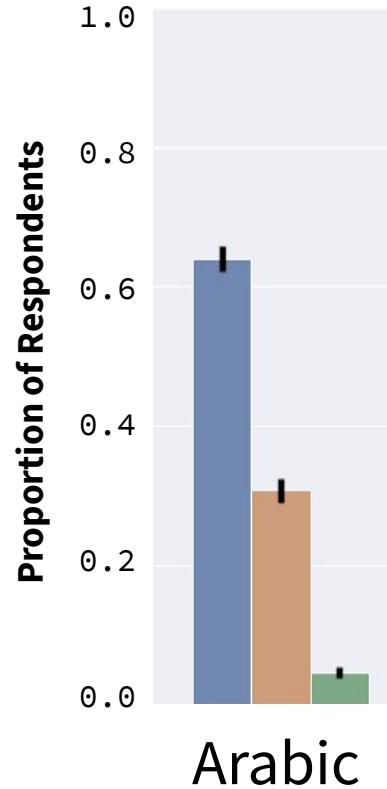
What is your native language?

- ☐ English -- en
- ☐ АԥԾԿԻՅԱ -- ab
- ☐ Bahasa Acèh -- ace
- ...
- ☐ isiZulu -- zu
- ☐ Other
- ☐ Prefer not to say

What is your second native language?

- ✕ I do not have a second native language
- ✕ Other... [open-text response]

Language of Respondents by Survey



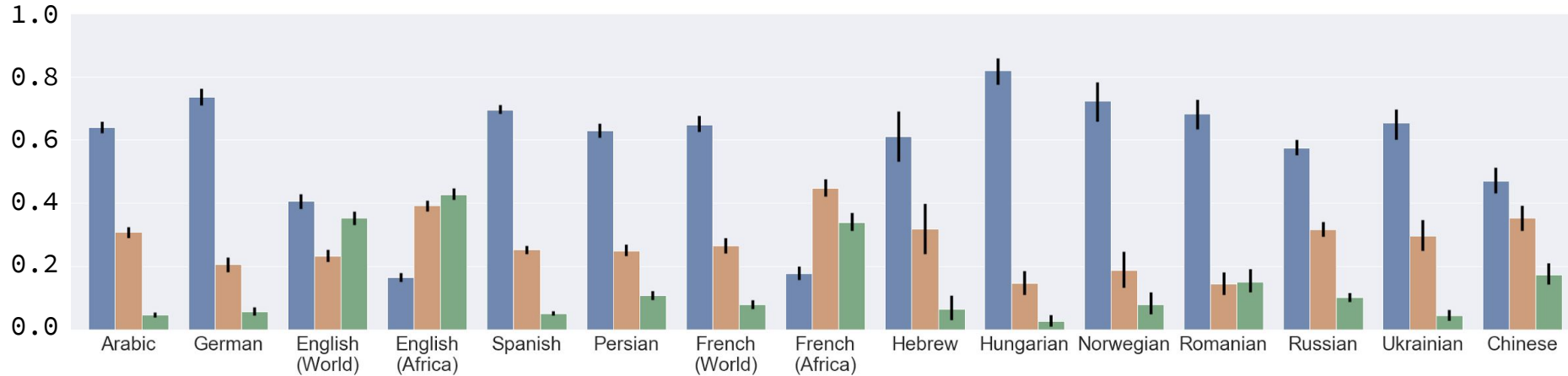
Legend

**Monolingual:
Native
Speaker**

**Multilingual:
Native
Speaker**

**Non-native
speaker**

Language of Respondents by Survey



- Wikipedias like Hungarian and Norwegian have readers who are predominantly **native speakers**
- Wikipedias like English and French are associated with much higher proportions of **non-native readers**

Larger version: <https://w.wiki/CJg>

Takeaways

Past Efforts

- 2008 UNU-MERIT Survey: <https://w.wiki/76w>
 - Survey distributed on wiki: 20 languages, 170k responses
 - Demographics, contributions, motivations, etc.
- 2011 Reader Surveys: <https://w.wiki/7Xx>
 - Survey distributed off wiki to representative panel in 16 countries
 - Demographics, usage, etc.
- 2016 New Readers Phone Surveys: <https://w.wiki/76v>
 - Phone survey to representative panel in Nigeria, India, Mexico, Brazil, Egypt, Iraq
 - Gender, age, location, usage, etc.
- 2017 Reader Motivation Surveys: <https://w.wiki/76u>
 - Survey distributed on wiki: 14 languages, 217k responses
 - Motivation, information need, familiarity

Reader Diversity

Wikipedia communities differ greatly -- most notably in whether their readers are monolingual, multilingual, and native/secondary speakers of the language.

Pipeline to Online Participation

In many regions, we cannot address gaps in our editor population without also addressing gaps in our reader population.

Readers vs. Page Views

While men and women may be equally aware of Wikipedia in many countries, this survey indicate that metrics like page views likely represent a more skewed population.

Representation Matters

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

Thank you! Questions?

Meta page with more info / data: <https://w.wiki/CJg>

Me: Isaac (WMF) -- isaac@wikimedia.org

Collaborators: Florian Lemmerich (RWTH Aachen), Diego Sáez-Trumper (WMF), Markus Strohmaier (RWTH Aachen / GESIS), Bob West (EPFL), Leila Zia (WMF)

...with additional help from: Abbad (Volunteer), Amire80 (Volunteer), Yury Bulka (Volunteer), Adélaïde Calais (WMFr), Kaganer (Volunteer), Liang-chih ShangKuan (Volunteer), Jon Harald Søby (WMNO), Strainu (Volunteer), Tgr (Volunteer), Morten Warncke-Wang (WMF), Readers Web, Legal, and probably a few others that I am forgetting