

Queering Wikidata

Early Insights from the Wikidata Gender Diversity Project

Daniele Metilli (they/them) University College London, UK

Beatrice Melis (she/her) University of Pisa, Italy

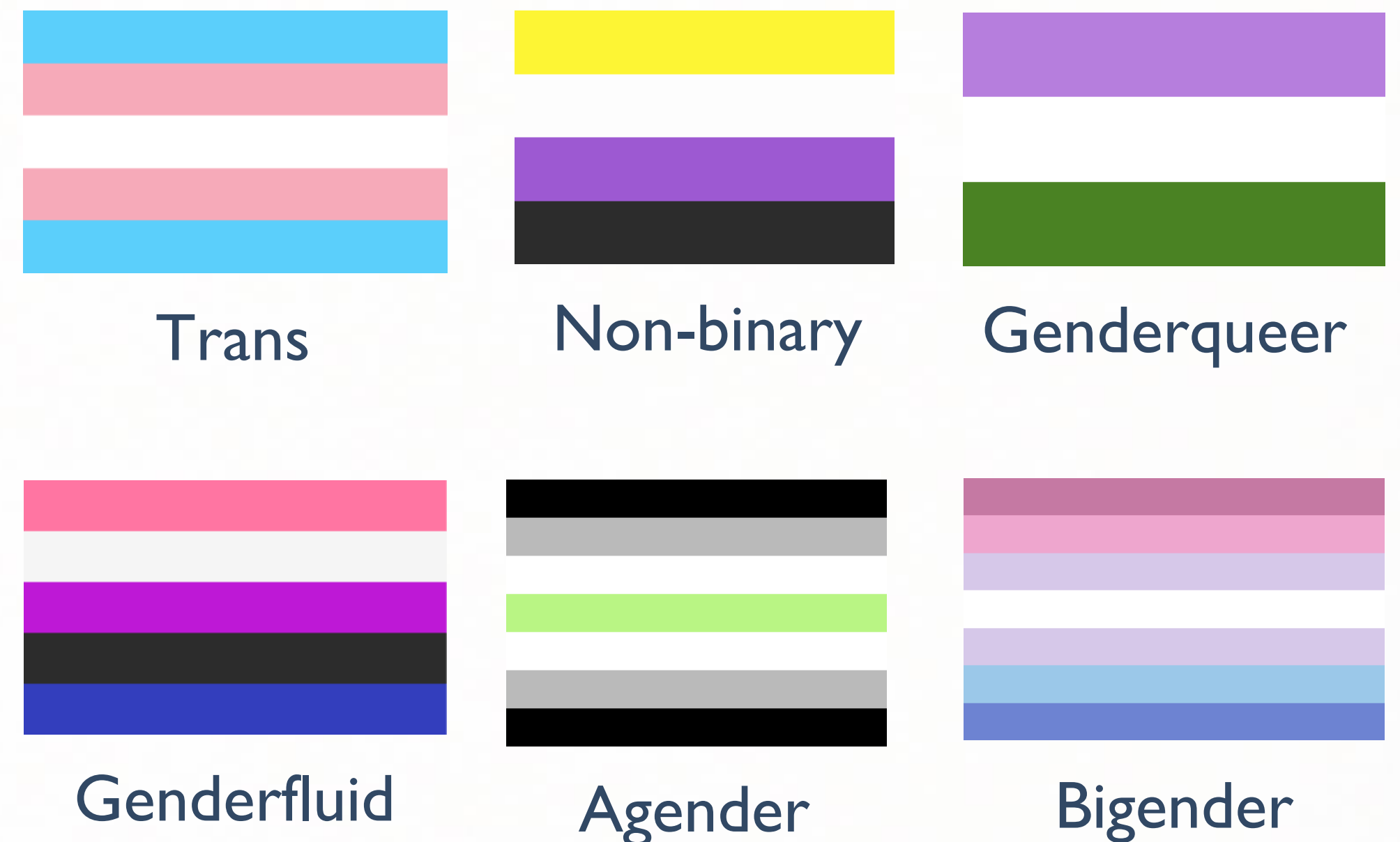
Marta Fioravanti (she/her) oio.studio, UK

Chiara Paolini (she/her) KU Leuven, Belgium

Wikidata Gender Diversity (WiGeDi)

Wikidata Gender Diversity is a one-year project funded through the **Wikimedia Research Fund**

Looking at **gender diversity** in **Wikidata**, focusing on **marginalized gender identities** from a queer and intersectional feminist perspective



...and many more

Our Motivation

Identities that fall outside the **gender binary** suffer **discrimination** and **systematic erasure** in society

This impacts **trans people, non-binary people, intersex people**, and anyone who does not conform to the traditional Western view of gender



“BLM-Demo Hof 20200620 DSC2232” by PantheraLeo1359531

(https://commons.wikimedia.org/wiki/File:BLM-Demo_Hof_20200620_DSC2232.jpg), cropped and color-corrected, Creative Commons BY 4.0

Some Facts

Gender is not a male/female binary

Gender is very complex

Gender diversity is not a new phenomenon



The Gender Unicorn Graphic by: **TSER**
Trans Student Educational Resources

Gender Identity

- Female/Woman/Girl
- Male/Man/Boy
- Other Gender(s)

Gender Expression

- Feminine
- Masculine
- Other

Sex Assigned at Birth

- Female
- Male
- Other/Intersex

Physically Attracted to

- Women
- Men
- Other Gender(s)

Emotionally Attracted to

- Women
- Men
- Other Gender(s)

To learn more, go to:
www.transstudent.org/gender

Design by Landyn Pan and Anna Moore

(from <https://transstudent.org/gender>,

CC-BY license)

Research Questions

There are **three research questions** that we are interested in:

1. **modeling of gender** in the Wikidata ontology



How

2. representation of **gender data** in the knowledge base



What

3. **discussions about gender** by the community



Why



Project Outcomes

Wikidata Gender Dashboard

A complete timeline
of gender modeling
in Wikidata

①

Wikidata Gender Timeline



Accurate statistics
about gender

identities in Wikidata

A corpus of user
discussions about
gender

③

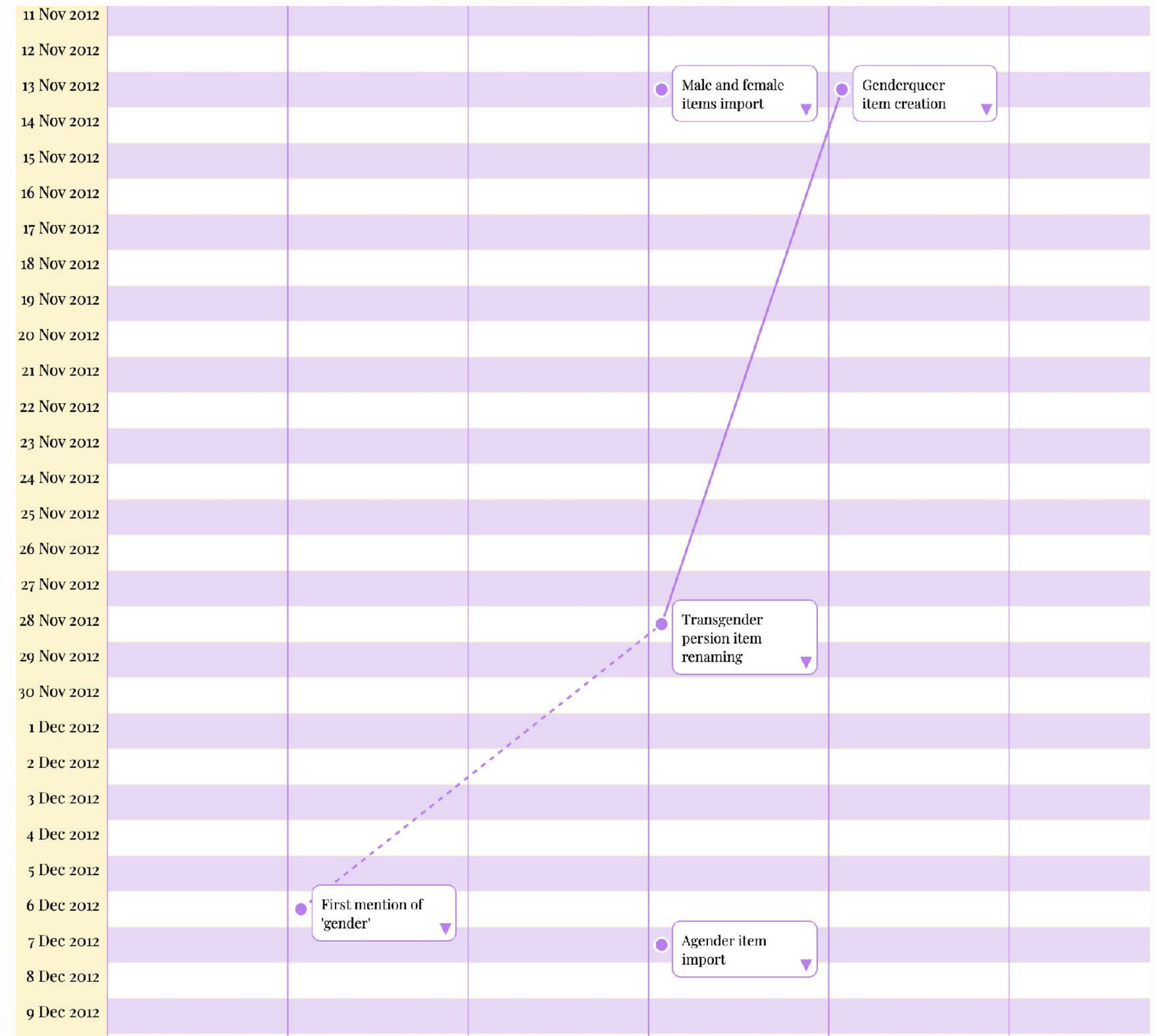
Wikidata Gender Talk

Wikidata Gender Timeline

Interactive timeline showing changes to the model, creation of relevant entities, edits, etc.

Events are **linked** to each other, **classified** and **contextualized**

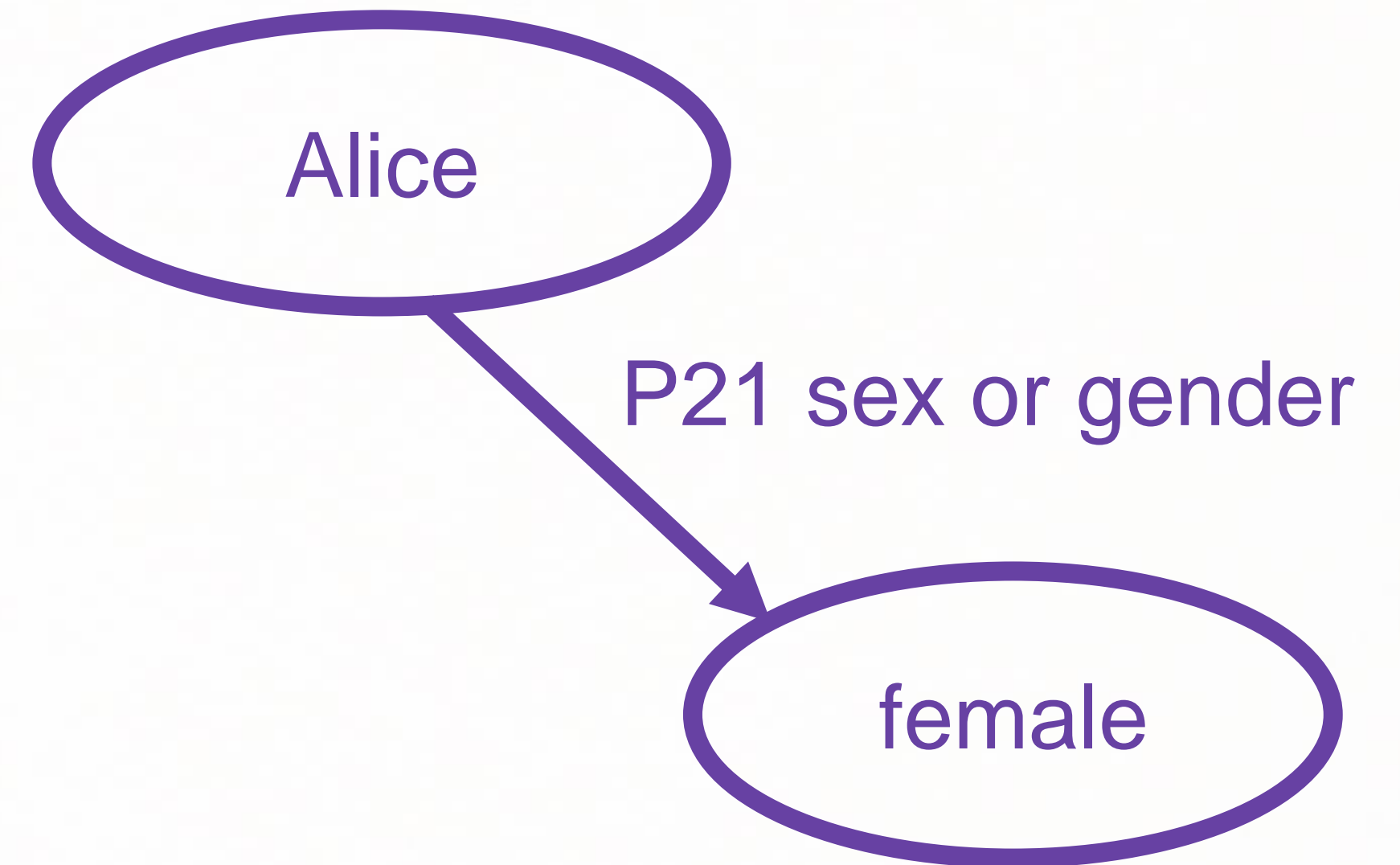
Will be released **publicly** after the dataset is finalized



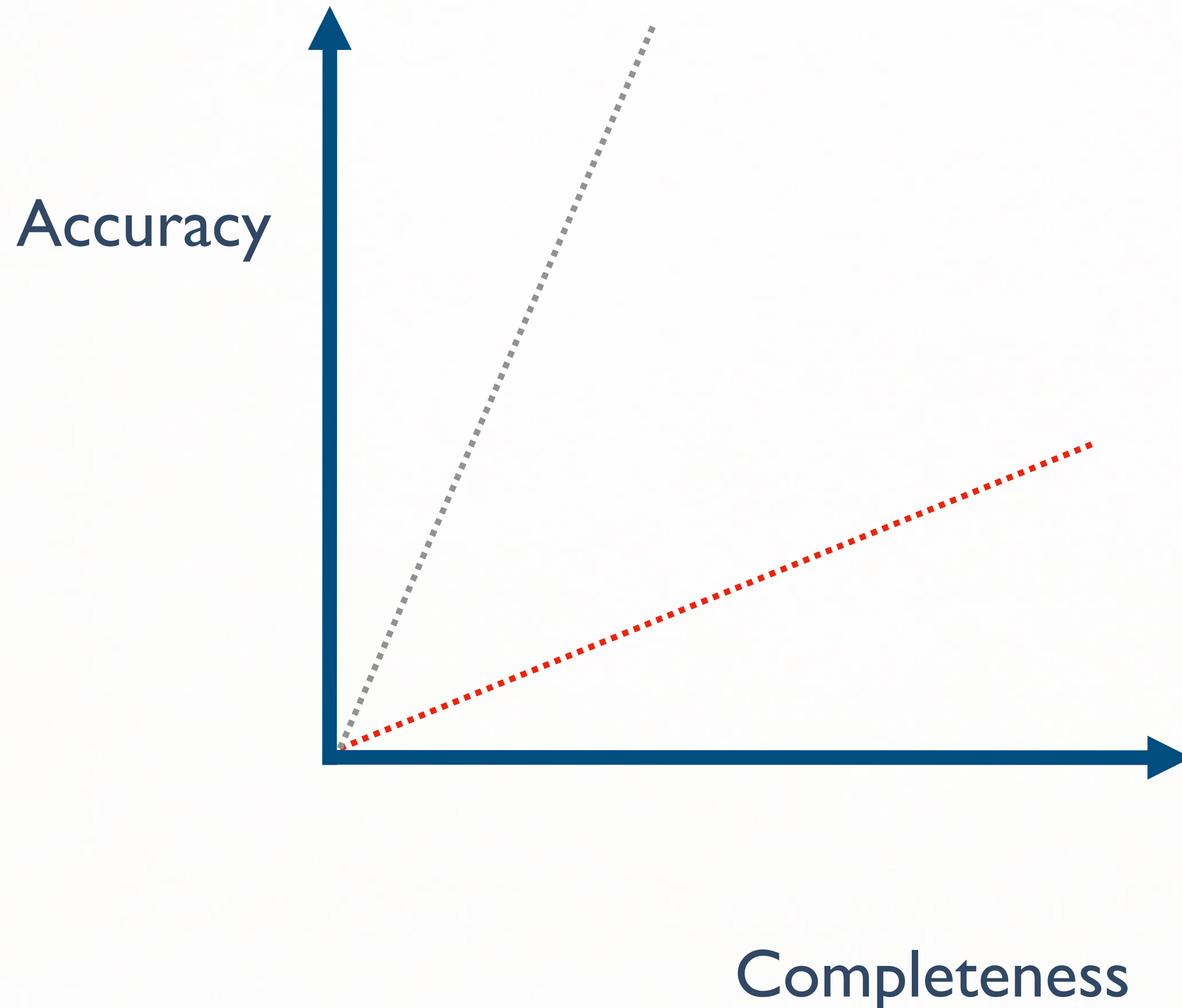
The P21 Property

The **property** that is used to represent gender in Wikidata is **P21**, which in English is labeled “**sex or gender**”

Ambiguity: does it refer to **sex assigned at birth** or to **gender identity**?



Completeness vs. Accuracy



Some users wanted to achieve **maximum accuracy...**

...but others were more interested in **maximum completeness**

For gender, completeness won:

- 70% in 6 months
- 95% in 2 years

... but at what cost?

Early Missteps

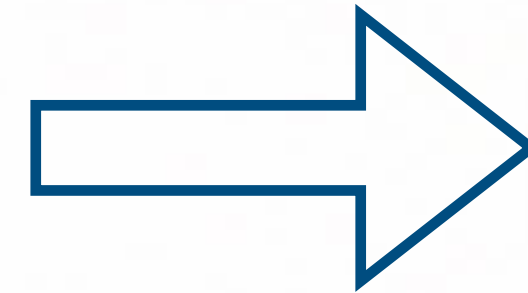
The initial (problematic) approach was:

- **every human** must have a **sex or gender...**
- ...and **only one**
- ...and only one of the **allowed values**
- in the beginning, only **male, female** and **intersex** were allowed



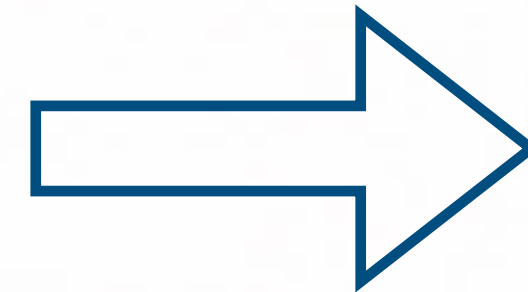
How Was Gender Populated?

Manually annotated



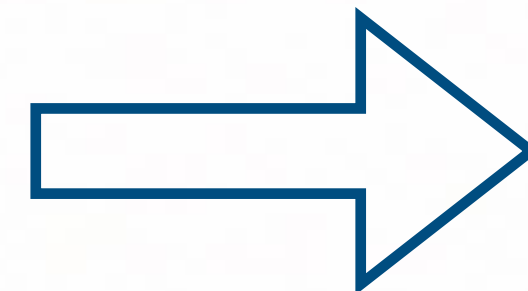
Gender is checked by the user

Imported from Wikipedia



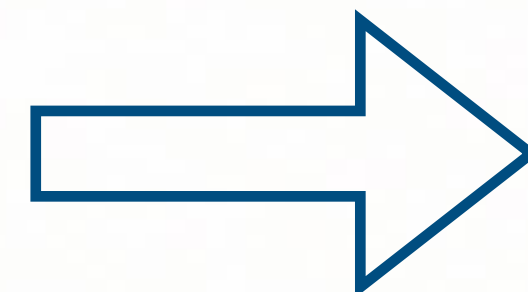
Not a reliable source

Imported from external databases



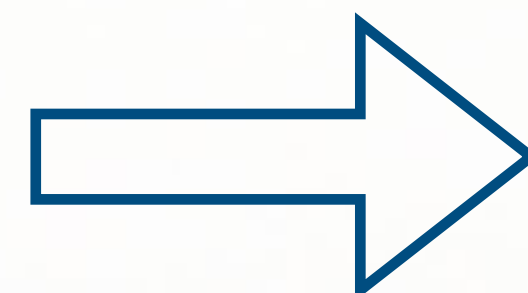
Reliable or not?

Imported from personal names



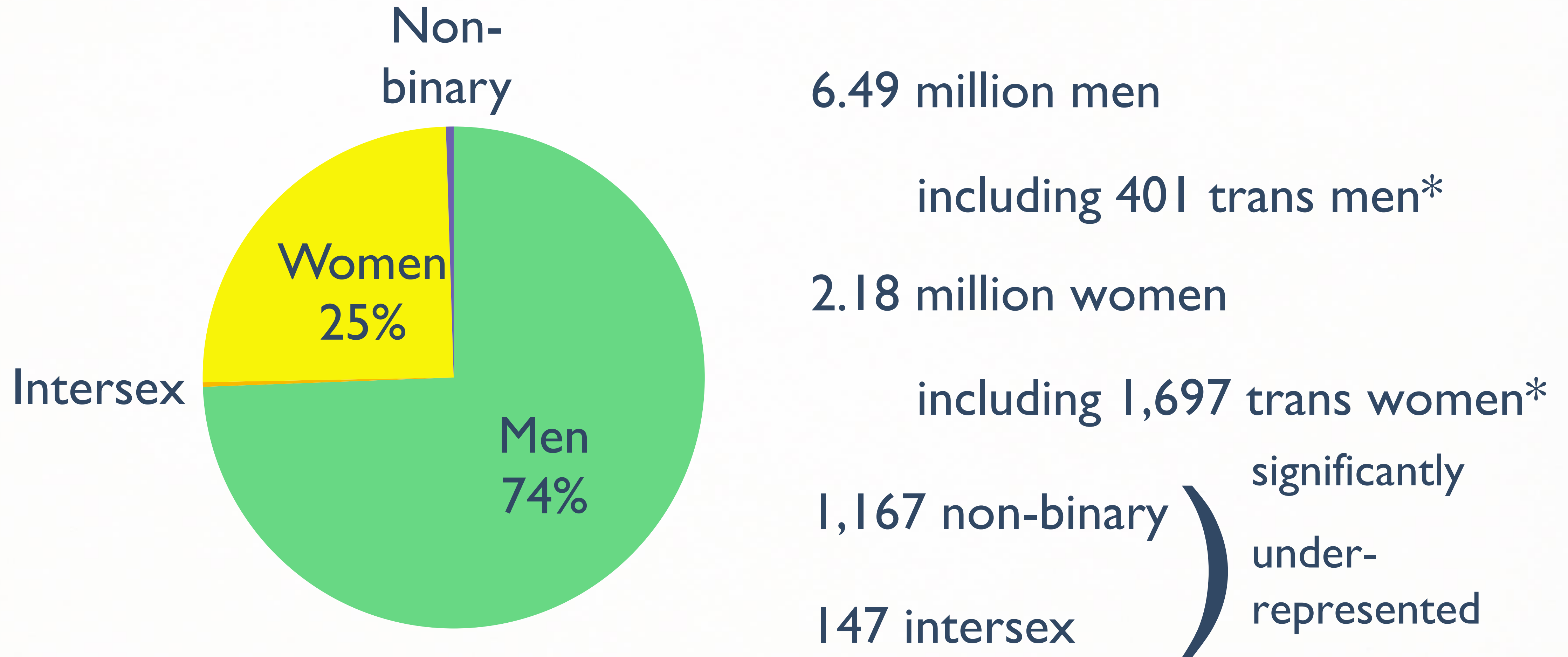
Very problematic!

Imported from pronouns



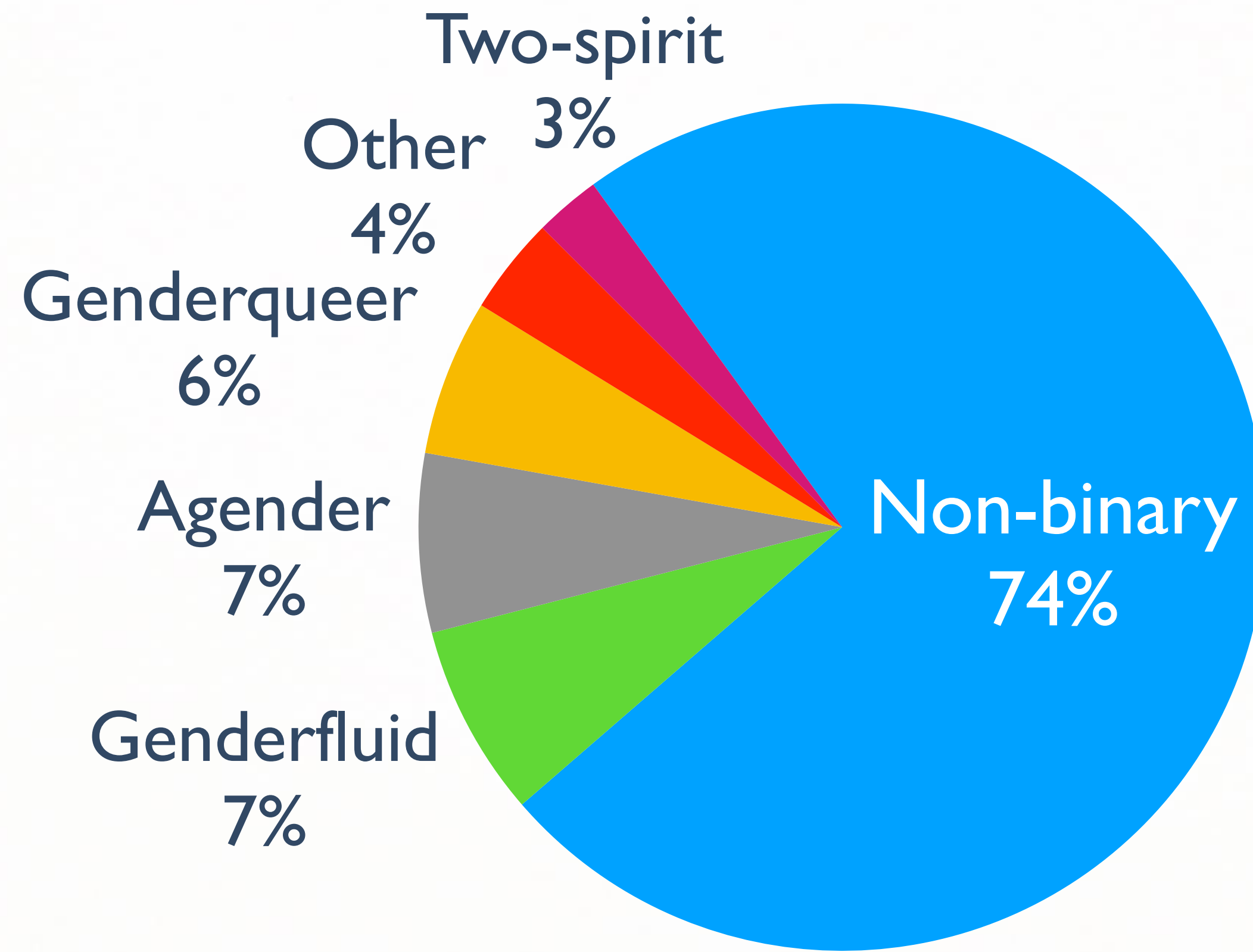
Very problematic!

Gender Identities in Wikidata



* Trans men are men, trans women are women

Non-binary Identities in Wikidata



852 non-binary

85 genderfluid

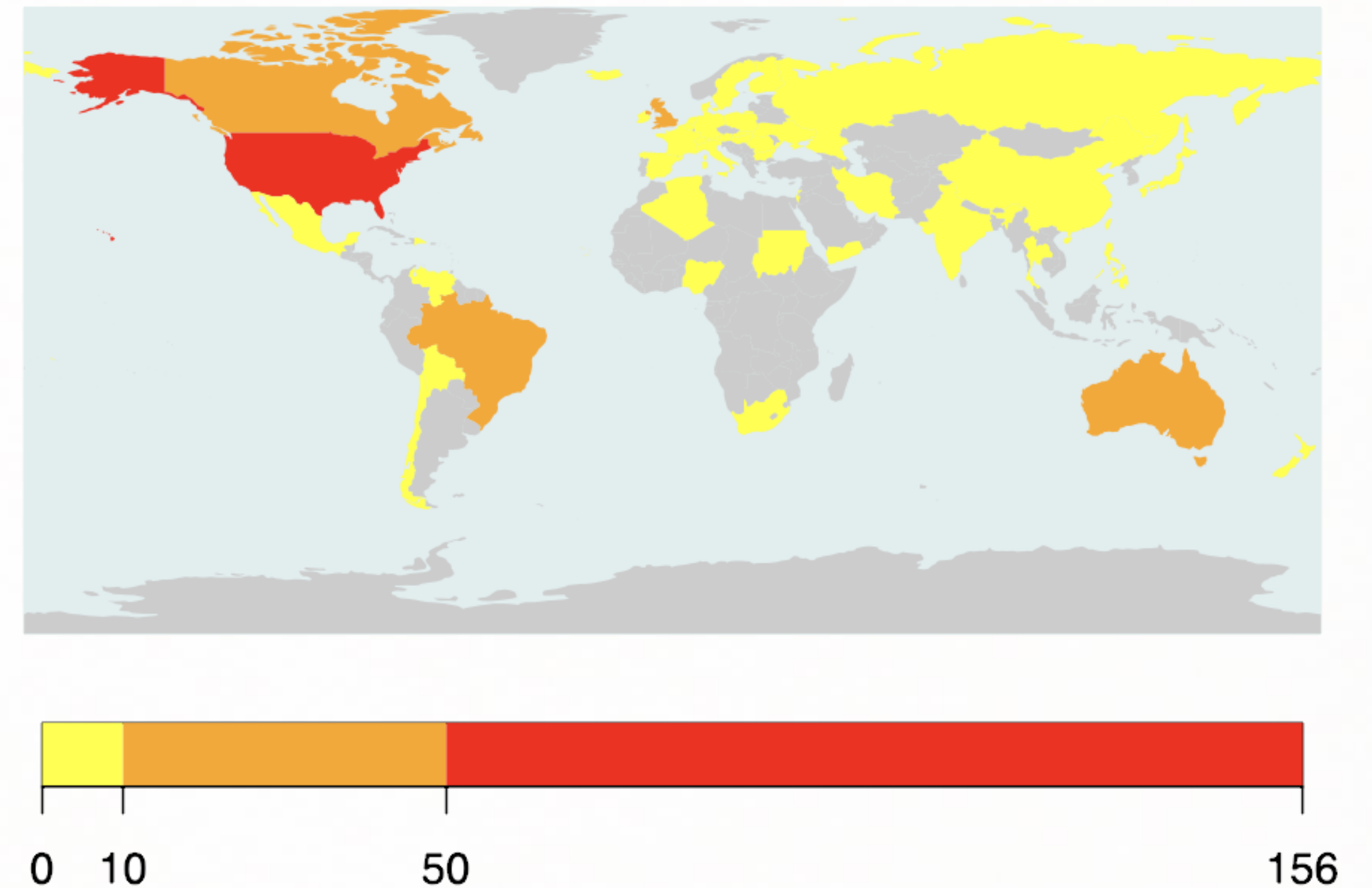
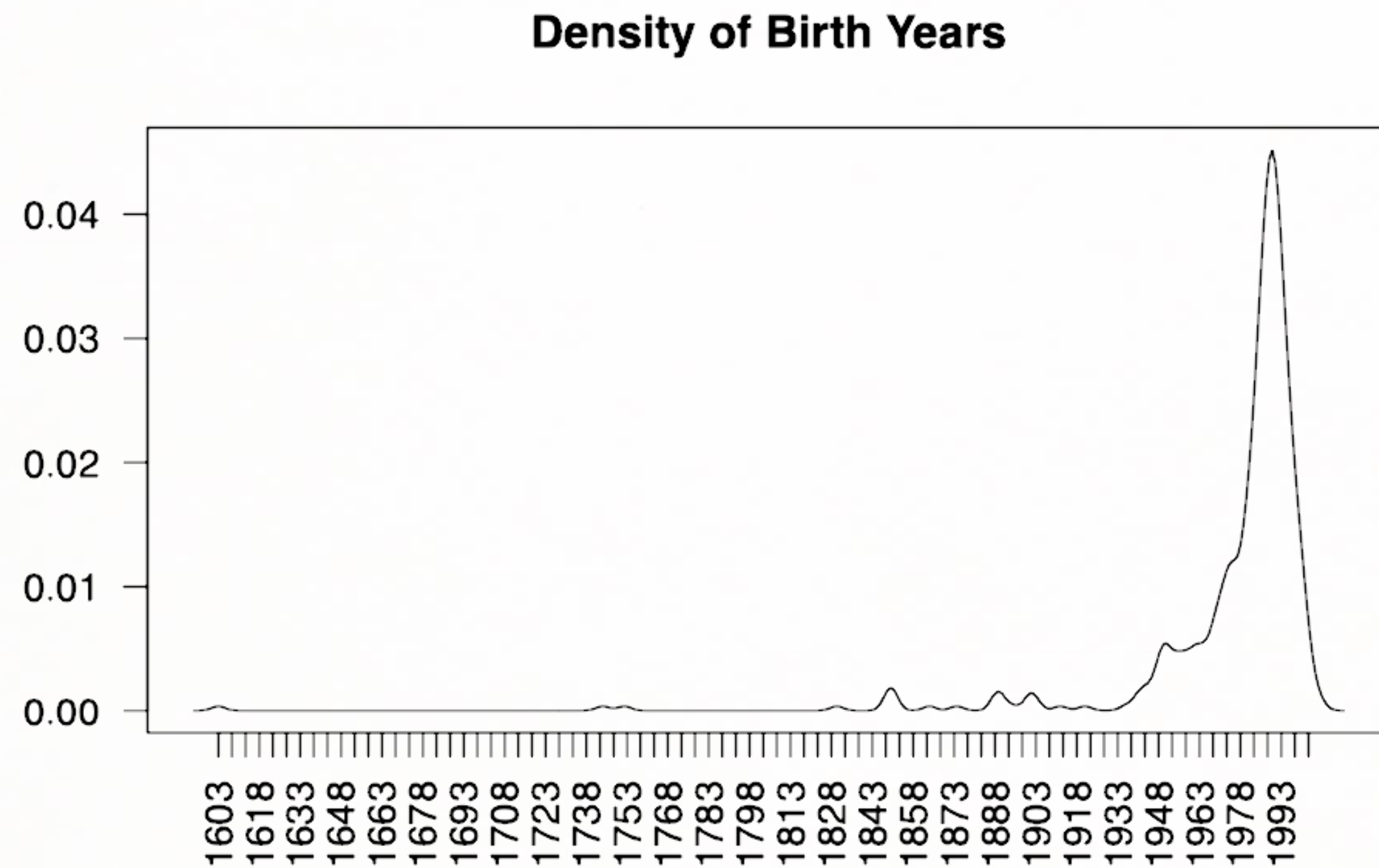
79 agender

69 genderqueer

29 two-spirit

+ 43 other

Non-binary People Through Time and Space



(based on data from early 2022, see slide 20 for our publication)

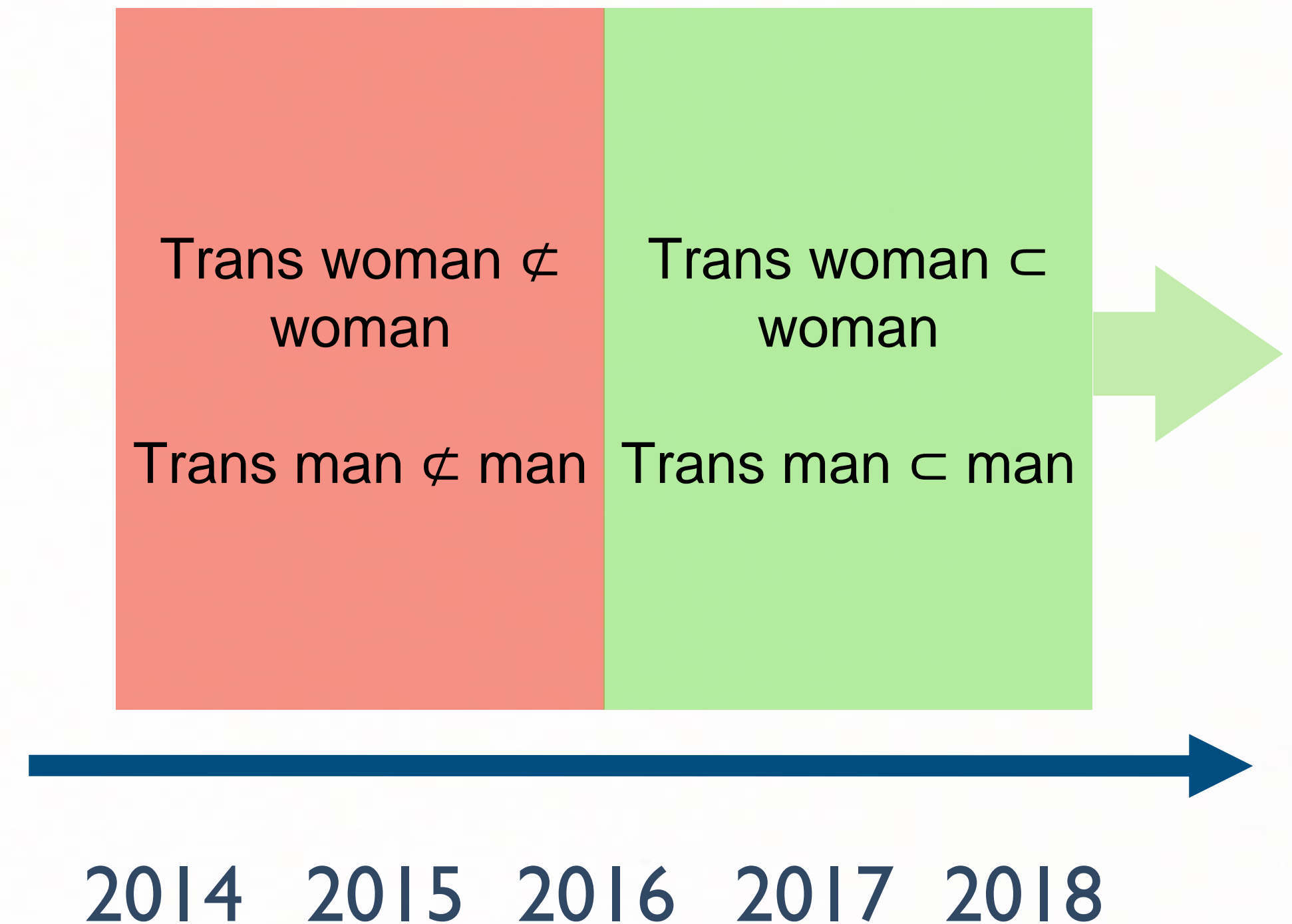
Trans Men and Women

Described using **transgender female** and **transgender man** entities

Recently re-labeled to **trans woman** and **trans man** in English (correct)

Wrongly classified until 2016

Several instances of **deadnaming** and edit wars related to **gender transition**



Gendered Properties

Gendered properties are properties that implicitly contain gender information

In the **early years** of Wikidata, these properties were created:

- mother/father  ~~2013 proposal~~ ~~2015 proposal~~ ~~2016 proposal~~
- brother/sister  replaced with “sibling” in 2017
- uncle/aunt  replaced with “relative” in 2013
- stepfather/stepmother  replaced with “stepparent” in 2017

Wikidata User Discussions

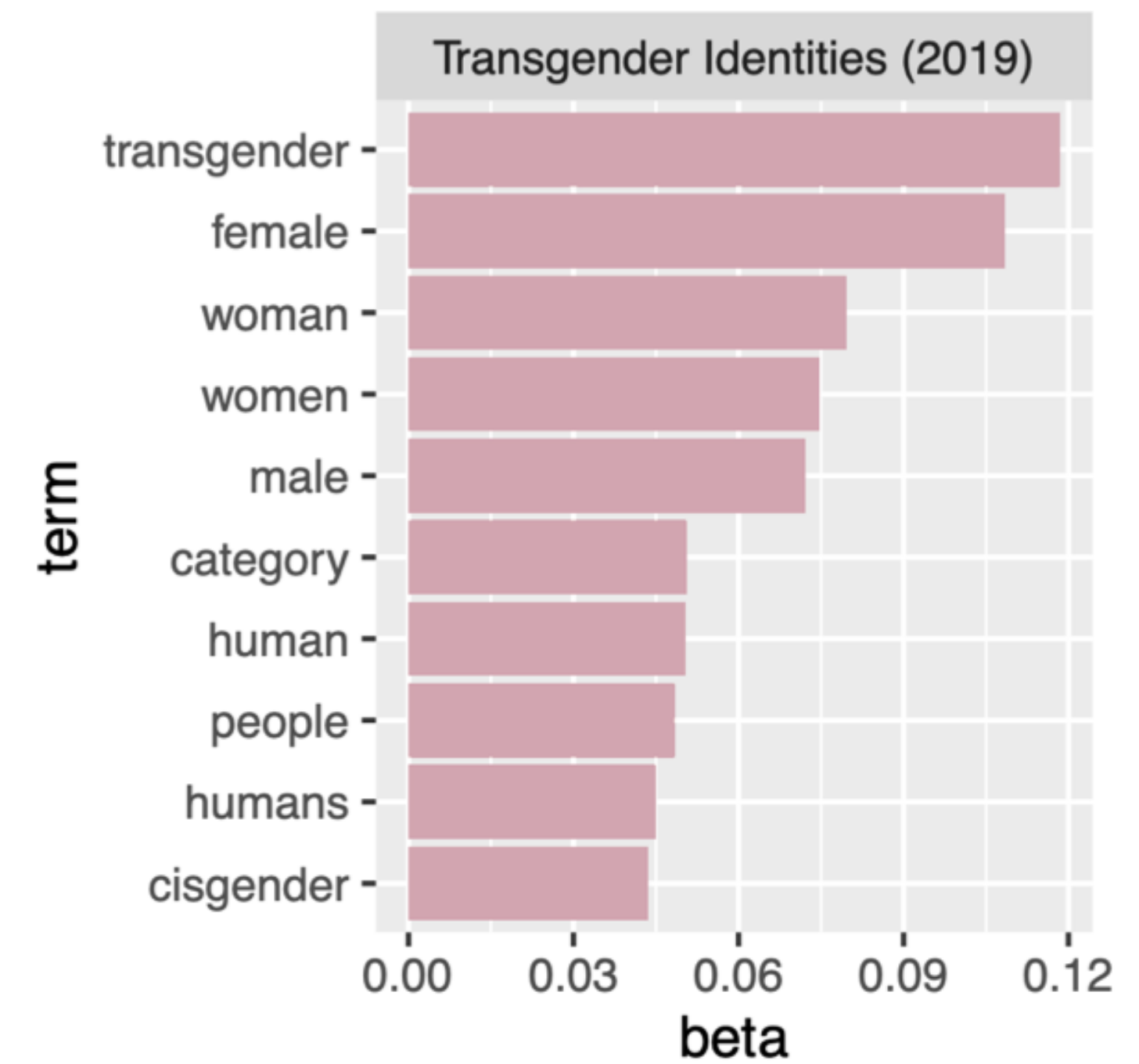
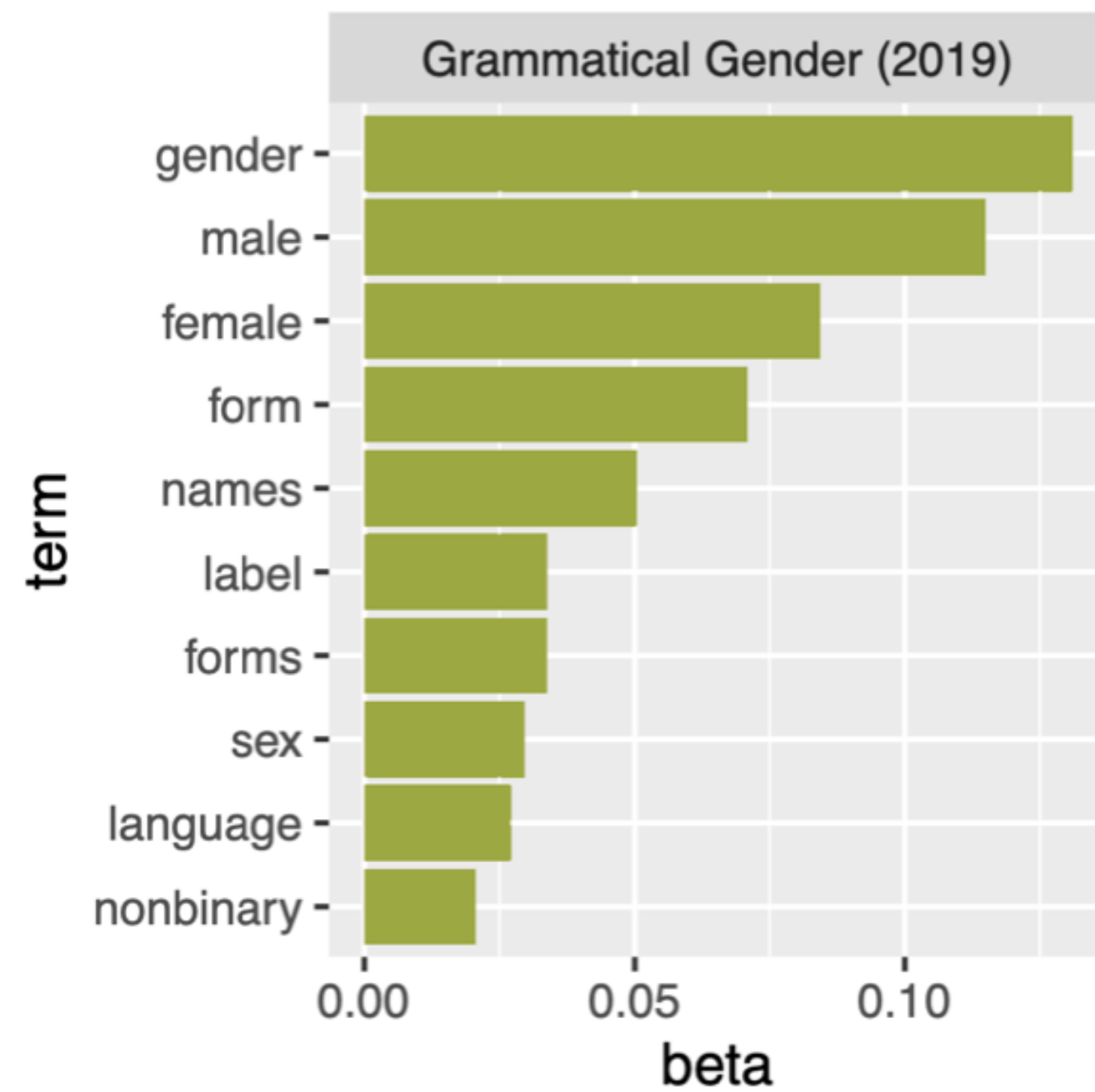
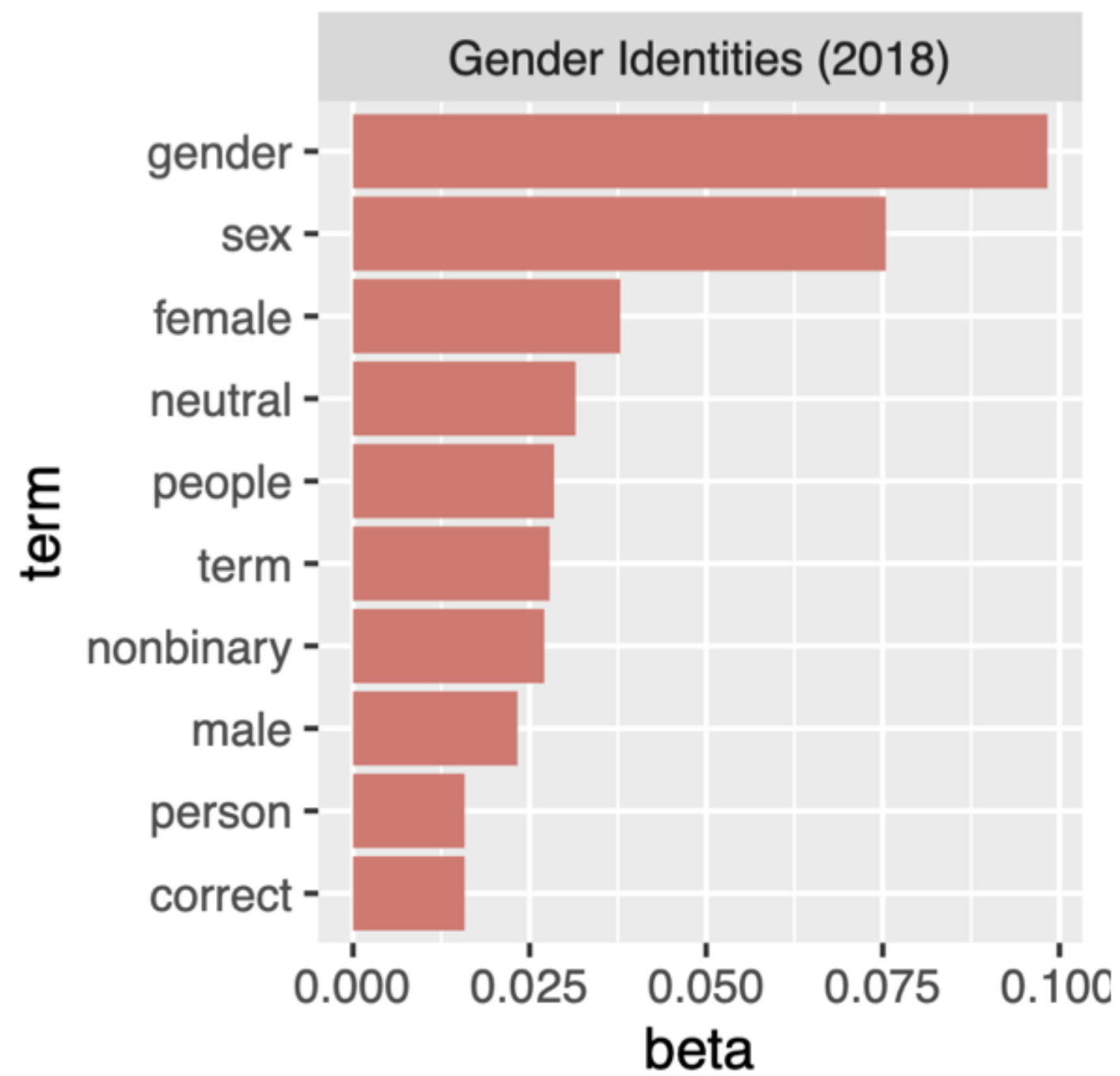
Corpus of Wikidata **user discussions**

Analysis based on **computational linguistics**,
using **discourse analysis** and **topic modeling** techniques

Quantitative and **qualitative** analysis of
discussions to reconstruct how the community
handled the representation of gender



Topic Modeling



(example from our initial publication, see slide 20)

Publications

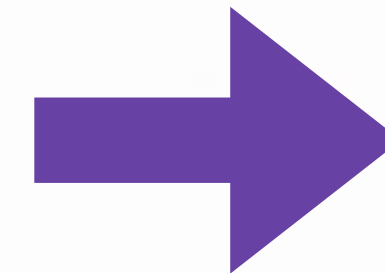
Ethics in Linked Data

(in press, Litwin Books)

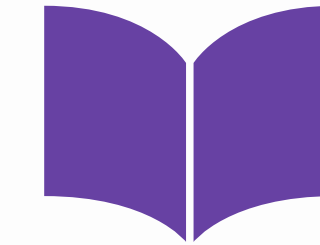


Editors:

- Alexandra Provo (she/her)
- Kathleen Burlingame (she/her)
- Bri M. Watson (they/m)



Our chapter



Authors:

Daniele Metilli & Chiara Paolini

“Non-binary gender representation in Wikidata”

<https://wigedi.com/chapter.pdf>

+ Abstract accepted for:

Special Issue of *Internet Histories* on Gender and Internet/Web Histories

Next Presentations

May 2023

- Wiki Workshop ✓
- Queering Wikipedia ✓

June 2023

- Nordic STS Conference
- Challenging the Binary
- Data Justice Conference
- Wikimedia Research Showcase

July 2023

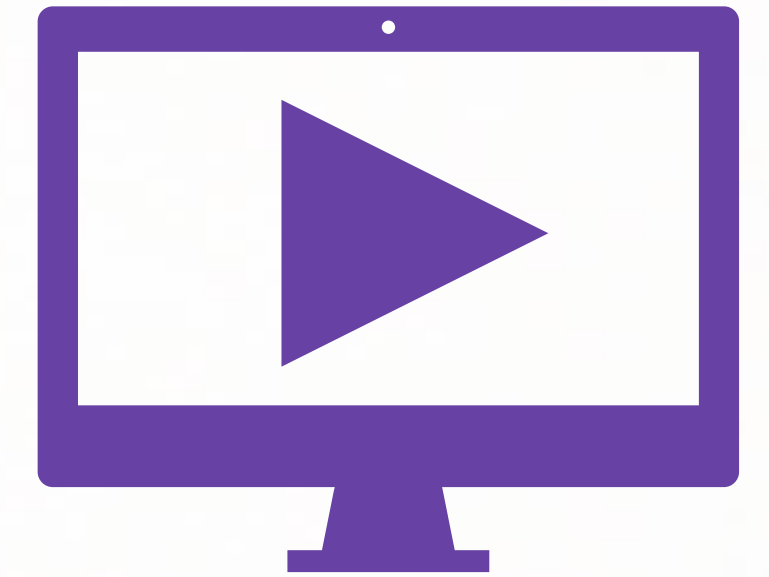
- IGALA Conference

August 2023

- Wikimania (submitted)

September 2023

- End-of-project event



Thank you

wigedi.com