



DATATHON

**Team
Centennial**



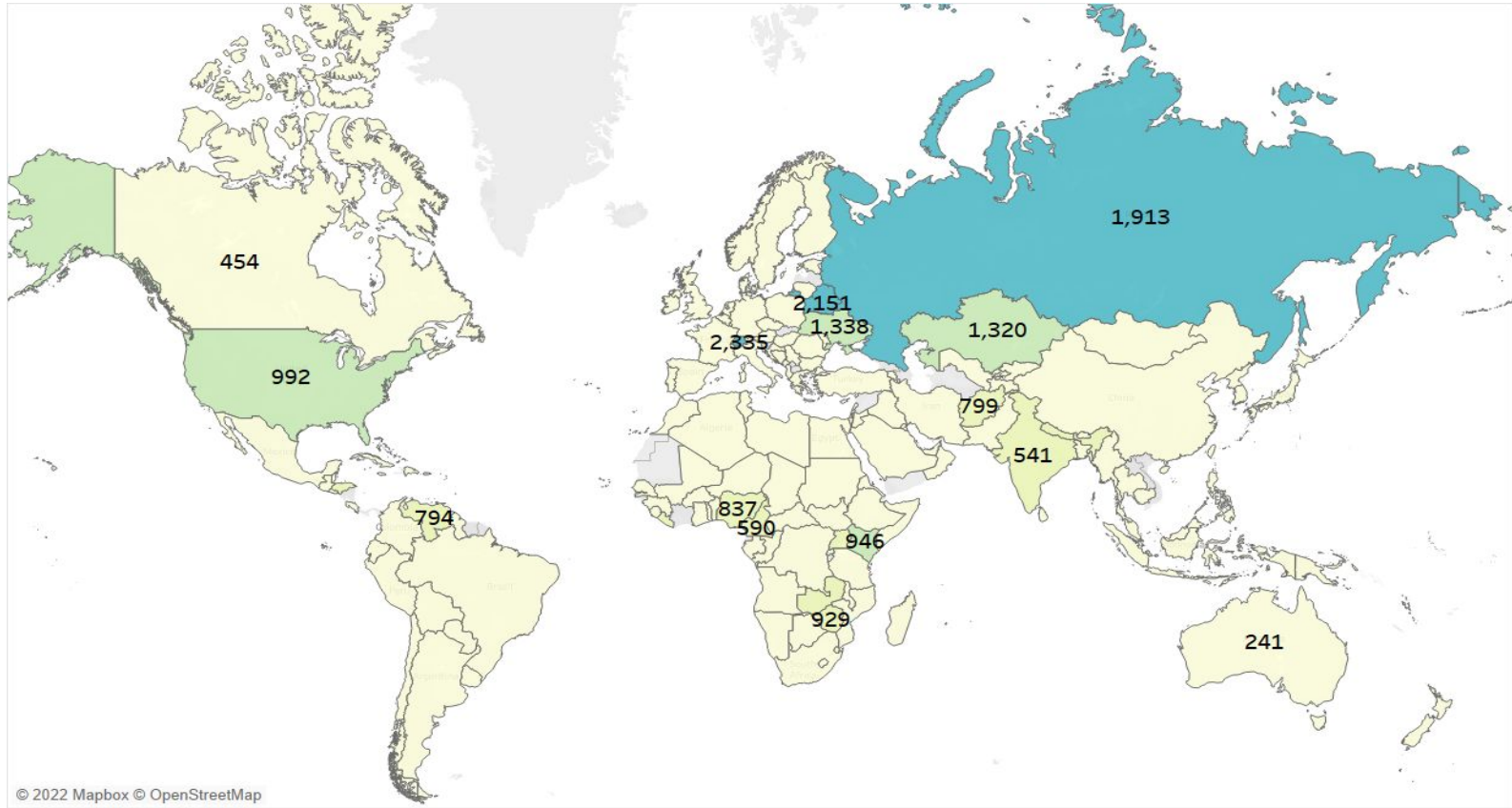
Data

Wrangling

Steps taken

- Approximately 7,000 records were removed because they were mostly incomplete.
- Data was transformed from wide form to long form.
- Several new calculated fields were created for analysis purposes. There were missing values for Gender - these were taken as Undisclosed. Similar steps were done for Race.

Responses by Country



24,051 total respondents

(not counting mostly incomplete surveys)

Large number of respondents from Switzerland, Belarus and Russia

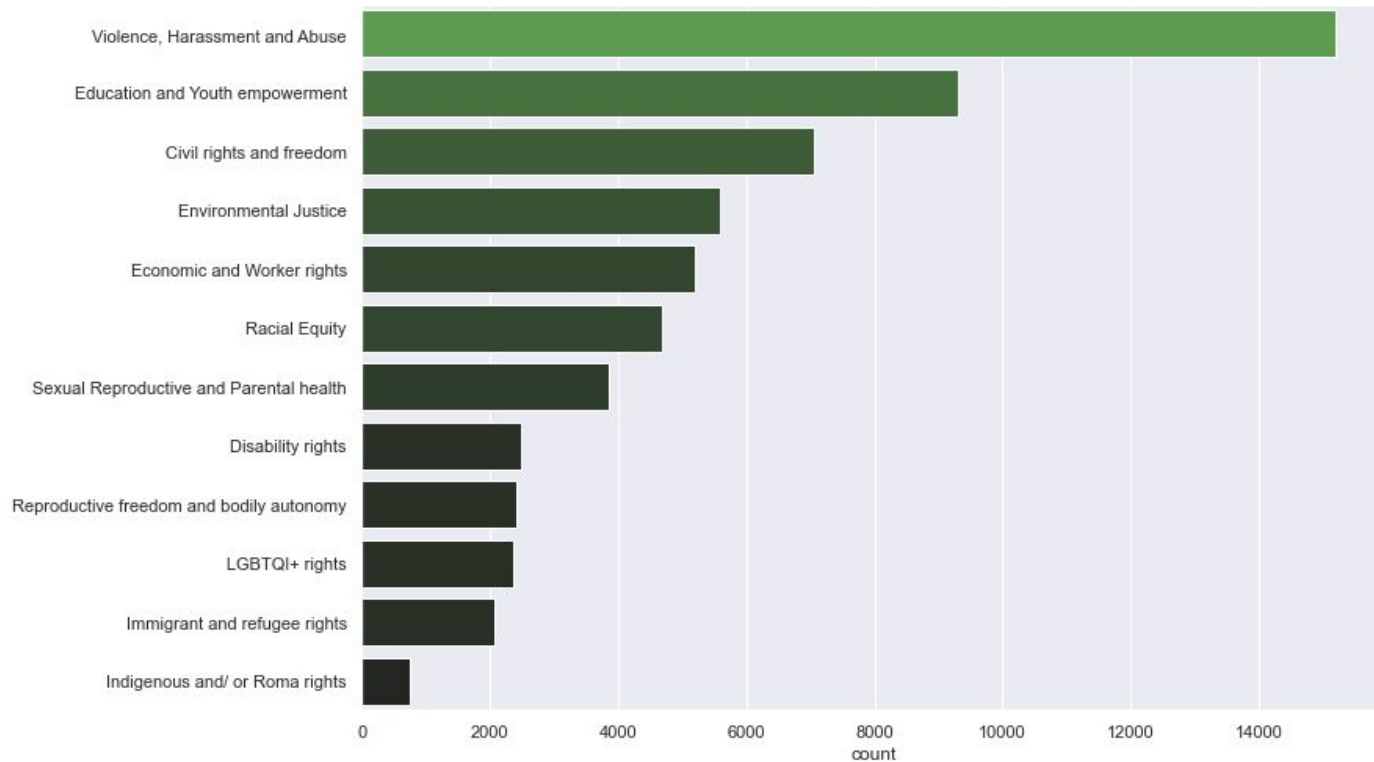
799 respondents from Afghanistan

1.

Critical Issues

Segment-wise
breakdown

Critical Issues faced



63% of respondents

chose Violence, Harassment & Abuse

39% of respondents

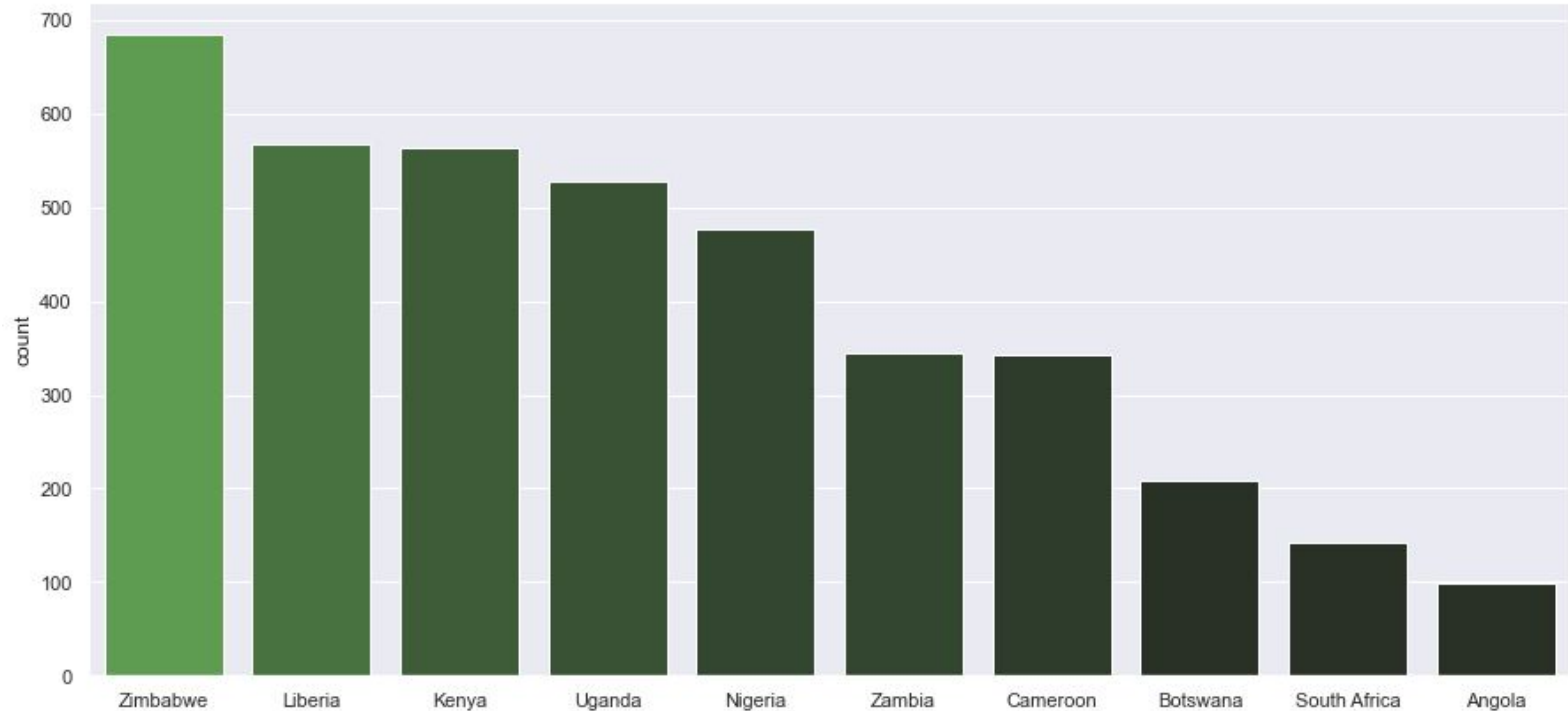
chose Education & Youth Empowerment

29% of respondents

chose Civil Rights and Freedom

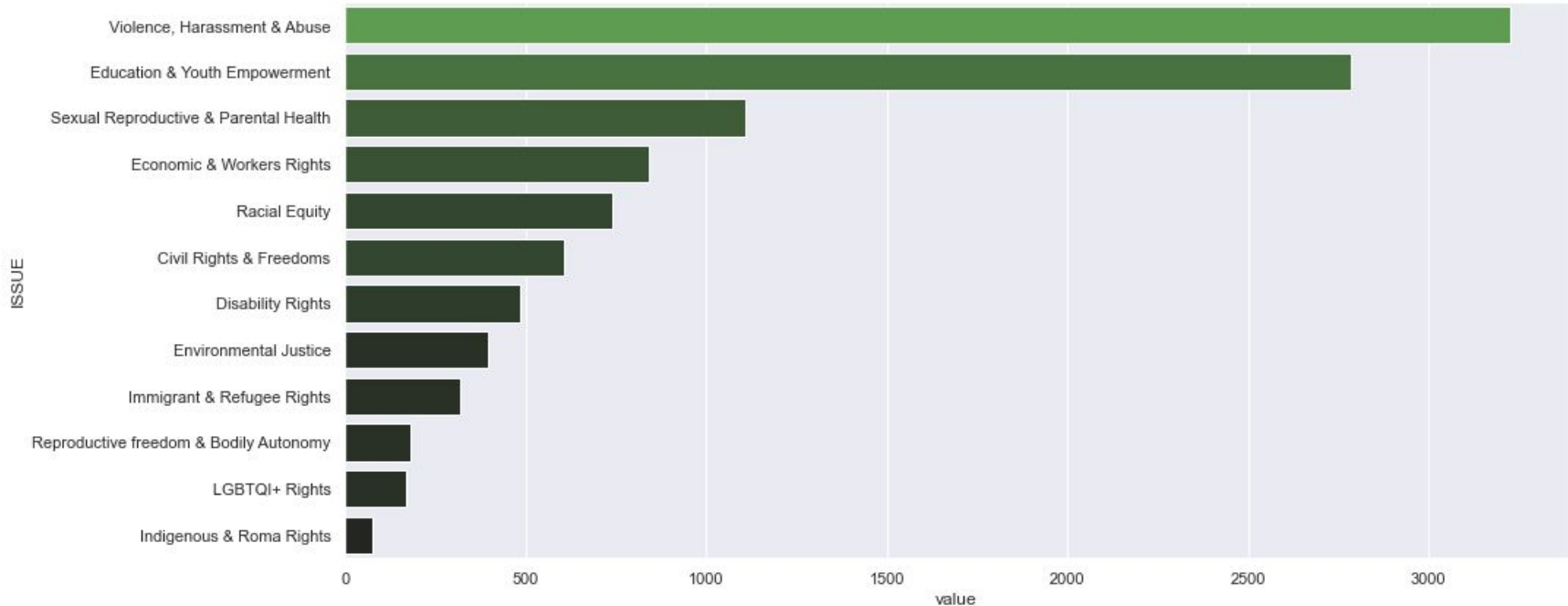
Black or of African Descent

Geographic Breakdown



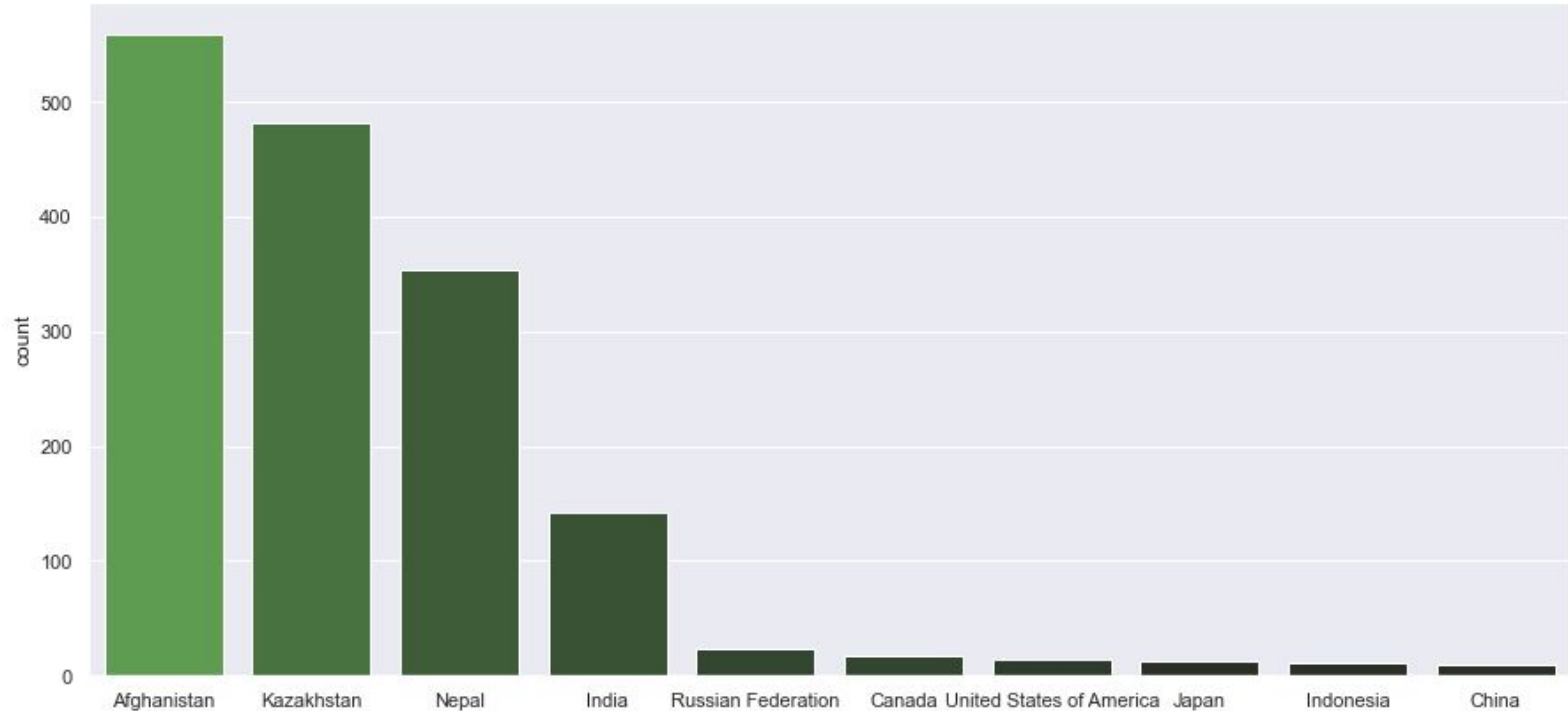
Black or of African Descent

Critical Issues



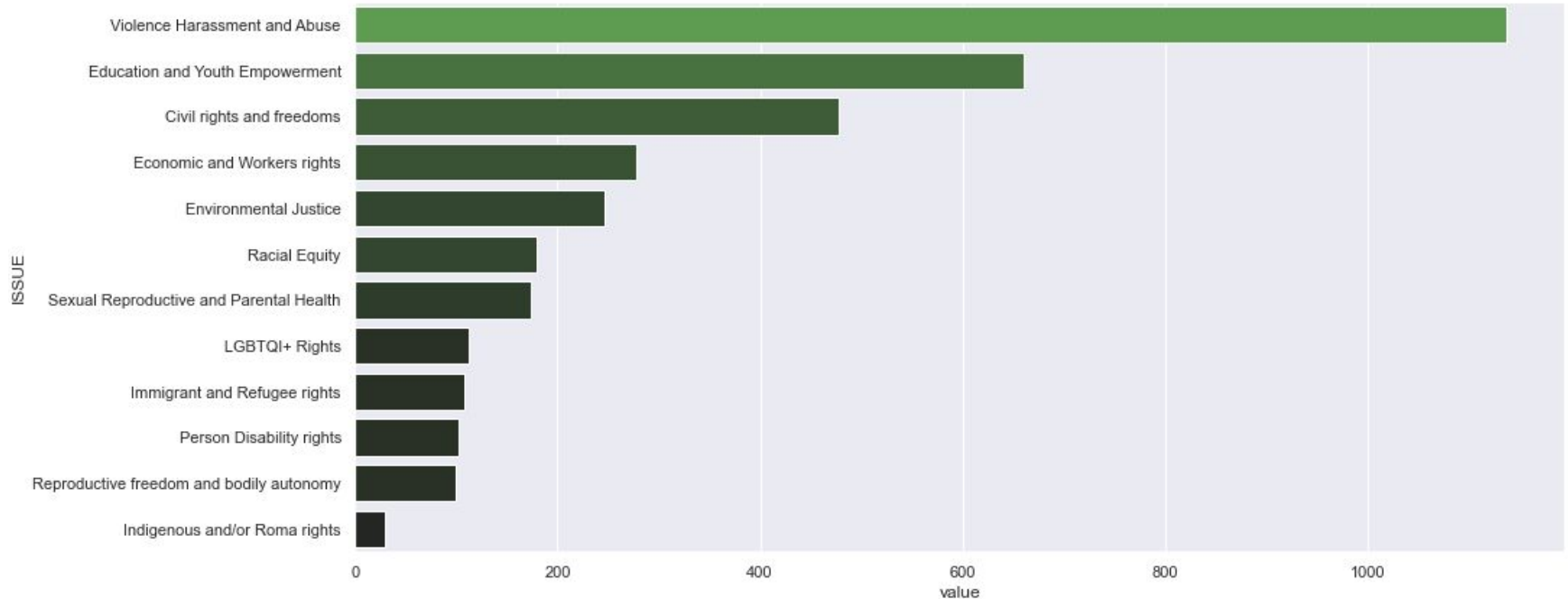
Asian

Geographic Breakdown



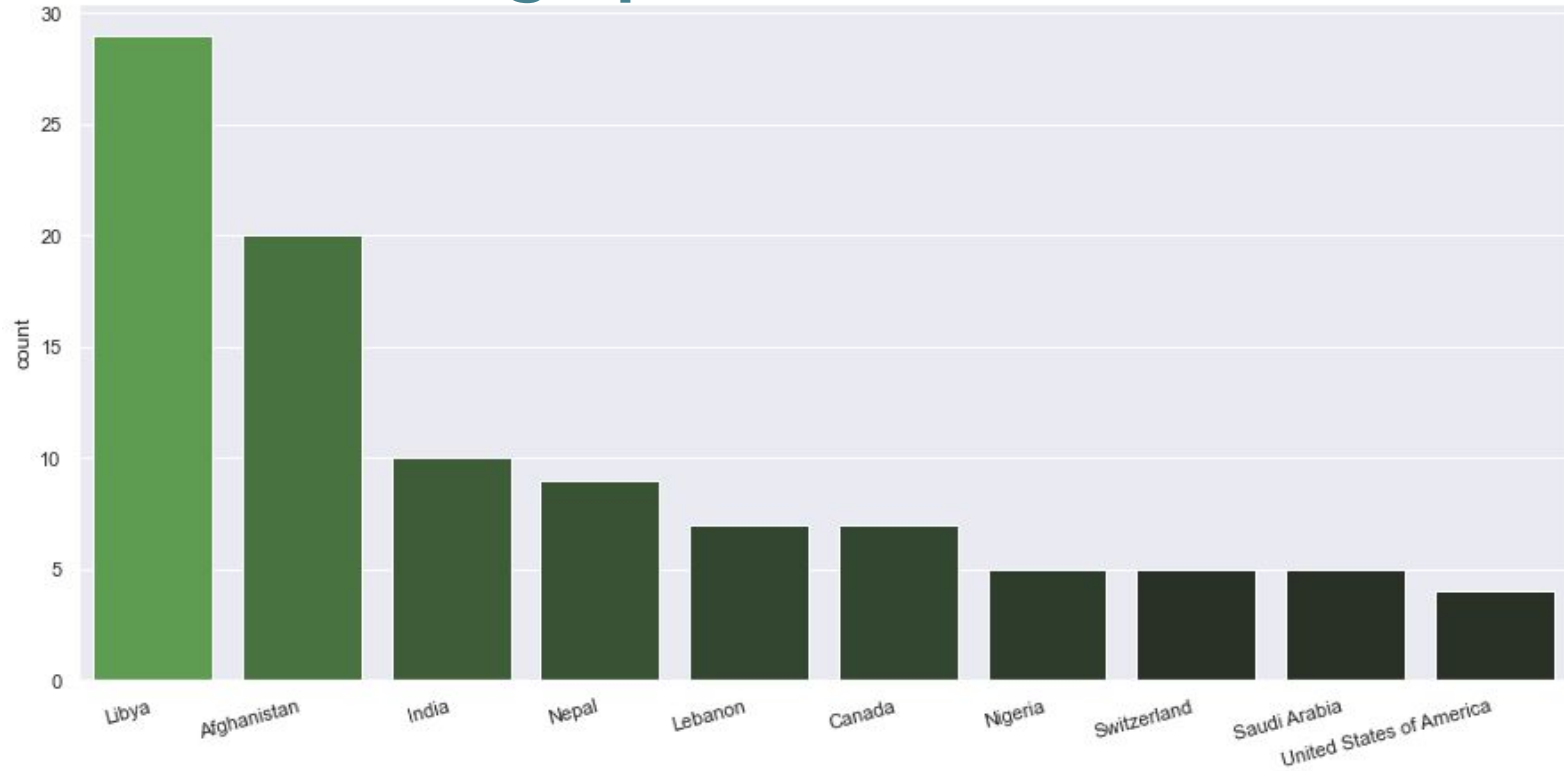
Asian

Critical Issues



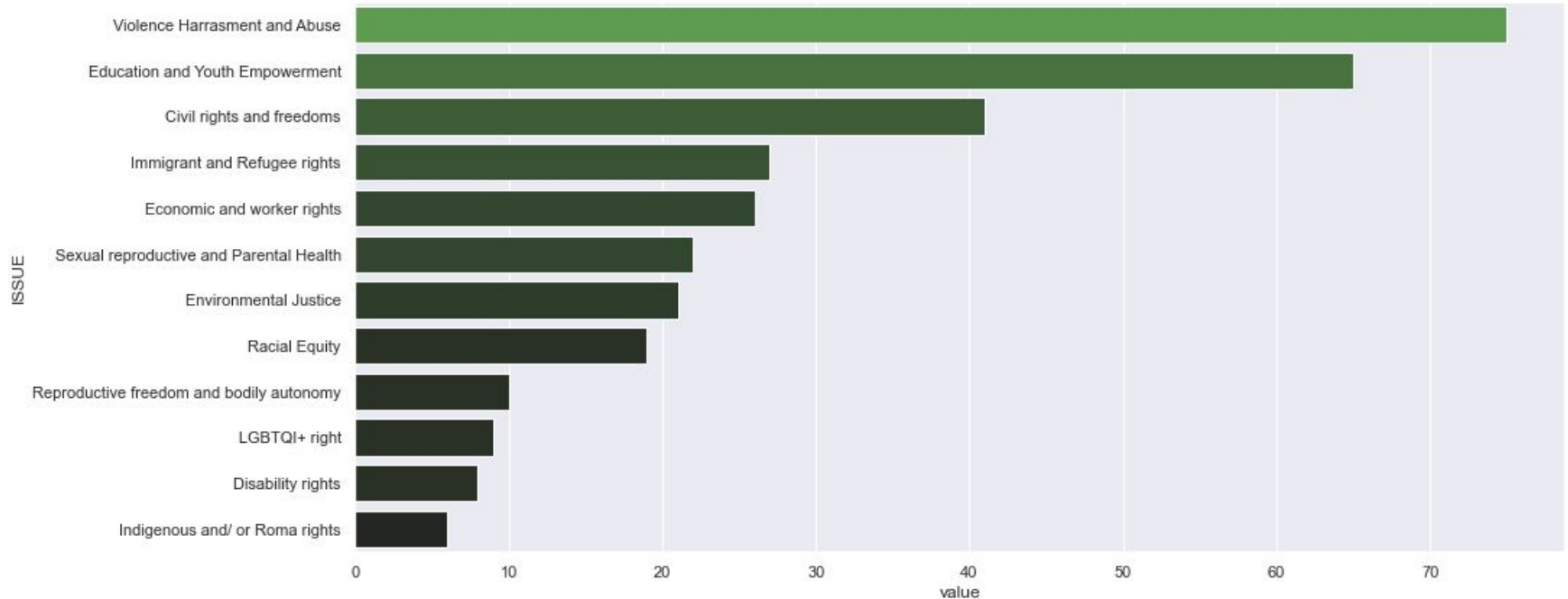
Middle Eastern

Geographic Breakdown



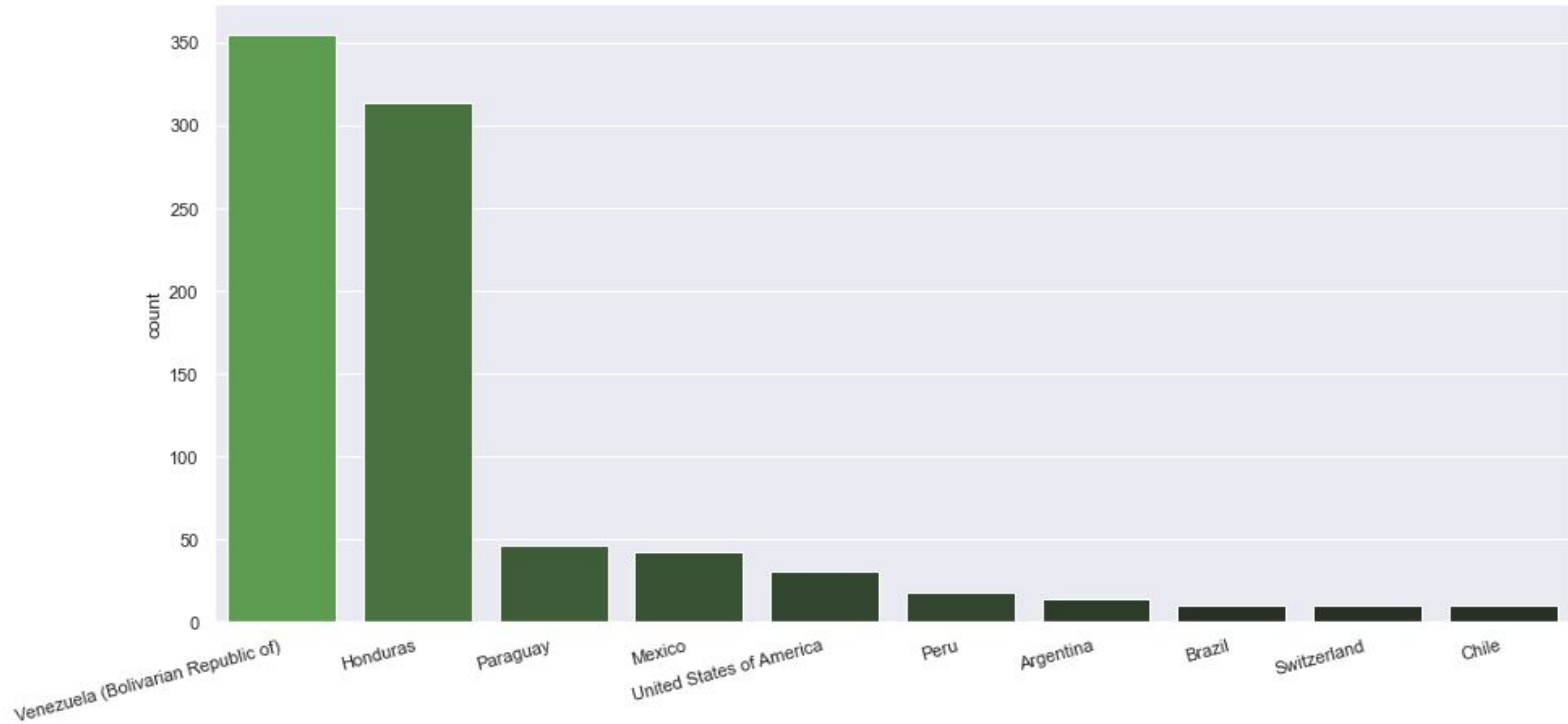
Middle Eastern

Critical Issues



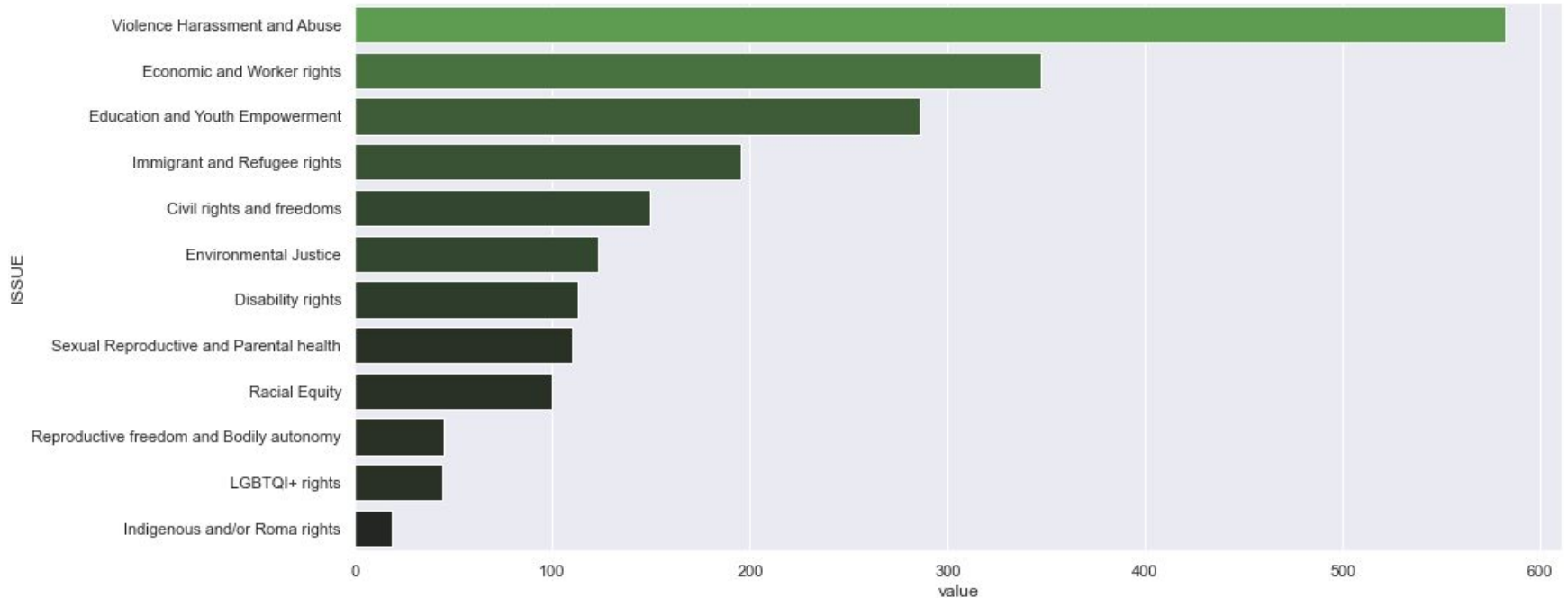
Hispanic or Latinx

Geographic Breakdown



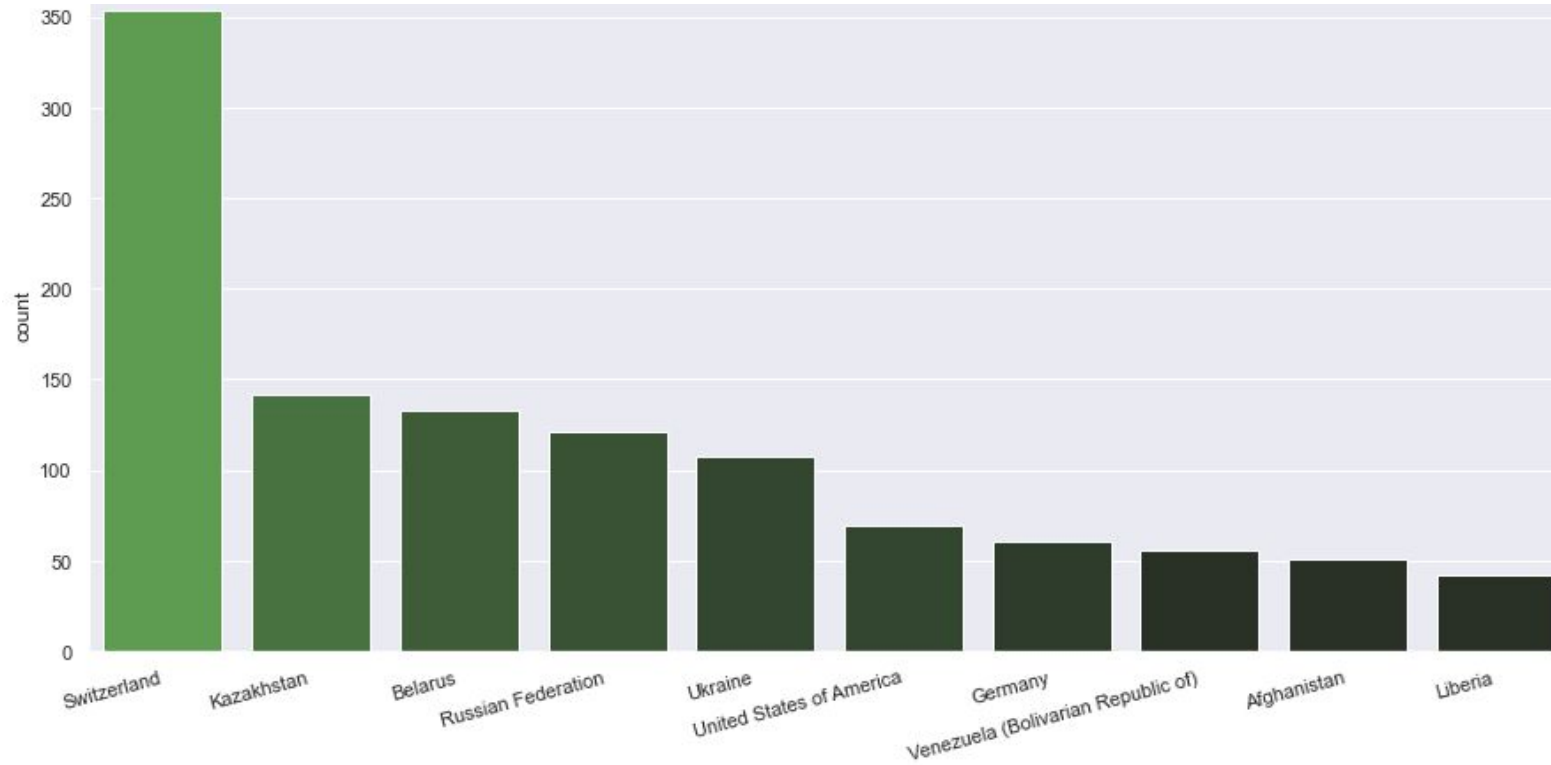
Hispanic or Latinx

Critical Issues



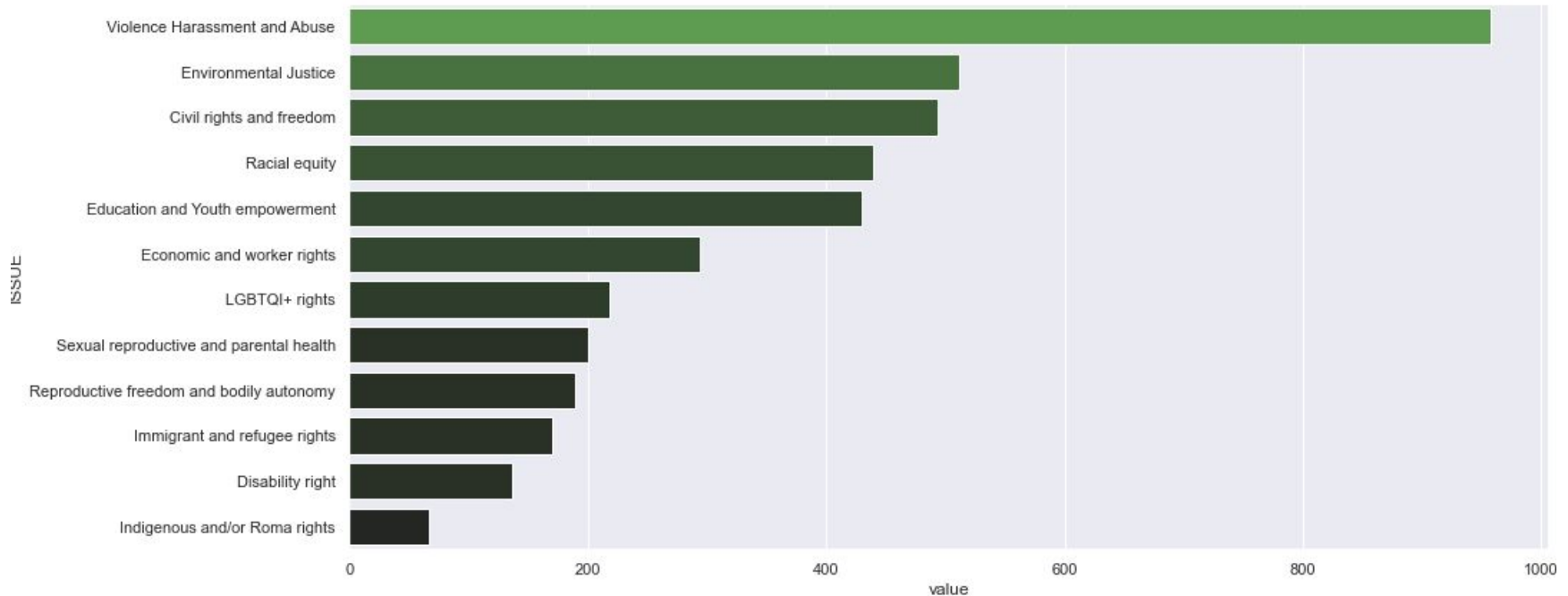
Multi-racial

Geographic Breakdown



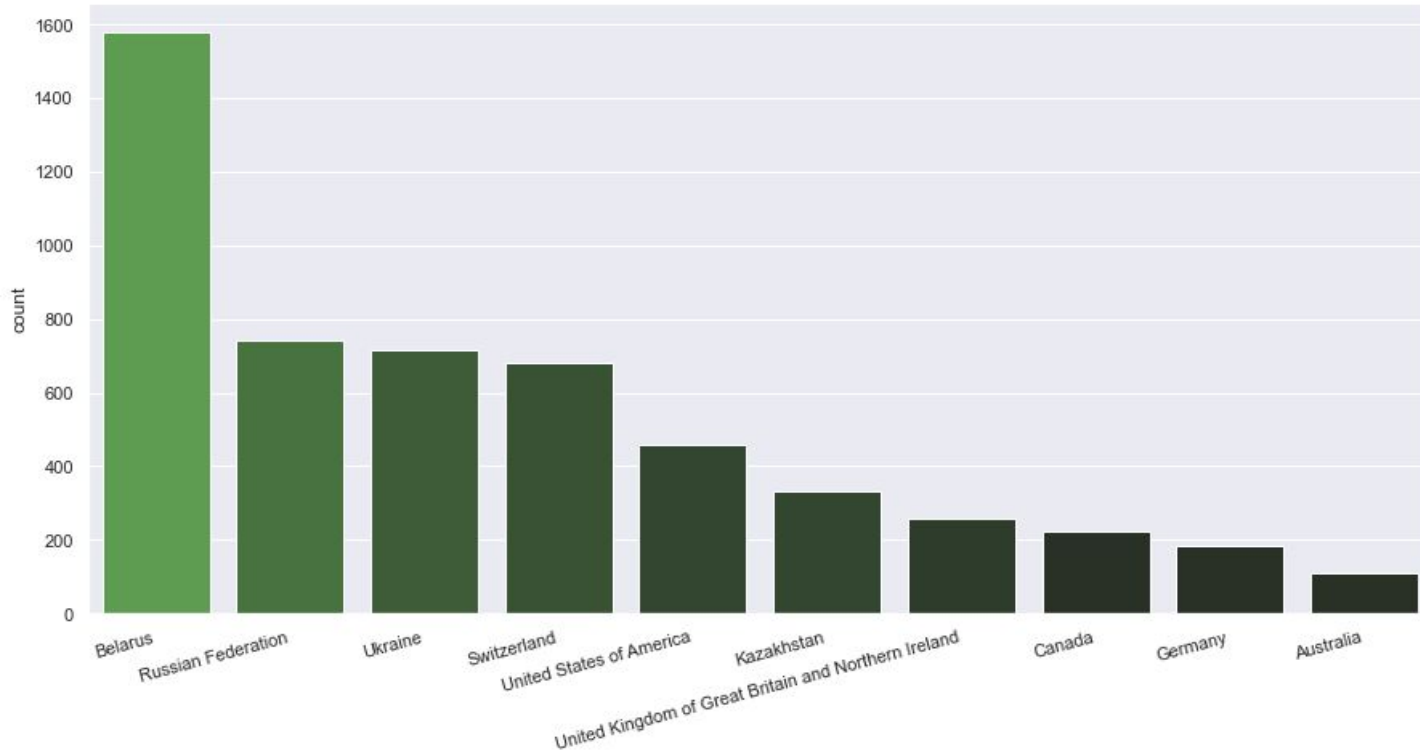
Multi-racial

Critical Issues



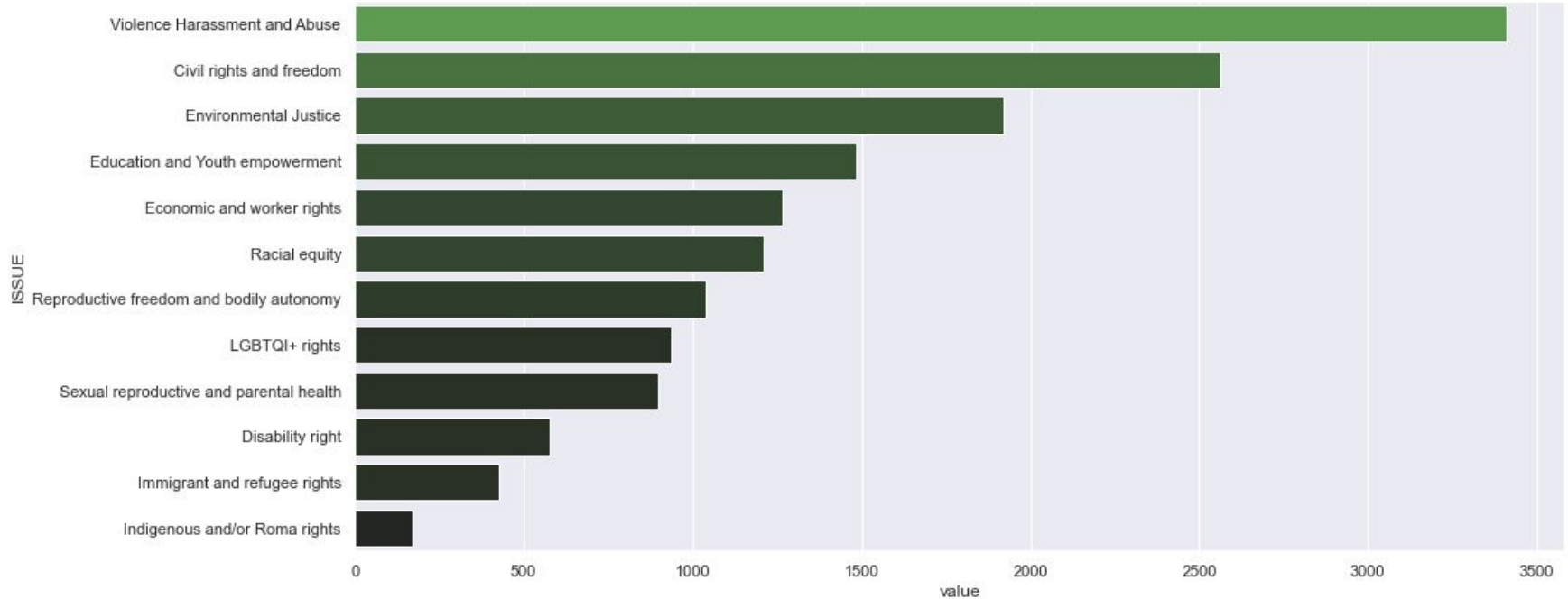
White

Geographic Breakdown



White

Critical Issues

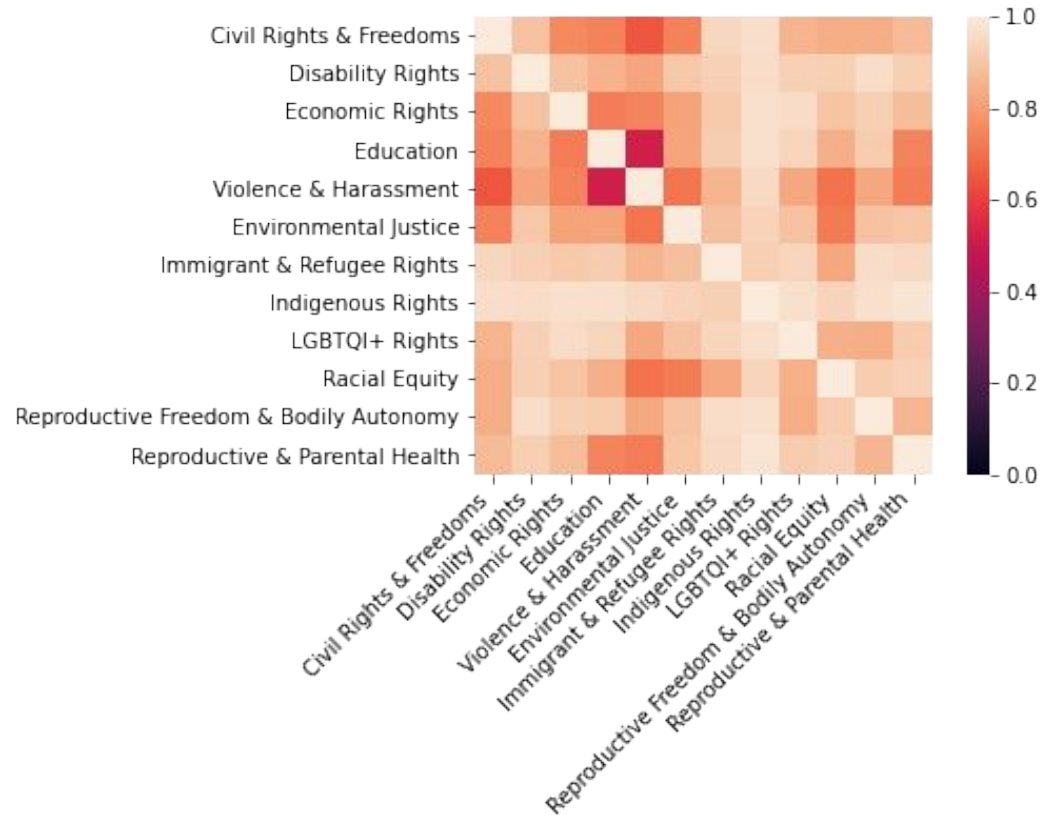


Similarity Heatmap

Relationship between issues picked in survey:
If a respondent picked Issue A, did they choose Issue B?

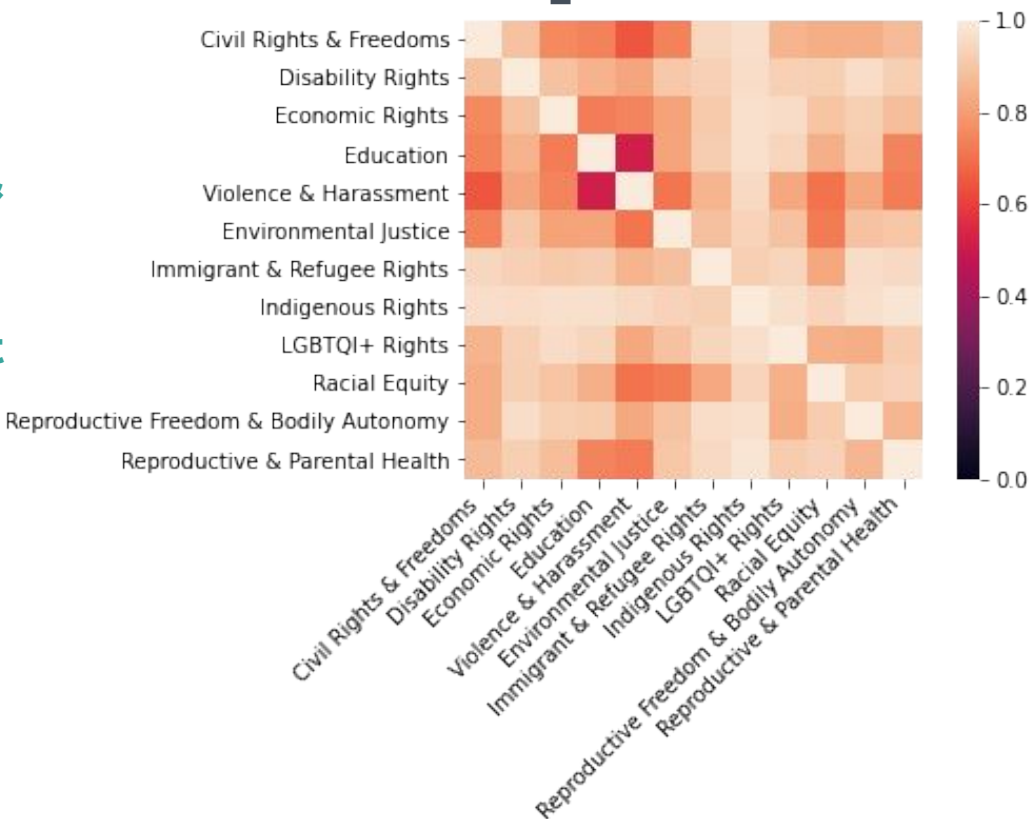
We used the Dice Similarity score between two logical vectors (Pearson's r may not produce good results with logical vectors).

Lower the score, the more dissimilar the choices.

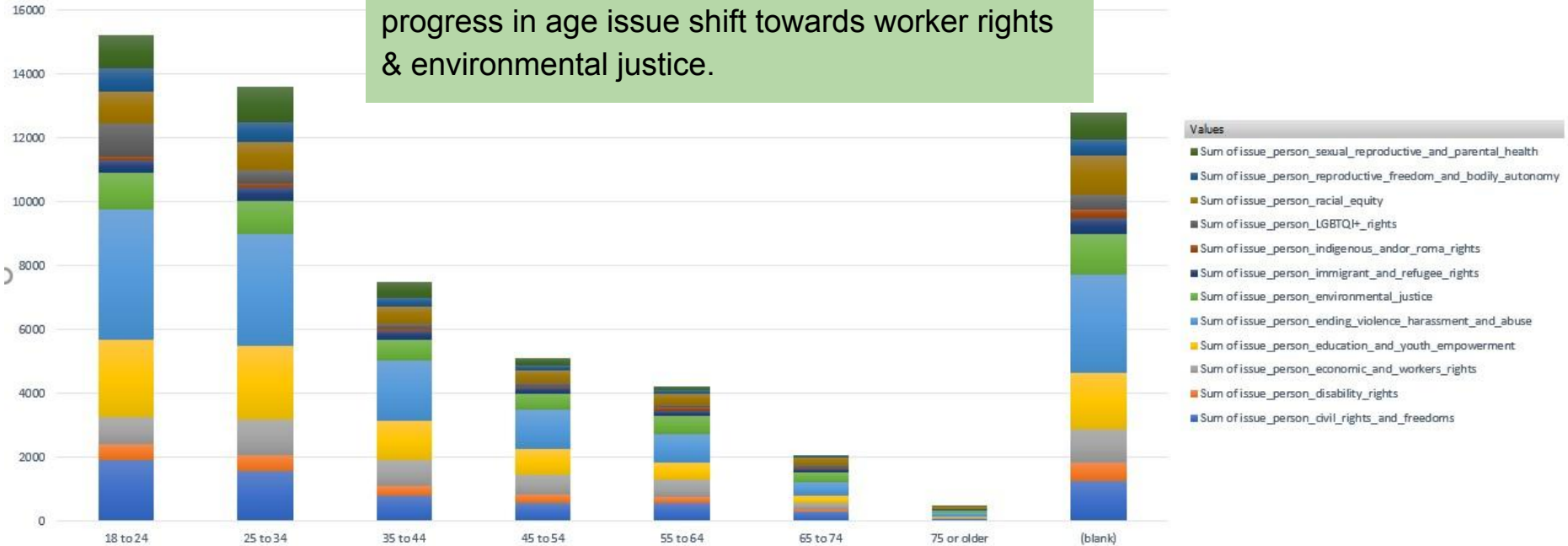


Similarity Heatmap

1. **Civil Rights & Freedom** dissimilar to **Violence & Harassment**
2. **Education** dissimilar to **Violence & Harassment**
3. **Racial Equity** dissimilar to **Violence & Harassment**
4. **Reproductive & Parental Health** dissimilar to **Environmental Justice**



Reproductive health is an important issue for people in age group 25-34 but as people progress in age issue shift towards worker rights & environmental justice.



Age wise breakdown of Issues

Takeaways

- 60% of respondents reported civil rights as a pressing issue in Russia.
- Racial inequity was one of top 3 issues in countries like Switzerland, USA and Canada.
- In addition to Violence & Education, some prominent issues per region are:
 - Africa : Reproductive & Parental Health
 - Asia & Middle east : Civil Rights & freedom
 - Hispanic : Economic & worker rights
 - Multi-racial & White : Environmental Justice and Civil rights & Freedom
 - Indigenous : environmental justice

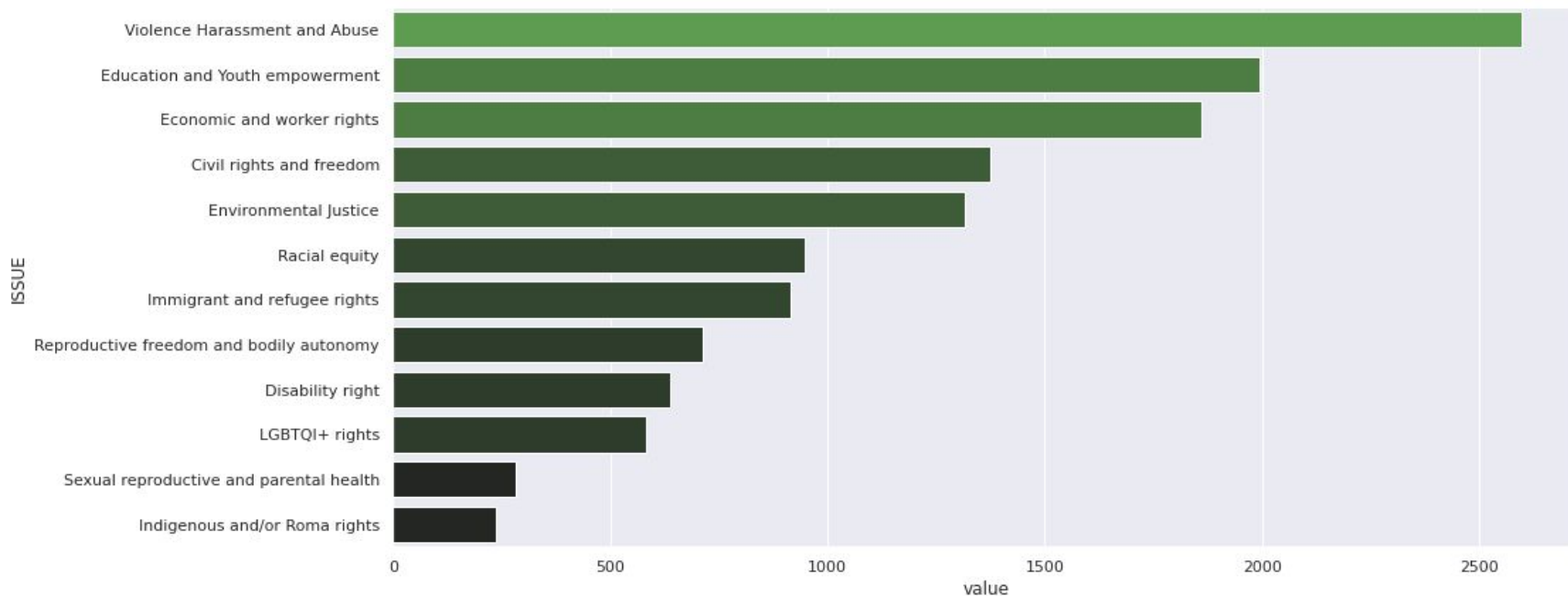
2.

Critical Issues

Gender wise
Breakdown

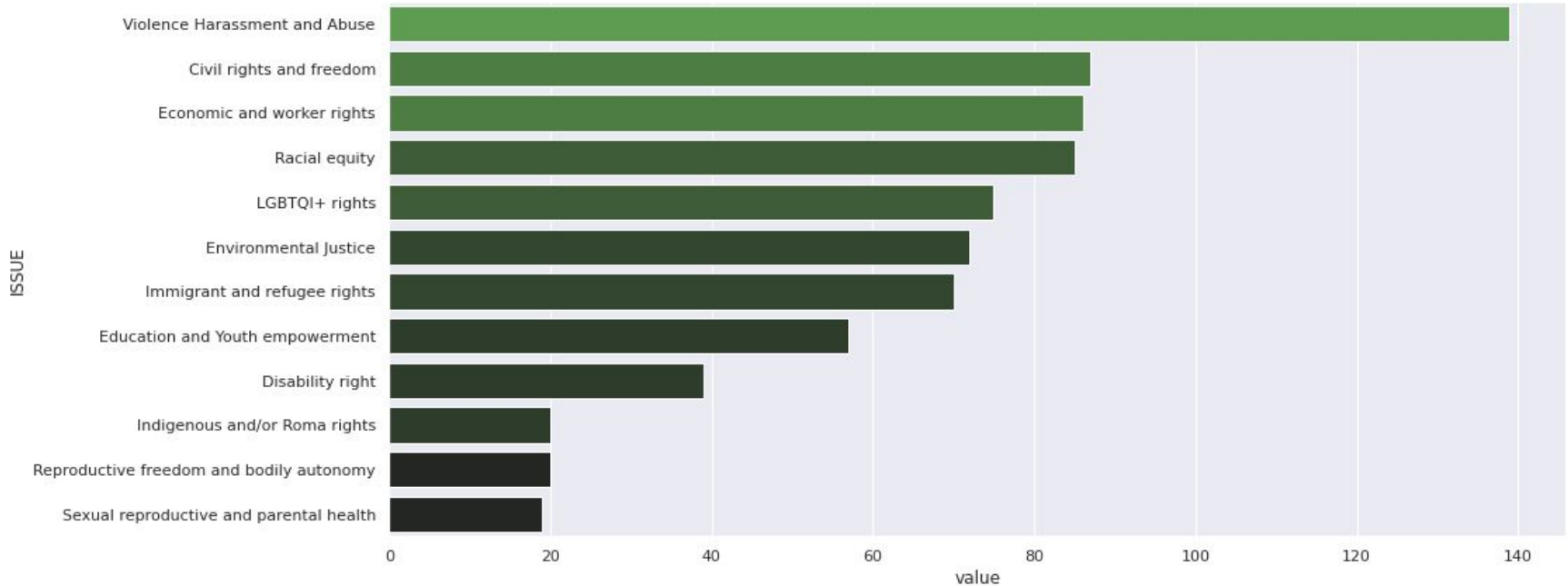
Women

Critical Issues



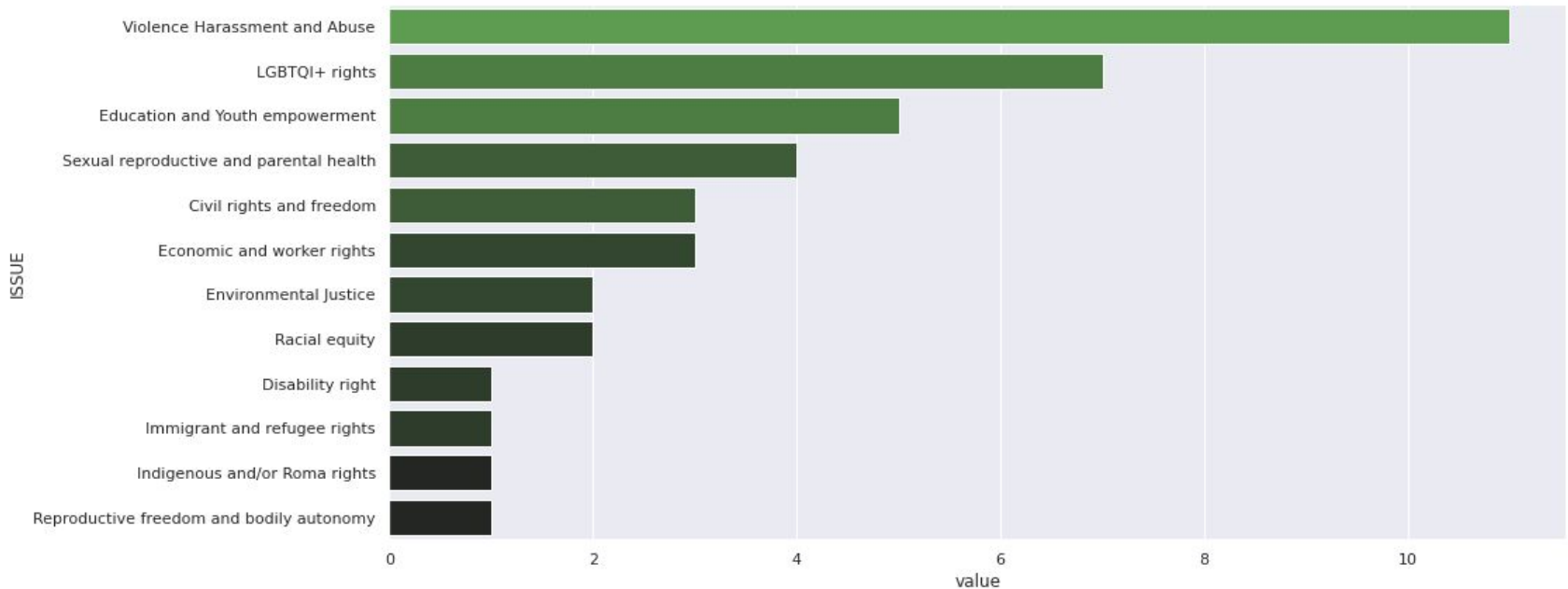
Cisgender-Women

Critical Issues



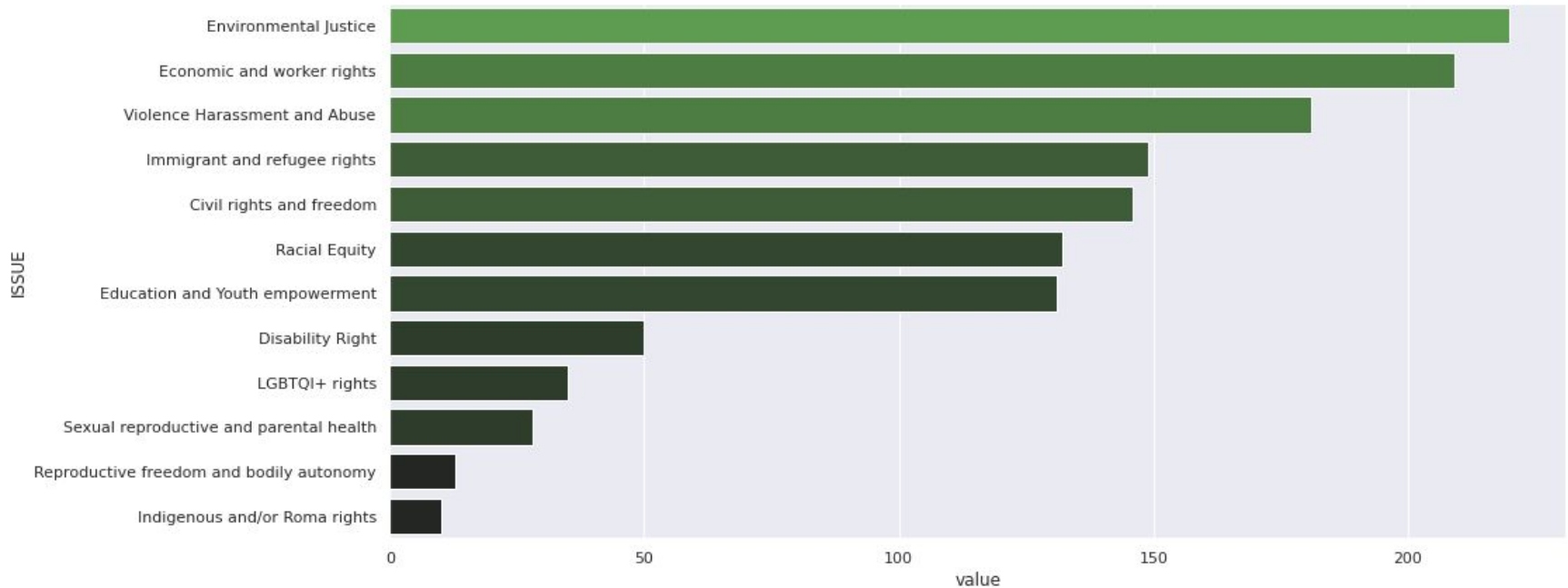
Transgender-Women

Critical Issues



Men

Critical Issues



Takeaways

- 24% of population reported Violence, education & youth empowerment as their major issue.
- 46% of non-binary people face LGBTQI+ rights issue.
- LGBTQI+ rights and Sexual reproductive and parental health were major issue for Transgender-women.
- For men, major issues are environmental justice & worker rights

3.

Feature Importance

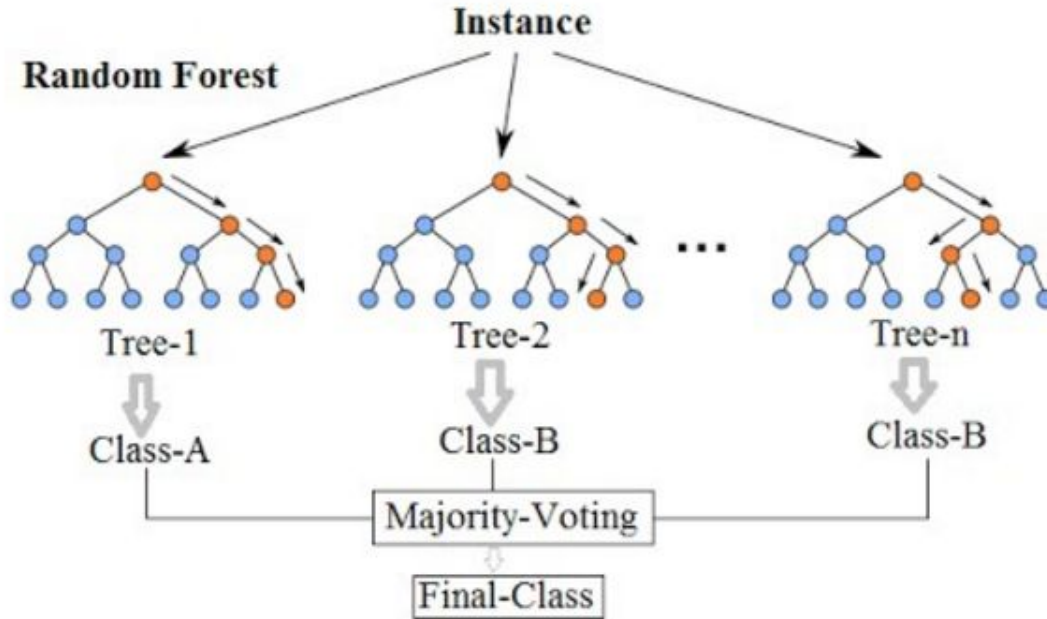
Using Random
Forests

Random Forests Classification

Random Forests is a supervised Machine Learning Algorithm whereby many decision trees are constructed, such that their outputs are combined in making a decision.

When used for classification, the class which is most commonly predicted by the trees is the output.

Random Forest Simplified

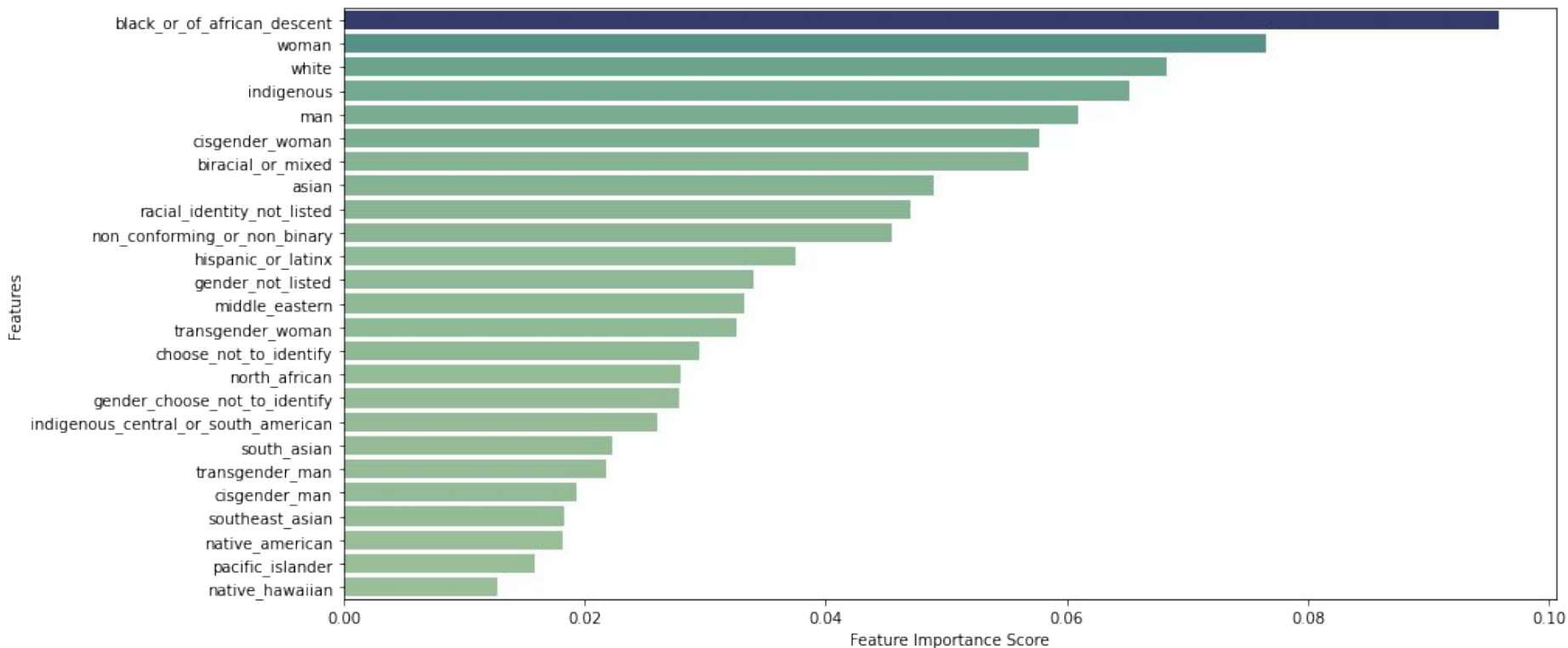


Source: Venkata Jagannath, Creative Commons (CC BY-SA 4.0)

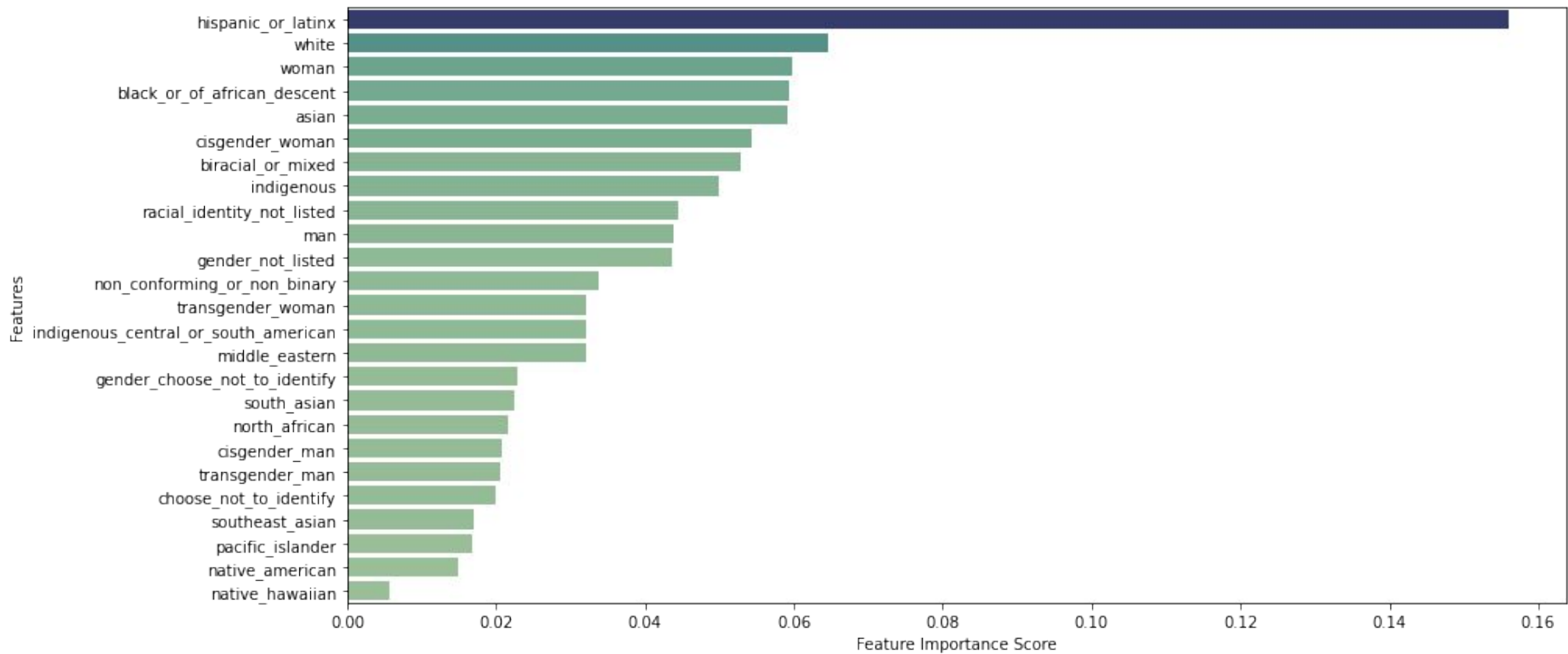
<https://community.tibco.com/wiki/random-forest-template-tibco-spotfirer-wiki-page>

Feature Importance

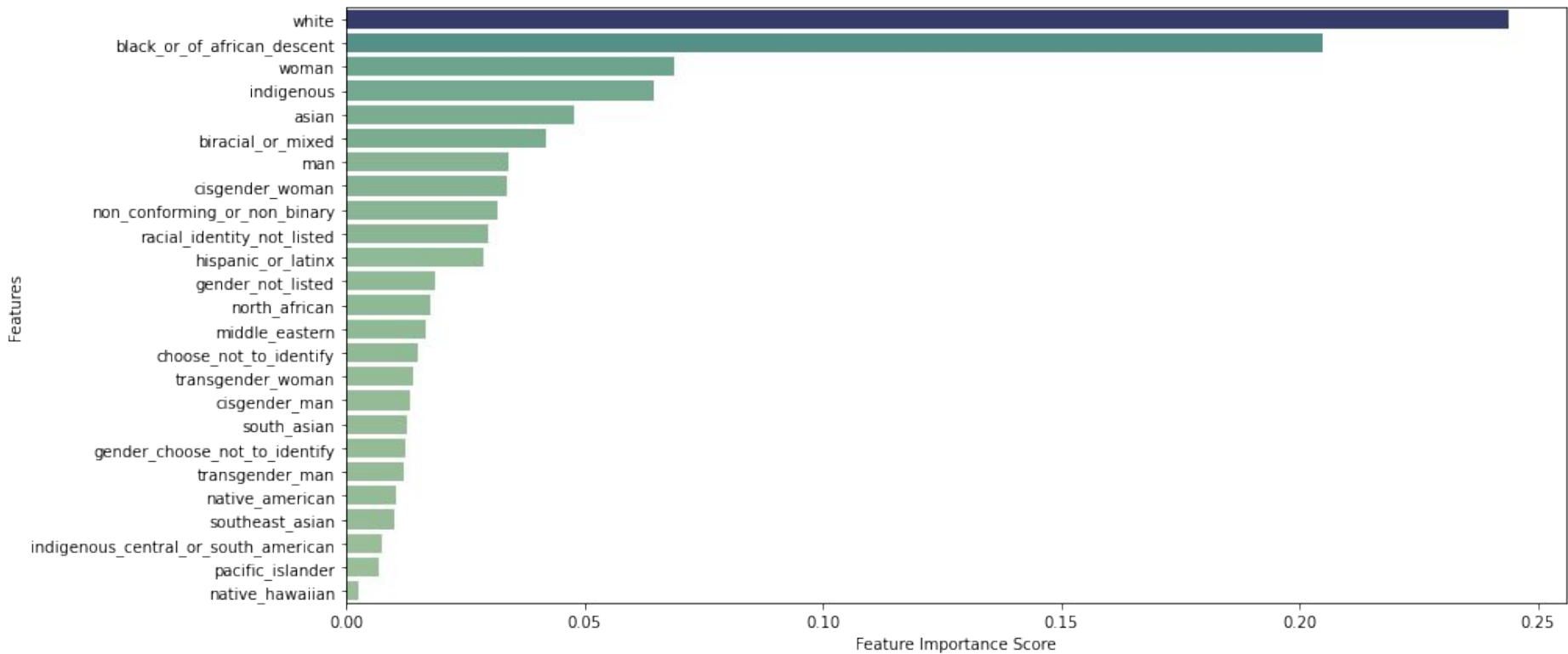
- **Violence, Harassment and Abuse** - Black or African descent
- **Economic & Worker Rights and Immigration** - Hispanic and Latinx
- **Environmental Justice** - White
- **Education & Youth Empowerment** - Black or of African descent
- **Reproductive Freedom and Bodily autonomy** - Cisgender Women



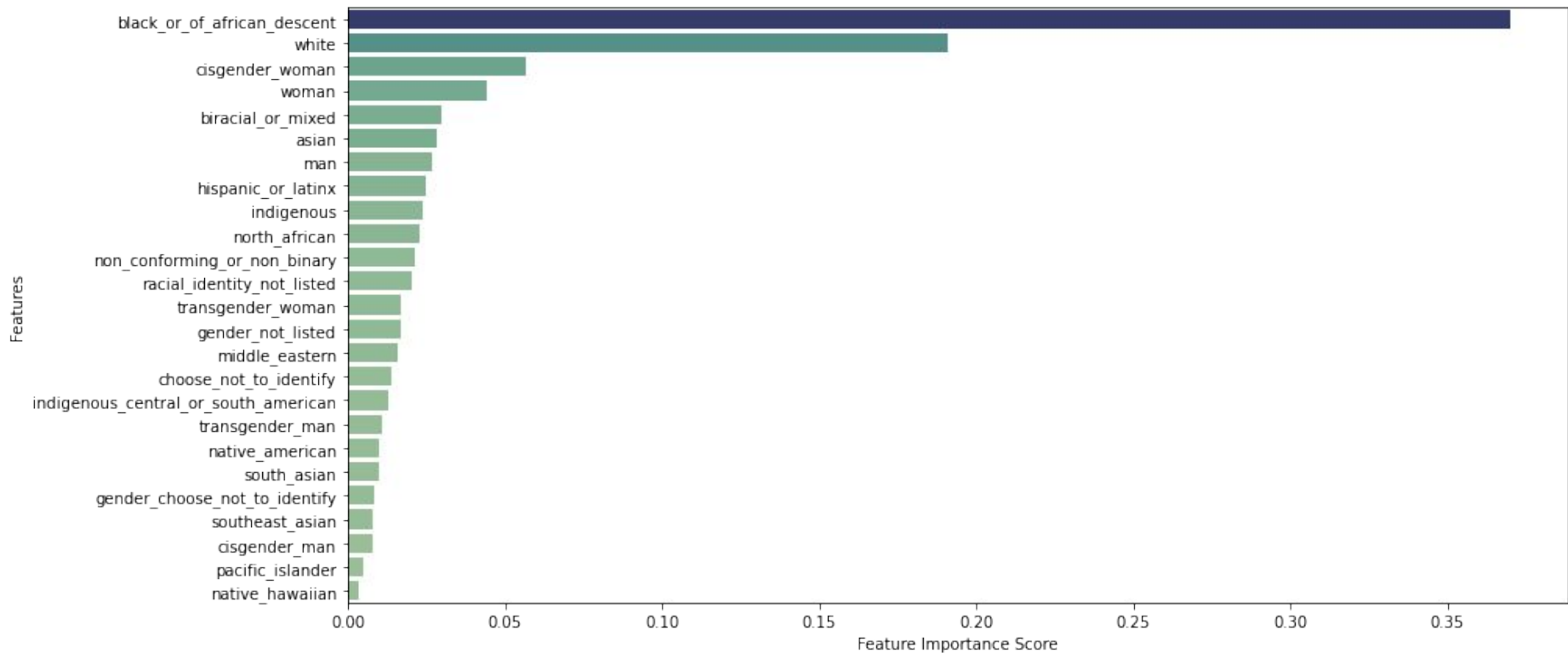
Feature Importance for Violence, Harassment & Abuse



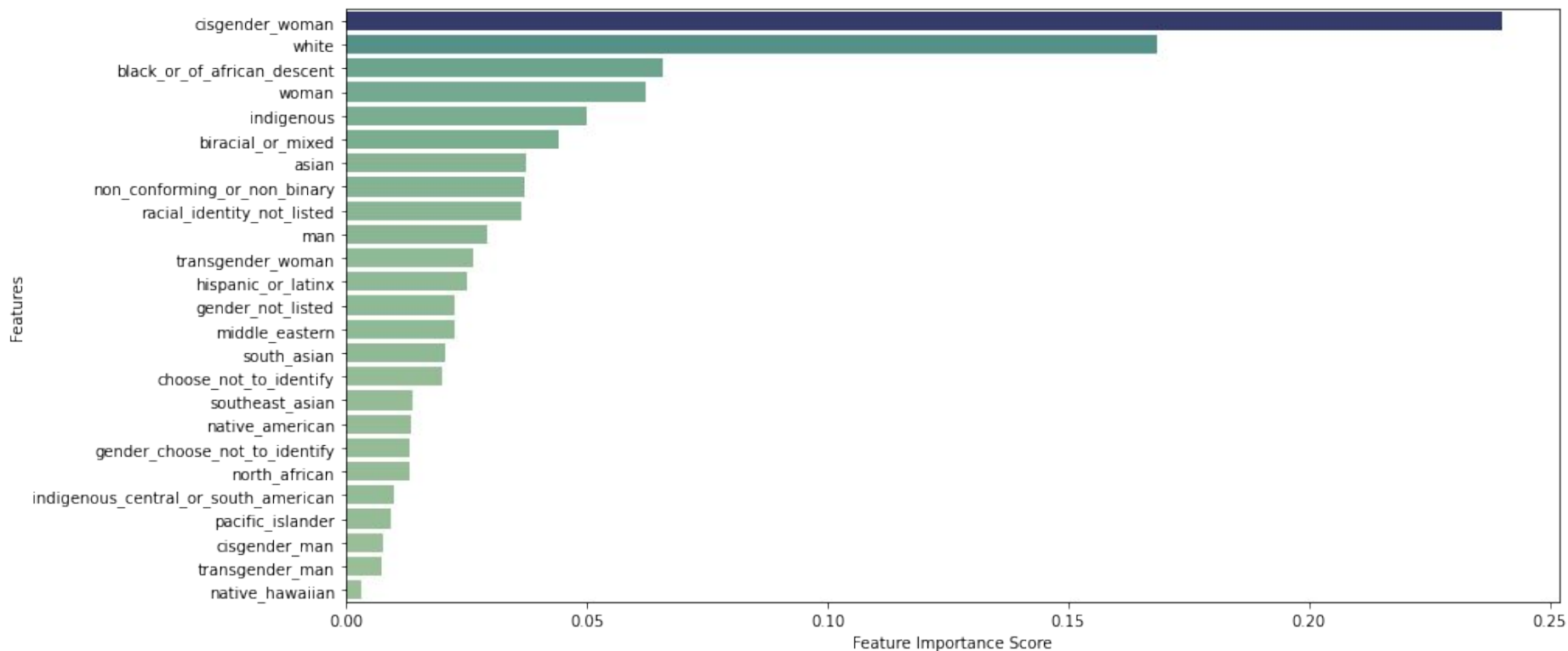
Feature Importance for Economic & Worker Rights



Feature Importance for Environmental Justice



Feature Importance for Education and Youth Empowerment



Feature Importance for Reproductive Health

4.

Common Barriers across Issues

What prevents
progress?

Political and Cultural/Social were the most commonly chosen barrier across all Issues.

Politics was a major barrier for issues like economic & worker rights, environmental justice, immigrant & refugee rights and indigenous rights.

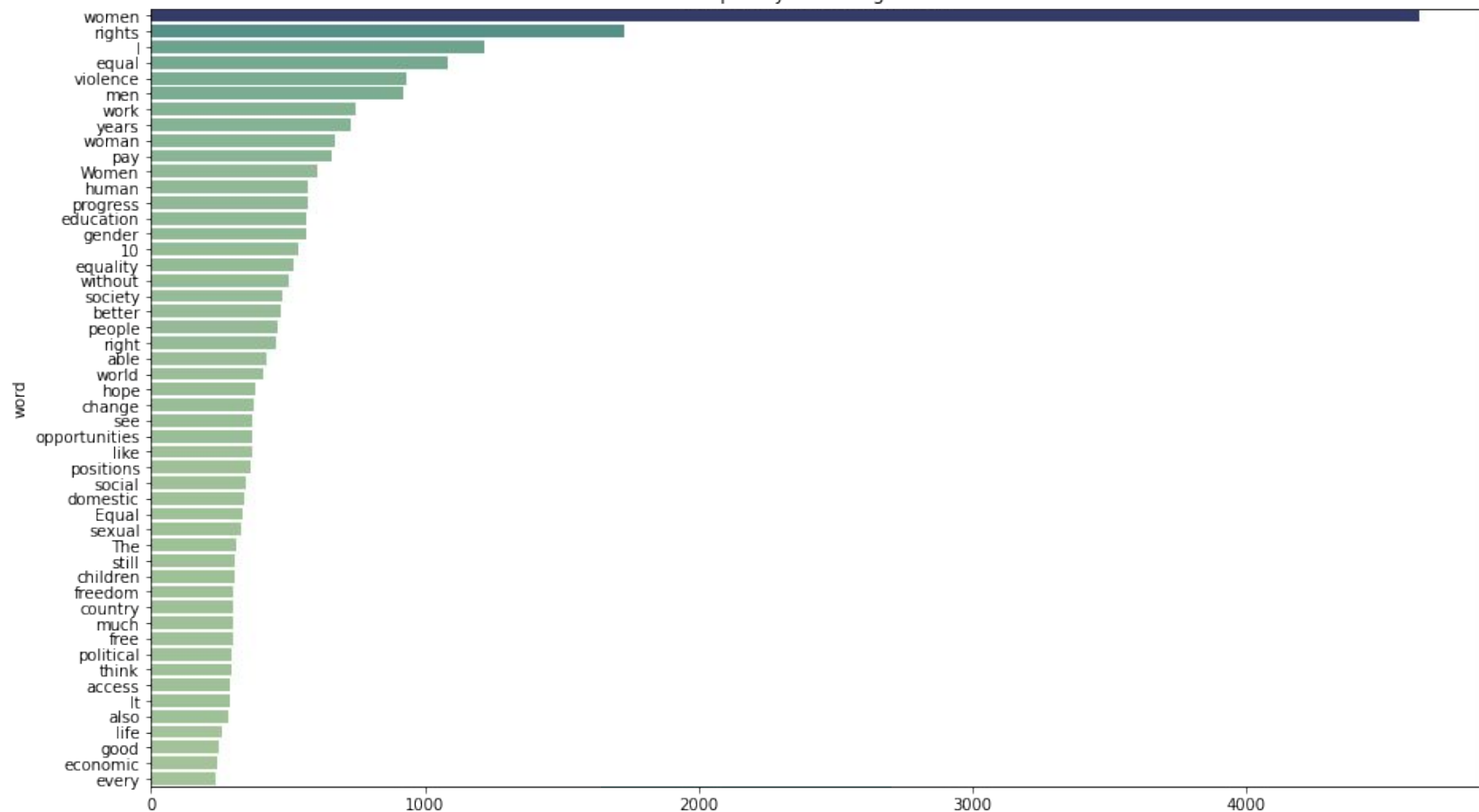
While most genders saw culture as barrier for violence, harassment and abuse, transgender-people experienced politics being the major barrier for progress.

5.

Comments about progress in 10 years

What did
people write?

Frequently Occurring Words



50 highest occurring words

Word Clouds

A word cloud (also known as a tag cloud) is a technique of summarizing vast amounts of text. The words featured in the Wordcloud are chosen based on frequency and relevance.

[illegible]

Recommendation for next steps



- Stricter rules and policies against violence, abuse & harassment.
- Data should be more inclusive of countries & genders for e.g. there were almost no entries from countries like China.
- If we want to extract order of importance of issues, then that information needs to be incorporated in the survey (through rank).

Thank you!

Team members:

Vincent Kwan

Tu Nguyen

Vivian Ogwuihe

Alka Bhambhu

Chaw Nandar