GROUP PROJECT PRESENTATION

Survey Research Methodology II

Irene García-Espantaleón, Candela Gómez,
Bradley McKenzie, Diego Paroli

uc3m | Universidad **Carlos III** de Madrid

Contents

Country level data

- GDP per Capita - World Bank
- LGBT+ Rights Index - Our World in Data
- Gender Inequality Index - UN Development Programme (UNDP)
    - Chosen over the GDI for its better fit for our analysis.
- Democracy Index - The Economist

Data cleaning

Process:

1. Remove unrelated sections of the questionnaire (ex. trade, energy policy...)
2. When more than 1 version of the same variable was available (ex. age, political ideology): choose the best one using visualizations vs target variable (still 200+)
3. Creating new measure by aggregating questions together to reduce dimensionality without losing too much information

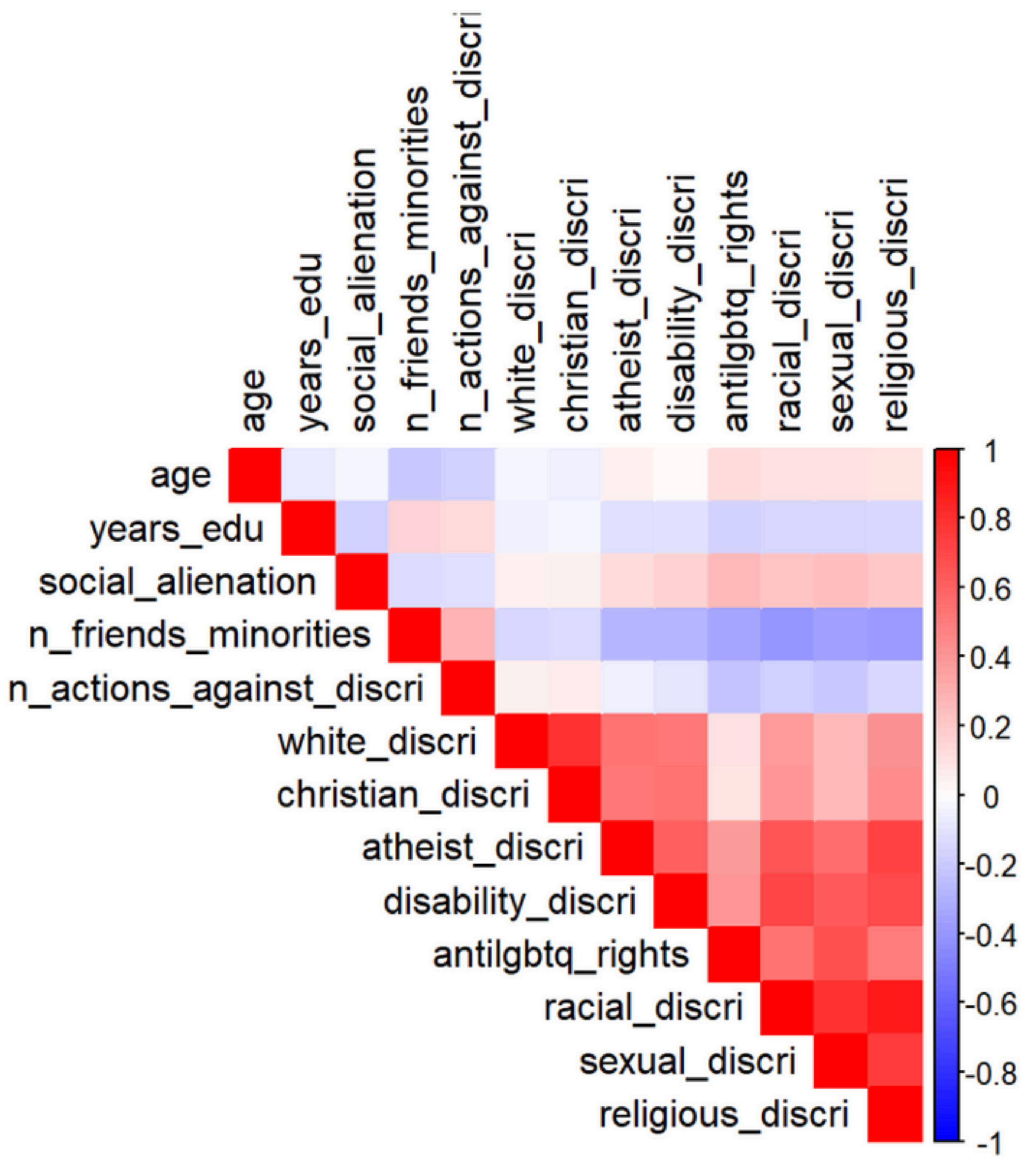
```
```{r}
data <- data |>
  mutate(across(starts_with("qc15"), ~ if_else(.x == 5, NA, .x))) |>
  mutate(antilgbtq_rights = round(rowMeans(cbind(qc15_1, qc15_2, qc15_3), na.rm = TRUE), 2)) |>
  select(-starts_with("qc15"))
# Scale of 1 to 4, 1 = supportive, 4 = homophobic
```
```

Variable creation for attitudinal questions

We created general variables to group attitudes across questions. For example:

- Discrimination scores for religion/minority groups
  - Combined between q12 and q13 response
- Anti LGBTQ+ discrimination score: qc15_1, qc15_2, qc15_3
- Social alienation: d72_1, d72_2

Later, we aggregate further due to very high collinearity to reduce dimensionality.



Missing data

- In general, refusals and DK's were recoded to NA.
- Identified all factor variables -> convert values to factor label

```
# Converting them to factors and assign them their labels automatically
data <- data |>
  mutate(across(all_of(factor_variables), labelled::to_factor))
```

- Then each "DK" code was consistent (vs 7,97,99)

- Our exception:
  - Spontaneous refusal to question on whether you had transgender friends.

- Imputation through MICE.

Literature Review and EDA

Individual Factors:

- Gender: women tend to be more favorable
- Age: Literature suggests less support from older individuals; EDA shows a nuanced picture with age affecting response rates.
- Religion: religious fundamentalism proved to be a significant factor; non-believers most accepting (Kanamori & Xu, 2020).
- Political Ideology: Left-wing more inclined to endorse transgender rights.

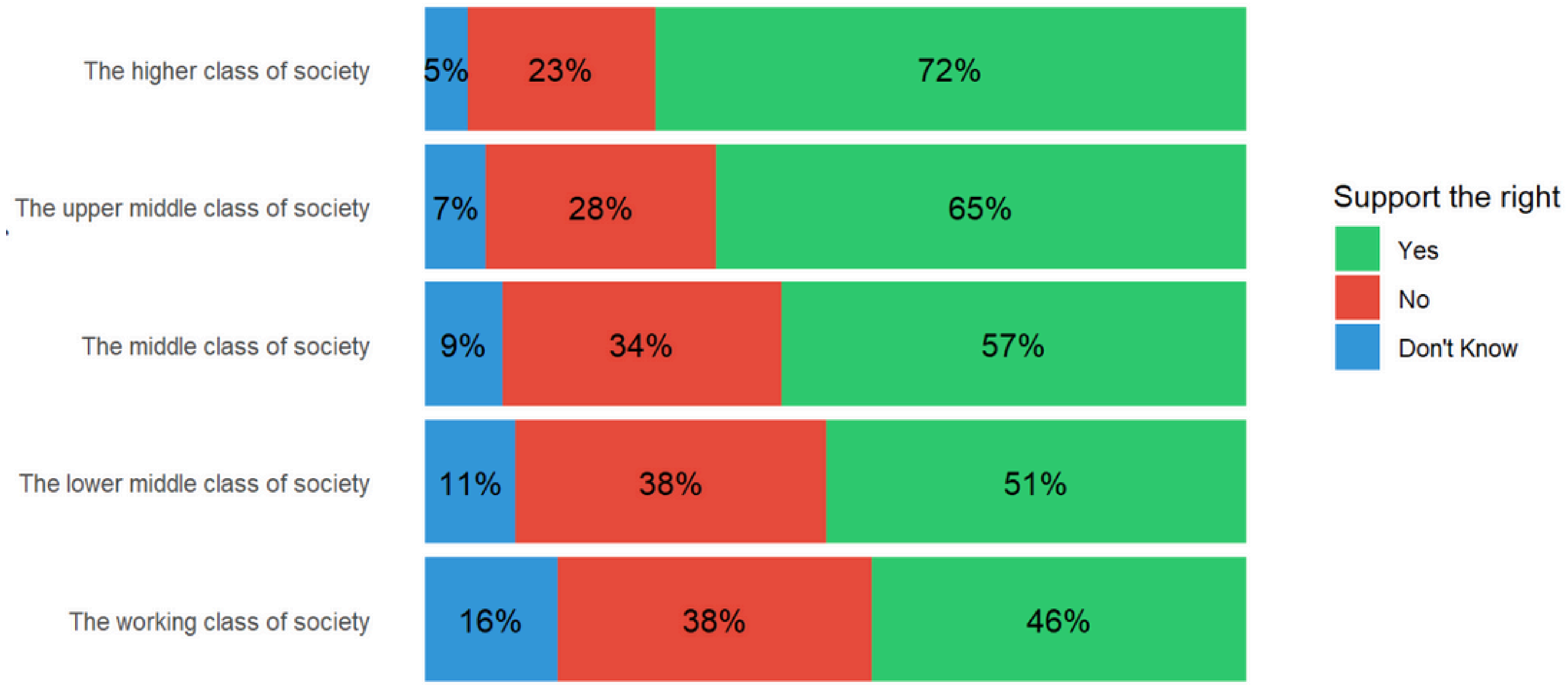
Social factors:

- Contact with transgender individuals strongly influences support (Aguirre-Sánchez-Beato, 2020).
- Education level shows ambiguous impact.
- Perceived social class has a positive correlation with support, something not so prominent in previous studies.

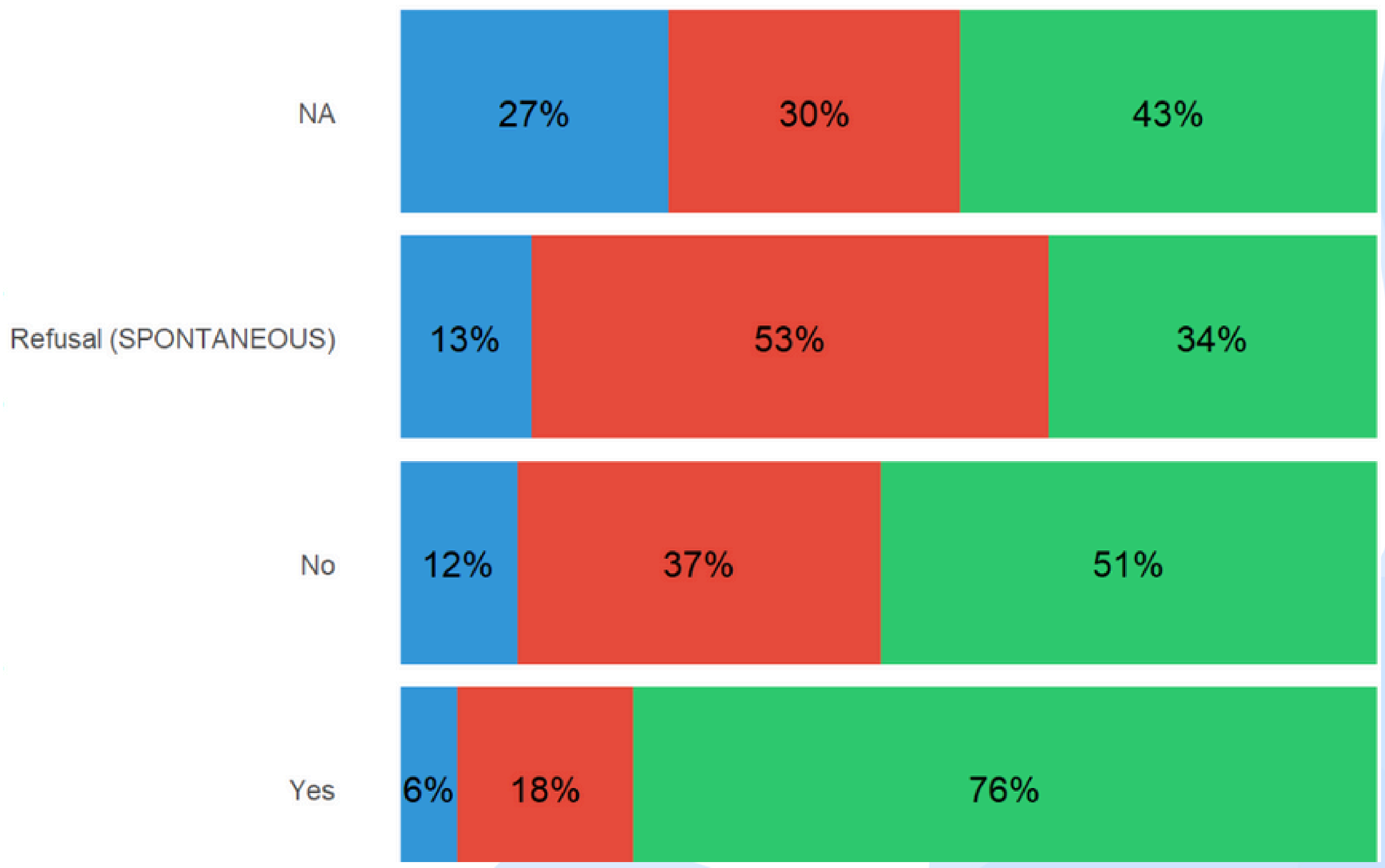
Literature Review and EDA

Correlations with target - many interesting relations

Perceived social class



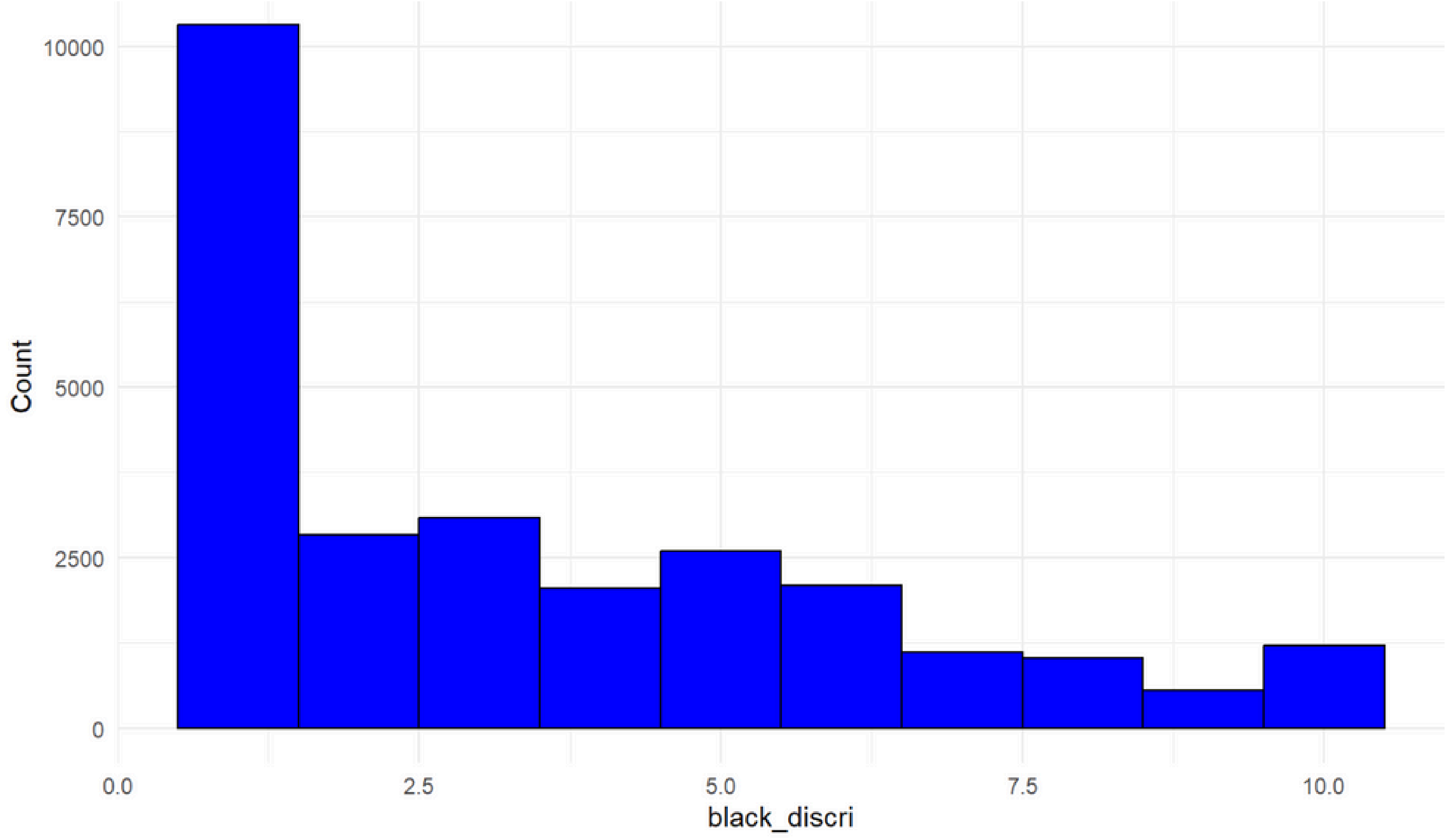
Friends with any transgender person?



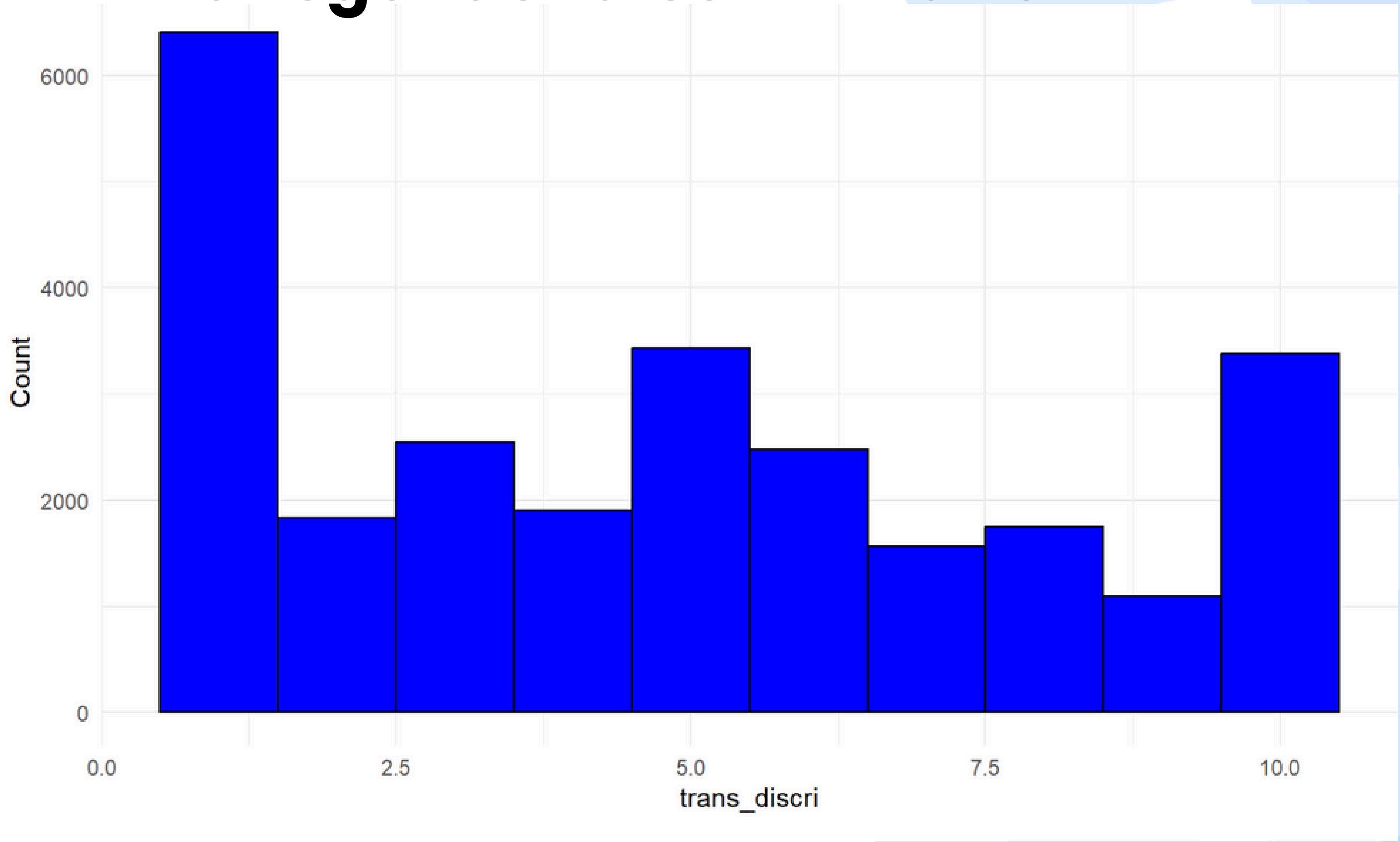
Literature Review and EDA

Distribution of discrimination variables - LGBTI+, transgender and intersex slightly higher in general than race and religion.

Black discrimination



Transgender discrimination

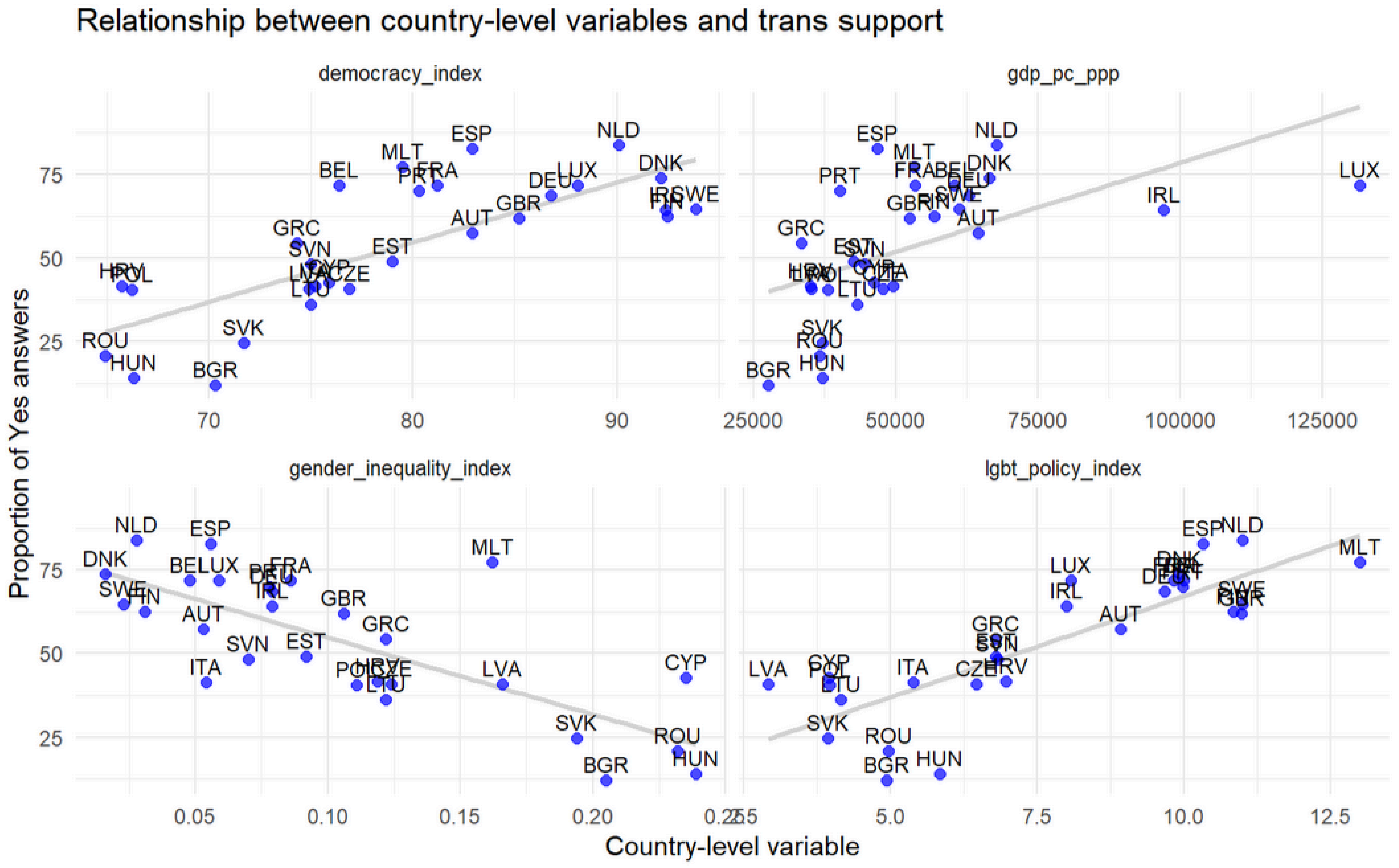


Note: these do use our recoded discrimination by minority variable.

Literature Review and EDA

Country-Level Factors

- Democracy & LGBT Policy: Positive relationship with support for trans rights.
- Gender Inequality Index: Negative relationship—more gender equality in a country leads to higher support.
- GDP per Capita: Non-linear relationship with diminishing returns beyond a certain threshold.



Survey Paradata

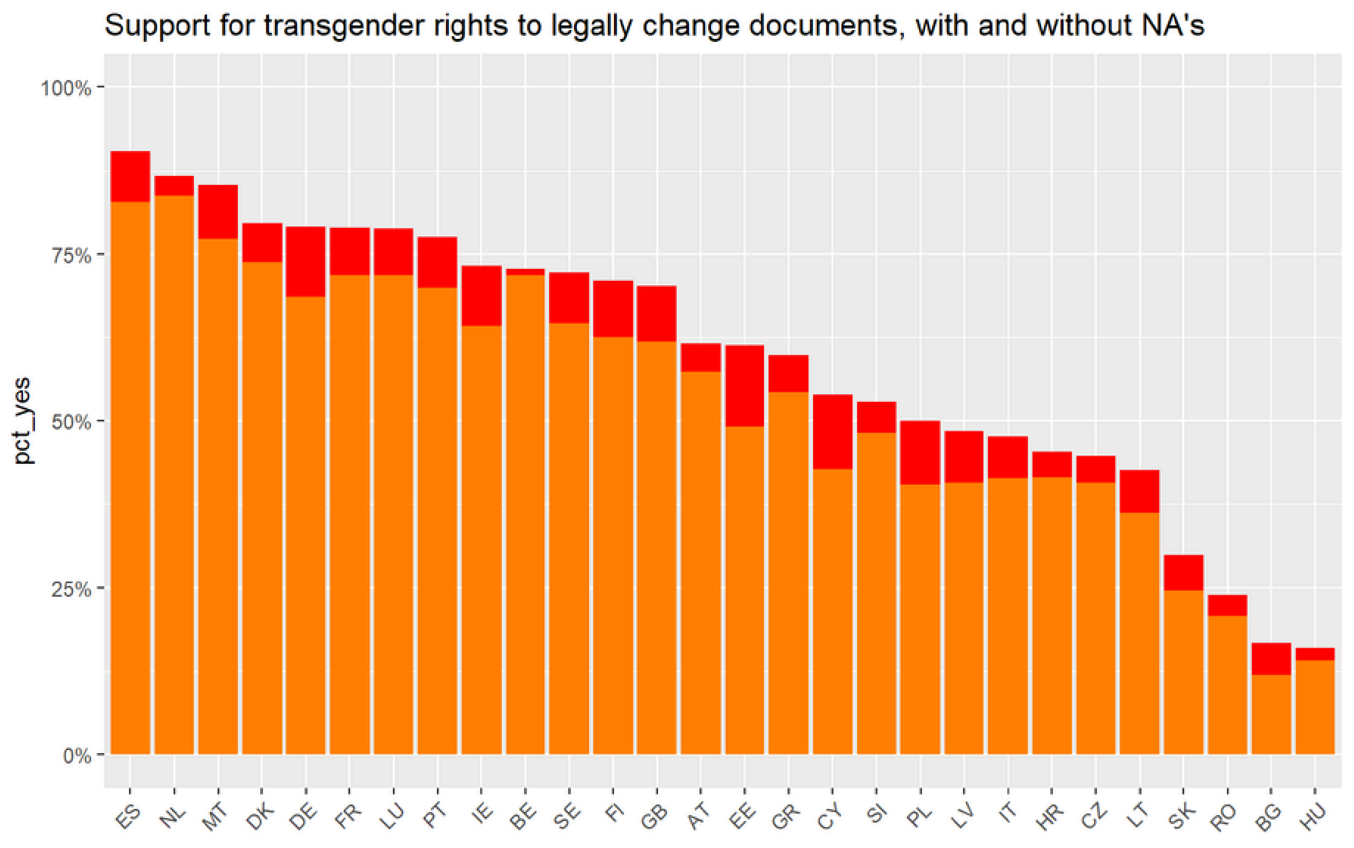
Longer interview duration and better survey engagement correlated with higher support for transgender right

Robustness checks - modelling NA predictability

Signficant variables

Treated NAs as binary outcome. We undersample and apply this logistic model.

- Women (+20%)
- Working class
- Anti-LGBTQ+ rights (+13%)
- Less than daily internet use



Paradata

Interview cooperation score (p5) significant.

Compared to the “excellent” cooperation rating:

- Fair -> 42% more likely to NA
- Average -> 74% more likely to NA
- Bad -> 349% more likely to NA resp

Explanatory models: our approach

Logistic regression with all the variables

Base individual data model. Higher support includes:

- older people
- women
- people who were self-employed
- unmarried (single) people
- non-believers (religious)
- people with a landline and mobile
- people who use the internet everyday/almost everyday
- people who reported being more left wing
- people who had friends in minority groups.

Dimensionality reduction models

Stepwise + lasso regularisation

Final variable selection

Linear mixed models - approach

Mixed model 1 - null model country level random effects only

↓

Mixed model 2 - adding in individual-level fixed effects

↓

Mixed model 3 - adding in country-level fixed effects

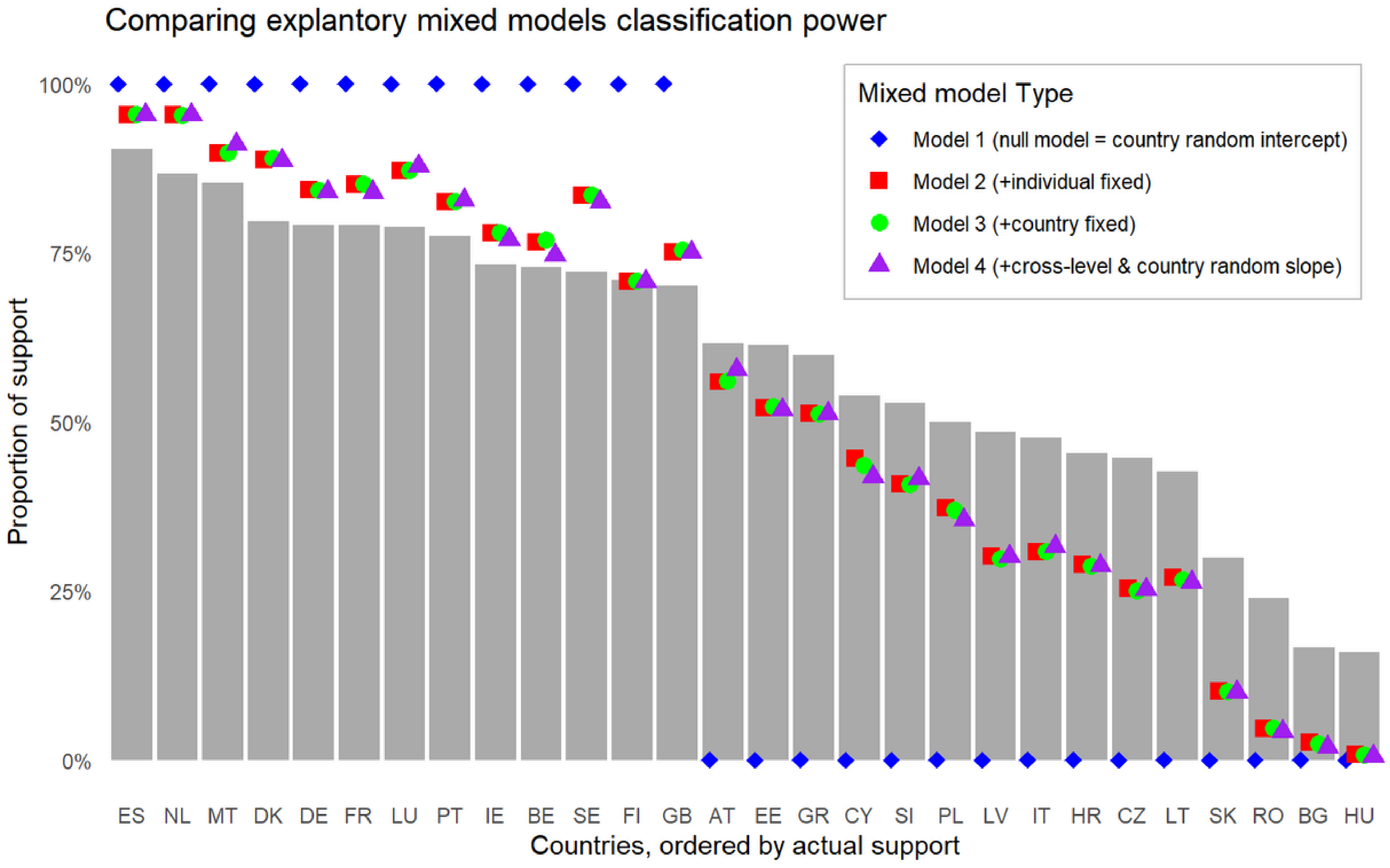
↓

Mixed model 4 - adding in cross-level interactions and random slopes

Linear Mixed models: performance

The best model was the full model with cross level interactions.

- However the improvements from the individual level effects are marginal.



Full model performance:

76% accuracy but better at identifying No (83.11%) than support (68.65%). This can be slightly improved by redcucing the threshold from 0.5 to 0.4

Cross-level effects

right_wing × lgbt_policy_index (negative)
genderWoman × lgbt_policy_index (positive)
non_believer × democracy_index (negative)
non_believer × gender_inequality_index (negative)

Predictive models - approach

Aggregating individual level data to country-level

↓

Set control method - LOOCV

↓

Predictive model 1 - Linear regression with Elastic Net

↓

Predictive model 2 - Random Forest

↓

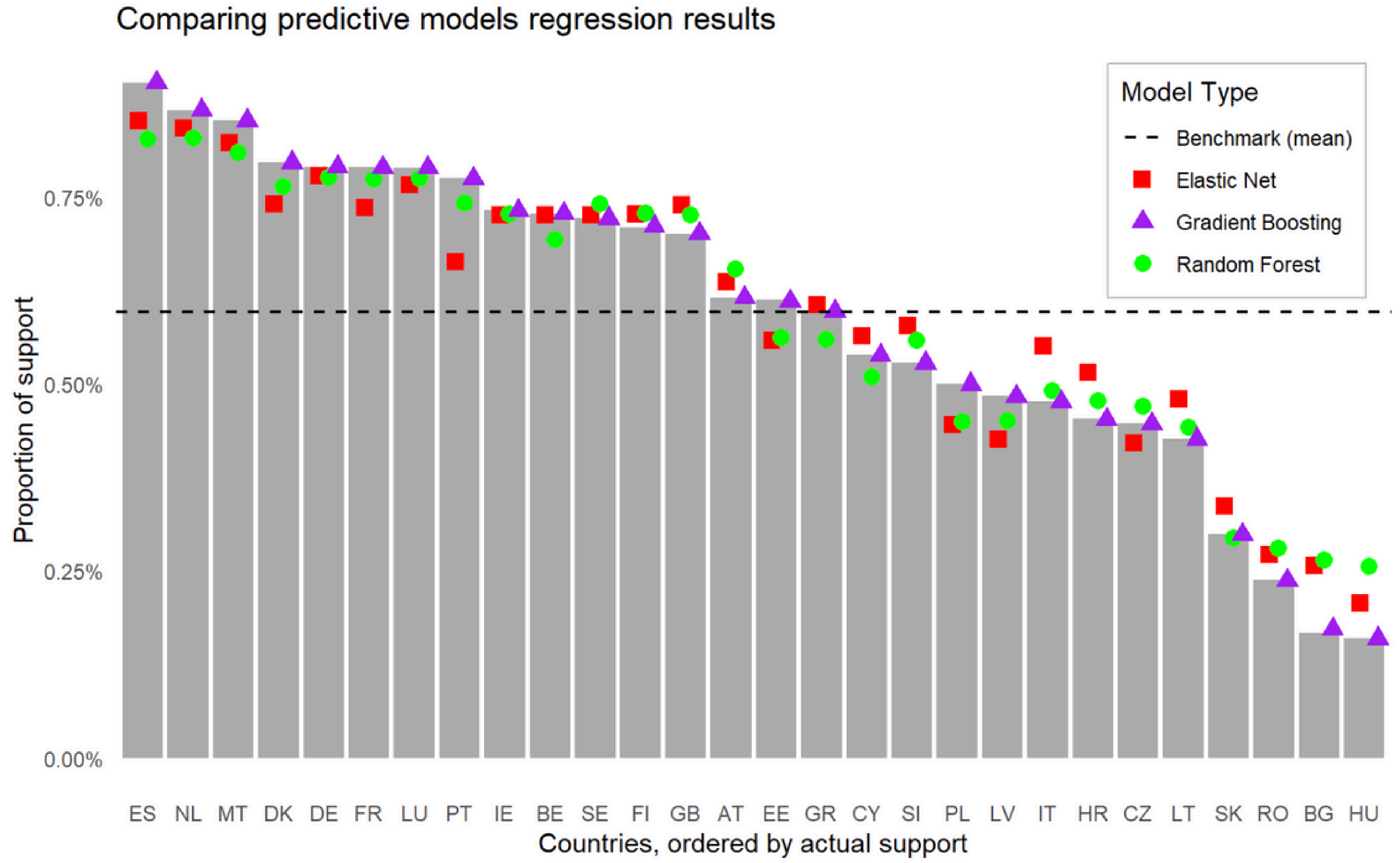
Predictive model 3 - Gradient boosting

Predictive models: comparison and findings

Performance metrics

	R squared	RMSE	MAE
Benchmark	0.00000	0.20707	0.17419
ElasticNet	0.94691	0.04771	0.04012
Random Forest	0.96040	0.04121	0.03413
Gradient Boosting	0.99996	0.00126	0.00061

Model fit by country



Final considerations

- Most of the variability of our mixed model was explained by **individual level** factors.

- **High predictive power** within the sample of our predictive models
- **Low generalizability** outside the sample due to the nature of the data

- All of our models found similar important variables.
  - Factors that were related to **supporting** the policy included:
    - being a woman
    - knowing a transgender person
    - being non-religious
    - being an everyday internet user
  - Factors related to **opposing** the policy included:
    - expressing anti-LGBTQ+ views
    - expressing more discriminatory views against minorities
    - having lower life satisfaction
    - expressing a more right wing ideology

Thank you!