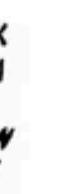


Grand Prix  
Eurovision  
de la Chanson  
Européenne  
1956

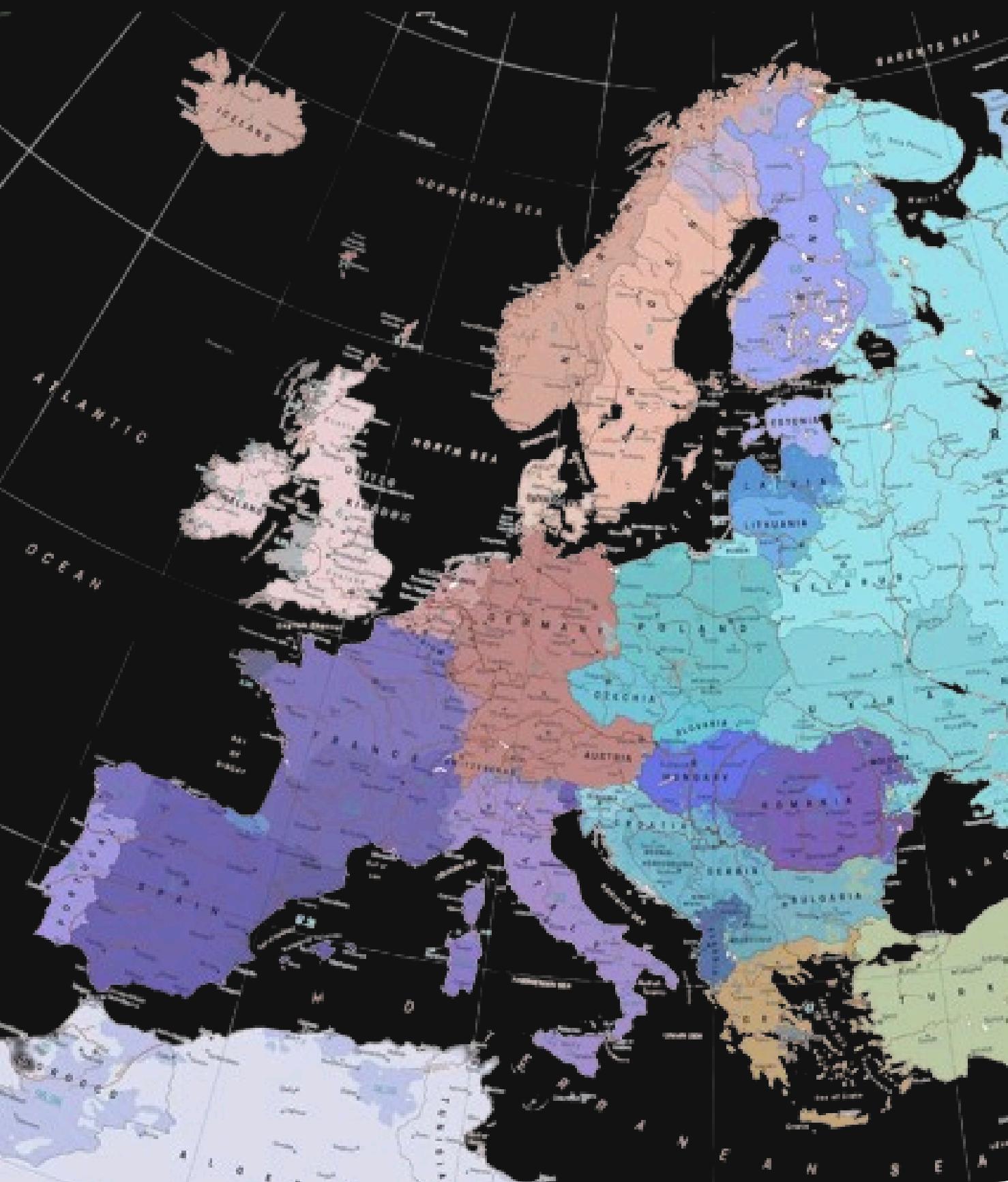


## ICS5115: Statistics for Data Scientists

# Eurovision Diplomacy: Investigating Neighbourhood Voting Through Data Analytics

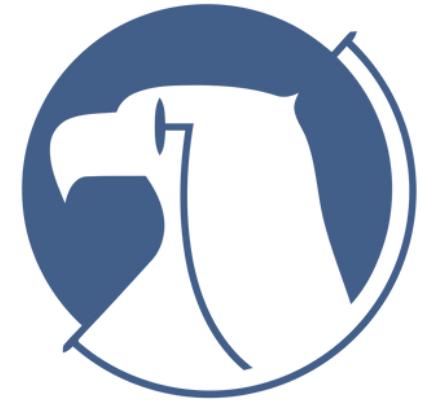
Nathan Portelli  
M.Sc. in AI

# Approach



- Simulating random vote allocation for a baseline
- Statistical analysis on ESC voting data and country profiles
- Correlations between voting patterns and factors such as;
  - Geographic Proximity
  - Linguistic Connections
  - Ethnic Affiliations
- Voting clusters

# Data Collection



GEO  
DATA  
SOURCE

Participating  
Countries &  
their Votes

- List of every participant between 1956 and 2023
- List of each individual vote from every country from 1957 until 2023
- [Metadata from EurovisionWorld](#)

Bordering  
States

- Land borders
- Major maritime borders manually included
- [GeoDataSource](#)

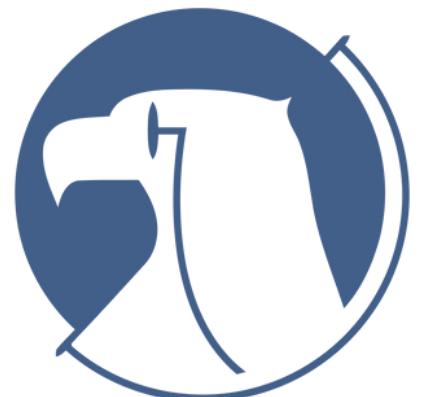
Major  
Languages

- The primary languages spoken in each country
- [Infoplease](#)

Major Ethnic  
Groups

- Every major ethnic group in each country
- [Scraped from CIA World Factbook](#)

# Data Cleansing



GEO  
DATA  
SOURCE

Participating  
Countries &  
their Votes

Bordering  
States

Major  
Languages

Major Ethnic  
Groups

- Removal of inconsistencies in country names
- Handling of historical complexities
- Exclusion of the WLD vote

- Standardisation of country names
- Addition of major maritime borders

- Extraction and categorization of language data
- Filtering data for language names
- Merging language information with country codes

- Extraction of ethnic data
- Filtering to remove unnecessary information

# Analysis of Voting Patterns

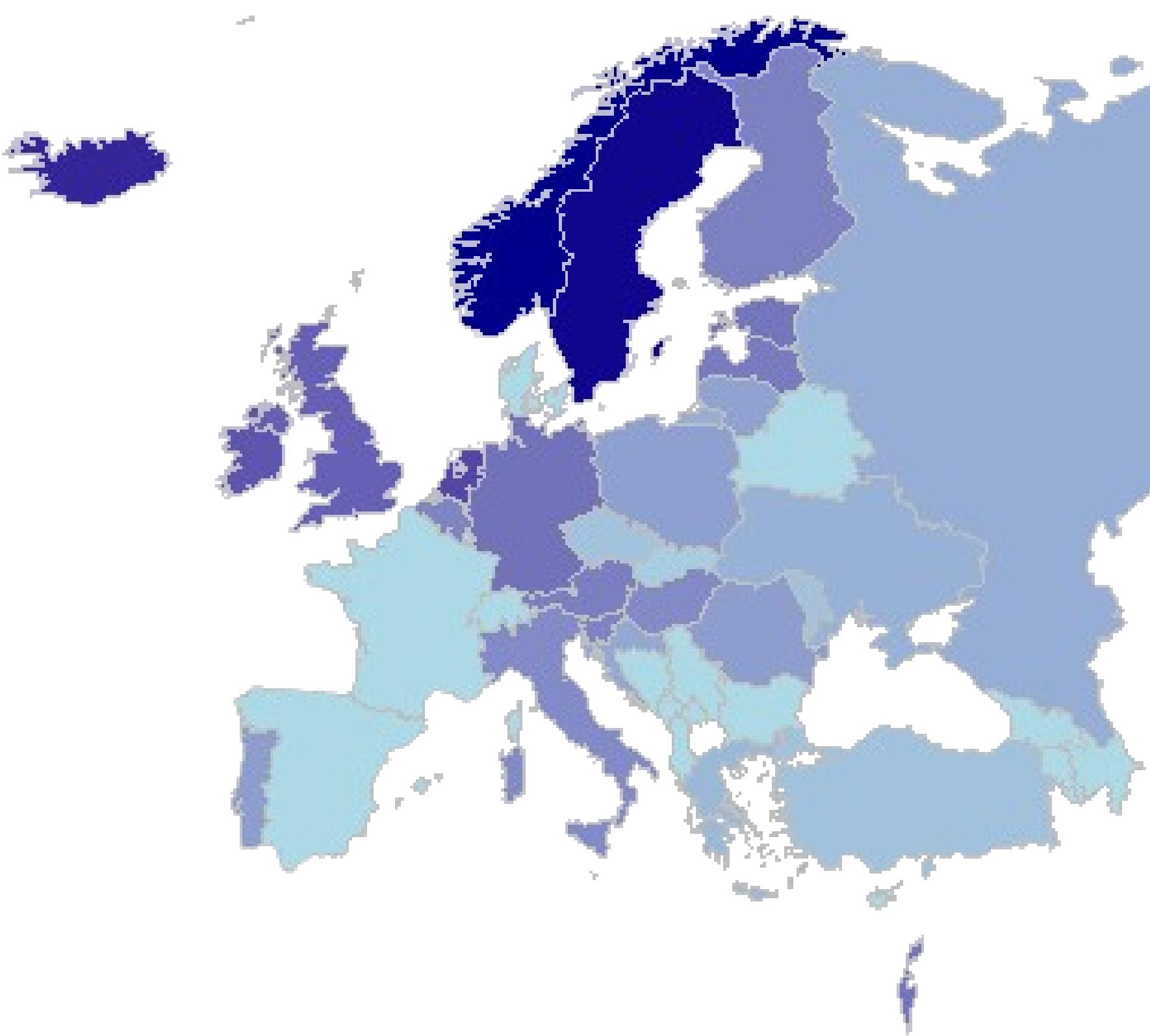
## Null Hypothesis ( $H_0$ )

Countries *do not* exhibit neighbourhood voting in the ESC; voting patterns are *random* and *not influenced* by geographic proximity, linguistic connections, or ethnic affiliations.

## Alternative Hypothesis ( $H_1$ )

Countries exhibit neighbourhood voting in the ESC; voting patterns are influenced by geographic proximity, linguistic connections, and ethnic affiliations

# Votes awarded to Denmark



# Voting Procedures

1957-1961  
1967-1970  
1974

- Each country awarded 10 points to distribute among other participants

1962-1963

- Maximum points set to 3 and 5 in respective years

1964-1966

- Points assigned from {1, 3, 5, 6, 9}

1971-1973

- Simplified rating system with points from 1 to 5

1975-2015

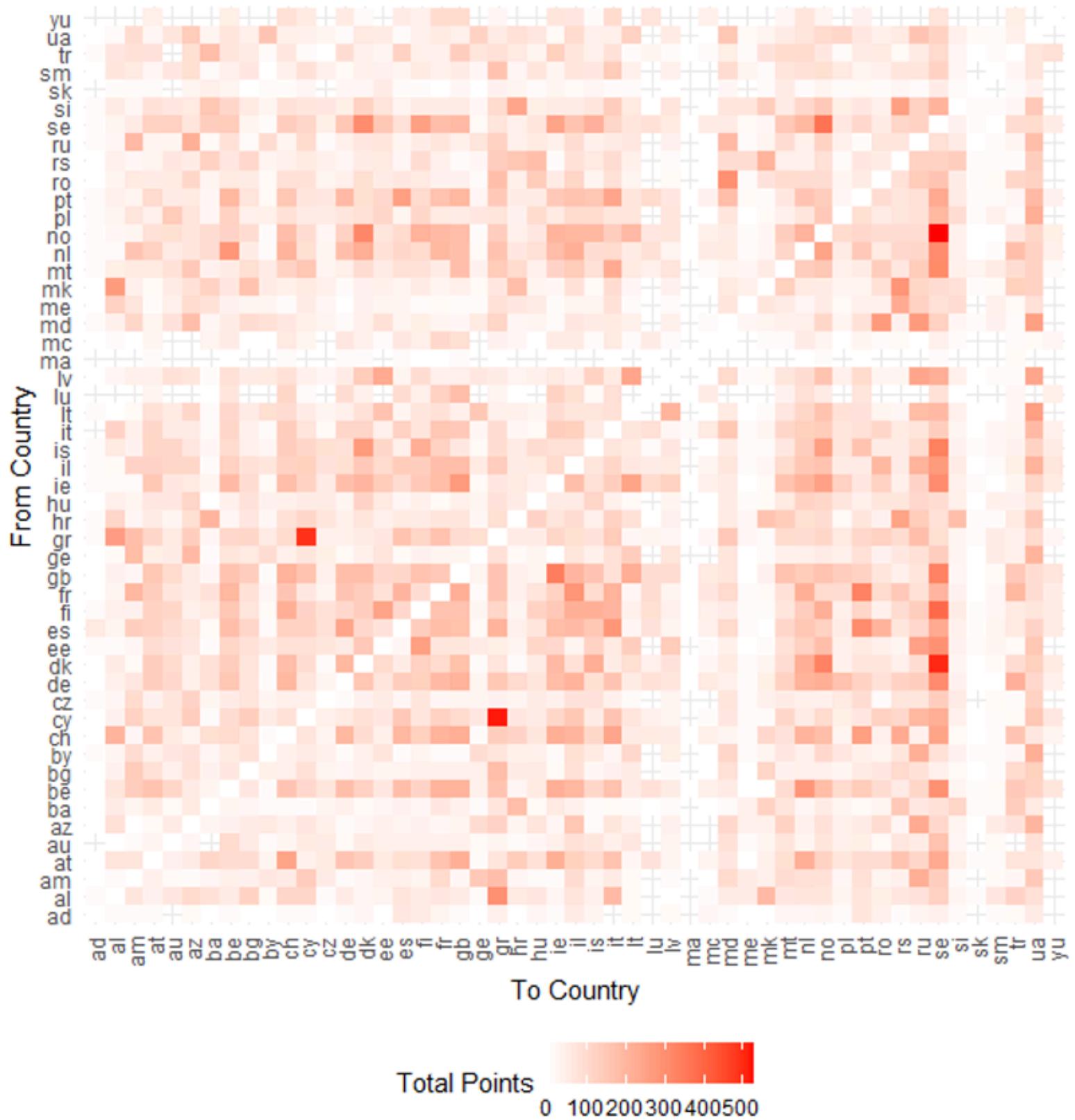
- Structured voting system with points from 1 to 12
- Integration of televoting alongside jury voting

2016-2023

- Continued use of structured voting system with points from 1 to 12
- Universal adoption of televoting alongside jury panels
- Separate but identical voting points for jury and televoters

# Randomised Voting

Actual Votes

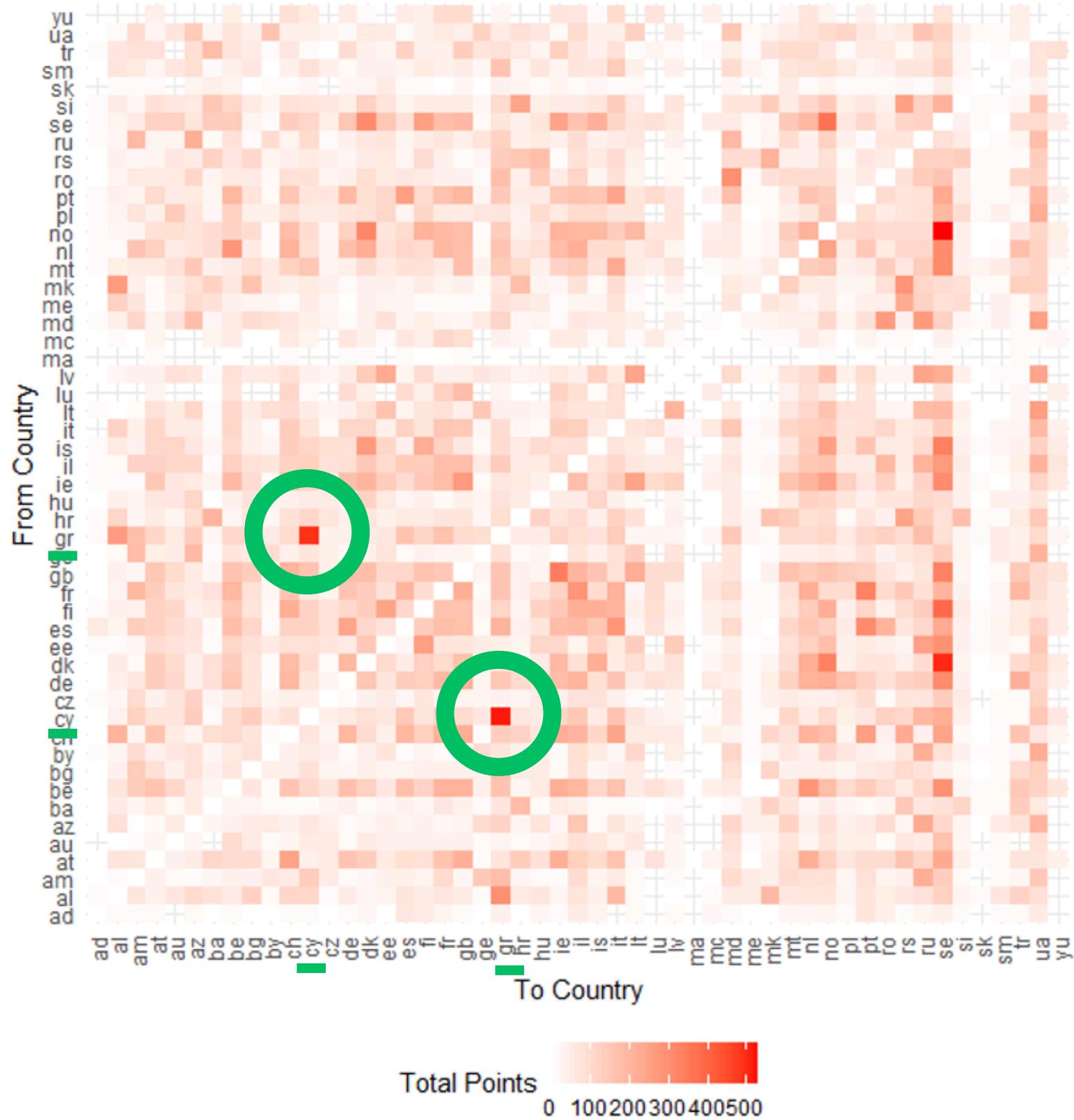


Random Votes

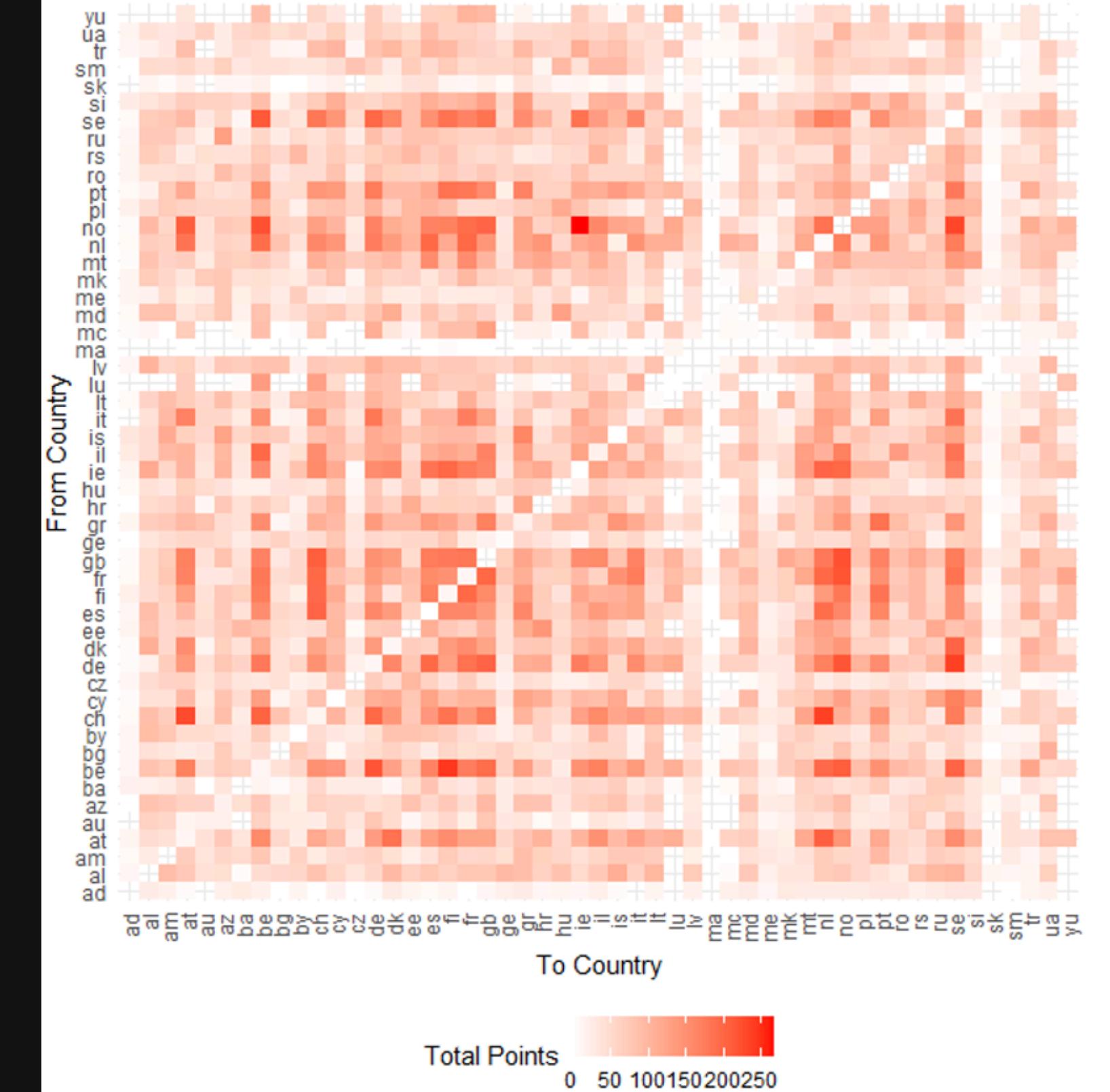


# Heatmap of Voting Patterns

## Actual Votes

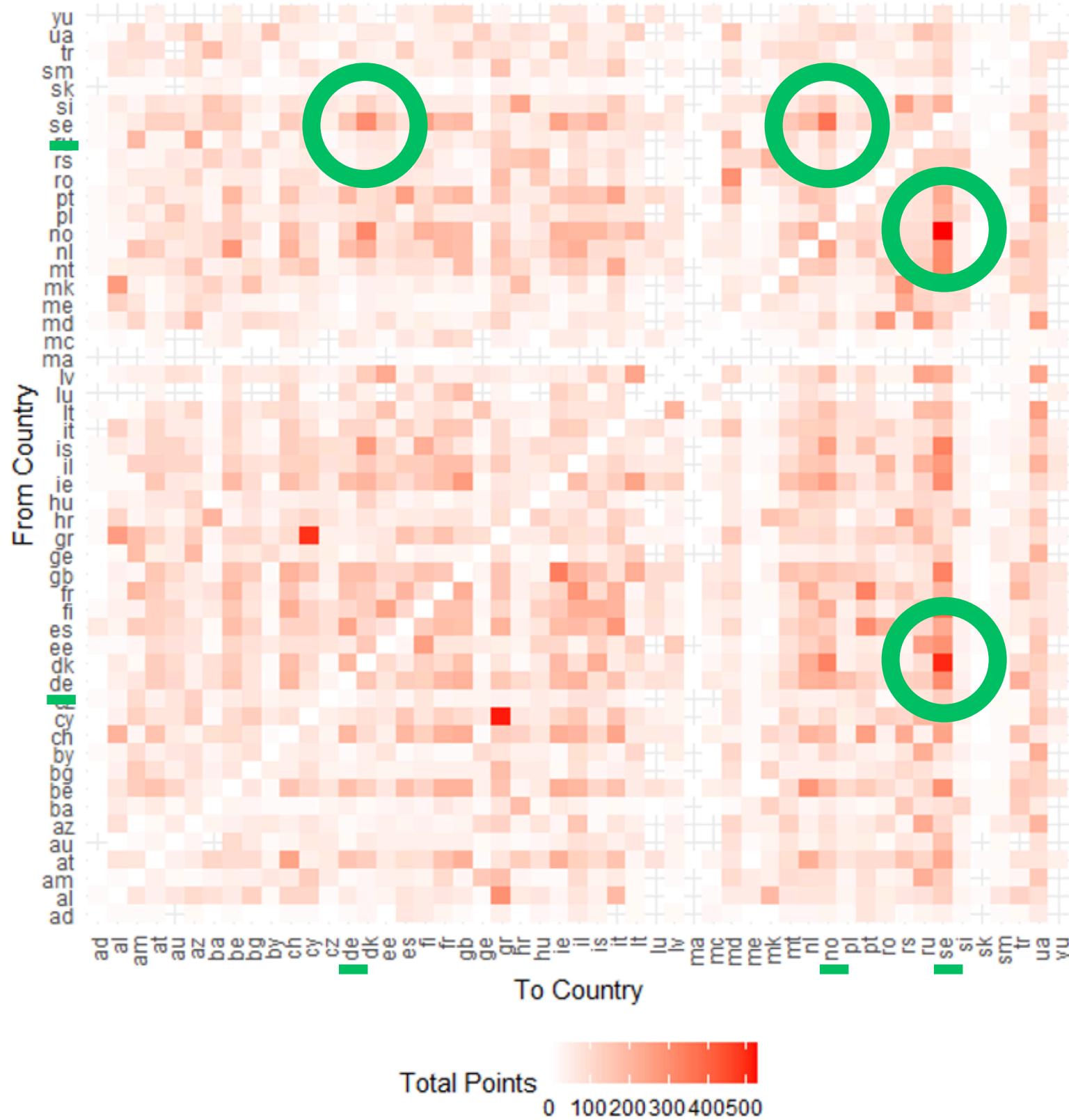


## Random Votes



# Heatmap of Voting Patterns

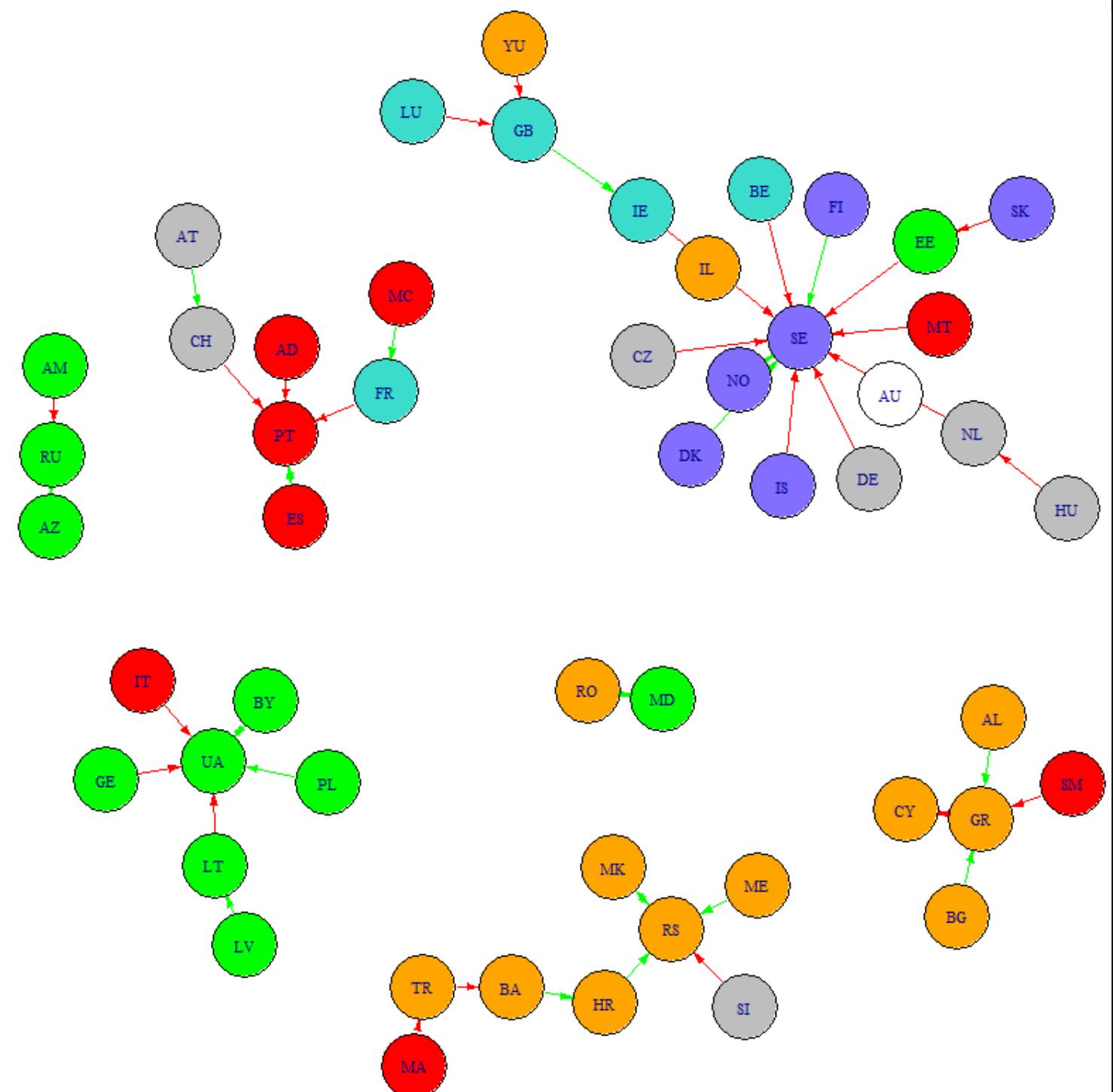
## Actual Votes



## Random Votes



# Network Analysis



## Key Findings

- Reciprocal high-voting tendencies between countries observed.
- Top Voters
  - 47.06% share borders
  - 23.72% share languages
  - 35.87% share ethnicities

## Network Centrality Analysis

- Degree Centrality
- Closeness Centrality
- Top Countries
  - Sweden
  - Norway
  - United Kingdom
  - Germany

South-West: Red  
North-West: Turquoise  
North: Blue  
Central: Grey

South-East: Orange  
East: Green  
Other: White

# Correlation Analysis

Geographic Analysis

0.8393

Linguistic Analysis

0.1021

Ethnic Analysis

0.1010

Implications

- Geographic proximity plays a significant role in voting behaviors, with border-sharing countries showing a strong tendency to reciprocate votes.
- Shared language and ethnicities contribute at a lesser extent.

# Clustering Analysis

CLUSTERING  
COEFFICIENTS

Mean Clustering Coefficient

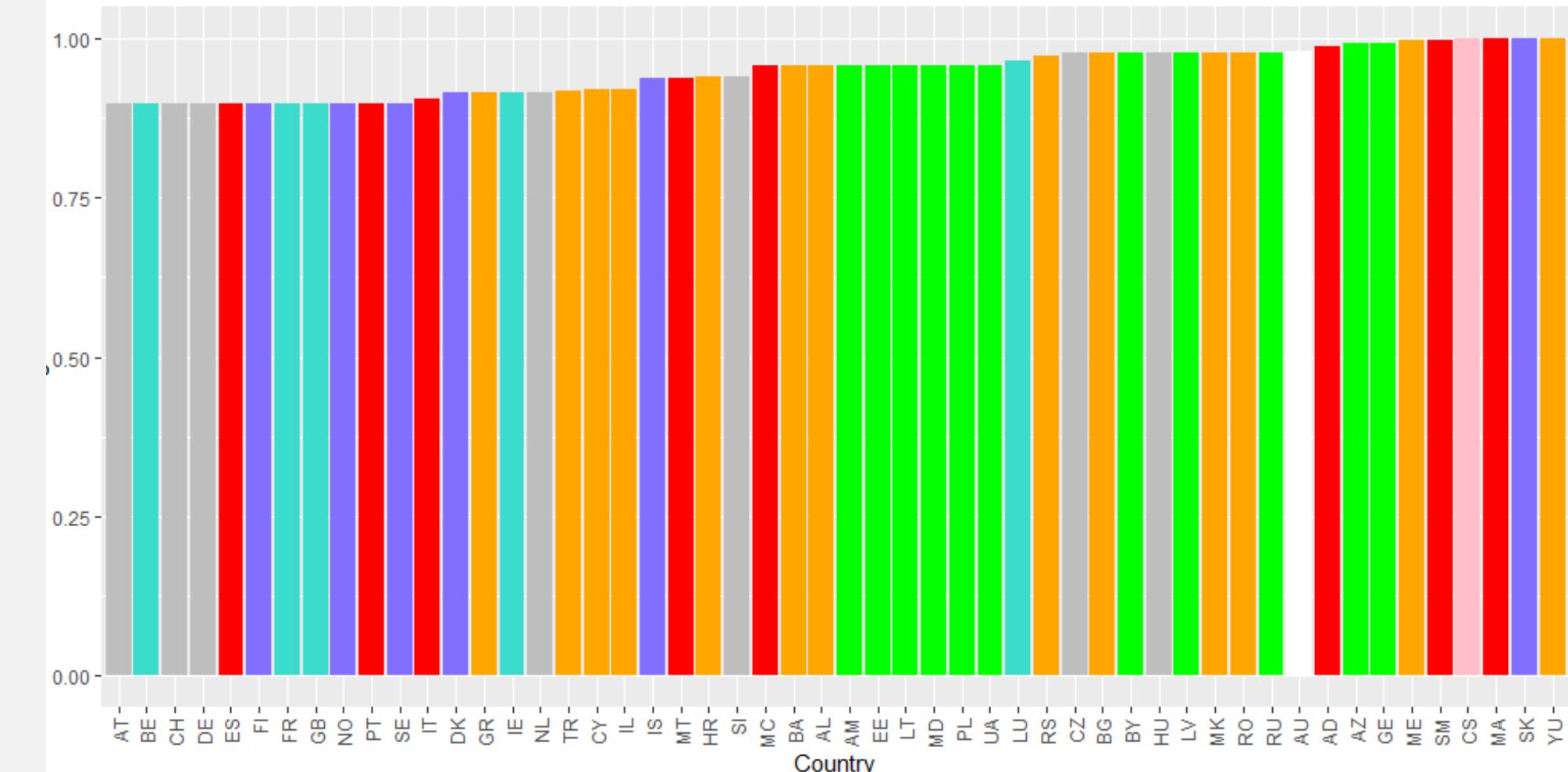
0.957

Median Clustering Coefficient

0.972

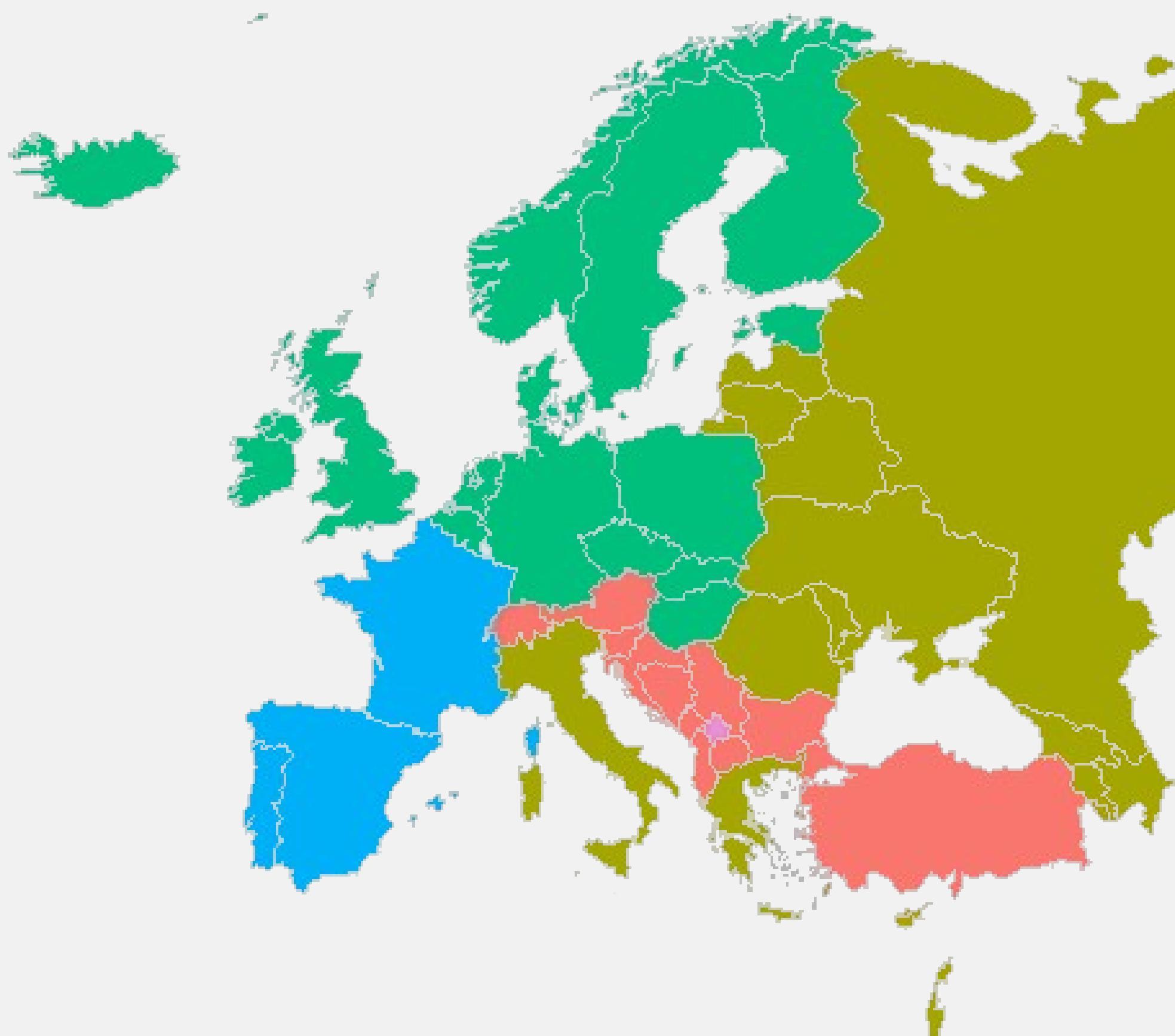
Standard Deviation

0.033



# Clustering Analysis

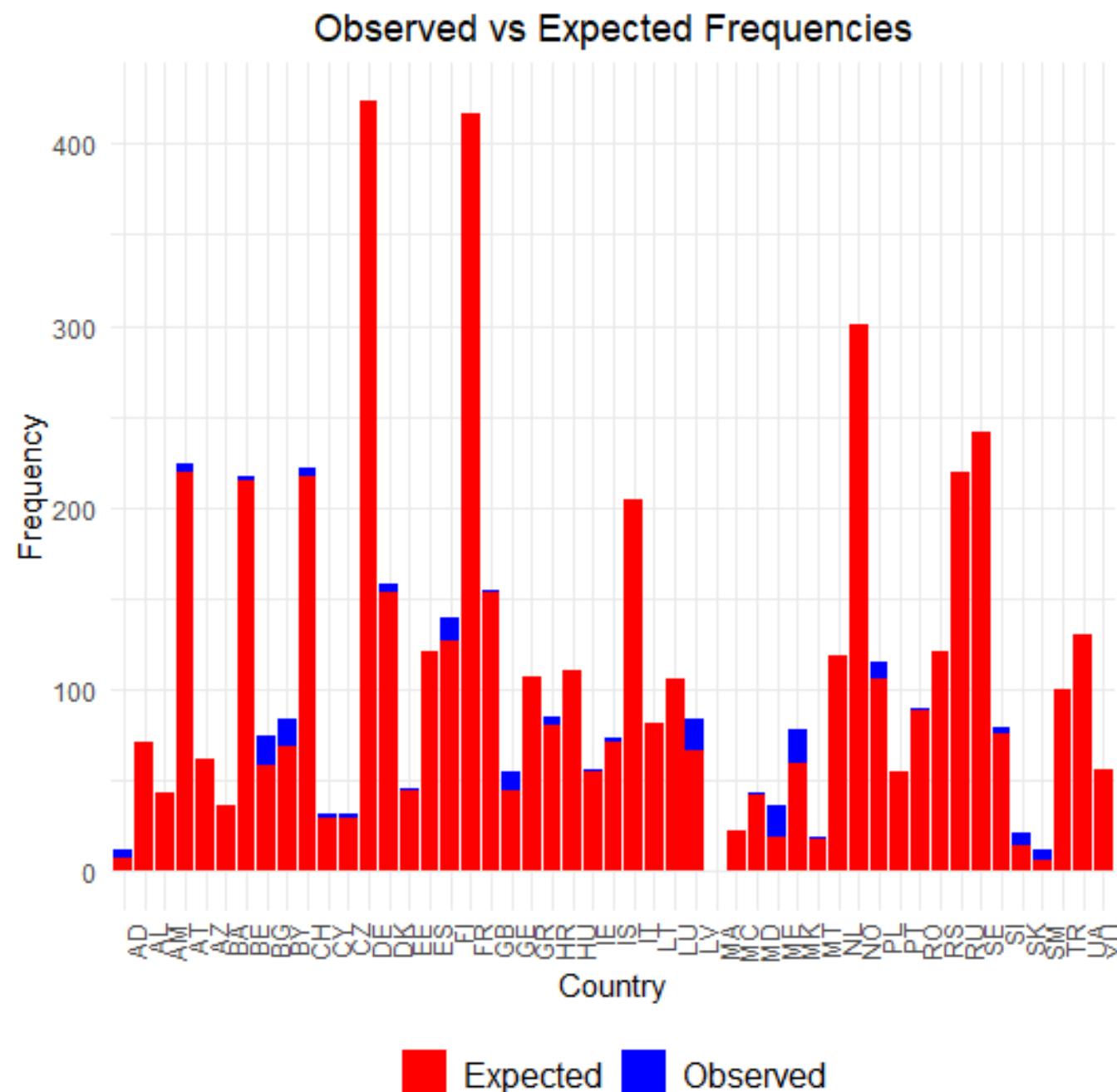
COMMUNITY  
DETECTION



# Statistical Tests

## Chi-Square Test

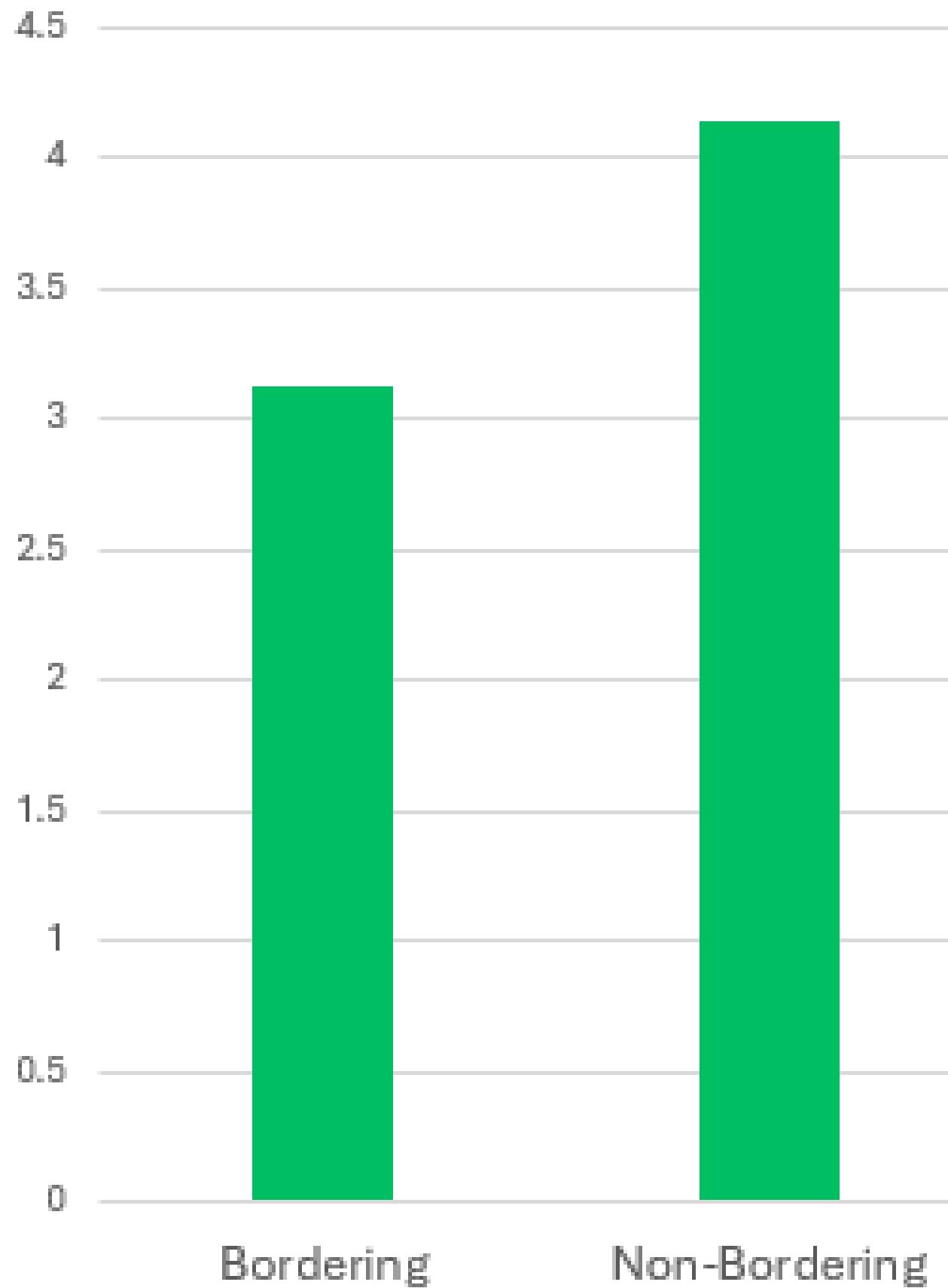
- Significant association found ( $p\text{-value} < 2.2\text{e-}16$ )
- Indicates non-random voting patterns among bordering countries
- Supports hypothesis of geographic influence on voting behaviour



# Statistical Tests

## T-Test

- Significant difference found ( $t = -37.248$ ,  $df = 102640$ ,  $p < 2.2 \times 10^{-16}$ )
- 95% confidence interval for difference in means: -1.0836 to -0.9752
- Mean points by bordering countries: 3.1221
- Mean points by non-bordering countries: 4.1515
- Proximity significantly influences voting preferences



# Conclusions

## Key Findings

- Voting patterns in ESC are significantly influenced by geographical proximity, as well as linguistic similarities and ethnic affiliations to a lesser extent.
- Analyses reveal non-random voting behaviors and voting blocs.

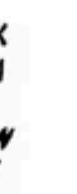
## Assumptions & Limitations

- Analysis based on primary languages spoken in each country, not songs.
- Focus on countries over performance.
- No analysis of news or social media.

## Future Direction

- Evolution & impact of voting patterns
- Integration of sentiment analysis
- Incorporation of performance quality
- Data on geopolitical events.

Grand Prix  
Eurovision  
de la Chanson  
Européenne  
1956



# ICS5115: Statistics for Data Scientists

## Eurovision Diplomacy: Investigating Neighbourhood Voting Through Data Analytics

Nathan Portelli  
M.Sc. in AI