Modeling and Predicting Conspiracy Belief

Capstone Sprint 1

Elliot Carter

The Problem Area

- Big challenges need cooperative solutions.

- Misinformation prevents cooperation.

- Solution requires understanding who believes misinformation and why.

How Data Science Can Help

- Much research on conspiracy theories has sample sizes too small to support ML approaches.

- Recent survey by Imhoff et al produced a large dataset (~100k respondents).

- My project: explore and model their data using ML.

Conspiracy mentality and political orientation across 26 countries

Roland Imhoff[1 ✉], Felix Zimmer[1], Olivier Klein[2], João H. C. António[3], Maria Babinska[4], Adrian Bangerter[5], Michal Bilewicz[4], Nebojša Blanuša[6], Kosta Bovan[6], Rumena Bužarovska[7], Aleksandra Cichocka[8], Sylvain Delouvée[9], Karen M. Douglas[8], Asbjørn Dyrendal[10], Tom Etienne[11], Biljana Gjoneska[12], Sylvie Graf[13,14], Estrella Gualda[15], Gilad Hirschberger[16], Anna Kende[17], Yordan Kutiyski[11], Peter Krekó[17], Andre Krouwel[18], Silvia Mari[19], Jasna Milošević Đorđević[20], Maria Serena Panasiti[21], Myrto Pantazi[22], Ljupcho Petkovski[23], Giuseppina Porciello[21], André Rabelo[24], Raluca Nicoleta Radu[25], Florin A. Sava[26], Michael Schepisi[21], Robbie M. Sutton[8], Viren Swami[27,28], Hulda Thórisdóttir[29], Vladimir Turjačanin[30], Pascal Wagner-Egger[31], Iris Žeželj[32] and Jan-Willem van Prooijen[18]
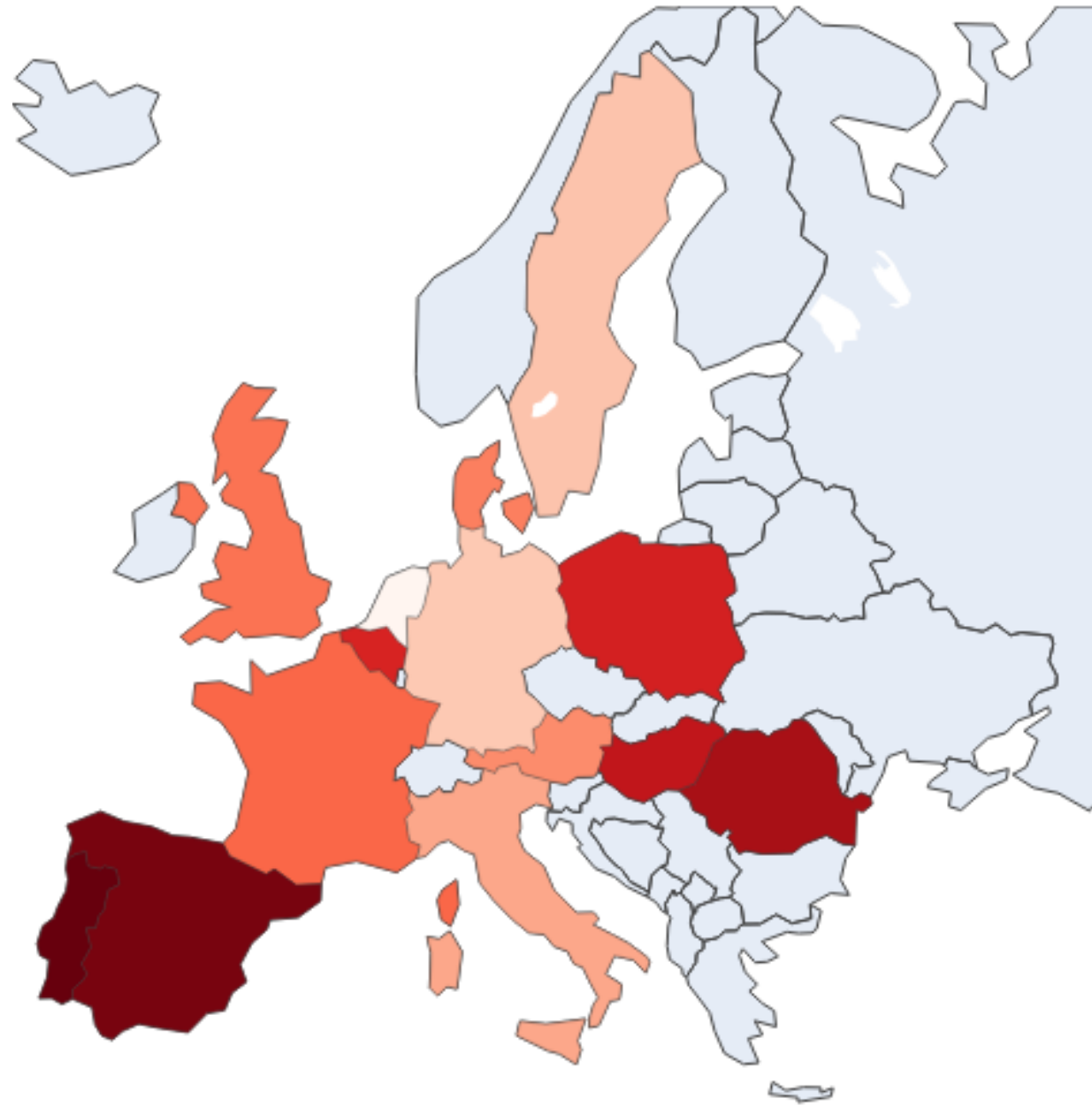
Introduction to the Dataset

- Two surveys, 26 countries, 100k respondents total

- Uses the *Conspiracy Mentality Questionnaire* (CMQ) to measure conspiracy susceptibility.

  - Ratings of agreement with five statements, like "I think that government agencies closely monitor all citizens."

Introduction to the Dataset

- Other variables:

  - Demographic features (age, sex, country, education)
  - Measures of political orientation
  - Belief in country-specific conspiracies

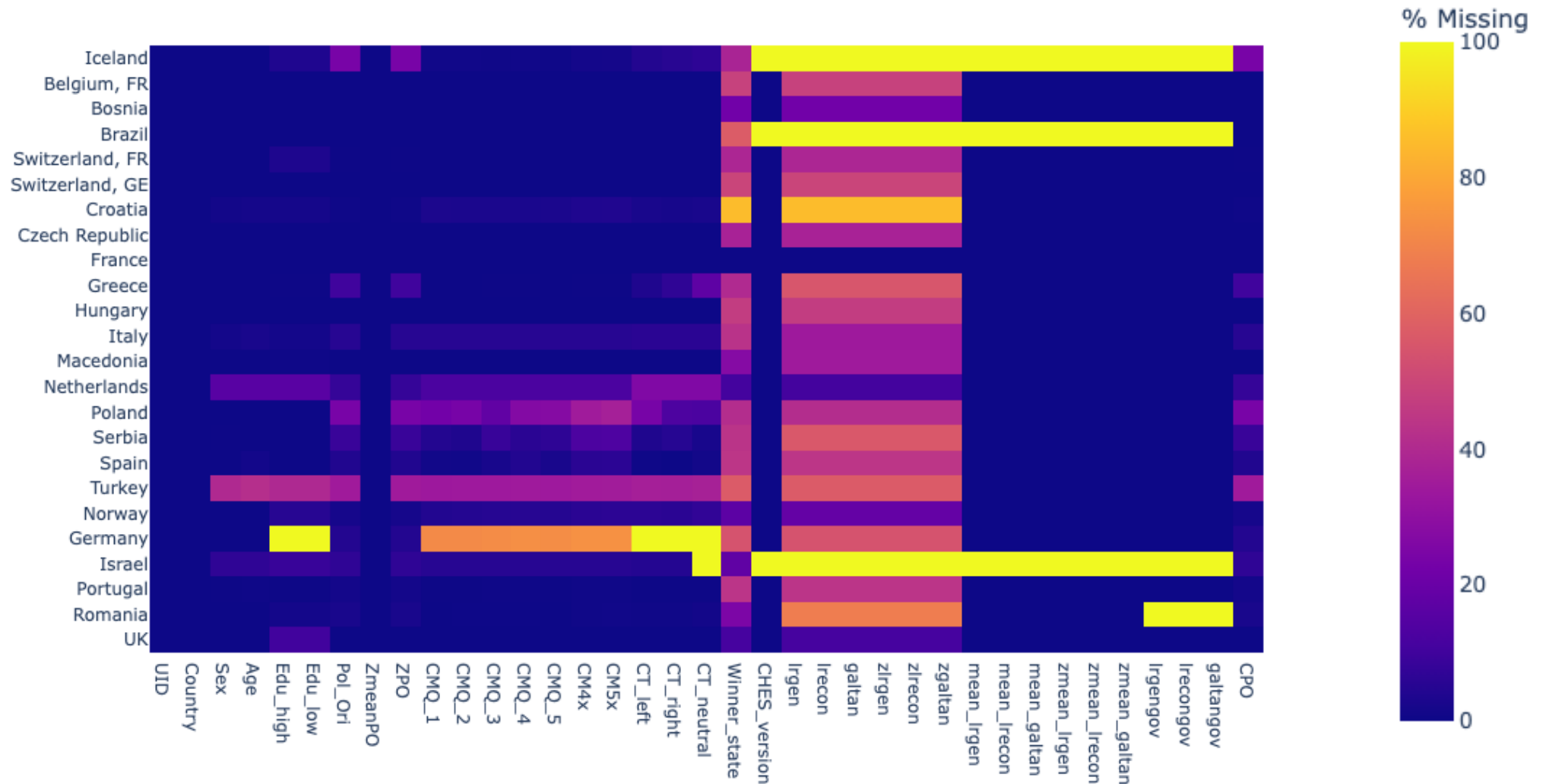**Average Conspiracy Mentality Score by Country**

Average CM Score Out of 10

Potential Concerns From Early Analysis

- Is the CMQ a valid measure?
  - And how to encode target? (Average? Average of certain questions? Low/high?)
- How to join data between two surveys, or split between countries
- Missing data
  - Both quantity and patterns…

Percentage of Missing Data by Country

Next Steps

- Determine what to impute (and how), what to drop (and for which purposes)
- Construct simple linear models; see where there's potential for ML
- Preprocessing features and statistical tests