

The background of the entire image is a photograph of a large, diverse crowd of people. Many individuals have their right hands raised, suggesting a protest or a rally. Some people are holding up smartphones to take pictures or record video. In the foreground, a person's hand is visible holding a clear plastic bottle. The scene is outdoors, with buildings and other protesters visible in the background.

SPRINGBOARD CAPSTONE PROJECT

MASS MOBILIZATION: PREDICTING STATE VIOLENCE

A Report by Jessica Montealvo

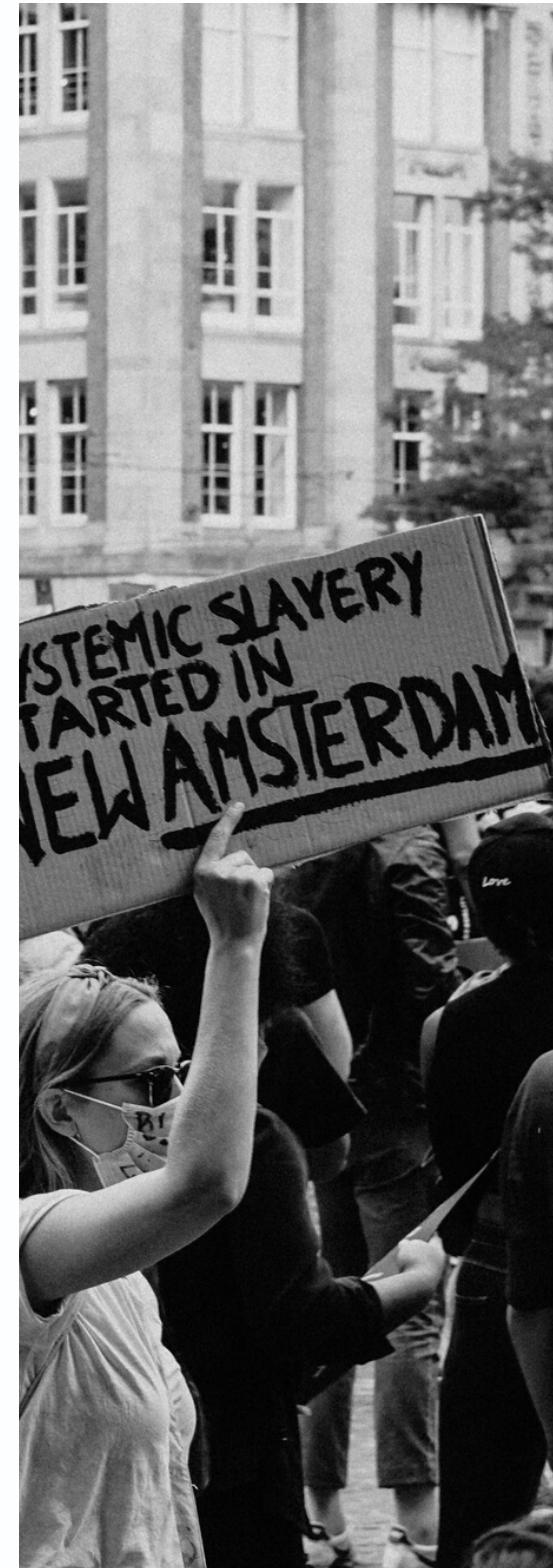
Introduction

WHY IS THIS TOPIC RELEVANT?

Since the dawn of democracy, protests have played an essential role in maintaining the power of the people. Research indicates an upwards trend of the rate of protests world wide. The outcome of a protest can vary between extremes of accommodation to violence with casualties.

Problem Statement

THIS PROJECT AIMS TO CREATE A MODEL THAT DISCERNS WHETHER A MASS MOBILIZATION WILL BE VIOLENT OR NOT AT ITS INCEPTION, TO DRIVE PROTESTER STRATEGIES AND INFORM DECISIONS WHEN DIALOGUE AND ENGAGEMENT -- POLITICAL OR COMMUNITY -- ARE NO LONGER AN OPTION.





The Data

"Mass Mobilization Data Project Dataverse"

David H. Clark (Binghamton University) and Patrick M. Regan (University of Notre Dame)

17,000 protests against governments across

- across 162 countries
- during 1990-2020
- Protester demands (up to 4)
- State response (up to 7)
- Country and Region
- Identity and number of protests

DATA WRANGLING

The unit of observation for this data set is the protest-country-year, where each demonstration is recorded individually within country and year.

- **Raw data set**

26 variables recorded for
17,145 events

- **Categorical columns**

Created boolean columns for
each of the protester demand
and state response types.
Any observations missing a
state response were dropped.

- **Cleaned data set**

53 variables and 14,482
events

- **Additional Columns**

`participants_category`
`response_combo`,
`demand_combo`,
`first_response`,
`last_response`,
`demands_count`,
`responses_count`



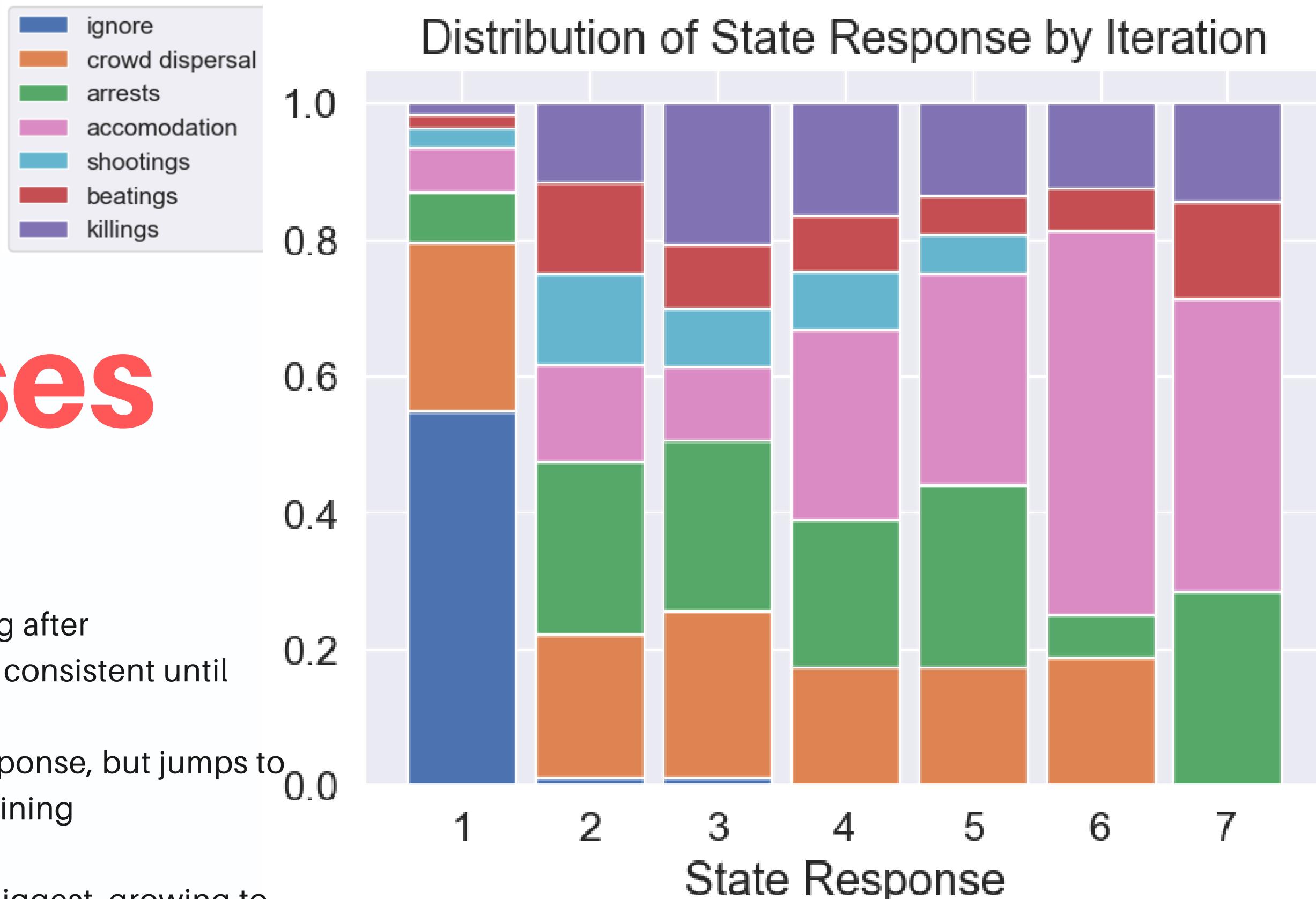


Exploration Data Analysis

State Responses

TRENDS

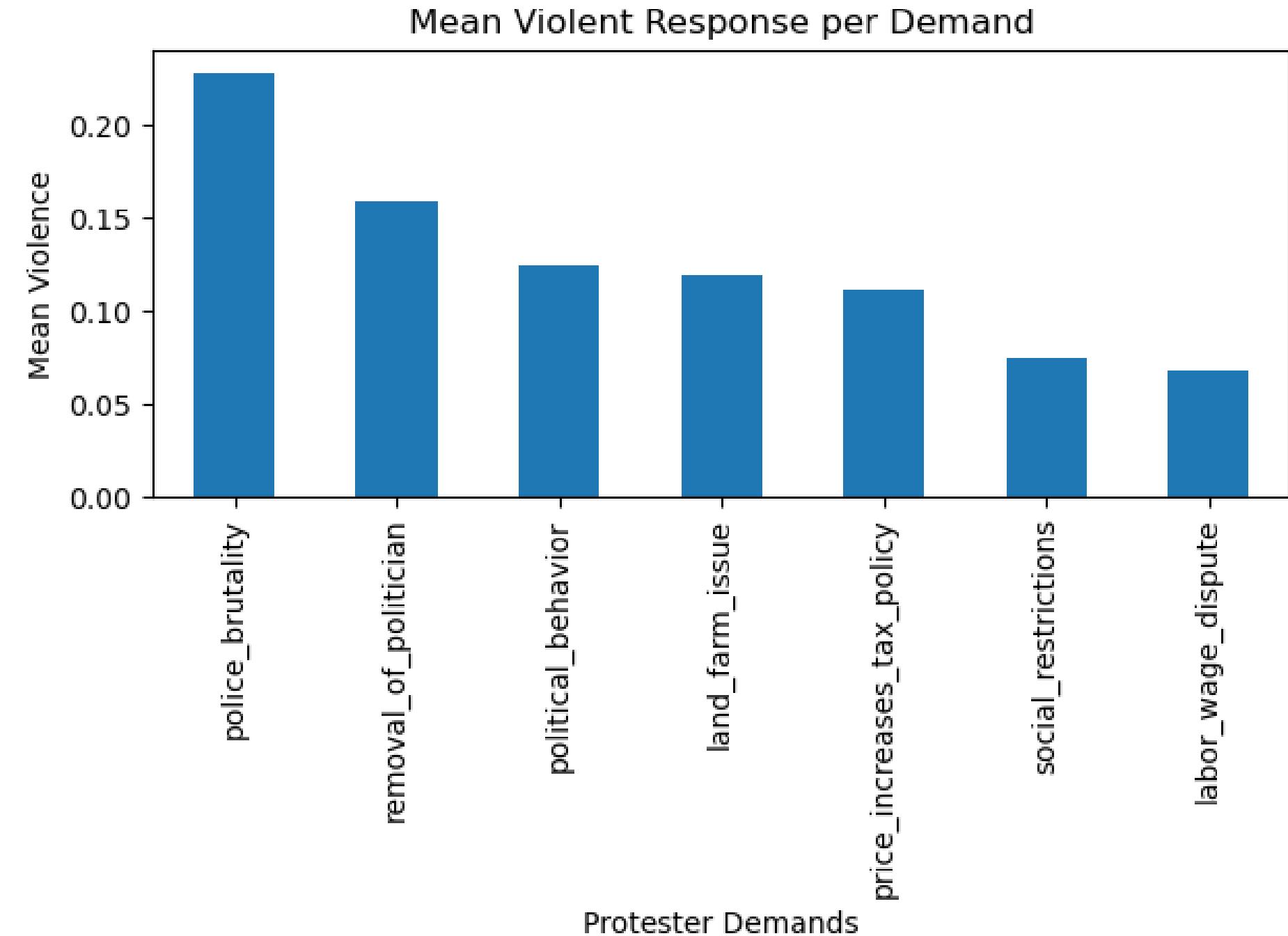
- 1) 50% ignore at first, drops to nothing after
- 2) Crowd dispersal remains relatively consistent until the last state response
- 3) Violence largely absent on first response, but jumps to 40% on subsequent responses, remaining relatively consistent
- 4) Accommodation seems to be the biggest, growing to represent the majority of responses after several iterations
- 5) Arrests relatively consistent after first response



Mean Violent Response per Demand

POLICE BRUTALITY, REMOVAL OF POLITICIANS, LABOR WAGE DISPUTE, AND SOCIAL RESTRICTIONS ARE ALL STATISTICALLY SIGNIFICANTLY ($P < 0.001$) DIFFERENT FROM THE MEAN VIOLENT RESPONSE OVERALL.

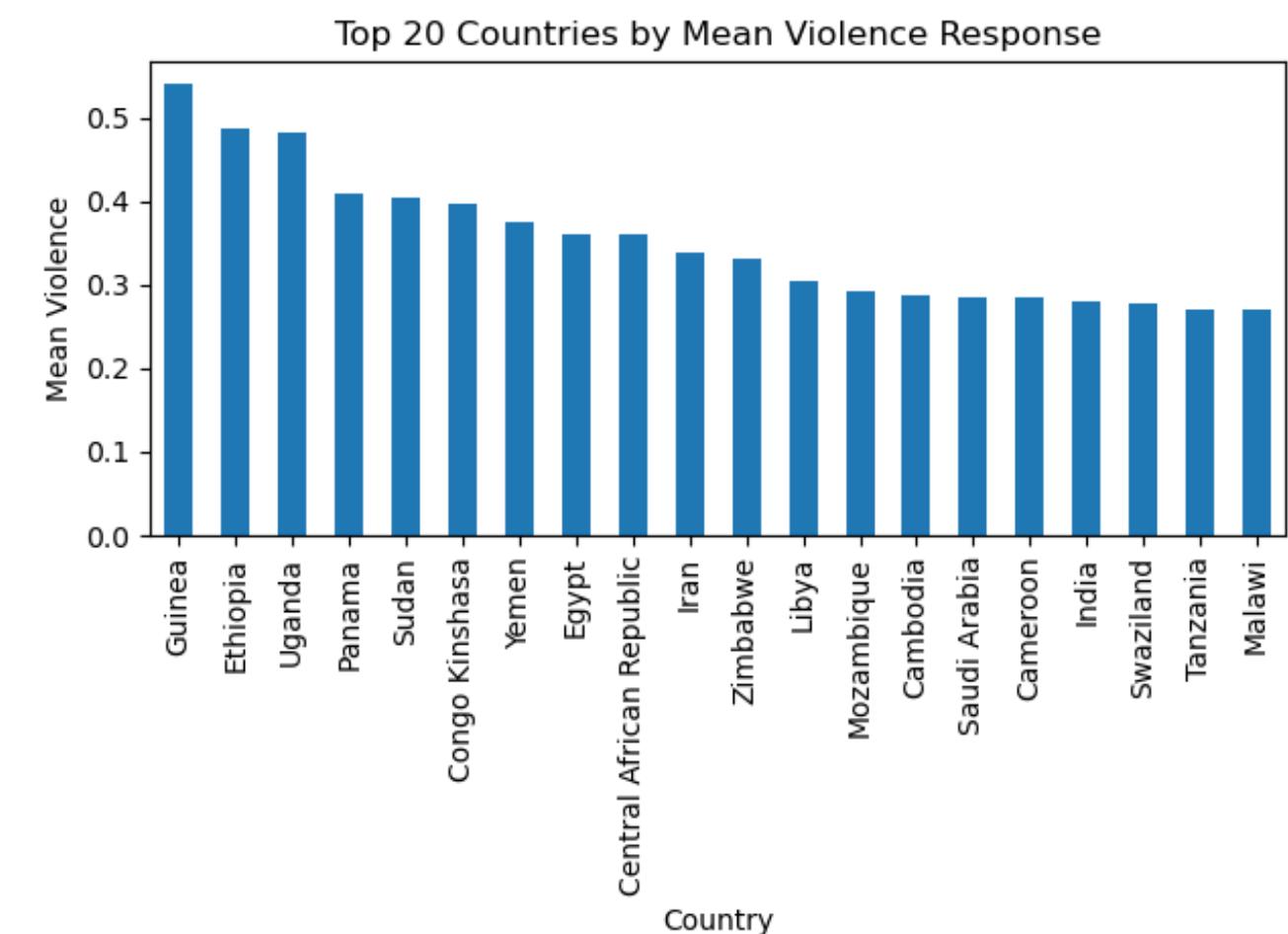
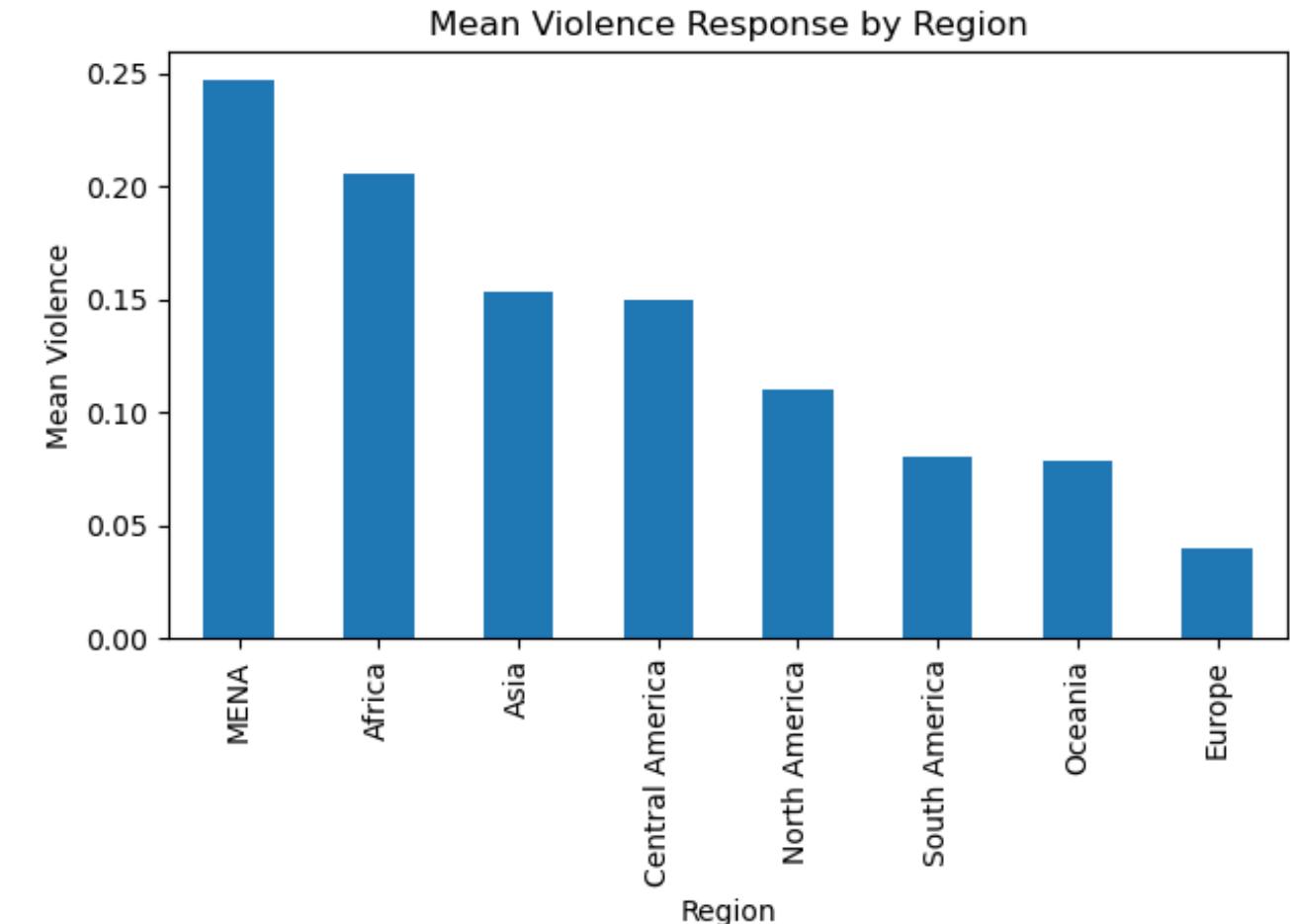
All seven categories of protest demands experienced violent responses, which means that although the protest demand(s) might be direct indicators of violent response, there are likely additional features driving this behavior.



Mean Violent Response by Region and Country

REGION AND COUNTRY ARE STATISTICALLY SIGNIFICANTLY ($P < 0.001$) DIFFERENT FROM THE MEAN VIOLENT RESPONSE OVERALL.

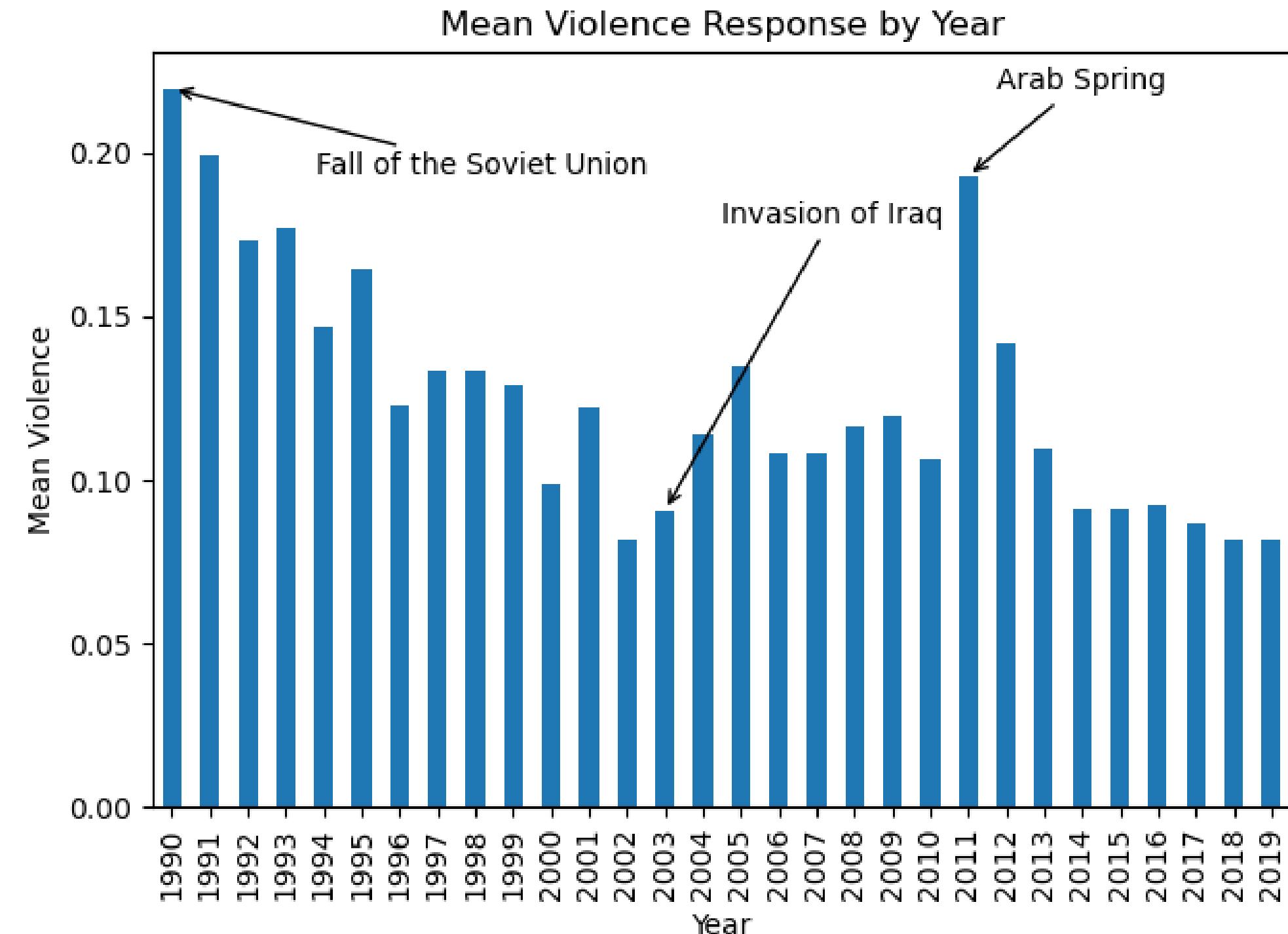
Even though most of the recorded demonstrations happened in Europe ($n=4808$), it is the Middle East (MENA, $n=1093$) that on average reacted with violence to mass mobilizations most, followed by Africa ($n=3037$) and Asia ($n=2977$).



Mean Violent Response by Year

YEAR IS STATISTICALLY SIGNIFICANTLY ($P < 0.001$) DIFFERENT FROM THE MEAN VIOLENT RESPONSE OVERALL.

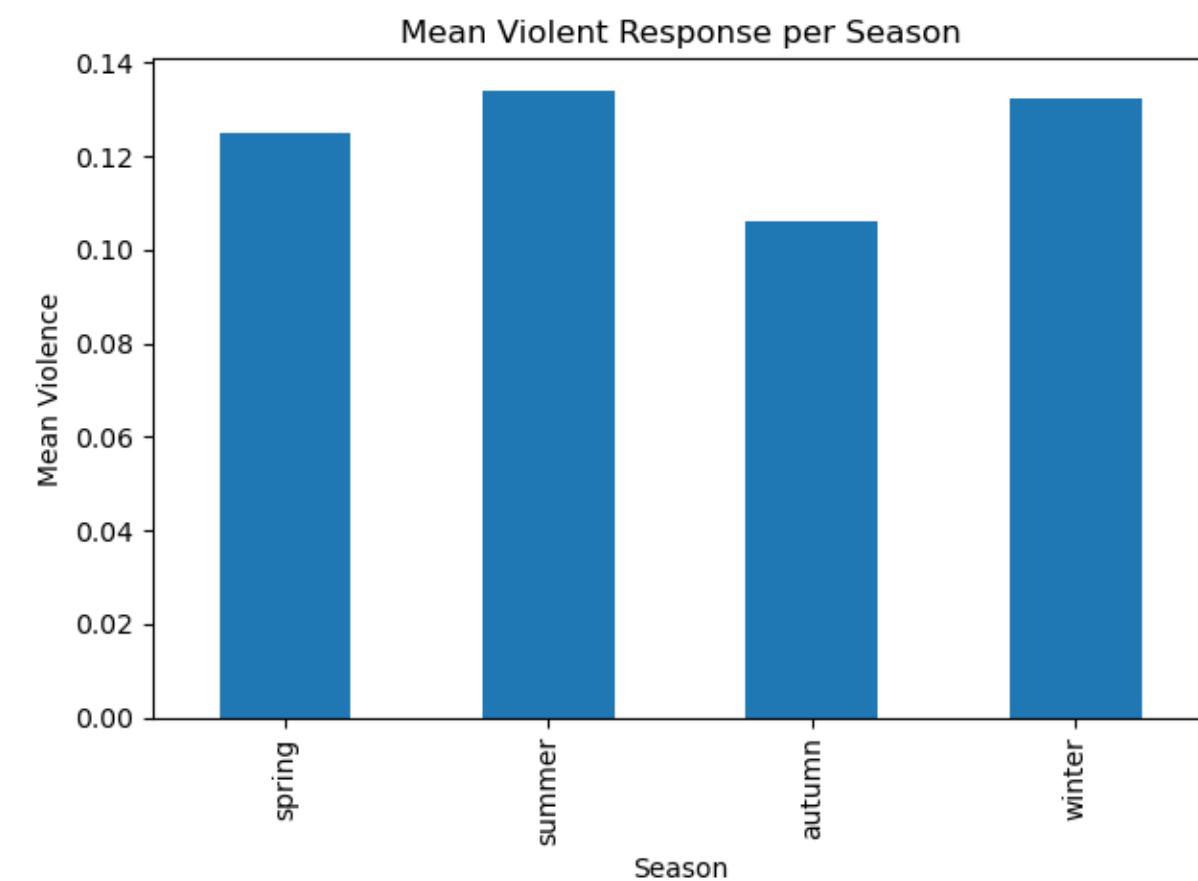
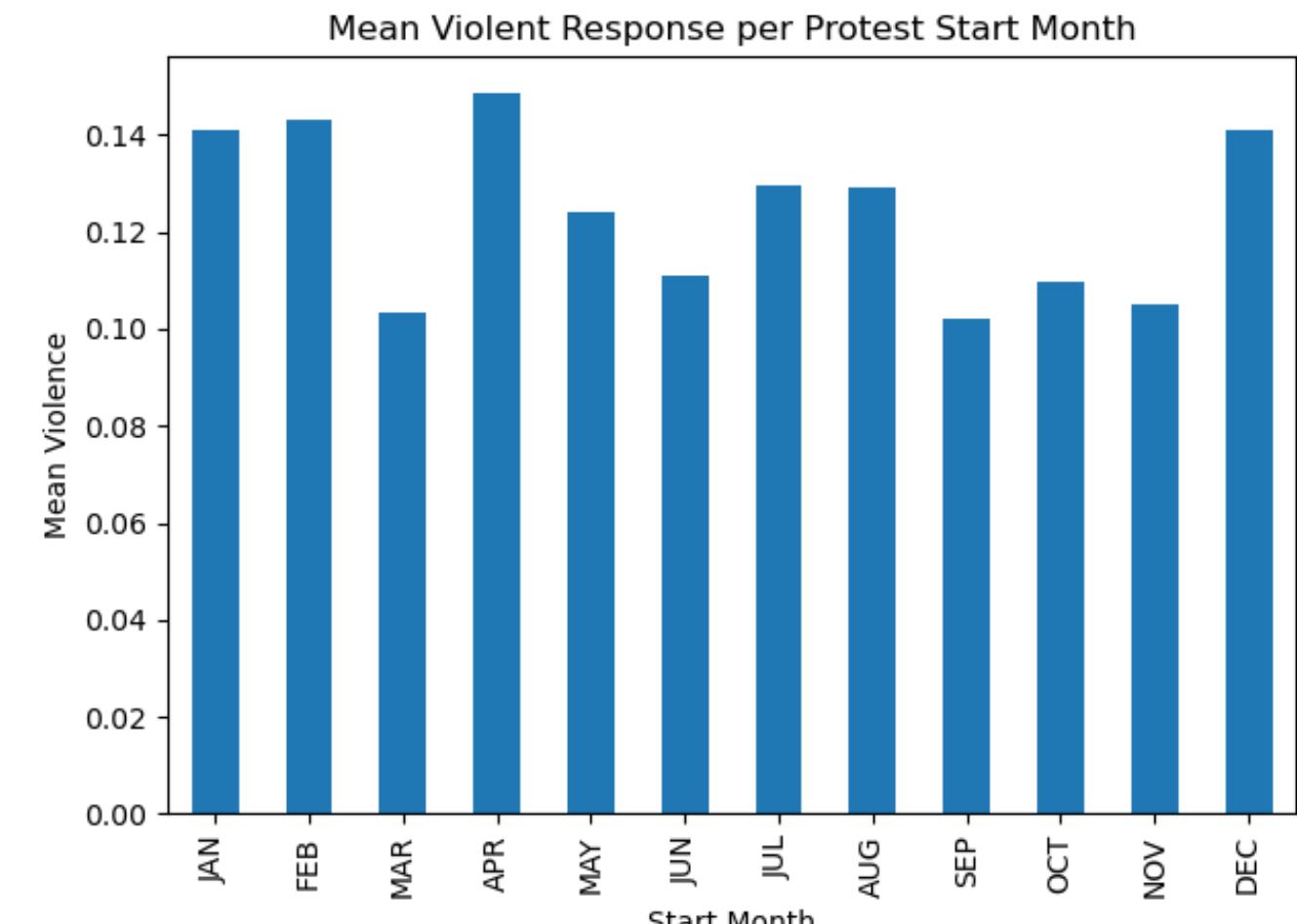
There are large shifts that can be explained by the context of ongoing world events, such as the demonstrations against the Soviet Union that preceded, and ultimately caused, its demise; the global opposition to the invasion of Iraq by the United States; and the collective activism that brought upon the Arab Spring.



Mean Violent Response by Start Month and Season

START MONTH AND SEASON ARE STATISTICALLY SIGNIFICANTLY ($P < 0.001$) DIFFERENT FROM THE MEAN VIOLENT RESPONSE OVERALL.

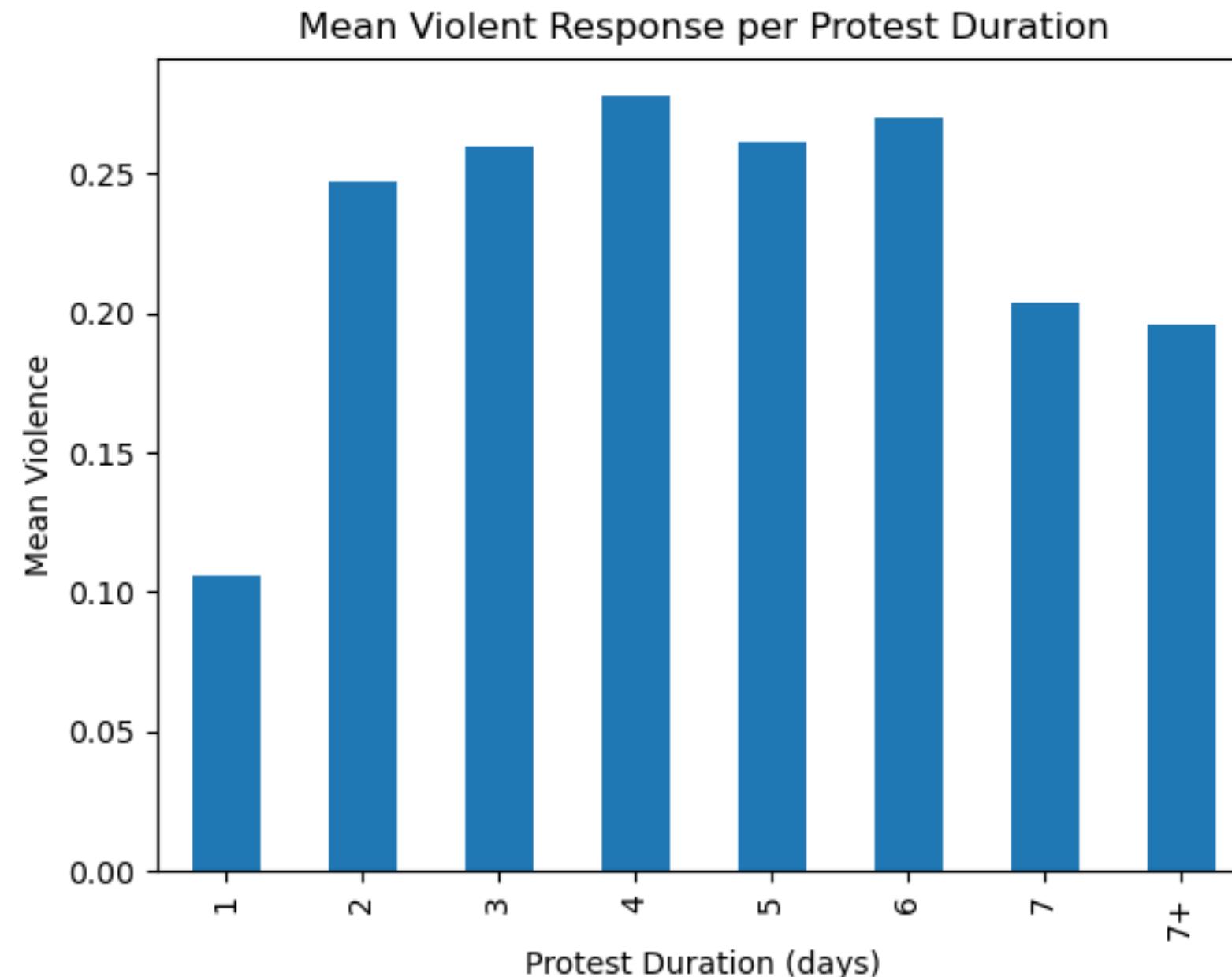
This shows a slight increase in violence in both the summer and winter, presumably the harshest seasons in terms of weather for both hemispheres, but also in line with known data that shows protest decline during the rainy seasons



Mean Violent Response by Protest Duration

PROTEST DURATION IS STATISTICALLY SIGNIFICANTLY ($P < 0.001$) DIFFERENT FROM THE MEAN VIOLENT RESPONSE OVERALL.

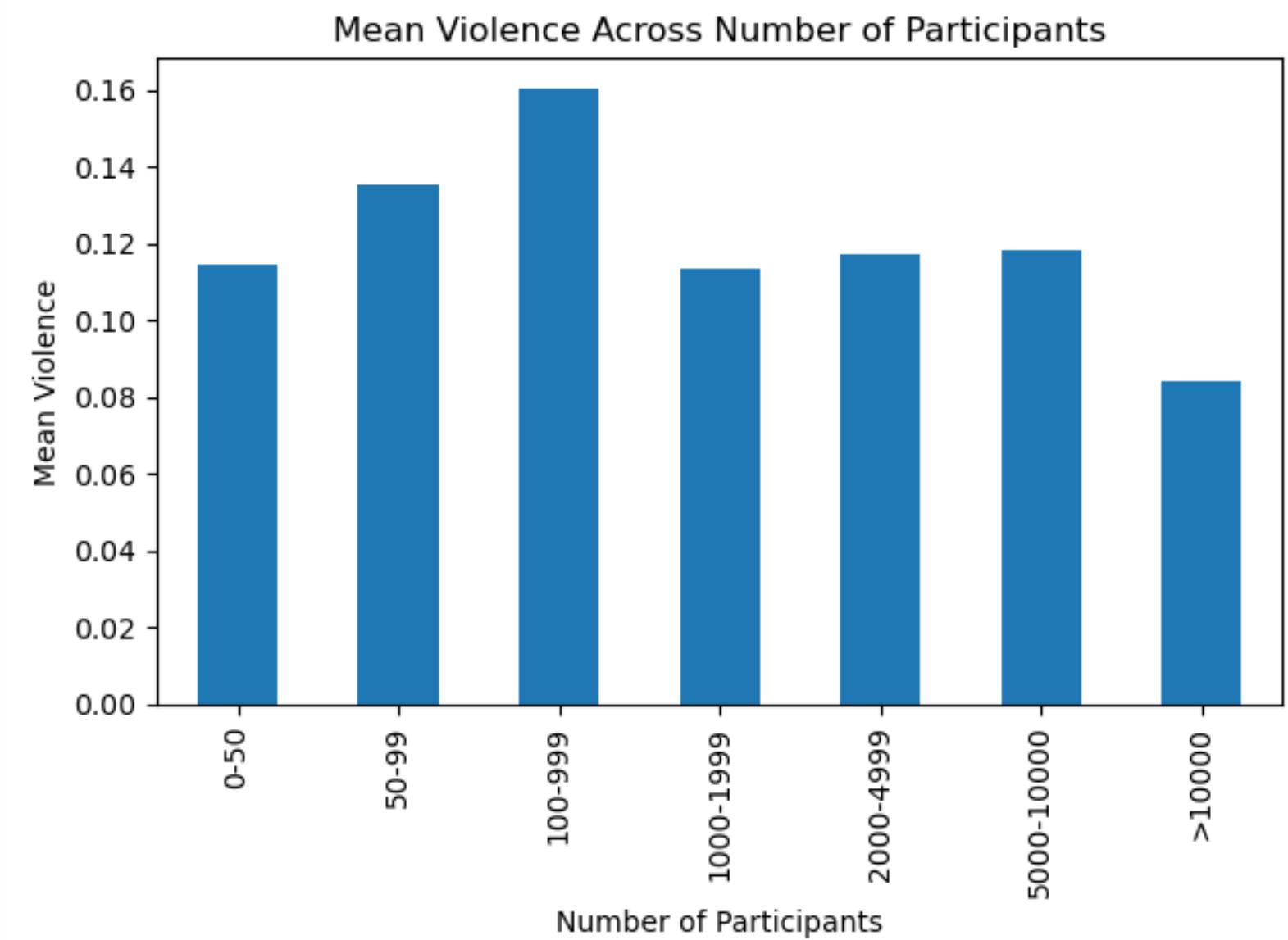
An overwhelming number of protests were only one day.
However, if a protest ran across multiple days, the chance of violence increased



Mean Violent Response by Number of Participants

NUMBER OF PARTICIPANTS IS STATISTICALLY SIGNIFICANTLY ($P < 0.001$) DIFFERENT FROM THE MEAN VIOLENT RESPONSE OVERALL.

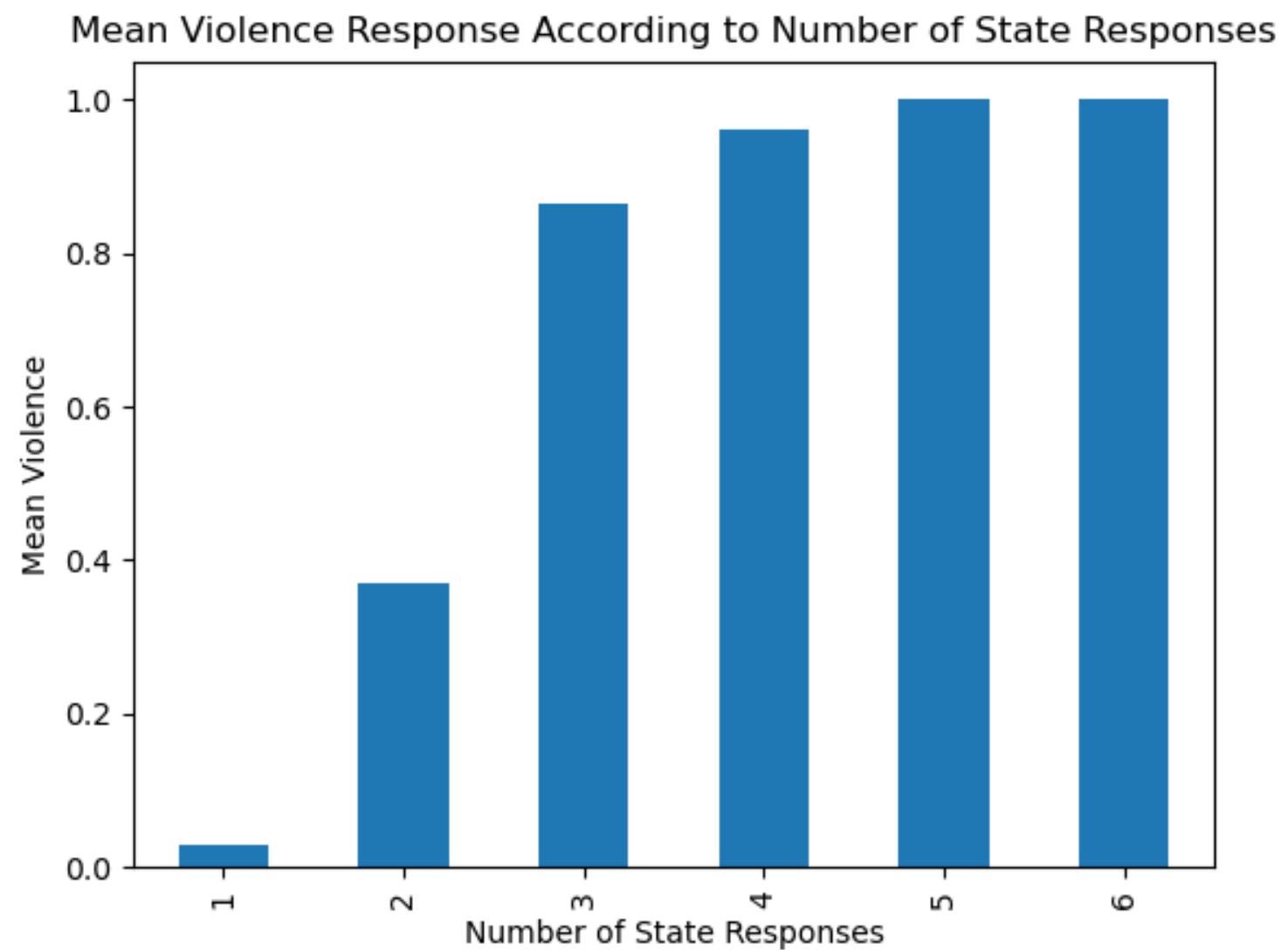
The majority of protests had between 1000 and 2000 participants. Curiously the largest protests are the ones that average the least violent responses, this could be because a larger mass of people might be more intimidating to the state, it could also mean more news coverage and a state's hesitance to react under the circumstances.

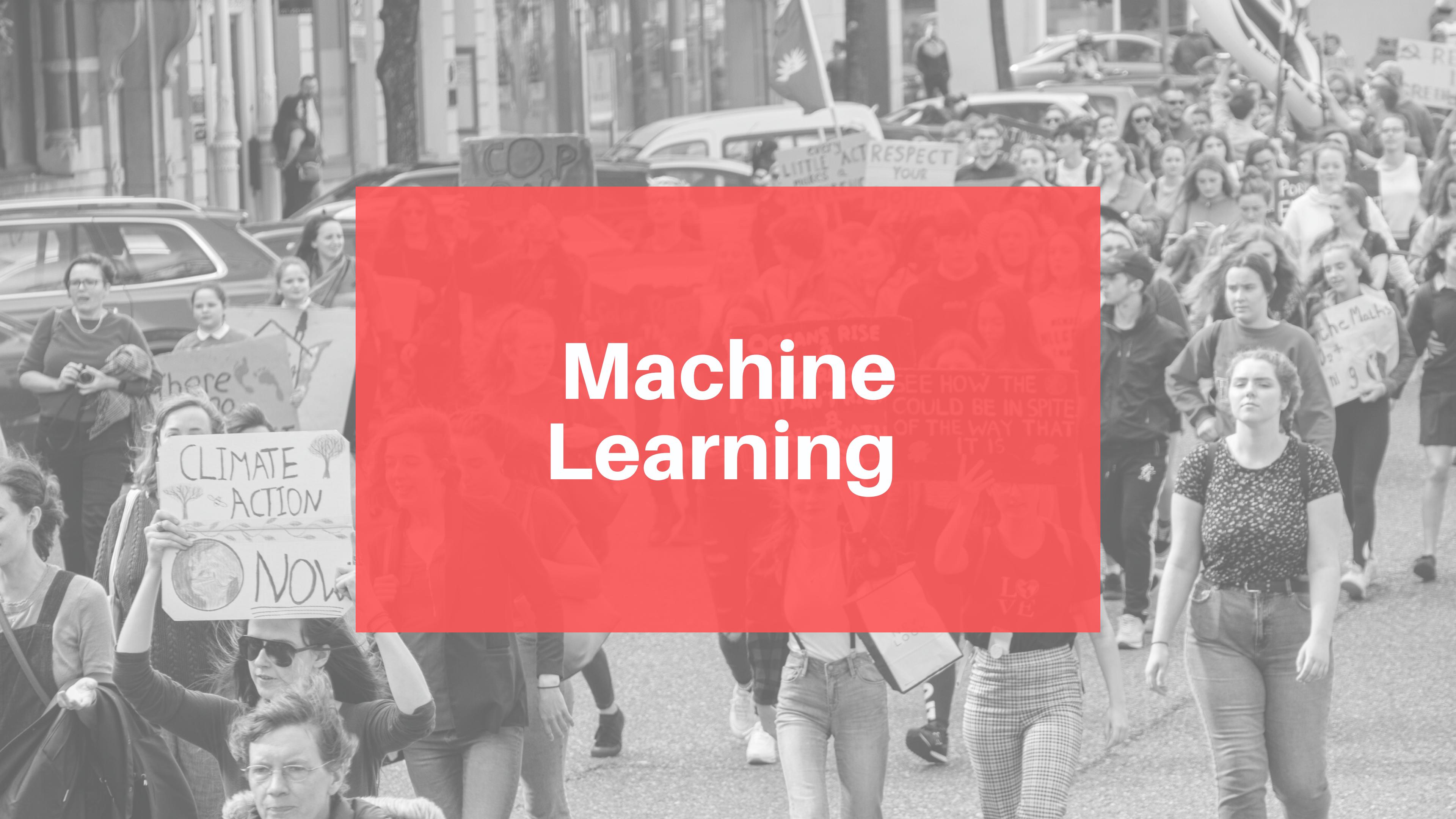


Mean Violent Response by Number of State Responses

NUMBER OF STATE RESPONSES IS STATISTICALLY SIGNIFICANTLY ($P < 0.001$) DIFFERENT FROM THE MEAN VIOLENT RESPONSE OVERALL.

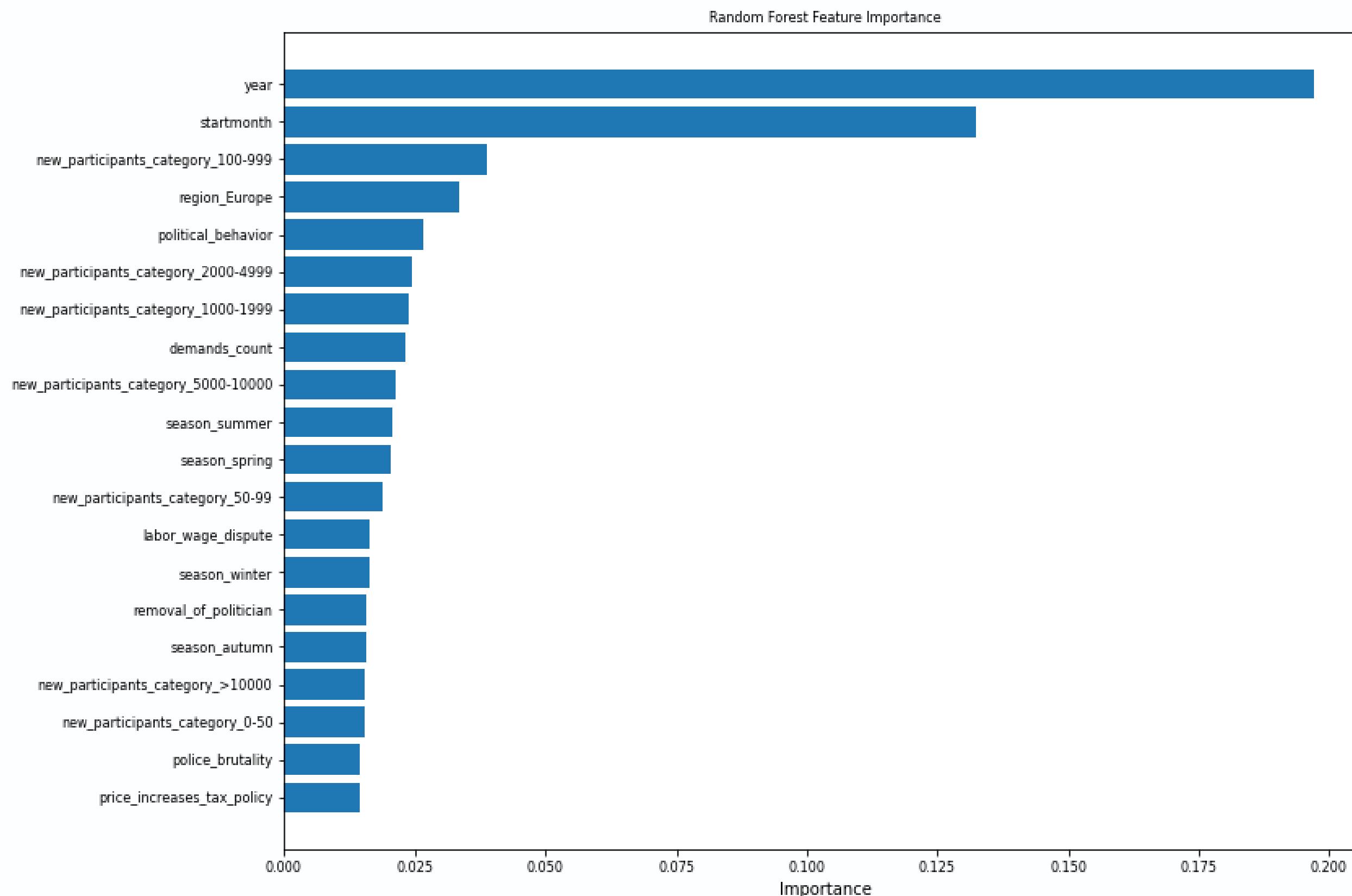
An overwhelming number of protests were only one day. However, if a protest ran across multiple days, the chance of violence increased





Machine Learning

Feature Importance



MODEL SELECTION

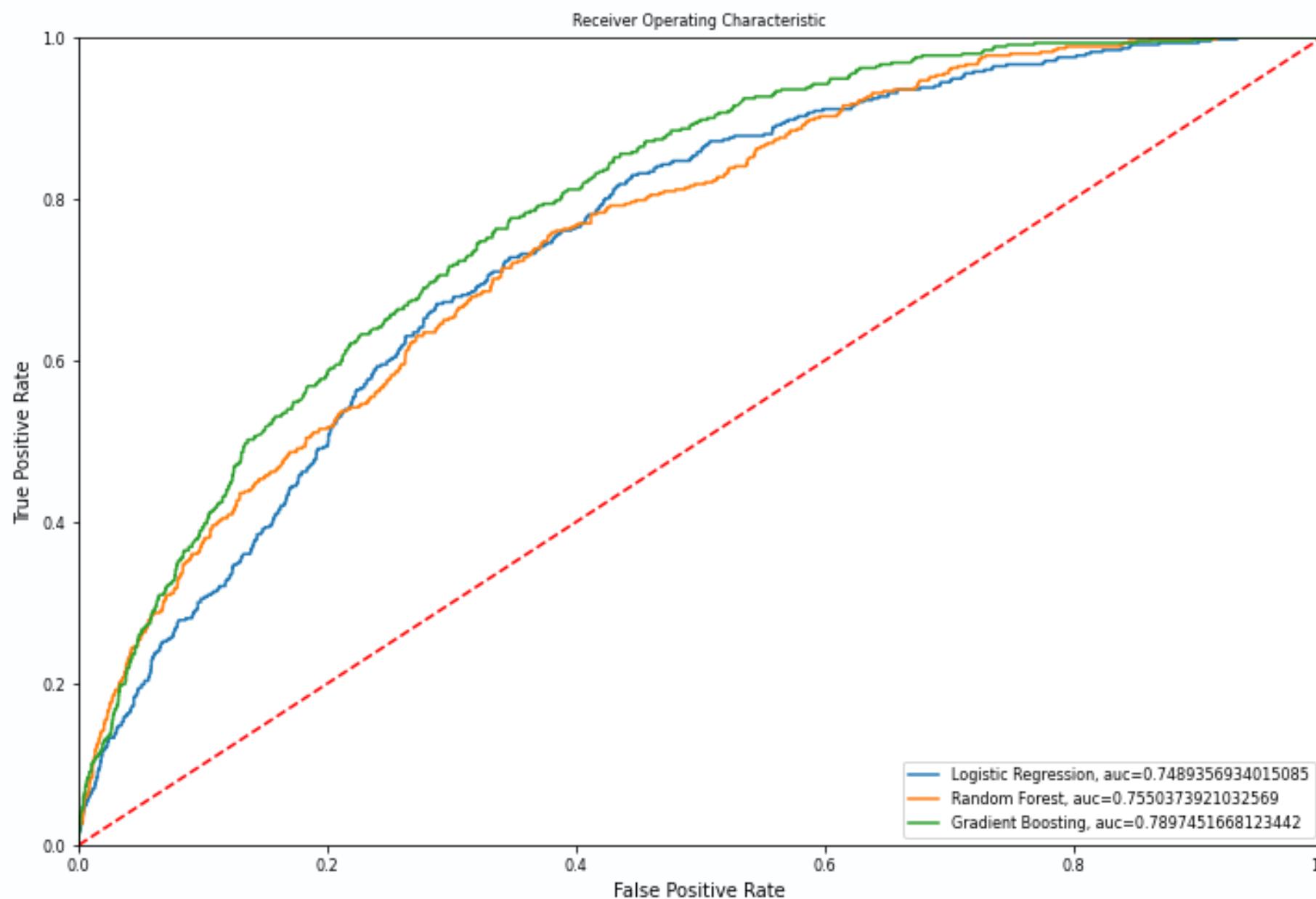
Three different models were created in order to try to predict state violence with the first day knowledge. GridSearchCV was used to find the best parameters and to test their ability to discern a violent response.

Classifier Performance

Classifier	ROC-AUC Score	Best Hyperparameter Values
Logistic Regression	0.7478675828490925	{'C': 0.5}
Random Forest	0.7671881738699942	{'max_depth': 7, 'n_estimators': 200}
Gradient Boosting	0.7911945848734335	{'n_estimators': 200, 'max_depth': 4, 'learning_rate': 0.1}



Model Selection & Threasholding





Model Selection & Threshholding

Based on the ROC curve above, we can see that Gradient Boosting performs best in every situation without the need to sacrifice our TRP nor FPR.

How might this be helpful?

This model may be used at the outset of a protest to understand the likelihood of a violent state response. Knowing this, the protestors could respond in a couple of ways.

- Dispersion:
- Protection/Preparation

List of References

ONLINE SOURCES

Mass Mobilization Data project protests against governments, all countries, 1990-2020. Visit the project page at <http://www.binghamton.edu/massmobilization/> (2019-02-07)

A. Murat Agdemir. "The Arab Spring and Israel's Relations with Egypt. Israel Council of Foreign Affairs, 2016. Vol.10, No.2, pp. 223-235

why-do-peaceful-protests-turn-violent
by

rights-repression-leaders-respond-violent-peaceful-protests/
by

how-violent-protests-change-politics
by