Mass Mobilization: Predicting state violence

A REPORT BY JESSICA MONTEALVO

INTRODUCTION

Since the dawn of democracy, protests have played an essential role in maintaining the power of the people. Unfortunately, the powers that be endeavor to maintain control over the people, and often do so violently. A report by the Center for Strategic and International Studies (CSIS) Risk and Foresight Group indicates an average annual increase of mass protests worldwide of 11.5 percent from 2009 to 2019 (report). The outcome of a protest can vary between extremes of accommodation to violence with casualties. Violent state reactions are often brought upon by what researchers refer to as "the law of coercive responsiveness"; in essence, when a government perceives a threat to its power, it will react with violent repression in order to suppress and weaken the opposition.

According to the CSIS, the rate of protests is only increasing. Thus, it is important to understand the factors that instigate violence and the effect of these demonstrations. In the 2010 "Arab Spring", several countries in North Africa and the Middle East experienced mass mobilizations calling for change in the face of oppressive regimes and low standard of living that were met with violent responses -- such as shootings, beating, and killings -- from authorities unwilling to cede their power. The death toll from these violent actions is estimated upwards of 61,000 people across 19 countries.

In Egypt specifically, peaceful protests began on January 25, 2011 and were met with a wave of escalating state responses trying to placate protesters that culminated with shootings and killings. 846 civilians died and 6,400 people were injured in the 18 days it took President Hosni Mubarak to step down, which happened on February 11, 2011 after almost 30 years in power. The ability to predict this particular state response could have saved hundreds of people and spared thousands of others the dangers and trauma of such an experience. The Egyptian government's response escalated to use of live ammunition against its own citizens, placing snipers on roofs of the areas surrounding the crowds and ordering vehicles to run over protesters in the streets. Although protestors knew of an increased likelihood of violence as tensions built, they did not have concrete data to inform what countermeasures to employ.

PROBLEM STATEMENT

This project aims to create a model that discerns whether a mass mobilization will be responded to violently or not at its inception, to drive protester strategies and inform decisions when dialogue and engagement are no longer an option.

THE DATA

In the data set explored ("Mass Mobilization Data Project Dataverse"), David H. Clark (Binghamton University) and Patrick M. Regan (University of Notre Dame), recorded individual instances of protests against governments in 162 countries, noting protester demands and government responses, with more than 17,000 observations made across 1990-2020. For each instance, the data set details protestor demands, government responses, the location of the demonstration (e.g., country, region), as well as the identity and number of protesters.

DATA WRANGLING AND CLEANING

The unit of observation for this data set is the protest-country-year, where each demonstration is recorded individually within country and year. There were 26 variables recorded for 17,145 events:

> ['id', 'country', 'ccode', 'year', 'region', 'protest', 'protestnumber','startday', 'startmonth', 'startyear', 'endday', 'endmonth', 'endyear','protesterviolence', 'location', 'participants_category', 'participants', 'protesteridentity', 'protesterdemand1','protesterdemand2','protesterdemand3','protesterdemand4', 'stateresponse1', 'stateresponse2','stateresponse3', 'stateresponse4','stateresponse5', 'stateresponse6', 'stateresponse7', 'sources','notes']

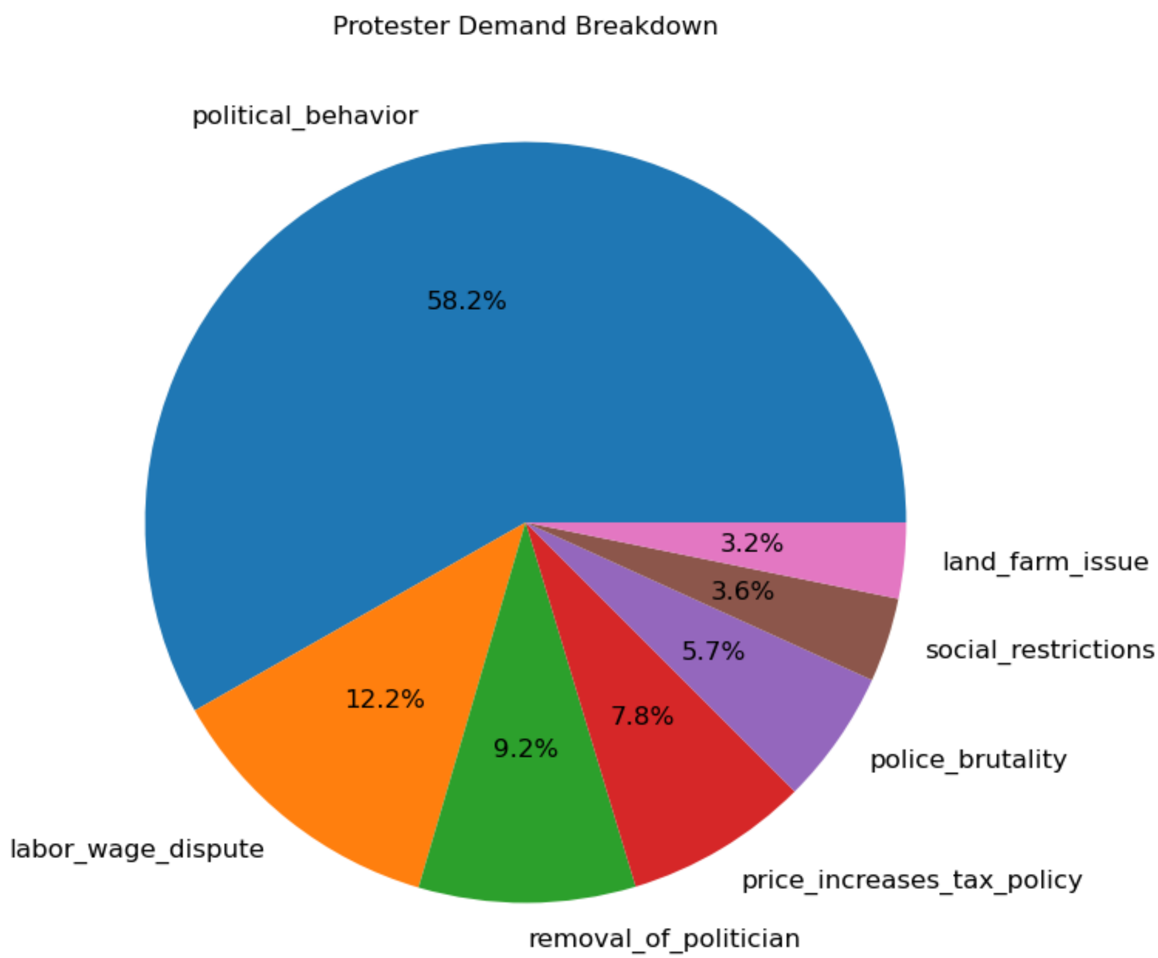
In addition to cleaning each column for data type, whitespace, and placeholder characters, new columns were created to denote the presence or absence (0/1) of each protester demand (n=4; **protesterdemand1** through **protesterdemand4**) and state response (n=7; **stateresponse1** through **stateresponse7**). I created new columns (**response_combo**, **demand_combo**) that combined demands and responses, respectively, in addition to columns (**demands_count**, **responses_count, first_response**, **last_response**) with counts of each, first, and last response to explore for significance.

The authors of the data set added a column, **participants_category**, part way through their original collection efforts. Since this column, along with the number of **participants**, provides valuable insight into a state's violent response, I made an effort to convert every observation to an integer in the ***participants*** field and to complete the **participants_category** column. In addition, since it's not possible to derive participants_category from a mix of data types, I standardized the **participants** column. I prioritized the use of **participants_category** overall, choosing the mid range of the category when available, i.e, if **participants** was noted as *100s*, and **participants_category** is *100-999*, the **participants_int** was recorded as *450*. Otherwise, I selected the lowest notated number, i.e, if **participants** was noted as *100s* with no **participant_category** data, participant_int was recorded as *100*.

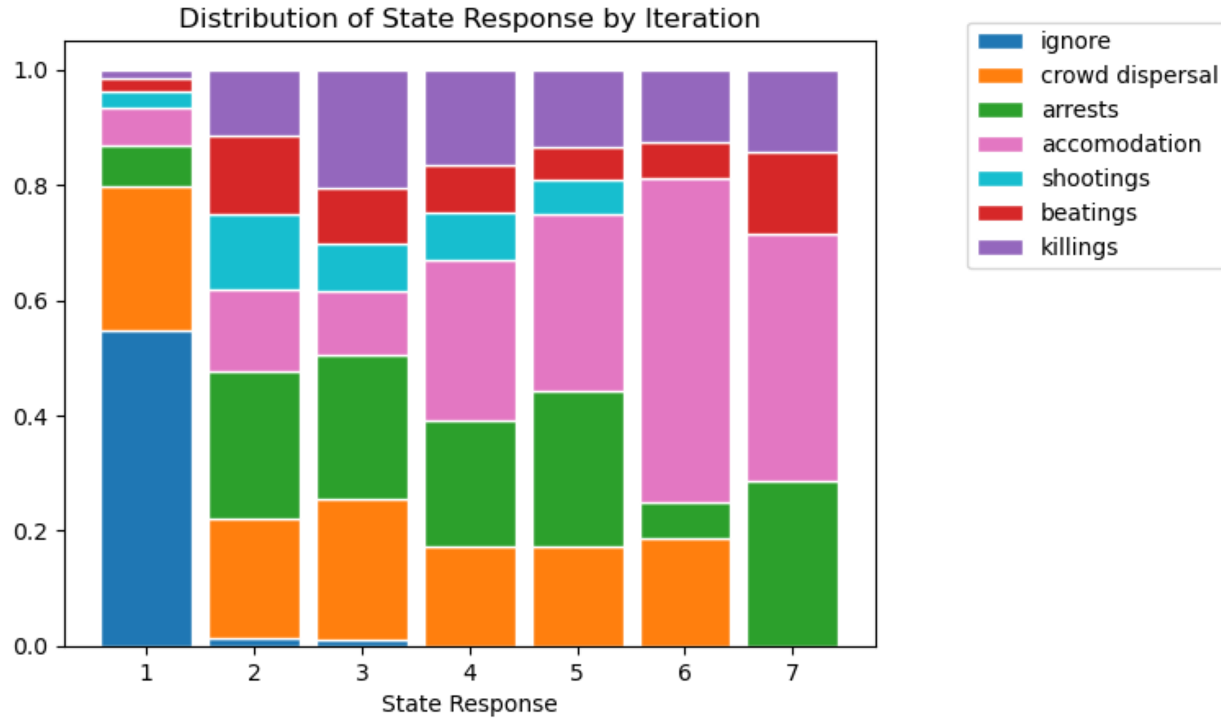
I dropped any observations missing a state response since it was the target variable. The cleaned data set became 53 variables and 14,482 events.

EXPLORATION DATA ANALYSIS

Potential protest motivations were categorized as seven types of issues that inspired protest behavior.Up to four were noted for each observation, as most, if not all, protests don't address more than one grievance at a time.



58.2% of protests were due to political behavior. This is consistent with the researchers' expectations, since the scope of this category is vast, including who rules and how, who can participate in elections, as well as choices made by leaders with both domestic and international impacts.

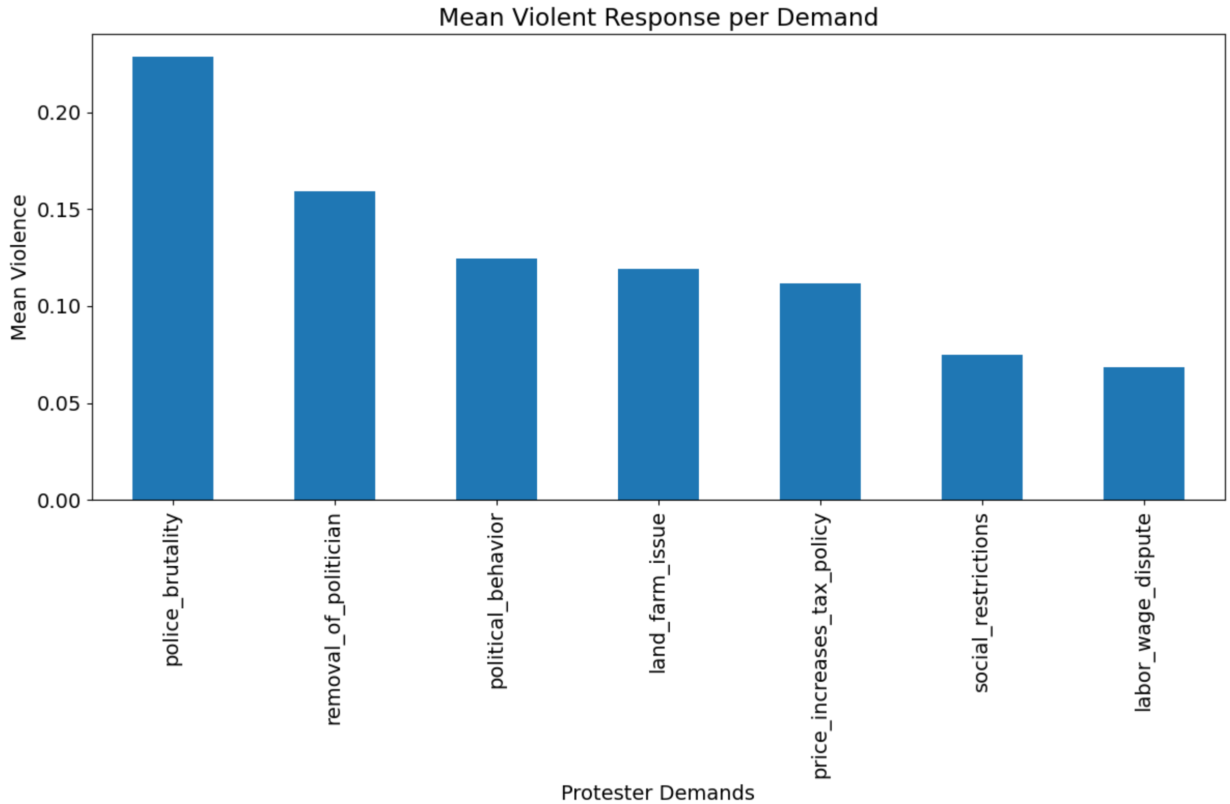


Every state response was categorized into seven classes, where up to seven responses were noted per protest.
The Distribution of State Response by Iteration stacked graph shows the following trends:

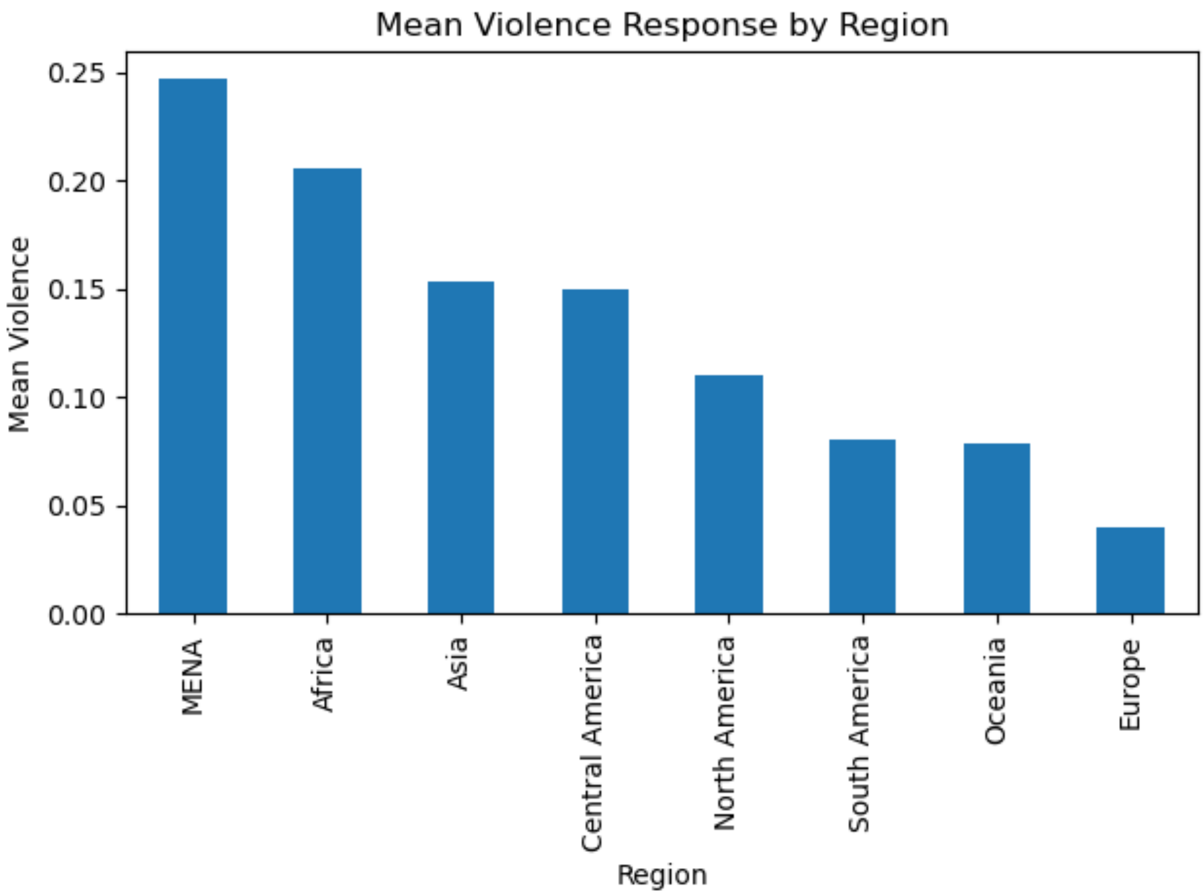
1) 50% ignore at first, drops to nothing after
2) Crowd dispersal remains relatively consistent until the last state response
3) Violence largely absent on first response, but jumps to 40% on subsequent responses, remaining relatively consistent
4) Accommodation seems to be the biggest, growing to represent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

The data reveals the most recorded demonstrations in Europe, followed by Africa and Asia. A majority of the protests were one day long, ranging from 50 to over 10,000 participants. Most protests had roughly 1,000 protesters.

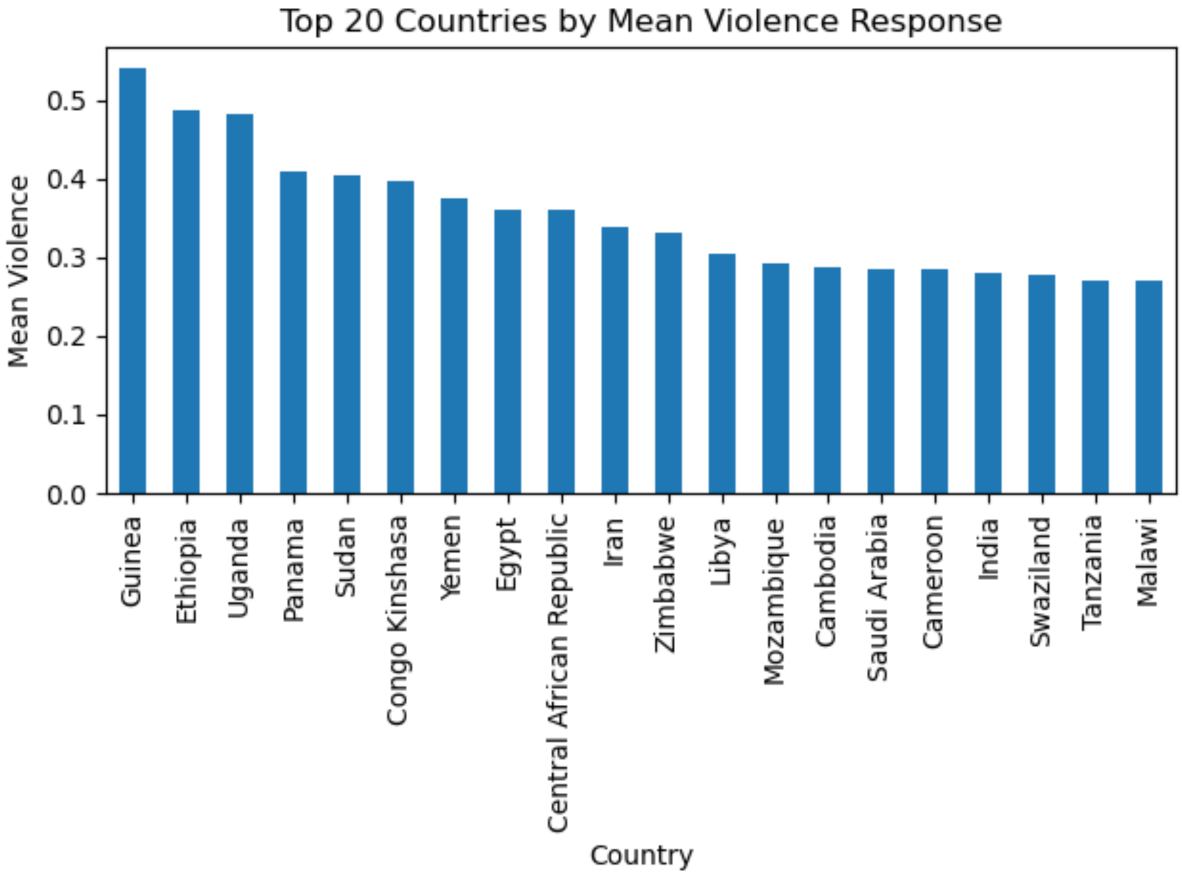
The picture the data paints is very different when seen through the lens of a violent state response.



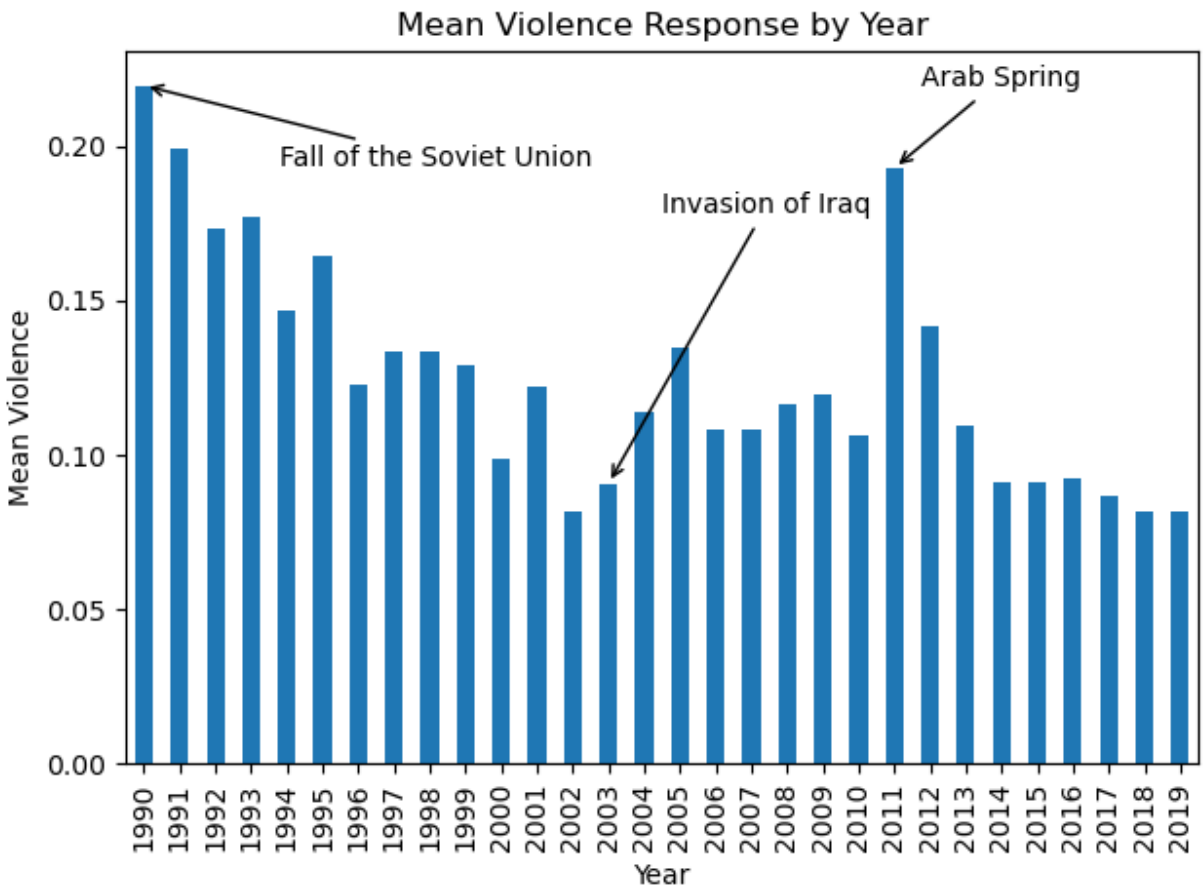
While political behavior was the most common reason to demonstrate, it was mass mobilizations against police brutality that elicited the most violence from states. Sadly, this is consistent with expectations of protesters fighting against an oppressive force. The demand that resulted in the second highest violence mean was removal of a politician. Unfortunately, this combination of demand-response is compatible with the reaction of a leader whose power is being threatened, who seeks to suppress anyone who opposes him. 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.
Police brutality, removal of politicians, labor wage dispute, and social restrictions were all statistically significantly ($p < 0.001$) different from the mean violent response overall.



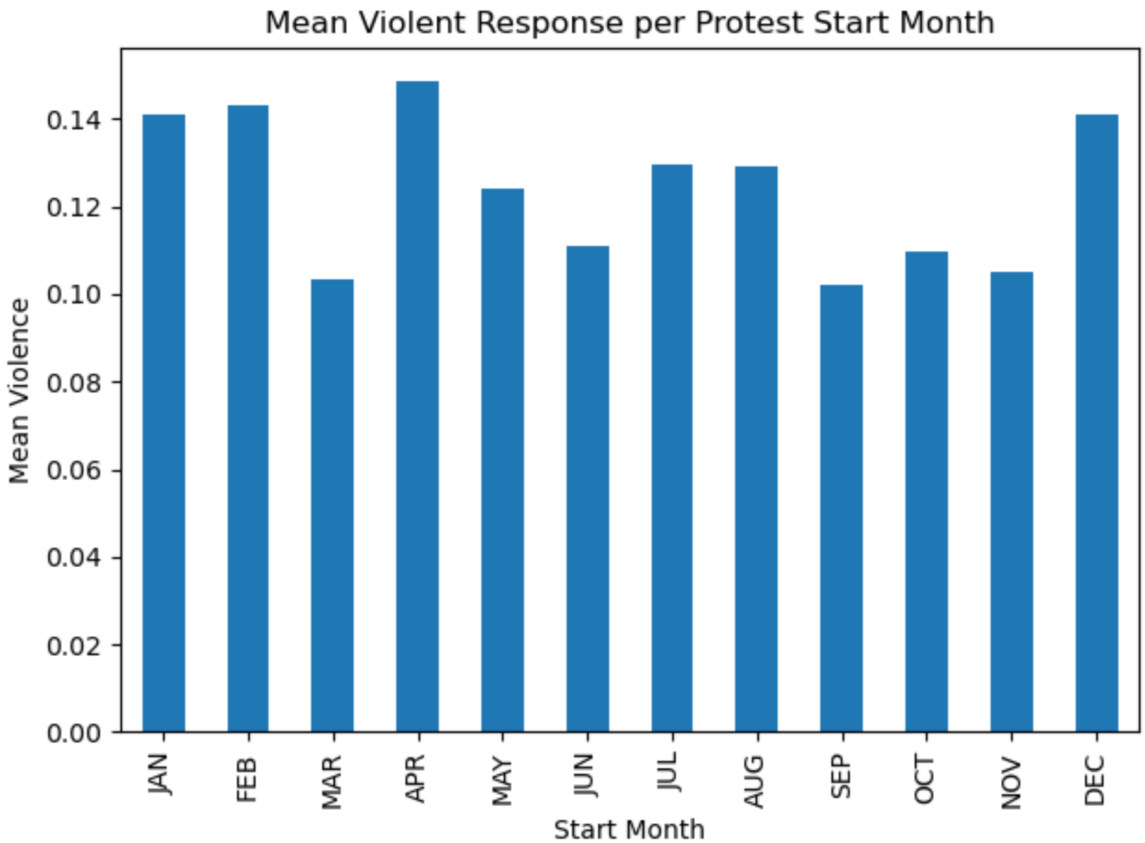
Even though most of the recorded demonstrations happened in Europe (n=4808), it is the Middle East (MENA, n=1093) that on average reacted with violence to mass mobilizations most, followed by Africa (n=3037) and Asia (n=2977). This could be due to the overwhelming presence of capitalist, industrial, wealthy, and developed countries within Europe, compared to the rest of the noted regions - though this is speculation. Another element to consider is the type of government in a specific country. There is an assumption for authoritarian and totalitarian regimes to respond more violently. This was observed from long-standing reigns, like that of Hosni Mubarak in Egypt (30 years) and Muammar Gaddafi in Libya (42 years), where large numbers of protester casualties were observed in the early 2010s when they were overthrown.



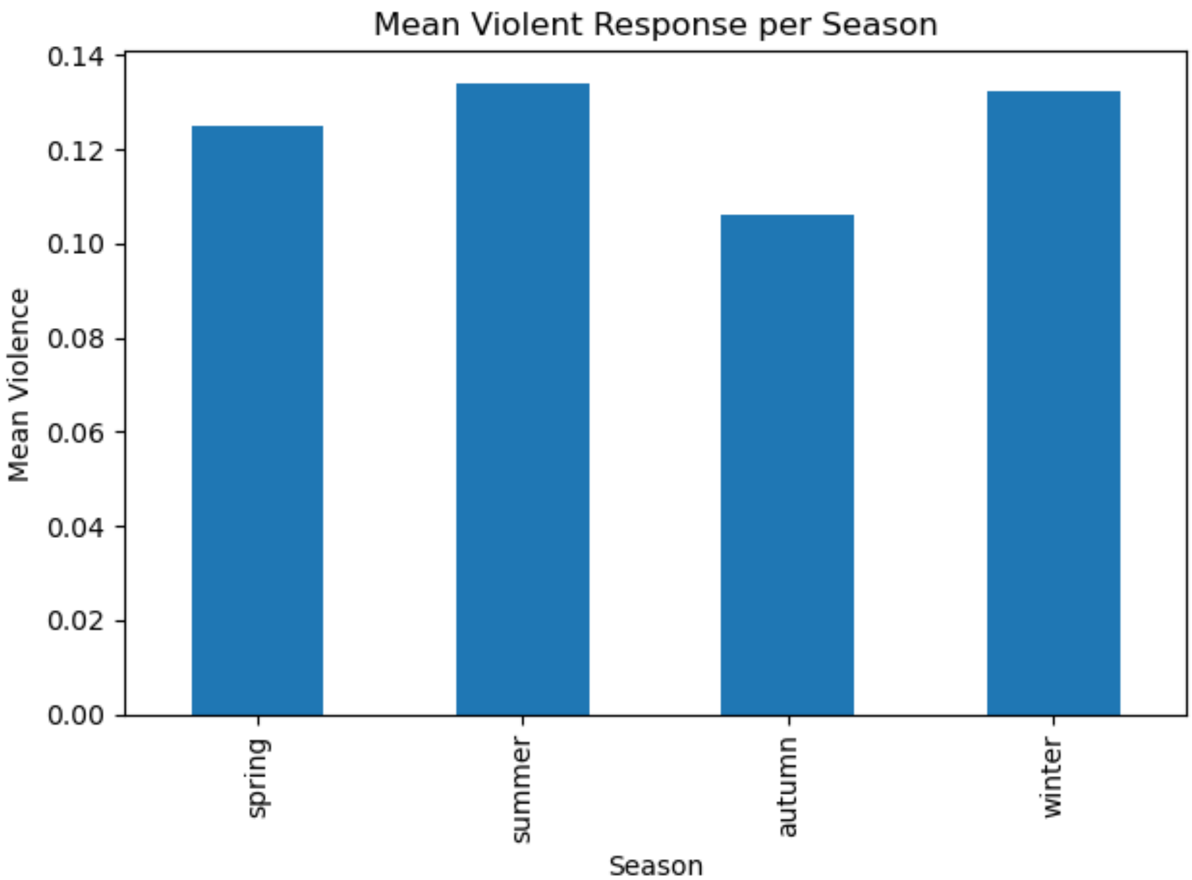
Looking at the breakdown of mean violence per country, the aforementioned influences are more confidently claimed with a $P < 0.001$. Guinea's Second Republic President, Lansana Conte, started as a military dictator, but sought to promote the protection of human rights and the improvement of the economy. After his death, Guinea's political instability trended towards state violence. A military junta seized control in 2009 instituting Guinea's Third Republic, which killed 157 civilians protesting their rise. The government of Guinea continues to respond violently to protests, as seen in demonstrations protesting political behavior concerning election fraud and transparency concerns in 2013 and again in 2020.



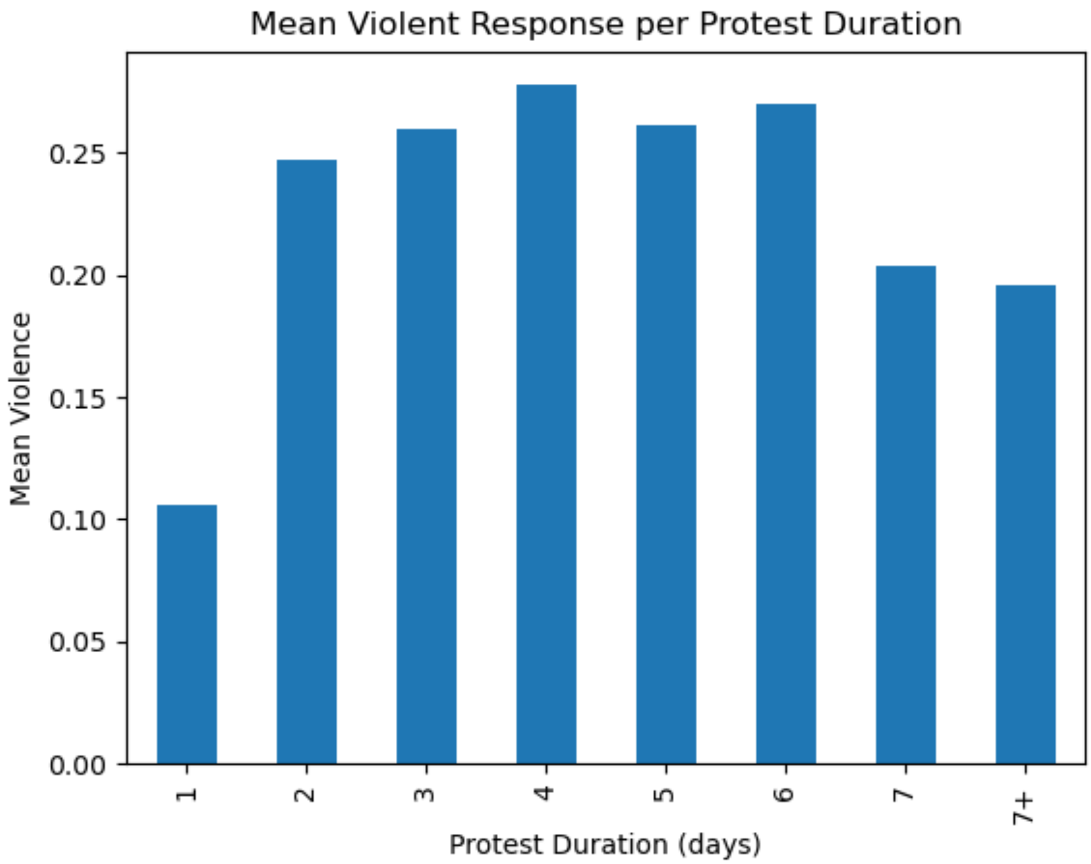
The year and date of each observation were explored for potential trends or influence of a violent state response. There are large shifts (noted in the graph above) that can be explained by the context of ongoing world events, such as the demonstrations against the Soviet Union that preceded, and ultimately caused, its demise; the global opposition to the invasion of Iraq by the United States (from 2002-2005); and the collective activism that brought upon the Arab Spring. It would be interesting to overlay a larger set of world events to see if the observed trends can explain our data. Year showed statistical significance with $p < 0.001$.



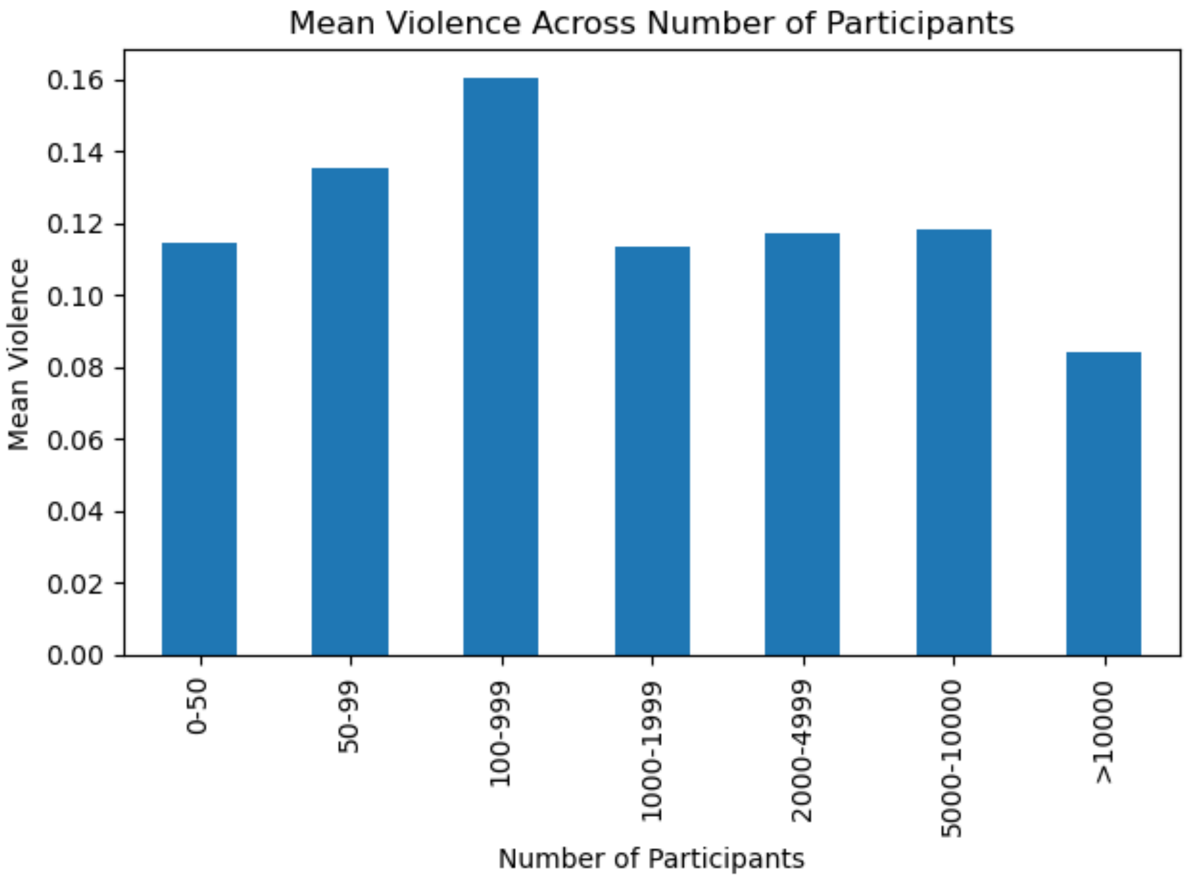
Another aspect noted by the researchers was the time of year that the demonstration took place. Depending on the hemisphere where a certain country is located it meant different weather and overall conditions for the participants and the states alike. Multiple potential behaviors could be expected, for example a decrease in demonstrations in the colder months, or protests during election months.



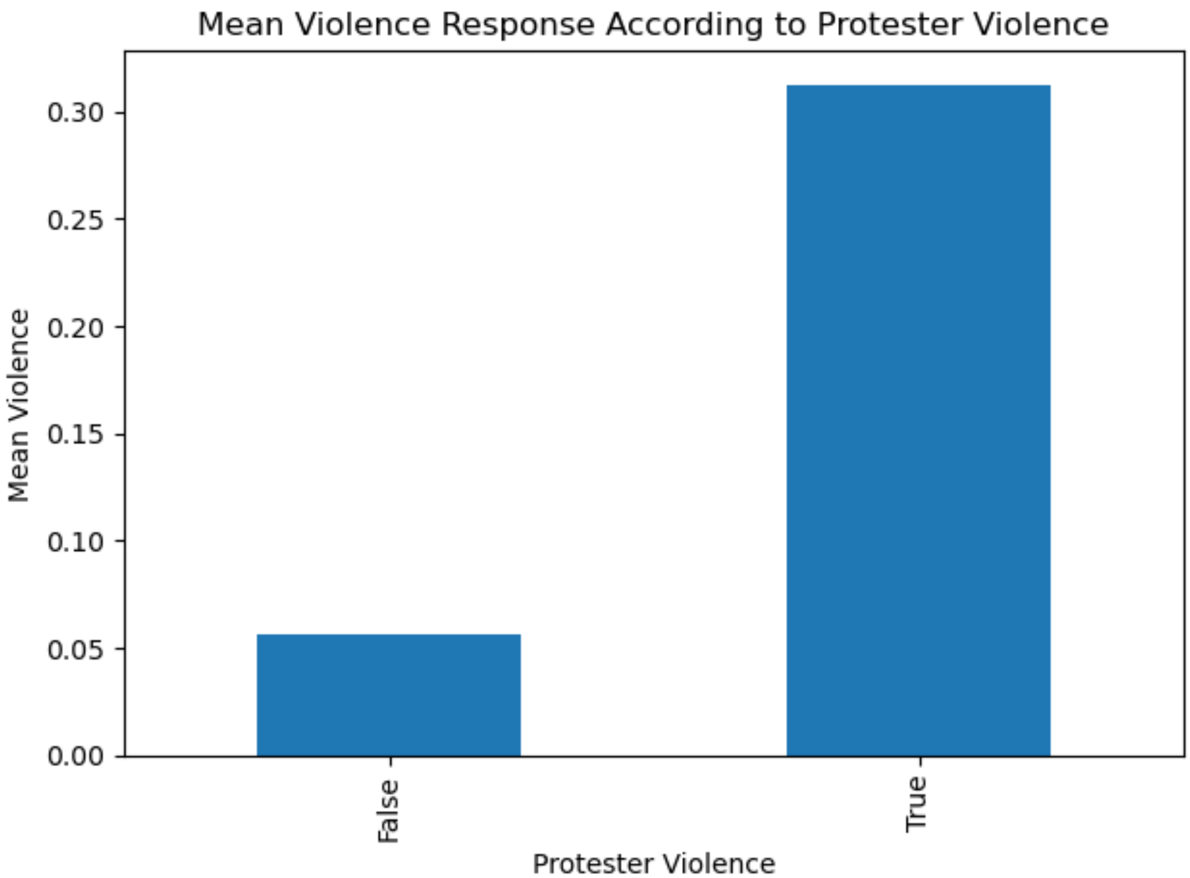
In order to be able to see how violence was distributed across different seasons, I filtered each country by its location in the southern or northern hemisphere. Then, I mapped the start date to a season. This shows a slight increase in violence in both the summer and winter, presumably the harshest seasons in terms of weather for both hemispheres, but also in line with known data that shows protest decline during the rainy seasons ($p < 0.001$).



The length of a demonstration was investigated in this data set. An overwhelming number of protests were only one day. However, if a protest ran across multiple days, the chance of violence increased ($p < 0.001$).

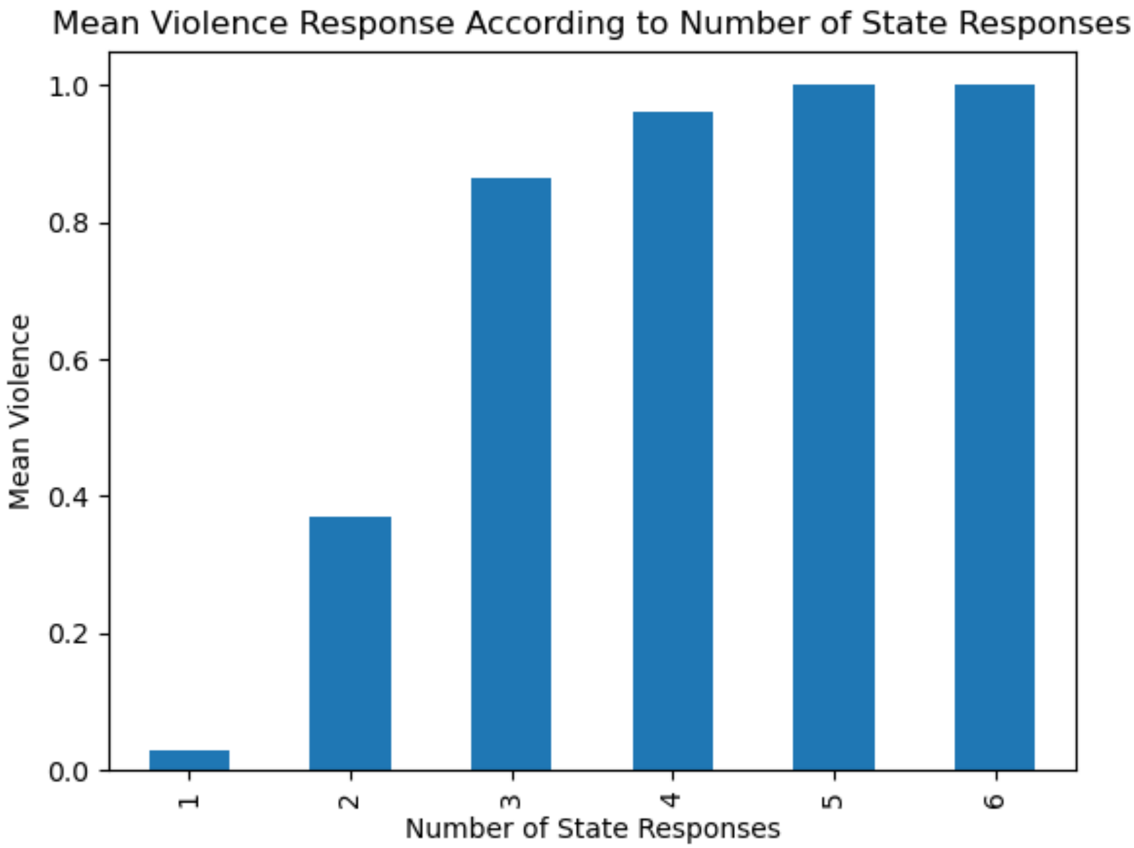


Another characteristic explored was the number of participants and the incidence of violence. As previously mentioned the majority of protests had between 1000 and 2000 participants. Curiously the largest protests are the ones that average the least violent responses, this could be because a larger mass 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.



An additional, pertinent aspect of protests was whether there was protester violence or not. The graph below does show that violence begets violence in the observed cases, there is a statistically significant difference between these two groups ($p < 0.001$).

A note made by J. Pinckney on his assessment of peaceful protests turned violent observes an increase in protester violence when there is a hierarchy and a more centralized protester campaign. Protester violence can especially be seen as a result of repression over a long period of time, especially when they approached the demonstration nonviolently and if they are met with violence again and again. Protester violence can also indicate a split between a moderate and a more radicalized protester base, but this aspect of the data was not explored.



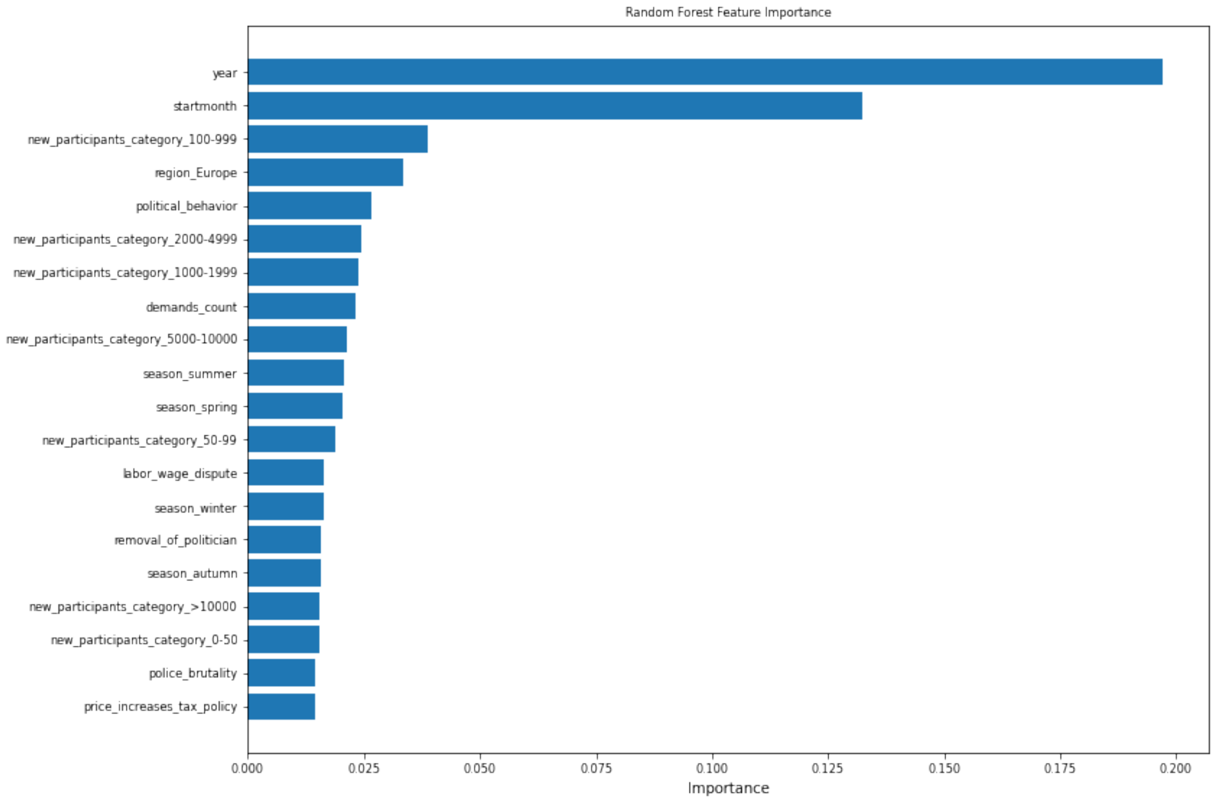
When observing whether a demonstration was brought upon by multiple demands, it seemed that if a group had more than three demands(of note, up to four were noted in this data set), this increased the chance of violence, but this observation was not statistically significant when compared to the mean.

The data show that an increased number of responses almost assuredly resulted in a violent final response. we can start to see this in the distribution of responses. There are no ignore responses after the third iterative response. Though statistically significant ($p < 0.001$), this wasn't a feature that could be used in a real world model given that it is not available from the start.

MACHINE LEARNING

Preprocessing

In order to create a data set that could model real life and be useful, the duration column, response count, and protester violence field were dropped from the data set since these were features that could not be known at the start of a demonstration.



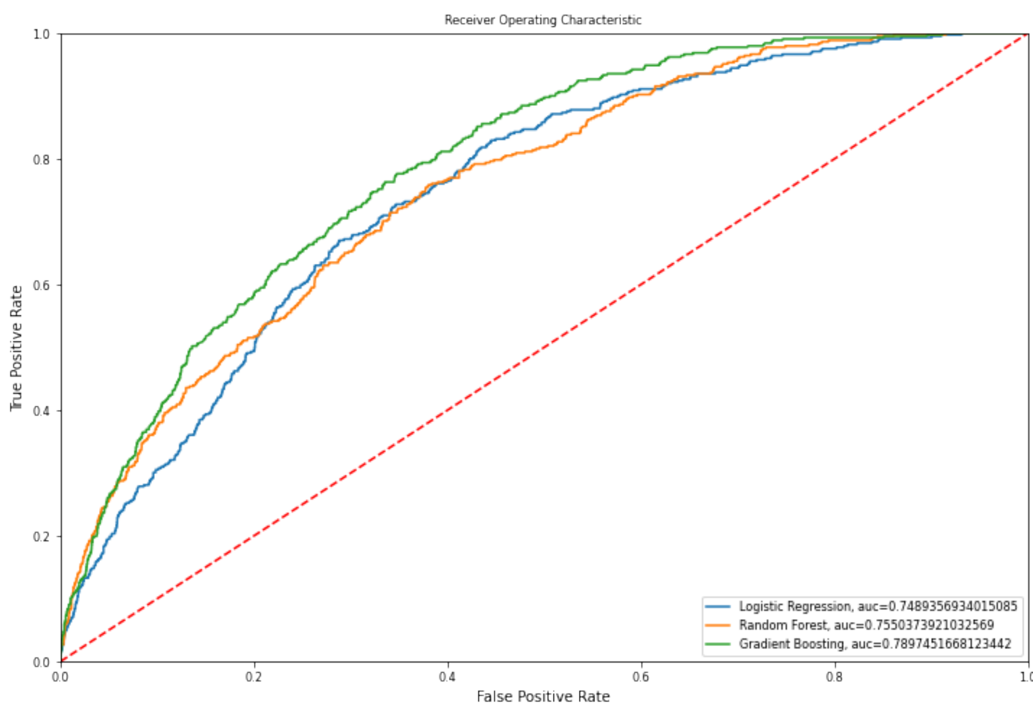
After converting categorical variables into dummy columns, a random forest feature identification analysis was subsequently run. The results coincide with some of the findings from previous data exploration analysis and show that year, start month and and a participant category of one hundred to one thousand protesters are most consequential.

Model Selection and Thresholding

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	0.7479	{'C': 0.5}
Random Forest	0.7672	{'max_depth': 7, 'n_estimators': 200}
Gradient Boosting	0.7912	{'n_estimators': 200, 'max_depth': 4, 'learning_rate': 0.1}



In order to understand which model to choose, we need to understand better how this model will be used. 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.

So how might this model 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 number of ways.

1) Dispersion: in this case, it's likely we'd want to prioritize a low FPR, as we wouldn't want to disperse unnecessarily out of fear, and so might choose the Gradient Boosting Model perhaps at a threshold where our FPR is .2 and TPR. is .6
2) Protection/preparation: if the protestors are willing to face state violence, and the knowledge of it potentially occurring would help them better prepare and allocate resources more effectively we could better withstand a high FPR in order to make sure our TPR is high so we aren't caught unaware. In this case, we might choose the Gradient Boosting Model with a TPR of ~.95 and a FPR of ~.65

CONCLUSIONS

In summary, there are features that are indicative of a violent state response at the beginning of a demonstration, but no true predictor. One major finding of this project is greater duration and frequency of state response to a demonstration increases the likelihood of violent response. Given the findings, citizens can shift their approach, when possible, in order to avoid the losses and perils associated with state violence.

Some variables that increase the likelihood of a state's violent response, such as the year or country/region cannot be modified and remain an opportunity for refining this model. Of note, this data set does not include any demonstrations in the USA, which has experienced an increase in demonstrations, both violent and peaceful with violent endings in the last 10 years. This project also disregards that some crowd dispersal techniques, like the use of rubber bullets or water cannons, may be considered violent reactions by protestors and observers.

Conversely, protesters can leverage these features and make an effort towards maximizing their impact with strategic (peaceful, short) demonstrations. With the advent of new technologies, controlled demonstrations become more visible, bypassing the inherent costs and dangers of a violent outcome. An interesting expansion on this analysis would additionally gather the type of government and information surrounding the economic well-being of a country at the time of each observation in order to better predict -- and thus, avoid -- aggressive responses.

REFERENCES

Citation: 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

https://www.csis.org/analysis/age-mass-protests-understanding-escalating-global-trend
https://politicalviolenceataglance.org/2016/10/25/why-do-peaceful-protests-turn-violent/
https://campaignforsocialscience.org.uk/news/rights-repression-leaders-respond-violent-peaceful-protests/
https://www.newyorker.com/news/q-and-a/how-violent-protests-change-politics