DSA 210 Term Project Report

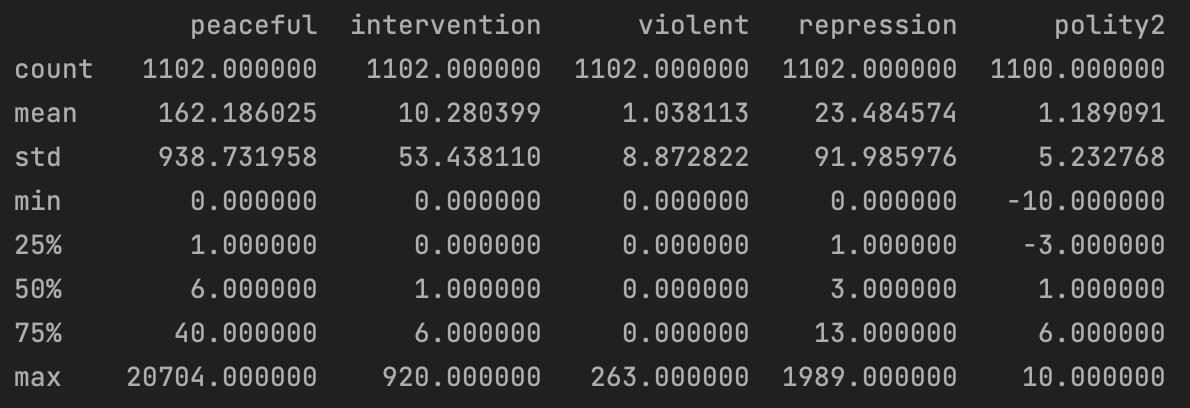
Mert Bilge Tosya - 31170

## Introduction

This project aims to explore the relationship between political unrest and changes in regime authoritarianism. The unrest data comes from ACLED, an organisation that collects data on conflicts and protests. The data is filtered down to only cases of protest (peaceful or violent) and regime repression. It is then aggregated to country-year data to allow comparison with the change in polity values for each country-year. This polity value comes from the Polity Project and signifies authoritarianism, or rather democratisation, as the values range from -10 (perfect autocracy) to +10 (perfect democracy). The two datasets are then merged.

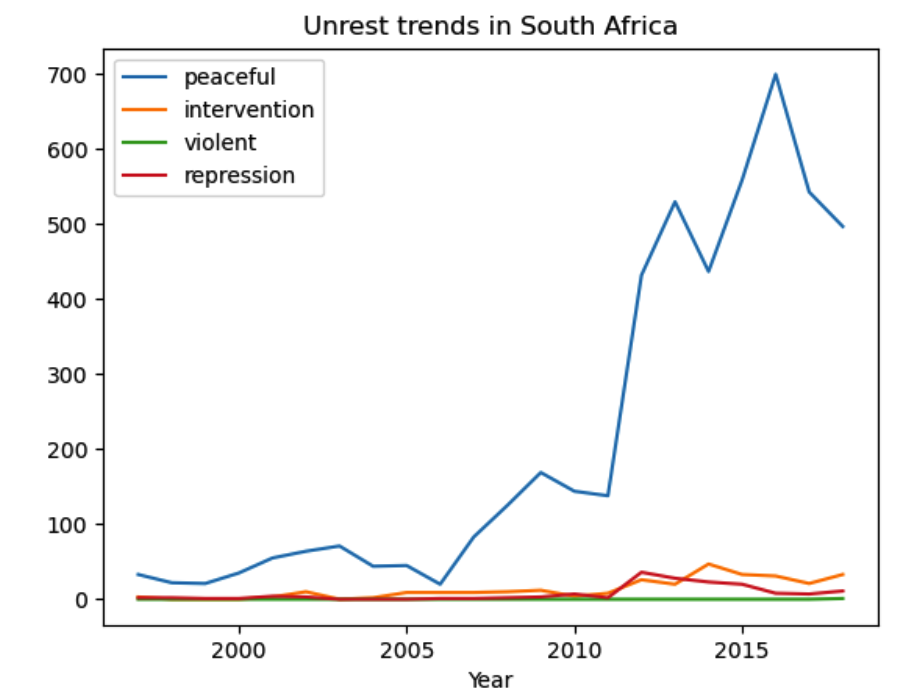
## Exploratory data analysis

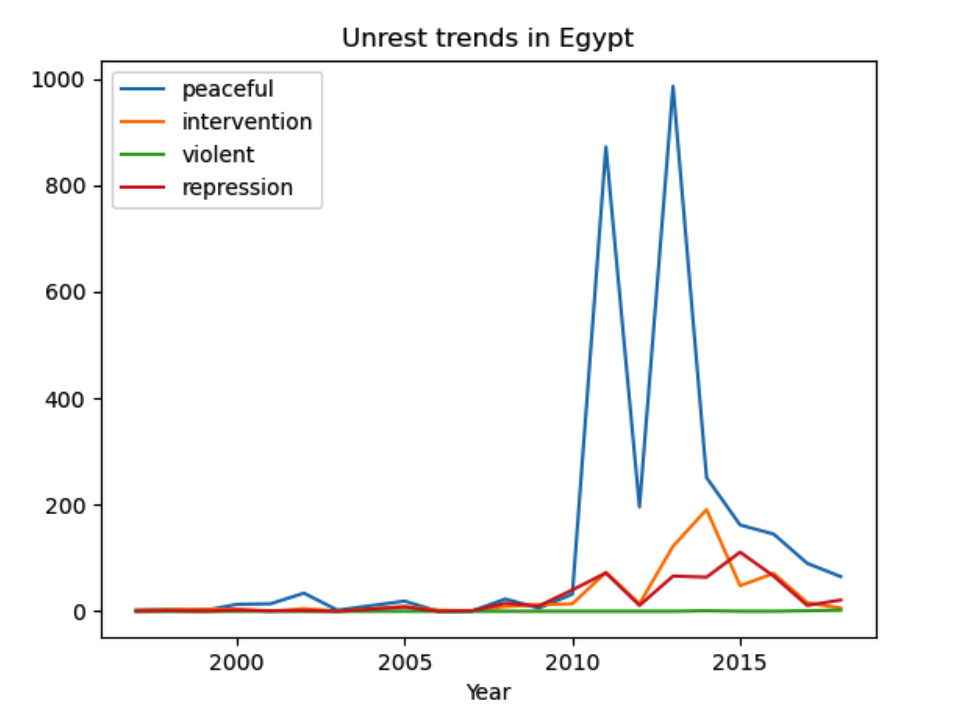
The data (for each country-year) is summarised in the table:

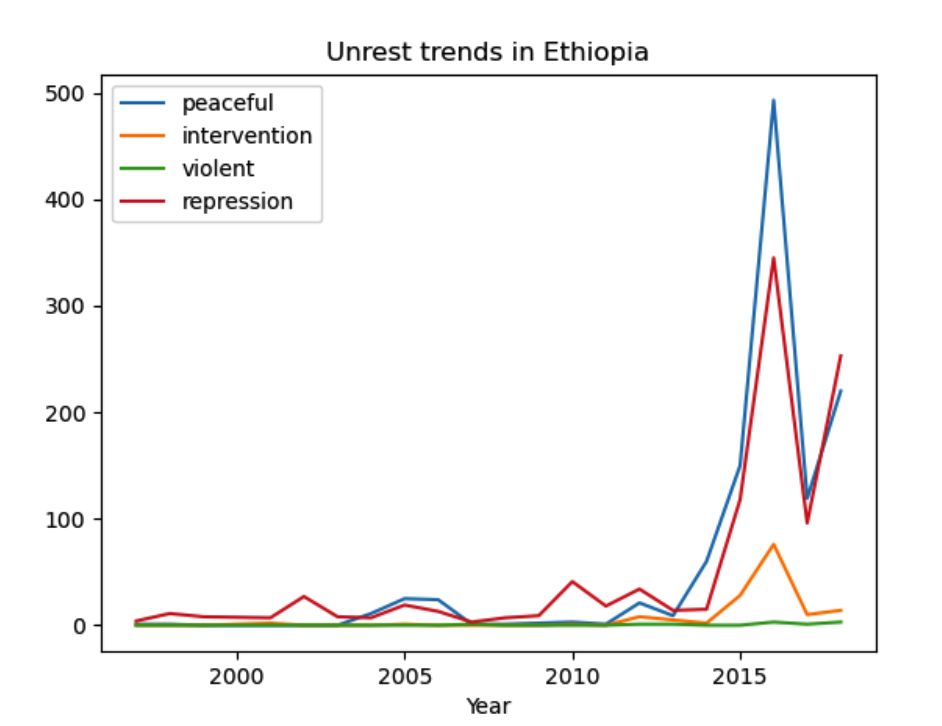


We see that peaceful protests comprise the majority of events of interest, and that specifically for violent protests our sample size is quite small - only one event on average in a country-year. Note that “intervention” comprises of peaceful protests that were violently suppressed by the authorities. These were separated in order to avoid the problems double-counting them could introduce (they are instances of both peaceful protest and state repression).

We present time-series graphs for each of the four incident types for three sample countries so that we may get a glimpse of their trends.

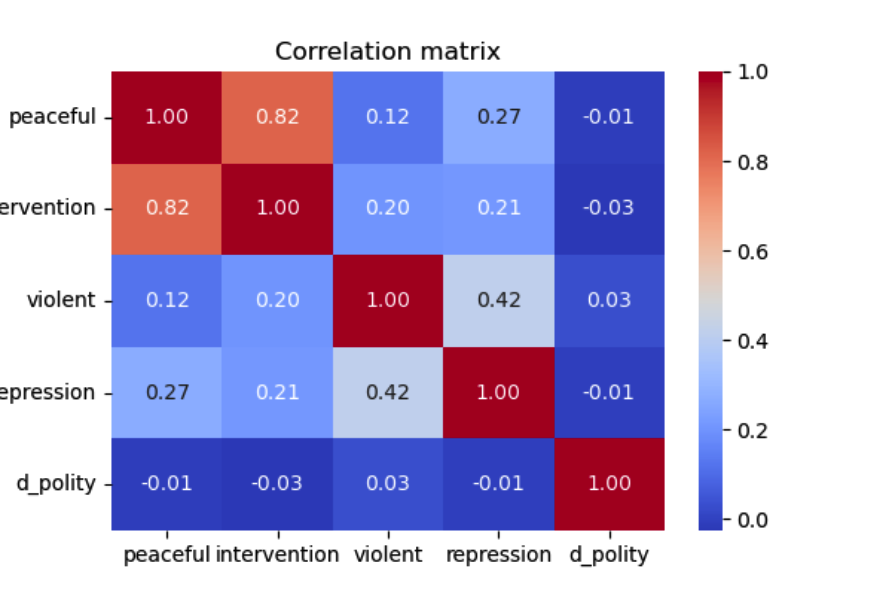






Note that these are all African countries since the overlap between ACLED and Polity data is only a few years for the other regions and they do not make good graphs.

We then create a correlation matrix and see that the correlations are very small. The column of interest here is the rightmost one.



## Hypothesis Testing

We conduct a t-test to compare the mean year-to-year change in Polity score (d\_polity) between ‘high-repression’ years (top 25 % of repression counts) and ‘low-repression’ years.

Null hypothesis*:* the average d\_polity is the same in high vs. low-repression years.

Result: t=0.89, p=0.372

Since p > 0.05, we fail to reject the null hypothesis and conclude there is no statistically significant difference in subsequent regime change associated with high levels of state repression.

We conduct the same t-test for the other variables. The results are:

Intervention: t=-1.57, p=0.116

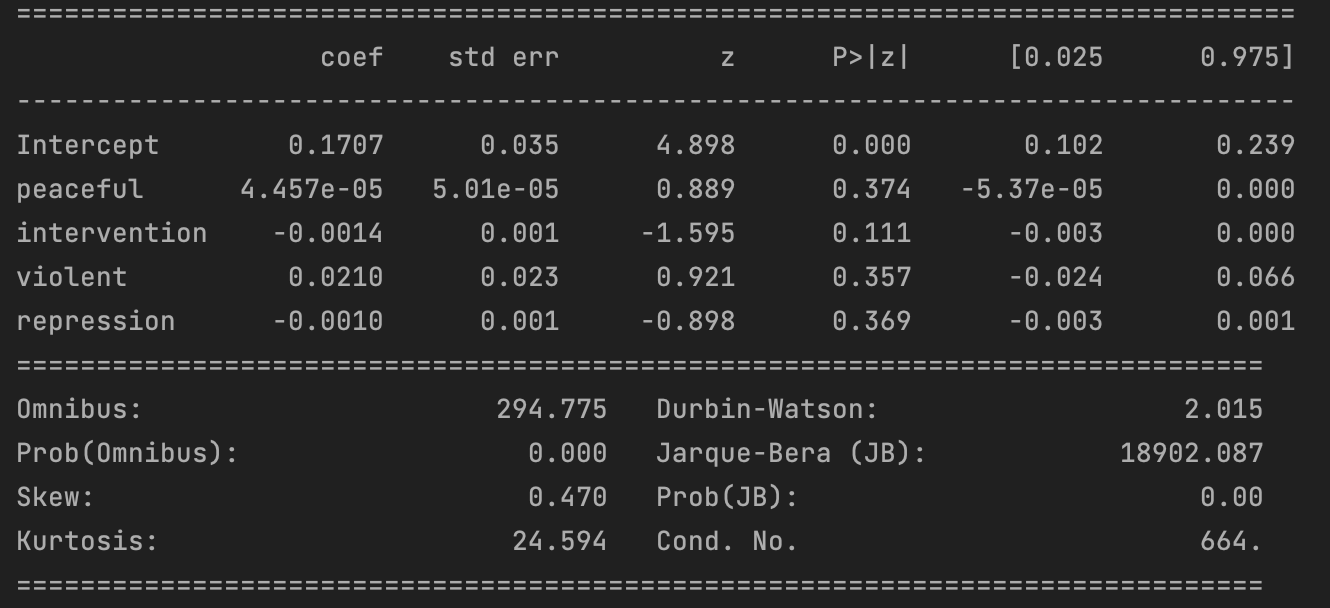
Peaceful: t=-1.59, p=0.112

Violent: t=-0.48, p=0.634

We see that the only somewhat acceptable p-values are for intervention and peaceful, and as their t-statistics are negative, their effect would be autocratisation, if we are to consider them significant.

## Regression Analysis

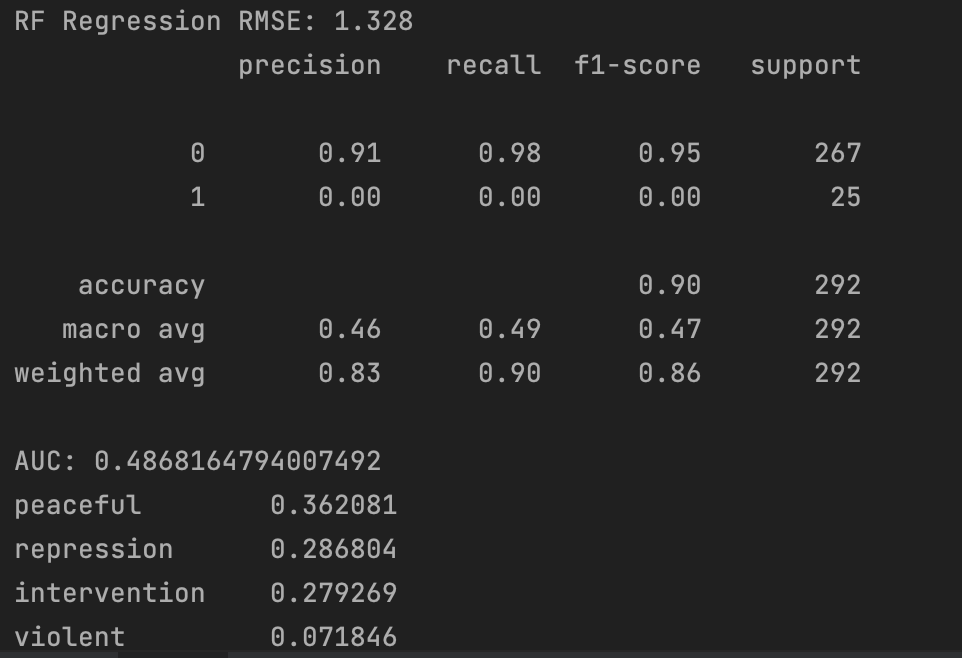
We run an OLS regression of the four event types on d\_polity. The results:



Again we see horrendous p-values, with only intervention being acceptable at a 15% significance level. In addition, R² is very low at 0.003.

## Machine Learning Analysis

We then run a random forest regression, and get the following results:



The RMSE of 1.328 indicates that model is off by about 1.3 points (on a -10 to +10 scale), which is fairly large and erodes our hopes of getting meaningful results, if any have remained until this point. For the following part, we had binarized d\_polity so that 1 indicates positive d\_polity and 0 non-positive. The accuracy of 0.91 for the predictions of 0 is very misleading as about 91% of cases are no-improvement i.e. a model that always predicts 0 would be as accurate as ours. Our model also never correctly predicts the positive cases with an accuracy of 0.00. The AUC of 0.486 confirms our doubts that the model is worthless, as this signifies a worse than random performance.

## Conclusion

We conclude that we cannot conclude anything about changes in short-term regime characteristics from the protests data we have. The only remotely acceptable p-values we find are 0.11 for peaceful protests and those with intervention, and 0.11 again for protests with intervention in the multivariate model. The ML model is completely useless and performs worse than random. One significant shortcoming is that we only consider year-to-year changes in polity score. It may be that the incidence of these events are significant for polity scores, just not within a year. Another shortcoming is our small sample size of about 1000 country-year pairs in the final data set, which is due to the fact that we have a very small overlap in the time span of the data for all regions except Africa, which goes back to 1997 in ACLED. Three other regions go back to 2018, one to 2015 and one to 2020 in ACLED, while polity data *ends* at 2018. In fact, the majority of our data is from African countries. One should choose different data sets with higher overlap to remedy this. One last shortcoming is that the numbers of each type of events in a country-year are not normalized. The model may be improved by normalizing them with respect to population.