

Vasily Vasilyevich Vereshchagin



Conflicts, Cause of War in 19th and 20th Century

Capstone Project by Mehmet J Oguz

July 2019

Why do wars occur and recur?

Two prerequisites for a war between actors.

- 1- At least one sides that involved has to expect that the anticipated gains from a war in terms of resources, power, glory, territory.
- 2- Lack of ability to enforce or commit to an agreement

DATA

- Correlates of War Data
- Oslo Peace Research Institute's armed conflict data
- Stockholm International Peace Research Institute's war data
- Our World in Data

The variables included in the UCDP GED version 18.1:

Variable name	Content	Type
id	A unique numeric ID identifying each event.	integer
year	The year of the event	integer
active_year	1: if the event belongs to an active conflict/dyad/actor- Integer 0: otherwise	integer
type_of_violence	1: state-based conflict 2: non-state conflict 3: one-sided violence	integer
conflict_new_id	A unique conflict identification code for each individual conflict in the dataset.	integer
conflict_name	Name of the UCDP conflict to which the event belongs. For non-state conflicts and one-sided violence this is the same as the dyad name.	string(9999)
dyad_new_id	A unique conflict identification code for each individual dyad in the dataset.	integer
dyad_name	Name of the conflict dyad creating the event. A dyad is the pair of two actors engaged in violence (in the case of one-sided violence, the perpetrator of violence and civilians).	string(9999)
side_a_new_id	A unique ID of side A.	integer
gwnoa	The Gleditsch and Ward number for Side A if the side is a state. Empty if Side A is not a state.	string(9999)
side_a	The name of Side A in the dyad. In state-based conflicts always a government. In one-sided violence always the perpetrating party.	string(9999)
side_b_new_id	A unique ID of side B.	integer
gwnob	The Gleditsch and Ward number for Side B if the side is a state. Empty if Side B is not a state.	string(9999)
side_b	The name of Side B in the dyad. In state-based always the rebel movement or rivaling government. In one-sided violence always "civilians".	integer
number_of_sources	Number of total sources containing information for an event that were consulted.	text
source_article	References to the names, dates and titles of the source material from which information on the event is gathered. This variable is highly streamlined for information collected since 2013, and is less so for older data. For such older data, abbreviations are sometimes used for source agencies. The most frequent are: R: Reuters News, BBC: BBC Monitoring, AP: Associated Press Newswires AFP: Agence France Presse, X: Xinhua, DOW: Dow Jones Wires	text
source_office	The name of the organizations publishing the source materials.	text
source_date	The dates the source materials were published on.	text
source_headline	The titles of the source materials.	text
source_original	The name or type of person or organization from which the information about the event originates in the original report.	string(9999)
where_prec	The precision with which the coordinates and location assigned to the event reflects the location of the actual event. 1: exact location of the event known and coded. 2: event occurred within at maximum a ca. 25 km radius around a known point. The coded point is the known point. 3: only the second order administrative division where an event happened is known. That administrative division is coded with a point representing it (typically the centroid). 4: only the first order administrative division where an event happened is known. That administrative division is coded with a point representing it (typically the centroid). 5: the only spatial reference for the event is neither a known point nor a known formal administrative division, but rather a linear feature (e.g. a long river, a border, a longer road or the line connecting two locations further afield than 25 km) or a fuzzy polygon without defined borders (informal regions, large radiiuses etc.). A representation point is chosen for the feature and employed. 6: only the country where the event took place is known. 7: event in international waters or airspace	integer
where_coordinates	Name of the location to which the event is assigned. Fully standardized and normalized.	string(9999)
adm_1	Name of the first order (largest) administrative division string(9999) where the event took place	string(9999)
adm_2	Name of the second order administrative division where the event took place	string(9999)
latitude	Latitude (in decimal degrees)	numeric(9,6)
longitude	Longitude (in decimal degrees)	numeric(9,6)
geom_wkt	An Open Geospatial Consortium textual representation of the location of each individual point. Formatted as OGC WKT (well known text) without SRID.	string(9999)
priogrid_gid	An Open Geospatial Consortium textual representation of the location of each individual point. Formatted as OGC WKT (well known text) without SRID.	integer
country	The PRIO-grid cell id (gid) in which the event took place. Compatibility with PRIO-grid (Tollefson, 2012) is guaranteed for both PRIO-grid 1 and 2.	string(9999)
country_id	Gleditsch and Ward number of the country in which the event takes place.	integer
region	Region where the event took place. One of following: [Africa, Americas, Asia, Europe, Middle East]	string(9999)
event_clarity	1 (high) for events where the reporting allows the coder to identify the event in full. 2 (lower) for events where an aggregation of information was already made by the source material that is impossible to undo in the coding process.	integer
date_prec	How precise the information is about the date of an event. 1: exact date of event is known; 2: the date of the event is known only within a 2-6 day range. 3: only the week of the event is known 4: the date of the event is known only within an 8-30 day range or only the month when the event has taken place is known 5: the date of the event is known only within a range longer than one month but not more than one calendar year.	integer
date_start	The earliest possible date when the event has taken place.	Date YYYY-MM-DD
date_end	The last possible date when the event has taken place.	Date YYYY-MM-DD
deaths_a	The best estimate of deaths sustained by side a. Always 0 for one-sided violence events.	integer
deaths_b	The best estimate of deaths sustained by side b. integer Always 0 for one-sided violence events.	integer
deaths_civilians	The best estimate of dead civilians in the event. For non-state or state-based events, this is the number of collateral damage resulting in fighting between side a and side b. For one-sided violence, it is the number of civilians killed by side a.	integer
deaths_unknown	The best estimate of deaths of persons of unknown integer status.	integer
best_est	The best (most likely) estimate of total fatalities resulting from an event.	integer

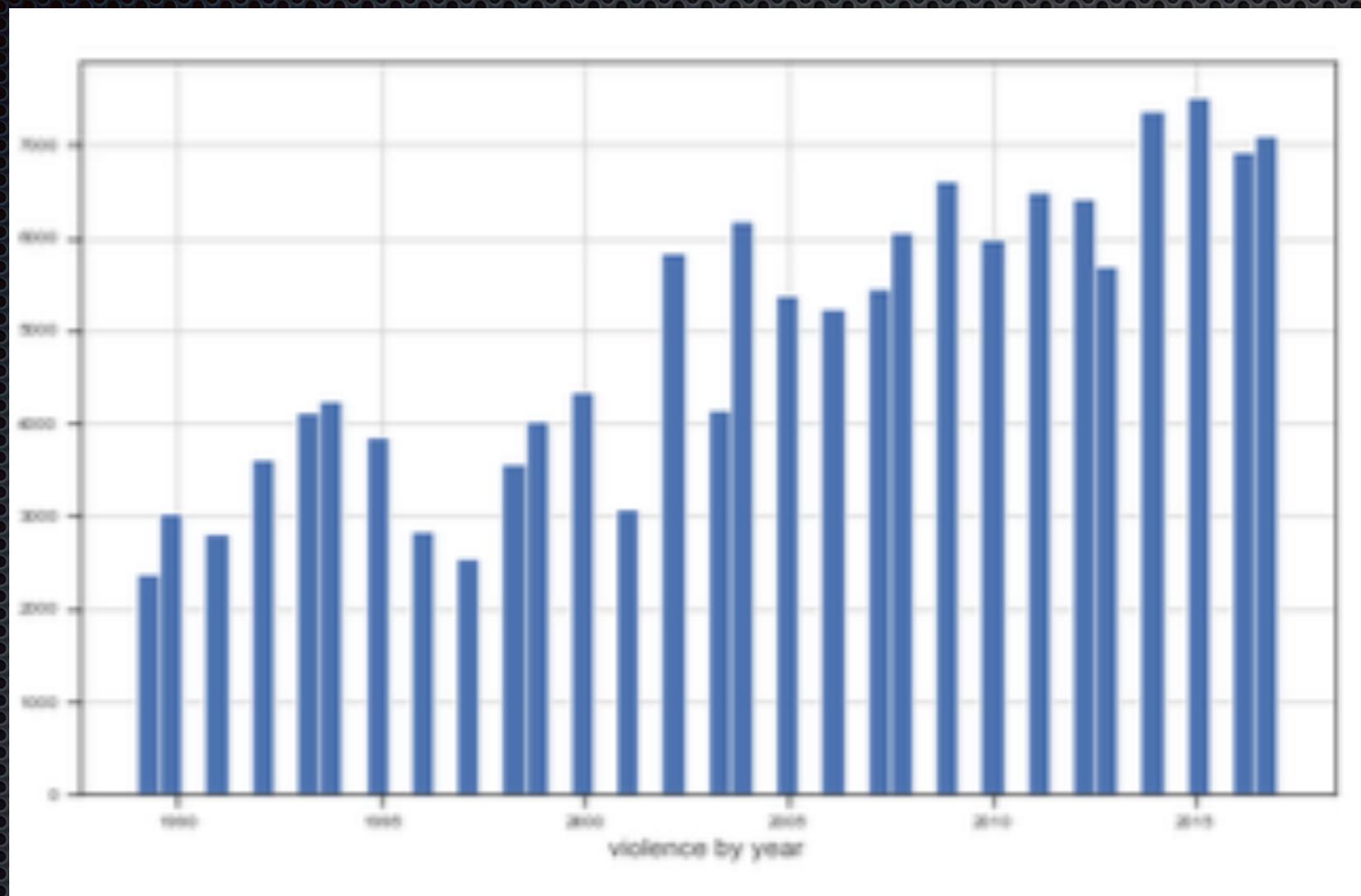
	conflict_name	side_a	side_b	where_coordinates	country	region	date_start	date_end
0	Sikh insurgents - Civilians	Sikh insurgents	Civilians	Punjab State	India	Asia	1989-01-01	1989-01-01
1	India:Punjab/Khalistan	Government of India	Sikh insurgents	Punjab State	India	Asia	1989-01-01	1989-12-31
2	JVP - Civilians	JVP	Civilians	Deniyaya town	Sri Lanka	Asia	1989-01-13	1989-01-13
3	Supporters of IFP - Supporters of UDF	Supporters of IFP	Supporters of UDF	Mpumalanga town	South Africa	Africa	1989-01-18	1989-01-18
4	NDFB - Civilians	NDFB	Civilians	Sonitpur district	India	Asia	1992-10-20	1992-10-20
5	LTTE - Civilians	LTTE	Civilians	North Eastern	Sri Lanka	Asia	1989-02-11	1989-02-11
6	IFP - Civilians	IFP	Civilians	Nqutu town	South Africa	Africa	1993-11-07	1993-11-07
7	Supporters of ANC - Supporters of IFP	Supporters of ANC	Supporters of IFP	Durban town	South Africa	Africa	1989-02-19	1989-02-19
8	Philippines:Government	Government of Philippines	CPP	Abra province	Philippines	Asia	1989-03-01	1989-03-03
9	Mohajir - Pashtun	Mohajir	Pashtun	Karachi Town	Pakistan	Asia	1989-03-10	1989-03-14

Preparation

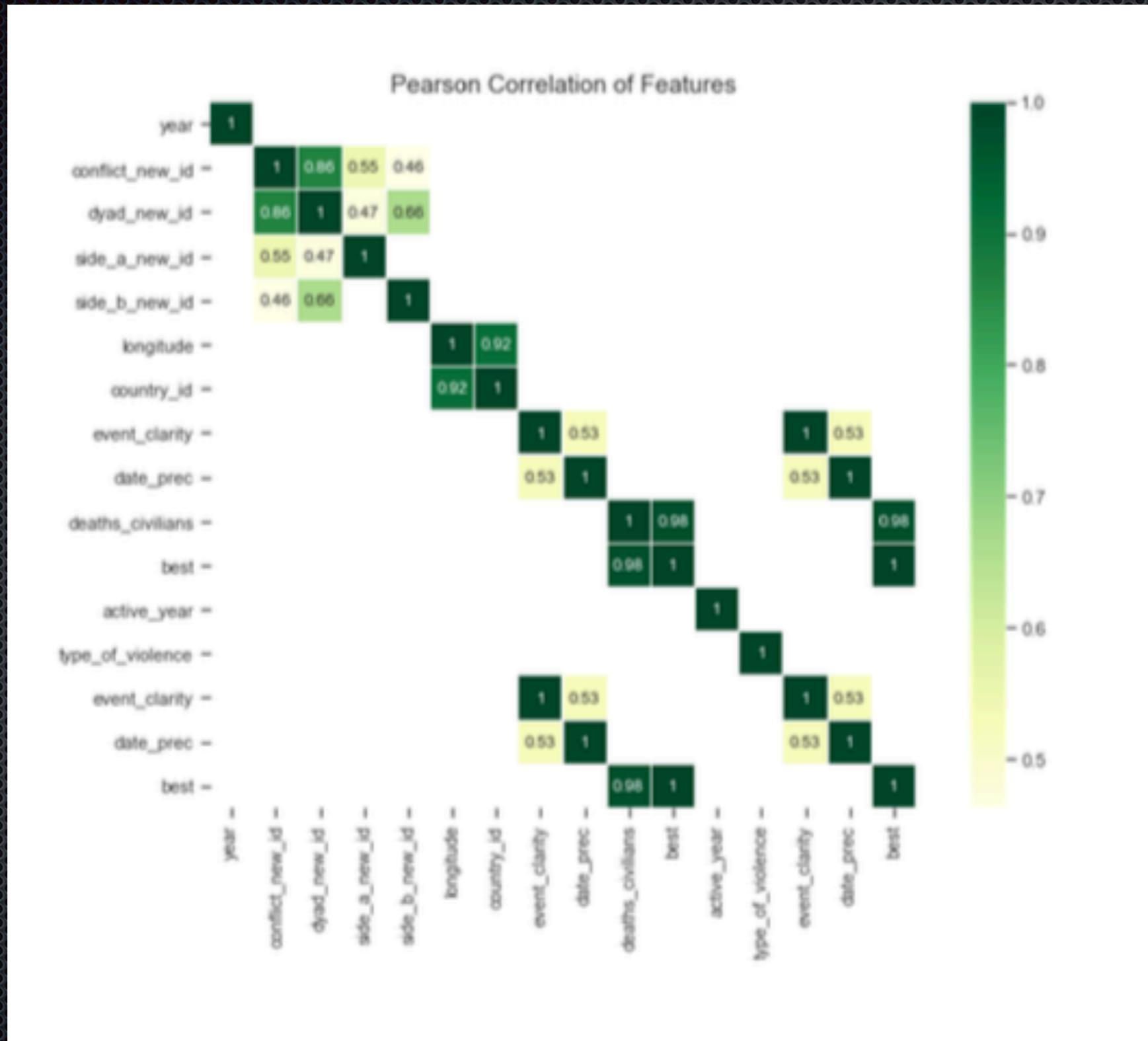
- All datasets were explored with Pandas methods such as .head(), .info(), and .describe(), and DataFrame attributes like .columns and .shape .
- Dataframes merged.
- Null Data filled
- Seaborn pairplot

Preliminary Exploration and Findings

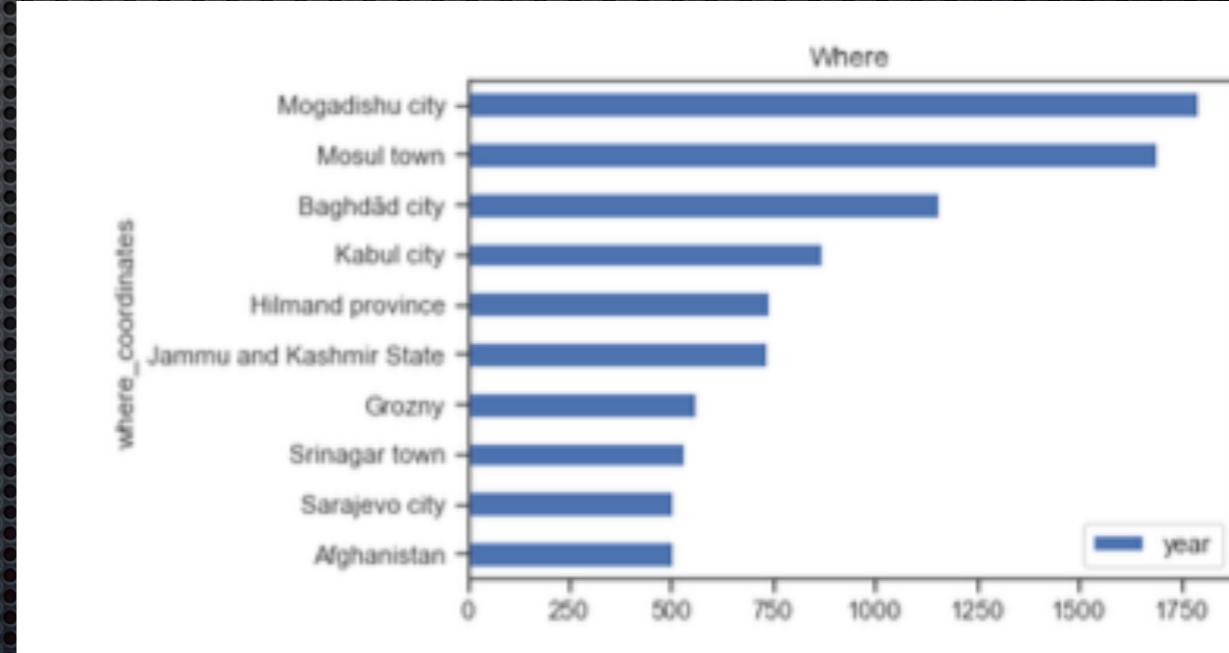
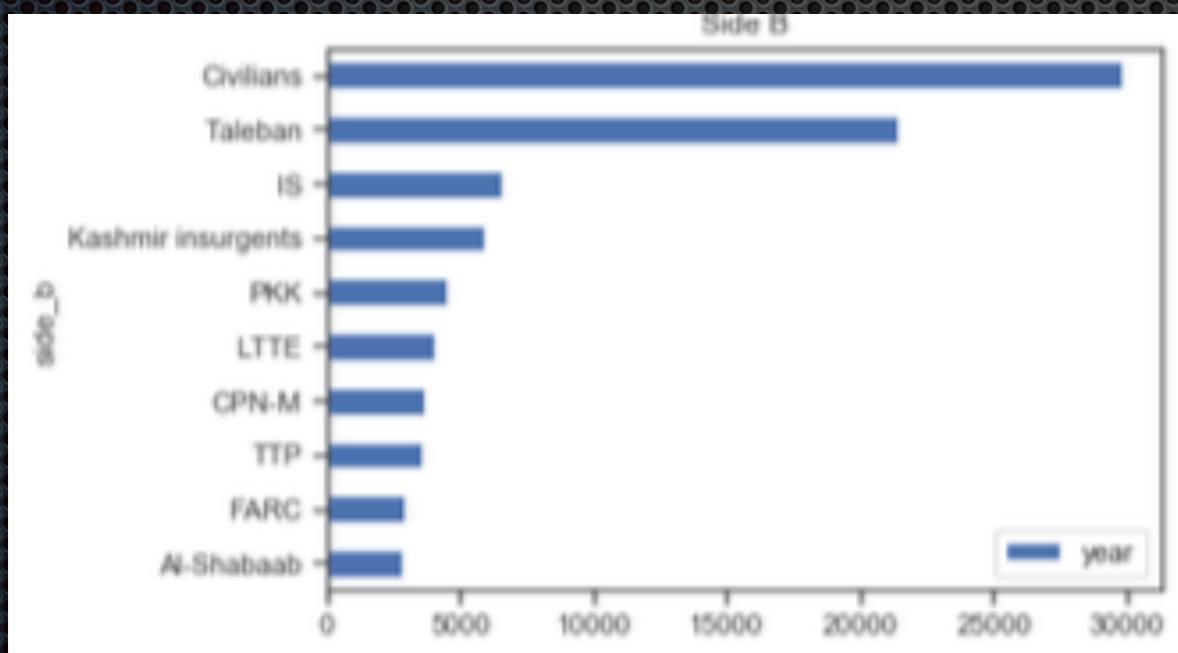
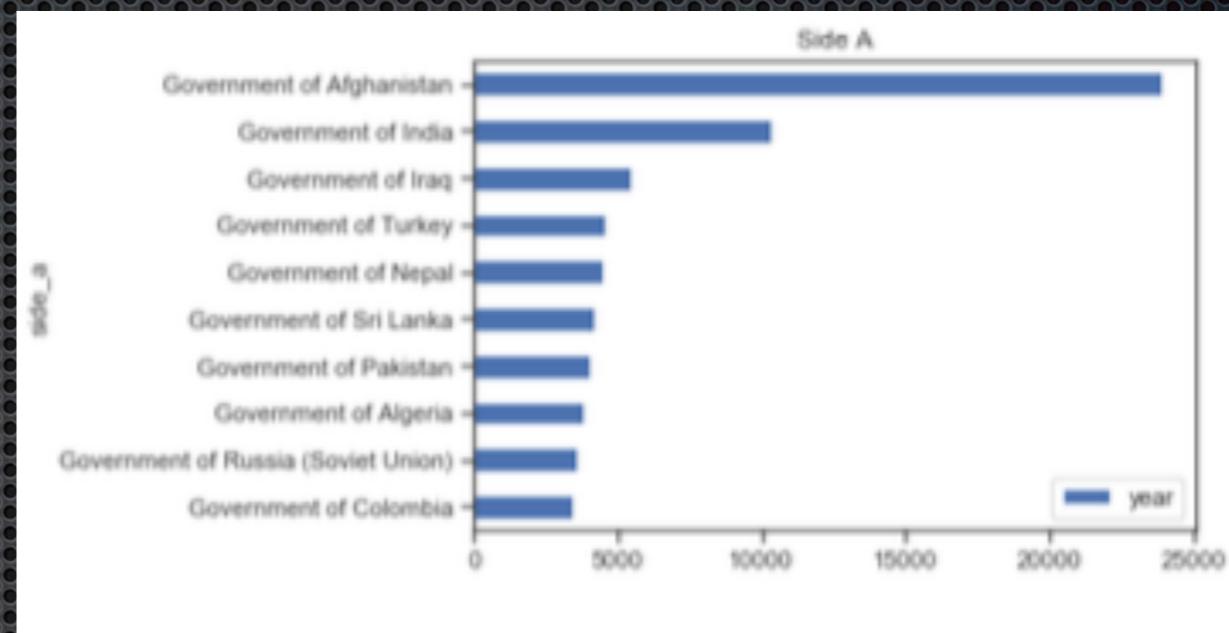
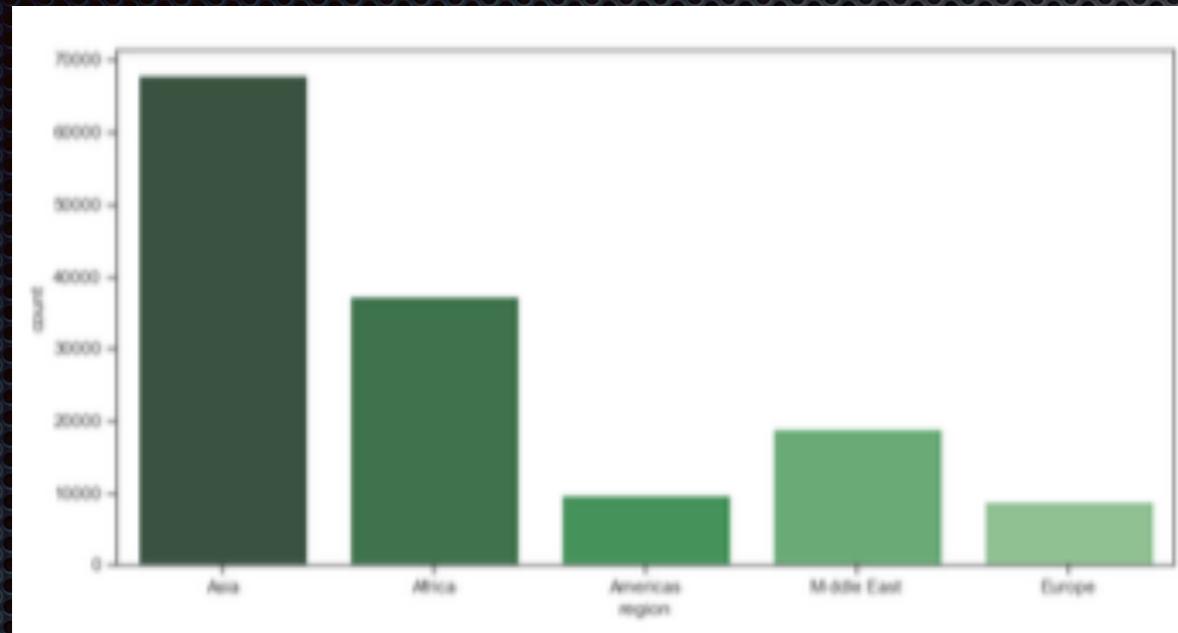
There is an increase in violence



Pearson Correlation graph shows that there is a strong relationship between the events clarity and reporting time. I take in consideration of correlation that is greater or equal to 0.5 or less than or equal to -0.4.



As seen in below graphic, 8 out of the first ten countries that involved in a conflict are in Asia. What is the main reason behind Asian conflicts?



MODEL APPROACH

Data split into 70% training and 30% testing set.

Mean absolute error (MAE) used as metric.

The baseline guess is a score of 45.8 and baseline performance on the test set is MAE = 32.3763. This shows our average estimate on the test set is off by about 32 points.

	year	conflict_new_id	dyad_new_id	side_a_new_id	side_b_new_id	longitude	country_id	event_clarity	date_prec	deaths_civilians
count	100031.000000	100031.000000	100031.000000	100031.000000	100031.000000	100031.000000	100031.000000	100031.000000	100031.000000	100031.0000
mean	2005.407264	1369.342504	2054.364117	296.585808	504.237017	45.539838	609.562286	1.132519	1.315792	7.0808
std	8.086251	2970.388842	3464.004630	666.758887	983.666835	45.180329	187.916910	0.387902	0.793880	961.8379
min	1989.000000	205.000000	406.000000	3.000000	1.000000	-117.046450	2.000000	1.000000	0.000000	0.0000
25%	1999.000000	333.000000	735.000000	112.000000	209.000000	29.448776	517.000000	1.000000	1.000000	0.0000
50%	2007.000000	364.000000	782.000000	130.000000	303.000000	45.872400	666.000000	1.000000	1.000000	0.0000
75%	2012.000000	499.000000	974.000000	154.000000	488.000000	74.167293	750.000000	1.000000	1.000000	1.0000
max	2017.000000	14333.000000	15538.000000	7046.000000	7014.000000	155.894681	940.000000	3.000000	5.000000	300559.0000

To develop a model that can predict the conflict.

A supervised regression task to develop a model that can predict the conflict.

The results presented below:

1. Linear Regression Performance on the test set:

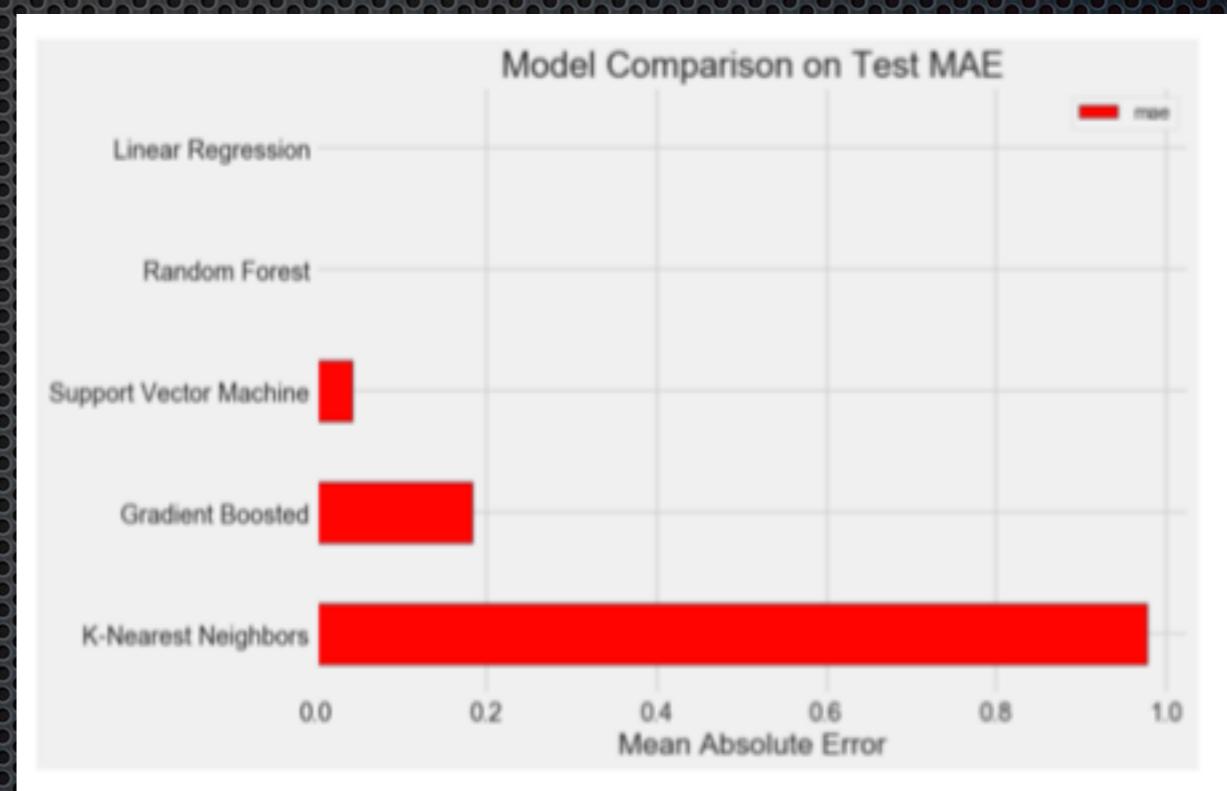
$$\text{MAE} = 0.0000$$

2. Support Vector Machine Regression Performance on the test set: $\text{MAE} = 0.0424$

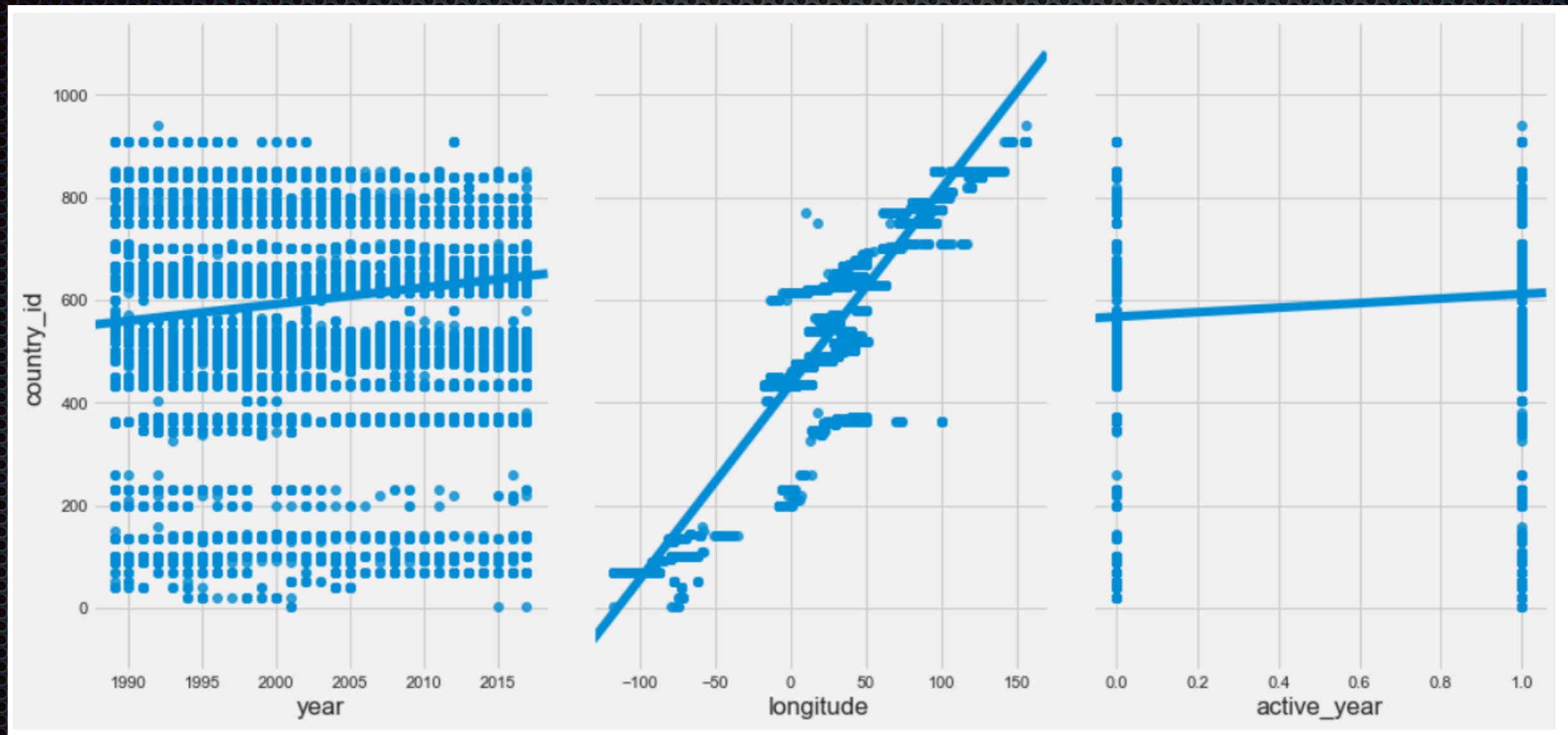
3. Random Forest Regression Performance on the test set: $\text{MAE} = 0.0018$

4. Gradient Boosted Regression Performance on the test set: $\text{MAE} = 0.1832$

5. K-Nearest Neighbors Regression Performance on the test set: $\text{MAE} = 0.9770$



Linear Regression used as a model optimization, and as seen in the graph, there is a strong relationship between longitude and country _id.



The data split into 70% of training, 30% testing set, and the slope and intercept of the data contained in the model's fit parameters. The interception of the model is 435.83437, and the slope is 3.80887. Means that for every one unit of change in country_id, the change in the longitude is about 3.80%.

Mean Absolute Error: 53.00359

Mean Squared Error: 5611.81805

Root Mean Squared Error: 74.91206

The root mean squared error is 74.91, which is higher than the mean value of the longitude of all states, which is 45.58. The result means that our algorithm was not very accurate but can still make reasonably good predictions.

CONCLUSION

Many factors may have contributed to this inaccuracy like we need more data, or we made the wrong assumption that this data has a linear relationship. Another reason might be the poor features we used may not have had a high enough correlation to the values we were trying to predict.

The essential causes of war are found in the nature and behavior of man. Conflicts result from selfishness, from misdirected aggressive impulses, from stupidity. If these are the primary causes of war, then the elimination of war must come through uplifting and enlightening men or securing their psychic-social readjustment.

RESOURCES

<https://www.prio.org/Data/Armed-Conflict/>

<https://ourworldindata.org/war-and-peace>

<https://ucdp.uu.se>

<https://www.theobjectivestandard.com/2014/10/causes-war-peace/>

Jupiter Notebook

Thank you

Mehmet J Oguz

www.linkedin.com/in/mjoguz