Are some ideologies more violent than others? Quantitative evidence from Asia and Africa

Bastian Lotz, Ruben van Zoest, Jon Vrushi

15 December 2017

Theory

The central research question that this data analysis tries to answer is whether certain political or religious ideologies are more violent than others.

To that end, we propose the following Hypothesis

H1 -> There is a statistically significant difference between the number of fatalities registered for groups with different ideologies.

We operationalise the independent variable of ideology by using as a proxy the presence of strings in actor names that can be used to mark a certain ideology. We then end up with 7 political and religious ideologies: *Christian*, *Islam*, *Ethnic-based*, *Clan-based*, *Revolutionary*, *Republican*, *Democratic* and *Liberationary* groups.

We operationalise the measurement of violence for each ideology based on a proxy which measures the number of fatalities for each conflict registered. This will serve as our dependent variable.

Preparations

We load the ACLED data of Africa 1997 - 2016 and 2017 as well as for Asia 2015 - 2017 into R (For all codes not included in this output, see the attached markdown file).

We then merge the data into one data frame.

In a next step, we assign a region variable as well as a GDP per capita variable to to every observation.

In order to analyze ideologies of actors in later steps, we need to transform our data frame from a wide into a long format (i.e. into a monadic file). In the resulting data frame, each observation is one actor and every event appears twice.

We use regular expressions and the grepl() function to find those ideologies in actor names that we are interested in, namely *Christian*, *Islam*, *Ethnic-based*, *Clan-based*, *Revolutionary*, *Republican*, *Democratic* and *Liberationary*. We then extract them as a new variable *Actor_Ideology* measuring the ideology of the group in question for every actor involved in a conflict.

In order to be able to work with the variable EVENT_TYPE, we transform it into factor variable. Note that the data includes some spelling error in *EVENT_TYPE* that lead to double entries of two categories. We account for this by first incorporating it in the transformation to a factor variable and then renaming the respective category so we yield the correct amount of categories.

We deactivate scientific notation to create more intuitive plot labels.

Exploratory Data Analysis

Summary statistics

Some summary statistics give us a first picture of our data.

We can see that the mean value of *FATALTIIES* is 3.75. For the standard deviation, we get a rather large value of over 70, which points to substantial variation within our data. Given that the standard deviation is much higher than the average, we need to also account for this in our model. Moreover, we compute the range of *FATALTIIES* (minimum and maximum values). We yield a very large range in *FATALTIIES*, where the minimum value is 0 and the maximumg value is 25000.

Table 1: Summary statistics

	vars	n	mean	sd	min	max	range	se
X1	1	193856	3.752363	70.36753	0	25000	25000	0.1598206

Finally, we calculate the proportion of zero *FATALTITIES* in our data and check if there are any missing values. We find that over 70% of observations have zero *FATALITIES*. This is an important finding to keep in mind for our model assumptions. However, there are no missing values in *FATALTITIES*.

Table 2: Proportion of Zero Fatalities and Count of missing values

| proportion_zerofatalities	NA_FATALITIES_Count
0.723372	0

Distributions

Distribution of Fatalities (overall)

Looking at the distribution of fatalities (frequency polygon), we can see that it is highly skewed as a result of a high share of zero fatalities and large outliers.

Years: 1997–2017 (Africa); 2015–2017 (Asia)
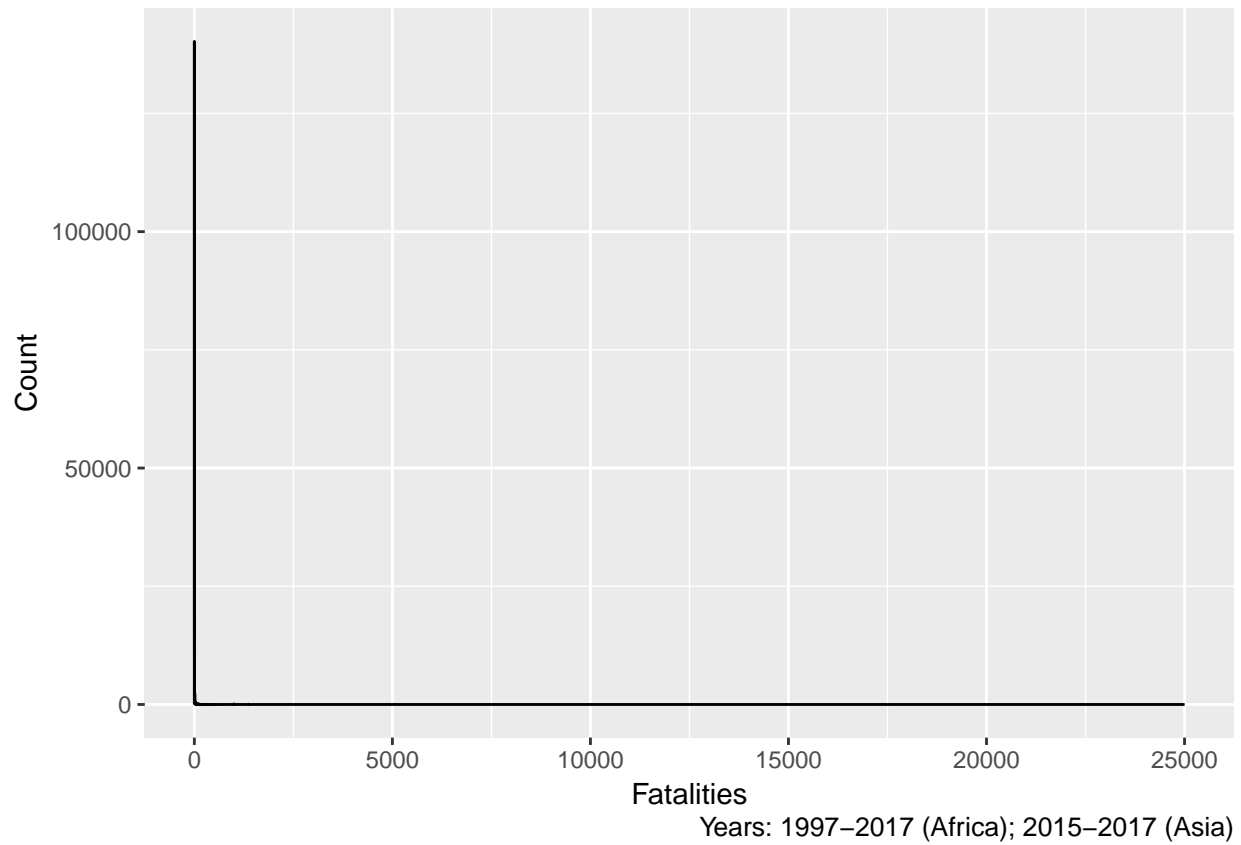
To get a better picture of the usual values of fatalities, we restrict the data for the frequency polygon to 30 fatalities. Only 2872 observations are above that threshold.

Years: 1997–2017 (Africa); 2015–2017 (Asia)
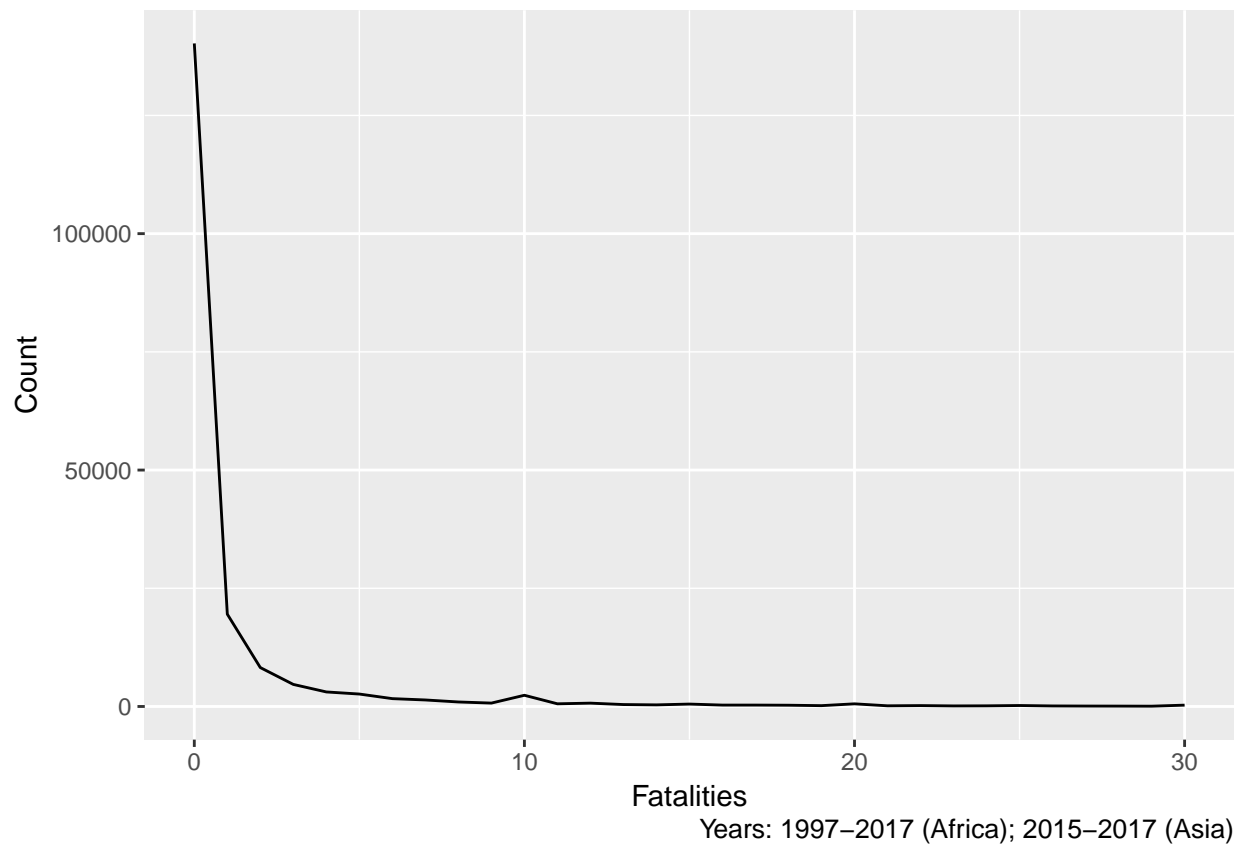
Distribution of Fatalities (subset by event type, region and actor_ideology)

We subset the distribution of *Fatality* by *EVENT-TYPE*, *region* and *Actor_Ideology*. Distributions are largely similar across any category and show the same level of skewness, large share of zeros and many large outliers.

Distribution of Fatalities subset by event type

Years: 1997−2017 (Africa); 2015−2017 (Asia)
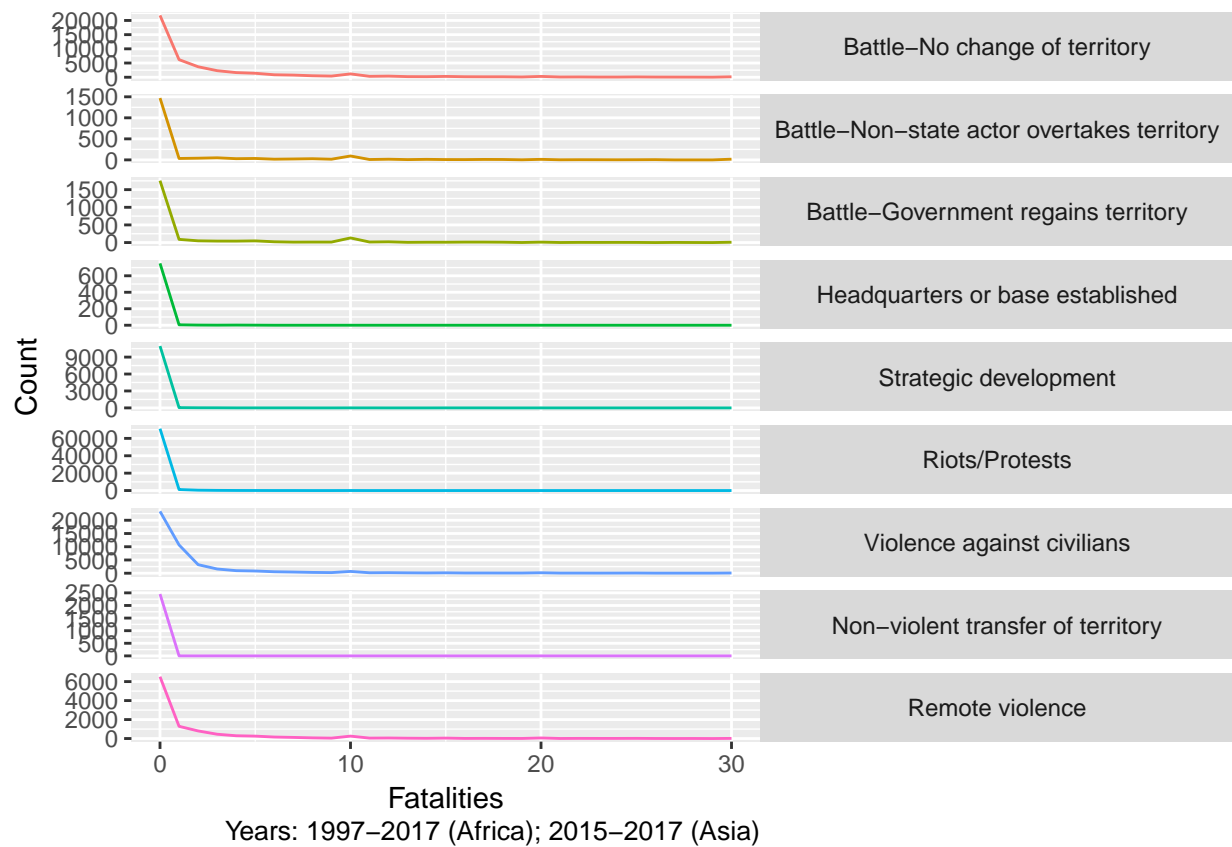
Distribution of Fatalities subset by region

Years: 1997–2017 (Africa); 2015–2017 (Asia)
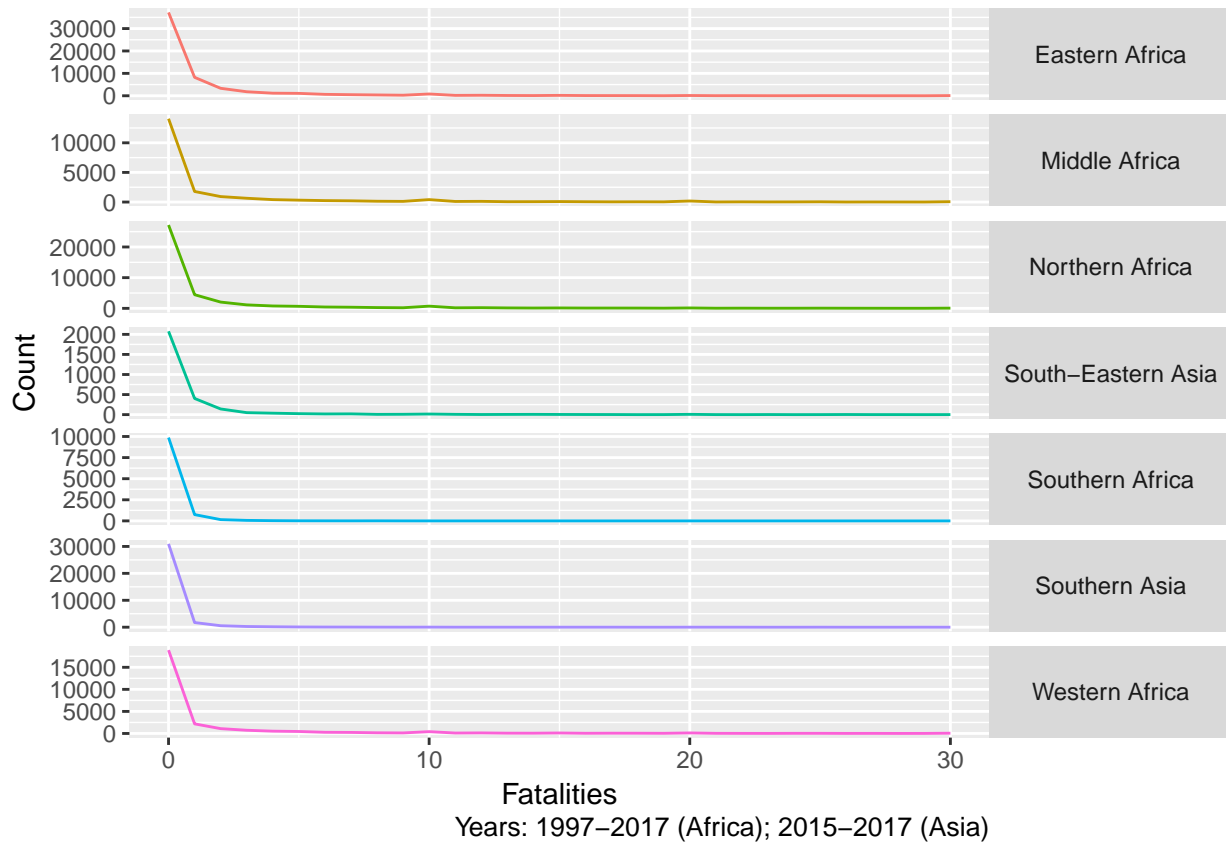
**Distribution of Fatalities subset by actor ideology**

We first clean the long data set by removing the unspecific actor ideologies, i.e. "Other", "Unidentified" and "None-Civilian" and then display the distribution of *FATALITIES* by the relevant categories of *ACTOR_IDEOLOGY*.

Years: 1997–2017 (Africa); 2015–2017 (Asia)
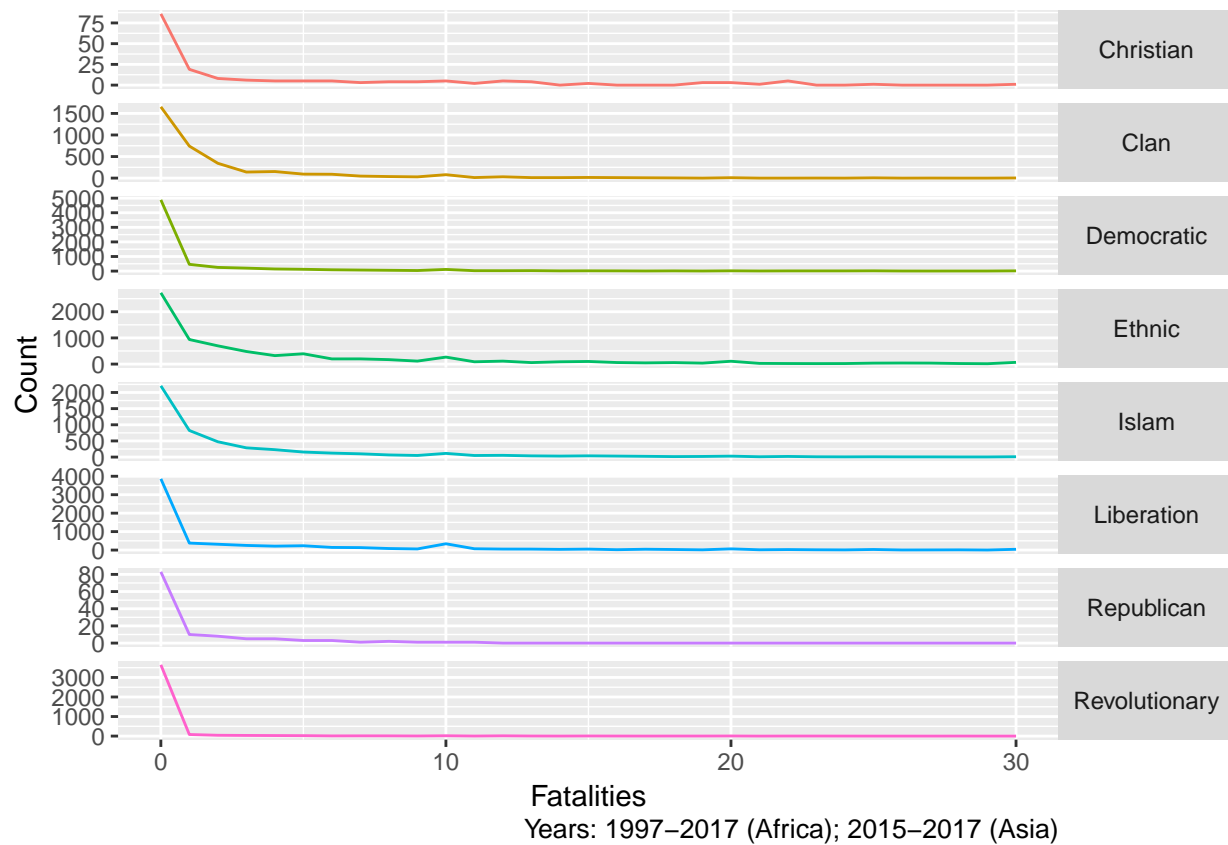
Distribution of Ideologies of Actors

We now take a look at the distribution of *ACTOR_IDEOLOGY*.

Actors with *ethnic*, *liberation* and *democratic* ideology are responsible for the highest number of conflict incidences. Groups who are associated with *Islam*, *Revolutionary* or *Clan* follow in terms of the amount of conflicts. *Republican* and *Christian* groups exhibit very few incidents of conflict.

Years: 1997–2017 (Africa); 2015–2017 (Asia)
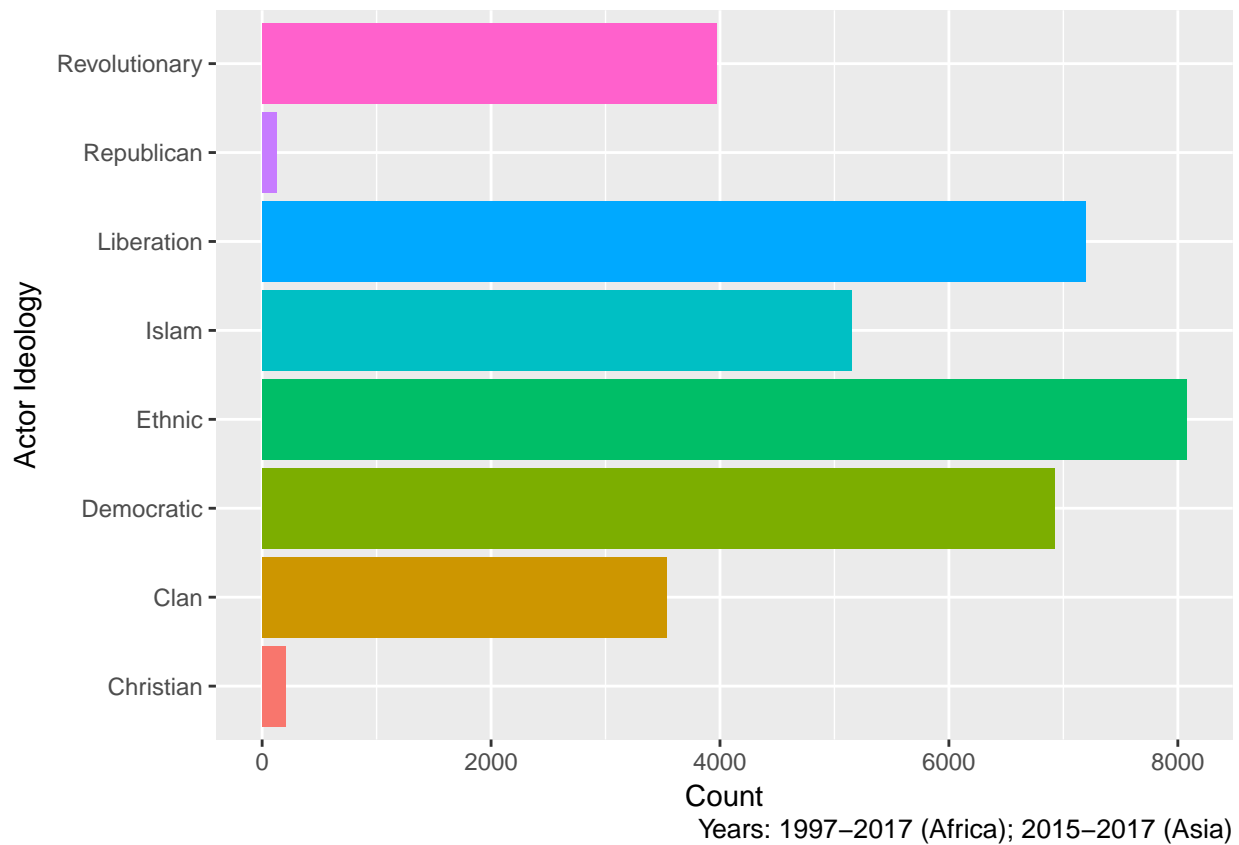
We subset the the distribution of *Actor_Ideology* by *region* in order to analyze potential regional variance. Please note that low counts in South-Eastern Asia and Southern Asia is mainly attributable to the lack of data for these region before 2015. We can see that *Clan* ideology is particularly present in Eastern Africa, whereas *Liberation* and *Islam* are the major actor ideologies in Northern Africa. In Western Africa, *revolutionary* ideology is prominent. Southern Africa does not show any pattern with respect to ideology. Moroever, *Islam* is the most important ideology for South-Eastern Asia and to lesser extent for Southern Asia.

Count
Years: 1997–2017 (Africa); 2015–2017 (Asia)
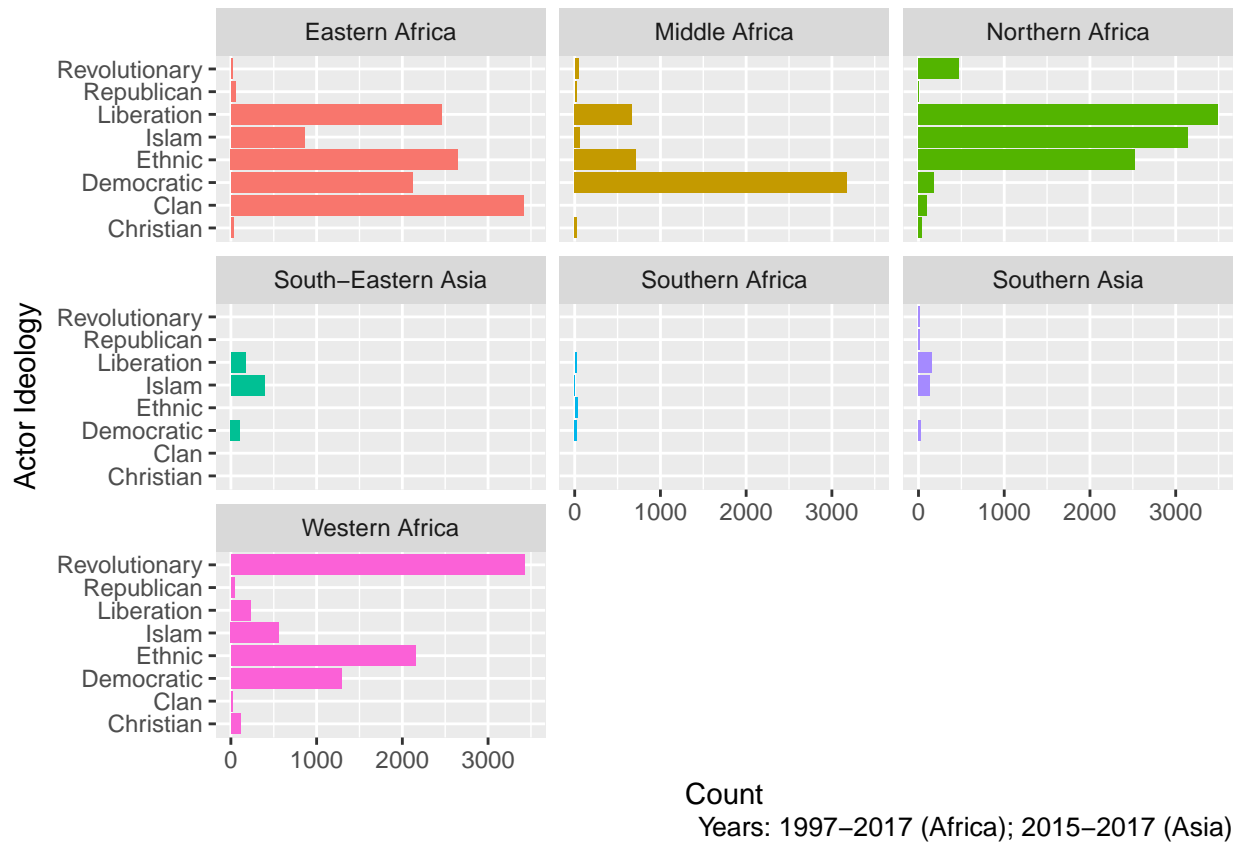
Distribution of EVENT_TYPE

We continue with an analysis of the distribution of *EVENT_TYPE*.

We first clean the data to remove missing values in *EVENT_TYPE*. Then, we plot the counts of *EVENT_TYPE* to get a better picture of the distribution of this variable.

We see that most conflicts are "riots/protests" (around 73 000 times), followed by "Violence against civilians" and "Battles that do not involve a change of territory" (around 45000 respectively). "Headquarters or base established", "Battle-Government regains territory", "Non-violent transfer of territory" and "Battle-Non state actor overtakes territory" produce relatively little victims.

Years: 1997–2017 (Africa); 2015–2017 (Asia)
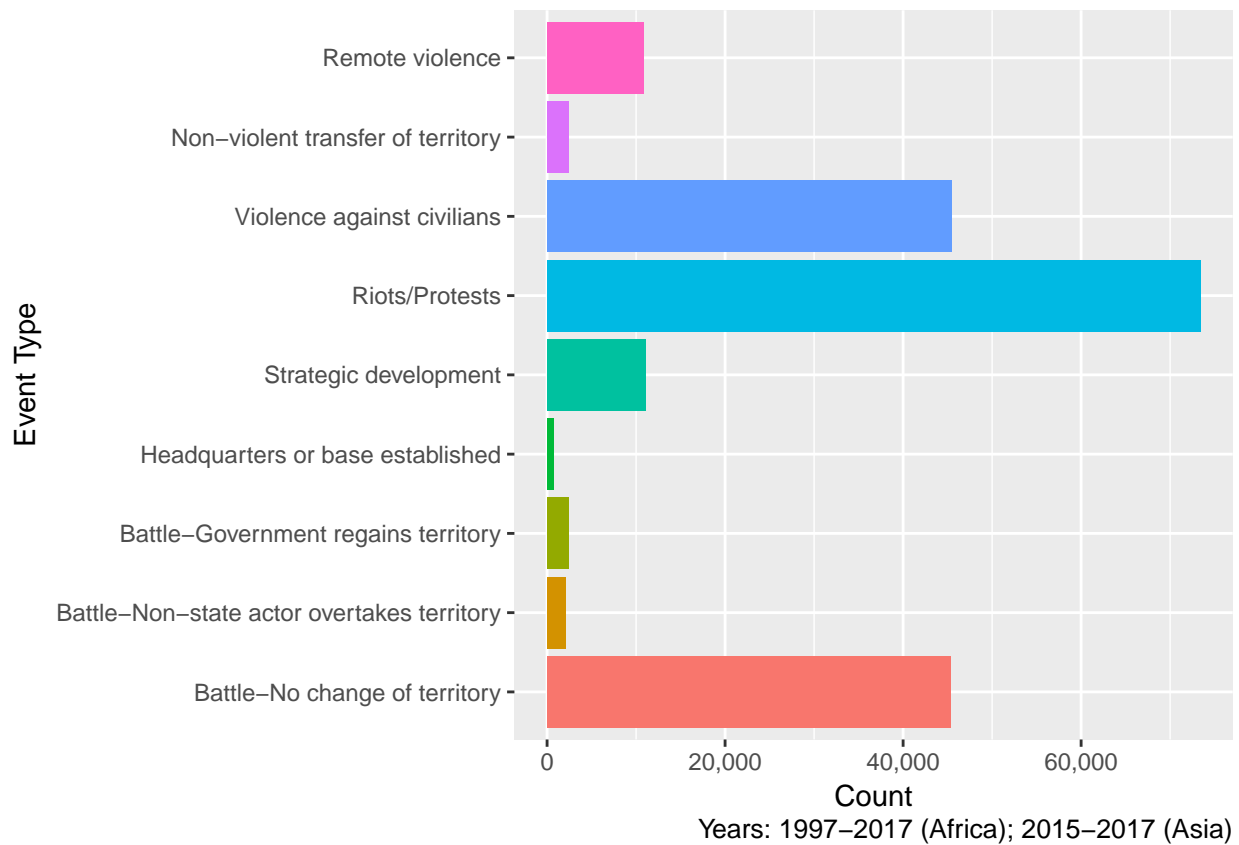
We take a more in-depth look by stacking the former graph by regions. Please note again, that Asia is significantly underrepresented in the data. Hence, low counts are at least partly attributed to this fact. However, we can still see that "Riots/Protest" are a particularly frequent event type for Southern Asia. Moreover, violence against civilians occurs comparatively often in Eastern Africa and Northern Africa. Moreover, "Battle-no change of territory" is dominated by incidents of conflict in Eastern Africa.

Years: 1997–2017 (Africa); 2015–2017 (Asia)
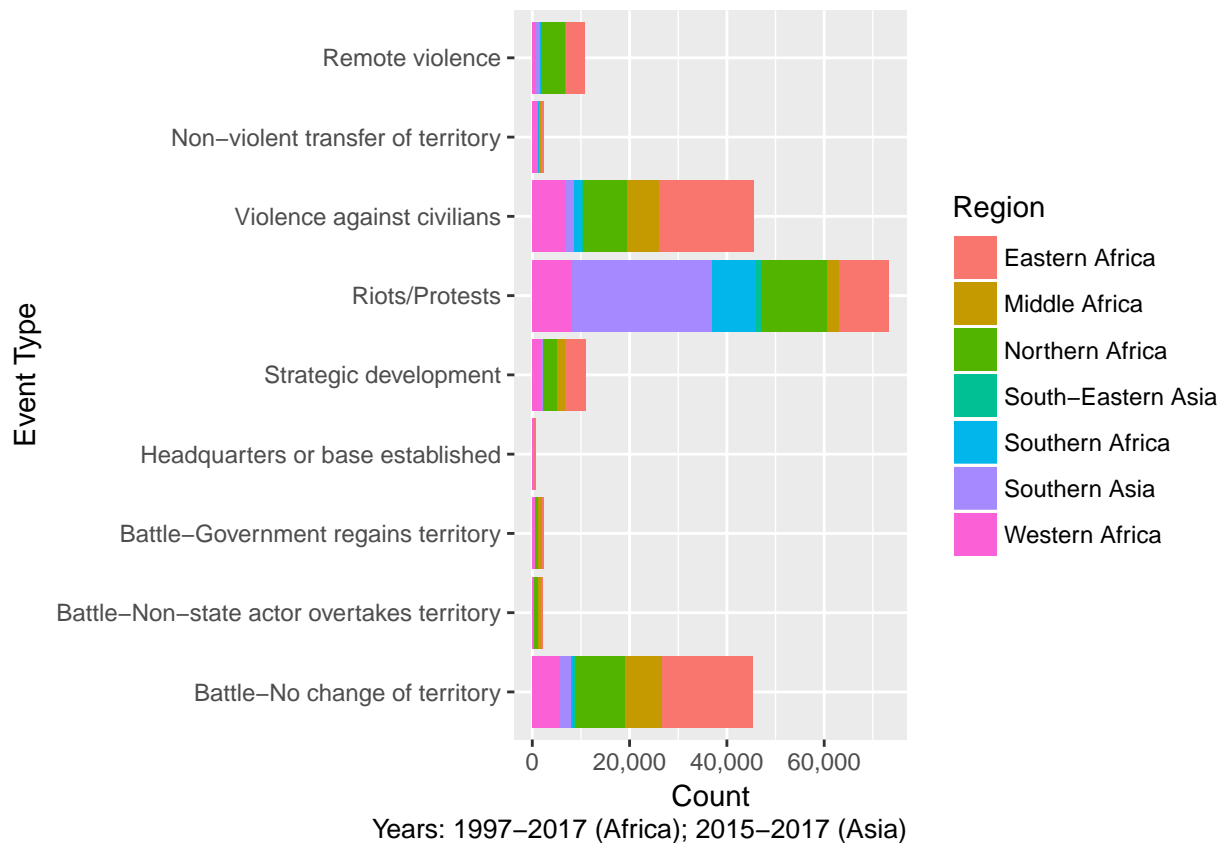
Distribution of Region

Looking at the overall regional distribution for all years, not taking into account that data for Asia is not available before 2015, Eastern Africa exhibits the most incidents of conflicts at 57 700. Northern Africa (ca. 40 000), Western Africa (ca. 27 000) and Eastern Africa (ca. 21 000) also demonstrate relatively high numbers of incidents. Surprisingly, Southern Asia has the third largest number of conflict incidents (ca. 34 000), even though data on this region is only available for the years 2015 to 2017.

Years: 1997–2017 (Africa); 2015–2017 (Asia)
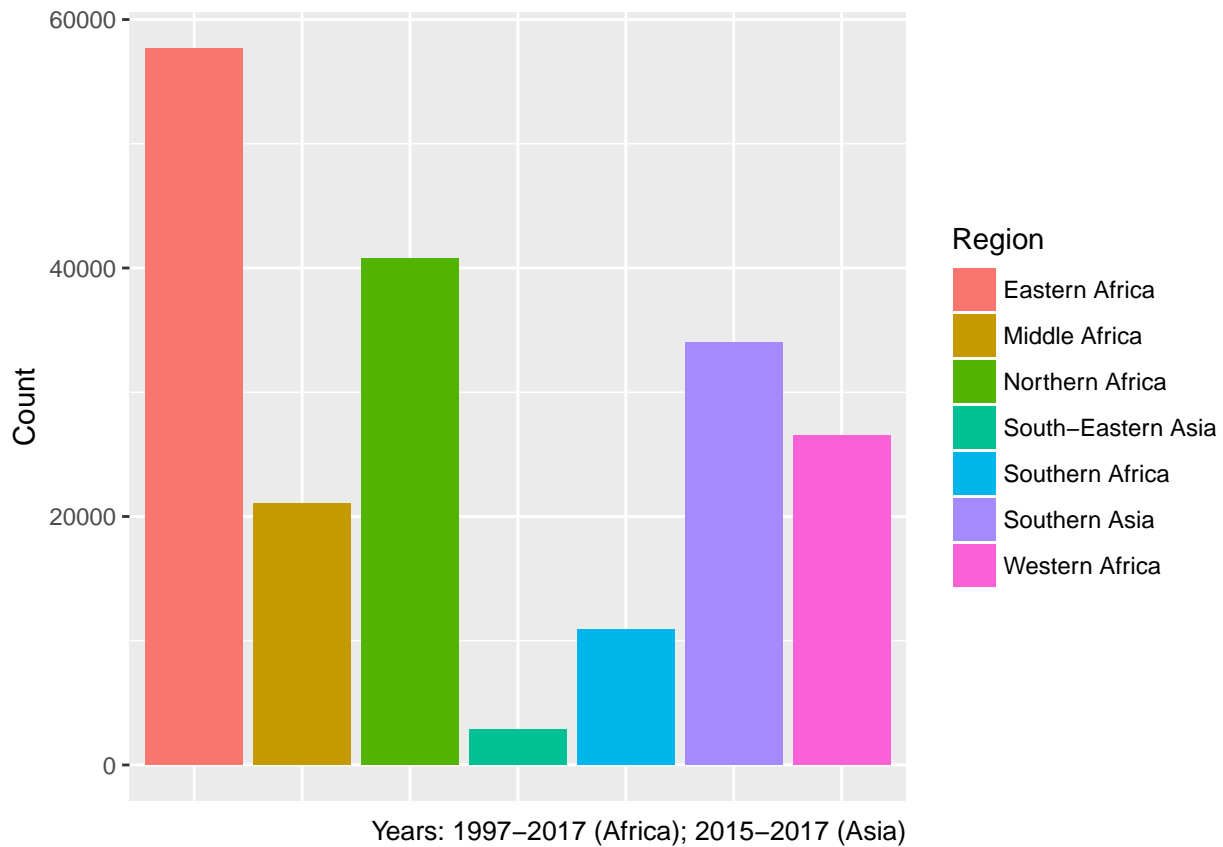
Covaration

In order to get a first idea of covariation between our dependent variable and relevant independent variables, we conduct the following analysis.
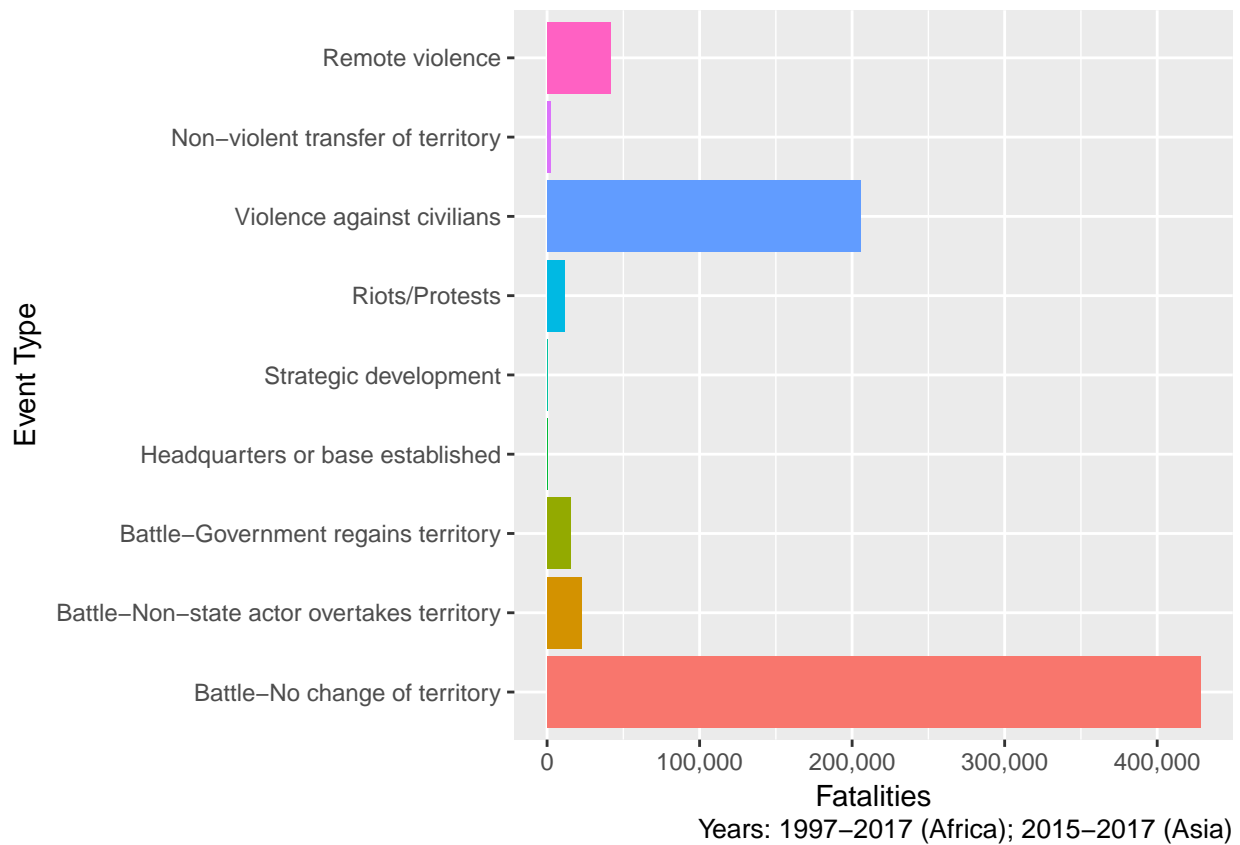
Between Fatalities and Event Type

The boxplot shows very similar distributions across event types. Median is at 0 for all event types. Violence against civilians has a particularly large outlier at 25 000 fatalities and Battle-No Change of territory has an outlier at around 6000.

Years: 1997–2017 (Africa); 2015–2017 (Asia)
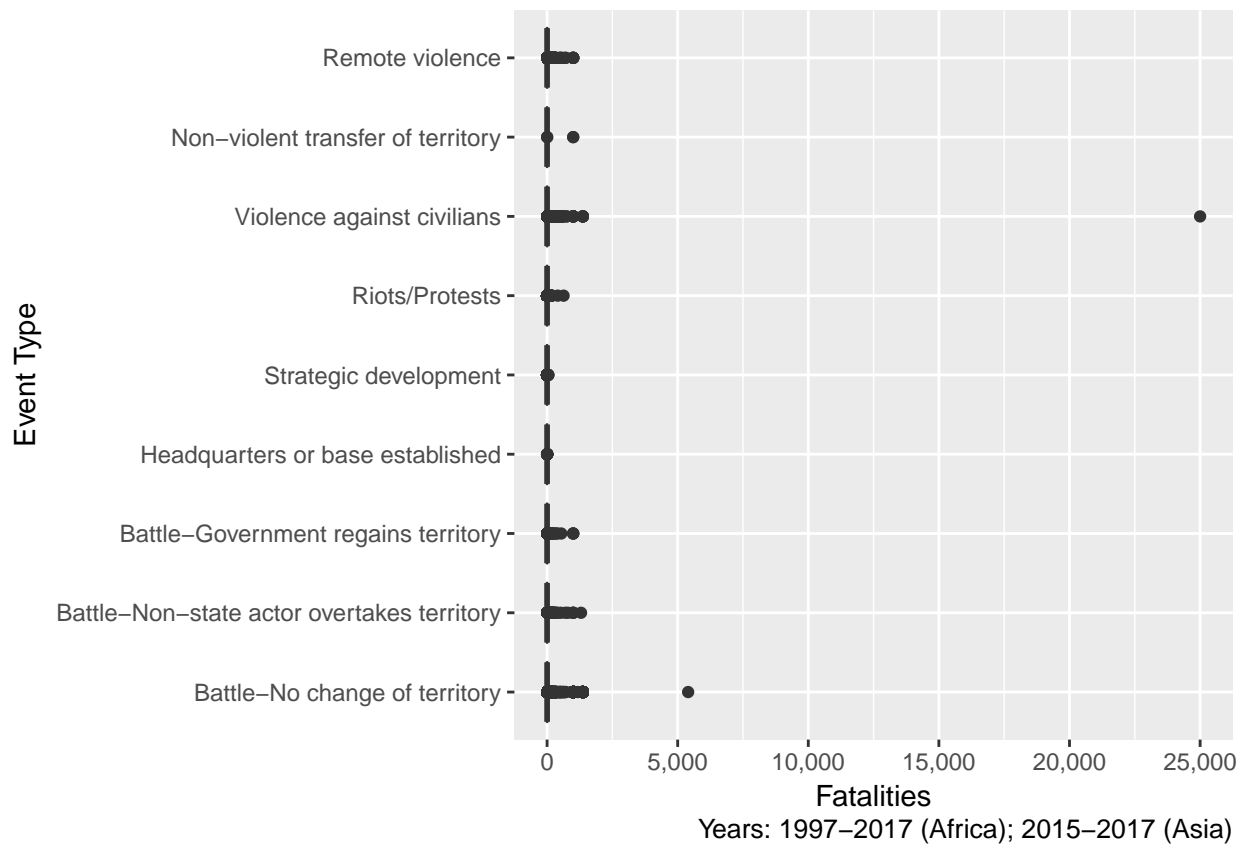
Filtering out the two outliers, gives the following boxplot, that allows to show more variation within the categories. The skewness of fatalities is clearly visible with most values at or around zero and many large outliers present.
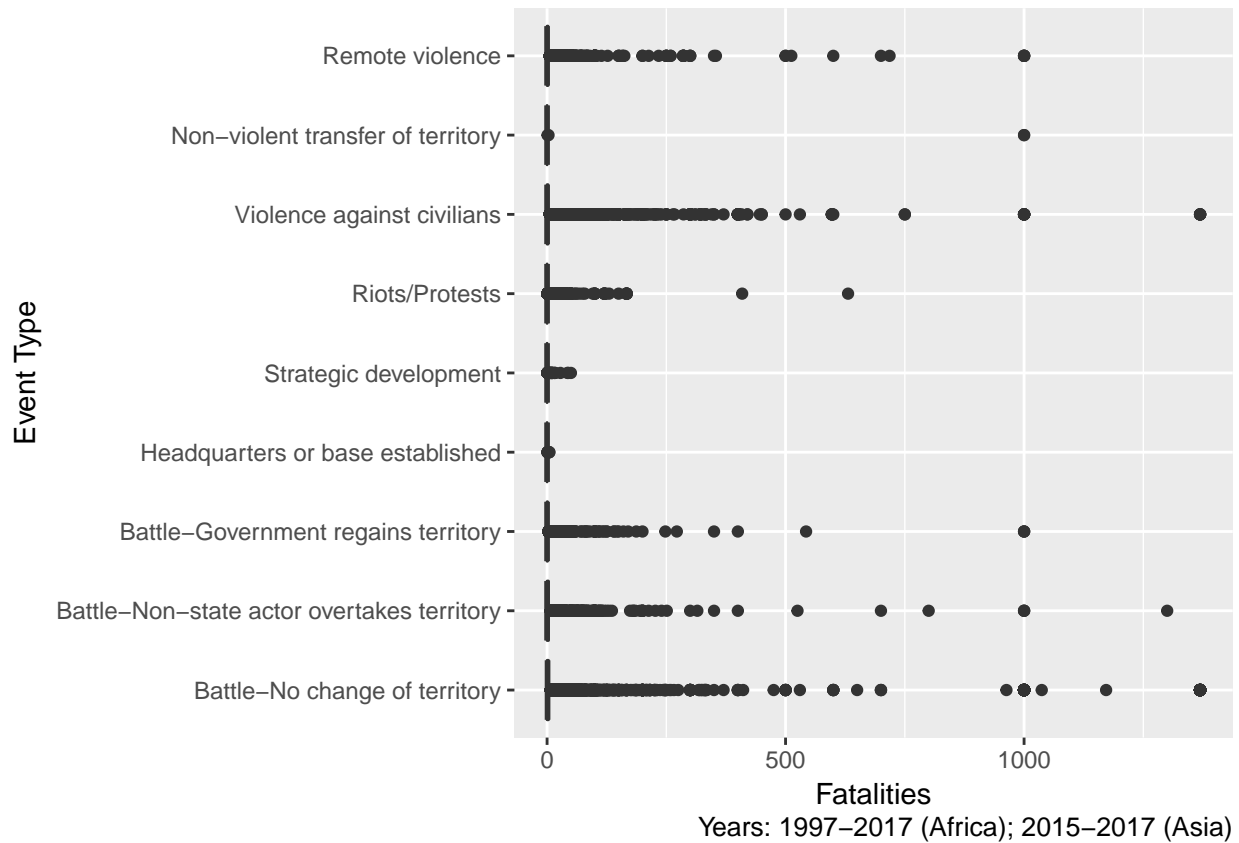
Event types "Violence against civilians" and "Battle-No change of territory" have most fatalities. However, these are also the two event types with the largest outliers. Other event types exhibit comparatively few victims.

Years: 1997–2017 (Africa); 2015–2017 (Asia)

Between Fatalities and Region

The boxplot shows very similar distributions across regions. Median is at 0 for all regions. The skewness of fatalities is clearly visible with most values at or around zero. Besides Southern Asia, South-Eastern Asia and Southern Africa, all regions exhibit large outliers.

Fatalities
Years: 1997–2017 (Africa); 2015–2017 (Asia)
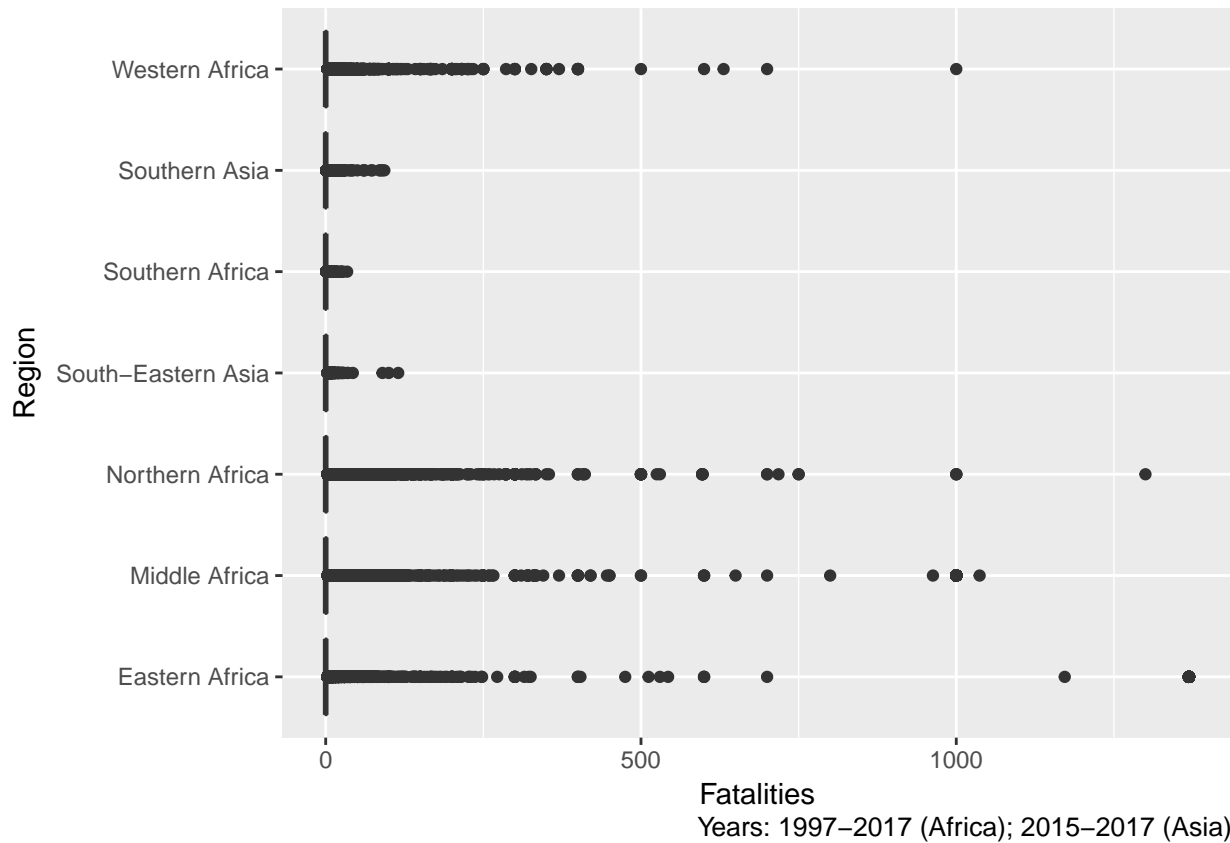
A bar graph demonstrating the relationship between *region* and *FATALITIES* shows us that Eastern Africa (ca. 23 000) and Middle Africa (ca. 25 000) have most fatalities. Northern Africa also features relatively high numbers of fatalities. Southern Africa has very few victims. For Asian regions numbers are also low, but data constraints need to be taken into account here.

Region

Years: 1997–2017 (Africa); 2015–2017 (Asia)
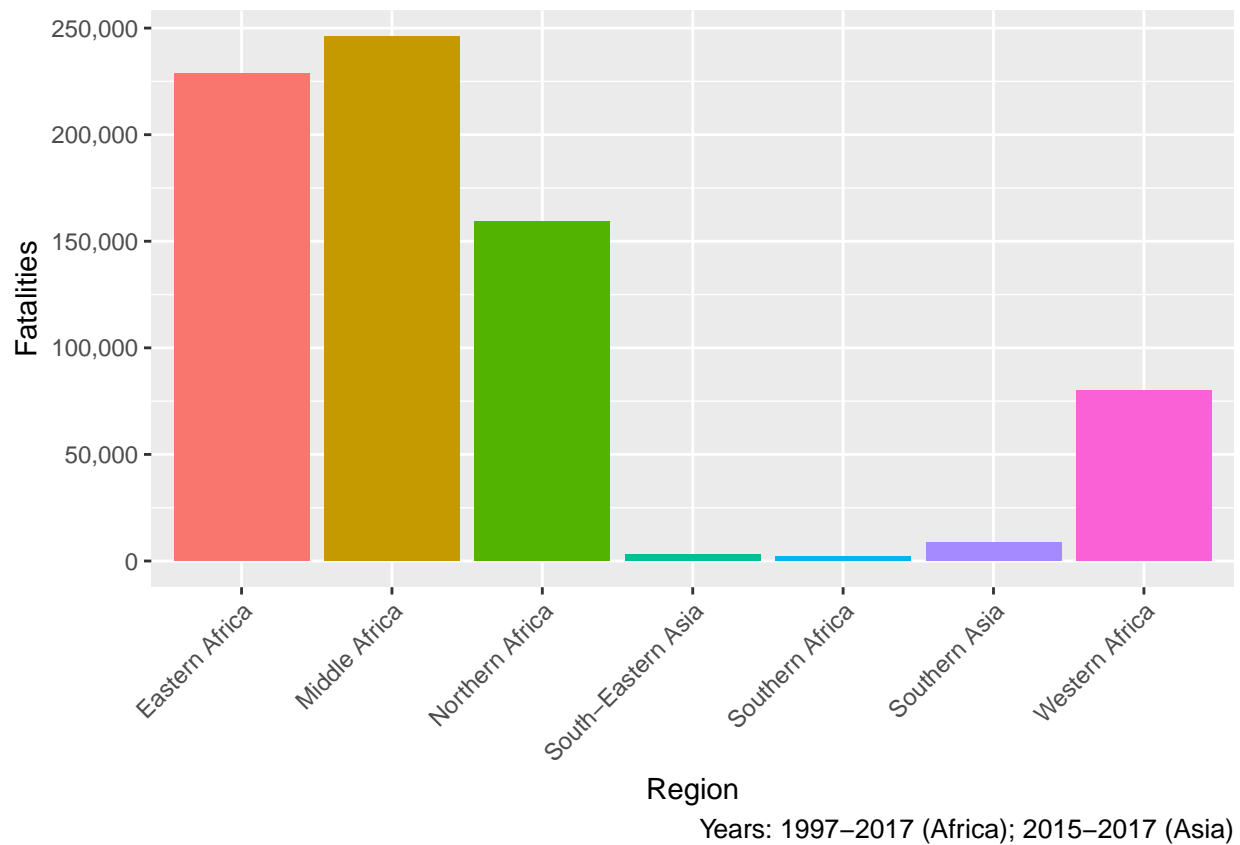
Between Fatalities and Actor Ideology

We create boxplots with no limit on fatalities, limit at 1400 and limit at 30 fatalities.

In the boxplot with no limit on *Actor_Ideology*, we see that the largest outlier in our data (25000 fatalities) is within the Liberation ideology.

Years: 1997–2017 (Africa); 2015–2017 (Asia)
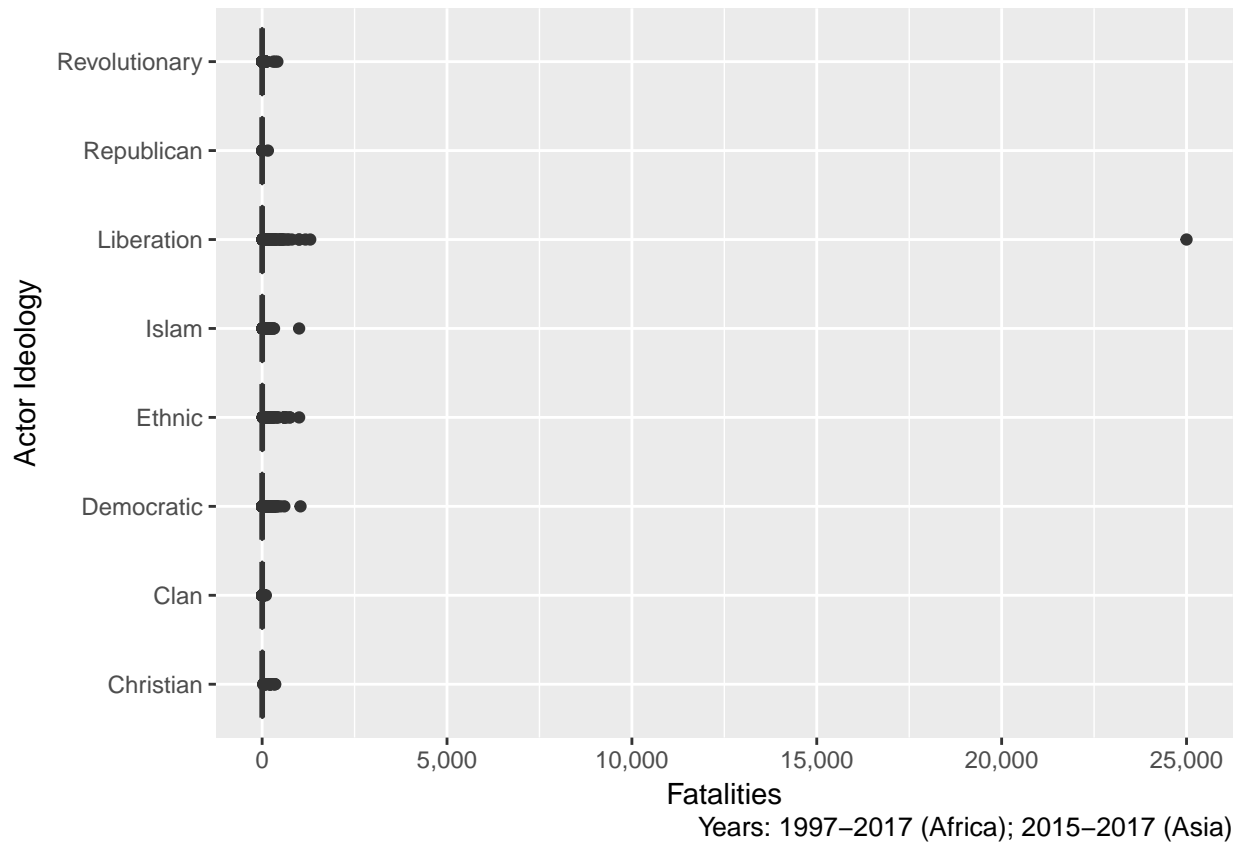
Limiting our results at 1400 to exclude this outlier, the boxplot shows very similar distributions across *Actor_Ideology*. Particularly large outliers exist for the *Liberation*, *Islam*, *Ethnic* and *Democratic* Ideologies.

Years: 1997–2017 (Africa); 2015–2017 (Asia)
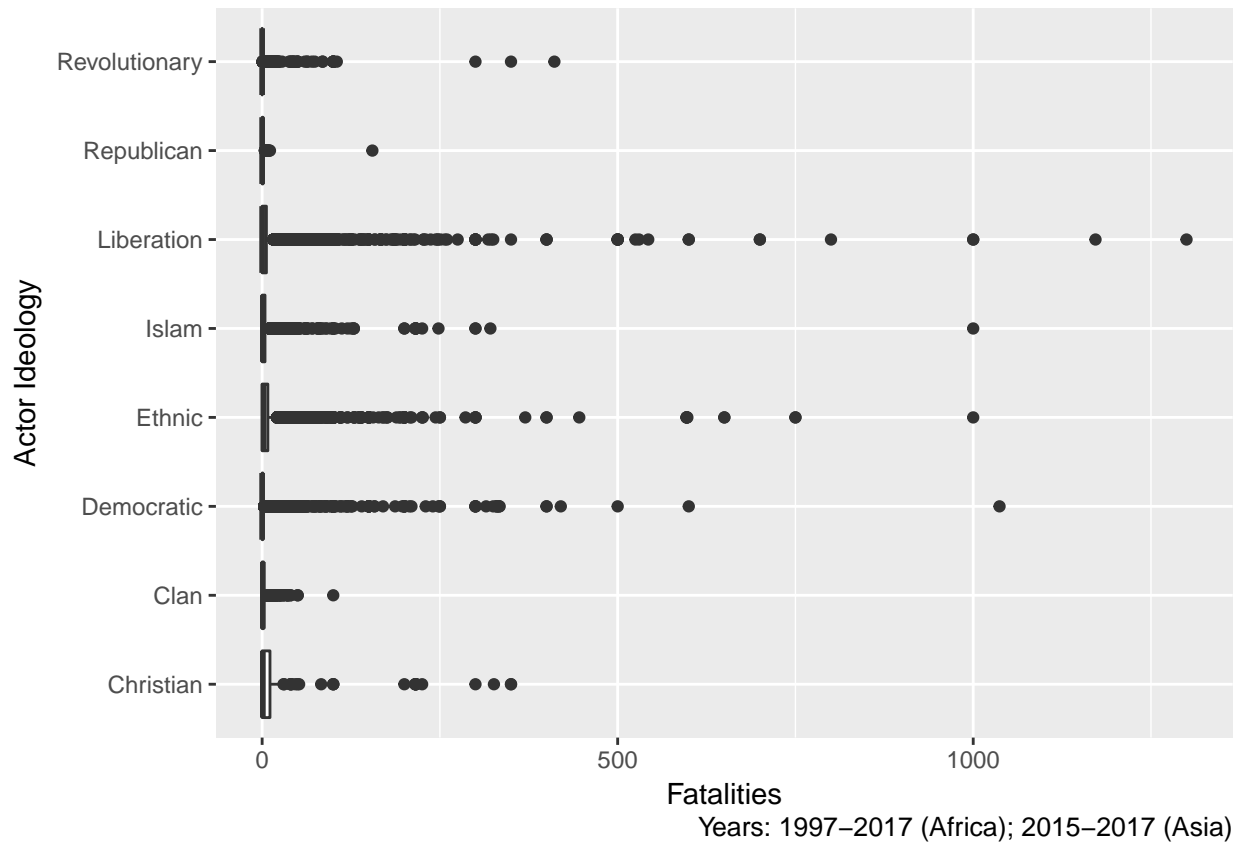
Taking an even closer look by limiting our boxplot to 30 fatalities, we can see that for the ideologies *Islam*, *Ethnic*, *Clan* and *Christian*, the median of fatalities is non-zero. In general, IQRs also vary across ideologies.

Years: 1997–2017 (Africa); 2015–2017 (Asia)
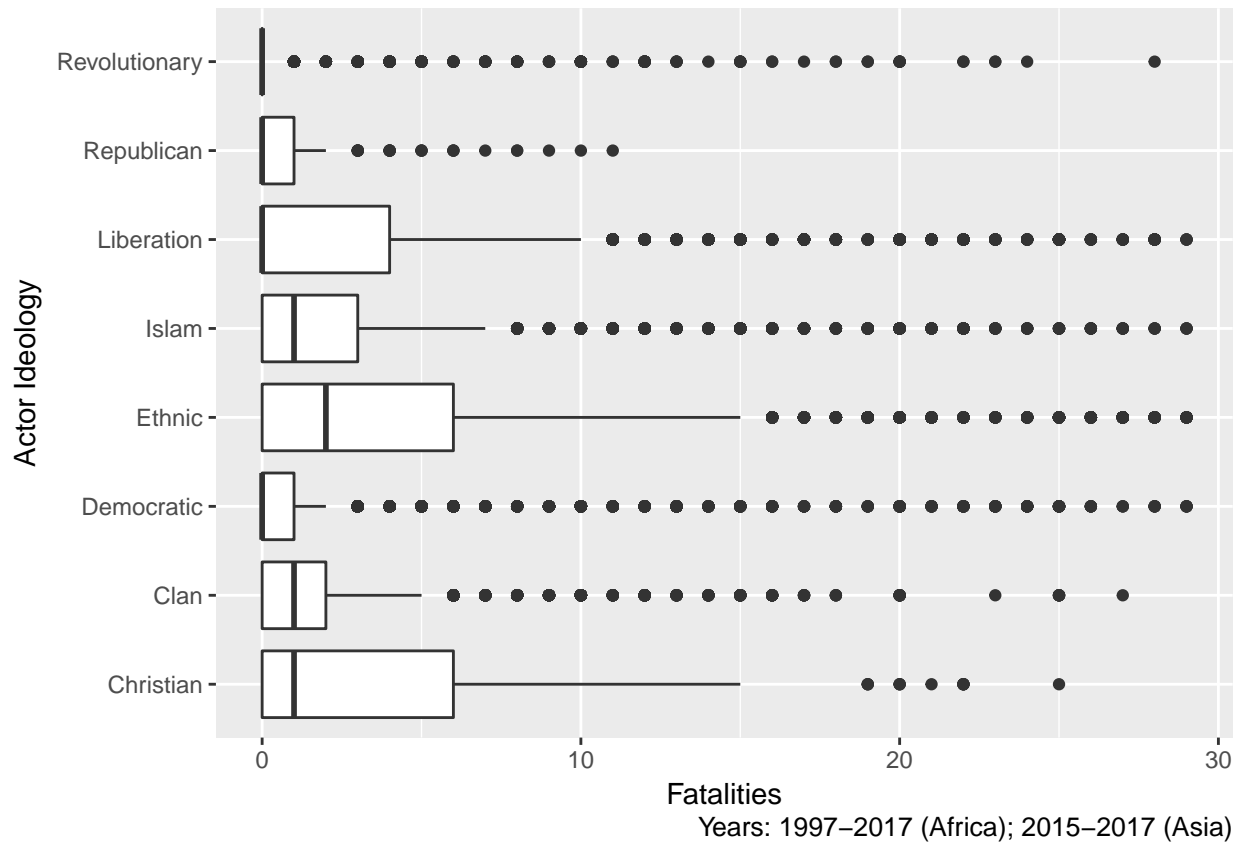
Looking at a barchart, we see that *Liberation* has most fatalities, but given the very large outlier of 25 000 fatalities in one single event and an overall median of 0, a clear correlation cannot easily be established. The correlations between the actor ideologies *Ethnic* and *Islam* seem more promising in this respect, because their data spreads around a non-zero median and outliers are not as large as in the case of Liberation and Democratic ideologies.

Years: 1997–2017 (Africa); 2015–2017 (Asia)
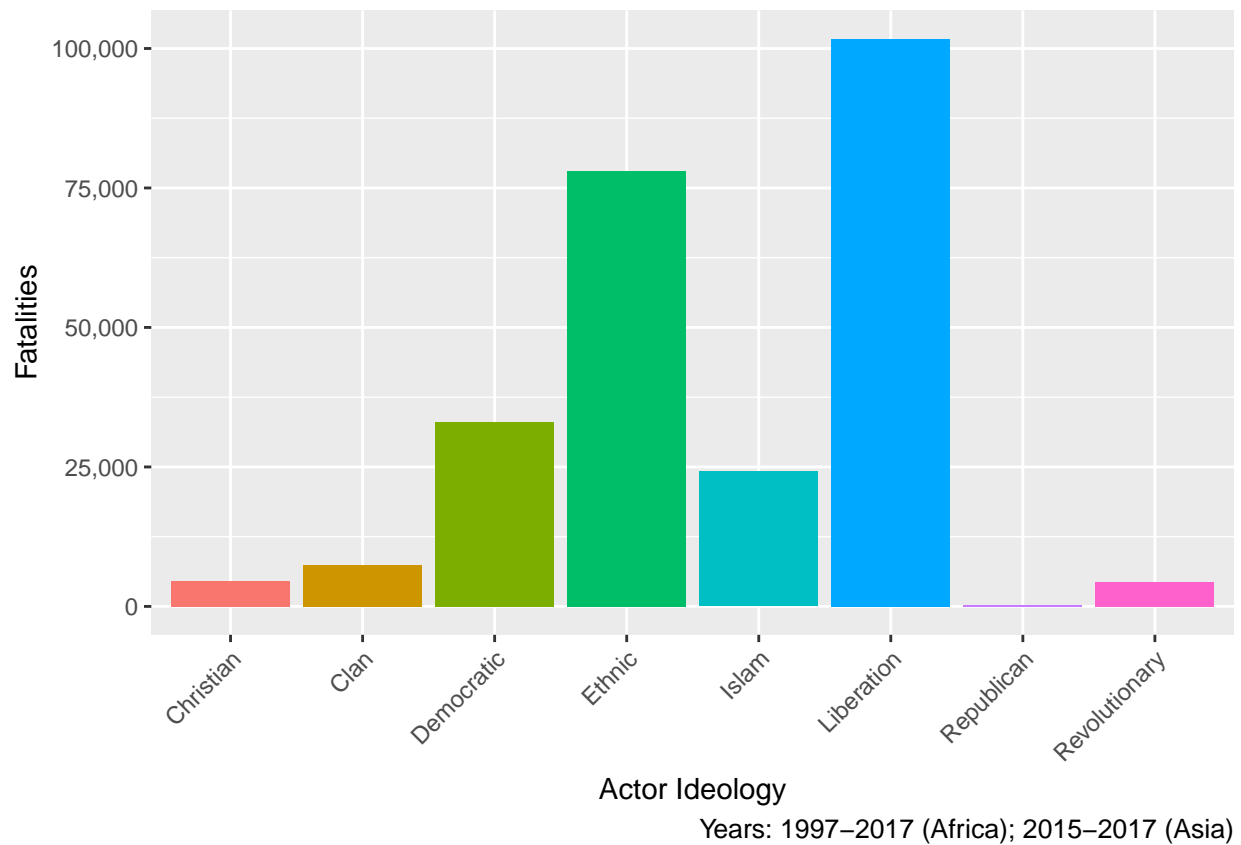
Time Series

We analzye the development of conflict incidents and their fatalities over time.

Total Number of Conflicts over time

The evolution of conflict incidences shows a clear upward trend. However, the sudden increase in 2015 is due to extra data for Asia only available after that time.

Years: 1997–2017 (Africa); 2015–2017 (Asia)
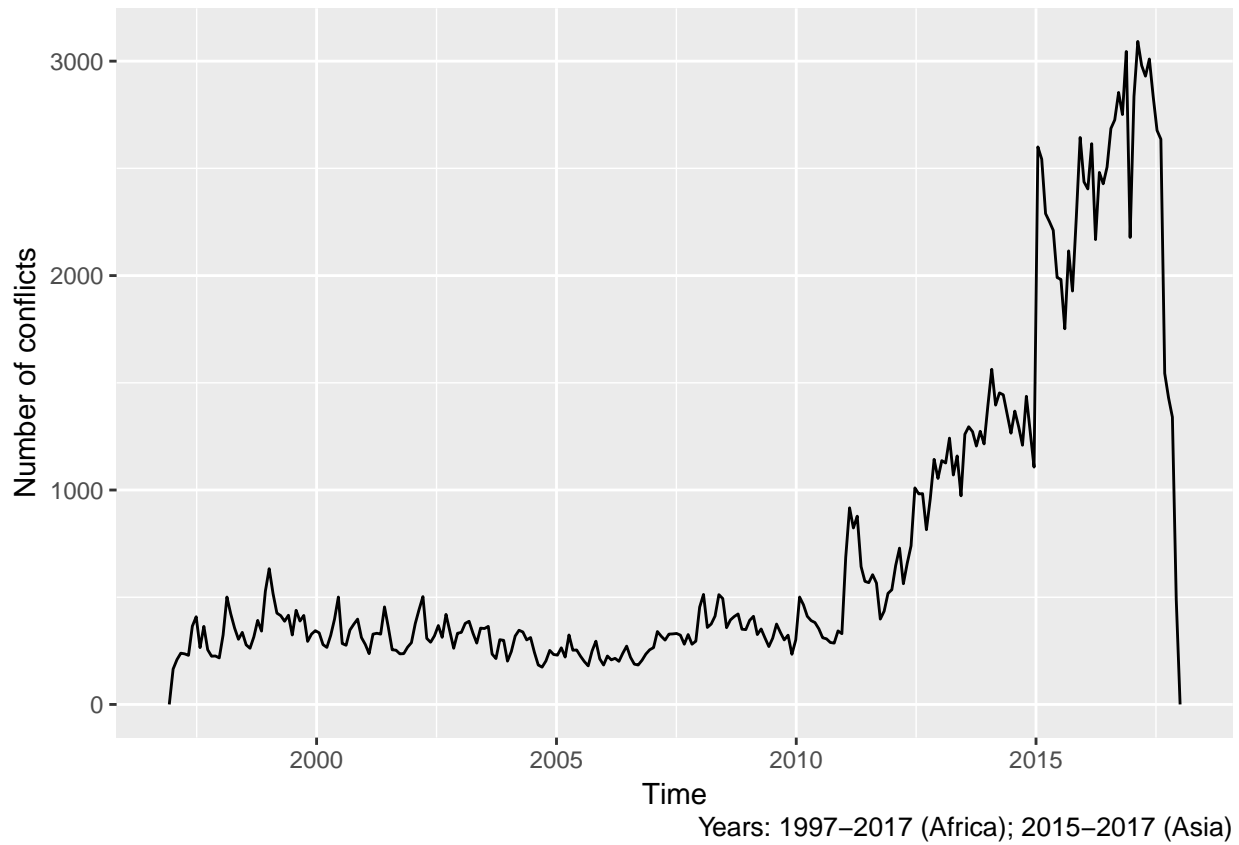
We take this into account by looking separately at the evolution in Africa and Asia. For Africa only, this yields the following time series (1997 - 2017) with similar upward trend as before. For Asia, we still see an upward trend, but not as strong as with the Africa data.

Years: 1997–2017 (Africa)

Years: 2015–2017 (Asia)
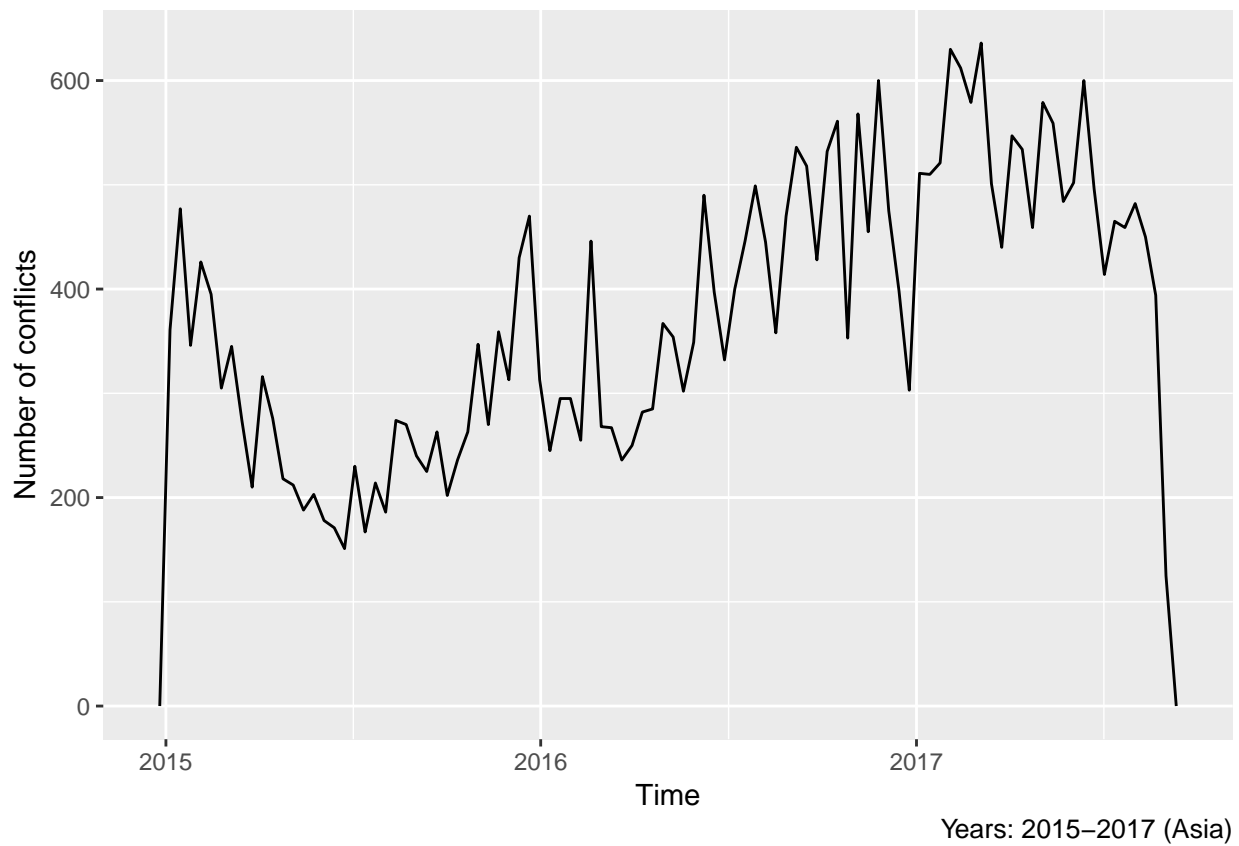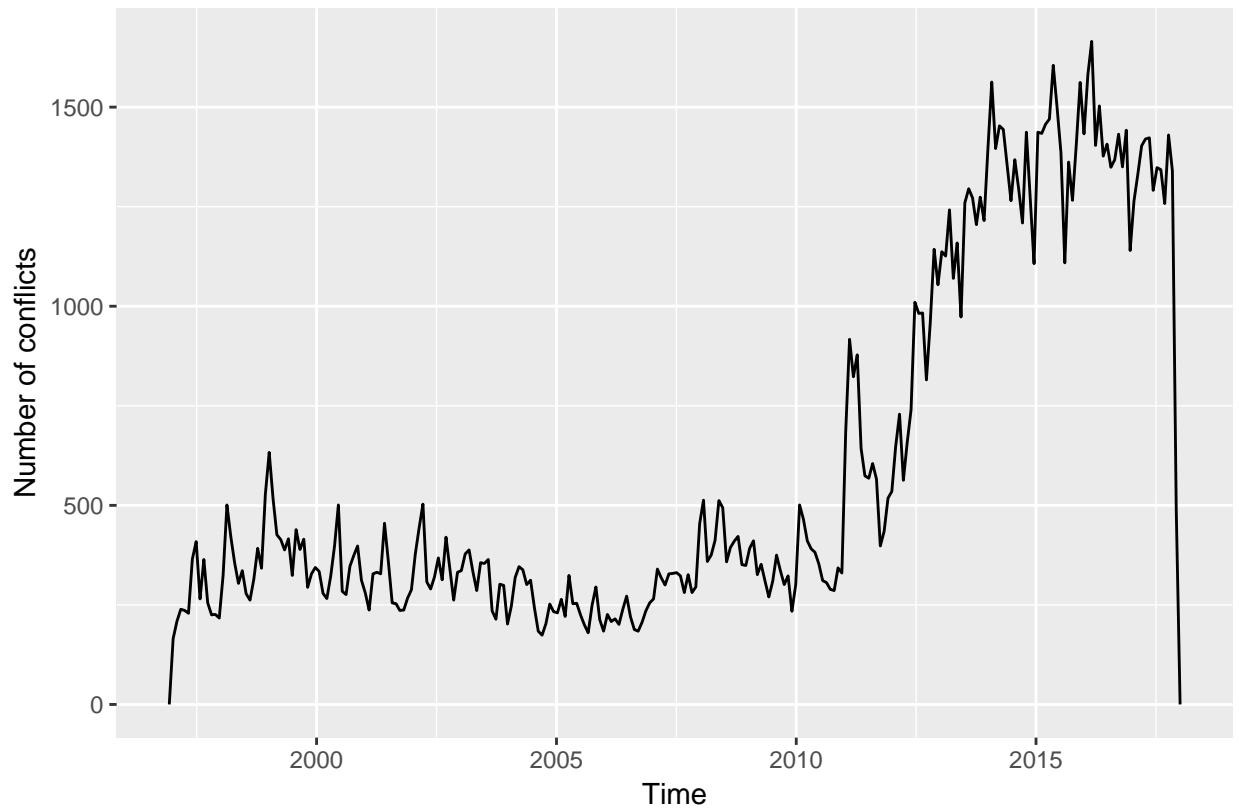
Number of Fatalities over time

Number of fatalities over time remain almost stable. While fluctuations from one year to the other remain relatively small, 1997 (Killing of 25 000 Hutu refugees in DRC) and 1999 (War Ethiopia against Eritrea) stand out with big sudden increases in fatalities.
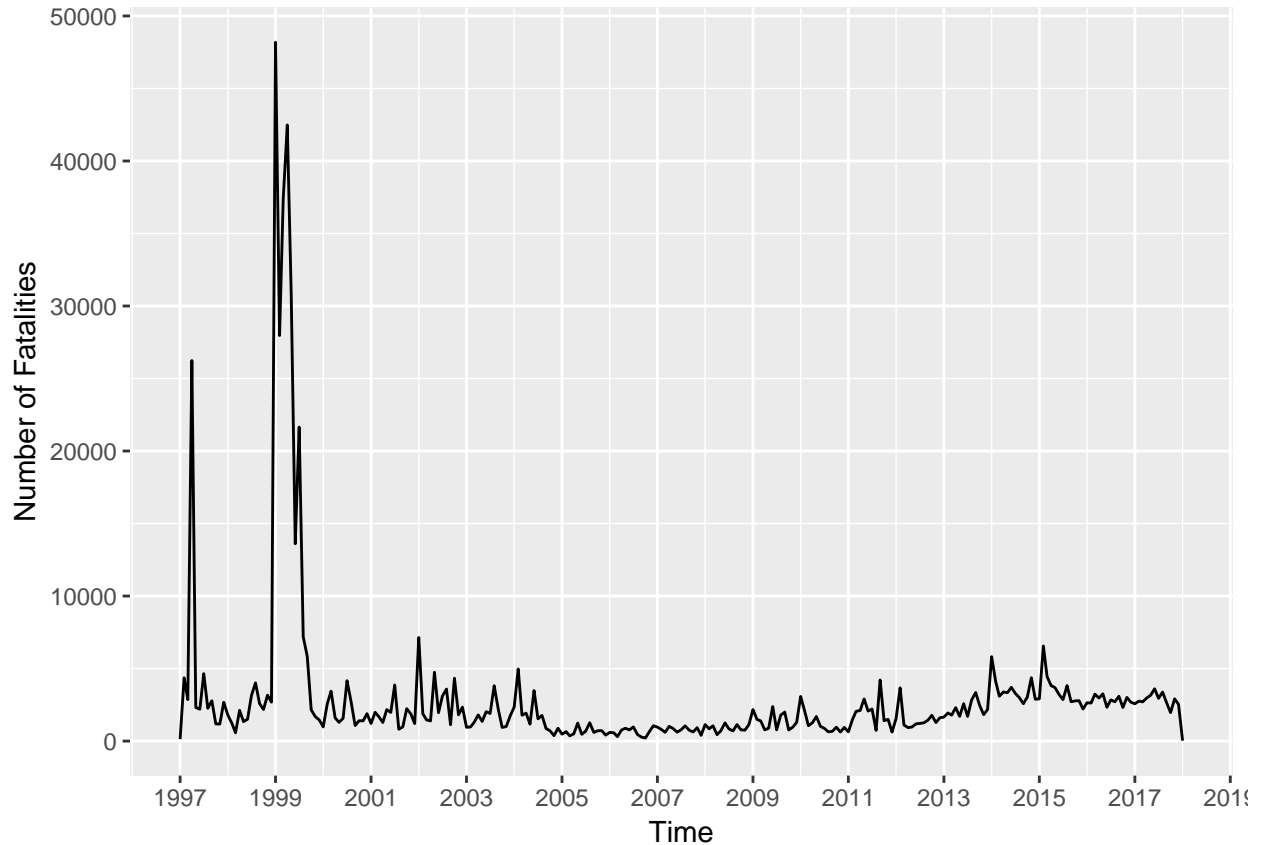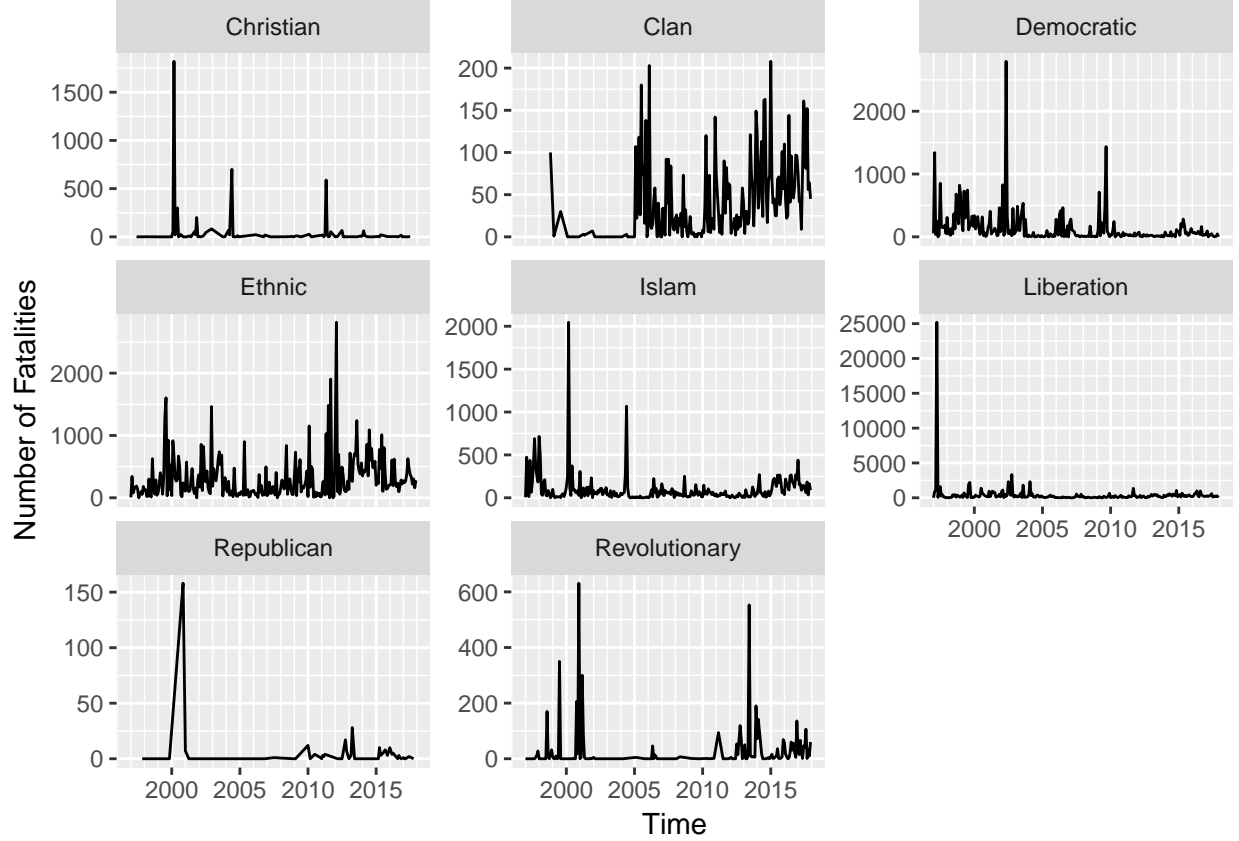

We also create an interactive version of this graph wit ggplotly() (see html output).

The following facet wrap plot gives an overview of the evolution of fatalities over time in the various regions:
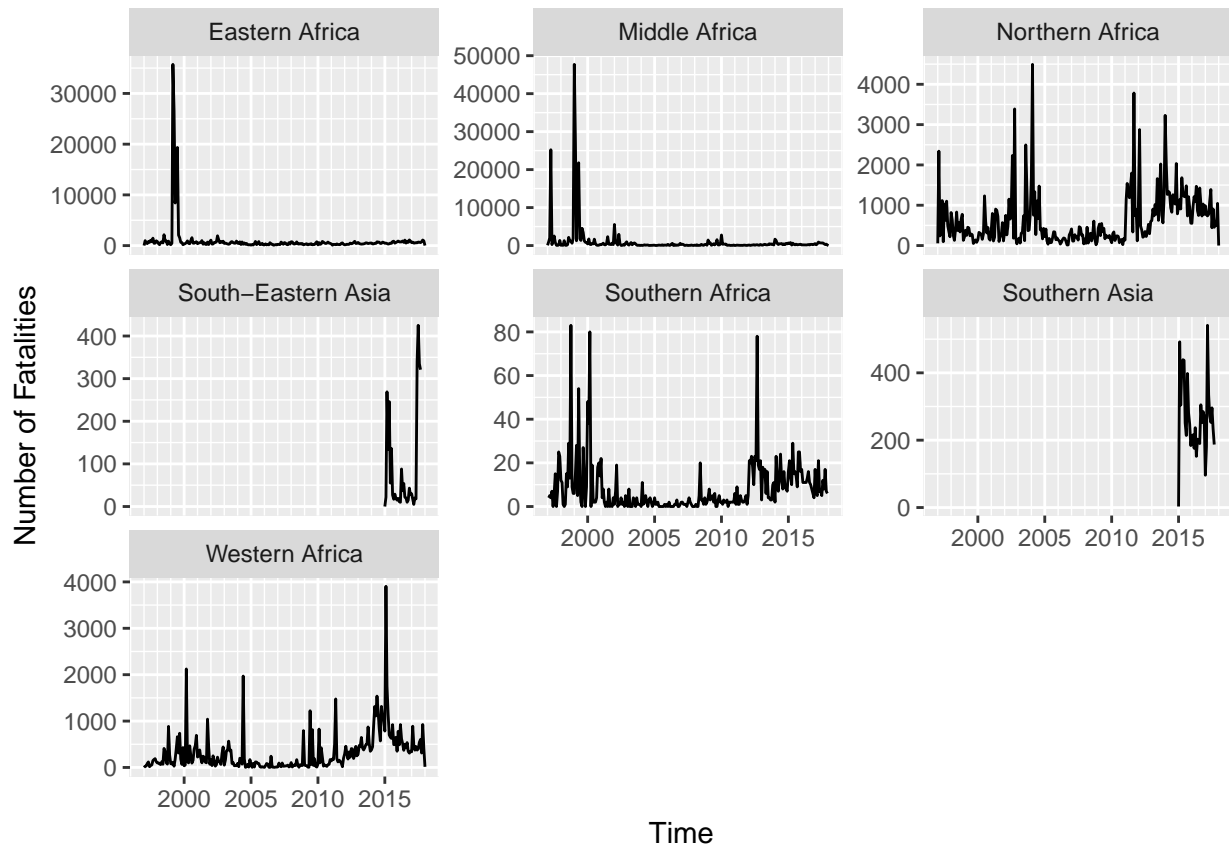
The following facet wrap plot gives an overview of the evolution of fatalities over time in the actor ideologies of interest.

Model

To test the relationship between the amount of fatalities and ideology of the actor(s) (involved), a 'Zero-inflated Negative Binomial model', i.e. a regression model for count data, is employed seeing that no values below zero are present (count data), a disproportional number of events have a death toll of zero,alongside it dealing better with overly dispersed data (seeing the great variance around the means). Because of computational limitations, we are not able to include additional controls, such as region, year or economic development.However, we compute separate models for Africa and Asia as well as model without big outliers, in order check for any obvious differences in the results.

As it turns out, all ideology coefficients are statistically significant. In the baseline scenario of no assigned ideology (Category: Others), any conflict is expected to lead to $\exp(\text{constant}) = \exp(2.107) = 8.35$ fatalities, holding all other factors constant. Looking at specific ideologies, it seems that in the case actors identifying with Christian ideology participate in a conflict, fatalities are $\exp(1.003) = 2.72$ times higher compared to a conflict without them, all else being equal. Moreover, ethnic ideology increases fatalities by a factor of $\exp(0,161) = 1.174$ and liberational ideology increases fatalities by a factor of $\exp(0.675) = 1.96$. All other ideologies show a negative correlation with *FATALITIES*. For example, identification with Islam is associated with a decrease of *FATALITIES* by a factor of $\exp(-0.560) = 0,57$, holding all other variables constant.

As a result, we conclude that there is a statistically significant effect of number of fatalities registered for groups with different ideologies.

Limitations

We recognise that our way of ascertaining the ideology of actors involved is far from perfect. It is more than likely that some actors may not include any of the terms which we identified, however they do espouse a

Table 3:

	Dependent variable:
	FATALITIES
Actor_IdeologyChristian	1.003***
	(0.175)
Actor_IdeologyClan	−1.366***
	(0.044)
Actor_IdeologyDemocratic	−0.078*
	(0.043)
Actor_IdeologyEthnic	0.161***
	(0.029)
Actor_IdeologyIslam	−0.560***
	(0.036)
Actor_IdeologyLiberation	0.675***
	(0.035)
Actor_IdeologyNone-Civilian	−0.660***
	(0.014)
Actor_IdeologyRepublican	−1.217***
	(0.230)
Actor_IdeologyRevolutionary	−0.313***
	(0.104)
Actor_IdeologyUnidentified	−1.536***
	(0.018)
Constant	2.107***
	(0.009)
Observations	387,712
Log Likelihood	−520,126.300
Note:	*p<0.1; **p<0.05; ***p<0.01

certain ideology.

There are ideologies which may have been left out of the analysis by omission. Further research would need to bring more robust theory in the construction of ideological categories.

Finally, our dependent variable may be confounded to an extent. For each violent event there are at least two actors involved. Our data shows fatalities per event, rather than who caused the fatalities. This means that we are not fully able to disaggregate the responsible perpetrators, and so the proxy in essence measures how many fatalities tend to happen when a certain actor is part of the event.