Final Project - Data Management with R

Shruti Gogia (166869); Vaishali Venkatesh Prasad (169435)

15 December 2017

Figure 1: Armed conflicts in Asia and Africa (2015-16)

Prerequisities

```r
library(readxl)
library(dplyr)
library(ggplot2)
library(ggmap)
library(knitr)
```

I. Introduction

In this report, we perform a comparative analysis of two continents using the Armed Conflict Location & Event Data Project (ACLED), a comprehensive source of information on political violence in developing countries (Raleigh et al. (2010)). The ACLED dataset is supported by funding from the Bureau of Conflict and Stabilization Operations of the United States Department of State, and the European Research Council. Spanning over 60 countries in the continents of Africa and Asia, the dataset contains information on violence from the year 1997 to present, and is updated real-time. The data is available for free download on the website: https://www.acleddata.com/data/

We are interested in examining the data on political violence in Africa and Asia for two years, 2015 and 2016. A politically violent event is defined as the "use of force by a group with a political purpose or motivation…a single altercation where often force is used by one or more groups to a political end" (ACLED, 2015).

In particular, we seek to answer questions such as the composition of politically violent events across the two continents, the aggregate number of fatalities, and the relationship between event type and the average number of fatalities. Further, we attempt to visually map the occurrence of politically violent events across the two continents and the actors involved in the event. We also run regression analyses in order to understand the differential impacts of variables such as the actor and event type on the number of fatalities across both continents.

The rest of the paper is divided as follows. Section II discusses the data cleaning techniques employed in this paper, Section III outlines the results of the preliminary analysis of the data. Section IV outlines the interactive data visualization techniques used, in particular Plotly , while Section V discusses the regression analysis conducted. Section VI concludes.

II.Data Cleaning

STEP 1: Importing Files

The datafiles are available in CSV format and are categorized according to continent. We imported the datafiles for Africa and Asia into R, and further subsetted the data so as to include only the years 2015 and 2016 in the analysis.

We also created a variable `continent` which takes two values: `Africa` and `Asia`. We then created a final dataset by horizontally merging both continents' datasets.

STEP 2: Check for missing values

The next step is to check for missing values in the variables of interest. We found that some values of the variables `LATITUDE` and `LONGITUDE` were missing, and we dropped these values as we would be utilizing the geolocation information in our spatial visualizations.

STEP 3: Recoding values within variables

We then recoded some values of the variables `EVENT_TYPE`, `ACTOR_TYPE1` and `ACTOR_TYPE2`. For example, two values of the variable `EVENT_TYPE` were "RIots/Protests" and "Riots/Protests". We recoded the former value into the latter, so as to avoid duplication of responses.

STEP 4: Transforming variables

We also transformed the variable `YEAR` as a character variable, and renamed the variables `INTER1` and `INTER2` as `ACTOR_TYPE1` and `ACTOR_TYPE2` respectively.

We then created two new variables `ACTOR_NAME1` and `ACTOR_NAME2` from `ACTOR_TYPE1` and '`ACTOR_TYPE2`, which were basically converting them from numerically coded variables to the names on the actor. For example, from ACTOR_TYPE1, 1 became "Government or mutinuous force"

The final dataset after cleaning contains 27 variables and 59,438 observations across two continents (Asia and Africa) and two years (2015 and 2016).

III.Preliminary Analysis

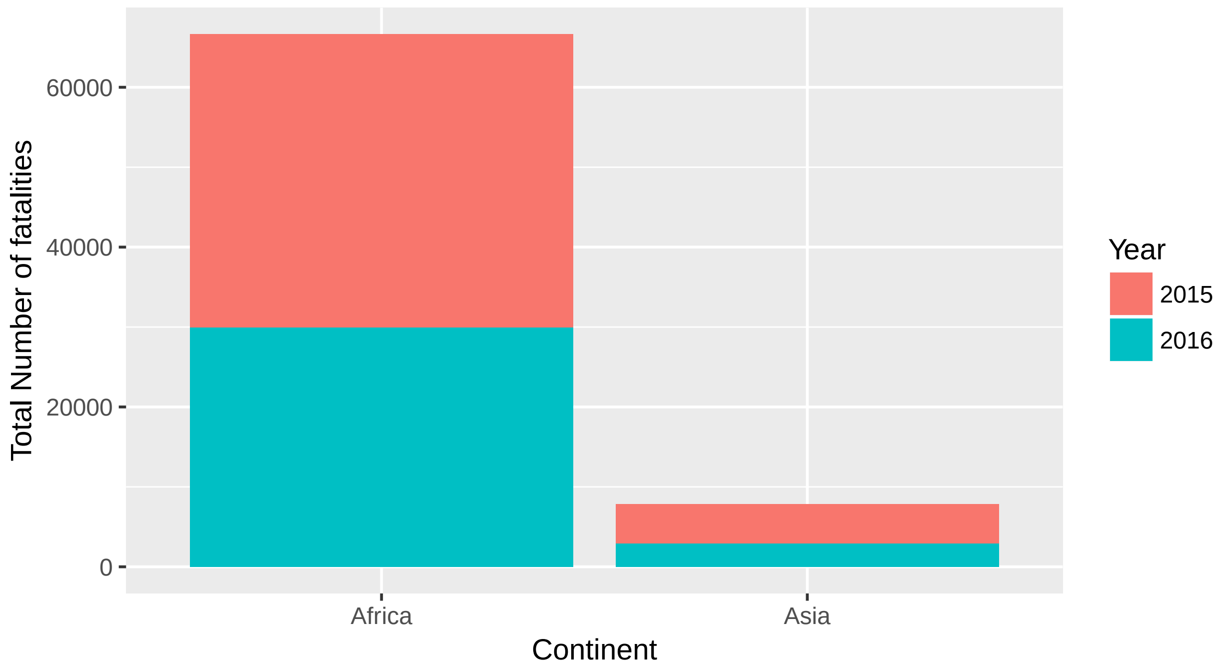
In this section, we outline the preliminary visualization of the data conducted using the `ggplot2` and the `ggmaps` packages. The objectives of this analysis is to understand the patterns of distribution of the main variables of interest, in particular the total number of fatalities by continent, average number of fatalities by the type of actor and event, and a geographic representation of fatalities by actor type (using `ggmaps()`). This also helps us in defining the dependent and independent variables for the regression analysis discussed in Section V.

The preliminary analysis has been conducted under the categories of number of fatalities, analysis by actor type and analysis by event type

Number of fatalities

Figure 1 displays the total number of fatalities across both continents in the years 2015 and 2016. The dataset clearly has much more information on African events of political violence than Asian – this could either be reflective of real life events, or a limitation of the dataset in itself. In both continents however, the number of fatalities is larger in the year 2015 than 2016.

Figure 1: Total number of fatalities by continent, across two years

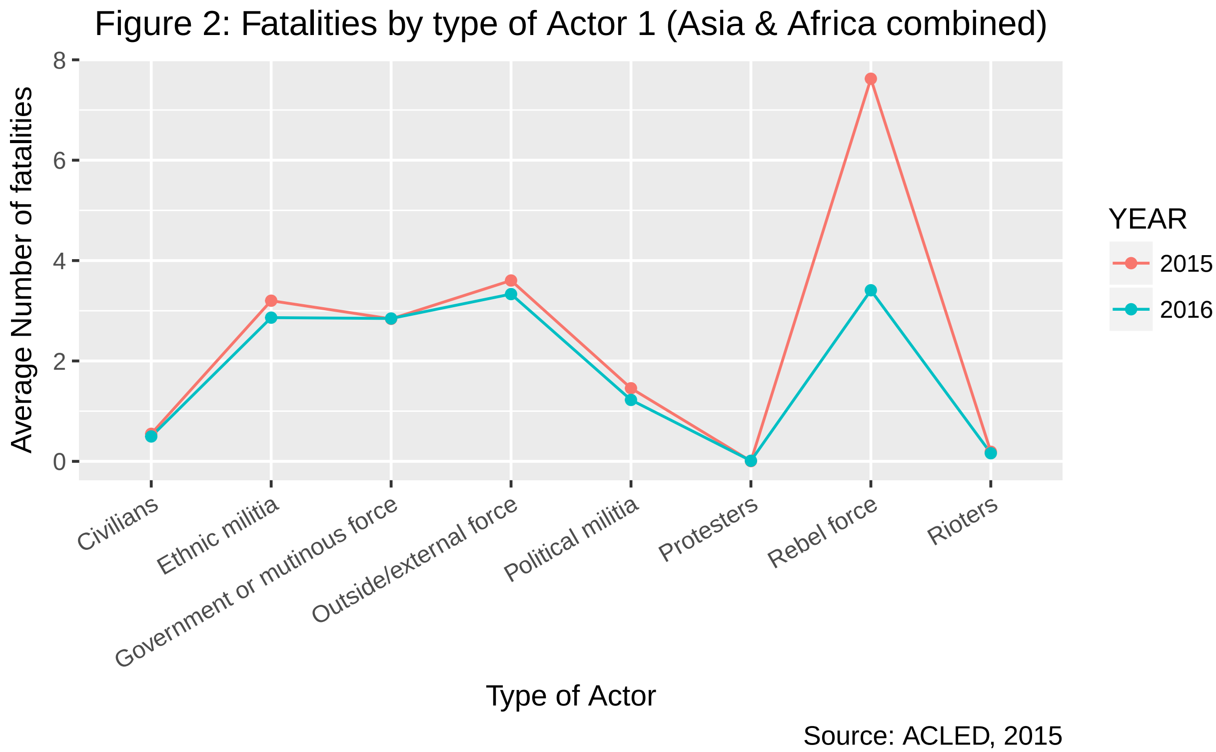


Source: ACLED, 2015

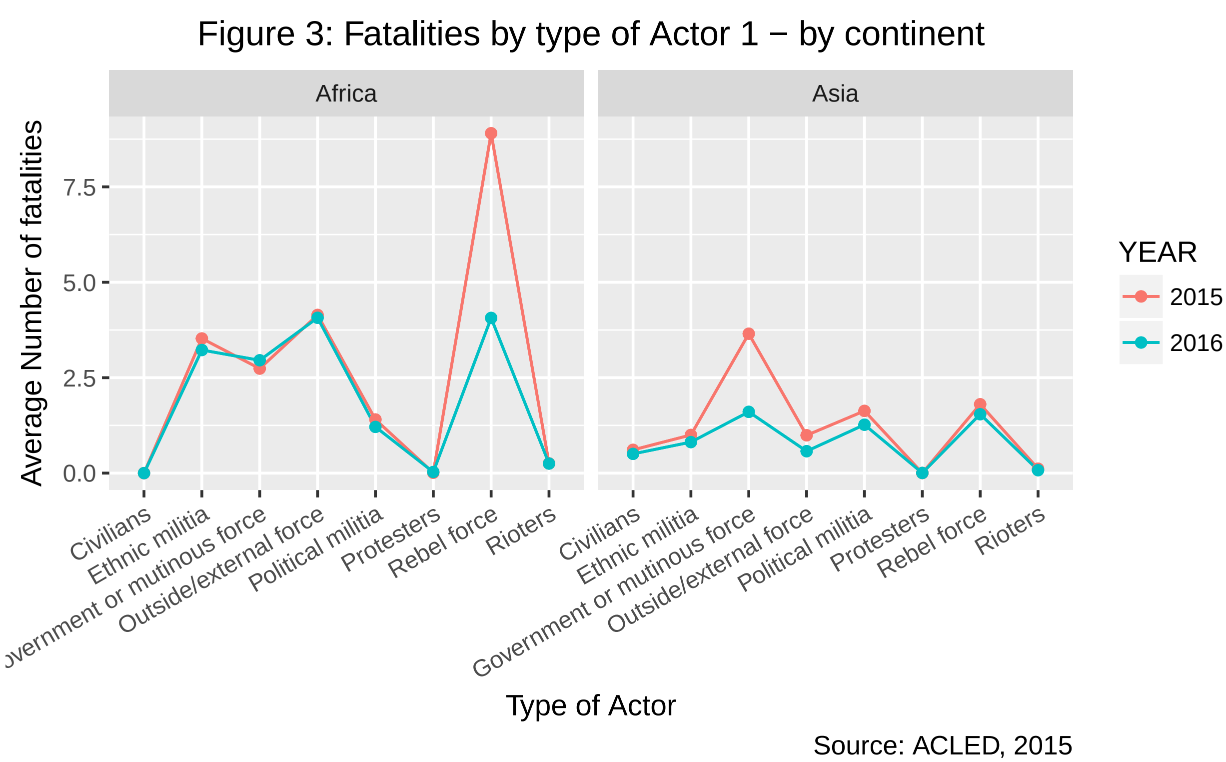
Analysis by actor type

`Actor 1` and `Actor 2` are two variables recording the named actors involved in the event. According to the ACLED coding methodology, politically violent actors include governments, rebels, ethnic groups, rioters and civilians, among others. The dataset also contains information on the interaction between actors, for instance the interaction code `12` representing the interaction between Military (`Actor 1`, coded as `1`) and Rebels (`Actor 2`, coded as `2`).

In this exploratory analysis, we plot the average number of fatalities by each actor type disaggregated by continent. Figure 2 demonstrates the overall average fatalities by each type of Actor 1, disaggregated by year. This shows that in the year 2015 on average, rebel forces were involved in scenarios with the maximum number of fatalites, which is around 8 fatalities per politically violent event. In 2016 however, the average number of fatalities dropped to around 4 for rebel forces.



This pattern is largely derived from the data on Africa, as is demonstrated in Figure 3, which disaggregates the average number of fatalities by Actor 1 across both continents and years. Figure 3 also demonstrates that apart from Rebel forces in Africa and Government/mutinous forces in Asia, the trends of average fatalities by Actor 1 remains similar across both years.



Similarly, Figures 4 and 5 display the average number of fatalities by Actor 2 overall and disaggregated by continent respectively.

Figure 4: Fatalities by type of Actor 2 (Asia & Africa combined)

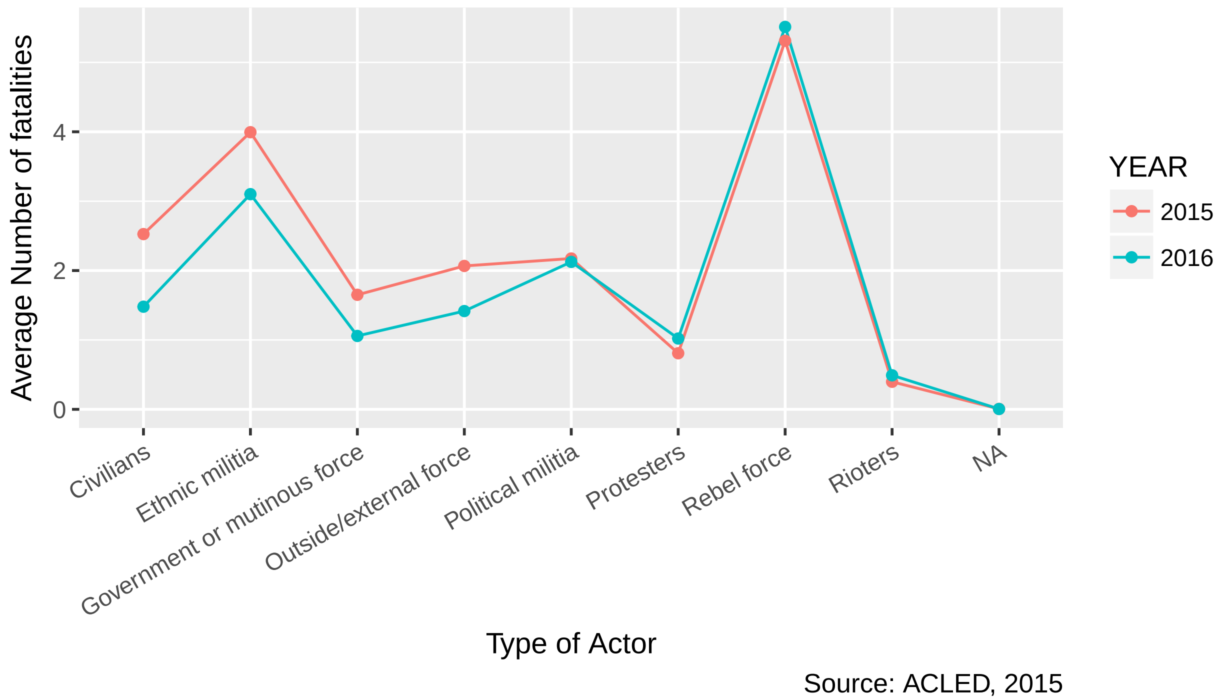
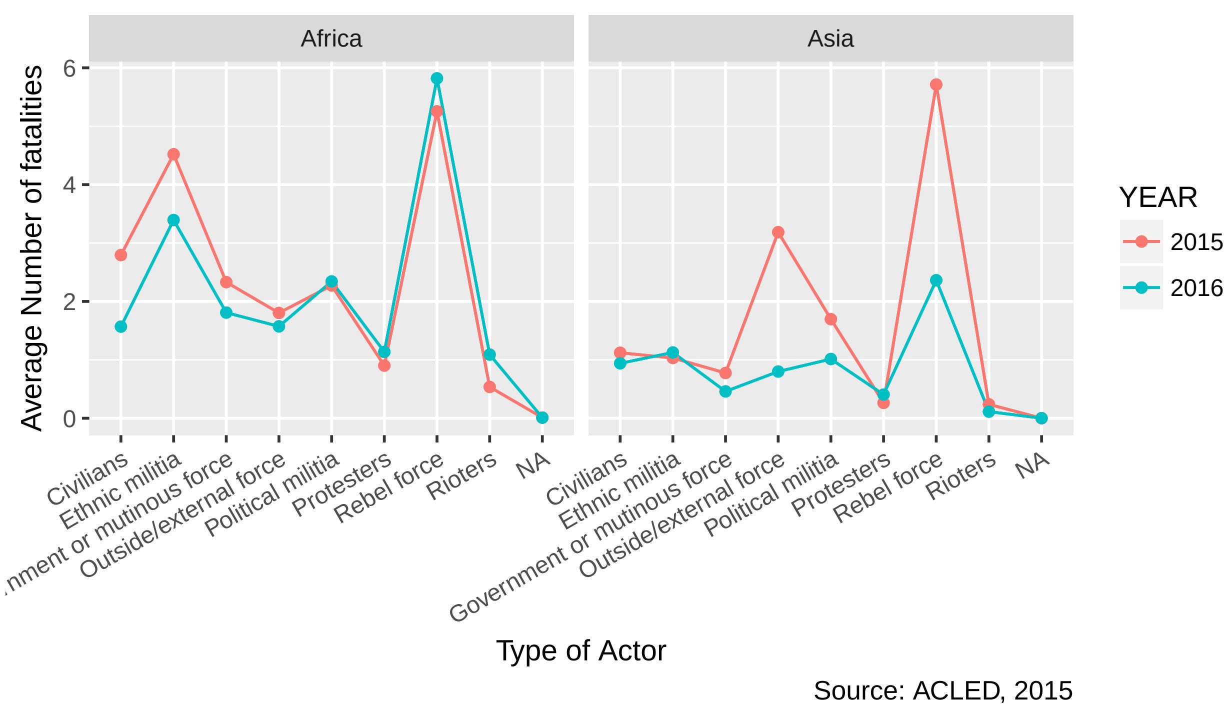


Figure 5: Fatalities by type of Actor 2 − by continent



Analysis by event type

The (politically violent) event is the main unit of observation in the dataset (ACLED Codebook, 2017) and occur between different actors types such as civilians and the government. 9 types of (violent and non-violent) events are coded under the variable `EVENT_TYPE` which include "Violence against civilians", "Strategic development" and "Riots/Protests" among others.

Figure 6 demonstrates the overall proportions of the occurrence of different type of events in both continents.This shows that the most occuring event type is that Riots/Protests, followed by a Battle with no change in territory.

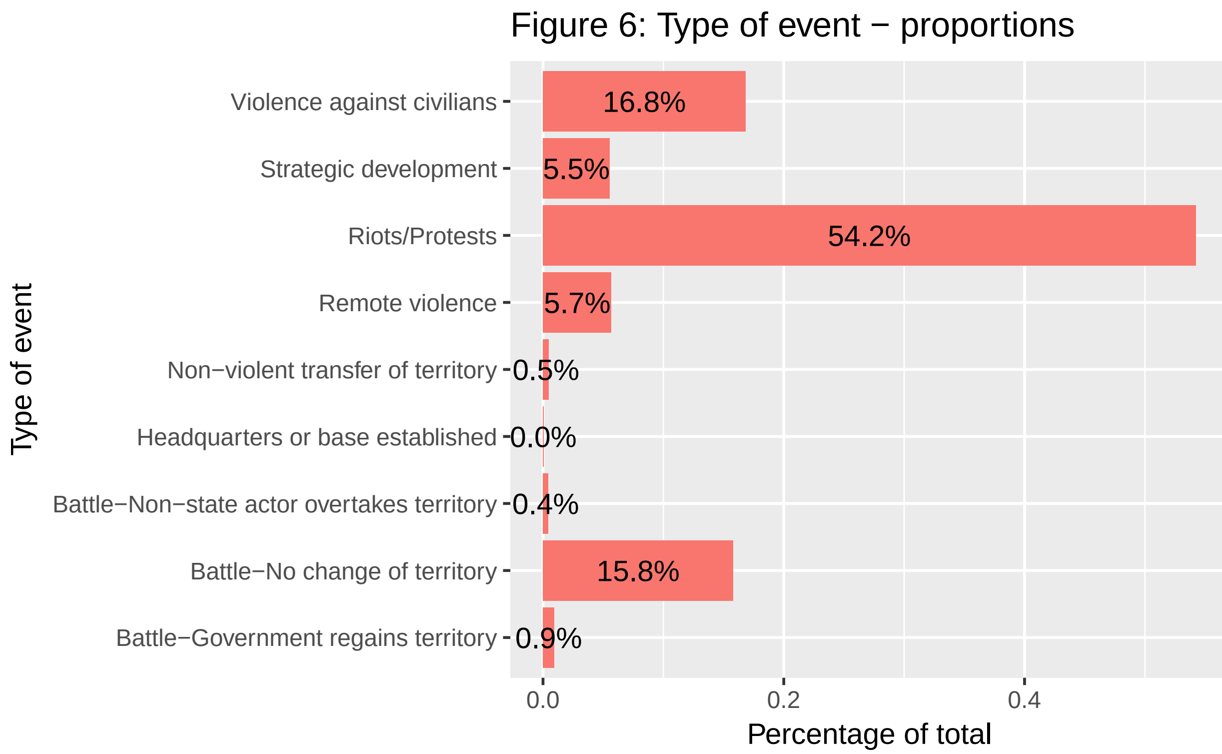
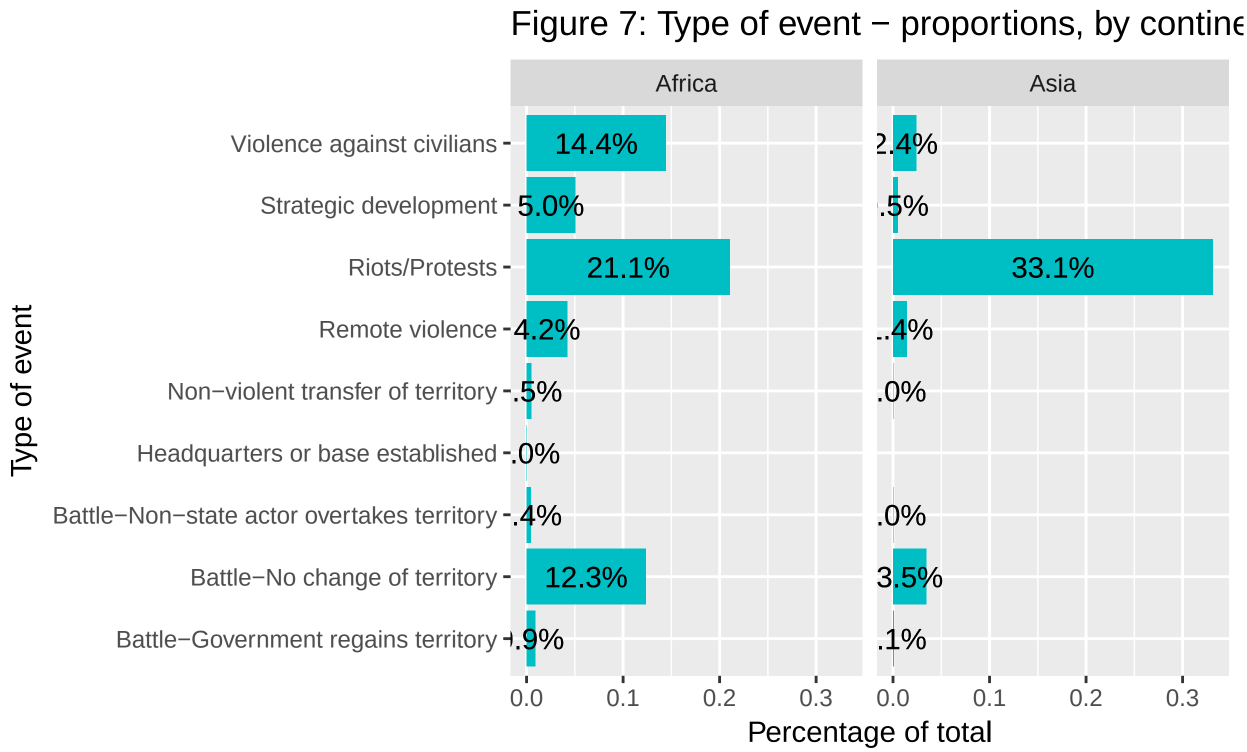
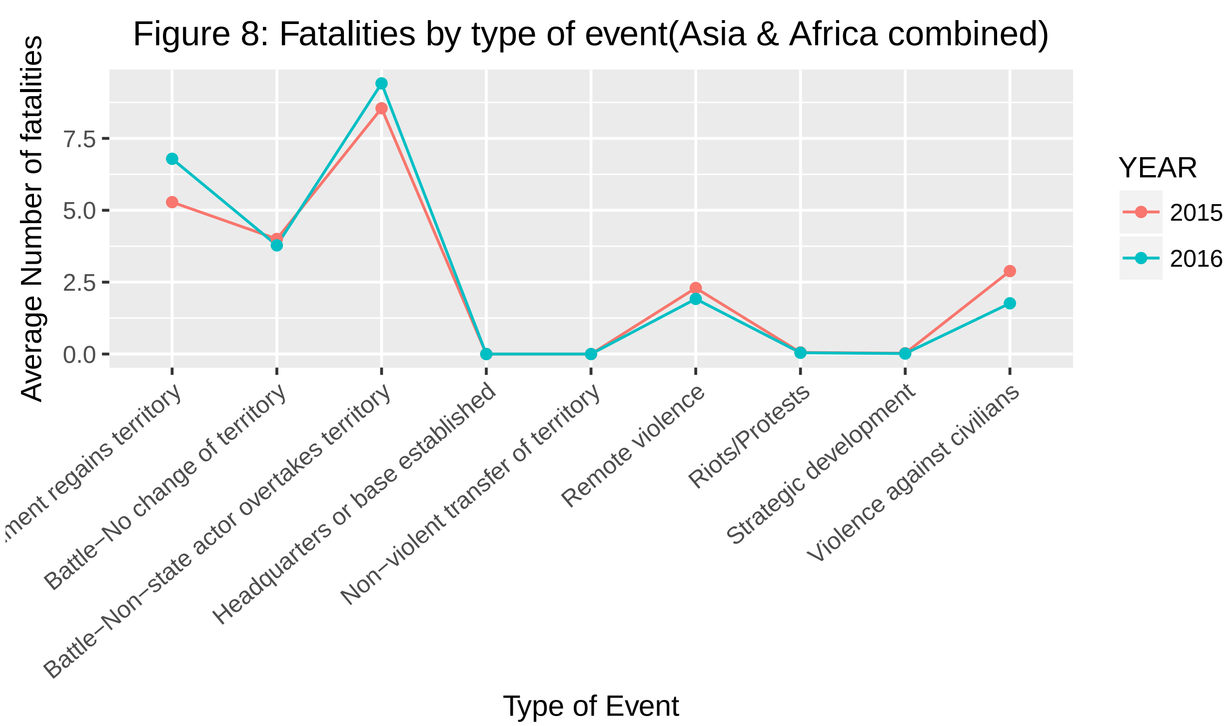


Figure 7 demonstrates the proportion of event types by continent. In both continents, Riots/Protests remain the most frequent event type. However in Africa, violence against civilians comprises a greater proportion of the total number of events than in Asia.



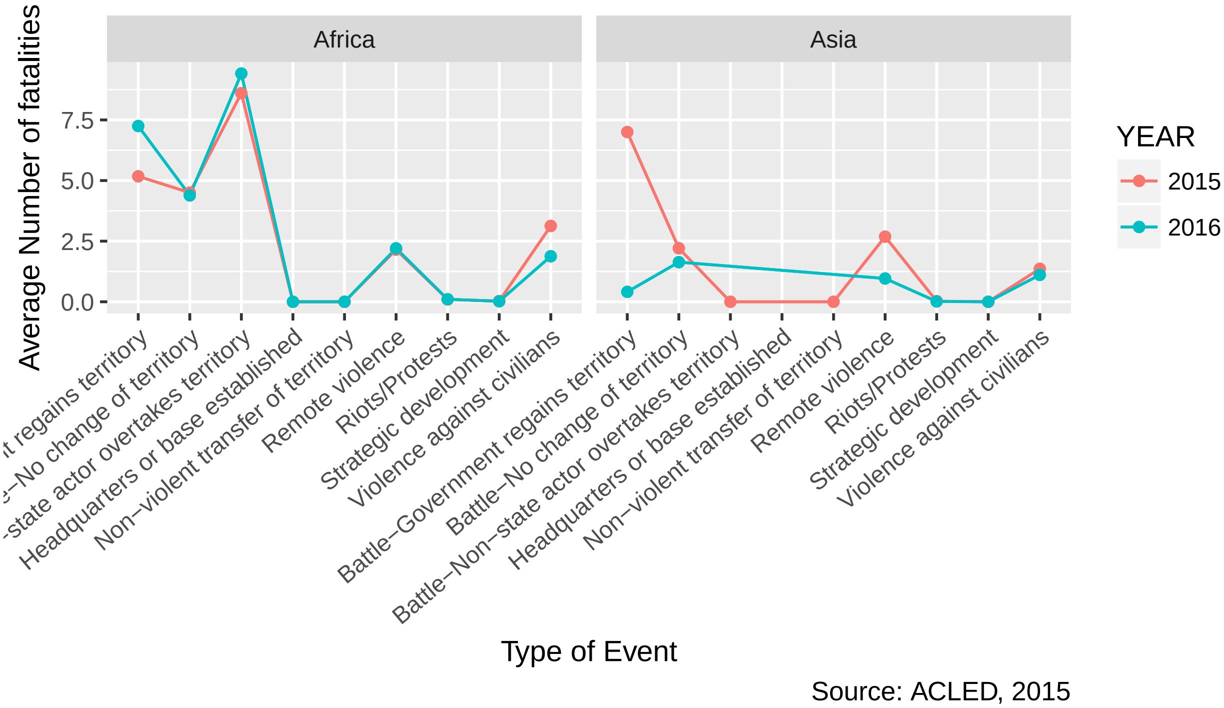
Figures 8 and 9 display the average number of fatalities by event type on the whole and across continent respectively. On the whole, the maximum average fatalities is caused when non-state actors overtake territory (around 8 fatalities), a pattern which remains fairly constant across both years. However it is more useful to disaggregate fatality data by continent given that data from Africa is much more than from Asia.



Source: ACLED, 2015

Figure 9 thus shows that in the case of Asia, maximum (average) fatalities is caused in events of government regaining territory, particularly in the year 2015. Both continents thus demonstrate significantly different trends of number of fatalities by event type.

Figure 9: Fatalities by type of event



IV.Interactive Visualization

Geographic visualization

In this section, we describe a spatial visualization of political events and associated actors in Africa and Asia. This visualization has the advantage of enabling one discover "clusters" of violent events, as well as a spatial distribution of different actors on the ground.

The maps are created with the `ggmaps()` package (Kahle and Wickham (2013)) and can be seen in the form of an interactive dashboard which have been plotted using the plotly function and thus allow the user to find the details of each conflict plotted on the map by navigating with a mouse pointer (Please find attached the rmd file for the interactive dashboard.) The dashboard has four maps, as follows:

Map 1 graphs the distribution of politically violent events across Asia. As indicated by Figure 7, the event types of maximum proportion in Asia are "Riots/Protests", "Battle - no change of territory" and "Violence against civilians" respectively. As shown in Map 1, countries such as Pakistan have witnessed many events of battle and violence against civilians. Bangladesh and Myanmar also display similar clusters of battles and violent events against civilians.

Map 2 displays the distribution of actors involved in political events across Asia. On the ground, protestors and rioters form a large portion of actors involved, especially in Sri Lanka and parts of North India.

Map 3 displays the distribution of fatalities across Africa and is particular useful in getting a sense of availability of fatality information for the continent. For instance, data on fatalities for Libya, Algeria and Egypt is largely concentrated in the North, while in Zimbabwe and Madagascar, it is 'spread' all over the country.

Finally, Map 4 contains the distribution of actors involved in political events across Africa. This shows that protestors are a main type of actor in Northern Algeria, while Governments/mutinous forces are more prominent in Northern Libya.

V.Regression Analysis

As can be seen from the graphs in the preliminary analysis, there is a significant amount of variation in the occurence of event types within the two continents. Further, there is also a huge difference between the number of fatalities related to different event types.Thus, the aim of the regression to find out if any event type has a significant relationship in singnificantly increasing the number of fatalities in Africa and Asia. For this purpose, we conduct multiple linear regressions. The first set of regressions has been conducted for the overall data. The first model regresses `EVENT TYPE` on `FATALITIES` and the second model conducts the same regression while controlling for continent. The regression results can be found in Table 1 in Appendix.

```r
reg1 <- lm(FATALITIES~ EVENT_TYPE, ACLED_Final, na.action = na.exclude)
summary(reg1)

reg2 <- lm(FATALITIES~ continent+EVENT_TYPE, ACLED_Final, na.action = na.exclude)
summary(reg2)
```

Then linear regression models were conducted only for Africa. Again there were two models - Model 1 only factoring `EVENT TYPE` and Model 2 controls for the effect of `COUNTRY`. The regression results can be found in Table 2 in the Appendix.

```r
#Africa regression
reg3 <- lm(FATALITIES~ EVENT_TYPE, AfricaMaps, na.action = na.exclude)
summary(reg3)

reg4 <- lm(FATALITIES~ EVENT_TYPE+ COUNTRY, AfricaMaps, na.action = na.exclude)
summary(reg4)
```

The final set of regressions were conducted only for Asia. The models were same as for Africa (Table 3, Appendix).

```r
#Asia regression
reg5 <- lm(FATALITIES~ EVENT_TYPE, AsiaMaps, na.action = na.exclude)
summary(reg5)

reg6 <- lm(FATALITIES~ EVENT_TYPE+COUNTRY, AsiaMaps, na.action = na.exclude)
summary(reg6)
```

VI. Discussion and Conclusion

The results from preliminary analysis show riots/protests to be the highest occuring events (Plot 6 & 7) in both Asia and Africa. However, the preliminary analysis also shows that when the number of fatalities are plotted against event type, overall riots/protests record the lowest number of fatalities. The highest fatalities are recorded for the events when non-state actor takes over a territory. Further, when the number of fatalities are plotted against event type separately for the two continents, some interesting findings emerge. The first one being that Africa has highest fatalities by non-state actors, while Asia records higher fatalities when government regains territory. Secondly, riots/protests maintain lowest fatalities in both continents. Also there is a significatly higher number of fatalities caused by events of violence against the civilians.

Thus, the regression analysis has been conducted to see if the event types bear a significance of the number of fatalities caused in Africa and Asia. The results from the multiple regression analysis are reported in the Appendix.

Table 1 shows the results of overall data for Asia and Africa. The results show that Asia has significantly lower number of fatalities than Africa.Further, when a non-state actor takes over a territory, everything else being equal, the number of fatalities significatly rise by 2.713 units. For all other event types, everything else being constant, the number of fatalities are always less, when compared to non-state actor events.

Table 2 shows regression results only for Africa. The patterns are very similar to the overall trends. The number of fatalities is the highest due to non-state actor events, as compared to other events.

Lastly, Table 3 shows regression results for Asia. The table indicates that the number of fatalities, everything else being constant, are significantly less than when government takes over territory.

The results from these analyses highlight the relationship between loss of human lives in various turbulent zones in Asia and Africa. In-depth analysis along these lines can be used to predict patterns and thus aid state preparation to deal with conflicts in ways that can minimize loss of human life.

However, there are some ***limitations*** to the present data. Primarily, the data between the two continents in not proportional and data from Africa predominates the overall data. Secondly, the data reported for Asia is not complete and only reports data for some countries, leaving out some hotspot areas of regional violence, like Afghanistan, in Asia. Including such data in Asia might lead to different results. Lastly, the number of fatalities reported in the data are only from various secondary and media sources and thus not primary data, leading to speculation regarding its authenticity. The results should be interpreted keeping these limitations in mind.

Appendix

Table 1: Regression Table for Asia and Africa combined

Dependent variable: Fatalities

Variable	Model 1	Model 2
Continent		
Asia		-0.570***
		(0.07)
Event Type (a)		
Battle - No Change of trty	-2.268***	-2.180***
	(0.33)	(0.33)
Battle-Non-state actr wins trty	2.747***	2.713***
	(0.58)	(0.58)
Headquarters or base established	-6.164***	-6.201***
	(1.54)	(1.54)
Non-violent transfer of trty	-6.164***	-6.197***
	(0.56)	(0.56)
Remote Violence	-4.052***	-3.944***
	(0.35)	(0.35)
Riots/Protests	-6.115***	-5.803***
	(0.33)	(0.33)
Strategic Development	-6.142***	-6.127***
	(0.35)	(0.35)
Violence against civilians	-3.840***	-3.796***
	(0.33)	(0.33)
Intercept	6.164***	6.201***
	(0.32)	(0.32)
N	59436	59436
Adjusted R- squared	0.044	0.044

Collinearity: (a) Battle - Government regains territory excluded

*p < 0.05; **p < 0.01; ***p < 0.001

Table 2: Regression Table for Africa

Dependent variable: Fatalities

Variable	**Model 1**	**Model 2**
Event Type (a)		
Battle - No Change of trty	-1.937***	-1.194**
	(0.44)	(0.44)
Battle-Non-state actr wins trty	2.563***	3.453***
	(0.75)	(0.74)
Headquarters or base established	-6.383**	-5.331**
	(1.97)	(1.96)
Non-violent transfer of trty	-6.383***	-5.265***
	(0.73)	(0.72)
Remote Violence	-4.207***	-3.424***
	(0.47)	(0.47)
Riots/Protests	-6.283***	-5.933***
	(0.44)	(0.44)
Strategic Development	-6.359***	-5.305***
	(0.46)	(0.47)
Violence against civilians	-3.878***	-3.354***
	(0.44)	(0.44)
Intercept	6.382***	5.534***
	(0.43)	(0.56)
N	35072	35072
Adjusted R- squared	0.036	0.051

*p < 0.05; **p < 0.01; ***p < 0.001 NOTES: 1. Collinearity: (a) Battle - Government regains territory excluded; 2. Model 2 controls for COUNTRY

Table 3: Regression Table for Asia

Dependent variable: Fatalities

Variable	Model 1	Model 2
Event Type (a)		
Battle - No Change of trty	-1.039**	-1.243***
	(0.37)	(0.37)
Battle-Non-state actr wins trty	-2.972	-3.301
	(2.24)	(2.22)
Non-violent transfer of trty	-2.972	-2.999
	(1.61)	(1.59)
Remote Violence	-1.047**	-1.194**
	(0.38)	(0.38)
Riots/Protests	-2.955***	-3.055***
	(0.37)	(0.37)
Strategic Development	-2.972***	-3.058***
	(0.39)	(0.38)
Violence against civilians	-1.742***	-1.800***
	(0.37)	(0.37)
Intercept	2.972***	2.819***
	(0.37)	(0.37)
N	24366	24366
Adjusted R- squared	0.083	0.100

*p < 0.05; **p < 0.01; ***p < 0.001 NOTES: 1. Collinearity: (a) Battle - Government regains territory excluded; 2. Model 2 controls for COUNTRY

References

Kahle, David, and Hadley Wickham. 2013. "Ggmap: Spatial Visualization with Ggplot2." *The R Journal* 5 (1): 144–61. http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf.

Raleigh, Clionadh, Andrew Linke, Håvard Hegre, and Joakim Karlsen. 2010. "Introducing Acled: An Armed Conflict Location and Event Dataset: Special Data Feature." *Journal of Peace Research* 47 (5). Sage Publications Sage UK: London, England: 651–60.