Fatalities and the role they play in Electoral Violence datasets

---

Stuart Morrison

Supervisor(s):
Prof Alence

A Capstone Project report submitted in partial fulfillment of the requirements for the degree of Master of Arts in the field of e-Science

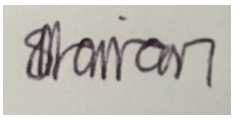
in the

School of Social Sciences
University of the Witwatersrand, Johannesburg

20 September 2022

Declaration

I, Stuart Morrison, declare that this report is my own, unaided work. It is being submitted for the degree of Master of Arts in the field of e-Science at the University of the Witwatersrand, Johannesburg. It has not been submitted for any degree or examination at any other university.



Stuart Morrison

20 September 2022

Abstract

The impact of fatalities on electoral violence datasets is something that is rarely explored despite it being quite a common debate within the field of electoral violence. This paper seeks to determine whether or not fatalities are useful in studying the dynamics of electoral violence. The paper found that the fatalities provide an important and useful theoretical grounding within which to build good electoral violence datasets that are not based on temporal dimensions. Furthermore the paper found that fatalities do not leave out or provide a skewed view of electoral violence by removing non-fatal electoral violence events.

Acknowledgements

The support of the DSI-NICIS National e-Science Postgraduate Teaching and Training Platform (NEPTTP) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NEPTTP. Furthermore the proposal acknowledges the support provide by Professor Rod Alence as supervisor of the project.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Conflict has been a core concern for social scientists since the ancient Greeks. Philosophers, politics scientist, sociologists, international relations scholars and the likes have all dedicated their lives to the field of conflict to not only better understand why conflicts are started but also to understand how to prevent them. However, one of the biggest challenges that scholars for centuries faced was the limited data that was available, often only having to rely on secondary sources, they could only engage in the theoretical world but never accurately test whether these theories best describe the empirical world. Fortunately, since the emergence of the internet and computers data has become more accessible making it easier to analyse and test these theories on conflict. One subfield of conflict which has benefitted immensely from the access to larger and better designed datasets is electoral violence, which in the past few decades has seen an increase the literature on the subject. The accessibility of data on violence that is linked to elections has helped redefine how electoral violence is conceptualised. As Birch, Daxecker, and Höglund (2020), argues this increase in accessibility to data on violence around elections allowed scholars to begin treating election violence as a unique and specialised form of political violence with its own specific dynamics independent of other forms of political violence. This shift in approach to electoral violence required that scholars use empirical data to study these processes and dynamics. One of the most popular datasets used is the Deadly Electoral Conflict dataset (DECO) which is an offshoot of the widely used conflict data set Uppsala Conflict Data Program Georeferenced Events Dataset (UCDP GED). Released in 2019, this dataset has changed how election violence is studied by measuring electoral violence events based on the interactions between actors and not on whether the violence takes place just before, on or right after an election day. However, while this dataset has helped

expand how election violence is understood, there are still several key gaps which have left some important questions unanswered. One such gap is the issue of fatalities; the DECO dataset only treats an event as election violence if there are fatalities during the event. Critics of the DECO dataset have suggested that this has limited what research can be done using the dataset and more importantly has the potential to distort results due to this limitation. However, whether fatalities limit the ability to fully understand electoral violence or not is something that has not been explored. As such this research report seeks to answer the question of what happens when one removes this fatalities filter? Are the critics correct in saying it distorts how electoral violence is understood or not? To test this the author constructed an election violence dataset as similar to the DECO dataset as possible with the only exception being that it did not limit events to those with fatalities. In order to build this dataset the author used data from the widely known political violence and conflicts dataset, the Armed Conflict Location Event Data Project (ACLED). The results were then analysed to see what impact fatalities played in the overall accuracy of the data.

1.1 Background

Electoral violence has increased in interest within the social sciences recently because it has fundamentally challenged one of the core assumptions with regards to democratisation, namely that as countries become more democratic, they become more peaceful. Jansen, du Plessis and Siebrits (2015) argues that despite the deepening of democratic principles in Africa since the 1990's, the number of conflict and violence incidents continues to remain high. Furthermore, in many cases, these democratic institutions are the site of violence and conflict, particularly elections. As such election violence has become more relevant in understanding democracy and conflict on the African continent. The literature on electoral violence can be systemically divided in two ways. Firstly, and more broadly the relevant literature can be divided into two categories namely, the theoretical understanding and conceptualisation of electoral violence, and how data on electoral violence is collected, and studied. Secondly, the literature can be grouped within each of these categories according to how they are defined, namely, election violence as a form of violence

that takes place in proximity to the election day and electoral violence as a political strategy.

1.1.1 Theoretical understanding and conceptualisation:

Firstly, when it comes to the theoretical understanding and conceptualisation of election violence, one needs to start at what elections are. Elections as Wojtasik (2013) argues are a medium through which leaders, governments and certain decisions can be selected by the people of the country. The primary goals of an election are to ensure peaceful political contestation, accountability, political participation from citizens, and promote the interests of the general public (Reynolds, Reilly and Ellies 2005; Heywood 2019). Electoral violence on the other hand is defined largely by two key ideas, timing and strategy (Birch, Daxecker and Höglund 2020). The literature is divided on which of these two is most influential, however it is important to emphasise that these two ideas often overlap with scholars defining electoral violence as a combination of these two ideas, the key difference is which idea they put more emphasis on.

Some scholars in defining electoral violence put a lot of emphasis on timing, arguing that electoral violence is largely defined by the fact that it takes place during the election period or more specifically around the election day (Birch, Daxecker and Höglund 2020; Brancati and Snyder 2013). Some scholars in the literature take the approach that violence is only considered electoral if it takes place in proximity to the election day and in some cases also the proximity to voting stations(). This has been particularly useful in studying relationship between time and place. On the other hand, some scholars determine whether violence is electoral based on proving whether the violence would still have taken place if the election was not being held (). Particularly, scholars in the latter group hold that by determining the temporal dimension based off proximity to the election day, it overlooks violence that may happen early in the election period or very late, which can and does limit the reliability of the data. However, a key weakness when it comes to literature looking at the temporal dimension of electoral violence is that in many cases it is difficult to determine whether violence is related to the election or not. Kovacs (2018) argues that because electoral violence usually intersects with other forms of violence, it can

make distinguishing between various forms of violence based on time and even location, quite difficult. Scholars have thus suggested that using micro-level analysis helps make this distinction between various forms of violence (Balcells and Justino 2014). Hence looking at how electoral violence is defined through timing has both its strengths and weaknesses.

Alternatively, there are scholars who argue that electoral violence can be understood as a means to an end, a strategic tool that is used to achieve some electoral outcome. The literature on electoral violence as a tool and strategy largely considers three key factors, namely electoral violence as a tool to specifically win elections, electoral violence as a way to gain legitimacy within an election and electoral violence as a way to avoid accountability (Smidt 2016; Birch 2020; Wilkinson 2004). For instance, scholars like Birch (2020) argue that electoral violence is a tool used by countries that have weak democratic institutions and high levels of corruption in order to stay in power. Violence in this form can vary from physical hard violence i.e. assassinations and riots, or soft violence i.e. intimidation, exclusion of certain opponents from election. Taylor, Pevehouse and Straus (2017) builds on this by arguing that incumbents rerunning for office are more inclined to use electoral violence to win elections than other political actors if the political competition becomes too close. Thus, scholars looking at electoral violence as a strategy to win elections consider not only failings of institutions which allow for this to happen but also the way in which democratic institutions can fuel electoral violence. Alternatively, scholars looking at electoral violence as a way to gain political legitimacy is usually based along various identities such as race, ethnicity or nationality. Müller-Crepon (2021) argues that regions with greater ethnic polarisation are more likely to have instances of electoral violence. This violence is largely used to legitimise one's own political interests by mobilising ethnic groups against opposition voters highlighting how electoral violence as a strategy can be used to gain political legitimacy.

However, one key gap in the literature on the definitions of electoral violence is that it fails to consider the implications that defining electoral violence has on how datasets are compiled and how they are used to analyse electoral violence. There is very little on the actual datasets themselves and how they are influenced by the electoral violence definitions. Most of the literature is just on the variables themselves and not on how these variables are defined and the nuances of how they are measured. Therefore, the paper seeks to address this gap by contributing

to the limited literature highlighting the implications of definitions when it comes to data. Furthermore, the reason this project is relevant and important in the field of electoral violence is because it can help better understand the role that fatalities play election violence.

1.1.2 How electoral violence is empirically studied:

Literature on how electoral violence is empirically studied has predominantly been centred around two key issues, firstly how the data is collected and secondly how datasets are used to study electoral violence. This subsection of the literature review will focus on how media-based conflict event data, is used to compile data on conflict and how this data, in the form of datasets, are used to develop datasets on electoral violence.

Media-based conflict data

Media-based Conflict Data is defined as data that provides disaggregated data on conflict events which can be used to better study the patterns and dynamics of violence at a micro-level (Weidmann 2014). Scholars such as Weidmann (2014), Fjelde and Höglund (2020) and Eck (2012) raise both weaknesses and strengths with using this approach to empirically study electoral violence. In terms of the weaknesses Weidmann (2014) in particular argues that the general concern with media-based conflict events data is that it can be quite biased. This bias is an issue for two reasons issues, firstly it is an issue when it comes to choosing which stories to report on and which to not report on, as this limits the data around the conflict and in some cases leave out important details which may result in the incorrect classification of conflict. As mentioned in the previous subsection, this can be an issue in trying to determine whether an event is electoral violence or whether it is another form of violence (Fjelde and Höglund 2020). Secondly, bias is an issue as it may distort empirical results thus reducing the reliability of the research. This bias is rooted in inequality around media coverage as typically news covered in the West is more likely to be better covered and investigated which means the sources are more reliable, as they have more information, however news that does not make it in international news, typically from the global south, is largely underreported which when combing all of this data together can get false results. However, they

also argue that this type of data is useful in providing information that is as comprehensive as one can get across a wide range of regions, countries continents, thus while there is inequality in global reporting, on a micro-level it is able to provide a wider range of data. Furthermore, scholars argue that this is important in understanding the dynamics behind the violence, as media-based conflict event data provides more specific data on actors and whether the violence is lethal or non-lethal which is important in discussing the strategy behind the electoral violence.

Datasets

There are a growing number of datasets on electoral violence and as such this literature review may not be able to address every one, however, this literature review will focus on some of the most prominent datasets used to study electoral violence. The various datasets can be categorised in a number of ways, however for this project to show the gap in the literature, and to justify why ACLED dataset is be used, the body of work will be categorised according to which element of electoral violence they address. Firstly, in terms of the temporal dimension of electoral violence there are two key datasets that can be used, namely Varieties of Democracy (V-DEM) and National Elections across Democracy and Autocracy (NELDA), these two datasets both use temporal dimensions as the basis for what is considered electoral violence. However, these electoral violence datasets fail to provide a micro-level analysis and just stay at the national level. This is useful when making cross national comparisons as they are able to determine patterns of electoral violence across different countries and they are able to explore the links between violence and across election periods at a national level. However, there are two key weaknesses, firstly the datasets are very specific in some areas and not specific in others which provides an imbalance that can distort the results. For example, the V-DEM data provides data on electoral dynamics particularly with regards to the intensity of the violence however this is limited because it contains data only at a national level, thus when looking for the more detailed electoral violence dynamics within specific countries, regions, communities, this data is not useful.

In terms of looking at electoral violence as a strategy, one can use the Social Conflict Analysis Database (SCAD) or the Deadly Electoral Conflict Dataset. Both these datasets look more broadly at election dynamics by looking at actors involved.

DECO particularly stands out because it looks at reported purposes as well. However, both these datasets are able to focus on more specific electoral violence incidences which allows studies using these datasets to infer patterns across subnational regions. However, a key difference between these two datasets is that DECO defines electoral violence very specifically as it only looks at events in which fatalities have transpired. Whereas SCAD does not make that distinction and rather looks at all incidents regardless of whether there was a fatality.

There are a number of datasets that one can use to measure electoral violence, which raises the question of why this project would use ACLED data to build the electoral violence dataset? Firstly, this is because the ACLED dataset unlike many other conflict datasets does not only consider an event as something whereby fatalities occur, rather it defines conflict and violence very broadly as to allow non-fatal events. Thus, because the research is focused on the effect of fatalities, it made sense to build the electoral violence dataset using ACLED data.

Furthermore, the ACLED dataset contains certain variables which are similar to that of the DECO dataset, making it easier to build a dataset that is similar to DECO but without the fatality's limitation. The ACLED dataset has a similar set of variables especially when it comes to how the two actors interacted with each other, which will help get a better sense of what is electoral violence and what is not without relying on the time period of when an election occurred.

The final reason for using ACLED data is that it is more established than the other electoral violence datasets meaning it is more reliable and is updated more regularly than other newer datasets.

1.2 Problem Statement

There has been quite a lot of mixed responses to using a dataset that includes or excludes non-fatal events. Some scholars argue that typically when there violence is involved there is a higher chance of a fatality has occurring thus excluding helps define the violence more. However, others argue that it does have an impact and are sceptical to use empirical data in their research because they do not trust the data's reliability. However, this debate has never been fully explored or tested. As such

the research will help explore and contribute to this debate to better understand the role of fatalities in electoral violence datasets.

1.3 Research Question

This research project seeks to understand the role that fatalities plays as a filter for what is considered electoral violence and what is not. More specifically it seeks to answer the question what happens when the fatality filter is removed? Does it improve the reliability of the data or does it worsen it. I put forward the following hypothesis to try answer this question: "Excluding non-fatal events in the electoral violence dataset limits the overall reliability of the dataset and studies done using this data because it creates a skewed perception of electoral violence."

1.4 Research Aims and Objectives

1.4.1 Research Aims

The aim of this research project is to develop a paper on how if at all, including non-fatal election violence events affects the reliability of datasets on this topic.

1.4.2 Objectives

The project has 3 key objectives that it seeks to obtain to achieve the broader aim of the project. More specifically the objectives of the research are to: 1) Clearly understand what the strengths of including non-violent events into the dataset are. 2) Clearly understand what the weaknesses of including non-violent events into the dataset are. 3) Gain insight into the nuances of elections and violence.

1.5 Limitations

There are three main limitations to consider for the research. Firstly, the project is limited by the availability and reliability of the ACLED dataset. One of the big criticisms of the ACLED is that because it relies on media sources as its main source of data collection, the data may not represent an accurate depiction of either what had happened, or it fails to account for certain events entirely. This can skew the results from the data which in turn hinders the ability to provide accurate generalisations from a sample size because there is information missing. In terms of how it relates to the project more specifically this means that pieces of data can be left out or overlooked due to a lack of information when it might be useful in studying the dynamics of electoral violence. Secondly, because the ACLED dataset is a conflict dataset not an election violence dataset it is almost impossible to get a very accurate electoral violence dataset from the ACLED data. Thirdly, the project is limited by the computational skills and knowledge of the researcher. The study would be improved if the author had a deeper understanding of computational and statistical techniques which could provide a far more elaborate analysis.

Chapter 2

Research Methodology

The next chapter will be providing a brief overview of the Research Methodology for the project. In order to do this the chapter will be considering and exploring all the relevant methods for the project, which methods the project will be using and why. The chapter will also consider the various limitations of the methodology to provide a clear understanding of the research methodology.

2.1 Research Design

Scholars studying conflict have largely preferred a more empirical, quantitative research approach than the more common qualitative approach that is taken in the Social Sciences. This is largely because empirical data enabled researchers and scholars to test and make conclusions about the plethora of theoretical concepts on conflict and organised violence and how it links to the population. Electoral violence is no exception to this with scholars since in the early days of research preferring empirical methods to other such methods and with the increasing accessibility to the internet these datasets are becoming more accurate and more accessible.

For this project I used also followed this trend by using a quantitative and empirical approach to the research however, unlike with a lot of the research done on electoral violence, the primary goal is not to show a causal relationship between two variables, and thus will be using an approach that looks at evaluating and addressing a gap or problem within the research environment. The most common of such broad research methods and the one this research project will be looking at is the Action research method. The reason this method is most applicable to this

study is because it is commonly used to evaluate and critically examine the epistemology of a specific field, specifically addressing the issues around the research and knowledge production process within a field. While typically this method is collaborative and conducted over several years with many participants, this approach is most relevant to the type of research the project seeks to conduct. The Action research design includes five key steps, namely identifying a problem or gap within the field, proposes a specific course of action to address problem, test the proposed 'action', evaluate and analyse the results and finally the processes is repeated. This research design is most appropriate because to test the impact of fatalities, I need to compare the DECO dataset to several versions of the ACLED built dataset. The project will build two datasets to be compared to the DECO dataset using different techniques to ensure that there is less bias. The course of action, following the research design will thus be build the one dataset, run various tests, reviewing and cleaning the dataset throughout the tests and then doing the same with the second dataset

2.2 Data

As mentioned briefly earlier in the paper, the project used two different datasets. The first being the DECO dataset which is explicitly an electoral violence dataset built off a more broader conflict and violence dataset, the Uppsala Conflict Data Program Georeferenced Events Dataset (UCDP GED). DECO defines electoral violence as "violence that is substantially linked to an electoral contest" (Fjelde, Höglund, Olafsdottir Renvall 2021: 5) which fits the focus that the dataset has on the interaction between actors rather than whether the event happens during an election cycle. Furthermore, the dataset tracks electoral violence events globally not restricting it to a certain region however because it is still quite new it only has data on electoral violence from 1989 to 2017. What makes the dataset useful to those studying electoral violence is that it not only removes the temporal limitation but also looks into the perpetrators desired targets and provides additional pieces of information not really considered before such as an electoral violence uncertainty score which assigns a score to each event based on how uncertain they are that it is electoral violence, with zero being most confident to 2 being most uncertain. DECO puts a lot

of emphasis on the interaction between the actors as the main parameter to determine what is electoral violence and what is not. Within this parameter the dataset uses three criteria for election violence, namely who are the perpetrators, who are the targets and what is the reported purpose behind the violence. Using these three main criteria it distinguished between electoral violence and other forms of violence. One key limitation to keep in mind when working with the DECO dataset is that because the definition and scope of electoral violence is so specific it has the potential to limit exclude various cases of electoral violence. Due to the scope of the study, I only extracted the most relevant variables to use namely, year, country, actors for side A involved in the conflict, actors for side B, most accurate estimated fatalities, and type of violence, which is broken up into state-based conflict (coded as 1), non-state conflict (coded as 2), one-sided violence (coded as 3), violence in civilian protests (coded as 4).

The ACLED dataset on the other hand is a conflict and violent events dataset, which similar to UCDP GED, tracks various forms of violence, conflict and aggression around the world. It is by far one of the most comprehensive datasets however what makes it useful to many different scholars is the fact that it defines violence very broadly, allowing one to include as many events as possible capturing a plethora of different forms of violence. In addition, the ACLED data also tracks events that are not necessarily incidents of violence but are events linked to groups involved in such violence like the signing of agreements between actors, changing of strategies or even movement of soldiers and military forces. ACLED was chosen to be used in the study because it not only includes non-fatal events, but also because it has very similar variables to DECO allowing the study to be more comprehensive and as accurate as possible. The variables from the ACLED data chosen for the electoral violence dataset includes: The data of event, country, the type of event (conflict or otherwise), type of sub-event (a more detailed breakdown on the type of violence it is), Actor 1 (which is the perpetrator), Actor 2 (victims or group receiving violence), an Interaction variable which is coded to describe the type of interaction between actor 1 and 2 and then additional notes and sources on each event. I also included the associated actor's variables (for both 1 and 2), which similar to DECO, can be used to help identify any actors that may not be directly linked to the election process but is due t0 their association, making it easier to determine what is electoral violence and what is not based on the interaction between actors.

The electoral violence datasets that I built are based on the DECO definition of electoral violence and helped to guide what events to include and what to not include, this was especially true for the second dataset which looked closely at the events themselves as opposed to just using the variables (like in the first dataset). The key word from this definition is "substantially" which in this case was interpreted to mean that there needs to be enough evidence that there is a link to electoral violence in order for it to be included in the data. Cases that were unclear on whether they were electoral violence or not were excluded. Since the DECO dataset is from 1989 to 2017 and ACLED is from 1997, my electoral violence dataset used the period from 1997 to 2017 to accommodate both ACLED and DECO. Additionally, because the ACLED dataset is a global dataset which means it tracks conflict from around the world. However, because using all this data would be too much for one project, I reduced the sample size to just the African continent. I did this for two reasons, firstly I have the most experience with conflict in Africa and thus could gauge better whether an event is electoral violence or not and being familiar with the region helped in the theoretical component of the research. Secondly, the African region is of relevant interest when it comes to electoral violence because the democracies are fairly new and have challenged many of the Western assumptions about liberal democracy, governance and elections that dominate the literature and scholarship more broadly. Thus, focusing on Africa would be most relevant as research is now predominantly focused on the continent, helping fill a gap for researchers looking to use DECO or ACLED data in their research on electoral violence.

Sample of the First ACLED Electoral Violence dataset

EVENT_DATE	YEAR	EVENT_TYPE	SUB_EVENT_TYPE	COU
1997-01-01	1997	Violence against civilians	Attack	Alge
1997-01-02	1997	Violence against civilians	Attack	Alge
1997-01-03	1997	Violence against civilians	Attack	Alge

2.3 Methods

In order to carry out the broad research design, the project built two datasets off of the ACLED data using various computational techniques such as indexing, text mining, various visualisations and descriptive statistics. These datasets were then analysed and explored. The reason for building two datasets was to ensure that bias from building the dataset is accounted for and to get better results by analysing two different samples. Each dataset is more specific allowing for a more comprehensive analysis of the impact that including or excluding fatalities will have on the accuracy and reliability of the data. The first dataset explores the effect of fatalities more broadly, looking at whether it helps further clarify what electoral violence is and is not or it simply creates too much noise thus does not have much impact at all. In doing so it helps test the hypothesis.

The second dataset is more specific looking at what happens if you remove the filter when doing a more comprehensive and detailed cleaning of the data as compared to a more general relying only on the structured variables. This dataset looks at all the electoral violence events in Nigeria during the 1997-2017 period. These results are then compared to DECO's to better understand the impact of excluding non-fatal electoral violence events. Looking at a specific country will help test the hypothesis on whether including non-fatal electoral violence events decreases the accuracy and reliability of the data by seeing whether it helps paint a more accurate picture of election violence in a country. Which in turn will help determine whether not having the filter is useful. Using these two datasets, I was able to analyse and explore the effects that the fatalities filter has on the accuracy and reliability of electoral violence datasets which in turn help answer whether including non-fatal events is useful in studying electoral violence.

2.3.1 Dataset 1

The first dataset based off ACLED data, used the actors, variables, events variables and interaction code variable to determine what is electoral violence and what is not. Without doing a deep dive into the specific rows and events themselves, I removed all the events that would definitely not be relevant or be election violence themselves. Starting off the events and sub-events variables, I first removed the event type called "Battles" because as defined by ACLED's codebook, this event type focused explicitly on warfare, between two politically organised groups or actors. This event does not fit the scope and relevancy of the paper because even though elections and electoral violence can happen during wars and conflicts, battles as defined by ACLED are specific events during a specific period and location and thus the ACLED data does not treat overarching wars (Like The Algerian Civil War) as events but rather only focuses on the specific battles in that war. Hence it made no sense to include this event type. The only other event type that I excluded was the Strategic development type because it included various non-violent events such as signing of agreements, non-violent transfer of territory, establishing of headquarters. All things that are not relevant to the study as they contradict election violence because they are nonviolent events, so it does not make logical sense to include them. The only one 'non-violent' event that I did include was looting

or destruction of property, because it can be used as intimidation which is a type of electoral violence. For example, polling or voting stations can be attacked, the votes can be stolen or damaged, or political figures property can be attacked. Thus, it was included.

The next thing to remove were specific sub-events from the remaining event types, namely, Explosions/Remote violence, Violence against civilians, Protests and Riots. In terms of the Explosions variable, I removed two sub-event types: the Air/drone strikes, and Shelling/artillery/missile attack the reason I removed these two is because they are not logically relevant to the study. Even though air/drone strikes do happen during election periods they are rarely used perpetrate election violence, and upon inspection of the ACLED data no events involving these two categories were linked to electoral violence. In terms of Violence against civilians I did not remove any of the categories because they all are relevant to electoral violence as in many cases groups use Sexual violence, Attacks against civilians, and Abduction/forced disappearance as mechanism to perpetrate this form of political violence. In terms of Protests and Riots I only removed peaceful protests because these are events in which no violence occurs and as such isn't very relevant in a study on election violence. Protest with intervention and Excessive force against protesters were kept in because these can also be used as election violence.

The next step was to remove the interaction codes which were not relevant to this study. I counted something as irrelevant if it did not fulfil the election violence criteria given by DECO, meaning if it did not involve some form of violence, with the potential of being used to affect an election outcome. In this case it includes non-physical acts of violence such as intimidation as our dataset does not have the fatalities filter. Some of the notable interaction codes I removed include sole military actions (which means it was peaceful military action, establishing a base, etc). I also removed Military versus military as a way to help clean up the armed clashes sub event and not pick up conflicts that are just two militaries. I also removed all the protestor interaction codes they referred to all peaceful events as well as the other interaction code which was also about things that were not relevant to the study such as international organisations.

The final thing I did was look at the actors, there weren't any specific actors that are not capable of election violence however I did remove all the unidentified

actors because we are using actors and their interactions with each other as the main determining factor for electoral violence. It did not make sense to include unknown actors as it will be impossible to know whether they are linked to electoral violence or not. To remove them I went down the list of all the African countries that the ACLED data had to also just check which countries were missing. This applied to the associated actors as well because we are determining electoral violence based on actors and their engagements, it is very difficult to determine whether something is electoral violence if there are no associated actors to help provide that substantial link.

Lastly because there was no data on São Tomé and Príncipe, Eswatini (or Swaziland), Congo-Brazzaville or Comoros. While doing this I also left out Eritrea which did not have elections for the years 1997 to 2017 and so it was also excluded. Overall, for the first dataset I used the structured variables to clean the data, looking at the various components from the codebook and indexing based off those definitions. I did this to explore the impact of fatalities on a more general dataset scale.

2.3.2 Dataset 2

To build the second dataset I recleaned the dataset but instead of using the approach that the DECO dataset took, I used a different approach, that being working through the various variables together to determine what is electoral violence and what is not. To do this I looked at each of the variables for the dataset and using all of them tried to do a more in-depth exploration to determine whether something is electoral violence. However despite using a different approach this time, I still used the core definition as DECO, meaning I only included events that had substantial links to electoral violence. The reason I used Nigeria is because within the DECO dataset it makes up the majority of cases for Africa. It is better to have a sample with more data and information than one with less because you can explore and understand a lot than with a smaller sample size. To begin I looked at each of the various variables such as event type, sub event type actors interaction codes and notes to cut out cases that were not relevant. After extracting the Nigeria data from the ACLED dataset, the original size was 10022 events. However, using the different event types as the basis for cleaning the data I was able to cut the events down to just under 300 events.

Chapter 3

Results and Discussion

3.0.1 Dataset 1 Results and analysis

By cleaning and sorting the data, there were a few key things that the study found. Firstly overall, it found that the approach is as important to building the dataset as the filters are. After cleaning the dataset as explained in the previous chapter, the dataset was riddled with issues, particularly was the fact that it was not picking up actual election violence events and populating the dataset with false positives instead. Table 3 shows how many mentions of election, polls or voting or ballots were found in the note's variable of the dataset. Out of 3047 events, only 177 of them picked up these after removing Looking Table 3, it shows that only 188 of the events mentioned at least one of the key words mentioned above. Even though using text mining can pick up false positives it is enough to gauge roughly what the data looks like.

Table 3: Number of election violence terms in Dataset 1

Var1	Elections	elections	polls	ballot	voting	vote
FALSE	3045	2954	3042	3030	3024	2999
TRUE	2	93	5	17	23	48

The reason why the results were so poor in the first dataset is largely because one key factor was overlooked, the fact that the fatalities filter in the DECO dataset works because it helps ground the approach that it took to building the dataset. As mentioned earlier the DECO dataset uses the actors and their interactions with each other as the core parameter for how to determine what is election violence rather than using the common temporal parameter. However, because the approach is to focus on the actors themselves, the fatalities filter helps to provide a core theoretical groundwork upon which these interactions are based. However, because we removed that theoretical basis for interaction, any event that had certain actors or interaction codes in the ACLED dataset was being picked up as electoral violence, even though it should not have been. This disproves the hypothesis that excluding the non-fatal electoral violence events limits the reliability of the dataset rather it helps to ground the dataset theoretically making it more likely to improve reliability. This shows that the criticisms about the DECO dataset being too limited are somewhat untrue as seen with this test, the filter has a core role in the construction of the dataset.

3.0.2 Dataset 2 Results and analysis

The second dataset was cleaned much more accurately as I did not rely on the structured variables, allowing me to provide a more comprehensive dataset. Even though the dataset was small only 200 or so events, there were a few insights from building this dataset. Firstly, by removing the fatalities filter in this instance, there were more cases of election violence than with the DECO which is to be expected. However, despite there being more events than the DECO dataset, it did not seem to make a drastic difference in the dataset as the difference was very minimal. This shows that when it comes to removing the fatalities filter it is not as great a difference in the data than some people would imagine. This further disproves the hypothesis that excluding non-fatal electoral violence events does not give a completely skewed and flawed view. However, it must be noted that this is just one country in many different countries, which means that if this dataset is built using a different country the differences might be greater. One interesting thing that is interesting from the findings is that even though for electoral violence the difference in fatal events and non-fatal events is minimal, the overall sample, has a greater number of non-fatal events, which does make sense in a way because there are typically more events that do not involve fatalities because there are far greater situations in which non-fatal violence can occur. However overall, the second dataset also shows that the impact of non-fatal electoral violence events is not as significant as some assume.

3.1 Summary

Overall this study sought to understand the impact if any that fatalities has on election violence datasets. To do this paper built two datasets exploring different components of this idea. The first dataset explored whether removing the fatalities filter will mkae it easier to identify and determine what is and what is not electoral violence. However the results showed that by removing the filter it makes it extremely difficult to determine what is and isnt electoral violence. This is because the fatalities filter grounds the dataset theoretically making it easier to determine the parameters for classifying electoral violence. The second dataset was a more comprehensive dataset, focusing only on Nigeria, and was built to test whether

the removing the fatalities filter with allow for more complete picture of electoral violence on the ground. However the results showed that there was very little difference between the DECO dataset events and my dataset events. By doing so it showed that the fatalities do not have as much an impact as some assume. Thus in both cases the datasets disproved the hypothesis that by excluding the non-fatal electoral violence events from electoral violence datasets it does limits the accuracy and reliability of the data.

Chapter 4

Conclusions and Future Work

4.1 Conclusions

Scholars have been debating whether the fatalities filter limits electoral violence datasets, with some arguing that it does not really matter because in many cases where there is violence typically means some fatality will occur. While others criticising dataset that do this argue that it limits the reliability and accuracy of the dataset. However, despite the various debates on the issue, no one has really tested to see if this is actually the case. This paper sought to conduct research in whether or not fatalities limits the reliability of electoral violence datasets. The aim of this research project is see if including non-fatal election violence events affects the reliability of datasets on this topic. In order to test whether this is true, the paper built two datasets without the fatality filter from ACLED data using these two datasets the study explored the hypothesis mentioned above. Despite the main hypothesis the results from the two datasets showed that this was not the case. Fatalities as a filter for electoral violence datasets is useful because it not only grounds the approach to determining what is electoral violence and what is not but it also makes it easier to build a dataset with this filter and does not completely skew results that emerge from it.

Chapter 5

Bibliography

Birch, Sarah. “Electoral Violence, Corruption, and Political Order”. Princeton: Princeton University Press, 2020. https://doi.org/10.1515/9780691203645

Birch, Sarah, Ursula Daxecker, and Kristine Höglund. “Electoral Violence: An Introduction.” Journal of Peace Research 57, no. 1 (2020): 3–14. https://doi.org/10.1177/00223433

Brancati, Dawn, and Jack L. Snyder. “Time to Kill: The Impact of Election Timing on Postconflict Stability.” The Journal of Conflict Resolution 57, no. 5 (2013): 822–53. http://www.jstor.org/stable/24545572.

Coppedge, Michael, Gerring, John, Knutsen, Carl Henrik, Lindberg, Staffan I., Teorell, Jan, Altman, David, Bernhard, Michael, Steven Fish, M., Glynn, Adam, Hicken, Allen, Lührmann, Anna, Marquardt, Kyle L., McMann, Kelly, Paxton, Pamela, Pemstein, Daniel, Seim, Brigitte, Sigman, Rachel, Skaaning, Svend-Erik, Staton, Jeffrey, Wilson, Steven, Cornell, Agnes, Alizada, Nazifa, Gastaldi, Lisa, Gjerløw, Haakon, Hindle, Garry, Ilchenko, Nina, Maxwell, Laura, Mechkova, Valeriya, Medzihorsky, Juraj, Römer, Johannes von, Sundström, Aksel, Tzelgov, Eitan, Wang, Yiting, Wig, Tore, 2020, Daniel Ziblatt.. “V-Dem Country–Year/Country–Date Dataset v10.” Varieties of Democracy (V-Dem) Project. Accessed May 23, 2022. https://doi.org/10.23696/v

Eck, Kristine, and Nick Dietrich. "Known unknowns: media bias in the reporting of political violence." International Interactions 46, no. 6 2020: 1043-1060. doi:https://doi.org/10.1080/03050629.2020.1814758.

Fjelde, Hanne, and Kristine Höglund. “Introducing the Deadly Electoral Conflict Dataset (DECO).” Journal of Conflict Resolution 66, no. 1 (2021): 162–85. https://doi.org/10.1177/00220027211021620.

Heywood, Andrew. Politics. 5th. London: Palgrave. 2019.

Hyde, Susan and Marinov, Nikolay. “Codebook for National Elections Across Democracy and Autocracy Dataset, 5.0.” Accessed May 23 2022. (2019). https://nelda.co/codebook

Müller-Crepon, Carl. “Local Ethno-Political Polarization and Election Violence in Majoritarian vs. Proportional Systems.” Journal of Peace Research 59, no. 2 (March 2022): 242–58. https://doi.org/10.1177/0022343320973724.

du Plessis, Sophia, Krige Siebrits, and Ada Jansen. 2015. Democratisation in Africa: The role of self-enforcing constitutional rules. 1. Accessed 05 24, 2022. https://www.accord.org.za/ajcr-issues/democratisation-in-africa/.

Reynolds, Andrew, Ben Reilly , and Andrew Ellis. 2005. “Electoral System Design: The New International IDEA Handbook”. Stockholm: International Institute for Democracy and Electoral Assistance.

Smidt, Hannah. “From a Perpetrator’s Perspective: International Election Observers and Post-Electoral Violence.” Journal of Peace Research 53, no. 2 (March 2016): 226–41. https://doi.org/10.1177/0022343315626240.

Söderberg Kovacs, Mimmi . 2018. "Introduction: The everyday politics of electoral violence in Africa." In Violence in African Elections: Between Democracy and Big Man Politics, edited by Mimmi Söderberg Kovacs and Jesper Bjarnesen, 1-26. Uppsala: Zed Books Ltd.

Taylor, Charles Fernandes, Jon CW Pevehouse, and Scott Straus. 2017. "Perils of pluralism: Electoral violence and incumbency in sub-Saharan Africa." Journal of Peace Research (Sage Publications, Inc) 54 (3): 397-411. https://www.jstor.org/stable/44511221.

Weidmann, Nils B. “On the Accuracy of Media-Based Conflict Event Data.” The Journal of Conflict Resolution 59, no. 6 (2015): 1129–49. http://www.jstor.org/stable/24546323.

Wojtasik, Waldemar. 2013. "Functions of Elections in Democratic Systems." PoliticalPreferences. no 4. https://www.researchgate.net/publication/255965201$_Functions_of_Election$