Women in insurgent groups: a quantitative assessment of women leaders' impact on civilian violence perpetration


Sources: Kurdistan au féminin, n.d. ; Kurdistan au féminin, n.d.

Bachelor Thesis: Political Science: Insurgency and Counterinsurgency

Name: Milano Katia

Student number: 3759830

Workgroup number: 105

Instructor: Dr. Corinna Jentzsch

Second reader: Dr. Diana Davila Gordillo

Date: 19/12/2025

8004 words

Reference Style: APA 7th edition

Use of ChatGPT to reference the sources in the bibliography in APA style (7th edition)

Public thesis

Universiteit Leiden

Table of content

List of figures

List of tables

I. Introduction

"What would we learn about rebel governance and civil wars if we included women and took their labour, participation and experience of wars seriously?" (Hedström, 2025, p. 1).

"If one fails to pay close attention to women—all sorts of women—one will miss who wields power and for what ends." (Enloe, 2014, p. 28).

Women occupied all positions in the Communist Party of Nepal–Maoist (CPN-M) insurgent group and their participation decreased civilian victimization (Mehrl, 2023, pp. 265-271). Onsari Gharti Magar or Devi Khadka are examples of women's critical leadership within the CPN-M. Yet, their names are much less familiar than men insurgent' leaders widely mentioned by mainstream scholars of political violence, such as Che Guevara or Mao Zedong (Duyvesteyn & Rich, 2012). Consequently, the topic of insurgent leadership omits women's agency from the argument, showcasing that the topic is predominantly researched with a patriarchal lens (Arjona, 2014; Duyvesteyn & Rich, 2012; Mampilly, 2012), over-emphasizing men's over women's roles as insurgent leaders.

Yet, the presence of women in the insurgent leadership is critical to comprehend insurgencies' outcomes (Henshaw et al., 2019, p. 1089), rendering the topic highly relevant to investigate. Indeed, Enloe (2014, p. 29) states that overlooking women's agency represents a danger, as it favors a representation of women through traditional norms rather than as agenda-setting actors striving to enforce their decisions. In this regard, a wide array of scholars has delved into women's involvement in rebel groups during civil wars, their roles within insurgent organizations, and their impact on conflict outcomes (Hedström, 2025; Loken, 2022; Myers, 2025; Utas, 2005; Viterna, 2015; Wood & Thomas, 2017). Nevertheless, few authors scrutinized women leadership positions in insurgent groups' governance (Loken, 2022, Loken & Matfess, 2023a; Myers, 2025). Therefore, additional research needs to be undertaken to enhance our comprehension of the topic. Loken and Matfess (2023a, p. 490) compiled a novel dataset, specifically encouraging research on the nexus between different types of women participation and their impact on insurgent groups' behaviors and strategies. Following their recommendation, I tackle the controversial and widely debated topic of women leadership in insurgent groups.

While scholars point to women participation's effect on insurgencies' dynamics (Loken & Matfess, 2023a), I narrow down insurgencies' outcomes to a puzzle relative to civilian violence variations (Hultman, 2012, p. 289; Kalyvas, 2006; Winward, 2016). Scrutinizing this phenomenon is fundamental as the question has been heavily debated by scholars and no definite answer relative to the determinants of varying levels of violence has been formulated. More specifically, civilian violence is important to examine as it engenders large human sufferings (Humphreys & Weinstein, 2006, p. 429). Additionally, little effort has been made to link civilian violence dynamics to women leadership (Loken, 2018).

Therefore, I aim to connect these two academic debates and shed light on a grey area from the literature by investigating the following research question: What is the effect of women leadership within insurgent groups on the groups' level of civilian violence perpetration? I attempt to answer this research avenue by using a quantitative analysis to systematically address my research agenda and to generalize my findings to a broad set of cases. I resort to a linear regression to comprehend the relationship between my two variables of interest.

I start this thesis by reviewing the literature on the topics of women leadership in insurgencies and civilian violence in civil wars. I continue by conceptualizing this paper's key concepts. I further develop a novel theory giving rise to a hypothesis. Methods and data are then discussed before presenting and analyzing my results. I find that insurgent groups with a leadership comprising at least one fifth of women on average commit significantly fewer civilian killings than groups with no women leaders, when controlling for other predictors. I hence conclude that a high proportion of women leaders corresponds to a meaningful reduction in the insurgent group's perpetration of civilian violence. Women leaders hence shape insurgencies' dynamics and groups' behaviors and strategies.

II. Literature review

The literature review starts by shedding light on the various ways in which women participate in insurgent groups during civil wars. Then, it underlines how women impact insurgencies' outcomes. Thereafter, civilian violence determinants are investigated. I further review gender perspectives on civil wars' dynamics. Lastly, this chapter summarizes scholars' arguments, identifies an under-researched gap and formulates the research question that guides the analysis.

1. Women during civil wars

Previous scholars have investigated the pathways that women follow during civil wars. Viterna (2015) finds that women's mobilization into armed groups encompasses a broad range of motives and is contingent upon the interaction of various factors, such as situational contexts, biographical availability and network ties. For example, women politicized before joining the Salvadoran guerilla participated in the conflict to defend their values (pp. 90-91). Similarly, women recruited during the conflict were enthusiastic to fight the war (pp. 93-95).

By contrast, Utas (2005) argues that female civilians socially navigate conflicts by oscillating between combative and non-combative participation with an armed group, to overcome the conflict's violence (p. 408). Yet, by reducing women's agency to a quest for survival, the author focuses on women's coerced positions and overlooks women's ideological and political commitment to armed groups (Hedström, 2025, p. 7). Therefore, I aim to showcase that women are crucial political actors who play an active and politically driven role in insurgencies.

While early scholarship limited their focus to women's participation in insurgencies as fighters (Alison, 2004; Jordan & Denov, 2007), later scholarship enlarged the scope of women's mobilization in insurgencies to non-combatant positions (Hedström, 2025; Loken, 2018, 2022). Indeed, Loken and Matfess (2023, p. 489) broaden women mobilization to their participation in combat, non-combat and leadership positions. Nevertheless, the authors do not account for various outcomes produced by women participation. In this regard, other authors delved deeper (Hedström, 2025; Henshaw et al, 2019; Loken, 2018, 2022; Myers, 2025).

2. Impact of women participation on insurgencies' outcomes

Women's mobilization in insurgencies takes many forms and shapes insurgencies' viability, success, and governing policies. First, women prevalence within armed groups enhances groups' viability (Hedström, 2025, p. 5; Loken, 2022, p. 156). Indeed, women occupy crucial supportive roles, such as daily gendered labor within the home, which is fundamental to sustain armed groups' organizations over time.

Second, women increase the likelihood of an armed group's success and post-conflict's prosperity (Henshaw et al., 2019; Loken, 2018, 2022, 2024; Braithwaite & Ruiz, 2018). Indeed, women participation in insurgencies heightens the armed group's legitimacy, increasing the group's likelihood of gaining territorial control and civilian support (Loken, 2018, 2022, 2024). Consequently, such dynamics favor insurgencies' success and post-conflict peace settlement prospects. Likewise, Henshaw et al. (2019) uncover that women participation in armed groups increases conflicts' duration and has a higher likelihood of negotiated settlement. Nonetheless, while voluntary women participation tends to engender insurgents' victory, forced women recruitment decreases the armed group's chances of success (Braithwaite and Ruiz, 2018). Consequently, the association between women participation in insurgencies and the armed group's victory is contingent upon women's willingness to participate in the groups.

Third, women insurgent leaders promote social care services policies. Myers (2025) finds a significant nexus between women leadership and direct social care services. Subsequently, women's mobilization in insurgencies shapes insurgencies' viability, success, and governing policies. While Loken and Matfess (2023a, p. 490) invite scholars to study how various levels of women prevalence and type of participation in armed groups influence rebel behavior and strategies, little effort has yet been made to comprehend how women participation affects insurgents' civilian violence strategies and perpetration. Yet, understanding the determinants of civilian violence constitutes a crucial academic debate (Hultman, 2012, p. 289).

3. Mainstream debate about civilian violence's determinants

Many scholars scrutinized civilian violence facilitators, but no consensus has been reached. Following a rational choice perspective, Humphreys and Weinstein (2006, p. 430) argue that sufficient incentives to restrain civilian targeting in exchange for anticipated future rewards, territorial control, and armed group's internal cohesion lessen civilian violence. According to the authors, these three key factors solve the collective action issues faced by armed groups. Alternatively, ideology is expected to shape rebel group's strategies and violence use (Thaler, 2012). Indeed, the choice of civilian targeting's scope lies upon the group's ideological commitments to restrain civilian violence. On the other hand, Fujji (2008) reorients the debate towards social embeddedness arguments. That is, indiscriminate violence is better explained by situational factors, such as immediate social context and social ties.

Conversely, Winward (2021) states that intelligence services with a better capacity reduce civilian targeting criteria, decreasing civilian violence. Additionally, Mampilly (2011) discredits ideological and economic arguments (pp. 13-15), showing how the interactive process of governance better accounts for violence propensity (p. 232). In fact, rebel governance conditions social, political and economic interactions with civilians, deeply impacting the group's ability to enhance voluntary participation, and consequently shape civilian violence. Arjona (2014) supports that insurgent groups wartime governing institutions impact violence's scope and brutality.

In summary, determinants of civilian violence remain debated. As identifying members of armed groups is necessary to understand the motives for extreme violence in civil wars (Cohen, 2013, p. 387), I study female participation to examine civilian violence. I hence scrutinize gender theories' findings.

4. *Gender perspectives on civil wars' dynamics*

While gender theorists touch upon civil war's dynamics, no systematic link is established between women leaders and civilian violence perpetration. Rather, authors delve into the impact of structural gender inequalities on civil wars' outcomes (Asal et al, 2013; Caprioli, 2005; Huber, 2018; Hudson, 2019). First, scholars find a positive association between gender inequality and intra-state conflicts onset (Caprioli, 2005; Hudson, 2019). Indeed, gender theories underline that socialization and gender norms sustain structural violence (Caprioli, 2005). Additionally, Hudson et al. (2019) link micro and macro processes, arguing that women domestic disempowerment generates political violence, thereby endangering international security.

Second, Asal et al. (2013) find a significant negative association between gender-inclusive ideology and violence. Similarly, Huber (2018) discovers a positive association between gender inequality and civilian violence. While gender equality condemns attacks against civilians, gender inequality triggers civilian violence. Despite the relevance to scrutinize patterns of women participation in armed groups to comprehend insurgencies' outcomes (Henshaw et al. 2019), gender theories did not yet systematically examine women participation's impact on civilian violence.

5. *Summary and research gap*

In summary, I identified two gaps in literature. First, a gray area remains relative to women leadership's inclusion and its impact on rebel order (Myers, 2025, p. 189). While mainstream theories study rebel governance to enhance our comprehension of violence perpetration (Mampilly, 2011, p. 232), they omit a deep understanding of women leadership's impact on insurgency's consequences. Yet, scrutinizing female leadership is highly relevant to comprehend insurgencies' outcomes (Henshaw, et al. 2019).

Second, understanding civilian violence variations within civil wars remains a puzzling phenomenon (Hultman, 2012, p. 289) which crucially matters due to the devastating human suffering experienced by civilians (Humphreys & Weinstein, 2006, p. 429). I therefore bridge the two gaps into a unique research agenda. Consequently, as the nexus between women leadership in armed groups and violence perpetration's scope has not yet been systematically investigated, I address the following research question:

What is the effect of female leaders within insurgent groups on the groups' level of civilian violence perpetration?

III. Theoretical framework

Firstly, I conceptualize the paper's key concepts: women leadership in insurgent groups and civilian violence. I then discuss theories exploring women leadership and civilian violence. I further assess the theories' strengths. Lastly, I present my own theory connecting this research's two variables, and leading to my hypothesis formulation.

1. Concepts

Firstly, w*omen leaders in insurgent groups* are defined as women exerting direct control over other participants or over the strategies, policies or ideologies of the group (Loken and Matfess, 2023a, pp. 490-491). This notion incorporates both military leaders such as commanders or lieutenants, and non-military leaders occupying auxiliary functions in political organizations (Loken & Matfess, 2023b, p. 10). I*nsurgent groups* are armed groups possessing fewer capabilities and resources than governments and using guerilla tactics to conduct warfare (Duyvestern & Rich, 2012, p. 11). More specifically, a*rmed groups* refer to formally organized groups having a name and resorting to armed force (Högbladh, 2025, p. 30; Pettersson, 2025, p. 3) to achieve political control (Malone, 2023, p. 3).

Secondly, *violence against civilians* is defined as the use of armed forces targeting civilians (Downes, 2006, P. 156) and leading to direct civilian killings (Högbladh, 2025, p. 30). Lastly, *civilians* are defined as unarmed individuals not actively participating in operations conducted by governments or armed groups (Balcells, 2017, p. 20; Downes, 2006, p. 157; Högbladh, 2025, p. 30; Thaler, 2012, p. 547), and posing no direct lethal threats to others (Barter, 2012, p. 546).

2. Women leaders in insurgent groups

This paper's argument draws on several authors' theories. To begin with, women in insurgent groups are expected to enhance the group's ability to gain civilian support, and to integrate into local communities (Loken, 2018). Indeed, gendered narratives convey a trustworthy image of women (Loken, 2018; Myers, 2025, p. 193). Hence, women participation has political and ideological legitimizing effects on civilians' perception of the armed group (Loken, 2018). Consequently, civilians comprehensively respond to the group's grievances, building up support and confidence in the group. This bonding process consequently integrates the group into local communities (Figure 2).

To continue, women in rebel governance adopt objectives based upon gender expectations and therefore promote direct care services policies (Myers, 2025). Securing leadership positions in armed groups is challenging for women, who are expected to provide traditional ‘women’s work’ (p. 190). Gilligan (as cited in Eisler & Loye, 1986, p. 96) confirms that women are morally incited to act in caring or compassionate ways. As women strive to preserve their high-ranking status, they respond positively to gendered duties by abundantly focusing on direct social services (Myers, 2025; see Figure 3).

In summary, these theories explain how women participation in insurgencies has a double effect on civilian support. On the one hand, women insurgents legitimize the movement grievances and hence gain trust and civilian support (Loken, 2018). On the other hand, women leaders shape the armed group’s social agenda, orienting policies towards civilian support (Myers, 2025).

3. Civilian violence

Wartime institutions conceptualized as ‘rebelocracies’ establish a broad social contract between insurgents and civilians, limiting insurgents’ brutality to selective violence (Arjona, 2014). A social contract creates predictability and order by ensuring that insurgents and civilians follow consented rules of conduct (p. 1374). Rebelocracies issue a social contract encompassing a large range of insurgent actions, reaching a de facto authority, and tackling political, economic and social policies. As civilians are broadly protected by insurgents, they collaborate with them and provide information enabling insurgents to select targets. Rebelocracies are therefore expected to perpetrate selective violence (Figure 4).

While selective violence does not necessarily entail absolute low levels of civilian violence, it does reduce civilians' targeting criteria, producing much lower levels of violence compared to indiscriminate violence (Koc-Menard, 2006, p. 333). First, while selective violence involves the strategic targeting of individuals collaborating with a rival group, indiscriminate violence does not restrict its targeting criteria, harming individuals regardless of their conduct or political affiliations. Secondly, selective actions are not intended to target public and social infrastructures such as schools or hospitals. Thirdly, insurgents’ violence perpetration is contingent upon civilian support for armed groups, which shapes groups’ capacity to use selective violence (Humphreys & Weinstein, 2006; Mkandawire, as cited in Kalyvas, 2006, p. 293; Wood, as cited in Rueda, 2017, p. 1628). Gaining civilian support is fundamental to gaining access to information and control. By better controlling civilians, insurgents no longer need to perpetrate massive violence as a reprisal (Rueda, 2017, p. 1638). Relative to the above explanations, groups selectively perpetrating violence are expected to have access to civilian support and information, which likely reduces the amount of human loss and violence levels compared to indiscriminate targeting (Figure 5).

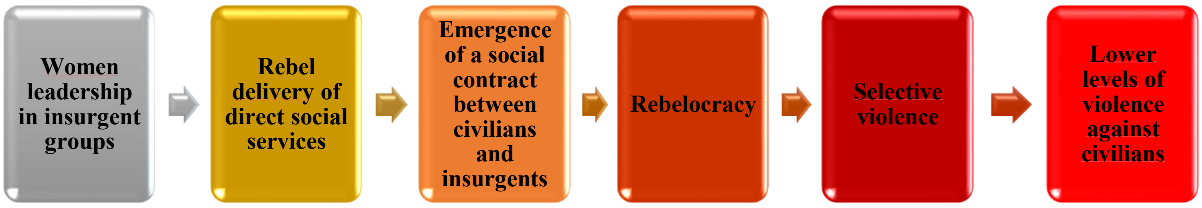
4. Theories' strengths

Firstly, the above theories are relevant for my research as they investigate the dynamics related to my variables of interest. Indeed, Loken's (2018) and Myers' (2025) argument are related to women participation in insurgent leadership positions, corresponding to this paper's explanatory variable. Further, Arjona's (2014) theory provides insight into selective violence, which Koc-Menard (2006) and Rueda (2017) link to civilian violence levels, my outcome variable. Secondly, the theories presented are parsimonious and generalizable, featuring good theories (Halperin and Heath, 2020, p. 130). Indeed, they provide detailed information with a limited number of variables, simplifying complex explanations by prioritizing variables deemed most meaningful. Moreover, the authors generalize their findings to large sets of cases, either using quantitative methods (Loken, 2018; Myers, 2025, Rueda, 2017) or detailing causal steps through qualitative methods (Arjona, 2014; Koc-Menard; 2006). Consequently, the above theories are relevant to investigating my research question.

5. My theory and hypothesis

I theorize that women insurgent leaders build civilian trust and increase social services, creating a broad social contract between insurgents and civilians, called rebelocracy, resulting in higher insurgent information capacity and thus in selective violence's use, reducing civilian violence. See Figure 1.

Figure 1: My suggested theory



To begin, women increase civilian support in the insurgent group and create bonds and trust with local communities (Loken, 2018). As women participation is a broad category which includes women leadership in armed groups, I deduce that women in leadership positions increases the group's inclusion into local communities as well. Such inclusion ensuring collaboration with civilians displays a situation of order between insurgents and civilians. Following Arjona's (2014, p. 1375) typology, the insurgent group hence establishes a social contract with civilians. As women in insurgent leadership promote social policies (Myers, 2025), the scope of the social contract expands beyond the sole military control of the means of violence. Such a broad social contract corresponds to Arjona's (2014) concept of rebelocracy. As rebelocracies decrease violence emergence and favor a selective use of violence (Arjona, 2014), overall levels of violence perpetration hence decrease (Koc-Menard, 2006; Rueda, 2017). See Figure (6).

Following this theory, I expect that larger proportions of women leaders in insurgent groups lead to a decrease in the levels of violence perpetration. A null hypothesis and a main hypothesis stem out of this theory:

H1: Higher proportions of female leaders within insurgent groups decrease the groups' violence perpetration levels.

H0: Female leaders' prevalence within insurgent groups has no effect on violence perpetration levels.

IV. Methodology

In this methodology section, I explain my choice of research design, then justifies how I select the model's data, and I lastly detail how the operationalized.

1. Research design

Quantitative analysis' strength lies in its ability to infer robust generalizations (Halperin & Heath, 2020, p. 410). Indeed, statistical methods explain associations between variables through inference (p. 395). Additionally, large-N designs increase the reliability of the findings, thereby enabling their generalization to a wider population of cases. I run a linear regression model to test my hypothesis. A linear regression summarizes how predictions of a dependent variable (DV) vary across different values of a set of predictor variables (Gelman et al., 2020, p. 4). Hence, a linear regression model delves into the relationship between the explanatory and outcome variables and indicates the direction of the association. Moreover, it indicates the strength of the potential association and the findings' significance. It further provides information about the dataset's standard deviation or error and about the findings' confidence intervals, highlighting the range of values that a certain percentage of cases are likely to fall within (Halperin & Heath, 2020, p. 412). Lastly, a linear regression fits a numeric dependent variable, matching this paper's DV. Consequently, this method adequately fits my research question.

2. Data selection

I am merging four datasets to obtain my final dataset. First, I use the Women's Activities in Armed Rebellion (WAAR) dataset (Loken & Matfess, 2023a), as it contains my independent variable (IV). It is cross-sectional, with data collected from primary and secondary sources, like non-governmental organizations (NGO) and international organizations (IO) reports, regional monitors, and secondary ethnographic and field research on interviews with rebels (Loken & Matfess, 2023b, p. 12). Using such a novel dataset is extremely relevant as its potential has not yet been fully explored.

Second, I use the Uppsala Conflict Data Program (UCDP) one-sided violence dataset (Davies et al., 2025; Eck & Hultman, 2007), as it contains my DV. It collects data from global newswire reporting, global monitoring and translation of local news, social media, NGO and IO reports (Pettersson, 2025, p. 8). I filter out cases where the government perpetrated civilian violence as this paper's unit of analysis is the insurgent group, I further remove cases beyond 2015 to match WAAR's timeline. Then, I merge these first two datasets and summarize the data by insurgent groups, my unit of analysis, to obtain a cross-sectional dataset, and to avoid autocorrelation issues.

Third, I use the Quality of Government (QoG) dataset (Dahlberg et al., 2025; Marshall & Gurr, 2020), containing one of my control variables, government's regime type. Its sources are mainly formal institutional and historical documents, such as legal documents and historical records (Marshall & Gurr, 2020). I filter out post-2015 cases, missing cases, and summarize the dataset by unit of analysis. Yet, no common variable exists between QoG and WAAR or one-sided violence. I hence resort to a dataset which contains a common variable between the datasets one-sided violence and QoG.

Therefore, I fourthly use the Georeferenced Event (GED) Dataset (Sundberg & Melander, 2013), because I can easily merge it with the one-sided violence dataset as it possesses a common variable with QoG and WAAR. Indeed, the variables ccodecow from QoG and gwnoa from GED code the state's country similarly. Consequently, I merge the four datasets and obtain a new dataset containing 74 cases, with each case representing a different insurgent group.

3. Operationalization

To start, I measure the DV, violence perpetrated against civilians, with civilian fatalities. Homicide is often chosen as violence´s proxy for simplicity purposes (Kalyvas, 1999, p. 246). Indeed, violence is conceptually proxied by killings which are expected to reflect overall violence patterns (Liem, 2022, pp. 1-13). This operationalization further fits my conceptualization. Therefore, I use the variable best_fatality_estimate from the one-sided violence dataset and rename it sum_fatalities. This variable indicates the most accurate estimate of the number of civilian deaths occurring during civil wars (Pettersson, 2025, pp. 3-4). Yet I am only interested in civilian deaths perpetrated by insurgent groups. Thus, I exclude fatalities perpetrated by state actors. Subsequently, I recode the binary variable called is_government_actor ranging from 0 to 1 by excluding all cases coded as 1, indicating state actors. Consequently, I obtain my numeric DV ranging from 25 to 19430 with higher values indicating higher civilian deaths thresholds. This variable is highly skewed with a cluster at lower values. Following Benoit's (2011, p. 2) insight, I log the DV before adding it to the linear regression to obtain more linear values, see Table 7 in Appendix B for more detail. The logged DV ranges from about 3.22 to 9.88. Lastly, it is renamed Civilian_deaths.

I measure the IV, women leadership in insurgent groups, with the variable lead_prev_best (Loken & Matfess, 2023b, p. 4). This variable is relevant as it is operationalized by the same authors who conceptualized women leaders in the concepts section. It is an ordinal variable ranging from 0 to 4 and measuring the best estimate of women participation in leadership roles. 0 indicates that no women participation is observed. 1, 2, 3, and 4 respectively indicate that fewer than 5%, between 5% and 9%, between 10% and 19%, more than 20% of women are estimated to have participated in leadership roles. The percentages respectively correspond to occasional, low, moderate, and high levels of women leaders. I factorize this variable before adding it to the model and rename it Women_Leaders.

Moreover, I include control variables in my model. While a vast array of control variables are mentioned in the literature, my dataset contains 76 cases, enabling the use of only three control variables. First, I control the number of conflict years that the dataset includes. Indeed, longer conflicts tend to produce more civilian deaths than shorter ones. Hence, I create a variable called Year, indicating how many years of fighting each unit of analysis contains. Year is numeric and ranges from 1 to 24.

Second, I control for the proportions of women in the insurgent group's non-leadership positions. Indeed, women soldiers are as aggressive and perpetrate the same atrocities as men (Utas, 2005, p. 403). Moreover, groups recruiting women fight longer on average (Henshaw et al., 2019, p. 23), as women participate in supportive and combative roles (Hedström, 2025), which increases civilian deaths. Consequently, I select the variables noncombat_prev_best and frontline_prev_best, respectively indicating the proportion of women in auxiliary roles and combative roles (Loken & Matfess, 2023b, p. 4). Both variables are ordinal and range from 0 to 4, with higher numbers indicating higher proportions of women in the respective positions. By summing these two variables, I create a new variable called Women_nonleaders, ranging from 0 to 8. It indicates the prevalence of women in insurgent groups' non-leadership positions, with higher values corresponding to higher proportions of women non-leaders.

Third, I control the government's regime type as it impacts conflicts' intensity and outcomes (Loken, 2018), further altering deaths perpetration. Indeed, democratic states perpetrate less brutal counterinsurgency campaigns and have better territorial control compared to their non-democratic counterparts (Loken, 2018). Therefore, democracies are less likely to perpetrate mass killings during counterinsurgencies, and are more likely to have the military capabilities to restrain rebels' actions (pp. 46-47). Such characteristics constrain insurgent groups' operations and decrease the armed groups' ability to perpetrate civilian killings. Therefore, I use the continuous variable p_polity2 from the QoG dataset and ranging from -7.2 to 9.5, with higher values indicating more democratic states and with lower values indicating more autocratic states (Dahlberg et al., 2025, p. 232). I rename it Regime.

In summary, my model contains five variables: Civilian_deaths, Women_Leaders, Year, Women_nonleaders, and Regime. Noticeably, the model violates the assumptions of linearity for the control and DV variables and has too many outliers.

V. Analysis

Firstly, I present the results of my research. I start with the descriptive statistics of the IV and DV. Then, I present the results of the linear regression model visualized by a regression table. Further, I analyze the coefficient plot of my model, the model fit determining the most parsimonious and meaningful model, and the predicted values of civilian deaths per prevalence of women leaders. Lastly, I discuss the results in relation to my theoretical framework and contextualize the findings to the broader academic debate.

1. Results

The descriptive statistics of the DV are presented in Table 1**.** The measures of central tendency and dispersion of the DV are detailed for each category of the IV on each line. I only explain how the first line can be understood, but the same logic can be applied to the other lines.

Table 1: Descriptive statistics of civilian deaths by prevalence of women leaders

	Women Leaders	mean	sd	N	median	min	max	Q1	Q3
1	0	707.7097	1506.8535	31	124.0	25	7884	45.0	474.50
2	1	3315.1818	5788.6357	11	1348.0	82	19430	229.0	2404.00
3	2	1081.4375	1038.7541	16	717.5	69	3550	371.5	1357.75
4	3	1794.6000	2340.4967	10	551.0	40	6042	316.5	2227.50
5	4	1003.1667	983.4627	6	911.0	34	2710	276.5	1268.75

Line 1 shows that 31 groups had no women in their leadership. On average, the groups perpetrated about 708 deaths per conflict. The deadliest group perpetrated 7884 civilian deaths compared to 25 for the least violent group. The median is 124, meaning that 15 groups perpetrated fewer than 124 civilian deaths and 15 civilian groups perpetrated more than 124 civilian deaths. 25% of the groups perpetrated fewer than 45 deaths and 75% of the groups killed less than 475 civilians. In other words, 25% of cases perpetrated more than 475 civilian deaths. These groups hence perpetrated much more civilian deaths than the remaining 75% of the groups considering the wide range that occurs between the highest civilian deaths toll and the third quartile. It can thus be expected that influential cases and outliers skew the results. To delve deeper into the association between the variables, the linear regression table presented in Table 2 provides meaningful information about the relationship between the variables.

Table 2: Ordinary linear regression: Predicting (logged) civilian deaths with women leadership, year, women prevalence and regime

	Just Women leadership	Full Model
Intercept	5.189***	3.788***
	(0.281)	(0.245)
Less than 5% of Women In leadership	1.669**	0.250
	(0.549)	(0.417)
Between 5-9% of Women in leadership	1.310**	-0.085
	(0.482)	(0.410)
Between 10-19% of Women in leadership	1.247*	-0.301
	(0.569)	(0.548)
More than of 20% Women in leadership	1.014	-1.310*
	(0.698)	(0.650)
Year		0.216***
		(0.028)
Women prevalence in other positions		0.294***
		(0.085)
Regime type		-0.085**
		(0.027)
Num.Obs.	74	74
R2	0.170	0.686
R2 Adj.	0.122	0.652

- p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

OLS coefficients with standard errors in parentheses. Reference group for Women_Leaders = Null

The first model, called Just Women leadership, only includes the IV and the DV. The second model, called Full Model, additionally controls for women's prevalence in non-leadership positions, insurgencies' duration, and countries' political regime characteristics. Relative to the logged version of the DV, the models' coefficients need to be exponentiated before interpreting the results (UVA Library, n.d.). Table 7 from appendix B refers to the calculation of the variables' coefficients interpreted in the findings below.

The Just Women Leadership model finds a positive association between larger proportions of women insurgent leaders and higher levels of violence perpetration against civilians. Yet, the model omits crucial control variables, undermining the reliability of the results. The lack of credibility of the model translates into unreliable coefficients. According to the model, insurgent groups with fewer than 5%, 5% to 9%, 10% to 19%, and more than 20% of female leaders on average kill 430.3%, 270.6%, 248.0% and 175.7% more civilians respectively than groups with no women leaders. While the results for the first three categories are statistically significant ($p<0.05$), they are not statistically significant for the fourth category ($p>0.05$). Moreover, no theoretical explanation revolving around women's leadership positions in insurgent groups supports such large increases in civilian deaths. Analyzing the full model's results is hence more meaningful and reliable than focusing on the incomplete model.

After adding the necessary control variables to the hypothesis, the findings point to a different direction. The Full Model shows that groups containing at least 20% of women leaders significantly reduce civilian violence. On average, groups with a higher prevalence of women leaders perpetrate fewer civilian deaths, but the results only become significant once the proportion of women leaders reaches a fifth of the leadership size.

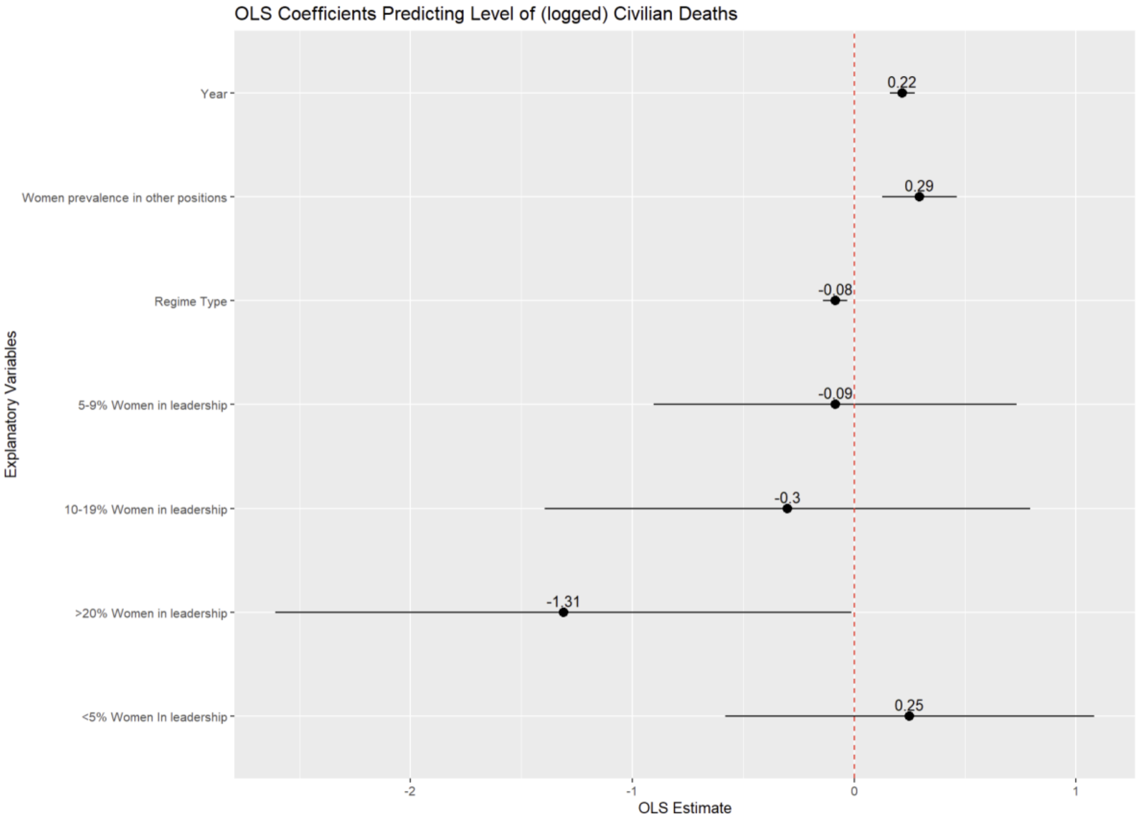
More specifically, insurgent groups with up to 5% of female leaders kill, on average, 28.4% more civilians than groups with no women leaders with equivalent women prevalence in non-leadership positions, insurgencies' duration, and occurring in countries with equivalent political regime characteristics. The difference is not statistically significant at conventional levels ($p>0.05$). On average, insurgent groups containing 5% and 9% of female leaders kill 8.2% less civilians than groups with no women leaders, keeping constant the effects of the three control variables mentioned above. The difference is not statistically significant at conventional levels ($p>0.05$). Moreover, insurgent groups having between 10% and 19% of female leaders on average kill 26% less civilians than groups with no women leaders, keeping constant the effects of the three control variables mentioned above. The difference is not statistically significant at conventional levels ($p>0.05$). Lastly, insurgent groups with more than 20% of female leaders on average kill 73% less civilians than groups who are not led by any women with equivalent women prevalence in non-leadership positions, insurgencies' duration, and occurring in countries with equivalent political regime characteristics. The difference is statistically significant at conventional levels ($p<0.05$). In short, the average perpetration of civilian violence by insurgent group decreases as the prevalence of women in the groups exceeds 20% while other explanatory factors remain constant.

I now look at the coefficient from the control variables. We expect the level of civilian deaths to increase as the conflict endures, holding constant the effect of other explanatory variables. Based on this model, the level of civilian deaths is expected to increase by 24.1% on average with each one-year increase, and keeping constant the effects of women prevalence in non-leadership positions, women prevalence in leadership positions, the country's political regime characteristics. This positive association is statistically significant ($p<0.001$). Further, we expect the level of civilian deaths to decrease as countries get more democratic, holding constant the effects of the other explanatory variables. Based on this model, the level of civilian deaths is expected to decrease by 8.2% on average with each one unit increase in the government's polity score, and keeping constant the effects of women prevalence in non-leadership positions, women prevalence in leadership positions, and insurgencies' duration. This negative association is statistically significant ($p<0.01$). Moreover, we expect the level of civilian deaths to increase as prevalence of women in non-leadership positions in insurgencies increases, holding constant the effects of the other explanatory variables. Based on this model, the level of civilian deaths is expected to have an average increase of 34.2% with each one unit increase in women's prevalence in non-leadership positions, and keeping constant the effects of women prevalence in leadership positions, the country's political regime characteristics and insurgencies' duration. This positive association is statistically significant ($p<0.001$).

In short, controlling for other explanatory factors, civilian deaths are expected to increase when female leaders comprise at least one fifth of the insurgent's leadership group, as the conflict endures, as more women occupy non-leadership roles, and as the government's type of regime becomes less democratic.

Analyzing the model's coefficient plot found Table 3 next, is useful as it depicts the graphical overview of the explanatory variables' coefficients and their confidence intervals (Demarest & Robison, 2025, Chapter 8). Such a plot facilitates the visualization of the association between the DV and the IV. It indicates the direction of the association with the coefficient's sign and the confidence intervals through the display of horizontal black lines. Lastly, it details the statistical significance of each association. Indeed, coefficients are statistically significant when the confidence interval does not intersect with the red line.

Table 3: Coefficient plot



Notes: Markers provide the OLS coefficient for each variable with 95% confidence intervals. See Table 2 for full results.

Firstly, the positive coefficient 0.25 indicates that insurgent groups containing up to 5% of women leaders perpetrate, on average, more civilian deaths than insurgent groups with no women leaders. Then, the negative coefficients -0.09 and -0.6 indicate that insurgent groups respectively containing 5%-9% and 10%-19% of women leaders perpetrate, on average, fewer civilian deaths than insurgent groups with no women leaders. The three associations between each of the first three IV's categories and the DV are not statistically significant as the 95% confidence intervals include the value of 0, failing to reject the null hypothesis of no effect. Lastly, the negative coefficient -1.31 indicates that insurgent groups containing more than 20% of women leaders on average perpetrate fewer civilian deaths than insurgent groups with no women leaders. The association is statistically significant as the 95% confidence interval does not include the value of 0, failing to support the null hypothesis of no effect.

Next, we look at the control variables' coefficients. The variable controlling for the conflict's length is associated with an increase in civilian deaths when keeping other variables constant, as indicated by the positive coefficient of 0.22. Additionally, the variable controlling for the prevalence of women in insurgent group's non-leadership positions is associated with an increase in civilian deaths when keeping other variables constant, as the positive coefficient of 0.29 shows. Last, more democratic countries are associated with a decrease in civilian deaths keeping other variables, as indicated by the negative coefficient of -0.08. The three associations between each control variable and the DV are statistically significant as the 95% confidence intervals do not include the value of 0, failing to support the null hypothesis of no effect.

In summary, longer conflicts with more women occupying non-leadership roles significantly increase civilian violence. Conversely, when insurgent groups comprise at least 20% of women leaders in their leadership and reside in more democratic countries, civilian violence decreases, supporting this paper's main hypothesis.

In the following, I compare the fit of different models against the null model with the Anova test, and look at the proportion of variance in civilian deaths that each model explains (see Table 4). I am testing five models with model 1 not having an explanatory variable and each further model having one extra predictor variable than the one before. A model in a row is more meaningful and complex than the model in the preceding row when the p-value of the F-statistic is statistically significant. It is hence concluded that this model is a better fit than all preceding models.

Table 4: Analysis of variance table

```
Model 1: Civilian_deaths ~ 1
Model 2: Civilian_deaths ~ Women_Leaders
Model 3: Civilian_deaths ~ Women_Leaders + Year
Model 4: Civilian_deaths ~ Women_Leaders + Year + Women_nonleaders
Model 5: Civilian_deaths ~ Women_Leaders + Year + Women_nonleaders + Regime
  Res.Df     RSS Df Sum of Sq       F    Pr(>F)    
1     73 203.517                                   
2     69 168.959  4    34.558  8.9134 8.117e-06 ***
3     68  81.026  1    87.933 90.7208 5.134e-14 ***
4     67  73.491  1     7.535  7.7737  0.006918 ** 
5     66  63.972  1     9.519  9.8208  0.002574 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value of the F-statistics of each model shows that model 2 has a better fit than model 1, model 3 has a better fit than model 2 and 1, model 4 has a better fit than the first three models, and model 5 has a better fit than the first four models. In summary, the best fitting and most parsimonious model is the fifth, corresponding to the Full Model that was used in the linear regression discussed above. The Full Model explains the highest variance in civilian deaths with the least number of predictors. As showcased earlier in Table 2, the Full Model's R-squared value explains 65.2% of the variance in civilian deaths.

Table 5: Predicted values

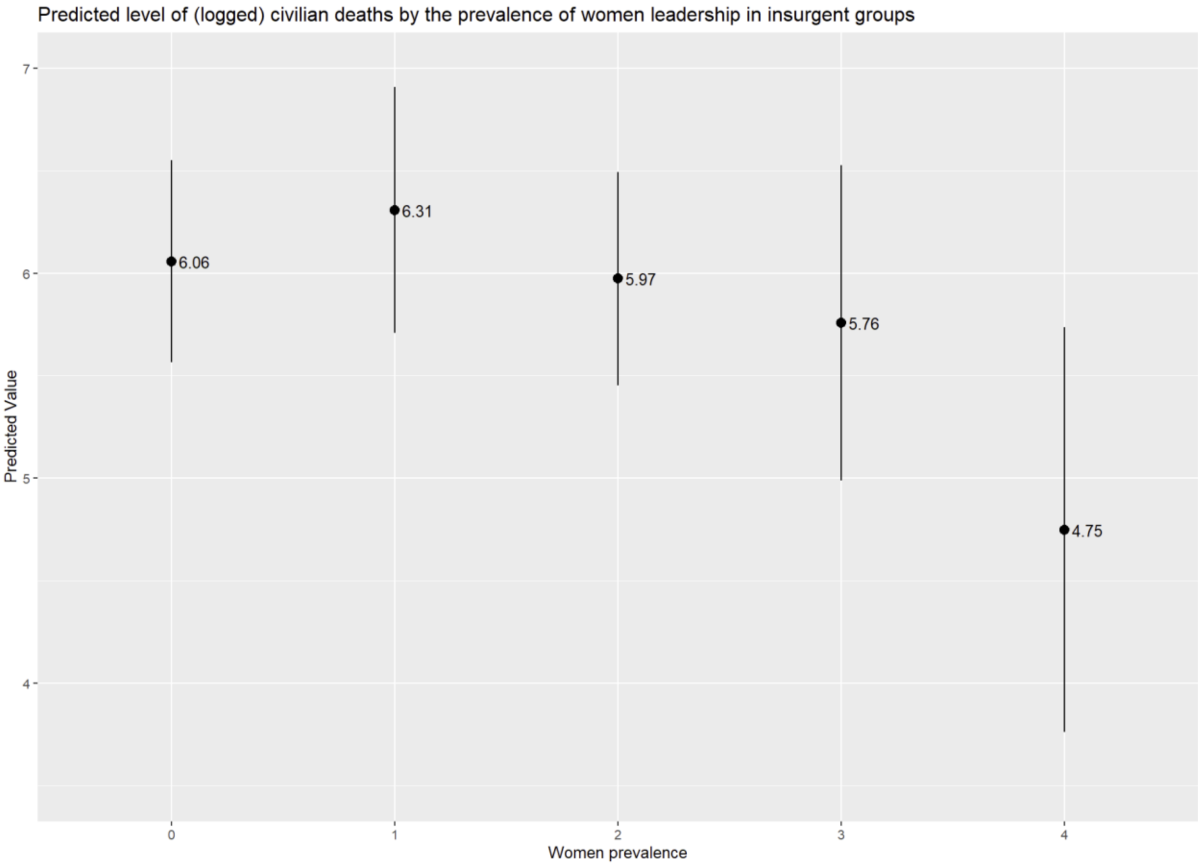


Table 5 above presents the results of the DV's predicted value plot per category of the IV to delve into the potential substantive meaning of the relationship between the two variables (Demarest & Robison, 2025, Chapter 8). The predicted values plot enables the visualization of the predicted effects that distinct proportions of women leaders in insurgent groups have on civilian deaths. The overall downward trend of the coefficients and of the confidence intervals, represented by the vertical lines, illustrate that the predicted value of civilian deaths decreases as more than 5% of women participate in leadership positions of insurgent groups, while keeping other explanatory factors constant. The predicted values of the four categories of women prevalence are compared to the baseline reference category with no women leaders. On the one hand, there is no clear separation between the confidence intervals of the reference category and of the first three categories as the vertical lines representing the confidence intervals overlap considerably. On the other hand, the reference category's confidence interval displays only limited overlap with the confidence interval of the fourth category. This suggests that when insurgent groups have more than 20% of women leaders, the predicted number of civilian deaths decreases considerably, and that the association is meaningful.

To summarize the results, the difference in civilian killings between groups with no women leaders and groups with women leaders becomes meaningful when the percentage of women leaders exceeds 20%, keeping other explanatory variables constant. Consequently, I predict that insurgent groups with at least one fifth of women leaders, or high levels of women leaders, perpetrate fewer civilian killings than groups without female leaders. In the following part, I analyze these results and contextualize them in the larger academic debate.

2. Discussion

The empirical results partly support and call for nuance in scrutinizing the hypothesis. Insurgent groups with at least one fifth of women leaders are less violent against civilians than groups with no women leaders, keeping other factors constant. Yet, below this threshold, there is no meaningful difference between groups with or without women leaders. Indeed, the results are not significant when women form fewer than 20% of the group's leadership size. Consequently, the findings support that larger proportions of women leaders decrease civilian violence, but this association occurs only when women leaders constitute at least 20% of the group's leadership. This finding is consistent with "critical mass" theories, holding that women must form sufficient proportions of political groups before being able to change political outcomes (Dahlerup, as cited in Beckwith & Cowell-Meyers, 2007, p. 553). More precisely, a minority of women constrains their ability to bring about political change. Thomas (2024, p. 844) confirms that larger groups of women with greater political participation have a larger influence on the insurgency's decisions and outcomes. Indeed, it is more challenging for small groups of women leaders to bring about political change. Subsequently, this paper's statistical results are explained by critical mass theories and conditionally support the hypothesis.

Further, comparing the explanatory variables' standardized coefficient indicates which predictor has the strongest association on civilian death levels (see Table 6).

Table 6: Standardized coefficients

Parameter	Std. Coef.	95% CI
(Intercept)	0.08	[-0.23, 0.38]
Women Leaders [1]	0.15	[-0.35, 0.65]
Women Leaders [2]	-0.05	[-0.54, 0.44]
Women Leaders [3]	-0.18	[-0.83, 0.47]
Women Leaders [4]	-0.78	[-1.56, -0.01]
Year	0.61	[0.45, 0.76]
Women nonleaders	0.47	[0.20, 0.74]
Regime	-0.25	[-0.41, -0.09]

On average, the shift from no women to high female participation in insurgent leadership is associated with a larger decrease on civilian deaths (-0.78 standard deviation (SD)) than one SD increase in any other explanatory variables, which all have effect sizes smaller in absolute value. Hence, an insurgency's duration, the prevalence of women in other positions, or groups acting in countries with more democratic regimes are on average less strongly associated with civilian killings than having 20% of leadership positions held by women in insurgent groups. These results fit both the theoretical framework and expectations. That is, high levels of women leaders decrease groups' civilian deaths perpetration and this variable's effect is larger than the effect sizes from other predictor variables.

This research engages with ongoing academic debates as it shows that female leaders shape insurgencies' violence perpetration levels. Hence, it provides empirical support for Mampilly's (2011, p. 232) argument that insurgent governance plays a crucial role in shaping violence perpetration. Moreover, it is consistent with Henshaw's (2019) argument which states that female leadership impacts insurgencies' outcomes. Additionally, the findings suggest one answer to Loken's and Matfess' (2023a) questioning about how various levels of women participation in leadership roles impact insurgents' strategies. Indeed, the results underline that high levels of women's participation in insurgent leadership roles shape the groups' decision to restrict violence against civilians. The study further sheds light on the debate relative to the association between women leadership and rebel order researched by Myers (2025).

Ultimately, the findings are consistent with this paper's theory and fail to support the null hypothesis of no difference between groups with high proportions of women leaders and groups with few to no women leaders. Rather, the findings indicate that leaderships of insurgent groups comprising at least one fifth of women play a key role in shaping civilian violence.

To complement the quantitative findings and contextualize this research, I turn to the case of the Lord's Resistance Army (LRA) in Uganda. The LRA is both an influential case and an outlier, highlighting the relevance to investigating this case's context. LRA's leadership included a small proportion of women, and the group perpetrated widespread civilian killings between 1989 and 2015, in line with the finding that a higher proportion of women insurgent leaders decreases civilian violence. This case stands out as an outlier due to its long duration and large number of civilian fatalities. Indeed, LRA insurgents killed about 8000 civilians over a time span of 24 years, corresponding to about 326 civilian killings per year, which makes this group deadlier than 75% of the dataset's insurgent groups. The group relied on extreme forms of violence against civilians, such as indiscriminate killings or mutilations (Annan et al., 2009, p. 11).

The LRA emerged in 1988 and reproduced gender inequalities in its organization and norms (Annan et al., 2009, p. 21). Indeed, the dataset shows that fewer than 5% of women reached leadership positions while women were comparatively more abundantly present in other positions; between 5% and 9% of women were frontline fighters and between 10% and 19% of women had noncombat roles. Overall, women occupied primary or secondary roles as fighters (p. 9). Others completed logistical tasks in the group, such as nursing, or overseeing radio communication. Following Myer's (2025) findings, such a small proportion of women leaders jeopardizes the group's issuing of direct social services, which in turn undermines civilian support for the group. Annan et al. (2009, p. 11) confirm that the LRA lost civilian support after resorting to abductions and indiscriminate violence. Hence, the loss of civilian support compromised the creation of an encompassing social contract (Arjona, 2014). Lacking considerable popularity and material resources, the LRA widely resorted to indiscriminate violence, leading to widespread civilian killings.

In short, LRA's case supports my theory. The lack of women leaders hindered the LRA's creation of a social contract with civilians. This encouraged the group to favor large-scale abductions and indoctrination to recruit members rather than civilian support (Annan et al., 2009, pp. 4-11), and to perpetrate indiscriminate violence to target civilians, which increased the civilian death toll. Therefore, this case fails to discredit the hypothesis, providing empirical support for a negative association between the presence of women as insurgent leaders and civilian violence perpetration.

VI. Conclusion

To conclude, women participation in insurgent groups' leadership is expected to shape insurgents' strategies and behaviors. More specifically, insurgent groups with high proportions of women in leadership positions on average commit significantly fewer civilian killings than groups with no women leaders. Hence the results fail to invalidate the null hypothesis of no difference, supporting this thesis' hypothesis that moving from low to high levels of women participation in leadership positions of insurgencies decreases civilian violence.

This research is relevant for several reasons. First, it bridges two academic gaps identified by scholars (Henshaw, 2019; Kalyvas, 2006; Loken & Matfess, 2023a). On the one hand, it studies how women in leadership roles impact insurgent groups' decisions and hence civil wars' outcomes. On the other hand, it explains how women leadership impacts the dynamics of civilian violence. Second, this empirical research provides new insights on the topic. Indeed, it compiles a novel theory explaining the causal mechanism linking women insurgent leaders to insurgent groups' restrained perpetration of civilian violence. Moreover, it develops a statistical model combining data from several datasets that allows for scrutinizing the thesis' hypothesis.

While this research finds meaningful results and is inscribed in the larger academic debate, it is necessary to consider some important limitations. First, collecting data on insurgent groups is challenging due to limited access to information. Hence, the final model was run with a limited number of cases, that is 74 cases for the full model. Having a small sample might introduce bias in the statistical model and provide inadequate results. Moreover, the limited sample size allowed for only adding three control variables, meaning that the model might miss some critical confounding factors, such as the insurgent group's size, the group's military capacity, or the type of terrain. Second, some ordinary linear regression assumptions are violated, such as the linearity of the control variables and the excess of outliers, which might bias the results. Third, while quantitative research has considerable assets, it also has downsides. Indeed, it prioritizes generalizability to contextual explanations, which might overlook some relevant mechanisms impacting the association between the variables and which hinders the possibility to test the suggested mechanism.

Relative to these limitations, I encourage future research to delve deeper into the topic to enhance our understanding of the impact of women leaders on insurgencies' outcomes. More specifically, this paper provides a novel theory that requires further testing. First, additional data collection efforts should be made to compile larger datasets to overcome the small-N issues that limit this research's findings. Second, studying the question from a qualitative lens would be most useful to further comprehend contextual dynamics impacting the relationship between female insurgent leaders and civilian violence perpetration. Process tracing would be a very relevant method to study this research agenda as the theory highlights a detailed causal mechanism.

Bibliography:

Alison, M. (2004). Women as agents of political violence: Gendering security. *Security Dialogue*, *35*(4), 447-463.

Annan, J., Blattman, C., Mazurana, D., & Carlson, K. (2009). Women and girls at war: 'Wives', mothers and fighters in the Lord's Resistance Army. *Households in Conflict Network Working Paper*, *63*.

Arjona, A. (2014). Wartime institutions: A research agenda. *Journal of Conflict Resolution, 58*(8), 1360–1389.

Armed Conflict Location & Event Data Project (ACLED). (2023). *ACLED codebook (Version 11)*. ACLED.

Asal, V., Legault, R., Szekely, O., & Wilkenfeld, J. (2013). Gender ideologies and forms of contentious mobilization in the Middle East. *Journal of Peace Research, 50*(3), 305–318.

Balcells, L. (2017). A Theory of Violence Against Civilians. In *Rivalry and Revenge: The Politics of Violence during Civil War* (pp. 19–46). chapter, Cambridge: Cambridge University Press.

Balcells, L., & Stanton, J. A. (2021). Violence against civilians during armed conflict: Moving beyond the macro- and micro-level divide. *Annual Review of Political Science, 24*(1), 45–69.

Barter, S. J. (2012). Unarmed forces: Civilian strategy in violent conflicts. *Peace & Change, 37*(4), 544–571.

Beckwith, K., & Cowell-Meyers, K. (2007). Sheer Numbers: Critical Representation Thresholds and Women's Political Representation. *Perspectives on Politics*, *5*(3), 553–565.

Braithwaite, A., & Ruiz, L. B. (2018). Female combatants, forced recruitment, and civil conflict outcomes. *Research & Politics, 5*(2), 2053168018770559. https://doi.org/10.1177/2053168018770559

Caprioli, M. (2005). Primed for violence: The role of gender inequality in predicting internal conflict. *International Studies Quarterly, 49*(2), 161–178.

Cohen, D. K. (2013). Female combatants and the perpetration of violence: Wartime rape in the Sierra Leone civil war. *World Politics*, *65*(3), 383-415.

Dahlberg, S., Sundström, A., Holmberg, S., Rothstein, B., Alvarado Pachon, N., Dalli, C. M., Lopez Valverde, R., Phiri, V. S., & Gerber, L. (2025). *The Quality of Government Basic Dataset* (Version Jan25) [Data set]. University of Gothenburg, Quality of Government Institute. https://www.gu.se/en/quality-government doi:10.18157/qogbasjan25

Davies, S., Pettersson, T., Sollenberg, M., & Öberg, M. (2025). Organized violence 1989-2024, and the challenges of identifying civilian victims. *Journal of Peace Research, 62*(4), 1223–1240.

Demarest, L., & Robison, J. (2025). *Reporting and presenting results* (Chapter 8). In *Statistical analysis and presentation using R: An overview of R functions used in Statistics II*. Statistical analysis and presentation using R

Downes, A. B. (2006). Desperate times, desperate measures: The causes of civilian victimization in war. *International Security, 30*(4), 152–195. https://doi.org/10.1162/isec.2006.30.4.152

Duyvesteyn, I., & Rich, P. B. (2012). *The study of insurgency and counterinsurgency.* In P. B. Rich & I. Duyvesteyn (Eds.), *The Routledge handbook of insurgency and counterinsurgency* (pp. 1–15). London, England: Routledge.

Eck, K., & Hultman, L. (2007). One-sided violence against civilians in war: Insights from new fatality data. *Journal of Peace Research, 44*(2), 233–246.

Eisler, R., & Loye, D. (1986). Peace and feminist theory: New directions. *Bulletin of Peace Proposals*, *17*(1), 95-99.

Enloe, C. (2014). *Bananas, beaches and bases : Making feminist sense of international politics*. University of California Press.

Fujii, L. A. (2008). The power of local ties: Popular participation in the Rwandan genocide. *Security Studies, 17*(3), 568–597.

Gelman, A., Hill, J., & Vehtari, A. (2020). *Regression and other stories*. Cambridge University Press.

Halperin, S. & Heath, O. (2020). *Political research: Methods and practical skills* (3rd edition). Oxford University Press.

Hedström, J. (2025). Feminist perspectives on rebel governance and civil wars. *Civil Wars, 27*(1), 1–14. https://doi.org/10.1080/13698249.2025.2463049

Henshaw, A., Eric-Udorie, J., Godefa, H., Howley, K., Jeon, C., Sweezy, E., & Zhao, K. (2019). Understanding women at war: A mixed-methods exploration of leadership in non-state armed groups. *Small Wars & Insurgencies, 30*(6–7), 1089–1116.

Högbladh, S. (2025). *UCDP GED codebook (Version 25.1)*. Department of Peace and Conflict Research, Uppsala University. https://ucdp.uu.se/downloads/

Huber, L. (2018). When civilians are attacked: Gender equality and terrorist targeting. *Journal of Conflict Resolution.* https://doi.org/10.1177/0022002719835601

Hudson, V., Bowen, D. L., & Nielsen, P. L. (2019). *The first political order: How sex shapes governance and national security worldwide.* New York: Columbia University Press.

Hultman, L. (2012). *Violence against civilians*. In P. B. Rich & I. Duyvesteyn (Eds.), *The Routledge handbook of insurgency and counterinsurgency* (pp. 289–299). Routledge.

Humphreys, M., & Weinstein, J. M. (2006). Handling and manhandling civilians in civil war: Determinants of the strategies of warring factions. *American Political Science Review, 100*(3), 429–447.

Jordan, K., & Denov, M. (2007). Birds of freedom? Perspectives on female emancipation and Sri Lanka's Liberation Tigers of Tamil Eelam.

Kalyvas, S. N. (1999). Wanton and senseless? The logic of massacres in Algeria. *Rationality and society*, *11*(3), 243-285.

Kalyvas, S. N. (2006). *The logic of violence in civil war.* Cambridge: Cambridge University Press.

Koc-Menard, S. (2006). Switching from indiscriminate to selective violence: The case of the Peruvian military (1980–95). *Civil Wars*, *8*(3-4), 332-354.

Kurdistan au féminin. (n.d.). *Sakine Cansiz organizing women* [Photograph]. *Kurdistan au féminin*. https://kurdistan-au-feminin.fr/sources/

Kurdistan au féminin. (n.d.). *Foundation of the PKK* [Photograph]. *Kurdistan au féminin.* https://kurdistan-au-feminin.fr/sources/

Lewis, J. I. (2017). How does ethnic rebellion start. *Comparative Political Studies, 50*(10), 1420–1450.

Loken, M. (2018). *Women in war: Militancy, legitimacy, and rebel success.* [Unpublished doctoral dissertation]. University of Wisconsin–Madison.

Loken, M. (2022). Noncombat participation in rebellion: A gendered typology. *International Security, 47*(1), 139–170.

Loken, M., & Matfess, H. (2023a). Introducing the women's activities in armed rebellion (WAAR) project, 1946–2015. *Journal of Peace Research, 61*(3), 489–499. https://doi.org/10.1177/00223433221128340

Loken, M., & Matfess, H. (2023b). *Women's activities in armed rebellion (WAAR) project dataset v1.0 codebook* [Codebook]. University of Amsterdam; Josef Korbel School of International Studies, University of Denver

Malone, I. (2022). Unmasking militants: Organizational trends in armed groups, 1970–2012. *International Studies Quarterly, 66*(3), Article sqac050.

Mampilly, Z. (2011). *Rebel rulers: Insurgent governance and civilian life* (Chapters 1 & 8). Ithaca, NY: Cornell University Press.

Marshall, M. G., & Gurr, T. R. (2020). Polity v project, political regime characteristics and transitions, 1800-2018.

Mehrl, M. (2023). Female combatants and rebel group behaviour: Evidence from Nepal. *Conflict Management and Peace Science*, *40*(3), 260–280. https://doi.org/10.1177/07388942221082234

Myers, E. (2025). Gendered governance: How women in insurgent government shape rebel rule. *Civil Wars, 27*(1), 189–222. https://doi.org/10.1080/13698249.2024.2397766

Pettersson, T. (2025). *UCDP one-sided violence codebook (Version 25.1)* [Data set]. Uppsala Conflict Data Program. (https://ucdp.uu.se/downloads/).

Rueda, M. R. (2017). Popular support, violence, and territorial control in civil war. *Journal of Conflict Resolution*, *61*(8), 1626-1652.

Slim, H. (2007). *Killing civilians: Method, madness and morality in war*. London: Hurst.

Sundberg, R., & Melander, E. (2013). Introducing the UCDP Georeferenced Event Dataset. *Journal of Peace Research, 50*(4), 523–532. https://doi.org/10.1177/0022343313484347

Thaler, K. M. (2012). Ideology and violence in civil wars: Theory and evidence from Mozambique and Angola. *Civil Wars, 14*(4), 546–567.

Thomas, J. L. (2024). Sisters Are Doing It for Themselves: How Female Combatants Help Generate Gender-Inclusive Peace Agreements in Civil Wars. *American Political Science Review*, *118*(2), 831–847.

Thomas, J. L., & Wood, R. M. (2017). *Women in Armed Rebellion Dataset (WARD).* https://www.wardproject.com

Utas, M. (2005). Victimcy, girlfriending, soldiering: Tactic agency in a young woman's social navigation of the Liberian war zone. *Anthropological Quarterly, 78*(2), 403–430.

UVA Library. (n.d.). Interpreting log transformations in a linear model. Interpreting Log Transformations in a Linear Model | UVA Library

Viterna, J. S. (2015). Women's mobilization into the Salvadoran guerilla army. In J. Goodwin & J. M. Jasper (Eds.), *The social movements reader* (pp. 83–99). Sussex, UK: Wiley Blackwell.

Winward, M. (2021). Intelligence capacity and mass violence: Evidence from Indonesia. *Comparative Political Studies, 54*(3–4), 553–584.

Appendix A- Figures

Figure 2: Loken's theory

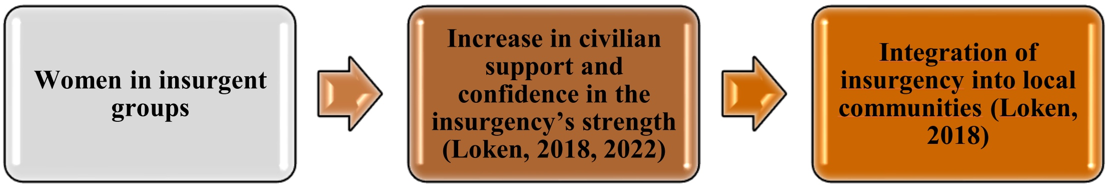


Figure 3: Myers' theory

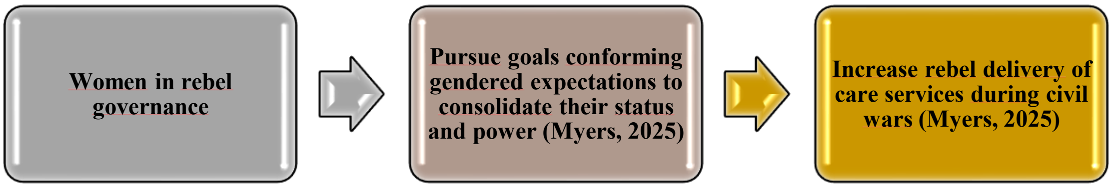


Figure 4: Arjona's theory

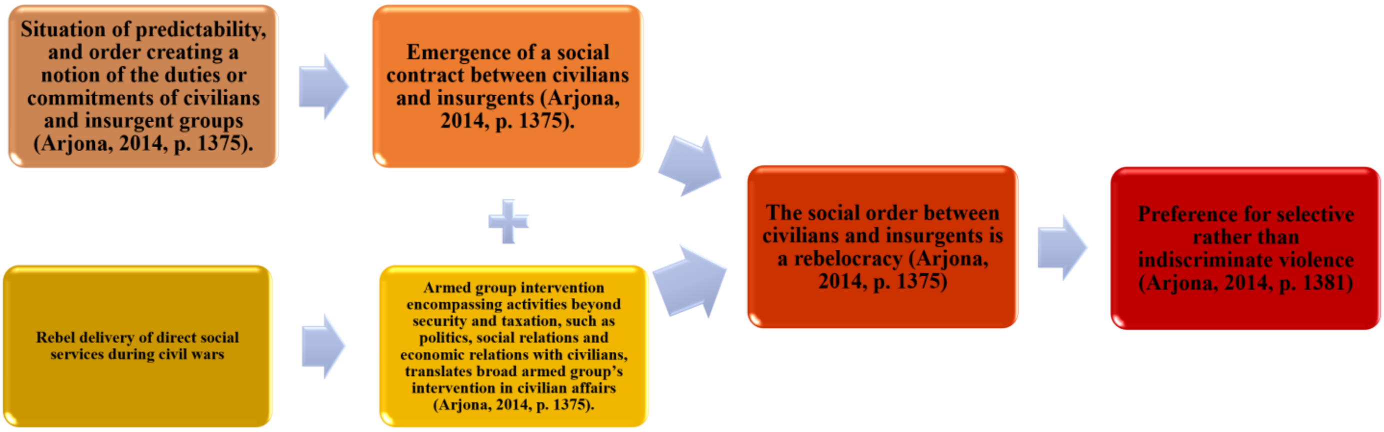


Figure 5: Koc-Menard´s and Rueda´s theories

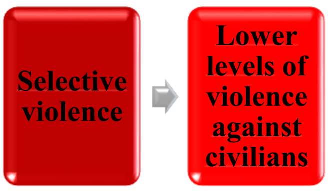
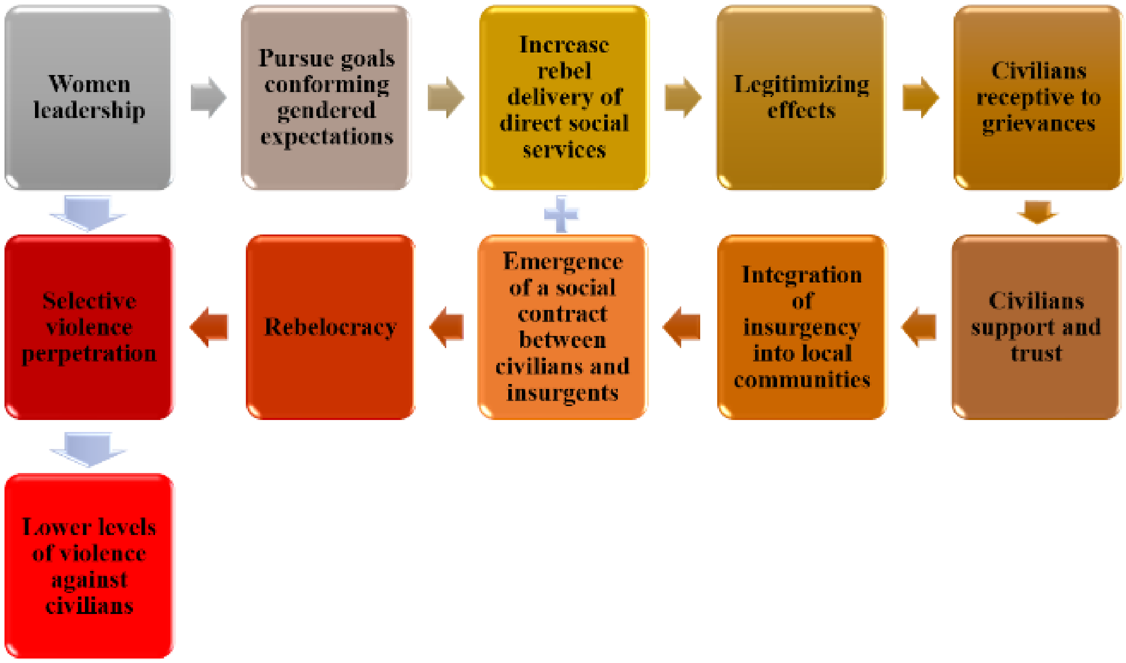


Figure 6: Detailed proposed theory



Appendix B – Tables and OLS assumptions

Table 7: Converting coefficient to interpret the results

Variables	Just Women leadership	Full Model
Less than 5% of Women In leadership	(exp(1.669) - 1) X 100 = 430.3	(exp(0.250) – 1) X 100 = 28.4
Between 5-9% of Women in leadership	(exp(1.310) - 1) X 100 = 270.6	(exp(-0.085) – 1) X 100 = -8.2
Between 10-19% of Women in leadership	(exp(1.247) - 1) X 100 = 248.0	(exp(-0.301) – 1) X 100 = -26.0
More than of 20% Women in leadership	(exp(1.014) – 1) X 100 = 175.7	(exp(-1.310) – 1) X 100 = -73.0
Year		(exp(0. 216) – 1) X 100 = 24.1
Women prevalence in other positions		(exp(0.294) – 1) X 100 = 34.2
Regime type		(exp(-0.085) – 1) X 100 = -8.2

1. ***Transformation of the DV***

I notice that the values are much dispersed across the DV. The median, 246.5, is much lower than the mean, 1338.0, and the highest value, 19430, is critically higher than the third quartile, 1413.2.

```
> summary(C1$sum_fatalities)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   25.0    88.0   246.5  1338.0  1413.2 19430.0
```

Hence, I compare the dot plots of the non-transformed DV and the logged transformed DV. See Figures 7 and 8.

Figure 7: Dot plot non-transformed DV

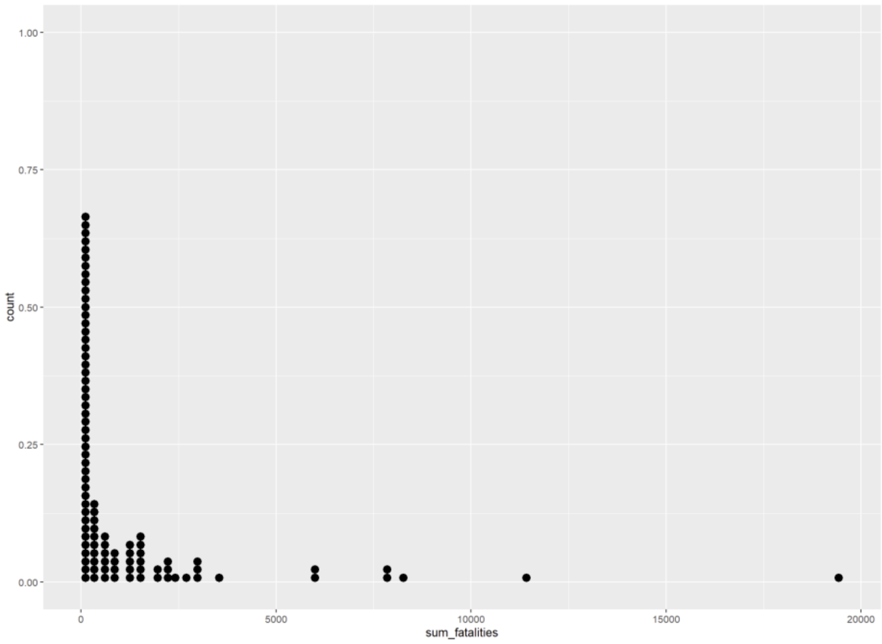
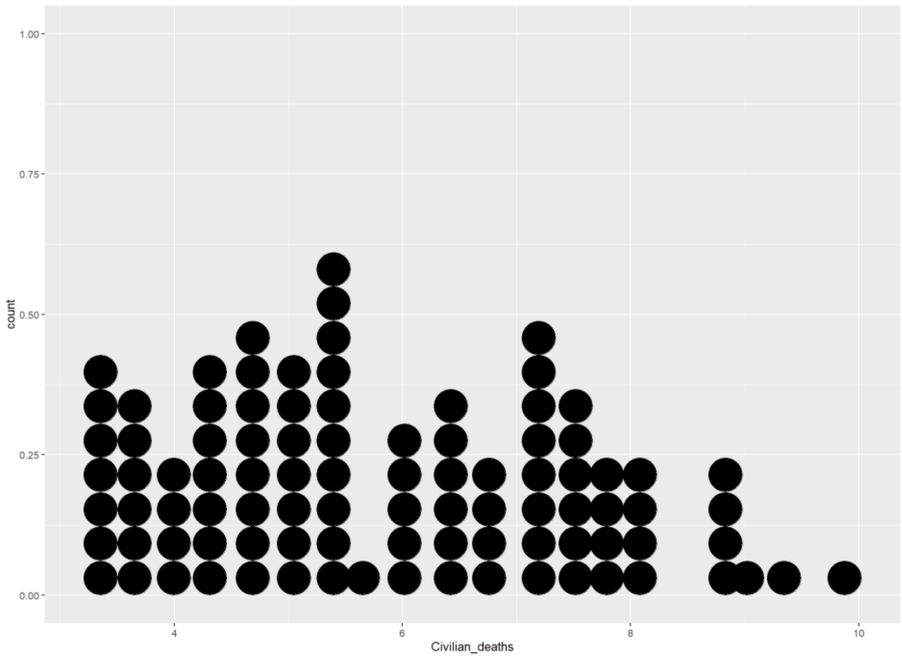


Figure 8: Dot plot logged DV



While the initial version of the DV is highly skewed on the left corner with an accumulation of dots at lower values, the logged transformed DV displays much more constant dots across the graph, with no substantial gap across values. I further investigate the summary of the logged transformed DV.

```
> summary(C2$Civilian_deaths)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  3.219   4.477   5.507   5.852   7.253   9.875 
```

The values of the median and of the mean are closer from each other than earlier, there is no substantial difference between the lowest value and the first quartile and between the highest value and the third quartile. Indeed, logarithmic transformations are convenient means of converting a highly skewed variable into a more normal one (Benoit, 2011, p. 2). Consequently, the logged version of the DV is more relevant than the non-transformed version.

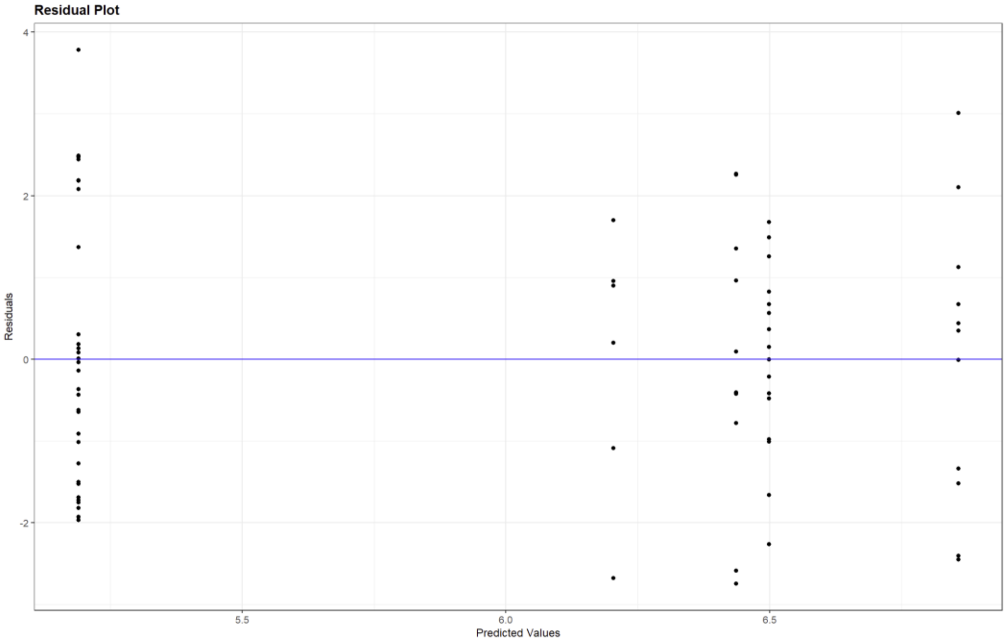
2. *OLS assumptions for the Just Women leadership model*

The model Just Women leadership from Table 2 is referred to as Test in the R codes. First, no test for multicollinearity is run. Indeed, the model contains no control variable. There is therefore no risk of encountering the issue of multicollinearity. Second, the model's Durbin Watson statistic has a value of about 2.01 and is therefore comprised between 1 and 3. Consequently, there is no issue with autocorrelation.

```
> durbinWatsonTest(Test)
 lag Autocorrelation D-W Statistic p-value
   1     -0.04069553      2.014761   0.978
 Alternative hypothesis: rho != 0
```

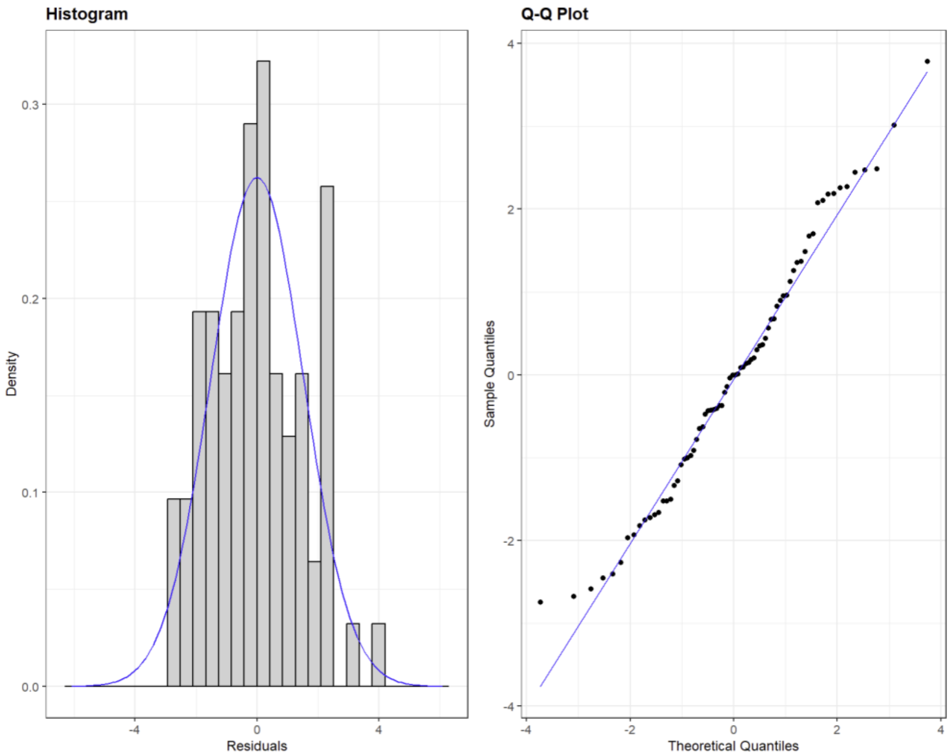
Third, the IV's residual plot, Figure 9, shows accumulation of dots at five values of the IV, consistent with the five categories of the variable. There appears to be an accumulation of dots in the left of the plot, indicating that some type of non-linear pattern might impact the linearity assumption. While the linearity assumption is not critically violated, the results do suggest some mild violation of the assumption.

Figure 9: Residual plot IV



Fourth, the histogram and Q-Q plot indicate that the model's residuals follow a normal distribution convincingly. The errors therefore appear to be normally distributed, and this assumption is not clearly violated.

Figure 10: Histogram and Q-Q plot (Test)



Fifth, the quantity of outliers is investigated. There are less than 5% of cases with a standardized residual higher than the absolute value of 1.96, and no case with a standardized residual higher than the absolute value of 2.58. Therefore, the model does not present excessive amounts of outliers.

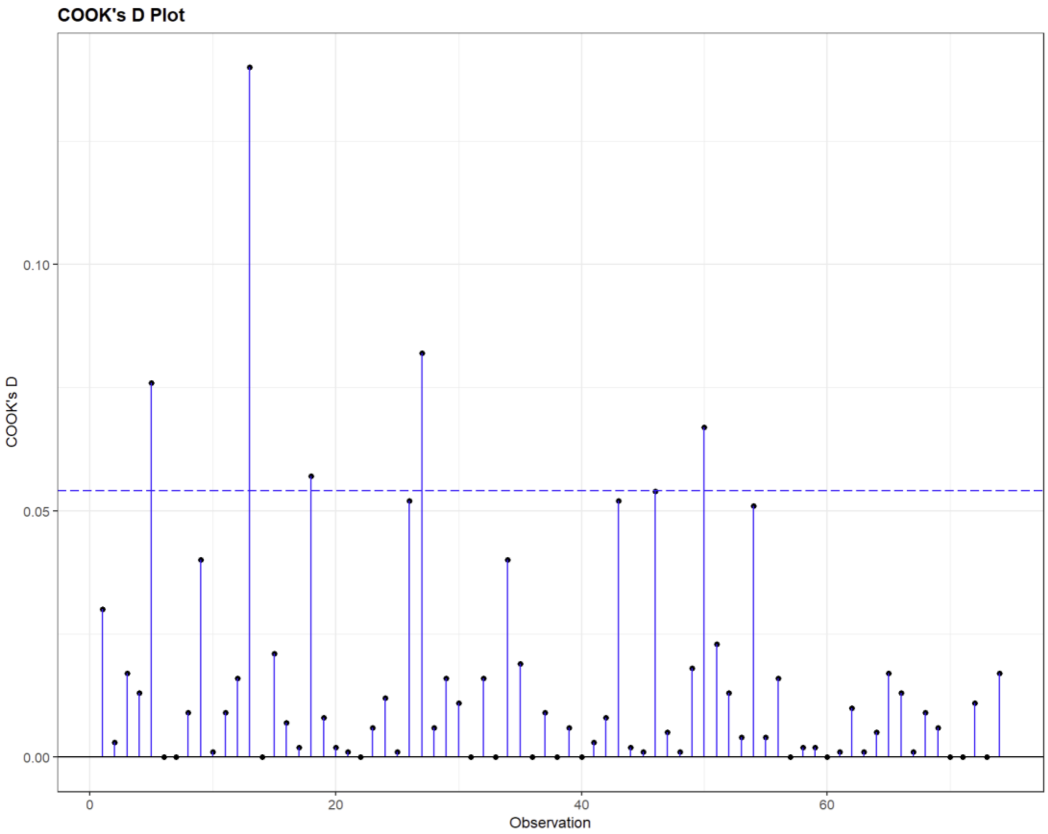
```
> summary(Test_agm$.std.resid)
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max. 
-1.873731 -0.735511 -0.003411  0.000000  0.664037  2.458050 
```

```
> fre(Test_agm$SRE1.96)

 | Test_agm$SRE1.96 | Count | Valid percent | Percent | Responses, % | Cumulative responses, % |
 | ---------------- | ----- | ------------- | ------- | ------------ | ----------------------- |
 |                0 |    72 |          97.3 |    97.3 |         97.3 |                    97.3 |
 |                1 |     2 |           2.7 |     2.7 |          2.7 |                   100.0 |
 |           #Total |    74 |         100.0 |   100.0 |        100.0 |                         |
 |             <NA> |     0 |               |     0.0 |              |                         |
> fre(Test_agm$SRE2.58)

 | Test_agm$SRE2.58 | Count | Valid percent | Percent | Responses, % | Cumulative responses, % |
 | ---------------- | ----- | ------------- | ------- | ------------ | ----------------------- |
 |                0 |    74 |           100 |     100 |          100 |                     100 |
 |           #Total |    74 |           100 |     100 |          100 |                         |
 |             <NA> |     0 |               |       0 |              |                         |
> fre(Test_agm$SRE3.29)

 | Test_agm$SRE3.29 | Count | Valid percent | Percent | Responses, % | Cumulative responses, % |
 | ---------------- | ----- | ------------- | ------- | ------------ | ----------------------- |
 |                0 |    74 |           100 |     100 |          100 |                     100 |
 |           #Total |    74 |           100 |     100 |          100 |                         |
 |             <NA> |     0 |               |       0 |              |                         |
>
```

Seventh, influential cases are delved into. No case has a Cook's D value exceeding 0.5. There is no direct risk with influential cases. One case does stand out visually, yet as the model is incomplete and does not include the relevant control variables, I do not investigate further the reasons causing this potential influential case.

```
> summary(Test_agm$.cooksd)
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max. 
1.220e-07 1.092e-03 6.436e-03 1.552e-02 1.669e-02 1.404e-01 
```

Figure 11: Cook's D plot (Test)



3. *OLS assumptions for the Full Women leadership model*

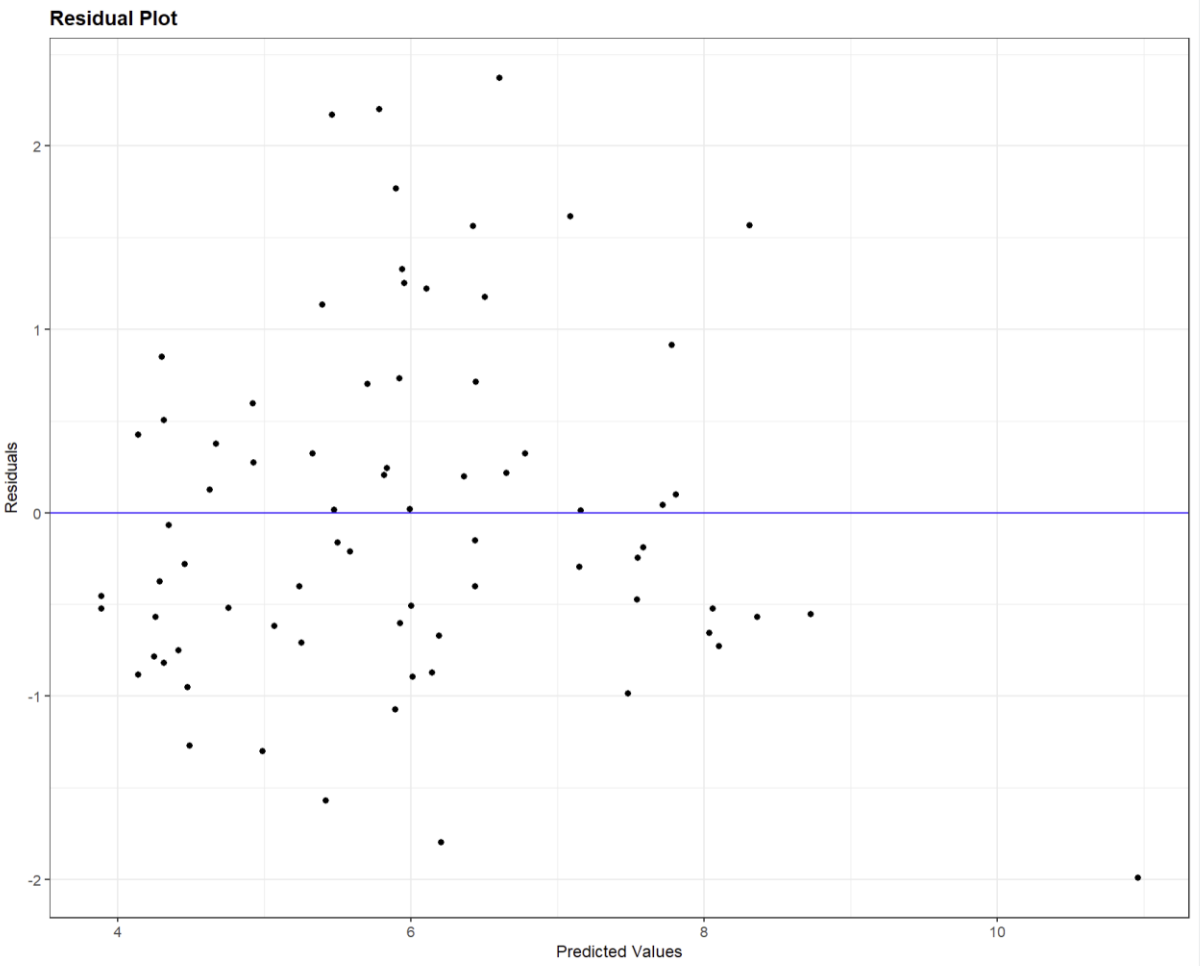
First, excessive multicollinearity is tested. No variable has a VIF value higher than 5. Consequently, the model presents no multicollinearity concerns.

```
> vif(Model)
                    GVIF Df GVIF^(1/(2*Df))
Women_Leaders   3.637857  4        1.175184
Year            1.279779  1        1.131273
Women_nonleaders 3.801774  1        1.949814
Regime          1.349958  1        1.161877
```

Second, the model's Durbin Watson statistic has a value of about 2.19 and is therefore comprised between 1 and 3. Consequently, there is no issue with autocorrelation.

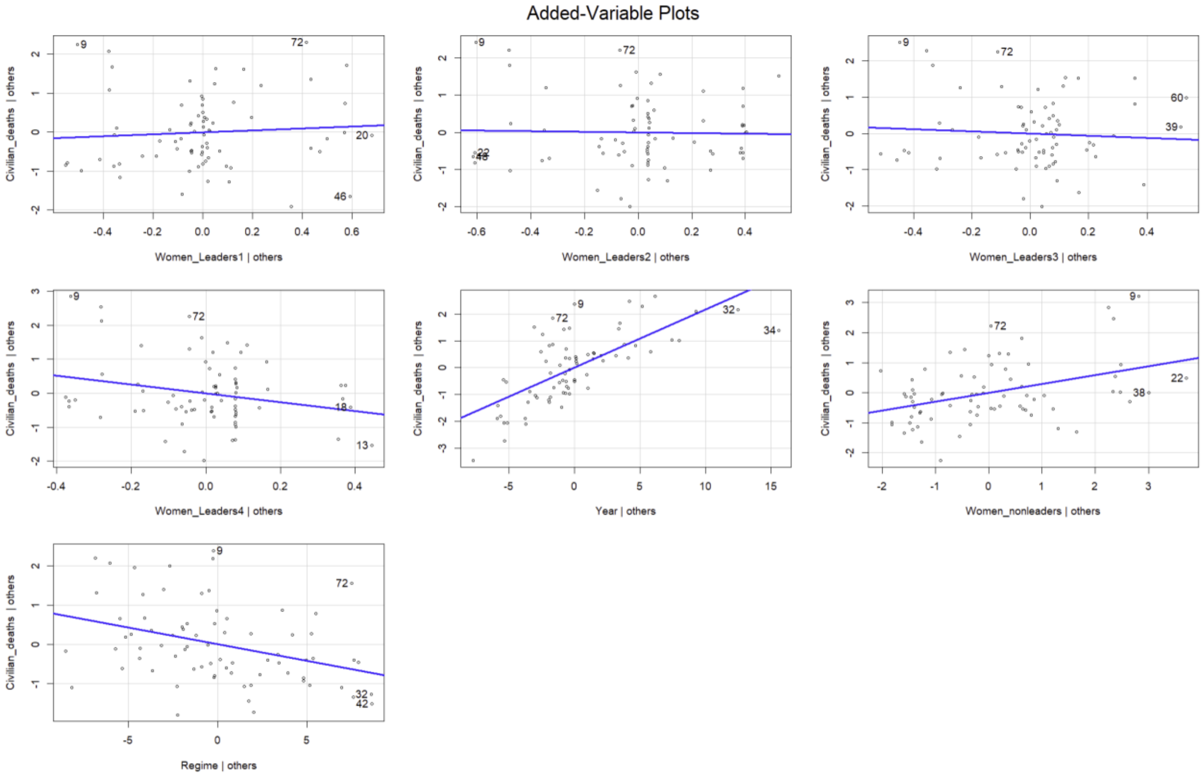
```
> durbinWatsonTest(Model)
 lag Autocorrelation D-W Statistic p-value
   1      -0.1220228      2.190924   0.428
 Alternative hypothesis: rho != 0
```

Figure 12: Residual plot (Full Model)



Third, the dots appear to be approximately randomly distributed on average. Yet, there might be a noticeable accumulation of dots in the lower left corner of the graph The results above do not suggest a strict violation of this assumption, but the dot's shape might stress a mild violation of the linearity and additivity assumptions.

Figure 13: Added-Variable plots



The added variable plots confirm that the association between some explanatory variables and the DV are not linear. We ignore the partial regression plots, or added value plots, for non-continuous variables (the IV). Looking at the variable of years, there is a steep upwards regression line with dots more or less spread randomly, suggesting no serious concern. The line is also going up but less steeply for women participating in non-leadership roles, and the dots are less randomly distributed around the line, suggesting potential issues of linearity. Last, a steep negative line with randomly distributed dots characterizes the relationship between countries' regime type and civilian deaths, after accounting for the relationship between the other predictor variables and the DV). I further look at the visually relationship between each predictor variable and the DV.

Figure 14: IV and DV

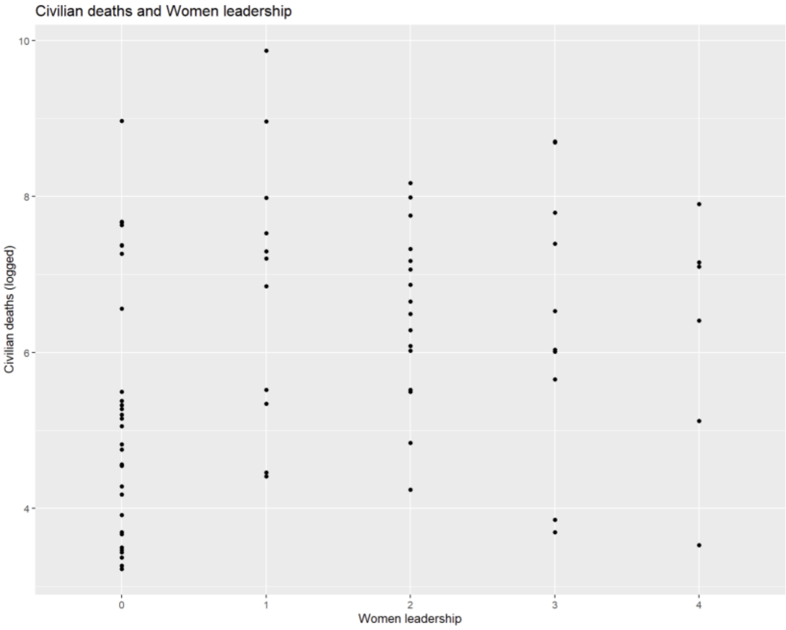


Figure 15: Year and DV

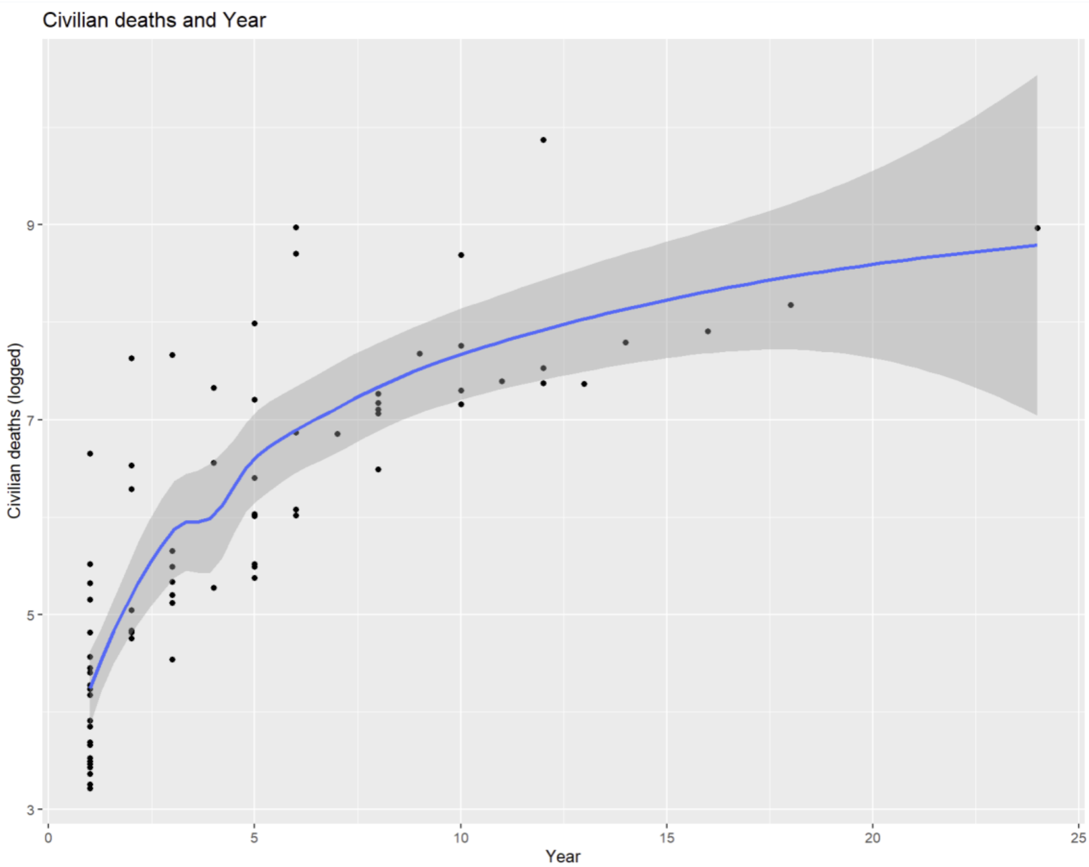


Figure 16: Regime and DV

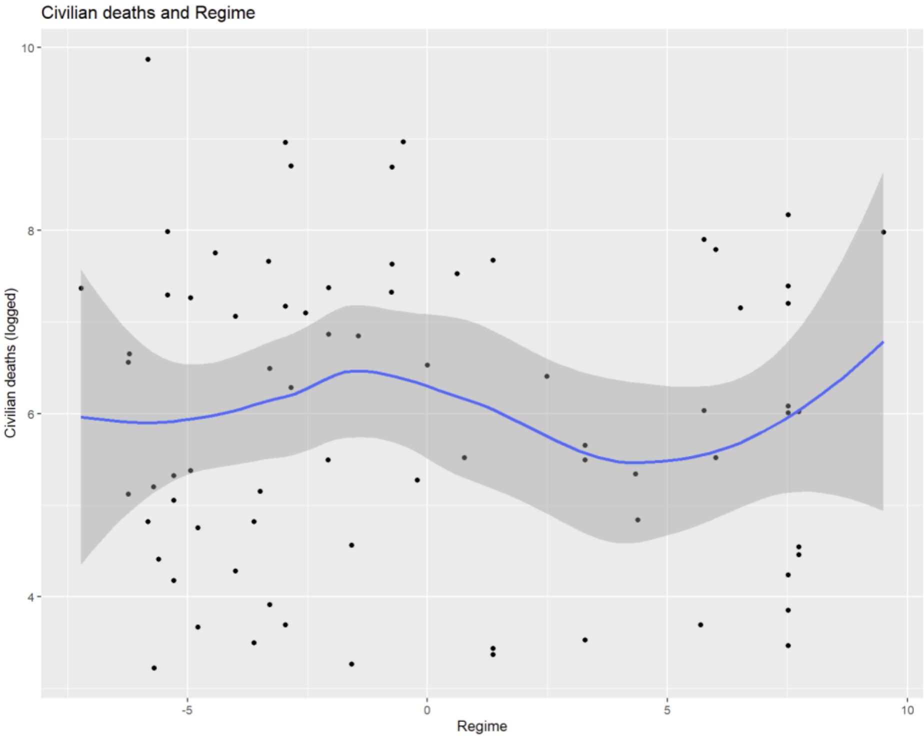
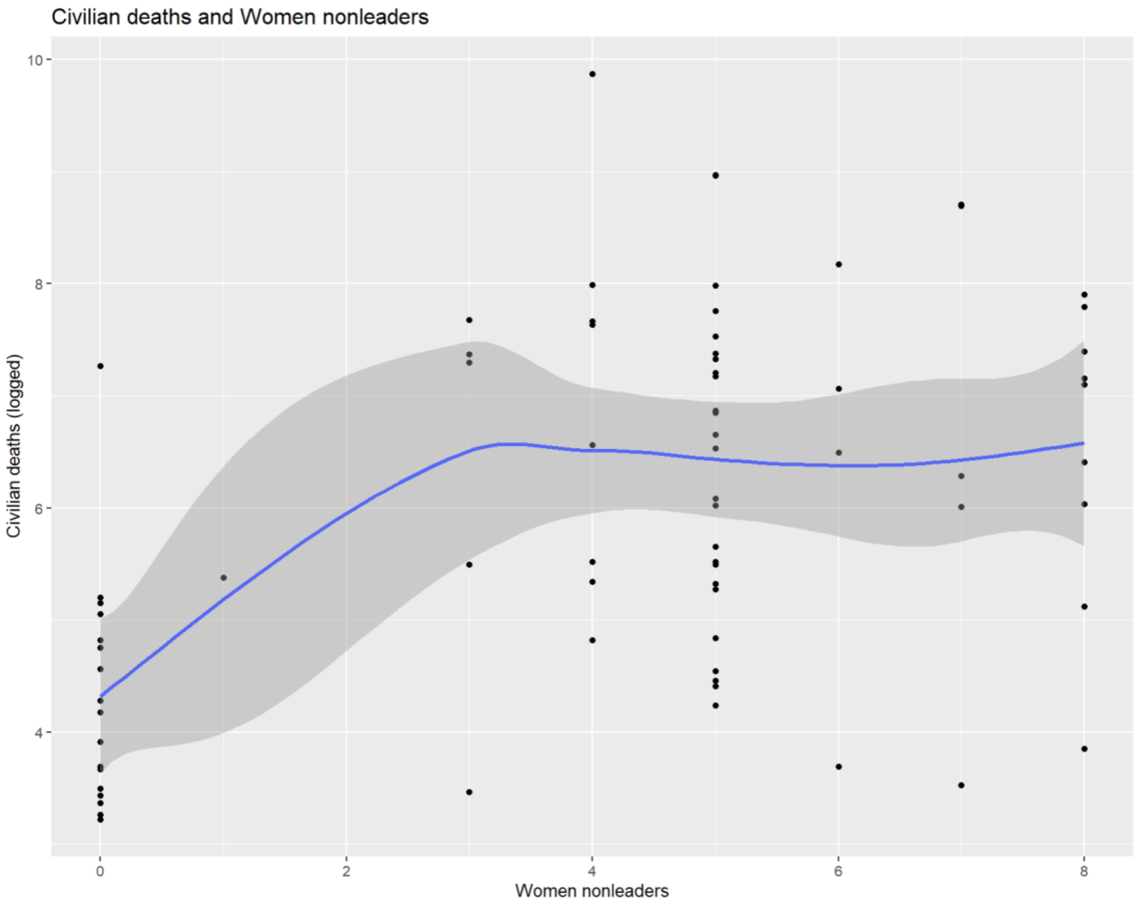
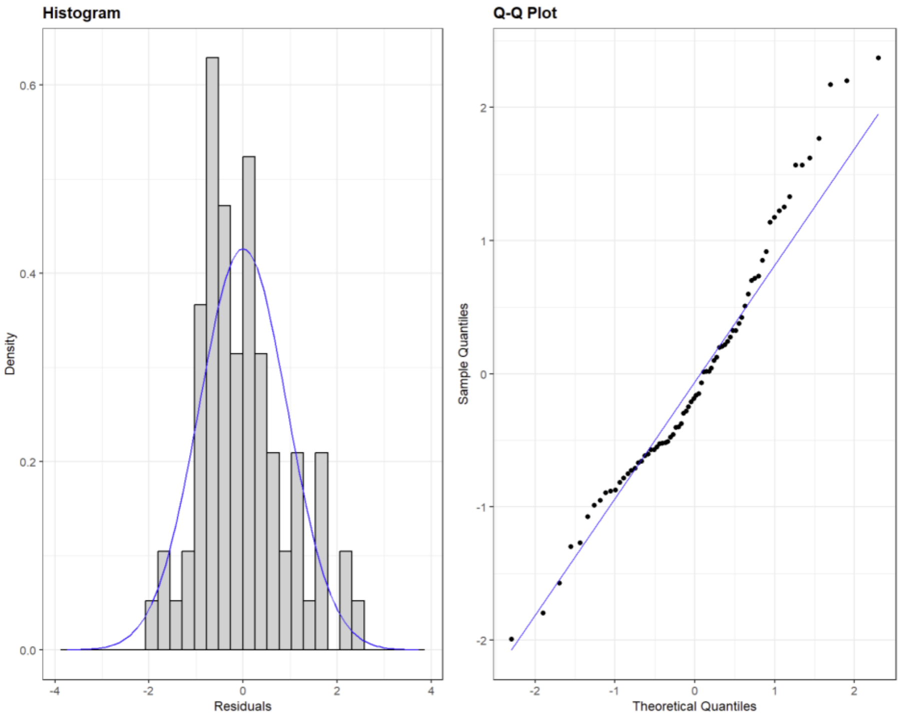


Figure 17: Women non-leaders and DV



Non of the four above plots seem to indicate linear relationships between the variables. The relationship between each predictor variable and the DV is hence non-linear. There is a serious issue with the linearity assumption for this model, calling for nuance when interpreting the results.

Figure 18: Histogram and QQ-Q plot



Fourth, the histogram and Q-Q plot indicate that the full model's residuals approximately follow a normal distribution. The errors therefore appear to be normally distributed, and this assumption is not strictly violated. Yet, some cases do deviate towards the right part of the distribution, suggesting the presence of potential outliers or influential cases.

```
> summary(Model_agm$.std.resid)
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max. 
-2.418349 -0.672396 -0.186915 -0.005773  0.508733  2.538605 
```

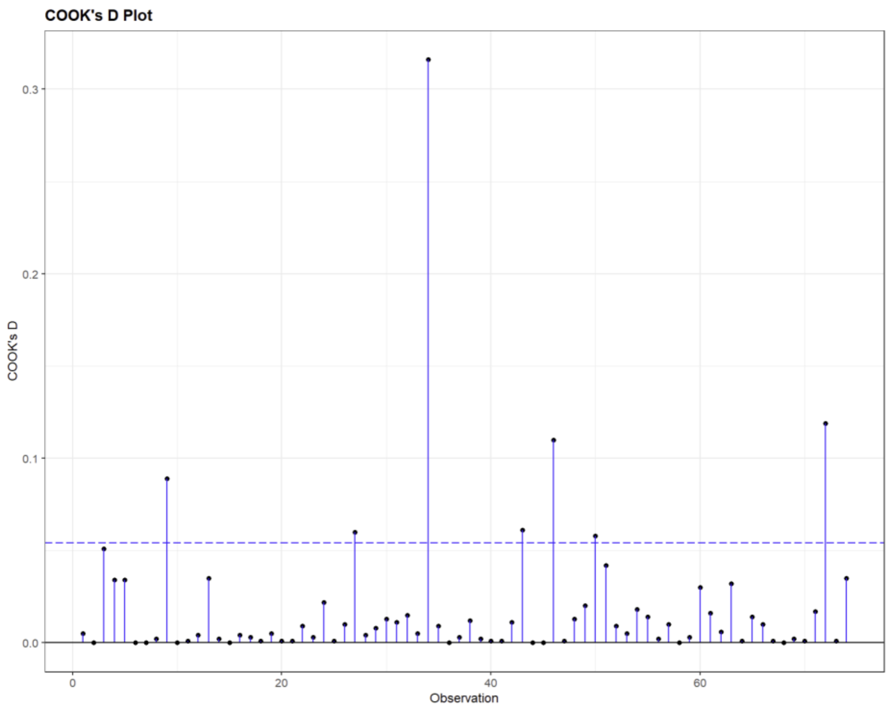
```
> fre(Model_agm$SRE1.96)

 | Model_agm$SRE1.96 | Count | Valid percent | Percent | Responses, % | Cumulative responses, % |
 | ----------------- | ----- | ------------- | ------- | ------------ | ----------------------- |
 |                 0 |    69 |          93.2 |    93.2 |         93.2 |                    93.2 |
 |                 1 |     5 |           6.8 |     6.8 |          6.8 |                   100.0 |
 |            #Total |    74 |         100.0 |   100.0 |        100.0 |                         |
 |              <NA> |     0 |               |     0.0 |              |                         |
> fre(Model_agm$SRE2.58)

 | Model_agm$SRE2.58 | Count | Valid percent | Percent | Responses, % | Cumulative responses, % |
 | ----------------- | ----- | ------------- | ------- | ------------ | ----------------------- |
 |                 0 |    74 |           100 |     100 |          100 |                     100 |
 |            #Total |    74 |           100 |     100 |          100 |                         |
 |              <NA> |     0 |               |       0 |              |                         |
> fre(Model_agm$SRE3.29)

 | Model_agm$SRE3.29 | Count | Valid percent | Percent | Responses, % | Cumulative responses, % |
 | ----------------- | ----- | ------------- | ------- | ------------ | ----------------------- |
 |                 0 |    74 |           100 |     100 |          100 |                     100 |
 |            #Total |    74 |           100 |     100 |          100 |                         |
 |              <NA> |     0 |               |       0 |              |                         |
>
```

Fifth, the quantity of outliers is investigated. There are more than 5% of cases with a standardized residual higher than the absolute value of 1.96, and no case with a standardized residual higher than the absolute value of 2.58. Therefore, the model does present excessive amounts of outliers. This assumption is violated.

Lastly, influential cases are delved into. No case has a Cook's D value exceeding 0.5. There is no direct risk of influential cases concerns. One case does stand out visually.

```
> summary(Model_agm$.cooksd)
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max. 
2.486e-06 1.184e-03 5.264e-03 1.950e-02 1.667e-02 3.156e-01 
```

Figure 19: Cook's D plot Full Model



```
  Civilian_deaths .cooksd
            <dbl>   <dbl>
1            8.97   0.316
```

This case is the LRA. I qualitatively delve into this specific case in this thesis' analysis.

Appendix C – Codes

Data frame: Data

ID	*Name*	*Label*	*Values*	*Value Labels*
1	sideb			<output omitted>
2	sum_fatalities		*range: 25-19430*	
3	Civilian_deaths		*range: 3.2-9.9*	
4	Women_Leaders			0 1 2 3 4
5	Year		*range: 1-24*	
6	Women_nonleaders		*range: 0-8*	
7	Regime		*range: -7.2-9.5*	

```r
library(rio)
library(tidyverse)
library(tibble)
library(sjPlot)
library(psych)
library(car)
library(broom)
library(ggResidpanel)
library(expss)
library(marginaleffects)
library(performance)
library(modelsummary)
library(expss)
#I start by merging my data sets
#import WAAR data set
data <- import("WAAR+Project+Dataset+v1.0.csv")
#Import one sided data set
dt <- import("OneSided_v25_1.rds")
#filter out cases where government perpetrates violence since my unit of
#analysis is the armed group and filter out cases after 2015 (since WAAR goes until 2015)
dt1 <- dt |>
  filter(!(is_government_actor == "1"),
         !(year > 2015))
#rename a common variable with same id and notation within my two data sets
dt4 <- dt1 |>
  rename(sideb_id = actor_id)
#inner join: only sideb_id that are in the two data sets
dtWAARcombined <- inner_join(x = dt4, y = data, by = c("sideb_id"))
#select my variables of interest
dtWAARcombined2 <- dtWAARcombined |>
  select(year, sideb_id, sideb, best_fatality_estimate,lead_prev_best,
         country_primary, actor_name_fulltext, frontline_prev_best,
         noncombat_prev_best)

#view my merged data set with selected variables
view_df(dtWAARcombined2)
table(dtWAARcombined2$year)
#summarize my data set so it becomes cross-sectional in order to avoid autocorrelation
#I make the sum of the values when the values differ, but when the values remain
#constant, I use first(). In many cases, the values are the same for my unit of
#analysis so I can just use first() to obtain the cross-sectional data.
```

```r
dtWAARcombined3 <- dtWAARcombined2 |>
  group_by(sideb) |>
  summarise(sum_fatalities = sum(best_fatality_estimate, na.rm=TRUE),
            sideb_id = first(sideb_id, na_rm = TRUE),
            sideb = first(sideb, na_rm = TRUE),
            lead_prev_best = first(lead_prev_best, na_rm = TRUE),
            country_primary = first(country_primary, na_rm = TRUE),
            year = n_distinct(year, na.rm = TRUE),
            frontline_prev_best = first(frontline_prev_best, na_rm = TRUE),
            noncombat_prev_best = first(noncombat_prev_best, na_rm = TRUE),
)
view_df(dtWAARcombined3)

#Adding control variables
#import QOG data to control for government capacity
QOG <- import("qog_bas_ts_jan25.csv")
table(QOG$year)
#filter out cases post 2015 since WAAR stops after 2015
QOG1 <- QOG |>
  filter(!(year > "2015"))
QOG12 <- QOG1 |>
  mutate(p_polity2 = na_if(p_polity2, -66)) |>
  mutate(p_polity2 = na_if(p_polity2, -77)) |>
  mutate(p_polity2 = na_if(p_polity2,-88))

QOG123 <- QOG12 |>
  filter(!is.na(p_polity2))

QOG2 <- QOG123 |>
  group_by(cname) |>
  summarise(ccodecow = mean(ccodecow, na.rm = TRUE),
            p_polity2 = mean(p_polity2, na.rm = TRUE))
view_df(QOG2)
#ccodecow matches gwnoa from UCDP GED dataset, that's the easiest way to merge
#QOG with dtWAARcombined3
#import GED UCDC data set
GED <- import("ged251-csv.zip")
GED2 <- GED |>
  select(gwnoa, dyad_name, side_b_new_id, country)
#rename side_b_new_id to merge UCDP one-sided with GED
GED3 <- GED2 |>
  rename(sideb_id = side_b_new_id)
```

```
GED4 <- GED3 |>
  group_by(sideb_id) |>
  summarise(gwnoa = first(gwnoa))

#merge GED with dtWAARcombined
dtWAARcombined3GED4 <- left_join(x = dtWAARcombined3, y = GED4, by = c("sideb_id"))
#the problem is that South Africa is not coded for the same dyad (different sideb)
#but the gwnoa number of South Africa is 560
#I attribute the missing value to the case of ANC found row 6
dtWAARcombined3GED4$gwnoa[6] <- 560
dtWAARcombined3GED4$gwnoa[46] <- 770
dtWAARcombined3GED4$gwnoa[47] <- 490
dtWAARcombined3GED4$gwnoa[56] <- 750
dtWAARcombined3GED4$gwnoa[65] <- 652
dtWAARcombined3GED4$gwnoa[89] <- 500

#rename ccodecow from QOG in order to have a common variable to merge the two datasets
QOG3 <- QOG2 |>
  rename(gwnoa = ccodecow)
#i can't merge my two data sets because one is numeric and the other is character
#i get an error message from R, so I need to have the same for both
#i decide to convert as numeric because BAAD also has the variable as numeric
dtWAARcombined3GED4$gwnoa <- as.numeric(dtWAARcombined3GED4$gwnoa)
A <- inner_join(x = dtWAARcombined3GED4, y = QOG3, by = c("gwnoa"))

#rename variables coding for the same unit of analysis but using different spelling
#I have to add dplyr:: before recode otherwise my packages clash
QOG4 <- QOG3 |>
  mutate(country_primary = dplyr::recode(cname,
                                         "Congo (the Democratic Republic of the)" = "DR Congo (Zaire)"
                                         "Syrian Arab Republic (the)" = "Syria"))
#merge by country name
B <- left_join(x = A, y = QOG4, by = c("country_primary"))
#combining the two variables of government capacity to get rid of missing value
C <- B |> mutate(p_polity2 = coalesce(B$p_polity2.x, B$p_polity2.y))

#Investigate the two indicators of my control variable Women_nonleaders
table(C$frontline_prev_best)
describe(C$frontline_prev_best)
class(C$frontline_prev_best)
summary(C$frontline_prev_best)
```

```
table(C$noncombat_prev_best)
describe(C$noncombat_prev_best)
class(C$noncombat_prev_best)
summary(C$noncombat_prev_best)

#Creating Women_nonleaders
C1 <- C |>
  mutate(Women_nonleaders = frontline_prev_best + noncombat_prev_best)

#inspect my the other variables with table, class, and summary
table(C1$lead_prev_best)
describe(C1$lead_prev_best)
class(C1$lead_prev_best)
summary(C1$lead_prev_best)

table(C1$year)
describe(C1$year)
class(C1$year)
summary(C1$year)

table(C1$Women_nonleaders)
describe(C1$Women_nonleaders)
class(C1$Women_nonleaders)
summary(C1$Women_nonleaders)

table(C1$p_polity2)
describe(C1$p_polity2)
class(C1$p_polity2)
summary(C1$p_polity2)

table(C1$sum_fatalities)
describe(C1$sum_fatalities)
class(C1$sum_fatalities)
summary(C1$sum_fatalities)
#The DV is highly skewed. The median (246.5) is much below the mean (1338), hence there are
#lots of smaller value but also very high values. It looks like the variable would
#be more linear with a logged transformation.
#To tackle this issue, I transform my DV to a logged version.
C2 <- C1 |>
  mutate(Civilian_deaths = log(sum_fatalities))
summary(C2$Civilian_deaths)
#The logged version appears to be much better and tackles the skewness issue
```

```r
#I visualise the two versions with a bar plot
ggplot(data = C2,
       mapping = aes(x = sum_fatalities)) +
  geom_dotplot(binwidth = 210)
ggplot(data = C2,
       mapping = aes(x = Civilian_deaths)) +
  geom_dotplot(binwidth = 0.30)
#The plots confirm that the logged version is more adequate

#Convert the IV, women leaders, to a factor variable before running the regression
C3 <- C2|>
  mutate(lead_prev_best1 = factor(lead_prev_best))
class(C3$lead_prev_best1)

#Renaming my variables' and data set names to easier names
C4 <- C3 |>
  rename(Women_Leaders = lead_prev_best1,
         Year = year,
         Regime = p_polity2)

#selecting variables of interest
Data <- C4 |>
  select(sideb, sum_fatalities, Civilian_deaths, Women_Leaders, Year,
      Women_nonleaders, Regime) |>
  filter(!is.na(Women_Leaders)) |>
  filter(!is.na(Women_nonleaders))
#Looking at the IV's categories
attributes(Data$Women_Leaders)

view_df(Data)

# running the linear regression models
Test <- lm(Civilian_deaths ~ Women_Leaders, data=Data)
summary(Test)

Data_complete <- Data |>
  filter(complete.cases(Civilian_deaths, Women_Leaders, Year,
                        Women_nonleaders, Regime))

model1 <- lm(Civilian_deaths ~ 1, data = Data_complete)
```

```r
model2 <- lm(Civilian_deaths ~ Women_Leaders, data = Data_complete)
summary(model2)

model3 <- lm(Civilian_deaths ~ Women_Leaders + Year, data=Data_complete)
summary(model3)

model4 <- lm(Civilian_deaths ~ Women_Leaders + Year + Women_nonleaders, data=Data_complete)
summary(model4)

model5 <- lm(Civilian_deaths ~ Women_Leaders + Year + Women_nonleaders + Regime,
             data=Data_complete)
summary(model5)

anova(model1, model2, model3, model4, model5)
#The best model fit is model 4. It makes sense theoretically to keep it.
Model <- lm(Civilian_deaths ~ Women_Leaders + Year + Women_nonleaders + Regime,
              data=Data)
summary(Model)

#I start the OLS assumptions
#I test the six ordinary linear regression assumptions with
#I start by testing the multicollinearity assumption of Model
vif(Model)
#No statistics exceed the value of 5. There is hence no issue of
#excessive multicollinearity.

#I continue by testing the independent errors assumption using the
#durbinWatson test
durbinWatsonTest(Test)
durbinWatsonTest(Model)
#These models have no issue of serial autocorrelation since the Durbin Watson
#statistic, revolving around 1.7, is comprised between 1 and 3

#I continue by checking the linearity and additivity assumptions.
resid_panel(Test, plots = c("resid"))
resid_panel(Model, plots = c("resid"))
avPlots(Test)
avPlots(Model)
```

```r
#This plots do not display clear violations of these assumptions
ggplot(Data, aes(x = Women_Leaders, y = Civilian_deaths)) +
  geom_point() +
  geom_smooth(method = "loess") +
  labs(title = "Civilian deaths and Women leadership",
       x = "Women leadership",
       y = "Civilian deaths (logged)")

ggplot(Data, aes(x = Year, y = Civilian_deaths)) +
  geom_point() +
  geom_smooth(method = "loess") +
  labs(title = "Civilian deaths and Year",
       x = "Year",
       y = "Civilian deaths (logged)")

ggplot(Data, aes(x = Regime, y = Civilian_deaths)) +
  geom_point() +
  geom_smooth(method = "loess") +
  labs(title = "Civilian deaths and Regime",
       x = "Regime",
       y = "Civilian deaths (logged)")

ggplot(Data, aes(x = Women_nonleaders, y = Civilian_deaths)) +
  geom_point() +
  geom_smooth(method = "loess") +
  labs(title = "Civilian deaths and Women nonleaders",
       x = "Women nonleaders",
       y = "Civilian deaths (logged)")

#Then, I test the assumption of the normally distributed errors.
resid_panel(Test, plots = c("hist", "qq"))
resid_panel(Model, plots = c("hist", "qq"))

#I now inspect outliers and influential cases for the model called Test
#Run the command and store results
Test_agm <- augment(Test)
summary(Test_agm$.std.resid)

#Some values appear to be concerning since they exceed the 1.96 threshold
#I delve deeper into this issue by running the following code
```

```
Test_agm <- Test_agm |>
  mutate(SRE1.96 = case_when(
    .std.resid > 1.96 | .std.resid < -1.96  ~ 1,
    .std.resid > -1.96 & .std.resid < 1.96 ~ 0),
    SRE2.58 = case_when(
      .std.resid > 2.58 | .std.resid < -2.58  ~ 1,
      .std.resid > -2.58 & .std.resid < 2.58 ~ 0),
    SRE3.29 = case_when(
      .std.resid > 3.29 | .std.resid < -3.29  ~ 1,
      .std.resid > -3.29 & .std.resid < 3.29 ~ 0
    ))
fre(Test_agm$SRE1.96)
fre(Test_agm$SRE2.58)
fre(Test_agm$SRE3.29)

#I continue by investigating influential cases.
summary(Test_agm$.cooksd)
resid_panel(Test, plots = c("cookd"))

#I inspect the cases individually.
Test_agm |>
  filter(.cooksd > 0.1) |>
  select(Civilian_deaths, .cooksd)

#I now inspect outliers and influential cases for the model called Model
#Run the command and store results
Model_agm <- augment(Model)
summary(Model_agm$.std.resid)

#Some values appear to be concerning since they exceed the 2.58 threshold
#I delve deeper into this issue by running the following code
Model_agm <- Model_agm |>
  mutate(SRE1.96 = case_when(
    .std.resid > 1.96 | .std.resid < -1.96  ~ 1,
    .std.resid > -1.96 & .std.resid < 1.96 ~ 0),
    SRE2.58 = case_when(
      .std.resid > 2.58 | .std.resid < -2.58  ~ 1,
      .std.resid > -2.58 & .std.resid < 2.58 ~ 0),
    SRE3.29 = case_when(
      .std.resid > 3.29 | .std.resid < -3.29  ~ 1,
      .std.resid > -3.29 & .std.resid < 3.29 ~ 0
    ))
```

```
fre(Model_agm$SRE1.96)
fre(Model_agm$SRE2.58)
fre(Model_agm$SRE3.29)

Model_agm |>
  filter(.std.resid > 1.96 | .std.resid < -1.96) |>
  select(Civilian_deaths, .std.resid)

#I continue by investigating influential cases.
summary(Model_agm$.cooksd)
resid_panel(Model, plots = c("cookd"))

#I inspect the cases individually.
Model_agm |>
  filter(.cooksd > 0.3) |>
  select(Civilian_deaths, .cooksd)

modelsummary(Test,
             stars = TRUE,
             coef_rename = c(
               "(Intercept)" = "Intercept",
               "Women_Leadership" = "Women in leadership"),
             gof_map = c("nobs", "r.squared", "adj.r.squared"),
             title = "Predicting (logged) civilian deaths with women leadership",
             notes = "OLS coefficients with standard errors in parentheses")

modelsummary(Model,
             stars = TRUE,
             coef_rename = c(
               "(Intercept)" = "Intercept",
               "Women_Leaders" = "Women in leadership",
               "Year" = "Year",
               "Women_nonleaders" = "Women prevalence in other positions",
               "Regime" = "Regime type"),
             gof_map = c("nobs", "r.squared", "adj.r.squared"),
             title = "Predicting (logged) civilian deaths with women leadership, year, women prevalence and regime",
             notes = "OLS coefficients with standard errors in parentheses. Reference group for Women_Leaders = Null")

#A list with names
model_list_named <- list(
  "Just Women leadership" = Test,
  "Full Model" = Model)
```

```
library(pandoc)
#Create the table
modelsummary(model_list_named,
             stars = TRUE,
             coef_rename = c(
               "(Intercept)" = "Intercept",
               "Women_Leaders1" = "Less than 5% of Women In leadership",
               "Women_Leaders2" = "Between 5-9% of Women in leadership",
               "Women_Leaders3" = "Between 10-19% of Women in leadership",
               "Women_Leaders4" = "More than of 20% Women in leadership",
               "Year" = "Year",
               "Women_nonleaders" = "Women prevalence in other positions",
               "Regime" = "Regime type"),
             gof_map = c("nobs", "r.squared", "adj.r.squared"),
             title = "Predicting (logged) civilian deaths with women leadership, year, women prevalence and regime",
             notes = "OLS coefficients with standard errors in parentheses.  Reference group for Women_Leaders = Null",
             output = "example_regression_table.docx")

#Calculating the table's coefficients
exp(1.669) - 1
exp(1.310) - 1
exp(1.247) - 1
exp(1.014) - 1
exp(0.250) - 1
exp(-0.085) - 1
exp(-0.301) - 1
exp(-1.310) - 1
exp(0.216) - 1
exp(0.294) - 1
exp(-0.085) - 1
#Store the model results as a tidied data object
#NEED to ask for the confidence interval
Model_tidied <- tidy(Model, conf.int = TRUE)
Model_tidied

library(dplyr)
#recoding term to get nicer names to plot:
Model_tidied <- Model_tidied |>
  mutate(term = dplyr::recode(term,
                              "Women_Leaders" = "Women Leaders",
                              "Year" = "Year",
                              "Women_nonleaders" = "Women prevalence in other positions",
                              "Regime" = "Regime Type"))
```

```r
Model_tidied


#recoding term to get nicer names to plot:
Model_tidied2 <- Model_tidied |>
  mutate(term = dplyr::recode(term,
                       "Women_Leaders1" = "<5% Women In leadership",
                       "Women_Leaders2" = "5-9% Women in leadership",
                       "Women_Leaders3" = "10-19% Women in leadership",
                       "Women_Leaders4" = ">20% Women in leadership",
                       "Year" = "Year",
                       "Women_nonleaders" = "Women prevalence in other positions",
                       "Regime" = "Regime Type"))
#making a coefficient plot
Model_tidied2 |>
  filter(term != "(Intercept)") |>
  ggplot(aes(x = estimate, y = term)) +
  geom_pointrange(aes(xmin = conf.low,
                      xmax = conf.high)) +
  labs(title = "OLS Coefficients Predicting Level of (logged) Civilian Deaths",
       x = "OLS Estimate",
       y = "Explanatory Variables") +
  geom_vline(xintercept = 0, linetype = "dashed", color = "red") +
  geom_text(aes(label = round(estimate, 2)), vjust = -0.5)

Model_tidied <- tidy(Model, conf.int = TRUE, conf.level = 0.95)

#predicted values plots
#obtain predicted values
demo_preds <- predictions(Model,
                          by = "Women_Leaders",
                          newdata = "mean")
demo_preds

#make a predicted values plot
ggplot(demo_preds, aes(x = Women_Leaders, y = estimate)) +
  geom_pointrange(aes(ymin = conf.low, ymax = conf.high)) +
  labs(title = "Predicted level of (logged) civilian deaths by the prevalence of women leadership in insurgent groups",
       y = "Predicted Value",
       x = "Women prevalence") +
  scale_y_continuous(limits = c(3.5,7)) +
  geom_text(aes(label = round(estimate, 2)), hjust = -0.25)

```

```r
#standardized coefficients
library(parameters)
multiple_std <- standardize_parameters(Model,
                                       method = "refit")
glimpse(multiple_std)

tidy(Model)

#Standardized
multiple_std

#I look at the number of deaths per year to compare the deaths toll prepetrated
#by insurgent groups
lila <- Data$sum_fatalities/Data$Year
lila
describe(lila)
summary(lila)

#I make a table of the descriptive statistics
Data1 <- Data |>
  group_by(Women_Leaders) |>
  summarise(
    mean = mean (sum_fatalities, na.rm = TRUE),
    sd = sd(sum_fatalities, na.rm = TRUE),
    N = n(),
    median = median(sum_fatalities, na.rm = TRUE),
    min = min(sum_fatalities, na.rm = TRUE),
    max = max(sum_fatalities, na.rm = TRUE),
    Q1 = quantile(sum_fatalities, 0.25, na.rm = TRUE),
    Q3 = quantile(sum_fatalities, 0.75, na.rm = TRUE),
  )
view_df(Data1)
```