



The Missing Link: How Languages Can Influence Conflict

Citation

Kregel, Megan Ashley. 2022. The Missing Link: How Languages Can Influence Conflict. Master's thesis, Harvard University Division of Continuing Education.

Link

<https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37373338>

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), as set forth at

<https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFIOWI2M2EwYTg>

Accessibility

<https://accessibility.huit.harvard.edu/digital-accessibility-policy>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#)

The Missing Link: How Languages Can Influence Conflict

Megan Ashley Kregel

A Thesis in the Field of International Relations
for the Degree of Master of Liberal Arts in Extension Studies

Harvard University

November 2022

Abstract

The purpose of this thesis research was to explore and assess the extent to which language diversity and conflict are related. For this global exploration, language diversity was measured by the number of languages used within countries. Conflict was measured by six different measures. Relationships between language diversity and the six conflict measures were explored for each of four years, 2004, 2010, 2015, and 2020 to account for changes over time. In addition to the visual geographical analysis of each of these measures, non-spatial regressions were conducted to determine whether statistically significant correlations were present.

Given the limited amount of prior research that incorporates conflict data into geographical analysis, this is exploratory research. This research drew upon multiple conflict data sets to create the six conflict measures, including monadic armed conflict, dyadic armed conflict, one-sided violence, non-state conflict, battle-related deaths, and deadly electoral conflict to illuminate what aspects of conflict may be related to the number of languages spoken.

The regression results identified three conflict measures with statistically significant correlations with language diversity, but the level of explanation was modest at best, with just 11% and 12% of the variance explained. The three measures that showed promise in explaining conflict were monadic conflict, one-sided violence, and non-state conflict. Further research is needed with additional years to build upon these exploratory results.

Frontispiece

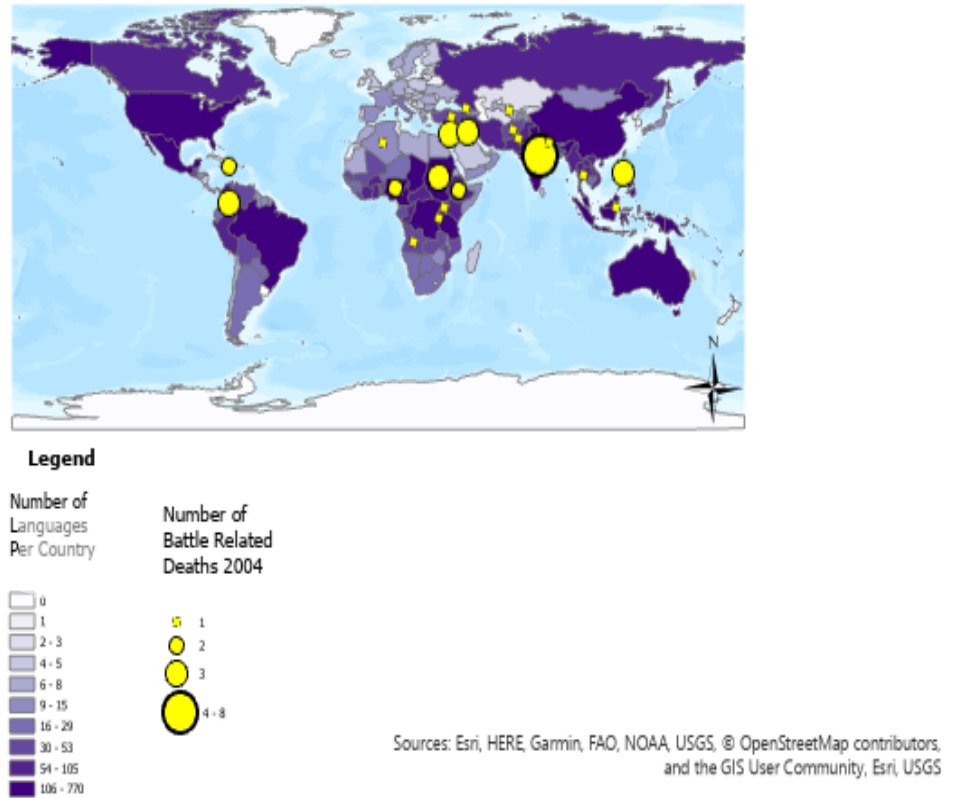


Figure 0-1 Battle Related Deaths 2004

Author's Biographical Sketch

Megan Ashley Kregel completed her undergraduate study at Indiana University-Bloomington, where she graduated in 2011 with degrees in History, Eastern European Languages and Literatures, and Eastern European Studies. During the last eleven years she has worked in the localization industry specializing in rare language recruitment, medical/pharmaceutical translations, and training initiatives. Her background and several Harvard Extension School courses led her to investigate if there is a link between language and conflict.

Dedication

I would like to dedicate this firstly to my parents. Your self-sacrifices my whole life have taught me the meaning of challenging work. Your drive, courage, and love are what have made me who I am today. I would not be where I am without you. I hope that I have made you and the ancestors proud.

My sibling: for being one of the very first people to challenge me, you have helped me to always grow, always question, always keep arguing my point and driving it home. Also, I acknowledge you received your MBA before I completed my masters even though I am older.

The librarians of the Hebron Public Library: for teaching me to never give up when doing research. That you need to keep digging until you find what you are looking for.

My teachers: for pushing me to continue, for helping me to learn and grow, for the inspiration to keep going.

My husband: your support in this has meant so much to me. You know how important this has been to me on this journey I started 5 years ago when I went back to school. You have been a champion with the “Chopped” inspired dinners the last few months and for listening to me speaking nonstop (even in my sleep) about maps and in foreign languages.

Acknowledgments

There are a number of people I would like to thank for helping me get this thesis to where it is today. Without your guidance, patience, and support, I would have never been able to complete this.

Doug Bond & Jill Kelly – thank you both for your willingness to collaborate with me on this endeavor. I honestly do not think I would have completed this without both of your encouragement, expertise, wisdom, and allowing me to discuss everything with you both.

Toni Haraldsen – for proofreading and your sage advice

Steven Sajkich – for all the inspirational quotes and encouragement sent my way,
I owe you a shipment of cookies for proofreading

Mandy & Daphne Fleming – your words of encouragement and the coffee care package have helped me through the last few weeks immensely. Thank you for being amazing friends. I cannot wait until we can all celebrate our new degrees together.

Rachel Kaunas – thank you for forcing me to leave the house and take mental health breaks.

Harvard's Geography department – Scott & Scott for their consultancy and for helping me with finding the geographic datasets that I needed.

Carolyn & Victoria Schulz – for sending me much needed funny GIFs as distractions and my self-care kits while I worked on this.

Jean & Katie Laut – for opening your home to me all those years ago. Jean, I would have never made it in the localization industry without you taking me under your wing. This thesis itself would have never happened without you inspiring me. Thank you for being my life-long mentor.

Table of Contents

Frontispiece.....	iv
Author's Biographical Sketch.....	v
Dedication	vi
Acknowledgments.....	vii
List of Tables	x
Table of Figures	xi
Chapter I Introduction.....	1
Chapter II Background of the Problem.....	6
Linguistics Background	6
Conflict Research.....	8
Political Issues	11
Geography	12
Chapter III Research Methodology.....	16
Chapter V. Results	23
UCDP/PRIO Armed Conflict	23
UCDP Dyadic Dataset	29
UCDP One-Sided Violence Dataset	34
UCDP Battle Related Deaths	44
Deadly Electoral Conflict	48
Regressions	522
Chapter VI. Conclusion	62
References.....	64

List of Tables

Uppsala Conflict Datasets Table 1.....	18
--	----

Table of Figures

Figure 0-2 PRIO Armed Conflict Dataset 2004	24
Figure 0-3 PRIO Armed Conflict 2010	26
Figure 0-4 PRIO Armed Conflict 2015	27
Figure 0-5 PRIO Armed Conflict 2020	28
Figure 0-6 Dyadic Dataset 2004	30
Figure 0-7 Dyadic Dataset 2010	31
Figure 0-8 Dyadic Dataset 2015	32
Figure 0-9 Dyadic Dataset 2020	33
Figure 0-10 One-Sided Violence 2004	35
Figure 0-11 One-Sided Violence 2010	36
Figure 0-12 One-Sided Violence 2015	37
Figure 0-13 One-Sided Violence 2020	38
Figure 0-14 Non-State Conflict 2004	40
Figure 0-15 Non-State Conflict 2010	41
Figure 0-16 Non-State Conflict 2015	42
Figure 0-17 Non-State Conflict 2020	43
Figure 0-18 Battle-Related Deaths 2004.....	44
Figure 0-19 Battle Related Deaths 2010.....	45
Figure 0-20 Battle Related Deaths 2015.....	46
Figure 0-21 Battle Related Deaths 2020.....	47
Figure 0-22 Deadly Electoral Conflict 2004.....	49

Figure 0-23 Deadly Electoral Conflict 2010.....	50
Figure 0-24 Deadly Electoral Conflict 2015.....	51
Figure 0-25: One-Sided Violence Regressions.....	54
Figure 0-26: One-Sided Violence Regressions Output.....	55
Figure 0-27: PRIO Armed Conflict Regressions	57
Figure 0-28: PRIO Armed Conflict Regressions Output.....	58
Figure 0-29: Non-State Conflict Regressions	59
Figure 0-30: Non-State Conflict Regressions Output.....	60

Chapter I

Introduction

The languages we speak help us to understand the world around us. They not only shape the nature of our society but impact us psychologically as well. Our “communication is based on features that describe an event and capture emotions, needs, interests and fears. Language is used to resolve or escalate dispute.”¹ But could language in fact be one of the reasons that a conflict occurs within a particular region at all?

While linguists have acknowledged that language could be one of the main causes of conflict, in addition to social, economic, or political factors, scholars in other fields have not fully considered or dismissed the linkage between them. I hypothesize that if there are more than ten languages spoken in any given country, the probability that a conflict will occur is greater than if residents speak fewer languages, given that communication will likely be far more difficult in those countries where the people speak more languages. For countries that have fewer languages, individuals can typically communicate more effectively, thus establishing a peaceful status quo. Language, in effect, could contribute to peace and democracy.

Due to the limited number of resources and literature on this assertion, the following chapters will define what information there is on this topic, and additional

¹ Denis Kisembe, “Language, culture and conflict resolution.” Research Paper, 2018. <https://www.grin.com/document/433194>.

chapters will detail the methodology used in testing. The final chapters will contain research findings and statistical outputs from exploratory regressions.

Glossary of Key Terms

Active Conflict: “A conflict, both state-based and non-state, is deemed to be active if there are at least twenty-five battle-related deaths per calendar year in one of the conflict’s dyads. This rule also applies to settle dyad activity and the activity of the primary warring parties. A secondary warring party is however considered to be active if it actively supports on of the primary parties with regular troops within the state incompatibility. In other words, a secondary warring party does not have to, on its own, incur or suffer twenty-five battle-related deaths to be classified as active. A variant of this coding rule is applied regarding one-sided violence. A one’ sided actor is deemed to be active if an organized group incurs at least twenty-five deliberate killings of civilians in a year.”²

Actor: “A state or a non-state formally organized or informally organized group.”³

Armed Conflict: “A state-based armed conflict is a contested incompatibility that concerns government and/or territory where the use of armed force between two parties, of which at least one is the government of a state, results in at least 25 battle-related deaths in one calendar year.”⁴

² “UCDP Definitions,” Definitions, Uppsala University, accessed April 30, 2021, https://www.pcr.uu.se/research/ucdp/definitions/#tocjump_8829967763543327_2.

³ Ibid.

⁴ Ibid.

Battle-Related Deaths: “Counted as battle-related deaths is the use of armed force between warring parties in a conflict dyad, be it state-based or non-state, resulting in deaths.”⁵

Best Estimate of Deaths (relates to state-based, non-state, one-sided, actors, dyads): “The best estimate consists of the aggregated most reliable number of deaths.”⁶

Civilians: “Civilians [...] are unarmed people who are not active members of the security forces of the state, or members of an organized armed militia or opposition group.

Government officials, such as members of parliament, governors, and councilors, are also excluded and are instead seen as representatives of the government of a state.”⁷

Conflict, Interstate (relates to state-based): “A conflict between two or more governments.”⁸

Conflict, Intrastate (relates to state-based): “A conflict between a government and a non-governmental party, with no interference from other countries.”⁹

Dyad: “A dyad is made up of two armed and opposing actors. In state-based conflicts a dyad is defined as two actors, with one or more being the government, and have a stated incompatibility. In a non-state conflict, a dyad is constructed by at least two organized actors, of which none is the government of a state, which oppose each other with arms. In

⁵ Ibid.

⁶ Ibid.

⁷ Ibid.

⁸ Ibid.

⁹ Ibid.

non-state conflicts it is possible for an alliance of non-state actors to enter a dyad with either an opposing group, or an alliance of opposing groups.”¹⁰

Fatalities: “Deaths incurred in the three categories of organized violence captured by the UCDP. For state-based armed conflict and non-state conflict these are defined as battle-related deaths (i.e., the use of armed force between warring parties in a conflict dyad, be it state-based or non-state, resulting in deaths). For one-sided violence these are deaths stemming from attacks conducted by organized actors, targeting unarmed civilians.”¹¹

Non-State Conflict: “The use of armed force between two organized armed groups, neither of which is the government of a state, which results in at least 25 battle-related deaths in a year.”¹²

One-Sided Violence: “The deliberate use of armed force by the government of a state or by a formally organized group against civilians which results in at least 25 deaths in a year.”¹³

State: “A state is either an internationally recognized sovereign government controlling a specified territory, or an internationally unrecognized government controlling a specified territory whose sovereignty is not disputed by another internationally recognized sovereign government previously controlling the same territory.”¹⁴

¹⁰ Ibid.

¹¹ Ibid.

¹² Ibid.

¹³ Ibid.

¹⁴ Ibid.

State-Based Armed Conflict: “A state-based armed conflict is a contested incompatibility that concerns government and/or territory where the use of armed force between two parties, of which at least one is the government of a state, results in at least 25 battle-related deaths in one calendar year.”¹⁵

¹⁵ Ibid.

Chapter II

Background of the Problem

This chapter discusses the linguistic background, and conflict research, associated with this research, along with selected political and geographical issues to illuminate how they interact and form the context of this research.

Linguistics Background

Language is all around us. It is not only our means of daily communication, but also acts as a part of our ethnic identity. Therein lies a complication and topic of great debate amongst scholars. “Language” is often used interchangeably with “ethnicity,” which Christopher Anderson and Aida Paskeviciute argue is not the case. They state that:

[e]mpirically speaking, researchers have commonly used language as a proxy for ethnic variation in a country and have utilized so-called measures of ‘ethnolinguistic fractionalization’ to capture variability within a population along ethnic lines. [...] Suffice it to say for our purposes that using language as a proxy for ethnicity or relying exclusively on ethnicity or language to measure population heterogeneity is likely to miss important dimensions of variation within a population for two reasons: First, it assumes that ethnicity maps onto language one-for-one; second, it assumes that language has the same consequences for citizenship behaviors as ethnicity.¹⁶

¹⁶ Christopher J. Anderson & Aida Paskeviciute, “How Ethnic and Linguistic Heterogeneity Influence the Prospects for Civil Society: A Comparative Study of Citizenship Behavior,” *The Journal of Politics*, Vol. 68, No. 4, November 2006, 787.

Interestingly, from a linguistic perspective, it is only recently in our post World War II world that linguists have begun to consider the social implications of languages, as a result of nation state lines being re-drawn after the war.¹⁷ Indeed, a recent article from 2017 penned by Mike Medeiros addresses the role of linguistics in conflict, in which he summarizes what research has been conducted. He explains that

[t]he traditional reasons to explain the motives that push ethnic group members to enter into conflict or not, as well as the level of intensity they choose to adopt, have centered on economic inequalities and political differences. However, the magnification of these two major grievances has led collective injustices specifically to ethnicity to be somewhat overlooked. Although ethnic social cleavages are centered on a variety of group markers, sometimes even on multiple ethnic distinctions, the overwhelming majority of ethnic conflicts possess a linguistic difference between the conflicting parties. This suggests that linguistic factors might be a principal element in inter-group tensions. Language-based ethnic tensions are far from uniform. They involve movements that vary greatly in their demands; ranging from the simple desire to protect the group's language, as with the Frisians in the Netherlands, to the will to form an independent country, as in the case of the Catalans of Spain. They also have a high degree of variance in their intensity with some linguistic groups using peaceful means in order to have their demands acquiesced, such as South Tyroleans in Italy, while others take up arms and resort to violence, a deadly situation exemplified by the Karen in Myanmar. Yet, the role of linguistic factors in intergroup conflicts has been underexplored by scholarly research. One reason for this neglect is due to the fact that studies in political science and sociolinguistics remain disjointed. This situation impedes the understanding of language-based ethnic tensions and has led to a call for more collaboration between these different fields.¹⁸

In supporting Medeiros's claims, Marc Shell has explained that "linguistic conflict and cooperation now take place in such a bewildering variety of specific spaces and cases that the importance of language itself is in danger of being lost. Yet, in an age

¹⁷ Frederick J. Newmeyer, "The History of Modern Linguistics," *Linguistic Society of America*, 2022, <https://www.linguisticsociety.org/resource/history-modern-linguistics>.

¹⁸ Mike Medeiros, "The Language of Conflict" *Ethnicities*, (October 2017), Vol.17, No.5, pp. 627-645.

when language is being rediscovered as an ‘historical determinant,’ many wars that were once considered simply ‘religious’ or ‘nationalist’ turn out, on further reflection, to have been ‘linguistic’ as well.”¹⁹ According to Shell, because so many languages exist and are complex, scholars and society have overlooked the linguistic root of conflict.²⁰ Jeroen Darquennes, a professor of German linguistics, supports Shell’s claims and acknowledges the lack of literature on the subject, explaining that “essential for a good understanding of societal language conflict is that in some cases the social or other divisions that in other cases would lead to language conflict either go unnoticed or are not experienced as being problematic.”²¹ Dr. Peter Nelde, a leading sociolinguist, published an article in 1987, which addressed language contact and conflict occurring between speakers of languages. He stated that: “conflict plays a role in many social sciences. Linguistically, conflict between different ethnic groups often results from language contact. Problems viewed as political, economic, or sociological in nature are often rooted in linguistic conflict. In the literature, however, contact has overshadowed conflict.”²²

Conflict Research

Nils-Christian Bormann, Lars-Erik Cederman, and Manuel Vogt’s study on religion, language, and ethnic war also delves into how conflict researchers will typically

¹⁹ Marc Shell, “Language Wars,” *CR: The New Centennial Review* 1, no.2 (2001):1-17.

²⁰ Nell Lake, “Language Wars,” *Harvard Magazine*, March-April 2002.

²¹ Jeroen Darquennes, “Language conflict research: A state of the art,” *International Journal of the Sociology of Language*, (2015) 235: 7-32.

²² Peter H. Nelde, “Language contact means language conflict,” *Journal of Multilingual and Multicultural Development*, (1987) 8:1-2, 33-42.

adhere to the Barthian school of ethnic thought, which “defines ethnicity in terms of boundaries or cleavages rather than in terms of its specific contents.”²³ They note how other scholars like political scientists Laitin and Juergensmeyer “have singled out religion as being more conflict-prone than other ethnic dimensions. In the aftermath of the terrorist attacks of September 11, 2001, this hypothesis has gained considerable currency beyond academia, as illustrated by Tony Blair’s recent assertion that ‘religions difference will fuel this century’s battles.’”²⁴ They caution analysts not to put religious conflict or linguistic conflict in separate dimensions. They then conclude that the existing documentation that supports a religious foundation for violence is too biased.²⁵ Laitin’s work is notable, however, because he acknowledges language as its own entity. I disagree with Laitin that “conflict over language is not a prescription for violence. In fact, under certain potentially incendiary conditions, language conflict can help to contain violence”²⁶ because of various differences in multiple languages spoken within any given area.

Tristan Mabry shares this opinion and questions Laitin’s conclusions, where he “evaluates quantitative variables ascribed to characteristics of language communities, variables developed by [him] and James Fearon and eventually incorporated into the minorities at risk (MAR) dataset as ancestral language scores (ALS). In sum, such

²³ Nils-Christian Bormann, Lars-Erik Cederman, and Manuel Vogt, “Language, Religion, and Ethnic Civil War,” *Journal of Conflict Resolution*, 2017, 744-745.

²⁴ Ibid., 745.

²⁵ Ibid., 764-766.

²⁶ David D. Laitin, “Language conflict and violence: the straw that strengthens the camel’s back,” *Arch.europ.sociol.*, XLI, 2000, 98.

variables are rejected as incapable of representing that which is nominally under consideration: the relative importance of *language difference* in relation to cultural-cum-national conflict. Therefore, any conclusions drawn from analyses that employ such data are questionable.”²⁷ It would be remiss to not give Laitin, a political scientist, credit for trying to understand language conflict; however, he mainly focused on the ancestry of languages to try and quantify differences between them. Laitin brought into question the validity of language data, as respondents are known to not always be truthful when filling out surveys or providing census data.²⁸ Mabry argues that, “the proponents of language as a critical component of conflict are abundant because the logic of language difference is simple and compelling: ethnic differences lead to ethnic conflict; language differences equal ethnic differences; therefore, language differences are associated with ethnic conflict.”²⁹ It is a compelling argument, and one that must be acknowledged.

Manuel Vogt’s study of ethnic group mobilization compares overlap between religion and linguistics within ethnic groups, in which he too deviated from Barth’s ambient views of ethnicity. Vogt looked at “the lasting political impact of distinct ethnic cleavage types related to specific markers, such as race, language, and religion. Specifically, in contrast to the stratifying force of race, linguistic and religious segmentation of ethnic groups increases the risk of violent ethnic conflict. Previous studies that focused on specific ethnic markers also tended to regard religious divisions as particularly violence prone. By contrast, this study’s results suggest that linguistic

²⁷ Tristan James Mabry, “Language and Conflict,” *International Political Science Review*, 2010, 190.

²⁸ Ibid., 195.

²⁹ Ibid., 191.

segmentation might actually be more critical than religious segmentation, especially for separatist violence.”³⁰ I sympathize with Vogt’s desire to glean the political impact of these types of conflicts.

Political Issues

Anastasia Smirnova and Rumen Iliev shed light on politicization of language.

They state that:

[o]ne factor that is particularly relevant for ethnic conflicts but received little empirical attention in studies on language attitudes is the politicization of language, where political views become closely linked to language ideology. Early proponents of nationalism have essentialized language as the key element of nation building, associating it with purity, authenticity, unity, and historical continuation of a cultural group. [...] Since the proliferation of nation-states as a form of governance in the 19th and 20th centuries, in many countries language has become closely related to nationality. In multiethnic societies, official recognition of a minority language often becomes a divisive issue, seen by some as a basic human right, but seen by others as an existential threat to the state. Nationalist politicians, accordingly, recognize the role that language plays in building a group identity and mobilizing supporters, and often make it a central symbol of the political debate.³¹

This too has a trickle effect into the media; if the media are only catering to a select number of language speakers in a given area, this politicization can cause linguistic misrepresentation.³² Charles King argues that while many are familiar with the “language of politics,” there has not been much done in regards to “the politics of

³⁰ Manuel Vogt, “Ethnic Stratification and the Equilibrium of Inequality: Ethnic Conflict in Postcolonial States,” *International Organization* 72, Winter 2018, 133.

³¹ Anastasia Smirnova and Rumen Iliev, “Political and Linguistic Identities in an Ethnic Conflict,” *Journal of Language and Social Psychology*, Vol. 36(2) 211-225, 2017, 213.

³² Philip Mpofu & Davie E. Mutasa, “Language policy, linguistic hegemony and exclusion in the Zimbabwean print and broadcasting media,” *South African Journal of African Languages*, 34:2, 2014, 226

language,” which is especially notable considering that in certain parts of the world, linguistic status has helped to shape inter-ethnic conflicts or states of harmony.³³ Indeed, many national governments purposefully discriminate against minority languages, making language control a policy objective. A perfect example is the continuing debate on whether states will support bilingual learning, as Spanish becomes a more prevalently spoken language within the United States.³⁴ King warns that “the challenge for security studies specialists and international relations experts, though, is to understand the ways in which language issues can become transformed into life-and-death struggles between contesting ethno-linguistic groups.”³⁵

Geography

In addition to sharing a common language, ethnic groups often also share a common homeland or geographic area.³⁶ Geography has been viewed as both an opportunity for conflict, as well as a motivating factor for conflict.³⁷ Homelands in particular have “the fundamentals of culture and identity. And, as such, [they are] about sustaining cultural boundaries and boundedness...The ‘other’ is always and continuously a threat to the security and integrity of those who share a common home.’ A ‘homeland’

³³ Charles King, “Policing Language: Linguistic Security and the Sources of Ethnic Conflict,” *Security Dialogue*, Vol. 28(4), 1997, 494.

³⁴ Ibid., 496.

³⁵ Ibid.

³⁶ The author of thesis acknowledges that people are often displaced and that diasporas exist.

³⁷ Nils B Weidmann, “Geography as Motivation and Opportunity,” *Journal of Conflict Resolution*, Volume 53 Number 4, August 2009, 529.

is therefore a special category of territory: it is not an object that can be exchanged, but an indivisible attribute of group identity. This feature explains why ethnic groups rationally view the right to control their homeland as a survival issue, regardless of a territory's objective value in terms of natural or man-made resources. [...] Homeland control ensures that a group's language can be spoken, its culture expressed, and its faith practiced. This intimate connection between homeland territory and the preservation of identity distinguishes ethnic groups from states.”³⁸ This phenomenon used to be studied in a non-geographic way through the Minorities at Risk database, but the data was not helpful in a geographic sense, as it was only a compilation of settlement characteristics. In the 1960s, cartographers from the former Soviet Union compiled the Atlas Narodov Mira. Even though the Atlas's data is outdated, Nils Weidmann still used it, with GIS software, in 2009 to understand ethnic conflicts more clearly.³⁹

Weidmann's research led me to focus on languages spoken in areas of conflict in general. However, unlike other researchers who have focused on drawing comparisons between languages within a certain area, I was more curious as to how many languages are spoken in total within a conflict zone. In a conflict of nation state versus nation state, do the opposing sides have similar linguistic diversity? Does the presence of linguistic diversity have the potential to cause a conflict?

To answer these questions, I completed Harvard University's GIS summer institute and have been working to create maps using conflict data and the languages of

³⁸ Monica Duffy Toft, “Indivisible territory, geographic concentration, and ethnic war,” *Security Studies*, 12:2, 2002, 87.

³⁹ Nils B Weidmann, “Geography as Motivation and Opportunity,” *Journal of Conflict Resolution*, Volume 53 Number 4, August 2009, 526-541.

the world, with the intention of determining whether or not my research could support or refute previous studies on language conflict. Through mapmaking, I worked to evaluate my contention that language is one of the key elements that needs more consideration when there is conflict in a region. With a greater focus on language, it is possible that conflict resolution would be far easier to achieve. By comparing the conflict area to the number of languages spoken, I hypothesize and will evaluate the following:

- Hypothesis: a country that has multiple languages spoken within its borders is more likely to encounter international or a national conflict than a single language speaking country

To evaluate the hypothesis above, I used data from Uppsala University's Conflict Data Program, provided through their Department of Peace and Conflict Research, established in the 1970s. The Program's mapping tool displays conflict information from 1975-2020, and the data are filterable by state-based violence, non-state violence, and one-sided violence. I then joined Uppsala data with linguistic data from Ethnologue, which is recognized as one of the world's leading linguistic resources.

Although data does exist on conflicts and languages, I wanted to be able to join datasets and project this data comprehensively onto a map. However, trying to display multiple languages within a particular area of a map is a challenge in GIS. Trying to "force language into the realm of points, lines, and polygons requires endless compromises between reality and representation."⁴⁰ Weidmann used maps from the 1960s in order to try to map out linguistic differences. It is no wonder monolingual mapping is more the status quo. Indeed, other GIS practitioners from Virginia

⁴⁰ Candice R. Luebbering, Korine N. Kolivras & Stephen P. Prisley, "Visualizing Linguistic Diversity Through Cartography and GIS," *The Professional Geographer*, 65:4, 581.

Polytechnic Institute and State University have stated that: “while linguistic diversity is an integral component of cultural landscapes, the spatial depiction of languages fails to represent all community members. Language is difficult to map and established guidelines are lacking. The perception of power conveyed is arguably the most meaningful design issue of in language mapping, as most language maps inaccurately show one language per place.”⁴¹ Despite these cautions, I was still determined to map a relationship between language and conflict. Therefore, I compromised and used counts of languages per country, rather than trying to map out what could very well be obsolete observer-reported data.

In joining the data, my hope was to evaluate my contention that language has been a piece of the puzzle that international relations theorists have overlooked, and perhaps discover that there is correlation between linguistic diversity and conflict. I hope that my findings can open a new pathway of research study for students of international relations.

⁴¹ Ibid., 580.

Chapter III

Research Methodology

This chapter discusses the language and conflict data used in this research, along with the visual analyses and regression methods. In addition, this chapter includes a discussion of this exploratory research's limitations.

The most accurate resource containing the language data needed to conduct both my visual analysis and to be able to run regressions in this research study is *Ethnologue's* language and shapefile data. The *Ethnologue* dataset that was used for the language shapefiles is from 2004, as that is the most recent version available to obtain from the Harvard Library System. All maps created are based on the 2004 data findings as they were known and captured by *Ethnologue* at that time. Therefore, this study is based on the reporting of the number of languages and country boundaries at that time. In other words, the count of languages used within a given country as of 2004 is my independent or explanatory variable.

The maps visuals I have created were made using ArcGIS®, which is a software created by Esri, the Environmental Systems Research Institute.⁴² My first step was to join the shape file of country boundaries from *Ethnologue* to the table of all the world's languages that existed in 2004. This was called a table join, as I joined a table of the languages to the countries they are spoken in spatially to areas on the Earth's surface. Throughout my map legends, the number of languages present in any given country are

⁴² Esri. 2022. www.esri.com.

listed and the corresponding colors are presented by country polygon in a choropleth map. The number of deaths per conflict type are denoted using a graduated point symbology in both the legend and on the maps themselves.⁴³

The next step in my process was a bit complicated, requiring joining the shape country boundaries spatially to the distinct types and number of conflicts that occurred within each year. It is here that my methodology changed. I was originally going to map out several decades worth of conflict and compare interstate to intrastate language conflicts, but due to the difficulty in doing that using ArcGIS®, I determined for the purpose of this study it would be better to map several years post-9/11, as the way we view conflict drastically changed after that event. I chose instead to map the following years: 2004, 2010, 2015, and 2020.

Since I wasn't focusing on just minority & ethnic groups, I chose not to use the GeoEPR, MAR or ANM datasets. In addition, the ANM data set is outdated.⁴⁴ I chose to utilize the Uppsala Data Program's conflict datasets, which are recognized for their high quality. I compared the following measures of conflict datasets in each of the four-year samples. In sum, my dependent variable is conflict and is comprised of six distinct measures of conflict. These diverse measures were used in recognition that conflict is complex and therefore I tested each of them individually. Each of these conflict measures is described below:

⁴³ Iberhard, David M., Gary F. Simons, and Charles D. Fenning (eds). 2004. *Ethnologue: Language of the World*. Dallas, Texas: SIL International.

⁴⁴ Wucherpfennig, Julian, Weidman, Nils B., Girardin, Luc, Cederman, Lars-Erik, and Wimmer, Andreas "Politically Relevant Ethnic Groups across Space and Time: Introducing the GeoEPR Dataset," *Conflict Management and Peace Science* Vol 28(5): 423–437.

Uppsala Data Program Conflict Datasets	Definitions
UCDP/PRIO Armed Conflict Dataset version 21.1 https://ucdp.uu.se/downloads/replication_data/2021_ucdp-prio-acd-211.pdf	a contested incompatibility that concerns government and/or territory where the use of armed force between two parties, of which at least one is the government of a state, results in at least 25 battle-related deaths in a calendar year
UCDP Dyadic Dataset version 21.1 https://ucdp.uu.se/downloads/dyadic/ucdp-dyadic-211.pdf	builds on the UCDP/PRIO Armed Conflict dataset, but goes beyond the conflict level and focuses on dyads within each conflict. As such, it constitutes a disaggregated version of the UCDP/PRIO Armed Conflict dataset
UCDP One-Sided Violence Dataset version 21.1 https://ucdp.uu.se/downloads/nsos/ucdp-onesided-211.pdf	the use of armed force by the government of a state or by a formally organized group against civilians which results in at least 25 deaths. Extrajudicial killings in custody are excluded
UCDP Non-State Conflict Dataset version 21.1 https://ucdp.uu.se/downloads/nsos/ucdp-nonstate-211.pdf	the use of armed force between two organized armed groups, neither of which is the government of a state, which results in at least 25 battle-related deaths in a year
UCDP Battle-Related Deaths Dataset version 21.1 https://ucdp.uu.se/downloads/brd/ucdp-brd-codebook-211.pdf	a contested incompatibility that concerns government and/or territory over which the use of armed force between two parties, of which at least one is the government of a state, has resulted in at least 25 battle-related deaths in one calendar year
Deadly Electoral Conflict Dataset (DECO) version 1 https://ucdp.uu.se/downloads/deco/DECO_codebook_1.0.pdf	violence that is substantially linked to an electoral contest.' Central to the definition is the direct connection between the violence and the dynamics of the electoral process enacted through political parties, voters, candidates, the polling process, or institutional arrangements surrounding elections

Uppsala Conflict Datasets Table 1

I have previously worked with the Uppsala Conflict Data Program, and I knew that when it came to conflict data, they would have the datasets that I needed to run this study. I also knew that the data would be able to be mapped out, since the program itself has a mapping feature where you can see the conflicts on the world map. Based on my previous experience with this tool, I focused on the UCDP datasets for this exploratory research.

The UCDP / Peace Research Institute Oslo (PRIO) armed conflict dataset was the first dataset reviewed. It was appropriate to map that data at the national level because

the government of a state must be one of the acting parties. The Dyadic dataset, also focusing on armed conflict, was the second dataset reviewed and I almost discarded it because its dyadic measure spanned two countries. However, I was curious to see if there would be different results from the PRIO armed conflict dataset as they both contained some of the same data from the same data source. I next analyzed One-Sided Violence, which involved attacks on civilians. I had previously looked at some of the data before and knew the value of this dataset. The fact that this dataset also yielded correlation with linguistic diversity when I ran regressions after mapping the conflicts was of little surprise. I surmised if there were any conflict measure to present a correlation between language and conflict, this would be it. I will be continuing my research with this dataset. I then reviewed the non-state conflict dataset, because I thought it would be interesting to analyze conflicts where a government was not involved. This is a dataset I may further explore, because of the encouraging preliminary correlation.

I had hoped that the Battle Related Deaths dataset would reveal a correlation with linguistic diversity since it is a direct measure of the intensity of a conflict in terms of fatalities, however, I was disappointed to find that there was no correlation with language diversity. I might eventually use this dataset again in the future, but in the interim, I will be focusing on the other datasets that yielded correlations. Finally, I decided to map out the Deadly Electoral Conflict dataset to see if its focused measure of conflict explicitly related to elections would yield any results. This dataset was not of a particular interest to me, but I did not want to discard it without testing it. As this dataset was last completed and published in 2017, I am interested in it, given the events of the last few years. I am

currently waiting to see when the next dataset will be released and will investigate the correlation with linguistic diversity to see if there is any change or not.

I downloaded and saved each of the six datasets of conflict measures, filtered them by year, saving the 4 years needed, and spatially joined each of these individual one-year datasets (23 in total) to the shapefile for the countries. This approach enabled me to visually see how many incidents, used here as a measure of conflict intensity, had occurred during each year within each country. I then joined the language count data to the conflict data before displaying them both together on a map to see if any inferences could be made from a visual inspection. Since I had created several tables in the process, I was able to easily join these two new attribute tables together in a table join and see all the data together in one table.

The next chapter presents these visual analyses between language diversity and each of the six measures over the four sample years. This analysis approach was conducted to be able to visually see on a map where conflicts took place and the number of languages that are spoken within those areas of conflict. Following the analysis of these visualizations, the results of the exploratory regression analysis follow to assess the strength of the bivariate relationship between language diversity and the six conflict measures.

Despite extensive research, I could find no record of a researcher having previously mapped conflict incidents together with counts of the languages spoken in the conflict zone to evaluate if there was a relationship between the two, and if any inferences could be made by looking at the maps. While there are conflict datasets and separate language diversity datasets, combining the data remains a challenge. To the

extent that my hypotheses are supported, I would like eventually to develop skills with coding and building custom algorithms within ArcGIS and potentially use it as a tool for predicting conflict.

Several of these limitations have been addressed in previous chapters. Future issues that I can foresee occurring include the possibility that, after mapping out more conflict data and then running regressions on that data, little or no significant correlations may be found. Another limitation is that there is little to no previous research that has been conducted and recorded in this field, so I do not have access to a different researcher's work to which I can compare my findings.

Perhaps the most limiting aspect of my exploratory research is the use of a single independent explanatory variable (language diversity) to assess conflict. Clearly there are social, political, economic, and environmental variables that contribute to the onset, intensity, and duration of conflict, and again, conflict is a complex phenomenon. Thus, to pursue this effort to test my hypothesis would require the consideration of multiple additional explanatory variables. In sum, the technical and temporal constraints of this thesis research limited the findings to preliminary status.

A question remains of what to do with these findings, and specifically, how to ensure that other students of international relations consider the role of language in conflict and that this approach continues to be utilized to address language conflict.

While in recent years maps have often been criticized as outdated, lacking in creativity, or boring,⁴⁵ I am hoping the maps in this research project will not be deemed

⁴⁵ Candice R. Luebbering, Korine N. Kolivras & Stephen P. Prisley, "Visualizing Linguistic Diversity Through Cartography and GIS," *The Professional Geographer*, 65:4, 581.

as such and perhaps someone in the field of linguistic geography might find this work to be of value. The following lengthy quotation succinctly captures the fractured state of the field of linguistic geography. As a linguist, I hope that this exploratory research contributes to both the fields of linguistics and conflict research.

[The] field of linguistic geography, in spite of the interdisciplinary overtones of its name, has been practiced primarily by linguists, with limited interaction by geographers. Whereas linguists are interested in the internal workings of language systems, geographers tend to treat language as a unitary variable and have not engaged significantly in research on its internal complexity. This state of affairs was lamented by Wagner (1958) and remained unchanged throughout the remainder of the 20th century in spite of his agitation for more geographer involvement in the sorts of questions that interest historical linguists. Even among linguists, however, linguistic geography is a fractured field. Spatial patterns and geographic relationships are typically secondary to linguists' other research aims such as diagnosing linguistic relationships, modeling the dynamics of change, or describing structural diversity among the world's language. As a result, geography is relegated to the periphery of various linguistic subfields more often than it is treated as a unified topic in diachronic linguistics.⁴⁶

⁴⁶ Hannah J. Haynie, "Geography and Spatial Analysis in Historical Linguistics," *Language and Linguistics Compass*, 2014, 344.

Chapter V.

Results

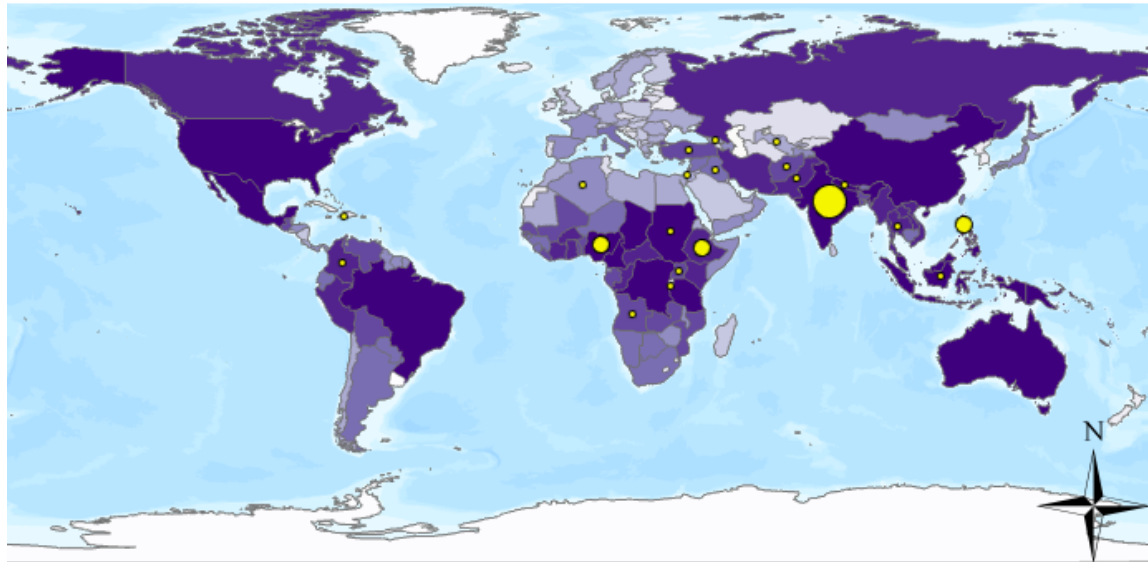
As I was mapping each of the datasets, I visually saw a pattern emerge on the maps: in countries that speak multiple languages, there were conflicts present vs those that only have one language spoken. I estimated that countries that speak fewer languages are less likely to experience conflict than those that speak more languages. Correspondingly, I presumed that a country with hundreds of languages was far more likely to have conflicts than a country that did not. As you can see from the graduated point symbols within countries, there was only one instance of a single language speaking country being involved in conflicts during these conflict years, which I will discuss later.

UCDP/PRIO Armed Conflict

I was most intrigued by the PRIO armed conflict dataset, which is “a conflict-year dataset with information on armed conflict where at least one party is the government of a state in the time period 1946-2020.”⁴⁷⁴⁸ I wanted to see if there would be many instances of conflict involving the government of a state, or if the likelihood of this happening would be lower in a context of more linguistic homogeneity.

⁴⁷ Pettersson, Therese, Shawn Davis, Amber Deniz, Garoun Engström, Nanar Hawach, Stina Högladh, Margareta Sollenberg & Magnus Öberg (2021). Organized violence 1989-2020, with a special emphasis on Syria. *Journal of Peace Research* 58(4).

⁴⁸ Gleditsch, Nils Petter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg, and Håvard Strand (2002) Armed Conflict 1946-2001: A New Dataset. *Journal of Peace Research* 39(5).



Esri, FAO, NOAA, USGS, Esri, USGS

Legend

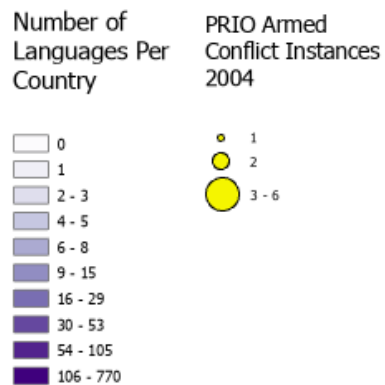


Figure 0-2 PRIO Armed Conflict Dataset 2004

Figure 0-2 shows that in 2004, India showed the highest number of armed conflicts. The government of India fought against groups of insurgents that year, including the National Democratic Front of Bodoland (NDFB),⁴⁹ who were fighting for sovereignty. Of particular note, the NDFB criticized the written script of the Bodo

⁴⁹ Ibid.

language. The group was primarily composed of Christians, who wanted Roman script to be used, rather than the Devnagri script.⁵⁰ This incident heavily supports the contention that linguistics was a major factor in conflict that year.

In comparing Figure 0-2 and Figure 0-3, India continued to have armed conflicts, but the number of conflicts decreased in 2010. In evaluating the PRIO dataset, the same

⁵⁰ “National Democratic Front of Bodoland,” South Asia Terrorism Portal, 2001.
https://www.satp.org/satporgtp/countries/india/states/assam/terrorist_outfits/ndfb.htm.

actors on both sides of the conflict were still involved.⁵¹⁵²

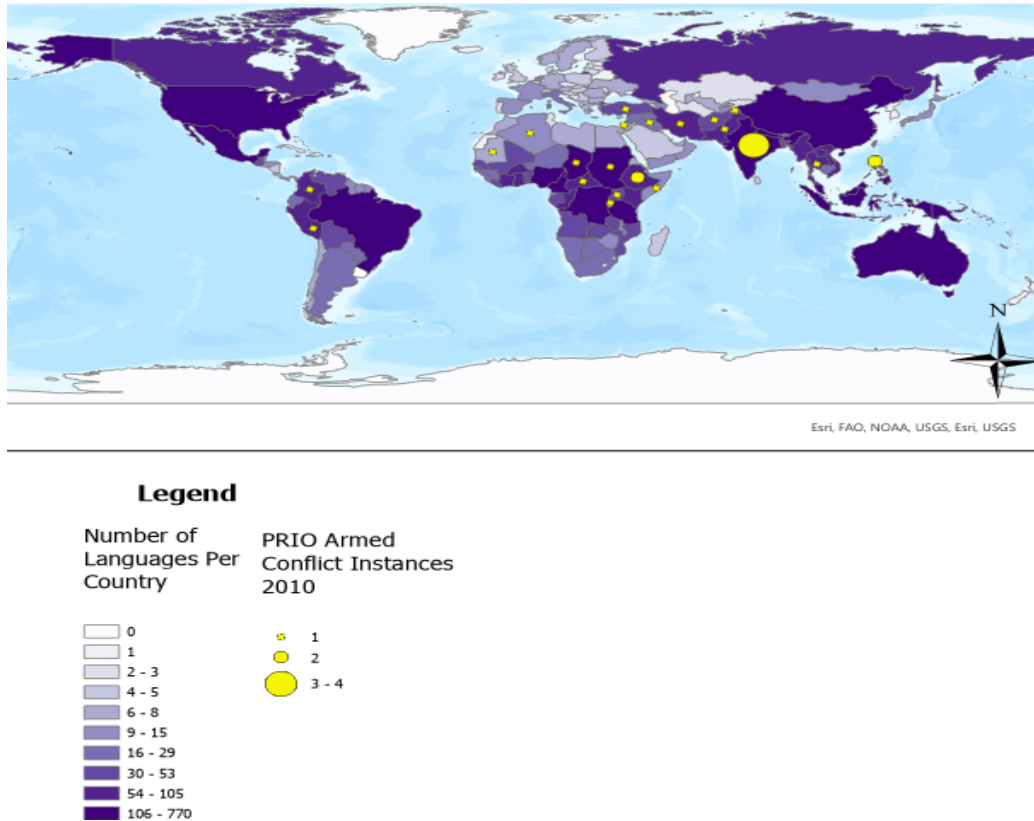


Figure 0-3 PRIO Armed Conflict 2010

Then in comparing Figure 0-3 to Figure 0-4 and evaluating the conflicts of 2010 and 2015, India still had a larger number of armed conflicts, but there was a rise in conflicts in Africa and Southwest Asia. In further exploring the PRIO dataset, we see that

⁵¹ Pettersson, Therese, Shawn Davis, Amber Deniz, Garoun Engström, Nanar Hawach, Stina Högladh, Margareta Sollenberg & Magnus Öberg (2021). Organized violence 1989-2020, with a special emphasis on Syria. *Journal of Peace Research* 58(4).

⁵² 52 Gleditsch, Nils Petter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg, and Håvard Strand (2002) Armed Conflict 1946-2001: A New Dataset. *Journal of Peace Research* 39(5).

these main conflicts occurred in Syria and Mali, both of which saw Islamist armed groups engaged in conflict with the governments of these countries during this time period.⁵³

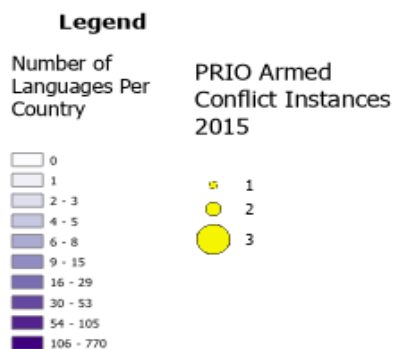
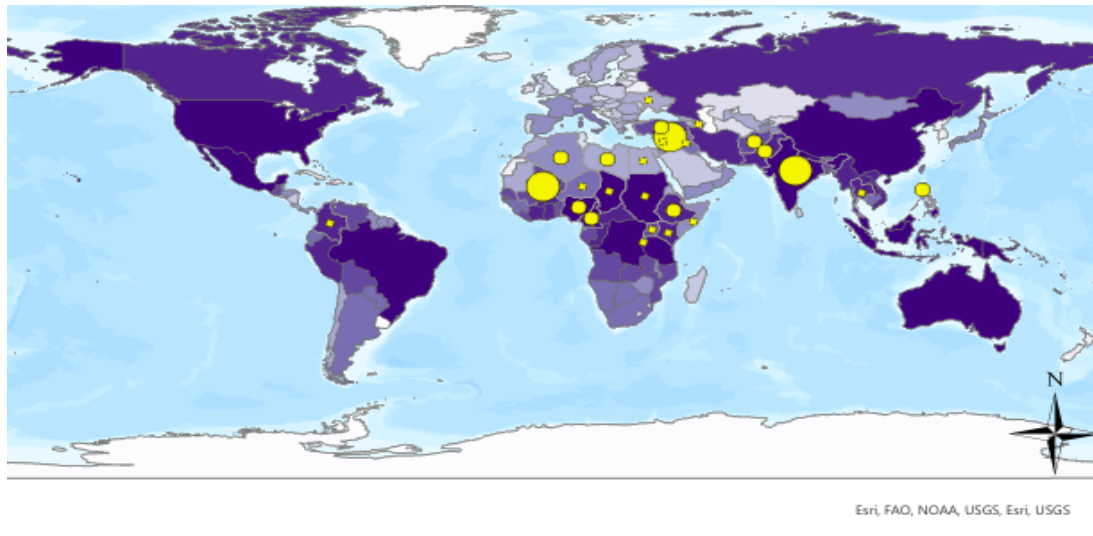


Figure 0-4 PRIO Armed Conflict 2015

Between 2015 and 2020, the counts go down in most cases, as Figures 0-4 and 0-5 demonstrate, but there is an increase in conflict in the Philippines between the government of the Philippines and the communist party of the Philippines and the Islamic

⁵³ Ibid.

State (IS).⁵⁴ This was at first a bit puzzling, but it is important to note that the government of the Philippines has been combating a war on drugs, and tensions escalated in 2020, which we can correlate to the COVID-19 pandemic and nationwide lockdown.

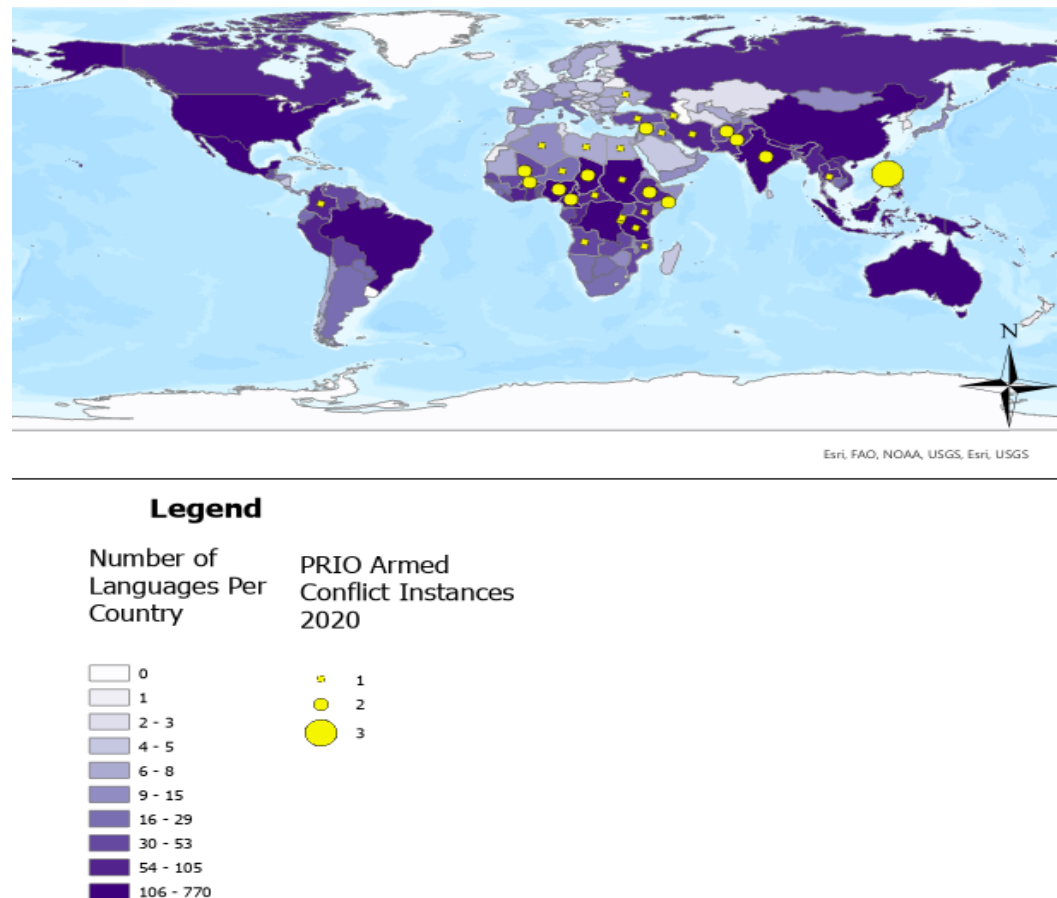


Figure 0-5 PRIO Armed Conflict 2020

As I suspected though, the general pattern in Figures 0-2 – 0-5 is that countries with more languages experienced more conflicts than those that speak fewer languages. This also corresponds to global events. I then wanted to compare this data to the Dyadic Dataset to see if there was a notable difference or not, as both datasets pertain to conflict with one of the actors being a government.

⁵⁴ Ibid.

UCDP Dyadic Dataset

This dataset is a “dyad-year version of the UCDP/PRIO Armed Conflict Dataset. A dyad consists of two opposing actors in an armed conflict where at least one party is the government of a state.”⁵⁵⁵⁶ I expected to see Figures 0-6 – 0-9 look very similar to Figures 02 - 05, and so I was surprised to see the number of conflicts higher than the PRIO dataset had reported, especially for India.

⁵⁵ Pettersson, Therese, Shawn Davis, Amber Deniz, Garoun Engström, Nanar Hawach, Stina Högladh, Margareta Sollenberg & Magnus Öberg (2021). Organized violence 1989-2020, with a special emphasis on Syria. *Journal of Peace Research* 58(4).

⁵⁶ Harbom, Lotta, Erik Melander & Peter Wallensteen (2008) Dyadic Dimensions of Armed Conflict, 1946-2007. *Journal of Peace Research* 45(5).

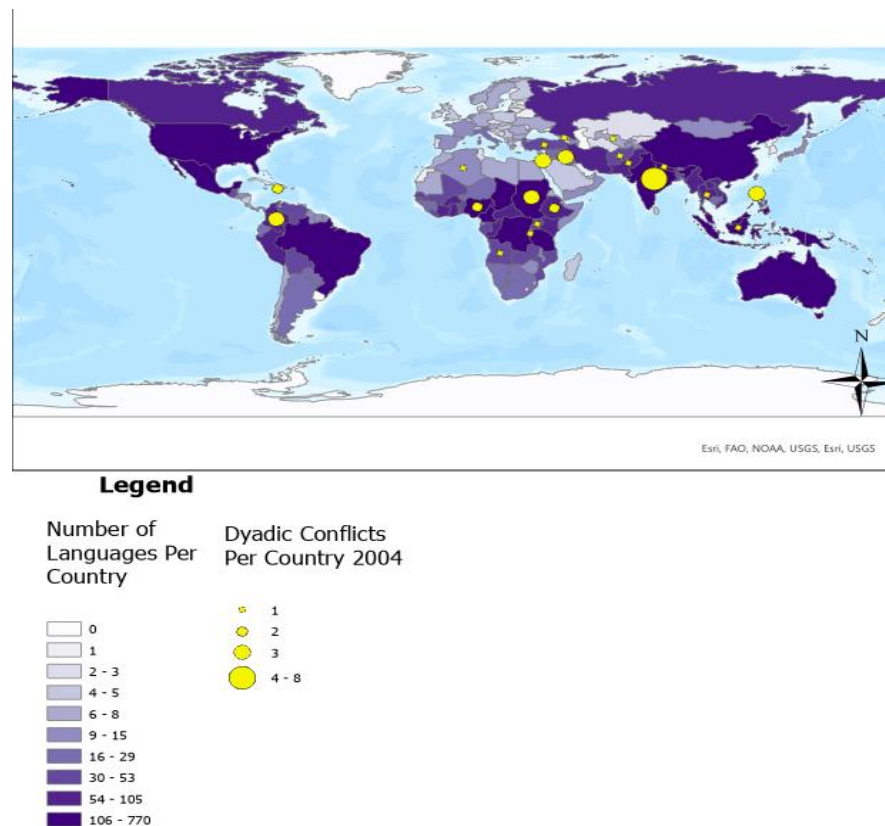


Figure 0-6 Dyadic Dataset 2004

The conflicts captured in Figure 0-6 aligned with the same groups as had been active in Figure 0-2 for India, although there were a few more instances of conflict captured with this dataset.⁵⁷ Figure 0-7 also aligned with Figure 0-3, again, same year, same groups, remaining consistent.⁵⁸

⁵⁷ Ibid.

⁵⁸ Ibid.

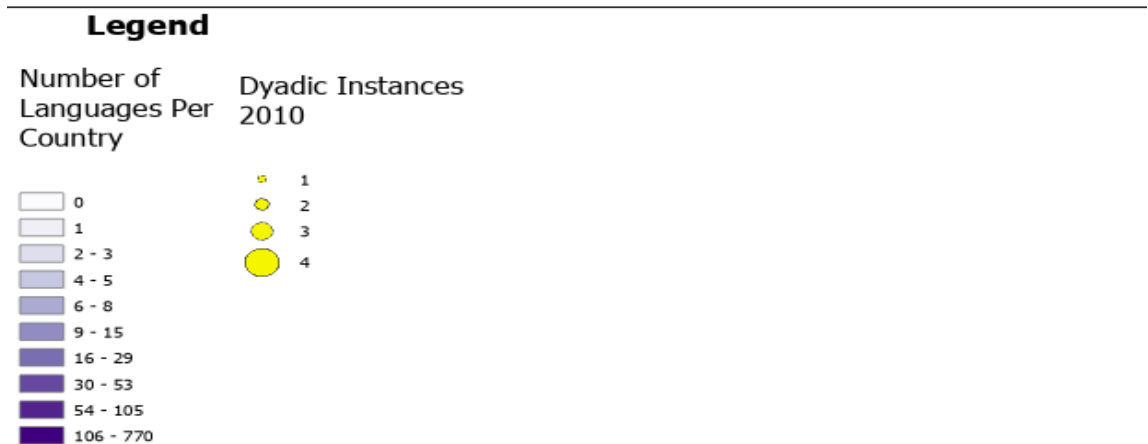
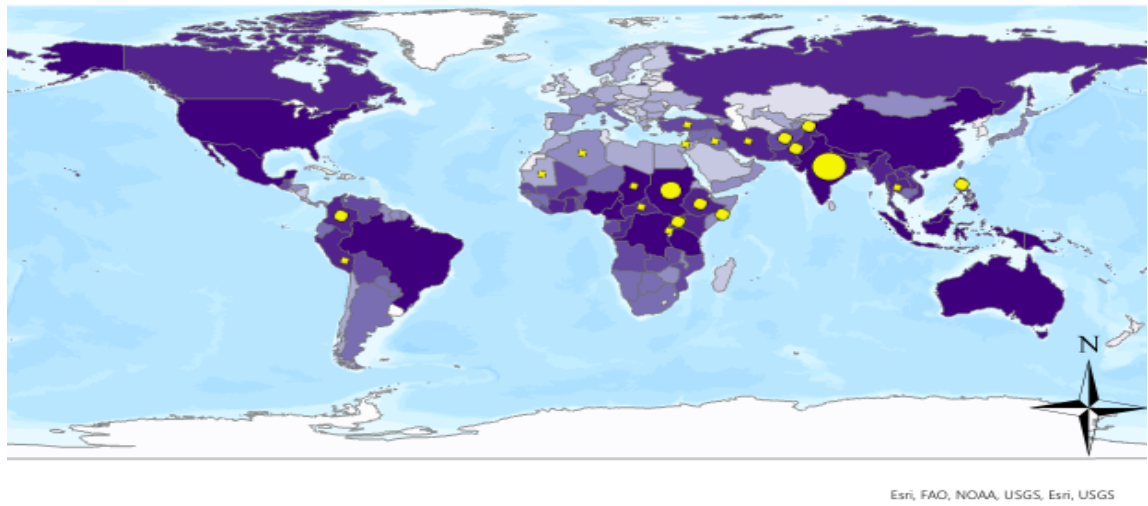
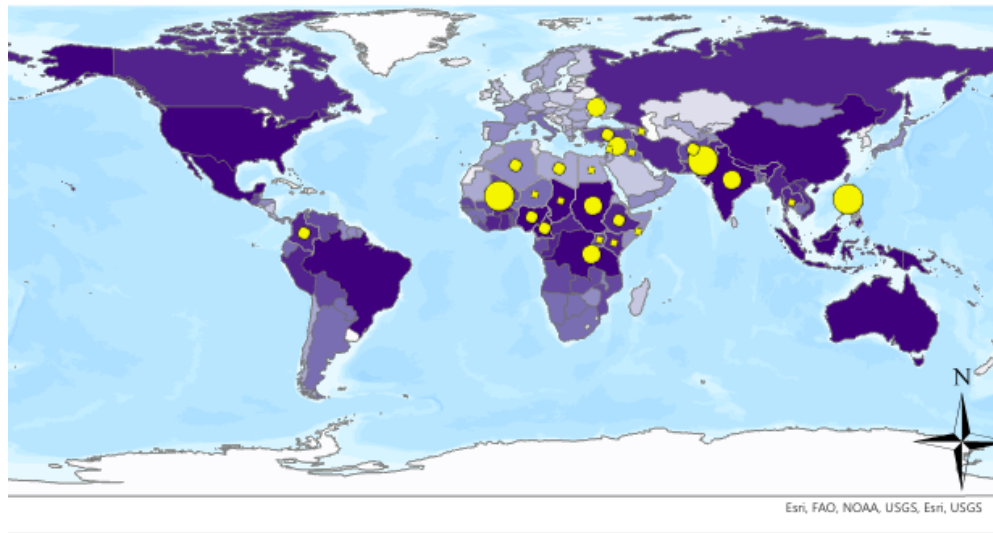


Figure 0-7 Dyadic Dataset 2010

An interesting pattern emerges in Figure 0-8 with four conflicts being prevalent in Mali, the Philippines, and Pakistan. While we know that based on the dataset the same actors were present in Mali and the Philippines as were in the PRIO dataset, it is notable that the conflicts that occurred in Pakistan in 2015 were between the Pakistani Urdu speaking government and different linguistic speaking insurgent groups, and these conflicts were initially begun by tribal members.⁵⁹

⁵⁹ Ibid.



Legend

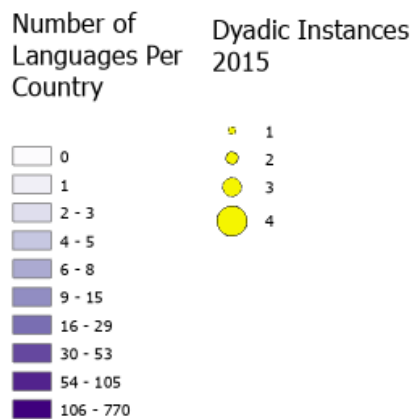


Figure 0-8 Dyadic Dataset 2015

Sadly, the incidents of violence that occurred between the government of Afghanistan, IS, and the Taliban in 2020 can now be considered foreshadowing for the events of 2021. Figure 0-9 in this regard is not surprising for Afghanistan. Aligning with the PRIO dataset, the same conflict between the same groups is again recorded for 2020. What is also unsurprising are the number of conflicts recorded below for the Democratic Republic of the Congo (DRC). Again, with the pandemic, it is not surprising that the government of the DRC and rebel groups like the Kata Katanga (which means secede)

clashed.⁶⁰ The fact that DRC also hosts so many of the world's internally displaced people, who all speak many different languages, likely is a cause of conflict.

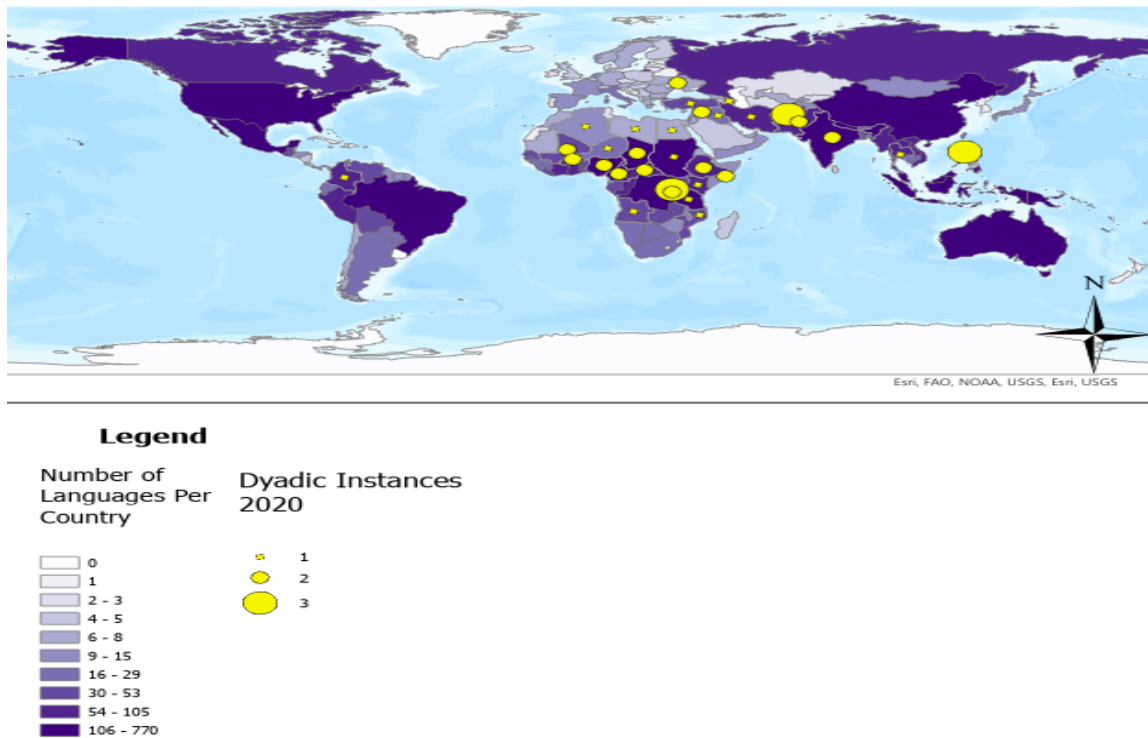


Figure 0-9 Dyadic Dataset 2020

After reviewing the Dyadic Datasets, and especially the findings of Figure 0-9, I began to consider what the attacks specifically on civilians look like, in contrast to the actor vs. actor type conflicts that one normally thinks about with conflicts. The human dimension of conflict is often forgotten, and that is something that I hope to research.

⁶⁰ Ibid.

UCDP One-Sided Violence Dataset

This dataset is “an actor-year dataset with information of intentional attacks on civilians by governments and formally organized armed groups.”⁶¹⁶² On a positive note, there were far fewer recordings of violence against civilians than I originally suspected within this dataset. However, I suspect there are actually far more instances of violence against civilians than were recorded or reported. For example, the lack of civilian attacks reported in the Middle East for Figure 0-10 stood out, especially since this was near the beginning of the Iraq War. A reason for this could very well be a lack of technology during that time, or simply nobody was recording this data. Although this explanation is anecdotal, it is based on what has been seen recently with the Russian invasion and subsequent attacks on Ukrainian civilians in Ukraine in 2022.

As seen in the other datasets, it was of no surprise that attacks on civilians in India in 2004 were launched by the NDFB, and other insurgents active within India at that time.⁶³

⁶¹ Pettersson, Therese, Shawn Davis, Amber Deniz, Garoun Engström, Nanar Hawach, Stina Högladh, Margareta Sollenberg & Magnus Öberg (2021). Organized violence 1989-2020, with a special emphasis on Syria. *Journal of Peace Research* 58(4).

⁶² Eck, Kristine & Lisa Hultman (2007) Violence Against Civilians in War. *Journal of Peace Research* 44(2).

⁶³ Ibid.

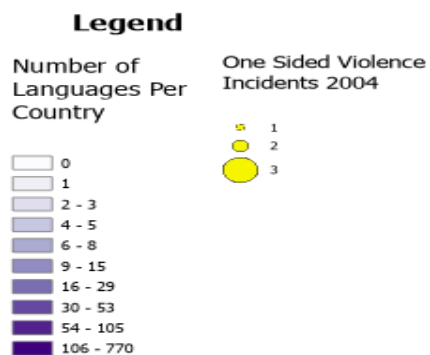
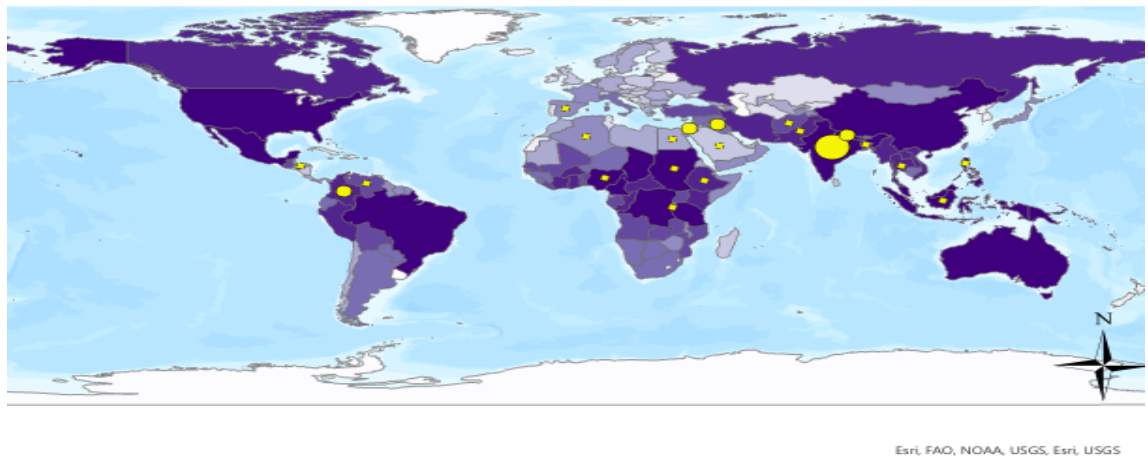
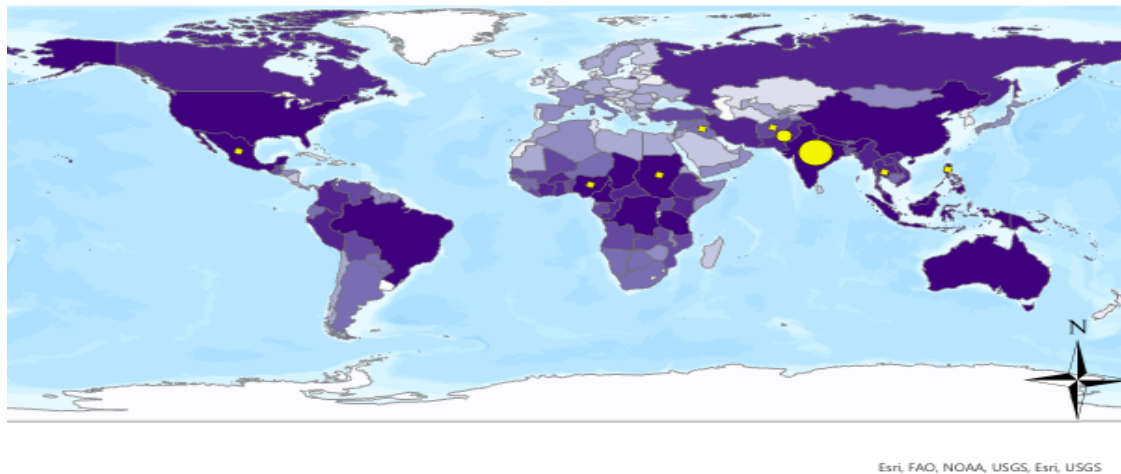


Figure 0-10 One-Sided Violence 2004

A slight change occurred between Figure 0-10 and Figure 0-11, highlighting my theory that the lower number of attacks was due to a lack of reporting and lack of technology. India again experienced attacks from insurgents and the NDFB in the fight over Bodoland.⁶⁴

⁶⁴ Ibid.



Legend

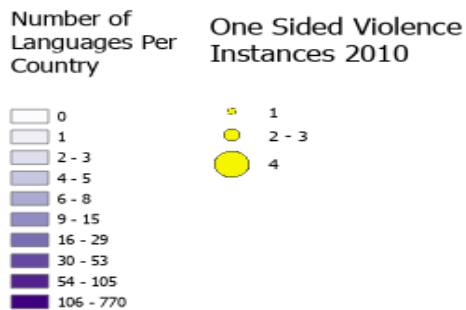


Figure 0-11 One-Sided Violence 2010

In comparing Figure 0-11 to Figure 0-12, attacks against civilians in India subsided, while attacks against civilians in the Central African Republic escalated. These attacks were by the Lord's Resistance Army, anti-Balaka, Popular Front for the Rebirth of the Central African Army, Union for Peace in the Central African Army, and Democratic Front for the People of Central Africa,⁶⁵ suggesting these attacks could be for religious or political reasons. However, it is of note that there are over 54-105 different

⁶⁵ Ibid.

languages spoken in this particular country, suggesting that the large number of languages could have contributed to these escalations.

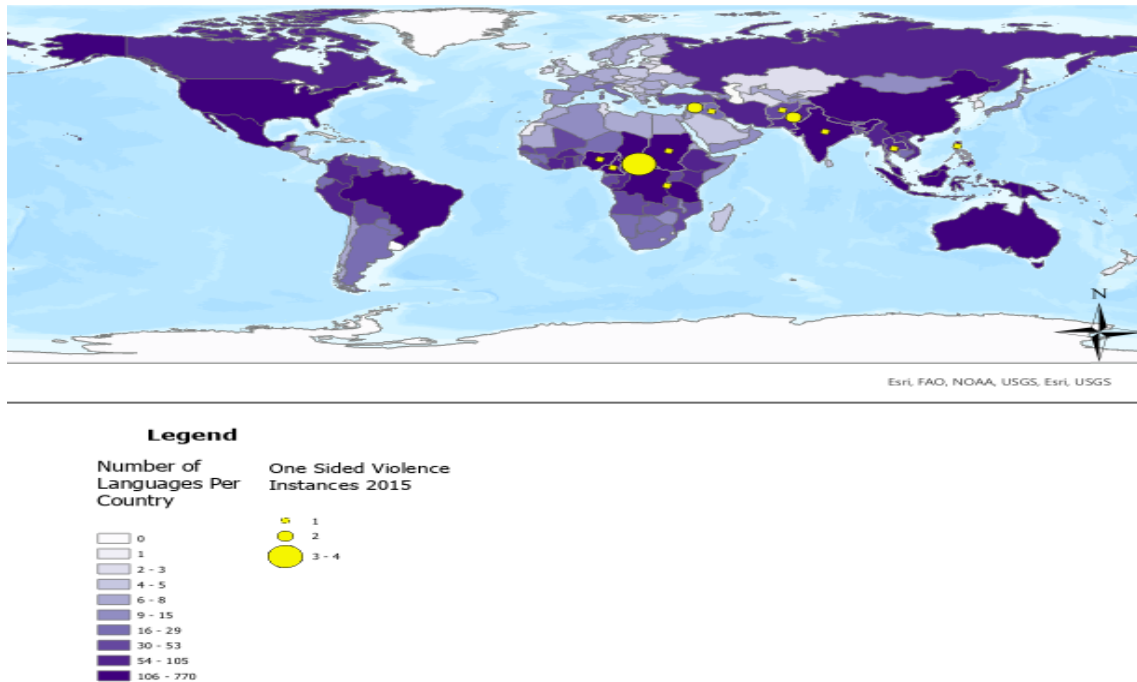


Figure 0-12 One-Sided Violence 2015

Figure 0-13 shows a shift from the Central African Republic to Ethiopia, where the governments of Ethiopia and ethnic militias from Eritrea launched attacks on civilians. Ethiopia's people also speak many different languages, and with ethnic troops attacking from a neighboring country, I can state with near certainty that linguistic divide has contributed to these conflicts.

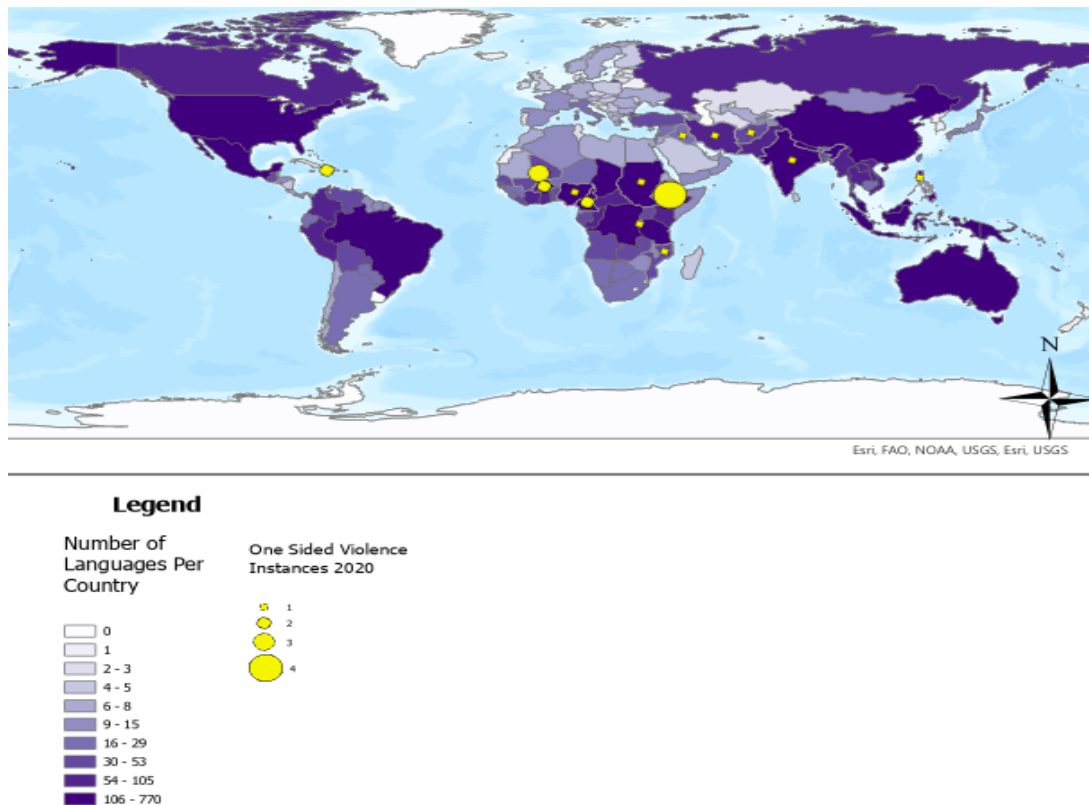


Figure 0-13 One-Sided Violence 2020

What this dataset in particular has revealed is the only instance of conflict in countries with fewer than ten languages involved in one-sided violence, and that was in Haiti. What could account for this? According to the Corruption Perceptions Index as reported by Transparency International, Haiti was most recently ranked in 2021 as 164/180 globally in terms of public sector corruption.⁶⁶ That alone, coupled with the economic hardships faced in the country, can account for the anomaly we see of violence happening in a country that speaks less than ten languages. It will be interesting to see if

⁶⁶ Transparency International (2022), Our Work In: Haiti, web, [Haiti - Transparency.org](https://www.transparency.org/en/haiti).

there are other instances of these one-sided conflicts happening within Haiti in other years between 2000 – present day.

UCDP Non-State Conflict

The Non-State Conflict dataset is “a conflict-year dataset containing information on communal and organized armed conflict where none of the parties is the government of a state.”⁶⁷⁶⁸ For this dataset, my initial hypothesis was to anticipate seeing larger numbers of terrorist organization activity. While incorrect, the number of incidents are higher than the other previously researched datasets, showing there has been a higher incidence of attacks from non-governmentally affiliated groups.

Figure 0-14 was surprising in the fact that there was no conflict data for India, as found in the previous datasets. The violence in Somalia at the time is attributed to clans attacking other clans, which again, given the high number of different languages spoken in Somalia, supports my linguistic theory. The incidents in Mexico were due to different cartels attacking other cartels.⁶⁹

⁶⁷ Pettersson, Therese, Shawn Davis, Amber Deniz, Garoun Engström, Nanar Hawach, Stina Högladh, Margareta Sollenberg & Magnus Öberg (2021). Organized violence 1989-2020, with a special emphasis on Syria. *Journal of Peace Research* 58(4).

⁶⁸ Lundberg, Ralph, Kristine Eck, and Joakim Kreutz (2012) Introducing the UCDP Non-State Conflict Dataset. *Journal of Peace Research* 49(2).

⁶⁹ Ibid.

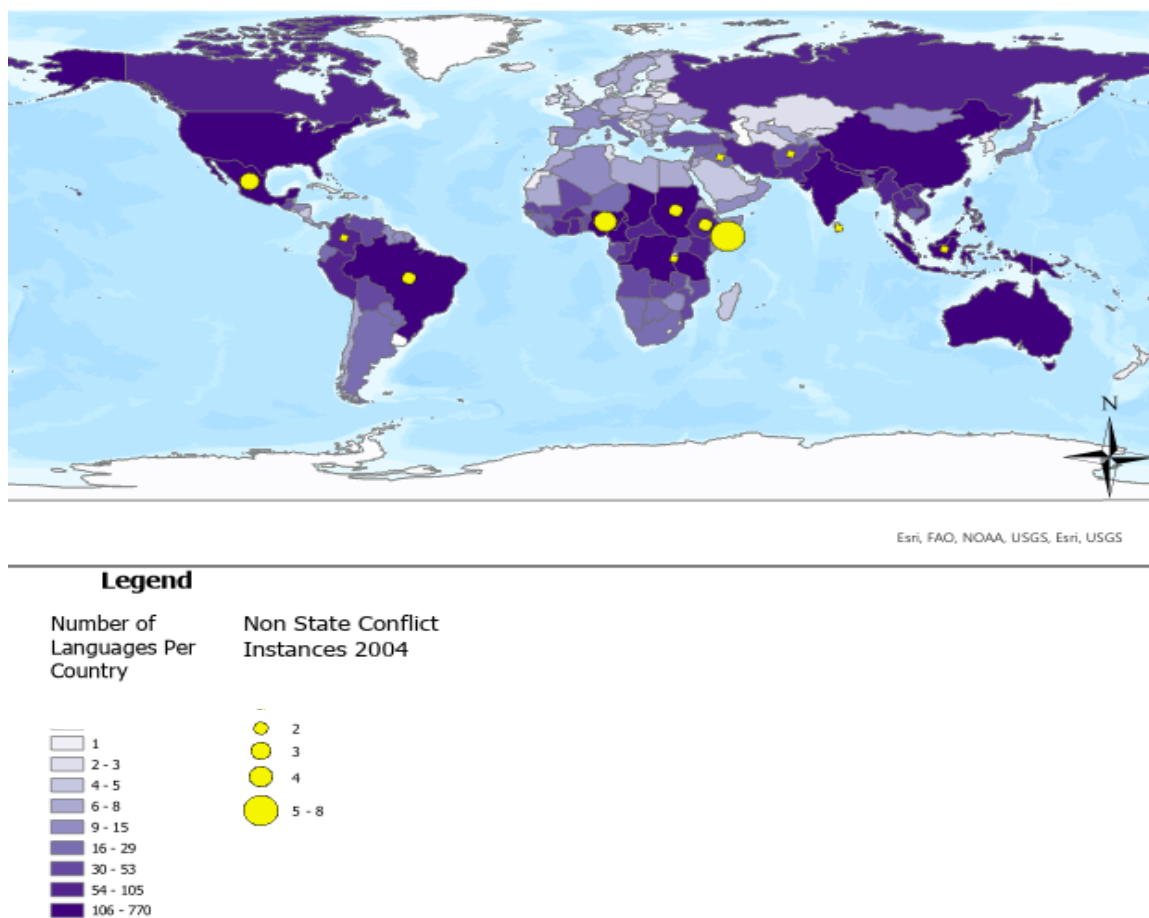


Figure 0-14 Non-State Conflict 2004

In Figure 0-15, the greatest number of conflicts are in the same countries again in 2010. In Mexico, it is still inter-cartel conflict. For Somalia, there is more conflict between the State of Somalia, the mujahideen, and the Islamic party, rather than more of the clan conflict that was recorded in 2004.⁷⁰

⁷⁰ Ibid.

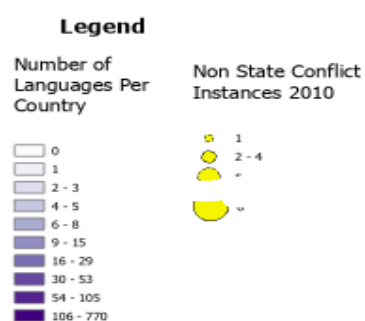
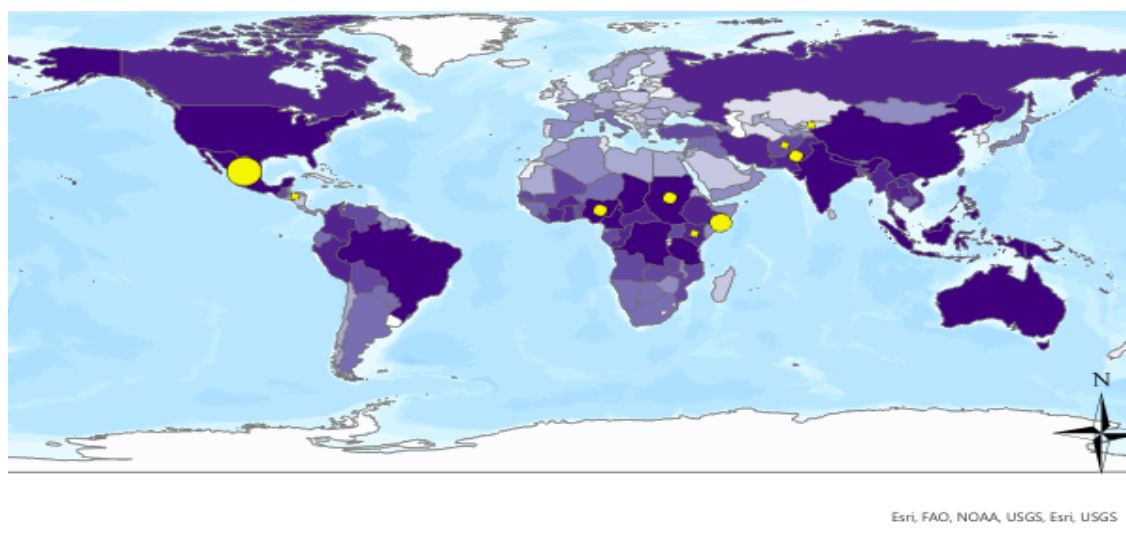


Figure 0-15 Non-State Conflict 2010

Figure 0-16 shows the highest number of conflicts thus far in any dataset. The conflicts in Mexico for this year were the same as in 2004 and 2010, being attributed to the cartel wars. The conflicts in Nigeria were between different ethnic groups and opposing Christians and Muslims. Syria's multiple conflicts are a direct result of the civil war and fighting with IS.⁷¹

⁷¹ Ibid.

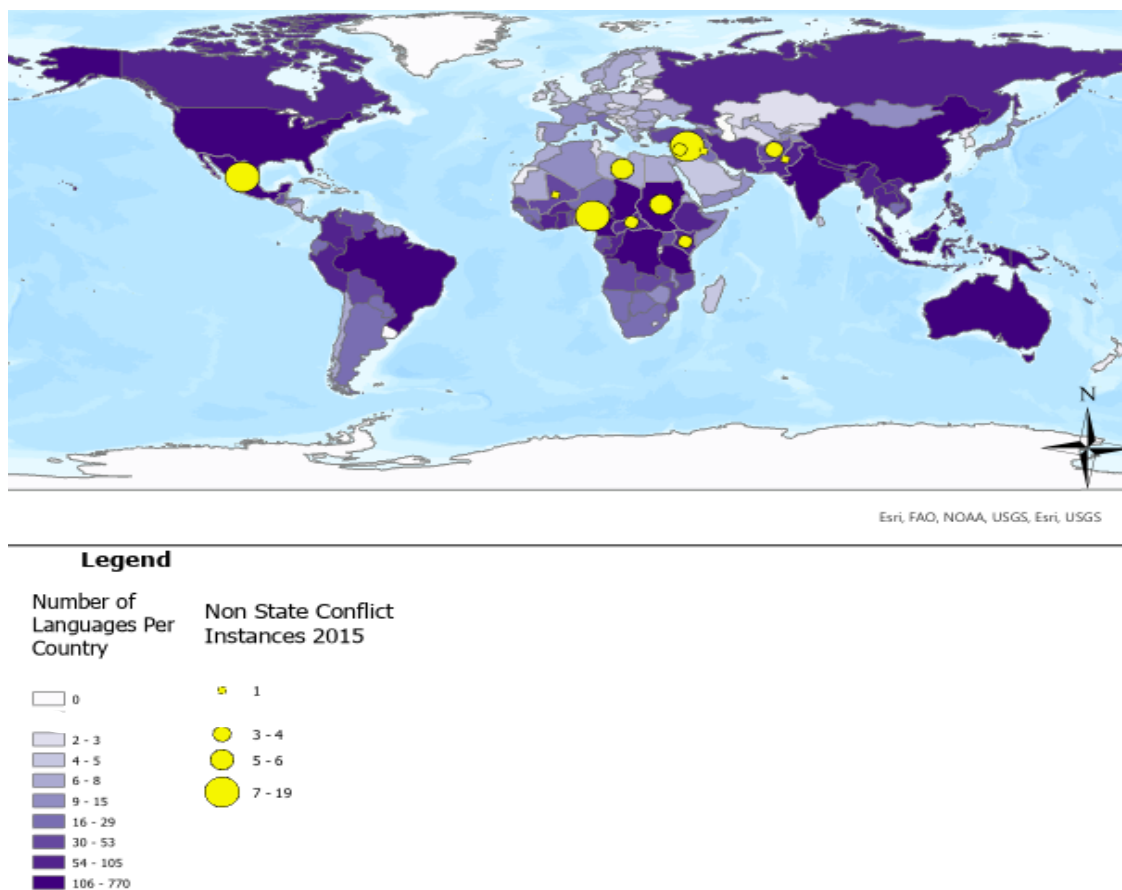


Figure 0-16 Non-State Conflict 2015

Based on the trends for Figures 0-14 – 0-16, Figure 0-17 was not surprising, as the conflicts between the cartels in Mexico continued, as well as for the ethnic groups of Nigeria.⁷²

⁷² Ibid.

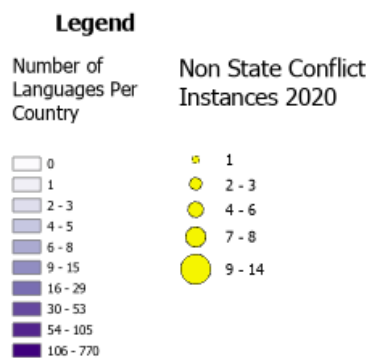
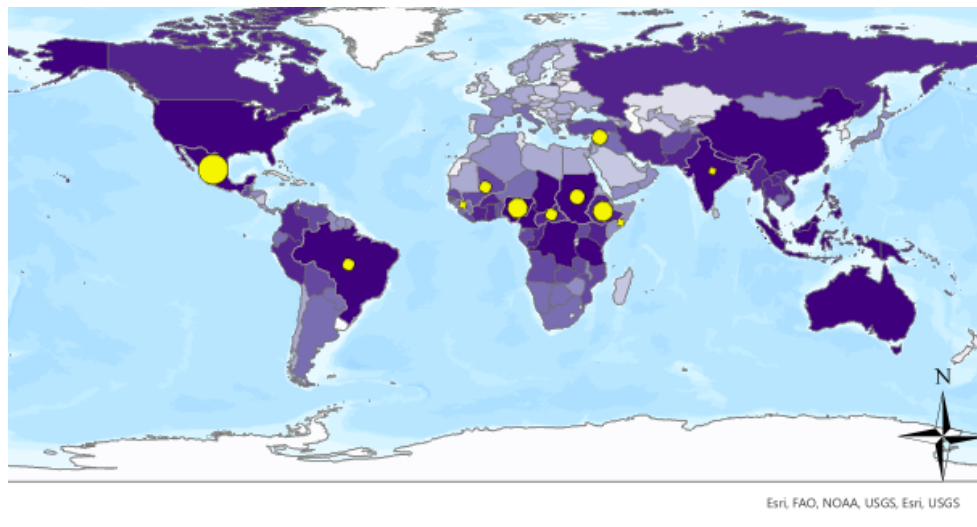


Figure 0-17 Non-State Conflict 2020

Based on this current trajectory, I can predict with some confidence that we will see more instances of conflict within Mexico and Nigeria within the next 5 years, and I expect the release of the 2025 data to confirm this prediction.

UCDP Battle Related Deaths

I was most curious about this dataset, out of any of the datasets reviewed. The “file contains a dyad-year dataset with information on the number of battle-related deaths in the conflicts from 1989-2020 that appear in the UCDP/PRIO Armed Conflict Dataset.”⁷³ I was curious to see where the battles took place and what languages were spoken in the area of conflict.

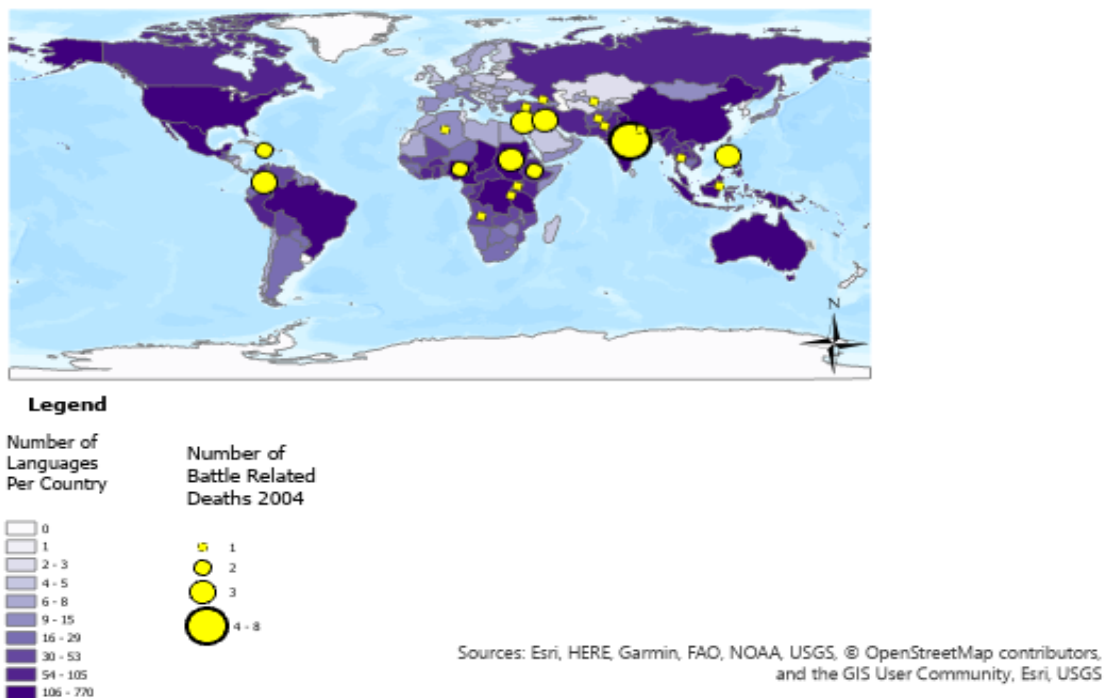


Figure 0-18 Battle-Related Deaths 2004

⁷³ Pettersson, Therese, Shawn Davis, Amber Deniz, Garoun Engström, Nanar Hawach, Stina Högladh, Margareta Sollenberg & Magnus Öberg (2021). Organized violence 1989-2020, with a special emphasis on Syria. *Journal of Peace Research* 58(4).

Knowing much of this data also corresponds to the PRIO dataset, Figure 0-18 did not come as a surprise. India was unsurprising, with battles between the government of India and insurgents in Bodoland, Manipur, Kashmir, Assam, and Tripura.⁷⁴

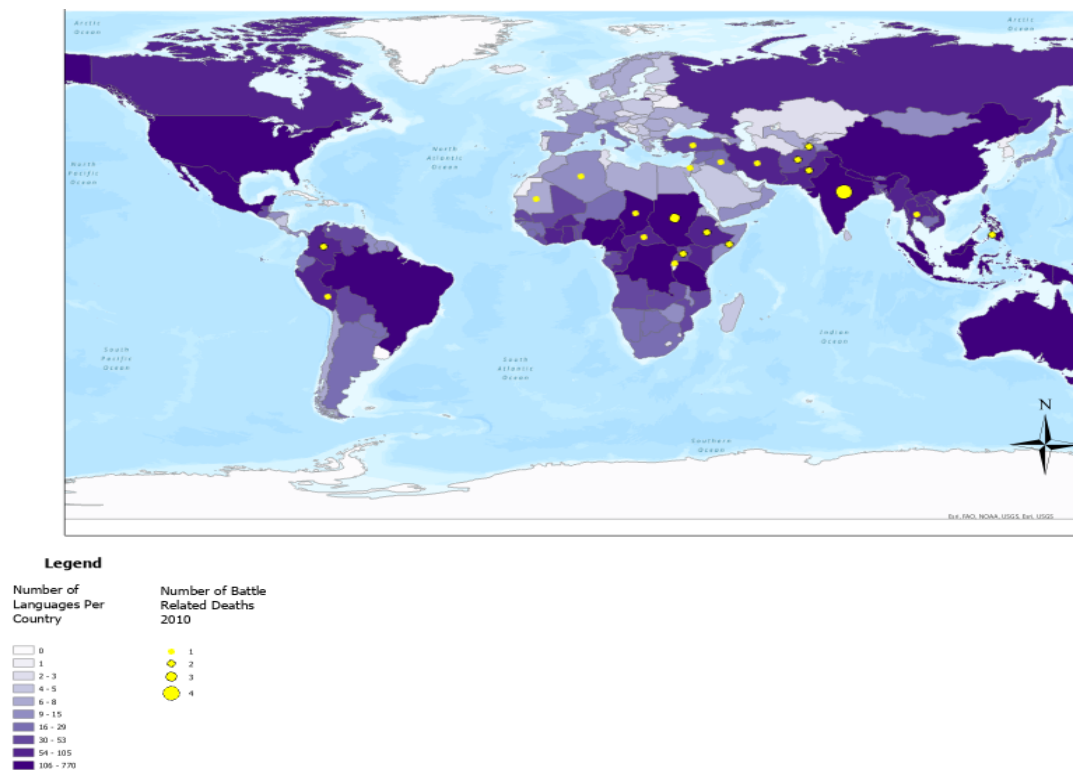


Figure 0-19 Battle Related Deaths 2010

⁷⁴ Ibid.

Figure 0-19 showed relatively minor changes that reflect the PRIO data, although the number of conflicts between the government of India and the insurgent groups declined during this period of time.⁷⁵

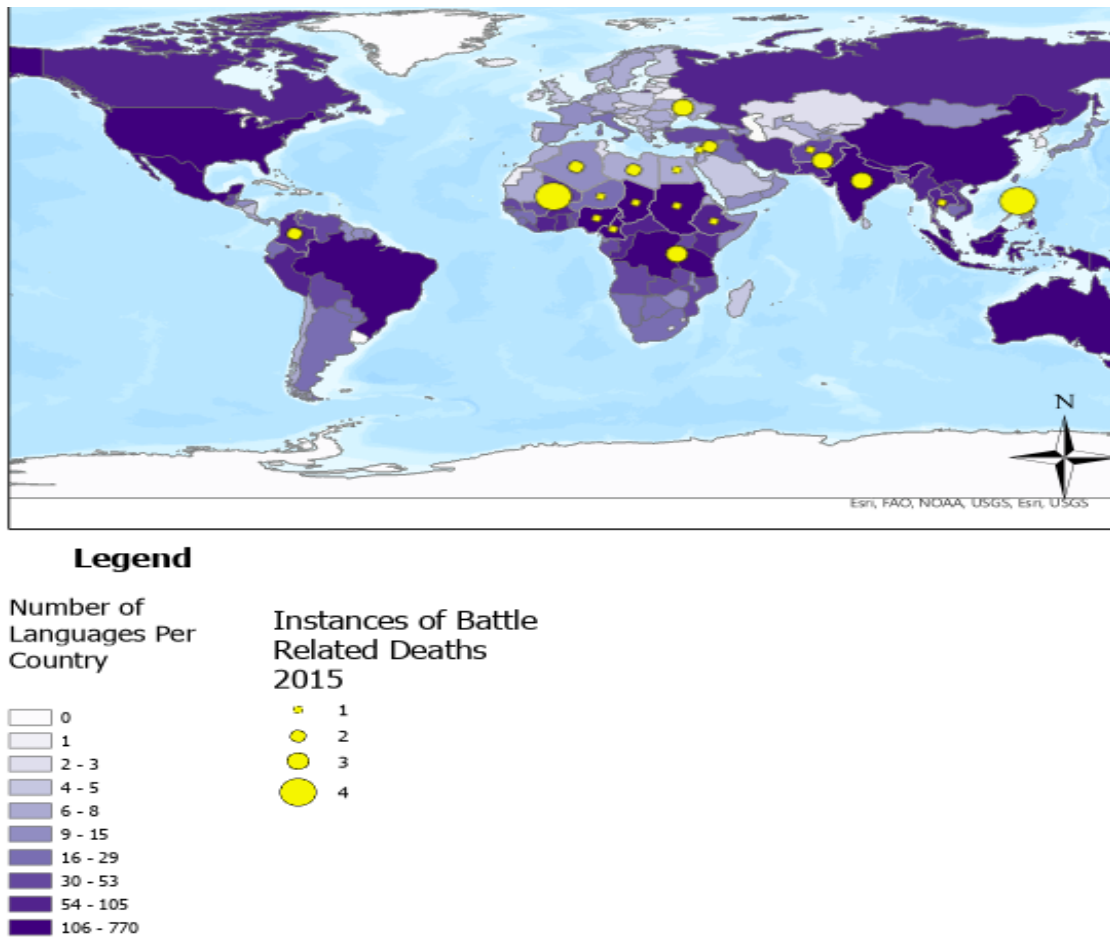


Figure 0-20 Battle Related Deaths 2015

⁷⁵ Ibid.

The year 2015 had a higher number of recorded conflict incidents that occurred globally as seen in Figure 0-20, which also correlated with the PRIO dataset for the countries of Mali, the Philippines, Nigeria, and India.⁷⁶ Figure 0-21 below aligned as well, with no surprising new conflict areas that differed from the PRIO dataset.⁷⁷

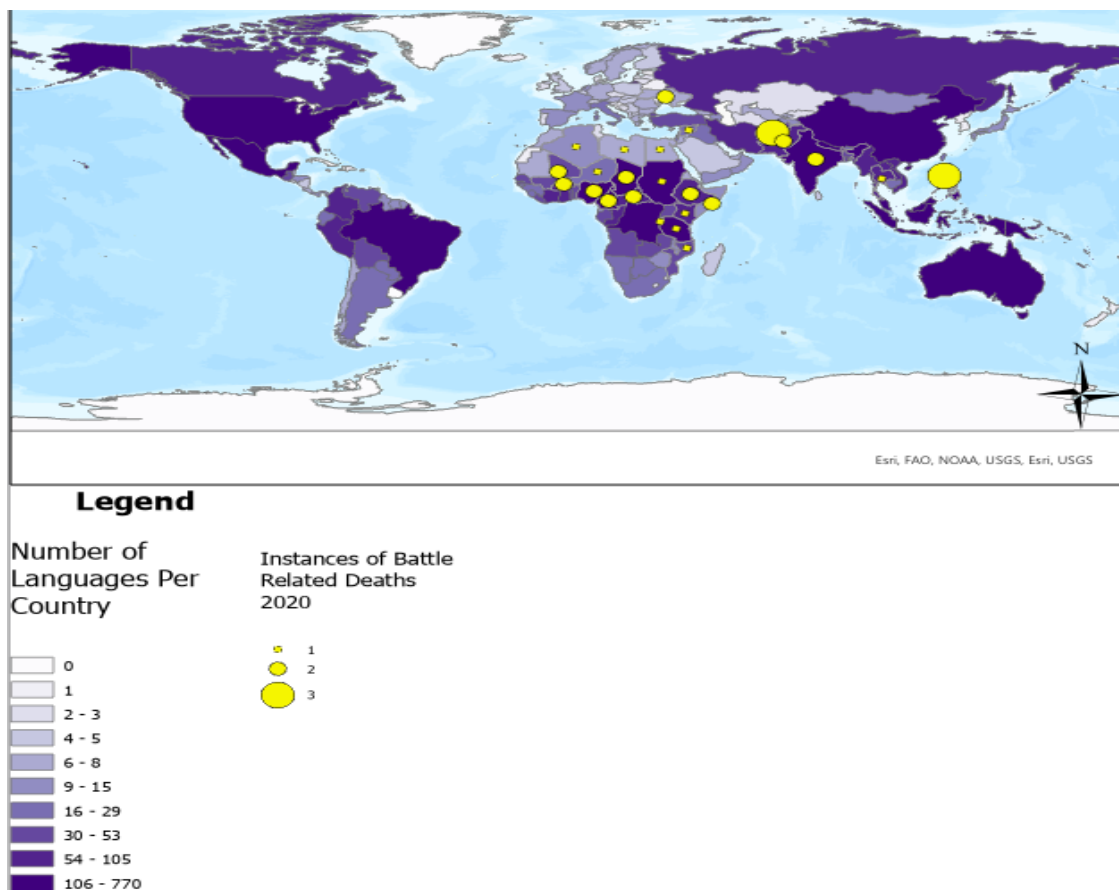


Figure 0-21 Battle Related Deaths 2020

⁷⁶ Ibid.

⁷⁷ Ibid.

If anything, it will be interesting to see what the 2025 data brings in comparison to the 2020 recorded data as seen above. It is notable that there are few battle-related deaths in North or South America in the years mapped. It may be worthwhile to obtain and review data prior to 2004 for these regions.

Deadly Electoral Conflict

In terms of conflict, I was unsure what results this dataset in particular would yield. It is “a global, georeferenced event dataset, based on UCDP data, identifying electoral violence with lethal outcomes from 1989 to 2017.”⁷⁸ As data is only available up to the year 2017, there is missing data for 2020, resulting in only three figures for this dataset.

Figure 0-22 reflected an alarming rate of electoral violence. In India, the violence was often between the Government of India and Kashmir insurgents, as well as Kashmir insurgents and civilians as a retaliatory measure. In Afghanistan, there is an analogous situation with the Government of Afghanistan and the Taliban, with resulting Taliban violence toward civilians. In both instances, the governments of these countries were not aggressors against civilians.⁷⁹

⁷⁸ Fjelde, Hanne and Kristine Höglund (2021) “Introducing the Deadly Electoral Conflict Dataset (DECO)” *Journal of Conflict Resolution*, <https://doi.org/10.1177/00220027211021620>.

⁷⁹ Ibid.

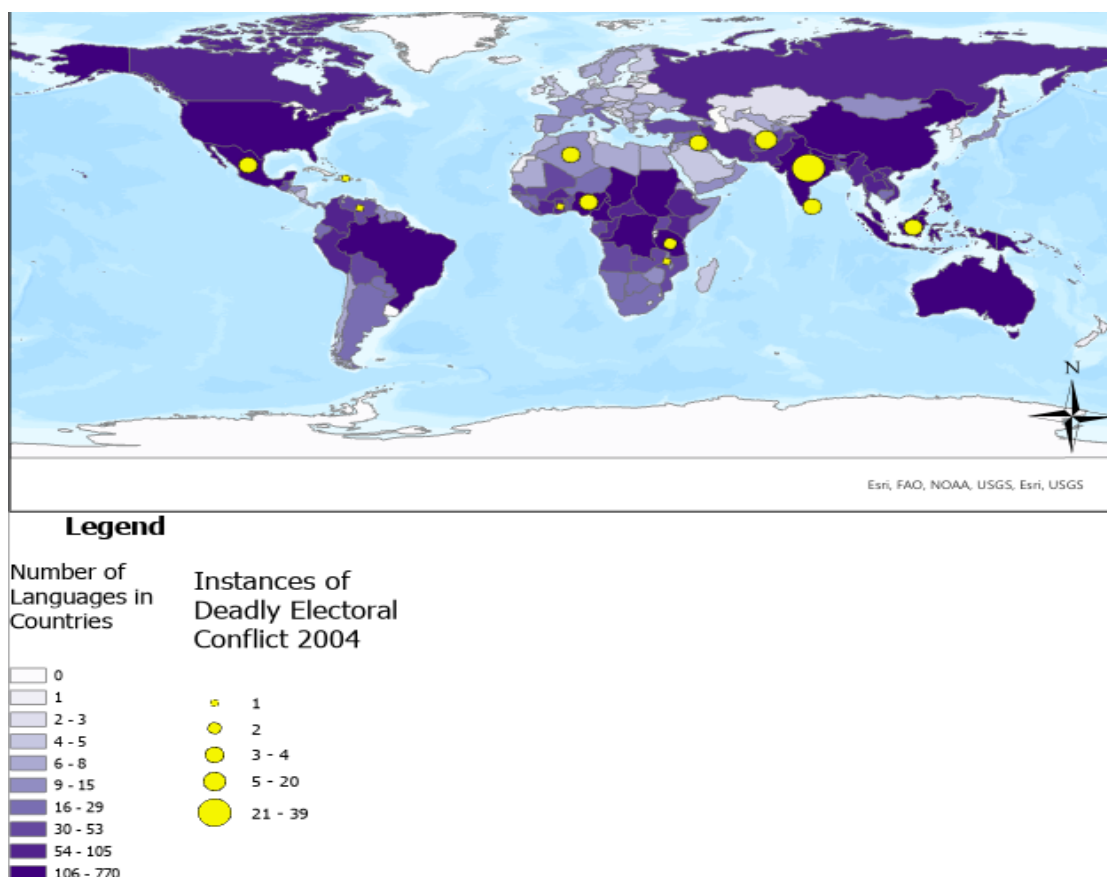


Figure 0-22 Deadly Electoral Conflict 2004

In Figure 0-23, there is an increase in Deadly Electoral Conflict particularly in Afghanistan and the Philippines. This dataset includes headlines for these incidents, and they are all repetitive. In Afghanistan, there were a number of blasts, roadside bombs, rocket attacks, and gunmen who killed civilians and election candidates. For the Philippines, most of the violence was between warring clans, and involving opposition to the election of different candidates in certain villages.⁸⁰

⁸⁰ Ibid.

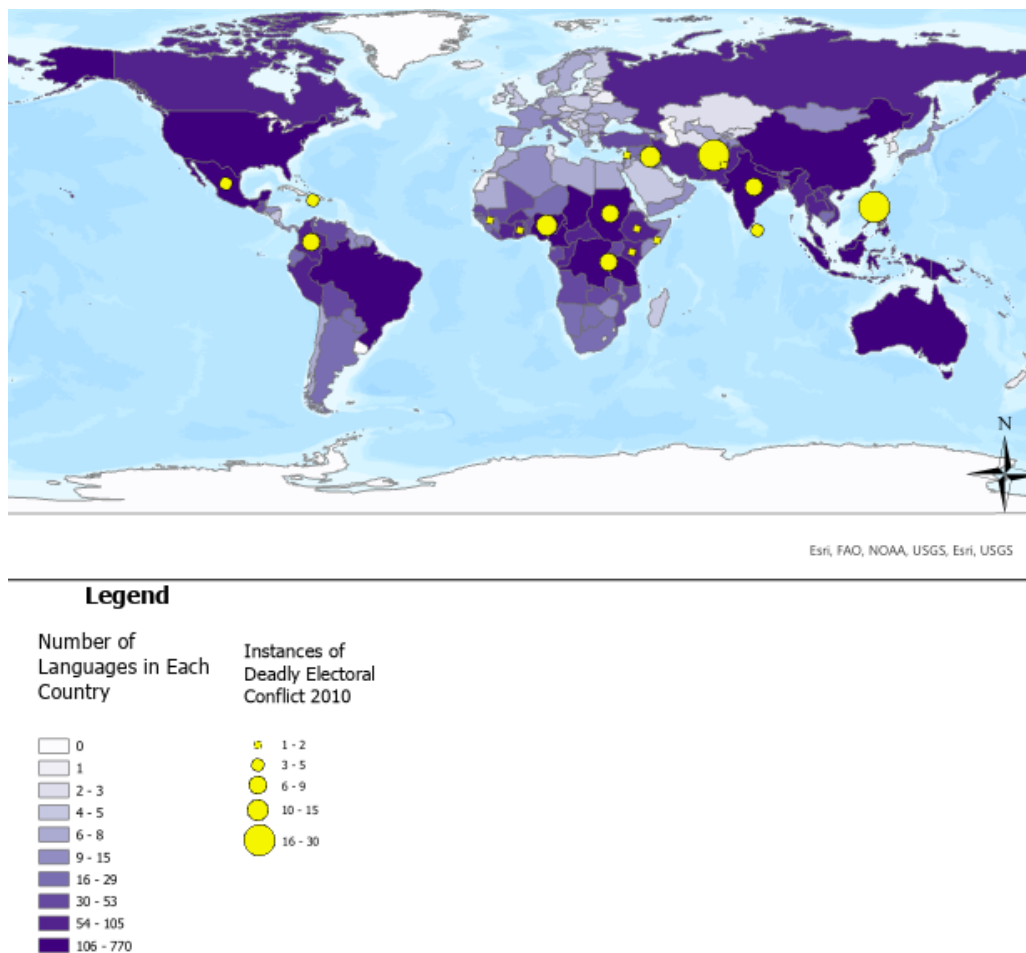


Figure 0-23 Deadly Electoral Conflict 2010

In Figure 0-24 there is a distinct set of Electoral Conflicts that emerge for 2010, which are vastly different than those seen in Figures 0-22 and 0-23. The number of conflicts seen in Bangladesh was due to political violence. Burundi's conflicts were between the Government of Burundi and opponents of Nkurunziza, as well as targeted

killings of civilians in a genocidal sweep. In all instances, there were electoral candidates and civilians both targeted.⁸¹

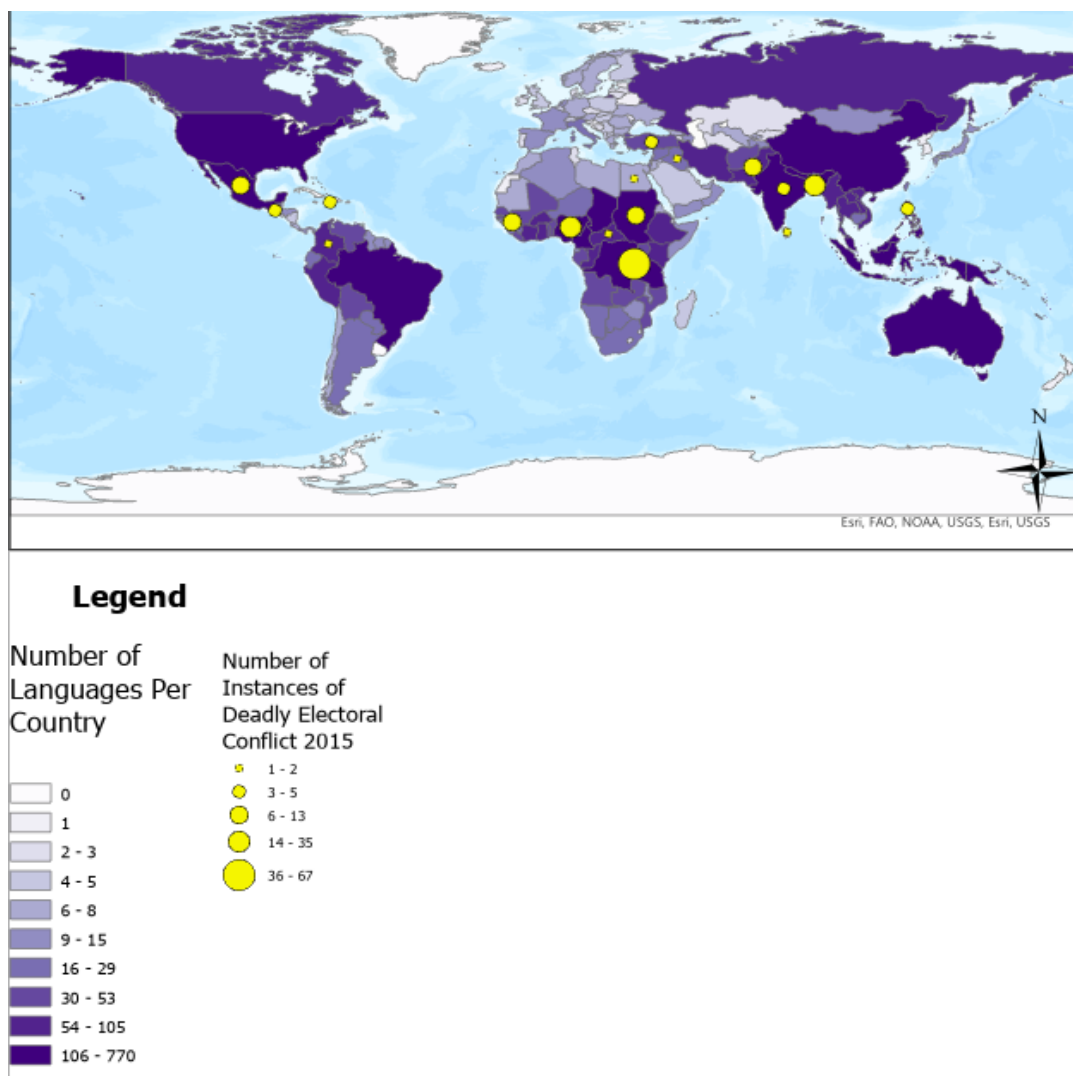


Figure 0-24 Deadly Electoral Conflict 2015

⁸¹ Ibid.

I am interested in reviewing the reports of violence for 2020 when the data is available to see if there is any further trending data, or if electoral conflict receded in 2020. Given the data gleaned from the other datasets, however, I predict that we will see prominent levels of electoral conflict here as well.

Regressions

The mapping and visual analysis of the above datasets was a prelude to the exploratory regressions presented in this section. The ESRI exploratory regression tool was run six times as a test of the bivariate correlations between language diversity and each of the six conflict measures. In other words, I combined the sums for each of the six conflict measures over the four sampled years to create six variations of the dependent variable. In addition to a regression summary, ArcGIS creates a map and shows which countries would have residuals above and below the mean, where the coefficients are used to predict Number of Languages. Five of the six were significant at the $p < 0.05$ level. Non-State Conflict was the exception.

The exploratory regression tool was run with the number of languages set as the independent/explanatory variable and the sum across the six conflict measures as the dependent variables. The results are admittedly modest with only twelve percent of the variance explained, but given the absence of social, economic, political, and environmental variables in the equation, this level of explanation was not surprising. The ArcGIS regression tool helps find the best Ordinary Least Squares (OLS) models by “using all possible combinations for a given list of candidate explanatory variables and assesses which models, if any, pass the necessary OLS checks. By evaluating all possible

combinations of variables, you greatly increase the chances that you will find the best model to solve your problem or answer your question. In addition, the Exploratory Regression tool creates summary reports comparing all passing models. If there are no passing models, it presents information that will help you figure out the problem areas.”⁸² Since spatial factors did not correlate with my response variable and my variables were not spatially autocorrelated as well, what I ran were non-spatial regressions; as such, AdjR2 was determined to be the best statistic to use. Therefore I ignored the Global Moran’s I test for Spatial Autocorrelation (SA), as well as the AICc, JB, and VIF models within the exploratory regression global summaries listed below.⁸³ As previously mentioned, I did not account for economics, religion, or politics as other variables, so in time, some of these other models could help yield meaningful results.

In looking the outputs, one-sided violence (OSV), PRIO armed conflict (PAC), and non-state conflict (NSC) presented statistically significant bivariate correlations with the number of languages. What was interesting to me though was the Koenker (BP) statistic. It reveals that the PAC, NSC, and OSV models, are consistent in both geographic and conflict spaces. The ESRI manual explains:

when the model is consistent in geographic space, the spatial processes represented by the explanatory variables behave the same everywhere in the study area (the processes are stationary). When the model is consistent in data space, the variation in the relationship between predicted values and each explanatory variable does not change with changes in explanatory variable magnitudes (there is no heteroscedasticity in the model.)⁸⁴

⁸² Rosenshein, Lauren , Lauren Scott, and Monica Pratt (2021) “Exploratory Regression A tool for modeling complex phenomena” Esri, <https://www.esri.com/news/arcuser/0111/files/exploratory.pdf>

⁸³ Esri, “Interpreting Exploratory Regression results,” web, [Interpreting Exploratory Regression results—ArcGIS Pro | Documentation](#).

⁸⁴ Esri, “How OLS regression works,” web, <https://pro.arcgis.com/en/pro-app/2.8/tool-reference/spatial-statistics/how-ols-regression-works.htm>.

It is great that K (BP) reported 0.00 for all three datasets, as that indicates there is no heteroskedasticity present, which if there had been, it would mean that the countries possessing more languages would not be as reliable if making predictions with the number of conflicts. The results and data outputs of the regressions are as follows:

One-Sided Violence results:

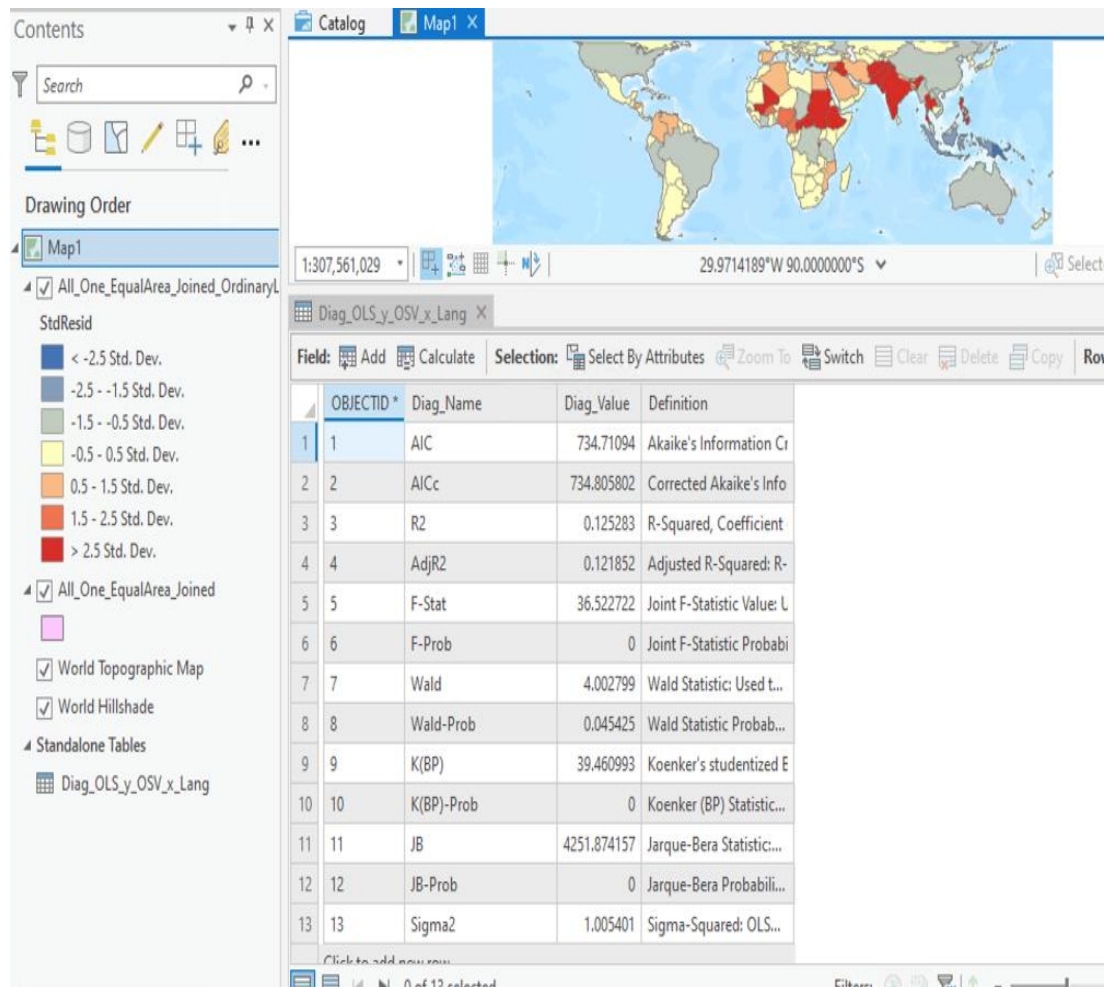


Figure 0-25: One-Sided Violence Regressions

One-sided violence (OSV) had the highest percentage, at 12% of the variance explained with 0.05% of the model variance significant.

```

Choose 1 of 1 Summary
  Highest Adjusted R-Squared Results
AdjR2  AICc  JB K(BP)  VIF  SA  Model
  0.12 734.81 0.00  0.00 1.00 0.00  +JOIN_COUNT**
    Passing Models
AdjR2 AICc JB K(BP) VIF SA  Model

    Passing Models
AdjR2 AICc JB K(BP) VIF SA  Model
*****
***** Exploratory Regression Global Summary (SUM_OSV) *****

          Percentage of Search Criteria Passed
Search Criterion          Cutoff Trials # Passed % Passed
Min Adjusted R-Squared    > 0.50      1         0      0.00
Max Coefficient p-value   < 0.05      1         1     100.00
Max VIF Value             < 7.50      1         1     100.00
Min Jarque-Bera p-value   > 0.10      1         0      0.00
Min Spatial Autocorrelation p-value > 0.10      1         0      0.00

          Table Abbreviations
AdjR2 Adjusted R-Squared
AICc  Akaike's Information Criterion
JB    Jarque-Bera p-value
K(BP) Koenker (BP) Statistic p-value
VIF   Max Variance Inflation Factor
SA    Global Moran's I p-value
Model Variable sign (+/-)
Model Variable significance (* = 0.10; ** = 0.05; *** = 0.01)

```

Figure 0-26: One-Sided Violence Regressions Output

This is interesting, as one would assume that battle related deaths (BRD) would have shown the most significance of all the datasets, not the dataset with attacks on civilians by armed groups and their governments. This will require further investigation and research as well, as it brings to light an interesting question of the difference in violence between interstate and intrastate conflicts and how language may influence the outcome of both of those types of conflicts as well. The fact that intrastate violence here

is showing the most significance is also interesting in the fact that the linguistic diversity of a region could very well be influencing much of the violence seen between different language speaking groups that reside there.

PRIO Armed Conflict Results:

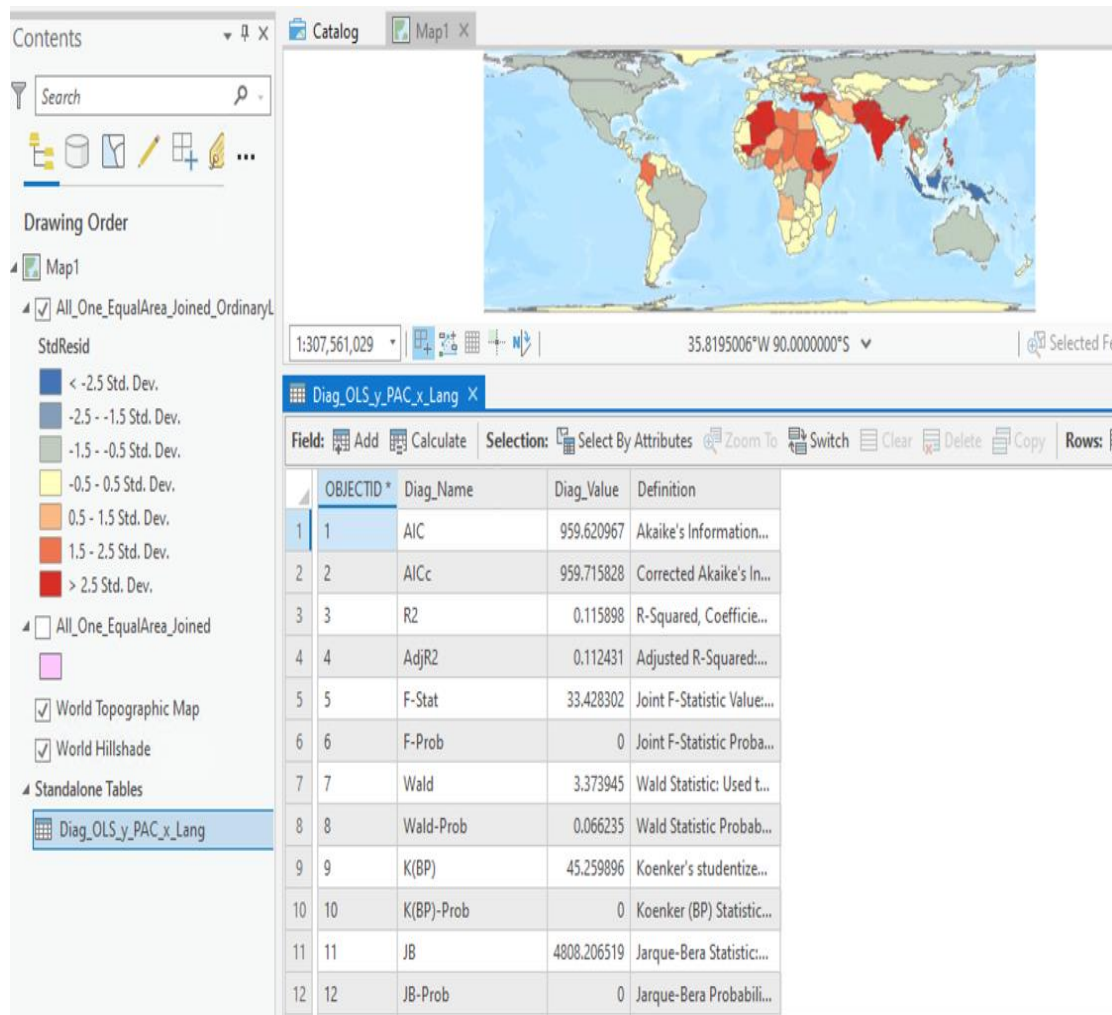


Figure 0-27: PRIO Armed Conflict Regressions

The PRIO armed conflict (PAC) results were just slightly lower than the OSV percentage, with 11% of the variance explained, but 0.10% model significance.

```
Choose 1 of 1 Summary
Highest Adjusted R-Squared Results
AdjR2  AICc  JB  K(BP)  VIF  SA  Model
0.11  959.72  0.00  0.00  1.00  0.00  +JOIN_COUNT*
Passing Models
AdjR2  AICc  JB  K(BP)  VIF  SA  Model
```

***** Exploratory Regression Global Summary (SUM_PAC) *****

Percentage of Search Criteria Passed				
Search Criterion	Cutoff	Trials #	Passed	% Passed
Min Adjusted R-Squared	> 0.50	1	0	0.00
Max Coefficient p-value	< 0.05	1	0	0.00
Max VIF Value	< 7.50	1	1	100.00
Min Jarque-Bera p-value	> 0.10	1	0	0.00
Min Spatial Autocorrelation p-value	> 0.10	1	0	0.00

Table Abbreviations

AdjR2 Adjusted R-Squared
AICc Akaike's Information Criterion
JB Jarque-Bera p-value
K(BP) Koenker (BP) Statistic p-value
VIF Max Variance Inflation Factor
SA Global Moran's I p-value
Model Variable sign (+/-)
Model Variable significance (* = 0.10; ** = 0.05; *** = 0.01)

Figure 0-28: PRIO Armed Conflict Regressions Output

Again, we see an armed conflict with one of the actors being the government of a state, but with conflicts occurring at both the interstate and intrastate level. This was more to be expected, as it demonstrates that conflicts occur with internal and external states, and language diversity can be seen within both types of conflicts. A comparative analysis over the span of more conflict years of this dataset might help researchers discover if one or both interstate/intrastate conflicts are affected more by language diversity. Based on that data then, counts of interstate vs. intrastate conflicts that have been caused because of linguistic diversity could be compared to see which type of conflict occurred more often than the other.

Non-State Conflict Results:

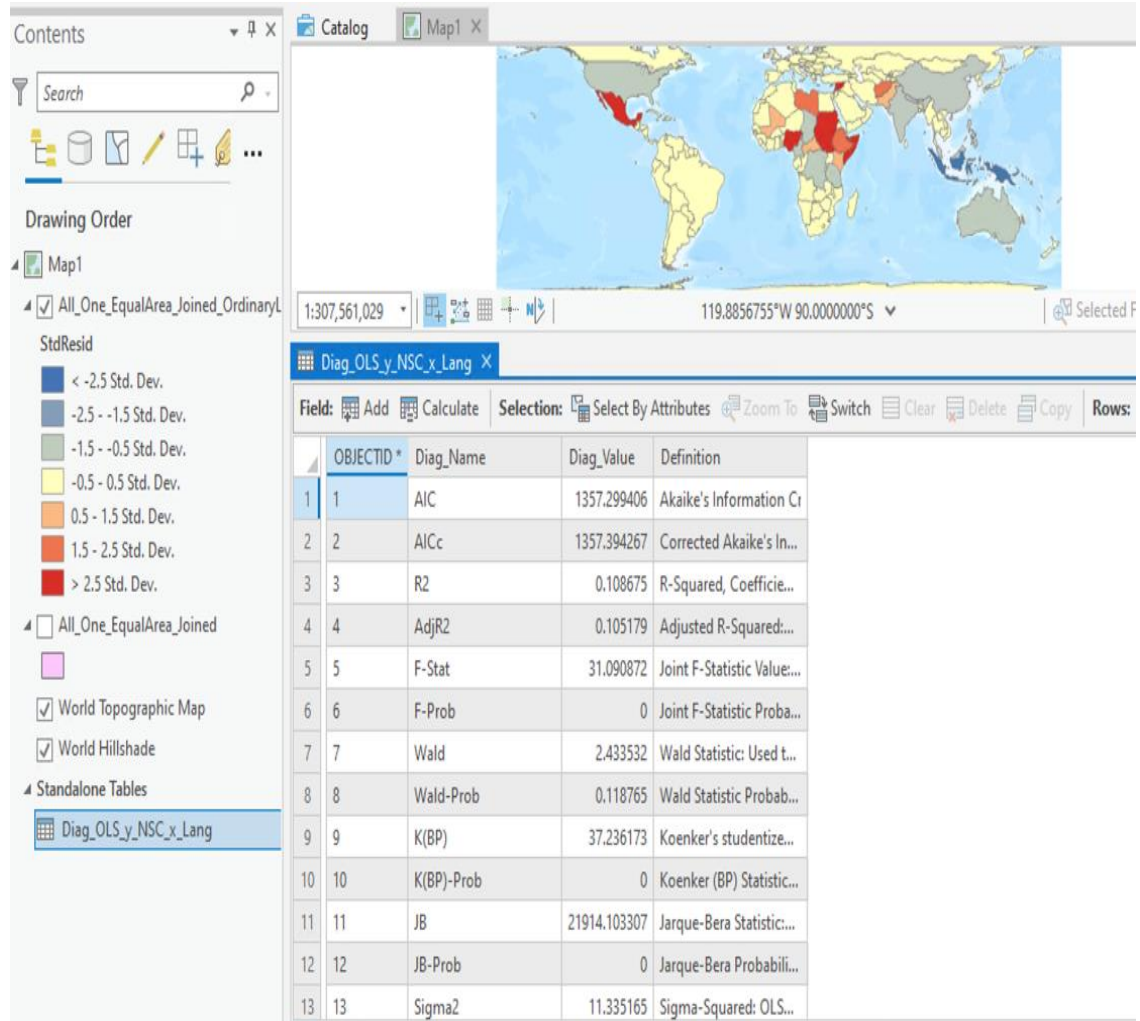


Figure 0-29: Non-State Conflict Regressions

Non-State Conflicts (NSC) yielded the same percentage result as PRIO, with 11% of the variance explained, but as previously stated, had 0 model variance significance.

```

Choose 1 of 1 Summary
Highest Adjusted R-Squared Results
AdjR2    AICc    JB K(BP) VIF    SA    Model
0.11 1357.39 0.00 0.00 1.00 0.59 +JOIN_COUNT
Passing Models
AdjR2 AICc JB K(BP) VIF SA Model

***** Exploratory Regression Global Summary (SUM_NSC) *****

Percentage of Search Criteria Passed
Search Criterion          Cutoff Trials # Passed % Passed
Min Adjusted R-Squared    > 0.50      1      0      0.00
Max Coefficient p-value   < 0.05      1      0      0.00
Max VIF Value             < 7.50      1      1     100.00
Min Jarque-Bera p-value   > 0.10      1      0      0.00
Min Spatial Autocorrelation p-value > 0.10      1      1     100.00

Table Abbreviations
AdjR2 Adjusted R-Squared
AICc Akaike's Information Criterion
JB Jarque-Bera p-value
K(BP) Koenker (BP) Statistic p-value
VIF Max Variance Inflation Factor
SA Global Moran's I p-value
Model Variable sign (+/-)
Model Variable significance (* = 0.10; ** = 0.05; *** = 0.01)

```

Figure 0-30: Non-State Conflict Regressions Output

NSC's being the third dataset to yield correlation results was a bit of a surprise, as I had fully expected to see correlation data more from datasets with the government of a state being one of the main actors in a conflict. This also brings to light the need for more study into conflicts between clans, cartels, ethnic and religious groups. This would be the initial dataset that I would introduce these other types of variables into and see what types of correlative data would then be generated.

Future Regression Results

As there was only one outlier previously mentioned in the visual analysis section, that being Haiti, I did not notice any impact on the regression results. I am most interested to see if other outliers may eventually affect the outcomes of future regressions as I merge and run the sums of more conflict years through the exploratory regression tool. Will this skew my data, or will the results mainly be the same because there will be too few outliers? At this time, I cannot with certainty give any predictions on this, other than the data for 2025 may reveal more corruption than we realized took place due to the pandemic.

Chapter VI.

Conclusion

As evidenced in the maps above, my visual analyses, and the results of the exploratory regressions, I have established a both a visual relationship and a correlation between language diversity and conflict, which supports, albeit weakly, my hypothesis that the number of languages that are spoken within a country can contribute to the incidence and/or intensity of conflict. There were no instances of conflict in countries whose people speak ten languages or fewer within the chosen timeframe, other than Haiti being the outlier. This overall absence also supports my hypothesis that countries that speak more than one language are more likely to experience conflict.

In some instances, seeing the conflicts on the maps can potentially help predict future conflicts, as was seen with Afghanistan. In continuing my work, my first step will be to run a comparative analysis of all data between 2000-2020. Because I know there were changes in the several year spans that I had bypassed in this analysis, I am especially interested to see the yearly changes on the map, and to join all of the yearly data together to see if my exploratory regression findings are altered. I am also keen in mapping other conflicts from previous decades as well and seeing if my results may perhaps shed new light on past historic events.

I also would like to take a more micro approach and map where different insurgent groups and terrorist organizations originate from within countries, to compare with the location of the civilians and governments that they attack. Such mapping could determine if a common language is spoken within that particular region of the country. This would further support my findings.

These exploratory findings further support the importance of foreign language education and linguistic learning in all levels of education. Ultimately, these findings show that there is another element – linguistic diversity – that is important to consider when studying conflict, and that language can be an influence at the root of conflicts.

References

- Anderson, Christopher & Aida Paskeviciute (2006) How Ethnic and Linguistic Heterogeneity Influence the Prospects for Civil Society: A Comparative Study of Citizenship Behavior. *The Journal of Politics*, Vol. 68, No. 4, November 2006, 787.
- Bormann, Nils-Christian, Lars-Erik Cederman, and Manuel Vogt (2017) Language, Religion, and Ethnic Civil War. *Journal of Conflict Resolution*, 744-745.
- Darquennes, Jeroen. "Language conflict research: A state of the art." *International Journal of the Sociology of Language*, (2015) 235: 7-32.
- Duffy Toft, Monica (2002) Indivisible territory, geographic concentration, and ethnic war. *Security Studies*, 12:2, 87.
- Eck, Kristine & Lisa Hultman (2007) Violence Against Civilians in War. *Journal of Peace Research* 44(2).
- Esri. www.esri.com.
- Esri. "Interpreting Exploratory Regression results," web, [Interpreting Exploratory Regression results—ArcGIS Pro | Documentation](#).
- Esri. "How OLS regression works," web, <https://pro.arcgis.com/en/pro-app/2.8/tool-reference/spatial-statistics/how-ols-regression-works.htm>.
- Gleditsch, Nils Petter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg, and Håvard Strand (2002) *Armed Conflict 1946-2001: A New Dataset*. *Journal of Peace Research* 39(5).

- Fjelde, Hanne and Kristine Höglund (2021) “Introducing the Deadly Electoral Conflict Dataset (DECO)” *Journal of Conflict Resolution*, <https://doi.org/10.1177/00220027211021620>.
- Glottolog. “Languages.” Languages. 2021. <https://glottolog.org/glottolog/language>.
- Harbom, Lotta, Erik Melander & Peter Wallensteen (2008) Dyadic Dimensions of Armed Conflict, 1946-2007. *Journal of Peace Research* 45(5).
- Haynie, Hannah J. (2014) Geography and Spatial Analysis in Historical Linguistics. *Language and Linguistics Compass*, 344.
- Iberhard, David M., Gary F. Simons, and Charles D. Fenning (eds). *Ethnologue: Language of the World*. Dallas, Texas: SIL International. 2004.
- King, Charles King (1997) Policing Language: Linguistic Security and the Sources of Ethnic Conflict. *Security Dialogue*, Vol. 28(4), 494.
- Kisembe, Denis. “Language, culture and conflict resolution.” Research Paper, 2018. <https://www.grin.com/document/433194>
- Laitin, David D. Laitin (2000) Language conflict and violence: the straw that strengthens the camel’s back. *Arch. europ.sociol.*, XLI, 98.
- Lake, Nell. “Language Wars.” *Harvard Magazine*, March-April 2002.
- Mabry, Tristan James (2010) Language and Conflict. *International Political Science Review*, 190.
- Luebbering, Candice R., Korine N. Kolivras & Stephen P. Prisley Visualizing Linguistic Diversity Through Cartography and GIS. *The Professional Geographer*, 65:4, 581.

- Nelde, Peter H. "Language contact means language conflict." *Journal of Multilingual and Multicultural Development*, (1987) 8:1-2, 33-42.
- Newmeyer, Frederick J. "The History of Modern Linguistics," *Linguistic Society of America*, 2022, <https://www.linguisticsociety.org/resource/history-modern-linguistics>
- Medeiros, Mike. "The Language of Conflict" *Ethnicities*, (October 2017), Vol.17, No.5, pp. 627-645.
- Mpofu, Philip & Davie E. Mutasa (2014) Language policy, linguistic hegemony and exclusion in the Zimbabwean print and broadcasting media. *South African Journal of African Languages*, 34:2, 226.
- Pettersson, Therese, Shawn Davis, Amber Deniz, Garoun Engström, Nanar Hawach, Stina Högladh, Margareta Sollenberg & Magnus Öberg (2021). Organized violence 1989-2020, with a special emphasis on Syria. *Journal of Peace Research* 58(4).
- Rosenshein, Lauren , Lauren Scott, and Monica Pratt. "Exploratory Regression A tool for modeling complex phenomena." Esri, 2021.
<https://www.esri.com/news/arcuser/0111/files/exploratory.pdf>.
- Shell, Marc. "Language Wars." *CR: The New Centennial Review* 1, no.2 (2001): 1-17.
- Smirnova, Anastasia and Rumen Iliev (2017) Political and Linguistic Identities in an Ethnic Conflict. *Journal of Language and Social Psychology*, Vol. 36(2) 211-225.
- South Asia Terrorism Portal. "National Democratic Front of Bodoland." *South Asia Terrorism Portal*, 2001.

https://www.satp.org/satporgtp/countries/india/states/assam/terrorist_outfits/ndfb.htm.

Sundberg, Ralph, Kristine Eck, and Joakim Kreutz (2012) Introducing the UCDP Non-State Conflict Dataset. *Journal of Peace Research* 49(2).

Transparency International. “Our Work In: Haiti,” web, [Haiti - Transparency.org](https://www.transparency.org/en/our-work-in-haiti).

Uppsala Conflict Data Program. “UCDP.” Department of Peace and Conflict Research. Accessed April 30, 2021. <https://ucdp.uu.se/>.

Uppsala University. “UCDP Definitions.” Definitions. Accessed April 30, 2021. https://www.pcr.uu.se/research/ucdp/definitions/#tocjump_8829967763543327_2

Vogt, Manuel (2018) Ethnic Stratification and the Equilibrium of Inequality: Ethnic Conflict in Postcolonial States. *International Organization* 72, 133.

Weidmann, Nils B (2009) Geography as Motivation and Opportunity. *Journal of Conflict Resolution*, Volume 53 Number 4, 526-541.

Wucherpfennig, Julian, Weidmann, Nils B., Girardin, Luc, Cederman, Lars-Erik, and Wimmer, Andreas “Politically Relevant Ethnic Groups across Space and Time: Introducing the GeoEPR Dataset,” *Conflict Management and Peace Science* Vol 28(5): 423–437.