Death, Destruction, and Data

An Intro to Using Data to Study International Conflict

Miles D. Williams

2026-01-18

Table of contents

Preface

My name is Miles Williams, and I authored this free and open access book for students taking a 300 level undergraduate class at Denison University called "Death, Destruction, and Data."

This provocatively named course is an introduction to using data to study international conflict. This is a specialized class that assumes some prior knowledge of the R programming language and, ideally, some experience with the `{tidyverse}` dialect. If you don't have prior experience with either, that's not the end of the world. I'll go slowly, and because this class is focused on studying one issue (international conflict) that also means that much of the material will center on conceptual issues rather than an emphasis on technical know-how (though there will be some of that, so don't rest too easy, and I will expect you to do some original data analysis and writing along the way.).

The book is divided into two parts, which I've summarized briefly below.

Part I: Trends in conflict

Part I is all about studying historical trends in international conflict. I begin with an introduction to conflict (including war) and how international relations scholars define armed conflict between countries.

I then talk about how conflict is measured using data, and show how to use this data to answer a pair of important questions: is war becoming more or less common, and is it becoming more or less deadly?

Both questions are foundational to the quantitative study of war, but they have faded in relevance as conflict scholars turned their attention to dyadic (country-pair) analyses of factors that explain why some countries tend to fight while others don't. These questions have witnessed some renewed interest, however, due to recent debates about the "decline of war theory."

On the technical side, we'll spend a lot of time thinking about how conflict is measured, how to visualize trends in conflict onset and deadliness, and detect whether changes in each have taken place.

Part II: Who fights whom and why?

Part II moves the focus to country dyads with the goal of answering the question: why do certain countries fight each other more than others?

The literature on why countries fight is much too broad to cover in a short book about using data to study conflict, but I will draw on a sample of prominent theories of conflict and peace to motivate a variety of analysis examples and expose you to some of the ins and outs of constructing datasets for studying international conflict. I'll also introduce you to the standard set of statistical tools conflict scholars use in their research (so get ready to learn about logit models!).

My hope

I will not be able to address every single technical issue or measurement question that comes along with conflict research. I do hope, however, that these instructional materials will give you a solid enough foundation to start asking your own questions, testing novel hypotheses, and conducting empirical research. This is the start of your journey rather than the end (assuming you choose to keep going long after this course wraps up).

1 What is War?

Main ideas:

- A useful way to think of war is as a continuation of politics by other means (but what does that mean, exactly?).
- Peace science is a research tradition in the field of international relations that seeks to use data and the scientific method to study war and peace.
- War, and international conflict more broadly, is treated by peace scientists as a specific kind of political event that can be measured and, about which, systematic comparisons can be drawn.
- The Correlates of War Project, started in the 1960s, had (and continues to have) outsized influence over how international conflict is defined and empirically operationalized.
- A recent controversy surrounding one of the most popular COW datasets has called into question the reliability of the data, and the outgrowth of this controversy is a brand new dataset that we'll use in the rest of this class.

1.1 Politics by Other Means

It's a bit foolhardy of me to think I can address the question "what is war?" in a single chapter of a set of open source lecture notes. If this were my goal, I'd have to affix a big asterisk to the chapter with the ubiquitous phrase spoken in nearly every *Star Wars* movie ever produced: "I've got a bad feeling about this."

War has been part of the human experience since well before recorded history. For thousands of years, countless thinkers have opined on the nature of war. One of the more recent takes (by historical standards) that you may have heard of is the idea that "war is a continuation of politics by other means." This idea comes from the famed Prussian General, Carl von Clausewitz (2003); though this sentence is sometimes is taken out of context, and he didn't actually say this exact phrase. Actually, he didn't write in English, so that also adds some layers of obscurity between what he actually said and what he meant.

Nonetheless, one thing that is true, and clear enough from Clausewitz's writing is that war is *political*. At least, this is true of the most common type of war. As the political scientist Bear Braumoeller (2013) nicely summarizes, Clausewitz drew a distinction between two kinds of war: *absolute* war and *real* war.

Absolute war is rare, and considered to be a kind of duel to the death. It also isn't the type of war Clausewitz considered an extension of politics. When he said this, he meant *real* war. So what's the distinction?

As Braumoeller (2013) further summarizes, real war is a form of political communication meant to gather information necessary to assess the hypothetical outcome of absolute war. Another, more intuitive way to put it, is that real war is about sizing up your opponent. By engaging in brief skirmishes, sides of a conflict learn information about the other side (its strengths, weaknesses, resolve, and so on). This information is necessary so that both sides can agree to the terms of a bargain.

In short, war is intimately connected with bargaining between countries. Countries have some issue, whether that's territory, trade, or resources, that they have a dispute over. As they try to bargain with each other to find a compromise that satisfies them both, there's some uncertainty about how much each side should actually ask for and deserves to get. Real war can serve as a means by which sides get this information. *Ergo*, war as a continuation of politics by other means.

The idea that war is related to bargaining would later be famously formalized by the political scientist James Fearon (1995), who developed a game theoretical model to help him show why countries would choose to go to war when negotiating a deal would be mutually better for both sides. A few years later, R. Harrison Wagner (2000) would extend Fearon's formalized logic of bargaining to include war as not just the result of bargaining failure, but also a costly means by which sides negotiate a deal.

In some ways, these game theoretic approaches nicely encapsulated the ideas Clausewitz proposed long ago. Braumoeller (2013) puts it most succinctly: "information, not attrition, is the goal of conflict" (9).

Now, if you're at all tuned in to current events, you probably have some doubts about this way of understanding war, and rightly so. You can make a good case that Israel's recent war with Hamas went far beyond communicating information. It seemed an awful lot like attrition. It also demonstrated that war can be a means of decisively changing the balance of power by weakening an opponent, which happened, by extension, with Israeli strikes against Hezbollah in Lebanon and against the Iranian regime.

These kinds of conflicts, however, align more with Clausewitz's definition of absolute war, which he considered rare, as these kinds of conflicts in fact are. But they do happen, and most observers agree that when you initiate a conflict of the continuation-of-politics variety you risk escalation to the duel-to-the-death sort. This is why German Chancellor Theobald von Bethmann-Hollweg referred to the start of war as *rolling the iron dice* on the eve of World War I. Actually, what he is supposed to have said exactly is: "If the iron dice must roll, may God help us."

Iron dice seems an apt analogy, because much to everyone's surprise at the time, the conflict that became known as World War I was predicted by nearly all Great Power elites at the time

to be a short conflict that would be over before Christmas (Tuchman 1994). The lesson of World War I is that wars of information can escalate to wars of attrition, even when those involved had no hopes or intentions of making that happen.

War, in short, is more than just a costly form of bargaining; it comes with existential risks. The choice to roll the iron dice is political, but what happens next can quickly fall beyond anyone's ability to control.

1.2 The Peace Science Tradition

Because war is a form of politics, and of a sort that has fatal implications, it's natural that a group of political scientists would be motivated to carve out a niche area of research focused on trying to study war in a systematic way. This area of research became known as "peace science."

Even up to the mid-20th century, the study of war was mostly the purview of historians, whose analyses generally focused on particular wars or periods in time. The focus was on close examination of historical documents and other source materials to construct a narrative about why a war happened, why one side won or lost, or the special role that particular individuals played.

These are all important things to document, and the field of peace science would not have been able to get started without the painstaking work of historians to document past conflicts.

However, peace scientists wanted to go further. They wanted to measure wars across time and space in a consistent and rigorous way so that they could then use statistical analysis and the scientific method to systematically test arguments about war and peace. They wanted to come up with theories, generate testable hypotheses, and empirically verify them against observable regularities in conflict.

The political scientist J. David Singer and the historian Melvin Small were among those at the forefront of the peace science tradition. Singer founded the Correlates of War Project in 1963, which remains alive and well today, and with the help of Small they began work to collect data on all wars that took place since 1816 (the post-Napoleonic-Wars period).

On the Correlates of War Project website (most people just refer to it as COW, and pronounce it like "cow"), it says that "[t]he original and continuing goal of the project has been the systematic accumulation of scientific knowledge about war." This project rests on the assumption that international conflicts can be *compared* across time and space. To make this possible, Singer and Small, as the COW website continues, "needed to operationally resolve a number of difficult issues such as what is a 'state' and what precisely is a 'war.' " Based on their work, Singer and Small published a book in 1972 called *The Wages of War*, which would establish a standard definition of war that would guide the work of countless future peace scientists.

As a result of their efforts, decades of research looking at trends in the data and assessing how various factors predict conflict onset have supported the accumulation of knowledge about war. One of the most comprehensive, yet succinct, resources you can find that summarizes all this accumulated knowledge is the edited volume *What Do We Know About War?*, which is now in its third iteration (Mitchell and Vasquez 2024). While there is still a lot that peace science has failed to figure out about war, the field knows a lot more now than they did in 1963 when Singer and Small began their work.

1.3 Defining War and Other forms of Conflict

So how do peace scientists define war? How do they define other instances of conflict between countries as well?

Singer and Small considered war an instance of "sustained combat involving substantial fatalities," according to documentation for the latest COW wars dataset. More precisely, they recorded that a war occurred if more than 1,000 battle deaths were documented, combat was sustained, and it involved organized armed forces.

They would go on to develop a typology of war based on the participants: (1) interstate war, (2) civil war, (3) extra-systemic war.

Interstate war involves wars between and among sovereign countries considered part of the international state system (we'll focus on interstate conflict in this class). Civil war involves conflicts between the government of a state and an armed group within its borders. Extra-systemic war involves conflict between a state and a separate actor not currently recognized as a state, such as a colony.

COW maintains datasets for each of these kinds of wars on their website.

There are other prominent datasets on international conflict as well. Probably the most popular is the Militarized Interstate Dispute (MID, pronounced "mid") dataset, which is also maintained by COW. MIDs are distinct from wars but are often used to study the potential for war, or warlikeness, of countries.

MIDs are a comparatively more recent innovation, and the intended goal of measuring MIDs is to capture more variation in international conflict than is possible using the more stringent criteria used to measure war. The concept and original dataset were proposed by Gochman and Maoz (1984), and the data would go through several rounds of updates and revisions. The most recent version of the dataset (version 5) just dropped in 2020 (Palmer et al. 2022).

COW notes on their webpage that MIDs:

> '...are united historical cases of conflict in which the threat, display or use of military force short of war by one member state is explicitly directed towards the government, official representatives, official forces, property, or territory of another state.

> Disputes are composed of incidents that range in intensity from threats to use force to actual combat short of war' (Jones et. al 1996).

MIDs, which are not wars, are the most common measure of international conflict used by peace scientists today, and this is for two key reasons.

The first is pragmatic: MIDs are more common than wars. That makes them more amenable to statistical analysis.

The second is conceptual. I think Braumoeller (2013) puts it best:

> [T]he best way to gauge the warlikeness of nations [is] to measure the frequency with which they run the risk of war. Initiating a war, or issuing a threat that creates the risk of war, is indicative of the willingness, in the worst case, to spill quite a bit of blood; that willingness is the best indicator of the propensity of the state to engage in violent conflict; and the individual conflict incident, or *militarized dispute*, is the most reliable objective indicator of that willingness (10, emphasis mine).

Linking back to the idea of *real* war (war as a costly means of bargaining), MIDs seem to capture the essence of this definition much better than the one proposed by Singer and Small, who require that a war involve actual use of armed military forces and substantial fatalities (at least 1,000). Real war, by Clausewitz's standards, might only need a minimum of a few shots across the bow at sea, or targeted strikes on military bases, for both sides to learn what they need to learn about the other side's ability, resolve, and the possible outcome of absolute war. Threats may even be enough.

Going further, Braumoeller (2019) later concluded in a book he wrote about trends in war that escalation to absolute war (like one of the World Wars) is hard to predict and can happen even when most or all the actors involved wish to avoid it. The time when purposeful political action is the most significant on the path the war is the initial stages when countries issue threats or make the decision to start shooting. MIDs capture this idea by definition. To relate this back to another idea in the last section, MIDs show the willingness of countries to roll the "iron dice."

1.4 How We'll Measure War In This Class

Going forward in this class, we'll primarily use the concept of the MID to study international conflict. However, as I'll discuss more in the next chapter, we won't use the COW MIDs dataset.

Without giving too much away, a recent controversy in peace science emerged in the last decade when a group of scholars undertook painstaking work to validate and recreate the COW MID dataset from scratch. What these scholars found surprised them.

They discovered thousands of data entry errors, including phantom conflicts (conflicts that had no supporting evidence to verify they happened) and double counting. So bad were these issues that this team of researchers claimed all MID data for years prior to 2002 were unreliable for analysis.

This was a bold claim, and a damning one if true. The maintainers of the MID dataset obviously responded, and the result was a series of publications that would slowly come out over several years documenting the argument that ensued. Eventually, the group of scholars who documented these data issues would get a National Science Foundation grant to reconstruct all the data from the ground up, and along the way they made some conceptual and technical improvements.

Perhaps the two most relevant updates are their rigorous reporting of battle fatalities and their choice to redefine wars as a subset of all conflicts rather than a distinct phenomenon. (In the MID dataset, an observation leaves the dataset the moment it escalates to a war.) Their new concept is called the militarized interstate confrontation (or MIC), which can include threats, shows, or uses of force short of war, but also all-out wars.

Even better, these scholars documented specific militarized interstate events (MIEs), their specific timing, and deadliness. These events are the individual threats, uses of force, and battles that comprise MICs.

In short, the data that I'll introduce in the next chapter is incredibly rich and represents an important new advancement in the study of international conflict. I therefore want you to become familiar and comfortable with this dataset as we work with it throughout this class.

References

Braumoeller, Bear F. 2013. "Is War Disappearing?" In *APSA Chicago 2013 Meeting*.

———. 2019. *Only the Dead: The Persistence of War in the Modern Age*. Oxford University Press.

Fearon, James D. 1995. "Rationalist Explanations for War." *International Organization* 49 (3): 379–414.

Gochman, Charles S, and Zeev Maoz. 1984. "Militarized Interstate Disputes, 1816-1976: Procedures, Patterns, and Insights." *Journal of Conflict Resolution* 28 (4): 585–616.

Mitchell, Sara McLaughlin, and John A Vasquez. 2024. *What Do We Know about War?* Bloomsbury Publishing PLC.

Palmer, Glenn, Roseanne W McManus, Vito D'orazio, Michael R Kenwick, Mikaela Karstens, Chase Bloch, Nick Dietrich, Kayla Kahn, Kellan Ritter, and Michael J Soules. 2022. "The MID5 Dataset, 2011–2014: Procedures, Coding Rules, and Description." *Conflict Management and Peace Science* 39 (4): 470–82.

Tuchman, Barbara W. 1994. *The Guns of August: The Outbreak of World War i; Barbara w. Tuchman's Great War Series*. Random House Trade Paperbacks.

von Clausewitz, Carl. 2003. *On War*. Penguin UK.

Wagner, R Harrison. 2000. "Bargaining and War." *American Journal of Political Science*, 469–84.