Beyond Body Counts:
Vietnam War Ground Combat Data

Gregory Smith, PhD

September 20, 2025

Abstract

This paper presents one of the first systematic efforts to clean, chronicle, and visualize data files related to ground combat operations collected by the United States government during the Vietnam War. These files document incidents initiated by both friendly and enemy forces. The paper begins by providing a high-level overview of MACV's data collection efforts in Southeast Asia and argues for the importance of open-source efforts to clean and preserve ground combat data. The second section gives a general overview of the twelve original data files that accompany this article. The third discusses how the data files can expand research programs on conflict dynamics and political violence. The fourth section documents the steps taken to process and clean the files. The article concludes with a quantitative overview of the ground war and the Tet Offensive. In general, the paper presents a new direction for IR and conflict scholarship.

Word Count: 10,000

Key Words:

Vietnam War, data, US foreign policy, political violence, COIN, battlefield effectiveness

I would like to thank Liwu Gan, Kevin Li, Maël van Beek, and the three anonymous reviewers for their helpful feedback, the archivists at the National Archives for their help accessing the original datafiles, and Bob Rudis for his help converting MGRS coordinates.

Assistant Professor, Department of Political Science, Syracuse University. Email: gsmith28@syr.edu

Introduction

Micro-level data on political violence and armed conflict have been central to advances in international relations and conflict scholarship over the past several decades. Such data are characterized by their fine-grained spatial and temporal resolution and typically record details of specific violent incidents, including the date, time, location, actors involved, and characteristics of the event — such as the number of individuals killed or wounded and the attack method. In particular, data sets such as the Armed Conflict Location and Event Data Project (ACLED), the Uppsala Conflict Data Program Georeferenced Event Dataset (UCDP GED), the Global Terrorism Database (GTD), and Significant Activities (SIGACTs) data from Iraq and Afghanistan, and new platforms and frameworks such as xSub and PRIO Grid have allowed scholars to incorporate these fine-grained event records as the foundation for their analyses (LaFree and Dugan, 2007; Raleigh et al., 2010; Tollefsen et al., 2012; Sundberg and Melander, 2013; Zhukov et al., 2019). These data have been used to study a wide range of topics, including the effects of terrain on rebel capacity, the determinants of civilian targeting, the effectiveness of conventional and unconventional strategies, and the impact of technology on battlefield effectiveness, among many other essential topics (Berman et al., 2018).

Many of these recent efforts to chronicle conflict at a granular level are based on the United States' Vietnam War-era initiatives, which pioneered one of the first systematic efforts to collect micro-level conflict data. As part of its ongoing war efforts in Southeast Asia, the Department of Defense implemented unprecedented efforts to collect data and "quantify" nearly every aspect of the Vietnam War (Harrison, 1988; Pierce et al., 1976; Thayer, 1975). Led by Secretary of Defense Robert McNamara, the US military attempted to devise metrics that would enable them to determine whether they were "winning" or "losing" and optimize their efforts to defeat the North Vietnamese military on the battlefield and pacify South Vietnam (Daddis, 2011; Thayer, 1975). More specifically, these files contain incident-level information about North Vietnamese Army (NVA) and Viet Cong (VC) attacks across Southeast Asia, Allied ground combat operations, and bombing sorties that collectively dropped over 7 million tons of ordnance. In addition, these files also document the locations of enemy base areas, naval surveillance and interdiction efforts, naval gunfire support, estimates of territorial control, and aggregate summaries of psychological

operations (PSYOPs), in addition to a variety of other topics (National Military Command System Support Center, 1969).[1]

Despite the impressive scale and scope of data collection efforts during the Vietnam War, most of the electronic records are inaccessible. Currently, only air combat files have been cleaned and publicly released as the Theater History of Operations Reports (THOR) database (Robertson and Barth, 2013). This data accessibility problem is troubling for several reasons. First, there are well-documented historical biases and errors in the data collection efforts. These issues can only be resolved if the underlying data are publicly available in an open-source format, where scholars can work together to identify problems and solutions at scale. Second, replicating existing studies is impossible without access to the original records, as scholars have not released cleaned versions of the source files that serve as the foundation for their analysis, and many of the findings may be wrong or misleading, given the problems with collection efforts. Third, new research pathways in conflict studies—related to conflict dynamics, military effectiveness, and the long-term effects of war—have been obstructed since scholars have been unable to leverage the wealth of ground combat data. Finally, the release of ground combat records will facilitate pairing rich qualitative and quantitative accounts of the war in a manner that was previously impossible.

This paper details the first effort to systematically chronicle, clean, and present a combined ground combat database that will serve as a complement to the THOR database, which documents the Allied bombing of Southeast Asia, and more modern databases, such as the SIGACTs data that document ground combat incidents in Iraq and Afghanistan. Two sets of data accompany this article. The first set of files includes cleaned versions of twelve original data files that document ground combat actions and incidents, combat support efforts, or estimates of territorial control.[2]The orig-

[1]Here, the terms "Allied" or "friendly" forces refer to the militaries from the United States, South Korea, Australia, New Zealand, Thailand, and the Philippines. Likewise, enemy forces is the terminology used in the files to refer to NVA and VC forces. Compared to the current conflict literature, the term' friendly forces' generally corresponds to counterinsurgents. In contrast, the term enemy generally refers to insurgents, particularly when referring to the VC. However, these terms are imperfect and imprecise, as the NVA relied on conventional actions and offensives throughout the war. In this respect, the terms "friendly" and "enemy" are not perfect, but do not give the false impression that the United States and its allies were exclusively fighting an insurgency. I rely on the terms friendly, allied, and enemy forces to remain consistent with the source files.

[2]Not every electronic record about the Vietnam War has been cleaned and released with this article. For example, data documenting airlift operations were not included since they detail combat sustainment operations.

inal files have been updated with informative column names, intelligible data entries (that replace poorly defined alphanumeric characters), latitude and longitude coordinates, and the removal of duplicate events, among other changes. The second data set is a master database that includes ground combat incidents initiated by Allied and enemy forces, combining the incident information from four databases referred to as VNDBA, SITRA, TIRSA, and VCIIA files, respectively. This data set standardizes the information in the individual data files and removes duplicate observations that appeared in multiple source files. This combined database will enable researchers to easily access each ground combat incident. Finally, several tools accompany the cleaned files to allow researchers to utilize the data. These include a GitHub website that contains downloadable files and the corresponding documentation, and an R package with a simple API that will enable researchers to integrate the data directly into their workflows and that includes all the functions and procedures used to replicate the data processing steps from the raw to finished files.

The article includes five major sections. The first section presents a high-level overview of MACV's data collection efforts in Southeast Asia and discusses the importance of open-sourced efforts to share the ground combat data. The second section gives a general overview of the twelve original data files that accompany this article. The third section reviews the literature on war termination, political violence, insurgencies, and counterinsurgency (COIN) strategies and discusses how the data files can be used to extend these research programs. The fourth section documents the steps taken to process and clean the files to be used in large-N analyses and serves as a template for future efforts to clean additional data files related to the Vietnam War. The fifth and final section uses the combined master database to provide a macro-level overview of ground combat operations across Southeast Asia from 1963 to 1973 and a micro-level overview of VC and NVA attacks during the Tet Offensive for the cities of Hue and Saigon. In general, the paper presents a new direction for Vietnam War scholarship.

Toward an Open-Source Platform for Vietnam Combat Data

The unprecedented scope and complexity of data collection efforts during the Vietnam War represented the first attempt to quantify modern warfare and COIN operations (Thayer, 1975). To

determine whether the United States was winning or losing the conflict, senior officials developed a variety of metrics to measure battlefield effectiveness, territorial control, and pacification progress. These efforts produced many indicators, including enemy kill counts, friendly-enemy kill ratios, and hamlet-district control scores derived from ground combat files, all intended to paint a picture for senior U.S. officials and ground commanders about the war's trajectory (Daddis, 2011). The military's monthly progress reports tracked countless variables ranging from force ratios and Viet Cong incidents to tactical air sorties, weapons losses, and population control measures. The scope of these files presents a veritable gold mine of evidence for scholars seeking to understand the complexities of modern counterinsurgency warfare, test theoretical frameworks about conflict dynamics, and draw systematic comparisons between the Vietnam War and contemporary military operations.

Despite the apparent importance of battlefield metrics during the Vietnam War, the data files used to calculate these statistics are inaccessible to most scholars because of the way they have been stored and maintained. These data files present significant technical challenges, including antiquated formats incompatible with modern software, missing delimiters, indecipherable alphanumeric coding systems that require extensive documentation to decipher, and zoned-decimal storage that prevents mathematical operations. Additional complications include incomplete military grid coordinates, which lack the necessary zone identifiers, fixed and period set data sets that require complex merging procedures, and duplicate entries that must be removed. In general, these challenges have prevented most scholars from utilizing these records. Ultimately, an open-source platform for sharing the cleaned files and documenting every processing step is essential for resolving the many technical challenges these records present and extending conflict studies.

Enabling scholars to access ground combat data from the Vietnam War will be invaluable for several reasons. First, these records will allow scholars to systematically document and correct problems with the source files, which are notoriously plagued by serious data collection issues. Second, the files will enable scholars to replicate previous studies that have not made the raw data files available, which serve as the foundation for their analysis, which may be affected by data collection issues. Third, this will enable scholars to resolve a variety of open questions and debates in the field of conflict studies and develop new lines of inquiry. Finally, scholars can pair this data with

rich qualitative evidence to improve our understanding of the Vietnam War. By making both data and cleaning scripts openly available, researchers can build upon existing work, verify findings, and develop new theoretical insights while maintaining the highest standards of scholarly rigor and transparency.

Although the scope of data collection efforts in Southeast Asia was impressive, many scholars have argued that the resulting data files have serious problems. In their analysis of the raw air combat records, High et al. (2013) document several issues with the source files. For example, they argue that the files contain falsified target locations, widespread double-counting of B-52 missions after 1971, frequent misclassifications of kinetic versus non-kinetic sorties, and tens of thousands of records with missing or miscoded weapon types. Despite noting these problems, High et al. (2013) did not release cleaned versions of the air combat records, and other problems likely remain unidentified. In addition, scholars have also identified issues with the ground combat files. Some have argued that official body count counts were routinely inflated because commanders faced strong incentives to exaggerate kills. At the same time, those compiling the Hamlet Evaluation System—which estimated the level of territorial control—also faced pressure to demonstrate progress by inflating estimates of Allied control and relied on sporadic village visits, ambiguous coding rules, and poorly translated questionnaires that produced noisy and often biased estimates of local control and security (Connable, 2012; Harrison, 1988; High et al., 2013; Pool et al., 1968; Thayer, 1975). Open-sourced efforts will enable scholars to identify, document, and correct such problems collaboratively.

The lack of clean and accessible source files undermines the transparency and replicability of several noteworthy studies that rely on Vietnam War data. During the last 20 years, there has been a growing interest in incorporating these records into quantitative analyses of the Vietnam War. However, these scholars have not released cleaned versions of the source files (Dell and Querubin, 2017; Kalyvas and Kocher, 2009; Kocher et al., 2011). For example, Kalyvas and Kocher (2009) analyzed the Hamlet Evaluation System, which measured territorial control at the village level, to identify patterns of selective insurgent violence. Similarly, Dell and Querubin (2017) and Kocher et al. (2011) study the effect that indiscriminate Allied bombing had on Viet Cong tactics and

shifts in local territorial control. The well-known collection issues and biases make it imperative for scholars to carefully examine the robustness of the original results, especially given the importance of the scholarly debates to which they speak. Unfortunately, such efforts are currently impossible without access to the original records.

In addition to preventing scholars from careful replications, scholars have been precluded from leveraging the information collected about the ground war to extend conflict studies. Currently, only data files documenting the air war over Southeast Asia have been cleaned and shared publicly. In 2015, the Air Force Research Institute released the Theater History of Operations Reports Database (THOR), which combines data from three Vietnam-era Joint Chiefs of Staff databases: the Combat Activities File (CACTA) 1965-1970, the Southeast Asia Database (SEADAB) 1970-1975, and the Strategic Air Command Activities Database (SACCOACT) 1965-1975. The THOR database contains records from the United States Air Force, the United States Navy, the United States Marine Corps, and missions flown by the Republic of Vietnam Air Force, the Royal Lao Air Force, the Korean Air Force, and the Royal Australian Air Force. These records detail a sortie's objective, attacking plane, coordinates, type of munition dropped, total ordnance dropped, take-off location, and mission objectives, among other details. Pairing the air and ground combat records will enable scholars to intervene in various debates and develop promising new research pathways in conflict studies. A variety of promising directions are discussed below.

Finally, the release of ground combat records will facilitate the pairing of quantitative studies with rich qualitative records, such as those housed at institutions like Texas Tech University.[3] In addition to enabling researchers to ground-truth statistical findings against historical narratives and eyewitness accounts, such efforts will enable scholars to better understand the lived experiences of people on the ground. Unfortunately, existing scholarship has prioritized studies of military actions and the combatants. Consequently, there remains a gap in our understanding of war as it is perceived by its victims (Lin, 2022). Future research should aim to connect the electronic records with local, civilian experiences of violence, to more richly incorporate victims' accounts and perspectives of the war (Kwon, 2006, 2008; Tang et al., 2018).

[3]For a link to the Texas Tech archives, see: `https://www.vietnam.ttu.edu/`.

An Overview of the Cleaned Ground Combat Files

The cleaned data files accompanying this article contain invaluable information about ground combat in Southeast Asia from 1963 to 1974 and support efforts to defeat VC and NVA forces.[4] The cleaned study data represent a small proportion of the data collected by the United States military and its allies during the war. The primary criterion for inclusion was that each of the following data files documented Allied ground combat operations or combat support operations during the Vietnam War. The following section provides an overview of each of the cleaned data files. The first subsection discusses the four incident-level data files, Vietnam Database (VNDBA) Files, the Terrorist Incident Reporting System (TIRSA), the Viet Cong Initiated Incidents Files (VCIIA), and the Situation Report Army (SITRA) Files, which form the basis for the combined ground combat database. The subsequent subsections summarize the information included in the remaining files and group them based on their substantive focus. The second subsection discusses incident-level naval gunfire and surveillance data. The remaining subsections detail US casualties, the monthly locations of maneuver battalions, enemy base areas, the hamlet evaluation files, and the data on psychological operations, respectively. Additional information about these files is presented in the Appendix, and a summary of these files is presented in Table 1.

Incident Level Ground Combat Data

The most granular data included in the ground combat files detail incident-level reports of friendly and enemy-initiated ground combat events that occurred from 1963 to 1973. Incident-level data on ground combat events appear in four separate data files. These sources include the Terrorist Incident Reporting System (TIRSA), the Viet Cong Initiated Incident (VCIIA) file, the Vietnam Database (VNDBA), and the Situation Reports Army (SITRA) files. Each of these files identifies specific attributes of ground incidents that were initiated by Allied or North Vietnamese forces, though the VCIIA and TIRSA files only detail enemy-initiated incidents. These details include the date and time the incident was started, the grid coordinates of the event, the target of the attack,

[4] In general, the periods contained in the incident-level files overlap with the start of those operations. However, in the case of the files documenting combat support operations such as PSYOPs, the source files are not particularly transparent about why US officials decided to begin recording that information. Some scholarship, such as the work by Thayer (1975) and Daddis (2011), provides particularly fruitful discussions of the origins of these data files and the accompanying use of statistics and metrics to evaluate progress in the war in South Vietnam.

the resulting physical damage, the number of casualties, the type of tactics that were employed (ex: "clear and hold" or "search and destroy), and the nationalities, services, and units of the aggressor and target forces.

On the Allied side, operations were broken down into several different categories. The most general distinction is made between small and large unit operations. Small unit operations involved "less than three companies of line troops but not less than a fire team," while large unit operations involved "three or more companies of line troops" (National Archives and Records Administration, 2005, pp 10). In addition to these general categories, friendly operations were labeled with specific operation types such as "search and destroy missions", "reconnaissance missions," and "clear and hold missions," among others. Similarly, incidents initiated by the enemy were divided into seven general categories, including entry, propaganda, sabotage, terrorism, mining, harassment, and attacks. Within these general categories, enemy actions were further divided into more specific actions such as bombing, kidnappings, and anti-aircraft fire, in addition to many other categories. Finally, many of the incidents contain detailed descriptions of the equipment and supplies that were expended or destroyed. These loss descriptions include information on the fired rounds, weapons destroyed, types of houses or structures damaged, and information on food and medical supplies.

Incident Level Naval Gunfire and Surveillance Data

In addition to incident-level data on Allied operations and North Vietnamese attacks, two additional incident-level files were cleaned detailing naval operations that supported Allied ground units. The first of these data files, the Combat Naval Gunfire Support File (CONGA), details gunfire support of ground units in Southeast Asia from 1966 to 1973. Naval gunfire support generally refers to using naval artillery to provide fire support to amphibious assaults and other troops operating within their range (Martin et al., 2020). During the Vietnam War, mainly destroyers and cruisers provided gunfire support to Army and Marine forces. Naval gunfire support was used to target preselected areas, often in preparation for an attack or to target known enemy ground positions, and in response to requests made by Allied units that were engaged with enemy forces. Ground or aerial spotters facilitated such calls for fire support and were often instrumental in preventing friendly units from being overrun. The CONGA files include information on the date and

time of the support mission, the name and type of ship that was firing the barrage, the name of the supported operation, the specific ordnance employed, target coordinates, and details about the target type and what was destroyed.

The Naval Surveillance Activities File (NASVA) documents US Navy and Coast Guard maritime surveillance efforts from February 1966 to December 1972, encompassing two complementary operations designed to counter North Vietnamese maritime logistics and deny Viet Cong access to strategic waterways. Operation Market Time aimed to prevent North Vietnamese ships from providing support to Viet Cong forces by using a combination of air and naval assets to intercept suspicious vessels. Likewise, Operation Game Warden served as the inland counterpart, focusing on denying Viet Cong access and resources in the strategically vital Mekong River Delta, which housed 40% of South Vietnam's population (approximately 8 million people), produced most of the country's rice, and contained more than 1,400 miles of natural waterways plus 2,400 miles of man-made canals across 15,000 square miles. The NASVA files contain incident-level details of particular interdiction missions, including information on vessel types (wood or steel), encounter times, ship direction, friendly vessel details, coordinates, and quantities of goods destroyed, in addition to weekly summaries of naval surveillance efforts providing overviews of the number and types of ships encountered.

Casualty Data

A devastating legacy of the Vietnam War was the deaths of more than 58,000 service personnel, in addition to thousands of soldiers who were wounded or missing in action. To help preserve a detailed record of US military casualties, a clean version of the Defense Casualty Analysis System (DCAS) has been included in the cleaned data files. Although incident-level ground combat data reference some details about the number of soldiers killed or wounded in action during particular events, these data are presented only in aggregate form, and there are questions about their reliability.[5] The DCAS file helps account for some of these deficiencies by detailing the military

[5]For example, the SITRA files have separate columns that list the total number of soldiers killed in action, missing in action, and wounded in action for each incident. However, there are questions about the reliability of the aggregate casualty data, such as whether or not particularly large casualty counts are valid or an erroneous entry, or if each of the casualties can be attributed to the listed unit and regiment.

background of the service member, such as service, unit, rank, and military occupation; personal background, such as gender, age, hometown, marital status, and religion; and details of the cause of death, including date, location, and specific cause of death. Although this database lacks information about personnel who were wounded in action, it can be paired with incident-level data, using time, location, and unit information to provide detailed accounts of the results of each battle.

Maneuver Battalion Location Data

MACV commanders created the Southeast Asia Friendly Forces File (SEAFA) to provide a comprehensive picture of the location of friendly maneuver battalions in South Vietnam from October 1966 to July 1972. Here, "friendly" refers to the forces of the United States and South Vietnam, as well as the "Free World" units comprised of South Korean, Australian, and Thai troops. Relatedly, "maneuver battalions" refer to units consisting of 300 to 1,000 soldiers designed to engage in direct combat with enemy forces and execute tactical movements on the battlefield. These units include infantry battalions, tank battalions, and armored cavalry squadrons, although reconnaissance battalions internally assigned to infantry, airmobile, and mechanized divisions were excluded from this measure. In addition, combat support units and logistics units are also excluded. Consequently, this is far from a complete record of every battalion. The SEAFA files contain monthly records, recorded on the last day of each month, about each maneuver battalion. The recorded information includes a unique unit ID, the name of each unit, the coordinates of each battalion, the unit service, the country of origin, the unit type (e.g., infantry, armored cavalry, or airmobile), the control headquarters unit name (e.g., "101st ABN DIV" or "25TH ARVN DIV"), and the name of each station.

North Vietnamese Base Areas

During the war, Viet Cong cadres became well known for taking refuge in highly fortified sanctuaries both within the territorial confines of South Vietnam and across the border in Laos and Cambodia. The Enemy Base Area File (BASFA) provides details of these base areas from July 1966 to June 1971, and for many of the base areas constructed before this period. The BASFA files provide a monthly record of the North Vietnamese base areas. The primary contribution of these

files is to provide the geographic coordinates of each base. These records provide a record of the central coordinate (or the estimated "center of mass") for each base and the coordinates required to plot the entire base area as a polygon. The latter set of coordinates contains up to nine possible values so that fairly intricate polygons can be plotted. In addition to detailing the location of base areas, the BASFA Files contain information on the region that the base was located in, the date that the area first appeared, whether a base was a campaign priority, the current status of the base, actions (such as search and destroy operations) that were initiated against the base, and the date that the area was last recorded. In total, 139 base areas are included in the data file. The coordinates of each center of mass and the perimeter of the base area are plotted in the Appendix.

Hamlet Evaluation System Data

The Hamlet Evaluation System (HES) is probably the most famous data collection project of the Vietnam War. In addition to being infamous for its efforts to quantify progress in South Vietnam by devising a numerical basis for success and failure, since many argue that it was fundamentally flawed or even intentionally misleading, it has been the subject of numerous research studies and histories (Daddis, 2011; Kocher et al., 2011; Dell and Querubin, 2017). The HES system was designed by analysts from the Office of the Assistant Secretary of Defense to allow military commanders from the United States and the Government of the Republic of Vietnam (GVN) to assess the ongoing war effort at the hamlet level, which comprised the smallest population units. The HES files contain monthly assessments of hamlet and village security that measured territorial control at the village level. These monthly reports documented security conditions, enemy presence, economic conditions, development scores, and estimates of government control for each of the 21,735 hamlets in the sample, enabling military commanders to track changes in territorial control over time.

Psychological Operations (PSYOPS) Data

Throughout the war, the US military and supporting agencies relied extensively on psychological operations (PSYOPS) to support pacification efforts. These operations aimed to influence the behavior and morale of both enemy forces and civilian populations by using propaganda, psycho-

logical tactics, and cultural sensitivities (Roberts, 2018).[6] The National Archives has preserved two data files that contain aggregate summaries of PSYOPS implemented in South Vietnam at the district level, which were part of the Psychological Operations Information System (PSYOPSIS) that was collected from March 1970 to February 1973. The first of these subsidiary files, the PSYOPSA file, contains information concerning ground operations, including radio and television transmissions, in addition to the demographic details of the targeted district. The second subsidiary file, the SORTIEA file, details aerial operations where newspapers, magazines, and leaflets were dropped over enemy positions. Together, these files contain information about the date of operations, the target audience, the number of distributed leaflets, posters, and publications, loudspeaker and radio broadcast hours, television transmission hours, and demographics of the intended audience broken down into specific operations. These district-level indicators were designed to augment the monthly security evaluations presented in the Hamlet Evaluation System.

[6]For example, the Chieu Hoi ("Open Arms") Amnesty Program encouraged defections by promising safety and reintegration into South Vietnamese society. Leaflets containing information about the program were dropped over areas believed to contain VC cadres by Allied air crews or via artillery shells. These invitations to defect were intended to act as safe conduct passes for those who wished to defect. Official military records claim that this program resulted in the defection of more than 100,000 communist forces. Relatedly, Operation Wandering Soul was another infamous operation that sought to weaken the Viet Cong's resolve by increasing desertions. This program attempted to leverage Vietnamese fears about death and burial rituals by playing eerie sounds and recordings that pretended to be the ghosts of killed Viet Cong soldiers or loved ones.

Table 1: Summary of the Cleaned Vietnam War Data Files

Abbreviated Name	**Full Name**	**Period Covered**	**Summary**
VNDBA	Vietnam Database	1/1963 - 12/1969	Data on ground combat operations in South Vietnam during the Vietnam War, part of OPSANAL system.
SITRA	Situation Report Army	5/20/1966 - 3/12/1973	Records of ground combat operations in Southeast Asia, detailing military operations, casualties, and logistics.
TIRSA	Terrorist Incident Reporting System	10/1967 - 2/1973	Records of Viet Cong and North Vietnamese incidents against South Vietnamese civilians, part of OPSANAL system.
VCIIA	Viet Cong Initiated Incidents Files	1/1/1965 - 12/31/1968	Data on enemy-initiated incidents during the Vietnam War, focused on Viet Cong activities.
CONGA	Combat Naval Gunfire Support File	3/1966 - 1/1973	Records of naval gunfire support missions during the Vietnam War.
NASVA	Naval Surveillance Activities File	2/22/1966 - 12/9/1972	Information on operations Market Time and Game Warden in South Vietnam.
DCAS	Vietnam Conflict Extract Data File	1956-2006	Records of 58,220 US military fatal casualties of the Vietnam War.
SEAFA	Southeast Asia Friendly Forces File	10/1966 - 7/1972	Information on American, South Vietnamese, and Allied maneuver battalions in Southeast Asia, part of OPSANAL system.
BASFA	Enemy Base Area File	7/1/1967 - 6/1/1971	Data defining enemy base area locations in South Vietnam, North Vietnam, and Cambodia.
HES	Hamlet Evaluation System	1967-1974	Monthly data on security, political, and socioeconomic status of South Vietnamese hamlets and villages.
SORTIEA	Psychological Operations: Aerial	3/1970 - 2/1973	Records of aerial psychological operations by US military during the Vietnam War.
PSYOPSA	Psychological Operations: Surface	3/1970 - 2/1973	Records of surface psychological operations by US military during the Vietnam War.

Advancing IR and Conflict Studies

The ground combat data provides a promising pathway to extend various research questions related to conflict studies, military effectiveness, and the long-term effects of war. These include the impact of bureaucratic incentives on battlefield performance, the effect of combat on wartime negotiations, the relationship between terrain ruggedness and insurgent violence, the long-term consequences of war on economic development, and the link between territorial control and political violence. Although these avenues of research are particularly promising, they are far from an exhaustive list of applicable topics.

The first line of research draws on debates about the validity of data collection efforts. Rather than rendering the data unusable, biased statistics used to demonstrate progress in the war effort can be transformed into a fruitful line of inquiry. Scrutinizing when and why officers filed inaccurate hamlet security assessments or inflated body count figures will allow scholars to uncover the bureaucratic incentives embedded in wartime reporting. These patterns and incentive structures are likely more complicated than analysts have predicted. Newly arrived officers, for example, may have found it advantageous to depict conditions pessimistically at the outset and then report steady improvement to bolster their performance evaluations, which are necessary for career advancement, rather than simply documenting inflated progress measures that would leave little room to demonstrate subsequent improvement (Tate, 1978). By linking incident-level records with Hamlet Evaluation System scores, unit rotation schedules, and the identities of the officers responsible for each report, researchers can trace these incentive structures empirically and test how they shaped battlefield behavior. This approach not only rehabilitates the data, but also enriches broader debates on bureaucratic politics and civil–military relations, demonstrating how organizational pressures filter into tactical decisions and ultimately influence military effectiveness (Avant, 2019; Teodoro, 2011).

Second, ground combat files present a unique opportunity to advance our understanding of the dynamic relationship between battlefield developments and diplomatic behavior during prolonged conflicts (Ito and Hinkkainen Elliott, 2020; Menninga and Prorok, 2021; Min, 2020, 2022). These

data contain the spatial and temporal granularity necessary to extend research on conflict negotiation dynamics beyond the aggregate-level measures typically used in existing scholarship. The geocoded incident data documenting both friendly and enemy ground operations across Southeast Asia from 1963 to 1974, in addition to the THOR bombing data, enable researchers to examine not only the overall volume of combat activity but also the precise location and timing of attacks and their effects on ongoing diplomatic efforts. This offers a way to extend Min's (2022) Korean War analysis, which relied primarily on aggregate casualties and movement measures, by allowing scholars to investigate how localized battlefield successes or failures in strategically important areas might influence negotiation behavior in ways that theater-wide metrics cannot capture. The availability of detailed geographic coordinates for combat incidents allows researchers to test whether proximity to key population centers, transportation networks, or symbolic locations produces different effects on diplomatic postures and rhetoric and identify the metrics that are most important to senior policy makers (Gartner, 1999).

Third, a wealth of recent research has studied the link between terrain and the patterns and success of various asymmetric warfare tactics. Terrain ruggedness, for example, is one prominent factor that affects a counterinsurgent's ability to target its opponent and the insurgent's ability to avoid those efforts. Many argue that insurgents tend to have higher levels of control in rough terrain than in smooth terrain, as it is easier to conceal their movements and protect their organization from counter-insurgents (Buhaug and Lujala, 2005; Buhaug et al., 2009; Carter et al., 2019; Fearon and Laitin, 2003; Shaver et al., 2019; Vogt et al., 2015). Similarly, recent scholarship has examined the determinants of insurgent diffusion (Hammond, 2018; Schutte and Weidmann, 2011; Zhukov, 2012; Zhukov and Stewart, 2013). The Vietnam War files offer a straightforward way of extending these studies to assess how insurgent strategies may have changed or been driven by particular terrain features. Southeast Asia has a variety of terrain features. For example, the terrain in the Central Highlands is generally rugged and has extensive vegetation. In contrast, the coastal lowlands are typically flat and are known for rice cultivation, while the Mekong Delta contains a dense network of rivers and canals.[7] Researchers could leverage the attack data in the THOR bombing records,

[7]Relatedly, the interplay between terrain characteristics and seasonal patterns of conflict may be particularly important, as agricultural cycles tied to specific geographic features can create predictable temporal variations in insurgent activity (Guardado and Pennings, 2025).

combined with the incident-level combat data and CONGA files, to model changes in insurgent attack locations and tactics across these varied terrain types. These detailed geographic and temporal records, when integrated with the details of enemy attacks, the Hamlet Evaluation System (HES) territorial control scores, and interdiction data from the NASVA files covering the Mekong Delta, would enable scholars to systematically test how terrain characteristics influenced both the spatial diffusion of insurgent activities and the effectiveness of counter-insurgency operations across different geographic regions of Southeast Asia.

A fourth arc for future research extends studies that examine the effects of war and political violence that shape the long-term political trajectories of participants (Charnysh, 2019; Charnysh et al., 2023; Cirone and Pepinsky, 2022; Dinas et al., 2021; Homola et al., 2020; Pepinsky et al., 2024; Walden and Zhukov, 2020; Wayne and Zhukov, 2022). One subset of this literature, particularly relevant to Southeast Asia, examines the impact that unexploded ordnance has on the long-term economic prosperity of the host country (Barthelme et al., 2024; Lin, 2024). Lin (2024) argues that short-term adjustments in agricultural methods accumulate to long-term underdevelopment and poverty. In particular, she finds that the most fertile land in Cambodia suffered the heaviest consequences and ultimately became the least productive, since bombs are less likely to explode in the comparatively softer land. Although this literature is compelling, there is room for expansion into other regions and to focus on different aspects of ground combat. The data files include records for naval gunfire support around South Vietnam and reference specific battles and events that could extend these studies to include different aspects of fighting in other regions of the war.

A fifth strand of literature examines the causes and consequences of territorial control as it relates to insurgent tactics and violence. Territorial control is notable because many conflicts are shaped by competition over particular places, and control is a significant determinant of how both rebel groups and counter-insurgents optimize their strategies (Balcells and Stanton, 2021; Berman et al., 2018; Condra et al., 2018; Condra and Wright, 2019; De la Calle and Sánchez-Cuenca, 2015; Kalyvas, 2006; Stewart, 2018). For example, many argue that areas in which political control is contested will tend to be particularly prone to acts of selective violence. In contrast, places that are

firmly under the control of rebels will be relatively peaceful, since they do not need to use additional violence to achieve their political objectives. The ground combat files offer a means of extending this literature, since a great effort was made to measure local and territorial control using the HES system. Although there have been numerous analyses of the HES system and several studies that use it to assess the implications of territorial control, this research has barely scratched the surface (Dell and Querubin, 2017; Kalyvas and Kocher, 2009; Kocher et al., 2011). Kocher et al., for example, correlate territorial control estimates with selective violence, but only use measures included in the HES data, and their analysis was limited to a narrow temporal domain and spatial window. The combined database can extend this analysis by assessing how this control is associated with particular tactics or understanding how these patterns varied across different regions and periods of the war in Southeast Asia.

Finally, the data files provide a way to extend existing research on the effects of various military strategies, particularly concerning the effectiveness of insurgents and counterinsurgent operations. One prominent example of recent scholarship that can be extended and re-assessed studies the effectiveness of "hearts and minds" vs "overwhelming firepower" oriented approaches to COIN (Dell and Querubin, 2017; Johnson, 2018; Long, 2016). The basis for this comparison lies in the different strategies used by the Marines and the Army during the war. The Army relied on overwhelming firepower, while the Marines favored a population-centric approach focused on securing the local population. Although there is emerging consensus that the Marines' approach was more effective, further research is required to assess the generalizability of these findings and to determine how these tactics shaped insurgent and non-combatant behavior. Similarly, ground combat data should be paired with data on bombing and naval gunfire support to assess the effectiveness of combined arms operations, how they shape insurgent tactics, and how the technology of conflict affects effectiveness. Although several studies have evaluated the causes and consequences of combined arms operations at the macro-level, few have tested these insights using a granular micro-level analysis (Allen and Martinez Machain, 2018, 2019; Horowitz and Reiter, 2001; Post, 2019).

Processing and Cleaning the Data: The Pipeline

The process of cleaning each of the data files can be broken into three major sub-components: (1) data processing, which involves converting the data files into a format that can be used with modern software tools, (2) data cleaning, which involves reformatting the data and updating the data values so that they can be used for research projects, and (3) the creation of the complete ground combat incident database. The central challenge with many of the data files can be traced back to the way the files were created and stored during the war, which relied on an early database system called the National Military Command System (NMCS) Information Processing System 360 Formatted File System (more commonly referred to as NIPS).[8] One of the defining attributes of this database is that the files related to each record are organized into one of several distinct elements, including the Control Set, Fixed Set, and Period Sets. The Control Set contains the unique record identifier that links to the Fixed Set and Periodic Sets, such as identifying information for a district or village. In contrast, the Fixed Set contains non-repetitive data, such as the names of military operations like Operation Attleboro or Cedar Falls. The Periodic Sets contain fields that can be repeated and include information such as friendly and enemy casualties, expended ordnance, and captured equipment. Data files frequently contain multiple Periodic Sets.

Efforts to convert the data files from their original proprietary format have led to several complications that must be resolved before the data can be cleaned and used. Internal NARA efforts in the late 1970s and early 1980s led to many files being converted to a binary flat-file format called the extended binary coded decimal interchange code (EBCDIC). The primary challenge with binary files is that they cannot be parsed easily using modern software programs, which typically rely on ASCII encoding. Even after binary files have been converted to ASCII characters, many of the numeric values are stored as zoned decimals, which was a way of preserving numeric data in mainframe environments and needed to be converted to integers. In addition, converted binary files lack delimiters, which are necessary for distinguishing data values. In practice, this means that users cannot easily distinguish between different pieces of information, such as values that detail a unit's tactical objective, versus information about its size or nationality. Subsequent NARA ef-

[8] `https://tinyurl.com/nara-vietnam-war-files`.

forts resulted in the creation of the NIPSTRAN software to convert the original database files into ASCII characters. However, the tool can only be applied to the files that remain in binary form. Even with this tool, researchers must merge the Control Set, Fixed Set, and Period Sets, which are returned separately. These data processing steps are summarized in Table 1.

Table 2: Summary of the Data Processing Steps

Step	**Description**
1	Download or purchase the files from NARA
2	Convert the NIPS Files to ASCII characters (when applicable)
3	Convert the EBCDIC files to ASCII encoding (when applicable)
4	Add delimiters to separate the columns
5	Convert zoned decimals to integers
6	Merge the control, fixed, and periodic set files

Even after data processing has been completed, extensive work is required for the data to be suitable for descriptive or statistical analysis. First, the column labels referenced in the data file use nondescript names that cannot be identified or interpreted without extensive documentation. Due to the database conventions at the time the files were constructed, the names were limited to five characters.[9] Therefore, hundreds of pages of documentation frequently accompany these files. Finally, records that reference the location of events, such as Allied combat operations or Viet Cong attacks, are presented using military grid reference system (MGRS) coordinates rather than latitude and longitude coordinates, which cannot be used directly by open-source plotting libraries. Furthermore, the listed coordinates lack the three-character UTM grid zone codes, which typically appear at the beginning of every coordinate string, necessary to convert MGRS coordinates to the more common latitude and longitude coordinates.[10] All this information must be updated to analyze and assess these data systematically. These data cleaning steps are summarized in Table 2 below.

[9]In the SITRA files, for example, which contain records of ground combat operations in Southeast Asia, the column named FNATL contains information about a unit's nationality, the column CNTRY refers to the country in which the operation occurred, and OPNAM refers to the name of a military operation. Similarly, the values used to record responses include hundreds of nondescript alphanumeric codes. In the force nationality column, for example, character A referred to the Republic of Vietnam (RVN) forces, character M referred to the US Marines, and character D referred to the US Coast Guard, in addition to many other values.

[10]A discussion of this conversion process, including the steps necessary to add in the missing UTM grid zones, appears in the Appendix.

Table 3: Summary of Data Cleaning Steps

Step	Description
1	Create informative column names
2	Correct and reformat the date and time vectors
3	Update the alphanumeric entries with informative values
4	Add missing values (e.g., 9999 to NA)
5	Convert the MGRS coordinates to latitude and longitude coordinates
6	Remove duplicate entries

Creating the Combined Database

After processing and cleaning the individual data files, the four files containing the ground combat incidents—the VNDBA, SITRA, TIRSA, and VCIIA Files—were combined into a single database. The creation of this combined database required several additional processing steps. First, the column names of each of the source files were standardized so that they shared a single naming convention.[11] Second, the entries for these actors were standardized, since the same actors were referenced inconsistently.[12] Third, a column was created to identify the name of the original data set so that researchers could refer back to the source file. Fourth, incidents were merged into a single file. Then duplicate entries were removed when it became apparent that thousands of incidents had redundant entries, since they shared identical actors, locations, and start times. Finally, a unique ID value was created for the combined database to identify each column. The original files contain additional information not combined in the master data file, so this is not a perfect replacement for the original files.

Table 4: Steps to Create the Combined Database

Step	Description
1	Column names were standardized
2	Actor labels were standardized
3	A column was added to identify the original data file
4	Incidents were merged
5	Duplicate entries (that appeared in multiple source files) were removed
6	A unique incident ID was created for each of the remaining observations

[11]For example, columns containing the specific identity of the aggressor force were converted to "aggressor_sub_group.

[12]For example, the VC were referred to as guerrillas or "local forces" in several of the files.

A Quantitative Overview of the Vietnam War

The combined ground combat files provide the basis for a quantitative overview of the Vietnam War. The following section uses the combined ground combat data to summarize spatial and temporal variations in the intensity of ground combat in Southeast Asia.[13] The quantitative analysis also provides a better sense of the information in the data files, a small window into the types of research questions that might be tested, and some concerns and observations about the underlying data. This analysis consists of three parts. The first is a temporal analysis of the intensity of the conflict that presents time-series plots of the number of Allied operations and ground attacks in North Vietnam that were launched daily. The second presents density maps that identify the intensity of Allied and North Vietnamese attacks across Southeast Asia on an annual basis from 1963 to 1973. The third and final section provides a micro-level overview of actions taken during the Tet Offensive around Hue and Saigon from January through April 1968, to showcase the granularity of the data.

Figure 1: Daily Incident Counts (1963 - 1973)

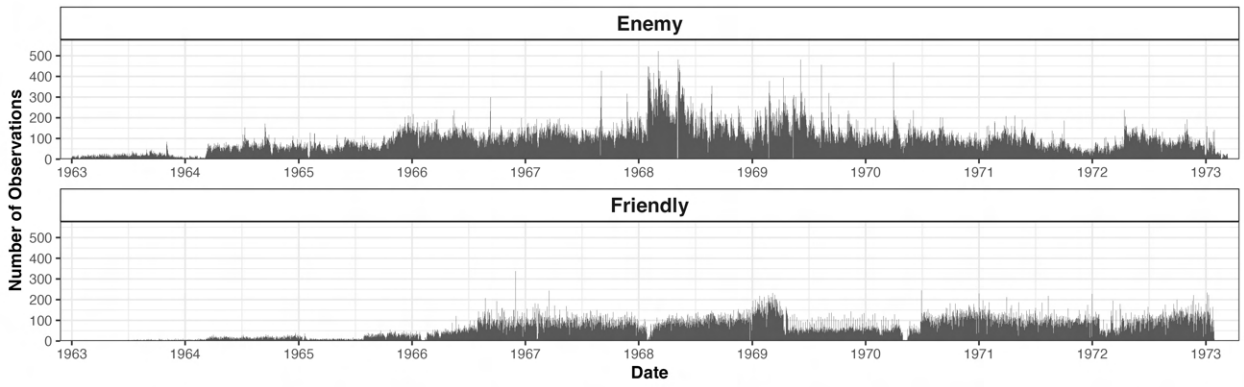


Figure 1 presents the number of North Vietnamese attacks and Allied ground operations initiated daily. The data reveal significant variation in attack frequency over the decade, with both enemy and friendly incidents generally increasing from 1964 onward, and these attacks tend to be correlated. The first three years in the sample correspond to a significant increase in the US force posture and a corresponding rise in violence. In 1963, the year that Ngo Dinh Diem was

[13]This limits the analysis to the four combined data files from the VNDBA, SITRA, VCIIA, and TIRSA files. An additional overview of each file is presented in the Appendix.

assassinated by his generals as part of a coup, the United States had approximately 16,000 military personnel in Southeast Asia (U.S. Army Center of Military History, 2023). This number was increased to approximately 23,000 in 1964 after the Gulf of Tonkin Resolution and the supposed attacks on the US destroyers Maddox and Turner Joy on August 2 and August 4. In March of 1965, US ground units landed in Da Nang, and over 184,000 military personnel were stationed in the theater by the year's end. Much of the heaviest fighting occurred from 1966 to 1969 when President Nixon started the withdrawal of US troops. The significant decrease in Allied operations corresponds to the reduction in US troops and the shift in strategies to Vietnamization, which sought to reduce American involvement in the Vietnam War by transferring all military responsibilities to South Vietnam. The prominent spikes, particularly in NVA and VC attacks, correspond to major operations such as the Tet Offensive in 1968, the massive surprise attack initiated across South Vietnam, and the Easter Offensive in 1972, which was an enormous offensive intended to improve North Vietnam's bargaining position in the ongoing Paris peace negotiations.

A prominent trend revealed in Figure 1 is that the number of enemy attacks was consistently higher than the number of Allied operations for the entire sample period. A major source of this discrepancy in incident counts was driven by each side's different approach to fighting the war. Allied forces fought a large conventional war effort, and their primary objective on the battlefield in South Vietnam was to destroy communist military forces by "grinding them down" in a war of attrition (Thayer, 1975, pp. 89). At least in the case of the US army, their approach was dominated largely by conventional operations called "search and destroy missions" intended to locate and eliminate enemy forces. The data files distinguish friendly small unit actions, which involved "less than three companies of line troops but not less than a fire team," from large unit operations, which involved "three or more companies of line troops" (National Archives and Records Administration, 2005, pp 11). Major operations generally refer to major search and destroy missions, such as Operation Cedar Falls and Operation Junction City (Hickey, nd). In contrast, small-unit operations focused on efforts to patrol or secure a small area. The Marine Combined Action Program (CAP) in the northernmost military region of South Vietnam provided a particularly famous example of these operations (Johnson, 2018).[14] The CAP program was intended to improve security in rural

[14]These distinctions preserve essential qualitative differences between ground operations. Major operations involved

areas by integrating American marines with local Vietnamese militia forces and to protect local villagers from VC attacks. Perhaps unsurprisingly, there were more minor operations than major ones.

For Communist forces, the main goal on the battlefield was to avoid critical losses while simultaneously inflicting enough punishment on Allied forces to prompt an American withdrawal. Following heavy losses and invaluable lessons learned at Ia Drang in 1965, the North Vietnamese adopted guerrilla tactics against the United States and its allies. In practical terms, the northern strategy attempted to prompt the US withdrawal via a death by a thousand paper cuts. Rather than confronting Allied ground forces in major head-to-head battles, northern forces perfected guerrilla tactics that allowed them to rapidly engage and withdraw while minimizing the loss of VC cadre members. In addition to their attacks on Allied military personnel and installations, northern forces coerced local civilians with terrorist attacks, eliminated local officials and leaders who opposed their actions, and had extensive propaganda efforts. These incidents were generally smaller and less intensive to initiate, but were more numerous (Thayer, 1975). Thayer notes that more than 95 percent of Communist ground assaults initiated between 1965 and 1972 were conducted by units smaller than a battalion, and consisted primarily of harassment, standoff attacks, and terrorism. The small unit size and limited tactical goals of most Communist actions are major reasons why the number of enemy attacks initiated daily consistently exceeded the number of Allied operations.

The data files categorized VC and NVA attacks into several main incident categories, the most common being direct attacks, harassing fires, mining operations, and terrorist attacks. Direct attacks were primarily ground assaults by smaller North Vietnamese units on Allied positions. Harassment operations involved stand-off attacks using indirect fire from various weapons. Mining tactics used explosive devices in areas frequented by US and ARVN troops. Terrorist attacks targeted civilians and officials through acts such as assassinations, abductions, and public bombings.

the coordination of multiple battalions under complex command structures. These operations integrated combined arms, such as artillery, air support, armor, and infantry units, with extensive logistic support, and were sustained for extended durations, lasting weeks to months. In contrast, minor operations were confined to the purely tactical level and were typically led at the squad, platoon, or company level. These operations had limited objectives, such as conducting reconnaissance, setting ambushes, or providing local security, and were usually short-duration, lasting only hours to a few days.

The combined files reveal a total of 388,238 enemy incidents, with harassment being the most common (125,888 cases), followed by terrorism (54,867), mining (30,557), direct attacks (17,137), sabotage (9,589), entry (3,685), and propaganda (3,097). In comparison, 143,418 enemy incident categories were not listed.[15]

[15]For a more granular comparison, see Figure 2 in the Appendix for a daily count of enemy incident counts grouped by attack type.

Figure 2: Heat Map of VC and NVA Attacks (1963 - 1976)

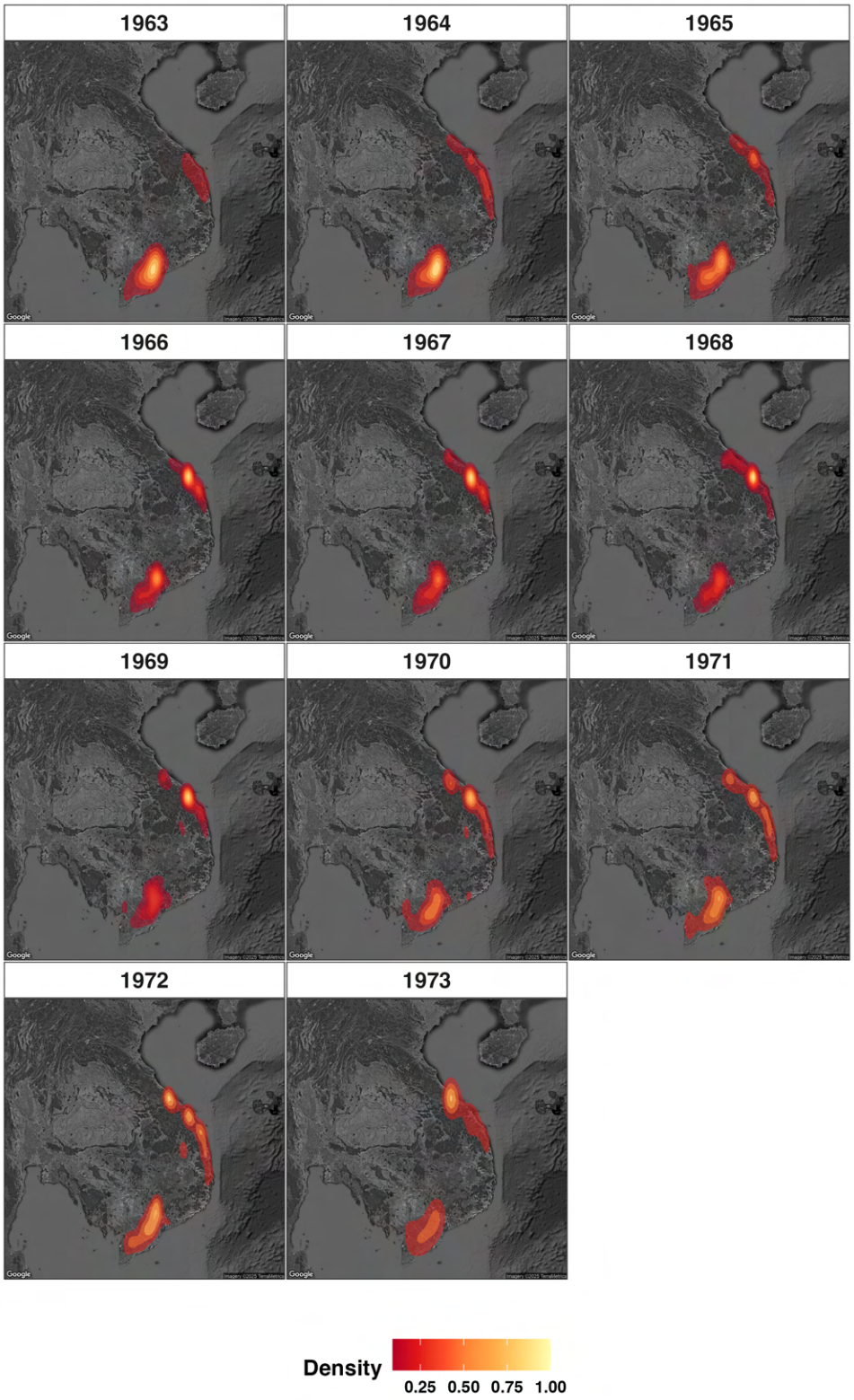
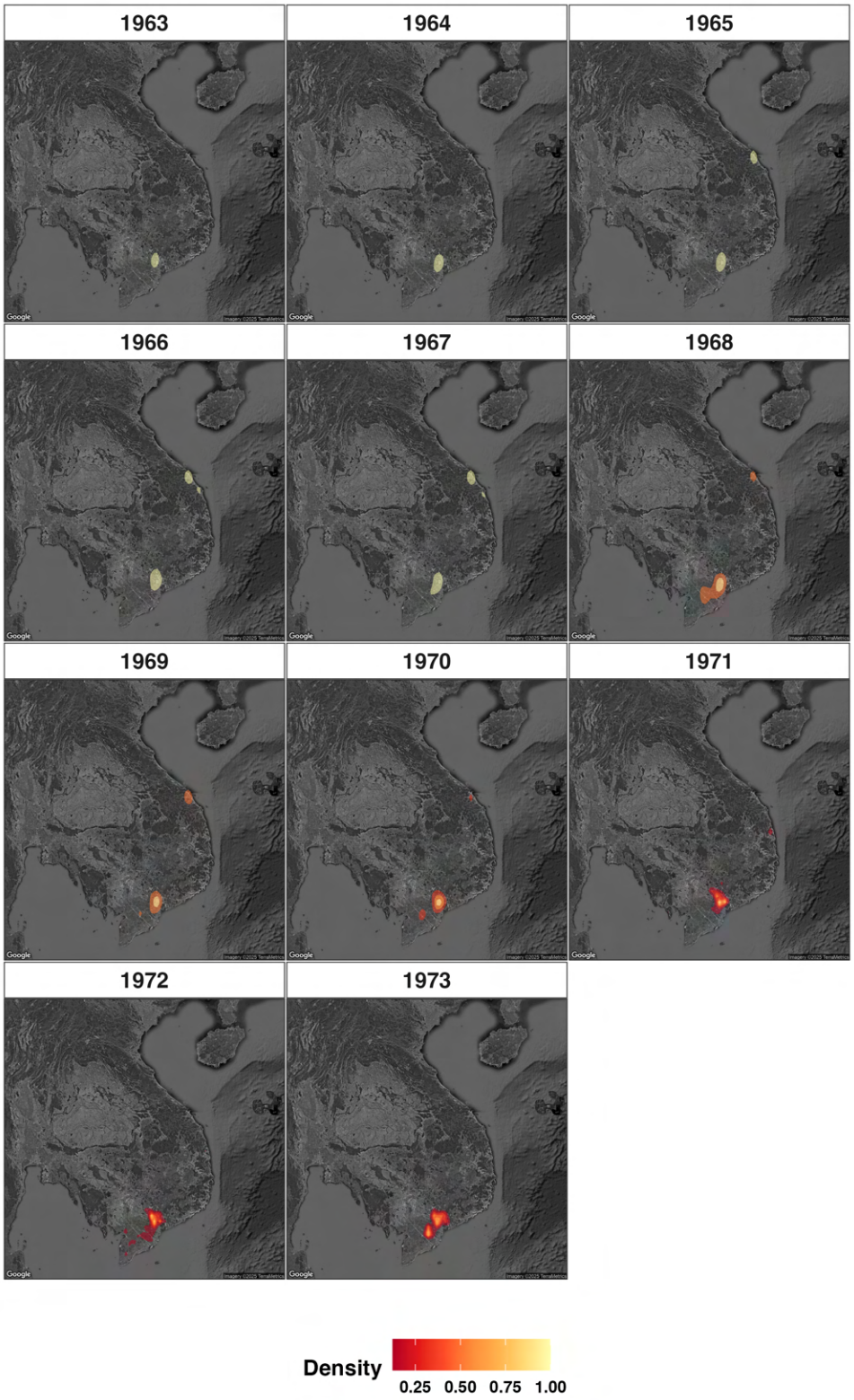


Figure 3: Heat Map of Allied Operations (1963 - 1976)



Figures 2 and 3 present the density estimates of the concentration of North Vietnamese and Allied attacks across Southeast Asia, respectively, annually from 1963 to 1973. These plots identify how concentrated the attacks were at a particular location on the grid. In this case, each facet reveals patterns for a particular year. The density scale ranges from 0.25 (red) to 1.00 (yellow), with brighter colors indicating higher concentrations of combat incidents. The most intense fighting appears as bright yellow-white spots on the map, while lower-intensity areas are shown in darker red shades. Note that the absence of color in some regions of these heat maps does not necessarily mean that there was no combat activity there. The minimum density threshold of 0.25 means that areas with combat activity below this threshold are not displayed. Consequently, combat incidents that were more sporadic or dispersed would fall below this threshold and thus appear as gray areas on the map. For example, the density maps do not highlight the First Cavalry Division's well-documented airmobile operations in the Central Highlands. The Appendix includes plots where each incident coordinate is overlaid on a map of South Vietnam. The attack coordinates are so numerous and spread throughout South Vietnam that it is challenging to determine the underlying characteristics of the terrain and the relative intensity of the fight.

The heat maps of the VC and NVA attacks reveal several key geographic distributions and temporal patterns that emerge over time. Throughout the war, two distinct combat hot-spots are consistently visible: a northern concentration in I Corps near the DMZ (Demilitarized Zone) and a southern concentration around the Mekong Delta region north through the south of Saigon. The intensity of combat evolved significantly over the decade, with the early period (1963-1965) showing that combat was initially concentrated in the south and moderate activity near the DMZ. During the mid-war period (1966-1969), the intensity of combat increased substantially in both regions, with particularly intense fighting along the DMZ and sustained activity in the Mekong Delta. In the later period (1970-1973), the pattern shows a gradual shift in intensity, with combat remaining significant but becoming more dispersed, particularly in the northern regions.

The heat maps of the Allied combat operations reveal stark differences with the North Vietnamese attack patterns. The most noticeable difference is that Allied ground combat operations tended to be focused around Saigon. This pattern changes later in the war after US ground forces

arrive in force, and a second area of dense operations appears in the northern part of South Vietnam, which is close to the demilitarized zone. Relatedly, Allied operations were far more concentrated geographically. Compared to the North Vietnamese-initiated incident map, the areas in operation are far more focused on particular points of interest and tend to be similar in intensity, with the density values close to the maximum value of one. In the map of enemy-initiated incidents, density estimates are widely dispersed and have lower intensity levels. The broader spread of events across South Vietnam is likely driven in part by the nature of VC attacks that targeted civilians with terrorist attacks, in addition to Allied military personnel.

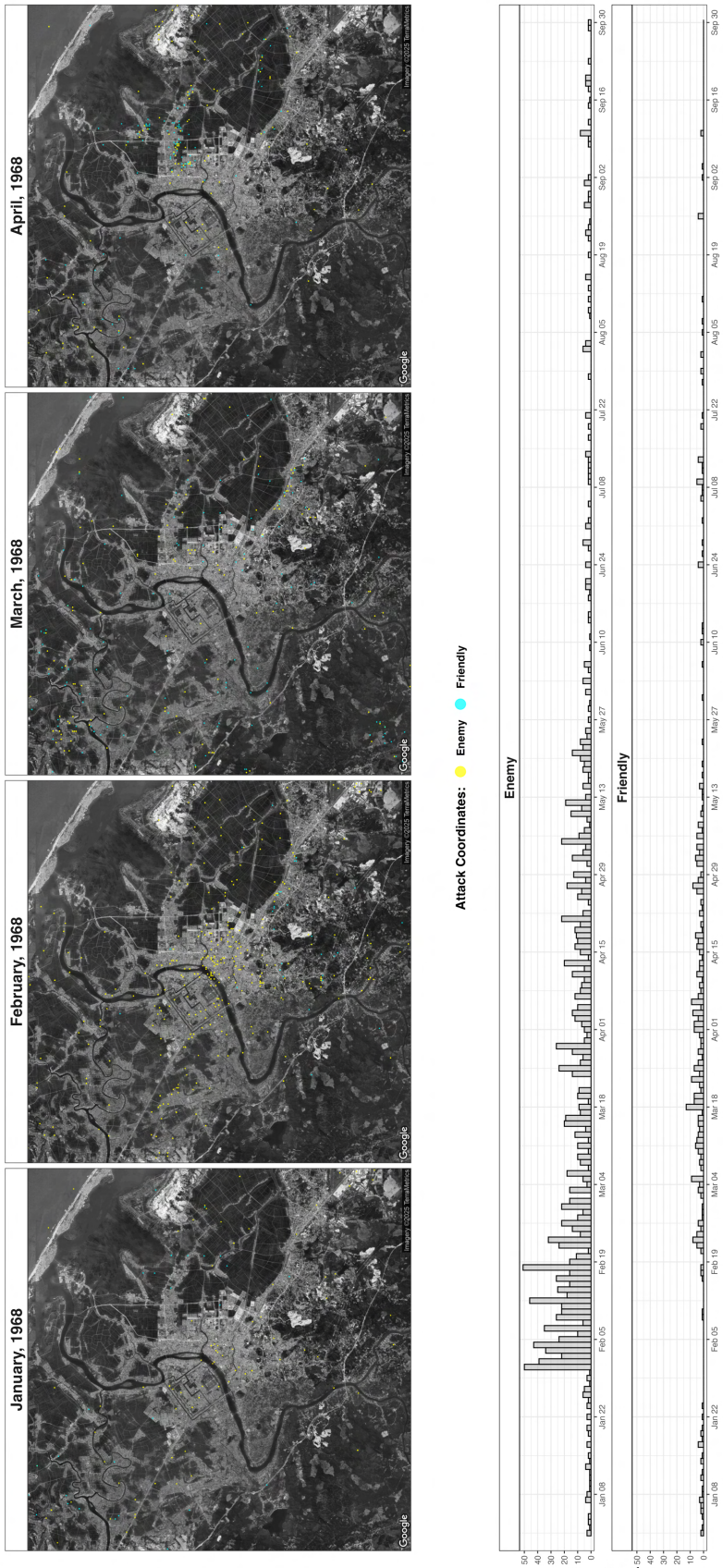


Figure 4: Incidents Near Hue During the Tet Offensive

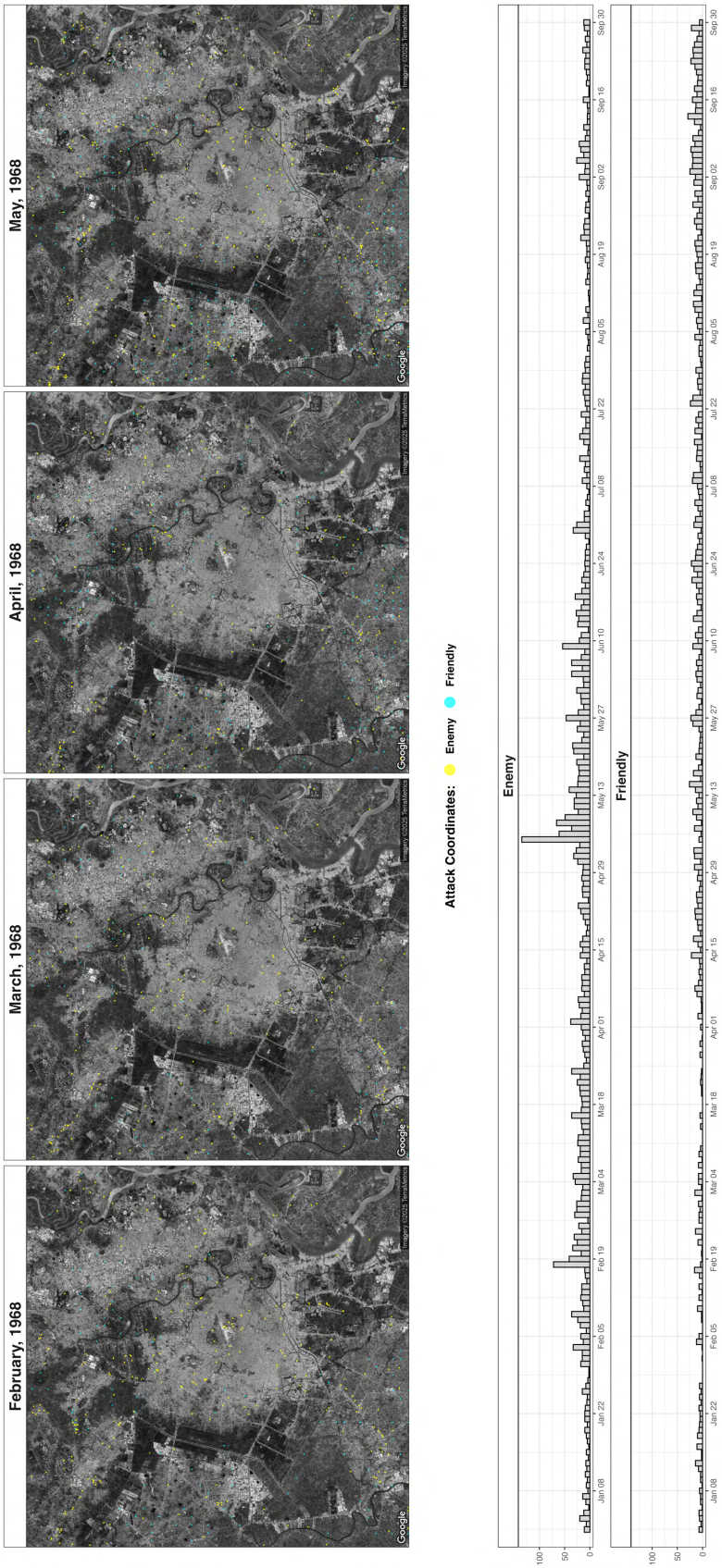


Figure 5: Incidents Near Saigon During the Tet Offensive

The final part of the quantitative review presents a micro-level examination of the Tet Offensive in the areas surrounding the cities of Hue and Saigon from January to April 1968. The Tet Offensive was a major hybrid military campaign launched by North Vietnamese and Viet Cong forces against South Vietnam, the United States, and their allies during the Vietnam War. This large-scale series of coordinated attacks began on January 30, 1968. They caught Allied forces by surprise because of their timing, which coincided with the Lunar New Year (Tet) holiday, as a large proportion of ARVN personnel were on leave. The scale of this attack was unprecedented and targeted civilian and military locations throughout South Vietnam. The Tet Offensive was selected for its historical significance. Many credit the shocking nature of this offensive with causing President Johnson not to seek reelection and US policymakers to withdraw military personnel from South Vietnam. The decision to visualize events around Hue and Saigon was made because of the heavy fighting in those areas and the political importance of those cities. Although more readers will likely be aware of Saigon, since it was the capital of South Vietnam, Hue was a major cultural center located close to the demilitarized zone at the northern end of South Vietnam. The Battle of Hue, which raged from January 31 to March 2, 1968, was one of the major battles of the Tet Offensive.

This visualization of attacks around Hue and Saigon is presented in Figures 4 and 5, respectively. Each figure presents the coordinates of each Allied and North Vietnamese attack overlaid on satellite maps of Hue and Saigon, as well as a daily time series of the number of enemy and Allied attacks initiated within the region displayed on each map. These time series plots span from January through September 1968, and the upper plot shows the frequency of enemy attacks while the lower plot displays Allied operations. The attack coordinates are presented using a two-color system: yellow dots mark enemy positions, whereas blue dots identify the position of friendly forces. Each satellite map represents one month from January to April 1968, with the earliest month located on the farthest left and the latest month on the farthest right.

At the most general level, Figures 4 and 5 reveal that the number of enemy-initiated incidents far exceeds the number of Allied operations in both Hue and Saigon. Despite this very general similarity, the plots reveal a variety of differences in the way the Tet Offensive played out over time. The time series plots demonstrate that the significant escalation of attacks occurred earlier

in Hue than in Saigon. The highest level of conflict intensity occurred around the Tet holiday and in the immediate days following, and steadily dissipated over time. The elevated period of fighting, relative to the pre-Tet period, persisted through May. In Saigon, the first significant increase in North Vietnamese attacks occurred on February 18, 1968. Attacks persisted steadily until early May 1968, when there was a major escalation in violence on May 5, 1968, when 135 attacks were initiated. In aggregate terms, far more attacks were initiated near Saigon than near Hue—a fact partially obscured by the fact that the y-axis units use different scales. 4,209 attacks were initiated near Saigon, and 1,646 attacks were initiated near Hue. In the case of Allied ground operations, the number of operations initiated daily remained roughly constant in both cities after the start of the Tet Offensive.

In addition to the different temporal patterns, the figures also reveal differences in the location and density of attack locations. The attacks in Hue were concentrated around the central area and the Perfume River, which runs through the city. The heavy concentration of attacks is consistent with the carefully orchestrated attacks that made the Battle of Hue one of the bloodiest of the war. In contrast, Saigon exhibits a more dispersed attack pattern. Attacks were widely launched in central urban areas and the surrounding countryside. The satellite maps also reveal a disparity in the frequency and locations of Allied ground operations. There appear to be far more enemy incidents than Allied operations in Hue, which are relatively sparse. The difference between Allied and North Vietnamese ground operations appears less stark in Saigon, though still dominated by enemy forces. Allied operations are spread more widely throughout the region and often seem to be clustered together. This observation is perhaps unsurprising since Saigon was the capital of South Vietnam and the location of the MACV headquarters.

Finally, Figures 4 and 5 provide the basis for several interesting observations about the differences between VC and Allied attacks. Enemy incidents conform to various topographical features, such as rivers and roads, whereas Allied operations do not follow the same pattern. The most obvious example of this appears in the enemy attacks initiated along Highway 4, which is located south of Saigon (in the bottom left quadrant). These attacks were most visible in February and March 1968. Unlike enemy operations, Allied operations frequently appear in evenly spaced grids. The

most obvious example of this grid pattern appears in April and May 1968 in the countryside south of Saigon. This pattern raises important questions about how Allied operations were recorded in the SITRA and VNDBA files, such as why this pattern appears sporadically in some places and not others, and whether this pattern betrays additional differences in how specific incidents were recorded.

Conclusion

This paper contributes to conflict scholarship by cleaning, releasing, and documenting Vietnam War ground combat data. This work addresses long-standing data accessibility issues that have limited researchers' ability to systematically study ground combat, in contrast to the air war, which has been well documented in the THOR database. This data provides a valuable foundation for replicating existing studies and new theories about conflict dynamics, insurgent tactics, and military effectiveness, among many other topics. In addition, these files help capture the details and legacy of one of the most devastating episodes in Southeast Asian and American history. By combining these quantitative data with qualitative records such as personal accounts, interviews, and historical narratives, researchers can create a holistic and nuanced way of preserving the war's memory and evaluating its long-lasting political implications.

References

Allen, S. H. and C. Martinez Machain (2018). Choosing air strikes. *Journal of Global Security Studies 3*(2), 150–162.

Allen, S. H. and C. Martinez Machain (2019). Understanding the impact of air power. *Conflict management and peace science 36*(5), 545–558.

Avant, D. D. (2019). *Political institutions and military change: Lessons from peripheral wars*. Cornell University Press.

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

Barthelme, P., E. Darbyshire, D. Spracklen, and G. R. Watmough (2024). Detecting vietnam war bomb craters in declassified historical kh-9 satellite imagery. *Science of Remote Sensing*, 100143.

Berman, E., J. H. Felter, and J. Shapiro (2018). *Small wars, big data: the information revolution in modern conflict*. Princeton University Press.

Buhaug, H., S. Gates, and P. Lujala (2009). Geography, rebel capability, and the duration of civil conflict. *Journal of Conflict Resolution 53*(4), 544–569.

Buhaug, H. and P. Lujala (2005). Accounting for scale: Measuring geography in quantitative studies of civil war. *Political Geography 24*(4), 399–418.

Carter, D. B., A. C. Shaver, and A. L. Wright (2019). Places to hide: Terrain, ethnicity, and civil conflict. *The Journal of Politics 81*(4), 1446–1465.

Charnysh, V. (2019). Diversity, institutions, and economic outcomes: Post-wwii displacement in poland. *American Political Science Review 113*(2), 423–441.

Charnysh, V., E. Finkel, and S. Gehlbach (2023). Historical political economy: Past, present, and future. *Annual Review of Political Science 26*(1), 175–191.

Cirone, A. and T. B. Pepinsky (2022). Historical persistence. *Annual Review of Political Science 25*(1), 241–259.

Condra, L. N., J. D. Long, A. C. Shaver, and A. L. Wright (2018). The logic of insurgent electoral violence. *American Economic Review 108*(11), 3199–3231.

Condra, L. N. and A. L. Wright (2019). Civilians, control, and collaboration during civil conflict. *International Studies Quarterly 63*(4), 897–907.

Connable, B. (2012). *Embracing the fog of war: Assessment and metrics in counterinsurgency*. Rand Corporation.

Daddis, G. A. (2011). *No Sure Victory: Measuring US Army Effectiveness and Progress in the Vietnam War*. Oxford University Press.

De la Calle, L. and I. Sánchez-Cuenca (2015). How armed groups fight: Territorial control and violent tactics. *Studies in Conflict & Terrorism 38*(10), 795–813.

Dell, M. and P. Querubin (2017). Nation building through foreign intervention: Evidence from discontinuities in military strategies. *The Quarterly Journal of Economics 133*(2), 701–764.

Dinas, E., V. Fouka, and A. Schläpfer (2021). Family history and attitudes toward out-groups: Evidence from the european refugee crisis. *The journal of politics 83*(2), 647–661.

Fearon, J. D. and D. D. Laitin (2003). Ethnicity, insurgency, and civil war. *American political science review 97*(1), 75–90.

Gartner, S. S. (1999). *Strategic Assessment in War*. Yale University Press.

Guardado, J. and S. Pennings (2025). The seasonality of conflict. *Conflict Management and Peace Science 42*(1), 56–81.

Hammond, J. (2018). Maps of mayhem: Strategic location and deadly violence in civil war. *Journal of Peace Research 55*(1), 32–46.

Harrison, D. F. (1988). Computers, electronic data, and the vietnam war. *Archivaria*, 18–32.

Hickey, L. J. (n.d.). Operation junction city. Technical report, HQ PACAF, Directorate, Tactical Evaluation, CHECO Division, Christiansburg, VA. Prepared by S.E. Asia Team.

High, H., J. R. Curran, and G. Robinson (2013). Electronic records of the air war over southeast asia: A database analysis. *Journal of Vietnamese Studies 8*(4), 86–124.

Homola, J., M. M. Pereira, and M. Tavits (2020). Legacies of the third reich: Concentration camps and out-group intolerance. *American Political Science Review 114*(2), 573–590.

Horowitz, M. and D. Reiter (2001). When does aerial bombing work? quantitative empirical tests, 1917-1999. *Journal of Conflict Resolution 45*(2), 147–173.

Ito, G. and K. Hinkkainen Elliott (2020). Battle diffusion matters: Examining the impact of microdynamics of fighting on conflict termination. *Journal of Conflict Resolution 64*(5), 871–902.

Johnson, J. L. (2018). *The Marines, Counterinsurgency, and Strategic Culture: Lessons Learned and Lost in America's Wars*. Georgetown University Press.

Kalyvas, S. N. (2006). *The Logic of Violence in Civil War*. Cambridge University Press.

Kalyvas, S. N. and M. A. Kocher (2009). The dynamics of violence in vietnam: An analysis of the hamlet evaluation system (hes). *Journal of Peace Research 46*(3), 335–355.

Kocher, M. A., T. B. Pepinsky, and S. N. Kalyvas (2011). Aerial bombing and counterinsurgency in the vietnam war. *American Journal of Political Science 55*(2), 201–218.

Kwon, H. (2006). *After the massacre: Commemoration and consolation in Ha My and My Lai*, Volume 14. Univ of California Press.

Kwon, H. (2008). *Ghosts of war in Vietnam*. Cambridge University Press.

LaFree, G. and L. Dugan (2007). Introducing the global terrorism database. *Terrorism and political violence 19*(2), 181–204.

Lin, E. (2022). How war changes land: Soil fertility, unexploded bombs, and the underdevelopment of cambodia. *American Journal of Political Science 66*(1), 222–237.

Lin, E. (2024). When the bombs stopped: The legacy of war in rural cambodia.

Long, A. (2016). *The Soul of Armies: Counterinsurgency Doctrine and Military Culture in the US and UK*. Cornell University Press.

Martin, B., B. Clayton, J. Welch, S. J. Bae, Y. Kim, I. Khan, and N. Edenfield (2020, April). Naval surface fire support. Technical report, RAND Corporation.

Menninga, E. J. and A. K. Prorok (2021). Battles and bargains: escalation, commitment, and negotiations in civil war. *International studies quarterly 65*(2), 406–422.

Min, E. (2020). Talking while fighting: Understanding the role of wartime negotiation. *International organization 74*(3), 610–632.

Min, E. (2022). Painful words: The effect of battlefield activity on conflict negotiation behavior. *Journal of Conflict Resolution 66*(4-5), 595–622.

National Archives and Records Administration (2005, October). Supplemental user note: Situation reports army (sitra) files, may 1966 - march 1973. User Note NN3-218-76-032, National Archives and Records Administration, College Park, Maryland. Prepared by J. Powell, Archives Specialist.

National Military Command System Support Center (1969, 9). *The Operation Analysis System (OPSANAL) User's Manual* (Revision A ed.). National Military Command System Support Center.

Pepinsky, T. B., S. W. Goodman, and C. Ziller (2024). Modeling spatial heterogeneity and historical persistence: Nazi concentration camps and contemporary intolerance. *American Political Science Review 118*(1), 519–528.

Pierce, J. H., G. Carter, J. Reiley, A. Peterson, and J. W. Ellis (1976). User's guide to southeast asia combat data.

Pool, I. d. S., G. Fowler, P. McGrath, and R. Peterson (1968). Hamlet evaluation system study.

Post, A. (2019). Flying to fail: Costly signals and air power in crisis bargaining. *Journal of Conflict Resolution 63*(4), 869–895.

Raleigh, C., A. Linke, H. Hegre, and J. Karlsen (2010). Introducing acled: An armed conflict location and event dataset. *Journal of Peace Research 47*(5), 651–660.

Roberts, M. E. (2018). *The Psychological War for Vietnam, 1960-1968*. University Press of Kansas.

Robertson, J. A., B. and B. Barth (2013). USAF THOR Database. Data retrieved from `http://www.au.af.mil/au/afri/thor/index.asp#.XGsKny2ZOV4`.

Schutte, S. and N. B. Weidmann (2011). Diffusion patterns of violence in civil wars. *Political Geography 30*(3), 143–152.

Shaver, A., D. B. Carter, and T. W. Shawa (2019). Terrain ruggedness and land cover: Improved data for most research designs. *Conflict Management and Peace Science 36*(2), 191–218.

Stewart, M. A. (2018). Civil war as state-making: Strategic governance in civil war. *International Organization 72*(1), 205–226.

Sundberg, R. and E. Melander (2013). Introducing the ucdp georeferenced event dataset. *Journal of Peace Research 50*(4), 523–532.

Tang, T. N., D. Van Toai, and D. Chanoff (2018). *A Vietcong memoir: An inside account of the Vietnam War and its aftermath*. Tantor Media, Incorporated.

Tate, G. W. (1978). Behavioral effects of officer evaluation in the us army.

Teodoro, M. P. (2011). *Bureaucratic ambition: Careers, motives, and the innovative administrator*. JHU Press.

Thayer, T. C. (1975). A systems analysis view of the vietnam war 1965-1972. volumes 1 - 12. population security. Technical report, Assistant Secretary Of Defense (Systems Analysis) Washington DC.

Tollefsen, A. F., H. Strand, and H. Buhaug (2012). Prio-grid: A unified spatial data structure. *Journal of Peace Research 49*(2), 363–374.

U.S. Army Center of Military History (2023). Vietnam war campaigns.

Vogt, M., N.-C. Bormann, S. Rüegger, L.-E. Cederman, P. Hunziker, and L. Girardin (2015). Integrating data on ethnicity, geography, and conflict: The ethnic power relations data set family. *Journal of Conflict Resolution 59*(7), 1327–1342.

Walden, J. and Y. M. Zhukov (2020). Historical legacies of political violence. *Oxford research encyclopedia of politics*, 1–66.

Wayne, C. and Y. M. Zhukov (2022). Never again: The holocaust and political legacies of genocide. *World Politics 74*(3), 367–404.

Zhukov, Y. M. (2012). Roads and the diffusion of insurgent violence: The logistics of conflict in russia's north caucasus. *Political Geography 31*(3), 144–156.

Zhukov, Y. M., C. Davenport, and N. Kostyuk (2019). Introducing xsub: A new portal for cross-national data on subnational violence. *Journal of peace research 56*(4), 604–614.

Zhukov, Y. M. and B. M. Stewart (2013). Choosing your neighbors: Networks of diffusion in international relations. *International Studies Quarterly 57*(2), 271–287.

Appendix: Beyond Body Counts

Contents

1 Introduction

This appendix accompanies *Quantifying the Conflict: Vietnam War Ground Combat Data*. It provides an overview of how the "master" incident-level data file was created, paying special attention to the deduplication procedure. In addition, it provides a descriptive summary of each cleaned data file. The complete list of cleaned files is available in Table 1 of the article.

2 Overview of the Combined Data

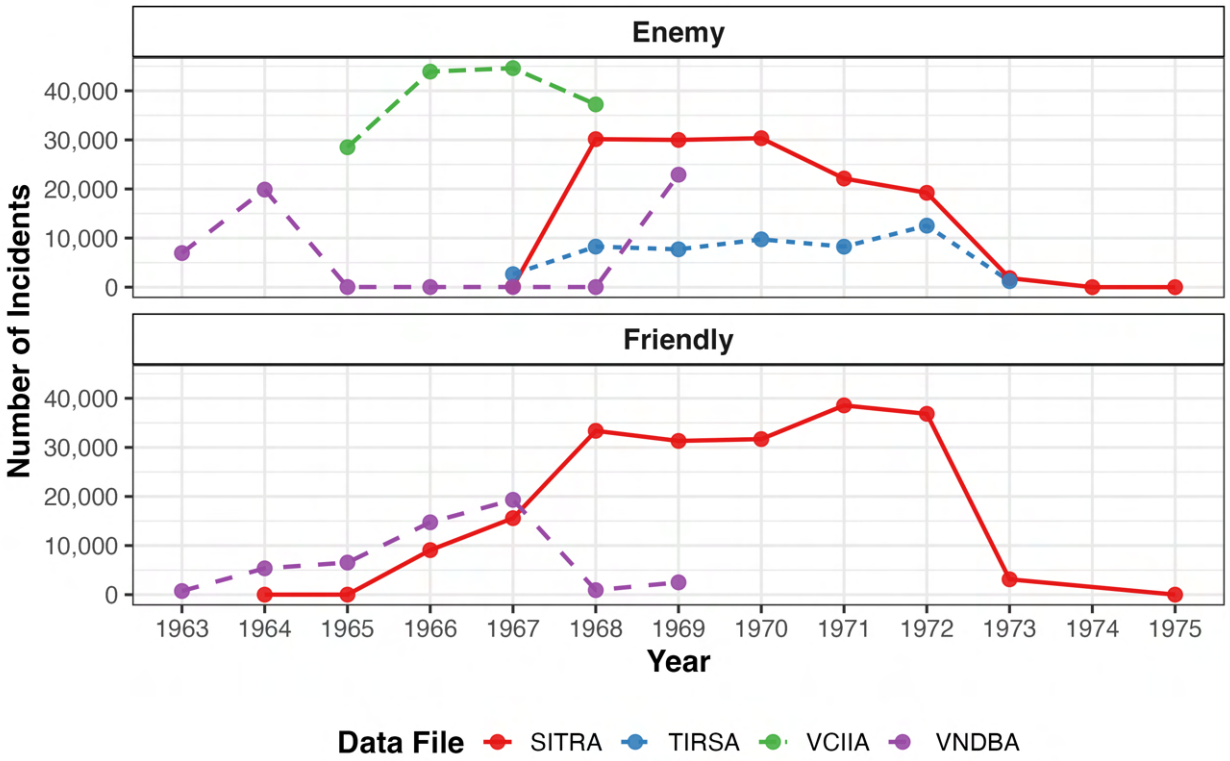
The following section details the incident-level combat data from the SITRA, TIRSA, VCIIA, and VNDBA files and discusses how they were combined and duplicated.

These data files are drawn from larger parent databases, the Operation Analysis System (OPSANAL), and from operational reports created by the Military Assistance Command Vietnam (MACV) (National Military Command System Support Center, 1969). The specific operational reports used to populate the data files were called OPREP-4 and OPREP-5 reports. The OPREP-4 reports contained brief narrative summaries of small-unit actions, large-unit operations, and enemy-initiated incidents. In contrast, the OPREP-5 report contained weekly statistical summaries of large unit operations, friendly attacks, enemy-initiated incidents, and personnel and weapons losses (National Archives and Records Administration, 2005).

The incident-level data files vary in their specific substantive focus. For example, the TIRSA and VCIIA files focus exclusively on attacks that were initiated by NVA forces and VC cadres. More specifically, the TIRSA files document acts of violence carried out against the civilian population of South Vietnam that were initiated by the Viet Cong and North Vietnamese forces from 1967 to 1973. Similarly, the VCIIA files document Viet Cong attacks on civilian and military targets in South Vietnam from 1965 to 1968. The VNDBA and SITRA files provide the most comprehensive overview of ground combat, as they contain information on both Allied and enemy-initiated incidents. The VNDBA files were taken from the OPSANAL system and include information dating back to the start of major US military involvement. In contrast, the SITRA files were constructed from operations reports focused on the later period of US involvement from 1966 to 1973. Although the periods contained in the VNDBA and SITRA files overlap, the scope of the VNDBA files decreases over time, and the SITRA files appear to be more comprehensive starting in 1967.

One of the primary challenges with creating a combined ground database was that incidents were duplicated across original source files. Although it is unclear why some periods have more duplicate values than others, the problem almost certainly derives from the fact that these records relied upon the duplicate source files and had slightly different objectives and target audiences. Figure S1 presents the count of incidents in each source file that are present in the combined file in a given year after duplicate events were removed. A more in-depth discussion of the deduplication process appears below.

Figure S1: The Number of Incidents From Each Source File



2.1 The Merge and Deduplication Procedure

The first step in the deduplication procedure was to rename column names to a consistent pattern across all source files. For example, VNDBA and SITRA files include columns to distinguish between friendly- and enemy-initiated incidents. In contrast, the TIRSA and VCIIA files include only enemy-initiated incidents and consequently lack columns to distinguish them. The column "aggressor side" was added to the TIRSA and VCIIA files to resolve this problem. Similarly, the files vary in how they store information about friendly and foreign soldiers who were wounded, killed, or missing in action. The loss records were standardized across the files so that each loss record now appears in a separate column.

After the files were standardized, they were merged into a data file that included each of the original events. Once the events were combined into a single file, they were de-duplicated based on their shared column values. Matches were identified and dropped using the columns: aggressor side, enemy action category, operation name, combined operation name, division designation, brigade designation, battalion ID, company, initiation date, initiation time,

latitude, and longitude. VNDBA records were removed from the combined file when they contained the same event ID, which was used to identify events in the source files. Following this procedure, a master event identification column was created that uniquely identifies each event.

The final stage of the data cleaning effort involved basic spatial processing to make it easier for researchers to analyze geographic patterns. Several columns were added to the cleaned data set. The first step was to determine whether an incident occurred outside South Vietnam. As a result of the escalation and expansion of the war into Cambodia, many events occurred outside of South Vietnam. This column allows scholars to identify such events easily. The second column identifies incidents that happened over the ocean. There is reasonable concern that some of these points may result from errors in how coordinates were recorded, and researchers may wish to omit these records from their analyses. Finally, the names of the provinces and military regions were assigned to events based on their geographical coordinates, providing valuable context for the spatial analysis of the conflict.

3 Problems With the Data Files

Researchers interested in using the data files should be aware of several challenges and issues in the original and the combined files. Researchers must decide how to address these concerns individually, as the answers will vary depending on the intended use case. Many of these challenges involve inconsistencies or errors in how data were coded, missing information such as event coordinates, questions about the consistency of coding practices, and challenges with grouping events (such as terrorist attacks vs. direct attacks) based on their listed characteristics.

A particularly obvious problem with the cleaned data files is that certain labels are missing for certain subsets of the data or are used inconsistently. For example, one of the general categories, which classifies Allied events as either "small unit" or "large operations", appears in the SITRA files but not in the VNDBA files. This limitation makes it challenging to compare trends and strategies across different eras of the war, particularly pre-/post-1967, because these basic distinctions are missing. Moreover, some classifications are clearly incorrect. For example, some incidents are listed in the source files as being initiated by enemy forces in the field that identifies the type of data/event, but are listed as being initiated by a small unit or large operations in the category field, which is only supposed to pertain to incidents initiated by Allied forces in the reference documents (National Archives and

Records Administration, 2005).

Table S1: Contradictory Aggressor Labels

Aggressor Side	**Data Type**	**Count**
Enemy	Friendly Small Unit	120
Enemy	Large Unit Operation	361
Friendly	Enemy Initiated Incident	365

Several additional issues are apparent in the records of enemy-initiated attacks. Events are grouped into action categories that identify specific incident types. These categories include harassment, terrorism, mining, attacks, sabotage, entry, and propaganda. The problem is that 143,421 of the 388,238 events are missing classifications (approximately 37%. Table S2 presents the total number of events present in each category. Although these events are missing, unformatted comments in several data files should allow researchers to identify the missing labels with some effort.

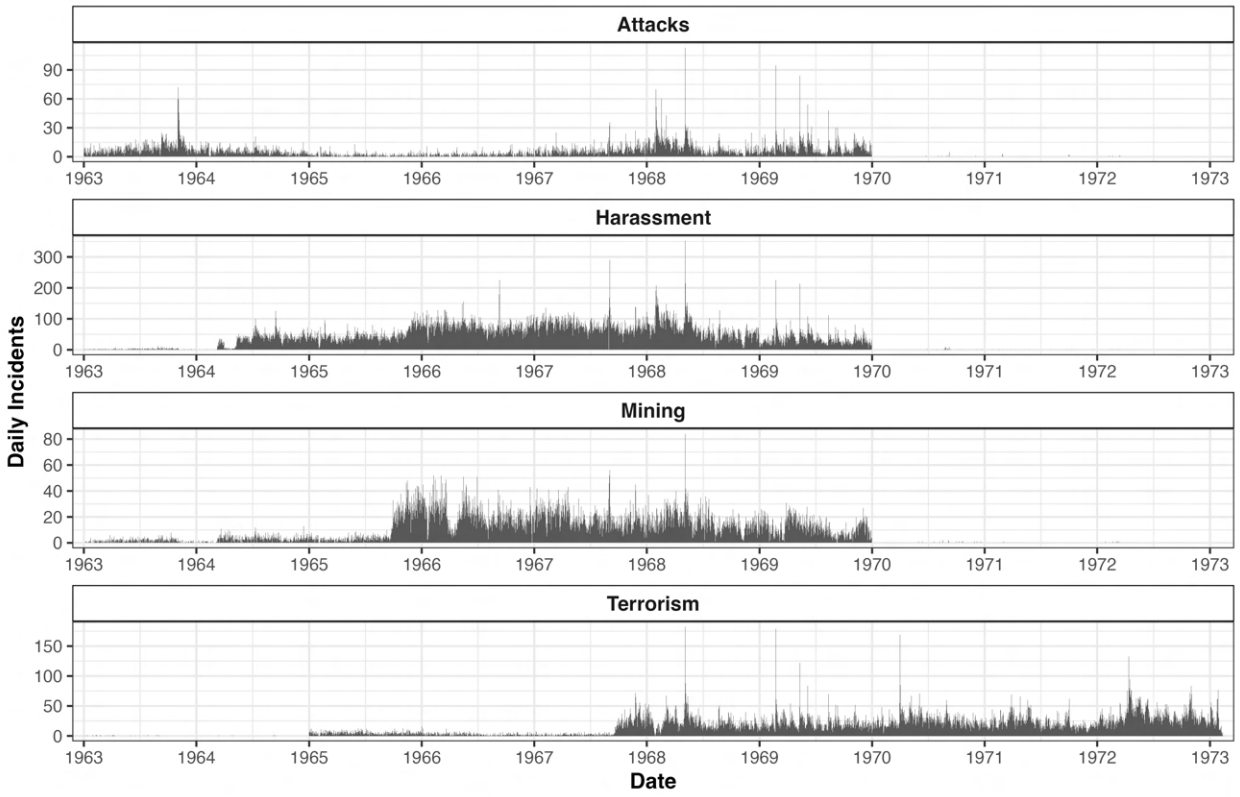
Table S2: Enemy Action Categories and Counts

Action Category	**Count**
Not Categorized	143,418
Harassment	125,888
Terrorism	54,867
Mining	30,557
Attacks	17,137
Sabotage	9,589
Entry	3,685
Propaganda	3,097

Similarly, the time series plot of daily attacks initiated, presented in Figure S2, raised questions about the consistency of coding practices over time. Although attacks show sporadic spikes throughout the period and have broad temporal coverage, harassment, terrorism, and mining appear to have disproportionate numbers of events in specific periods. In the case of harassment, the most heavily concentrated period is between 1964 and 1970. Likewise, mining incidents are concentrated between 1966 and 1970, and terrorism from 1968 to

1973. The fact that these periods do not perfectly overlap suggests that the coding strategies may have changed over time. These concerns could be alleviated if future research could add missing event categories.

Figure S2: The Number of Daily Enemy Incidents by Type



Finally, the plots of each coordinate recorded in the data files indicate that some coordinates are likely incorrect. For example, some coordinates locate incidents over the ocean or far away from the center of fighting in South Vietnam, Laos, or Cambodia. Maps that include every Allied operation and enemy incident coordinate appear below.

Figure S3: The Coordinates of Each Allied Ground Operation

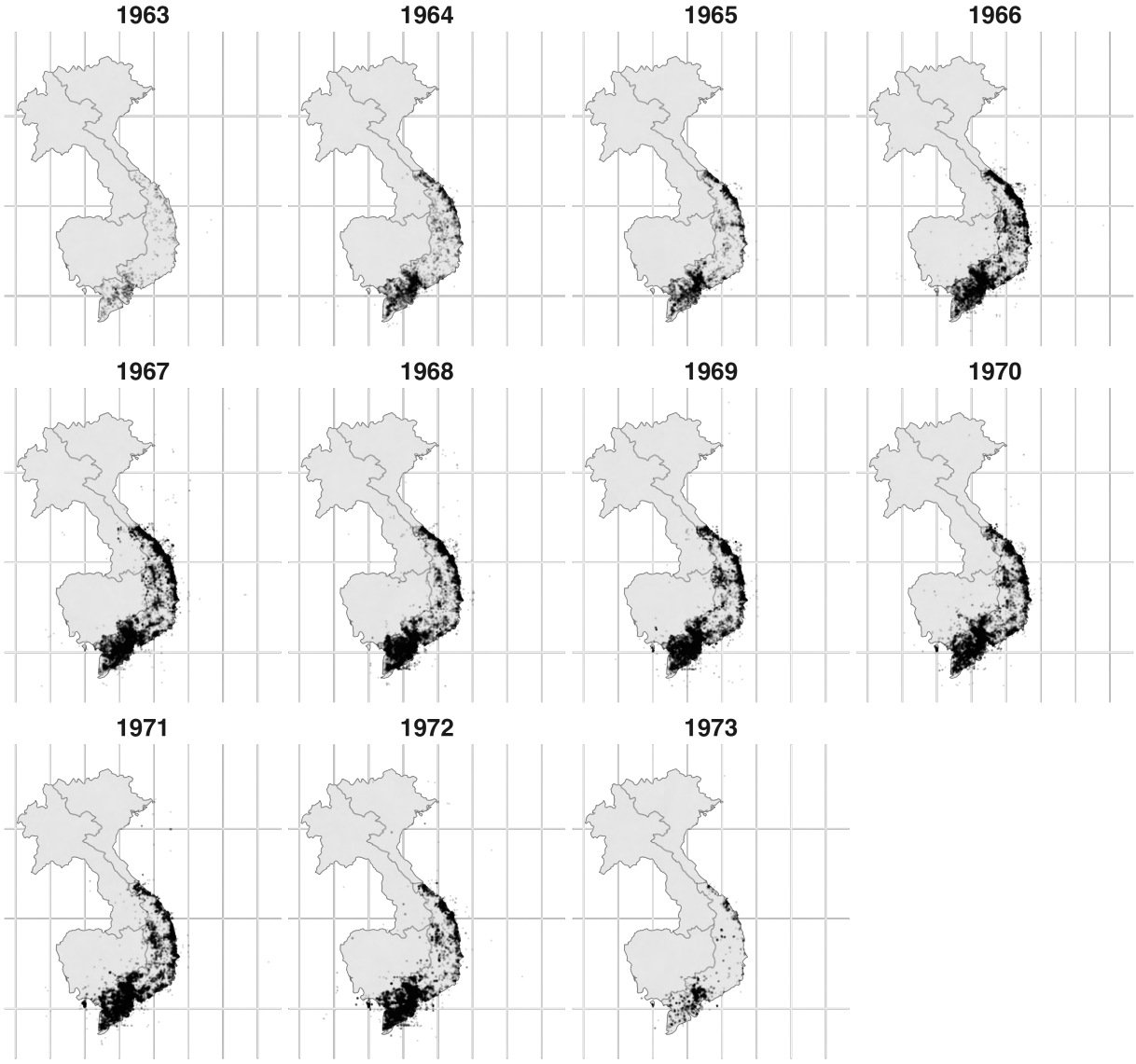
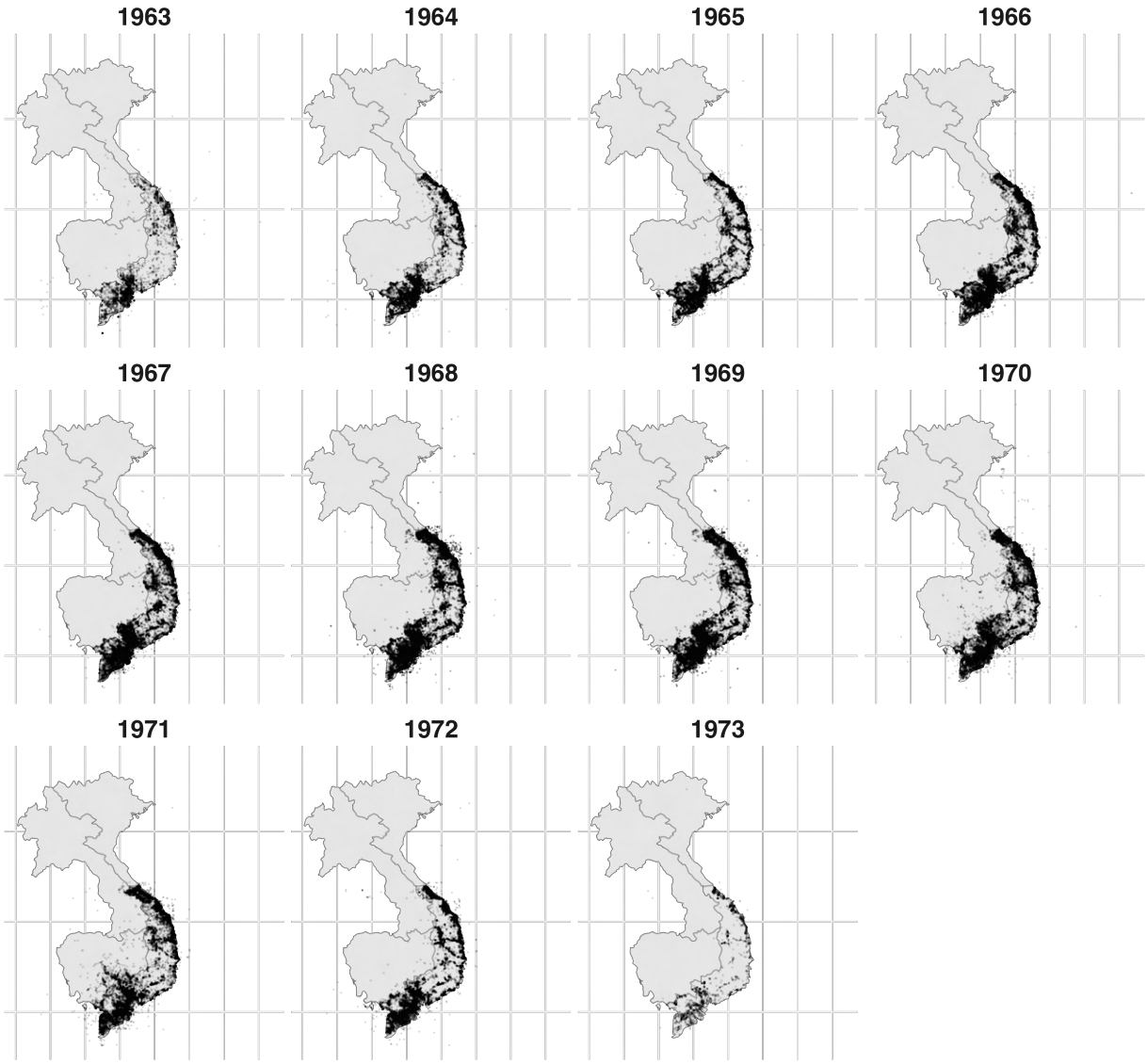


Figure S4: The Coordinates of Each Enemy Incident



4 Additional Descriptive Plots of the Combined Data

The following section provides a descriptive overview of the Allied, VC, and NVA-initiated operations in the combined data file.

4.1 An Overview of Allied Operations

The following figures provide counts of the most common data entries for records of friendly ground combat operations. These figures include counts of the most frequent aggressors, the most common operation names recorded, the types of operation designation, and the number of enemy casualties resulting from Allied operations.

Figure S5: The Number of Incidents Initiated by Each Sub-Group Actor

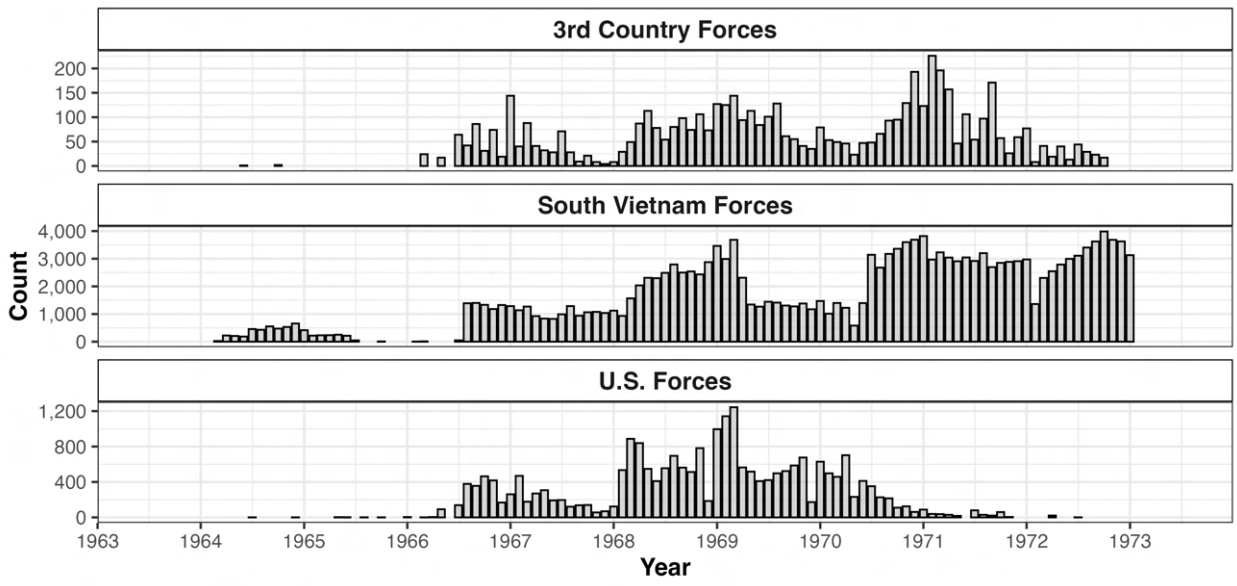


Figure S6: The Number of Incidents Initiated by Each Aggressor

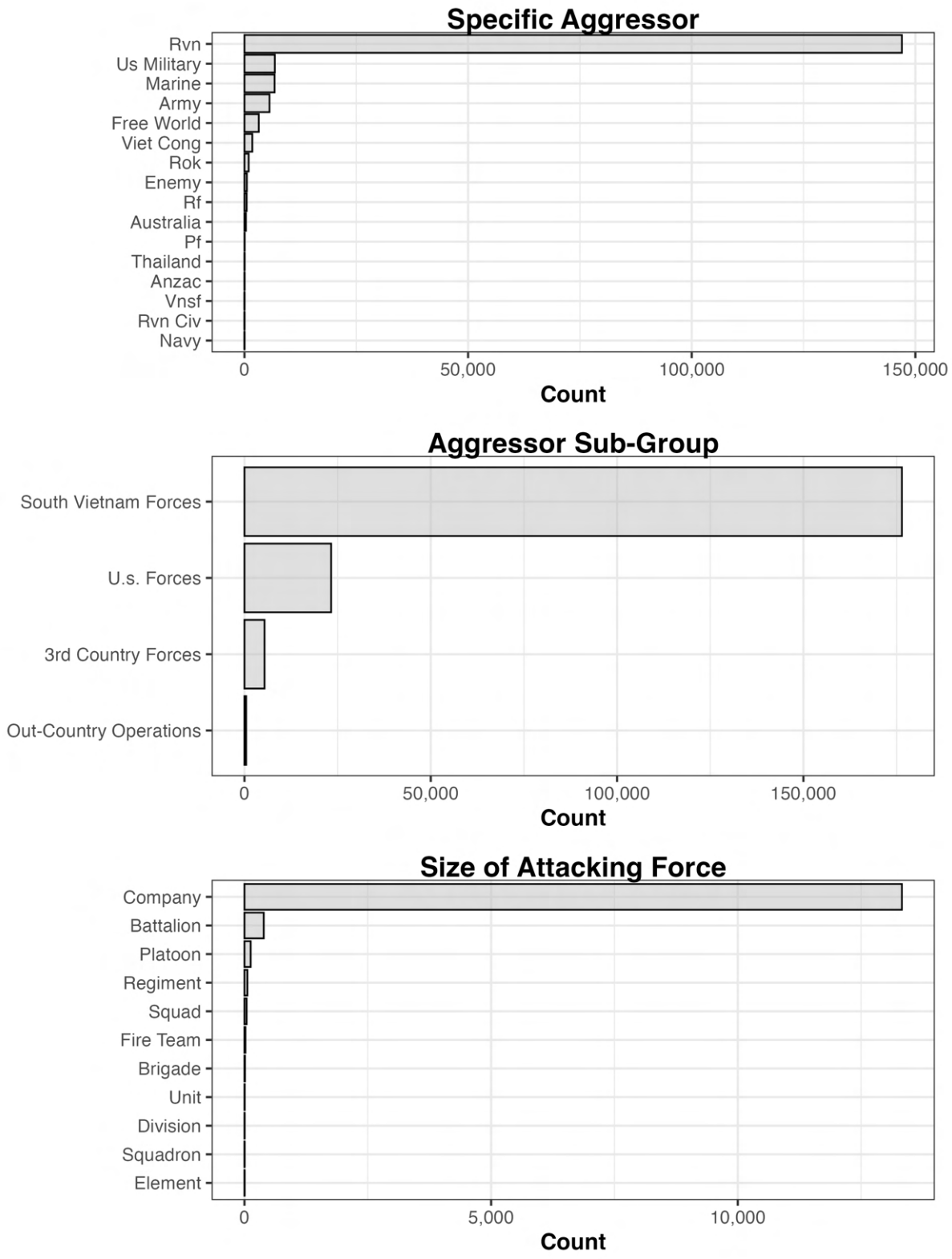


Figure S7: Count of Operation Classifications and Target

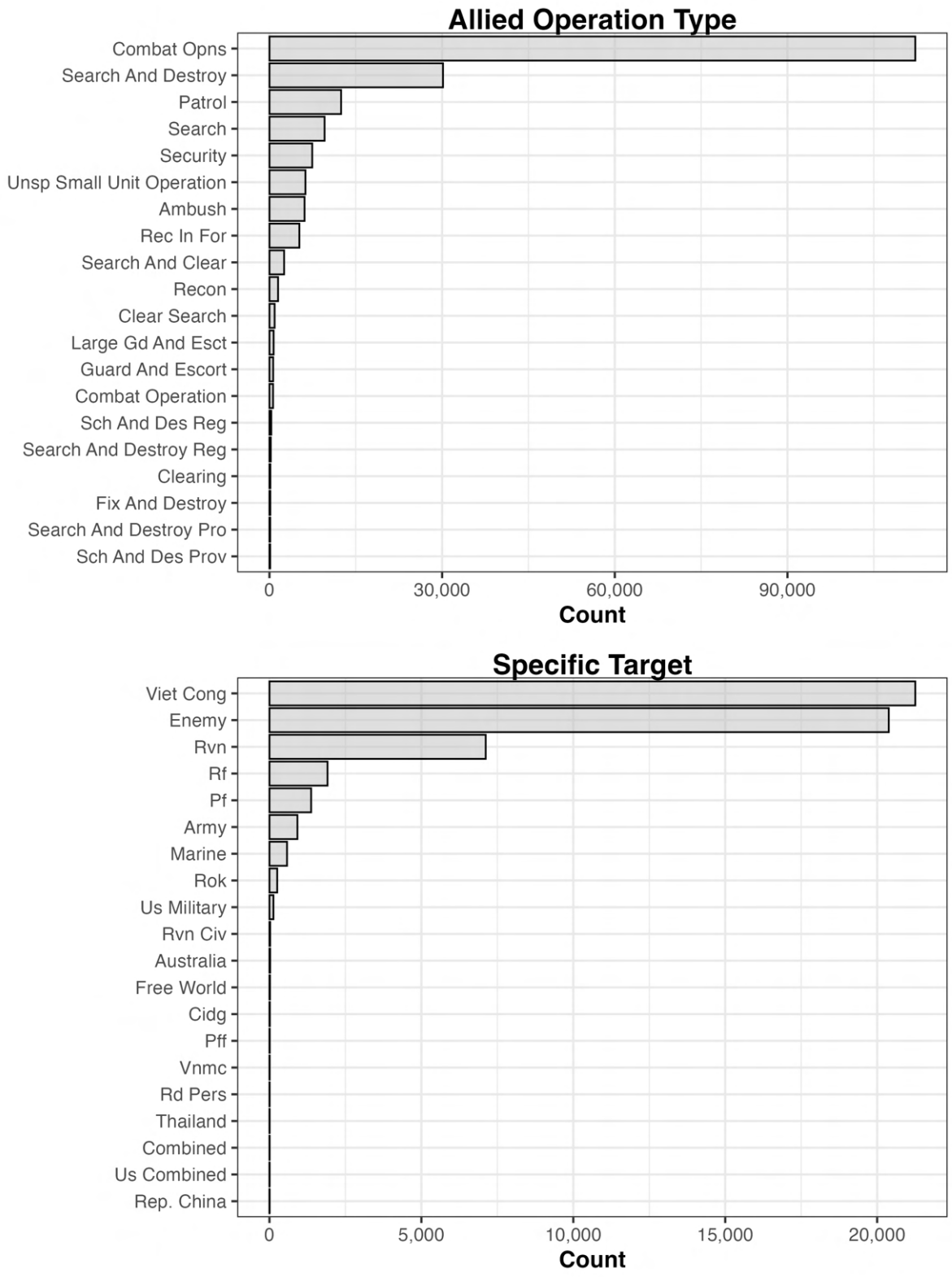


Figure S8: Count of Allied Operation Names

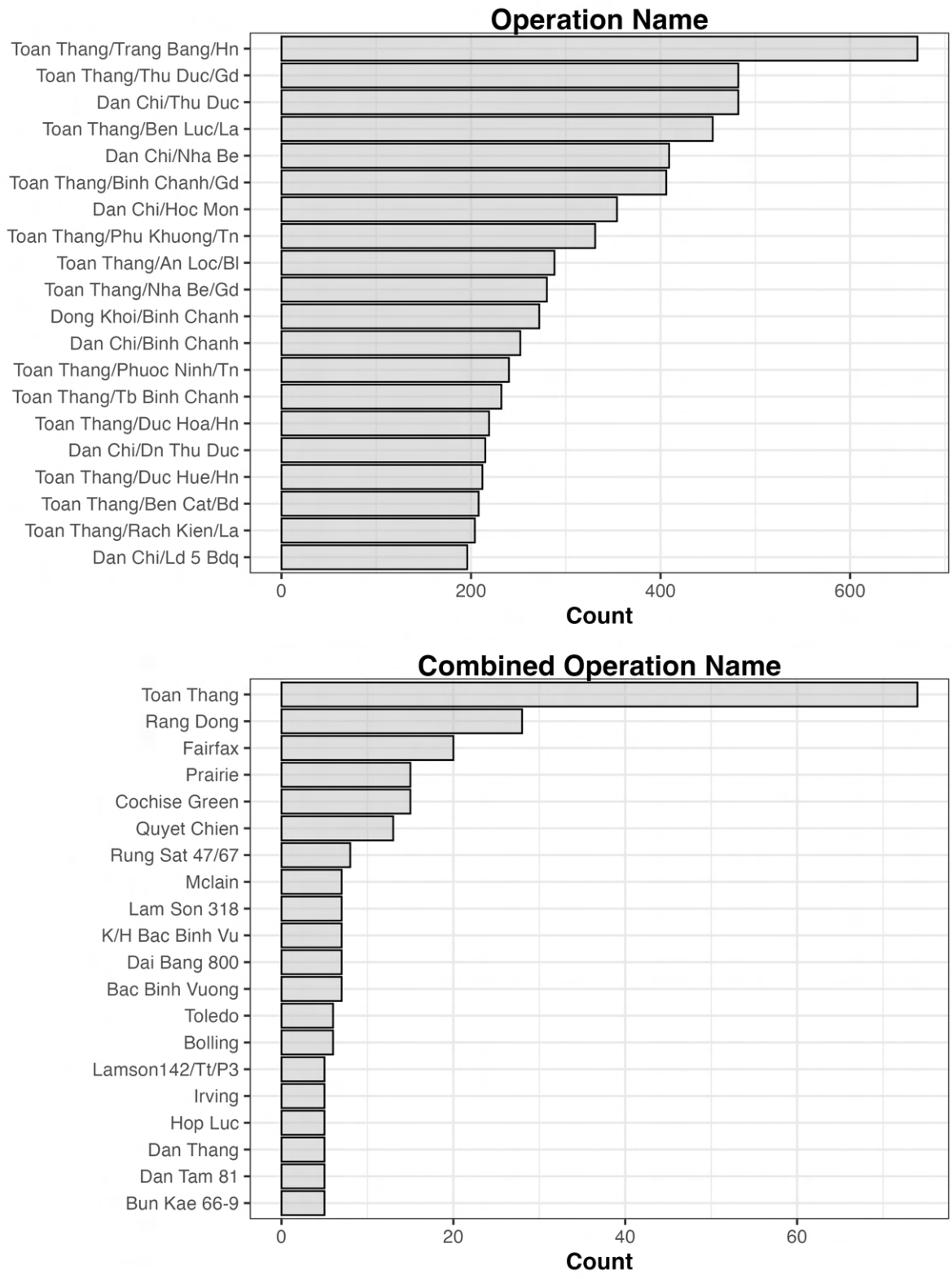


Figure S9: Count of Allied Unit Designations

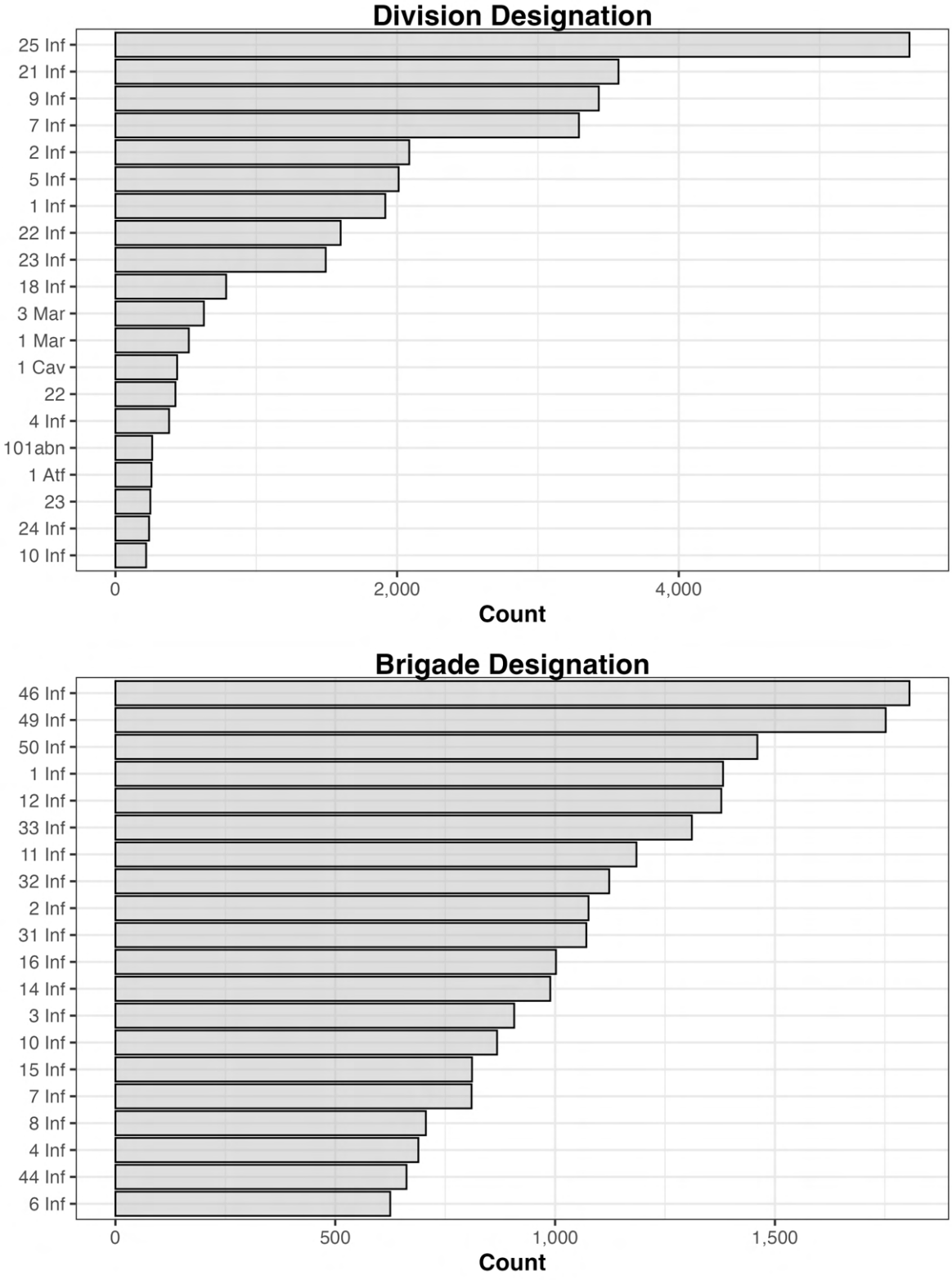
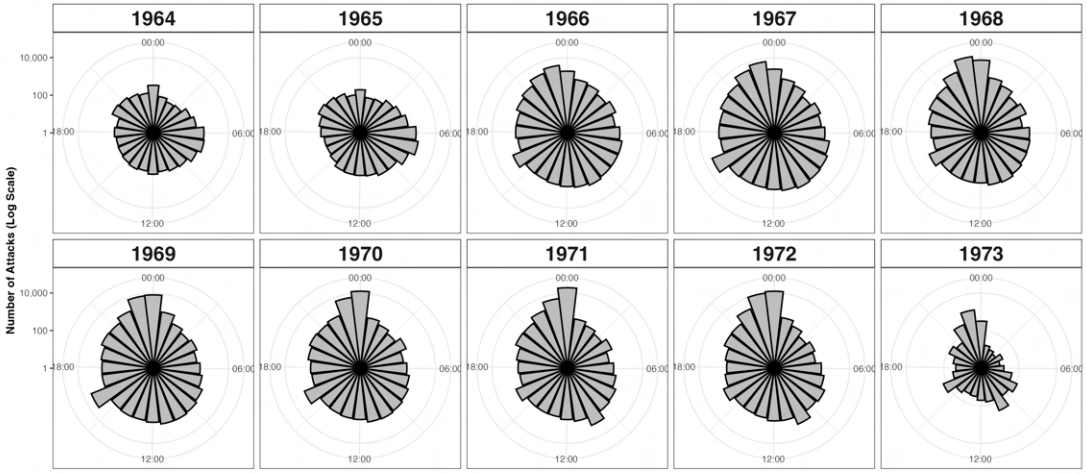


Figure S10: Count of Allied Initiation Times



4.2 An Overview of Enemy Operations

The following figures provide counts of the most common data entries for records of enemy ground combat operations. These figures include counts of the most frequent targets, types of attack designation, and the number of Allied casualties that resulted from enemy operations.

Figure S11: Count of Enemy Action Categories

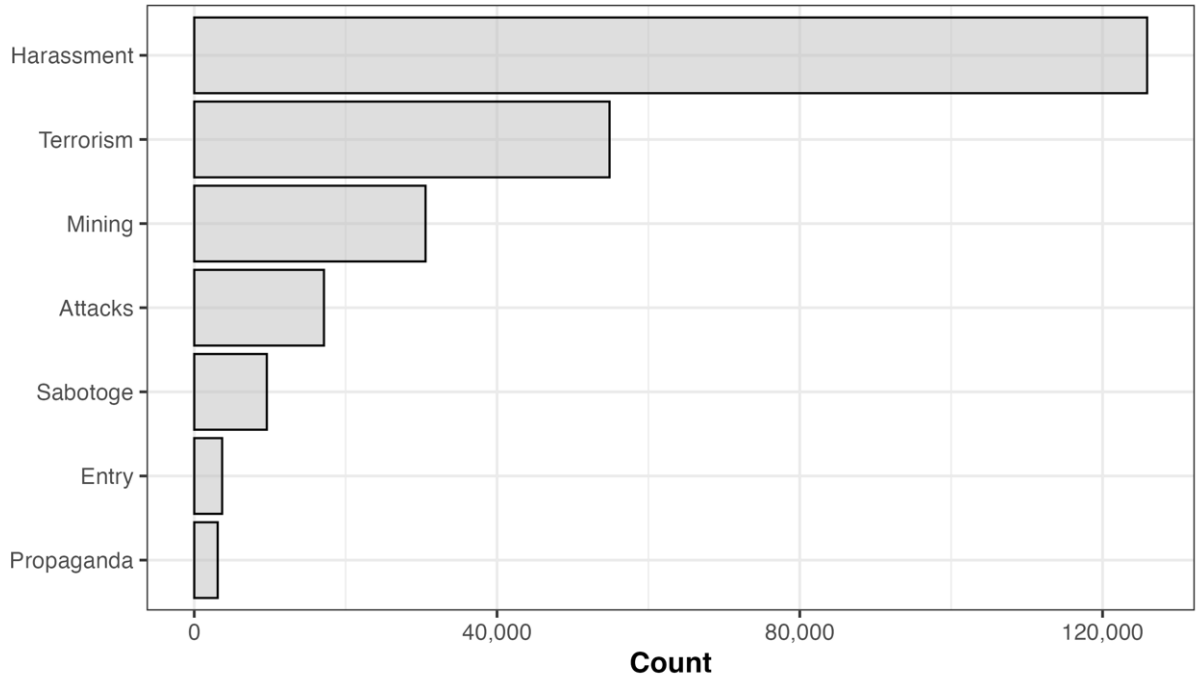


Figure S12: Count of Specific Enemy Targets

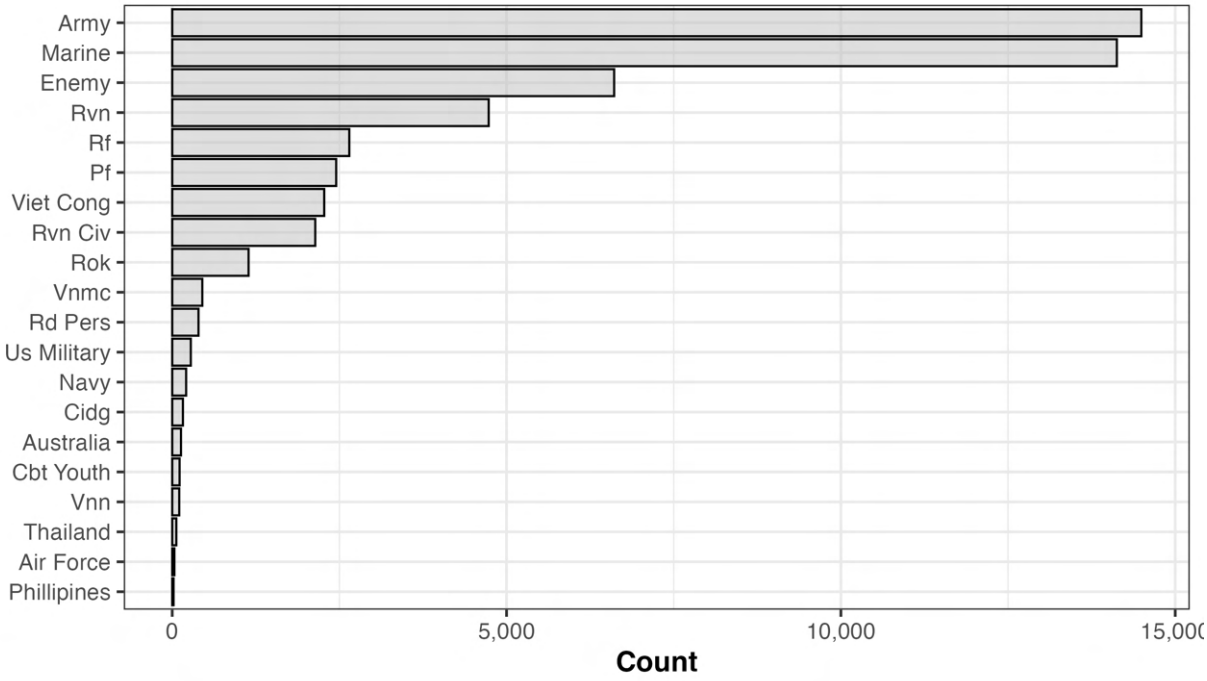


Figure S13: Count of the Size of Enemy Attacking Forces

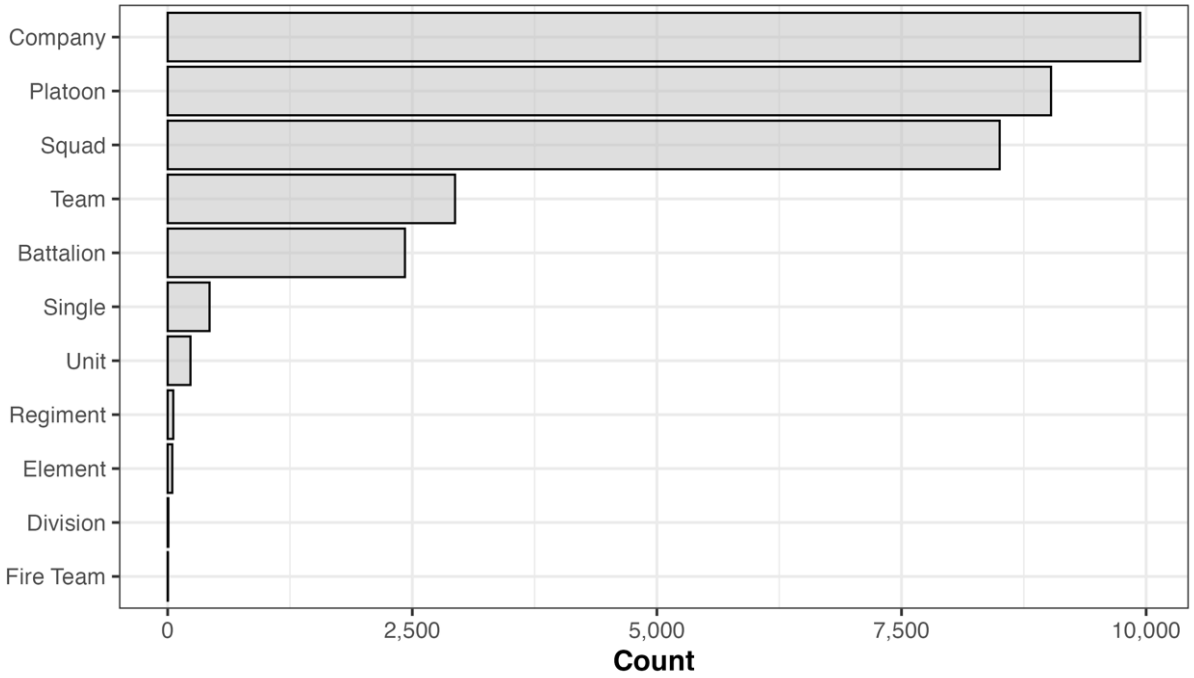


Figure S14: Count of Enemy Objectives

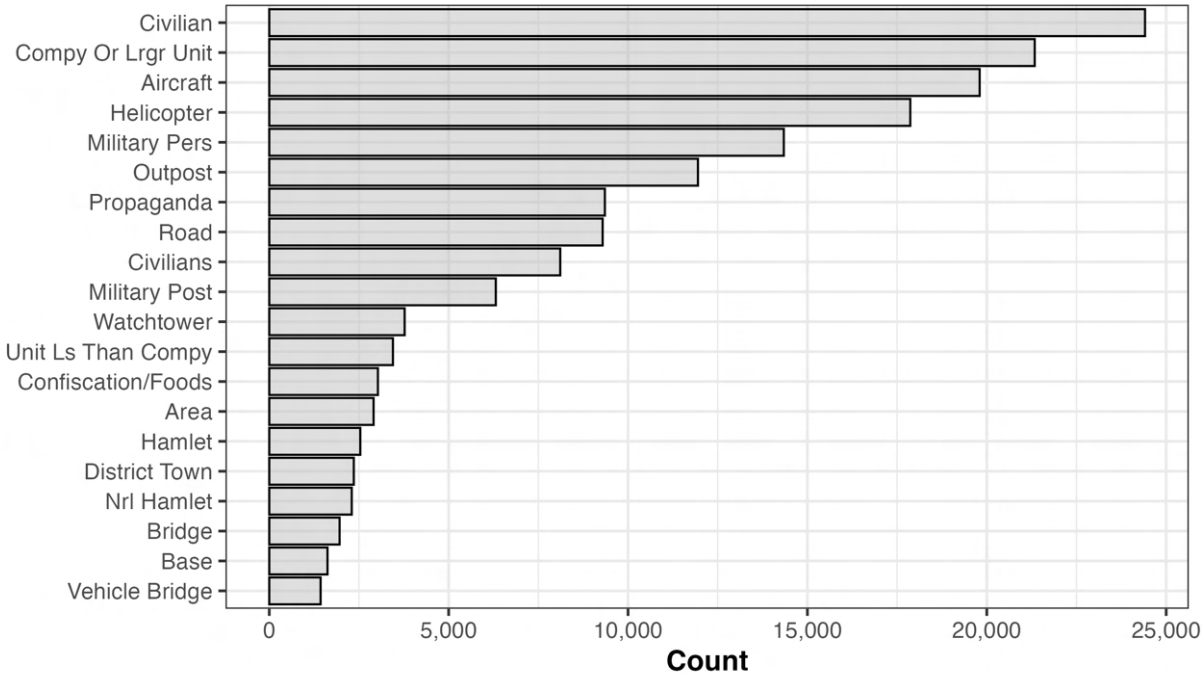


Figure S15: Count of Enemy Operation Names

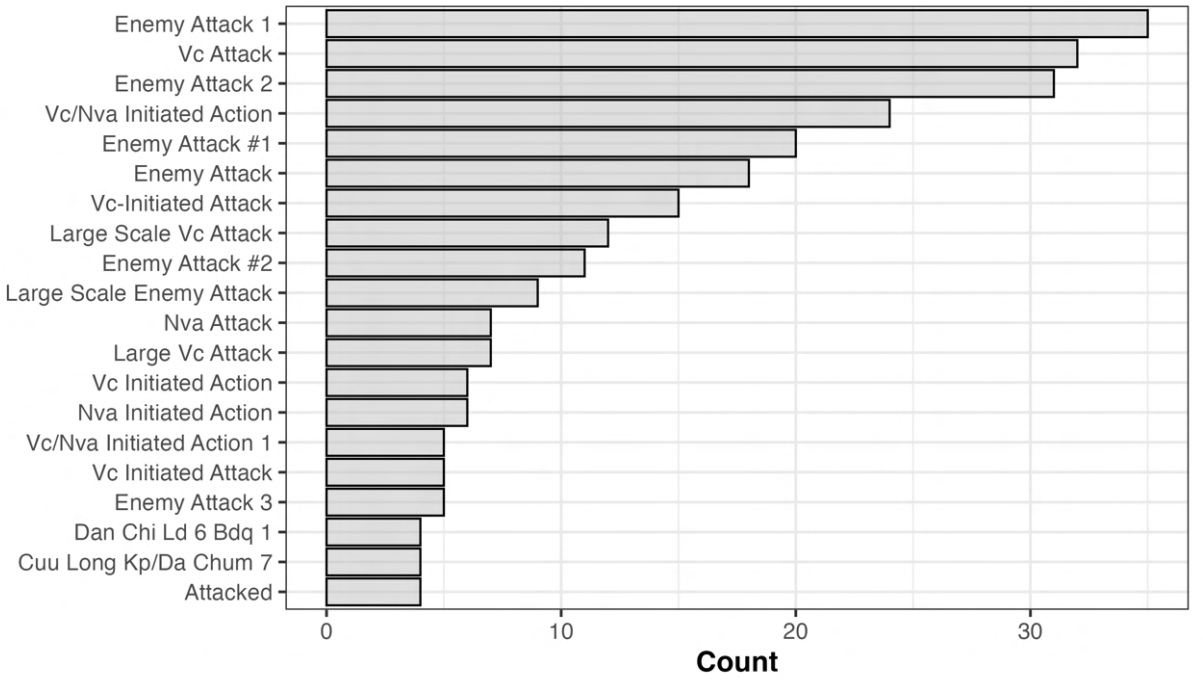


Figure S16: Count of Enemy Attack Times

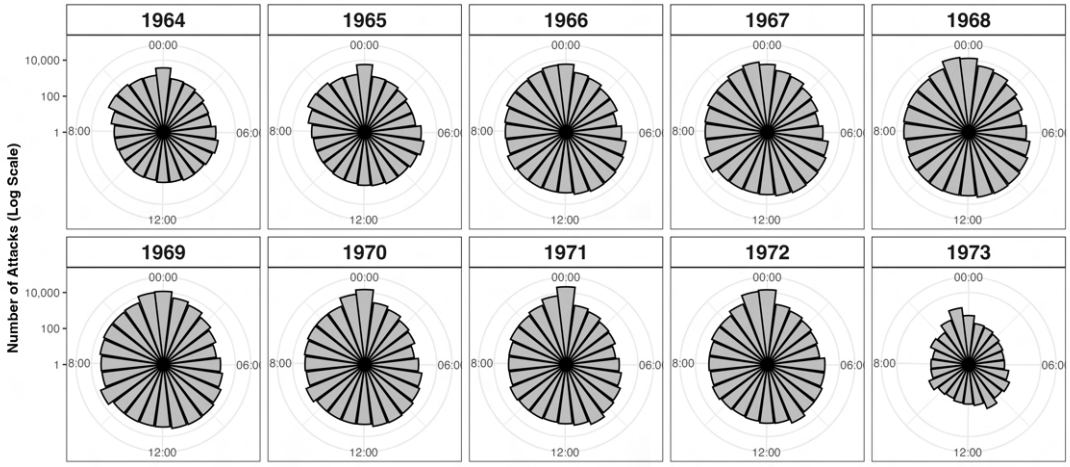


Figure S17: Count of Annual Allied Casualties From Enemy Attacks

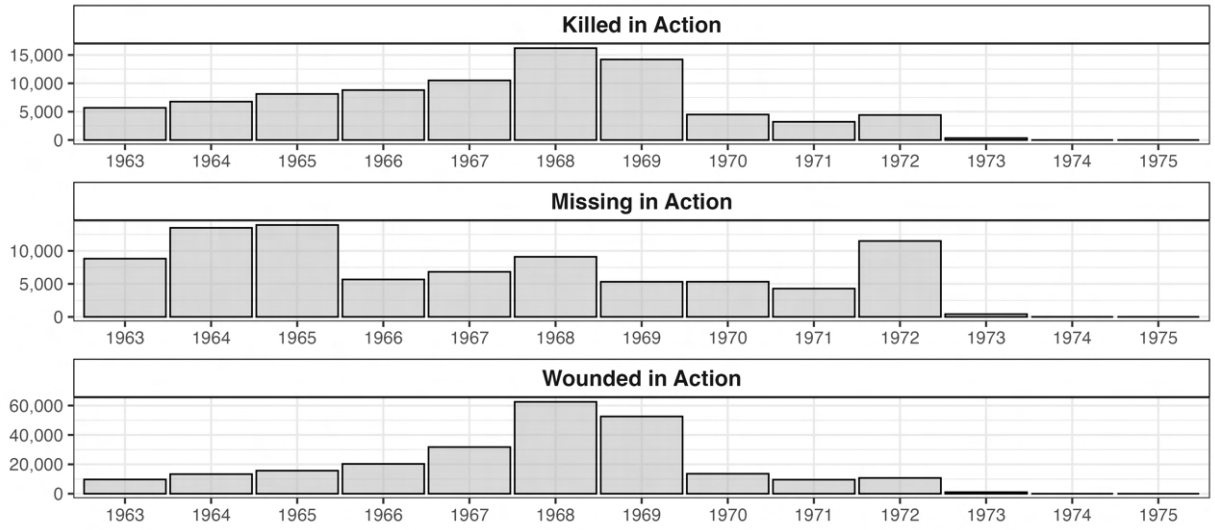
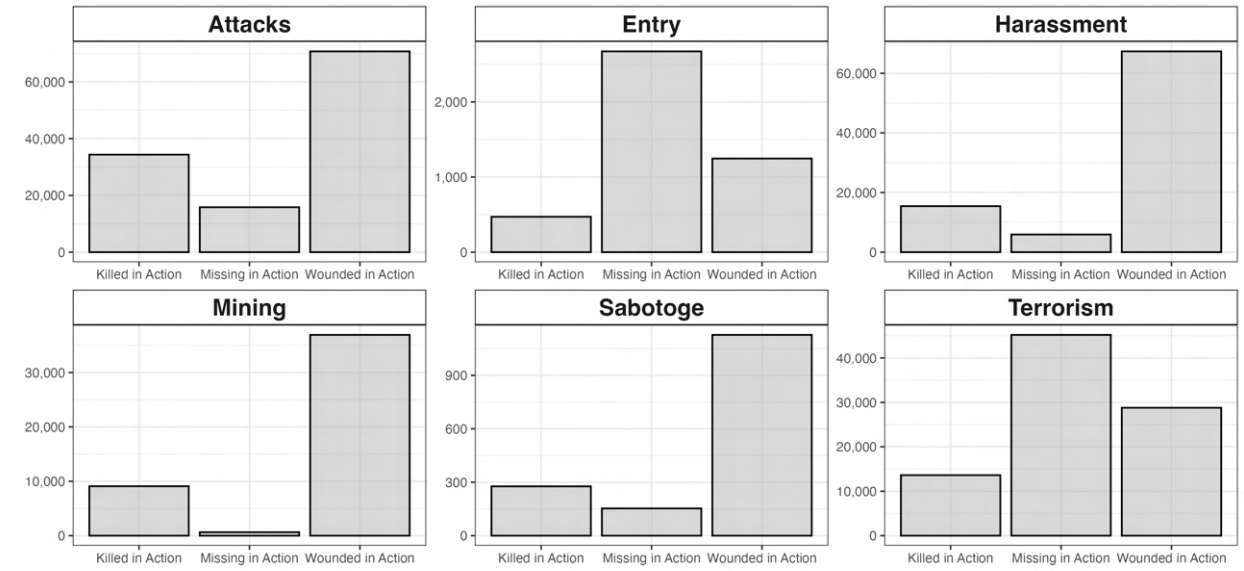


Figure S18: Count of Casualties Per Enemy Attack Type

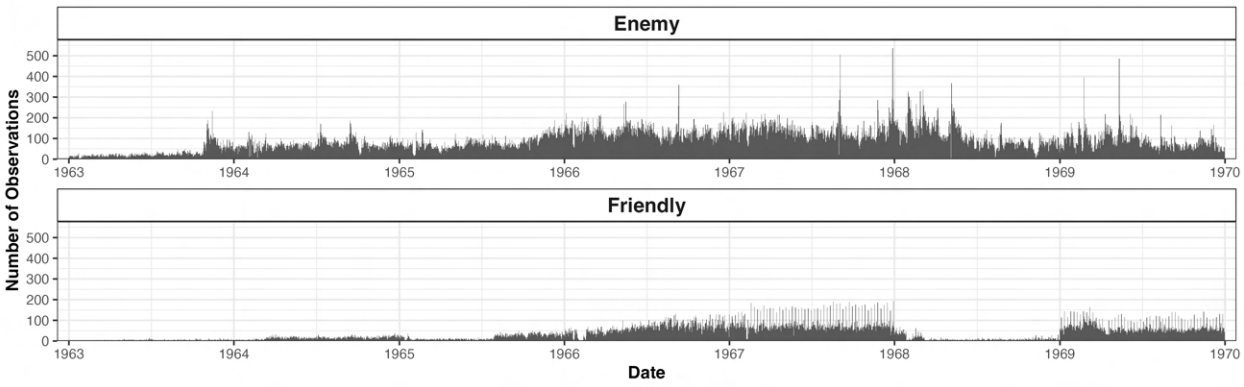


5 Descriptive Overview of the Separate Incident Files

The following section provides a descriptive overview of the information contained in each data file. The overviews of the VNDBA, STRIA, TIRSA, and VCIIA are condensed, as their contents largely overlap with the information in the combined files. The important thing to note is that the data files cover different periods. In addition, there are apparent gaps in the data files during periods that should be covered. In the VNDBA files, for example, the majority of Allied operations initiated during 1968 are missing.

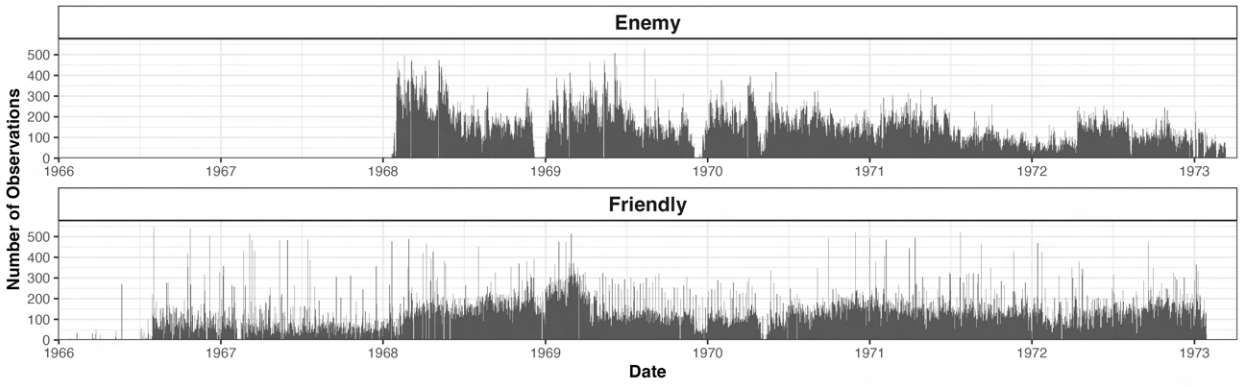
5.1 VNDBA Files

Figure S19: VNDBA Daily Incident Counts (1963 - 191969)



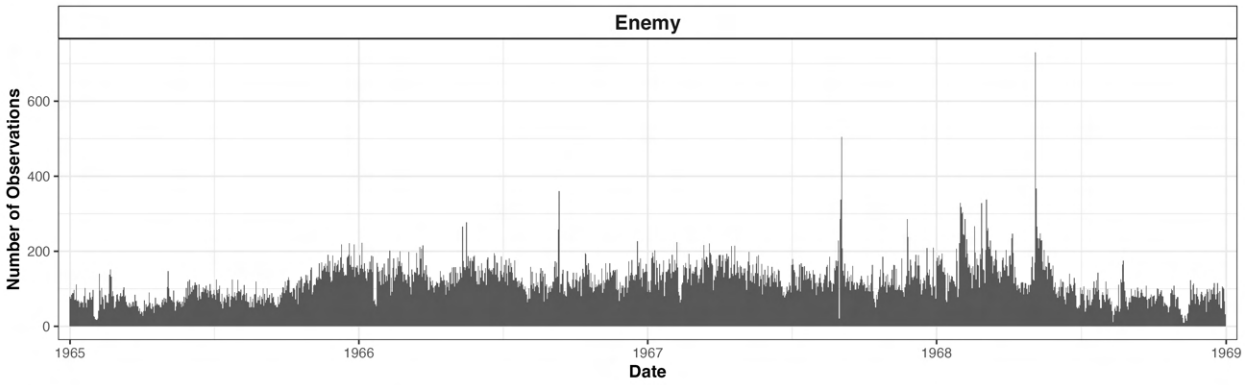
5.2 SITRA Files

Figure S20: SITRA Daily Incident Counts (1966 - 1973)



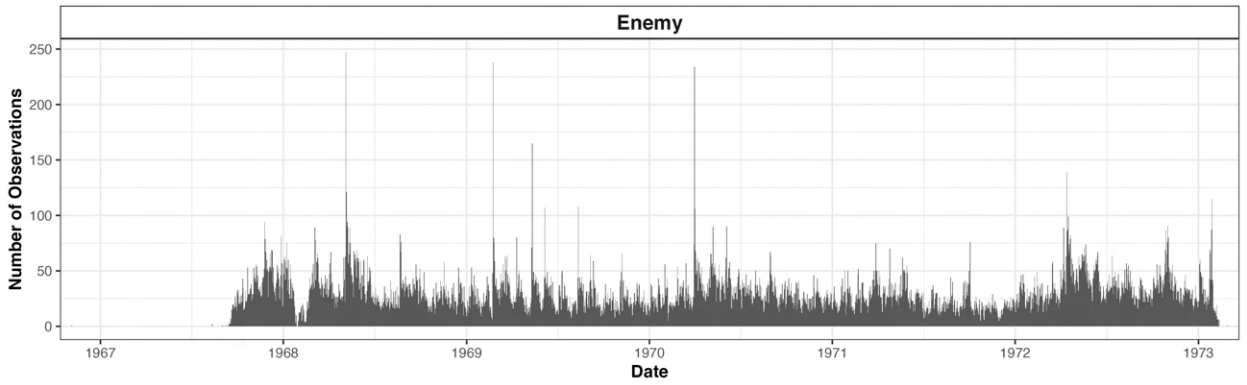
5.3 VCIIA Files

Figure S21: VCIIA Daily Incident Counts (1965 - 1968)



5.4 TIRSA Files

Figure S22: TIRSA Daily Incident Counts (1966 - 1973)



6 CONGA Files

The following figures provide an overview of naval gunfire support data. There are several notable problems with these files. The most obvious is that support missions initiated in 1969, 1970, and 1971 all have missing target coordinates, as do many of the incidents commenced in 1972. Moreover, many of the fields lack a large number of operations.

Figure S23: Daily CONGA Mission Counts (1966 - 1973)

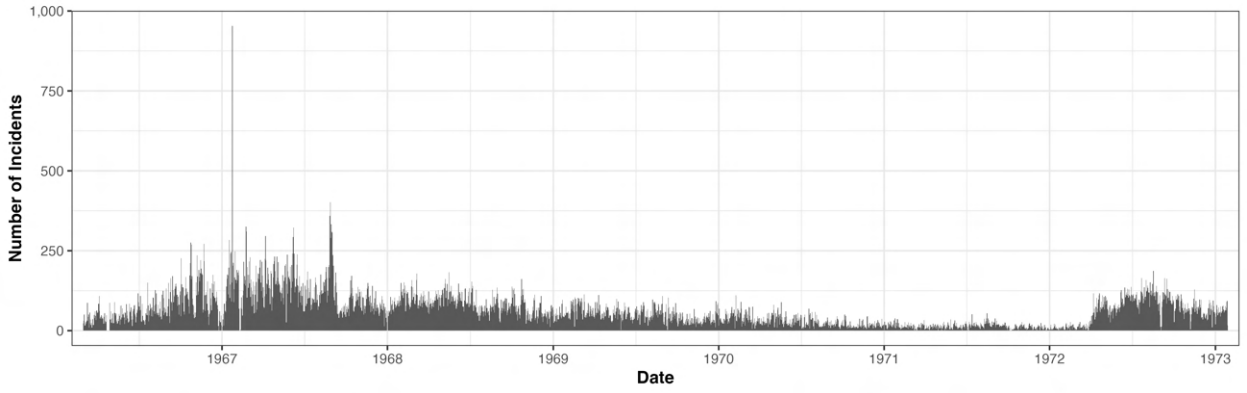


Figure S24: The Coordinates of Each CONGA Mission

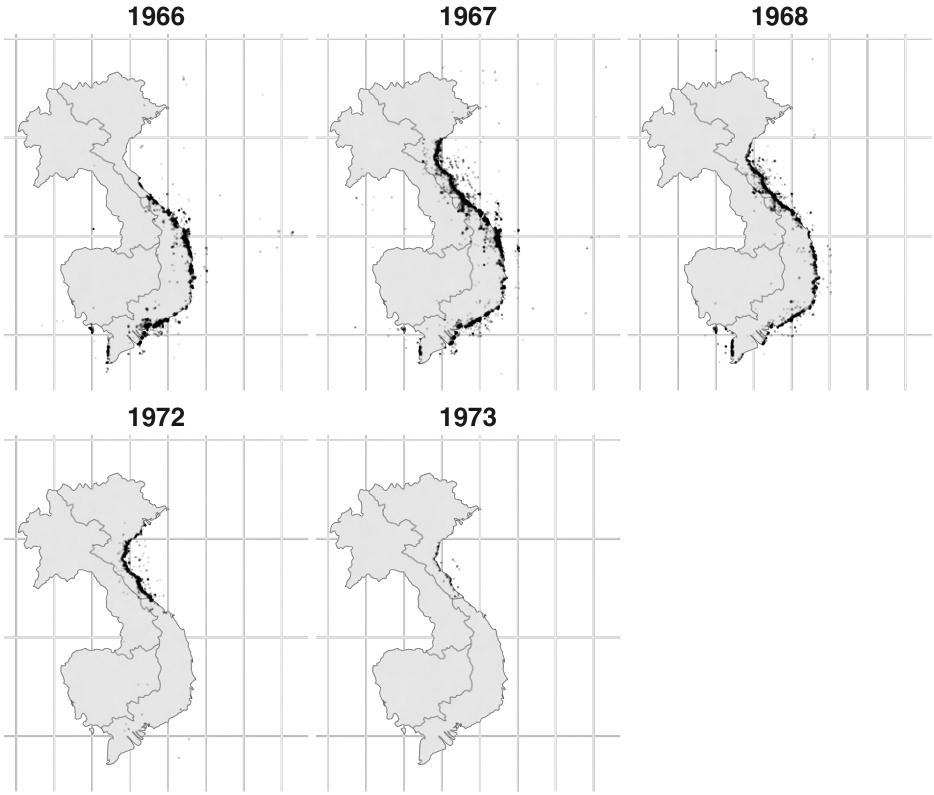


Figure S25: The Number of CONGA Missions Missing Target Coordinates

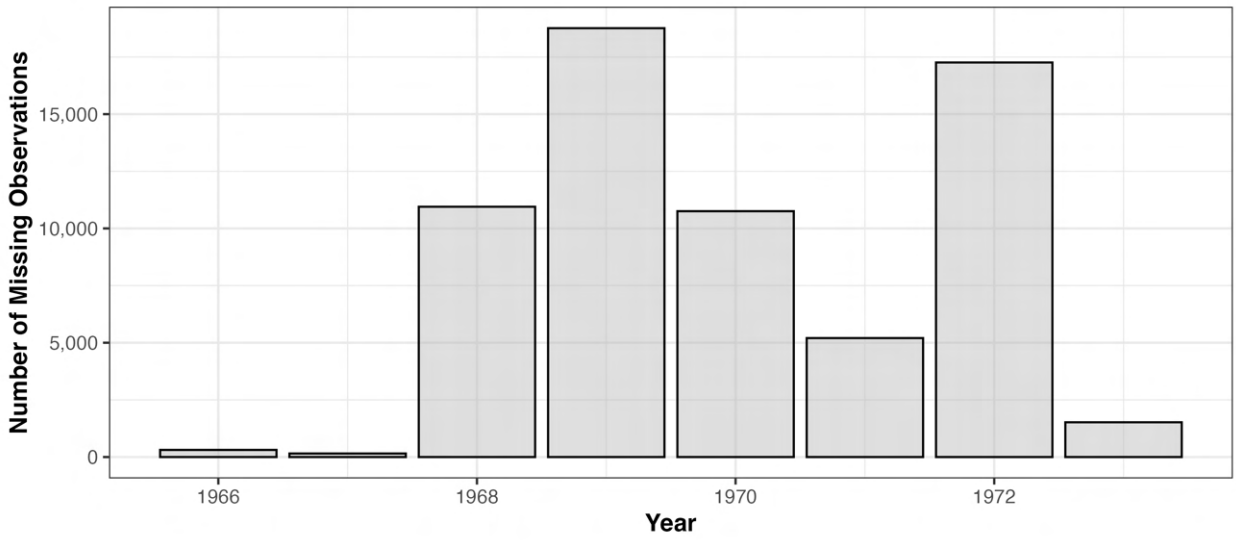


Figure S26: The Thirty Most Common Ships

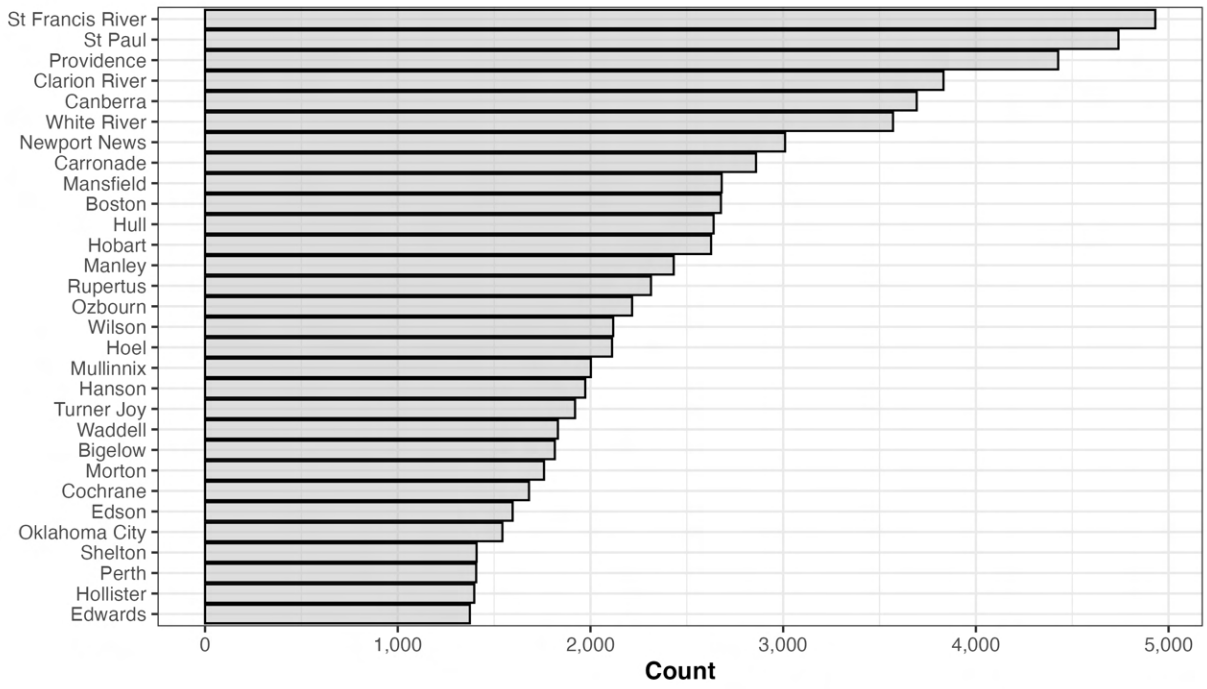


Figure S27: The Most Common Ship Types

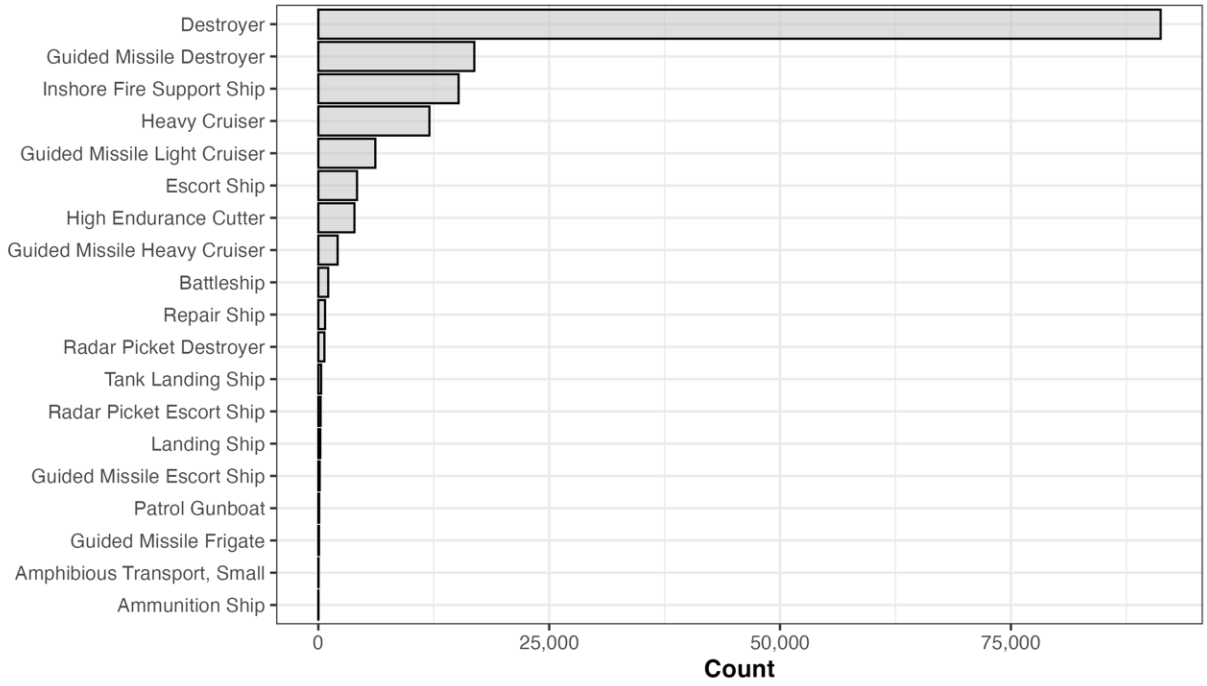


Figure S28: The Most Common Operation Types

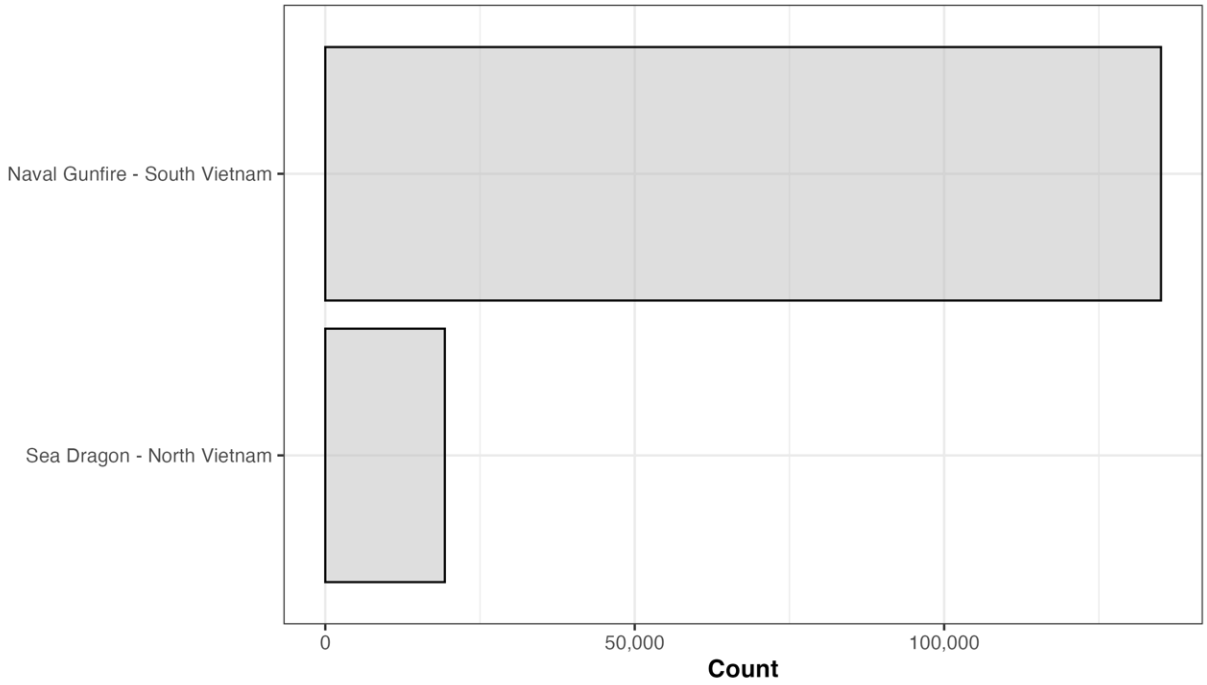


Figure S29: The Most Common Operation Supported

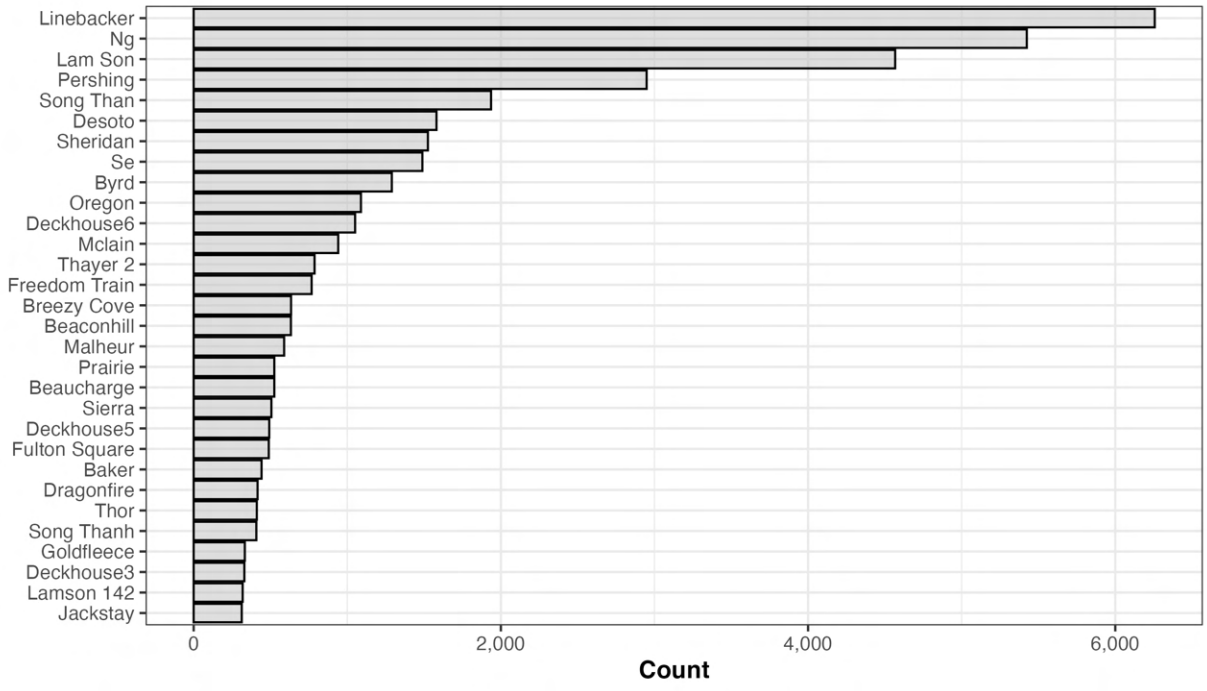


Figure S30: The Most Common Operation Times

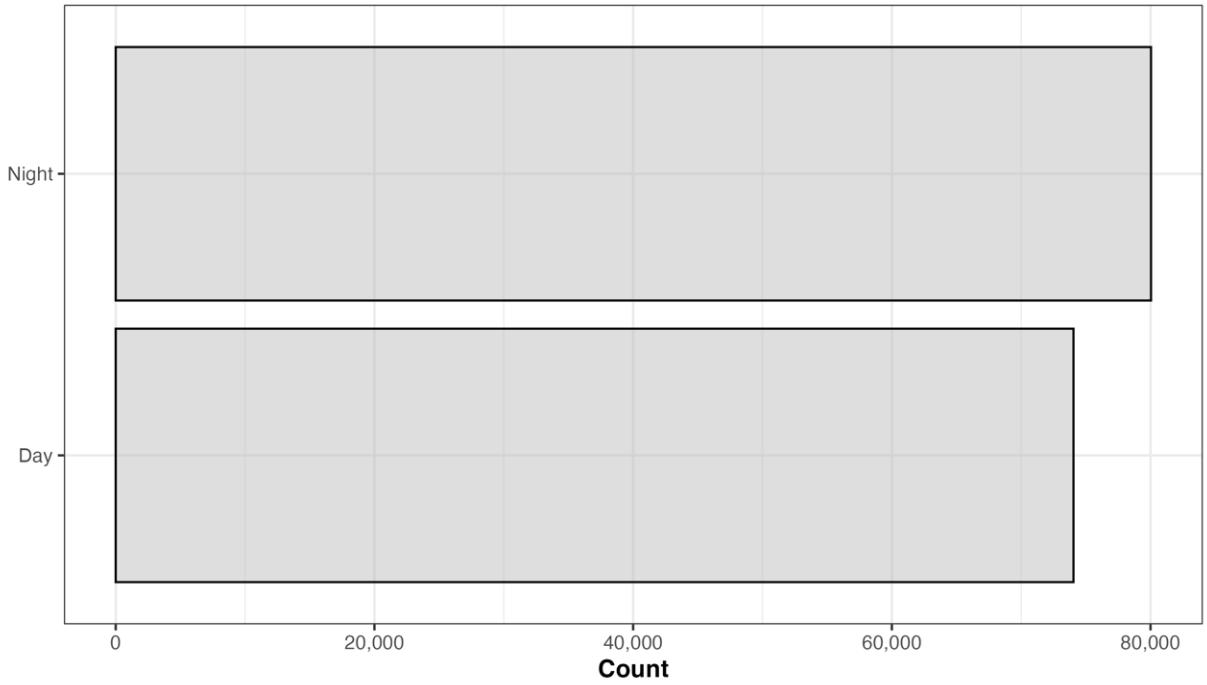


Figure S31: The Most Common Target Types

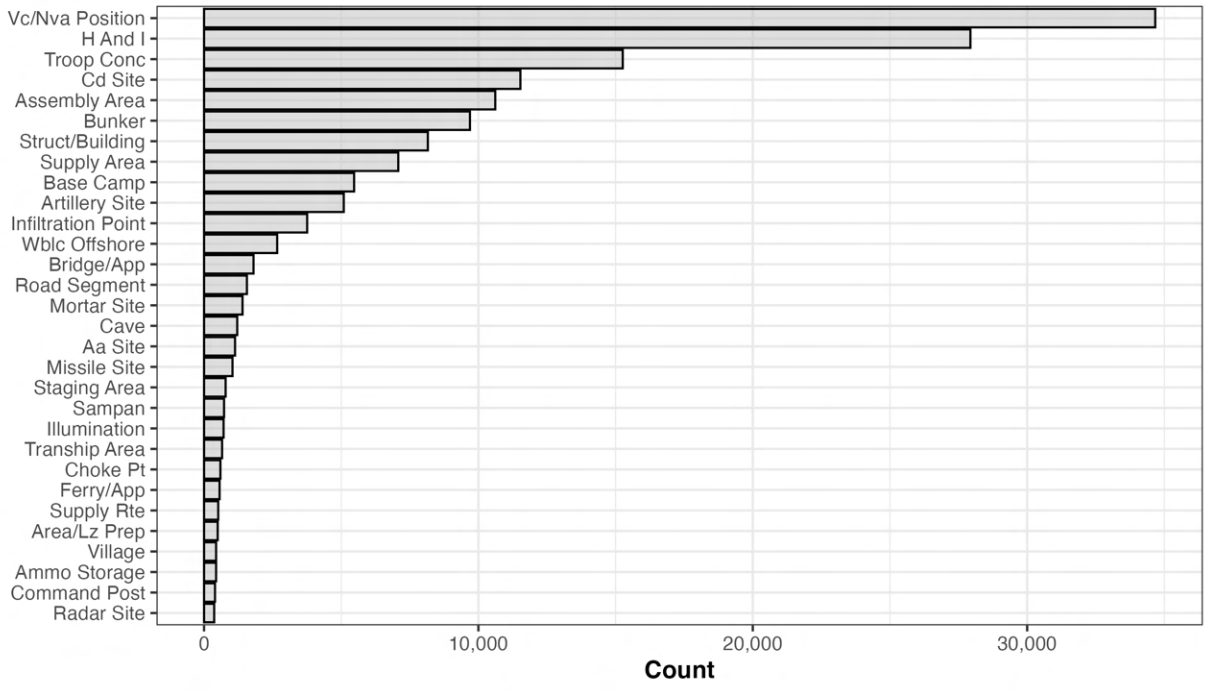


Figure S32: The Most Common Ordnance Types

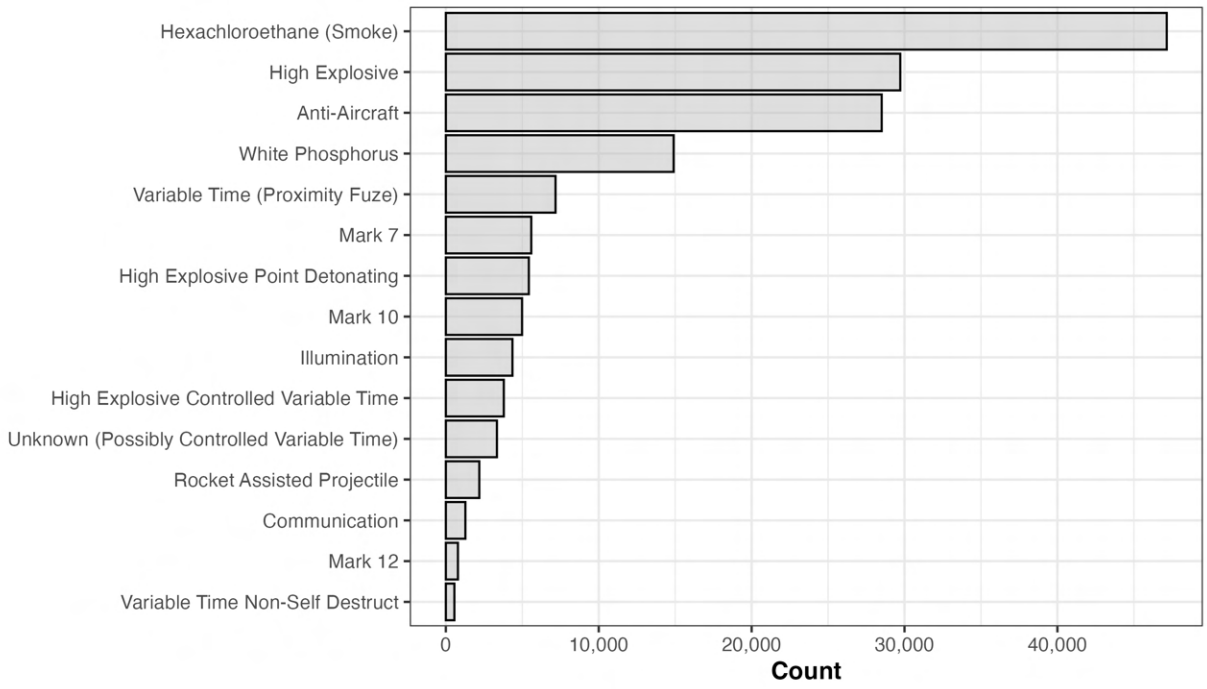


Figure S33: The Distribution of Ordnance Quantity

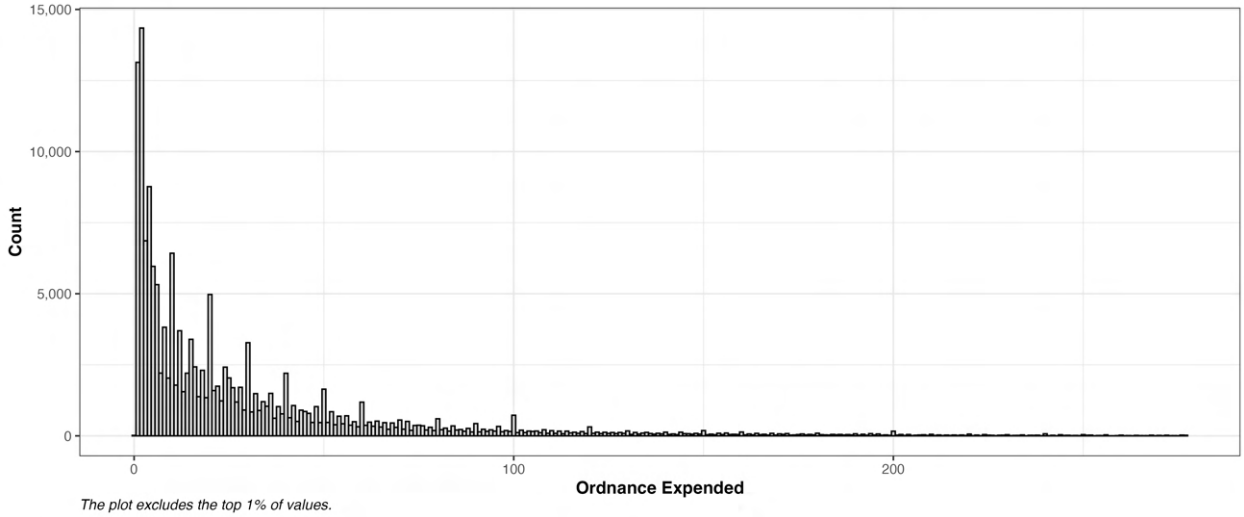
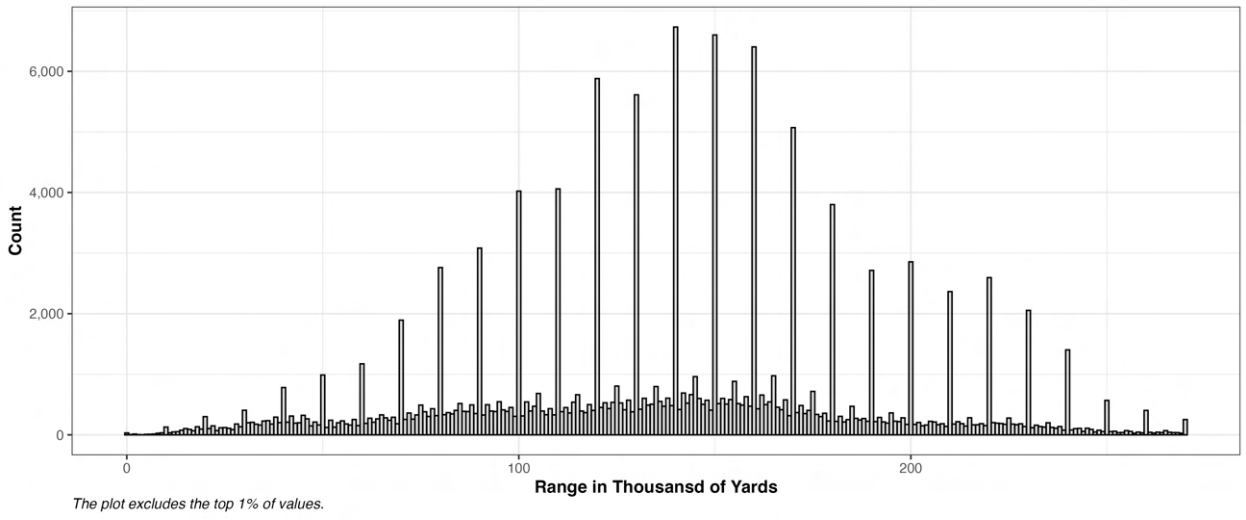


Figure S34: The Distribution of Mission Ranges



7 NASVA Files

The following figures provide an overview of the Naval Surveillance Activities (NASVA) files. These files are challenging to visualize and work with because they contain weekly summaries and daily geolocated data but lack clear column labels to easily distinguish them. The files feature two primary operations, Operation Market Time and Operation Game Warden.

The primary goal of Operation Market Time was to prevent North Vietnamese ships from providing support to Viet Cong forces from the sea. These efforts to stop the flow of supplies resulted in three separate bands/barriers, each relying on a different interdiction method to identify North Vietnamese vessels. Naval patrol aircraft patrolled the outermost barrier, further from the coast, to identify suspicious vessels. The reports from aerial observers were subsequently passed to US Coast Guard cutters that patrolled the area forty miles from shore. The remaining vessels that were interdicted by Coast Guard cutters were stopped by smaller patrol boats that remained closer to the coastline (Fanell, 1997; Naval History and Heritage Command, 2025). According to the US Navy's historical records, this operation was responsible for forcing North Vietnamese forces to shift their logistics lines inland through Laos and Cambodia over the Ho Chi Minh Trail.

Operation Game Warden was the inland counterpart to Operation Market Time, which was designed to deny Viet Cong access and resources in the Mekong River Delta (Robinette, 1987). The Mekong Delta had great strategic importance for both sides, as it lay south and

west of South Vietnam's capital, Saigon, which was home to 40% of the South Vietnam population (approximately 8 million people) and produced most of the country's rice. In total, the area covers 15,000 square miles and contains approximately 1,400 miles (2,300 km) of natural waterways in the Delta, complemented by an additional 2,400 miles (3,900 km) of man-made canals (Carhart, 1984). The primary concern for Allied war planners was that local Viet Cong cadres could easily traverse these waterways without being detected, since more than 50,000 vessels operated in the region. The early riverine operations consisted mainly of 31-foot river patrol boats. Still, they later expanded to include landing ships, mine sweepers, landing ship docks, and helicopters, working in tandem to intercept any vessel (Schreadley, 1992). Although the effect of these riverine operations remains debated, some historians have credited them with helping to secure South Vietnam during the Tet Offensive in 1968 (Sherwood, 2015).

The following section provides an overview of the information included in these files. The figures are separated by operation unless specifically stated otherwise.

Figure S35: NASVA Operation Count

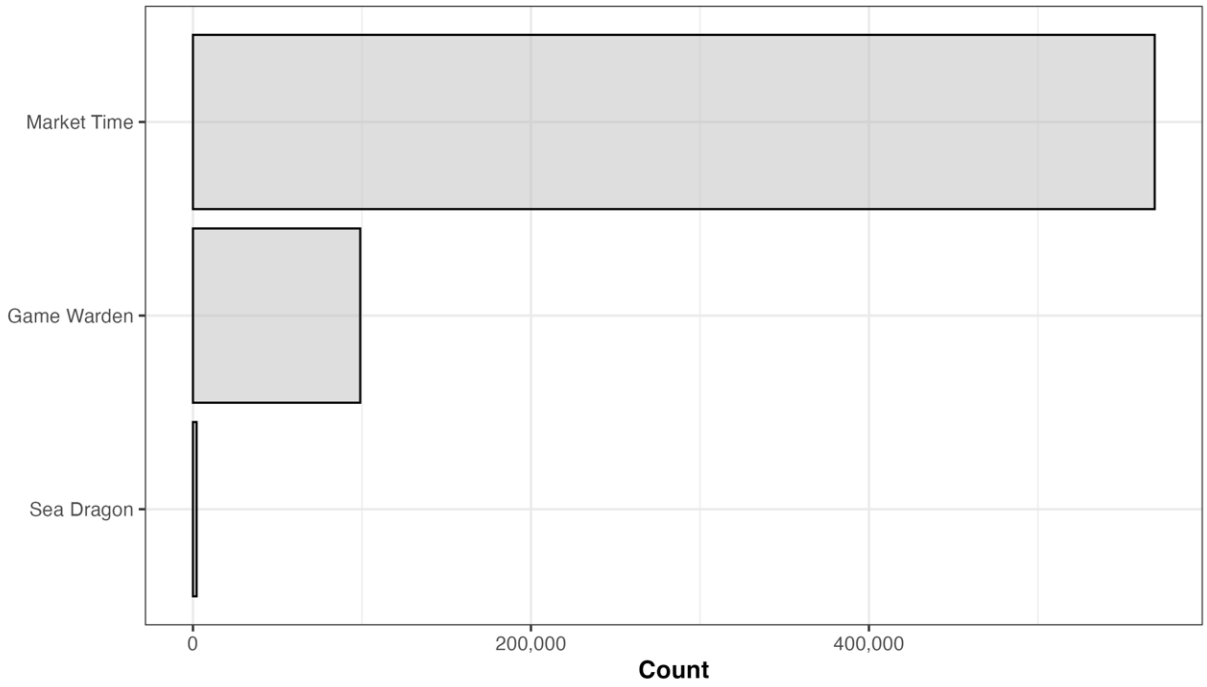


Figure S36: NASVA Enemy or Friendly Vessel Counts

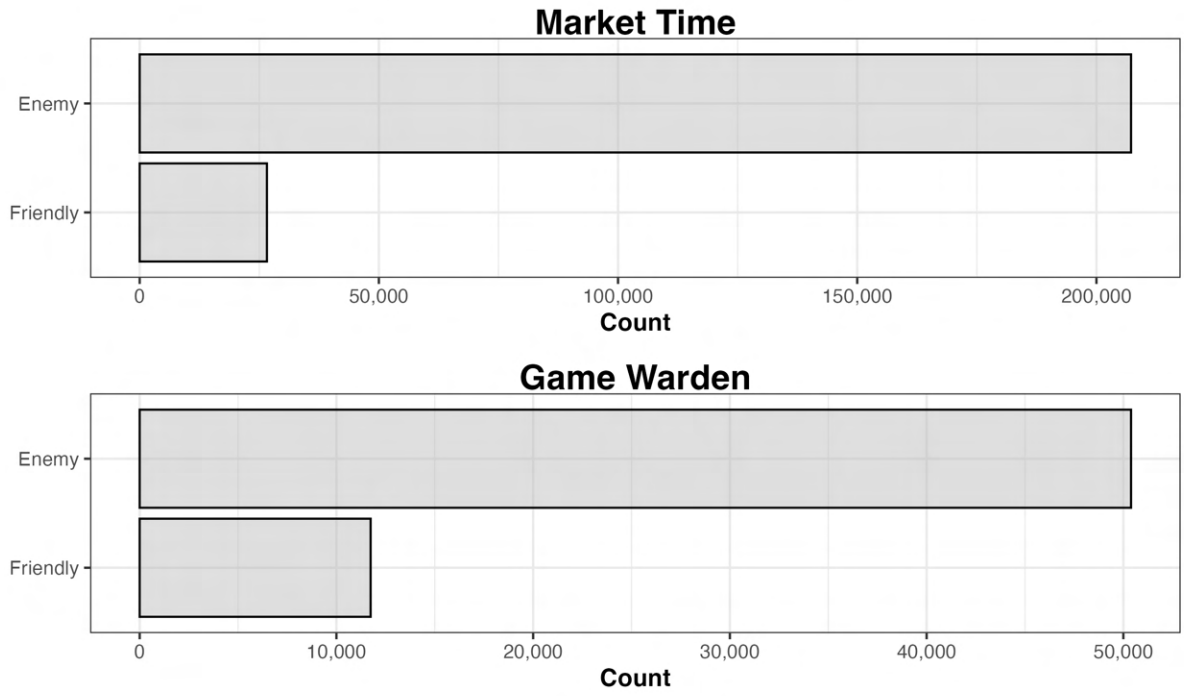


Figure S37: NASVA Full Ship Type Count

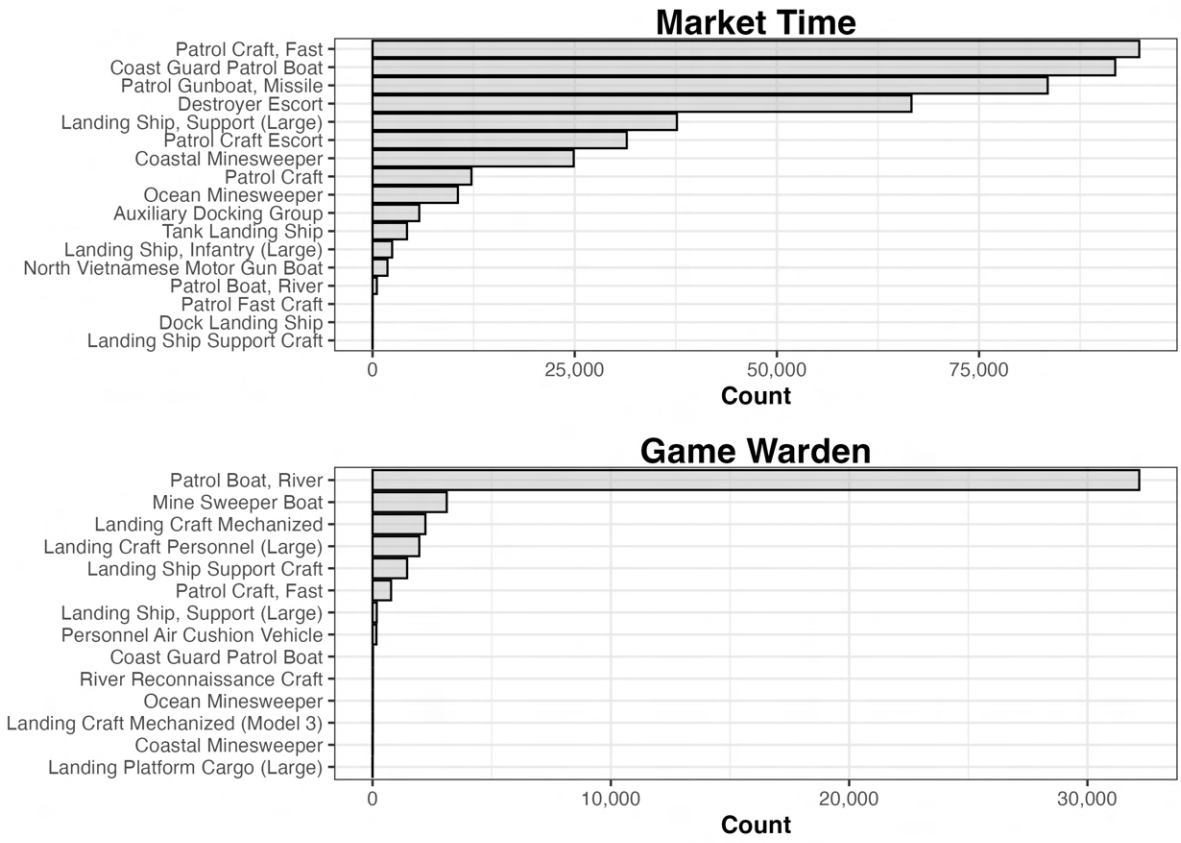


Figure S38: NASVA Count of Ships Involved by Name

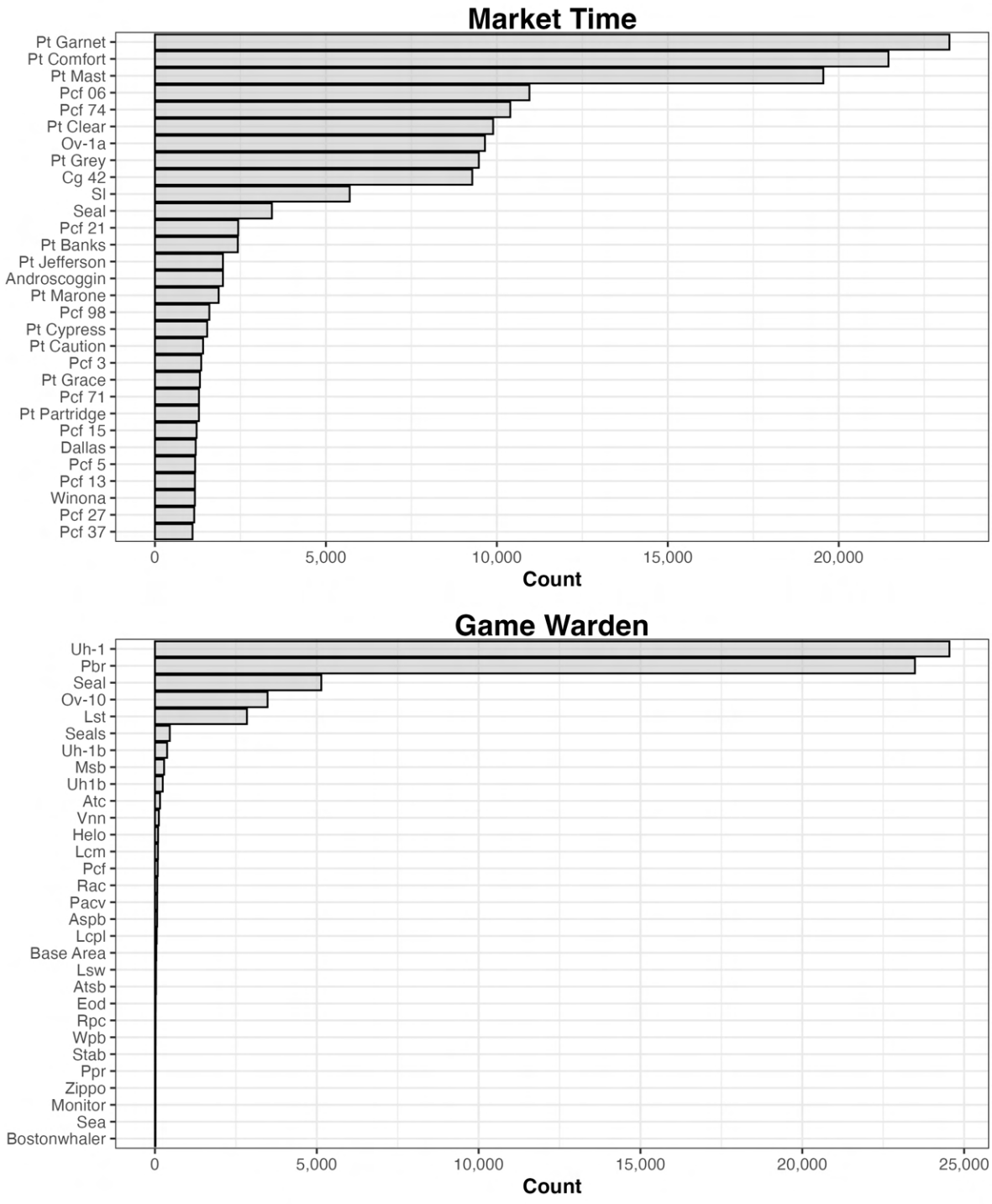


Figure S39: NASVA Market Time Aircraft Type Count

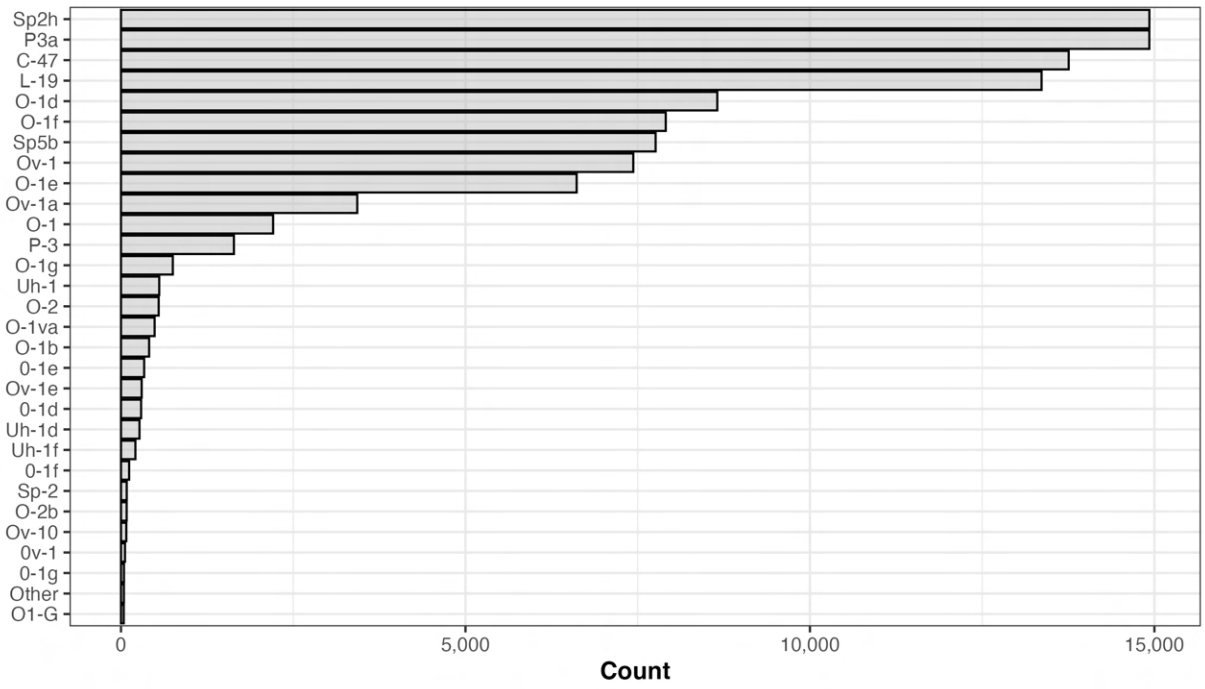


Figure S40: NASVA Material Loss Classification Count

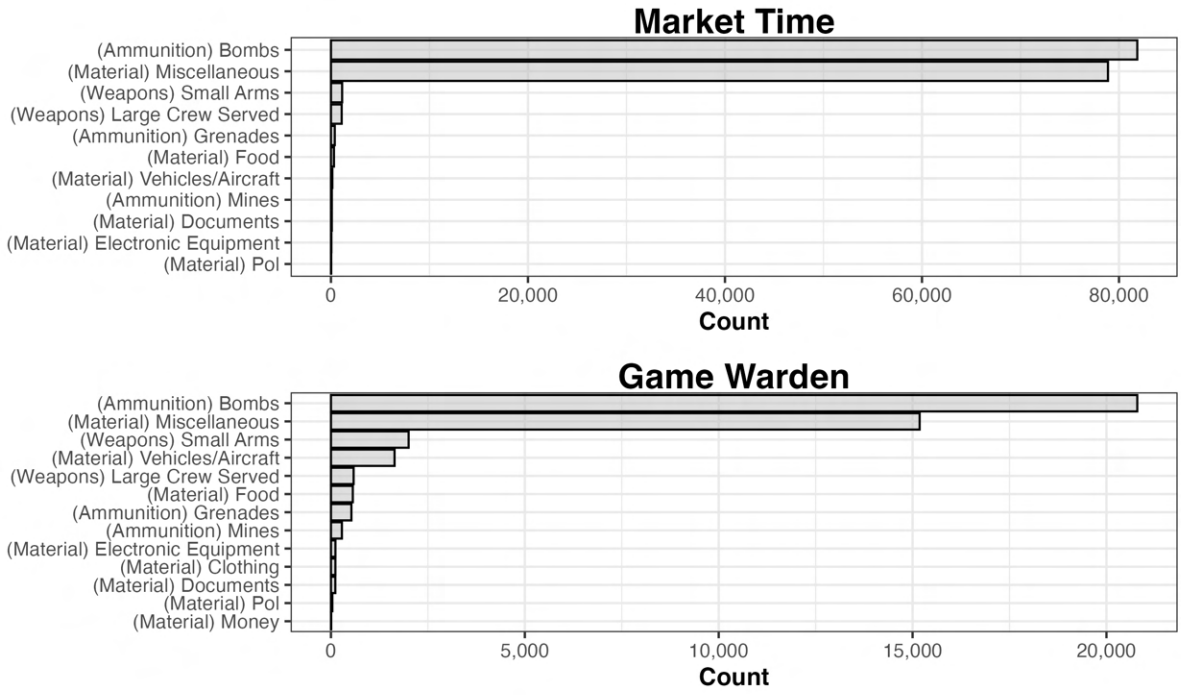


Figure S41: NASVA Ship Nationality Distribution

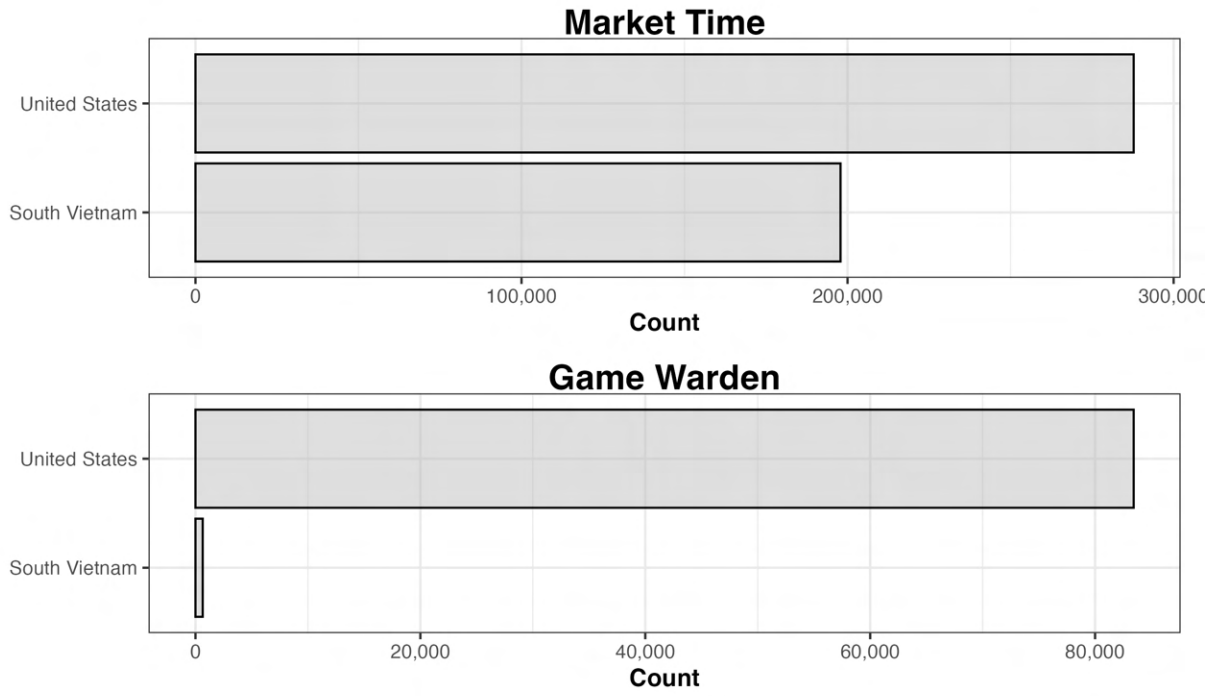


Figure S42: NASVA Vessel Types Count

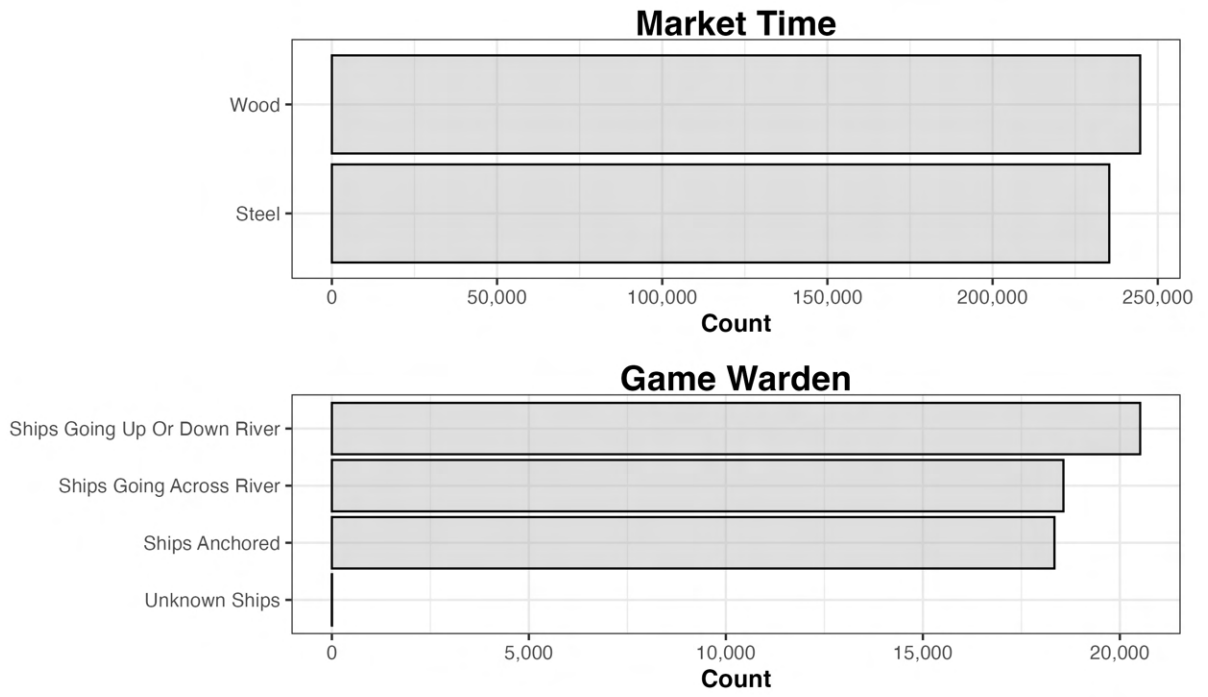


Figure S43: The Location of Every Coordinate in the NASVA Files

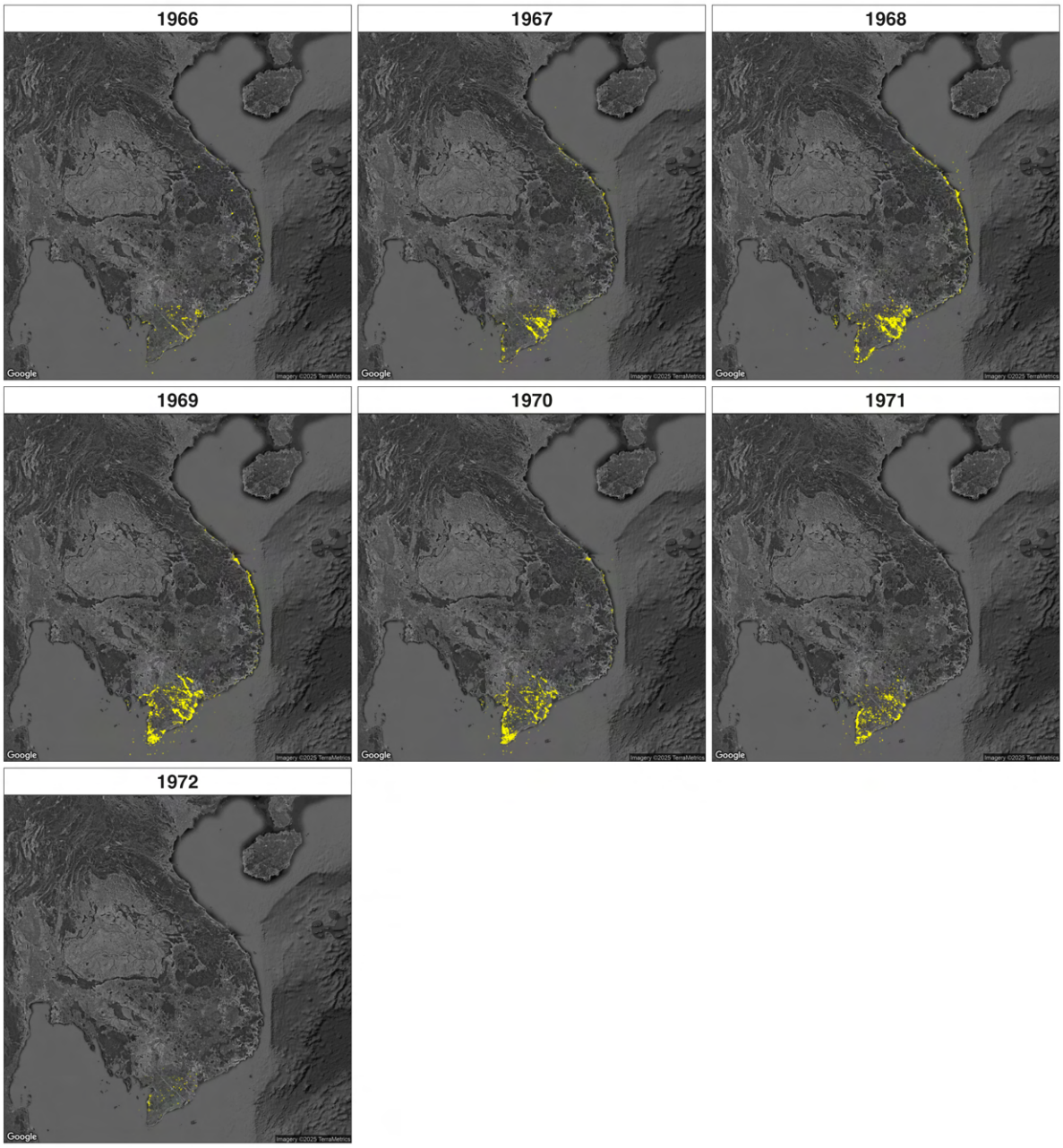


Figure S44: NASVA Market Time Incident Locations

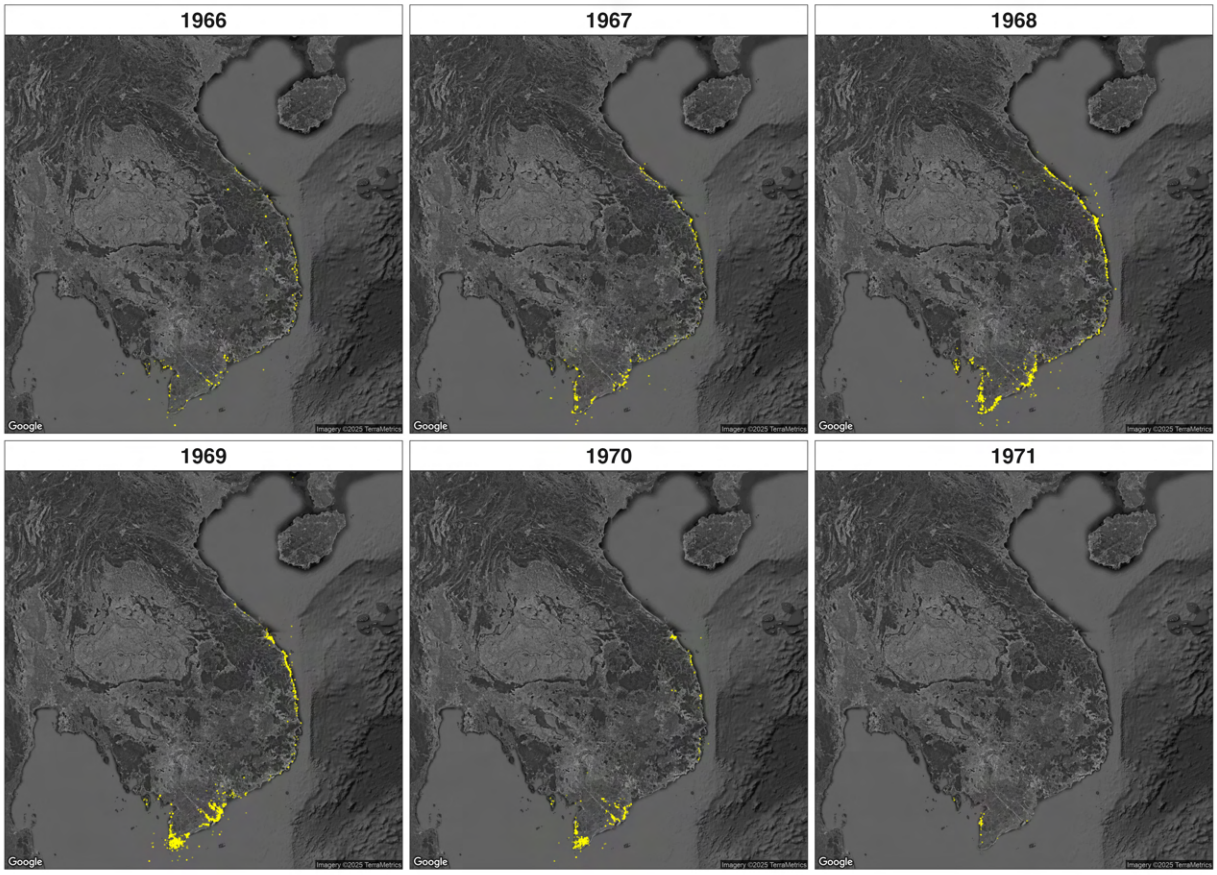
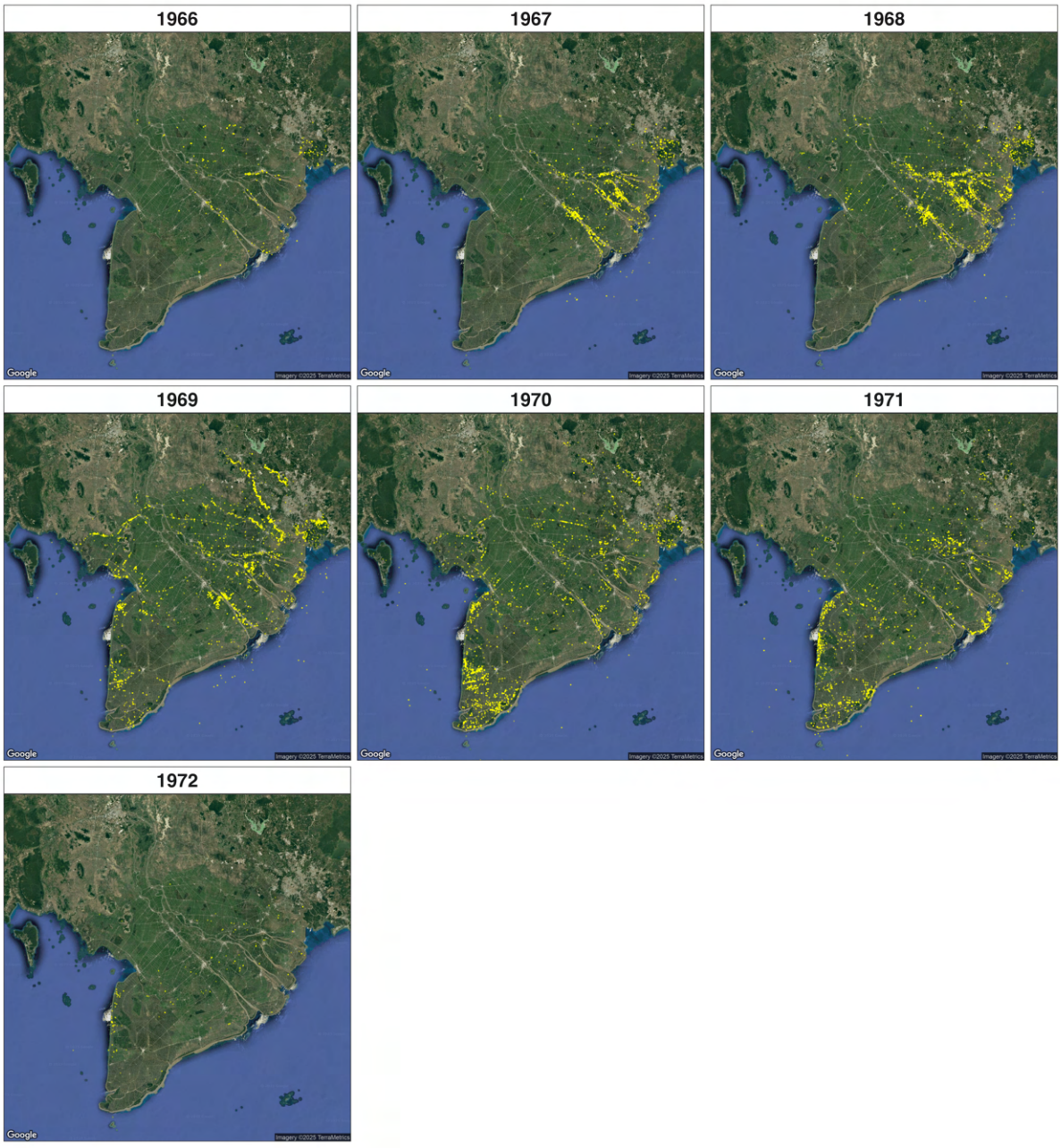


Figure S45: NASVA Game Warden Incident Locations



8 DCAS Files

The following figures provide an overview of the Vietnam War Defense Casualty Analysis System. This file includes background information for US military personnel who were killed during the Vietnam War. The plots display the 10 most common values for each demographic

description, unless otherwise specified.

Figure S46: Timeline of Deaths and Incidents (1963-1973)

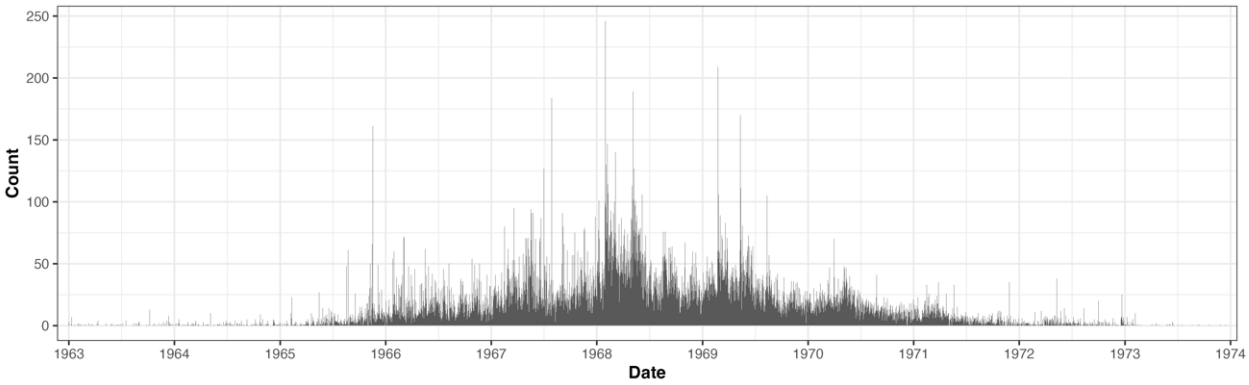


Figure S47: Distribution of Age at Death

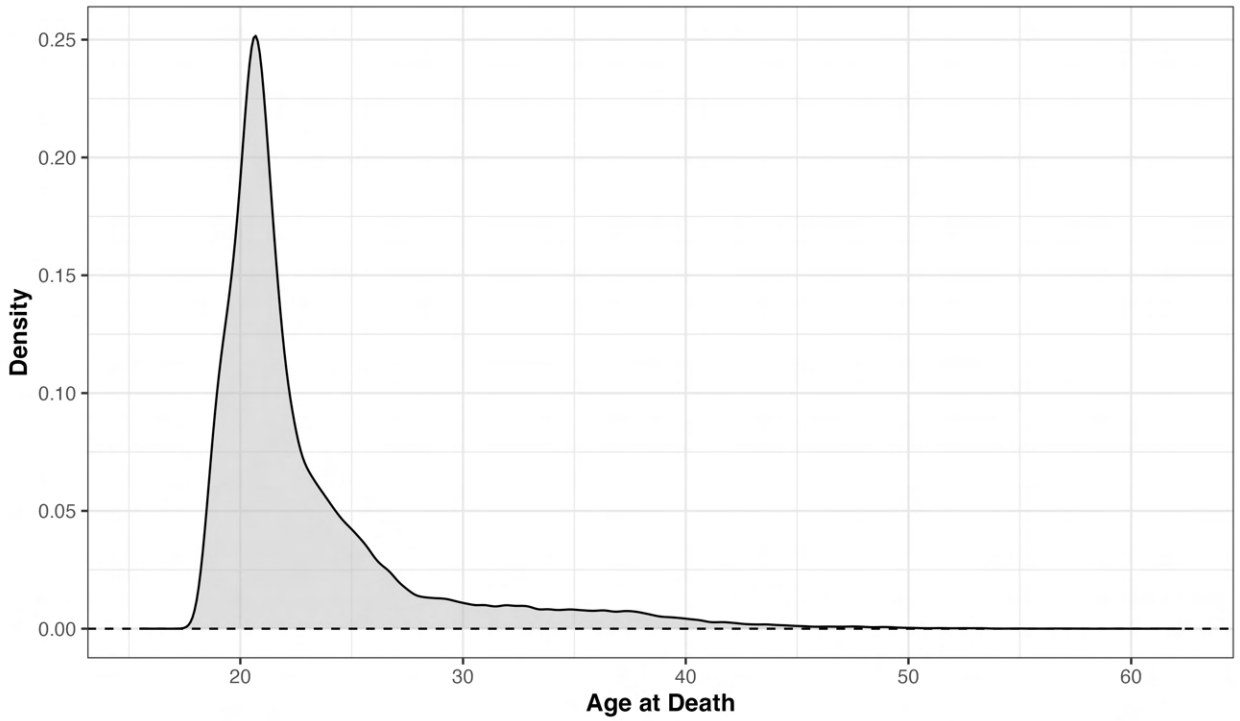


Figure S48: Count of Casualty Reasons

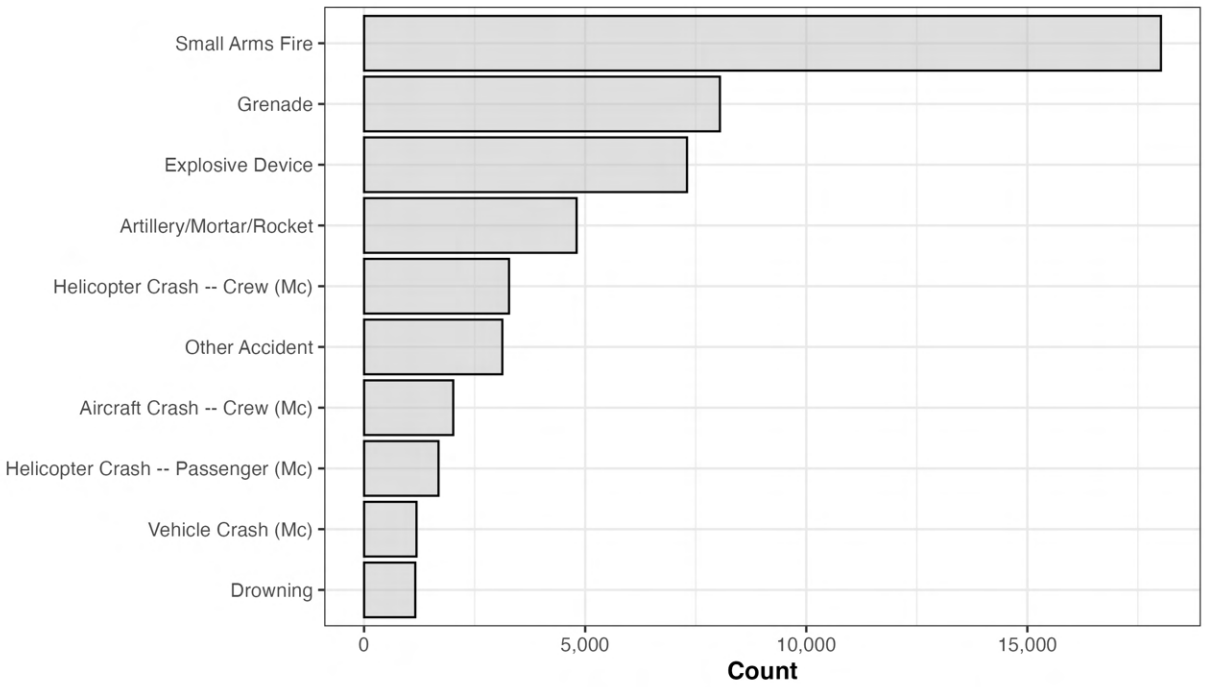


Figure S49: Count by State or Province

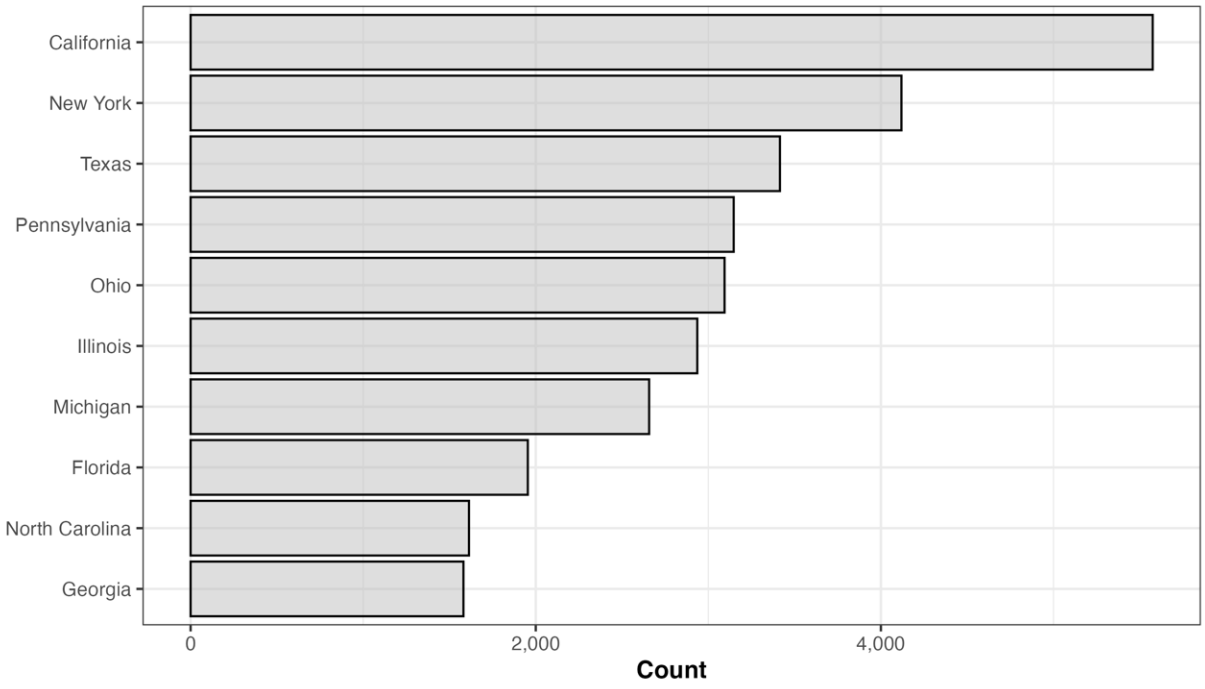


Figure S50: Count of Home Cities

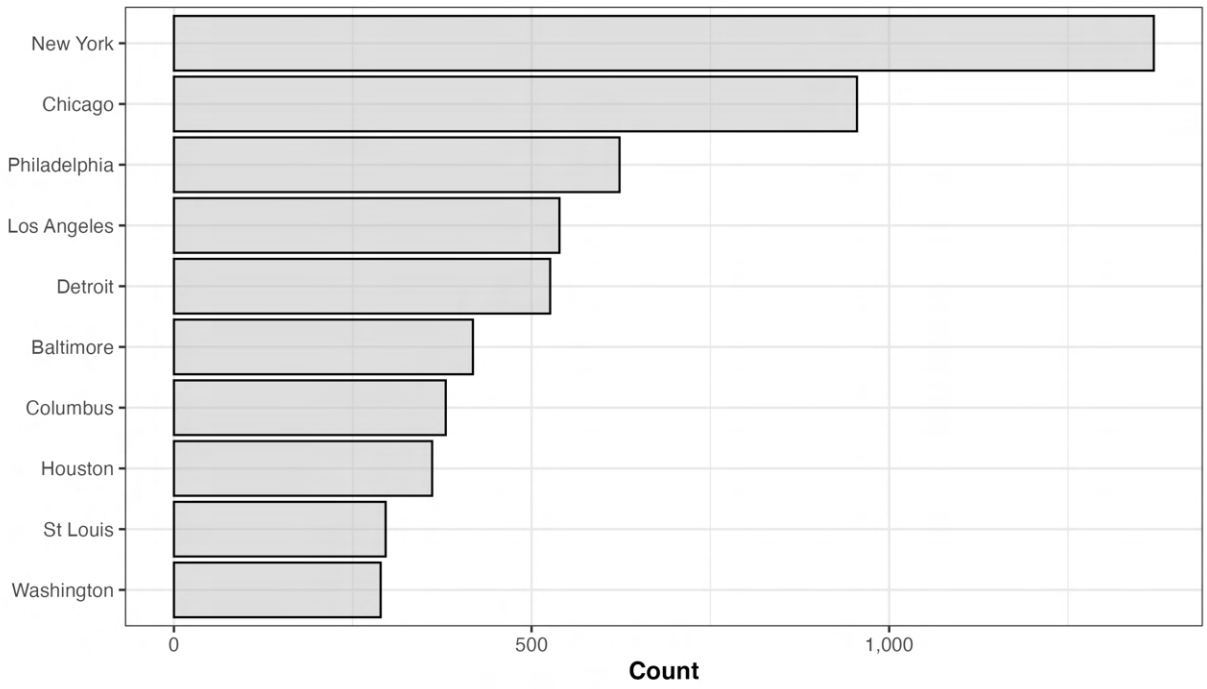


Figure S51: Count of Military Occupations

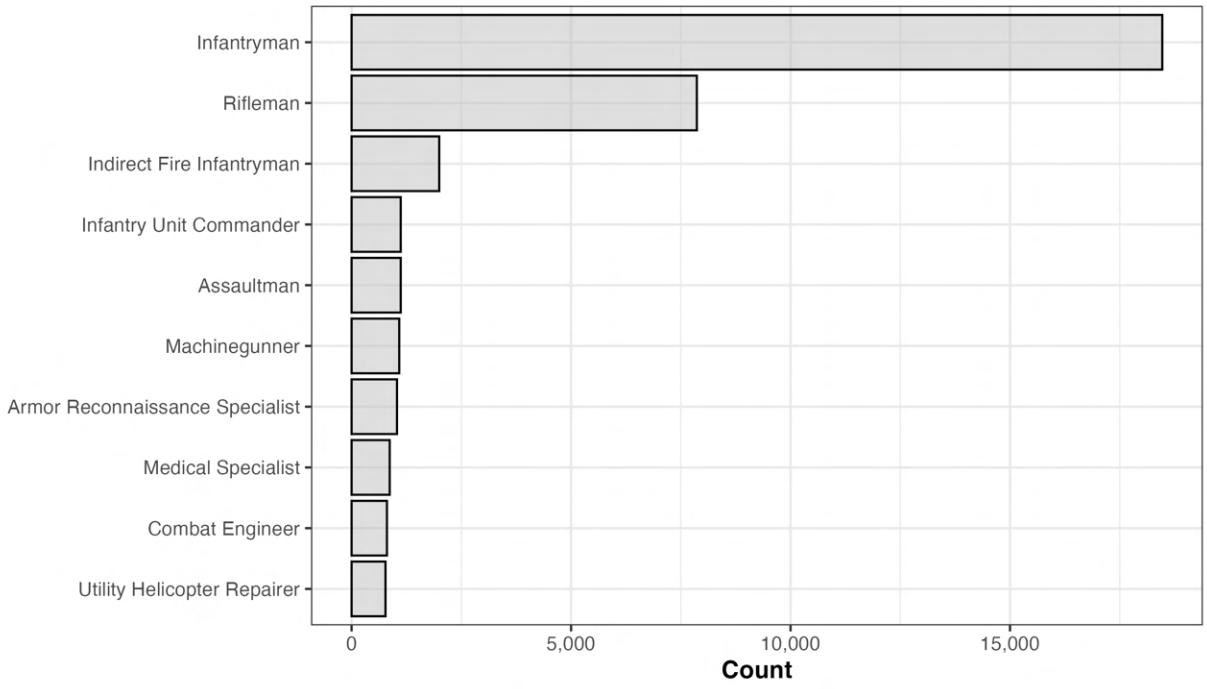


Figure S52: Count of Personnel Types

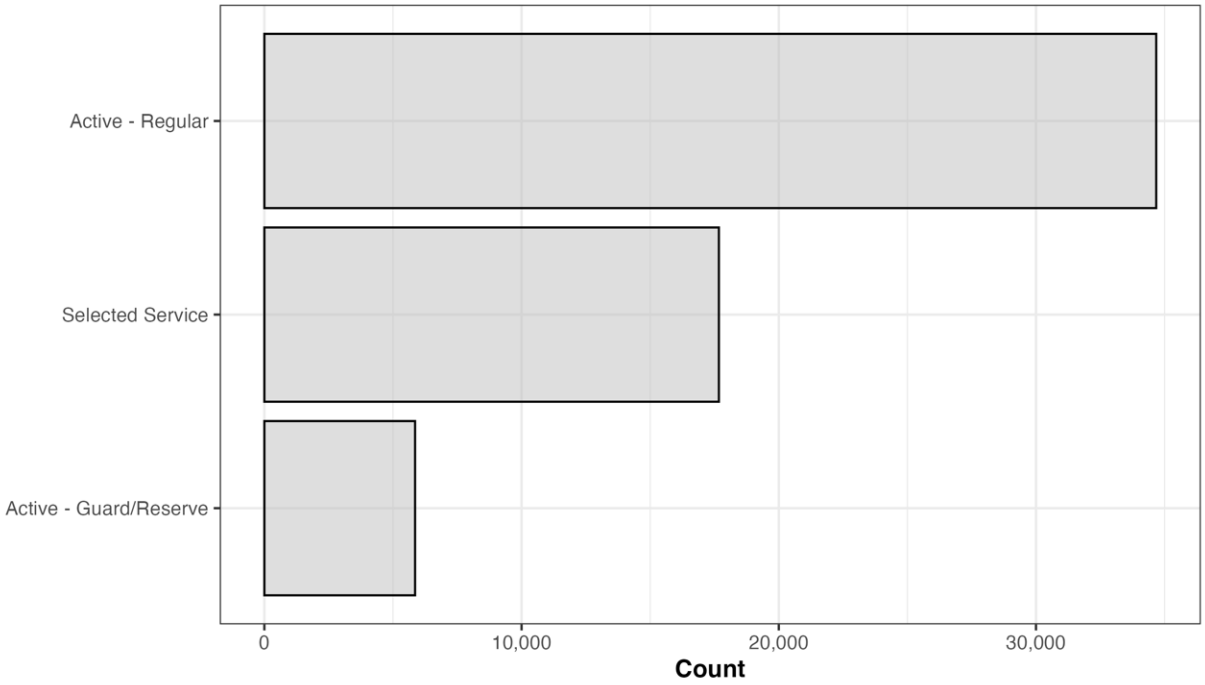


Figure S53: Count by Race

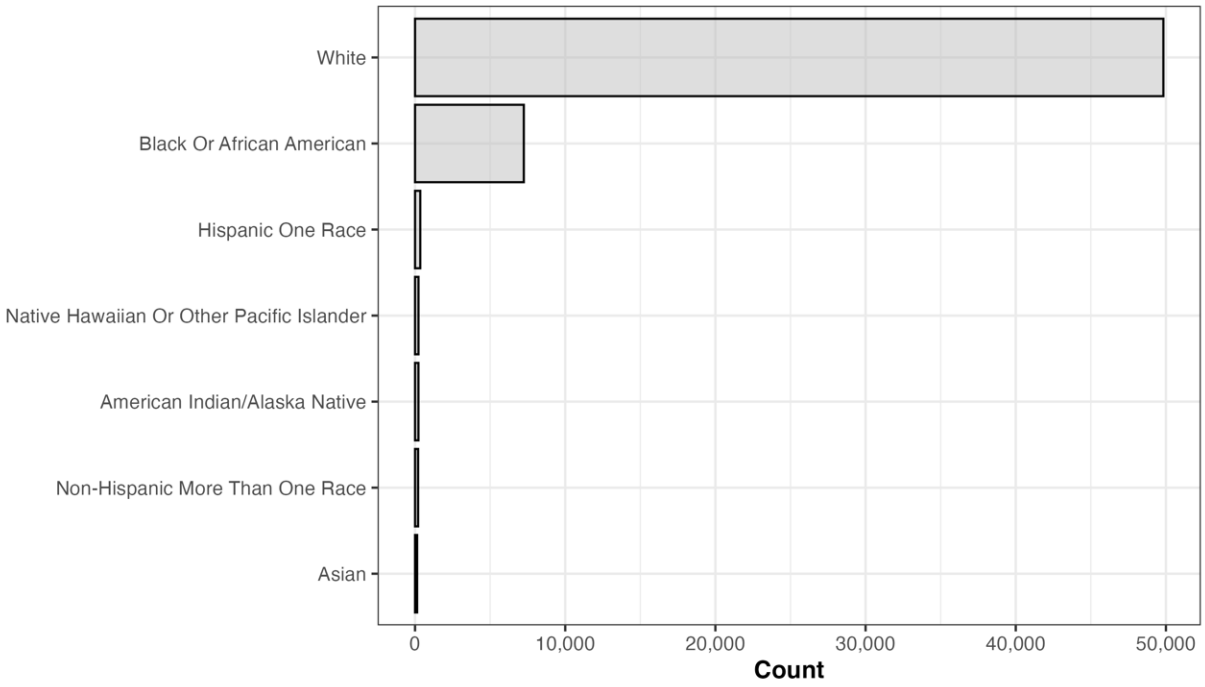


Figure S54: Count Per Military Rank

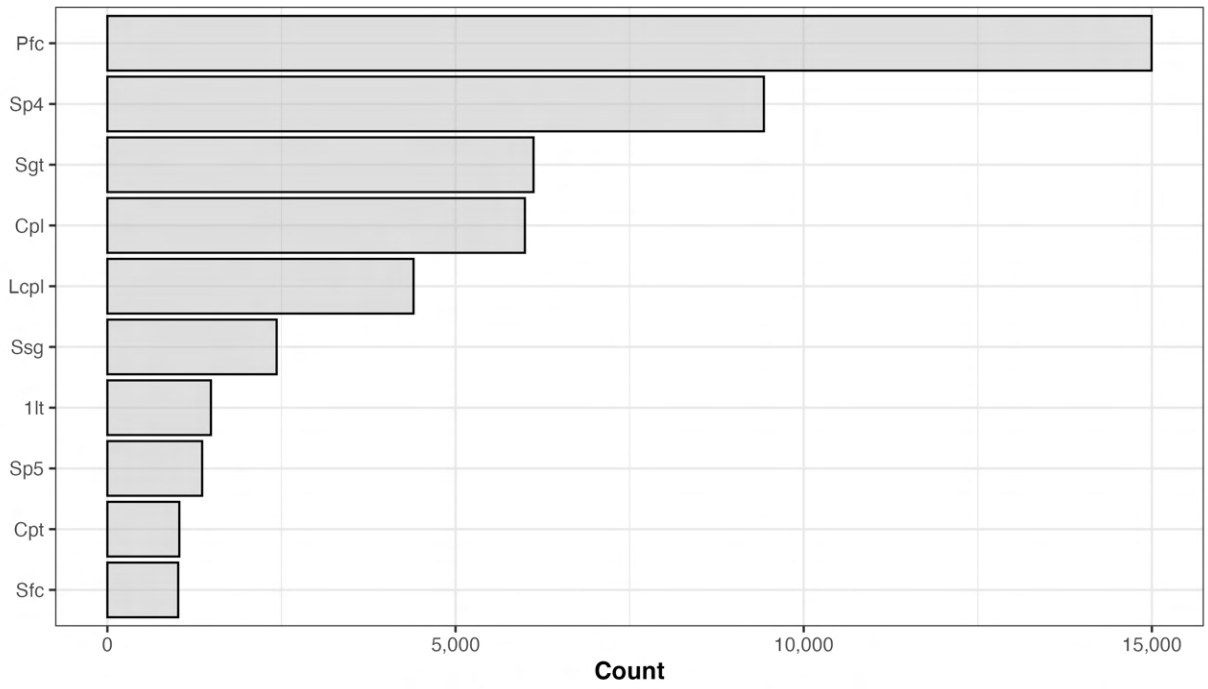
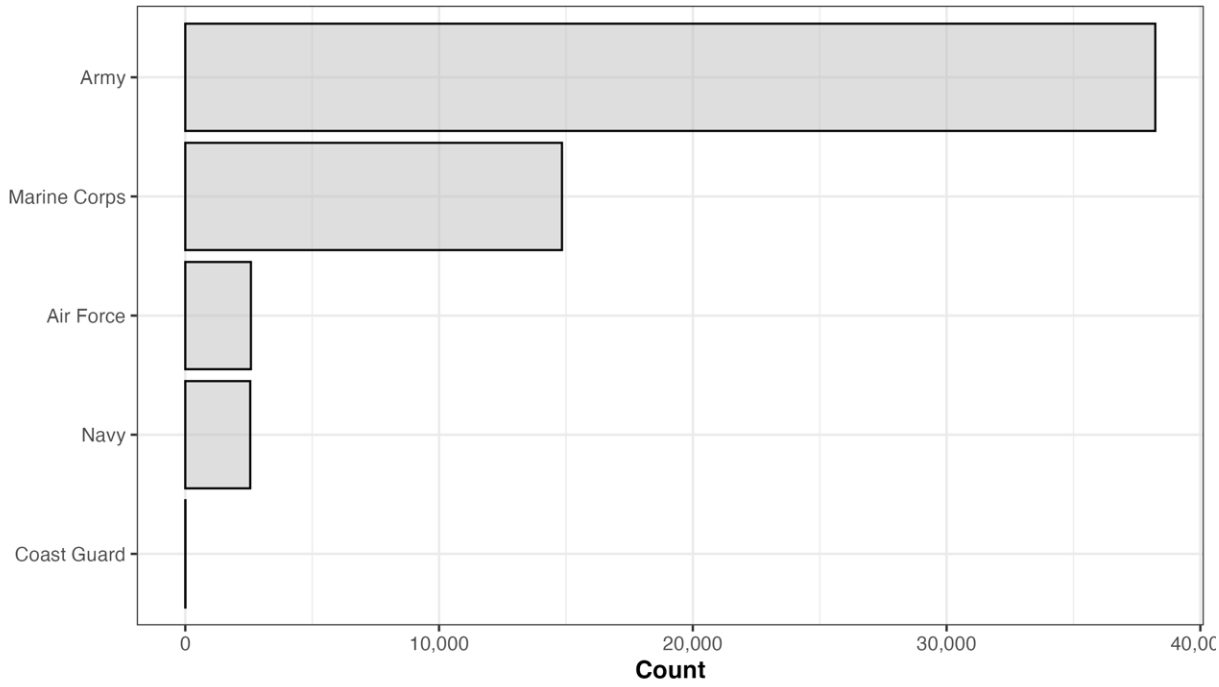


Figure S55: Count by Service Branch



9 SEAFA Files

The following figures provide an overview of the Southeast Asia Friendly Forces file. This data contains monthly information on the identity and location of American, South Vietnamese, and Allied maneuver battalions deployed in Southeast Asia during the Vietnam War. Unit locations were recorded at the end of each month. Figure S58 plots changes in the station location of the US Army's 11th ACR to demonstrate how the SEAFA files can be used to track the monthly locations of maneuver battalions.

Figure S56: The Location of Every SEAFA Station

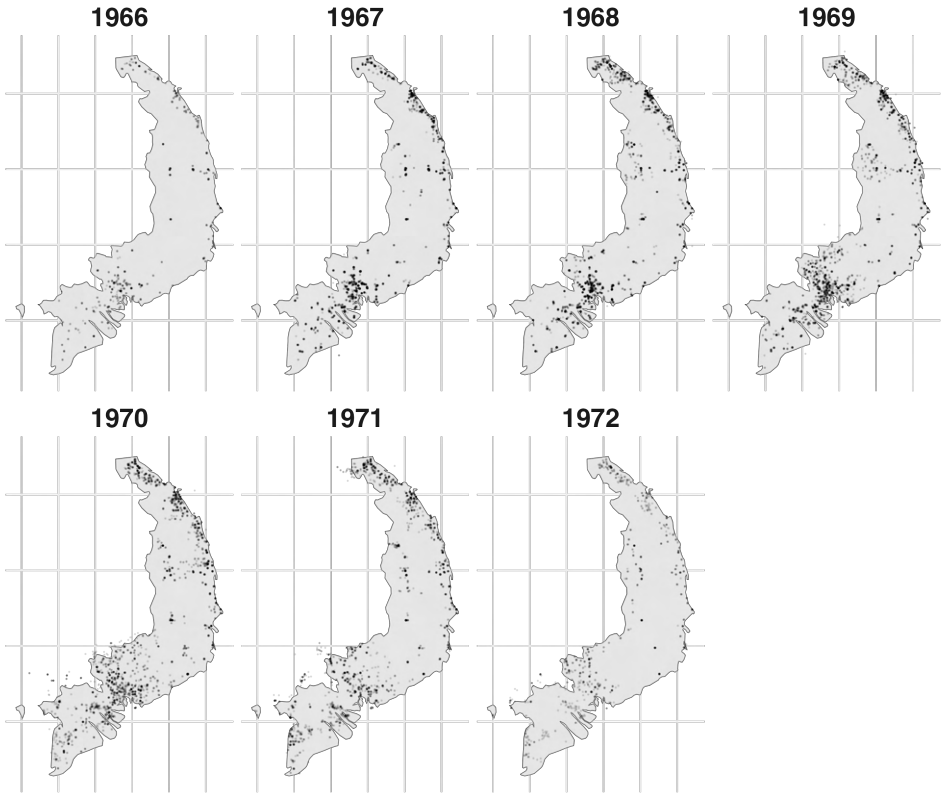


Figure S57: The Twenty Most Frequently Recorded SEAFA Stations

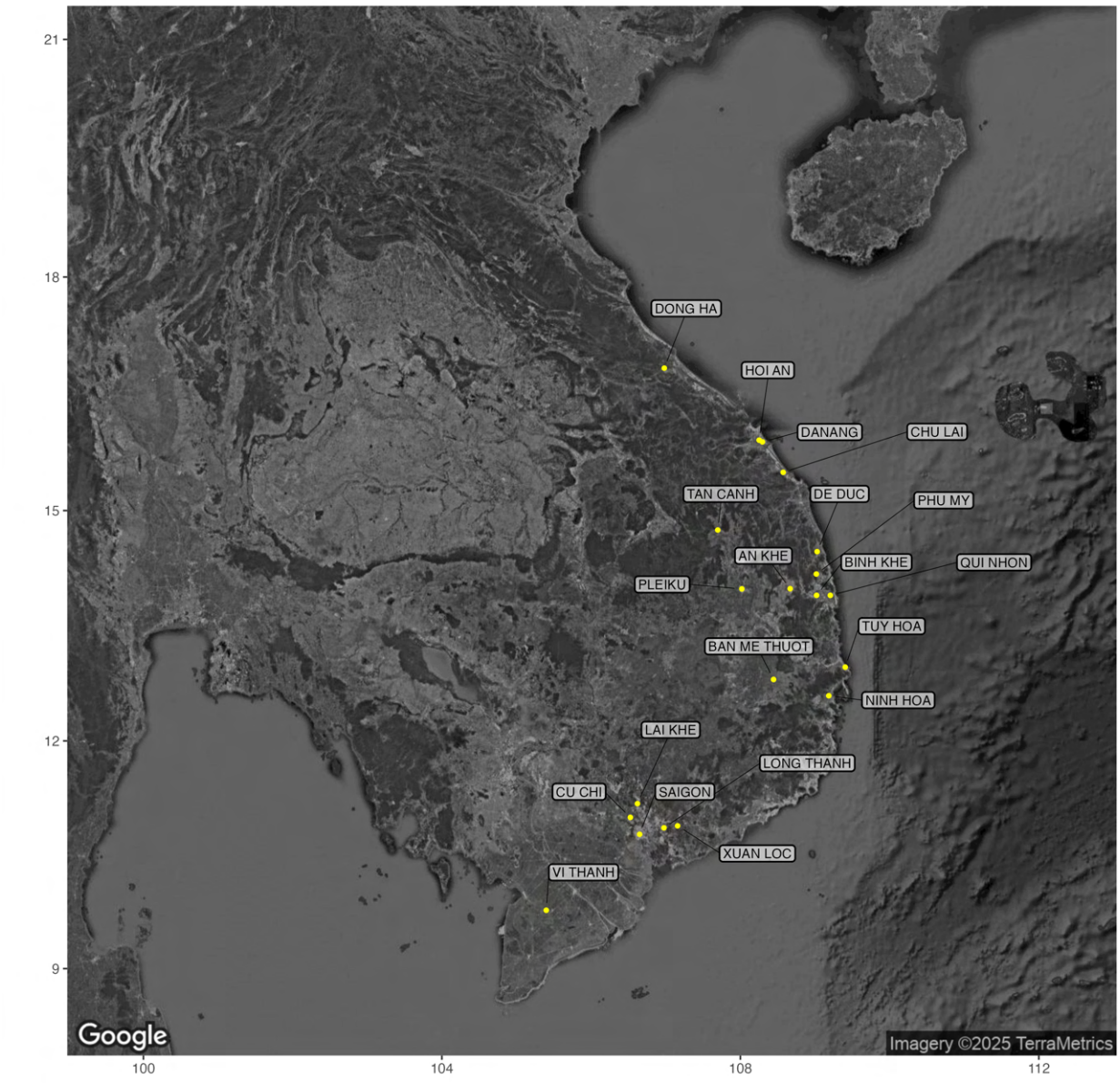


Figure S58: The Path of the US Army's 11th Armored Cavalry Regiment

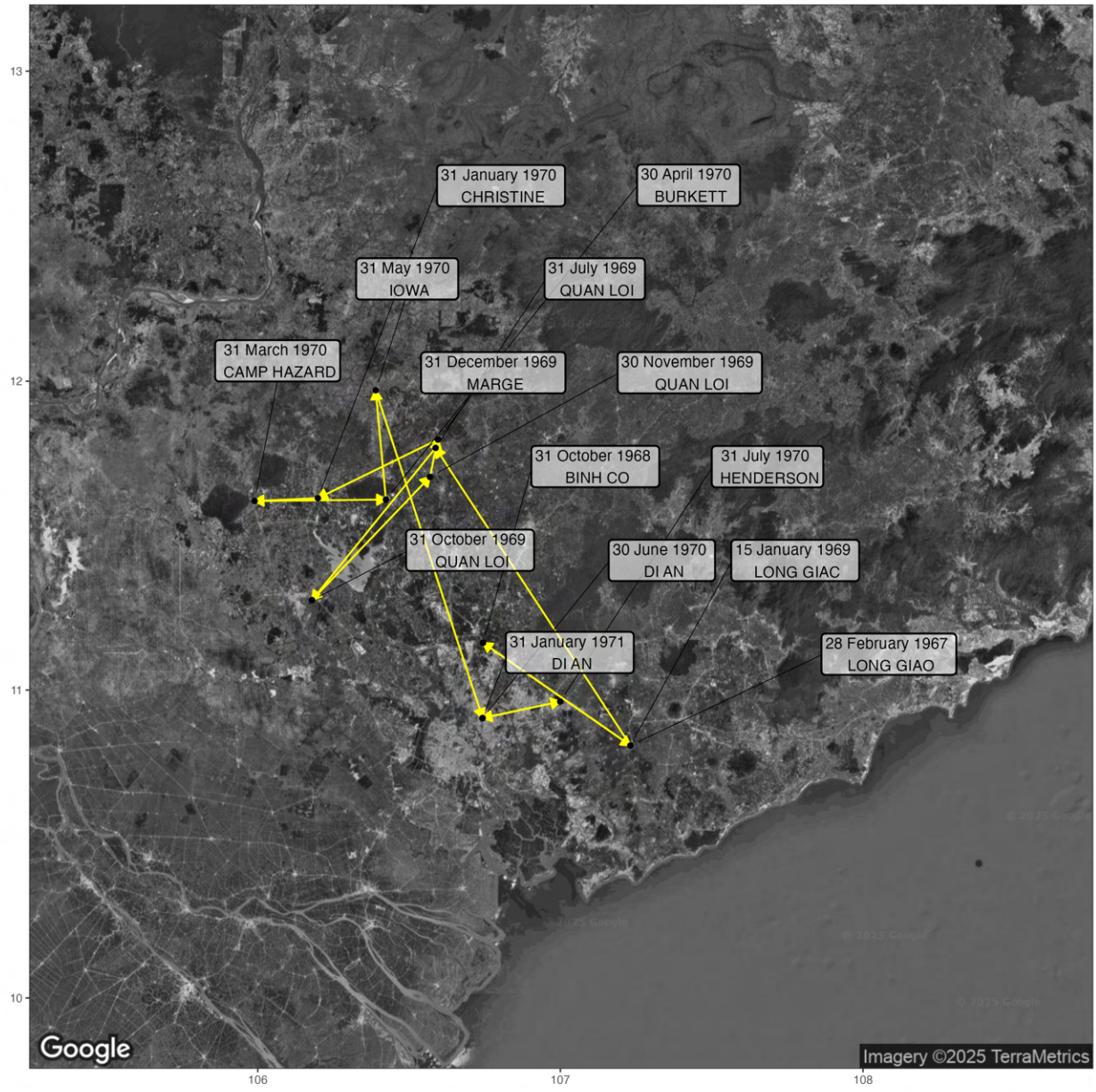


Figure S59: The 30 Most Common Battalions

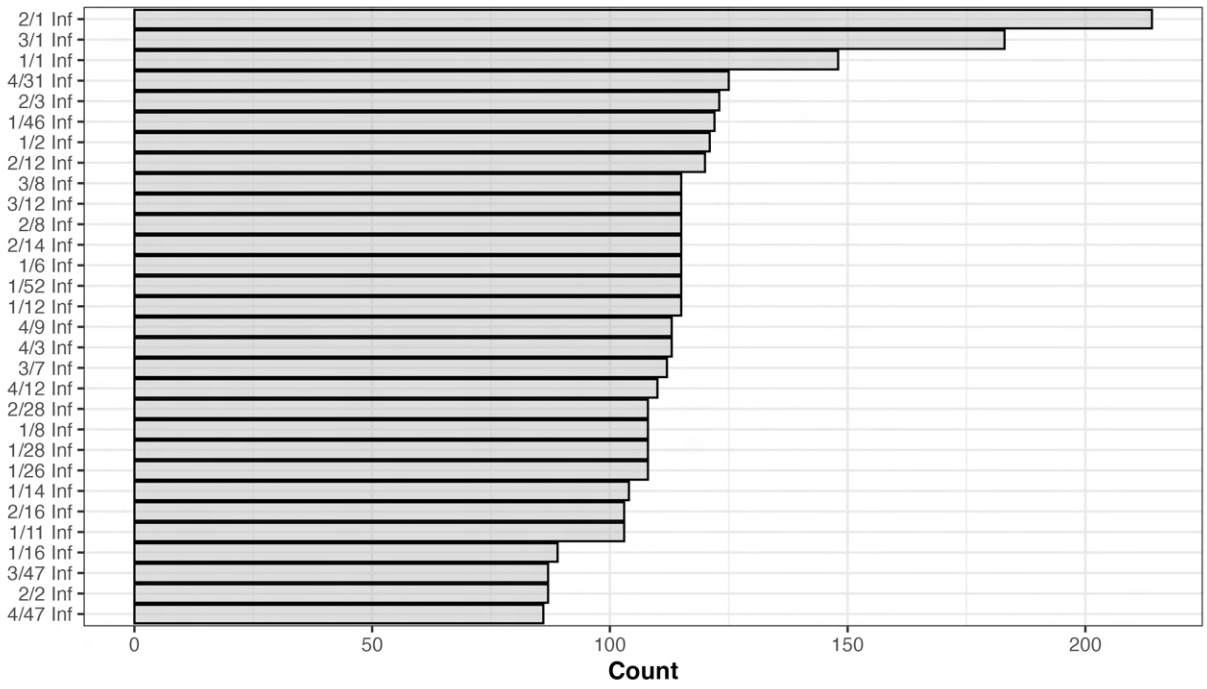


Figure S60: The Total Number of Battalion-Months Per Country

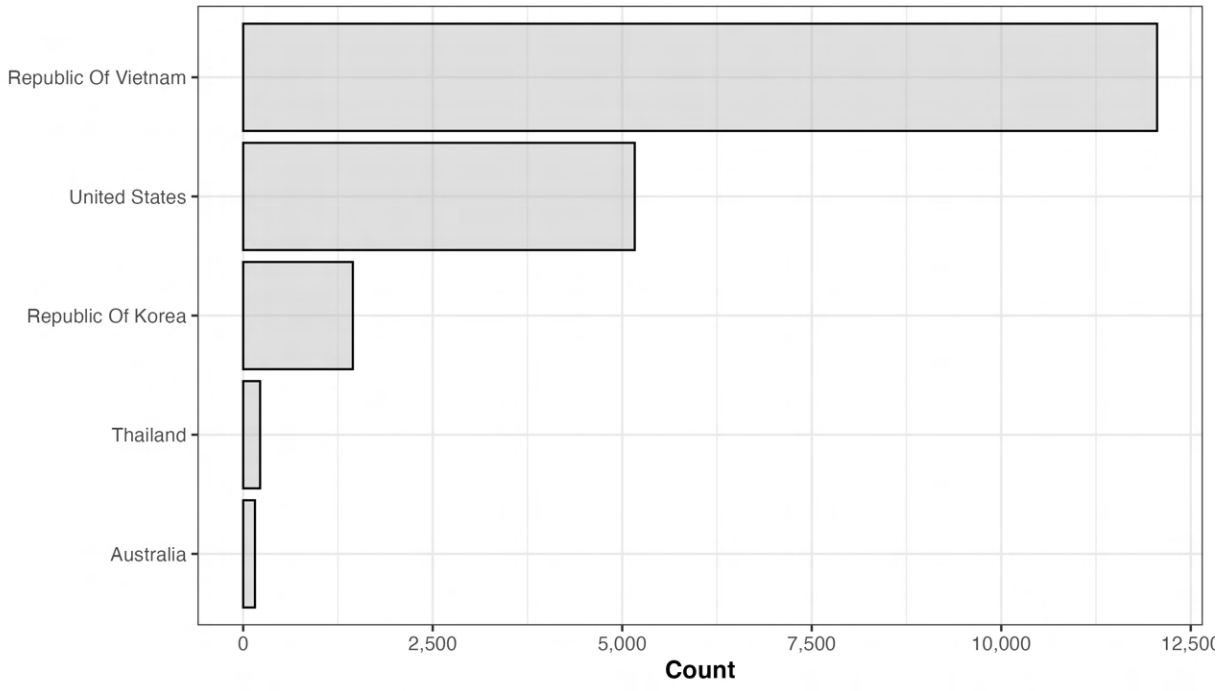


Figure S61: The Count of Battalion Codes

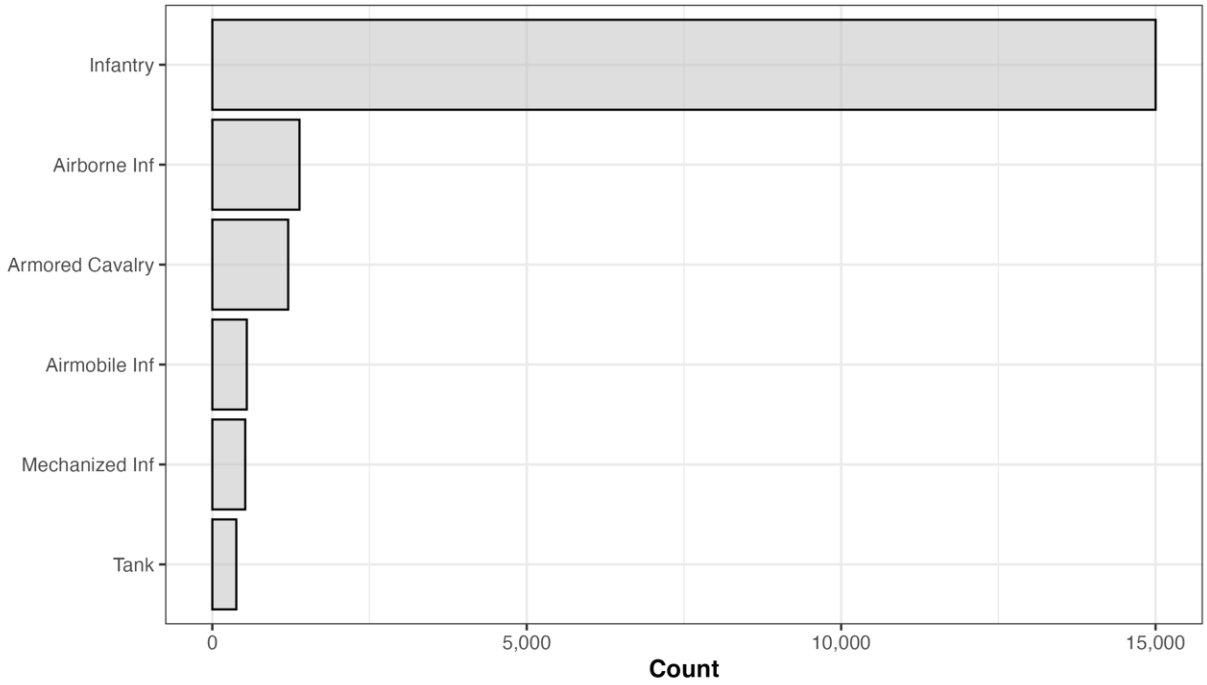


Figure S62: The Count of Headquarter Units

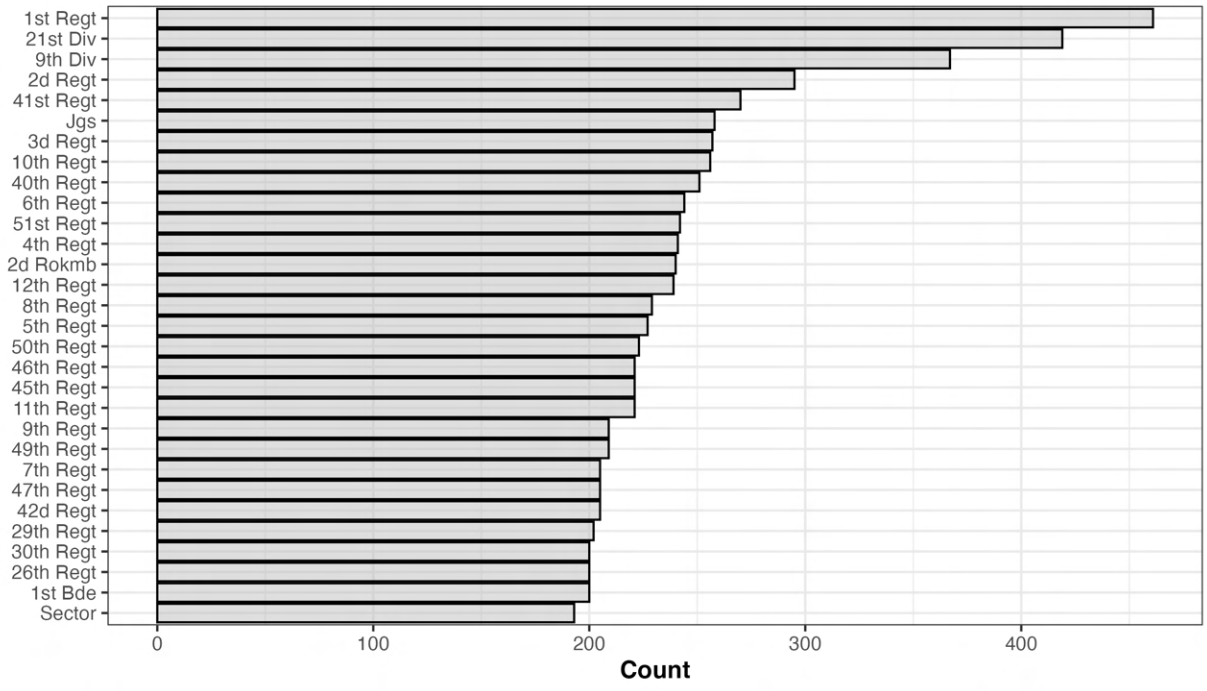


Figure S63: The 30 Most Common Stations

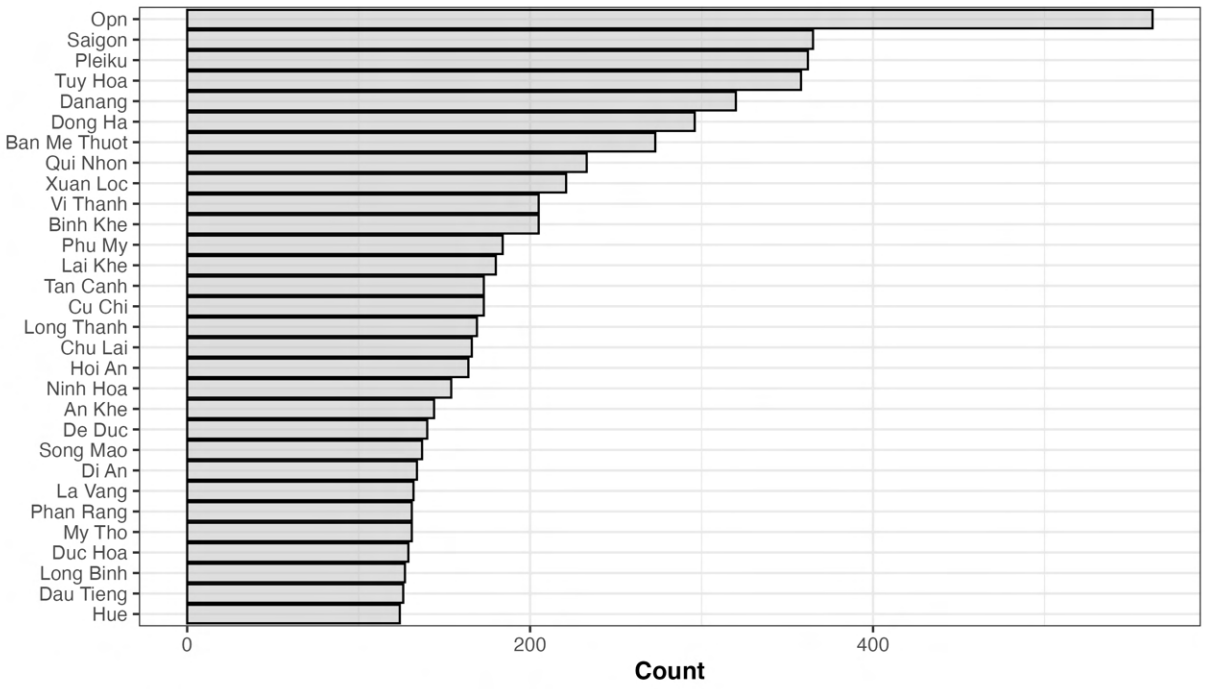
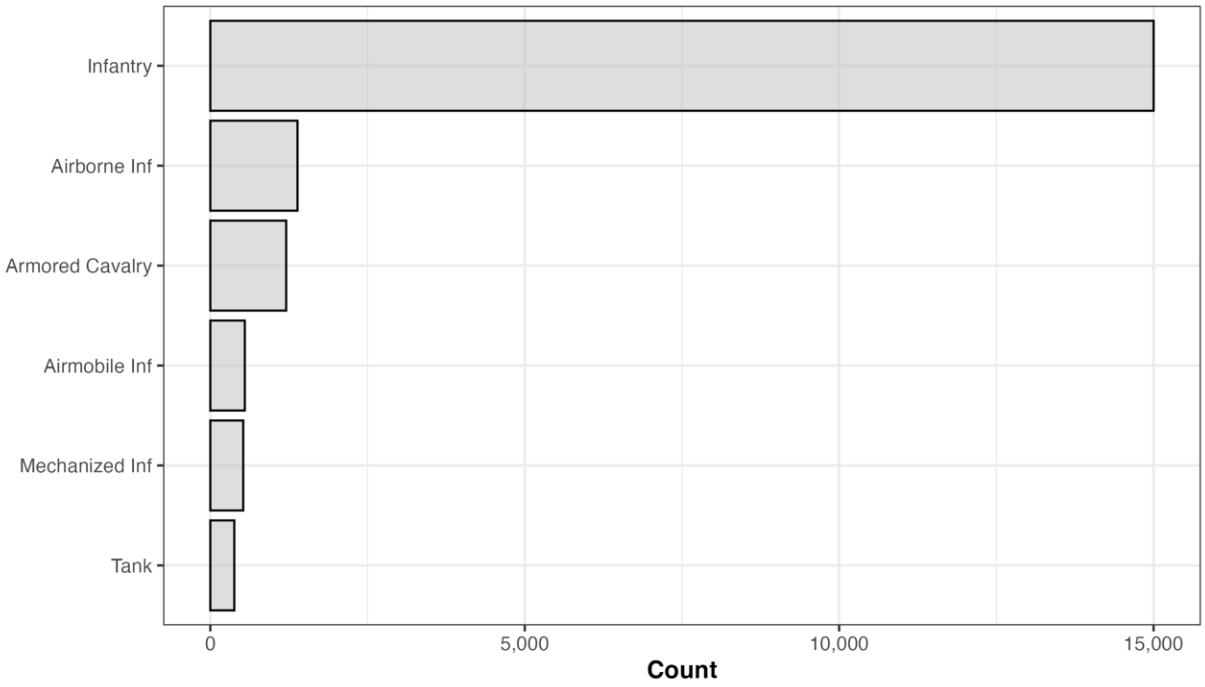


Figure S64: The Count of Unit Sub-Types



10 BASFA Files

The following figures provide an overview of the Enemy Base Area File. These base areas were areas where the VC insurgents and NVA forces took refuge. Consequently, they were frequently the targets of major Allied ground operations. For example, War Zone C, which was located in South Vietnam's Tay Ninh province between Cambodia and Saigon, served as a stronghold for the Viet Cong (VC) and North Vietnamese forces (Hickey, n.d.). This location, combined with its extensive jungle terrain that provided cover for troops and supplies, made it ideal for launching attacks into South Vietnam and receiving supplies from the Ho Chi Minh Trail. Due to its strategic importance, the zone housed important communist facilities, including the Central Office for South Vietnam (COSVN), which directed VC operations in the south. Not coincidentally, War Zone C became the target of one of the US Army's most extensive "search and destroy" missions as part of Operation Birmingham.

The BASFA files identify the location and center of mass of the North Vietnamese base areas in Southeast Asia. The black coordinate points represent the center of the base, while the yellow polygons represent the base area.

Figure S65: The Location of Every Recorded North Vietnamese Base Area

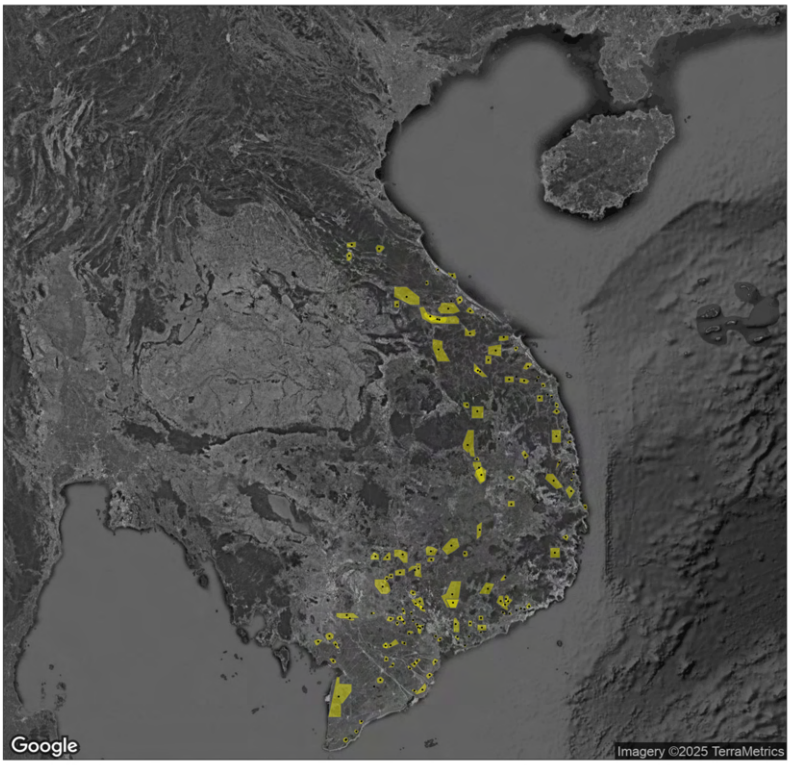
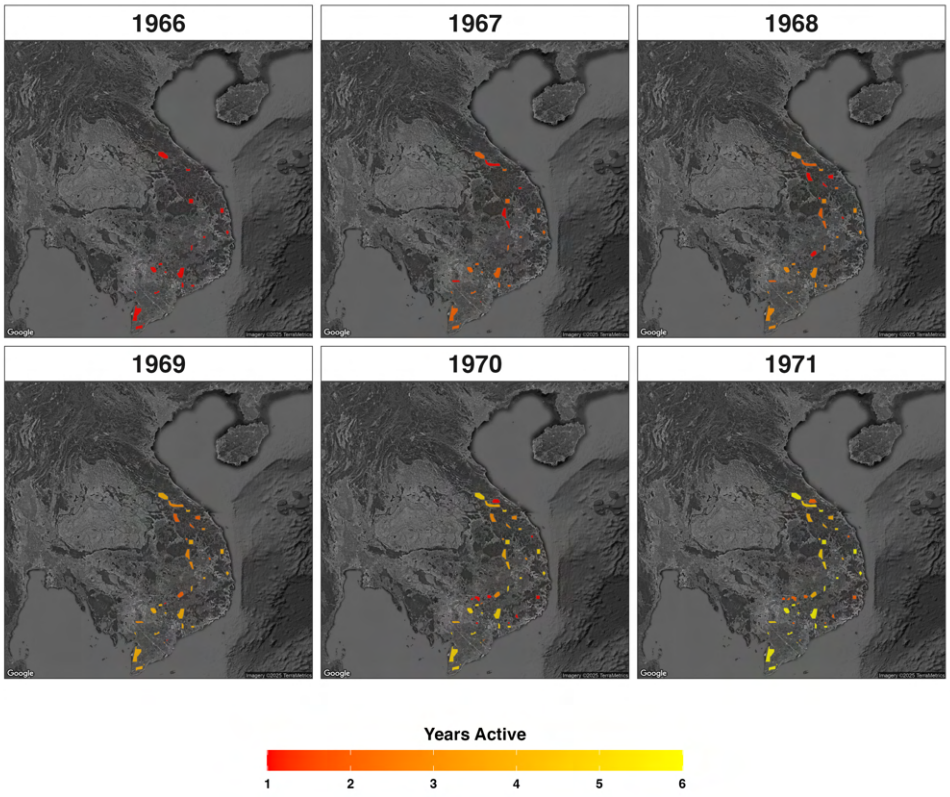


Figure S66: The Age of Each Active North Vietnamese Base



11 HES Files

The following figures provide an overview of the Hamlet Evaluation System. The HES system was designed by analysts from the Office of the Assistant Secretary of Defense to allow military commanders from the United States and the Government of the Republic of Vietnam (GVN) to assess the ongoing war effort at the hamlet level, which comprised the smallest population units. The HES files contain monthly assessments of hamlet and village security that measured territorial control at the village level. These monthly reports documented security conditions, enemy presence, economic conditions, development scores, and estimates of government control for each of the 21,735 hamlets in the sample, enabling military commanders to track changes in territorial control over time.[1]

[1]According to the Civil Operations and Rural Development (CORDS) Support Research and Analysis Directorate, the formal definition of the term hamlet varies slightly across South Vietnam depending on the actual area of interest (Civil Operations and Rural Development Support, 1971). In some districts, hamlets are identified as geographic areas with surveyed boundaries that are contiguous throughout a village. In other areas, hamlets are identified only as an area of a massed population without clearly defined boundaries." [2] Moreover, "hamlets exist for internal village administration, and the number of hamlets within a specific village may fluctuate over time according to its administrative needs."[3] In contrast to hamlets, villages

The Hamlet Evaluation System actually refers to three related efforts to estimate progress at the local level that include the Hamlet Evaluation System Files (HAMLA), Hamlet Evaluation System 1970 (HES70), and Hamlet Evaluation System 1970 (HES71) systems. Over time, members of the MACV introduced new and improved surveys to identify broad patterns of public support. Consequently, the HAMLA files represent a comparatively basic and more primitive attempt to assess local levels of territorial control at the hamlet and village levels in South Vietnam. Initially, efforts to evaluate territorial control in South Vietnam placed an inordinate emphasis on measures of enemy military activity. Later iterations added more nuanced measures of socioeconomic factors. The HES70 and HES71 files are similar in construction and can be seamlessly combined into a comprehensive data file. The differences between the HAMLA and HES files are more substantial. Consequently, the cleaned versions of the HAMLA and HES files are preserved separately.

In practice, each evaluation system relied on army officers, typically majors or higher, and local district and provincial leaders to answer monthly and quarterly questions about the security, political, and economic conditions of Vietnamese villages and hamlets. In the HES71 files, for example, these officers had to enter responses to 169 questions. These questions were then classified into 19 submodels that estimated social, political, and military control, and Bayes' rule was used to aggregate responses within each submodel into a continuous score ranging from A (very secure) to E (least secure).[4] Similarly, the assessors also documented whether the hamlets were entirely controlled by the Vietcong (denoted with a V) and instances in which a village was not evaluated during a particular month (denoted with an N). Of particular interest are MOD5A, which models enemy military control of a hamlet; MOD6A, which models political control; MOD7B, which models community and economic development; and most importantly, MOD8, which is the aggregate HES rating that aggregates across the 19 submodels.

These files identify the locations of hamlets and villages in South Vietnam, along with security, socioeconomic, and political evaluations for each location. The following figures represent a small proportion of the included information. These figures show the monthly count of measured hamlets, the average monthly security score per military region, and the locations of each hamlet recorded in the data files. There are several obvious problems

typically have "surveyed boundaries and form contiguous coverage of the area of each district in the country." The average Vietnamese village contains five hamlets, though the number can range from one to more than 25. A legal entity with administrative autonomy, the village is the basic administrative unit in rural areas of Vietnam.

[4]Dell and Querubin (2017)

with the HES files. The first is that there are several periods with missing observations. These issues appear to be the result of data preservation issues at NARA. In addition, the coordinates listed for some villages and hamlets in the HES files are clearly incorrect. The most obvious example of incorrect coordinate values is for the hamlets located over the ocean that appear in Figure S69.

Figure S67: The Count of Hamlets Assessed Per Month

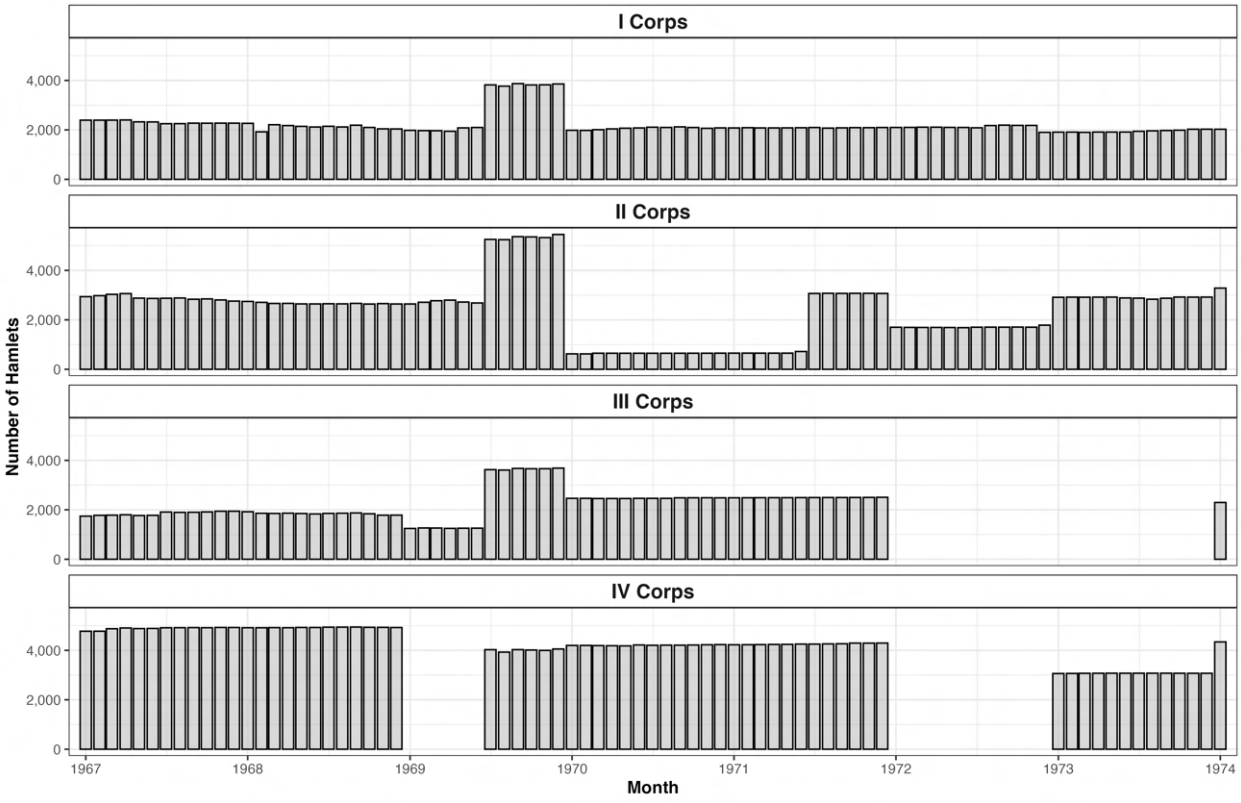


Figure S68: Average Hamlet Classification Score by Military Region

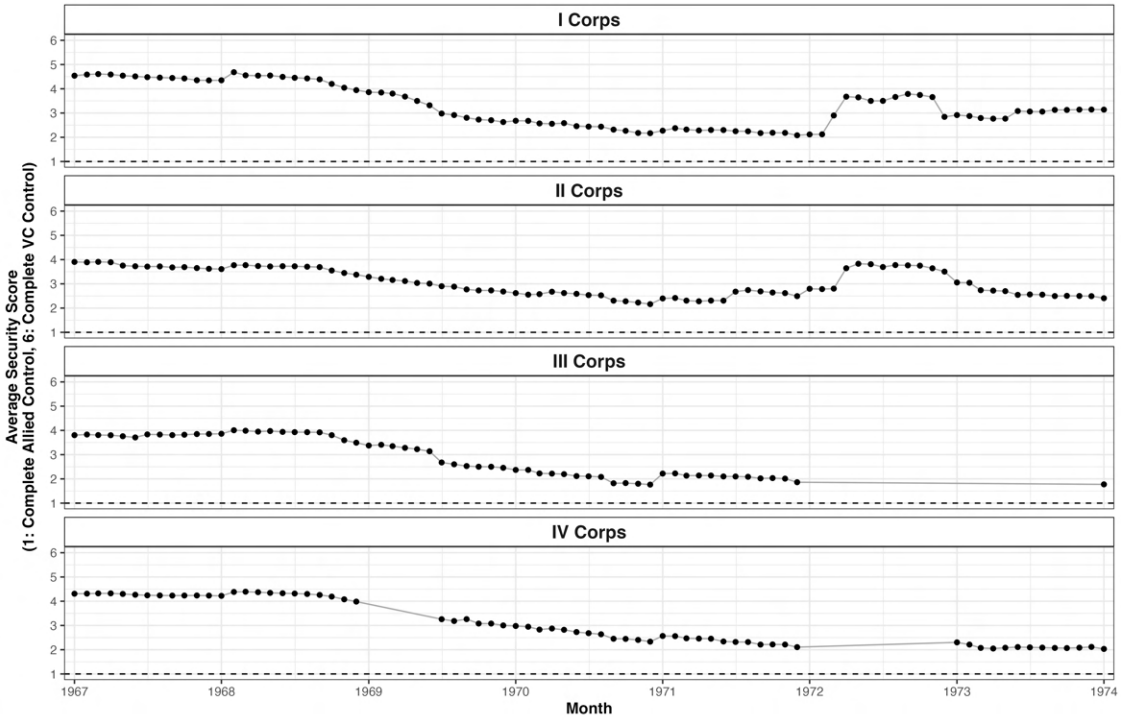
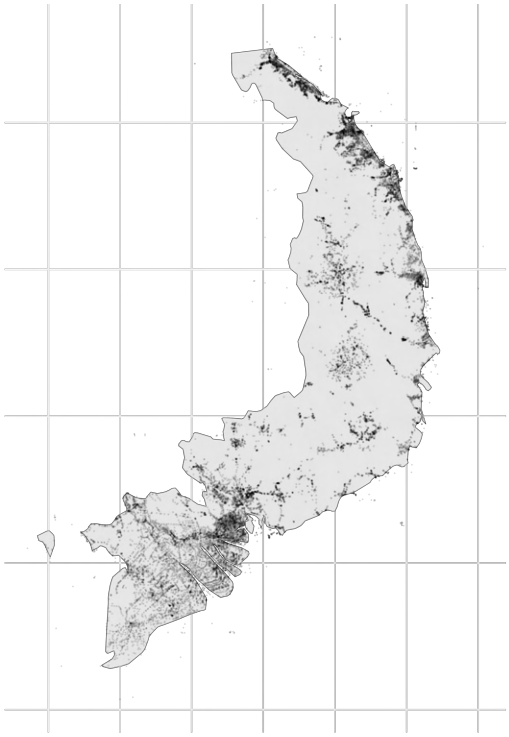


Figure S69: The Coordinates of Each Hamlet in the HES Files



12 SORTIEA Psychological Operations Files

The following figures provide an overview of the SORTIEA files. These files contain information about aerial activity associated with psychological operations, including the date of the operation, aircraft type used, flight hours, and the amount of leaflets, newspapers, and magazines dropped. The records include the date of the operation, the target audience, the number of brochures, posters, and publications distributed, the hours of radio and loudspeaker broadcast, the hours of transmission of television, and the demographics of the audience. The observations in the SORTIEA files appear to be updated daily. The information can be paired with the HES files to add additional historical context.

Figure S70: Aerial PSYOP Weekly Incident Count (1969-1973)

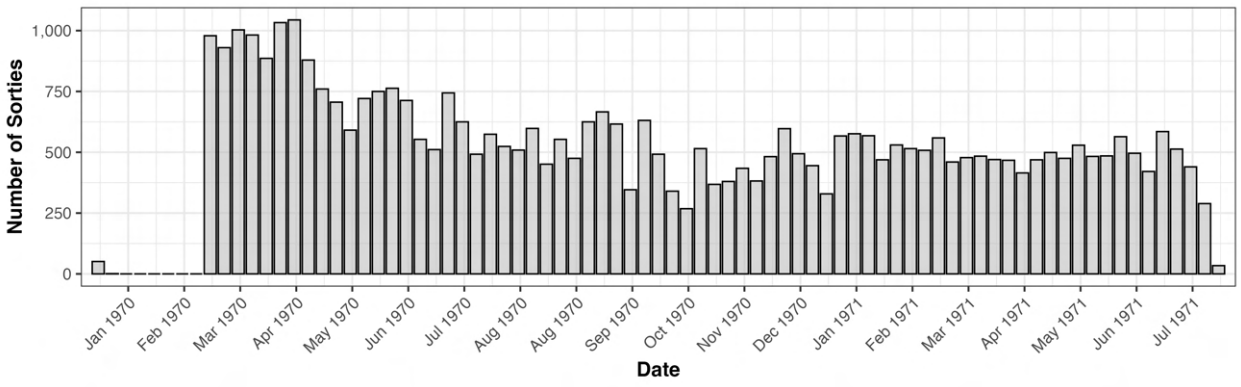


Figure S71: Aerial PSYOP Conducting Agency Count

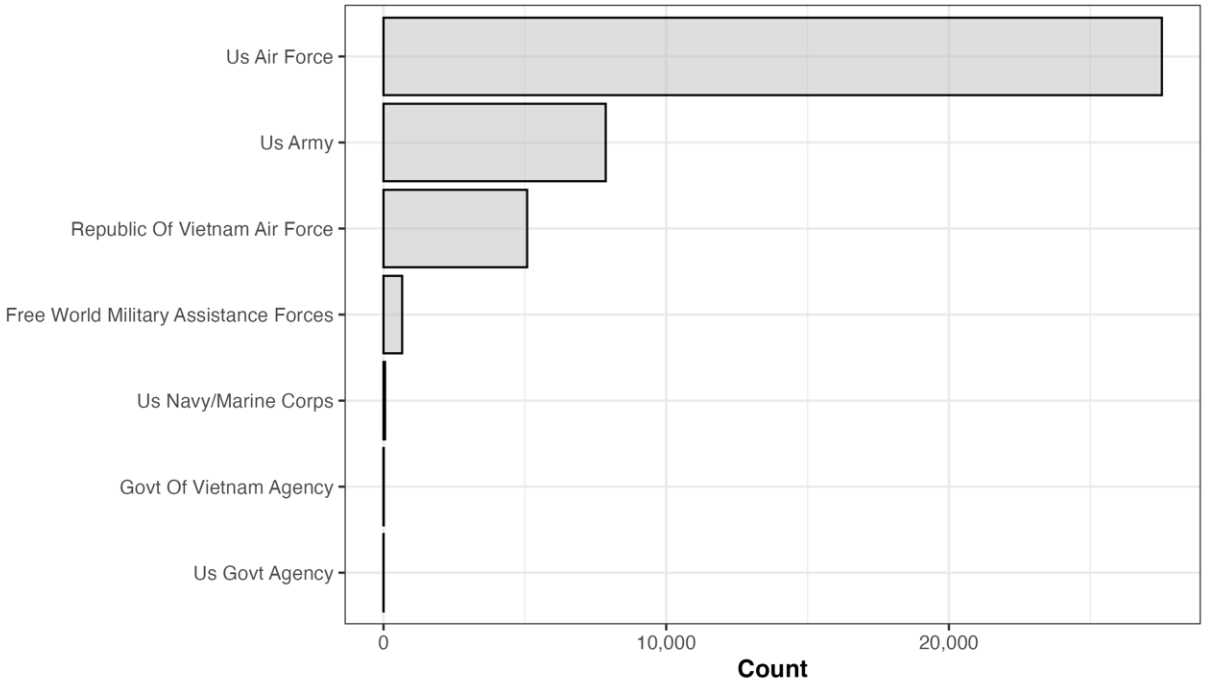


Figure S72: Aerial PSYOP Aircraft Type Count

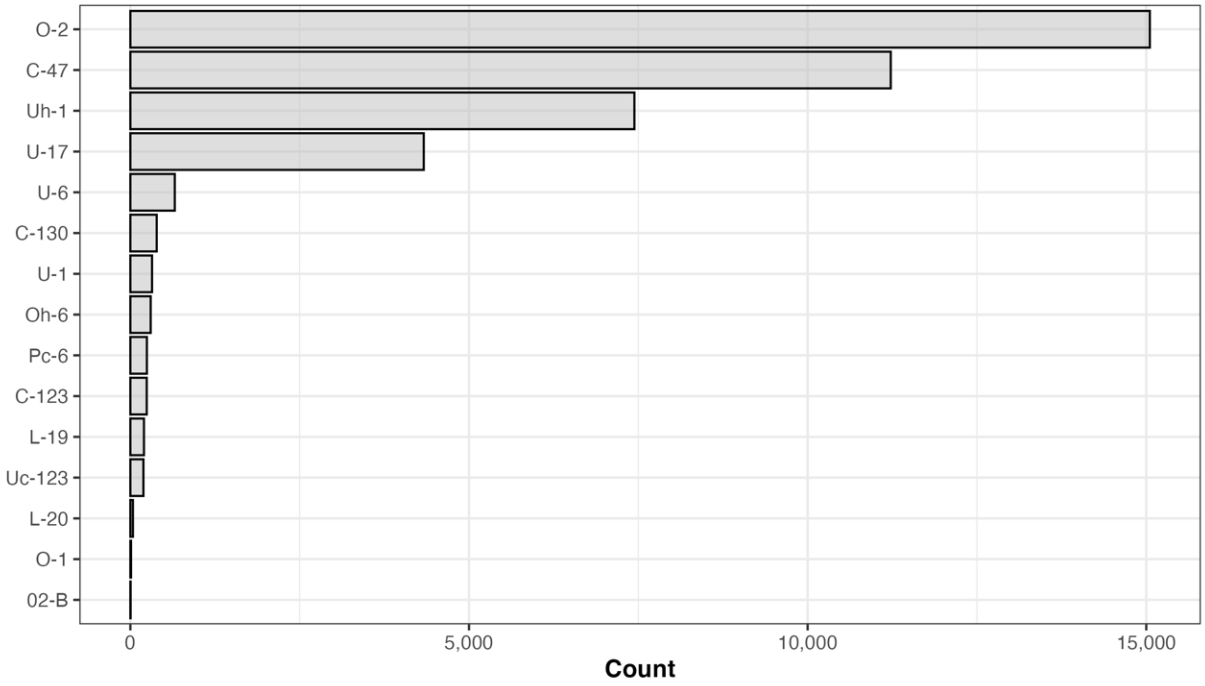


Figure S73: Aerial PSYOP Operation Type Count

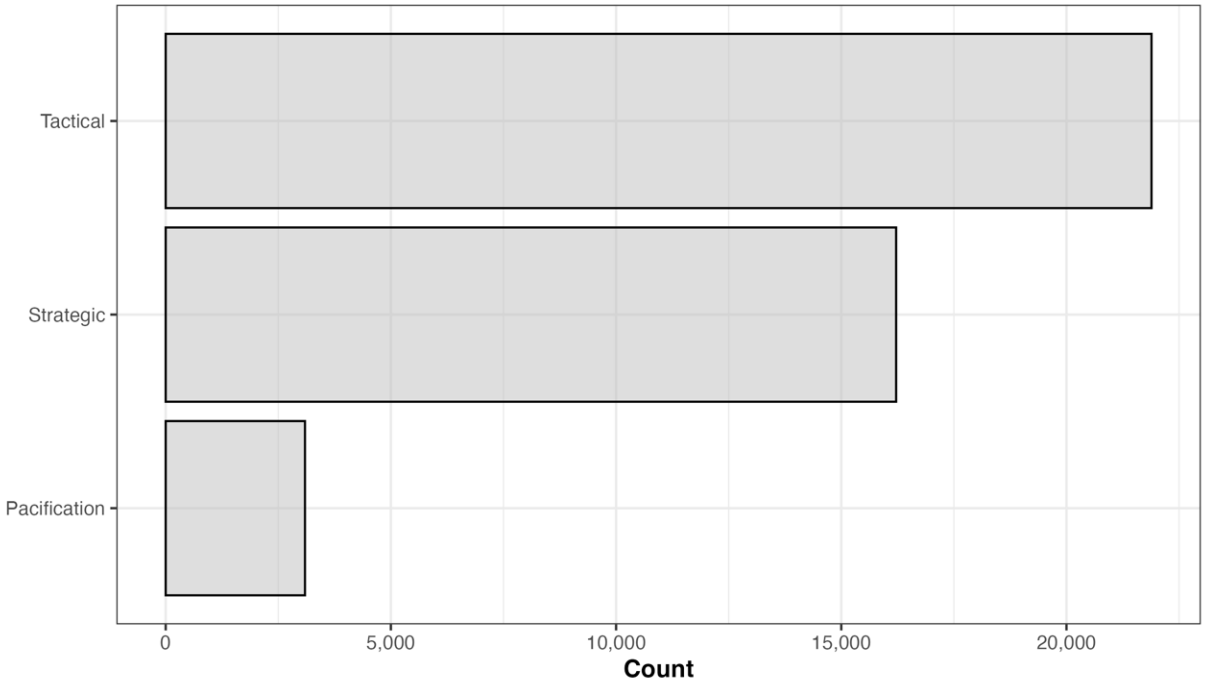


Figure S74: Aerial PSYOP Campaign Type Count

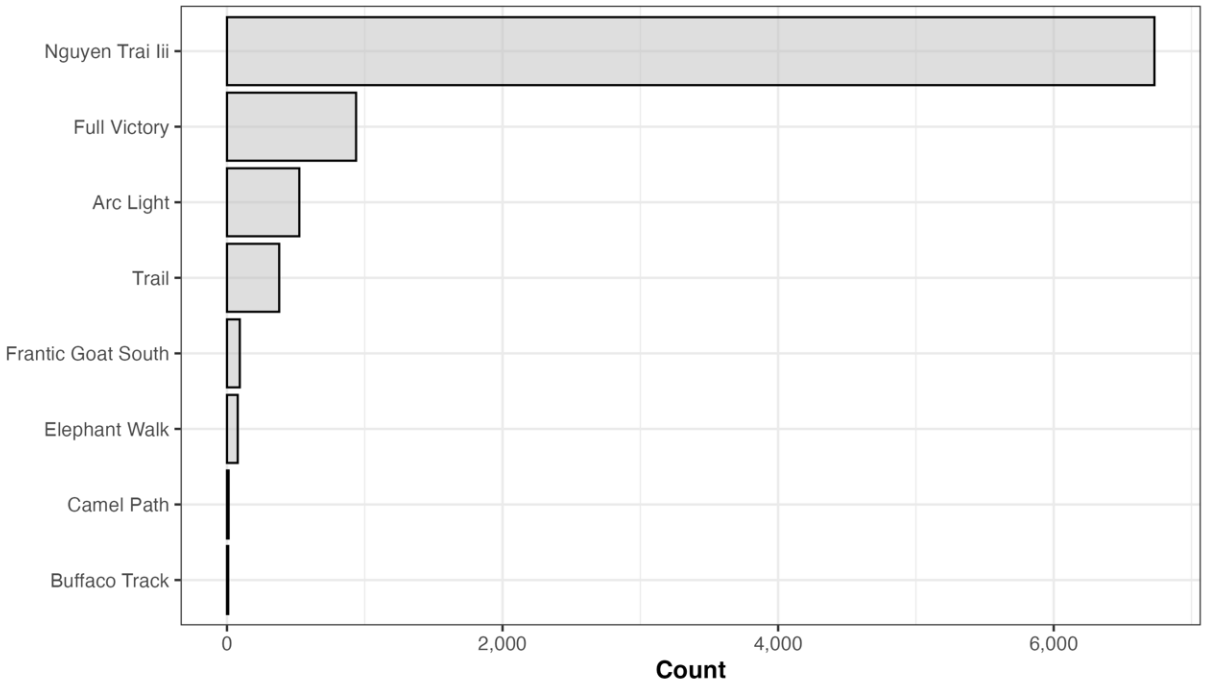


Figure S75: Aerial PSYOP Operation Theme Count

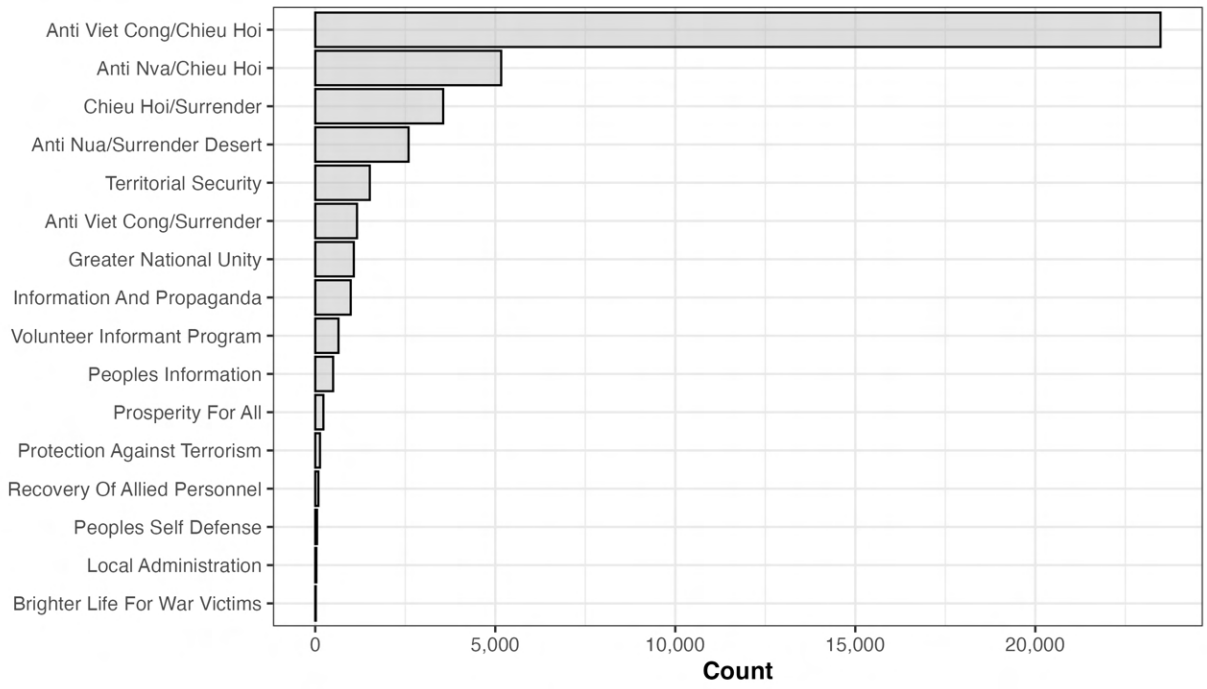


Figure S76: Aerial PSYOP Operation Sub-Theme Count

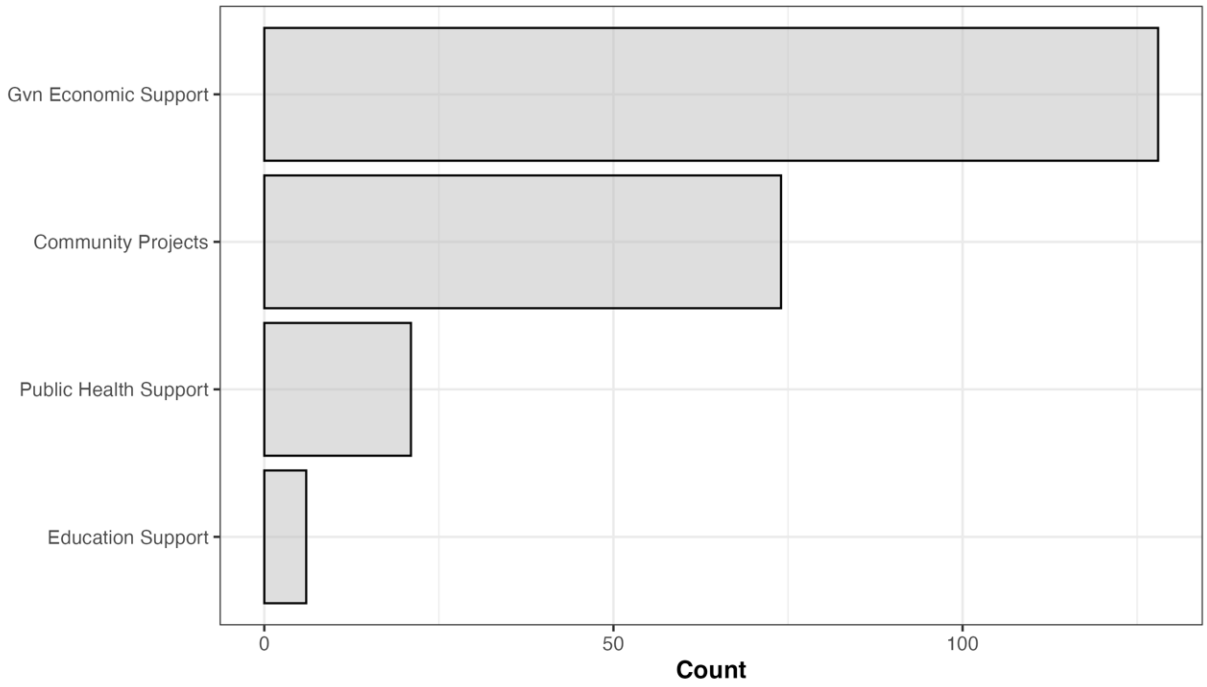


Figure S77: Aerial PSYOP Distribution of Numeric PSYOPS Columns

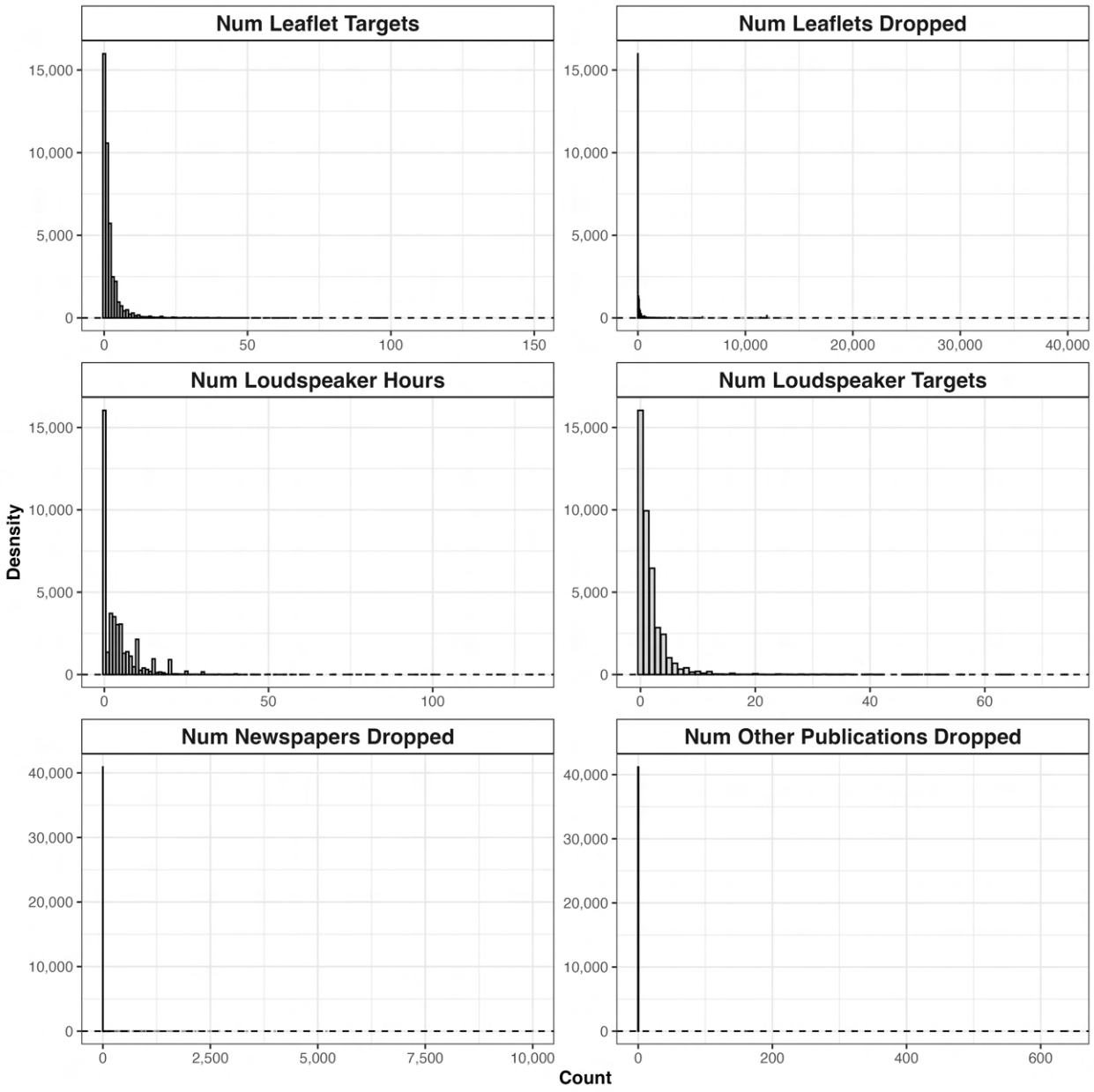
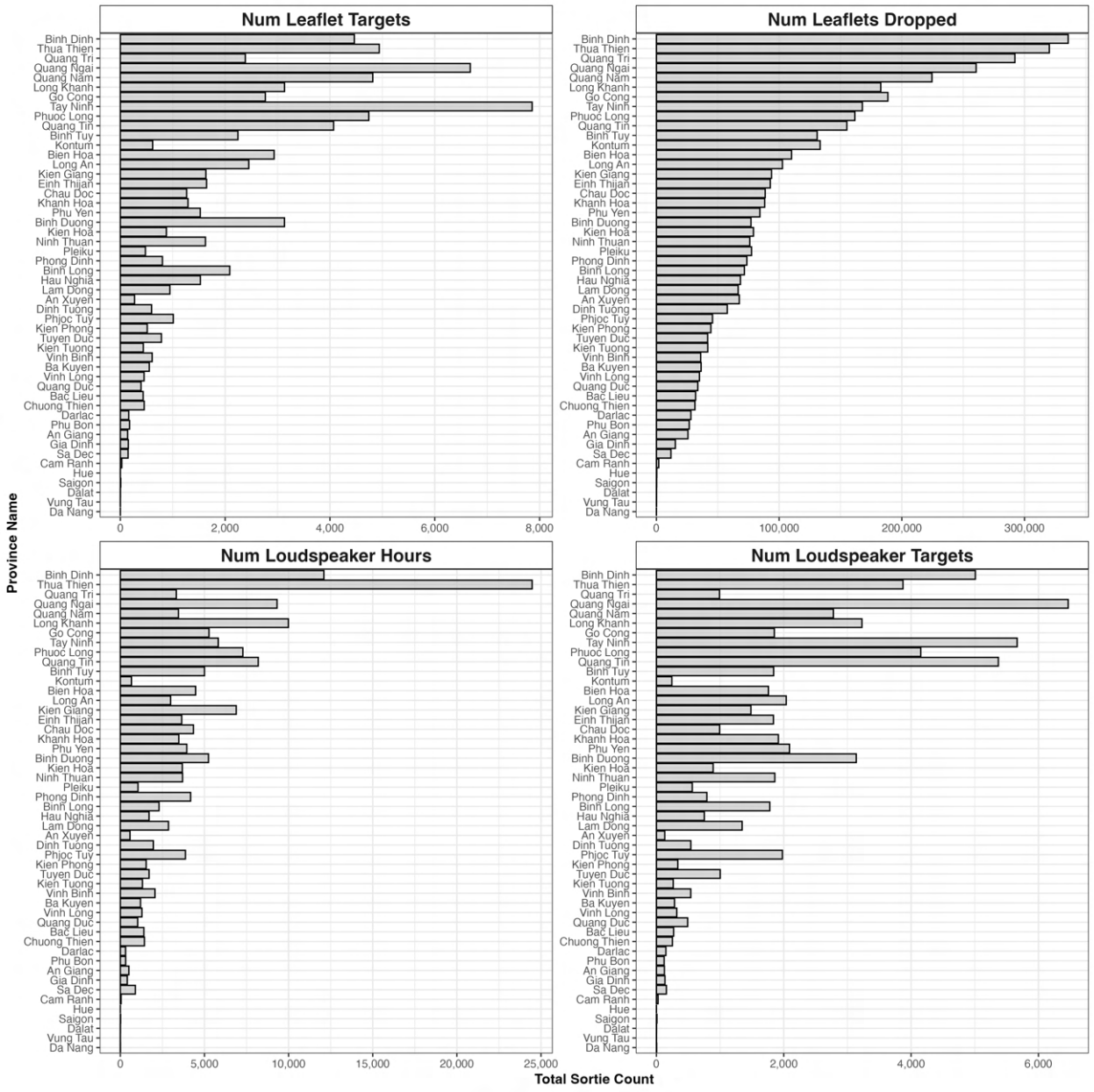


Figure S78: Aggregate Aerial PSYOP Effort per Measure by Province



13 PSYOPSA Psychological Operations Files

The following figures provide an overview of the PSYOPSA files that focus on ground-based psychological operations. The PSYOPSA file contains information on surface activity, radio and television transmission, and demographic statistics collected at the district level. These files appear to be recorded every couple of weeks and are therefore less granular than the SORTIEA files. The information can be paired with the HES files to add additional historical

context.

Figure S79: Ground PSYOPS Campaign Type Count

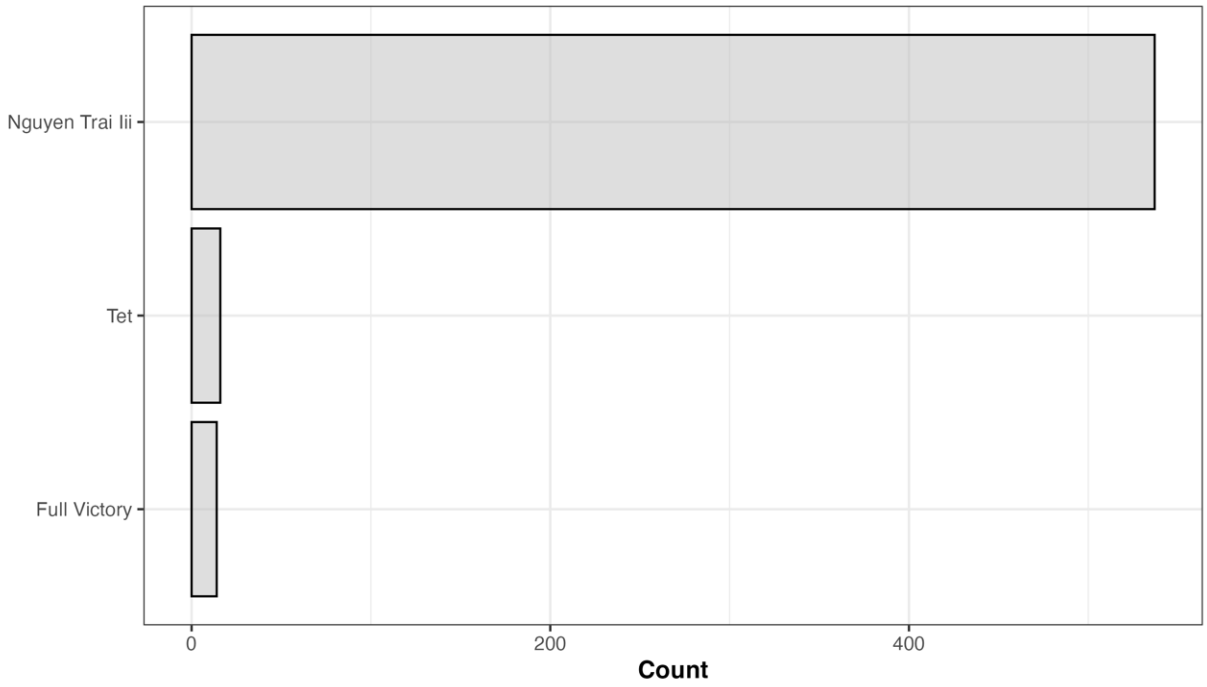


Figure S80: Ground PSYOPS Conducting Agency Count

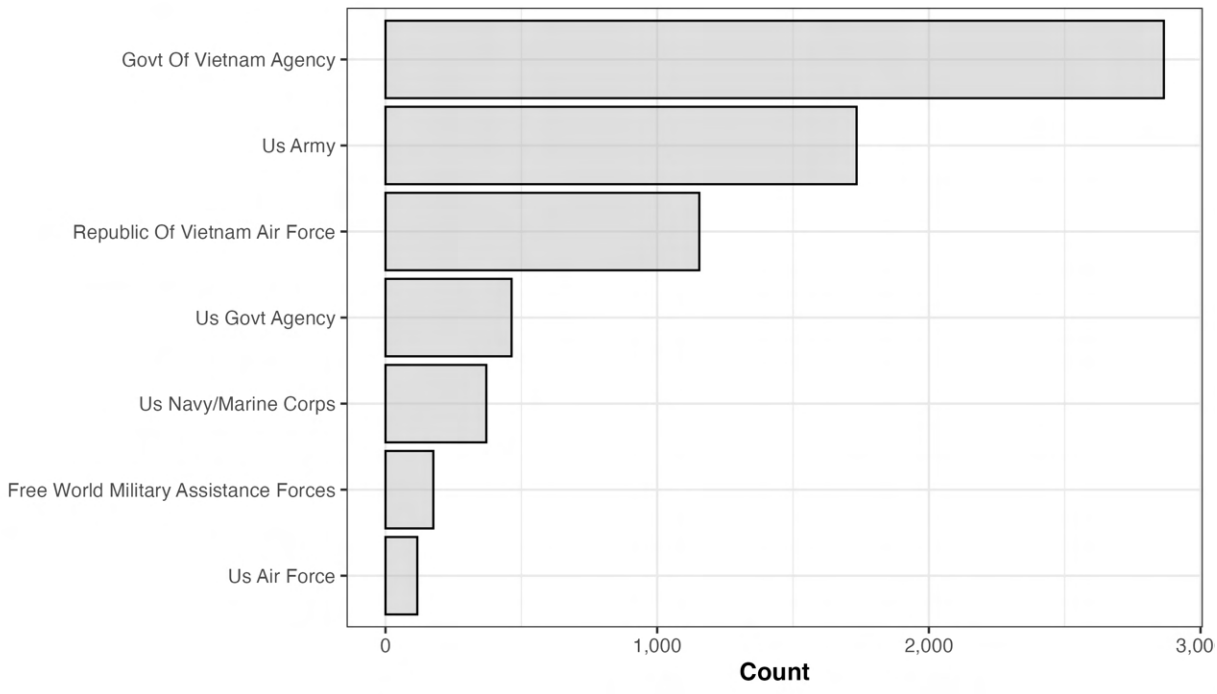


Figure S81: Ground PSYOPS Operation Theme Type Count

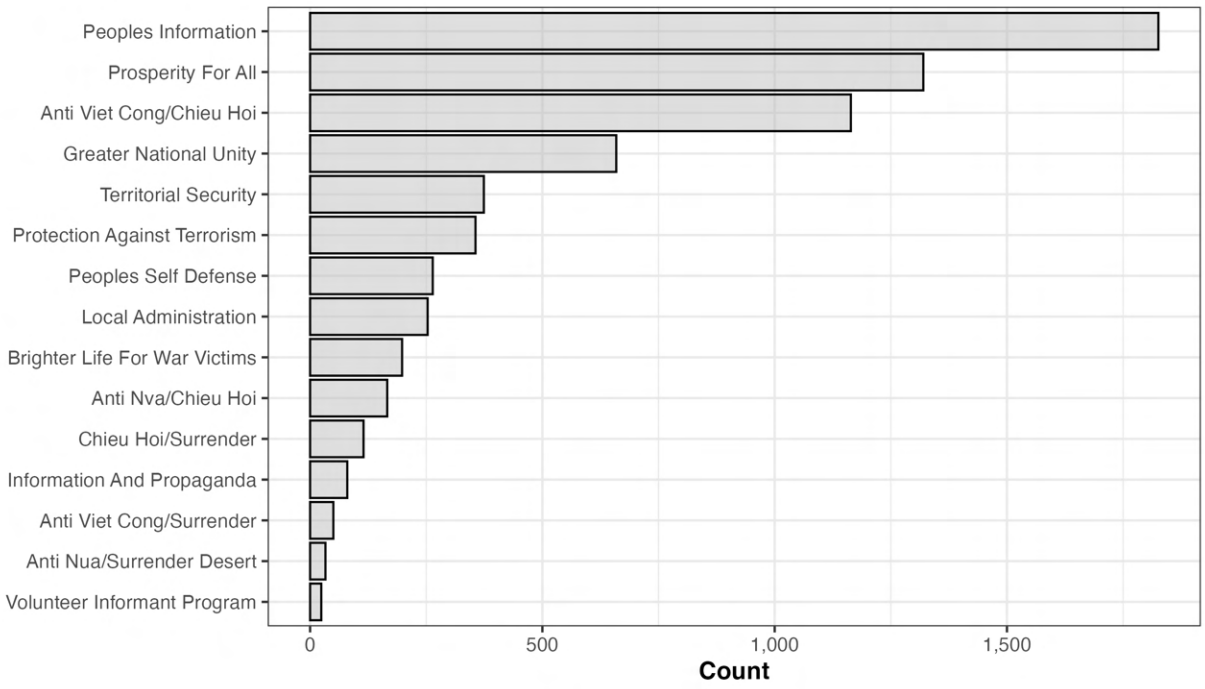


Figure S82: Ground PSYOPS Operation Sub-Theme Count

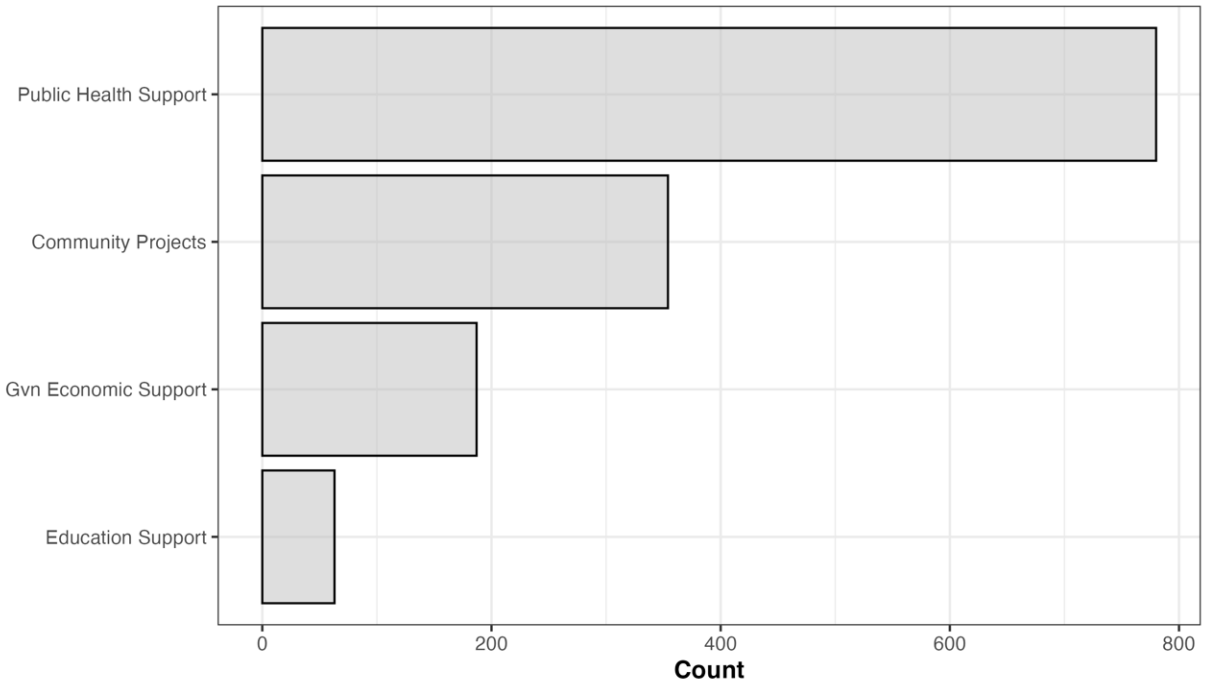


Figure S83: Ground PSYOPS Operation Type Count

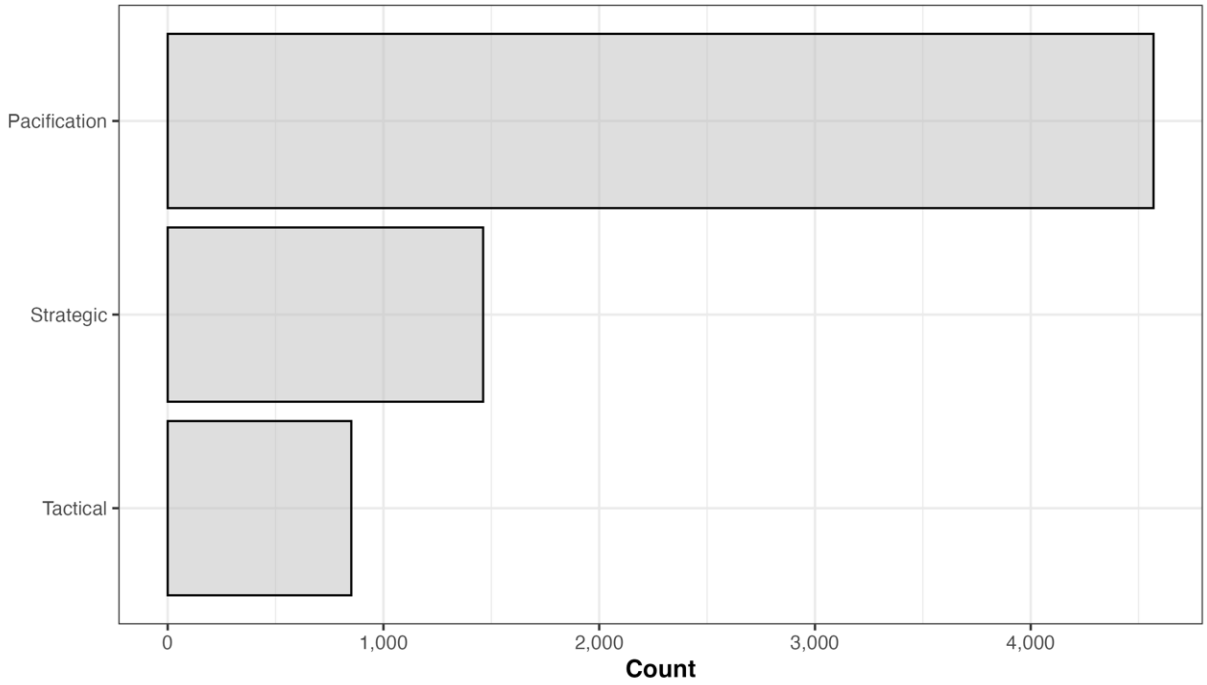


Figure S84: Ground PSYOPS Target Audience Count

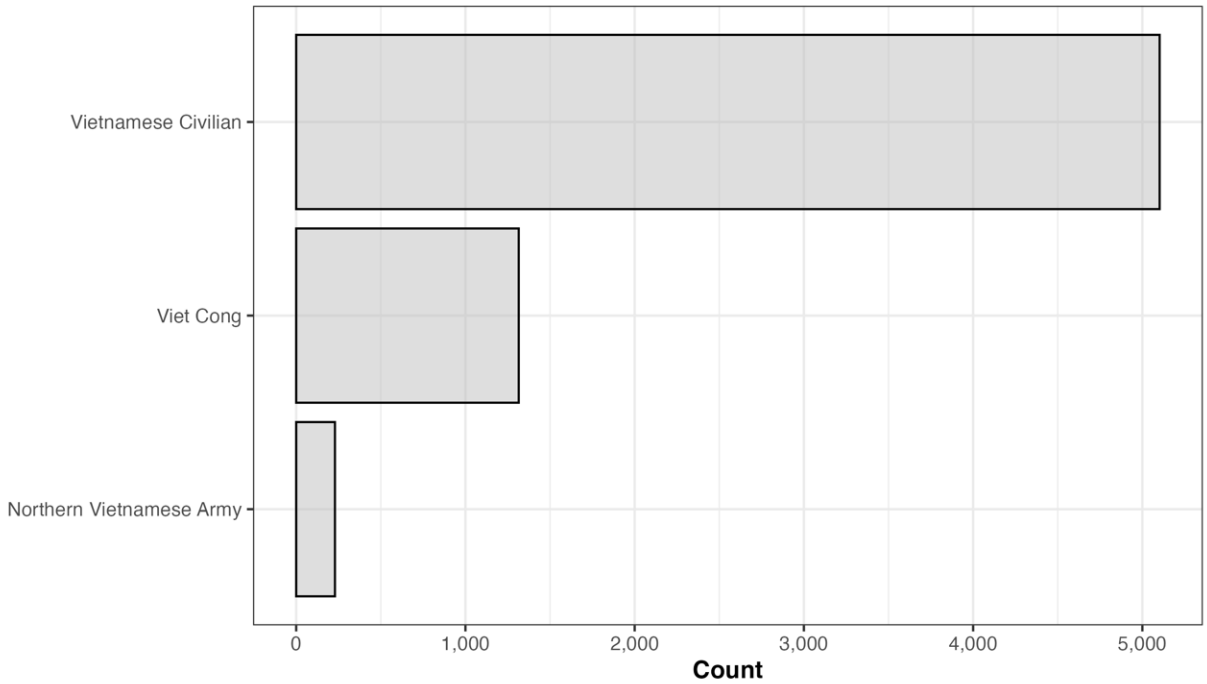


Figure S85: Total Quantity for Each Numeric Ground PSYOPS Measure

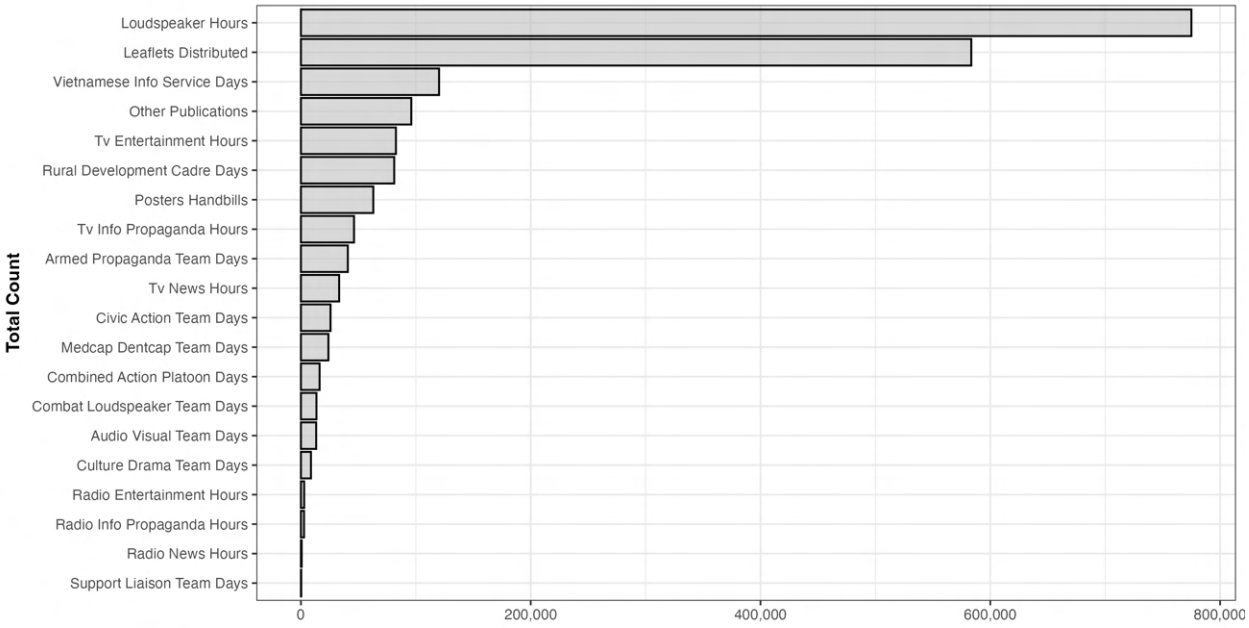
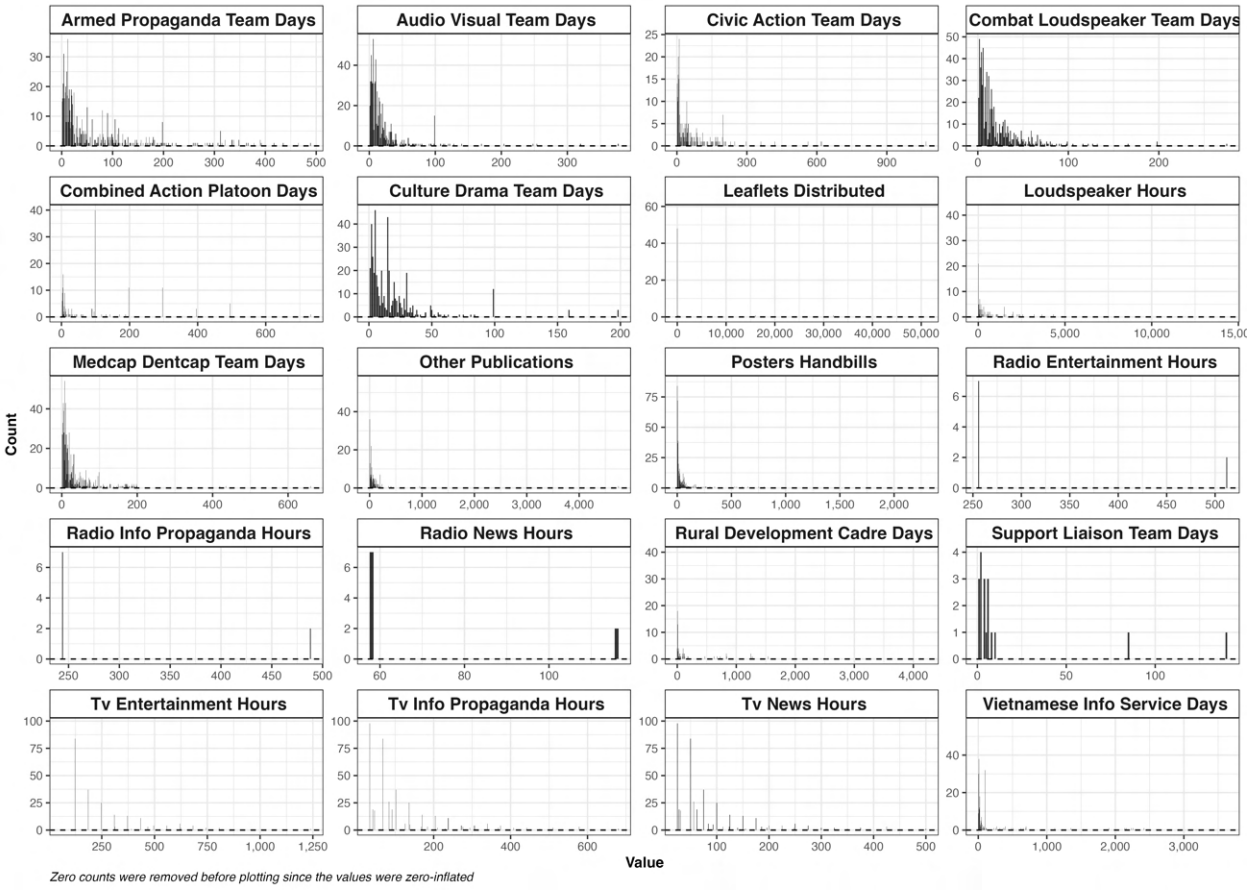


Figure S86: Distribution of Numeric Ground PSYOPS Columns



References

Carhart, Tom. 1984. *Battles and Campaigns in Vietnam: 1954–1984*. New York: Military Press.

Civil Operations and Rural Development Support. 1971. Hamlet Evaluation System (HES) Command Manual. Technical Manual DAR R70-79, CM-01B Military Assistance Command, Vietnam.

Dell, Melissa and Pablo Querubin. 2017. "Nation building through foreign intervention: Evidence from discontinuities in military strategies." *The Quarterly Journal of Economics* 133(2):701–764.

Fanell, James E. 1997. Operation Market Time: The United States Navy in Coastal Vietnam, 1965-1968 Thesis University of Hawai'i at Manoa. ProQuest Dissertations Publishing.

Hickey, Lawrence J. n.d. Operation Junction City. Technical report HQ PACAF, Directorate, Tactical Evaluation, CHECO Division Christiansburg, VA: . Prepared by S.E. Asia Team.

National Archives and Records Administration. 2005. Supplemental User Note: Situation Reports Army (SITRA) Files, May 1966 - March 1973. User Note NN3-218-76-032 National Archives and Records Administration College Park, Maryland: . Prepared by J. Powell, Archives Specialist.

National Military Command System Support Center. 1969. *The Operation Analysis System (OPSANAL) User's Manual*. Revision a ed. National Military Command System Support Center.

Naval History and Heritage Command. 2025. "Operation Market Time.". Accessed January 12, 2025.
URL: *https://www.history.navy.mil/content/history/museums/hrnm/explore/VietnamExhibitPage/Sur*

Robinette, Hillary. 1987. "Game warden: A report from the field." *Naval History* 1(1):19.

Schreadley, R. L. 1992. *From the Rivers to the Sea: The U.S. Navy in Vietnam*. Annapolis, Maryland: Naval Institute Press.

Sherwood, John D. 2015. *War in the Shallows: U.S. Navy Coastal and Riverine Warfare in Vietnam, 1965–1968*. Washington, D.C.: Naval History and Heritage Command.