MYANH TRAN
HONGJOON KIM

November 23th, 2021

The Trends of Global Terrorism from 1999 to 2019

Group 21: Kiwi Bomb

STAT 3355.001
Data Analysis for Statisticians and Actuaries
Department of Statistics
University of Texas at Dallas

TABLE OF CONTENTS

INTRODUCTION

a. Synopsis

The purpose of this research project is to implement the data analysis tool such as R programming and its library such as ggplot2 to extract insights about trends of global terrorism over the last 20 years, from 1999 to 2019. The driving motivation behind the investigation is to identify the common themes in the frequencies, locations, methods, targets, and lethality of terrorism during such time.

b. Data Set Information

The Global Terrorism Database (GTD) is managed by the National Consortium for the Study of Terrorism and Responses to Terrorism at the University of Maryland (START). Our group began with the most updated form of the data set from GTD, which includes 201,183 observations that span from January 1, 1970, to December 31, 2019, and up to 135 variables per incident. Some of the notable variables include information on the time and location (iyear, country, region, latitude, longitude).

We begin with the discussion of contextual information. The GTD Codebook, the official guide to GTD, defines *terrorism* as "the threatened or actual use of illegal force and violence by a non-state actor to attain a political, economic, religious, or social goal through fear, coercion, or intimidation" (START, p.10). The following requirements must be met for a given incident to be considered for inclusion in the data set: "The incident must be intentional... entail some level of violence or immediate threat of violence...[and the] perpetrators... must be sub-national actors" (START, p. 12) In addition, at least two of the following criteria must be satisfied for an incident to be included in the GTD:

1. “The act must be aimed at attaining a political, economic, religious, or social goal.”
2. “There must be evidence of an intention to coerce, intimidate, or convey some other message to a larger audience (or audiences) than the immediate victims.”
3. “The action must be outside the context of legitimate warfare activities.” (START, p.12)

For a given terrorism incident to be included in the data set, the conspirators need to be en route to execute the attack. Thus, “[p]lanning, reconnaissance, and acquiring supplies do not meet this threshold” (START, p. 13). A set of incidents occurring at the same time and geographical location is counted as a single *incident*. Hence, an event that is discontinuous in terms of time and/or location is divided into separate incidents, even if it is part of a series of terrorism events (START, p.13).

The Global Terrorism Database is well-suited for our investigation. First, its *large quantity* enabled us to sustain a good sample size, even though we have put certain constraints in our investigation (See Data Cleaning). Second, its *excellent quality* is assured because every source for a given incident goes through a rigorous vetting process so that every incident included is confirmed by at least one high-quality source (START, p. 10). Finally, its *global scope* enables us to easily investigate terrorism on a global scale with a single data set without compiling multiple data sets.

c. Interest

The inspiration for this investigation was sparked by the withdrawal of the United States troops from Afghanistan. This event marked the end of 20 years of the war between the United States and Afghanistan, which is a direct result of the War on Terrorism. As the tragedy of these wars continues, the question of whether they paid off has been raised to the surface. Therefore, we want to investigate the actual impact of the War on Terrorism on global terrorism. We measure the impacts through specific metrics: frequency, locations, methods, targets, and lethality.

d. Roadmap

Our roadmap was to use the proposed ten most common questions regarding such a topic as our guideline for the analysis. We realize that all of the questions boil down to these specific metrics: frequency, location, methods, targets, and lethality, we change the focus to analyzing these metrics. By carefully examining each of these variables through many aspects, we obtain the fullest picture the data provides. Frequently, we encounter variables that have more than 100 subcategories. In response, we filter the top 5 or 10 highest volumes and do the analysis on them. Firstly, we need an overview of where and when these attacks took place. Then, we want to see how the methods and targets of terrorism attacks change over time. Finally, we arrived at the impact of terrorism by graphing the lethality.

2. DATA CLEANING

a. The Process and The Rationale

Despite a few variables needing some combing through, the data set did not need too many adjustments. We began by investigating variables that were not pertinent to the inter-related questions initially posed. The fact that there are up to 135 variables per incident rendered the original data set superfluous and inefficient for the scope of our investigation. Thus, we deleted 76 impertinent variables, resulting in 59 variables useful for our investigation.

In addition to identifying relevant variables, we made more qualifications to the scope of our data set as a whole. First, we deleted 251 observations that were missing entries in the "iday" variable, just in case we wanted to investigate specific days of a month. Next, we considered the nature of the older observations (from 1970 to 1997), which rely on retrospective data entry methodology and suffer from inconsistent data entry due to the lack of access to reliable sources. Thus, we chose to focus on post-1999 observations by deleting all pre-1999 observations. Finally, we decided to focus on the observations that are true incidents of terrorism, as defined by GTD. This is necessary since GTD recognizes the ambiguous cases, where incidents may have some definitional overlap with other forms of crime or have insufficient evidence from the sources to deem them as an act of terrorism. Still, GTD does include ambiguous cases in which they are likely, but not beyond all doubts, true cases of terrorism (START, p. 12). Getting only the true terrorism incidents was accomplished by filtering out observations whose entry in "doubtterr" (the variable indicating the presence of doubt as to an incident's identity as an act of terrorism) is "1" (corresponding to "Yes"); this is so that for all of our observations, "[there] is essentially no doubt as to whether the [incidents are acts] of terrorism" (START, p. 17).

All of the data cleaning steps thus far narrowed the data set from 201,183 to 110,096 observations and from 135 to 59 variables. The next major step was fine-tuning individual variables. From the beginning, we made sure that all blanks are NA's, as necessary. Using "as.logical ()", we turned the entries in the variables success and suicide into TRUE or FALSE values. We converted targtype1 to a factor vector, using "as.factor ()." Repetitive data entries in three variables categorizing the choices of tactics (attacktype1, attacktype2, attacktype3) were cleaned out by identifying all indices where such repetitions occurred through "which ()" function and setting the repeated values of those indices at the variables to NA. Next, we dealt with some inconsistent data entries in variables relating to the number of deaths (nkill, nkillus, nkillter) and the number of wounded victims (nwound, nwoundus, nwoundter). There were several NA and "0" entries in nkill and nwound (relating to total number of confirmed deaths and wounded victims, respectively), for sometimes reliable estimates of these values could not be identified by the START researchers. In latter cases in which nkillus and nwoundus (number of deaths and wounded victims from the US) were available, we set the values of nkill and nwound to those values, so that our analysis can be based on all available confirmed numbers.

Finally, we constructed other data classes to aid our investigation. We made a vector apart from the main data set that records occurrences of US states in the data set (refer to index_states in Section 4.c.i). We also made a series of data frames (refer to Section 4.c.i). These data frames were created for simplification of multiple related variables, organization of cumulative sums, ease of use for certain visual analysis, or numerical analysis. These were created through several vectors, for-loops, "rbind ()", "cbind ()", "as.data.frame ()", "select ()" from plyr / dplyr packages, and “melt ()” from reshape2 package. More details on the nature and development of these data frames are delineated in Section 3: Data Analysis.

In the next section, we present motivating questions and curious results through data visualization and numerical analysis. We explore the trends in the geography and frequency of global terrorism, as well as the chronological and regional patterns in types of victims. We also explore the general methods (also called tactics) of terrorism in terms of region and time. Finally, consider some explanations for why certain tactics are more prevalent than others.

3. DATA ANALYSIS

---

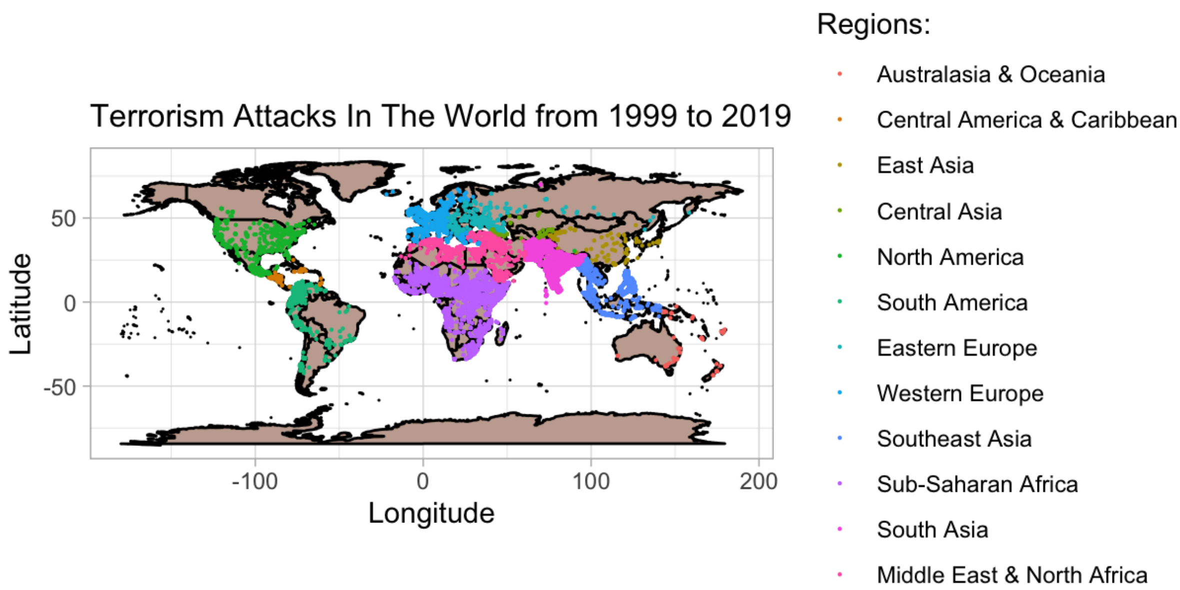
a. Data Description

There are many stereotypes about terrorism, such as the claim that it only exists in the Middle East, that it is only conducted by some specific ethnicities, that all terrorists care about is harming and murdering others for the sake of doing so. We first want to conduct our own overview of global terrorism trends and secondly, counter-check these myths by letting the data represent itself. Interpretation of the analysis is carefully fact-checked and based on academic resources in order to protect the analysis against unconscious bias.

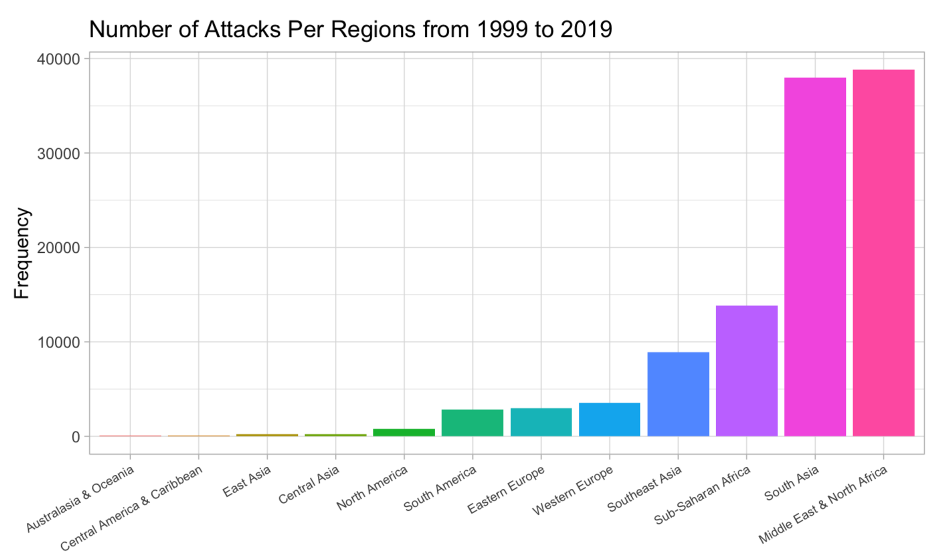
b. Topics and Questions

i. Location

1. What Regions Have The Most Frequency of Attack?

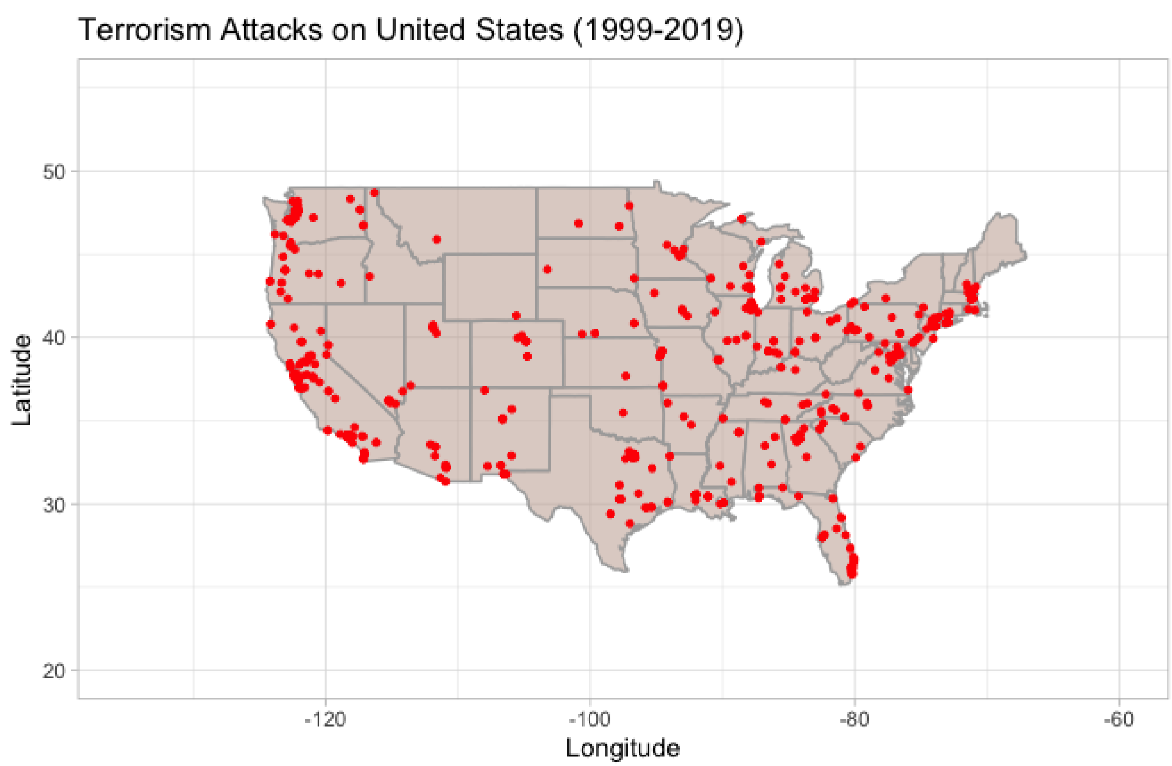


First, we want to get a big picture of how terrorist attacks spread throughout all the regions in the world. Every point in the graph is one observation. Some points are in the middle of the ocean because there are attacks on boats and ships. However, due to the enormous amount of observations, the points are too dense. They cover entire continents and we cannot distinguish the difference in frequencies between regions anymore.



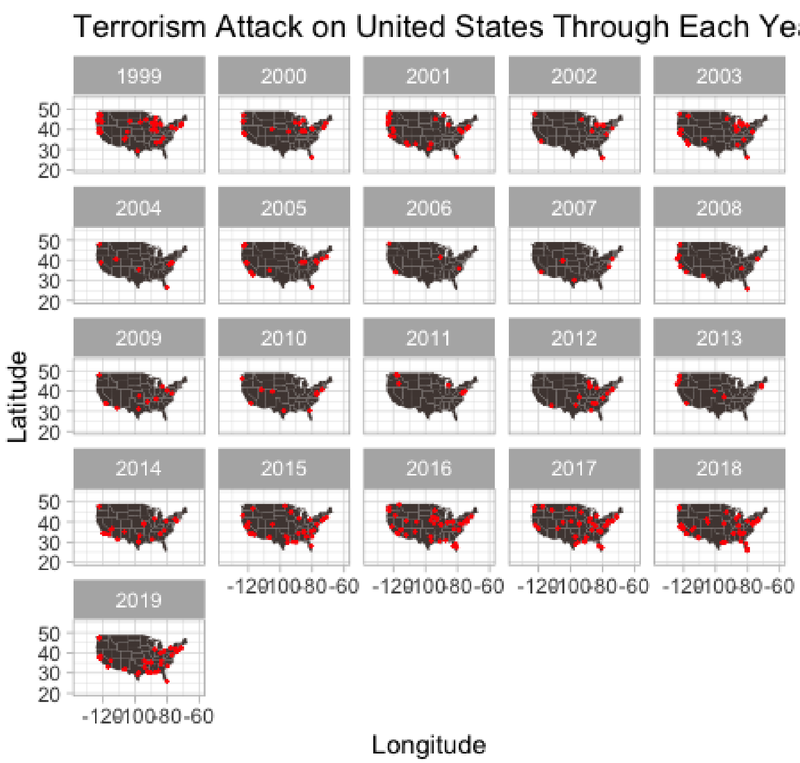
Then I return to the good old histogram. It demonstrates perfectly the imbalance between the top four most attacked regions that are not reflected in the first graph. The Middle East's high frequency of attacks can be explained by the wars resulting from War on Terrorism such as War on Iraq, War on Afghanistan. The Africa region as a whole, even though it is not in the spotlight of the media, has undergone many civil wars and crises such as Somali Civil Wars, Congo Wars,... Despite being civil war or conflict, they have inter-state support from other countries and organizations, which check up the requirement of GTD to be a global terrorism event. On the other hand, the unexpectedly high frequency of attacks in South and Southeast Asia demonstrates the profound impacts of terrorist organizations. For example, Al Qaeda forces fled from Afghanistan with their Taliban supporters to link with indigenous Pakistani terrorist groups. Southeast Asia, by itself, has been home to the indigenous Islamic militant groups for decades, many have transformed into legal opposition to the local regime. The Al Qaeda terrorist network showed evidence of making significant inroads into this region as early as the beginning of the 1990s.

2. What About The Distribution in America?



Now, we take a closure look at the impact of global terrorism by zooming in on America. California leads the top states having the most terrorism attack by 95, leaving New York with 66 and Texas along with Washington 44 cases during the span of 20 years. However, most of the top cities are not inside the state of California. New York City led the chart with 51 attacks, followed by Washington City and Los Angeles. Being the financial hub and the cultural symbol, the representation of America, New York City has been chosen to be an established target, a statement of many terrorist organizations. Perhaps the idea that if you can make it in New York City, you can make it anywhere is also applied to these organizations. On the other hand, Washington City and Seattle being in the top 5 most attacked cities demonstrate the political aspects of many attacks in the United States. Capital and the biggest city in a country are landmarks for any global terrorism attacks.

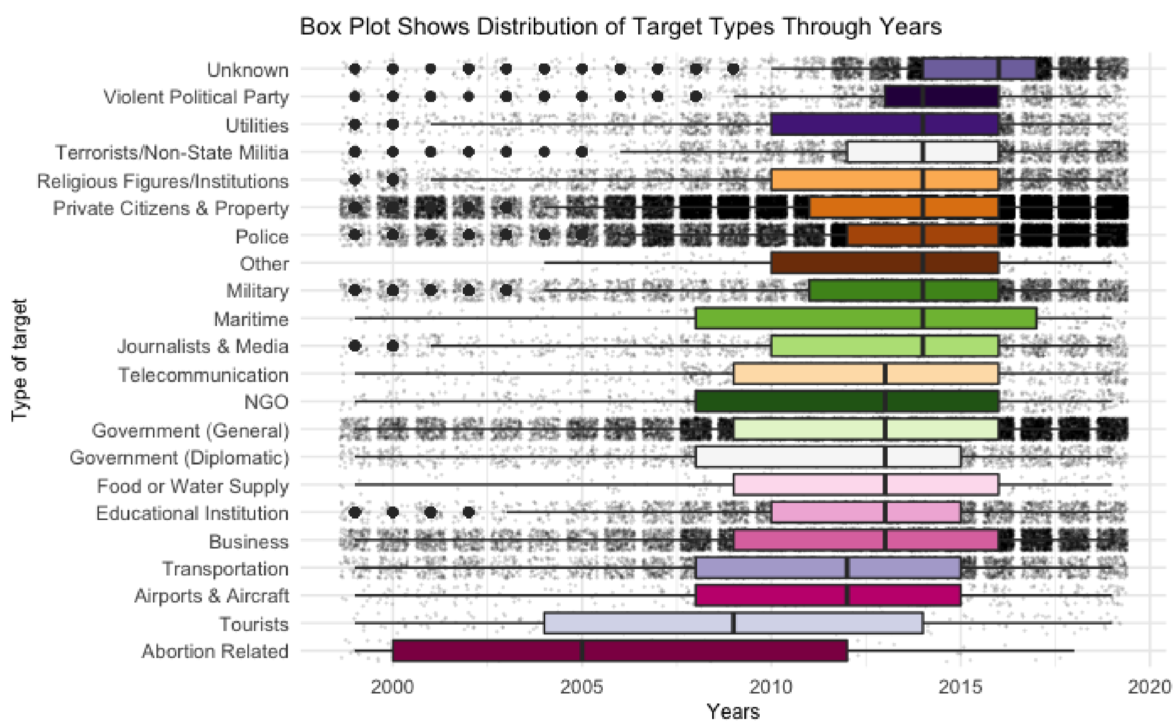
3. What Is The Trend of Terrorist Attacks in America Throughout 1999-2019?



This graph debunks the stereotype that the only terrorist attack in America is the September 11th event. Starting from 1999, attacks have been surging across America's East and West coast and scattered through the Middle West region. There are three important time periods to be examined: before 9/11, the start of the war on Afghanistan and Iraq, and the death of Osama bin Laden. After the incident of 9/11/2001, America immediately declared War on Terror, which led to the war with Afghanistan starting on 7/10/2001 that ended in 2021 and war with Iraq from 3/23/2003 to 12/5/2011. During the period of time from 2004 and 2011, we observed an obvious drop in the number of attacks. The frequency is reduced and mainly focused on the coasts. National security such as restrictions of airline access and immigrant programs have been improvised much more than before 2001. Keep in mind, in this period, the Iraq War begins, Saddam Hussein and Bin Laden were killed. On May 2nd, 2011, the death of Osama bin Laden, the organizer of 9/11 which started the war on Terror, marked a partial end to the War on Terror itself. Despite our expectation that killing the mastermind of the most famous terrorist attack will reduce the terrorism attacks, the number of these attacks has been increased more than ever. 2012 sparked the return rising of terrorism attacks, which have been continuously rising since then.

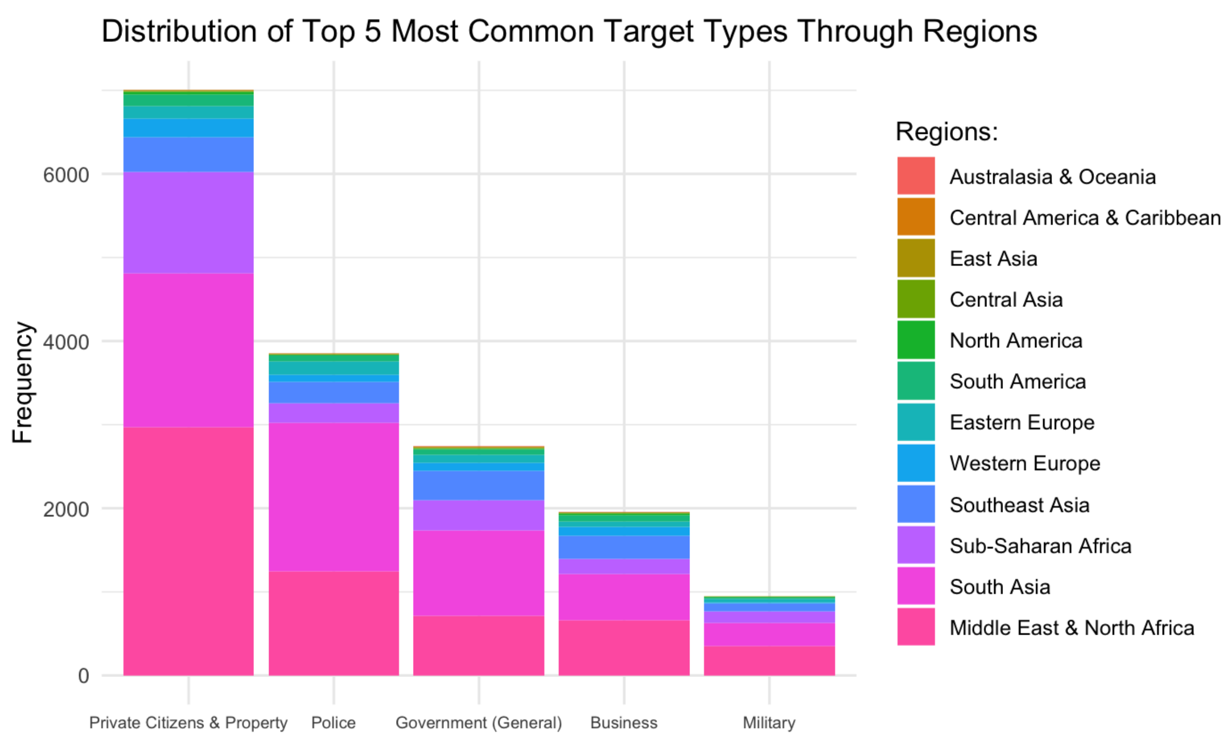
ii. Target

1. What is Distribution of Target Type Through Years?



GTD is a well-documented dataset that categorizes up to 22 types of terrorist targets. Box plot demonstrates the upper and lowers bound of each type of target in terms of years. Jitter plot shows the distribution of observations. Even though some subcategories have the same means, their numbers of observations can be extremely different. I rank all types of targets by the most common years they occur and almost all of them follow the same pattern. 2014 and 2013 seem to be the most significant, violent years as most of these attacks happen in these year. We will later explain this pattern as we observe them through the report.

2. Is There A Preferred Target Type For Each Region?



We want to know if a type of target is more likely to be attacked because of its location. We want to know if the location of the region plays into account what type of target is more attacked. After taking an individual look at each category of target type, now we examine the distribution of the top 5 most attacked targets among all the regions in the world. Terrorist groups in the Middle East and North Africa are more likely to attack the “Private Citizen and Property” than any other region by dominating 40%. On the other hand, South

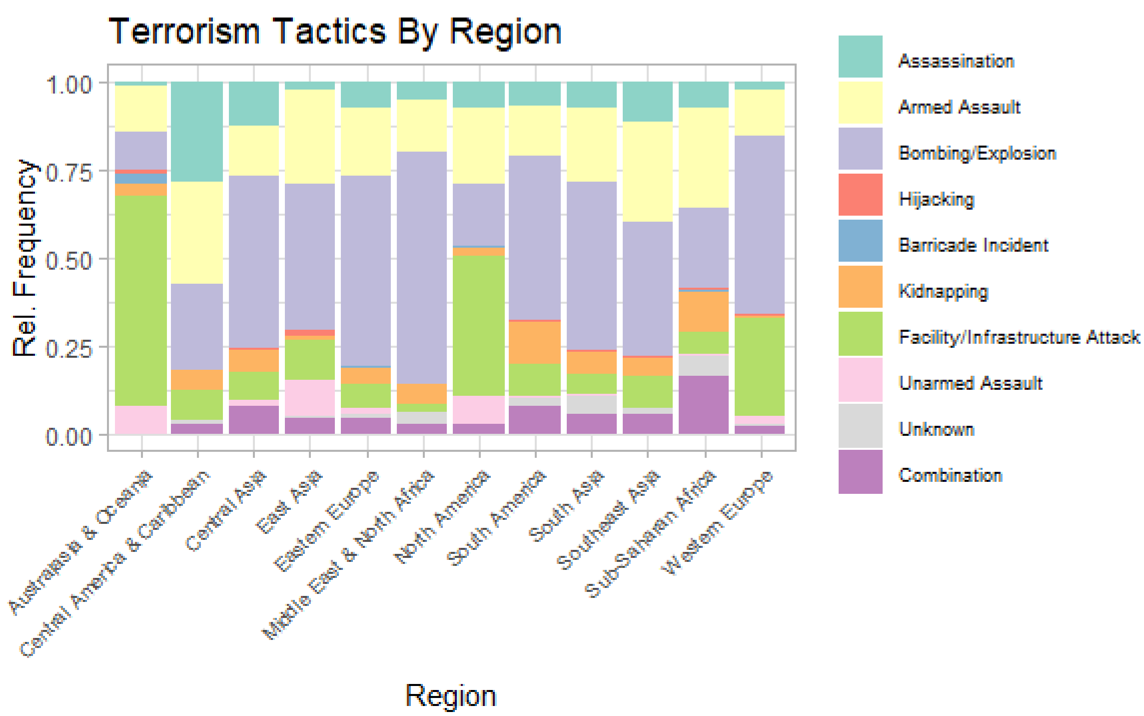
Asia's terrorists prefer to attack the "Police" and "Government" more than any other target type.

iii. Tactics

Now that we have considered the trends in geographical and chronological patterns of frequency and preferred target types, we now turn our attention to general methods (or tactics) used by perpetrators worldwide. For the sake of national security, we believe there is importance in recognizing various tactics used by the perpetrators, for if we want to prevent and combat terrorism we must study the tactics.

Before diving into analysis, some contextual information needs to be imparted. First, the variables relating to tactics in the original data set is such that multiple tactics can be inputted; however, the order in which the tactics were inputted was neither based on the chronological sequence of tactics nor the primacy of certain tactic in a given terrorism incident. Rather, GTD inputs multiple tactics according to an arbitrary hierarchy of tactics set by GTD (see START, p. 24). Second, a derivative data frame was necessary to ease the process of analyzing the tactics; thus, with a combination of for-loops, reshape2, and dplyr/plyr packages, we made a smaller data frame, *gt_1*, with a new variable: *newattack*. This data frame, *gt_1*, and its other derivatives are the basis for the data analysis from this point forward. The difference between the pre-existing data set and *gt_1* is that this accounts for a "combination" of tactics. "Combination" (or "Combo" for short) signifies that a given incident of terrorism involved a chronological sequence of distinct tactics used. This does not include cases where a specific, primary tactic was carried out through the means of or support from other tactics. For example, an assassination carried out through explosives is inputted strictly as "Assassination," even though explosives were involved (see START, p. 24). Now, we are ready to explore our findings regarding tactics of terrorism.

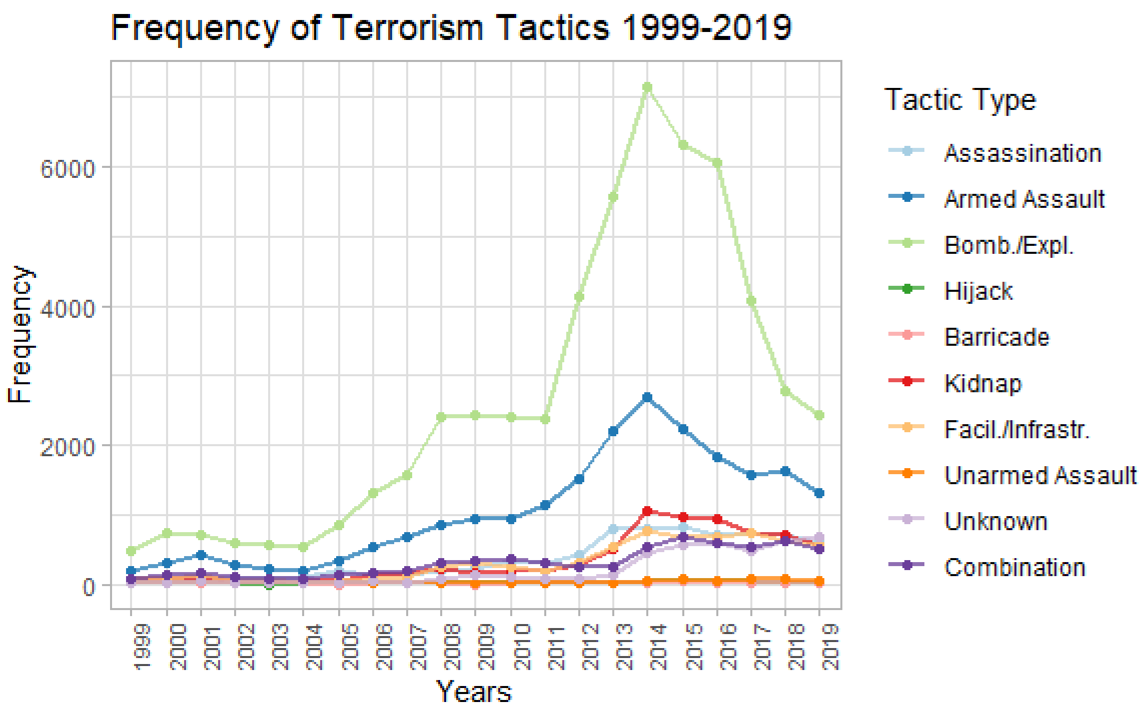
1. What are the Most Common Tactics for Each Region?



The relative frequency bar graph separates the globe into 12 different regions. Of these regions, we see that in at least seven of them, bombing and explosion occupy the largest proportion of all incidents recorded in each respective region. We also see that in almost all 12 regions, both armed assault and bombing/explosion is in the top three most common types of terrorism tactics in each respective region. Notice that not only does the Middle East and North Africa have the highest proportion of bombing/explosion incidents, in relation to the whole scope of its own incidents, but also have a greater number of them than all other 11 regions, especially when we compare this with the regional frequency bar graph in Section 3.b.i.1.

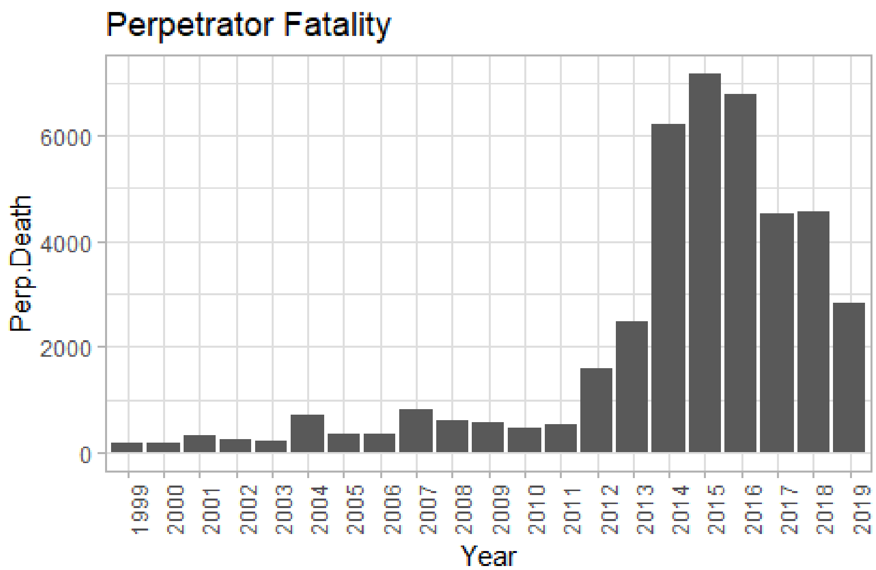
Finally, we see an anomaly in Australasia / Oceania and North America, where facility/infrastructure attacks (attacks against non-human targets, like buildings or monuments) dominate each region's proportion of tactics used by its perpetrators. A possible explanation is the small sample size. According to the same bar graph in Section 3.b.i.1, both regions have the lowest occurrences of true terrorism; therefore, the small sample size might result in unfair comparison with other regions. However, that alone does not appear to address it satisfactorily, since Australasia / Oceania has 92 total incidents and North America has 766 total incidents since 1999, which seems not too little. At least with respect to the United States, we can speculate that the relatively stable government, economic security, and military strength play a factor, but this cannot be confirmed by the data set.

2. Which Tactics are the Most Popular of All Time?



Next, we consider the global frequency of each tactic over time. The multiple line graph is based on data frame *gt_3, which is* derived from *gt_1*.
The purpose of *gt_3* was to gain a cumulative sum of the frequency of each tactic on a yearly basis. We see that, in the similar spirit as the previous graph, bombing/explosion and armed assault dominates in frequency across two decades. Why is it that both tactics not only dominate the regional proportions but also the yearly frequency globally? We will explore possible explanations very soon.

However, we note two other interesting observations. First, notice that most tactics reach their peak around 2014 and drop, with the two most frequent types dropping most dramatically. Hence, terrorism, in general, dropped in frequency. A possible explanation of this is, for whatever is happening at the political level in each country, the number of perpetrator fatalities has increased dramatically in 2014. Note the bar chart detailing perpetrator death count over the years below:



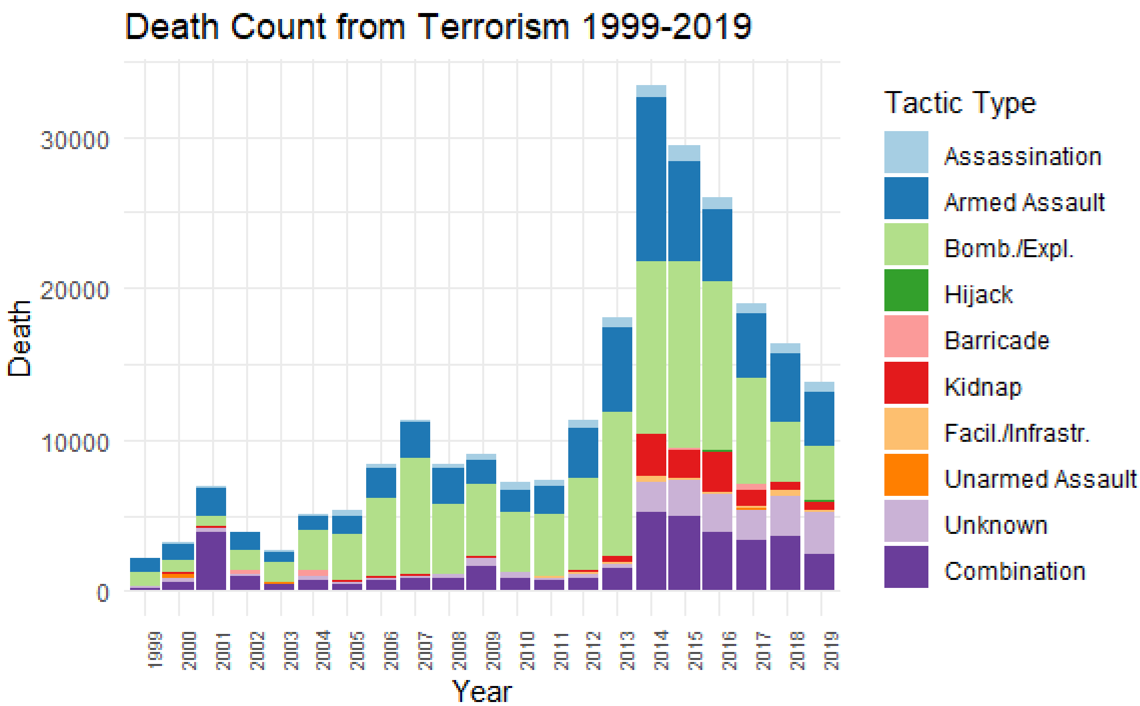
Notice that the perpetrator fatality more than doubled from 2013 to 2014, and from 2014 to 2015 the total fatality increased by a thousand. Consider the sudden drop in the number of terrorism in 2015 in the previous multiple line graph; perhaps, the countries under attack have adapted to the major terrorism tactics (particularly bombing/explosion and armed assault), but we still have to consider the proportions of that death that are suicide attacks (where a perpetrator does not intend to emerge from the incident alive) and accidental deaths of the perpetrators. The current data set is insufficient in determining the proportion because of the lack of organized data for accidental deaths.

Finally, observe the slow and steady rise of the unknown tactic. According to GTD Codebook, these are tactics that " cannot be determined from the available information" (START, p. 26). Although the data set does not shed much light on the nature of these tactics, we do think it might be significant for national security, since the frequency of these tactics is on the rise. Of course, we can speculate that these do not fit the standard GTD categorization, but it could signal a rise in a new tactic. We will explore the potential significance of this tactic further.

3. Why are Certain Tactics More Prevalent than Others?

Having considered the prevalence of certain tactics over others, particularly bombing / explosion and armed assault chronologically and regionally, we now ponder the reason behind it? Can such a reason be determined from the current dataset? We now begin by examining two areas: the lethality and the success rate of each tactic.

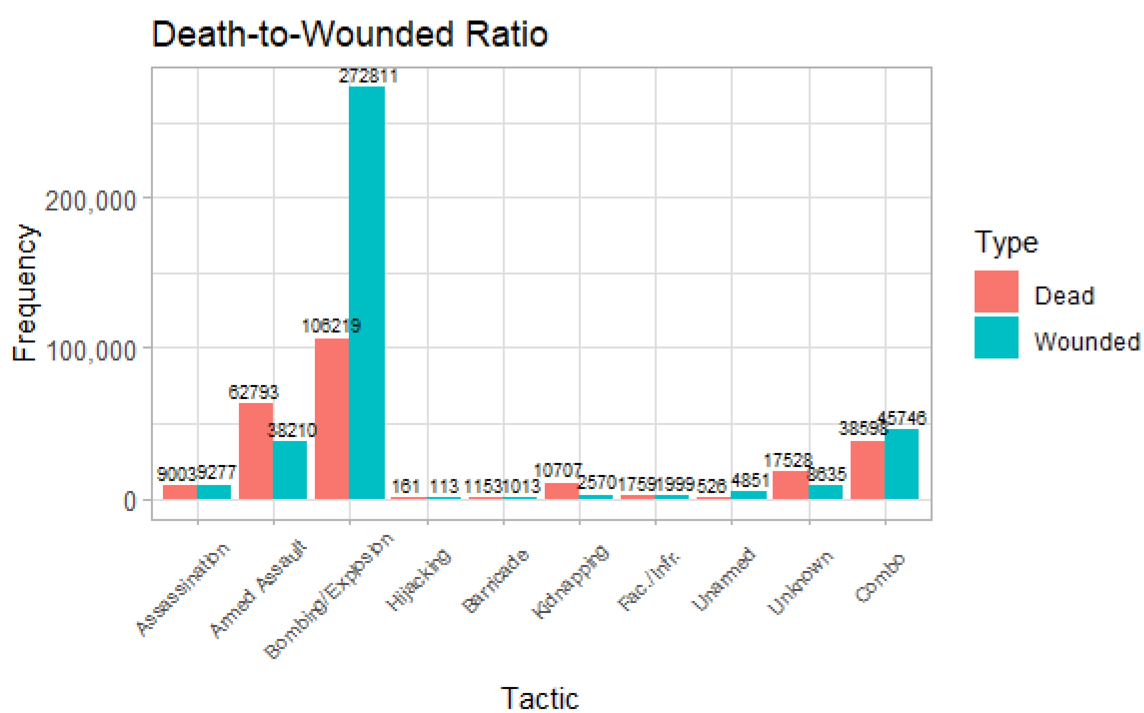
a. Let's Consider: Lethality of Tactics



One of the possible reasons for the prevalence of certain tactics is how effective it is at killing its victims. The stacked bar chart above may perhaps explain why. Certainly, the incidents with the top two most common tactics (bombing / explosion and armed assault) are associated with more than half of all death counts for most years, but what about the combination of tactics and the unknown tactics? Despite having the lowest frequencies, they appear to occupy a relatively large portion of death count, especially in 2019, where both occupy near 50% of all death counts. This raises a suspicion. Perhaps the reason why the top two most common tactics are associated with the most death counts is precisely because they *are* the most frequent! More occurrences, more deaths. We now consider the deaths-per-incident ratio.

	Deaths-Per-Event
Assassination	1.2032879
Armed Assault	2.8607289
Bomb./Expl.	1.9162728
Hijack	0.5532646
Barricade	6.3351648
Kidnap	1.4874965
Facil./Infrastr.	0.2717442
Unarmed Assault	0.8945578
Unknown	4.1822954
Combination	6.1159880

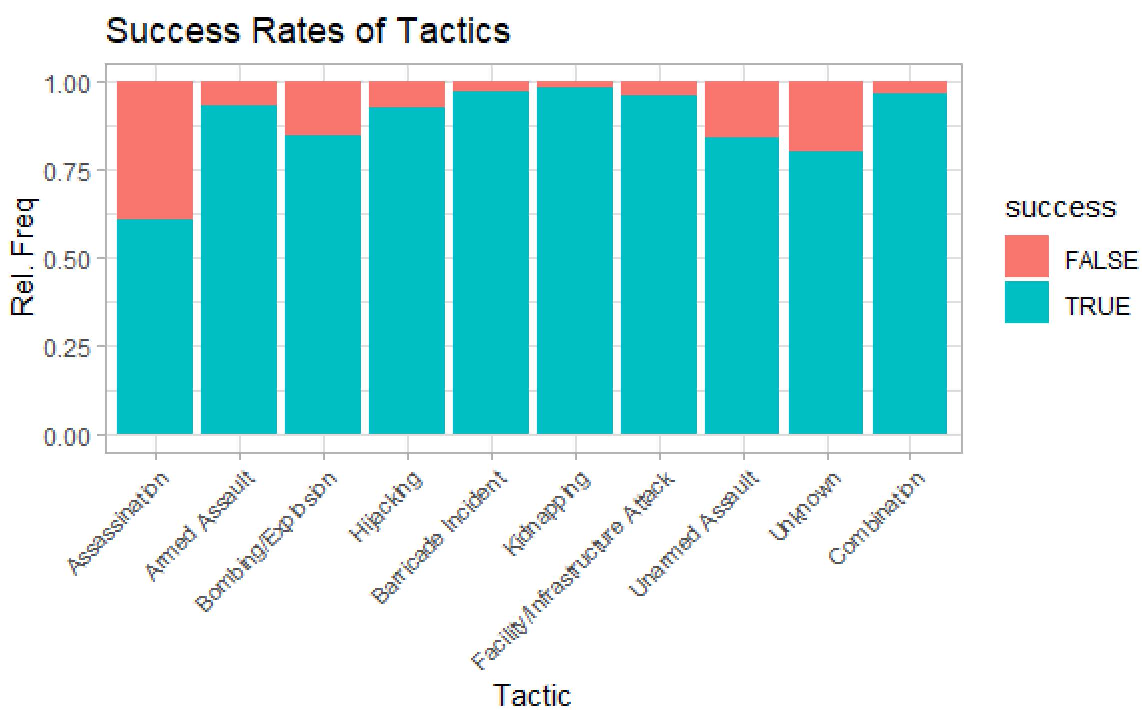
The table is generated from a small data frame, *gt_8*, which is based on *gt_1*. It was constructed by running a for-loop to (1) find the sum death count corresponding to a given tactic and to (2) divide the latter value by the number of total occurrences of incidents for that tactic. What is surprising is that deaths resulting from the incidents with unknown tactics is 25% higher than that from armed assault and more than double that from bombing / explosion. The sum of the deaths-per-event of two most common tactic types combined cannot surpass that of the combination of tactics. Thus, we can conclude that greater lethality does not imply greater popularity. Finally, we want to consider how survivable each of these are.



In the side-by-side bar chart above, generated by *gt_7*, which is a derivative of *gt_6*, a form of *gt_1* in which the NA values in number killed or wounded are swapped out for 0. We make some startling discoveries. First, the number of wounded victims from bombing and explosion triples the fatality from the same tactic. In contrast, incidents from the second most popular tactic, armed assault, have ~30% more fatality than the number of wounded victims from it. Also, notice that several of the less popular tactics have nearly a 50-50 split between death count and wounded count. This confirms our suspicion that lethality is not *the* reason certain tactics are popular. Before we close, we explore another aspect: success rate.

b. Let's Consider: Success Rate of Tactics

Initially, one might guess that success of an incident depends on the achievement of the political, economic, religious, or ideological goal of the perpetrators behind the incident. Actually, GTD defines success in terms of a given tactic being executed successfully. Each type of tactic has different criteria, as to what makes it a successful one or not. That being said, we now consider the success rate of each tactic:



The relative bar chart above shows that likelihood of execution of a certain tactic does not relate to its frequency of practice. Surely, armed assault and bombing / explosion have near 90% success rate. However, consider the two hostage incidents: barricade incident (where perpetrators hold a victim hostage with no intent to do so in the long-term) and kidnapping (where perpetrators take hostage with intent to move the victim from one place to another). Despite having a low frequency, the success rate is very high.

In sum, we have shown that the popularity of certain tactics have little to do with the lethality or likelihood of success. There are more factors to consider that are beyond the scope of this research. Despite the lack of positive reason, however, we do have a new insight: Terrorism does not happen simply because the perpetrators takes pleasure in killing people. It appears that our data suggests that simply wounding them is sufficient. Why? Perhaps, the answer is where it began. We recall that *terrorism* is defined in a specific way: "the threatened or actual use of illegal force and violence by a non-state actor to attain a political, economic, religious, or social goal through fear, coercion, or intimidation" (START, p.10). Thus, the data is consistent with the proposition that, for the perpetrators, making a memorable statement to their audience about their power and potential.

4. CONCLUSION

---

a. Further questions

i. Domestic compare to Global

For further research, I want to investigate the connection between domestic terrorist groups with global groups. I want to know whether they influence each other in terms of financing or popularizing. I want to research the rise of extreme right-wing terrorist groups in the recent political climate. For example, the rise of alt-right groups and the number of attacks related to hate crimes.

ii. Political impact

I want to have more evidence or a closer look into the impact of politics on the terrorist attack. For example, I want to find the relationship between

American politics to the trend of attacks in the Middle East. Like that, I want to examine how influential is American to the movement of terrorists. If not, then who and what, are the factors behind these changes.

iii. Rise of the Unknown

As we have seen in the data analysis, there is not only a significant rise in frequency of the unknown but also the fatality from it. Could this be the rise of some tactics we do not know about? We would like to identify any trends in these attacks by means of other data sets.

iv. The Heart of the Problem

Despite the wealth of details on the various aspects of terrorism, the data set did not allow us to explore the common motives behind global terrorism. What is the most common motivation? Is it political, religious, or ideological? We would like to identify any trends in these attacks by means of other data sets.

b. Reference

"The Softest Of Targets: A Study On Terrorist Target Selection". Taylor & Francis, 2021, https://www.tandfonline.com/doi/full/10.1080/19361610902929990?scroll=top&needAccess=true.
"Terrorism In South Asia". Everycrsreport.Com, 2021, https://www.everycrsreport.com/reports/RL32259.html.

"Global Terrorism Database: Codebook: Methodology, Inclusion Criteria, and Variables". START, 2021, https://start.umd.edu/gtd/downloads/Codebook.pdf.

c. Code

i. Data cleaning

```
gt <- read.csv("~/R/R wd/globalterrorismdb_0221dist.csv")
# Read Data Set

# iday
gt<-gt[which(gt$iday != 0),]
# Deleted "iday" entries with 0's.

gt[c(1, 3,      4,      5,      6,      7,      13,     16,     17,     18,     19,     20,     21,
22,     24,     25,     26,     47,     48,     55,     56,     59,     60,     61,     62,     63,
64,     65,     66,     67,     68,     69,     70,     71,     72,     73,     74,     75,     76,
77,     78,     79,     80,     81,     98,     105, 106,       107,    108,    109,    110,    111,
```

```
112,    113,    114,    115,    116,    117,    118,    119,    120,    121,    122,    123,    124,
125,    126,    127,    128,    129,    130,    131,    132,    133,    134,    135)] <- NULL
# Deleted Unwanted Variables

gt <- gt[which(gt$iyear >= 1999),]
#Now, we have data from 1999 to 2019.

gt <- gt[which(gt$doubtterr == 0),]
# Filtered out doubtful cases.
```

1. Hongjoon's Portion

```
#success
gt$success <- as.logical(gt$success)
#Logical values assigned for "1" and "0".

#suicide
gt$suicide <- as.logical(gt$suicide)
#^^ ^^Logical values assigned for "1" and "0".

#attacktype1, attacktype2, attacktype3
gt$attacktype1[gt$attacktype1 == ""] <- NA
gt$attacktype2[gt$attacktype2 == ""] <- NA
gt$attacktype3[gt$attacktype3 == ""] <- NA
# Replace blanks with NA's.

which(gt$attacktype1 == 1 & gt$attacktype2 == 1)
# No repeating "1" for attacktype1 & attacktype2.
which(gt$attacktype1 == 1 & gt$attacktype3 == 1)
# No repeating "1" for attacktype1 & attacktype3.
which(gt$attacktype1 == 1 & gt$attacktype2 == 1 &
      gt$attacktype3 == 1)
# No repeating "1" for attacktype1 & attacktype2 & attacktype3

which(gt$attacktype1 == 2 & gt$attacktype2 == 2)
gt$attacktype2[which(gt$attacktype1 == 2 &
                     gt$attacktype2 == 2)] <- NA
which(gt$attacktype1 == 2 & gt$attacktype2 == 2)
# No repeating "2" for attacktype1 & attacktype2.

which(gt$attacktype1 == 2 & gt$attacktype3 == 2)
gt$attacktype3[which(gt$attacktype1 == 2 &
                     gt$attacktype3 == 2)] <- NA
which(gt$attacktype1 == 2 & gt$attacktype3 == 2)
```

```
# No repeating "2" for attacktype1 & attacktype3.

which(gt$attacktype1 == 2 & gt$attacktype2 == 2 &
      gt$attacktype3 == 2)
# No repeating "2" for attacktype1 & attacktype2 & attacktype3

which(gt$attacktype1 == 3 & gt$attacktype2 == 3)
gt$attacktype2[which(gt$attacktype1 == 3 &
                  gt$attacktype2 == 3)] <- NA
which(gt$attacktype1 == 3 & gt$attacktype2 == 3)
# No repeating "3" for attacktype1 & attacktype2.

which(gt$attacktype1 == 3 & gt$attacktype3 == 3)
gt$attacktype3[which(gt$attacktype1 == 3 &
                  gt$attacktype3 == 3)] <- NA
which(gt$attacktype1 == 3 & gt$attacktype3 == 3)
# No repeating "3" for attacktype1 & attacktype3.

which(gt$attacktype1 == 3 & gt$attacktype2 == 3 &
      gt$attacktype3 == 3)
# No repeating "3" for attacktype1 & attacktype2 & attacktype3

which(gt$attacktype1 == 4 & gt$attacktype2 == 4)
# No repeating "4" for attacktype1 & attacktype2.

which(gt$attacktype1 == 4 & gt$attacktype3 == 4)
# No repeating "4" for attacktype1 & attacktype3.



which(gt$attacktype1 == 4 & gt$attacktype2 == 4 &
      gt$attacktype3 == 4)
# No repeating "4" for attacktype1 & attacktype2 & attacktype3

which(gt$attacktype1 == 5 & gt$attacktype2 == 5)
# No repeating "5" for attacktype1 & attacktype2.

which(gt$attacktype1 == 5 & gt$attacktype3 == 5)
# No repeating "5" for attacktype1 & attacktype3.

which(gt$attacktype1 == 5 & gt$attacktype2 == 5 &
      gt$attacktype3 == 5)
# No repeating "5" for attacktype1 & attacktype2 & attacktype3
```

```
which(gt$attacktype1 == 6 & gt$attacktype2 == 6)
gt$attacktype2[which(gt$attacktype1 == 6 &
                     gt$attacktype2 == 6)] <- NA
which(gt$attacktype1 == 6 & gt$attacktype2 == 6)
# No repeating "6" for attacktype1 & attacktype2.

which(gt$attacktype1 == 6 & gt$attacktype3 == 6)
gt$attacktype3[which(gt$attacktype1 == 6 &
                     gt$attacktype3 == 6)] <- NA
which(gt$attacktype1 == 6 & gt$attacktype3 == 6)
# No repeating "6" for attacktype1 & attacktype3.

which(gt$attacktype1 == 6 & gt$attacktype2 == 6 &
      gt$attacktype3 == 6)
# No repeating "6" for attacktype1 & attacktype2 & attacktype3

which(gt$attacktype1 == 7 & gt$attacktype2 == 7)
# No repeating "7" for attacktype1 & attacktype2.
which(gt$attacktype1 == 7 & gt$attacktype3 == 7)
# No repeating "7" for attacktype1 & attacktype3.
which(gt$attacktype1 == 7 & gt$attacktype2 == 7 &
      gt$attacktype3 == 7)
# No repeating "7" for attacktype1 & attacktype2 & attacktype3

which(gt$attacktype1 == 8 & gt$attacktype2 == 8)
# No repeating "8" for attacktype1 & attacktype2
# Note: attacktype3 only goes up to 7.

which(gt$attacktype1 == 9 & gt$attacktype2 == 9)
# No repeating "9" for attacktype1 & attacktype2

which(gt$attacktype2 == 9)
gt$attacktype2[which(gt$attacktype2 == 9)]
gt$attacktype1[which(gt$attacktype2 == 9)]
gt$attacktype2[which(gt$attacktype2 == 9)] <- NA
# Deleted "9" (Unknown from non-"9" attacktype1 observations)

which(gt$attacktype1 == 9 & gt$attacktype2 != 9)
which(gt$attacktype1 == 9 & gt$attacktype3 != 9)
# No "unknown" attacktype1 with known attacktype2 or attacktype3.


# nkill, nkillus
gt$nkill[gt$nkill == ""] <- NA
```

```
# Replace blanks with NA's
which(gt$nkill == 0 & gt$nkillus > 0)
# Vector indicating where nkill is 0 and nkillus is not 0.
gt$nkill[which(gt$nkill == 0 & gt$nkillus > 0)] <- gt$nkillus[
  which(gt$nkill == 0 & gt$nkillus > 0)]
# Let nkill == non-0 entry of the nkillus, if nkill == 0


which(is.na(gt$nkill) == TRUE & gt$nkillus > 0)
# Vector indicating where nkill is NA and nkillus is not 0.
gt$nkill[which(is.na(gt$nkill) == TRUE & gt$nkillus > 0)] <- gt$nkillus[
  which(is.na(gt$nkill) == TRUE & gt$nkillus > 0)]
# Let nkill == non-0 entry of the nkillus, if nkill == NA
which(gt$nkill == 0 & gt$nkillter > 0)
# Vector indicating where nkill is blank and nkillter is not 0; NA

# nwound, nwoundus
gt$nwound[gt$nwound == ""] <- NA
# Replace blanks with NA's
which(gt$nwound == 0 & gt$nwoundus > 0)
# Vector indicating where nwound is 0 and nwoundus is not 0.
gt$nwound[which(gt$nwound == 0 & gt$nwoundus > 0)] <- gt$nkillus[
  which(gt$nwound == 0 & gt$nwoundus > 0)]
# Let nwound == the non-0 entry of the nwoundus, if nwound == 0.

which(is.na(gt$nwound) == TRUE & gt$nwoundus > 0)
# Vector indicating where nwound is NA and nwound is not 0.
gt$nwound[which(is.na(gt$nwound) == TRUE & gt$nwoundus > 0)] <- gt$nkillus[
  which(is.na(gt$nwound) == TRUE & gt$nwoundus > 0)]
# Let nwound == the non-0 entry of the nwoundus, if nwound == NA
which(gt$nwound == 0 & gt$nwoundter > 0)
# Vector indicating where nwound is blank and nwoundter is not 0
```

2. Myanh ’s Portion

```
gt$region_txt[gt$region_txt == ""] <- NA
gt$longitude[gt$plongitude == ""] <- NA
gt$latitude[gt$latitude == ""] <- NA
gt$provstate[gt$provstate == ""] <- NA
# Filter out empty observations by filling in NA.

index_states <- gt$provstate[which(gt$country_txt == "United States")]
# Create a vector with US states with terrorism incidents.
```

```
gt$targtype1_txt <- factor(gt$targtype1_txt)

gt$targsubtype1_txt[gt$targsubtype1_txt == ""] <- NA
# Filter out empty observations by filling in NA.

sort(table(gt$targsubtype1_txt), decreasing = TRUE)
gt$targsubtype1_txt <-
  factor(gt$targsubtype1_txt, levels = names(sort(table(gt$targsubtype1_txt),
                                                  decreasing = TRUE)))

sort(table(gt$targsubtype1), decreasing = TRUE)
gt$targsubtype1 <-
  factor(gt$targsubtype1,
         levels = names(sort(table(gt$targsubtype1), decreasing = TRUE)))

sort(table(gt$targtype1), decreasing = TRUE)
gt$targtype1 <-
  factor(gt$targtype1, levels = names(sort(table(gt$targtype1),
                                           decreasing = TRUE)))

gt$country_txt <- factor(gt$country_txt,
                         levels = names(sort(table(gt$country_txt),
                                             decreasing = TRUE)))
sort(table(gt$country_txt), decreasing = TRUE)

# Factorize "txt" variables.
# Assign levels for each.
# Sort them in decreasing order.
# Chose the top 5 or 10 subcategories and do the analysis on them instead of all subcategory.
# Repeat the process if the variables have the same context.
```

ii. Data analysis

1. Myanh's portion

a. Location's world map

```
ggplot(data = world) +
  geom_polygon(aes(x = long, y = lat, group = group),
               color = "black", fill = "#4F4440") +
  geom_point(data = gt,
             mapping = aes(x = longitude, y = latitude, color = region_txt),
             size = 0.1) +
  labs(x = "Longitude", y = "Latitude", color = "Regions: ",
       title = "Terrorism Attacks In The World from 1999 to 2019") +
```

```
  coord_quickmap() + theme_light() +
  theme(plot.title = element_text(size = 12))
```

b. Location's histogram

```
gt$region_txt <- factor(gt$region_txt,
                        levels = names(sort(table(gt$region_txt),
                                            decreasing = FALSE)))
ggplot(gt,aes(x = region_txt, fill = region_txt)) +
        geom_histogram(stat = "count") + theme_light() +
  theme(axis.text.x = element_text(size = 7, angle = 90)) +
        labs(x = "", y = "Frequency",
           title = "Number of Attacks Per Regions from 1999 to 2019") +
  theme(legend.position = "none") +
  theme(axis.text.x = element_text(angle = 30, hjust = 1))
```

c. Location's US map

```
ggplot() +
  geom_polygon(data = usa, aes(x = long, y = lat, group = group, alpha = 1),
             fill = "#C5ab9f", color = "dark grey") +
  geom_point(data = subset(gt, gt$country_txt == "United States"),
           mapping = aes(x = longitude, y = latitude),
           color = "red", size = 1) +
  coord_quickmap() + theme_light() +
  labs(x = "Longitude", y = "Latitude",
      title = "Terrorism Attacks on United States (1999-2019)") +
  theme(legend.position = "none") + xlim(-135, -60) + ylim(20, 55)
```

d. Location's US map through years

```
ggplot() +
  geom_polygon(data = usa, aes(x = long, y = lat, group = group),
             fill = "#4F4440") +
  geom_point(data = subset(gt, gt$country_txt == "United States"),
           mapping = aes(x = longitude, y = latitude),
           color = "red", size = 0.2) +
  coord_quickmap() + theme_light() +
  labs(x = "Longitude", y = "Latitude",
      title = "Terrorism Attack on United States Through Each Year") +
  theme(legend.position = "none") + xlim(-135, -60) + ylim(20, 55) +
  facet_wrap(~iyear)
```

e. Target's boxplot

```
library(RColorBrewer)
mycolor = c(brewer.pal(name = "PiYG", n = 11), brewer.pal(name = "PuOr", n = 11))
ggplot(data = gt, mapping = aes(x = reorder(targtype1_txt, iyear, FUN = median),
                   y = iyear, fill = targtype1_txt)) +
  geom_jitter(size = 0.1, alpha = 0.1) +
  geom_boxplot() +
  labs(title = "Box Plot Shows Distribution of Target Types Through Years",
      x = "Type of target", y = "Years") +
```

```
  theme_minimal() + coord_flip() + theme(legend.position = "none") +
  scale_fill_manual(values = mycolor) + theme(title = element_text(size = 9))
```

f. Target's stacked bar plot

```
library(forcats)
ggplot(data = subset(gt, targtype1 == c(14,2,3,4,1))) +
  geom_bar(aes(x = fct_infreq(targtype1_txt),
               fill = region_txt), position = "stack") +
  labs(title = "Distribution of Top 5 Most Common Target Types Through Regions",
       x = "", y = "Frequency", fill = "Regions: ") + theme_minimal() +
  theme(axis.text.x = element_text(size = 5), title = element_text(size = 10))
```

2. Hongjoon's Portion

a. Data Frames Used

i. gt_1

```
gt_1 <- select(gt, iyear, region_txt, attacktype1, attacktype2, attacktype3, nkill, nwound, success)
# Create a sub-data set.

newattack <- rep(0,length(gt_1$attacktype1))
# Create a new vector to add as a variable to gt_1 data frame.

for (i in 1:length(gt_1$attacktype1)){
  if (is.na(gt_1$attacktype1[i]) == FALSE &
      is.na(gt_1$attacktype2[i]) == FALSE) {
    newattack[i] <- 10
  }
  else {
    newattack[i] <- gt_1$attacktype1[i]
  }
}
# Enter any observation with two or more distinct tactics as "10" in newattack.
gt_1$newattack <- newattack
# Create a new attacktype variable reflecting a combination of tactics.
```

ii. gt_3

```
# Create a new data frame
for (i in 1999:2019) {
  if (i == 1999){
    x2 <- c(i,unname(table(factor(gt_1$newattack[which(gt_1$iyear == i)], levels = 1:10))))
  }
  else if (i < 2019){
    x2 <- rbind(x2,c(i,unname(table(
      factor(gt_1$newattack[which(gt_1$iyear == i)], levels = 1:10
      )))))
  }
```

```
  else{
    x2 <- rbind(x2,c(i,unname(table(
      factor(gt_1$newattack[which(gt_1$iyear == i)], levels = 1:10
      )))))
    gt_3 <- as.data.frame(x2)}
}

# Re-named the Columns:
colnames(gt_3) <- c("Year","Assassination","Armed Assault","Bombing/Explosion",
"Hijacking", "Barricade Incident", "Kidnapping",
"Facility/Infrastructure Attack","Unarmed Assault","Unknown",
"Combination")

# Melt the data frame:
gt_3 <- melt(gt_3,id.vars = c("Year"))
```

iii. gt_8

```
x6 <- c(0)
for (i in 1:10){
a <- sum(gt_1$nkill[which(gt_1$newattack == i)], na.rm = TRUE)/length(which(gt_1$newattack
== i))
x6[i] <- a
}
gt_8 <- as.data.frame(x6)
colnames(gt_8) <- c("Deaths-Per-Event")
rownames(gt_8) <- c("Assassination","Armed Assault","Bomb./Expl.",
"Hijack", "Barricade", "Kidnap",
"Facil./Infrastr.","Unarmed Assault","Unknown","Combination")
```

iv. gt_6

```
gt_6 <- gt_1
for (i in 1:110096) {
  if (is.na(gt_1$nkill[i]) == TRUE) {
    gt_6[i,6] <- 0
  }
  if (is.na(gt_1$nwound[i]) == TRUE) {
    gt_6[i,7] <- 0
  }
}
# Turn all NA's in either nkill or nwound into zeroes
```

v. gt_7

```
for (i in 1:10){
  if (i == 1){
    x6 <- c(i,sum(gt_6$nkill[which(gt_6$newattack == i)]),
```

```
        sum(gt_6$nwound[which(gt_6$newattack == i)]))
  }
  else {
    x6 <- rbind(x6, c(i,sum(gt_6$nkill[which(gt_6$newattack == i)]),
        sum(gt_6$nwound[which(gt_6$newattack == i)])))
  }
}
gt_7 <- as.data.frame(x6)
# Create a data frame with cumulative sums of nkill and nwound per tactic.

colnames(gt_7) <- c("Tactic", "Dead", "Wounded")
# Renamed the columns

gt_7$Tactic <- factor(gt_7$Tactic, levels = c(1:10),
              labels = c("Assassination","Armed Assault","Bombing/Explosion",
"Hijacking", "Barricade", "Kidnapping",
"Fac./Infr.","Unarmed","Unknown", "Combo"))
# Renamed the tactics variable.

gt_7 <- melt(gt_7, id.vars = c("Tactic"))
# Melted the data frame

colnames(gt_7) <- c("Tactic", "Type", "Frequency")
# Renamed the columns
```

vi. gt_cum

```
gt_test <- select(gt,iyear, nkillter)
for (i in 1999:2019) {
  if ( i == 1999) {
    gt_cum <- c(i,
          sum(gt_test$nkillter[which(gt_test$iyear == i)], na.rm = TRUE))
  }
  else {
    gt_cum <- rbind(gt_cum, c(i,sum(gt_test$nkillter[which(gt_test$iyear == i)],
             na.rm = TRUE)))
  }
}
gt_cum <- as.data.frame(gt_cum)
colnames(gt_cum) <- c("Year", "Perp.Death")
```

b. Graphs Used

i. Terrorism Tactics By Region (Bar)

```
ggplot(data = gt_1, aes(x= region_txt, fill = as.factor(newattack))) +
geom_bar(position = "fill") + theme_light()+
labs(x = "Region",y = "Rel. Frequency",
```

```
title = "Terrorism Tactics By Region") +
theme(axis.text.x= element_text(angle=45, size=7.5, vjust = 1,
                                hjust = 1),
      legend.text = element_text(size = 6.5)) +
scale_fill_brewer(palette = "Set3", name = "Tactic Type",
labels = c("Assassination","Armed Assault","Bombing/Explosion",
"Hijacking", "Barricade Incident", "Kidnapping",
"Facility/Infrastructure Attack","Unarmed Assault","Unknown", "Combination")) + guides(fill =
guide_legend(title = NULL))
# Relative Frequency Bar Graph of Regional Tactics Proportions.
```

ii. Frequency of Terrorism Tactics 1999-2019 (Line)

```
ggplot(data = gt_3, aes(x = as.factor(Year), y = value, color = variable,
  group = variable)) +
  theme_light() + geom_line(size = 0.8, alpha = 0.75) + geom_point()+
  labs(title = "Frequency of Terrorism Tactics 1999-2019", x = "Years",
       y = "Frequency")  +
  theme(axis.text.x= element_text(angle=90, size=8),
        panel.background = element_rect(color = "black")) +
  scale_color_brewer(palette = "Paired", name = "Tactic Type",
labels = c("Assassination","Armed Assault","Bomb./Expl.",
"Hijack", "Barricade", "Kidnap",
"Facil./Infrastr.","Unarmed Assault","Unknown", "Combination"))
```

iii. Perpetrator Fatality (Bar)

```
ggplot(data = gt_cum, aes(x = as.factor(Year),
                          y = Perp.Death)) + theme_light() +
  geom_col() +
  labs(x = "Year", title = "Perpetrator Fatality") + theme(axis.text.x = element_text(angle = 90))
```

iv. Death Count from Terrorism 1999-2019 (Stacked Bar)

```
ggplot(data = gt_1, aes(x = as.factor(iyear), y = nkill,
                        fill = as.factor(newattack))) +
  geom_col() +
  scale_fill_brewer(palette = "Paired", name = "Tactic Type",
labels = c("Assassination","Armed Assault","Bomb./Expl.",
"Hijack", "Barricade", "Kidnap",
"Facil./Infrastr.","Unarmed Assault","Unknown", "Combination")) +
  labs(title = "Death Count from Terrorism 1999-2019",
                   x = "Year", y = "Death") + theme_minimal()+
  theme(axis.text.x = element_text(angle = 90, size = 7))
```

v. Death-to-Wounded Ratio (Side-by-Side Bar)

```
ggplot(data = subset(gt_7),
      aes(x = Tactic, y= Frequency, fill = Type)) + theme_light() +
  geom_col(position = "dodge") + theme (axis.text.x = element_text (angle = 45, vjust = 0.8,
hjust = 0.75, size = 7.5), axis.text.y =  element_text(size = 10)) +
  scale_y_continuous(labels = comma) + labs(title = "Death-to-Wounded Ratio") +
geom_text(aes(label = Frequency), position = position_dodge(1), size = 2.25, vjust = -0.4)
```

vi. Success Rates of Tactics (Stacked Bar)

```
ggplot(data = gt_1) +
geom_bar(mapping = aes(x= as.factor(newattack), fill = success),position = "fill") +
theme_light() + labs(title = "Success Rates of Tactics", y= "Rel. Freq", x = "Tactic") +
  scale_x_discrete(labels = c("Assassination","Armed Assault","Bombing/Explosion",
"Hijacking", "Barricade Incident", "Kidnapping",
"Facility/Infrastructure Attack","Unarmed Assault","Unknown",
"Combination")) +
  theme(axis.text.x = element_text(angle = 45, size = 8,
                                   hjust = 1))
```