Terrorism Trends in Urban Environments

Lukas Bretzinger, Cameron Reed, Sascha Schuster

Tuesday, December 09, 2014

Contents

X	Date... Place	On.UC	Killed... Wounded	Rel..Light	Rel..Access	Rel..Prox..Coast	Rel..Density
113575	2003-11-14, Istanbul	yes	162	96.83	97.62	98.26	81.97
113739	2008-07-27, Istanbul	yes	171	96.83	97.62	98.26	81.86
113414	1999-07-30, Gurpinar	no	4	77.78	0.00	100.00	72.20
113490	1999-03-05, Cankiri	no	14	52.38	99.60	74.08	2.50
113582	2003-08-01, Ankara	yes	11	100.00	98.81	65.96	7.58
113393	1998-11-27, Kirikkale	no	24	77.78	99.40	61.12	6.03

X	Date... Place	On.UC	Killed... Wounded	Rel..Light	Rel..Access	Rel..Prox..Coast	Rel..Density
113492	1999-10-18, Bademli	no	1	11.11	57.54	84.91	9.39
113403	1998-09-09, Kuyucak	no	4	19.05	97.02	80.08	6.58
113671	2006-08-28, Antalya	no	69	95.24	99.80	99.23	4.04
113675	2006-08-28, Antalya	no	23	95.24	99.80	99.23	4.04

Topic and Relevance

In the past five years, violent extremist groups have attracted significant attention using urban centers as stages. In 2008, Pakistan-based Lashkar-e-Taiba launched a coordinated assault on Mumbai, India which lasted three days and resulted in 164 civilian deaths. In 2013, Somalia-based Al-Shabaab infiltrated the Westgate mall in Nairobi, Kenya, killing 67 and injuring 175 civilians. Terrorist groups are a tremendous threat to vulnerable cities. In both cases, the groups exploited with relative ease the complex flow of systems in the city to achieve desired outcomes. This sparked a debate about terrorist groups' targeting practices and on how to locate, understand, protect against, and mitigate attacks in the face of megatrends like urbanization, population growth, migration, and connectedness. The urban environment emerges as a unit of analysis for acts of political violence. Future conflicts are expected to take place in crowded, urban, coastal and connected environments instead of landlocked, remote and rural ones. Especially megacities present fertile grounds for inequality and conflict, as well as the most connected, large, yet vulnerable human and physical sub-system for terrorist attacks. Since 9/11, the fields of geography, urban studies, political science, and many others have contributed to the understanding of terrorist attacks, strategic decision-making, and geospatial aspects of terrorists' targeting behavior. We conduct a large-N pattern analysis using the Global Terrorism Database [@START2013] to illuminate trends in targeting behavior. Our study aims to fill a gap: On a global scale, is there a trend for terrorists to target urban over rural environments?

Literature Review

General

Urban terrorism is by no means a new concept, as it arises in guerilla, riot, and insurgency literature [@Crenshaw1981; @Grabosky1979; @Grabosky1988; @Karber1971; @Laqueur1977; @Lupsha1967]. Indeed, the urban context was a feature of Walter Laquer's early study of guerilla warfare. Groups like the Mau Mau of Kenya, the Irish Republican Army, and the Front de Liberacion Nationale employed tactics against colonial powers that exhibited characteristics of urban life and space [@Laqueur1977]. Martha Crenshaw's seminal piece on the causes of terrorism highlighted the enduring importance of modernization and urbanization, which she saw as an ever growing opportunity for terrorist groups to execute attacks [@Crenshaw1981]. Others highlight the urban space as a ground for recruitment, hiding and communication, as well as the variety of possible targets [@Grabosky1979]. Most of the literature leading up to 9/11 recognized the urban element, but did not fully address linkages between urbanization and state security. The literature explains various reasons for targeting practices, such as instrumental means, i.e., resources and capabilities driving strategy, organizational survival, management, human resources, and funding constraining strategic decisions, or ideological and religious motivation. More specifically, targeting behavior has been growing in importance in terrorism literature due to its policy relevance and potential for prediction. Martha Crenshaw has shed light on how groups shape motivations behind targeting, such as religiosity, communal ties, or other less tangible, intrinsic motivations [@Crenshaw1981]. Todd Sandler built strategic game theory models that parse potential decisions with constraints of resources and capabilities taken into consideration [@Enders2000]. Other theorists believe that organizational survival drives decisions, such as Jacob Shapiro's economic insights into how covert organizations are limited by their ability to fund operations, also known as rational choice [@Shapiro2007]. Finally, scholars have thoroughly explored characteristics of organizations that elongate or decline lifespans, including analysis of favorable conditions for organizational survival [@Blomberg2011].

Of course, these theories are not mutually exclusive and they collectively strengthen the understanding of strategic considerations in terrorist decision-making. However, few theories of strategy take into consideration geospatial aspects or variation in targeting behavior. Consider the following questions. Does the organization want to control territory? Or do they rather want to make a point, provoke a government in the realm of national symbolism, or attract attention? Apart from explaining internal factors of the organization, we can attempt to explain geography-bound terrorism, such as considerations of spatial and hierarchical diffusion, currently nascent in the literature on terrorist group dynamics [@Bahgat2013]. Spatial diffusion refers to one base of operations close to a series of attacks, whereas hierarchical diffusion is the existence of several hot spots from which attacks emanate. These phenomena are supported by findings on geolocated IRA attacks in Northern Ireland and Great Britain. The difference here in spatial and hierarchical diffusion refers to a growing sense-making literature on the spatial logic of terrorism, but also growing importance of, as Bahgat and Medina put it, how cities of high population and administrative worth to the government appear to have become the main targets of modern-day terrorism for a variety of strategic and cost-effective reasons. [-@Bahgat2013].

Prominent theorists of conflict re-problematize the issue, taking into consideration megatrends of urbanization, population growth and connectedness. Megacities are their unit of analysis for studying conflict [@Kilcullen2013]. Geographers have offered a research agenda for their field in the face of the rising importance of the ancient social phenomena of terrorism [@Cutter2003]. It is often assumed that growing and developing urban environments become increasingly attractive as targets for violent extremists, resulting in more attacks on urban systems, but mostly single case studies like in the Mumbai 2008 case are referenced in support of that claim-larger empirical studies are rare [@Beall2006; @Glaeser2002; @Graham2008; @Sassen2010; @Savitch2001]. It said to be a traditional characteristic of terrorism studies that much is written on the basis on little empirical analysis [@Jongman1988]. To our knowledge, a comparable study exists only on the geolocation of terror attacks on the U.S. level [@Webb2009].

Statistical Methods, Research, and GTD Use

The Global Terrorism Database (GTD) has been employed in different fields and various ways to study terrorism. Also, it has been used to study specific categorical phenomena or regions, e.g, hostage-taking or weapon types, as well as sweeping trends, such as casualty rates due to terrorist attacks. The research methods followed include geographical mapping, descriptive statistics, and qualitative inquiry [@LaFree2009a]. There are essentially three camps of researchers in the applied fields using the GTD:

1. Geography
2. Political Science/International Relations
3. Terrorism Studies

In most of the cases, there is a convergence of the three, albeit to different extents. For example, geographers use the data to make sense of geospatial path dependencies of terrorist groups and social network analysis (physical and human geography), whereas political scientists undertake more rigorous qualitative analysis and counter conventional beliefs about security and war. Often, terrorism scholars will focus on one category, region, group, or trajectory. LaFree is producing the most comprehensive review of uses of the GTD, which comes out in 2015. He also has written extensively on how to use the GTD to counter widely held beliefs regarding terrorism [@LaFree2009]. Correlation work between terrorism and economic, legal, or group characteristic data also exists [@LaFree2010]. This body of work might be viewed as functional. Geospatial analysis of these trends gained traction after 9/11 and have since been growing in importance and policy relevance. Operating under the assumption that large-scale attacks are planned, geospatial analysis, or mapping other types of analysis onto geographical illustrations, provide insights into strategies of targeting and decision-making. In fact, applied geographers have discovered trends in the environment that signal a trend toward targeting areas with high populations [@Bahgat2013]. The cross-fertilization of the three fuels our study.

Research question: What are the trends in terrorism targeting urban vs. rural spaces? This is a purely descriptive question, looking at what the data can show us up to this point.

Data

Data Set

Data: Global Terrorism Database (GTD) [@START2013]

We will use the START Global Terrorism Database (GTD), as it is the most comprehensive open source database on terrorist attacks [@LaFree2006]. The data ranges from 1970-2013, logs 125,000 terrorist attacks, and uses 45 - 120 variables per attack. Among other information, the GTD holds records on the location, the target, and the damage caused by attacks [@START2014]. It is a simple .xls file, available after creating an account on the GTD Projects website and it is already tuned towards being turned into a .csv, as close to no excel functions are layered over the data entry. It contains both numeric and factor variables for describing the attacks characteristics. All categorical variables have both categorical numbers and a respective text variable for each number. This creates a lot of redundant information and needs a long tidying process. The current GTD is the product of several phases of data collection efforts, each relying on publicly available, unclassified source materials. These include media articles and electronic news archives, and to a lesser extent, existing data sets, secondary source materials such as books and journals, and legal documents [@START2014]. We are aware of the entangled problems of reliability and comparability. The original set of incidents that comprise the GTD occurred between 1970 and 1997 and were collected by the Pinkerton Global Intelligence Service (PGIS) a private security agency. PGIS data collection efforts are remarkable in that they were able to develop and apply a similar data collection strategy for a 28-year period [@LaFree2006]. After START completed digitizing these handwritten records in 2005, they collaborated with the Center for Terrorism and Intelligence Studies (CETIS) to continue data collection beyond 1997 and expand the scope of the information recorded for each attack [@START2014]. CETIS collected GTD data for terrorist attacks that occurred from January 1998 through March 2008, after which ongoing data collection transitioned to the Institute for the Study of Violent Groups (ISVG). ISVG continued as the primary collector of data on attacks that occurred from April 2008 through October 2011.

GTD staff based at START headquarters at the University of Maryland integrated and synthesized data collected across the entire 1970-2013 time span with the goal of ensuring that the definitions and methodology are as consistent as possible across all phases of data collection. In addition, GTD staff at START retroactively coded several key variables not originally available for the PGIS cases, conducted numerous quality control projects, and supplemental data collection efforts. These supplemental data collection efforts involve systematically comparing a variety of additional sources of terrorism incident data to the GTD to identify any missing events that satisfy GTD inclusion criteria. GTD staff research these missing events to identify primary sources of information and code the attack details for addition to the GTD. Beginning with cases that occurred in November 2011, all ongoing GTD data collection is conducted by START staff at the University of Maryland. Additional information on the history and data collection methodology of the database can be found on the GTD website [@START2014]. Given the varied context of GTD data collection, the database is another source of general inconsistency-legacy problems. The GTD now includes incidents of terrorism from 1970 to 2013, however a number of new variables were added to the database beginning with the post-1997 data collection effort. Wherever possible, values for these new variables were retroactively coded for the original incidents, however some of the new variables pertain to details that were not recorded in the first phase of data collection. For any newly added variables that were not retroactively coded, they only exist for post-1997 cases. GTD is based on PGIS and PGIS is the most granular and comprehensive. To illustrate how consequential these coding differences are we compare terrorism event counts for 1997 between the PGIS database and the U.S. State Department terrorism database. In that year, the Department of State records 304 acts of international terrorism, which caused 221 deaths and 683 injuries. For the same year, the PGIS data reports on 3,523 acts of terrorism and political violence that claimed 3,508 lives and inflicted 7,753 injuries [@LaFree2006].

Statistical Methodology

Variable: Urbanity of terroris' targets. The GTD defines a terrorist attack 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. In order to consider an incident for inclusion in the GTD, all attributes must be present. The database does not include acts of state terrorism. In order to use the database, we have to stick to the definitions given, therefore limiting us to non-state terrorism.

We will build on definitions of urbanization from the standpoint of urban studies and security, while geography will help us mostly in finding measurable indicators. To speak to literature on conflict, we conceptualize urbanity to include a static, geospatial dimension, not purely socially co-constructed spaces of human interaction / communication alone. How can we still cover connectedness, embeddedness, or a sense of being networked or plugged in? In other words, what parts of the constructed environment area the human environment, and how can we justify and measure that? It has to be permanent, relatively densely populated, and confined in terms of economic, cultural and social characteristics. The interconnectedness and interdependency (partly expressed in infrastructure of different kinds) of urban life is part of our concept, on top of being a permanent and dense human settlement. Urbanity is measured on a continuous spectrum. It is not a binary concept. The measure for the urbanity of the target has two components for us: location and target. To give an idea on how to measure both:

1. *Location*: There are many ways to determine, if an attack took place in an urban environment. As GPS coordinates are available for the time after 2001, potentially one could find a way to include land use, population density and proximity to urban center in the analysis of the attack. This likely exceeds resources and time available to us. If we include the idea that our research interest lays in the intention to target urban life and space, then we can use another measure. This would be the relative size and importance of the city compared to the national and regional environment. We assume that a terrorist organization can choose between targets at least on a national level, and put together our own probably quickly hand-collected dataset with city like: 100 largest cities of the for each decade; largest cities for each county (between 2 and 25 depending on the country's population determined with a simple mathematical formula); coastal megacities for each decade; world Cities for each decade. To combine the lists in a data frame, the vector names might be: Name of the city, hundred_largest_70s, hundred_largest_80s, hundred_largest_90s, hundred_largest_20s, 100 largest 21s, large_national_decade, megacity_decade, worldcity_decade, costal, capitals whereas anything but the name is a 0-1 binary.
2. *Target*: The target of the attack is coded in the GTD dataset with target and subtarget type. In order to weight them in terms of their representation of urbanity, we created categories of targets that represent aspects of urban life and space, with varying degrees: Expression of urban life (restaurant, hotel, apartment, etc.); infrastructure necessary to sustain urban life (water supply, port, electricity, etc.); employment (construction, factory, multinational corporation); police and governance; military and rural.

Merging the datasets over cities will be easy, but special cases of city renaming must be accounted for, like East and West Berlin/Berlin, Bombay/Mumbai, etc. The combined dataset basically adds to the GTD a variable for potential urbanity of the targets and adds variables of quality to various cities due to their respective characteristics in relation to their national and regional peers.

Analysis

The analysis will focus on multivariate statistical regression models for describing a potential trend in the urbanity of terror attacks in terms of the chosen city and the chosen target over time. Descriptive inference such that our results will leads us to reject a null hypothesis that attacks have not changed over time in targeting urban life and space that goes beyond the growth of this space over time.

To include the success of the attacks, weights for killed and injured humans will be added to each attack, as well as property damage in a second round. It might be necessary to manipulate the property damage in terms of the relative wealth of the country to account in country variation in the potential wealth that can be destroyed. For all our results, we will conduct a significance-test.

Necessary controls are the:

- Relative growth of urban life (populations and space on the country scale [@worldbank])
- Civil war (1-0 dummy from the Correlates of War [@COW2007] Project, Intra-State War Database 4.0 which is the most updated one)
- Capital Cities (coded by hand)
- Changes of collecting entities: PIGS until 1997, CETIS until 2008, ISVG until 2011 and START since 2011 (all in the GTD)

To elaborate on the last point, we created three preliminary charts to shop how transition in data collecting entities influences variables of our interest.

After finding a possible correlation between time and targets urbanity and locations urbanity, analyzing trends in targeting behavior on the regional, country, or group level will be easy to execute as the GTD contain variables for all attacks.

Data Collection Process

Data Categories & Sources

We use three main categories of data, which stem from a different number of sources and serve different purposes.

Global Terrorism Database

We have introduced the [GTD] (http://www.start.umd.edu/gtd/) [@GTD] extensively in the last assignment. It gives qualitative data on about over 120k terrorist attacks, including (in about 2/3 of the observations) information that can be used to georeference the attack.

Geolocated City-Data

We used two open-source datasets of city level data that we need to establish a relation between the place of the attacks and their urbanity.

a. "world.cities" from the R package 'maps'. The database "is primarily of world cities of population greater than about 40,000. Also included are capital cities of any population size, and many smaller towns." [@mapspackage] The variables include the city name, country name, approximate population (as of January 2006), latitude, longitude, and capital status indication.

b. "worldcities2013" from MaxMind Inc.[@worldcities]. This data set provides similar information, but is updated more regularly.

c. "Urban Centers" from wikipedia. In the absence of a free data set on urban centers, we scraped a list with around 500 urban centers (>1 million inhabitants) of the [respective Wikipedia page] (http://en.wikipedia.org/wiki/List_of_urban_areas_by_population)[@wikiUCs]. It draws from seven different types of sources and is put together in terms of defining urban space and urban centers. We added a hand-coded "coastal city" variable to indicate if a city is close to the coastline and has a port.

Note: All three datasets are time invariant. Since we did not find comprehensive data containing city-level data over the past years (which is a crucial requirement for our analysis), we need another data set for country-level data.

Country-Level Data

Our source for country-level data is the set of World Development Indicators (WDI) provided by the World Bank. We download them using the WDI package for R, a shortcut to the World Bank's API that provides data already formatted in long country-year format [@WDIpackage].

Additional Data

In the future, we plan to include additional data that helps us control for phenomena affecting our analysis.

a. For example, we are working on including civil war dummy variables because civil wars are likely to exponentially increase the amount of terror attacks in a given year and city. It comes from the Correlates of War project and is called the Intra-State War Database 4.0 [@COW].

Data Cleaning

Challenges in All the Data Sets

Missing Information

None of the datasets used can be considered complete with regard to the individual observations. In fact, they contain a huge number of NAs. The subset of the GTD the we use for our analysis (containing only 18 of the original 123 variabels, and only successful terror attacks) has 107143 NA values, summing up to a total of 5.2% of all values. We aim not to have a drastically increased share of NA values in the dataset used for the final analysis. All datasets are very comprehensive and stem from sources with high reputation. An extensive cleaning process was necessary nonetheless.

Spelling Inconsistencies

The main challenge across and within all datasets is the huge variation in spelling of countries and cities, which triggered an extensive hand recoding process. We developed a standardized style for country and city names and applied that to all data sets.

- GTD ~ 120k rows
- cities.a. ~ 50k rows
- cities.b. ~ 50k rows
- WDI ~ 10k rows
- Urban Centers ~ 500 rows
- War ~ 500 rows

Coding Gaps, Information Inconsistencies, and Lack of Detail

All datasets containing georeferenceable data contained this information on varying scales and for different time periods. For example, while some attacks in the GTD were probably geolocated using GPS guidance, others lack their own geoposition and are only presented using the central point of the city or district. When possible, we tried to define position data.

A huge gap existed between the WDI data and the GTD. The GTD assigns attacks to the countries they took place in at the time they happened. However, these countries (Soviet Union, Yugoslavia, GDR,etc.), in some cases, do not exist anymore. The WDI, on the other hand, contains country-level data back to 1960 in the form of countries as they are today.

Data Cleaning Process

We brought all country names to the standard of the World Bank data as a point of reference and because we will draw most of our country level data from there.

Although we combine the two world city datasets, we decided not to bring the city names to the same standard before merging them into the GTD. This has to do with the sort and amount of inconsistencies mentioned above: The more (even inconsistent, wrong, or outdated) city names we have in the world city datasets, the higher our chances to match them with cities mentioned in the GTD (even if by the coincidence of matching typos that we may have overlooked).

Because of the abysmal quality of the city_txt variable in the GTD, at least 750 lines of code were necessary to bring the ~2,5k unique city names to a level in which we could work. Codings like "somewhere at the border" or up to 10 typos (from "Buen%%s Eir$" to "Buenos Aires") for a heavily targeted city are not unusual.

Merging Process and Current Status

First, we merged the WDI country level data into the GTD by country and year. These indicators contain information on population sized in different settings (living in largest city, living in urban environment, etc.) per country and year.

Second, we merged the two city data sets. We eliminated duplicates, keeping either the city entry that was truthfully coded as capital or the one with the higher population (we ended with ~50k rows + ~50k rows = ~80k rows). As we use them to merge with the cleaned GTD city_txt variable, the more cities in our dataset, the better.

The third step is the most computing intensive one so far: We merged the urban center dataset with the now combined city dataset, assigning each city to its nearest urban center. The reasoning behind this step is that, while we have around 50k different cities in our GTD, only a share of them fulfills the requirement of being "urban" the way we understand it. A small or big distance between the city the attack took place and its closest urban center may serve as a rudimentary indicator for an intent to attack urbanity.

Therefore, we include lat/long data for each urban center using the google maps API. Then, the distance from each urban center to each city was calculated. The merged dataset assigns the closest urban center to each city (and the respective distance). The necessity comes from the way cities are coded in the GTD. While an attack on Tokyo, which is rarely attacked, is usually coded using "Tokyo", attacks in often targeted cities are usually localized more precisely - assigned to districts. Good examples for this phenomenon are Lima, or the urban area around Tel Aviv. Both are attacked often and the GTD delivers predominantly the sub-municipality as the place of attack.

With the new dataset, we can set a parameter of distance around each urban center (as a place holder we currently use $2 * sqrt(urban - centers - area/pi)$, and later decide to count any attack that falls into that parameter as an attack on the urban center itself. If the GTD codes "New York City", it finds both the urban center and the city - but as the GTD sometimes codes "Manhattan", we now have a match on the urban center "New York City" as the distance between the two falls within our parameter.

Finally, we merge the GTD and the combined city-urbancenter dataset. We use a merging varaiable which is a clean character string of the form of *countrynamecityname*, in order to avoid false positives of similar city names across countries. Thanks to our previously unified country and city coding in all datasets, we find a

city (thus, population size and also closest urban center) for around 60% of all 120k terror attacks in the GTD. As the GTD often lacks any city name and has "unknown" or area codings (e.g. "District xyz"), 60% is a satisfying result given complexity and resource constraints.

Currently, We are working on increasing the robustness of our observations by cleaning further. The google maps API might provide for further analysis over lat/long calculated distances to cities within, e.g., Arabic-speaking countries with rivaling city names in the latin alphabet.

To Do Before Analysis

1. Include the 0-1 war variable in the country level data.
2. Continue cleaning process in order to increase usable observations (e.g. Drive up the matches between our combined city-urbancenter dataset and the GTD by analysing "messy" coded countries like India, Sri Lanka, and Arab countries. We did this already with Iraq and had promising results.)
3. Defend assumptions on choice of urban center radius, population growth on the city level, aggregation or disaggregation of variable values, etc.
4. Look for further helpful sources for control variables and other phenomena impeding our analysis.
5. Include population data 1970 - ~2000 into the GTD by combining WDI with city data. This could happen in the following form.

Preliminary Analysis

We have a large amount of information on each incident in the GTD already. This includes:

- Time
- Country
- WDI Data for county and year
- City
- Data on the City for 60% of the incidents including
- Population estimate
- Capital City or not
- Distance to Urban Center
  - Population of that Urban Center
  - Area and Population Density of that Urban Center
  - Coastal Location of the Urban Center
- Attack Type (Bomb, Assault, Hostage Taking etc.)
- Target Type (eg. Restaurant, Electricity Grid, Military Installation)
- Number of Killed and Wounded
- Economic Damage from the Attack

Examples of what we can say (tables, figures)

With our preliminary GTD (PreGTD), we can already look at a lot of different information that helps us understand the distribution of terror attacks across either time *or* space (time and space is not possible since we do not have population data available on a city level across time so far). We use this information to reveal pecularities in the data that we need to investigate. We can use this to either explain features or direct us towards further data cleaning.

So far, our analysis has given us results along the line that we suspected to find or that is in line with other resarch. For us, this means that our way of approaching the problem and conducting our research is suitable.

References