

## RESEARCH ARTICLE

# Exploring Twitter to Analyze the Public's Reaction Patterns to Recently Reported Homicides in London

Ourania Kounadi<sup>1\*</sup>, Thomas J. Lampoltshammer<sup>1,2</sup>, Elizabeth Groff<sup>3</sup>, Izabela Sitko<sup>1</sup>, Michael Leitner<sup>1,4</sup>

**1** Doctoral College GIScience, Department of Geoinformatics-Z\_GIS, University of Salzburg, Salzburg, Austria, **2** School of Information Technology and Systems Management, Salzburg University of Applied Sciences, Salzburg, Austria, **3** Department of Criminal Justice, Temple University, Philadelphia, Pennsylvania, United States of America, **4** Department of Geography and Anthropology, Louisiana State University, Baton Rouge, Louisiana, United States of America

\* [raniakounadi@gmail.com](mailto:raniakounadi@gmail.com)



## OPEN ACCESS

**Citation:** Kounadi O, Lampoltshammer TJ, Groff E, Sitko I, Leitner M (2015) Exploring Twitter to Analyze the Public's Reaction Patterns to Recently Reported Homicides in London. PLoS ONE 10(3): e0121848. doi:10.1371/journal.pone.0121848

**Academic Editor:** Lidia Adriana Braunstein, UNMdP-CONICET, ARGENTINA

**Received:** July 29, 2014

**Accepted:** February 19, 2015

**Published:** March 26, 2015

**Copyright:** © 2015 Kounadi et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** This research was funded by the Austrian Science Fund (FWF) through the Doctoral College GIScience at the University of Salzburg (DK W 1237-N23). Part of the research was also funded by the Salzburg University of Applied Sciences.

**Competing Interests:** The authors have declared that no competing interests exist.

## Abstract

Crime is an ubiquitous part of society. The way people express their concerns about crimes has been of particular interest to the scientific community. Over time, the numbers and kinds of available communication channels have increased. Today, social media services, such as Twitter, present a convenient way to express opinions and concerns about crimes. The main objective of this study is to explore people's perception of homicides, specifically, how the characteristics and proximity of the event affect the public's concern about it. The analysis explores Twitter messages that refer to homicides that occurred in London in 2012. In particular, the dependence of tweeting propensity on the proximity, in space and time, of a crime incident and of people being concerned about that particular incident are examined. Furthermore, the crime characteristics of the homicides are analysed using logistic regression analysis. The results show that the proximity of the Twitter users' estimated home locations to the homicides' locations impacts on whether the associated crime news is spread or not and how quickly. More than half of the homicide related tweets are sent within the first week and the majority of them are sent within a month of the incident's occurrence. Certain crime characteristics, including the presence of a knife, a young victim, a British victim, or a homicide committed by a gang are predictors of the crime-tweets posting frequency.

## Introduction

This study contributes to the existing literature of crime perception. The main objective is to explore citizen's perception of homicides. In particular, how the characteristics and proximity of the event affect the public's concern about it as reflected by the timing and content of tweets. Our research fits into the topics of fear of crime, which is addressed as a psychological reaction [1–3] and concern about crime, which is considered a social problem [4–5].

Regarding the spatio-temporal aspects of crime perception, we want to examine if there is any dependence on the proximity in space or time of a crime incident, and the people who are concerned about that particular incident. The study of Lewis and Maxfield [6] surveyed residents in four Chicago neighborhoods and revealed that the fear of crime is a problem independent of the actual crime rates. In addition, Furstenberg [5] revealed that people living in low crime areas are more concerned about crime than people living in high crime areas. Further studies examined how landscape patterns affect the fear of crime and crime prevalence estimates. The results of Wang and Taylor [7] showed that safety concerns varied as respondents walk down two different urban alleys that varied in terms of their lengths, land use patterns, and site features. Controlling for position in the alley, two elements were linked with the amount of fear respondents experienced. These were the expected direction to refuge for attackers, a composite element of the proximity of escape routes and how well one can see what is coming. McCord et al. [8] focused on variations in land use and confirmed a positive association between the “crime-generator index” or the “crime-attractor index” and of perceived incivilities and crime. Similarly, Williams [9] surveyed residents of a small geographical area in London regarding concern about crime types. The results were mapped so as to reveal variations in the level of fear of crime in space. The author and the residents interpreted these variations based on environmental and landscape characteristics (e.g. “the streets are dirty”). Furthermore, to the best of our knowledge there is only one study that addresses the relationship between the temporal behavior of people concerned about a crime, and the temporal attributes of the crime itself. Heverin and Zach [10] examined Twitter messages that referred to the shooting of four police officers in Washington State, USA in November 2009. Relevant tweets were retrieved for the next two and a half days after the event, for five time periods of twelve hours each. Then, the tweets were categorized by types (e.g. emotion or opinion) and the percentages of each type per time period were analyzed. The authors proposed that text mining of microblogging is a promising method to disseminate crisis-related information. With respect to crime characteristics, we want to investigate if certain characteristics are associated with the public's higher or lower concerns. Empirical studies showed that characteristics of crimes affect people's reactions, feelings, and opinions about them. For instance, official rates of violent crimes [11] and direct victimization [5] are positively related to fear of crime. Other explanatory variables of crime perception that have been examined in the literature are: a) locality, sensationalism, and randomness of crime [12–14], b) types of crime visualization [15], and c) whether justice is restored or not [16].

To address our research objective, we use two data sources. The first one is crime reports from online newspapers. Newspapers were chosen since previous studies in several regions worldwide [17–19], as well as in the UK, which is the study area of this research [14], [20], [21], suggest that media affects peoples' perception. More important, both in the UK [21] and elsewhere [15] media, and in particular newspapers, are the primary source of crime-related information. The second data source is the Twitter social network and microblogging service that is used to explore public's concerns about crime reports. Recent publications applied data mining techniques to social media content to examine perceptions and other behavioral patterns of individuals over several topics of interest. Amongst these topics, the use of social media regarding health related research [22–24], risk management and natural disasters [25–27] have gained increasing interest in the current literature. Also, social media platforms have been used as a tool to sense collective human behavior and uncover mobility patterns at different geographical scales [28], [29]. Other studies focused on developing topic detection techniques [30], [31], which identify clustering of words that frequently appear together in large volumes of unstructured text. Others explored the geo-coded portion of social media to show its novel potential for geographical analysis and visualization [32–35]. Furthermore, context

from social media has been used to mine information for crime analysis and investigation. White and Roth [36] presented the application “TwitterHitter” that creates geovisual analytics to harvest topics of interest. The application was demonstrated for the purposes of crime analysis and the authors provided examples of how it may be used for criminal investigative analysis, intelligence analysis, tactical analysis, strategic analysis, and administrative analysis. Sun and Ng [37] developed the “Comment Arrival Model”, which reveals the social media messages’ popularity. They analyzed the “drug abuse” topic and suggested that the model can be used for crime detection. Concerning crime prediction, Wang et al. [38] used Twitter data combined with natural language techniques and liner modeling as an alternative approach to traditional methods for crime prediction. Said et al. [39] extracted and analyzed social networking data files using different types of smart phones and proposed to forensic analysts and security specialists the great potential of social media mining for digital crime investigation. Last, with regards to text mining in crime intelligence analysis, Helbich et al. [40] proposed an innovative approach using self-organizing maps and point pattern analysis. The objective of their research was to explore information extracted from people’s statements regarding the last victim of a homicide series in Jennings, Louisiana in the USA.

Our motivation to explore crime perception is based on literature findings that reveal variations of perception in several dimensions such as spatial, temporal, and crime characteristics. Our approach is to combine data sources (i.e. media reports and social media text) as an alternative to traditional sources (surveys or census data). Finally, our research utilizes methods that were developed by Lampoltshammer et al. [41]. The authors analyzed variations of crime perception by crime type, demographic, and spatial characteristics. In our study we adopt the data-processing methods “*Links Correspondence Method*” and “*Home Estimation Method*” and build upon the limitations and suggestions mentioned by the aforementioned authors. Firstly, the validation of the previous study’s results was limited due to the small available sample size. Secondly, empirical results indicate that violent crime has a stronger effect on people’s perceptions than other crime types [13], [41]. Consequently, we analyze and validate an “as complete as possible” dataset that contains homicides reports. Also, we develop an extended version of the analytical methodology. We believe that the additional analytical parts of this study account for the most important aspects influencing the public’s concern about crime. The first aspect is the temporal variations of concerns and the second aspect is the spatial variations of concerns. Here we want to investigate the spatio-temporal relationships between homicides-related tweets and homicide incidents. The third aspect is the variations of concerns due to variations in crime characteristics. In particular, we want to examine whether the characteristics of homicides can explain the public’s reaction to crime news on Twitter. Details about the study’s data sources and primary datasets are described in **Section 2**. Then we present, in separate sections (3–5), the additional databases, subsets, methods, and analyses that were used to address each of the three aspects. Last, **Section 6** discusses the study’s findings.

## Data

This section presents the acquisition and processing steps of the data sources to obtain the study’s primary datasets. As the basis for our study we stored all homicide cases (in total 98 incidents) from “[www.murdermap.co.uk](http://www.murdermap.co.uk)” from the year 2012 [42] in our *Homicide-Dataset*. This website is dedicated to mapping every single homicide in London, starting from the crime itself up to the conviction of the perpetrator. The operators of the website acquire the information directly from the police, reports in media, and court reports. All information shared via the website is freely available. The information contains details about the victim (name, gender, date of crime, nationality, and the homicide’s weapon), as well as about the perpetrator (name,

gender, and nationality). In addition, the status of the crime (unsolved/solved), the location of the incident, and a category of the general type of crime are provided.

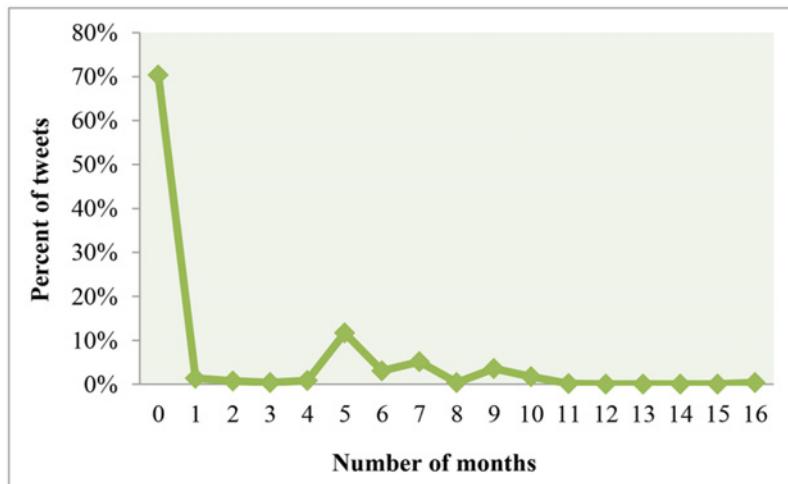
We searched up to ten web links that pointed to respective media reports. We selected, and then verified, the links based on a maximum of the first 10 URLs (uniform resource locator) that were found with Google search using keywords from the incidents. Keywords included but were not limited to combinations such as: perpetrator's name and victim's name or weapon and location and victim's gender and date. Several combinations were applied to obtain as many URLs as possible for each incident. The majority of the URLs were retrieved from UK-based online newspapers and few URLs from blogs and other websites located in the UK or in other countries. Even though all incidents were publicly available and reported on the internet, we retrieved different numbers and types of links for each homicide. For instance, we found many links for some homicides while for others only a few. Some were covered by worldwide and well-known newspapers, while others by local newspapers or small blogs. The number of links varied significantly also for the incidents that were covered by well-known newspapers. So far evidence shows that homicides vary in popularity, which justifies the initial objective to compare them.

In total 781 web crime links were collected, which served as a look-up table to connect Twitter messages with the homicide cases. In the next step, we searched for Twitter messages that contain those news links and therefore spread the information within the social network community. Twitter messages were collected from the social media reseller 'TOPSY' [43]. TOPSY is a social search analytics company and also a certified Twitter partner. Its commercial application (*Topsy Pro Analytics*) allows users to collect relevant twitter messages with keyword-based queries. In our case the keywords were the 781 web crime links. The data as such were collected on February 23rd 2014 from the TOPSY service. As this service provides access to all tweets back from 2006 up to now, completeness can be assumed for our data set. The results were then exported from TOPSY as a CSV file for further processing. The exported Twitter messages dataset was cleaned in terms of removing tweets from professional organizations, newspapers, and commercial sources. This formed our *HomicideTwitter-Dataset*. The [S1 Data-set](#) includes the 98 crime incidents (Homicide-Dataset) and the 781 web crime links (Murder links) that were used to extract the HomicideTwitter-Dataset. The *HomicideTwitter-Dataset* contains 216 links from the look-up table, 70 homicides incidents from the *Homicide-Dataset*, and in total 3,272 tweets. The tweets were sent by 1,939 users posting on an average 1.7 tweets (S.D = 2.4) from the 2<sup>nd</sup> of January 2012 until the 4<sup>th</sup> of February 2014.

## Temporal Analysis

In this section we examined the public's temporal reaction on Twitter regarding homicide incidents. The analysis aims to detect and discuss temporal peaks. Furthermore, the results may indicate a time frame for which Twitter can be mined in future research to assist homicide crime investigations. Firstly, we analyzed the length of the time period in days and months between the actual homicide incidents and the public's reaction to them on Twitter (Figs. 1 and 2).

As it can be seen from [Fig. 1](#), 70% of all tweets were posted within one month of the occurrence of the incidents. In addition, the figure reveals a small peak occurring after five months (151–180 days) with about 12% of tweets. This reflects the reaction of people to additional breaking news about the homicides such as confessions and convictions. In our data set, a mother pleaded guilty of killing her children. A similar case occurred in the smaller peak after seven months (211–240 days), when a brother confessed the murder of his sister. Changing the temporal scale, [Fig. 2](#) zooms in on the daily distribution of the tweets that were posted within one month (30 days) of the homicide occurrences. It can be seen that about 81% of the tweets

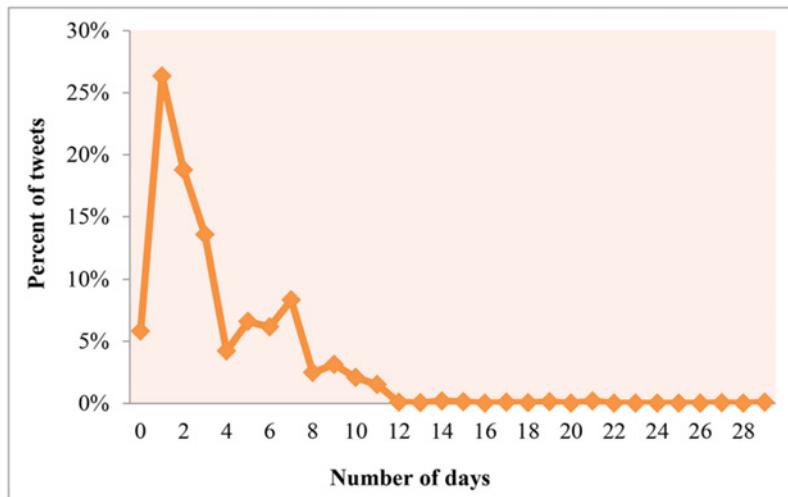


**Fig 1. Distribution of Twitter responses to homicide incidents in months.** 0 shows the percent of tweets that were sent within 30 days of the homicide's occurrence, 1 shows the percent of tweets that were sent within 31–60 days, etc.

doi:10.1371/journal.pone.0121848.g001

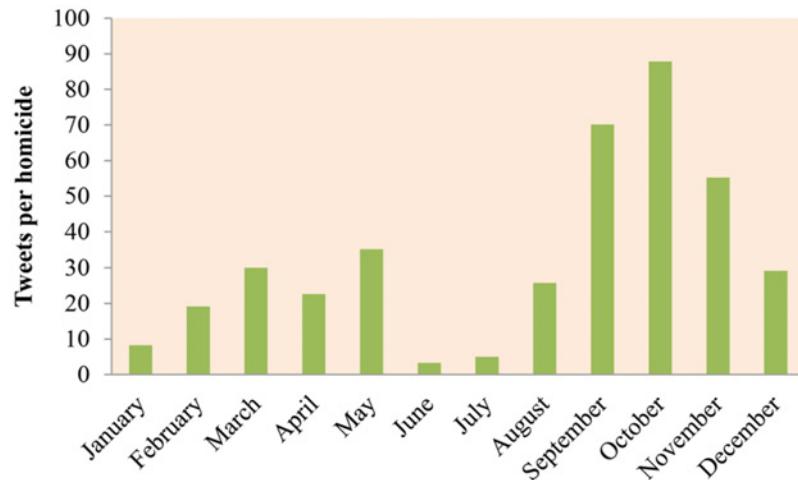
within the first month were sent within the first week of the homicide occurrence. The highest peak appears on the first day after the incident with about 26%. Altogether, this indicates that Twitter users disseminate information about homicides close to the date when the actual incident occurred.

In the next step we calculated the percent of tweets that occurred in the same month as the homicides that they refer to (e.g. homicide occurred in June and the tweet that refers to the homicide was posted in June). This was 65% of the tweets, which implies that the homicides are being discussed on Twitter mainly at the month of their occurrence. Then, the tweets per homicide ratios were calculated over a period of twelve months in 2012 (Fig. 3). These ratios are



**Fig 2. Distribution of Twitter response to homicide incidents in days.** The distribution refers to tweets that were sent within the first month of the incident. 0 depicts the percent of tweets that were sent the same day of the homicide's occurrence, 1 indicates the percent of tweets that were sent the next day, etc.

doi:10.1371/journal.pone.0121848.g002



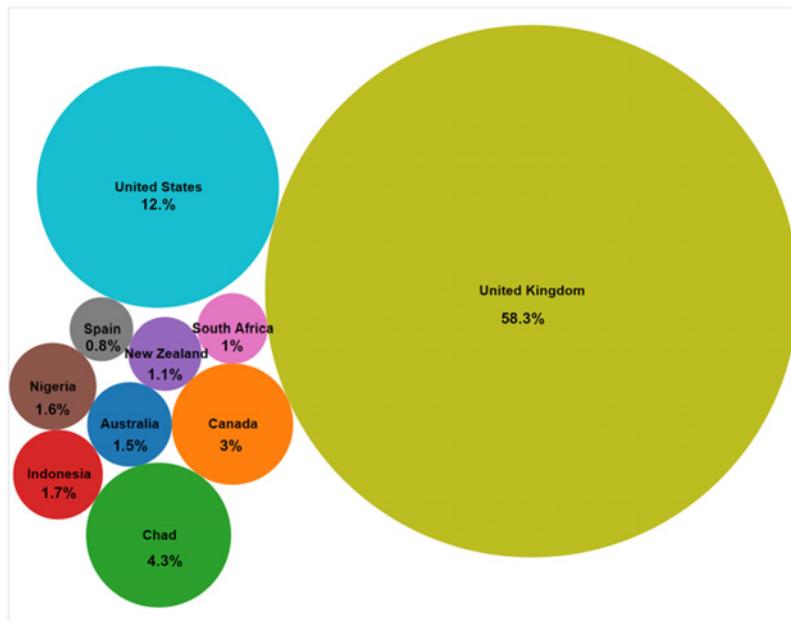
**Fig 3. Homicide tweets ratio by month in 2012.**

doi:10.1371/journal.pone.0121848.g003

calculated as the number of homicide-tweets per month divided by the number of homicides that occurred in the same month. Equal interest or concern about homicides would yield similar ratios per month. However, in contrast, a great variation in the ratios across the twelve months is observed. In June we have the lowest ratio with 3.3 tweets per homicide (46 tweets and 14 homicides) while in October the ratio is 87.9 tweets per homicide (703 tweets and 8 homicides). Also, during the autumn a higher response is observed compared to other seasons. This is a first indication that certain homicides receive more attention than others.

## Spatial Analysis

This section presents an analysis of the relationship between the users' locations and the locations of the homicides. For a general overview of the worldwide response to local crime news, we relied on the location information extracted from the users' profiles in Twitter. After cleaning the data to remove empty and irrelevant entries (e.g. jokes) from the location field on the users' profiles, 1,088 tweets were used for the global analysis (ca. 33% of the original *HomicideTwitter-Dataset*). The locations referred to different geographical scales, from a national to a neighborhood level. In order to achieve an equal granularity we extracted the name of the countries using the free Web mapping service "Nominatim" [44]. In total, homicides that occurred in London in 2012 were commented by messages coming from 74 countries. Nevertheless, only 10 countries accounted for more than 85% of all tweets (Fig. 4). The results clearly show that even though London is one of the world's largest cities and a metropolitan area, the influence of the information about the crimes strongly depends on the locality, the spoken language, and the particular attributes of the crime incident. As it was anticipated, the majority of tweets were sent from the UK. Among other countries with a high frequency of crime tweets are countries where English is the first official language, or one of the official languages (USA, South Africa, and Nigeria). Furthermore, after comparing information in the *Homicide-Data*-set with some of the countries with high frequency (e.g. Australia, Canada, and New Zealand), we observed that these countries are related to the nationality of the victim or the perpetrator of the homicides. To a minor extent, observed frequencies may be affected by the unequal adoption of Twitter among different nations, with the generally higher popularity of this medium in wealthier countries, especially in the UK and the USA [45], [29].

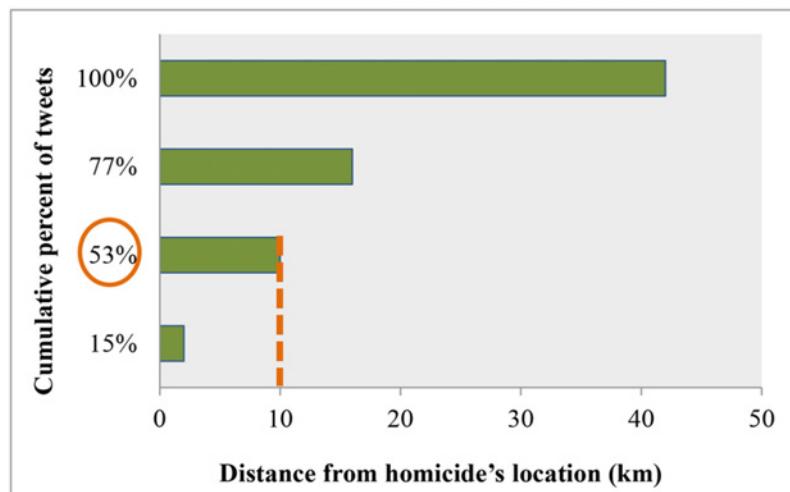


**Fig 4. Share of tweets that refer to London's homicide news among different countries.**

doi:10.1371/journal.pone.0121848.g004

Furthermore, a local spatial analysis was performed to examine whether people living closer to locations of crime incidents exhibit higher tendency to discuss them on Twitter. In order to estimate home locations of users who created messages gathered from the *HomicideTwitter-Dataset* we examined their geo-located posts from an additional database, the *geo-located Twitter database*. The *geo-located Twitter database* is a subset of the Twitter database that consists of tweets with location information at the level of geographic coordinates. It includes all geo-located messages from the UK in 2012 (in total 68,579,232), collected directly from the Twitter Streaming API [46]. This snapshot from the entire Twitter database was cleaned from errors and artificial tweeting noise, similarly as described by Hawelka et al. [29]. Since the analysis was restricted to London, all geo-located tweets outside the city were excluded. Also, in order to obtain a realistic subset of the users' potential home locations we excluded daytime hours (07:00 to 17:59), since some users may frequently post from their work locations. Finally, we excluded all users who had less than ten geo-located messages. In total, the locations of 13,830 geo-located tweets formed the *night activity spaces* of 94 users who posted about 124 tweets related to the 70 homicides. The number of geo-located tweets per user ranges from 10 to 886.

This set of 13,830 locations was further analyzed to estimate the home location of each user from the *HomicideTwitter-Dataset*. The estimation method consisted of three steps and was repeatedly applied to each user. Firstly, a spatial clustering was performed using the "Nearest Neighbor Hierarchical Clustering" method [47]. The following parameters were used with this method: one standard deviational ellipses to outline the clusters, a minimum of 10 points per cluster, only first-order clusters, and a search radius based on the random nearest neighbor distance. Secondly, locations contained in the highest frequency cluster were selected. Finally, the center of minimum distance of these locations was calculated. This is the point that possesses the smallest accumulated distance to all other points in the cluster and from here on it is called "central feature". In few cases where no significant clusters were found, the central feature was directly calculated. We assume that central features are likely to be the user's home locations, since tweets are sent during evening times. This is an approximation based on reasoning of



**Fig 5. The cumulative percent of crime tweets at different distance intervals.** Distance is defined as the distance from users' estimated home locations to the crime incidents' locations.

doi:10.1371/journal.pone.0121848.g005

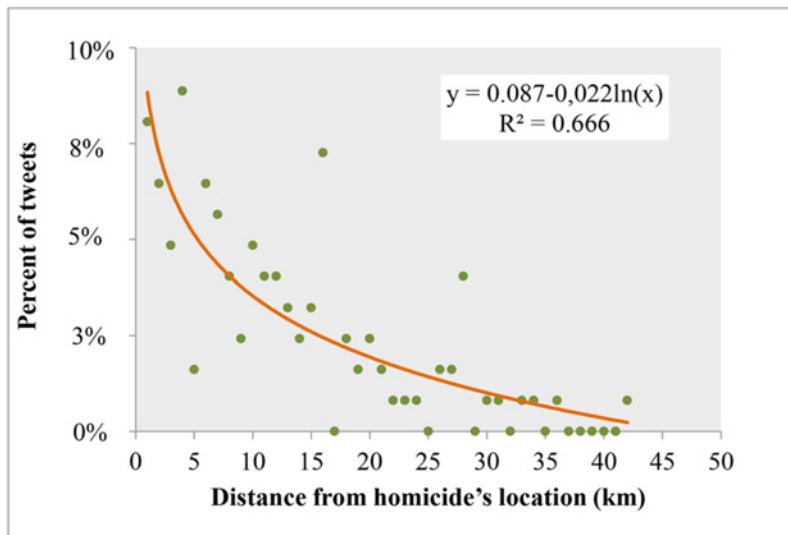
common human habits and may not be the case for all users. If some of these places are not the “true” home locations, they are still the most frequently visited places and locations that are of high interest to users.

Then, the Euclidean distances between the users' estimated home locations and the locations of crimes were calculated and further explored. The results of the distance analysis between the users' locations and the homicides locations confirm that the proximity to the crime incident affects the user's decision on posting about it. More specifically, authors of 53% of crime-tweets live within a 10 km buffer zone around the crime's location, while authors of the remaining tweets live within a buffer zone of greater than 10 km to 42 km (Fig. 5). The cumulative distribution shows that tweets are more associated with shorter distances to homicide locations, rather than with longer ones. Furthermore, we plotted the frequency of the tweets against the distance from users' estimated home locations to the homicide locations (Fig. 6). The graph clearly shows a decreasing trend, which also indicates spatial dependency. In other words, the closer a user resides to a crime location the more likely it is that he or she will spread homicide news via Twitter. Lastly, a distance decay model was fitted for this type of relationship at a distance aggregation of 1 km (Fig. 6). Among several functions, the one that best fits the data is the logarithmic function ( $R^2 = 0.666$ ,  $F = 79.685$ ,  $b = -0.22$ , constant = 0.087, at  $p < 0.001$ ).

## Frequency Analysis

The results of the temporal analysis showed that the monthly distribution of crime tweets does not follow the monthly distribution of crimes. This indicates that even though all crimes are of the same type, namely homicides, they do not have the same impact on Twitter users. A high volume of tweets for a particular crime could imply high popularity of this topic, great interest, and consequently higher concern about this incident. To analyze the distribution of tweets and understand what makes some incidents receive more attention than others we use an ordinal logistic regression (OLR) analysis.

The dependent variable of the analysis is the frequency of the tweets of each crime incident. Ordinal logistic regression requires the dependent variable to be of ordinal type. However, it



**Fig 6. The distance decay fitted model between the frequency of the tweets and the distance from crime incidents.** Distance is defined as the distance from users' estimated home locations to the crime incidents' locations.

doi:10.1371/journal.pone.0121848.g006

can be used when a categorical variable has been derived from a measured continuous variable [48]. According to Agresti [49] logistic regression is the most important model for categorical response data and it has been used in a variety of applications such as biomedical studies, as well as social science research and marketing. To use ordinal logistic regression we categorized the number of tweets per crime incident using the “natural breaks” classification, where the lowest values after peaks in a distribution are used to define classes’ limits. The “natural breaks” model was proposed by Jenks & Caspall [50] and follows a similar concept to the approach of Fisher [51], which minimizes the variance within the classes to create homogeneous groups. We selected a number of classes that balances detailed information with sufficient observations per class. This resulted in the following five frequency classes: homicides with a) 0–8 tweets, b) 9–26 tweets, c) 27–55 tweets, d) 56–114 tweets, and e) 115–420 tweets.

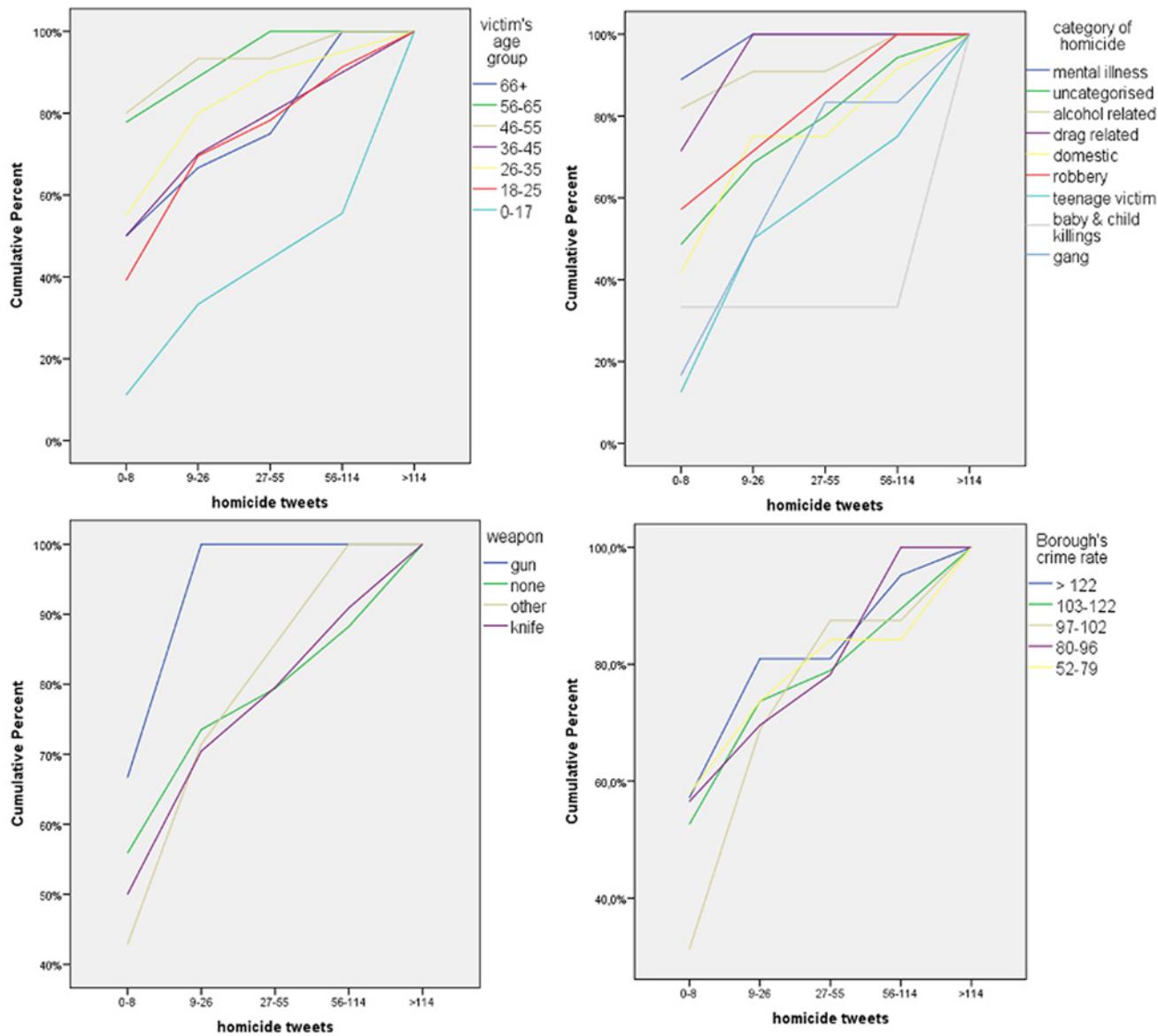
As for the independent variables, we used information about crimes that was obtained from the source website [42]. The information was organized in nine variables: 1) victim's gender, 2) victim's age, 3) victim's nationality, 4) perpetrator's gender, 5) perpetrator's nationality, 6) number of arrested perpetrators, 7) weapon, 8) category of homicide, and 9) status of conviction. In particular, the first three variables describe demographic characteristics of victims, the next two describe demographic characteristics of perpetrators, and the last four describe specific characteristics of the crime itself. The weapon variable defines the type of the weapon that was used (e.g. none or knife), the “category of homicide” defines the type of the homicide (e.g. drug related or domestic), and the last variable “status of homicide” clarifies whether the crime incident is a solved or an unsolved case. According to “[www.murdermap.co.uk](http://www.murdermap.co.uk)” the cases are updated regularly to show the progress of the investigation, (solved or unsolved case), however, the exact time interval of the updates is not stated. In addition we consider one more variable, i.e. the number of Borough's offences per 1,000 of the population. The rates were retrieved from the Metropolitan Police Service [52] to examine whether the overall crime density of the Borough that includes the location of the incident may have an impact on the frequency of the tweets. In other words do crimes occurring in areas of high crime rates receive respectively higher public attention or vice versa?

The first step in developing the model was to observe the cumulative percentages of the tweets' frequency by the categories of each variable. This was particularly useful for the non-binary categorical variables (more than two categories) since visual observations outlined which categories may act as predictors in the model and therefore should be selected as the categories which the others will be compared to. [Fig. 7](#) illustrates four variables of the model (due to space constraints only four of the ten variables are illustrated). A general observation is that when a line is located more to the right side of the graph compared to the other lines then the respective category is associated with a higher tweets frequency. Hence, the victim's age group "0–17" was selected to be compared with the remaining categories, since on the graph it greatly differs from the other age groups. For the category of homicide the "baby & child killings" and "teenage victims" categories were not selected because they correspond to a large degree with the victim's age group "0–17" category. Therefore, the "gang" category (the next one with high tweets' frequency) was selected. Regarding the weapon variable, both "knife" and "none" have similar trends, which are higher than the other two categories. The "knife" category was finally selected since it has more homicide tweets than the "none" category, in total. The last variable of [Fig. 7](#), the Borough's crime rate, does not have a clear trend in terms of its categories. In this case the highest class (>122) was selected to be compared with the others based on the assumption stated in the previous paragraph.

The model was developed using a stepwise procedure similar to a backward elimination of the variables with the largest p-value [53]. After a sequential removal of four of the ten variables (perpetrator's gender, status of conviction, number of perpetrators, and victim's gender) we obtained the optimal model. The model fit was significant and the Goodness-of-fit tests were insignificant (this means that the null hypothesis is true and the model is a good fit). Furthermore, the Pseudo R-square decreased insignificantly with the removal of each of the four variables (Negelkerke of original model with all ten variables: 0.738, Negelkerke of final model with six variables: 0.714). Finally, 71.4% of the variance of the model's outcome is explained by the remaining six explanatory variables.

From the six variables of the final model the perpetrator's nationality and the crime rate of the Borough were insignificant variables. The remaining four variables yielded significant results and are presented in [Table 1](#). More specifically, for an additional victim being in any of the age groups: 18–25, 26–35, 46–55, 56–65, and 66+, controlling for the other variables in the equation, the odds of that homicide being in a higher tweets' class are lower by a minimum of 99.7% compared to the age group of 0–17 years old. Similarly, for an additional homicide in the category of "teenage victim", "drug related", or "mental illness", controlling for the other variables in the equation, the odds of that homicide being in a higher tweets' class are lower by a minimum of 99.67% compared to the "gang" category. Also, for an additional homicide performed by a gun, controlling for the other variables in the equation, the odds of that homicide being in a higher tweets' class are lower by 99.69% compared to the homicide being performed by a knife. Finally, for an additional victim being British, controlling for the other variables in the equation, the odds of that homicide being in a higher tweets' class are more than 100 times higher than for a non-British victim.

As a check of the regression homicide data, the significant variables as predictors of public concern are compared with the frequency of the categories in the *Homicide-Dataset* ([Fig. 8](#)). Nationality of a victim and weapon seem to be in line with their frequencies. For instance 84% of the victims were British and therefore it makes sense to expect more tweets regarding British victims. In the same manner, the weapon most commonly used in homicides was the knife, and respectively the likelihood of referring to such a crime is higher than posting about a homicide performed by a gun. On the other hand, the age and the crime's category variables are not in accordance with their frequencies implying that certain topics receive high public concern



**Fig 7. Cumulative percentages of the tweets' frequency by categories of each variable.**

doi:10.1371/journal.pone.0121848.g007

even though they are not as frequent as others. In particular, homicides where the victim's age is above 18, apart from the age group of 36–45, are less likely to be mentioned on Twitter, even though they are as frequent, or in most cases more frequent than the homicides where the victim was underage. Also, a “gang” homicide seems to trigger more concern than categories of higher occurrence such as “teenage victim”, “drug related”, and “mental illness”.

## Discussion

This study utilized Twitter data to explore the spatio-temporal dimensions of homicide incidents' impact on the public, as well as the influence of the characteristics. To perform the analyses the data-processing methods 'Links Correspondence Method' and 'Home Estimation Method' by Lampoltshammer et al. [41] were used. The validation of the proposed approach

**Table 1. Parameter estimates of the OLR model (only significant variables are shown).**

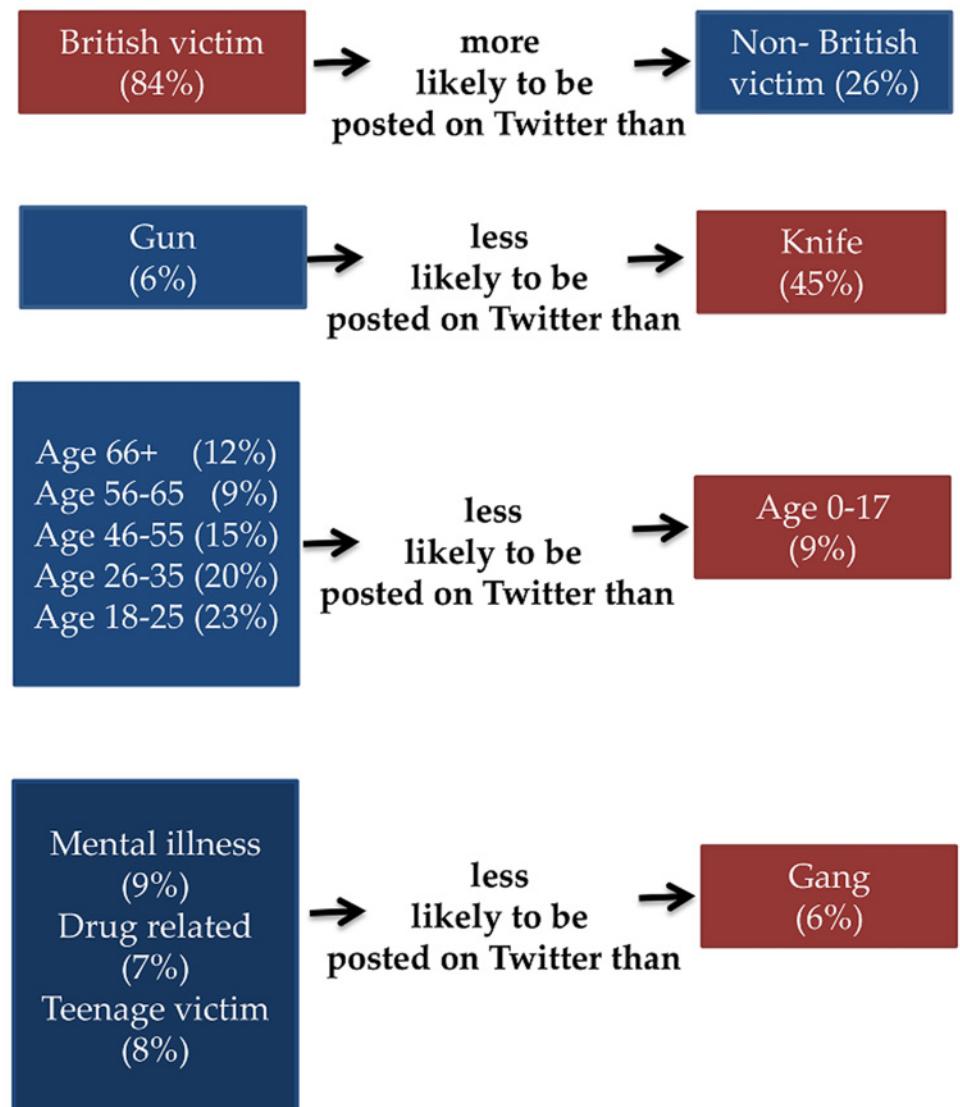
Explanatory variables	Estimate	Std. Error	Wald	Sig.	Exp(B)	
<b>Age group</b>	66+	-14.538	4.235	11.781	.001	<b>0.000</b>
	56–65	-6.960	2.997	5.392	.020	<b>0.001</b>
	46–55	-9.324	2.827	10.874	.001	<b>0.000</b>
	36–45	-4.526	2.723	2.762	.097	
	26–35	-8.498	2.542	11.175	.001	<b>0.000</b>
	18–25	-5.822	2.099	7.694	.006	<b>0.003</b>
<b>0–17</b>						
<b>Nationality</b>	victim is British	4.735	1.675	7.991	.005	<b>113.878</b>
<b>victim is not British</b>						
<b>Category</b>	mental illness	-5.723	2.367	5.847	.016	<b>0.003</b>
	not categorised	-2.624	1.508	3.027	.082	
	alcohol related	-25.077	7.291.646	.000	.997	
	drug related	-6.802	2.426	7.861	.005	<b>0.001</b>
	domestic	.098	1.703	.003	.954	
	robbery	-2.464	1.938	1.617	.204	
	teenage victim	-5.784	2.229	6.730	.009	<b>0.003</b>
	baby & child killings	12.675	0.000			
<b>gang</b>						
<b>Weapon</b>	gun	-5.784	2.219	6.793	.009	<b>0.003</b>
	none	.471	.870	.293	.588	
	other	2.717	1.798	2.284	.131	
	<b>knife</b>					

doi:10.1371/journal.pone.0121848.t001

has been analyzed separately and can be found in the [S1 Text](#). In short, the 'Home Estimation Method' was validated and proved to be an appropriate technique to estimate home locations of Twitter users. The aspects of crime concerns (i.e. temporal, spatial, and crime characteristics) were validated as well and the findings show that they are the main themes of discussion in homicide related tweets.

With regards to the temporal response of homicide news on Twitter, a clear temporal decay was observed with more than 2/3 of all tweets related to a particular homicide posted within one month of the incident. 81% of those were posted within a week and 26% on the day after the homicide. Furthermore, smaller tweeting peaks exist after five and seven months a crime has happened. This is likely explained by the release of new information about a case. Finally, additional analysis shows that temporal variations in crime-tweets frequencies are not related to the number of homicides in the respective months.

The spatial analysis revealed a strong spatial dependency between the estimated home locations of users who tweeted about homicide related news and the locations of these incidents. More precisely, the distance decay model of distances between these paired locations suggests that the number of messages regarding a particular homicide increases as the users' locations become closer to the location of the homicide. This finding seems to be in line with findings of previous studies which confirmed that local crime news are associated with a higher level of fear of crime for local residents compared to non-local crime news [\[12\]](#), [\[13\]](#). If people have a greater fear about homicides happening in their neighborhoods, then they are more likely to discuss them, also via Twitter posts, rather than addressing other crime news. It is also consistent with Tobler's [\[54\]](#) first law of geography which states that *near things are more related*



**Fig 8. Posting likelihood on Twitter by category of the explanatory variable.** Likelihood is expressed as odds, namely the likelihood of an event occurring to the likelihood of not occurring. Blue rectangles indicate categories of lower posting likelihood compared to the categories in red rectangles. The percentages in parenthesis show the frequency of this category in the *Homicide-Dataset*.

doi:10.1371/journal.pone.0121848.g008

than distant things. People may be more affected by crime stories when they concern their familiar environment.

Analysis of crime characteristics indicates that some of them are associated with a higher frequency of tweets than others. In addition, previous studies used socio-demographic characteristics of people who are exposed to crime news as predictors of fear of crime. Here we used the characteristics of crime incidents as predictors of concern about crime and operationalized that concern via higher levels of posting frequency. The results showed that homicides that have at least one of these characteristics, namely being committed by a gang, involving an underage or a British victim, and using a knife as a weapon, are positively related with a higher frequency of tweets. Furthermore, homicides committed by a gang or underage victims were

categories less numerous than others in the *Homicide-Dataset*. This finding implies that either the public is more sensitive to these types, or that the estimated crime prevalence for these types is higher than the actual crime prevalence. The latter explanation is consistent with findings from the literature regarding the crime rates perception gap [55–58].

On the other hand, other crime characteristics do not have a significant impact on the posting frequency. For instance, the crime rate of the Borough where the incident occurs does not have an impact on people's decision on posting about it. This is consistent with findings from the fear of crime literature which suggest that personal risks do not consistently correspond with official statistics of neighborhoods [6]. However, the relationship that is examined here is rather specific. It shows that the background risk (crime rate of an area) does not affect human reactions to homicides. The results of the model may have been different if human reactions to other or all crime types were examined. For instance, tweets with general crime content may be strongly related with the underlying crime clustering in space. Also, tweets with general crime content may be somehow related with the underlying crime clustering with evidence of spatial variations. This relationship seems to be an unexplored opportunity that can be associated or assist current research efforts on crime prediction. Literature on predictive mapping utilizes past crime incidents and quantitative findings about repeat victimization to predict future incidents [59]. Crime related geo-referenced tweets could be an alternative data source for prediction. The advantage of social media data is that they can be collected and analyzed almost in real time. Thus, they may reveal trends and changes in the spatial distribution of crime related discussions that is important or correlated with future crime trends.

Legitimacy of Big data analytics and prediction has been already confirmed in many research areas, for instance “infodemiology” [60], [61] [23], stock market [62], consumer preferences [63], and global mobility patterns [29]. Through the practice of computational social science, digital traces that we leave behind have the potential to enhance our understanding of our lives and societies or even transform our current knowledge [64]. Our study and in particular the agreement between the received results and existing crime literature grounded in traditional survey approaches, proves that Twitter data may also be successfully applied for crime analysis. The spatial analytical approach that was used in this study and, in general, the examination of the spatial impact of events is perhaps only viable with user-generated georeferenced content. The practice of social sciences is moving into a new scientific era where human sensor data such as social media are not only effective and reliable for investigation, but also necessary to address key research questions.

## Supporting Information

### S1 Dataset. Aggregated anonymized data of the study.

(XLSX)

### S1 Text. Validation of the approach.

(DOCX)

## Author Contributions

Conceived and designed the experiments: OK TJL. Performed the experiments: OK TJL. Analyzed the data: OK TJL. Contributed reagents/materials/analysis tools: IS. Wrote the paper: OK TJL EG IS ML. Data processing and analysis: OK TJL. Coordinated the study and consulted on the geographic techniques to be used: EG. Identified and provided source material: IS. Consulted on the presentation of the results and the geographic techniques to be used: ML. Contributed design/analysis/interpretation/drafting: OK TJL EG IS ML.

## References

1. Ferraro KF, Grange RL (1987) The Measurement of Fear of Crime. *Sociological Inquiry*. 57 (1). 70–97.
2. Warr M, Ellison CG (2000) Rethinking Social Reactions to Crime: Personal and Altruistic Fear in Family Households. *American Journal of Sociology*. 106 (3). 551–578.
3. Jackson J (2006) Introducing Fear of Crime to Risk Research. *Risk Analysis*, 26 (1), 253–264. PMID: [16492196](#)
4. Amerio P, Roccato M (2005) A predictive model for psychological reactions to crime in Italy: an analysis of fear of crime and concern about crime as a social problem. *Journal of Community & Applied Social Psychology*. 15 (1). 17–28.
5. Furstenberg FF (1971) Public reaction to crime in the streets. *The American Scholar*. 40 (4). 601–610.
6. Lewis DA, Maxfield MG (1980) Fear in the neighborhoods: An investigation of the impact of crime. *Journal of Research in Crime and Delinquency*. 17(2). 160–189.
7. Wang K, Taylor RB (2006) Simulated walks through dangerous alleys: Impacts of features and progress on fear. *Journal of environmental psychology*. 26(4). 269–283.
8. McCord ES, Ratcliffe JH, Garcia RM, Taylor RB (2007) Nonresidential crime attractors and generators elevate perceived neighborhood crime and incivilities. *Journal of Research in Crime and Delinquency*. 44(3). 295–320.
9. Williams C (2008). Mapping the fear of crime—a micro-approach in Merton, London. *Crime Mapping Case Studies: Practice and Research*, 103.
10. Heverin T, Zach L (2010) Microblogging for Crisis Communication: Examination of Twitter Use in Response to a 2009 Violent Crisis in the Seattle-Tacoma. Washington. Area: ISCRAM.
11. Warr M (1995) Public opinion on crime and punishment. *Public Opinion Quarterly*. 59(2). 296–310.
12. Heath L (1984) Impact of newspaper crime reports on fear of crime: Multimethodological investigation. *Journal of Personality and social Psychology*. 47(2). 263.
13. Liska AE, Baccaglini W (1990) Feeling safe by comparison: crime in the newspaper. *Soc. Probs.* 37. 360.
14. Williams P, Dickinson J (1993) Fear of Crime: Read All About it? The Relationship between Newspaper Crime Reporting and Fear of Crime. *British Journal of Criminology*. 33 (1) 33–56.
15. Groff ER, Kearley B, Fogg H, Beatty P, Couture H, Wartell J (2005). A randomized experimental study of sharing crime data with citizens: Do maps produce more fear? *Journal of Experimental Criminology*. 1 (1). 87–115.
16. Wakshlag J, Vial V, Tamborini R (1983) Selecting crime drama and apprehension about crime. *Human Communication Research*. 10(2). 227–242.
17. Gebotys RJ, Roberts JV, DasGupta B (1988) News media use and public perceptions of crime seriousness. *Canadian J. Criminology*. 30. 3.
18. Smolej M, Kivivuori J (2006) The relation between crime news and fear of violence. *Journal of Scandinavian Studies in Criminology and Crime Prevention*. 7 (2). 211–227.
19. Chiricos T, Eschholz S, Gertz M (1997) Crime news and fear of crime: Toward an identification of audience effects. *Soc. Probs.* 44. 342.
20. Schlesinger P, Tumber H, Murdock G (1991) The media politics of crime and criminal justice. *British Journal of Sociology*. 397–420.
21. Duffy B, Rowden L (2005) You are what you read? London: MORI.
22. Paul MJ, Dredze M (2012) A model for mining public health topics from Twitter. *Health*. 11. 16–16. doi: [10.1002/xd.20024](#) PMID: [23097359](#)
23. Culotta A (2010) Detecting influenza outbreaks by analyzing Twitter messages. *arXiv preprint arXiv:1007.4748*.
24. Ghosh D, Guha R (2013) What are we 'tweeting' about obesity? Mapping tweets with topic modeling and Geographic Information System. *Cartography and Geographic Information Science*. 40(2). 90–102. PMID: [25126022](#)
25. Kent JD, Capello HT Jr. (2013) Spatial patterns and demographic indicators of effective social media content during the Horse thief Canyon fire of 2012. *Cartography and Geographic Information Science*. 40(2). 78–89. PMID: [25126022](#)
26. Neubig G, Matsubayashi Y, Hagiwara M, Murakami K (2011) Safety Information Mining—What can NLP do in a disaster. Paper presented at the IJCNLP.

27. De Longueville B, Smith RS, Luraschi G (2009) OMG. from here. I can see the flames!: a use case of mining location based social networks to acquire spatio-temporal data on forest fires. Paper presented at the Proceedings of the 2009 International Workshop on Location Based Social Networks.
28. Sagl G, Resch B, Hawelka B, Beinat E (2012) From social sensor data to collective human behaviour patterns: Analysing and visualising spatio-temporal dynamics in urban environments. Paper presented at the Proceedings of the GI-Forum 2012: Geovisualization. Society and Learning.
29. Hawelka B, Sitko I, Beinat E, Sobolevsky S, Kazakopoulos P, Ratti C (2014) Geo-located Twitter as the proxy for global mobility patterns. *Cartography and Geographic Information Science*, 41(3), 260–271.
30. Cataldi M, Di Caro L, Schifanella C (2010) Emerging topic detection on twitter based on temporal and social terms evaluation. Paper presented at the Proceedings of the Tenth International Workshop on Multimedia Data Mining.
31. Zhao WX, Jiang J, He J, Song Y, Achananuparp P, Lim EP, et al. (2011) Topical keyphrase extraction from twitter. Paper presented at the Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1.
32. Crampton JW, Graham M, Poorthuis A, Shelton T, Stephens M, Wilson MW, et al. (2013) Beyond the geotag: situating 'big data' and leveraging the potential of the geoweb. *Cartography and Geographic Information Science*. 40(2). 130–139. PMID: [25126022](#)
33. Stefanidis A, Cotnoir A, Croitoru A, Crooks A, Rice M, Radzikowski J (2013) Demarcating new boundaries: mapping virtual polycentric communities through social media content. *Cartography and Geographic Information Science*. 40(2). 116–129.
34. Hofer B, Lampoltshammer TJ, Belgiu M (2015) Demography of Twitter Users in the City of London—An Exploratory Spatial Data Analysis Approach. In Vondrakova A., Brus J. & Vozenliek V. (Eds.), *Modern Trends in Cartography*. Springer International Publishing. pp. 199–211.
35. Cheng Z, Caverlee J, Lee K, Sui DZ (2011) Exploring Millions of Footprints in Location Sharing Services. *ICWSM*, 2011, 81–88.
36. White JJ, Roth RE (2010) TwitterHitter: Geovisual analytics for harvesting insight from volunteered geographic information. Paper presented at the Proceedings of GIScience.
37. Sun B, Ng VT (2011). Lifespan and popularity measurement of online content on social networks. In *Intelligence and Security Informatics (ISI)*, 2011 IEEE International Conference on (pp. 379–383). IEEE.
38. Wang X, Gerber MS, Brown DE (2012) Automatic crime prediction using events extracted from twitter posts *Social Computing. Behavioral-Cultural Modeling and Prediction* (pp. 231–238): Springer.
39. Said H, Yousif A, Humaid H (2011). iPhone forensics techniques and crime investigation. Paper presented at the Current Trends in Information Technology (CTIT). 2011 International Conference and Workshop on.
40. Helbich M, Hagenauer J, Leitner M, Edwards R (2013) Exploration of unstructured narrative crime reports: an unsupervised neural network and point pattern analysis approach. *Cartography and Geographic Information Science*, 40(4), 326–336.
41. Lampoltshammer TJ, Kounadi O, Sitko I, Hawelka B (2014) Sensing the public's reaction to crime news using the 'Links Correspondence Method'. *Applied Geography*, 52, 57–66. Available: doi: [10.1016/j.apgeog.2014.04.016](#)
42. murdermap. Investigate. Available: <http://www.murdermap.co.uk/investigate.asp>. Accessed 2014 Feb 1.
43. Topsy. Search and Analyze the Social Web. Available: <http://topsy.com/>. Accessed 1 April 2014.
44. Nominatim. Reverse Geocoding. Available: <http://wiki.openstreetmap.org/wiki/Nominatim>. Accessed 2014 Mar 1.
45. Mocanu D, Baronchelli A, Perra N, Gonçalves B, Zhang Q, Vespignani A (2013) The Twitter of Babel: Mapping World Languages through Microblogging Platforms. *PLoS ONE* 8(4): e61981. doi: [10.1371/journal.pone.0061981](#) PMID: [23637940](#)
46. Twitter Streaming API. Available: <https://dev.twitter.com/docs/streaming-apis>. Accessed 2015 Jan 2.
47. Everett B (1974) Cluster Analysis. London: Heinemann Educational Books Ltd.
48. Hosmer DW Jr., Lemeshow S (2004) Applied logistic regression (pp. 293–305). New York: John Wiley & Sons.
49. Agresti A (2002) Categorical data analysis ( Vol. 359) (pp.165) New York: John Wiley & Sons.
50. Jenks GF, Caspall FC (1971) Error on choroplethic maps: definition, measurement, reduction. *Annals of the Association of American Geographers*, 61(2), 217–244.
51. Fisher WD (1958) On grouping for maximum homogeneity. *Journal of the American statistical Association*, 53(284), 789–798.

52. Metropolitan Police Service. Crime Mapping: Data tables. Available: <http://maps.met.police.uk/tables.htm>. Accessed 2014 Feb 1.
53. Goodman LA (1971) The analysis of multidimensional contingency tables: Stepwise procedures and direct estimation methods for building models for multiple classifications. *Technometrics*, 13(1), 33–61.
54. Tobler WR (1970) A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 234–240.
55. Duffy B, Wake R, Burrows T, Bremner P (2008) Closing the gaps—crime and public perceptions. *International Review of Law Computers & Technology* 22(1–2): 17–44. PMID: [17411780](https://pubmed.ncbi.nlm.nih.gov/17411780/)
56. Kemshall H (1997) Sleep Safely: Crime Risks May Be Smaller than you Think. *Social Policy & Administration* 31 (3):247–259.
57. Pfeiffer C, Windzio M, Kleimann M (2005) Media use and its impacts on crime perception. sentencing attitudes and crime policy. *Eur J Criminol* 2: 259–285.
58. Weatherburn D, Matka E, Lind B (1996) Crime Perception and Reality. *Crime and Justice Bulletin* 28.
59. Bowers KJ, Johnson S, Pease K (2004) Prospective hotspotting: The future of crime mapping? *British Journal of Criminology* 44, 641–58.
60. Chew C, Eysenbach G (2010) Pandemics in the age of twitter: Content analysis of tweets during the 2009 h1n1 outbreak. *PLoS ONE* 5(11):e14118. doi: [10.1371/journal.pone.0014118](https://doi.org/10.1371/journal.pone.0014118) PMID: [21124761](https://pubmed.ncbi.nlm.nih.gov/21124761/)
61. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L (2008) Detecting influenza epidemics using search engine query data. *Nature*, 457(7232), 1012–1014.
62. Moat HS, Curme C, Avakian A, Kenett DY, Stanley HE, Preis T (2013) Quantifying Wikipedia usage patterns before stock market moves. *Scientific reports*, 3.
63. Goel S, Hofman JM, Lahaie S, Pennock DM, Watts DJ (2010) Predicting consumer behavior with Web search. *Proceedings of the National academy of sciences*, 107(41), 17486–17490. doi: [10.1073/pnas.1005962107](https://doi.org/10.1073/pnas.1005962107) PMID: [20876140](https://pubmed.ncbi.nlm.nih.gov/20876140/)
64. Lazer D, Pentland A, Adamic L, Aral S, Barabási AL, et al. (2009) Computational social science. *Science* 323: 721–723 doi: [10.1126/science.1167742](https://doi.org/10.1126/science.1167742) PMID: [19197046](https://pubmed.ncbi.nlm.nih.gov/19197046/)