ESSAY COVER SHEET 2018

NIR605: Critical Data Studies

Full Name: **Sean O'Riogain**

Student Number: **18145426**

Date Submitted: **4th April 2019.3311**

Course Title & Lecturer: **Critical Data Studies (NIR605), Dr. Rob Kitchin.**

Essay/Project title: **A Critical Comparison on US Officer-Involved Shooting (OIS) Data Sources**

NIR605 – Critical Data Studies - Assignment 2

A Critical Comparison on US Officer-Involved Shooting (OIS) Data Sources

Study Overview

Context

Since the infamous [Rodney King](#) incident in Los Angeles (1991), during which the eponymous victim was severely beaten by 4 police officers, police forces across the United States have been struggling to come to grips with managing their members' use of force while on duty. While that incident did not involve an Officer-Involved Shooting (OIS), it did give rise to 6 days of rioting the following year, after the trial of the officers in question resulted in their acquittal, during which 63 people were killed and 2,363 were wounded.

However, without salient and reliable data it is not possible to assess what progress, if any, has been made – and where – during the intervening years. Although some police departments, such as [Dallas PD](#), have been collecting and publishing OIS data for several years, there has been no concerted approach to doing so, and to consolidating it, across all US law enforcement agencies. In fact, it was not until January of this year (2019) that the Uniform Crime Reporting ([UCR](#)) Program of the US Federal Bureau of Investigation (FBI) has launched its National Use-of-Force Data Collection ([NUoFDC](#)) initiative to capture data of this type (including OIS data) from the 1,800 plus police forces that historically supply it with crime data. But it will take several years before sufficient data have been collected to enable meaningful analysis to be performed, trends to be identified, and corrective action programs to be launched. In the meantime, various interested parties have been endeavouring to address this dearth of UoF/OIS data.

Objective

The aim of this study is to critically compare (at a high level) a sample of those extant OIS data sources in order to assess the usefulness of the data that they contain in order to determine if any of them could be used immediately to expedite the corrective action programs referred to above.

See the References section below for links to those data sources. Henceforth in this document they will be referred to by the following shorthand labels: [Fatal Encounters](#), [Killed by Police](#), [Police Shootings](#), [Police Violence](#) and [The Counted](#).

Methodology

This assessment is performed under the following headings: motivation, collection, curation, access, documentation and quality.

Findings

As can be seen in the Study Details below, the [Fatal Encounters](#) was the leader under virtual all of the above headings and is therefore. It shows that its data covers the greatest timespan , offers the highest volumes, that it uses a greater variety of data collection methods, is being constantly updated, supports open data sharing, appears to be the best designed, curated (to ensure data quality) and documented, and is used as a source of data by a number of the other data sources.

In contrast, there is a risk of political bias (albeit in opposite directions) being attached to using the [Killed by Police](#) and [Police Violence](#) data sources, while the [The Counted](#) data source has not been updated since the end of 2016. The [Police Shootings](#) data source is not as open as it might be since it requires the payment of a subscription (albeit a small one) to obtain data for the most recent years.

NIR605 – Critical Data Studies - Assignment 2

Study Details

Motivation

Under this heading, this study attempts to assess the stated motivation of the parties responsible for the creation and maintenance of each of the data sources under review in order to identify any bias which would require the data they contain to be used with caution. In other words, the objective here is to establish the overall level of trustworthiness of each data source.

Data Source	Organisation	Org. Type	Political Wing	Motivation	Mission Statement
Fatal Encounters	N/A	Journalist	Centre	"Journalistic Curiosity"	See here.
Killed by Police	Unknown	Unclear	Right	Unclear	None found.
Police Shootings	Washington Post	Newspaper	Centre	Public Service	See here and here.
Police Violence	We The Protestors	Black Activist	Left	Protest	See here and here and here.
The Counted	The Guardian	Newspaper	Centre	Public Service	See here and here.

Having researched within and around the data source websites, whose links are in the first column above, it appears that 4 out of 5 of have a public service-type motivation (of which 3 are politically centrist and 1 is adjudged to be most likely leftist). The exception is the Killed by Police data source which may be associated with an unspecified right-wing organisation judging by its feature on telescopic sights for various types of assault rifles (see here). If there is political bias in any of the other data sources, it is most likely to underlie the Police Violence data source.

Based on the contents of the table above, including that of its linked material, the following data sources appear to have clearest and most credible mission statements (in order of transparency): Fatal Encounters. The Counted, Police Violence.

Collection

How a data source's content is collected – both originally and on an ongoing basis – can provide another reliable indicator of its potential veracity and, therefore, utility. For instance, whether or not the source is a primary one, or whether or not it uses open data collection methods such as crowdsourcing, are important considerations in this regard.

Data Source	Type	Method	Notes
Fatal Encounters	Primary / Secondary	Multiple	"*Fatal Encounters is a complex and rigorous project that uses several processes of data collection to ensure a high level of validity. Media news sources have predominantly focused on the crowdsourcing aspects of our project. While some of our data is crowdsourced, we have three main methods of collecting information. They are listed below in order of numbers of records in the database: 1) Paid researchers; 2) Public records requests; 3) Crowdsourced data.*" (See here.)
Killed by Police	Unknown	Unclear	No information available.
Police Shootings	Secondary	Multiple	"*This database is based on news reports, public records, Internet databases and original reporting.*" (See here.)

NIR605 – Critical Data Studies - Assignment 2

Data Source	Type	Method	Notes
Police Violence	Secondary	Consolidation	*"This information has been meticulously sourced from the three largest, most comprehensive and impartial crowdsourced databases on police killings in the country: FatalEncounters.org, the U.S. Police Shootings Database and KilledbyPolice.net."* (See here.)
The Counted	Secondary	Multiple	*"So far, we count with traditional reporting on police reports and witness statements, by monitoring regional news outlets, research groups and open-source reporting projects such as the websites Fatal Encounters and Killed by Police. But our intention is to progress to a verified crowdsourced system."* (See here.)

Based on the contents of the table above, including its linked material, the following data sources appear to be making the best efforts to optimise data collection, including acquiring new data (in order of merit): Fatal Encounters. The Counted, Police Violence.

Curation

Here, this paper searches for evidence that the content of each data source is curated in a manner that is likely to ensure that the veracity of their contents will be achieved and maintained. As a minimum a well-curated data source should offer its contributors facilities to review existing content (to avoid duplicating it), to create that content, to update it (to correct errors) and to delete it (to address data privacy issues) using well-managed processes.

Data Source	Create	Review	Update	Delete	Notes
Fatal Encounters	See here and here	See here	See here	No	No privacy policy was found on this database's website whereby, for instance a victim's family could request the deletion of his/her incident details.
Killed by Police	No	No	No	See here and here	*"…we do not maintain any procedures for you to review or request changes to the information that we collect about you, except that you may request that we remove all information about you from our database by contacting us…"* (see here).
Police Shootings	No	No	No	No	No indication of these facilities was found on this database's website.
Police Violence	No	No	No	No	
The Counted	See here.	See here	See here	No	No privacy policy was found on this database's website.

The information, and that linked to, in the table above show that Fatal Encounters appears to have put the most thought and effort into the curation of its content, followed by The Counted (as a distant second).

Access

Any data source will be of limited benefit if it is not made available, both openly and easily, to the parties that are likely to make productive use of it – at local, regional and national levels. In the context of this study, those constituencies include law enforcement management, their political oversight authorities, interested non-governmental organisations, and civil society in general.

NIR605 – Critical Data Studies - Assignment 2

Data Source	Downloadable	Searchable	Notes
Fatal Encounters	See here	See here	This website provides a good online search facility and the entire database is available for download in Google Docs (although the download mechanism is not intuitive).
Killed by Police	No	No	At first glance, the data appears to be searchable but, in practice, only the data for the last month of each year is searchable.
Police Shootings	See here	See here	The entire database is available for download on github. The search facility is very basic, and by year, using filters and the most recent years require payment of a subscription.
Police Violence	See here	No	The entire database can be downloaded in MS Excel format but there is no online search capability offered on the website, which concentrates solely on visualising it.
The Counted	See here	No	The entire database can be downloaded in CSV format but the website does not offer an online browse or search facility for that data.

The contents of the table (and its linked material) show that the closest one to being classed as an open data source is Fatal Encounters followed by Police Shootings as a distant second and Killed by Police in last position. While the other two data sources support download of their content, surprisingly, they do not enable online browsing or searching of same. The requirement of a subscription payment (albeit a small one) tarnishes the open data source credentials of the Police Shootings website.

Documentation

The quality of the documentation of any deliverable can be a good indicator of the quality of the deliverable itself. In the case of a data source, this documentation should cover the overall design of the database (including the metadata for its content) and tools to visualise its content.

Data Source	Design	Visualisation	Notes
Fatal Encounters	See here	See here	Limited visualisation is provided using Tableau.
Killed by Police	No	No	The visualisations provided here were culled from the Police Violence website.
Police Shootings	No	See here	Some good visualisation provided but no design material.
Police Violence	No	See here	Strong visualisations here (as there is a data scientist on the Planning Team).
The Counted	No	See here	Limited visualisation is provided in the homepage heading and via a map button there.

Fatal Encounters is the only data source that makes any attempt to document its design approach, including its selection of data elements and also provides some visualisations of its contents. Of the other data sources, Police Violence makes the best use of visualisation of its underlying data.

Quality

In this section of the document, the data quality aspects of in-scope data sources is reviewed by repurposing the 8 elements of Kitchin's big data taxonomy (2013).

NIR605 – Critical Data Studies - Assignment 2

Volume

The quantity of data provided by a data source – both in terms of the number of observations (rows) and variables (columns) can provide an early, if not definitive, indication of the relative utility of a number of competing data sources.

Data Source	Rows	Columns	Cells	Notes
Fatal Encounters	25,777	26	670,176	Could not open file when downloaded in MS Excel format but CSV format was fine.
Killed by Police	0	0	0	This data source does not support data download.
Police Shootings	4,054	13	52,702	Downloaded as a single CSV file.
Police Violence	6,838	24	164,112	Downloaded as a single MS Excel file.
The Counted	2,239	14	31,346	Downloaded as 2 CSV files (for the years 2015 & 2016, respectively).

The table above shows that the Fatal Encounters data source can provide the highest data volume, followed by Police Violence as a distant second. The remaining 2 data sources that support data download each provide less than 10% of the data volume (in terms of the number of cells) of the leading data source.

Velocity

The frequency of data update can provide another indicator as to level of the timeliness of a data source's content.

Data Source	Latest Incident Date	Age (in Days)	Notes
Fatal Encounters	25/03/2019	6	Determined from the downloaded data.
Killed by Police	28/03/2019	3	As recorded on the website (download not supported).
Police Shootings	14/02/2019	46	Determined from the downloaded data.
Police Violence	31/12/2018	90	
The Counted	31/12/2016	821	

The table above shows that the Fatal Encounters and Killed by Police data sources are achieving the highest update velocity (with data less than 7 days old) at the time of writing, followed by Police Shootings and Police Violence (with data 1-3 months old). In this context, The Counted is the outlier (with data more than 2 years old).

Variety

The availability of structured, semi-structured and unstructured data in a particular data source could be another important selection consideration. In this context, for instance, photographic images of the victims could be of particular benefit.

Data Source	Notes
Fatal Encounters	Supports links to photos, news articles and video footage.
Killed by Police	Supports links to news articles on its website. (Data download is not supported.)
Police Shootings	Supports links to links to photos and news articles.
Police Violence	Provides no support in this area.
The Counted	

NIR605 – Critical Data Studies - Assignment 2

A quick review of the datasets of the 4 data sources that support download reveals that only 3 of them provide links to unstructured data – namely, Fatal Encounters and Police Shootings. In the case of the former, where available, it supports links to a photo of the victim, a news article or photo of an official document, and video footage of the incident (although this is populated in only 2 instances). The latter data source can provide a link to an image (albeit invariably on the former data source's website) and to a news article or photo of an official document. Killed by Police also provides links to news articles on its website. This review reveals that Fatal Encounters is the leader in this category.

Exhaustivity

The level of ambition that a particular data source aspires to – and achieves – will also determine its potential suitability to a particular use. For instance, a data source that claims to include all OIS incidents for the past 10 years could be of more use to one that maintains them for most recent 5-year period. On the other hand, a data source which seeks to - or, by default, does – maintain data for only a sample subset of incidents for its target period may be of less interest.

Data Source	From (Year)	To (Year)	Notes
Fatal Encounters	2000	2019	Determined from the downloaded data.
Killed by Police	2013	2019	As recorded on the website (download not supported).
Police Shootings	2015	2019	Determined from the downloaded data.
Police Violence	2013	2018	
The Counted	2015	2016	

While all 5 data sources claim to strive to record the details of all OIS incidents in a particular year, the contents of the table above show that the number of years covered can vary greatly, with the Fatal Encounters data source offering by far the greatest coverage.

Resolution

Depending on the objectives of the data user, a data source which maintains data at individual incident level could be preferable to one that maintains it at some aggregate level – e.g. at precinct, city, county, state, zip code or national level.

Data Source	Level	Notes
Fatal Encounters		Determined from the downloaded data.
Killed by Police		As recorded on the website (download not supported).
Police Shootings	Incident	
Police Violence		Determined from the downloaded data.
The Counted		

As all 5 data sources maintain their data at incident level, there is no clear leader in this area.

Indexicality

The presence of a unique identifier (primary key index) in a particular data set can indicate that consideration has been given to the identification and elimination of duplicate records. In the case of OIS data, the case number of the originating police force would not be sufficient as this could not be guaranteed to be unique following consolidation at the national level.

NIR605 – Critical Data Studies - Assignment 2

Data Source	Unique Id	Notes
Fatal Encounters	Unique ID	Determined from the downloaded data.
Killed by Police	-	Not displayed on the website, if present (download not supported).
Police Shootings	id	Determined from the downloaded data.
Police Violence	-	
The Counted	uid	

As shown by the table above, 3 of the 5 data sources support a unique identifier for each incident. In the case of the Police Violence data source, while no unique identifier is apparent in its dataset, in many cases it does provide a value in its WaPo ID field, which seems to suggest that it is using the Washington Post's Police Shootings data source as a data provider. While no unique id is displayed on the Killed by Police website, we cannot say that one is not present in its underlying database.

Relationality

The utility of a data source is often proportionate to its end-user's ability to link its contents with those of other data sets. In the case of OIS data, for example, it could be useful to be able to link the incident details to the original case files of the relevant law enforcement agency – and the use of some (federal) standard identifier for those organisations could also be of benefit.

Data Source	Foreign Key Field Name	Notes
Fatal Encounters	URL of image of deceased; Location (State Code); Location (ZIP Code); Latitude; Longitude; Link to news article or photo of official document (URL); Video.	Determined from the downloaded data.
Killed by Police	State (code); Source (URLs).	As recorded on the website (download not supported).
Police Shootings	state (code).	Determined from the downloaded data.
Police Violence	State (code); Zipcode; Link to news article or photo of official document (URL); URL of image of victim (URL); WaPo ID.	
The Counted	state (code).	

While none of the 5 datasets have a column containing a link that uniquely identifies the relevant law enforcement agency's case file (most probably because that piece of information is not being divulged), they do contain some useful foreign key values which can be used to link the incident details to photos, images and video footage. All of them also provide location codes to at least state level, with some of them going down to ZIP code level which would be useful to tools that are capable of performing map-based visualisation. The Police Violence data source appears to be making use of URL-based coding to link to material on the Fatal Encounters and Police Shootings websites. Only Fatal Encounters dataset has been geocoded down to latitude-longitude level.

Flexibility

Another hallmark of a well-designed and realised database is the capability to be adapted easily and quickly to incorporate additional information as end-user requirements evolve. This capability is often dependent on how/where the data source is hosted. For instance, a database that is hosted on the cloud is very likely to be more flexible in this way than a spreadsheet that is maintained on somebody's laptop and distributed via a mailing list on a regular basis.

Data Source	Extensible	Scalable	Notes
Fatal Encounters	Yes	Yes	
Killed by Police			
Police Shootings			
Police Violence			
The Counted	-	-	Not applicable. (This dataset is no longer being maintained.)

Because all of the datasets listed in the table above are hosted on the worldwide web, it is reasonable to assume that additional data elements could be supported relatively easily (extensibility). Because the volume of OIS incidents in the US appears to be averaging about 1,200 per annum, it is also reasonable to assume that their respective databases would be sufficiently scalable to accommodate many years of future data even if that annual rate increased substantially. However, if/when a particular data source's data volumes reach MS Excel's technical limit, its use of this mechanism for sharing data would need to be reviewed.

NIR605 – Critical Data Studies - Assignment 2

References

Data Sources

Fatal Encounters: http://www.fatalencounters.org/ (accessed 2 April 2019);

Killed by Police: http://killedbypolice.net/ (accessed 2 April 2019);

Washington Post Police Shootings: https://www.washingtonpost.com/graphics/national/police-shootings/ (accessed 2 April 2019);

Mapping Police Violence: http://mappingpoliceviolence.org/ (accessed 2 April 2019);

The Counted: http://www.theguardian.com/us-news/ng-interactive/2015/jun/01/the-counted-police-killings-us-database (accessed 2 April 2019).

Others

Dallas Police Department (2018) *On-Duty Officer Involved Shootings Data* [online]. Available at: http://dallaspolice.net/ois/ois (accessed 2 April 2019).

Federal Bureau of Investigation (2019) *National Use-of-Force Data Collection* [online]. Available at: https://www.fbi.gov/services/cjis/ucr/use-of-force (accessed 2 April 2019).

Federal Bureau of Investigation (2019) *Uniform Crime Reporting (UCR) Program* [online]. Available at: https://www.fbi.gov/services/cjis/ucr (accessed 2 April 2019).

Kitchin R (2013) Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography* 3(3): 262–267.

Wikipedia (2019) *Rodney King* [online]. Available at: https://en.wikipedia.org/wiki/Rodney_King (accessed 2 April 2019).