Using GIS to Understand the Relationship of Community Factors and Police Shootings in the United States: A First Look

Full paper

Hasan Arslan, PhD
Pace University
harslan@pace.edu

Daniel Farkas, PhD
Pace University
dfarkas@pace.edu

Abstract

While many municipalities and counties have good collections of crime data, the lack of availability of national data limits the potential research on gun violence related to officer-involved shootings. Using an open source data collection data model, a data repository of police shootings is being created (SHOT - Statistics Help Officer Training). With incident location information, this paper takes a “first look” at understanding community factors and their relationship to police “use of force”. The paper will briefly describe how GIS has been used to study crime, the problem related to data and data collection for police shootings and the SHOT method of data collection. Preliminary results of a GIS analysis of incidents, income, diversity and education will be described.

Keywords

GIS, Geographic Information Systems, Spatial Analysis, Police, Criminal Justice, Crime

Introduction

Interest in police shooting incidents have increasingly become the focus of the media and criminal justice research. Unfortunately, there is no national database that coordinates the collection of incident information with variables on suspects and officers that can be used to better understand these incidents and develop strategies to reduce gun violence by law enforcement.

A number of municipalities and news outlets do provide data on shootings within their jurisdiction, but these are local only, may be incomplete, or have limited attributes on each incident. One of the best of these is from the Las Vegas Review-Journal in which data on police shootings has been collected from 1990 (LVRG, 2015). In this type of repository, while the data may be complete for the jurisdiction and has simple mapping of incidents, it is not amenable to analysis for trends or insight beyond the municipality. Another approach is a “crowd-sourced” database in which the public contributes incidents to the database. An example of this type of collection is the *Deadspin Police-Shooting Database* (Wagner, 2014). In this case, a data entry capability has been established for the public to fill out a form that includes information about the suspect, officer and the incident (e.g. shots fired, weapon, etc.). Another crowd-sourced database is *Fatal Encounters* (Kyle, 2015), which also provides a data entry and query capability. In both of these cases while the data is national, one can’t rely on its completeness, validity or reliability.

SHOT is a multi-faceted project to develop a national database of police shooting incidents with robust attributes on the suspect, officer and incident. The data, which will be useful to government agencies, police departments and researchers, can be statistically, spatially and temporally analyzed to better understand the dynamics of these types of incidents.

One aspect of collecting and analyzing national data for statistical significance is that its open source methodology can never be complete and for statistical analysis, the collection method can’t insure that the collected incidents are random. However, the goal of the project is to inform policy. If the data shows a trend or points to specific attributes as contributing ones, further analysis would be indicated. In the case where data is complete (e.g. within a jurisdiction), GIS models can be applied with confidence. Since the data collection is ongoing – as incidents are collected (both current and historic), the reliability of the data will continue to improve.

GIS and Crime

As GIS development systems became more widespread as a result of more powerful and less expensive hardware and software, there are been many reports of GIS application to policing and crime. Most of these studies map occurrences in order to understand the location dynamics and in some cases predict future crime. This gives law enforcement agencies the knowledge to better deploy their resources.

Lebeau (2000) published a report describing the use of GIS for police operations. The report analyzed the changing distribution of service calls based on street size, intensity of crime and race. A second report investigated locations where officers were injured. Herchenrader (2015) indicated that GIS or mapping improves efficiency and can by historical analysis determine where crime may occur in the future. Caplan (2011) developed risk terrain maps to show risk of future shootings and then tested and found statistically significant the predictive capability of the risk maps to forecast future incidents. Hiropoulos (2014) used GIS to understand property crime patterns in South Africa with the objective of understanding the spatial variations and geographic contexts of incidents. In another interesting application, Thomas (2009) used GIS to track crime and campus safety at a mid-sized state university. The results of these types of studies can, for example, highlight were more lighting is needed, identify hotspots and help law enforcement develop crime abatement strategies. Looking at high auto theft areas of Tampa, Florida, based on clustering compared to the traditional police grids based on city streets, Aleksa (2006) investigated what makes these areas vulnerable (e.g. cul-de-sacs, types of businesses, etc.). Kumar (2002) mapped crime locations with proximity to alcohol consumption locations with the intent to help law enforcement agencies better allocate resources. The results correlated with an exponential decline in crime as the distance from the establishments increased.

In most of the literature and in the field, the applications of GIS for crime analysis and prevention are increasing. The SHOT project uses GIS models specifically for police shooting incidents with the objective of understanding the dynamics of these encounters and with goals of prevention, police training and research.

Research Questions

As incidents of police shootings become more prominent in national conscience, it has become necessary to better understand the motivation for these incidents, provide insight into to preventing them and give policy makers tools to reduce the unnecessary use of police force. Because there is no national database of police shooting incidents in the United States, research at the national level is sparse while the possibilities for study are multidimensional. As a first look, this study investigated the relationship between community variables and the number and proximity of incidents at the county level.

1. Do incidents cluster in specific geographic areas?
2. If so, is there a relationship between the number of police shooting incidents and income, education, or diversity levels of these communities?

To address these questions, 10 community variables described later in this paper were compared to the incidence of shootings to see if there is a correlation.

Data Collection

The authors have collected specific data about the characteristics of incidents (location, time of day, etc.), officers and subjects (age, race, mental state, etc.). Additionally, there is "community" data that describes, at various geographic granularities (e.g., state, county, zip, census block), attributes relating to income, diversity, education and more. This data is available through the US Census Bureau and ESRI, Inc. (ESRI 2016a, 2016b)

The police shooting data is collected from open source news sources and has 4 components. The data collection protocol involves the source or where the incident data originates (all entries in the database are sourced) and contains the name of the source, the author, the date of the article and when possible the full-text. More than one source may contribute to the incident record. The incident contains a complete set of attributes/variables including date, time of day, weapon, day of week, bullets fired, police agency involved, whether the target was hit and most importantly, the location. Demographic information is on both the officer(s) and suspect. For officers, along with other variables, we collect race, gender, age and experience. A particular source may not contain, and usually won't, all the information about an incident, and so the data collection protocol allows for adding missing data (even years later) from additional sources. A particular source may also contain information relating to several incidents. Table 1. Summarizes most of the attributes of the collected data upon which GIS and other analysis can be done.

Attribute Category	Attributes
Incident	Date (day, month, season, year), approximate time (early morning, morning, noon, afternoon, evening, night and midnight), number of officers on the scene, number of officers, who fired their guns, number of rounds hit the target, part of target being hit (head/neck, torso, limbs) and geographical information (region, state, city, address, location type).
Suspect	Age, race, gender, fatality, mental status, weapon possession, type of aggression, use of vehicle, occurrence of any foot or car chase, gang affiliation, nationality, fatality/injury and whether a lawsuit was filed.
Officer	Race, gender, experience, affiliation, department type, assignment, status (on/off duty) and type of police call.

Table 1. Database component with associated variables

A prototype enterprise database is under construction that will support data entry and editing, search, administrative operations (e.g. printing), and security that includes user administration and transaction logging.

Utilizing data from ESRI, Inc. Business Analyst Online (ESRI 2016a), demographic and economic data on more than 3000 counties was incorporated into a GIS model. The data is aggregated by ESRI from among other sources, US Census Bureau (ESRI 2016b). These datasets are listed in Table 2.

Data Files from ESRI, Business Analyst Online	Description
2014 Total Crime Index	Crime Indexes data incorporates information from the AGS national Crime Risk database that is based on an extensive analysis of several years of crime incidents reported by most US law enforcement jurisdictions.
2015 Diversity Index	The index shows the likelihood that two persons, chosen at random from the same area, belong to different race or ethnic groups. A Diversity Index of 65 translates to a probability of 65 percent that two people randomly chosen from the US population would belong to different race or ethnic groups.
2015 Median Household Income (Index)	The ratio of a local percent (rate) to a national percent (rate). An index at the national level is 100
2015 Minority Population (%)	% of the population
2015 Per Capita Income (Index)	The ratio of a local percent (rate) to a national percent (rate). An index at the national level is 100
2015 Pop-1 Race_ Black (%)	% of the population
2015 Pop-1 Race_ White (%)	% of the population
2015 Pop 25+ by Educ 9th Grd (%)	% of the population
2015 Pop 25+ by Educ HS Grad (%)	% of the population
2015 Unemployment Rate (Index)	The ratio of a local percent (rate) to a national percent (rate). An index at the national level is 100

Table 2. Data from ESRI, Business Analyst Online

Spatial and Temporal Models

Geographic Information Systems are called *models*. They visualize, analyze and process data features that have location attributes as a part of each record. Features are of different types including "points" (e.g. city or address), "lines" (e.g. roads or rivers) and "polygons" (e.g. state or county boundaries). Maps or spatial visualizations are created when several layers are superimposed upon each other. So for example a road map has street, city and water body layers. Beyond just visualizing data, spatial analytic techniques can be used to investigate properties of the data that are difficult to discern in other ways. This includes proximity, clustering, trending, etc. There are spatial statistical techniques that can be used to test the significance of these relationships (Mitchell 1999, 2005; Pick 2004; Farkas, 2016).

Descriptive statistics can give insight into the number of incidents, gender and race aspects of officers and suspects, and experience of the officer to name a few. Spatial and temporal analysis will allow researchers to look at where incidents occur and track the trends and changes over time. For example, if some departments initiate additional training for the use of firearms or race relations, it will be possible to visualize whether the training was successful in comparison to other departments and locations. It will also be possible to develop a model that shows temporal change.

GIS Methods

The collected data is currently in a spreadsheet with incidents (approximately 2000+) that span more than 10 years. The first step was to geocode, or apply a coordinate system (e.g. latitude/longitude) to the addresses of incidents. Texas A&M University provides a highly accurate (and for researchers, affordable) geocoding service (TAMU 2015).

Once the addresses are geocoded, it is possible to incorporate the incident data into a general geospatial model. There are a number of development environments for GIS development but for this project, ESRI, Inc. ArcGIS Desktop 10 was used. These tools provide the ability to process and create layers to investigate relationships to the incidents (e.g. proximity, community variables).

This "first look" study looked at the location of incidents and their relationship to community variables at the county level such as ethnicity, income and education.

Results

By county, each of the demographic variables listed were correlated to the number of incidents as well as each other. This is shown in Table 3 (positive correlation) and Table 4 (negative correlation). While the correlation was done on all the incidents and counties – the tables below limits the result to those counties with 10 or more incidents. While the correlation values are low, they do stand out for certain variable combinations (highlighted values are > .20 and < -.20). It will interesting to see whether, as the data collection grows, the correlations will become stronger. The value that starts the label is the year of the dataset. "NoHSgrad" was assigned to completed 9th grade data.

	Count	14TotCrime	15DivIndex	15MedIncome	15Minority	15Percapita	15Black	15White	15NoHSgrad	HSGrad	15Unemp
Count	1.000										
14TotCrime	0.097	1.000									
15DivIndex	0.272	0.336	1.000								
15MedIncome	0.010	-0.132	0.197	1.000							
15Minority	0.277	0.332	0.951	0.135	1.000						
15Percapita	-0.107	-0.028	-0.036	0.866	-0.095	1.000					
15Black	-0.072	0.364	0.128	-0.161	0.285	-0.029	1.000				
15White	-0.203	-0.409	-0.844	-0.095	-0.928	0.068	-0.548	1.000			
15NoHSgrad	0.252	0.000	0.730	0.002	0.779	-0.271	-0.042	-0.594	1.000		
HSGrad	0.056	-0.368	-0.368	-0.054	-0.260	-0.024	0.263	0.212	-0.066	1.000	
15Unemp	0.071	0.119	0.500	0.120	0.576	-0.077	0.191	-0.496	0.618	0.267	1.000

Table 3. Correlation table of incidents with positive community demographic and economic factors.

	Count	14TotCrime	15DivIndex	15MedIncome	15Minority	15Percapita	15Black	15White	15NoHSgrad	HSGrad	15Unemp
Count	1.000										
14TotCrime	0.097	1.000									
15DivIndex	0.272	0.336	1.000								
15MedIncome	0.010	-0.132	0.197	1.000							
15Minority	0.277	0.332	0.951	0.135	1.000						
15Percapita	-0.107	-0.028	-0.036	0.866	-0.095	1.000					
15Black	-0.072	0.364	0.128	-0.161	0.285	-0.029	1.000				
15White	-0.203	-0.409	-0.844	-0.095	-0.928	0.068	-0.548	1.000			
15NoHSgrad	0.252	0.000	0.730	0.002	0.779	-0.271	-0.042	-0.594	1.000		
HSGrad	0.056	-0.368	-0.368	-0.054	-0.260	-0.024	0.263	0.212	-0.066	1.000	

Table 4. Correlation table of incidents with negative community demographic and economic factors.

The results highlight several factors:

1. There is a positive correlation of incident count to diversity, minority and no high school graduate.
2. There is a negative correlation of incident count to White communities although effectively no correlation to Black communities.
3. There was no correlation of per capita income to incident count.

The implications are that policy focus (and police training) should be directed to diverse and minority communities. For most of these variables, future research can refine the granularity of the geography (e.g. down to the zip code) targeting specific police jurisdictions as well as refine specific community attributes (e.g. different minority groups).

We can visualize where these communities are located using GIS. Figures 1 is a Density map. Figures 2, 3 and 4 show thematic maps displaying county-level diversity, minority, and high school graduate percentages and indices. While the values are different for each community factor, the locations that police shootings are concentrated are similar. A future line of investigation is why some communities with high demographic/economic indicators do not correlate to high police shooting incident levels while others do.

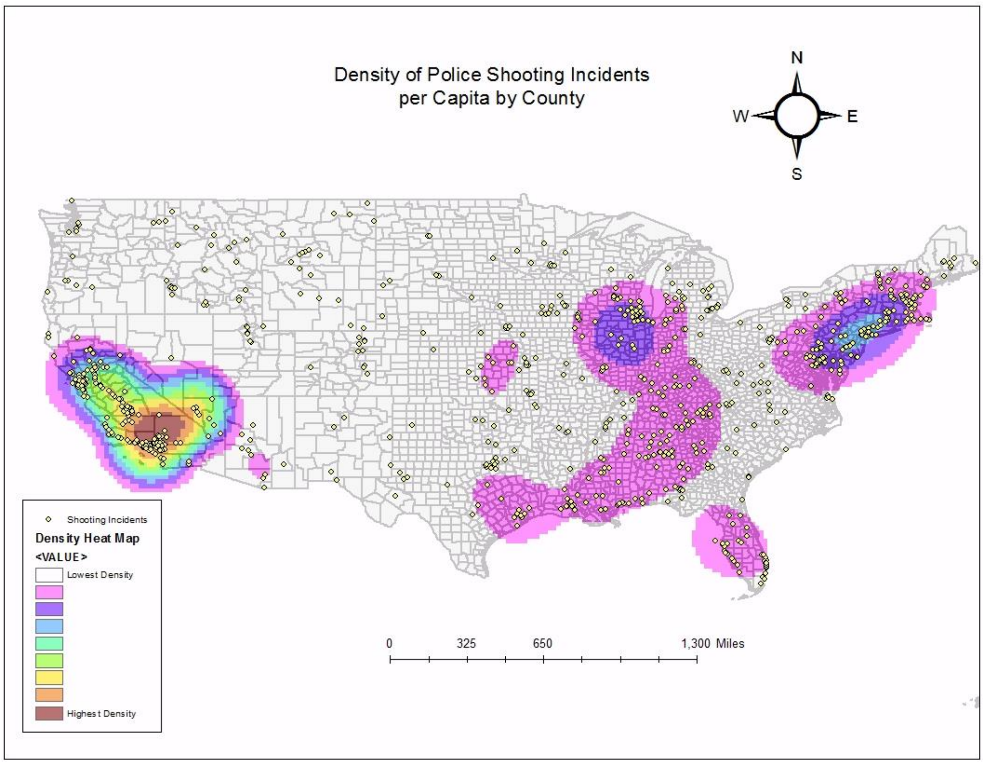

Figure 1. Incident Density Heat Map

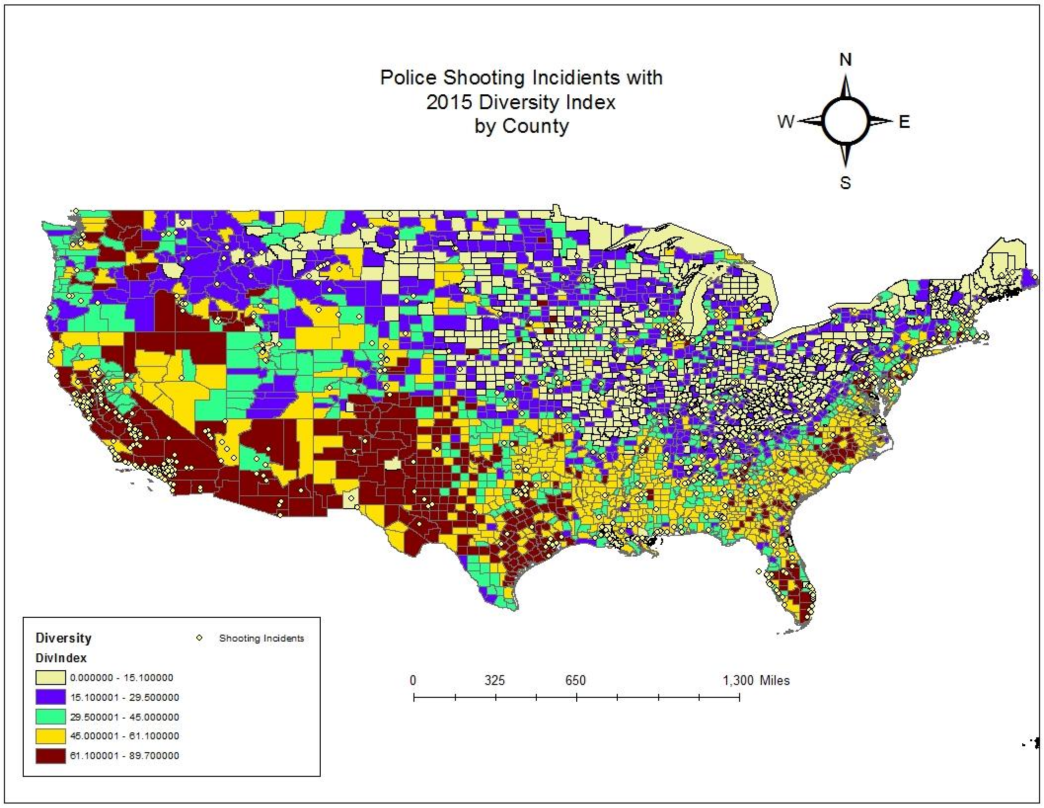

Figure 2. Incidents and Diversity

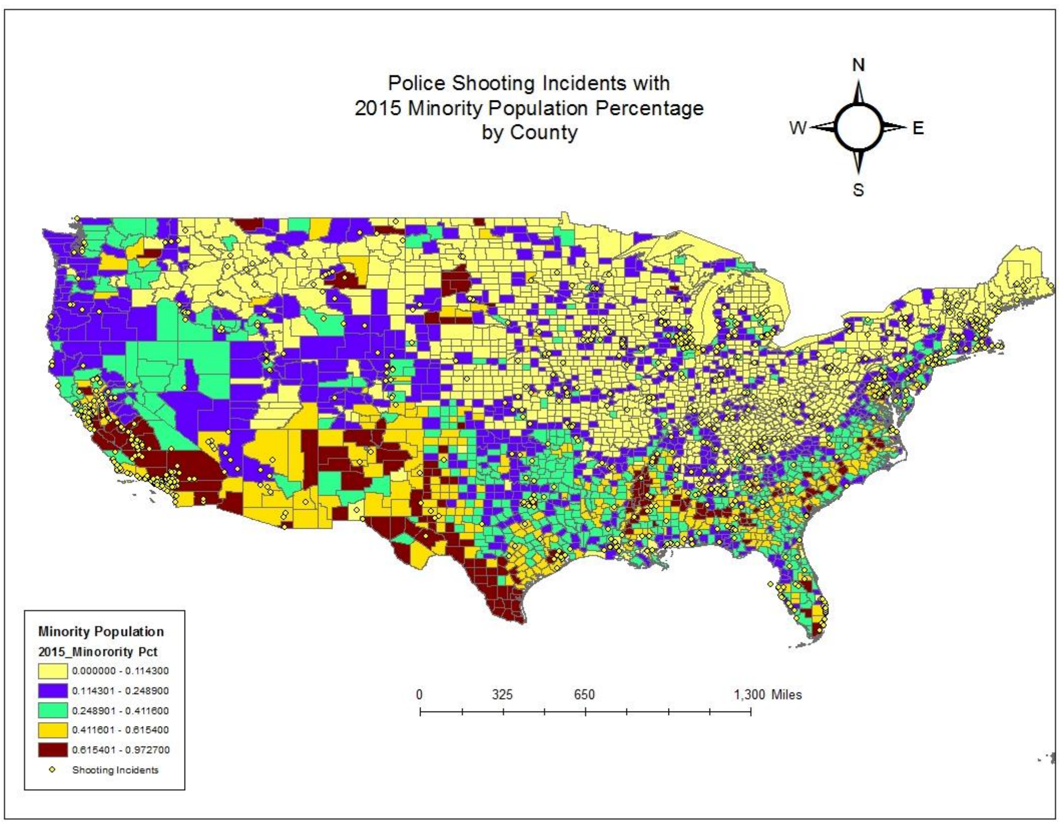

Figure 3. Incidents and minority population

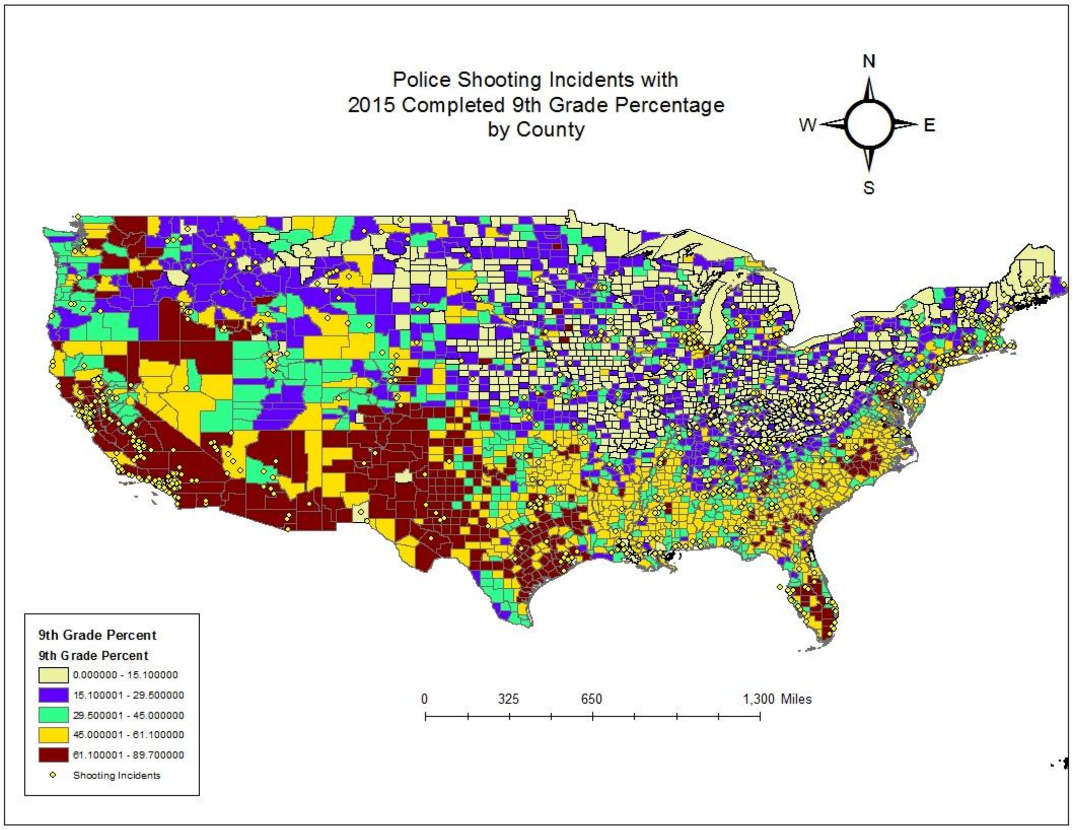

Figure 4. Incidents and high level of non-high school graduates.

Limitations

The main limitation of the project is the nature of the shooting data and its collection. Since it is based on public records and media news feeds, it is incomplete and can't be guaranteed to be a random selection for most statistical analysis. However, as clusters of activity appear, it would indicate to agencies and researchers areas for further study. From this study, for example, it would be interesting to break down the Total Crime variable (e.g., larceny, auto theft, etc.) as well as look further in to "highest education level" achieved. Data exists for college graduation, graduate school, etc.

A related limitation is the amount of data collected – reliability of the data and the trends or patterns it reveals will depend on growing the number of incidents collected.

Future Research

In addition to hundreds of data sets based on community factors, there are more than 40 variables in the shooting collection for which location-based (and non-location based) relationships can be compared with one or more of the others depending on the research question:

- Incidents with officers with different experience levels
- Incidents at night
- Incidents mentally impaired suspects
- Incidents with mentally impaired suspects using a knife
- Incidents with mixed race participants
- Incidents with mixed race participants and officers with different levels of experience

Adding community and other landmark data will give additional insight. For example, even though the data doesn't contain points such as schools, police stations or shopping malls, similar to the study on crime and alcohol consuming establishments, they can be added to the study for investigating incident proximity (Kumar 2002).

Ultimately, SHOT is a multifaceted project that will involve the management of an enterprise web-based database with the capability to extract data for multiple variables and analyze it both statistically and spatially.

References

Aleksa V. 2006, "The GIS Study of the City of Tampa: What Makes Certain City Areas Vulnerable to Auto Theft Crime?" *Conference Papers -- American Society Of Criminology*.

Burghart D.B., 2015. "Fatal Encounters", URL: http://www.fatalencounters.org/, retrieved: 2/2/2015.

Caplan J, Kennedy L, Miller J. 2011. "Risk Terrain Modeling: Brokering Criminological Theory and GIS Methods for Crime Forecasting," *JQ: Justice Quarterly*, (28:2), April pp.360-381.

ESRI, 2016a. Business Analyst Online, URL: http://bao.arcgis.com

ESRI, 2016b. Business Analyst Online Data Descriptions, URL:http://help.arcgis.com/en/businessanalyst/online/newgen/data-descriptions/index.html, retrieved: 3/2/2016

Farkas, D., Hilton, B., Pick, J., Ramakrishna, H., Sarkar, A., Shin, N., "A Tutorial on Geographic Information Systems: A Ten-year Update", *(38)*.

Herchenrader T, Myhill-Jones S. 2015 "GIS supporting intelligence-led policing. *Police Practice & Research* (16:2), April, pp. 136-147.

Hiropoulos A, Porter J. 2014,"Applying GIS to crime pattern theory," *SA Crime Quarterly*, (47), March, pp. 17-28.

Kumar N, Waylor C. 2002, "Proximity to alcohol-serving establishments and crime probabilities in Savannah, Georgia: A statistical and GIS analysis," *Southeastern Geographer*, (42:2) January, pp.125-142.

Lebeau J. 2000, Demonstrating the Analytical Utility of GIS for Police Operations: A Final Report, Southern Illinois University, Carbondale, IL.

LVRJ, 2015. "Deadly Force: When Las Vegas Police Shout, and Kill," *Las Vegas Review-Journal*, URL: http://www.reviewjournal.com/news/deadly-force, retrieved: 2/1/2015.

Mitchell, A. 1999. *The Esri Guide to GIS Analysis. Volume 1: Geographic Patterns and Relationships*, Redlands, CA: Esri Press.

Mitchell, A. 2005. *The Esri Guide to GIS Analysis. Volume 2: Spatial Measurements and Statistics*, Redlands, CA: Esri Press.

Pick, J. B. 2004. "Geographic Information Systems: A Tutorial and Introduction", *Communications of the Association for Information Systems*, (16), pp. 307-33.

QGIS 2015, Quantum GIS, URL: http://www.qgis.org, retrieved: 2/2/2015

TAMU 2015, Texas A&M Geoservices, URL: http://geoservices.tamu.edu/Services/Geocode/, retrieved: 2/2/2015

Thomas M, Patten R. 2009, "Mapping the Campus: Using GIS to Track Crime and Public Safety Concerns at a University" *Conference Papers -- American Society of Criminology* Annual Meeting

Wagner, K., 2014. "Deadspin Police-Shooting Database Update: We're Still Going Strong", *Regressing*, URL: http://regressing.deadspin.com/deadspin-police-shooting-database-update-were-still-go-1627414202, retrieved: 2/1/2015.