Correlation, Causation and Trend Analysis of Crime Data using Police killings Dataset

Venkata Mohan Pallapothu
Software Engineering Dept
Arizona State University
Tempe, Arizona
vkumarpa@asu.edu

Akshay Reddy Kola
Software Engineering Dept
Arizona State University
Tempe, Arizona
akola1@asu.edu

Sri Shashank Mandava
Software Engineering Dept
Arizona State University
Tempe, Arizona
smanda38@asu.edu

Anish Unnikrishnan Nair
Software Engineering Dept
Arizona State University
Tempe, Arizona
anair47@asu.edu

Wasim Ahamed Syed
Software Engineering Dept
Arizona State University
Tempe, Arizona
wsyed4@asu.edu

***Abstract*—The dataset, stemming from the 2014 Michael Brown killing in Ferguson, is a comprehensive record of U.S. police shootings since 2015, including victim demographics and circumstances [1]. Augmented by socio-economic data, it enables user queries to shed light on police fatalities in relation to socio-demographic factors. In this research endeavor, we introduce the concept of a Crime Data Visualization Dashboard, a web application designed to provide a holistic view of crime trends and patterns in different locations. By integrating data from multiple sources, this dashboard offers valuable insights for law enforcement agencies, policymakers, researchers, and the general public. A key aspect of the study is to investigate potential biases in police actions, particularly examining whether the ethnic group most frequently targeted in police killings aligns with the ethnic group predominantly involved in criminal activities. This inquiry is crucial for understanding and addressing issues of racial profiling and systemic prejudice in law enforcement.**

I. Introduction

In the realm of criminal justice and public safety, data-driven insights are pivotal for informed decision-making. The Crime Data Visualization Dashboard is an innovative web application crafted to harness and interpret crime data, providing a lucid and comprehensive picture of criminal activities and law enforcement responses. Racial profiling and discrimination by law enforcement agencies are longstanding concerns in the United States. There is a growing awareness that certain racial and ethnic groups may be disproportionately targeted by the police, leading to tragic consequences in some cases. This project aims to shed light on this issue by examining the intersection of two significant data points: the ethnic group most frequently subjected to police shootings and the ethnic group predominantly responsible for crimes. To achieve this, we harness the power of data, employing two specific datasets: "Uniform Crime Reporting Program Data-2018" and "Police Killings in US."

These datasets collectively form the foundation for our demographic analysis, offering the opportunity to explore whether racial bias or prejudice exists within the context of police shootings. We have expanded our scope to all the 50 States and thus we endeavor to provide a focused yet comprehensive assessment of the issue while demonstrating the potential of the Crime Data Visualization Dashboard as a valuable tool for data-driven decision-making in the realm of law enforcement and public policy. This endeavor aims not only to uncover patterns and correlations but also to advocate for policy changes that promote equity and justice within the criminal justice system.

II. Problem Definition

The primary problem addressed in this study revolves around the presence of racial prejudice in policing, particularly in the context of police killings. The key question to be explored is whether the ethnic group that experiences the highest frequency of police shootings in regions aligns with the ethnic group primarily responsible for crimes in these cities. Should such an alignment exist, it may indicate a concerning pattern of racial profiling by law enforcement agencies, echoing a well-documented issue in the United States.

Through a meticulous examination of the crime datasets and the police killing dataset, combined with a rigorous demographic analysis, our goal is to unveil any possible connections between incidents of police shootings and the ethnic demographics of criminal involvement. If such correlations are identified, it would raise pertinent questions about the presence of a systemic problem rooted in racial prejudice. This research endeavor strives to enrich the ongoing conversation surrounding law enforcement practices, racial biases, and the role of data-driven methodologies in shaping policies and decisions. In doing so, it underscores the indispensable worth of the Crime Data Visualization Dashboard as an instrumental aid in addressing and rectifying these significant societal challenges, ultimately contributing to a more just and equitable society.

III. Related Literature

A. GNIS- LD: Serving and Visualizing the Geographic Names Information System Gazetteer as Linked Data.

This paper describes a method for converting the Geographic names information system gazetteer to linked data. The GNIS gazzetter is a database of geographic names which is maintained by the united states geological survey [2]. This could be used in context of the crime data in the following ways: It can be used to geocode crime data, which will allow the users to visualise the data on a map. This is particularly useful for identifying crime hotspots and trends of police shootings/ brutality. The greatest advantage would be the ability to link crime data to other datasets like census data, in order to find correlation between poverty, unemployment, median house income and crime. Another way it would be useful is the creation of a comprehensive map through time, where we can map crime and census data through the years and notice shifting trends or migration patterns of certain demographics.

B. Police Violence and Citizen Crime Reporting in the Black Community.

Several seminal studies have delved into the multifaceted aspects of police violence. Desmond, Papachristos, and Kirk (2016) investigated the dynamics of police violence within Black communities, emphasizing the intricate relationship between law enforcement and citizens. This study underscores the need for a nuanced understanding of the interactions between marginalized communities and law enforcement agencies, particularly in the context of crime reporting [3]. We will draw inspiration from this paper in the following ways: We can identify the impact of police violence on crime reporting in different communities. We can also draw conclusions from our findings and find a more nuanced understanding of the relationship between police shootings and crime. This would inturn help identify communities that are at risk of experiencing crime and police violence, and to correlate if these are a result of discrimination or other factors.

C. A Bird's Eye View of Civilians Killed by Police in 2015: Further Evidence of Implicit Bias.

Nix, Campbell, Byers, and Alpert (2017) provided further insights by conducting a comprehensive analysis of civilians killed by the police in 2015. Their research uncovered implicit bias in police shootings, raising important questions about the use of force and its potential racial disparities [4]. This work highlights the urgency of addressing implicit bias and its implications in police encounters leading to fatalities. This paper could be used in our web application to include a map that shows locations of all civilians killed by the police, along with filters or a legend to see data by race, ethnicity, gender and other factors. The application could also include a chart that shows the trend of police killings over time, including filters to see crime based on location. Overall this paper is a valuable resourrce for the project, and can be used to inform the development of the web ap to track the issue.

D. A Multi-Level Bayesian Analysis of Racial Bias in Police Shootings at the County-Level in the United States

Ross (2015) took a geospatial approach, offering a multi-level Bayesian analysis of racial bias in police shootings at the county level in the United States [5]. The research highlighted geographic disparities in police violence, emphasizing the importance of considering local contexts when examining these incidents. We will draw inspiration from this paper in the following ways: We can analyze why there is a racial factor included in these shootings and how much of a role this plays during any of these situations and how this can be avoided or atleast diminished in order for these type of instances to not happen again. By doing this might help everyone understand the risk of being on either side of the system.

E. An Empirical Analysis of Racial Differences in Police Use of Force.

Fryer's empirical analysis (2019) of racial differences in police use of force further underscored the disparities in police interactions. This study is particularly valuable for its quantitative approach, providing insights into the extent of racial disparities in the use of force by law enforcement [6]. We will draw inspiration from this paper in the following ways: We can identify the use of additional force and the main reason behind this and how this can be avoided from further escalating. We can also form a relation between the crimes committed and the amount of force used against different groups of people which further makes the case easier to analyze.

F. Perils of police action: a cautionary tale from US data sets

Miller, Lawrence, Carlson, Hendrie, Randall, Rockett, and Spicer (2017) cautioned against the perils of police action, drawing attention to the limitations and challenges in analyzing police violence based on available data sets [7]. This work raises the need for improved data collection and analysis methods in the field. We will draw inspiration from this paper in the following ways: We can analyze this data sets more clearly and maybe also include recent data and analyze and compare if there are any changes among them and provide the updated data which can help in better understanding of the situation in recent years.

G. Summary of related literature:

These essays examine the complex nature of police brutality and how it affects underprivileged groups. By introducing a technique for transforming geographic data into linked data, GNIS-LD allows users to see patterns of police violence and crime hotspots. Research by Nix et al. and Desmond et al., respectively, explore implicit prejudice in police shootings and the dynamics of police brutality within Black communities. Geographic variations in police violence are highlighted by Ross's geospatial research, highlighting the significance of local circumstances. Fryer's empirical research highlights the differences in police use of force between races even more, while Miller et al. warn of the shortcomings of the current data sets and call for better ways to collect data.

IV. Approach And High-Level System Design

The methodology initiates with the acquisition of open data from two pivotal sources: "Uniform Crime Reporting Program Data-2018" and "Police Killings in the USA." During the data preparation phase, our focus is on extracting only the essential data fields that align with our specific objective – assessing the potential presence of racial bias in police responses relative to local crime data. Subsequently, we employ the Stardog-software[11] for ontologies to create structured ontologies, which, in turn, facilitate the formation of data instances. These data instances play a pivotal role in shaping dynamic SPARQL queries, the core of our crime trend analysis process.

From a high-level system design perspective, our approach delineates three primary components: Frontend, Backend, and Ontologies. For the Backend infrastructure, we opt for the python FAST framework to craft the necessary APIs. Ontologies, a critical part of our system, are hosted within a cloud environment. When a user inputs their query, our system dynamically generates and dispatches a SPARQL query to the server for execution. Subsequently, the results are retrieved and seamlessly presented through the user interface, ensuring a user-friendly and efficient experience

V. Ontology design and visualization of ontology

This report discusses the design and visualization of the ontology for our application of a crime data visualization dashboard. The ontology is designed to represent key concepts and relationships related to the crime data. Information about the type of crime and it's relationship with the type of perpetrator and victim with respect to the location is trying to be assessed to infer a bias in society due to race of a person.

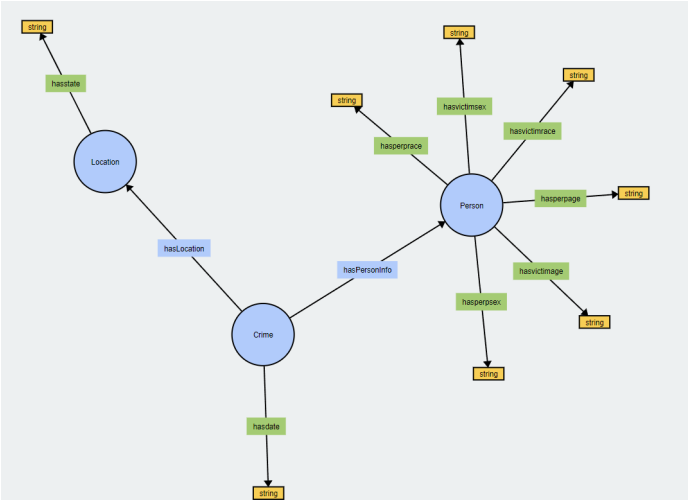
We have the following classes: Crime, Location, Person. We have also used the following data properties: hasperprace, hasvictimsex, hasperpsex, hasdate, hasvictimage, hasperpage, hasvictimrace, hasstate. The following object properties have been defined: hasLocation, hasPersonInfo.

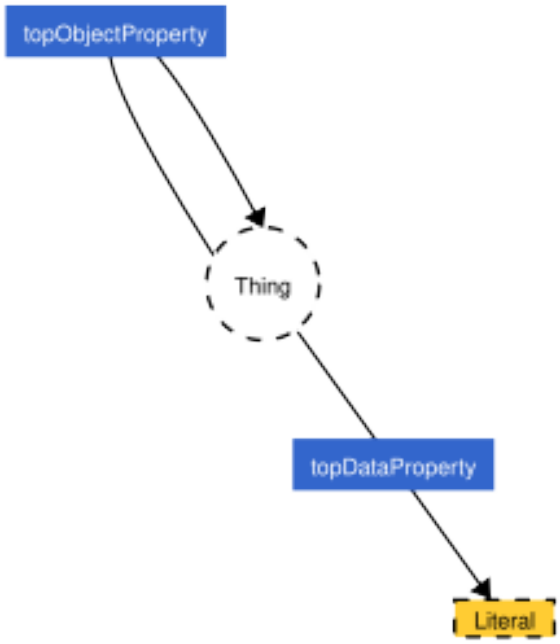
Inorder to convert our data into triplet format, we use Stardog with protege that takes our csv data as input. With the aid of this software, we choose the fields that are necessary for the transformation and it establishes the axioms from triplet data by displaying the dataset's data.

This ontology is designed to be modular and extensible, meaning that new classes, data properties and object properties can be added to the ontology as and when needed. This ontology has also been designed to be consistent with other crime data ontologies, thus can be used to exchange crime data with other systems.

VI. Architecture

The primary objective of our system architecture is to retrieve information according to specified criteria, encompassing state, start date, and end date, subsequently delivering a response aligned with these requirements. Here is a detailed breakdown: Via the user interface (UI), the user chooses the State, start date, and end date. This user data request is then

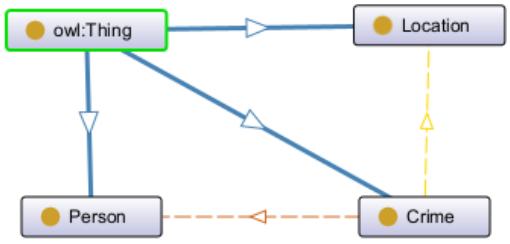




transmitted to the server, signaling a demand for data based on these specific preferences extracted through the React-powered front-end website. Upon receiving the input, the server triggers the relevant SPARQL query function to acquire filtered data from datasets containing information on homicides and police killings. Tailored to the user's specifications, this query collects all pertinent data from the RDF dataset and furnishes the results back to the server. Once the server processes this data, it is displayed to the front end as an HTTP response, allowing the user to view the outcome.

VII. Data Collection And Processing

Police Killings Collection [1], Uniform Crime Reporting Program Data-2018 [8] provide the open-source data sources we use. From these sources, we identify the datasets required



to create the dataset required for our application by cleaning the datasets.

The data required from these datasets primarily comprises details such as the date of the incident, victim information, including race, location of the incident, and the type of crime. Additionally, we also omitted information such as MO codes, weapon used, weapon description, status from the crime dataset, and victim name, manner of death, flee, and body camera data from the police killings dataset, which are not relevant to the application's purpose.

Once the initial data has been collected and ontologies are built using the free open source tool known as protege, data cleaning is the next step. Open Refine, a free and open-source tool, is employed for data cleansing. This tool is effective in handling unclean data, cleaning it, converting it to different formats, and augmenting it with web services and external data sources. During this process, redundant information and null values are removed, and data is adjusted to meet our requirements.

After data refinement, it needs to be adapted to adhere to the ontology specifications. This adjustment can be achieved by exporting the data as an Excel file and making the necessary modifications. The stardog-software [11] can then be used to create instances based on this refined data and use the CSV files to create RDF triplets.

We collect metadata using FAST API with the REST API[13] and data pertaining to these fields are extracted:

- STATE
- OCCDATE
- OFFAGE
- OFFSEX
- OFFRACE
- VICTAGE
- VICTSEX
- VICTRACE

VIII. Semantic Querying

SPARQL query for filtering homicide data based on date, start date and end date:

```
PREFIX ns: <http://example.com/ontology#>
    SELECT ?occ_date ?off_age ?off_sex ?off_
victict_age ?vict_sex ?vict_race ?file_name WHE
        ?homicide ns:OCC_DATE ?occ_date ;
            ns:OFF_AGE ?off_age ;
            ns:OFF_SEX ?off_sex ;
            ns:OFF_RACE ?off_race ;
            ns:VICT_AGE ?vict_age ;
            ns:VICT_SEX ?vict_sex ;
            ns:VICT_RACE ?vict_race .
        FILTER (?occ_date >= "{from_date}"^^xsd:da
te && ?occ_date <= "{to_date}"^^xsd:date && ?homic
ide/ns:state = "{state}"^^xsd:string && ?homicide/
ns:file_name = "hm"^^xsd:string)
```

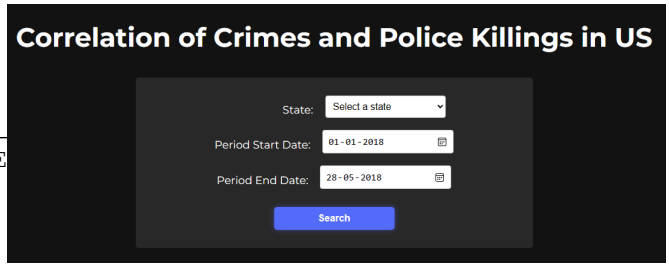
SPARQL query for filtering policekillings data

```
PREFIX ns: <http://example.com/ontology#>
    SELECT ?occ_date ?vict_name ?vict_age ?vict_
sex ?vict_race ?file_name WHERE {{
        ?policekillings ns:OCC_DATE ?occ_date ;
            ns:VICT_NAME ?vict_name ;
            ns:VICT_AGE ?vict_age ;
            ns:VICT_SEX ?vict_sex ;
            ns:VICT_RACE ?vict_race .
        FILTER (?policekillings/ns:state =
"{state}"^^xsd:string && ?policekillings/ns:
file_name = "pk"^^xsd:string)
```

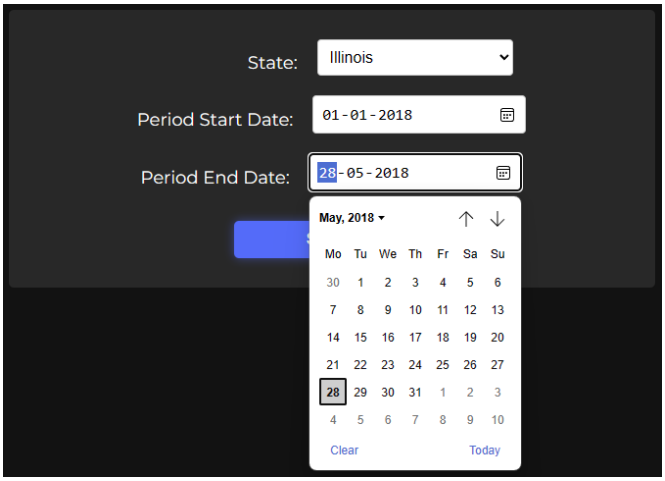
IX. Implementation Process

- Perform extensive data preprocessing, encompassing data validation, cleaning, and feature engineering, adhering to specific requirements and data quality standards.
- Employ the Stardog-software to seamlessly integrate and encapsulate the existing datasets into Protege ontologies, ensuring semantic consistency and harmonization of data.
- Engage in advanced SPARQL query formulation, optimizing query efficiency and precision to interrogate the constructed ontologies effectively.
- Architect a sophisticated user interface with dynamic features, including a hotspot map visualization, catering to user input submissions. Leverage React to create an interactive and visually appealing frontend, offering a seamless user experience.
- Undertake the architectural design of the backend system, strategically employing Django to create robust APIs capable of handling user inputs, processing queries, and retrieving contextually relevant results.
- Seamlessly integrate the frontend and backend components to form a cohesive system, allowing for fluid data flow and efficient user interaction.

X. Application Overview

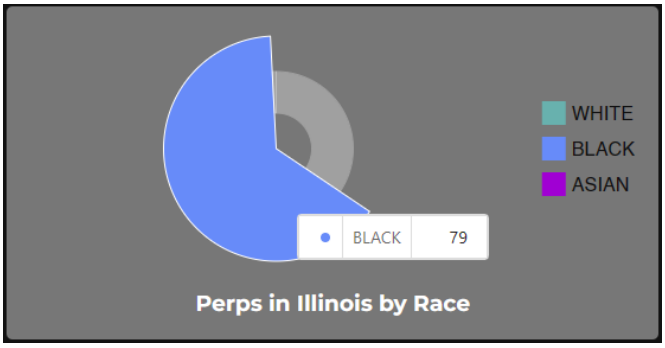


Three input boxes appear with the following inputs: State, period start date, period end date.

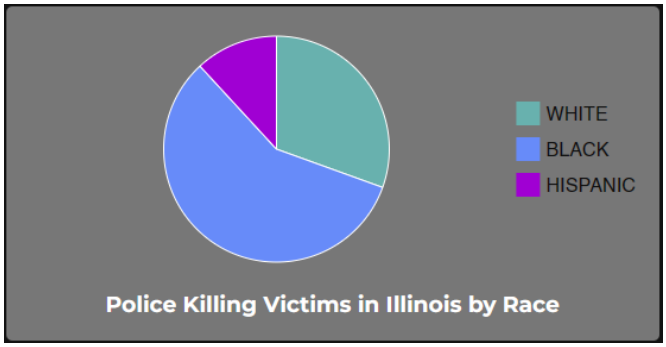


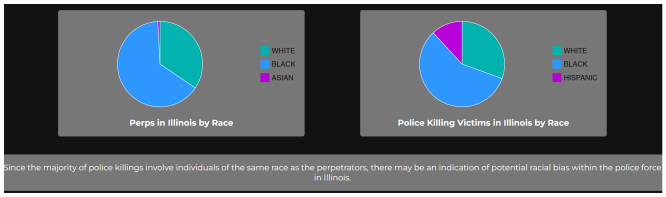
After inserting necessary input conditions for both fields you can visualise data in a tabular format, with information about the perpetrator's age, sex, race and the victim's age, sex and race.

Date	Perp Age	Perp Sex	Perp Race	Victim Age	Victim Sex	Victim Race
31-01-2018	30	MALE	WHITE	BB	FEMALE	WHITE
31-03-2018	39	MALE	BLACK	26	MALE	BLACK
31-03-2018	37	MALE	BLACK	34	MALE	BLACK
31-01-2018	26	MALE	BLACK	21	MALE	WHITE
30-04-2018	22	MALE	WHITE	2	FEMALE	WHITE
31-01-2018	26	MALE	BLACK	30	MALE	BLACK
31-01-2018	23	MALE	BLACK	19	MALE	BLACK
31-01-2018	31	MALE	BLACK	26	MALE	BLACK
31-03-2018	21	MALE	BLACK	31	FEMALE	BLACK
28-02-2018	18	MALE	BLACK	20	MALE	BLACK
31-01-2018	22	MALE	WHITE	57	MALE	WHITE
31-01-2018	33	MALE	BLACK	48	MALE	WHITE
28-02-2018	21	MALE	WHITE	29	MALE	WHITE



You can also visualise pie charts of perpetrated crimes and police killings that display information as follows: racial demographic of crimes committed over the input period and





state selected and racial demographic of police killings in the selected time period and state. Using this information, one can visualize the correlation between crimes committed by the majority of the demographic race creating a racial bias and leading to police prejudice and discrimination leading to increased police shooting on the same demographic.

XI. Project Overview

Through this project, we gained a great deal of knowledge about semantic web programming and its many uses. Finding relevant datasets or getting data via web scraping was the first issue. Following the integration of the two approaches, we had to clean the data, which meant closely examining the table data and figuring out how each field affected the overall interpretation. We then created ontologies to ascertain the unique characteristics and relationships between each of our datasets. Linking the data was the next step, which involved carefully assessing and figuring out how datasets were connected to one another. We used ReactJS for the UI and MUI for the pie charts. We used FAST api and integrated it with stardog where we fetched the data by running SPARQL queries.

XII. Roles and Responsbilities

Venkata Sai Mohan Kumar Pallapothu - In my primary role, I played a key part in developing a knowledge graph using Stardog. This involved undertaking essential mapping data to the knowledge graph, and contributing to the formulation of SPARQL queries. Additionally, I extended support in the development of backend code.

Wasim Ahamed Syed - In my position, my primary focus was on backend development. I leveraged FastAPI to establish a seamless connection between our knowledge graph and the frontend. I played a role in updating the knowledge graph from previous comments. Moreover, I offered support in data cleaning.

Akshay Reddy Kola - My responsibility involved constructing the frontend of the application using React. I also developed APIs and provided assistance in the overall integration process. Additionally, I was involved in data cleaning, ensuring that the data adhered to the specified requirements.

Anish Unnikrishnan Nair - In my position, my primary focus was on creating and updating the ontology, establishing connections to an upper-level ontology. I played a significant role in data cleaning, SPARQL queries and contributed to project documentation by writing and updating sections of the project report.

Sri Shashank Mandava - I focused mainly on supplying information for the report and carrying out data cleaning.

XIII. Future Scope

Our project is built keeping scalability in mind, and can be expanded upon using additional datasets and more data to obtain greater insights upon the demographic, with a greater census data we will be able to get a greater correlation and this information can be used to create awareness. There is other ways to improve upon the app we made, we could add a map to visualise location. One other way to improve upon this project is obtaining data from recognised government websites in real time and updating the database to provide more accurate and up-to-date insights.

XIV. Conclusion

This app allows us to view irrefutable factual evidence linking police killings to racial prejudice and racial bias created by the criminal demographic in particular areas. This can also be used to identify and reflect upon crime hotspots and how they are the areas to focus upon. We can also find correlation between more number of policeshootings in an area with respect to the crime rate drop over the years. This can be a useful educational tool for public offices and be used to make a change in the way police and even common people think.

References

[1] https://www.kaggle.com/datasets/kwullum/fatal-police-shootings-in-the-us?select=PoliceKillingsUS.csv

[2] Regalia, Blake, Krzysztof Janowicz, Gengchen Mai, Dalia Varanka, and E. Lynn Usery. "GNIS-LD: Serving and Visualizing the Geographic Names Information System Gazetteer as Linked Data." In European Semantic Web Conference, pp. 528-540. Springer, Cham, 2018.

[3] Desmond, M., Papachristos, A. V., & Kirk, D. S. (2016) - "Police Violence and Citizen Crime Reporting in the Black Community." American Sociological Review.

[4] Nix, J., Campbell, B. A., Byers, E. H., & Alpert, G. P. (2017) - "A Bird's Eye View of Civilians Killed by Police in 2015: Further Evidence of Implicit Bias." Criminology & Public Policy.

[5] Ross, C. T. (2015) - "A Multi-Level Bayesian Analysis of Racial Bias in Police Shootings at the County-Level in the United States, 2011–2014." PLOS ONE.

[6] Fryer, R. G. (2019) - "An Empirical Analysis of Racial Differences in Police Use of Force." Journal of Political Economy.

[7] Miller, T. R., Lawrence, B. A., Carlson, N. N., Hendrie, D., Randall, S., Rockett, I. R. H., Spicer, R. S. (2017) - "Perils of police action: a cautionary tale from US data sets." Injury Prevention.

[8] https://www.icpsr.umich.edu/web/NACJD/studies/37838/datadocumentation

[9] https://catalog.data.gov/dataset/crimes-one-year-prior-to-present

[10] https://catalog.data.gov/dataset/nypd-shooting-incident-data-historic

[11] https://docs.stardog.com/

[12] https://mappingpoliceviolence.org/about

[13] https://pypi.org/project/djangorestframework/