|  |  |  |
| --- | --- | --- |
| Exploring Racism in the U.S. Criminal Justice System using DCAT | August 3  2015 | |
| [Type the abstract of the document here. The abstract is typically a short summary of the contents of the document. Type the abstract of the document here. The abstract is typically a short summary of the contents of the document.] | | Semantic Web Search Engine |

Term Project: Exploring Racism in U.S. Criminal Justice System using DCAT

# Objective

I am basing my course project on DCAT- Data Catalog Vocabulary. I will be using semantic web programming combined with data analytics to conduct my course project in proving certain effects of Racial Profiling in America based on various perspectives and sources. My goal is to build a semantic search engine for seeking specific statistical data regarding racial profiling and police brutality.

# Significance

Over the past years there have been an escalating number of unjustifiable deaths to American citizens caused by the police and federal law enforcement agencies within this country. Allegations of the use of excessive force by U.S. police departments are generating today’s headlines. In Staten Island, NY, July 2014, the death of Eric Garner died because of the apparent use of a chokehold by an officer sparked outrage. A month later Michael Brown, in Ferguson MO, got shot by police officer Darren Wilson and the grand jury’s decision not to indict him triggered further unrest. In November, Tamir Rice, who was only 12 years old and playing with a toy pistol, got shot by police in Cleveland, Ohio. On April 4, 2015, Walter L. Scott was shot by a police officer after a routine traffic stop in North Charleston, S.C. The same month, Freddie Gray died while in police custody in Baltimore, setting off widespread unrest.

A total of 623 people have already been killed this year throughout the US according to ["The Counted"](http://www.theguardian.com/us-news/ng-interactive/2015/jun/01/the-counted-map-us-police-killings). More disturbingly, half of the people that were killed by these police officers and federal agents were minorities, in which make up no greater than 38% of the US’s total population. Being a young African American man residing in Florida, which has one of the highest numbers of people killed by the police in the US as to date, I am affected greatly by these statistics.

The purpose of this project is to be able explore and examine statistical data analysis on topics dealing with Racial Profiling in America. I will use multiple perspectives and sources in order to perform data analytics to answer questions such as: Are minorities in America more likely to be killed by cops than whites? Are minorities in America more likely to be injured by cops than whites? Are minorities in America more likely to die during medical operations than whites? The statistical data obtained from the provided datasets created by my search engine will prove that the U.S. criminal justice system is a race-based institution where African-Americans are directly targeted and punished in a much more aggressive way than whites.

# Tools and Languages;

I will use DCAT, CKAN, Java, and Jena to combine collected data into one dataset I will be able to examine in order to prove my hypothesis.

## DCAT

DCAT is an RDF vocabulary designed to facilitate interoperability between data catalogs published on the Web. DCAT will be used to describe my datasets in data catalogs. This will increase the discoverability and enable applications easily to consume metadata from multiple catalogs. The namespace for DCAT is [http://www.w3.org/ns/dcat#](http://www.w3.org/ns/dcat). DCAT itself defines a minimal set of classes and properties of its own including dcat, dct, dctype, foaf, rdf, rdfs, skos, vcard, and xsd. DCAT is well-suited for representing government data catalogs such as Data.gov, which I will be using to provide some of the data for my study.

## CKAN

CKAN is the underlying platform used by Data.gov. It is a powerful data management system that makes data accessible by providing tools to streamline publishing, sharing, finding and using data. CKAN will help make the data I collect be open and available.

# Datasets

The datasets included in this study contain statistical information regarding a racial demographic breakdown of the U.S. population, the number of police shootings by race, the amount of justified and unjustified shootings by police, the percent of people killed by police, the total number of police shootings ordered by race, and the number fatal injuries by legal intervention.

* **A racial demographic breakdown of the U.S. population**

<http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk>

* **Justifiable Homicides by Police**

<https://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/2013/crime-in-the-u.s.-2013/offenses-known-to-law-enforcement/expanded-homicide/expanded_homicide_data_table_14_justifiable_homicide_by_weapon_law_enforcement_2009-2013.xls>

* **Arrest-Related Deaths**
* **The number of police shootings by race**

<http://www.bjs.gov/index.cfm?ty=pbdetail&iid=5259>

* **The number fatal injuries by legal intervention**

<https://wisqars.cdc.gov:8443/cdcMapFramework/mapModuleInterface.jsp>

# Inputs and Outputs

The main sources I will be collecting my data from will be Data.gov, the United States Census Bureau, and FBI.gov. This data will be going into the created search engine involves combing multiple sources in order to acquire a broader view of various datasets. In return, I will be able to perform data analytics by cross examining and comparing the retrieved stats.

The first input dataset being used is **a racial demographic breakdown of the U.S. population,** provided by **Data.gov**. This dataset will provide the information on the Annual Estimates of the Resident Population by Sex, Race, and Hispanic Origin for the United States, States, and Counties.

The second input dataset will be the **Justifiable Homicides by Police** data provided by **FBI.gov**. This data contains the number of justifiable deaths recorded by police and which weapon was used.

The third input dataset will be the total number of **Arrest Related Deaths** by police. This dataset, provided by the “Bureau of Justice Statistics” (BJS.gov), will have information on the number of police shootings by race, gender, demographics, etc.

The fourth input dataset being used is **The Number of Fatal Injuries by Legal Intervention,** provided by **CDC.gov,** will show information on number of people that died due to police brutality and fatal injuries.

The output of this data will answer the 3 questions laid out earlier in this paper. If the provided data is manipulated properly and the proposed questions are correctly answered according to the produced output, the results should provide a beneficial correlation on data regarding racial profiling and police brutality.

### Questions

1. Are minorities in America more likely to be killed by cops than whites?

2. Are minorities in America more likely to be injured by cops than whites?

3. Are minorities in America more likely to die during medical operations than whites?

# Possible Challenges

A challenge I may run across is the lack of police shooting data. This is due to the reason that law enforcement agencies lack “sufficient incentives” to report officer-involved shootings. Local agencies also don’t properly report the injuries and deaths of police offenders at times either. The Federal Bureau of Investigation captures data on justifiable homicides by law enforcement officers, but reporting is **voluntary** and limited to instances in which a civilian is killed while committing a felony.