

Investigation into Over Policing and Racial Bias: Evidence from Traffic Stops in Nashville, Hartford, and Stockton

Alex, Conroy (08324841)

IFN703/4 Assessment 1, due 11:59pm Sunday 28 March 2021

Abstract

Modern policing has begun to reap the rewards of the second and third waves of industrial revolution [1], with multiple big data tools at the fingertips of officers. However, there is mounting evidence to suggest that rather than help drive out human based biases, such as those involving race. These tools are reinforcing them [2]. This combined with the often complete lack of transparency around policing data in general, has left police-community relations at what appears to be all time lows, particularly in the in the United States of America (US) [3].

This project aims to analyse traffic data from around the US using statistical tools like the Generalised Linear Models to drive an Outcome based biasing model. To determine if the likelihood that Minorities in the US are being pulled over and subjected to over policing at higher rates than their Caucasian counterparts.

The project will look to accomplish this by analysing the cities of Nashville, Hartford, and Stockton with data sources from the Stanford Open Policing dataset combined with National Census data. These cities were chosen as they meet the variables that have been determined are necessary to undertake the Outcomes modelling, and each provide a look into a city that has a statistically larger percentage of a major ethnical/racial group than the US at large.

Title

Proposed Title: Investigation into Over-policing and Racial Bias: Evidence from Traffic Stops in Nashville, Hartford, and Stockton

Introduction

Modern policing is one of the many professions that has benefited greatly in the age of big data. However, due to either a real or perceive lack of transparency in application and reporting, this use has helped to exacerbate underlying tensions between the community and the police force. This is particularly apparent in the US. A country in which police interactions are often deadly [4] and have long been linked to racial biases [5].

This is exemplified in the Black Lives Matter protests and riots following the police related deaths of George Floyd and Breonna Taylor [3] with a reportedly nearly 8% of the total population of the US

participating [6]. Despite or because of this, governmental research into this issue have seen heavy politicisation.

This has given rise to several independent attempts to collect, quantify, and analyse data in this area. Such as the Stanford Open Policing Project (SOP) [7] and Data for Black Lives [8]. These datasets were created with the hope of better understanding the outcomes and affects of the current state, and realised impacts of changes to the current state of community interaction with police through a racial lens.

The data that will be used for this project is a subset of the Stanford Open Policing Project. The SOP began in 2015 and involved the collection and standardization of traffic stop related data for several cities and most states in the US.

The specific dataset that will be used in this project consisting of two years of data from 2013 – 2014, for the cities of Nashville, Hartford, and Stockton. These cities were selected as they contained information related to; Race, geographical location of stop, at least half of the indicators on stop outcome, and the reason that the stop occurred. Above this however, they were selected as each city has a higher than statistically expected population of the three major ethnicities found in the US shown as outlined in figure 1.

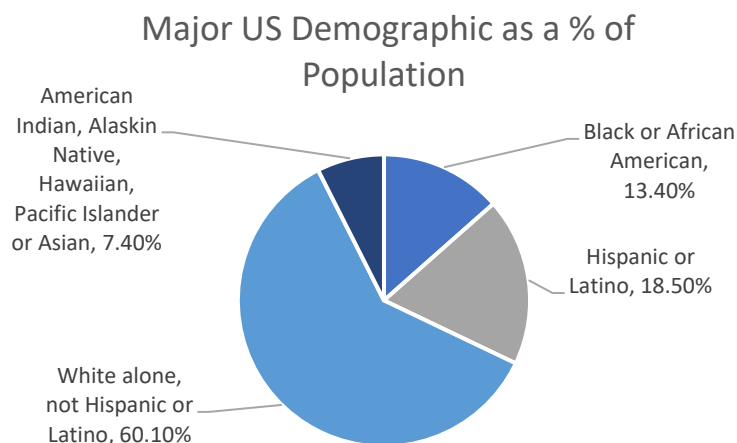


Figure 1 Major Racial Demographics reported in the United States of America [9]

With Stockton reporting 40% of its residents identify as Hispanic [10], Hartford reporting 38% of residents identify as Black or African American [11], and Nashville reporting 64% of its residents identify as White [12]. The years chosen to analyse are the closest available across all the cities to the 2010 census that will be used to link into the SOP data.

Once exploration and cleaning has occurred, the project will look to use an optimised Generalized Linear Model (GLM). This will produce the most likely outcome of a traffic stop at both a suburb and city level. When combined with the census data this will provide the ability to decipher the difference, if any between what statistically should occur in that location and what is actually occurring.

Literature Review

Racial Bias and the Police

The public can interact with the police in many ways. One of the most frequent ways [13] these interactions do occur is through a traffic stop. As such traffic stops are considered one of the most used applications in the policing toolkit [14].

Research into any racial biasing in traffic stop data has generally shown that African Americans and Hispanics are both pulled over at higher rates than Caucasians [13]. This higher rate was found even after controlling for factors around neighbourhood benchmarking through the US census, or the argument that minorities commit more crime [14].

That is not to say there are not circumstances in which Caucasian driver stop rates are found to be more than other ethnic groups. A study from Cambridge University on drivers in Missouri found that instances of financially motivated or 'revenue raising' fines are more likely to be targeted towards White drivers. With the effects being found to be particularly strong in areas where over policing of Black drivers was already present. It is believed that this occurs as the Black drivers already have 'too many' fines to pay, and the White drivers are presumed to be wealthy enough to afford the cost [15].

A study by Simoiu, Corbett-Davies and Goel posit that some of these results could be attributed to the shortcomings on the benchmarking style used in determining bias. Instead an outcomes based model should be used to understand if a bias is occurring in multifaceted decision making such as that found in a traffic stop [16].

Public Perception of Policing

Colloquially negative perceptions of policing have been around since policing began. The US has even had a term coined after the act of police detaining a member of the public with little evidence, a 'Terry Stop' for over 50 years. Named after a supreme court case in the late 60s [17].

Interestingly research conducted by Boehme, Cann and Isom [18] suggests that the public's perception of police is more intrinsically tied to the stability and ethnical uniformity of the neighbourhood's makeup rather than actual police outcomes. Further, when considering minority communities, the negative perception of police is present and directed towards both under and over policing. Drawing conclusions that both have negative effects on these communities [19].

Studies into policing of multicultural communities have confirmed that minorities do indeed suffer under both conditions. With further evidence to suggest that over policing does little to prevent crime but does show evidence that undermines police-community relations [19].

Project Definition

Research Purpose

This project aims to analyse traffic stop data from the different American cities, cross referenced with demographic data collected from the US census. With the purpose of producing a model that is capable of understanding if areas within these cities are experiencing any over or under policing. The project will then attempt to further the analysis by understanding if there are any ethnic markers that might aid in understanding what is driving these areas of under or over police presence.

Research Problem & Significance

Many organisations are now independently attempting to create or corroborate policing related data. This project aims to utilise the Stanford Open Policing Project (SOP) in analysing data in three US cities. These cities represent a cross section of the geographic and demographic subsection of the United States of America. While there have been many papers already published on police interactions using this dataset, this project aims to discover general links between poor policing outcomes, before investigating if these poor outcomes can then be linked to any racial markers. Rather than looking for instances of unequal racial outcomes and then looking if certain interactive or judicial changes can be linked to these outcomes.

Research Objective

The main objective of this research is to understand if there are areas that can be characterised as over policed in the cities of Nashville, Hartford, and Stockton. These cities were selected due to their populations having an overrepresentation off one of the three main racial demographics in the US: Hispanic, African American, and Caucasian. Then building off these insights to see if it is possible to indicate that those ethnicities are overrepresented in areas identified as inadequately policed.

Scope

The scope of this project is to analyse areas within Nashville, Hartford, and Stockton at a suburb level that mostly identify as one of the three target demographics using census data for a two-year period. This will then be used to cross examine the outcomes at both suburb and city levels of granularity.

Project Plan

Research Approach

The proposed approach for this project is to perform the modelling and analytics in an agile sprint form, with the project broken down into five two-week sprints. The sprint objectives are shown in the table below.

Sprint	Objectives
1	Exploratory analysis of Nashville data and build of basic GLM model and hypothesis testing of alternative models. Outcome: Basic model tested and running on SOP data
2	Exploratory analysis of Hartford and Stockton. Run chosen model against full dataset. Outcome: Tested model running on full dataset
3	Run full model and dataset through optimisation trials. Outcome: Optimised model running on full dataset
4	Explore model results, conclude results from study, any future work required or recommended, and build appropriate visualisations. Outcome: Optimised model provides link to an overall outcome and visualisations are ready for reporting.
5	Prepare project presentation and reporting. Outcome: Finalised project outcomes presented for review and grading

Table 1 Sprint Objectives with Outcomes

This approach has been chosen to help gain a more rapid insight into potential results and best practices while working on larger subsets of the data, that will allow for a smoother transition for the scaling of the model inputs. This will mean that lessons learnt from previous sprints will be put into practice immediately.

It should be noted that while it is being called the 5th sprint, it is not technically an agile sprint. It is hoped that smaller chunks of reporting will be conducted during sprints 1-4. To allow for sprint 5 to act as a buffer for any time overruns, or if there is a need to complete a more intensive reporting process. If the model proves to be inconclusive, inappropriate, or otherwise needing a dedicated time spend.

Research & Analysis

This section looks to expand and define the initial assumptions and models that the project will use as a basis for the analysis. These will be sense checked and validated during the project.

Policing level criteria

The continued research for the project will be centred around finalising the definitions on what can appropriately be called over/under policing. As a starting point however, it will be defined as anytime the model produces a likelihood estimation for a given area where the outcome of a stop is that there was no outcome, i.e. that the person(s) was stopped for no reason. Similarly as a starting point, the motivations will be attributed to some form of racial bias if the likelihood of the ethnicity of the driver

is over representative given the percentage of the population attributed to any of the ethnical demographic for the area.

Data Modelling

The modelling approach that will be used to produce the outcomes of this project is the Generalized Linear Model (GLM). The GLM is a unified formula that will allow for the model to not only express a maximum-likelihood estimate but do so with the most appropriate distribution model for the data [20]. The use of a GLMM or Generalized Linear Mixed Model will also be considered, to understand if factoring the random effects in the linear predictor will achieve a more appropriate result [21].

Given there are two main variables involved in the question being asked by this project. The models will initially be run using just the outcome, or ethnic variables, along with a combination of both.

Research Criteria

The research criteria outlined in the below table will be applied to make sure any potential data sources are of a high enough standard for the level of analytics and research being conducted in this project. While also ensuring that the research topics within are relevant to policing analytics and racial biasing.

Criteria	Inclusion	Exclusion
Year Published	Ideal timeframe: 2015 – 2021 Acceptable timeframe: 2005+	Pre-2000
Language	English	All non-English or translated sources
Source minimum	Published in a peer-review Journal, conference paper.	All other sources.
Idealised source	Highly regarded journal	N/A
Topic	Policing Data, Biasing in Data, Biasing in Policing	All other research.
Idealised Topic	Racial Biasing in Policing that use the Stanford Open Policing Database.	Publications that do not use some form of data to as source for any claims made.

Table 2 Inclusion & Exclusion Criteria

Time Planning

The project will take place over semester 1 2021 at Queensland University of Technology. The major timeframe for the project is approximately 10 weeks. Considering the time to select and propose a research topic and the 1-week mid-semester break. Linking with the key deliverables, and sprints described above, a flight plan is provided below that represents these deliverables and their sub streams.

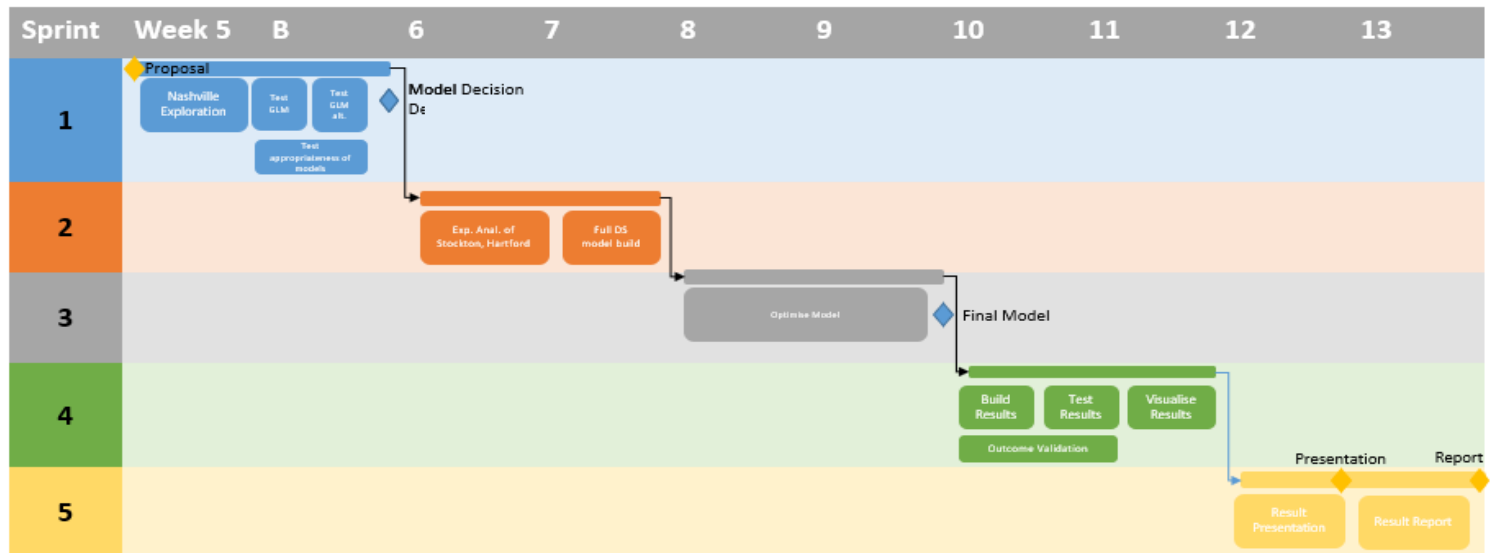


Figure 2 Project Flight Plan

Risk Management

A risk assessment matrix is provided below, that highlights the risk management strategies for the potential hazards identified for the lifecycle of this project. These have been provided using the RAID methodology and include a mitigation strategy [22].

Raid Rating	Title (Area)	Description/Mitigation
Unlikely/Minor	Illness/Lockdown (Risk)	The current global pandemic is still in effect. However, Australia is currently in a good position compared to other parts of the world, and all members of the project team have received a vaccine. Making both infection and adverse effects from infection low. Risk of lockdown would still have some effect on individual movement. Mitigation: Follow all Governmental instructions on how to protect project members from the virus. Enable the project to be completed remotely from the start.
Likely/Moderate	Programming Issues (Risk)	The project is programming based and all deliverables are reliant on successful coding. Mitigation: Ensure an appropriate control software such as GitHub is used for version control and offsite backups.
Likely/Major	Timeslip (Risk)	The unit has large times between deliverables, that might cause milestones to go missed. Leading to a major time crunch near the end of semester. Mitigation: Attendance of scheduled supervisor meetings, to allow for feedback on progress.
Unlikely/Major	Depth of Research (Assumption)	During the undertaking of this project it is determined that there is either; not enough research/data necessary to conduct a proper analysis, or that the research being undertaken is too derivative of other published articles. Mitigation: Conducting of a thorough literature review and completion of a list of pivot areas that would allow the research being conducted to adjust to any unforeseen articles that are too closely related in nature.

Table 3 RAID Risk Register

References

- [1] A. G. Ferguson, *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*. NYU Press, 2019.
- [2] N. Turner Lee, 'Detecting racial bias in algorithms and machine learning', *J. Inf. Commun. Ethics Soc.*, vol. 16, no. 3, pp. 252–260, Jan. 2018, doi: 10.1108/JICES-06-2018-0056.
- [3] J. Chernega, 'Black Lives Matter: Racialised Policing in the United States', *Comp. Am. Stud. Int. J.*, vol. 14, no. 3–4, pp. 234–245, Oct. 2016, doi: 10.1080/14775700.2016.1267322.
- [4] J. M. Shane, B. Lawton, and Z. Swenson, 'The prevalence of fatal police shootings by U.S. police, 2015–2016: Patterns and answers from a new data set', *J. Crim. Justice*, vol. 52, pp. 101–111, Sep. 2017, doi: 10.1016/j.jcrimjus.2017.05.001.
- [5] A. Gelman, J. Fagan, and A. Kiss, 'An Analysis of the New York City Police Department's "Stop-and-Frisk" Policy in the Context of Claims of Racial Bias', *J. Am. Stat. Assoc.*, vol. 102, no. 479, pp. 813–823, Sep. 2007, doi: 10.1198/016214506000001040.
- [6] 1615 L. St NW, Suite 800 Washington, and D. 20036 USA 202-419-4300 | M.-857-8562 | F.-419-4372 | M. Inquiries, 'Majorities Across Racial, Ethnic Groups Express Support for the Black Lives Matter Movement', *Pew Research Center's Social & Demographic Trends Project*, Jun. 12, 2020. <https://www.pewresearch.org/social-trends/2020/06/12/amid-protests-majorities-across-racial-and-ethnic-groups-express-support-for-the-black-lives-matter-movement/> (accessed Mar. 28, 2021).
- [7] 'The Stanford Open Policing Project', *openpolicing.stanford.edu*. <https://openpolicing.stanford.edu/> (accessed Mar. 28, 2021).
- [8] 'Data 4 Black Lives'. <https://d4bl.org/> (accessed Mar. 28, 2021).
- [9] U. C. Bureau, 'Population', *The United States Census Bureau*. <https://www.census.gov/topics/population.html> (accessed Mar. 28, 2021).
- [10] 'U.S. Census Bureau QuickFacts: Stockton city, California'. <https://www.census.gov/quickfacts/fact/table/stocktoncitycalifornia/INC110219> (accessed Mar. 28, 2021).
- [11] 'U.S. Census Bureau QuickFacts: Hartford city, Connecticut'. <https://www.census.gov/quickfacts/fact/table/hartfordcityconnecticut/PST045219> (accessed Mar. 28, 2021).
- [12] 'U.S. Census Bureau QuickFacts: Nashville-Davidson (balance), Tennessee'. <https://www.census.gov/quickfacts/fact/table/nashvilledavidsonbalancetennessee/IPE120219> (accessed Mar. 28, 2021).
- [13] E. Pierson *et al.*, 'A large-scale analysis of racial disparities in police stops across the United States', *Nat. Hum. Behav.*, vol. 4, no. 7, pp. 736–745, Jul. 2020, doi: 10.1038/s41562-020-0858-1.
- [14] A. Chohlas-Wood, S. Goel, A. Shoemaker, and R. Shroff, 'An Analysis of the Metropolitan Nashville Police Department's Traffic Stop Practices', p. 10.
- [15] A. P. Harris, E. Ash, and J. Fagan, 'Fiscal Pressures and Discriminatory Policing: Evidence from Traffic Stops in Missouri', *J. Race Ethn. Polit.*, vol. 5, no. 3, pp. 450–480, Nov. 2020, doi: 10.1017/rep.2020.10.
- [16] C. Simoiu, S. Corbett-Davies, and S. Goel, 'The problem of infra-marginality in outcome tests for discrimination', *Ann. Appl. Stat.*, vol. 11, no. 3, Sep. 2017, doi: 10.1214/17-AOAS1058.
- [17] S. Goel, M. Perelman, R. Shroff, and D. A. Sklansky, 'Combatting Police Discrimination in the Age of Big Data', *New Crim. Law Rev.*, vol. 20, no. 2, pp. 181–232, May 2017, doi: 10.1525/nclr.2017.20.2.181.
- [18] H. M. Boehme, D. Cann, and D. A. Isom, 'Citizens' Perceptions of Over- and Under-Policing: A Look at Race, Ethnicity, and Community Characteristics', *Crime Delinquency*, p. 0011128720974309, Dec. 2020, doi: 10.1177/0011128720974309.
- [19] G. Ben-Porat, 'Policing multicultural states: lessons from the Canadian model', *Polic. Soc.*, vol. 18, no. 4, pp. 411–425, Dec. 2008, doi: 10.1080/10439460802094686.

- [20] J. J. Faraway, *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models, Second Edition*. CRC Press, 2016.
- [21] W. W. Stroup, *Generalized Linear Mixed Models: Modern Concepts, Methods and Applications*. CRC Press, 2012.
- [22] D. Gibson and A. Igonor, *Managing Risk in Information Systems*. Jones & Bartlett Learning, 2020.