Small-Town Police Accountability: Challenges and Recommendations

Claire Kelling[*], Anna Haensch[†‡], Ariana Mendible[§], Spencer Brooks[¶], Alex Wiedemann[‖], Manuchehr Aminian[**], Wade Hasty[¶], Jude Higdon[¶††]

July 13, 2023

Abstract

According to the 2020 US Census more than 60% of the US population lives in towns with fewer than 50,000 residents, yet this is not in proportion with the research and public data surrounding policing, which focus on large and dense urban areas. One reason for this disparity is that studying small-town police departments presents unique obstacles. We present some of the challenges that we have encountered in studying small-town police activity such as data availability, quality, and identifiability, and our solutions to these challenges using computational tools. Finally, we give our recommendations in getting involved in this space based on our efforts to-date.

Keywords: data science, quantitative social justice, police reform, small-town, natural language processing, spatial statistics

[*]Department of Mathematics and Statistics, Carleton College, Northfield, MN
[†]Corresponding author: anna.haensch@tufts.edu
[‡]Data Intensive Studies Center, Tufts University, Medford, MA
[§]Department of Mathematics, Seattle University, Seattle, WA
[¶]Institute for Quantitative Study of Inclusion, Diversity, and Equity, Williamstown, MA
[‖]Department of Mathematics, Randolph-Macon College, Ashford, VA
[**]Department of Mathematics, California State Polytechnic University, Pomona, CA
[††]The authors gratefully acknowledge *please remember to list all relevant funding sources in the unblinded version*

1 Introduction

According to the 2020 census, roughly 60% of the US population lives in towns with fewer than 50,000 people [36], yet the overwhelming majority of data and scholarship related to policing is focused on large cities. While crime and police violence in urban area receives most of the attention from scholars and mainstream media, the need for research on small-town policing has never been greater. Not only has the rate of fatal police shootings steadily increased in rural areas over the last decade, but that rate of fatal police shootings per criminal homicide is known to increases sharply as town size decreases [28]. The history of policing in the United States is wrought with abuses of power, and recent events have acutely highlighted the need for oversight and transparency in policing. Since the high-profile murder of George Floyd in 2020, there have been widespread calls for police reform, but lack of data has been a considerable bottleneck in translating these calls to action into concrete policies and recommendations [25].

With the goal of enabling citizens to hold their governments accountable, the federal Freedom of Information Act (FOIA) codifies the rights of the general public to access data on their governance, including policing. While access to such data is legally guaranteed, significant challenges exist in acquiring and using it. In small and under-resourced towns, infrastructure and personnel constraints make data collection even more difficult which places additional burdens on researchers, activists, and policymakers who want to understand the shape of policing in their town. Official FOIA requests must be made and often require persistence and legal expertise to be successful. Once the data are obtained, further technical skills may be needed to process, understand, and draw meaningful insights from them. For these reasons transparency in policing is often limited to the data requested and aggregated by highly-resourced groups.

There are several large-scale data collection efforts on policing across the United States. These include initiatives such as the Police Data Initiative [21], Mapping Police Violence [8], Fatal Encounters [6], Stanford Open Policing Project [26], FiveThirtyEight's Police Misconduct Settlement database [35], the Police Scorecard [30], the Chicago-based Citizens Police Data Project [15] and others. Comparisons of the availability, amount of coverage, and topics described in these data initiatives are given in Table 1.

Table 1: Comparison of national police data collection initiatives, including number of cities represented in the initiative, the average population of those cities, the average number of datasets per cities (for example, cities could include both use of force and calls for service), and the data type represented.

Data Initiative	# of Cities	Avg Pop	Avg # of Data Sets per City	Data Type
Fatal Encounters	5,635	31,038	1	Single-incident reports of people killed by or in the presence of police
Police Scorecard	13,469	44,216	1.3	Summary statistics of incidents, departments, and incarceration
Mapping Police Violence	3,303	45,920	1	Single-incident reports of police violence
Five Thirty Eight	48	457,091	1	Full-coverage, incident-level reports of police settlements
Police Data Initiative	71	534,686	3	Full-coverage, incident-level reports of police interactions, arrests, and misconduct
Open Policing	52	606,996	1	Full-coverage, incident-level reports of police stops
Citizens Police Data Project	1	2,693,976	1	Full-coverage, incident-level reports of officer complaints in Chicago

These initiatives provide a degree of transparency to the public about how communities are being policed. Access to these data sets allow the public to understand patterns, to describe policing behavior in detail, and to advocate for change. However, most of these initiatives focus on large cities. Figure 1 highlights the distribution of the population sizes of census-designated places (dashed blue) compared to the distribution of population sizes of cities/towns where detailed information is available within the collections highlighted above (solid orange). Out of the full-coverage, incident-level data sets in national and inter-regional data collection initiatives listed in Table 1, only 10.4% of the data covers cities under 50,000 people. From this figure, we see that the distribution of the population of census-designated places in the United States tends to be far lower than the population of places that are fully represented in some of the largest policing data collection initiatives. This highlights the need to collect data and better understand policing in small towns that are more representative of the US population. Small towns sometimes collect similar data (time, location, individuals involved, description, charges, etc.) but the resulting information is often not made publicly available, especially not at the incident level. This creates a barrier for community members and for researchers seeking to better understand how small towns are being policed.

As an effect of this data limitation, much of the research on policing and criminology also focuses on the study of large cities. We conducted a systematic literature review of the last several years of papers in two top policing journals. This included 145 papers from the *Journal of Quantitative Criminology* from 2017-2021 and 336 papers from the journal *Policing and Society* from 2016-2021. From the *Journal of Quantitative Criminology*, 67 cities/towns were named in these analyses and only three of these cities had populations of less than 100,000 people (Wilmington, DE; New Castle, DE; Somerville, MA) and 12 cities with population less than 500,000 people. From *Policing and Society*, of the 89 cities/towns identified, 18 had populations of less than 100,000 people, 34 with less than 500,000 people. Several papers redacted location names due to privacy issues. We believe that increased availability of public data and discussion on how to collect data from small towns will lead to increased research and therefore knowledge about how small towns are being policed.

Recent work on small-town policing has illustrated the importance of studying small

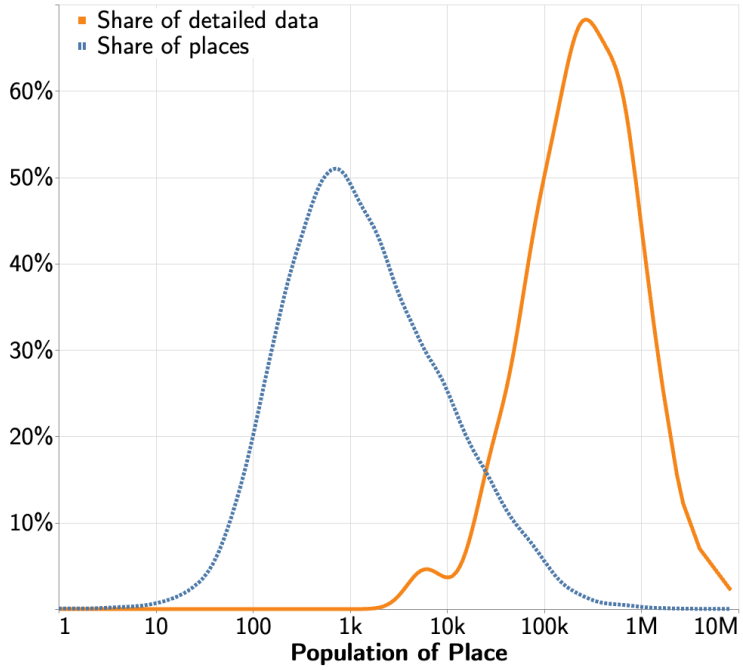


Figure 1: Comparison of the distribution of towns and distribution of comprehensive data for various population sizes. Share of places (dashed blue) shows the relative share of census-designated places of a given size. Share of detailed data (solid orange) shows the share of full-coverage, incident-level data from the data initiatives in Table 1 for the same population sizes. As some cities are represented in multiple data sources, e.g. New York City having both arrest and police misconduct data, each unique data set within the data initiatives is represented as one observation.

towns but also the lack of available data to statistically analyze policing dynamics. For example, there have been studies on the use of social media by small-town police departments to understand what information is being released to the public [4] as well as studies on the fear of crime in a survey of a small town [5]. Call for service and officer-completed surveys have been used to study response to mental-health related calls in rural areas [40]. Interview data has been used to analyze the disorder and conflict in a college town [39]. Other qualitative studies have also been pursued with small-town police departments (cf. [1, 24]). Notably, most of the above studies rely on interview data rather than publicly available, event-level data, such as call for service data, to analyze policing dynamics in small towns.

In this work, we present challenges and potential approaches to collection and analysis of different aspects of these small-town policing data sets. We hope that what we have learned through our work with the Small-Town Policing Accountability (SToPA) Lab will be useful to researchers, community members, and others looking to better understand how their communities are being policed. We present a pair of case studies: one small town in New England and one large town in the Southeast. Each of these is accompanied by its own set of processing and analysis challenges and opportunities. In Section 2, we introduce and contextualize the case studies. In Section 3, we describe the analysis pipeline for the data sets in these two towns. We include details such as data collection from a police department, preprocessing from messy formats, and insights into statistical analysis of cleaned data sets. Lastly, we discuss our findings and recommendations for researchers and community stakeholders in Section 4.

2 Case Studies

Although the focus of our work is small-town policing, in what follows we share two case studies: one small town and one large town. In doing so we hope to highlight some of the specific differences between the two and outline some of the challenges a scholar is likely to encounter when dealings with municipalities similar to these. The two locations are dissimilar not only in their size but also their socioeconomic and demographic makeup; some key features are shown in Table 2. One attribute that the locations share is the presence of one or more colleges or universities that have a significant impact on the region in terms of employment, emerging industry and political leaning.

The data sets were collected and driven by a variety of motivations by local activists and community organizers. A common concern is investigating and bringing patterns of police misconduct into the spotlight. Another important motivation is to investigate disproportionately high policing of Black communities. We discuss the details of each data collection process in the sections that follow.

Table 2: Key Features of Case Study Municipalities. Demographic data is as reported by the U.S. Census Bureau 2021 Population Estimates Program [38].

Municipality	Area (km^2)	Pop.	Pct. White	Median Household Income
Williamstown, MA	121.40	7,813	82.9%	$95,682
Durham, NC	300.92	326,126	54.5%	$67,000

2.1 A Small New England College Town

Williamstown, Massachusetts is home to a private liberal arts college and is the smaller and less diverse of the towns in this study. In the recent past, the Williamstown Police Department became the target of scrutiny and media attention surrounding several instances and accusations of anti-Semitism, sexual harassment, racism, and other instances of misconduct [32, 33, 16, 17]. Discussions with community activists led to specific questions about the policing of historically marginalized communities, specific department-wide practices, and the behavior of specific officers with documented patterns of misconduct.

The data for Williamstown covers police responses in 2019 and 2020. The acquisition of this data set required several months of communication and legal action, including a formal FOIA records request and escalation to the state attorney general. Ultimately, access to the data was granted after a change in departmental leadership. Though legally entitled to digitized data, due to the limited resources of this department, the records come in the form of printed call logs; there were 2,372 and 1,472 pages of records for 2019 and 2020, respectively. The data covers the majority of Williamstown PD's 2019-2020 interactions with the public, including publicly-initiated incidents like 9-1-1 phone calls, police-initiated incidents like traffic stops, and routine officer activity like building checks. According to the 2020 Williamstown Annual Town Report [37], a total of 15,973 and 10,921 calls occurred in 2019 and 2020, of which 12,298 and 7,509 are represented in the data. Absent from the data are intermittent missing calls, identified by skipped call numbers, a number of days in June 2019, and all calls past October 24, 2020. Logs may include details about officer actions, including the reason for initiation, type of incident, location, response time, officer names, call outcome, and free-form narratives, among others.

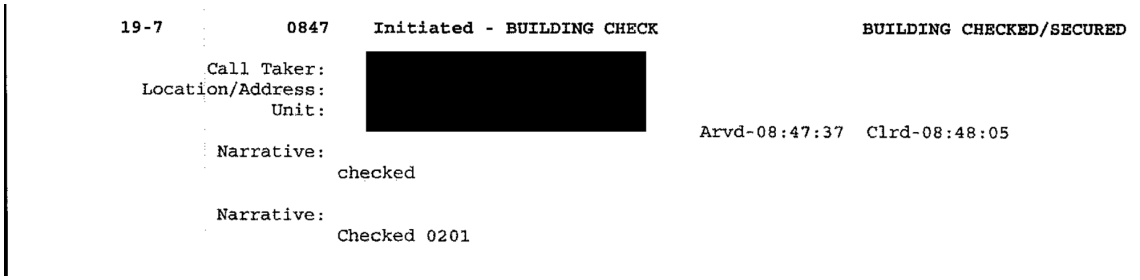


Figure 2: Partial sample Williamstown PD call log entry

2.2 A Southern Town with a Booming Tech Industry

Durham, North Carolina is part of the Research Triangle Region and is home to research universities with an economy driven by research facilities in the biomedical and pharmaceutical sciences [9]. Durham is the larger of the case study towns in both population and area. Although Durham is decidedly large, our analysis and workflow here has helped us to develop tools that could be repurposed in the analysis of small towns.

The data was originally gathered as part of a summer research project for the Duke "Data+" program during which students put together an initial framework for mining and processing this data. Broadly, the motivating question was to investigate potential disparities in policing with respect to race.

The data for Durham consists of reports gathered from the open data Durham P.D. Police to Citizen Portal [10]. The open-access portal can be used to search for arrest/incident reports for up to 30 days at a time, and download reports individually. With the help of SToPA's Durham affiliates, this process was automated via a browser driver using RSelenium [13], allowing for batched downloads. Presently, the lab has processed all entries reported in 2019. These consist of incident reports and arrest reports. A sample of each type of report with all identifying information redacted is shown in Figures 3 and 4. The complete 2019 data consists of 18,129 reports, of which 11,026 are incidents and 6,720 are arrests. Both the incident and arrest reports contain structured (address, dates, and other case ID values) and unstructured (incident descriptions and narrative) data fields.

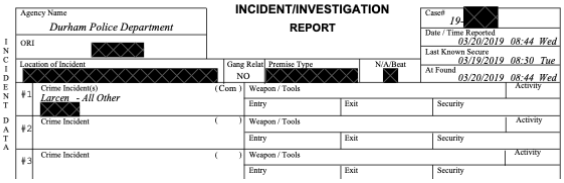


Figure 3: Partial sample incident report.

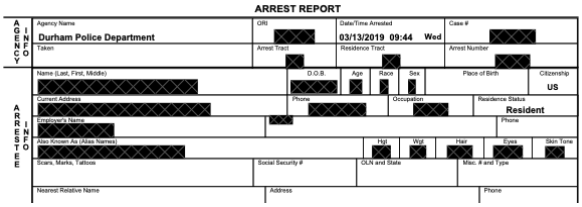


Figure 4: Partial sample arrest report.

3 Methods for Small-Town Policing Research

Though data collection is often an arduous step in the research process, the challenges are particularly acute in small towns. Some larger cities offer easy-to-access structured tables downloadable via an API (Application Programming Interface, which in general terms allows one to directly interfacing with a database). In small towns, availability is likely more limited. These case studies highlight the disparity in effort and expertise needed to collect and process small-town policing data. We describe the processes used in these instances, which we hope will serve as a guide for future work in similar locales.

3.1 Data Acquisition

The first step in this work is to gather the necessary data, which is a multi-step process outlined in Figure 5. When we set out to get data, it is important to understand the local laws around data entitlement which vary from state to state. Each state's website should contain the relevant details including the state-specific definition of *public records* and what exactly is included under that umbrella. Some non-profits like the National Freedom of Information Coalition compile these details across multiple states all in one place [20], but a state's governmental website is usually a reasonable place to start.

Once it is confirmed that the desired data is eligible for a public records request, there are multiple pathways. Often the path of least resistance is through public data portals such as the one maintained by the Durham Police Department [10]. Data from these portals can be accessed by manually downloading the data of interest, or by building automated web crawlers and scrapers such as the one described in Section 2.2. More complicated pathways involve human-managed processes such as online forms, emails, or other written

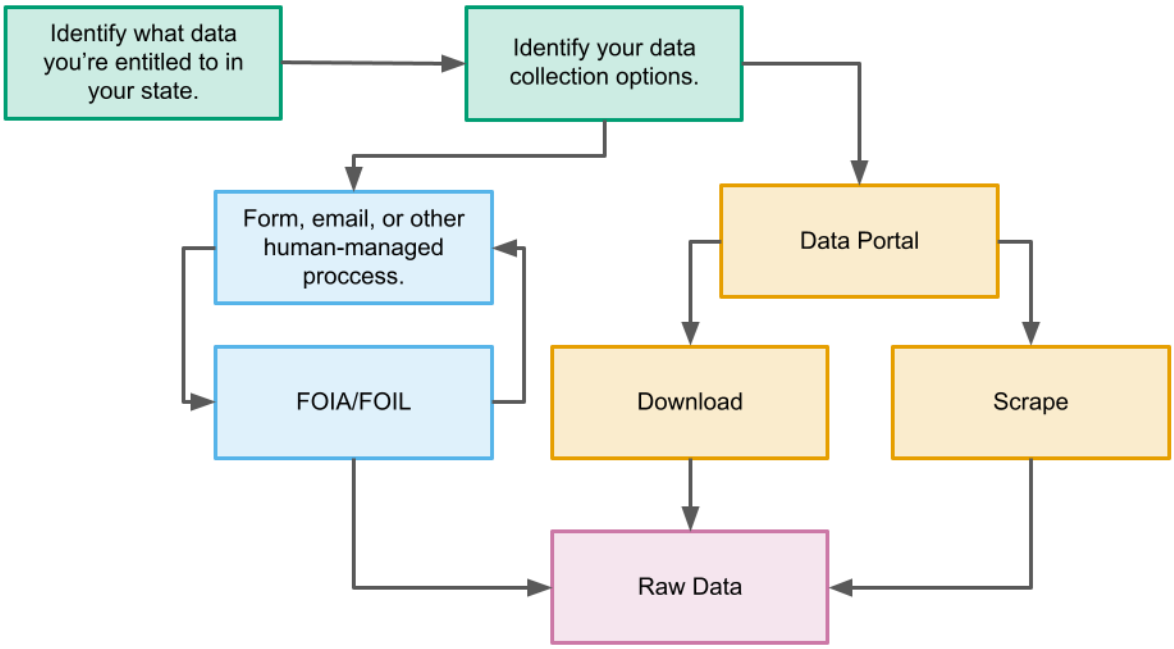


Figure 5: Data Gathering Workflow

communication. Small towns often lack the infrastructure to maintain data portals, so the human-managed pathway is the more common approach. In our experience, this method often requires multiple iterations of discussions and negotiating with local and regional government authorities before the proper data is in hand.

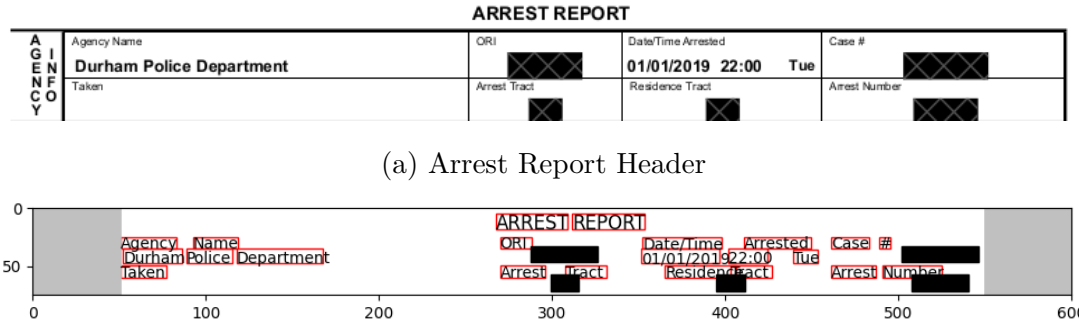
Depending on how the data is obtained, it will arrive in a variety of formats. The data might be digitally rendered pdf tables (as is the case with the Durham PD data). It could be documents printed out from a digital records system (as with the Williamstown, PD). It could even be copied, printed, or scanned versions of hand-written documents. Depending on the type and quality of data the task of turning this raw data into analyzable tabular data can vary in difficulty. In the following section we will describe some possible tools for carrying out this step.

3.2 Data Processing

One of the significant challenges of this work is parsing and preparation of multimodal data. We outline some of the specific challenges that arise in this pipeline.

3.2.1 Text Extraction

In the cases of a digitally generated pdf (e.g. a pdf form that was filled out using a computer program) there are numerous libraries that can be used to extract text. For the Durham PD data, text was easily extracted from report pdf's using the Python library PDF Plumber [29]. For manually generated pdf files it is necessary to use a more powerful optical character recognition (OCR) library such as Tesseract [22]. An OCR engine like this allows for the reading of both handwritten characters as well as scanned pdfs, as was the case in the Williamstown PD data. For either of these tools, the text extraction process is essentially the same. First, the pdf page is clipped to remove excess margins and scanning boundaries such as the thin black line seen on the far left of Figure 2. Then, text items are spatially identified on a pdf page within a bounding box as in Figure 6.



(a) Arrest Report Header

(b) Extracted Text

Figure 6: **Top:** A header from a typical arrest report. This is an example of a computer generated pdf. **Bottom:** Text extraction from an arrest report after the hatched grey regions of the document are cropped out. Words identified in the documented are shown with red bounding boxes.

3.2.2 Text Identification and Correction

Once the text has been extracted and placed within a bounding box, the next step is to identify the extracted text with the features of the original data. Extraction relies on spatial methods; typically via either identifying regions on the page where features typically appear, or relying on the known co-occurrence of certain features. For example, in records

such as the one shown in Figure 6, the case number (redacted) is always the upper and right-most integer, and the weekday abbreviation of the arrest always falls immediately to the right of the time of arrest. A drawback to these methods is that they require some knowledge of original pdf page layouts, as well as consistent formatting across records. For digitally filled pdf forms this isn't typically a major issue, but it becomes quite complex for manually generated pdfs (i.e., hand-scanned printed pages).

However, even in the situation where the data come from digitally created forms, the so-called *reading order* of a pdf file, i.e., the order of instructions to render the document exist within the pdf, can be significantly different than it appears to a human. Simply put, the order seen by the human is not how the data is stored. This means that even well-meaning attempts by police departments to digitize their data practices can be equally unhelpful to "low-tech" approaches; even a digitally generated pdf with form fields typically still requires scanning via OCR technology. We strongly recommend future efforts from law enforcement organizations towards data transparency include input from a data practitioner toward best practices.

After setting a baseline best guess for text using spatial methods, the next step is to confirm text against libraries of known values. For example, in a small town, it is typically easy to find a list of known officers to check against extracted words. This is particularly useful for manually generated pdfs, where a name like "LT. BROWN" might be returned as one of the following:

LT BROWN, LT. 8R0WN, BROWN

This sort of discrepancy can be overcome using Levenstein distance and fuzzy matching, implemented for example, in the Python package fuzzywuzzy [27]. In the case where BROWN is returned, this value is compared against a known officer X, fuzzywuzzy's `partial_ratio` function

$$\texttt{fuzz.partial\_ratio}(\text{“BROWN”}, X) = \max_{\{X_n\}} \frac{2n - \text{Levenstein}(\text{“BROWN”}, X_n)}{2n}$$

where $\{X_n\}$ is the set of all length n substrings of X, and Levenstein(“BROWN”, X_n) is the Levenstein distance. From here it is possible to simply choose the office whose name is the best fit.

3.2.3 Data Validation

An important final step in this process is to validate the data. If one is using an OCR engine like Tesseract, a good place to begin is with the page level confidence scores returned at the text extraction step. Tesseract, and it's Python based wrapper Pytesseract, predict words as a two-pass process: first it passes through the page and matches identified word-blobs with words in a known dictionary, these words are then passed to an adaptive classifier as prototypes and classification continues in this manner down the page, the second pass goes through the full page a second time with the fully trained adaptive classifier. The confidence of a word level prediction is then the product of the normalized outline length of the word-blob and the the normalized distance between the word-blob and the prototype [31]. In this way, the page level confidence can be taken as the mean of all word level confidence scores on a page. Another option is to validate the generated data against some known ground truths. For example, checking that numerical log numbers are in sequential order, verifying that the parsed datetimes when calls are received chronologically preceeds the parsed datetimes when calls are dispatched, en route, and cleared. Finally, if time permits, the data can also be directly validated by a human comparing random samples.

3.3 Exploratory Data Analysis

With processed and validated data in hand, the next step is analyzing the data. In what follows we present some methods that we consider to have broad utility. At this stage of the analysis, there are typically more questions than answers. Hence, we lay out frameworks for analysis that researchers might use in their own towns.

3.3.1 Natural Language Processing

In policing call logs that include large amounts of free-form narrative data, it can be advantageous to identify themes in the narratives so that each theme may be understood separately. Topic modeling is a promising approach toward this end. Depending on the data at hand, thematic organization may already be included to some degree. For example, the Williamstown data includes call-type labels for each call log, such as "building check," "public service," or "motor vehicle stop." These themes can be fairly broad, however, with

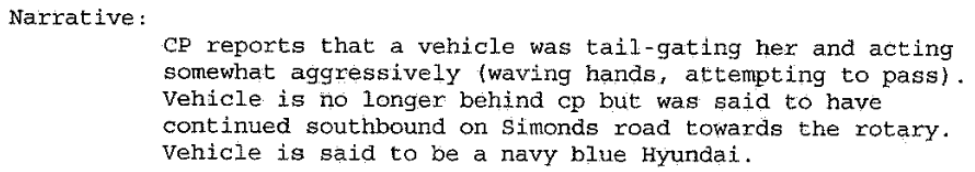


Figure 7: Sample narrative text from Williamstown PD call logs

"public service" containing narratives which range from helping a child assemble their bicycle to escorting a citizen home after they received threats of violence. Topic modeling has the potential to provide deeper and more granular understanding than what is readily available in the data (e.g. call types) or to find relationships humans might not immediately recognize.

Topic models partition a collection of documents into groups based on latent statistical word and phrase patterns. Intuitively, given a human-defined topic, say "vehicles," we associate certain words with that topic (e.g. "SUV," "plate," "driver,"). Documents written by humans can be intuitively thought of as a mixture of topics: we might think of a call log requesting occasional well-being checks during a power outage due to a medical issue as a mixture of topics like "well-being checks", "utilities", and "health". A topic model can identify possible "topics" based on which words occur together in documents (given a sufficiently large body of documents with sufficiently robust vocabulary). Unlike a human who can represent topics as an abstract category with an explicit name, a topic model represents topics without an explicit label, typically as a set of words that frequently co-occur or as a clustering of documents that use similar vocabularies. A topic model may cluster words into word clusters, documents into document clusters, or both. Documents which consist of a similar mixture of human-recognizable topics are expected to be clustered together by a topic model. A topic model with ability to finely distinguish between many topics could, in theory, divide the broader call-type label "motor vehicle stop" into more precise topics, such as stops for speeding, stops for DUIs, or stops for broken taillights. This machine-assisted topic detection could significantly reduce the amount of time researchers spend reading complicated documents and could detect topics humans reasonably cannot in extremely large document collections. In the remainder of this section we will walk through several examples of topic modeling carried out with the Williamstown data.

To reduce noise and improve topic cohesion in the topic modeling process, the following preprocessing steps were implemented: non-alphanumeric characters were removed; words of length 2 or less were removed; the text was tokenized (i.e. separated into individual words); and words were lemmatized (i.e. multiple forms of the same word reduced to one form) using the Natural Language Toolkit [3]. Lemmatizing allows words like "fires" and "fired" to be combined together with their basic grammatical stem "fire." Then, words that only appeared in one document were removed, and documents were converted from lists to sets of words (i.e. duplicate words in the same document were removed). About one-third (6,481) of the 19,975 Williamstown police logs were removed during preprocessing, leaving 13,494, the lengths of which are summarized in Figure 8. Most of the removed logs had empty narratives in the primary data sets, consisted of a single character, or consisted only of the words "narrative" and "done". A few contained typographical or OCR errors that were removed due to a low number of occurrences (e.g. "cheeked" instead of "checked"). Of the remaining logs, 30% were of length 1 word, and an additional 18% were of length 2. Extremely short narratives of length 0, 1, or 2, which are predominant in the data set, offer little or no information about what occurred in a police officer's action. The efficacy of topic modeling is limited by the policing practice of writing extremely short narratives, if any at all. This feature of police record-keeping makes it difficult for any reader, machine or human, to understand what occurred in a police action.

Even disregarding extremely short logs, the short average length of call log narratives could present an obstacle: popular models which use statistical approaches to determine word co-occurrence patterns are limited by data sparsity. *Short text topic modeling* algorithms designed specifically for short text documents have emerged to address this issue. For a review, taxonomy, and qualitative analysis of such short text topic models see [19] and references therein.

We modeled the topics in the Williamstown Police logs using a hierarchical stochastic block model (hSBM) [11, 14] and the Gibbs Sampling algorithm for the Dirichlet Multinomial Mixture model (GSDMM) [41]. The hSBM represents a state-of-the-art topic model that avoids many of the pitfalls of other popular models. For example, the popular Latent Dirichlet Allocation (LDA) model is limited by its tendency to overfit and its inability to

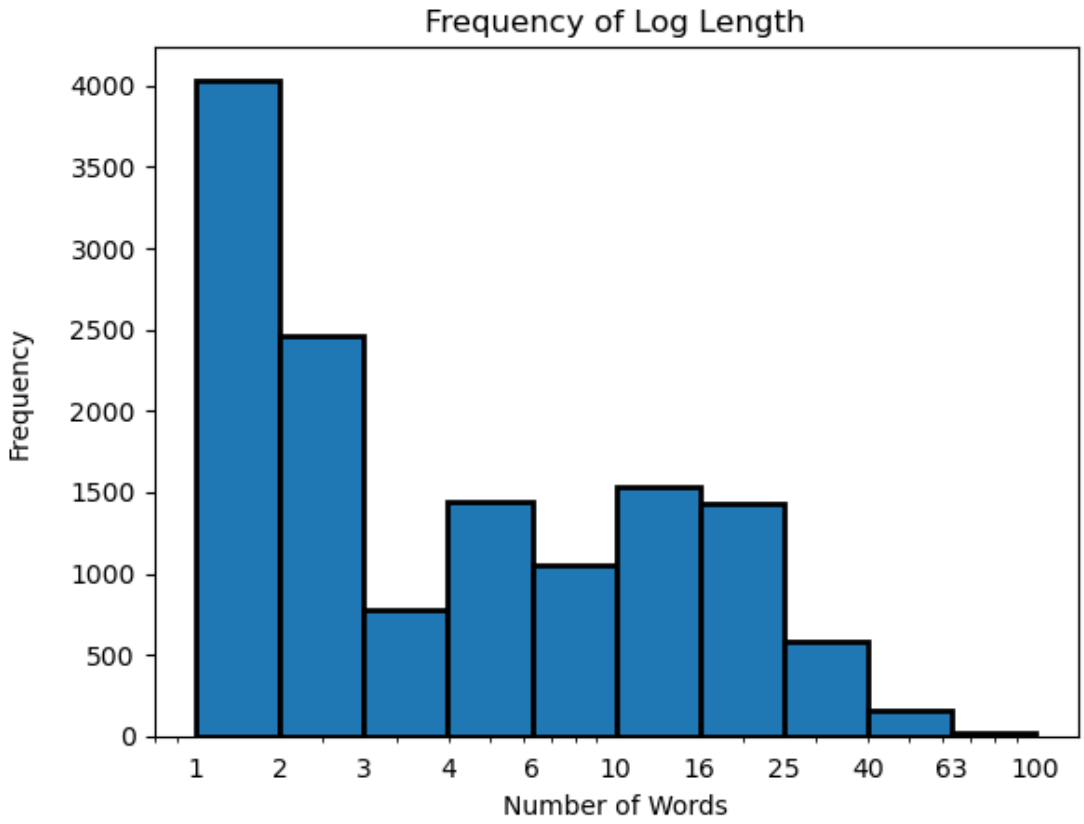


Figure 8: About one-third of the Williamstown police logs were removed during preprocessing, and 48% of the additional logs had fewer than 3 words.

capture complex, heterogeneous structure that does not follow assumed statistical distributions (see e.g. [11] and references therein). An hSBM represents a document collection as a bipartite network of documents and words so that the problem of inferring topics becomes a problem of inferring communities in a network. Community detection methods from network analysis, which have few assumptions about the structure of the underlying data, are then used to identify word clusters (topics) and document clusters. On the other hand, in order to model topics in short text documents, GSDMM makes the assumption that each document is associated with only one topic. After random initial assignment of documents to clusters, document clusters are generated by iteratively reassigning each document to the cluster with the most similar vocabulary to that document. An excellent explanation of the mechanics of GSDMM can be found in [41]. The hSBM was considered because of its status as a state-of-the-art model, while GSDMM was considered because of

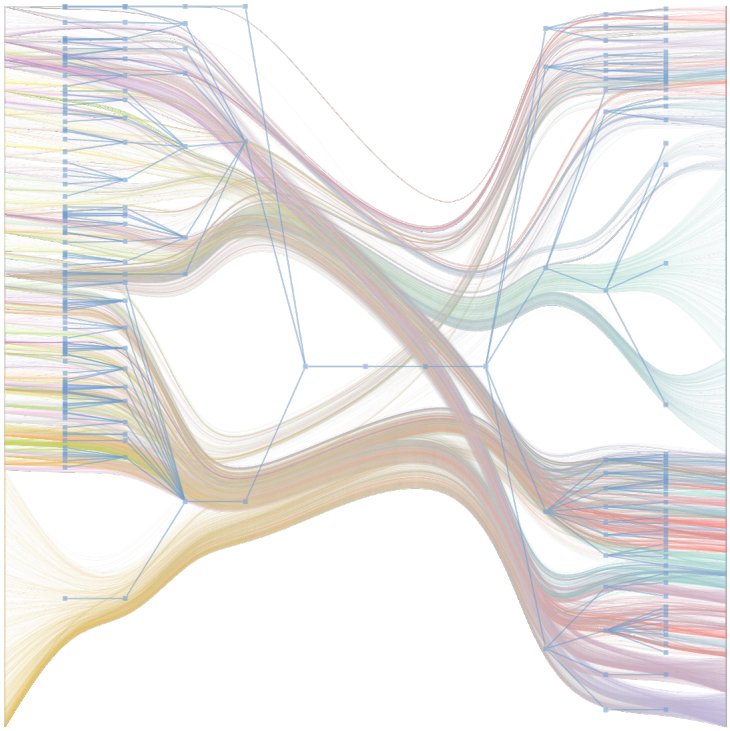


Figure 9: A visualization of the network generated by the hSBM for the Williamstown data set, with document clusters on the left and word clusters (topics) on the right. The largest document cluster (yellow on the left) consists of building and vehicle checks. The blue graph overlaid illustrates the hierarchical community structure detected by the model: on the coarsest level, the nodes are divided into words and documents. Deeper levels of the community structure (closer to the left and right sides of the diagram) represent the smaller, more specific groups detected at lower levels of the hierarchy.

the potential of its short-text specific approach.

For the Williamstown police logs, the hSBM and GSDMM document clusters exhibited a great deal of overlap. For example, there was a large number of logs written for simple building or vehicle checks. The narrative for this category of police action usually consisted of the word "checked," or perhaps "checked/secured" or "checked area." Both models identified a cluster of several thousand logs of this type, with significant overlap between the hSBM cluster and the GSDMM cluster (see Table 3).

Additional common themes were identified by both topic models in the police logs, as

Table 3: Common Topics detected by models with associated cluster sizes (in number of logs).

Topic Theme	hSBM	GSDMM
Simple building and vehicle checks	5,940	5,753
Misdials, telemarketers, and test calls	535	439
Obstruction of roadways	644	541
Disabled motor vehicles	–	533
Other (various clusters)	6,375	6,228
Total	13,494	13,494

Table 4: Narratives that share no words in common but were correctly placed in the same cluster by both models.

`Narrative: Roads snow covered and becoming slippery.`
`Dispatch to contact [`*`name redacted`*`]. Mass DOT is already out.`
`Narrative: [`*`name redacted`*`] Advised.`
`Narrative: Reports of a tree on the wires, possible downed wires.`
`Narrative: Nat Grid on scene.`

summarized in 3. These include misdials, telemarketers, and test calls (a single cluster) and obstruction of roadways. The category on obstruction of roadways was robust enough that similar logs were clustered together even if they did not share many vocabulary words. For example, the narratives that follow in Table 4 share no words at all after preprocessing, yet were correctly clustered together by both models. These words have a high rate of co-occurrence in the high dimensional vocabulary space, but also seem to share some common themes related to adverse weather incidents.

These results on topic modeling for call logs have provided interesting preliminary information on how the Williamstown community is being policed, beyond information that is easily gleaned from the provided categorization of calls by the PD. This work could be expanded to other small towns and cities in the analysis of log descriptions or other text fields, such as the description of use of force incidents.

Moving forward, one should be aware of several issues that may arise when using topic modeling on any text data, and especially short text data. If a topic model worked exactly as intended, each document cluster would represent a particular latent topic or distribution of latent topics. Because of data quality issues and random noise, however, at least two inaccuracies can occur: a latent topic can be distributed among several clusters in a nonsensical way, and a cluster may group together multiple unrelated topics. In practice, both of these issues occur simultaneously to differing degrees.

Since clustering is an unsupervised method, the connection between resulting clusters and topics of interest isn't always one-to-one. For example, in the GSDMM model, there were two distinct document clusters we characterized as logs related to "wild animals" including bears, raccoons, possums, deer, etc. (the GSDMM model also identified a separate topic for logs related to domestic animals like dogs and cats). It is worth noting that the hSBM model successfully grouped these two sets of documents together into one larger cluster. On the other hand, hSBM failed to group some clusters that the GSDMM grouped successfully, such as "disabled motor vehicles," a group which was instead distributed among several document clusters related to motor vehicles.

Across all clusters, frequent overlap between topics (e.g. two overlapping topics about vehicles and traffic) made topics difficult to interpret. For example, the GSDMM "wild animals" cluster referenced above, while largely coherent, also contained a number of unrelated logs that had to do with needles and syringes. In addition, logs that did not contain enough information to be identified with a latent topic seemed to be randomly distributed into other existing clusters. For example, narratives consisting of only the words "daughter home" were grouped together with "suspicious behavior" and "trespassing" in the GSDMM model, and with "public records requests" in the hSBM model. The fact that clusters contained some degree of narratives irrelevant to the main theme (we term "noise"), decreased the overall cohesion of each cluster. We have also observed such noise could have been coherently grouped in a new cluster, if it were to be introduced. It is worth noting that in some topic models, the number of resulting clusters is specified as a hyperparameter. However, an advantage of hSBM and GSDMM is that the exact number of clusters does not need to be specified beforehand, though other hyperparameters may implicitly affect

the number of clusters discovered. Therefore, any resulting clustering is a result of the choice of algorithm or the nature of the corpus of text; not because there were too few clusters specified to reflect all latent themes. Producing high quality clusters in the domain of limited-length narratives in the context of police logs is of ongoing interest to us.

The fact that both models succeeded and failed in similar ways suggests that data quality is the most impactful source of inaccuracies in clustering, and it would be worthwhile to try to understand and address precisely the data attributes that limit the effectiveness of each topic model. For example, human labeling, and removal of extremely short logs, could prevent those short logs with little contextual meaning from being sorted randomly into clusters. A more sophisticated lemmatization step – one capable of distinguishing different denotations of a word, such as "parked" car compared to green space "park" – would improve identification of word co-occurrence. Finally, improving the OCR and parsing segments of the data pipeline would decrease miscategorization in the detected narratives and therefore decrease noise in the final topic clusters. Approaches such as these would improve data quality and thereby cluster cohesion.

3.3.2 Spatial Analysis

Spatial information can be useful to understand the distribution of policing events, such as arrests or use of force incidents, especially when compared to socioeconomic and/or demographic characteristics of communities. For example, community members and activists may be interested in determining if some neighborhoods are being highly policed or subject to high rates of violence force by police. The level of detail in spatial information that is made public varies by town due to factors such as data collection processes and privacy processes in data release. For example, Williamstown releases primarily the name of the road where a given event occurred. Releasing a single street name for an incident provides little useful spatial information. This is particularly problematic for rural areas, because roads can be quite long, which makes statistical inference via spatial aggregation difficult. Williamstown in particular releases some events with "spatial codes," which give more precise spatial information such as the name of a specific location (a numbered address, or building name), but only 29% of events in logs have these spatial codes. The meaning

of these spatial codes is also not obvious (ex: `WIA_123`) and a table mapping codes to locations required a follow-up request from the police department.

Another nuance of spatial analysis is privatization methods. It is a relatively common practice for cities to spatially privatize the data before making information available; most commonly by jittering GPS coordinates, adjusting addresses, or by redacting some information from the address of an incident. As an example which we have studied, the police department in Minneapolis, Minnesota appears to spatially privatize their latitude/longitudes that are released to the public, by moving all events to the center of a given block. We illustrate this in Figure 10 where there are many locations of police use of force incidents that are all coded at the same block centroids. Although information on the privatization process was not found to be publicly available, this practice is evident when plotting the police use of force data from Open Minneapolis [23]. Although this is a relatively mild example of spatial privatization in police data, people studying policing data should be cognizant that spatial privatization steps that are not declared may obscure accountability efforts.

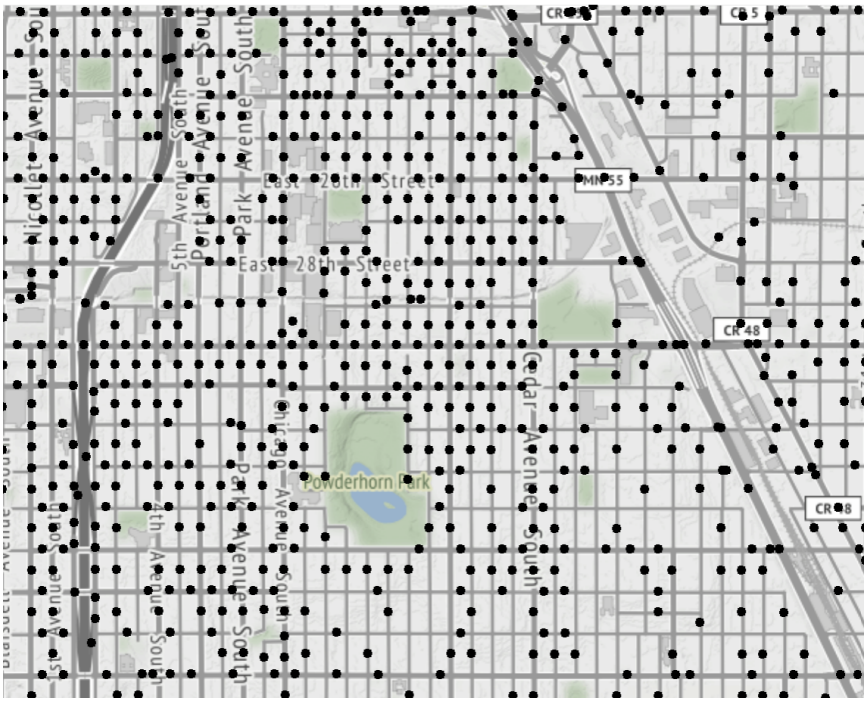


Figure 10: Minneapolis Police Department Use of Force Data from the 'Open Minneapolis' Initiative for an area of southern Minneapolis [23].

Rather than releasing privatized or partially redacted spatial data, it is more common for police departments to release data on a more aggregated level, such as the police beat/district of an event.In some cases, the exact location of an original arrest/event is still obtainable. For example, in Durham we have complete address information in the arrest records that can be geocoded to obtain exact latitude/longitude coordinates. This makes precise geocoded information readily available, removing many of the complications of spatial statistical analysis.

We note that even when it seems there are exact addresses available, data processing can reveal further issues. For example, we found the Durham data set coded many incidents as taking place at the police headquarters. Datasets from cities or towns may have a similar "default" location near the center of the municipality or county. These locations may not be of interest, given that coding an event to be located at the policing headquarters may not accurately describe the policing dynamics of a city (for example, where police officers are patrolling).

Generally in the spatial analysis of event information, there are two common modeling frameworks: areal and point process. We present a very simple summary of these two approaches to illustrate the challenges in spatial analysis of incomplete or anonymized spatial information that is often released in small-town policing data.

$$Y(s) = X(s)'\beta + e \tag{1a}$$

$$\lambda(s) = \exp(X(s)'\beta), \tag{1b}$$

Equation 1a, represents an "areal" approach which relies on aggregate analysis on spatial levels such as counties, census tracts, or census block groups (which we call areal units) as a function of abstract location s. The count of events per areal unit (or a rate per population), $Y(s)$, is related to the spatial covariates, $X(s)$, (e.g. census socioeconomic/demographic information) and the error distribution, e, typically has spatial structure or dependency. Equation 1b illustrates the general framework of point process modeling, in this case via a nonhomogeneous Poisson process. The spatial "intensity," $\lambda(s)$, indicates generally how likely events are to happen at a given location, s. This depends again on some spatial information, $X(s)$, at that location s, which is oftentimes again census information at areal units. In both models, we are interested in inferring or learning about regression

coefficients, β, as well as other parameters that may be incorporated in more complex versions of these models.

There are many statistical challenges presented through spatially uncertain geographic point locations. For example, when only street segments are released, analysts may consider population-weighted placement of events along the streets to analyze events spatially, or it may be necessary to avoid spatial analysis of events altogether. When blocks are released (e.g. '12XX Main Street'), similarly one may consider only analyzing the data based on areal units (such as block groups through Equation 1a) rather than point process (event-level) analysis. Additionally, an assumption of spatial point process models is that at most one point of the process is observed at each location, which means that there is a fundamental problem when spatial anonymization is pursued where points are placed in the same (anonymized) location, such as the center of a block as in the city of Minneapolis. Minimally, these points must be spatially jittered in some way to try to approximate the original locations, but areal-level analysis may be more appropriate in some cases.

In the rare case that the exact address is released, the analyst must convert the address to a latitude/longitude point for use in statistical analysis in a process called geocoding. The process of geocoding introduces positional errors, where points are not always placed correctly at their true location. This is particularly problematic in rural settings where positional error can be quite large (and therefore is a topic of concern for small towns) [18]. If exact latitude/longitude coordinate values are released by the police department, the analyst should be aware of potential anonymization applied to the data set by the police department that is not disclosed to the user. If the coordinates are exact, *which is rare*, or if geocoded exact addresses are used, the analyst must be conscious of the need to protect the privacy of the original location. These addresses may represent sensitive information, such as home addresses of the people involved in the policing event. Spatial privacy methods may be considered to protect the original location (for example, when plotting the dataset) without having too large of an impact on the statistical utility of the perturbed data set. When spatial information has been modified in some way before analysis, one must be very cautious of interpretation of regression coefficients, β, when using spatial models to analyze the relationship between community characteristics and policing outcomes.

In summary, statistical analysis of the spatial features of small-town policing datasets is often difficult. Largely, these difficulties arise from lack of data availability: spatial information is either completely missing, or is redacted in some way. It is important to be careful in the analysis of these redacted locations and to not treat them as if they were the original data locations. However, even anonymized location information can help communities learn about the policing dynamics in relation to other spatial information.

4 Discussion and Recommendations

One of the primary challenges in studying small town policing is the availability of high quality data. As we've seen in our work, larger and more resourced towns such as Durham are able to make data available online in a format that can be downloaded and parsed. Smaller towns like Williamstown may not currently have the necessary infrastructure to make this possible. In many cases, the barriers presented here– repeated records requests, low quality data, the necessity of a technically challenging digitization pipeline, and uncertain quality of the resulting data– can prove insurmountable to individuals working alone at the local level. Our hope is to help lower the barrier to entry by working alongside these individuals and creating well-documented and easy-to-replicate tools and processes in our public repository `https://qsideinstitute.github.io/SToPA/`.

One end goal of much of this work is to make processed, cleaned, and validated data available for analysis. Depending on the nature of the town and the data, the analyses of these resulting datasets can go in many directions. Sometimes descriptive statistics (e.g. how many arrests occurred in each census block group and what were the primary reasons) is already of great interest to community activists. If the data quality and quantity is sufficient, there is also opportunity for analysis using more robust statistical methods such as clustering, classification, and regression. For example, it is possible to model response times between 911 calls and officer arrival on the scene in a manner similar to analysis done on fire department turnout times [2]. Alternatively, one could model wait times between calls initiated and calls cleared using frameworks similar to those that have been used to analyze hospital emergency department wait times [7]. One might also be interested in modeling the relationship between community characteristics and policing outcomes in

those neighborhoods to understand potential biases or patterns in policing behavior.

Often the questions we most want to ask about bias, such as the effect of race on policing incidents and outcomes, cannot be directly modeled because race data is not recorded. This brings up the need for methods that allow for the estimation of individual or group level statistics from aggregate data, such as with ecological inference and Bayesian simulation. For example, an important question to ask is whether different communities are being policed in different ways (i.e. is there racial or socioeconomic polarization in the rates of policing). Recently, [12] attempts to answer this type of question using a Bayesian modeling to estimate posterior probabilities of racial polarization using Markov Chain Monte Carlo (MCMC) sampling. Methods such as this allow researchers to achieve reasonable estimates for demographic distribution among log subjects when that information is not directly available in the data.

For police accountability efforts to be successful and impactful, a critical component is the participation of citizens, activists, and researchers across domains. We are, primarily, mathematicians, statisticians and data scientists with an interest in social justice. But working with local activists in the Williamstown area has lent insight into artifacts in the data, such as historical racial tensions in specific neighborhoods, that could not possibly have been visible to outsiders like us. Additionally, it is important that we bring expertise from researchers in the sociology of policing to understand the patterns and norms in policing across the United States to call attention to the many accepted but undocumented practices in policing. Additionally, researchers in urban policy and planning can lend insight into the ways that cities and places shape behavior and the way that human behaviors shape their environments. Individuals working in law enforcement and the criminal justice system also carry knowledge that could be helpful in interpreting our work and placing more targeted data requests.

For small-town communities seeking insights into the way that their municipalities are policed, this work is intended to provide a launching point for discussions of local law enforcement. Whether the goal is to reinforce trust that members of the community already hold, to regain confidence in a department where trust has been lost, or to help start conversations to more progressive change, we believe that transparency and accountability

are vital.

This work equips the community and the local law enforcement agencies with an opportunity to discuss highly sensitive topics grounded in structured findings. In the case that there are distinct facets of policing that community members would like to see improved, results from analysis of policing data can be brought to the attention and/or discussed with the local policing agencies themselves and the municipality executive members.

For local law enforcement agencies facing challenges regaining the confidence of the communities they serve, they can seek assistance from The United States Department of Justice (DOJ) Community Relations Service (CRS). The CRS, through their Strengthening Police and Community Partnerships (SPCP) process [34], can facilitate community-led discussions. These conversations are intended to platform community leaders and members from diverse backgrounds by targeted outreach. An advantage of the SPCP process is that it can be quickly adopted by the local police since it operates from within the DOJ and therefore requires no additional mechanisms for implementation.

In more extreme cases, the findings can augment disciplinary cases against some law enforcement personnel. Often, the municipality executive members and police leadership have challenges of their own in dealing with police unions that protect substandard law enforcement officers. Community members can use the findings to address the facets by petitioning their grievances to elected and appointed municipal leaders. Regardless of the approach and/or the predisposition of community members, the methods outlined in this research can aid community members to better understand how they are being policed and empowers them to approach this data acquisition and analysis process with both guidance and concrete tools in-hand.

References

[1] Joshua L Adams. "I Almost Quit": Exploring the Prevalence of the Ferguson Effect in Two Small Sized Law Enforcement Agencies in Rural Southcentral Virginia. *The Qualitative Report*, 24(7):1747–1764, 2019.

[2] Madison Arnsbarger, Joshua Goldstein, Claire Kelling, Gizem Korkmaz, and Sallie

Keller. Modeling response time to structure fires. *The American Statistician*, 75(1):92–100, 2021.

[3] Steven Bird, Edward Loper, and Ewan Klein. Natural language processing with Python. *O'Reilly Media Inc.*, 2009.

[4] Francis D Boateng and Joselyne Chenane. Policing and social media: A mixed-method investigation of social media use by a small-town police department. *International Journal of Police Science & Management*, 22(3):263–273, 2020.

[5] Michelle A Bolger and P Colin Bolger. Predicting fear of crime: Results from a community survey of a small city. *American Journal of Criminal Justice*, 44(2):334–351, 2019.

[6] D. Brian Burghart. Fatal encounters, 2023. Last Accessed 2023-02-27.

[7] Bill Cai and Iris Shimizu. Negative binomials regression model in analysis of wait time at hospital emergency department. In *Proceedings. American Statistical Association. Annual Meeting*. NIH Public Access, 2014.

[8] Campaign Zero. Mapping Police Violence, 2023. Last Accessed 2023-02-27.

[9] Durham Economic Development. Community Data.

[10] Durham Police Department. Durham Police to Citizen Portal. `https://durhampdnc.policetocitizen.com/eventsearch`, 2023. Accessed 2023-01-02.

[11] Martin Gerlach, Tiago P. Peixoto, and Eduardo G. Altmann. A network approach to topic models. *Science Advances*, 4(7):eaaq1360, 2018.

[12] Anna Haensch, Daanika Gordon, Karin Knudson, and Justina Cheng. A multi-method data science pipeline for analyzing police service in the presence of misconduct. *SocArXiv*, November 2022.

[13] John Harrison. *RSelenium: R Bindings for 'Selenium WebDriver'*, 2022. R package version 1.7.9.

[14] Charles C. Hyland, Yuanming Tao, Lamiae Azizi, Martin Gerlach, Tiago P. Peixoto, and Eduardo G. Altmann. Multilayer networks for text analysis with multiple data types. *EPJ Data Science*, 10(1):33, Jun 2021.

[15] Invisible Institute. Citizen Police Data Project, 2023. Last Accessed 2023-02-27.

[16] Josh Landes. Williamstown police department admits to illegally searching critics' records. *WAMC Northeast Public Radio*, Mar 2021.

[17] Josh Landes. Independent investigations of Williamstown Police Department detail systemic failures at "at many levels," confirm sexual misconduct, racism claims. *WAMC Northeast Public Radio*, Jan 2022. https://www.wamc.org/news/2022-01-28/independent-investigations-of-williamstown-police-department-detail-systemic-failures-at-at-many-levels- confirm-sexual-misconduct-racism-claims.

[18] Shengde Liang, Bradley P Carlin, and Alan E Gelfand. Analysis of Minnesota colon and rectum cancer point patterns with spatial and nonspatial covariate information. *The Annals of Applied Statistics*, 3(3):943, 2008.

[19] Belal Abdullah Hezam Murshed, Suresha Mallappa, Jemal Abawajy, Mufeed Ahmed Naji Saif, Hasib Daowd Esmail Al-ariki, and Hudhaifa Mohamme Abdulwahab. Short text topic modelling approaches in the context of big data: taxonomy, survey, and analysis. *Artificial Intelligence Review*, 56:5133–5260, 2023.

[20] National Freedom of Information Coalition. State Freedom of Information Laws. `https://www.nfoic.org/state-freedom-of-information-laws/`, 2023. Accessed 2023-02-09.

[21] National Policing Institute. Police Data Initiative. Last Accessed 2023-02-27.

[22] Jeroen Ooms. *tesseract: Open Source OCR Engine*, 2022. https://docs.ropensci.org/tesseract/ (website) https://github.com/ropensci/tesseract (devel).

[23] Open Minneapolis.

[24] Stephen S Owen, Tod W Burke, April L Few-Demo, and Jameson Natwick. Perceptions of the police by LGBT communities. *American Journal of Criminal Justice*, 43(3):668–693, 2018.

[25] Lynne Peeples. What the data say about police brutality and racial bias—and which reforms might work. *Nature*, 583:22–24, 2020.

[26] Emma Pierson, Camelia Simoiu, Jan Overgoor, Sam Corbett-Davies, Daniel Jenson, Amy Shoemaker, Vignesh Ramachandran, Phoebe Barghouty, Cheryl Phillips, Ravi Shroff, et al. A large-scale analysis of racial disparities in police stops across the United States. *Nature Human Behaviour*, 4(7):736–745, 2020.

[27] Seat Geek Inc. *fuzzywuzzy: Fuzzy String Matching in Python*, 2014.

[28] Lawrence W Sherman. Reducing fatal police shootings as system crashes: Research, theory, and practice. *Annual Review of Criminology*, 1:421–449, 2018.

[29] Jeremy Singer-Vine and The pdfplumber contributors. pdfplumber, 11 2022.

[30] Samuel Sinyangwe. Police Scorecard, 2023. Last Accessed 2023-02-27.

[31] Ray Smith. An overview of the tesseract ocr engine. In *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*, volume 2, pages 629–633. IEEE, 2007.

[32] Scott Stafford. Lawyers group, Pittsfield NAACP branch seek termination of Williamstown, Mass. police officer. *The Bennington Banner*, Aug 2021.

[33] Scott Stafford. Reports confirm hostile work environment at Williamstown Police Department, lay the blame on former chief and former sergeant. *The Berkshire Eagle*, Jan 2022.

[34] The United States Department of Justice. Strengthening police and community partnerships, Dec 2022.

[35] Amelia Thomson-DeVeaux, Laura Bronner, and Damini Sharma. Cities spend millions on police misconduct every year. here's why it's so difficult to hold departments

accountable., Feb 2021. https://fivethirtyeight.com/features/police-misconduct-costs-cities-millions-every-year- but-thats-where-the-accountability-ends/.

[36] Amel Toukabri and Lauren Medina. America: A Nation of Small Towns. *United States Census Bureau*, May 2020.

[37] Town of Williamstown Massachusetts. 2020 Annual Town Report. `https://williamstownma.gov/wp-content/uploads/2021/05/ATR-FINAL-1.pdf`, 2020. Accessed 2023-01-25.

[38] U.S. Census Bureau. Quickfacts: United States. `https://www.census.gov/quickfacts/fact/table/US/PST045222`, 2023. Accessed 2023-01-02.

[39] Rachael A Woldoff and Karen G Weiss. Studentification and disorder in a college town. *City & Community*, 17(1):259–275, 2018.

[40] Sue-Ming Yang, Charlotte Gill, L Caitlin Kanewske, and Paige S Thompson. Exploring police response to mental health calls in a nonurban area: A case study of Roanoke County, Virginia. *Victims & Offenders*, 13(8):1132–1152, 2018.

[41] Jianhua Yin and Jianyong Wang. A Dirichlet multinomial mixture model-based approach for short text clustering. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14, page 233–242, New York, NY, USA, 2014. Association for Computing Machinery.