CPDP Project Report

The Wise Lobsters

Theme

Identifying allegations early in a repeater officer's career and exploring patterns that lead to increasing allegations over time.

Introduction

Using the data from the Citizen Police Data Project (CPDP), We aimed to find components in a police officer's early career that lead to misconduct allegations and explore possible areas of interest that signify future repeater behavior. We examined areas of the data to determine relationships in the officers and their repeating type of allegations. Along with exploring different aspects of the data, we utilized different approaches to database management and exploration to aid in our search. This project served as a valuable data science learning experience. We tackled answering our questions about the data through means of relational analytics, visualizations, integration of external data, graph analytics, natural language processing, and machine learning. Our focus areas throughout this project are officer's first/early career allegations, allegation categories, repeater identification, and the metric for years on the force.

We believe our most prominent findings were in the exploration of the data itself, including visualization, graph analytics, and a new method of document tagging. We visualized allegation categories and how officers tend to stack up categories over time. Integrating settlement and arrest data, we explored how both relate to the count of allegation. With graph analytics, we constructed a co-accusal network with officer's first allegation with repeaters. Lastly, we explored a new way of tagging allegation documents. This paper will discuss our process for exploration and how we attempted to answer questions found in our initial project proposal.

Relational Analytics

Through our initial exploration of the database provided by the Invisible Institute, we started to frame our data for our problem space. We started by exploring how we filter officers and allegations in order to have a consistent window into an officer's early career. We did this by segmenting officers' age, years on the force, count, and types of allegations over time. Knowing that the Citizens Police Data Project (CPDP) dataset only has a complete allegation collection from 2000 onward, we started our subset their. Since we were looking at an officer's early career, we needed to make sure we were also able to see enough of their career to know their later behavior. For this, we picked the upper bound appointed date of December 31, 2007, to ensure that we had at least ten years of data for every officer in our subset. The dataset ends in early 2018 at the time of this writing. We also filtered to make sure that all officers in the subset are active at the last data extraction. We wanted to focus on behavior that allowed officers to repeat allegations and stay on the force for a large portion of their career.

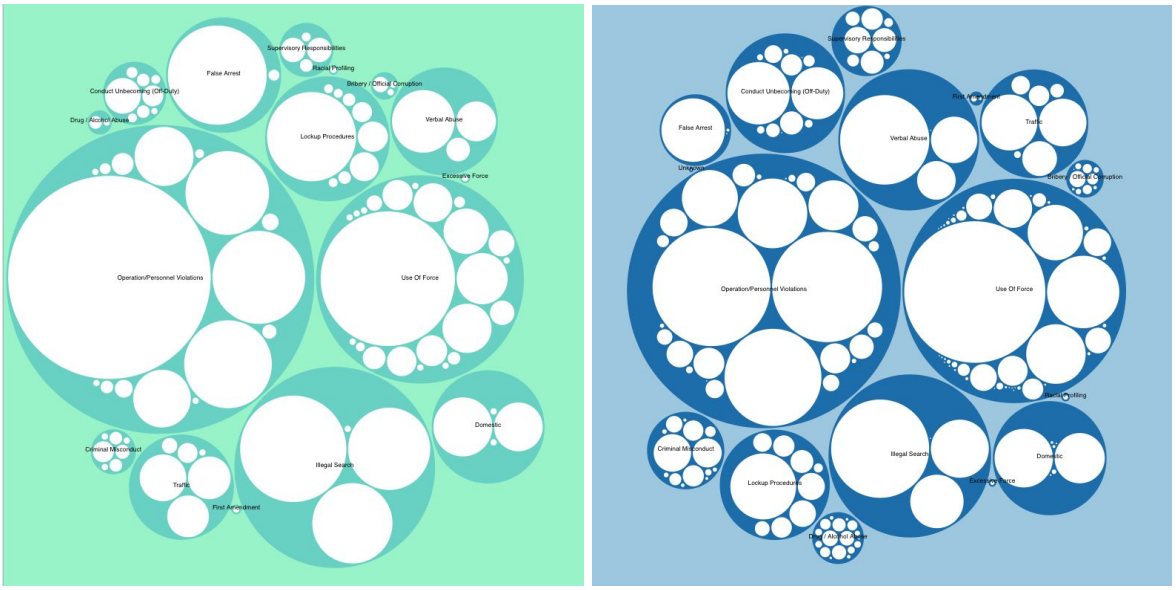
We made assumptions about how to identify repeaters. Generally, we considered ten severe allegations to mean a repeater, but the threshold can depend on the context. We also explored allegation percentage and number of total allegations over time.

Next, we identified an officer's first allegation, another metric we used throughout this project. Initially, we did little filtering and just identified absolute first allegation, but later filtered this on category and time. We also limited allegations to civilian reported allegations, this is a better focus for our theme, but also, police-reported allegations are often minor. Generally, we wanted to gain insight into significant or violent allegations.

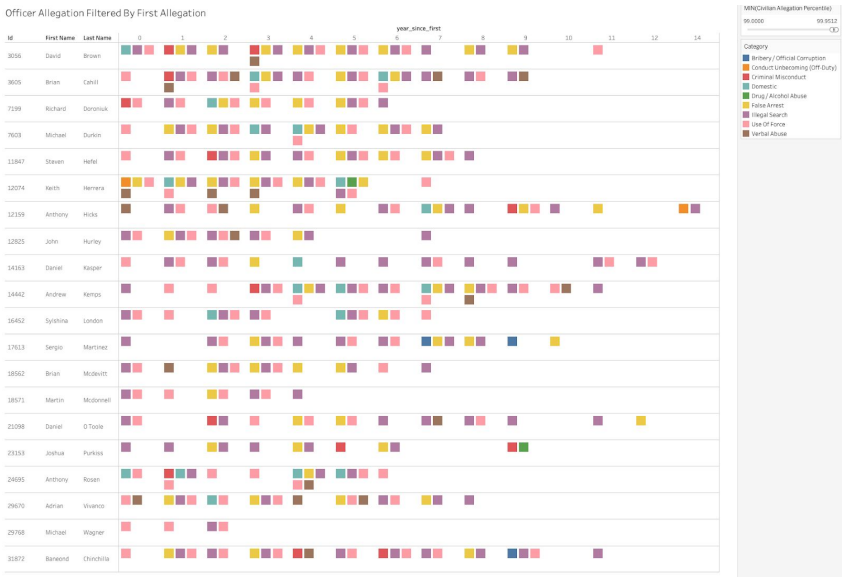
We addressed the statement to study allegations among officers during the inception of their career, studying patterns and metrics as they start in the force, exploring over time, and evaluating networks comprising of repeaters. Setting the parameters and a subset of officers from the mentioned range, we studied count of allegations over time, with an emphasis on a defined pattern of higher than ten allegations to be marked as a repeater for our research. Another split against allegations over the year has revealed insight that the increase is prominent in years two and three years of employment, with the highest volume of allegations in the first five years. This study further highlighted the average for each cluster of officer age groups as well. Our exploration showed that the officers in the 35-44 age group committed the most allegations.

The study also helped to narrow down on prominent categories, from Illegal Search to Force and Operations based allegations. Our theme explored this in further research for not only younger sections but repeated behavior and clusters formed among these officers. The queries also provided metrics and inputs like how 18.62% of allegations for first-time or early incidents are with repeaters and how this influence plays a role in more allegations later in the officer's career. These findings helped pave a path towards guided analysis when exploring the arrest dataset, outlined in a later section.

Visualization



Expanding upon the queries from the earlier section, we developed visualization to help us explore and understand the data. We developed a nested bubble chart (illustrated above, left: allegation categories --link, right: first allegation categories -- link) for civilian allegations, a Gantt chart (below) for allegations by categories over the years on the force and a histogram for allegations centered around a media event. We also developed an additional interactive scatter plot to illustrate trends in category reporting over time. The aggregate data for these graphics are generated directly from the CPDP dataset using insights gained from initial exploration. Visualizations help us see patterns in the data, but also identify parameters for future research and analysis.



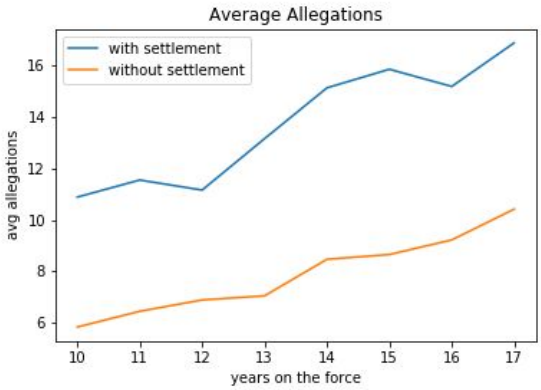
The most prominent and visually pleasing visualization were centered around allegations categories and are the ones listed above. The nested bubble charts helped us visualize what categories were overwhelmingly dominate and that we should filter some out for future analysis. Explicitly, Operation/Personnel Violations, which is the most common allegation overall, does not provide much information since it can be a variety of non-severe complaints. The Gantt chart furthers this analysis by seeing how allegations reported overtime per officer. The chart shifts allegation data per officer to a year-by-year mapping for years on the force and lists a color square for each allegation color by category. Filtering by top repeaters identified by allegation percentiles illustrates that severe allegations are prominent throughout the officer's career.

Our interactive scatter plot, while not showing us what we initially hypothesized, did result in some insights on how allegations can shift reported categorization over the years. Our initial assumption is that these categories are well defined and follow a distribution over the dataset. What we see from this visual is that categories that are abundant in one year may shift from year to year. We are unsure if this change reflects behavior, but we hypothesize that it is related to reporting behavior rather than actual officer behavior. We also attempted to explore how allegation reporting tracked around the release of the Laquan McDonald shooting video in 2017. We were hoping to find an external factor to bring into our analysis, but we did not find promising results.

Data Integration and Data Cleaning

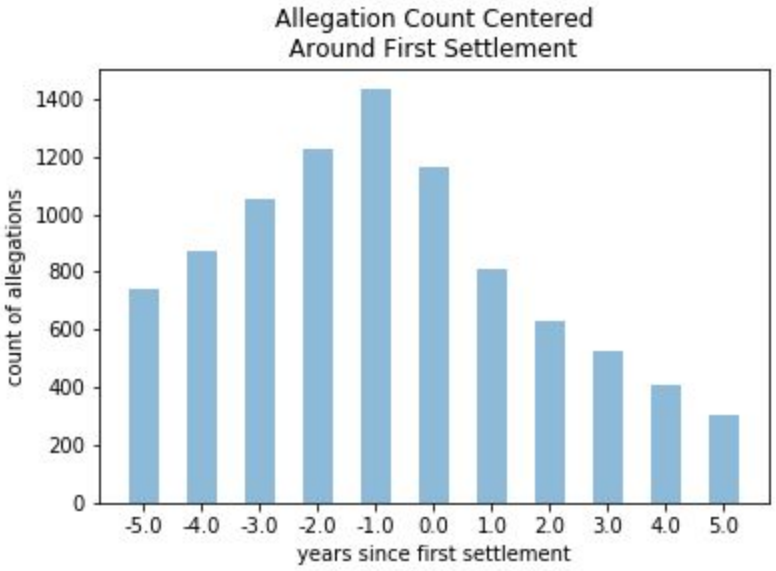
External datasets for arrest and settlement allowed further exploration of events early in an officer's career. Arrest did not have a direct link to allegations, but we were still able to perform general analysis. Settlement did relate directly to allegations, and we found some interesting trends centered around when an officer is involved in a settlement.

Settlement Data



First, we analyzed the average number of complaints against an officer with any settlement compared to officers involved in no settlements and grouped by years on the force. The trend of

this graph follows an expected pattern; officers with settlement have a much higher average allegation count than those without any settlements. It also shows that there is a slight increase in the number of average allegations of overtime for officers with no settlements. With settlements has a range from 10.89 to 16.89 (6 allegation spread, or 35% increase) compared with no settlements with to 5.83 - 10.43 (4.6 allegation spread or 44% increase). In summary, having a settlement is a valid indication of bad behavior and overall allegation count.



Second, we discovered that after a settlement, the average number of allegations decreased, see the plot above. This trend presents a potentially significant trend in the behavior of officers. From this analysis, the count of allegations trend down after the first settlement. Note that these officers are active at the time of the last data update, so there are no cases of officers leaving the force as a result of an allegation or settlement. Allegation counts steadily increase until the year a settlement is filed, then decreases at a higher rate than the increase (48% increase, 78% decrease from the high point). The assumption is that year 0 follows the downward trend given that the trend started around the time the settlement is filed, which is sometimes within year 0.

The decrease in allegation highlights two crucial points. First, the obvious, settlements tend to decrease allegations. Second, a year over year increase in allegation counts could be a good indication of a risk for a settlement. Settlements are a complicated process, but a big concern for the city and its residents. Reducing settlements mean reducing misconduct cases. If the events leading up to an event resulting in a settlement can evaluate a risk factor, it could be valuable information.

Arrest Data

One of the critical challenges with the arrest dataset is linking them to officers. Our initial assumption linked arrests to allegations, but there is no such linkage to support a confident connection. The date seems to be the only potential link, but we decided it is not enough to

conclude a connection. There can be multiple arrests or allegations in a day, and dates could be inaccurate.

In our analysis, we determined the average number of arrests per officer over years of employment and how this compares with allegation counts. There is a trend of average arrests peaking in the three to six-year mark. For allegations, we see a similar trend but slightly shifted in the range of 2 to 7 years. Although there are small variations in the peaks, it generally follows the assumption that more arrests lead to more allegations.

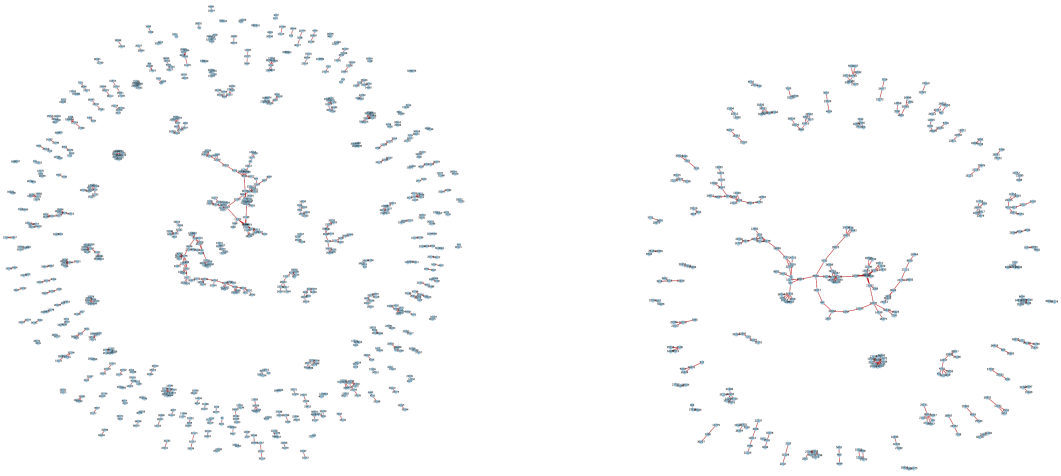
In terms of an officer's early career, we compared allegations to arrests in an attempt to explore if arrests serve as an indication of allegation count and, therefore, unsatisfactory behavior. Similar to our previous analysis, we see that more allegations correlate to more arrests. More arrests mean more interaction with civilians and more opportunities for misconduct. When looking into an officer's career, arrest counts could apply as a weighting metric for the expected count of allegations. For example, if an officer is in 90% of arrests, more miscount could be expected. Although arrests should not be the only metric as allegations are an indication of bad behavior and arrests can be the opposite.

Graph Analytics

Analyzing networks was a fascinating topic for our theme, providing new ways to view the data outside the context of relational databases. We identified two main areas to explore in this space, co-accusals and linking allegation to awards. Our social network analysis yielded exciting results that warrant further exploration. Our award analysis established awards, often occur around allegations and should not blindly count as the only indication of behavior.

We wanted to build a social network that was different from some we have seen on this dataset and focused on our theme. We created a network where the nodes are officers in our subset along with repeaters. We did a simple repeater indication of 10 severe allegation, not included a few non-severe categories. We also noted the time an officer became a repeater as a node attribute. The edges of the graph indicate an officer's first allegation co-accused with a repeater. The allegation had to happen after the repeater hit the ten allegation threshold. We were delighted with the findings in the network output.

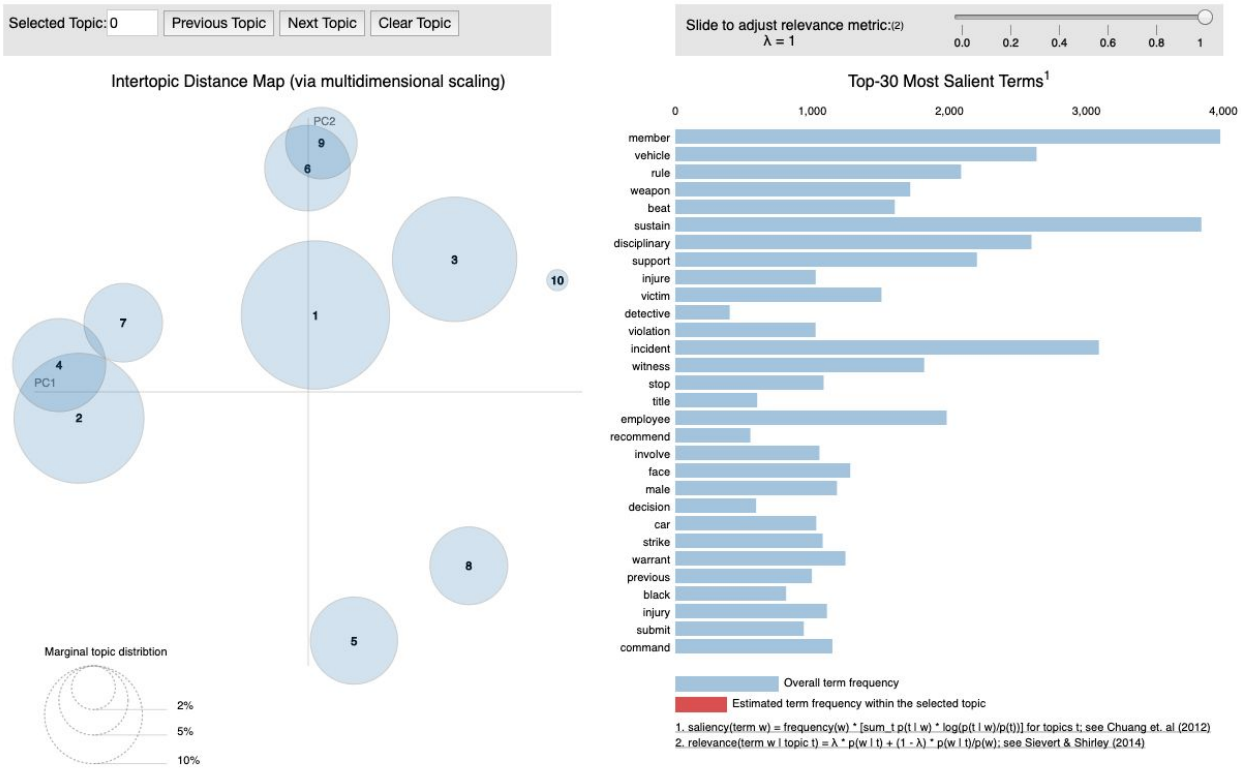
Below illustrates our network. The left depicts the entire network, and the right is focusing on top co-accusing repeaters with three-degrees pf connections. Locating top co-accusers is as simple as finding a node with the most in-degrees. Out of all of our findings, we believe this is the most important to our theme. If an officer is part of one of these networks early in their career, it could be an indication of a breeding ground of repeaters.



Our second exploration of graph analytics was constructing a different type of graph. This graph is one with few nodes and many edges, with the nodes being allegation categories and awards types. The edges are linking an allegation to an award within 60 days of the incident date. To our surprise, this was not an uncommon occurrence with thousands of instances. Unfortunately, the paths to awards and the fact that most awards are generic Honorable Mentions, we were not able to draw any conclusions to use as an indication. We go as far as to say awards add minimal insight to the analysis of officer's careers. It seems from our findings that awards and allegations are entirely separate, and often allegations are overlooked awarding good behavior.

Text Analytics and Machine Learning

With using text extracted from supplementary documents detailing allegations, our goal was to attempt to extract a scale of violence. Once we received access to the data, unfortunately, we realized that labels were far too sparse to extract a scale. We decided to train a model to observe the type of separation among the documents. To achieve this, we used a Latent Dirichlet Allocation model, an unsupervised grouping method, breaking the documents into ten topics.



	id	term1	term2	term3	term4	term5	term6	term7	term8	term9	term10
0	0										
1	1	injure	member	support	disciplinary	sustain	professional	medical	property	injury	visible
2	2	weapon	member	discharge	battery	fire	incident	force	firearm	injury	first
3	3	detective	homicide	supplementary	attorney	murder	inspector	taxi	government	emma	strip
4	4	rule	decision	follow	hear	hearing	conduct	write	guilty	suspension	agreement
5	5	go	incident	leave	male	sustain	back	strike	take	face	door
6	6	rule	violation	involve	copa	accountability	write	work	would	give	find
7	7	sustain	member	disciplinary	support	warrant	employee	incident	command	witness	file
8	8	vehicle	stop	car	drive	take	driver	individual	arrest	involve	seat
9	9	incident	witness	employee	face	audio	victim	command	preliminary	weapon	sustain
10	10	beat	victim	title	recommend	concur_recommend	black	unknown	age	event	male

We found that categorizing these documents yielded a new and in some cases, human-readable labeling of the documents. In some cases, the level of violence is noticeable, for example, with terms like homicide, weapon, and discharge. There was not enough data to link this back to an officer's first career allegations. However, we believe this is a start to providing an automated labeling pipeline or at least assisting in active learning of supplementary model.

Although we believe the model is generalizing topics within the data, there are concerning patterns when looking at how the documents separate over time. Looking at the Use of Force

and Illegal Search allegation categories, there are points in the eighteen years of the data where document topics are more prominent than others. Alternatively, in the case of Use of Force, the majority of the documents come from a specific time. We do not initially relate the problem with the model. We hypothesize that this is an inconsistency in the data provided and that the dataset is not complete. Even still, this model's pipeline continues to train as more documents are collected. An enhancement to the pipeline for future research could combine other processes to attempt to automate the labeling of the documents.

Conclusions and Future Analysis

Through our analysis, we can see recurring patterns around what early careers of repeaters look like and some assumptions about what type of allegations. We are also able to identify where the data is lacking and how assumptions can be dangerous. From what we have seen throughout this project, behavior resulting in misconduct that starts early in an officer's career tends to continue throughout their career. Of course, there are far too many variables in this problem to pinpoint why officers continue misconduct, but, admittedly, entire representation is missing lacking the dataset. We are only looking at a tiny window of information to try and infer behavior. Aspiring to build out an early risk assessment with only the data presented would not be possible. However, we do believe we highlighted some interesting findings that could be explored in the future to better piece together the story of how repeaters continue to exist in the Chicago police force.

The main goal of this project was to gain familiarity with the core concepts of data science while also extracting meaningful information out of the Chicago Police Department database. In total, we conducted eighteen separate analyses. Out of them, we believe the most promising for continuing future research is our social network of repeaters and first-time allegations and our document tagging model.

Further graph analysis on the first time offending officers and how they interact with repeaters on allegations could lead to the social aspects of growing repeater behavior. Areas we did not cover with our research, but we believe could yield results are the location of the officers, race or rank. It would be interesting to know if any of these networks span multiple beats or what the average rank distance is between the rookie and repeater. Also, comparing these networks to previous research to see if top page ranked officers appear prominent in both networks. This network is the most suitable direction we found to explore further behavior in officer's early careers. It is a start to gain insight into social patterns which are not well accounted for in the dataset.

A large portion of our analysis bases assumptions on allegation categories and allegation names. There are two big problems we see in the data with these categorizations. For one, the categories are too vague, and names are two specific or inconsistent. Additionally, there is only one name and one category associated with an allegation, with no secondary or tertiary component considered. Second, we have seen shifts in the data on how allegations are categories over time. Excessive Force and Use of Force are good examples of how categories

can overlap. It becomes hard to tell if an increase is just a reporting anomaly or an actual increase in events. Building out a more granular and consistent reporting system, if possible, could help in understanding these events further. Potentially through allegation report categorization or mapping categorization to a consistent distribution overtime.

If the allegation document is to continue to grow, it will become essential to work out an automated or partially automated tagging system. There would be many benefits to a tagging system that is accurate at identifying documents, underlying allegation categories, or scale of severity. Although, since these are inconsistent human-created documents, humans in the loop learning will likely be required. Limiting interaction with labeling will allow for relevant or information-rich documents to be read at load time. The addition of new dimensions to the database and the already existing data could serve as an early warning indication of bad behavior and warrant correction to avoid repetition.