Sarah Ahmad (saz848)
Vikram Kohli (vkp260)
Junhao Li (jlo708)

Project Report - A Look into Chicago's Police Officers

Introduction

The focus of our project was to learn about the police officers within the Chicago Police Database and in particular about the ways in which factors of identity affect them.

The interactions between police officers and civilians are shaped not just by actions and individuals but by discrepancies in social location including race, gender, socioeconomic location, and age. While many studies focus on the standpoint of the accuser or victim, little research is done with respect to the standpoints of the officers. Our goal was to figure out how factors like race and gender affected officers in areas such as implicated numbers of allegations or length of appointment.

Relational Analytics

Our initial questions largely consisted of demographic information about the officers relating to the nature of the complaints against them; whether officers are by and large alleged of similar offenses by race and gender, whether the percentages of TRRs and complaints were consistent between officer groups, and finally more generally whether the breakdown of percent complaints against officers of particular graphics matched up with the percentage of complainants.

Our first major finding was that allegations are disproportionately levied against officers on the basis of race; this differential is most starkly seen in the context of white and black officers. The complainants in most allegations are black, whereas the officer being alleged of improper conduct is disproportionately white; 59% of all allegations are from black complainants, and 61% of allegations were against White officers.

Second, we found that, for the most part, the percent of officers in complaints generally matches up with the percent involved in TRRs, including when separated into groups based on race and gender. A notable exception we found to this was with regards to black men, who were involved in 23% of allegations, but only in 13% of tactical response reports.

Finally, we found several interesting discrepancies in the types of complaints levied against officers between officers of different race and gender. Considering gender alone, women were alleged of use of force and illegal search substantially less than the average officer, but received complaints concerning personnel violations substantially more. For black women in particular, the percentage of complaints alleging unbecoming off-conduct duty was substantially higher than both officers overall and for white women / women overall, as well as higher than for black men. On the flip side, white men seem to receive allegations of racial profiling substantially more than the average officer --- and no allegations of profiling at all were alleged against Asian/Pacific or Native officers.

Although very few complaints in total alleged racial profiling, white officers seemed to receive these allegations noticeably more than other officers. No allegations of racial profiling were levied against Asian/Pacific or Native officers, though this may simply be due to their relatively small representation in the police force.

Something that was surprising about the proportion of allegations that ended up with a punishment was the same for both male and female officers, though the proportion of allegations against women was slightly less than the proportion of allegations against males.

Data Cleaning and Integration

In our integration of the settlement database with the CPDB we examined the relationship between the amount of money an officer has costed the city in settlements and the amount of allegations they have received, as well as how race and gender affected this.

Looking at both the individual and average costs to city across number of complaints, we found that where was essentially no correlation between an officer's number of complaints and the amount paid by the city for settlement cases they were involved in. There are low payment cases for officers across the full range of complaint numbers while highest payment cases seemed to be scattered randomly over number of complaints.

We produced similar results for each group of officers when considering the race and gender of officers separately. There again was no correlation between the number of complaints an officer received and the amount of money paid out in settlements concerning the officer. One interesting thing of note is the officers involved in the most costly settlements were more likely to be women for officers with less than 25 complaints while these were almost always men, otherwise. Of course, female officers with high payment settlements and high complaint counts might not exist because of there are few female officers in settlements, in general.

We also looked at whether there was a correlation between number of complaints received before an officer's first settlement case and the number of complaints received after. Our results showed that there was no major correlation between the two and that officers received around the same number of allegations after their first settlement despite whatever number of allegations they had received before. The majority of officers involved in a settlement received 0 to 12 complaints afterwards but ranged from 0 to 78 complaints before.

We also found that only 1 in 7 officers received allegations after their first settlement even after removing officers that resigned in less than one year after their first settlement closing date. More surprising was that only 1 in 3 officer received allegations before their first settlement, suggesting that officers who become involved in settlements are often not the type of officers that receive complaints.

Workflow Analytics

Within our work in Databricks, we first tried to determine whether there was a correlation between race, gender, and length of appointment for officers. We looked at distribution of lengths of appointment for officers of specific races and genders and saw that the distributions were fairly similar across race and gender with many officers appointed between 0 and 10 years and the number of officers in subsequent bins of appointment length decreasing exponentially.

We later encoded officer race and gender to integers and tested for correlation between the encoded values with length of appointment but found weak Pearson correlations between -0.2 and -0.3. Overall, we believe that there is no significant correlation between race and gender and length of appointment.

We next examined the race and genders of officers with the highest allegation rates by beat. We found the officer with the greatest number of complaints for each beat and analyzed the distributions of race and gender across this select group. As expected, white officers, the largest group of officers by race, made up more than half of the officers with the highest allegations by beat. Black male officers made up 25% of the highest allegations officers despite only 15% of officers being black men. However, we did not normalize these results against the demographics of the beats themselves so it is unclear whether this is due to a large number of beats having mostly black male officers.

In an attempt to understand why black officers had been overrepresented in our analysis, we looked at complaint sources, allegation categories, and allegations by year hoping to find a

discrepancy that might explain our observation. We did not find major discrepancies in complaint sources and allegation categories but for allegations per year, allegations against black officers made up 4% more allegations per year on average in the early 90's than in the past 4 years. We think this could mean that there may have been factors in the 90's causing black officers to receive more allegations.

The final topic that we learned about in Databricks was how allegations on the races of involved officers and the race of the victim were distributed in individual beats. We compared these distribution for individual beats to the demographics of the officers and the civilians in beats. We found although this distribution varied widely across beats, they generally matched the demographics of the officers and civilians in the beat.

Machine Learning

We wanted to create a machine learning model to predict which officers complaints would come in against using a combination of their race, gender, rank and age. The machine learner we ended up using was a decision tree because our attributes were primarily categorical and because the different factors and splits are easy to visualize using decision trees.

Our initial decision tree algorithm reached a test accuracy of around 79%; it was able to predict whether or not an officer had been subject to a complaint after 2009 with a relatively high accuracy. Interestingly, we found that the first attribute that our decision tree split on was not race or gender, but rather age; an aspect of officer identities that we hadn't at all considered during our previous checkpoints. It appeared that officers particularly between the age ranges of 37.5 and 47.5 were predicted to be subject to complaints.

In order to connect our predictor to our previous research which centered largely around race and gender, we then decided to remove the ages of the officers from our dataset, and rerun the decision tree. This resulted in a slightly lower test accuracy of around 76%; which seems to suggest that, even without age, race gender and officer position still seem to provide strong predictive power for whether an officer is likely to be subject to a complaint. Here we found that the decision tree now initially split on gender as its first attribute, followed by race. In the case of both this decision tree and the other, it appeared that the status/position of the officer did not provide much information with regards to our question, and was not included as an attribute that was split over.

Overall, it appears that the social location of officers can be used as an effective predictor of whether or not they are likely to be subject to a complaint. One interesting application of this

could be to apply such a predictor to the dataset of all officers (who have yet to be subject to a complaint) and to see if over, for example, the next 5 years, the predictions made by the decision tree turn out to correctly identify officers that are 'set for' a complaint.

We also wanted to use machine learning to predict whether an officer will be the subject of a settlement using their frequency and nature of complaints. Again, we chose to use a decision tree because our attributes were categorical and it would be clear which attributes resulted in the highest information gain at each split.

Our decision tree resulted in an accuracy of approximately 82%. The decision tree first chose to split on the number of allegations for each officer and then the categories of complaints. The cases where the learner predicted that the officer would be involved in a settlement were the following:

1. If allegation count is greater than 47.5 and the category is Operation/Personnel Violations, Lockup Procedures, False Arrest, Criminal Misconduct or Domestic
2. If category is Illegal Search and allegation count is between 28.5 and 36.5
3. If category is Use Of Force and allegation count is between 36.5 and 47.5
4. If category is Illegal Search, Traffic, Drug/Alcohol Abuse, Bribery/Official Corruption, First Amendment then regardless of count the officer will be involved in a settlement

These results seem to make sense; less severe complaint types such as Operational/Personnel Violations will only result in a settlement with a very large amount of total allegations, whereas more severe complaints such as Bribery/Official Corruption or violation of the First Amendment would result in a settlement regardless of the number of allegations.

Looking at the decision tree that was created, it seems like the tree relied heavily on the number of allegations as opposed to the type of complaint. Previously in checkpoint two, we found that there wasn't a correlation between the amount that an officer owed in a settlement and the number of complaints an officer was involved in. However, our decision tree seems to rely heavily on the number of allegations to determine whether an officer will be involved in a settlement or not. From this, it seems that how much an officer owes in a settlement and whether an officer will be involved in a settlement are not related by number of allegations.

Neural Networks

For this checkpoint, we attempted to create a neural language model which could, based on the summary of a complaint, be able to determine whether or not an officer who had received a complaint would likely be subject to another complaint in the future. In order to accomplish this,

similar to the last task, we picked a year to choose as a 'checkpoint'; we first queried for all the officers who had received a complaint at all, chose their most recent complaint before the year 2014 to use as the summary text, and made the task to predict whether, given the text of the complaint, the officer had received another complaint after 2014. Due to running time and implementation constraints, we had chosen exclusively to use the summary text as an attribute of the network.

Implementing this task required the use of a neural language model; we accomplished this by using a model already written in PyTorch by OpenAI, which uses a combination of transformers and unsupervised pre-training followed by discriminative fine-tuning, and had been shown to achieve state of the art results on a variety of tasks. We rewrote large parts of this model in order to take our datasets as inputs and perform binary classification based on whether or not an officer will receive a complaint in the future given the summary of their previous complaint.

An issue that immediately became clear was that the summaries were all fairly similar in structure, and difficult to discern elements of from one another in meaningful ways, which resulted in the model seemingly not being able to improve performance very much. Additionally the existence of complaints without summaries may have confounded the model's ability to make predictions effectively.

Ultimately the performance of our model was relatively variable, and could likely be improved; it achieved around 78% accuracy on the validation set, but only around 62% accuracy on the test set. While the former number is substantially higher than ZeroR / choosing the most frequent result, the latter is about the same, and actually even slightly worse. Explanations for why this may be the case are above; it's likely that, given the lack of seeming information, the model either chose to just classify everything the same or even to choose spurious features of the summaries, which resulted in worse performance. We could have likely resolved this by running the model for more iterations; this would have been easier with access to GPU / other tools that would allow the model to run faster.

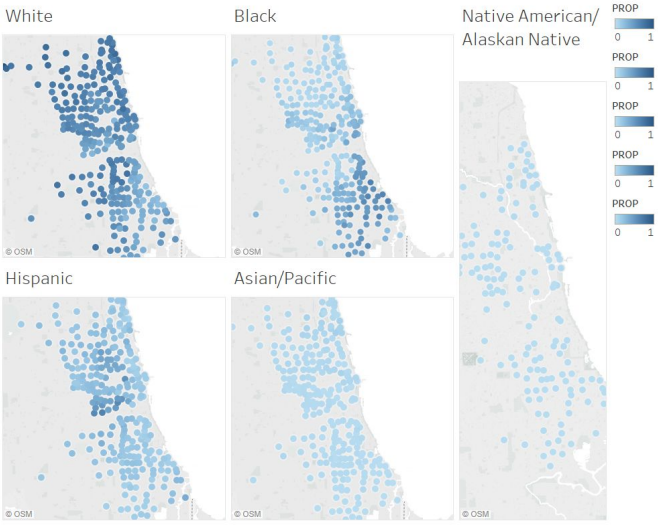
Finally, including other features such as race, gender, rank and age of officer may have supplemented the text predictor and allowed for the neural language model to predict more effectively.

Visualization

The first relationship we wanted to visualize was geographic-based. We wanted to look at the different areas of Chicago and the race, gender, and salary of officers who had complaints placed

in these areas, and the extent that medium income in these areas matched up to these areas. We made maps showing the proportion of allegations against officers of each race and each gender per beat, as well as maps showing the median income of civilians and the average salary of officers in a beat.

The maps that showed the most interesting disparity were our maps showing the proportions of allegations involving officers of a particular race by beat, the maps with the most notable disparity between areas of Chicago are those for white, black, and hispanic officers. The majority of allegation counts are against white officers for most beats but an area of beats in South Chicago near the shore appear to be an exception. Allegation counts involving black officers are shown to make up a higher proportion of all allegation counts in parts of South Chicago but never in North Chicago. There is a rise in the proportion of allegation counts for Hispanic officers in a specific region that seem to correspond to South Lawndale and the Lower West Side which both have high hispanic populations.

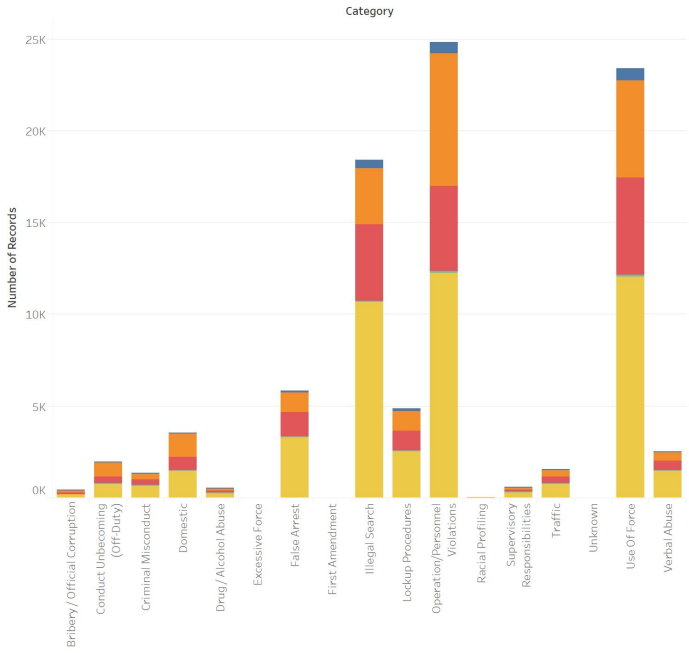


Proportion of allegations involving officers of particular race by beat

We were not able to find any major trends looking at gender and average salary, but we have now seen explicitly that the proportions of officer race in allegations vary between areas of Chicago and how they vary. Assuming that the visualizations reflect reality, we learned from them that there is not an obvious connection between the race, gender, or pay of officers that get allegations in an area and the income of the civilians in the area.

The second relationship we looked at was the relationship between the gender of the complainant, race of the officer, and the nature of the complaint. One interesting disparity we found was when looking at the race of the officer and the nature of the complaint. Overall the

distribution of allegations by race matched with the distribution of officers' race in the police force, but two interesting types of complaints were Domestic and Conduct Unbecoming (Off-Duty). In these types of complaints the split between white officers and black officers was almost equal, almost 40-40.



Officer race and nature of complaints

The Domestic complaint type is particularly interesting because according to an ABC 7 article written February 17, 2017, many of the domestic violence complainants are family of the police officer in question. The family members often hesitate to issue a complaint against the officer because they fear that if they call the police to report the violence, they will only encounter friends or co-workers of the police officer and will not be taken as seriously (Most Chicago police accused of domestic violence go undisciplined, ABC7 Chicago). This seems like it might influence the disparity in our findings as there is the possibility that many cases are not being recorded. We also then looked at the relationship between the gender of the complainant and the type of complaint and found that the gender matched up for the Domestic case, with a larger proportion of the complainants being women. It would be useful in the future to look at whether there is a correlation between race of the complainant and officer and whether there are groups of people who are not coming forward to report domestic violence cases.

Conclusion and Future Work

From our research, we've been able to discern different areas in which aspects of officer identity (most notably race and gender) seem to have and have not affected officers. We saw that race plays a role in the number of allegations that officers receive on average, both race and gender affect the categories of allegations officers receive, and age may be an influencer on an officer's likelihood to receive allegations. On the other hand, race and gender did not affect officer length of appointment or the proportion of allegations from peers. We believe that the discrepancies in number and type of allegations would be great areas to delve deeper in understanding how factors of identity shapes officer experiences.

Now that we've looked into the officers involved in allegations in Chicago, we'd like to perform a similar analysis for officers in other cities to try to understand whether trends in officer allegations are localized or apply to a greater context than just Chicago. One reason this would be interesting to look into is because of the recent attention Chicago has received from political rhetoric highlighting crime in Chicago. Looking at trends in officer allegations across various cities would allow for us to potentially determine whether the officers themselves can be considered as factors in the increased crime rate found in Chicago versus other cities in the United States.