Predicting Bad Cops Based on Their Seniority

The Glorious Gorillas

Introduction

We are interested in investigating the trends between the seniority of an investigator in terms of their age, rank, and their history of complaints. The focus of this investigation is twofold: First, how does the history of complaints correlate with each investigator over time? For example, as an investigator rises in the ranks, does the rate of complaints increase for this officer? Similarly, is there a trend in the type of complaint that the investigator receives? Second, can we understand potential biases that police officers might hold against their victims? Additionally, while making these investigations we are sensitive to keep an eye out for related trends that may not necessarily be confined to the scope of this topic, but are interesting and important nonetheless. For example, trends between officers that rise in rank and their complaint volume might reveal some information about how the Chicago PD operates as an institution. That is, an officer may receive complaints from the public after an incident, but may be awarded or promoted internally by the institution.

Development of Investigations

Our first series of investigations helped our group to find concrete data surrounding aspects of our theme which can be further analyzed for insights. The first set of data we sought to gather was the total amount of complaints an officer had over their career. This helps to showcase a timeline of complaints as measured over an entire officer's career. This is supplemented by the number of complaints per officer every year, over an entire career. This information can warrant a deeper investigation into these particular officers, and so thus we can narrow down why the volume of complaints are so high at the officer level. This was a great introduction to the data as it was our first encounter understanding the distribution of allegations made per officers and what 'a lot' of allegations against an officer really looks like and also what was a normal amount for an officer to have.

We also wanted to understand the nature of complaints for these officers. This helps us understand if an officer holds any biases against certain demographic groups. To achieve this, we performed a query to find the most frequent demographic that comprised of each officers complaint history. Doing this helped us investigate potential causes for racism or bias throughout these officer's careers. Since race isn't the only potential source for bias, we also investigated allegations in terms of the most commonly accused gender per officer. This was a very interesting finding as we were then able to see the disaggregated demographic groups that correlate to certain officers that appeared in the total count table.

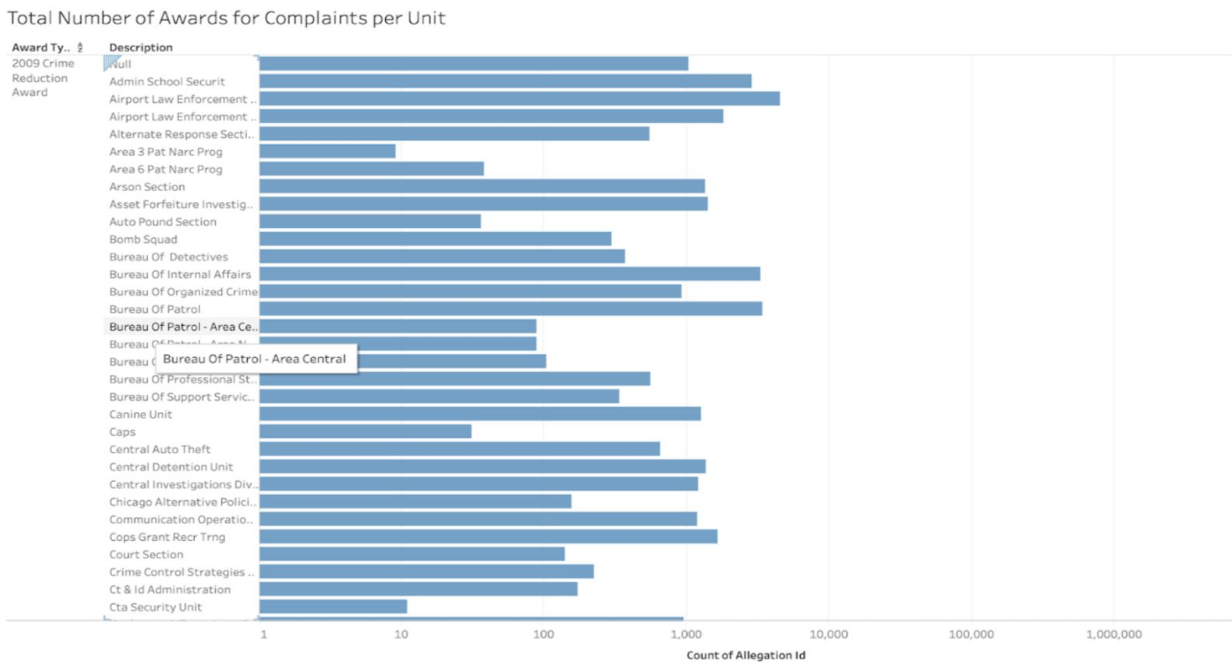
After reviewing our results from our initial queries, we wanted to explore the magnitudes of complaints against officers visually. By disaggregating and separating complaints at the group level, as well as visualizing the outcomes of complaints on a unit level, we were able to investigate and draw conclusions about how the police department reacts to certain complaints.

The first proponent we wanted to visualize the scale of complaints per unit, to get an idea of the magnitude of the unit with the highest amount of complaints against all other units.

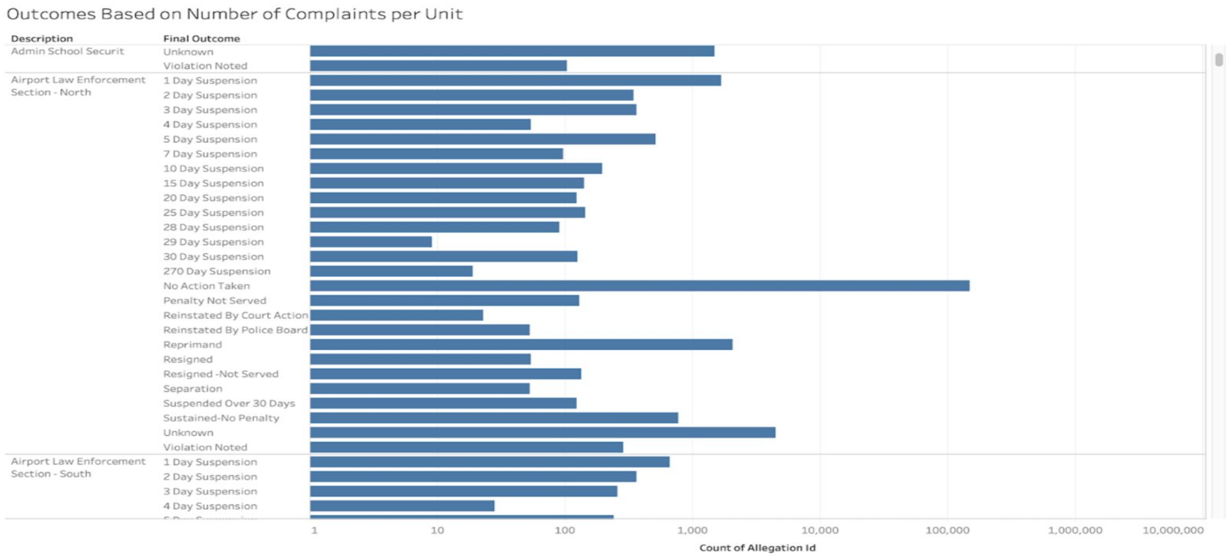
Number of Complaints per Unit

Above is a tree-map of the total number of allegations per unit, with larger and darker squares representing larger shares of the data. Here, we see that the Recruit Training Unit has the most complaints out of all units, with District 007 and 011 being second and third. This intuitively makes sense, as recruits are more prone to error while getting used to the job.

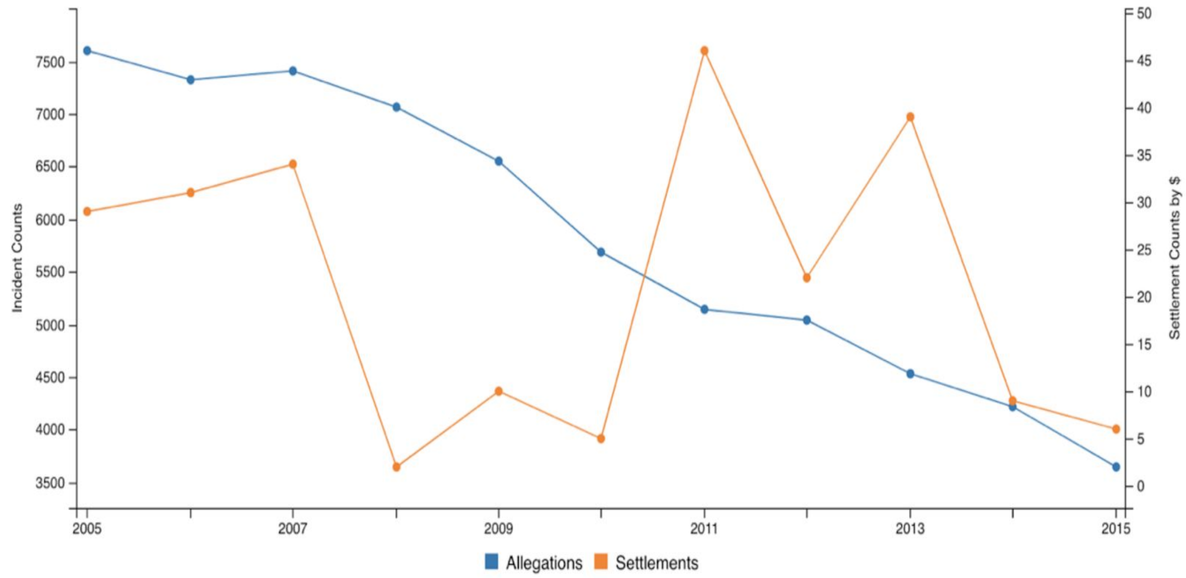
A subsequent line of thinking brought our attention to the awards that police officers receive for various forms of excellence in their field. We wanted to understand how a police department was handling police officer allegations through the perspective of professional and personal gratification through awards. Perhaps this could uncover foul play or bias in the department. We investigated this by looking at the number of officers per unit that have received awards for corresponding complaints. This is interesting as the count of the number of allegations are represented on the bars, with each corresponding unit and award type (as you scroll down, you find segments for each award). Thus, the count of allegation ID's represents the count of award "x" received per allegation. However from just looking at the graphs below, there does seem to be a relationship under certain departments that leads us to believe that departments may ignore or worse yet, condone irresponsible behavior through positive reinforcement with awards.



Next, we wanted to see if we could take another look at understanding complaints and reward systems in police departments. Specifically, looking at the outcomes of allegations for departments. This is a bar chart depicting the final outcome of each complaint by unit. The Unit is presented in the first column, and the outcome on the second, with the count of allegations corresponding to each outcome. This shows us if and how officers per unit are disciplined on a per complaint basis. One thing we noticed was how the “no action taken” field was generally the highest among all units.



Finally, we wanted to visualize how allegations correlate with settlements. The reason for this is to investigate on the complainant side, how many allegations actually resulted in settlements? On the left- hand side, you can see we have incident count from 3500 to 7500. The blue line is associated with such dimensions as aided by the legend at the bottom as indicated ‘Allegations’. On the right-hand side, you can see Settlement Count by $ amount from 0 to 50. This is measured with the orange line graph as indicated ‘Settlements’. This is an easy way to show allegations and total dollar settlement amounts per year from 2005 to 2015. We find that there is not an apparent correlation between allegations and settlements. Allegations as shown are actually steadily decreasing over the 10-year span, while settlement counts seem to go in several year bursts of high activity, then followed by an equal amount of years with low settlement dollar counts. Note that this data is also not representative of the entire dataset as the we used a sample size of n=1000 to yield a palatable graph.



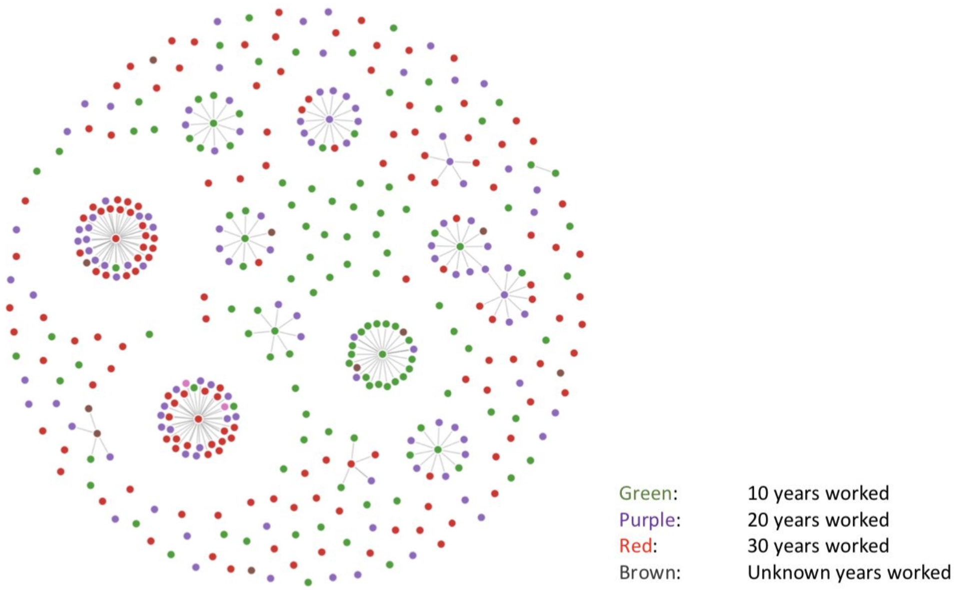
After visualizing and performing initial queries, we wanted to clean and integrate the data to better fit the scope of our objectives. The first question our group sought to investigate how years worked as a police officer correlate to the amount of arrests they have over their career. Interestingly, there does seem to be a trend in 'bad apples'. There are a select few officers who account for a significant portion of arrests, and average over 15 years of work in the field. This contributed to our research question because we can see that longer working officers not only arrest more over time, but the 'bad apples' seem to all be in the late stages of their careers and are more senior in their experience and at times, more senior in their positions. For future directions of inquiry, it would be interesting to understand the rankings and awards for these 'bad apple' cops to see how these people progress in their careers with the complaints and arrests they have. In addition, we wanted to understand if there is any significance between how many complaints an officer has with the amount of arrests they have? Interestingly, the amount of arrests made per officer seem to be held closely in tandem with complaints made against officers. In some cases, there are as many arrests as there are complaints. This is a natural follow up to question 1 as this data tell us how many complaints and arrests each officer had. We wanted to expand our research by not only understanding a timeline of years served and arrests but also what kind of story the data can tell around an officer's proportion of arrests with complaints.

Moreover, we looked at the data to understand any biases in officer judgments for arrest by understanding arrests as a proportion for each ethnic/racial demographic. Through our queries, we found that there are 5 racial groups which have counts of complaints from each group as well as the count of arrests. It should be noted that a lot a large portion of this data is a result of not reporting the race of the complainant. When entering this data, this lack of information contributed greatly to having 'dirty data'. However, from the vast amount of data that we as researchers can understand we can see that the number of complaints and arrests strongly resemble one another. A hypothesis can be drawn that a large number of complaints coming in were a result of an arrest due to police misconduct which is a significant finding if proven to be true with more supporting data. This data helped advance our understanding for our research question as we wanted to understand any ‘hidden’ trends. This certainly seem to show biases for certain demographic groups in the Chicago area. The city of Chicago is made up of 45.3% people of the white demographic and 32% of people in the black demographic group (according to the http://worldpopulationreview.com/us-cities/chicago-population/ census). There seems to be a disproportionate amount of arrests being made for people of color. This unfortunately (and fortunately) sheds light on trends that a minority of Chicago citizens seem to be facing disproportionately. It would certainly be interesting for future investigations to understand the demographic groups of people making complaints to see if there is a proportionate amount of people making complaints for wrongful arrests.

The next step in our analysis was to explore the relationships between co-accusal networks of officers as grouped by the most commonly accused demographic group per officer, and we decided the best way to represent this was through network graphics.

What we learned from this analysis is that an officer does seem to share accusations with other officers who have the same grouping category (with of course, exceptions). This matched our expectations that officer groups would be co-accused with other officers of the same grouping label. Another notable observation from the visualization is that there is a larger proportion of nodes labeled as ‘Black’ as compared to other label types. This means there is a much higher proportion of complainants that are of one racial group over all others as shown by the nodes (orange label). It also can be said that there is a strong correlation between officers to be co-accused with other officers of a similar label. Even if you look at the blue label, which is the ‘Unknown’ label, you can see that is fairly randomly spread out and random in terms of its connections to other nodes and groups.

To explore our initial theme regarding seniority, we wanted to understand the network analytics of co-accused officers based on how long they have served for and who they are co-accused with. It was important to create a question that moved our research forward in helping understand officer seniority from officer’s with complaints against them.



What we learned from this analysis is that an officer seems to share accusations with other officers within the same category. This is of course a generalization but still holds true nonetheless. This came as a surprise as the graph tells us that officers seem to be accused with other officers of the same ranking. This might make sense logically as officers of the same rank are likely to be assigned to work together and patrol together. Moreover, we see aggregation of some higher-ranking individuals in many of the networks. This could be due to the fact that sometimes senior officials get listed in the reports as a co-accuser even if they were not directly involved.

Finally, we wanted to leverage machine learning and text analytics to gather further insights and predictions into officer behavior as alleged from each complaint, and how this may relate to their biases. We first wanted to utilize machine learning to see if we could model patterns of behavior to make predictions about officer allegations. We wanted to understand what kinds of features were relevant in predicting just how many allegations an officer could expect to have over their career. We believe that officer rank, total years worked per officer, officer race, total officer arrests, and most commonly targeted officer race were contributing factors to predicting how many allegations an officer would have over a career. It turns out, there is a high likelihood chance of predictions based on these features. Using an SVM model, we achieved an accuracy of 93.7% for officer allegations. This was a huge surprise and big finding. We tested our model on a dataset size of around 9,000. This might be an interesting dimension to investigate whenever police departments are looking to hire new officers. It's exciting to think that a model such as this can help our community leaders and advisors make a more efficient and clear analysis on officers.

```
Results for predicting amount of allegations per officer:
---------------------------------------------------------
F1 score:         0.9525559658566799
Accuracy Score:   0.9365466748017084
```

Moreover, we wanted to see if this model could make predictions on officer bias since it worked so well to predict allegations. We utilized the same model and dataset but exchanged predicting allegations for predicting the most commonly targeted racial group per officer. The only change is that we included officer allegations in the feature set rather than the target set. The results are as follows:

```
Results for predicting most commonly targeted racial group per officer:
-----------------------------------------------------------------------
F1 score:         0.7750373692077728
Accuracy Score:   0.6327028676021964
```

The same model had an accuracy of around 63.3%. While this is above chance, we concluded that this model isn't accurate enough to be able to properly say that it makes good predictions. This makes sense because it might not be quite as easy to predict something abstract like bias. The model did concur with our gathered findings from previous work in predicting 'Black' as the most common label for each officer. Our previous work gave insights that officers certainly have a bias against black Chicago citizens. We would definitely be interested in developing this model further to understand perhaps other features needed or other labels to be able to meaningfully predict bais.

Utilizing text analytics, we wanted an effective way to first tag and label each document, and second, categorize and summarize the details of each document to draw meaningful conclusions. We first tried to label documents under "racial slurs" and get the count of those documents. The reasoning for this was to find in proportion to all the available complaints, how many of them contained content which involved racial slurs from the officer. The final count we managed to achieve was 44 cases versus a total of 1220 documents, roughly 4% of all available complaint documents. This is interesting as it is perhaps the closest that we could get to straightforward evidence of racism, as looking at statistics and queries alone cannot help draw definite conclusions. However, to help support this finding and to gather more evidence into racism, we then wanted to go a step further and check for how many of these documents categorized as racial slurs involved violence? That is to say, how many complainants that filed for racial slurs also experienced some form of aggression or violence during their encounter? The result we find here is surprising as it seems that out of the 44 documents we found for racial slurs, 30 of them, or 68% of them also involved some sort of violence. As a result, we believe that this provides more evidence towards the racism argument, and the next step from this would be to check for which officers were involved in these complaints, and are there any repeat officers within this subsample?

Conclusion

While our group only had a few months to begin an investigation into the CPDB dataset, we consider our findings to be a huge win and source of developments in terms of the amount of insights we were able to uncover and synthesize. This was a superb opportunity to not only learn how to use database tools, but to also investigate a real world problem while doing so. Our roadmap (checkpoints) helped give structure to different perspectives that allowed us to tackle our problem through unique and valuable lenses. What first struck as interesting was just how many allegations Chicago police officers have. We didn't have much of an idea of what was ‘normal’ until we saw all of the totals. This helped us get a foundational expectation as to how the officers in the dataset are behaving at work. What really struck us was the top percentage officers who had a hugely higher proportion of allegations against them but still continued to not only work, but also get appraisal in the form of awards by the police department. We were surprised that we could actually visualize patterns of ideology and pathology, entangled in a network of police officers within and between departments. These would have been too abstract of concepts to see through a table but Tableau and especially our resulting network graphs really helped show us in an intuitive manner how police officers tended to create communities of recurrent behavior amongst one another. What we learned was that bias isn’t just a term. The data doesn’t lie and we found out in real numbers just how much police officers seem to take advantage of their authority to behave in an inappropriate manner in many cases. While again, this is not the overwhelming majority of officers as most officers seem to be ‘good apples’, this was an investigation seeking out ways in which our public servants are abusing power which in turn seems to be harming many of its citizens wrongfully. We certainly have more questions than answers at this point and certainly have future lines of inquiry such as: what is a police beat made up of? Are they of similar ranking and experience level? What can we understand broad-scale about the temperment of these police officers from the data? Is there a certain kind of personality that draws police officers to become officers?

To be transparent, before starting this investigation it was difficult to believe that some numbers and tables on a screen could really tell us any sort of meaningful narrative about the world. The fact that we can now draw insights at a remarkable level into the patterns of behavior is our biggest surprise this quarter. We have just scratched the surface from a breadth of yet to be seen knowledge on topics such as behavior of individuals, networks of people, network of units, biased ideologies and patterns, textual context, geographic analytics, interactive visualizations and other similar insights which help illuminate an otherwise vast and incomprehensible source of data. This project and these checkpoints helped show us how you can utilize remarkable tools in data science and engineering to bring more truth and understanding to a messy and complex world.