Use of Force by Officers in the Chicago Police Department

The Flying Koalas
Meenakshi Kommineni, Madison McClellan, Archana Ramasubramaniam

COMP_SCI 396/496: Data Science Seminar
Professor Rogers
Fall 2021

Theme and Motivation
The high-level theme that we wanted to explore this quarter was the conditions under which officers in the Chicago PD use force. Ideally we wanted to isolate a set of variables that predict higher frequency of use of force such as subject demographic or the location of the incident. Our motivation for choosing this topic was due to the strong media coverage of use of force incidents in police forces. We wanted to use data to understand these incidents on a deeper level and without the bias of media sources. Although we analyzed complaint reports (CRs) for a source of comparison, the majority of our work was focused on tactical response reports (TRRs).

Checkpoint 1: Relational Analytics
Questions
- Are certain types of use of force more common in certain districts?
- Does age play a role in what action the officer takes against the victim?
- Are there trends in officers' actions based on the lighting conditions of the incident?
- Are officers more likely to use a firearm on an unarmed victim if they are off duty?
- Are there trends in officers' actions based on the race of the subject?

Methodology
For this checkpoint, we used simple SQL queries and grouped them by the different attributes that we found interesting. To understand the police presence in each district, we started by writing a query to count the total number of TRR incidents in each of the 25 districts. To understand the role age plays in the type of action taken against the subjects, we wrote a query to count the number of TRR incidents that occurred against subjects of each age. This piqued our interest to understand if the type of force taken by officers differ based on the age of the subject. We grouped subjects into three age ranges to simplify the results: 0-29, 30-59, and 60+. For each of these age groups, we wrote a query to divide the total number of TRR incidents based on the action type. Similarly, we wrote queries to count the number of TRR incidents based on the conditions we wanted to analyse.

Findings
First, we understood that the 11th district had the greatest number of TRR incidents with 29261 incidents whereas the 20th district had the fewest with only 4275 incidents. When we tried to find out the most commonly used type of force across the different districts, we found that the same types of actions were used at similar frequencies in each district. For example, verbal commands were the most common TRR action in almost every single district, which made sense given that this is a more mild use of force.

When we tried to find out if age of the subject played a role in the actions taken against a subject, we found that the subject age with the highest number of TRR incidents is 22, with a total of 15838 incidents. Ages 21, 23, 20, and 24 followed. When we tried to find out if the type of action taken against subjects differs with different age groups, to our surprise we noticed that verbal command was the most commonly used force against people of all age groups. This follows the conclusion from question one that verbal commands were the most commonly used type of force in almost every district.

The next question we wanted to answer was the role of lighting conditions in the type of force used by officers. The results for this query indicated that a vast majority of TRR incidents occur in good artificial lighting and daylight. Furthermore, when dividing TRR incidents by the type of action taken by the officer, a majority of each type of use of force occurred in good artificial lighting or daylight. These findings were inconsistent with our initial assumption that many TRR incidents would occur at night or in poor lighting conditions.

To understand the usage of firearms on unarmed subjects by officers who are off-duty, we first tried to see the number of instances where firearms were mostly used against armed and unarmed subjects. We learnt that 17% of incidents occurred on unarmed subjects (166 of 993 total instances) and the remaining 83% of incidents occurred on armed subjects (827 of 993 total instances). On further investigation, we found out that only 14% of instances of firearm use on an unarmed subject were performed by off-duty officers. This was contrary to our assumption that there would be many instances of firearm use by off-duty officers.

Our final analysis concluded that even though the Black population does not constitute the majority of Chicago's population (30% according to US Census Bureau in 2019), 75% of subjects involved in TRR incidents were Black.

Checkpoint 2: Data Exploration

Questions

- [Scatterplot] Is there an overlap between officers with a high number of TRRs filed against them and officers with a high number of complaint reports?
- [Correlation matrix] Is there a correlation between an officer's use of a firearm and the lighting conditions of the incident?

Methodology

For this checkpoint, we used Tableau to construct a scatter plot to check if there was an overlap between officers with a high number of tactical response reports (TRRs) and officers with a high number of complaint reports (CRs) filed against them. We then constructed a correlation matrix to analyse the connection between an officer's use of a firearm and the lighting conditions of the incident.

Findings

The goal of the scatterplot was to determine what type of correlation, if any, exists between the number of TRRs and the number of CRs filed against officers. Without looking at the data, we assumed that we would find a positive correlation between these variables. Each mark on the scatterplot represented a single officer based on their unique ID. From the scatterplot we noticed that there is no correlation between the number of TRRs and the number of CRs filed against officers. Rather, we see that most officers fall within a lower range of both TRRs (< 30) and CRs (< 70), but that we cannot use the number of TRRs to predict the number of CRs or vice versa (i.e. there is no correlation between the variables). This disproves our initial hypothesis that we would find a positive correlation.

For the correlation matrix, we extracted data from the TRR table, grouping by lighting condition and use of firearm. Our initial assumption was that we might find more frequent use of firearms under poor lighting conditions or nighttime because it might lead to more ambiguity in the situation. For each combination of firearm use and lighting condition (dawn, daylight, dusk, good artificial, night, and poor artificial), we counted the number of instances and labeled the corresponding box in the matrix. We also used a color gradient to visually represent the counts. The first conclusion that we drew from the matrix was that most TRRs did not involve a firearm. Secondly, we could conclude that most TRRs occur under good artificial lighting and daylight, regardless of whether a firearm was involved. This disproves our initial hypothesis that we would find many instances of firearm use under poor lighting or at night.

Figure 1: Number of CRs vs. Number of TRRs

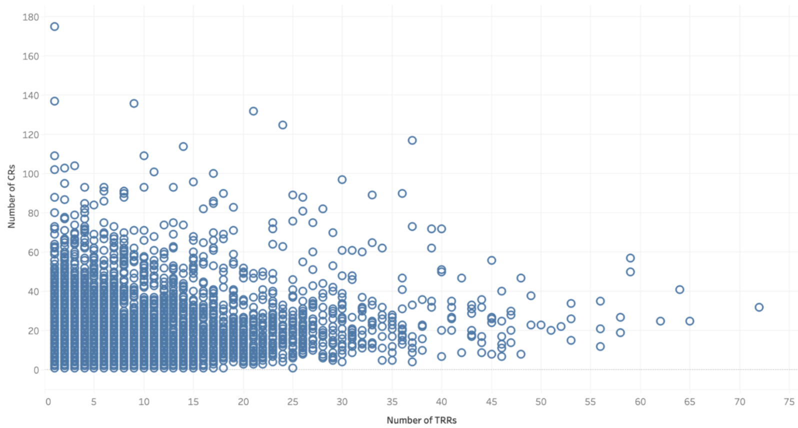


Figure 2: Correlation Matrix for Firearm Use vs. Lighting Condition

Checkpoint 3: Interactive Visualization

Questions

- An interactive bar graph to observe how the distribution of TRRs by action type (e.g. verbal commands, taser, etc.) changes in relation to police district, subject age, subject race, lighting condition, whether the officer was on or off duty, and the year. Furthermore, we want to observe whether this distribution is dependent on the group of officers being observed. Do officers in the highest percentile of TRRs use specific types of force more than officers in a lower percentile of TRRs?

Methodology and Findings

To explore these ideas and extend upon our Checkpoint 1 findings we designed and implemented interactive visualizations using D3.js. First we built a bar chart showing the frequencies of the number of TRR instances of each action type. We created a dropdown menu for each variable we wanted to consider, allowing the user of the visualization to select specific values to filter the TRR data. The bar chart automatically updates to reflect the filtered data. After observing the interactions between each of these variables we observed that the distribution of TRRs across action types remains relatively stable across different districts, subject ages, subject races, lighting conditions, officer status, and years. This visualization illustrates how the proportion of TRRs in each category remains the same as the data is filtered across these variables.

After observing the results in the above interactive visualization we were able to develop an intuition about the data used and wanted to further determine whether the distribution remains stable when looking at officers that have TRRs filed against them at different rates. We built a second bar chart that only shows the top five action types for each percentile range of officers. We included a drop down menu of percentile ranges so the user can filter the data and see how it varied. Our initial assumption was that officers with a high TRR percentile might be more prone to using more aggressive types of force (e.g. firearm, taser) as opposed to more mild types of force (e.g. verbal commands). This visualization disproved our initial hypothesis and showed that the distribution of TRRs is the same (i.e. officers use the same types of force in the same proportions) regardless of how often they use force. This is an interesting finding because it suggests that officers who are more prone to using force do not have different tendencies than those who are less prone to using force. It would be much more problematic if the officers in the highest TRR percentile also used more severe types of force.

Figure 3: TRR Distribution by Action Type

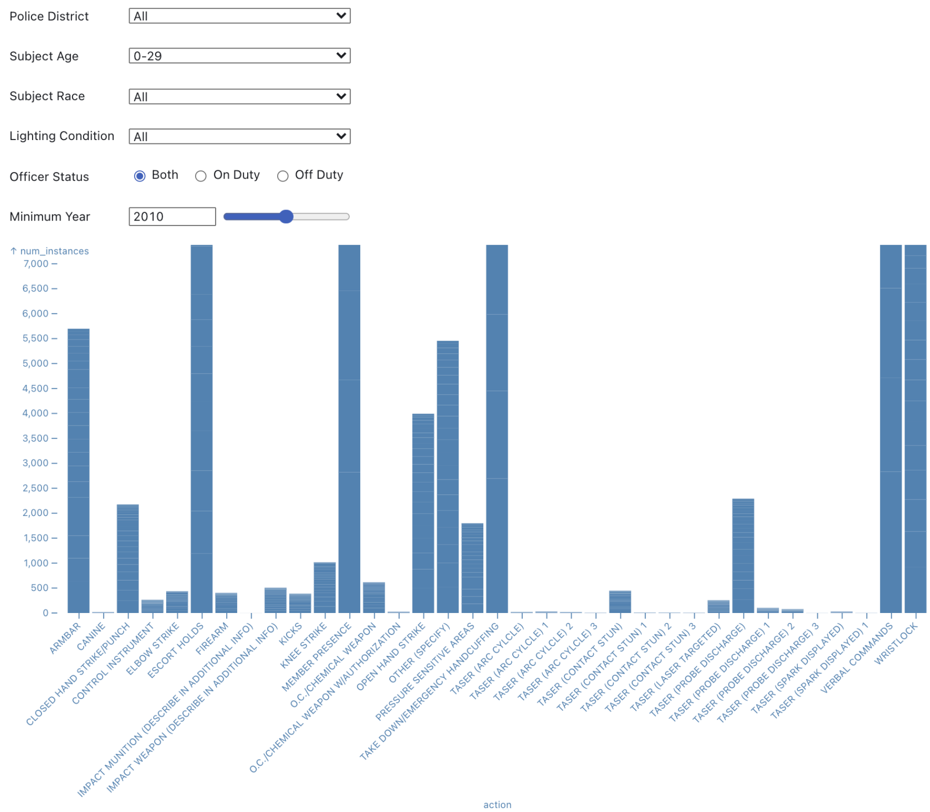
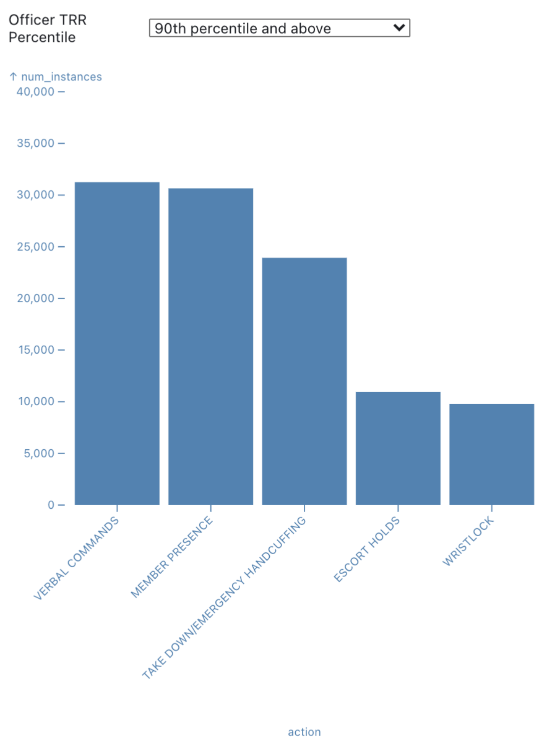


Figure 4: Top Five TRR Action Types, Filter by Officer TRR Percentile



Checkpoint 4: Graph Analytics

Questions

- If we construct a graph where nodes represent officers and edges signify that two officers were involved in the same allegation, can we then determine whether specific officers are more involved than others in these incidents?
- Can we determine whether the graph described above contains many clusters (i.e. officers are highly connected in allegations) or whether officers are more isolated? How does this compare to a graph of officers involved in TRRs?

Methodology and Findings

Using the data from tables data_officer and data_officerallegation we constructed a graph where the nodes were queried by selecting all of the values for officer_id from data_officer. The edges were queried by joining data_officerallegation with itself and selecting pairs of officers where the allegation_id matched signifying the officers being co accused in an allegation. Upon running the page rank algorithm over this graph we wanted to determine which subset of officers were highly connected (co-accused) with other officers using Pagerank values. Further investigation into these highly connected officers would be needed to draw any definitive conclusions, but we could potentially infer that these officers have a greater negative presence in the community.

We then wanted to determine whether these highly connected officers have any specific traits in common. To do so we selected the top 20 officers based on page rank values and queried these officers IDs from the data_officer table. All 20 officers in this group are males born in or before 1951. The racial distribution is 50% Black (10/20), 40% White (8/20), and 10% Hispanic (2/20). The average complaint percentile among the group is 92.5. Perhaps the most interesting insight out of these is that all 20 officers are male. This is consistent with what we learned from the Invisible Institute's policy recommendation of hiring more female officers. We can note that these officers are not necessarily "outliers". The next 20 officers' values are not drastically lower than the top 20. This trend continues as we look at groups of 20 officers lower and lower down the list. This is inconsistent with the "bad apple" theory that suggests the majority of officer misconduct occurs among a small group of easily isolated officers.

Next we ran the Triangle Count algorithm over the graph constructed above where the resulting counts for each officer ID represent the number of triangles a given node is involved in. We calculated the average count over all the nodes, which was 246.5. This value tells us that on average, a node is involved in 246.5 triangles. We then constructed another graph with nodes as officers and edges signify officers involved in the same TRR and ran the triangle count over this graph and found that in general, there are much fewer triangles. For example, the officers identified by the IDs 21371, 13313, and 11615 have the highest counts, with 206, 182, and 180 triangle involvements, respectively. The average count over all the nodes was only 1.8, compared to 246.5 for the allegations graph. From this comparison, we can conclude that officers are more often co-accused in allegations than they are in use of force incidents.

Checkpoint 5: Natural Language Processing

Questions

- Do civilian narratives of complaint reports against officers in a high percentile of TRRs (90th percentile and above) contain more negative sentiment than those against officers in a low percentile of TRRs (below 50th percentile)?

Methodology
To answer this question, we used the built-in sentiment analyzer in Python's Natural Language Toolkit (NLTK). We gathered data by connecting to the PostgreSQL database from within the Jupyter Notebook and querying the desired data.

Findings
The results of the sentiment analysis over CRs indicated that approximately 11% of narratives contained positive sentiment and 89% contained negative sentiment for both the >=90th percentile and <=50th percentile groups. We were surprised to see that CRs were equally split between positive and negative sentiment for both percentile groups. Initially we expected that CRs against officers in the >=90th percentile group would have much more negative sentiment than those against officers in the <=50th percentile group.

We also wanted to evaluate how accurate the NLTK models are against the CR narratives. To do this, we gathered a small sample of CRs and examined how the model classified them. The narratives that were classified as negative contained stronger language, mentioned weapons, and included words that are generally perceived as negative. However, the narratives that were classified as positive did not contain many words that are generally perceived as positive. Rather, these narratives were "less negative" than those classified as negative. For example, they did not mention weapons and typically involved less severe incidents such as traffic stops.

Main Takeaways
One of our primary goals was to identify conditions under which use of force is more likely to occur. Our findings from Checkpoints 1, 2, and 3 provided a basis for answering this question. In Checkpoint 1, we found that use of force, measured by the number of TRRs, occurs much more frequently in certain districts. Furthermore, there are drastic differences in the number of TRRs between districts. For example, the 11th district has the greatest number of TRR incidents with 29,261 whereas the 20th district has the fewest with only 4,275 incidents. A per capita measure of TRRs might be a helpful measure to consider in the future, but we can still conclude that TRRs do not occur proportionately in different areas of Chicago. We also found that a majority of TRRs occur against Black males in their early 20s. This finding was consistent across districts. Another variable that we considered was lighting condition. We found that a majority of TRRs occur in good artificial lighting and daylight. Finally, we looked at whether TRRs are more likely to occur when officers are on- or off-duty and when subjects are armed or unarmed. We found that most TRRs occur when officers are on-duty and when subjects are armed. All of these results tell us that use of force does occur more frequently in certain circumstances. However, after completing Checkpoint 1 we realized that using the results of SQL queries to draw conclusions is difficult without visualizing the results. Thus, we used interactive visualizations in Checkpoint 3 to further solidify these findings.

Through the first three Checkpoints we also found that officers use different types of force (e.g. verbal commands, taser, firearm) proportionately as the variables mentioned above differ. For example, although most TRRs occur against young, Black males, officers use different types of force proportionately across all racial and age groups. Similarly, although there are vast differences in the total TRR counts across districts, verbal commands are the most commonly used type of force in almost every district. To build off these findings that officers use different types of force in similar proportions regardless of the situation they are in, in Checkpoint 5, we wanted to look at differences in civilian narratives against officers who are more and less prone to using force. We found that narratives against officers in the highest percentile of TRRs and officers in the lowest percentile have almost the exact same amount of positive and negative sentiment. Although more investigation would be needed, this finding suggests that officers who are more prone and less prone to using force are perceived similarly by civilians. One limitation of this finding is that we used Python's Natural Language Toolkit (NLTK) for sentiment analysis, which might not be a great model for the language used in civilian narratives. Although we manually checked the model's classification for several narratives and it appeared to be relatively reliable, a custom model would be helpful to draw stronger conclusions about positive and negative sentiment. These two findings about officers' use of different types of force and the proportion of positive and negative sentiment in CRs against officers in different TRR percentiles suggest that officers exhibit similar behavior across the police force. One policy implication of this finding is that implementing behavioral training across the police force might not be effective in solving the issue of excessive use of force. Rather, stricter punishment for officers who use force excessively might be more effective.

Checkpoint 4 slightly deviated from our overall theme as we analyzed both CRs and TRRs. We constructed two graphs, one representing co-accusals in CRs and one representing co-accusals in TRRs. Through the results of the Triangle Count algorithm, we found that officers are more likely to be co-accused in CRs than they are in TRRs. Because TRRs are more likely to be committed alone, this might suggest that pairing or grouping officers when they are in situations that use of force might occur could be a beneficial tactic in reducing excessive use of force. For example, our findings from Checkpoint 1 revealed that use of force occurs most often in District 11, so ensuring that officers are paired or grouped in this district could be beneficial. One limitation of this finding is that we treated all TRRs the same, meaning that we did not distinguish between those that were an appropriate use of force and those that might be considered excessive use of force. We might be able to determine this by examining factors such as the type of force used and whether or not the subject was armed and resisting. However, this would require extensive manual effort for the amount of data used.

Future Research Directions

Throughout this project, we focused on examining use of force at particular points in time. More specifically, most of the data we considered involved gathering totals or averages up to a certain point in time. We did not consider taking a different approach and looking at how use of force changes over time with respect to different officers or conditions. For example, do officers in the highest TRR percentile use force consistently across their careers or are there turning points where they begin to use force more frequently? Examining use of force with respect to time would provide an interesting new perspective to build on the work we completed in this project.