CPDB Injury Trends

By: Hawkins Gay, Alex Leidner, Ramsey Wehbe

Data Science Seminar Fall 2020

Introduction:

As members of the healthcare community our team chose to focus on complaints of use of force and their propensity for injury, as this provides a potential intersection between the CPDB and Chicago healthcare systems. We hypothesized that those that experienced the most of force would also experience more severe force resulting in injury. Given national and international protests, investigating inequality through data and observational evidence is important not for accusation but to hypothesize methods to determine root cause as well as means of improvement. The following checkpoints will stratify injuries by different demographics including, race, age, sex, and police district. Information and discussion thereafter will be presented both as a total amount of injuries and use of force, as well as the percentage of use of force. Ideally the synthesis of all of this information could fuel real world experiments that would lead to root cause analysis and reform in the Chicago police department that hopefully would be valid for the rest of the country that also struggle with these issues.

Checkpoint 1, Injury Trends:

The following questions have been answered using SQL queries. These methods can be inspected in Checkpoint 1, The Incredible Pirates folder.

Question 1: What percentage of use of force incidents result in injury for citizens and police Officers, broken down by race, age, neighborhood, and type of force.

Here we found that throughout the 67019 CPDB use of force reports 26% result in subject injury, and 22% result in officer Injury. In addition out of 14,494 subject alleging injury, 18% of those allegations were not confirmed by the database. Next we stratified these injuries by race (Table 1). Black's experience more use of force accounting for 74.2% of the total use of force, but injury profiles when stratified by race remain consistent at 25-30% of subjects of any race being injured, with Hispanics being injured the most often at 33%. The 2010 census data shows that Black's account for 30% of the population of Chicago. Alleged injuries that were subsequently rejected amongst races were similar ranging from 16-19%. Officers cause injuries regardless of their race at a rate of 22-24%. Finally, when stratified by officer race and subject race, an Asian officer and Hispanic citizen use of force was more likely to result in injury (Appendix 1).

Table 1:

subject_race	total_use_of_force_ev...	subject_injuries	alleged_injuries
BLACK	49747	12217	10377
HISPANIC	9369	3150	2285
WHITE	6540	2116	1536
<null>	878	238	198
ASIAN/PACIFIC ISLANDER	431	124	92
NATIVE AMERICAN/ALASKAN NATIVE	54	9	6

subject_race	percent_subject_inju...	percent_subject_alleged_injuries_not_co...
HISPANIC	33	15
WHITE	32	13
ASIAN/PACIFIC ISLANDER	28	13
<null>	27	18
BLACK	24	19
NATIVE AMERICAN/ALASKAN NATIVE	16	16

Subject Injuries broken down by Race, Total(Top) Percentage(Bottom)

Percent Injury increases with subject age with 21% of 0-18 year olds being injured and 28% of >65 year olds being injured. Females are injured less than Males with 13% and 28% respectively. Females are less often the subject of a use of force report. Districts and their difference are better explored in Checkpoints 2 and 3.

Question 2: Are there difference in injury pattern in relation to different types of uses of force stratified by subject race.

Firearms were much more likely to cause injury at 64% than Taser at 40%. Race did not seem to be correlated with percent of reports reporting use of these weapons. While few use of fire arm events in the CPDB(1,030) Whites had more frequent injury at 76%. Taser was similar across races at ~48% however Black individuals were injured less often at 37%. Additional stratification of the types of force included in which force is more likely to cause injury is better explored in Checkpoint 3.

Question 3: Are neighborhoods with higher rates of officer injury reports more likely to be associated with subject injuries or total number of events.

Beat 1134 had the highest number of subject injuries (251 and 28% of events) however officer injuries in this beat were not the most or highest percent injured. Checkpoint 2 and 3 will further explore injury per district. The beat with the most officer injuries,621, had a low rate of subject injuries at 21% (Appendix 2)

Question 4: Are individual officers more likely to be involved in use of force incidents that lead to injury.

Yes. When evaluating officers with more than 10 force reports, there were some officers with percent subject injuries vastly outpacing others. Some officers had average subject injuries such as officer 10583 with 72 use of force reports who had 26% subject injury, consistent with the database. However there were other officers with 23 use of force reports and 74% injury.

Table 2:

	officer_id	total_use_of_force_events	subject_injured	as_percent_of_events
1	<null>	287	86	29
2	23132	47	30	63
3	20756	47	24	51
4	16551	31	23	74
5	20038	49	22	44
6	2470	56	22	39
7	32105	56	21	37
8	30290	45	20	44
9	6097	58	20	34
10	15631	56	20	35

Individual officer percent injuries

Question 5: Are individual officers more likely to underreport injuries – ie are they less likely to report injury in TRR when injury is alleged by a complaint.

The average alleged injuries that were not counted as true injuries in the database again was 18%. There were some officers whose reports discounted as many as 50% of injuries (Appendix 3).

Checkpoint 1 Conclusion:

This data indicates many important themes that will be further explored. First is that Black's experience much more use of force events than the remainder of the population. This is additionally concerning due to the fact that Blacks are a minority of the Chicago population at 30% compared to Whites at 49%. Next, we have shown that most subcategories of use of force reports result in injury around 25-30% of the time. There are not many significant outliers no matter the stratification; this shows that there is likely a separation between the decision to use force and the intent to cause injury. One area that there are outliers is in individual officers and beats. These may be areas for focused improvements, while it is possible that the overall r ate of 25-30% is too high, if the Chicago Police as a whole can keep to that level of injury, then individual officers should not be at 50 and 70% injury over a large sample size. Officer injury was not the impetus for increased subject injury.

Checkpoint 2 and 3 – Data Visualization and Interactive Graphics:

We analyzed some questions utilizing visualization tools for better clarity and depth. We had specific questions about how injuries were distributed amongst demographic groups and districts. To investigate these, we employed different visualization tools including circle graphs, stratified bar charts, choropleth maps and interactive graphics.

<u>Question 1 (CP1)</u>: How are use of force incidents distributed by demographics (race, gender, age) and what are the percentage that result in injury?

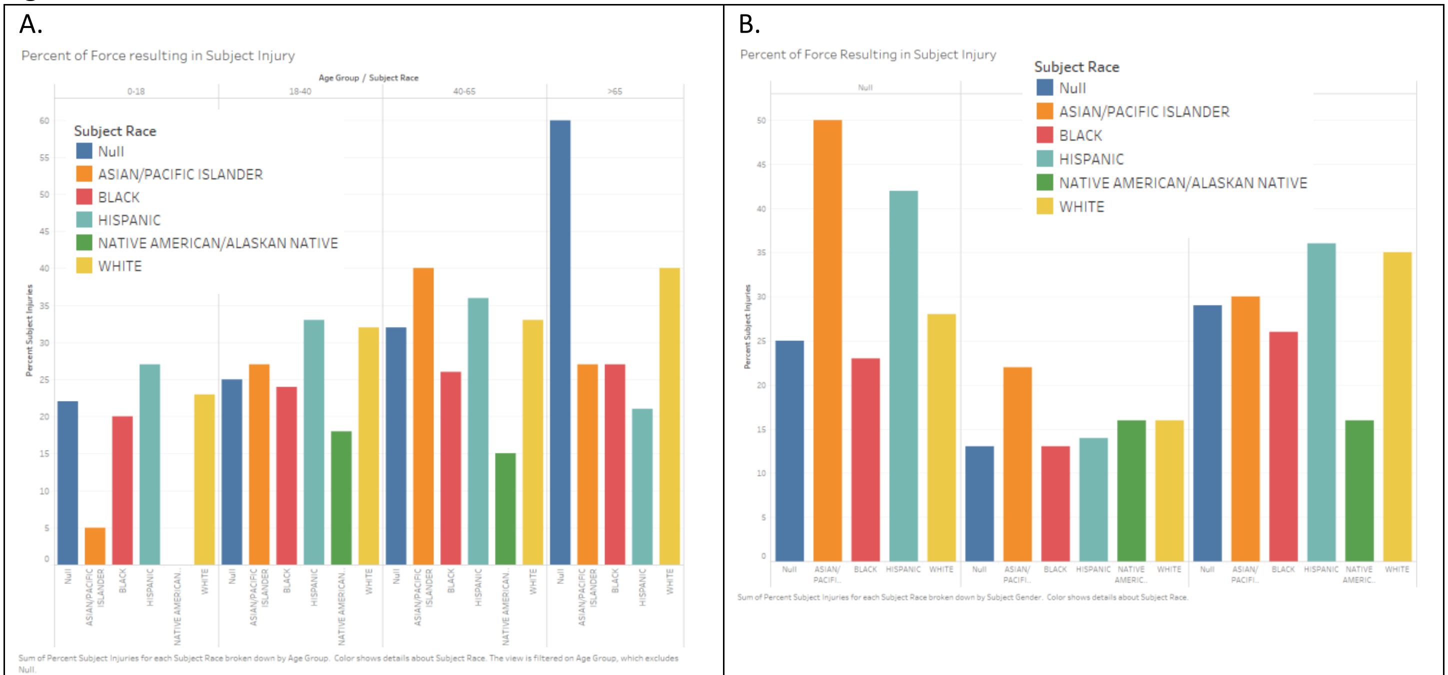
To look at distribution by race we used a circle graph and found that black individual had the highest number of total injuries (Figure 1A), which was expected as they also had the highest number of use of force events enacted against them. Our hypothesis that black individuals would have a higher percent of injuries was incorrect. Figure 1B shows that injuries as a percent of total use of force events was similar across race / ethnicities, with Hispanic and white individuals having the highest percent injury at 33% and 32%, respectively. It is difficult to assess from our data whether these individuals are truly injured more often during use of force events, or whether injuries go under-reported in some groups. This is a questions requires further investigation.

Figure 1:

Circle graphs evaluating total injuries (A) and injuries as a percent of total use of force events (B), stratified by race.

We examined injuries by age and gender as well. When examining injury by age group, we grouped individuals into age ranges: 0-17 years, 18-40 years, 41-65 years and >65 years. Gender and age were further broken down into categories stratified by gender and displayed as bar charts. These visualizations helped identify a few patterns. Young individuals, in the age range of 18-40 years had the highest number of use of force events and injuries (14,815). However, it was the higher age ranges that were injured more frequently when force was used. Individuals in the 41-65 and >65 year age ranges were injured in ~30% of use of force events. There was a greater discrepancy when looking at percent of injury by age when further stratified by race, specifically Hispanic individuals in the 41-65 year age range and white individuals over the age of 65 years, had injuries rates of ~40% when involved in a use of force event (Figure 2A). With regard to gender, men had a higher number of injuries, especially black men. As noted above, Hispanic and white men suffered a higher rate of injury, with rates as high as 35% when stratified by gender in this way (Figure 2B). Again, with such a large number of black men being injured, but at a lower rate, it would be interesting to examine this pattern more with follow-up investigations – it is unclear if this is a true relationship or just based on selective reporting of injury. It does seem plausible that older individuals would have a higher likelihood of injury during any use of force event, just given high susceptibility towards injury in general we age.

Figure 2:



Bar charts, displaying injury rates (as percent of total use of force) by (A) age and race, and by (B) gender and race.

<u>Question 2 (CP2)</u>: How are injuries from use of force events distributed geographically and are there differences in injury rates by district?

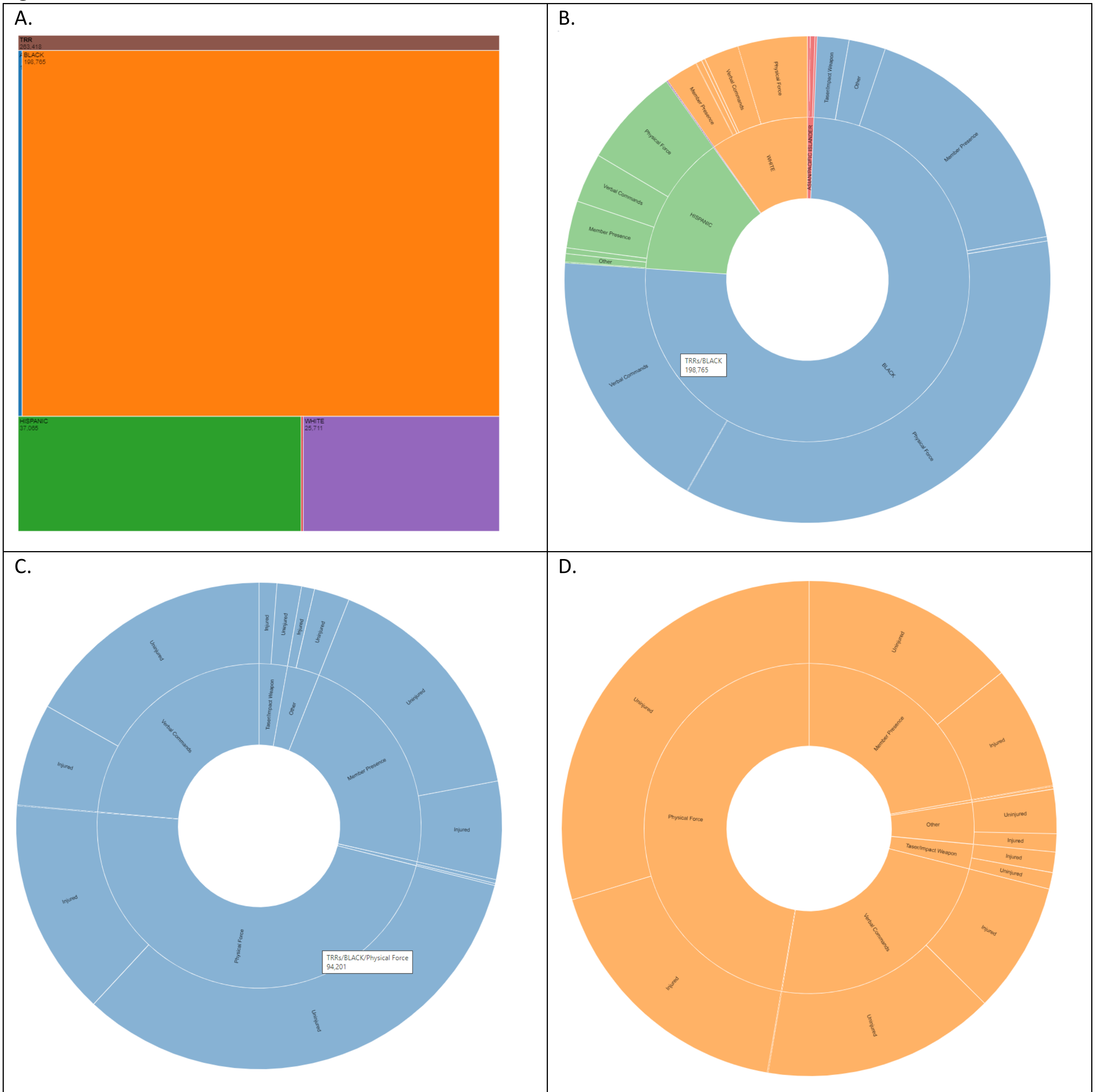
To answer this question, we looked at total injures at the police beat level. We then aggregated beats into districts to better reflect neighborhood lines and possibly demographics. With this district level data, we were able to see that some districts had a much higher rate of use of force events and therefor total injuries, specifically district 11. This district is known as the Harrison District and is one of the most violent in Chicago. There were more murders in this district in 2019 (70) and year-to-date 2020 (90) than any other district in Chicago. This area has been analyzed in the past for this very reason, and hypotheses for the high rate of violence and crime include a high level of poverty, high rates of drug trafficking, and numerous vacant lots and empty housing where these operations are being direct. Chicago has been trying to make investments into this area in an effort to reduce violence here. It is not necessarily surprising then that there are high levels use of force events and injury in this district. More interestingly perhaps, is that injuries as a percent of total use of force events were higher in other districts, especially 14, 16 and 9. The etiology of elevated injury rates in these districts is unclear and is another finding that would benefit from further investigation. These districts could possibly fit demographic profiles for individuals more likely to have an injury (elderly white and Hispanic) but more information is needed prior to any conclusion.

<u>Question 1 (CP3)</u>: What are the relationships between individual race, type of use of force, use of weapon and sustained injuries (alleged or documented)?

We analyzed these complex connections with interactive graphs, specifically starburst maps, treemaps and circle packing graphs. The most significant finding when looking at these graphs, is an appreciation for the discrepancy in total injury numbers for black individuals. Proportional graphs like these really display uneven numbers quite dramatically (see Figure 3, A and B). When clicking into the individual groups, stratified by race, the numbers appear more even, which mirrors the findings from the questions above, the percent injury is more evenly distributed among race. Another key observation that is made from clicking through to the second lay of these interactive graphs are that the rates of type of force used tends to be similar between groups (see Figure 3, C and D). Specifically, physical force makes up about 50% of the use of force events and verbal events make about 25%, this relationship appears to be true in all race / ethnicity categories. Utilization of a taser or other weapon was low, and rates similar, for all race / ethnicity groups (see

Figure 3, C and D) – it would have a notable finding if a specific demographic had experienced higher rates of use of force with by means of a physical weapon.

Figure 3:

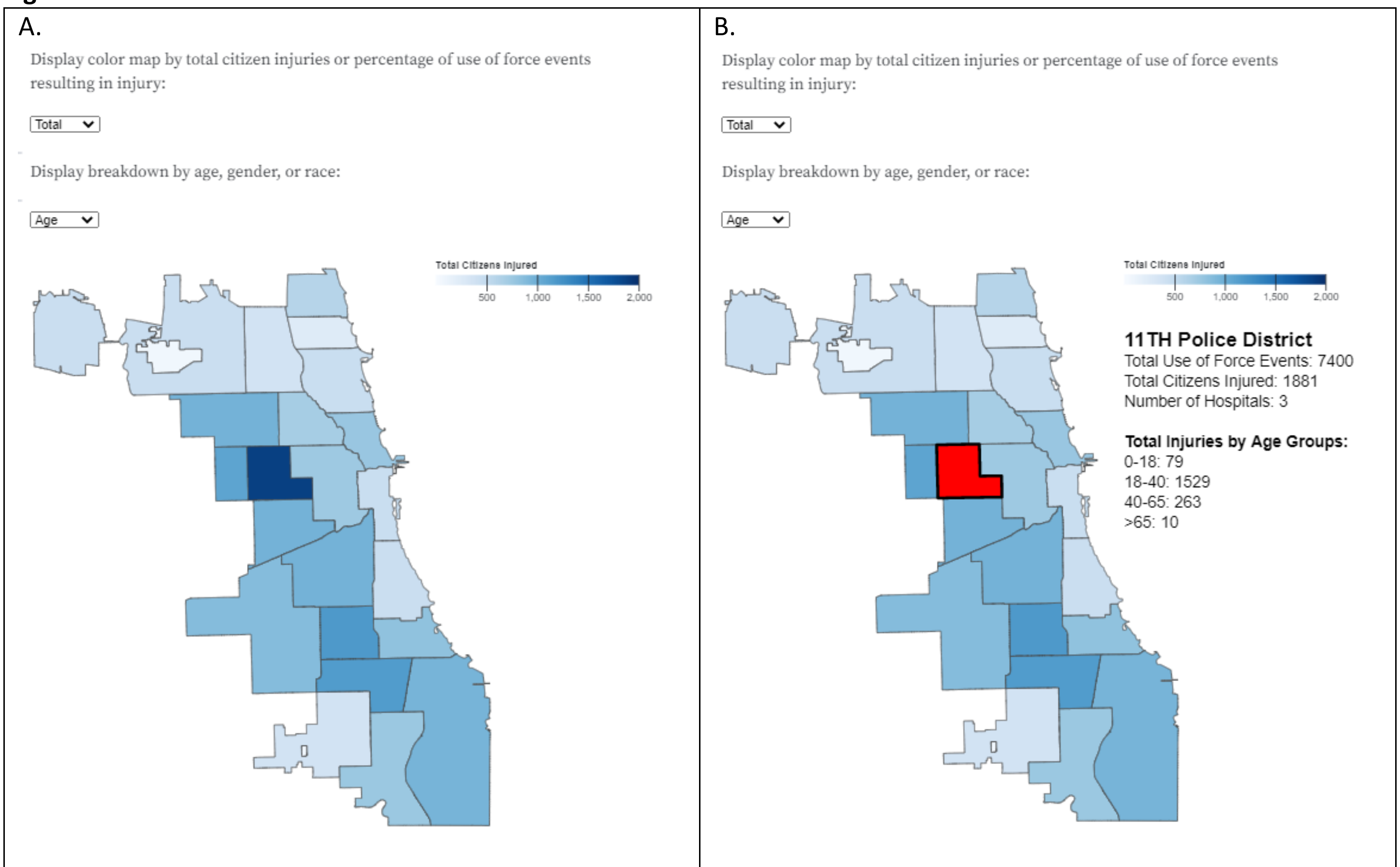


(A) Treemap showing use of force events by race / ethnicity (orange = black, purple = white, green = Hispanic, red = Native American, blue = Asian); (B) Starburst map showing use of force events by race / ethnicity (blue = black, green = Hispanic, orange = white, red = Asian, purple = Native American); (C) Starburst map showing types of force used and injury rates for black individuals; (D) Starburst map showing types of force used and injury rates for black individuals.

<u>Question 2 (CP3)</u>: How are injuries from use of force events distributed geographically and are there differences by demographic group or Hospitals in a specific district?

We created an interactive choropleth map that shows number of injuries or percent of injuries (controlled by a dropdown menu) by district, as discussed in Question 2 (CP2) above, but will also display a number of other statistics when each individual district is selected. Specifically, clicking a district will display the name, total use of force events, total injuries and number of hospitals. There is also a second dropdown menu that allows for toggling between district level data as stratified by race, gender and age (see Figure 4, A and B). The findings from this display are similar to those discussed above, namely the high number of injuries and total force events in district 11 and high rate of injury in districts 14, 16 and 9. Information that is new in this graphic are the number of hospitals per district. The districts with the most hospitals were 12 and 19, each with 6 hospitals. Our hypothesis was that districts with more hospitals may be more likely send victims of an injury in for medical care and therefor to have higher rates of injury reporting. This hypothesis was wrong, and district 12 and 19 were not among the districts with high injury rates.

Figure 4:



(A) Choropleth interactive map showing total injuries by district or percent of injuries by district with total represented as a color scale. (B) Choropleth interactive map with a specific district chosen showing the selected statistics.

<u>Checkpoint 2 and 3 Conclusion</u>:

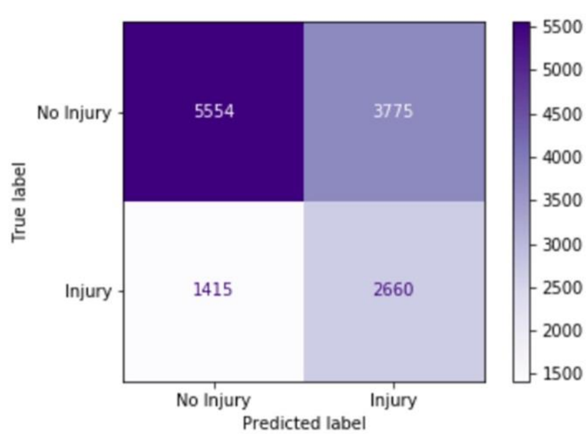
The visualization and interactive features served to highlight a lot of the patterns identified in checkpoint 1. It shows, sometimes in dramatic fashion, the heavy burden of use of force events and injuries that are experienced by the black community. Importantly, however, the high number of injuries is directly correlated to the high number of use of force events. Breaking injuries down into rates, there doesn't appear to be a greater rate of injuries in the black community. In fact, it was the Hispanic and white categories that had the highest rates of injury noted in the data. This is an important area for further investigation as it would be interesting to confirm these findings or see if this is due to other phenomena such as selective reporting in certain districts or for certain groups. We did find, not surprisingly, that older individuals suffered the highest rates of injury if involved in a use of force event even younger individuals had more total injuries. The district with the highest number of use of force events also had the highest number injuries, and happens to be considered one of Chicago's most violent districts given rates of crime, drug arrests and murder. This district was not the once with highest injury rates however, which was another area we identified for further investigation – why do some districts have higher rates of injury then others? Finally, the presence of a hospital and the number of hospitals did not appear to be related the number of rate of injuries in the area.

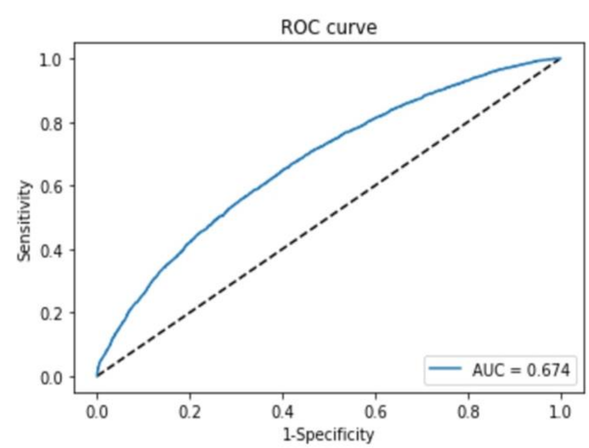
Checkpoint 4: Machine Learning
For this checkpoint, we sought to delve deeper into our data to go beyond simply recognizing trends in the data to using the data to predict important medically relevant outcomes such as subject injury and police officer use of lethal force (e.g. firearm use).

<u>Question 1 (CP4)</u>: Predict police encounters (based on citizen and police demographics, alleged crime, neighborhood) where use of force incidents are likely to result in injury.

For our first question, we were interested in predicting which use of force events that would result in injury. To answer this question we employed a strategy of testing several different machine learning algorithms and choosing the best performing model. We used subject injury as a label/target and used 19 hand-picked features felt likely to be relevant for predicting the outcome of interest including demographics (gender, race, and age), lighting condition, indoor or outdoor, weather condition, officer in uniform, officer injured, officer rank, and type of use of force (firearm, taser/impact weapon, chemical weapon, other weapon, member presence, and verbal commands). We used 60%/20%/20% split of 67,019 tactical response reports for train/validation/test sets and compared the performance of 3 different models on the test set for predicting subject injury – random forests, penalized logistic regression, and K nearest neighbors. Hyperparameters were manually tuned using the validation set. We noted that there was significant class imbalance in our dataset with only 4075/13404 subjects having been injured. There are many approaches to handle the problem of imbalanced classes, the so-called rare event phenomenon, including oversampling methods and class weighting, whereby examples from the minority class are weighted more heavily during the learning process in order to force the model to optimize discriminative performance (https://www.tensorflow.org/tutorials/structured_data/imbalanced_data). We used class weighting to account for this class imbalance problem.

The highest performing model was a random forests model with an accuracy of 61%, precision 41%, recall 65%, F1 score of 51%, and area under the receiver operating characteristic curve (AUC) of 0.674. There were more false positives then true positives, which means that the model was very sensitive for predicting injury, though at the cost of a high misclassification rate. By examining the AUROC you get a better sense of the models true discriminative performance; the black dotted line represents a classifier that is no better than a coin toss and our model was slightly better than this, suggesting it is not that successful of a classifier.



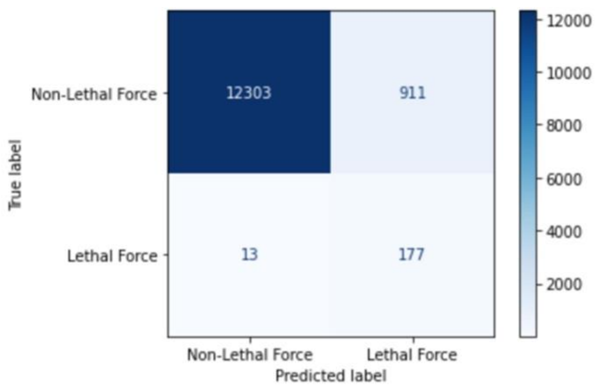


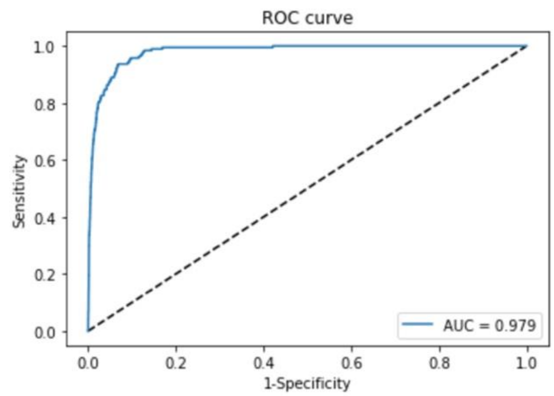
Unfortunately, it doesn't appear that the features we included were strong predictors of injury in this dataset based on our modeling. Of the data that we did assess, the strongest features using impurity-based feature importances were subject age (0.129) and officer age (0.122) (Appendix 4).

<u>Question 2 (CP4)</u>: Predict police encounters more likely to result in lethal uses of force vs. less-than- lethal uses of force.

For this question we were interested in predicting the factors that could result in a police officer deciding to use lethal force. To answer this question, we employed a different strategy. Rather than using a breadth of models, we used the best performing model from our first question (random forests model) and explored a breadth of hyperparameters using grid search for automated hyperparameter optimization. Features considered were the same as for the first question, with the exception of dropping subject injury and firearm use. For this problem, we still set aside 20% of the data for testing, but used 5-fold cross-validation for hyperparameter tuning. Once again, to handle the problem of severe class imbalance (~1,000/67,000 with firearm use) we used class weighting.

The resulting model exhibited remarkable performance for predicting lethal use of force, with an accuracy of 93%, sensitivity (recall) of 93%, and specificity of 93%. While precision was low at 16%, we specifically tuned the model to detect as many lethal use of force events as possible, as for this problem we felt that false negatives (missing a potential lethal use of force event) were more costly than false positives. The model also showed exceptional discriminative performance with an AUC of 0.98.





To no surprise, the most important features for lethal use of force were if the subject was armed and whether the officer engaged with the subject physically. Other factors such as the subject's age, officer age, and subject gender also provided some predictive weight (Appendix 5).

Checkpoint 4 Conclusion:

Interestingly, while predicting subject injury by a police officer proved to be a difficult task, predicting lethal use of force was easier. This makes some intuitive sense, as subject injury can come in many forms and can range from severe to very minor injury, whereas lethal use of force is a severe, hard outcome. Additionally, documentation issues may account for these differences, as injury documentation by an officer may be incomplete, whereas it is difficult to avoid paperwork after deploying one's firearm. We had previously observed that while black subjects are involved with by far an overall greater number of use of force events, they are not more likely to be injured in a given use of force event. It was therefore no surprise that race was not a strong predictor of subject injury during a use of force event. The two strongest predictors were actually subject and officer age – looking back at our initial checkpoint, we did observe that older patients were more likely to be injured (likely because they are more frail) when involved in a use of force event with an officer, and younger officers are more likely to injure a subject in a use of force event (possibly due to younger officers being more aggressive). When it comes to predicting lethal use of force, it is logical that an officer is more likely to brandish his/her firearm when a subject is armed and when engaging physically with that armed subject. However, the order of events is not entirely clear (e.g. does physical engagement preceed firearm exposure by subject and officer?). While there are some times that officers must brandish their gun in response to armed subjects there is also the potential for scrutiny of gun use as an escalation of physical force. This could serve as a starting point for further research into officers that are frequently involved in physical conflict, as they may be a higher risk for using lethal force in the future.

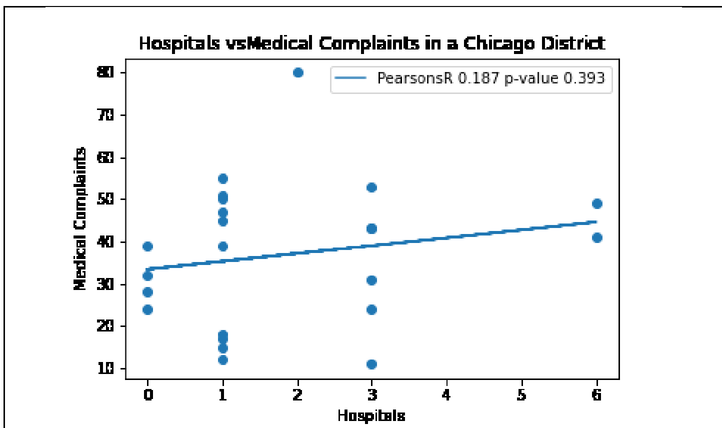
Checkpoint 5: Natural Language Processing

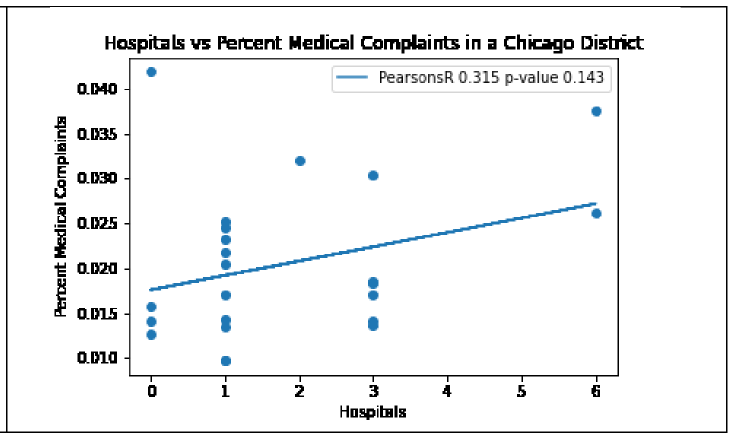
Question 1 (CP5): Use NLP from the narratives within the CPDB to identify encounters that result in emergency medical care and if possible mode and outcome of that care – EMS (ambulance), hospital admission, emergency room.

Unfortunately, we recognize that alleged injury by subjects in use of force events is not always documented by an officer in the tactical response report. We were also interested in identifying encounters with police officers that might result in encounters with the medical system, including emergency medical care and hospital admission/medical attention. To this end, we analyzed complaint reports from lawsuit filings, using NLP tools to identify mentions of medical care. Text from these reports was preprocessed for analysis including removing punctuation and line breaks, converting to lower case, lemmatization, removing stop words, and tokenization.

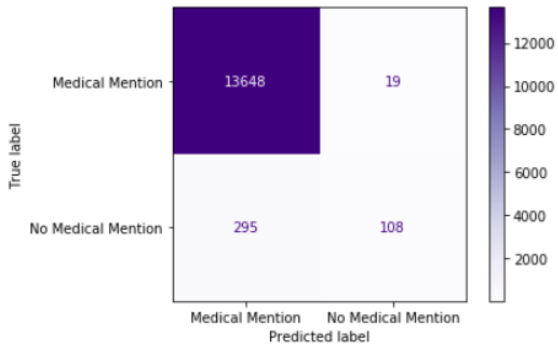
Following text preprocessing we used 2 approaches to identify medical mentions in the complaint reports:

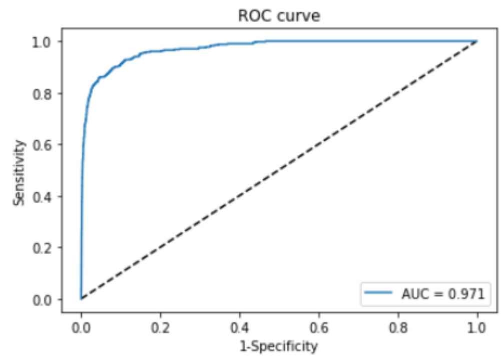
First, we used a semi-supervised approach (essentially using a form of active learning) by first identifying words from the corpus that might signify a medical encounter. We first built a regex of hand-picked words that we felt would be indicative of medical attention (e.g. 'EMS', 'hospital', 'ambulance'). Next, we trained word2vec word embeddings using continuous bag of words to create meaningful word vector representations based on context. Via an iterative process, we were able to identify additional terms that might signify medical interaction by calculating cosine similarities to words in our original regex to words from the entire corpus. We were able to then enrich our regex list to include additional terms we did not initially consider, including for example "surgery" and "fracture". Using this regex list, we then identified all of those complaints where there was at least one mention of a medical term. These were used as the positive label for medical mention. We looked for an association between hospital locations and the medical complaints by plotting the percent of narratives and total narratives containing medical complaints in each district by the number of hospitals in that district. The results indicated a trend towards a significant correlation for percent of complaints containing a medical mention with number of hospitals in a district (Pearson correlation coefficient with 0.315 and a p value of 0.143). While this is not statistically significant it is possible with gold standard labels or curated data that we may find a correlation. This could indicate that people in or closer to a medical district have more medical problems, are seeking medical care, or that police exhibit less restraint in these districts.



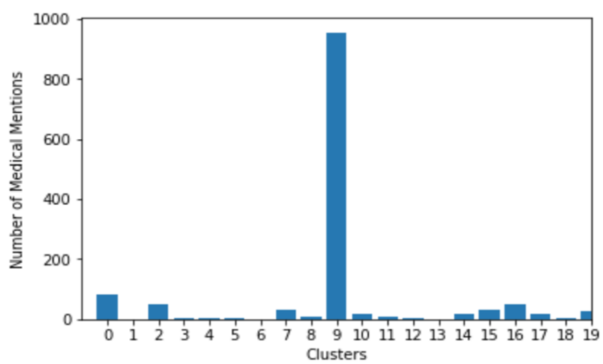


Next, we trained a naïve bayes classifier to create a predictive algorithm for classifying complaint reports as 'medical complaint' or not. The goal of training this classifier is to create a robust model to identify medical mentions in future (not yet seen) complaint reports without having to establish a simple regex list. Notably, we used our labels defined as above to train this classifier given the difficulty in labeling such a large dataset, but future efforts could be made to label a portion of the corpus (for example using crowdsourcing applications like Amazon's Mechanical Turk). To develop input features for this model, we used a term frequency–inverse document frequency (TF-IDF) bag-of-words model to build vectors from the corpus vocabulary. After splitting the preprocessed data into a train set and test set, the vectorizor was applied to the training data to create a feature matrix. After training, we applied our tuned model to the testing set for assessment. The naïve bayes classifier was very accurate, at 98%. There were very few false positives, which is reassuring as we did not want to misclassify complaints inappropriately as seeing medical attention. AUROC shows the model's results for sensitivity and specificity, with strong performance in both.





Finally, recognizing the difficulties in labeling such a large corpus and the limitations of using labels derived from simple information extraction using a regex search, we used an unsupervised approach for topic modeling to identify a topic cluster that might correspond to medical mentions. We used 4 different approaches to unsupervised clustering including latent Dirichlet allocation (LDA), word2vec embeddings/PCA/K-means clustering, BERT embeddings/PCA/K-means clustering, and doc2vec embeddings/K-means clustering. Interestingly, the "shallow" ML approach of LDA seemed to perform best in terms of identifying a cluster that corresponds to medical mentions. We found that the ideal number of topics by LDA using perplexity score was 20. Below is each topic graphed by number of medical mentions derived from our regex derived labels.



Cluster 9 clearly contained the majority of complaints with medical mentions. Looking more closely at this cluster, we determined that this topic cluster included complaints that were more violent/physical than others (see Appendix 6 and 7). While these results are promising, this cluster included over 6,000 complaint reports – therefore, this method was a sensitive, but imprecise way to identify medical mentions. Nonetheless, this method could be used in the future in an active learning framework, to identify reports in this cluster which might benefit from manual review and labeling.

Checkpoint 5 Conclusion:

This checkpoint carries forward observations from previous checkpoints to include actual mentions of interactions with the medical system. Interestingly, we found the most successful approach given lack of labels to identify subject complaints with medical mentions was a simple method using word embeddings in a semi-supervised approach to build a regex, then using simple search to identify those reports with medical mentions. Of course, there are weaknesses to this approach, as not every mention of a medical term corresponds to medical interaction, these terms are likely not exhaustive, and this does not consider negation. We used a Naïve-Bayes model to build a classifier, but this method relied on labels derived from our regex list. Perhaps a way forward is using an unsupervised approach to identify topic clusters for manual labeling using a crowdsourcing tool. More advanced NLP methods using transformers were not successful for this information extraction task, however, alternative methods including using sentence rather than paragraph vectors and exploring different clustering algorithms for topic identification may improve performance.

Medical mentions were surprisingly sparse. Out of over 50,000 complaints, only a little over 1,000 contained a mention of medical attention by our regex method. Once we can confirm the accuracy of these labels, these medical mentions can further be used to explore the relationship between subject injury, medical attention, and reporting of injury by officers in the tactical response reports.

Discussion:

This project made significant observations to the factors involved in subject injury throughout the CPDB, and developed ways to predict the types of situations that are more likely to result in injury, and furthermore the type that result in the necessity for medical care.

The demographics involved in use of force resulting in injury are of particular concern. Blacks in Chicago are subject to much more total use of force and therefore much more total injury. However, the rates of injury when stratified upon race are generally similar at 25-30%. This is important when considering how to study and fix this discrepancy in the future. First root cause analysis must be performed internally at the Chicago police department to determine why force is initiated in this case and how to prove if it is appropriate or inappropriate as this discrepancy should otherwise not exist. A similar analysis should be done in regards to sex as it should be possible to police genders equally, without injuring males more often, especially when they do not lead to significantly more officer injury.

Next, the individual aspects of policing must be examined. When broken down by district level, we can see that the 14th police district causes injury in 35% of reports, whereas the 22nd police district causes injury in only 22% of reports. Similarly, individual officers with a higher propensity for injury should be looked at, although reform in individual officers may have a smaller impact on the community than identifying district and command wide problems. Efforts to inspect officers from the way they file reports to the way they conduct themselves in an encounter that requires force should be undertaken. At best, this discrepancy could be a difference in standards of filing forms and reporting, at worst, it could show that some commands are more motivated to cause injury to those that are being apprehended. Regardless this sort of discrepancy amongst peers would not be allowed to go unnoticed and unexplained for in any other sector of society.

A particular area of scrutiny that our machine learning algorithm has evaluated is predicting officer use of firearms. This class balanced random forest can determine based upon these factors somewhat reliably if a firearm will be used. While some cases may justify a firearm (such as the subject using a firearm), other factors including age, and the officer using physical force, that were important in prediction of firearm use should not. These situations and observations could be included in training as ways to curb potential further firearm use and injury. While our model for injury was less accurate it did find correlations between the age of the officers, the age of the subjects and the propensity to cause injury when subjects were older.

Finally, we were able to label, through NLP, allegations that involved medical issues. Using these labels we were able to make a Naïve Bayes Classifier capable of labeling future allegations. While infrequent in total allegations these represent severe claims that should be taken seriously. As we have seen in national events, medical emergencies happen in police custody due to pre-existing conditions or due to sustained injury, and are not always handled properly as police do not have extensive medical expertise. Simply by accurately labeling these allegations we can draw attention to events that, regardless of outcome, have the potential for lethality. This technique could be trialed when allegations that are received so that they could undergo appropriate scrutiny and quality improvement.

Appendix 1: Subject injuries broken down by officer and subject race

subject_race	officer_race	percent_subject_inju... 1	percent_officer_injuries
HISPANIC	Asian/Pacific	40	24
WHITE	Hispanic	35	25
HISPANIC	White	33	21
NATIVE AMERICAN/ALASKAN NATIVE	Black	33	0
<null>	Native American/Alaskan Native	33	33
HISPANIC	Black	32	23
HISPANIC	Hispanic	32	23
ASIAN/PACIFIC ISLANDER	White	31	19
WHITE	Asian/Pacific	31	25
WHITE	White	31	23
HISPANIC	Native American/Alaskan Native	30	21
<null>	Hispanic	30	24
WHITE	Native American/Alaskan Native	29	29
<null>	Black	28	23
WHITE	Black	27	25
ASIAN/PACIFIC ISLANDER	Black	26	12
<null>	White	26	23
BLACK	Black	25	22
BLACK	Hispanic	25	23
BLACK	Native American/Alaskan Native	25	24
BLACK	White	24	22
BLACK	Asian/Pacific	23	22
ASIAN/PACIFIC ISLANDER	Asian/Pacific	21	25
ASIAN/PACIFIC ISLANDER	Hispanic	21	18
<null>	Asian/Pacific	20	29

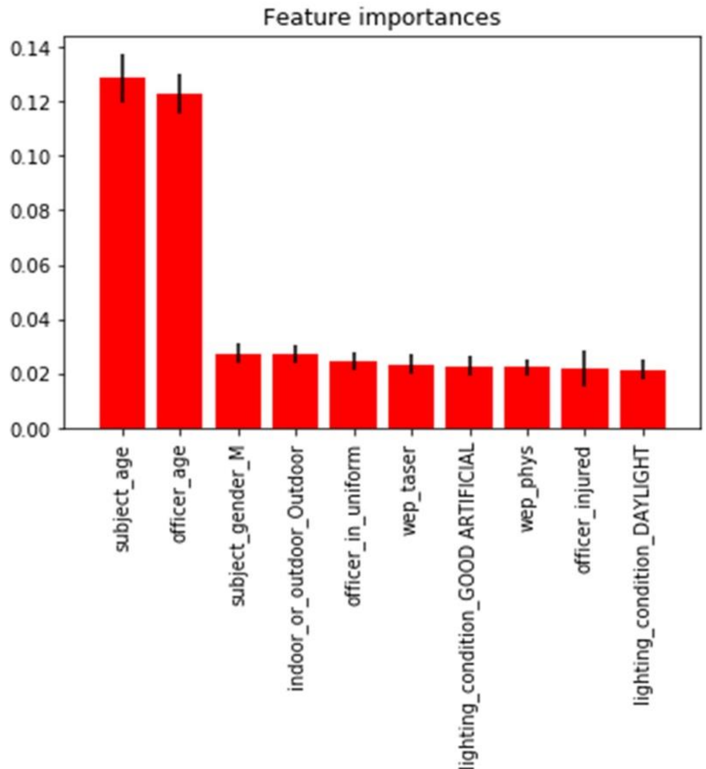
Appendix 2: Officer and Subject Injury by beat

	beat	total_events	officer_injured	percent_officer_injured	subject_injured	percent_subject_injured
1	621	854	190	22	182	21
2	1112	816	173	21	200	24
3	1134	881	160	18	251	28
4	713	718	155	21	173	24
5	1533	663	152	22	192	28
6	1522	679	144	21	172	25
7	1824	599	135	22	178	29
8	624	623	135	21	170	27
9	1822	460	132	28	106	23
10	531	618	131	21	151	24

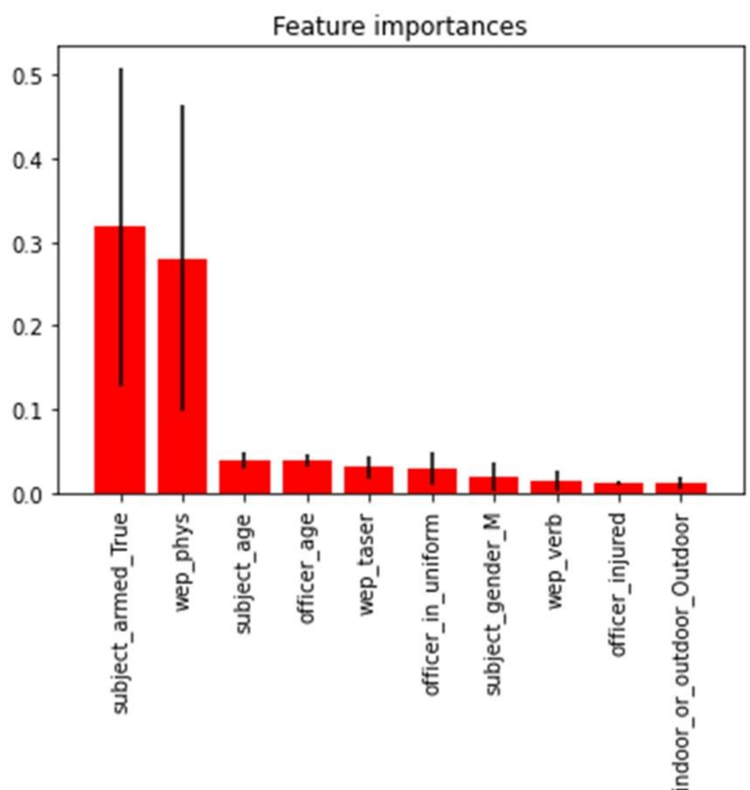
Appendix 3: Alleged Injuries not counted per officer.

	officer_id	alleged_injuries	subject_injuries	percent_subject_alleged_injuries_not_counted
1	18894	10	5	50
2	22929	11	6	45
3	30290	19	11	42
4	26846	10	6	40
5	30352	10	6	40
6	4549	15	9	40
7	25155	10	6	40
8	13473	13	8	38
9	14706	11	7	36
10	8176	11	7	36
11	25177	11	7	36
12	11257	14	9	35
13	26018	12	8	33
14	31127	10	7	30
15	22392	10	7	30
16	15987	10	7	30
17	18959	10	7	30
18	28797	10	7	30
19	3082	10	7	30
20	22150	13	9	30

Appendix 4:



Appendix 5:



Appendix 6:

Top words by TF-IDF in each topic cluster

```
Topics in LDA model:
Topic #0: fail provide vehicle traffic return license driver phone inventory accide
nt
Topic #1: offender weapon situation type armed batter flee authorization discharge
service
Topic #2: state regard case respond fail unknown threaten telephone time manner
Topic #3: false use direct profanity bag proper robbery murder statement pant
Topic #4: remove order child action court contact school landlord document effect
Topic #5: record motorist photo lane change mr ago ms motorcycle thompson
Topic #6: area time present september january james basement john lllinois duty
Topic #7: citation witness issue state incident justification fuck supervisor tell
ticket
Topic #8: file complaint state information steal girlfriend obtain occur pocket sel
l
Topic #9: tell place arrest home handcuff gun charge arrive battery year
Topic #10: search residence warrant district justification th damage apartment prop
erty door
Topic #11: arrest falsely justification member department cause possession probable
family burglary
Topic #12: daughter rd father comment pretense mail unwarranted ts cite demeanor
Topic #13: enter officer home business discover check tenant computer burglarize le
tter
Topic #14: car drive strike street pull stop sign squad light traffic
Topic #15: hour service approximately unit duty itis employee march fail july
Topic #16: vehicle male stop white unknown harass uniformed black state subject
Topic #17: number fail star request complete attempt inventorie reference behalf af
fidavit
Topic #18: location work leave department hour june february november october emplo
yment
Topic #19: refuse rude unprofessional state tell allow want ass leave request
```

Appendix 7:

```
after work day construction contractor cadle visit friend home south cadle park car
friend home begin talk friend garage soon approach ask car garage cadle reply deman
d license registration cadle ask question falsely traffic violation deny allegation
cadle provide license registration friend ask cadle cooperate cadle allow search ca
r sobriety test pass find contraband cadle car falsely open alcohol backseat grab c
adle slam nearby gate shove ground strike rib cadle feel pain chest difficulty brea
the handcuff cadle shackle leg throw backseat car drive cadle station lock despite
repeat request medical care hour transport cadle paddy wagon hospital cadle diagnos
e puncture lung break rib fractured collarbone cadle falsely charge resist arrest t
raffic citation include illegal transportation alcohol month charge citation cadle
drop

mccambry purse vehicle bailey pursue mccambry foot  bailey marked vehicle east  th
street south calumet avenue bailey overtake mccambry drive sidewalk strike injure b
ailey vehicle mccambry tell hip break need hospital bailey ignore plea yell mccambr
y beat bailey drag foot squad car leano nichols arrive drive mccambry station mccam
bry continue request care  tell  shut  mccambry arrive station walk nichols lean
o drag station handcuff wall mccambry tell watts extreme pain need medical care wat
ts tell mccambry shut hospital threaten gun case allow hospital hour custody medica
l treatment mccambry hospital diagnose multiple fracture follow month nichols false
ly testify mccambry injure fall fence squad car
```