Why Are Some Complaints More Likely to Lead to Discipline?

Jacob Beran

Theme

I want to investigate why some complaints wind up being used to discipline officers and why others are dismissed without disciplinary action. Looking at factors such as previous misconduct allegations, misconduct allegation category, seniority, complaint location, network proximity to other officers accused of misconduct, and complaint sentiment could illuminate how circumstances change the disciplinary and legal reactions to accusations of police misconduct.

It is important to examine factors behind police disciplinary decisions because it could lead to a more just system for disciplining officers and keeping them accountable while also illuminating biases and unjust disciplinary practices currently occurring.

Checkpoint 1: Relational Analytics

The goal for this section is to numerically explore different factors that lead to officer discipline or sustained allegations.

	percent_allegations_given_disciplined_prior	percent_allegations_given_undisciplined_prior
1	35.64217610768368	44.559040202062725

Figure 1: Percentage of officers with an allegation given that the officer was or was not disciplined before 2007

The above table (Figure 1) shows the percentage of officers that had allegations of misconduct against them given that they were disciplined before 2007. Interestingly, those that were not disciplined but had at least one misconduct allegation against them had a much higher rate of getting more allegations. Some of this could be due to officers that are disciplined either being monitored more closely, behaving better to avoid further discipline, or being removed from the police force. This trend does show us that an officers disciplinary status may not necessarily indicate their likelihood of being accused of misconduct.

	percent_sustained_given_disciplined_prior	percent_sustained_given_undisciplined_prior
1	3.954010095344924	3.9360134708482426

*Figure 2: Percentage of officers with an allegation **sustained** given that the officer was or was not disciplined before 2007*

Figure 2 shows that there is very little difference in whether an officer has their misconduct allegations sustained or not based on a previous discipline before 2007. It seems that previous disciplinary actions do not have a significant impact on whether an officer will be disciplined again, will receive a misconduct allegation again, or will have a misconduct allegation sustained.

According to the data in Figure 3, 6.6988859% of non-officer complaints were sustained in total. This is a useful number to use as a baseline.

	percent_sustained	total_allegations	sustained_allegations
1	6.698885927663934	(235263)	(15760)

Figure 3: Percentage of allegations sustained against officers.

	number_of_allegations	number_of_officers	total_allegations_sustained	percent_sustained
1	1	3139	828	26.377827333545717
2	2	2175	818	18.804597701149426
3	3	1808	841	15.505162241887904
4	4	1549	833	13.444157520981278
5	63	12	99	13.095238095238097
6	5	1350	794	11.762962962962963
7	7	1202	940	11.171856429759924
8	6	1218	814	11.13847837985769
9	9	903	872	10.729666543620032
10	8	1020	872	10.686274509803921
11	11	740	799	9.815724815724817
12	12	676	758	9.344181459566075
13	10	826	755	9.14043583535109
14	18	369	594	8.94308943089431
15	17	400	592	8.705882352941176
16	13	595	673	8.70071105365223
17	14	543	660	8.681925808997633
18	23	244	485	8.642195295794727
19	15	496	631	8.481182795698924
20	42	36	124	8.201058201058201

Figure 4: Percentage of allegations sustained grouped by number of allegations, top 20 in percentage of allegations sustained.

Figure 4 has columns that show the number of allegations which act as our bins, the number of officers that have had that number of misconduct allegations, the total number of misconduct allegations sustained, and the percentage sustained. As can be seen above, officers that have fewer misconduct allegations seem to have a larger number sustained. Additionally, officers that have many misconduct allegations seem to sustain a much lower percentage of them, as can be seen below. These findings seem to support the idea that officers that sustain many misconduct allegations are held less accountable, which is disappointing since you would hope that they would actually sustain more allegations as it became clearer that their behavior was clearly inappropriate and harmful to the communities they are policing.

The data in Figure 5 includes officer allegations as well as civilian ones. Data incompleteness aside, it seems that the higher-ranking officers sustain significantly less complaints when compared to other ranks of officers. This supports the hypothesis that accountability is more difficult the further up the promotion ladder you go and that police officers may be more willing to abuse their disciplinary power so that they will not be disciplined in similar situations. While these percentages are not necessarily surprising, they are disappointing. More research into this could include graphical representations of this data and even a breakdown of which misconduct categories higher ranking officers find themselves involved in.

The big takeaways from checkpoint 1 were that previously disciplined officers were less likely to have allegations against them but no less likely to have sustained allegations, officers with lower numbers of allegations were more likely to have those allegations sustained, frequently disciplined officers have more complaints and are generally more likely to have those allegations sustained (but the disciplined allegations are all sustained allegations, so those variables are not independent), and percentage of allegations sustained seems to be negatively correlated with officer rank.

	percent_sustained	rank_description
1	30	Unknown
2	26.02829878249424	Police Officer / Mounted Patrol Officer
3	20.43010752688172	<null>
4	14.457831325301203	SR Evidence Technician
5	12.03155818540434	Police Forensic Investigator I
6	11.178247734138973	Police Officer Assign Youth Off
7	9.458023379383635	Police Officer Assigned Traffic Specialist
8	9.397810218978101	Police Officer Assigned Security Specialist
9	9.10209102091021	Police Agent
10	8.713227037951942	Police Officer / Field Training Officer
11	8.622847194813382	Police Officer
12	8.339990413804122	Police Officer Assigned Evidence Technician
13	7.301630068705375	Captain of Police
14	6.848401922604638	Lieutenant of Police
15	6.492454277349868	Sergeant of Police
16	5.872340425531915	Police Officer / Marine Officer
17	5.684796854521625	Commander
18	5.5132390279289085	Gang Crimes
19	5.392198521288349	Police Officer as Detective
20	4.960460100647016	Police Officer / Explosives Detective Canine Handler
21	4.893092105263158	Deputy Chief

Figure 5: Percentage of sustained allegations given officer rank

Ultimately, these results demonstrate that some officers have a lower percentage of allegations sustained against them than others based on how many allegations they have received and seniority. This fact backs up the idea of the "Blue Wall of Silence[1]" since the disciplinary procedures seem to protect officers that have had more allegations against them or have more ranking within the command hierarchy.

Checkpoint 2: Data Exploration

In this checkpoint, we seek to answer to create two different visualizations that will help us explore further explore which factors lead officers to have allegations sustained against them.

The first visualization (Figure 6) is designed to show how both the racial disparities in who submits complaints of officer misconduct and racial disparities in the sustaining of those misconduct allegations. The top percentages show the percentage of allegations that are not sustained and that are sustained, respecitively. The colors show what percentage of the not sustained and sustained pools of allegations for that neighborhood are from that race, so a darker color means that a race contains a higher number of either sustained or not sustained allegations as compared to the other races in that neighborhood. Therefore, the colors show how the race of the person filing the complaint impacts the way in which it is received and ultimately acted on. The table is also ordered in descending order from most total allegations to least, with the cutoff at 1500 total allegations. However, this cutoff includes many allegations with race unlabeled so the totals in the table are generally lower.

[1] https://theintercept.com/2016/10/06/in-the-chicago-police-department-if-the-bosses-say-it-didnt-happen-it-didnt-happen/

Looking at Englewood, for example, we see that only 7.25% of misconduct allegations filed by African Americans are sustained while both Hispanics and White people have their allegations sustained at at least 25%. Additonally, a huge proportion of the not sustained complaints are from African Americans as the color indicates at first glance and the count backs up upon further inspection. However, Englewood is over 94% African American so the number of complaints is proportional to the population and so it may seem that this disparity may be due to the demographics of the neighborhood. However, other neighborhoods have a plurality of not sustained allegations filed by African Americans but are much more diverse, including Rogers Park (27% African American), Lake View (9.2% African American), and River North(10% African American). This disparity suggests a possible bias in the way in which police examine complaints from African Americans as compared to other people and may be a contributing factor to police officers remaining on the force after committing misconduct. Additionally, the fact that the neighborhoods at the top of the list containt the largest raw number of misconduct allegations and are predominantly African American suggests that they also bare the brunt of policing and therefore police misconduct in Chicago, worsening the lives of the people in those neighborhoods. While there is always a possiblity that these all of these allegations are being evaluated fairly and this is just how the numbers shake out, it seems far more likely that there is extreme bias in how allegations are either sustained or not sustained based on the race of the person filing the complaint.

Sustained vs. Not Sustained Allegations by Neighborhood and Race

Name	Race / Final Finding: Black		Hispanic		White		Asian/Pacific Islander		Native American/Ala..	
	Not Sust..	Sustained	Not Sust..	Sustained	Not Sust..	Sustained	Not Sust..	Sustained	Not Sust..	Sustained
Englewood	92.75% 1,356	7.25% 106	64.29% 18	35.71% 10	74.58% 44	25.42% 15		100.00% 2		
Austin	94.37% 922	5.63% 55	57.58% 19	42.42% 14	56.90% 33	43.10% 25		100.00% 2		
Auburn Gresham	93.92% 818	6.08% 53	85.71% 6	14.29% 1	37.50% 6	62.50% 10	100.00% 2			
Garfield Park	97.33% 730	2.67% 20	83.33% 20	16.67% 4	75.61% 31	24.39% 10	66.67% 2	33.33% 1		
Humboldt Park	97.77% 438	2.23% 10	90.91% 150	9.09% 15	87.80% 36	12.20% 5	100.00% 8			
North Lawndale	95.25% 441	4.75% 22	73.08% 38	26.92% 14	73.33% 22	26.67% 8				
Grand Crossing	94.19% 503	5.81% 31			75.00% 6	25.00% 2				
Chicago Lawn	89.47% 340	10.53% 40	94.44% 68	5.56% 4	84.62% 22	15.38% 4			100.00% 2	
Roseland	94.71% 412	5.29% 23		100.00% 4	64.29% 9	35.71% 5				
South Shore	93.98% 390	6.02% 25	25.00% 4	75.00% 12	57.14% 8	42.86% 6				
New City	94.65% 283	5.35% 16	91.30% 63	8.70% 6	94.12% 64	5.88% 4				
South Chicago	92.07% 360	7.93% 31	84.00% 21	16.00% 4	20.00% 2	80.00% 8				
Woodlawn	96.55% 336	3.45% 12	100.00% 5		53.70% 29	46.30% 25				
Chatham	91.52% 356	8.48% 33	100.00% 2		50.00% 2	50.00% 2				
River North	96.12% 223	3.88% 9	84.62% 33	15.38% 6	77.42% 72	22.58% 21				
Belmont Cragin	89.00% 89	11.00% 11	98.37% 181	1.63% 3	75.38% 49	24.62% 16				
Douglas	86.69% 228	13.31% 35	63.64% 21	36.36% 12	41.67% 10	58.33% 14	100.00% 4			
Loop	85.82% 115	14.18% 19	87.50% 35	12.50% 5	81.34% 109	18.66% 25		100.00% 1		100.00% 1
Little Village	92.59% 50	7.41% 4	95.08% 174	4.92% 9	91.11% 41	8.89% 4				
Lake View	94.21% 114	5.79% 7	94.29% 33	5.71% 2	80.18% 89	19.82% 22	100.00% 4		100.00% 6	
Logan Square	76.81% 53	23.19% 16	91.38% 106	8.62% 10	82.54% 52	17.46% 11				
West Pullman	87.61% 191	12.39% 27			76.92% 10	23.08% 3				
Grand Boulevard	91.16% 196	8.84% 19		100.00% 2	83.33% 10	16.67% 2				
Rogers Park	96.88% 155	3.12% 5	100.00% 8		88.24% 30	11.76% 4	50.00% 2	50.00% 2		
Uptown	98.46% 128	1.54% 2	100.00% 12		95.45% 42	4.55% 2	100.00% 4			
Little Italy, UIC	86.59% 71	13.41% 11	63.64% 14	36.36% 8	70.83% 17	29.17% 7	100.00% 1			

% Total by Race: 0.24% – 98.89%

% of Total Count of Final Finding and count of Final Finding broken down by Race and Final Finding vs. Name. Color shows % of Total Count of Final Finding. The marks are labeled by % of Total Count of Final Finding and count of Final Finding. The view is filtered on Race and Name. The Race filter keeps Asian/Pacific Islander, Black, Hispanic, Native American/Alaskan Native and White. The Name filter has multiple members selected.

Figure 6: Sustained vs. Not Sustained Misconduct Allegations by Neighborhood and Race

The second visualization (Figure 7) shows allegations broken down by category. Immediately, what stands out is that Use of Force and Operation/Personnel Violations seem to make up a majority of allegations when combined. However, Operation/Personnel Violations seem to be sustained significantly more often than Use of Force allegations. While we can't tell the specifics of each complaint based on

just this graphic, we can make some guesses as to why Use of Force allegations lead to significantly less sustained allegations than Operations/Personnel Violations and most other categories of misconduct. These reasons include police refusing to hold each other accountable, which seems to be a big theme throughout the qualitative investigations into the Chicago Police Department. Additionally, Operations/Personnel violations seem to be easier to prove than Use of Force cases, which presumably rely on civilian and police accounts of the incident, with the police typically probably siding with their fellow officers unless a video or medical records can prove serious misconduct. This graphic also helps show that the best way to cut down on police misconduct would be to attempt to cut down on the use of force and generally instructing officers to be less combative towards the civilians they police since many of the categories such as Illegal Search, Verbal Abuse, and Use of Force all occur when officers interact with civilians.

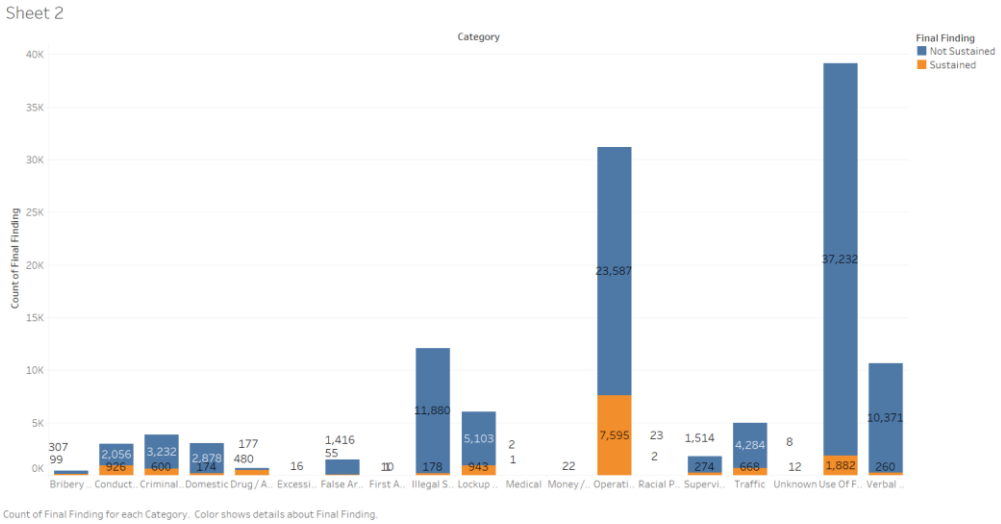

Figure 7: Sustained and Not Sustained Allegations Broken down by Misconduct Category

Checkpoint 3: Interactive Visualization

Heat Map: https://observablehq.com/@trickortea/heat-map-of-disciplined-percentage-for-misconduct-allegat

First looking at all allegation categories and filtering by race, it seems clear that allegations of misconduct do not result in disciplinary action when the complaint is filed by a black person as much as it does when the complaint is filed by a white person. Additionally, complaints filed by Hispanic people also seem to result in more disciplinary action as well. Another notable aspect of this graphic is that there are not many allegations from people who identify as Asian/Pacific Islander, making it difficult to gain insight into how complaints are handled when the race of the person complaining is Asian/Pacific Islander. Breaking down allegations geographically, it appears that there aren't a lot of geographic disparities between areas when the complaints are from someone who identifies as black. However, there are a few areas in the North of the city that are clustered with higher discipline ratios than elsewhere in the city. I am curious as to why this is the case and if it has anything to do with the surrounding area or the specific police officers in those areas. When filtering by white people, it appears that complaints on the south side generated more disciplinary action than complaints on the north side of the city. This geographic disparity might be due to demographic differences in those areas since a police force in a largely minority neighborhood might subconsciously take complaints from white people more seriously since they don't interact with them as much and police just being racially biased wouldn't explain the geographic disparity. The heat map for complaints from Hispanic people also showed a similar pattern with higher discipline ratio in the areas in the south of the city.

When filtering by category with race constant, it is first apparent that there is only a lot of data for the first two categories of misconduct, Operation/Personnel Violations and Use of Force. From the map, it seems that the former results in more disciplinary action than the latter but only for a few areas. Interestingly, the category with sizeable data that lead to the most disciplinary action was Conduct Unbecoming (Off Duty). It is a little disappointing to see that it is easier to hold officers accountable when they are off duty than when they are in uniform since it is clear that Conduct Unbecoming (Off Duty) complaints generate significantly more disciplinary action than Use of Force complaints. Examining different misconduct categories geographically, it seems that there is not a huge disparity in Operation/Personnel Violations disciplinary percentage between the north and south of the city but there is definitely a higher disciplinary percentage towards the east than the west. There were no discernible geographic differences for Use of Force complaints, which is interesting since there were some for other misconduct allegation categories. Interestingly, sustained traffic violations seem to occur more frequently and have increasing discipline the further from the center of the city they occur. This may be due to an increased ability for police to make traffic stops in areas with less traffic but may also lead to a higher number of stops for reasons that could garner a complaint that leads to disciplinary action. Traffic complaints led to disciplinary action a similar amount for all races based on looking at the map as well so race may not be a factor in traffic complaints when compared to other categories.

Ultimately, this graphic is interesting in the context of racialized policing because it shows that the process of acting on complaints is still dependent on the race of the person submitting the complaint and is sometimes dependent on which neighborhoods of the city the complaints occurred.

Checkpoint 4: Graph Analytics

For the following graphs, I chose to only include officers with ten or more allegations because I did not want the data to be clouded by officers with small numbers of allegations. The Allegation Network consisted of edges between officers that were in the same allegation while the Disciplined Allegation Network only included edges from allegations leading to discipline.

Figure 8 shows the results from running the PageRank algorithm on both the Allegation Network and the Disciplined Allegation Network. It is apparent from the graph and the statistical tests that there is no meaningful correlation between the PageRanks in either network for a given officer. A lack of correlation here shows that an officer's PageRank centrality in one network has no bearing on their ranking in another network, so trying to predict an officer's centrality in the Disciplined Allegations Network is not possible with just their centrality in the Allegations Network, making it difficult to predict which officers will become central in the Disciplined Allegations Network.

Figure 9 shows the results of graphing the PageRank in the Allegations Network against the number of allegations resulting in disciplinary action for a given officer. As we can see, there is a moderate positive correlation between the variables. More importantly this correlation is higher than the correlation between allegations and allegations resulting in discipline for a given officer, which can be seen two graphs below in the graph titled *Allegation Count vs. Disciplined Allegation Count*. This is important because it shows that the PageRank centrality of an officer in the allegations network can be more predictive of the number of allegations this officer has been disciplined for and possibly will be disciplined for in the future when compared to predicting an officer's disciplined allegation likelihood using just a fit from the relationship between allegation counts and disciplined allegation counts. This improvement shows that graph analytics can help in the predictive power of which officers may or may not commit misconduct that will lead to disciplinary action.

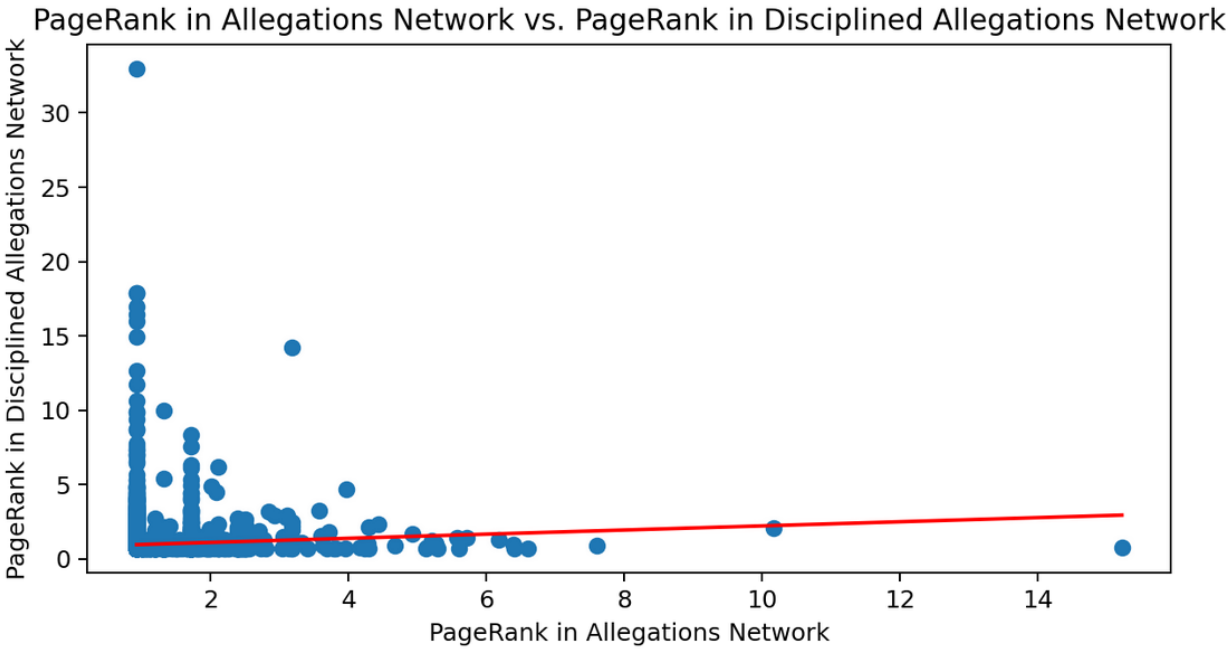

Figure 8: PageRank in Allegations Network vs. PageRank in Disciplined Allegations Network

```
R Squared: 0.003844919275253489
Pearson Correlation Coefficient 0.062007413066935094
Spearman Correlation Coefficient 0.10532044311289324
```

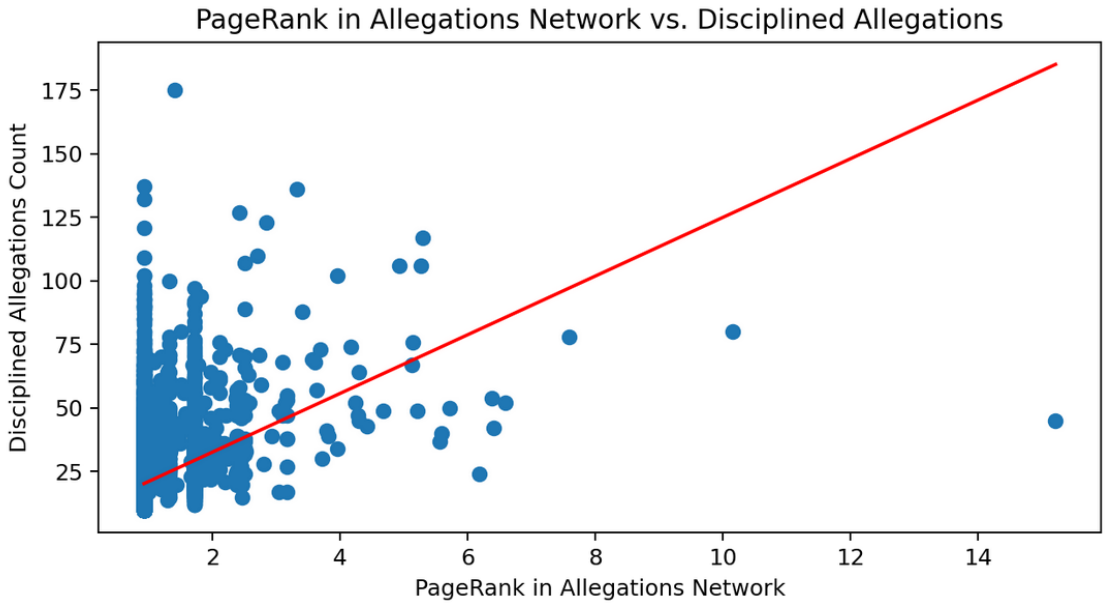

Figure 9: PageRank in Allegations Network vs. Disciplined Allegations

```
R Squared: 0.10905566180371778
Pearson Correlation Coefficient 0.3302357669964258
Spearman Correlation Coefficient 0.3219471033631776
```

Figure 10 serves as a baseline for the other relationships so that it is easier to see which methods of graph analysis will lead to better predictions of which officers will commit misconduct. Allegation and Disciplined Count have a moderate positive correlation as it would make sense that officers that have more allegations against them would also probably have more disciplined allegations as well.

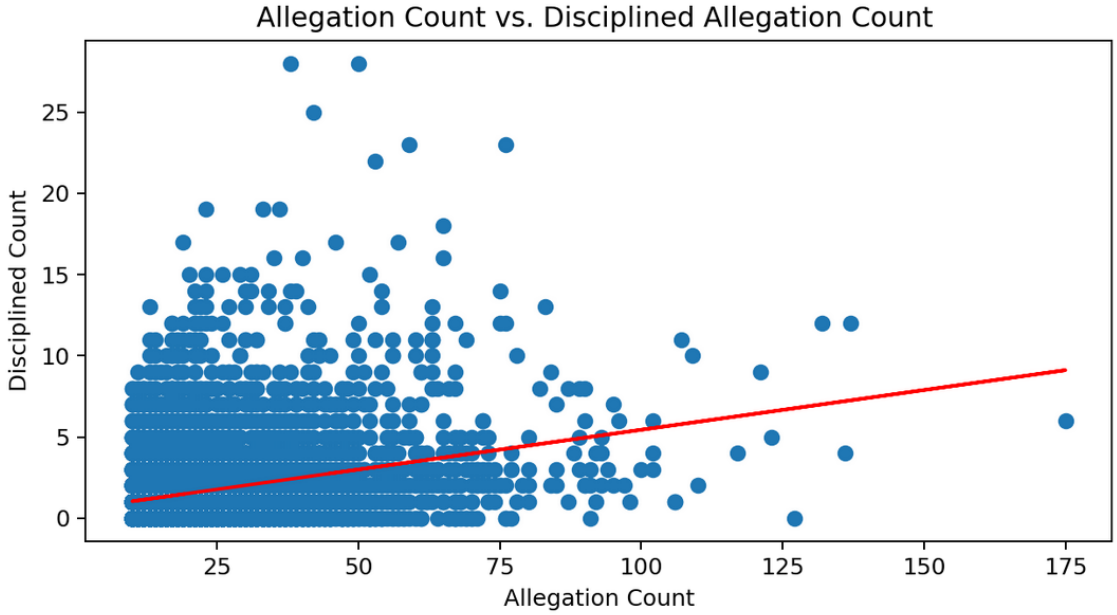

Figure 10: Allegations Count vs. Disciplined Allegations Count

```
R Squared: 0.08547698751276421
Pearson Correlation Coefficient 0.29236447717321
Spearman Correlation Coefficient 0.24444004127077873
```

Ultimately, the degrees of a given officer node are not useful for predicting whether they will be disciplined. However, their PageRank in the Allegations network has a moderate positive correlation that is higher than the correlation between allegations and allegations resulting in discipline, showing that there is good reason to believe that utilizing PageRank centrality in a graph of co-accused police officers may be more predictive of which officers receive misconduct than just linearly fitting a graph of allegations vs. disciplined allegations.

Checkpoint 5: Natural Language Processing

I used tokenization and transformations to break up the allegation notes and then performed sentiment analysis to get the sentiment scores used throughout this checkpoint. Since there were so few allegations with reports that resulted in discipline, I also tried to see if there was a correlation between sentiment score and allegation count since it may be reasonable to believe that complainants would be more upset with officers that tend to have more allegations.

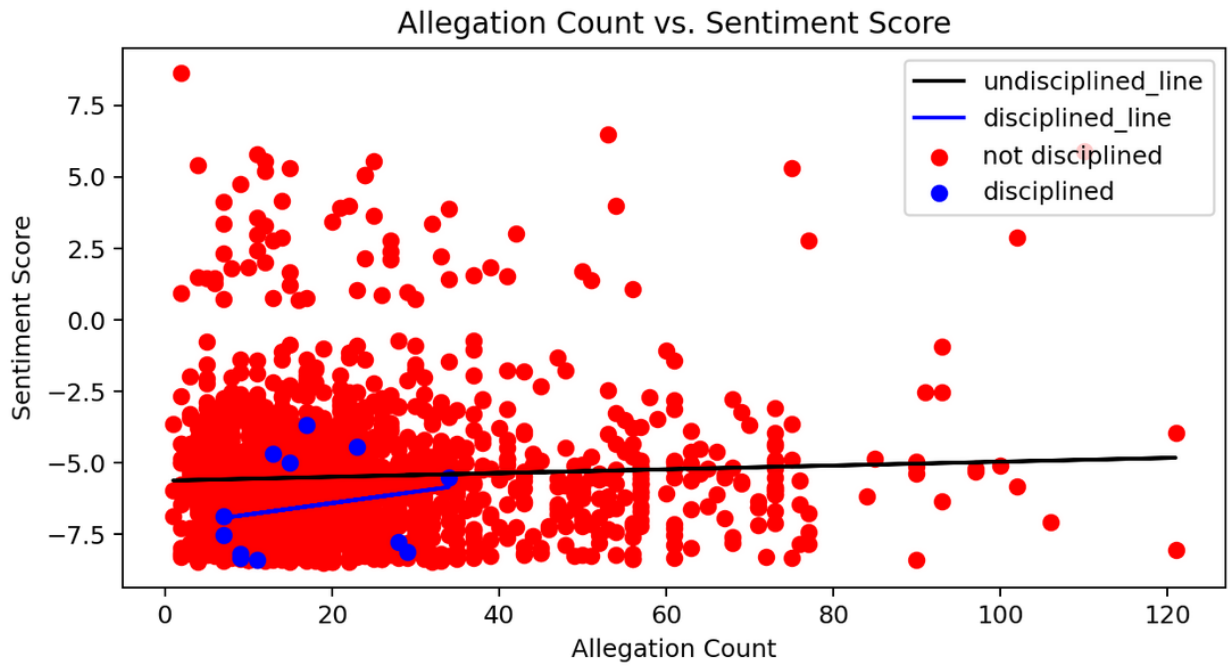

Figure 11: Allegations Count vs. Sentiment Score

```
R Squared Not Disciplined: 0.002467671233613637
Pearson Correlation Coefficient Not Disciplined 0.04967566037420778
Spearman Correlation Coefficient Not Disciplined 0.03011874668301905
R Squared Disciplined: 0.045865923107537274
Pearson Correlation Coefficient Disciplined 0.2141633094335659
Spearman Correlation CoefficientDisciplined 0.30175624350367664

KS Test: Ks_2sampResult(statistic=0.30701754385964913, pvalue=0.17147898551166485)
AVERAGE DISCIPLINED SENTIMENT: -6.54
AVERAGE NOT DISCIPLINED SENTIMENT: -5.47

 DISCIPLINED STD SENTIMENT: 1.6799528763232219
 NOT DISCIPLINED STD SENTIMENT: 2.3838242942844228
```

The above graph (Figure 11) shows the relationship between allegation count and sentiment analysis for different allegations. As can be seen, the relationship between these two variables is very noisy. Additionally, there were very few examples of allegations that resulted in disciplinary action in which the allegation text was recorded, making it difficult to infer anything predictive about how sentiment analysis can be used to predict which officers will be disciplined. Also of note is that these distributions have a high enough p value so that we cannot reject the null hypothesis that they come from different distributions, but the p value is fairly low. There is also a moderate positive correlation between sentiment score and allegation count for allegations leading to disciplinary action but it is very difficult to make general inferences from 12 data points.

Although not included, I did analyze the relationship between allegation count and sentiment score partitioned by both race and gender and found that there was no definitive correlation, which was surprising. Additionally, the lack of data for reports resulting in disciplinary action make it difficult to find any meaningful measure of correlation and limits the opportunities for logistic regression and other supervised classification algorithms with sentiment score as a feature vector along with race, gender, and allegation count since there is no way to prevent overfitting a model when one class has three times the magnitude of the other class.

Conclusions:

This project has been able to identify multiple variables that show relations illuminating why some complaints are sustained and lead to discipline and why others are not through use of relational analytics, exploratory data analysis, interactive visualizations, graph analytics, and natural language processing. It is apparent that complainant race plays a major role in which allegations are sustained with African American complainants having their complaints lead to disciplinary action significantly less than other races regardless of location in the city. Additionally, officer seniority by position and number of allegations seem to be negatively correlated with sustained allegation percentage, showing that officers that elicit complaints at a higher rate and officers holding higher positions seem to escape accountability for those complaints more easily. It is also apparent that PageRank centrality in the network of allegations for officers with high numbers of allegations against them is positively correlated with allegations leading to disciplinary action. Taken together, these correlations paint a picture of a police department that does not necessarily do a good job of holding officers accountable who are accused of large levels of misconduct. Since this is not a controlled experiment testing causation, we cannot rule out that these allegations are being evaluated objectively and the results show these correlations as an artifact of a just process. However, the fact that much of the data paints the picture of a racialized officer complaint handling process that allows for officers with large numbers of complaints to continue racking up complaints, coupled with anecdotal evidence of racialized policing in Chicago[2], we can reasonably infer that the system needs significant improvement in how these complaints are handled and how officers are disciplined and held accountable.

Future Work:

From this project, we can now use some of the tools seen such as network analysis using PageRank to infer which officers may have higher likelihoods of garnering complaints that lead to disciplinary action, which could be helpful for any predictive measures built to predict officer discipline. Future work may include obtaining more text complaint reports from reports which led to officer discipline, since the number is extremely low in the sample and makes implementing a machine learning classification model utilizing different features ranging from complaint sentiment analysis to complainant gender and race extremely difficult. Additional statistical significance testing for some of the correlations in checkpoint 1 would also be important in further showing how these variables are correlated.

[2] https://theintercept.com/2018/10/16/chicago-police-shooting-video-ricky-hayes/