Tracking Effectiveness of Police Reform

Data Pirates: Kavya Jaganathan, Clarissa Cheam and Ayushi Mishra

December 7, 2021

1. Introduction

The theme of this project is to track various police reforms and assess whether there are improvements after reformative interventions (like COPA, TRR, etc.). Since August 15, 1835, the creation of the Chicago Police Department, a lot of reformative policies have been deployed to constantly improve the service to Civilians. Despite the efforts, many civilians have fallen victim to instances of police brutality and negligence. An institution that should have been a symbol of security, fostered a lot of unrest and distrust.

Tracking the effectiveness of these reformation efforts will provide insight on questions like-

(1) Are substantial improvements being made across the board?
(2) Identifying neighborhoods that might require additional supervision/intervention?
(3) Are officers being held accountable appropriately?

Our goal is to leverage the information available on types of complaints made by both civilians and officers alike against police officers, neighborhood data, complaint outcomes, the extent of discipline, and information on the subsequent investigation including investigation/complaint summaries to gain insight on if and how accountability is being enforced. Armed with this information, statistically informed decisions can be made to ensure policies are updated accordingly, resources aren't squandered or wasted and future reform can be perfected as well.

2. Checkpoint1: Data Exploration using SQL analytics

To start off our project, we drafted up four questions to explore: 1) What are the top 3 complaints in each neighborhood before COPA, 2) What are the top 3 complaints in each neighborhood after COPA, 3) For all officers that had sustained allegations, what were the disciplinary actions taken?, and 4) For each type of allegation, how many were sustained before and after COPA? The bird's eye view of this checkpoint explores (1) What the effect of COPA is geographically on types of complaints (2) How the officers are held accountable for their actions (3) What the effect of COPA is on the types of complaints.

	name	category
2	Albany Park	Operation/Personnel Violations
3	Albany Park	Use Of Force
4	Albany Park	Illegal Search
5	Andersonville	Use Of Force
6	Andersonville	Operation/Personnel Violations
7	Andersonville	Verbal Abuse

Sample output from Question 1

	name	category
2	Albany Park	Operation/Personnel Violations
3	Albany Park	False Arrest
4	Andersonville	Operation/Personnel Violations
5	Andersonville	Conduct Unbecoming (Off-Duty)
6	Armour Square	False Arrest
7	Armour Square	Use Of Force

Sample output from Question 2

id	first_name	last_name	category	allegation_name	final_outcome
5	Carmen	Abbate	Operation/Personnel Violations	Inventory Procedures	No Action Taken
5	Carmen	Abbate	Operation/Personnel Violations	Slow / No Response	No Action Taken
5	Carmen	Abbate	Operation/Personnel Violations	Miscellaneous	Reprimand
5	Carmen	Abbate	Criminal Misconduct	Conspiracy To Commit A Crime	No Action Taken
6	Anthony	Abbate	Use Of Force	Excessive Force / On Duty - No Injury	No Action Taken
6	Anthony	Abbate	Lockup Procedures	Prisoners Property	No Action Taken
6	Anthony	Abbate	Operation/Personnel Violations	Weapon / Ammunition	Penalty Not Served
6	Anthony	Abbate	Use Of Force	Unnecessary Display Of Weapon / On Duty	Unknown
6	Anthony	Abbate	Operation/Personnel Violations	Inadequate / Failure To Provide Service	No Action Taken
6	Anthony	Abbate	Operation/Personnel Violations	Insubordination	Penalty Not Served
9	Michael	Abbey	Illegal Search	Improper Search Of Person	No Action Taken
9	Michael	Abbey	Drug / Alcohol Abuse	Positive Drug Screen - Originated From Complaint	Separation
9	Michael	Abbey	Operation/Personnel Violations	Inventory Procedures	No Action Taken

Sample output from Question 3

	category	pre_copa_allegations_per_year	post_copa_allegations_per_year
2	Operation/Personnel Violations	226.4929859719438878	10.2564102564102564
3	Domestic	8.6372745490981964	15.3846153846153846
4	Use Of Force	40.9619238476953908	57.6923076923076923
5	Excessive Force	0.02004008016032064128	8.9743589743589744
6	Illegal Search	3.6472945891783567	5.1282051282051282
7	Conduct Unbecoming (Off-Duty)	45.250501002004008	1.2820512820512821
8	False Arrest	0.64128256513026052104	7.6923076923076923
9	Lockup Procedures	33.6272545090180361	6.4102564102564103
10	Verbal Abuse	6.4328657314629259	8.9743589743589744

Sample output from Question 4

Conclusion:

The most glaring observation made about the dataset is about how skewed it is. We have 49.9 years' worth of data pre COPA whereas we have 0.78 years' worth of data post COPA. This adds bias and assumptions to the observations made above that can be rectified by normalizing the sustained vs nonsustained allegations over the number of police officers active in the force for every year. The effectiveness of reform can be checked for a few months before and after the timestamp '2017-09-15' to get a clearer picture. This can be further explored by using time series data visualization.

The most prevalent allegations in Chicago were found to be Operation/Personnel Violations(33%) and Use of Force(25%). In addition, by comparing the findings from Question 1 and 2, we found that allegations in the categories of 'Operation/Personnel Violations' (11% increase i.e. from initial 33% to 44%) and 'False Arrests' (8% increase i.e. from initial 2% to 10%) both increased after the founding of

COPA, while in 'Use of Force' (13% decrease i.e. from initial 25% to 12%) and 'Verbal Abuse' (7% decrease i.e. 7% to 0%) decreased. However, we are unsure of whether the decrease can be attributed to a lack of data or the actual success of reform. We assume that "SU" in final_findings denotes Sustained and the rest of the notations are variations of the non-sustained allegation, therefore we acknowledge that if this assumption proves to be false then there would be a huge no. of misclassification cases in our results. This could result in over-reporting of non-sustained cases painting the reform made by COPA in a poorer light in comparison to the reality due to this "bias" introduction caused by our assumption.

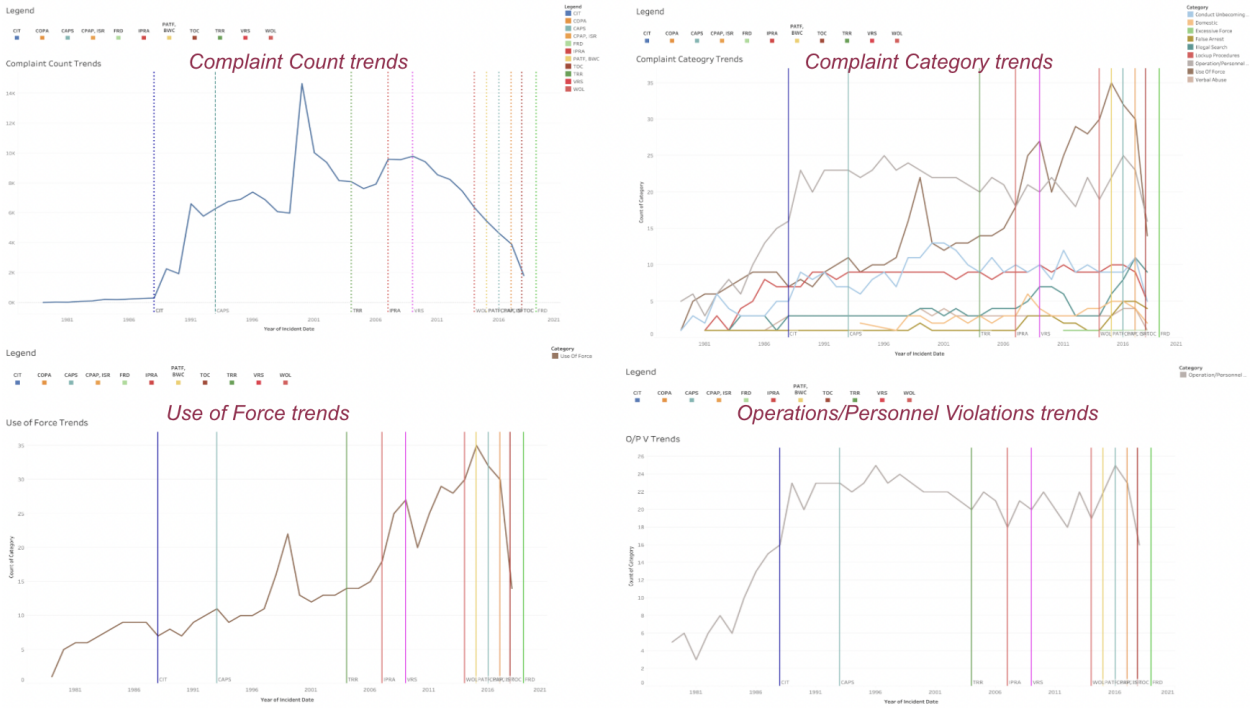
Out of all sustained allegations, 82% had no action taken by the Chicago Police Department, and the extent of punishment only extends to either a reprimand or a 1 or 2 day suspension. For sustained allegations that had no action taken as their final outcome, 14% were of the 'Use of Force' category with allegation 'Excessive Force/On Duty Injury), 11% were of 'Illegal Search' with allegation 'Search of Premise Without Warrant' and 9% were of 'Operation/Personnel Violations' with allegation 'Inadequate/Failure to Provide Service'.

We can also further reconcile our assumption on the final_finding attribute by getting subject matter experts' opinions.

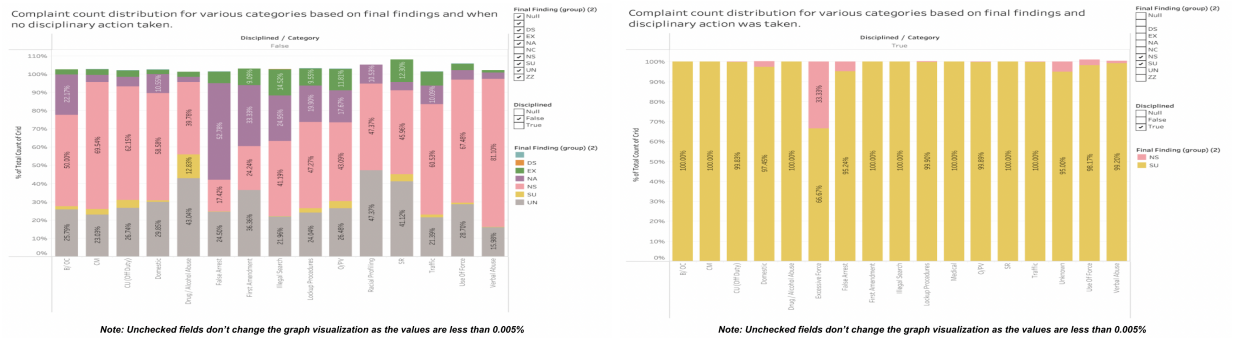
3. Checkpoint 2: Data Visualization

Why track Police Reform? Since August 15, 1835, the creation of the Chicago Police Department, a lot of reformative policies have been deployed to constantly improve the service to Civilians. Despite the efforts, many civilians have fallen victim to instances of police brutality and negligence. An institution that should have been a symbol of security, fostered a lot of unrest and distrust. This checkpoint aims to answer the following questions:

1. What are the trends in complaint count overall and per complaint category over time and what are the effects on these counts with the introduction of various reforms?
2. What is the percentage of Total Count of Complaints for each Category of final_findings ("SU", "NA", "NS" etc.) broken down by Disciplinary action?



Visualizations generated for question 1.



Visualizations generated for question 2

The Chicago Police Department was founded on August 15, 1835. Since then, instances of public outcry due to the unjust treatment of and the public failure in the handling of cases such as Laquan McDonald, Emmett Till, Diane Bond, and Rekia Boyd, etc. called for reformative practices. Some of these notable reformative efforts over time are:

1.	Crisis Intervention Team (CIT)	1988
2.	Chicago Alternative Policing Strategy (CAPS)	1993
3.	TRR Reports (TRR)	2004
4.	Independent Police Review Authority (IPRA)	2007
5.	Violence Reduction Strategy (VRS)	2009
6.	Watch Operations Lieutenant (WOL)	2014
7.	Body Worn Cameras (BWC)	2015
8.	Police Accountability Task Force (PATF)	2015
9.	Investigatory Stop Report Training (ISR)	2016
10.	Community Policing Advisory Panel (CPAP)	2016

11.	Civilian Office of Police Accountability (COPA)	2017
12.	Training Oversight Committee (TOC)	2017
13.	Force Review Division (FRD)	2019

These reforms vary from investigative oversight agencies to mental health wellness awareness training of police officers.

The answers to our questions are as follows:

1. **What misconducts were targeted by a certain reform?**
    a. We can see that most of the reforms target excessive force, use of force, violations, and very few that target operations/personnel violations. This indicates that department restructuring, disciplining/policies are required in every police department to deter operation/personnel violations.
2. **Did it incite a positive change?**
    a. Most of the reforms barring VRS, COPA have not affected the complaint counts indicating the need for restructuring, iterative improvements of the other reformative policies or actions.
3. **Are these reformation efforts being enforced?**
    a. CIT, CAPS, TRR, IPRA, WOL, despite being introduced, don't indicate that they have been enforced due to the unchanged complaint trend.

The third answer raises the question: why are these reforms failing despite being deployed? We can localize this question to neighborhoods, and this, in turn can allow for targeted rectification and supervision of reform enforcements.

4. Checkpoint 3: Interactive Visualization

In our previous checkpoint, we visualized complaint category trends with a simple static line graph with the incident date attribute as the x-axis and the count of category attributes on the y-axis. However, it is hard to really picture how complaint category trends change over time on a static visualization, and so we have decided to use a bar chart race to illustrate the leading complaint category trends up till 2018. Using a bar chart race we can intuitively visualize which complaint categories had the top counts in each year and its progression through the years.

Analysis of Bar Chart Race:
Through the years 1919 - 1927 Verbal Abuse and Operation/Personnel Violations were the prevalent misconducts in accordance with the data.
During these years the CPD was in unrest. [1]

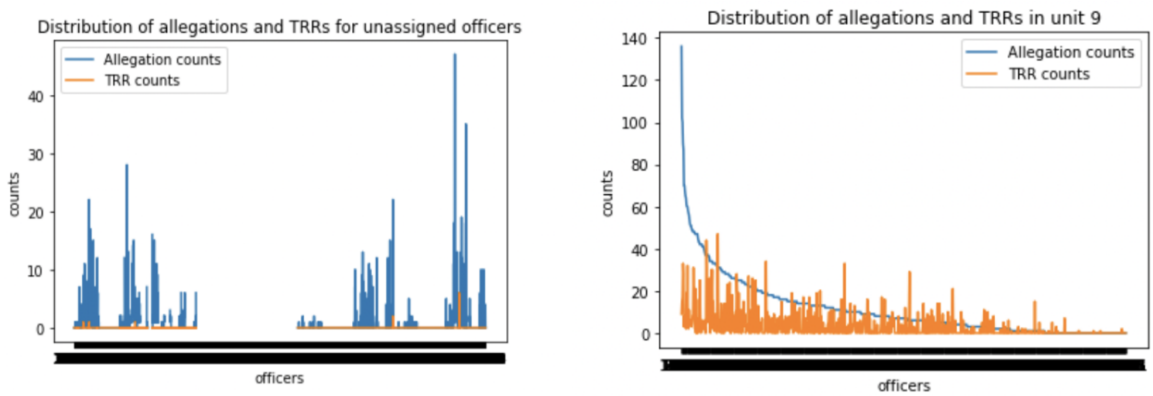
- Change in Rank Title – Ranks were reinstated several times but were abolished in 1909, 1917, and 1921. For an extremely hierarchical institution, this instability can cause inadvertent harm.
- The Chicago Police Reserves were organized - the members of this reserve ranged from officers to men. The Reserve was used instead of regular officers or to supplement them. As representatives of the CPD, interacting a lot with the general public, these supplemental reserves of officers do not have enough experience or training. These could have been contributing factors to the Verbal Abuse allegations.

Like so, the bar chart race allows us to split the data into eras based on when the bar graph rearranges its top three misconducts. These characteristics make the graph more intuitive compared to the previous checkpoint. From the race bar and the previous checkpoint, we can confidently conclude that the top 3 misconducts across Chicago are Use of Force, Operation/Personnel Violations, False Arrest but on further examination of data, we observe that these categories are blanket statements to more intricate complaints.

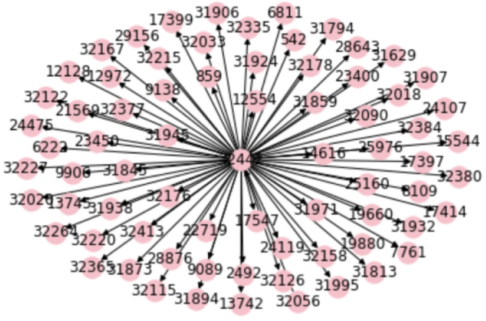
The categories of misconduct in Operation/Personnel Violations that affect civilians are; Failure to Identify, Unfit for duty, Neglect of duty, property damage, and misuse of departmental equipment. The total count of allegations amount up to 1,748,516! This is not a small number but pales in comparison to Use of Force allegations. The circle packing chart shows only the counts for the top 3 Use of Force allegations grouped as; Physical Violence, Weapon use, and Threats which still count up to 25,228,588 counts of allegations. These blanket categories and corresponding blanket reforms don't work due to the different natures of misconduct. Threats are due to misuse of power whereas various factors such as the mental state of police officers, fear of mortality, etc. can contribute to physical violence and weapon use. Further research into these subcategories and better grouping and reform development would move the scale in the right direction

5. Checkpoint 4: Graph Analytics

In this checkpoint, we test the hypothesis, "Are certain officers the cause of high misconduct counts in a certain district, or is it a unit-wide issue? If so, how does this affect reform choices?" We test this hypothesis to investigate whether reform needs to be implemented unit-wide or would retraining/removing certain officers improve the situation.
Question 1: What are the distributions of allegation counts and trr counts in units? Are the counts uniformly distributed, indicating a unit-wide problem, or is the distribution skewed, showing the presence of certain problematic officers?
Question 2: If certain officers are the problem in their respective units we want to magnify this network and investigate which officers are the ring leaders, and at the center of co-accusals, influencing other officers?



Visualization for Question 1



Officer id '32412' (Adam Wazny) coaccusal network graph with pagerank score of 165.07

Officer id '32442' (John Zinchuk) coaccusal network graph with pagerank score of 35.032

Visualization for Question 2

We first choose unit id 9 that operates in the 8th district - Chicago Lawn. As a unit, they have contributed to 10079 allegations in the total of Chicago allegations. This unit has a total of 895 officers. This unit makes up 2% of the officers in Chicago. Their allegation counts make up 4% of all the allegations in the Chicago Police Database. It is evident from the plot distribution for unit 9 that the counts are extremely high for some of the officers whereas it is almost 0 for others. We can infer that the elimination of these officers from this dataset will reduce the allegation contribution of this unit drastically. On the other hand from the plot distribution of unassigned officers, we can infer that the plot is pretty erratic with one extreme peak with an allegation count>40. We cannot really draw any significant conclusions from this data separately but 40 is still a large number and should be addressed with relevant data to that officer. To do so, it is imperative that the dataset is improved to provide clarity on why these officers don't have a unit id. On inspecting the rest of the units we further observed that all the units have a distribution similar to unit 9. From this, we can infer that none of the units are entirely engaging in misconduct. Instead, it is certain officers of the unit that are contributing to the allegations.

Possible reasons could be, the officers with high counts are the "rotten apples" in that unit or the officer could be a beat officer who has more engagement with the public than a supervisory agent who will consequently have a low allegation count due to the nature of the job.

These distributions are also muddled due to the assumptions we made with respect to allegation counts. To demystify these assumptions we want to know how important these officers with high allegations are, in the network of officers.

Despite the possibility that post-reform interventions like COPA there could be an increase in the count of misconduct in various categories due to more reporting by citizens and inadvertent increase in false allegations made against the officers; having observed the findings of pagerank algorithm we can conclude that if we were to eliminate the identified ring leaders or focus on reforming them particularly it would positively impact the police units and the neighborhood misconducts in these cases. We say this on the basis that just these 20 officers are involved in a total of 2581 coaccusal cases which is a significant number.

Conclusion:

We know from research and inferences from previous checkpoints that all the reforms have a broad target with respect to implementation and focus. We also know that disciplinary actions against officers are weak in comparison to their sustained allegations. This checkpoint drives in the fact that the contributors to allegation counts are certain officers in a unit and not unit-wide. We further also identify certain officers who are ring leaders in the CPD network. Reform targeted towards these officers can dismantle the vicious network of engaging in bad behavior. The conclusion from the previous question should translate into reformation activities. The CPD advised by data should focus on these certain officers and reprogram the CPD network to affect a positive change in the allegation counts.

6. Checkpoint 5: Natural Language Processing

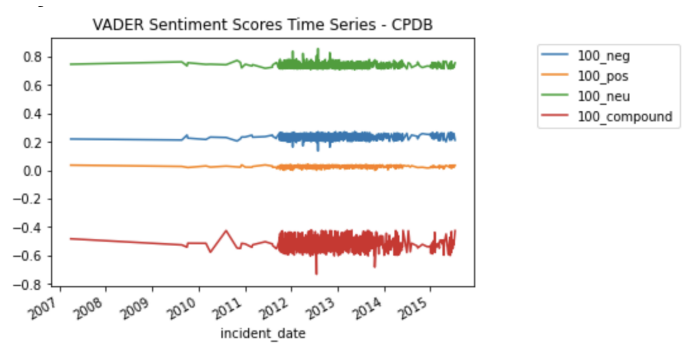
In this checkpoint, we explore the text content available in the tables data_attachmentnarrative and data_allegation. We study both these tables since the data_attachmentnarrative table's text content gives an account of the allegation, along with the victim perspective in terms of words uttered by the accused to them, tone/demeanor of accused that they observed, etc. whereas the data_allegation table's text content gives a report written by an officer without much of the victim's direct words.

<u>The question we explore:</u>

What was the sentiment in the CPDB complaint data pre-major reforms (before 2016) and post-major reforms?

Note: Major reforms here indicate reforms like COPA and the introduction of TRR

Time Series complaint analysis of data_attachmentnarrative using incident date:



In the data_attachmentnarrative table, the text content gives an account of the allegation along with the victim's perspective in terms of words uttered by the accused to them, tone/demeanor of the accused that they observed, etc. We perform sentiment analysis on this data to understand the prevalent emotion/sentiment. With a compound score in the range of (-0.8, -0.4) we can understand that the sentiment was majorly negative. We cannot perform further analysis on this data to understand the effect of reform since the data in this table dates up to 2015. Despite the lack of data, analyzing this data is essential as understanding the history of data is important in forming any new reforms.

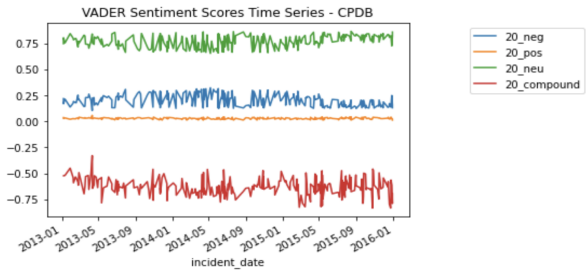
PCA ANALYSIS

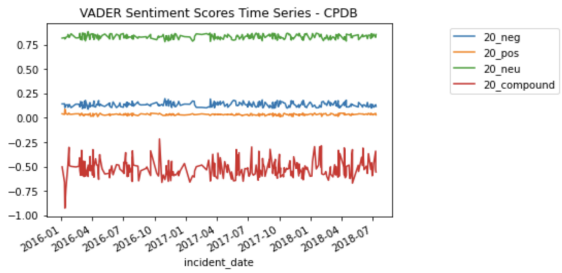
The PCA tool was utilized since the high-level view of frequent words does not give much insight.

We observe a high prevalence of aggressive, unprofessional behavior of the accused, misconduct in terms of procedure highlighted by the frequent appearance of words like 'false', 'ignore', 'harass', 'inappropriate', 'supervisor' and a lot of rude verbal abuse and unprofessional conduct including profanities. These

findings are supported by the bigrams and trigrams as well. Ignoring common words that appear on all reports like alleged, accused, complaint, report, etc. we can see the bigrams like "offic,rude","fals,arrest","without,justif" are all indicative of unprofessional and improper conduct with civilians and callousness towards procedure. The trigrams "search,vehicl, without", "vehic, without,justif", "fail,inventori,return","without,justif,complain" all support the findings from bigrams indicating lack of attentiveness bordering on complete disregard of procedure.

Time Series complaint analysis of data_allegation table using incident date:





Pre major reform Analysis

The range of compound score lies between (-0.25, -0.80). Standalone, this data displays a pretty negative atmosphere. There are also high neutral scores generic of reports. This data combined with future data provides a better perspective.

Post major reform Analysis

Contrary to expectations of reform the compound scores are observed to be further negative, going as low as 0.92. The range of the compound score has shifted to (-0.25, -1.0) indicative of excessive negativity.

This highly negative score could be indicative of harsher allegations and tonality of reports. We further look into the words to pick out contributors to this score.

PCA Analysis

All PCA topics, in general, seem to indicate excessive violence with frequent words being strike, discharge, head, foot, offend, alterc, control, punch, knee, gun, knife, etc.

Conclusion:

From this Checkpoint, we can infer that major reforms fostered further negativity due to the prevalence of more reports displaying escalated aggression. We cannot claim the reform completely failed or succeeded due to minimal data post-reform but, the spike in negativity due to reports with escalated aggression could indicate that the reforms fostered more reporting of misconduct. This can be a partial success of the reform. Although, there was barely any downward trend in the negative sentiment. The only success of the reform could be encouraging reporting but not correcting the officers/preventing misconduct.

7. Discussion and Future work

This project tracks police reform from 1835 to 2018 and documents the successes and failures of the same. In conclusion, our findings are as follows-

1. **Complaint Category Analysis:**
   a.Use of Force, Operations/Personnel Violations, and False Arrests are the **top three allegations.**
   b.Most of the reforms target excessive force, use of force, violations, and very few that target operations/personnel violations. This indicates that **department restructuring, disciplining/policies are required in every police department** to deter operation/personnel violations.
2. **Reform performance:**
   a.Reforms have a **broad target** with respect to implementation and focus.
   b.Most of the reforms barring VRS, COPA have not affected the complaint counts indicating the need for restructuring, iterative improvements of the other reformative policies or actions.
   c.CIT, CAPS, TRR, IPRA, WOL, despite being introduced, don't indicate that they have been enforced due to the unchanged complaint trend.
3. **Disciplinary actions against officers are weak** in comparison to their sustained allegations.
4. The sub-categorization of violations grossly undermines violence against civilians and the counts are gravely misrepresented.
5. **Allegation Contributors**
   a.Contributors to allegation counts are certain officers in a unit and not unit-wide.
   b.**Certain officers are ring leaders** in the CPD network. Reform targeted towards these officers can dismantle the vicious network of engaging in bad behavior.
6. **Language Analysis:**
   a.Post Major reforms spike in negativity due to the prevalence of more reports displaying escalated aggression is observed.
   b.The **only success of the reform could be encouraging reporting** but not correcting the officers/preventing misconduct.

Future works:

COPA has jurisdiction only over some of the CRs. The remaining CRs go through the CPD's Bureau of Internal Affairs. We can **explore the difference in outcomes for those CRs that go through COPA and those that don't.** We have looked at the effects of major documented reforms on complaint counts. We can expand this question by **exploring the data after high-profile cases** that bring a lot of attention and media pressure to the CPD.

8. References

Hyperlinked on the text where required.