Data Science Seminar - MSAI 339

Checkpoint 1

November 1st, 2021

Professor
Jennie Rogers

Students
Anery Patel
Chukwueloka Obi
Ana Cheyre

Research questions

Introduction

With the following relational analysis, we intend to find variables that can be related to the declaration of unsustained allegations and go one step further. Understand if the number of years in the force of an officer, the number of changes of units that an officer has made, the number of awards, the neighborhood where the person who creates the allegation lives, among others, may have some kind of correlation to this type of particular allegations.

- **What is the correlation factor between the number of years in the force of an officer and the number of unsustained allegations that a police officer has?**

We see that the correlation factor is 56% meaning that the number of years in the force of an officer and the number of unsustained allegations that a police officer has is somewhat moderately correlated since the correlation factor is greater than 50%. This table was assembled by selecting the id, appointed_date, resignation_date, and unsustained_count from the data_officer table. This question is pertinent to us because we felt that by seeing the correlation factor of these two variables we are better able to understand if this police brutality matter is a deep seated institutional matter rather than just a matter of a few bad police. The reason for this correlation factor we obtained could be there are quite a few police officers that have had a long file of allegations filed against them throughout their career on the police force. Many of these officers that contribute to this correlation are veteran officers with more than 20,30 and 40 years of experience on the police force. Many of these men are more accustomed to an old way of policing but there are also some police officers with less experience who have a number of unsustained police allegations. Could the high number of unsustained allegations be due to increases in crime in recent times in the South Side Chicago area? Do police officers have to use more violent and brutal tactics in order to keep order in these neighborhoods during recent times? Or should we judge police officers as individuals who can either have bad characteristics or good characteristics. There are many officers who have as many years in the force as the officers with multiple unsustained allegations and have very low amounts or almost no allegations. So it really depends on who the officer is. Are we not able to compile good data on all officers over the last 40 years due to technology and recording data not being advanced enough or unsustained allegations just not being recorded at all over the past 40 years.

- **What is the average number of times a frequently accused police officer changes from one department unit to another?**

As we read in The Intercept, many of the most egregious examples of police misconduct come from tight-knit groups of officers. Misconduct is said to spread from one officer to another like an “infectious disease”. And the same behavior can lead police officers to break the rules often. This is why we want to start by seeing if there is a relationship between the number of times an officer changes from one unit to another and the number of unsustained allegations he or she has. This could indicate that an officer who has many allegations changes his location, but instead of changing his behavior, he continues to possibly spread his misconduct, which would be proved in the next analyzes.

First, to define what a frequently accused police officer is, we are going to divide the police officers in groups by the number of unsustained allegations each one have by quartiles: Q1: 0-25 percentile, Q2: 25-50 percentile, Q3: 50-75 percentile, Q4: 75-100 percentile, being Q4 the group with the highest frequency of allegations.

To do this query, we created a new table joining **data_officerhistory** and **data_officer.** From **data_officerhistory** we created a new field **unit_count,** which counts all the units a police officer has been in his or her career, also we added the **officer_id** to track each member. From **data_officer** we created a new field **sum_allegations**, which sums all the unsustained allegations a police officer has.

Quartile	Number Unsustained Allegations	Avg Unit Count
Q1	0	1.82
Q2	1	2.35
Q3	6	3.47
Q4	16	5.78

Table 1: Average changes of unit according to the quartile of the number of unsustained allegations per officer

For the highest quartile of the number of unsustained allegations that an officer has (Q4), on average an officer has been in 5.78 units. Then for the third quartile (Q3), we see that on average an officer has been in 3.37 units, almost half of the highest quartile

does. While further down the quartile, both the number of allegations and the number of units an officer has been in decreases in an accelerated way.

This shows that the officers who have more unsustained allegations also change frequently from one district to another, which can potentially lead to widespread misconduct.

As a next analysis, it would be interesting to normalize by time in the force, since perhaps an officer who has been in the force for more years, has committed more unsustained allegations and has also been in more units in his career.

- **Which are the top 5 neighborhoods with the most number of civilian complaints that have been unsustained? What percentage of unsustained complaints have each one from the total?**

With this query, we aim to understand if demographics of civilians residing in designated geographic areas has any relation with the complaint outcome. We can focus on the areas to identify recurring patterns of injustice demonstrated by the Chicago police department. Since we are looking only at those complaints filed by the civilians and where not followed through, we identify civilian unsustained complaints in each geographic location.

We use with **data_allegation_areas** to group by the area_id and count the number of unsustained civilian complaints using the relations **data_officerallegation** and **data_allegation**. Furthermore, we normalize the count of complaints in each area by the per capita count so as to have a fair comparison. This is done by calculating the total population by summing the count for each race in the **data_racepopulation** table.

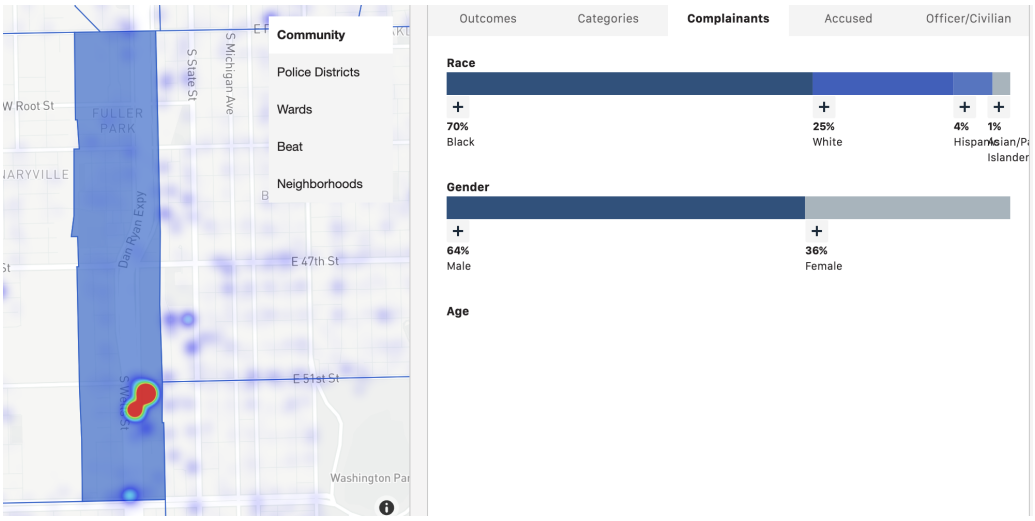
Area_Id	Area_Location	% of civilian unsustained complaints
428	Fuller Park	9.3
472	Pullman	3.8
490	West Engelwood	2.79
453	East Garfield Park	2.46
463	Loop	2.43

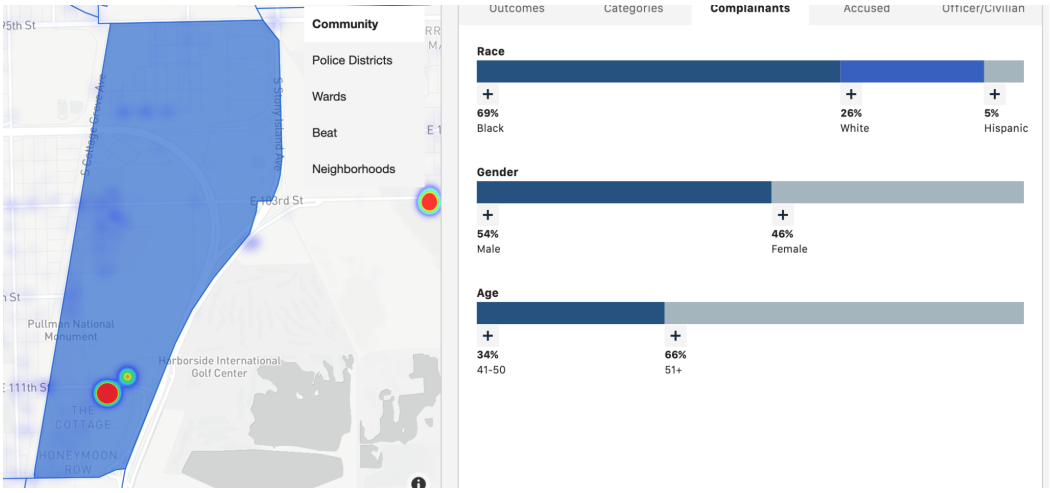
Table 2: Distribution of unsustained allegations per area

We notice that there are some neighborhoods with a high percentage of unsustained complaints. Ideally speaking, for 99 areas, the rate should be 1.01% if the complaints are evenly distributed. This gives us a direction to probe these areas in multiple ways. We can look into the functioning of the police units active there, the group of officers serving the areas and also the civilian nature: race, income, age, etc. We can identify factors that prevent the civilian complaints from being sustained.

When we look at the demographics in the areas, we notice three things. The population has the majority of the people belonging to the **'Black' race**. Secondly, the **median income per household** is low. Lastly, we look at the officer groups in force in each neighborhood.

1. **Race population in each area**





When we probe the race distribution in the top 5 neighborhoods, we notice that Fuller Park has 70% complainants that are Black. Similar case with the neighborhood Pullman. East Garfield park had 90% Black complainants while West Engelwood has 94%. Racial disparity in policing is clear at this point.

2. Median Household Income in each area

FULLER PARK

Population	Race	
2,457	Black	92.3%
Median Household Income	Hispanic	6.2%
$19,589	White	1.5%
	Asian	0.0%
	Other	0.0%

PULLMAN

Population	Race	
6,501	Black	82.8%
Median Household Income	Hispanic	8.5%
$36,622	White	7.1%
	Other	1.1%
	Asian	0.5%

WEST ENGLEWOOD

Population	Race	
32,156	Black	93.2%
Median Household Income	Hispanic	4.9%
$27,203	Other	1.2%
	White	0.7%
	Asian	0.1%

EAST GARFIELD PARK

Population	Race	
20,665	Black	91.3%
Median Household Income	White	4.2%
$21,307	Hispanic	3.5%
	Other	0.7%
	Asian	0.3%

We notice the median income per household is low in the neighborhoods reporting most number of unsustained complaints. However, it would be interesting to perform an analysis to learn how much ‘median household income’ compares with ‘race’ factor in the demographics. We can weight the influence of two factors to learn how each one contributes to the problem at hand.

3. Police Officer race population serving in each area

An interesting case opens up when we look at the group of officers involved in the unsustained complaints. First, we check the race of the officers and their disciplined rate in the top three areas.

From the CPDB website, we see that in Fuller Park, 51% of the officers with unsustained complaints against them are White while in Pullman 59% of the officers are Black. However, 0% of the officers are disciplined in the case of Use of Force complaints in Fuller Park but 1% of the officers in Pullman are disciplined for the same

category. This draws an important insight that even among the officers, there is a racial disparity when it comes to disciplining them for misconducts.

4. **Community Policing**

When we look at the disciplined rate in each neighborhood, we observe a lot of difference.

West engelwood has maximum disciplined officers for each category of accusation. However, when we look at the statistics about the officers involved, we notice that officers belonging to the age group 20-30 are 30%, the maximum we have seen among the five neighborhoods. This throws light on possibly a notable flaw that older officers are less likely to be disciplined thus causing them to contribute the most to the number of misconducts.

We can safely conclude that the officer dynamics within the police units in different neighborhoods play an important role in determining the rate of unsustained complaints filed by the civilians.

- **What is the correlation between the number of unsustained allegations against a police officer and the number of awards that officer has?**

As we aim to identify various factors and relationships in the policing system to understand why police officers are not held accountable for their misconduct, awards bagged by the police officer during his/her tenure play an important role. Intuitively speaking, it is likely that an officer with multiple achievements asserts his/her standing due to their capabilities and thereby puts themselves at a position higher than the fellow officers. At such a position, the officers are at an advantage when it comes to getting away with misconduct. They are less likely to be questioned, blamed and disciplined albeit not in an ideal policing environment.

To determine the role of achievements, we find a correlation factor between the number of awards an officer receives versus the number of unsustained allegations he/she has during their tenure. The query includes calculating the total number of unsustained complaints grouped by the officer ids from the tables **data_officerallegation**. We select the complaints 'Unsustained', 'exonerated' and 'not sustained' and count the allegation id count. Thereafter, from the **data_award** relation table we count the number of tracking_no for each award received by the officer. Through this, we get two columns in a table: allegation_count for each officer versus the awards he/she has received in the second column. We use the CORR() function between these two columns in Postgresql to result in a correlation factor.

On executing the query, we get 0.98 as the correlation factor. Hence, we notice a strong correlation between the number of unsustained allegations against an officer and the number of awards he/she has. This makes sense because an officer is less likely to be blamed or disciplined by their fellow officers for misconduct if he/she has received many accolades during their career. It is an interesting yet quite simple takeaway. The system needs to have provisions in place that does not give a benefit of doubt to officers with awards because they are the ones to misuse it in the first place.

- **What is the average percentage of civilian and officer unsustained allegations an officer has?**

We want to understand how relevant unsustained allegations, both civilian and officer's, are in an officer, to know if one of them has more weight than the other. This can be a key indicator to understand if unsustained allegations are really relevant in this whole investigation.

To do this query, we used the tables **data_allegation** and **data_officerallegation**. From **data_allegation**, the field **is_officer_complaint** tells us if the allegation was originated by an officer (TRUE) or by a civilian (FALSE). We used the field **final_finding** from **data_officerallegation,** to determine if the allegation was unsustained or not.

We can see that most of the allegations that an officer has correspond to civilian unsustained allegations, with an average of 87.9%, so it is important to understand why this is. What is the profile of these civilians who originated the allegation? What kind of cause of allegation has to be most unsustained? Would there be any percentage of these complaints that have been erroneously classified as unsustained? All of these questions would be very interesting to answer for a further analysis and would help to have a better understanding of the kind of allegations that an officer receives.

	Civilian	Officer
Unsustained	87.9%	3.3%
Sustained	6.3%	2.5%

Table 3: Average percentage distribution by type of allegation an officer has

Conclusion

This was our first step of the analysis. After evaluating the result, we can find out which variables do have a relationship and which don't, with the unsustained allegations. In addition, we understood the great weight that this type of allegation has in the record of an officer, so it makes it more relevant to investigate.

The variables that we can declare as related are:

- **Number of awards of an officer:** The more awards he or she has, the more unsustained allegation he or she has.
- **Number of unit changes of an officer**: The more number of units he or she changed in his or her career, the more unsustained allegation he or she has.
- **Neighborhoods:** There is a difference between one neighborhood and another in terms of the amount of unsustained allegations per capita. The variables related to this difference are mainly related to race and median income. Neighborhoods with more unsustained allegations tends to have a higher percentage of black race and lower income.

The variable that we declare is not significantly related is:

- **Number of years in the force of an officer:** We did not find that this variable was strongly related to the number of unsustained allegations, so we can say that the number of years in the force is independent of the number of unsustained allegations that an officer may have.

It was a good first approach with correlations that may exist between variables since we began to delimit where there really may be some bias that causes allegations to be declared unsustained. Of course, there is still much to be found, so the next analyzes will help to further narrow this investigation.