Faraaz Beyabani, Andrew Vagliano, Johnny Trinh
October 6, 2020

Checkpoint 1: SQL Analytics

The theme of our project is to analyze the activities of officers across neighborhoods, units, and beats to examine which, if any, areas are overpoliced. Overpolicing can typically be categorized by a large influx of complaints from a given geographical location. However, we can also use other metrics such as area demographics, TRRs filed, beat assignments and composition, and settlement payouts. Therefore, we aimed to answer the following questions by using the Citizens Police Data Project (CPDP) dataset.

Question 1: What is the number of assigned officers, complaints, and settlements across all of the Chicago area?

Chicago policing is asymmetric across neighborhoods and city areas. With some neighborhoods being overpoliced, notable geographic spikes in complaints, settlements, and TRRs can be found. While the citywide total of complaints fell between 2000 and 2015, the clusters of complaints in the south and west areas of the city were constant. Seven of the top ten police beats with the most civilian complaints against them in 2015 were stationed in the South Side, and six of the top ten in 2000 were similarly clustered. These spikes in complaints show increased aggression from officers in specific areas of the city.

  The number of officers with complaints against them jumped between 2005 and 2015. Accused officers more than doubled from about 3,700 to over 8,000, showing a massive increase in abuse reports. It is unclear whether this increase is due to more public engagement with the abuse reporting system or more widespread abuse from officers, but the rapid increase shows that more work needs to be done to understand the apparent rise in abuse.

  Allegation payouts follow the same trend. In 2011, a little over $32M was paid out to victims at an average rate of ~$261,000 per case. In 2019, that total grew to over $45M, at an average of $314,000. The across-the-board rise in abuse, cases, and payouts is notable and will require a deeper dive into which more precise geographic areas are the primary drivers of this change.

Question 2: What is the racial and gender breakdown of the policing force (aggregating across all of Chicago)? Which units have the highest male-to-female ratio of police officers?

The CPD skews significantly white and male. Of the 22,000 officers active in 2000, 70% were white and 86% were male. In 2015, those ratios normalized slightly to 54% and 79%, respectively, but are still skewed. We have not found exhaustive data for individual beats to analyze the relationship between diversity and abuse allegations, but this path will be explored

in the future as the asymmetric demographics of the police and the citizens they are tasked to protect may produce interesting findings.

	race	gender	count
0	Asian/Pacific	F	56
1	Asian/Pacific	M	440
2	Black	F	1713
3	Black	M	3489
4	Hispanic	F	891
5	Hispanic	M	3226
6	Native American/Alaskan Native	F	12
7	Native American/Alaskan Native	M	43
8	Unknown		23
9	Unknown	F	50
10	Unknown	M	111
11	White	F	2247
12	White	M	10347

2000 Results

Question 3: What does the distribution of officer race look like for every district?

Next, we examine the role of officer race for each district. More specifically, we were curious if districts with a larger than average population of a given race (say, White) were more likely to accumulate complaints or certain types of complaints compared to districts that were different (predominantly another race or no significant majority).

To find this relation, we decided to start by extracting the number and races of officers in active beats, determined by the start time of the shift. We then used the geometric polygons in the area table to map each beat to the police district it was contained in. By combining these two tables, we were able to count the number of officers of each race, across every police district. We then normalized the data for each district by the number of officers to better compare across districts. A small snippet of the data collected can be seen in the image below.

Here, in the first column, we can see the name of the district. We can use this to join back on the area table if we have need of the geometric polygon or other relevant data. The second and third columns display the race, and what ratio/percent that race contributes to the overall officer population of that district. The ratio has been normalized to better allow for comparison between districts with wildly varying officer counts. Furthermore, sub-queries also exist to fetch the number of officers in each district, in the event that information is needed.

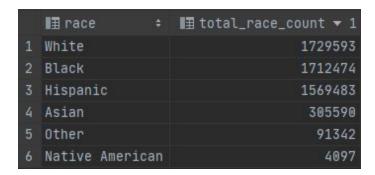|	district_name	race	ratio
1	10th	Amer Ind/Alaskan Native	0
2	10th	Asian/Pacific Islander	0.03
3	10th	Black	0.09
4	10th	Hispanic	0.5
5	10th	Unknown	0.01
6	10th	White	0.37
7	11th	Amer Ind/Alaskan Native	0
8	11th	Asian/Pacific Islander	0.04
9	11th	Black	0.12
10	11th	Hispanic	0.39
11	11th	Unknown	0.01
12	11th	White	0.43
13	12th	Amer Ind/Alaskan Native	0
14	12th	Asian/Pacific Islander	0.04
15	12th	Black	0.12

Even this small snippet of data has some interesting observations to be made. While there are a few districts that have a majority non-White population, the majority of districts still have a significant (> 30%) White population. While this may just be due to a majority of the officers in the CPD being White, it may also be that officers of certain races are more often assigned to certain districts, which may tie in to low-income neighborhoods being assigned more officers, for instance. This will be easier to visualize in Checkpoint 2, with the use of heatmaps.

In the future, we are planning to use this data to cross reference between neighborhoods and complaint data to visualize how many complaints might arise in neighborhoods of similar socioeconomic status that are in different districts. This would allow us to examine the relationship between district racial make-up and any resulting complaints levied against officers.
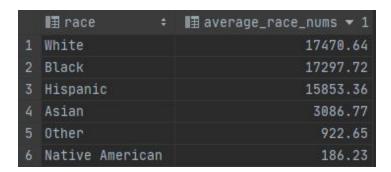
Question 4: What is the city's racial breakdown for its population? Which neighborhoods are predominantly White, Black, Asian, Hispanic (compared to the city average)?

As an extension of the previous question, knowing about the racial distribution of individual neighborhoods in comparison to the overall city demographics is especially important for analyzing overpolicing. For example, it could be that White officers are more often assigned to beats in majority-Black neighborhoods, and these officers might have more complaints filed against them than officers of other races and/or genders who patrol other neighborhoods. This might be due to officers overpolicing or it could just be citizens being biased against police presence. Regardless, we felt this data was very important to reach our final goal.

Therefore, we first performed a query on the data_racepopulation table to examine the total racial makeup of the city, the results of which are shown in the image below.

	race	total_race_count ▼ 1
1	White	1729593
2	Black	1712474
3	Hispanic	1569483
4	Asian	305590
5	Other	91342
6	Native American	4097

Here, the racial distribution is typical. One thing that did jump out to us was the 'Other' statistic. It may be that many people are selecting this option for racial data collection, such as the Census, because they are either uncomfortable picking other options, or feel that other options do not adequately describe their race. Nevertheless, these statistics are important because they let us calculate the average counts of each race across all neighborhoods:

	race	average_race_nums ▼ 1
1	White	17470.64
2	Black	17297.72
3	Hispanic	15853.36
4	Asian	3086.77
5	Other	922.65
6	Native American	186.23

The rankings of average racial distribution are identical to the total racial distribution of Chicago: people who identify as White are the most common, while Native Americans are the least common. With this, we can query for the areas which have a much higher racial population than the average. Right now, we are classifying this parameter as 20% more than the average racial population.

	race	area_id	name	count	average_count
1	Other	454	Near West Side	1308	922.65
2	White	454	Near West Side	25876	17470.64
3	Asian	454	Near West Side	9626	3086.77
4	White	442	Dunning	28340	17470.64
5	Black	455	North Lawndale	31397	17297.72
6	Asian	459	Near South Side	4450	3086.77
7	Other	482	Lake View	2709	922.65
8	White	482	Lake View	78186	17470.64
9	Asian	482	Lake View	6764	3086.77
10	Hispanic	484	New City	24990	15853.36
11	White	461	Norwood Park	29164	17470.64
12	Black	470	Roseland	40846	17297.72
13	Black	491	Englewood	24810	17297.72
14	Other	462	Near North Side	2214	922.65
15	White	462	Near North Side	61334	17470.64
16	Asian	462	Near North Side	9723	3086.77
17	Black	492	Greater Grand Crossing	31077	17297.72
18	Asian	463	Loop	5745	3086.77
19	Other	450	West Town	2572	922.65
20	White	450	West Town	49892	17470.64
21	Hispanic	450	West Town	23239	15853.36
22	Black	451	Austin	81066	17297.72
23	Hispanic	1527	17th	62232	15853.36
24	Other	1527	17th	3645	922.65
25	Asian	1527	17th	17373	3086.77

From this data, we can note the area IDs of locations with higher than average racial populations. Now, further specification of this data is necessary, as it may be that we get all areas with a higher-than-average general population, but we can note areas such as Englewood, which are majority Black or the 17th being mostly non-White. In the future, this data can be used to directly compare two areas with similar demographic makeups and compare the number and nature of all complaints originating from them.