

## **CFJJ Team 2: Research on Lowell, MA**



Dec 10th 2021

**Project Goal:**  
Research the hidden relationship between traffic stops and racial biases of police officers.

Akira Kato, Baiqing Lyu, John Yin, Jacqueline Lin

Sponsored by CITIZENS FOR JUVENILE JUSTICE

---

## DATA COLLECTION/PREPROCESSING OF NEW DATA

The datasets we used are provided by our client, Eileen Morrison, from the Committee for Public Counsel Services. Two data files contain specific offenses that police officers gave out via citation numbers (file DOT LOWELL) and the specific location for each citation (file Location Analysis Report). In DOT LOWELL, there are a totally of 1457 records and multiple fields: in field citation\_type, 1.24% for ARREST, 79.00% for CIVIL, 3.84% for CRIM, and 15.92% for WARN; in field gender, 38.1% for FEMALE and 59.9% for MALE while the rest stand for missing or not present; in field race, there are nine different races, with the four top commonest are white (48.0%), Hispano (25.7%), black (11.2%) and Asian(7.1%). Plus, one citation number may exist multiple times because a citation can offend multiple laws at the same time.

The “DOTLowell\_April2021” contains 1457 lines of offenses recorded by Lowell Police, having columns issuing agency, Officer ID, Event Date, specific time, citation number, citation type, Offence number, Offence description, location name, offenders’ Race, Gender, Year of Birth, Plate Type, vehicle type. The overall dataset is clean. In addition to the data received, we classified each offense by severity level according to the Massachusetts Master Crime List. The Master Crime List ranks all possible offenses by severity with a rank from 1 to 7, but the offenses included in the data is between 1-3 where there is an approximate 65-35 split between rank 1 and 2 and rank 3 having only 2 cases so it is essentially negligible.

For document “Location Analysis Report - April 2021”, there are 1496 records of offenses in the Lowell area and 58 distinct offenses. It contains the columns of Date, specific time, citation numbers, offense, and location. There are only prefixes of offense in offense, and the rest of the offense is combined with the location in the Location column. The exact name has a tab space from the locations. The adjustment on this document is that we separate the location column into apparent reason and location. Moreover, we add the longitude, latitude, and Zip Code according to the location by using Google API, which helps plot the offense on the map.

Based on the two documents above, we generate a new excel sheet “merge,” which merges the “DOTLowell\_April2021” and “Location Analysis Report - April 2021”. This document contains 2083 lines of records, having 1290 different citation numbers. The columns of “merge” include all the columns of papers above. However, there are still 111 records not fully merged, lacking information like offenders’ demographic information, specific time, and Officer ID.

The last document is the “offence-race\_makeup,” we extract columns with detailed reason and race from “merge.” After that, we group the dataset by exact reason and race, counting the number of people by race for each crime and the number of people who committed each crime. In the last Column is the race makeup percentage for each crime. There are 58 distinct offenses in this document. Also, there are 26 offenses conducted by only one race.

## QUESTION 1: EXISTENCE OF CLUSTERED ENFORCEMENT IN SPECIFIC NEIGHBORHOODS:

We plotted and found several hot spots of police enforcement throughout the Lowell area, and we identified the same streets to be Auburn street, England street, and Hale street. Those three locations have seen significantly more activity, so more knowledge is needed on those three locations. There is also a hot spot within the UMass Lowell area; however, that is disregarded in this analysis as university campuses are seen as outliers compared to the rest of the residential areas in the town. More specifically, when analyzing clusters, we ignore the university spot but still add in the racial percentage of each offense in our other analysis.

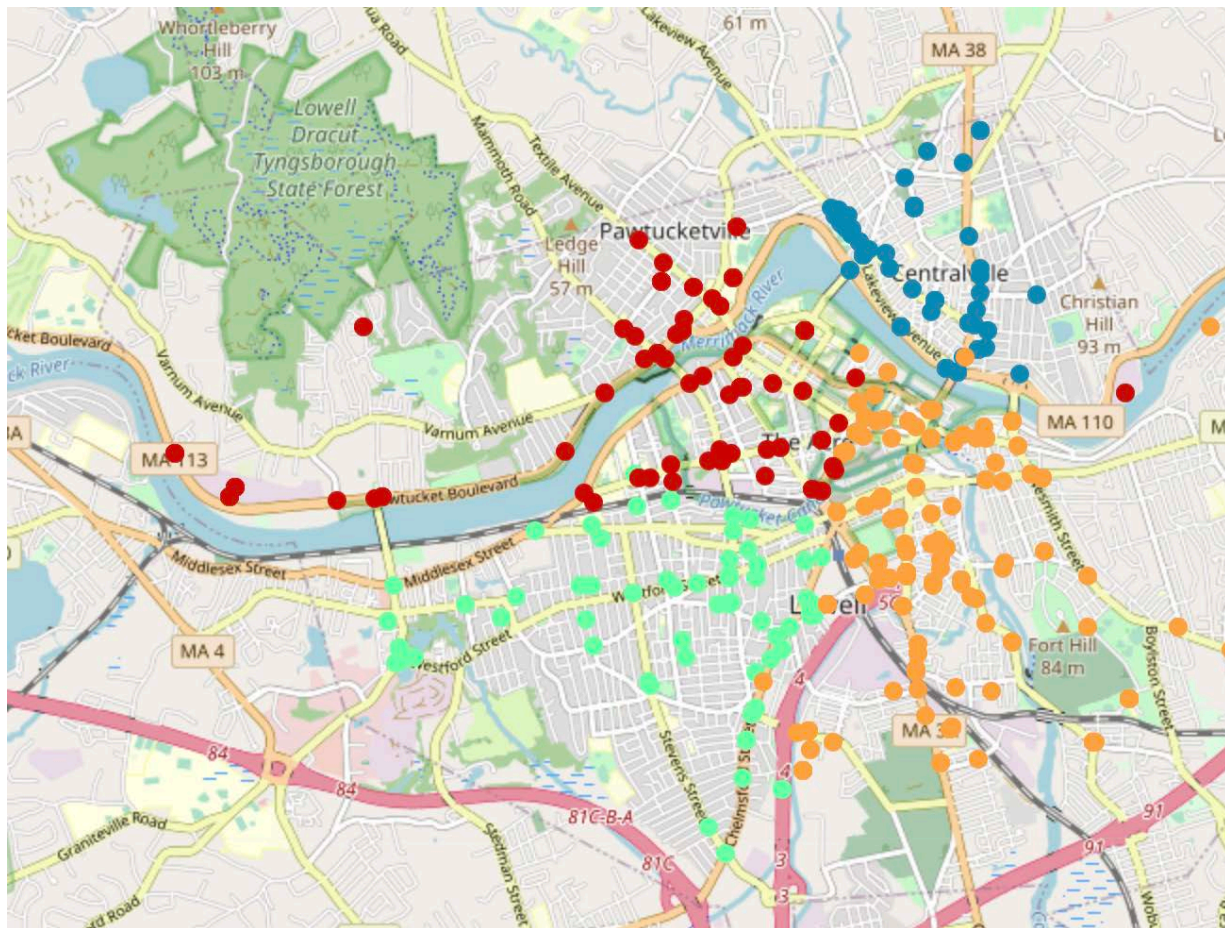
The map below shows the location of traffic stops classified by Zipcode:

237 blue plots represent Zipcode 01850 Centralville;

196 green plots represent Zipcode 01851 Highlands;

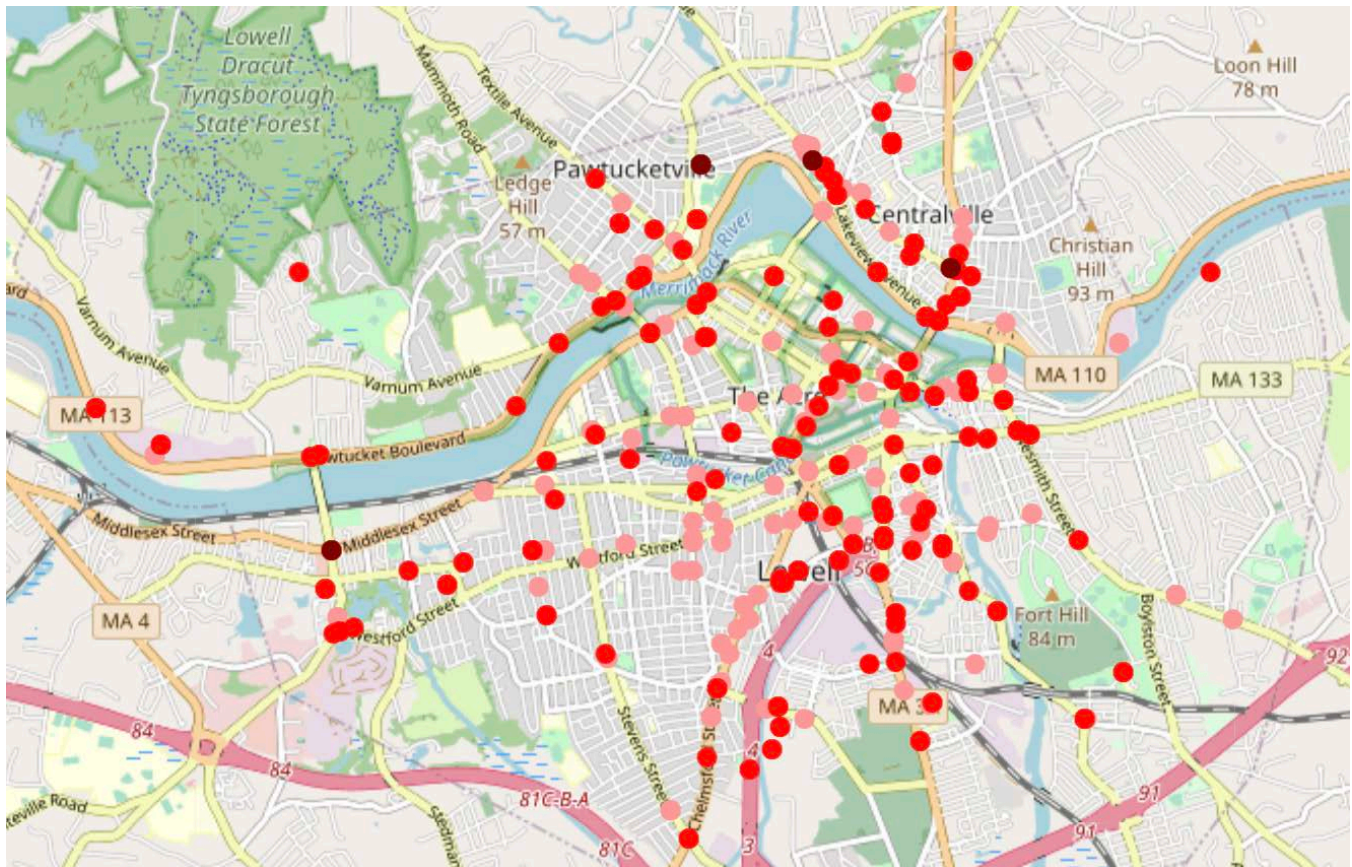
537 orange plots represent Zipcode 01852 South Lowell and Belvidere;

and 249 red plots represent Zipcode 01854 Pawtucketville and The Acre.





After we label the severity of each offense, 64.4% have the severity of 1 (light), 35.1% have the severity of 2 (moderate), and 0.5% have severity 3 (severe). The severity is determined by cross examining the Massachusetts master crimes list and the labels given there against the offenses that were given out in our dataset. Moreover, we created a map to present the severity on the map, pink represents light severity(label 1), red represents moderate severity(label 2), and dark red represents high severity(label 3):



---

## QUESTION 2: SUGGESTION OF RACIAL BIAS IN HIGH ENFORCEMENT CLUSTERS:

There are no immediate reports on the specific racial makeup of neighborhoods in Lowell. However, there is census data on the percentage of each race that lives in Lowell. Namely the distribution is as follows: White=64,240, Asian=21,513, Hispanic=18,396, Black=7,328. By using the general data in combination with the percentage of each race in police stops, we can identify if there is a certain skew to any population.

The following numbers have been computed by dividing the number of criminal cases for each race of the crime severity by the total population of that race in Lowell, a crime density. Preferably we will do this breakdown over each crime cluster using fine neighborhood race makeup data but we did not receive it or find it. Assuming that an individual of any race is equally likely to commit a crime, then the larger the crime density the more arrests are being made by enforcements implying racial bias in action. As you can see from the table below, the ordering of crime density is consistent between all severity of crime with the Black and Hispanic population, having a significantly higher crime density than White or Asian. The fact that the ordering is consistent implies that the enforcements are biased relatively consistent across different levels of crime. There isn't enough data on level 3 crimes to draw any conclusions. The crime levels were assigned by referencing the master crime data. It is possible that our assumption, that each race is equally likely to commit a crime, is incorrect. This will be addressed later.

Crime Severity	Black	Hispanic	White	Asian
1	0.015474	0.013916	0.006569	0.002510
2	0.007046	0.006305	0.004296	0.002324
3	0.000108	0	0.000031	0
All	0.023917	0.020344	0.010903	4838

Table 1: Crime Density for Races over crime severity

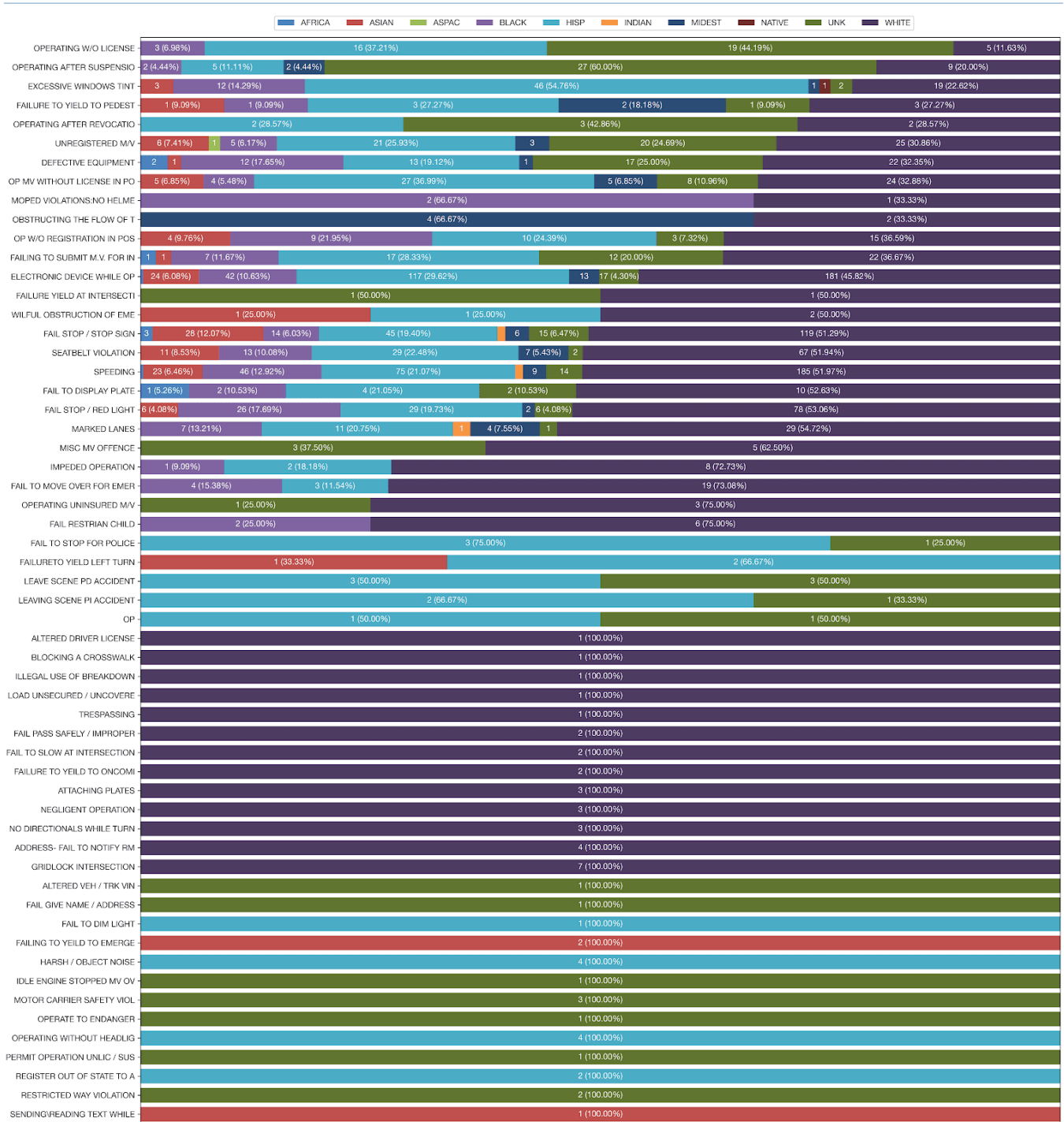
---

### QUESTION 3: DOES THE MAKEUP OF DRIVERS STOPPED IN LOWELL SUGGEST THAT CERTAIN NEIGHBORHOODS ARE TARGETED FOR ENFORCEMENT?

From the DATA USA: Lowell, MA, we could know that 48.7% of the overall Lowell population is white, 23% as Asian, 7.19% as black or African American, 16.2% as Hispanic, 0.145 as American Indian & Alaska Native. We can tell that the race ratio for each offense type is relatively stable and similar to the race ratio of Lowell population contribution.

We notice that there are nearly 45% of the offenses were conducted by single race, which have 26 out of 58. In this part, half of the offenses are conducted by white people, which are altered driver license, blocking a crosswalk, illegal use of breakdown, load unsecured/uncovered, trespassing, fail to pass safely/improper, fail to slow at the intersection, failure to yield to oncoming traffic, attaching plates, negligent operation, no directional while turn, address - fail to notify rm and gridlock intersection.

In addition, in order to visually see if enforcements are targeting specific neighborhoods depending on race and hot areas for enforcements to arrest citizens, we created a heatmap for each of the races, Black, White, Asian, Latino/Hispanic, over the location of crimes.





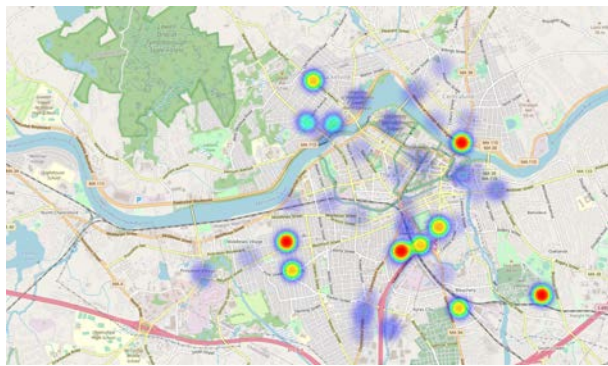


Figure 3: Asian Crime Heat Map

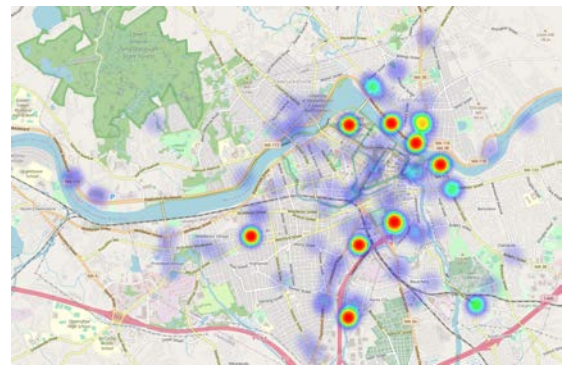


Figure 4: Hispanic/Latino Crime Heatmap

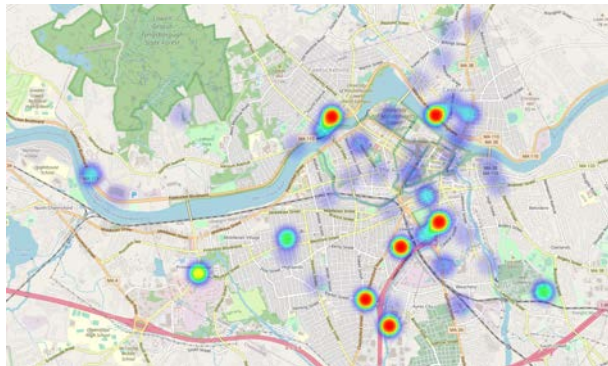


Figure 5: Black Crime Heatmap

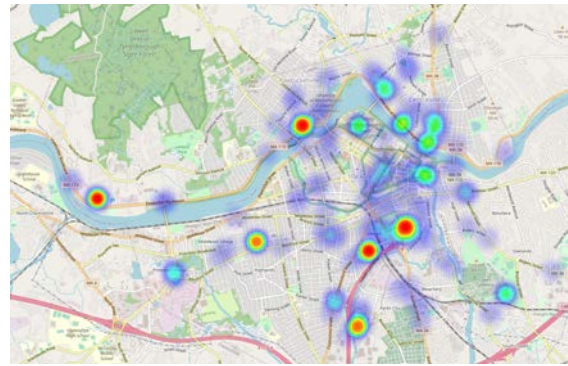


Figure 6: White Crime Heatmap

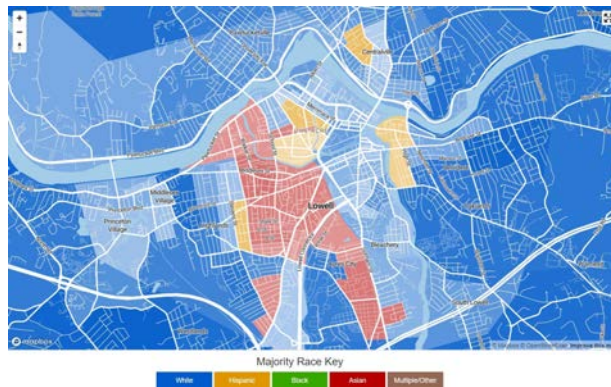


Figure 7: Majority Race Population Map

All race types have a concentrated police encounter center point at Princeton Blvd. Middlesex Village vicinity along with focus on the highway, intersection bridges with roads, Central Lowell, and Highlands. These clusters indicate that enforcement in general are focusing on the borders of Lowell and the central neighborhood of Lowell.

We first observed that the clusters for White and Asian populations are mostly the common clusters (shared between all races) and those locations are marked as Asian dominated by the majority race map so as far as we know those points are not racially correlated. In other words the enforcements



---

are not deliberately targeting those areas for racial reasons. However, the Black and Hispanic population have a highly disjoint set of clusters compared to Asian and White population.

Since the only data we have for neighborhood makeup is this majority race map that has only large area neighborhood, the Black population racial makeup is not represented at all so we can only look at Hispanic population for this analysis. For this hispanic population, their disjoint crime clusters overlap perfectly with the Hispanic population clusters as shown in the majority race map. With that being said, if it is also true for Black population (if we get population race makeup data) then we can conclude the enforcements are targeting those disjoint neighborhoods for potentially racially motivated reasons. We can say this because according to table 1, we see that Black and Hispanic populations have a disproportionately high crime density meaning those outlier clusters, a.k.a where Hispanic and Black population live the most, are accounting for the disproportionate number of arrest cases.

## REFINED SCOPE AND LIMITATIONS:

We are unable to find the specific race ratio of every race in each individual neighborhood at Lowell, as the census data was unable to provide insight at that granularity. Therefore, we went forward with more general statistics of the entire racial makeup in Lowell as a compromise.

Figure 7 is from Race, Diversity, and Ethnicity in Lowell, MA | [bestneighborhood.org](http://bestneighborhood.org), and we have requested the data they have used to come up with race distribution in areas of Lowell in order to find population clusters for each race more confidently, namely for the Black population.

## FOLLOWING PLAN:

If we had an extra month, we would have more communication with our clients and ask for more clarifications about the LPD Lawsuit Civil Rights, since some column names are missing, and we cannot tell from the contents in that column. Specifically, the data is corrupt at the point of uploading to google sheets and missing all rows after 100. The 101th row is a repeating data that goes on for 10,000+ columns, likely some overflow corruption.

Two questions that can be answered with clarified data and more time are:

- Are there any way of classifying officers that have been filed a complaint that was justified, officers that have been complained and unjustified, and the rest of the officers using their arrest history and officer data?
- Are there any way of classifying officers that have been filed a complaint that was justified, officers that have been complained and unjustified, and the rest of the officers using their arrest history and officer data?

Furthermore, we emailed Dr. Persily for detailed race makeup for the smaller areas in Lowell, but we did not hear anything from him so far. If people who continued this project received this data, they could compare the race makeup of traffic stops in smaller parts of the Lowell area and the overall race makeup in smaller parts of the Lowell area. Specifically, we can figure out how statistically

---

significant the clusters of the crimes produced for each race correlate with the corresponding race's habitation clusters in Lowell. The higher correlation or lower mean square error suggests less bias by the enforcement. Therefore, more research can be conducted on the racial bias of traffic stops

In addition, for our analysis of question 2, we assumed that the chance for an individual to commit a crime is the same for all races which is not necessarily true. In order to find out the true crime commit rate for each race in order to scale the crime density, an approach could be to analyze the income of each race in Lowell and consult research on income vs crime rate (independent of race). This requires, at the minimum, the income map of Lowell and together with the low level racial makeup of the Lowell population we can find out the income level of each race.