

Making Sense of Police Field Interrogation and Observation (FIO) Data

Citizens for Juvenile Justice (CFJJ)

Farid Karimli, Akshey Nischal, Noah Jean-Baptiste, Christina Xu, Ruiding Feng

December 2022

Table of Contents

I.	Introduction.....	1
II.	Individual Municipality Analysis.....	2
III.	Comparative Municipality Analysis.....	33
IV.	Limitations.....	36
V.	Conclusion.....	37

Introduction

This study was done in collaboration with Citizens for Juvenile Justice (CfJJ) to investigate police field and interrogation data (FIOs). Their mission is to advocate for statewide systemic reform that achieves equitable youth justice. CfJJ is the only independent, non-profit, statewide organization working exclusively to reform and reimagine the juvenile justice system in Massachusetts. CfJJ aims to educate the public on important juvenile and justice issues through community engagement and conduction of research.

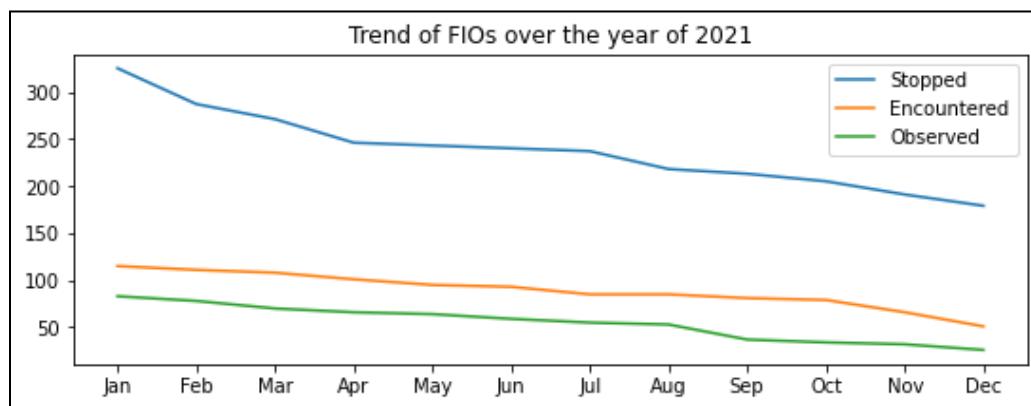
The motivation behind this study was to support their mission by trying to gain insight into policing practices and policies. Police departments in Massachusetts record FIOs which are typically interactions with members of the public that do not lead to an arrest. However, there is a large discrepancy between police behavior and data collection. While there is a lot of data on arrests, there exists a lack of data in police interrogations. Additionally, FIOs can be misleading as they can simply record a police sighting of a person. Furthermore, when CfJJ requested FIO data from police departments in some towns in Massachusetts, they refused to give them access which raises questions about their ethics. FIOs can be problematic as they can be misused to justify stop and frisks and misidentify individuals who are gang members which can potentially lead to negative outcomes for those individuals in society. FIO data has been gathered and put together on nearly every municipality in the state over the past few years. Although they are not uniform across municipalities, they share similar features such as race, age, location, and time data. Thus, the objective of this study is to draw conclusions about policing in specific municipalities, patterns between municipalities, demographics on who are stopped, and trends around time of day they are stopped based on the people that are subject to FIOs.

Individual Municipality Analysis

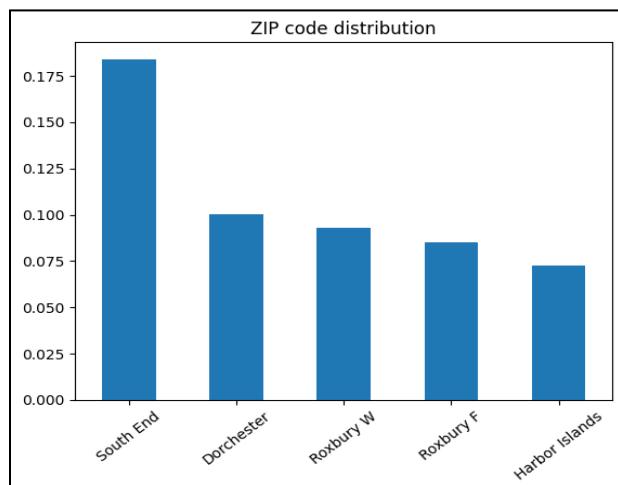
Boston

Data description: The data tables came with several columns that had no values whatsoever, which were deleted. The analysis on Boston focused on trends in the number of FIOs, the relationship between FIO and race proportions, and location.

Results: The total number of FIOs in Boston in 2021 was 4490. If we look below, we can see the trend in the number of FIOs color-coded by circumstance. Circumstance is the attribute that describes the nature of an FIO. It takes three values: Encountered, Observed and Stopped. We can observe that FIOs for each circumstance have been steadily decreasing over the course of 2021 with “Stopped” FIOs being much more frequent in comparison.

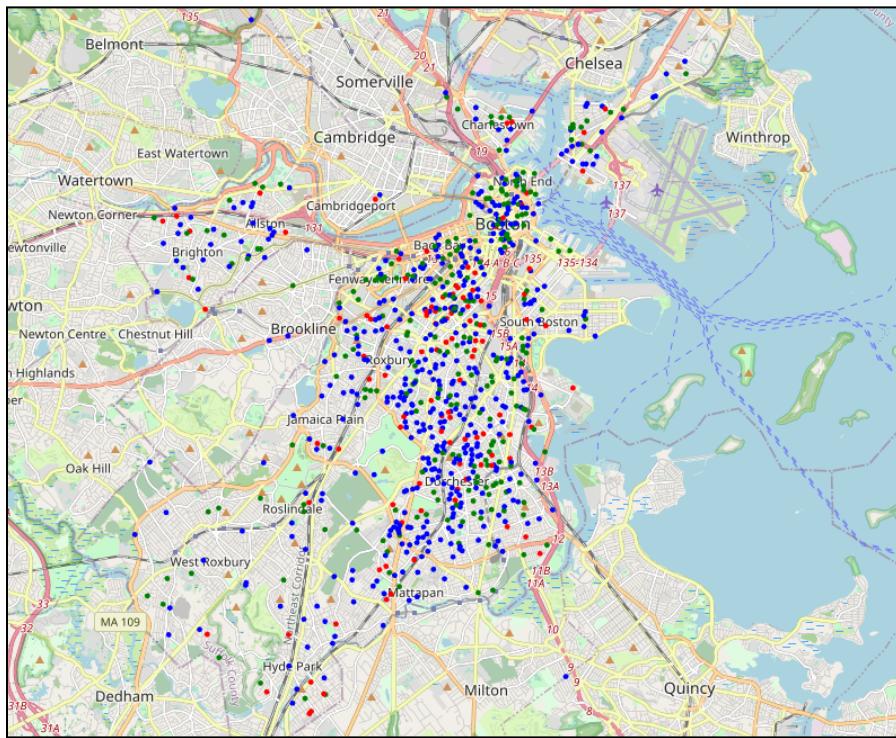


We then proceeded to do location analysis. The Boston tables came with clean street values. Using the Google Maps API, we were able to get the coordinates of each FIO’s street, and mark it on a map of Boston. The markers were color-coded by circumstance as before. Also, the tables had a ‘zip’ attribute, which corresponded to the zip code of the town where the FIO occurred. So, we also plotted the 5 regions in Boston with the highest (normalized) number of FIOs (the regions came from the zip code).



By looking at the bar chart we see that South End had the most FIOs in 2021 (around 18%), followed by Dorchester. The W and F letters for the two Roxbury bars correspond to Washington Park and Franklin Field regions in Roxbury.

The map of Boston visualization confirms that, we see a heavy concentration of markers in South of Boston, Roxbury and Dorchester. The vast majority of markers are blue, reinforcing the fact that most of the FIOs resulted in a stop. Red markers indicate ‘Encountered’, and green markers indicate ‘Observed’ FIOs. An interactive map is available [here](#).



We collected census data for each region to get the population proportions of each race so that we can draw comparisons between FIO and race proportions. The following table displays race information for each of the 5 regions displayed above.

Region	White	African American	Hispanic	Asian	Other
South End	53%	10%	15%	19%	3%
Dorchester	23%	43%	20%	9%	5%
Roxbury W (02119)	6.7%	49.6%	27.7%	2.1%	13.9%
Roxbury F*** (02121)	5%	60%	32%	1%	2%
Harbor Islands	29%	31%	21%	11%	9%

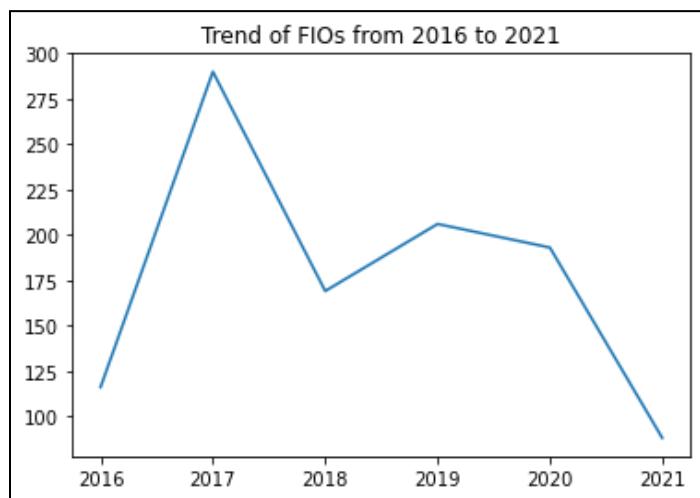
*** The margin of error is 10%

Out of the five, South End has the highest White population proportion, lowest Black population proportion and the highest number of FIOs. Every other neighborhood has a generally much higher Black population proportion than that of white, but they have at least two times less FIOs than South End. The two Roxbury neighborhoods have the highest Hispanic population proportions, and their number of FIOs is very close (each almost 10% of all FIOs).

Brockton

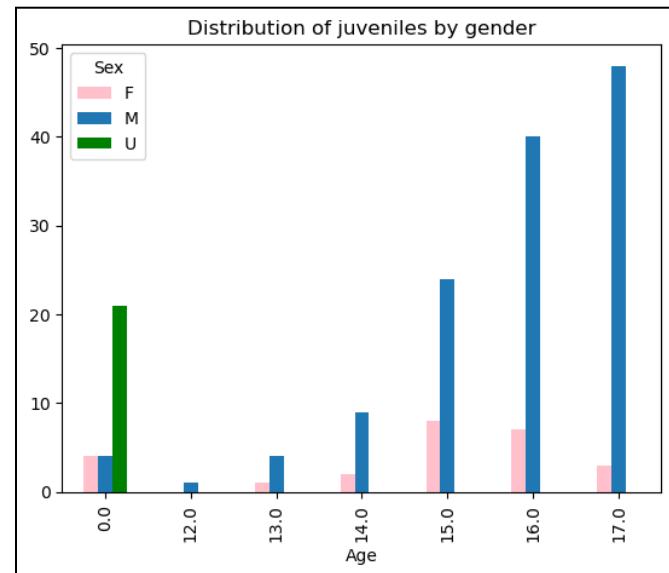
Data description: Brockton tables were some of the cleanest of the entire set of FIO data. It had date, race, and age information for each FIO, but no location data. So, we were able to perform more effective analysis for Brockton than we did for Boston. The results are shown below.

Results:



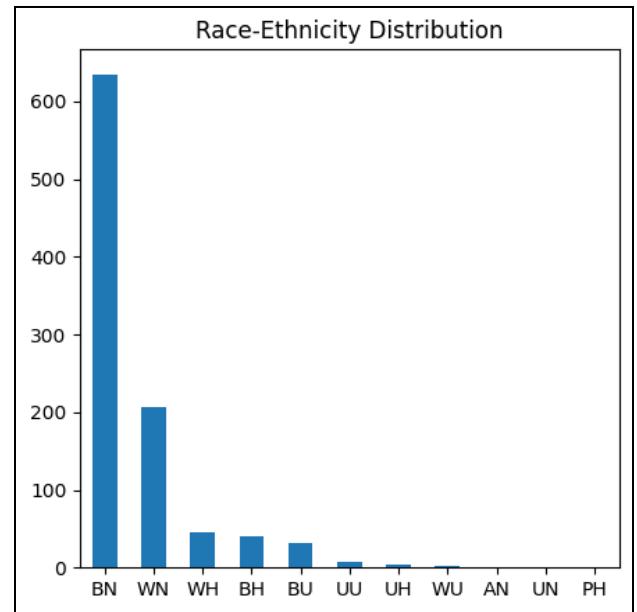
The data included FIOs in the years 2016-2021. The number of FIOs in Brockton peaked in 2017, and then dropped sharply. There's also a decrease starting from 2020, which we can account for the COVID-19 pandemic.

Since Brockton data had information about age, we were able to gain insight into the FIO involvement of juveniles. On the left, we can observe the distribution of the ages of juveniles involved in FIOs in Brockton since 2016. Ages 16-18 dominate the numbers, with a really small number of FIOs involving 12 and 13 year-olds. There is also some irregular data. Some data points had missing gender values with the age of 0, as seen by the green bar, and some data points that did have gender had age values of 0.



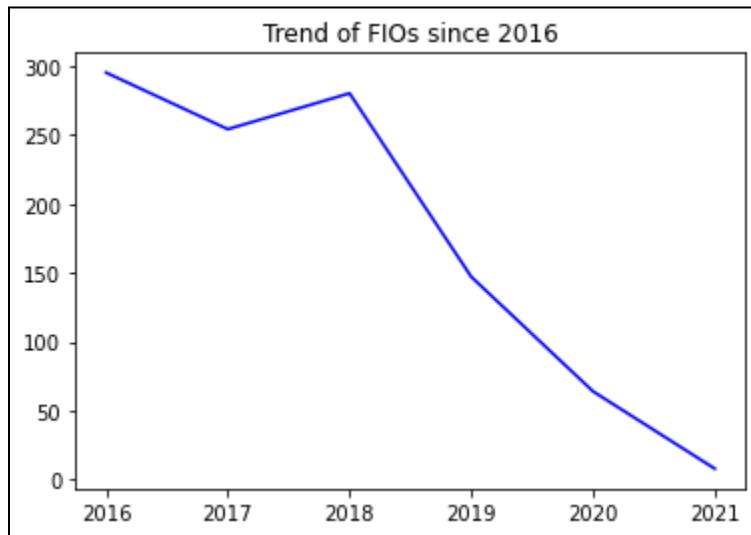
Brockton's dataset contained really clean race and ethnicity values. Since the ethnicity values were only H for Hispanic and N for non-Hispanic, we concatenated the race and ethnicity fields into one column. The column contained values like "WH" for White and Hispanic, "BN" for Black and non-Hispanic and etc. "A" corresponds to Asian, "P" to Pacific and "U" for unknown.

The resulting distribution can be seen on the right. As we can see, the vast majority of FIOs involve Black non-Hispanics, over 600 out of a total of 992 starting from 2016 (around 60%). The black population percentage of Brockton is nearly 44%, according to the U.S. Census Bureau. White non-Hispanics are the second largest, around 20% of total FIOs. White population proportion of Brockton is almost 34%. Even though the Black population proportion is approximately 10% points larger than that of the White population, Black non-Hispanics were involved in twice more FIOs than White non-Hispanics.

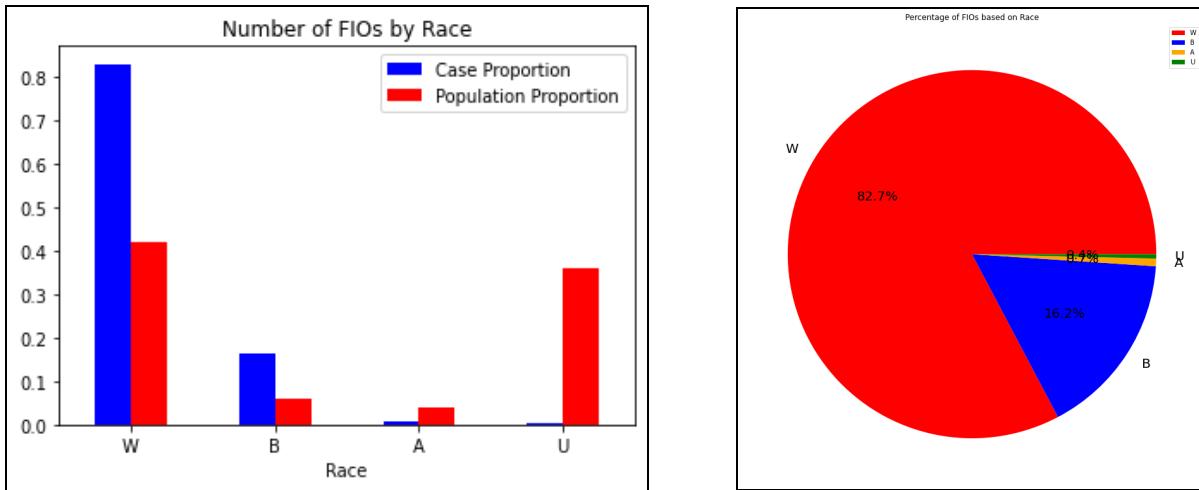


Chelsea

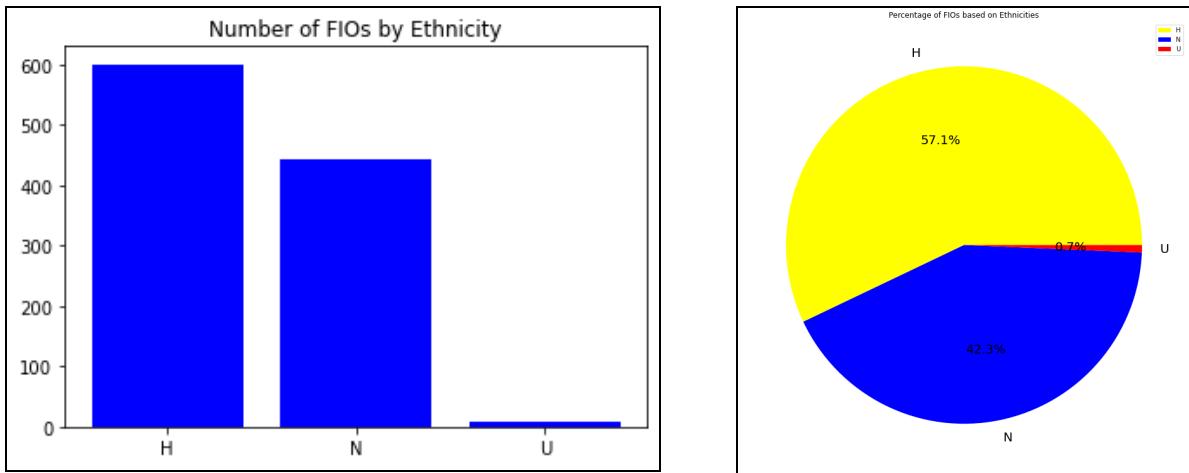
Data description: The data provided to us for Chelsea includes the following relevant data for FIOs: FieldNum, GangRelated, Age, Sex, Race, Ethnicity, and Location. With this data, we are able to draw conclusions on various fields of FIO data and plot frequencies and trends based on these observations.



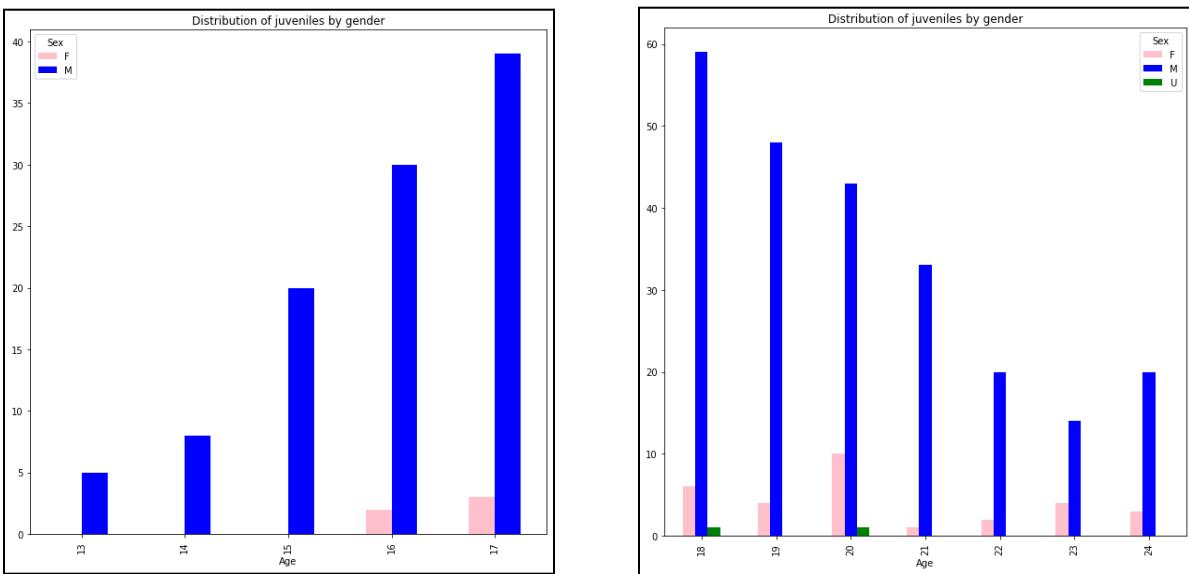
Results: Let's start by looking at the total distribution of cases within Chelsea. In total, there were 1052 cases in Chelsea from January 2016 until May 2021. Of these cases, 295 were in 2016, 255 were in 2017, 281 were in 2018, 149 were in 2019, 64 were in 2020, and 8 were in 2021. Excluding the jump from 2017 to 2018, we notice that there is an overall decline in the number of cases per year. It is worth noting that 2020 and 2021 have skewed data as the Covid-19 pandemic severely impacted the number of cases occurring.



When taking a look at the Race distributions for these cases, we are provided with 7 main classifications of race: Asian, Black, Unknown, and White. Of these race classifications, 7 are Asian, 170 are Black, 4 are Unknown, and 871 are White. There exists a large discrepancy between white cases and the rest of the races. Thus, it is important to look at the census data to see if the trends match the population distributions. According to the US 2020 census data, there are 16,910 white individuals, 2,491 black individuals, 1,575 asians, and 14,520 unknown individuals. Taking a look at the proportions we have 83% of the cases being committed by 41% of the population, 16% of the cases being committed by 6% of the population, .007% of the cases being committed by 4% of the population, and .003% of the cases being committed by 36% of the population. The two obvious discrepancies occur within the White population and Unknown. The skewness shown in Unknown races is due to the lack of classifications done by police.

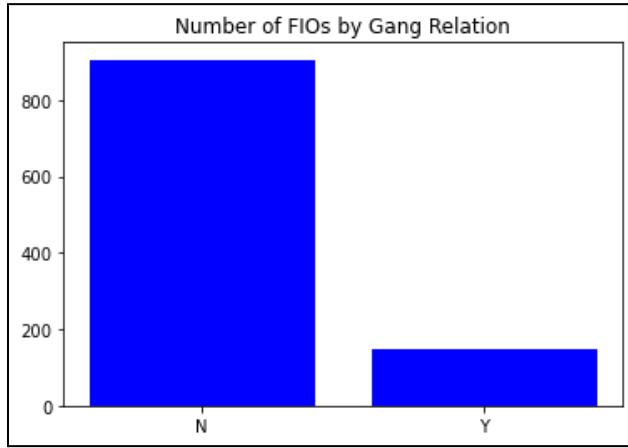


The same way we looked at race, we can also look at ethnicity. The ethnicities within Chelsea are identified as Hispanic 600 cases, Non-Hispanic 444 cases, and unknown 8 cases. Looking at the census data, there are 26,999 hispanics and 12,879 non-hispanics. The cases numbers shown for hispanic, non-hispanic, and unknown people show a strong correlation with the population distribution within the municipality.

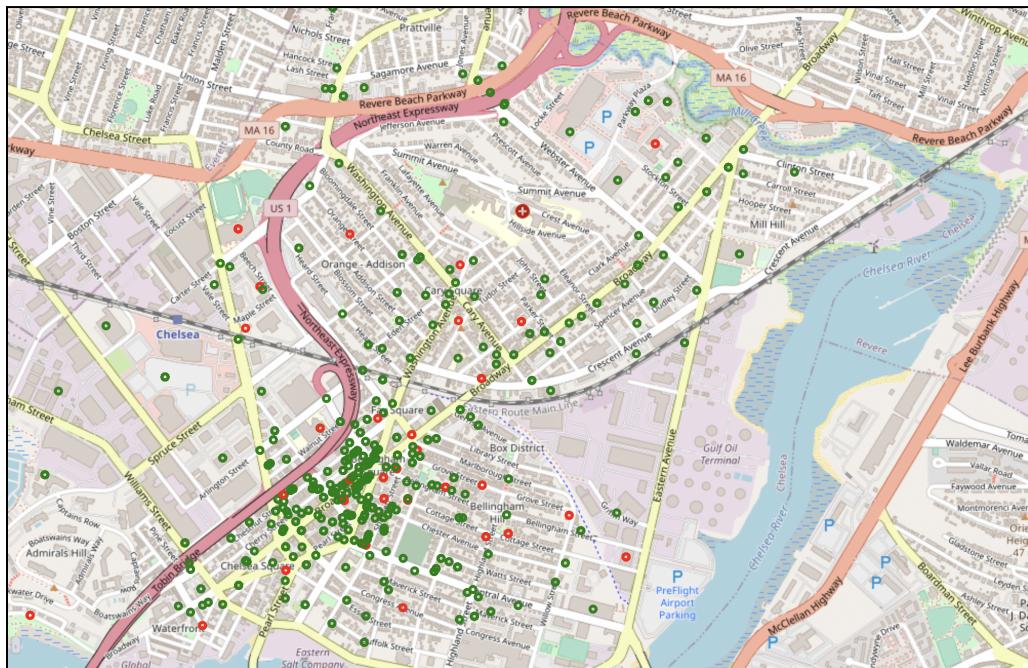


The data for Chelsea can be interpreted through age ranges of cases. Lets start by looking at those cases which occur amongst juveniles. In Chelsea, we see cases from ages 13-17. Of these cases, a majority are male and a small amount of cases involve females. This trend is consistent with the rest of the data as there is a large skew towards males in the dataset. If we

take a look at the cases for ages 18-24, we see a similar trend in skewness towards males. However, unlike in juvenile cases, female cases occur more frequently and follow a similar trend to that of male cases. Despite following the trend, there is still quite a big disparity in cases based on genders.



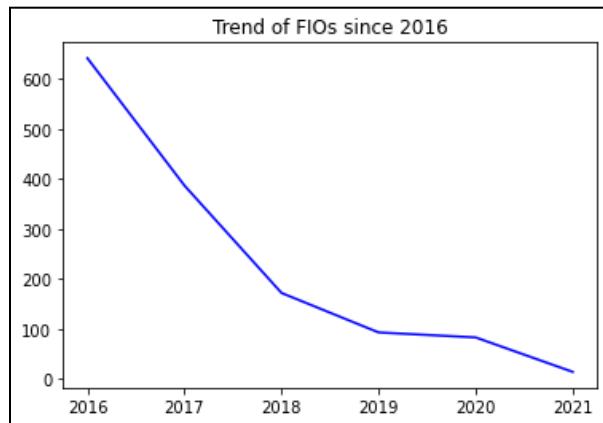
Chelsea is interesting to look at since the data also includes gang related cases. In this case most of the cases are not gang related. However, those that were gang related, tend to occur in the same areas. Thus, we have produced a map showcasing all the cases and identifying by being gang related or not.



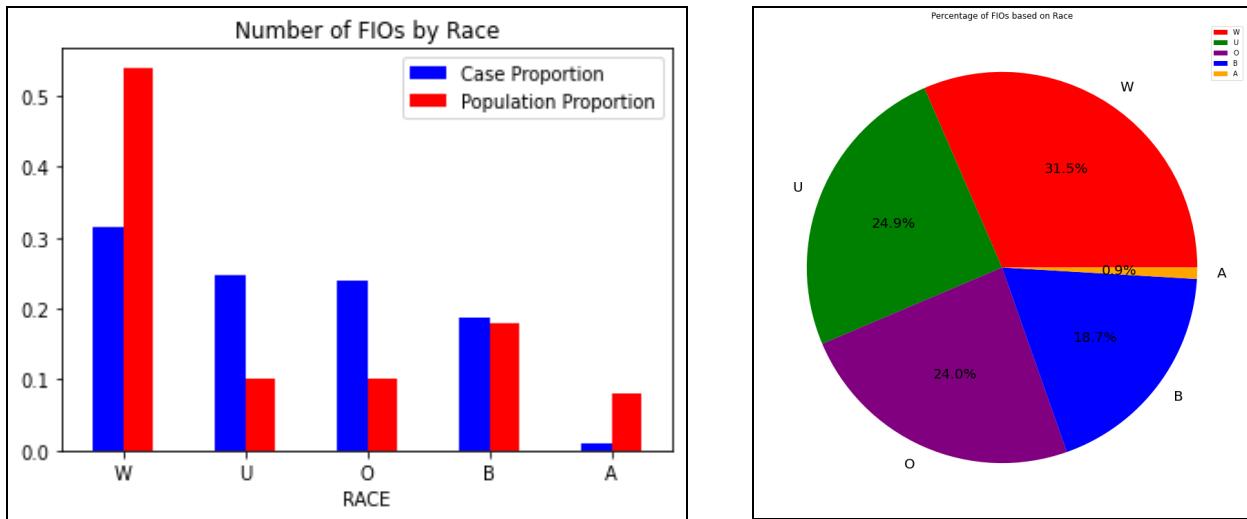
Given that Chelsea was one of the few municipalities to give us gang data and locations, we decided to plot the location of Fios based on whether or not they were gang related. The map above indicates that an FIO was gang related if the circle is red and not gang related if the circle is green. By plotting these points, we see that most gang related cases occur near others. While the green dots don't show any grouping, we see the red ones occur in clusters, thus indicating potential gang activities occurring in those areas.

Everett

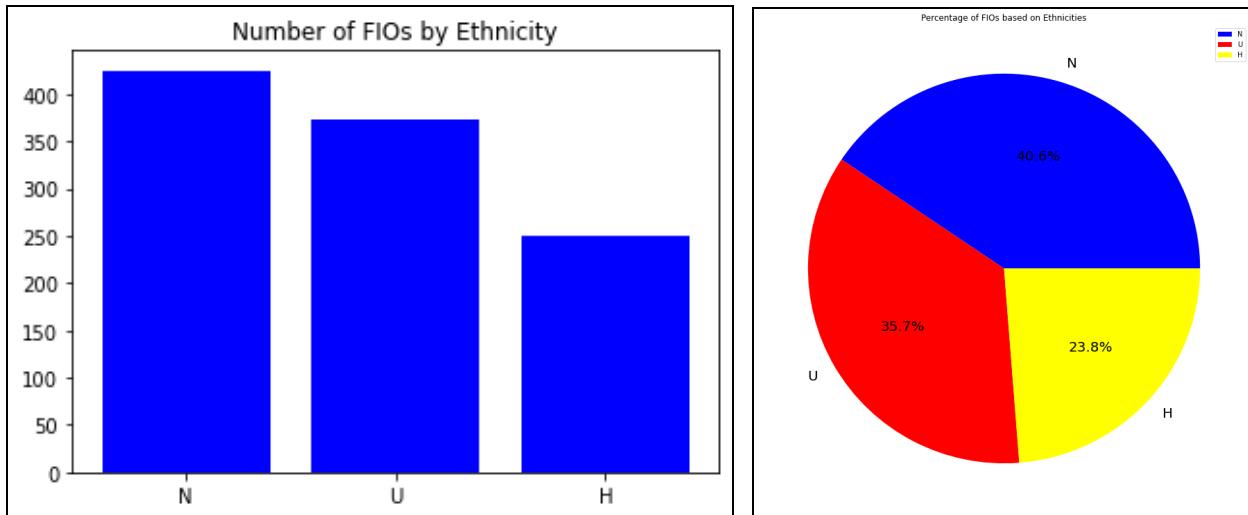
Data description: The data provided to us for Everett includes the following relevant data for FIOs: FioNum, Age, Sex, Race, Ethnicity, and Location. With this data, we are able to draw conclusions on various fields of FIO data and plot frequencies and trends based on these observations.



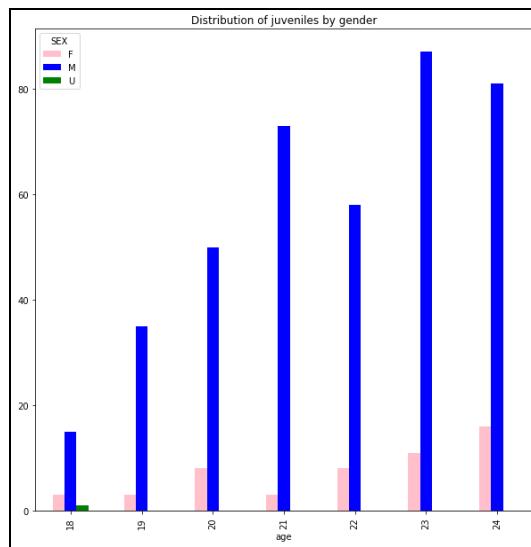
Results: Let's start by looking at the total distribution of cases within Everett. In total, there were 1052 cases in Everett from 2016 until 2021. Of these cases, 642 were in 2016, 387 were in 2017, 172 were in 2018, 93 were in 2019, 83 were in 2020, and 14 were in 2021. We notice that there is an overall decline in the number of cases per year. It is worth noting that 2020 and 2021 have skewed data as the Covid-19 pandemic severely impacted the number of cases occurring.



When taking a look at the Race distributions for these cases, we are provided with 7 main classifications of race: Asian, Black, Unknown, White, Mixed, Hispanic, and American Indian. For the sake of normalizing the data, we will include Mixed, American Indian, and Hispanic under the Unknown classification. Of these race classifications, 13 are Asian, 259 are Black, 345 are Unknown, 437 are White, and 333 are other. When looking at the graph of this data, it is quite apparent that the outlier is Unknown. This is due to the normalization we have done, primarily in categorizing Hispanics under unknown, rather than under whichever respective category they fall under. Thus, it is important to look at the census data to see if the trends match the population distributions. According to the US 2020 census data, there are 25,017 white individuals, 8,246 black individuals, 3,640 asians, 4,703 other individuals and 4,669 unknown individuals. Taking a look at the proportions we have 31% of the cases being committed by 54% of the population, 25% of the cases being committed by 10% of the population, 24% of the cases being committed by 10% of the population, 19% of the cases being committed by 8% of the population 36% of the cases. It is interesting to note that those grouped in other and unknown have similar case numbers as well as population distributions.

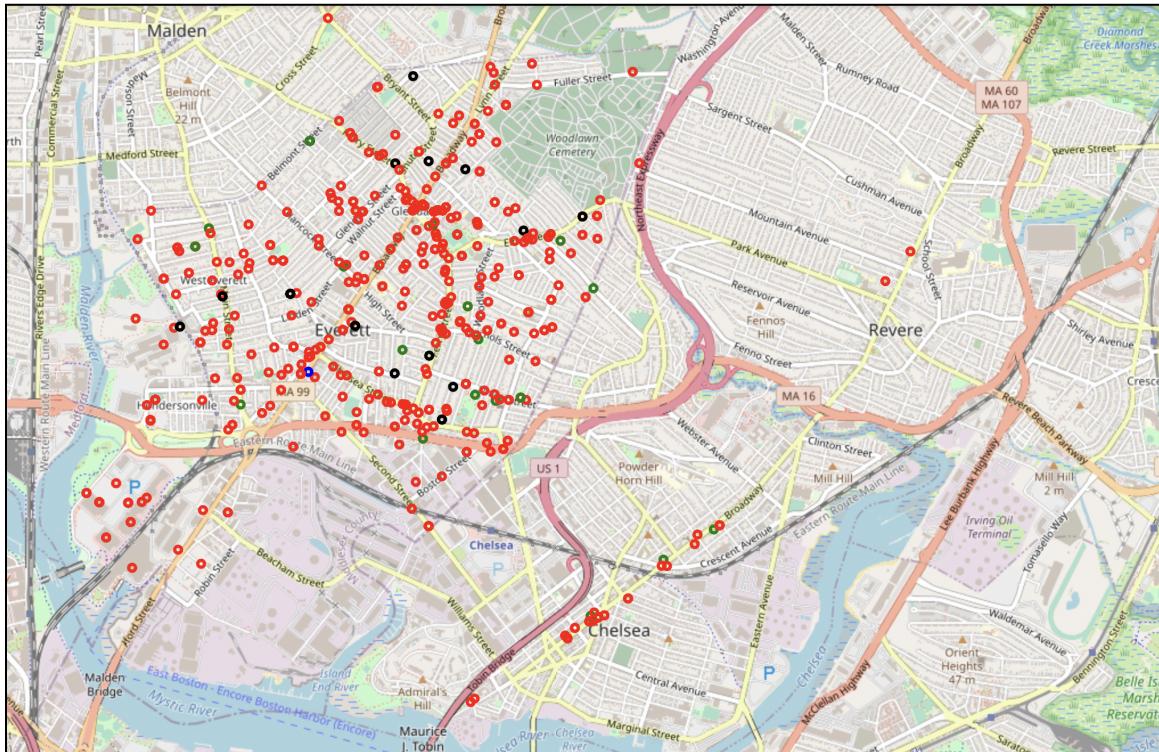


The same way we looked at race, we can also look at ethnicity. The ethnicities within Everett are identified as Hispanic 249 cases, Non-Hispanic 425 cases, and unknown 373 cases. In order to normalize the data, we will interpret the unknown as non-hispanic. Thus, we have 249 Hispanic cases and 798 Non-Hispanic cases. Looking at the census data, there are 13,925 hispanics and 32,750 non-hispanics.



The data for Everett can be interpreted through age ranges of cases. Let's start by looking at those cases which occur amongst juveniles. In Everett, we see cases primarily for the age of 17,

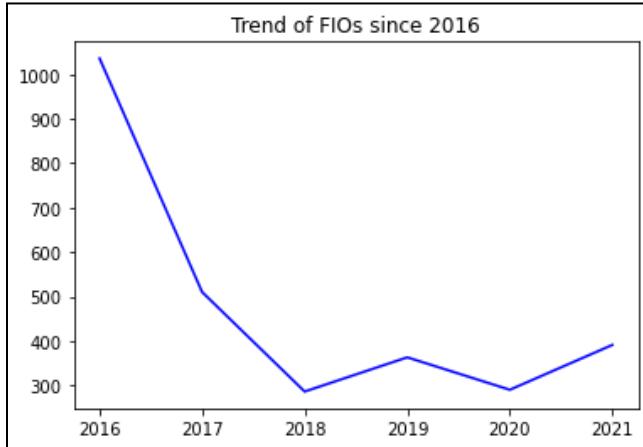
with a few outliers occurring at 6 and 9. Of these cases, all those occurring at 17 are male while the females make up the two outlier cases. This trend is consistent with the rest of the data as there is a large skew towards males in the dataset. If we take a look at the cases for ages 18-24, we see a similar trend in skewness towards males. However, unlike in juvenile cases, female cases occur more frequently and follow a similar trend to that of male cases. Despite following the trend, there is still quite a big disparity in cases based on genders.



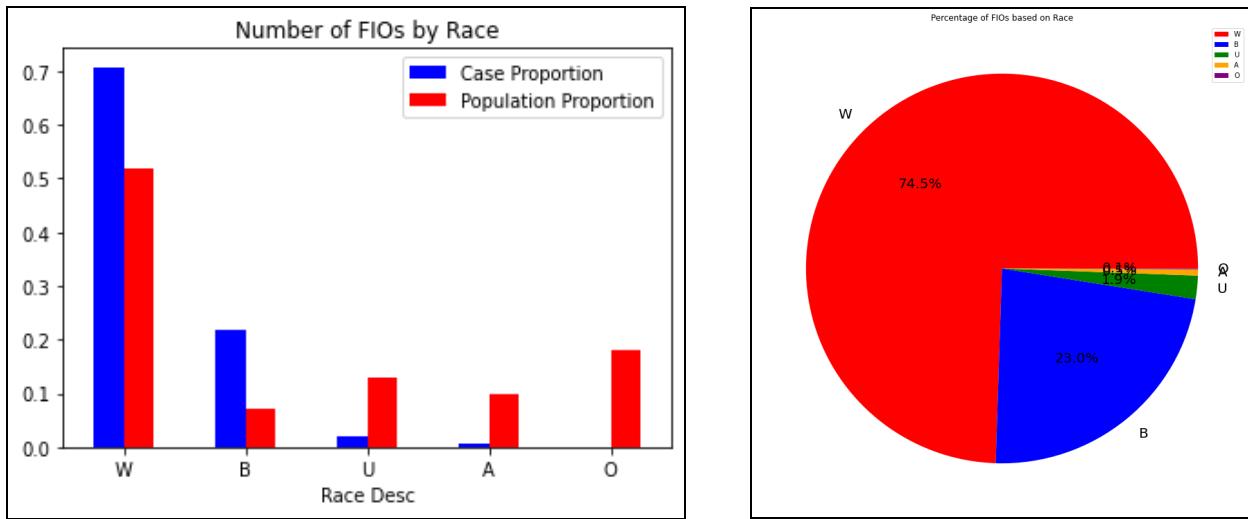
Everett provided us with data regarding locations of Fios, thus we decided to plot these locations based on Fio type. The red dots represent type I, the blue dots represent type A, the green dots represent type O, and the black dots are type S. Based on the map, it is quite apparent that the dominant form of Fios in Everett are of type I, while the others happen quite rarely in comparison to type I.

Framingham

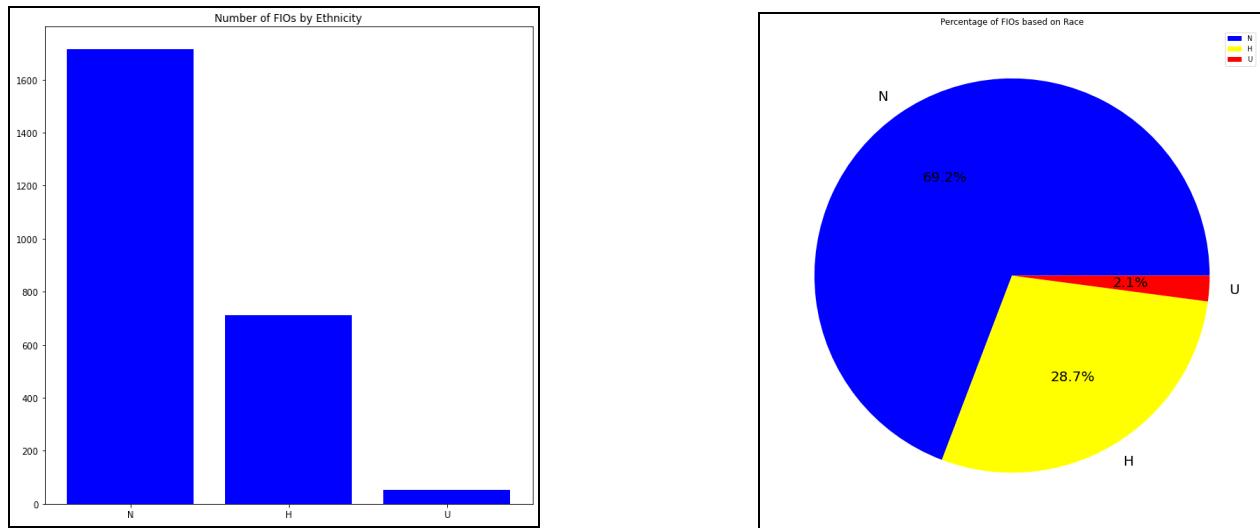
Data description: The data provided to us for Framingham includes the following relevant data for FIOs: FioNum, Age, Race, Hispanic, and Location. With this data, we are able to draw conclusions on various fields of FIO data and plot frequencies and trends based on these observations.



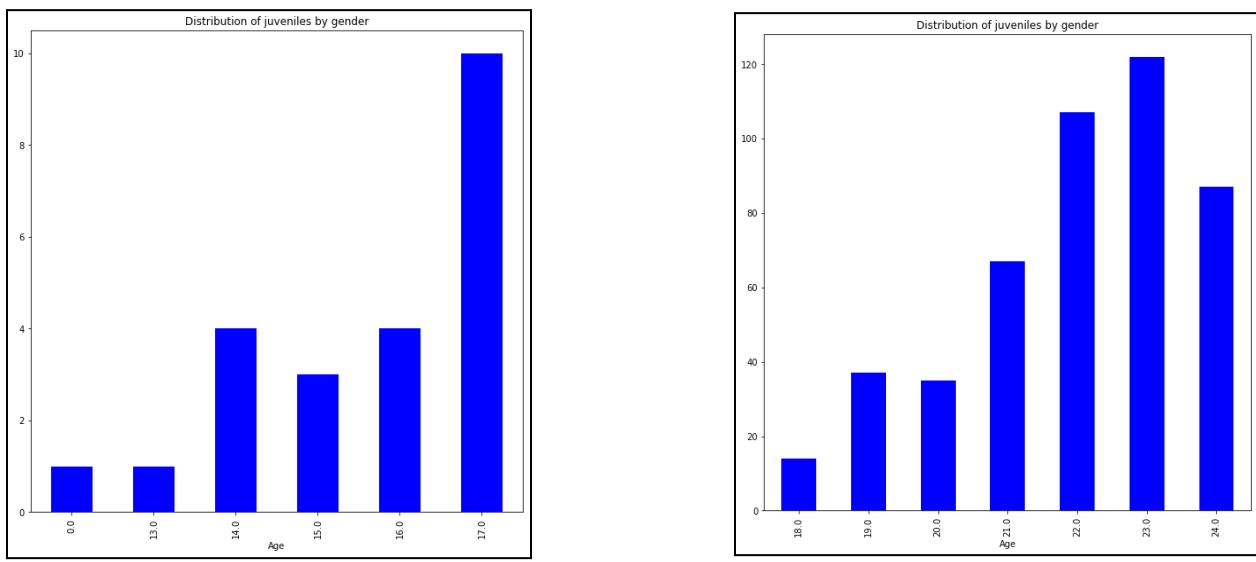
Results: Let's start by looking at the total distribution of cases within Framingham. In total, there were 1052 cases in Everett from 2016 until 2021. Of these cases, 1,036 were in 2016, 510 were in 2017, 286 were in 2018, 363 were in 2019, 290 were in 2020, and 391 were in 2021. We notice that there is an overall decline in the number of cases per year. It is worth noting that 2020 and 2021 have skewed data as the Covid-19 pandemic severely impacted the number of cases occurring.



When taking a look at the Race distributions for these cases, we are provided with 7 main classifications of race: Asian, Black, Unknown, White, and Native Hawaiian. For the sake of normalizing the data, we will include Native Hawaiian under the Unknown classification. Of these race classifications, 14 are Asian, 628 are Black, 55 are Unknown, and 2,035 are White. When looking at the graph of this data, it is quite apparent that the outlier is White. There exists a large discrepancy between white cases and the rest of the races. Thus, it is important to look at the census data to see if the trends match the population distributions. According to the US 2020 census data, there are 37,396 white individuals, 4,771 black individuals, 7,133 asians, 12,740 other individuals, and 9,220 unknown individuals. Taking a look at the proportions we have 71% of the cases being committed by 52% of the population, 22% of the cases being committed by 7% of the population, 2% of the cases being committed by 13% of the population, 1% of the cases being committed by 10% of the population, and .01% of the cases being committed by 18% of the population. In this township, there is a heavy bias on black FIO cases as they account for 22% of cases, while only making up 7% of the population.

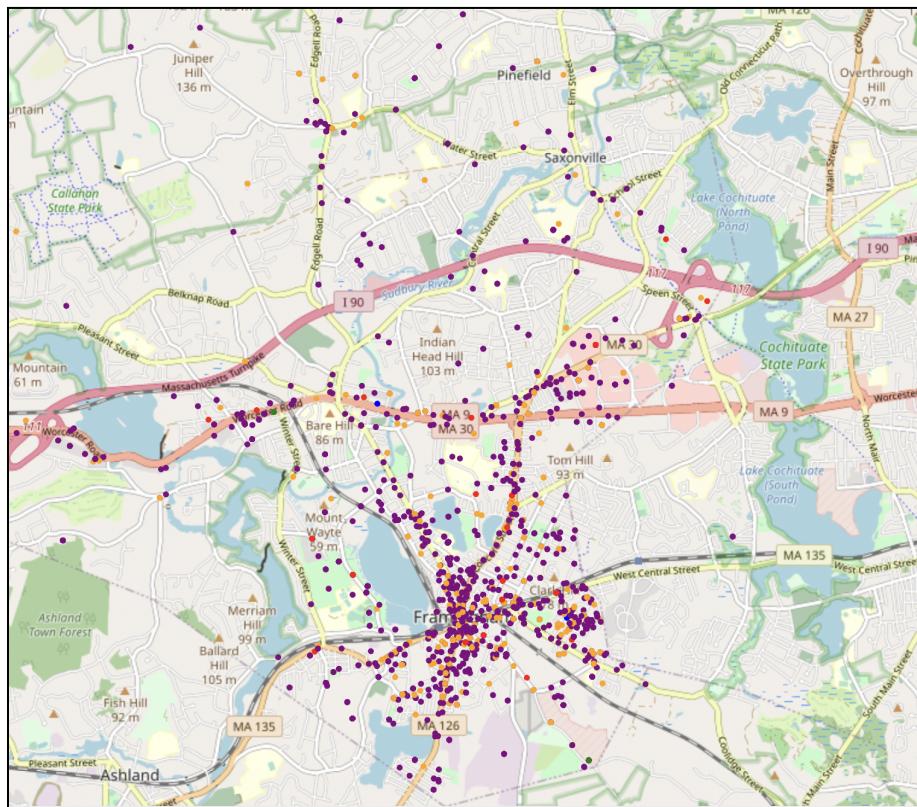


The same way we looked at race, we can also look at ethnicity. The ethnicities within Framingham are identified as Hispanic 710 cases, Non-Hispanic 1,714 cases, and unknown 308 cases. In order to normalize the data, we will interpret the unknown as non-hispanic. Thus, we have 710 Hispanic cases and 2,022 Non-Hispanic cases. Looking at the census data, there are 13,757 hispanics and 57,503 non-hispanics.



Unlike the other municipalities, we were only given ages for Framingham and not genders. Thus, we have plotted the age ranges for juveniles and ages 18-24 without classifying gender.

Based on the graphs the number of cases don't seem to spike until we reach age 21. From there onwards, the number of cases based on age range from 70 up to around 120.

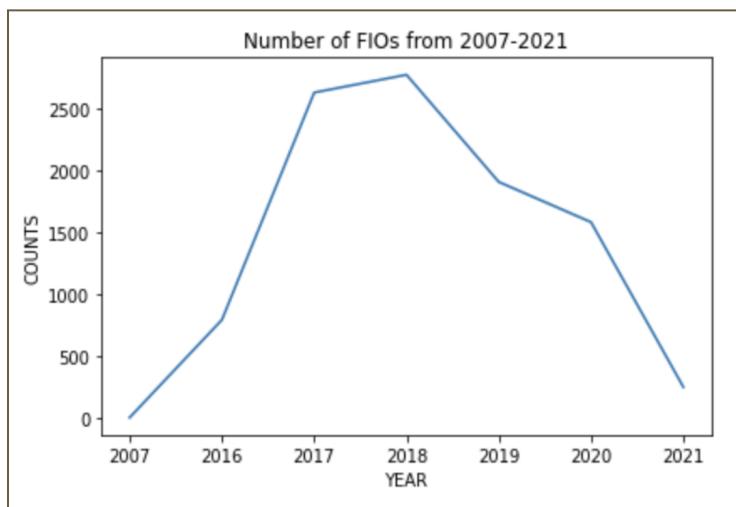


Framingham provided us with data regarding locations of Fios, thus we decided to plot these locations based on race. The blue dots represent Asians, orange dots represent Black, red dots represent Unknown, purple dots represent White, and green dots represent other. The map reveals that a majority of Fios occur in the central part of Framingham with a large portion also occurring in the eastern most part of Framingham. It also reveals that races are clustered together, as we see large numbers of purple and orange dots grouped together, thus indicating a large portion of that region's population being of that race.

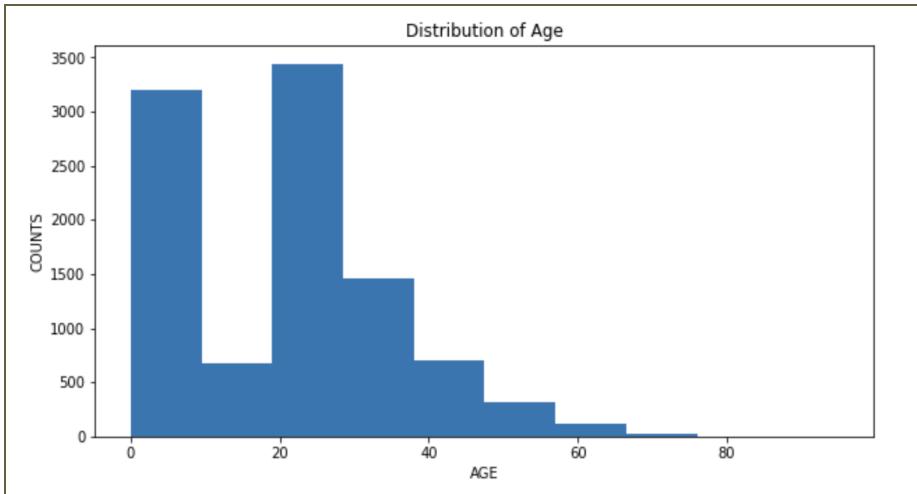
Lowell

Data description:

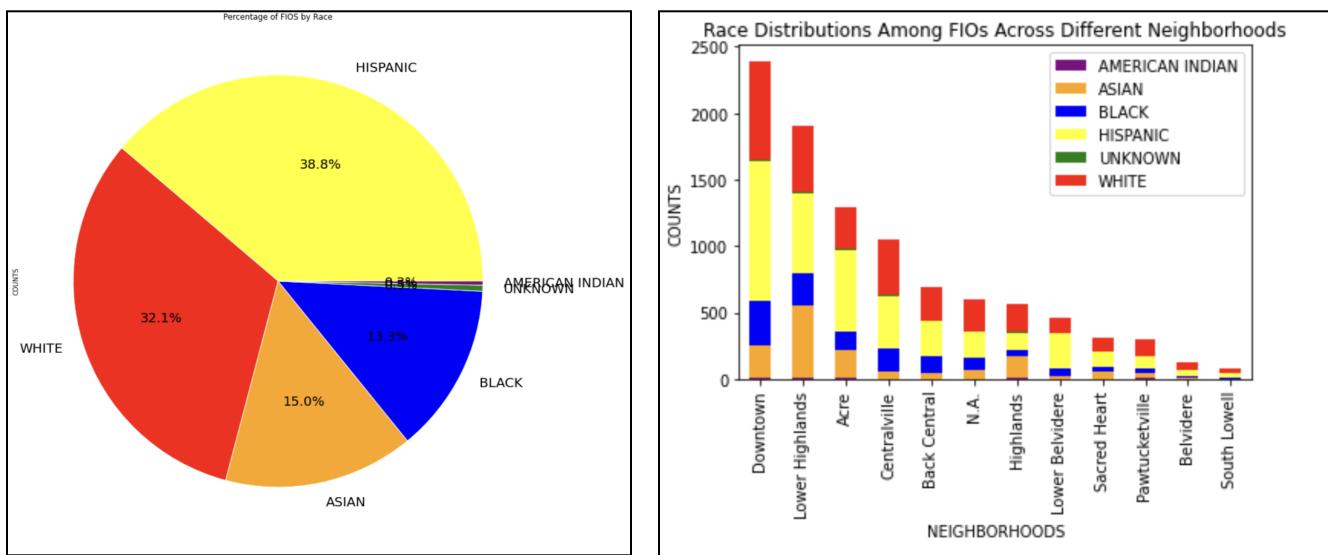
The FIOs for Lowell featured the report number, name of the individual recorded, address of the interaction, APT (unsure of what this field represents), police sector and neighborhood of the interaction, age, gender and race of the individual recoded, the reporting officer's name, whether the individual was determined to have gang relations, and the interaction date. Out of these fields the name of the individual recorded and APT were unusable because they had over 90% null values.



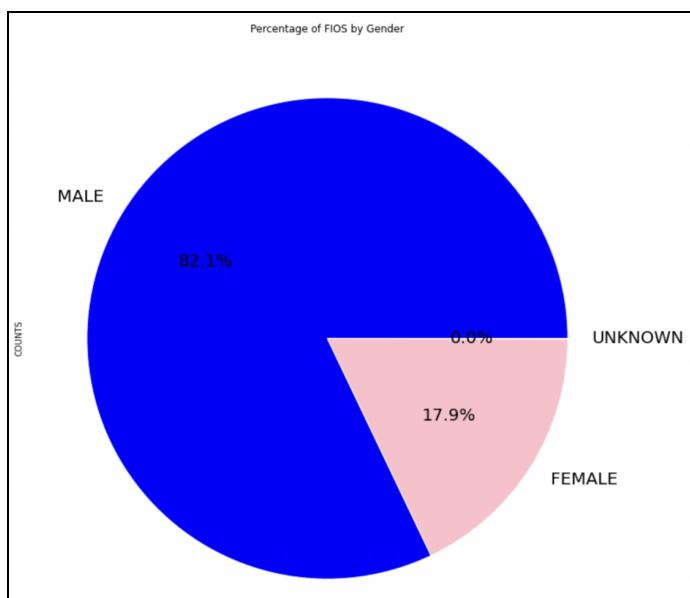
FIOs in Lowell were unique in that they dated back to 2007 whereas in most towns, they dated back to only 2016. From 2007 to 2018, there was an increase in the number of FIOs in Lowell and beyond 2018 to 2021, there was a decrease, similar to the trend observed in other towns across Boston.



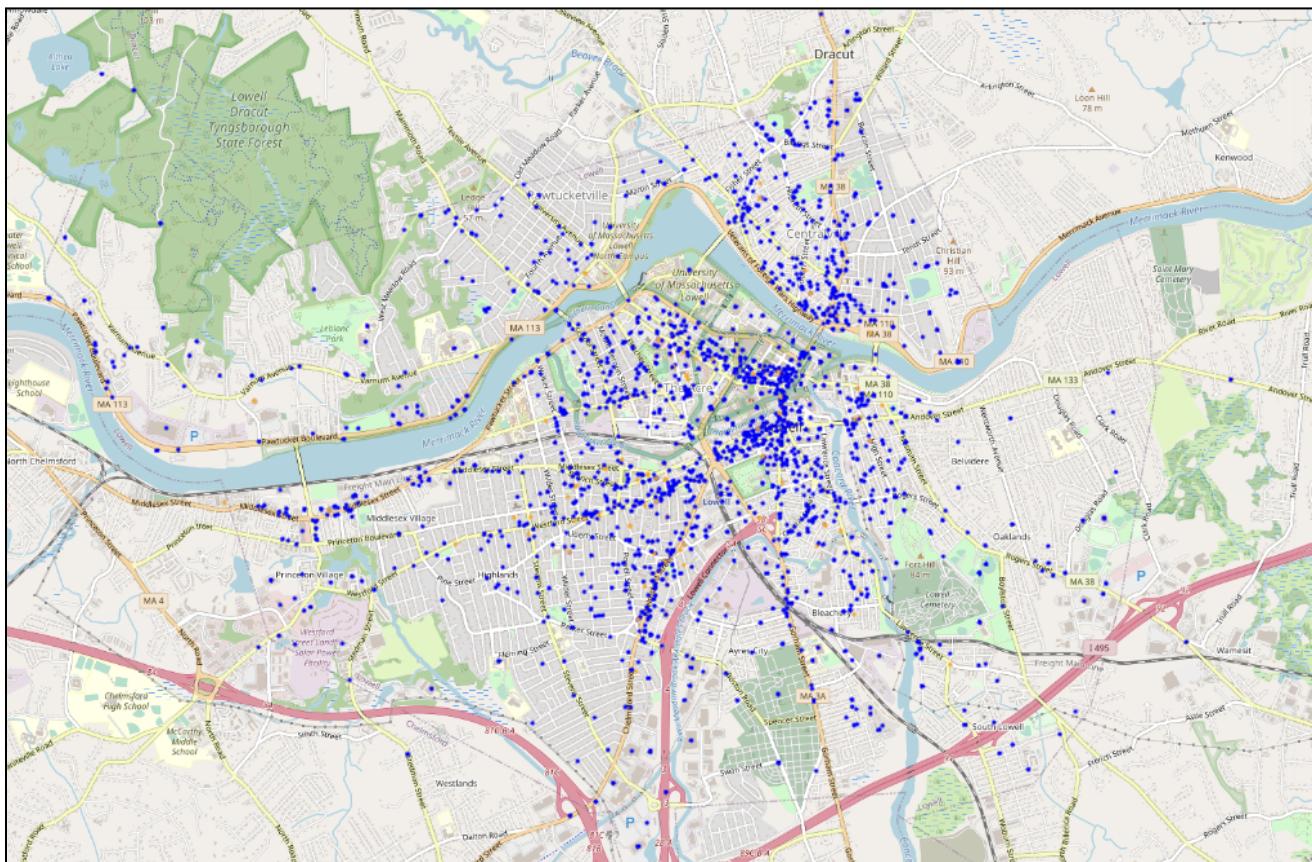
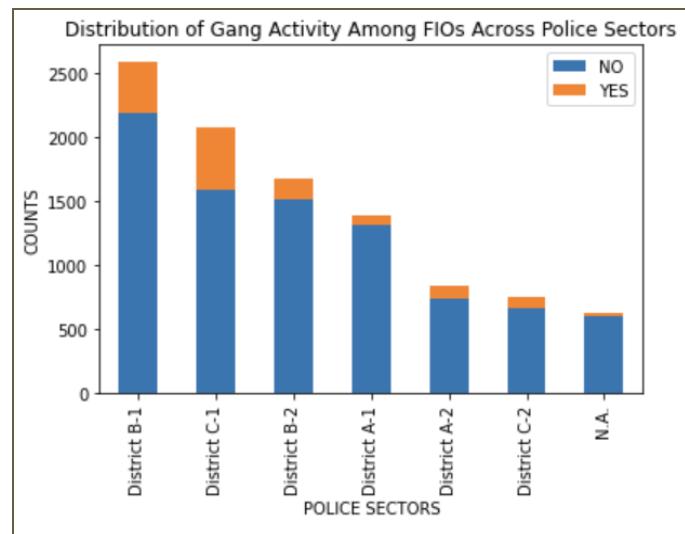
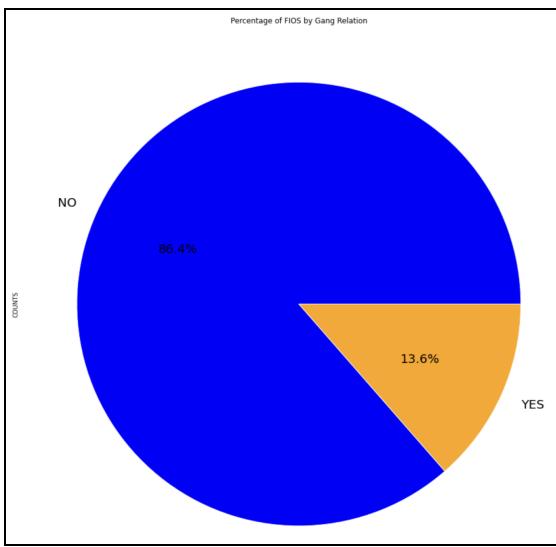
The distribution of ages in the individuals who were stopped was interesting because some of the values were illogical. More than 3000 individuals who were stopped aged 0 to 10 years old, making up the second largest age group in FIOs for Lowell. Clearly this is attributed to a data quality issue and is one of the limitations of this study.



The majority race of individuals recorded in the FIOs in Lowell was Hispanic (38.8%). The second largest group of individuals that were stopped were White (32.1%), followed by individuals who were Asian (15.0%), Black (13.3%), of an unknown race (0.5%) and Native American (0.3%). These proportions are reflected across the neighborhoods across Lowell, as Hispanic and white individuals made up the majority of FIOs that were collected while those of an unknown race and Native American made up the minority in each neighborhood.



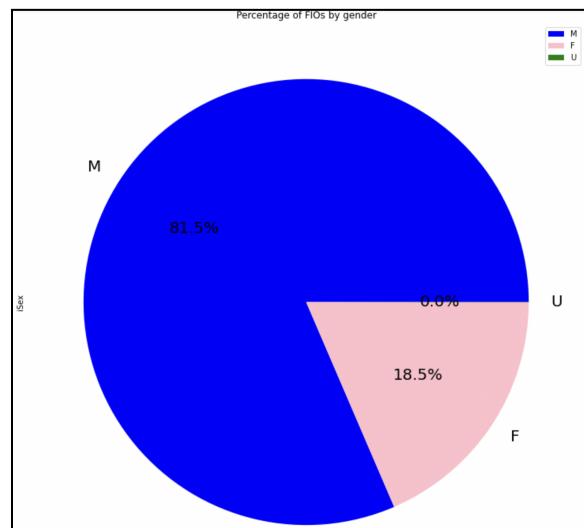
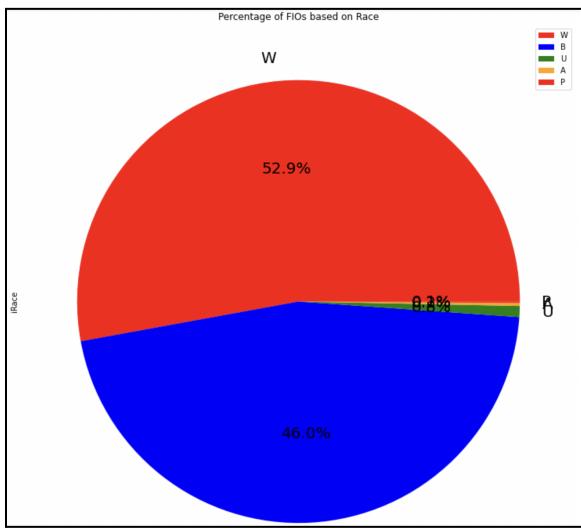
In terms of gender, the FIOs collected in Lowell were highly imbalanced between men and women. 82.1% of all individuals who were stopped were identified as men whereas 17.9% of them were women. There was a negligible percentage of individuals whose gender was recorded as unknown.



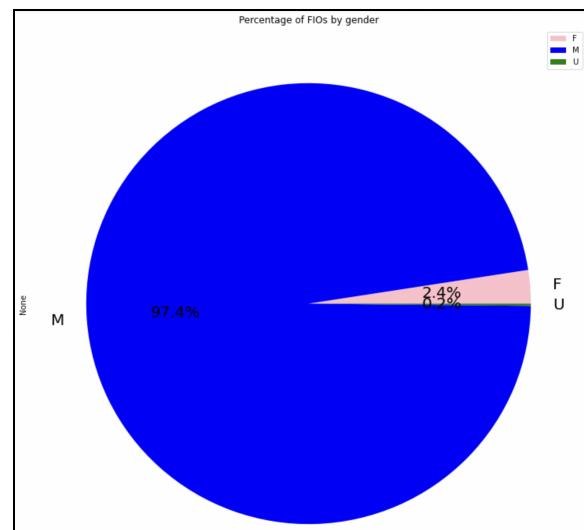
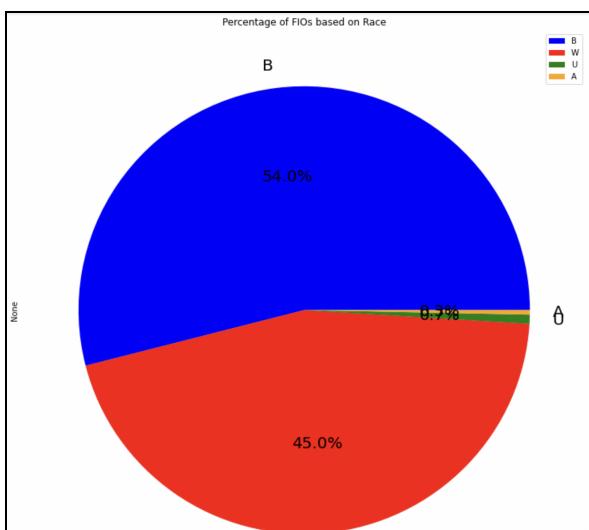
The majority of individuals who were recorded in the FIOs were not identified as being gang-affiliated. 86.4% of individuals present in the FIOs were not identified as gang members while 13.6% were identified as such. However, in comparing the distribution of FIOs and the distribution of individuals who were recorded as being gang-affiliated across the different police sectors, it appears that there is a relationship between sectors with high counts of FIOs and reported gang members. District B-1, District C-2, and District B-2 had the highest count of FIOs across all police sectors in Lowell and they also had the highest counts of reported gang members.

New Bedford

In New Bedford, there are two reports, and they are lists of FIRs and gangs. The two New Bedford forms have many similarities. It has date, race and gender information for each FIO. Therefore, we were able to compare the two New Bedford tables. The results are shown below.



I discovered that there are more white people than black people in the FIR. Males make up the majority of each gender, although there are also some female. Blacks were 46% of the cases, while whites were 52% of the cases. Men were 77% of the cases, while women were 21% of the cases.

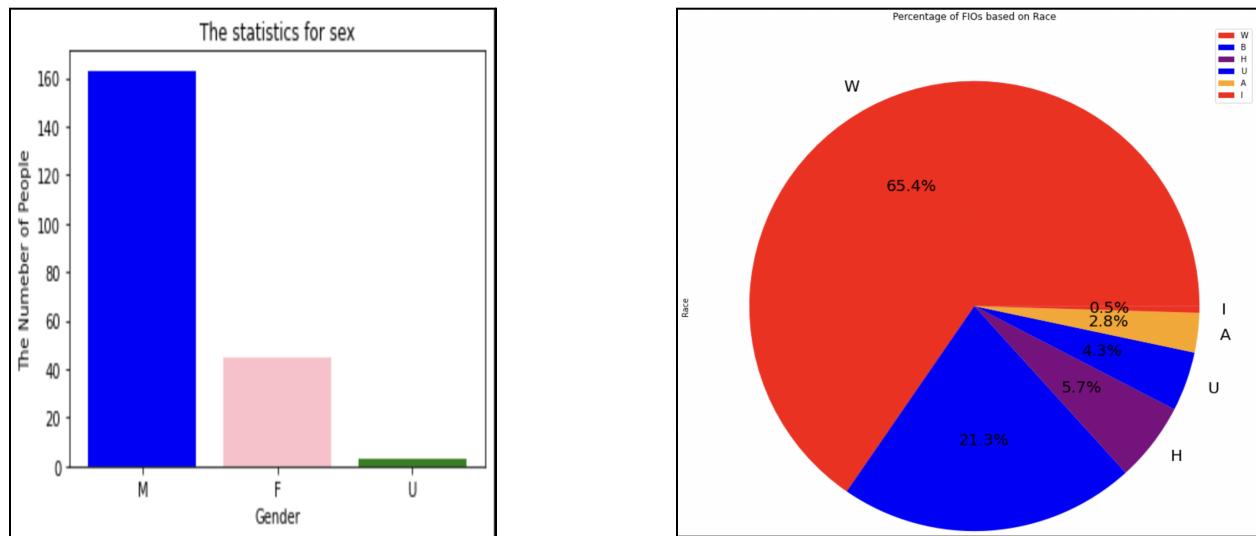


More blacks than whites are found in the gang data, but men predominate, and women are few. Blacks were 54% of the cases, while whites were 45% of the cases. Men were 97% of the cases, while women were 2%.

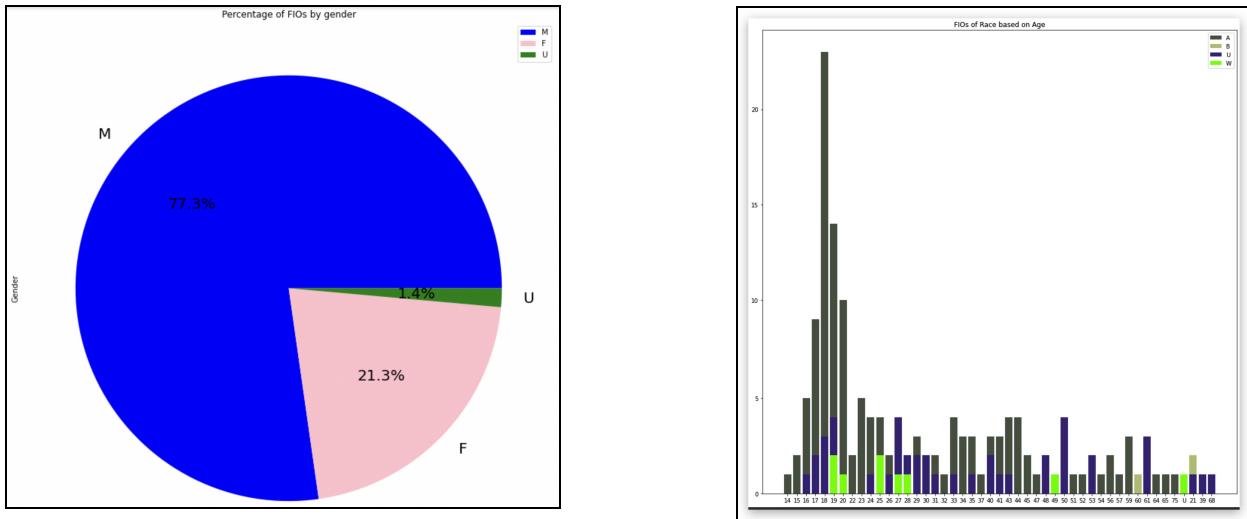
Newton

Data description: The following pertinent information for FIOs is included in the data that was given to us for Newton: FioNum, Age, and Race. Based on these observations, we can use this data to infer patterns and trends in a variety of FIO data fields.

Results: By looking at the racial distribution of these cases, we get six major racial breakdowns. Asian, Black, Unknown, White, and Native Hawaiian. We can find that proportionally Whites make up 65% of the cases, Blacks make up 21% of the cases, Hispanics make up 5.7% of the cases, and Unknowns make up 4.3%



The ratio of gender from the other graph is much higher for men than for women. The percentage of males is at 77% and females are only 21%. This is about three times as much as women.

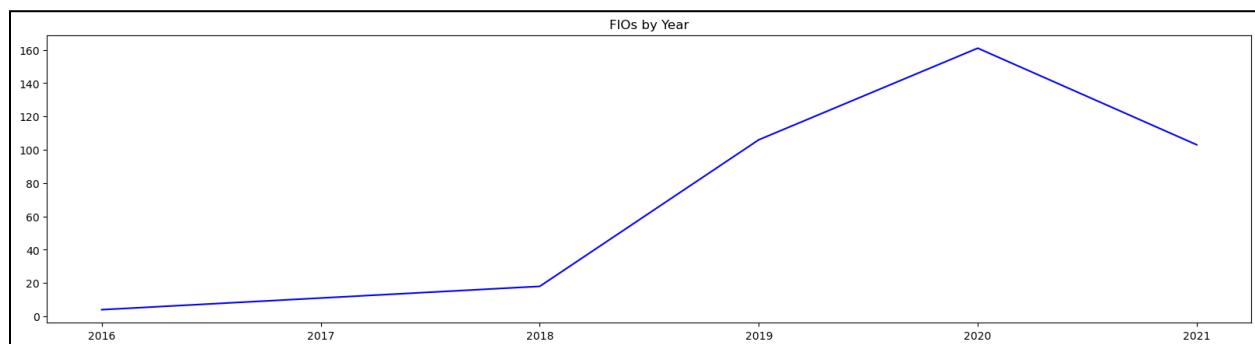


We also did a comparison in terms of age, and in this case, we found that the age around 18 was the highest number, and the percentage of whites in the age around 18 was also much higher than other races.

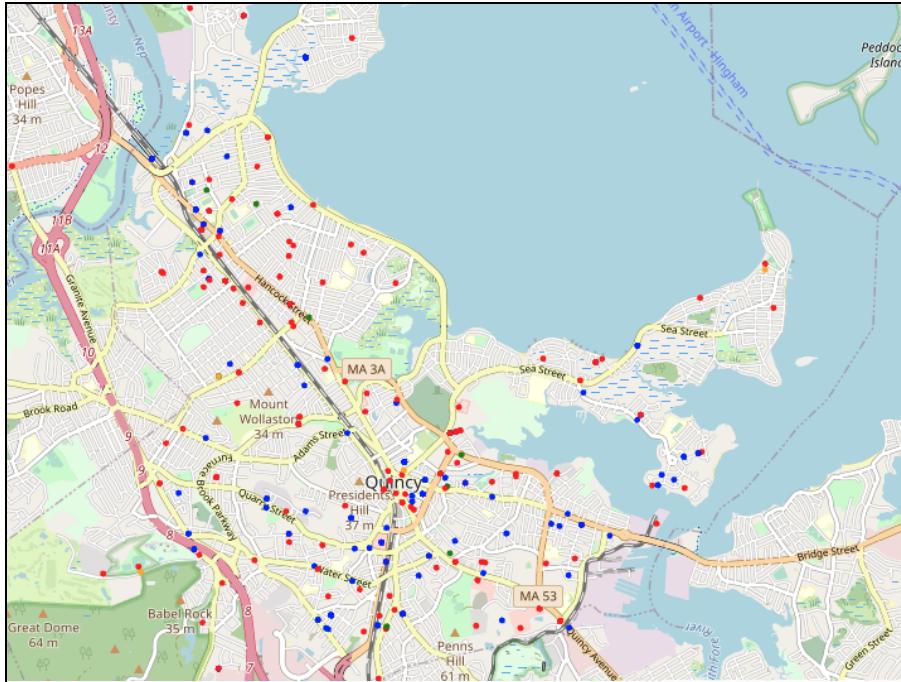
Quincy

Data description: The Quincy data set contained a number of columns relevant to analysis of FIOs. This includes Sex, Age, Race, Ethnicity, and Location. With this information, various trends can be observed in the FIOs that have occurred in Quincy.

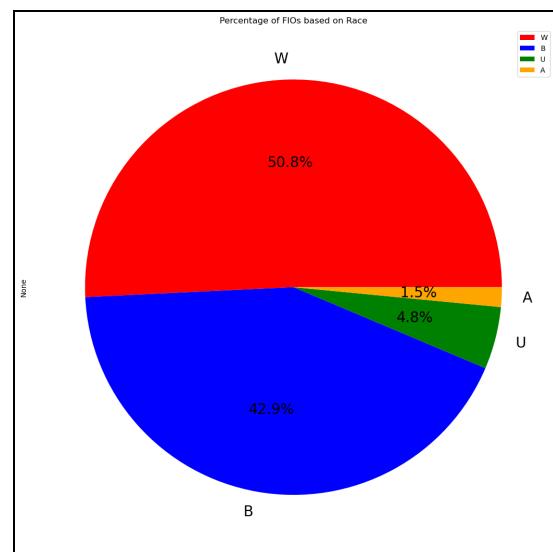
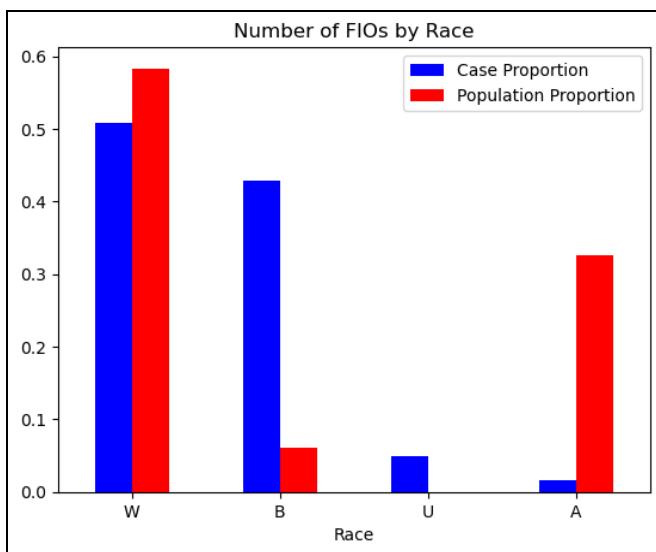
Results: In Quincy the total number of FIOs was 392 from 2016 to 2021. In the graph below we can see how the total number of FIOs has changed over the years. It can be seen that the number of reported FIOs drastically increased after 2018 and peaked in 2020. The number of FIO drops after 2020 which coincides with the Covid-19 pandemic.



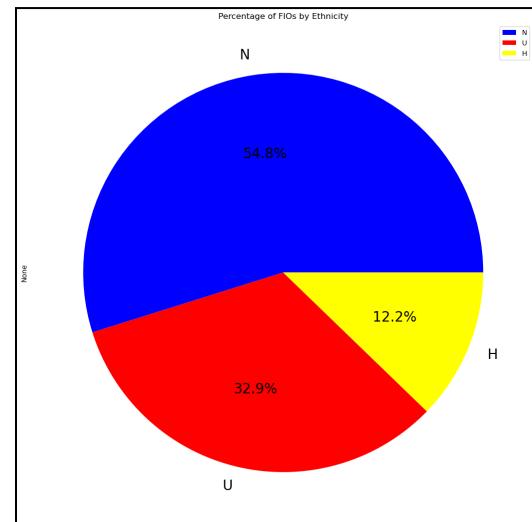
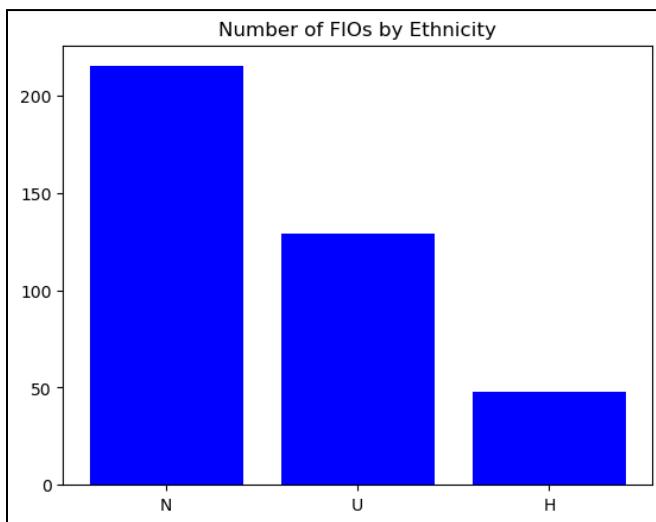
We can do analysis of the location data provided in the data set. Using the addresses of the FIOs, We can show where in Quincy the FIOs took place. Additionally, we can color code this map based on race. The colors indicate races as follows: red dots represent White, blue dots represent Black, green dots are unknown, orange dots represent Asian. From this we can observe trends such as a large quantity of Black cases are in south Quincy.



We can continue to look at the distribution of different races across Quincy's FIOs. We can refine the data to include 4 main race classifications: White, Black, Asian, and Unknown. We can visualize the proportion of the total data each race takes up in the graphs below. Using the census data from Quincy, we can also compare how the distribution of race across FIOs compare the distribution of races in Quincy. According to Census Data, the population is around 58% White, 6% Black, and 32% Asian. In contrast, the reported FIOs are distributed as follows: 51% White, 43% Black, and 1.5% Asian. The biggest discrepancy is that the reported FIOs that are Black is a far greater proportion than the Black population in Quincy. Additionally, the reported Asian FIO's is far lower than the population of Asians in Quincy.



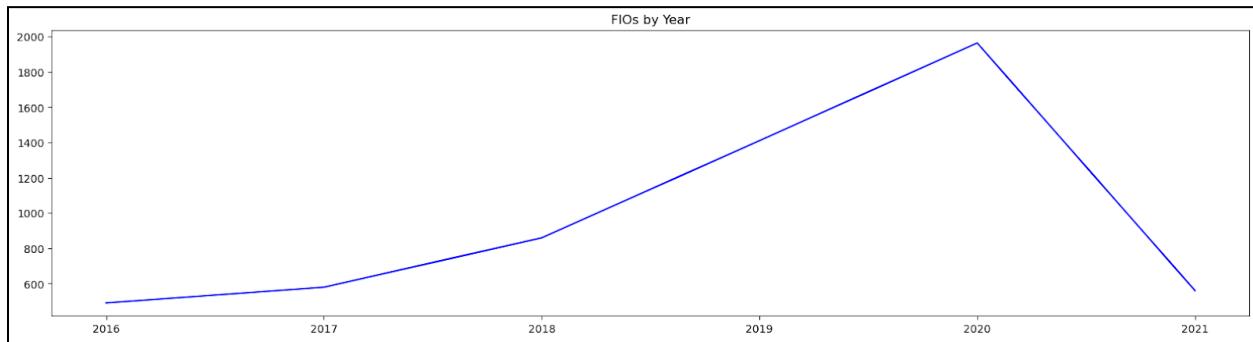
We can observe ethnicity in a similar way as we looked at race. Unfortunately a large portion of the ethnicities are unknown for the Quincy data set.



State Police

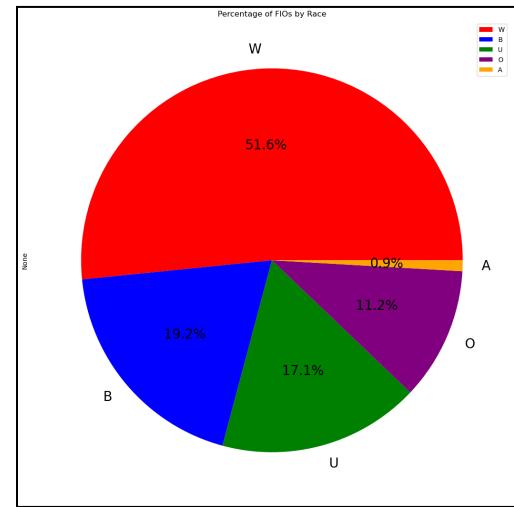
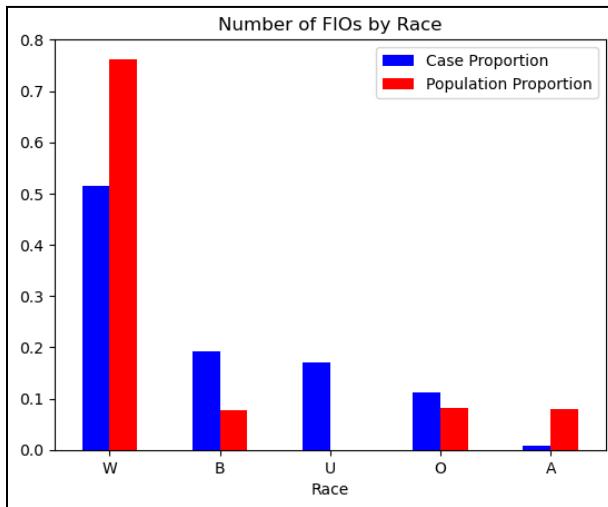
Data description: The State Police data set contained a number of columns relevant to analysis of FIOs. This includes Sex, Age, Race, Ethnicity, and Subclass. With this information, various trends can be observed in the FIOs that were reported by the State Police.

Results: In State Police data, the total number of FIOs was 392 from 2016 to 2021. In the graph below we can see how the total number of FIOs has changed over the years. It can be seen that the number of reported FIOs drastically increased after 2018 and peaked in 2020. The number of FIO drops after 2020 which coincides with the Covid-19 pandemic.

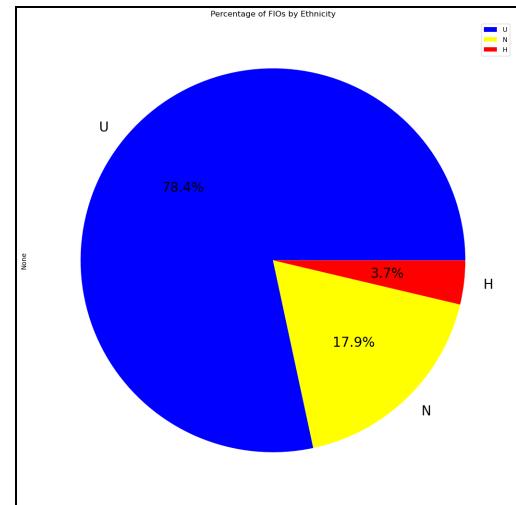
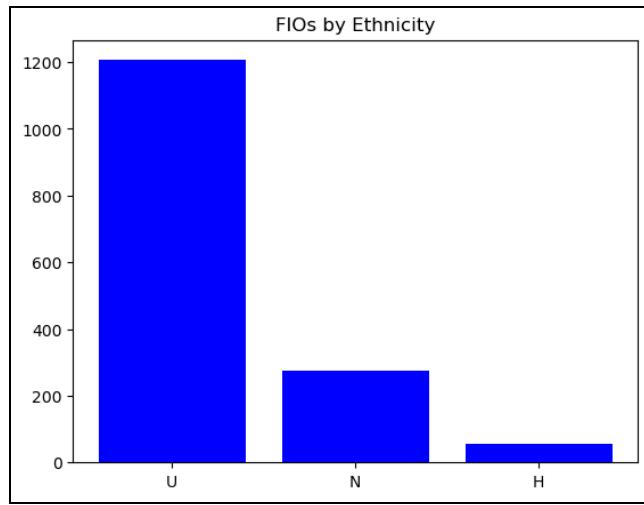


We can look at the distribution of different races across State Police's reported FIOs. We can refine the data to include 4 main race classifications: White, Black, Asian, and Unknown. Additionally we can provide an Other classification as there are a lot of races with a low quantity of reports in this data set. We can visualize the proportion of the total data each race takes up in the graphs below. Using the census data from Massachusetts, we can also compare how the distribution of race across FIOs compare the distribution of races in the State Police data. According to Census Data, the population is around 76% White, 7.7% Black, 8.1% other, and 7.9% Asian. In contrast, the reported FIOs are distributed as follows: 52% White, 19% Black, 11% other, and 1% Asian. There is a large discrepancy between the census and reported data for each of the White, Black, and Asian races. The White and Asian cases are distributed

significantly less than in the census data, while the Black cases are more than double the Black population proportion.

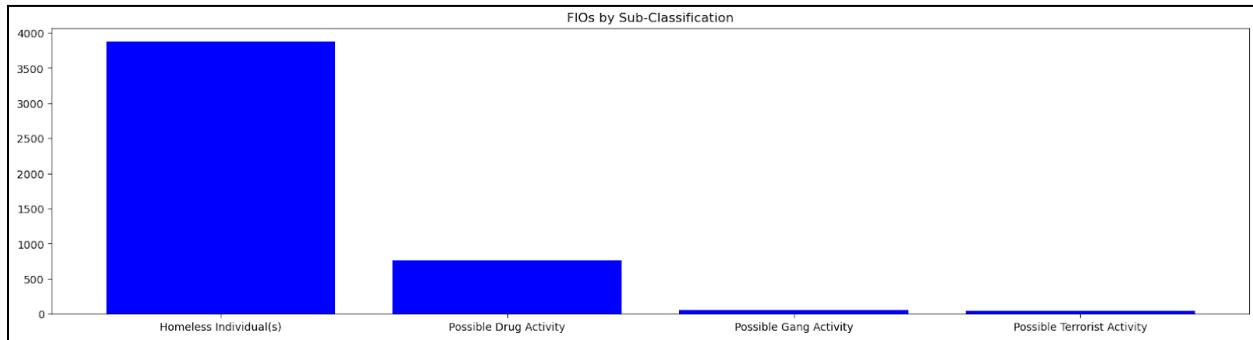


We can observe ethnicity in a similar way as we looked at race. Unfortunately a large portion of the ethnicities are unknown for the State Police data set.



An interesting portion of the State Police data set is the subclass column, This Further divides the FIOs into 4 categories regarding the context of the FIO case: Homeless Individuals, Possible Drug Activity, Possible Gang Activity, and Possible Terrorist Activity. The vast majority of the cases were homeless individuals and there was not a significant amount of gang or

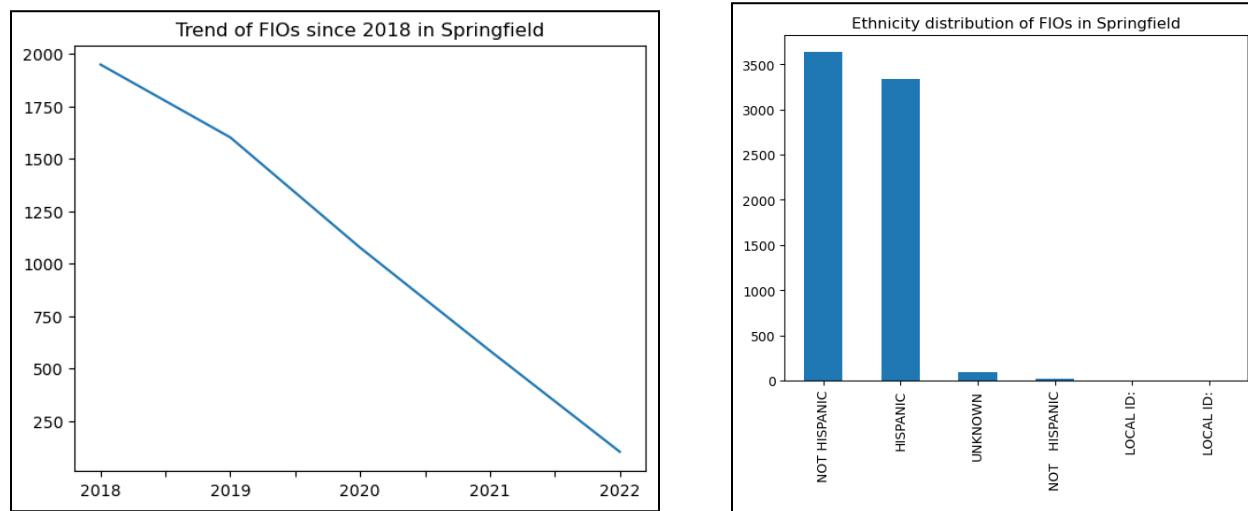
terrorist related cases. There was however a large portion of drug related cases although not to the same extent as homeless individuals.



Springfield

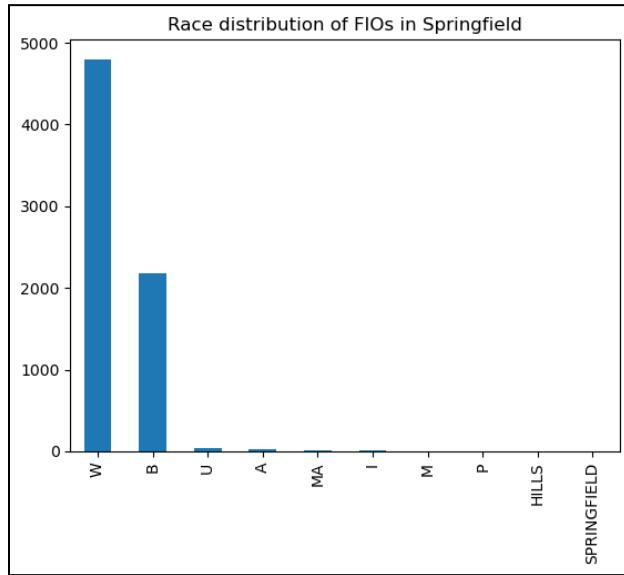
Data description: Springfield data has 3 date fields: ‘entered’, ‘modified’ and ‘approved’.

Unfortunately, around 1600 out of the 6599 data points had missing date values. We produced a plot for the portion of the dataset that did not have missing date values. Thus, the resulting plot is not a perfect representation of the trend of FIOs. The plot is shown below, starting from 2018. Interestingly, the number of FIOs seems to be steadily decreasing since 2018.

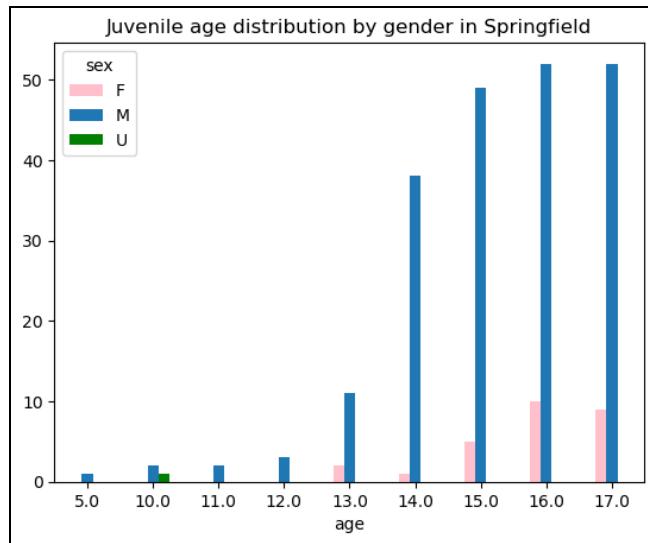


Springfield's race and ethnicity values have some irregularities. Below, we have plotted the distribution of race and ethnicity values separately. As we can see, there are a number of weird or redundant values, like "Local ID" for ethnicity and 'MA', 'I', 'HILLS' for race. Because of those, we couldn't concatenate the two fields like we did for Brockton.

Results: The distributions tell us that the most reported ethnicity is non-Hispanic, but the number for Hispanic is a close second. The two most reported races are White (W) and Black (B). There is a tiny fraction of FIOs with unknown ethnicities or unknown races.

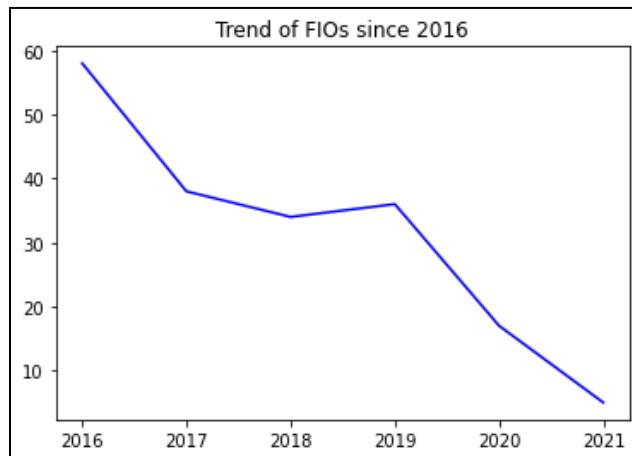


Just like Brockton, we were able to produce a distribution of FIOs involving juveniles by gender, which is included below. The shape is very similar to that of Brockton, however the number of FIOs involving 16 and 17-year old males is nearly the same. Also, the number for 15-year old male is much higher in Springfield than in Brockton. The numbers for females are nearly the same as that of Brockton.



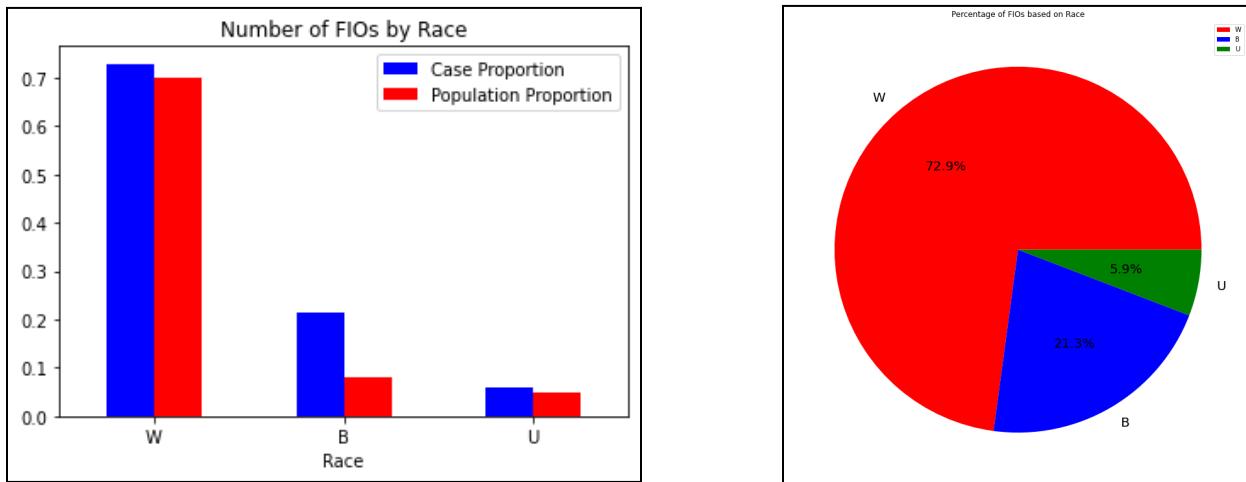
Waltham

Data description: The data provided to us for Framingham includes the following relevant data for FIOs: FioNum, date, and race.. With this data, we are able to draw conclusions on various fields of FIO data and plot frequencies and trends based on these observations.



Results: Let's start by looking at the total distribution of cases within Framingham. In total, there were 188 cases in Everett from 2016 until 2021. Of these cases, 58 were in 2016, 38 were in 2017, 34 were in 2018, 36 were in 2019, 17 were in 2020, and 5 were in 2021. We notice that

there is an overall decline in the number of cases per year. It is worth noting that 2020 and 2021 have skewed data as the Covid-19 pandemic severely impacted the number of cases occurring.



When taking a look at the Race distributions for these cases, we are provided with 3 main classifications of race: Black, Unknown, and White. Of these race classifications, 44 are Black, 11 are Unknown, and 137 are White. When looking at the graph of this data, it is quite apparent that the outlier is White. There exists a large discrepancy between white cases and the rest of the races. Thus, it is important to look at the census data to see if the trends match the population distributions. According to the US 2020 census data, there are 44,130 white individuals, 4,752 black individuals and 3,292 unknown individuals. Taking a look at the proportions we have 73% of the cases being committed by 70% of the population, 21% of the cases being committed by 8% of the population, and 6% of the cases being committed by 5% of the population. The only discrepancy which exists in this example is in the observance of black FIO cases.

Comparative Analysis

Town	Population	Number of FIOs	FIOs per 10k	Most frequent race (if applicable)	Time Period
Boston	654,776	4,490	69	N/A	2021
Brockton	105,446	1,062	101	Black	2016 - 2021
Chelsea	39,878	1,052	263	White	2016 - 2021
Everett	46,275	1,391	301	White	2016 - 2021
Framingham	71,260	2,876	404	White	2016 - 2021
Lowell	113,994	10,000	11	Hispanic	2007 - 2021
New Bedford	100,941	4,997	49	White	2015 - 2020
Newton	87,453	211	2	White	2015 - 2020
Quincy	101,636	392	39	White	2016 - 2021
State Police	7,029,917	5,863	8	White	2016 - 2021
Springfield	154,789	7,305	472	White	2018 - Apr 2022
Waltham	62,597	188	30	White	2016-2021

Relative Rate Index

The right column for each town is the RRI. The left is FIO proportion/race proportion.

	Brockton		Chelsea		Everett		Framingham		Lowell **	
White	1.05	1	2.02	1	.574	1	1.37	1	1.00	0.47
Black	1.75	1.67	2.67	1.32	1.06	1.84	3.14	2.29	1.27	1.00
Asian	0.005	0.0047	.002	.0009	.125	.2175	.1	.073	0.74	0.58

	New Bedford		Newton		Quincy		State Police		Springfield		Waltham	
White	0.85	1	0.814	1	0.87 1	1	0.67 7	1	1.15	1	1.04	1
Black	6.88	8.12	6.09	7.47	7.10 5	8.15 7	2.49 5	3.68 6	1.48	1.28	2.63	2.53
Asian	0.06	0.07	0.215	0.26	4.60 3	5.28 5	1.13 3	1.69 8	0.11	0.095	N/A	N/A

*Boston dataset did not contain race data

**Average from 2020 to 2021, race population data was unavailable for the remaining years

Limitations

The data quality and inconsistency across datasets was a major concern. FIOs for two municipalities were presented in a scanned PDF that had an irregular formatting style, making it impossible to scrape data from them. Additionally, most of the datasets did not share enough fields needed to perform proper comparative analysis across all the municipalities. For example, in the case of Lynn, information on individuals was aggregated, making it difficult to derive any meaningful insights, much less comparisons. Out of the FIOS that did share fields, some did not record a field the same way, or had a lot of missing and nonsensical values. For example, while some municipalities recorded both race and ethnicity (eg. Brockton), others only recorded race (eg. Springfield), while others included the Hispanic ethnicity as a race (eg. Lowell). These irregularities prevented us from gaining an accurate evaluation of potential biases in police officers and comparing police behavior in regards to race across multiple municipalities. Additionally, some of the FIOs contained nonsensical values for age. For example, approximately 30% of the individuals recorded in the Lowell FIO were aged from 0-10 years old. Clearly, this prevented us from providing accurate insights about police behavior in terms of age and which age group was most targeted in the FIOs.

Due to these data quality issues, we strongly recommend an implementation of a standardized data collection method. Without consistency between the various municipalities, it makes it difficult for researchers to make accurate evaluations about police behavior and raises suspicion about their methods. Furthermore, there needs to be a lot more transparency about how the police determine traits of individuals such as race, age, and gang relation and whether their interaction led to an arrest or not. A lot of FIO data can be misleading since the police can simply record a sighting of a person, leaving room for error in the police's description of the individual versus the individual's actual traits. It is impossible to express that margin of error without insight into policing methods.

Conclusion

Even with the issues of data quality and inconsistency, we were still able to perform substantial individual analysis and drew conclusions for each municipality separately. In our analyses, we were able to (1) visualize trends in FIOs over time, (2) map out FIO locations, (3) extract FIO proportions based on race and compare to the demographic proportions and (4) plot frequency charts for age (for datasets that contain age data).

In every municipality with the exception of Quincy, the number of FIOs per year followed a decreasing trend. We observed a rapid decrease in most of the FIOs in 2021 which coincided with the COVID-19 pandemic. We also saw that within each of the municipalities, FIOs were concentrated in a specific area. For future analyses, it would be helpful to get data regarding the boundaries of police sectors with the number of police officers assigned to them to understand these spatial patterns further.

The most frequent race observed across all of the FIOs was white people followed by black people. However, after calculating the RRIs, we found that black people are disproportionately recorded in the FIOs compared to white people as the RRIs for black people were greater than the RRIs for white people across all municipalities. This suggests racial bias amongst the police and needs to be further studied.

The vast majority of FIOs involved male juveniles aged 16-18 years old or in their early 20s. However, we do not claim that this is an accurate representation of the actual ages and gender of individuals in the FIOs because some of the municipalities such as Lowell and Framingham had nonsensical values of age. Furthermore, for the gender field, some municipalities had “Unknown” as a value. It is unclear whether this is because the individual recorded identified as nonbinary or the police officer who recorded the sighting had trouble determining their gender. This speaks to both data quality issues and the lack of transparency of

methods in FIO data. In order for more accurate race and gender analyses in the future, we strongly recommend more data quality control and documentation of methodology.