Sam Malcolm Kirti Pande Evan David Maggie Yu

Homicides Report

Description of Project Goals

The primary goal of this project is to investigate homicides in major cities in the United States over the last decade. Specifically, we want to know if we can predict where the Federal Government should allocate resources for homicide prevention and investigation.

The dataset is very interesting because we are able to examine the frequency of homicides in these cities, demographics about the victims, and the status of the disposition i.e. whether the case was closed with an arrest or is still open with no arrest made. This allows us to see relationships between these factors and gain important insights that may aid in preventing future homicides.

If we can decrease the number of homicides by allocating more law enforcement resources in cities most in need, we can benefit society and save lives. By allocating more investigative resources in places where homicide cases are still open, we can solve more of these cases and catch the murderers, thus benefiting the community and leading to a safer society.

Exploratory Analysis

For our analysis we obtained data from the Washington Post which they claim is more specific than the data released annually by the FBI. In addition, we used population data from the US census and state crime rate statistics from the FBI.

The data contains information about 52,179 homicides in 50 American cities since 2007 and includes 14 variables per homicide including:

- Reported date
- Victim first and last name
- Victim age, race, and sex
- City and state
- Latitude and longitude (where homicide occurred)
- State-wise population for every year
- City-wise population from 2010
- Crime rate of every state
- Disposition, which consisted of three options:
 - "Open/No Arrest"
 - "Closed by arrest"

 "Closed without arrest" - According to the Washington Post, "these are cases in which there is sufficient evidence but an arrest is not possible, for example, if the suspect has died."

We first took a broad look at the total number of homicides over time, in what areas they occurred most frequently, and the proportion which were classified as each disposition. From there, we attempted to discern patterns in age, race, and sex demographics.

We observed a steady rise in homicides committed annually¹ with 2016 containing the greatest number², approximately 1,000 more than the years preceding and following it. Further research shows that there wasn't anything particularly unique that caused this. Rather, there was simply an astonishing increase in already high-crime metropolises such as Chicago and Baltimore.

Homicides occurred least frequently³ in February, in part we're assuming because it has the fewest days of any month. Homicides then begin to accumulate at an increasing rate until July where they are most frequent. From August until December, the rate decreases steadily and slightly.

We continued by plotting⁴ the amount of homicides per region⁵ and state⁶. We observed that California and Texas were the states with the greatest number of homicides. However, we found that Missouri had the highest homicide rates after adjusting for population differences. We thought this was unique, but Missouri also had the highest crime rate adjusted by population as well. We decided not to pursue this information further. First, the data represents just 50 cities, so there is the possibility that certain larger areas could be over or underrepresented in the data. Second, we believe that greater impact can be made at more local levels. We will explore this more in the Solutions & Insights portion of the report.

In terms of disposition⁷, 44.4% remained "Open/No arrest," 47.8% were "Closed by Arrest", and 7.8% were "Closed without arrest." We weren't sure whether or not to be surprised by the amount of homicide cases that remained open, but we were certainly disappointed, and it reevaluated the importance of our goal.

When examining demographic information, a clear pattern emerged. Young, African American males are the most likely by far to be homicide victims. 22.7% of victims in the data were African American males in their 20s. That number increases to 32.8% if you also include African American males in their teens.

Overall, 63.94% of homicide victims were African American⁸, 13.23% were Hispanic, 12.14% were White, 8.05% were of Unknown race, and other races were less than 2% each.⁹

78% of victims were $Male^{10}$, 13.82% were Female, and 8.11% of unknown sex. 11

Finally, 34.5% of victims were in their 20s, 18.9% were in their 30s, 18.1% were between the ages of 40 and 60, and 15.6% were teens. 12

Solution and Insights

When evaluating how to allocate federal resources, we wanted to consider two different metrics. First, where are homicides occurring most frequently? It is in these areas that we hope to make an effect via prevention. Second, where are homicide cases not being solved? Here is where we can make a significant improvement in investigation and closure. To gauge both of these metrics, we will focus on cities. As previously stated, the data isn't completely reliable for states and regions, and we believe the greatest effect can be made at a micro level.

To evaluate frequency, we simply looked at where the most homicides have been committed. There were six cities with greater than 2,200 homicides reported (with no other city above 1,600):¹³

- Chicago 5535
- Philadelphia 3037
- Houston 2942
- Baltimore 2827
- Detroit 2519
- Los Angeles 2257

When adjusting for population, the order changes slightly, but it is consistent with frequency in general. The cities with the highest homicide rate were:

- Kansas Citv
- Baltimore
- Detroit
- Philadelphia
- Houston
- Chicago
- Los Angeles

To evaluate effectiveness, we judged cities based on a measure we refer to as "Closure Rate." Closure Rate is the proportion of a city's homicides that were closed (whether by arrest or without arrest). There were 14 cities with a Closure Rate of 0.5 or below 15 (with no other city below 0.53 but 9 cities above 0.70). Of those 14, we found 4 that also had a considerable quantity of homicides: 17

- Chicago 0.33
- Baltimore 0.41
- Detroit 0.41
- Los Angeles 0.50

These four cities alone account for 25% of the total homicides¹⁸ and 34% of the total open cases in the data.¹⁹ It makes sense that there would be some overlap, especially at the extreme end of the frequency list. If a city has an exceptional number of homicide victims, they will have fewer resources (money, time, manpower) per victim to solve the case. Rather than being an issue of collinearity that would discount either measure, though, the relationship can almost be viewed as a multiplier and as further evidence that these are the areas deserving federal assistance.

When examining the demographics of each city, they are generally in line with the demographics of the data as a whole, but there are a couple exceptions. Victims in Detroit²⁰ and Los Angeles²¹ in their 40s - 60s are represented slightly more frequently than average. Also, the proportion of Hispanic victims is much higher in Los Angeles, most likely because the Hispanic population is proportionately much greater in Los Angeles (roughly 47%) than other cities and the nation as a whole.²² However, when considering that African Americans make up only about 10% of the total population in Los Angeles, that group is still being killed at the highest rate, consistent with the other cities we examined.

After narrowing down the cities, we wanted to predict the number of homicides for the year 2018 so that we could advise the federal government accordingly. We used a time series forecasting model called ARIMA(Autoregressive Integrated Moving Average). We can anticipate the following trends in each city:

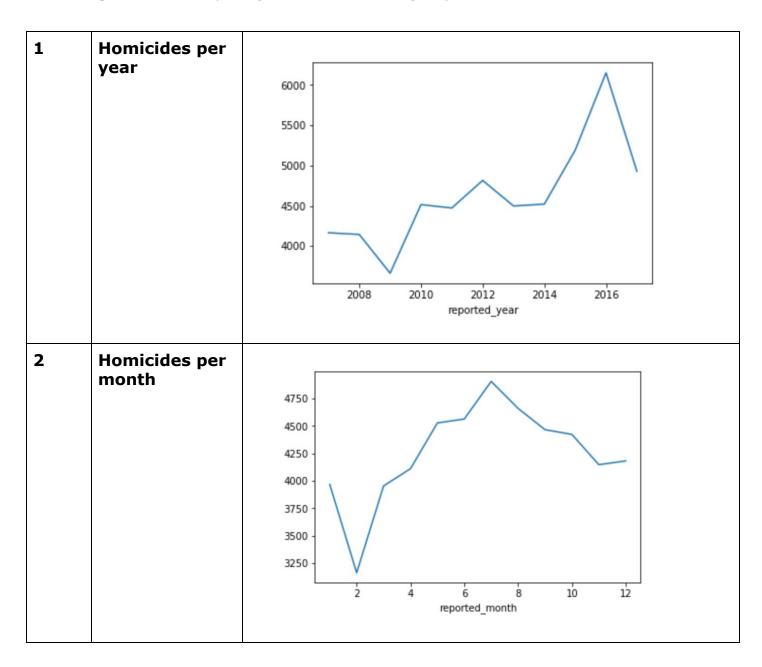
- Chicago²³ Increased homicides, particularly in the last quarter
- Detroit²⁴ Increased homicides, again rising in the last quarter
- Los Angeles²⁵ Very consistent with 2017
- Baltimore²⁶ Steadier rate of homicides per month (no dramatic spikes and dips like 2016)

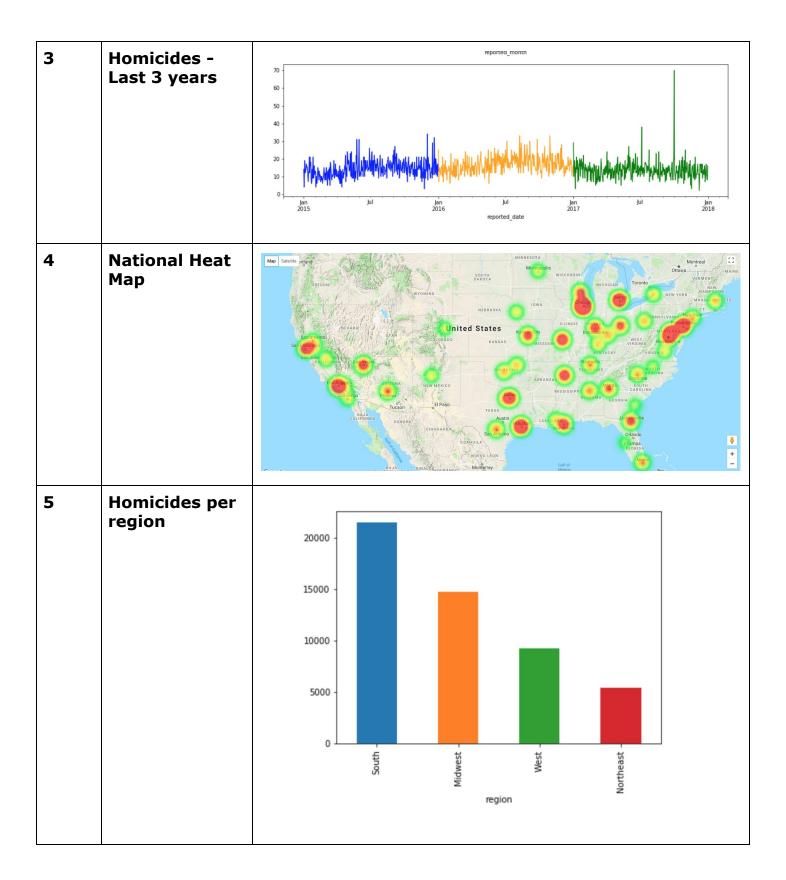
As a whole this analysis supports what already might have been expected. The areas with the greatest number of homicides and the lowest rate of cases being closed are those that are predominantly low-income and African-American. With this in mind, we would make the following suggestions:

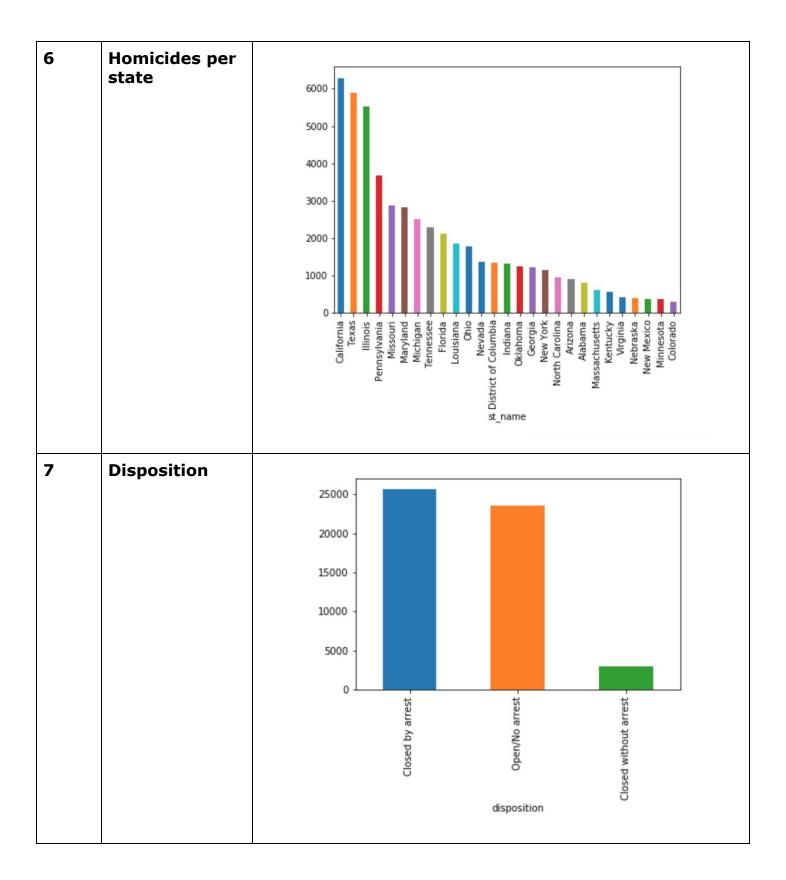
- 1. Expedite distribution to help prevent the anticipated rise in homicides in the next 2 months (and final quarter) of 2018
- 2. Prioritize allocation of resources meant to to mitigate gang violence
- 3. Funnel resources and advanced forces / manpower to specific neighborhoods with the worst economic conditions in each city

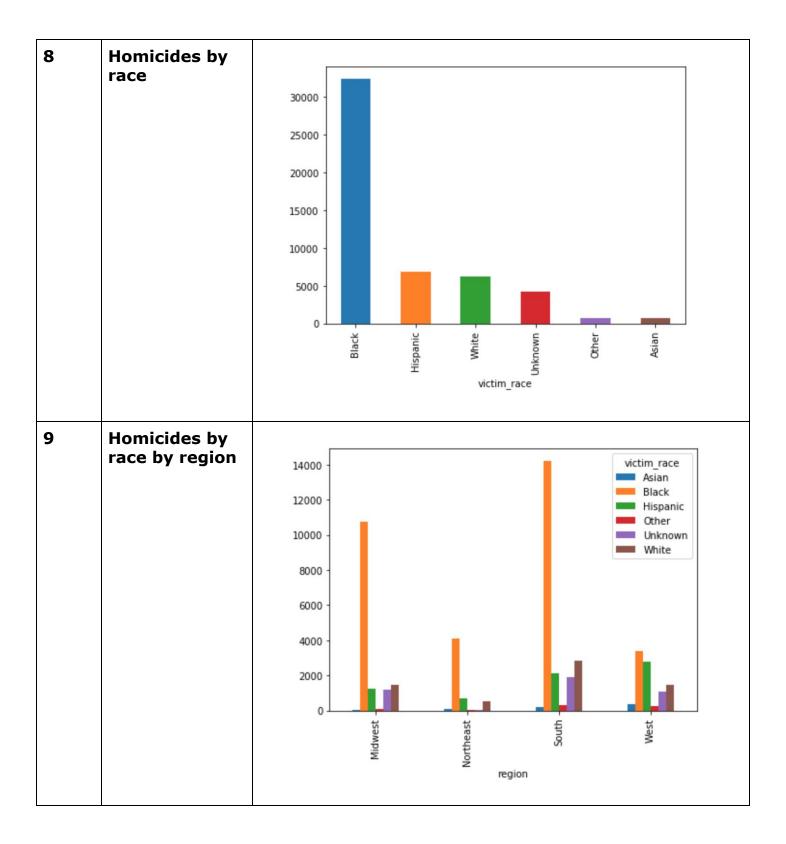
Appendix

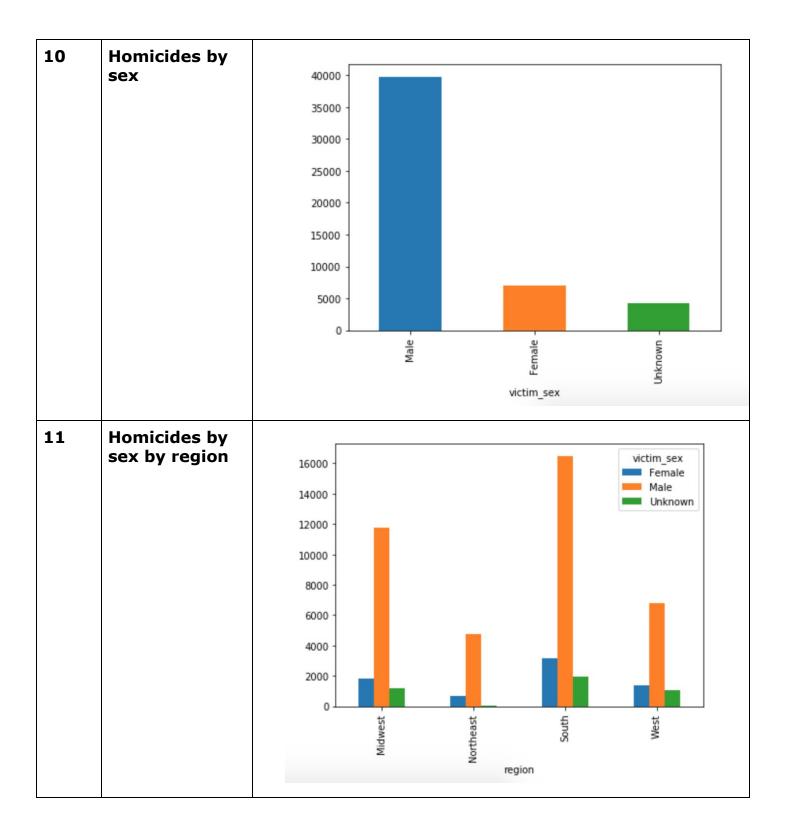
Original Data: https://github.com/washingtonpost/data-homicides

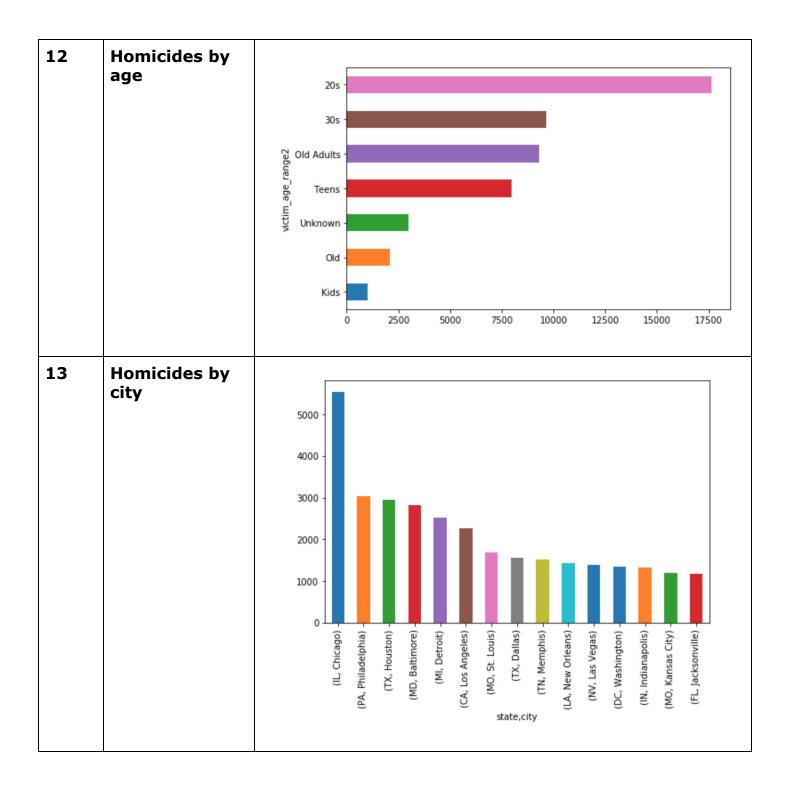






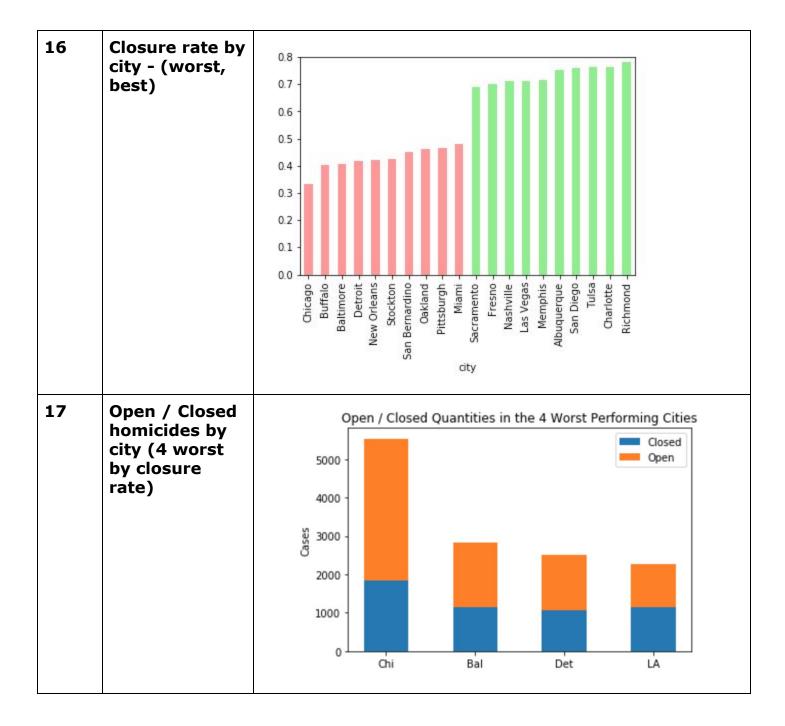


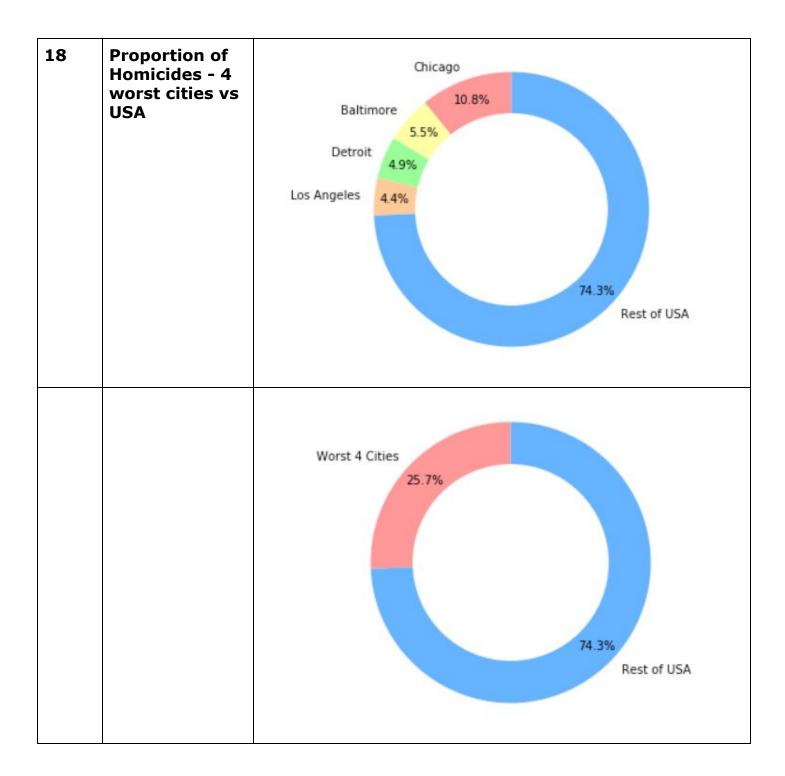




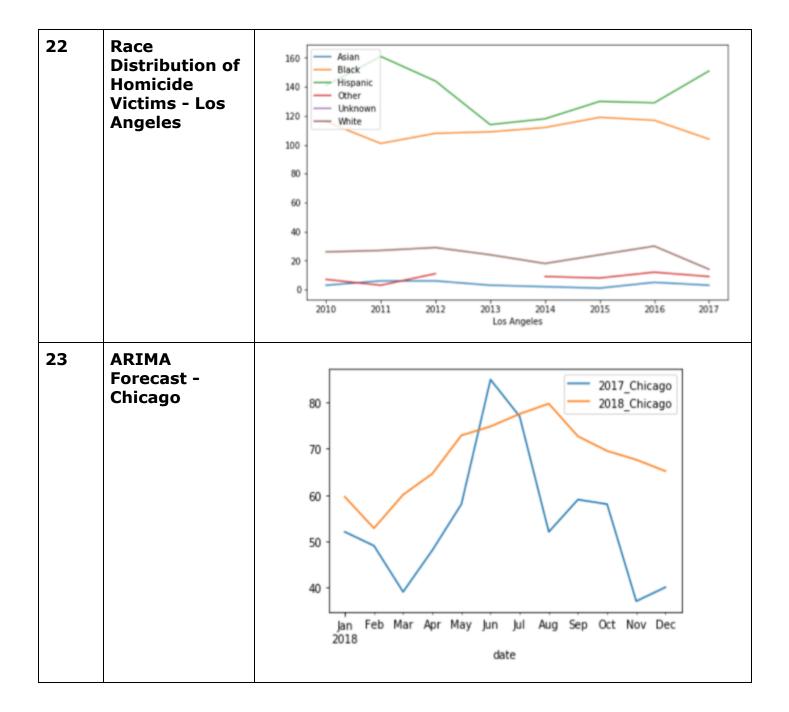
		Chicago 5535	
		Philadelphia 3037 Houston 2942	
		Baltimore 2827	
		Detroit 2519	
		Los Angeles 2257	
		St. Louis 1677	
		Dallas 1567	
		Memphis 1514	
		New Orleans 1434	
		Las Vegas 1381	
		Washington 1345	
		Indianapolis 1322	
		Kansas City 1190 Jacksonville 1168	
		Columbus 1084	
		Atlanta 973	
		(SECON ACTION AC	
14	Closure rate by		
	city (worst)	0.8	
		0.7 -	
		3.7	
		0.6 -	
		0.6 -	
		0.6 - 0.5 -	
		0.6 -	
		0.6 - 0.5 -	
		0.6 - 0.5 - 0.4 -	
		0.6 - 0.5 - 0.4 - 0.3 - 0.2 -	
		0.6 - 0.5 - 0.4 - 0.3 - 0.2 - 0.1 -	
		0.6 - 0.5 - 0.4 - 0.3 - 0.2 - 0.1 -	
		0.6 - 0.5 - 0.4 - 0.3 - 0.2 - 0.1 -	
		0.6 - 0.5 - 0.4 - 0.3 - 0.2 - 0.1 -	
		0.6 - 0.5 - 0.4 - 0.3 - 0.2 - 0.1 -	
		0.6 - 0.5 - 0.4 - 0.3 - 0.2 - 0.1 -	

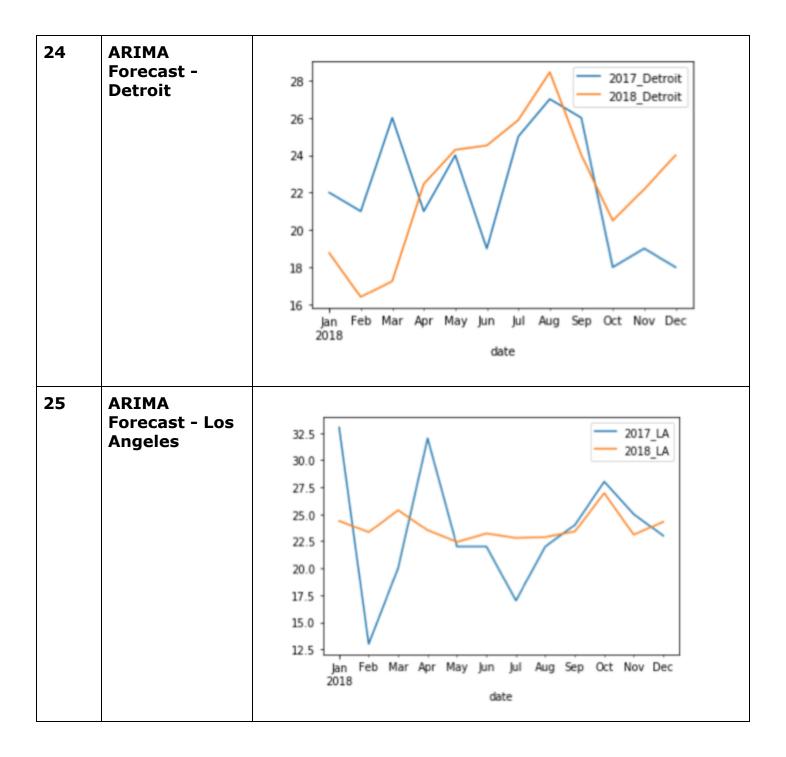
		city Chicago Buffalo Baltimore Detroit New Orleans Stockton San Bernardino Oakland Pittsburgh Miami St. Louis San Francisco Boston Los Angeles	0.334056 0.403071 0.408207 0.418023 0.419805 0.425676 0.450909 0.463569 0.465927 0.479839 0.479839 0.484198 0.494721 0.495114 0.509969	
15	Open / Closed homicides by city (14 worst by closure rate)	Open / Closed Quantities in the 14 Worst Performing Cities Closed Open		
		2000 - 1000 - Ohi Bal Det LA	STL NO Oak Mia SF PBG Bos Buf StoSan B	

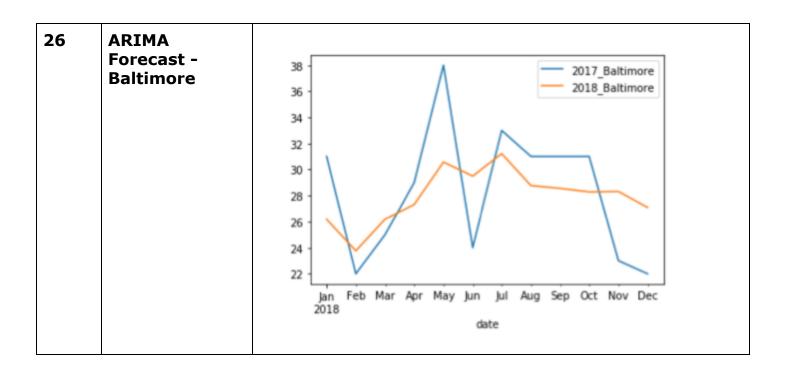






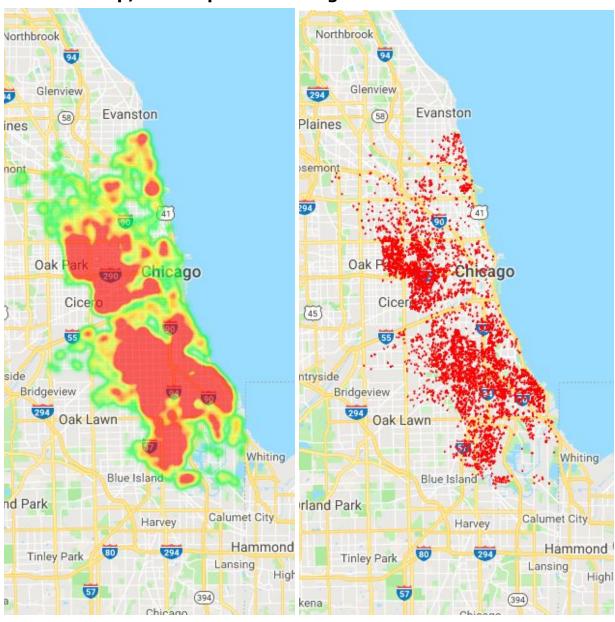




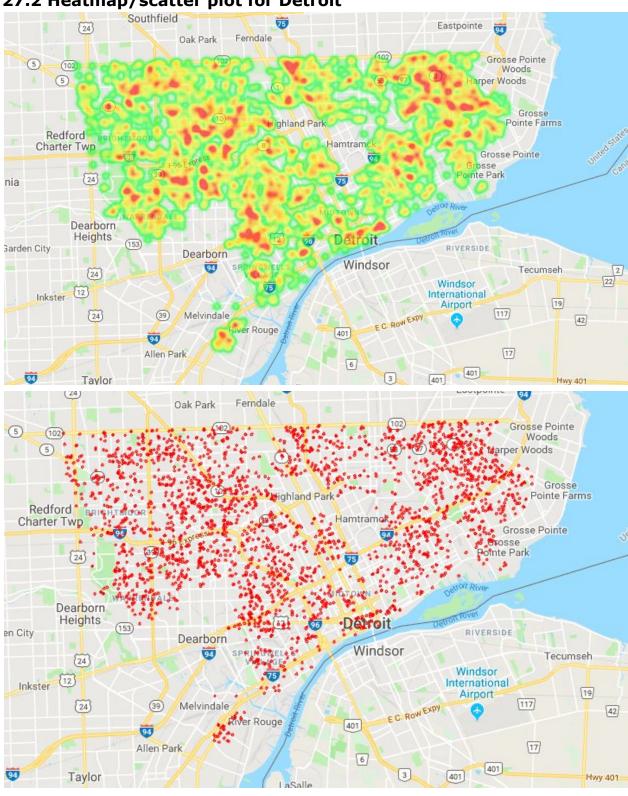


Heat Maps for each city (4 worst)

27.1 Heatmap/scatter plot for Chicago



27.2 Heatmap/scatter plot for Detroit



27.3 Heatmap/scatter plot for Baltimore White Mars Parkville Pikesville Nottingha Istown Overlea 95 LAND PARK Rossville Milford Mill Lochearn Midd HAMPDEN WINDSOR MILL (150 Woodlawn (702) Rosedale Essex tapsco (150) /alley Baltimore (40) Catonsville City (166) Arbutus Dundalk (157) Edgen Halethorpe Brooklyn ? Sparrows Point White Parkville Pikesville Not vn Overlea 95 Rossville ord Mill Locheam NDSOR MILL Woodlawn (702) Rosedale Essex CO (150) y ark 40 Catonsville (166) Arbutus Dundalk (157)

Halethorpe

Sparrows Po

27.4 Heatmap/scatter plot for Los Angeles

