Final Project: Arrest Record of LA County in the COVID-19 Era

Haoran Zhang

12/8/2021

Introduction

Los Angeles is a very vibrant city with lots of neighborhoods. The fusion of different races and ethnics makes everyone nice and friendly. However, the crime issue of Los Angeles is something that cannot be ignored. Various kinds of crimes, including robbery, sexual harassment, burglary, and battery were reported near USC since the beginning of the semester which is concerning.

Main Question

This report aims to analyze the patterns of crimes in the era of COVID-19(2020-present) and the association between arrest counts and the unemployment rate.

Method

The main data which will be analyzed is the arrest data acquired from the LA city database website, including the following key variables:

Variable Name	Description
Report ID	ID for the arrest
Report Type	Booking = the person is booked at a detention facility RFC = the person is released from custody
Age	In years
Sex Code	F = Female, M = Male
Descent Code	Races, B = Black, W = White, C = Chinese etc.
Charge/Description	The charge/description the individual was arrested for
Lat/Lon/Location	The location where the crime incident occurred

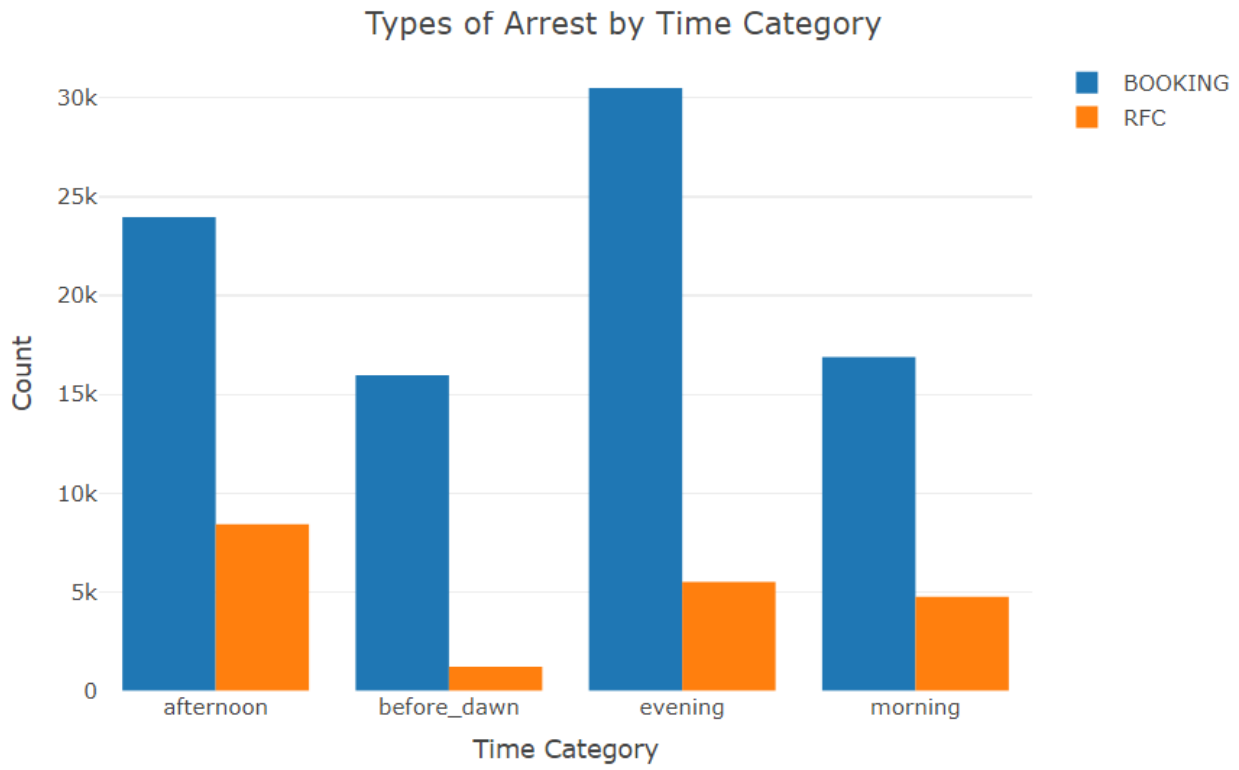
The unemployment data set was acquired from the Fred Economic database website, which contains the monthly unemployment rate collected from Current Population Survey (CPS). The Los Angeles council district area coefficients was captured from ArcGIS hub with further process using QGIS by inputting shape parameters and arrest counts as fields.

Data Cleaning and Wrangling

We first merge the arrest data of 2020 and 2021 together in order to get an overall dataframe containing the arrest data since 2020. Then we modify the date variable by adding month and year

variables for calculating monthly average data. We further exclude those arrest records with unknown location(both latitude and longitude equal 0) and those that are unlikely(age<8). Also we create a time category which partitions a day into four categories with equal length.

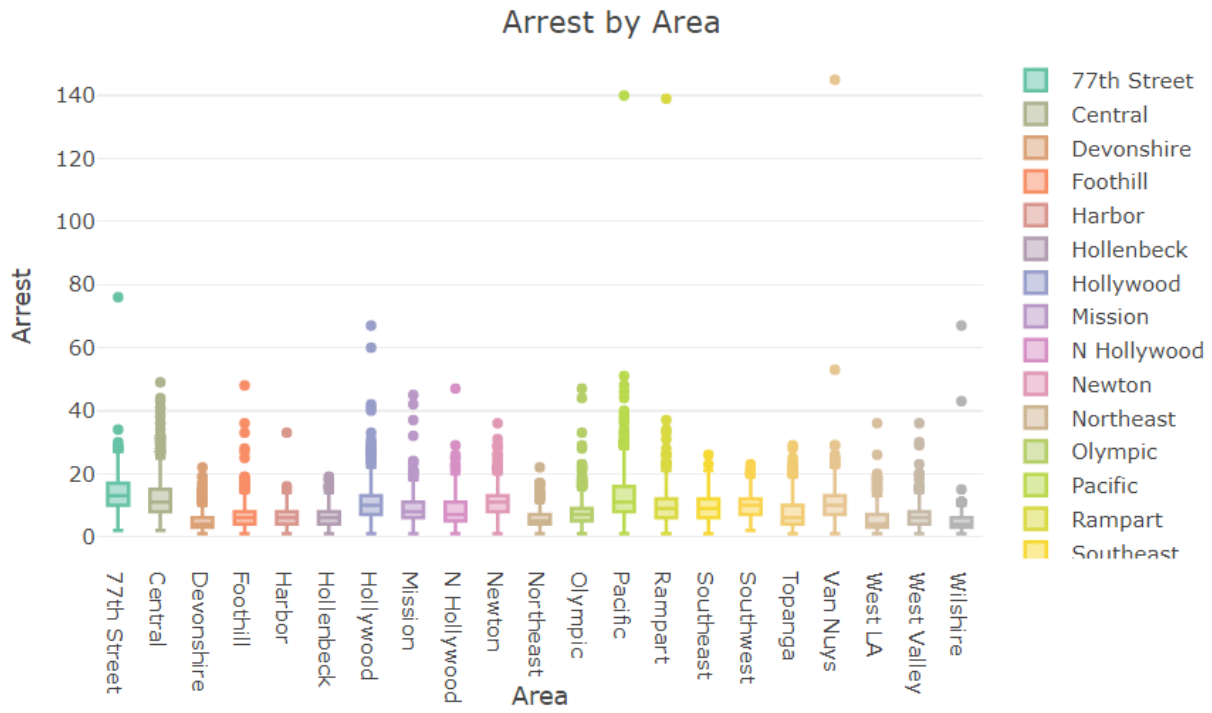
Preliminary results



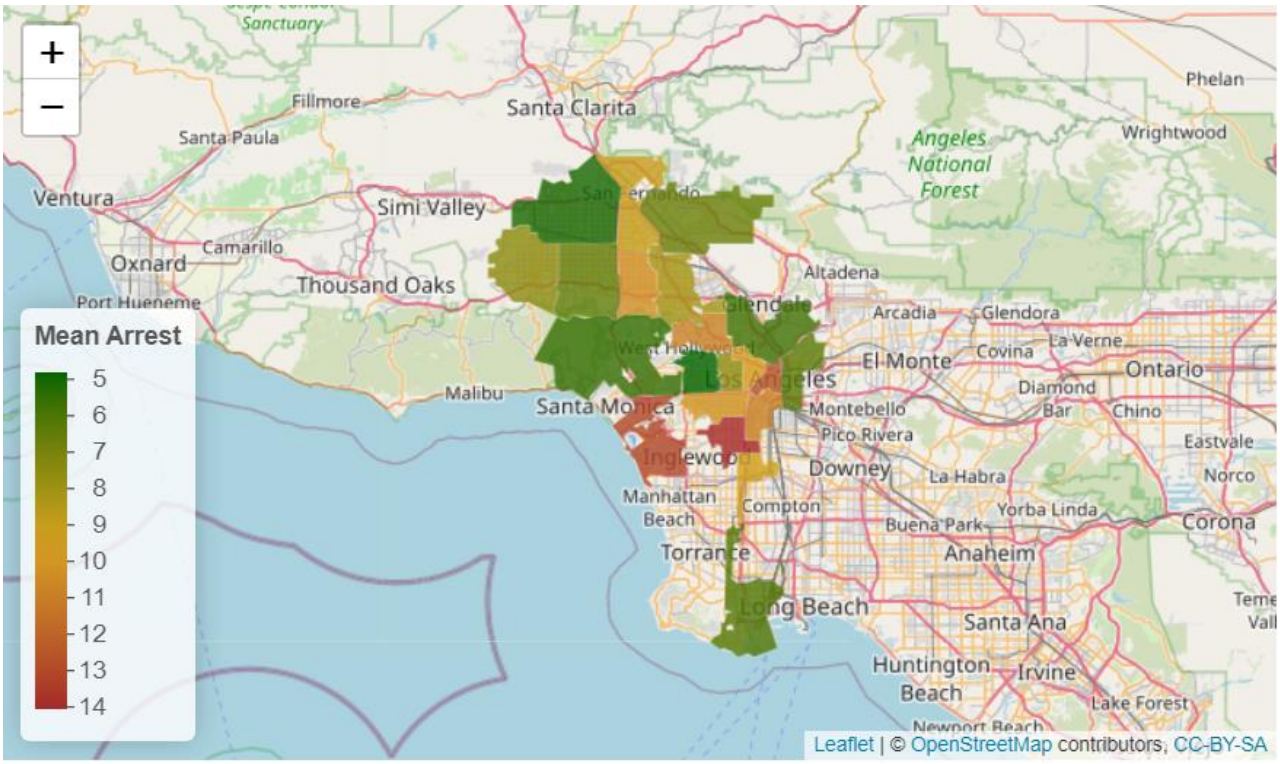
This graph depicts a bar chart of time of arrest by type of arrest. Time of arrest is categorized by partitioning a whole day into 4 categories with equal length starting from midnight. There are two arrest types, booking means the person is booked at a detention facility and RFC means the person is released from custody. From this figure, we can see that despite the number of arrests is the least in the before dawn time category, the ratio of booking comparing to RFC is the largest, indicating that persons are more likely to be booked at a detention facility for arrests occur at night, while persons have higher probability to be released if arrests occur in the day.

Area_Name	Mean_arrest	Max_arrest	Min_arrest	std_arrest
77th Street	14.033389	76	2	5.924741
Central	12.457429	49	2	7.032969
Devonshire	5.022071	22	1	3.186043
Foothill	6.670588	48	1	4.164027
Harbor	6.102694	33	1	2.979620
Hollenbeck	6.317114	19	1	3.141232
Hollywood	10.879800	67	1	6.597271
Mission	8.918197	45	1	4.498047
N Hollywood	8.173623	47	1	4.554013
Newton	10.891486	36	1	4.586943
Northeast	5.726351	22	1	3.120715
Olympic	7.619366	47	1	4.644775
Pacific	13.091820	140	1	9.017444
Rampart	9.742475	139	1	7.178782
Southeast	9.192630	26	1	3.820829
Southwest	10.086811	23	2	3.730800
Topanga	7.557047	29	1	4.731276
Van Nuys	10.275459	145	1	7.437088
West LA	5.549053	36	1	3.970115
West Valley	6.455611	36	1	3.653644
Wilshire	4.832203	67	1	3.758045

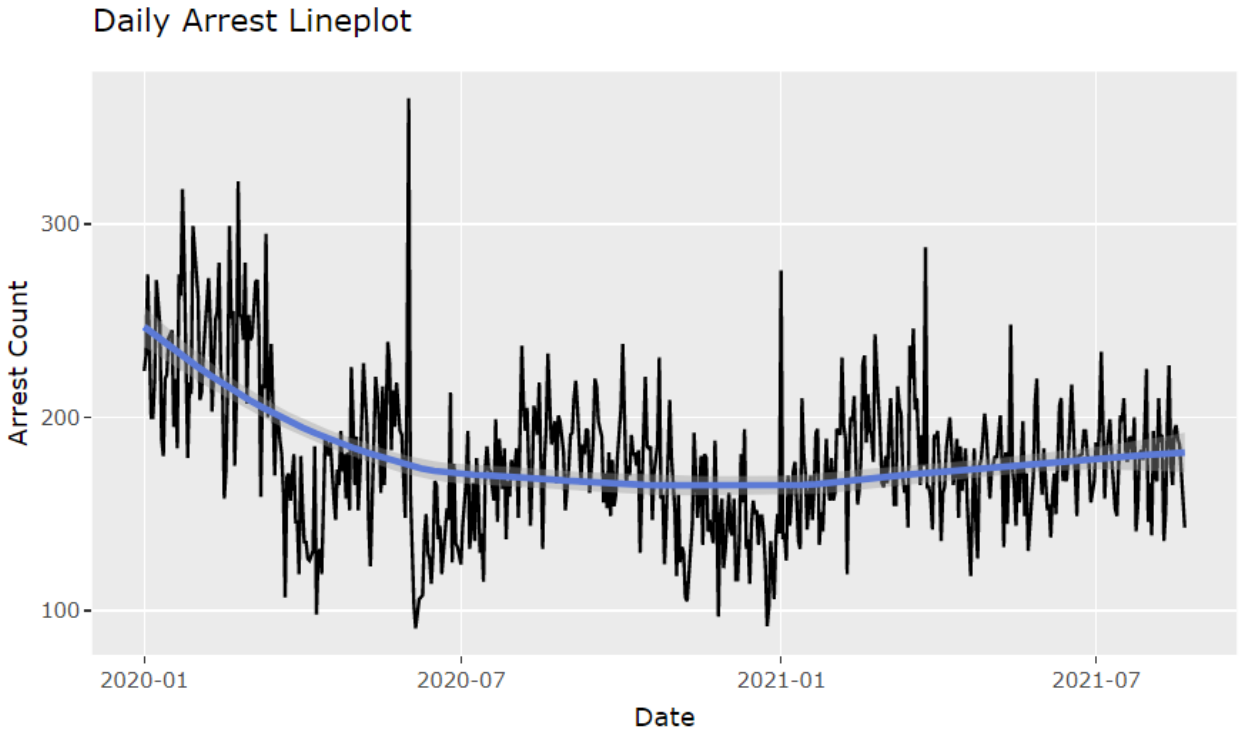
This table provides a brief summary of arrests by Districts. The LAPD has 21 Community Police Stations referred to as Geographic Areas within the department. The figure below shows the boxplot of daily arrest group by the 21 police stations. From the figure we can see that 77th street has the greatest number of reported cases followed by Pacific and Central divisions.



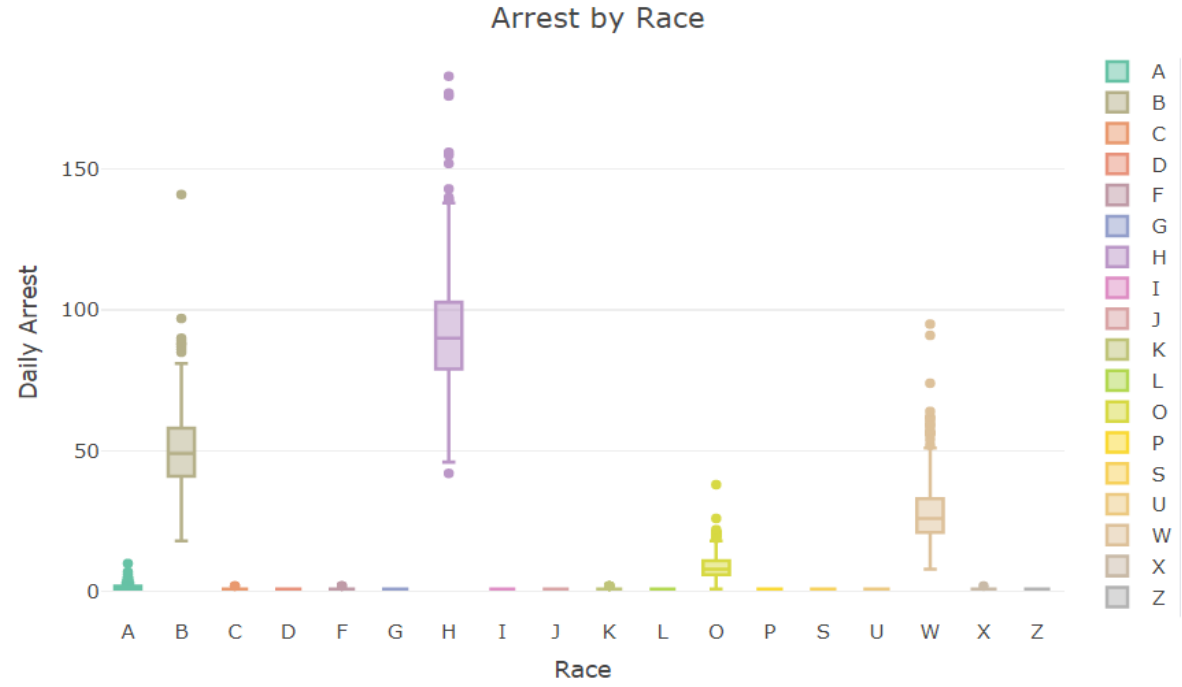
By inputting the geographic parameters of council districts data into QGIS and get a shapefile of LA districts, we are able to construct a choropleth map showing the daily average arrest of each district.



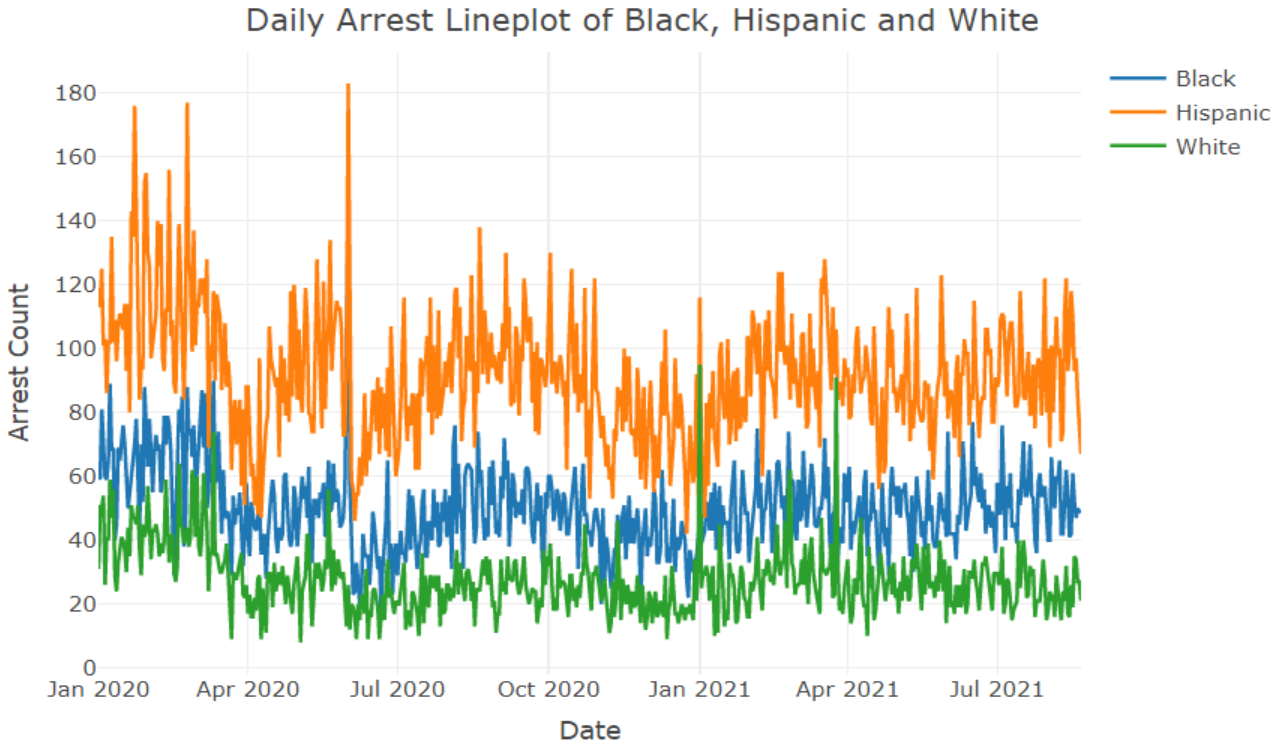
Now we consider the arrest records along timeline. If we build a line chart of arrest counts along the time, we are able to know the variation of arrest counts starting from 2020.



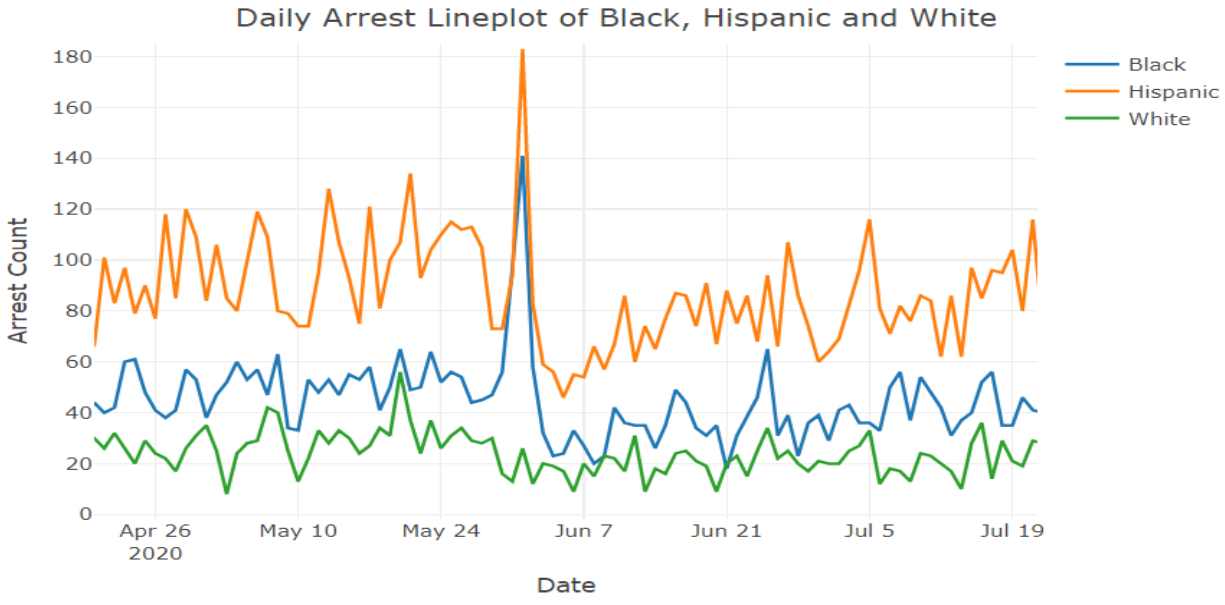
The arrest count decreased drastically in March 2020 but got a sudden burst in late May and early June 2020(May 30 - Jun 3). We claim that this was due to the BLM crime wave after the death of George Floyd on May 25, 2020, and we will show some evidence next.



This figure depicts the boxplot of arrest counts by races. In this graph, we can observe that Hispanic, Black, and White appears to have the three highest arrest records. We will pick these three races as our sample and prove our claim that the sudden increase of arrest records was caused by the BLM crime rate.

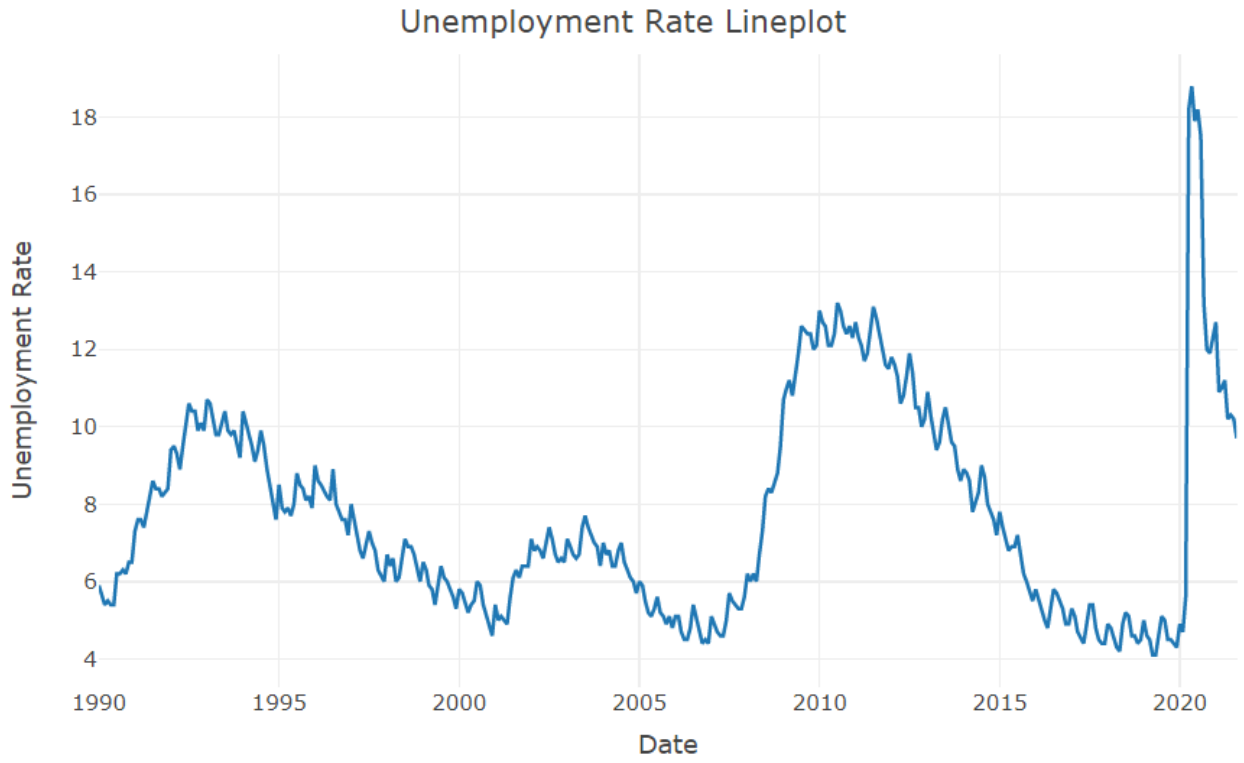


The line chart above depicts the arrest count of Black, Hispanic, and White. If we focus on the time around late May and early June 2020 which is shown below, we can see that the arrest counts of Black and Hispanic increased suddenly while the arrest counts of White continued to decrease. This provides strong evidence between the burst around late May and the BLM crime wave.

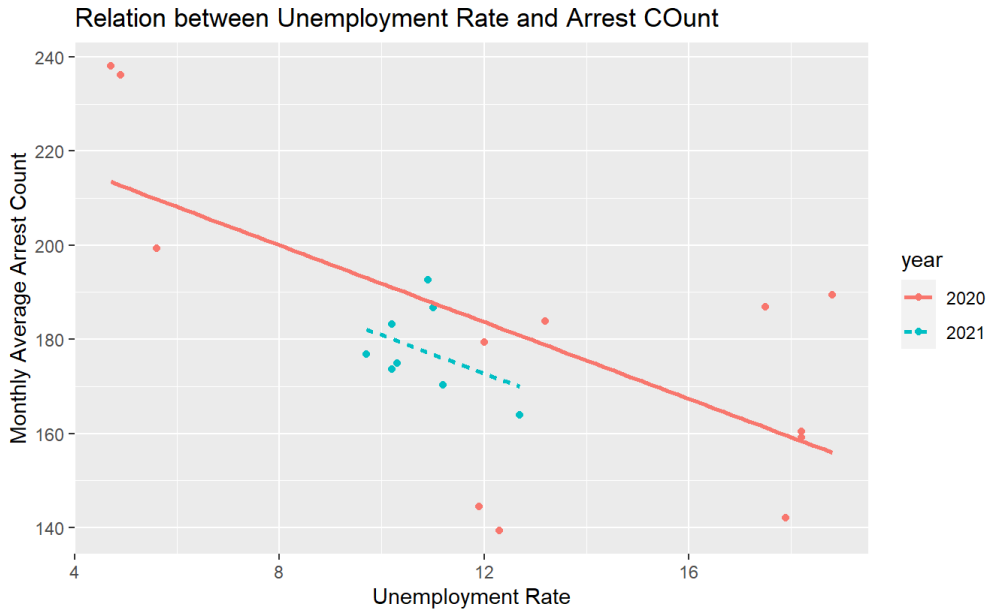


Unemployment Rate

This data table depicts the monthly unemployment rate starting from 1990.



From the plot above, we can see that the unemployment rate fluctuated since 1990 but increased suddenly in early 2020. We will analyze if there is some association between the arrest count and unemployment rate as intuitively, high unemployment rate could cause an increase of crime rate which would result in rise of arrest count.



This graph depicts the monthly average arrest count by the unemployment rate . Beyond expectation, the regression line shows a negative association between unemployment rate and arrest counts. First, the unemployment rate was low while the arrest count was high before the pandemic, which can be seen on the left-up corner of the graph. Then during the lockdown, the unemployment rate grew very high, but due to the lockdown and the lack of population flow after the lock down, the arrest count decreased, such plots can be seen on the right down corner of the graph. In all, the lock down and lack of population flow led to a negative association.

Conclusion

Among the overall arrest record from 2020 to present, most arrests were recorded in the afternoon and evening, those in the afternoon were more likely to be released from custody, while those in the evening and before dawn are more likely to be booked into retention. Number of males that is arrested is much higher than the number of females. By age, males between 20-40 years old contribute the most record. 77th Street District has the largest arrest counts follow by Pacific and Central. By races, Hispanic, Black, and White had the highest number of arrest record, maybe because of their large population base. The overall arrest count was decreasing at the start of the pandemic, probably caused by the lockdown, with sudden bursts of arrests due to the BLM crime wave around June 2020. If we combine the arrest record with the unemployment rate, we can see that the arrest record decreased as the unemployment rate increased surprisingly. Maybe this is because of the lockdown not only caused unemployment, but also caused a large decrease of population flow. Few people on street maybe associate with less crime which finally lead to less arrest records.

Reference Websites

Arrest Data:
https://data.lacity.org/Public-Safety/Arrest-Data-from-2020-to-Present/amvf-fr72
Unemployment Data:
https://fred.stlouisfed.org/series/CALOSA7URN
LA District parameters:
https://hub.arcgis.com/datasets/lahub::lapd-divisions/explore?location=34.020187%2C-118.410104%2C10.37&showTable=true