Arrest Cases Situation in the City of Los Angeles in 2010 and 2019

Jiqing Wu

Nov 18, 2020

Here is the link to the data: https://data.lacity.org/A-Safe-City/Arrest-Data-from-2010-to-2019/yru6-6re4

Introduction

This dataset reflects arrest incidents in the City of Los Angeles from 2010 to 2019. There are 1,320,000 rows in this dataset, each row represents an arrest. In my project, I would focus on the arrest data in 2010 and 2019. In order to have a better understanding of the data in these two years, I raised two questions: Whether there are difference in the number of arrest incidents in the City of Los Angeles in 2010 and 2019, and is there any relationship of age with variables in sex, area of patrol divisions, arrested time and arrest type in the crime situation in 2019? The five main variables I choose to analyze are subject age, subject sex, patrol divisions location, arrested time and the type of charge the individual was arrested for repectively.

Methods

I have separated the data in 2010 and 2019 from the original dataset. There are 162344 rows and 26 columns in crime data 2010, 88296 rows and 26 columns in crime data 2019. After check the main variables (Sex, Age, Area, Time and Arrest Type), I found there are missing values in the Time variable. So I imputed the missing value using the mean value grouped by sex, area and arrest type. I also found a case which the arrested position (LAT=0.00, LON=0.00) is too far away from the others, In order to make map looks more clear, I removed this case only in the mapping step. There is no implausible values in these main variables. I have recoded the sex, arrest type and created a new variable "Part of Day" based on Time for the analysis in the next step.

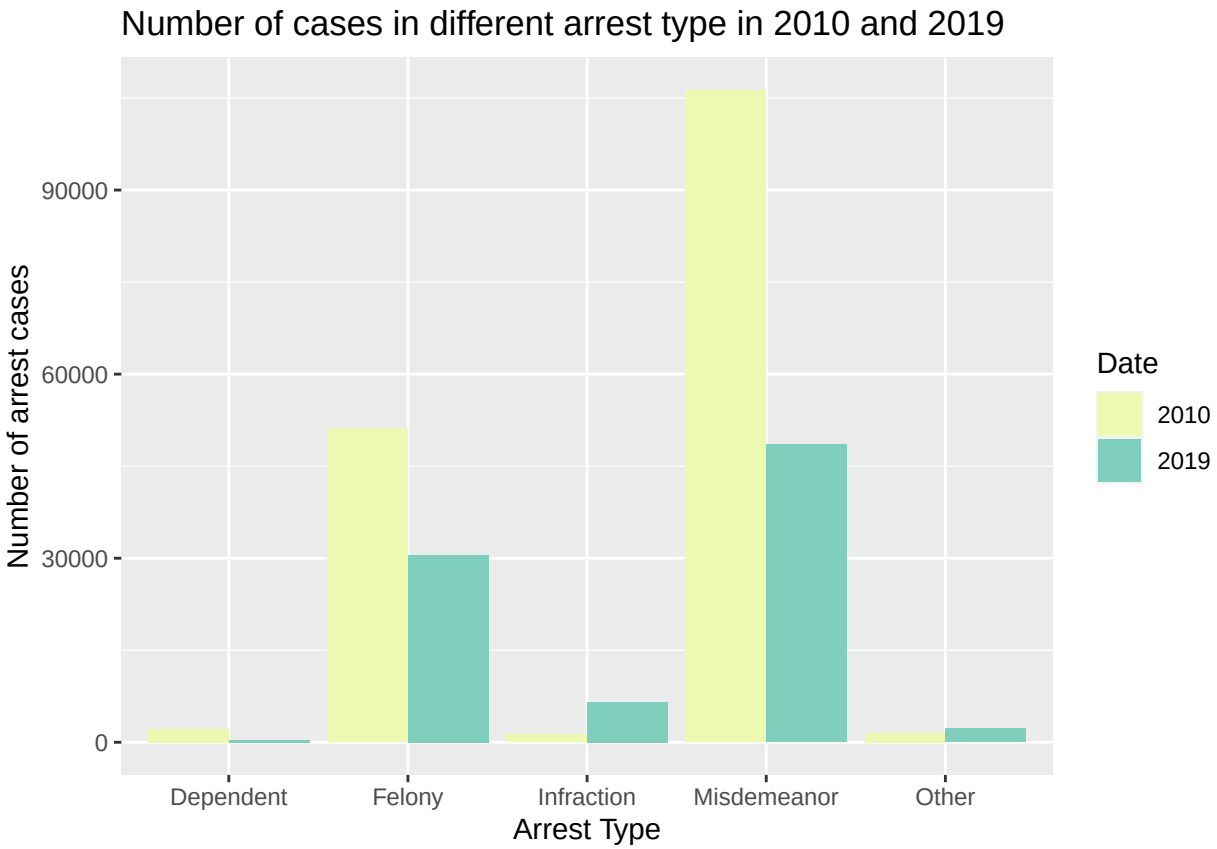
For the difference between 2010 and 2019, I choose to make three barcahrts to show the difference of number of cases in arrest type, subject sex, and area. I also made a map to

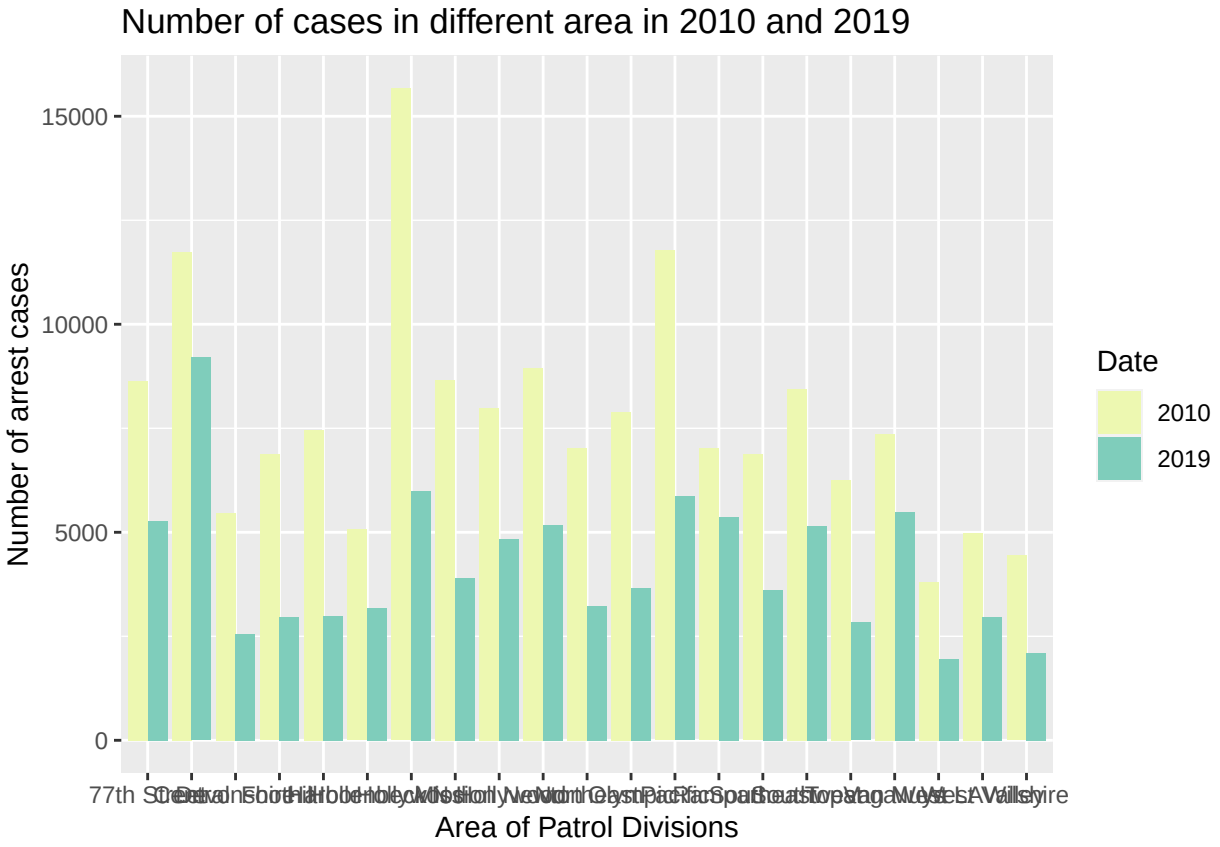
show difference in the distribution of arrested position. However the map for the whole dataset in these two years are too slow to show up (the dataset is too large). I took the distribution of arrested position arrested by Patrol Divisions in central area as an instance. Here is the link to see the map: https://jiqingwu1997.github.io/PM566_Final/.

For the relationship of age with sex, arrested time, area and arrest type in 2019, I used summary tables and several plots to show the variation of age in each variable. The summary tables shows the overall distribution of age in each variables by mean and standard deviation, also the female propotion and male propotion grouped by area, arrested type and arrested time. I used the histogram to show the correlations between age with arrest type and arrested time. The histogram is transparent, therefore the overlaps of age range can be seen clearly. The boxplot shows the distribution of age in males and females in different arrest types. The relationship between area and age is shown in scatter plot. I used the average of age in each area and arrest type. Finally I calculate the p-value and prove the relationship statistically.

Results

After reviewing the summary of data, I found that in this dataset, the total number of arrest cases are seems to be droped. In 2010, there are 129346 males are arrested and 32998 females. While in 2019, there are only 69491 males and 18805 females are arrested, almost half. In 2010, the mean and median age of criminals are 32.16 and 29 while in 2019 increase to 35.25 and 33 respectively. Misdemeanor is the most common type of arrest. There are 106249 subjects in 2010 were arrested because of that while only 48533 in 2019. Patrol divisions in Hollywood has the most arrest cases, in 2010 (15671), while in 2019, Central area became the most. Arrest incidences decrease from 0:00 to 5:00, arrive its lowest point at about 5:00 and highest point at about 16:00 both in 2010 and 2019.


Number of cases in different arrest type in 2010 and 2019
Number of arrest cases
90000
60000
30000
0
Dependent
Felony
Infraction
Misdemeanor
Other
Arrest Type
Date
2010
2019


Number of cases in different area in 2010 and 2019
Number of arrest cases
15000
10000
5000
0
Area of Patrol Divisions
Date
2010
2019

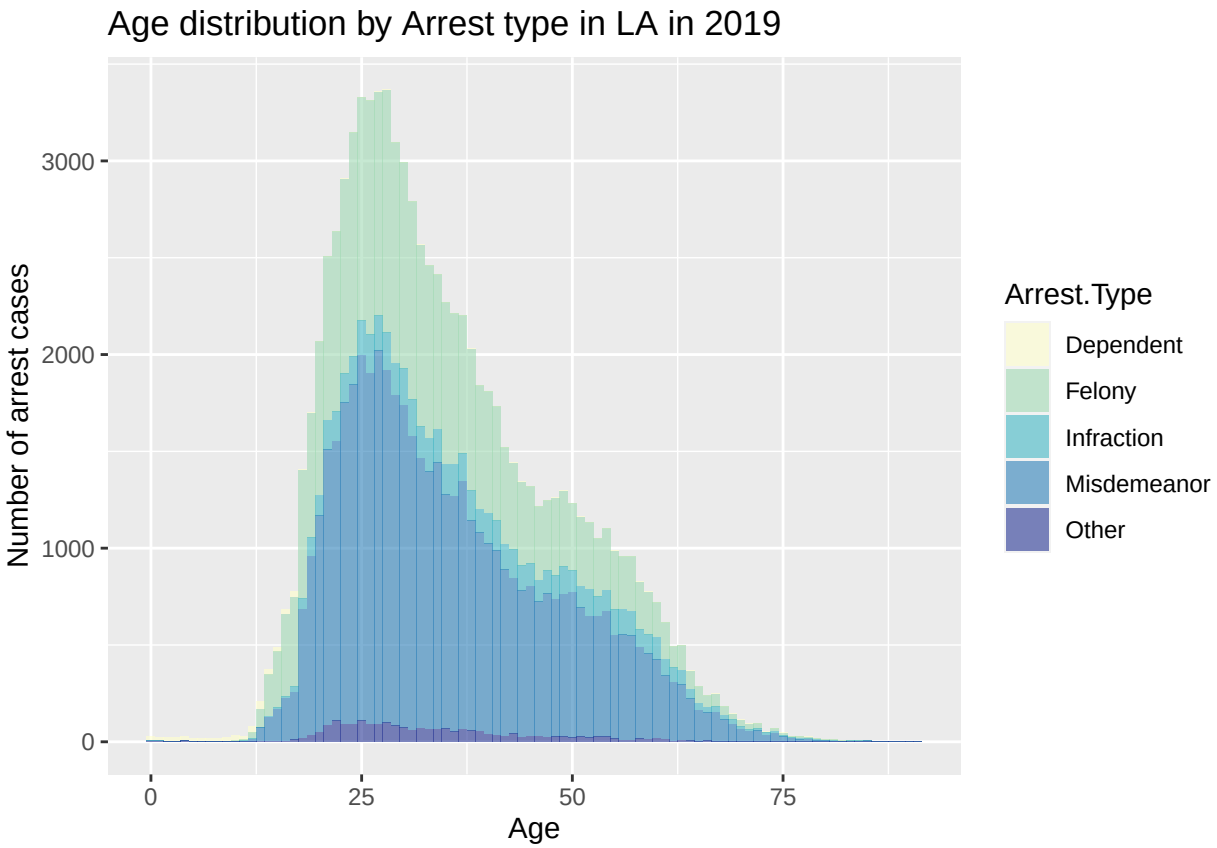
The barcahrts above shows the same results. Felony and misdemeanor are the two main arrest type in both 2010 and 2019. In both of these two arrest type, number of arrest cases in 2019 seems to be less that in 2010. In all of the area and subject sex, number of arrest cases in 2019 show the same results.

```
## # A tibble: 5 x 5
##   Arrest.Type Age_avg Age_sd Male_prop Female_prop
##   <fct>         <dbl>  <dbl> <chr>     <chr>
## 1 Dependent      9.25   5.31 41.81%    58.19%
## 2 Felony        33.8   12.4  81.57%    18.43%
## 3 Infraction    39.9   13.7  82.81%    17.19%
## 4 Misdemeanor   35.8   12.8  76.09%    23.91%
## 5 Other         33.2   11.9  90.86%    9.14%
```

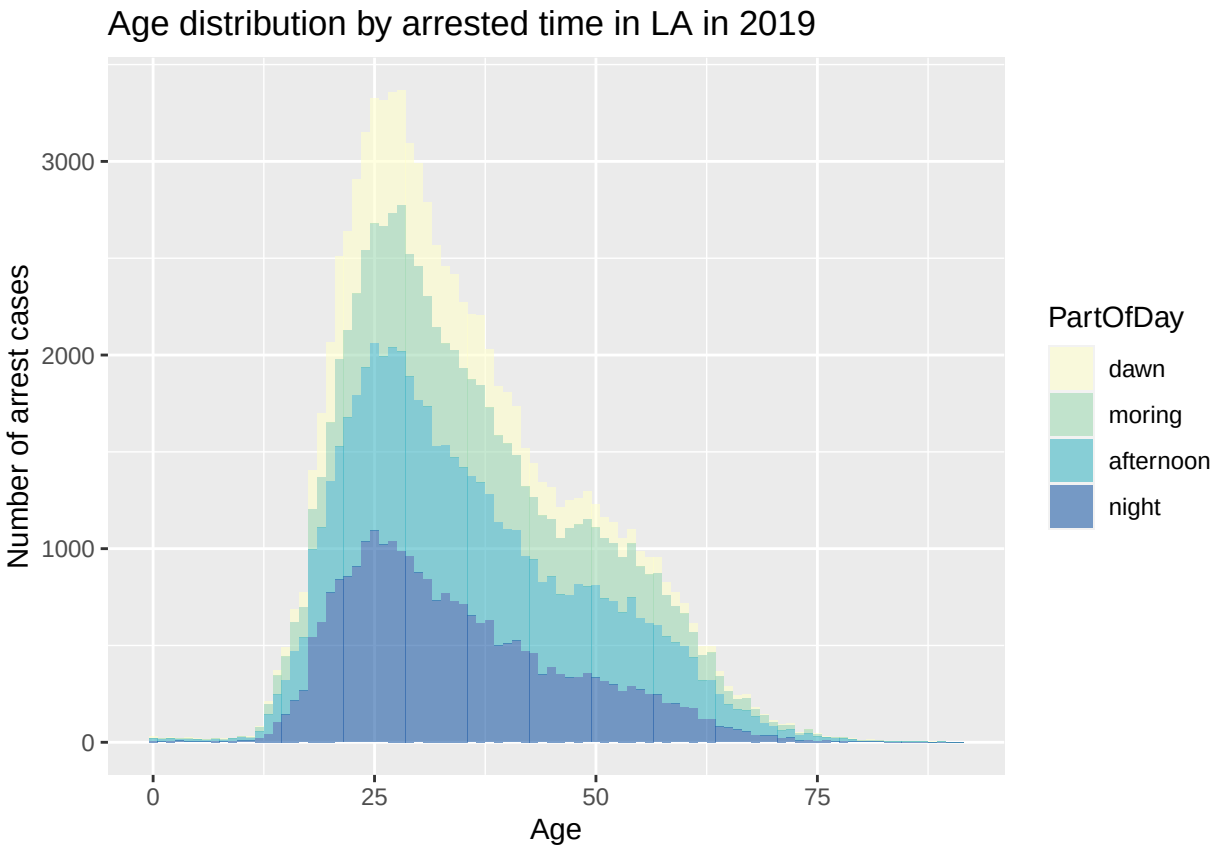
```
## # A tibble: 4 x 5
##   PartOfDay Age_avg Age_sd Male_prop Female_prop
##   <fct>       <dbl>  <dbl> <chr>     <chr>
## 1 dawn         32.4   11.0 76.11%    23.89%
## 2 moring       36.7   13.0 78.59%    21.41%
## 3 afternoon    36.6   13.7 78.37%    21.63%
## 4 night        34.2   12.5 80.46%    19.54%
```

```
## # A tibble: 21 x 5
##    Area.Name   Age_avg Age_sd Male_prop Female_prop
##    <fct>         <dbl>  <dbl> <chr>     <chr>
##  1 77th Street    31.8   12.7 67.95%    32.05%
##  2 Central        40.0   13.6 81.87%    18.13%
##  3 Devonshire     33.9   13.4 71.93%    28.07%
##  4 Foothill       34.4   12.1 80.94%    19.06%
##  5 Harbor         35.4   13.1 74.09%    25.91%
##  6 Hollenbeck     33.1   12.6 81.74%    18.26%
##  7 Hollywood      33.1   11.1 81.7%     18.3%
##  8 Mission        33.0   11.9 80.61%    19.39%
##  9 N Hollywood    35.8   12.0 78.88%    21.12%
## 10 Newton         34.8   12.9 81.69%    18.31%
## # ... with 11 more rows
```

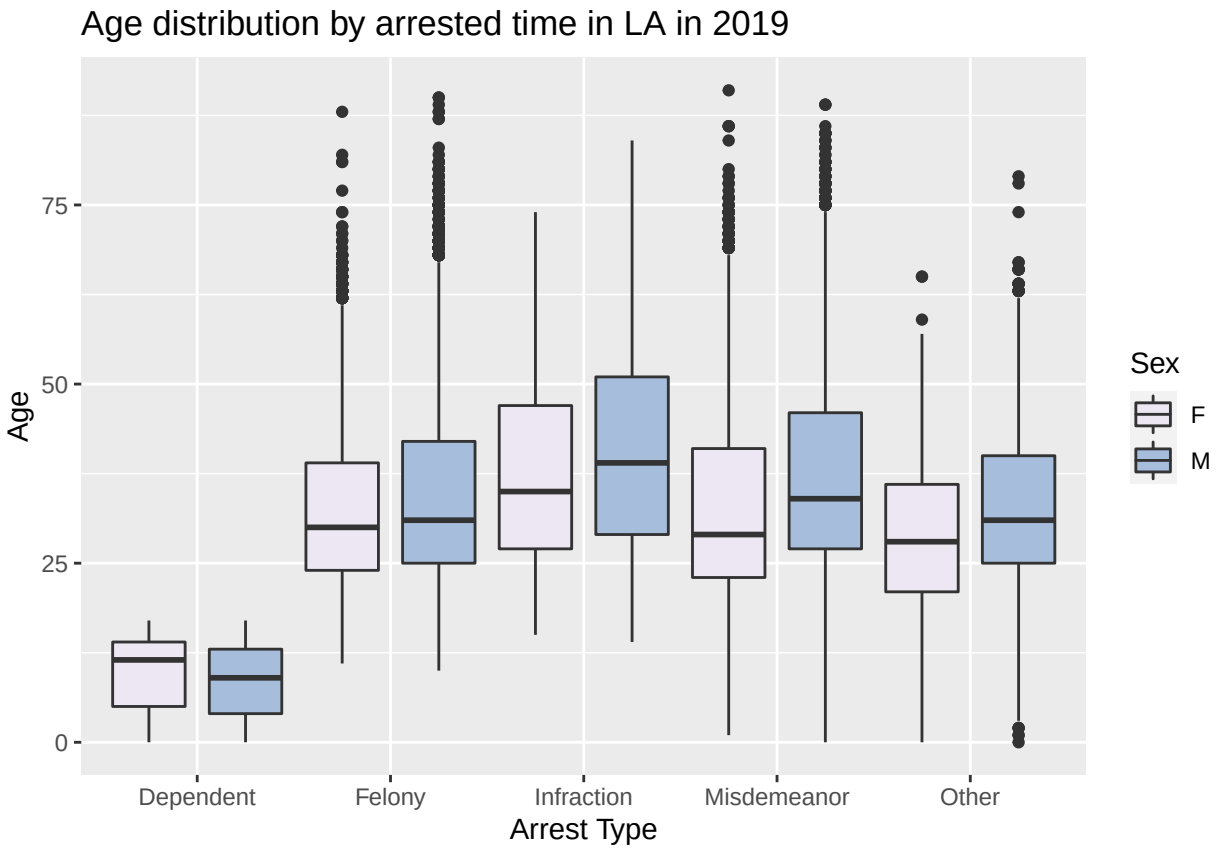
The summary tables show the variation of subject age in different arrest type, arrested time, sex and area statistically. For example, mean age in dependent type is only 9.25 while that in other types are all more than 30.



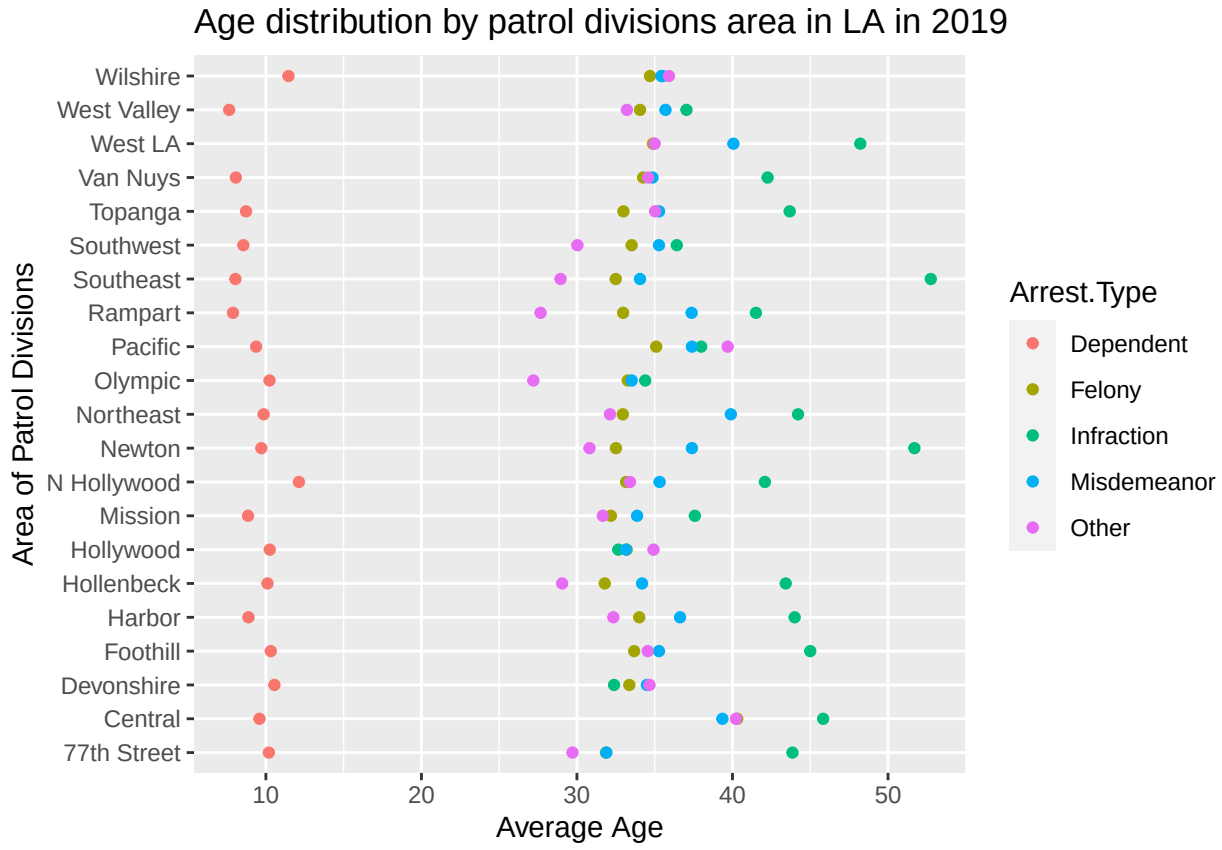
Histogram of age by arrest types indicates that there are still overlaps of age range, especially in felony, infraction, misdemeanor and other type. Most of the subjects age are concentrated at 20 to 30.



It can be clearly seen from this histogram that most of the arrest cases were happening in dawn. The age distbution are highly overlapped.



From the boxplot, in most arrest types, the mean age of males tend to be older than females.



The scatter plot does not show any pattern for average age in each area of patrol divisions.

```
##                Df   Sum Sq Mean Sq F value Pr(>F)
## Arrest.Type     4   504230  126057   783.1 <2e-16 ***
## Residuals   88291 14212458     161
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##                Df   Sum Sq Mean Sq F value Pr(>F)
## PartOfDay       3   229989   76663   467.2 <2e-16 ***
## Residuals   88292 14486698     164
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##                Df   Sum Sq Mean Sq F value Pr(>F)
## Sex.Code        1   141564  141564   857.6 <2e-16 ***
## Residuals   88294 14575123     165
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##                Df   Sum Sq Mean Sq F value Pr(>F)
```

```
## Area.Name      20   472731   23637   146.5 <2e-16 ***
## Residuals   88275 14243957     161
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The results can be conclued in the table below.

Variables	F-value	p-value	Related with age?
Arrest Type	783.1	<0.05	Yes
Arrest Time	467.2	<0.05	Yes
Criminal Sex	857.6	<0.05	Yes

Brief Conclusion

In summary, the number of arrest incidences in the City of LA between 2010 and 2019 seems to be different. In 2019, the number of arrest cases seems to be less than that in 2010 from the aspects of different sex, area and arrest type. The age of arrest subjects is correlated in arrest type, arrest time, criminal sex, and patrol divisions area in the City of LA in 2019.