

testing

Shangjun Jiang

2 August 2016

Introduction

For the final project, our group cleaned, explored, and analyzed four different data sets from the City of Berkeley that contained information on stops, calls for service, arrests, and jail bookings made by the police department in 2016 (and 2015 as well for the stop data). Rather than limiting the project to one set of data, four sets were chosen in order to gain a more holistic and comprehensive understanding of the data. The additional variables included expanded the project's capacity for manipulating data, examining relationships, and improving result reliability.

Data Collection

- Stop Data (16,000)
- Arrest (200)
- Jail Bookings (250)
- Calls for Service (4,000)

With the variety of resources and information made available by collating multiple data sets, the objective for the project was to study differences in police activity (in terms of call requests and patrols) and intensity of assessed offenses based on time, race, gender, age, and mental health. As a challenge, another project target was to create a map applet depicting the density of police activity in Berkeley with an interactive component allowing the visitor to input an address and observe their proximity and observe the types of incidences that occurred most commonly in the area.

For the stop data, there were 16,255 incidents assessed by the Berkeley police. In each case, the call

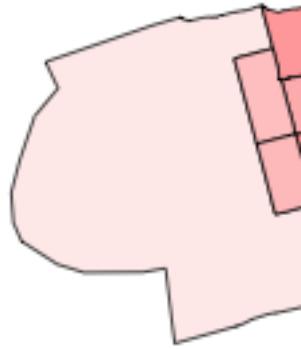
The calls for service data (4913 offenses) contained information on the case number, the offense type w

The data for arrests (205), and jail bookings (223) contained similar information on case/arrest/booking

Data Analysis

```
## [1] ""          "00000"      "AR"         "AR, M"       "AR, M, P"  
## [6] "AR, P"     "FC"        "FC, M"       "IN"         "M"  
## [11] "M, P"       "MH"        "MH, AR, P"   "MH, M"       "MH, P"  
## [16] "P"          "TOW"        "TOW,"        "TOW, AR"    "TOW, AR, M"  
## [21] "TOW, AR, P" "TOW, CO, P"  "TOW, FC"     "TOW, IN"     "TOW, IN, AR"  
## [26] "TOW, M"     "TOW, P"  
  
## Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=37.865887,-122.276384&zoom=14&size=600x300
```

Population by census block group:



We found the census data online and map the Berkeley population in block groups.

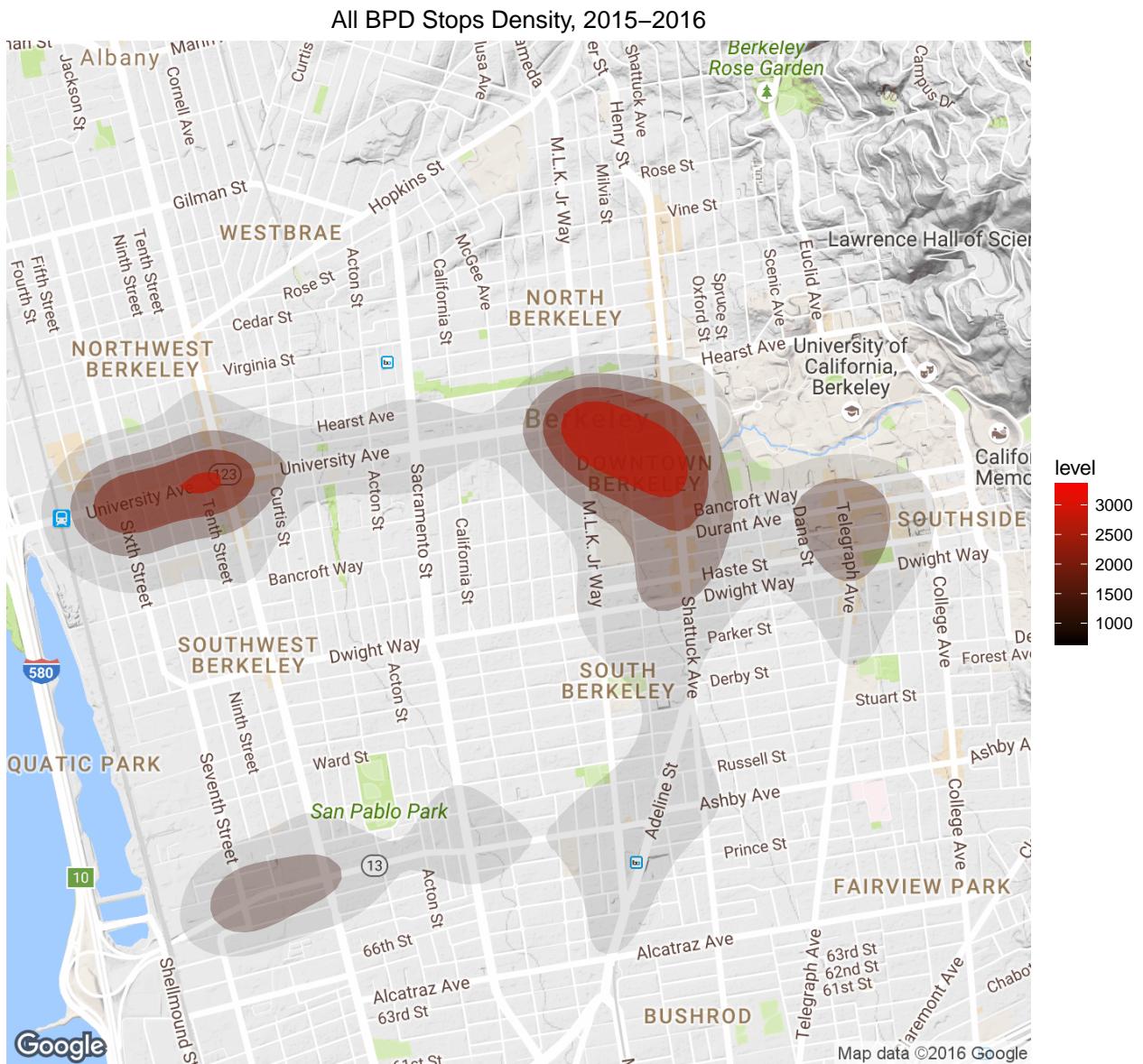
About the Berkeley PD Stop Data

The Berkeley PD stop data reflects accidents that have been visited by the Berkeley Police Department. Raw data set is downloaded from City of Berkeley Open Data website and provided by Berkeley Police Department.

Analysis in this part focuses on the information provided by the dispositions variable. The dispositions variable can be divided into three types. The first type only includes more than one six-character person message. Each character in each person message represents the race, gender and age range of the person, reason and enforcement of the stop, and the car search information during the stop. The second type of the dispositions includes only additional dispositions, such as P for primary cases. The third type of dispositions includes both person information and additional dispositions.

MAP 1:All BPD Stops Density, 2015-2016

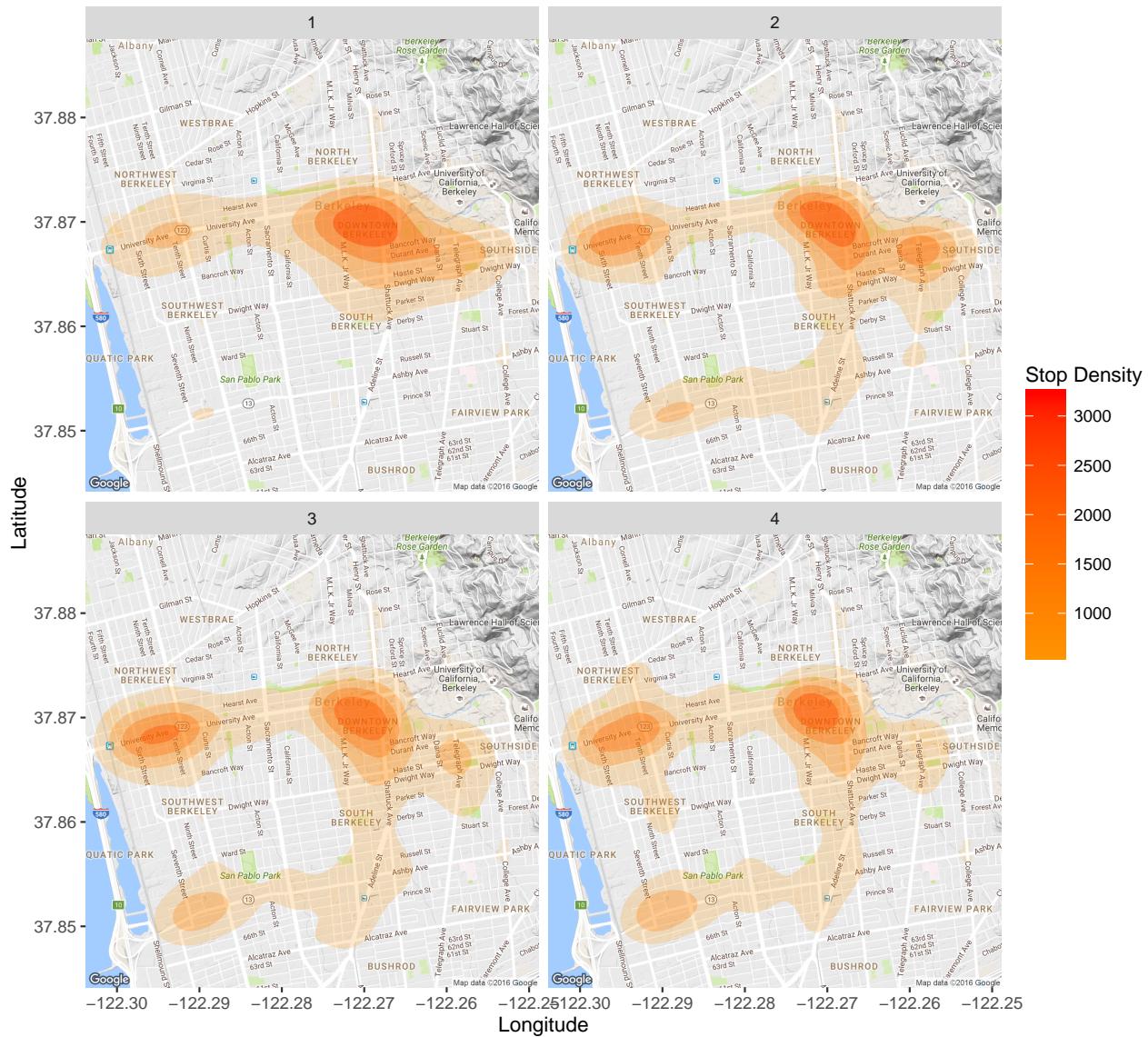
```
## Warning: Removed 933 rows containing non-finite values (stat_density2d).
```



MAP 2: BPD Stop Contour Map of Berkeley by Age Range

```
## Warning: Removed 933 rows containing non-finite values (stat_density2d).
```

BPD Stop Contour Map of Berkeley by Age Range



Analysis of Race

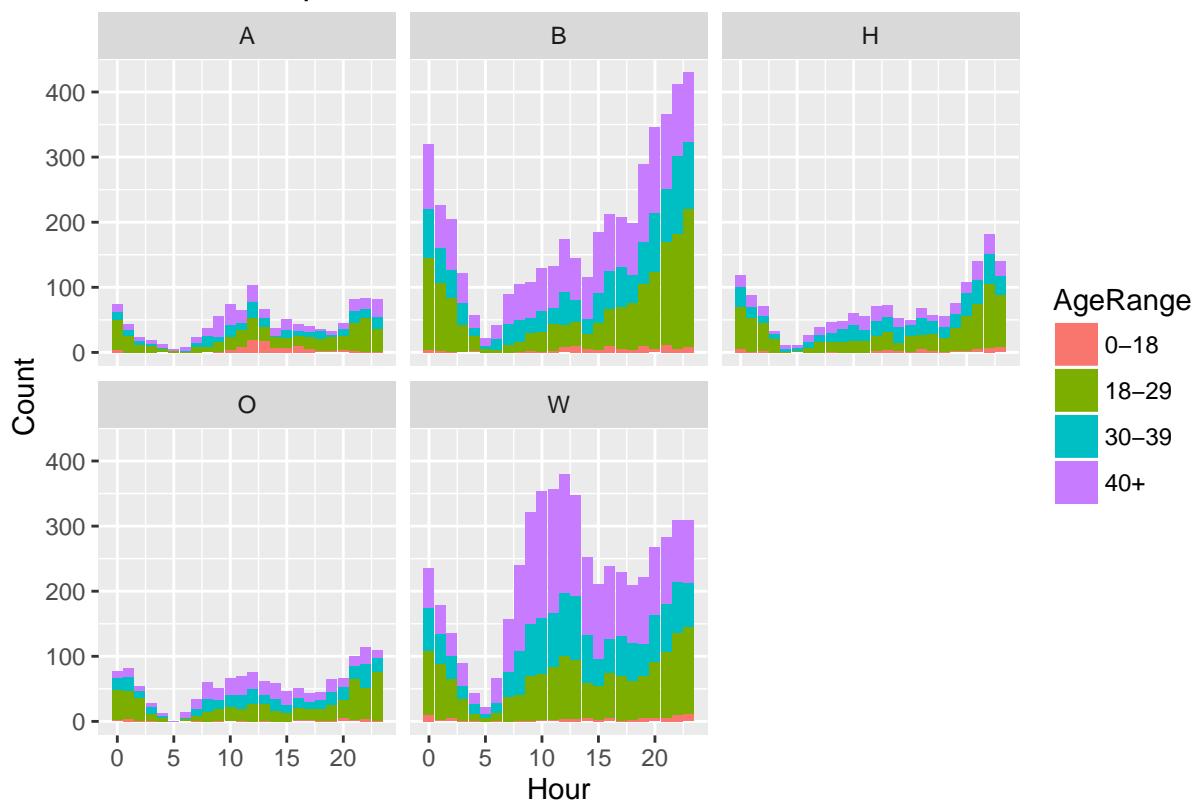
Picture: count of person recorded in each race

Table: count of person recorded in each race

Race	Count	0-18	18-29	30-39	40+
A (Asian)	1141	8.50%	43.47%	21.03%	26.99%
B (Black)	4636	2.29%	34.97%	25.09%	37.66%
H (Hispanic)	1676	3.22%	46.12%	29.18%	21.47%
O (Other)	1384	2.67%	43.28%	29.17%	26.87%
W (White)	5454	1.72%	29.45%	25.39%	43.44%

Picture: count of person recorded of each race in each hour

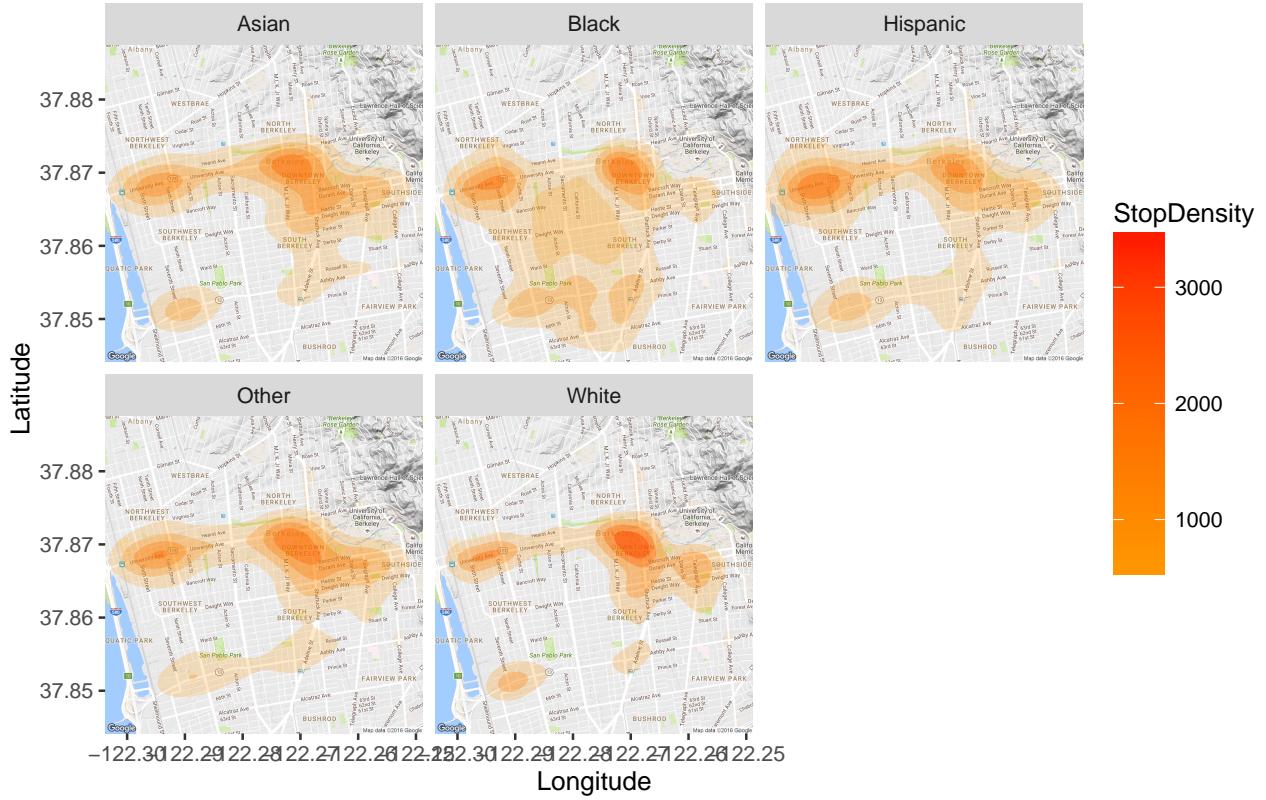
Count of person recorded of each race in each hour



MAP 3: BPD Stop Contour Map of Berkeley by Race

```
## Warning: Removed 933 rows containing non-finite values (stat_density2d).
```

BPD Stop Contour Map of Berkeley by Race



- (1) Among the 14291 personnel information recorded, white people contributes the largest percentage. 38.16% of all records, namely 5454 people are white. black people contributes the second largest percentage. 32.44% of all records, namely 4636 people are black.
- (2) The percentage of people aged from 0 to 18 of each race is less than 9%. People aged from 0 to 18 contributes the smallest percentage in each race, which makes sense in daily life.
- (3) Among white people and black people recorded by the BPD, people aged greater than 40 contribute the largest percentage, which are 43.44% and 37.66% respectively.
- (4) Among Asian, Hispanic and other people recorded by the BPD, people aged from 18 to 29 contribute the largest percentage, which are 43.37%, 46.12% and 43.28%, respectively.
- (5) Black people contribute a especially high percentage of the incidents at night, while the white people contribute a especially high percentage of the incidents at the noon.

Analysis of Day

Picture: count of person recorded in each day in a week

Picture: count of person recorded in each day in a week

Day	Count	A	B	H	O	W
1	1692	116	653	221	159	543
2	1665	149	543	182	166	625
3	2344	189	645	231	236	1043

Day	Count	A	B	H	O	W
4	2287	206	616	251	244	970
5	1941	152	592	223	168	806
6	2231	186	764	268	203	810
7	2131	143	823	300	208	657

- (1) Among all the day in a week, the number of incidents happened on Sunday and Monday are much less than those happened through Tuesday and Saturday.
- (2) The count of records of Asian people is much less than that of any other race of people in each day and the whole week.
- (3) Asian people are less liable to commit an incident on Sunday, and more liable to commit an incident on Wednesday.
- (4) Black and Hispanic people are less liable to commit an incident on Monday, and more liable to commit an incident on Saturday.
- (5) White people are less liable to commit an incident on Sunday, and more liable to commit an incident on Tuesday and Wednesday, which is similar to Asian people.

Analysis of incidents of each hour in a day

Picture: count of person recorded in each hour in a day

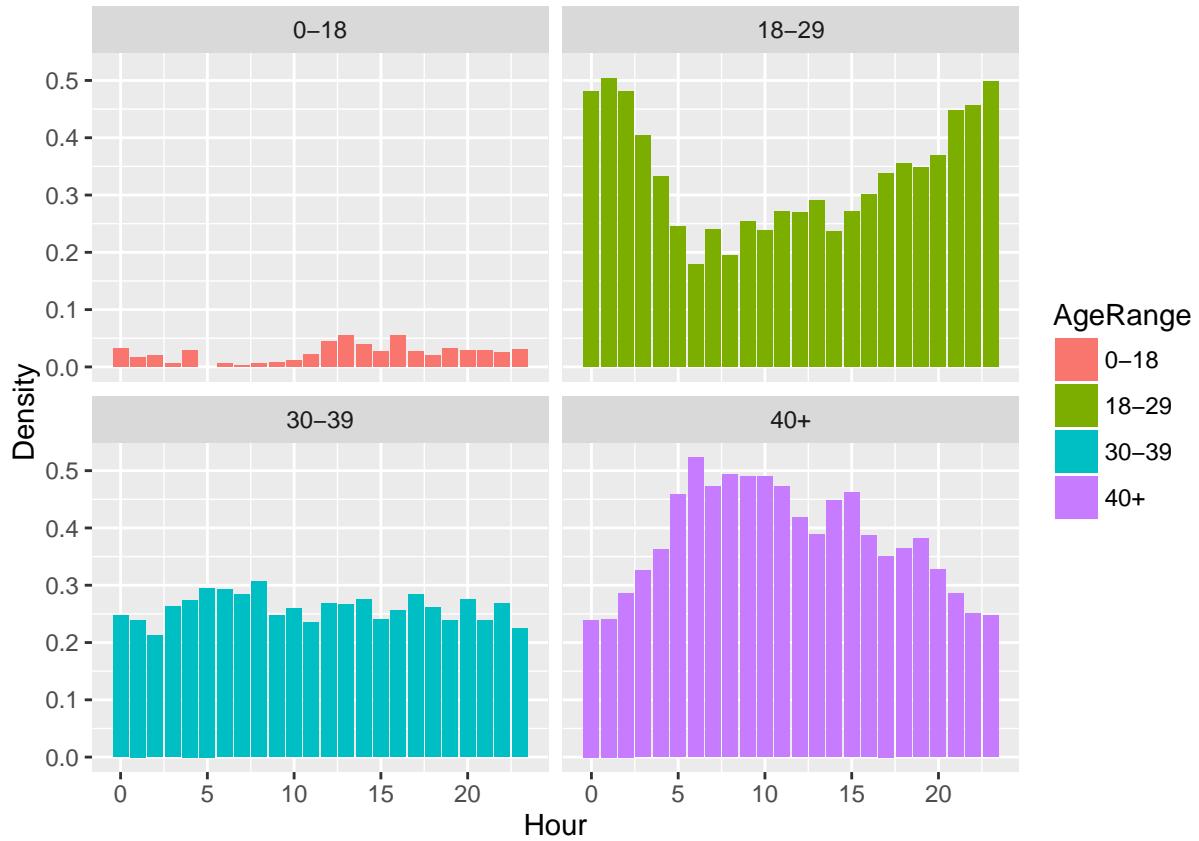
Note: you can change the value of the input slider to see the number of incidents happened in specific time span of the day.

Table: count of person recorded in each hour in a day

Hour	Count	Hour	Count	Hour	Count	Hour	Count
0	823	6	157	12	802	18	544
1	618	7	341	13	695	19	685
2	489	8	489	14	515	20	834
3	289	9	582	15	542	21	971
4	135	10	682	16	612	22	1098
5	61	11	678	17	578	23	1071

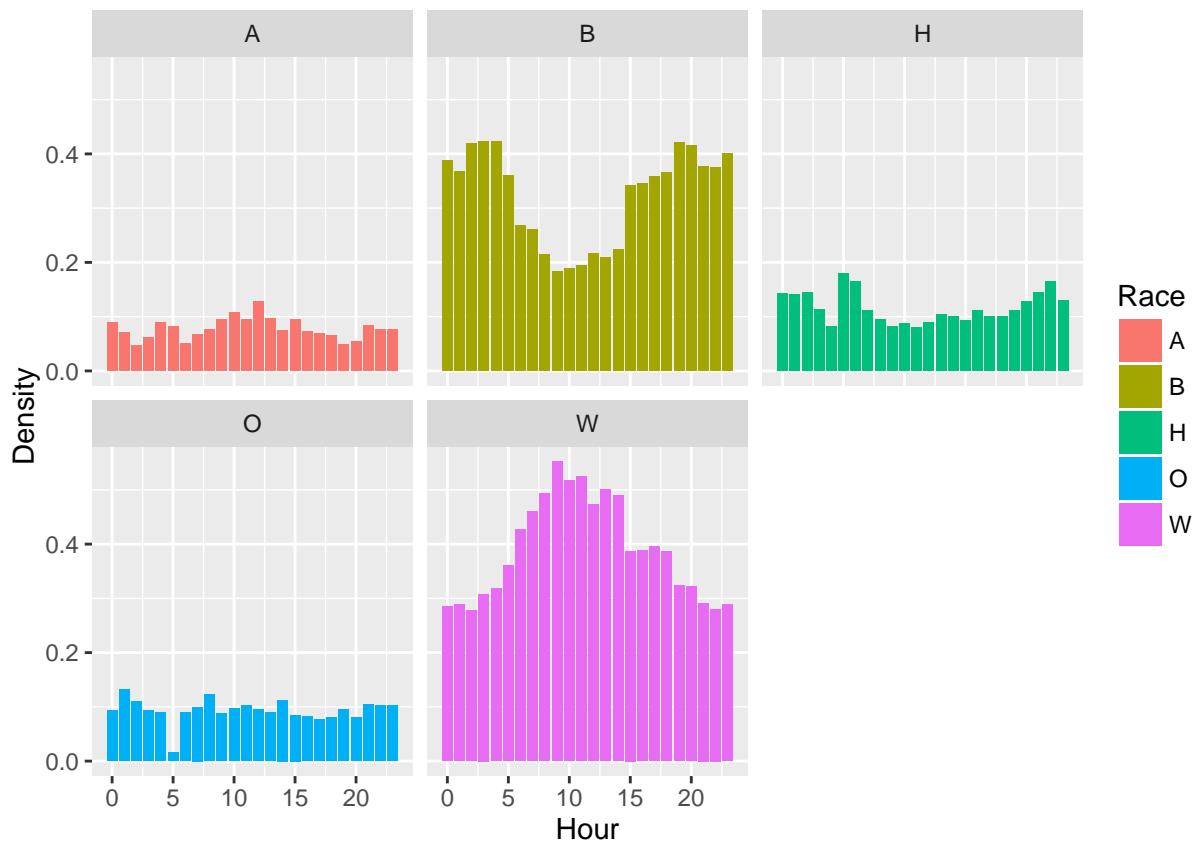
- (1) time span of 22:00 to 22:59 in a day has the highest incidents count, and 23:00 to 23:59 has the second highest, which means that Berkeley area is most dangerous from 10pm to 12am in a day, which corresponds to daily life experience.
- (2) From 12am to 5am, incidents count decreases gradually. A reasonable conjecture of this fact is that more and more people choose to sleep as time goes by in this time span. For the similar reason, incidents count increases gradually from 5am to 11am.
- (3) After the evening until the midnight, the incidents number increases gradually again in a day. The tally of the statistics and observation in daily life is in a good agreement.

Picture: Probability of BPD stop for a specific age range in a specific hour



- (1) People aged from 18 to 29 are obviously liable to commit incidents at night. The average ratio of people aged from 18 to 29 to all people stoped by the Berkeley Police Department at night is greater than 40%.
- (2) People aged greater than 40 are obviously liable to commit incidents in the daytime. The average ratio of people aged greater than 40 to all people stoped by the Berkeley Police Deparment during the day is greater than 40%.
- (3) Ratio of incidents commited by people aged between 0 and 18 and people aged from 30 to 39 fluctuates during the daytime and the night, with an average ratio of 2.5% and 25% respectively.

Probability of BPD stop for a specific race in a specific hour



- (1) Black people are obviously liable to commit incidents at night. The average ratio of black people to all people stoped by the Berkeley Police Department at night is about 40%.
- (2) White people are obviously liable to commit incidents in the daytime. The average ratio of white people to all people stoped by the Berkeley Police Deparmtent during the day is about 45%.
- (3) Ratio of incidents commited by Asian people, Hispanic people and other people fluctuate during the daytime and the night, with an average ratio of 8%, 11% and 9% respectively.

Analysis of preference of Berkeley Police Department arresting people

Picture: Probability of arrested by BPD of a specific race for a specific reason

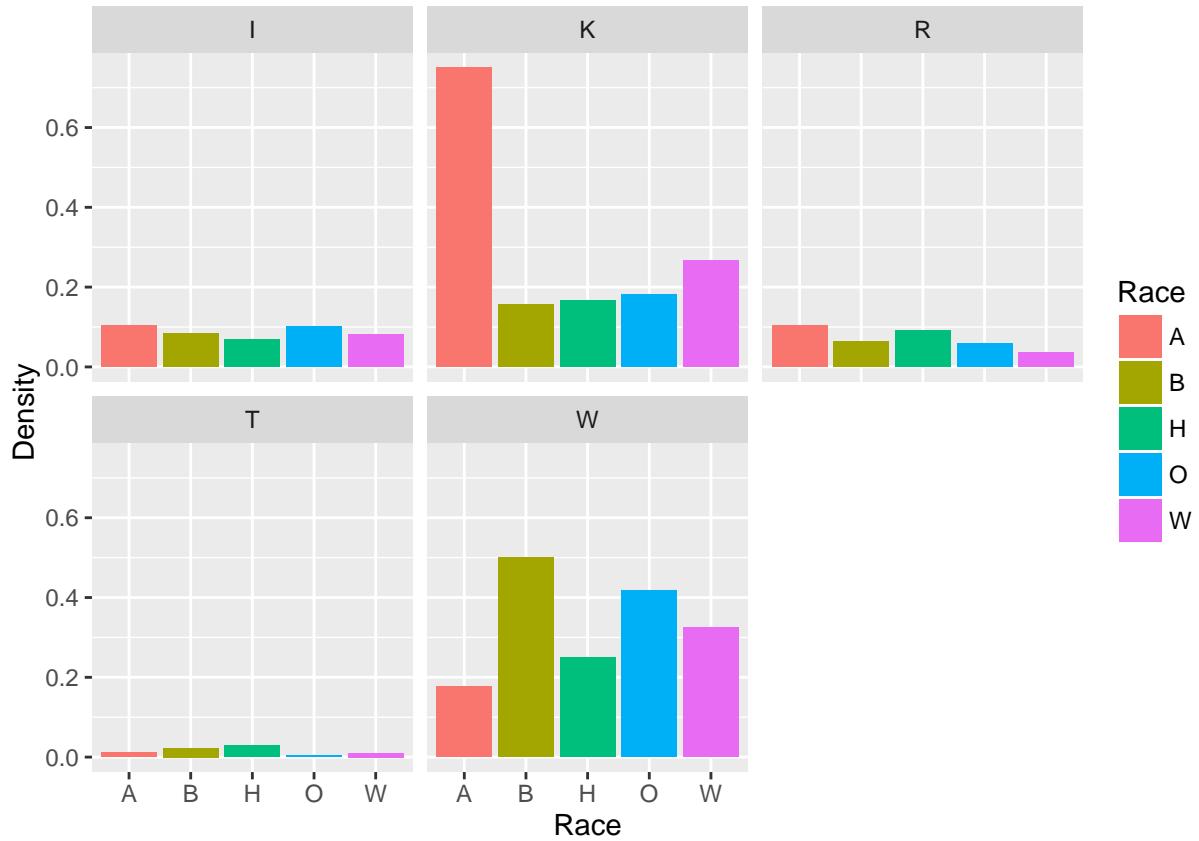


Table: Probability of arrested by BPD of a specific race for a specific reason

P	I	K	R	T	W
A	10.34%	75%	10.34%	1.24%	17.60%
B	8.31%	15.79%	6.44%	2.28%	50.00%
H	6.98%	16.67%	9.09%	2.90%	25.00%
O	10.14%	18.18%	5.80%	0.41%	41.67%
W	8.14%	26.79%	3.40%	1.04%	32.56%

Note: in the picture and table above, I for Investigation, T for Traffic, R for Reasonable Suspicion, K for Probation/Parole, W for wanted.

- (1) The probability of arrested in a stop with the reason Traffic is much lower than other reasons. The average conditional probability of arrested given reason is traffic is 1.58%.
- (2) The probability of arrested in a stop with the reason Wanted is much higher than other reason. The average conditional probability of arrested given reason is wanted is 33.33%.
- (3) An interesting fact is that the conditional probability of arrested in a stop with the reason Probation or Parole and race Asian is 75%, which is much higher than that of any other race. Asian people is much more liable to be arrested by Berkeley Police Department during Probation or Parole if stopped by the BPD.

Picture: Probability of arrested by BPD of a specific age range for a specific reason

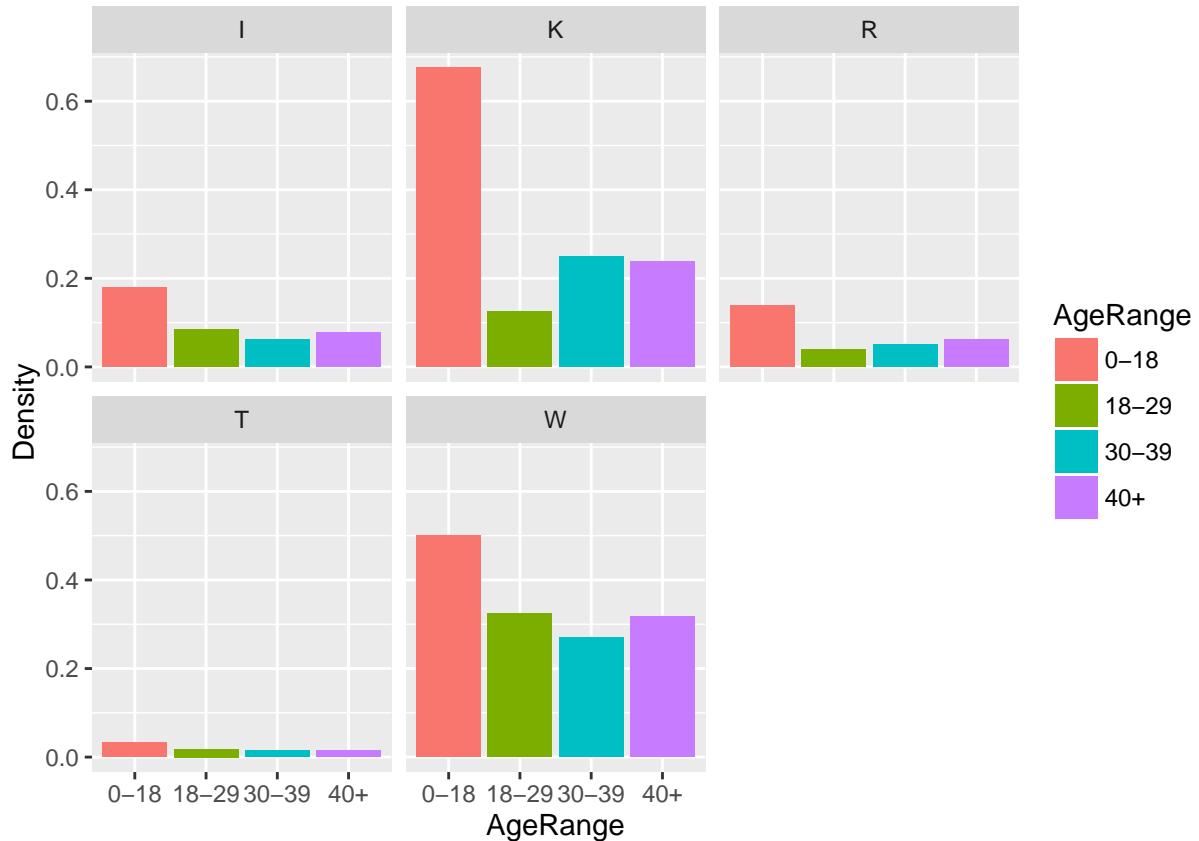


Table: Probability of arrested by BPD of a specific age range for a specific reason

P	I	K	R	T	W
0-18	17.95%	67.57%	13.95%	3.24%	50.00%
18-29	8.45%	12.50%	3.87%	1.82%	32.35%
30-39	6.32%	25.00%	5.04%	1.49%	26.92%
40+	7.87%	23.86%	6.13%	1.41%	31.82%

- (1) Similar to the conclusions above, the probability of arrested in a stop with the reason Traffic is the lowest and wanted is the highest.
- (2) Again similar to the conclusion above, an interesting fact is that the conditional probability of arrested in a stop with the reason Probation or Parole and age range 0-18 is 67.56%, which is much higher than that of any other age range.
- (3) The conditional probability of arrested given person involved aged from 0 to 18 in a stop given any reason is more than that of given person with any other age range. Therefore, in a stop, teenagers is more liable to be arrested by Berkeley Police Department.

Picture: Probability of arrested by BPD of a specific gender for a specific reason

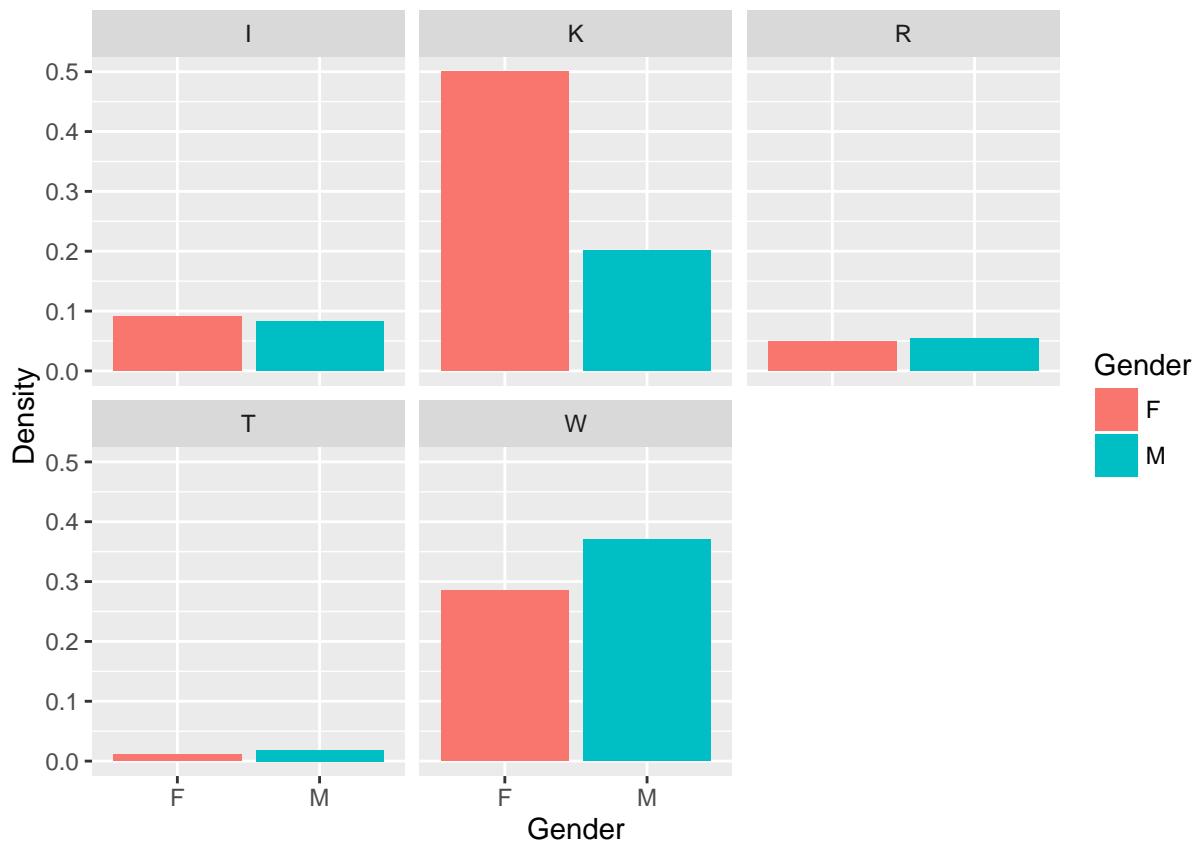


Table: Probability of arrested by BPD of a specific gender range for a specific reason

P	I (Investigation)	K (Probation/ Parole)	R (Reasonable Suspicion)	T (Traffic)	W (Want)
F	9.05%	50.00%	5.00%	1.10%	28.57%
M	8.26%	20.11%	5.42%	1.85%	37.04%

- (1) Similar to the conclusions above, the probability of arrested in a stop with the reason Traffic is the lowest and wanted is the highest.
- (2) Again similar to the conclusion above, an interesting fact is that the conditional probability of arrested in a stop with the reason Probation or Parole of female is 50.00%, which is much higher than that of male, which is 20.11%.

Conclusion