CRIME ANALYTICS: BOSTON, MA

MAY 19, 2019

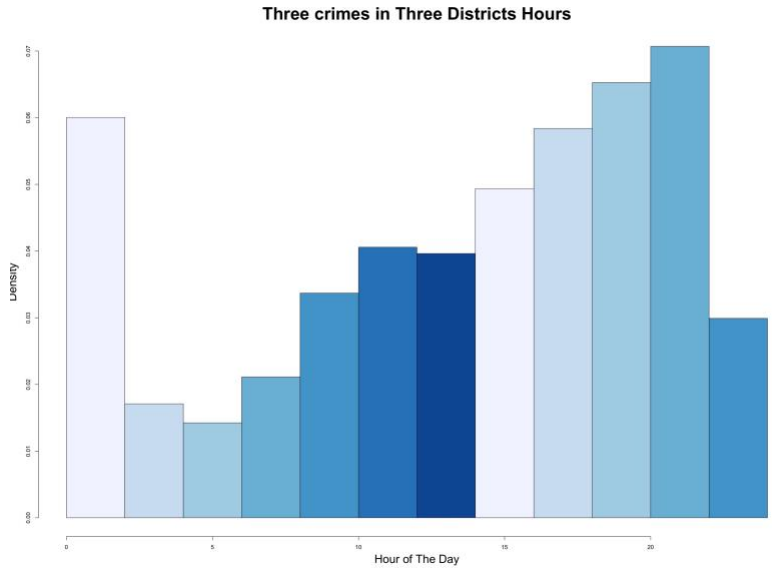
MATH 345

Mikhail Khayretdinov

What shift schedule law enforcement agencies could employ as the most efficient response to the temporal patterns of crime in given districts?

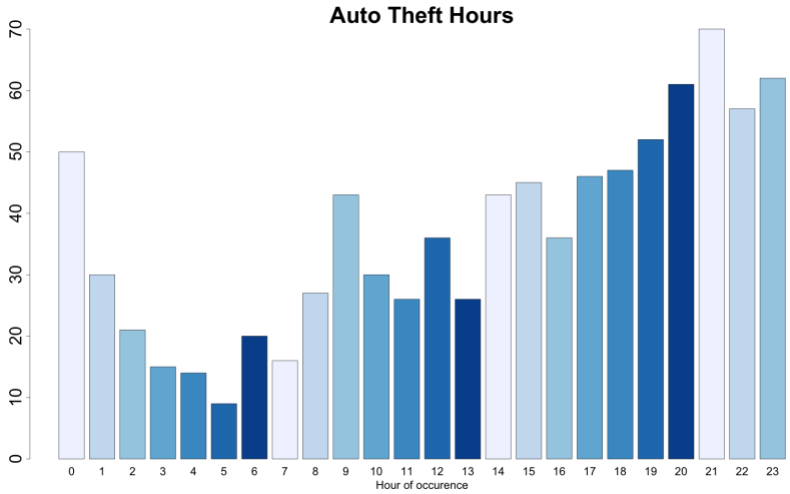
Temporal analysis of crime data can help in defining the patterns of criminal behaviors as a function of time. The resulting analytics can be applied in development of crime preventive tactics as well as advanced management of law enforcement agencies leading to an increase of its efficiency. Crime data analysis is the backbone of the emerging field of predictive policing, which already employs some algorithms that combine temporal, spatial and environmental context and that has already demonstrated up to 50% increase in crime detection in some cities. While most of these algorithms are beyond the scope of our course, we can still employ methods learned during this semester to detect certain temporal patterns of criminal behavior and to draw some useful conclusions.

Let's start by verifying that hourly analysis is actually feasible with the dataset under consideration. To do this we could examine all the three crimes in all the three districts and check if any trends can be spotted. Histogram can provide the required visualization:



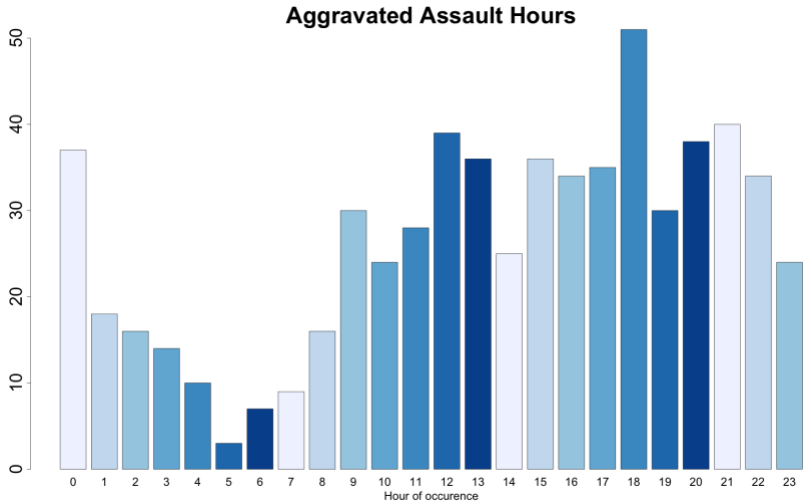
The resulting histogram clearly indicates sufficient change in the frequency distribution depending on the hour of the day with the general trend of crime frequencies increasing with the progression of the day, making early mornings the safest time and late evening the most dangerous time of the day.

We might further want to verify if the trend persists for each type of crime separately, starting with the distribution of time for auto thefts in all three districts:



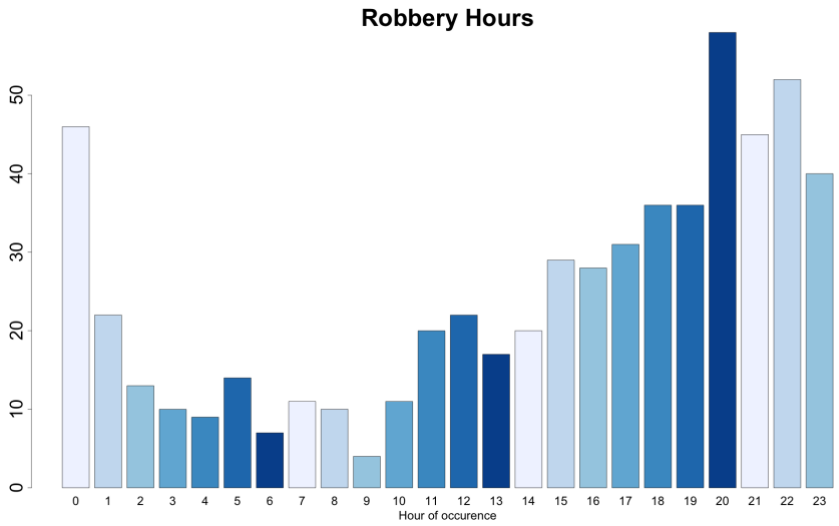
This bar plot demonstrates that an hour of the day has a strong impact on the chances of occurrence of an auto theft, with overall 9 cars stolen around 5 a.m., compared to 70 cars at 9 p.m. Most likely a car theft will occur in the evening time, with the ~45% of all auto thefts happened between 18.00 and 00.00 inclusively, peaking at 21.00.

Distribution of time of aggravated assaults in all three districts:



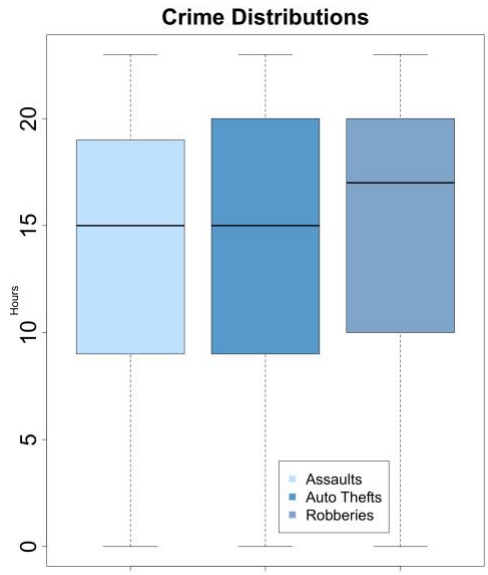
Clearly, the time of the day has also a major impact on the rate of aggravated assaults happening in all three districts, with 51 assaults occurred around 18.00, compared to 3 happened at 5.00 in the morning. Overall assaults are most likely to happen during the day, with ~58% of all of them happened between 9.00 and 18.00 inclusively.

Distribution of time of robberies in all three districts:



Robberies are also affected by time, with evening hours demonstrating the most likelihood of a robbery to occur. In this time of the day ~52% of them took place between 18.00 and 00.00 inclusively, peaking around 20.00 with 58 robberies. Morning hours are the safest, with only 4 robberies happened around 9.00 a.m.

While there are some minor differences in specific hours for each crime, the general pattern of growth from morning to evening persists, granting quite similar distribution for all three crimes with the median in the afternoon around 15.00 or 17.00 and evening hours being denser with criminal activity compared to early morning hours. Notable is that the interquartile range for all three types of crime corresponds to the usual work schedule from around 9 am to 19 pm, so half of all crimes happen during the regular business hours.

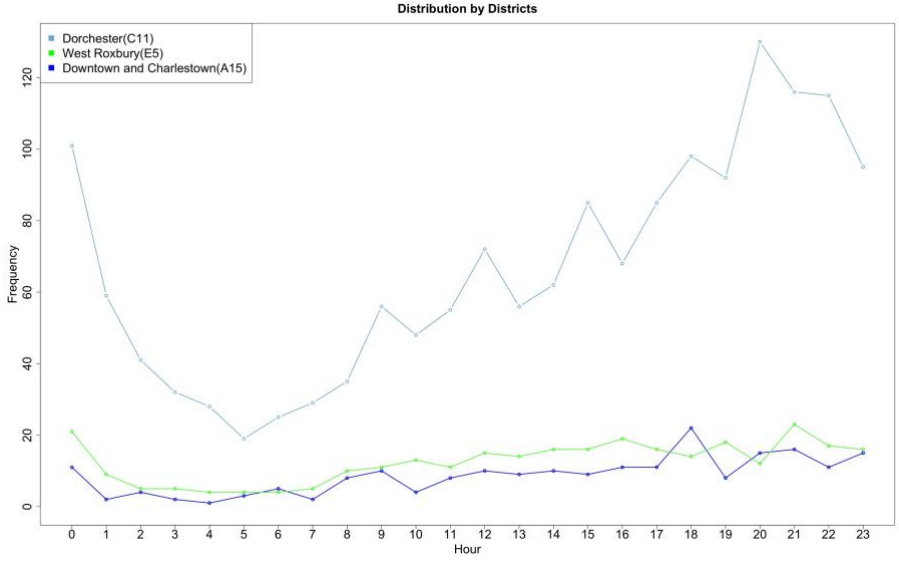


With a closer look at the criminal statistics in each separate district it becomes evident, that only a quarter of all crimes happen from midnight to 9.00. Half of the crimes occur by 16.00 and another half happening in the last 8 hours, making this shift hours the most intense.

Summary table for crime frequencies, percent and cumulative percent in each selected district

Time	Dorchester (C11)			West Roxbury (E5)			Dntn. & Charlestown (A15)		
	Freq	%	∑%	Freq	%	∑%	Freq	%	∑%
00.00	101	6.3%	6.3%	21	7%	7%	11	5.3%	5.3%
01.00	59	3.6%	9.9%	9	3%	10%	2	0.9%	6.2%
02.00	41	2.5%	12.5%	5	1.6%	11.7%	4	1.9%	8.2%
03.00	32	1.9%	14.5%	5	1.6%	13.4%	2	0.9%	9.2%
04.00	28	1.7%	16.3%	4	1.3%	14.7%	1	0.5%	9.6%
05.00	19	1.2%	17.4%	4	1.3%	16.1%	3	1.4%	11.1%
06.00	25	1.6%	19%	4	1.3%	17.4%	5	2.4%	13.5%
07.00	29	1.8%	20.8%	5	1.6%	19.1%	2	0.9%	14.5%
08.00	35	2.2%	23%	10	3.3%	22.4%	8	3.8%	18.3%
09.00	56	3.5%	**26.5%**	11	3.7%	**26.1%**	10	4.8%	23.1%
10.00	48	3%	29.5%	13	4.3%	30.5%	4	1.9%	**25.1%**
11.00	55	3.4%	32.9%	11	3.7%	34.2%	8	3.8%	28.9%
12.00	72	4.5%	37.4%	15	5%	39.2%	10	4.8%	33.8%
13.00	56	3.5%	40.9%	14	4.7%	43.9%	9	4.3%	38.1%
14.00	62	3.8%	44.8%	16	5.3%	**49.3%**	10	4.8%	42.9%
15.00	85	5.3%	**50.1%**	16	5.3%	54.7%	9	4.3%	47.3%
16.00	68	4.2%	54.3%	19	6.3%	61%	11	5.3%	**52.6%**
17.00	85	5.3%	59.6%	16	5.3%	66.4%	11	5.3%	57.9%
18.00	98	6.1%	65.7%	14	4.7%	71.1%	22	10.6%	68.5%
19.00	92	5.7%	**71.5%**	18	6%	**77.1%**	8	3.8%	**72.5%**
20.00	130	8.1%	79.6%	12	4%	81.2%	15	7.2%	79.7%
21.00	116	7.2%	86.9%	23	7.7%	88.9%	16	7.7%	87.4%
22.00	115	7.1%	94%	17	5.7%	94.6%	11	5.3%	92.8%
23.00	95	6%	100%	16	5.4%	100%	15	7.2%	100%

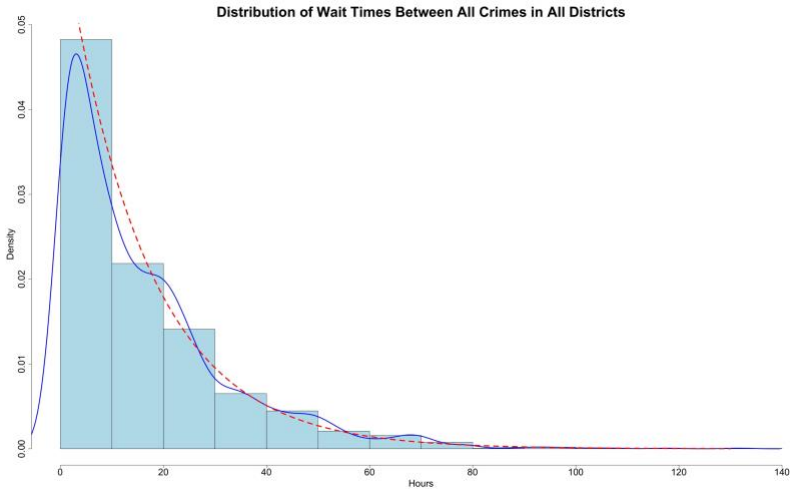
All Criminal Records for Each District Separately:



Slight variation in time patterns among the three districts is insignificant, what eloquent is the discrepancy between the frequency of occurrence of crimes in Dorchester compared to West Roxbury or Downtown & Charlestown. It might be worthwhile to examine this situation in more details by comparing the waiting times between the crimes happening in each district. But first let's check the overall distribution of wait times between crimes in all districts.

Five-Number + Mean and Standard Deviation Summary for the Wait Times Between All Crimes in All Three Districts

Sample Minimum	1-st Quartile	Median	3-rd Quartile	Sample Maximum	Mean	Standard Deviation
0.0	3.75	10.59	22.32	131.45	15.93	16.56



Since the most of the area under the density function is located near the origin, and the density function drops gradually as y increases, we can employ the gamma probability distribution function to approximate wait times between the crimes. In particular, exponential function could model this distribution fairly accurately as we observe the good fit of exponential density curve to the data histogram. We can also note that mean and standard deviation of the dataset are almost similar, which is also a good indicator of exponential distribution.

$$f(y) = \begin{cases} \frac{1}{\beta} e^{-y/\beta}, & 0 \le y < \infty \\ 0, & elsewhere \end{cases}$$

Having this observation in mind we may now try to analyze the demand for safety and measures required in each district. For the purposes of our analysis, let's assume that the regular shift is equal to 8 hours and see what results would bear the exponential distribution if applied to each of the districts separately:

Five-Number + Mean and Standard Deviation Summary
for the Wait Times in Each District

District #	Sample Minimum	1-st Quartile	Median	3-rd Quartile	Sample Maximum	Mean	Standard Deviation
C 11	0.0	4.917	14.650	28.700	189.733	20.953	21.88
A15	0.0	42.00	95.98	191.90	956.98	160.26	185.72
E5	0.0	33.12	81.83	155.12	670.87	112.56	108.92

Probability of encountering a crime during a single shift:

District	Exponential Distribution	Resulting Probability
Dorchester (C11)	$\int_0^8 \frac{1}{20.953} e^{-y/20.953} dy$	0.317
Downtown and Charlestown (A15)	$\int_0^8 \frac{1}{160.26} e^{-y/160.26} dy$	0.049
West Roxbury (E5)	$\int_0^8 \frac{1}{112.56} e^{-y/112.56} dy$	0.069

The probability of crime happening during the 8-hour shift varies significantly in each of the districts. From the obtained results we can see that the demand in Dorchester is much higher compared to other districts, specifically it requires sevenfold amount of law enforcement units compared to Downtown & Charlestown and fivefold compared to West Roxbury. So, if we were to assign 100 patrol units among these three areas, the distribution would have to be the following: 73 units to Dorchester, 11 to Downtown and Charlestown and 16 to West Roxbury. Furthermore, half of all units in each group would have to take a shift from 16.00 to 00.00 to fully meet the demand, another 30% of the units would have to take shift during the day from 8.00 to 16.00, while the remaining 20% would take the night shift between 00.00 and 08.00.

Units Assigned In Districts During Each Shift

Shift	Units in Dorchester	Units in West Roxbury	Units in D&C*
00.00 - 08.00	14	3	2
08.00 - 16.00	22	5	3
16.00 - 00.00	37	8	6

*Downtown and Charlestown