CRIME ANALYTICS: BOSTON, MA

MAY 19, 2019

Gursimar Singh, Igli Ajdni, Daniel Werminghausen, Erald Dervishi, Mikhail Khayretdinov

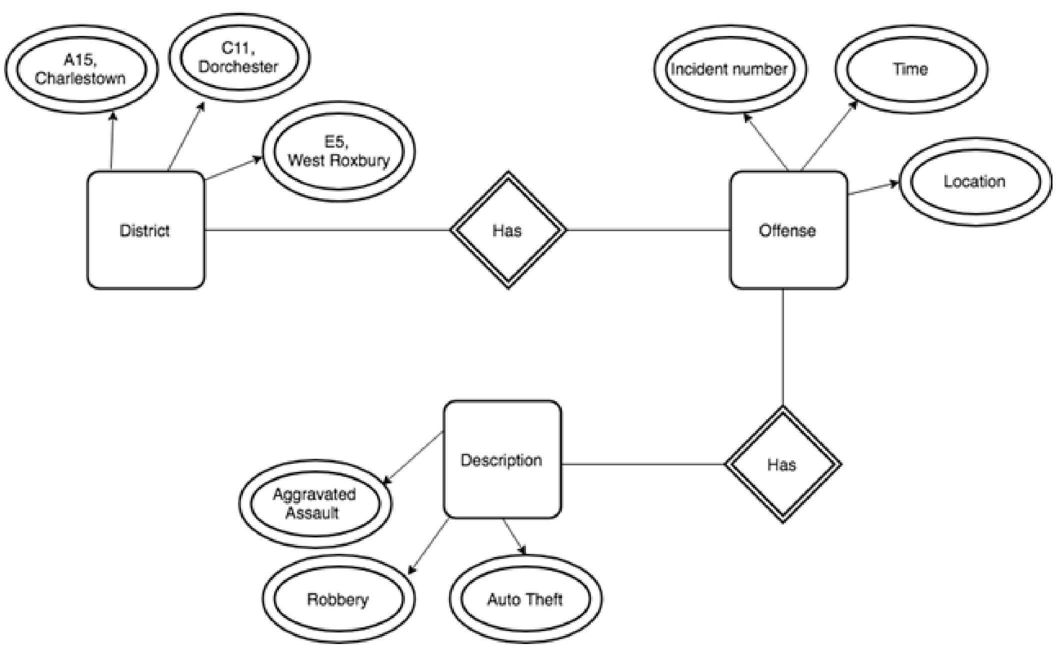
Table of Contents

Introduction:

Using statistical methods learned in class, we aim to compare crime rates in different neighborhoods in Boston, MA. Through our analysis we aim to calculate the probability of a crime happening next in neighborhoods- Charlestown, Dorchester and West Roxbury. By comparing and analyzing crime rates of aggravated assaults, auto theft and robbery we aim to provide a comparison between the three.

Data Description:

The data was collected from data.boston.gov, featuring the crime incident reports provided by the Boston Police Department from the years 2015-present. From there, we alter the dataset to focus on three districts- A15 (Charlestown), C11 (Dorchester), and E5 (West Roxbury). Furthermore, focusing on three crimes- aggravated assaults, auto theft and robbery, happening in those districts.



The figure above better explains how the dataset used for the analysis was set up. There are three main entities that we look at: district, offense and description. The three entities are connected such that, each district has an offense occurring and for each offense there is a description along with it. Each of the three entities has some attributes along with it. The district entity contains three districts, offense contains the unique identifier incident number, time and location. Description contains the type of crime that was committed, in this case we only focus on three.

Research Questions:

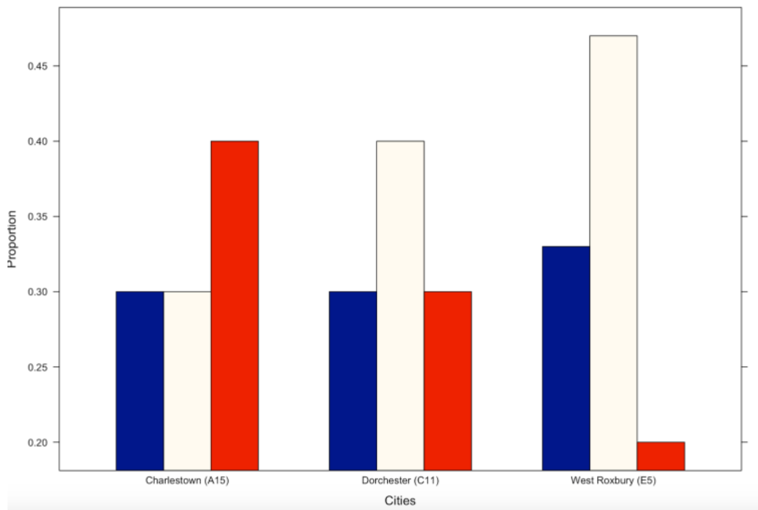
1. What is the most popular crime in each neighborhood and how do the neighborhoods compare against each other in terms of aggravated assault occurences?
2. What is the probability that randomly selected individuals can be a victim of a crime (Auto theft, Robbery, Aggravated Assault) two or more time during 2015-2019 in West Roxbury?
3. What is the chance of walking by someone living in this district having been a victim of one of the 3 crimes in 2015 - 2019 with the notion that each crime is independent per victim?
4. Given the probability of a particular crime in district C11, if a crime occurs today, what are the chances that another crime will occur in a week or more?
5. What shift schedule law enforcement agencies could employ as the most efficient response to the temporal patterns of crime in given districts?

Section 1: What is the relative frequency of crime, by crime type and by neighborhood?

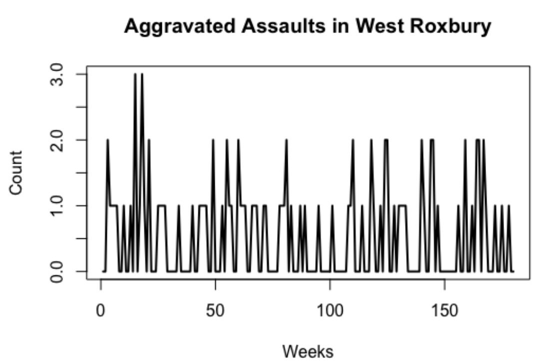
The table below compares the crime occurrences in the three districts. From the data we find that in West Roxbury there are more Auto Thefts than there are Aggravated Assaults or Robberies. In Dorchester we also have more Auto Thefts**,** while in Charlestown the most frequent crime are Robberies.

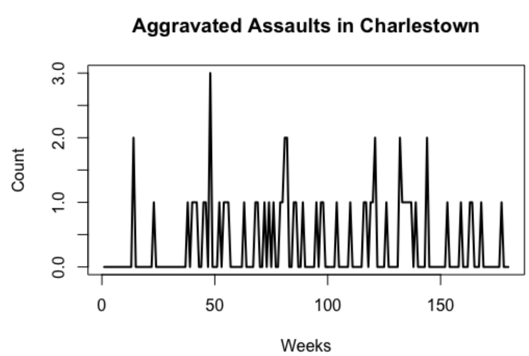
West Roxbury (E5)			
Crime type	**Auto Theft**	**Aggravated Assault**	**Robbery**
Proportion	0.47	0.33	0.2
Dorchester (C11)			
Proportion	0.4	0.3	0.3
Charlestown (A15)			
Proportion	0.3	0.3	0.4

The histogram below visualizes the table above. Blue represents aggravated assault, white is for auto theft, and red is for robbery. As we see, the most common crime in Charlestown is Robbery. In both, West Roxbury and Dorchester the most common crime is Auto Theft. We note that in both, Charlestown and Dorchester the proportion of aggravated assaults is similar, but is lower than what it is in West Roxbury.

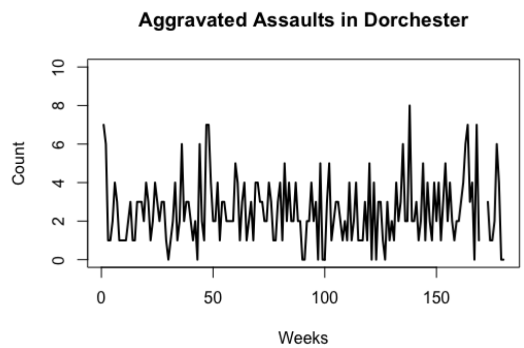


To analyze the weekly number of aggravated assaults happening in all three districts the first thing we need to do is separate the data into weeks (we get ~180 weeks). Next, we put in the numbers of crimes happening in each week, in all three districts and visualize using line graphs.

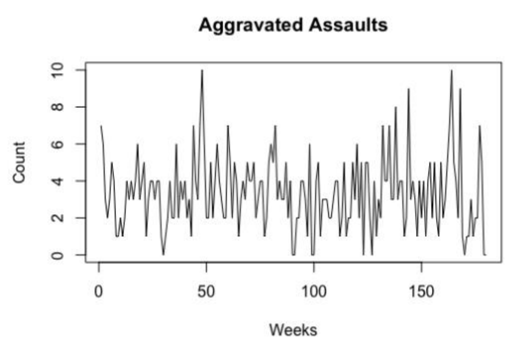




From the graphs above we can see that the most Aggravated Assaults that happened in West Roxbury in these 4 years in one week is 3, and it only happened twice. Most of the times, the count per week is a 0 or a 1. The total count of Aggravated Assaults in West Roxbury happening in these 180 weeks is 95. We note that Charlestown and West Roxbury share similar patterns, with the highest 3 assaults being the highest in one week and most frequently it also been a 0 or a 1. The total count of Aggravated Assaults in Charlestown happening in these 180 weeks is 59.



On the other hand, Dorchester as we can see its different, the highest in one week is 8 assaults and out of the 180 weeks, only in 10 weeks we've had **0** Aggravated Assaults. The total count of Aggravated Assaults in this district is 450.



Finally, we combine all three districts in one to show the total aggravated assaults. We can see the highest count being a 10 which happened on the fourth week of July 2016 and on the fourth week of December 2018.

Section 2: Probability of being a victim of a crime 2 or more times?

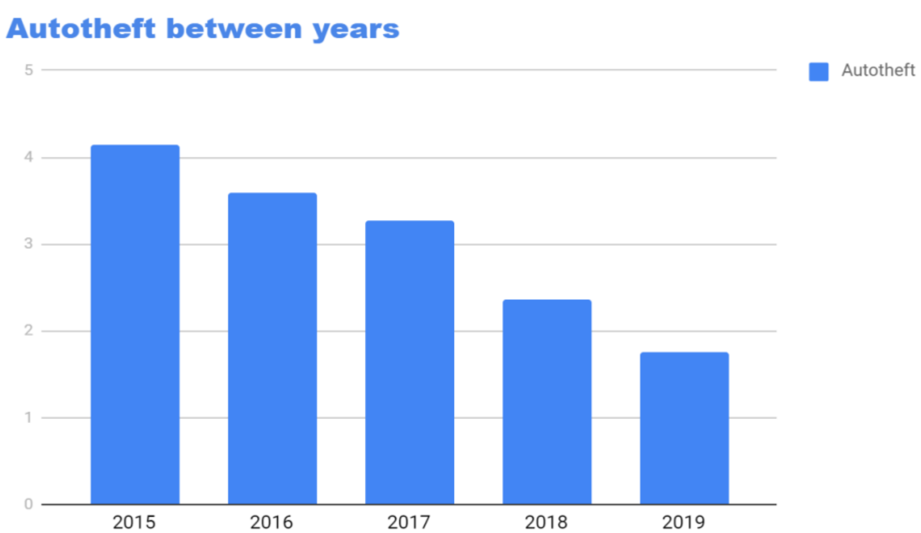
It might be useful to be able to predict the number of crimes potentially happening in a particular district. For this specific research question, I will be taking a closer look at the district of West Roxbury. There are a total of 30,446 people living in West Roxbury. For the data we have, there have been 141 reported cases of auto theft, 95 aggravated assaults and 62 robberies. In order to find the rate per month, we simply divide the number of occurrences by the total number of months (47).

Taking a look at the 47 months we find:

Rate auto theft	141/47 = 3.00
Rate Aggr Assault	9547= 2.0212
Rate robberies	6247= 1.3191

The table below lists the average number of auto thefts per month per year and is visualized using the histogram below

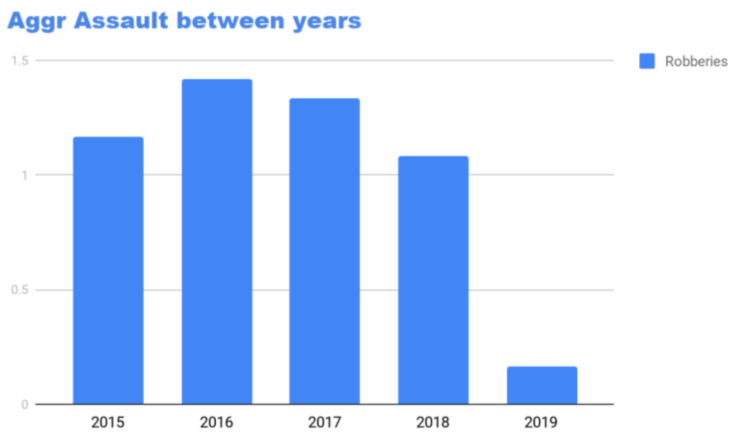
YEAR	2015	2016	2017	2018	2019
AVERAGE	4.142857	3.583333	3.272727	2.363636	1.750000



As we can see, the graph clearly shows a decline in auto theft over the span of 2015-2019. It has turned to less than half the amount of auto thefts per month in 2015. This shows a decline in auto theft and the beginning of a safer district districts of Boston

The table below lists the average number of aggravated Assault per month in different years and is visualized using the histogram below

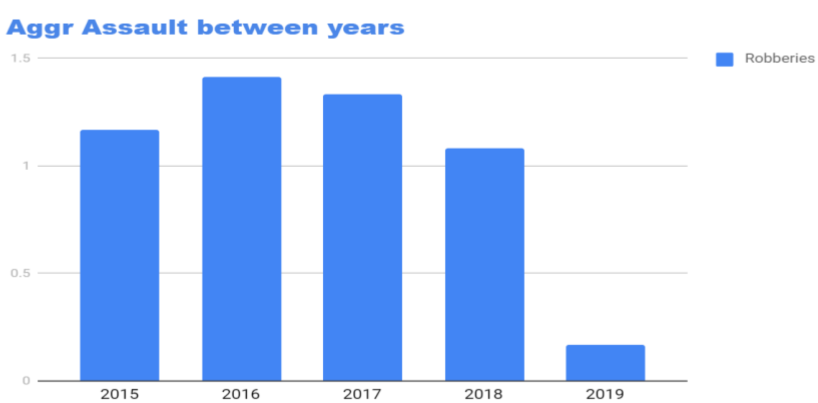
YEAR	2015	2016	2017	2018	2019
AVERAGE	1.3333333	2.2500000	1.4166667	2.2500000	0.6666667



We can see that in 2016 there were the most number of assaults. We se that between 2016, 2017 and 2018, the assaults had a downward trend. For 2015 and 2019 we can’t say much since the data between those years is partially available.

The table below lists the average number of robberies per year and is visualized using the histogram below.

YEAR	**2015**	**2016**	**2017**	**2018**	**2019**
AVERAGE	1.1666667	1.4166667	1.3333333	1.0833333	0.1666667



Using Poisson distribution we can find the probability of certain number of crimes to happen within a month. The poisson distribution is given by the formula:

P(Y=y) = e-λλyy!, y N.

We find that the average number of auto thefts per month is 3, aggravated assault is 2.02 and robberies 1.32 per month.

What is the probability of 2 or more crimes to happen in West Roxbury within a month?

P(X2) = 1-[P(X=0) + P(X=1)]
Formula = 1- [e-λ+(e-λ*λ)]
λ= 3+2.0212+1.3191**=6.3303**
λ=6.3303
P(X2) = 1-[P(X=0) + P(X=1)]

Formula = 1- [e-6.3303+(e-6.3303*6.3303] = 0.9869

What is probability of 1 to 5 crimes to happen in West Roxbury within a month?

dpois(1, 6.33) = 0.0112
dpois(2, 6.33) = 0.0357
dpois(3, 6.33) = 0.0753
dpois(4, 6.33) = 0.1192
dpois(5, 6.33) = 0.1509

Conclusion:

In conclusion, I used poisson distribution for this part of our project because it's the most effective way to answer my research question. We have to keep in mind that data victims of each crime are independent. In our data we do not have names individuals, which means that one event does not affect the probability of the other event when it will occur. *"The Poisson distribution is used to model the number of events occurring in a given time period, geographical area, etc(Chapter 3: Discrete Probability Distributions,p29)*". The dataset that we are analyzing, fulfill the requirements needed to use poisson distribution because in our data we have number of events that are independent. We have a time period and graphical areas and districts that these crimes happen in our dataset. To conclude my findings, we can tell that crimes are sufficiently decreased over the years for the particular three crimes listed. In West Roxbury we can see that the probability of crime happening per month is greater than 2 crimes. This is not a lot of crime but this is only looking at robbery, aggravated assault and auto theft.

Completing this project I found interesting information’s regarding the topic. One of them is how these crimes change during the years from 2015 to 2019. As mentioned previously one can see through the histograms how these crimes change throughout the years. Through this project, I learned how to analyze dataset given, but the most important part for me that I believe is going to help me in the future is how to implement these datasets with R language and use those in code. It was fascinating due to the fact that R language is a new statistical programming language for me. Last but not least, the importance of connecting what we learned during the semester with our dataset was how to use the right distribution or formula to tackle the questions given as assignments by applying distributions and tactics learned.

Section 3: Victims of crime and the chance of meeting one or more victims

Walk through a neighborhood in Boston and think about all the people you pass on the street. Some have been a victim of certain types of crime. What are the chances of passing people in neighborhood of Charlestown, Dorchester, and West Roxbury who has been a victim of the three crimes - aggravated assault, robbery? Since our data does not have any victim names listed, we have to assume every reported crime is independent and that the entire population of a neighborhood is in the lines of the neighborhood's location. Table below lists the district's in question and their respective population size.

District	**Population**
Charlestown (A15)	18,901
Dorchester (C11)	87,585
West Roxbury (E5)	30,446

The table below lists the number of crimes in each district over 47 months of data (about 4 years)

District	**Auto Theft**	**Aggravated Assault**	**Robbery**
Dorchester (C11)	649	479	474
Charlestown (A15)	91	60	55
West Roxbury (E5)	141	94	61

In order to calculate the rate per person at which the crimes happen in each of the three districts, we look at the total number of crimes over the period of 47 months, and divide that by the population size for each district. Equation:

$$rate\ per\ person = \frac{total\ number\ of\ crimes}{population\ of\ the\ district}$$

District	Total number of crimes	Rate calculation	Rate per person
C11	1602	$\frac{1602}{87,585}$	0.0182908032
A15	206	$\frac{206}{18901}$	0.010898894
E5	296	$\frac{296}{30,446}$	0.00972213099

Using Poisson to find p, the probability of success that someone is a victim of one of these three crimes

Poisson Distribution **=** $\mathbf{e}^{-\lambda}\ \mathbf{(}\frac{\lambda^{-y}}{\lambda!}\mathbf{)}$

Each rate of crime for each district gets placed as lambda (□) to find the probability P of success

We find:

$\lambda_{Charlestown}$ = 0.0109

$\lambda_{Dorchester}$ = 0.0183

$\lambda_{West\ Roxbury}$ = 0.0097

To find P (probability) using formula $1 - e^{-\lambda}$

Dorchester = $1\text{-}e^{-0.0182908032}$ = 0.0181

Charlestown = $1\text{-}e^{-0.010898894}$ = 0.10839

West Roxbury = $1\text{-}e^{-0.00972213099}$ = 0.00975

Now, we can use a geometric distribution to model the chance of walking by someone living in the same district also having been a victim of one of the three crimes.

Geometric distribution is given by the formula: $f(x) = (1 - p)^{N-1}p$

Where P (X = 0), P (X = 1), P (X = 2), ………...P (X = N)

Meeting 1-10 people affiliated as a victim of the 3 types of crime 2015-2019:

As a test I test the P (2), P (5), P (10) to see the change in chance of meeting more than one victim. As the results show, the probability decreases as the more people you pass. We have to keep in mind that the probability is affected by the population and each district has its induvial small or large population size.

<u>**Charlestown**</u>**(A15)** = $f(x) = (1 - p)^{N-1}$ (p)

Sample Calculations for P(Y=2), & P(Y=5) & P(Y=10)

$P(X=2) = P(2) = (1 - 0.0108)^{2-1}(0.0108)$	= 0.01072 x 100 = 1.072%
$P(X=5) = P(5) = (1 - 0.0108)^{5-1}(0.0108)$	= 0.01072 x 100 = 1.072%
$P(X=10) = P(5) = (1 - 0.0108)^{10-1}(0.0108)$	= 0.0097 x 100 = 0.97%

<u>**Dorchester**</u>**(C11)** = $f(x) = (1 - p)^{N-1}$ (p)

Sample Calculations for P(Y=2), & P(Y=5) & P(Y=10)

$P(X=2) = P(2) = (1 - 0.018)^{2-1}(0.018)$	= 0.01747 x 100 = 1.747%
$P(X=5 = P(5) = (1 - 0.018)^{5-1}(0.018)$	= 0.01654 x 100 = 1.654%
$P(X=2) = P(10) = (1 - 0.018)^{10-1}(0.018)$	= 0.015 x 100 = 1.5%

<u>**West Roxbury**</u> **(E5)** = $f(x) = (1 - p)^{N-1}$ (p)

Sample Calculations for P(Y=2), & P(Y=5) & P(Y=10)

$P(X=2) = P(2) = (1 - 0.00975)^{2-1}(0.00975)$	= 0.00956 x 100 = 0.956%
$P(X=5) = P(5) = (1 - 0.00975)^{5-1}(0.00975)$	= 0.00928 x 100 = 0.928%
$P(X=10) = P(10) = (1 - 0.00975)^{10 - 1}(0.00975)$	= 0.0088 x 100 = 0.88%

How sure can you be that you have passed a victim of one of the three crimes in a given district. To calculate the probability of being 60% certain in Charlestown, 70% in Dorchester, and 80% in West Roxbury, we use the formula P(Y>=y) = q_y,

Being 60% sure that you have walked past a victim of one of the three crimes in Charlestown:

Mean = $\frac{1}{p} = \frac{10000000000}{975023859} = \frac{975023859}{10000000000}$ = 102.56

Variance = $\frac{1-p}{p^2} = \frac{1-0.00975}{(0.00975)^2} = 10416.32$

Standard deviation = $\sqrt{\frac{1-\mathrm{p}}{\mathrm{p}^2}} = \sqrt{\frac{1-0.00975}{0.00975^2}}$ = 102.06

In order to be 60% certain to have passed a victim of crime in the given district we need to have passed at least 47 people in Charlestown: P(X>=47) = 0.60 = 60% (47 trails)

Being 70% sure that you have walked past a victim of one of the three crimes in Dorchester:

Mean = $\frac{1}{p} = \frac{10000000000}{10839716} = \frac{10839716}{10000000000}$ = 92.25

Variance $= \frac{1-p}{p^2} = \frac{1-0.018}{(0.018)^2} = 8418.425$

Standard deviation = $\sqrt{\frac{1-\mathrm{p}}{\mathrm{p}^2}} = \sqrt{\frac{1-0.018}{0.018^2}}$ = 91.75

In order to be 70% certain to have passed a victim of crime in the given district we need to have passed at least 20 people in Dorchester: P(X>= 20) = 0.70 = 70% (20 trails)

The number of persons it takes to be 80% sure to have walked past a person affected by crime

Being 80% sure that you have walked past a victim of one of the three crimes in West Roxbury:

Mean = $\frac{1}{p} = \frac{10000000000}{975023859} = \frac{975023859}{10000000000}$ = 102.56

Variance = $\frac{1-p}{p^2} = \frac{1-0.00975}{(0.00975)^2} = 10416.32$

Standard deviation = $\sqrt{\frac{1-\text{p}}{\text{p\^{}2}}} = \sqrt{\frac{1-0.00975}{0.00975\text{\^{}}2}}$ = 102.06

In order to be 80% certain to have passed a victim of crime in the given district we need to have passed at least 23 people in West Roxbury: P(X>=23) = 0.80 = 80% (23 trials)

Average # of people to pass in districts to pass a victim of one of three crimes:

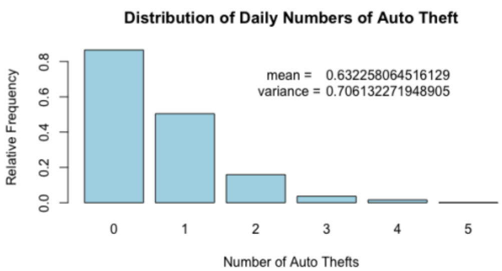
District	**Calculations**	**Average # of people**
Charlestown	0.01083 = 1.083%, 100/1.083	92
Dorchester	0.01812 = 1.812%, 100/1.812	55
West Roxbury	0.00975 = 0.975%, so 100/0.975	103

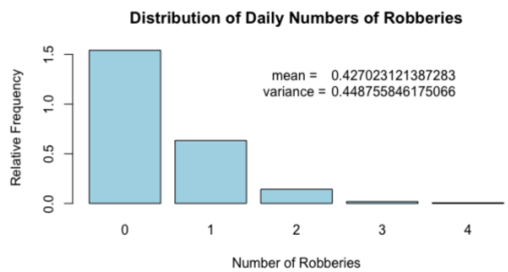
In Charlestown 19,901 people live and there are 206 reports of crime for the particular crimes of auto-theft, aggravated assault, and robbery. The chance of walking by a person on the street who has been a victim of one of these three crimes from (2015 - 2019) is one in every 92 people living in the district of Charlestown are a victim of one of the three crimes listed. In Dorchester there are 87,585 people living and 1,602 crimes reported for the particular three crimes. The chance of walking by a person who has been a victim of one of these three crimes from (2015 - 2019), is about one in every 55 people are victims in Dorchester. In West Roxbury there are 30,446 people living and 296 reports for the particular three crimes listed above. The chance of walking by a person who has been a victim of one of the three crimes from (2015 - 2019) is one in every 103 people in West Roxbury are victims of one of the three crimes.

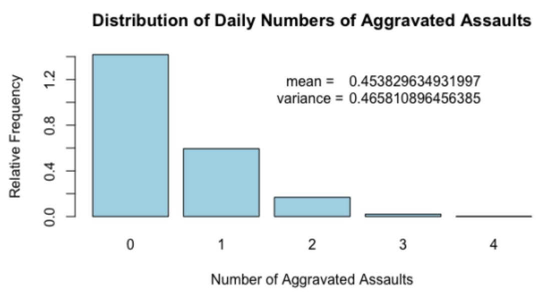
West Roxbury is the safest district out of the three according to the particular three crimes chosen. With geometric distribution we can see that meeting more than one victim in a specific district decreases the probability of meeting multiple victims. Using geometric distribution, we can check how sure we are to have passed a person who has been a victim by being sure of a certain amount of percentage of having passed a victim.

Section 4: Probability of a crime happening again a week or more later

To model the probability of crimes happening again in a week we would need to first look at the distribution of daily number of crimes happening.







The figures above show the relative frequencies of the three crimes, auto theft, robberies and aggravated assaults. We note that for all three crimes the mean and variance are relatively close. For instance, for aggravated assaults the mean is ~0.4538 and the variance is 0.4685. This relative closeness suggests that a Poisson distribution would be a good probability distribution model for daily counts of crime.

In order to calculate the rate at which the crimes happen we need to first, get the total amount of crimes reported. In this case, there were 474 robberies, 479 assaults, and 649 auto thefts. Now, to answer the question we need to look at the rate at which each of the crimes happen per day. The data we have is from June 2015 to April 2019. In order to find out how many crimes happen per month, we simply divide the total number of that crime with the total number of months. After we derive the rate per month we would divide it by 31 in order to get the rate per day. We find that there are 0.3253 robberies per day, 0.3287 assaults and 0.4454 auto thefts per day in district C11 (Dorchester).

With the per day rate calculated we can use a Poisson distribution to model the probability distribution to answer the question as to what is the probability that the next crime occurrence will be a week or more later. For poisson, we will take the λ = rate per day for the crime * 7. For the three crimes we find that:

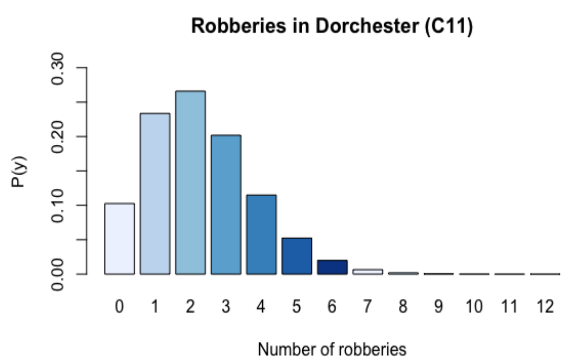
$$\lambda_{robberies} = 0.3253 * 7 = 2.2771$$

$$\lambda_{assaults} = 0.3287 * 7 = 2.3009$$

$$\lambda_{auto\ theft} = 0.4454 * 7 = 3.1178$$

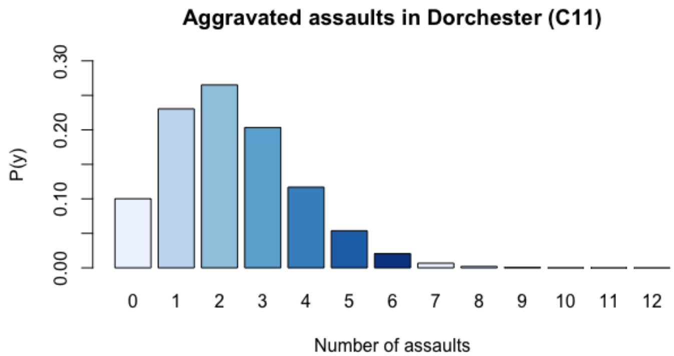
$$\text{Poisson Distribution} = e^{-\lambda}\left(\frac{\lambda^{-y}}{y!}\right)$$

P(robbery elapsing a week or more later) = $e^{-2.2771}\left(\frac{2.2771^{-y}}{y!}\right)$



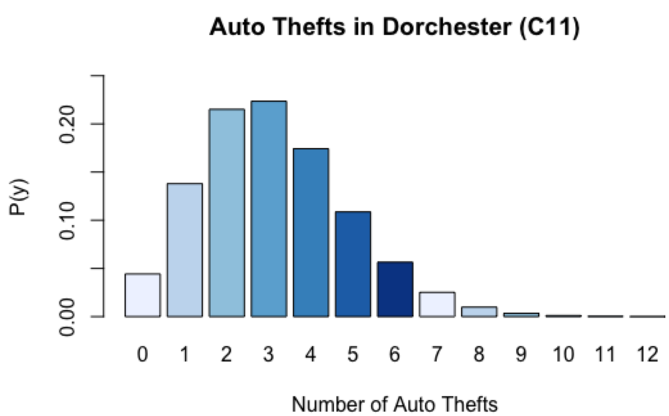
From the above figure, we see that in a week, the most likely number of robberies is 2 with a probability of ~0.2659. That is followed in closely with the probability of 1 robbery with probability ~0.2335 and 3 robberies with probability 0.2018.

P(another assault occurring a week or more later) = $e^{2.3009}\left(\frac{2.3009^{-y}}{y!}\right)$



The figure above shows that the most likely number of assaults happening in a week is 2 with the probability ~0.2651. In addition, the probability of there being 1 assault is ~0.2304 and 3 assaults is ~0.2033.

P(another auto theft occurring a week or more later) = $e^{3.1178}\,(\frac{3.1178^{-y}}{y!})$

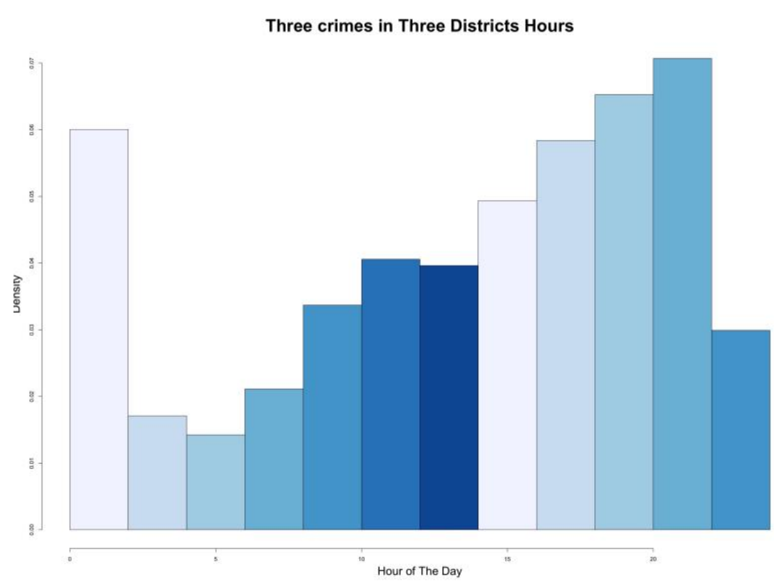


The figure above shows that the most likely number of auto thefts happening in a week is 3 with a probability of ~0.2235. In addition, the probability for 2 auto thefts is ~0.2150 and the probability for 4 auto thefts ~0.1742.

Section 5. What shift schedule law enforcement agencies could employ as the most efficient response to the temporal patterns of crime in given districts?

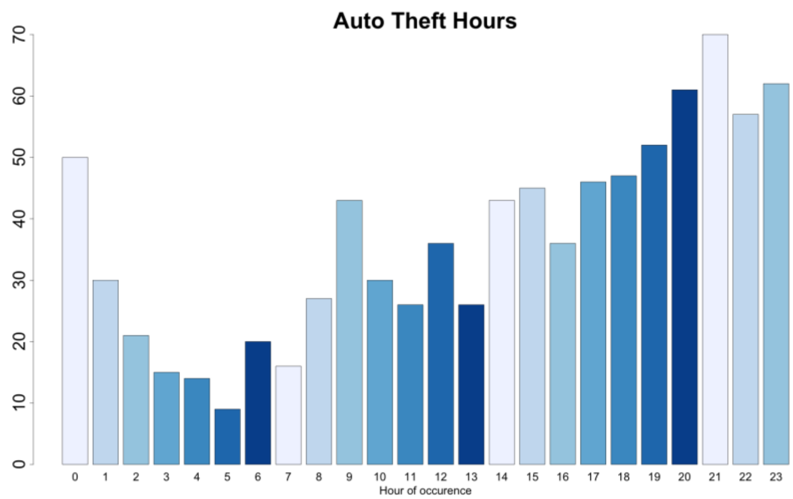
Temporal analysis of crime data can help in defining the patterns of criminal behaviors as a function of time. The resulting analytics can be applied in development of crime preventive tactics as well as advanced management of law enforcement agencies leading to an increase of its efficiency. Crime data 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:



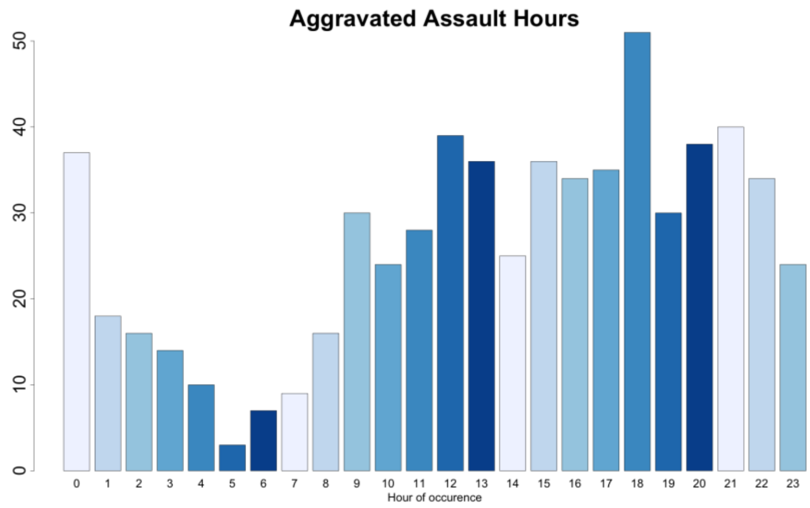
The resulting histogram clearly indicates sufficient change in the frequency distribution depending on the hour of the day with the general trend of crime frequencies increasing with the progression of the day, making early mornings the safest time and late evening the most dangerous time of the day.

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:



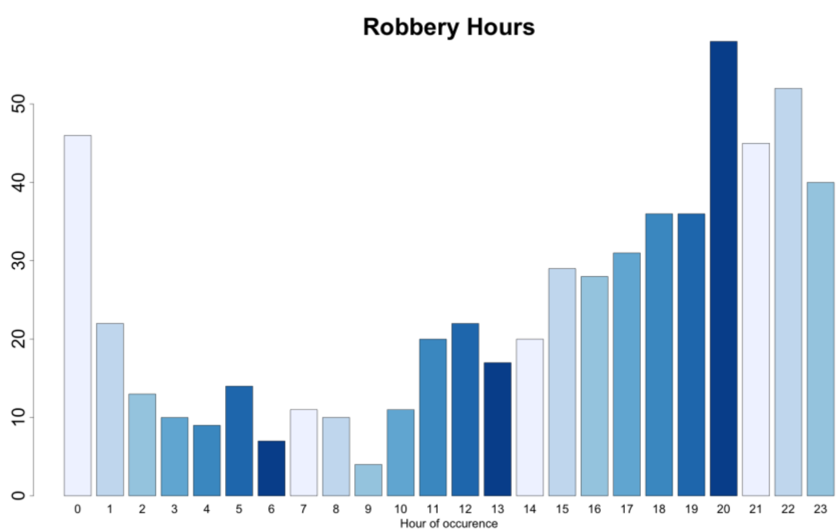
This bar plot demonstrates that an hour of the day has a strong impact on the chances of occurrence of an auto theft, with overall 9 cars stolen around 5 a.m., compared to 70 cars at 9 p.m. Most likely a car theft will occur in the evening time, with the ~45% of all auto thefts happened between 18.00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

Distribution of time of **aggravated assaults** in all three districts:



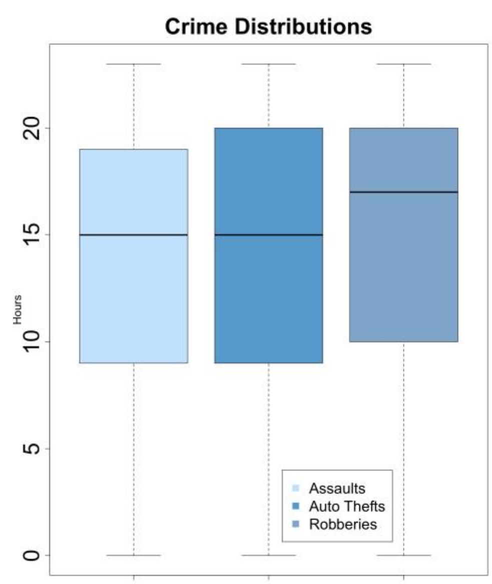
Clearly, the time of the day has also a major impact on the rate of aggravated assaults happening in all three districts, with 51 assaults occurred around 18.00, compared to 3 happened at 5.00 in the morning. Overall assaults are most likely to happen during the day, with ~58% of all of them happened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



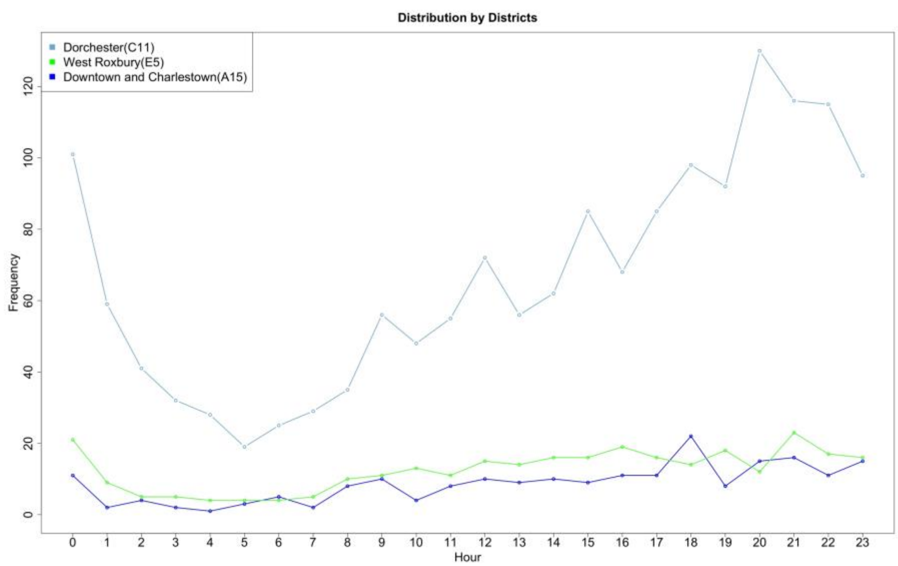
With a closer look at the criminal statistics in each separate district it becomes evident, that only a quarter of all crimes happen from midnight to 9.00. Half of the crimes occur by 16.00 and another half happening in the last 8 hours, making this shift hours the most intense.

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%

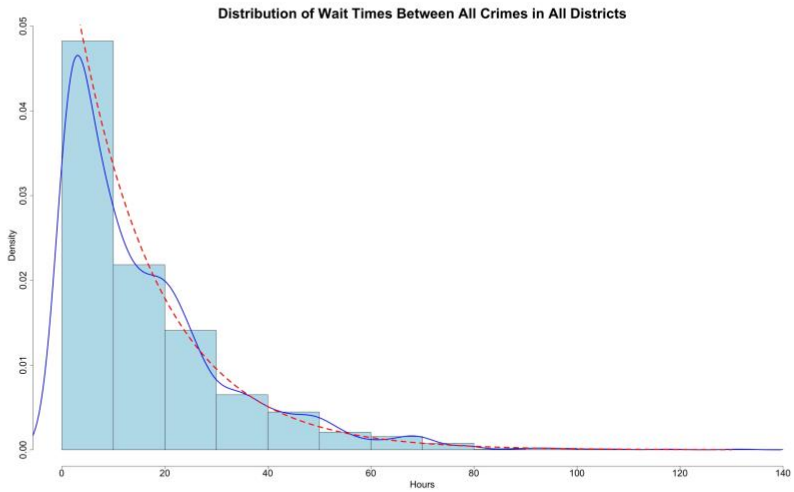
All Criminal Records for Each District Separately:



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.

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

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

Conclusion

In this project we demonstrated some ways in which statistics can be utilized for the purposes of crime analytics. Each part of the report targeted quite specific focus on the problem of lawbreaking which collectively builds into a fairly precise description of the crime situation in given districts in regard to the chosen set of crimes.

The first section grants a broad overlook on the given dataset and introduces the probability distributions of the selected crimes. The second section offers a standpoint of a victim, and proposes the statistically justified observation of chances of a regular person to experience a crime committed against him or her. The main tool of this section was the Poisson distribution, which seemed to be the most suitable approach in the framework of this question. In the third part of our report, the focus on victims of the crime was continued with the detailed statistical investigation of the chances of meeting a victim in each district. The main method employed in that part was the geometrical distribution, which corresponded with the goals of the question and was perfectly applicable. The fourth part of the project shifted focus from victims back to crimes, starting the analysis of crime occurrence in a particular timeframe, specifically targeting the chances of an offence to reoccur in a period of a week or more. And again, the Poisson distribution was the best fitting method to tackle this question. In the fifth part of our report the idea of temporal analysis was continued and the viewpoint of law enforcement agencies was introduced. In particular, this part covered some practical application of statistical analysis by modeling the patrolling shift schedule of a hypothetical fleet of police cars which would meet the demand for law enforcement in different districts defined by hours. Exponential distribution was the best fitting instrument in modeling the situation of this section.

Our project demonstrated just a fraction of possible applications of statistics in addressing the problem of crime, however it has already produced some unequivocal results that could have various use. For instance, the discovery of chances of being a victim in West Roxbury could help the real estate dealers to better asses the attractiveness of a property in this area. The distribution of crime probabilities could motivate city police to reassess their priorities in this location. The time distributions combined with chances to be a victim could help people in avoiding a crime, as well as to be a useful resource to law enforcement agencies in addressing the crime patterns more effectively by avoiding understaffing at peak times and overstaffing at safer hours.

Teamwork in this project

The teamwork started immediately after we were assigned the project and our group was formed. We started by collectively brainstorming the set of ideas and choosing a good topic with a suitable dataset. After selecting the theme of the project, we carefully went through the possible questions that would help us in discovering the dataset. In this phase we actively collaborated and helped each other to shape the questions of our sections and focus on appropriate probability distributions for the problems we chose to cover. Eventually each of us picked one section of the project and focused their effort on properly addressing in the report. Igli picked the first section, Erald was on the second one, Daniel was on third one, Gursimar did the fourth part and Mikhail did the fifth. We all collaborated via Google Docs, maintaining communication through messengers and Google Hangout and on meetings which happened both in and out of class.