CSDA 1010
Final Report
Group 10

Group 10 Members:
Michael Conyette
Dolly Gerrior
Xiaoman Lu
Bobby Muttathumankal Babu
Surin Singh

Contents

Business Understanding

Business Objective

Background:

Crime generates substantial costs to society at individual, community, and national levels. In the United States, more than 23 million criminal offenses were committed in 2007, resulting in approximately $15 billion in economic losses to the victims and $179 billion in government expenditures on police protection, judicial and legal activities, and corrections (U.S. Department of Justice, 2004). Programs that directly or indirectly prevent crime can therefore generate substantial economic benefits by reducing crime-related costs incurred by victims, communities, and the criminal justice system. Indirect losses suffered by crime victims, including pain and suffering, decreased quality of life, and psychological distress. Measuring losses across these four components provides an estimate of the economic cost of individual crimes. The broad societal perspective is appropriate for economic analysis and program evaluation because more narrow perspectives (e.g., crime victim, criminal justice agency, community organization) apply to specific stakeholders or agendas (McCollister, French, & Fang, 2012).

Canada unfortunately has not been spared the experience of these losses and costs. Over the last few years for example, there has been a dramatic increase in automobile thefts across the Greater Toronto Area (GTA). But both police and the Insurance Bureau of Canada (IBC) say owners are making it too easy for thieves to steal their vehicles. Toronto Police from the 53rd division in the Downtown core reveal a 92 per cent jump in claims for automobile theft compared to this time last year (Shum, 2018).

Fraud and theft are significantly contributing to the high automobile insurance premiums we see today. This is affecting the provincial economy and the cost of living of residents. Therefore, the need for an aggressive enforcement strategy to deal with automobile theft is growing every day (Quan, 2018).

Canada is home to the king of bike thieves. In 2008, a quirky Toronto bike-shop owner named Igor Kenk turned out to have 2,865 stolen bikes squirreled away in his store and various warehouses. Kenk, considered the most prolific bike thief in the world, is a messy, dyspeptic Slovenian intellectual who before he was caught lived in a fancy house and was associated with the classical music scene (Symmes, 2012).

The Toronto Police Services use a variety of methods to prevent crime. These include educating the public in proactive measures, examining the influence of built environment on behaviour and cognitive-behavioural strategies. (Toronto Police Service, n.d.). The Toronto Police Service also collects data on major crimes as they are reported. While this may serve to increase transparency of crimes to the public, it can also be used as part of a data mining exercise such as clustering.

There are controversies about whether crime reduction by regions is beneficial and effective. While some might argue that offenders will simply go to other locations, a review of 206 studies has found that crime displacement is the exception instead of the norm, proving the effectiveness of regional crime reduction (Guerette, & Bowers, 2009). Marcus Felson's seven-level situational crime prevention framework specifically proposes strategies by the distance from crime events and aims to reduce crime locally. It follows the principle that forces beyond the immediate crime site can reduce local crime opportunity. The seven categories are:

1. Site: categorized by individual units
2. Spot: clusters of properties or small local regions consisting most active crime sites
3. Zone: regions consisting of a fair number of businesses and/or households, for example, an entertainment zone, or a residential neighborhood
4. Metro: metropolis or large-scale region
5. Corporate: multisite corporations for crime prevention
6. National or
7. International

For instance, in the case of automobile theft, at the site level, using entry cards and control gates can properly protect parking units; at the spot level, there needs to be a sufficient amount of surveillance cameras installed, especially at sites where automobile theft frequently happens ensuring bright artificial lighting at entertainment zone will reduce crime at zone level. At metro level, it is beneficial to pass design ordinances for parking structures. Working with owners of multiple parking sites to reduce risks, establishing national laws on parts replacement via insurance, and working out international agreements on locking device manufacturing can reduce automobile theft at the large scale (Felson, 2018).

The strategy of increasing surveillance at the spot level is confirmed by a 40-year systematic meta-analysis. Particularly, it shows that closed-circuit television (CCTV) surveillance cameras is associated with a significant and modest decrease in crime in car parks and residential areas, leading in a reduction of all relevant crimes. In addition, incorporating multiple schemes is demonstrated to be much more effective than having no prevention scheme or one scheme alongside with CCTV (Piza, Welsh, & Farrington, 2019).

Business Objective:

In business, clustering methods are used to identify distinct groups within their customer base. This same methodology could be used to identify patterns in crime data and gain meaningful insights that could be used for crime prevention. Such crime patterns, once identified, could be used as part of a problem-solving process so that a tailored response based on the specific pattern information provided in analysis could be formulated. This would aid in improving the efficiency, effectiveness, and accountability of Toronto Police Services' crime reduction efforts.

Assumptions:

1. Model is based on assumptions that clustering techniques are effective in reduction or prevention of crime.
2. Crime patterns are useful in explaining why crimes are committed.
3. Crime acts with different legal codes are classified by similarity in nature so that they are assigned to the same crime pattern.

Ethical ML Framework:

Problem Definition and Scope

In order to address concerns regarding personally identifiable information (PII), our data collection is subjected to a privacy review and privacy impact assessment (PIA). The resulting dataset has been stripped of personal details and exact crime location information. This ensures the outputs from our models do not compromise individual privacy.

Design

The data mining project will factor in security and privacy impacts in the collection of data for the dataset. Data input into the model will be generated from police reports which are stored in a secure database accessed only by internal staff and requiring two factor authentication. A standardized and rigorous process for categorization along with a profile for those resources used to train the system and interpret the model will be put in place to minimize importing biases into selecting labels and options to train the system.

Data Collection and Retention

Appropriate data governance policies will be developed to manage data retention timelines and procedures for deleting data based on data retention schedules that comply with industry guidelines. Datasets made available to the public will follow ISO standards for managing personally identifiable information and data security.

Data Processing

A risk analysis will be conducted to review types of inferred features that may be present in the dataset. The application of this risk-based approach to the dataset will comply with the Information and Privacy Commissioner of Ontario's recommended process to de-identify data (IPC, 2016).

Model Prototyping and Quality Assurance

Security considerations for the data mining and modeling process will be applied to identify cybersecurity vulnerabilities with databases and the data science teams' workstations, systems access and external access to systems. This will ensure threats to datasets and the models developed are minimized.

Deployment, Monitoring and Maintenance

Development of a detailed process for updating models to ensure the risk of re-identification of an individual is mitigated. Systems design will ensure only authorized team members are able to make queries and train the model through authentication, firewall security and anomaly detection protocols.

Data Mining Objective

This study will employ different analysis methods such as clustering and Principal Components Analysis to identify crime patterns across the city which can be used by law enforcement for preventative measures or the basis for future research initiatives.

Data Understanding

Describe the data

We chose to evaluate crime statistics in the City of Toronto from thefts to assaults and other crimes. We obtained a dataset from the Police department's Public Service Data Portal (Toronto Police Service, n.d.). In this data the Toronto Police Service has taken the necessary measures to protect the privacy of individuals involved in the reported occurrences. No personal information is released as open data and the location of crime occurrences have been deliberately offset to the nearest road intersection node to protect the privacy of parties involved in the occurrence. All location data must be considered as an approximate location of the occurrence.

This dataset includes Major Crime Indicators (MCI) occurrences and their related offences reported between the years 2014 and 2018, inclusively. The dataset consists of 167,525 observations with 29 attributes. The MCI are categorized as Assault, Break and Enter, Auto Theft, Robbery and Theft Over.

The Toronto Police Department provides this data to:

- Develop positive relations between the police and the community through a culture of openness and transparency in policing;
- Inform the community about the police department policies and their measures in curbing crime incidents;
- Educate the public and make them aware of these MCI and their frequent occurrence in neighbourhoods;
- Ensure the public cooperates with them and help uphold law and order in our community.

Verify data quality

We eliminated duplicated entries with the following R code.

```
nrow(toronto)
toronto <- subset(toronto, !duplicated(toronto$event_unique_id))
nrow(toronto)
```

We dropped certain variables that were deemed unimportant for clustering analysis with the following R code.

```
drops <- c("X", "Y", "Index_", "ucr_code", "ucr_ext", "reporteddate", "reportedmonth",
"reportedday", "reporteddayofyear", "reporteddayofweek", "reportedhour", "occurrencedayofyear",
"Division", "Hood_ID", "FID")
toronto <- toronto[, !(names(toronto) %in% drops)]
head(toronto)
unique(toronto$occurrenceyear)
unique(toronto$reportedyear)
```

Grouping crime by year produced the following data:

```
occurrenceyear     n
         <int> <int>
 1        2000    13
 2        2001    10
 3        2002     7
 4        2003     8
 5        2004     9
 6        2005     8
 7        2006     7
 8        2007    16
 9        2008    23
10        2009    28
11        2010    49
12        2011    66
13        2012   117
14        2013   452
```

15	2014	27829
16	2015	28045
17	2016	28274
18	2017	29746
19	2018	31070

Table 1: Crime occurrence by year

In Table 1 we see incidents occurring in each year with a large increase from 2013 to 2014. The following discussion explores what happened in the most recent year 2018.

```
toronto <- toronto[toronto$occurrenceyear == 2018, ]
summary(toronto)
```

There were some missing values in a few columns, so we removed them, and Table 2 shows the resulting data.

```
indicator_group <- group_by(toronto, MCI)
crime_by_indicator <- summarise(indicator_group, n=n())
crime_by_indicator <- crime_by_indicator[order(crime_by_indicator$n, decreasing = TRUE),]
crime_by_indicator
```

1 Assault	15635
2 Break and Enter	7392
3 Auto Theft	4118
4 Robbery	2816
5 Theft Over	1109

Table 2: Top crimes by type in 2018

Table 2 shows the most prominent major crime indicator in 2018. Assault is the most prevalent form of violent crime in Toronto. Assault is an attempt to initiate harmful or offensive contact with a person, or a threat to do so.

Explore the data

To explore the different types of assault we used this R code:

```
assault <- toronto[toronto$MCI == 'Assault', ]
assault_group <- group_by(assault, offence)
assault_by_offence <- summarise(assault_group, n=n())
assault_by_offence <- assault_by_offence[order(assault_by_offence$n, decreasing = TRUE), ]
```

assault_by_offence

We created the following graphing code to produce Figure 1.

```
ggplot(aes(x = reorder(offence, n), y = n), data = assault_by_offence) +
  geom_bar(stat = 'identity', width = 0.6) +
  geom_text(aes(label = n), stat = 'identity', data = assault_by_offence, hjust = -0.1, size = 3) +
  coord_flip() +
  xlab('Types of Assault') +
  ylab('Number of Occurrences') +
  ggtitle('Assault Crimes in Toronto 2018') +
  theme_bw() +
  theme(plot.title = element_text(size = 16),
        axis.title = element_text(size = 12, face = "bold"))
```

It could be argued that the most serious assaults are assault with a weapon, assault that inflicts bodily harm and assaulting a police officer.

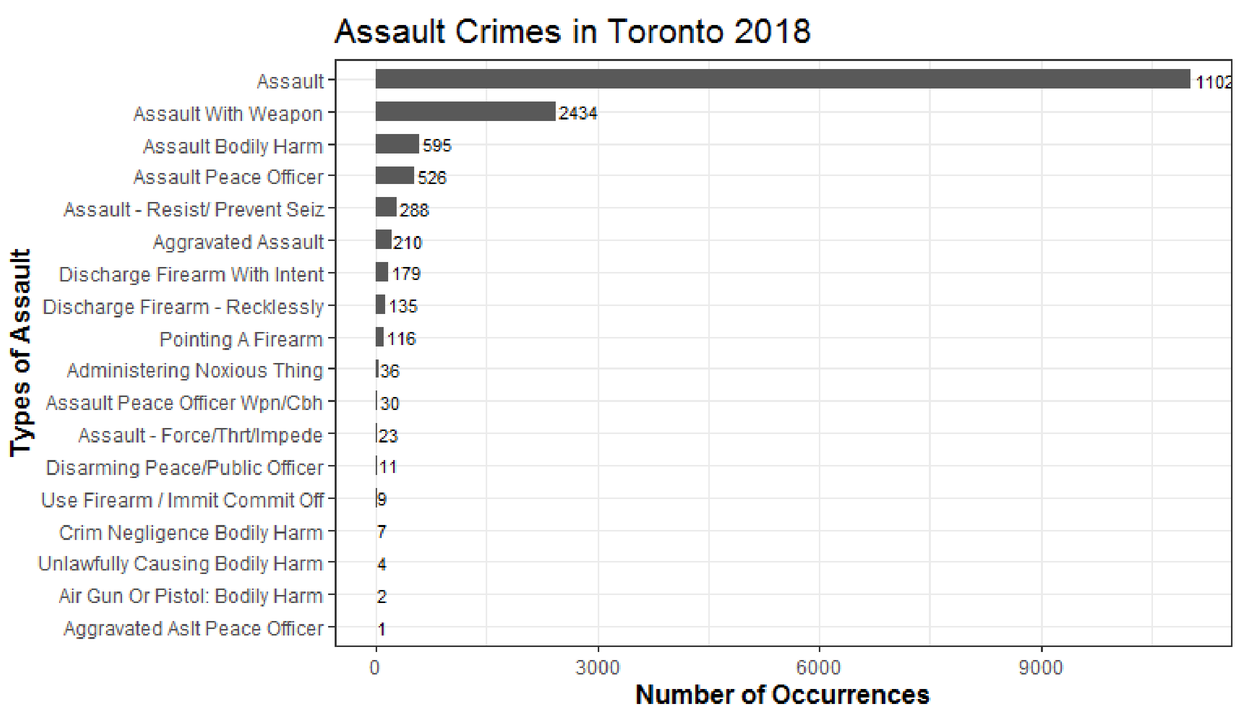

Figure 1: Assault types in 2018

All reported offences in Toronto in 2018 were collated in the following Figure 2.

```
offence_group <- group_by(toronto, offence)
crime_by_offence <- summarise(offence_group, n=n())
crime_by_offence <- crime_by_offence[order(crime_by_offence$n, decreasing = TRUE), ]
crime_by_offence
```

```
ggplot(aes(x = reorder(offence, n), y = n), data = crime_by_offence) +
  geom_bar(stat = 'identity', width = 0.7) +
  geom_text(aes(label = n), stat = 'identity', data = crime_by_offence, hjust = -0.1, size = 2) +
  coord_flip() +
  xlab('Types of Offence') +
  ylab('Number of Occurrences') +
  ggtitle('Offence Types in Toronto 2018') +
  theme_bw() +
  theme(plot.title = element_text(size = 16),
        axis.title = element_text(size = 12, face = "bold"))
```

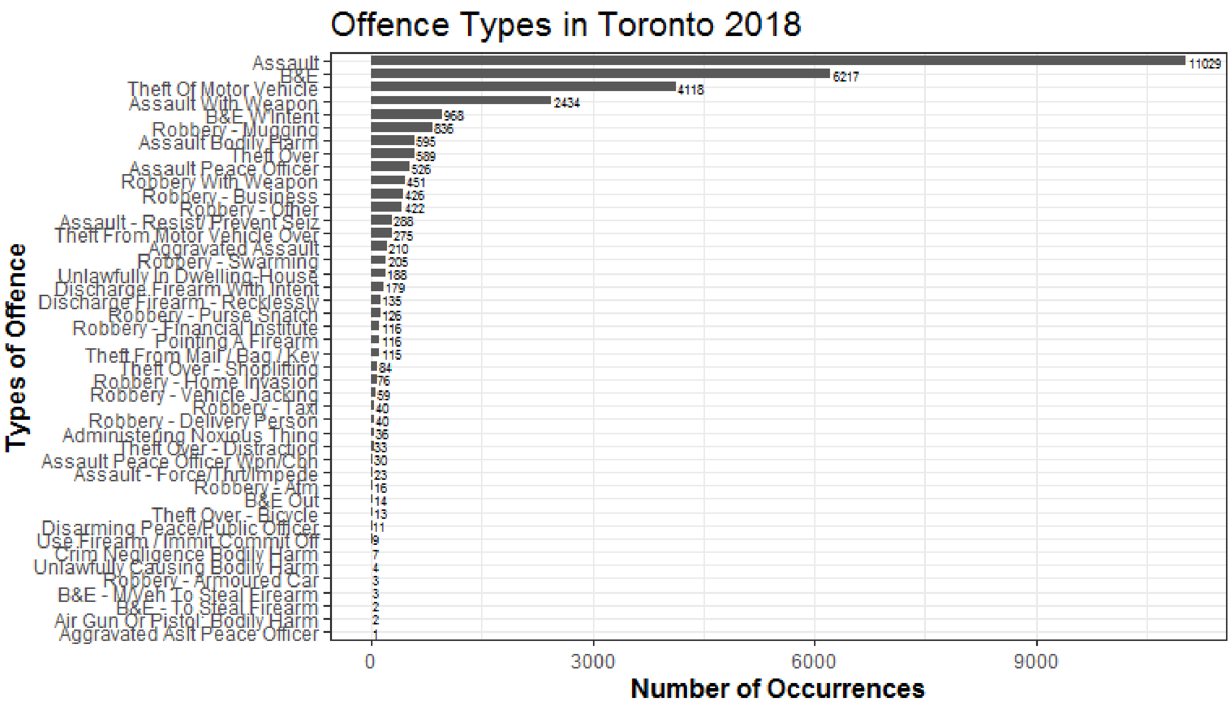

Figure 2: All offences by occurrence in 2018

In Toronto Assault is the most common offence followed by Break and Enter (B & E). Break and Enter refers to the criminal act of entering a residence or enclosed property through the slightest amount of force without authorization. While burglary is usually classified as a felony, breaking and entering is usually classified as a misdemeanor, in a similar manner to criminal trespassing.

The offence of break and enter encompasses situations where the accused was or attempted to trespass on private property with an intent to commit an indictable offence. The most typical form of break and enter is a break into a commercial or private residence in order to steal property. This suggests that break and enters are more likely to occur when there is no one at home.

In a study by Bunge, Johnson & Baldé on B & E in North America, houses are more likely to be targeted than apartments due to the relative ease of approaching a house. In addition to this, it is estimated by the break and enter statistics recorded across North America that more than 67 percent of break-ins happen at the doors of a house, with the front door being the entry point of choice at just over 18 percent. The reason behind is that most thieves look for quick and easy access and do not want a confrontation. The report shows firearms (33%) and knives (30%) to be the weapon of choice for most intruders. Also, around 28% of the victims of this crime are young residents in the age group (18 to 29). Careful analysis of break-ins reveals patterns in which burglars access a home by walking through an unsecured area in the neighbourhood and essentially 'trick or treating' for opportunities to break in (Bunge, Johnson & Baldé, 2005).

According to Bunge, Johnson & Baldé, the weather also plays its part in such crimes. For instance, while most people vacation out during the summer (July and August in particular), their homes remain easy targets due to unreliable surveillance and inadequate security. Though there were no clear patterns when it comes to days of the week a break and enter was reported, it was seen in the study that Fridays had slightly more break-ins than the rest with slightly lower numbers on Saturdays and Sundays when people are home.

In our report below we analyze times of day when crimes occur, days of the week, and occurrences by month, also by neighbourhood along with some explanation of the contributors to these crimes and crime prevention strategies.

Moreover, fighting and preventing crime seems to require efforts at various points - the individual, community, and national levels. Various programs mounted by multiple stakeholders need to work simultaneously directly or indirectly to prevent crime. Involvement needs to take place by the crime victim, criminal justice system, community organization and other specific stakeholders. In addition, educating the public in proactive measures, examining the influence of the environment on behaviour and cognitive-behavioural strategies will all contribute to crime prevention and reduction. In the Appendix our paper includes a wider examination of these issues including the role that AI could play in crime prevention and enhancing security.

To determine the crime by time of the day, this code was used:

```
hour_group <- group_by(toronto, occurrencehour)
crime_hour <- summarise(hour_group, n=n())
 ggplot(aes(x=occurrencehour, y=n), data = crime_hour) + geom_line(size = 2.5, alpha = 0.7, color =
"mediumseagreen", group=1) +
  geom_point(size = 0.5) +
  ggtitle('Total Crimes by Hour of Day in Toronto 2018') +
  ylab('Number of Occurrences') +
  xlab('Hour(24-hour clock)') +
  theme_bw() +
```

```
  theme(plot.title = element_text(size = 16),
        axis.title = element_text(size = 12, face = "bold"))
```

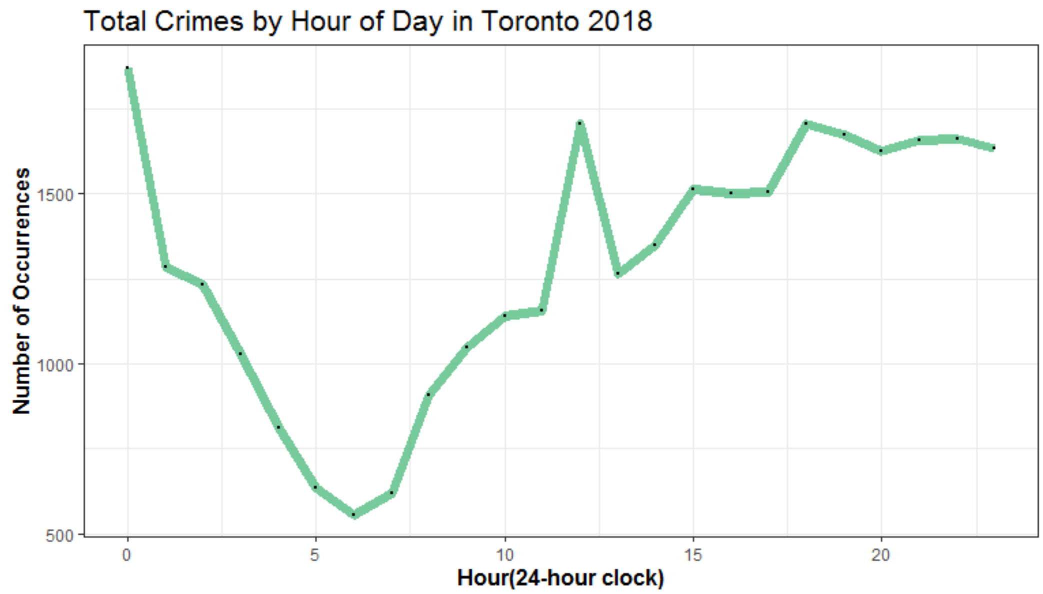

Figure 3: Total Crime occurrence by hour

Here is our R code to determine what types of crime are most frequent at each hour:

```
hour_crime_group <- group_by(toronto, occurrencehour, MCI)
hour_crime <- summarise(hour_crime_group, n=n())

ggplot(aes(x=occurrencehour, y=n, color=MCI), data = hour_crime) +
  geom_line(size=1.5) +
  ggtitle('Crime Types by Hour of Day in Toronto 2018') +
  ylab('Number of Occurrences') +
  xlab('Hour(24-hour clock)') +
  theme_bw() +
  theme(plot.title = element_text(size = 16),
        axis.title = element_text(size = 12, face = "bold"))
```

Assaults are the top crime and they happen more frequently in the afternoon and evening rather than during the day. On the other hand, Break and Enter happened more frequently very early in the morning until around 6 am when people are asleep or during the day when people are likely not at home and then it levels off in the evenings. Robberies and auto thefts are more likely to happen very early in the morning and at night. Restriction on alcohol sales in entertainment districts can reduce violence, by reducing trading hours and banning alcohol takeout and rapid intoxication drinks after a certain time (Devilly, Hides & Kavanagh, 2019).

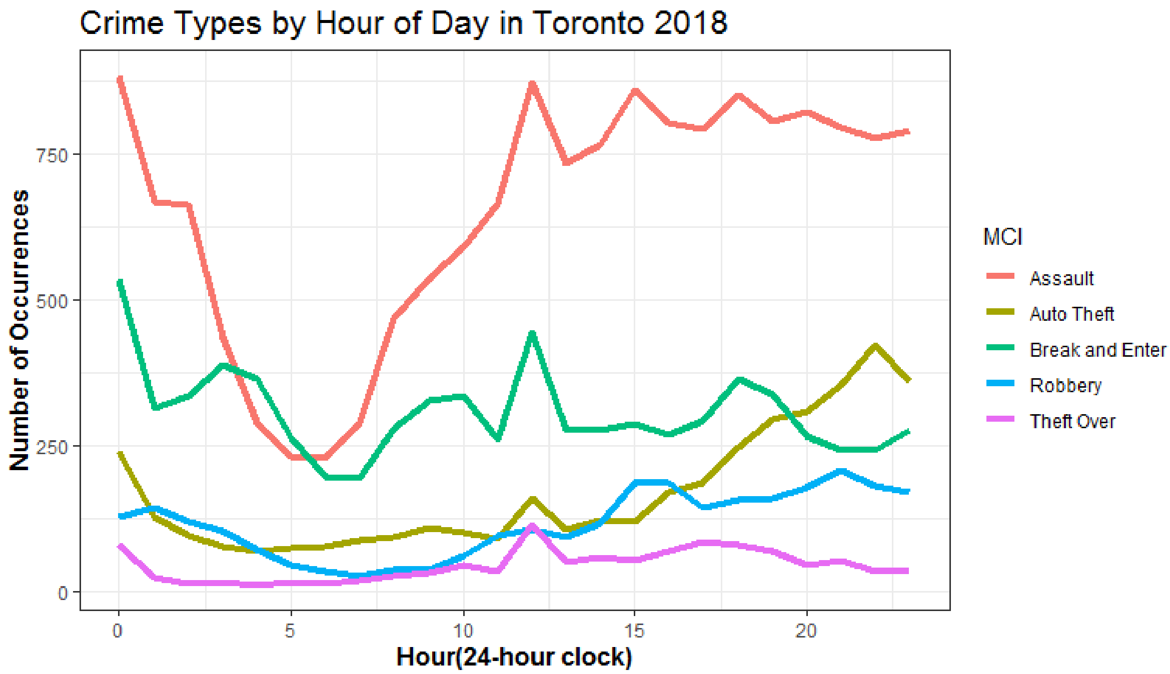

Figure 4: Types of Crime occurrence by hour

Figure 5 shows where those crimes are most likely to occur in Toronto.

```
location_group <- group_by(toronto, Neighbourhood)
crime_by_location <- summarise(location_group, n=n())
crime_by_location <- crime_by_location[order(crime_by_location$n, decreasing = TRUE), ]
crime_by_location_top20 <- head(crime_by_location, 20)
crime_by_location_top20

ggplot(aes(x = reorder(Neighbourhood, n), y = n), data = crime_by_location_top20) +
  geom_bar(stat = 'identity', width = 0.6) +
  geom_text(aes(label = n), stat = 'identity', data = crime_by_location_top20, hjust = -0.1, size = 3) +
  coord_flip() +
  xlab('Neighbourhoods') +
  ylab('Number of Occurrences') +
  ggtitle('Neighbourhoods with Most Crimes - Top 20') +
  theme_bw() +
  theme(plot.title = element_text(size = 16),
        axis.title = element_text(size = 12, face = "bold"))
```

The most dangerous neighbourhood is Church-Yonge Corridor then Waterfront. The Church-Yonge Corridor is popular with students because of the location of Ryerson and it's Toronto's Gay Village, so the area has its share of crime problems. That's a bit disconcerting given how close it is to downtown.

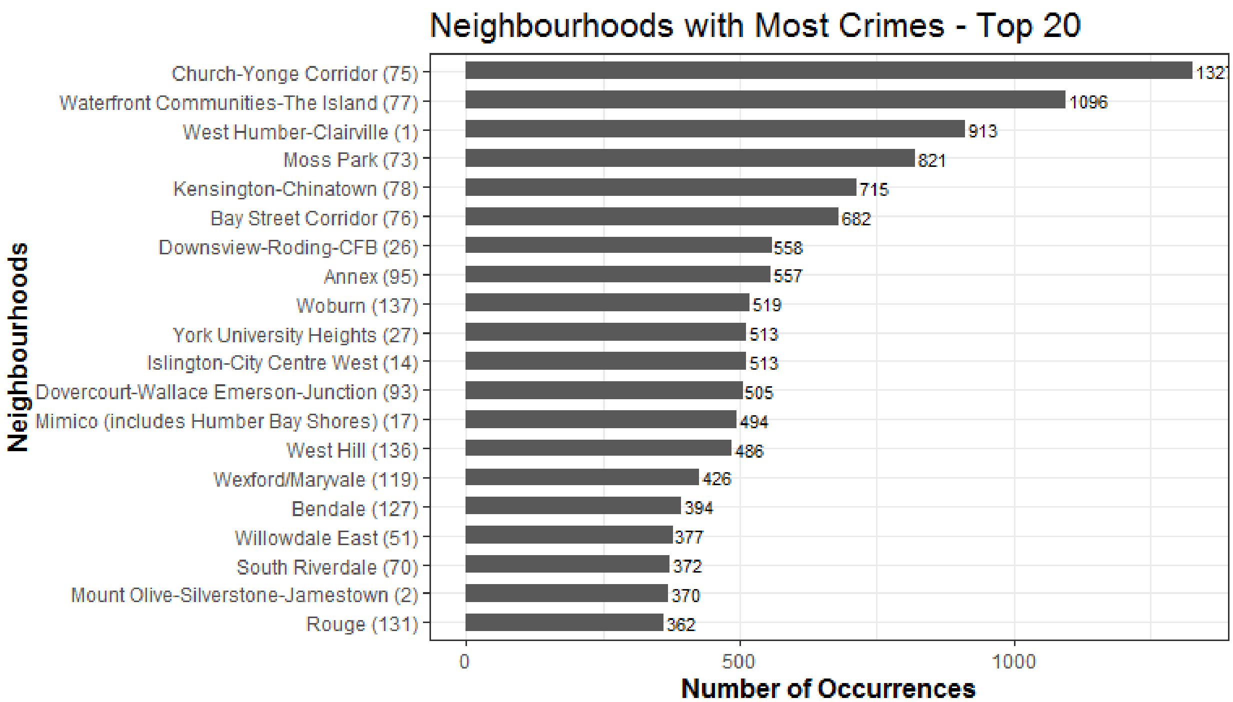

Figure 5: Crimes by neighbourhood

A future study could compare each year of crime statistics to examine how much crimes may move around from neighbourhood to neighbourhood. If it does shift, could it be because an active neighbourhood watch in one neighbourhood helps reduce crime there because their 'watch' is proactive? Or is it because the Police Department cracks down on certain neighbourhoods more than others? The sprawling downtown catch-all includes a densely packed condo-land and the bacchanalian revelry of the entertainment district. The result seems to be a staggering number of violent crimes and arsons. A good question to ask is why crime moves from one neighbourhood to the next. There are controversies about whether crime reduction by regions is beneficial and effective. While some might argue that offenders will simply go to other locations, a review of 206 studies has found that crime displacement is the exception instead of the norm, proving the effectiveness of regional crime reduction (Guerette, & Bowers, 2009). Thus, a good strategy perhaps for the City of Toronto to undertake is to look carefully at the neighbourhoods that have less crime and to use these as benchmarks to emulate in other neighborhoods that have more crime. They could evaluate the characteristics of the low crime neighbourhoods and the crime reduction strategies that seem to be successful there and roll out similar strategies in higher crime neighbourhoods.

Table 3 shows the safest neighbourhoods. Out of this list Eringate-Centennial seems relatively safe. In previous years it could be areas such as Forest Hill South which is a safe and affluent neighbourhood in Toronto that boasts many beautiful homes and mansions. What explains why these neighbourhoods are so crime free? We propose more ideas later in this report that may explain these differences.

tail(crime_by_location, 5)

```
  Neighbourhood                              n
 <fct>                    <int>
1 Caledonia-Fairbank (109)                   44
2 Broadview North (57)                       42
3 Lambton Baby Point (114)                   39
4 Yonge-St.Clair (97)                        38
5 Eringate-Centennial_West Deane (11)         6
```

Table 3: Safest neighbourhoods in Toronto based on incidence of crime

Our code to explore neighbourhoods vs. offence types:

```
offence_location_group <- group_by(toronto, Neighbourhood, offence)
offence_type_by_location <- summarise(offence_location_group, n=n())
offence_type_by_location <- offence_type_by_location[order(offence_type_by_location$n, decreasing =
TRUE), ]
offence_type_by_location_top20 <- head(offence_type_by_location, 20)
ggplot(aes(x = Neighbourhood, y=n, fill = offence), data=offence_type_by_location_top20) +
  geom_bar(stat = 'identity', position = position_dodge(), width = 0.8) +
  xlab('Neighbourhood') +
  ylab('Number of Occurrence') +
  ggtitle('Top Offence Type vs. Neighbourhood Toronto 2018') + theme_bw() +
  theme(plot.title = element_text(size = 16),
          axis.title = element_text(size = 12, face = "bold"),
          axis.text.x = element_text(angle = 90, hjust = 1, vjust = .4)
```

In Figure 6 we see that besides assaults, Church-Yonge Corridor and Waterfront had the most Break and Enters. Incidences of burglary depend on ethnic heterogeneity, percentage of single-family dwellings, and closeness to offenders' residences (Bernasco & Nieuwbeerta, 2005). As stated, the Church-Yonge Corridor and Waterfront have the most Break and Enters. West Humber-Clairville had the most theft of vehicles. It is likely due to flux of travellers from or to Pearson International Airport.

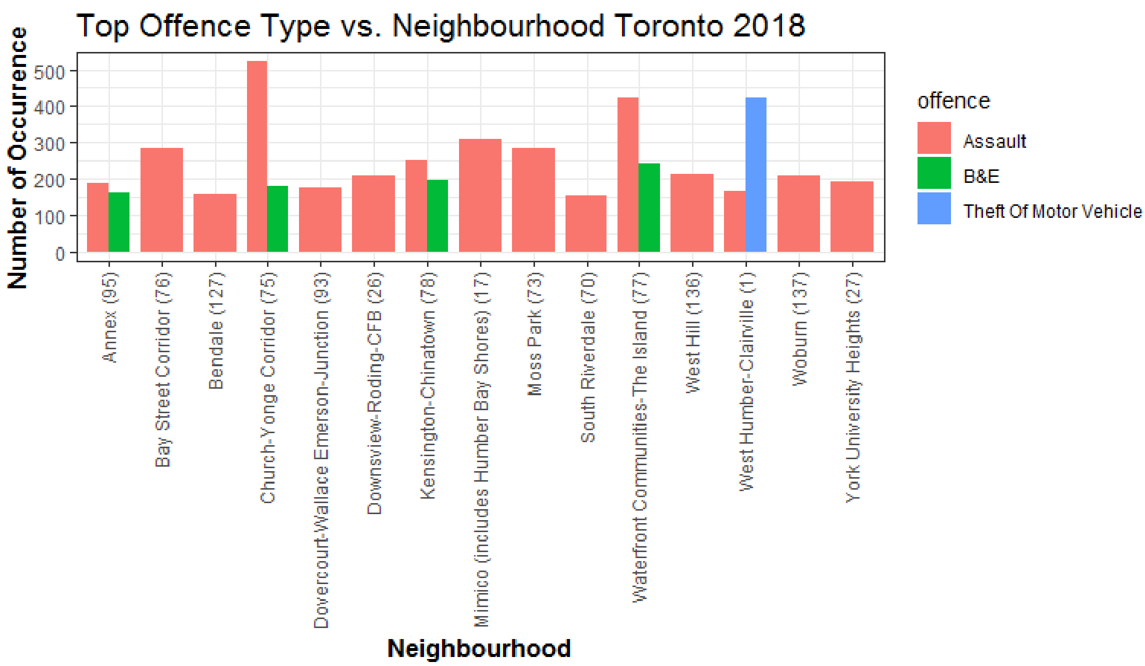

Figure 6: Offence type in different neighbourhoods in Toronto

To create a graphic depicting crime by month we used the following code:

crime_count <- toronto %>% group_by(occurrencemonth, MCI) %>% summarise(Total = n())
crime_count$occurrencemonth <- ordered(crime_count$occurrencemonth, levels = c('January', 'February', 'March', 'April', 'May', 'June', 'July', 'August', 'September', 'October', 'November', 'December'))

ggplot(crime_count, aes(occurrencemonth, MCI, fill = Total)) +

geom_tile(size = 1, color = "white") +

scale_fill_viridis() +

geom_text(aes(label=Total), color='white') +

ggtitle("Major Crime Indicators by Month 2018") +

xlab('Month') +

theme(plot.title = element_text(size = 16),

axis.title = element_text(size = 12, face = "bold"))

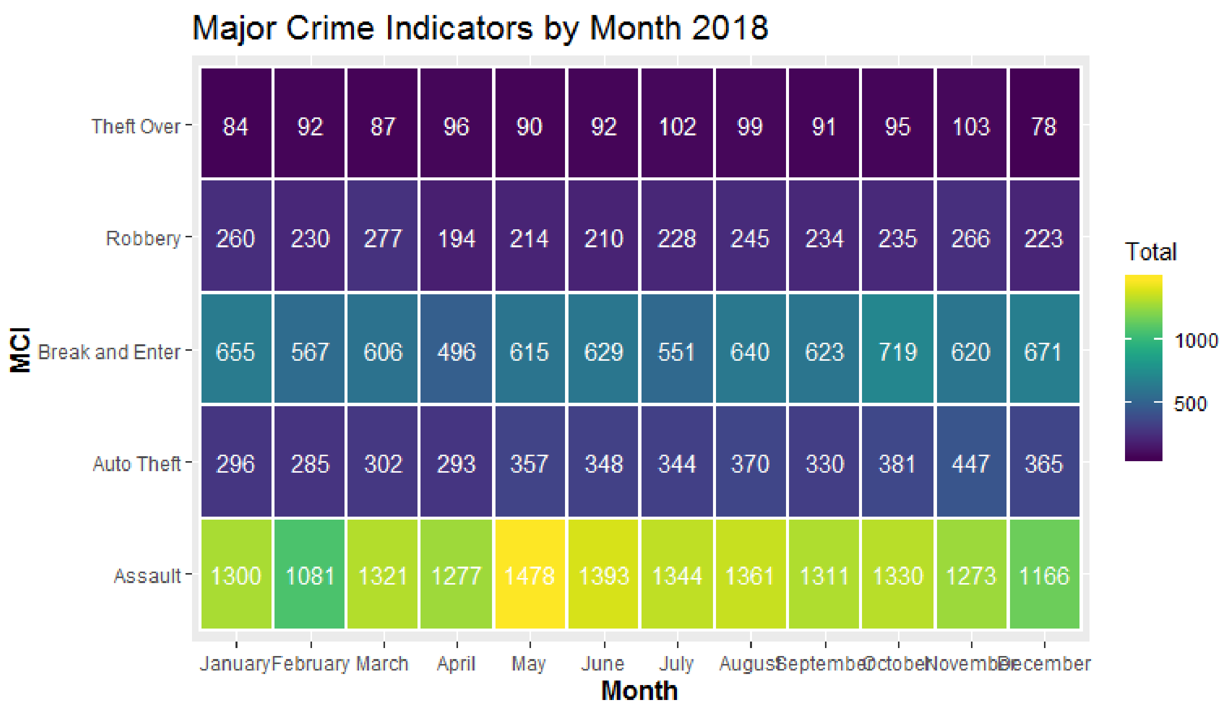

Figure 7: Crimes by month

Figure 7 shows assault is a common crime every month of the year with no exception. However, the weather contributes to automobile theft. Winter is the most frequent time of the year when vehicles are stolen. Thieves frequently help themselves to cars whose drivers leave them to warm up and are less probable to be spotted by other people in the neighbourhood (Gorzelany, 2018).

To create a graphic depicting crime by day of the week we used the following code:

```
day_count <- toronto %>% group_by(occurrencedayofweek, MCI) %>% summarise(Total = n())

ggplot(day_count, aes(occurrencedayofweek, MCI, fill = Total)) +

geom_tile(size = 1, color = "white") +

scale_fill_viridis()  +

geom_text(aes(label=Total), color='white') +

ggtitle("Major Crime Indicators by Day of Week 2018") +

xlab('Day of Week') +

theme(plot.title = element_text(size = 16),

axis.title = element_text(size = 12, face = "bold"))
```

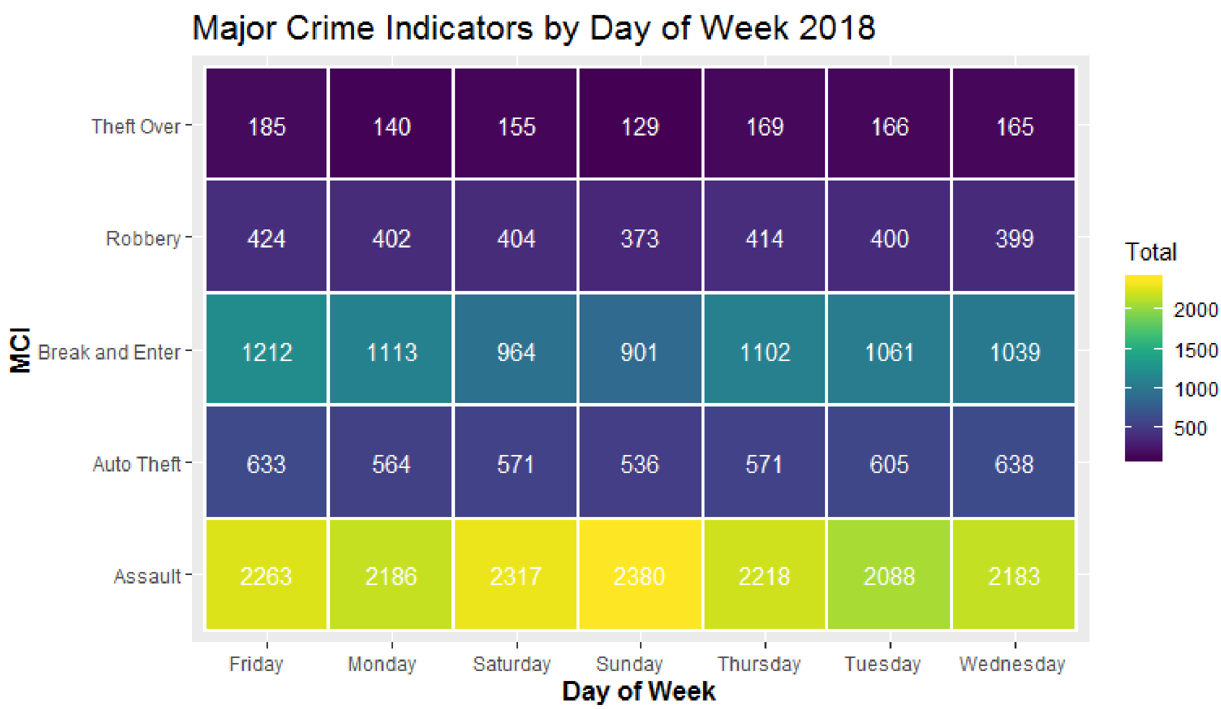

Figure 8: Offence type by days of the week

Essentially, Saturdays and Sundays had more assaults than any other days and had less 'Theft Over' than any other days. Auto thieves are busy almost equally every day of the week.

As far as occurrences of crime goes on a monthly basis, we used this code to create Figure 9.

```
#plot(dataset$occurrencemonth,horiz=TRUE,las=2)
par(las=2) # make label text perpendicular to axis
par(mar=c(5,8,4,2))
counts <- table(dataset$occurrencemonth)
barplot(counts, main="The Distribution of Crimes in the Months of 2018", horiz=TRUE,xlab="Count")
```

Most crimes can be seen taking place from May to October in 2018 during the summer season which can be expected since the temperature is great to be outdoors. Lower crimes seem to occur in the winter months from December to April where the temperature is low, and people are forced to stay indoors. In a future study that compares crime over numerous years we would likely to see the same pattern.

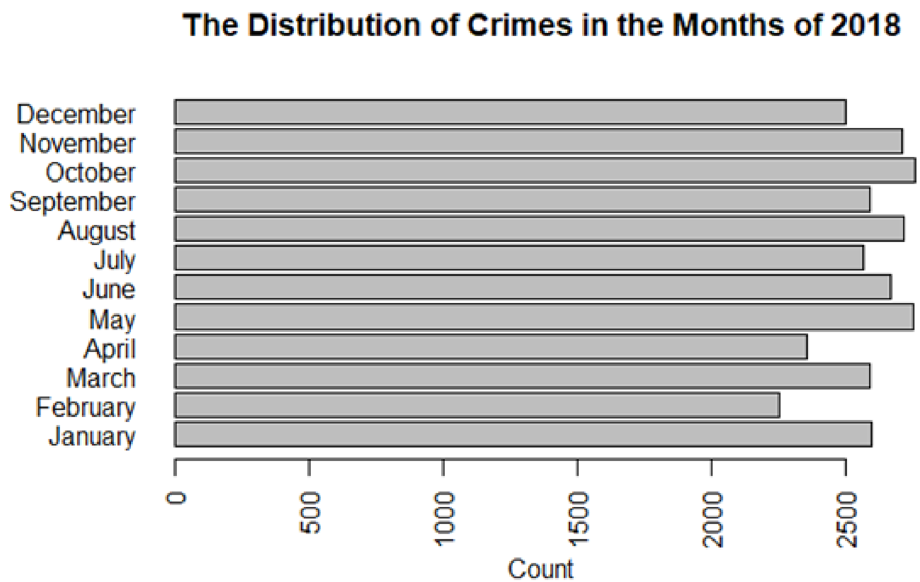

Figure 9: The distribution of the number of crimes in the different months in 2018.

We generated maps showing where crimes occurred using the following code:

```
lat <- toronto$Lat

lon <- toronto$Long

crimes <- toronto$MCI

to_map <- data.frame(crimes, lat, lon)

colnames(to_map) <- c('crimes', 'lat', 'lon')

sbbox <- make_bbox(lon = toronto$Long, lat = toronto$Lat, f = 0.01)

my_map <- get_map(location = sbbox, maptype = "roadmap", scale = 2, color="bw", zoom = 10)

ggmap(my_map) +

geom_point(data=to_map, aes(x = lon, y = lat, color = "#27AE60"),

size = 0.5, alpha = 0.05) +

xlab('Longitude') +

ylab('Latitude') +
```

ggtitle('Location of Major Crime Indicators Toronto 2018') +

guides(color=FALSE)

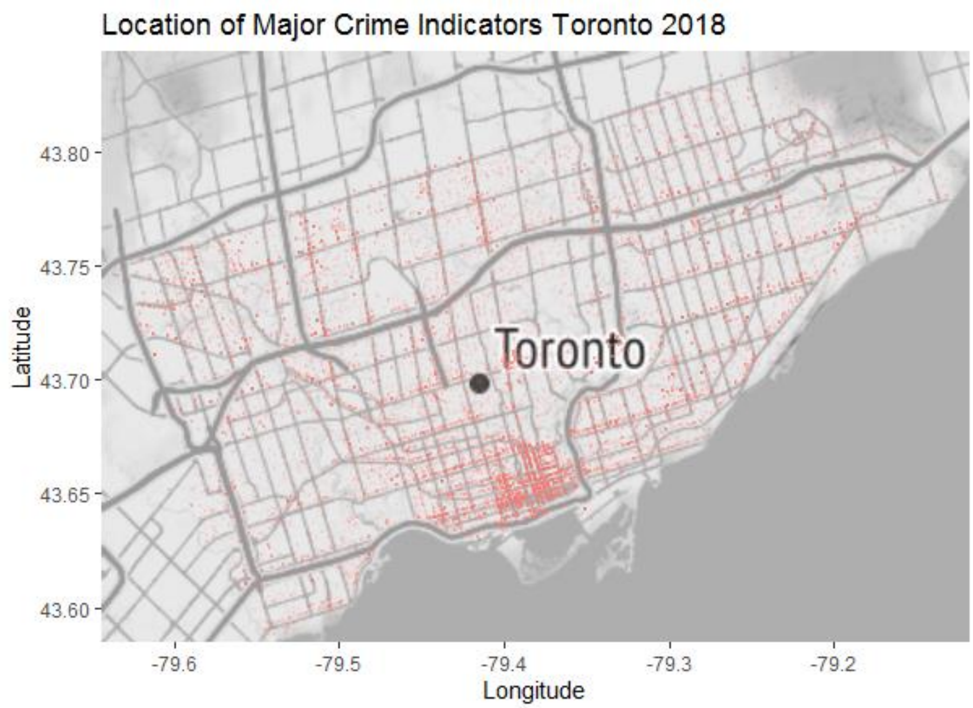

Figure 10: Location of crimes (MCI) in Toronto

Figures 10 and 11 show locations where crimes took place in Toronto, firstly overall crimes, then crimes by types. Figure 11 shows more detail and we could see some crimes, such as Assaults, and Break and Enter occur all over the city but there is a concentration in the downtown areas. Auto theft seems to occur everywhere. Robbery and Theft Over primarily have clusters in the Harbour front area.

```
library(ggmap)

ggmap(my_map) +
  geom_point(data=to_map, aes(x = lon, y = lat, color = "#27AE60"),
           size = 0.5, alpha = 0.05) +
  xlab('Longitude') +
  ylab('Latitude') +
  ggtitle('Location of Major Crime Indicators Toronto 2018') +
  guides(color=FALSE) +
  facet_wrap(~ crimes, nrow = 2)
```

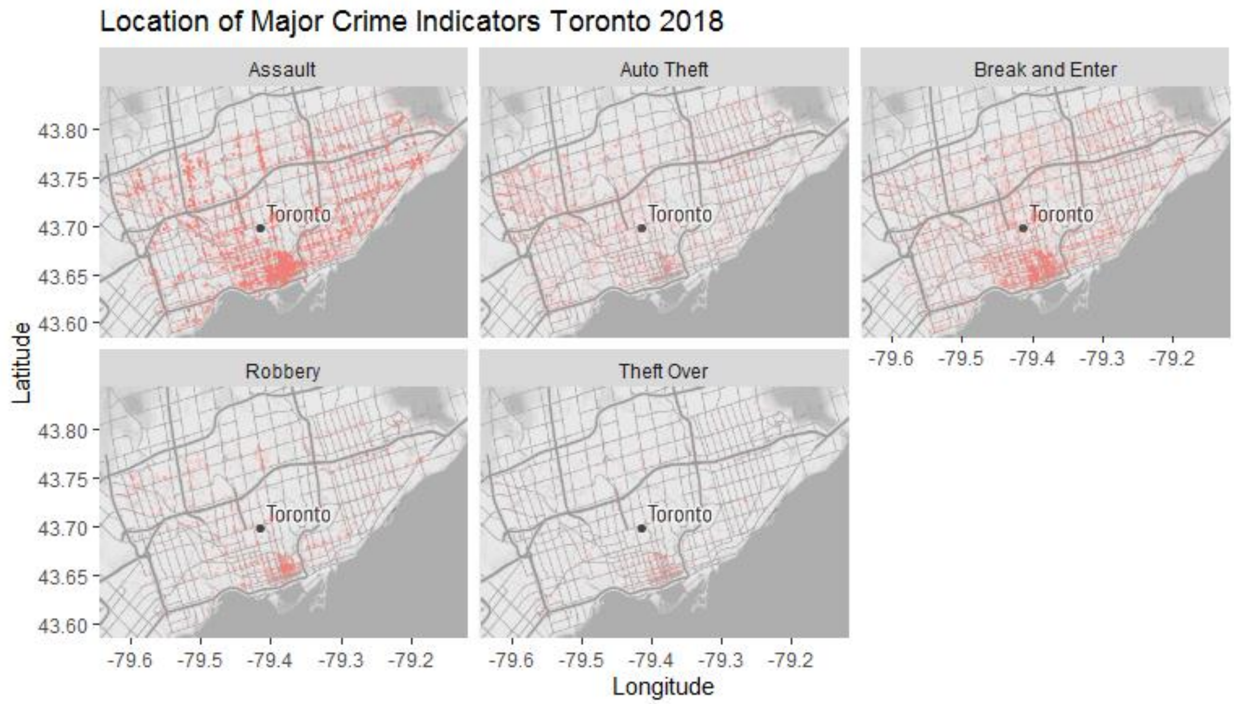

Figure 11: Location of crime types in Toronto

As discussed earlier, crime may move from one neighbourhood to the next and there are controversies about whether crime reduction by regions is effective. However, Marcus Felson's seven-level situational crime prevention framework could help a neighbourhood experiencing more crime. It specifically proposes strategies by the distance from crime events and aims to reduce crime locally. It follows the principle that suggests forces beyond the immediate crime site can reduce local crime opportunity (Felson, 2018). For instance, in the case of automobile theft, at the site level, using entry cards and control gates can properly protect parking units; at the spot level, there needs to be enough surveillance cameras installed, especially at sites where automobile theft frequently happens ensuring bright artificial lighting at an entertainment zone will reduce crime at the zone level. At the metro level, it is beneficial to pass design ordinances for parking structures. Working with owners of multiple parking sites to reduce risks, establishing national laws on parts replacement via insurance, and working out international agreements on locking device manufacturing can reduce automobile theft at the large scale (Felson, 2018).

We created a histogram in Figure 12 showing the number of crimes in 2018 for different hours of the day. With occurrencehour 0 meaning 12 o'clock in the night and 23 meaning 11 pm. Most crimes in 2018 took place from 18 (6 pm) to 23 (11 pm) with many crimes continuing to 0 (12 o'clock midnight) due to it being nighttime where crime is easier to take place. Surprisingly, a lot of crime took place at 12 midday and this can be due to it being the lunch hour where it is quite busy. Low crime can be seen from 3 (3 am) to 9 (9 am) which is due to most people being indoors sleeping.

```
# Color by group
library(ggplot2)
ggplot(dataset, aes(factor(occurrencehour),
fill=factor(occurrencehour)))+geom_bar()
```

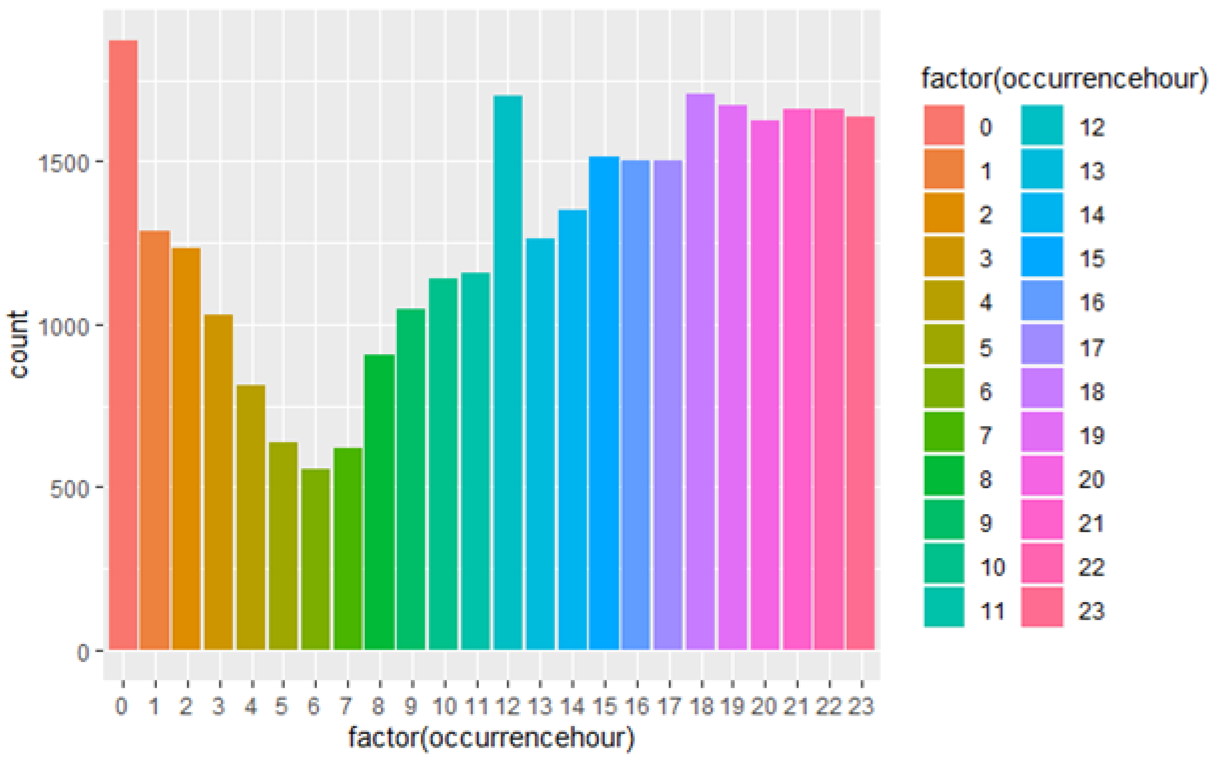

Figure 12: Histogram showing the number of crimes in 2018 for the different hours of the day.

We also analyzed where crimes took place such as outside, in apartments, houses, etc., in Figure 13.

```
plot(dataset$premisetype,main="Premise Type where crime occurred in 2018",xlab="Premise Type",ylab="Count")
```

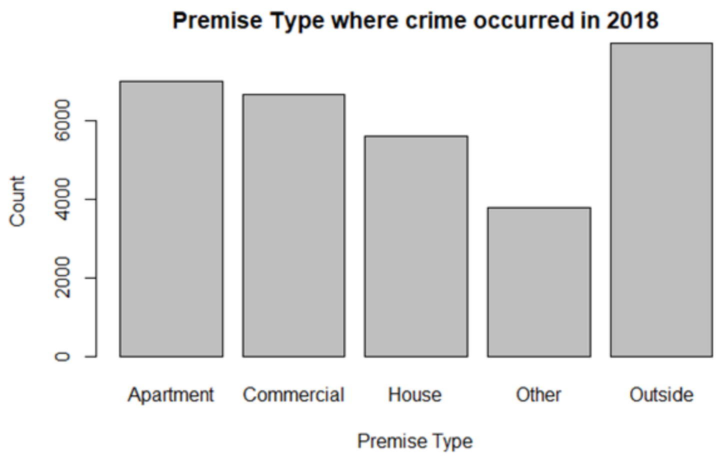

Figure 13: Barplot showing the number of crimes in 2018 for different Premise Type.

Most crime in Toronto took place in the open (Outside) which can be expected, especially in the night. With the large number of apartments present in Toronto in 2018 just under 7000 crimes took place at this type of premise. To combat crimes wherever they occur, the strategy of increasing surveillance at the spot level is confirmed by a 40-year systematic meta-analysis. Particularly, it shows that closed-circuit television (CCTV) surveillance cameras is associated with a significant and modest decrease in crime in car parks and residential areas, leading in a reduction of all relevant crimes. In addition, incorporating multiple schemes is demonstrated to be much more effective than having no prevention scheme or one scheme alongside with CCTV (Piza, Welsh, & Farrington, 2019).

Data Preparation

We intend to use *K*-means clustering which is a type of unsupervised learning.

The data can not have any missing values, so we use the following code.

z <- z[complete.cases(z),]

The data must be scaled for comparison, so we scale it as follows:

m <- apply(z, 2, mean)

s <- apply(z, 2, sd)

z <- scale(z, m, s)

Modelling

Select modelling technique

K-means clustering is a type of unsupervised learning, which is used when you have unlabeled data (i.e., data without defined categories or groups). The *K*-means clustering algorithm is used to find groups which have not been explicitly labeled in the data. The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable *K*. The algorithm works iteratively to assign each data point to one of *K* groups based on the features that are provided. Data points are clustered based on the similarity of features. Clustering allows you to find and analyze the groups that have formed organically rather than defining groups before looking at the data. Each centroid of a cluster is a collection of feature values which define the resulting groups. Examining the centroid feature weights can be used to qualitatively interpret what kind of group each cluster represents (Trevino, 2016).

Once the algorithm has been run and the groups are defined, any new data can be easily assigned to the correct group.

We employ K-Mean clustering using the following code:

```
by_groups <- group_by(toronto, MCI, Neighbourhood)

groups <- summarise(by_groups, n=n())

groups <- groups[c("Neighbourhood", "MCI", "n")]

groups_wide <- spread(groups, key = MCI, value = n)

z <- groups_wide[, -c(1,1)]
```

Next, we determine the number of clusters that appear, and then fit a model.

```
wss <- (nrow(z)-1) * sum(apply(z, 2, var))

for (i in 2:20) wss[i] <- sum(kmeans(z, centers=i)$withiness)

plot(1:20, wss, type='b', xlab='Number of Clusters', ylab='Within groups sum of squares')

kc <- kmeans(z, 2)

kc
```

Generate Test Design

This resulting scree plot in Figure 14 shows a plot with a very definitive elbow. Based on the plot, we can say with confidence that we do not need more than two clusters (centroids).

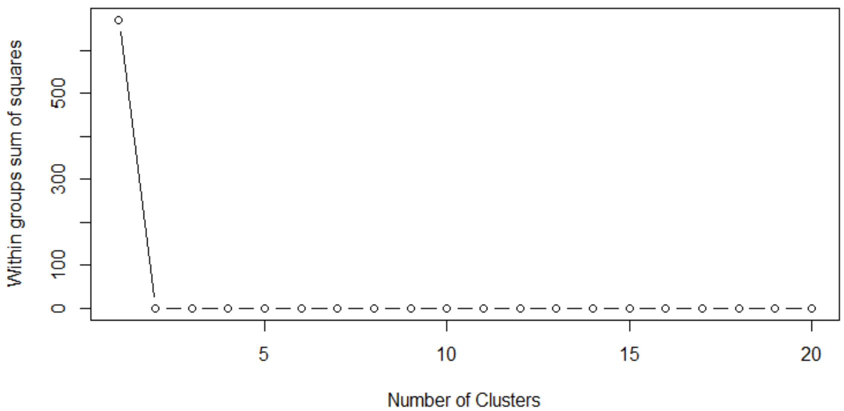

Figure 14: Two clusters emerge

Table 4 below shows the resulting outputs from K-means clustering.

K-means clustering with 2 clusters of sizes 10, 125

Cluster means:

	Assault	Auto Theft	Break and Enter	Robbery	Theft Over
1	2.4065841	1.4836283	2.6307557	2.2985550	2.9355147
2	-0.1925267	-0.1186903	-0.2104605	-0.1838844	-0.2348412

Clustering vector:
 [1] 1 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 [53] 1 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[105] 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1

Within cluster sum of squares by cluster:
[1] 175.5833 183.2225
 (between_SS / total_SS = 46.4 %)

Available components:

[1] "cluster" "centers" "totss" "withinss" "tot.withinss" "betweenss" "size"
[8] "iter" "ifault"

Table 4: Output from K-Means clustering with 2 clusters

In the first cluster we see a number of 10 neighbourhoods, and in the second cluster there are 125 neighbourhoods.

We standardized the data before performing the cluster analysis so the ranges of these numbers in Cluster means seem a bit strange. The negative values mean "lower than most" and positive values mean "higher than most". In Figure 15, the profile of Cluster 1 (small circles in red oval) are 10 neighbourhoods with high means of assault, high auto theft, high break and enter, high robbery and high theft over. Cluster 2 (triangles in blue oval) represents 125 neighbourhoods with low means of assault, low auto theft, low break and enter, low robbery and low theft over. These two groups have a significant variance in every variable. It indicates that each variable plays a significant role in categorizing clusters.

The Clustering vector in Table 4 shows first (West Humber-Clairville (1)), second (Mount Olive-Silverstone-Jamestown (2)) and third (Thistletown-Beaumond Heights (3)) neighbourhoods should all belong to Cluster 1, the fourth neighbourhood (Rexdale-Kipling (4)) should belong to Cluster 2, and so on.

As stated earlier, based on this cluster analysis a reasonable strategy for the City of Toronto to pursue is to look carefully at the neighbourhoods that have less crime (Cluster 2) and to use these as benchmarks to match in other neighborhoods that have more crime (Cluster 1). What are the characteristics of the low crime neighbourhoods and the crime reduction strategies that seem to be successful there? How easily could these characteristics and strategies be replicated in higher crime neighbourhoods?

Build and Assess the Model

A measurement that is more relative would be the withinss and betweenss. Withinss tells us the sum of the square of the distance from each data point to the cluster center. Lower is better. Betweenss tells us the sum of the squared distance between cluster centers. Ideally, we want cluster centers far apart from each other.

Plotting the results.

```
z1 <- data.frame(z, kc$cluster)
clusplot(z1, kc$cluster, color=TRUE, shade=F, labels=0, lines=0, main='k-Means Cluster Analysis')
```

It appears that our choice of number of clusters is good, and we have little noise. Figure 15 illustrates that the two clusters emerge explaining 84.09% of point variability.

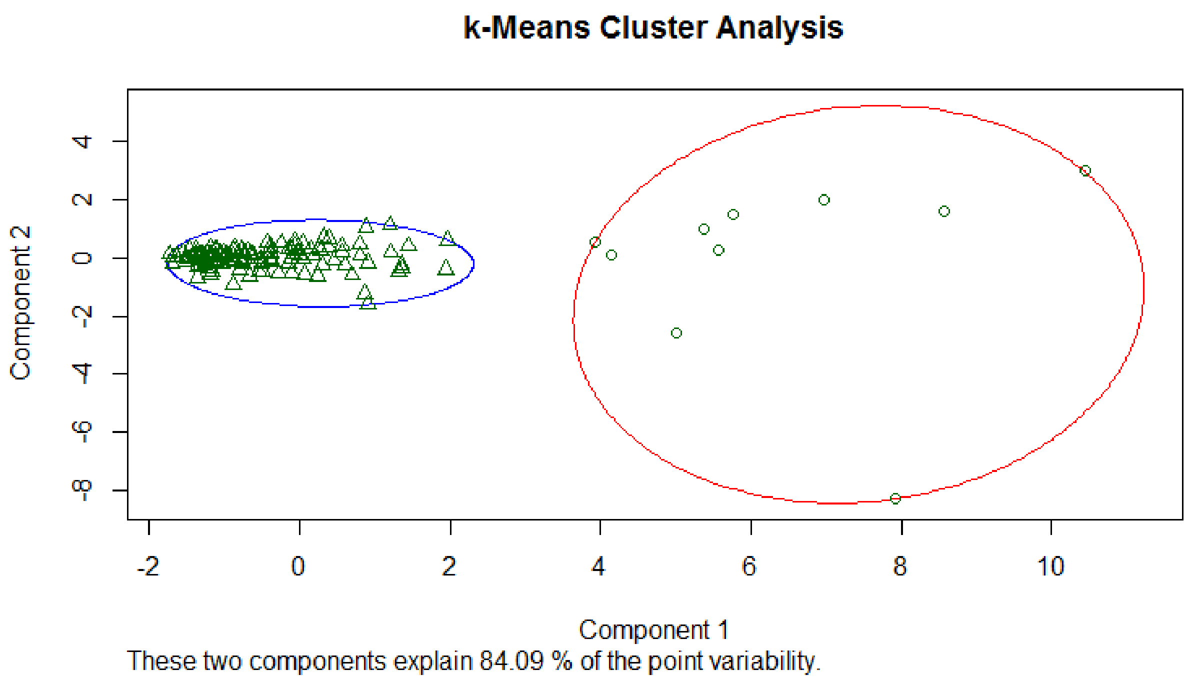

Figure 15: Two clusters emerge explaining 84.09% of point variability

As was suggested, we tried fitting a model with 3 clusters using this code:

```
kc <- kmeans(z, 3)
kc
```

The resulting K-means clustering reveals 3 clusters of sizes 5, 29, 101.

Cluster means:

	Assault	Auto Theft	Break and Enter	Robbery	Theft Over
1	549.6000	39.80000	201.00000	99.80000	38.000000
2	210.1034	57.79310	80.82759	36.75862	15.344828
3	65.0000	21.63366	38.37624	12.03960	4.623762

Clustering vector:

```
  [1] 3 3 3 2 3 3 1 3 3 3 2 2 2 3 3 3 3 3 3 3 1 2 3 3 3 3 3 3 2 2 2 3 2 3 2 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3
 [53] 3 3 2 3 3 2 1 3 3 2 3 3 3 3 3 3 3 2 3 3 3 2 3 1 3 2 3 2 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 2 2 3 3 3
[105] 3 2 3 3 3 3 3 3 3 3 3 3 3 1 2 2 3 3 3 2 2 3 3 2 3 3 3 3 3 2 2
```

Within cluster sum of squares by cluster:

[1] 159562.8 314323.4 195418.7

(between_SS / total_SS = 71.7 %)

Table 5: Output from K-Means clustering with 3 clusters

The Clustering vector in Table 5 shows first (West Humber-Clairville (1)), second (Mount Olive-Silverstone-Jamestown (2)) and third (Thistletown-Beaumond Heights (3)) neighbourhoods should all belong to Cluster 3. The fourth neighbourhood (Rexdale-Kipling (4)) should belong to Cluster 2, the seventh neighbourhood (Willowridge-Martingrove-Richview (7)) should belong to Cluster 1 and so on. Oddly, the 75th (Church-Yonge Corridor (75)), and 77th (Waterfront Communities-The Island (77) show as part of Cluster 3. Cluster 1 has the highest crime. This may be a good reason to prefer a two-cluster solution. Tables 6 and 7 below explore this peculiarity further.

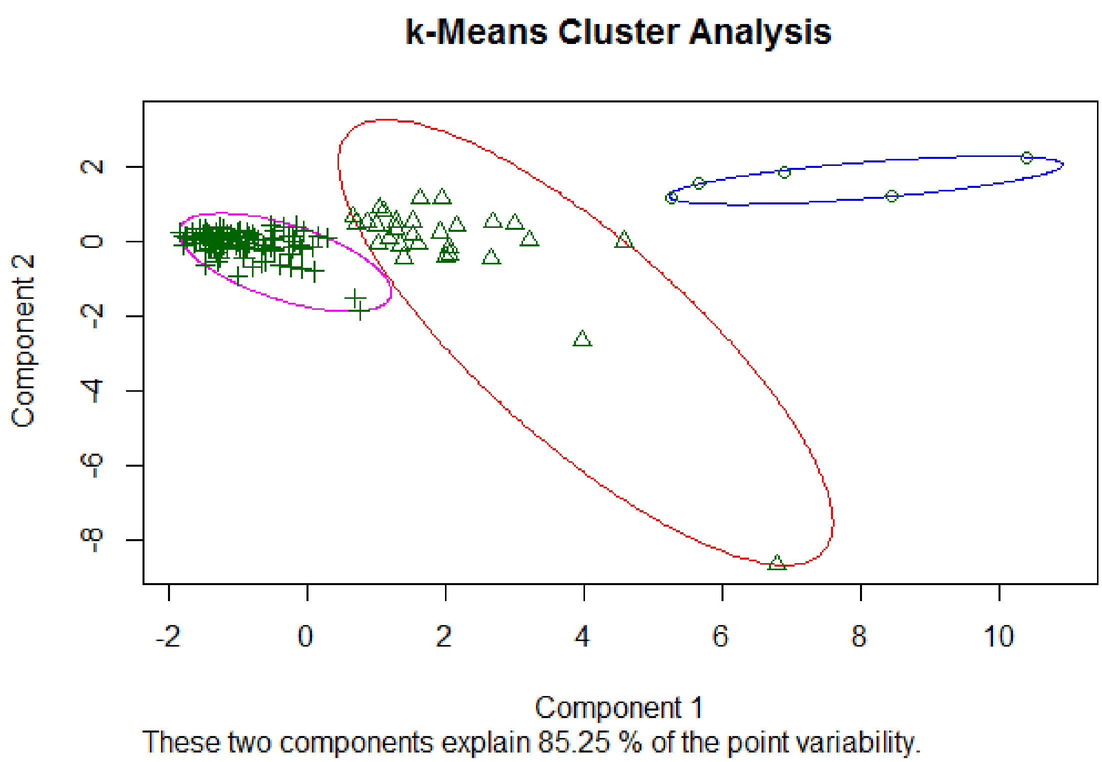

Figure 16: Three clusters emerge explaining 85.25% of point variability

Figure 16 shows three clusters with two components explaining 85.25% of point variability. Also, three clusters reveal a noteworthy insight as described below. Although the location of the elbow in the scree plot in Figure 14 suggests two as a suitable number of clusters for the k-means. Three clusters of neighbourhoods are neighbourhoods (29 triangles in red oval)) with a moderate amount of crime, as well as those (101 small crosses in pink oval) with little, and those with high crime rates (5 small circles in blue oval).

An intriguing observation could be made using three clusters as opposed to two; comparing Table 4 and Table 5 highlights the differences. Cluster 2 is lower in most of the crime indicators than Cluster 1 except Auto Theft. Based on this cluster analysis the strategy for the City of Toronto remains largely the same - look carefully at the neighbourhoods that have less crime and use these as benchmarks to match in other neighborhoods that have more crime. Observe the characteristics of the low crime neighbourhoods and the crime reduction strategies that are successful there and transfer these to higher crime neighbourhoods wherever possible. They could also investigate the oddity of relatively high Auto Theft in Cluster 2 neighbourhoods when compared to Clusters 1 and 3. Could Cluster 2 have a unique demographic where there are a lot of youth for instance? The prime motive among automobile thefts across North America seems to be joyrides by youth (Gorzelany, 2018). As we noted earlier in the report, West Humber-Clairville had the most theft of vehicles and this could be due to the flux of travellers from or to Pearson International Airport. This sort of dynamic in Cluster 2 neighbourhoods could skew auto thefts upwards when compared to other neighbourhoods.

We used this code to create a large Table 9 - MCI's in various neighbourhoods. This is in Appendix 2.

table(toronto$MCI,toronto$Neighbourhood)

From data in the Table of MCI's in various neighbourhoods we created Tables 6 and 7 below to compare the two-cluster and three-cluster solutions in terms of the means of various crimes. A quick calculation of the average in the high crime cluster of Table 6 demonstrates that this cluster has more crimes than the low crime cluster in most crime categories. In low crime neighbourhoods there is on average more Auto Theft (56.23) and Robbery (21.10) than in high crime neighbourhoods where these respective averages are 36 and 18. We calculated average means of Low crime by subtracting High crime total crimes from Total crimes, and then dividing that subtotal by 125, the number of low crime neighbourhoods. Total crimes in all neighbourhoods in 2018 are shown in Table 2 earlier in the report.

Two-cluster solution Cluster 1 (high crime means of 10 neighbourhoods)	Assault	Auto Theft	B & E	Robbery	Theft Over	Total
Rexdale-Kipling (4)	58	50	28	19	4	159
Willowridge-Martingrove-Richview (7)	68	45	65	11	3	192
Humber Summit (21)	100	104	57	34	15	310

Brookhaven-Amesbury (30)	58	36	28	12	6	140
Thorncliffe Park (55)	81	10	23	3	3	120
Danforth East York (59)	59	11	11	7	3	91
Bay Street Corridor (76)	444	26	122	50	40	682
Tam O'Shanter-Sullivan (118)	95	28	66	9	3	201
Clairlea-Birchmount (120)	204	37	62	28	10	341
Highland Creek (134)	58	16	35	6	2	117
High crime totals	1225	363	497	179	89	2353
Averages High crime means	**123**	**36**	**50**	**18**	**9**	**235**
Total crimes (Table 2)	15635	7392	4118	2816	1109	31070
Table 2 Total crimes-High crime totals	14410	7029	3621	2637	1020	28717
Averages Low crime means (Total crimes-High crime total)/ 125	**115.28**	***56.23***	**28.97**	***21.10***	**8.16**	**229.74**

Table 6: Means of crimes in two cluster solution K-Means

We performed the same sort of calculation of averages for the three-cluster solution. These numbers show in Table 7 of Appendix 3. There we see that low crime means are not actually lower than moderate crime means with the exception of B & E. In fact, the means of Auto Theft and Robbery in the Low crime means is actually higher than those respective values of High crime means. This does not seem logical. From this it appears that Table 6 makes more sense and that a two-cluster solution is best.

Furthermore, in Figure 17 we produced a silhouette plot of a two-cluster and three-cluster solution and the two-cluster solution had a higher width score of 0.75 compared to a three-cluster solution of 0.69. According to scikit-learn.org, the two-cluster solution is preferred (Scikit-learn.org, n.d.). As a result, we recommend a two-cluster solution.

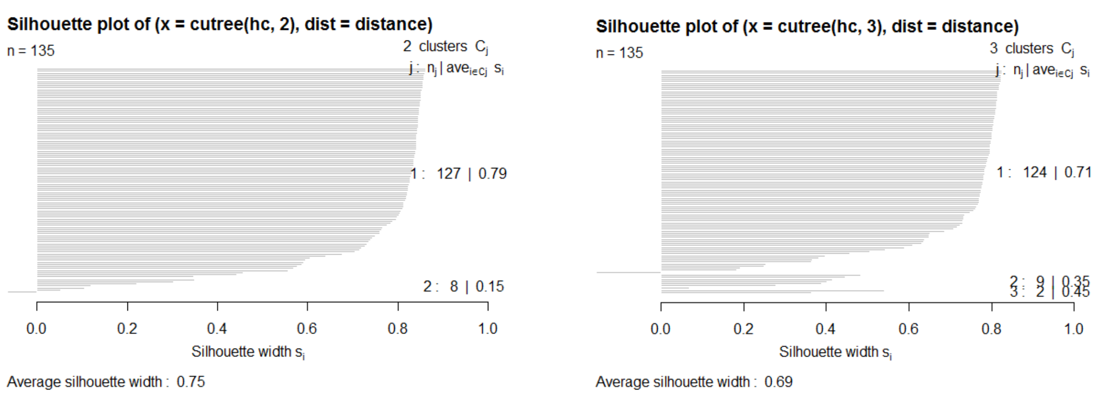

Figure 17: Silhouette plots of two-cluster and three-cluster solutions

Select modelling technique

In hierarchical clustering, each object (data point) is assigned to a separate cluster and then compute the distance (similarity) between each of the clusters and join the two most similar clusters (Joseph, 2018).

The dendrogram is the main graphical tool in hierarchical clustering for getting insight into a cluster solution.

We used the following code:

```
z2 <- data.frame(z)
distance <- dist(z2)
hc <- hclust(distance)
groups_wide <- groups_wide[complete.cases(groups_wide), ]
plot(hc, labels = groups_wide$Neighbourhood, main='Cluster Dendrogram', cex=0.65)
```

Figure 18's Dendrogram shows two clusters in our cluster solution in a bit of a cluttered graphic. If we choose any height along the y-axis of the dendrogram and move across the dendrogram counting the number of lines that we cross, each line represents a cluster. For example, if we look at a height (h) of 12, and move across the x-axis at that height, we'll cross two lines. That defines a two-cluster solution; by following the line down through all its branches, we can see the names of the neighbourhoods that are included in these two clusters. Looking at the dendrogram for the Toronto's crimes data, we can see our data points are very imbalanced. There are two distinct groups; one group consists of branches with branches and more branches, while another group only consists of few data points, and we can see these are Toronto's most dangerous neighbourhoods.

Using R's cutree() function to cut the tree with hc as one parameter and the other parameter as height (h) = 12 or cluster (k) = 2, we can see which observation went into which cluster (HackerEarth, n.d.).

```
cut_avg <- cutree(hc, k = 2)
cut_avg
```

```
 [1] 1 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

 [55] 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

[109] 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
```

To visually see the clusters on the dendrogram you can use the function abline() in R to draw the cut line and cover rectangular compartments for each cluster on the tree with the rect.hclust() function as shown in this next code:

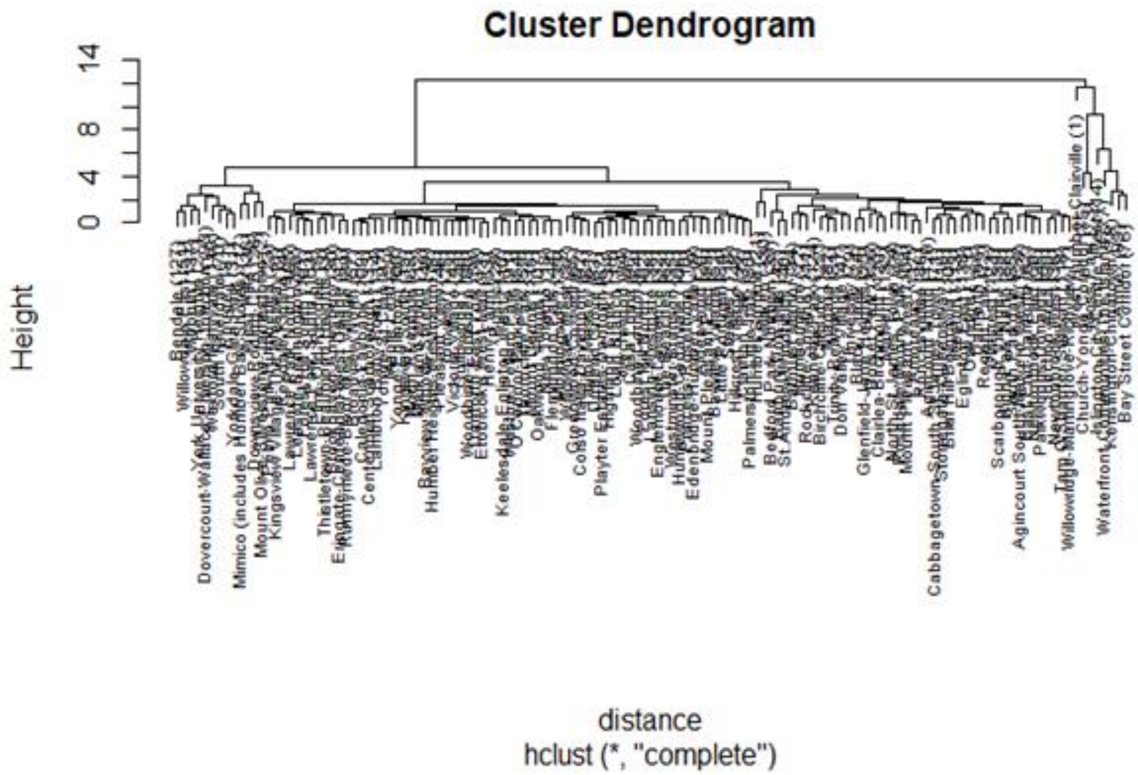

Figure 18: Dendrogram showing two clusters

plot(hc,labels = groups_wide$Neighbourhood,cex=0.5)
rect.hclust(hc , k = 2, border = 2:6)
abline(h = 12, col = 'red')

As you can see, when Height (h) equal to 12 was used to achieve two clusters. The first cluster is shown in the red rectangular box and shows low crime neighbourhoods and the second cluster is shown in the green rectangular box and shows high crime neighbourhoods such as Church-Yonge Corridor and Waterfront Communities - The Island.

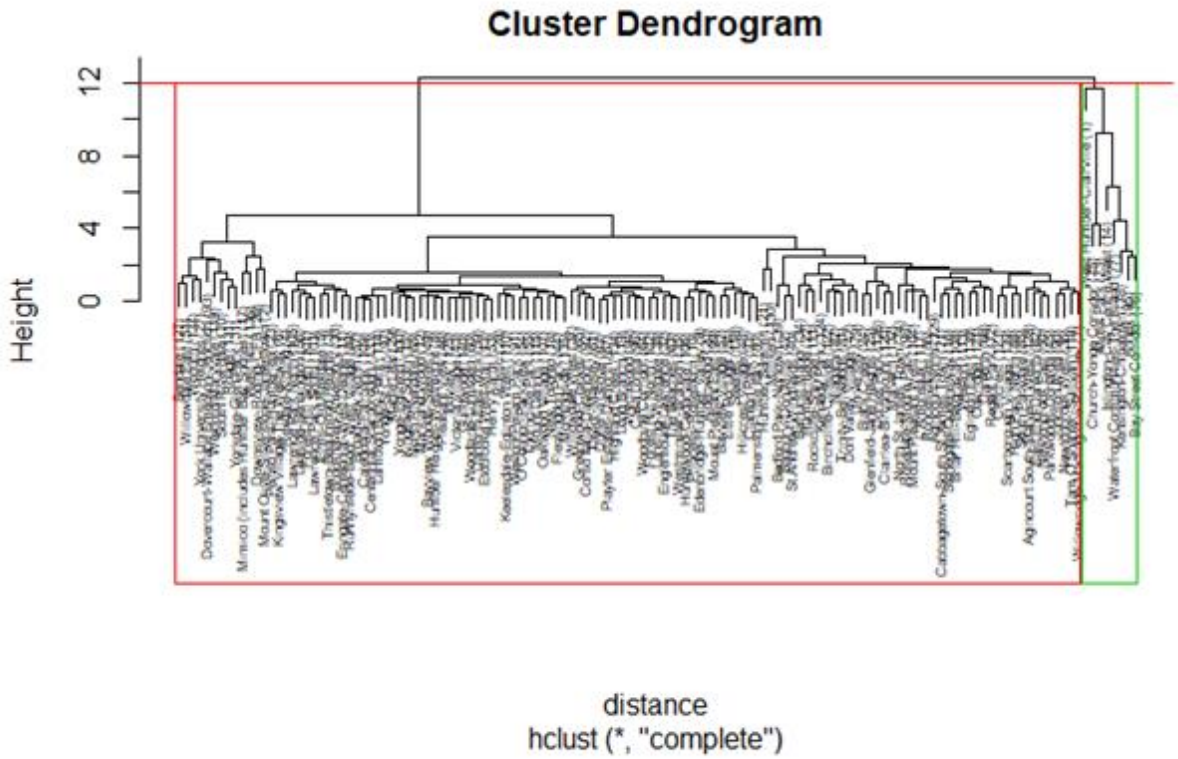

Figure 19: Cluster Dendrogram showing the 2 clusters with line at height (h) = 12

From the dplyr package, we append the cluster results obtained back in the original dataframe under the column name 'cluster' by using the function mutate(). By using the count() function, we count how many observations were assigned to each cluster.

library(dplyr)
seeds_df_cl <- mutate(z2, cluster = cut_avg)
count(seeds_df_cl,cluster)

Table 8 shows the two-cluster solution with 127 observations of low crime neighbourhoods, in this case Cluster 1, and high crime, Cluster 2 with 8 observations. To survey the trend between two features and extract more useful information from the data cluster-wise, we can analyse the movement between 'Assault' and 'Robbery' with the help of ggplot2 package (Pathak, 2018).

cluster <int>	n <int>
1	127
2	8

Table 8: We can see how many observations were assigned in each cluster.

```
library(ggplot2)
ggplot(seeds_df_cl, aes(x=Assault, y = Robbery, color = factor(cluster))) + geom_point()
```

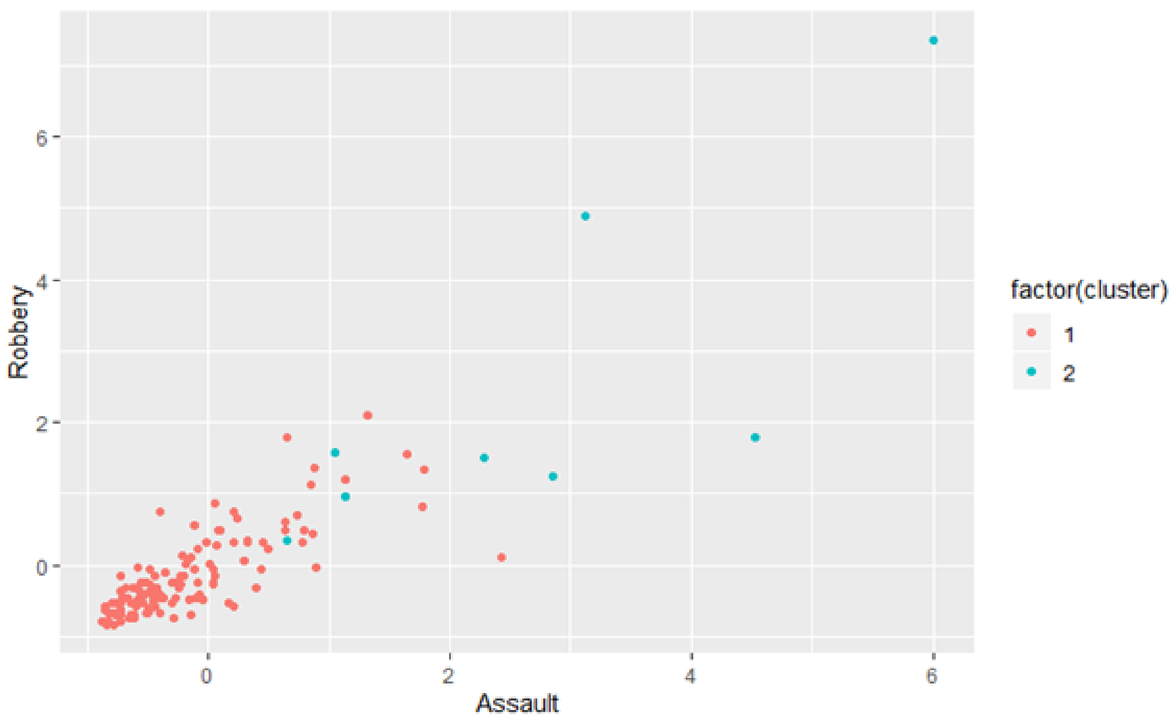

Figure 20: Relationship between Assault and Robbery.

In Figure 20 we notice there seems to be a linear relationship between Assault and Robbery. When theft is carried out with violence or threats of violence, the resultant crime is called a robbery. While money is the prime motivation for such incidents, they could also involve assaults. The more valuable the property being stolen, the more serious the assaults involved seem to be. Such assaults, called aggravated assaults are those with a deadly weapon used by thieves to intimidate their victims into submission. In some cases, assaults happen from self defense measures and the resultant stand-off a victim would have with the perpetrator of the crime (Cook, 1986).

Visualizing high dimensional data in one plot is challenging and it would therefore be more useful using Principal Component Analysis (PCA). PCA is a useful technique for datasets with many variables allowing one to better visualize the variation present. It is most useful for "wide" datasets, where for each sample many variables are present (Hayden, 2018).

Using the princomp() function, we implement PCA and only take the first two components for our advantage.

```
#pca
pcmp <- princomp(z2)
pred_pc <- predict(pcmp, newdata=z2)[,1:2]
```

A data frame with principal components and their corresponding clusters are created and a plot is created using ggplot2.

```
comp_dt <- cbind(as.data.table(pred_pc),cluster = as.factor(groups), Labels = toronto$Neighbourhood)
ggplot(comp_dt,aes(Comp.1,Comp.2))+ geom_point(aes(color = cluster),size=3)
```

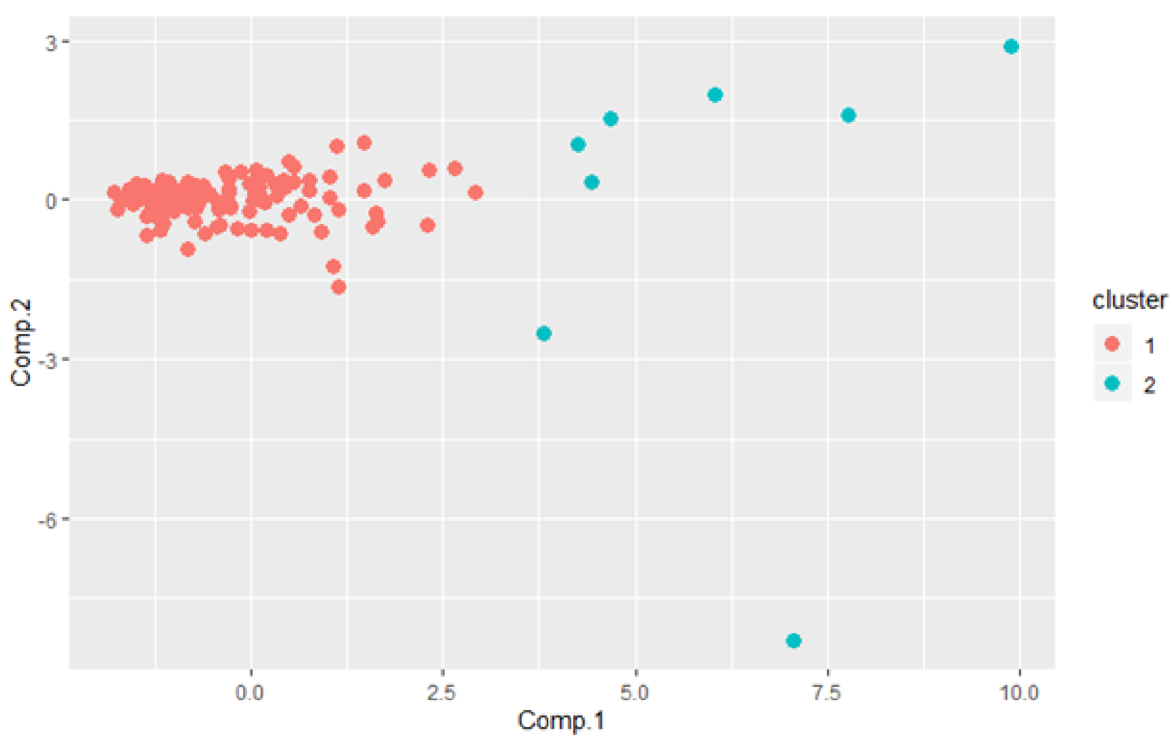

Figure 21: Principal Component Analysis for the two clusters.

In Figure 21 we see the data through one dimension, and it is usually better to make that dimension the principal component with the most variation. We get a better visualization of Cluster 1, low crimes separated clearly from Cluster 2, high crimes.

Evaluation

This study aimed to identify crime patterns with the intent of uncovering useful techniques for crime prevention. Through our initial assessment of the data we were able to conclude:

1. In general, crime is most likely to occur from the period of 6 pm to Midnight and then peak at Noon.
2. When broken down by type of crime, assaults are the top crime and are most likely to occur in the afternoon. Whereas, Break & Enters are more prevalent in the early morning and during the workday. This overlaps with robberies and auto theft which have higher rates of occurrence in the early morning hours.
3. The most serious assaults in 2018, were assault with a weapon, assault that inflicts bodily harm and assaulting a police officer.

The use of K-means clustering and hierarchical clustering are beneficial approaches to detect crime patterns. Our modeling technique was able to identify the crime patterns from many crimes and would be helpful to Toronto Police Services and other stakeholders in their crime prevention efforts. From our results, we believe crime data mining is an effective approach to crime analysis and that this study has met the business criteria.

Here we applied clustering techniques and provided some explanations that could be further developed into specific hypotheses. In a future study we could test hypotheses and apply supervised learning techniques to uncover causal relationships among different crimes and develop deeper insights into crime patterns. For instance, such a study could compare each year of crime statistics to examine how much crimes may move around from neighbourhood to neighbourhood along with causal agents. Another study could examine how clustering may differ if fewer than all 5 crime indicators were used. Also, the TPS has other datasets such as homicide and bike thefts so if these added crime indicators were used in conjunction with the 5 MCI's from our dataset, then other insights could be obtained.

Deployment

For the model to provide value to the TPS, it must be integrated into other police management techniques such as resource planning or integration into other detection

methods. Coordination with other stakeholders would enhance crime prevention approaches. Also, since there is a need for accuracy in the detection of patterns, the dataset used to train the model should be kept up to date in terms of occurrences and types of crimes.
As such, maintenance of the model developed would be required on a monthly basis to ensure that it continued to provide insights into where and when crimes occur in Toronto.

The model would also need to be kept relevant in terms of the types of crimes committed. Additional categories of crime may need to be periodically updated in the dataset.

Fighting and preventing crime seems to require efforts at the individual, community, and national levels and through various programs mounted by different stakeholders such as the crime victim, police services, criminal justice agency, community organization and other specific stakeholders. Throughout this report we provided insights into crime prevention and in the Appendix, we list numerous approaches that could be adopted by various stakeholders. The role of AI in enabling security and crime prevention is also shown.

References

Baraniuk, C. (2018, November). Exclusive: UK police wants AI to stop violent crime before it happens. Retrieved from https://www.newscientist.com/article/2186512-exclusive-uk-police-wants-ai-to-stop-violent-crime-before-it-happens/

Bernasco, W., Nieuwbeerta, P. (2005). How Do Residential Burglars Select Target Areas? A New Approach to the Analysis of Criminal Location Choice. Retrieved from https://www-jstor-org.myaccess.library.utoronto.ca/stable/23639319?pq-origsite=summon&seq=1#metadata_info_tab_contents.

Bunge, V., Johnson, H., & Baldé, T. (June 2005). Exploring Crime Patterns in Canada. Retrieved from http://www.publications.gc.ca/Collection/Statcan/85-561-MIE/85-561-MIE2005005.pdf

Chianis, A. (2018, November). 7 Ways to Avoid Getting Your Car Stolen. Retrieved from https://www.safewise.com/blog/7-ways-avoid-getting-car-stolen/

Conger, C. (2008, June). 10 Simple Ways to Discourage Break-ins. Retrieved from https://home.howstuffworks.com/home-improvement/household-safety/discourage-break-ins1.htm

Conger, C. (2008). Background paper: Residential break and enter. (August, 2011). Department of Justice. Retrieved from http://www.crimeprevention.nsw.gov.au/Documents/break_and_enter_background_paper.pdf

Cook, P. J. "The Relationship between Victim Resistance and Injury in Noncommercial Robbery." The Journal of Legal Studies, vol. 15, no. 2, June 1986, pp. 405–416., doi:10.1086/467819.

Devilly, G.J., Hides, L., Kavanagh, D.J. (2019). PLoS ONE. A big night out getting bigger: Alcohol consumption, arrests and crowd numbers, before and after legislative change. Retrieved from https://doaj.org/article/3efc360e23904b26b6f68b2f4b76d321.

Dhar, R. (2012). Priceonomics.com. What Happens to Stolen Bicycles? Retrieved August 20, 2019 from https://priceonomics.com/post/30393216796/what-happens-to-stolen-bicycles.

Dixit, P. (2019, May). Prioritizing Home Security Residual Risks & Improvements: Toronto. Retrieved from http://www.securehouse.ca/priority-home-security-residual-risks-improvements-toronto.html

Faggella, D. (2019, February). AI for Crime Prevention and Detection - 5 Current Applications. Retrieved August 28, 2019 from: https://emerj.com/ai-sector-overviews/ai-crime-prevention-5-current-applications/

Felson, M. (2018). The Annals of the American Academy of Political and Social Science. Policy Levels for Situational Crime Prevention. Retrieved August 22, 2019 from https://journals-scholarsportal-info.myaccess.library.utoronto.ca/pdf/00027162/v679i0001/198_plfscp.xml

Gorzelany, J. (2018, September). Hottest Wheels: The Most Stolen New And Used Cars In The U.S. Retrieved from https://www.forbes.com/sites/jimgorzelany/2018/09/18/hottest-wheels-the-most-stolen-new-and-used-cars-in-the-u-s/#19741c9c5258

Guerette, R.T., Bowers, K.J. (2009). Criminology (Beverly Hills). Assessing the Extent of Crime Displacement and Diffusion of Benefits: A Review of Situational Crime Prevention Evaluations. Retrieved August 22, 2019 from https://heinonline-org.myaccess.library.utoronto.ca/HOL/Page?handle=hein.journals/crim47&id=1345&collection=journals&index=

HackerEarth. (n.d.). Retrieved August 23, 2019 from https://www.hackerearth.com/practice/machine-learning/machine-learning-algorithms/clustering-algorithms-evaluation-r/tutorial/.

Haskins, C. (2019, February). Dozens of Cities Have Secretly Experimented with Predictive Policing Software. Retrieved from https://www.vice.com/en_us/article/d3m7jq/dozens-of-cities-have-secretly-experimented-with-predictive-policing-software

Haworth, C.D., Petillot, Y.R., Trucco, E. (2006). Image processing techniques for metallic object detection with millimetre-wave images. Retrieved from https://journals-scholarsportal-info.myaccess.library.utoronto.ca/pdf/01678655/v27i0015/1843_iptfmodwmi.xml

Hayden, L. (2018, August). Principle Component Analysis in R. Retrieved August 23, 2019 from https://www.datacamp.com/community/tutorials/pca-analysis-r.

Information and Privacy Commissioner of Ontario. (2016). De-identification Guidelines for Structured Data. Retrieved August 28, 2019 from https://www.ipc.on.ca/wp-content/uploads/2016/08/Deidentification-Guidelines-for-Structured-Data.pdf

Joseph, R. (2018, June). Learn with an example: Hierarchical Clustering. Retrieved on August 29, 2018 from https://medium.com/@rohanjoseph_91119/learn-with-an-example-hierarchical-clustering-873b5b50890c.

Martínez-Ballesté, A., & Solanas, A. (2010). Advances in Artificial Intelligence for Privacy Protection and Security. Hackensack, N.J.: World Scientific.

McCollister, K.E., French, M. T., & Fang, H. (2010). The Cost of Crime to Society: New Crime-Specific Estimates for Policy and Program Evaluation. *Drug Alcohol Depend. 108(1-2):* 98–109. doi: 10.1016/j.drugalcdep.2009.12.002. Retrieved August 20, 2019 from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2835847/.

Neistat, C. (2012). NY Times. 'Bike Thief'; Op-Docs**.** Retrieved August 20, 2019 from https://www.nytimes.com/2012/03/13/opinion/bike-thief.html?_r=1.

Pathak, M. (2018, July). Retrieved August 23, 2019 from https://www.datacamp.com/community/tutorials/hierarchical-clustering-R.

Piza, E.L., Welsh, B.C., & Farrington, D.P. (2019). Wiley Online Library. CCTV surveillance for crime prevention. Retrieved August 22, 2019 from https://onlinelibrary-wiley-com.myaccess.library.utoronto.ca/doi/full/10.1111/1745-9133.12419

Quan, D. (2018, December). Many cars stolen in Canada end up smuggled to distant destinations like Asia, Middle East: study. Retrieved from https://nationalpost.com/news/canada/many-cars-stolen-in-canada-end-up-smuggled-to-distant-destinations-like-asia-middle-east-study

Scikit-learn.org. Selecting the number of clusters with silhouette analysis on KMeans clustering. Retrieved August 31, 2019 from https://scikit-learn.org/stable/auto_examples/cluster/plot_kmeans_silhouette_analysis.html

Shum, D. (2018, December). 'Target harden' your vehicles as auto thefts up 29% in Toronto: police. Retrieved from https://globalnews.ca/news/4758366/toronto-police-vehicle-theft-trends/

Symmes, P. (2012). Outsideonline.com. Who Pinched My Ride? Retrieved August 20, 2019 from https://www.outsideonline.com/1922671/who-pinched-my-ride?page=1.

Takahashi, T. (2016, October). China's Hikvision taps Movidius to bring better A.I. to smart cameras. Retrieved August 28, 2019 from: https://venturebeat.com/2016/10/24/movidius-partners-with-chinas-hikvision-to-bring-better-a-i-to-smart-cameras/

Toronto Police Service; Public Service Data Portal. (n.d.). MCI Indicators. Retrieved August 6, 2019 from http://www.torontopolice.on.ca/d32/.

Toronto Police Service; Crime Prevention (n.d.). Retrieved August 20, 2019: http://www.torontopolice.on.ca/crimeprevention/

Trevino, A. (2016). Learn Data Science, Machine Learning; Introduction to K-means Clustering. Retrieved August 20, 2019 from https://www.datascience.com/blog/k-means-clustering.

U.S. Department of Justice, Bureau of Justice Statistics. Profile of Jail Inmates (2004). Bureau of Justice Statistics; Washington, DC.

Appendix 1

Crime Prevention and Reduction Strategies

As mentioned earlier in the report, fighting and preventing crime seems to require efforts at the individual, community, and national levels and through various programs mounted by different stakeholders such as the crime victim, criminal justice agency, community organization and other specific stakeholders. Also, educating the public in proactive measures, examining the environment on behaviour and cognitive-behavioural strategies will all contribute to crime prevention and reduction. Below we discuss some strategies in more detail.

AI for crime prevention and security

As day to day life becomes more dependent on technology and interconnectivity in cyberspace, policing services can leverage technological advancements to further their efforts in crime prevention and crime detection. Some of the technology at the forefront of crime detection use machine learning algorithms to triangulate the location of a gunshot (Faggella, 2019). The ShotSpotter uses multiple sensors to capture the sound of a gunshot and subsequently record data on the time of the shot, the noise level and echo from nearby buildings. This data is then used by ShotSpotter's machine learning algorithm to triangulate the location of the shot which is then used to alert authorities. The company has a client list of over 90 cities in the United States and has recently expanded to Cape Town, South Africa. While surveillance cameras have been around for decades, Hikvision has taken them to the next step with the introduction of a computer chip with the ability to run deep neural networks. This would enable the camera to scan license plates, run facial recognition and detect anomalies in crowds (Takahashi, 2016).

Machine Learning clustering algorithms are also being used for crime prevention. Using historical data on crime locations, Predpol uses its machine learning algorithm to predict the date and location crimes are most likely to occur in order to identify locations where police patrol resources can be increased.

Security systems have become ubiquitous in airports, important government buildings, museums, etc. The use of artificial intelligence can speed up the process and increase accuracy. Haworth, Petillot, and Trucco develop a robust classification model to identify frames containing threat and locating threat region. It allows an average time of 0.83 second per frame (Haworth, Petillot, & Trucco, 2006).

When a person is charged with a crime, more than likely they would be released until they stand trial. Judges are the ones in the past who decides whether a person can be released before trial and how much bail should be set at. Using their best judgement, judges only have several minutes to establish if an individual is a danger to society, a flight risk or can harm a witness if released; this therefore creates a system open to bias.

In the United Kingdom, the city of Durham is using AI to enhance their present system which decides whether to release a suspect. A software Harm Assessment Risk Tool (Hart), used five years worth of criminal data to determine if a suspect is a low, medium or high risk. Its purpose is to guide authorities on which individuals are more likely to commit another crime. Since 2013, the city has been experimenting with the system and analyzing it with real world results. They said Hart's predictions that a person would be high risk were accurate 88% of the time and determining whether a person would be low risk were accurate 98% of the time (Faggella, 2019).

The potential of AI to enable governments to gather, track, and analyze information for the objective of policing has triggered the spread of AI technology for surveillance & criminal prediction across the world. Its prepended use by law enforcement and security professionals to predict crimes has promise but is still much more of an unknown. But, according to Daniel Flagella of the Emerj Artificial Intelligence Research, the vast computational resources an AI boasts to process and cluster data for pattern recognition, it is wise to reckon that someday we can eventually create more accurate predictions of crimes (Flagella, 2019). PredPol® is considered to be a step in this direction.

PredPol® is a cloud-based software developed by the Predictive Policing Company based in Santa Cruz, CA. It uses a machine-learning algorithm (called real time epidemic type aftershock sequence crime forecasting) to predict critical events and gain actionable insights using data points namely crime type, crime location and crime date with time. These data points considered the three pillars of predictive policing, when coupled with a competent artificial intelligence, help guide their clients in law enforcement and corporate security with actionable insight into where and when specific crimes are most likely to occur (Baraniuk, 2018).

PredPol® algorithm uses point-and-click reporting, crime analytics and graphical crime analysis in predicting future crime spots. It is based on the observation that specific crime types are likely to cluster in space as well as time. By utilizing historical data and paying attention to places where new crimes occur, the Predictive Policing Company claim they can accurately forecast potential hotspots (Haskins, 2019).

The successful implementation of PredPol® in Tacoma, Washington in 2013, saw a 22% decline in residential burglaries through 2015. Furthermore, researchers at Predpol® deduced that police patrols based on their algorithm in Tacoma resulted in a 7.4 % reduction in crime volume during the same time frame. Using PredPol®, law enforcement focusses patrols on potential hotspots and are deemed more effective in reducing crime rates and victimization (Haskins, 2019).

Such systems, called Predictive Policing, have been in use both in the US and the UK. Meanwhile, the Netherlands Police have been using an alternative solution that analyses crime data as well as social data in specific areas to predict where in a city specific types of crime are more likely to occur (Baraniuk, 2018).

Using artificial intelligence to design smart sensors or machines for perimeter security has attracted increased attention recently, such as an acoustic recognizer of running vehicles. This acoustic recognizer is intended for integration into a larger security context. It is normally assumed that there exists a fixed asset to protect and a perimeter that defines the vicinity around that asset for surveillance, whenever using security by humans is dangerous or expensive. The sound recognizer of incoming vehicles is developed for perimeter protection in national, agricultural, airport, prison, military sites, and residential areas. For example, the recognizer can be used to detect approaching vehicles that may be engaged in nefarious activity (Martínez-Ballesté & Solanas, 2010).

Martínez-Ballesté and Solanas also discuss an integrating system that combines different aggression detection systems to make the overall system become more effective. Firstly, the system automatically detects unusual events without being an invasion of individual privacy. Secondly, an application for smart phones enables the best aggression detectors (human beings) to anonymously supply in-put to the detection system. Finally, an underlying framework allows for the fusion of information and delivers the infrastructure to transparently integrate all the systems. To train and the system, several hours of audio and video recordings were taken of aggressive and common scenarios in a real Dutch international train compartment. The researchers expect that the system will contribute to public safety by increasing the chance of detection and thus helping to prevent all forms of aggression and terrorism (Martínez-Ballesté & Solanas, 2010).

<u>Bike theft prevention</u>

Solutions to the bike theft problem are hard to find. More bike racks in better-lit areas, stronger locks and bike garages all help. But ultimately, greater public awareness may be the only way to substantially curb theft. If someone saw a car being stolen, they would surely call the police. Why should a bike be any different? (Neistat, 2012).

The futility of locking is hard to believe. We live in an age of surveillance but there doesn't appear to be a good fix, a tool, gadget, or technology solution. Every technical panacea seems to have its own flaw. Victims of bike theft have created online registries for stolen bikes, but these are obituaries, not a way to pre-empt the crime. Some riders have urged manufacturers to install cheap RFID tags inside every bike they manufacture, like those on clothing; with unique digital signatures, bikes would be completely traceable. However, RFID tags can't be tracked via satellite, only by handheld reader (Symmes, 2012).

Pegasus Technologies, a company in California, created a long-distance system for tracking bikes, which Sacramento police installed in the handlebars of a bait bike. It worked: when the wired bicycle was stolen, police located it across town and arrested the thief (Symmes, 2012).

Part of the solution is to punish such crimes more heavily so that the economics of stealing a bike is removed. Chicago economist Gary Becker introduced the notion that criminal behavior could be modeled using conventional economic theories. Criminals were just rational actors engaged in a careful cost-benefit analysis of whether to commit a crime. Is the potential revenue from the crime greater than the probability of getting caught? (Dhar, 2012).

Criminal activity (especially crime with a clear economic incentive like theft) could therefore be modeled like any financial decision on a risk reward curve. If you are going to take big criminal risk, you need to expect a large financial reward. Crimes that generate more reward than the probability weighted cost of getting caught create expected value for the criminal. Criminals try to find "free lunches" where they can generate revenue with little risk. The government should respond by increasing the penalty for that activity so that the market equilibrates and there is an "optimal" amount of crime (Dhar, 2012).

Using this risk-return framework for crime, it begins to be clear why there is so much bike theft. For all practical purposes, stealing a bike is risk-free crime. It turns out there is a near zero chance you will be caught stealing a bike and if you are, the consequences are minimal (Dhar, 2012).

Bike thievery is essentially a risk-free crime. The other side of the coin is there has to be customer demand and a liquid market for the product in order for the criminal to turn their contraband into revenue. So, how exactly does a criminal go about converting a stolen bicycle to cash? Amateur bike thieves sell their stolen goods at local fencing spots and are typically drug addicts or down on their luck homeless. It is estimated that the overwhelming majority of bike thefts are driven by drug addicts and end up being sold on the street for 5 to 10 cents on the dollar (Dhar, 2012). "Bikes are one of the four commodities of the street — cash, drugs, sex, and bikes… You can virtually exchange one for another." On the other end of the spectrum are professional bike thieves. Instead of opportunistically targeting poorly locked bicycles, these thieves target expensive bicycles. They have the tools that can cut through u-locks and aim to resell stolen bikes at a price near their "fair market value." These thieves acquire the bicycles from the streets, but then resell them on online markets to maximize the selling price. To reduce the incidence of bike theft by both these types of thieves the penalty and fines need to be raised significantly.

Another crime prevention approach is to deal with the demand side of things. Encouraging bike buyers to only buy bicycles from legitimate stores for example.

To help protect a bike, register the bicycle with the Toronto police for instance. On-line registration could be done at www.torontopolice.on.ca. A bike could also be registered at Bikeindex.org. This registry was established by Cofounded by Seth Herr and Bryan Hance in 2013, Bike Index is a non-profit business that is the most widely used and successful bicycle registration service in the world with over 268,000 cataloged bikes, 800 community partners and tens of thousands of daily searches. Bryan developed and ran a community driven bicycle recovery service (StolenBikeRegistry.com) that recovered bikes from the first week it was created in 2004. Seth and Bryan created the universal bike registration service they both dreamed of — a database used and searched by individuals, bike shops, police departments and other apps. A bike registry that gives everyone the ability to register and recover bicycles (Bikeindex.org).

Auto theft prevention

A growing trend sees thieves use more ultramodern tools and electronic override techniques to gain access to vehicles. Yet, most of these incidents are still attributed to thefts by opportunity wherein cars and trucks are an easy target due to an owner's ineptitude or negligence (Gorzelany, 2018).

Lately, luxury vehicles such as Lexus, Mercedes and Land Rover have been the main targets. Investigators say the increased trend in luxury auto thefts is similar in areas surrounding the city including York, Halton and Peel region. Thefts were up two per cent nationally in 2017, and 15 per cent in Ontario, the hardest-hit province (Shum, 2018).

The prime motive among automobile thefts across North America seems to be joyrides by youth. Yet, there have been astounding revelations linking many robberies to mafias and terrorist organisations across the world as a medium for organised crime and terror (Gorzelany, 2018).

Parking intelligently after taking one's surroundings into consideration and securing the vehicle is what we can do to prevent their falling easy prey to thieves. In addition to this, the use of a custom-made anti-theft device and reliance on a tech-based auto recovery tool are essential to deal with sophisticated tools robbers use today (Chlanis, 2018).

Break & Enter prevention

One defining characteristic of residential break and enter is that it is a difficult crime for the authorities to prosecute. The background paper on residential break and enter by the crime prevention group of New South Wales, Australia states that very few of these crimes are reported due to fear of an increase in home insurance premiums. Out of those reported crimes, only about 15 percent of the total reported incidents eventually make it to court (Johnson, 2005),

Moreover, of the people charged with breaking and entering, only about 57 per cent are found guilty due to lack of sufficient evidence. This indicated the difficulty in getting a clear picture of the offenders. However, property crimes were most notable among young offenders. Young male offenders marginally tend to be more pronounced than young female offenders (Johnson, 2005).

Investigating residential break-ins can be challenging if there isn't surveillance or an eyewitness. It is often difficult for police to know exactly what time of day a break-in happens, as victims typically give only a time range. Yet, there are two windows of opportunity: when people are at school or work and when they're sleeping. Therefore, the rate of break-ins reported overnight and during the day are evenly split.

Measures to prevent and control such crimes from happening frequently are extremely important to promote safety and well-being in a neighbourhood. According to Cristen Conger, a writer for Discovery News, the factors in a neighbourhood that increase the rates of breaking and entering are weak regular doors and windows, insufficient detection and reporting of suspicious people, relying solely on a basic alarm system, obscured view of doors and windows to see offenders coming, and giving away information about your routines to strangers thereby enhancing their ability to determine if you are away (Conger, 2008). These are crime factors one can control by hardening their house to reduce target attractiveness all the while supported by other interventions, such as awareness campaigns and natural surveillance to reduce opportunities in the neighbourhood for burglary and home invasion. Awareness campaigns alone as a sole intervention are inadequate and needs to be used in conjunction with other means. Natural surveillance encourages people who use the area to monitor activity as part of their daily life. It works best in areas where there is a motivated group of residents such as a Neighbourhood Watch Group prepared to monitor their local area (Dixit, 2019).

Residents should report to the police and community watch all suspicious activity even if there was nothing stolen. This would help the police understand if there is a pattern of behaviour and prevent such incidents from happening (Johnson, 2005).

Appendix 2

Table of MCI's of various neighbourhoods.

	Assault	Auto Theft	B & E	Robbery	Theft Over
Agincourt North (129)	67	39	70	38	4
Agincourt South-Malvern West (128)	89	31	93	24	7
Alderwood (20)	22	17	25	8	7
Annex (95)	246	32	188	43	48
Banbury-Don Mills (42)	68	22	82	5	7
Bathurst Manor (34)	86	48	62	15	16
Bay Street Corridor (76)	444	26	122	50	40
Bayview Village (52)	56	8	55	6	8
Bayview Woods-Steeles (49)	29	7	28	6	2
Bedford Park-Nortown (39)	37	47	106	10	9
Beechborough-Greenbrook (112)	44	14	37	13	0
Bendale (127)	213	61	56	47	17
Birchcliffe-Cliffside (122)	160	23	66	13	8
Black Creek (24)	188	57	30	35	7
Blake-Jones (69)	38	6	13	10	0
Briar Hill-Belgravia (108)	103	39	41	26	3
Bridle Path-Sunnybrook-York Mills (41)	23	11	53	0	6
Broadview North (57)	29	3	7	2	1
Brookhaven-Amesbury (30)	58	36	28	12	6
Cabbagetown-South St.James Town (71)	94	23	47	22	6
Caledonia-Fairbank (109)	24	8	6	5	1
Casa Loma (96)	46	11	24	8	2
Centennial Scarborough (133)	25	8	7	4	2
Church-Yonge Corridor (75)	810	70	214	194	39
Clairlea-Birchmount (120)	204	37	62	28	10
Clanton Park (33)	51	23	45	8	7
Cliffcrest (123)	106	19	47	11	2
Corso Italia-Davenport (92)	91	13	34	17	4
Danforth (66)	120	19	45	41	7
Danforth East York (59)	59	11	11	7	3
Don Valley Village (47)	98	24	55	23	12
Dorset Park (126)	152	65	68	28	12
Dovercourt-Wallace Emerson-Junction (93)	246	41	153	49	16
Downsview-Roding-CFB (26)	321	95	81	52	9
Dufferin Grove (83)	61	13	38	10	4
East End-Danforth (62)	149	26	78	22	6
Edenbridge-Humber Valley (9)	18	15	43	5	5
Eglinton East (138)	151	22	51	29	5
Elms-Old Rexdale (5)	62	34	7	13	1
Englemount-Lawrence (32)	71	29	42	10	2
Eringate-Centennial-West Deane (11)	34	26	31	13	3
Eringate-Centennial_West Deane (11)	2	0	2	1	1
Etobicoke West Mall (13)	38	10	18	3	3
Flemingdon Park (44)	106	8	14	11	1
Forest Hill North (102)	42	18	49	13	1
Forest Hill South (101)	15	23	34	6	1
Glenfield-Jane Heights (25)	206	49	36	32	10
Greenwood-Coxwell (65)	87	18	39	14	7
Guildwood (140)	43	4	18	3	1
Henry Farm (53)	43	8	20	4	3
High Park-Swansea (87)	46	18	32	3	0
High Park North (88)	92	13	61	21	5
Highland Creek (134)	58	16	35	6	2
Hillcrest Village (48)	54	21	45	15	5
Humber Heights-Westmount (8)	40	16	29	4	3
Humber Summit (21)	100	104	57	34	15
Humbermede (22)	81	44	38	15	11
Humewood-Cedarvale (106)	23	18	24	1	6
Ionview (125)	62	5	20	7	1
Islington-City Centre West (14)	189	137	111	29	47
Junction Area (90)	119	34	63	19	6
Keelesdale-Eglinton West (110)	49	15	16	15	3
Kennedy Park (124)	165	17	41	19	18
Kensington-Chinatown (78)	379	19	232	56	29
Kingsview Village-The Westway (6)	87	49	39	17	3
Kingsway South (15)	29	9	19	12	4
L'Amoreaux (117)	138	28	85	38	8
Lambton Baby Point (114)	16	14	6	1	2
Lansing-Westgate (38)	46	18	48	13	3
Lawrence Park North (105)	18	38	17	2	3
Lawrence Park South (103)	15	21	34	7	3
Leaside-Bennington (56)	29	33	37	4	2
Little Portugal (84)	61	11	47	11	9
Long Branch (19)	39	10	38	8	2
Malvern (132)	200	45	58	37	12
Maple Leaf (29)	42	18	12	8	5
Markland Wood (12)	16	13	22	2	2
Milliken (130)	57	78	118	14	16
Mimico (includes Humber Bay Shores) (17)	396	20	46	23	9
Morningside (135)	102	9	18	10	2
Moss Park (73)	476	35	140	136	34

Mount Dennis (115)	79	18	22	8	1
Mount Olive-Silverstone-Jamestown (2)	216	71	27	53	3
Mount Pleasant East (99)	29	15	39	8	8
Mount Pleasant West (104)	139	13	95	28	8
New Toronto (18)	88	10	22	14	1
Newtonbrook East (50)	119	25	74	15	7
Newtonbrook West (36)	119	35	54	14	4
Niagara (82)	139	12	64	7	12
North Riverdale (68)	31	7	23	10	3
North St.James Town (74)	172	9	72	26	10
O'Connor-Parkview (54)	97	13	25	4	4
Oakridge (121)	123	15	65	32	4
Oakwood Village (107)	82	24	15	10	1
Old East York (58)	76	9	30	8	0
Palmerston-Little Italy (80)	53	22	43	11	6
Parkwoods-Donalda (45)	101	18	77	19	6
Pelmo Park-Humberlea (23)	47	58	23	12	3
Playter Estates-Danforth (67)	61	10	37	9	4
Pleasant View (46)	23	13	28	5	4
Princess-Rosethorn (10)	12	20	22	2	5
Regent Park (72)	112	9	34	28	6
Rexdale-Kipling (4)	58	50	28	19	4
Rockcliffe-Smythe (111)	104	42	24	15	16
Roncesvalles (86)	125	26	84	32	8
Rosedale-Moore Park (98)	188	30	95	32	9
Rouge (131)	167	58	88	28	21
Runnymede-Bloor West Village (89)	30	23	23	17	2
Rustic (28)	79	17	10	15	2
Scarborough Village (139)	133	17	44	8	3
South Parkdale (85)	121	9	69	17	6
South Riverdale (70)	217	29	83	20	23
St.Andrew-Windfields (40)	56	26	83	5	5
Steeles (116)	29	21	62	8	5
Stonegate-Queensway (16)	72	31	40	18	7
Tam O'Shanter-Sullivan (118)	95	28	66	9	3
Taylor-Massey (61)	85	8	37	13	4
The Beaches (63)	68	25	38	10	4
Thistletown-Beaumond Heights (3)	46	42	17	9	1
Thorncliffe Park (55)	81	10	23	3	3
Trinity-Bellwoods (81)	109	14	57	9	9
University (79)	116	13	76	21	14
Victoria Village (43)	55	7	23	5	2
Waterfront Communities-The Island (77)	639	49	297	63	48
West Hill (136)	319	25	98	40	4
West Humber-Clairville (1)	235	422	152	58	46
Westminster-Branson (35)	64	26	37	13	3
Weston-Pellam Park (91)	62	14	24	17	2
Weston (113)	122	43	37	27	6
Wexford/Maryvale (119)	214	68	100	31	13
Willowdale East (51)	189	40	70	63	15
Willowdale West (37)	46	8	32	20	5
Willowridge-Martingrove-Richview (7)	68	45	65	11	3
Woburn (137)	304	49	97	57	12
Woodbine-Lumsden (60)	44	2	20	7	1
Woodbine Corridor (64)	67	10	21	10	2
Wychwood (94)	60	18	31	12	2
Yonge-Eglinton (100)	25	13	11	8	2
Yonge-St.Clair (97)	17	4	13	1	3
York University Heights (27)	267	63	92	70	21
Yorkdale-Glen Park (31)	142	44	65	36	18

Table 9: MCI's of various neighbourhoods

Appendix 3

Three-cluster solution

Cluster 1 (high crime means)	Assault	Auto Theft	B & E	Robbery	Theft Over	Total
Willowridge-Martingrove-Richview (7)	68	45	65	11	3	192
Humber Summit (21)	100	104	57	34	15	310
Danforth East York (59)	59	11	11	7	3	682
Bay Street Corridor (76)	444	26	122	50	40	682
Tam O'Shanter-Sullivan (118)	95	28	66	9	3	201
Averages High crime means	**153**	**43**	**64**	**22**	**13**	**413**
Total High crimes	766	214	321	111	64	1476
Cluster 2 (moderate crime means)						
Rexdale-Kipling (4)	58	50	28	19	4	159
Eringate-Centennial-West Deane (11)	36	26	33	14	4	113
Markland Wood (12)	16	13	22	2	2	55
Etobicoke West Mall (13)	38	10	18	3	3	72
Humbermede (22)	81	44	38	15	11	189
Maple Leaf (29)	42	18	12	8	5	85
Brookhaven-Amesbury (30)	58	36	28	12	6	140
Yorkdale-Glen Park (31)	142	44	65	36	18	305
Clanton Park (33)	51	23	45	8	7	134
Westminster-Branson (35)	64	26	37	13	3	143
Victoria Village (43)	55	7	23	5	2	92
Thorncliffe Park (55)	81	10	23	3	3	120
Old East York (58)	76	9	30	8	0	123
East End-Danforth (62)	149	26	78	22	6	281
South Riverdale (70)	217	29	83	20	23	372
North St.James Town (74)	172	9	72	26	10	289
Kensington-Chinatown (78)	379	19	232	56	29	715
Palmerston-Little Italy (80)	53	22	43	11	6	135
Roncesvalles (86)	125	26	84	32	8	275
Yonge-Eglinton (100)	25	13	11	8	2	59
Forest Hill South (101)	15	23	34	6	1	79
Humewood-Cedarvale (106)	23	18	24	1	6	72
Wexford/Maryvale (119)	214	68	100	31	13	426
Clairlea-Birchmount (120)	204	37	62	28	10	341
Kennedy Park (124)	165	17	41	19	18	260
Ionview (125)	62	5	20	7	1	95
Agincourt South-Malvern West	89	31	93	24	7	244

Highland Creek (134)	58	16	35	6	2	117
Morninside (135)	102	9	18	10	2	141
Averages Moderate crime means	**98**	**24**	**49**	**16**	**7**	**194**
Total Moderate crimes	2850	684	1432	453	212	5631
Total crimes (Table 2)	15635	7392	4118	2816	1109	31070
Table 2 Total crimes- (High + Moderate crime totals)	12019	6494	2365	2252	833	23963
Cluster 3 (low crime means)						
Averages Low crime means (Total crimes- (High crime + Moderate total))/ 101	**119.0**	**64.3**	*23.4*	**22.3**	**8.2**	**237.3**

Table 7: Means of crimes in three cluster solution K-Means