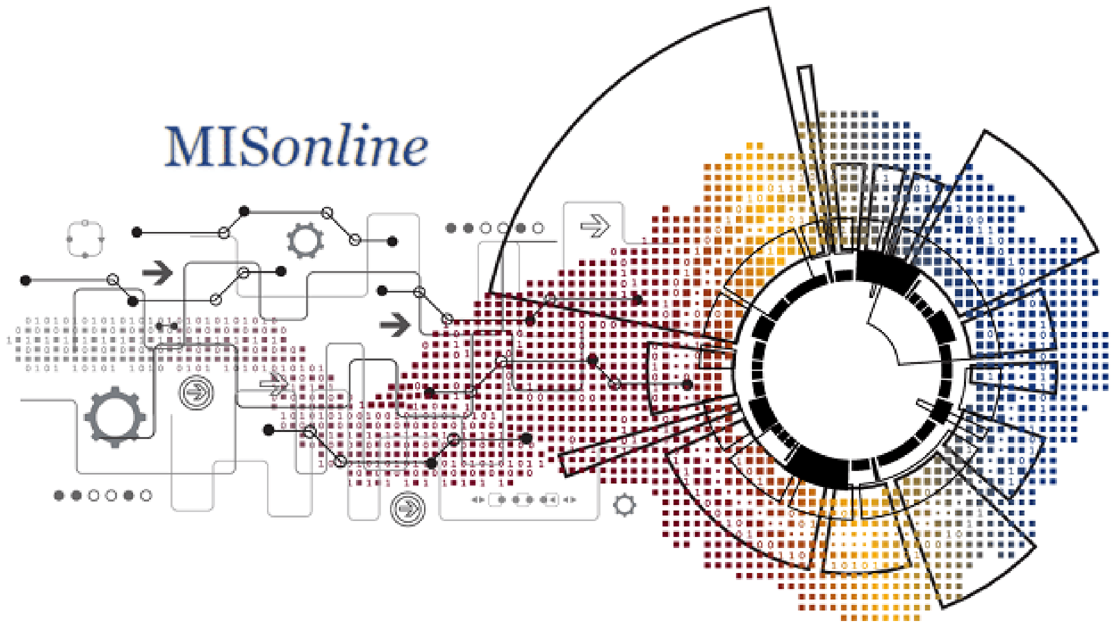


Course: MIS 545

Toronto Crime Data Analysis

Executive summary of results and findings

The major crime classification in 2016 was assault followed by breaking and entering and then auto theft. The next question to be answered was to look at the top crime classification and break it down into types. *What are the different types of crime?* The main classification was assault followed by breaking and entering (B&E) and then Theft of a Motor Vehicle. *Next question was what time of the day did the crimes occur and what were the peak times of crime?* The peak crime hour was found to be around midnight, another peak time is around noon, then again at around 8pm. Assaults had two peak times 11pm and 2pm. Breaking and entering occurred more often in the early mornings. Robberies and auto thefts were more likely to occur in the late evenings and nights. The next question to answer was *where are the top crimes are most likely to occur?* What are the safe neighborhoods and what are the dangerous one? This helped me to highlight some of the most dangerous areas of the city. Cluster 1 indicates neighborhoods with low assault, low auto theft, low break and enter, low robbery and low theft. Cluster 2 indicates neighborhoods with high assault, high auto theft, high break and enter, high robbery and high theft. The most dangerous neighborhood in Toronto was the *Waterfront Communities*. The sprawling downtown area is not only densely packed but also a busy entertainment district. This might explain the higher crime rate. The results indicate a staggering number of violent crimes and arsons. Finally I compared neighborhoods and crime types. This highlighted which areas have a problem with a specific type of crime. *Church-Yonge Corridor* and *Waterfront* had the most break and enter. *West Humber-Clairville* had the most auto theft. What were the safest areas of the city to live in? Our results indicate that the *Malvern, Mount Olive* and *South Parkdale* area where the safest.

The Dataset

A section about the dataset: What is the data about, what are the records and attributes, what kind of pre-processing did it require, etc.

I am going to use various data mining techniques to examine the Toronto Police Service Major Crime Indicators (MCI) data base. The database is available for download at http://data.torontopolice.on.ca/datasets/mci-2016. The information contained in this dataset refers to current Year-to-Date as well as previous full Year End content. Current Year-to-Date data was not available for download so I will be using the 2016 database. The database contains **32,613** records. Each record represents an individual crime report. There are 29 columns.

A glossary of each of these terms and what they mean is provided here. To reduce the complexity of dealing with the full source data I will remove various data as needed.

Data selection and transformation

First task will be to check for duplicated **event_unique_id**. A quick inspection of the data indicated that there were multiple instances of duplicated records. (see below) If any are found they will be removed. After I ran this process, we are now down to **28,147** records. Also some of the crimes may have been reported in 2016 but happened much earlier. We are only interested in crimes that happened in 2016 so we will remove the other reports. We will do this by checking the **occurrenceyear**. The occurrence year ranged from 2000 to 2016. I ran a report to see how many late reported incidents are present. The vast majority of reports were in 2016. A total of 27705 total. These are the reports we are interested in. The rest will be removed. To reduce the complexity of dealing with the full source data I will remove other columns that we do not need. We are now down to 14 columns.

```
> groups_headers
# A tibble: 140 x 6
   Neighbourhood                        Assault `Auto Theft` `Break and Enter` Robbery `Theft Over`
   <fct>                                  <int>        <int>             <int>   <int>        <int>
 1 Agincourt North (129)                     73           16                55      12            4
 2 Agincourt South-Malvern West (128)       112           23                66      18           11
 3 Alderwood (20)                            31           14                25       8            8
 4 Annex (95)                               261           21               131      44           22
 5 Banbury-Don Mills (42)                    67            6                53       8            9
 6 Bathurst Manor (34)                       39           20                28       4            2
 7 Bay Street Corridor (76)                 406           13                92      48           33
 8 Bayview Village (52)                      56           13                33       7            4
 9 Bayview Woods-Steeles (49)                36           16                31      NA            1
10 Bedford Park-Nortown (39)                 26           25                81       8            8
# ... with 130 more rows
> 
```

You can see from this view above that there is qualitative data (neighborhood) and missing values present. Any missing value in the data must be removed or estimated. **The data must be standardized (i.e., scaled) to make variables comparable.** Standardization consists of transforming the variables such that they have mean zero and standard deviation one. We are now ready to begin our analysis of the data. With the data finally cleaned, integrated, selected and transformed, the actual data mining will begin.

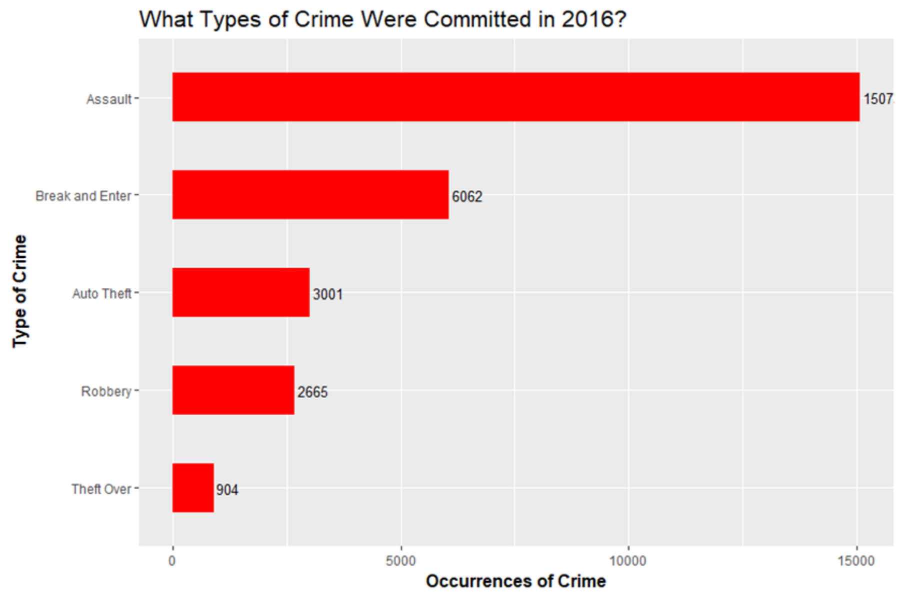
Implications answering these questions will have.

A section about the questions, the implications answering these questions will have, etc.

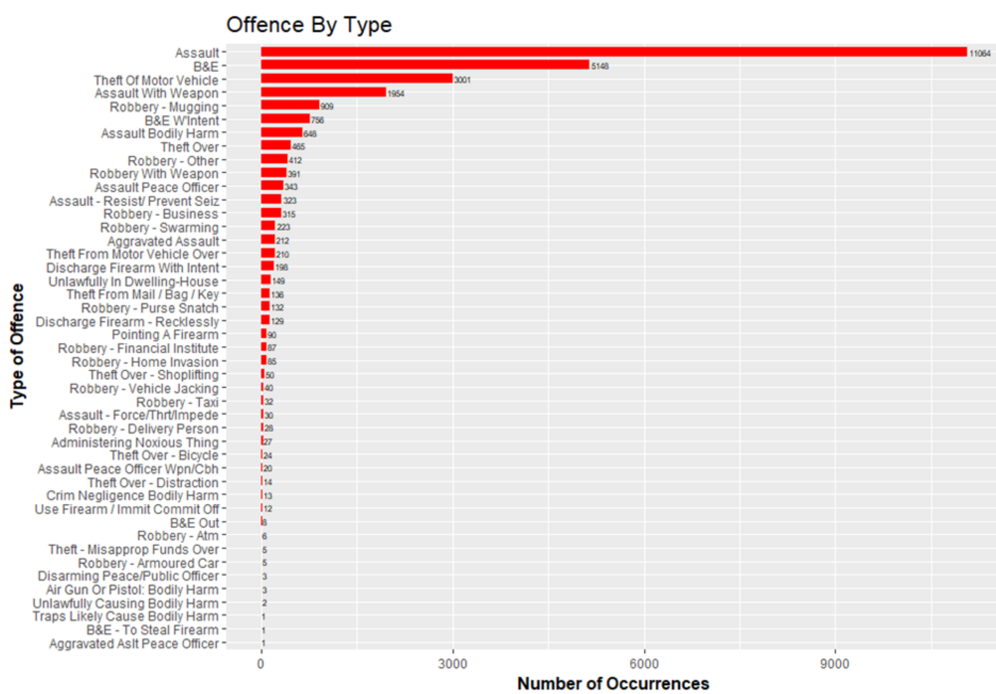
I want to answer three main questions. **Where are the main high crime areas in Toronto? Where are the low crime areas in Toronto and what classification of crime is being committed**. In the process, I will also answer a variety of other questions. I will start with some simple plots of variables I processed using the powerful ggplot2. We will then use k-means clustering. As one of the unsupervised learning algorithms, I will use K-Mean to build models that help me understand the data better. The purpose of unsupervised learning with clustering is to find meaningful relationships in the data, preferably where you could not have seen them otherwise. In addition, I will attempt to use Naïve Bayes to predict a class, given a set of features using probability. Crime analysis and prevention is a useful tool for identifying and analyzing patterns and trends in crime. We will attempt to predict regions, which have high probability for crime occurrence and can visualize crime prone areas. This will help Law enforcement officers with the process of protecting neighborhoods. Using the concept of data mining, we can extract previously unknown, useful information from an unstructured data.

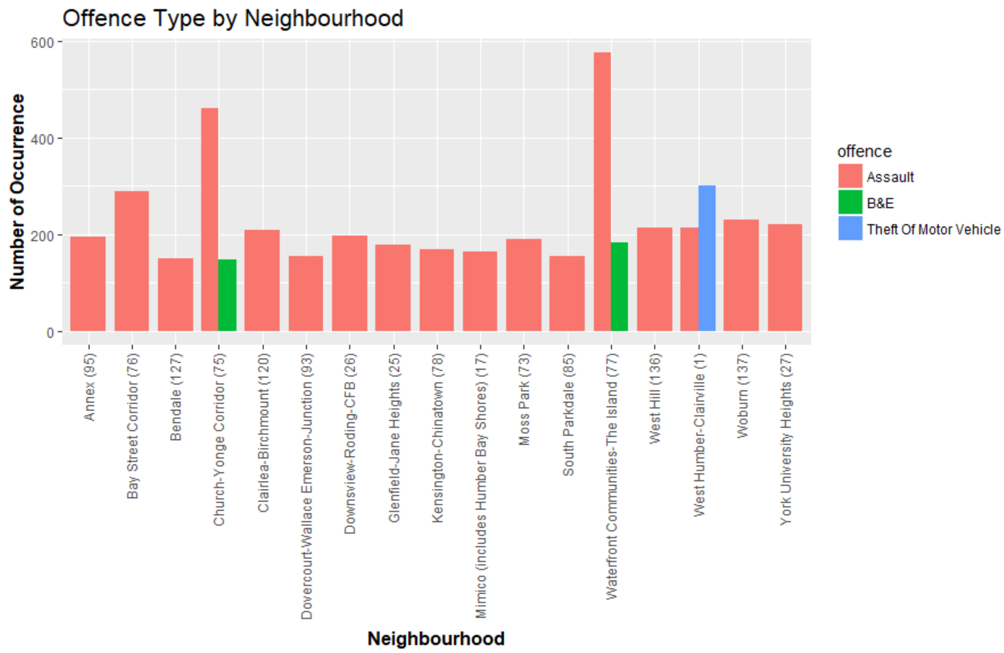
Summary statistics and descriptive analysis of data

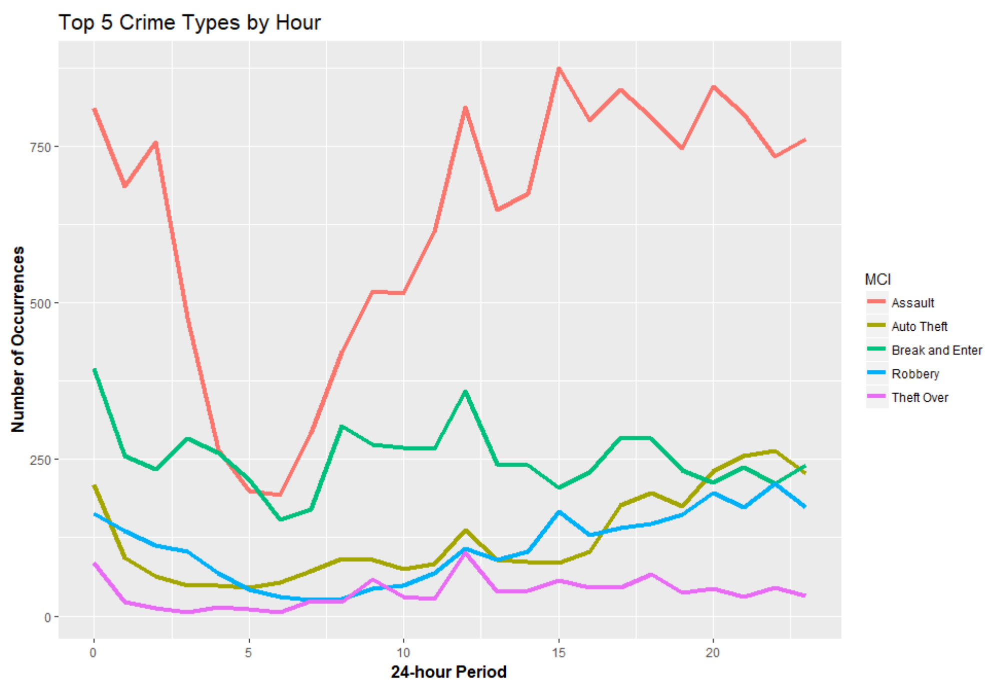
Visualizing data is a powerful way to derive high-level insights about the underlying patterns in the data. To see a few examples, we start with some plots processed using the powerful ggplot2.



Visualizations provide helpful clues as to where we need to look for information. We are interested in MCI (Major Crime Indicators) and Neighborhoods. The MCI classification is is made up of assault, auto theft, break and enter, robbery and theft over.







Results from model executions

Each model needs to answer one specific question that you identified earlier. Models can be classification, clustering, association rule mining, etc. You need to explain each model and justify the operators that you use.

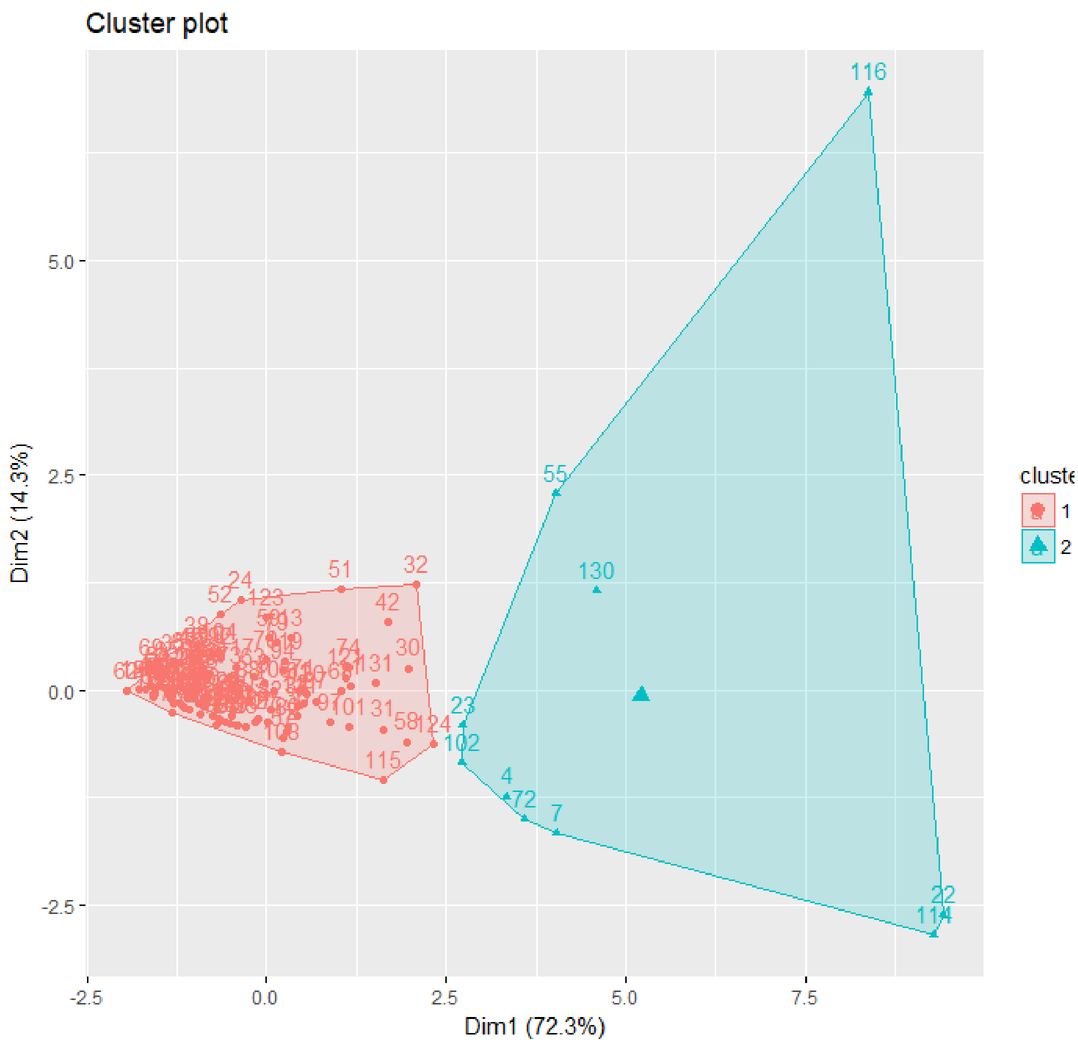
K-means Clustering

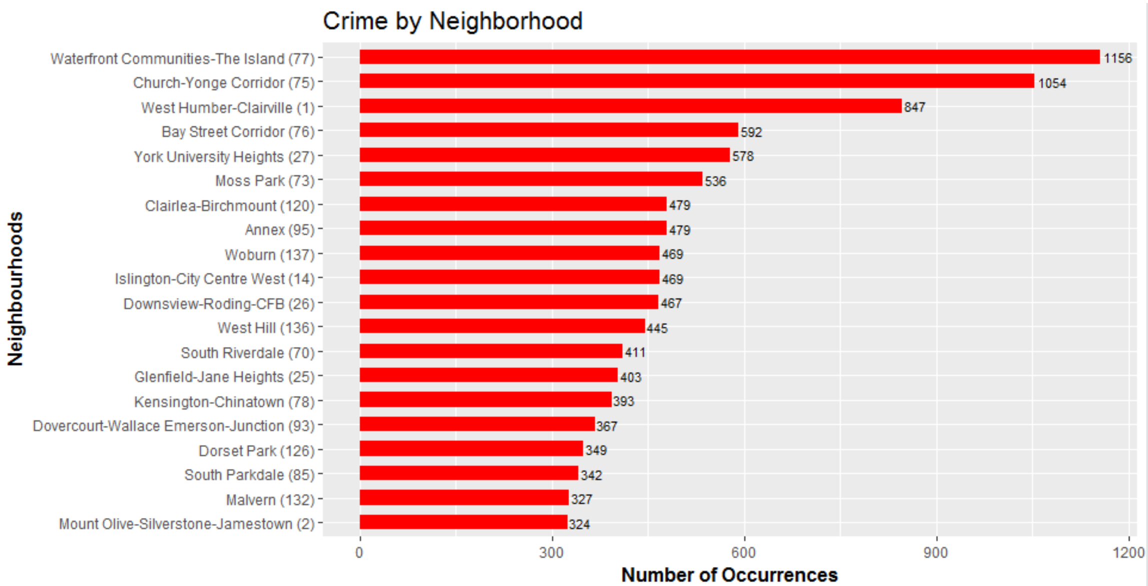
```
K-means clustering with 2 clusters of sizes 10, 121

Cluster means:
    Assault Auto Theft Break and Enter    Robbery Theft Over
1  2.335808   1.639429       2.6337422  2.1521148  2.7689425
2 -0.193042  -0.135490      -0.2176646 -0.1778607 -0.2288382

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

Within cluster sum of squares by cluster:
[1] 170.2395 183.3436
 (between_SS / total_SS =  45.6 %)

Available components:

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

First cluster has 10 neighbourhoods, and the second cluster has 121 neighbourhoods.

```
> str(kc)
List of 9
 $ cluster     : int [1:131] 1 1 1 2 1 1 2 1 1 1 ...
 $ centers     : num [1:2, 1:5] -0.193 2.336 -0.135 1.639 -0.218 ...
  ..- attr(*, "dimnames")=List of 2
  .. ..$ : chr [1:2] "1" "2"
  .. ..$ : chr [1:5] "Assault" "Auto Theft" "Break and Enter" "Robbery" ...
 $ totss       : num 650
 $ withinss    : num [1:2] 183 170
 $ tot.withinss: num 354
 $ betweenss   : num 296
 $ size        : int [1:2] 121 10
 $ iter        : int 1
 $ ifault      : int 0
 - attr(*, "class")= chr "kmeans"
>
```

K-means clustering will enable me to learn groupings of unlabeled data points. Here I will attempt to measure the number of assaults and other indicators. Neighborhoods with a high number of assaults will be grouped together**. In this project the goal of clustering is to assign a cluster to each data point (neighborhood).** I will first partition datapoints (neighborhoods) into k clusters in which each neighborhood belongs to the cluster with the nearest mean, serving as a prototype of the cluster. If we examine the Cluster Means, the negative values mean “lower than most” and positive values mean “higher than most”. **Cluster 1 indicates neighborhoods with low assault, low auto theft, low break and enter, low robbery and low theft. Cluster 2 indicates neighborhoods with high assault, high auto theft, high break and enter, high robbery and high theft**. If we examine the Clustering vector: The first, second and third neighborhoods should all belong to cluster 1, the fourth neighborhood should belong to cluster 2 and so on. Withinss is a Vector of within-cluster sum of squares, one component per cluster. Lower is better. The between-cluster sum of squares. Ideally we want cluster centers far apart from each other. We can also view our results by using fviz_cluster. This provides a nice illustration of the clusters. If there are more than two dimensions (variables) fviz_cluster will perform principal component analysis (PCA) and plot the data points according to the first two principal components that explain the majority of the variance.



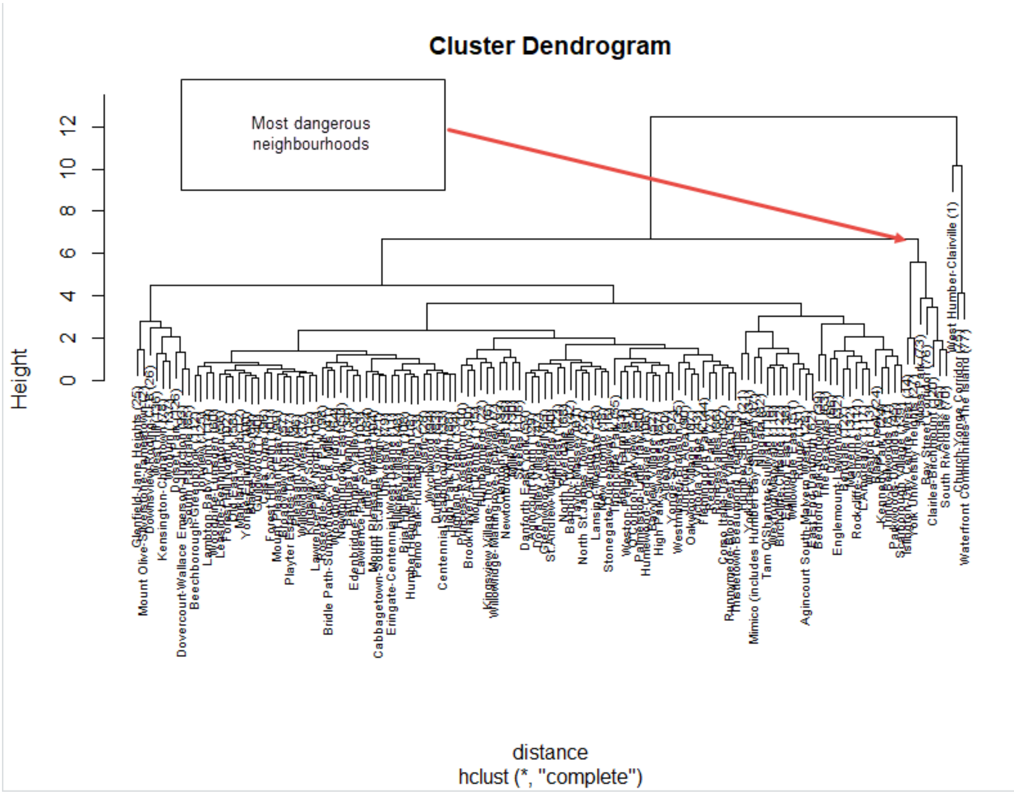


Hierarchical clustering

In hierarchical clustering we do not specify the number of clusters upfront. These were determined by looking at the dendogram after the algorithm had done its work. I will undertake some hierarchical clustering. Hierarchical clustering is a method of cluster analysis which seeks to build a hierarchy of clusters. I will use hierarchical clustering to create a sequence of nested clusters to explore deeper insights from the data.

```
> hc <- hclust(distance)
> hc

Call:
hclust(d = distance)

Cluster method   : complete
Distance         : euclidean
Number of objects: 131

> summary(hc)
            Length Class  Mode
merge       260    -none- numeric
height      130    -none- numeric
order       131    -none- numeric
labels        0    -none- NULL
method        1    -none- character
call          2    -none- call
dist.method   1    -none- character
>
```

The denogram below represents a two-cluster solution; by following the line down through all its branches, we can see the names of the neighborhoods that are included in these two clusters. From the top of the tree, there are two distinct groupings. One group consists of many groups within groups. The other group consists of only a few neighborhoods. **These neighborhoods are high crime rate neighborhoods**.



Naïve Bayes

A sample size is calculated and I randomly decide which ones are training data. The next step is to divide the available data into training and test datasets. The former will be used to train the algorithm and produce a predictive model. The effectiveness of the model will then be tested using the test dataset. An important consideration is that both sets must contain records that are representative of the entire dataset. Next, we invoke the Naive Bayes method from the e1071 package. The first argument uses R's formula notation. In this notation, the dependent variable (to be predicted) appears on the left hand side of the ~ and the independent variables (predictors or features) are on the right hand side. Now that we have a model, we can do some predicting. We do this by feeding our test data into our model and comparing the predicted data with the known ones. The latter is done via the confusion matrix – a table in which true and predicted values for each of the predicted classes are displayed in a matrix format. Below is the model showing crime category by neighborhood.

```
> toronto.model

Naive Bayes Classifier for Discrete Predictors

Call:
naiveBayes.default(x = X, y = Y, laplace = laplace)

A-priori probabilities:
Y
       Assault     Auto Theft Break and Enter        Robbery     Theft Over
    0.56138602     0.09981218      0.19805282     0.10912645     0.03162252

Conditional probabilities:
                 i..X
Y                      [,1]       [,2]
  Assault         -79.38934 0.10447589
  Auto Theft      -79.44047 0.11100928
  Break and Enter -79.38837 0.09791290
  Robbery         -79.40079 0.11050821
  Theft Over      -79.40154 0.09869889

                 Neighbourhood
Y                  Agincourt North (129) Agincourt South-Malvern West (128) Alderwood (20)   Annex (95)
  Assault                   0.0050525741                       0.0078519732   0.0021166189 0.0182985115
  Auto Theft                0.0057603687                       0.0065284178   0.0049923195 0.0076804916
  Break and Enter           0.0090961873                       0.0094832591   0.0038707180 0.0210954132
  Robbery                   0.0049174570                       0.0063224447   0.0031612223 0.0182648402
  Theft Over                0.0060606061                       0.0121212121   0.0084848485 0.0230303030
                 Neighbourhood
Y                  Banbury-Don Mills (42) Bathurst Manor (34) Bay Street Corridor (76) Bayview Village (52)
  Assault                    0.0038918476        0.0023897310             0.0266967090        0.0040284037
  Auto Theft                 0.0023041475        0.0061443932             0.0049923195        0.0042242704
  Break and Enter            0.0087091155        0.0044513257             0.0143216567        0.0056125411
  Robbery                    0.0035124693        0.0014049877             0.0196698279        0.0035124693
  Theft Over                 0.0084848485        0.0024242424             0.0387878788        0.0036363636
                 Neighbourhood
Y                  Bayview Woods-Steeles (49) Bedford Park-Nortown (39) Beechborough-Greenbrook (112) Bendale (127)
  Assault                        0.0021166189              0.0018435068                  0.0030725113  0.0131776594
  Auto Theft                     0.0057603687              0.0088325653                  0.0023041475  0.0107526882
  Break and Enter                0.0056125411              0.0133539772                  0.0017418231  0.0079349719
  Robbery                        0.0000000000              0.0024587285                  0.0017562346  0.0151036178
  Theft Over                     0.0000000000              0.0072727273                  0.0036363636  0.0109090909
                 Neighbourhood
Y                  Birchcliffe-Cliffside (122) Black Creek (24) Blake-Jones (69) Briar Hill-Belgravia (108)
  Assault                         0.0100368701     0.0134507715     0.0028676772               0.0030042332
  Auto Theft                      0.0053763441     0.0149769585     0.0019201229               0.0046082949
  Break and Enter                 0.0110315464     0.0048383975     0.0021288949               0.0048383975
  Robbery                         0.0042149631     0.0105374078     0.0017562346               0.0052687039
  Theft Over                      0.0072727273     0.0072727273     0.0012121212               0.0048484848
```

For the complete model see the appendix.

```
> toronto.model <- naiveBayes(MCI ~ . , data = train)
> summary (toronto.model)
        Length Class  Mode
apriori 5      table  numeric
tables  2      -none- list
levels  5      -none- character
call    4      -none- call
>
> str(toronto.model)
List of 4
 $ apriori: 'table' int [1:5(1d)] 16459 2951 5739 3278 923
  ..- attr(*, "dimnames")=List of 1
  .. ..$ Y: chr [1:5] "Assault" "Auto Theft" "Break and Enter" "Robbery" ...
 $ tables :List of 2
  ..$ i..X         : num [1:5, 1:2] -79.4 -79.4 -79.4 -79.4 -79.4 ...
  .. ..- attr(*, "dimnames")=List of 2
  .. .. ..$ Y    : chr [1:5] "Assault" "Auto Theft" "Break and Enter" "Robbery" ...
  .. .. ..$ i..X: NULL
  ..$ Neighbourhood: table [1:5, 1:140] 0.00462 0.00474 0.00924 0.00458 0.00542 ...
  .. ..- attr(*, "dimnames")=List of 2
  .. .. ..$ Y            : chr [1:5] "Assault" "Auto Theft" "Break and Enter" "Robbery" ...
  .. .. ..$ Neighbourhood: chr [1:140] "Agincourt North (129)" "Agincourt South-Malvern West (128)" "Alderwood (20)" "Annex
(95)" ...
 $ levels : chr [1:5] "Assault" "Auto Theft" "Break and Enter" "Robbery" ...
 $ call   : language naiveBayes.default(x = X, y = Y, laplace = laplace)
 - attr(*, "class")= chr "naiveBayes"
>
```

Model evaluation

K-Means

It is sometimes difficult to decide how many clusters to use. While one solution may be technically correct, the two-cluster solution may seem to give better results. If you increase the number of clusters beyond three, your predictions' success rate starts to break down. It can be seen that as the value of K increases, distortion decreases.

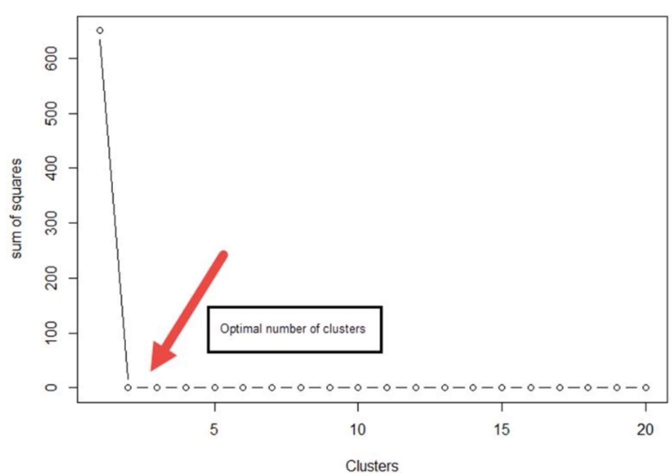
```
> kmeans.totwithinss.k(z, 2)
[1] 353.5831
> kmeans.totwithinss.k(z, 3)
[1] 257.0361
```

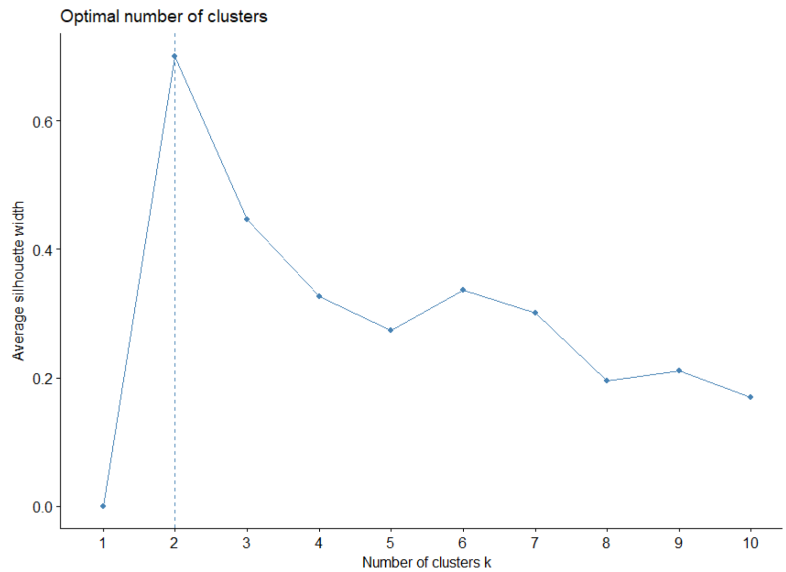
Evaluating the performance of an algorithm requires a label that represents the expected value and a predicted value to compare it with. Remember that when you apply a clustering algorithm to an unsupervised learning model, you do not know what the expected values are — and you don't give labels to the clustering algorithm. The algorithm puts data points into clusters on the basis of which data points are similar to one another; different data points end up in other clusters. **The basic idea behind cluster partitioning methods, such as k-means clustering, is to define clusters such that the total within-cluster sum of square is minimized.**

Elbow method:

We can implement this in R with the following code. The results suggest that 2 is the optimal number of clusters as it appears to be the bend in the knee (or elbow).

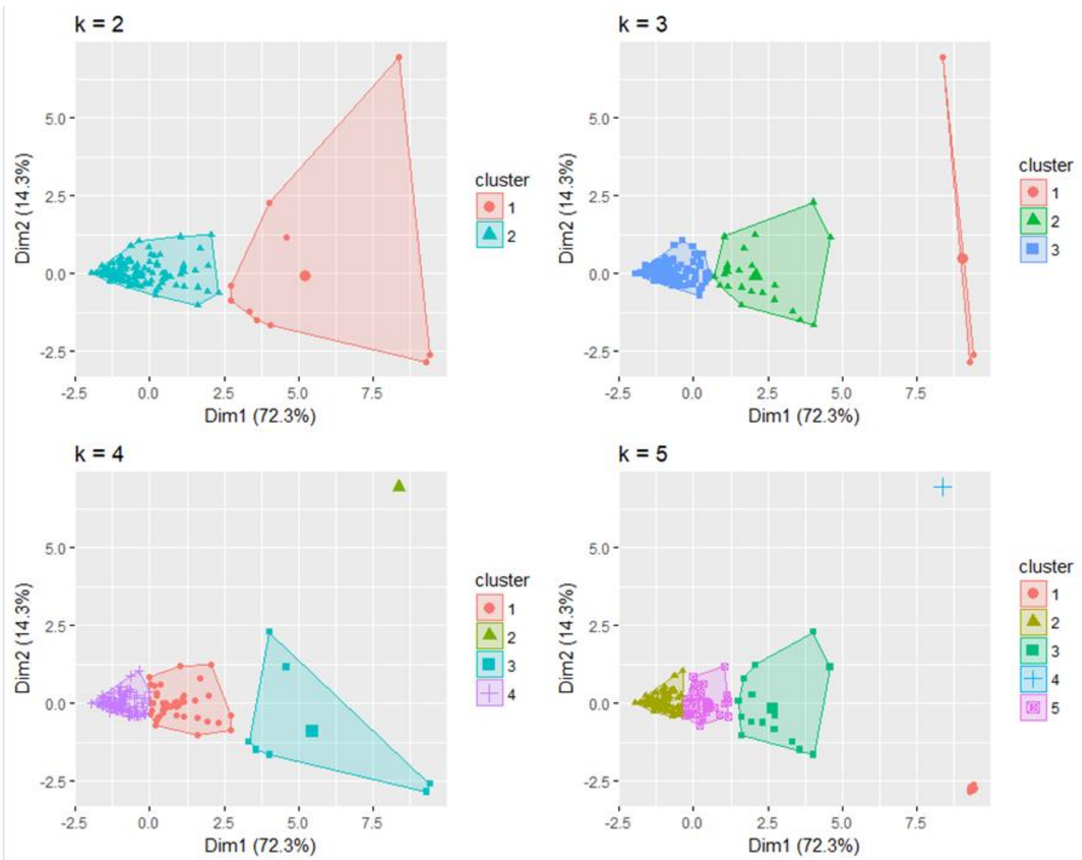
```
> wss <- (nrow(z)-1) * sum(apply(z, 2, var))
> for (i in 2:20) wss[i] <- sum(kmeans(z, centers=i)$withiness)
> plot(1:20, wss, type='b', xlab='Clusters', ylab='sum of squares')
```

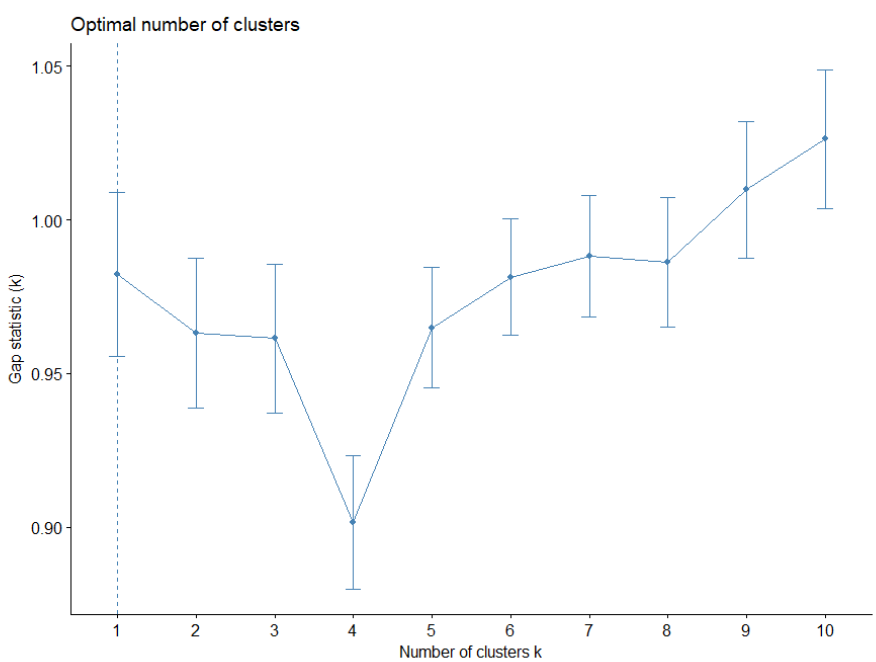




Average Silhouette Method

The average silhouette approach measures the quality of a clustering. It determines how well each object lies within its cluster. A high average silhouette width indicates a good clustering. The average silhouette method computes the average silhouette of observations for different values of k. The optimal number of clusters k is the one that maximizes the average silhouette over a range of possible values for k.2 **The results show that 2 clusters maximize the average silhouette values with 4 clusters coming in as second optimal number of clusters.** Because the number of clusters (k) must be set before we start the algorithm, it is often useful to use several different values of k and examine the differences in the results. We can execute the same process for 3, 4, and 5 clusters, and the results are shown below:





Gap Statistic Method

The gap statistic approach can be applied to any clustering method (i.e. K-means clustering, hierarchical clustering). The gap statistic compares the total intracluster variation for different values of k with their expected values under null reference distribution of the data (i.e. a distribution with no obvious clustering). This one indicates that 4 is the best choice.

Based on the above plots, we can say with confidence that we do not need more than two clusters (centroids). Ways I could improve on the process could include merging neighboring clusters if the resulting cluster's variance is below the threshold. I could also isolate elements that are "far" if a cluster's variance is above the threshold and move some elements between neighboring clusters if it decreases the sum of squared errors.

K-means clustering is a very simple and fast algorithm. Furthermore, it can efficiently deal with very large data sets like the Toronto MCI database. However, there are some weaknesses of the k-means approach. One potential disadvantage of K-means clustering is that it requires us to pre-specify the number of clusters. Hierarchical clustering is an alternative approach which does not require that we commit to a particular choice of clusters. Hierarchical clustering has an added advantage over K-means clustering in that it results in an attractive tree-based representation of the observations, called a dendrogram which I have included in this report. An additional disadvantage of K-means is that it's sensitive to outliers and different results can occur if you change the ordering of your data.

Naïve Bayes

In the confusion matrix (as defined below), the **true values are in columns** and the **predicted values in rows**.

```
> toronto.predict <- predict(toronto.model, test, type = 'class')
> results <- data.frame(Predicted = toronto.predict,  Actual = test[,'MCI'])
> table(results)
                 Actual
Predicted         Assault Auto Theft Break and Enter Robbery Theft Over
  Assault            1777        250             556     292         82
  Auto Theft           47         50              35      19         14
  Break and Enter      44         18              51       6          6
  Robbery               5          7               0       1          2
  Theft Over            0          0               0       0          0
```

A simple measure of efficacy would be the fraction of predictions that the algorithm gets right. The simplest way to calculate this in R is:

```
> mean(toronto.predict==test[,'MCI'])
[1] 0.576027
>
```

The total accuracy is calculated as follows.

$$Accuracy = \frac{TP+TN}{Total} = \frac{1879}{3262} = 0.576 \text{ 58\%}$$

	Assault	Auto Theft	Breaking & Entering	Robbery	Theft Over
Sensitivity	**0.94**	**0.15**	**0.08**	**0.003**	**0.0**
Specificity	**0.15**	**0.96**	**0.98**	**1.0**	**0.1**
FP	**0.85**	**0.30**	**0.03**	**0.005**	**0.0**
FN	**0.05**	**0.85**	**0.92**	**0.997**	**1.0**
Precision	**0.06**	**0.30**	**0.40**	**0.7**	**0.0**
Recall	**0.94**	**0.15**	**0.08**	**.003**	**0.0**
F Score	**0.74**	**0.20**	**0.13**	**.006**	**0.0**

```
> cm
Confusion Matrix and Statistics

                 Reference
Prediction        Assault Auto Theft Break and Enter Robbery Theft Over
  Assault            1777        250             556     292         82
  Auto Theft           47         50              35      19         14
  Break and Enter      44         18              51       6          6
  Robbery               5          7               0       1          2
  Theft Over            0          0               0       0          0

Overall Statistics
                                                   
               Accuracy : 0.576                    
                 95% CI : (0.5589, 0.5931)
    No Information Rate : 0.5742
    P-Value [Acc > NIR] : 0.4231

                  Kappa : 0.0911
 Mcnemar's Test P-Value : <2e-16

Statistics by Class:

                     Class: Assault Class: Auto Theft Class: Break and Enter Class: Robbery
Sensitivity                  0.9487           0.15385                0.07944      0.0031447
Specificity                  0.1505           0.96084                0.97176      0.9952446
Pos Pred Value               0.6009           0.30303                0.40800      0.0666667
Neg Pred Value               0.6852           0.91120                0.81160      0.9023714
Prevalence                   0.5742           0.09963                0.19681      0.0974862
Detection Rate               0.5448           0.01533                0.01563      0.0003066
Detection Prevalence         0.9065           0.05058                0.03832      0.0045984
Balanced Accuracy            0.5496           0.55735                0.52560      0.4991946
                     Class: Theft Over
Sensitivity                    0.00000
Specificity                    1.00000
Pos Pred Value                     NaN
Neg Pred Value                 0.96812
Prevalence                     0.03188
Detection Rate                 0.00000
Detection Prevalence           0.00000
Balanced Accuracy              0.50000
> 
```

My accuracy rate could possibly be improved from 58% by adjusting the classifier's tunable parameters. I could also apply some sort of classifier combination technique (eg, boosting, bagging). In addition I could look at the data used in the project and either add more data, improve my basic parsing, or refine the features I select from the data. Naïve

Bayes is an algorithm that allows us to predict a class, given a set of features using probability. Naïve Bayes operates on the common principle, that every feature being classified is independent of the value of any other feature. Features, however, are not always independent. This can be a disadvantage of using the Naive Bayes algorithm. I suppose this is why it is called Naïve. However, the model I used for this report was relatively simple to understand and build. In addition, it was also easily trained and did not require a huge dataset.

Implications and conclusion

This report has helped to highlight some of the most dangerous and safest areas of Toronto. It also highlighted what types of crimes were committed in these neighborhoods. For example, the most dangerous neighborhood in Toronto was the *Waterfront Communities*. *West Humber-Clairville* had the most vehicle theft. Our analysis indicated these neighborhoods also have high assault rates and a staggering number of violent crimes. The safest areas of the city to live in were *Malvern, Mount Olive and South Parkdale*. I used various approaches to measure the quality of my clustering. The correct number of clusters is often ambiguous and depends on the shape and scale of the distribution of points in a data set and the desired clustering resolution of the user. I attempted to make the optimal choice of clusters that would strike a balance between maximum compression of the data using a single cluster, and maximum accuracy by assigning each data point to its own cluster. My Naïve Bayes classifier had an accuracy of 58%. In future, I may want to focus on my data and the quality of my pre-processing and feature selection to help improve the accuracy. Perhaps identifying and separating out segments could give me increased performance and focus on the elements of the problem that are more difficult to model. Perhaps future research could include comparisons with previous years, observing the characteristics of a particular region over time. In addition, certain crimes types have very different characteristics from others. The 5 MCI (Major Crime Indicators) I used may have been too broad. Crimes like murder, aggravated assault, and rape, may need special attention and much more specific modeling to be really useful. Crime data is not easy to work with. It has both spatial and temporal attributes. Processing them can be a challenging task. The challenge is not limited to handling spatial and temporal data but also deriving information from them at these levels. In other words what information is useful? The purpose behind building these models and analyzing this data is to build a resource that will help law enforcement agencies deploy their limited resources more proactively and efficiently.

Appendices

List of Libraries

library(ggplot2)
library(ggthemes)
library(dplyr)
library(viridis)
library(tidyr)
library(cluster)
library(ggmap)
library(maps)
library(factoextra) # clustering algorithms & visualization
library(tidyverse) # data manipulation
library(cluster) # clustering algorithms

Naive Bayes Classifier for Discrete Predictors

```
Call:
naiveBayes.default(x = X, y = Y, laplace = laplace)

A-priori probabilities:
Y
        Assault      Auto Theft Break and Enter         Robbery      Theft Over
     0.56138602      0.09981218      0.19805282      0.10912645      0.03162252

Conditional probabilities:
                 ï..X
Y                       [,1]       [,2]
  Assault         -79.38934 0.10447589
  Auto Theft      -79.44047 0.11100928
  Break and Enter -79.38837 0.09791290
  Robbery         -79.40079 0.11050821
  Theft Over      -79.40154 0.09869889

                 Neighbourhood
Y                  Agincourt North (129) Agincourt South-Malvern West (128) Alderwood (20)
   Annex (95)
  Assault                  0.0050525741                       0.0078519732   0.0021166189
```

```
 0.0182985115
  Auto Theft              0.0057603687                           0.0065284178   0.0049923195
 0.0076804916
  Break and Enter         0.0090961873                           0.0094832591   0.0038707180
 0.0210954132
  Robbery                 0.0049174570                           0.0063224447   0.0031612223
 0.0182648402
  Theft Over              0.0060606061                           0.0121212121   0.0084848485
 0.0230303030
                 Neighbourhood
Y                 Banbury-Don Mills (42) Bathurst Manor (34) Bay Street Corridor (76) Bay
view Village (52)
  Assault                   0.0038918476        0.0023897310             0.0266967090
     0.0040284037
  Auto Theft                0.0023041475        0.0061443932             0.0049923195
     0.0042242704
  Break and Enter           0.0087091155        0.0044513257             0.0143216567
     0.0056125411
  Robbery                   0.0035124693        0.0014049877             0.0196698279
     0.0035124693
  Theft Over                0.0084848485        0.0024242424             0.0387878788
     0.0036363636
                 Neighbourhood
Y                 Bayview Woods-Steeles (49) Bedford Park-Nortown (39) Beechborough-Green
brook (112) Bendale (127)
  Assault                        0.0021166189               0.0018435068
0.0030725113  0.0131776594
  Auto Theft                     0.0057603687               0.0088325653
0.0023041475  0.0107526882
  Break and Enter                0.0056125411               0.0133539772
0.0017418231  0.0079349719
  Robbery                        0.0000000000               0.0024587285
0.0017562346  0.0151036178
  Theft Over                     0.0000000000               0.0072727273
0.0036363636  0.0109090909
                 Neighbourhood
Y                 Birchcliffe-Cliffside (122) Black Creek (24) Blake-Jones (69) Briar Hil
l-Belgravia (108)
  Assault                         0.0100368701     0.0134507715     0.0028676772
     0.0030042332
  Auto Theft                      0.0053763441     0.0149769585     0.0019201229
```

```
    0.0046082949
  Break and Enter                0.0110315464     0.0048383975     0.0021288949
    0.0048383975
  Robbery                        0.0042149631     0.0105374078     0.0017562346
    0.0052687039
  Theft Over                     0.0072727273     0.0072727273     0.0012121212
    0.0048484848
                 Neighbourhood
Y                  Bridle Path-Sunnybrook-York Mills (41) Broadview North (57) Brookhaven-
Amesbury (30)
  Assault                                   0.0008876144         0.0025945651
 0.0051891301
  Auto Theft                                0.0019201229         0.0019201229
 0.0096006144
  Break and Enter                           0.0077414360         0.0029030385
 0.0032901103
  Robbery                                   0.0003512469         0.0035124693
 0.0077274324
  Theft Over                                0.0048484848         0.0012121212
 0.0024242424
                 Neighbourhood
Y                  Cabbagetown-South St.James Town (71) Caledonia-Fairbank (109) Casa Loma
 (96)
  Assault                                 0.0055305203             0.0044380718   0.00218
48969
  Auto Theft                              0.0030721966             0.0038402458   0.00153
60983
  Break and Enter                         0.0071608283             0.0021288949   0.00329
01103
  Robbery                                 0.0035124693             0.0080786793   0.00035
12469
  Theft Over                              0.0060606061             0.0012121212   0.00727
27273
                 Neighbourhood
Y                  Centennial Scarborough (133) Church-Yonge Corridor (75) Clairlea-Birchm
ount (120) Clanton Park (33)
  Assault                        0.0032090673             0.0435613819              0.
0187081797      0.0032773454
  Auto Theft                     0.0019201229             0.0122887865              0.
0168970814      0.0176651306
  Break and Enter                0.0036771821             0.0280627056              0.
```

```
0149022644      0.0067737565
  Robbery                         0.0042149631              0.0509308044              0.
0122936424      0.0021074816
  Theft Over                      0.0012121212              0.0472727273              0.
0169696970      0.0060606061
                 Neighbourhood
Y                 Cliffcrest (123) Corso Italia-Davenport (92) Danforth (66) Danforth Eas
t York (59)
  Assault             0.0058719104                0.0049842961  0.0058036324
0.0039601256
  Auto Theft          0.0030721966                0.0026881720  0.0026881720
0.0034562212
  Break and Enter     0.0089026514                0.0019353590  0.0067737565
0.0079349719
  Robbery             0.0063224447                0.0084299262  0.0105374078
0.0031612223
  Theft Over          0.0024242424                0.0000000000  0.0060606061
0.0036363636
                 Neighbourhood
Y                 Don Valley Village (47) Dorset Park (126) Dovercourt-Wallace Emerson-Ju
nction (93)
  Assault                    0.0053256862      0.0107196504
0.0141335518
  Auto Theft                 0.0049923195      0.0157450077
0.0103686636
  Break and Enter            0.0092897232      0.0135475131
0.0116121541
  Robbery                    0.0035124693      0.0196698279
0.0122936424
  Theft Over                 0.0036363636      0.0157575758
0.0121212121
                 Neighbourhood
Y                 Downsview-Roding-CFB (26) Dufferin Grove (83) East End-Danforth (62) Ed
enbridge-Humber Valley (9)
  Assault                      0.0180253994        0.0045746279            0.0066912468
              0.0010924485
  Auto Theft                   0.0257296467        0.0015360983            0.0072964670
              0.0057603687
  Break and Enter              0.0087091155        0.0044513257            0.0092897232
              0.0069672924
  Robbery                      0.0126448894        0.0049174570            0.0091324201
```

```
              0.0003512469
  Theft Over                  0.0109090909        0.0036363636           0.0096969697
              0.0024242424
                 Neighbourhood
Y                 Eglinton East (138) Elms-Old Rexdale (5) Englemount-Lawrence (32)
  Assault              0.0093540899         0.0042332377             0.0065546907
  Auto Theft           0.0057603687         0.0046082949             0.0069124424
  Break and Enter      0.0090961873         0.0019353590             0.0116121541
  Robbery              0.0080786793         0.0052687039             0.0136986301
  Theft Over           0.0072727273         0.0000000000             0.0060606061
                 Neighbourhood
Y                 Eringate-Centennial-West Deane (11) Etobicoke West Mall (13) Flemingdon
 Park (44)
  Assault                             0.0030725113             0.0029359552         0.
0095589239
  Auto Theft                          0.0096006144             0.0042242704         0.
0015360983
  Break and Enter                     0.0042577898             0.0029030385         0.
0023224308
  Robbery                             0.0031612223             0.0014049877         0.
0080786793
  Theft Over                          0.0024242424             0.0000000000         0.
0024242424
                 Neighbourhood
Y                 Forest Hill North (102) Forest Hill South (101) Glenfield-Jane Heights
(25) Greenwood-Coxwell (65)
  Assault                   0.0012290045            0.0002048341                0.016455
0048          0.0055987983
  Auto Theft                0.0023041475            0.0026881720                0.022657
4501          0.0042242704
  Break and Enter           0.0030965744            0.0034836462                0.004838
3975          0.0098703309
  Robbery                   0.0017562346            0.0003512469                0.019318
5810          0.0024587285
  Theft Over                0.0024242424            0.0036363636                0.008484
8485          0.0036363636
                 Neighbourhood
Y                 Guildwood (140) Henry Farm (53) High Park-Swansea (87) High Park North
(88) Highland Creek (134)
  Assault              0.0021848969    0.0058719104            0.0036870135         0.004574
6279         0.0043015158
```

```
  Auto Theft         0.0015360983    0.0038402458           0.0038402458          0.001920
1229         0.0026881720
  Break and Enter    0.0017418231    0.0027095026           0.0058060770          0.003870
7180         0.0042577898
  Robbery            0.0024587285    0.0042149631           0.0028099754          0.003863
7162         0.0038637162
  Theft Over         0.0012121212    0.0036363636           0.0048484848          0.002424
2424         0.0000000000
                 Neighbourhood
Y                 Hillcrest Village (48) Humber Heights-Westmount (8) Humber Summit (21)
Humbermede (22)
  Assault                   0.0027311211                0.0020483408       0.0083299194
   0.0049842961
  Auto Theft                0.0076804916                0.0061443932       0.0176651306
   0.0130568356
  Break and Enter           0.0044513257                0.0034836462       0.0061931488
   0.0038707180
  Robbery                   0.0045662100                0.0035124693       0.0077274324
   0.0063224447
  Theft Over                0.0024242424                0.0048484848       0.0169696970
   0.0000000000
                 Neighbourhood
Y                 Humewood-Cedarvale (106) Ionview (125) Islington-City Centre West (14)
Junction Area (90)
  Assault                     0.0034139014  0.0039601256                    0.0120852110
      0.0062815786
  Auto Theft                  0.0042242704  0.0023041475                    0.0395545315
      0.0057603687
  Break and Enter             0.0054190052  0.0019353590                    0.0172246952
      0.0052254693
  Robbery                     0.0038637162  0.0028099754                    0.0119423955
      0.0035124693
  Theft Over                  0.0060606061  0.0036363636                    0.0351515152
      0.0048484848
                 Neighbourhood
Y                 Keelesdale-Eglinton West (110) Kennedy Park (124) Kensington-Chinatown
(78)
  Assault                           0.0035504575       0.0131776594              0.015703
9465
  Auto Theft                        0.0049923195       0.0038402458              0.008448
5407
```

```
  Break and Enter                   0.0027095026       0.0071608283              0.011805
6900
  Robbery                           0.0080786793       0.0094836670              0.014401
1240
  Theft Over                        0.0000000000       0.0048484848              0.013333
3333
                 Neighbourhood
Y                 Kingsview Village-The Westway (6) Kingsway South (15) L'Amoreaux (117)
Lambton Baby Point (114)
  Assault                              0.0076471392        0.0013655606     0.0096954800
          0.0008876144
  Auto Theft                           0.0138248848        0.0026881720     0.0080645161
          0.0003840246
  Break and Enter                      0.0058060770        0.0040642539     0.0085155796
          0.0013547513
  Robbery                              0.0094836670        0.0017562346     0.0126448894
          0.0017562346
  Theft Over                           0.0048484848        0.0012121212     0.0036363636
          0.0012121212
                 Neighbourhood
Y                 Lansing-Westgate (38) Lawrence Park North (105) Lawrence Park South (10
3) Leaside-Bennington (56)
  Assault                  0.0060767445              0.0010241704              0.00170695
07            0.0015021166
  Auto Theft               0.0080645161              0.0038402458              0.00614439
32            0.0011520737
  Break and Enter          0.0079349719              0.0036771821              0.00599961
29            0.0040642539
  Robbery                  0.0031612223              0.0007024939              0.00000000
00            0.0024587285
  Theft Over               0.0060606061              0.0024242424              0.00121212
12            0.0012121212
                 Neighbourhood
Y                 Little Portugal (84) Long Branch (19) Malvern (132) Maple Leaf (29) Mar
kland Wood (12)
  Assault                 0.0059401884     0.0032773454  0.0121534890    0.0019117848
   0.0010241704
  Auto Theft              0.0026881720     0.0034562212  0.0099846390    0.0034562212
   0.0046082949
  Break and Enter         0.0059996129     0.0048383975  0.0090961873    0.0030965744
   0.0025159667
```

```
  Robbery                 0.0028099754     0.0014049877  0.0158061117    0.0021074816
   0.0024587285
  Theft Over              0.0024242424     0.0000000000  0.0060606061    0.0000000000
   0.0012121212
                 Neighbourhood
Y                 Milliken (130) Mimico (includes Humber Bay Shores) (17) Morningside (13
5) Moss Park (73)
  Assault           0.0041649597                                 0.0120169330      0.00641813
46   0.0204834084
  Auto Theft        0.0096006144                                 0.0096006144      0.00345622
12   0.0046082949
  Break and Enter   0.0098703309                                 0.0077414360      0.00232243
08   0.0193535901
  Robbery           0.0094836670                                 0.0038637162      0.00175623
46   0.0302072357
  Theft Over        0.0109090909                                 0.0121212121      0.00000000
00   0.0121212121
                 Neighbourhood
Y                 Mount Dennis (115) Mount Olive-Silverstone-Jamestown (2) Mount Pleasant
 East (99)
  Assault               0.0058036324                              0.0131093814             0.
0019800628
  Auto Theft            0.0069124424                              0.0099846390             0.
0026881720
  Break and Enter       0.0056125411                              0.0046448616             0.
0036771821
  Robbery               0.0115911486                              0.0249385318             0.
0010537408
  Theft Over            0.0012121212                              0.0036363636             0.
0024242424
                 Neighbourhood
Y                 Mount Pleasant West (104) New Toronto (18) Newtonbrook East (50) Newton
brook West (36) Niagara (82)
  Assault                      0.0068960808      0.0059401884           0.0035504575
   0.0082616414 0.0082616414
  Auto Theft                   0.0015360983      0.0030721966           0.0007680492
   0.0153609831 0.0049923195
  Break and Enter              0.0063866847      0.0073543642           0.0059996129
   0.0085155796 0.0102574027
  Robbery                      0.0052687039      0.0052687039           0.0021074816
   0.0052687039 0.0021074816
```

```
  Theft Over                   0.0048484848     0.0048484848          0.0024242424
   0.0072727273 0.0157575758
                 Neighbourhood
Y                 North Riverdale (68) North St.James Town (74) O'Connor-Parkview (54) Oa
kridge (121)
  Assault               0.0040966817              0.0075105831            0.0060767445
0.0053939642
  Auto Theft            0.0007680492              0.0049923195            0.0023041475
0.0038402458
  Break and Enter       0.0073543642              0.0083220437            0.0052254693
0.0077414360
  Robbery               0.0084299262              0.0066736916            0.0073761855
0.0066736916
  Theft Over            0.0060606061              0.0072727273            0.0060606061
0.0084848485
                 Neighbourhood
Y                 Oakwood Village (107) Old East York (58) Palmerston-Little Italy (80) P
arkwoods-Donalda (45)
  Assault                0.0065546907       0.0015021166                 0.0031407893
        0.0082616414
  Auto Theft             0.0072964670       0.0019201229                 0.0034562212
        0.0069124424
  Break and Enter        0.0021288949       0.0036771821                 0.0040642539
        0.0065802206
  Robbery                0.0035124693       0.0017562346                 0.0049174570
        0.0063224447
  Theft Over             0.0024242424       0.0012121212                 0.0048484848
        0.0024242424
                 Neighbourhood
Y                  Pelmo Park-Humberlea (23) Playter Estates-Danforth (67) Pleasant View
(46) Princess-Rosethorn (10)
  Assault                    0.0025262870                  0.0032090673       0.0021848
969             0.0019800628
  Auto Theft                 0.0072964670                  0.0023041475       0.0019201
229             0.0080645161
  Break and Enter            0.0042577898                  0.0027095026       0.0036771
821             0.0054190052
  Robbery                    0.0035124693                  0.0021074816       0.0014049
877             0.0098349139
  Theft Over                 0.0024242424                  0.0024242424       0.0036363
636             0.0048484848
```

```
                 Neighbourhood
Y                  Regent Park (72) Rexdale-Kipling (4) Rockcliffe-Smythe (111) Roncesvall
es (86)
  Assault              0.0063498566        0.0022531749            0.0080568073      0.006
5546907
  Auto Theft           0.0007680492        0.0038402458            0.0092165899      0.003
8402458
  Break and Enter      0.0042577898        0.0011612154            0.0056125411      0.006
7737565
  Robbery              0.0056199508        0.0021074816            0.0133473832      0.007
3761855
  Theft Over           0.0024242424        0.0000000000            0.0048484848      0.002
4242424
                 Neighbourhood
Y                  Rosedale-Moore Park (98)  Rouge (131) Runnymede-Bloor West Village (89)
  Rustic (28)
  Assault                      0.0021166189 0.0090809777                      0.0034139014
 0.0051891301
  Auto Theft                   0.0019201229 0.0061443932                      0.0026881720
 0.0042242704
  Break and Enter              0.0073543642 0.0112250823                      0.0040642539
 0.0030965744
  Robbery                      0.0042149631 0.0087811732                      0.0070249385
 0.0035124693
  Theft Over                   0.0060606061 0.0157575758                      0.0012121212
 0.0012121212
                 Neighbourhood
Y                  Scarborough Village (139) South Parkdale (85) South Riverdale (70) St.A
ndrew-Windfields (40)
  Assault                       0.0105830944        0.0129045473         0.0119486549
        0.0034821794
  Auto Theft                    0.0026881720        0.0053763441         0.0103686636
        0.0023041475
  Break and Enter               0.0048383975        0.0123862977         0.0253532030
        0.0096767950
  Robbery                       0.0073761855        0.0112399017         0.0136986301
        0.0077274324
  Theft Over                    0.0012121212        0.0096969697         0.0242424242
        0.0096969697
                 Neighbourhood
Y                  Steeles (116) Stonegate-Queensway (16) Tam O'Shanter-Sullivan (118) Tay
```

```
lor-Massey (61)
  Assault          0.0024580090             0.0045746279                 0.0059401884
   0.0071691930
  Auto Theft       0.0103686636             0.0049923195                 0.0088325653
   0.0034562212
  Break and Enter  0.0075479001             0.0079349719                 0.0092897232
   0.0071608283
  Robbery          0.0045662100             0.0042149631                 0.0080786793
   0.0073761855
  Theft Over       0.0024242424             0.0084848485                 0.0096969697
   0.0084848485
                 Neighbourhood
Y                 The Beaches (63) Thistletown-Beaumond Heights (3) Thorncliffe Park (55)
 Trinity-Bellwoods (81)
  Assault             0.0071691930                     0.0030042332          0.0051891301
           0.0128362693
  Auto Theft          0.0034562212                     0.0053763441          0.0007680492
           0.0099846390
  Break and Enter     0.0162570157                     0.0038707180          0.0040642539
           0.0094832591
  Robbery             0.0066736916                     0.0070249385          0.0042149631
           0.0091324201
  Theft Over          0.0024242424                     0.0024242424          0.0084848485
           0.0048484848
                 Neighbourhood
Y                 University (79) Victoria Village (43) Waterfront Communities-The Island
 (77) West Hill (136)
  Assault            0.0052574082          0.0073057490                           0.05209
61355    0.0232828076
  Auto Theft         0.0015360983          0.0065284178                           0.01344
08602    0.0053763441
  Break and Enter    0.0067737565          0.0030965744                           0.03580
41417    0.0104509386
  Robbery            0.0031612223          0.0038637162                           0.02318
22972    0.0210748156
  Theft Over         0.0048484848          0.0048484848                           0.06181
81818    0.0072727273
                 Neighbourhood
Y                 West Humber-Clairville (1) Westminster-Branson (35) Weston-Pellam Park
(91) Weston (113)
  Assault                      0.0188447358             0.0087395876            0.004369
```

```
7938 0.0102417042
  Auto Theft                0.1033026114             0.0084485407             0.004224
2704 0.0107526882
  Break and Enter           0.0212889491             0.0036771821             0.002322
4308 0.0050319334
  Robbery                   0.0323147172             0.0042149631             0.003863
7162 0.0119423955
  Theft Over                0.0436363636             0.0096969697             0.008484
8485 0.0036363636
                 Neighbourhood
Y                 Wexford/Maryvale (119) Willowdale East (51) Willowdale West (37)
  Assault               0.0094906459         0.0097637580         0.0024580090
  Auto Theft            0.0096006144         0.0145929339         0.0023041475
  Break and Enter       0.0112250823         0.0123862977         0.0050319334
  Robbery               0.0059711978         0.0105374078         0.0014049877
  Theft Over            0.0096969697         0.0121212121         0.0024242424
                 Neighbourhood
Y                 Willowridge-Martingrove-Richview (7) Woburn (137) Woodbine-Lumsden (60)
 Woodbine Corridor (64)
  Assault                              0.0047111839 0.0216441349         0.0008876144
         0.0024580090
  Auto Theft                           0.0149769585 0.0092165899         0.0007680492
         0.0019201229
  Break and Enter                      0.0058060770 0.0129669054         0.0013547513
         0.0063866847
  Robbery                              0.0154548648 0.0186160871         0.0007024939
         0.0007024939
  Theft Over                           0.0048484848 0.0121212121         0.0072727273
         0.0060606061
                 Neighbourhood
Y                 Wychwood (94) Yonge-Eglinton (100) Yonge-St.Clair (97) York University
Heights (27)
  Assault          0.0032090673         0.0019117848        0.0029359552
0.0206199645
  Auto Theft       0.0057603687         0.0011520737        0.0026881720
0.0337941628
  Break and Enter  0.0048383975         0.0025159667        0.0029030385
0.0149022644
  Robbery          0.0028099754         0.0021074816        0.0024587285
0.0214260625
  Theft Over       0.0036363636         0.0024242424        0.0072727273
```

0.0290909091

	Neighbourhood
Y	Yorkdale-Glen Park (31)
Assault	0.0092175338
Auto Theft	0.0138248848
Break and Enter	0.0112250823
Robbery	0.0059711978
Theft Over	0.0206060606

References:

Crime Analysis using K-Means Clustering, *International Journal of Computer Applications* (0975 – 8887), Volume 83 – No4, December 2013,
http://research.ijcaonline.org/volume83/number4/pxc3892579.pdf

Journal of Computational and Applied Mathematics, Volume 20, November 1987, Pages 53-65
Silhouettes: A graphical aid to the interpretation and validation of cluster analysis
Author links open overlay panelPeter J.Rousseeuw
https://www.sciencedirect.com/science/article/pii/0377042787901257?via%3Dihub

Cluster Analysis with R, Gabriel Martos,
https://rpubs.com/gabrielmartos/ClusterAnalysis

Exploring, Clustering, and Mapping Toronto's Crimes, R-BLOGGERS, November 2, 2017, By Susan Li
https://www.r-bloggers.com/exploring-clustering-and-mapping-torontos-crimes/

Analysis and Prediction of Crimes by Clustering and Classification, (IJARAI) International Journal of Advanced Research in Artificial Intelligence, Vol. 4, No.8, 2015
https://pdfs.semanticscholar.org/3643/74119cd633ac6396f81959700912acdf30ee.pdf

Crime Series Identification and Clustering, Michael D. Porter, 2015-09-19
https://cran.r-project.org/web/packages/crimelinkage/vignettes/crimeclustering.html

Crime Analyses Using R, Anindya Sengupta*, Madhav Kumar*, Shreyes Upadhyay{
https://irgn452.files.wordpress.com/2016/03/3-s2-0-b9780124115118000141-main.pdf
UC Business Analytics R Programming Guide, K-means Cluster Analysis
https://uc-r.github.io/kmeans_clustering#silo

Eight to Late, *Sensemaking and Analytics for Organizations*, A gentle introduction to Naïve Bayes classification using R
https://eight2late.wordpress.com/2015/11/06/a-gentle-introduction-to-naive-bayes-classification-using-r/