Toronto's Underpriviledged Neighbourhoods are Close to Each Other

Ke-Li Chiu & Diego Mamanche Castellanos

29/02/2020

Abstract

Abstract nnnnn nnnnnn

Introduction

There are 140 neighbourhoods officially recognized by the City of Toronto and the divisions are used for internal planning purposes. Many shorcomings of economic or social policy planning and making come from the over-reliance on "one-size-fits-all" approach that overlooks their differences in socioeconomic status. However, customizing policies for 140 neighbourhoods is also improbable. Can we then create groups of neighbourhoods that share similar socioeconomic traits so policy makers can plan strategically and efficiently? Can we identify the group of neighbourhoods that are vulnerable in their social and economic development so they receive more attention from the Government? In this study, a clustering analysis is conducted to be used to define neighborhood types and display their characteristics and spatial patterns.

Troubling differences in poverty, education and diversity exist among neighbourhoods in the City of Toronto. According to David Hulchanski, a professor of the University of Toronto who uses the 2016 census to create demographic charts of Toronto, the city is segregated by race and income. For example, visible minorities are concentrated in low-income neighbourhoods and white residents are dominating affluent areas in numbers far higher than their share of the population. The segregation pointed out by Hulchanski leads to our choices of traits we use to conduct the clustering analysis for the neighbourhoods — low-income rate, population percentage of residents who have no education certificate, and visible minority population percentage. We intend to then articulate the social stratification and spatial stratification in the City of Toronto.

Dataset and Method Description

Limitation of the Dataset and Approach

In 2016, the long form Census became mandatory once again, and was distributed to one out of every four households (25%). Income data were gathered solely by linking with administrative data (Canada Revenue Agency). Census data is subject to error such as coverage, non-response, and sampling errors. Moreover, althought the Census is mandatory, it is still likely to be affected by Response Bias because people do not always provide accurate or truthful information. This is especially common in collection of income data. Correct income data are often more difficult to obtain because the richest households are more prompted to underreport their incomes. Therefore, the income inequality based on data from household surveys are likely to be underestimated.

Another limitation lies in the lack of time-series data to be compared with the data sourced from 2016. This prevents a holistic observation in the change of socioeconomic dynamic in the neighbourhoods across time. The neghbourhoods may not be assigned to the same clusters based on data sourced from past years. Such shifts would provide insights that is not available in this study.

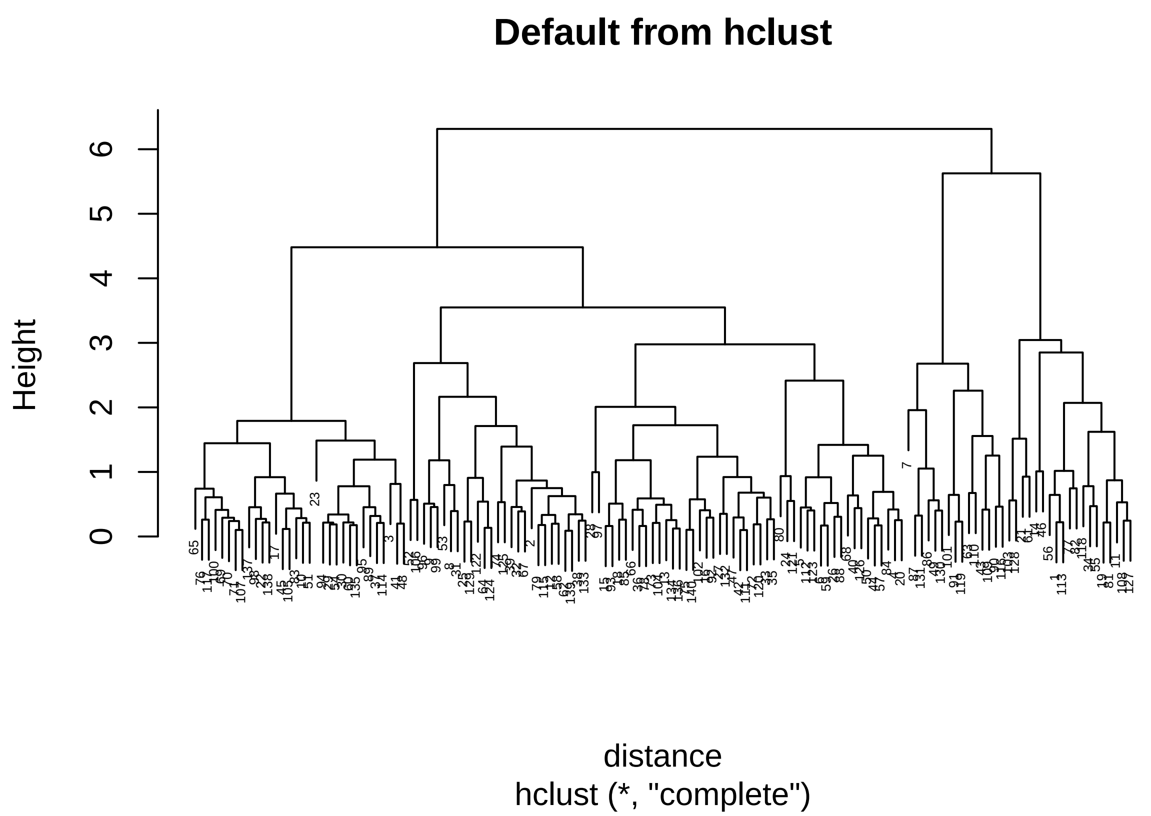
Ethical Considerations

Features such as race or income can be highly predictive for certain problems. The three variables in poverty, education, and visible minority population in this clustering analysis does not provide details of underlying contributors for the social division in the neighbourhoods. Without further investigation, the study possibley shapes the categories of perception and classification through which readers internalize social divisions. These categories shape the way we envision class structure and social problems associated with the neighbourhoods and paved the way for implicit stigmatization. More crucially, the incorporation of these categories structures decisions to buy property and, for parents, to send children to the local schools.

Hierarchical Clustering of Toronto's Neighbourhoods

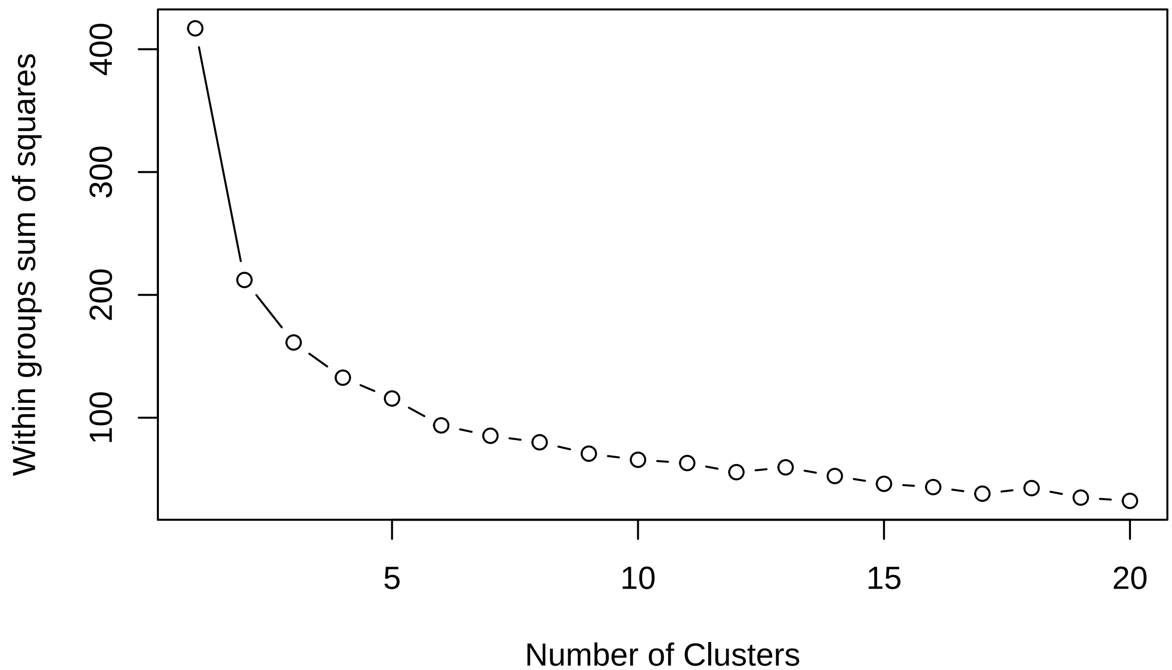
Calculating distance between observations (Neighbourhoods)

Firstly, Euclidean Distance Between Toronto's Neighbourhoods is calculated and used to construct a dendrogram. The height of the branch points indicates how similar or different the neighbourhoods are from each other; the greater the height, the greater the difference.



Determine optimal number of clusters

The Elbow Method is used to determine the optimal number of clusters. Viewing the Scree Plot (Figure X), the "elbow" on the arm is "2" on the x-axis. Therefore, the neighbourhoods will be assigned to two different clusters.



Assessing Characteristics of Neighbourhood Clusters

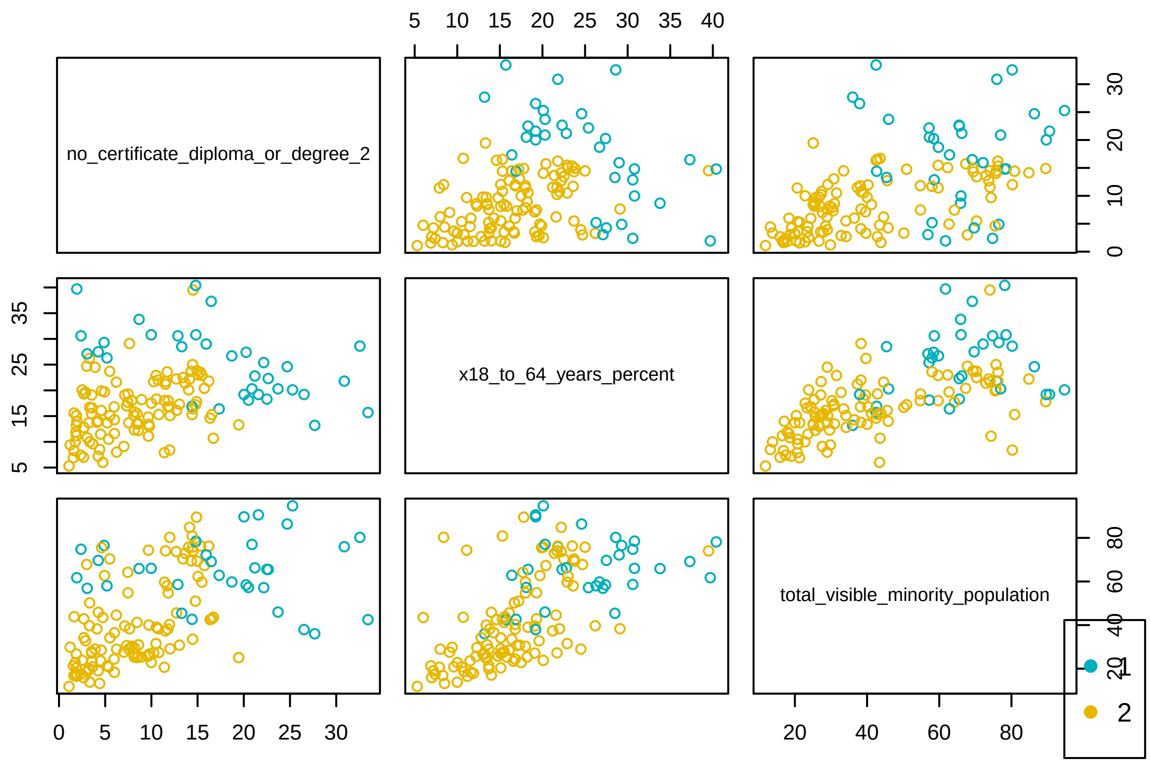
The two clusters has different numbers of members. In Group 1, there are 34 neighbourhoods while the other 106 are assigned to Group 2. Table x. showcases the distances between the means of the two clusters with normalized data values. The distinction between the two clusters is clear—all values in Group 1 are positive and all values in Group 2 are all negative. The positive values in Group 1 indicates its characteristics in higher percentage in poverty, lack of education and visibile minority population.

```
## member
##   1   2
##  34 106
```

```
##   Group.1 no_certificate_diploma_or_degree_2 x18_to_64_years_percent
## 1       1                          0.9872369               1.0508852
## 2       2                         -0.3166609              -0.3370764
##   total_visible_minority_population
## 1                         0.9036849
## 2                        -0.2898612
```

Clusters Distribution

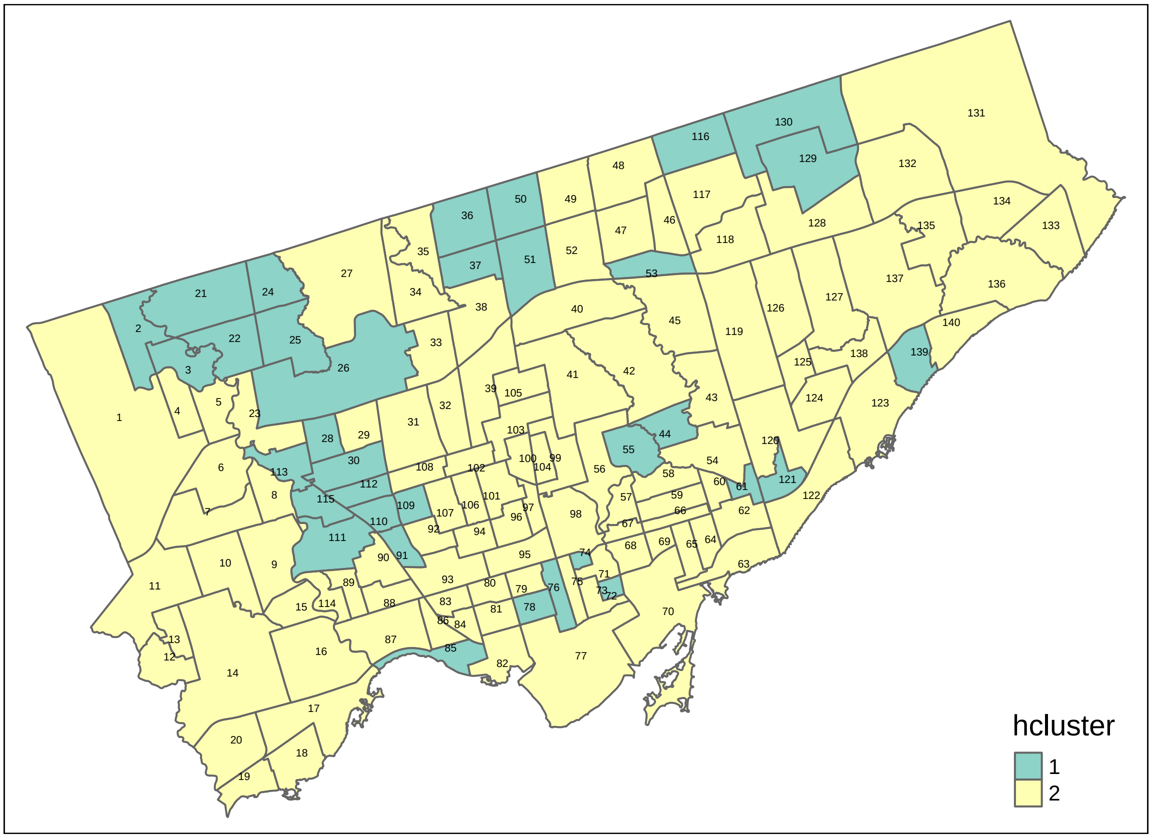
The grouped scattor plots shows that there is a pattern of neighbourhood distribution within the three selected variables. In general, the 34 neighbourhoods in Group 1 are located at the top-right area of the graphs, which visualizes its characteristics in higher percentage in poverty, lack of education and visibile minority population.



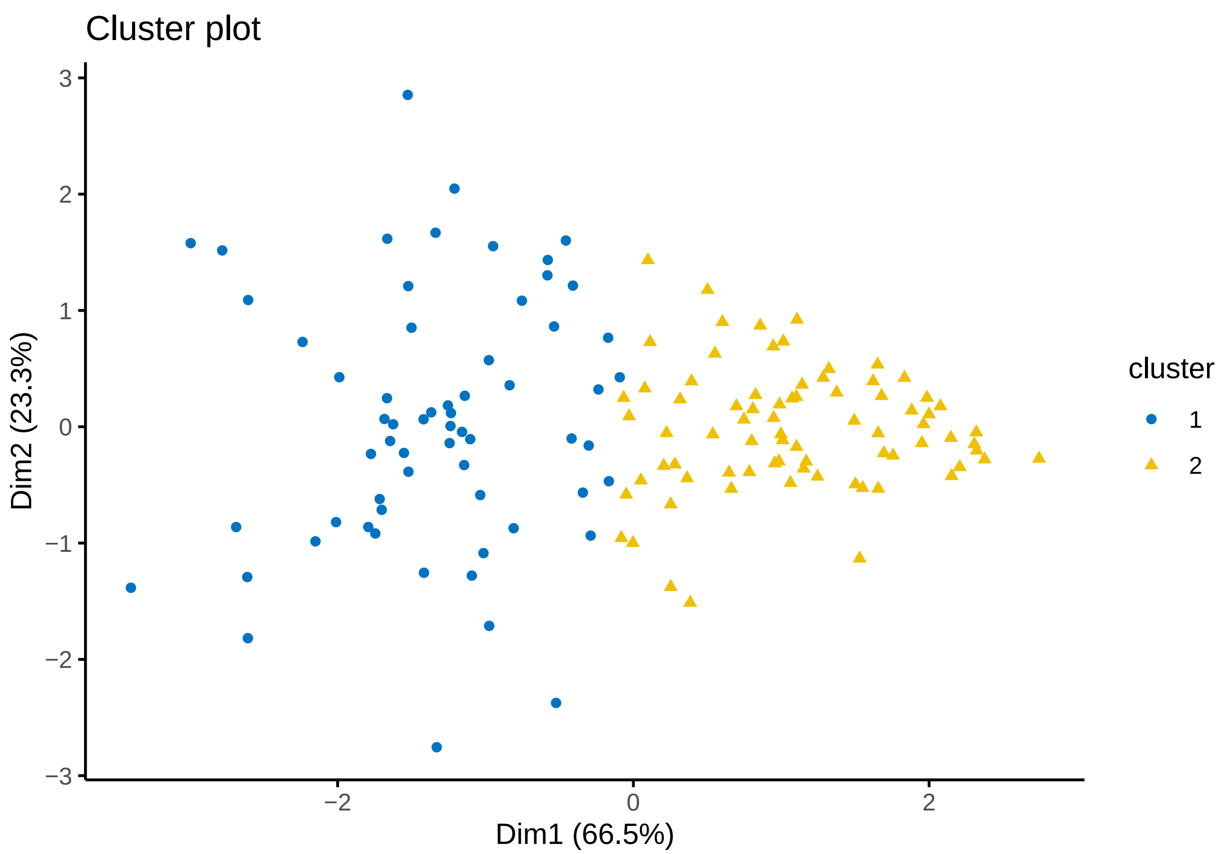
Toronto's Neighbourhoods Clusters Map

Through the visualization of a map, we can see that the neighbourhoods in Group 1 not only share similar socieconomic traits, they are also geographically close to each other. The geographic proximity of the vulnerable neighbourhoods [to be continued…]

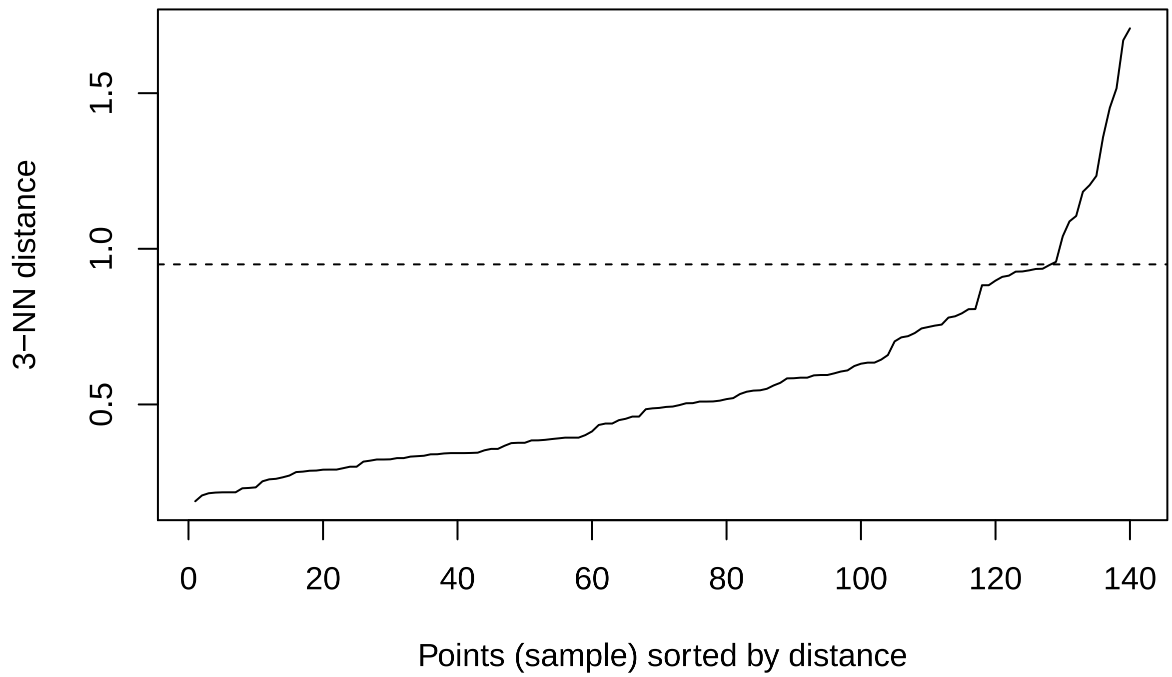
```
## tmap mode set to plotting
```




```r
set.seed(123)
km.res <- kmeans(num_data, 2, nstart = 25)
fviz_cluster(km.res, num_data,  geom = "point",
             ellipse= FALSE, show.clust.cent = FALSE,
             palette = "jco", ggtheme = theme_classic())
```



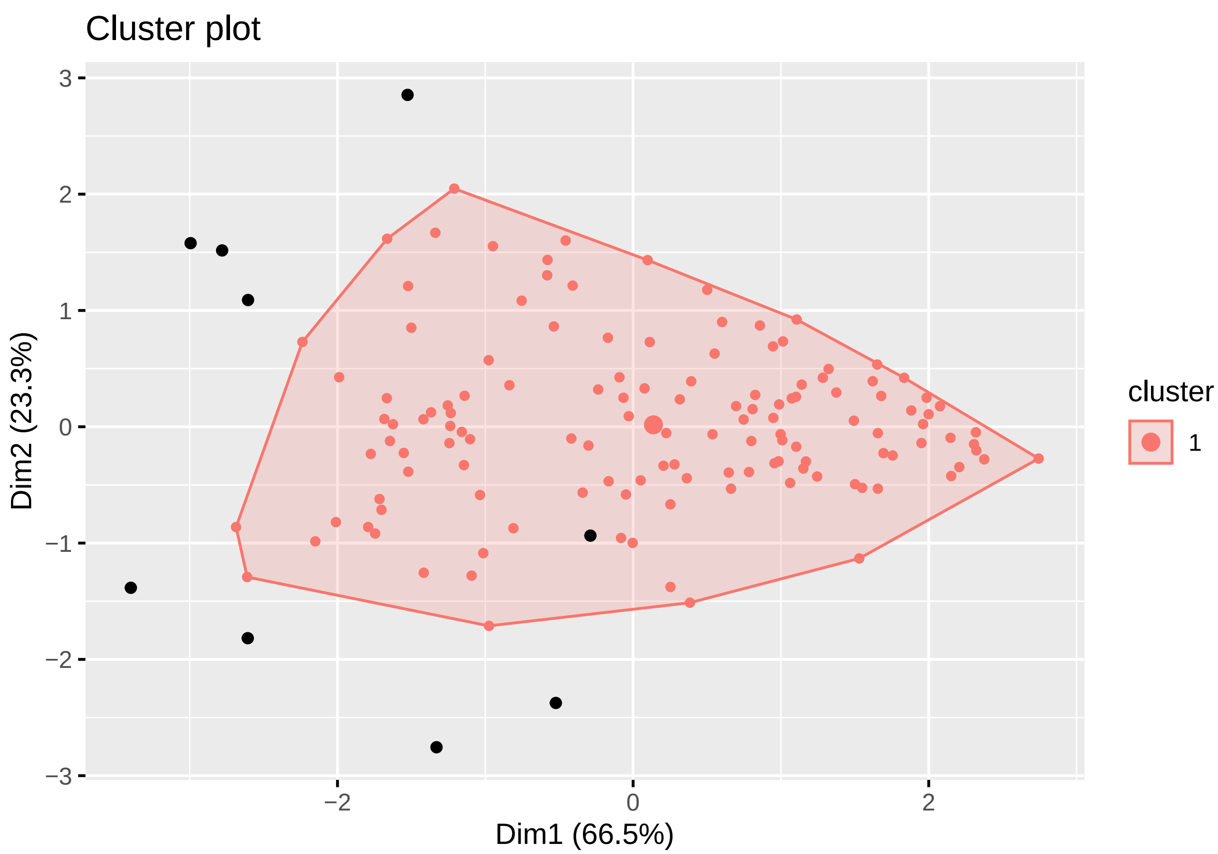

```r
dbscan::kNNdistplot(num_data, k =  3)
abline(h = .95, lty = 2)
```




```r
pred_dbscan <- fpc::dbscan(num_data, .95, 5)
pred_dbscan
```

```
## dbscan Pts=140 MinPts=5 eps=0.95
##         0   1
## border 9  10
## seed   0 121
## total  9 131
```

```r
fviz_cluster(pred_dbscan, num_data, geom = "point")
```




```
?kNN
```

To be included in references: https://www.justice.gc.ca/eng/rp-pr/cj-jp/yj-jj/yj1-jj1/p1_6.html https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3901 https://www.datanovia.com/en/lessons/dbscan-density-based-clustering-essentials/