Universiteit Leiden

ICT in Business

Comparing Municipalities: Who is your partner in crime?

Name: Thijs van der Velden
Student-no: s1645676

Date: 17/08/2017

1st supervisor: Dr. Cor Veenman
2nd supervisor: Prof. Dr. Aske Plaat

MASTER'S THESIS

Leiden Institute of Advanced Computer Science (LIACS)
Leiden University
Niels Bohrweg 1
2333 CA Leiden
The Netherlands

Master's Thesis

Comparing Municipalities: Who is your partner in crime?

Author:
Thijs van der Velden

Supervisor:
Dr. Cor Veenman
2nd Supervisor:
Prof. Dr. Aske Plaat

Abstract

In the Netherlands, each municipality is strongly encouraged to write an Integral Safety Plan (ISP). An ISP is a document that contains problems, goals, and priorities related to safety. This research proposes a method to select an effective and relevant ISP for a municipality out of all available ISPs. It focused on the ISPs of the 390 municipalities in the Netherlands, and aimed to determine ways to assist/support a new written ISP by suggesting relevant ISPs to municipalities in order to learn effective methods from the suggested ISPs. This was done by scoring all currently existing ISP's to get the best performing plans, focusing on crime statistics. Two dimension-reduction techniques were applied to descriptive statistics of the municipalities, and the best performing technique (t-SNE) was used. The municipalities were then clustered with K-means in the t-SNE space to retrieve groups of similar municipalities. In addition, k-nearest neighbor algorithm was used to find the K most similar municipalities from a starting point. This research showed a solution to score ISPs and two solutions to find similar municipalities in order to suggest relevant ISPs. The results showed groups of similar municipalities, where several groups could be identified for their similarity (e.g. student cities). Domain experts have to be consulted for further validation.

August 17, 2017

Contents

1 Preface

For the Master study ICT in Business at Leiden University (Leiden, Netherlands), and in co-operation with Atos (Zoetermeer, Netherlands), a research plan for a Master's Thesis has been created to help municipalities optimize the development process of an ISP (Integral Safety Plan), gain insight in the performance of an ISP, and improve the quality of an ISP. The research has been conducted by Thijs van der Velden, in close cooperation with Patrick de Koning, who is a fellow MSc student doing his thesis on a similar problem.

2 Introduction

2.1 Background

An ISP is a document written by a municipality to describe safety issues, priorities and possible solutions. Municipalities in the Netherlands are strongly encouraged to make an ISP, and there is a law proposed in the Netherlands to make it mandatory[22]. Municipalities have people assigned to this job who investigate which safety issues should be prioritized. Atos has direct contact with municipalities and wants to provide insight on how to improve their ISPs to fit their current situation.

Currently, the quality of the ISPs is checked by the Dutch institution CCV[1] to make sure the documents are meeting their standards. The process used by the CCV focuses on structure and mainly provides guidance in writing an ISP. This does not ensure that the documents correctly reflect the reality of the problems the municipalities face in terms of safety issues. Municipalities could verify the problems and priorities stated in their ISPs by consulting related (crime)statistics.

2.2 Interviews

To get insight into the safety concerns of a certain municipality, the municipality employees have to look at a great variety of data (criminality statistics, population surveys, council-meeting transcripts). This is a time consuming process. Currently there are 390 municipalities in the

[1] https://hetccv.nl/onderwerpen/veiligheidsplannen/

Netherlands, which all have their own ISP. It is difficult for a municipality to look at all the different ISPs to find inspiration for solutions to their problems, if they want to learn from each other.

In order to improve the ISP development process and ISP quality, the current process that is used by municipalities needed to be identified. In addition, several questions needed to be answered in order to eliminate assumptions (see Appendix D). For example, it was assumed that municipalities consulted the crime statistics when writing an ISP, but this had to be verified by speaking with policymakers. This was done through interviews with three municipalities. These interviews were conducted in collaboration with Patrick de Koning.

The interviews made clear that the municipalities have a similar way of developing and evaluating an ISP. The development process can be translated to the model in Figure 1. This model is based on the Policy Making Process model [13].

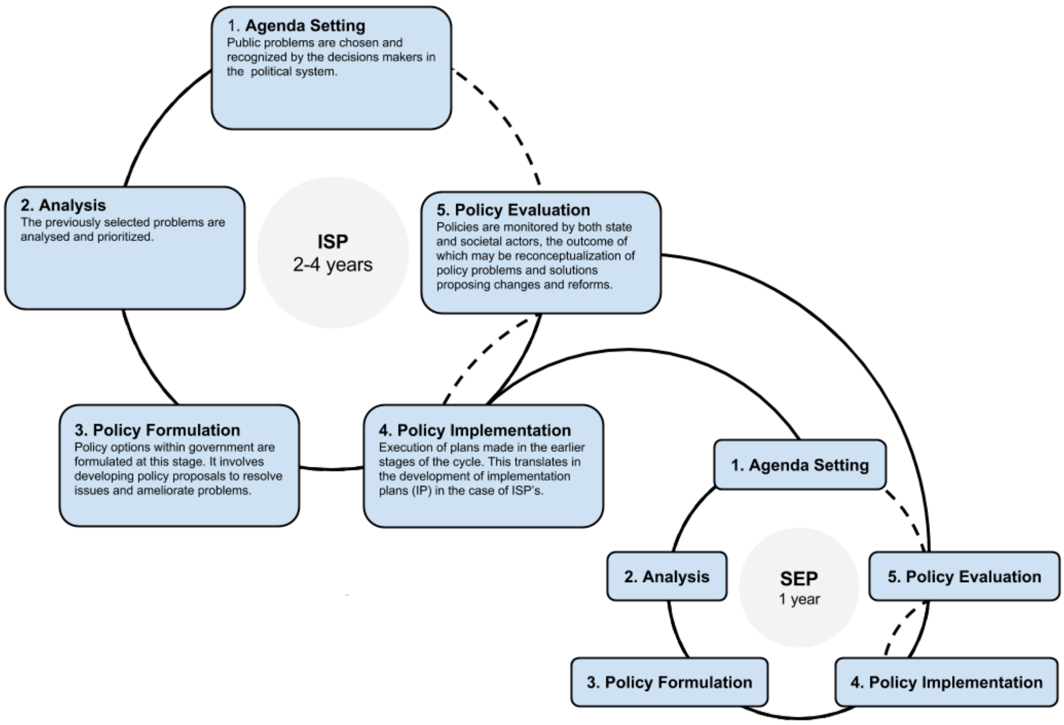

Figure 1: The ISP development process for municipalities.

The model shows two main processes which the municipalities use. The municipalities first develop an ISP, and in the implementation process develop an Safety Execution Plan (SEP). SEPs were used to enforce the actions described in the ISP specifically.

Another result from the interviews was the different sources the municipalities used to evaluate their ISPs. Municipalities used third parties that investigated the livability of their municipality[2], the social planning bureau[3], as well as the "AD Newspaper crime monitor[4]" to determine the development of crime in their municipality, in combination with statistics provided by the police and the Centraal Bureau voor de Statistiek[5] (CBS).

The municipalities compared themselves to other municipalities inside their police region. One municipality mentioned it was comparing itself with a municipality outside of their police region, but the only provided concrete reasoning behind this comparison was population size. When asked about other properties they used when comparing with other municipalities, urbanity was the only other property mentioned.

2.3 Research outline

The purpose of this research was to develop methods that could be used to improve the process of finding solutions for crime related problems in the ISPs of other municipalities, and provide comprehensible statistics. Both the solutions and the statistics should assist them in the writing of an ISP. The main question that this research had to answer was:

How to automatically suggest relevant ISPs of other municipalities?

The result of this research question is a method that allows a municipality to find several relevant ISPs that can help them in writing a new ISP. In order to answer this question, the relevance of an ISP to a municipality has been defined by two measures.

The first measure was the effectiveness of an ISP. The assumption was made that less effective plans are less relevant. The effectiveness of an ISP can be quantified by examining the crime statistics during its active time. This leads to the first subquestion:

How to measure the effectiveness of an ISP?

The second measure is the similarity between municipalities. Identifying which municipalities

[2]http://www.lemoninternet.nl/

[3]https://www.scp.nl/

[4]http://www.ad.nl/binnenland/ad-misdaadmeter-2016-de-ranglijst-en-alle-cijfersãa3fa577/

[5]https://www.cbs.nl/nl-nl/onze-diensten/methoden/classificaties/overig/gemeentelijke-indelingen-per-aar/indeling%20per%20jaar/gemeentelijke-indeling-op-1-januari-2016

are similar could provide insight for a municipality to see which ISPs are relevant to look at. This can be done by looking at descriptive statistics for each municipality, and determining the similarity between different municipalities. This leads to the second subquestion:

How to compare municipalities based on descriptive statistics?

The answers to these questions lead to solutions to identify the effectiveness of an ISP, display meaningful statistics, and show which municipalities are comparable. This could be used to suggest relevant ISPs.

3 Related work

There is similar work done that describes evaluating the performance of ISPs, where it was described as a difficult task[7], where the introduction of safety measures was often too slow to measure an effect during the active time of an ISP. In addition, a large part of what security policies were composed of were the continuation of already in place measures enacted by neighborhoods. This created uncertainty whether measured effects were from the ISP or from other sources. Only when new policies were described in an ISP, the effects that can be linked to those could be measured. However, to measure the effect of an ISP there needs to be a baseline. Police/crime statistics should be consulted to check the development of registered crime. In addition, to measure the number of victims of crime, the fear of crime, nuisance, and livability of the neighborhood, surveys need to be conducted among the population. The value of population surveys to measure the effects of an ISP was mentioned in other literature as well[24]. This approach may be better than only looking at the statistics reported by the police. It is suggested that only consulting crime statistics is not adequate, since it was suggested that there are other causes for the decline or increase in crime that are not related to an ISP [25]. When using the crime statistics as a performance measure for an ISP, a relationship between an ISP and the crime statistics should be present. At the moment there is no scientific evidence that describes this relationship, but it is assumed that this relationship is present since ISPs are used to increase the safety within a municipality by reducing crime.

Similar work done on comparing municipalities was based on their financial status[18], describing that there is a large difference in spending between municipalities suggesting it is an indicator for (dis)similarities between municipalities. Benchmarks were also mentioned when comparing municipalities, where the benchmark is based on descriptive statistics such as economic development, public health, police (crime statistics) and traffic management [3]. It is described how to measure these statistics, and not necessarily how to compare municipalities based on the benchmarks.

Peer city identification and comparison shows close resemblance with comparing municipalities, where a peer city is defined as a city that shows similarities in descriptive statistics. For example, geographic, demographic, and economic characteristics are mentioned[6]. There is a tool developed that automatically identifies peer cities, by doing a hierarchical clustering analysis with 300 cities resembling the data points [11]. Some of the variables used in the clustering that may be applicable to this research are: *poverty rate*, *percent with a bachelor's degree*, *unemployment rate*, *median family income*, *population*, *vacancy rate*.

In order to standardize the way describing statistics are selected, ISO[7] has introduced a way to measure the performance of cities. It can be used by any city, municipality or local government wishing to measure its performance in a comparable and verifiable manner, irrespective of size and location or level of development[23]. They introduced 46 performance measures. These measures are used to measure the performance of a city to allow comparison. However, it is noted that most of the indicators are not currently registered at most cities.

A way to compare smart cities is proposed that consists of evaluating multiple benchmarking approaches. They identified the following themes that the approaches had in common: *people*, *government*, *economy*, *mobility*, *environment* and *living*[4].

[6]http://fyi.uwex.edu/downtown-market-analysis/understanding-the-market/peer-city-comparison/
[7]https://www.iso.org/home.html

4 Data collection/exploration

The goal of the data collection was to retrieve all possible ISPs for each municipality in the Netherlands. This ensured the data can be classified as a population, instead of a sample from the population. The source code that corresponds to this section can be found at Appendix A.

4.1 ISPs

At the start of this research, there were 390 different municipalities in the Netherlands according to the CBS. The CBS provides access to all municipality names from the year 2016.

The CCV has a website containing most ISPs [8]. The municipalities that had an ISP on the website of the CCV were matched with the data of the CBS. This step resulted in 177 ISPs (01-12-2016), out of the potential 390 ISPs. The start- and end-date of each ISP was saved.

Since there was no other central repository where the ISPs could be found, the ISPs for the other municipalities were searched manually. The e-mail addresses of the municipalities whose ISP was missing were found through accessing their website. An e-mail was sent requesting the municipalities to send all of their ISPs (the most recent version as well as older versions). A web search was conducted, which added more IPS's resulting in ISPs for 350 municipalities. The final number of municipalities with at least one ISP was 377 out of the 390 municipalities. Municipalities often had more than one ISP, resulting in a total number of 430 ISPs.

To identify if there was a pattern present in the missing ISPs, a geographical visualization was made of the Netherlands showing which municipalities had an ISP and which did not (see Figure 2). From the visualization can be noted that the municipalities which do not have an ISP are relatively scattered. However, three Waddeneilanden (Texel, Vlieland, Terschelling) do not have an ISP. This was probably due to the fact that these municipalities are very small, and use a regional ISP.

4.2 Crime statistics

In this research the assumption was made that ISPs have influence on the crime statistics since they are used to increase the safety within a municipality by reducing crime [8]. The CBS

[8] https://hetccv.nl/onderwerpen/veiligheidsplannen/

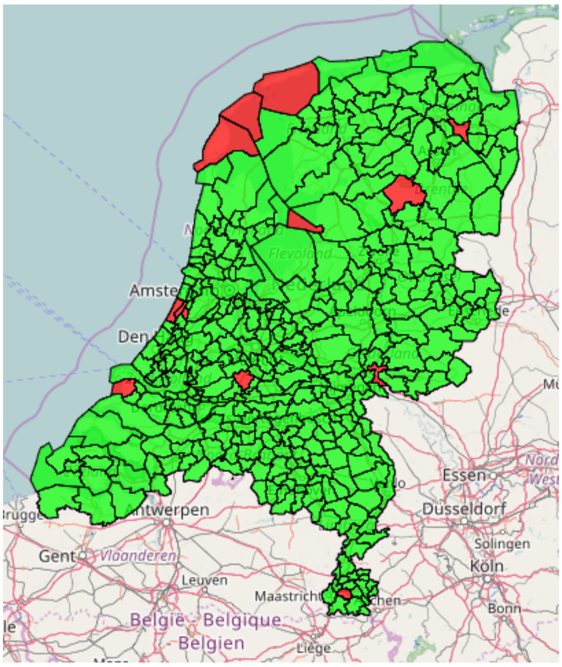

Figure 2: The different municipalities in the Netherlands which have an ISP (green), and which do not (red).

provides crime statistics for each municipality which can be found at CBS statline [9]. A selection was made of all available crimes and years. This research used all crimes for the municipalities in the Netherlands in 2016 from 2009 to 2016. Some manual string replacements were done on the CBS data (e.g. *Beek (L.)* has been renamed to *Beek*).

The crime files contained several crime indicators for each municipality (see Appendix B). The crime indicators consisted of 9 main indicators, with each main indicator containing several sub indicators. *Misdrijven, totaal* is the sum of all indicators. The names of the crime indicators have been left in Dutch throughout this research. A translation of the main indicators can be found in Appendix C. This research has focused on the 9 main indicators:

- Misdrijven, totaal
- 1 Vermogensmisdrijven
- 2 Vernielingen, misdr.openb.orde/gezag
- 3 Gewelds- en seksuele misdrijven
- 4 Misdrijven WvSr (overig)

[9]http://statline.cbs.nl/Statweb/dome/?TH=50480&LA=nl

- 5 Verkeersmisdrijven
- 6 Drugsmisdrijven
- 7 (Vuur)wapenmisdrijven
- 9 Misdrijven overige wetten

To see how the crime statistics were distributed, a visualization of the trends of the first 16 ISPs can be seen in Figure 3.

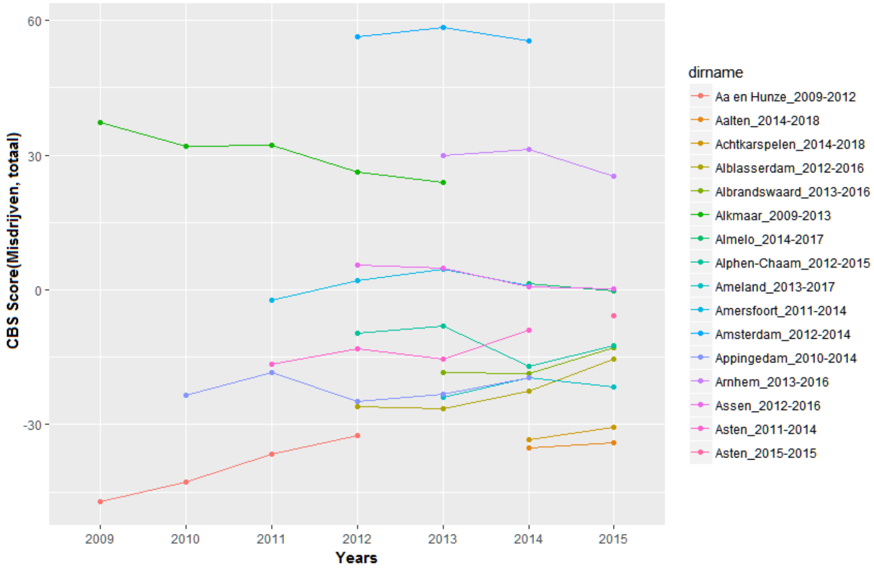

Figure 3: The trends of 16 ISPs for the indicator *Misdrijven, totaal*. The CBS crime statistics are transformed by crimes per 1000 citizens compared to the mean of the Netherlands.

Throughout this research all figures contain data for the municipality Leidschendam-Voorburg because this municipality had 3 ISPs and had fluctuating data.

5 ISP effectiveness

This section describes how the effectiveness of an ISP was measured. The crime statistics were analyzed in order to get an answer to the question: *How to measure the effectiveness of an ISP?*

An ISP is effective when the goal(s) that are stated within an ISP were realized after the ISP has completed its active time. In most ISPs the active time is four years. Since the goals of the ISP are within the text, in this research the assumption has been made that every ISP had the goal of reducing crime. Therefore, a reduction in crime statistics during the active time of an ISP for the corresponding municipality meant the ISP has successfully completed its goal of reducing crime. The effectiveness of an ISP was measured for each indicator. The source code that corresponds to this section can be found at Appendix A.

5.1 Data preparation

The data was standardized to number of crimes per 1000 citizens. In order to see if an ISP was effective, it can be compared with the mean of the Netherlands (see Figure 4). The reason it was important to make this comparison was because there might be a national trend present in the development of crime in the Netherlands which may be due to external factors like (inter)national legislation, which means that when the number of crimes is getting lower every year throughout the Netherlands the majority of the municipalities should also see a reduction in crime numbers over the years. To cancel out the national trend growth, the number of crimes of every municipality was subtracted with the mean of the Netherlands. Meaning that when a municipality follows the exact same trend as the mean of the Netherlands, the effectiveness will be 0.

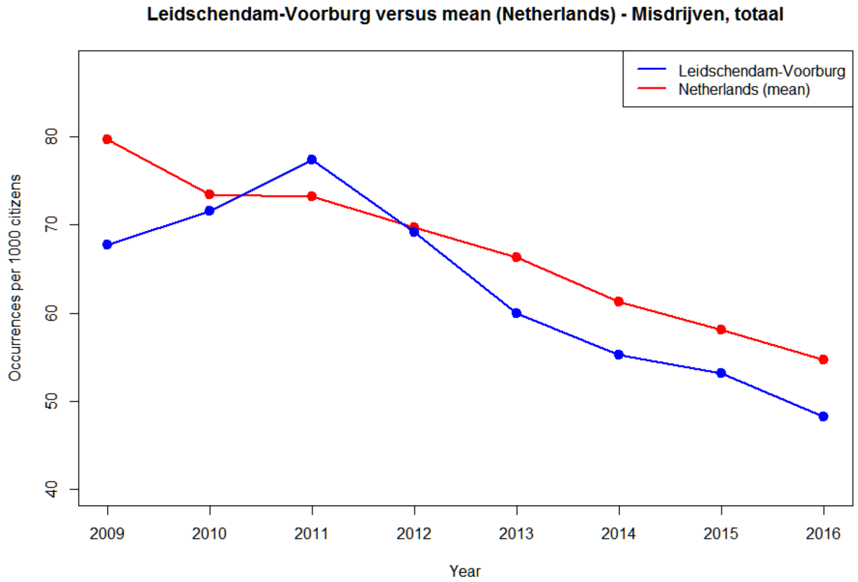

Figure 4: The crime numbers for the indicator *Misdrijven, totaal* for Leidschendam-Voorburg (blue), and the mean of the Netherlands (red).

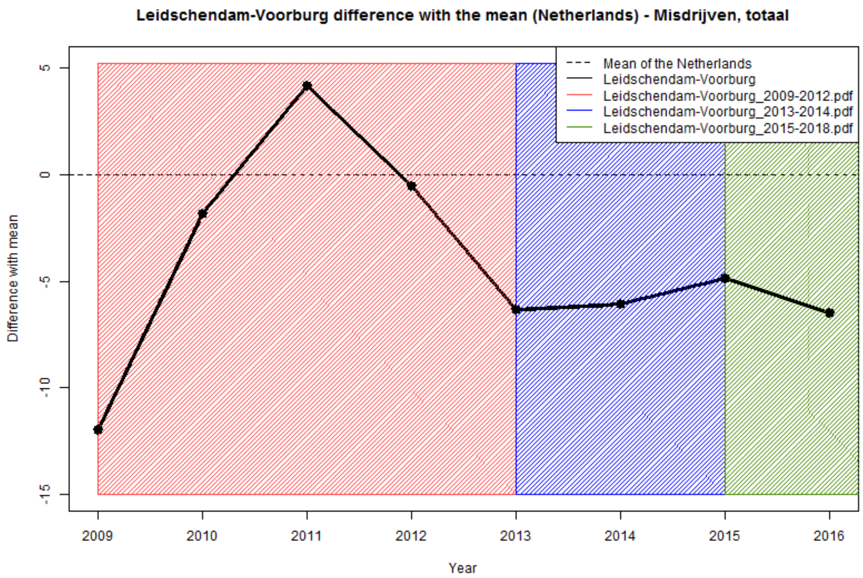

Figure 5: The number of crimes for Leidschendam-Voorburg for indicator *Misdrijven, totaal*, in comparison to the mean of the Netherlands. Different background colors represent the active time of different ISPs.

5.2 Determining the effectiveness of an ISP

5.2.1 Method

After standardization of the data the performance of the different ISPs of a municipality can be calculated. An example of a performance visualization for Leidschendam-Voorburg is shown in Figure 5. This shows the progress of the crime numbers for the indicator *Misdrijven, totaal* over the years 2009-2016, with the value of the crime statistic for the active years of the different plans visualized.

In order to rate the actual performance of an ISP and compare the scores between different ISPs, the value of a crime statistic at the start-year of an ISP (startValue) and the end-year of an ISP (endValue) were taken. The ISP endValue could then be subtracted from the ISP startValue to get the growth or decline in crime for the duration of an ISP (crimeScore). See equation (1). If the endValue was higher than the startValue, the crimeScore was negative, indicating a growth in crime. Therefore the effectiveness of an ISP could only be measured when there was atleast two years of data available. For example, an ISP that had an active time of 2016-2019 only has one year of data available (2016). This eliminated some ISPs from the dataset and resulted in a total of 361 ISPs that could be rated on performance. In order to account for the abnormal distribution of scores between the different indicators, a normalization factor has been introduced that is relative to the maximum and minimum crimeScores of all the ISPs for an indicator to normalize the score. See equation (2). In order to make the score interpretable for municipalities, the score has a maximum of 10 and a minimum of 1. A baseline of 6 is used to ensure that when an ISP has followed the national trend the ISP score is 6. A 6 in the European grading system means average. The normalization factor is multiplied with the crimeScore and added to the to the baseline to calculate the ISPScore. See equation (3).

$$crimeScore = startvalue - endvalue \tag{1}$$

$$normalizationfactor = \frac{10}{max(crimeScore) - min(crimeScore)} \tag{2}$$

$$ISPScore = normalizationfactor \times crimeScore + 6 \tag{3}$$

5.2.2 Results & analysis

The scores for each of the 9 indicators for the ISP *Leidschendam-Voorburg_2009-2012* are shown in Figure 6. If a bar is red, this means the score was below 6 indicating it was below average. Green means average or above. *Misdrijven, totaal* and *1 Vermogensmisdrijven* have similar scores since *1 Vermogensmisdrijven* accounts for the largest share of crimes in *Misdrijven, totaal*. It should be noted that indicators 4, 6, 7 and 9 have incomplete and unreliable data, which may result in scores that do not reflect the reality.

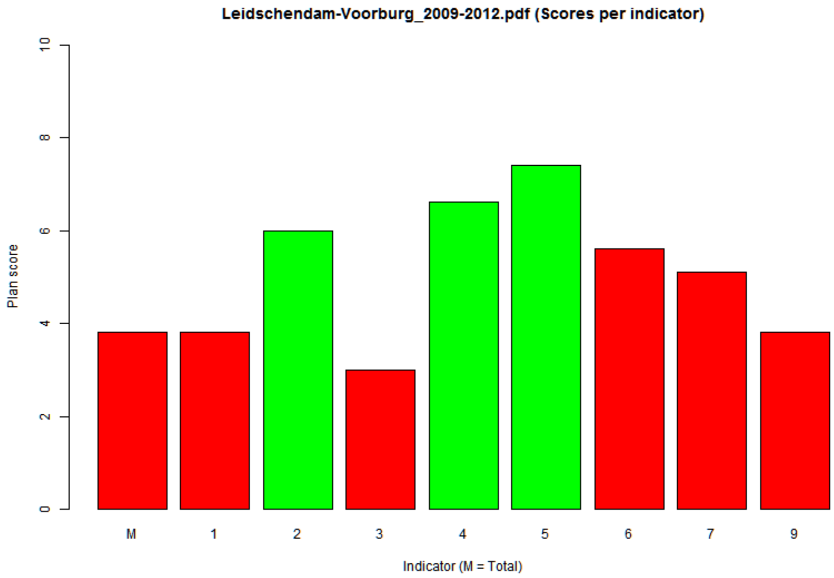

Figure 6: The ISP scores of the ISP *Leidschendam-Voorburg_2009-2012.pdf* for all indicators. (M = *Misdrijven, totaal*, numbers represent the corresponding main indicator.)

The increases and decreases over time in Figure 5 show the progress of an ISP, which is valuable to a municipality to see how the ISP(s) has performed over time. For example, in the case of *Leidschendam-Voorburg_2009-2012* at first there is an increase which is followed by a decrease. This allows a municipality to understand that a new ISP can have a temporary increase before the measures described in the plan are active and effective. It should be noted that there may be other factors that can also influence the crime statistics, such as criminal gangs traveling through the country.

Figures 5 and 6 can be used in combination with each other to show the effects of an ISP overtime. This allows a municipality to see which plans were most effective and helps to identify and improving the next ISP.

6 Comparing municipalities based on descriptive statistics

In order to suggest a relevant ISP to a municipality, the suggested ISP should come from a municipality that shows similarities. For example, a large municipality may have different budgets compared to a smaller municipality, which means that the ISP may not be relevant due to the policies mentioned requiring a higher budget. Descriptive statistics like budgets can be used to find similar municipalities. In this research is assumed that an ISP is relevant to another municipality when the municipality who wrote the ISP shows the same descriptive statistics as the target municipality. This assumption came from the understanding that there is a large difference in budget, types of crime, geography, and number of citizens between municipalities. For example, there was more drug-related crime in the municipalities near the border of the Netherlands [5] suggesting these municipalities described measures to counter drug-related crime. These measures might not be relevant for a municipality in the north of the Netherlands where the number of drug-related crimes was very low.

In this section is described how it is possible to determine descriptive statistics objectively, and how these measures were used to find similar municipalities. This resulted in an answer to the question: *How to compare municipalities based on descriptive statistics?*. The source code that corresponds to this section can be found in Appendix A.

6.1 Selecting relevant descriptive statistics

6.1.1 Method

Descriptive statistics are statistics which describe an individual municipality. The descriptive statistics were selected through CBS statline [10], by looking for statistics which had a specific format in order to be eligible for comparison. The format requirements for the statistics were that they had to be present for every one of the 390 municipalities, and had to have data between 2009-2016. Every statistic was standardized to per 1000 citizens. Problemyouth was already standardized by the CBS.

To objectively determine which descriptive statistics are relevant when comparing municipalities, a solution is proposed that weight the descriptive statistics for their relevance to the crime

[10] http://statline.cbs.nl/Statweb/

indicators, by changing the scale of the variables[16]. The reason for this is that descriptive statistics are conditional on a certain crime. For example, a high number of students may have a lot of influence on indicator *5 Verkeersmisdrijven*, but not on indicator *1 Vermogensmisdrijven*.

To identify which descriptive statistics were relevant for municipalities in relation to crime statistics, the correlation was calculated between the descriptive statistics and the crime statistics. These correlations could then be used to weight the descriptive statistics for all municipalities. Through this method, every descriptive statistic could be added and be evaluated for its relevance to crime statistics. The weighting was done by multiplying the correlation with the measures. This resulted in an objective weighting of each measure corresponding to the relevance to each indicator.

6.1.2 Results & analysis

This section describes the results of selecting the similarity measures (descriptive statistics), clustering, and the analysis of the clustering methods.

For this research, a total of 10 descriptive statistics were selected:

- Jobs (per 1000 citizens, 2015)
- Population (2015)
- Problemyouth (2015)
- Surface (per 1000 citizens, ha, 2015)
- Welfare (per 1000 citizens, 2015)
- Buildings (per 1000 citizens, 2015)
- Water (per 1000 citizens, 2015)
- Cars (per 1000 citizens, 2015)
- WO Students (per 1000 citizens, 2015)
- MBO Students (per 1000 citizens 2015)

Some of these statistics were selected because it was suggested they already had some relationship with crime, such as the amount of welfare [10], which can also be seen in table 1. Others were selected because municipalities think they are relevant when comparing themselves with others. Problemyouth was selected because it is often described in ISPs [26]. For each descriptive statistic, the correlation with each indicator was calculated (see Table 1). The correlation was converted to an absolute value. For indicator *Misdrijven, totaal*, Welfare, Population, and Surface area had the highest correlation.

An interesting observation from the table is that the amount of jobs per 1000 citizens has a very low correlation with the crime statistics. This is interesting because unemployment has a relationship with criminality[6][1].

Table 1: Correlation between similarity measures and crime statistics for the first 3 indicators.

Measure	Misdrijven, totaal	1 Vermogensmisdrijven	2 Vernielingen, misdr.openb.orde/gezag
Jobs (per 1000 citizens, 2015)	0.04194488	0.06371811	0.03373170
Population (2015)	0.62081905	0.63680145	0.34287806
Problemyouth (2015)	0.14605824	0.08815838	0.22873920
Surface (per 1000 citizens, ha, 2015)	0.47526567	0.45778133	0.40869639
Welfare (per 1000 citizens, 2015)	0.62890714	0.57845395	0.57946164
Buildings (per 1000 citizens, 2015)	0.01023983	0.04240081	0.06894129
Water (per 1000 citizens, 2015	0.04030544	0.05477961	0.05082376
Cars (per 1000 citizens, 2015	0.26322018	0.25717574	0.27214669
WO Students (per 1000 citizens, 2015	0.31725858	0.36678656	0.10090320
MBO Students (per 1000 citizens 2015)	0.26119265	0.30241160	0.06725458

6.2 Data evaluation

To identify whether the data was usable and if there were patterns present, two dimension reduction techniques were used. There are several dimension-reduction techniques[9], where the most notable difference between those is whether they use a linear function or a non-linear function [12]. To find which dimension-reduction technique produces the best result, a non-linear as well as a linear technique should be used and compared [17]. Out of the linear techniques, principal component analysis (PCA) is most often used. PCA is a multivariate technique that analyzes a data table in which observations are described by several inter-correlated quantitative dependent variables [2]. Out of the non-linear ones, t-Distributed Stochastic Neighbor Embedding (t-SNE) is the best performing technique [20]. Both techniques result in a visualization of the data in a two-dimensional setting. This opens the possibility of manual identification of groups of similar municipalities, after which further analysis can be done.

After visualizing the data, and if the data proves to be useful, the similar municipalities should automatically be extracted. One way to do this is to determine the groups (clusters) that are present in the data[14], through clustering. Out of the different cluster algorithms available, K-means was suggested as the main approach.

Another way to find the most similar municipalities, is to use a starting point and selecting a number of most similar (or closest) points in the data. For this, an unsupervised nearest neighbor algorithm can be used[19].

Figure 7 shows the visualization of the PCA analysis on the data. The figure shows no clear patterns or groups in the data, with the center containing a large number of datapoints close together. At the edges small groups can be identified that are distinguishable from the other datapoints.

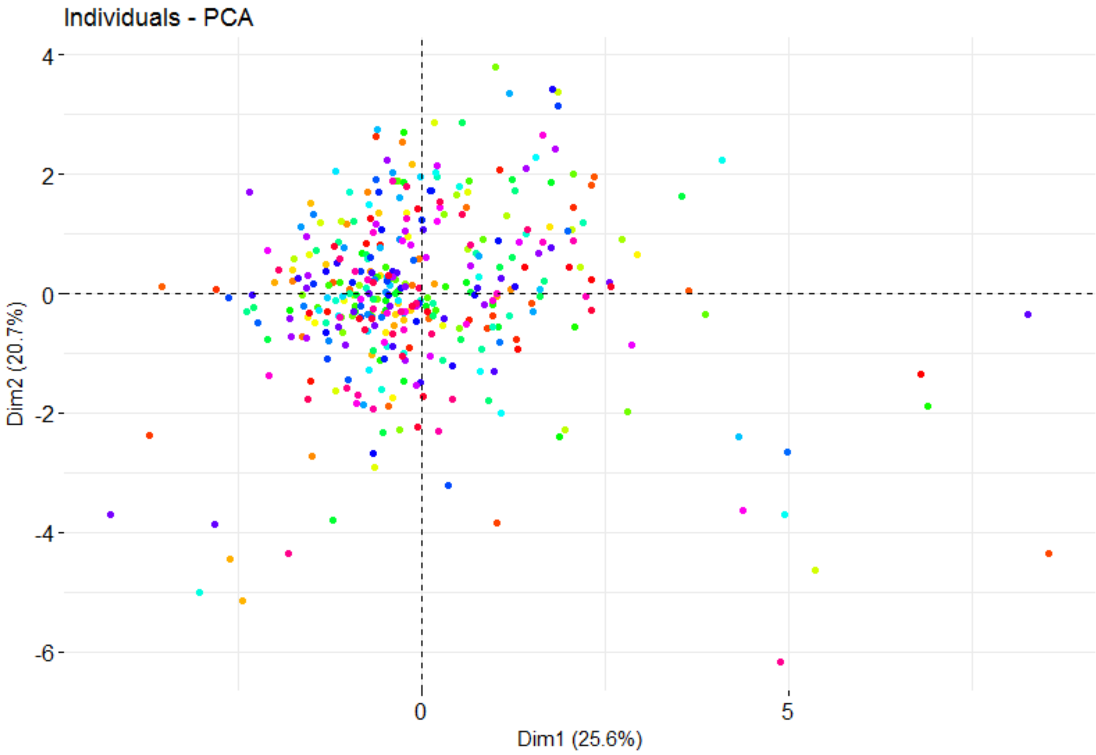

Figure 7: Visualization of the principal components that explain the most variation in the data for indicator textitMisdrijven, totaal.

Figure 8 shows the visualization of t-SNE on the data (dims = 2, perplexity = 5). The dimensions have been set to two, since this was the desired amount. The perplexity has been set at 5 since this resulted in the best visualization, as mentioned by the author of the algorithm[11]. In addition, it was also suggested to use a lower perplexity when dealing with a small dataset. In the figure distinct groups can be identified. When looking at the labels it becomes clear that

[11]https://lvdmaaten.github.io/tsne/

these municipalities show similarities when comparing the data manually. For example, at the top of the visualization a group can be identified that only contains municipalities with a high number of students.

The t-SNE visualization allows for further analysis to identify groups automatically or find the most similar municipalities from a starting point.

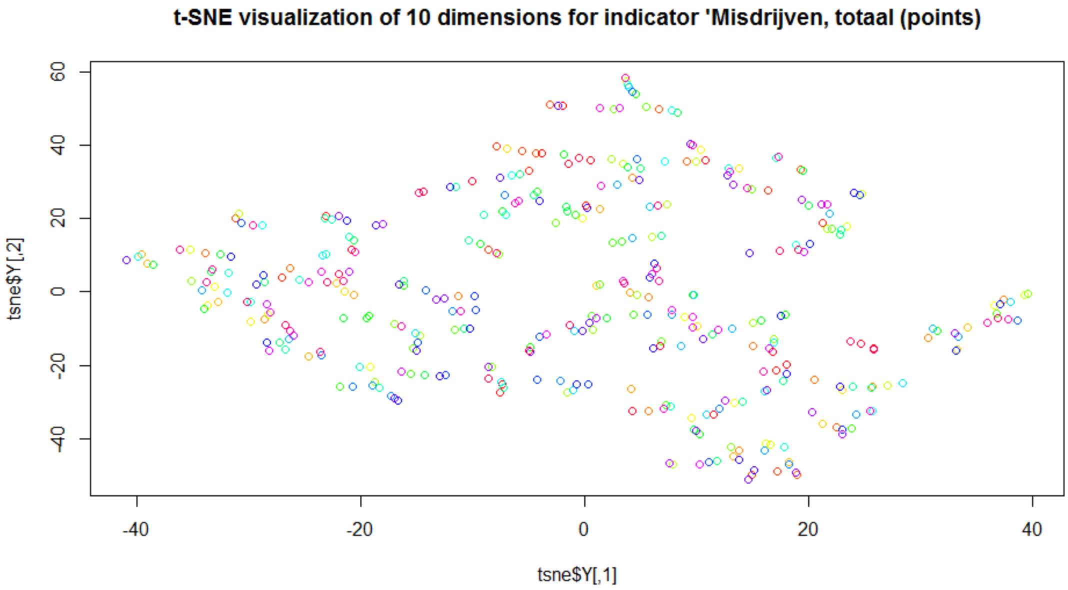

Figure 8: Visualization of in two dimensions for indicator *Misdrijven, totaal*.

6.3 Finding similar municipalities

6.3.1 Method

In order to automatically determine groups in the two dimensional t-SNE space, clustering was used (K-means [21], default parameters, K = 25). The municipalities were compared within the Netherlands, and within their police region[12], as the municipalities use the police region space for comparison. The clustering was performed on scaled data[13], which was then weighted by the method suggested in the previous section.

[12] https://www.regioatlas.nl/indelingen/indelingen_indeling/t/politie_eenheden
[13] https://stat.ethz.ch/R-manual/R-devel/library/base/html/scale.html

The amount of clusters used in the clustering was set to 25, based on an estimation of the number of groups in the t-SNE visualization (Figure 8). The same estimation was made for the different police regions, which have different amounts of municipalities. To visualize the clustering results, a function within R called *fviz_cluster* from the package 'FactoExtra' was used.

An additional way to identify similar municipalities is by using the k-Nearest Neighbors algorithm [19] (KNN). This algorithm allows the user to select his own municipality, and the number of neighbors that is desired. It then searches for the K nearest neighbors. When applying this algorithm to the two-dimensional t-SNE space, the user can specify a municipality, and the desired number of similar municipalities. The user is then presented with a list of K municipalities.

6.3.2 Results & analysis

This section describes the results of using K-means clustering with 25 clusters, and KNN in the two-dimensional t-SNE space. The results were produced for each indicator, but in this section only the results for indicator *Misdrijven, totaal* are shown. They can also be produced for each different police region, but only the results for police region *Den Haag* are shown (containing *Leidschendam-Voorburg*).

Clustering

Figure 9 shows the different clusters created. The patterns that were visible in the t-SNE visualization are now clearly defined in clusters, without overlap. The cluster labels are shown in Appendix E. When looking at the labels, clusters can be identified containing municipalities which would be grouped when selecting the clusters manually. For example, cluster 23 contains municipalities that have a high number of students, and are categorized in the Netherlands as student cities [14]. There are no clusters which contain only one or two municipalities, which is desired when presenting a municipality with ISPs of similar municipalities.

The effect of the weighting method can be seen by comparing figure 9 with 10. The cluster labels are shown in Appendix F. The figures show that the weighting results in two different distributions of the points, and different cluster contents. For example, the weighting for indicator

[14] https://www.studentenwegwijzer.nl/studentensteden/

Misdrijven, totaal results in the municipality *Enschede* residing in the cluster with the student cities. The weighting for indicator *5 Verkeersmisdrijven* results in *Enschede* not residing in the cluster with the student cities. This suggests that the ISPs of the student cities are relevant for *Enschede* when looking for solutions for *Misdrijven, totaal*, but not when looking for solutions for *5 Verkeersmidrijven*.

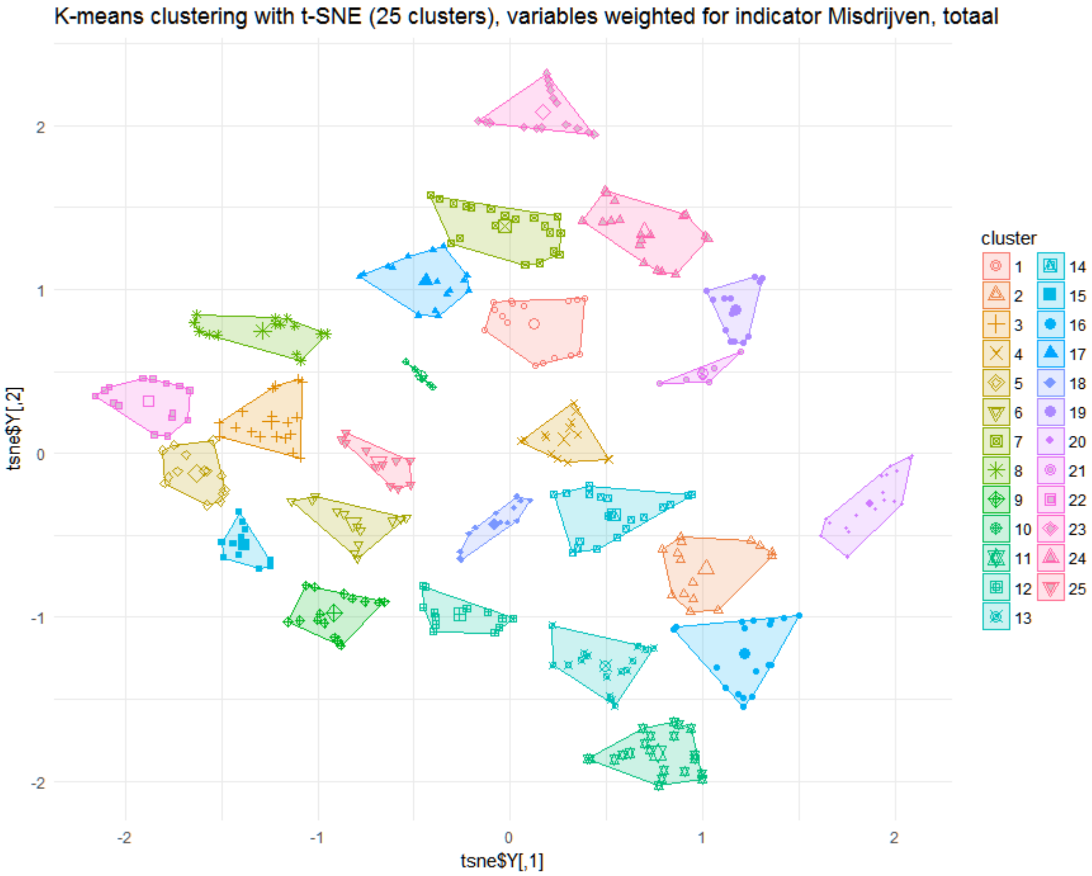

Figure 9: K-means clustering with 25 clusters in two-dimensional t-SNE space, variables weighted for *Misdrijven, totaal*.

Figure 11 shows the clustering result of K-means clustering with 4 clusters in the two-dimensional t-SNE space with the municipalities that belong to police region *Den Haag*. The figure shows distinct groups, but further analysis of the groups by a domain expert is necessary.

KNN

An example result can be seen in figure 12. In this example the starting point is Leidschendam-

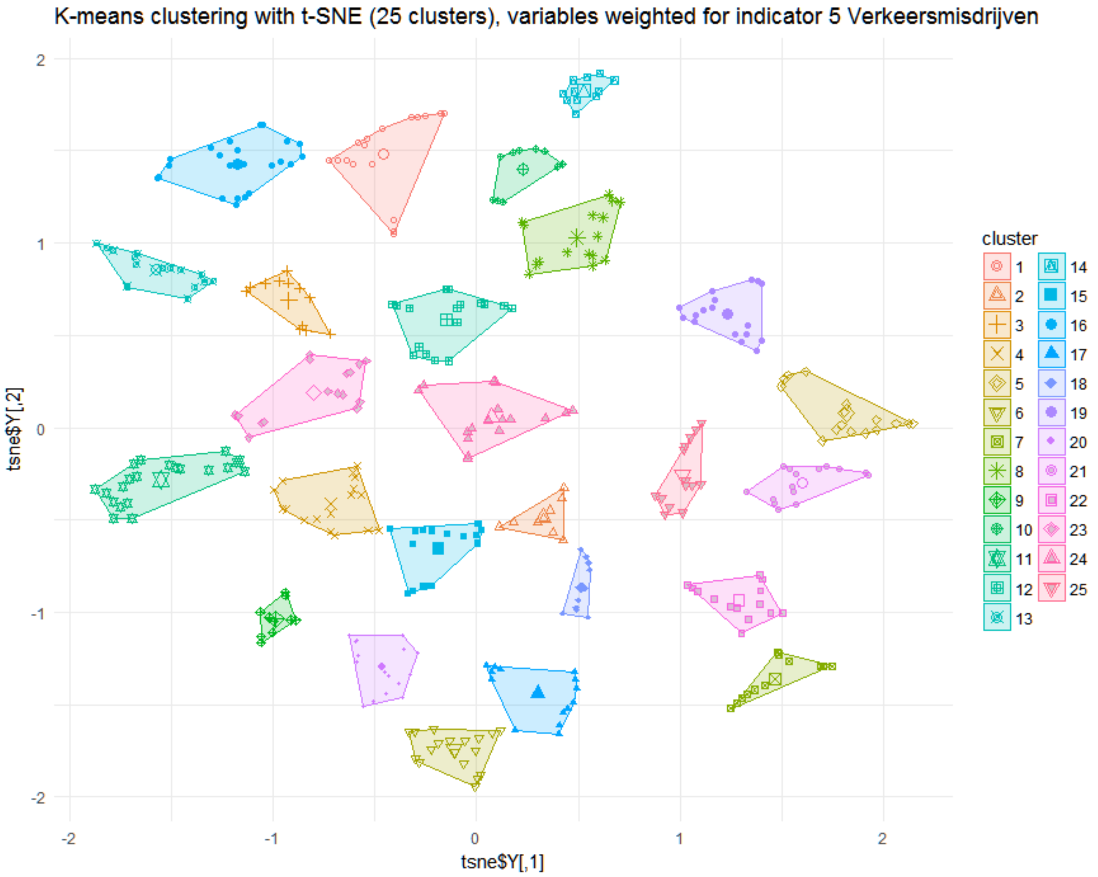

Figure 10: K-means clustering with 25 clusters in two-dimensional t-SNE space, variables weighted for *5 Verkeersmisdrijven*.

Voorburg, and the 10 most similar municipalities are requested. Figure 13 shows another example result with starting point Leidschendam-Voorburg, and the 5 most similar municipalities are requested in its police region.

Comparison

The advantage of K-means clustering over KNN is that ideally no dissimilar municipalities will be suggested. Since with KNN the number of similar municipalities has to be defined, it is possible that when the number is too high, similar as well as dissimilar municipalities will be suggested. KNN has the advantage over K-means clustering in that every municipality has the possibility of requesting as many similar municipalities as desired.

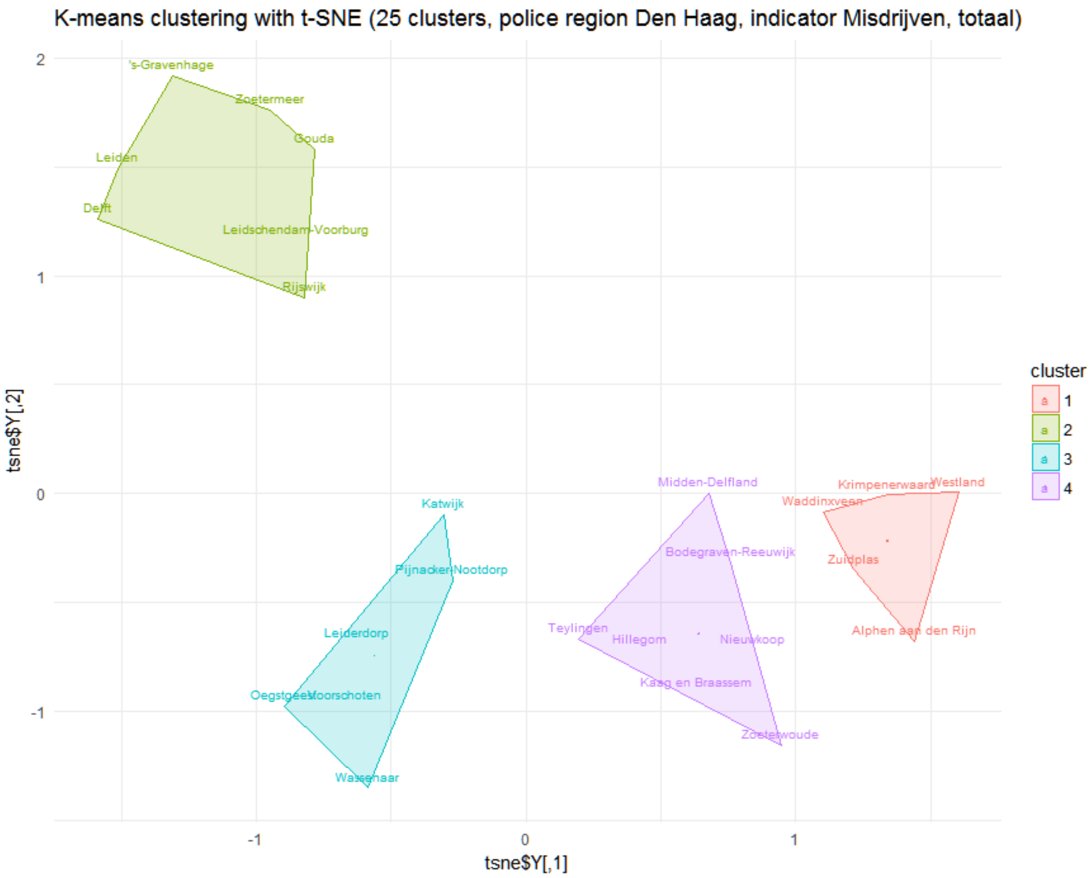

Figure 11: K-means clustering with 4 clusters in two-dimensional t-SNE space for police region *Den Haag*.

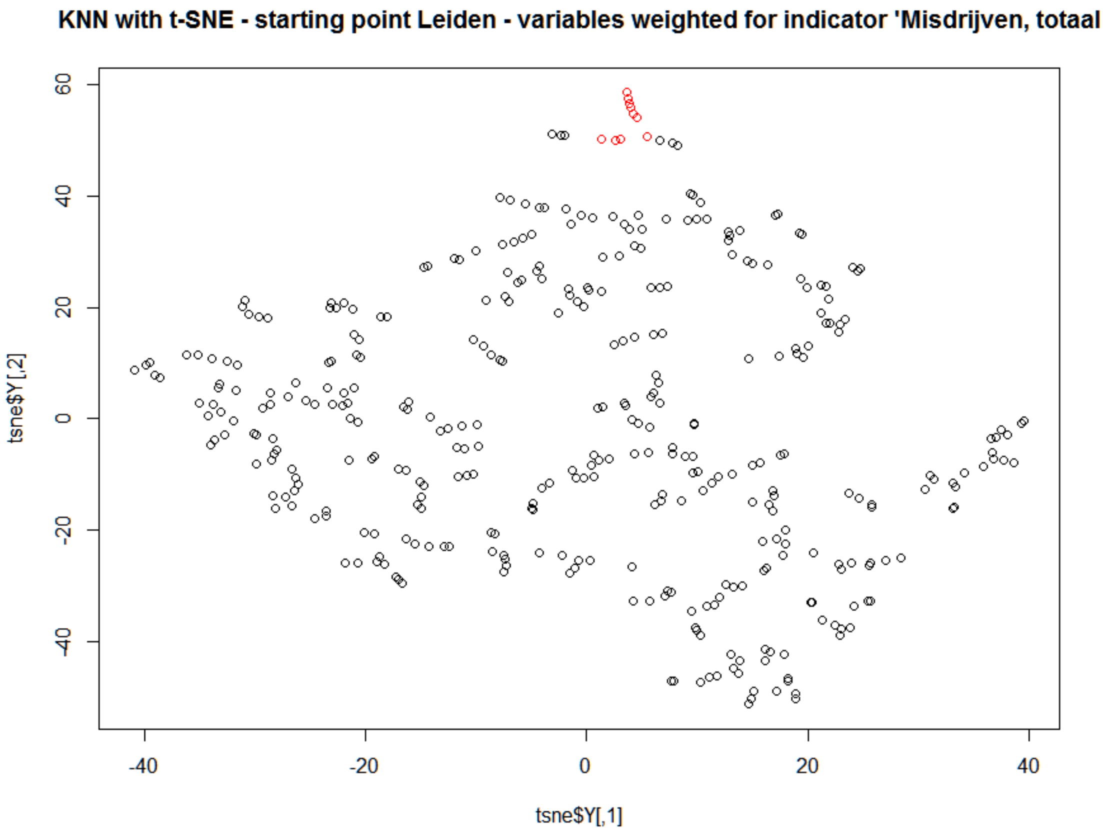

Figure 12: 10 Nearest Neighbors (red) for Leiden in the two-dimensional t-SNE space, variables weighted for *Misdrijven, totaal*.

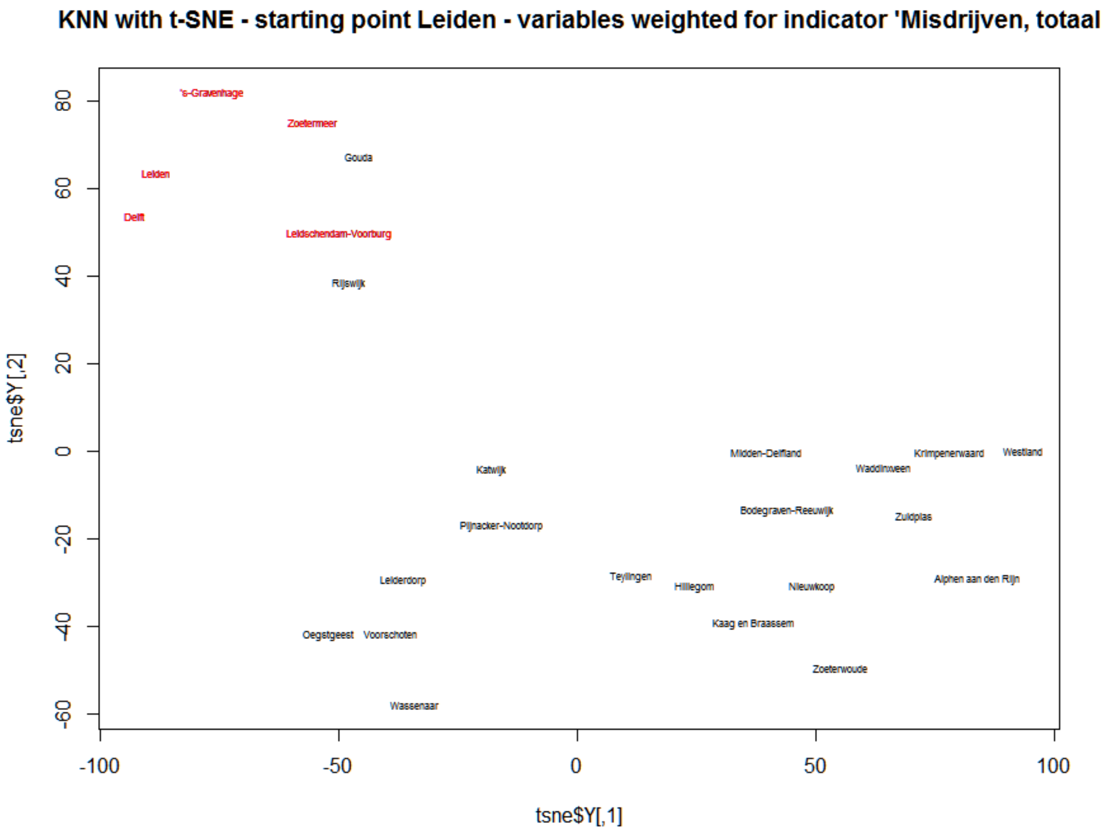

Figure 13: 5 Nearest Neighbors (red) for Leiden in the two-dimensional t-SNE space for police region *Den Haag*, variables weighted for *Misdrijven, totaal*.

7 Discussion

The effectiveness of the ISPs has been measured by taking the value of a crime statistic at the year before an ISP starts, and at the year an ISP ends. Then the values were scaled to let the least effective ISP correspond to 1 and the most effective ISP correspond to 10, in order to make the values interpretable. However, it could be the case that an ISP starts having effect one or two years after the start, and the effect finishes two years after the ISP has ended. A conclusion that clarifies the time an ISP really has an effect on crime development needs to be established in further research.

The suggestion of relevant ISPs does not account for the year an ISP was written in. However, this could pose problems when an ISP has been effective in previous years and the solutions proposed in those ISPs were outdated. For example, an ISP that was successful from 2009-2012, decreased vandalism by removing public mailboxes. But in 2017 there are already a lot less mailboxes because the use of physical mail keeps decreasing.

The result of this research is a way to suggest relevant ISPs, where the relevance is determined by the effectiveness and the similarity between municipalities. It could be the case that municipalities think an ISP is relevant when it meets other criteria, and ultimately the relevance of an ISP remains subjective. However, this research aims to eliminate the need of thinking about relevant ISP's. Further research should establish that the methods that were used in this research provide relevant ISPs, and identify other relevant descriptive statistics.

The results of both the relevance of documents and the groups of similar municipalities can not be validated objectively. A subjective validation method for the groups is to let a large amount of municipalities look at their groups, but they may disagree and/or not understand the contents of their group. This may be the case because they compare themselves with municipalities this research contradicts. A subjective validation method for the relevance of the suggested documents would be to suggest documents to municipalities, and let them decide whether the documents were relevant or not. Both methods are not ideal and are very time consuming and resource intensive.

One could argue that instead of suggesting ISPs and scoring those, using SEP's can be better since these contain more concrete actions. This means that it is easier to find solutions for crime related problems in SEPs. Nevertheless, ISPs often have a four year active time where SEP's have a one to two year active time. One or two years may not be enough to measure its performance accurately, and determining their relevance. An addition to suggesting relevant ISP's is suggesting the corresponding SEP's with the ISP's.

8 Conclusion

The first subquestion: *How to measure the effectiveness of an ISP?*, has been answered by calculating a score for each ISP. This resulted in a measure which could be used to sort the best performing ISPs. The score per indicator was scaled between 1 and 10, allowing it to be be interpretable by policymakers in municipalities.

The second subquestion: *How to compare municipalities based on descriptive statistics?*, was answered by using two algorithms that were able to produce similar municipalities, where the similarity measures used were descriptive statistics. This research also proposed a way to weight descriptive statistics based on their correlation to an indicator. This opens the possibility to suggest different ISP's for different indicators, where a municipality may be similar to another for indicator *Misdrijven, totaal*, but not for indicator *Verkeersmisdrijven*.

The main research question: *How to automatically find relevant ISP's using municipality statistics?*, has been been answered by combining the answers to the subquestions. The result is a solution that scores ISPs, which makes sure only effective ISPs are suggested, and a method that produces similar municipalities to ensure that only ISPs are suggested that are relevant.

9 Future work & other applications

9.1 Future Work

This research can be continued by adding more descriptive statistics, with keeping the curse of dimensionality[15] in mind (adding too many dimensions reduces the clustering quality). This can be done by using the weighting technique proposed in this research by eliminating descriptive statistics that have a low correlation with the crime statistics.

Another continuation would be using action-plans instead of or in combination with the ISPs. This allows for more comprehensive solution to crime related problems, with action-plans possibly containing more specific solutions. The combination of ISPs and action-plans might be the best approach, further research should confirm or refute this.

9.2 Other applications

The methods described in this research are applicable on every sort of plans, when the effect of the plans can be measured accurately. The relationship between the plans and the entity to which the plans belong must be clearly definable, where the entity has descriptive statistics in order to be compared to other entities.

An example of another application is for environmental plans. This can be done on a national scale where the entities are represented by countries, allowing for the use of sufficient descriptive statistics. The effect of environmental plans could be measured by using statistics such as the total reduce of CO2 emission, percentage of green in a country gained, and biodiversity gained during an environmental plan.

Another example is for time-bound business plans. The entities are represented by businesses, and effect of the plans can be measured by statistics such as revenue gained, and share price increase.

Other applications are:

- Healthcare plans
- Zoning/destination plans
- Traffic plans

References

[1] M. Aaltonen, J. M. MacDonald, P. Martikainen, and J. Kivivuori. Examining the generality of the unemployment–crime association. *Criminology*, 51(3):561–594, 2013.

[2] H. Abdi and L. J. Williams. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4):433–459, 2010.

[3] D. Ammons. *Municipal Benchmarks: Assessing Local Performance and Establishing Community Standards*. Routledge, 2014.

[4] L. G. Anthopoulos, M. Janssen, and V. Weerakkody. Comparing smart cities with different modeling approaches. In *Proceedings of the 24th International Conference on World Wide Web*, pages 525–528. ACM, 2015.

[5] B. Bieleman, S. Biesma, J. Snippe, and A. Beelen. Drugsgerelateerde overlast. Technical report, Intraval, 2009.

[6] C. L. Britt. Reconsidering the unemployment and crime relationship: Variation by age group and historical period. *Journal of Quantitative Criminology*, 13(4):405–428, 1997.

[7] W. J. M. de Haan. Evaluatie integraal veiligheidsbeleid. 1997.

[8] E. Devroe et al. Het kopje naast het schoteltje. over de gestructureerde versnippering in het integrale veiligheidsbeleid. *Cahiers Politiestudies*, 2014(1):22, 2014.

[9] Y. J. Fan and C. Kamath. On the selection of dimension reduction techniques for scientific applications. In *Real World Data Mining Applications*, pages 91–121. Springer, 2015.

[10] C. F. Foley. Welfare payments and crime. *The review of Economics and Statistics*, 93(1):97–112, 2011.

[11] T. George, S. Longworth, and M. O'Dell. Introducing, understanding, and using the ici 300 peer cities identification tool. *population*, 199(56,762):128–009, 2016.

[12] D. Gering. Linear and nonlinear data dimensionality reduction. *fullfillment of the Area Exam doctoral requirements*, 2002.

[13] M. Howlett, M. Ramesh, and A. Perl. *Studying public policy: Policy cycles and policy subsystems*, volume 3. Oxford University Press Toronto, 1995.

[14] L. Kaufman and P. J. Rousseeuw. *Finding groups in data: an introduction to cluster analysis*, volume 344. John Wiley & Sons, 2009.

[15] E. Keogh and A. Mueen. Curse of dimensionality. In *Encyclopedia of Machine Learning*, pages 257–258. Springer, 2011.

[16] K. Kerdprasop, N. Kerdprasop, and P. Sattayatham. Weighted k-means for density-biased clustering. In *International Conference on Data Warehousing and Knowledge Discovery*, pages 488–497. Springer, 2005.

[17] H. Kwon, J. Fan, and P. Kharchenko. Comparison of principal component analysis and t-stochastic neighbor embedding with distance metric modifications for single-cell rna-sequencing data analysis. *bioRxiv*, page 102780, 2017.

[18] C. Lammam, J. Emes, and H. MacIntyre. Comparing municipal government finances in metro vancouver. 2014.

[19] D. T. Larose. k-nearest neighbor algorithm. *Discovering Knowledge in Data: An Introduction to Data Mining*, pages 90–106, 2005.

[20] L. v. d. Maaten and G. Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(Nov):2579–2605, 2008.

[21] J. MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA., 1967.

[22] S. Overkamp and A. Tollenaar. *Integraal veiligheidsbeleid en planverplichting: versteviging gemeentelijke regierol?* University of Groningen [host], 2013.

[23] R. Steele and I. S. General. Iso 37120 standard on city indicators–how they help city leaders set tangible targets, including service quality and quality of life. *Centre for Liveable*

Cites, Singapore [cited 26 October 2015]. Available from Internet: http://www. clc. gov. sg/documents/Lectures/2014/CLC-2014-Rob-steele-Terry-Hill. pdf, 2014.

[24] T. Van den Broeck. De veiligheidscontracten: Evalueerbaarheid en een stand van zaken—. *Panopticon*, 16(1):58, 1995.

[25] G. VandeWalle, E. Wayenberg, A. Dormaels, and M. Easton. De plaats en betekenis van evaluatie in de interbestuurlijke veiligheids-en preventieplannen. *Panopticon*, 31(3):40–57, 2010.

[26] K. Wittebrood. Trends in jeugdgeweld. *Justitiële verkenningen*, 26(1):21–34, 2000.

A Source code

The python script to the ISP file locations from the CCV can be found at:

https://www.pcdekoning.com/thesis/thijs/1_get_ccv_data.py

The python script that retrieves all ISPs from the TSV file can be found at:

https://www.pcdekoning.com/thesis/thijs/2_retrieve_ivp.py

The python script to retrieve and preprocess the crime statistics from the CBS can be found at:

https://www.pcdekoning.com/thesis/thijs/3_transform_cbs_data.py

The complete dataset of gathered ISP files in PDF format can be found at:

https://www.pcdekoning.com/thesis/thijs/ivp.zip

The R script that called the functions and produces the results can be found at:

https://www.pcdekoning.com/thesis/thijs/Main.R

The R script that contains the functions can be found at:

https://www.pcdekoning.com/thesis/thijs/Main_functions.R

B CBS crime indicators

1	Misdrijven, totaal
2	1 Vermogensmisdrijven
3	1.1 Diefstal/verduistering en inbraak
4	1.1.1 Diefstal en inbraak met geweld
5	1.1.2 Diefstal en inbraak zonder geweld
6	1.2 Bedrog
7	1.2.1 Oplichting
8	1.2.2 Flessentrekkerij
9	1.2.3 Bedrog (overig)
10	1.3 Valsheidsmisdrijf
11	1.3.1 Muntmisdrijf
12	1.3.2 Valsheid in zegels en merken
13	1.3.3 Valsheid in geschriften
14	1.4 Heling
15	1.5 Afpersing en afdreiging
16	1.6 Bankbreuk
17	1.7 Witwassen
18	1.8 Vermogensmisdrijf (overig)
19	2 Vernielingen, misdr.openb.orde/gezag
20	2.1 Vernieling en beschadiging
21	2.1.1 Vernieling aan auto
22	2.1.2 Vernieling aan openbaar gebouw
23	2.1.3 Vernieling middel openb.vervoer
24	2.1.4 Dierenmishandeling
25	2.1.5 Vernieling, beschadiging (overig)
26	2.2 Openbare orde misdrijf
27	2.2.1 Openlijke geweldpleging
28	2.2.1.1 Openlijk geweld tegen persoon
29	2.2.1.2 Openlijk geweld tegen goed
30	2.2.2 Huisvredebreuk

31	2.2.3 Lokaalvredebreuk
32	2.2.4 Computervredebreuk
33	2.2.5 Discriminatie
34	2.2.6 Openbare orde misdrijf (overig)
35	2.3 Brandstichting/ontploffing
36	2.4 Openbaar gezag misdrijf
37	2.4.1 Niet opvolgen van ambtelijk bevel
38	2.4.2 Wederspannigheid
39	2.4.3 Valse aangifte
40	2.4.5 Verblijf ongewenste vreemdeling
41	2.4.6 Openbaar gezag misdrijf (overig)
42	3 Gewelds- en seksuele misdrijven
43	3.1 Mishandeling
44	3.2 Bedreiging en stalking
45	3.2.1 Bedreiging
46	3.2.2 Stalking
47	3.3 Seksueel misdrijf
48	3.3.1 Aanranding
49	3.3.2 Verkrachting
50	3.3.3 Schennis der eerbaarheid
51	3.3.4 Ontucht met minderjarige
52	3.3.5 Pornografie
53	3.3.6 Ontucht met misbruik van gezag
54	3.3.7 Seksueel misdrijf (overig)
55	3.4 Levensmisdrijf
56	3.5 Vrijheidsbeneming/gijzeling
57	3.6 Mensenhandel, mensensmokkel
58	3.7 Geweldsmisdrijf (overig)
59	4 Misdrijven WvSr (overig)
60	5 Verkeersmisdrijven
61	5.1 Verlaten plaats ongeval
62	5.2 Rijden onder invloed

63	5.3 Rijden tijdens ontzegging
64	5.4 Rijden tijdens rijverbod
65	5.5 Voeren vals kenteken
66	5.6 Joyriding
67	5.7 Weigeren blaastest/bloedonderzoek
68	5.8 Verkeersmisdrijf (overig)
69	6 Drugsmisdrijven
70	6.1 Harddrugs
71	6.2 Softdrugs
72	7 (Vuur)wapenmisdrijven
73	9 Misdrijven overige wetten
74	9.1 Militair misdrijf
75	9.2 Misdrijf (overig)
218	6.3 Drugsmisdrijf (overig)

C English translation of the 9 main crime indicators

Indicator in Dutch	English translation
Misdrijven, totaal	Total crimes
Vermogensmisdrijven	Property crimes
Vernielingen, misdr.openb.orde/gezag	Destruction, crimes public order/authority
Gewelds- en seksuele misdrijven	Violence and sexual crimes
Misdrijven WvSr (overig)	Crimes criminal law (Other)
Verkeersmisdrijven	Traffic crimes
Drugsmisdrijven	Drug crimes
(Vuur)wapenmisdrijven	(Fire)arms crimes
Misdrijven overige wetten	Crimes other laws

D Municipality interview questions

Vragen

Wordt er bij de totstandkoming van een integraal veiligheidsplan gekeken naar (historische) cijfers van de gemeente?

Zo ja, op welke cijfers en uit welke bronnen?

Wordt er bij de totstandkoming van een integraal veiligheidsplan gekeken naar de verhouding tussen de cijfers van de gemeente en de cijfers van het landelijk- en/of regionaal gemiddelde?

Wordt er bij de totstandkoming van een integraal veiligheidsplan gekeken naar eerdere eigen plannen en/of plannen van andere gemeenten?

Zo ja, op basis van welke criteria kiest u gemeenten om uw gemeente mee te vergelijken?

Zijn er statistieken die door de gemeente wel als problematisch worden gezien, maar niet worden verwerkt in het integraal veiligheidsplan omdat de focus ergens anders gelegd wordt?

Kunt u aangeven welke soorten misdrijven relevant zijn bij de totstandkoming van een integraal veiligheidsplan?

Soort misdrijf

Toelichting

Wel / niet relevant

Misdrijven, totaal

Totaal aantal misdrijven binnen de gemeente

Vermogensmisdrijven

Diefstal, inbraak, afpersing, witwassen

Vernielingen, misdragingen openbare orde / gezag

Vernielingen, dierenmishandeling, geweldpleging, huisvredebreuk, discriminatie

Gewelds- en seksuele misdrijven

Mishandeling, bedreiging en stalking, seksuele misdrijven, mensenhandel

Misdrijven WvSr

Misdrijven Wetboek van Strafrecht

Verkeersmisdrijven

Verlaten plaats ongeval, rijden onder invloed, joyriding,

Drugsmisdrijven

Harddrugs, softdrugs

(Vuur)wapenmisdrijven

(Vuur)wapenmisdrijven

Misdrijven overige wetten

Militaire misdrijven, overige misdrijven

```
Hoe wordt ervoor gezorgd dat het plan goed wordt nageleefd door verschillende
    instanties en de gemeente zelf?
Hoe wordt het succes van een integraal veiligheidsplan bij uw gemeente gemeten?
Wat kan uw gemeente op het gebied van veiligheid leren van andere gemeenten? Zijn hier
    al samenwerkingen voor aanwezig?
```

E Cluster labels for K-means clustering with t-SNE with variables weighted for *Misdrijven, totaal*

Table 2: Cluster labels for K-means clustering with t-SNE with variables weighted for *Misdrijven, totaal*. Refers to figure 9.

Municipality name	Cluster
Beverwijk	1
Culemborg	1
Doesburg	1
Doetinchem	1
Enkhuizen	1
Geldrop-Mierlo	1
Goes	1
Gorinchem	1
Heemskerk	1
Hellevoetsluis	1
Hoogeveen	1
Maassluis	1
Meppel	1
Rheden	1
Westervoort	1
Zwijndrecht	1
Aa en Hunze	2
Aalten	2
Barneveld	2
Bedum	2
Duiven	2
Hollands Kroon	2
Leeuwarderadeel	2
Raalte	2
Ten Boer	2
Voorst	2
West Maas en Waal	2
Wierden	2
Zeewolde	2
Zuidhorn	2

Continued on next page

Table 2 – continued from previous page

Municipality name	Cluster
Zwartewaterland	2
Albrandswaard	3
Barendrecht	3
Beuningen	3
Borne	3
Brielle	3
Hendrik-Ido-Ambacht	3
Krimpenerwaard	3
Langedijk	3
Lingewaard	3
Oostzaan	3
Stede Broec	3
Uitgeest	3
Uithoorn	3
Veldhoven	3
Waddinxveen	3
Wijk bij Duurstede	3
Wormerland	3
Zuidplas	3
Beesel	4
Best	4
Boxtel	4
Druten	4
Etten-Leur	4
Gilze en Rijen	4
Losser	4
Oosterhout	4
Oss	4
Terneuzen	4
Valkenswaard	4
Waalwijk	4
Weert	4
Wijchen	4
Bergen (NH.)	5
Bodegraven-Reeuwijk	5
Bunnik	5
Bunschoten	5
Castricum	5
De Ronde Venen	5
Hardinxveld-Giessendam	5
Lansingerland	5
Leusden	5
Midden-Delfland	5
Pijnacker-Nootdorp	5
Stichtse Vecht	5
Utrechtse Heuvelrug	5
Woerden	5
Dongen	6
Eemnes	6
Elburg	6
Epe	6
Grave	6
Haarlemmerliede en Spaarnwoude	6
Hattem	6
Heusden	6
Loon op Zand	6

Continued on next page

Table 2 – continued from previous page

Municipality name	Cluster
Nunspeet	6
Overbetuwe	6
Rijssen-Holten	6
Voerendaal	6
's-Hertogenbosch	7
Alkmaar	7
Almere	7
Amersfoort	7
Apeldoorn	7
Bergen op Zoom	7
Breda	7
Deventer	7
Dordrecht	7
Haarlem	7
Helmond	7
Hengelo	7
Hilversum	7
Middelburg	7
Nissewaard	7
Roosendaal	7
Vaals	7
Zaanstad	7
Zoetermeer	7
Zwolle	7
Alblasserdam	8
Amstelveen	8
Diemen	8
Heerhugowaard	8
IJsselstein	8
Katwijk	8
Krimpen aan den IJssel	8
Leiderdorp	8
Oegstgeest	8
Papendrecht	8
Ridderkerk	8
Sliedrecht	8
Soest	8
Voorschoten	8
Cromstrijen	9
Drimmelen	9
Giessenlanden	9
Goeree-Overflakkee	9
Heerde	9
Koggenland	9
Korendijk	9
Molenwaard	9
Montfoort	9
Nijkerk	9
Oldebroek	9
Oudewater	9
Reimerswaal	9
Scherpenzeel	9
Strijen	9
Alphen aan den Rijn	10
Ede	10
Haarlemmermeer	10

Continued on next page

Table 2 – continued from previous page

Municipality name	Cluster
Houten	10
Westland	10
Alphen-Chaam	11
Ameland	11
Baarle-Nassau	11
Bergeijk	11
Bladel	11
Boekel	11
Dalfsen	11
De Wolden	11
Dinkelland	11
Eersel	11
Hilvarenbeek	11
Landerd	11
Mill en Sint Hubert	11
Nederweert	11
Oirschot	11
Renswoude	11
Reusel-De Mierden	11
Schiermonnikoog	11
Sint Anthonis	11
Staphorst	11
Tubbergen	11
Aalburg	12
Aalsmeer	12
Drechterland	12
Eijsden-Margraten	12
Kapelle	12
Lingewaal	12
Lopik	12
Neder-Betuwe	12
Oisterwijk	12
Olst-Wijhe	12
Opmeer	12
Sint-Michielsgestel	12
Woudenberg	12
Beemster	13
Bernheze	13
Buren	13
Cranendonck	13
Geldermalsen	13
Haaren	13
Heeze-Leende	13
Hulst	13
Laarbeek	13
Maasdriel	13
Neerijnen	13
Sint-Oedenrode	13
Son en Breugel	13
Veere	13
Zoeterwoude	13
Zundert	13
Brummen	14
Cuijk	14
Dronten	14
Ermelo	14

Continued on next page

Table 2 – continued from previous page

Municipality name	Cluster
Gennep	14
Haaksbergen	14
Hardenberg	14
Hellendoorn	14
Medemblik	14
Menameradiel	14
Montferland	14
Noordenveld	14
Oude IJsselstreek	14
Putten	14
Schagen	14
Tholen	14
Twenterand	14
Tytsjerksteradiel	14
Uden	14
Zaltbommel	14
Binnenmaas	15
Hillegom	15
Kaag en Braassem	15
Meerssen	15
Nieuwkoop	15
Oud-Beijerland	15
Stein	15
Teylingen	15
Waalre	15
Werkendam	15
Wijdemeren	15
Asten	16
Berkelland	16
Borsele	16
Bronckhorst	16
Deurne	16
Gemert-Bakel	16
Hof van Twente	16
Horst aan de Maas	16
Leudal	16
Littenseradiel	16
Lochem	16
Moerdijk	16
Noord-Beveland	16
Ommen	16
Peel en Maas	16
Schouwen-Duiveland	16
Sluis	16
Someren	16
Veghel	16
Gouda	17
Harderwijk	17
Hoorn	17
Huizen	17
Kampen	17
Leerdam	17
Leidschendam-Voorburg	17
Nieuwegein	17
Purmerend	17
Renkum	17

Continued on next page

Table 2 – continued from previous page

Municipality name	Cluster
Rijswijk	17
Veenendaal	17
Velsen	17
Weesp	17
Zandvoort	17
Zeist	17
Echt-Susteren	18
Geertruidenberg	18
Gulpen-Wittem	18
Halderberge	18
Maasgouw	18
Nuth	18
Roerdalen	18
Rucphen	18
Schijndel	18
Steenbergen	18
Tynaarlo	18
Woensdrecht	18
Woudrichem	18
Achtkarspelen	19
Dantumadiel	19
Delfzijl	19
Dongeradeel	19
Franekeradeel	19
Heerenveen	19
Kollumerland en Nieuwkruisland	19
Menterwolde	19
Oldambt	19
Pekela	19
Smallingerland	19
Stadskanaal	19
Veendam	19
Bellingwedde	20
Bergen (L.)	20
Borger-Odoorn	20
Boxmeer	20
Coevorden	20
De Marne	20
Eemsmond	20
Ferwerderadiel	20
Grootegast	20
Loppersum	20
Marum	20
Midden-Drenthe	20
Noordoostpolder	20
Oost Gelre	20
Ooststellingwerf	20
Slochteren	20
Venray	20
Vlagwedde	20
Weststellingwerf	20
het Bildt	21
Leek	21
Opsterland	21
Rijnwaarden	21
Steenwijkerland	21

Continued on next page

Table 2 – continued from previous page

Municipality name	Cluster
Winsum	21
Winterswijk	21
Baarn	22
Blaricum	22
Bloemendaal	22
De Bilt	22
Edam-Volendam	22
Haren	22
Heemstede	22
Heiloo	22
Landsmeer	22
Laren	22
Ouder-Amstel	22
Rozendaal	22
Vught	22
Wassenaar	22
Waterland	22
’s-Gravenhage	23
Amsterdam	23
Arnhem	23
Delft	23
Eindhoven	23
Enschede	23
Groningen	23
Heerlen	23
Leeuwarden	23
Leiden	23
Maastricht	23
Nijmegen	23
Rotterdam	23
Tilburg	23
Utrecht	23
Wageningen	23
Almelo	24
Appingedam	24
Assen	24
Brunssum	24
Capelle aan den IJssel	24
Den Helder	24
Emmen	24
Harlingen	24
Kerkrade	24
Landgraaf	24
Lelystad	24
Roermond	24
Schiedam	24
Sittard-Geleen	24
Tiel	24
Venlo	24
Vlaardingen	24
Vlissingen	24
Zutphen	24
Beek	25
Goirle	25
Heumen	25
Mook en Middelaar	25

Continued on next page

Table 2 – continued from previous page

Municipality name	Cluster
Nuenen, Gerwen en Nederwetten	25
Oldenzaal	25
Onderbanken	25
Rhenen	25
Schinnen	25
Simpelveld	25
Vianen	25

F Cluster labels for K-means clustering with t-SNE with variables weighted for *5 Verkeersmisdrijven*

Table 3: Cluster labels for K-means clustering with t-SNE with variables weighted for *5 Verkeersmisdrijven*. Refers to figure 10.

Municipality name	Cluster
Barneveld	1
Berkelland	1
Bronckhorst	1
Dalfsen	1
De Wolden	1
Deurne	1
Dinkelland	1
Goeree-Overflakkee	1
Hof van Twente	1
Hollands Kroon	1
Horst aan de Maas	1
Landerd	1
Leudal	1
Medemblik	1
Peel en Maas	1
Staphorst	1
Veghel	1
Barendrecht	2
Beuningen	2
Brielle	2
Harderwijk	2
Kampen	2
Leerdam	2
Stede Broec	2
Uitgeest	2
Wormerland	2
Aa en Hunze	3
Bernheze	3
Geldermalsen	3
Laarbeek	3
Lopik	3
Olst-Wijhe	3
Opmeer	3
Strijen	3
Voorst	3
West Maas en Waal	3
Wierden	3
Zeewolde	3
Boxtel	4
Cromstrijen	4
Eemnes	4
Etten-Leur	4
Grave	4
Heumen	4
Korendijk	4
Loon op Zand	4
Mook en Middelaar	4

Continued on next page

Table 3 – continued from previous page

Municipality name	**Cluster**
Nieuwkoop	4
Oosterhout	4
Sint-Michielsgestel	4
Terneuzen	4
Valkenswaard	4
Voerendaal	4
Waalwijk	4
Werkendam	4
Almelo	5
Appingedam	5
Assen	5
Capelle aan den IJssel	5
Den Helder	5
Deventer	5
Dordrecht	5
Emmen	5
Hengelo	5
Leeuwarden	5
Lelystad	5
Nissewaard	5
Schiedam	5
Venlo	5
Vlaardingen	5
Zwolle	5
Amstelveen	6
Baarn	6
Bergen (NH.)	6
Bunnik	6
De Bilt	6
Haren	6
Heemstede	6
Heiloo	6
Oegstgeest	6
Ouder-Amstel	6
Renkum	6
Utrechtse Heuvelrug	6
Voorschoten	6
Vught	6
Waalre	6
Weesp	6
Wijdemeren	6
Zandvoort	6
's-Gravenhage	7
Amsterdam	7
Delft	7
Eindhoven	7
Groningen	7
Leiden	7
Maastricht	7
Nijmegen	7
Rotterdam	7
Tilburg	7
Utrecht	7
Wageningen	7
Aalten	8
Bedum	8

Continued on next page

Table 3 – continued from previous page

Municipality name	Cluster
Dronten	8
Duiven	8
Grootegast	8
Haaksbergen	8
Hardenberg	8
Leeuwarderadeel	8
Littenseradiel	8
Marum	8
Menameradiel	8
Ten Boer	8
Twenterand	8
Tytsjerksteradiel	8
Zuidhorn	8
Zwartewaterland	8
Binnenmaas	9
Goirle	9
Hillegom	9
Kaag en Braassem	9
Meerssen	9
Nuenen, Gerwen en Nederwetten	9
Rhenen	9
Schinnen	9
Simpelveld	9
Stein	9
Bergen (L.)	10
Borger-Odoorn	10
Boxmeer	10
Midden-Drenthe	10
Oost Gelre	10
Oude IJsselstreek	10
Raalte	10
Slochteren	10
Steenwijkerland	10
Winterswijk	10
Aalburg	11
Aalsmeer	11
Beek	11
Echt-Susteren	11
Eijsden-Margraten	11
Geertruidenberg	11
Gulpen-Wittem	11
Halderberge	11
Kapelle	11
Lingewaal	11
Maasgouw	11
Nuth	11
Roerdalen	11
Rucphen	11
Schijndel	11
Steenbergen	11
Tynaarlo	11
Woensdrecht	11
Woudenberg	11
Woudrichem	11
Zaltbommel	11
Beesel	12

Continued on next page

Table 3 – continued from previous page

Municipality name	Cluster
Brummen	12
Cuijk	12
Ermelo	12
Gennep	12
Gilze en Rijen	12
Hellendoorn	12
Losser	12
Montferland	12
Noordenveld	12
Rijnwaarden	12
Rijssen-Holten	12
Schagen	12
Tholen	12
Uden	12
Weert	12
Winsum	12
Beemster	13
Buren	13
Cranendonck	13
Eersel	13
Haaren	13
Heeze-Leende	13
Hulst	13
Lochem	13
Maasdriel	13
Neerijnen	13
Schouwen-Duiveland	13
Sint-Oedenrode	13
Son en Breugel	13
Veere	13
Zoeterwoude	13
Zundert	13
Bellingwedde	14
Coevorden	14
De Marne	14
Eemsmond	14
Ferwerderadiel	14
Loppersum	14
Noordoostpolder	14
Ooststellingwerf	14
Venray	14
Vlagwedde	14
Weststellingwerf	14
Albrandswaard	15
Borne	15
Bunschoten	15
Culemborg	15
Geldrop-Mierlo	15
Goes	15
Haarlemmerliede en Spaarnwoude	15
Hardinxveld-Giessendam	15
Langedijk	15
Oostzaan	15
Soest	15
Uithoorn	15
Waddinxveen	15

Continued on next page

Table 3 – continued from previous page

Municipality name	Cluster
Wijk bij Duurstede	15
Zuidplas	15
Alphen-Chaam	16
Ameland	16
Asten	16
Baarle-Nassau	16
Bergeijk	16
Bladel	16
Boekel	16
Borsele	16
Gemert-Bakel	16
Hilvarenbeek	16
Mill en Sint Hubert	16
Moerdijk	16
Nederweert	16
Noord-Beveland	16
Oirschot	16
Ommen	16
Renswoude	16
Reusel-De Mierden	16
Schiermonnikoog	16
Sint Anthonis	16
Sluis	16
Someren	16
Tubbergen	16
Blaricum	17
Bloemendaal	17
Hilversum	17
Houten	17
Huizen	17
Laren	17
Leiderdorp	17
Leidschendam-Voorburg	17
Nieuwegein	17
Rijswijk	17
Rozendaal	17
Vaals	17
Wassenaar	17
Zeist	17
Alblasserdam	18
Diemen	18
Heerhugowaard	18
Hendrik-Ido-Ambacht	18
IJsselstein	18
Katwijk	18
Krimpen aan den IJssel	18
Papendrecht	18
Ridderkerk	18
Sliedrecht	18
Achtkarspelen	19
Dantumadiel	19
Delfzijl	19
Dongeradeel	19
Franekeradeel	19
Harlingen	19
Heerenveen	19

Continued on next page

Table 3 – continued from previous page

Municipality name	**Cluster**
het Bildt	19
Kollumerland en Nieuwkruisland	19
Leek	19
Menterwolde	19
Oldambt	19
Opsterland	19
Pekela	19
Smallingerland	19
Stadskanaal	19
Veendam	19
Bodegraven-Reeuwijk	20
Castricum	20
De Ronde Venen	20
Edam-Volendam	20
Enkhuizen	20
Landsmeer	20
Lansingerland	20
Leusden	20
Midden-Delfland	20
Oud-Beijerland	20
Pijnacker-Nootdorp	20
Stichtse Vecht	20
Teylingen	20
Veldhoven	20
Waterland	20
Woerden	20
Arnhem	21
Bergen op Zoom	21
Brunssum	21
Enschede	21
Heerlen	21
Helmond	21
Kerkrade	21
Landgraaf	21
Middelburg	21
Roermond	21
Roosendaal	21
Sittard-Geleen	21
Tiel	21
Vlissingen	21
Zutphen	21
's-Hertogenbosch	22
Alkmaar	22
Almere	22
Alphen aan den Rijn	22
Amersfoort	22
Apeldoorn	22
Breda	22
Ede	22
Haarlem	22
Haarlemmermeer	22
Oss	22
Westland	22
Zaanstad	22
Zoetermeer	22
Drechterland	23

Continued on next page

Table 3 – continued from previous page

Municipality name	Cluster
Drimmelen	23
Elburg	23
Giessenlanden	23
Hattem	23
Heerde	23
Koggenland	23
Molenwaard	23
Montfoort	23
Neder-Betuwe	23
Oisterwijk	23
Oldebroek	23
Oudewater	23
Putten	23
Reimerswaal	23
Scherpenzeel	23
Best	24
Doetinchem	24
Dongen	24
Druten	24
Epe	24
Heusden	24
Hoogeveen	24
Krimpenerwaard	24
Lingewaard	24
Meppel	24
Nijkerk	24
Nunspeet	24
Oldenzaal	24
Onderbanken	24
Overbetuwe	24
Vianen	24
Wijchen	24
Beverwijk	25
Doesburg	25
Gorinchem	25
Gouda	25
Heemskerk	25
Hellevoetsluis	25
Hoorn	25
Maassluis	25
Purmerend	25
Rheden	25
Veenendaal	25
Velsen	25
Westervoort	25
Zwijndrecht	25