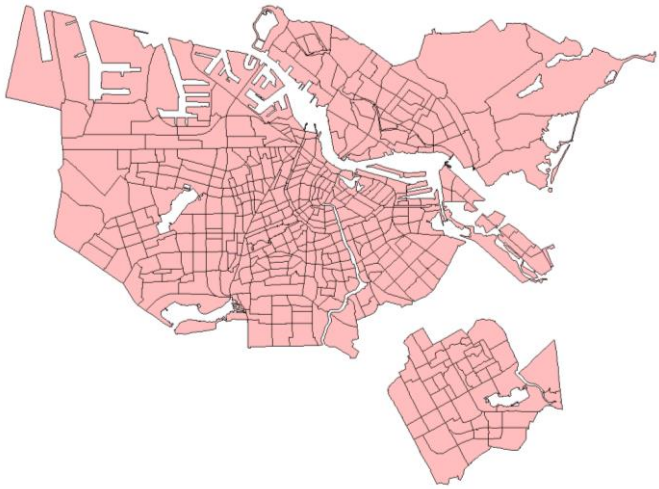
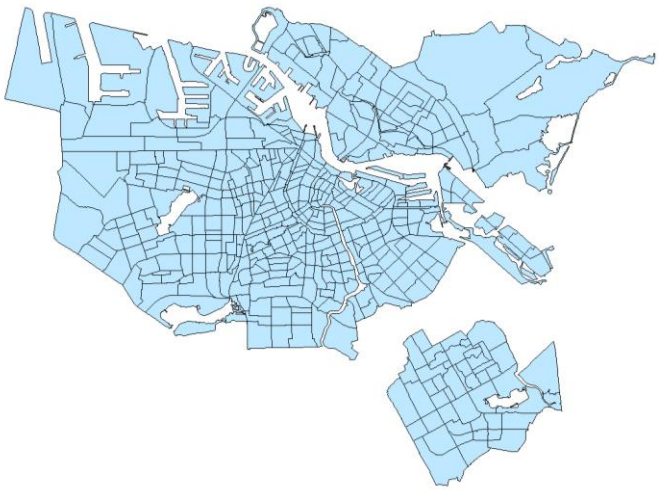
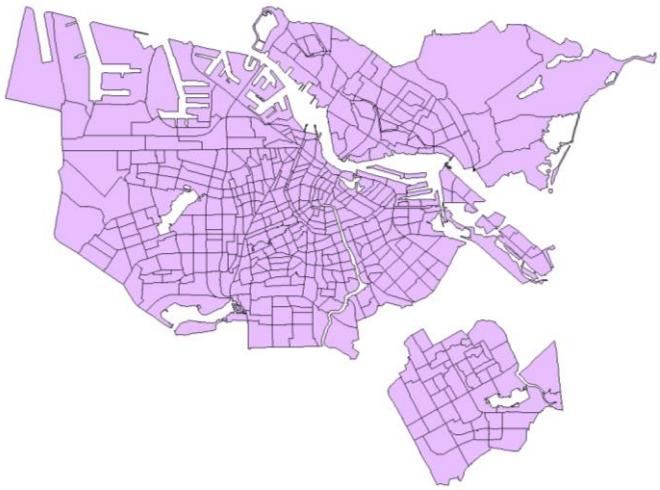
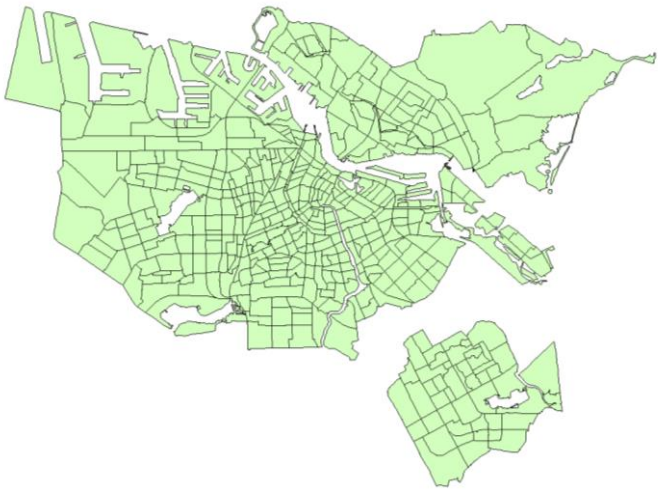
Where people vote what

Creating a model that predicts voting behaviour per neighbourhood









Eva van der Meulen
Rapport for the GI minor: the internship at Kieskompas
Vrije Universiteit Amsterdam
evavandmeulen@gmail.com

Preface

This rapport is written because of the internship I did at Kieskompas from November 2016 until February 2017. I would like to thank Eric Koomen for his guidance and advice. And also Eduardo Dias for the guidance in the beginning. And last but not least I want to thank Kieskompas for providing the internship.

Content

The problem

Predicting voting behaviour has been a popular subject for a very long time. But predicting voting behaviour on a small geographical area is a rather untouched area. Especially in the Netherlands since it doesn't use a district voting system. But knowing the chance that a certain party has of collecting votes in a certain neighbourhood can be rather useful even though it doesn't matter where voters live in the end. Imagine the labour party of the Netherlands wants to go into the country to talk to the people to collect more votes. Knowing where these potential voters live might help in this process. But how can this be done? There are a lot of factors that influence voting behaviour. Personal characteristics influence voting (Johnston & Pattie, 2006; Walks, 2006). But location factors also influence voting behaviour, the location of residence can also have an influence on voting behaviour (Walks, 2010; McAllister, ?;Van Gent, Jansen & Smits, 2012; Walks, 2006). A neighbourhood where a lot of poor people live of Dutch decent will probably have another dominant party than a neighbourhood where a lot of poor people live of non-western decent.

But is it possible to create a model that can calculate the chance that a certain party has to collect votes in a certain neighbourhood? Is there enough data available with a sufficient quality? This rapport will look into these questions. The research question of this paper is:

To what extent is it possible to make a plausible model to predict voting behaviour per neighbourhood?

The goal of this rapport is to create a model that will calculate the voting behaviour of every neighbourhood. What chance does a specific party have to get votes in a certain neighbourhood? In the first section the literature about location and personal characteristics influencing voting behaviour will be reviewed. The second section will look into the available data and the quality of the data. The third section will determine the factors that influence voting behaviour and to what extent. In the last section the outcomes will be discussed and also compared to reality.

Literature

The main goal of this rapport is to create a model that can predict voting behaviour per neighbourhood and municipality. But how can you predict voting behaviour. According to Walks (2006) personal and location characteristics of voters play a significant role in predicting voting behaviour.

Who votes what

People with different personal characteristics are likely to also vote in a different way. People with a different age, a different gender, a different income, a different ethnical background, a different education level, a different occupation or a different religion might all also vote different. According to the Columbia and later Michigan school of political science, political position is tied to personal sociodemographic characteristics and these factors are seen as the most important ones (Walks, 2006). But what do these sociodemographic say about voting behaviour? Who votes what if one would look at personal characteristics?

According Johnston and Pattie (2006) age matters when it comes to voting. They state that, in general, older people are more likely to vote conservative than younger people. For example because older people usually have more property and the conservative parties tend to protect these people more than progressive parties. The older people get, the more likely they are to vote conservative. But older people are less likely to vote PVV, which is the conservative populist party of the Netherlands, than their counterparts according to Van Gent et al. (2013). On the other side of the spectrum, young people are more likely to vote liberal progressive. To conclude, Walks (2006) states that age is an important factor for party preferences.

Man and women also tend to vote different. Women are less likely to vote PVV than man according to Van Gent et al. (2013). They found that man are more likely to vote anti-immigration than women. Van der Burg (2003) came to the same conclusions about LPF supporters, an anti-immigration populist party in the Netherland that received quite some votes in the 2002 elections. It isn't a surprise then that women also tend to vote more left wing and progressive than man.

Education level can also influence voting behaviour. Highly educated people are more likely to vote on the liberal progressive parties than lower educated people according to Johnston and Pattie (2006). Walks (2006) also came to the same conclusion. In relation to this, they are less likely to vote right wing populist parties (Van Gent, 2013; Van der Burg).

Income level can also influence voting behaviour. According to van den Burg (2003) people with a low income are more likely to vote right wing populist parties. And people with a high income are on the other hand less likely to vote right wing populist parties (Van Gent et al., 2013). Walks (2006) states that income appears most important together with occupation for driving preferences at the extremes of left and right.

According to Walks (2006) occupation is the most important variable to understand differences in party preferences. Like income, it says something about the extremes of left and right. He also states that people who are employed in the artistic or teaching sector are less likely to support the centre right conservative parties. People that are employed in the white collar sector are more likely to support conservatives than people that are employed in the blue collar sector. The liberal democratic parties were also more likely to get support from white collar workers (Johnston & Pattie, 2006).

Religion can also play a role in party preference. People that are active participants of religions tend to be influenced by religion in their party preference. The core components and ideology of a religion can link to specific parties (Johnston & Pattie, 2006). Although nowadays this will probably not have an influence on a lot of people since half of the Dutch population is not religious anymore (CBS, 2016). But religious people are less likely to support right wing populist parties (Van Den Burg, 2003; Van Gent et al., 2013). There are still quite some religious parties. One could argue that non-religious people would probably not support these parties.

The last sociodemographic variable that can influence party preference is ethnic background. After occupation this is the strongest predictor of party preferences according to Walks (2006). It comes with no surprise that people with a non-western background are less likely to vote PVV, an anti-immigration party (Van Gent et al., 2013). Also, people with a non-western background are usually a religious minority, if they are religious, and they tend to vote Labour (Jonhston & Pattie, 2006).

Where do people vote what?

The location of a voter can also have an influence on their voting behaviour. According to Johnston and Pattie (2006) where a voter lives can have a strong impact, and it may even be the major influence. But which location factors have an influence?

In the last decades, inner cities and suburbs have become polarized in their party preferences and political attitudes (Walks, 2006). He states that suburban residents in Canadian cities are six times more likely to vote conservative than inner city residents. According to Johnston and Pattie (2006) this is also the case in the United Kingdom. In the UK labour became the party of the industrial regions and the large cites , while the conservatives was the party of the country side and the suburban regions. In the Netherlands right-wing radical populist party support is mostly concentrated in suburban areas (Van Gent et al., 2013). So the degree of urban density matters. However, whether these patterns exist due to self-selection of residence area or the political influence of the area (Walks, 2006; Van Gent et al., 2013; Johnston & Pattie, 2006).

The local can influence voting behaviour in two ways. It can influence opinions through personal observation and experience, or it can influence opinions through social interaction, the so called neighbourhood effect (Walks, 2006). This neighbourhood effect has been discussed by many electoral geographers (Pattie & Johnston, 2000; McAllister, Johnston, Pattie, Tunstall, Dorling and Rossiter 2001). But several scholars had trouble finding proof that people influence each other by talking over the garden fence (McAllister et al., 2001; Johnston & Pattie, 2006). Johnston and Pattie (2000) state that political talk does not appear to be a significant mechanism for the existence of a neighbourhood effect. So it can't be proven that people who talk together vote together. And if they do, it is usually only in the at the level of the family and the household (Walks, 2006).

However the neighbourhood can have an influence on voting behaviour in other fields. For example neighbourhood status. Members of each social class were much more likely to vote Labour than conservative in low-status areas than in a high-status areas (McAllister et al., 2001). Whether a voter lives in Amsterdam Zuid Oost (a lower class area) or in Amsterdam Zuid (an upper class area) matters when it comes to voting behaviour, even after taking sociodemographic factors into account. So McAllister et al found some sort of neighbourhood effect. However, they couldn't tell which specific process(es) play a role in this result.

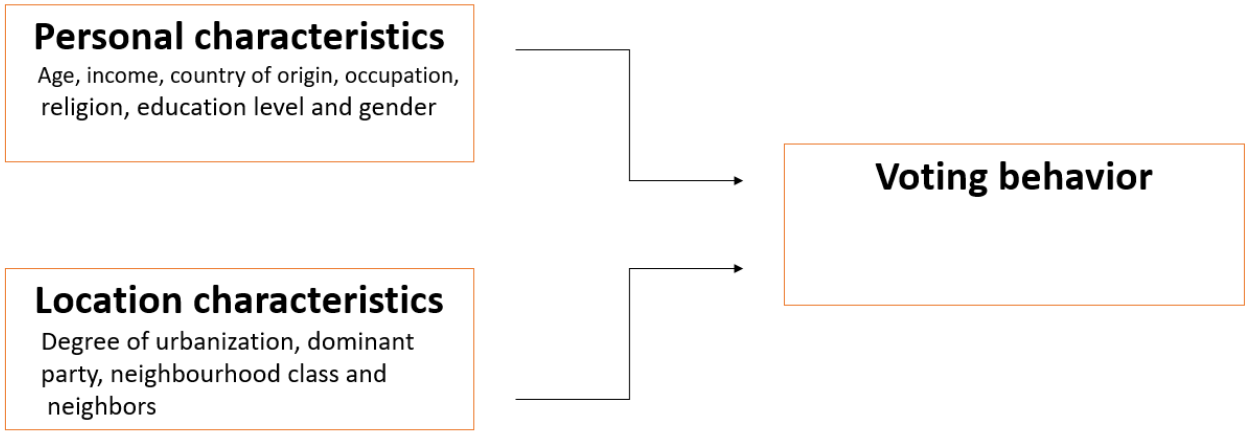
The Dutch political system and electoral geography

The translation of votes into seats, parties, coalitions and governments vary across the world (Agnew, Mamadouh, Secor and Sharp, 2015). In some countries, constituencies and electoral districts are used while in other countries one vote counts as one vote. In the Netherlands, one vote counts as one vote, independent of where a voter lives. So knowing where voters live is not as important as in for example the United states where it does matter where a voter lives. That is why electoral geography is still rather new in the Netherlands.

Another characteristic of the Dutch political system is that it has a multiple party system. There is not a single conservative, progressive, left-wing or right-wing party like in the United Kingdom or the United states. But there are several parties that can be of importance. Since 2012, the PvdA, the left wing social democratic party, and the VVD, the right wing liberal party, are the two parties that govern the Netherlands. They are also the biggest parties at present time. But the next elections in march 2017 might change this (parlement.com, 2017).

Conceptual model

The goal of this rapport is to create a model to predict voting behaviour by using the location characteristics and personal characteristics that are named above. The conceptual model is based on the variables that influence voting behaviour according the literature.



The data

Two datasets are used to conduct this research. The first dataset is used to perform a statistical analysis, the 'nationaal kiezersonderzoek 2012'. This research presents quantitative information behind the voting behaviour of the electoral (Nationaal kiezersonderzoek, 2012). The views of voters about political issues, the sociodemographic characteristics of voters and their voting behaviour are asked and reviewed. The dataset is available as an sav. file. The second dataset is used to combine the outcome of the statistical analysis with the neighbourhoods of the Netherlands, the 'wijken en buurten' dataset (CBS, 2015). This dataset has several characteristics of the neighbourhoods of the Netherlands like number of inhabitants and degree of urbanisation. This dataset is available as a shapefile.

Nationaal kiezersonderzoek 2012

The national voters research of 2012 is conducted over a period from 2006 to 2012. It was published in 2015. This research was done by *stichting Kiezersonderzoek Nederland (SKON).* This is a collaboration of several political scientist from different Dutch universities. The central bureau of statistics of the Netherlands, a government funded research institution, was also involved in this research. Since 1971 this research is done every time before and after the national elections take place. In the 2012 election research the respondents are interviewed at home after the elections, so the voting behaviour is already known. The sample of this dataset is formed by selecting people from the central bureau of statistics sampling frame. Respondents can be selected by personal characteristics within the dataset. In this manner a stratified sample is created so people with all sorts of personal characteristics will be represented in the sample (Van De Vocht, ??). So this dataset can be used to see which personal characteristics will result in specific voting behaviour and to what extent. But what is the quality of the dataset?

The sample size of this dataset is 1677, where 292 observations are missing. So the sample size is of this research is 1385. The response rate of this research was lower than the previous one in 2006. And the different groups of the Dutch population are also less well represented in the 2012 research than the 2006 research. They tried to improve the sample by doing a second wave. However, only people of the populations that were already represented in the sample reacted. So that didn't help to decrease the selectivity. However, there was no difference in response between the different degrees of urbanization. Looking at party choice during the election, the sample is reasonable representative. There is a difference of ten seats in the overall sample. Only the PPV, the anti-immigration populist party, is underrepresented with five seats to little than the outcome of the elections in 2012. Because of these representative problems. The NKO2012 dataset is weighted by the central bureau of statistics. How did they do this?

Certain groups that were underrepresented in the sample count heavier and the groups that are overrepresented count less heavy. Gender, age, degree of urban density, area of the country, marital state, voting behaviour and country of origin were part of the weighting model. All of these factors were weighted for the 2012 research. Some factors like age needed a weight factor, but that wasn't an extremely high factor. But for some characteristics the weight factor was very high. Non-voters counted two times heavier after weighting. They weigh 24,4% more. The same problems was present for the PPV voters. The problem that occurs in these cases is that it is taken for granted that the respondents that do vote PPV are reflection of the true population. The weight model that is used, is the linear weighting method. The goal of this method is to reduce the selectivity of the sample relative to the population. The data of the population is used and is obtained via the municipalities.

The quality and representativeness of the dataset can be argued. However, the weighting is done by following academic standards and correct and up to date data was used. So the outcomes can be used if the outcome of an analysis is reviewed carefully. Analysis about the non-voters and

PVV voters might not be very correct since they are strongly underrepresented in the dataset. So there are some limitations to this dataset since it is probably not trustworthy enough to perform an analysis about non-voters and PPV voters. Other limitations might arise during the analysis.

CBS wijken en buurten

The CBS 'wijken en buurten' datasets are published every year. This dataset is a shapefile and can be opened with geographical information systems like ArcMap. The geometry of this dataset is retrieved from the Dutch Kadaster, which is the government institution that keeps track of the Dutch cadastre, and the neighbourhood and borough data is from the central bureau of statistics. The Netherlands is divided into four scale levels by the CBS. The first one is the province, the second one municipality, the third one is borough (wijken) and the last one is neighbourhood. This dataset contains information about the boroughs and neighbourhoods.

This dataset is very useful to create a model to predict voting behaviour because it contains sociodemographic information about the boroughs and neighbourhoods. This data is retrieved from the basis administration of persons (BRP). So the data contains the whole population of the Netherlands and is therefore representative for the population. This data contains a lot of information about the Dutch inhabitants and it also contains addresses. So it has a geographical content. Some sociodemographic factors are only shown if an area has more than a certain amount of inhabitants. Income for example, is only shown when an area has more than a 100 inhabitants. So this dataset limits analysis to areas that have 100 inhabitants or more.

The dataset about 2014 is used for this research. There is a more recent dataset available. However, this dataset does not contain income levels. And as described in the literature review, income is one of the most important factors. So the choice is made to perform the analysis by using the 2014 dataset instead of the 2015 dataset.

There is one problem with the 'wijken en buurten' dataset. Sometimes neighbourhoods and/or boroughs change in size over time. It can happen that municipalities change because they are fused. Or addresses suddenly belong to a different neighbourhood. So sometimes it is impossible to compare certain areas to different years. The problem of ecological fallacy can also take place. If the average income of a certain area is rather high, it doesn't mean that everybody has a high income. This can influence the probability score for a party.

The variables used

As described in the literature review, there are several factors that influence voting behaviour. Sociodemographic factors but also location factors can influence voting behaviour. A condition to include a variable into the model is that it is present in both the NKO2012 file and the CBS file. Only age, income, country of origin and degree of urbanization where present in both files and had a significant influence on voting behaviour according to the literature. But some variables needed to be recoded to perform the analysis. Country of origin and degree of urbanisation were already the same in both files. However, age and income weren't. The age variable in the NKO2012 file was recoded into the categories of the CBS files by using recode into same variables in SPSS, a statistical analysis program. Income was also recorded in the NKO2012 file. In the CBS wijken en buurten file the income variables were divided between low income, medium income and high income. The NKO2012 income was a ratio variable so it had to be recoded into a ordinal variable. The CBS published the amount of incomes that were seen as low, medium and high in the explanation file about the wijken en buurten file (CBS, 2017). These were used to recode the ratio variable into the interval variables of the CBS wijken en buurten file. After these steps the data was ready to use for the analysis.

The method

This section will explain the process of creating a model to predict voting behaviour per neighbourhood. The statistical analysis that is performed, the probability formula that is used and the process of making the maps in a geographical information system program like Arcmap.

The multinomial logistic regression analysis

In the literature review, the conceptual model was created that contains the variables that influence voting behaviour. But how do these variables influence voting behaviour and to what extent? The multinomial logistic regression is used to answer these questions. When one wants to know the influence of one or more independent nominal/interval variables on more than one dependent nominal/interval variable the logistic regression analysis is the correct statistical analysis to use (Sieben & Linssen, 2009). The data used to perform this statistical analysis is not dichotomous and therefore a multinomial logistic regression must be used.

The recoded data of the NKO2012 dataset is used to perform this statistical analysis. The dependent variable is the party voted on during the 2012 elections and the independent variables are age (15 – 25, 25 – 44, 45 – 64, 65 and up), income (Low income, medium income, high income), country of origin (Dutch decent, western foreigner, non-western foreigner) and degree of urbanisation (Very high, high, medium, low, very low). 12 parties are included in the NKO 2012 research. So the dependent variable contains 12 items. This is quite a lot and there might not be enough degrees of freedom per item combination. That is why a small party was added to the variable 'others'

SPSS was used to perform this analysis. The outcome of this analysis can be seen in table 1. The CDA, the Christian democratic party, is used as the reference category. This party used to be very important. But nowadays it has lost its importance. Today it's only of importance by the elderly with a Christian background (NKO2012, 2015). In the table the Beta's of the multinomial logistic regression are showed. After using the weighting factors described in the previous section, all the variables were significant. The Beta's did not change after applying the weight factors, just the significance. So the weight factors can be seen as reliable. However, some Beta's were incorrect and came with an error. For example the 50plus party, the party for the elderly, has an error with the variables of people with the age of 15 – 24 and 25 – 44. This error probably happened due to the fact that there are not enough respondents in these categories. The same counts for GreenLeft. An error occurred here with the country of origin variables. There were also some other outcomes that didn't make sense. According to the outcome of the analysis, non-western foreigners were more likely to vote PVV than people of a Dutch origin with a Beta of -0,24. There were some other outcomes that weren't reliable as well. However, the Beta's of the VVD, the right wing party, and PvdA, the labour party, did make sense. That is why the choice is made to focus on these two parties only. Another option was adding all the parties with insufficient data to the 'other' category and keep some of the other parties. However, this resulted in rather incorrect Betas as well.

Table 1 Beta's of multinomial logistic regression analysis, 2015

	Beta									
Variable	PvdA	VVD	GroenLinks en PvdD	SP	D66	CUSGP	PVV	50PLUS	other	CDA
Age 15 - 24	1,154	1,176	1,373	2,712	2,018	0,743	1,431	-17,76	3,408	-
Age 25 -44	1,578	1,482	2,159	2,52	2,175	1,494	1,794	-17,24	3,289	-
Age 45 - 64	1,586	1,056	1,531	2,46	1,492	1,475	1,418	0,801	2,327	-
Age 65>	0*	0*	0*	0*	0*	0*	0*	0*	0*	-
Low income	-0,388	-1,109	-0,077	0,65	-1,299	-0,38	0,044	0,692	1,289	-
Medium income	0,373	-0,386	0,202	0,866	-0,57	0,288	0,268	1,325	1,047	-
High income	0*	0*	0*	0*	0*	0*	0*	0*	0*	-
Very high urban	1,928	1,402	3,01	1,733	2,571	1,4	1,621	2,895	2,708	-
high urban	0,382	0,315	0,833	1,123	0,851	0,829	0,676	1,355	1,235	-
medium urban	0,401	0,888	1,1	1,027	0,687	0,691	0,418	1,33	0,919	-
lightly rural	0,01	0,224	0,723	0,914	0,672	0,722	0,164	0,017	0,766	-
very rural	0*	0*	0*	0*	0*	0*	0*	0*	0*	-
Dutch decent	-1,09	-0,083	18,14	0,043	-0,259	0,588	-0,24	16,80	-0,585	-
Western foreigner	0,052	0,544	19,51	0,558	0,4	0,92	0,167	17,95	-0,049	-
Non-western foreigner	0*	0*	0*	0*	0*	0*	0*	0*	0*	-

The reference category is CDA
**this parameter is set to zero because it is redundant*

In table 2. the Beta's of the multinomial logistic regression analysis of the VVD and PvdA can be seen. The other parties were put together in the 'other' category. The other category is used as the reference category. Because all these parties combined are very divers, it is believed that this can be used as a correct reference category. The Beta's are smaller than in the previous analysis but they present a logical result. The result of the analysis will be discussed in the result. The Cox and Snell Psedo R-square is 0,77. This means that 7,7% of the voting behaviour is predicted by the model. This is not a perfect R-square, but considering the fact that only four variables are included it is reasonable.

Table 2 Beta's of multinomial logistic regression analysis for PvdA and VVD, 2015

	Beta		
Variables	PvdA	VVD	Other
Age 15 - 24	-0,273	-0,15	-
Age 25 -44	0,014	-0,029	-
Age 45 - 64	0,342	-0,147	-
Age 65>	0*	0*	-
Low income	-0,117	-0,866	-
Medium income	0,278	-0,505	-
High income	0*	0*	-
Very high urban	0,124	-0,412	-
high urban	-0,297	-0,38	-
medium urban	-0,211	0,301	-
lightly rural	-0,406	-0,258	-
very rural	0*	0*	-
Dutch decent	-1,165	-0,145	-
Western foreigner	-0,629	-0,058	-
Non-western foreigner	0*	0*	-

The reference category is Other
**this parameter is set to zero because it is redundant*

The probability formula

The Beta's that were the result of the multinomial logistic regression analysis can be implemented into a probability formula. This is where the CBS 'Wijken en Buurten' file is used for the first time. This datafile is a shapefile and couldn't be opened with excel straight away. The attribute table was exported from Arcmap and then opened in excel. The formula is seen in image 1. The first part of the calculation is multiplying the Beta of the analysis with the percentage that that certain category has in each neighbourhood.

The CBS 'wijken en buurten file' is edited so that all the categories are in percentages. For example the percentage of people with a medium income is 0,41 for a specific neighbourhood. This is the same for the categories income, age and country of origin. For the category of degree of urbanization the degree of urbanization is set to 100% per neighbourhood, the other degrees of urbanization is set to zero. A neighbourhood can only have one degree of urbanisation, not several like the other categories. The second part of this formula is dividing the outcome with 1+(e_pvda+e_vvd). The calculation can be seen in the appendix and are done per neighbourhood and municipality (berekeningen_buurt and berekeningen_gem).

Image 1. The probability formula for the multinomial logistic regression analysis

$$\mathbf{Pr}(Y_i = c) = \frac{e^{\boldsymbol{\beta}_c \cdot \mathbf{X}_i}}{\sum_{k=1}^{K} e^{\boldsymbol{\beta}_k \cdot \mathbf{X}_i}}$$

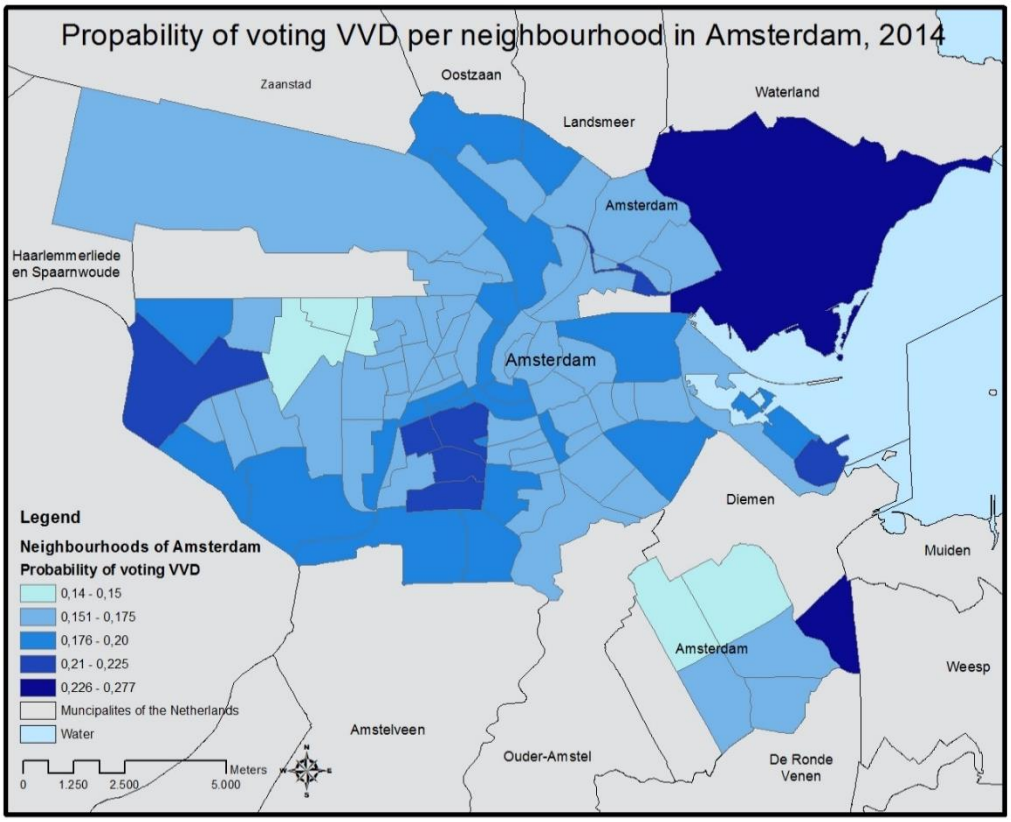
From data to a map

The GIS program Arcmap is used to go from data to maps. The probability formula is executed for every neighbourhood. For 8230 it was possible to calculate the probability score. The other neighbourhoods didn't meet the requirement of having more than a 100 inhabitants. The outcome was saved as an .xls datafile. This datafile is joined with the CBS neighbourhood file. Each neighbourhood has a specific code and this code is used to join the data together. Now the shapefile of the neighbourhoods contains the probability scores of the VVD and PvdA.

The analysis

In this section the result of the model will be shown and discussed. The choice is made to only show the neighbourhood probabilities for Amsterdam instead of the whole of the Netherlands. The reason that this decision is made is because the neighbourhood map for the whole of the Netherlands was not clear. A municipality map will be shown for the whole of the Netherlands.

The VVD map

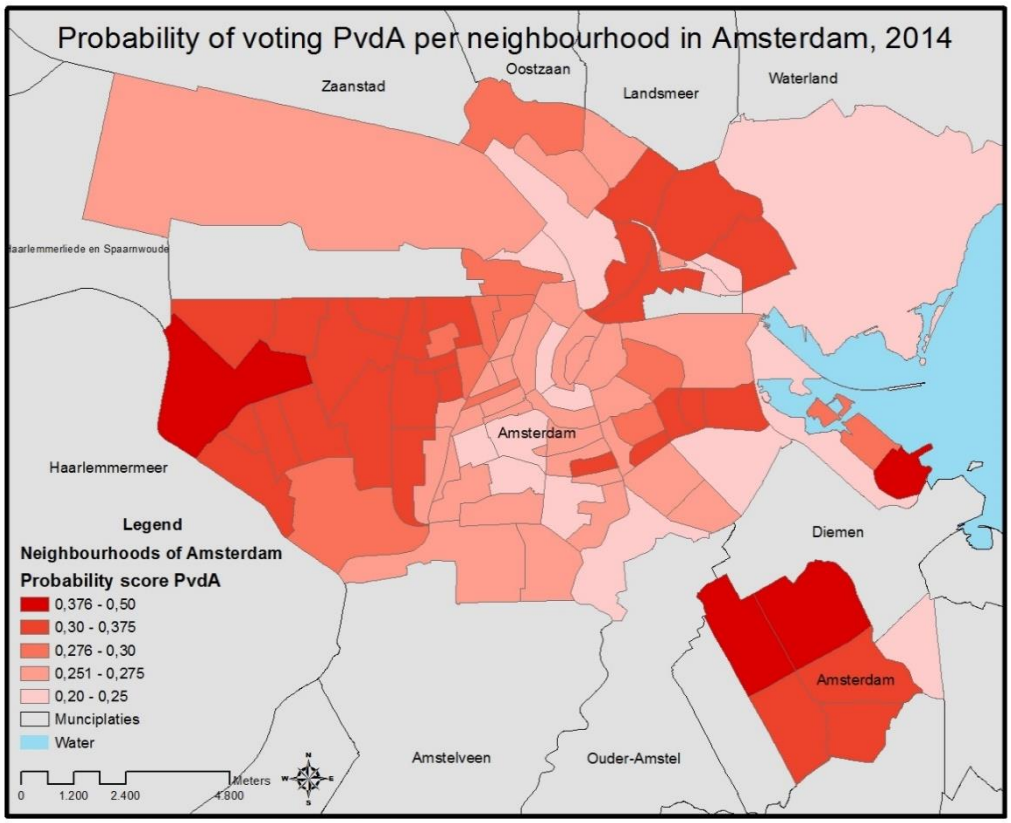


Source: SKON & CBS (2012) and CBS (2014)

This map shows the probability map of voting VVD per neighbourhood. The VVD scores very high in some of the neighbourhoods of Amsterdam South. It also scores high in two rich neighbourhoods in Amsterdam north, Waterland and Nieuwendammerdijk, and also in the neighbourhood Driemond in Amsterdam Zuidoost. This result is as expected. The VVD has lowest scores in some neighbourhoods in Amsterdam West and New-west and in Amsterdam Zuidoost. But overall VVD scores rather low in Amsterdam compared to PvdA.

When the outcomes are compared to the real outcomes of the 2012 elections, it is sometimes very accurate and sometimes not. In the neighbourhoods of Amsterdam South the VVD is indeed more popular than in other neighbourhoods in Amsterdam. The same counts for the neighbourhood of Driemond, 34% of the people voted VVD here. But for the Nieuwendammerdijk in Amsterdam North, the probability score is not very accurate. Not a lot of people voted VVD in this neighbourhood. This is a neighbourhood with a lot of rich people, but there might be other factors that have influenced the voting behaviour (OIS Amsterdam, 2012). In the other neighbourhoods of Amsterdam Zuidoost the VVD is indeed very unpopular. The same counts for the neighbourhoods of Amsterdam North. With an average support of 14,6% this is rather low.

The PvdA map

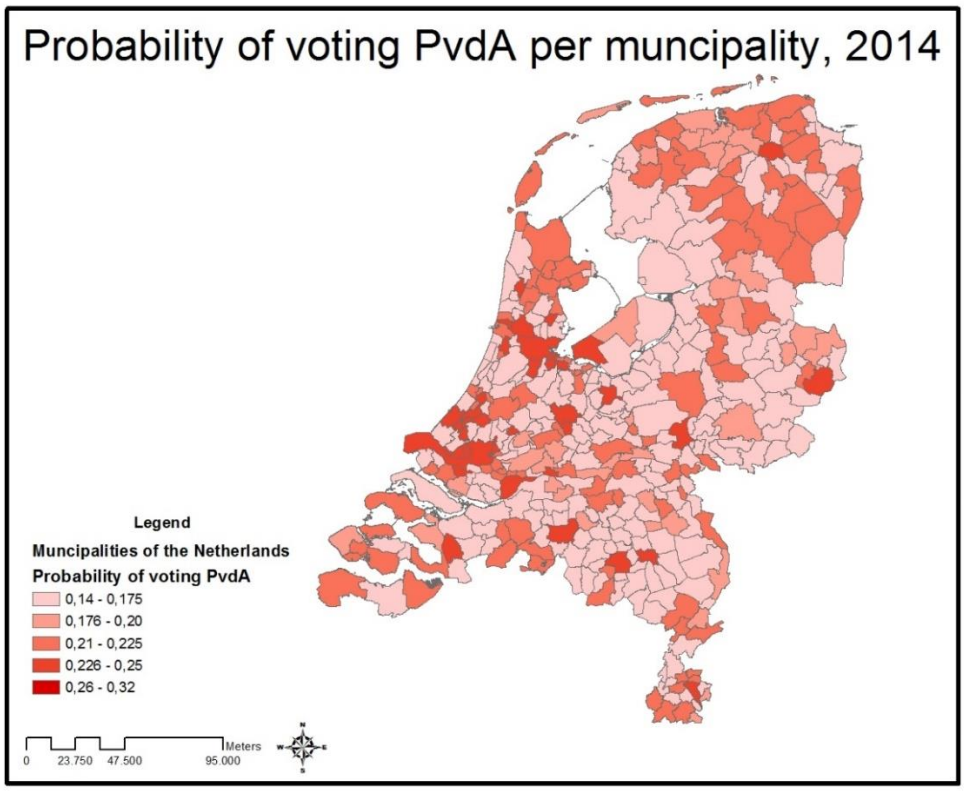


Source: SKON & CBS (2012) and CBS (2014)

This map shows the probability map of voting PvdA per neighbourhood. Overall the PvdA scores a lot higher than the VVD. This is as expected, since Amsterdam has always been a city that was a big supporter of the PvdA. The biggest support is found in Amsterdam North, Amsterdam West, Amsterdam Nieuw West and Amsterdam Zuidoost. The PvdA scores the lowest in the neighbourhoods of Amsterdam south and Amsterdam centrum. But even in these neighbourhoods the PvdA has a rather high probability score with 0,20 as the lowest score.

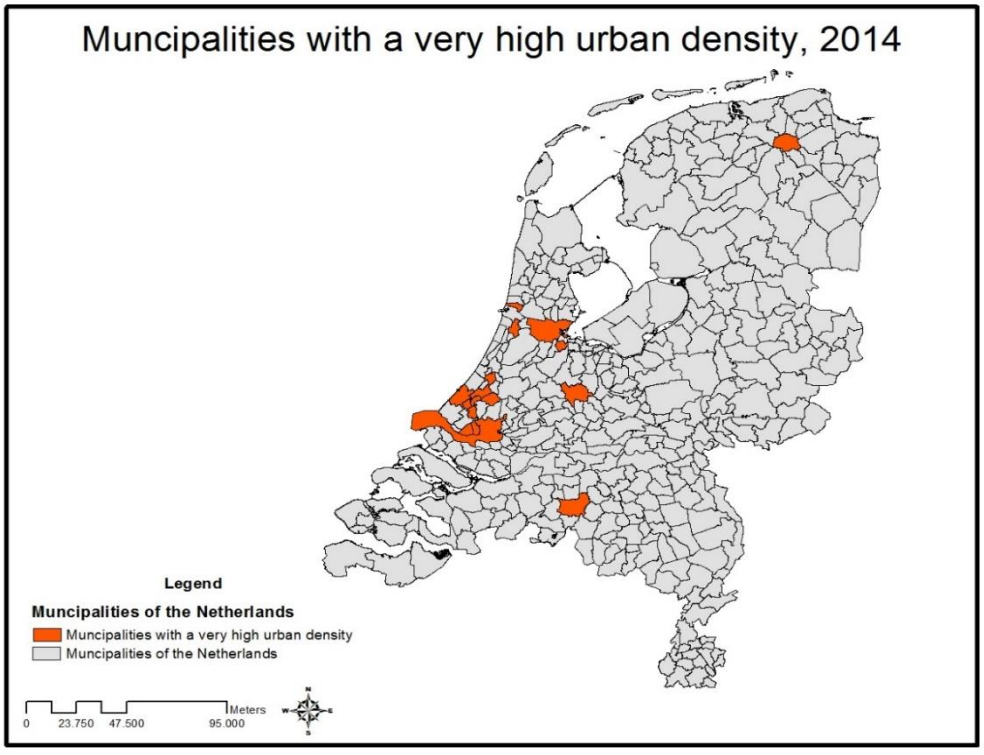
When the outcomes are compared to the real outcomes of the 2012 elections, the probability scores are rather accurate. The biggest PvdA support can be found in Amsterdam Zuidoost. The model predicts the same result. But the neighbourhoods with low PvdA support in real live don't really have a low PvdA support. Amsterdam has always been a bulwark of PvdA supporters. So some neighbourhoods with a low probability score still have a high real life PvdA support. In the Nieuwendammerdijk neighbourhood 32% of the voters voted PvdA. So even though the model predicted a low level of PvdA support, this was actually rather high. But overall the neighbourhoods with a low income and with citizens of non-western decent have the highest probability scores.

The municipality maps

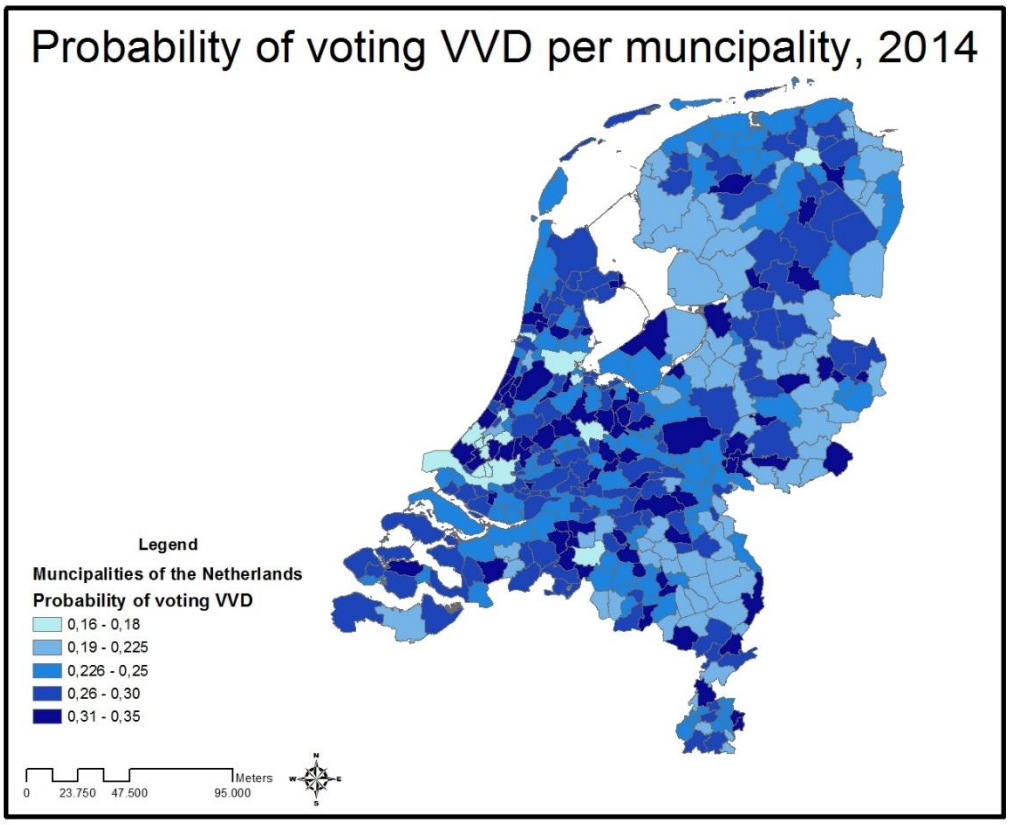


Source: SKON & CBS (2012) and CBS (2014)

The following map shows the probability of voting PvdA per municipality. As expected, Amsterdam has a very high probability score of voting PvdA. The same counts for almost all the other municipalities with a very high urban density. The PvdA scores low in the rural areas and high in the urban areas. In the map beneath the areas that are very urban can be seen. The pattern of PvdA voting an urban density can be seen when these maps are compared.



Source: SKON & CBS (2012) and CBS (2014)



Source: SKON & CBS (2012) and CBS (2014)

This map shows the probability of voting VVD per municipality. In the Netherlands the VVD is more popular than the PvdA. This is the same as the real life outcome of the 2012 elections, where the VVD was also the most popular party (tweedekamer.nl, 2017). The VVD is popular in the richer areas like the 'Gooi'. In this area almost all the municipalities are dark blue. This area is located a little bit to the south east of Amsterdam. The most northern provinces of the Netherlands in the eastern part, in the province of Groningen, are also a bulwark of the PvdA. So the probability outcomes of the map in these area are probably not too correct. But the overall probability scores are very accurate. It is interesting to see that the VVD has a higher overall probability score than the PvdA. This corresponds with the outcome of the 2012 Dutch elections, since the VVD was the winning party.

Conclusion

The goal of this rapport was to create a model that predicts voting behaviour per neighbourhood in the Netherlands. And this goal is achieved. However, how trustworthy is this model? The variables that were used to create this model were age, income, country of origin and degree of urbanization. There are many other variables that play a significant role in voting behaviour. However, these variables weren't available on a spatial level. So it wasn't possible to use some variables like education level and religion. So it must be noted that quite some variables are missing in this model.

However, the outcomes of the model are rather accurate. The variables that are included in the model apparently have the ability to predict voting behaviour quite accurate. And some variables that influence voting behaviour cannot be recorded in hard data. Why people vote on a specific party is a very personal process, so it is rather difficult to predict it by using hard location data. But this model can be altered and information can be added if it is available on the correct spatial level. The dataset that is used is also from the 2012 elections, but this has been the last elections so it's the most recent data available. So although the data is incomplete and not vey recent, the result is accurate.

But it was not possible to create a model to predict voting behaviour for all the parties of the Netherlands. As the Netherlands has a political system with a lot of parties, there are also a lot of categories in the model. If there are several independent variables and 12 dependent variables, the model has trouble with the analysis because the N is rather small in each combination of the variables. That is why the model was conducted for only two parties and the rest of the parties were left behind. The model works well with the PvdA and VVD, but the other parties are not taken into account while they are of importance. This model would be very suitable for a country that only has two parties that play a significant role. And especially in countries like the United States, where it matters where a voter lives since they have a system of electoral districts where it matters where a voter lives.

But what can this model mean for the host, Kieskompas? This model can be enhanced by using the Kieskompas data. Kieskompas has its own panel and collects their own opinion data. So instead of using the NKO2012 data that is used, Kieskompas can perform the analysis with their own opinion data. This data is more up to date and specific questions can be asked that can make the model more accurate. By asking for postal code the location of the Kieskompas panel members is known. So in the future it might be possible to create an even more accurate model to create voting behaviour per neighbourhood by combining Kieskompas data and CBS data.

To conclude, it is possible to create a model to predict voting behaviour per neighbourhood. But because of the limitations of the data it is not entirely accurate. More sociodemographic data with a geographic content and location data is needed. But this model gives a rough indication of where potential PvdA and VVD voters might live.

Literature

Agnew, J. A., Mamadouh, V., Secor, A., & Sharp, J. (Eds.). (2015). *The Wiley Blackwell companion to political geography*. John Wiley & Sons.

CBS. (2014). Toelichting wijken en buurten. Retrieved from https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/geografische%20data/wijk-en-buurtkaart-2014

CBS. (2014). Wijken en buurten kaart. Retrieved from https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/geografische%20data/wijk-en-buurtkaart-2014

De tweede kamer. (2017). Retrieved from https://www.parlement.com/id/vhnnmt7ih7yh/tweede_kamer

De Vocht, A. (2011). Basishandboek SPSS 19 IBM statistics.

Johnston, Ron, and Charles Pattie. *Putting voters in their place: Geography and elections in Great Britain*. Oxford University Press, 2006.

MacAllister, I., Johnston, R. J., Pattie, C. J., Tunstall, H., Dorling, D. F., & Rossiter, D. J. (2001). Class dealignment and the neighbourhood effect: Miller revisited. *British journal of political science*, *31*(01), 41-59.

OIS Amsterdam. (2012). Tweede kamer definitieve uitslag. Retrieved from https://www.ois.amsterdam.nl/nieuws/download/1039/2012_tweedekamer_definitieve_uitslag_1.pdf

Pattie, C., & Johnston, R. (2000). 'People who talk together vote together': An exploration of contextual effects in Great Britain. *Annals of the Association of American Geographers*, *90*(1), 41-66.

SKON. (2015). Nederlands kiezersonderzoek 2012. Retrieved from https://www.cbs.nl/nl-nl/publicatie/2015/12/nationaal-kiezersonderzoek-2006-2012

SKON., CBS., Kolk, H., Tillie, N., Erkel, P., Velden, M., Damstra, A. (2012). Dutch Parliamentary Election Study 2012. Retrieved from https://doi.org/10.17026/dans-x5h-akds

Sieben, I., & Linssen, L. (2009). Logistische regressie analyse: een handleiding. Retrieved from www. ru. nl/publish/pages/525898/logistischeregressie. pdf.

Van Gent, W. P., Jansen, E. F., & Smits, J. H. (2014). Right-wing radical populism in city and suburbs: An electoral geography of the Partij Voor de Vrijheid in the Netherlands. *Urban Studies*, *51*(9), 1775-1794.

Van der Brug, W. (2003). How the LPF fuelled discontent: Empirical tests of explanations of LPF support. *Acta Politica*, *38*(1), 89-106.

Walks, R. A. (2006). The causes of city-suburban political polarization? A Canadian case study. *Annals of the Association of American Geographers*, *96*(2), 390-414.