

# Case Study Elections Netherlands

*Ilse van Beelen, Floor Komen and Lotte Pater*

*december 14, 2018*

## Abstract

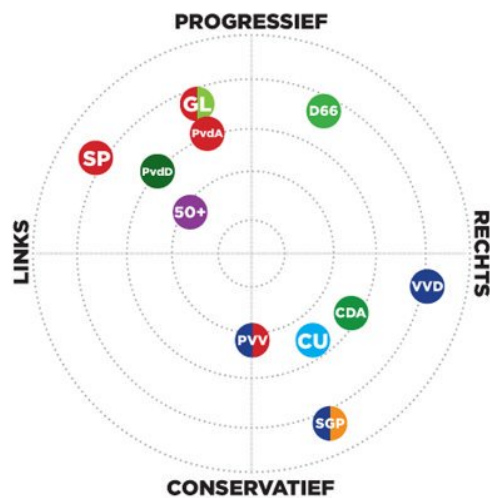
Put the abstract over here

## 1. Introduction

In this report the election

### 1.1 Summary data

In this chapter the data is summarized and explained how the data is collected. The percentage of votes per party per municipality were found on <https://data.overheid.nl/data/dataset/verkiezingsuitslag-tweede-kamer-2017>. The demographics; *amount of non-west residents per municipality, the urban index of a municipality and the standardized income per municipality*, were found on the CBS site. This is the Dutch central office of statistics.



**Dutch political parties**

In this landscape the difference between the parties is graphically displayed in this figure. In this research, two parties are chosen to investigate. These two parties had to be different, so that some comparisons could be made. The parties should not be too extreme left/right/conservative/progressive, so that the model will be proven to work on less-extreme parties. Therefore, the chosen parties are: CDA and GroenLinks. After the data-cleaning the CDA will be researched first, afterwards GroenLinks will be researched.

### Demographics

In this research the above described demographics are chosen because of their influence on a municipality level. The thought is that a more non-western municipality for example votes differently than a less non-western municipality. This is the same for the other two demographics.

Other demographics are also researched, for example gender, but on a municipality level there is no big difference between the amount of men and women per municipality. So that is a more interesting demographic to research on an individual level. *The standardized income per municipality* are given in thousands. *the urban index of a municipality* is a database with five categories per municipality. These five categories are; really strong urban (more than 2500 addresses per km2), strong urban (1500-2500 addresses per km2), moderate urban (1000- 1500 addresses per km2), little urban (500-1000 addresses per km2) and not urban (less than 500 addresses per km2). Per municipality the amount of km2 per category is given. The *non-west residents per municipality* is given in an amount per municipality, also the total amount of residents is given per municipality.

## 1.2 Data cleaning

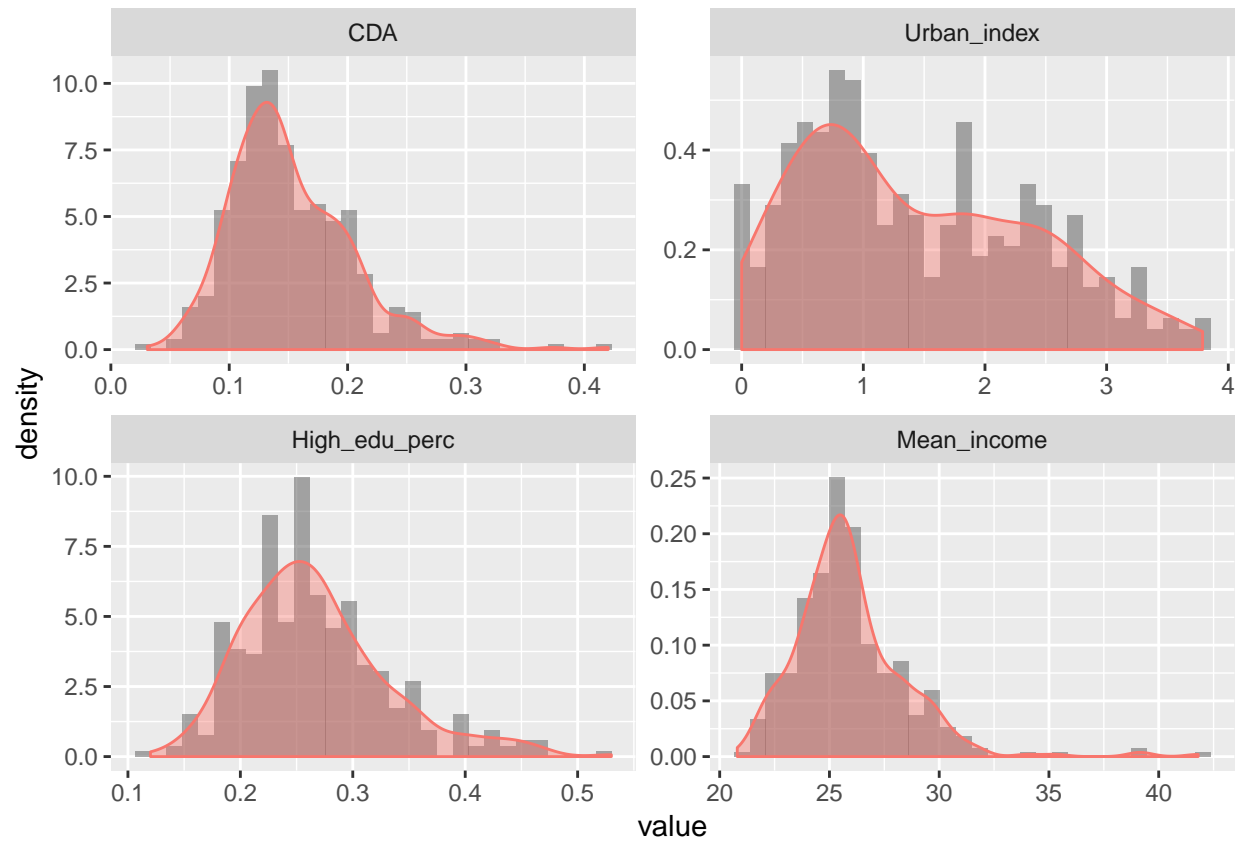
The variable *non-western residents* are divided in three groups. Municipalities with less than 5 % non-western residents, 5-10 % non-western residents and municipalities with more dan 10 % non-western residents.

```
Data <- read.csv("1_clean_data/voting_and_demographics.csv", stringsAsFactors = F,
  header = T)
Data <- Data[, -15]
colnames(Data) <- c("Muni", "VVD", "CDA", "PVV", "D66", "SP", "GL", "PvdA",
  "CU", "50PLUS", "PvdD", "SGP", "FvD", "DENK", "Urban_index", "High_edu_perc",
  "Mean_income", "Dutch_perc", "West_perc", "Non_west_perc")
Data$Non_west <- ifelse(Data$Non_west_perc < 0.05, 1, NA)
Data$Non_west <- ifelse(Data$Non_west_perc >= 0.05 & Data$Non_west_perc < 0.1,
  2, Data$Non_west)
Data$Non_west <- ifelse(Data$Non_west_perc >= 0.1, 3, Data$Non_west)
Data$Non_west <- as.factor(Data$Non_west)
Dat_cda <- Data[, c(1, 3, 15, 16, 17, 21)]
Dat_cda <- Dat_cda[complete.cases(Dat_cda), ]
```

## 1.3 Data visualisation

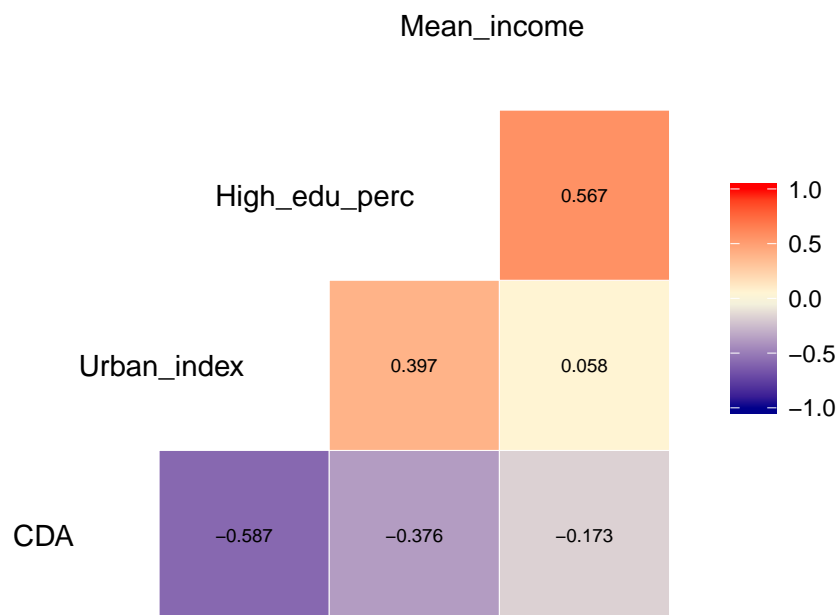
### CDA

In this part the cleaned data is visualized, so that a good picture can be obtained of the current data. First of all some demographics of data will be showed. In these histograms the density of the *CDA*, *the urban index*, *the percentage of highly educated residents* and *the mean income* are plotted.



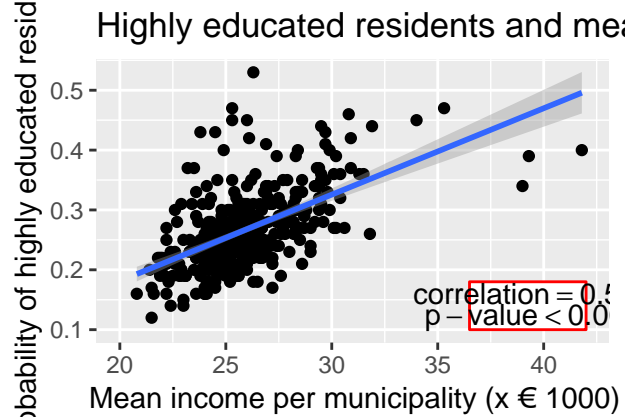
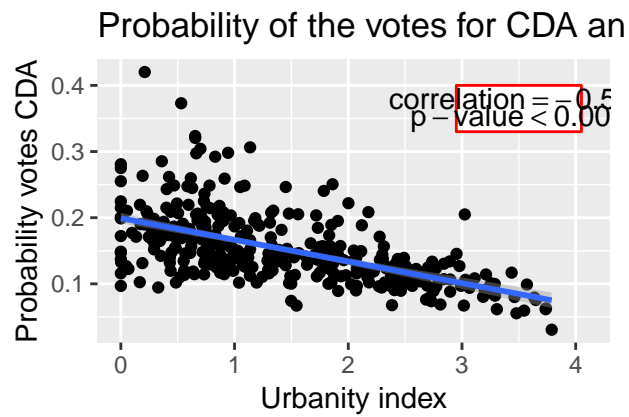
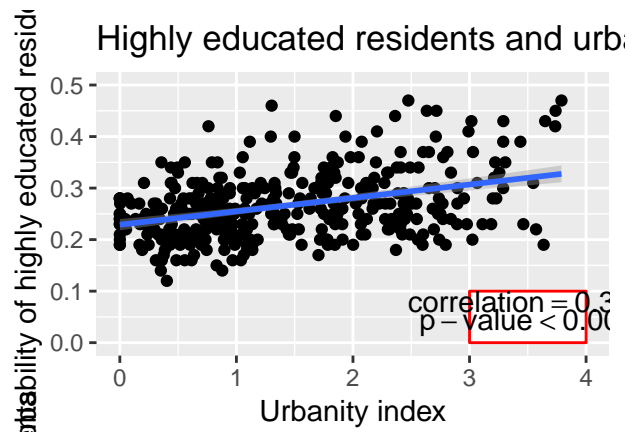
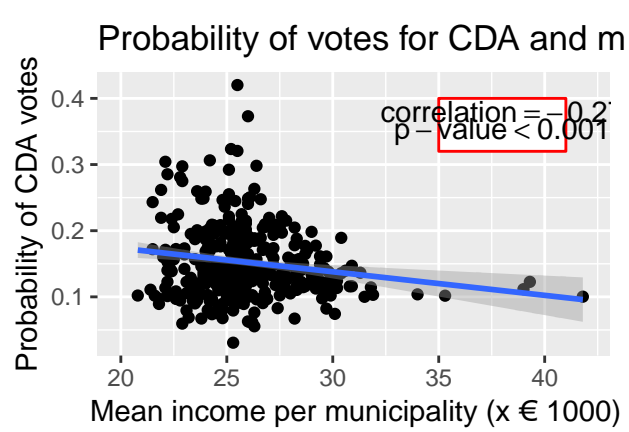
**Correlation heatmap** In this heatmap the correlation between explanatory and response variable are shown. The red color means a positive relation, the purple color means a negative relation. The relation between *mean income* and *percentage highly educated* is the highest.

## Correlation between explanatory & respons variables

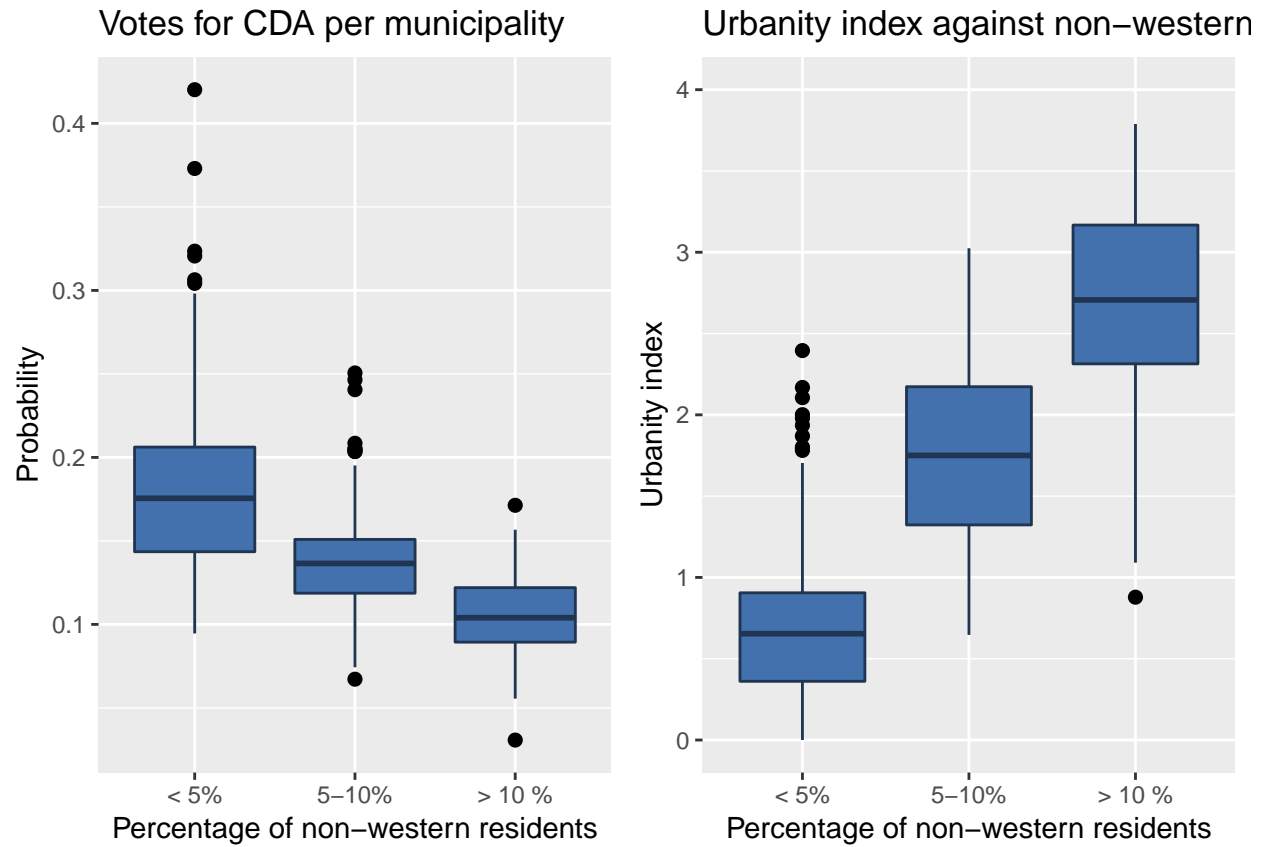


```
## [1] 0.02985029
```

**multiplot**



boxplots



Groenlinks

## 2. Formulate model

Model formuleren

**CDA**

**GroenLinks**

### **3 Final model**

**CDA**

**Groenlinks**

### **4 Analyse output**

### **5 Discussion**

#### **5.1 Limitations**