The Predictors of the Dutch Municipality Election Turnout

Universiteit Utrecht

ADS - Spatial Statistics and Machine Learning

Max van den Elsen, Sander Engelberts, Thom Venema

The Case

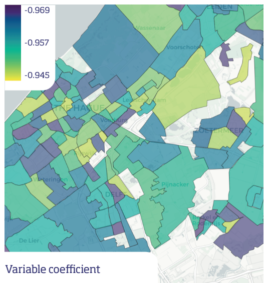
Dutch municipal council election turnout has been decreasing since the eighties, with a record low of **50.3%** earlier this year. In order to explain this trend, we investigated **socio-economic and socio-demographic determinants** of turnout prediction. Turnout research on these determinants has been done before, but not for The Netherlands.

The Data

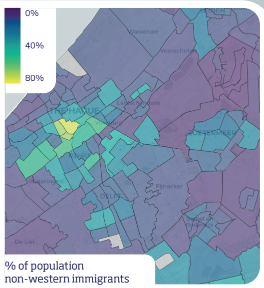
Three main datasources were used to allow for this; a dataset containing the number of people that voted at each polling station, a dataset containing the locations of these stations, and lastly and most importantly a dataset by CBS (Dutch Statistics Bureau) containing over 30 different socio- and economic variables for each neighbourhood. On this poster you will find and explaination of our methods,

GWR - Coefficients

Spatial distribution of our variables, % non-western immigrants and their coefficient values, which are significant for (almost) all neighbourhoods in The Netherlands. While the **variable** is more clustered, the coefficients of this and other predictors vary over space, with nearby neighbourhoods often containing relatively similar values.



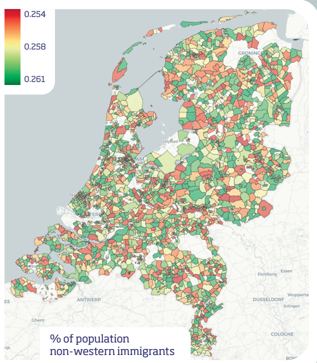
Variable coefficient



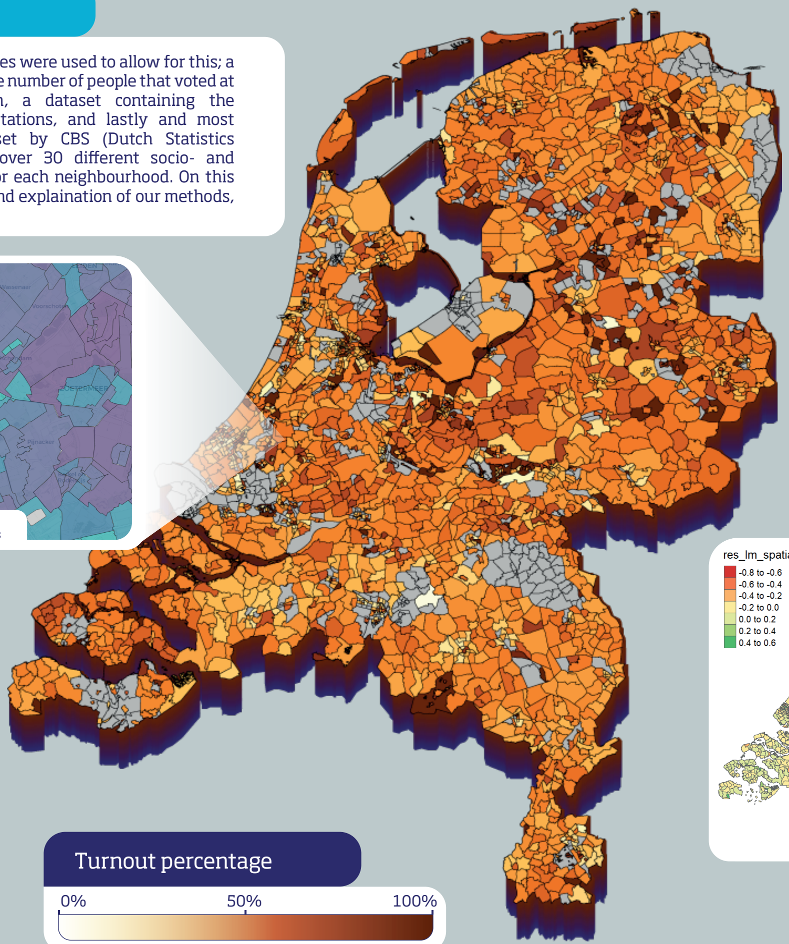
% of population non-western immigrants

GWR - Local R^2

Local explained variance (R^2) of GWR model based on all predictors. These values differ over space but are all very low (around 30%), also for the random forest models. This indicates a complex system of voter turnout on neighbourhood level, and the for further investigation at different levels and with different predictor variables.



% of population non-western immigrants



Turnout percentage

0% 50% 100%

Methods

To confirm if there indeed exists spatial correlation between the independent variables and the election turnout by estimating a simple linear regression model [21], and calculating its used Moran's I Metric. After this confirmation, a selection of dependent variables was to be made due to the presence of multicollinearity. This was done by repeatedly calculating the VIF value of all variables, dropping the one with the highest VIF.

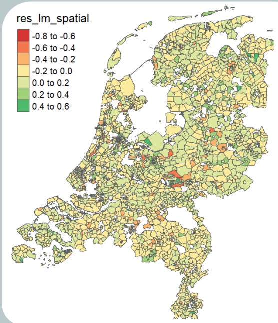
The goal was to explain how the relations are locally varying using maps. This could be done with a geographically weighted regression (GWR) model, which was fit [21, 22]. For this, an adaptive kernel was utilized with a Gaussian shape to weight the N closest polling stations based on their distance from a regression point. Further, its optimal bandwidth has been determined using cross validation [21, 22].

The significance of the local variations of relationships were checked with a t-test, and the explained variance R2 by the model was mapped [22]. This way conclusions could be drawn about the effects of the independent variables on the election turnout for different areas.

Lastly, a random forest regression was fit on the data [24]. This allowed for the inspection of feature importance based on permutation.

All these models were trained for different combinations of predictor variables. In specific these were grouped into socio-economic, socio-demographic, and spatial related predictors, and a composite of the most relevant ones.

LM - Residuals



Logistic regression residuals using an inverse distance weight matrix. These values vary over space, with some neighbourhoods having large negative or positive values. Also, a significant Monte-Carlo simulation of Moran's I p-value indicates the existence of spatial autocorrelation.Lorem ipsum dolor sit amet, cons ectetuer adipiscing CBS_turnout_neighborhood.shp

Result 4

Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue

Conclusion

Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi.
Lorem ipsum dolor sit amet, cons ectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, Lorem ipsum dolor sit amet, cons ectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam,

Variable importance

Variabele 3
Effect: **0.73**
Significantie: **0.001**

Variabele 4
Effect: **0.73**
Significantie: **0.001**

Variabele 3
Effect: **0.73**
Significantie: **0.001**

Variabele 4
Effect: **0.73**
Significantie: **0.001**

References

[21] C. Brunsdon, A. S. Fotheringham, and M. E. Charlton, "Geographically weighted regression: A method for exploring spatial nonstationarity," Geographical Analysis, vol. 28, no. 4, pp. 281-298, 1996.
[22] C. D. Lloyd, Local Models for Spatial Analysis. Boca Raton: CRC Press, 2 ed., 2010