

The Reprise of the Far Right

Investigating Spatial Heterogeneity in Far-Right Voting Behaviour in the 2025 German Federal Elections

Marion Späth	2772981
Katia Tseliou	2500434
Morgan Arima	9374752
Joris Burger	0859427

Utrecht University
Applied Data Science
Spatial Statistics and Machine Learning

Link to Github Repository:

<https://github.com/marion-spath/Spatial-Statistics-and-ML---Case-Study-2025-German-Federal-Elections>

7-4-2025

Word count: 1609



Source: <https://www.ft.com/content/8384228d-8156-4134-8eb4-035c068704b9>.

Introduction and Literature Review

Current concerns about the far-right rise in Germany -and the European Union more generally- spark growing interest in the political and socioeconomic dynamics of right-wing populism. In the recent German federal elections, the right-wing AfD (Alternative for Germany) party achieved a historic result garnering 20.8% of the vote, being Germany's second strongest political party¹. Thus, understanding far-right voting behaviour stands crucial for the preservation of future democracy.

Previous research explored and identified the regional voting patterns in the 2021 Federal Elections, concluding that AfD support is regionally uneven, with higher backing in rural and economically struggling areas, especially in eastern Germany (Götzel, 2023). Additionally, Schulte-Cloos (2021) utilized hierarchical linear and spatial error models to identify the key driving forces of voters, which are summarised among rural economic decline, population aging, historical political structures and lower immigration rates. Other research (Mellacher and Lechner, 2023) leveraged supervised machine learning to predict voter ideology by training models on expert-assigned ideological placements. Nevertheless, little focus has been placed on the incorporation of explicit spatial dependencies when leveraging machine learning techniques where the observations are deemed independent. This study fills the research gap of spatial variations of far-right voting behaviour and investigates the localised relationships between independent variables and AfD support. Thus, our research questions are: What factors influenced far-right voting behavior for the AfD in the 2025 German general elections? To what extent does the explanatory power of different factors vary across space?

To answer these questions, we use German census and administrative data and utilize spatially-explicit models (i.e., Geographically Weighted Regression and Spatial Random Forest). The findings contribute to both practical and academic understanding of the far-right's rise in the 2025 German elections. We begin by introducing the dataset and study area, then explain our model choices, present and interpret the results, and conclude with a summary.

¹ Federal Returning Officer (Die Bundeswahlleiterin), 2025. Available at: <https://www.bundeswahlleiter.de/en/bundestagswahlen/2025.html> (accessed March 18, 2025).

Data

To answer the research questions, we draw on three data sources. First, the Federal Returning Officer (Die Bundeswahlleiterin, 2025) publishes a spatial vector map of the constituencies and the official election results of 2021 and 2025. The definition of the constituencies takes into account population density and serves as the unit of analysis in this study. Second, we draw on administrative and census data published by the Statistical Offices of the Federal Government and the States. The 2022 census data is stored as point (centroid) data, based on an equally spaced 1 km raster (StaBuLa, 2025a). We aggregate the point data at the constituency level by either calculating the variable's mean (i.e., for proportional variables) or taking the sum (i.e., for count variables). The administrative data from the Regional Atlas (StaBuLa, 2025b) contains additional variables and is stored as polygon data at the county and independent city level. These data were transformed to the constituency level using weighted areal interpolation to account for the proportion of overlap between different regions.

All layers were reprojected to the appropriate EPSG:25832 for spatial analysis in Germany. Table 1 provides an overview of the specific variables by source. Selection considered both findings from previous research and current issues discussed in German public media (e.g., terrorist attacks ahead of the election; Tagesschau, 2025).

Table 1. Overview of Included Variables by Data Source

Data Source (Format)	Variables
Federal Returning Officer (csv)	<ul style="list-style-type: none">- Election results from 2021 and 2025- State name → used to create a dummy variable “east” indicating whether a constituency belongs to an Eastern or Western State
Census Data (Centroid of 1km grid)	<ul style="list-style-type: none">- Average age- Proportion of 18-29 year olds- Proportion of ≥ 65 year olds- Average rent per square meter- Average living space per person- Proportion of foreigners (defined as individuals without a German passport)- Proportion of house owners- Population count
Regional Atlas of Germany	<ul style="list-style-type: none">- Unemployment rate- Proportion of residential and transport area- Proportion of agricultural land use- Disposable income

Methods

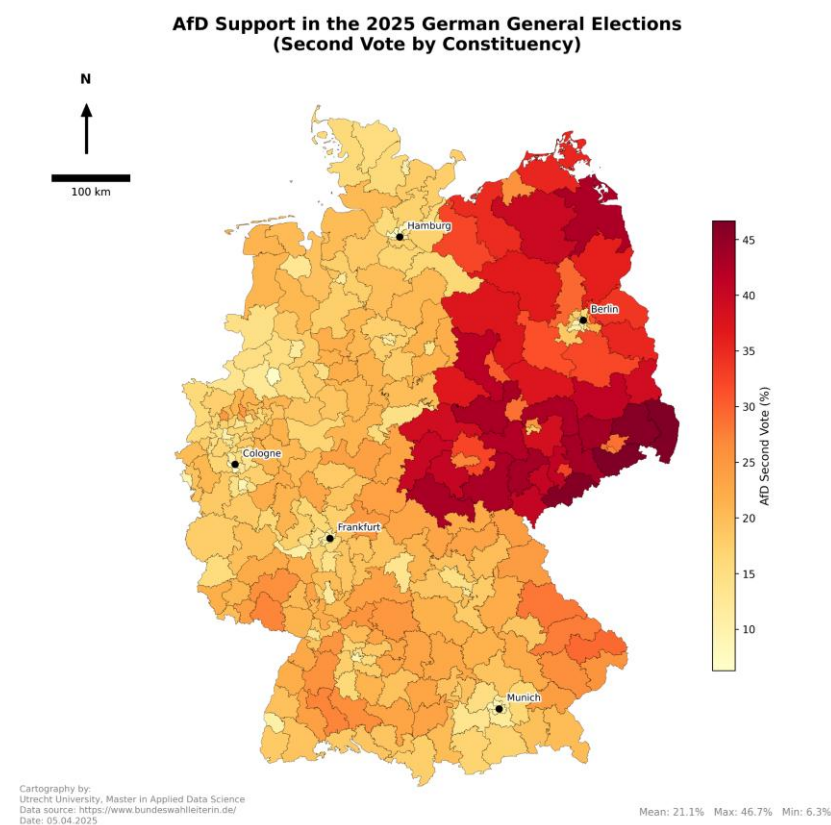
Data preprocessing

Prior to fitting any models, multicollinearity among predictor variables was assessed using Variance Inflation Factor (VIF) analysis. Features with a VIF score ≥ 10 were regarded as multicollinear variables, and were removed from subsequent analyses (see Appendix B). To ensure comparability of the coefficients, numeric predictor variables were standardized.

Models

Figure 1 maps AfD support in the 2025 elections. The visualisation and the Global Moran's I of 0.75 ($p=0.001$) indicates strong spatial clustering. Appendix A shows the spatial distribution and Global Moran's I of the independent variables. While clustering in the outcome variable itself is not yet problematic, it is suggestive of potential spatial dependence and heterogeneity. Thus, to answer the research questions, in addition to a standard model (i.e., non-spatial Random Forest (RF)) we also run other models, namely a Spatial Random Forest (SRF) and a Geographically Weighted Regression (GWR), which explicitly account for spatial effects.

Figure 1. AfD Support in the 2025 German Election.



Spatial Random Forest

SRF is a spatially-aware adaptation of RF and includes spatial lag predictors, here based on Queen contiguity. For both, the R^2 and Mean Absolute Error (MAE) are used to assess model performance. Feature importance scores were extracted for both models. Global Moran's I is calculated on the residuals of each model to evaluate to what extent the model accounted for spatial effects. Hyperparameters (i.e., `n_estimators`, `max_depth`, `max_features`) are tuned using grid-search, and cross-validation (spatially-aware for SRF) was used to find the optimal specification.

For the RF, the data was split randomly into a train and test set. For the SRF, a spatially aware strategy was used instead. For this, a dendrogram helped identify natural breakpoints of hierarchical spatial clusters. We chose 25 clusters as a balance between granularity and generalization (see Appendix C). We used a randomized cluster-based split to ensure diversity and prevent overfitting to a specific region. Thus, we randomly assigned 5 out of 25 clusters to testing while keeping the rest for training (see Appendix C).

Geographically Weighted Regression

While a SRF accounts for spatial autocorrelation, its feature importance scores are reported globally in Python and do not show spatial variations of the relationships across space. To combat this, GWR is conducted, as it allows for relationships to vary over space and a clear visualisation of the coefficients per constituency (for further explanation, see Appendix D).

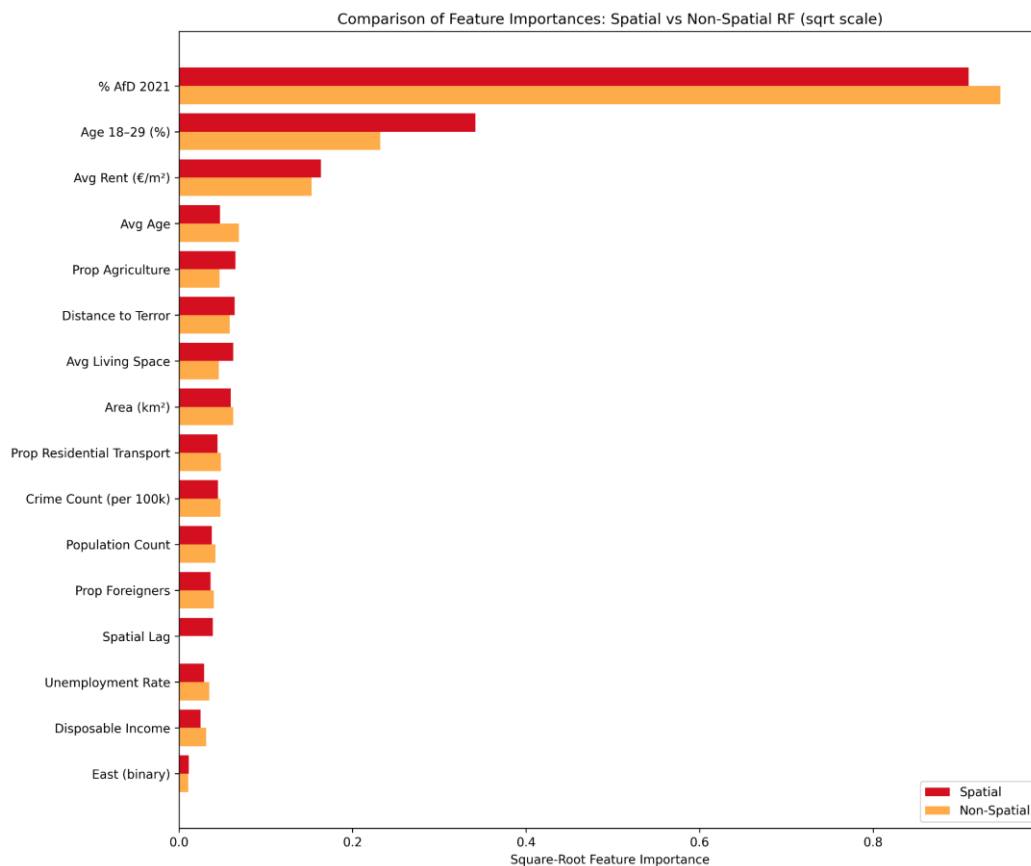
To optimize the model, a grid search was conducted to determine the best-performing hyperparameters (i.e., kernel function, bandwidth, and adaptivity of the kernel). Model selection was based on the Akaike Information Criterion (AICc) and spatial variation of coefficients, ensuring a balance between model complexity and goodness of fit. The optimal model resulted in a fixed Gaussian kernel with a bandwidth of 108,870.76 meters. For further discussion see Appendix D.

Results

(Spatial) Random Forests

The (non-spatial) RF achieves a R^2 score of 0.98 and a MAE of 0.86. The feature importance scores (log-transformed plotted in Figure 2) indicate that AfD support in the previous election in 2021 was the dominant predictor, accounting for 89.6% of the model's variance. The Global Moran's I is 0.21 with a p-value of 0.091, suggesting no statistically significant spatial autocorrelation in the residuals.

Figure 2. Log-Transformed Feature Importance Scores Spatial vs. Non-Spatial Random Forest



The R^2 of the fine-tuned SRF yields 0.96 (see Figure 3a), the MAE is 1.04 which is substantially worse than the previous RF. The feature importance plot indicates that the contribution of the spatial lag to the model is weak. The Global Moran's I for the residuals is 0.29 with a p-value of 0.002, indicating statistically significant positive spatial autocorrelation. The strength of the predictors show only little deviation from the previous RF, the strongest still being AfD support in 2021, proportion of individuals aged 18-29, and average rent per sqm. Thus, overall, this suggests that the spatial lag added noise rather than valuable explanatory power. Figure 3b visualises the spatial distribution of residuals of the RF and SRF.

Figure 3a. Visualisation of AfD Support

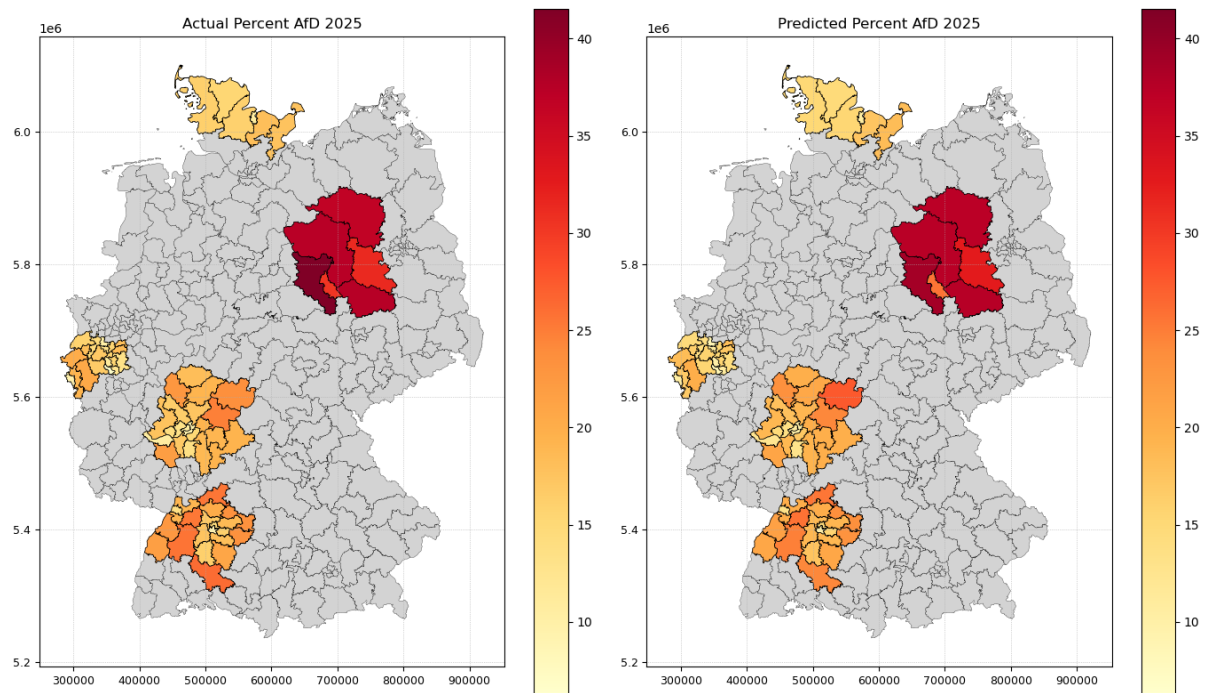
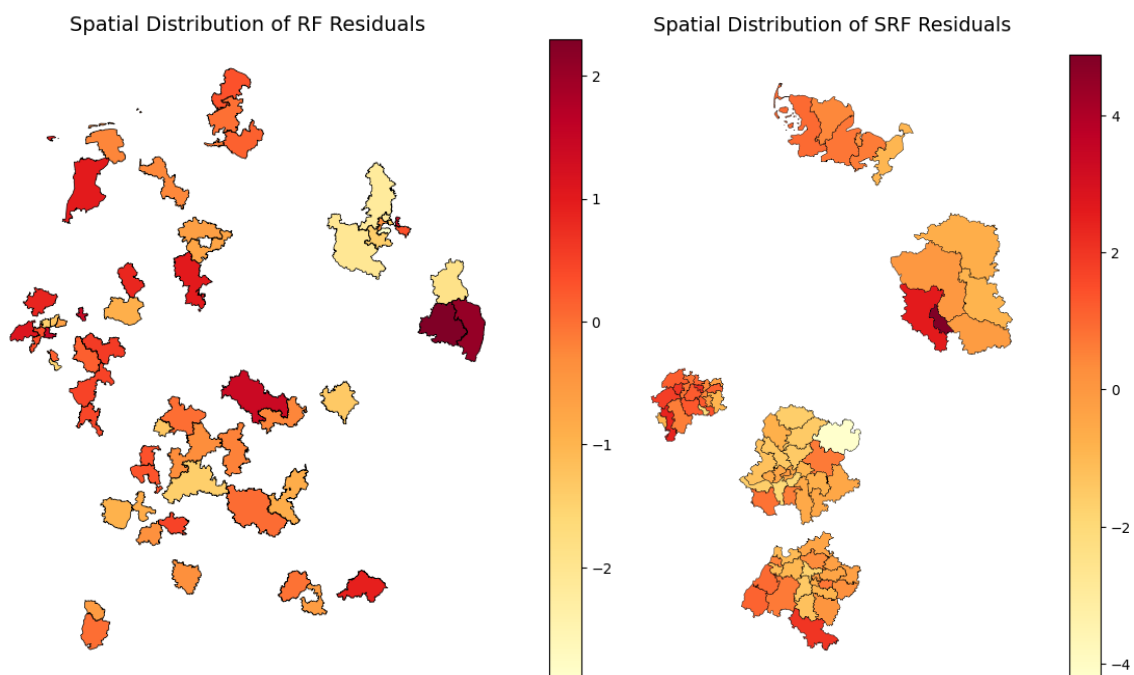


Figure 3b. Residuals of the RF and SRF



Geographically Weighted Regression

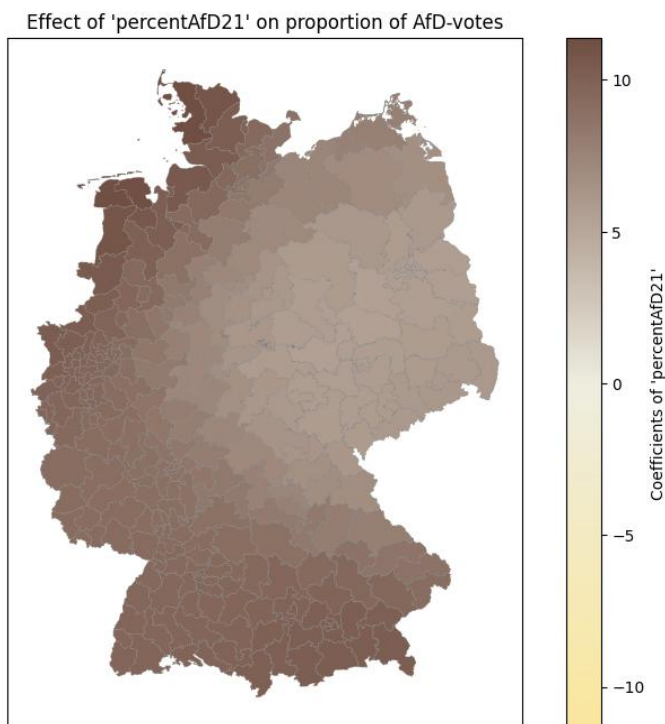
The tuned GWR achieved an AICc of 899.71 and a R^2 of 0.99. This model also accounted for the spatial autocorrelation according to a Moran's I test on the residuals (Moran's I = -0.016; p-value=0.117). Figures 4-6 map the effect coefficients per constituency of the GWR for the three strongest predictors (i.e., AfD support in 2021, proportion of 18-29 year olds, average rent per square meter) of the (S)RF (see Appendix E for the remaining predictors and a comprehensive table (E2) of the spread of the effect coefficients).

AfD support in the previous election has a moderate to strong positive effect (min=5.59, max=11.38) across all constituencies. However, values in the East appear to be consistently lower although AfD support has been highest in this area in the 2021 election which could be suggestive of a potential ceiling effect.

Although the effect of the proportion of 18-29 year olds at first glance appears to be spatially heterogeneous (i.e., overall negative and stronger in central Germany, but positive in the far East), the effect sizes are consistently very small around zero and hence practically unrelated² (mean=-0.42, min=-0.94, max=0.77).

The average effect of rent price is weak but negative (mean b=-1.09, min=-2.45, max=0.17). In Southern Germany, the coefficients are largely around zero, indicating that rent price is unrelated to AfD support. Stronger negative effect sizes in the North-East indicate that higher average rent is associated with lower AfD support.

Figure 4. Effect Coefficients of AfD Support in the Previous Election in 2021 on AfD Support in the 2025 Election per Constituency.



² A 1-standard deviation in-/decrease in the predictor is associated with a change in AfD support of less than 1 percentage point.

Figure 5. Effect Coefficients of Proportion of 18-29 Year Olds on AfD Support in the 2025 Election per Constituency.

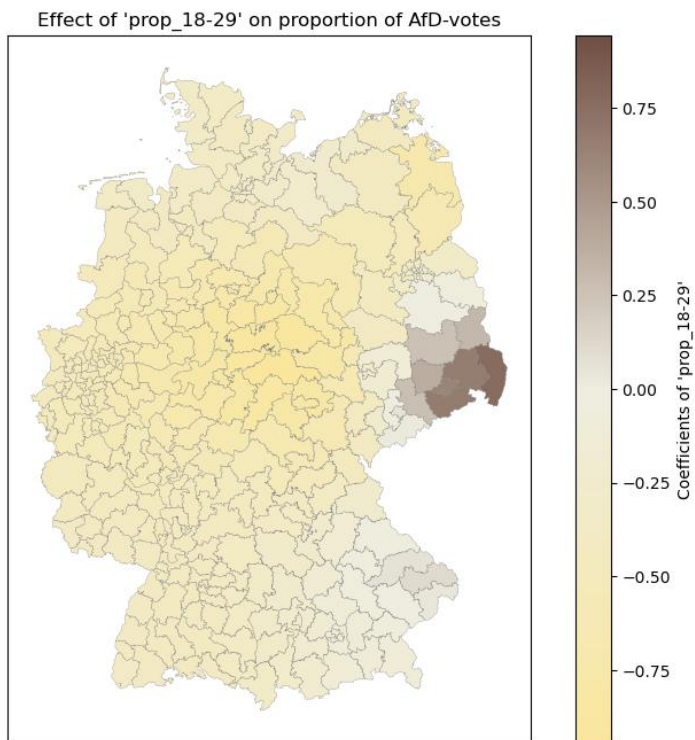
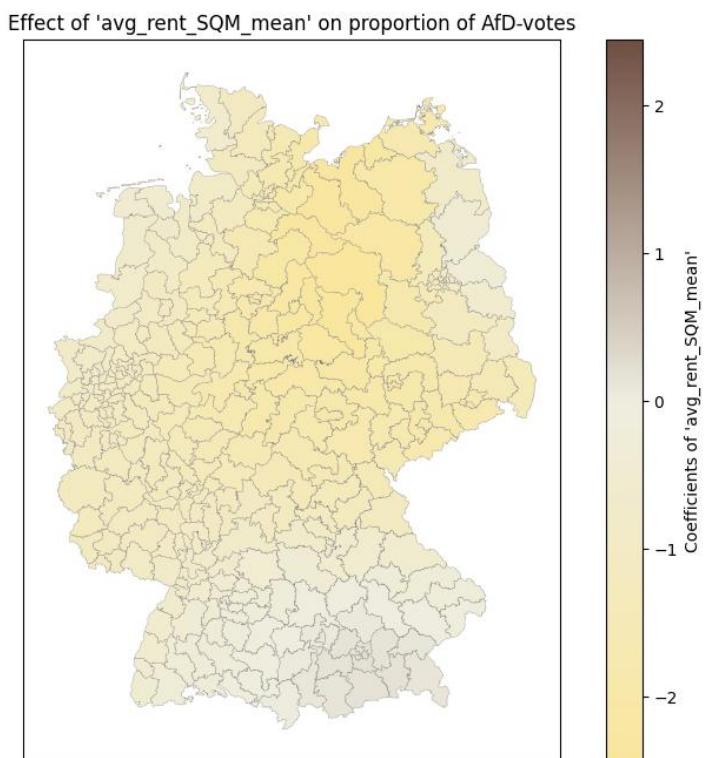


Figure 6. Effect Coefficients of Average Rent per Square Meter on AfD Support in the 2025 Election per Constituency.



Conclusion

This study aimed to investigate what factors influenced far-right voting behavior for the AfD in the 2025 German general elections with specific focus on the spatial variations of the associations between predictors and outcome. We explicitly incorporate spatial effects into our models to account for spatial dependence and heterogeneity, thereby advancing the field of electoral geography. Our main findings are:

1. Spatial effects are less prevalent than expected, in some cases adding more noise than explanatory power.
2. Prior AfD support was the strongest predictor of 2025 support, variations in coefficients suggest a potential ceiling effect as the effect is weaker for those constituencies with previously higher values.
3. Spatially aggregated administrative data may be suboptimal for studying this phenomenon. Individual-level data measuring attitudes directly might be more informative (Schulte-Cloos, 2021). For example, we hypothesised that higher rent prices might indicate higher costs of living driving political dissatisfaction. However, rent prices were often unassociated or negatively associated with AfD support, possibly because those are areas with higher-income populations less affected by inflation and less inclined toward right-wing politics.

Notwithstanding those limitations, the study generated valuable insights which can be used for future research.

References

Die Bundeswahlleiterin. (2025). *Karte der Wahlkreise zum Download*. Available at: <https://www.bundeswahlleiterin.de/> (last accessed 06.04.2025).

Götzel, S. (2023). When the East goes to the poll: Explaining the regional differences in AfD-voting in the 2021 federal election. *dngps – der moderne Staat*, 9(1). <https://doi.org/10.3224/dngps.v9i1.01>

Mellacher, L., & Lechner, L. (2023). Predicting voter ideology using machine learning. *OSF Preprints*. <https://doi.org/10.31235/osf.io/6jmga>

Schulte-Cloos, J. (2021). Political potentials, deep-seated nativism and the success of the German AfD. *Frontiers in Political Science*, 3, Article 698085. <https://doi.org/10.3389/fpos.2021.698085>

Statistische Ämter des Bundes und der Länder (StaBuLa). (2025a). Die Ergebnisse des Zensus. Available at: https://www.zensus2022.de/DE/Ergebnisse-des-Zensus/_inhalt.html#_6mryptng5 (last accessed 06.04.2025).

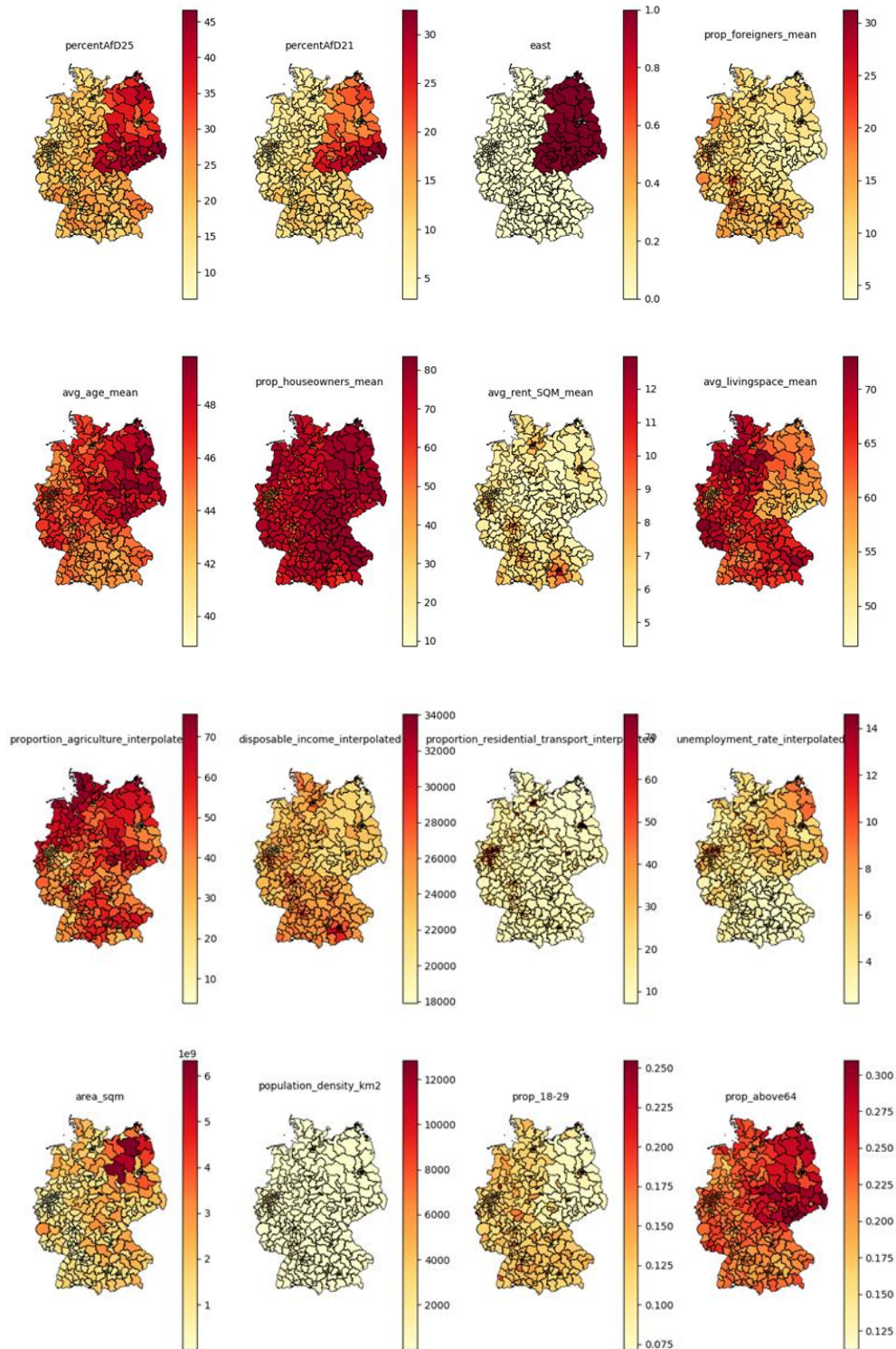
Statistische Ämter des Bundes und der Länder (StaBuLa). (2025b). *Die Ergebnisse des Zensus*. Available at: <https://regionalatlas.statistikportal.de/> (last accessed 06.04.2025).

Tagesschau. (28.01.2025). *Wie die Tat den politischen Diskurs verändert hat*. Available at: <https://www.tagesschau.de/inland/bundestagswahl/aschaffenburg-wahlkampf-100.html> (last accessed: 06.04.2025).

Appendix A:

Figure A1 shows the spatial pattern of all explanatory variables. Table A1 reports the respective Global Moran's I.

Figure A1 Spatial Distribution of All Variables



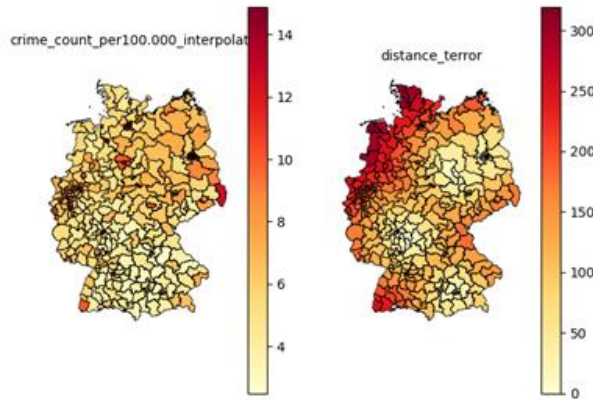


Table A1. Global Moran's I and p-value for all variables

Variable	Global Moran's I (p-value)
AfD support 2025	0.75 (0.001)
AfD support 2021	0.83 (0.001)
East	0.85 (0.001)
Proportion of Foreigners	0.57 (0.001)
Average Age	0.60 (0.001)
Proportion of Houseowners	0.57 (0.001)
Average Rent per Square Meter	0.73 (0.001)
Average Living Space	0.62 (0.001)
Proportion of Agriculture	0.57 (0.001)
Disposable Income	0.70 (0.001)
Proportion Residential and Transport Area	0.56 (0.001)
Unemployment Rate	0.66 (0.001)
Population Density	0.54 (0.001)
Proportion Age 18-29	0.3 (0.001)
Proportion Age above 64	0.59 (0.001)
Crime Count per 100.000	0.49 (0.001)
Distance to Terror Attack	0.93 (0.001)

Appendix B:

Data preprocessing - Multicollinearity

To ensure only non-multicollinear variables were in this analysis, VIF scores were calculated. Only the variables with a score lower than 10 were used in further analysis. Table B1 reports the VIF scores per variable.

Table B1. VIF scores for each variable in the dataset

	Variable	VIF
0	East	3.72
1	AfD support 2021	7.87
2	Proportion of Foreigners	3.58
3	Average Age	9.66
4	Proportion of Houseowners	14.28
5	Average Rent per Square Meter	7.95
6	Average Living Space	8.19
7	Population count	1.78
8	Proportion of Agriculture	3.27
9	Disposable Income	5.49
10	Proportion of Residential and Transport	9.38
11	Unemployment Rate	5.77
12	Crime Count per 100.000	3.54
13	Area in Square Meters	2.91
15	Distance to Terror Attack	1.63
16	Proportion of Individuals Aged 18-29	4.02
17	Proportion individuals Aged Above 64	10.79

Appendix C:

Figure C1 shows the random train/test split for the non-spatial Random Forest. Figure C2 shows the dendrogram which we based our decision for the hierarchical clusters for a spatially-aware train/test split and cross-validation on. Figure C3 shows the cluster solution (25 clusters). Figure C4 shows the spatially-aware train and test set split. 227 constituencies across 20 clusters depicted in grey belong to the training set. 72 constituencies across 5 clusters depicted in red belong to the test set.

Figure C1. Random Spatial Split of Constituencies for Non-Spatial Random Forest

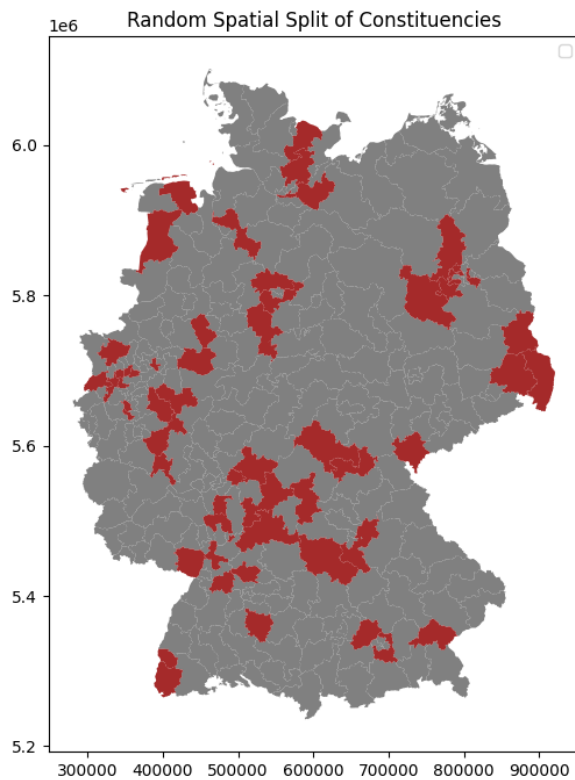


Figure C2. Dendrogram for Hierarchical Clustering.

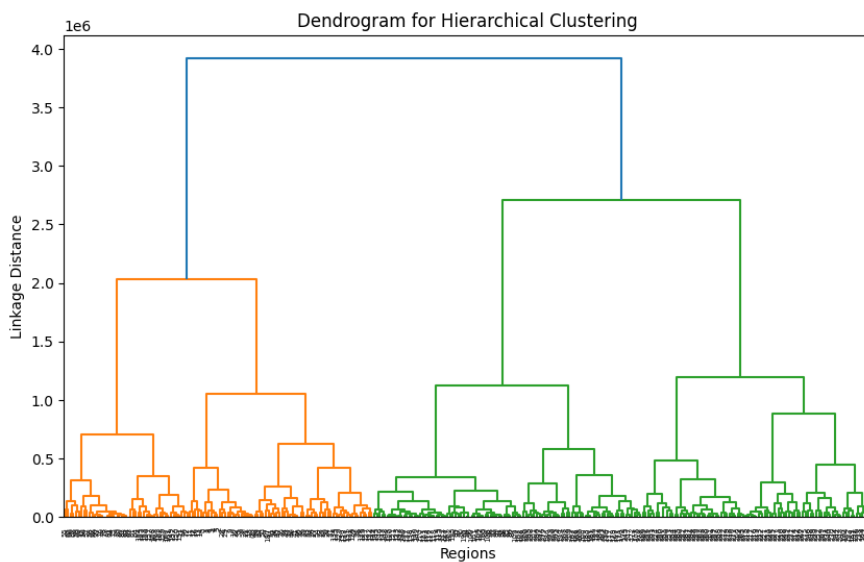


Figure C3. Map of the Clustering Solution with 25 Clusters.

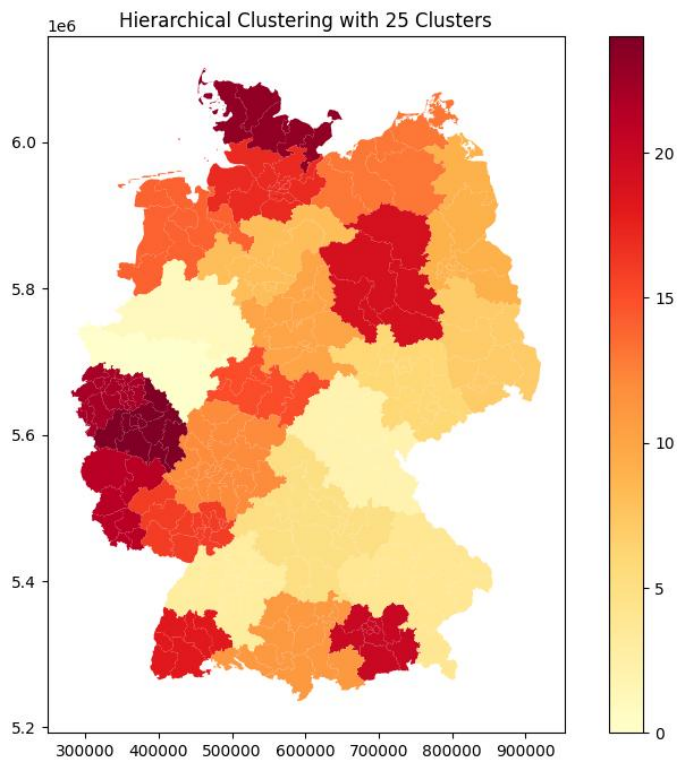


Figure C4. Train (Grey) vs Test (Red) Split for Spatially Random Forest.

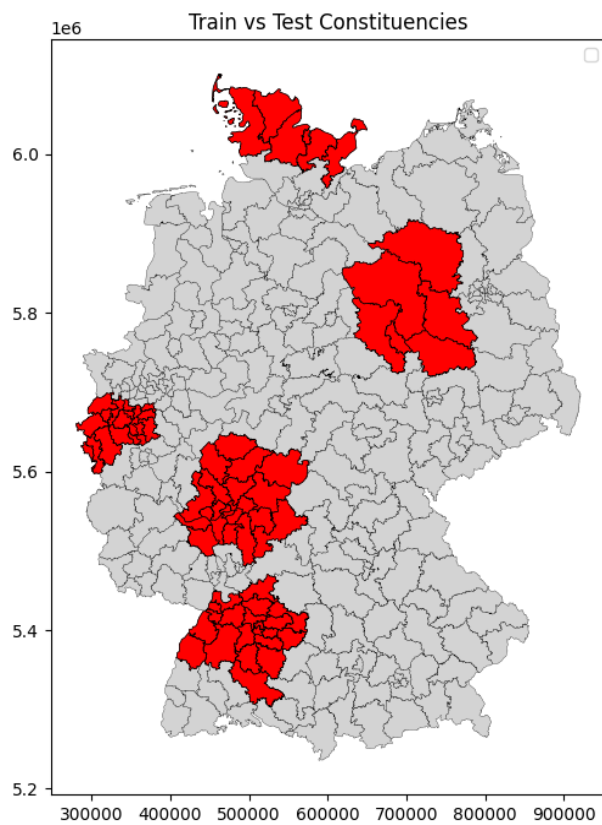
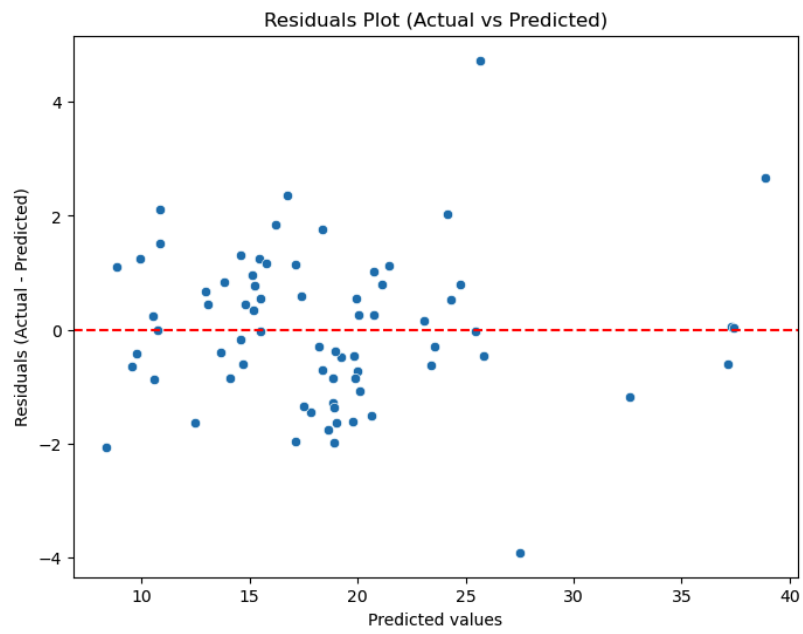


Figure C5. Plot of Actual vs Predicted Residuals for Spatially Random Forest.



Appendix D:

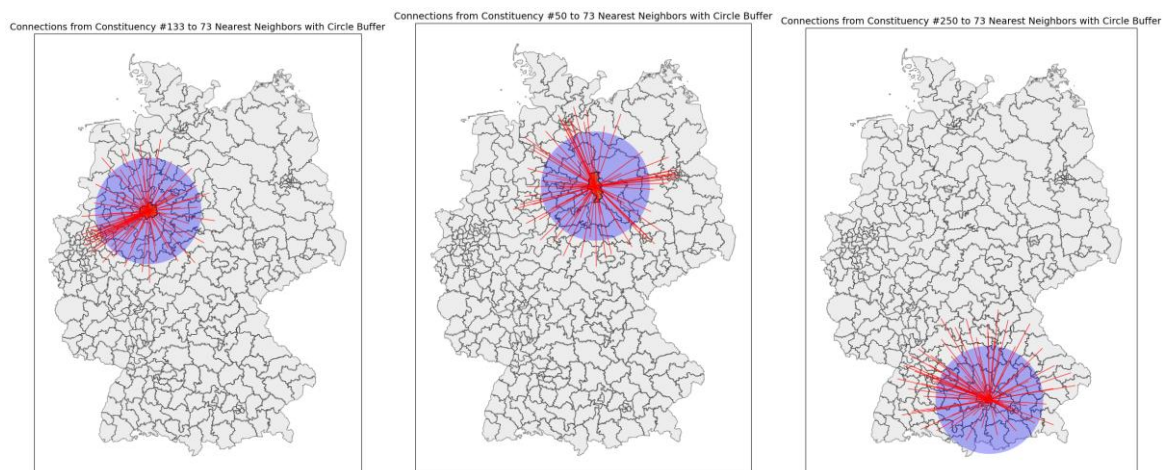
GWR operates by moving a search window from a regression point to the next through all the existing regression points in the dataset. The region within the moving window is fitted by a spatial kernel to the data. The weight of a point within the kernel decreases gradually as the distance between the point and a regression point increases. A regression model is thus calibrated locally for each individual constituency in the study area.

Table D1 shows the results of the hyperparameter tuning. Note that a bisquare kernel consistently resulted in an error message and could not be evaluated. In theory, adaptive compared to fixed kernels have the advantage of flexibly handling differences in data point densities. This advantage could even justify choosing a hyperparameter combination with a higher AICc. However, we chose the Gaussian-Fixed combination for our analysis for two reasons. First, the AICc is substantially lower than any other adaptive solution. Second, there are 299 constituencies in Germany. A flexible bandwidth of 73 practically translates to almost a quarter of the sample which leads to strong smoothing effects in the coefficients hindering meaningful interpretations. Figure D1 exemplifies this by showing both the 73 neighbors as well as the fixed buffer for selected constituencies. Taken together, this is a strong argument for choosing a fixed kernel for subsequent analysis. Nonetheless, for the interested reader we also report effect coefficients and respective maps for the Gaussian adaptive solution in appendix E.

Table D1. Results of Hyperparameter Tuning for the GWR.

Kernel	Bandwidth	Fixed vs. Adaptive	AICc / R ²
Gaussian	73	Adaptive	959.93 / 0.98
Gaussian	106670.76 meters	Fixed	899.71 / 0.99
Exponential	73	Adaptive	961.96 / 0.99
Exponential	106364.51	Fixed	924.79 / 0.99

Figure D1. Bandwidth for a Adaptive and Fixed Kernel Exemplified Using Selected Constituencies



Appendix E: Results of the GWR models

In table E1 a summary of the statistics of the GWR coefficients is given. Here we can see what effect the variables have on percent of AfD-votes.

Table E1. A Summary of the Distribution of the Effect Coefficients of the (Gaussian-Adaptive) GWR.

	Variable	Mean	STD	Min	Median	Max
0	Intercept	21.17	0.34	20.53	21.22	21.63
1	AfD support 2021	7.55	0.96	6.12	7.50	8.98
2	Area in Square Meters	0.53	0.29	-0.14	0.59	0.91
3	Unemployment Rate	0.45	0.18	0.07	0.41	0.76
4	Proportion of Foreigners	0.32	0.10	0.16	0.29	0.59
5	Proportion of Agriculture	0.29	0.13	0.08	0.26	0.55
6	Disposable Income	0.25	0.16	0.00	0.23	0.60
7	Distance to Terror Attack	0.19	0.20	-0.15	0.20	0.50
8	Average Age	0.08	0.14	-0.22	0.09	0.34
9	Population count	0.06	0.09	-0.09	0.06	0.20
10	Average Living Space	-0.10	0.33	-0.72	-0.09	0.52
11	Proportion of Residential and Transport	-0.16	0.16	-0.49	-0.13	0.07
12	Crime Count per 100.000	-0.25	0.21	-0.54	-0.27	0.22
13	Proportion of Individuals Aged 18-29	-0.49	0.10	-0.67	-0.49	-0.32
14	East	-1.17	1.51	-4.01	-0.56	0.89
15	Average Rent per Square Meter	-1.28	0.26	-1.77	-1.24	-0.72

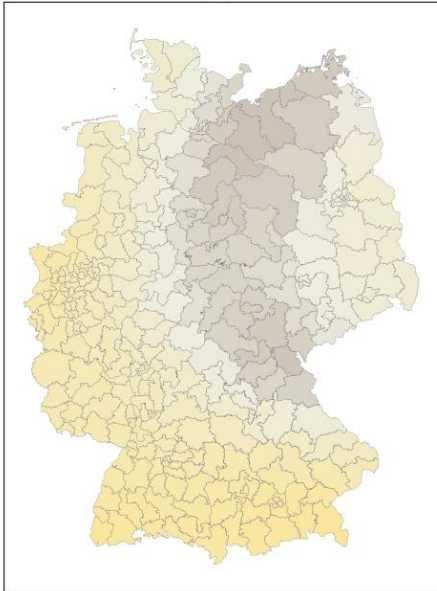
Table E2. A Summary of the Distribution of the Effect Coefficients of the (Gaussian-Fixed) GWR.

	Variable	Mean	STD	Min	Median	Max
--	-----------------	-------------	------------	------------	---------------	------------

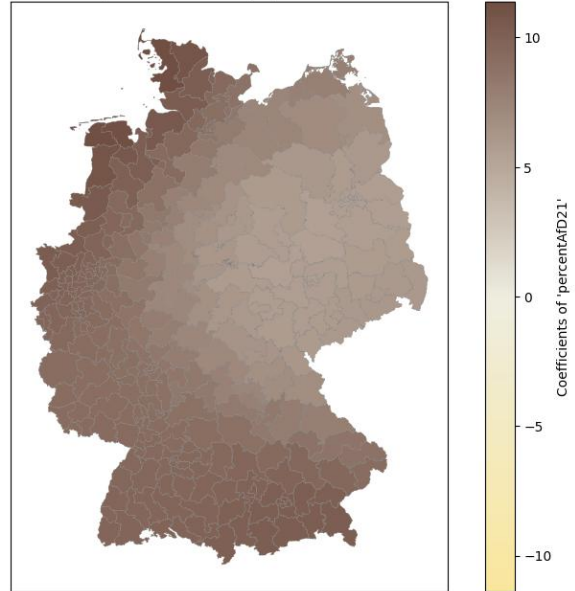
0	Intercept	21.36	0.82	19.10	21.65	23.02
1	AfD support 2021	8.44	1.58	5.59	9.08	11.38
2	Area in Square Meters	0.36	0.41	-0.38	0.42	1.17
3	Proportion of Agriculture	0.28	0.33	-0.22	0.16	1.57
4	Unemployment Rate	0.27	0.30	-0.36	0.24	1.48
5	Disposable Income	0.24	0.33	-0.31	0.09	1.54
6	Distance to Terror Attack	0.23	0.51	-1.40	0.46	0.67
7	Proportion of Foreigners	0.13	0.23	-0.60	0.17	0.92
8	Crime Count per 100.000	0.10	0.22	-0.37	0.08	0.62
9	Average Age	0.07	0.37	-0.90	0.03	1.00
10	Population count	-0.04	0.25	-0.99	-0.04	0.22
11	Average Living Space	-0.12	0.68	-2.36	0.01	1.07
12	Proportion of Residential and Transport	-0.26	0.38	-1.64	-0.20	0.36
13	Proportion of Individuals Aged 18-29	-0.42	0.25	-0.94	-0.42	0.77
14	Average Rent per Square Meter	-1.09	0.61	-2.45	-1.08	0.17
15	East	-2.82	2.83	-7.91	-3.28	2.14

Below the maps of the **fixed** GWR coefficients for each variable are shown.

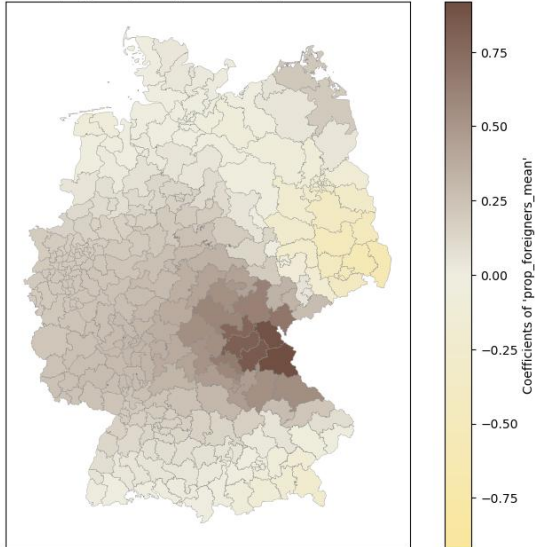
Effect of 'east' on proportion of AfD-votes



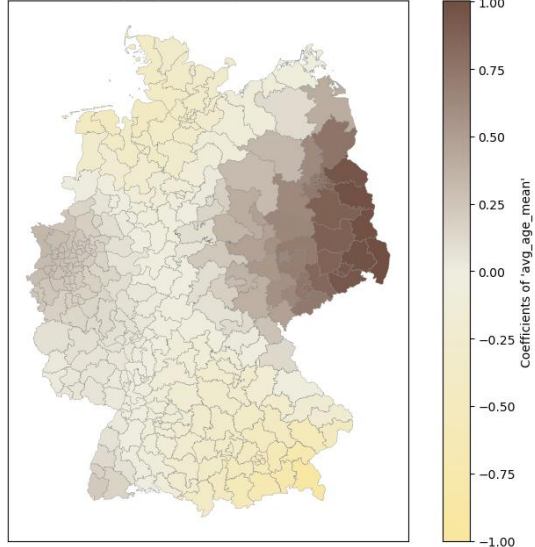
Effect of 'percentAfD21' on proportion of AfD-votes



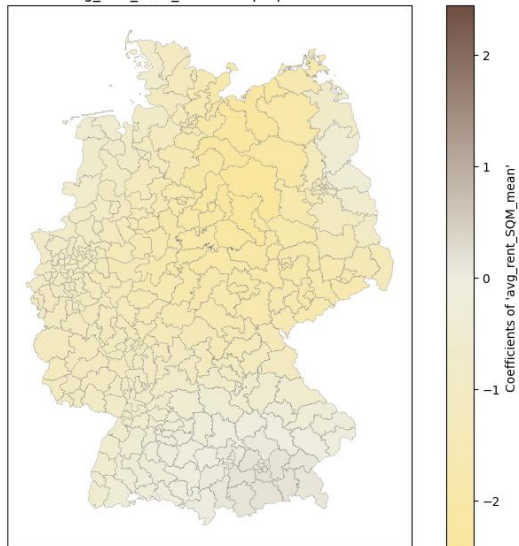
Effect of 'prop_foreigners_mean' on proportion of AfD-votes



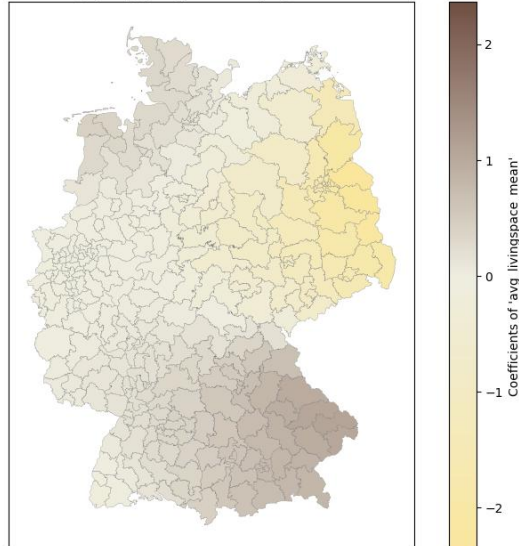
Effect of 'avg_age_mean' on proportion of AfD-votes



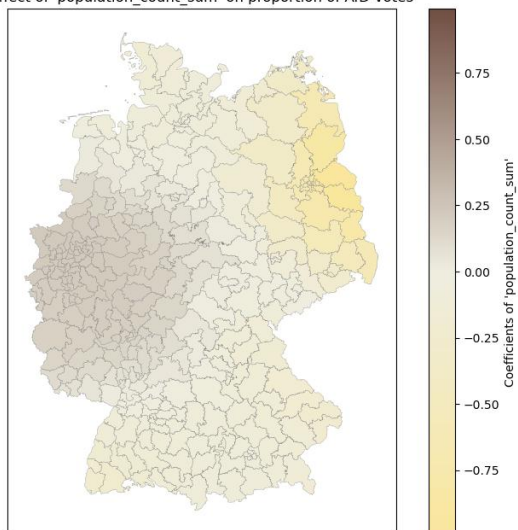
Effect of 'avg_rent_SQM_mean' on proportion of AfD-votes



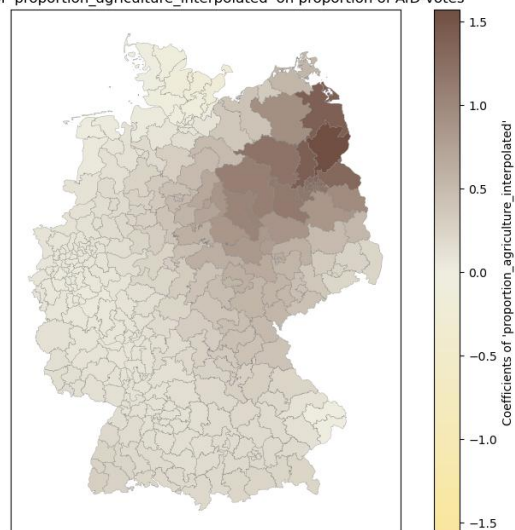
Effect of 'avg_livingspace_mean' on proportion of AfD-votes



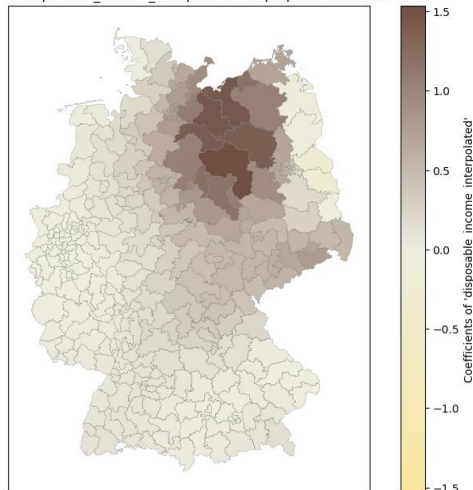
Effect of 'population_count_sum' on proportion of AfD-votes



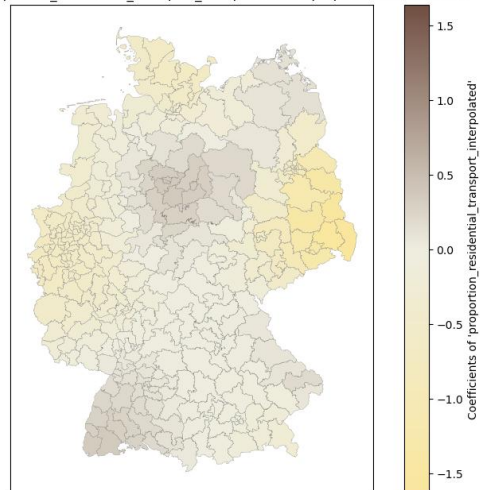
Effect of 'proportion_agriculture_interpolated' on proportion of AfD-votes



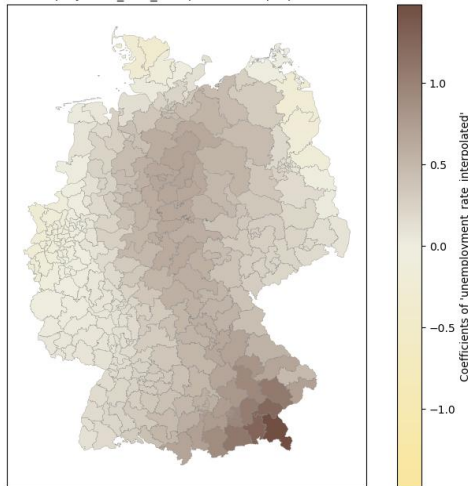
Effect of 'disposable_income_interpolated' on proportion of AfD-votes



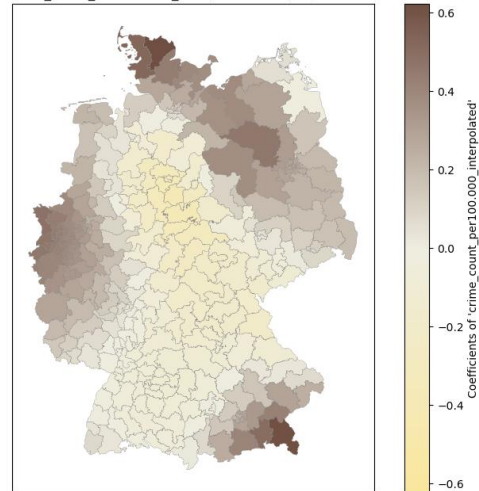
Effect of 'proportion_residential_transport_interpolated' on proportion of AfD-votes



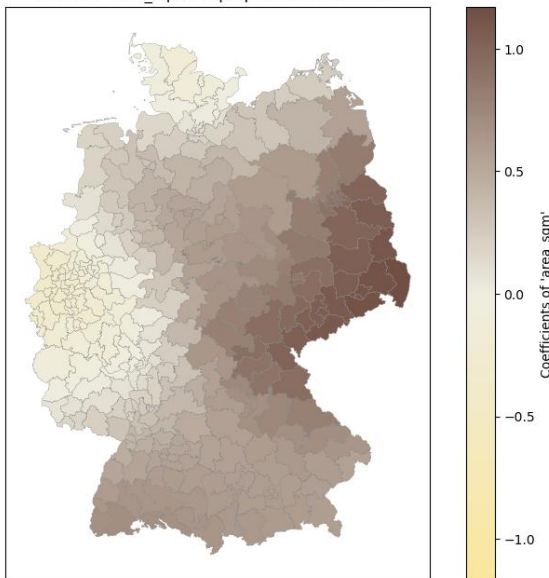
Effect of 'unemployment_rate_interpolated' on proportion of AfD-votes



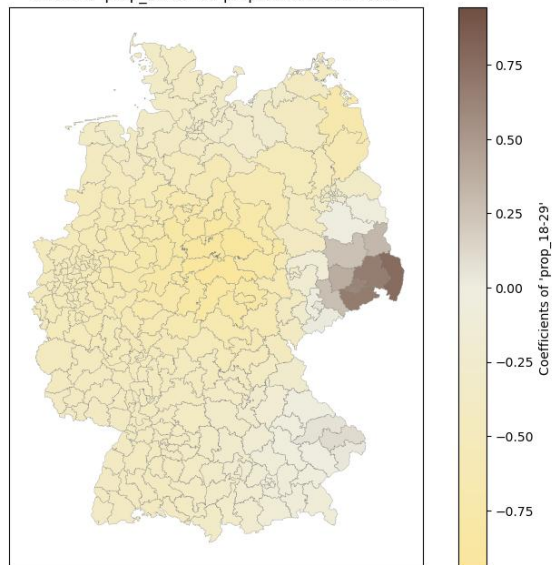
Effect of 'crime_count_per100.000_interpolated' on proportion of AfD-votes



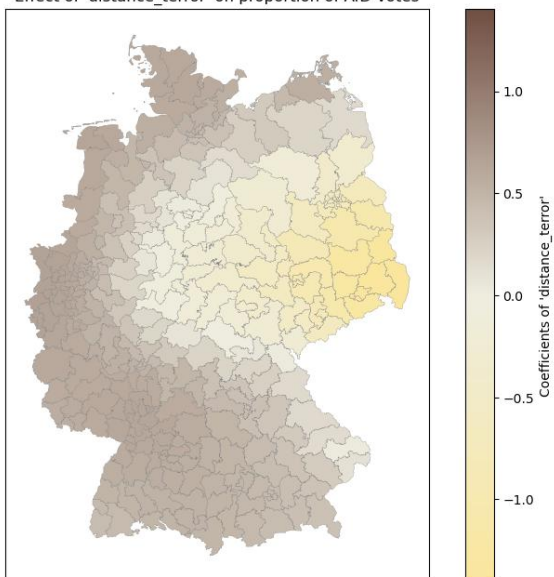
Effect of 'area_sqm' on proportion of AfD-votes



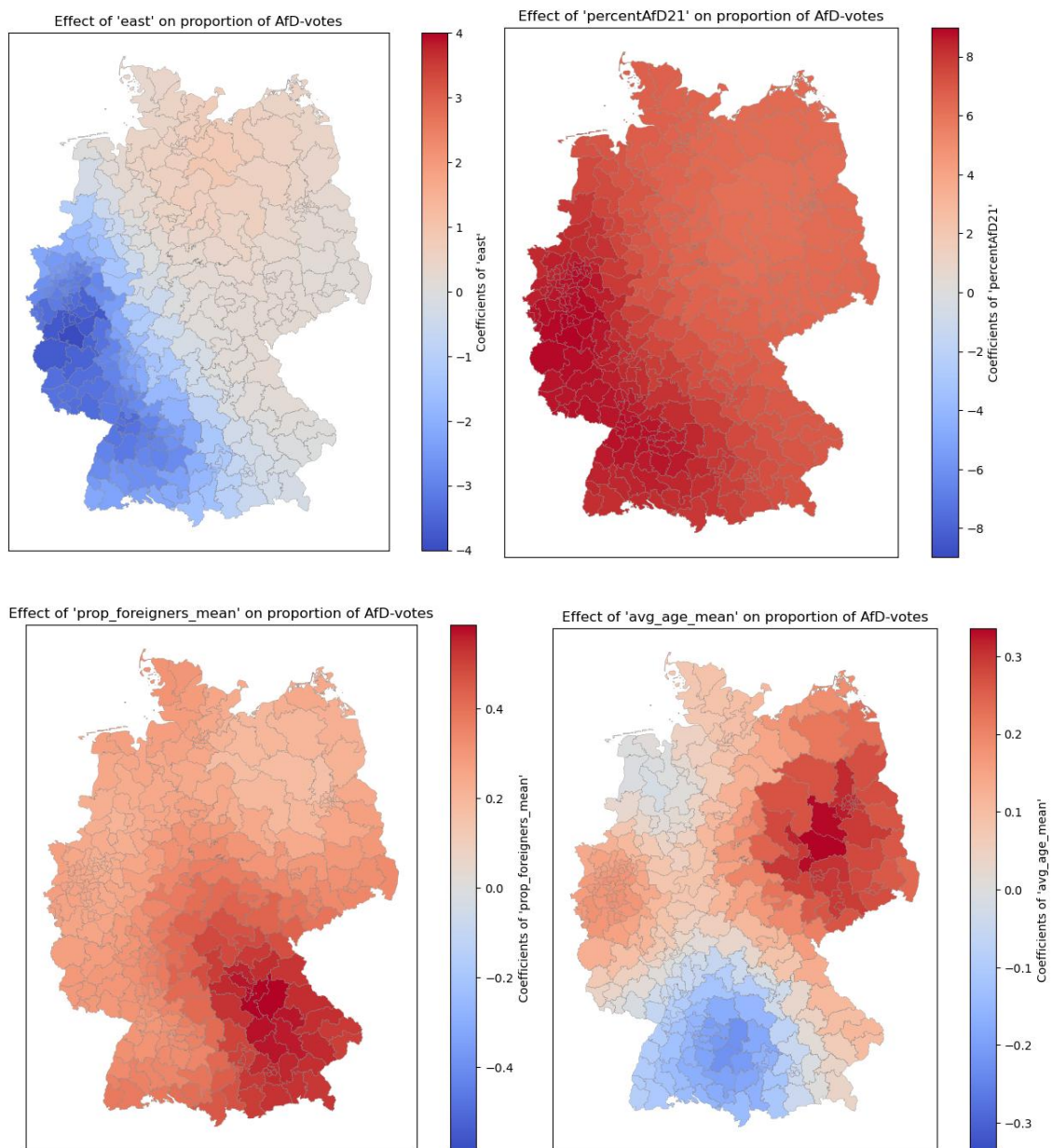
Effect of 'prop_18-29' on proportion of AfD-votes



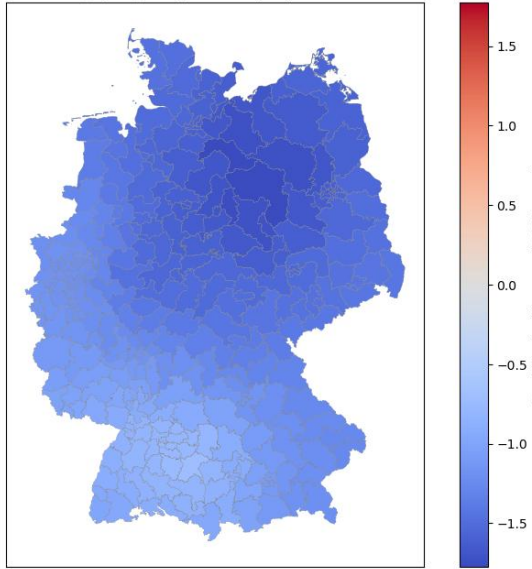
Effect of 'distance_terror' on proportion of AfD-votes



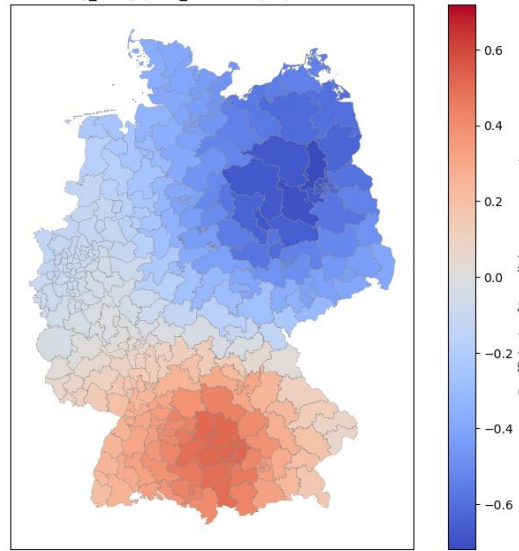
Below the maps of the **adaptive** GWR coefficients for each variable are shown.



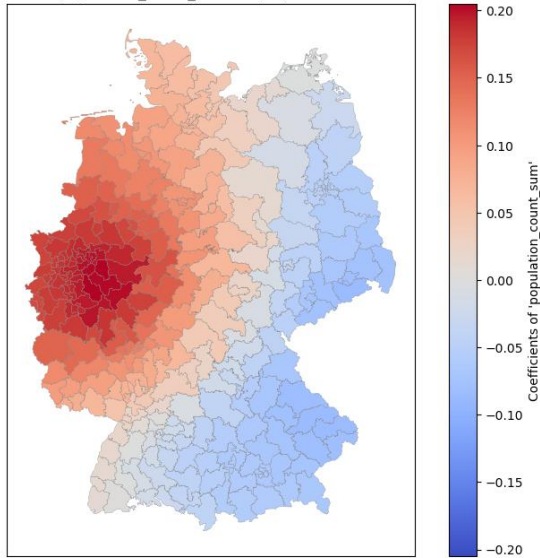
Effect of 'avg_rent_SQM_mean' on proportion of AfD-votes



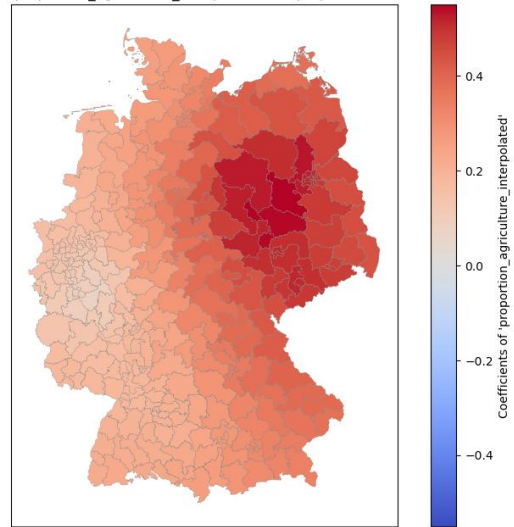
Effect of 'avg_livingspace_mean' on proportion of AfD-votes



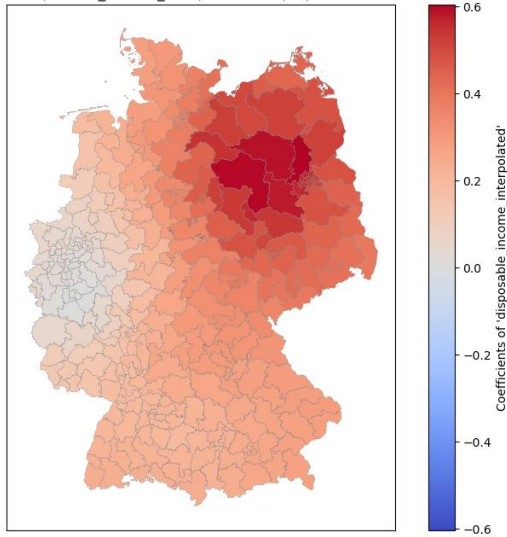
Effect of 'population_count_sum' on proportion of AfD-votes



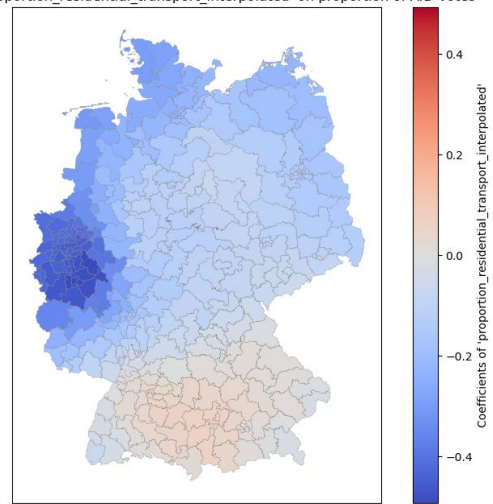
Effect of 'proportion_agriculture_interpolated' on proportion of AfD-votes



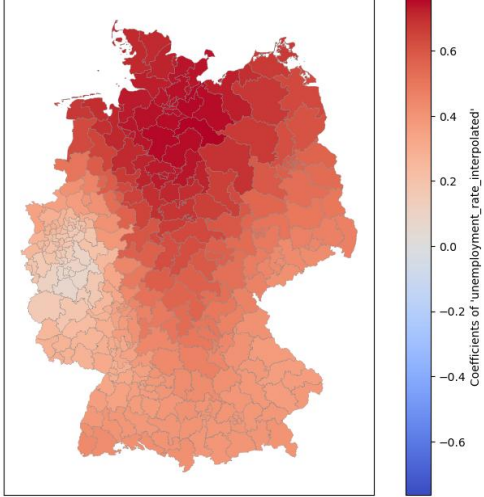
Effect of 'disposable_income_interpolated' on proportion of AfD-votes



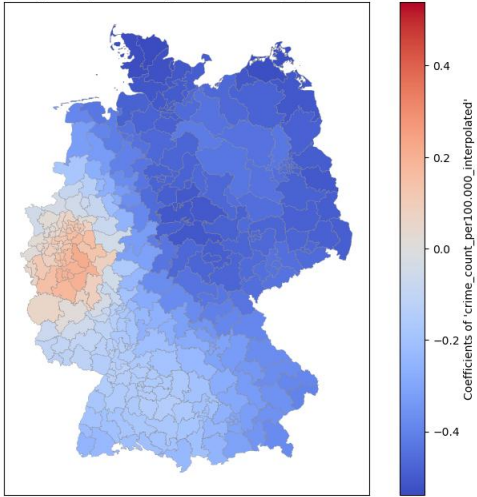
Effect of 'proportion_residential_transport_interpolated' on proportion of AfD-votes



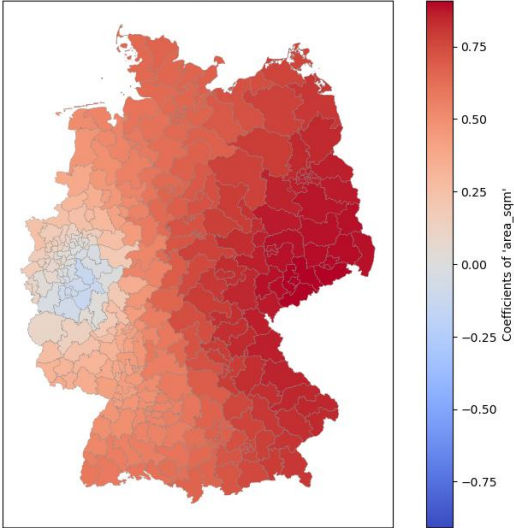
Effect of 'unemployment_rate_interpolated' on proportion of AfD-votes



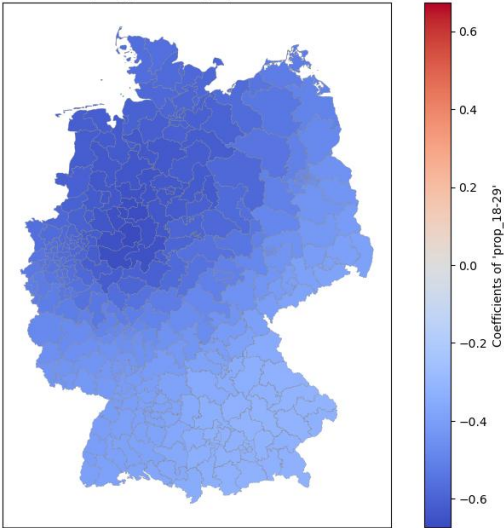
Effect of 'crime_count_per100.000_interpolated' on proportion of AfD-votes



Effect of 'area_sqm' on proportion of AfD-votes



Effect of 'prop_18-29' on proportion of AfD-votes



Effect of 'distance_terror' on proportion of AfD-votes

