

Does Urbanisation Predict Election Outcomes?

A Bayesian's Perspective

Eldaleona Odole Leonor Cunha Anarghya Murthy

Faculty of Statistics

January 26, 2025

Table of Contents

Table of Contents

Introduction

- **Research Question:** How does urbanization of a particular district affect result of an election in the US?
- **Variable of Interest:** Winning party in the House of Representatives 2022 General Election (binary)

Table of Contents

Dataset

We wanted to consider different factors in the analysis, with our primary focus being the urbanization of each House district. These factors included:

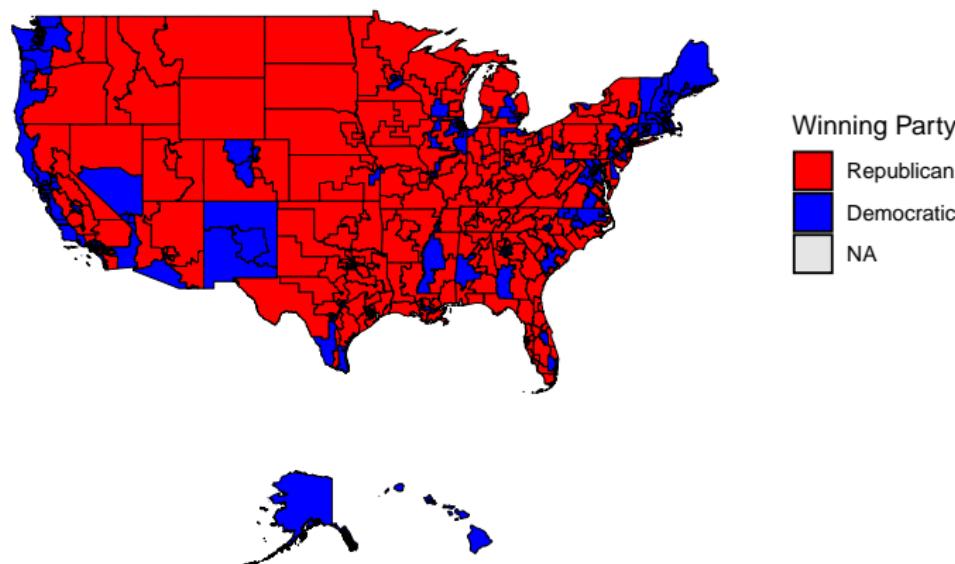
- ① Demographic Data (US Census Bureau)
- ② Urbanization (FiveThirtyEight)
- ③ Regional Information (US Census Bureau)
- ④ Election Results (FiveThirtyEight)

We combined different sources in order to create our data set containing 435 instances of 16 unique covariates.

Winning Party

Our independent variable is Winning party in the 2022 Election.

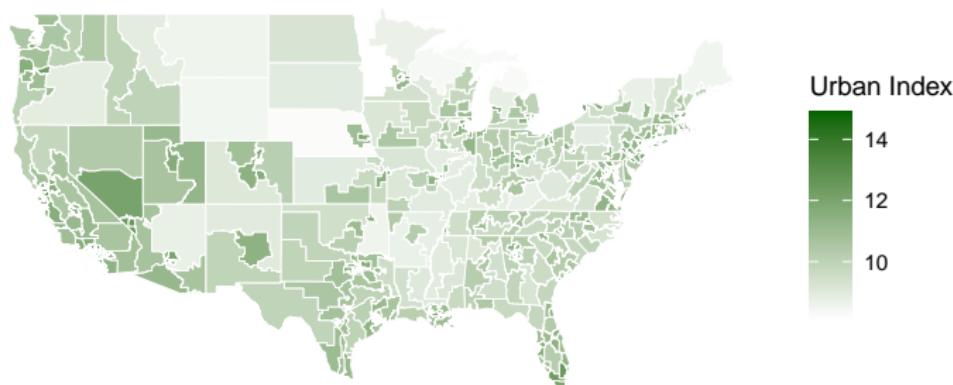
Winning Party by Congressional District



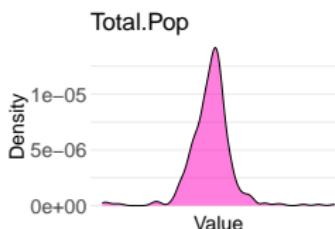
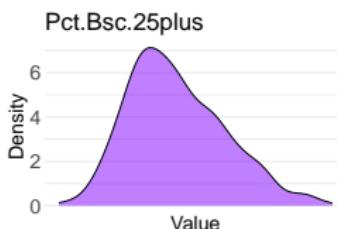
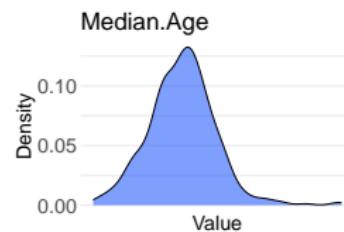
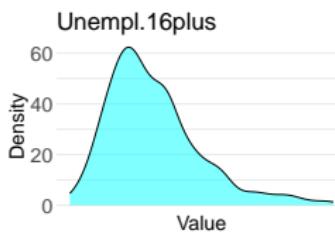
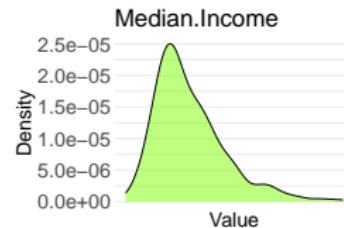
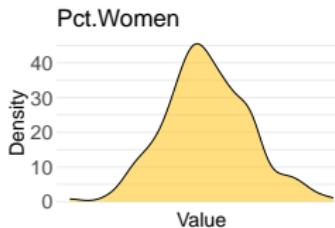
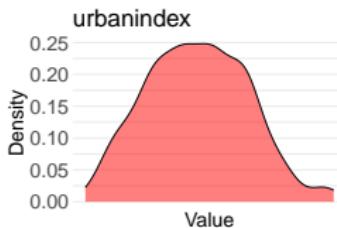
Urban Index

Our dependent variable of interest is the Urban Index from FiveThirtyEight.

Urban Index by Congressional District



Densities



Correlation Matrix

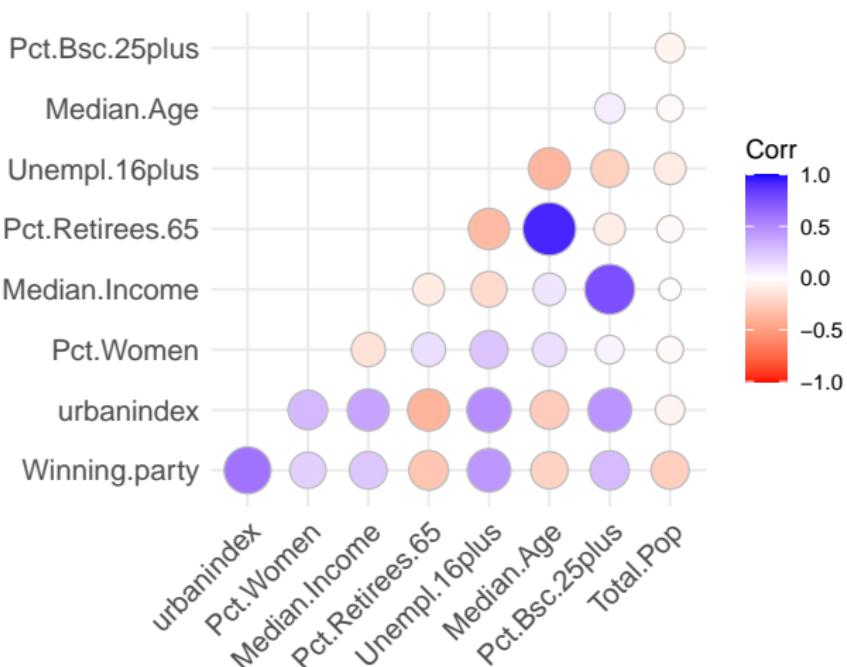


Table of Contents

Model Assumptions

There are many people trying to predict US election outcomes, from the wealth of data available about voters. However we wanted to look at the voters in relation to their geography. In order to do this we assumed

- District voting outcomes can be modeled via logistic regression
- Districts are exchangeable within each state and each state is exchangeable within its region
- We assume that the variation is explained by our predictors, and were there no additional predictors $\theta = .5$ across all levels of geography

Table of Contents

Model 1

Let the response variable 'Winning Party' be y , the predictor of interest 'Urban index' be x , and the other covariates be a 15-dimensional vector z . Let i , j , and k be the indices for the district, region, and state respectively.

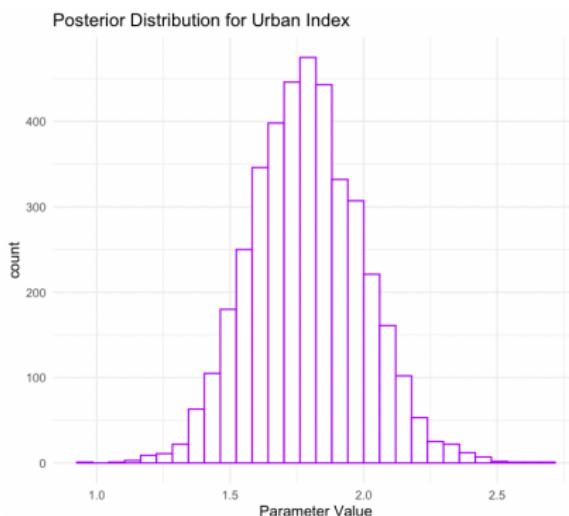
$$y_{i,j} \sim Ber.(logit^{-1}(\theta_j))$$

$$\theta_j := \beta_0, j + x_{i,j} * \beta_{1,j} + z_{i,j}^T * \gamma_{1,j}$$

$$\beta_{1,j} \sim Gam.(1, \tau)$$

$$\tau \sim Normal(0, 1)$$

Model 1: Results



Regression Coefficients:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS	
Intercept	-18.00	2.52	-22.94	-13.01	1.00	2713	2434	
urbanindex	1.79	0.21	1.40	2.20	1.00	3427	2864	
Percentage.Women	0.08	1.01	-1.94	2.08	1.00	7100	2504	
pct.bach	0.35	1.67	-2.87	4.51	1.00	3277	1433	
Median.Household.Income	-1.37	0.76	-2.90	0.06	1.00	4236	3037	universität
pct.retirees	-3.84	4.41	-15.57	0.97	1.00	2060	1858	

Model 2

Let the response variable 'Winning Party' be y , predicted by the variables Urban Index and the Percentage of Retirees. Let i , j , and k represent the indices for district, state, and region, respectively.

$$y_{i,j,k} \sim \text{Ber.}(\text{logit}^{-1}(\theta_{j,k}))$$

$$\begin{aligned}\theta_{j,k} = & \beta_0 + \beta_1 \cdot \text{Urban_Index} + \beta_{1,j} \cdot \text{Urban_Index} \\ & + \beta_{1,j:k} \cdot \text{Urban_Index} + \beta_2 \cdot \text{Pct_Retirees}\end{aligned}$$

$$\beta_0 \sim \text{Normal}(0, 0.5)$$

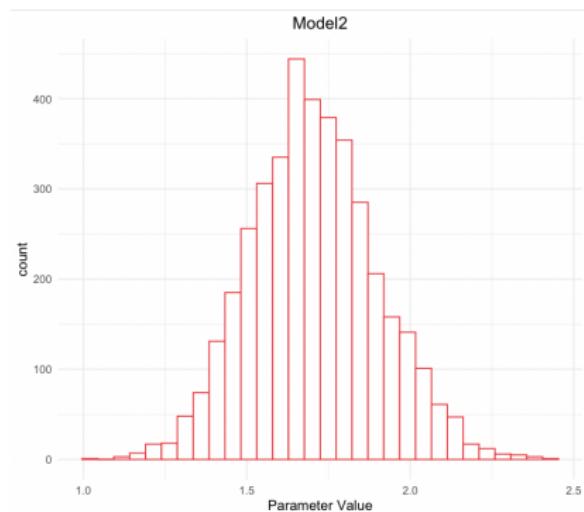
$$\beta_1 \sim \text{Normal}(0, 1)$$

$$\beta_{1,j} \sim \text{Normal}(0, \sigma_j)$$

$$\beta_{1,j:k} \sim \text{Normal}(0, \sigma_{j:k})$$

$$\sigma_j; \sigma_{j:k} \sim \text{Halfcauchy}(10)$$

Model 2: Results



Regression Coefficients:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS	universität
Intercept	-17.98	2.37	-22.70	-13.36	1.00	3212	3406	
urbanindex	1.71	0.19	1.34	2.10	1.00	4960	3321	
Percentage.Retirees..65..	-4.55	4.18	-16.44	-0.29	1.00	2240	1966	

Model 3

Let the response variable 'Winning Party' be y , predicted by the variables Urban Index and the Percentage of Retirees. Let i and j represent the indices for district and state respectively.

$$y_{i,j} \sim \text{Ber.}(\text{logit}^{-1}(\theta_j))$$

$$\theta_j = \beta_0 + \beta_1 \cdot \text{Urban_Index} + \beta_{1,j} \cdot \text{Urban_Index} + \beta_2 \cdot \text{Pct_Retirees}$$

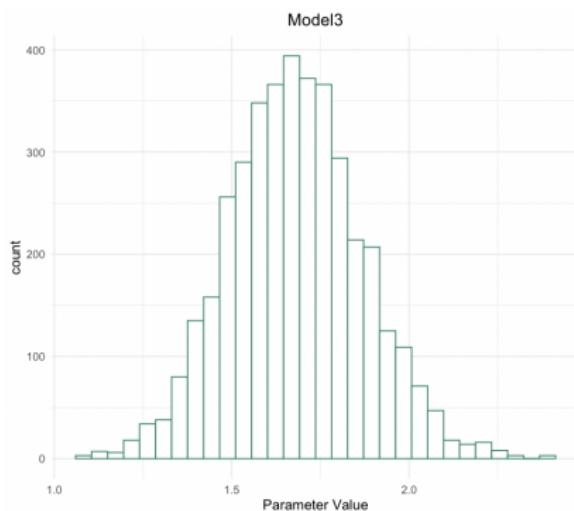
$$\beta_0 \sim \text{Normal}(0, 0.5)$$

$$\beta_1 \sim \text{Normal}(0, 1)$$

$$\beta_{1,j} \sim \text{Normal}(0, \sigma_j)$$

$$\sigma_j \sim \text{Halfcauchy}(10)$$

Model 3: Results



Regression Coefficients:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS	universität
Intercept	-17.77	2.27	-22.20	-13.18	1.00	2303	2442	
urbanindex	1.68	0.19	1.32	2.06	1.00	2509	2291	
Percentage.Retirees..65..	-4.28	3.82	-15.08	-0.11	1.00	1768	1379	

Model 4

Let the response variable 'Winning Party' be y , predicted by the variables Urban Index and the Percentage of Retirees. Let i and k represent the indices for district and region respectively.

$$y_{i,j} \sim \text{Ber.}(\text{logit}^{-1}(\theta_j))$$

$$\theta_j = \beta_0 + \beta_1 \cdot \text{Urban_Index} + \beta_{1,k} \cdot \text{Urban_Index} + \beta_2 \cdot \text{Pct_Retirees}$$

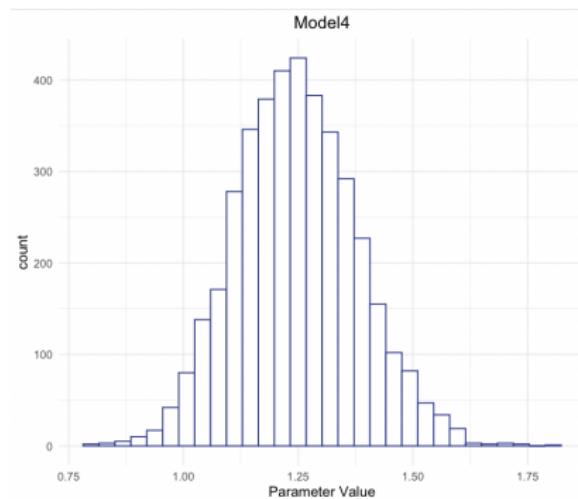
$$\beta_0 \sim \text{Normal}(0, 0.5)$$

$$\beta_1 \sim \text{Normal}(0, 1)$$

$$\beta_{1,k} \sim \text{Normal}(0, \sigma_k)$$

$$\sigma_k \sim \text{Halfcauchy}(10)$$

Model 4: Results



Regression Coefficients:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	-13.24	1.62	-16.53	-10.09	1.00	2028	2228
urbanindex	1.25	0.13	1.00	1.52	1.00	2548	2402
Pct.Retirees.65	-3.79	2.83	-10.98	-0.30	1.01	1460	1075

Table of Contents

Model Comparison: R^2

Model	Estimate	Estimate Error	Q 2.5	Q 97.5
1	0.568	0.0246	0.516	0.612
2	0.534	0.0255	0.479	0.579
3	0.529	0.0255	0.474	0.574
4	0.408	0.0234	0.359	0.449

Model Comparison: RMSE

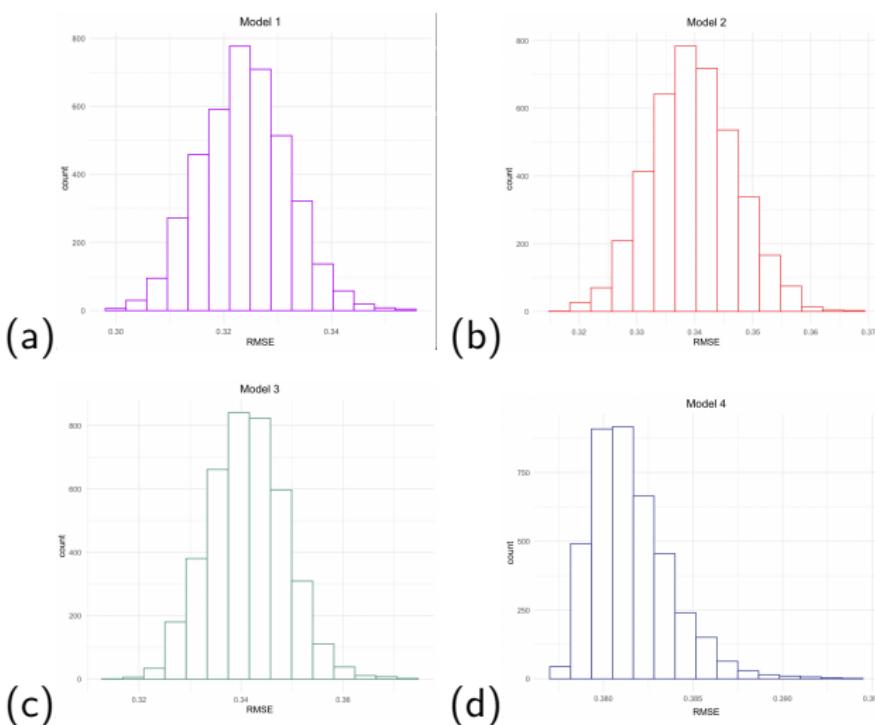


Figure: RMSE Comparison

Model Comparison: Data Log-likelihood

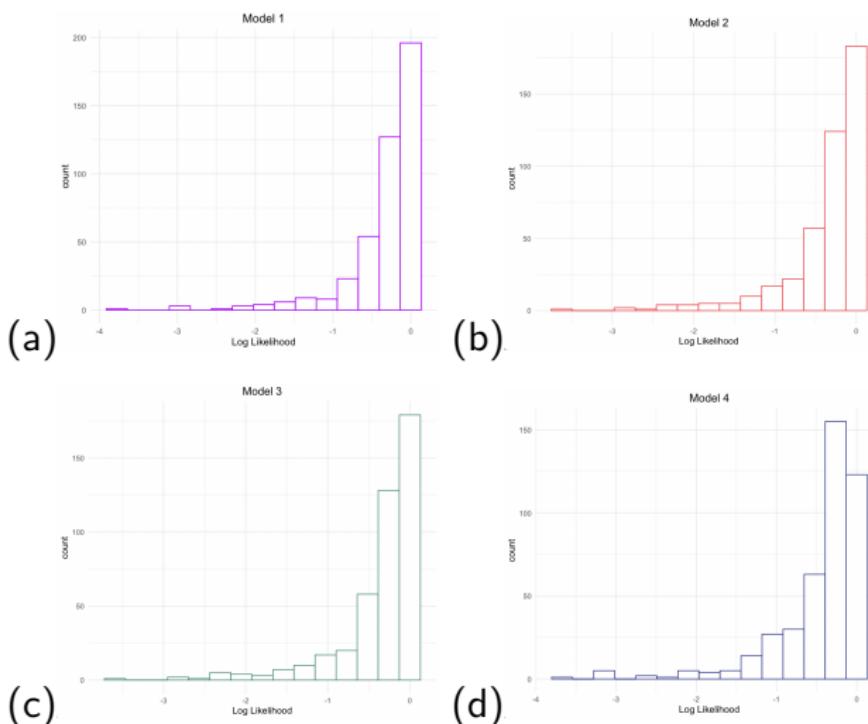


Figure: Log-Likelihood Comparison

Model Comparison: Leave One Out

Table:

	elpd_diff	se_diff	elpd_loo	se_elpd_loo	p_loo	se_p_loo	looic	se_looic
model1.final	0	0	-165.575	11.670	30.124	3.210	331.149	23.339
model3.final	-13.051	2.469	-178.625	11.915	30.126	3.028	357.250	23.830
model2.final	-13.397	2.507	-178.972	12.101	31.864	3.319	357.944	24.202
model4.final	-36.792	7.355	-202.367	12.056	5.383	0.508	404.734	24.112

Introduction
oo

Dataset Description
oooooo

Model Setup
oo

Models
oooooooo

Model Comparison
oooo

Conclusion
●ooo

Table of Contents

Introduction
oo

Dataset Description
oooooo

Model Setup
oo

Models
oooooooo

Model Comparison
oooo

Conclusion
o●oo

Best Model

Possible Improvements Next Steps

- Scaling all covariates to same scale for better parameter estimation interpretability
- More prior experimentation
- Varying intercept models, as we currently only have varying slopes
- Prior sensitivity analysis

Conclusions

In all of our models we found that urbanization has a small positive impact on the logit probability of voting democratically.

Contact:

{eldaleona.odole, leonor.cunha, anarghya.murthy}@tu-dortmund.de