Does Urbanisation Predict Election Outcomes?

A Bayesian's Perspective

Eldaleona Odole Leonor Cunha Anarghya Murthy

Faculty of Statistics

January 25, 2025

technische universität dortmund

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:

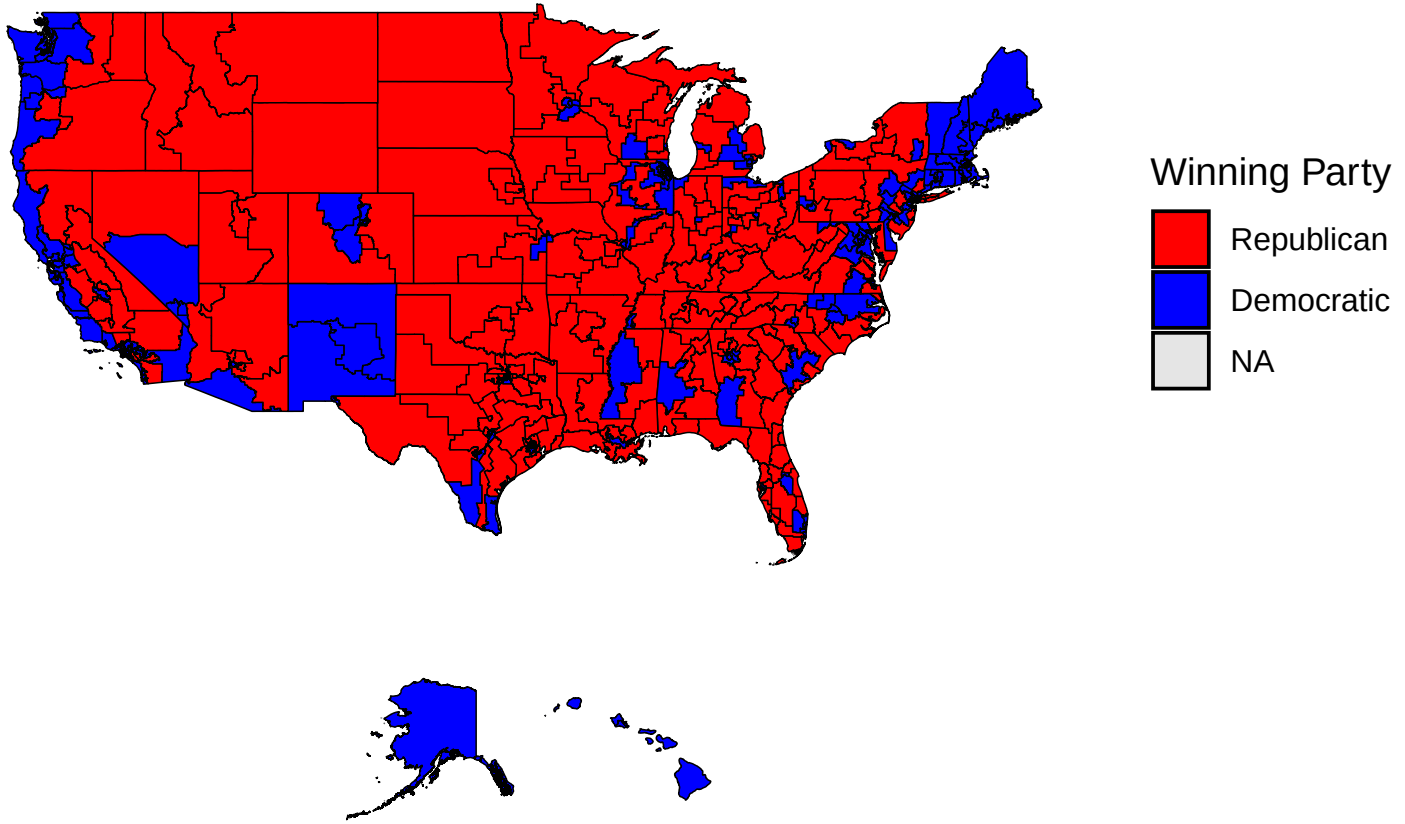
1. Demographic Data (US Census Bureau)
2. Urbanization (FiveThirtyEight)
3. Regional Information (US Census Bureau)
4. Election Results (FiveThirtyEight)

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.

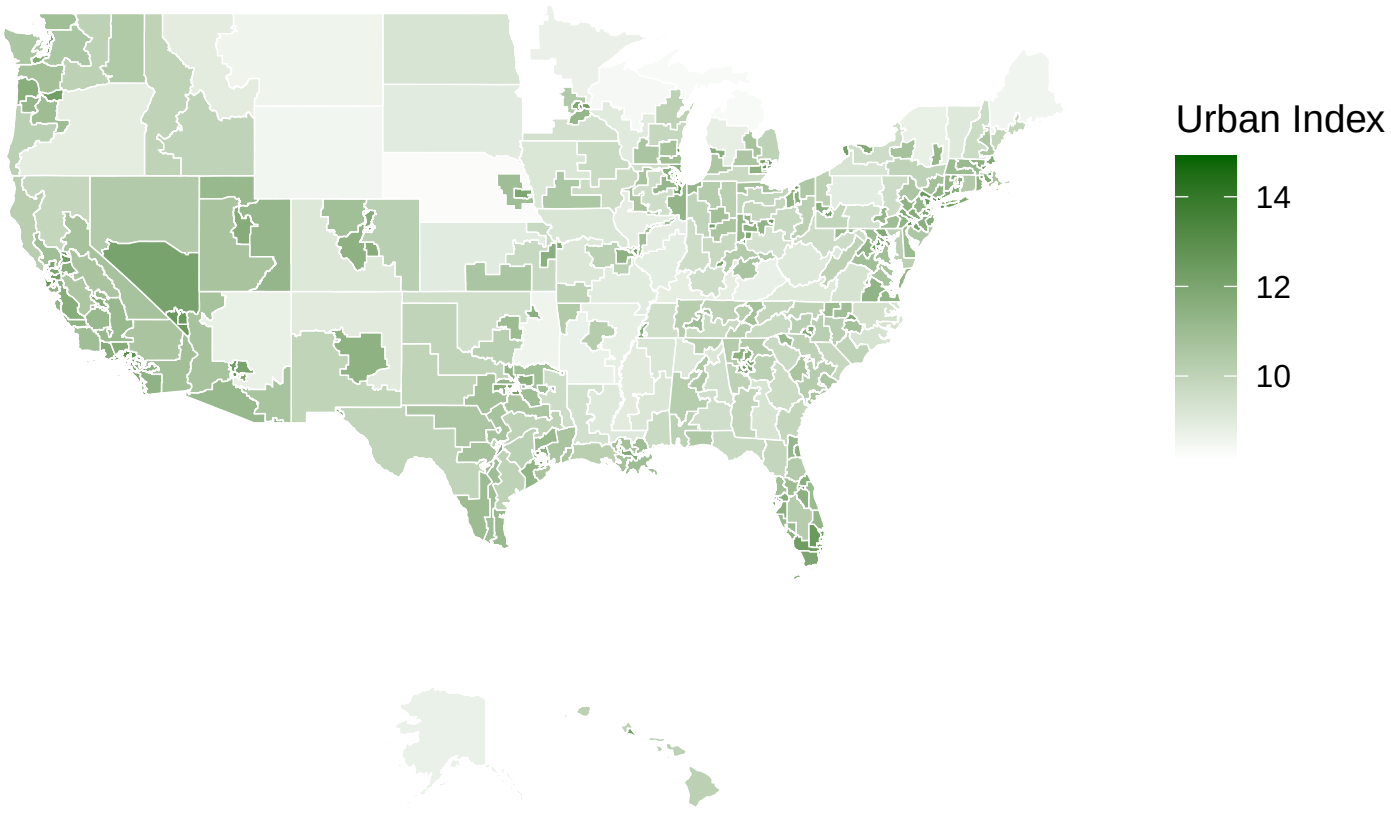
Winning Party by Congressional District



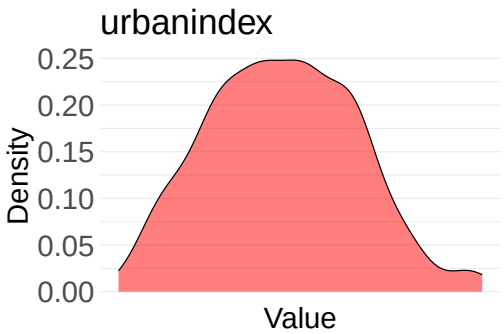
Urban Index

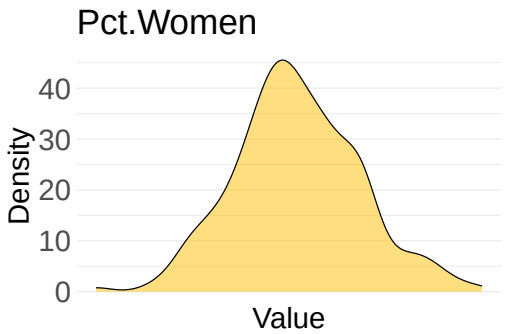
Our dependent variable of interest is the Urban Index from FiveThirtyEight.

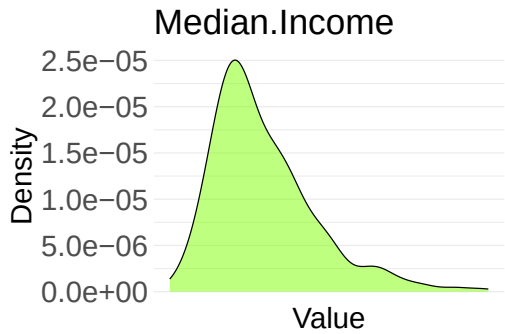
Urban Index by Congressional District

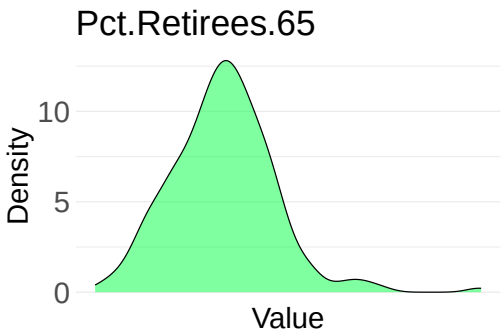


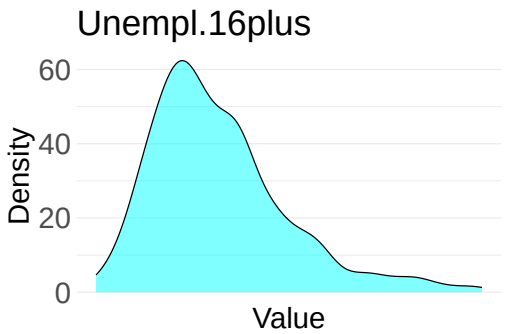
Densities

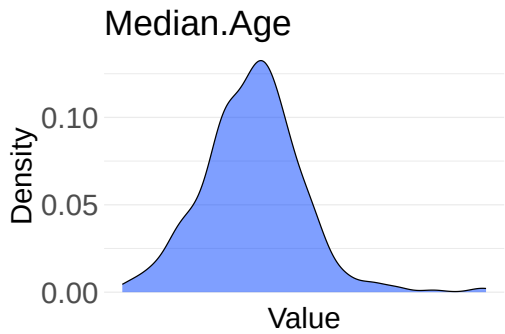

urbanindex
Density
0.25
0.20
0.15
0.10
0.05
0.00
Value

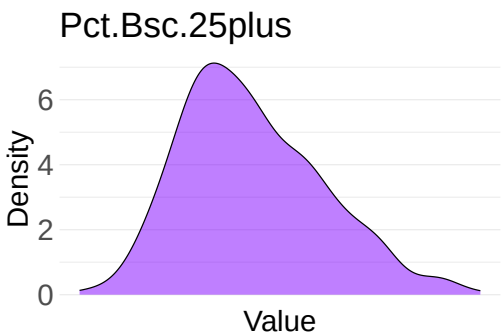

Pct.Women
Density
40
30
20
10
0
Value

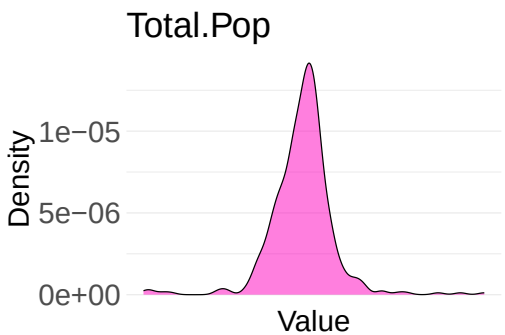

Median.Income
Density
2.5e−05
2.0e−05
1.5e−05
1.0e−05
5.0e−06
0.0e+00
Value


Pct.Retirees.65
Density
10
5
0
Value

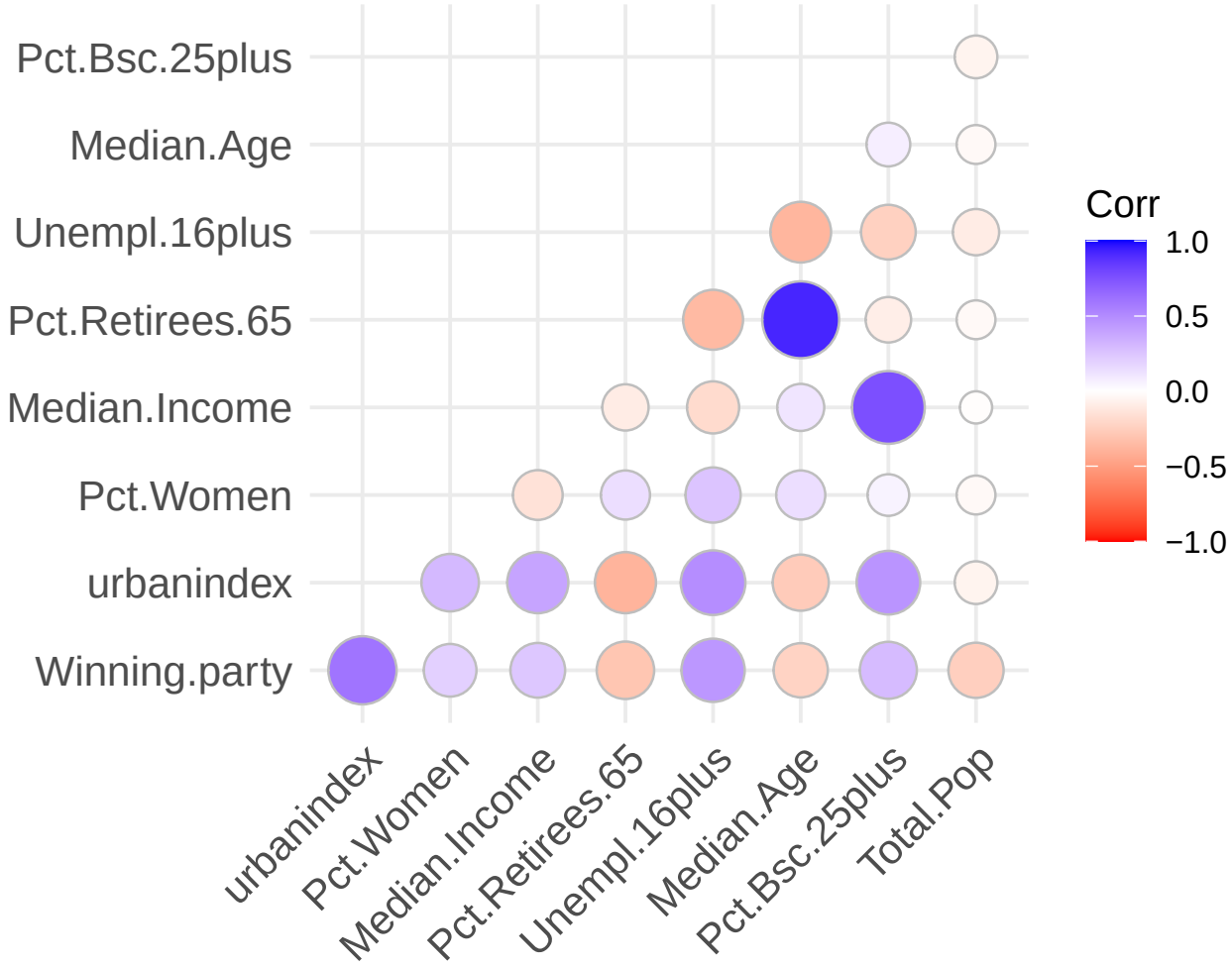

Unempl.16plus
Density
60
40
20
0
Value


Median.Age
Density
0.10
0.05
0.00
Value

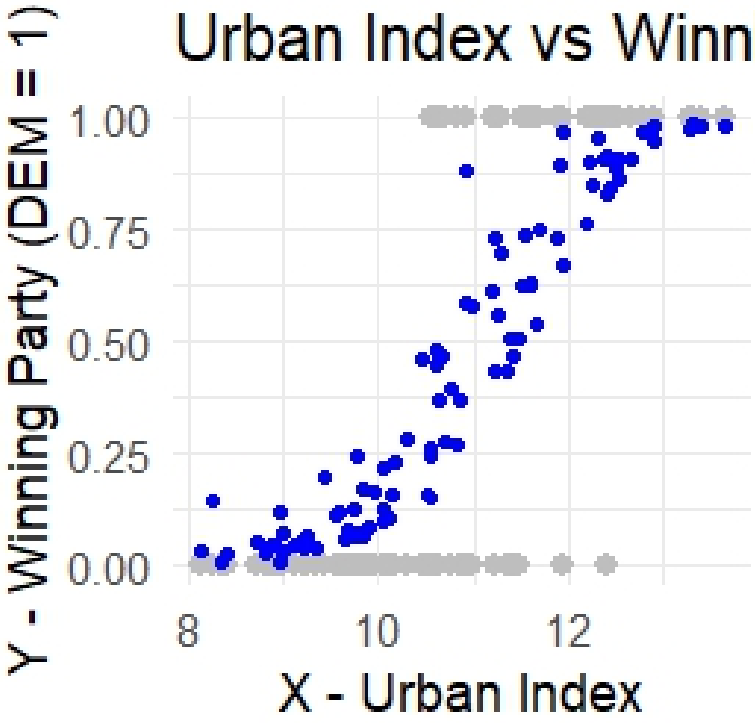

Pct.Bsc.25plus
Density
6
4
2
0
Value

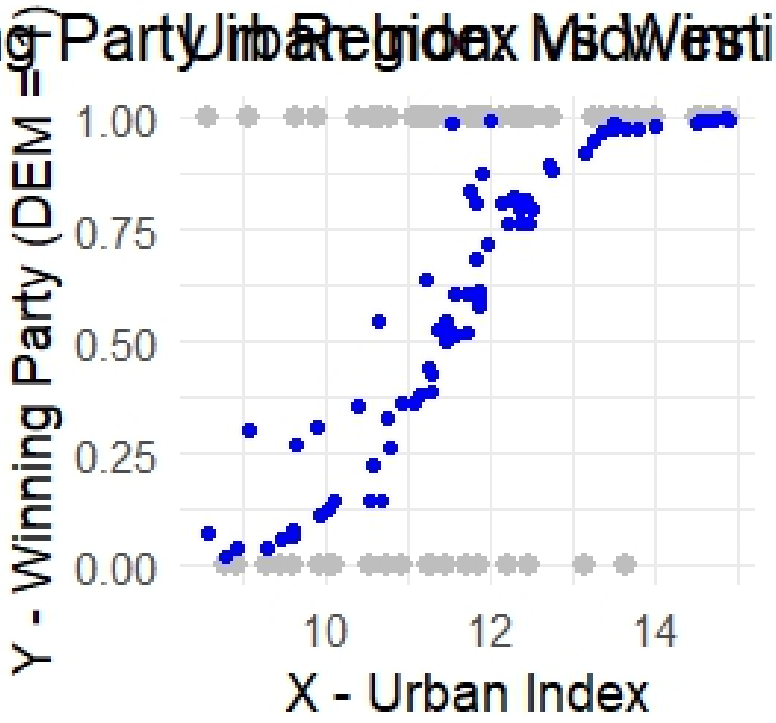

Total.Pop
Density
1e−05
5e−06
0e+00
Value

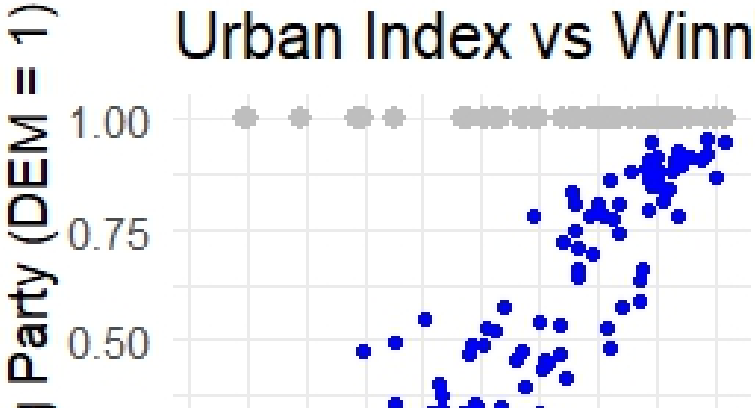
Correlation Matrix



Motivation for Hierarchical Modelling







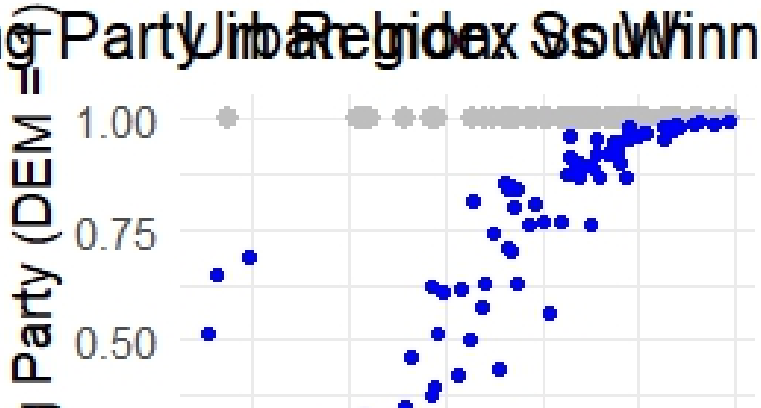


Table of Contents

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 exchangable within each state and each state is exchangable within its region
- ???

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.

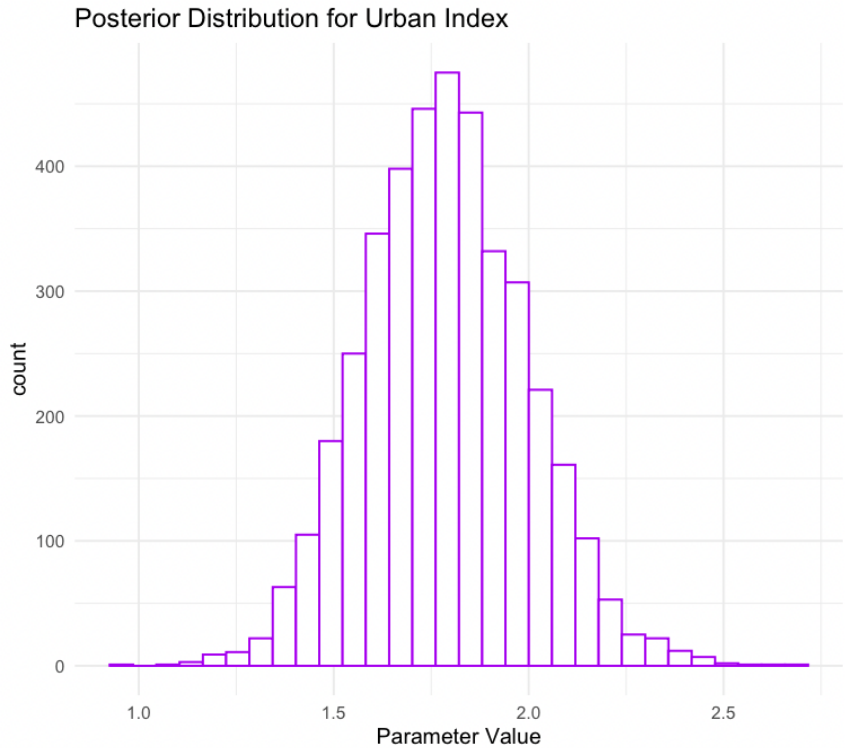
$$y_{i,j} \sim Ber.(logit^{-1}(\theta_j))$$

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

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

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

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
pct.retirees	-3.84	4.41	-15.57	0.97	1.00	2060	1858

Model 2

Model 2: Results

Model 3

Model 3: Results

Model 4

Model 4: Results

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

Table of Contents

Alternative Priors

What's changed

Table of Contents

Raw references

- stargazer
- tidybayes
- brms, stan