Does Urbanisation Predict Election Outcomes?

A Bayesian's Perspective

Eldaleona Odole Leonor Cunha Anarghya Murthy

Faculty of Statistics

January 21, 2025

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)

Dataset

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

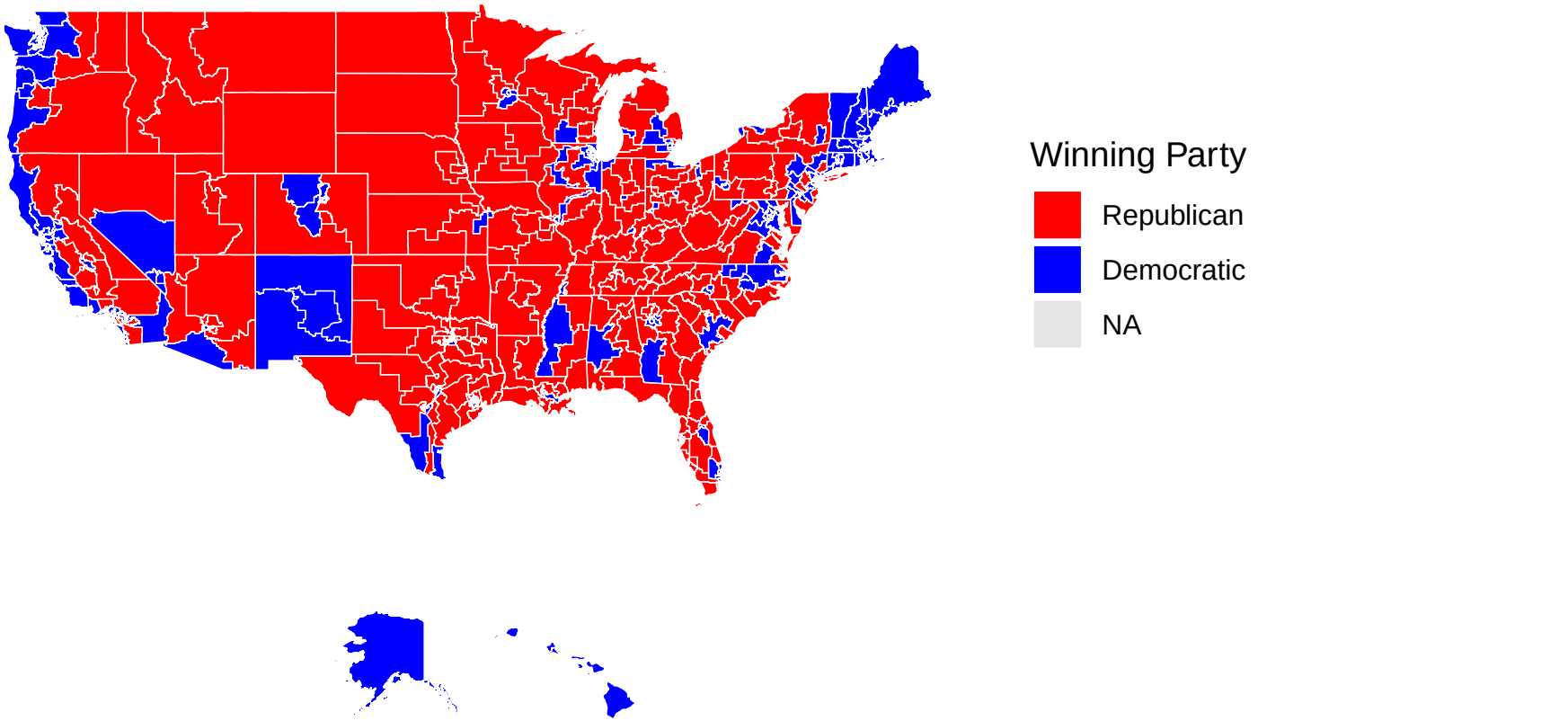
1. Demographic Data (US Census Bureau)
2. Urbanization (FiveThirtyEight)
3. Regional Information (US Census Bureau)
4. Election Results (FiveThirtyEight)

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.

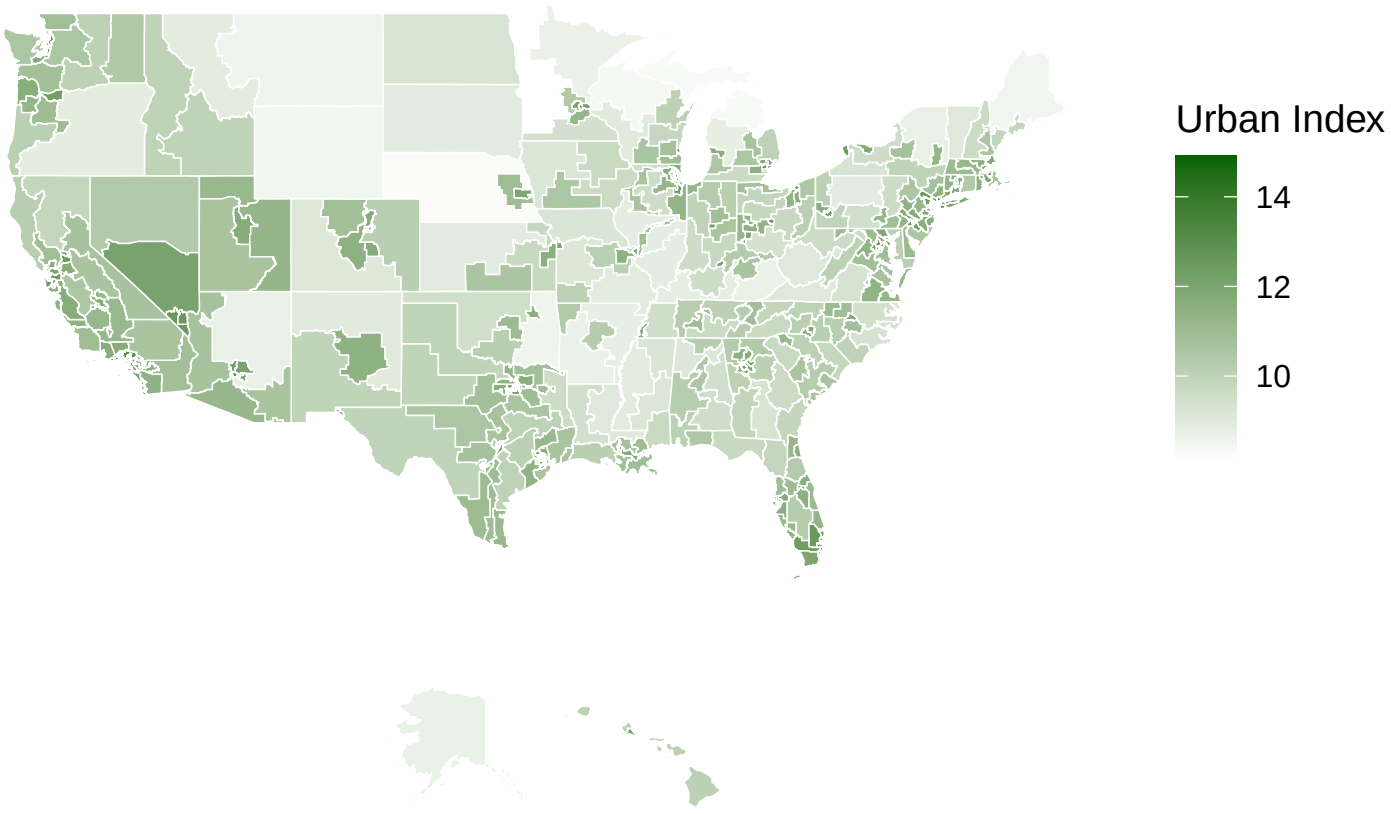
Winning Party by Congressional District



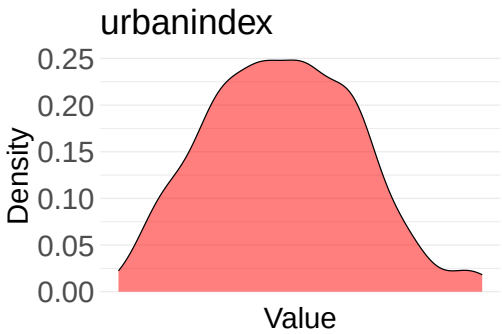
Urban Index

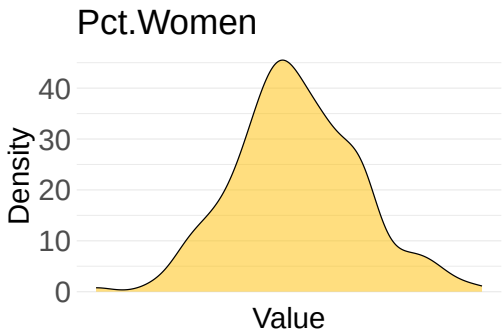
Our dependent variable of interest is the Urban Index from FiveThirtyEight.

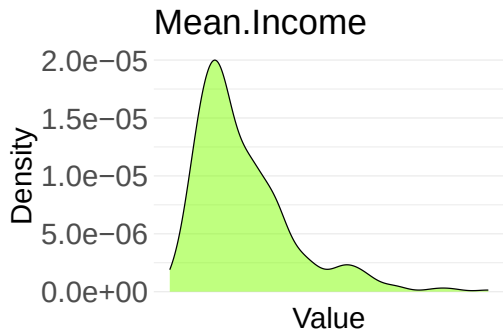
Urban Index by Congressional District

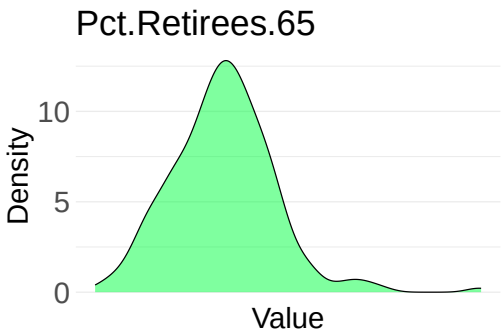


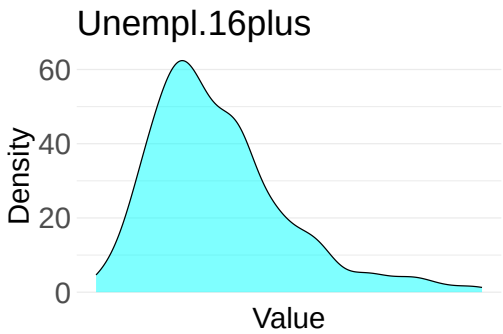
Densities

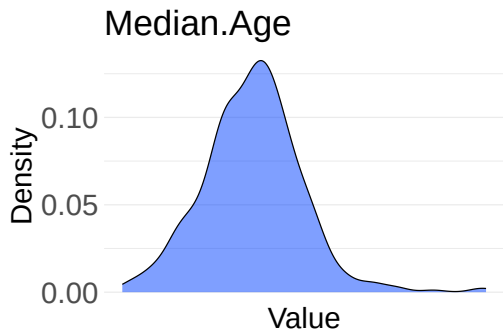

urbanindex
Density
0.25
0.20
0.15
0.10
0.05
0.00
Value

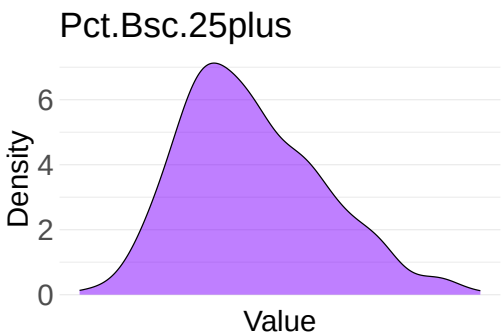

Pct.Women
Density
40
30
20
10
0
Value

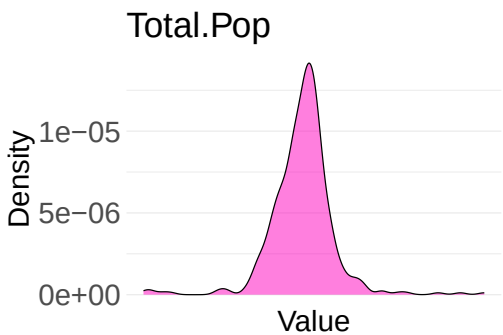

Mean.Income
Density
2.0e−05
1.5e−05
1.0e−05
5.0e−06
0.0e+00
Value


Pct.Retirees.65
Density
10
5
0
Value

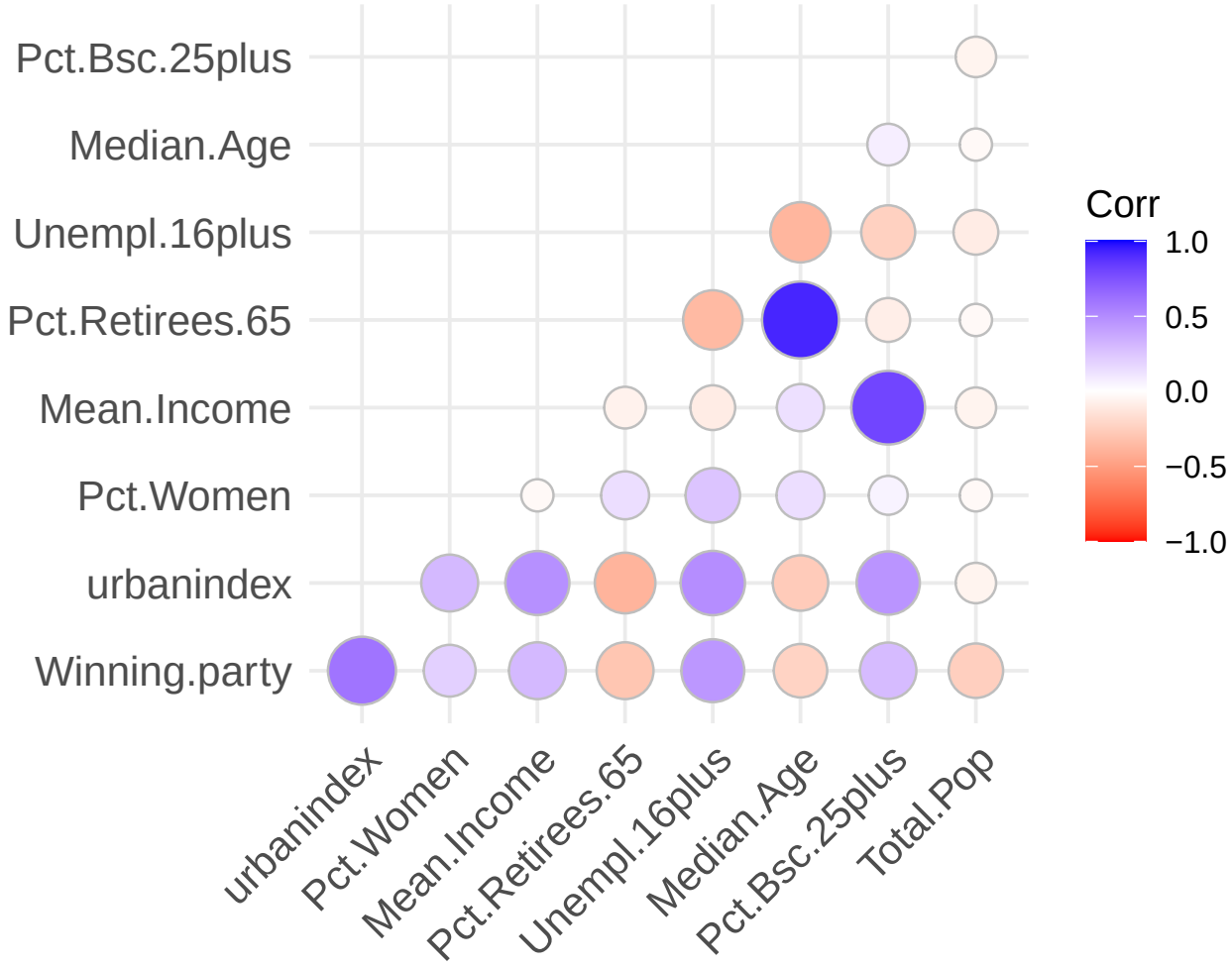

Unempl.16plus
Density
60
40
20
0
Value


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



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
- ???

Model

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.

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

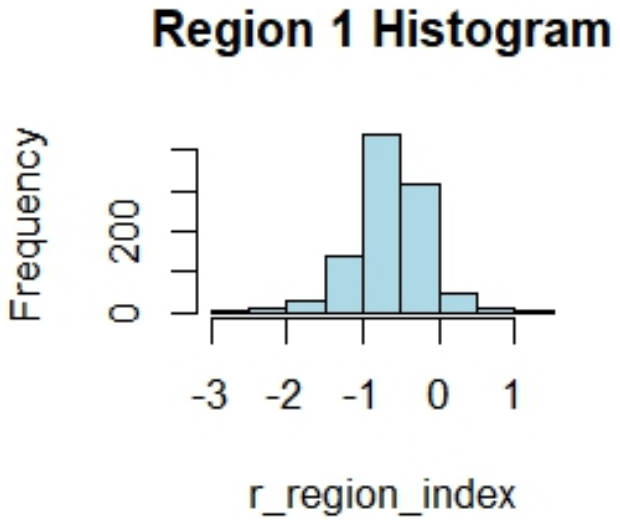
$$\theta_j := \beta_0, j + x_{i,j} * \beta_{1,j} + z_{i,j}^T * \gamma_{1,j}$$

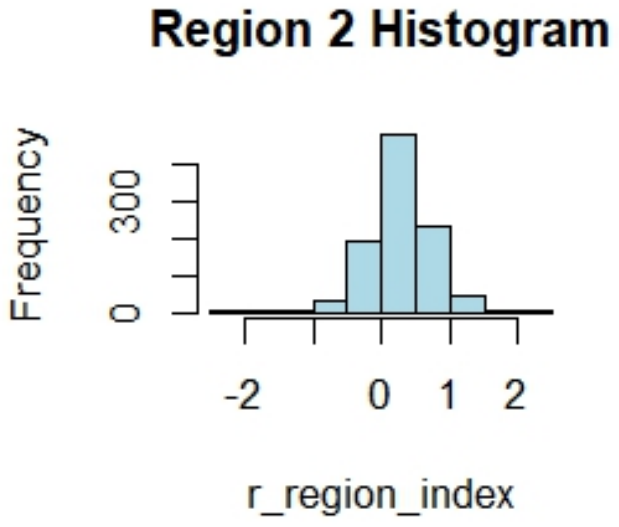
Factor name

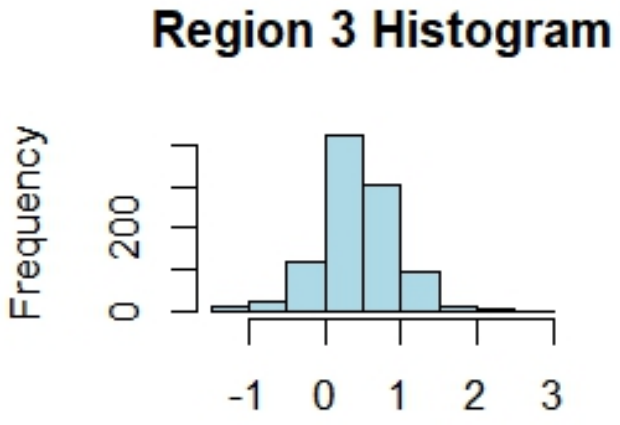
Urbanization Index

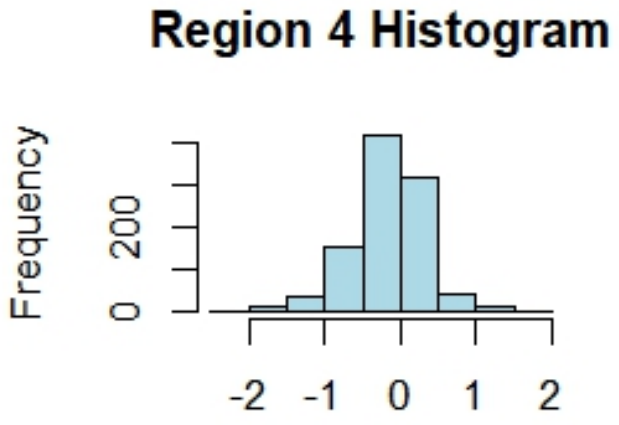
Urbanization Index is a measure created by fivethirtyeight calculated based on populatin density of a given district. We assume that urbaniation has some positive impact on the likelihood of voting democrat, because urban areas tend to lean more democratic [source]. However, we assume a regional effect as this relationship is likely more pronounced in more rural regions generally.

Varying Intercept Model









Varying Intercept Model - II

STAN Code

Raw references

- stargazer
- tidybayes
- brms, stan