

Does Urbanisation Predict Election Outcomes?

A Bayesian's Perspective

Eldaleona Odole Leonor Cunha Anarghya Murthy

Faculty of Statistics

January 22, 2025

Table of Contents

- 1 Introduction
- 2 Dataset Description
- 3 Model Setup
- 4 Priors
- 5 Pooled Model
- 6 Unpooled Model
- 7 Results - Hierarchical Model
- 8 Results - Varying Intercept Model
- 9 Centred vs Non-centred Parameterisation/ Sampling
Investigation with STAN
- 10 Raw References

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:

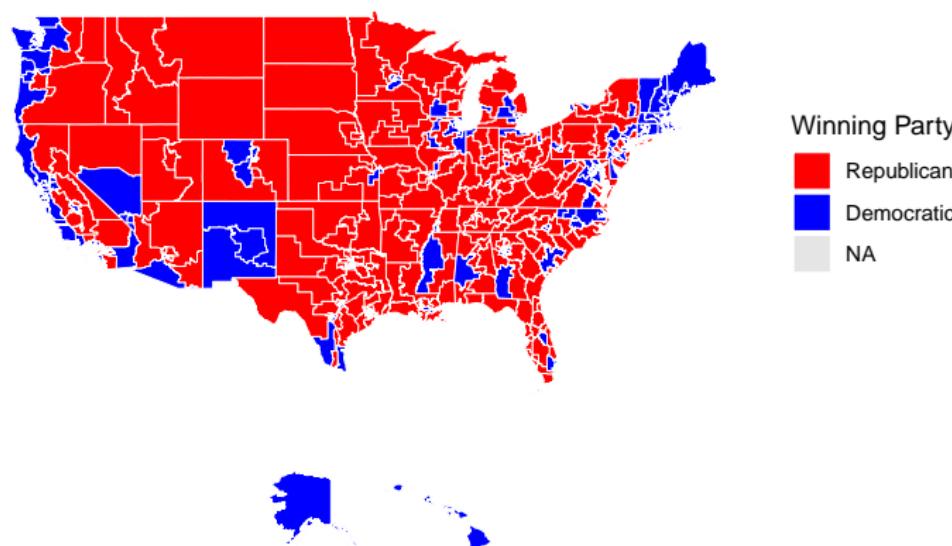
- ① Demographic Data (US Census Bureau)
- ② Urbanization (FiveThirtyEight)
- ③ Regional Information (US Census Bureau)
- ④ Election Results (FiveThirtyEight)

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.

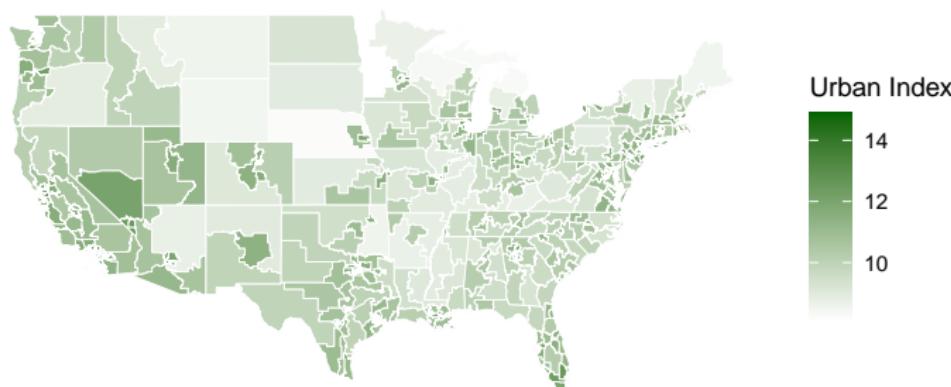
Winning Party by Congressional District



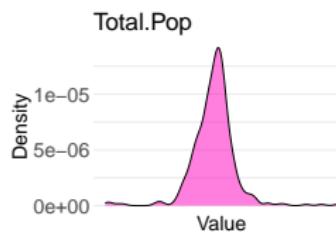
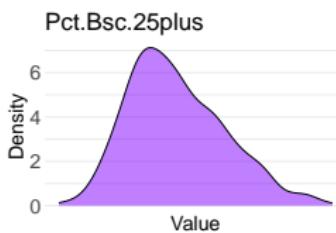
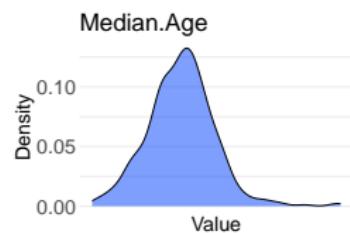
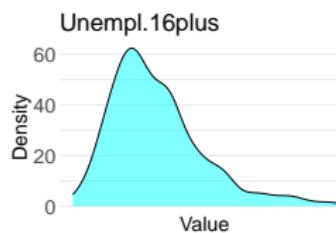
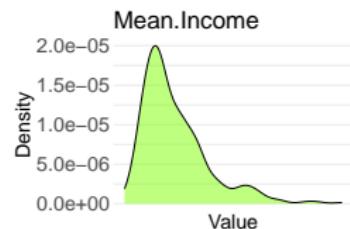
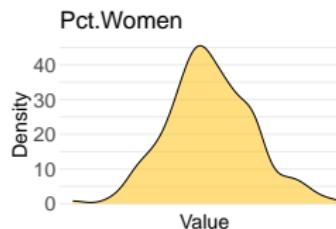
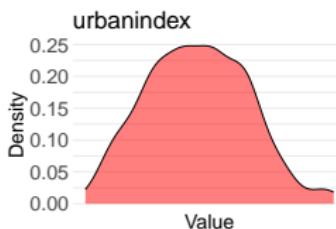
Urban Index

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

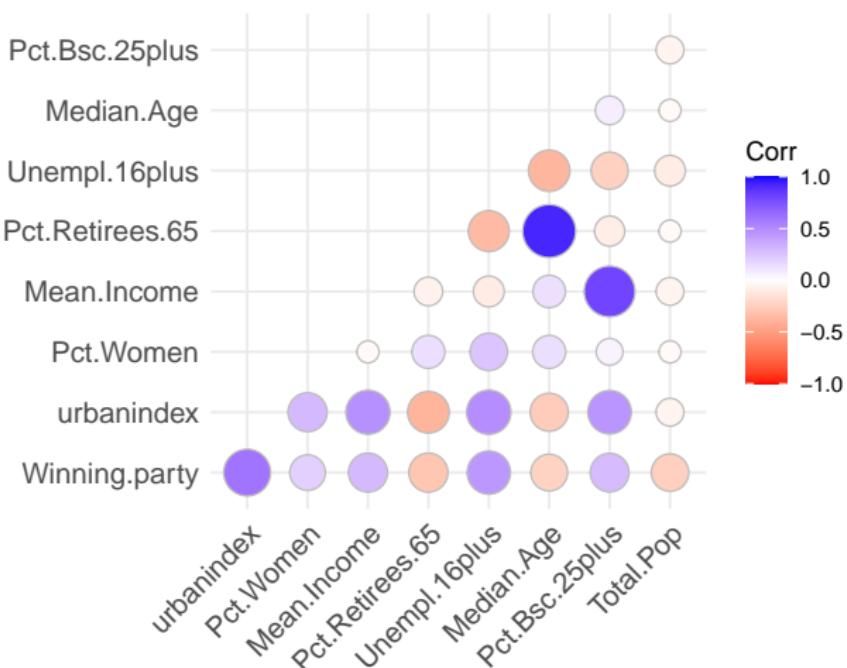
Urban Index by Congressional District



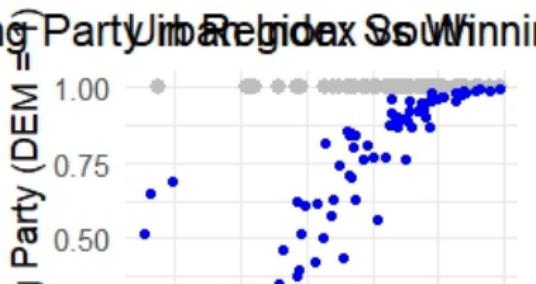
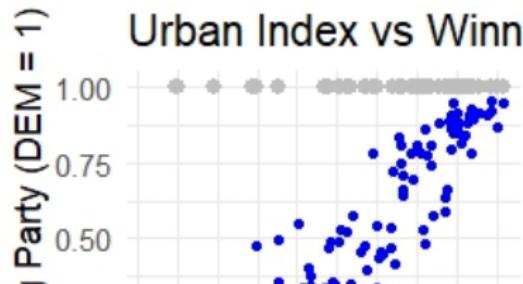
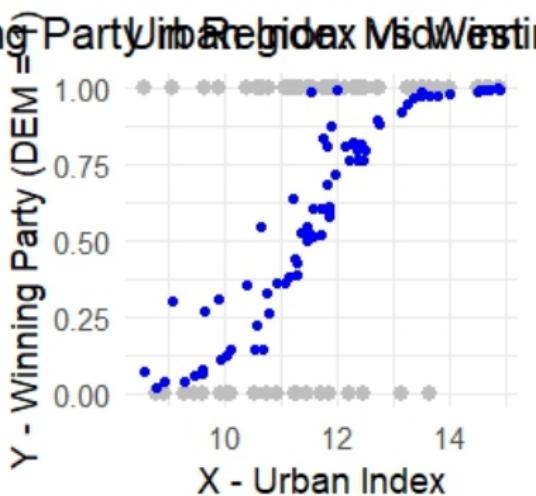
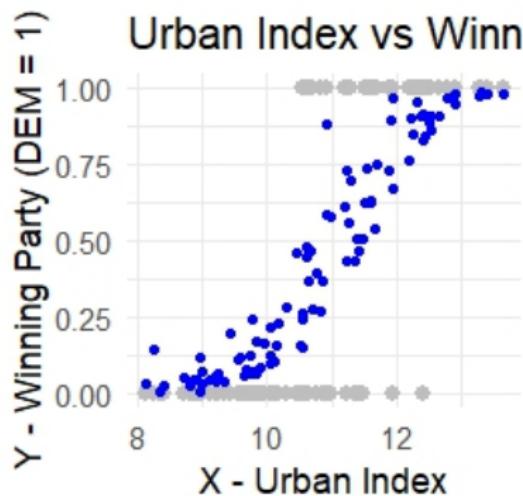
Densities



Correlation Matrix



Motivation for Hierarchical Modelling



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
- ???

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.

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

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

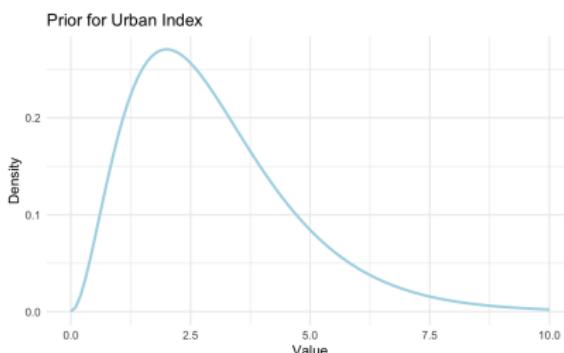
$$\beta_{1,j} \sim Gam.(1, \tau)$$

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

Factor name

Urbanization Index

Urbanization Index is a measure created by fivethirtyeight calculated based on population density of a given district. We assume that urbanization 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.



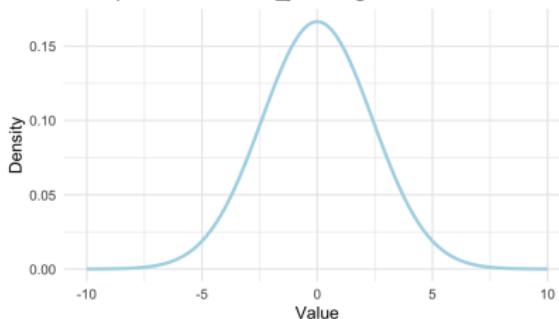
Urbanization Category

Urbanization category is the measure created by fivethirtyeight that bounds the numerical values of urban index into several buckets. We assume that each of these categories with have a positive association

Total Population

Since each district is drawn to have roughly the same population within each state, we think that this covariate will have a very small influence as there should be nearly no variance within a state. However, this effect maybe be state dependent.

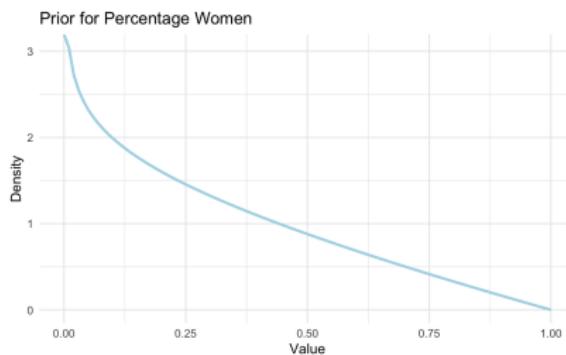
Example Prior for State_j with sigma = 2.394



$$\beta_{total.pop|state} \sim N(0, \sigma_{state})$$

Percentage Women

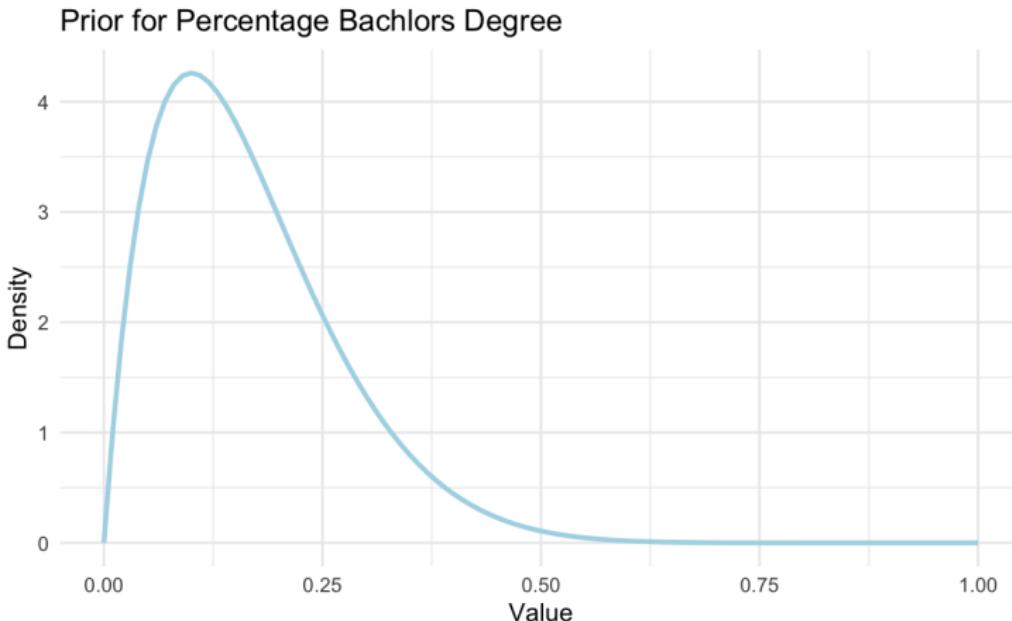
We found that women tend to lean more democratic and have higher turnout than men. Leading us to believe there should be a positive association between percentage women in a district and probability of voting democratic. However percentage of women isn't a variable that varies greatly with geography so we assume if there is an effect that it will be quite small.



$$\beta_{pct.women} \sim Beta(\alpha = \frac{6}{7}, \beta = 2)$$

Percentage of Population over 25 with Bachelors Degree

We found that those holding a bachelors degree or higher education lean democratic however they make up a small part of the population



Pooled Model - Description

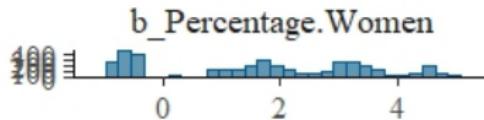
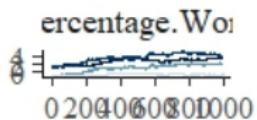
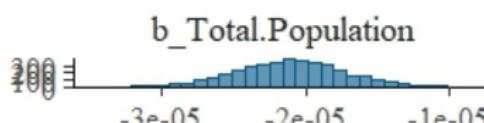
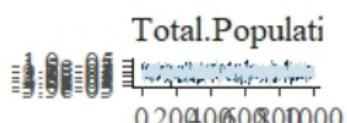
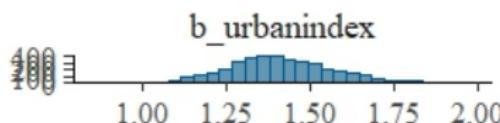
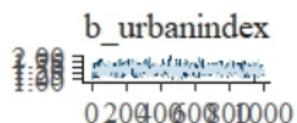
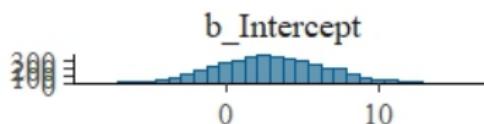
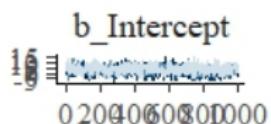
The simplest type of model where no hierarchies are taken into account:

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

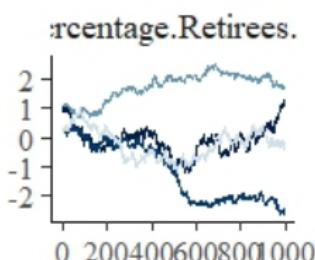
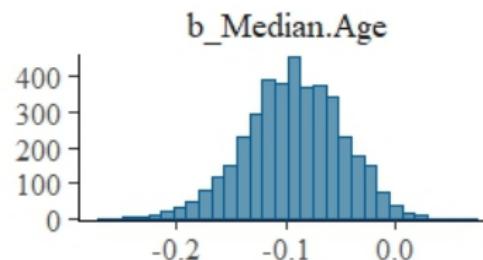
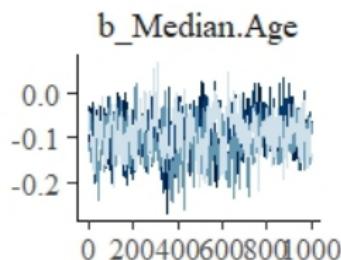
$$\theta := \beta_0 + x_i * \beta_1 + z_i^T * \gamma_1$$

$$\beta_{1,j} \sim Gam.(3, 1)$$

Pooled Model - Trace Plot

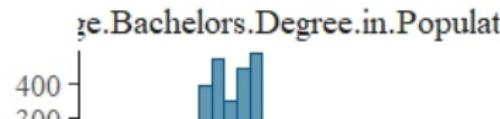
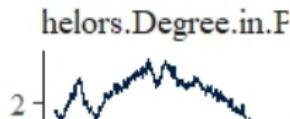
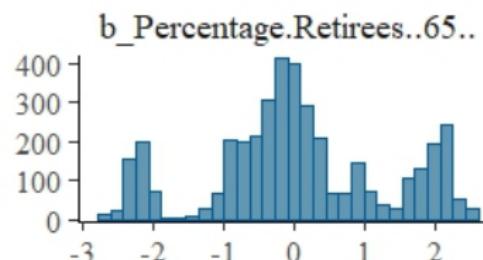


Pooled Model - Trace Plot 2



Chain

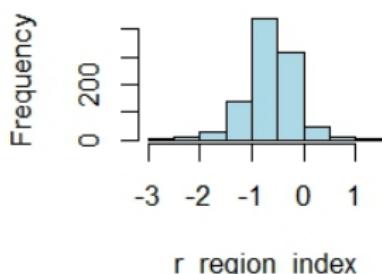
- 1
- 2
- 3
- 4



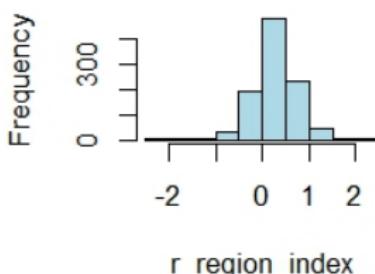
Description - Unpooled Model

Varying Intercept Model

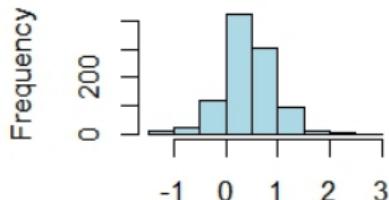
Region 1 Histogram



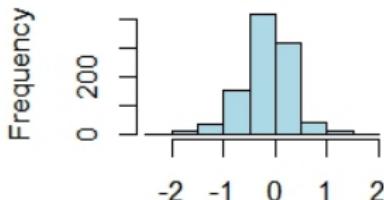
Region 2 Histogram



Region 3 Histogram



Region 4 Histogram



Varying Intercept Model - II



STAN Code

Raw references

- stargazer
- tidybayes
- brms, stan