

**BST 701: Bayesian Modeling in Biostatistics**  
**Breheeny**

Assignment 4  
Due: Thursday, April 25

The data set `nes2000` contains data from the National Election Study (NES); this particular data set contains responses from 861 individuals who responded to the survey prior to the presidential election in 2000. The outcome of interest is `rVote`, whether the respondent intended to vote for the Republican candidate (in 2000, this was George W. Bush) or not. `rVote = 1` indicates that the respondent did intend to vote for Bush, `rVote = 0` indicates that the respondent intended to vote for the Democratic candidate, Al Gore. Individuals who did not intend to vote, or who intended to vote for a different candidate, have been excluded from this data set.

The primary focus of this analysis is the complex relationship between income and voting preferences. Specifically, the data set includes the variables `StateIncome`, which measures the per capita annual income of the state that the respondent lives in, and `Income`, which measures the individual respondent's annual income on the following ordinal scale:

- 1: Under \$10,000
- 2: \$10,000–\$25,000
- 3: \$25,000–\$50,000
- 4: \$50,000–\$105,000
- 5: \$105,000 and above

A rather interesting phenomenon is that, if you perform a logistic regression of `rVote` vs. state income, you find that wealthier states were significantly more likely to support Gore. However, if you regress `rVote` vs. individual income, you find that wealthier individuals were significantly more likely to support Bush. This was true in the 2000 election, and has generally been true to varying extents for the past 20 years.

The two findings seem contradictory, and many have referred to this as a “paradox”. Your main task in this assignment is to use multilevel modeling to explain the above results. Specifically, fit a varying-intercept, varying-slope model that allows the effect of personal income to vary by state, with state income as a group-level predictor.

Write a report containing your analyses of the election data. The report should contain the following sections:

- **Models:** Briefly write out your models (see the “Extension” section) in mathematical notation.
- **“National” parameters:** For any parameters in your model that do not vary by state (*i.e.*, exclude the varying slopes and intercepts in this section), report their posterior and comment briefly on its meaning in terms of how it pertains to voter preferences.
- **State-level parameters:** There are too many states to comment individually on all their coefficients, but choose three states: a relatively wealthy state, a relatively poor state, and a state in the middle. Comment on their coefficients and how they compare to each other.

- **The paradox:** Provide an explanation for the apparent paradox mentioned earlier. Feel free to refer to earlier remarks and figures from the previous sections. This section does not necessarily have to be long, depending on whether you've touched on this issue in your earlier explanations.
- **Extension:** As in the previous project, extend the model in some way. One possible extension is to look at demographic variables: **Race**, **Age**, and gender (as an indicator for **Female**) are included in the data set. Another is to include an additional layer of the hierarchy using the **Region** variable (if you wish to alter the NES regions, feel free to do so). In this section, comment on what you find.