Bayesian Hierarchical Modeling to Understand Election Forecasting & Pollster Bias



Haley Reed, Professor Jake Price University of Puget Sound

Introduction

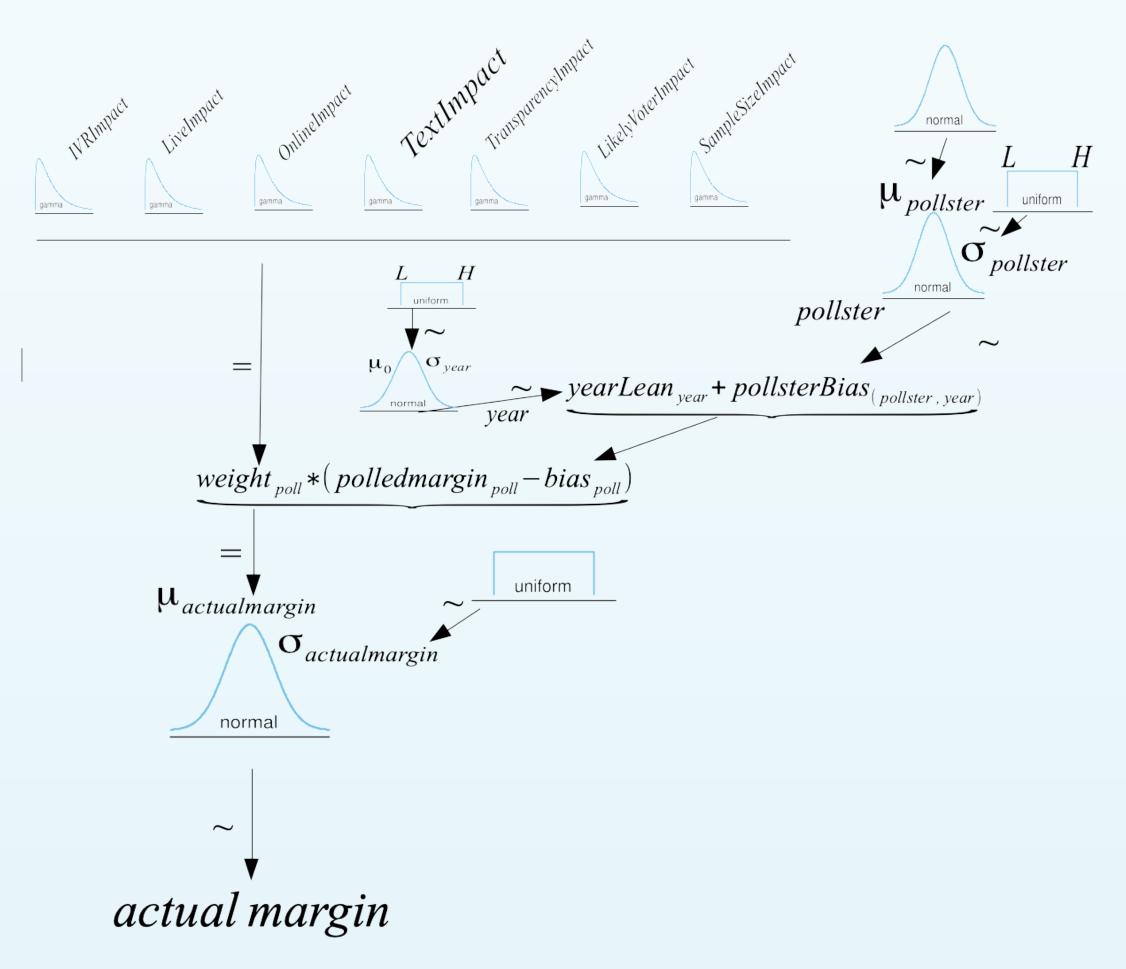
Given the closeness of recent elections, correct and accurate modeling is imperative to understanding and interpreting data from political pollsters, who often report contradictory results based on methodological differences. Even though there are many predictions available, there is still a demand for transparency in the field surrounding the mechanics and algorithms used in election forecasting.

This project employs Bayesian analysis, coupled with hierarchical modeling, to predict election outcomes, select variables, and estimate and correct pollster bias.

The hierarchical model was able to estimate distributions of true vote share with a combination of nominal and numerical variables.

Methodology

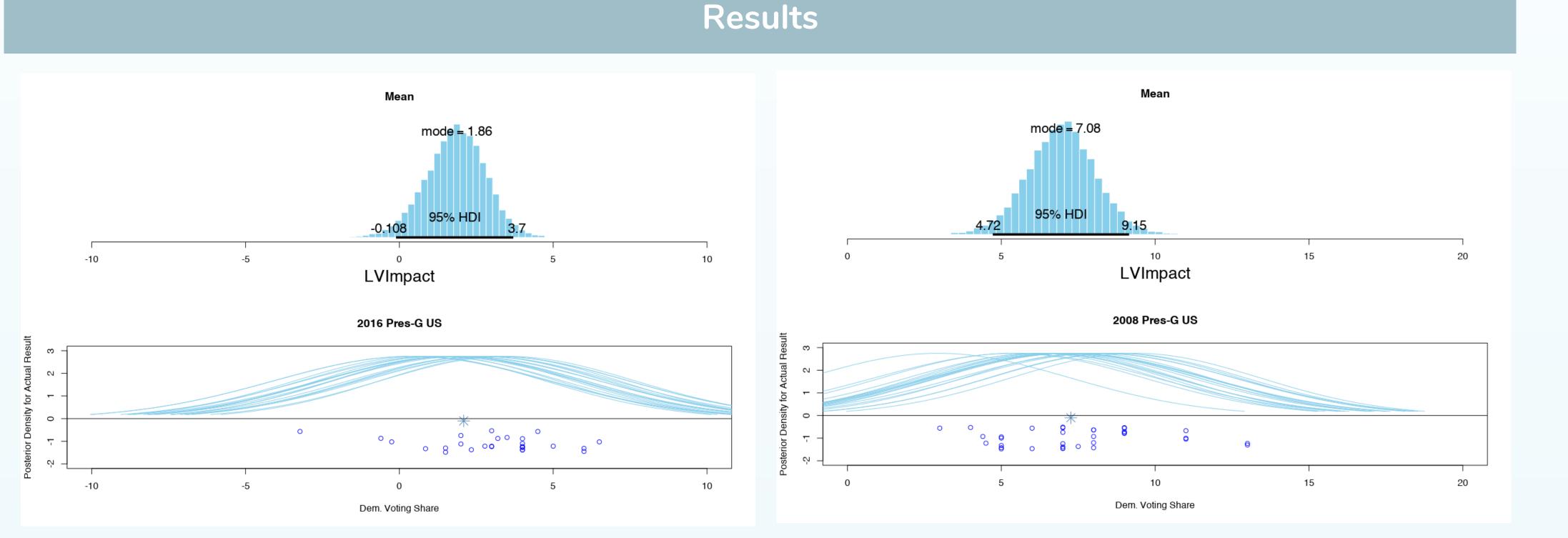
This project utilized the statistical processing systems of R and RStudio. Additionally, JAGS (Just Another Gibbs Sampler) was used extensively to create Markov Monte Carlo samples of posterior distributions for the hierarchal Bayesian models.



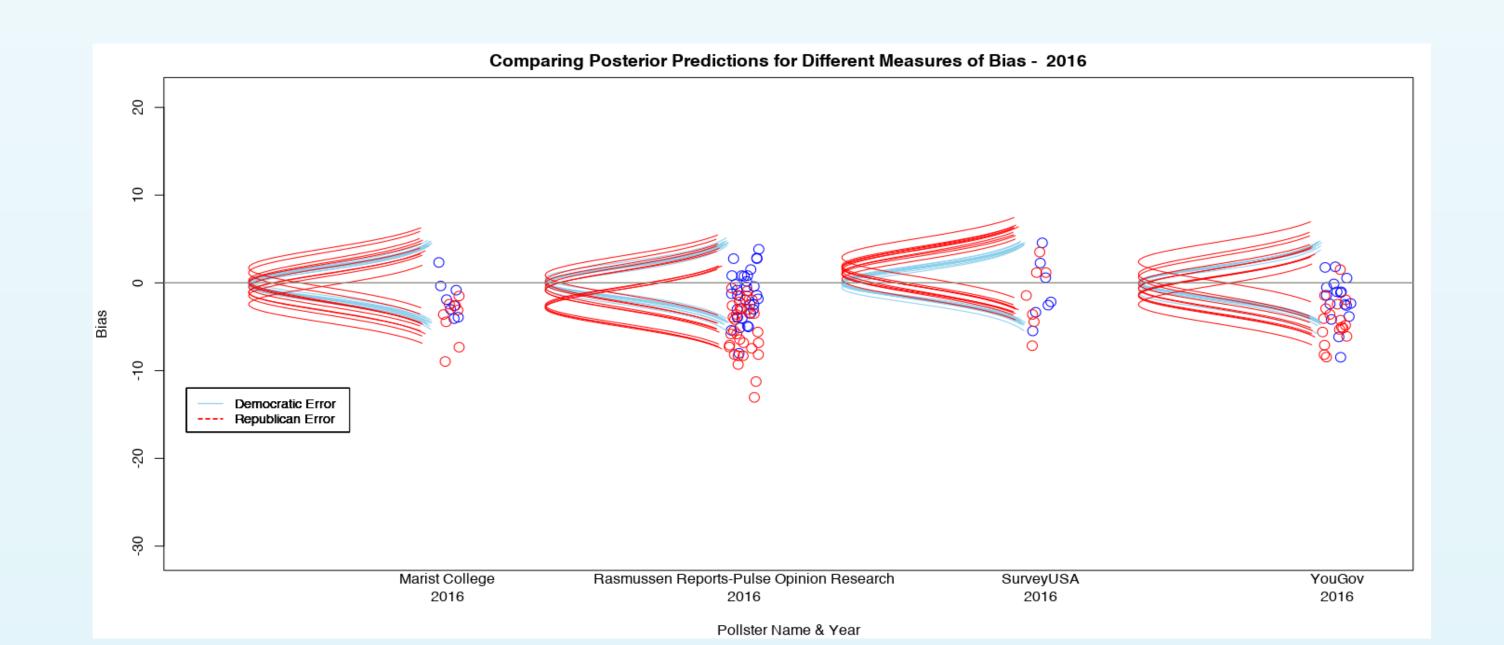
With the help of JAGS, model diagrams like those above, when converted to model statements, are used to predict the most credible parameter values for every parameter at every step in the model.

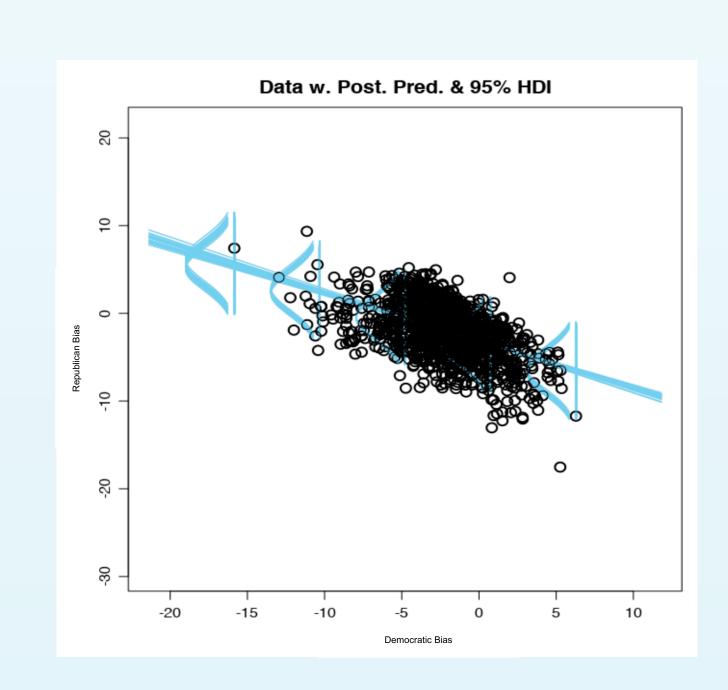
The data was retrieved from publicly available repositories. The final models summarized here used FiveThirtyEight's collection of pollster's reported results.

What's in a Model? What election year was it? Years 2000, 2004, 2008, 2012, and 2016 are included in this model. Year How the poll was conducted – Online? Over the phone? With a live operator? A robot caller? **Delivery Mode** What company or institution is conducting the poll? Pollster A yes/no dummy variable measured by if the pollster participates in the American Association for **Transparency** Public Opinion Research's Transparency Initiative and/or the Roper Center for Public Opinion Research (two organizations that promote transparency in polling). How many respondents to the poll? Sample Size Pollsters try to subset registered voters into classifications of "likely voters", these adjustments, or lack Likely Voter versus Registered Voter Model thereof, can effect the accuracy and bias of the model.



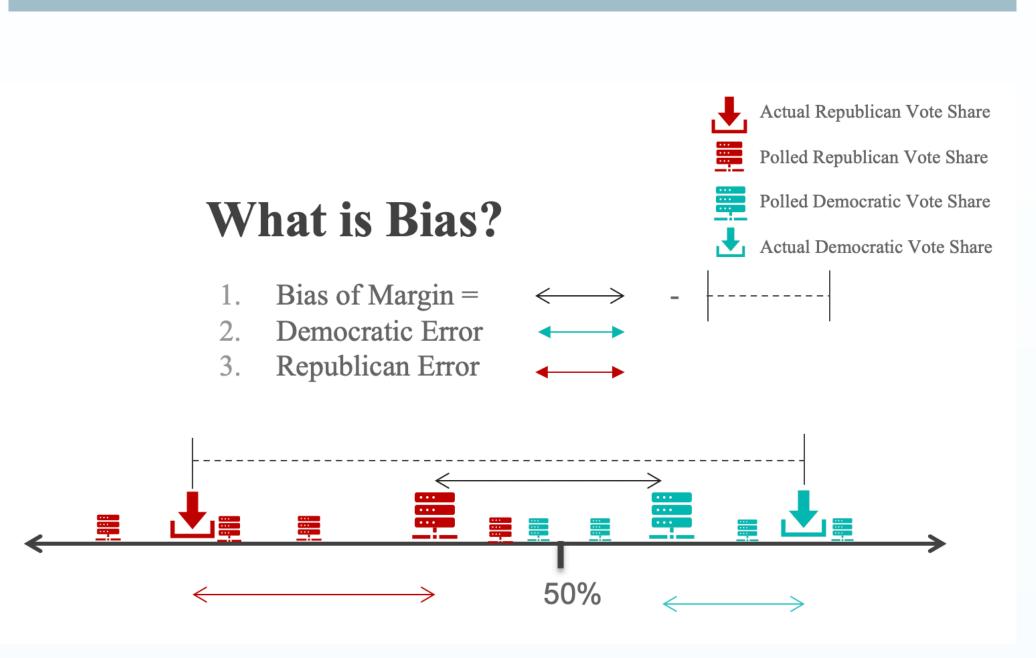
The final models, shown visually here as posterior predictive checks, weigh polls already corrected for bias. The distributions as a result of the hierarchal modeling can reconcile some of the differences between polled opinion and eventual vote share by using parameter values obtained from modeling a total of 953 independent polls since 2000.





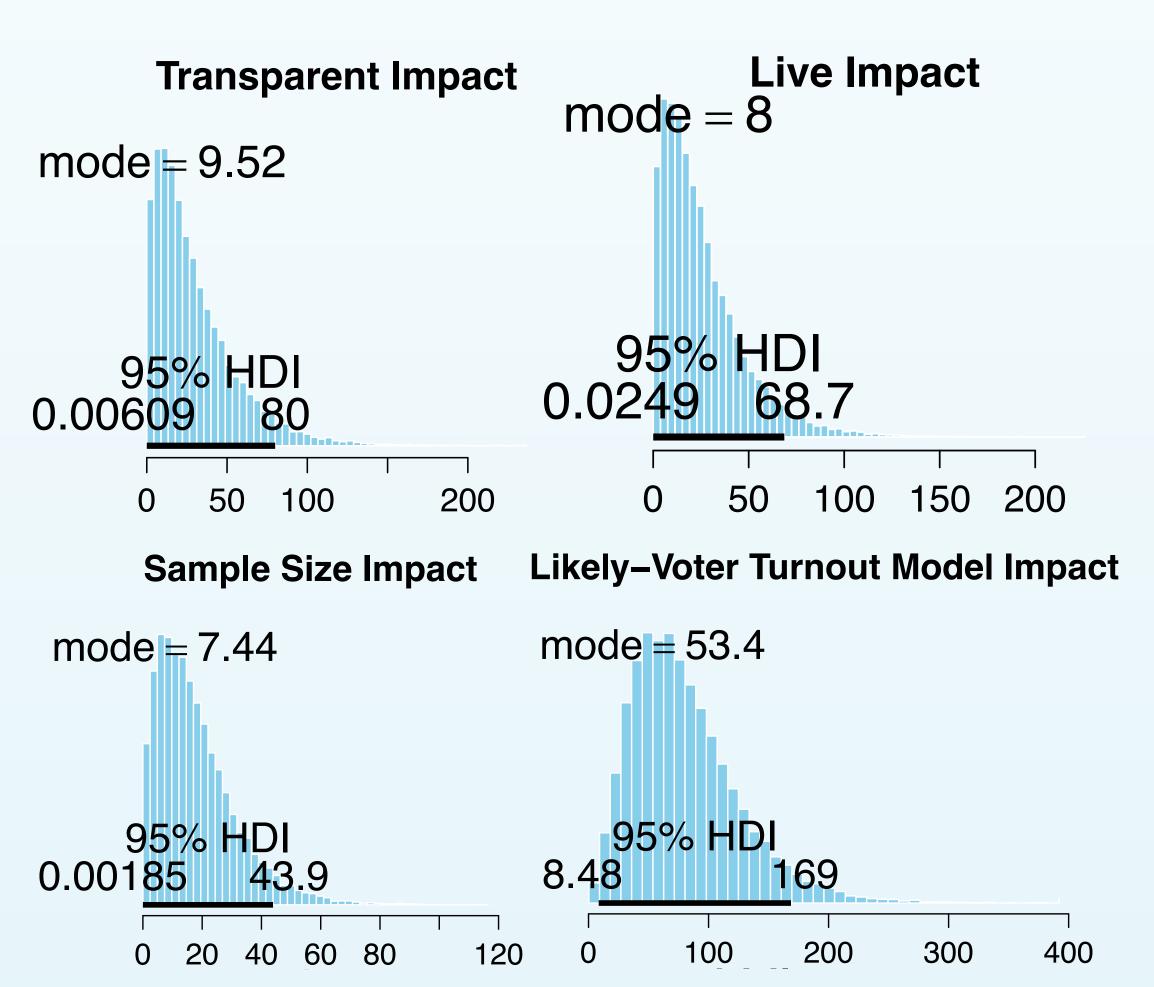
There is a relationship between democratic error and republican error, but it's certainty not a perfect one-to-one relationship. Because of this relationship, republican error can also be modeled for any given poll with just two numerical predictors: democratic error and undecided voter percentage.

Conclusion



How bias and accuracy varies by pollster and polling methods, and to what extent can be understood with a Bayesian hierarchical model.

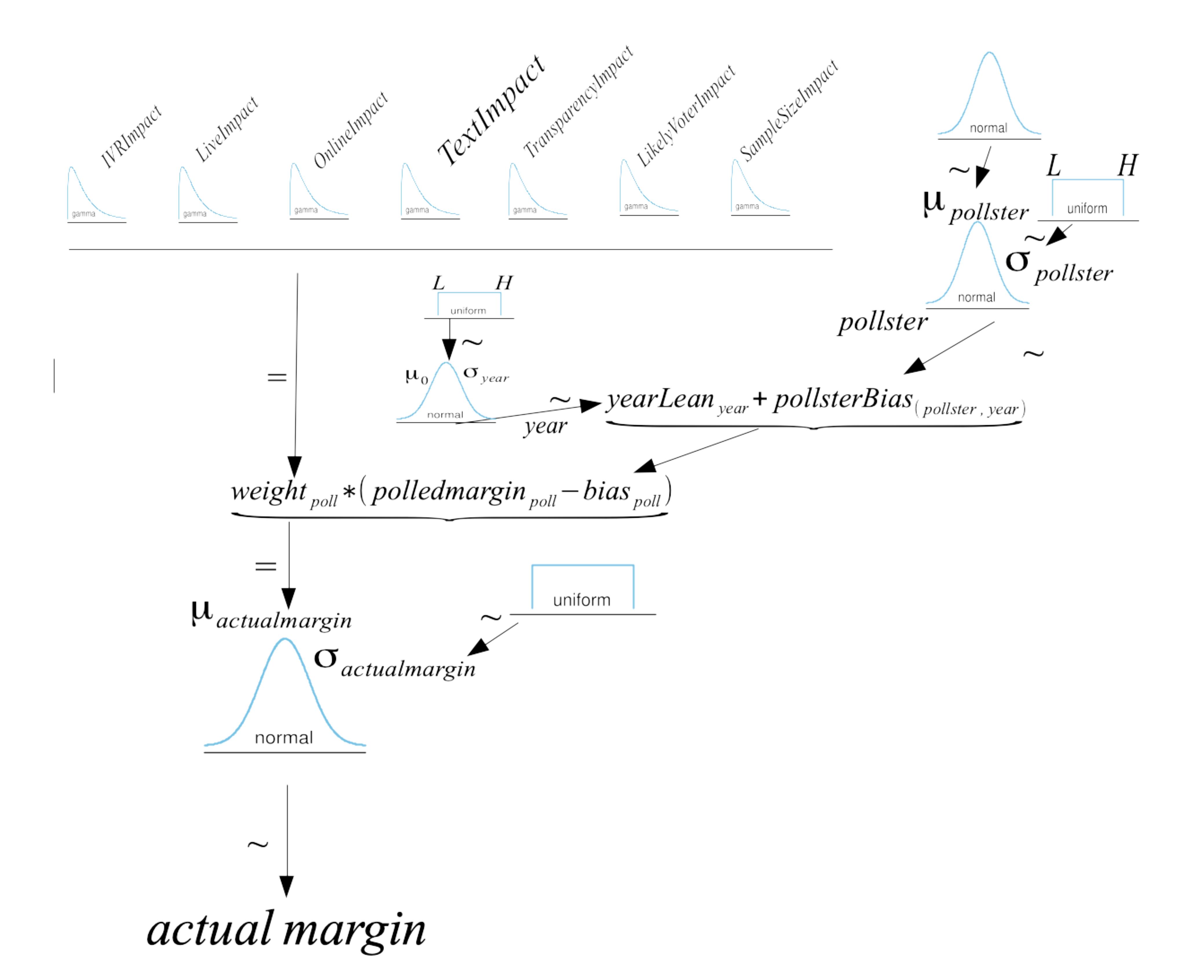
When you see a poll, things that should effect how much credence you give it include the delivery mode, the pollster, the pollster's commitment to transparency, the sample size, and the type of voting model used.



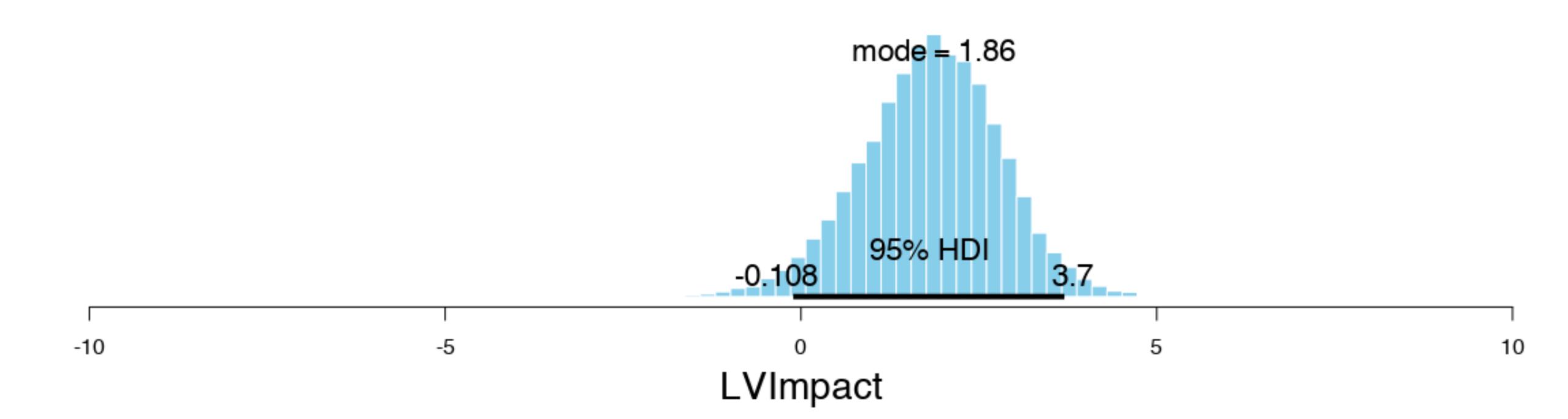
The pollster and the year themselves also produce systematic error in polls that you should be aware of.

Acknowledgements

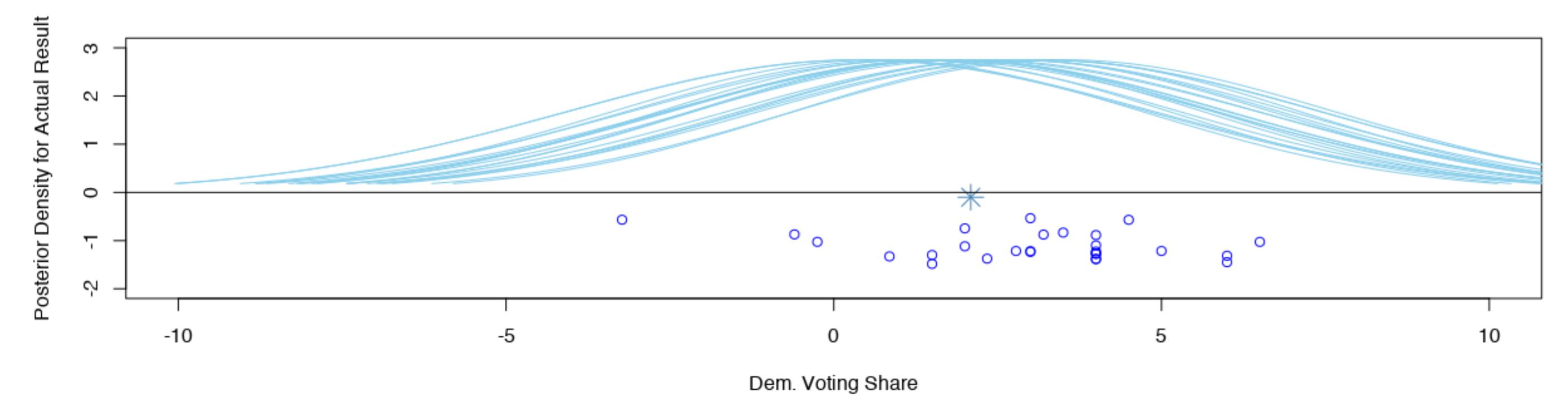
A special thanks goes to Jake Price for being a wonderful research advisor and to the Agricola fund donors for funding my research.



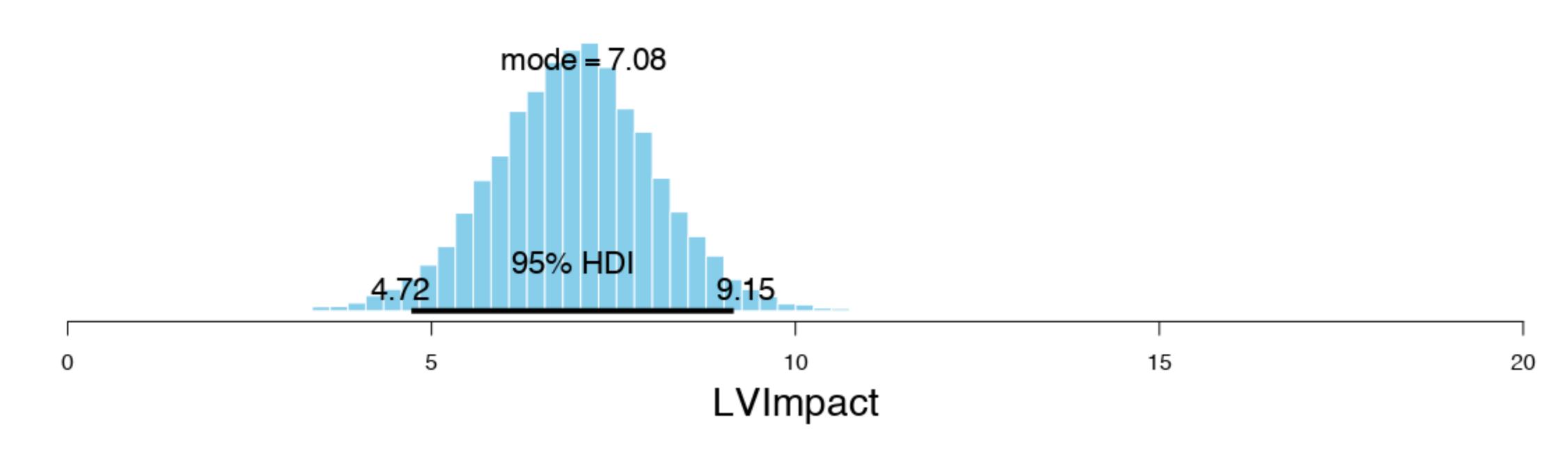




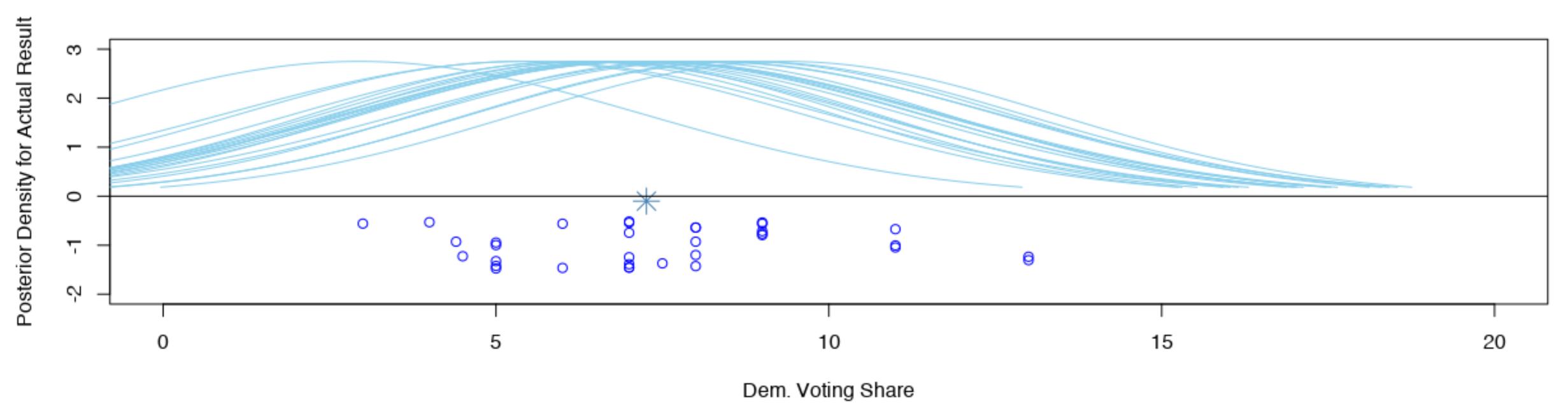
2016 Pres-G US

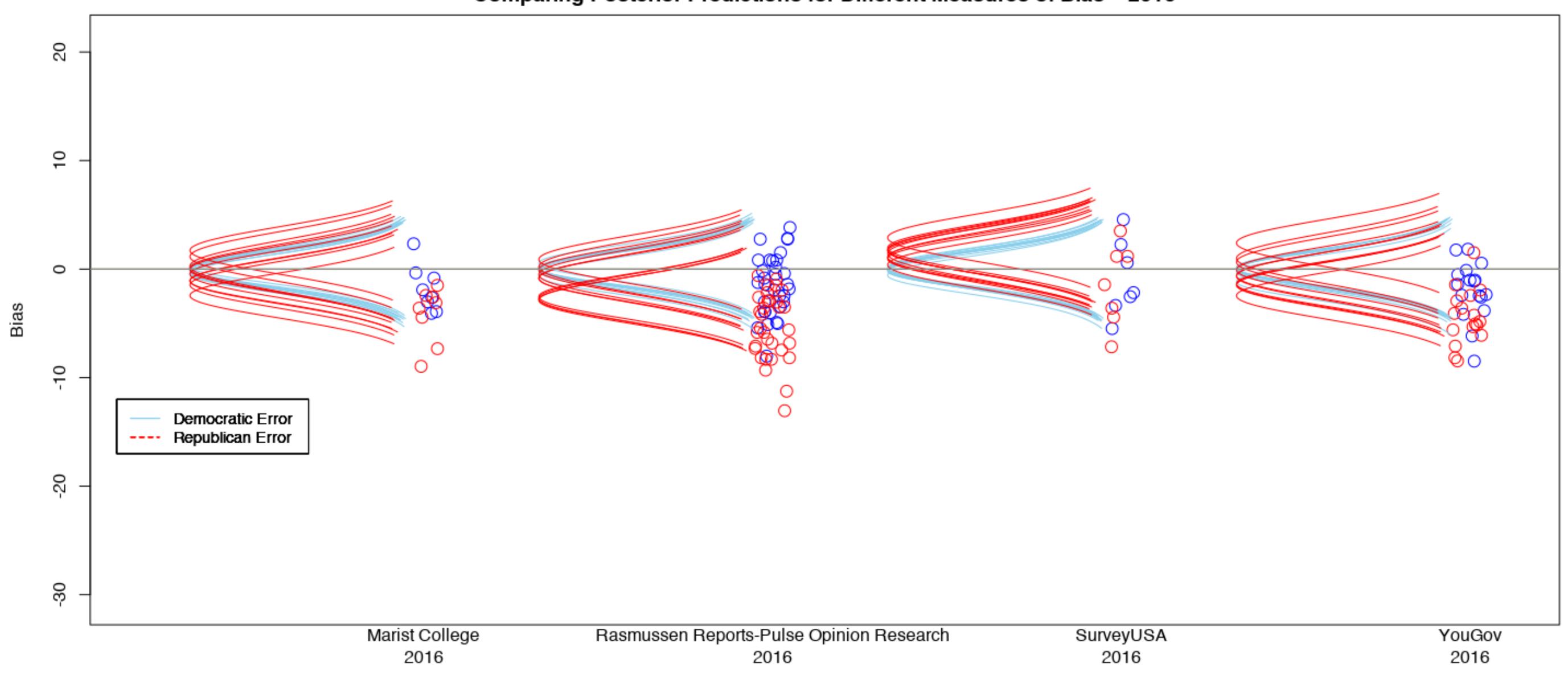


Mean



2008 Pres-G US





Pollster Name & Year

