Predicting elections with Bayesian "Mister P" (MRP)

MRP is "Multi-level regression and post-stratification"

Andrew Enfield, Data 512 Autumn 2017

Our samples are biased, and we should care

We usually work with samples

All samples are biased/not representative (or at least most are)

Biased samples can lead to incorrect results, under-representation of disadvantaged groups, bad things

So, adjust the sample to be more representative of the population

  Not a magic bullet, but it can help

Some options, for example

  Post-stratification

  Propensity score matching

Post-stratification overview, 2016 election

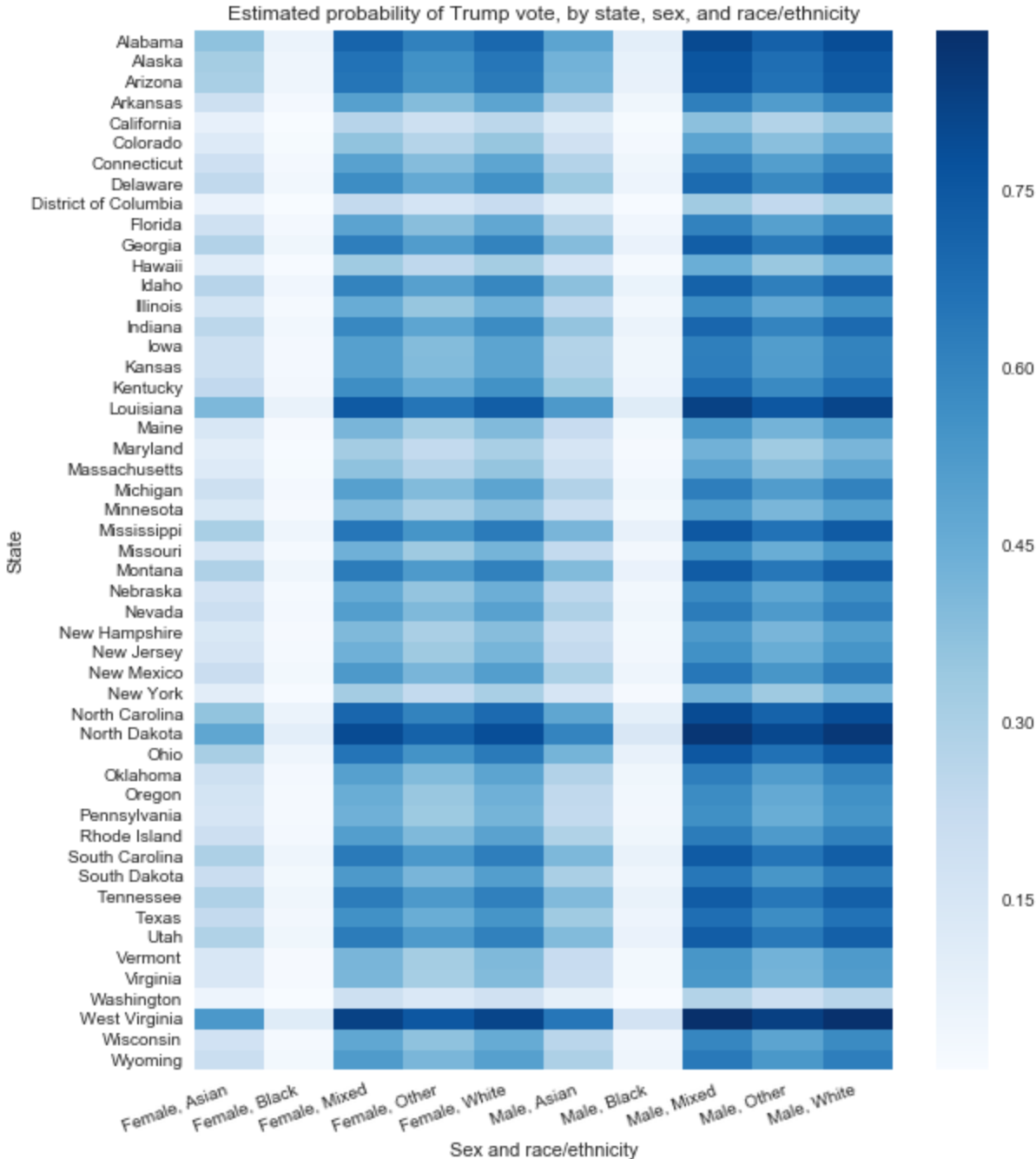|	
1. Divide population into "cells", one for each combination of values in the data	Sex, race/ethnicity, state 2 * 5 * 51 = 510 cells
2. **For each cell**, estimate contribution from that cell using the:	
Proportion voting for Trump, based on the sample	**Data** Pew Research Center poll data, 10/2016 **Method** Bayesian multi-level regression; could use simple logistic regression, others
Size of population	**Data** US Census "Current Population Survey", 11/2016
3. Final estimate is an average of all cells, weighted by size of the population in each cell	**Method** Post-stratification

Probability of voting for Trump, by cell

Notes

Sex and race categories defined by Pew Research Center poll data

Estimated probability of Trump vote, by state, sex, and race/ethnicity

It didn't work as well as I'd hoped it would

Ideal: MRP applied to unweighted raw data matches Pew results based on sophisticated weighting

Actual

   MRP results are different – 43.3% vs 46.4%

   MRP results no better than simple logistic regression

Possible improvements, for example

   Additional variables: age, income, region, educational level…

Method	Clinton	Trump
(baseline) Pew Research Center, weighted	53.6%	46.4%
Bayesian MRP	56.7%	43.3%
Bayesian MRP (adjusted for likelihood of voting)	56.8%	43.2%
Logistic regression (OLS)	56.5%	43.5%
Pew Research Center, unweighted	53.1%	46.9%

Thanks

Data, code, references, details at https://github.com/aenfield/Data512