
1992 US Presidential Election

FoDS S2019
Elina Lahtinen, Veikko Haakana

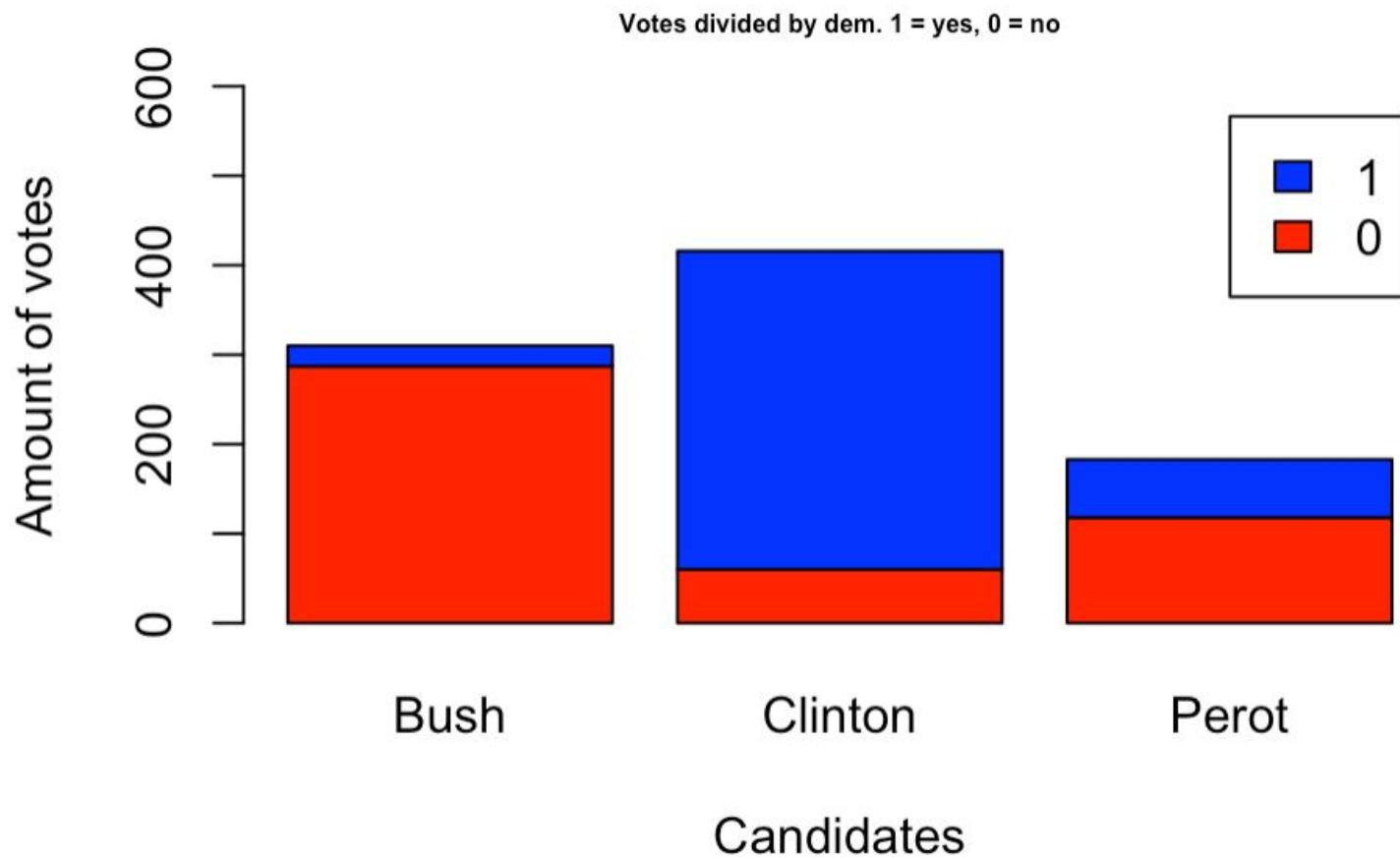
Data set: 1992 US Presidential election

- Task: Predict whether respondent voted for Clinton
- Data set: 9 variables and 909 observations
 - vote, female, dem, rep, persfinance, natlecon, bushdis, clintondis, perotdis.
- 6 new variables
 - vote1, cr.bushdis, cr.clintondis, cr.perotdis, voteclinton, other

Summary of data problems

- Outliers and skewed distribution in bushdis, clintondis, perotdis variables
 - Treated with cube root and replaced outliers with mean +1 standard deviations
- No missing values, no empty cells

clintondis	bushdis	perotdis
Min. : 0.0004	Min. : 0.1024	Min. : 0.2401
1st Qu.: 0.9604	1st Qu.: 0.4624	1st Qu.: 0.2401
Median : 1.0404	Median : 1.7424	Median : 2.2201
Mean : 3.5777	Mean : 3.2201	Mean : 2.0937
3rd Qu.: 4.0804	3rd Qu.: 5.3824	3rd Qu.: 2.2801
Max. :16.1600	Max. :18.6620	Max. :12.1800



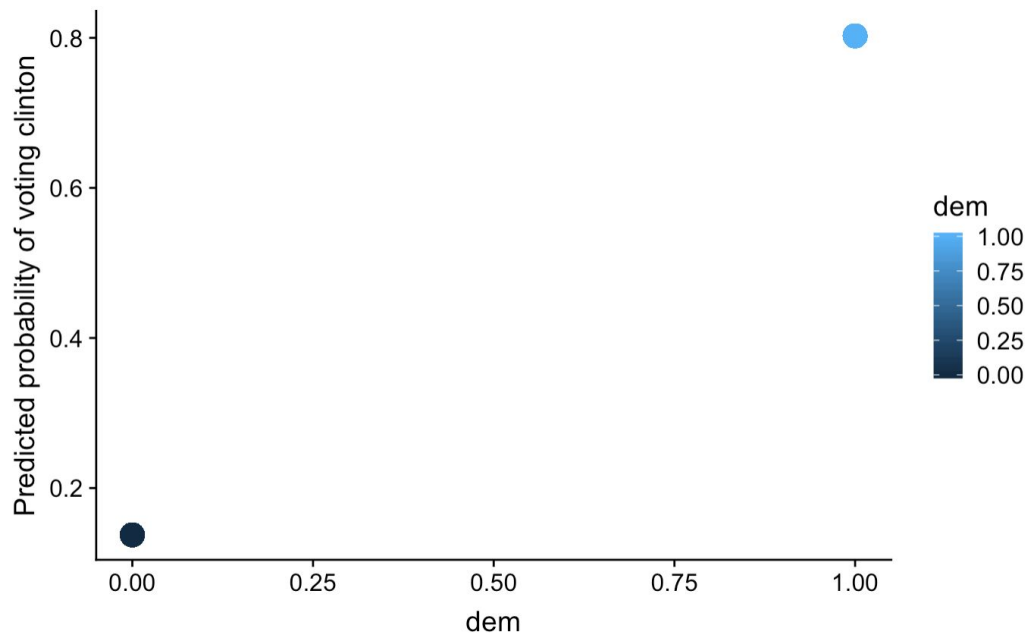
Models

- Three different logistic regression models
 - First model: using only dem as predictor
 - Second model: dem, natlecon, cr.clintondis as predictors
 - Third model: female, dem, rep, other, cr.clintondis, cr.bushdis, cr.perotdis, natlecon, persfinance as predictors

Model 1

- Using only dem as predictor
- AUC: 0.849
- RMSE: 0.358
- ACC: 0.846

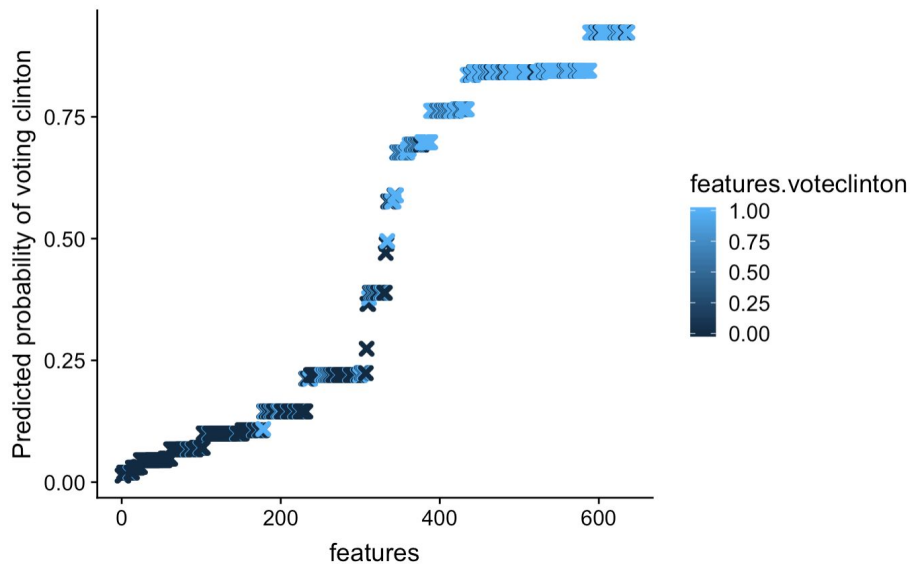
	reals	
preds	0	1
0	123	15
1	27	108



Model 2

- Using dem, natlecon and cr.clintondis as predictors
- AUC: 0.887
- RMSE: 0.35
- ACC: 0.85

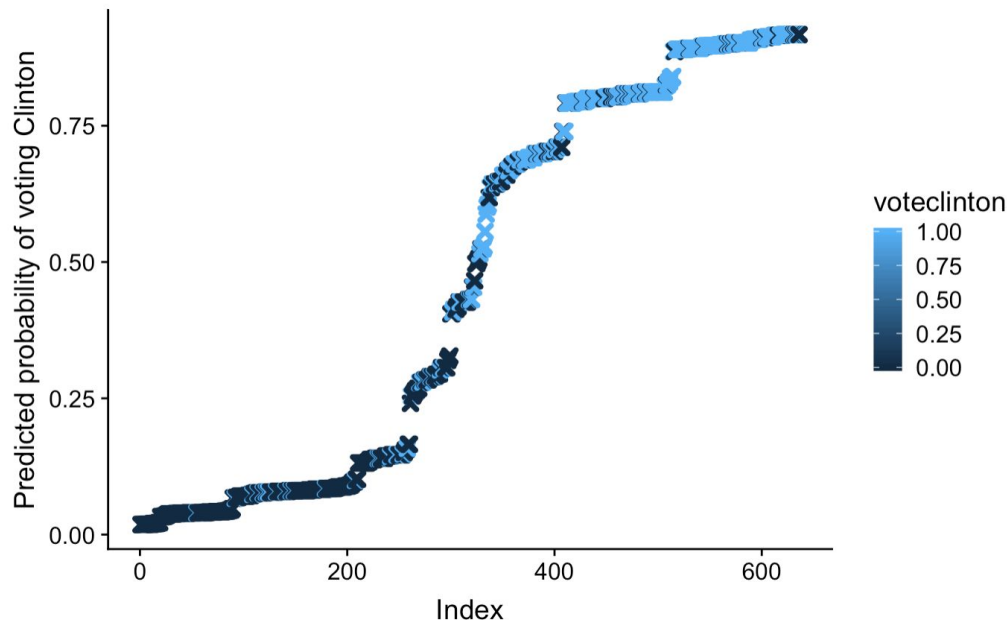
	reals	
preds	0	1
0	127	18
1	23	105



Model 3

- Using dem, rep, other, cr.clintondis, cr.bushdis, cr.perotdis, female, natlecon and persfinance as predictors
- AUC: 0.896
- RMSE: 0.337
- ACC: 0.865

	reals	
preds	0	1
0	126	13
1	24	110



Thank you!