The Effects of Voter Demographics on the 2016 Primary Elections

Alm, Jordan; Arriaza, Laura; Ejaz, Nida; Hunt, Carson; Loeblich, Aithne and Wain, Cydney
Donatello, RA, DrPH Department of Math and Statistics, California State University Chico

Introduction

Primary elections are preliminary elections held by each political party to select a candidate for the presidency. Both Republican and Democratic parties hold their own individual primary in each state. These primary results are used in culmination with the results from all states to determine party presidential nominees.

The purpose of this investigation is to determine if there is a relationship between the candidates that received the most votes in each county and the demographics of that specific county.

Methods

To understand and analyze voter demographic statistics, we took 2016 election data from Kaggle[1].They provided combined datasets of 2016 primary election data taken from several sources including: US Census Bureau data on all counties and primary election results reported from CNN.

We transformed our data sets using R^2 in RStudio[3], a programming language and software environment for statistical computing and graphics. We subset our total county data into just data from California and Texas, combining demographics and the majority winners of the primary. We then performed logistical regression models, univariate and bivariate analyses to create graphics, tables and charts in order to get actionable results in which we could draw our conclusions.

Univariate Analysis

- Univariate analysis indicated normal distribution of variables. Summary Table displayed at right.

Bivariate Analysis

- Higher Education: Counties with high percentage of individuals with a Bachelor's degree or higher tend to vote more for candidate Bernie Sanders. The mean of Bernie voters well above the maximum for each of the other three candidates.

- Hispanic/Latino Vote: Counties with high percentage of Hispanic individuals tend to vote more for candidate Hillary Clinton, both her mean and minimum are well above the maximum of the other three candidates.

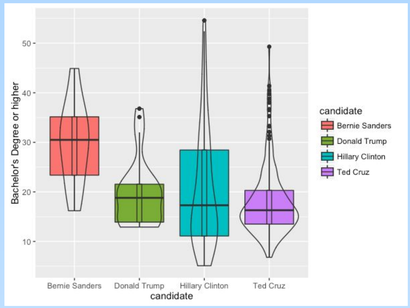
Regression Analysis

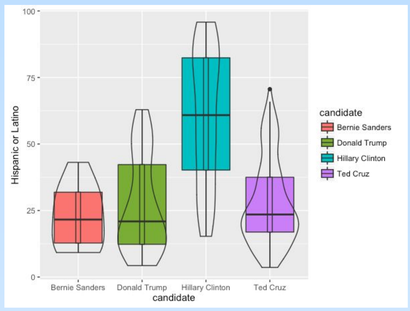
- Higher Education/Per Capita Income: There is a positive trendline for every candidate, with a noticeably large increase at higher per capita income and education for Hillary Clinton. (P-value < .0001. R^2 = .60) 60% of the variation in income can be explained by the regression line equation. Y= -6.912 + 1100x

- Hispanic/Poverty: There is a positive trendline for Hillary Clinton and Donald Trump. The trend for Ted Cruz is increasing and then decreases. The trend for Bernie Sanders decreases. (P-value < .0001. R^2 = .22) Only 22% of the variation in the percentage of Hispanics in the counties can be explained by the regression line equation. Y= 4.06 + 1.69x

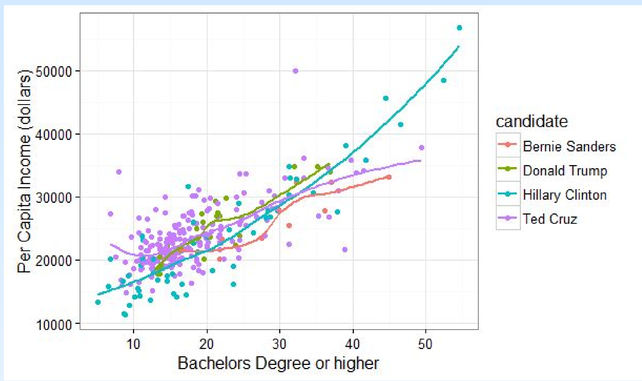
Results

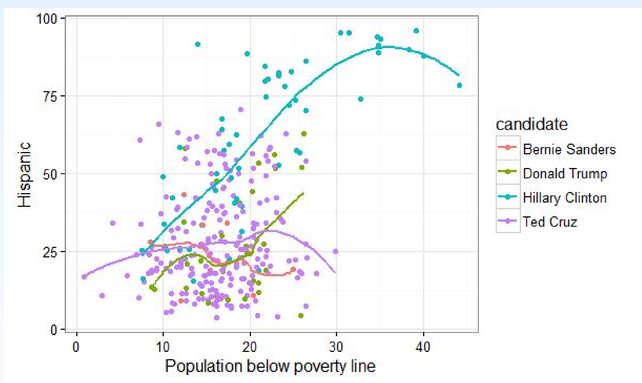
DataSet	Min	1st Qu.	Median	Mean	3rd Qu.	Max	STDEV
His./Lat.	3.60	17.40	26.20	33.34	47.63	95.80	21.89
Education	5.10	13.50	16.65	19.09	0.23	54.60	8.42



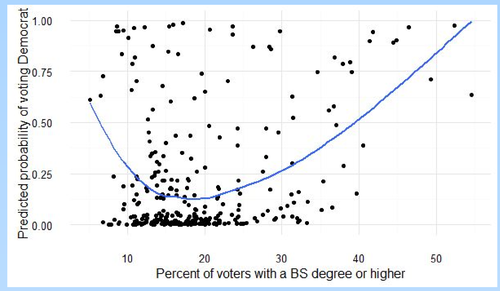


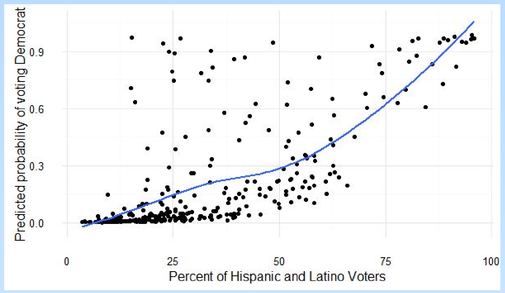
Regression Results





Multiple Logistic Regression





Logistic Regression Results

- The predicted probability for a county to vote democrat in the primary election increased as the proportion of college educated or hispanic and latino voters increased.
- The most significant increase in the predicted probability of a county to vote democrat by proportion of college educated voters happens beyond 30%
- The correlation between college education and probability to vote democrat was 0.26.
- The proportion of Hispanic and Latino voters and the probability of that county voting democrat showed a fairly strong, positive, linear relationship with a correlation of 0.70.

Conclusion

Based on our results of the comparison between Hispanic/Latino vote and percentage of individuals with a Bachelor's degree or higher, we can see two obvious trends. Counties with higher percentage of Hispanic/Latinos and higher educated individuals tended to vote Democratic. However, Hispanic/Latino's more often voted Hillary while individuals with Bachelor's degrees tended to vote for Bernie Sanders.

We can see when comparing Per Capita Income of counties with the percent of individuals with Bachelor's degree or higher, there is a significant positive relationship with voters. As the education levels increase, the yearly income increases (p<.0001). When comparing the percentage of Hispanic/Latino voters and the percentage of those below the poverty line there is a positive but not significant relationship (p<.0001).

Based on California and Texas county data, the predicted probability of a county to vote democrat increases linearly with respect to percentage of Hispanic and Latino voters and parabolically with respect to percentage of college educated voters. The correlation is weaker for college educated voters than with hispanic and latino voters with r-values of 0.26 and 0.70 respectively.

References

1. Kaggle Inc., Website. [Updated 2016; cited 2016, Nov 14].
2. R, Programming Language, version 3.3.1. [Updated 2016, Jun 21 cited 2016, Dec 5].
3. RStudio, Integrated development environment for R, version 3.3.1. [Updated 2016; cited 2016, Dec 5].
4. Diez, David; *et. al.*, *OpenIntro Statistics*, 3rd edition. [Updated 2016, Jan 13; cited 2016, Dec 5].
5. Smithsonian, Political Animals Graphics, Website, Smithsonian Institution. [Updated 2012, cited 2016, Dec 5].
6. Chico State Seal, [Updated 2016, cited 2016, Dec 5].