

MEMORANDUM

To: POSC 202A

From: Loren Collingwood

Subject: Final

Date: Due Friday

Open the dataset `sen.csv`, which is stored on the course website. This is a dataset of senate candidates' Latino mobilization in 2010 and 2012.

1. Descriptives:
 - (a) What are the dimensions of the dataset?
 - (b) Which state has the most observations, which the fewest?
 - (c) Which candidate(s) received the most votes in 2010? What about in 2012. Use the `which.max()` function to help answer this question and document it.
2. Combine the two variables `policy` and `website_ads` into one variable. Plot a histogram of this new variable. Include the mean value in the plot either as a line or in the text of the plot. Is the variable approximately normal?
3. Create a scatterplot between anti-immigrant sentiment and citizen voting age population Latino. Include a lowess regression line to show the trend. Is there much of a correlation here? Why would this be the case?
4. Do the same as question number 3 except between anti-immigrant sentiment and Democratic vote average between 2004 and 2008. This time, however, include the state labels instead of points. What does the graph tell us, and what is the correlation between these two variables? Is there face-validity to these data and why?
5. What is the correlation between percent latino and `policy`; `outreach`; and `policy/outreach` variables? Create a correlation matrix of these variables and produce a table. In general, what do these results tell us about the relationships between these variables?
6. Conduct separate regression analyses for `policy_outreach` and `cvap_pct_latino`, `policy` and `cvap_pct_latino`, and `website_ads` and `cvap_pct_latino`. What do these regressions tell us? Be sure to incorporate discussions of the intercept, the coefficient estimate for `cvap_latino`, t-value, and R-squared. Which model is "best" and why?