

## Problem Set 4: Interpreting Univariate Statistics and Plotting Political Variables

For this assignment, I'm asking you to perform some basic analysis of some data from the November 2008 election. The data set for this assignment can be found on the website (right below the assignment listing). There are three versions of the data: Stata, Excel, and text. You will need to read the data into R in a fashion identical to what you did for problem set 2. Once you attach the data in R, please answer the following questions/do the following tasks (this assignment is due Wednesday, Nov. 26 by class time):

The main variables you will be working with are: `obamapercent`; `proportionforprop8`; `proportionforprop4`.

Task 1: In a single plot, create a dot chart of: 1) Obama vote percentage; 2) Support for Proposition 8; 3) support for Proposition 4. Save the plot and turn it in.

Question 1: What are the basic features of the dot charts? How do you interpret? What do you learn about California voting patterns by looking at the dot chart? (10 points)

Task 2: In a single plot, create a histogram for the same variables as in task 1. Save the plot and turn it in.

Question 2: What are the basic features of the histograms? How would you describe the distribution of the data? What is the overall interpretation of the histograms? (10 points)

Task 3: In a single plot, create a box plot for the same variables as in task 1. Save the plot and turn it in.

Question 3: What are the basic features of the box plots? What information is conveyed in these plots? How are they related to the points you made in question 2? (10 points)

Task 4: For the Proposition 4 and Proposition 8 variables, compute: mean, standard deviation, and the five number summary and neatly display it in a table of your own creation.

Question 4: Provide some interpretation of these statistics? What do we learn by inspecting them? How are the descriptive statistics similar or different across the two propositions? (10 points)

Task 5: For one of the ballot propositions (you choose), compute the Z-score for each county and the one-tail  $p$ -value for this Z-score. In a single plot, create a dot chart for the Z-scores and the  $p$ -values.

Question 5: What is the relationship between the value of the  $z$ -score and the value of  $p$ ? Based on an analysis of these quantities, which California counties are the “least” like the “typical” California counties (i.e. which have extreme  $z$ -scores and small  $p$ -values)? What political factors may be related to these counties being “atypical”? (10 points)