Predicting the 2016 US Presidential Election

The U.S. president is elected by the electoral college – 538 electors corresponding to 435 members of congress, 100 senators, and 3 additional electors allocated to Washington D.C.. The number of electoral votes allocated to each state is equal to the size of its congressional delegation. And most states cast all their electoral votes for the candidate receiving a plurality of the state's votes in the general election (the *winner-takes-all* rule). Nebraska and Maine are the only two exceptions. These states allocate two electoral votes to the candidate receiving a plurality of the state's votes, and each of their remaining electoral votes go to the candidate receiving a plurality of votes within each of the states' congressional districts. But these are small and relatively homogeneous states. Maine has never actually split its electoral votes and Nebraska did it only once, casting a vote for Obama in 2008.

A candidate must receive a simple majority of electoral college votes (270 votes) to be elected. But, as we have seen in 2000, it is possible for a candidate to win the election without receiving a plurality of the popular vote. In today's precept we will analyze state-level polls downloaded from the Huffington Post's Pollster (http://elections.huffingtonpost.com/pollster/polls) and 3 additional polls for Washington D.C. available at (http://www.electoral-vote.com/evp2016/Pres/pres_polls.txt) to predict the outcomes of the 2016 presidential election. We will predict the distribution of electoral college votes according to the *winner-takes-all* rule and using only the 3 most recent polls in each state and examine how this distribution changed over time, starting at 90 days before the election.

The dataset we will be using this week (`polls2016.csv`) has 905 observations, each representing a different poll, and includes the following 7 variables:

Name	Description
`id`	Poll ID
`state`	U.S. state where poll was fielded
`Clinton`	The poll's estimated level of support for Hillary Clinton (in percentage points)
`Trump`	The poll's estimated level of support for Donald Trump (in percentage points)
`days_to_election`	Number of days before November 4, 2016.
`electoral_votes`	Number of electoral votes allocated to the state where the poll was fielded (a state-level variable)
`population`	The poll's target population, which may be `Adults`, `Registered Voters`, or `Likely Voters`

Question 1

We will begin by restricting our poll data to the 3 most recent polls in each state and computing the average support for each candidate by state. Create a scatterplot showing support for Clinton vs. support for Trump. Use state abbreviations to plot the results. Briefly interpret the results.

Hint: To do this see the code in Section 4.1.3 of QSS. The only difference is that you will have to sort the polls by the `days_to_election` variable within each state. Use the `sort()` function to sort the polls from the latest to the oldest. When the `index.return` argument is set to `TRUE`, this function will return the ordering index vector, which can be used to extract the 3 most recent polls for each state.

Question 2

Based on the average support you calculated for Clinton and Trump, predict the winner of each state and allocate the corresponding electoral college votes to the predicted winner. While two states, Maine and Nebraska, do not apply the *winner-takes-all* rule to allocate their electoral votes, for the sake of simplicity, we will apply this rule uniformly across these states as well. If the support for the two candidates in a given state is identical, split the state's electoral votes. Who do you predict will win the election? How many electoral college votes do you predict each candidate will receive?

Question 3

Let's examine how our predictions may have differed if we had used only polls based on *likely voters*. Since we have fewer polls that are based on likely voters, for each state compute the average of the most recent poll (based on the `days_to_election` variable) and those conducted within 30 days from it. In addition, assume that Clinton will win Washington DC. How does the result change when compared to Question 2? Repeat the question but this time using the polls based on *registered voters*. Briefly interpret the results.

Question 4

Finally, we examine how poll predictions have changed over the past few weeks. Starting at 60 days before the election, and for each day, repeat the same analysis as the one conducted for Question 1. That is, for each day, we take the 3 latest polls (or fewer if only one or two is available) for each state and compute the average support separately for Clinton and Trump within each state. We then allocate the electoral votes of that state based on its predicted winner. Use a time series plot to present the predicted total number of electoral votes for each candidate. Add the winning line, i.e., an absolute majority of 270 votes, as a horizontal line. Briefly describe the results.

Hint: You may want to create a nested loop, in which an outer loop is for each day, starting 60 days before the election and stopping on the day of the most recent poll, and an inner loop is used to calculate support for each candidate by state using the 3 most recent polls on any given day.