2018 U.S. Midterm Popular Vote Estimation

Matt Klapman

11/30/2018

Introduction

The 2018 United States midterm elections took place on Nov. 4, 2018. The goal of this project was to analyze polling data related to these elections, attempting to estimate the national popular vote and quantify some type of error along with it. In the past, polling has been an important aspect of election forecasting in political science. But, in the 2016 U.S. presidential election, some doubt was cast on the accuracy of polling. This was due to the perceived failure of polls and polling aggregators like the website *FiveThirtyEight* by the media and the populous. While *FiveThirtyEight* pointed out in their 11-part series **The Real Story of 2016** that the error in 2016 was mostly within the realm of normal polling error and the story around the "failure" of the polls centered around a predominant misunderstanding of probability, I thought it would be interesting to take a stab at predicting this year's election using only polls.

I was able to find polling data from the website *Real Clear Politics (RCP)*. *RCP* had been recording and aggregating polls for this year's election from March 2017 until just days before the election. This data was presented as a table on their website, so it took some work to upload into R. There were a total of 309 polls taken in this time span by a range of entities such as news organizations, non-profit research firms, and partisan political organizations. The polls contained information about the polling organization, date of the poll, sample size, sample type, and the results for the Democratic Party and the Republican Party in a "generic ballot" poll. A "generic ballot" poll just asks a surveyee which party they would vote for in the upcomming election, assigning no politician to the question. I did not have access to information about those who chose neither of the parties.

<– ADD SOMETHING HERE –>

Methods

Results

Discussion