Five Point Three Eight

A Shot at Political Analysis with Data, FiveThirtyEight-style

1. The Problem

American primary elections, unlike general elections, are a lot looser in structure and are definitely not uniform across states. This, of course, makes predictions much more difficult, because all the different systems make it hard to assess the attitudes of the electorate. Two interesting aspects of elections that are relevant here are:

a. To figure out how to interpret sentiment through choices
b. How the structure of the choices make the voter feel like they can express themselves accurately by voting (as opposed to staying home on Election Day).

If the choices don't at least have one option that aligns well with the voter, the voter will be unenthusiastic. Election structure, therefore, is a nontrivial issue, both from the standpoint of actually figuring out what people want, and from giving citizens the ability to express what they want.

The flip side of the disorganization that is the primary election process is that we have a wealth of different ways of implementing the two basic principles above, which enables us to compare and contrast different approaches, and hopefully come up with and adopt better systems. [Possible side project: examine causes of low voter turnout, which Americans are famous for; could that (b) above has a lot to do with it?]

2. Possible Clients

State and local governments and official election staff are the obvious clients: they want to make sure that they can get out the vote in their state, not just to vote for national candidates, but also to make sure that the state can push through initiatives, propositions, referendums, and the like for local projects. Campaigns also will have use for this data, so candidates can tailor their campaign strategy to fit the state's particular system. For example, it is well known that Bernie Sanders outperformed Hillary Clinton in caucus states, whereas it was the other way around for states with actual primaries (in this document, "primary" actually will refer to both "caucus" and "primary" electoral processes). The same was true back in 2008, where Barack Obama outperformed in caucus states, and Hillary in primary states. This obviously raises the question of how structure influences results!

3. The Data: Where am I going to get it?

Initial scouting indicates that Huffington Post has a good API for polling data, and it is an officially-sanctioned, open process. They have approximately 24,000 polls available (2431

pages when counting what was possible with the HTTP GET requests), going all the way back to 2007 (which means, for example, we can compare the primary process for Sanders vs. Clinton and Obama vs. Clinton).

4. Initial Approach

The main thing to get off the ground is to figure out how to comb through the data in a systematic manner: I hope the script is able to get more than 10 polls at a time (otherwise there might be some major for loop action going on).

5. Deliverables

Our basic goal is to come up with useful metrics that can determine what "election structure" means (which may be a difficult question). Results will be summarized in a report with charts and comparisons, and be presented via slides.