Monitoring the 2008 Election Visually

Lawrence Mosley, Dianne Cook, Heike Hofmann, Chris Kielion, Barret Schloerke Department of Statistics, Iowa State University

February 23, 2010

Introduction

Every election year we are bombarded with statements from candidates, campaign policy positions and pundit viewpoints, accusations and responses, and polls, polls, polls. Generally, there is a lot of conflicting information. The 2008 election season was as dizzying as usual, but it was different from the previous election years in at least one respect: several web sites made pollsters' data available almost immediately upon its production. The Statistical Graphics Working Group at Iowa State University, consisting of professors and students (both graduate and undergraduate) with an interest in statistical visualization, explored how to visually display this election data with the goal being to help sort through information reported in the media.

Every few days during the election process we extracted data from several web sites on nationwide and state polls, and combined these with map data. There were several main themes we wanted to learn and communicate with the help of charts such as: tracking the national popularity of the candidates, locating strongholds for each party as well as "swing" states that might tip to either party, examining the variation between polls and lastly, predict who would be the likely winner on election day. The results are displayed on the web site www.public.iastate.edu/~dicook/Election/Home.html.

This article describes some of the plots designed to monitor the election, and what we learned during the election process. We will describe the data sources, and how data were merged from different sources to make it available for plotting. Furthermore, we will describe the plots, the choices made in each design, and their purpose. The last section details our observations and ties together what was learned during the election season from these efforts.

Data

Data from polls in each state were extracted from two web sites: www.electoral-vote.com and pollster.com. National tracking poll data were extracted from the latter web site.

General plot design choices

To create our plots, we used the R package ggplot2 written by Hadley Wickham, which offers flexibility and good defaults for statistical graphics. Since displaying data in a clear, coherent, and efficient manner does not occur by happenstance, careful consideration was taken in the design of each graphical display relative to intended information to communicate. For all plots, a light grey background was chosen to offer a backdrop where the graphical elements stand apart from the page. Soft, white color grid lines were used to guide the reader's eyes, but not obstruct the data. Other lines were used to guide but not distract the reader. To give the plot a neat systematic appearance, special attention was given to the order of aspects of the data. Where possible, color was used to elucidate the contrast between the two candidates or pollsters

and the use of shading allowed us to differentiate between levels of intensity, where we could show stronger or more prominent effects. Working with these principles combined with a little creativity, we were able to make graphical displays of the election data that were aesthetically pleasing, informative, and entertaining.

Tracking candidate popularity

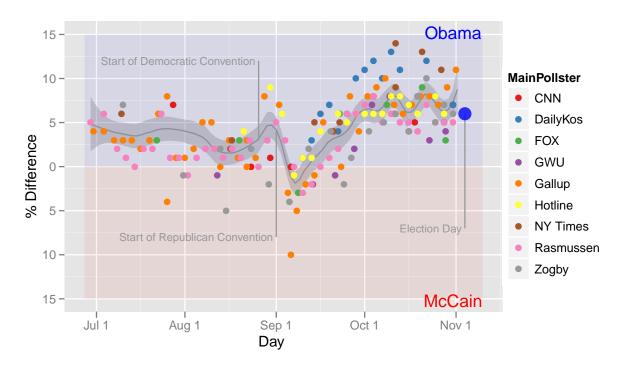


Figure 1: Representation of the tracking polls, monitoring the popular vote. The horizontal axis is day of the year and the vertical axis shows the difference in percentage between Obama and McCain. Each colored dot is one poll result, colored by pollster, and the grey band is a loess smoother showing the trend. The election day result is shown as a large blue dot. Generally, Obama had the lead for most of the election period, and the lead stabilized to about 5-8% in the last month. Obama received a small boost from the Democratic convention, but suffered a big decline in popularity from the start of the Republican convention. There is astonishing variability in the poll results! Some of this is due to pollster bias. Daily Kos tended to over-estimate Obama's lead, although were very close to the mark on their final poll. The New York Times was also overly optimistic about Obama. Rasmussen and Fox tended to be on the low side, favoring McCain. Gallup's results are extremely varied, dipping even down to a 10% lead for McCain at one point. Hotline was fairly close to the average trend after the conventions.

Our first objective was to establish a way to track the popularity of the two main candidates. A time series plot is the natural choice for accomplishing this goal and in Figure 1 we have plotted the percentage difference between Obama and McCain on the Y-axis against the day of the year on the X-axis. The percent difference is simply the difference in percentage points between the votes for Obama and McCain in a given poll. Positive differences correspond to a lead for Obama and are represented as points plotted above the zero line. The converse is true for McCain. Our tracking began in July 2008 and continued until the day of election on November 4th, 2008. Each pollster is color-coded, and a smoother was applied to the data to give a rough estimate of the trend.

From the beginning, then-Senator Obama had a small lead. This lead remained steady until September, the start of the Republican National Convention. The decline of Obama's favorability was abrupt. Gallup even announced that McCain had amassed a ten percentage point advantage, although this conflicts with the other poll results of the time. Unfortunately for McCain, from this point on until election day, McCain failed to maintain his popularity and Obama's lead stabilized in the month leading to the election.

Shifting our focus to the pollsters, we expose a rather curious pattern. There is a lot of variability in poll results, and some of this can be explained by pollster bias. The polls done by the pollster for the Daily Kos, represented by blue points, tend to return the largest percentage difference in favor of Obama, up until the last poll before the election. Conversely, Rasmussen polls, as represented by pink points, nearly always have the smallest percentage difference between the two candidates. Furthermore, the points for each pollster are outside of the (point-wise) confidence bounds of the loess smoother, suggesting there may be a reason behind the disparity between the polls. It is an open secret that the Daily Kos is a left-leaning establishment publishing articles in support of the Democratic agenda, whereas Rasmussen has a more right-leaning bias aiding the Republicans. Both pollsters seem to find and publish results that are more favorable to their supported candidate, an interesting fact indeed as each pollster is trying to paint the rosiest picture possible! For the reader we take the message that Daily Kos results need to be down-sized by a few percentage points, and Rasmussen results need to be up-sized to more closely reflect reality. Some of the other pollsters show a lot of variability. We can see that Gallup's poll results are sometimes way up, sometimes way down - a lot of variation. During the election process there was a lot of discussion about pollsters techniques, and whether the cell-phone-only vote could be accurately predicted. Perhaps Gallup had some problems in their sampling methods related to this, or perhaps they were experimenting with different techniques. One thing that we learned about Gallup is that they only operate on the national level, not in the state electoral vote polls. This surprised our group.

Monitoring the electoral votes

The U.S. presidential election is won by winning the largest number of electoral votes, which might differ from the popular vote. Each state has an allotted number of votes based on its population, for example, Alabama has 9, Iowa 7, Alaska 3, Florida 27, and New York 31. When a candidate wins a state, they take all the electoral votes for that state, with the exception of Maine and Nebraska. These two states have the potential to share the electoral votes depending on the margin between candidates. In the electoral process, states where there is a small margin between the candidates can have a big impact on the election, particularly, if they hold a lot of electoral votes. Monitoring the polls in each state is important for understanding how the results might tip come election day.

The dot plots in Figure 2 show the last two weeks of polls (left) and the final election results for each state (right). We used this type of plot to provide an overview of the states' voting patterns, and identify states that might be considered to be swing states, i.e. ones that might tip the election towards McCain or Obama. These plots are based on the dot plots described in Visualizing Data by William Cleveland. The horizontal axis shows the difference in percent points between votes for either candidates. The vertical axis displays the states, ordered from the most favorable for McCain at the top down to the most favorable for Obama at the bottom.

In the plot of the polls (left) multiple poll results in a state are represented by large grey dots, and the median of the polls is shown as a smaller black point. The yellow vertical bar indicates results less than 5% in difference between candidate, revealing states that might swing the election. States considered in doubt tend to also have more poll results. States that were considered safer had far fewer polls taken, and actually the result which is the most recent might have been much earlier than the final two weeks. Washington, DC, was the most favorable for Obama, with a huge difference, along with other Democratic strong-holds of Hawaii and some of the North-Eastern states. Republican strong-holds included Oklahoma, Wyoming, and generally the big sky and Southern states. States that might have been considered to be swing states from this display include Montana, Arizona, Louisiana, Georgia, North Dakota, Missouri, Indiana, and North Carolina. Florida, Ohio, Virginia, Nevada, and, Colorado were trending more assuredly for Obama at this

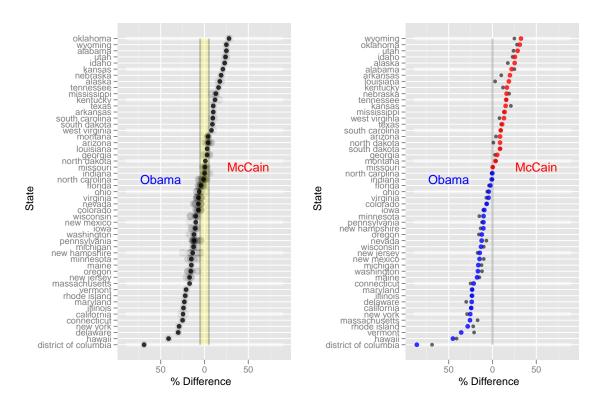


Figure 2: Dot plots of poll (left) and election (right) results in each state, ordered top to bottom from most Republican to most Democratic. This is the monitoring plot that we used throughout the election period displaying two weeks of polls in each state. Here, the two weeks leading up to the election are shown. Dark points indicate the median, light grey transparent points represent all poll results. The yellow region indicates the 5% margin – poll results within this region indicate swing states, such as Missouri and Indiana. States in the middle of the display also tended to have a lot more polls taken, obviously due to the desire to get a good read on the potential for the state to swing either way. The majority of states were tending to be in Obama's favor. The plot on the right shows final results, colored according to party, along with the very last poll results for each state shown as grey. The order is different from those of the polls, as the final results were discordant with the polls for some states.

time. Some states that have been swing states in previous elections were polling strongly for Obama, for example, Michigan, Pennsylvania, and Iowa. The majority of states were tending to be in Obama's favor.

In the plot of the final election (right), results are shown as red or blue, indicating the outcome. The median of the last poll results is shown in black. There was an over 80% difference between Obama and McCain in Washington, DC, showing overwhelming support for Obama in the nation's capital. Most of the poll results were on target, except for a few states. States where the polls missed the mark include Louisiana, North Dakota, and Arizona. These states were in the media in the week leading up to the election with "shocking" news that they might finish in favor of Obama, but these news reports clearly arose from a spurious poll in each state. Louisiana, though, might be explained as suffering from a Bradley Effect. The Bradley effect is a hypothesis proposing that voters feel pressured by political correctness to tell pollsters that they are undecided or likely to vote for a black candidate, yet on election day, underlying racial bias wins and the vote goes to the white opponent. This effect is often very difficult to verify and many analysts even deny its existence. Yet, when a study is properly done on a representative sample and the final results are so much out of line with recent polls the integrity of the poll is suspicious. Other states, where it looks like the polls missed the mark including Vermont, Massachusetts, and Wyoming, are better explained by old poll results – at some point pollsters stop polling in states where the final results are very predictable.

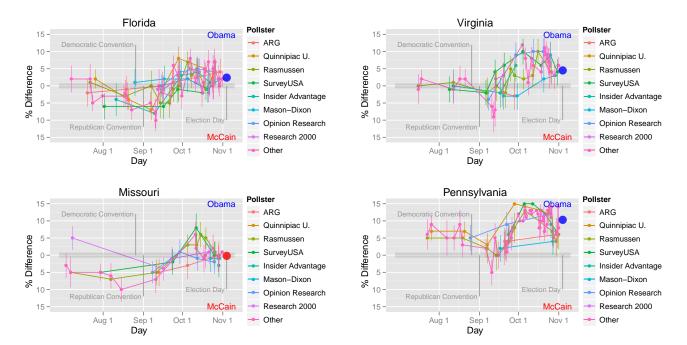


Figure 3: Poll results in four of the states. In each plot the horizontal axis is day of the year, and the vertical axis is difference between the percentage of votes for Obama and McCain. Color indicates pollster. Not all pollsters operate in every state. The final election result is included as a dot in the victor's color. The vertical lines displays the reported margin of error for the poll - when they cross the 0 line it indicates that no clear winner can be predicted. Virginia and Pennsylvania were fairly safely cast into Obama's camp. Florida had a tight race, but Obama had a consistent small advantage leading up to election day. Missouri was a real cliff-hanger!

Figure 3 shows the results of all polls between the conventions and election day, for four different states. Here, we show the plots for Florida, Virginia, Missouri, and Pennsylvania. We chose this sample of states because a lot of polls were being conducted and the polls exhibit interesting patterns about the trends in these states. The vertical axis shows the difference in percent points between votes for either candidate. Pollsters are, again, represented by color, and polls from the same pollster are connected by a line. The

vertical lines display the reported margin of error for each poll. The large red or blue dots indicate the actual election result. These plots were created to examine trends, while accounting for the differences between pollsters. Florida, Missouri, and Virginia were considered by many pundits to be swing states. Pennsylvania, for this election, appeared to produce a safe Democrat outcome. However, McCain poured substantial funds into advertising in the last month hoping to swing it back his way, and we can see that this had some effect. Obama's final percentage in Pennsylvania was actually higher than polls predicted. Although Florida and Virginia were consistently polling in Obama's favor, in the weeks before the election, the margins of error covered the 0 difference, which indicates a possibility that the final result might have been for either candidate. Missouri was a real nail-biter and the final outcome wasn't announced until days after election day.

Some pollsters' bias is also visible here: Mason-Dixon (light blue) tended to low-ball the difference, favoring McCain. Quinnipiac (puce) tended to over-estimate the difference, in favor of Obama. Rasmussen (khaki) tended to be more on target, which is different from their national tracking polls' bias.

Building up to an election victory

The "tower plot" in Figure 4 shows the candidates' total support as area. It is inspired by a graphic Building an Electoral Victory in The New York Times used for the 2000 election. The tower plot stacks blocks for each candidate, with heights proportional to a state's electoral votes, building towards the critical 270 electoral votes needed to win the election. Like an actual tower, the foundation is created at the location where each candidate has the most support with the width of each block displaying the percent margin between the two candidates in that state. The top of each tower holds the states with smaller support for each candidate. Color is also, redundantly used to indicate support. This tower illustrates the election results, showing that Obama won substantially more electoral votes than McCain, 365 to 173. Throughout the election process, we used this plot to gauge the growth of each candidates support. States were only added to the tower if the margin was greater than 10%. Obama consistently had a higher tower than McCain.

Mapping the political landscape

A traditional way to display poll and election results is to use a chloropleth map, with color coding the percentage margin in each state (top plot of Figure 5). It exposes the "mood" of the electorate – the more red the country, the more Republican the leaning. A problem with this display is that the small area of some states hides their contribution to the electoral count. In the map view the race looks tight: there is as much red as blue. Some states with large geographic area, dominating the map, have small populations and thus contribute a smaller amount to the election count. In comparison, DC barely registers on the map compared to Wyoming. Both regions have three electoral votes, though.

A cartogram (bottom plot of Figure 5) shows a more faithful rendering of the mood of the country. The rough geographic position of the state is maintained but the size of the state is re-scaled to match its number of electoral votes, for example, Rhode Island increases in size while Montana shrinks dramatically. (Information about cartograms can be found in cartography books such as *Map Use and Analysis*.) We now see how much of the electorate is in favor of the Democratic party and can safely infer that the nation as a whole is leaning more to the left.

During the election process we produced maps and cartograms of the percentage difference between candidates. When there were multiple polls available in a state, several displays were created: the median difference, the most favorable to McCain, and the most favorable to Obama. At times these extremes were far apart, showing the large variance in poll results in key states, and the extremes of potential final results.

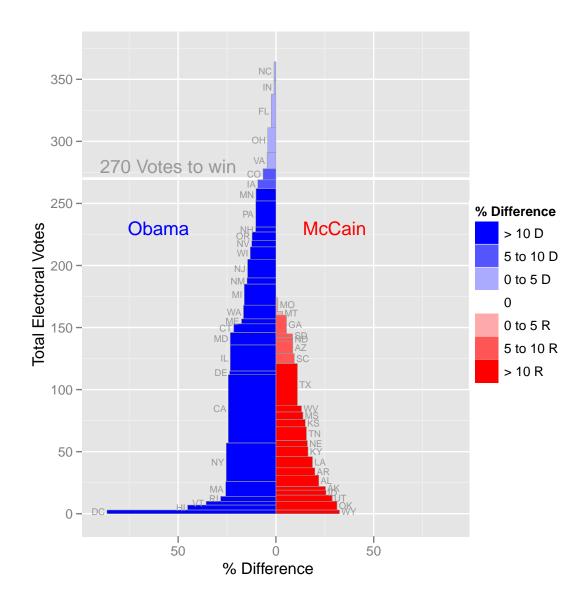


Figure 4: A New York Times inspired electoral votes' tower showing the state-by-state support for each candidate. The horizontal axis displays the difference in percentage points for the two candidates and the vertical axis displays the electoral votes of each state. States with the most support are at the base and color represents percentage difference. To win the election a candidate needs 270 votes. This tower shows the final election results. Obama amassed 365 electoral votes across 28 states and territories, while McCain came up much shorter at 173 votes across 22 states.

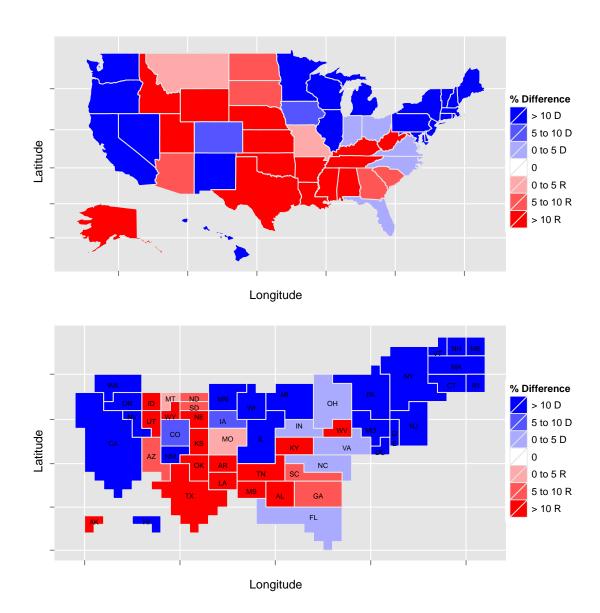


Figure 5: Electoral map and cartogram displaying the election results. States are colored according to difference in % between the two candidates. These displays show the geographic distribution of the vote. The map on the top is the traditional chloropleth display, but the presence of small areas, such as D.C. and Rhode Island, make it difficult to see the real lay of the land. The cartogram, in which the map has been warped so that the area of a state corresponds to the number of its electoral votes makes area-wise smaller states' contribution more explicit. On the traditional map red is more prominent, tempting us to interpret the election results as closer than it was.

Tying it all together

Our approach to monitoring the election differed substantially from that of other election websites. For example, www. vethirtyeight.com aimed to accurately predict the election results. It provided a wealth of information on different aspects of the electoral process, very brilliant observations, and accurate predictions. A lot of information, such as the algorithms used to make predictions, however, was kept private. Our purpose was different. Beyond using this as an opportunity to experiment with different graphical summaries, the main purpose was to enable the web site visitor to study the variation in publicly available data, and to arrive at their own conclusions. These plots also helped members of our group to digest, and be amused by, the media reports of the election.

Some of the main things about the pollsters and the election process that we learned are as follows:

- There is a lot of variation in poll results. Some pollsters have a consistent bias: Daily Kos overestimated the Democratic vote, Rasmussen (in the popularity data) was biased towards Republicans. Interestingly, both converge to be near the final count close to election day.
- The swing states for this election were Missouri, North Carolina, Indiana, Florida, Ohio, Virginia, Montana, and Georgia. Despite being in the news leading up to the election, Pennsylvania and Michigan were not battle ground states and were firmly Democratic in 2008. This was obvious in the state polling data early, so it was possible to tune out news to the contrary.
- The District of Columbia is a Democratic strong-hold.
- The tracking (popularity) data is not the same as election data. Although the time series plot gives a good sense of the general electoral mood, the tower plot gives a more accurate picture of the likely outcome. Throughout the pre-election season the tower for Obama, built from the states firmly in his camp, was consistently higher than for McCain.
- The map showed the geographic distribution of voting tendencies, the strong Democratic Northeast and West, and the Republican South and mid-section.
- In the final days before the election the news media dropped some shockers: Louisiana, Montana, North Dakota, and Arizona were in danger of going Democratic. The final results show a different story. Louisiana was solidly for McCain and polls for North Dakota and Arizona were clearly inaccurate. One problem for these states is that there were few polls conducted close to the election, so the news stories must have been based on a single maverick poll result.
- About a month before election day, the local news in Iowa proposed that the race was closing in, and McCain could win Iowa. This was not supported by the bulk of the data, so that particular piece of news was easy to ignore. It, too, was based purely on a single spurious poll result.

Overall, we were quite happy with the plots that were designed to monitor the election, and plan to do the same in 2012. We will make a major implementation improvement by automating the production of the graphics and the website.

Additional Resources

Campbell, J. (2001) Map Use and Analysis, McGraw-Hill, New York.

Cleveland, W.S. (1993) Visualizing Data, Hobart Press.

Ericson, M. (2005) Building an Electoral Victory, New York Times, Personal Communication.

R Development Core Team (2006) R: A Language and Environment for Statistical Computing, www.R-project.org.

Wickham, H. (2009) ggplot2: Elegant Graphics for Data Analysis, Springer.