Breaking down the 2008 vote[1]

Andrew Gelman[2]

22 April 2010[3]

As political scientists, we typically try to understand public opinion and voting by using least squares, logistic regression, and other statistical tools for estimating the effects (predictive if not always causal) of demographic factors (such as age, sex, income, education), political affiliation and ideology, and aggregate variables (most importantly, economic trends and presidential approval). We sometimes look at exit poll results or at color-coded maps of national elections but typically consider such descriptive summaries as amusements for the newspapers rather than as adjuncts to serious research.

Recently, however, there has been a renaissance in descriptive analysis of electoral data, accompanying the widespread availability of large datasets, free high-quality statistical graphics software, and outlets for rapid dissemination of color graphics on the internet. In many ways, this new wave of political graphics harkens back to the classic descriptive work of political scientists such as V.O. Key in the middle of the twentieth century. Along with this, the work of technologically-savvy outsiders such as Nate Silver (www.fivethirtyeight.com) has bridged the gap between journalism and political science.

In this article we give some examples from our recent analyses of polling and electoral data from the 2008 presidential election (Gelman et al., 2009, Gelman and Ghitza, 2010, Gelman, Lee, and Ghitza, 2010, Gelman and Su, 2010). We present two displays: a grid of maps and a grid of line plots, each summarizing a different breakdown of the vote in the general election. The maps show vote by ethnicity, income, and state; the line plots show vote by education, within categories defined by age and ethnic groups.

The two displays address different aspects of class-based voting. The maps dramatically reveal the different national voting patterns of rich, middle-income, and poor Americans. Separate rows of maps for each ethnic group make clear the distinctive voting patterns of poor whites, who were somewhat Democratic-leaning in most of the country but strongly Republican in the deep south. African-Americans and Latinos, in contrast, showed little variation by state of residence or income level.

The line plots expand upon the well-known pattern in recent years that Democrats do best among the least and most-educated voters, with Republicans being most successful with the voters in the middle. (This is *not* the same as the pattern with income; richer voters consistently vote more Republican.) Education is highly correlated with ethnicity and also with age, hence the separate

[1] For *Atlas of the 2008 Election*, ed. Stan Brunn. We thank the National Science Foundation , Department of Energy, and Institute for Education Sciences for financial support of this work and the Pew Research Center for their survey data. Further graphs of a similar nature appear at our blog: http://www.stat.columbia.edu/~cook/movabletype/archives/political-science/

[2] Department of Statistics and Department of Political Science, Columbia University, New York. gelman@stat.columbia.edu

[3] The original grid of maps had some errors. We inserted a corrected version on 12 June 2011.

plots for each category, which reveal, among other things, that the Obama vote was strongly correlated with education among the young but not among older voters.

We now briefly discuss some of the choices involved in the statistical modeling and graphical display. For the maps we used hierarchical Bayesian inference to get stable estimates for all states and categories, whereas the line plots were simple enough that we worked with simple weighted means, using error bars to indicate the large uncertainties for some of the smaller groups. We used the technique of small multiples (Bertin, 1967, Tufte, 1990) to allow a large number of quick comparisons. For the maps, we used a continuous blue-to-gray-to-red color scale to display estimated vote proportions; for the line plots we put all graphs on a common scale for ease of interpretation.

The two displays shown here certainly do not represent the future or even the current state of the art in political mapping or analysis. What they do indicate, we hope, is the way in which current technology allows us to prepare simple maps and graphs to directly summarize important aspects of demographic and geographic variation in public opinion and voting.

References

Bertin, J. (1967, 1983). *Semiology of Graphics*. Translated by W. J. Berg. Madison: University of Wisconsin Press.

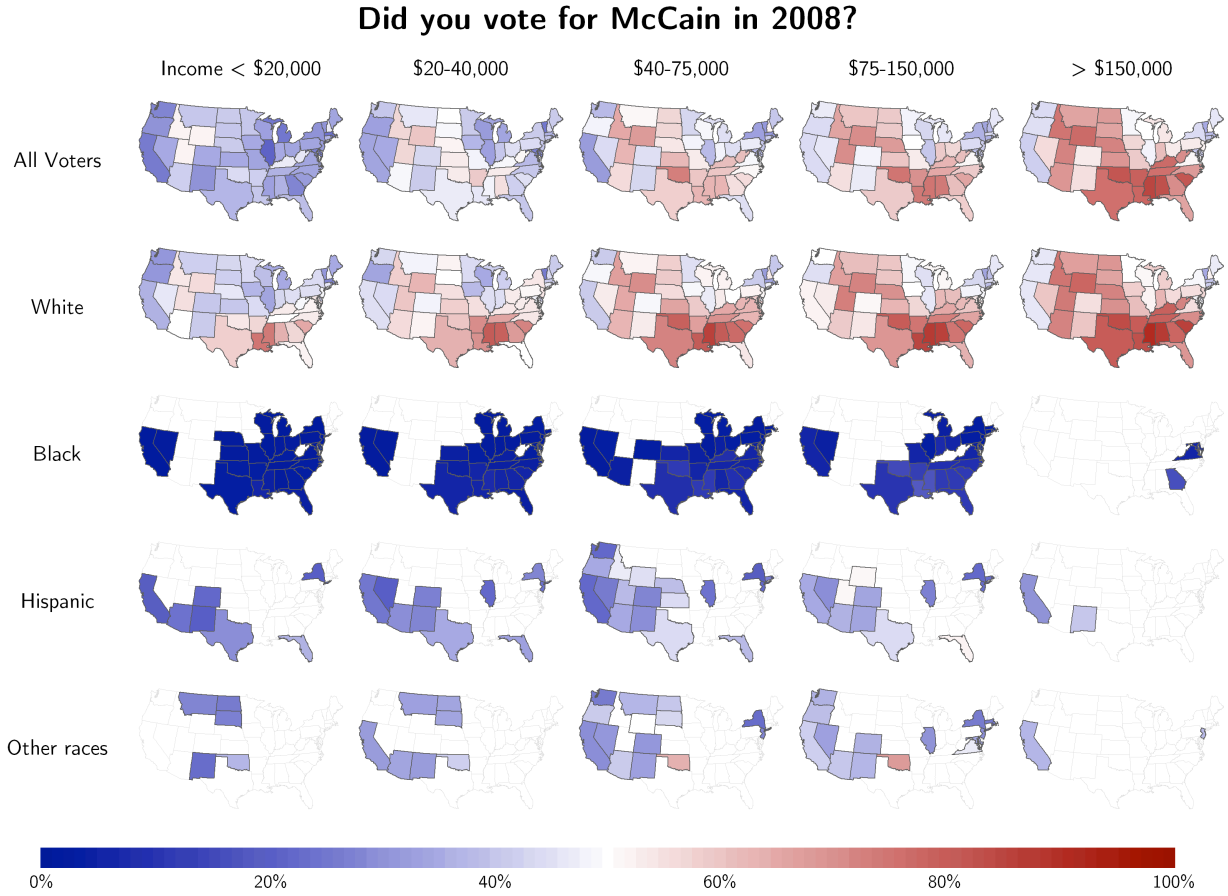
Gelman, A., and Ghitza, Y. (2010). Who votes? How did they vote? And what were they thinking? Technical report, Department of Political Science, Columbia University.

Gelman, A., Lee, D., and Ghitza, Y. (2010). A snapshot of the 2008 election. *Statistics, Politics, and Policy*, to appear.

Gelman, A., Park, D., Shor, B., and Cortina, J. (2009). *Red State, Blue State, Rich State, Poor State: Why Americans Vote the Way They Do,* second edition. Princeton University Press.

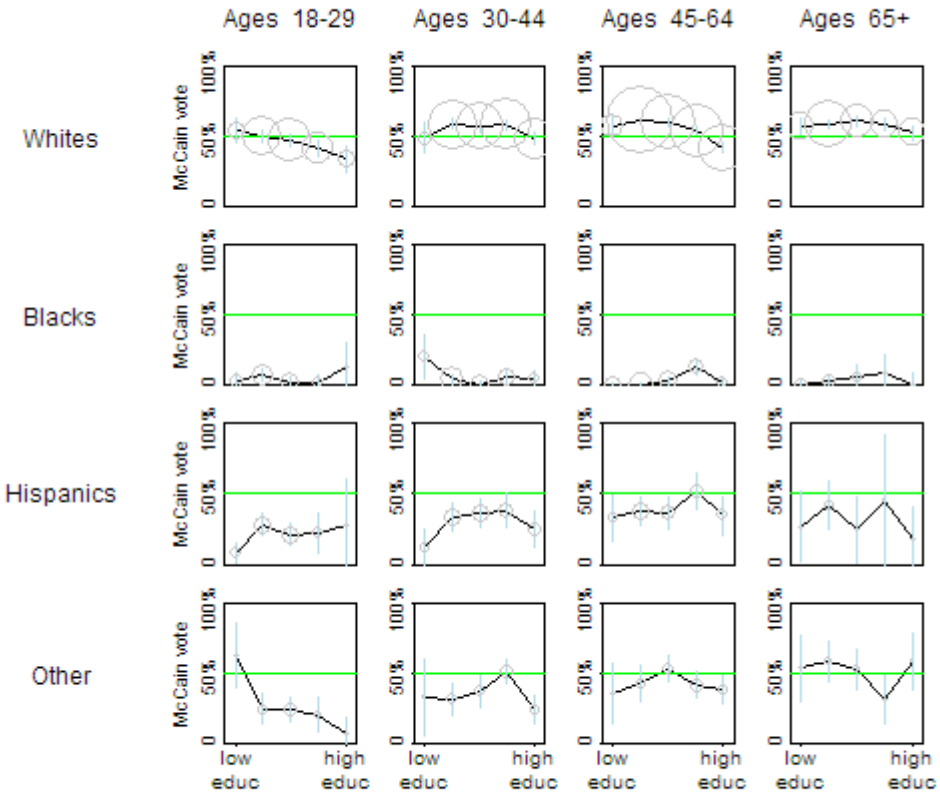
Gelman, A., and Su, Y. S. (2010). Voting by age in 2008. *Chance*, to appear.

Tufte, E. R. (1990). *Envisioning Information*. Cheshire, Conn.: Graphics Press



When a category represents less than 1% of the voters in a state, the state is left blank

Republican vote in 2008 by education, among age/ethnic groups



Education categories: no high school, high school, some college, college grad, postgrad.
Estimates based on Pew Research pre-election polls. Error bars show +\- 2 standard errors.
Area of each circle is approximately proportional to number of voters in the category.