Statistics, Politics, and Policy

Volume 1, *Issue* 1 2010 *Article* 3

A Snapshot of the 2008 Election

Andrew Gelman, *Columbia University*
Daniel Lee, *Columbia University*
Yair Ghitza, *Columbia University*

Recommended Citation:

Gelman, Andrew; Lee, Daniel; and Ghitza, Yair (2010) "A Snapshot of the 2008 Election," *Statistics, Politics, and Policy*: Vol. 1: Iss. 1, Article 3.
DOI: 10.2202/2151-7509.1006
Available at: http://www.bepress.com/spp/vol1/iss1/3

©2010 Berkeley Electronic Press. All rights reserved.

A Snapshot of the 2008 Election

Andrew Gelman, Daniel Lee, and Yair Ghitza

Abstract

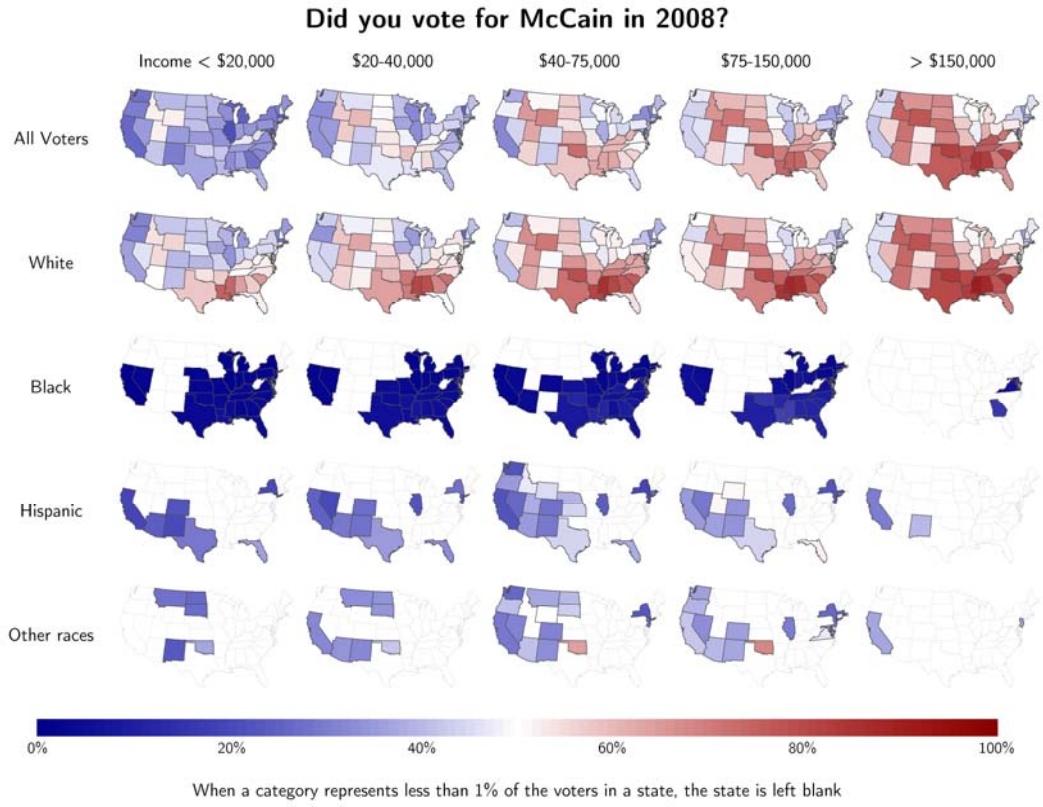
We present maps of the 2008 presidential voting bases on ethnicity, income, and state, and discuss the challenges involved in statistical modeling and the graphical presentation of the results

KEYWORDS: statistical graphics

Author Notes: We thank the National Science Foundation, Department of Energy, and Institute for Education Sciences for financial support of this work, the Pew Research Center for their survey data, and John Rolph for helpful comments. Further graphs of a similar nature appear at our blog: http://www.stat.columbia.edu/~cook/movabletype/archives/political-science/

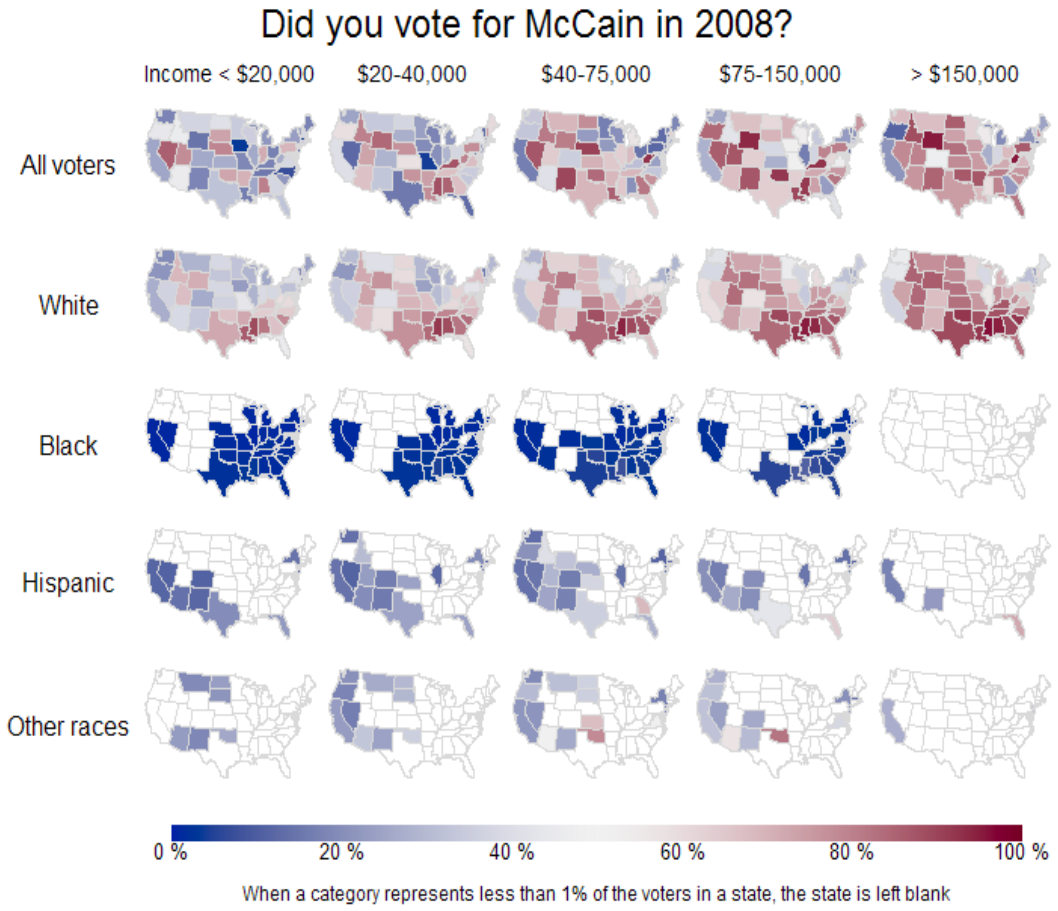
Erratum

Note from the authors: We made some computational mistakes in preparing the maps. The revised maps show the estimated percentage of the two-party vote received by John McCain and Barack Obama for voters characterized by income, ethnicity, and state. We thank Russell Lyons for pointing out our error.



We started producing graphs of the 2008 election starting around 1 am on election night and haven't stopped yet.

The purpose of the present article is to demonstrate the potential of statistical graphics to help us understand the political process, and also to illustrate some of the difficulties involved in constructing an effective graph (and in performing the statistical analysis leading up to it).



Having decided to illustrate our points with a single graph, we decided to cheat and display a grid of 25 maps, thus illustrating the power of the "small multiples" idea advocated by pioneering graphics researchers Bertin (1967) and Tufte (1990) and incidentally demonstrating how a large amount of information can be displayed effectively in a small space if one can make use of the reader's background knowledge about incomes, ethnic groups, and U.S. geography.

We produced the maps by processing Pew Research Center post-election polls using multilevel regression and poststratification, as described in detail in Gelman and Ghitza (2010). We fit a logistic model predicting the probability of voting for McCain (conditional on voting for one of the two major-party candidates), given ethnicity, income, and state. The model includes main effects for each of these factors and also two-way and three-way interactions. Large as our survey is (about 50,000 respondents nationally), we still only have small sample sizes in the individual cells defined by the interactions of our three factors. To get reasonable estimates, we use hierarchical Bayes inference, in which the observed data in each cell are partially pooled toward a regression model based on a combination of the one-way factors and continuous predictors (individual income, state income, and historical vote patterns by state).

In addition to fitting the model, we developed graphs specifically for the purpose of revealing discrepancies between model and data and for revealing estimates that might be surprising enough to cast doubt on our inferences. Indeed, earlier iterations of our maps had some serious problems—revealing serious model errors—which we corrected before getting to this current version (see Gelman, 2009). Our inferential procedure gives estimates even for small-sample-size categories such as high-income whites in Wyoming (as estimated by some combination of results for the income category, the ethnic group, the state, and other states with similar partisanship and income). We also use Census demographic numbers and state-by-state election returns to correct for known differences between sample and population that were handled using weighting adjustments in the original survey.

Three small graphical innovations in the present graph (at least, compared to our own earlier work) are: (1) A continuous blue-to-gray-to-red color scale that is labeled directly on the bottom of the display; (2) Gray state borders rather than the default black, which crowd the graph and distract the eye; and (3) Direct labeling, with the demographic identifiers on the left and top rows and the survey question right at the top.

The maps reveal that, among whites, income is strongly correlated with Republican voting, especially in the center of the country but not so much in the west coast and northeast. This is similar to what we found in 2000 and 2004 (Gelman et al., 2009), but this time our statistical analysis was more sophisticated, moving beyond simple multilevel logistic regression to add additional levels of variation corresponding to departures from linearity on the logistic scale. This model improvement was motivated by earlier graphical model checks.

Future researchers (maybe even us) will have many opportunities to improve on this work, most notably by increasing the number of variables that can be studied and coming up with tools to display the correspondingly more complex

DOI: 10.2202/2151-7509.1006

inferences. And we believe these methods can also be effective in studying the relation between public opinion and policy, following up on work such as that of Lax and Phillips (2009) who used multilevel regression and poststratification to get estimates of state-level opinion on gay rights and then link changes in opinion to state-level changes in policy. This sort of analysis had not been possible before, and the underlying statistical techniques were much more powerful in practice when tied to sophisticated visualizations—not necessarily eye-catching graphics that would win design awards, but structured displays tied to the political science and policy questions at hand.

Beyond their ability to reveal unexpected patterns in data and inference, and to summarize results for consumers of the research, detailed maps and graphs allow researchers to better understand the implications of their own models and to be better self-critics. A famous empiricist once said, "With great power comes great responsibility." Graphical methods allow us to exercise this responsibility.

References

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

Gelman, A. (2009). How did white people vote? Updated maps and discussion. Statistical Modeling, Causal Inference, and Social Science blog, 11 May. http://www.stat.columbia.edu/~cook/movabletype/archives/2009/05/discussion_and.html

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., 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.

Lax, J. R., and Phillips, J. H. (2009). Gay rights in the states: Public opinion and policy responsiveness. *American Political Science Review* **103**, 367-386.

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