

## 4. Generalizing from sample to population

- ▶ Big Data need Big Model
- ▶ Examples from polls

me social tv video games music apps



**UFC** 153

Press Conf



New Trials



Kinect Cent



ADVERTISEMENT

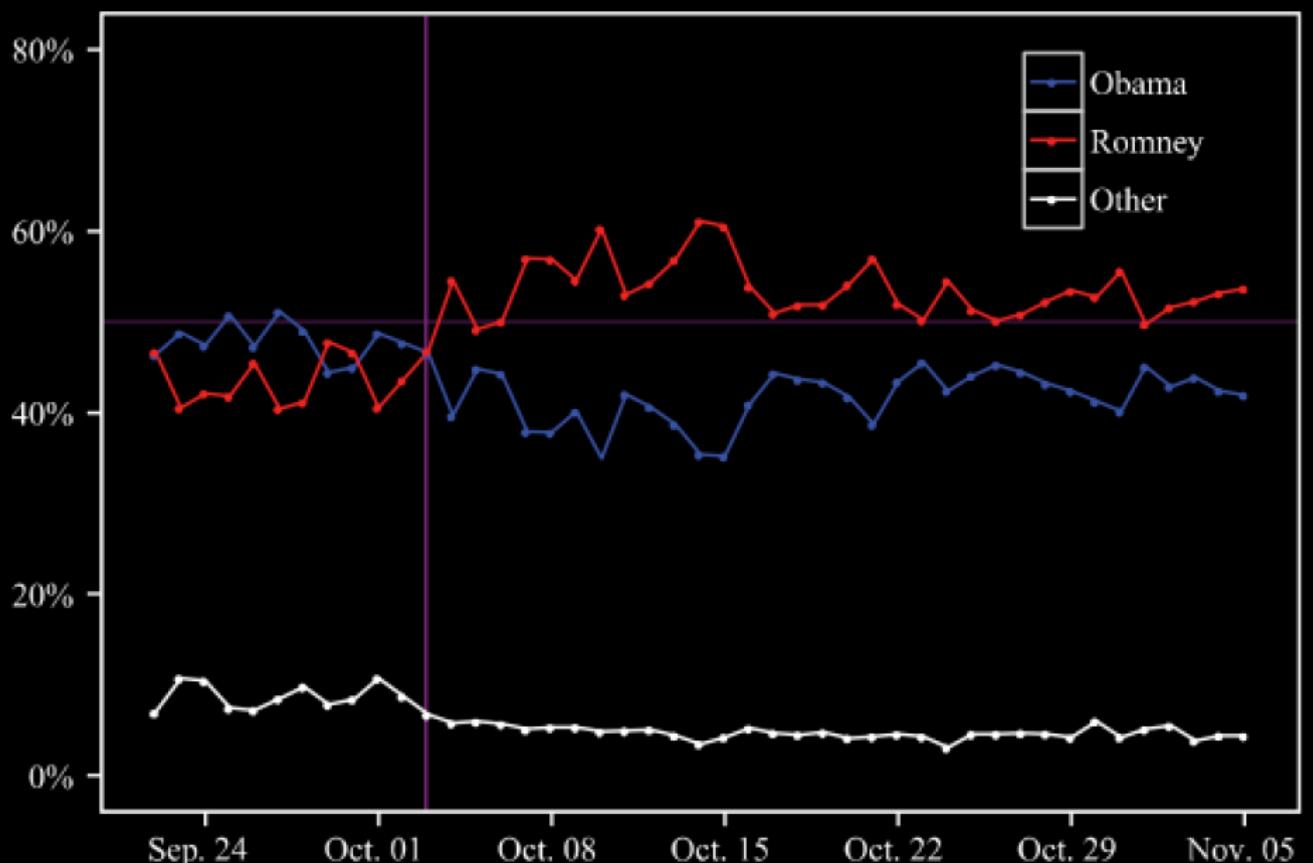
If the election were held today, who would you vote for?

Barack Obama

Mitt Romney

Other

Not sure



"This week, the New York Times and CBS News published a story using, in part, information from a non-probability, opt-in survey sparking concern among many in the polling community. In general, these methods have little grounding in theory and the results can vary widely based on the particular method used."

— Michael Link,

President, American Association for Public Opinion Research

**Michael W. Link** is Chief Methodologist for Research Methods at The Nielsen Company. Michael has a broad base of experience in survey research, having worked in academia (University of South Carolina, 1996-1999), not-for-profit research (RTI International, 1999-2004), government (Centers for Disease Control and Prevention, 2004-2007), and the private sector (Nielsen, 2007-present). He received his Ph.D. in Sociology from the University of South Carolina. Michael's research centers around developing new methodologies for confronting some of the most pressing issues facing survey research, including techniques for improving survey participation and data quality (use of address-based sampling, telephone and mail call screening technologies), methodological issues involving use of multiple modes in data collection (e-mail, CATI, field, mobile, meters), and obtaining participation from hard-to-survey populations (e.g., isolated, racial and ethnic groups). His numerous research articles have appeared in *Public Opinion Quarterly* and other leading scientific journals.

An AAPOR member since 1993, Michael served as AAPOR Conference Chair in back-to-back years (2009 & 2010), a member of both the Cell Phone and Online task forces, an instructor for an AAPOR short course, numerous short-courses, a reviewer for the student paper competition on several occasions, and a regular reviewer for *Public Opinion Quarterly*. He is a member of SAPOR, serving from 2008-2010 as President, Conference Chair, and Student Paper Competition Organizer and also a member of the

In 2011 he, along with several research colleagues, received AAPOR's Warren J. Mitofsky Award for their work on address based sampling designs. His current research focuses on the use of new technologies, such as mobile and social platforms, as vehicles for measuring and understanding attitudes and behaviors. He will be teaching a short course on "The Role of New Technologies in Augmenting, or Replacing Traditional Surveys" at the 2012 AAPOR conference.

# Nielsen feels the heat of competition as it flubs its ratings of news broadcasts, putting ABC ahead of NBC

In spite of the goof, its global president took time to slam rival Rentrak, which collects different kind of data from viewers

NEW YORK DAILY NEWS / Sunday, October 19, 2014, 2:00 AM

BY DON KAPLAN



MEDIA

## *TV Ratings by Nielsen Had Errors for Months*

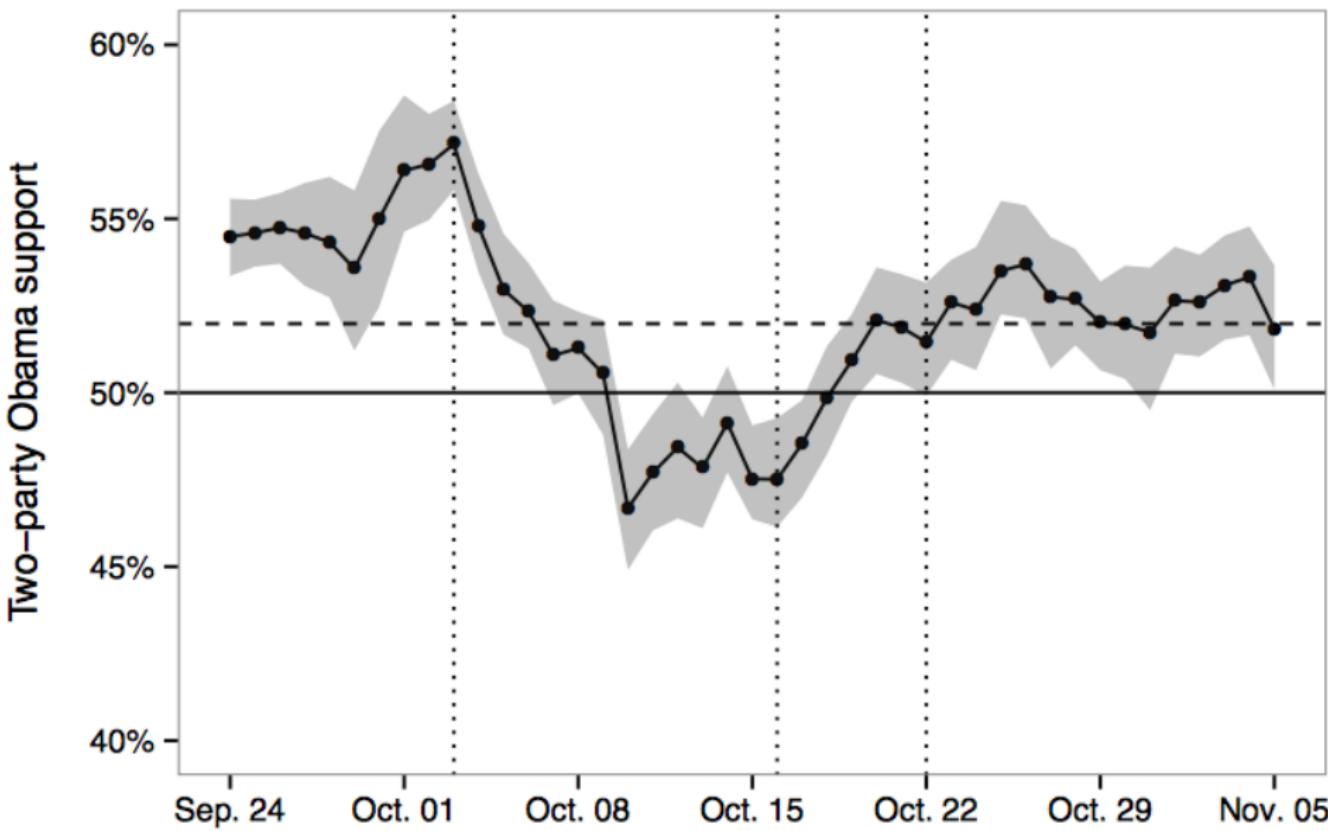
By BILL CARTER and EMILY STEEL OCT. 10, 2014

Email

Share

Nielsen, the television research firm, acknowledged on Friday that it had been reporting inaccurate ratings for the broadcast networks for the last seven months, a mistake that raises questions about the company's increasingly criticized system for measuring TV audiences.

## Xbox estimates, adjusting for demographics

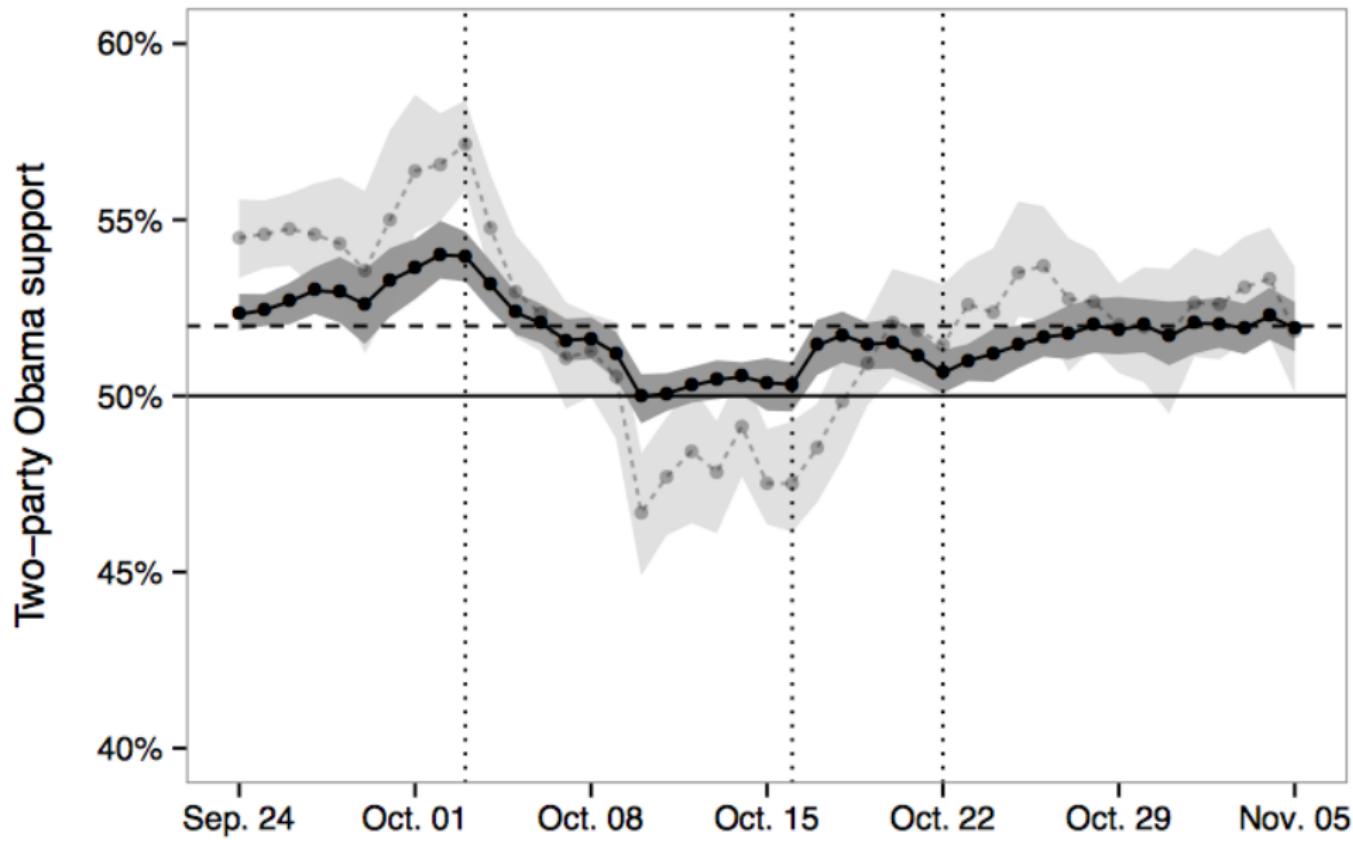




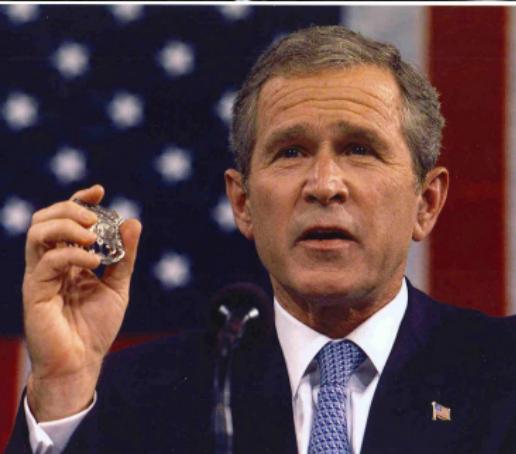


- ▶ Karl Rove, *Wall Street Journal*, 7 Oct: “Mr. Romney’s bounce is significant.”
- ▶ Nate Silver, *New York Times*, 6 Oct: “Mr. Romney has not only improved his own standing but also taken voters away from Mr. Obama’s column.”

# Xbox estimates, adjusting for demographics and partisanship

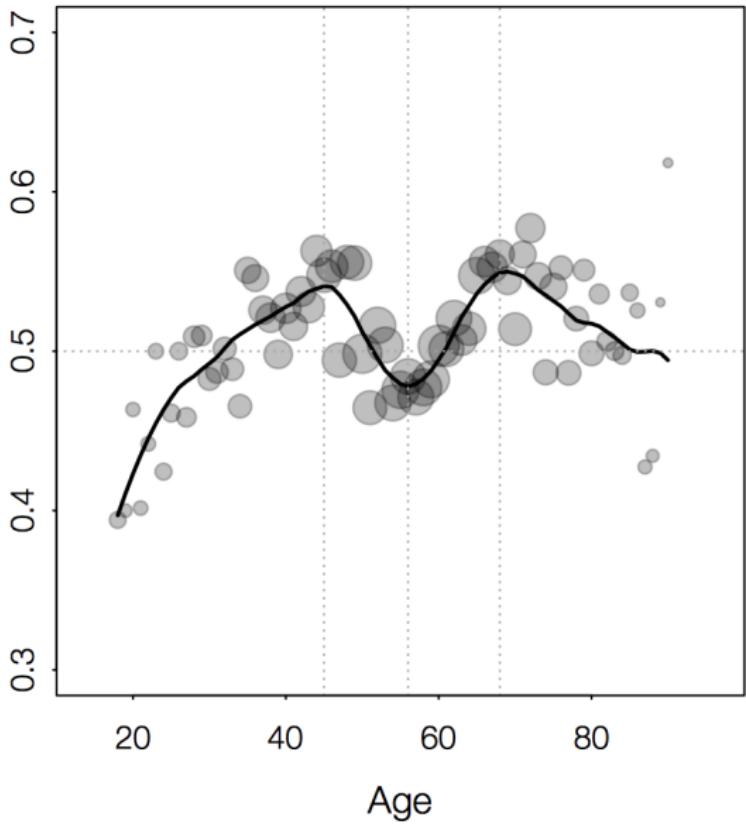


# Jimmy Carter Republicans and George W. Bush Democrats

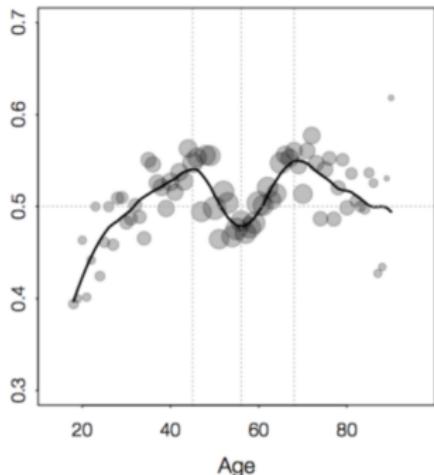


Republican Vote

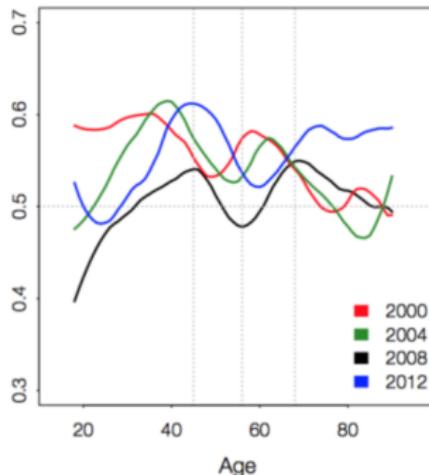
Non-Monotonic Age Curve in 2008



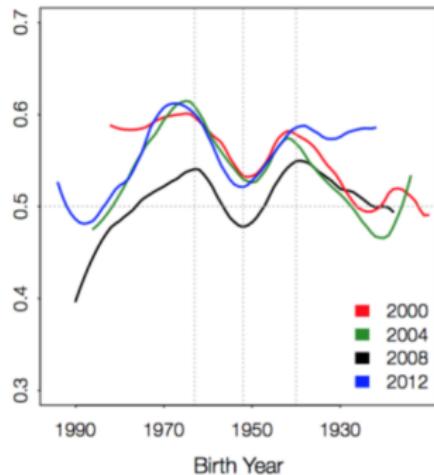
Non-Monotonic Age Curve in 2008



Non-Monotonicity in Other Elections

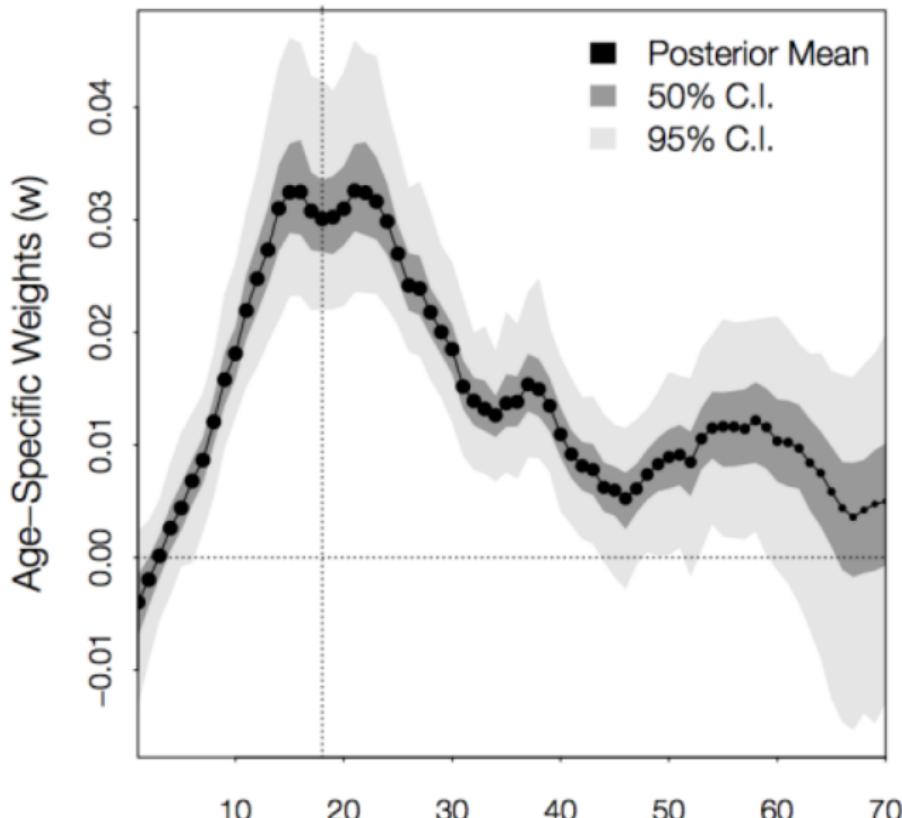


Lining up by Birth Year



# The Formative Years

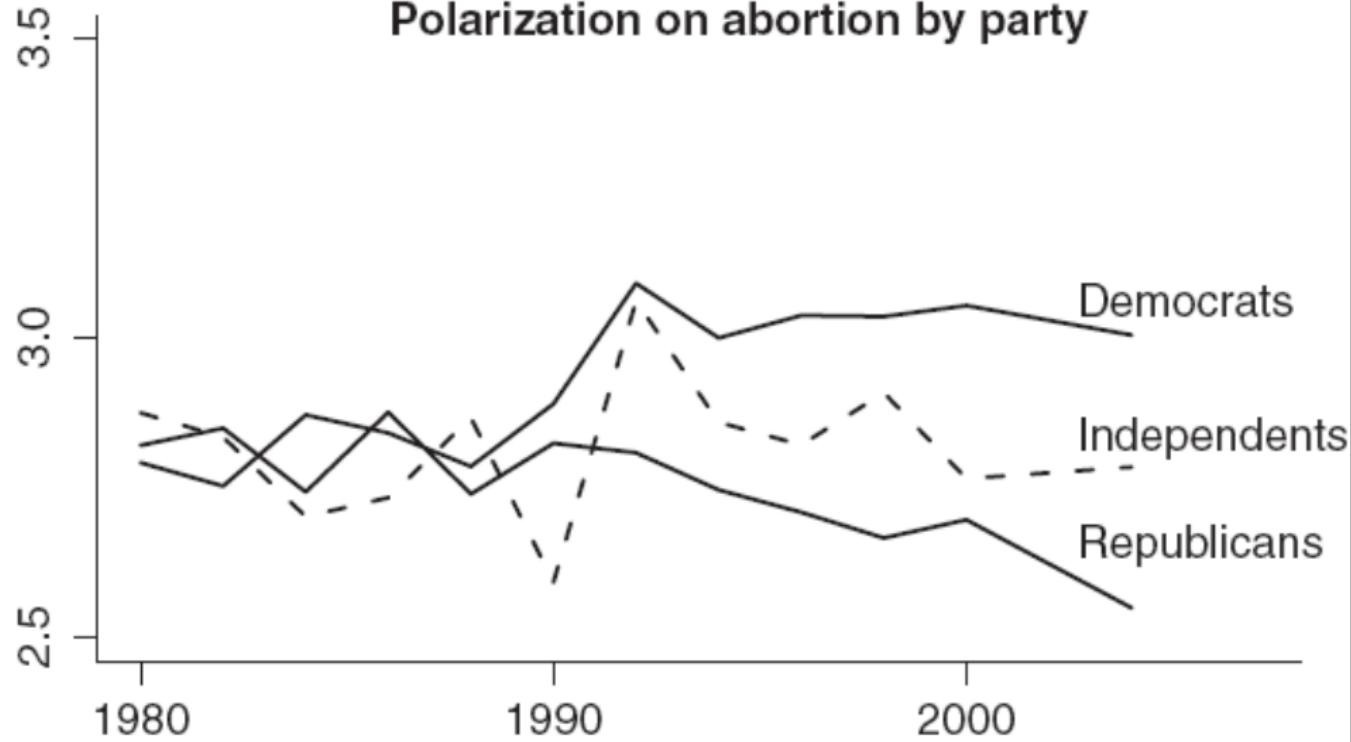
18



Adjusting for known differences between sample and population:

- ▶ Include more predictors
- ▶ Multilevel regression
- ▶ Poststratification

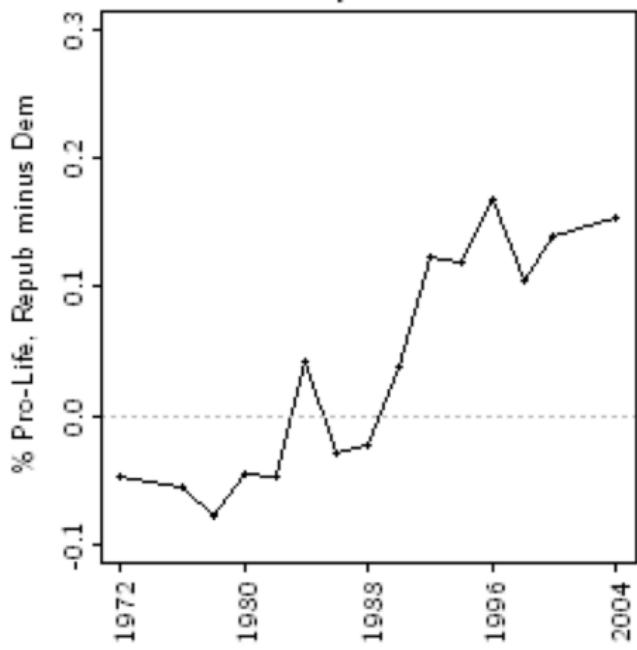
## Polarization on abortion by party



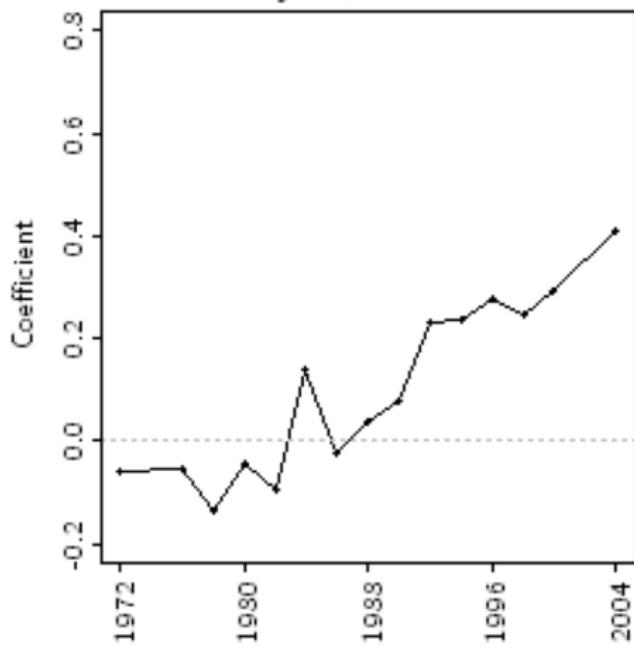
# G.O.P. Pursues Hispanic Votes With Abortion Stance



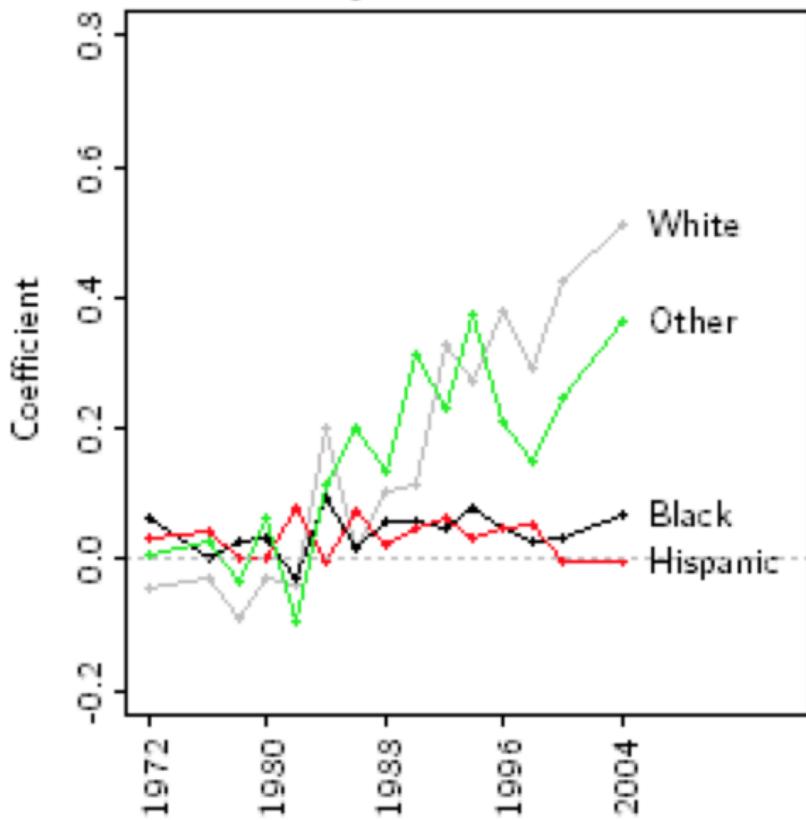
Pro-Life Tendency of  
R's Compared to D's



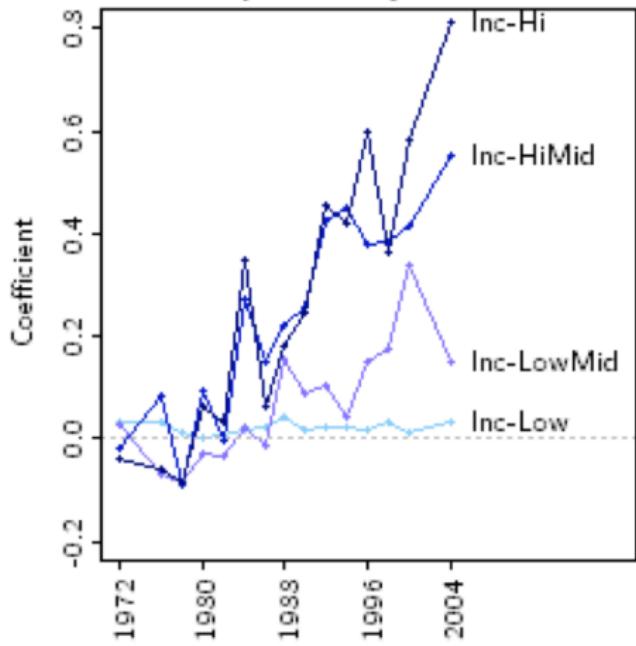
Coef of Abortion Opinion  
on Party ID, fit with HLM



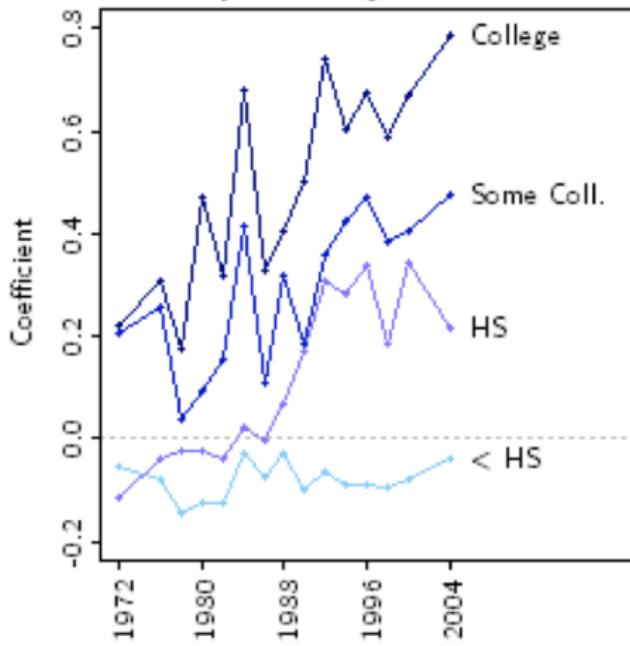
### Coef of Abortion Opinion on Party ID, fit with HLM



Coef for Whites Only,  
Separated by Income



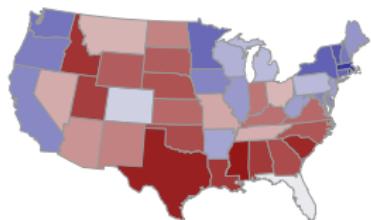
Coef for Whites Only,  
Separated by Education



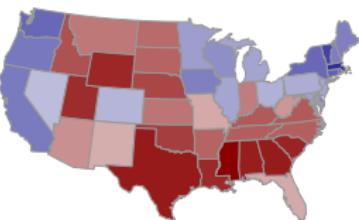
## Some Bayesian issues:

- ▶ Regression with lots of predictors
- ▶ Dense multilevel models
- ▶ Deep interactions
- ▶ Informative priors
- ▶ Computing
- ▶ Confidence building
- ▶ Understanding the models we've fit

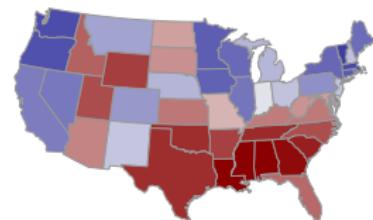
Annenberg 2000: Logit



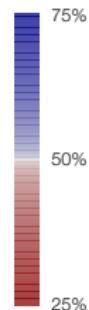
Annenberg 2004: Logit



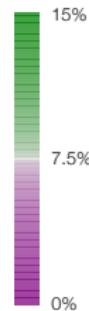
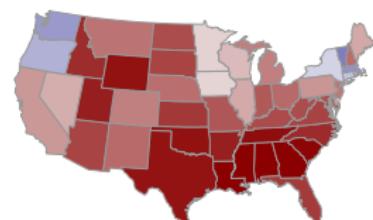
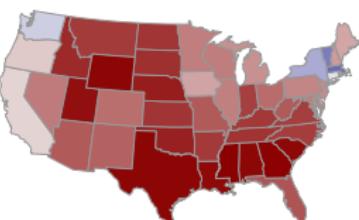
Annenberg 2008: Logit



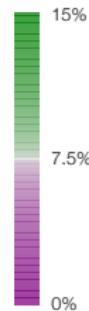
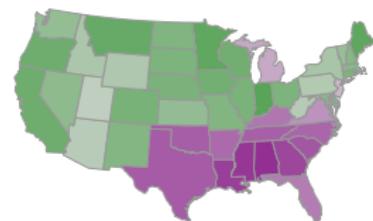
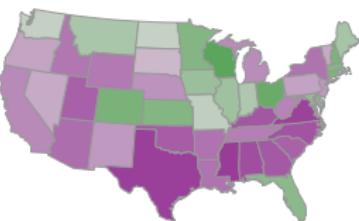
White Female



White Male



White Gender Gap



## White Gender Gap

