

Ronald Reagan was a statistician
and other examples
of learning from diverse sources of information

Andrew Gelman
Department of Statistics, Columbia University, New York

21 Mar 2013

“Are you better off than you were four years ago?”



“But the simplest way for each of you to judge recovery lies in the plain facts of your own individual situation. Are you better off than you were last year?”





Andrew Gelman

Ronald Reagan was a statistician

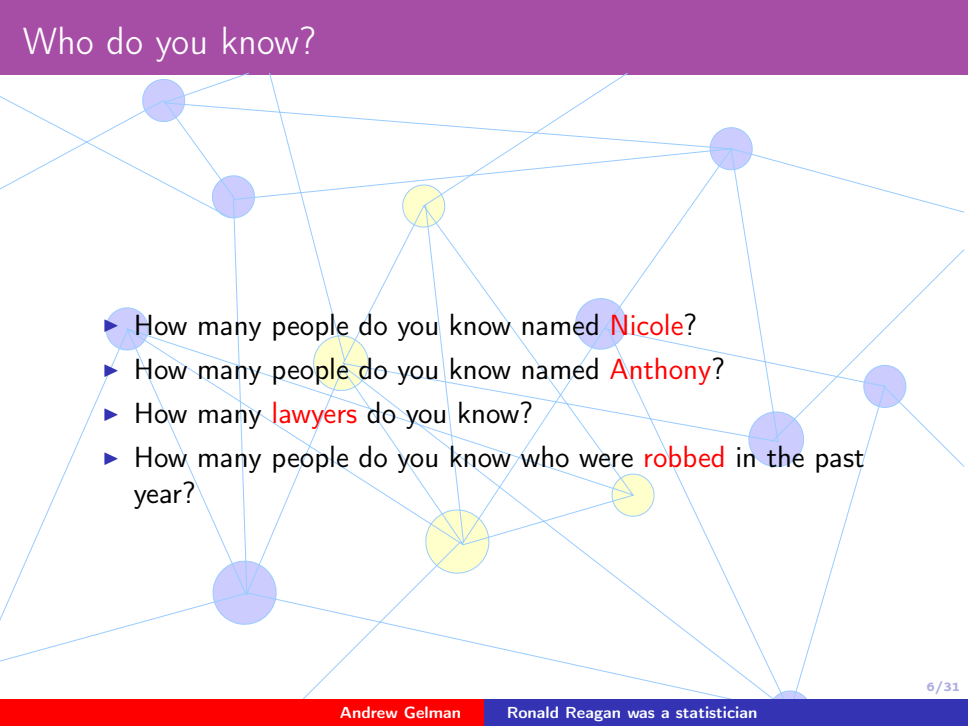
Go, Data Go!



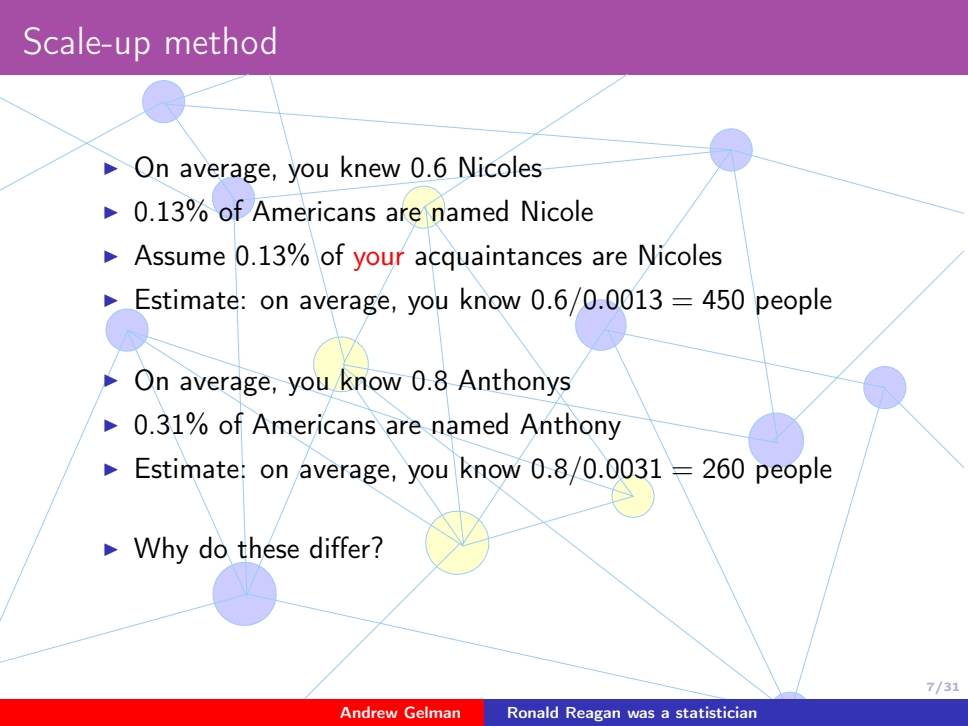
by

Andrew Gelman

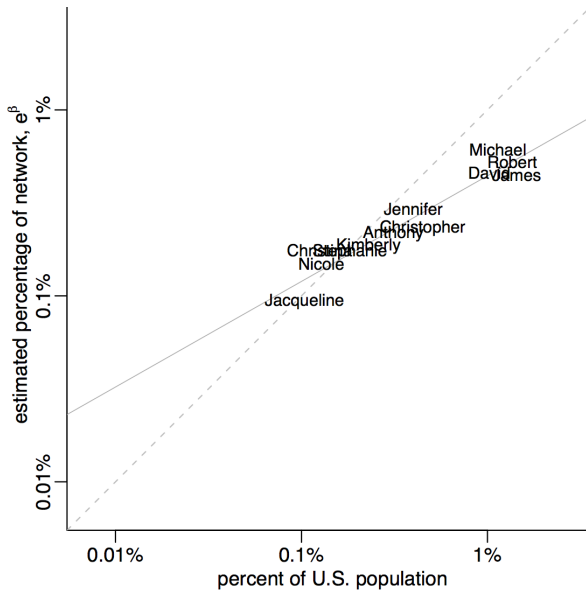
Who do you know?

- 
- ▶ How many people do you know named **Nicole**?
 - ▶ How many people do you know named **Anthony**?
 - ▶ How many **lawyers** do you know?
 - ▶ How many people do you know who were **robbed** in the past year?

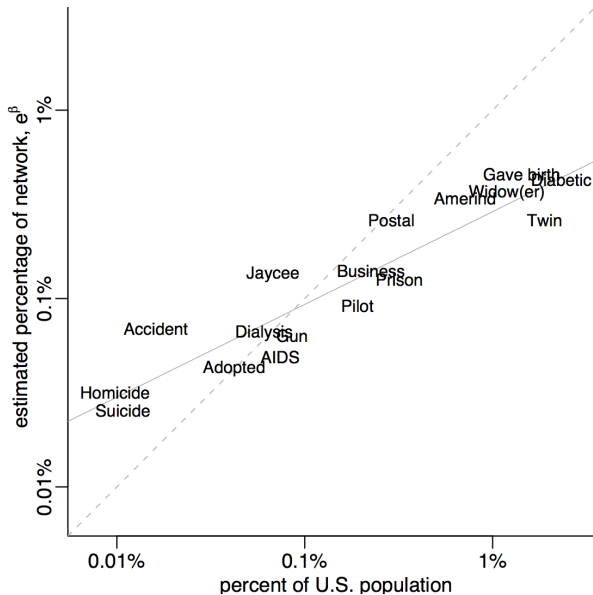
Scale-up method

- 
- ▶ On average, you knew 0.6 Nicoles
 - ▶ 0.13% of Americans are named Nicole
 - ▶ Assume 0.13% of **your** acquaintances are Nicoles
 - ▶ Estimate: on average, you know $0.6/0.0013 = 450$ people
 - ▶ On average, you know 0.8 Anthonys
 - ▶ 0.31% of Americans are named Anthony
 - ▶ Estimate: on average, you know $0.8/0.0031 = 260$ people
 - ▶ Why do these differ?

Perception vs. reality: Names

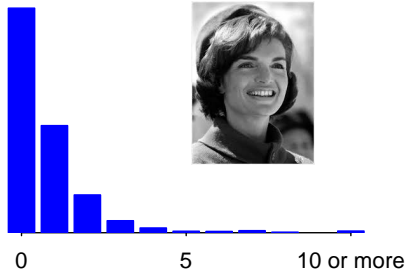


Perception vs. reality: Groups

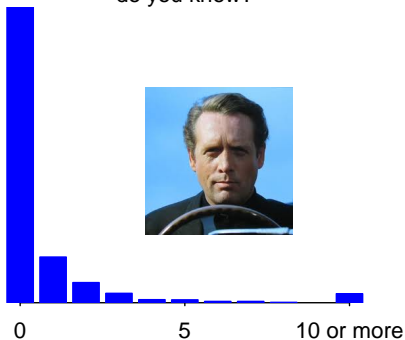


Social Structure!

How many Jacquelines do you know?



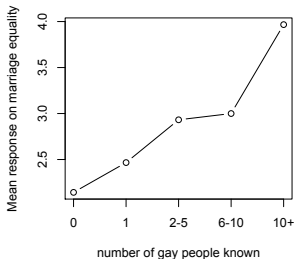
How many men in prison do you know?



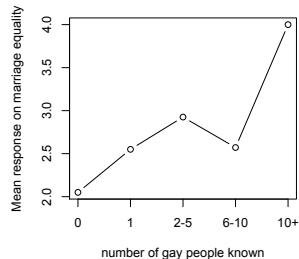
The Gay Penumbra



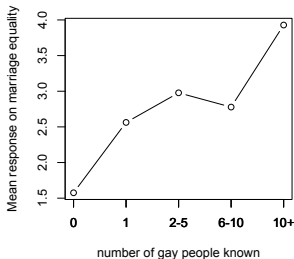
Marriage Support: Everyone



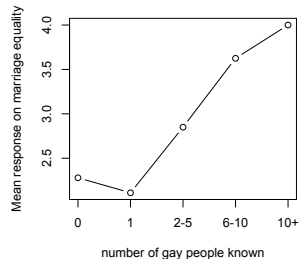
Marriage Support: Democrats



Marriage Support: Republicans



Marriage Support: Independents

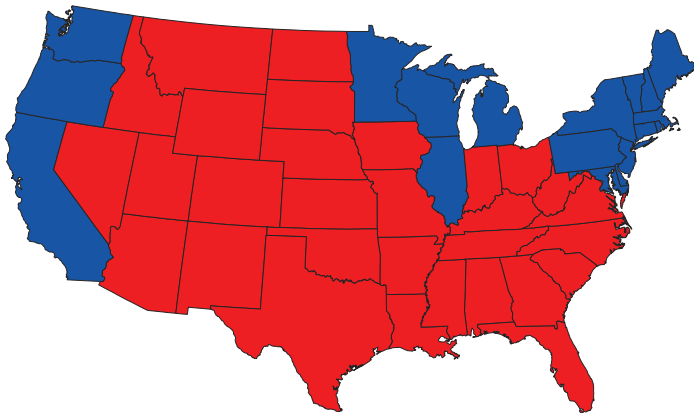


How did this happen?

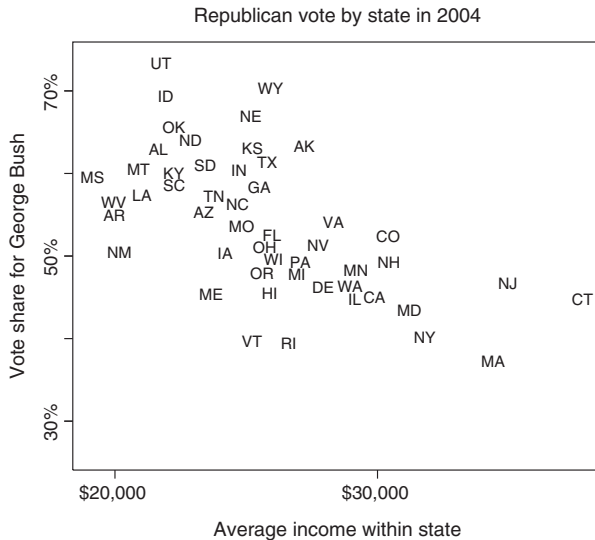


It makes no sense!

2004 election

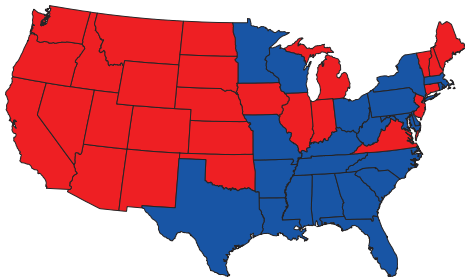


It's no illusion

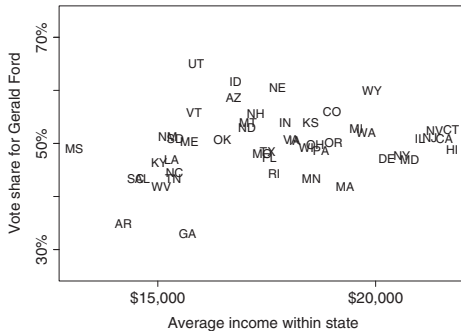


What has changed?

1976 election

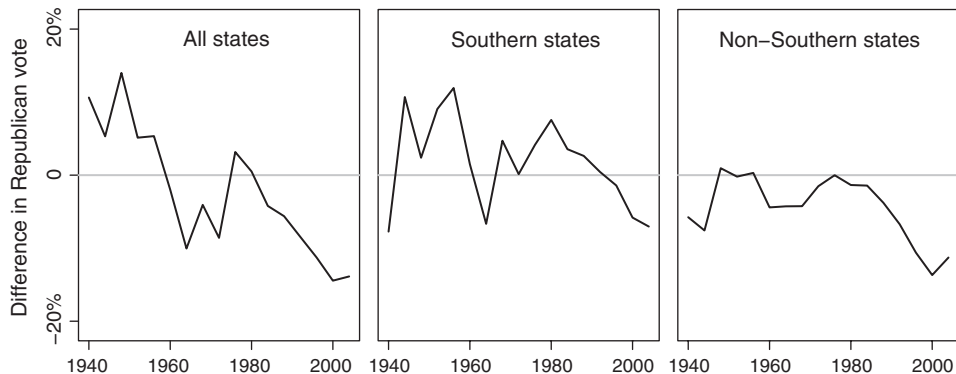


Republican vote by state in 1976



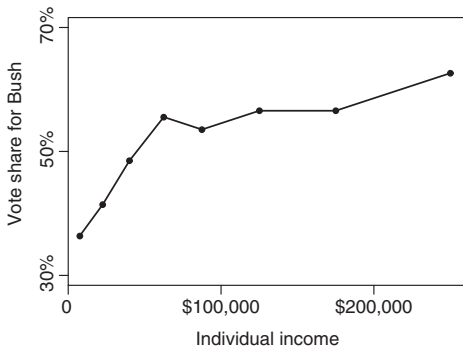
It's not just the South

Republican vote in rich states minus Republican vote in poor states

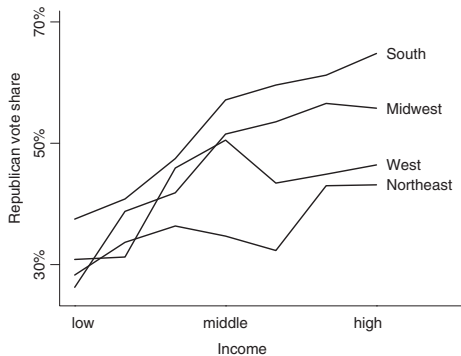


The rich still vote Republican

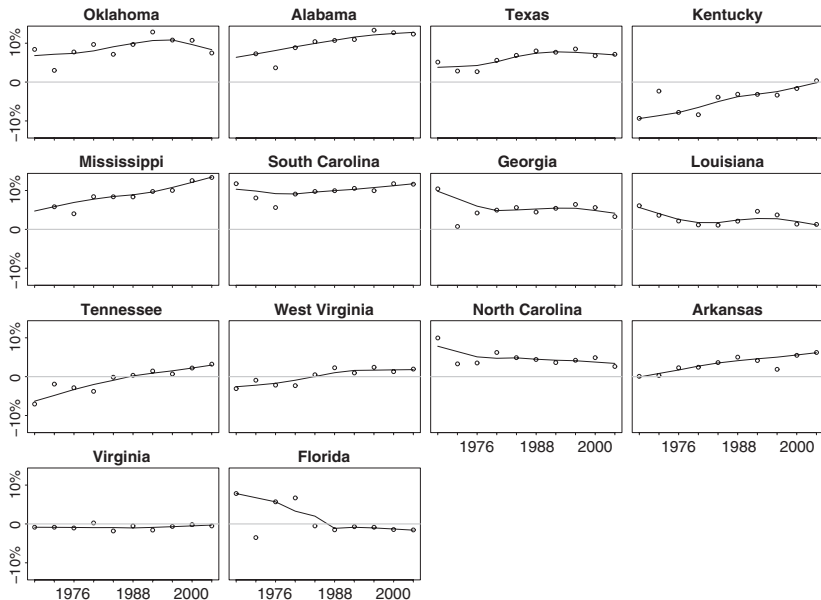
Bush vote in 2004 by income



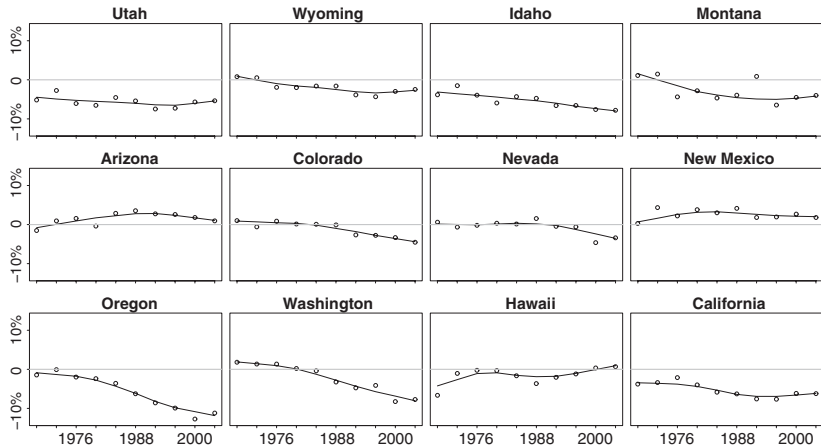
2006 House exit polls



Rich counties are more Republican



But not in the West!



“Opiate of the masses” vs. “Postmaterialism”

- ▶ “I don’t know that atheists should be considered citizens, nor should they be considered patriots. This is one nation under God.” — George H. W. Bush

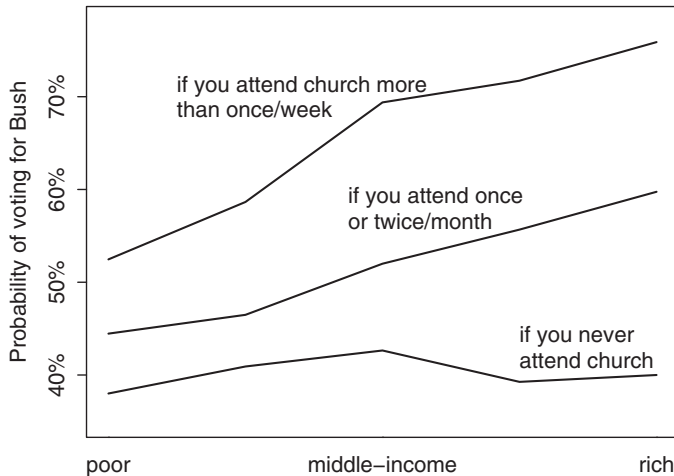


- ▶ “Very few of us . . . could name even five NASCAR drivers, although stock-car races are the best-attended sporting events in the country.” — David Brooks
- ▶ Two theories:
 - ▶ *Opiate of the masses*: Rich people vote their interests, poor people vote “Gods, guns, and gays”
 - ▶ *Postmaterialism*: Poor people vote based on economics, rich people have the luxury to vote on social issues

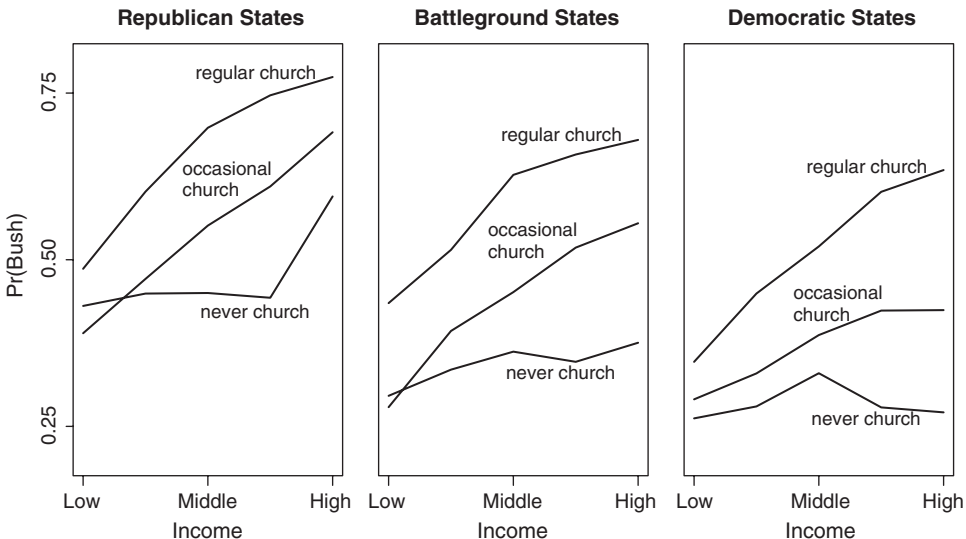


“Opiate” no, “Postmaterialism” yes

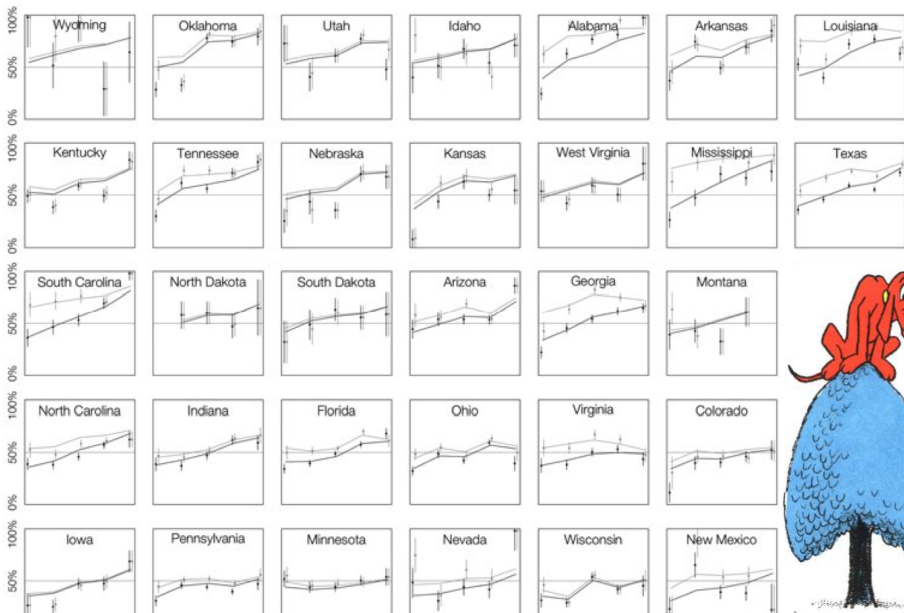
Bush vote in 2004 by income and religious attendance



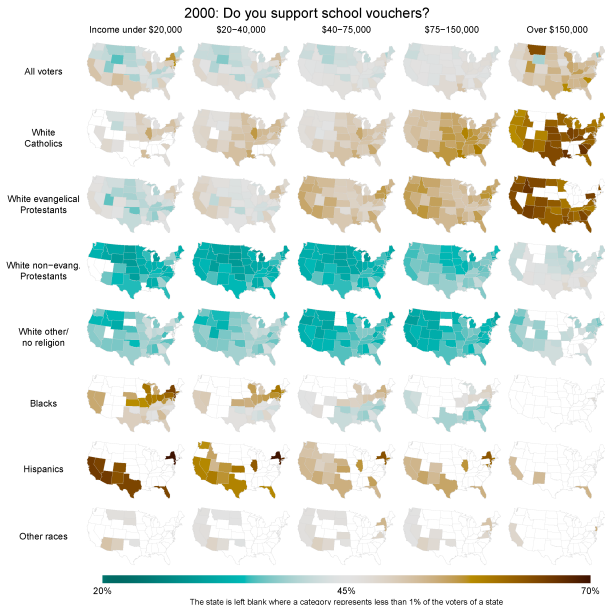
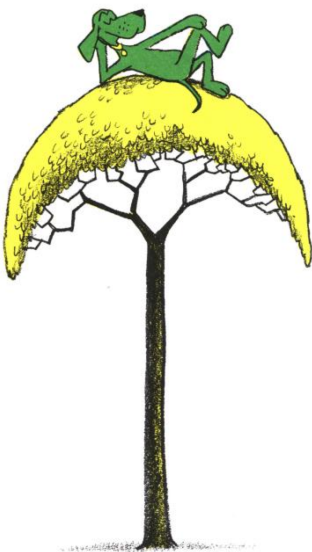
Similar patterns in red and blue America



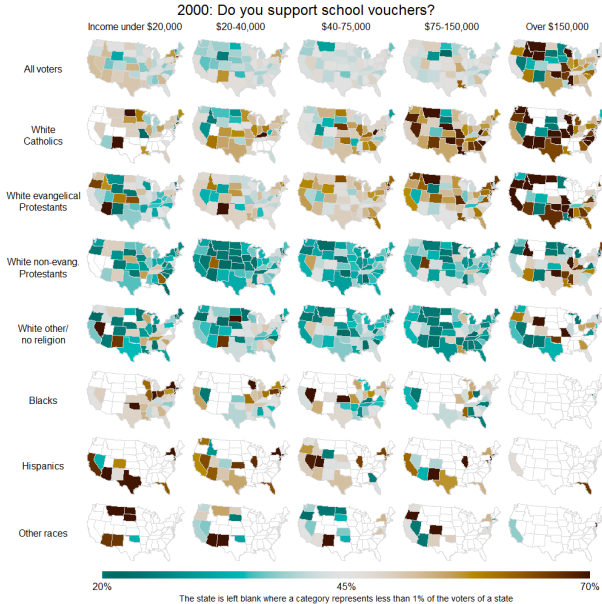
Big Model, Sparse Data



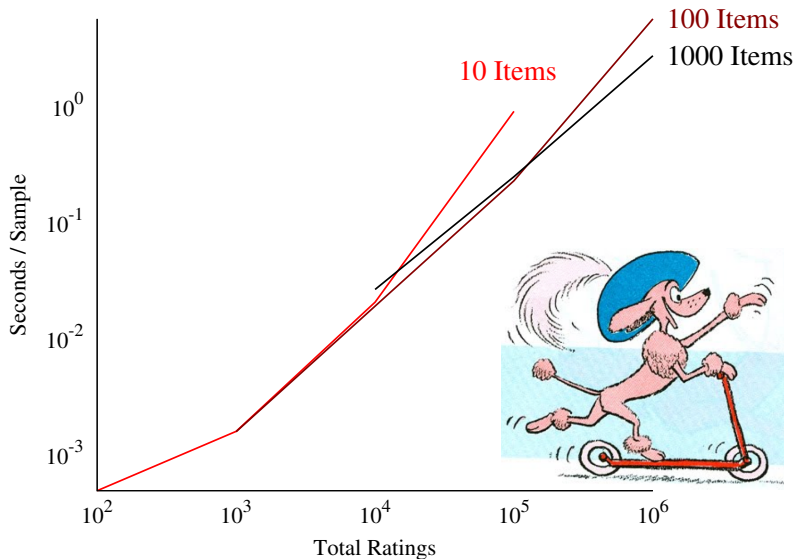
Big Data and Big Model



Data don't always “speak for themselves”



Big Data, Big Model, Scalable Computing



The Perils of Pooling

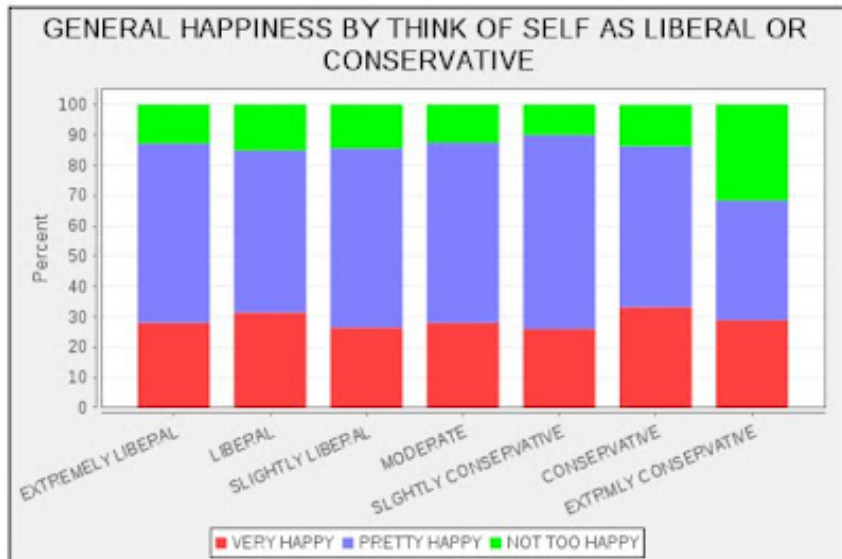
Arthur “not David” Brooks in the *New York Times*:

“People at the extremes are happier than political moderates. . . . none, it seems, are happier than the Tea Partiers . . .”

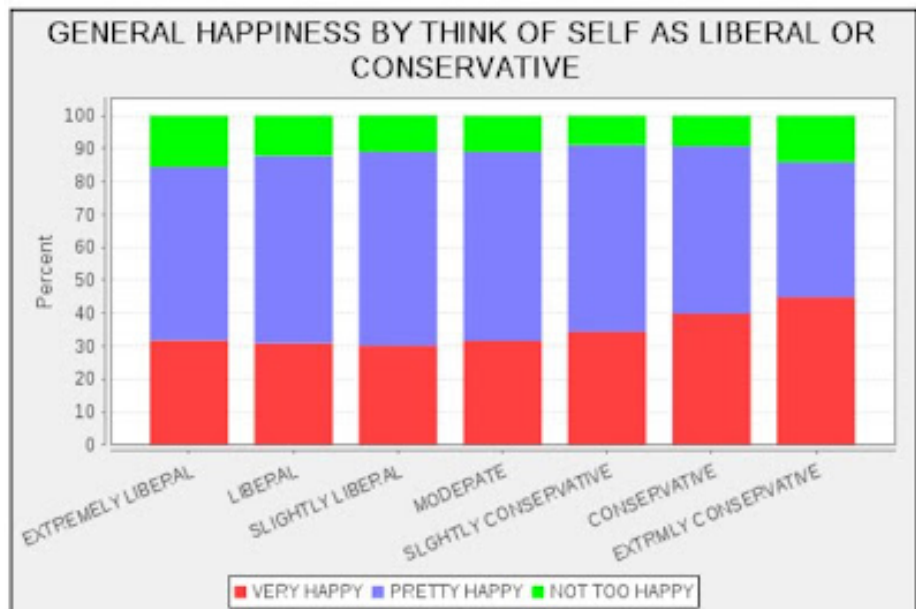
Jay Livingston (sociology, Montclair State University) looks up the data in the General Social Survey . . .

“None, it seems, are happier than the Tea Partiers ...” ??

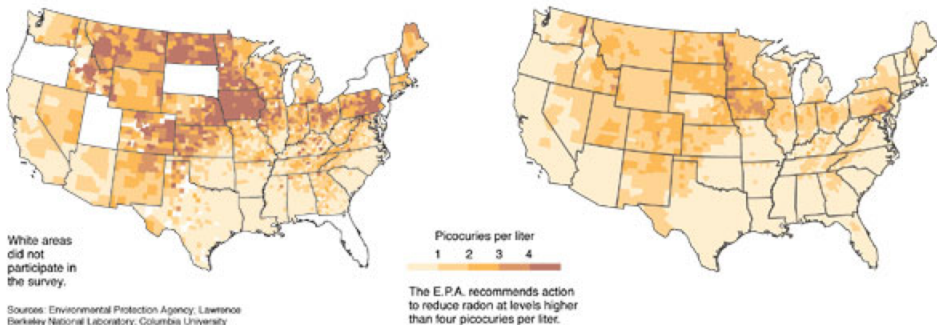
Chart for YEAR = 4(2009-2010)



Pooling, 1972–2010

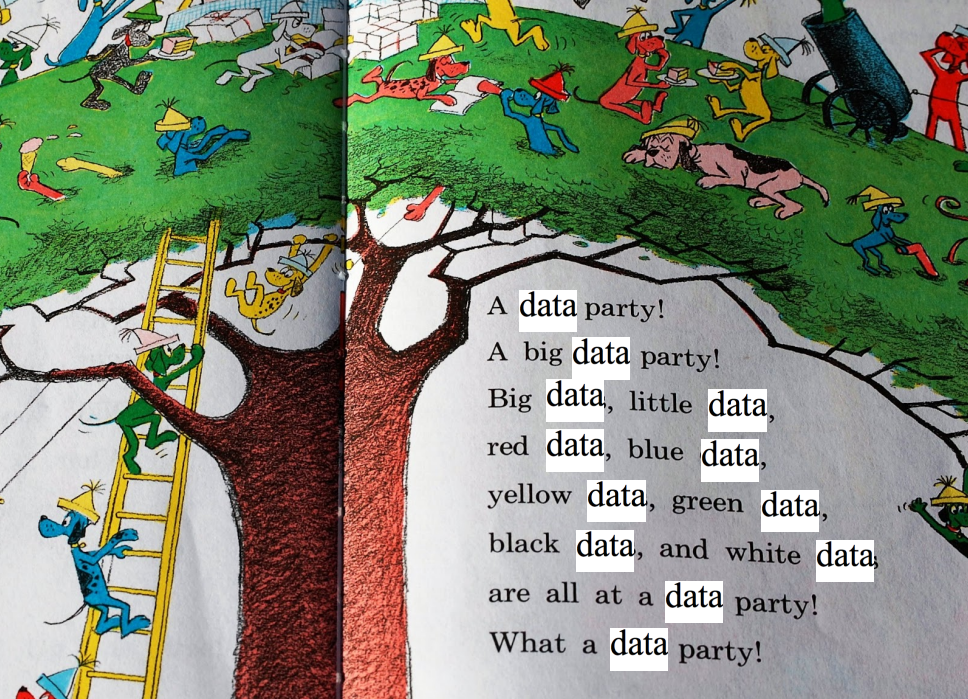


Local Data, National Coordination, Local Decisions



The New York Times

Older Model, left: Observed indoor radon levels, based on short-term measurements collected by the E.P.A. and most states. The measurements, taken in basements in the winter, may overestimate true exposure. New Model, right: The predicted average long-term levels of radon that people may actually breathe in their homes.



A data party!

A big data party!

Big data, little data,
red data, blue data,
yellow data, green data,
black data, and white data,
are all at a data party!
What a data party!