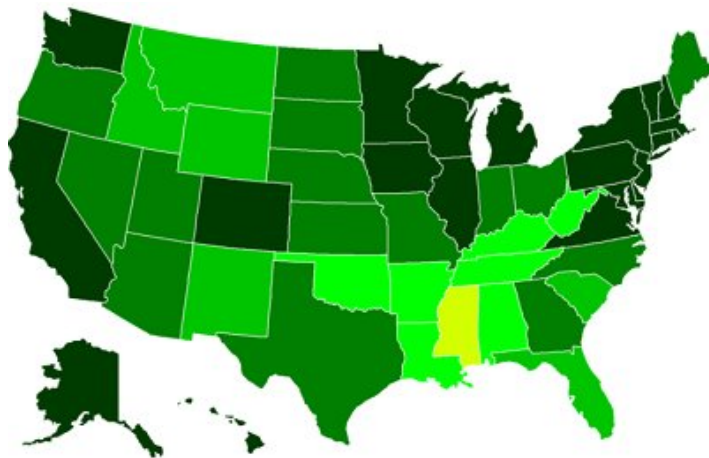# Design and Analysis of Sample Surveys

Andrew Gelman
Department of Statistics and Department of Political Science
Columbia University

Class 10b: Understanding and displaying data

Human Development Index

Human Development Index by State

|				
.950+	.900-.949	.850-.899	.800-.849	.750-.799

Human Development Index

- Three lowest values are Louisiana at .801, West Virginia at .800, and Mississippi at .799
- Four highest values are Connecticut, Massachusetts, New Jersey, and . . . D.C.??
- Components of HDI:
    - Life expectancy at birth
    - Knowledge and education, as measured by the adult literacy rate (with two-thirds weighting) and the combined primary, secondary, and tertiary gross enrollment ratio (with one-third weighting)
    - Standard of living, as measured by the natural logarithm of gross domestic product (GDP) per capita at purchasing power parity (PPP) in United States dollars
- What would Kumail Nanjiani say?

HDI vs. avg. income by state
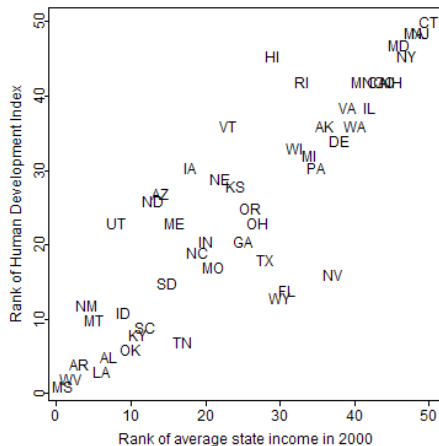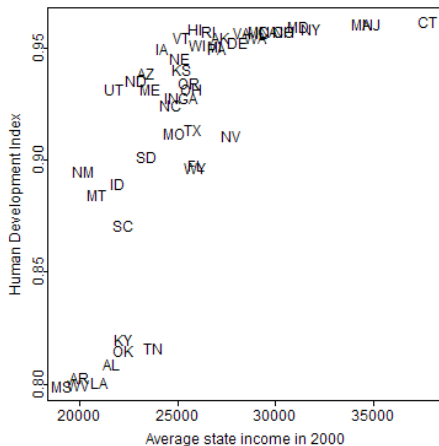
What's the matter with Portugal?
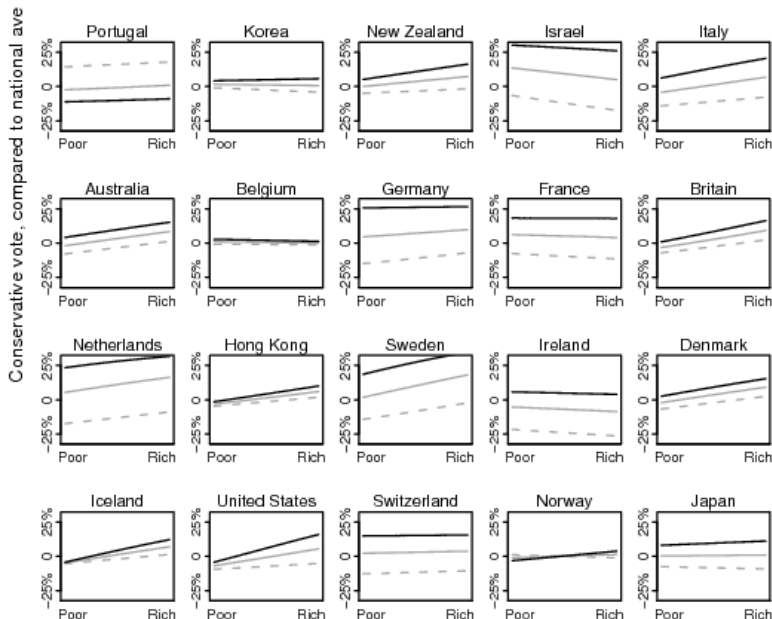
Statistical inference

- ▶ Averages, differences, and standard errors
- ▶ Dotplots and line plots of data
- ▶ Linear and logistic regression
- ▶ More complicated models
- ▶ Plots of parameter estimates
- ▶ Plots of fitted models
- ▶ Survey weights

Interpreting statistical findings

- Data quality
- Generalizing from sample to population
- The statistical model
- The steps of inference
- Causal inference and the implicit "treatment"
- Real-world implications

"Will Ohio State's football team decide who wins the White House?"

- Recent study: "a win in the 10 days before Election Day causes the incumbent to receive an additional 1.61 percentage points of the vote in Senate, gubernatorial, and presidential elections"
- County-level vote analysis
- More evidence:
  - Larger effects "for teams with stronger fan support"
  - "placebo tests based on postelection games"
  - Survey data: "surprising wins and losses affect presidential approval"
  - Consistent with literature on shark attacks etc.

"Will Ohio States football team decide who wins the White House?"

"The key to victory could come down to ... Florida, Ohio, and Virginia. On Oct. 27th, a little more than a week before the election, the Ohio State Buckeyes have a big football game against Penn State. The University of Florida Gators have a huge match up against the University of Georgia Bulldogs. If the election remains razor close, these games in these two key battleground states could affect who sits in the White House for the next four years. Can you imagine Ohio State head coach Urban Meyer getting a late night call from the Obama campaign suggesting a particular blitz package? Or maybe Romney has some advice for how the Gators can bottle up Georgia's running game. The decision of whether to punt or go for it on that crucial fourth down could affect the job prospects of more than just the football teams coaching staff." — Tyler Cowen and Kevin Grier, Slate.com

- Assume that a win in the two weeks before the election gives the incumbent an extra 2% of the vote
- What does this imply about the election?
- Not so much . . .
  - Diffusion: A $\pm 1\%$ shift of the vote in the county containing Columbus, Ohio, corresponds to a 0.1% shift in the state (i.e., 5000 votes)
  - Averaging: In the two weeks before the 2012 election, there were 4 major-college football games involving Ohio teams, and many more in Florida, Virginia, Colorado, etc.
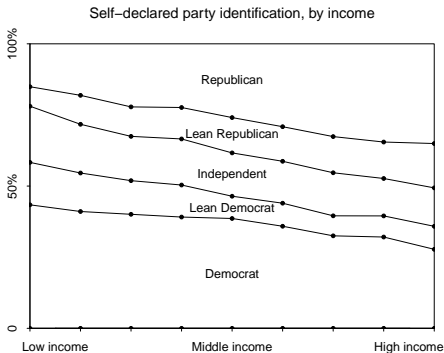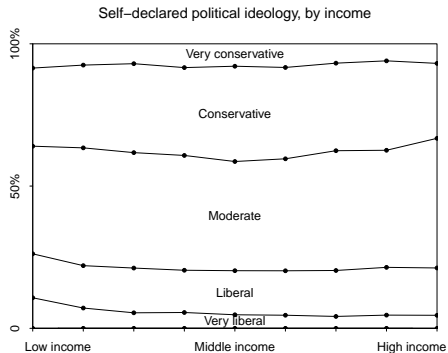
A married person having sexual relations with someone other than their spouse is...
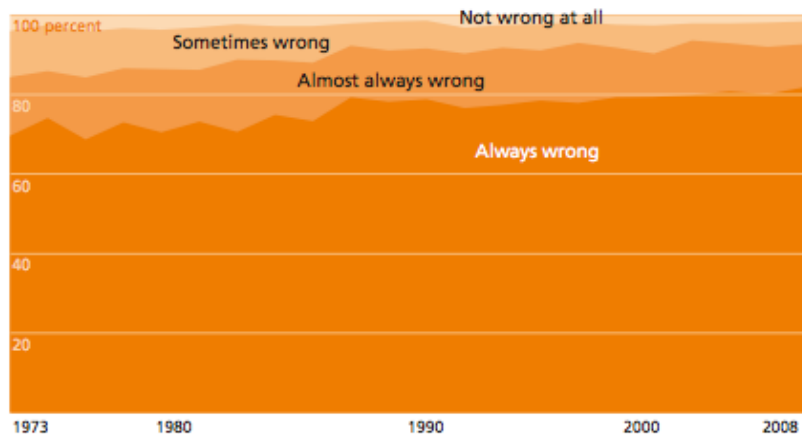
Source: General Social Survey (data collected biennially)

Multiple plots

Grid of plots

Republican vote in 2008 by education, among age/ethnic groups
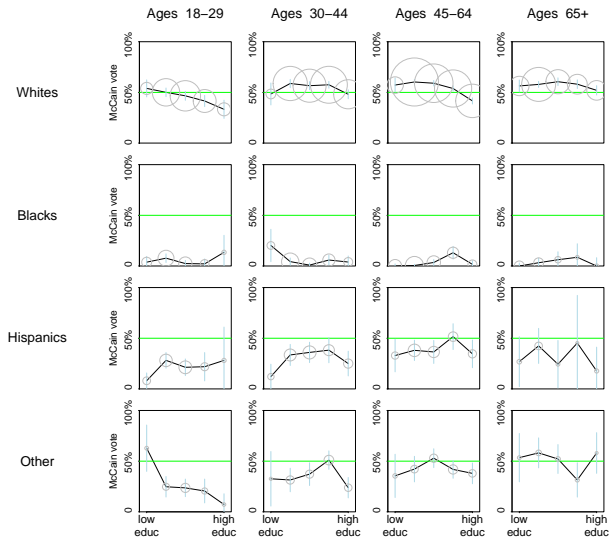
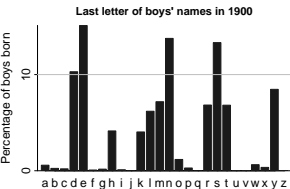Education categories: no high school, high school, some college, college grad, postgrad.
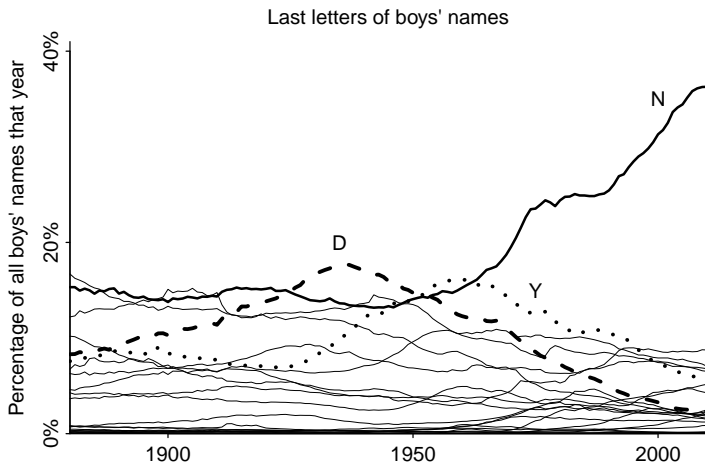
Even simpler . . .

Last letter of boys' names in 1906

Small multiples

Last letter of boys' names in 1906

Last letter of boys' names in 1956

Last letter of boys' names in 2006

Slightly prettier

Last letter of boys' names in 1900

Last letter of boys' names in 1950

Last letter of boys' names in 2010

Last letters of boys' names
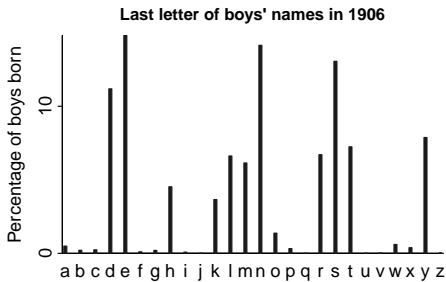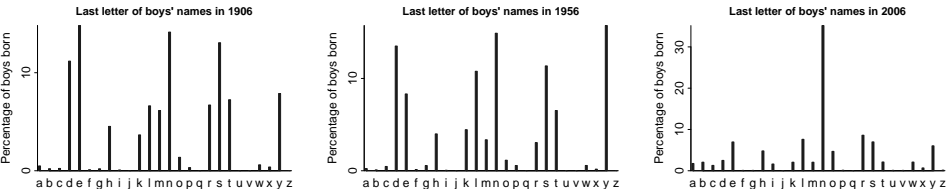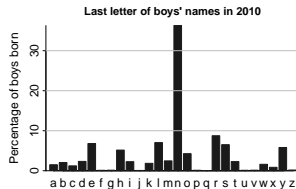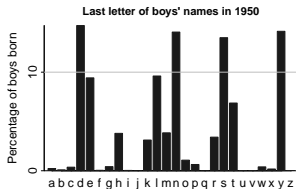
Real simple
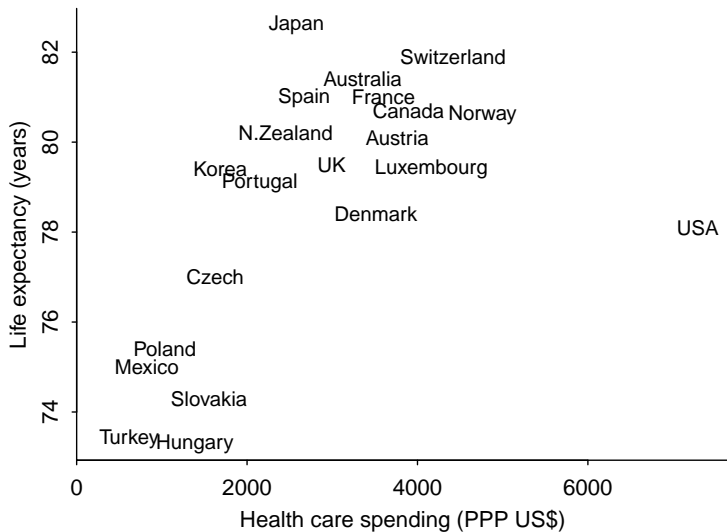
Raw data

Test scores in control classes   **Test scores in treated classes**

Grade 1 — control: mean = 69, sd = 13; treated: mean = 77, sd = 16

Grade 2 — control: mean = 93, sd = 12; treated: mean = 102, sd = 10
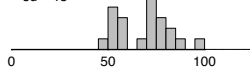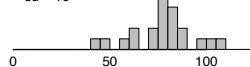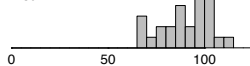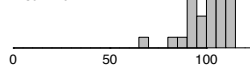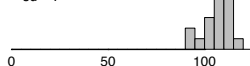
Grade 3 — control: mean = 106, sd = 7; treated: mean = 107, sd = 8
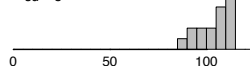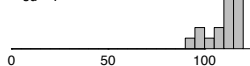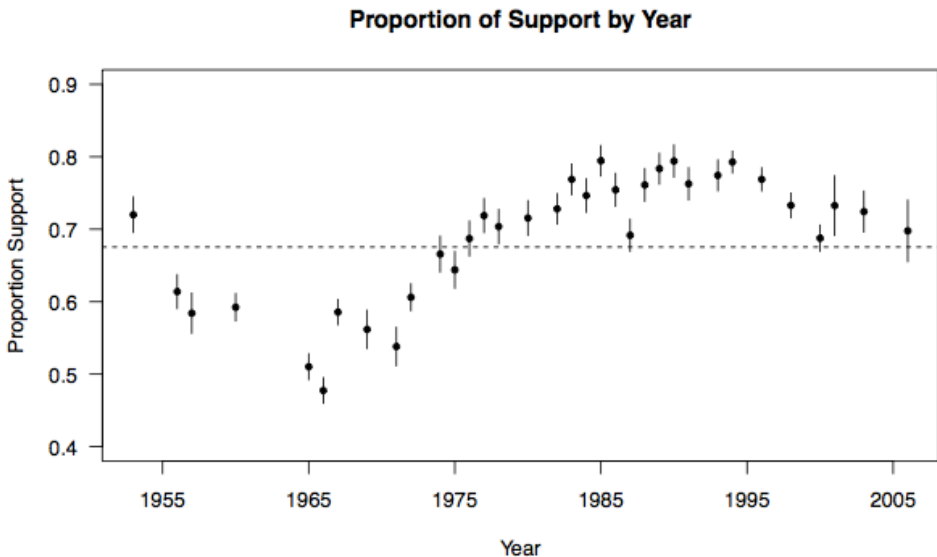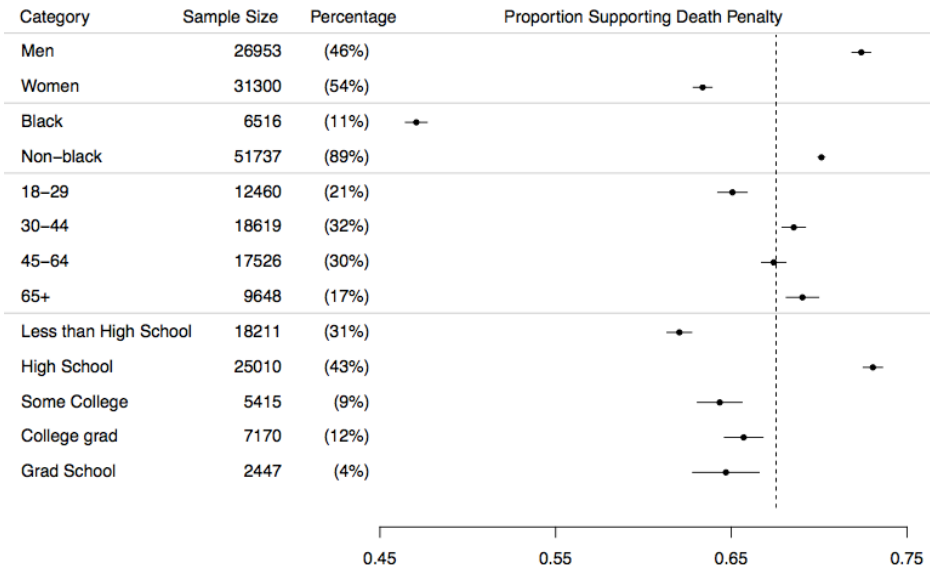
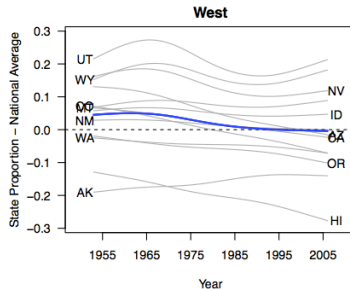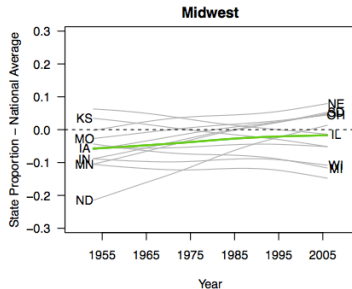Grade 4 — control: mean = 110, sd = 7; treated: mean = 114, sd = 4
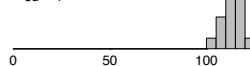
Re-expression saves space and adds clarity
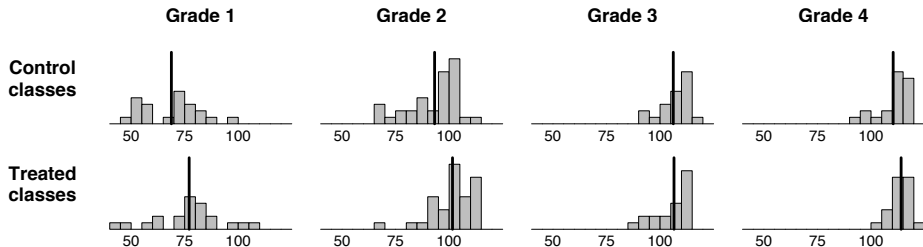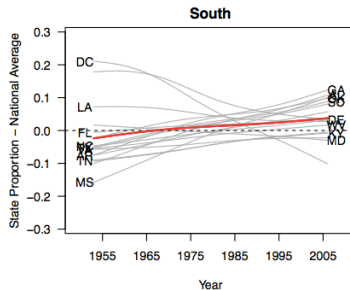
Death penalty time series

Proportion of Support by Year

National demographic breakdown

Category	Sample Size	Percentage	Proportion Supporting Death Penalty
Men	26953	(46%)	
Women	31300	(54%)	
Black	6516	(11%)	
Non–black	51737	(89%)	
18–29	12460	(21%)	
30–44	18619	(32%)	
45–64	17526	(30%)	
65+	9648	(17%)	
Less than High School	18211	(31%)	
High School	25010	(43%)	
Some College	5415	(9%)	
College grad	7170	(12%)	
Grad School	2447	(4%)	

0.45   0.55   0.65   0.75

Trends by state

State intercepts and slopes

State Intercepts

State Slopes

% above average in 1980

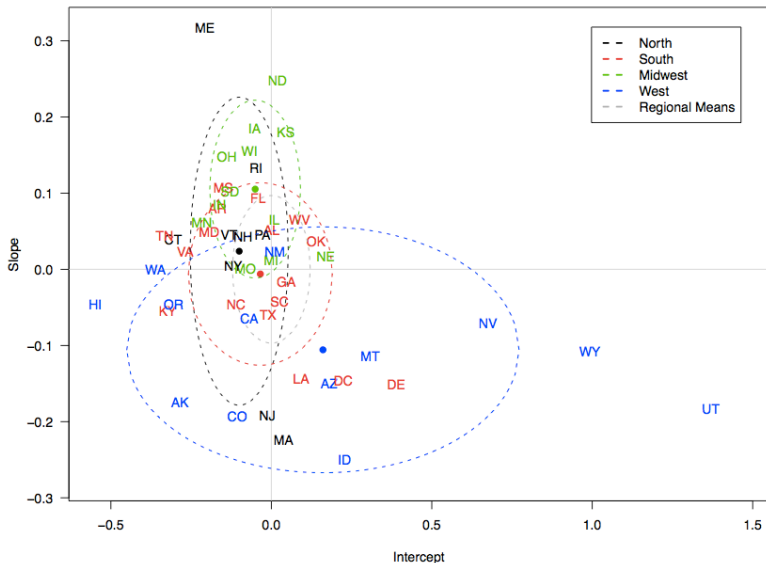−22%   −11%   0%   +11%   +22%

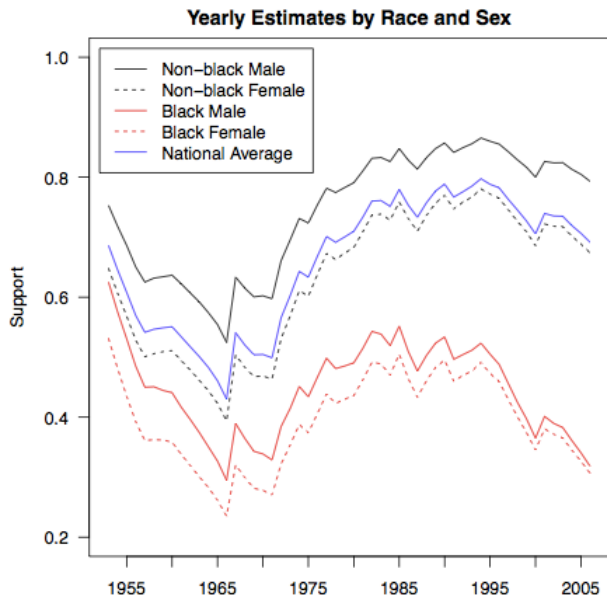% change per decade

−6.5%   −3.25%   0%   +3.25%   +6.5%
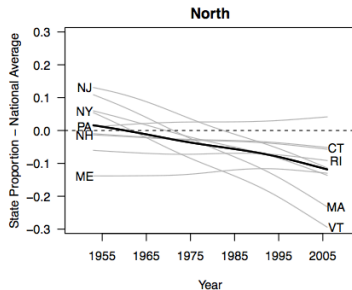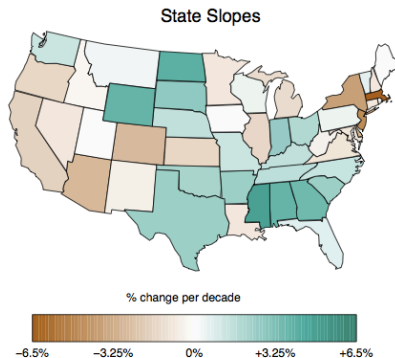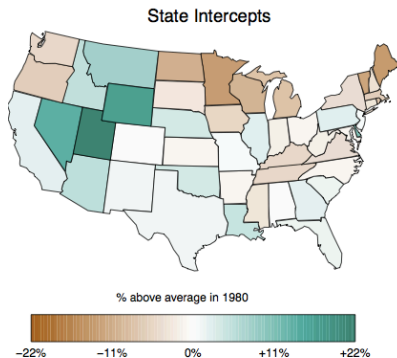
Unmodeled variation in state intercepts and slopes

Slopes vs. Intercepts by State (within region): State-specific random effects

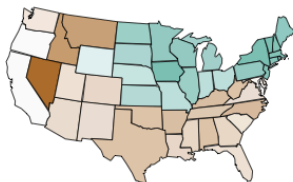# Trends by race and sex

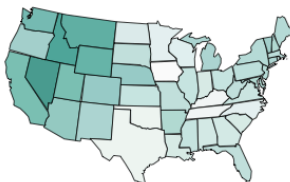**Yearly Estimates by Race and Sex**

Support among different education levels

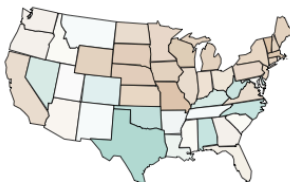Degree: Less than High School   Degree: High School   Degree: Some College

Degree: College Grad   Degree: Grad School
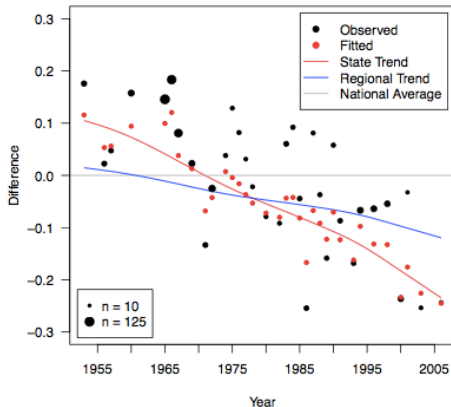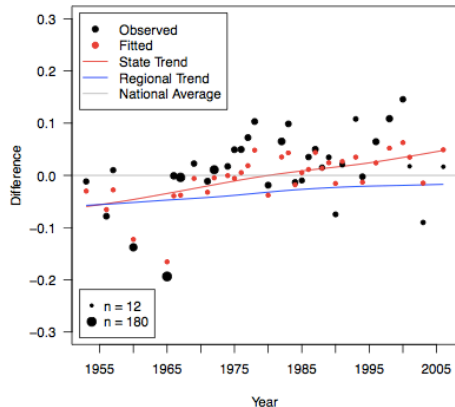
+25%
+12.5%
0%
−12.5%
−25%
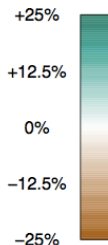
Close-up on two states (compared to U.S. avg)

Some recent examples of surveys in the news

- Big corporations are more popular than you might realize
- Trends in the economy
- Where are the larger-than-life athletes?
- Average personality differences between men and women
- Risk perception and political polarization
- Libertarians in space
- Surveys show Americans are populist class warriors, except when they aren't
- Gingrich or Romney as Obama opponent
- Atheists are less popular than rapists