Data 102 Lecture 13:

Causal inference I

Unit overview

- Lecture 13: Problems with associations
- Lecture 14: Randomized experiments
- Lecture 15: Observational studies

Lecture 13 overview

- Correlation and causation
  - Examples of mistaking correlation for causation
  - Spurious correlations
  - Explanations of association, confounders
- Quantifying association
- Two “statistical paradoxes”
  - Simpson’s paradox
  - The Red-Blue paradox

Simpson’s paradox

Kidney stones

	Treatment A	Treatment B
Failure	273	289
Success	77	61

From Charig et al. (1986)

Kidney stones

	Treatment A helps	Treatment B helps
All patients	**83%** (289 / 350)	78% (273 / 350)
Large kidney stones	69% (55 / 80)	**73%** (192 / 263)
Small kidney stones	87% (234 / 270)	**93%** (81 / 87)

Gender bias in Berkeley graduate school admissions

	All		Men		Women	
	Applicants	**Admitted**	**Applicants**	**Admitted**	**Applicants**	**Admitted**
Total	12,763	41%	8442	**44%**	4321	35%

" YES, ON THE SURFACE IT WOULD APPEAR TO BE SEX-BIAS BUT LET US ASK THE FOLLOWING QUESTIONS..."

From Wikipedia and Freedman, Pisani, Purves (2007)

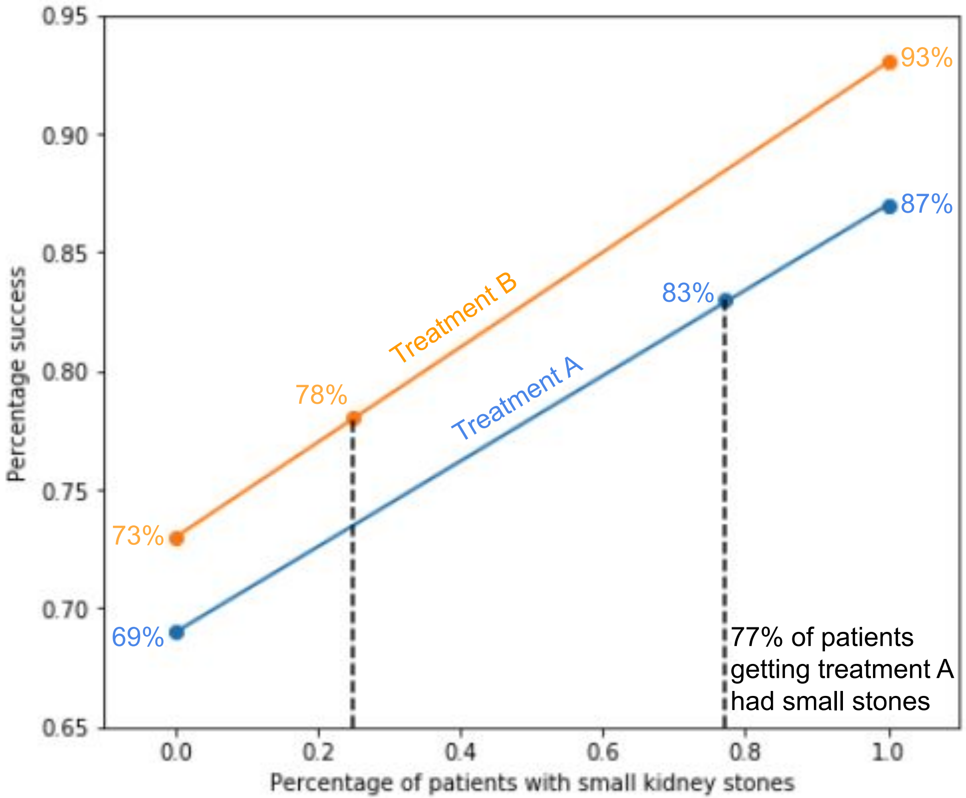
Gender bias in Berkeley graduate school admissions

Department	All		Men		Women	
	Applicants	Admitted	Applicants	Admitted	Applicants	Admitted
A	933	64%	***825***	62%	108	**82%**
B	585	63%	***560***	63%	25	**68%**
C	918	35%	325	**37%**	***593***	34%
D	792	34%	417	33%	375	**35%**
E	584	25%	191	**28%**	***393***	24%
F	714	6%	373	6%	341	**7%**

From Wikipedia

Explaining Simpson's paradox

	Treatment A helps	Treatment B helps
Large kidney stones	69% (55 / 80)	**73%** (192 / 263)
Small kidney stones	87% (234 / 270)	**93%** (81 / 87)
All patients	**83%** (289 / 350)	78% (273 / 350)

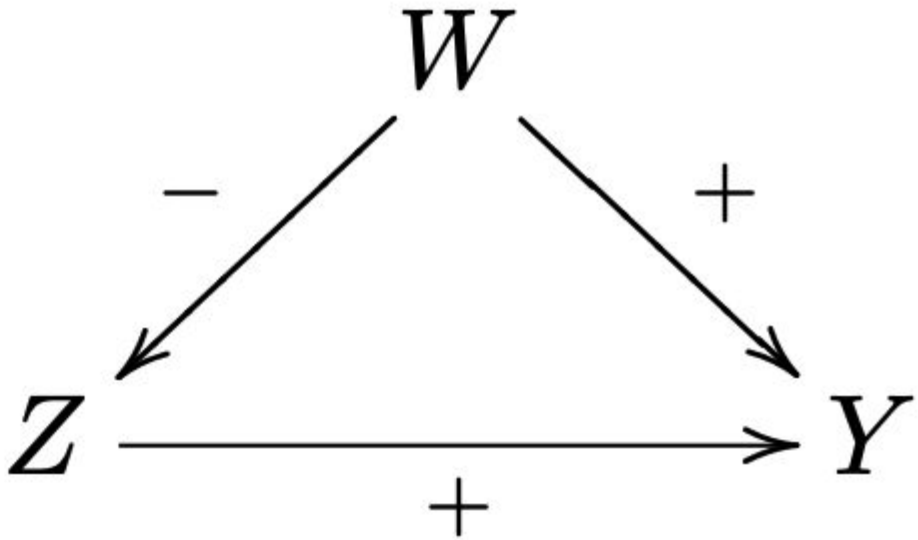


Explaining Simpson's paradox

Z = 1(Treatment B)

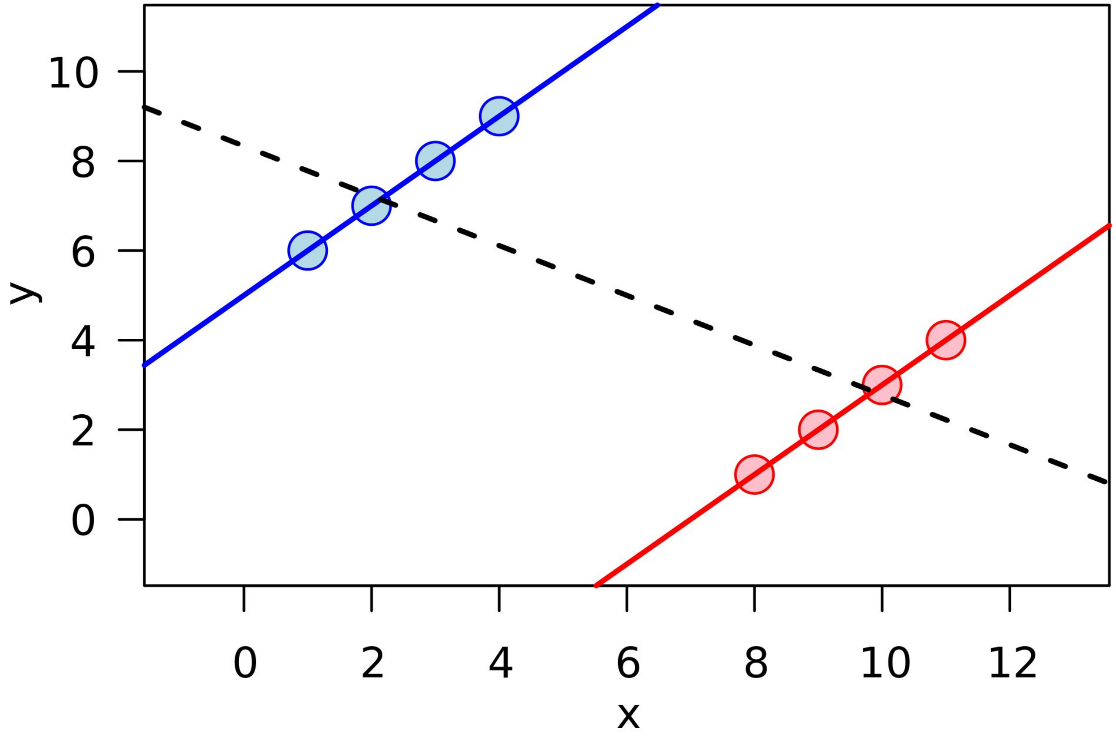
Y = 1(Success)

W = 1(Small stones)



From Peng Ding's notes

Simpson's paradox for continuous variables



X = years of employment

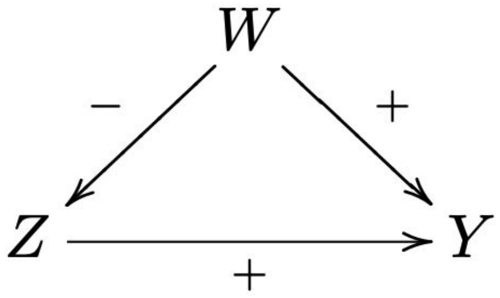
Y = salary

Color:

- Blue = CEOs
- Red = janitors

From Wikipedia

Simpson's paradox and causal inference



Simpson's paradox is about association. Nothing causal yet.

But based on the evidence, it is intuitive to choose Treatment B. Why?

- Treatment comes after the kidney stones chronologically, so the graph also reflects (partial) causal structure.
- Conditioning on W (kidney stone size) removes its influence
- After conditioning, if we believe that there are no significant differences between patients getting treatment A vs those getting treatment B, then any difference in success rate is purely due to the choice of treatment

May still have unobserved important variables...

Department	All		Men		Women	
	Applicants	Admitted	Applicants	Admitted	Applicants	Admitted
A	933	64%	***825***	62%	108	**82%**
B	585	63%	***560***	63%	25	**68%**
C	918	35%	325	**37%**	***593***	34%
D	792	34%	417	33%	375	**35%**
E	584	25%	191	**28%**	***393***	24%
F	714	6%	373	6%	341	**7%**

Possibility: For a given department, women applicants are more competent than male applicants

In which case, there is still gender bias despite parity in admission rates

From Wikipedia

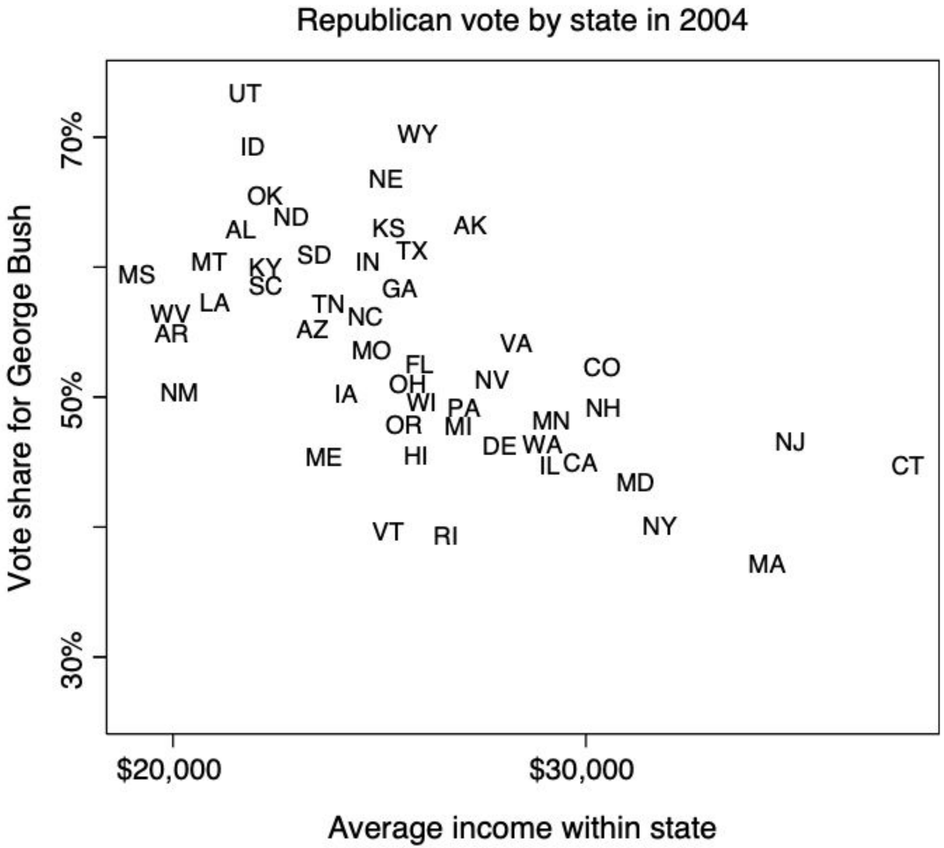
The Red-Blue paradox

From Gelman (2009)

OK, but here's the fact that nobody ever, ever mentions—Democrats win rich people. Over $100,000 in income, you are likely more than not to vote for Democrats. People never point that out. Rich people vote liberal. I don't know what that's all about.

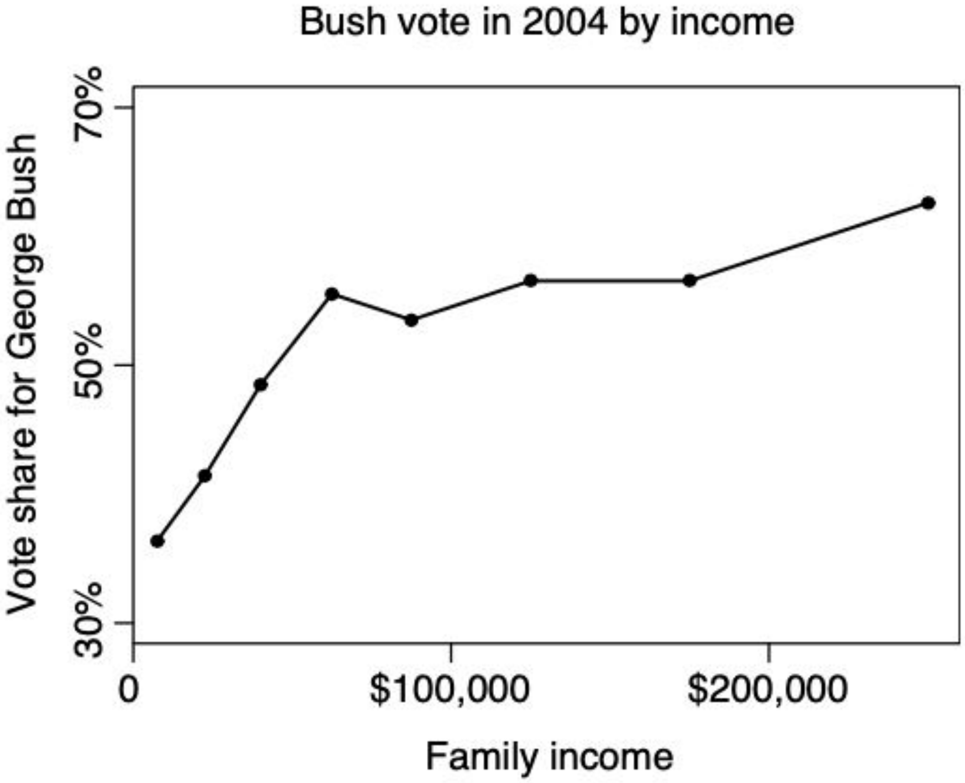
—Tucker Carlson, 2007

Support for Republicans is negatively correlated with income for states



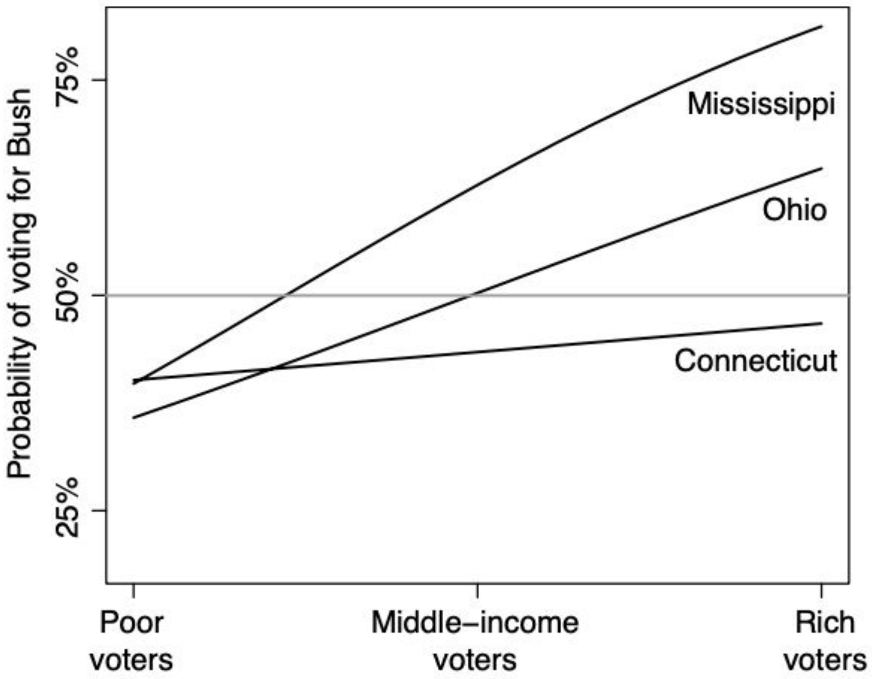
From Gelman (2009)

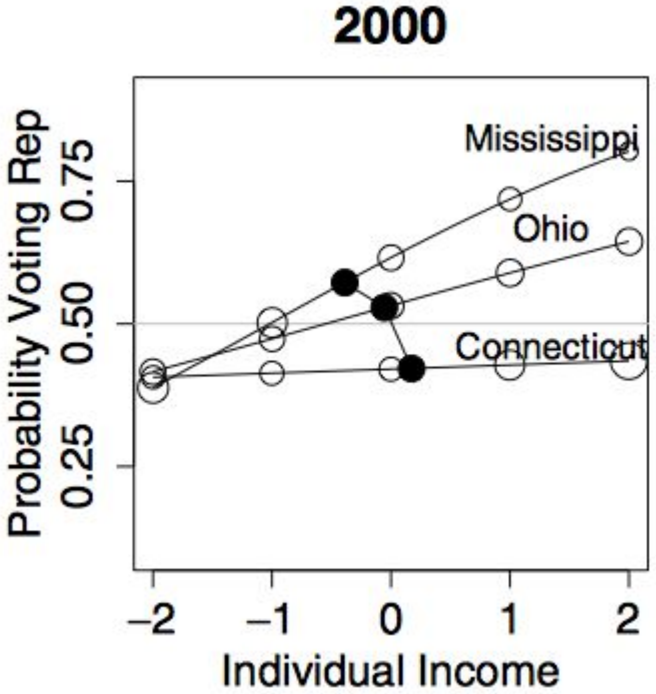
Support for Republicans is positively correlated with income for individuals



From Gelman (2009)

Resolving the paradox





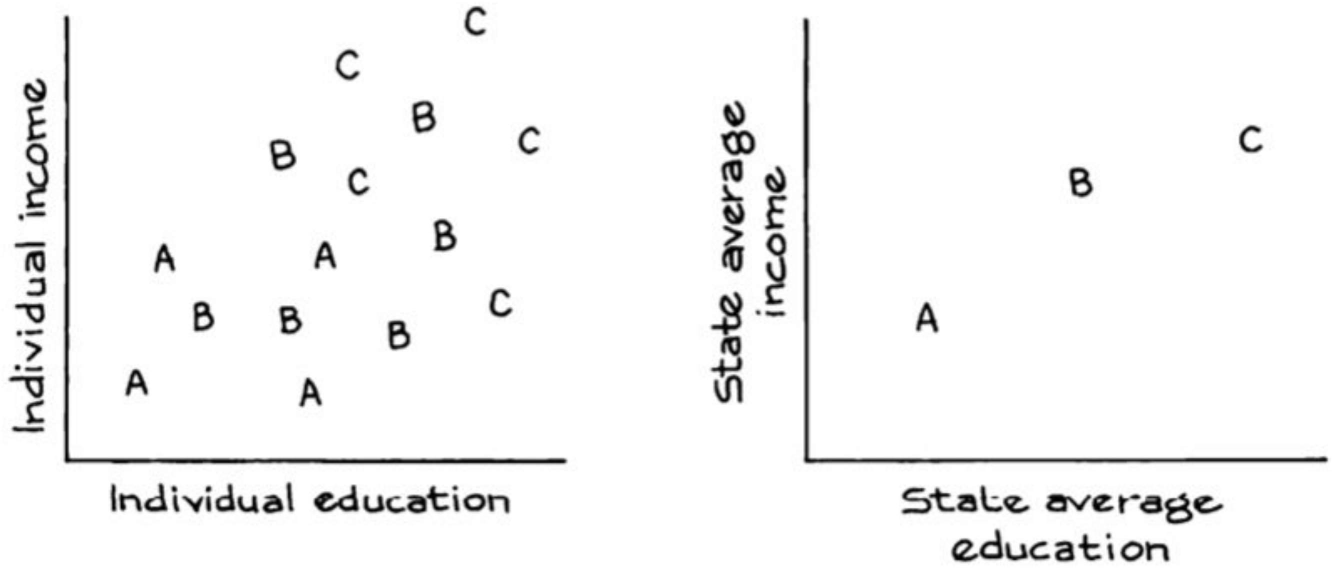
From Gelman (2009) and Gelman (2014)

Misinterpreting the Red-Blue paradox is an e.g of an ecological fallacy

Ecological fallacy

When inferences about the nature of *individuals* are deduced from inferences about the *group* to which those individuals belong

Ecological correlations usually too big



From Freedman, Pisani, Purves. (2007)