SOME FUN TOPICS IN PROBABILITY

Conditional probability

Independence and the math of winning streaks

Paradoxes, mixtures, and the rule of total probability

PROBABILITY NOTATION

- Our notation: P(some event) = Number between 0 and 1
- 1 means certain, 0 means impossible
- For example:

  P(coin lands heads) = 0.5

  P(flight departs on time) = 0.79

  P(rain in Austin) = 0.2

  P(rain in Dublin) = 0.4

  P(cold day in Hell) = 0.0000000001

  etc.

CONDITIONAL PROBABILITY

- A conditional probability is the chance that one thing (A) happens, given that some other thing (B) has already happened.
- Our notation: P(A | B) — "Probability of A, given B"
- Conditional probabilities reflect our uncertainty in light of partial knowledge:

  P(rain this afternoon | cloudy this morning)

  P(UT beats OU | UT ahead by a touchdown at halftime)

  P(accepted to medical school | college GPA > 3.6)

  etc.

EXAMPLE: RECOMMENDER SYSTEMS

- Instagram: P(follow @LeoMessi | follow @Cristiano)
- Amazon: P(buy organic dog food | buy GPS dog collar)
- Netflix: P(watch *Tinker Tailor Soldier Spy* | watch *Sherlock*)

EXAMPLE: RECOMMENDER SYSTEMS

- Instagram: P(follow @LeoMessi | follow @Cristiano)
- Amazon: P(buy organic dog food | buy GPS dog collar)
- Netflix: P(watch *Tinker Tailor Soldier Spy* | watch *Sherlock*)
- YouTube
- Google
- Spotify
- *The New York Times*
- Twitch
- Facebook
- EBay

KEY FACT:
$P(A \mid B) \neq P(B \mid A)$

*perhaps the single most important fact to remember about conditional probabilities!

A: YOU CAN DRIBBLE A BASKETBALL
B: YOU PLAY IN THE NBA


P(CAN DRIBBLE BASKETBALL | PLAYS IN NBA) = 1


P(PLAYS IN NBA| CAN DRIBBLE BASKETBALL) ≈ 0

KEY FACT:
$P(A \mid B) \neq P(B \mid A)$

Moral of the story:

Always be specific about what's on the left-hand side, and what's on the right-hand side.

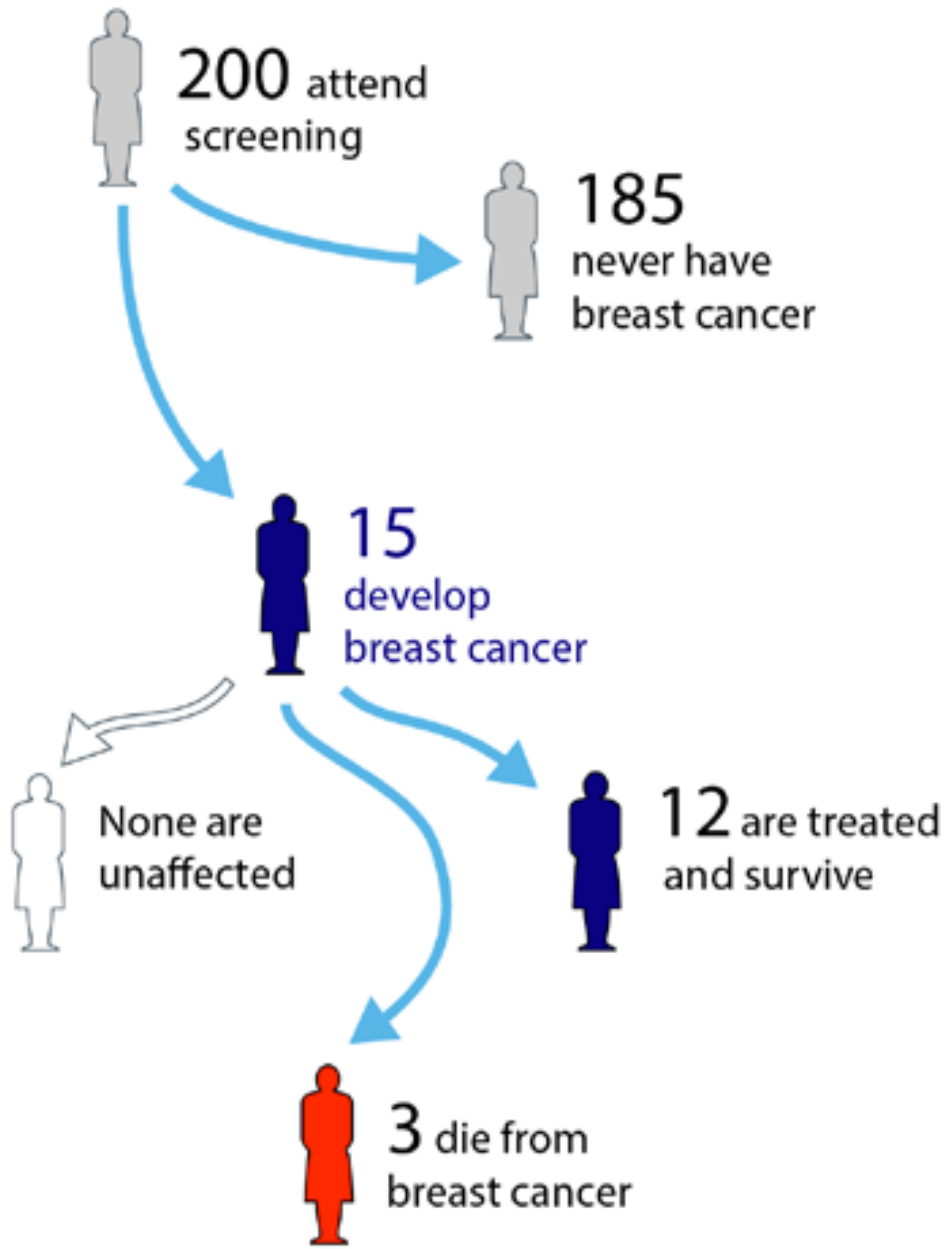
PROBABILITY TREES

- P(cancer) = 15/200
- P(die, cancer) = 3/200
- P(die | cancer) = 3/15
- In general, we can estimate P(A | B) as:

$$P(A \mid B) = \frac{\text{Frequency of A and B both happening}}{\text{Frequency of B happening}}$$

- Probability trees are a great way to make probability more intuitive for nonexperts.

200 women between 50 and 70 who attend screening



MULTIPLICATION RULE

- The multiplication rule just expresses this idea in general terms:

$$P(A \mid B) = \frac{P(A, B)}{P(B)}$$

MULTIPLICATION RULE

- We can also use the alternate version below if we want to go in reverse, from a conditional probability to a joint probability.
- It says the same thing, with the terms re-arranged:

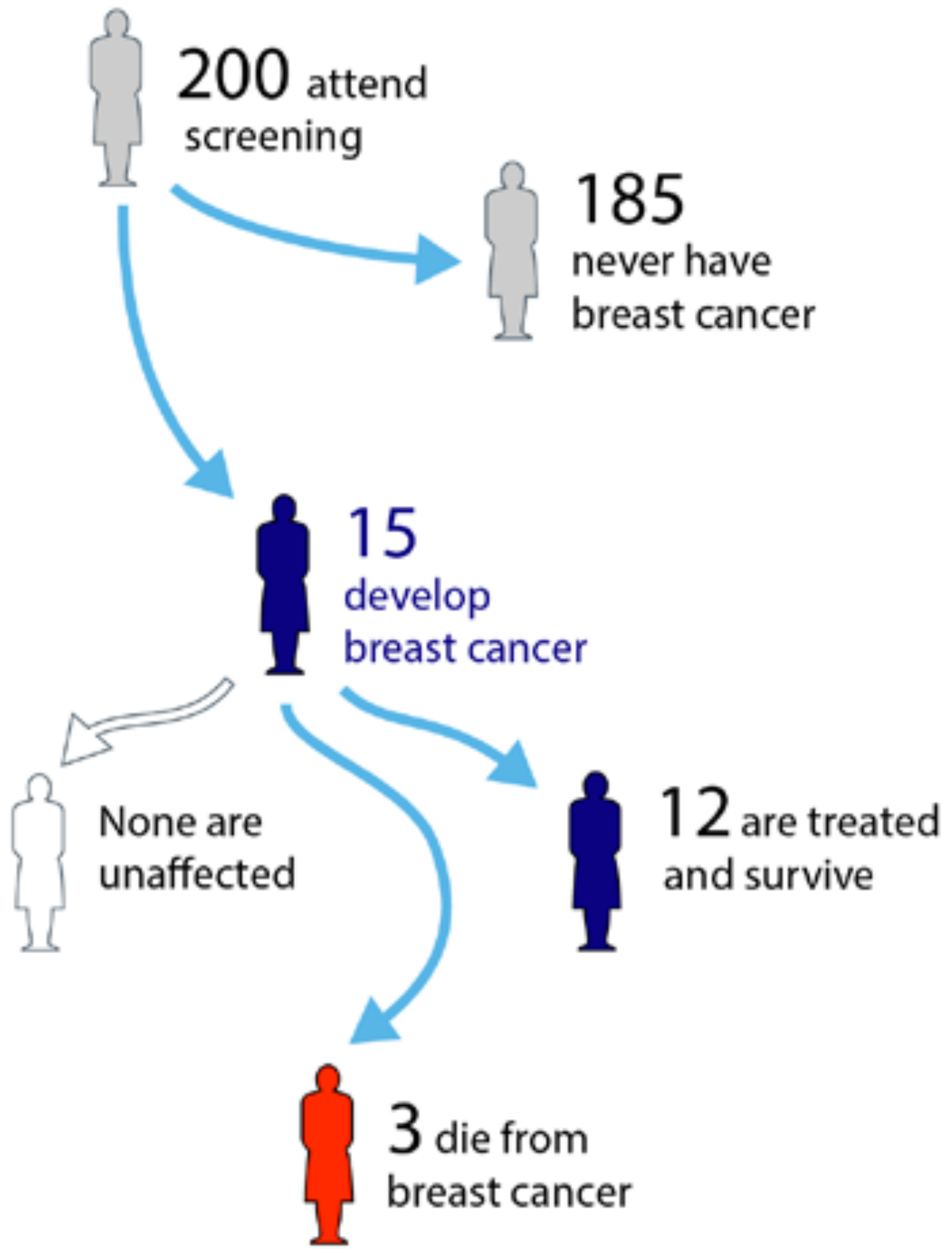
$$P(A, B) = P(A \mid B) \cdot P(B)$$

EXAMPLE: MAMMOGRAMS

- P(cancer) = 15/200
- P(die, cancer) = 3/200
- P(die | cancer) = 3/15
- Using the multiplication rule:

$$P(\text{die} \mid \text{cancer}) = \frac{P(\text{die,cancer})}{\text{cancer}} = \frac{3/200}{15/200}$$

200 women between 50 and 70 who attend screening

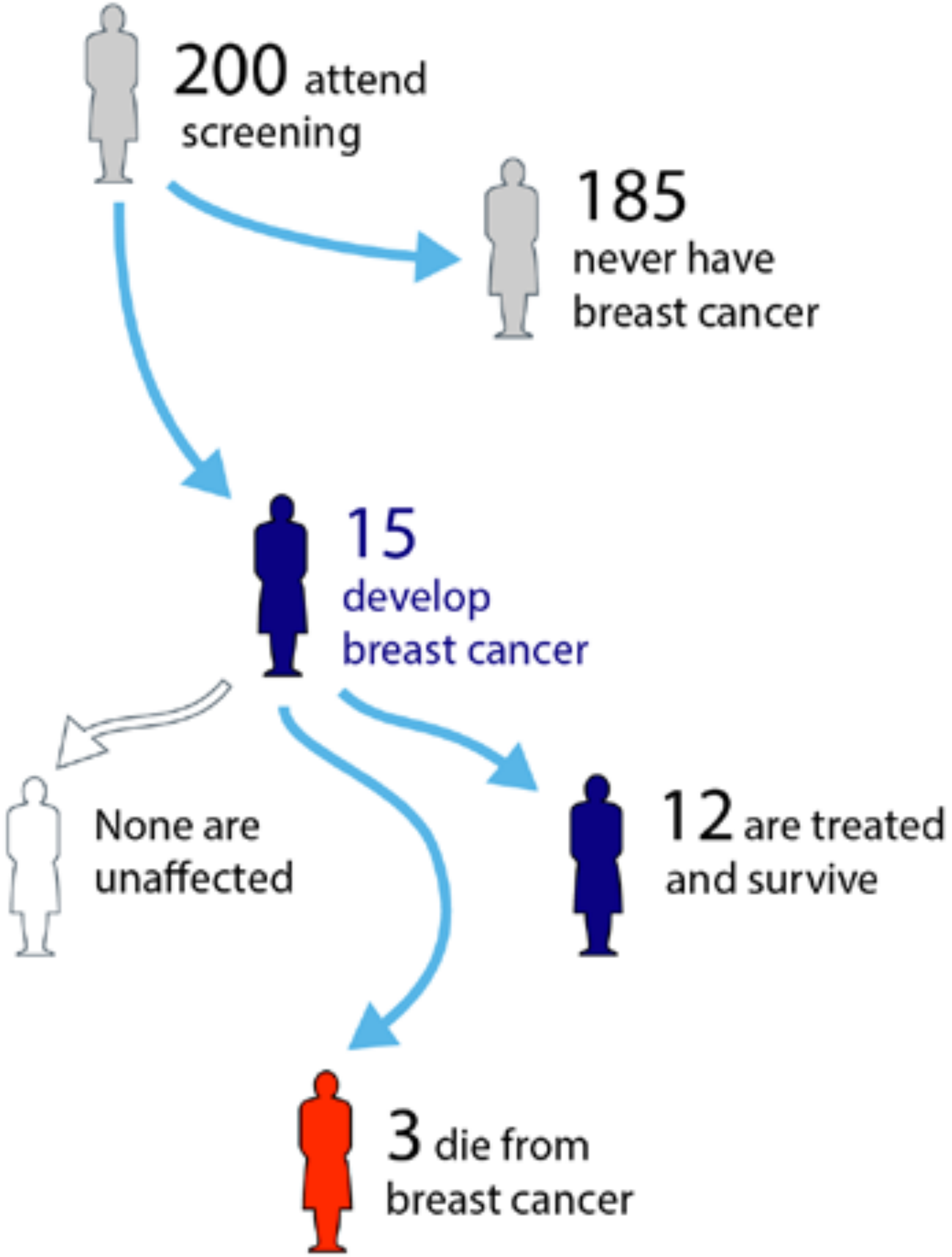


EXAMPLE: MAMMOGRAMS

- P(cancer) = 15/200
- P(die, cancer) = 3/200
- P(die | cancer) = 3/15
- Or the re-arranged version:

$$
\begin{aligned}
P(\text{die, cancer}) &= P(\text{die} \mid \text{cancer}) \cdot P(\text{cancer}) \\
&= \frac{3}{15} \cdot \frac{15}{200} \\
&= 3/200
\end{aligned}
$$

200 women between 50 and 70 who attend screening



PROBABILITIES FROM CONTINGENCY TABLES


BAND OF BROTHERS


SAVING PRIVATE RYAN

PRETEND YOU'RE NETFLIX...

- What we want to know (A): Maria likes *Saving Private Ryan?*
- What we already know (B): Maria likes *Band of Brothers*
- Key question: what is P(A | B)?
- Answer: go to the data and use the multiplication rule!

$$P(A \mid B) = \frac{\text{Frequency of A and B both happening}}{\text{Frequency of B happening}}$$

Subscriber	Liked Saving Private Ryan?	Liked Band of Brothers?
1. Ben Armstrong	Yes	Yes
2. Alejandra Contreras	Yes	Yes
...	...	...
99. Rashid Tannous	No	No
100. Anna Yeo	Yes	No

	Liked *Saving Private Ryan*	Didn't like it
Liked *Band of Brothers*	56	6
Didn't like it	14	24

$$P(\text{Likes SPR} \mid \text{Likes BB}) = \frac{56}{56+6} \approx 0.9$$

INDEPENDENCE

- Two events A and B are said to be independent if P(A | B) = P(A)
- Equivalently, A and B are independent if P(B | A) = P(B)
- In words, A and B convey no information about each other:

    P(coin 2 lands heads | coin 1 lands heads) = P(coin 2 lands heads)

    P(stock market up | bird poops on your car) = P(stock market up)

    P(God exists | Longhorns win title) = P(God exists)

    etc.

INDEPENDENCE

- So if A and B are independent, then:

  $P(A, B) = P(A) \cdot P(B \mid A) = P(A) \cdot P(B)$

- For this reason, independence is often something we *choose to assume* to make probability calculations easier.
- This assumption might be sensible:

  P(flip 1 lands heads, flip 2 lands heads) = P(flip 1 lands heads) • P(flip 2 lands heads)

  P(AAPL up today, AAPL up tomorrow) = P(AAPL up today) • P(AAPL up tomorrow)

- Or it might not…

  P(rain, high winds) ≠ P(rain) • P(high winds)

  P(sib 1 colorblind, sib 2 colorblind) ≠ P(sib 1 colorblind) • P(sib 2 colorblind)


103


ZTE
28


56

THE STREAK

- On 11/17/2014, the UConn women's basketball team lost a close road game to Stanford.
- They beat Creighton in their next game, a week later…
- Then they beat Charleston, and Notre Dame, and UCLA, and Duke…

THE STREAK

- On 11/17/2014, the UConn women's basketball team lost a close road game to Stanford.
- They beat Creighton in their next game, a week later...
- Then they beat Charleston, and Notre Dame, and UCLA, and Duke...
- And they kept winning... 111 times in a row!

THE STREAK

- On 11/17/2014, the UConn women's basketball team lost a close road game to Stanford.
- They beat Creighton in their next game, a week later…
- Then they beat Charleston, and Notre Dame, and UCLA, and Duke…
- And they kept winning… 111 times in a row!
- Then on March 31, 2017, they finally met their match…


FINAL FOUR
FOUR IT ALL
2017 NCAA WOMEN'S FINAL FOUR
FINAL FOUR
UCONN
Lost each of last 5 overtime games (Last: Nov. 17, 2014 at Stanford)
2 Mississippi St 64
1 UConn 64
OT 12.3
33-4
BONUS
36-0
BONUS
POSS: MISSISSIPPI ST

HOW IMPROBABLE WAS THE STREAK?

- During the 2010s, the UConn Huskies had a 327-17 record (95%).
- Suppose successive games are independent, with a 0.95 probability of victory.

  (Remember: independence is something we can *choose to assume* to make calculations easier.)
- What are the chances of a 111-game winning streak? To the board!

"WINNING STREAK" RULE: A SUMMARY

- Suppose the probability of a single win is P and that each game is independent.
- Then the probability of N wins in a row is P^N: that is, P raised to the Nth power.
- "Games" and "wins" need not involve sports:

  A mutual-fund manager beats the market 15 years in a row.

  A student attends school 180 days in a row without catching a cold.

  A World War II pilot goes 25 missions in a row without getting shot down.

  etc.
- But beware! <u>The winning streak rule is misapplied all the time.</u>
- See the excerpt from "AIQ" on the class website.

PARADOXES, MIXTURES, AND THE RULE OF TOTAL PROBABILITY

PARADOX 1

	Low-risk (easier)	High-risk (harder)	Overall
Senior doctor	0.052	0.127	0.076
Junior doctor	0.067	0.155	0.072

Complication rates across 3,690 deliveries at a large maternity hospital in Cambridge, UK

	Low-risk (easier)	High-risk (harder)	Overall
Senior doctor	0.052	0.127	0.076
Junior doctor	0.067	0.155	0.072

Complication rates across 3,690 deliveries at a large maternity hospital in Cambridge, UK

THE PARADOX

- **Senior doctors are…**
  - **better at easy cases…**
  - **better at hard cases…**
  - **yet worse overall.**
- **This is an example of *Simpson's paradox*. How is it possible?**

PARADOX 2

Rank	State	Median income	2016 winner
1	Washington, D.C.	$85,203	Clinton
2	Maryland	$83,242	Clinton
3	New Jersey	$81,740	Clinton
4	Hawaii	$80,212	Clinton
5	Massachusetts	$79,835	Clinton
6	Connecticut	$76,348	Clinton
7	California	$75,277	Clinton
8	New Hampshire	$74,991	Clinton
9	Alaska	$74,346	Trump
10	Washington	$74,073	Clinton

Ten richest states with 2016 electoral college result

Rank	State	Median income	2016 winner
42	Tennessee	$52,375	Trump
43	South Carolina	$52,306	Trump
44	Oklahoma	$51,924	Trump
45	Kentucky	$50,247	Trump
46	Alabama	$49,861	Trump
47	Louisiana	$47,905	Trump
48	New Mexico	$47,169	Clinton
49	Arkansas	$47,062	Trump
50	Mississippi	$44,717	Trump
51	West Virginia	$44,097	Trump

Ten poorest states with 2016 electoral college result

HIGH-INCOME STATES VOTE BLUE
LOW-INCOME STATES VOTE RED

"Farmers, factory workers, Truck Drivers, waitresses..."

vs.

"the know-it-alls of Manhattan and Malibu... who lord it over the peasantry with their fancy college degrees"

"Average Americans, humble, long-suffering, working hard, who buy their coffee already ground"

vs.

"The wealthy latte-swilling liberal elite"

"Real Americans, with a lawnmower in the garage and a flag on the front stoop"

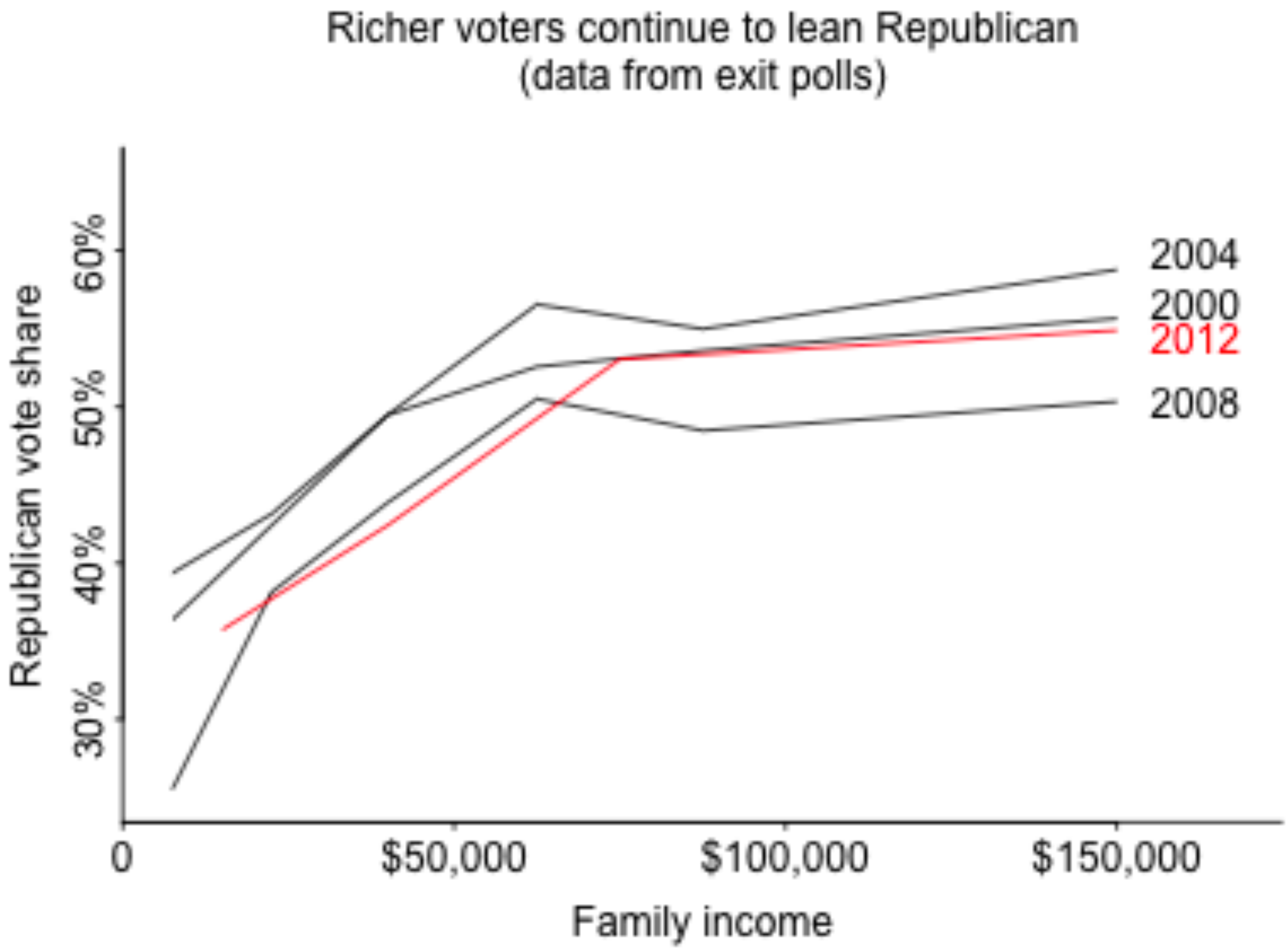
vs.

"Wealthy condo-dwellers with contempt for those who feel chills up their spines at 'The Star Spangled Banner.'"

AND YET...

	under $50K		over $50K	
	Dem.	Rep.	Dem.	Rep.
2004	0.55	0.44	0.43	0.56
2008	0.60	0.38	0.49	0.49
2012	0.54	0.44	0.44	0.54
2016	0.52	0.41	0.47	0.49

Presidential vote share by personal income



Presidential vote share by family income

THE PARADOX

- For states:
  - higher income means more likely to vote Democrat.
  - lower income means more likely to vote Republican.
- Yet for people:
  - higher income means more likely to vote Republican.
  - lower income means more likely to vote Democrat.
- How is this possible?

BACK TO PARADOX 1

	Low-risk (easier)	High-risk (harder)	Overall
Senior doctor	0.052 (231)	0.127 (102)	0.076 (315)
Junior doctor	0.067 (3169)	0.155 (206)	0.072 (3375)

Complication rates and sample sizes across 3,690 deliveries at a large maternity hospital in Cambridge, UK

THE RULE OF TOTAL PROBABILITY

RULE OF TOTAL PROBABILITY

- In words: the probability of an event is the sum of the probabilities for all the different ways in which that event can happen:

  P(rain) = P(rain, wind) + P(rain, no wind)

  P(complication) = P(complication, easy delivery) + P(complication, hard delivery)

RULE OF TOTAL PROBABILITY

- In words: the probability of an event is the sum of the probabilities for all the different ways in which that event can happen:

  P(rain) = P(rain, wind) + P(rain, no wind)

  P(complication) = P(complication, easy delivery) + P(complication, hard delivery)

- Suppose that B_1, B_2, … B_N are mutually exclusive events whose probabilities sum to 1:

$$P(B_i, B_j) = 0 \quad (i \neq j) \quad \text{and} \quad \sum_{i=1}^{N} P(B_i) = 1$$

- Then for any event A:

$$P(A) = \sum_{i=1}^{N} P(A, B_i) = \sum_{i=1}^{N} P(A \mid B_i) \cdot P(B_i)$$

The second part of the equation comes from the multiplication rule: P(A, B) = P(A | B) • P(B)

RULE OF TOTAL PROBABILITY

- So, for example, the overall (total) probability of a complication is:

$$
\begin{aligned}
P(\textbf{comp}) &= P(\textbf{comp, easy}) + P(\textbf{comp, hard}) \\
&= P(\textbf{easy}) \cdot P(\textbf{comp | easy}) + P(\textbf{hard}) \cdot P(\textbf{comp | hard})
\end{aligned}
$$

RULE OF TOTAL PROBABILITY

	Low-risk (easier)	High-risk (harder)	Overall
Senior doctor	0.052 (231)	0.127 (102)	0.076 (315)
Junior doctor	0.067 (3169)	0.155 (206)	0.072 (3375)

- So, for example, the overall (total) probability of a complication is:

$$\begin{aligned} P(\textbf{comp}) &= P(\textbf{comp, easy}) + P(\textbf{comp, hard}) \\ &= P(\textbf{easy}) \cdot P(\textbf{comp} \mid \textbf{easy}) + P(\textbf{hard}) \cdot P(\textbf{comp} \mid \textbf{hard}) \end{aligned}$$

- For senior doctors:

$$P(\textbf{comp}) = \left(\frac{231}{231 + 102}\right) \cdot 0.052 + \left(\frac{102}{231 + 102}\right) \cdot 0.127 = 0.076$$

- For junior doctors:

$$P(\textbf{comp}) = \left(\frac{3169}{3169 + 206}\right) \cdot 0.067 + \left(\frac{3169}{3169 + 206}\right) \cdot 0.155 = 0.072$$

PARADOX 1 RESOLVED

- Senior doctors are…

  better at easy cases and better at hard cases…

  yet worse overall.

- This is an example of *Simpson's paradox*. Here's how it's possible:

  P(comp | easy) and P(comp | hard) are both lower for senior doctors…

  yet senior doctors work fewer easy cases: P(easy) is lower in the mixture!

- Moral of the story:

  Make sure you're asking the right question!

  Always be sensitive to whether probabilities are conditional or unconditional (overall/total), and which type of probability makes more sense for your situation.

PARADOX 2 AGAIN...

- For states:

    higher income means more likely to vote Democrat.

    lower income means more likely to vote Republican.

- Yet for people:

    higher income means more likely to vote Republican.

    lower income means more likely to vote Democrat.

- How is this possible?

PARADOX 2 AGAIN...

- For states:

    higher income means more likely to vote Democrat.

    lower income means more likely to vote Republican.

- Yet for people:

    higher income means more likely to vote Republican.

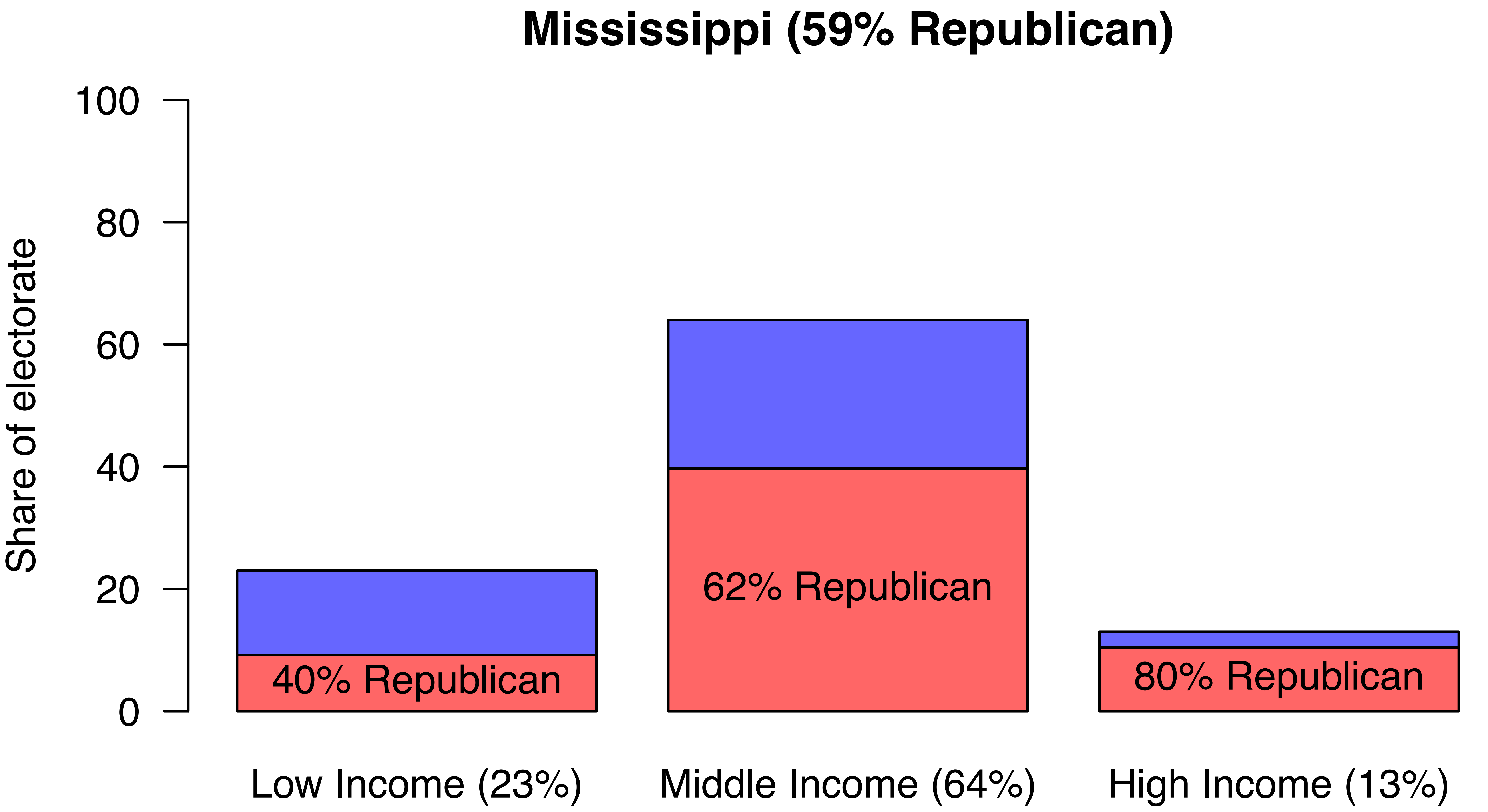
    lower income means more likely to vote Democrat.

- How is this possible? Use the rule of total probability! Let's compare two states:

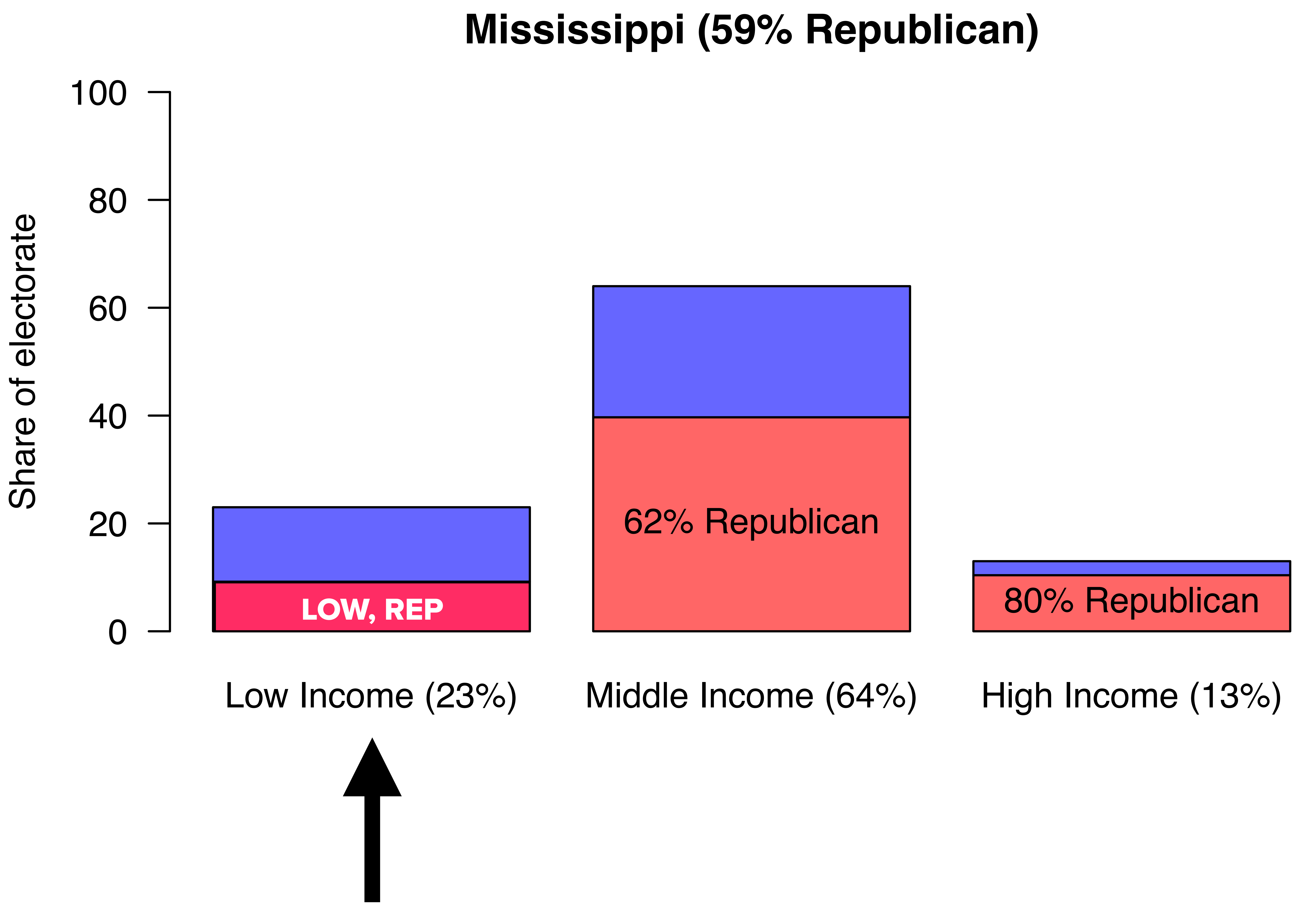
    Mississippi

    Connecticut


Mississippi (59% Republican)
Share of electorate
100
80
60
40
20
0
40% Republican
62% Republican
80% Republican
Low Income (23%)
Middle Income (64%)
High Income (13%)



Mississippi (59% Republican)
Share of electorate
100
80
60
40
20
0
40% Republican
62% Republican
80% Republican
Low Income (23%)
Middle Income (64%)
High Income (13%)


Mississippi (59% Republican)
Share of electorate
100
80
60
40
20
0
LOW, REP
62% Republican
80% Republican
Low Income (23%)
Middle Income (64%)
High Income (13%)


Mississippi (59% Republican)
Share of electorate
100
80
60
40
20
0
40% Republican
62% Republican
80% Republican
LOW, REP
0.40*0.23
Low Income (23%)
Middle Income (64%)
High Income (13%)


Mississippi (59% Republican)
Share of electorate
100
80
60
40
20
0
40% Republican
MIDDLE, REP
80% Republican
LOW, REP
0.40*0.23
Low Income (23%)
Middle Income (64%)
High Income (13%)


Mississippi (59% Republican)
Share of electorate
100
80
60
40
20
0
40% Republican
62% Republican
80% Republican
Low Income (23%)
Middle Income (64%)
High Income (13%)
MIDDLE, REP
LOW, REP
0.62*0.64
0.40*0.23

Mississippi (59% Republican)
Share of electorate
100
80
60
40
20
0
40% Republican
62% Republican
HIGH, REP
MIDDLE, REP
LOW, REP
0.62*0.64
0.40*0.23
Low Income (23%)
Middle Income (64%)
High Income (13%)


Mississippi (59% Republican)
Share of electorate
100
80
60
40
20
0
40% Republican
62% Republican
80% Republican
HIGH, REP
MIDDLE, REP
LOW, REP
0.80*0.13
0.62*0.64
0.40*0.23
Low Income (23%)
Middle Income (64%)
High Income (13%)

$$P(A) = \sum_{i=1}^{N} P(A, B_i)$$

$$= \sum_{i=1}^{N} P(A \mid B_i) \cdot P(B_i)$$

Mississippi (59% Republican)

Connecticut (44% Republican)
100
80
60
40
20
0
Share of electorate
39% Republican
43% Republican
47% Republican
Low Income (15%)
Middle Income (45%)
High Income (40%)

MISSISSIPPI AND CONNECTICUT

- Here’s P(Republican | income) for each state:

	Low-income	Middle-income	High-income
Connecticut	0.39	0.43	0.47
Mississippi	0.40	0.62	0.80

- Conclusion:

  Income differences cannot explain why CO is blue and MS red. Why?

  It must be something else. What might that something else be?

THE ECOLOGICAL FALLACY

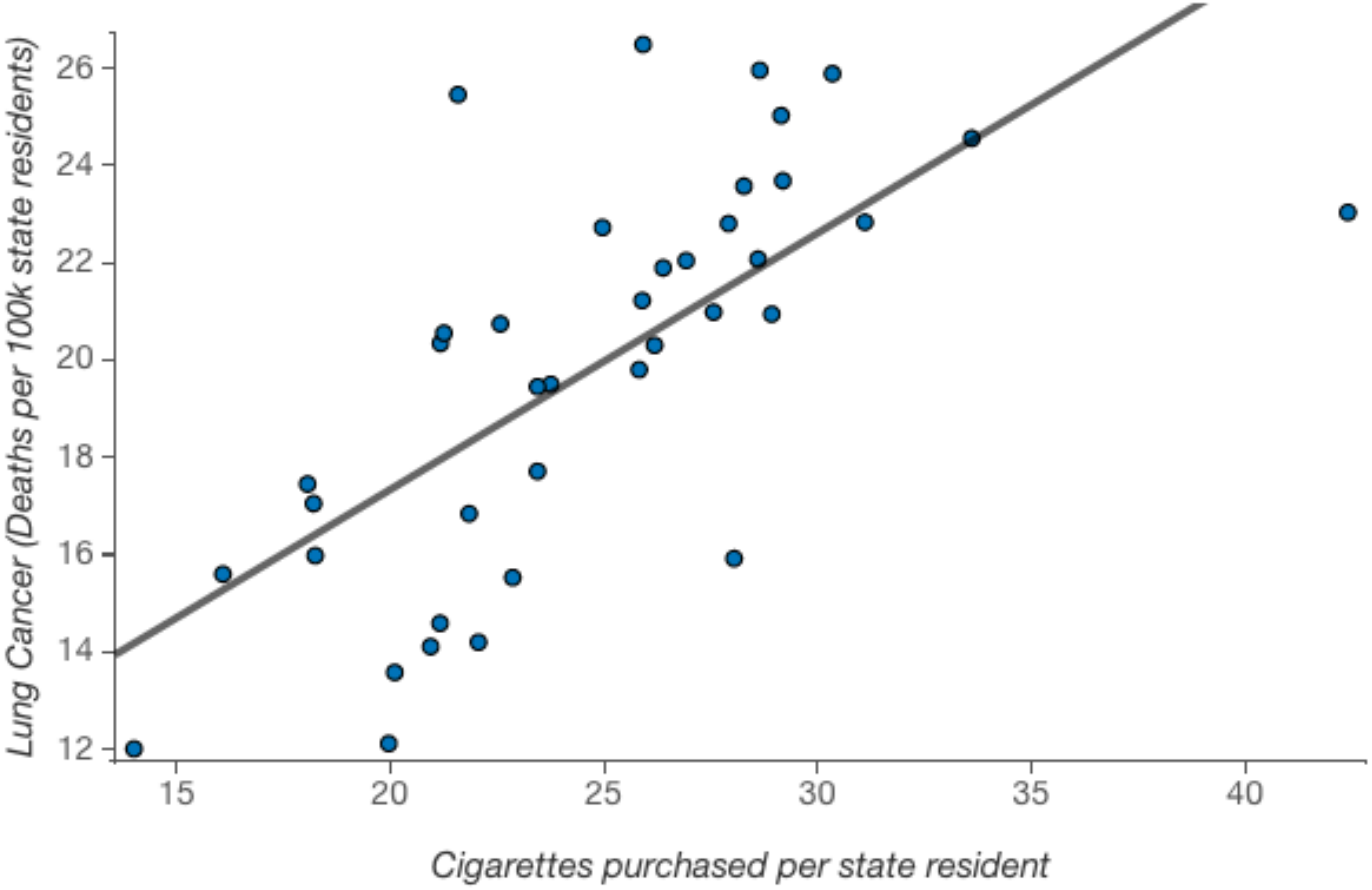
- *Ecological inference:* looking for associations between cause and effect at the level of groups or populations.

  "Do groups with higher average levels of A also tend to have higher B?"

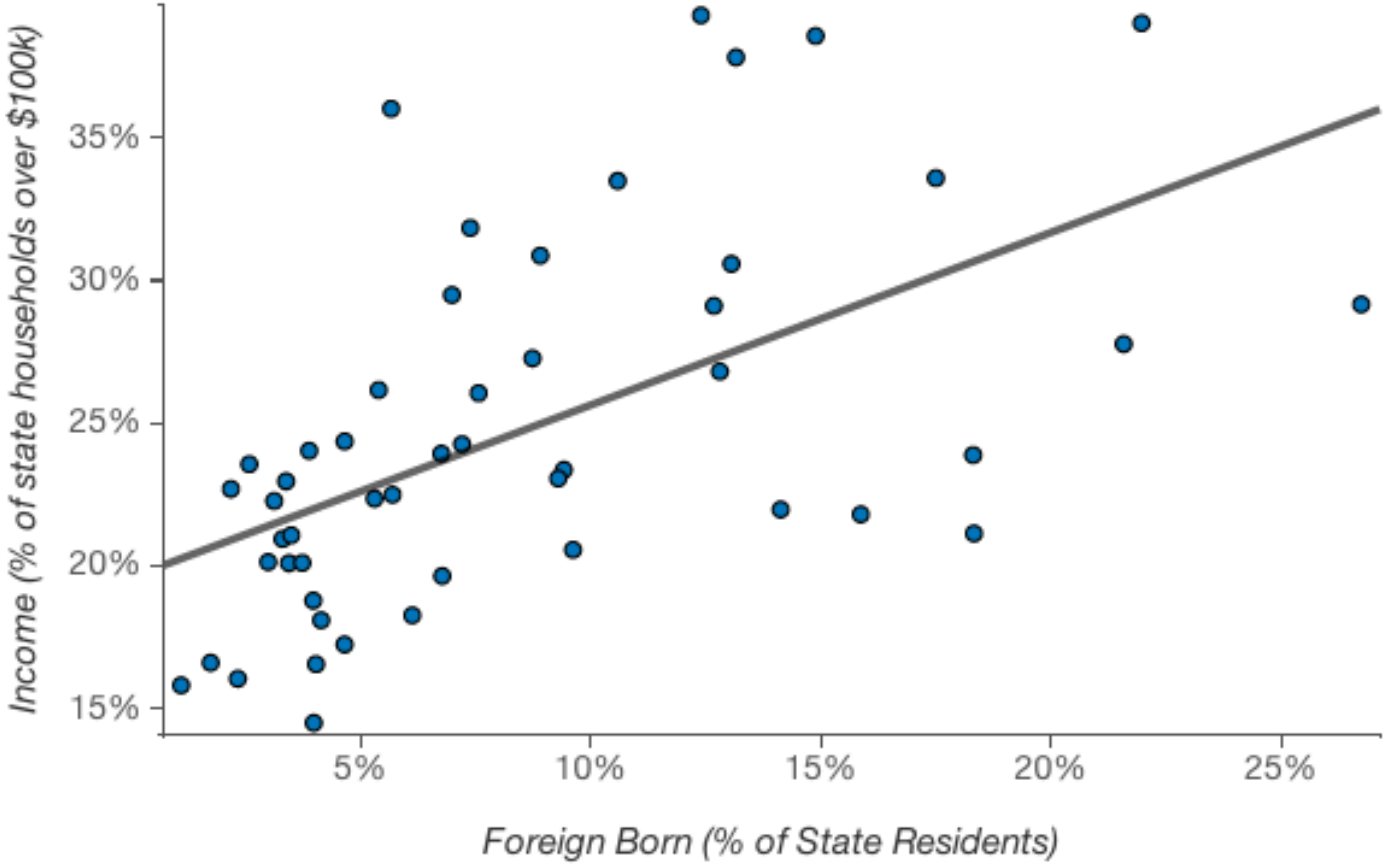
- The *ecological fallacy*: assuming, without further justification, that group-level associations accurately reflect individual-level associations.

  "Groups with higher A have higher B, on average. Therefore individuals with higher A have higher B, on average." Not necessarily!

- Ecological inference *can* be done well, but often it's not. It's much more useful for *generating* a hypothesis rather than *proving* it.



Smoking more cigarettes really does increase an individual's risk of lung cancer. This ecological association accurately reflects an individual-level trend…



...but this one doesn't. At the individual level, 22.1% of foreign-born residents make more than $100K, versus 26.1% of U.S.-born residents.

TAKE-HOME MESSAGES

- **A trend that appears when data are *separated into individuals/smaller groups* can look different, or even reverse entirely, when the data are *aggregated into larger groups.***
  - **Senior doctors perform worse overall, but better at every specific type of delivery (separating vs. aggregating deliveries by type).**
  - **Richer states lean Democratic, but richer people lean Republican in all states (separating vs. aggregating people by state).**

TAKE-HOME MESSAGES

- A trend that appears when data are *separated into individuals/smaller groups* can look different, or even reverse entirely, when the data are *aggregated into larger groups.*
  - Senior doctors perform worse overall, but better at every specific type of delivery (separating vs. aggregating deliveries by type).
  - Richer states lean Democratic, but richer people lean Republican in all states (separating vs. aggregating people by state).
- So what to do? Remember the rule of total probability!
  - Pay attention: the level of aggregation/grouping in a data set matters a lot.
  - Ask questions: “Do we care about an overall or a conditional probability? Are we missing any lurking variables?”
  - Avoid the *ecological fallacy*: learn to be skeptical when a group-level trend (voting vs. income for states) is casually applied to individuals (voting vs. income for people).

YOUR TURN

	Stones < 2cm	Stones ≥ 2cm	Overall
Open surgery	0.93	0.73	[redacted]
Percutaneous nephrolithotomy	0.83	0.69	[redacted]

Success rates in a medical study of two modes of treatment for kidney stones (from Julious and Mullee, BMJ 1994)

	Stones < 2cm	Stones ≥ 2cm	Overall
Open surgery	0.93	0.73	0
Percutaneous nephrolithotomy	0.83	0.69	0

Success rates in a medical study of two modes of treatment for kidney stones (from Julious and Mullee, BMJ 1994)

Which procedure has the higher *overall* success rate?

- **Open surgery**
- **Percutaneous nephrolithotomy**
- **It is impossible to tell.**
- **It is possible to tell, but we must know the sample sizes by procedure and stone type.**

	Stones < 2cm	Stones ≥ 2cm	Overall
Open surgery	0.93	0.73	0
Percutaneous nephrolithotomy	0.83	0.69	0

Success rates in a medical study of two modes of treatment for kidney stones (from Julious and Mullee, BMJ 1994)

Which procedure has the higher *overall* success rate?

- **Open surgery**
- **Percutaneous nephrolithotomy**
- **It is impossible to tell.**
- **It is possible to tell, but we must know the sample sizes by procedure and stone type.**

	Stones < 2cm	Stones ≥ 2cm	Overall
Open surgery	0.93 (87)	0.73 (263)	0.78 (350)
Percutaneous nephrolithotomy	0.83 (270)	0.69 (80)	0.83 (350)

Success rates in a medical study of two modes of treatment for kidney stones (Julious and Mullee, BMJ 1994)