

Causal Learning

Reading Group, Summer 2023

Logistics



Not a course (or credited)



Discussion based, not lecture/seminar series

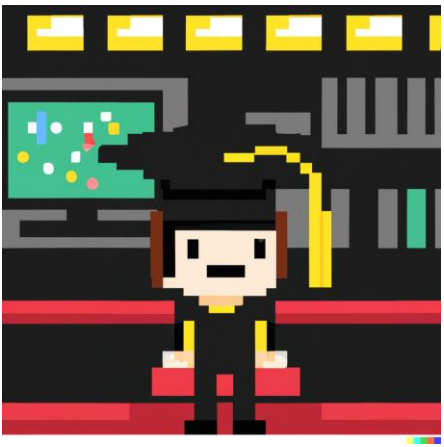


Tentative schedule: June 7 – August 9

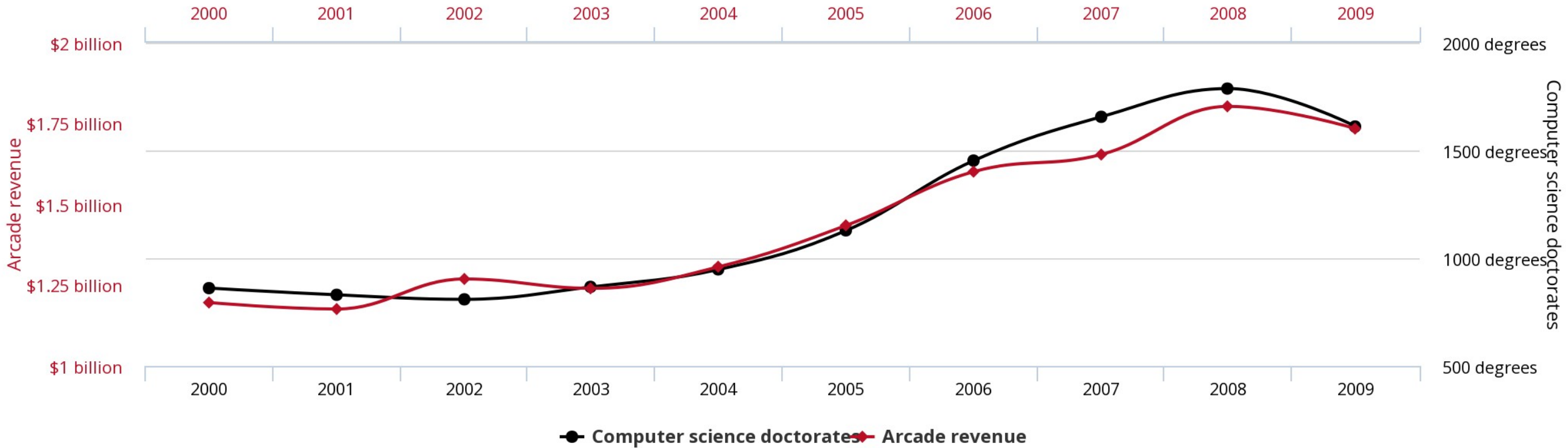


Format:

- Anshuman & Hannah cover basic concepts over the first 5-6 weeks
- Members volunteer and present advanced topics/papers



Total revenue generated by arcades correlates with Computer science doctorates awarded in the US



¹Source: <https://www.tylervigen.com/spurious-correlations>

²Generated with DALL-E prompt “a freshly graduated PhD student wearing a mortarboard hat, at a gaming arcade, pixel art, arcade obviously visible”

Simpson's Paradox

Gender	Drug	No drug	
Men ←	81/87 (93%))	234/270 (87%)	☑ .
Women ←	192/263 (73%))	55/80 (69%))	☑ .
Combined	273/350 (78%))	289/350 (83%))	X

Male: $P(\text{recovery}|\text{drug}) > P(\text{recovery}|\text{no drug})$

Female: $P(\text{recovery}|\text{drug}) > P(\text{recovery}|\text{no drug})$

Overall: $P(\text{recovery}|\text{drug}) < P(\text{recovery}|\text{no drug})$

D'oh! Nothing makes sense anymore!

Actual outcome (real-world): Drug is effective



Simpson's Paradox

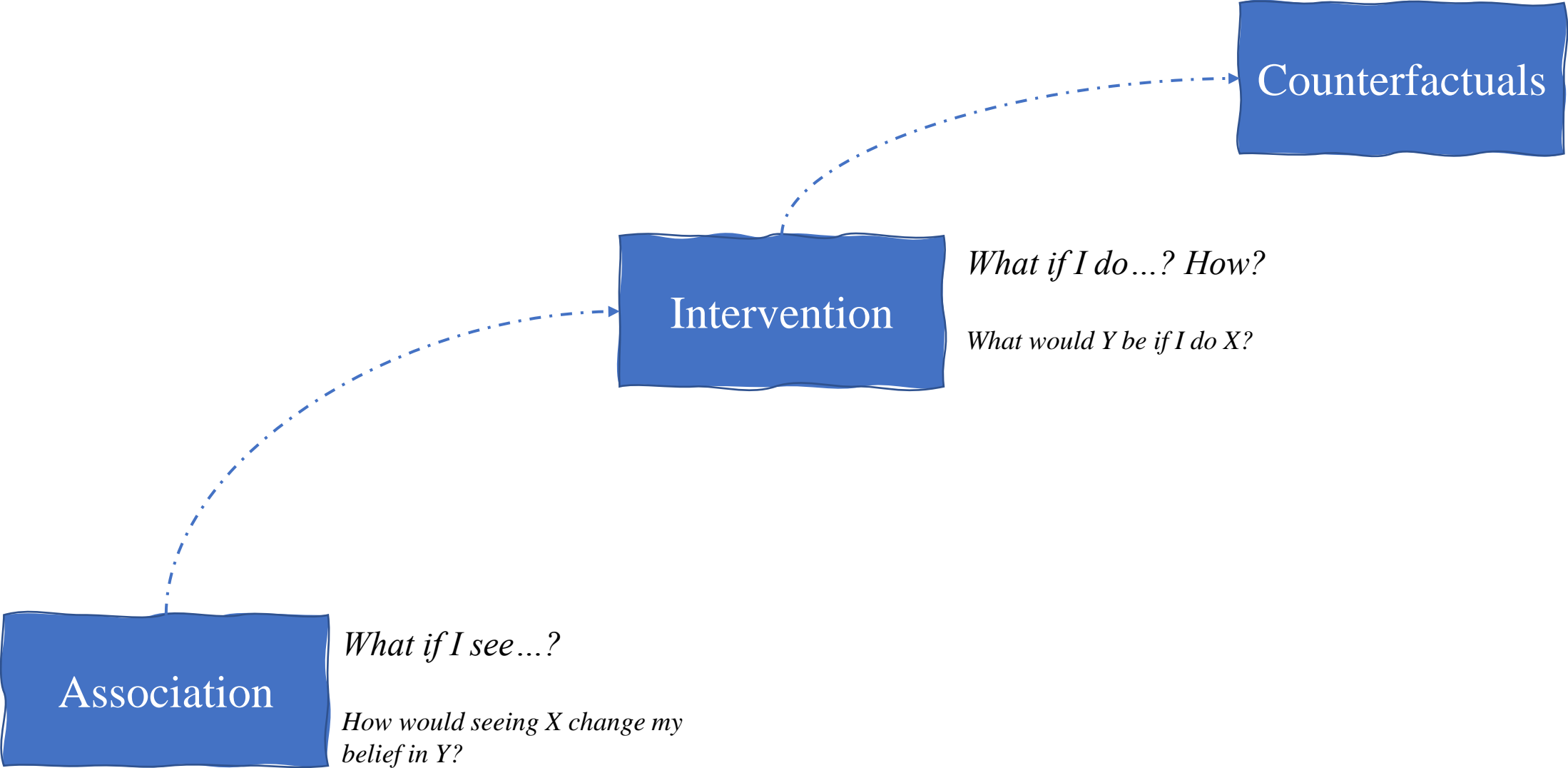
Post-treatment BP	Drug	No drug
Low BP	81/87 (93%)	234/270 (87%)
High BP	192/263 (73%)	55/80 (69%)
Combined	273/350 (78%)	289/350 (83%)

} Same

Actual outcome (real-world): Drug is not effective



Ladder of Causation



Ladder of Causation

Counterfactuals

What if I had done...? Why?

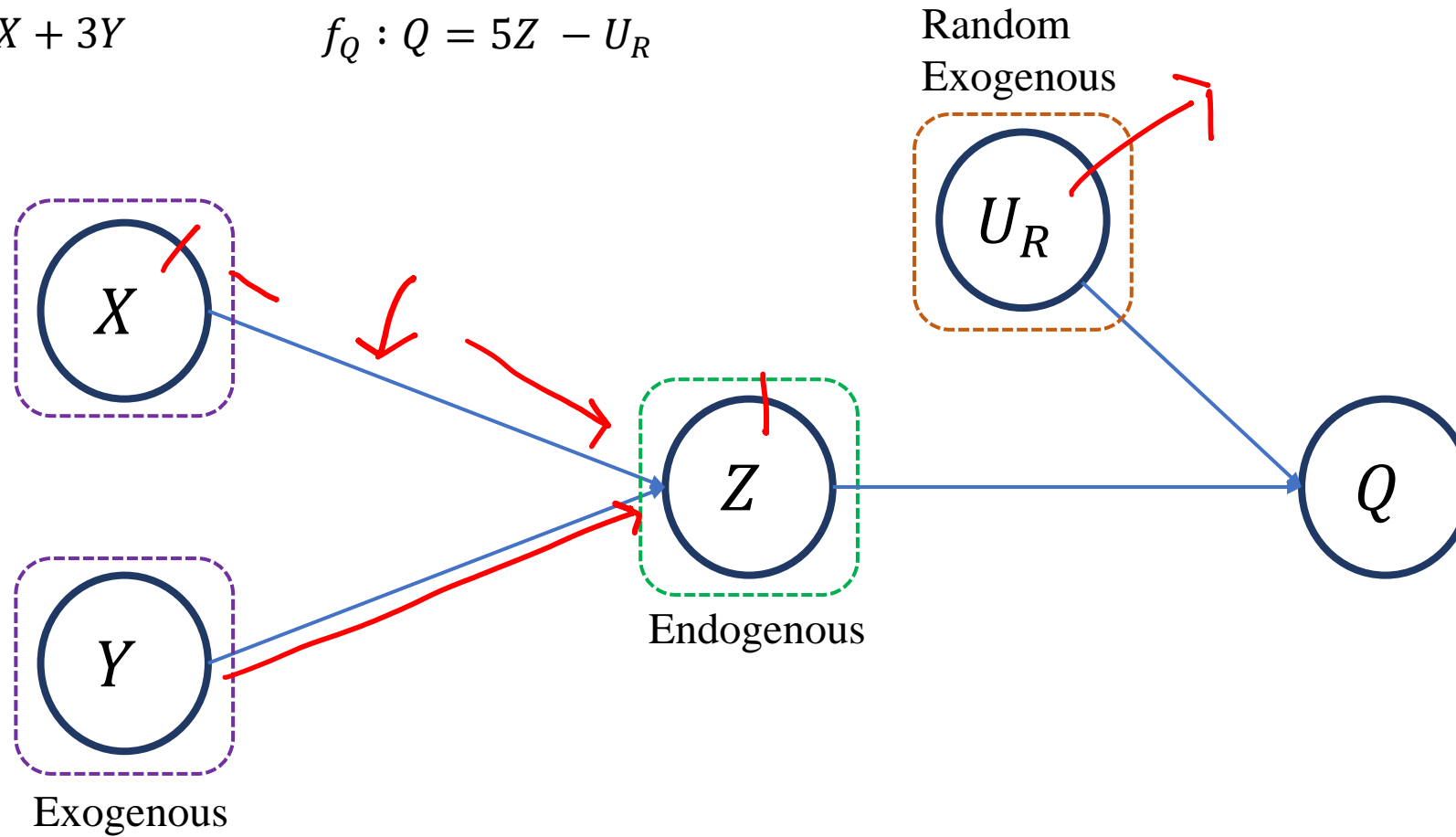
*Was it X that caused Y? What if X
did not occur?)*

“Imagination”

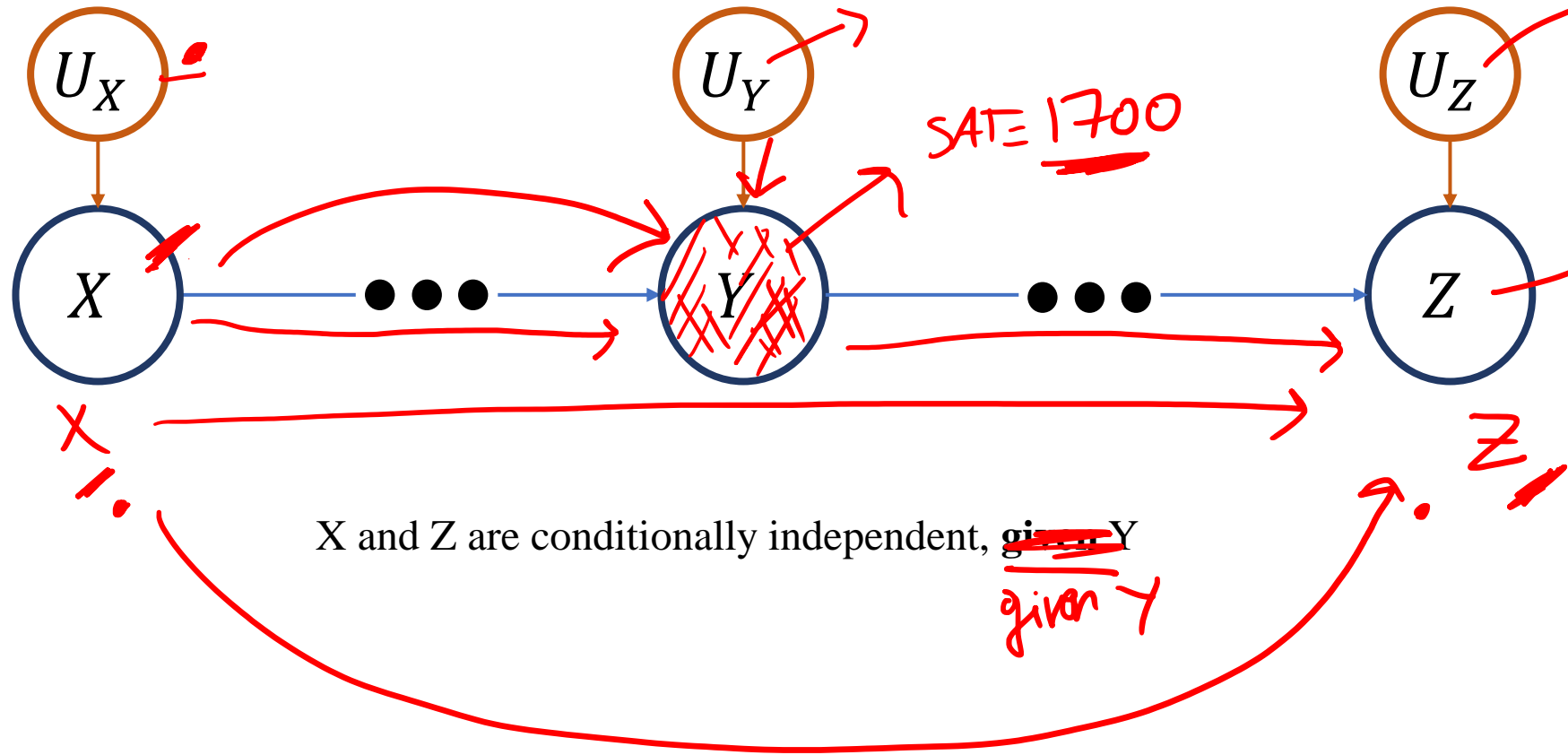
Structural Causal Models

$$f_Z : Z = 2X + 3Y$$

$$f_Q : Q = 5Z - U_R$$



Chains



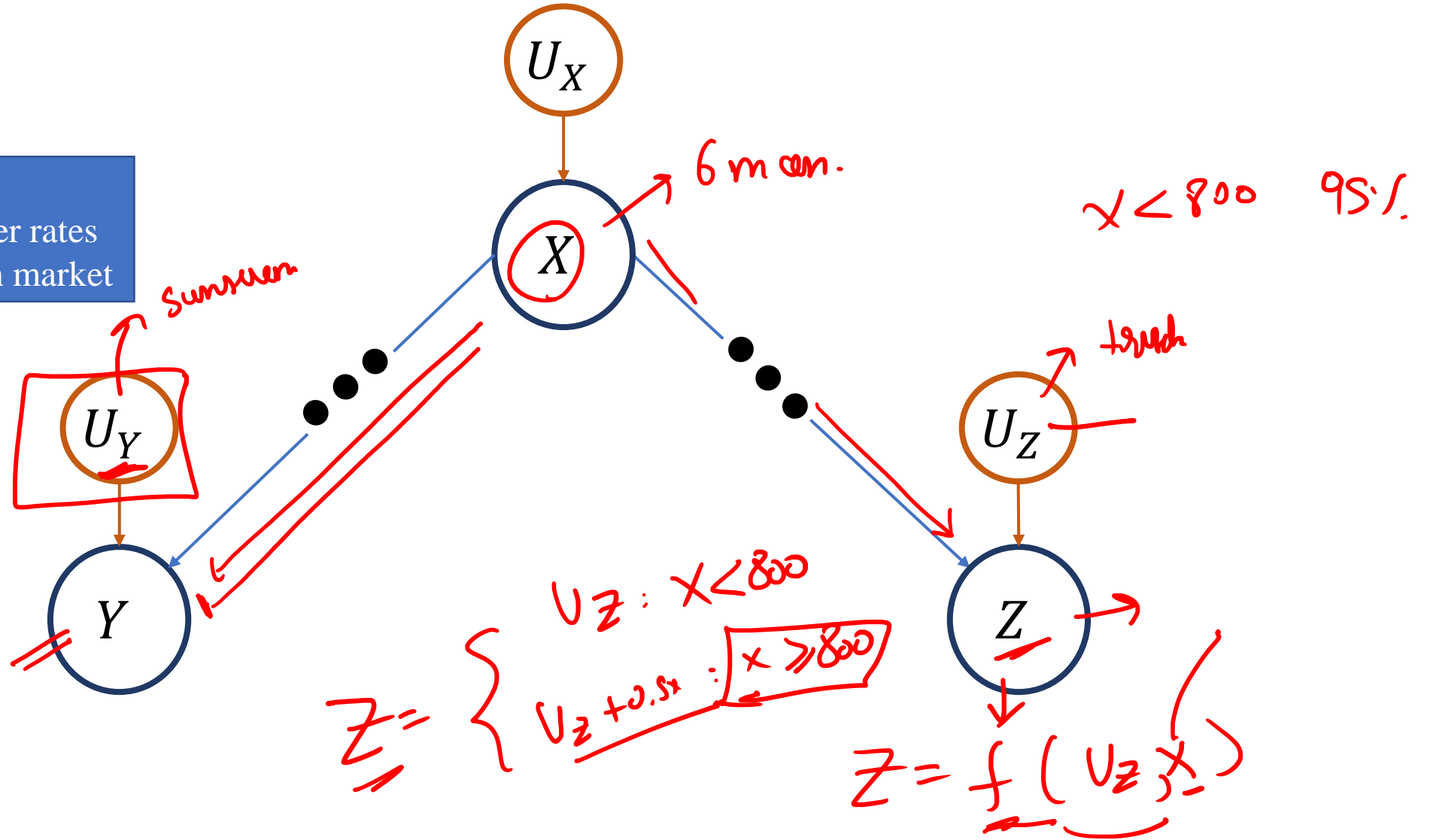
Example:

X: High school funding (\$)
Y: Average SAT Score
Z: College acceptance rate

Forks

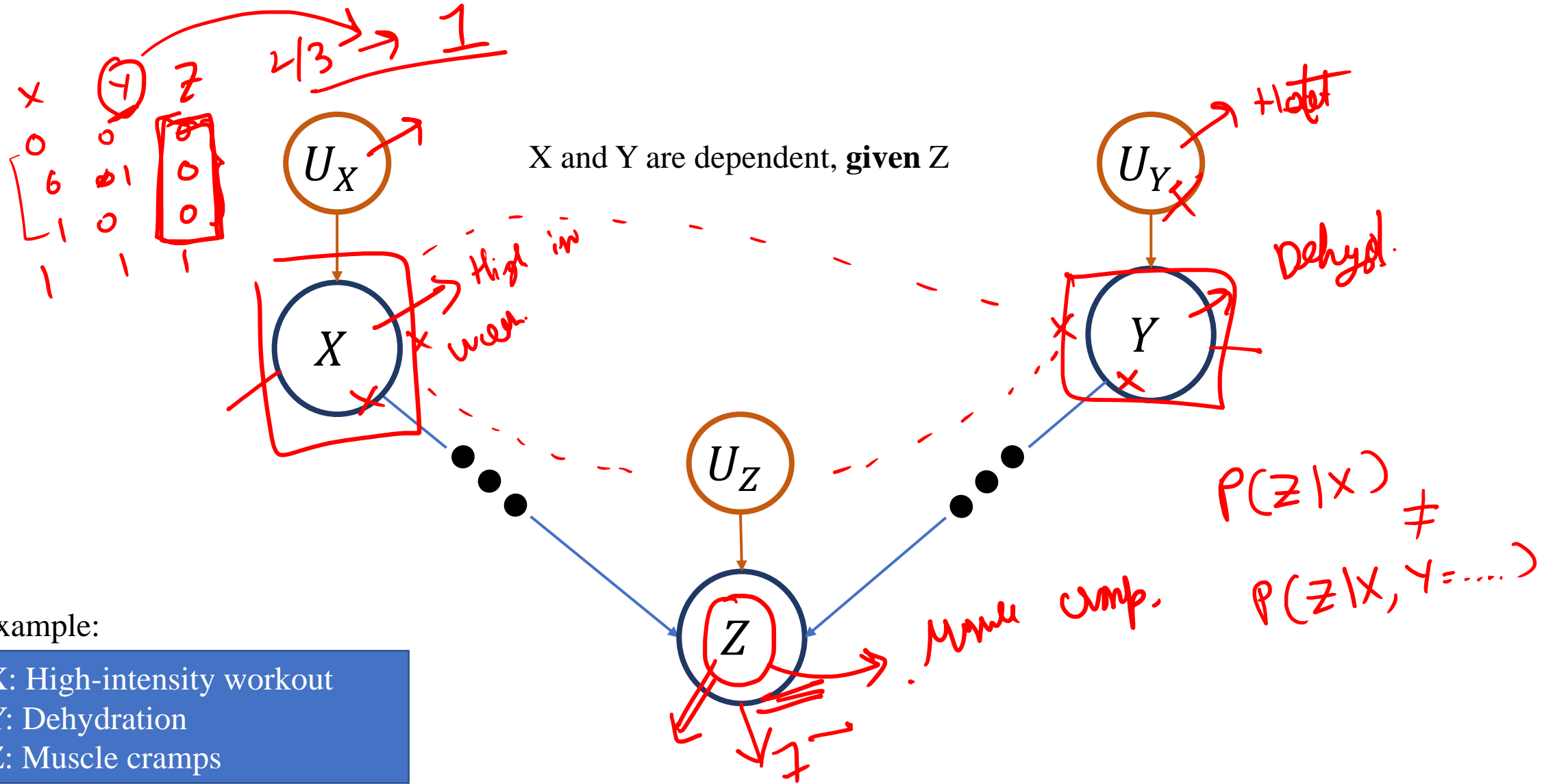
Example:

X: High UV Sunlight
Y: Increased skin cancer rates
Z: Better blueberries in market

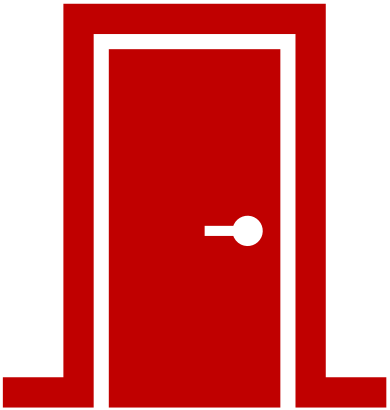


Y and Z are conditionally independent, **given** X

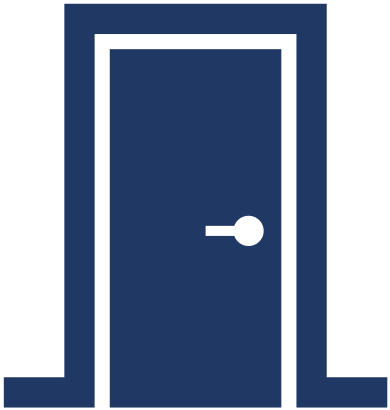
Colliders



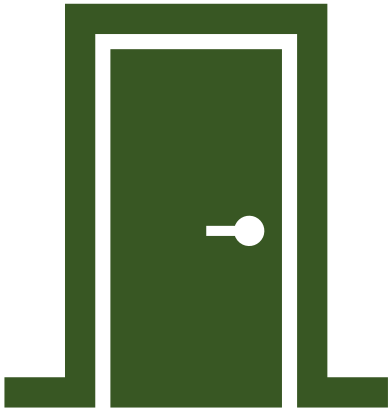
Monty Hall- Causal Perspective



A



B

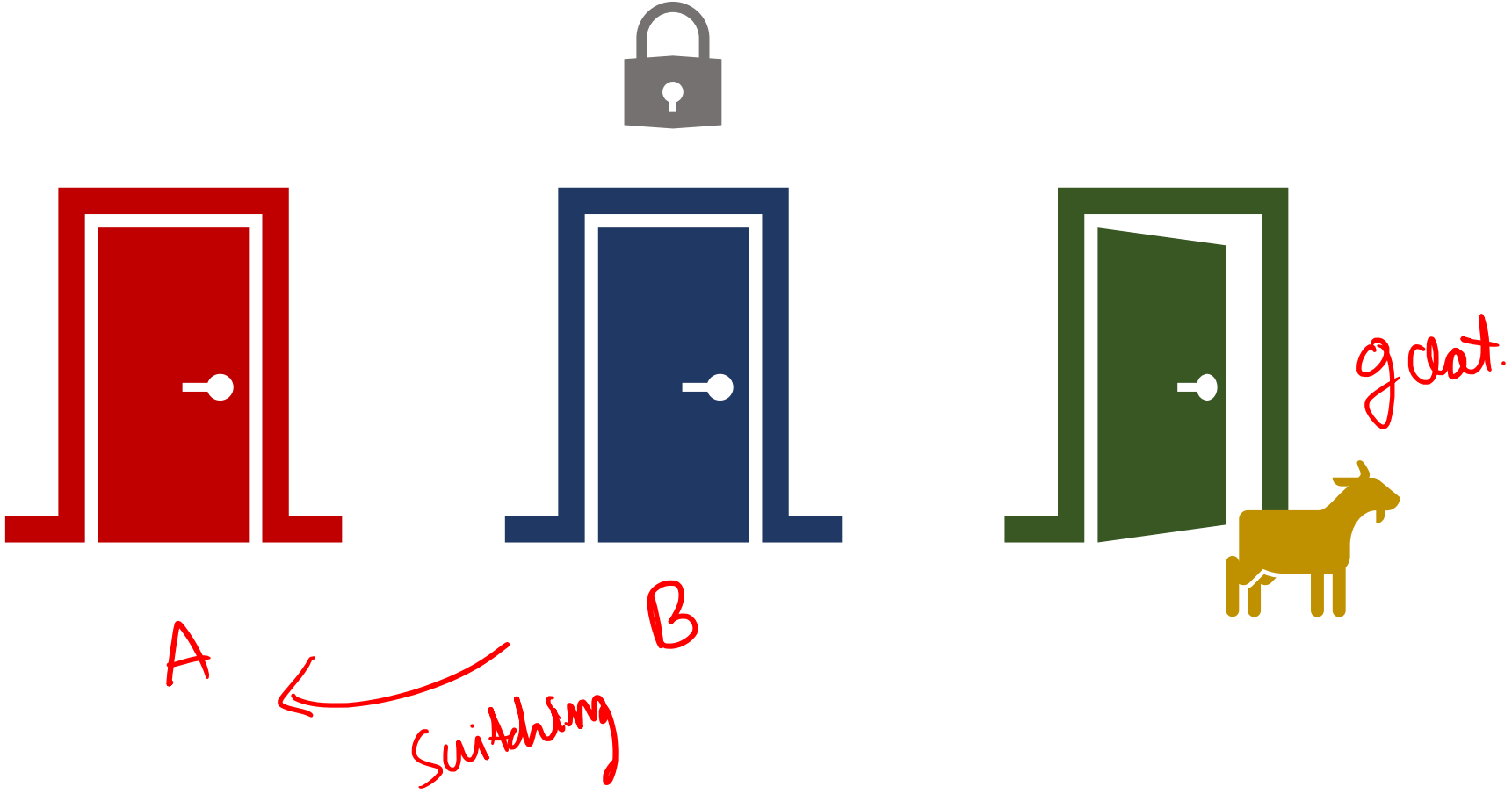


C

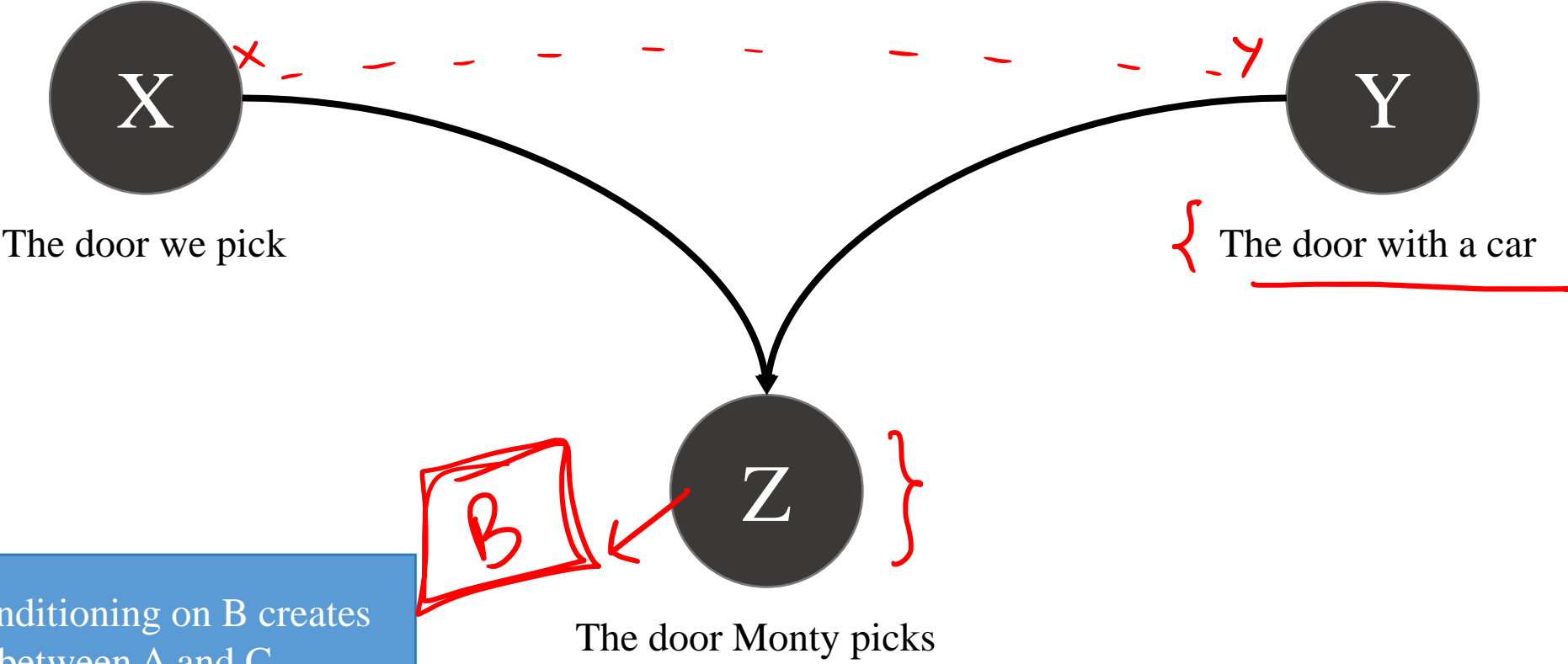


Monty Hall

Monty Hall- Causal Perspective



Monty Hall- Causal Perspective



Collider: Conditioning on B creates dependency between A and C

Monty Hall- Causal Perspective

Door A	Door B	Door C
Goat ✓	Goat	Car
Goat ✓	Car	Goat
Car ✓	Goat	Goat

Switch

Stay

1

0

1

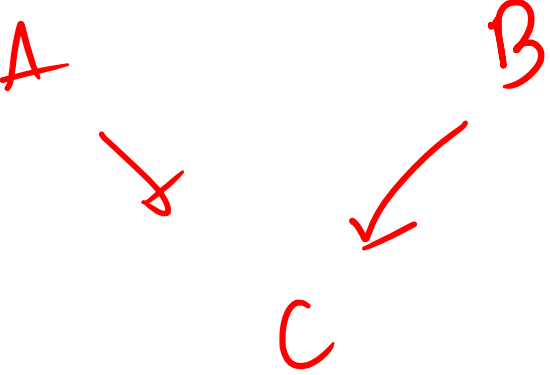
0

0

1

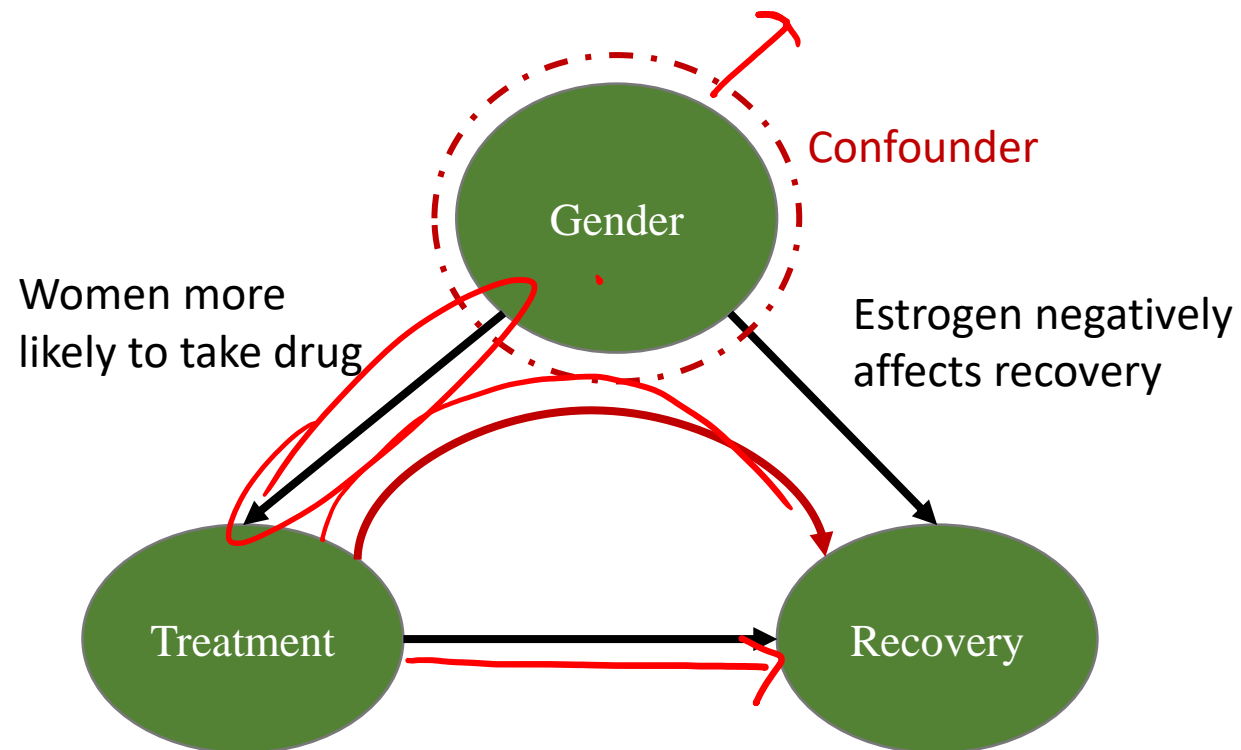
$\frac{2}{3}$

$\frac{1}{3}$



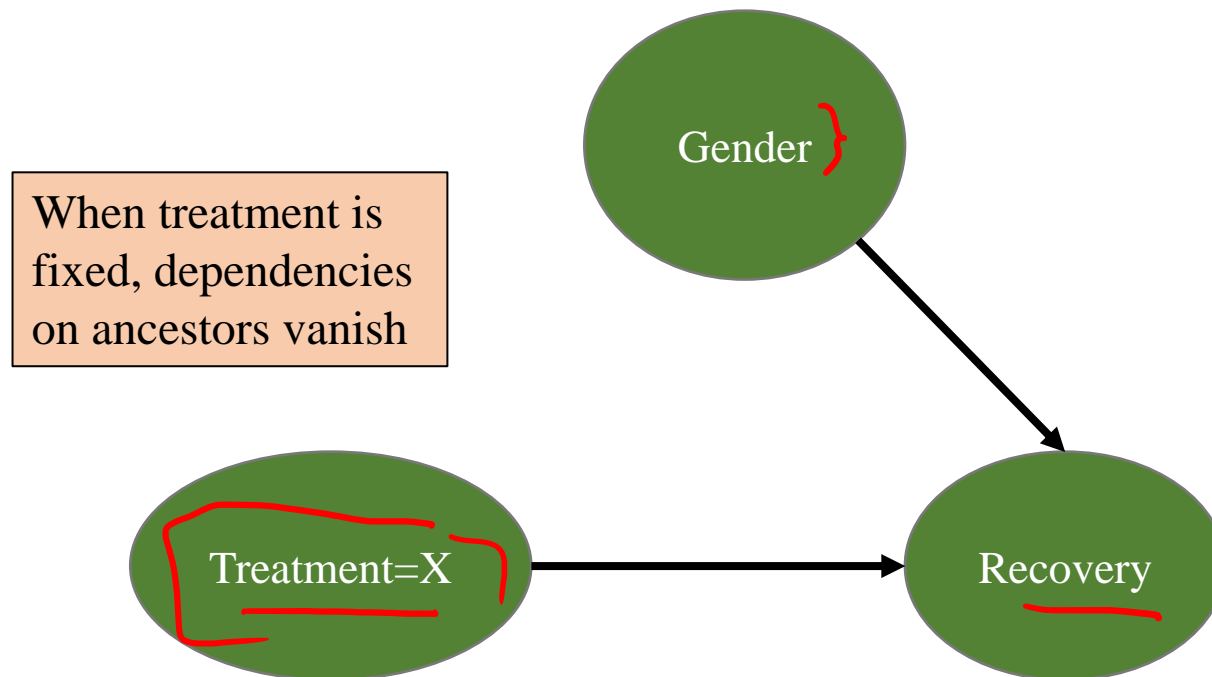
Simpson's Paradox: Revisited

Gender	Drug	No drug
Men	81/87 (93%)	234/270 (87%)
Women	192/263 (73%)	55/90 (69%)
Combined	273/350 (78%)	289/350 (83%)



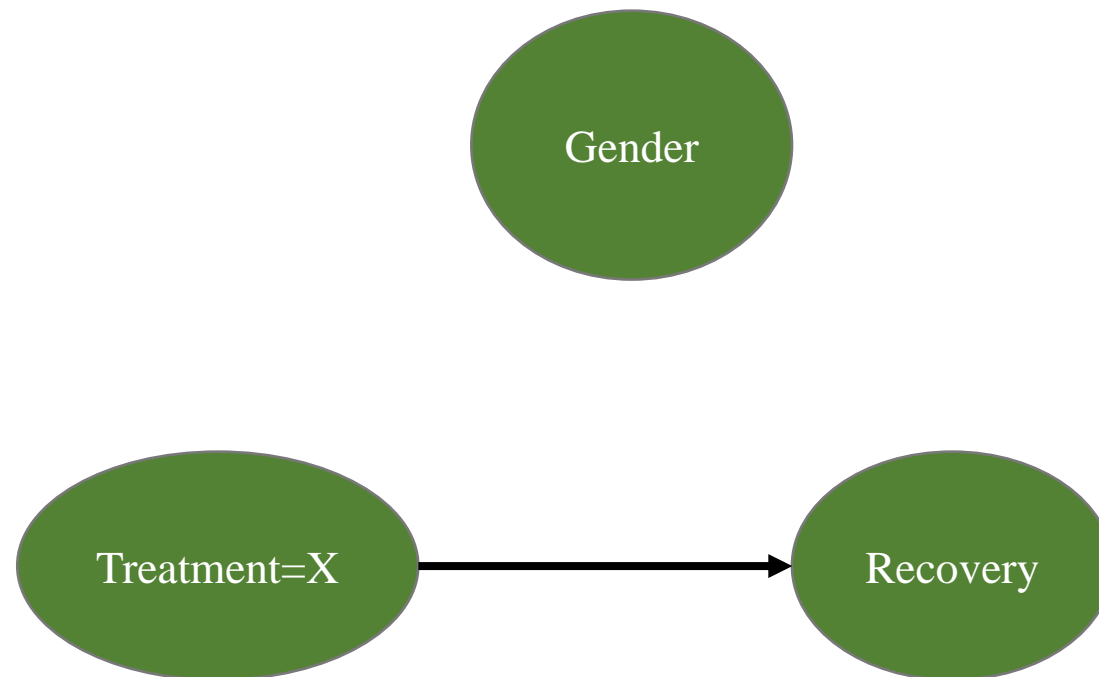
Simpson's Paradox: Revisited

Gender	Drug	No drug
Men	81/87 (93%)	234/270 (87%)
Women	192/263 (73%)	55/80 (69%)
Combined	273/350 (78%)	289/350 (83%)



Simpson's Paradox: Revisited

Gender	Drug	No drug
Men	81/87 (93%)	234/270 (87%)
Women	192/263 (73%)	55/80 (69%)
Combined	273/350 (78%)	289/350 (83%)

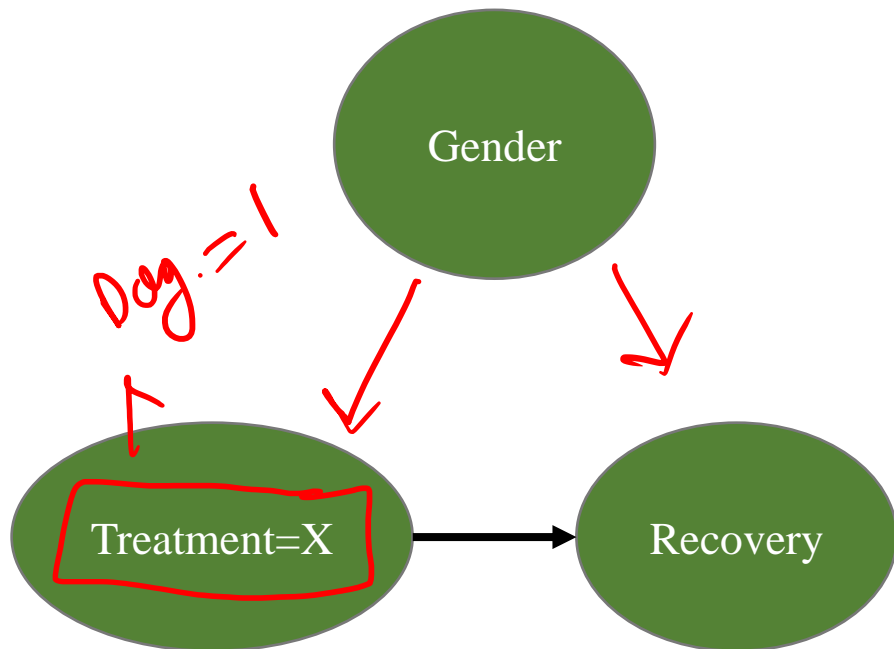


When gender is fixed,
only treatment effects
recovery

Simpson's Paradox: Revisited

Gender	Drug	No drug
Men	81/87 (93%)	234/270 (87%)
Women	192/263 (73%)	55/80 (69%)
Combined	273/350 (78%)	289/350 (83%)

Handwritten notes: A large red curly bracket on the right side of the table. To its right, a vertical sequence of red arrows pointing downwards, with a '0.' written above the first arrow and a box containing a '0' at the bottom.

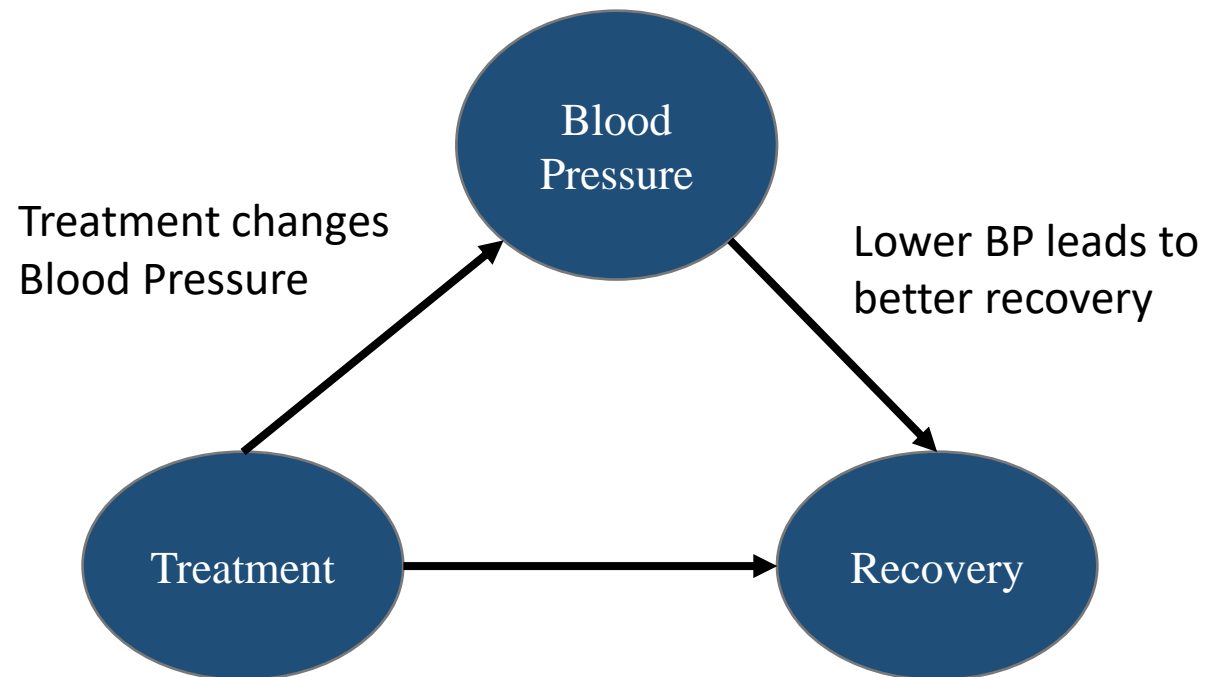


$$\begin{aligned}
 P(R|D) &= P(R|D, M) * P(M) + P(R|D, F) * P(F) \\
 &= 0.93 * \left(\frac{87 + 270}{700} \right) + 0.73 * \left(\frac{263 + 80}{700} \right) \\
 &= 0.832
 \end{aligned}$$

$$\begin{aligned}
 P(R|\neg D) &= P(R|\neg D, M) * P(M) + P(R|\neg D, F) * P(F) \\
 &= 0.87 * \left(\frac{87 + 270}{700} \right) + 0.69 * \left(\frac{263 + 80}{700} \right) \\
 &= 0.782 \quad \times
 \end{aligned}$$

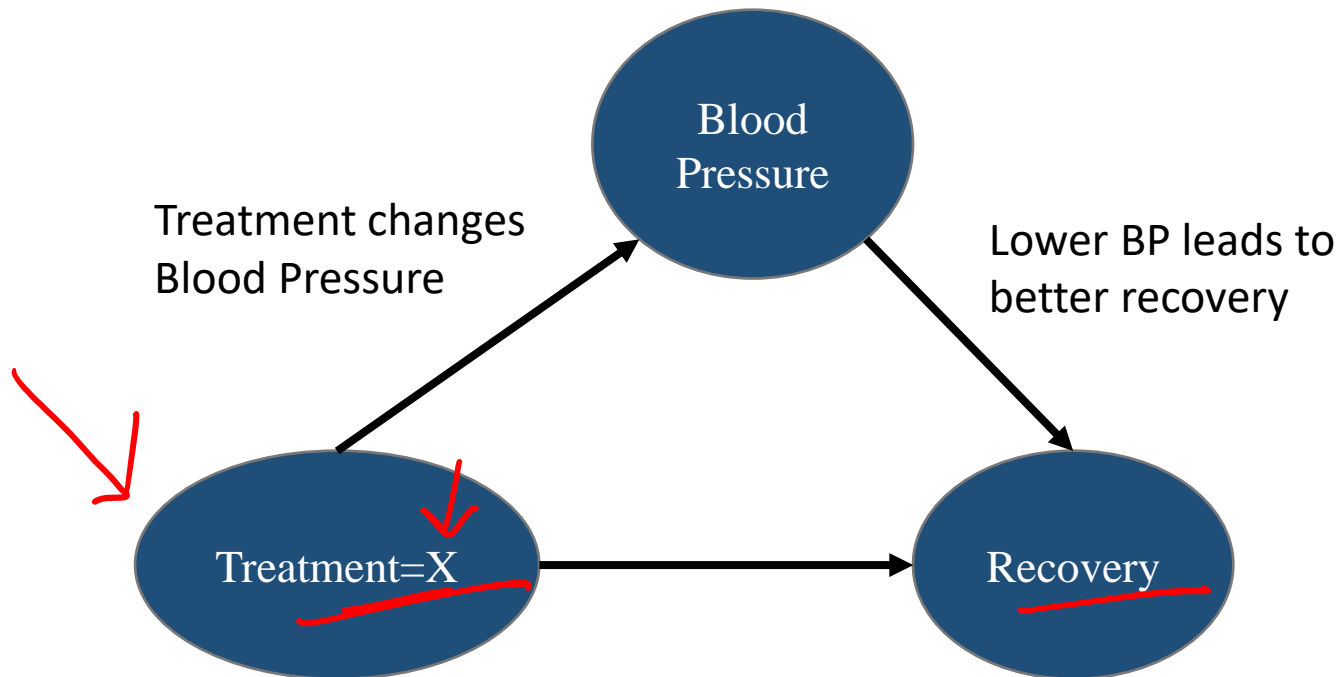
Simpson's Paradox: Revisited

Post-treatment BP	Drug	No drug
Low BP	81/87 (93%)	234/270 (87%)
High BP	192/263 (73%)	55/80 (69%)
Combined	273/350 (78%)	289/350 (83%)



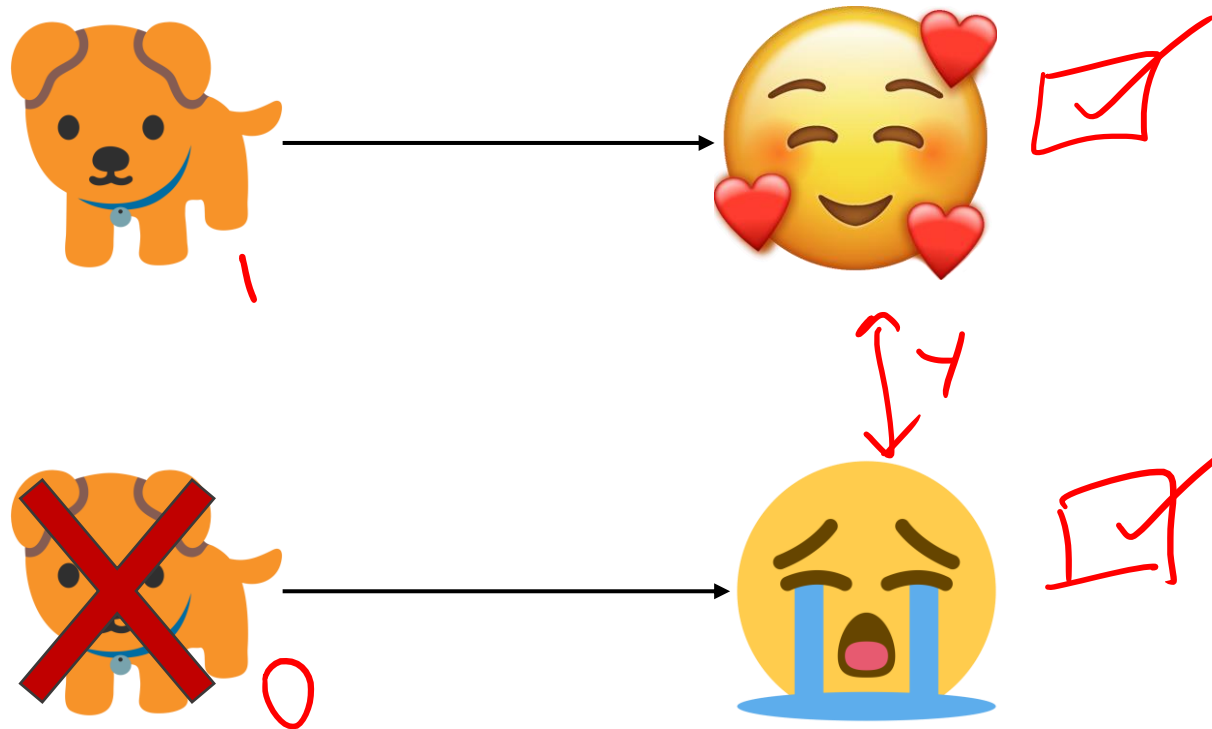
Simpson's Paradox: Revisited

Post-treatment BP	Drug	No drug
Low BP	81/87 (93%)	234/270 (87%)
High BP	192/263 (73%)	55/80 (69%)
Combined	273/350 (78%)	289/350 (83%)



On fixing treatment, interventional graph remains same as data generating graph

Fundamental Problem of Causal Inference



Individual Treatment Effect (ITE)

$$\tau_i \triangleq Y_i(1) - Y_i(0)$$

Can only observe one of $Y_i(0)$ or $Y_i(1)$

Outcomes not observed: Counterfactuals

How about average (expectation) treatment effect?

Next Week

Hannah: Potential Outcomes

Please fill out form if you want to volunteer (topic/paper)

Happy Global Running Day!

