THE UNIVERSITY OF SYDNEY

Advanced Machine Learning

(COMP 5328)

Causal Inference

Tongliang Liu

Review

RL Framework (Procedure)

An agent interacts with its environment according to following procedure.

- At each time step t, the agent is in some state $S_t = s_t$.
- It chooses an action $A_t = a_t$.
- Based on s_t and a_t , it receives reward $R_t = r_t$ and goes into state $S_{t+1} = s_{t+1}$.

In general, the action A_t , reward R_t and next state S_{t+1} have the following probability distributions,

$$\pi(A_t|S_t = s_t), \quad P(S_{t+1}|A_t = a_t, S_t = s_t), \quad P(R_t|A_t = a_t, S_t = s_t).$$

RL Framework (Objective)

Given current state is s_0, we want to find the optimal policy (best strategy) π^* that maximises the expected cumulative reward, i.e.,

$$\pi^* = \arg\max_{\pi} \mathbb{E}\left[\sum_{t\geq 0} \gamma^t r_t | S_0 = s_0, \pi\right],$$

With

$$a_t \sim \pi(A_t|S_t = s_t),\ s_{t+1} \sim P(S_{t+1}|A_t = a_t, S_t = s_t),$$

and a discounted factor

$$\gamma \in (0, 1)$$

Q-value function

Q-value function $Q : S \times A \rightarrow \mathbb{R}$ determines the expected cumulative reward by following the policy π given current state is s_0, and action is a_0 , specifically,

$$Q(s_0, a_0) = \mathbb{E}\left[\sum_{t \geq 0} \gamma^t r_t | S_0 = s_0, A_0 = a_0, \pi\right],$$

with

$$a_t \sim \pi(A_t | S_t = s_t),\ s_{t+1} \sim P(S_{t+1} | A_t = a_t, S_t = s_t),$$

$$r_t \sim P(R_t | A_t = a_t, S_t = s_t)$$

and γ is the discounted factor.

Optimal Q-value function

The optimal Q-value (denoted as Q^*) is the the maximum expected cumulative reward achievable by following the optimal policy π^* given that the current state is s_0 and action is a_0, specifically,

$$Q^*(s_0, a_0) = \max_{\pi} \mathbb{E}\left[\sum_{t \geq 0} \gamma^t r_t | S_0 = s_0, A_t = a_0, \pi\right]$$

with

$$a_t \sim \pi(A_t | S_t = s_t),\ s_{t+1} \sim P(S_{t+1} | A_t = a_t, S_t = s_t)$$

and r^t is the immediate reward for the state-action pair (s_t, a_t)

Update Q-value function

Empirically, we could use the following equation to update Q-value function

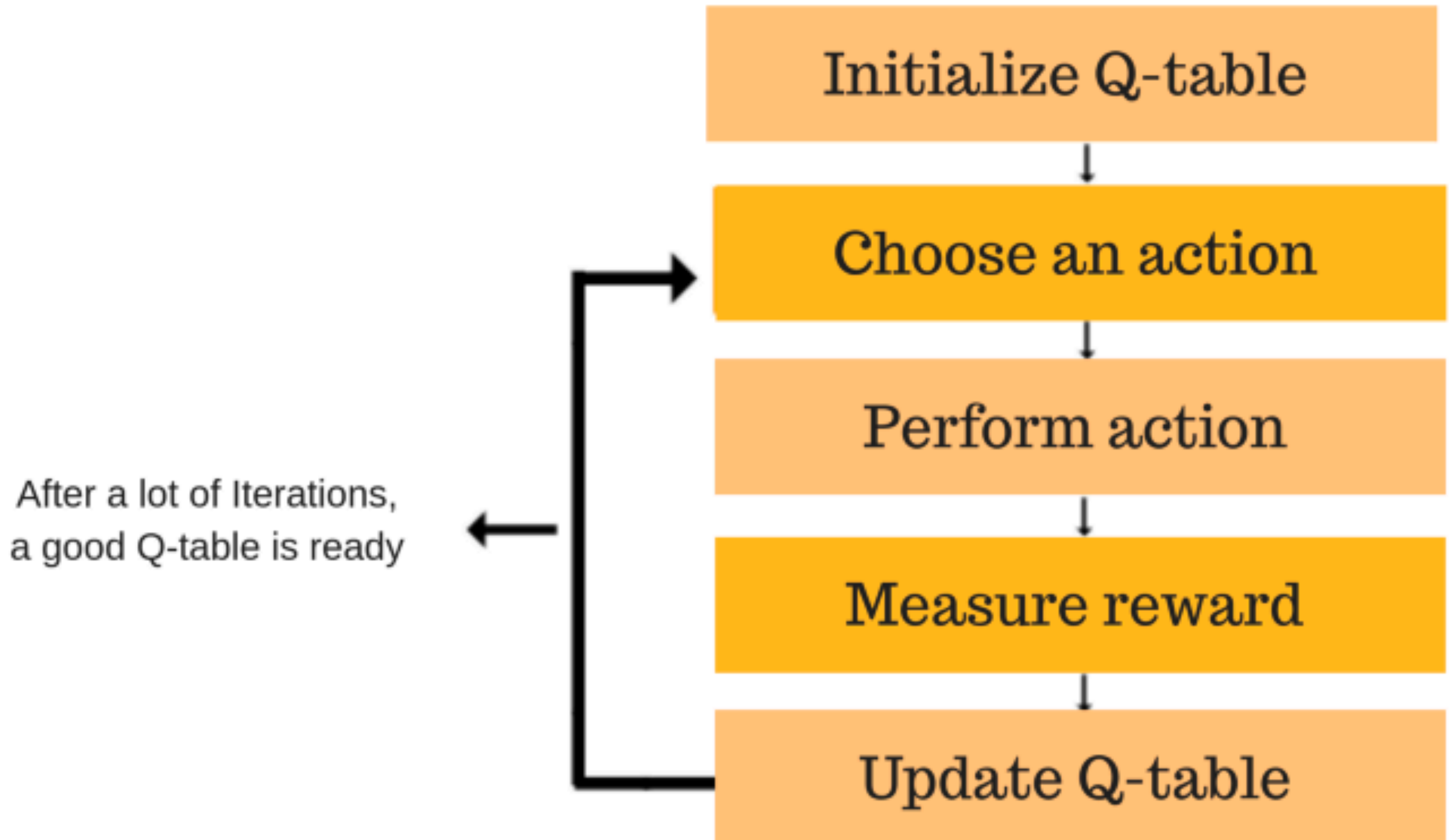
$$Q_{i+1}(s,a) = Q_i(s,a) + \eta \left(r + \gamma \max_{a'} Q_i(s',a') - Q_i(s,a) \right),$$

where $\eta \in (0,1)$, and i is the number of iterations. Intuitively, $r + \gamma \max_{a'} Q_i(s',a')$ is a more accurate approximation of the Q-value compared to $Q_i(s,a)$ because it contains a true value r.

We add a hyper-parameter η because the estimation $r + \gamma \max_{a'} Q_i(s',a')$ can not be fully trusted. If i goes to infinity, Q_i will converge to Q^* (see proof [1]).

[1] Proofs: http://users.isr.ist.utl.pt/~mtjspaan/readingGroup/ProofQlearning.pdf).

Q-Table



Credit: https://www.freecodecamp.org/news/a-brief-introduction-to-reinforcement-learning-7799af5840db/

Deep Q-learning Network (DQN)

Deep Q-learning Network

For some tasks both state and action could be continues. It is infeasible to store Q-value for every state-action pair into the Q table. Therefore, we use a neural network as a function approximator to store Q-values.

Deep Q-learning Network

We parameterise Q by weight w of neutral network, i.e, Q_w .
The objective function is to iteratively minimise

$$(r + \gamma \max_{a'} Q_w(s', a') - Q_w(s, a))^2,$$

where state s' is observed by substituting action a and s into the environment.
Note that gradient is applied only to $Q_w(s, a)$, not to $Q_w(s', a')$

Policy Gradients

- Sometimes Q-function can be very complicated! For example, a robot grasping an object has a very high-dimensional state (image), and it is hard to learn exact value of every (state, action) pair. However the policy can be much simpler.

Why don't we directly learning the optimal policy?

- we can use the similar way as modelling Q-values, i.e., using a neural network to model policy π_θ , where, θ is the parameter of the network.

Policy Gradients

The expected reward of a policy θ is:

$$J(\theta) = \mathbb{E}\left[\sum_{t\geq 0} \gamma^t r_t | \pi_\theta\right],$$

and we want to find the optimal policy (parameters)

$$\theta^* = \arg\max_\theta J(\theta).$$

Calculate Gradient on θ

Could we use gradient ascent on policy parameters?

Let $r(\tau)$ be the reward of a trajectory:

$$\tau = (s_0, a_0, r_0, s_1, a_1, r_1, \dots)$$

mathematically, we could write the expected reward $J(\theta)$ as:

$$J(\theta) = \mathbb{E}\left[\sum_{t\geq 0} \gamma^t r_t | \pi_\theta\right] = \mathbb{E}_{\tau\sim p(\tau;\theta)}[r(\tau)] = \int_{\tau\sim p(\tau;\theta)} r(\tau)p(\tau;\theta)d\tau.$$

Calculate Gradient on θ

Suppose we have some trajectories $\tau_1, \tau_2, \ldots, \tau_n$, could we approximate:

$$\nabla_\theta J(\theta) = \int_{\tau \sim p(\tau;\theta)} r(\tau) \nabla_\theta p(\tau;\theta) d\tau,$$

by summation (Monte Carlo methods)

$$\nabla_\theta \hat{J}(\theta) = \frac{1}{n} \sum_{i=1}^{n} r(\tau_i) \nabla_\theta p(\tau;\theta)?$$

No! The integration is not an expected value! Because non of $r(\tau)$ and $\nabla_\theta p(\tau;\theta)$ are distributions! Then we can not approximate it by empirical mean!

Calculate Gradient on θ (Log Derivative Trick)

$$\begin{aligned}\nabla_\theta J(\theta) &= \int_{\tau \sim p(\tau;\theta)} r(\tau)\nabla_\theta p(\tau;\theta)d\tau,\\ &= \int_{\tau \sim p(\tau;\theta)} r(\tau)p(\tau;\theta)\frac{\nabla_\theta p(\tau;\theta)}{p(\tau;\theta)}d\tau,\\ &= \int_{\tau \sim p(\tau;\theta)} r(\tau)p(\tau;\theta)\nabla_\theta \log p(\tau;\theta)d\tau,\\ &= \mathbb{E}_{\tau \sim p(\tau;\theta)}\left[r(\tau)\nabla_\theta \log p(\tau;\theta)\right].\end{aligned}$$

Then, we can proximate the expectation by

$$\nabla_\theta \hat{J}(\theta) = \frac{1}{n}\sum_{i=1}^{n} r(\tau_i)\nabla_\theta \log p(\tau_i;\theta).$$

Calculate Gradient on θ

Assume the Markov condition holds, i.e.,

$$P(s_{t+1}, a_t | s_t, s_{t-1}, \ldots, s_0) = P(s_{t+1}, a_t | s_t),$$

then $p(\tau;\theta) = \prod_{t\geq 0} p(s_{t+1}|s_t, a_t)\pi_\theta(a_t|s_t)$. We have that,

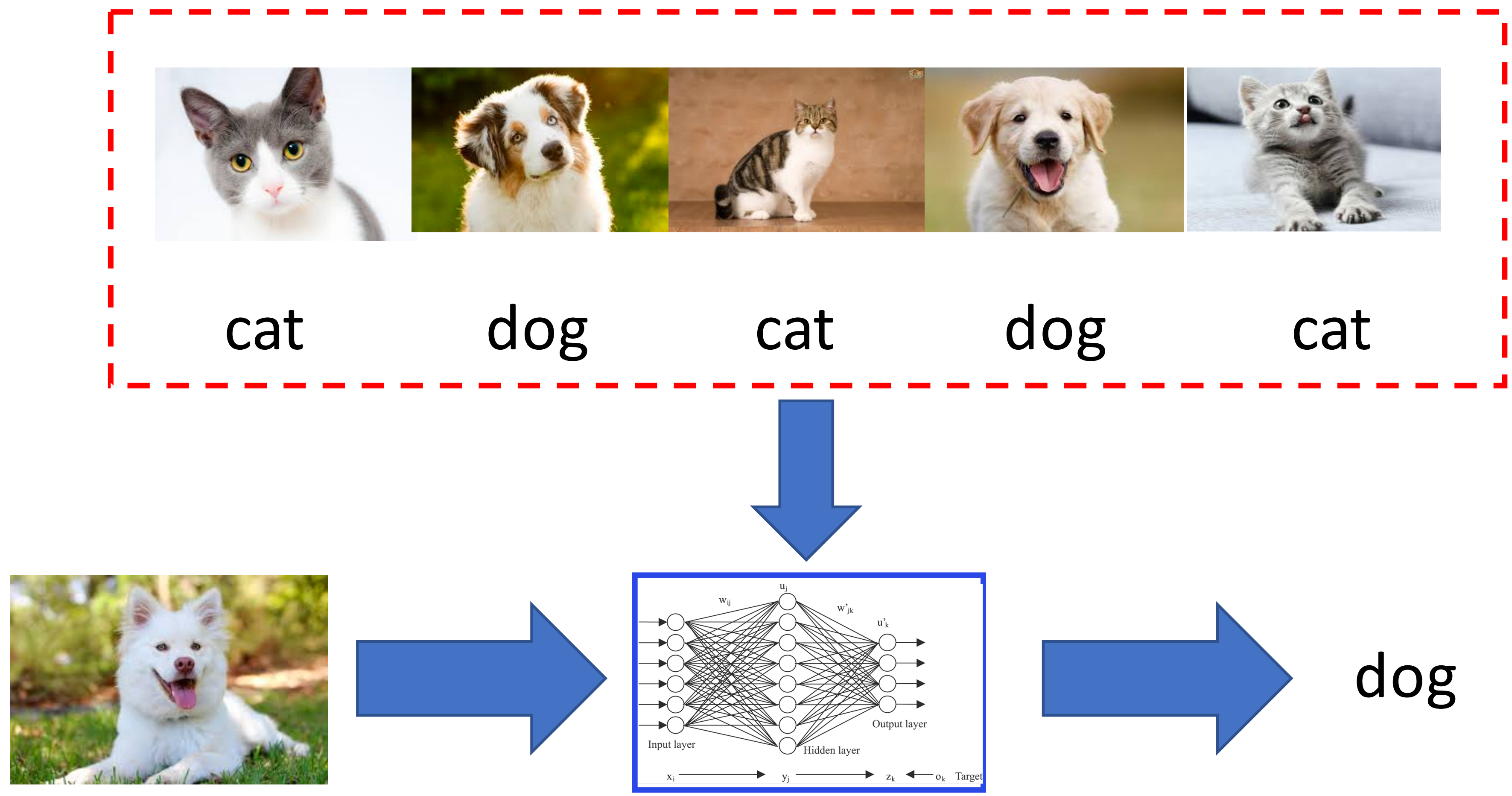
$$\begin{aligned}
\nabla_\theta \log p(\tau;\theta) &= \nabla_\theta \sum_{t\geq 0} \log p(s_{t+1}|s_t, a_t) + \log \pi_\theta(a_t|s_t), \\
&= \sum_{t\geq 0} \nabla_\theta \log p(s_{t+1}|s_t, a_t) + \nabla_\theta \log \pi_\theta(a_t|s_t), \\
&= \sum_{t\geq 0} \nabla_\theta \log \pi_\theta(a_t|s_t).
\end{aligned}$$

Finally, we have a gradient estimator:

$$\nabla_\theta \hat{J}(\theta) = \frac{1}{n}\sum_{i=1}^{n} r(\tau_i) \sum_{t\geq 0} \nabla_\theta \log \pi_\theta(a_t|s_t) = \frac{1}{n}\sum_{i=1}^{n}\sum_{t\geq 0} r(\tau_i)\nabla_\theta \log \pi_\theta(a_t|s_t).$$

Dependence

Machine learning systems are usually driven by statistical dependence.



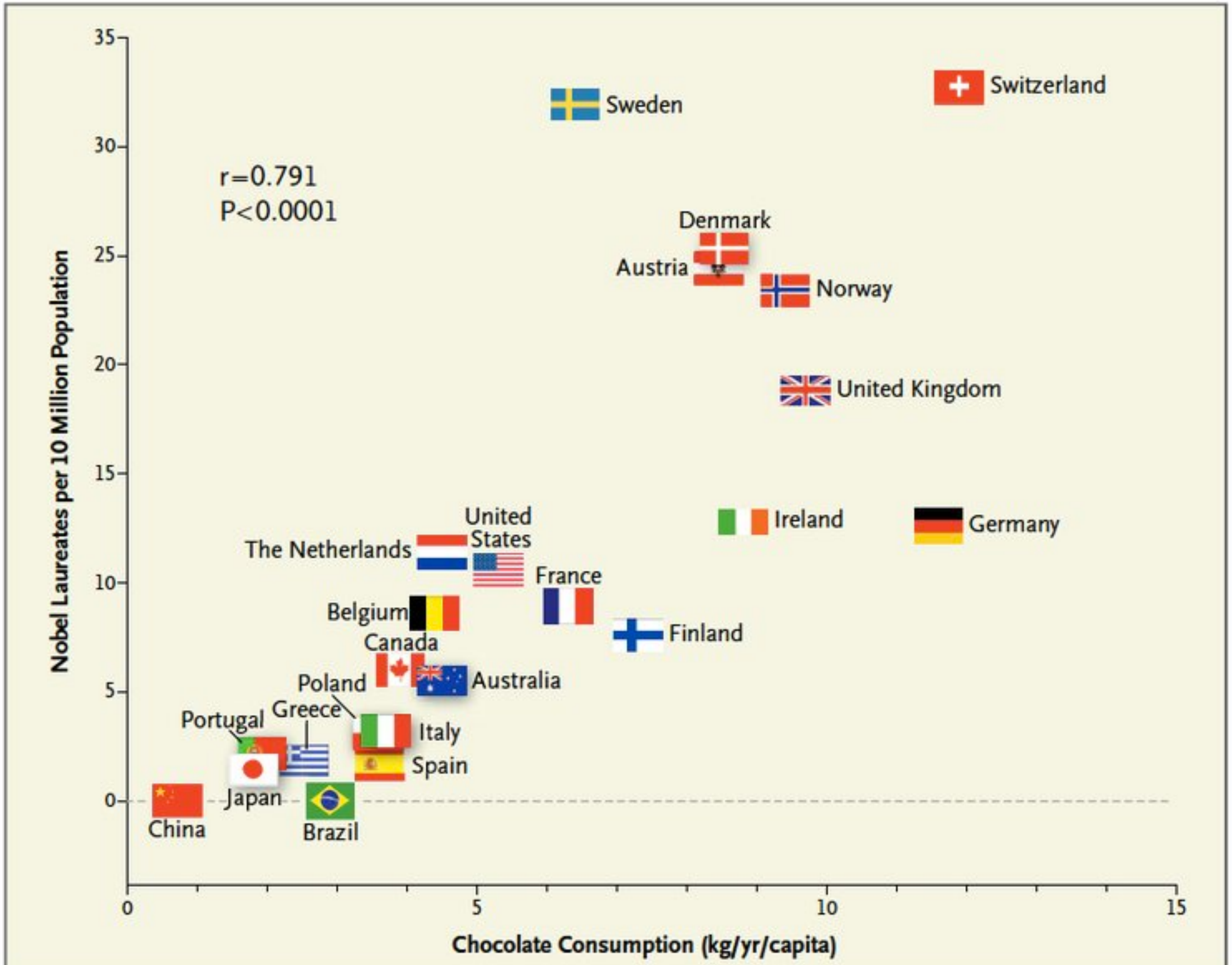


Figure 1. **Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.**

Messerli, Franz H. "Chocolate Consumption, Cognitive Function, and Nobel Laureates." (2012).

THE UNIVERSITY OF SYDNEY

The NEW ENGLAND JOURNAL of MEDICINE

HOME | ARTICLES & MULTIMEDIA | ISSUES | SPECIALTIES & TOPICS | FOR AUTHORS | CME

ORIGINAL ARTICLE

A Correction Has Been Published

Association of Coffee Drinking with Total and Cause-Specific Mortality

Neal D. Freedman, Ph.D., Yikyung Park, Sc.D., Christian C. Abnet, Ph.D., Albert R. Hollenbeck, Ph.D., and Rashmi Sinha, Ph.D.

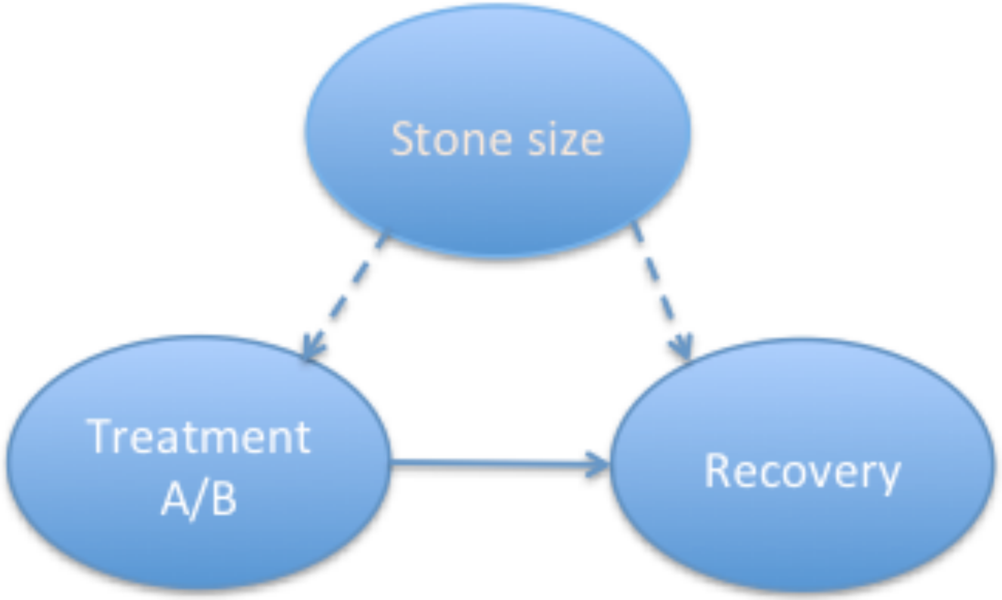
N Engl J Med 2012; 366:1891-1904 | May 17, 2012 | DOI: 10.1056/NEJMoa1112010

- The risk of death was increased among coffee drinkers
- Coffee drinkers were also more likely to smoke
- After adjustment for smoking, there was a significant inverse association between coffee assumption and mortality

Causal thinking

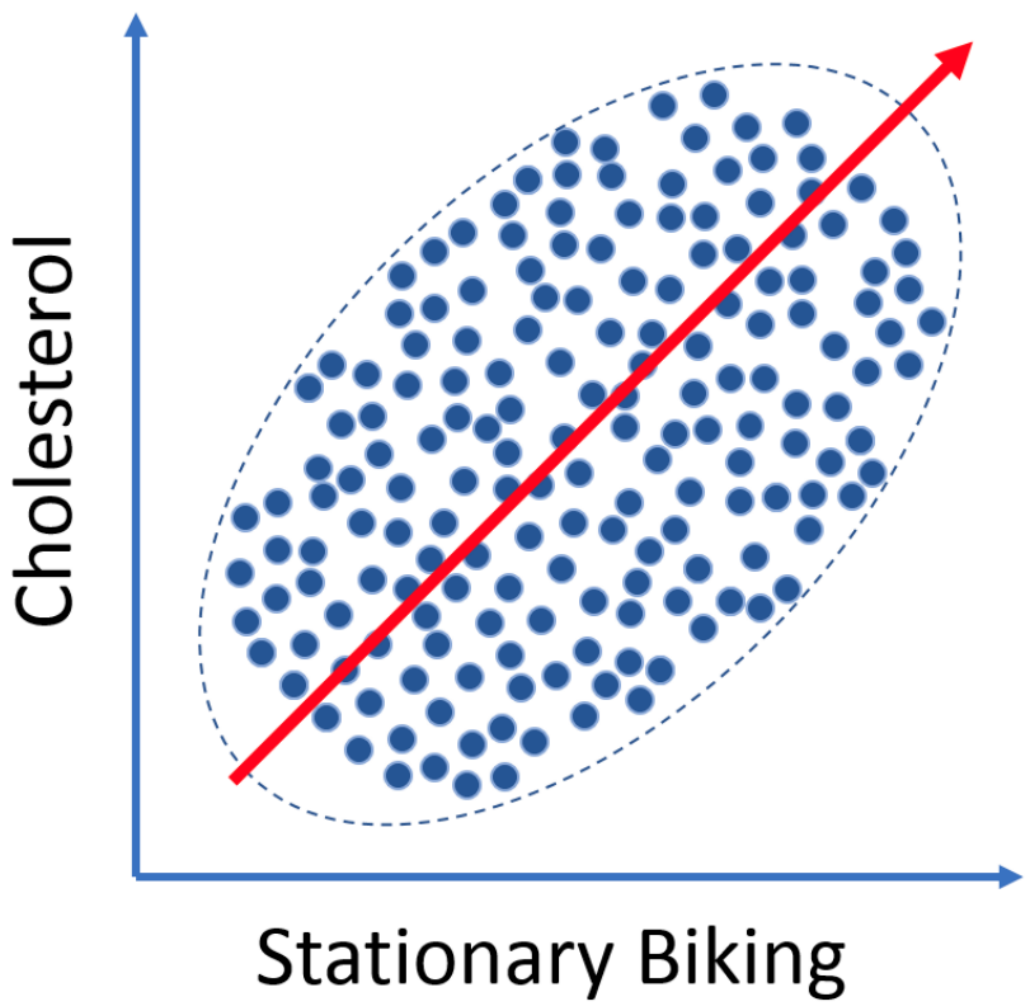
Simpson's Paradox

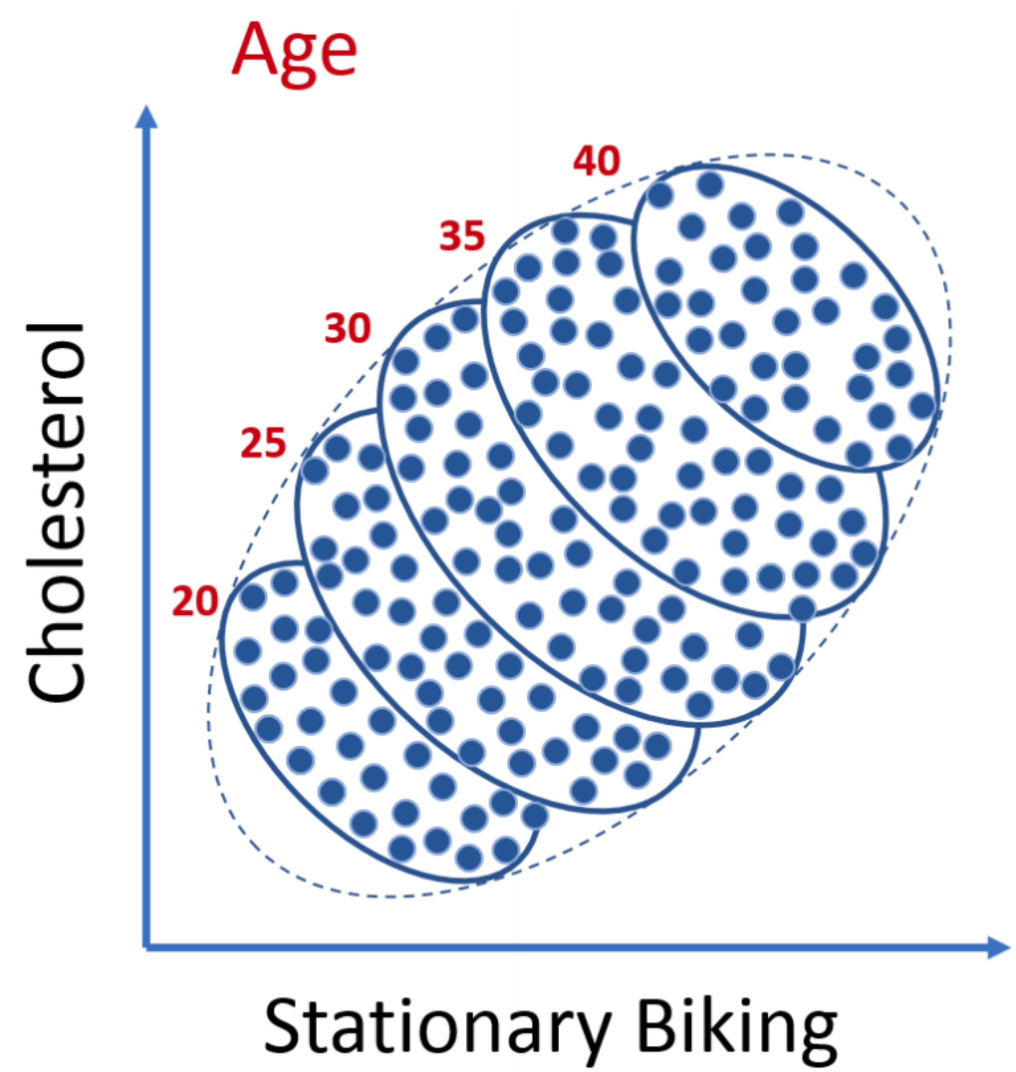
	Treatment A	Treatment B
Small Stones	*Group 1* **93% (81/87)**	*Group 2* 87% (234/270)
Large Stones	*Group 3* **73% (192/263)**	*Group 4* 69% (55/80)
Both	78% (273/350)	**83% (289/350)**



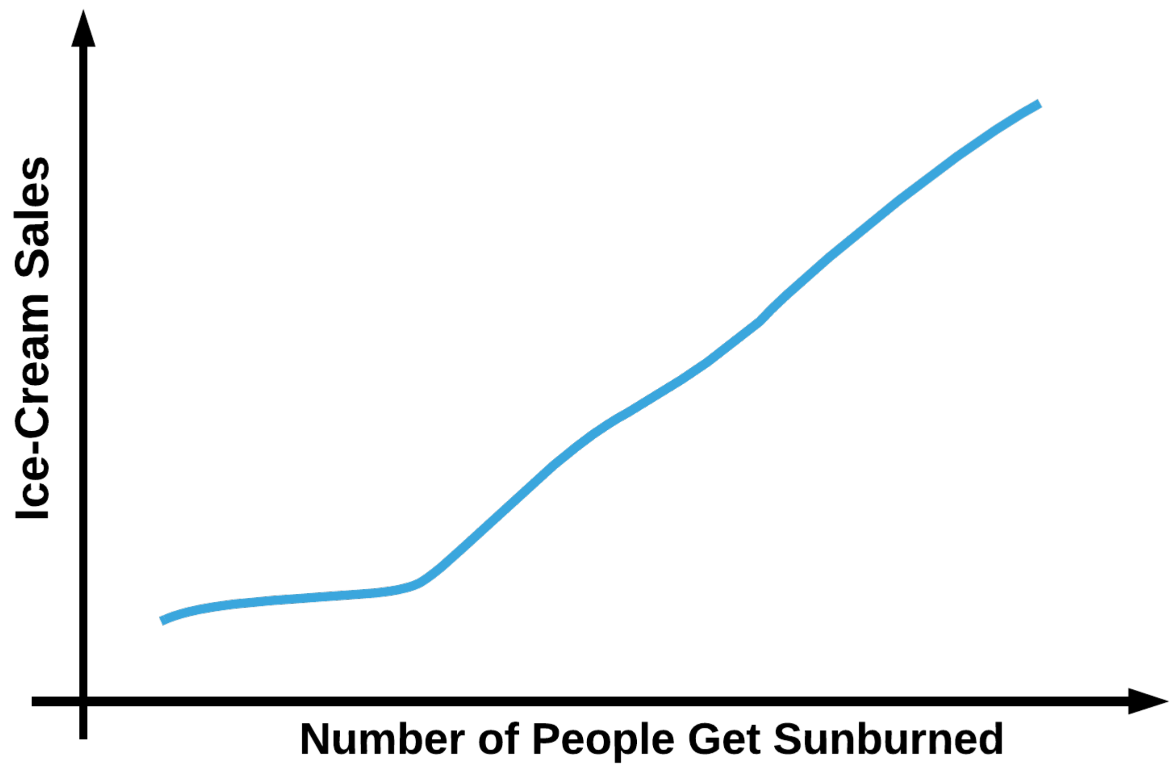
Would you make recommendations based on correlation or something else?

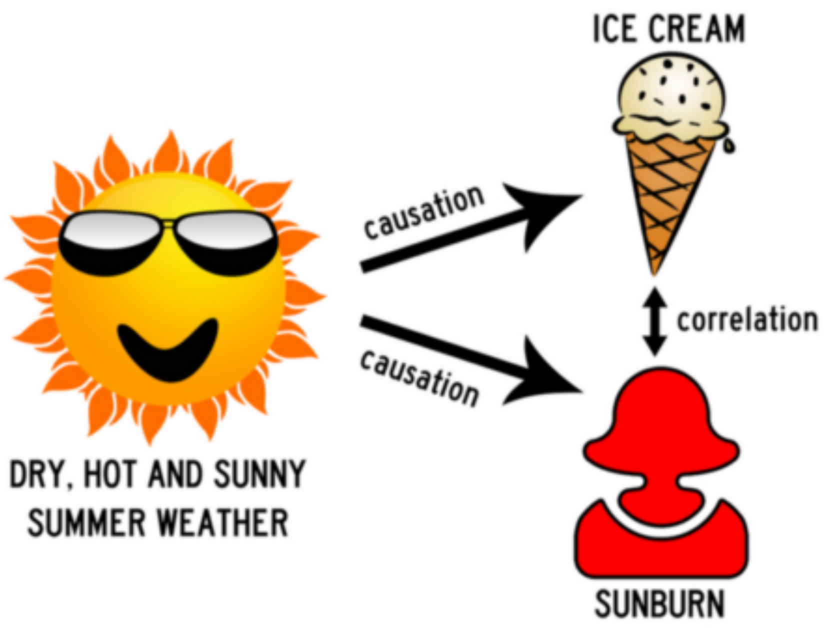
“Strange” Dependence





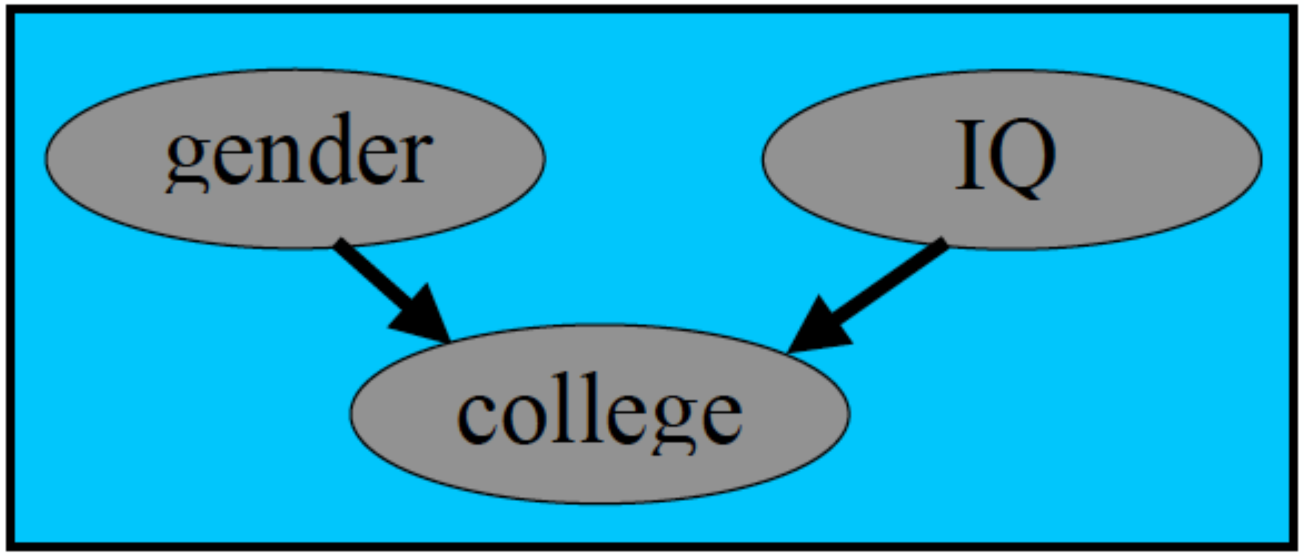
“Strange” Dependence





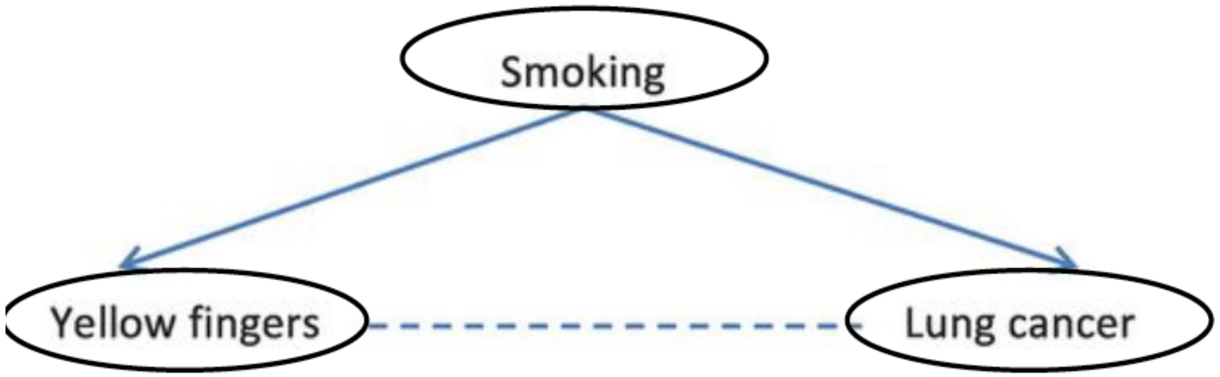
“Strange” Dependence

Let’s go back 50 years; maybe you’ll find female college students are smarter than male ones on average. Why?

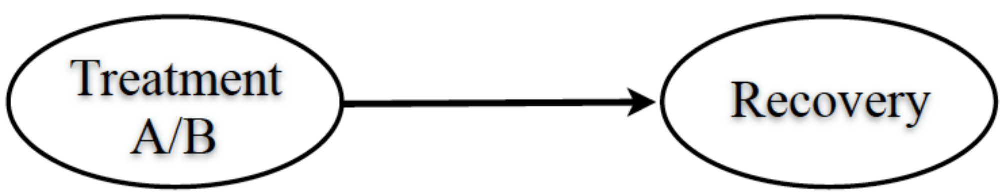


Selection bias!

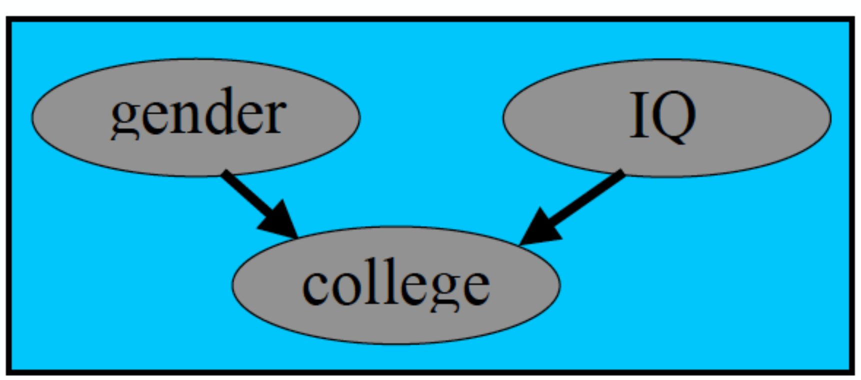
Ways to Produce Dependence



Common Cause



Causal relation



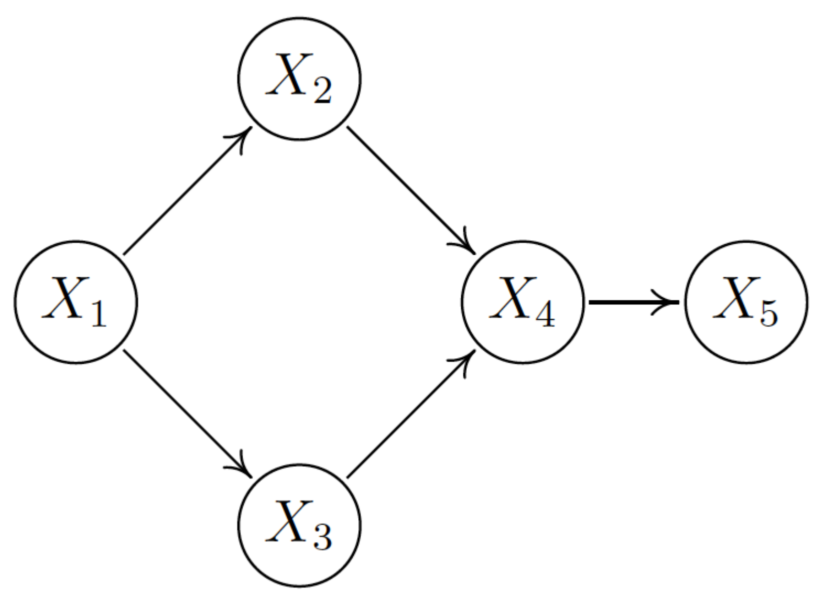
Conditional dependence given common effect

Causation vs Dependence

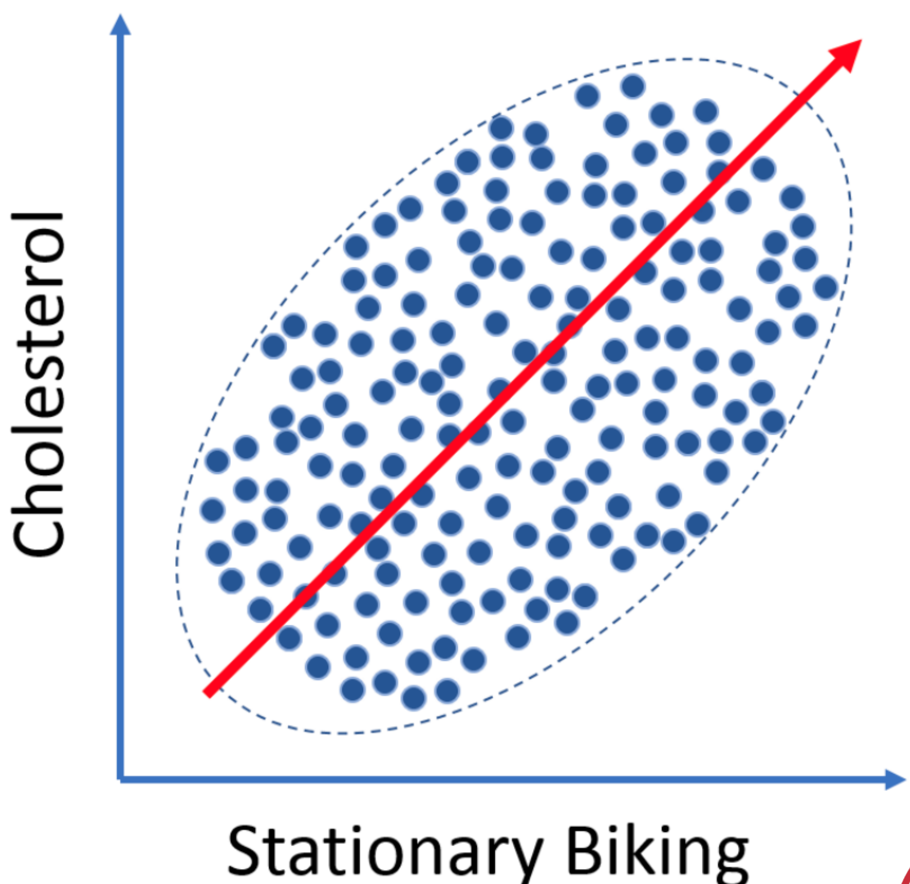
Causal representation

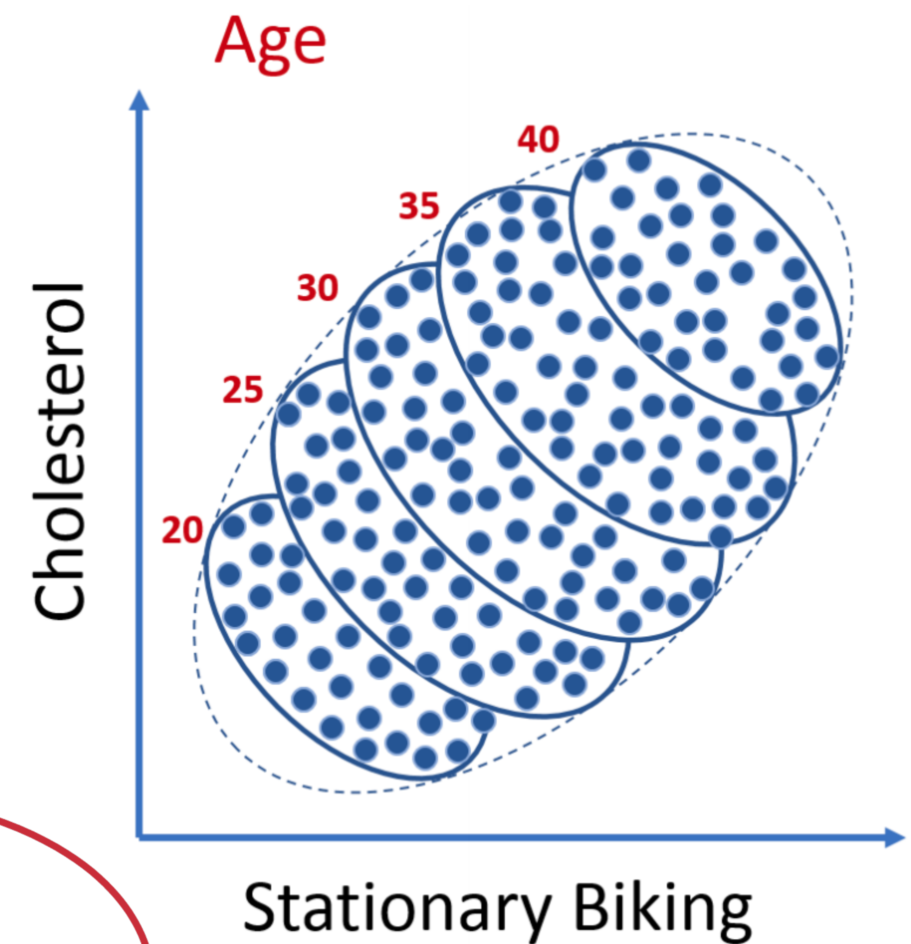
Directed Acyclic Graph (DAG)

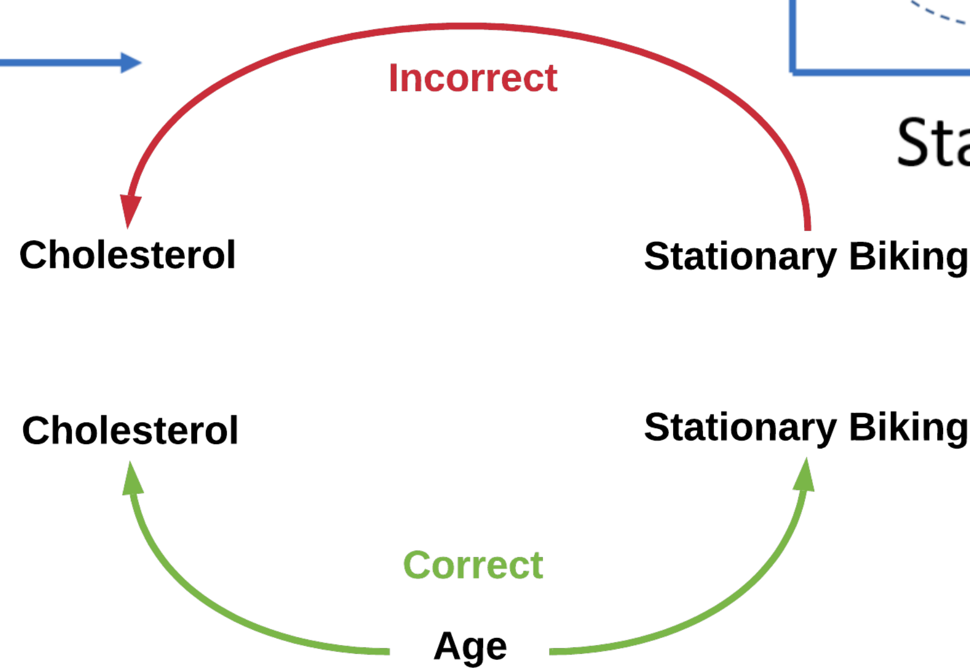
- **DAG**
  - graph with directed edges, no cycles
- **Path**
  - a sequence of connected vertices
  - X1 - X2 - X4 - X5
  - X1 - X2 - X4 - X3



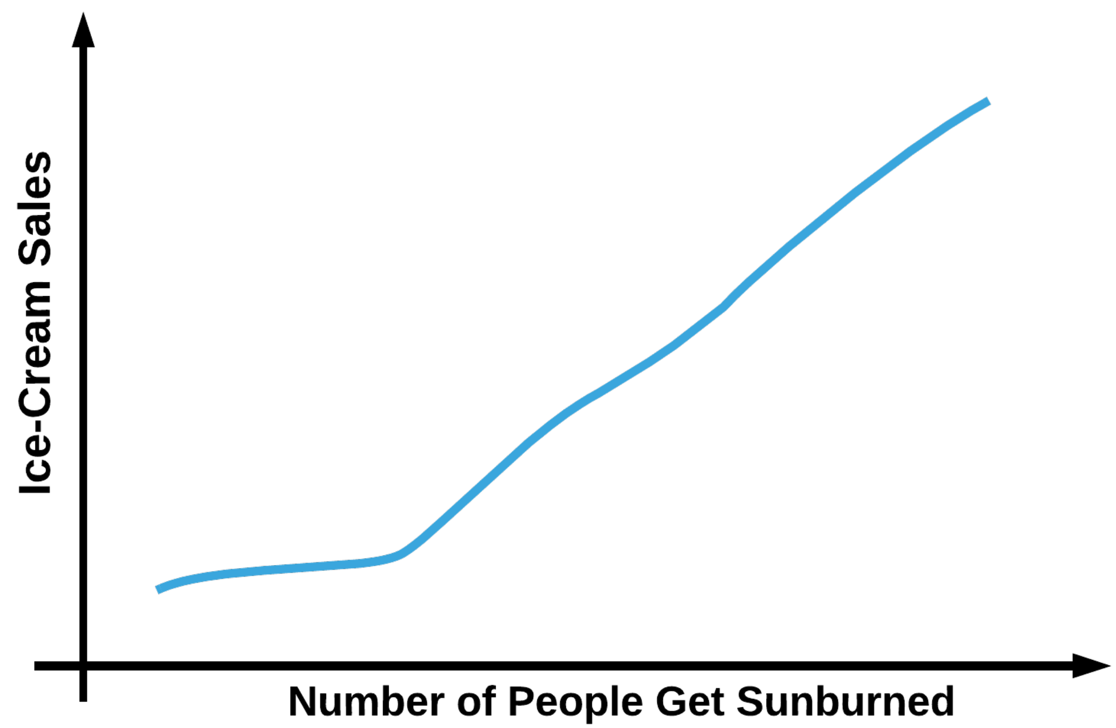
“Strange” Dependence

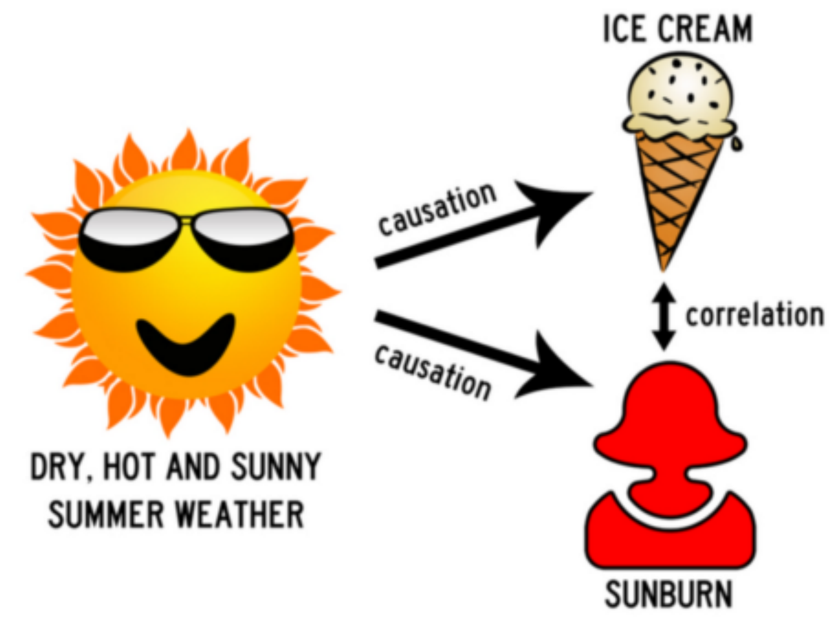


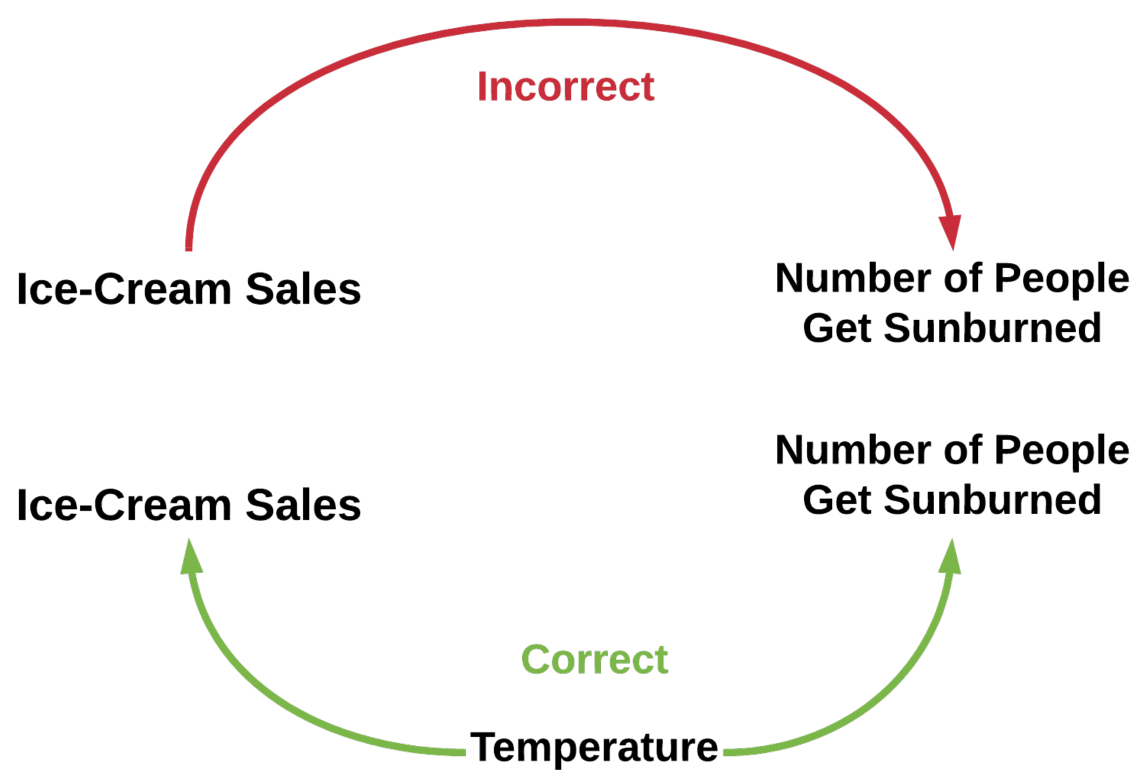




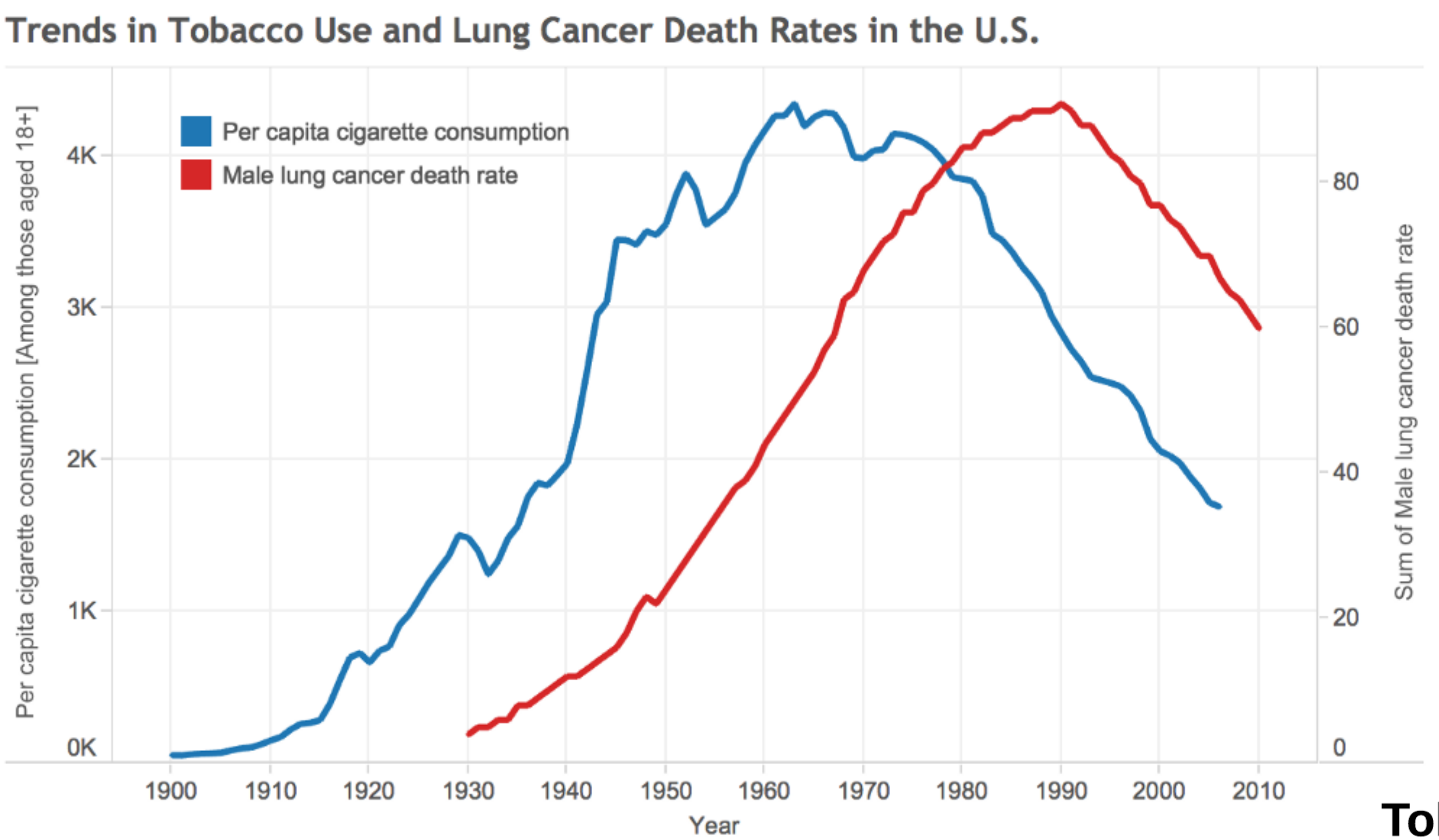
“Strange” Dependence



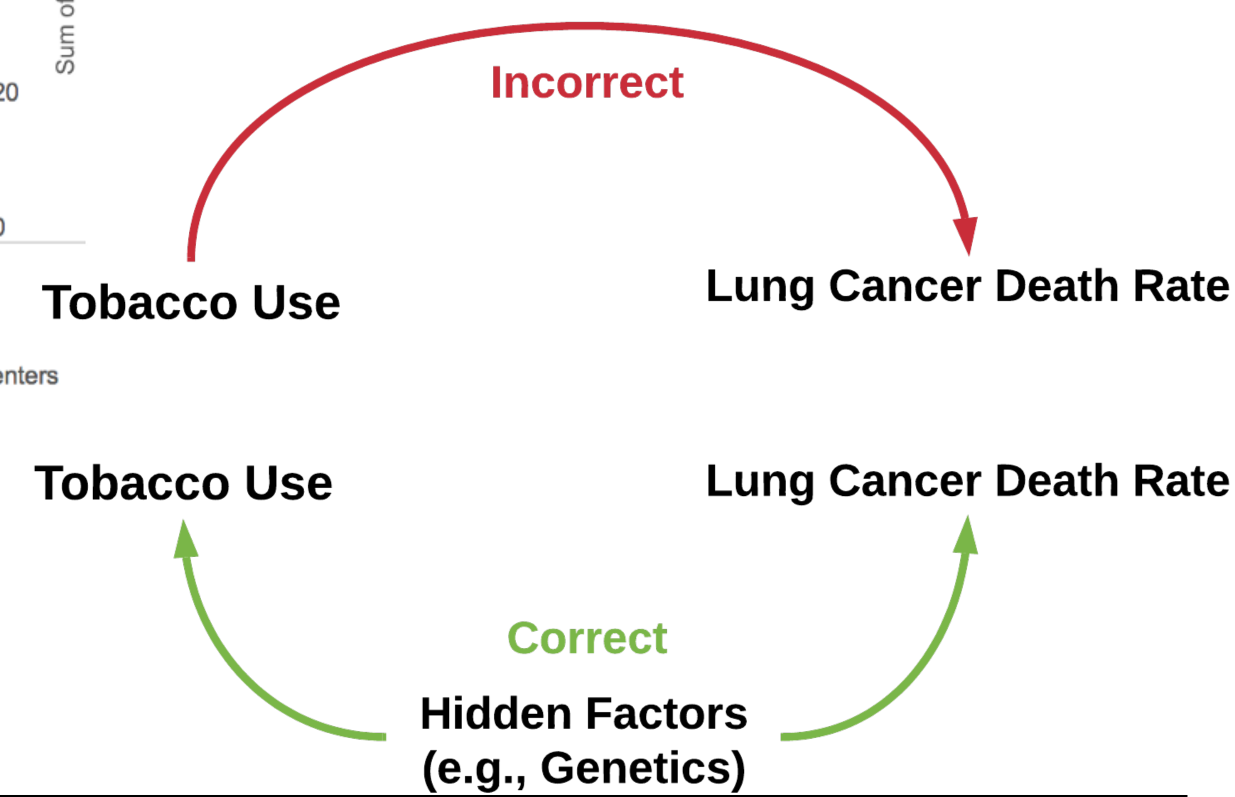




“Strange” Dependence



Death rates source: US Mortality Data, 1960-2010, US Mortality Volumes, 1930-1959, National Center for Health Statistics, Centers for Disease Control and Prevention.
Cigarette consumption source: US Department of Agriculture, 1900-2007.



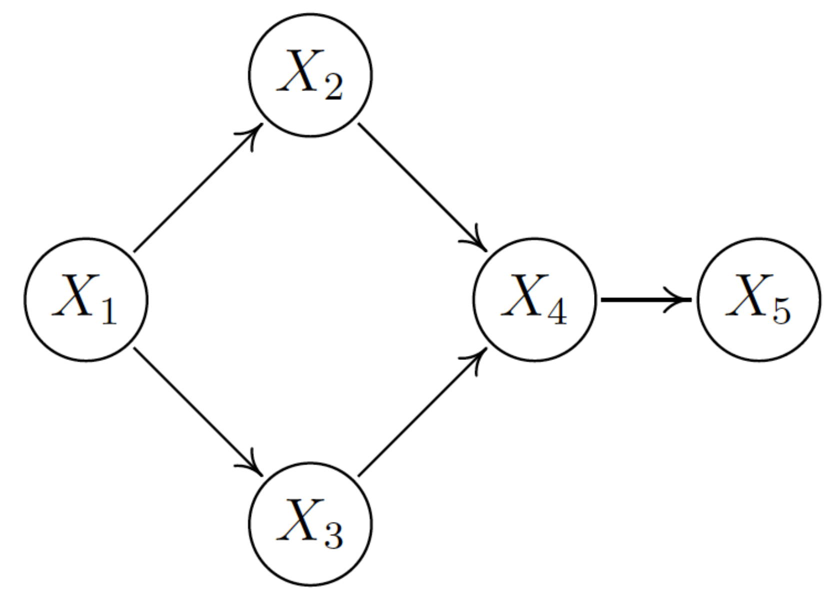
Markov Conditions

- Markov conditions state that if a certain graph property holds true, then a certain statistically independence holds true.

- There are local Markov condition and global Markov condition.

Local Markov Condition

Every variable X_i, in a directed acyclic graph, is independent of its non-descendant Y conditional on its parents, i.e., $(X_i \perp\!\!\!\perp Y \mid PA_i)_G$, which also implies $(X_i \perp\!\!\!\perp Y \mid PA_i)_p$.



For example:
$(X_4 \perp\!\!\!\perp X_1 \mid \{X_2, X_3\})_G \implies (X_4 \perp\!\!\!\perp X_1 \mid \{X_2, X_3\})_p$.

By the local Markov condition,
we could obtain a casual factorisation of the joint distribution as follows

$$P(X_1, \ldots, X_5) = P(X_1)P(X_2 \mid X_1)P(X_3 \mid X_1)P(X_4 \mid X_2, X_3)P(X_5 \mid X_4)$$

Global Markov Conditions (D-Separation)

For three disjoint sets of variables $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{S}$, $\mathbf{X}$ is d-separated from $\mathbf{Y}$ conditional on $\mathbf{S}$ if and only if all paths between any member of $\mathbf{X}$ and any member of $\mathbf{Y}$ are *blocked* by $\mathbf{S}$.

A path q is said to be blocked by the set $\mathbf{S}$ if

- q contains a chain $i \to m \to j$ or a fork $i \leftarrow m \to j$ such that the middle node m is in $\mathbf{S}$, or
- q contains a collider $i \to m \leftarrow j$ such that the middle node m is not in $\mathbf{S}$, and no descendant of m is in $\mathbf{S}$.

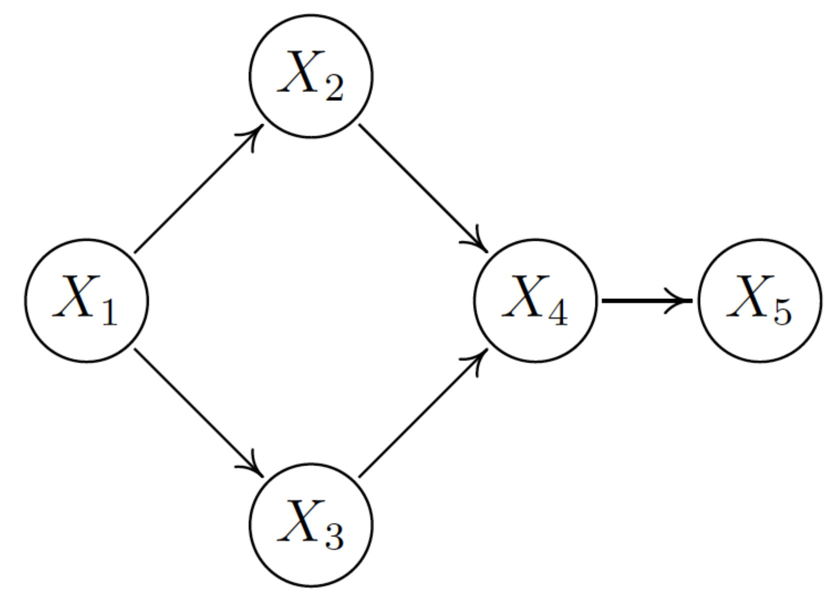
Formally, we use $(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{S})_G$ to denote that $\mathbf{S}$ d-separates $\mathbf{X}$ and $\mathbf{Y}$ in the DAG G, which also implies $(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{S})_p$.

Global Markov Conditions (D-Separation)

- Let $\mathbf{X} = \{X_1\}, \mathbf{Y} = \{X_4, X_5\}$, then $\mathbf{X}$ is d-separated from $\mathbf{Y}$ conditional on $\mathbf{S} = \{X_2, X_3\}$ (Because all paths from $\mathbf{X}$ to $\mathbf{Y}$ are blocked by conditioning on $\mathbf{S}$).

- Let $\mathbf{X} = \{X_2\}, \mathbf{Y} = \{X_3\}$, then $\mathbf{X}$ is not d-separated from $\mathbf{Y}$ conditional on $\mathbf{S} = \{X_1, X_4\}$.

- Let $\mathbf{X} = \{X_2\}, \mathbf{Y} = \{X_3\}$, then $\mathbf{X}$ is d-separated from $\mathbf{Y}$ conditional on $\mathbf{S} = \{X_1\}$.



Casual faithfulness Assumption

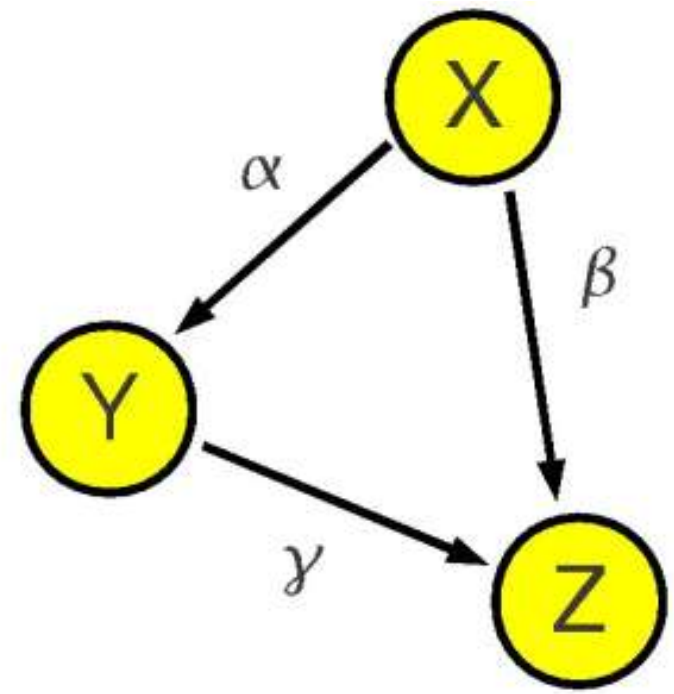
The probability distribution may have additional conditional independence relations that are not entailed by d-separation applied to a graph. When no such extra conditional independence relations hold the distribution is said to be faithful to the graph, i.e.,

$$(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{S})_p \implies (\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{S})_G,$$

then the distribution p is said to be faithful to the graph.

Casual faithfulness Assumption

Here is an example of a unfaithful distribution

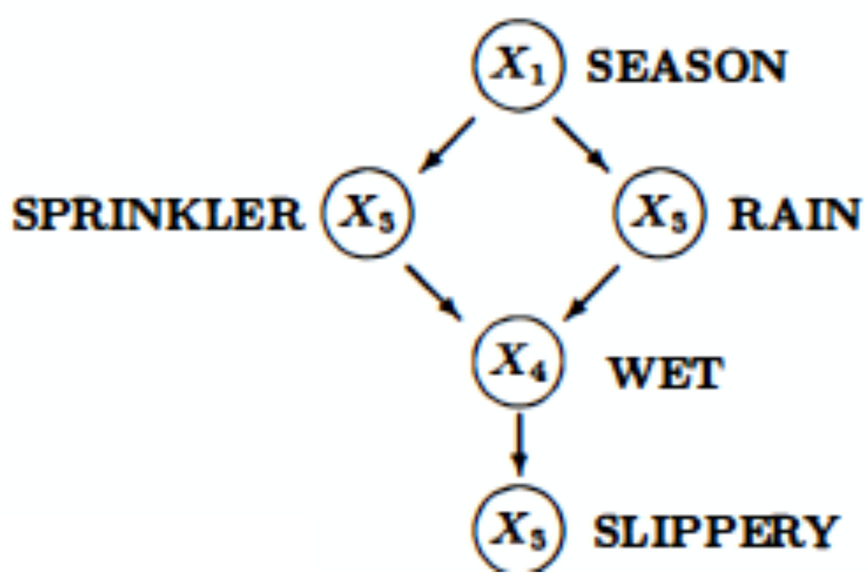


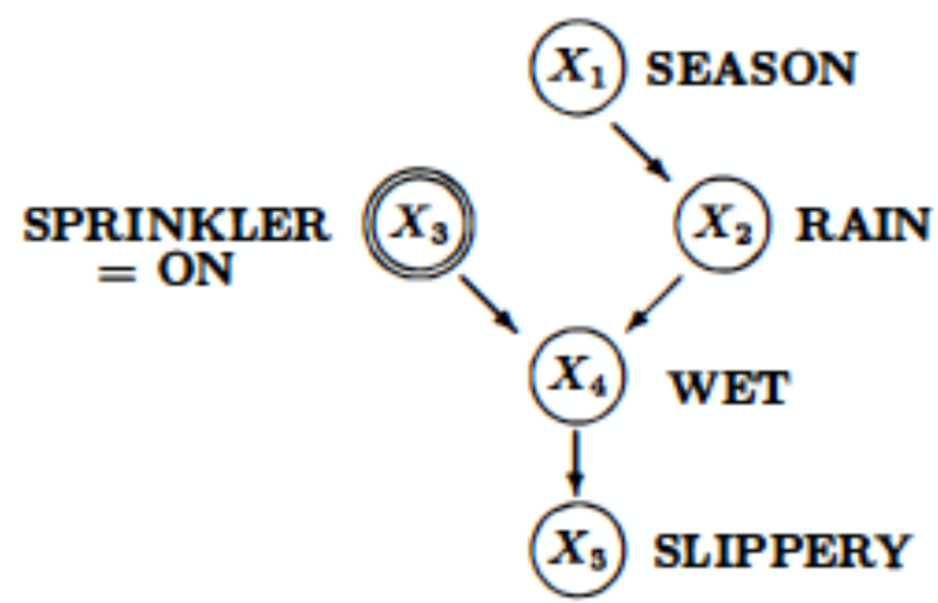
$X = U_x$,
$Y = \alpha X + U_y$,
$Z = \beta X + \gamma Y + U_z$.

From the graph $(\{X\} \not\perp\!\!\!\perp \{Z\})_G$, however, if $\beta + \alpha\gamma = 0$, then $(\{X\} \perp\!\!\!\perp \{Z\})_p$, and $(\{X\} \perp\!\!\!\perp \{Z\})_p$ does not imply $(\{X\} \perp\!\!\!\perp \{Z\})_G$.

Causal Bayesian Net

- **Causal Bayesian Network**
  - directed edges representation causal direction, causal DAG
  - more meaningful and represent external changes





Conditioning, Intervention, Counterfactual

- **Prediction/Conditioning**
  - would the pavement be slippery if we find the sprinkler off?

$$P(slippery \mid Sprinkler = off)$$

- **Intervention**
  - would the pavement be slippery if we turn off the sprinkler?

$$P(slippery \mid do(Sprinkler = off))$$

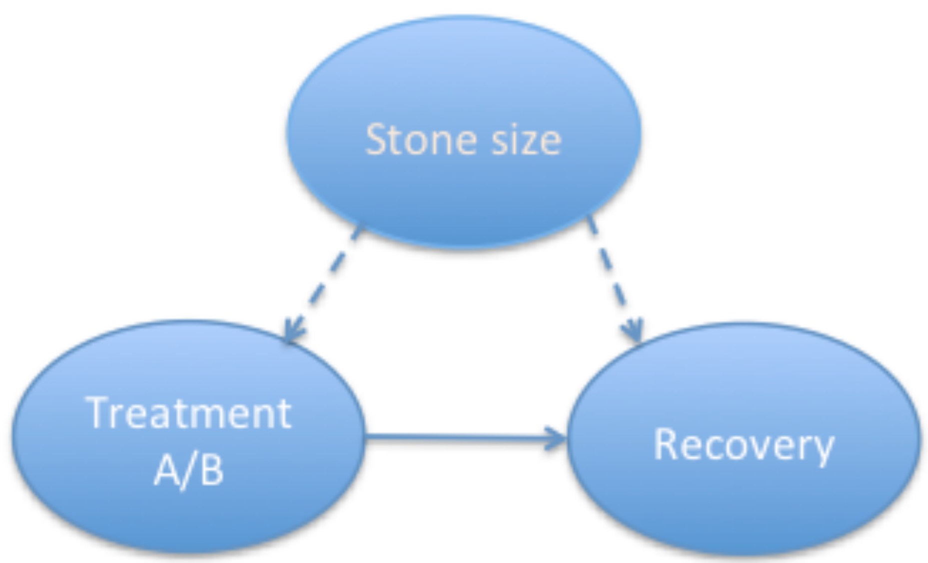
- **Prediction/Counterfactual reasoning**
  - would the pavement be slippery, had the sprinkler been off, given that the pavement is in fact not slippery and the sprinkler is on?

$$P(slippery_{sprinkler=off} \mid Sprinkler = on, Slippery = no)$$

Identification of Causal Effects

$$P(Recovery \mid do(Treatment = A))$$

- **“Golden standard”: randomised controlled experiments**
  - All the other factors that influence the outcome variable are either fixed or vary at random, so any changes in the outcome variable must be due to the controlled variable



- Usually expensive or infeasible to do!

Example

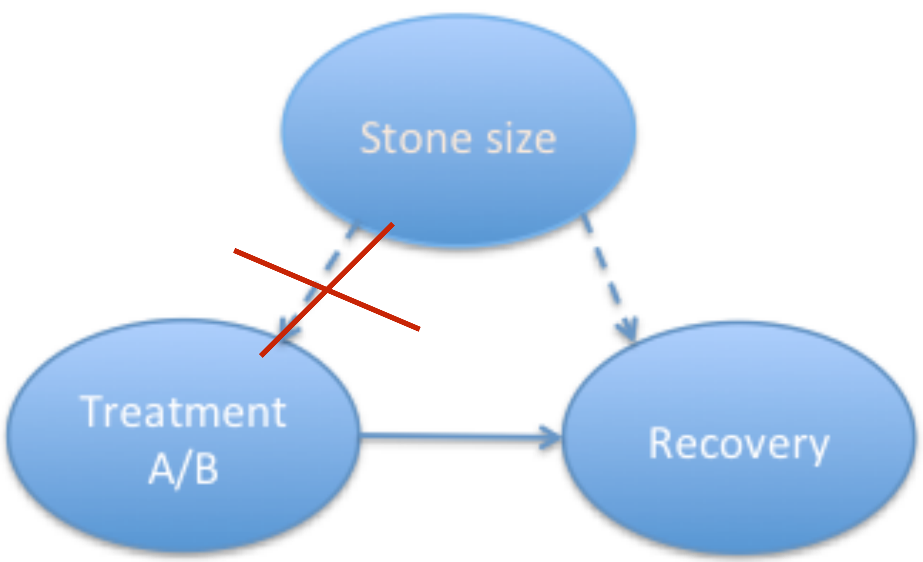
	Treatment A	Treatment B
Small Stones	*Group 1* **93% (81/87)**	*Group 2* 87% (234/270)
Large Stones	*Group 3* **73% (192/263)**	*Group 4* 69% (55/80)
Both	78% (273/350)	**83% (289/350)**

T - Treatment

R - Recovery

$$P(\mathrm{R} \mid T) = \sum_S P(R \mid T, S) P(S \mid T)$$

$$P(\mathrm{R} \mid do(T)) = \sum_S P(R \mid T, S) P(S)$$



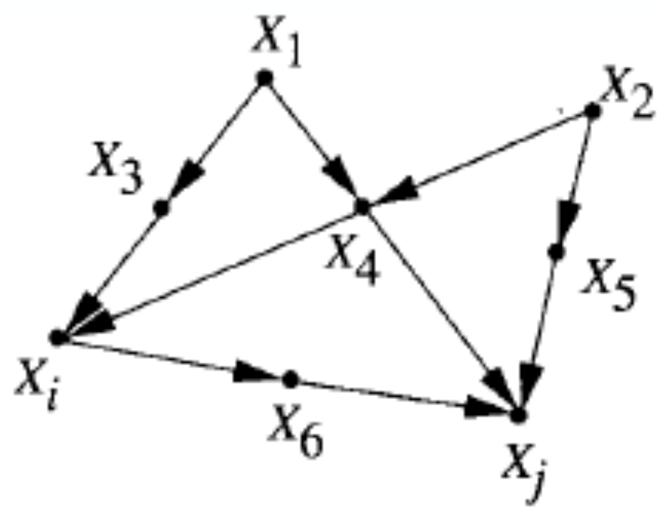
Intervention vs. conditioning

Back-Door Criterion

Definition 3.3.1 (Back-Door)
A set of variables Z satisfies the back-door *criterion relative to an ordered pair of variables (X_i, X_j) in a DAG G if:*

(i) *no node in Z is a descendant of X_i; and*

(ii) *Z blocks every path between X_i and X_j that contains an arrow into X_i.*



- What if $Z = \{X_3, X_4\}$?
  $Z = \{X_4, X_5\}$?
  $Z = \{X_4\}$?
- What if there is a confounder?

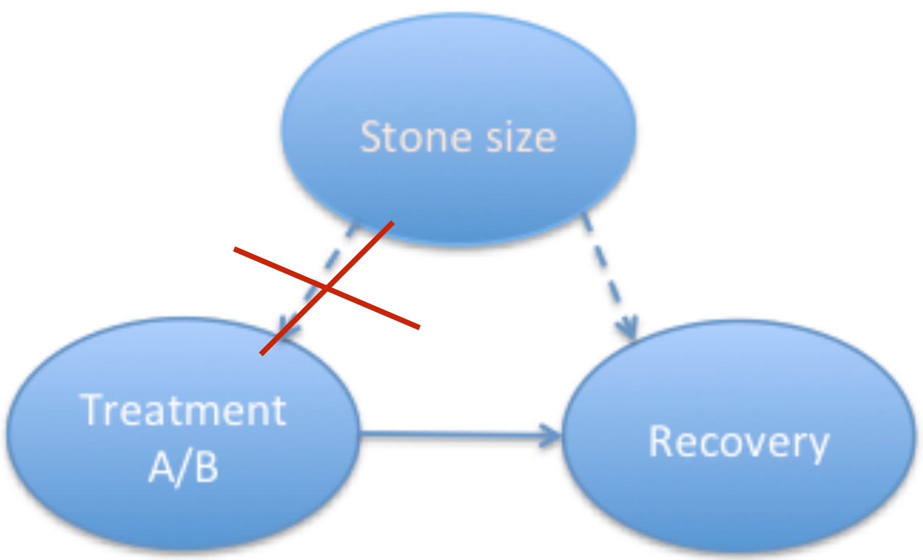
Theorem 3.3.2 (Back-Door Adjustment)
If a set of variables Z satisfies the back-door criterion relative to (X, Y), then the causal effect of X on Y is identifiable and is given by the formula

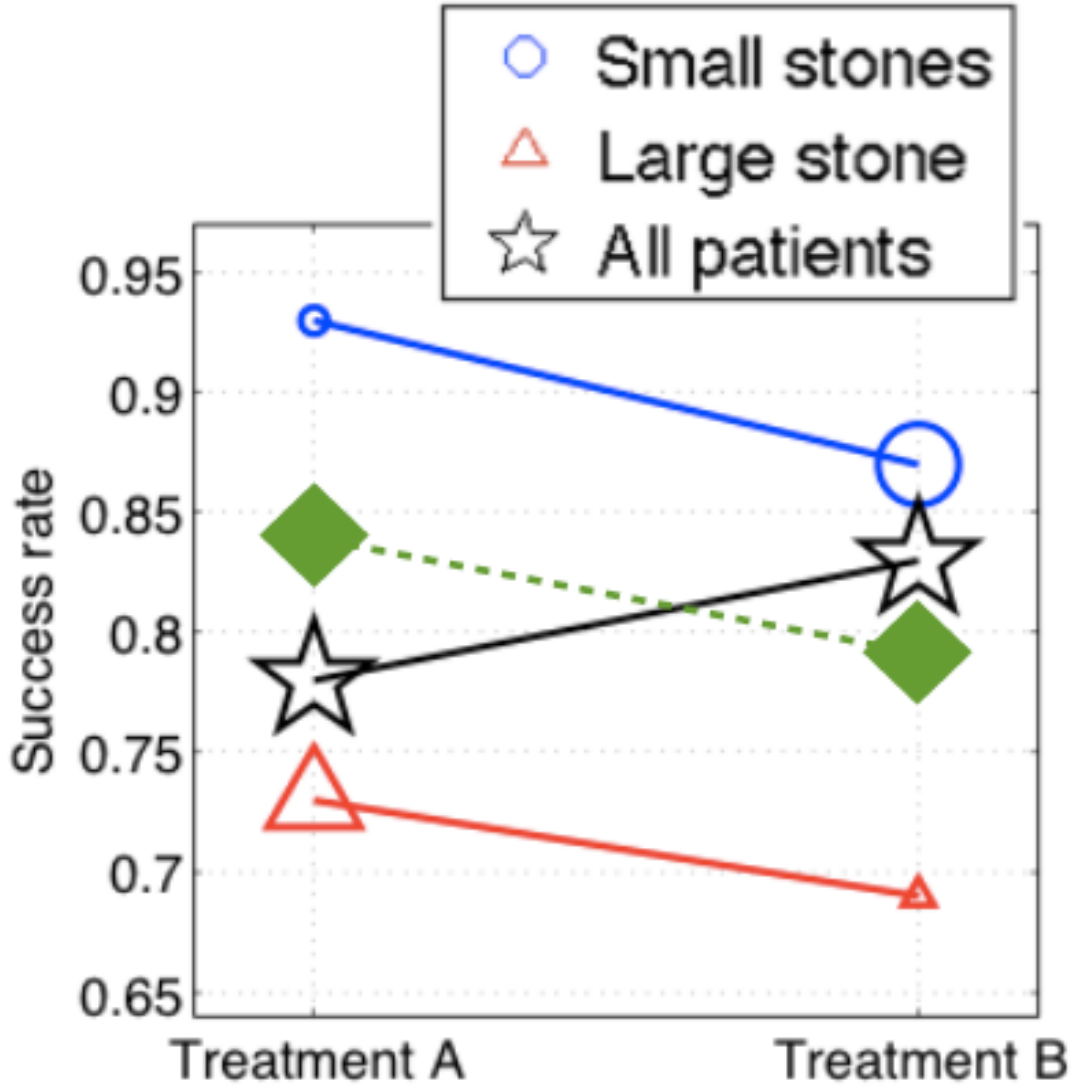
$$P(y \mid \hat{x}) = \sum_z P(y \mid x, z) P(z).$$

Example

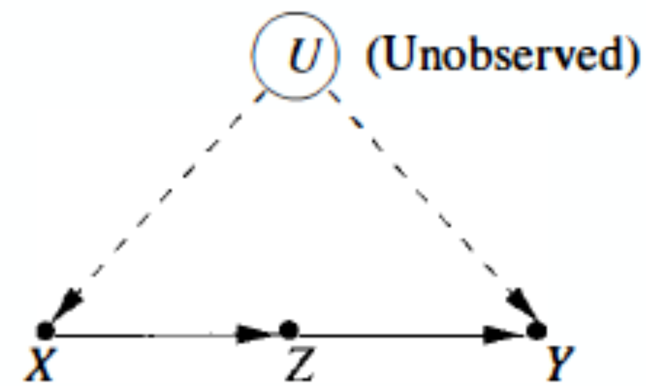
	Treatment A	Treatment B
Small Stones	*Group 1* **93% (81/87)**	*Group 2* 87% (234/270)
Large Stones	*Group 3* **73% (192/263)**	*Group 4* 69% (55/80)
Both	78% (273/350)	**83% (289/350)**

T - Treatment
R - Recovery





Front-Door Criterion



Definition 3.3.3 (Front-Door)

A set of variables Z is said to satisfy the front-door *criterion relative to an ordered pair of variables* (X, Y) *if:*

(i) *Z intercepts all directed paths from X to Y;*

(ii) *there is no back-door path from X to Z; and*

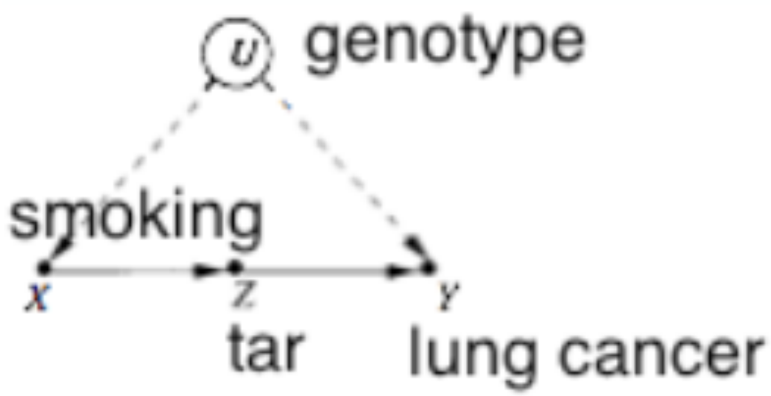
(iii) *all back-door paths from Z to Y are blocked by X.*

Theorem 3.3.4 (Front-Door Adjustment)

If Z satisfies the front-door criterion relative to (X, Y) *and if* $P(x, z) > 0$, *then the causal effect of X on Y is identifiable and is given by the formula*

$$P(y \mid \hat{x}) = \sum_z P(z \mid x) \sum_{x'} P(y \mid x', z) P(x'). \tag{3.29}$$

Example: Smoking

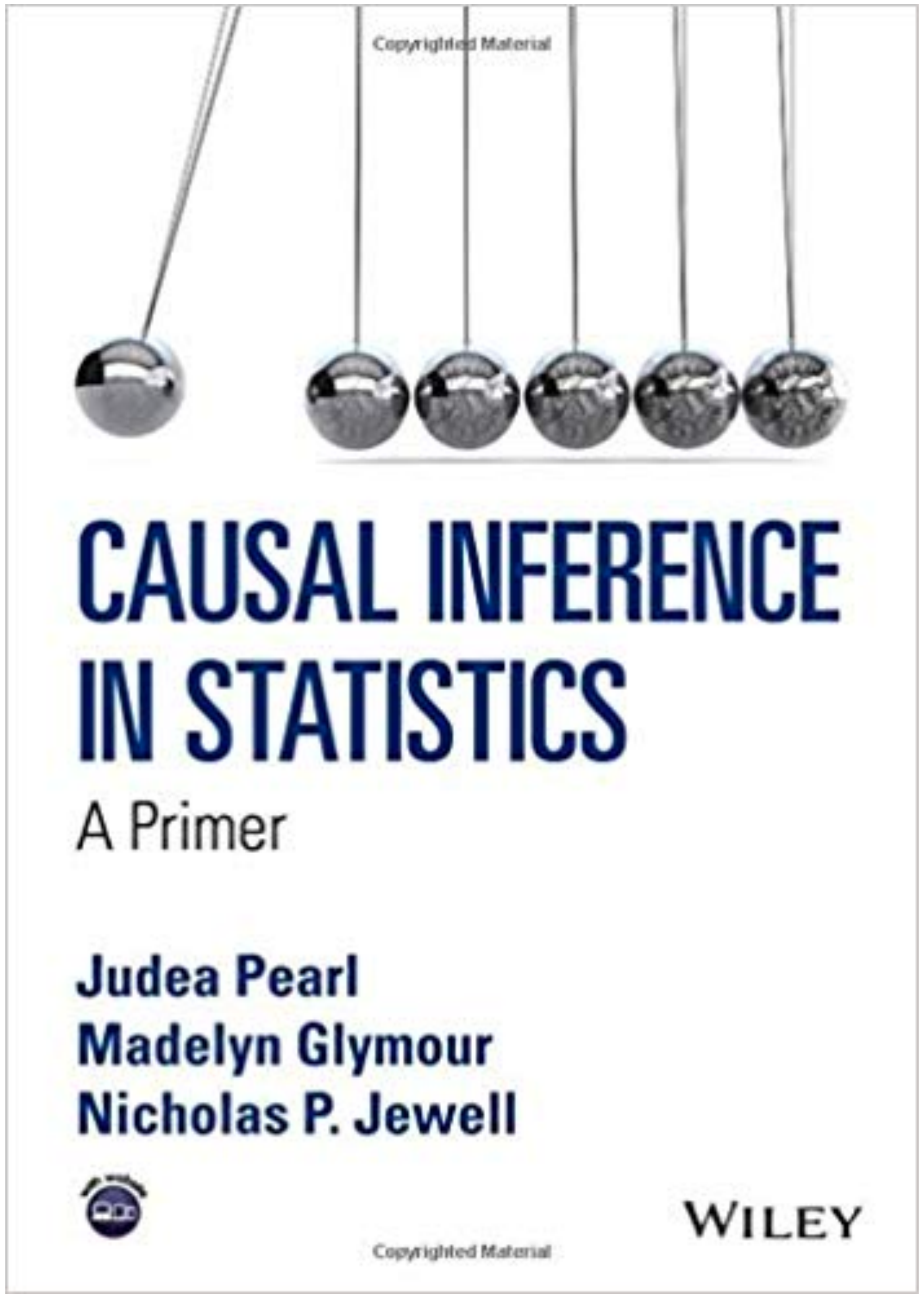


	Group Type	$P(x, z)$ Group Size (% of Population)	$P(Y = 1 \mid x, z)$ % of Cancer Cases in Group
$X = 0,\ Z = 0$	Nonsmokers, No tar	47.5	10
$X = 1,\ Z = 0$	Smokers, No tar	2.5	90
$X = 0,\ Z = 1$	Nonsmokers, Tar	2.5	5
$X = 1,\ Z = 1$	Smokers, Tar	47.5	85

$$\begin{aligned} P(Y = 1 \mid do(X = 1)) &= .05(.10 \times .50 + .90 \times .50) \\ &\quad + .95(.05 \times .50 + .85 \times .50) \\ &= .05 \times .50 + .95 \times .45 = .4525, \end{aligned}$$

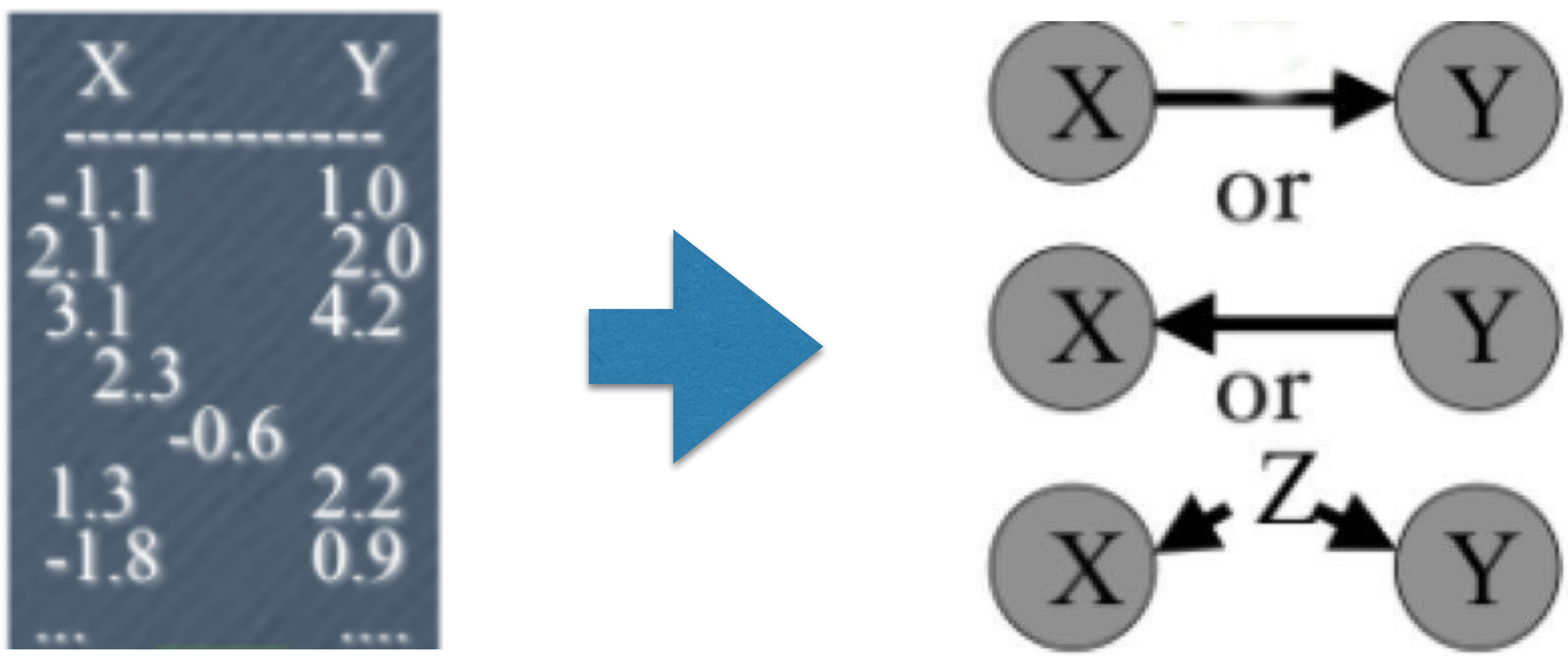
$$\begin{aligned} P(Y = 1 \mid do(X = 0)) &= .95(.10 \times .50 + .90 \times .50) \\ &\quad + .05(.05 \times .50 + .85 \times .50) \\ &= .95 \times .50 + .05 \times .45 = .4975. \end{aligned}$$

Causal Inference



Causal Discovery

Is it possible to discover causal structure from purely observational data?



Summary

- Dependence does not equal causality!!!
- Causality is essential in interventional studies.
- Given a causal graph, causal effects can be estimated from observational data.
- Causal graph can be learned from observational data.