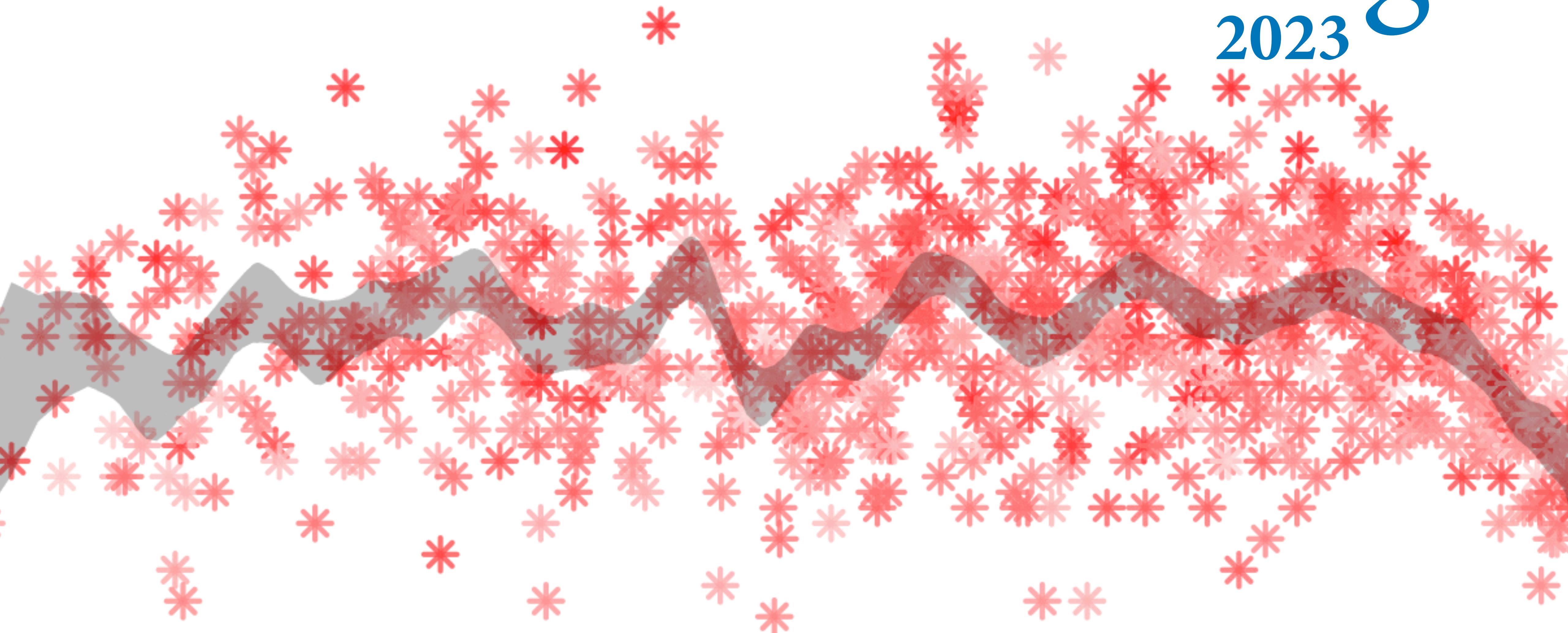
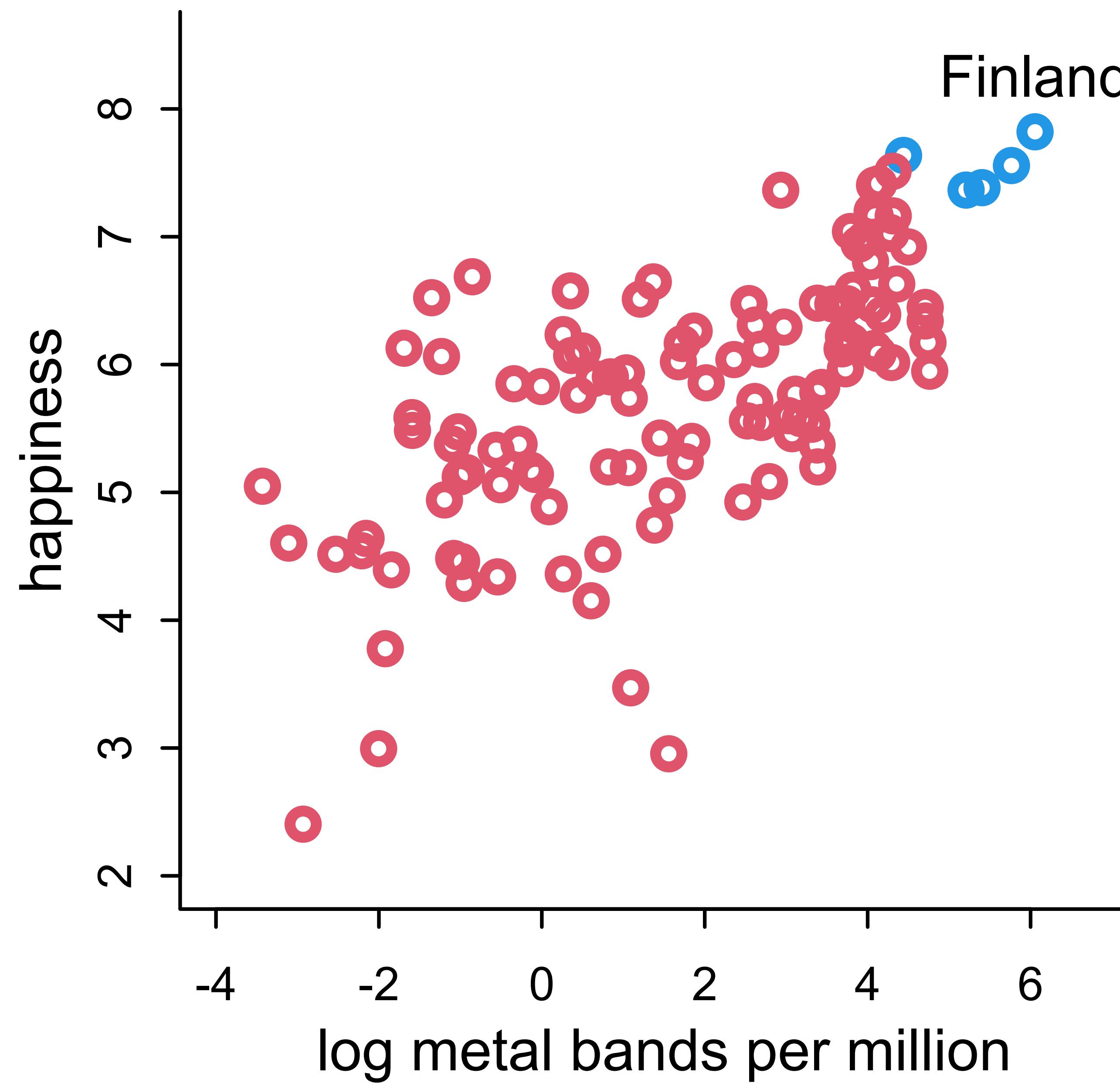


Statistical Rethinking

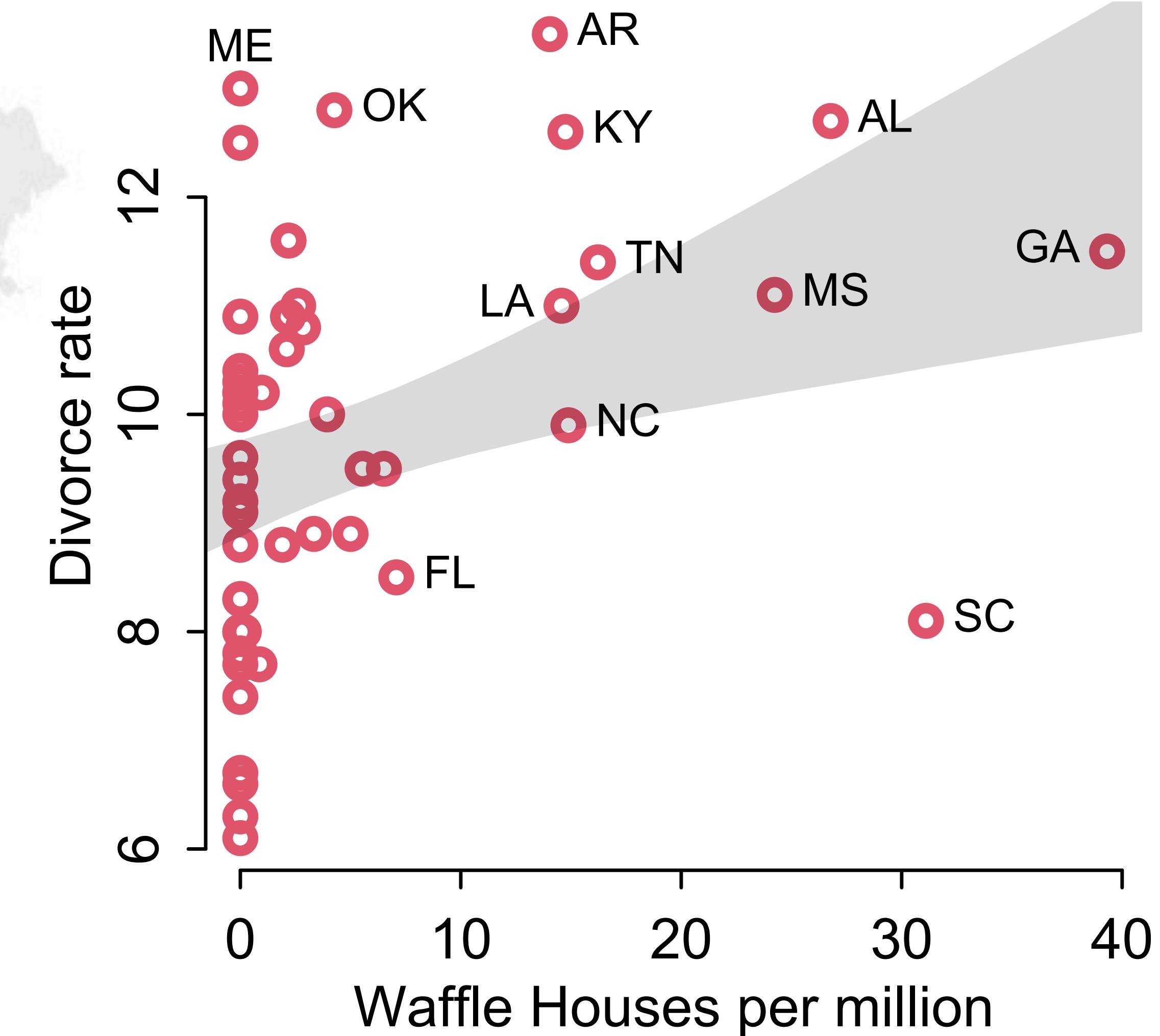
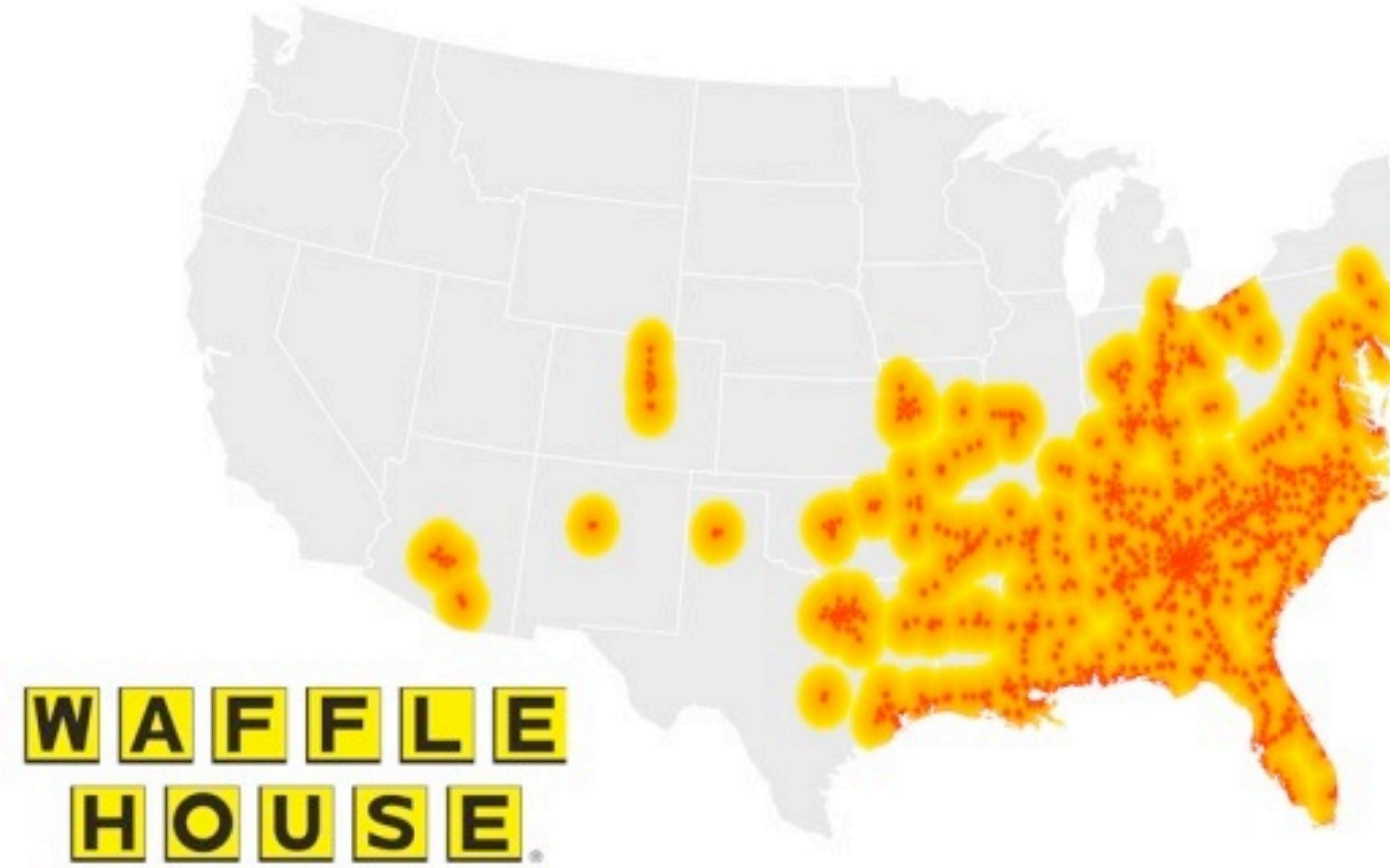
2023



5. Elemental Confounds



Does Waffle House cause divorce?



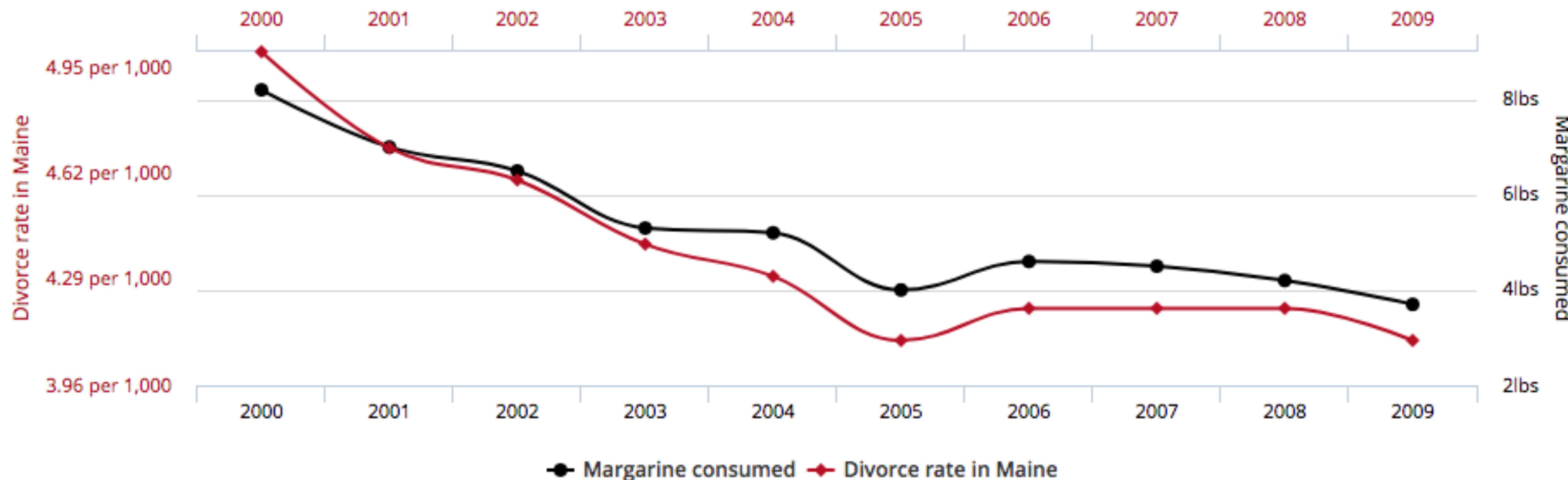
Correlation is commonplace

Divorce rate in Maine

correlates with

Per capita consumption of margarine

Correlation: 99.26% ($r=0.992558$)



ESTIMAND



ESTIMAND



ESTIMATOR

Ingredients

150g unsalted butter	1. Heat oven to 160C.
150g chocolate pieces	Grease 1 liter glass
150g all-purpose flour	baking pan. Line a 450g
1/2 tsp baking powder	loaf tin with baking paper.
1/2 tsp baking soda	2. Melt butter and
200g brown sugar	chocolate in a saucepan
2 large eggs	over low heat.

ESTIMAND



ESTIMATOR

Ingredients

150g unsalted butter
150g chocolate pieces
150g all-purpose flour
1/2 tsp baking powder
1/2 tsp baking soda
200g brown sugar
2 large eggs

Directions

1. Heat oven to 160C. Grease 1 liter glass baking pan. Line a 450g loaf tin with baking paper.
2. Melt butter and chocolate in a saucepan over low heat.

ESTIMATE

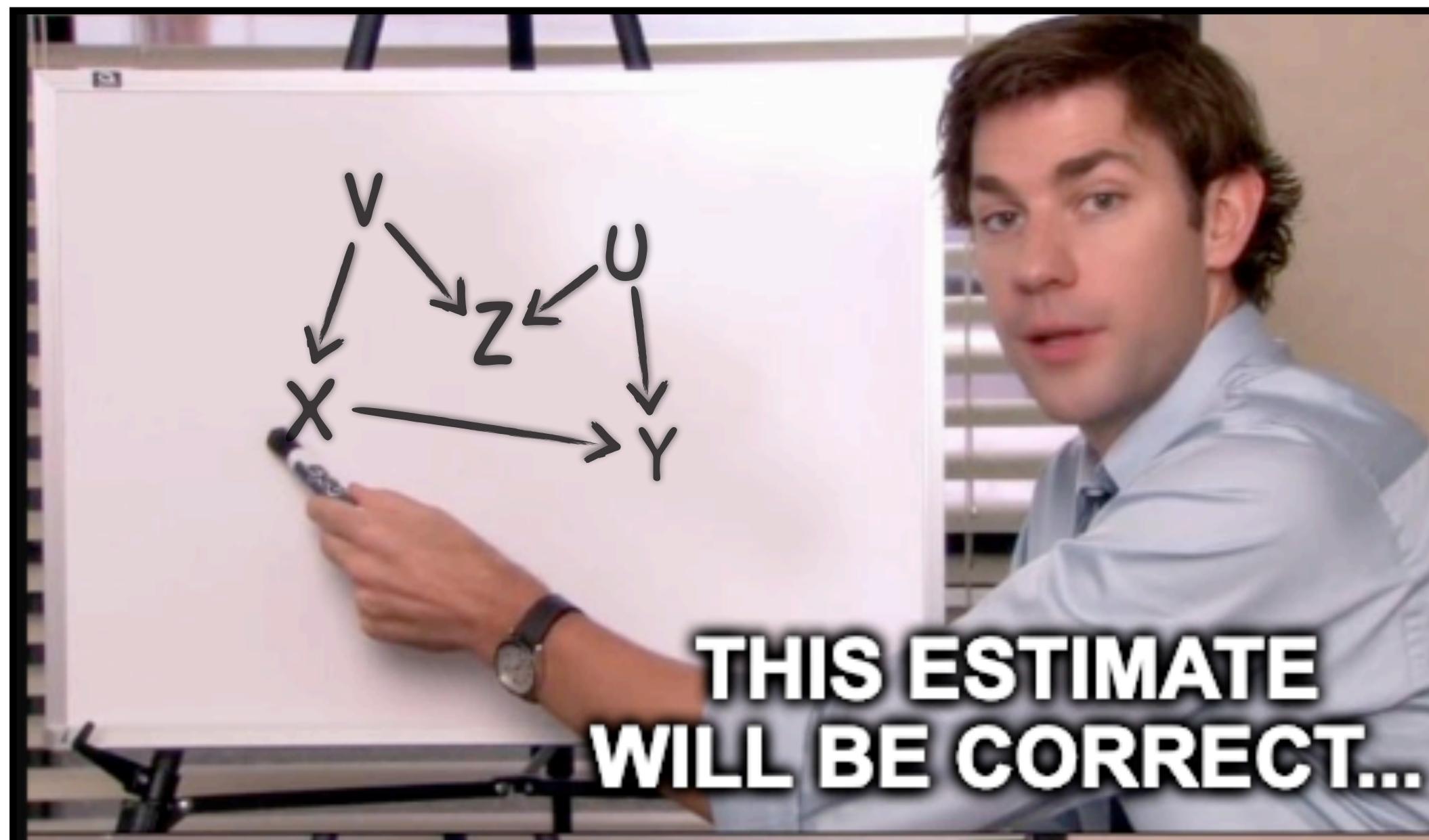


Association & Causation

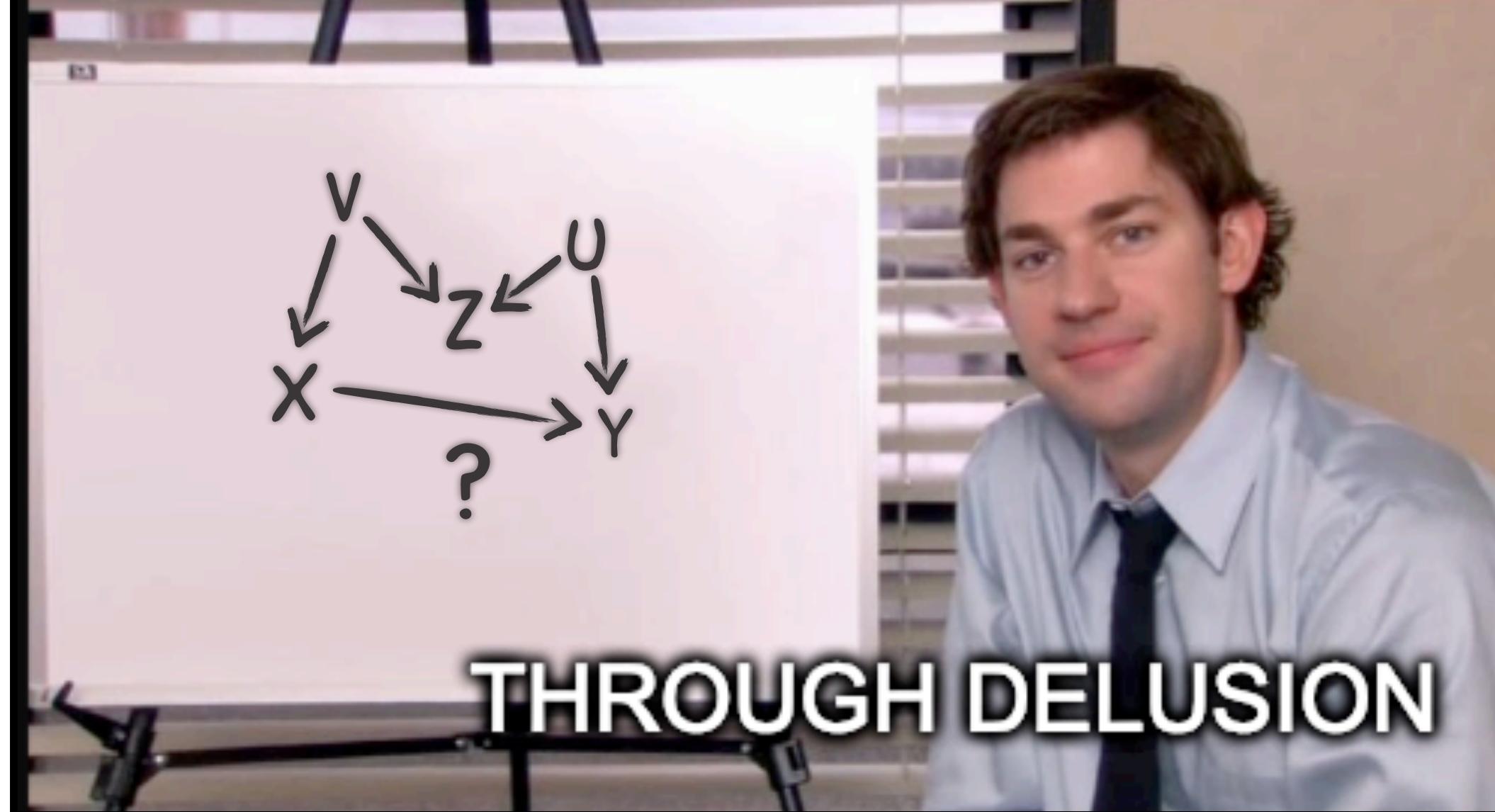
Statistical recipes must defend
against **confounding**

“Confounds”: Features of the sample
& how we use it that mislead us

Confounds are diverse



**THIS ESTIMATE
WILL BE CORRECT...**



THROUGH DELUSION

Ye Olde Causal Alchemy

The Four Elemental Confounds



The Fork



The Pipe



The Collider



The Descendant



↓
A



The Fork



Z is a “common cause”

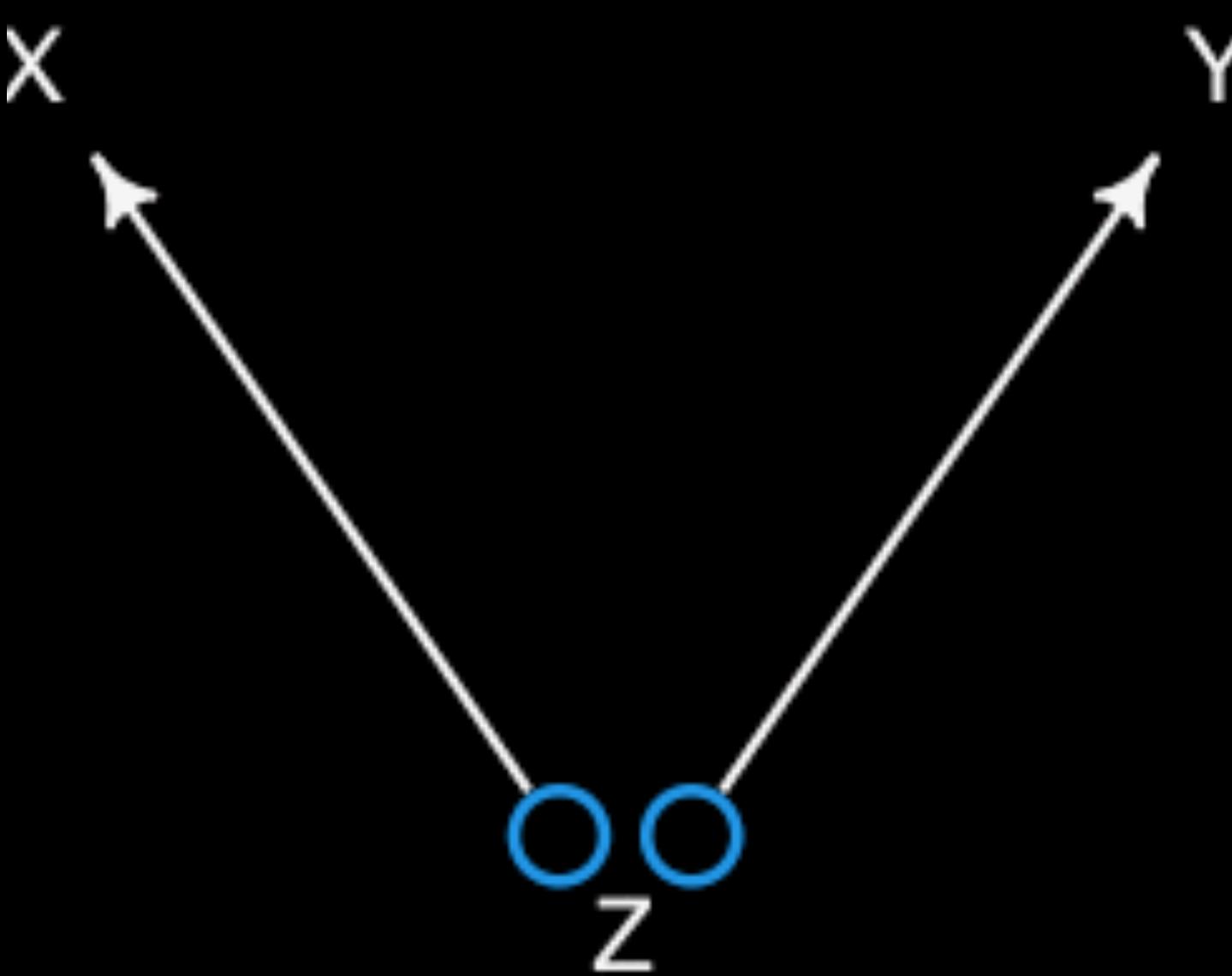
X and Y are associated

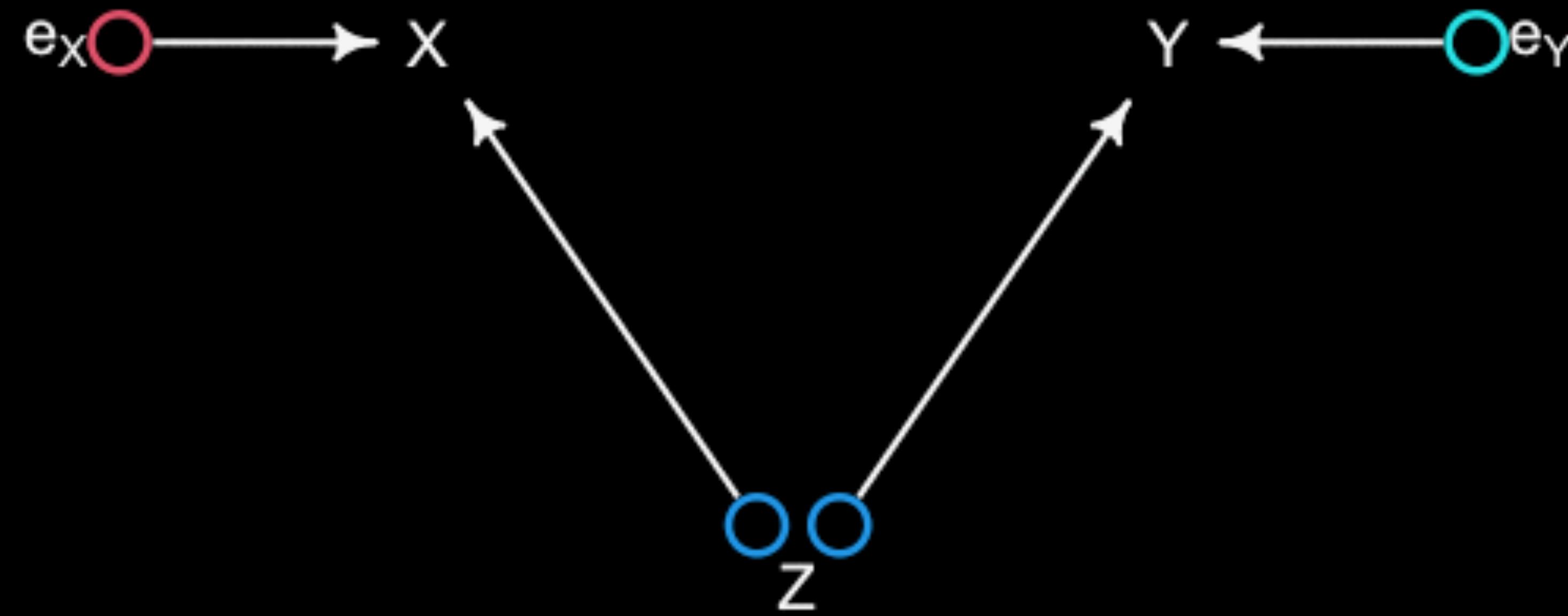
$$Y \not\perp\!\!\!\perp X$$

Share a common cause Z

Once stratified by Z, no association

$$Y \perp\!\!\!\perp X \mid Z$$





$$X \leftarrow Z \rightarrow Y$$

```
n <- 1000  
Z <- rbern( n , 0.5 )  
X <- rbern( n , (1-Z)*0.1 + Z*0.9 )  
Y <- rbern( n , (1-Z)*0.1 + Z*0.9 )
```

		Y
X	0	1
0	397	84
1	100	419

Y ↤ X

```
> cor(X,Y)  
[1] 0.63
```



```
n <- 1000
Z <- rbern( n , 0.5 )
X <- rbern( n , (1-Z)*0.1 + Z*0.9 )
Y <- rbern( n , (1-Z)*0.1 + Z*0.9 )
```

	Y	
X	0	1
0	397	84
1	100	419

$Y \not\perp\!\!\!\perp X$

```
> cor(X,Y)
[1] 0.63
```

Z = 0		
	Y	
X	0	1
0	390	43
1	44	5

Z = 1		
	Y	
X	0	1
0	7	41
1	56	414

$Y \perp\!\!\!\perp X | Z$

```
> cor(X[Z==0],Y[Z==0])
[1] 0.003
> cor(X[Z==1],Y[Z==1])
[1] 0.024
```

$X \leftarrow Z \rightarrow Y$

```
cols <- c(4,2)

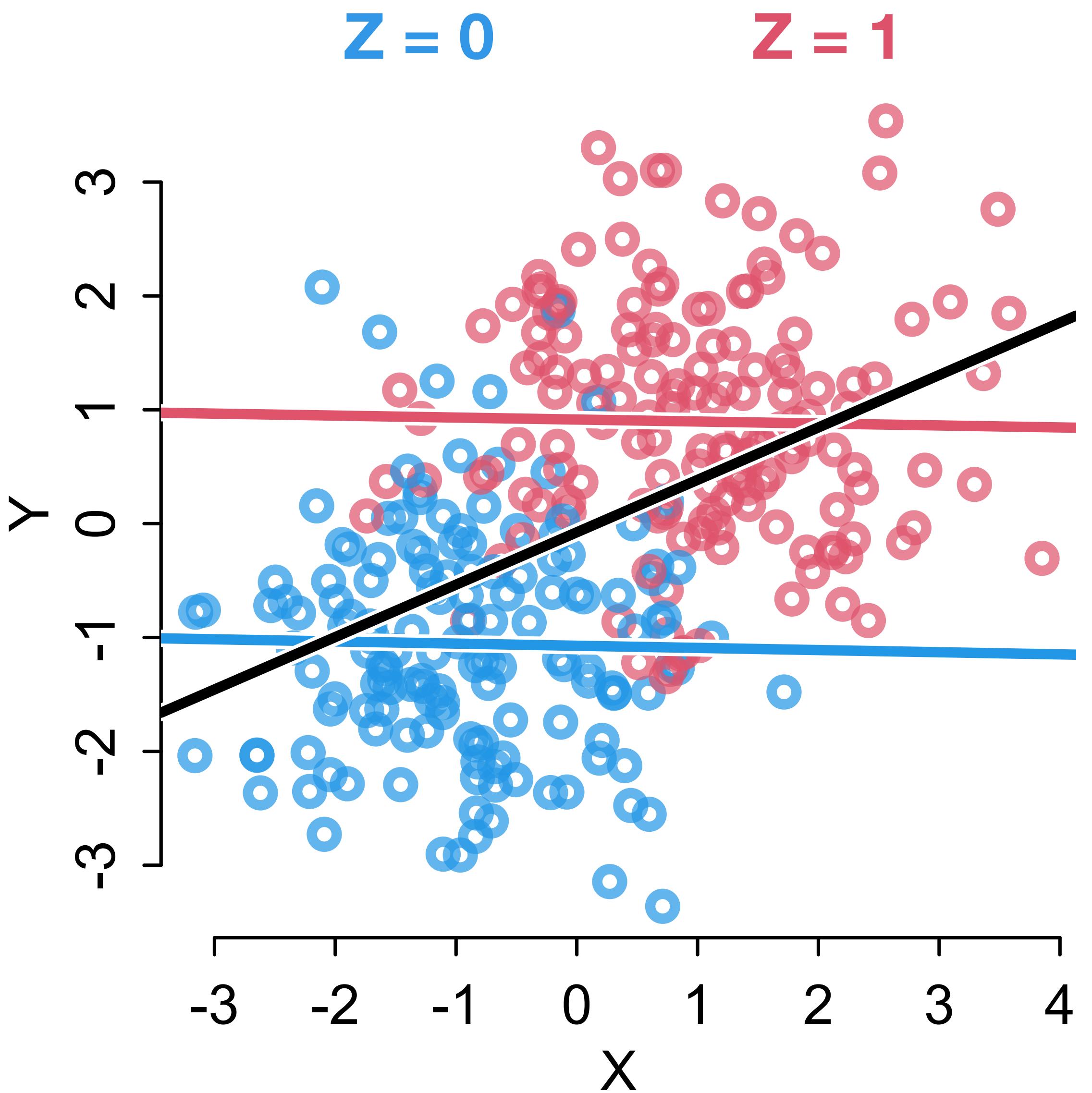
N <- 300
Z <- rbern(N)
X <- rnorm(N,2*Z-1)
Y <- rnorm(N,2*Z-1)

plot( X , Y , col=cols[Z+1] , lwd=3 )

abline(lm(Y[Z==1]~X[Z==1]),col=2,lwd=3)

abline(lm(Y[Z==0]~X[Z==0]),col=4,lwd=3)

abline(lm(Y~X),lwd=3)
```



Fork Example

Why do regions of the USA with
higher rates of marriage
also have
higher rates of divorce?

$$M \xrightarrow{?} D$$

```
library(rethinking)  
data(WaffleDivorce)
```



Marrying the Owl

(1) Estimand: Causal effect
of **marriage** rate on **divorce**
rate

$$M \xrightarrow{?} D$$

(2) Scientific model

(3) Statistical model

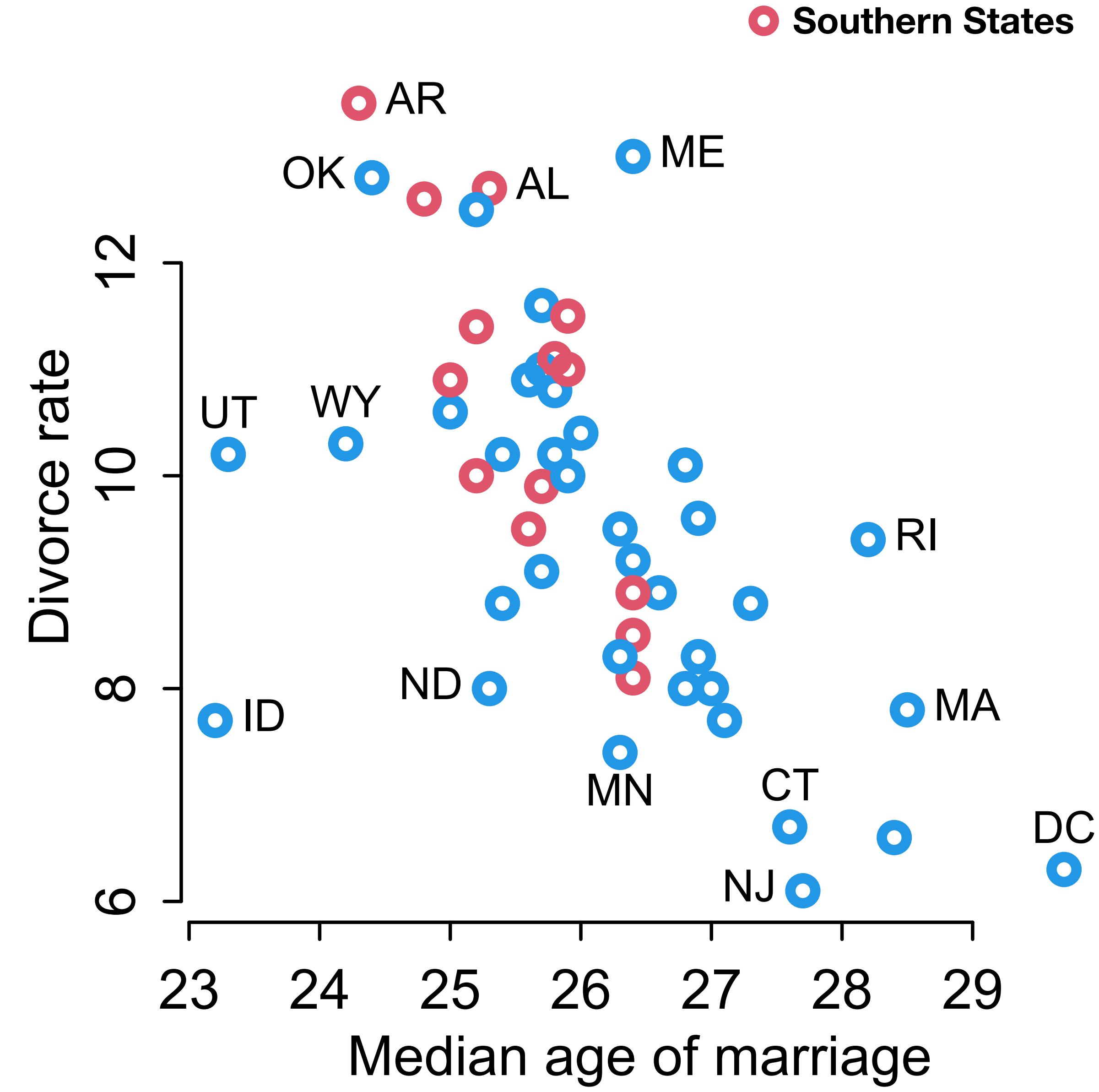
(4) Analyze

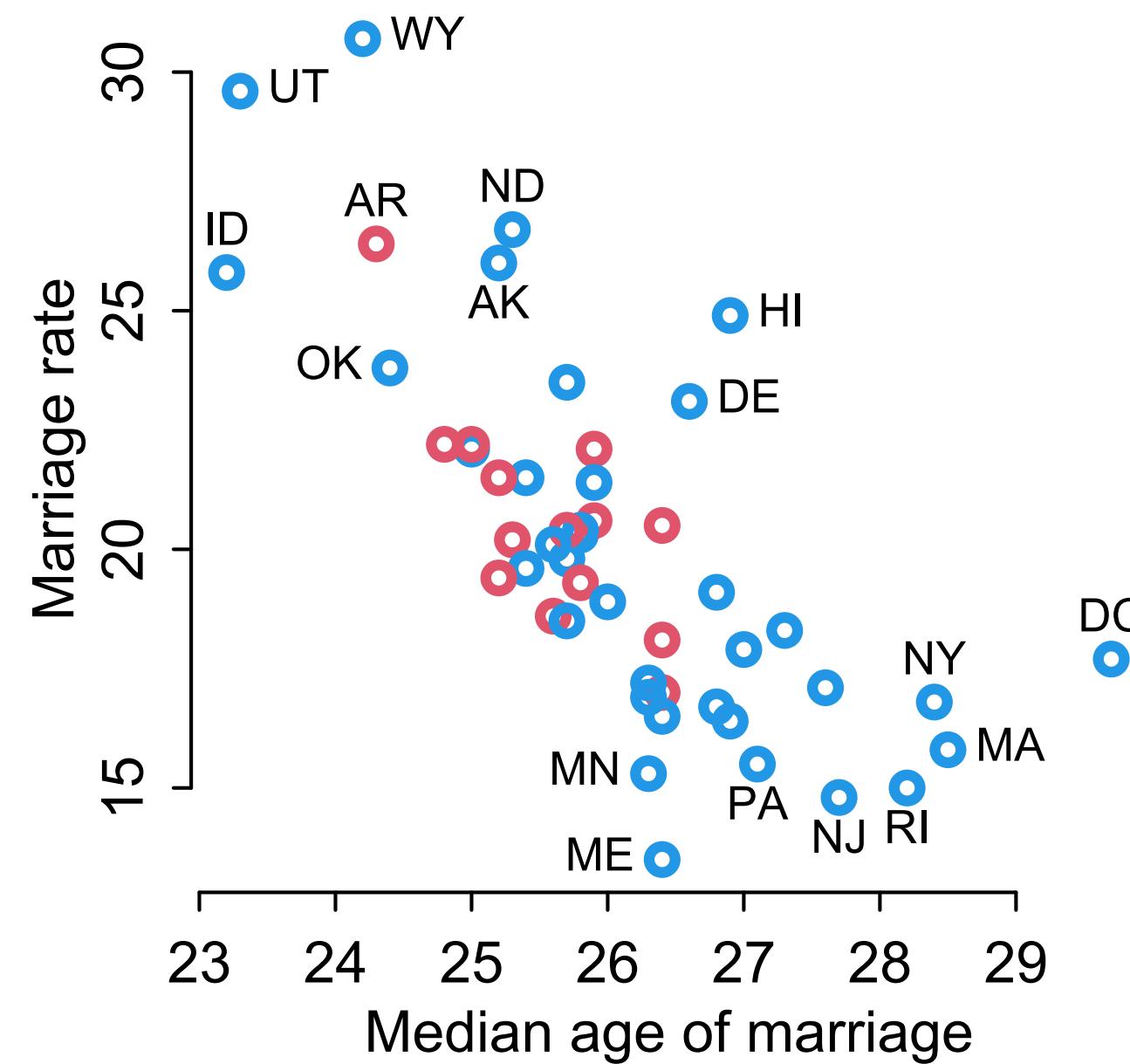
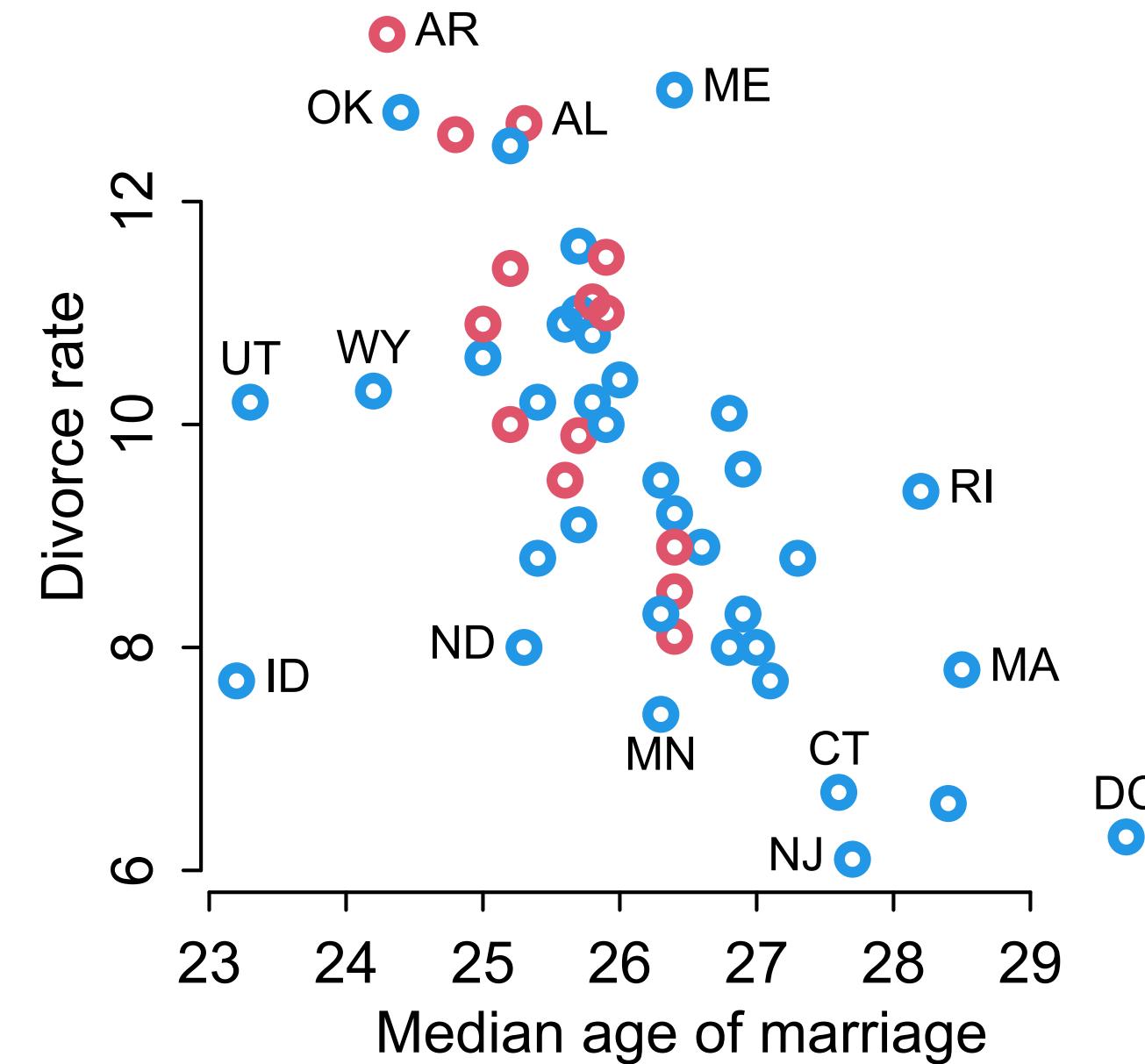
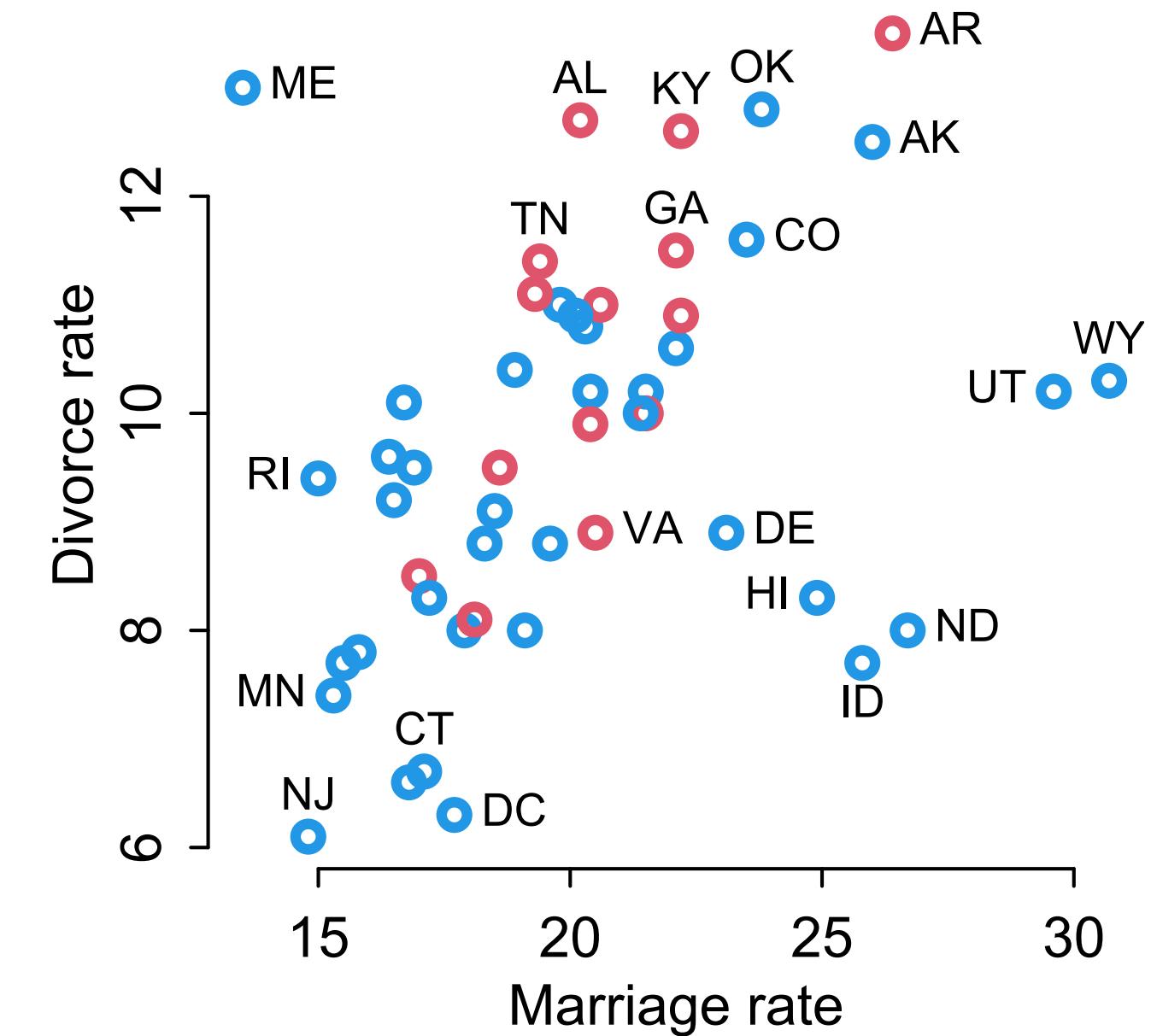
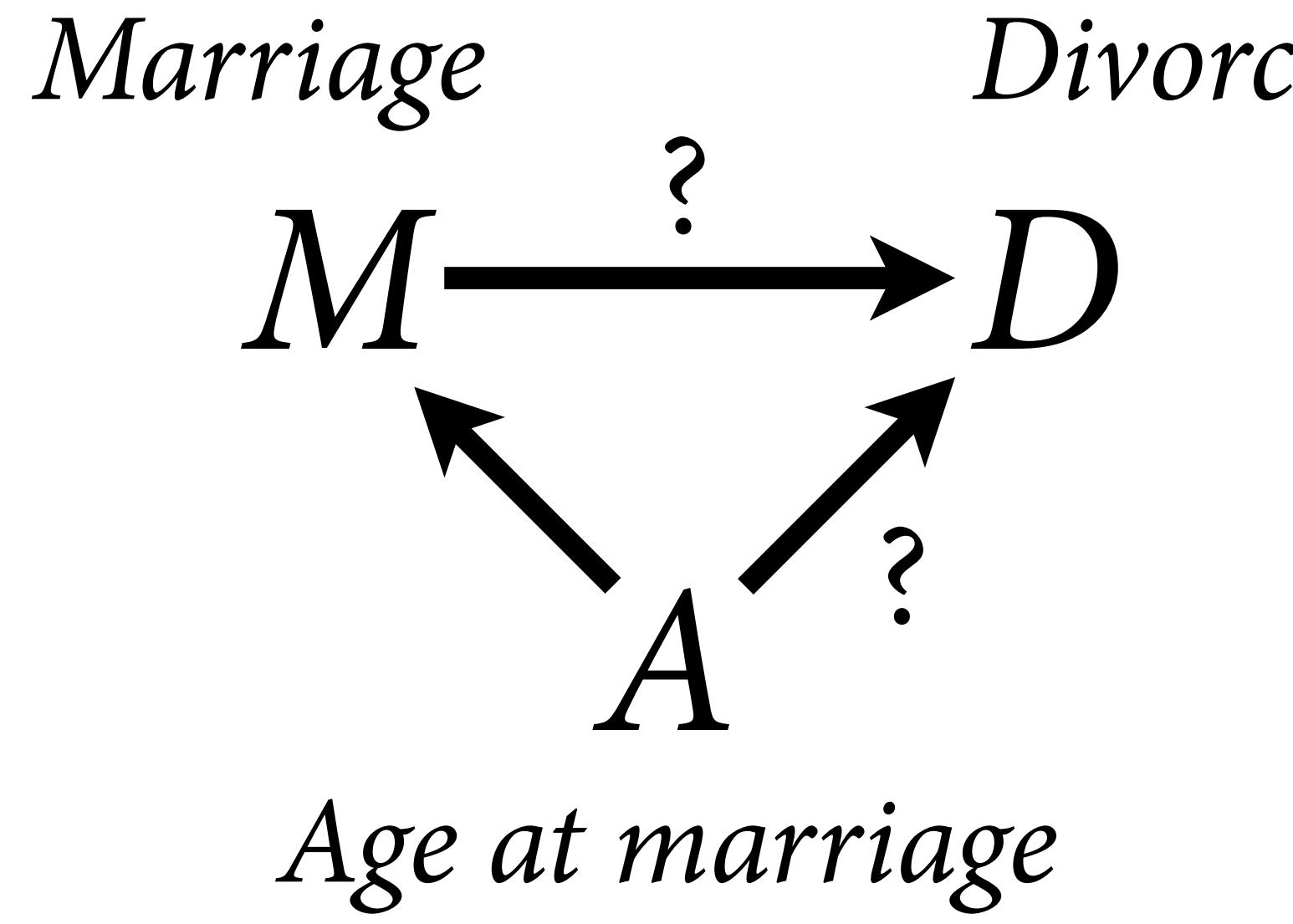


Divorce

A \nearrow ?

Age at marriage





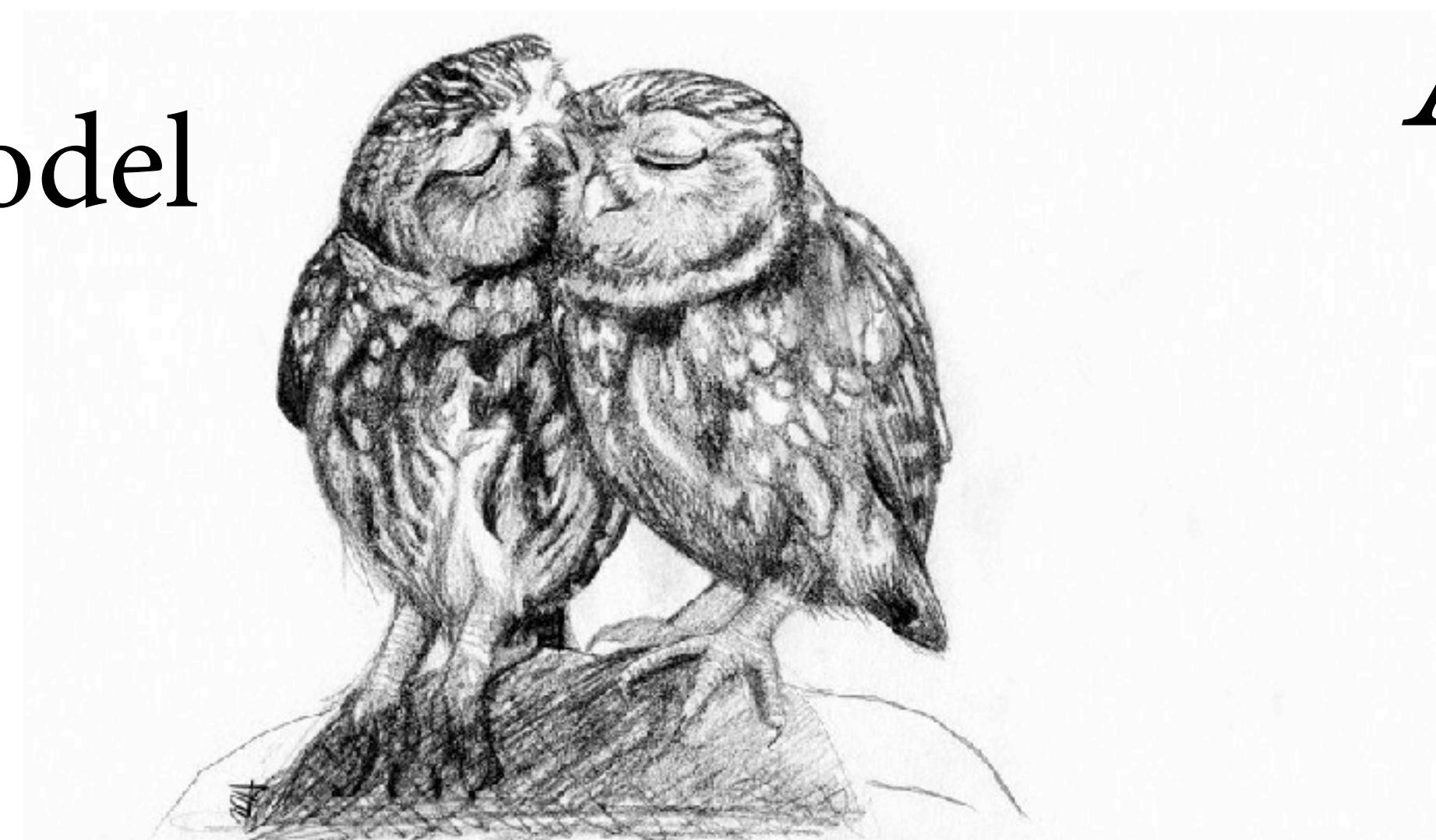
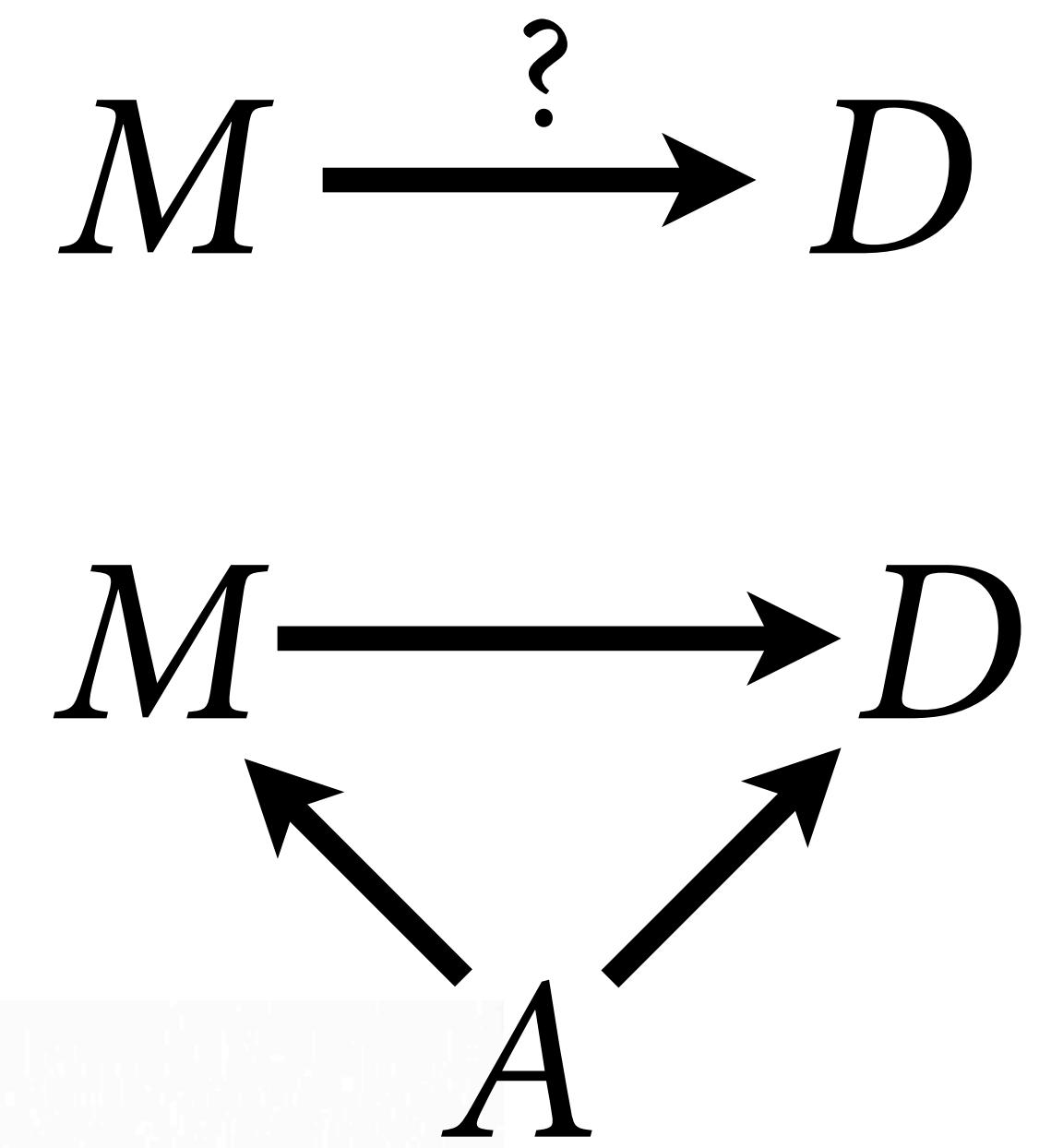
Marrying the Owl

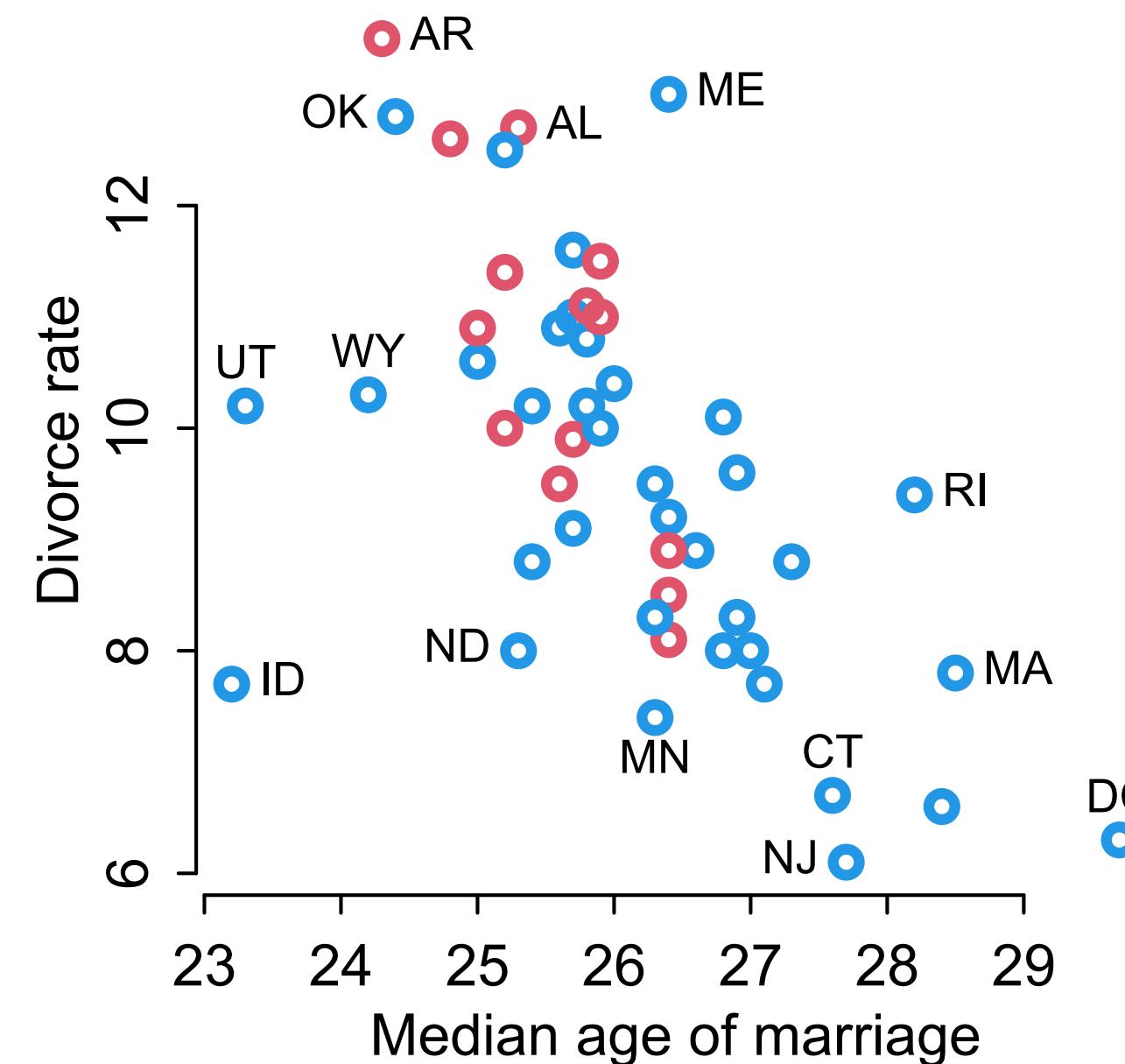
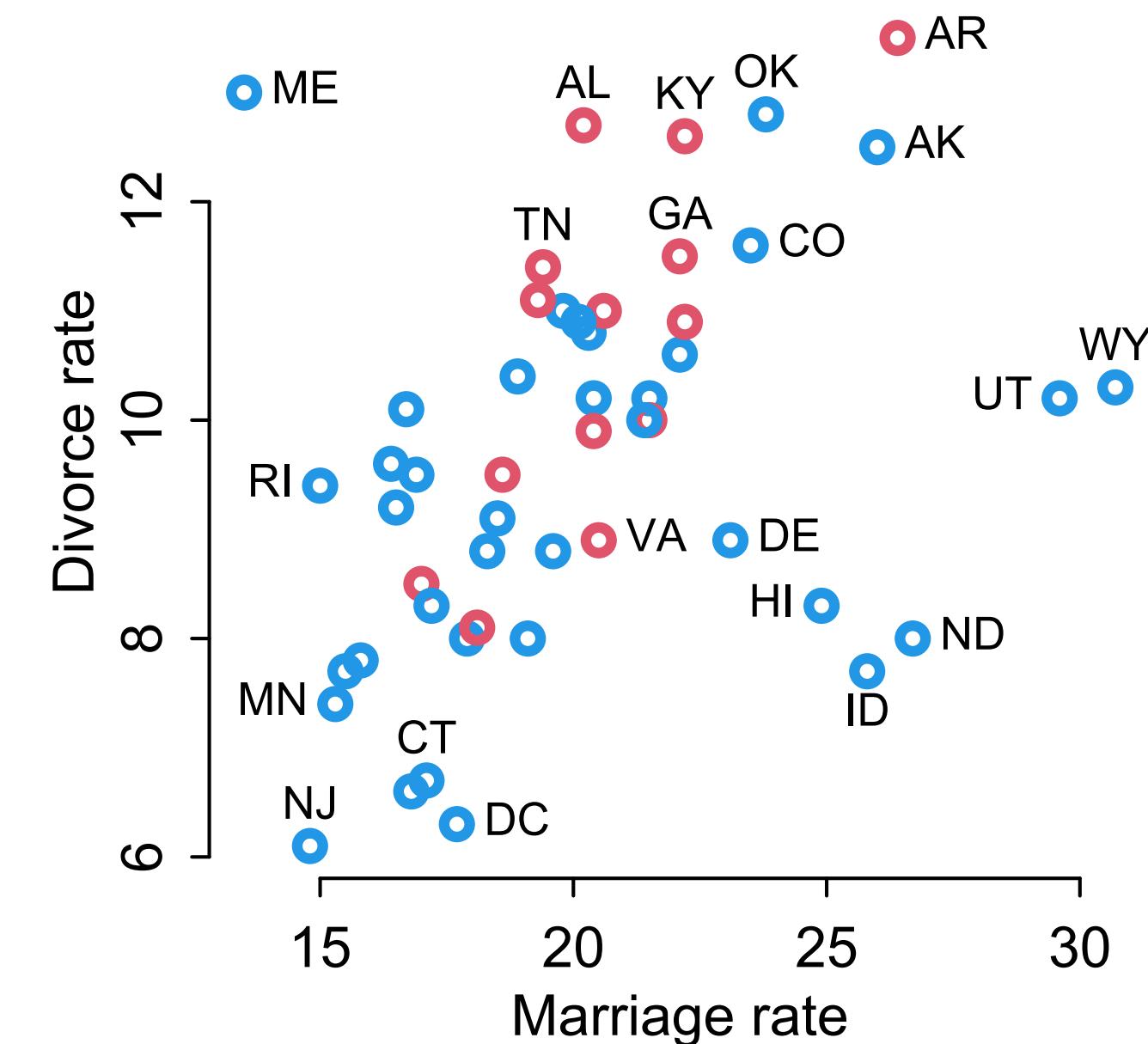
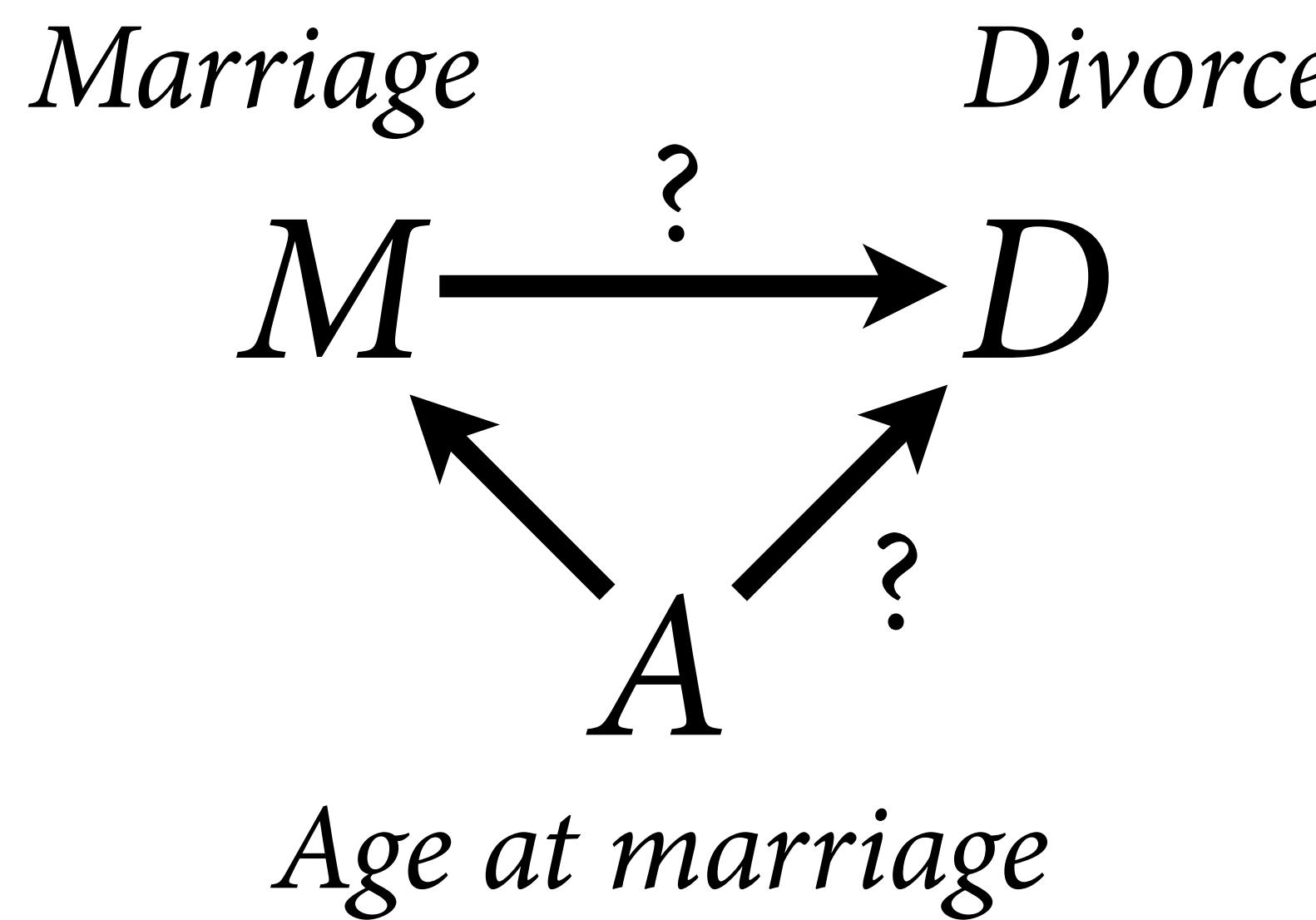
(1) Estimand: Causal effect
of **marriage** rate on **divorce**
rate

(2) Scientific model

(3) Statistical model

(4) Analyze

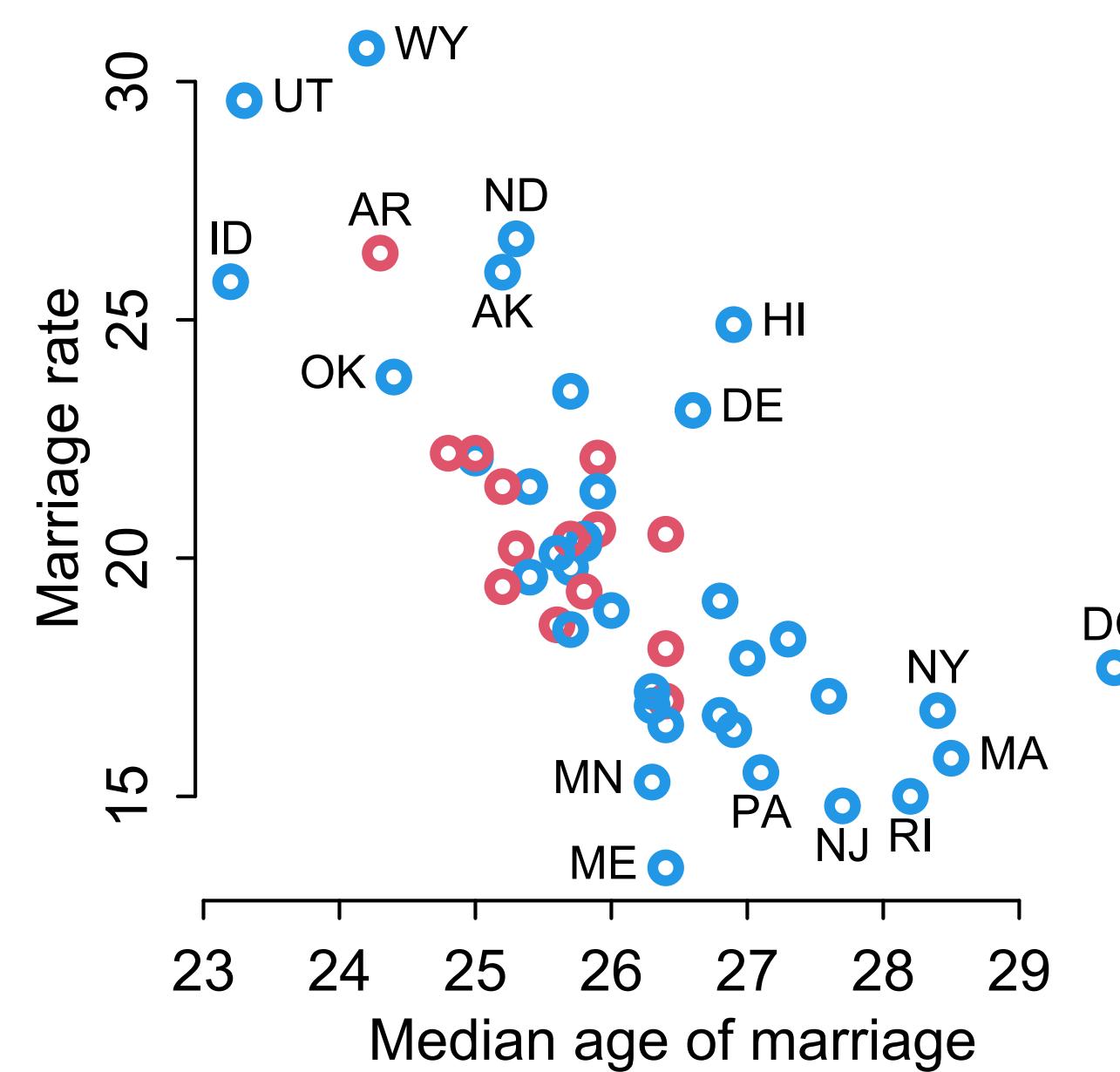




Fork: $M <- A -> D$

To estimate **causal effect** of M , need to break the fork

Break the fork by **stratifying** by A



What does it mean to stratify by a continuous variable?

It depends

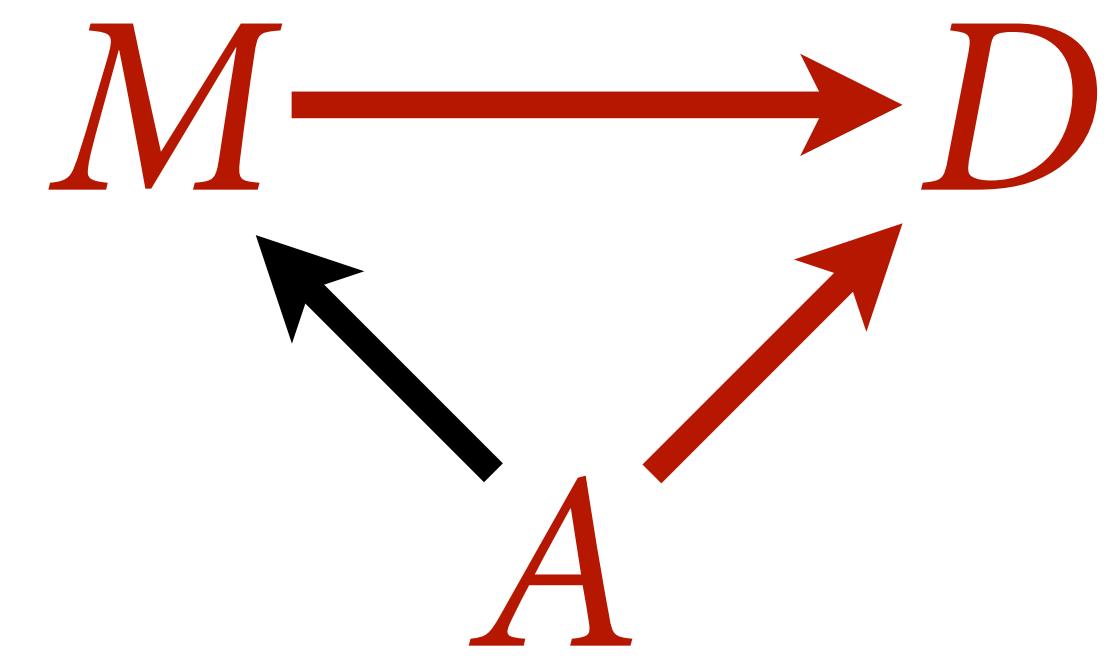
How does A influence D ?

What is $D = f(A, M)$?

In a linear regression:

$$D_i \sim \text{Normal}(\mu_i, \sigma)$$

$$\mu_i = \alpha + \beta_M M_i + \beta_A A_i$$



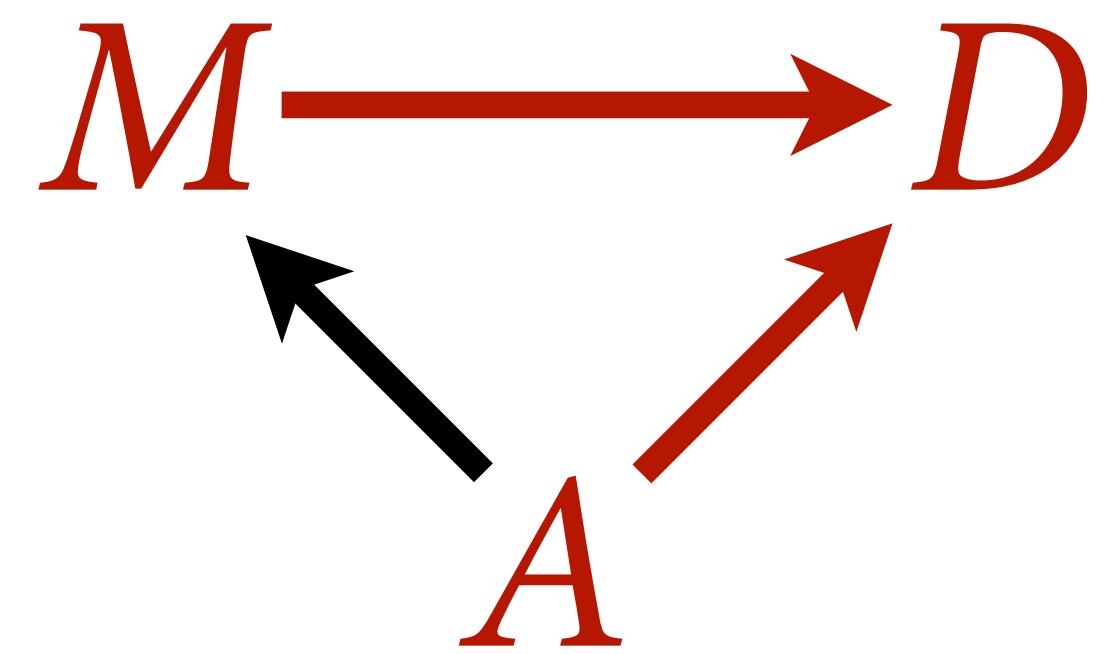
What does it mean to stratify by a continuous variable?

$$\mu_i = \alpha + \beta_M M_i + \beta_A A_i$$

Every value of A produces of different relationship between D and M :

$$\mu_i = (\alpha + \beta_A A_i) + \beta_M M_i$$

intercept



Statistical Fork

To stratify by A (age at marriage),
include as term in linear model

$$D_i \sim \text{Normal}(\mu_i, \sigma)$$

$$\mu_i = \alpha + \beta_M M_i + \beta_A A_i$$

$$\alpha \sim \text{Normal}(?, ?)$$

$$\beta_M \sim \text{Normal}(?, ?)$$

$$\beta_A \sim \text{Normal}(?, ?)$$

$$\sigma \sim \text{Exponential}(?)$$

*We are going to
standardize the data*

Standardizing the Owl

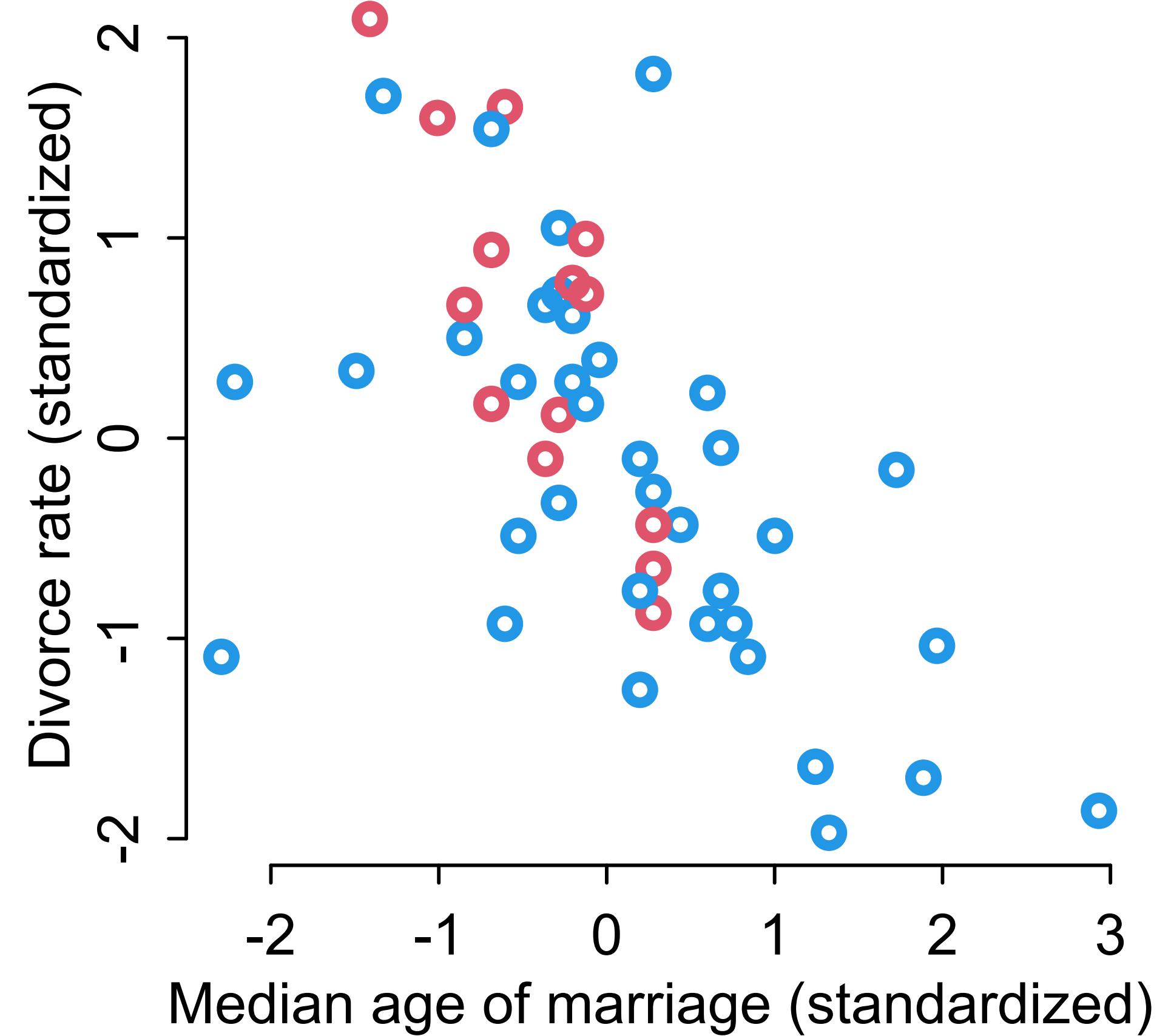


Often convenient to **standardize** variables in linear regression

Standardize: Subtract mean and divide by standard deviation

Computation works better

Easy to choose sensible priors



Prior predictive simulation

Some default priors

$$D_i \sim \text{Normal}(\mu_i, \sigma)$$

$$\mu_i = \alpha + \beta_M M_i + \beta_A A_i$$

$$\alpha \sim \text{Normal}(0,10)$$

$$\beta_M \sim \text{Normal}(0,10)$$

$$\beta_A \sim \text{Normal}(0,10)$$

$$\sigma \sim \text{Exponential}(1)$$

```
# prior predictive simulation
n <- 20
a <- rnorm(n,0,10)
bM <- rnorm(n,0,10)
bA <- rnorm(n,0,10)
plot( NULL , xlim=c(-2,2) , ylim=c(-2,2) ,
xlab="Median age of marriage (standardized)" ,
ylab="Divorce rate (standardized)" )
Aseq <- seq(from=-3,to=3,len=30)
for ( i in 1:n ) {
  mu <- a[i] + bA[i]*Aseq
  lines( Aseq , mu , lwd=2 , col=2 )
}
```

Prior predictive simulation

Some default priors

$$D_i \sim \text{Normal}(\mu_i, \sigma)$$

$$\mu_i = \alpha + \beta_M M_i + \beta_A A_i$$

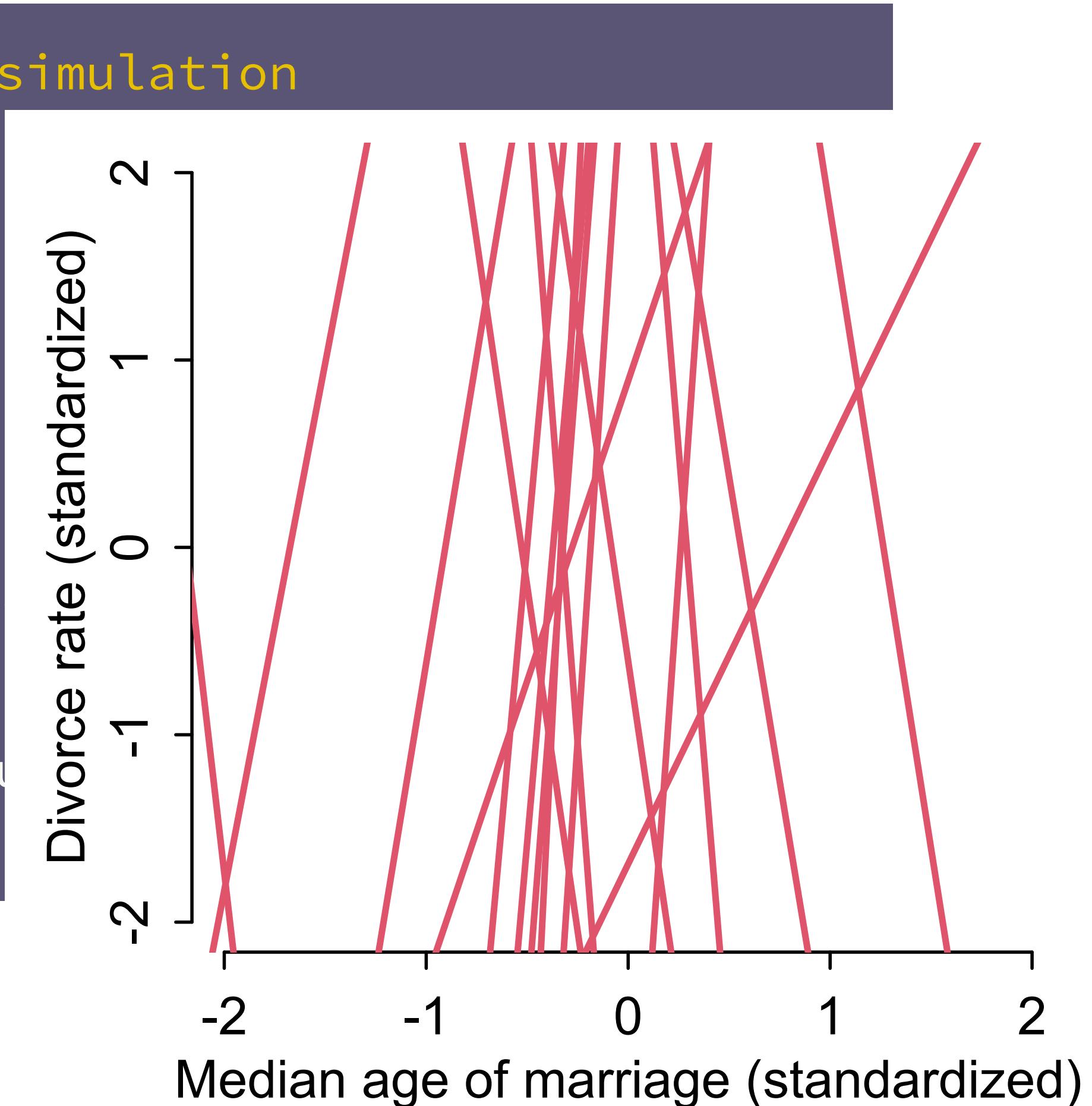
$$\alpha \sim \text{Normal}(0, 10)$$

$$\beta_M \sim \text{Normal}(0, 10)$$

$$\beta_A \sim \text{Normal}(0, 10)$$

$$\sigma \sim \text{Exponential}(1)$$

```
# prior predictive simulation
n <- 20
a <- rnorm(n, 0, 10)
bM <- rnorm(n, 0, 10)
bA <- rnorm(n, 0, 10)
plot( NULL , xlim=c,
      xlab="Median age of marriage (standardized)",
      ylab="Divorce rate (standardized)")
Aseq <- seq(from=-3, to=3, length=n)
for ( i in 1:n ) {
  mu <- a[i] + bA * Aseq[i] + bM * median(Aseq)
  lines( Aseq , mu )
}
```



Prior predictive simulation

Better priors

$$D_i \sim \text{Normal}(\mu_i, \sigma)$$

$$\mu_i = \alpha + \beta_M M_i + \beta_A A_i$$

$$\alpha \sim \text{Normal}(0, 0.2)$$

$$\beta_M \sim \text{Normal}(0, 0.5)$$

$$\beta_A \sim \text{Normal}(0, 0.5)$$

$$\sigma \sim \text{Exponential}(1)$$

```
# better priors
n <- 20
a <- rnorm(n, 0, 0.2)
bM <- rnorm(n, 0, 0.5)
bA <- rnorm(n, 0, 0.5)
plot( NULL , xlim=c(-2,2) , ylim=c(-2,2) ,
xlab="Median age of marriage (standardized)" ,
ylab="Divorce rate (standardized)" )
Aseq <- seq(from=-3,to=3,len=30)
for ( i in 1:n ) {
  mu <- a[i] + bA[i]*Aseq
  lines( Aseq , mu , lwd=2 , col=2 )
}
```

Prior predictive simulation

Better priors

$$D_i \sim \text{Normal}(\mu_i, \sigma)$$

$$\mu_i = \alpha + \beta_M M_i + \beta_A A_i$$

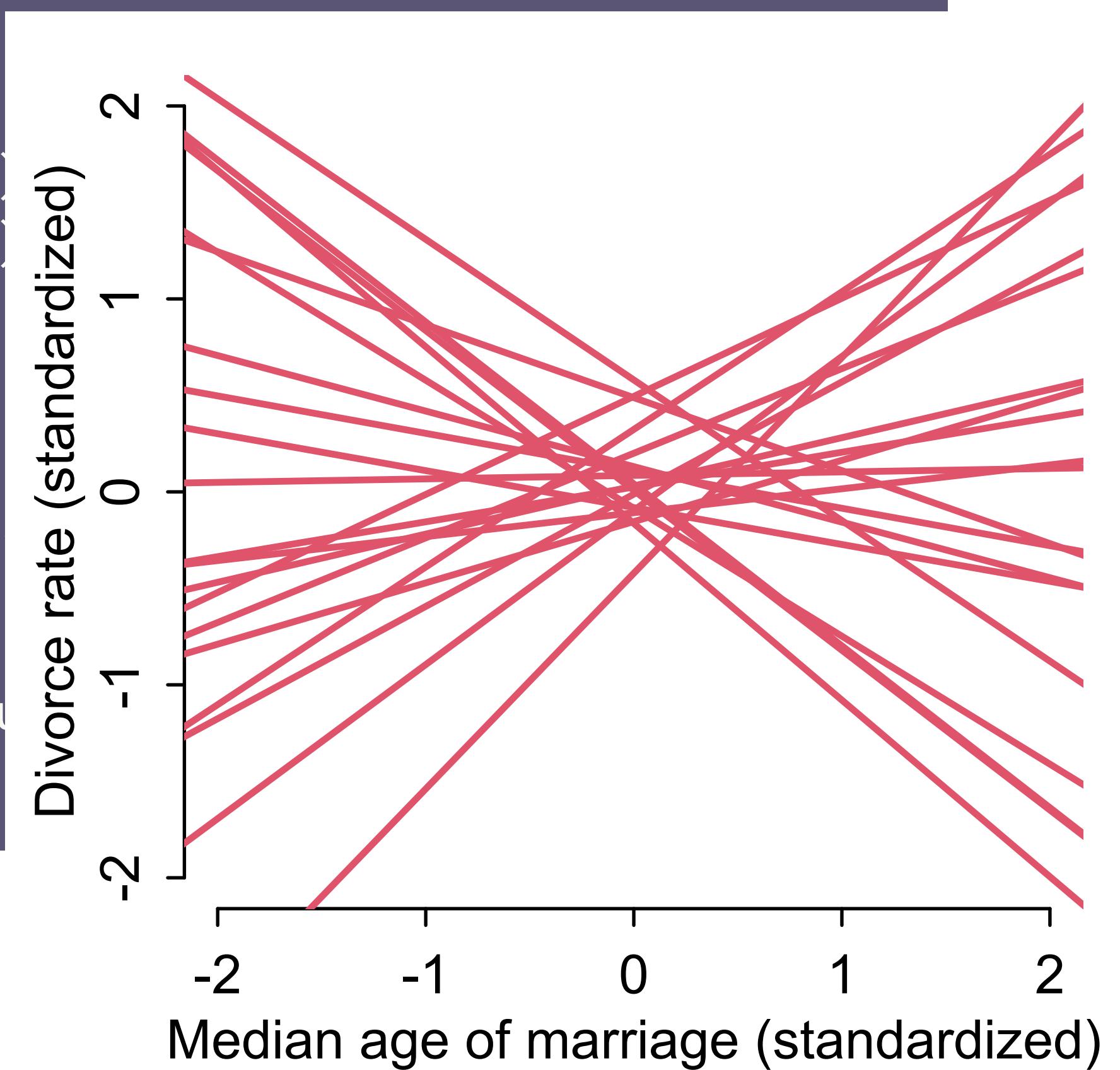
$$\alpha \sim \text{Normal}(0, 0.2)$$

$$\beta_M \sim \text{Normal}(0, 0.5)$$

$$\beta_A \sim \text{Normal}(0, 0.5)$$

$$\sigma \sim \text{Exponential}(1)$$

```
# better priors
n <- 20
a <- rnorm(n, 0, 0.2)
bM <- rnorm(n, 0, 0.5)
bA <- rnorm(n, 0, 0.5)
plot( NULL , xlim=c(-3,3),
      xlab="Median age of marriage (standardized)",
      ylab="Divorce rate")
Aseq <- seq(from=-3, to=3, by=0.1)
for ( i in 1:n ) {
  mu <- a[i] + bA * Aseq + bM * i
  lines( Aseq , mu )
}
```

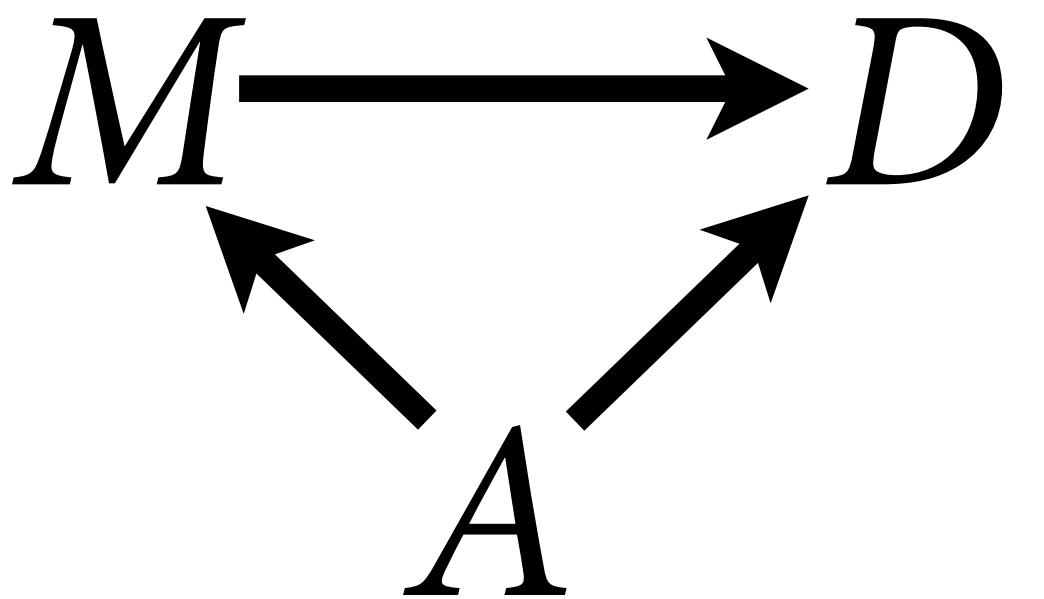


Marrying the Owl

(1) Estimand: Causal effect
of **marriage** rate on **divorce**
rate

$$M \xrightarrow{?} D$$

(2) Scientific model



(3) Statistical model

$$\mu_i = \alpha + \beta_M M_i + \beta_A A_i$$

(4) Analyze



Analyze data

```
# model
dat <- list(
  D = standardize(d$Divorce),
  M = standardize(d$Marriage),
  A = standardize(d$MedianAgeMarriage)
)
m_DMA <- quap(
  alist(
    D ~ dnorm(mu,sigma),
    mu <- a + bM*M + bA*A,
    a ~ dnorm(0,0.2),
    bM ~ dnorm(0,0.5),
    bA ~ dnorm(0,0.5),
    sigma ~ dexp(1)
  ) , data=dat )
```

$$D_i \sim \text{Normal}(\mu_i, \sigma)$$

$$\mu_i = \alpha + \beta_M M_i + \beta_A A_i$$

$$\alpha \sim \text{Normal}(0,0.2)$$

$$\beta_M \sim \text{Normal}(0,0.5)$$

$$\beta_A \sim \text{Normal}(0,0.5)$$

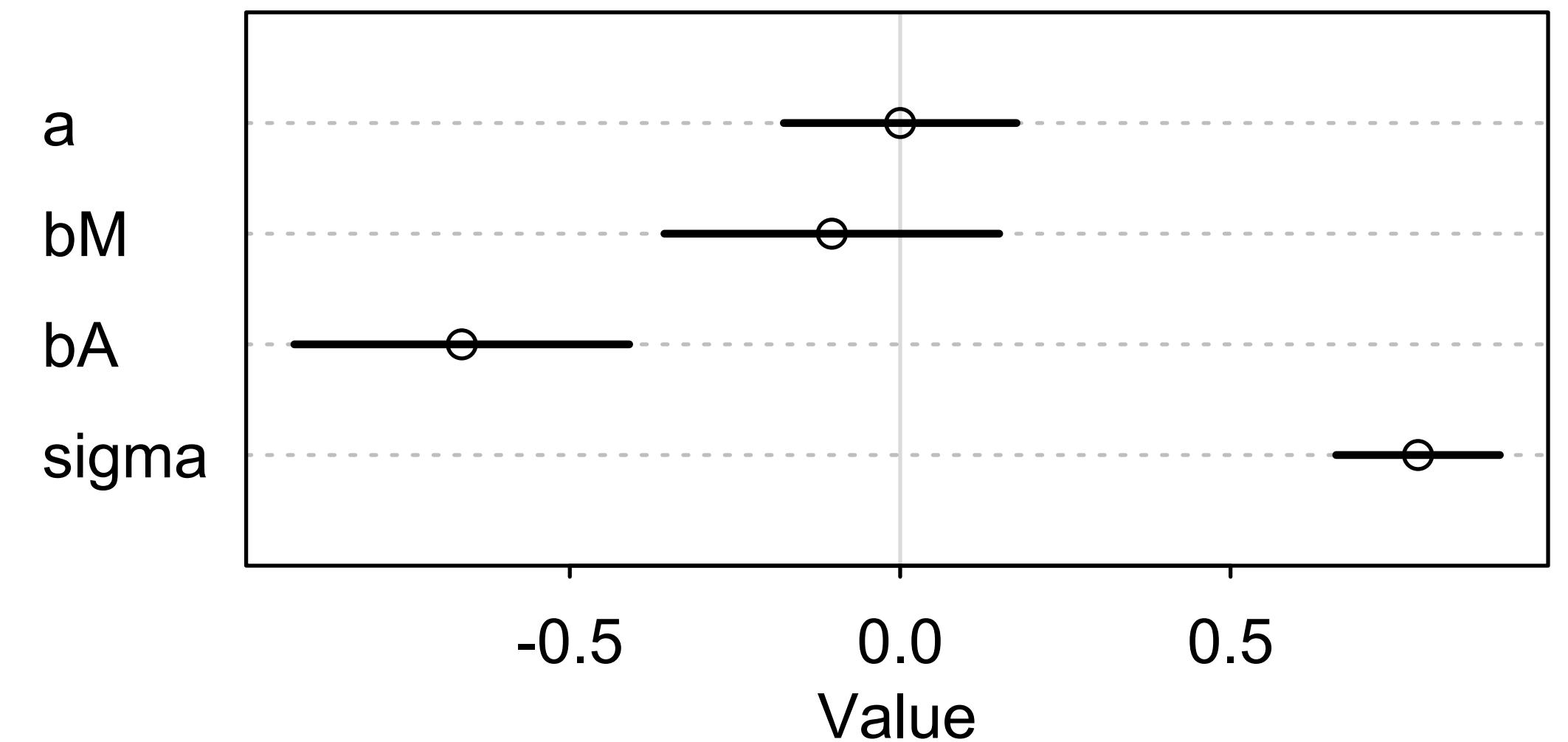
$$\sigma \sim \text{Exponential}(1)$$

Analyze data

```
# model
dat <- list(
  D = standardize(d$Divorce),
  M = standardize(d$Marriage),
  A = standardize(d$MedianAgeMarriage)
)

m_DMA <- quap(
  alist(
    D ~ dnorm(mu,sigma),
    mu <- a + bM*M + bA*A,
    a ~ dnorm(0,0.2),
    bM ~ dnorm(0,0.5),
    bA ~ dnorm(0,0.5),
    sigma ~ dexp(1)
  ), data=dat )
```

```
plot(precis(m_DMA))
```



What's the causal effect of M?

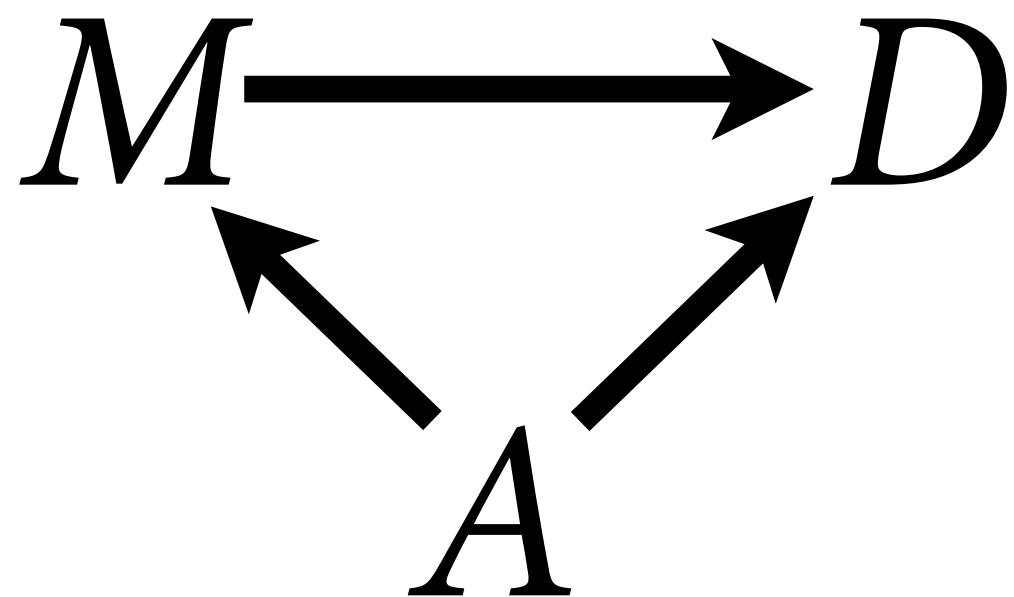
Simulating interventions

A causal effect is a manipulation of the generative model, an intervention.

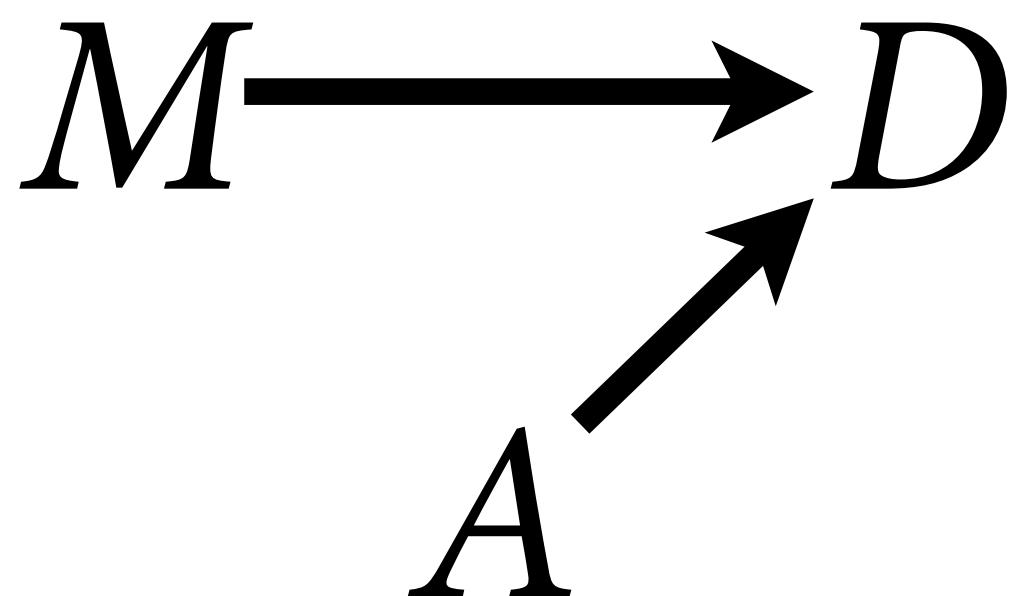
$p(D|\text{do}(M))$ means the distribution of D when we intervene (“do”) M

This implies deleting all arrows into M and simulating D

no intervention

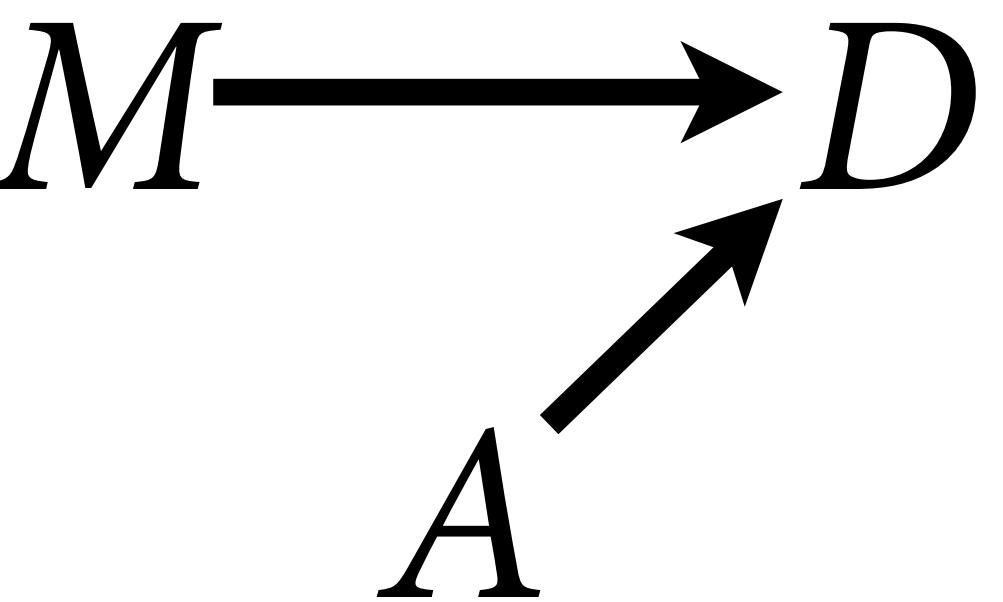


do(M)



```
post <- extract.samples( m_DMA )
# sample A from data
n <- 1e3
As <- sample(dat$A, size=n, replace=TRUE)
```

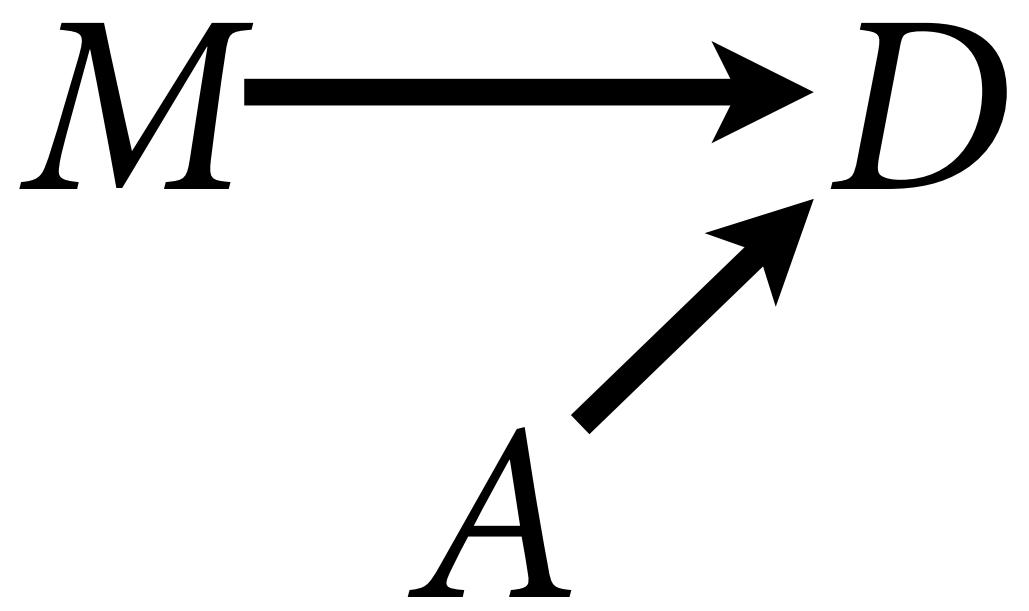
$\text{do}(M)$



```
post <- extract.samples( m_DMA )
# sample A from data
n <- 1e3
As <- sample(dat$A, size=n, replace=TRUE)

# simulate D for M=0 (sample mean)
DM0 <- with( post ,
  rnorm(n, a + bM*0 + bA*As , sigma ) )
```

$\text{do}(M)$



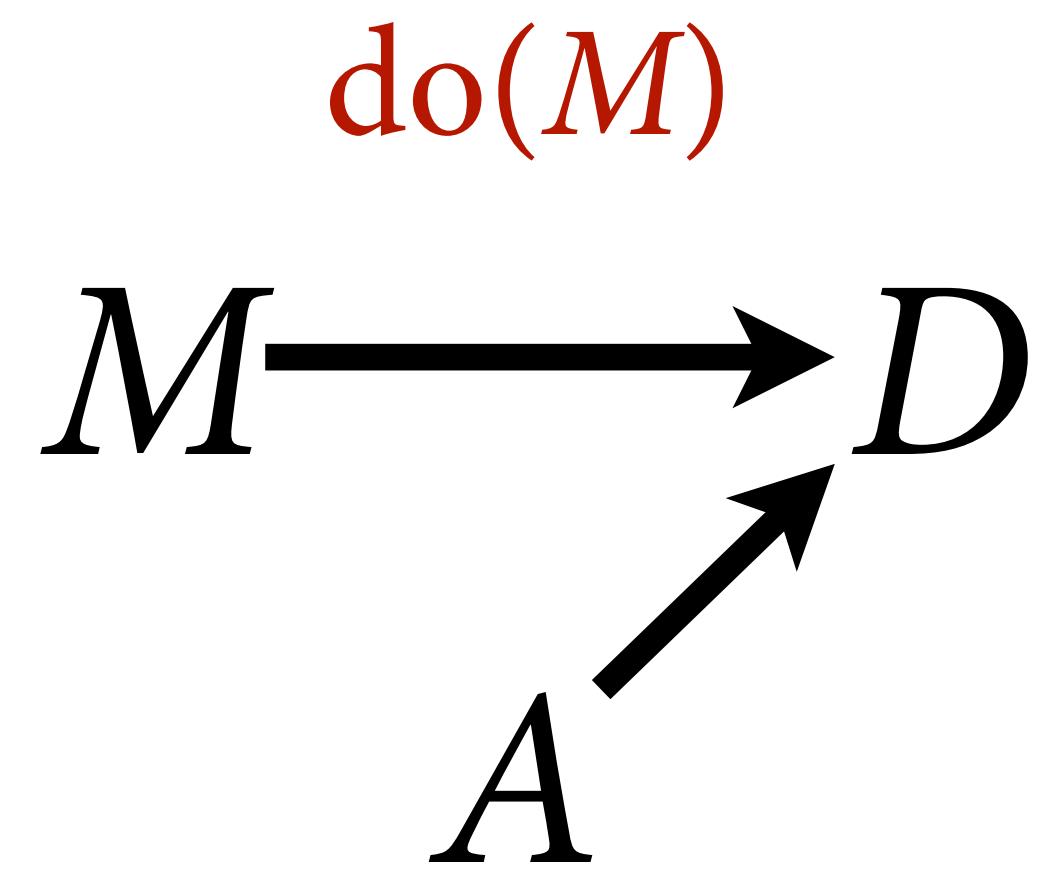
```

post <- extract.samples( m_DMA )
# sample A from data
n <- 1e3
As <- sample(dat$A, size=n, replace=TRUE)

# simulate D for M=0 (sample mean)
DM0 <- with( post ,
  rnorm(n, a + bM*0 + bA*As , sigma ) )

# simulate D for M=1 (+1 standard deviation)
# use the *same* A values
DM1 <- with( post ,
  rnorm(n, a + bM*1 + bA*As , sigma ) )

```



```

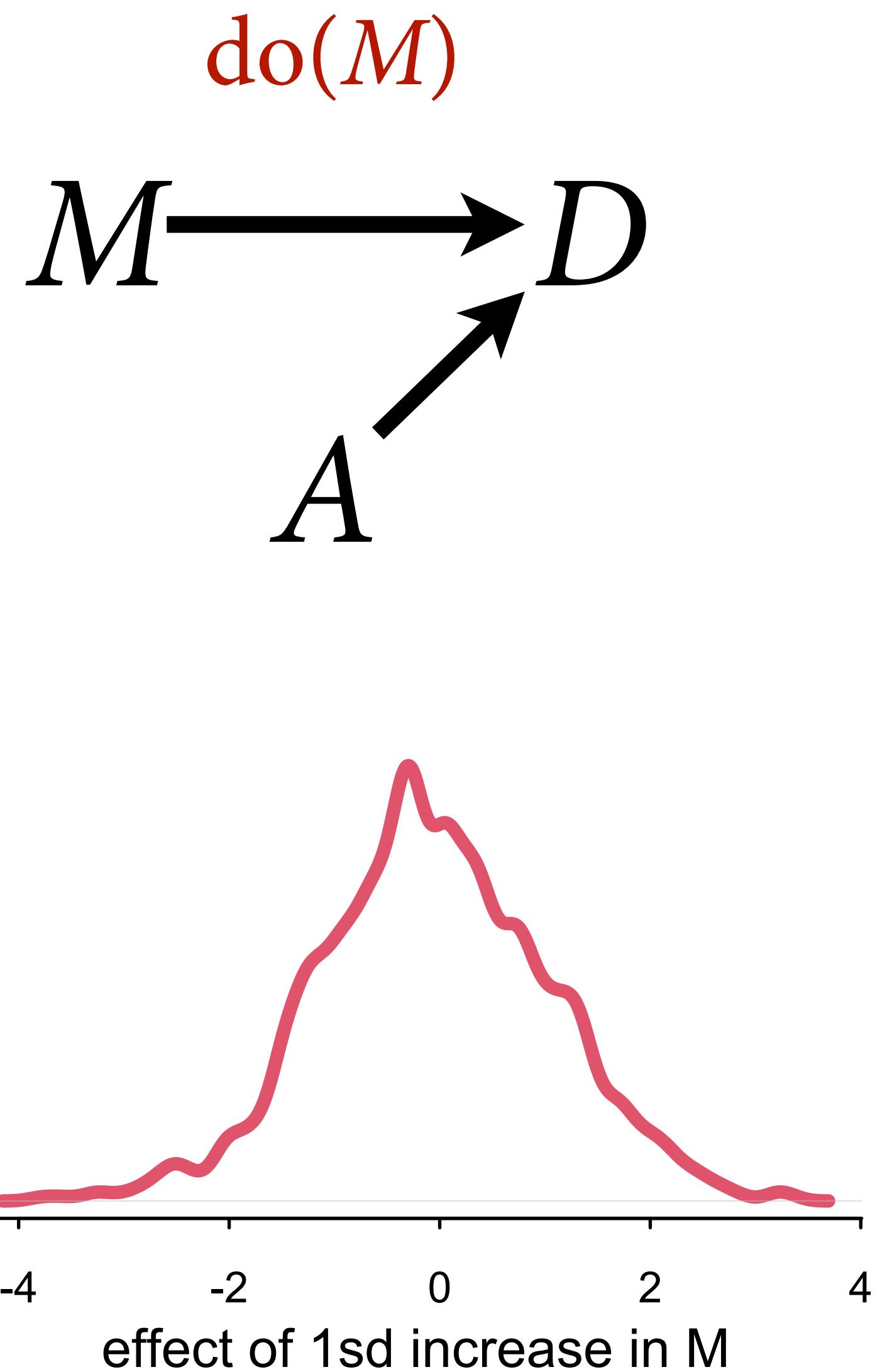
post <- extract.samples( m_DMA )
# sample A from data
n <- 1e3
As <- sample(dat$A, size=n, replace=TRUE)

# simulate D for M=0 (sample mean)
DM0 <- with( post ,
  rnorm(n, a + bM*0 + bA*As , sigma ) )

# simulate D for M=1 (+1 standard deviation)
# use the *same* A values
DM1 <- with( post ,
  rnorm(n, a + bM*1 + bA*As , sigma ) )

# contrast
M10_contrast <- DM1 - DM0
dens(M10_contrast, lwd=4, col=2, xlab="effect of 1sd
increase in M" )

```



Causal effect of A ?

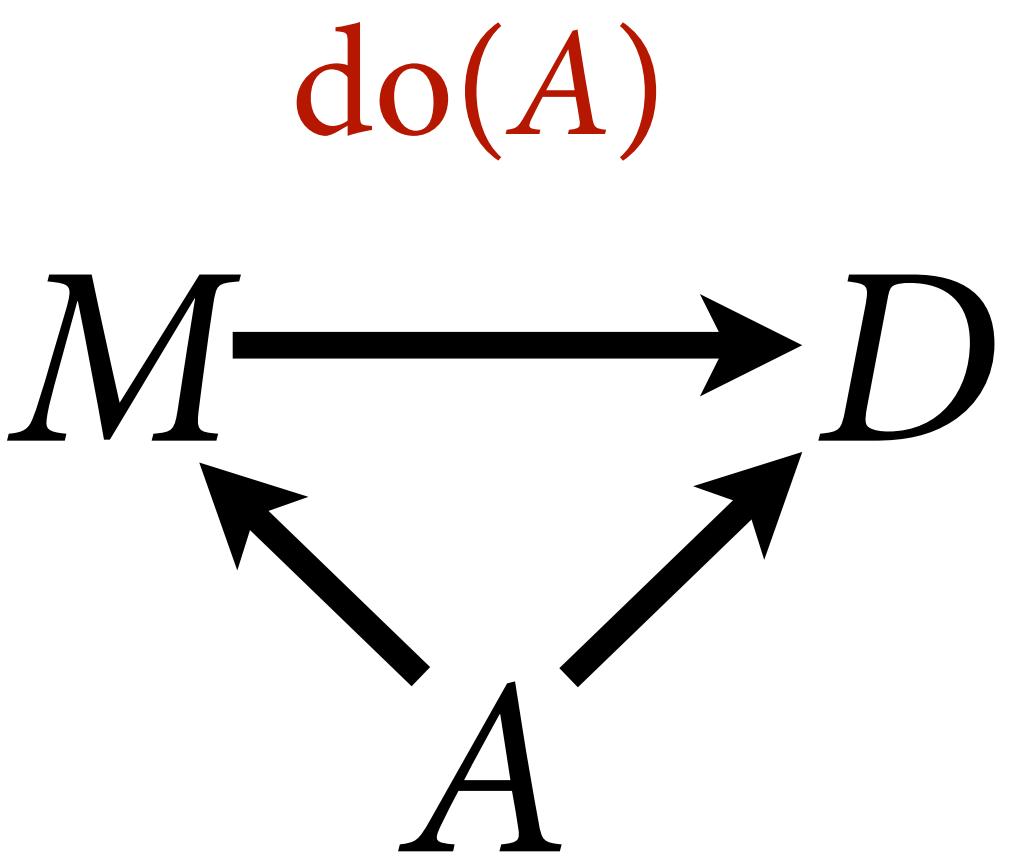
How to estimate causal effect of A ,
 $p(D|\text{do}(A))$?

No arrows to delete for intervention

Fit new model that ignores M , then
simulate any intervention you like

Why does ignoring M work?

Because $A \rightarrow M \rightarrow D$ is a “pipe”



PAUSE

Ye Olde Causal Alchemy

The Four Elemental Confounds



The Fork



The Collider



The Pipe



The Descendant



The Pipe

$$X \rightarrow Z \rightarrow Y$$

Z is a “mediator”

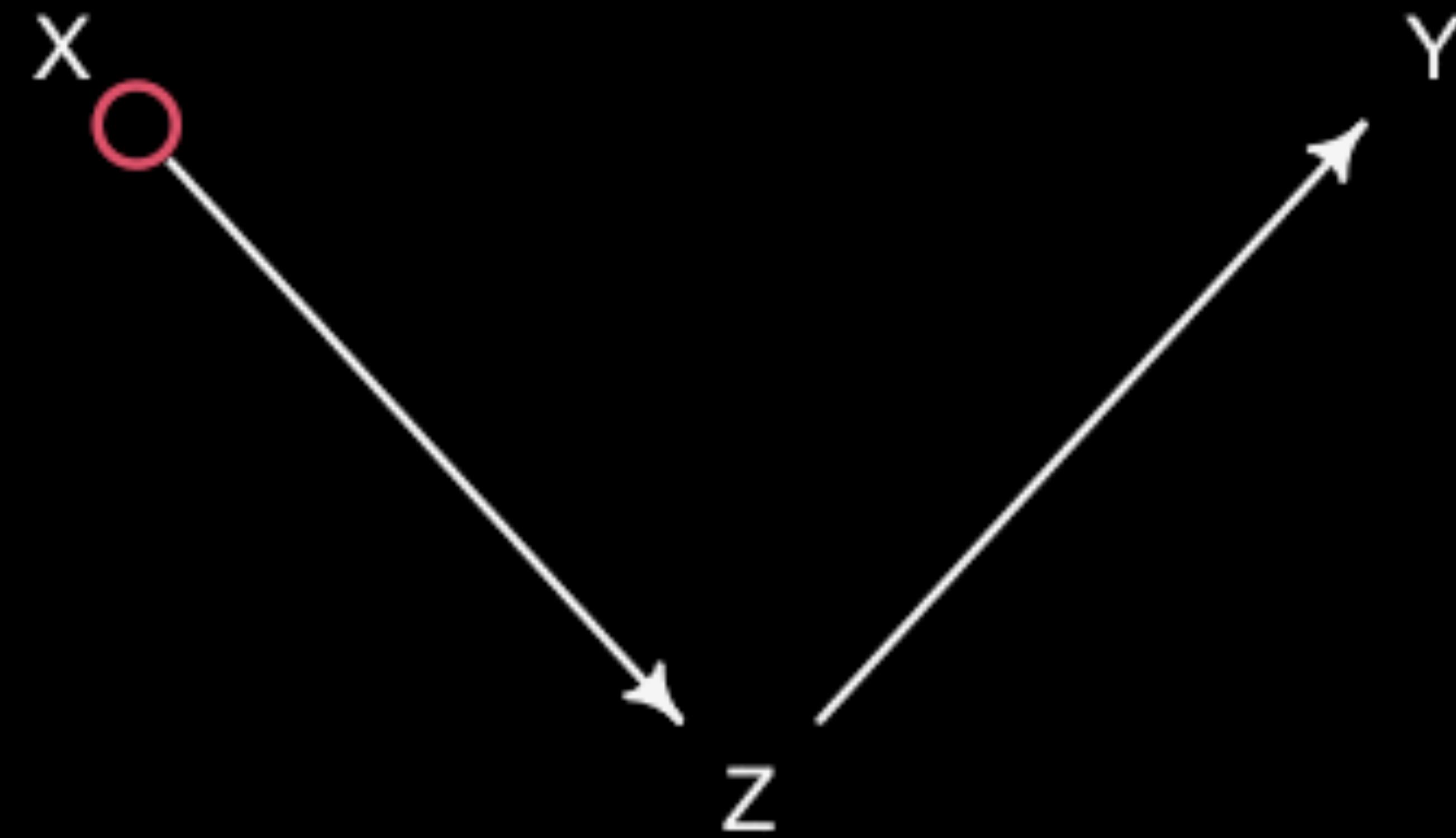
X and Y are associated

$$Y \not\perp X$$

Influence of X on Y transmitted through Z

Once stratified by Z , no association

$$Y \perp\!\!\!\perp X \mid Z$$



$X \rightarrow Z \rightarrow Y$

```
n <- 1000
X <- rbern( n , 0.5)
Z <- rbern( n , (1-X)*0.1 + X*0.9 )
Y <- rbern( n , (1-Z)*0.1 + Z*0.9 )
```

		Y
X		0 1
0	430	87
1	93	390

$Y \not\perp X$

```
> cor(X,Y)
[1] 0.64
```

$X \rightarrow Z \rightarrow Y$

```
n <- 1000
X <- rbern( n , 0.5)
Z <- rbern( n , (1-X)*0.1 + X*0.9 )
Y <- rbern( n , (1-Z)*0.1 + Z*0.9 )
```

		Y
		X
		0 1
0	430	87
1	93	390

$Y \not\perp\!\!\!\perp X$

```
> cor(X,Y)
[1] 0.64
```

		Z = 0
		Y
		X 0 1
0	422	39
1	53	5

$Z = 1$ $Y \perp\!\!\!\perp X | Z$

		Z = 1
		Y
		X 0 1
0	8	48
1	40	385

```
> cor(X[Z==0],Y[Z==0])
[1] 0.002
> cor(X[Z==1],Y[Z==1])
[1] 0.052
```

$X \rightarrow Z \rightarrow Y$

```
cols <- c(4,2)

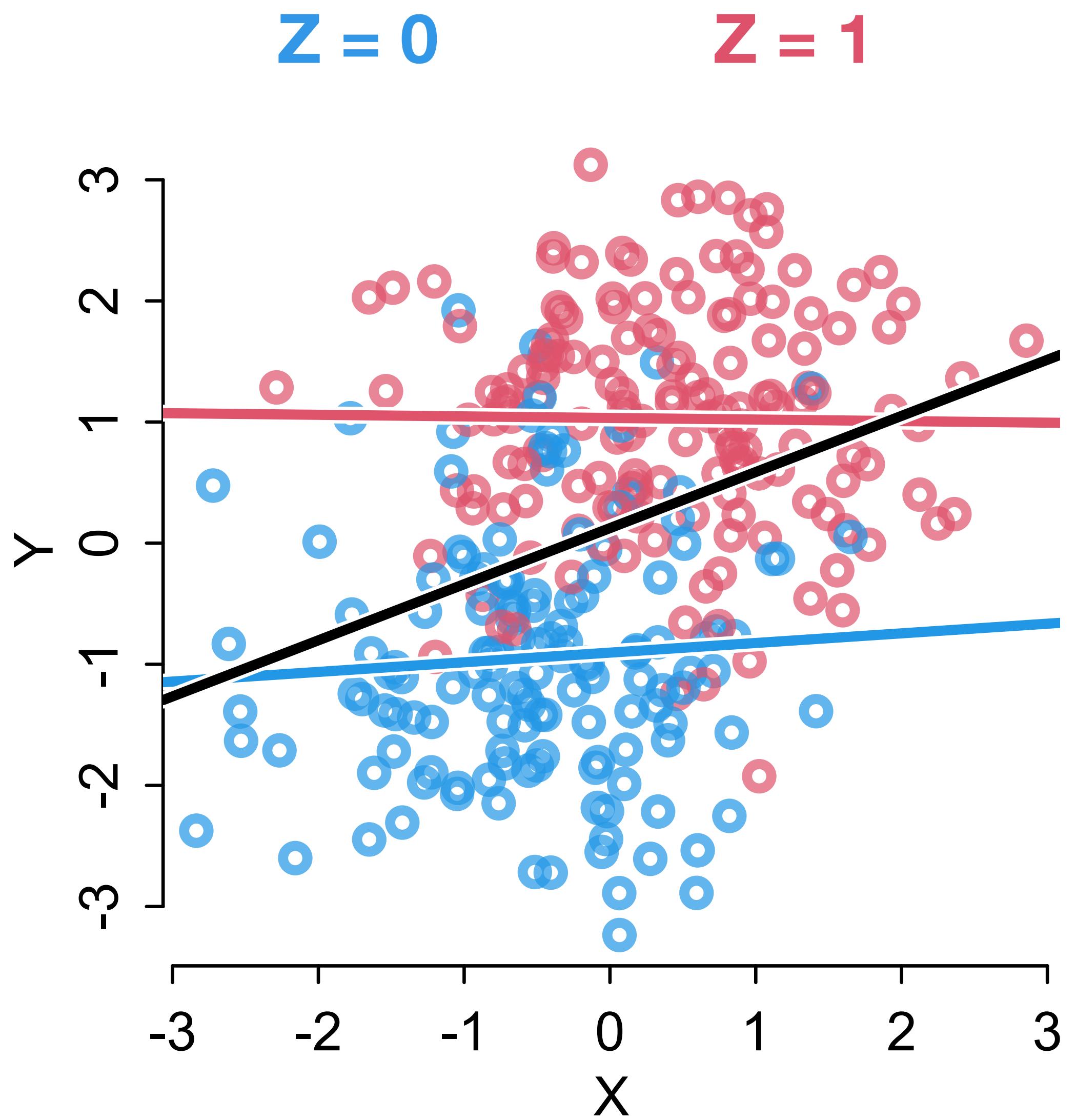
N <- 300
X <- rnorm(N)
Z <- rbern(N,inv_logit(X))
Y <- rnorm(N,(2*Z-1))

plot( X , Y , col=cols[Z+1] , lwd=3 )

abline(lm(Y[Z==1]~X[Z==1]),col=2,lwd=3)

abline(lm(Y[Z==0]~X[Z==0]),col=4,lwd=3)

abline(lm(Y~X),lwd=3)
```



Pipe example

Plant growth experiment

100 plants

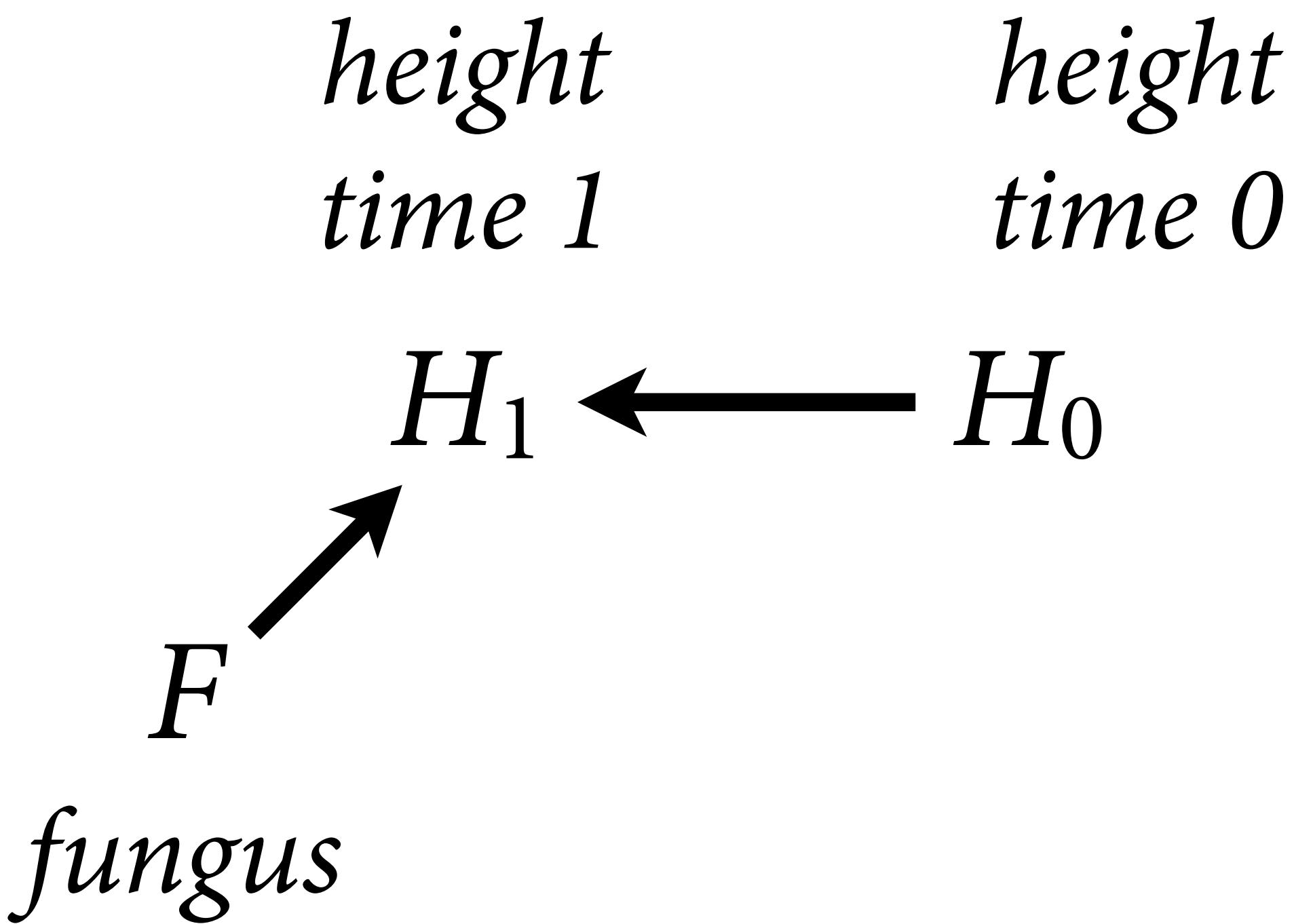
Half treated with anti-fungal

Measure growth and fungus

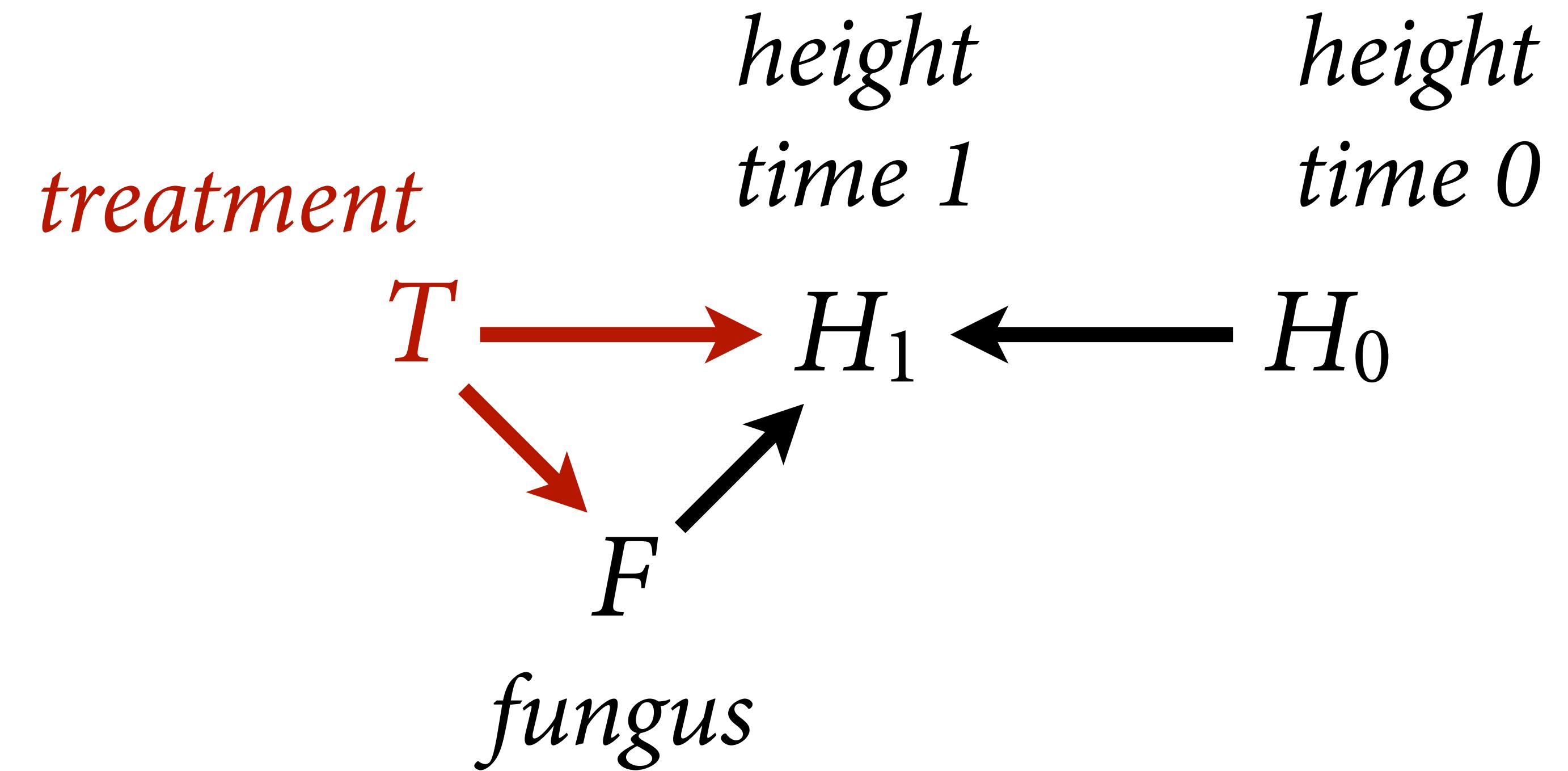
Estimand: Causal effect of treatment on plant growth



Scientific model



Scientific model



Statistical model

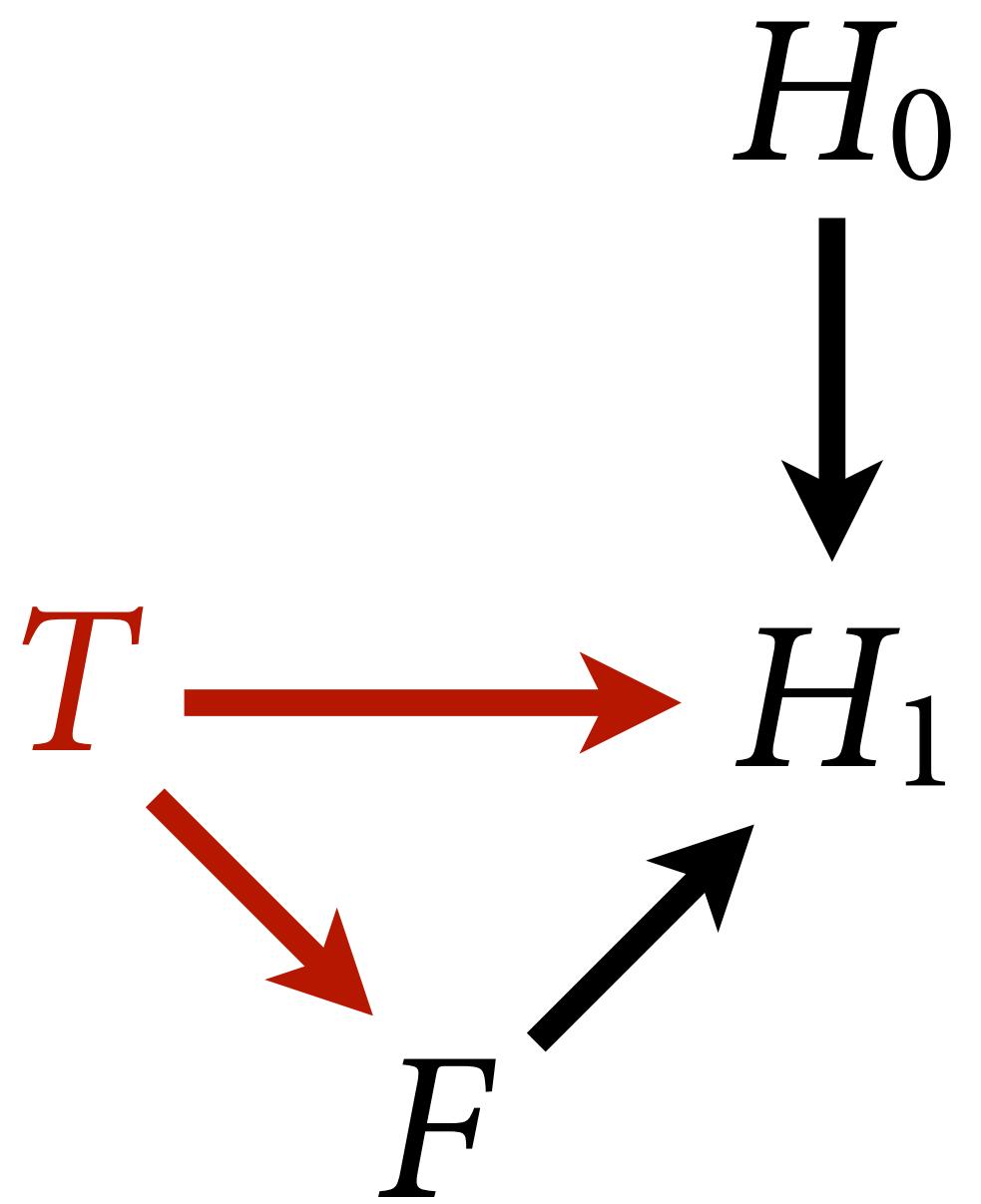
Estimand: Total causal effect of T

The path $T \rightarrow F \rightarrow H_1$ is a pipe

Should we stratify by F ?

NO — that would block the pipe

See pages 170–175 for complete example



*The treatment
must flow*

Post-treatment bias

Stratifying by (conditioning on) F induces **post-treatment bias**

Mislead that treatment doesn't work

Consequences of treatment should not usually be included in estimator

Doing experiments is no protection against bad causal inference

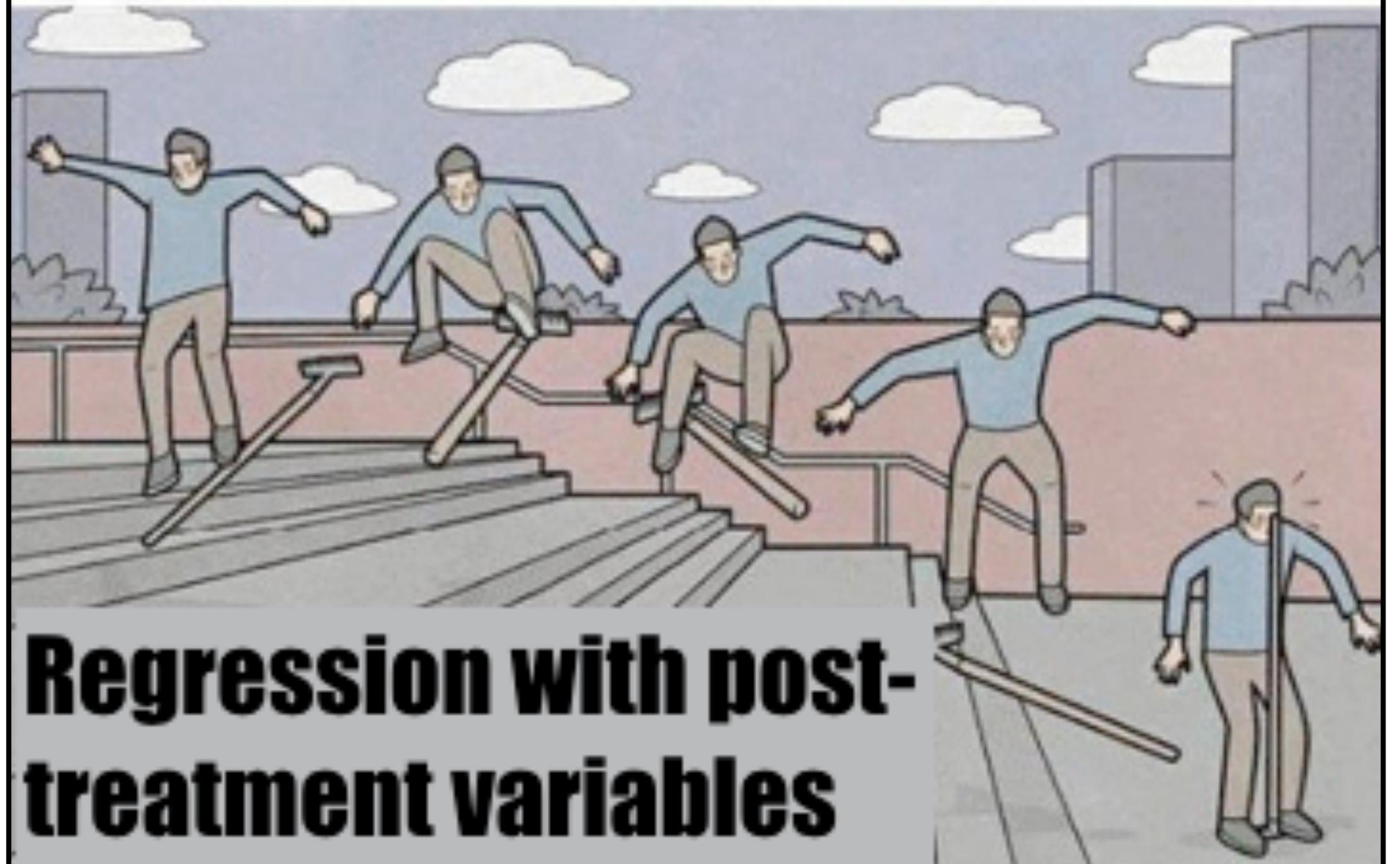
TABLE 1 Posttreatment Conditioning in Experimental Studies

Category	Prevalence
Engages in posttreatment conditioning	46.7%
Controls for/interacts with a posttreatment variable	21.3%
Drops cases based on posttreatment criteria	14.7%
Both types of posttreatment conditioning present	10.7%
No conditioning on posttreatment variables	52.0%
Insufficient information to code	1.3%

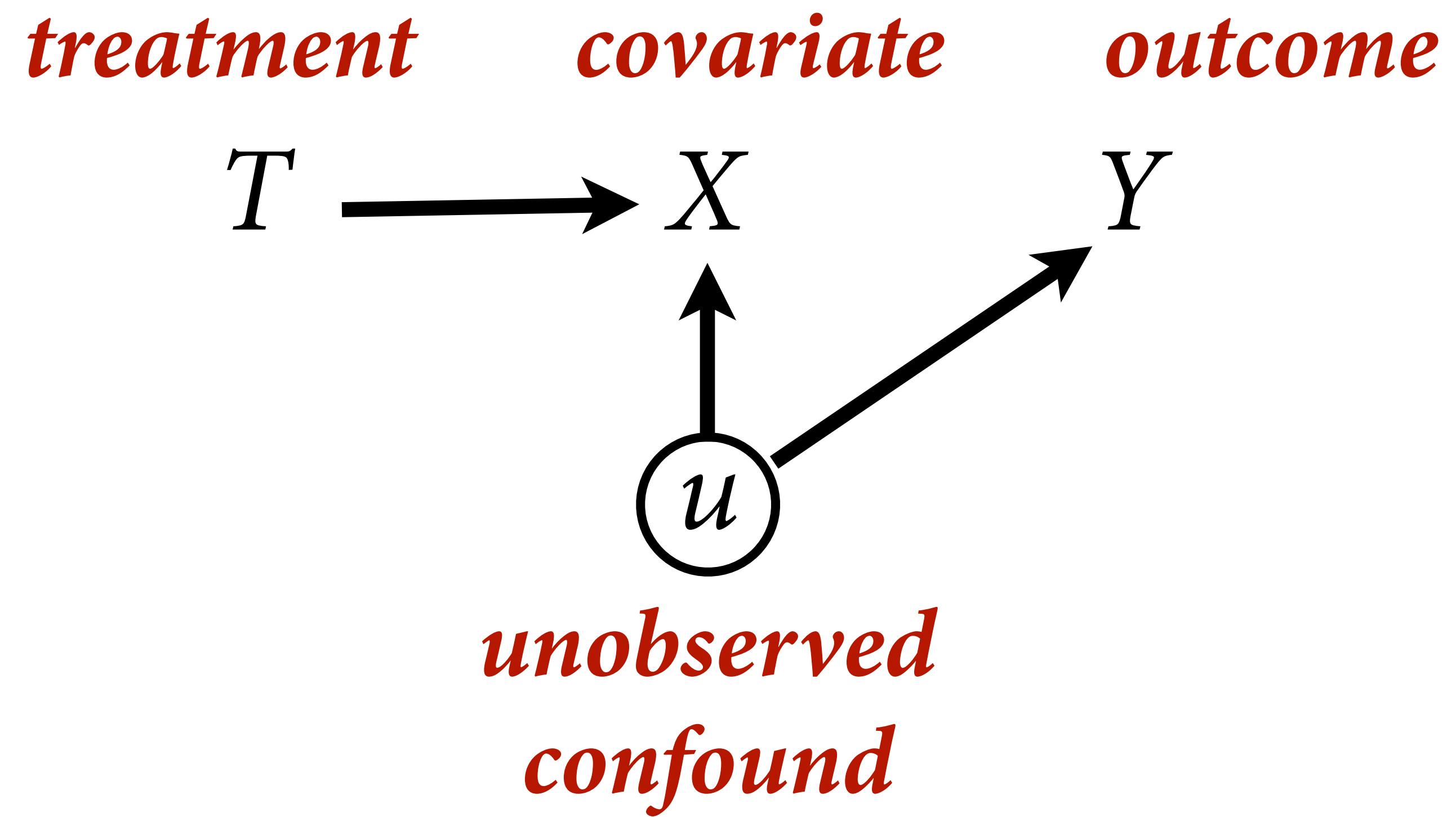
From Montgomery et al 2018 “How Conditioning on Posttreatment Variables Can Ruin Your Experiment and What to Do about It”



Regression with confounds



Regression with post-treatment variables



Ye Olde Causal Alchemy

The Four Elemental Confounds



The Fork



The Collider



The Pipe



The Descendant



The Collider

$$X \rightarrow Z \leftarrow Y$$

Z is a “collider”

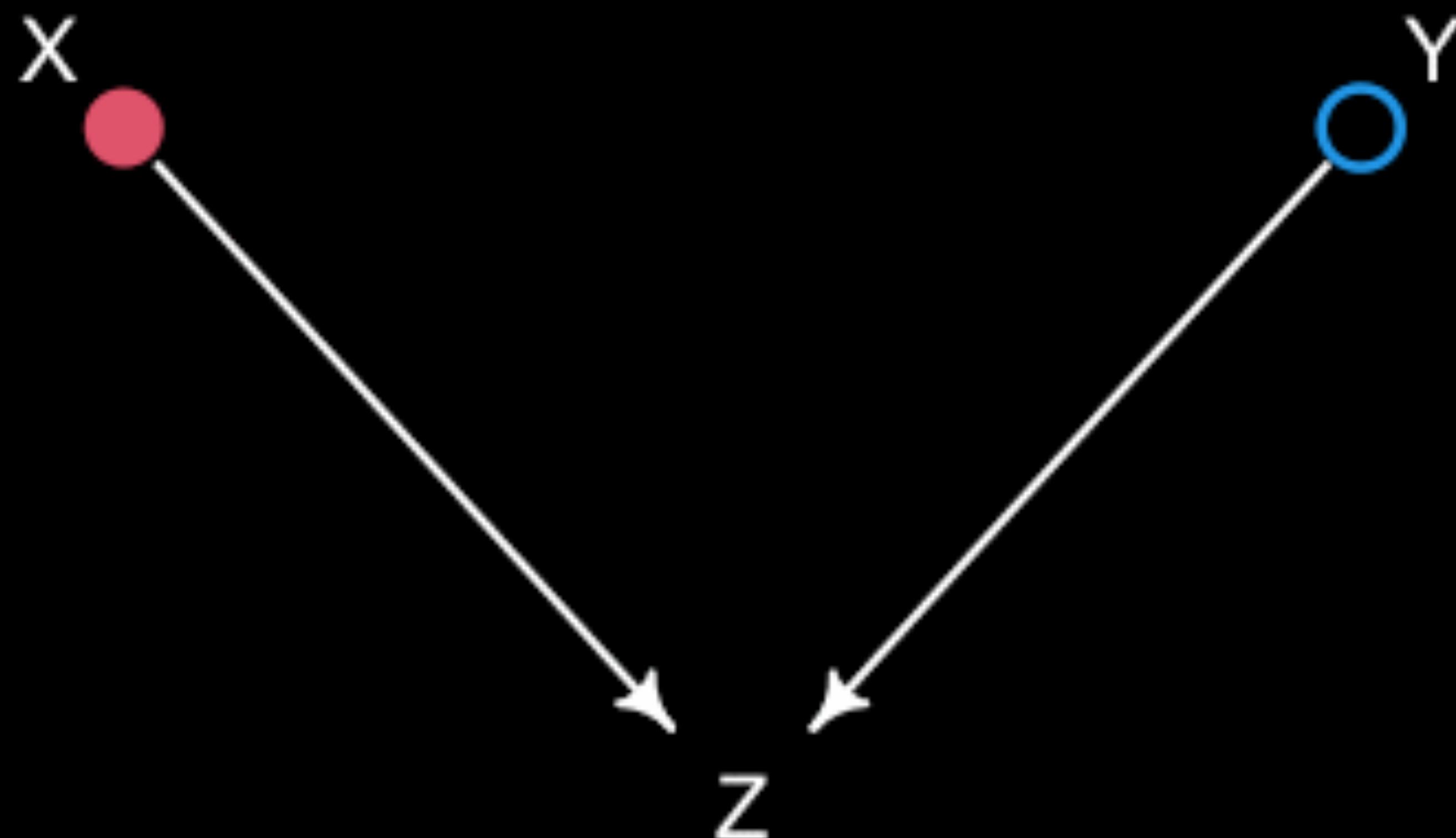
X and Y are not associated (share no causes)

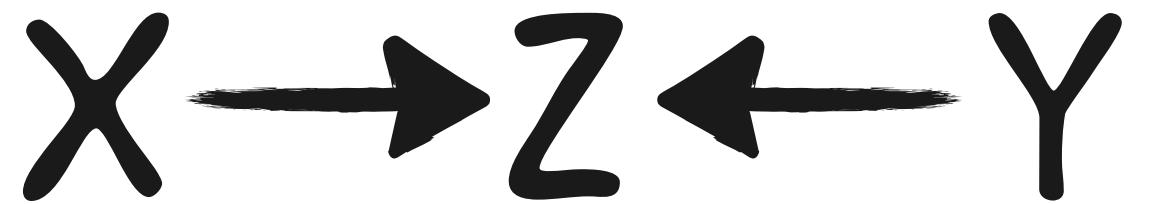
$$Y \perp\!\!\!\perp X$$

X and Y both influence Z

Once stratified by Z , X and Y associated

$$Y \not\perp\!\!\!\perp X \mid Z$$



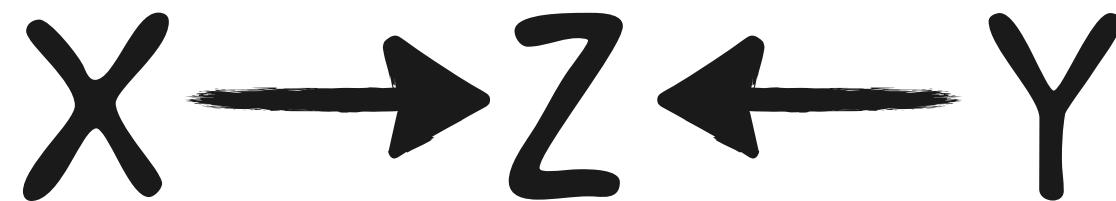


```
n <- 1000
X <- rbern( n , 0.5 )
Y <- rbern( n , 0.5 )
Z <- rbern( n , ifelse(X+Y>0,0.9,0.2) )
```

	Y	
X	0	1
0	243	236
1	250	271

$$Y \perp\!\!\!\perp X$$

```
> cor(X,Y)
[1] 0.027
```



```
n <- 1000
X <- rbern( n , 0.5 )
Y <- rbern( n , 0.5 )
Z <- rbern( n , ifelse(X+Y>0,0.9,0.2) )
```

	Y	
X	0	1
0	243	236
1	250	271

```
> cor(X,Y)
[1] 0.027
```

$Y \perp\!\!\!\perp X$

z = 0		
	Y	
X	0	1
0	200	19
1	32	29

z = 1		
	Y	
X	0	1
0	43	217
1	218	242

$Y \not\perp\!\!\!\perp X | Z$

```
> cor(X[z==0],Y[z==0])
[1] 0.43
> cor(X[z==1],Y[z==1])
[1] -0.31
```

$$X \rightarrow Z \leftarrow Y$$

```

cols <- c(4,2)

N <- 300
X <- rnorm(N)
Y <- rnorm(N)
Z <- rbern(N,inv_logit(2*X+2*Y-2))

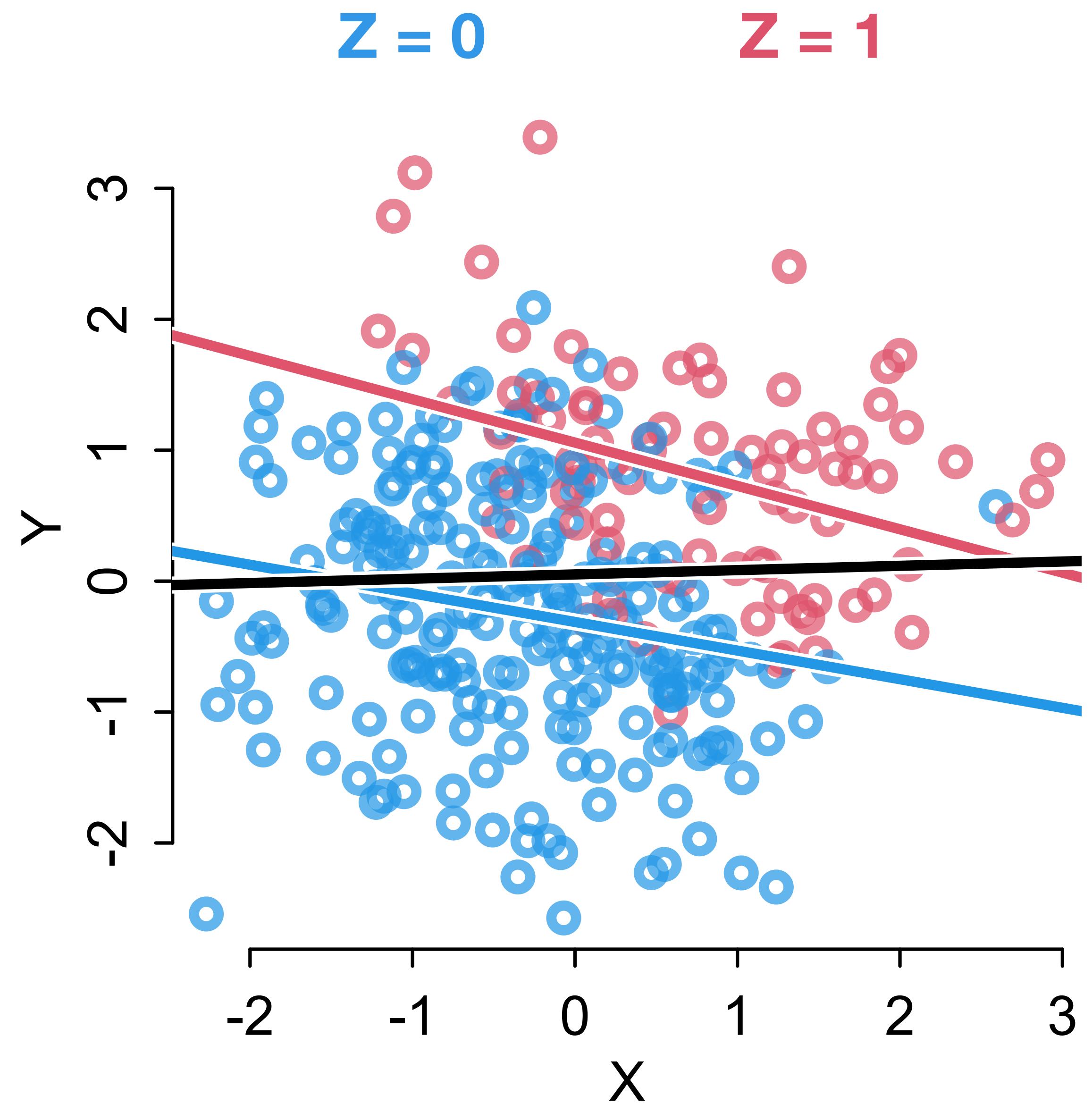
plot( X , Y , col=cols[Z+1] , lwd=3 )

abline(lm(Y[Z==1]~X[Z==1]),col=2,lwd=3)

abline(lm(Y[Z==0]~X[Z==0]),col=4,lwd=3)

abline(lm(Y~X),lwd=3)

```



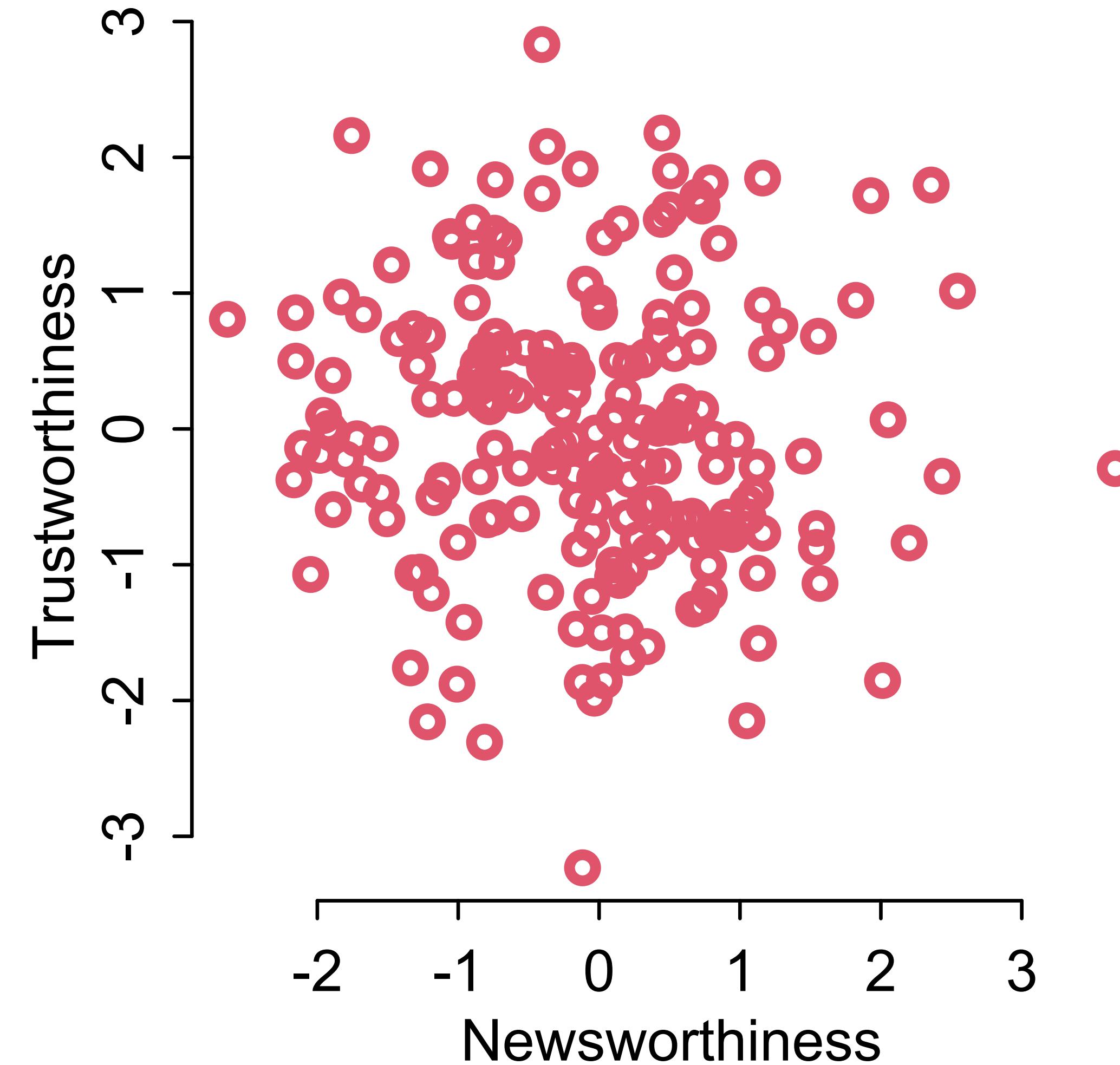
Collider example

Some biases arise from selection

Suppose: 200 grant applications

Each scored on newsworthiness
and trustworthiness

No association in population



Collider example

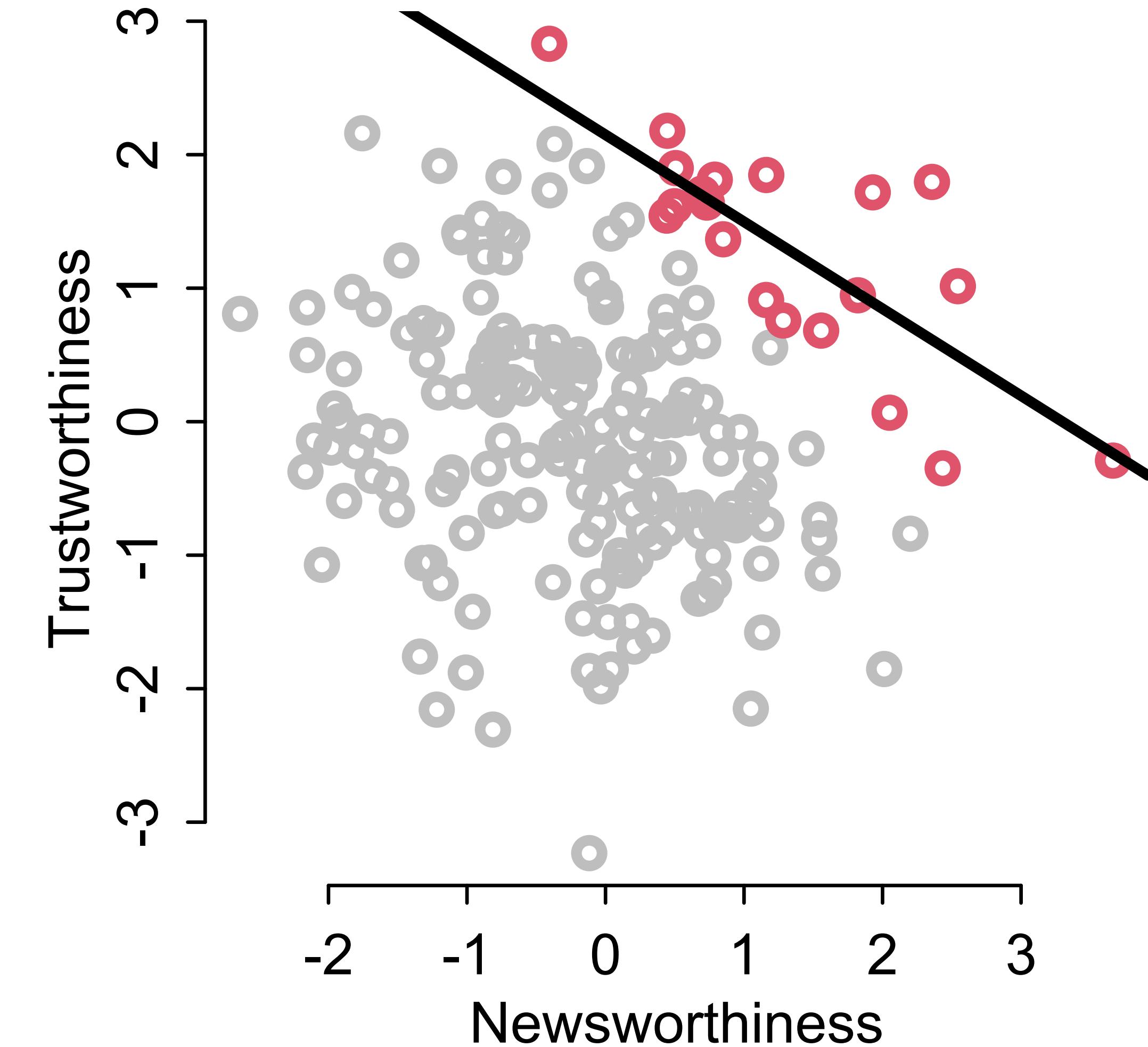
Some biases arise from selection

Suppose: 200 grant applications

Each scored on newsworthiness
and trustworthiness

No association in population

Strong association after selection



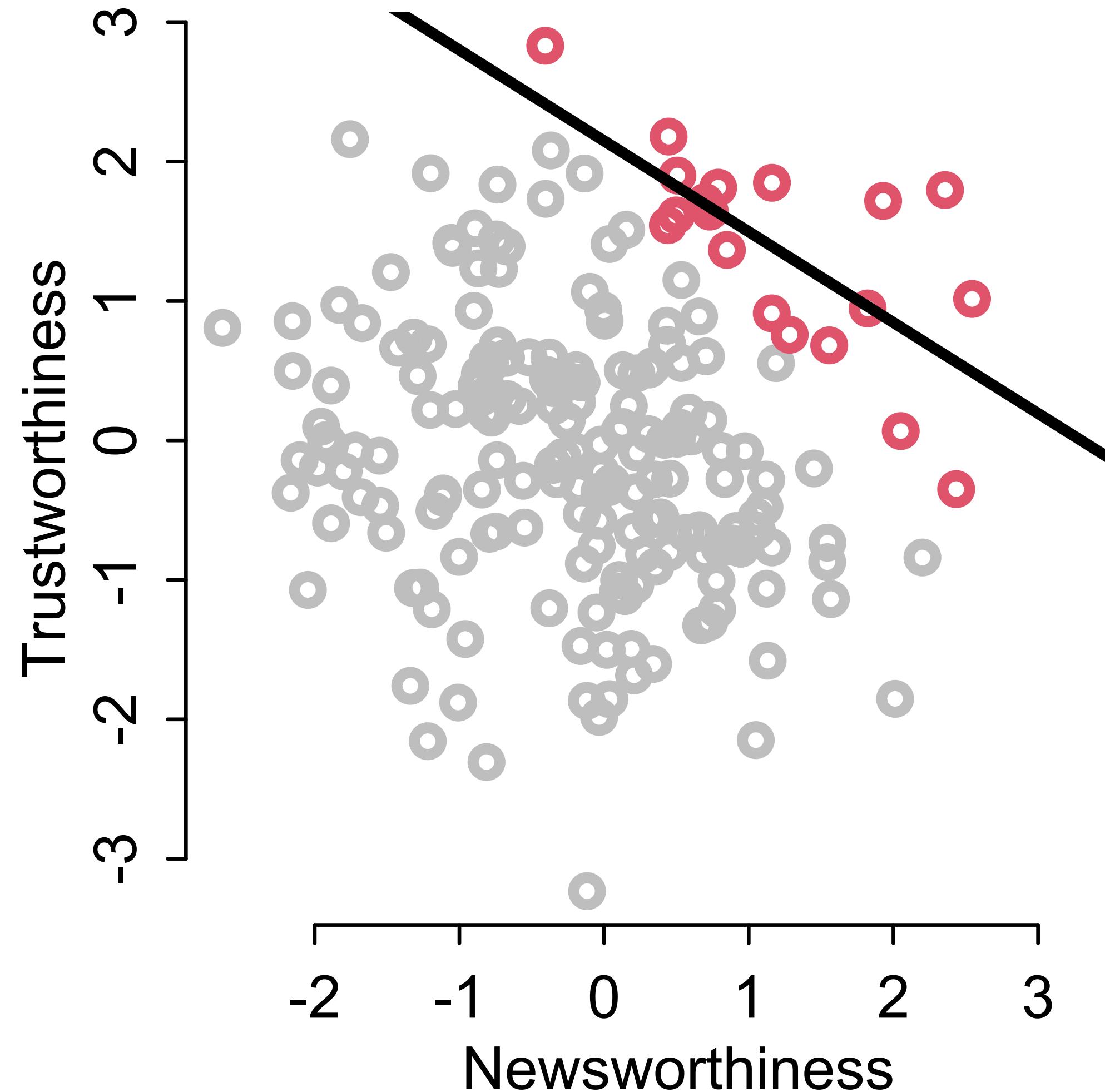
Collider example

$$N \rightarrow A \leftarrow T$$

Awarded grants must have been sufficiently **newsworthy** or **trustworthy**

Few grants are high in both

Results in **negative** association, conditioning on award



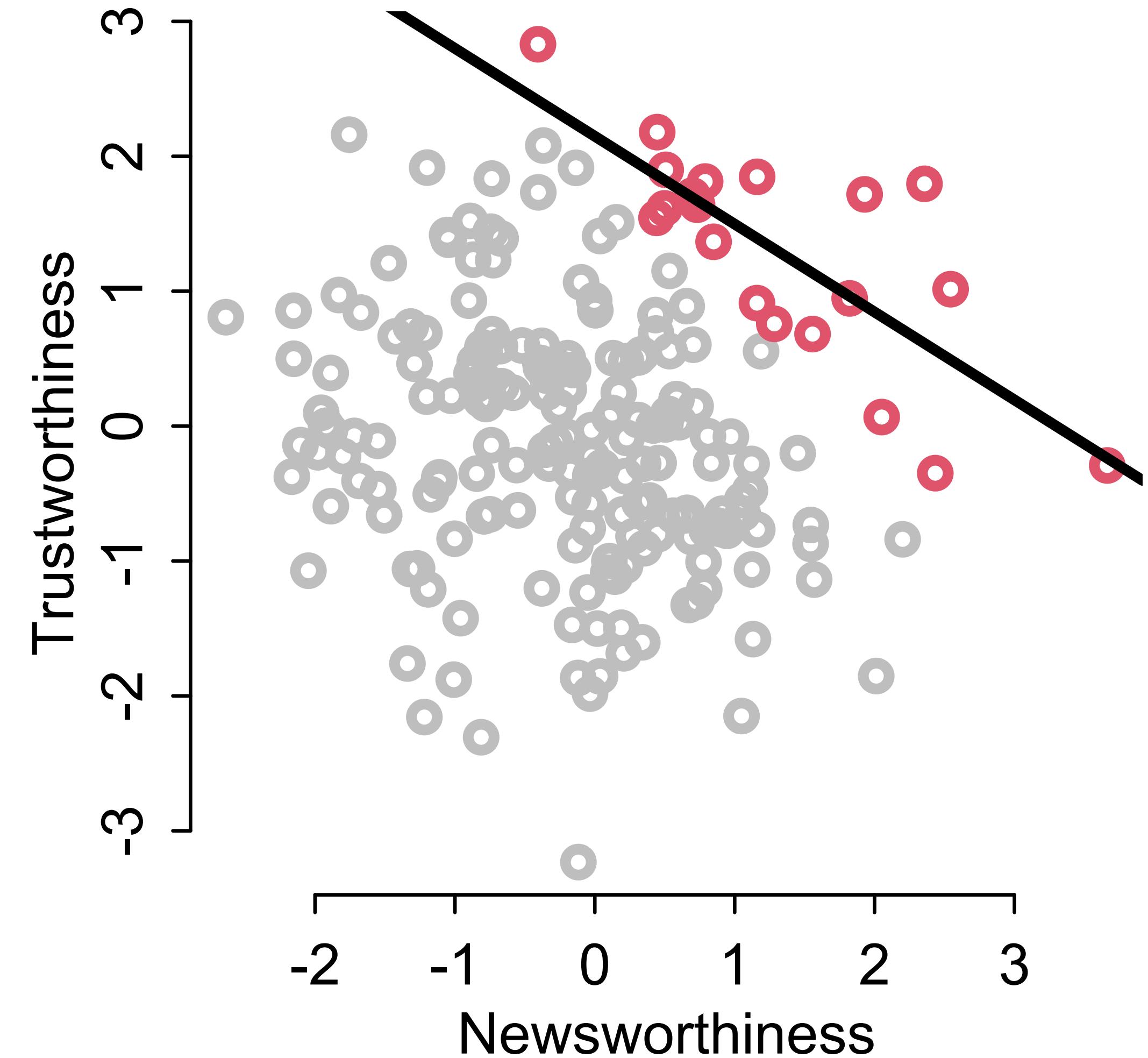
Collider example

$$N \rightarrow A \leftarrow T$$

Similar examples:

Restaurants survive by having good food or a good location => bad food in good locations

Actors can succeed by being attractive or by being skilled => attractive actors are less skilled



Endogenous Colliders

Collider bias can arise through statistical processing

Endogenous selection: If you condition on (stratify by) a collider, creates phantom non-causal associations

Example: Does age influence happiness?



Age and Happiness

Estimand: Influence of age on happiness

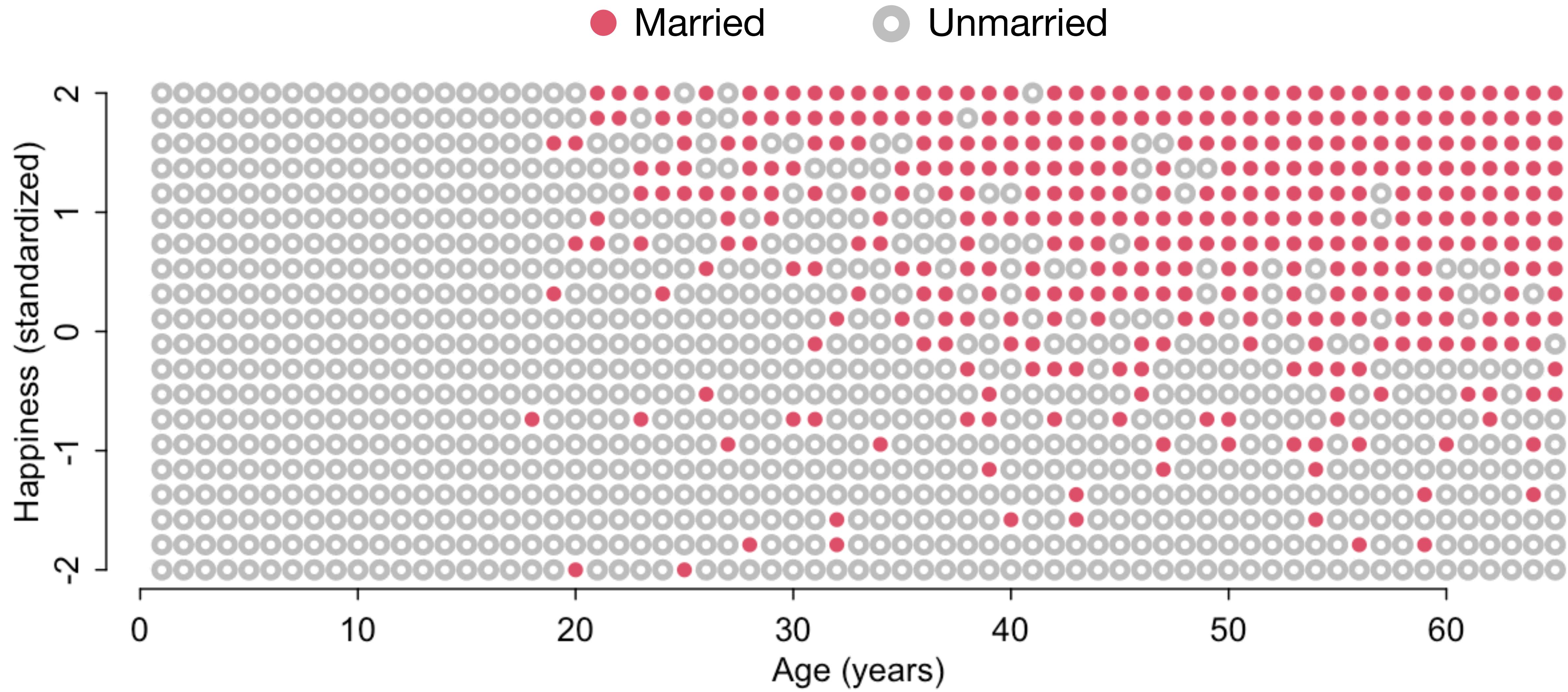
$$H \rightarrow M \leftarrow A$$

Possible confound: Marital status

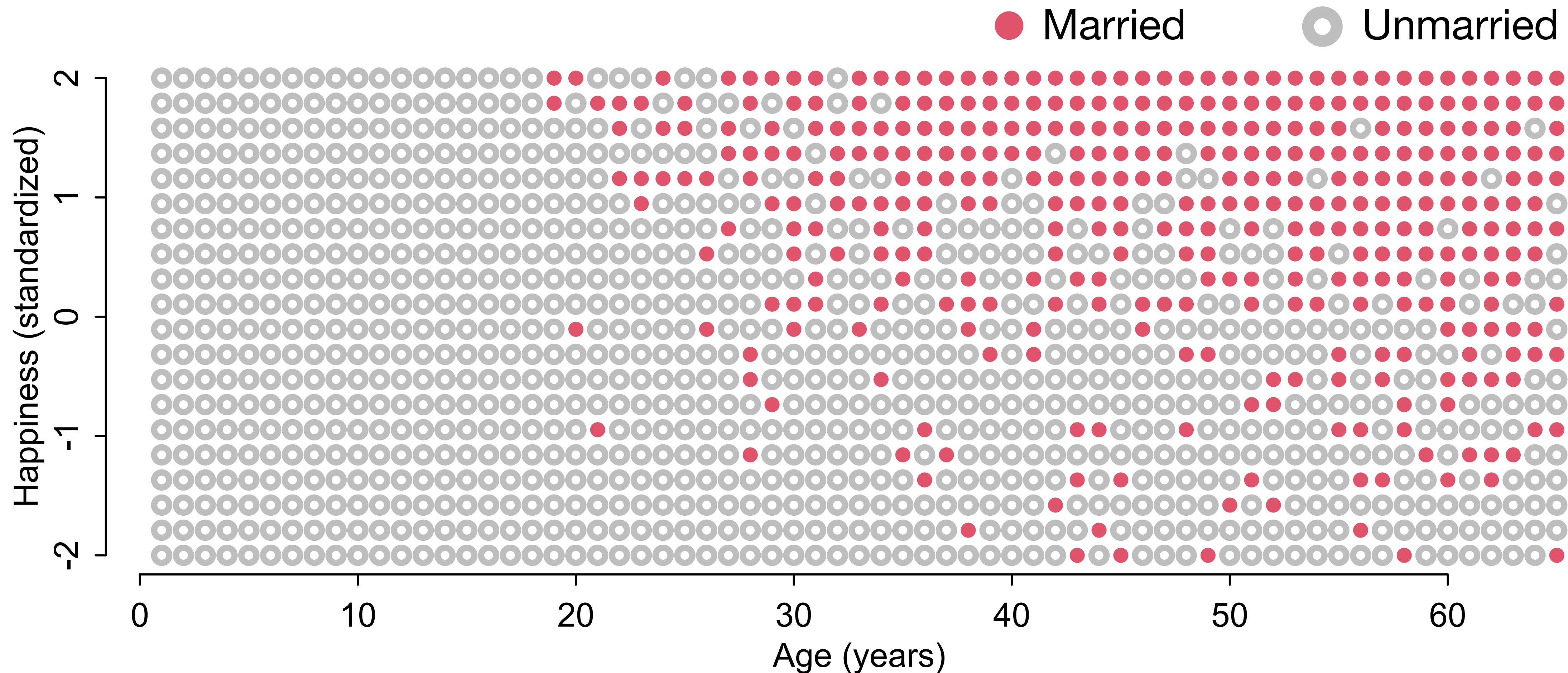
Suppose age has zero influence on happiness

Happiness
Married
Age

But that both age and happiness influence marital status



Stratified by marital status, negative
association between age and happiness



Full workflow starting on page 176

Ye Olde Causal Alchemy

The Four Elemental Confounds



The Fork



The Collider



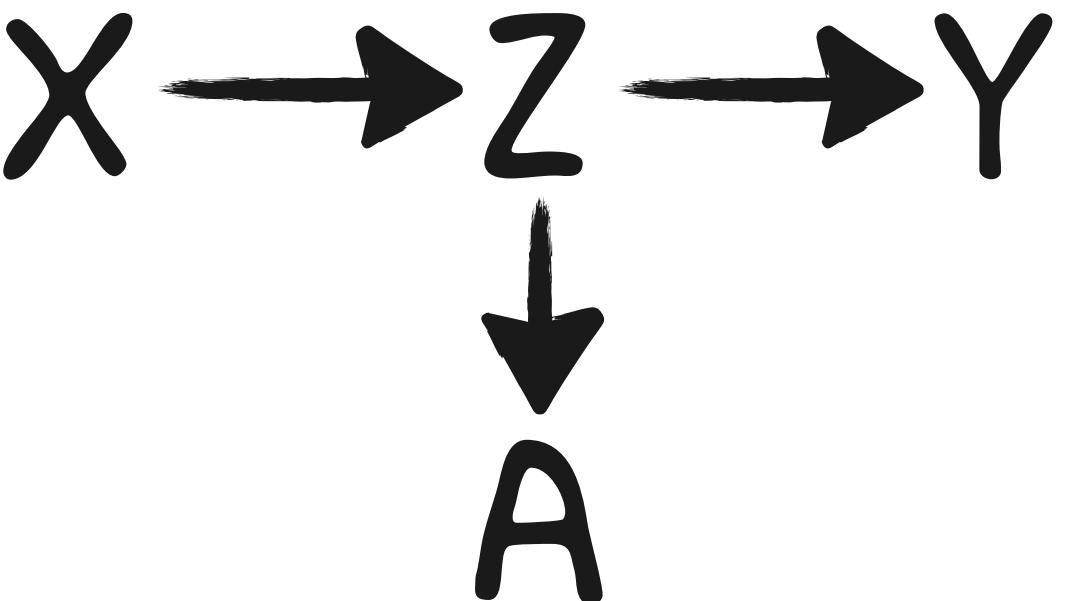
The Pipe



The Descendant



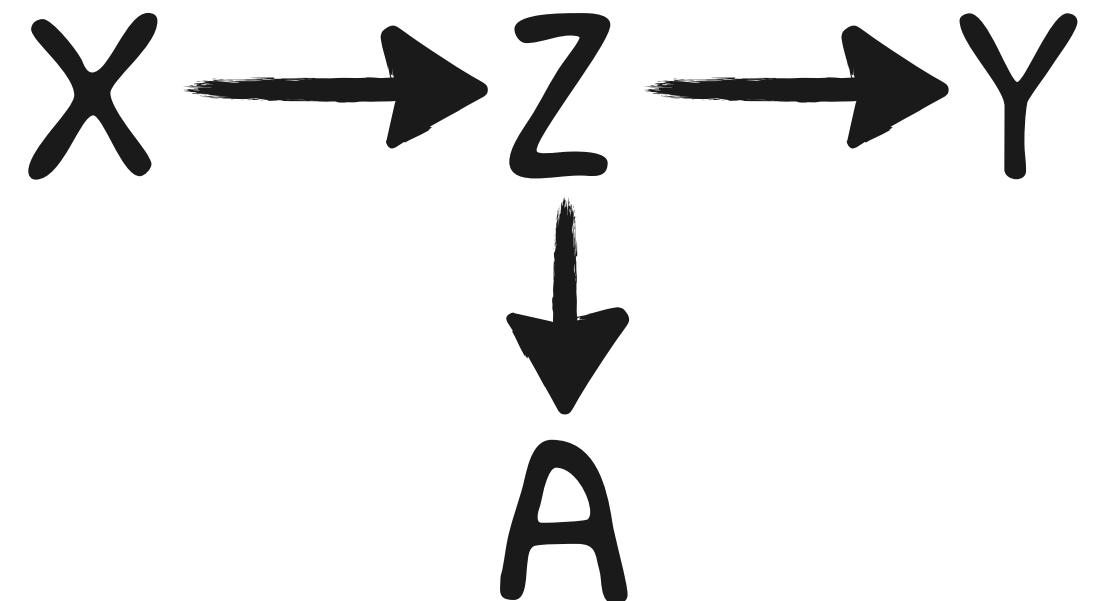
The Descendant



A is a “descendant”

How a descendant behaves
depends upon what it is
attached to

The Descendant



A is a “descendant”

X and Y are causally associated through Z

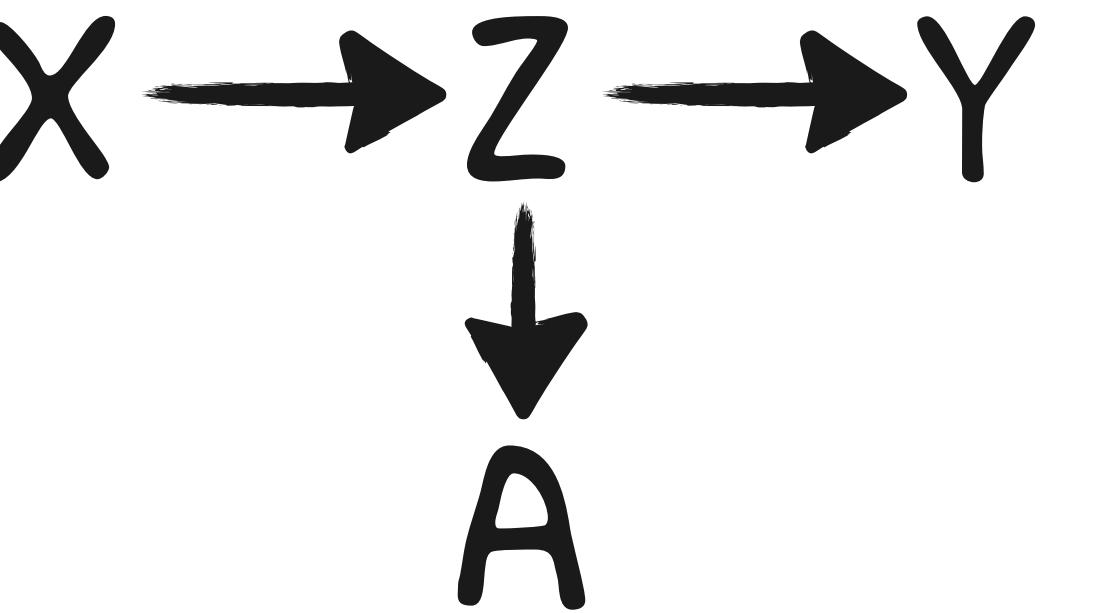
$Y \not\perp\!\!\!\perp X$

A holds information about Z

if strong enough

Once stratified by A, X and Y less associated

$Y \perp\!\!\!\perp X \mid A$



```

n <- 1000
X <- rbern( n , 0.5 )
Z <- rbern( n , (1-X)*0.1 + X*0.9 )
Y <- rbern( n , (1-Z)*0.1 + Z*0.9 )
A <- rbern( n , (1-Z)*0.1 + Z*0.9 )

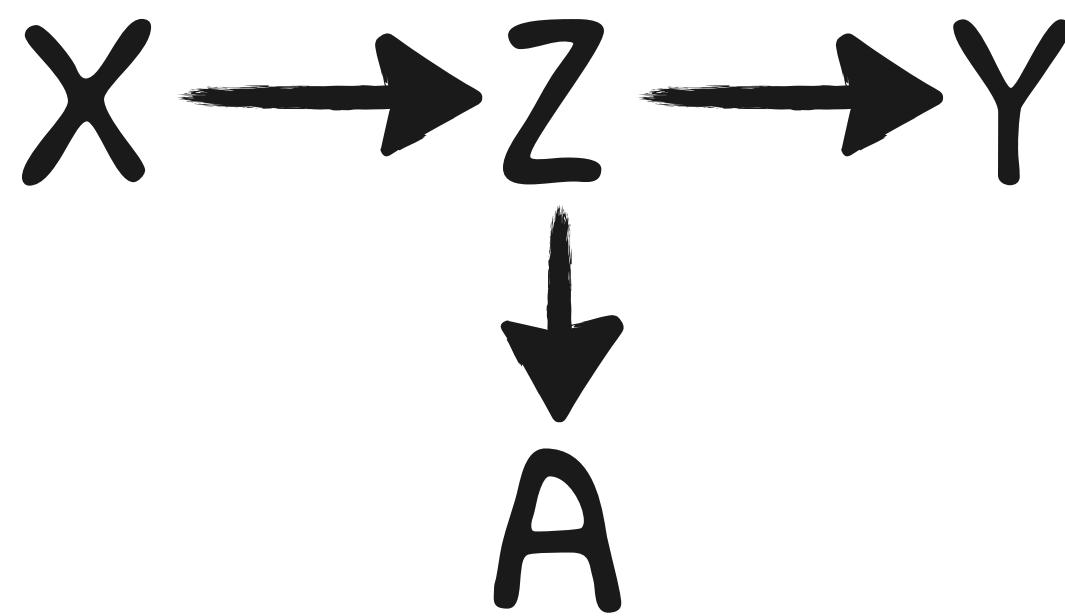
```

		Y
X		0 1
0	418 97	
1	98 387	$Y \not\perp X$

```

> cor(X,Y)
[1] 0.61

```



```

n <- 1000
X <- rbern( n , 0.5 )
Z <- rbern( n , (1-X)*0.1 + X*0.9 )
Y <- rbern( n , (1-Z)*0.1 + Z*0.9 )
A <- rbern( n , (1-Z)*0.1 + Z*0.9 )

```

	Y	
X	0	1
0	418	97
1	98	387

```

> cor(X,Y)
[1] 0.61

```

$Y \not\perp\!\!\!\perp X$

A = 0		
	Y	
X	0	1
0	387	54
1	50	32

if strong enough
 $Y \perp\!\!\!\perp X | Z$

A = 1		
	Y	
X	0	1
0	31	43
1	48	355

```

> cor(X[A==0],Y[A==0])
[1] 0.26
> cor(X[A==1],Y[A==1])
[1] 0.29

```

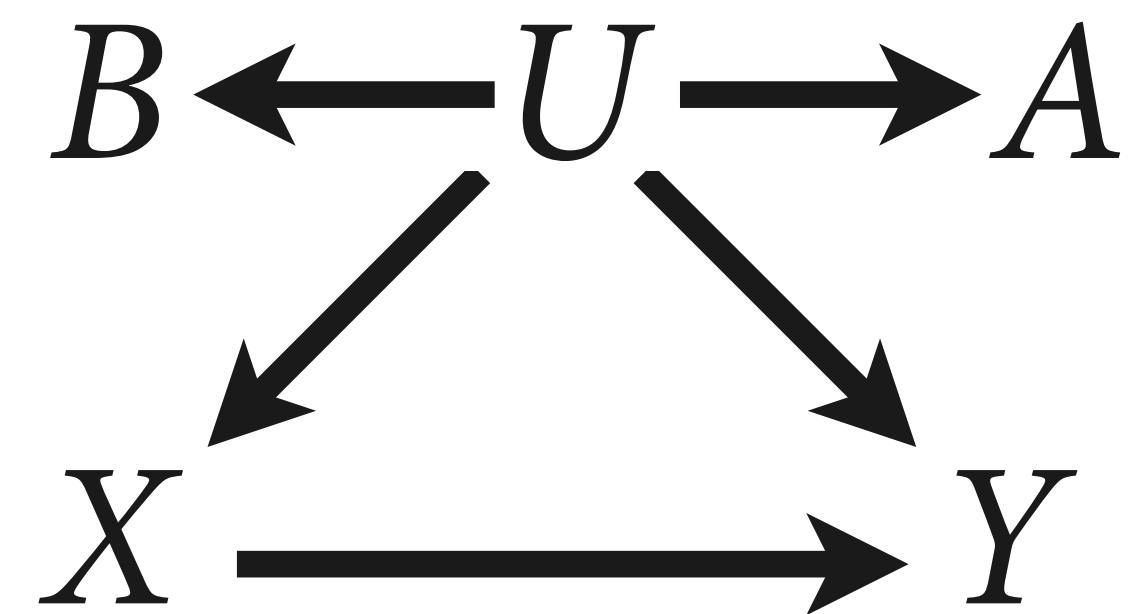
Descendants are everywhere

Many measurements are **proxies** of what we want to measure

Factor analysis

Measurement error

Social networks



U: Unobserved confound

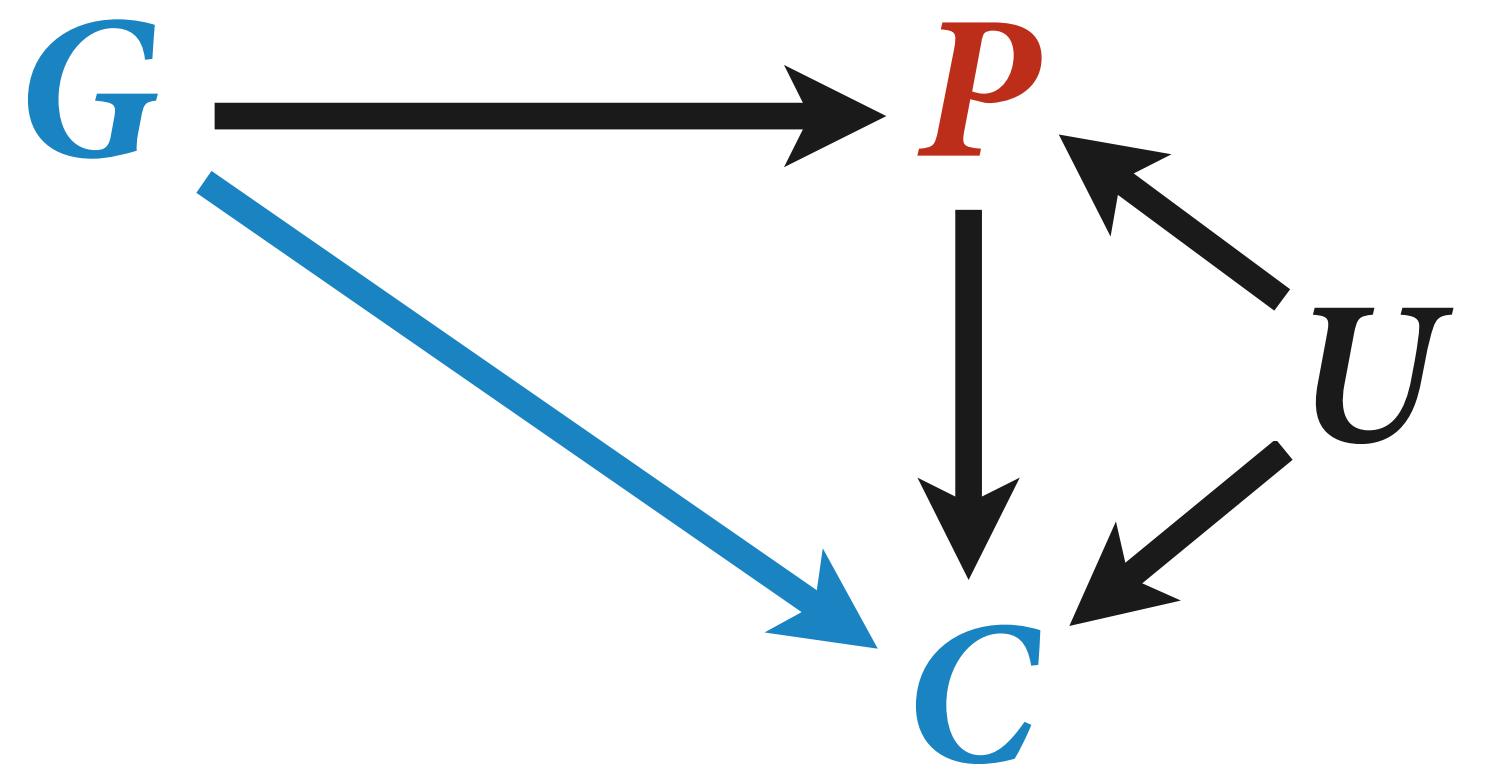
Unobserved Confounds

Unmeasured causes (U) exist and can ruin your day

Estimand: **Direct** effect of grandparents G on grandchildren C

Need to block pipe $G \rightarrow P \rightarrow C$

What happens when we condition on P ?



Course Schedule

Week 1	Bayesian inference	Chapters 1, 2, 3
Week 2	Linear models & Causal Inference	Chapter 4
Week 3	Causes, Confounds & Colliders	Chapters 5 & 6
Week 4	Overfitting / MCMC	Chapters 7, 8, 9
Week 5	Generalized Linear Models	Chapters 10, 11
Week 6	Integers & Other Monsters	Chapters 11 & 12
Week 7	Multilevel models I	Chapter 13
Week 8	Multilevel models II	Chapter 14
Week 9	Measurement & Missingness	Chapter 15
Week 10	Generalized Linear Madness	Chapter 16

